Genetic diversity of the Organic Cation Transporter 1 gene within the Cape Coloured Population

A thesis submitted in partial fulfilment of the requirements for the degree of Magister Scientiae in the Department of Biotechnology, University of the Western Cape

Brendon Pearce

May 2012

Supervisor: Prof. Mongi Benjeddou (PhD)

UWC 2012

Keywords

Pharmacogenetics

Polymerase chain reaction (PCR)

Organic Cation Transporter (OCT)

High-resolution Melt

Single Nucleotide polymorphism

Allele-specific PCR

Genotyping

Multiplex PCR

Internal validation

Population Studies

Allele frequencies

Abstract

Genetic diversity of the Organic Cation Transporter 1 gene within the Cape Coloured Population

B. Pearce

MSc Thesis, Department of Biotechnology, University of the Western Cape

The aim of this study was to investigate the genetic diversity of the SLC22A1 gene and to deduce its possible pharmacogenetic implications within the Cape Coloured population of South Africa; a uniquely admixed population of immigrant Europeans, Asians and the indigenous populations. Recent studies have reported an abundance of polymorphic variants within this solute carrier transporter gene encoding for the organic cation transporter 1, as well as evidence linking these variants to an effect on metformin uptake.

This study included establishing baseline frequency distribution of previously reported alleles for 20 SNP variants within the SLC22A1 gene, as well as the development of SNaPshot® and Multiplex AS-PCR genotyping assays, and also exploring the possibility of using High-resolution melt (HRM) analysis as a cost-effective alternative for SNP genotyping.

Ethics clearance was obtained from the Ethics Committee of the University of the Western Cape. Biological samples in the form of buccal (oral) swabs were collected from 132 unrelated voluntary donors from the Cape Coloured population residing in the Cape Metropolitan area.

Two SNaPshot® Multiplex Systems were specifically designed for the study, successfully optimized and used for genotyping. Hundred genetic profiles were then

generated for a total of 20 SNP variants on SLC22A1 gene, using this primer extension-based genotyping method that enables multiplexing up 10 SNPs.

Population genetics data obtained for the investigated SNPs were analysed using various statistical analysis software. Important population genetic parameters were calculated, and possible pharmacogenetics implications were then discussed. Among others, allelic and genotypic frequencies, as well as linkage disequilibrium were determined and compared with world populations. Minor deviation from Hardy-Weinberg equilibrium was observed in the Cape Coloured population. No significant Linkage Disequilibrium between the investigated SNPs was observed in this population.

A Multiplex allele specific – PCR (MAS-PCR) genotyping system was successfully designed and optimized for the genotyping of 10 SNPs from the SLC22A1. This system, also developed specifically for this study, was made of 2 multiplexes each covering 5 SNPs. It is an inexpensive genotyping assay that allows for efficient discrimination of SNP polymorphisms in one reaction tube with standard PCR conditions.

A pilot study was conducted to explore the possibility of using High-resolution melt (HRM) analysis as a cost-effective alternative for SNP genotyping. In addition to genotyping, HRM analysis can be used to scan large numbers of samples for novel genetic variations.

Declaration

I declare that 'Genetic diversity of the Organic Cation Transporter 1 gene within the Cape Coloured Population' is my own work that has not been submitted for any degree or examination in any other university and that all the sources I have used have been indicated and acknowledged by complete references.

Full Name: Brendon Pearce

Signature:

Date: 04/05/2012

Acknowledgements

First and foremost I would like to thank my Creator for granting me the opportunity and then helping me fulfil a dream of doing research in Pharmacogenetics.

To my supervisor, Prof. Benjeddou, thank you for always being readily available to help me understand and accept the complexities of being a Masters student. To the rest of the technical staff members, thank you for making me laugh and listening to me especially those times when I sounded brainless.

To the National Research Foundation and the Medical Research Council, I thank you for funding this research project.

To my family and friends, thank you for keeping me grounded, motivated and inspired.

List of Abbreviation

DNA	Deoxyribonucleic acid
RFLP	Restriction fragment length polymorphism
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
PCR	Polymerase chain reaction
RFU	Relative fluorescence units
HRM	High-Resolution Melt
SLC	Solute Carrier Transporter
OCT	Organic Cation Transporter
SAP	Shrimp Alkaline Phosphatase
EXO I	Exonuclease I
HWE	Hardy-Weinberg equilibrium
Ho	Observed heterozygosity
He	Expected heterozygosity

List of Tables

Table 3.3

Minor allele frequencies (MAF) of single nucleotide polymorphisms in Cape Coloured (C), African American (AA), Caucasian American (CA), Asian (A), Mexican (ME), Pacific Islander (P) and Sub-Saharan African (SSA). Minor allele frequencies for populations other than the Cape Coloured population were taken from the Pharmacogenomics Knowledgebase (http://www.pharmgkb.org), the UCSF-PMT (http://www.pharmacogenetics.ucsf.edu/) database and the NCBI-SNP database (http://www.ncbi.nlm.nih.gov)

Table 3.4

Summary of Heterozygosity and Fixation Indices generated using GenAlEx

Table 3.5

Hardy-Weinberg Chi-Square test

Table 3.6

Population differentiation between Cape Coloured and global populations. Significant data is shown only. All monomorphic or absent data has been removed. Where C – Cape Coloured, CA - Caucasian American, AA – African American, A – Asian, ME – Mexican and P – Pacific Islander

Table 4.3

HRM primer sequences for newly reported Exon 4 SNPs

List of Figures

Figure 2.3

Multiplex one genotyping result on 2.5% agarose gel.

Figure 2.4

Multiplex two genotyping results on 2.5% agarose gel.

Figure 2.5

Multiplex one genotyping result on 1.8% agarose gel.

Figure 2.6

SNaPshot® validation of the SNPs genotyped by MAS-PCR.

Figure 2.7

Genotyping results for multiplex 1 used for the blind test.

Figure 2.8

Genotyping result for Multiplex 2.

Figure 3.1:

Typical Multiplex 1 electropherogram. Note a shift of 12bp and 20bp in SNP1 and SNP3 respectively.

Figure 3.2:

Multiplex 2 electropherogram. A large peak at position 58 to 65 for the A base was observed.

Figure 3.3

Minor allele frequencies across various populations.

Figure 3.4

LD map of the genotype frequencies for all 20 SNPs.

Figure 3.5

LD map of the 10 heterozygous SNP markers.

Figure 3.6

LD map of the 13 heterozygous SNP markers under Confidence Bounds setting.

Figure 4.1

Raw fluorescence plots for each of the SNP amplifications. A: RS34163122, 6 reactions displayed late amplification where as one displayed no amplification. B: RS35446461, only one reaction displayed no amplification.

Figure 4.2

RS34163122 HRM mutation screen. Alleles 1 and 2 are indicated at the target site. Note a secondary target marked *X*.

Figure 4.3

Normalised plot (A) and difference plot (B) for RS34163122 using the wild type (WT) genotype as a reference. Note the larger divergence depicted by the mutant (Mut) in blue.

Figure 4.4

RS35446461 HRM mutation scan. A single target peak is observed. This SNP was determined to be homozygous in previous chapters in this study.

Figure 4.5

Normalised plot (A) and difference plot (B) for RS3544646. All samples exhibit wild type genotype (Alelle 1 homozygous).

Figure 4.6

Clustered data for RS54163122 using ScreenClust. Note the samples falling inside of the blue cluster. These samples underwent incomplete amplification resulting in the formation of non-specific or incomplete products.

Figure 4.7

Clustered data for RS35446461 using ScreenClust. Note only one sample appears in the red cluster as it did not amplify at all.

Table of Contents

Chapter 1 – Literature Review

Chapter One

Literature Review

1.1. Introduction

Human genetics commonly refers to the study of gene inheritance patterns occurring in human beings. Human genetics includes a multitude of overlapping fields. These include classical genetics, cytogenetics, molecular genetics, biochemical genetics, genomics, population genetics, developmental genetics, clinical genetics, and genetic counselling (Nussbaum *et al*, 2007). Studying human genetics is useful in that it can provide possible solutions to questions about human nature. From this we might also elucidate diseases, and develop effective disease treatment, and understand the genetics and mechanics of human life. (Nussbaum *et al*, 2007)

1.1.1. Major forms of Human genetic variation

Genetic variation has been defined as the existence of two or more genetic variants at a considerable occurrence in a specific population. This includes alleles, phenotypes, sequence variants and chromosomal structure variants (Strachan & Read, 2004). Human genetic variation may be subdivided into two major forms, namely sequence polymorphisms and length polymorphisms. Sequence polymorphisms are defined as a change in one or more bases in the DNA sequence at a specific locus. Length polymorphisms are defined as variations in the length of a repetitive DNA sequence (Rudin & Inman, 2002; Butler, 2005). Many techniques have been developed to test for and characterise DNA polymorphisms, including restriction fragment length

polymorphisms (RFLP) analysis, single nucleotide polymorphism (SNP) analysis, short tandem repeat sequences (STRs) and mitochondrial DNA (mtDNA) sequencing.

1.1.2. Single Nucleotide Polymorphisms

Single nucleotide polymorphisms (SNPs) are the most abundant form of genetic variation in the human genome, with a total of approximately 12 million SNPs identified (Brockmoller and Tzvetkov, 2008). SNPs are generally used as markers to aid identification of the genes underlying complex diseases (Shastry, 2007; Caulfield, 2001). As the name suggests, the site of variation lies within a single nucleotide which may undergo specific mutations termed base substitutions in which one nucleotide is replaced by another. Two forms of base substitutions can occur, transition substitutions or transversion substitutions. Transition substitution occurs between either two purines, Adenosine (A) and Guanine (G), or between two pyrimidines, Thymine (T) and Cytosine (C) and is also thought to have resulted in two thirds of all SNPs in the human genome. Transversion substitutions occur when shift from a purine to pyrimide, or vice versa, is evident, such as an A to T shift. (Smith, 2002)

It is accepted that most SNPs are not thought to impact of gene function. However, SNPs found within coding regions (cSNPs) or within regulatory regions are considered to be significantly more likely to result in functional changes (Gaudet *et al*., 2007). cSNPs found within coding regions are defined as being either synonymous or non-synonymous. Synonymous polymorphisms alter the coding sequence of the coding region, but that alteration does not result in a change in the amino acid encoded by that region. It is thus said to be a "silent" polymorphism. Non-synonymous polymorphisms, however, alter the coding sequence and thereby a codon

change. The result of this is a change in amino acid coded by that region. These are referred to as missense mutations and form the basis for genome studies (Gaudet *et al*., 2007). Therefore, detecting SNPs which alter gene function has become critically important to a broad spectrum of biological and medical disciplines (Benjedoou, 2010). Moreover, because SNPs are the most abundant genetic markers in the human genome, they have become useful for genome evolution studies, high density genetic maps, disease diagnostics, association studies and marker-assisted selection (Gaudet *et al*, 2007).

1.2. Interindividual Variability and Membrane transporters

It has been established that genetic interindividual variability, genetic variation between members the same population, plays a significant role in drug response and toxicity (Choi & Song, 2008). These variations may be as a result of numerous factors; which include genetic, environmental, physiological and pathophysiological (Eichelbaum *et al*, 2006). It is estimated that genetic factor may account for as much as 95% of interindividual variability in drug response and effects (Eichelbaum *et al*, 2006; Evans & Relling, 2004; Weinshilboum, 2003).

A prodigious expanse of interindividual variability; resulting from genetic, environmental, physiological and pathophysiological factors, exists in drug response and toxicity (Choi and Song, 2008). It is estimated that between 15 and 30% of interindividual variation in drug disposition and response is a result of genetic factors. Novel genetic variations in drug transporters, drug targets, effector proteins and metabolizing enzymes, have been discovered over the last decade as a result of the improvement of genotyping assays and technologies (Avery *et al*, 2009).

Recently, much attention has been focused on the genetic variation of drug transporters. This is a result of the roles these transporters play in gastrointestinal absorption, biliary and renal elimination and distribution to target sites of their specific substrates (Choi and Song, 2008). Moreover, emerging pharmacogenetic evidence suggests that membrane transporters are subject to both genotypic and phenotypic variation. Additionally, it is also suggested that variation in drug transporters may bet he underlying cause of interindividual variability in pharmacokinetic disposition, efficacy and toxicity of drug transporter substrates (Choi and Song, 2008).

Variations in drug transporters are thus believably responsible for, to a great extent, the interindividual variability in pharmacokinetic disposition, efficiency, and toxicity of drug transporter substrates (Choi & Song, 2008). The organic cation transporter (OCT) family contains three subtypes of facilitated transporters called OCT1, encoded by SLC22A1, OCT2 (SLC22A2) and OCT3 (SLC22A3) (Koepsell *et al*, 2007).

1.2.1. Drug Transporters

Membrane transporters are crucial for the maintenance of cellular and organismal homeostasis and act by importing essential nutrients for cellular metabolism and by exporting waste products and toxic compounds (Leabman *et al*, 2003). Furthermore, membrane transporters are also important in drug response as they are major determinants of drug absorption, distribution and elimination, as well as providing the targets for many commonly used drugs (Leabman *et al*, 2003). The ATP-binding

Cassette (ABC) transporters and Solute Carrier (SLC) transporters are the two major superfamilies of membrane transport proteins (Leabman *et al*, 2003). The ABC transporters include MDR-1 protein, responsible for the pumping of xenobiotics from cells, while the SLC transporters are responsible for the uptake and importing of neurotransmitters, nutrients, heavy metals and other substrates into cells (Leabman *et al*, 2003; Benjeddou, 2010)

1.2.2 Genetic Polymorphisms of Solute Carrier Transporters

SLCs are mutually regulated by exogenous and endogenous expressional factors and they are influenced by genetic polymorphisms. Genetic polymorphisms affect the expression and transport activity of these transporter genes. Many polymorphisms have been identified as the factors contributing to the interindividual variation in the disposition, distribution and response to substrate drugs (Kerb *et al*, 2002; Shu *et al*, 2003)

It has been reported, by numerous studies, that polymorphic variations exist within the SLC22A superfamily from different ethnic groups. Kerb *et al*. (2002) has reported that 25 single nucleotide polymorphisms (SNPs) have been identified within 57 Caucasian individuals in human SLC22A1. Furthermore, three of the SNPs (R61C, C88R, and G401S) showed reduced transport activities as they were non-synonymous variants. A total of 15 protein-altering variants from SLC22A1 from diverse ethnic backgrounds were reported by Shu *et al*. (2003). Further findings indicated that four SNPs (R61C, G220V, G401S, and G465R) showed reduced transport function, whereas S14F exhibited increased transport function (Shu *et al*, 2003). Two non-synonymous SNPs found in Korean and Japanese populations (P283L and P341L)

showed reduced transport activity (Sakata *et al*, 2003; Takeuchi *et al*, 2003). Conversely, some non-synonymous SNPs (S14F, S189L, P341L, M420del) showed diverse functional changes depends on the substrates examined, for example MPP+, a toxic compound implicated in cell death, and the anti-diabetic drug metformin. Therefore, depending on the substrate administered, genetic variants within Organic Cation Transporter 1 (OCT1) may also differ pharmacokinetically (Choi & Song, 2008).

Kang *et al*. (2007) identified three SNPs (M165I, R400C, K432Q) from eight amino acid changing variants of SLC22A2, which displayed reduced transport activity. Furthermore, separate studies have come to the conclusion that variants (T199I and T201M) in SLC22A2, identified in Japanese and Korean populations, showed a significant decrease in MPP＋ and metformin uptake (Leabman *et al*, 2002; Song *et al*, 2008). The A270S variant has been identified in all ethnic groups, and has the highest allele frequency. The Km for MPP＋ and Ki value for TEA inhibition of MPP＋ uptake was doubled (Leabman *et al*, 2002; Fujita *et al*, 2006). Kang *et al*. found that the intrinsic clearance of MPP＋ was decreased more than 20-fold in A270S overexpressing oocytes compared with that of the reference group. Interestingly, this indicated that the Vmax of Organic Cation Transporter 2 (OCT2), rather than binding affinity itself, was more greatly affected by this variant.

Conversely, in SLC22A3, several genomic variants (R120R and A411A) have been identified which do not cause amino acid substitutions (Lazar *et al*, 2003). Therefore, in relation to genetic polymorphisms, limited information is available for functional changes in Organic Cation Transporter 3 (OCT3). Therefore, our understanding of

interindividual drug responses and the pharmacokinetics of cationic drugs is aided by the information gained about the functional genetic polymorphisms of the Organic Cation Transporter family.

Table 1.1. Summary of published medically relevant SNPs

SNP ID	Family	Effect	Drug	Disease
rs622342	SLC22A1	Efficacy	metformin[a,b,d]	Diabetes Mellitus
rs1123617	SLC14A2	Efficacy	Nifedipine*	Hypertension
rs4149056	SLCO1B1	Other	Methotrexatec*	Precursor CLLL
rs11045819	SLCO1B1	Efficacy	fluvastatin[a]	Hypercholesterolemia
rs1051266	SLC19A1	Toxicity/ADR	methotrexate[c]	Psoriasis
rs1051266	SLC19A1	Efficacy	methotrexate[c]	Arthritis, Rheumatoid
rs4795541	SLC6A4	Efficacy,Toxicity/ADR	Escitalopram*	Depression,Depressive Disorder
rs4149081	SLCO1B1	Toxicity/ADR	methotrexate[c]	Precursor CLLL
rs11045879	SLCO1B1	Toxicity/ADR	methotrexate[c]	Precursor CLLL
rs2306283	SLCO1B1	Other	repaglinide[a]	
rs2306283	SLCO1B1		pravastatin[a]	
rs25531	SLC6A4	Efficacy	Fluoxetine*	Depressive Disorder, Major
rs316019	SLC22A2	Toxicity/ADR	cisplatin[a,c]	Neoplasms
rs4149056	SLCO1B1		nateglinide[a]	

rs4149056	SLCO1B1	Toxicity/ADR	Simvastatin*	Muscular Diseases,Myopathy, Central Core
rs4149056	SLCO1B1	Efficacy	repaglinide[a]	
rs4149056	SLCO1B1	Efficacy	Repaglinide[a]	
rs4149015	SLCO1B1	Efficacy	pravastatin[a]	
rs12422149	SLCO2B1	Efficacy	Montelukast*	Asthma
rs714368	SLC22A16	Dosage,Toxicity/ADR	doxorubicin,doxorubicinol*	Breast Neoplasms
rs2301159	SLC10A2	Toxicity/ADR	docetaxel,thalidomide*	Prostatic Neoplasms

Where CLLL is an abbreviation for Cell Lymphoblastic Leukemia-Lymphoma
[a] Fahrmayr,*et al*, 2010
[b] Kroetz *et al*, 2010
[c] Nakanishi & Tamai, 2011
[d] Choi and Song, 2008
*** Pharmacogenomics Knowledgebase (http://www.pharmgkb.org/gene/PA329)**

1.2.3. Solute Carrier Transporters (SLC)

The Solute Carrier Transporter superfamily is comprised of over 300 members further subdivided into 47 families. These transporters are primarily expressed in the liver, lungs, kidneys and intestines but can be found in most tissues. They are localized to either the basolateral or apical plasma membrane of polarized cells but can be expressed in mitochondria and other organelles (Wojtal *et al*, 2009; Chen *et al*, 2001; Sweet *et al*, 2001). Because they are membrane-associated transporters, they are crucial for the facilitating the passage of solutes such as peptide, amino acids, bile acids, xenobiotics, ions, drugs and other biological compounds (Hediger *et al*, 2004; Koepsell *et al*, 2007).

The SLCs are critically involved in drug absorption, primarily in the intestine, and are therefore major determinants of drug distribution and the pharmacokinetic characteristics of many drugs (Meier *et al*, 2007). The majority of transporters in the SLC22 family are polyspecific, indicating that they are specific for a myriad of substrates. They are further subdivided into three groups; organic cation transporters (OCTs), organic anion transporters (OATs) and organic zwitterion/cation transporters (OCTNs) (Meier *et al*, 2007). However, for the purpose of this study, the OCTs will be focused upon.

The OCTs, include OCT1 (SLC22A1), OCT2 (SLC22A2) and OCT3 (SLC22A3). Human OCTs (hOCT1, hOCT2, hOCT3) facilitate the electrogenic transport of small organic cations, many having different molecular structures, independent of sodium gradient. Included in these organic cation substrates are therapeutics (Metformin), endogenous compounds (dopamine) and toxic substances (TEA) (Kang *et al*, 2007;

Leabman *et al*, 2002; Fujita *et al*, 2006; Song *et al*, 2008). Clustered together on the long arm of chromosome 6 are the genes encoding the human isoforms of the three organic cation transporters (hOCT1, hOCT2, hOCT3). Furthermore, numerous SNPs have been identified in these three genes by various population genetics analyses (Tzvetkov *et al*, 2009).

1.3. Organic Cation Transporter Genes

1.3.1 Structure of the OCTs

Organised within the SLC22A superfamily, three isoforms of OCT exist; these include OCT1, OCT2 and OCT3. Similar membrane topologies are noted between homologues of this family. These consist of 12 transmembrane domains (TMDs), an intracellular N-terminus, a large glycosylated extracellular loop between TMD 1 and 2, a large intracellular loop with phosphorylation sites between TMDs 6 and 7, and an intracellular C-terminus. (Koepsell *et al*, 2007; Gorboulev *et al*, 2005; Gorboulev *et al*, 1999)

Substrate specificities have been proposed, by use of the tertiary structure model of the transmembrane helixes (TMHs) of rat Oct1 and site-directed mutagenesis of amino acids, substrate-binding cleft and critical amino acids. The substrate-binding pocket is believed to consist of three amino acids on the fourth TMH (W218, Y222, and T226), three amino acids on the tenth TMH (A443, L447, and Q448), and one amino acid on the eleventh TMH (D475). These amino acids are located in one region surrounding a large cleft which opens to the intracellular side. This region is highly conserved among the subtype of the OCT superfamily (Popp *et al*, 2005; Gorboulev *et al*, 2005; Gorboulev *et al*, 1999).

The organic cation transporter 1 (OCT1) was first cloned from rats in 1994 (Grundemann *et al*, 1994) and has subsequently been cloned from humans, rabbits, and mice (Green *et al*, 1999; Terashita *et al*, 1998; Zhang *et al*, 1997). The homologues of OCT1, OCT2 and OCT3 have also subsequently been cloned from humans, rabbits, rats, and mice (Green *et al*, 1999; Terashita *et al*, 1998; Zhang *et al*, 1997; Gorboulev *et al*, 1997; Okuda *et al*, 1996; Mooslehner & Allen, 1999; Grundemann *et al*, 1998; Kekuda *et al*, 1998; Wu *et al*, 2000).

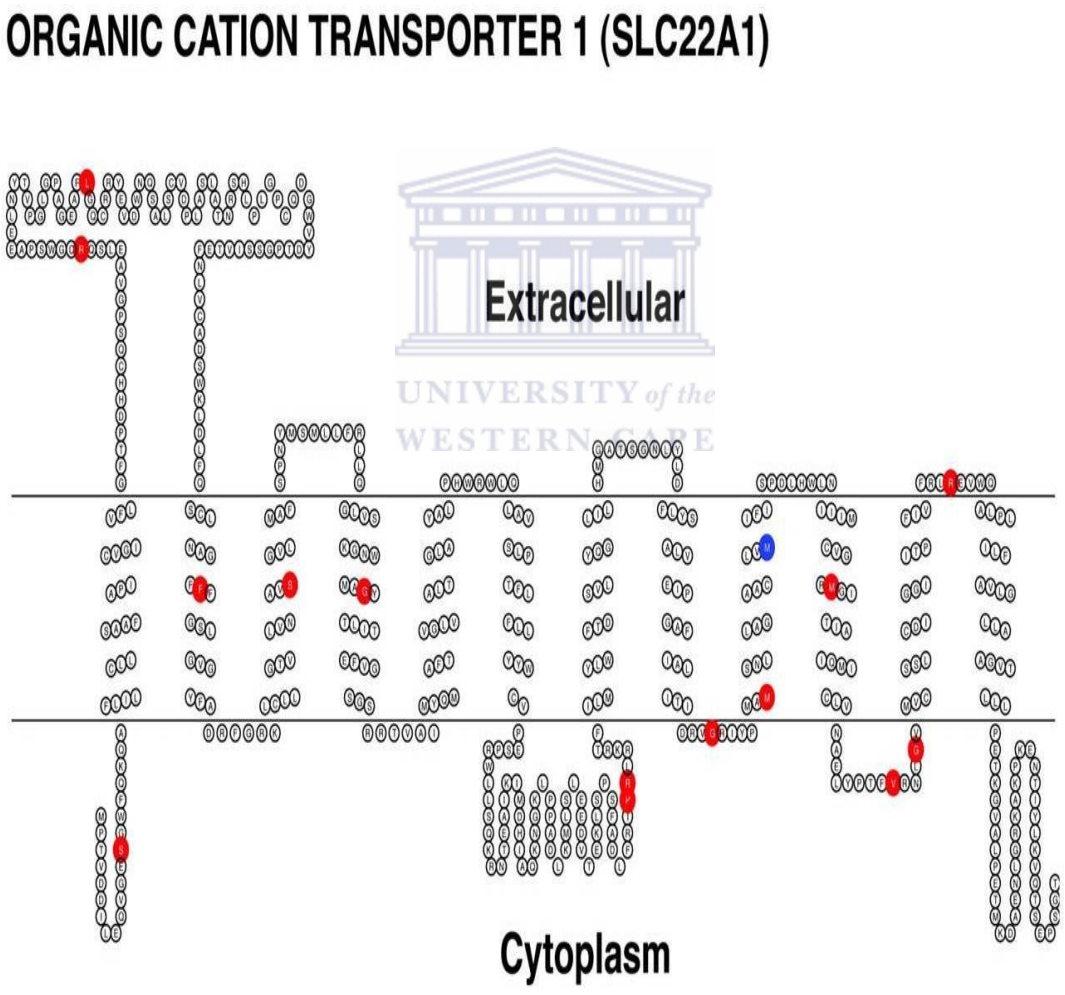


Figure 1.1: Human Organic Cation Tranporter 1 amino acid organisation at the membrane. (Wright & Dantzler, 2004)

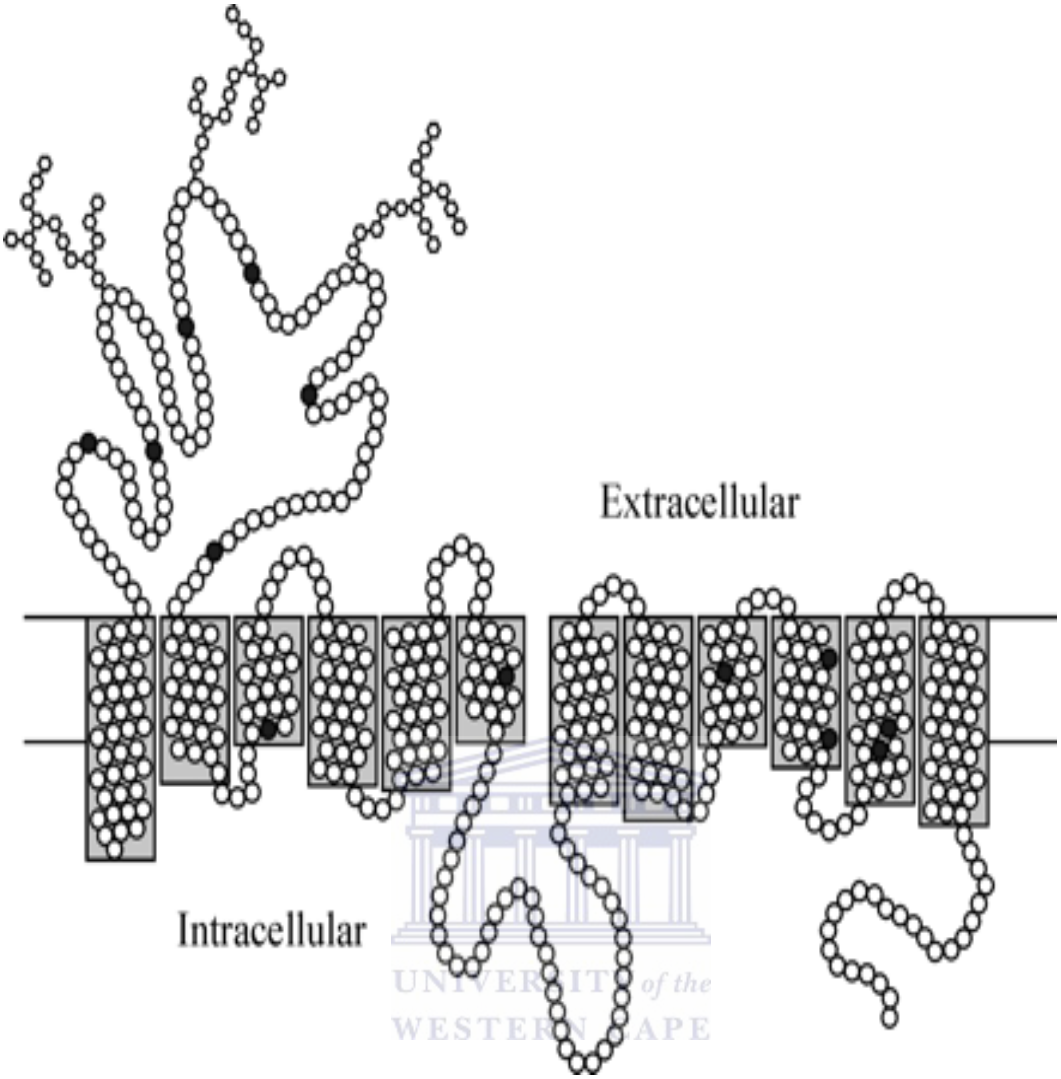


Figure 1.2: Membrane organisation of the human Organic Cation Transporter 2 protein. (Pelis *et al*, 2006)

1.3.2 Tissue Distribution of OCTs

The OCT isoforms mediate the electrogenic transport of small organic cations. These molecules have been determined to be structurally unrelated, however, they share broad substrate specificity and they include model organic cations (TEA, decynium-22), clinically important therapeutic drugs (metformin, procainamide, cisplatin, citalopram, cimetidine), endogenous compounds (dopamine, norepinephrine), and toxic substances (RHPP＋, HPP＋, MPP＋) (Koepsell *et al*, 2007; Gorboulev *et al*, 1997; Zhang *et al*, 1997; Kang *et al*, 2006).

However, despite the sequence homology and broad substrate specificity, OCT1 and OCT2 transporters are discernible from each other as a result of their selective substrates, as well as the difference in tissue distribution (Arndt *et al*, 2001; Koepsell *et al*, 1999). The OCT subtypes show distinctive tissue distributions. OCT1 is primarily expressed in the sinusoidal membrane of hepatocytes and is present in the epithelial membrane of the intestine at minimal levels. OCT2 is primarily expressed in the basolateral membrane of the kidney proximal tubules. OCT3 generally has an extensive tissue distribution as it is found in numerous organs, skeletal muscles, blood vessels, and the placenta (Koepsell *et al*, 2007).

1.3.3 Physiological role of OCTs

OCT1, OCT2 and OCT3 are considered to play an important role in the biliary and renal excretion of their substrates and the distribution of organic cationic drugs in the liver, kidney, heart, and brain (Jonker & Schinkel, 2004). This notion is largely based on their substrate specificities and tissue distributions. Knockout mouse models were generated for the Slc22a1, Slc22a2, and Slc22a3 genes to illuminate the physiological function of OCTs. Oct1-, Oct2-, and Oct3-deficient mice were viable and displayed no obvious phenotypic abnormalities (Jonker & Schinkel, 2004; Jonker *et al*, 2003; Zwart *et al*, 2001). Conversely, Oct1-knockout mice showed considerably reduced hepatic uptake of TEA and metformin (Jonker *et al*, 2001; Wang *et al*, 2002; Wang *et al*, 2003). In Oct1/Oct2 double-knockout mice, the renal secretion of TEA was eradicated and significantly amplified plasma levels of TEA were noted (Jonker *et al*, 2003). These animal studies accentuated the role of the OCT superfamily in hepatic and renal secretion of organic cations (Choi & Song, 2008).

1.3.4 Regulation of OCTs

Regulation of OCTs is affected on an MRN as well as a protein level and as such dictates the cellular tissue distribution of the substrate drugs. Ciarimboli *et al*. (2005) have proposed mechanisms of posttranscriptional regulation of OCT1 by its phosphorylation status and several molecules including activators or inhibitors of protein kinase A (PKA), Src-like p56, and calmodulin (CaM) have been confirmed to modify OCT activity (Cetinkaya *et al*, 2003; Ciarimboli *et al*, 2005; Denk *et al*, 2004; Biermann *et al*, 2006). It has recently been reported that the promoter of contains two adjacent DNA-responsive elements (DR-2) for hepatocyte nuclear factor-4a (HNF-4a). Additionally, HNF-4a effectively activated a luciferase reporter construct containing the OCT1 promoter. Furthermore, mRNA expression of OCT1 was amplified by the over expression of HNF-4a (Saborowski *et al*, 2006).

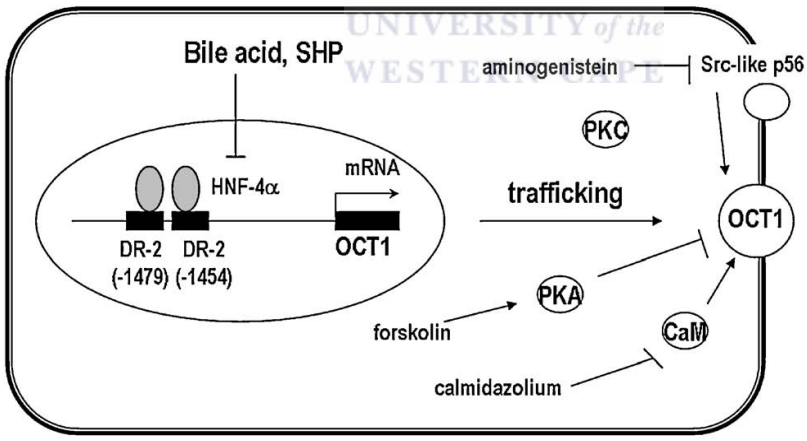


Figure 1.3: Schematic illustration of transcriptional and post trancriptional regulation of OCT1 (Choi & Song, 2008).

OCT 2 transcription is activated through basal promoter elements such as the CAAT-box, located at -92 and the E-box, located at -87. Human upstream stimulatory factor-

1 (USF-1) has been found to transactivate OCT2 promoter activity (Asaka *et al*, 2007).

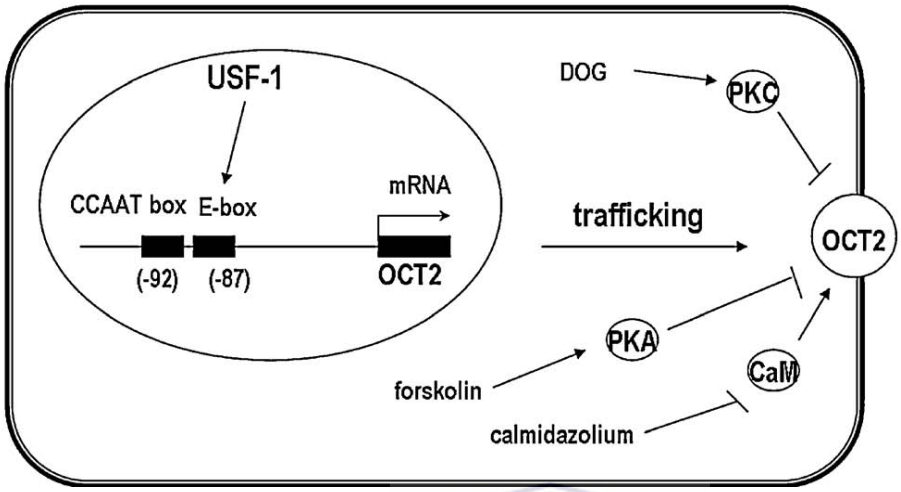


Figure 1.4: Schematic illustration of transcriptional and post trancriptional regulation of OCT2 (Choi & Song, 2008)

1.3.5. Organic Cation Transporter One

1.3.5.1. Genetic Variation in OCT1

Approximately 20 SNPs have been identified in the SLCs (Fahrmayr,*et al*, 2010; Kroetz *et al*, 2010; Nakanishi & Tamai, 2011; Choi and Song, 2008; Pharmacogenomics database). The specified data suggests that although more variations (polymorphisms) exist in the SLC superfamily, only a few are medically relevant. More specifically for OCT1, only one SNP (rs622342) is highlighted by the data proposed by the Pharmacogenomics Knowledgebase. RS622342 (SLC22A1) was determined, by Becker *et al*. (2009) to have a significant effect on the efficacy of metformin uptake in Type-2 Diabetes Mellitus patients.

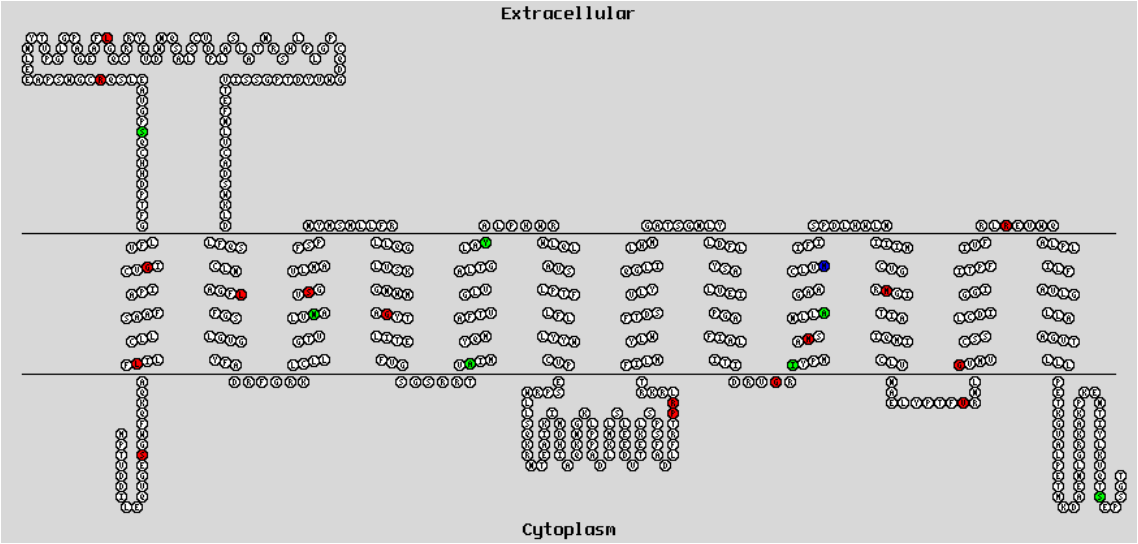


Figure 1.5: Red indicates non-synonymous amino acid changes, indels (insertions and deletions) are indicated in blue, and synonymous changes in green. Exon(s) are indicated by black outlines. (http://pharmacogenetics.ucsf.edu/cgi-bin/Gene.py?hgncSymbol=SLC22A1)

However, recent studies have reported an abundance of polymorphic variants in the SLC22A1 gene, each having an effect on metformin uptake. Kerb *et al*. (2002) identified 25 SNPs in SLC22A1. Shu *et al*. (2003) identified 15 non-synonymous variants in SLC22A1 from different ethnicities. Voluminous other polymorphisms have been identified by Sakata *et al.*(2004), Itoda *et al.* (2004), and Kang *et al*. (2007). Table 1.2. (below) is a summary of the clinically relevant variants in SLC22A1. All information in the table was taken from the PMT database.

Table 1.2. Single nucleotide polymorphisms previously reported as clinically relevant for OCT 1 in different population groups, included are the allele frequencies for each locus.

SNP	*African American*	*Caucasian*	*Asian*	*Mexican*	*Pacific Islander*	*Sub-Saharan African*
RS_2282143	C- 0.918 T- 0.082	C-1 T- 0	C-0.883 T- 0.117	C – 1 T -0	C – 1 T- 0	C- 0.916 T- 0.084
RS_36103319	G- 0.995 T- 0.005	G-1 T- 0	G-1 T- 0	G-1 T- 0	G-1 T- 0	G- 0.997 T- 0.003
RS_34104736	C-1 T- 0	C-0.995 T- 0.005	C-1 T- 0	C-1 T- 0	C-1 T- 0	C- 0.998 T- 0.002
RS_35167514	A - 0.971 Del - 0.029	A - 0.815 Del - 0.185	A - 1 Del - 0	A - 0.786 Del - 0.214	A - 1 Del - 0	A - 0.895 Del - 0.105
RS_34130495	G - 0.993 A - 0.007	G - 0.989 A - 0.011	G - 1 A - 0	G - 1 A - 0	G - 1 A - 0	G - 0.992 A - 0.008
RS_2297373	C - 1 G - 0	C - 1 G - 0	C - 1 G - 0	C - 1 G - 0	C - 1 G - 0	C - 1 G - 0
RS_16891138	A - 0.965 C - 0.035	A-1 C - 0	A-1 C - 0	A-1 C - 0	A-1 C - 0	A - 0.951 C - 0.049
RS_34305973	T - 0.971 Del - 0.029	T - 0.815 Del - 0.185	T - 1 Del - 0	T - 0.786 Del - 0.214	T - 1 Del - 0	T - 0.895 Del - 0.105

RS_1867351	T - 0.737	T - 0.794	T - 0.567	T - 0.722	T - 0.714	
	C - 0.263	C - 0.206	C- 0.433	C - 0.278	C - 0.286	
RS_12208357	T - 1	T - 0.928	T - 1	T - 0.944	T - 1	T - 0.989
	C - 0	C - 0.072	C - 0	C - 0.056	C - 0	C - 0.011

Minor alleles and minor allele frequencies for all populations were determined from the NCBI 1000Genome reference data.

1.4. Genotyping Techniques

1.4.1. PAMSA Genotyping and Limitations

PCR Amplification of Multiple Specific Alleles (PAMSA) involves the use of at least two allele-specific primers in the same reaction and allows for the detection of all SNP alleles, at that locus, present in a sample (Ishiguro *et al*, 2005; Hansson and Kawabe, 2005). The principle behind allele-specific PCR has remained constant while the execution of the method has been altered and improved numerous times in recent years. Many PAMSA methods allow for the discrimination of alleles by gel electrophoresis. Dutton and Sommer (1999), described a PAMSA method in which one allele-specific primer was 31 nucleotides longer than the other. A variant of the PAMSA method is the bidirectional-PASA, described by Liu *et al*. (1997). In this method one allele is amplified n one direction while the other allele is amplified in the opposite direction. This method, however, requires four primers and rigorous optimization (Liu *et al*, 1997; Sasvari-Szekeley *et al*, 2000; Waterfall and Cobb, 2001; Waterfall and Cobb 2002).

In another variation of the PAMSA method, described by Okimoto and Dodgson (1996), a primer length difference of between 2 and 5 nucleotides was used. Alleles

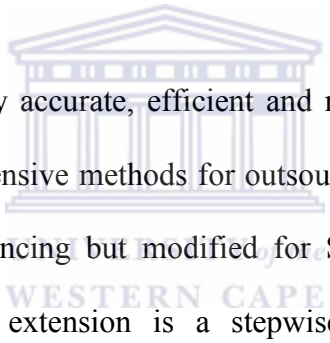
were resolved on acrylamide gel and the molar ratio of allele-specific primers needed to be optimized. In the method described by Gaudet *et al*. (2007), three unlabelled primers were used to amplify specific alleles. The alleles-specific primers included a destabilising mismatch within five bases of the 3' end and a 5' tail of randomly selected bases for the amplification of different length PCR products, which were resolved on agarose gel.

The addition of a destabilising mismatch and a 5' tail was meant to confer higher allele specificity and allow for performing PCR without optimization. However, when used in multiplex these features, with the exception of the destabilising mismatch, reduce the efficacy of the PCR reaction (Edward and Gibbs, 1995; Markoulatos *et al*, 2002). Therefore, non-specific amplification became prevalent. It was determined that the design of the primers was the cause of non-specific amplification. This is the fundamental limitation of this method. Since a random tail was added to each primer, primer-primer interactions, secondary structures and dimerization would have been virtually impossible to avoid. Therefore, rigorous optimization of the primer design and PCR reaction conditions was necessary.

1.4.2. The SNaPshot® Genotyping Assay

Interest in single nucleotide polymorphisms (SNPs) has increased in recent years. SNPs are the most abundant type of genetic variation and they are extensively characterized within the genome (Bertoncini, *et al*, 2011; Quintans, *et al*, 2004). Numerous applications in medical genetics, evolutionary genetics and forensics have been developed for the estimated 3 million identified SNPs (Evans, *et al*, 1999; Gwee, *et al*, 2003; Jobling, *et al*, 2000; Jobling, *et al*, 1997). Classically, SNP analysis was

performed with conventional PCR followed by restriction fragment analysis and gel electrophoresis. This method is, however, expensive, time-consuming and requires hefty amounts of template DNA (Quintans, *et al*, 2004). Moreover, countless high-throughput methods for studying SNPs have been developed in recent years including; DNA Microarrays, FRET analysis, MALDI-TOF spectrometry, Pyrosequencing and TaqMan assays or Molecular beacons (Syvanen, 2001; Carracedo, *et al*, 1998; Lareu, *et al*, 2001). The SNaPshot® assay, also based on conventional PCR, is a high-throughput method which requires little more than well designed primers. It is also less time-consuming and relatively rapid while remaining robust and accurate (Syvanen, 1999).



Despite it being exceptionally accurate, efficient and rapid, SNaPshot® genotyping remains one of the more expensive methods for outsourced genotyping. The assay is based on conventional sequencing but modified for SNP analysis by using chain terminator ddNTPs. Primer extension is a stepwise process that involves the hybridization of a probe to the bases immediately upstream of the SNP. This is followed by a 'mini-sequencing' reaction, in which DNA polymerase extends the hybridized primer by adding a base that is complementary to the SNP nucleotide. This newly incorporated base is detected and determines the SNP allele (Syvanen, 2001; Tully *et al*, 1996; Pastinen *et al*, 1996).

Since primer extension makes use of highly accurate high-fidelity DNA polymerase enzymes, the method is generally very consistent. Primer extension is able to genotype most SNPs under very similar reaction conditions, thus in addition, making it highly flexible. The primer extension method is used in numerous assay platforms.

These platforms use a wide range of detection techniques that include MALDI-TOF Mass spectrometry and ELISA-like methods. (Rapley & Harbron 2004)

1.5. Genetic Ancestry and its role in Pharmacogenetics

Pharmacogenetic discovery and our understanding of population level differences in drug response have long been strongly influenced by the concepts of race, ethnicity and ancestry (Urban, 2010). Population based studies can help to establish baseline frequency distribution of SNPs of genes important in drug metabolism and transport. Extrapolations of possible clinical implications of the baseline frequency distribution of some of the alleles based on established phenotypic characteristics could guide doctors in drug prescription decision-making and possibly provide explanations of ethnic-specific adverse effects in metropolitan medical practice (Matimba *et al*., 2008). As an example, angiotensin converting enzyme (ACE) inhibitors initially designed and tested for the Caucasian American population were blindly used for the treatment of hypertension in the African American population. However, no compensation was made for the highly prevalent Nitric Oxide Synthase deficiency in the African American population (Kalinowski *et al*, 2004). The medication, therefore, was ineffective and deemed dangerous. However, after numerous case-studies and genetic analysis an improved form of the drug was developed for the African American population (Tate and Goldstein, 2004; Echols and Yancy, 2006).

Genomic diversity within the African continent, and more specifically within sub-Saharan Africa, is relatively under-studied, although it is home to significant portion of human genomic diversity (Hardy *et al*., 2008). There is thus a considerable amount to be learnt from characterizing human genomic variation in this part of Africa,

particularly pertaining to health applications (Hardy *et al*., 2008). South Africa in particular contains a wealth of different population groups. The National Biotechnology Strategy Report for South Africa (2002) recognized this fact and recommended that the country focus on documenting the genomic diversity contained within the local indigenous and immigrant populations (Hardy *et al*., 2008). The country is indeed home to various indigenous populations groups, including the Khoisan, Xhosa, Zulu, Venda, and Sotho Pedi groups, the Afrikaaners and the Cape Coloured.

The self-designated Cape "Coloured" population is a uniquely admixed population of immigrant Europeans, Asians and the indigenous populations (Hardy *et al*, 2008). Furthermore, this population has been shown to have the greatest level of recent admixture and almost exclusively resides within the Western Cape region. It is currently estimated that 9% of the South African population is comprised of Coloured individuals, translating to an estimated four million individuals (Quintana-Murci *et al*, 2010).

Admixed groups, such as the Cape Coloured population, share varying proportions of different ancestral populations and their genetic complexity can potentially complicate biomedical research studies (Via *et al*, 2009). The mixed ancestry can potentially provide the intrinsic variability needed to unravel intricate gene-environment interactions, which may help explain the population differences in the epidemiology of complex disease (Via *et al*, 2009) and they will, therefore, form the focus of this study.

1.6. Statistical Analyses

In addition to the number of people typed, population data sets should include the main statistical parameters such as p-value(s) for Hardy-Weinberg equilibrium testing and the Bonferroni Value, as well as confidence intervals and heterozygosity value(s). Additional quality criteria such as the use of various population-specific indices could also be presented.

1.6.1. The Hardy-Weinberg Equilibrium

The Hardy-Weinberg equilibrium (HWE) states that both the allele and genotype frequencies in a population remain constant (Butler, 2005; Nei & Kumar, 2000). This principle assumes that they are in equilibrium from generation to generation unless influenced by external selective pressures (Butler, 2005; Nei & Kumar, 2000). These influences include: mutation, mating, selection, population size, random genetic drift and gene flow. However, this implies static allele frequencies which assume that random mating, no migration, no mutation, infinitely large population sizes and no selective pressure occurs which is, of course, invalid as no such population exists (Butler, 2005; Nei & Kumar, 2000). However, the HWE is still used to determine genotype frequencies.

1.6.2. Observed Heterozygosity

The number of homozygotes plus the numbers of heterozygotes equals 100% of the samples tested in a population (Butler, 2005). Heterozygosity displays the proportion of heterozygous individuals in the population. The higher the heterozygosity value the more allele diversity exists for that particular locus (Butler, 2005). This therefore means that there is less of a chance of a random sample matching.

1.6.3. Shannon Index

Shannon's index (I) is a measure of the differentiation across the loci. It measures the differentiation based on the number of different alleles present in the sample. This is critical to chi-square based statistical test for allele frequency differences (Peakall and Smouse, 2009).

1.6.4. Fixation Index

The Fixation index, also called the Inbreeding Coefficient, is a function of the heterozygosity of a locus. It displays values between -1 and 1 where negative values indicate excess heterozygosity due to selection for heterozygotes (Peakall and Smouse, 2009). Positive values, specifically those near zero, are expected under random breeding while positive values approaching 1 are indicative of undetected null alleles (Peakall and Smouse, 2009).

Table 1.3. Summary of Genetic Formulae

HWE	$\mathbf{p^2 + 2pq + q^2 = 1,}$ where: p = frequency of the dominant allele and, q = frequency of the recessive allele
Ho	$\mathbf{1 - \sum_{i}^{n} P_i^2,}$ where P_i = frequency of i^{th} allele in a population of n samples
He	1 - Sum pi^2 where sum pi^2 is the sum of the squared population allele frequencies
I	-1* Sum (pi * Ln (pi))
F	(He - Ho) / He = 1 - (Ho / He)

Genetic formulae taken from Butler, 2005; Peakall and Smouse, 2009.

1.7. Statistical Analysis Software

1.7.1. GenAlEx

Described by Peakall and Smouse, (2006) this an in-MS-Excel application allows for various statistical analyses to be calculated, including allele and genotype frequencies, the Hardy-Weinberg equilibrium, the Shannon Index and the Fixation Index, as well as numerous others. It is also particularly useful for integration of data into various formats including GenePop, PHASE, PopGen, and various others.

1.7.2. SNPTools

Created by Chen *et al*, SNPTools is an in-MS-Excel application which allows raw data to be converted into a format suitable for various population genetics software, including Haploview, PLINK, Phase and SNPHAP. Furthermore, it also has built-in basic statistical analyses for genetic and epidemiological studies, including Odds Ratios, Confidence Intervals and ChiSquare (Chen *et al*, 2009).

1.7.3. GenePop

The software GenePop, originally described by Raymond & Rousset (1995), filled a gap in practical tools for computing exact tests. Currently, it implements a combination of traditional methods and a few more focused developments (Rousset, 2008). It computes exact tests for Hardy–Weinberg equilibrium, for population differentiation and for genotypic disequilibrium among pairs of loci (Rousset, 2008). Furthermore, it is able to compute estimates of F-statistics, null allele frequencies, allele size-based statistics for microsatellites and number of immigrants by Barton &

Slatkin's (1986) private allele method. Additionally, it performs analyses of isolation by distance from pairwise comparisons of individuals or population samples, including confidence intervals for neighbourhood size (Rousset, 2008).

1.7.4. Haploview

Haploview is used to generate Linkage-Disequilibrium (LD) maps. This is achieved through a scoring system designed to make use of the Log of Odds Ratio (LOD) and D-prime (D') value for each SNP combination (Barrett *et al*, 2005). The LOD score compares the probability of obtaining the same data if the two loci are indeed linked, to the probability of observing that same data purely by chance (CardioGenomics, 2005). Positive LOD scores indicate the presence of linkage, while negative LOD scores indicate that linkage is less likely. The amount of LD between the paired SNPs is defined as D-prime ($0 < d < 1$). A value of 0 indicates that the two loci are in complete equilibrium and a value of 1 represents the highest amount of disequilibrium possible is present (CardioGenomics, 2005).

Therefore, If two SNPs are statistically (LOD>2) related they are coloured in shades of red. Darker shades imply more statistically significant the association. This indicates an allelic association exists between the SNP pair, thus strongly suggesting LD (CardioGenomics, 2005). White squares indicate no relevant statistical significance, as D' values are below 1, and therefore exhibit no LD. The blue squares represent pairwise comparisons where the D' value is equal to 1 but no statistical relevance exists (CardioGenomics, 2005).

1.8. Objectives of the study

This study forms part of a comprehensive project aimed at investigating the genetic diversity of the solute carrier transporter genes and its pharmacogenetic implications within the South African and Sub-Saharan African populations, and consequently it will contribute in filling the gap of missing important pharmacogenetics data from African populations. This includes establishing baseline frequency distribution of previously reported alleles for the SLC22A1, SLC22A2, SLC22A3, SLC47A1 and SLC47A2, SLC22A4 and SLC22A5 genes, as well as the discovery of new genetic variants within the investigated populations for these members of the super-family of solute carrier transporters.

In addition to providing an up-to-date review of the literature concerning the pharmacogenetics of the solute carrier transporters (Chapter 1), the specific objectives of the study described in this thesis were as follows:

1- The development of a MAS-PCR genotyping system suitable for the genotyping of 10 SNP variants from the SLC22A1 (Chapter 2). It is an inexpensive genotyping assay that allows for efficient discrimination of SNP polymorphisms in one reaction tube with standard PCR conditions.

2- The development of SNaPshot® Multiplex Systems for the genotyping of a total of 20 SNP variants from the SLC22A1 gene (Chapter 3). This primer extension-based genotyping method that enables multiplexing up 10 SNPs.

3- The generation of important population pharmacogenetic data for the investigated SNPs (Chapter 3). Among others, allelic and genotypic frequencies, as well as linkage disequilibrium were to be determined and compared with world populations.

4- Conduct a pilot study to explore the possibility of using High-resolution melt (HRM) analysis as a cost-effective alternative for SNP genotyping.

Chapter Two

Development of Allele-Specific PCR Genotyping Multiplexes for 10 SNP Variations on the SLC22A1 gene

2.1. Introduction

Human organic cation transporters mediate electrogenic transport of small organic cations with different molecular structures, independent of sodium gradient. These organic cation substrates include therapeutics, endogenous compounds, as well as toxic substrates (Kang *et al*, 2007). Genetic polymorphisms in drug transporter genes are increasingly being recognised as possible mechanisms responsible for variation in drug response (Shu *et al*, 2007). Furthermore, pharmacogenetic evidence increasingly suggests that membrane transporters are subject to both genotypic and phenotypic polymorphisms. Variations in drug transporters are thus believably responsible for, to a great extent, the interindividual variability in pharmacokinetic disposition, efficiency, and toxicity of drug transporter substrates (Choi & Song, 2008). An extensive assortment of methods has been developed to detect these genetic variations, including RFLP analysis, SNaPshot® mini-sequencing and direct sequencing (Gaudet *et al*, 2007). However, these methods require expensive equipment and rank higher in development costs. The aim for this part of the project was therefore to develop inexpensive Multiplex Allele Specific-PCR (MAS-PCR) genotyping systems for SNP variations on the SLC22A1 gene.

The method described in this section is essentially based on the one developed by Gaudet *et al*. (2007) with an improved multiplexing capability. As many as five SNPs

have been multiplexed in a single reaction. It is an efficient and inexpensive method that enables laboratories to analyse SNPs with standard PCR protocols and is suitable for projects with modest budgets or where sophisticated equipment is not available (Gaudet *et al*, 2007).

2.2. Materials and Methods

2.2.1. Sample Collection

A total of 132 samples in the form of buccal (oral) swabs were collected from healthy, unrelated male donors from the Cape Coloured population residing within the Cape Town metropolitan area. The term Coloured is used to define a uniquely admixed population which was established with contributions from Asian, European and indigenous populations (Hardy *et al*, 2008; Abrahams, *et al*, 2010). This population is known to have the highest rate of admixture in the world and comprises approximately 8.9% (~4 million) of the South African population (Ikediobi *et al*, 2011).

Ethical clearance for the project was obtained from the Senate Research Committee of the University of the Western Cape prior to the sample collection exercise. Once swabs were obtained, they were immediately placed into coded envelopes and immersed in ice for transport. Samples that were not used immediately were stored at -20 °C.

2.2.2. DNA Extraction

DNA was extracted by a combination of two techniques; salt lysis and phenol/chloroform extraction. The cotton tip of each swab was cut off and immersed

in an eppendorf tube containing 600 µl salt lysis buffer and 3 µl 20 mg.ml^{-1} Proteinase K. The tubes were then incubated at 56 °C overnight. The swabs were discarded and the solution containing the DNA was retained. To precipitate the DNA, 200 µl 6M NaCl was added and the tubes were vigorously shaken for 15 s. The tubes were then centrifuged at 5 000 rpm for 15 min and the supernatant was collected. The supernatant was mixed with isopropanol using a 1:1 ratio and incubated at -80ºC for 30 min. DNA was then pelleted by centrifugation at 14 000 rpm for 30 min. The pellet was washed with 100 µl 70% Ethanol and centrifuged at 14 000 rpm for a further 15 min. The pellets were allowed to dry briefly before being re-suspended in a final volume of 30 µl SABAX water. Samples were immediately stored at -20°C for future use. The DNA yield was determined using a Nanodrop ND 1000 UV-Vis Spectrophotometer.

2.2.3. Primer Design

All primers were designed using the online software BatchPrimer 3 (http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi), Primo (http://www.changbioscience.com/primo) and PrimoSNP (http://www.changbioscience.com/primo/primosnp.html). Primers were designed to genotype the selected SNPs in two multiplexes, namely Multiplex 1 (M1) and Multiplex 2 (M2), each covering 5 SNPs. The primers were designed according to the method described by Gaudet *et al*, 2007. A total of three primers were designed for each SNP. This includes a forward primer for each allele, and a common reverse primer (Figure 2.1).

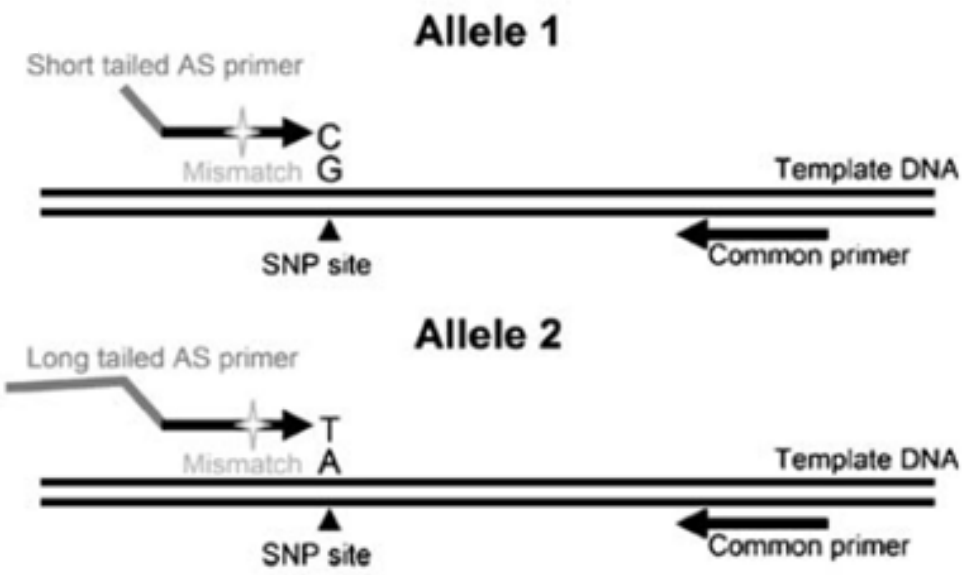


Figure 2.1: The mechanism of Allele-Specific PCR. Each primer is specific for one allele. A size difference is imparted by the addition of 5'-tails and a destabilising mismatch incorporated contiguous to the 3'-end. (Gaudet *et al*, 2007)

2.2.4. PCR Amplification

All DNA samples were diluted to a final concentration of 5 ng.μl^{-1} and used for the PCR amplification. The primers used for the PCR amplification for the two multiplexes are listed in table 2.1. All primers were synthesized by Integrated DNA Technologies (IDT) Munich, Germany. Primers were diluted to a working stock of 2 µM. PCR amplifications were performed using the Quiagen Multiplex kit in a final volume of 10 µl according to manufacturer's instructions. PCR amplifications were performed using 7 µl of the Qiagen Master Mix and 15 ng DNA adjusted to a final volume of 10 µl using RNA-free water. The Qiagen Master Mix used consisted of 1X Quiagen Multiplex Master Mix, 2 µM Primers, 0.25X Q Solution, adjusted to 7 µl with RNA-free water.

Amplification was performed on a GeneAmp 2700 (Applied Biosystems) thermal cycler or Supercycler (Kyratec). Thermocycling conditions were as follows: 95°C for 15 min, 94°C for 30 s, 52.5°C for 60 s, 72°C for 90 s, for a total of 30 cycles, followed by 60°C for 30 min and a final hold at 4°C. PCR products were confirmed by agarose gel electrophoresis. An aliquot of 3 µl of each product was run on a 2% agarose gel at $5V.cm^{-1}$. The gel was then stained with 0.1 $\mu g.ml^{-1}$ GelRed and visualised under UV light.

2.3. Results and Discussion

SNP variations were selected from the Pharmacogenomics Knowledgebase, UCSF-PMT, Japanese SNP and NCBI databases. As listed on the database, primarily SNPs which had been annotated and determined to be of clinical relevance were selected. However, in the Pharmacogenomics Knowledgebase only one such SNP existed. Thus, other databases were consulted for SNPs which have been cited in literature and investigated for possible clinical relevance or implications and pharmacogenetic importance. The UCSF-PMT database provided a further 7 SNPs, although many had not yet been annotated. The final SNPs were selected from a Japanese and NCBI databases to complete the ten SNPs for M1 and M2.

2.3.1. Primer Design and Optimization

Primers for both multiplexes were initially designed manually without using any primer software, and based on the PAMSA method principles (Gaudet *et al*., 2007). Initial trials using the PAMSA method for primer design were proven to be less suited to multiplex genotyping (data not shown). This approach was therefore improved by the use of multiple methods of primer design. The primer sets were then designed

using freely available online tools BatchPrimer3, Primo and PrimoSNP. BatchPrimer3 is an online tool similar to Primer3, which allows multiple primers to be designed at once, with the ability to design allele-specific primers (You *et al*, 2008). It automatically inserts the 3'-mismatch to increase specificity. Allele-specific mismatches were manually moved to fall in different positions within the final five bases contiguous to the 3'-end. A tail of random bases was added to the 5'-end of each forward primer, as prescribed by Gaudet *et al.* (2007). The difference in tail length between the allele-specific primers was designed to be between ten and twenty bases. The primer orientation was changed when needed in order to further increase specificity. The result was three SNP sites having allele-specific forward primers and two SNP sites having allele-specific reverse primers.

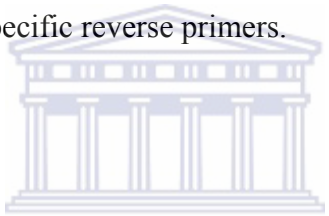


Table 2.1. Allele-Specific primer sets for 10 SNPs

SNP Name	Orientation	Primer Sequence	Allele	Product Size
Multiplex 1				
RS_34130495	R 1	GCCATGGGGTAGATGCGGCA	T	50 bp
	R 2	***GACTGCTGATCGATG***GCCATGGGGTAGATGCGGCG	C	70 bp
	F	AGCCCTCATCACCATTGACCG		
RS_34104736	F1	GTGCTGGTCAACGCGGTGTC	G	100 bp
	F2	***TGTCAGACTGCTGATCGATG*** GTGCTGGTCAACGCGGTGTT	A	120 bp
	R	AGTTGCCCTTGCTGACCAGG		
RS_36103319	F1	AAGGGCAACTGGATGGCTGG	C	210 bp
	F2	***TGTCAGACTGCTGATCGATG*** AAGGGCAACTGGATGGCTGT	A	230 bp
	R	AGAGAGGAGGCCATTCTAGCC		
RS_34305973	F1	AGGTTCACGGACTCTGTGCTC	G	250 bp
	R	CAACTTACCAGGTGAGATAAAAATCA		
RS_2282143	F1	GCAGACCTGTTCCGCACGCC	G	300 bp
	F2	***TGTCAGACTGCTGATCGATG*** GCAGACCTGTTCCGCACGCT	A	320 bp
	R	CTCATCCCACTGGTGATCTCC		
Multiplex 2				
RS_683369	F1	TTTTTTTGAATGCGGGCTTCATC	G	70 bp
	F2	TTTTTTTTTTTTTTTTTTTTTTTTGAATGCGGGCTTCCTG	C	90 bp
	R	GGCCTTTACATACCTGTCTGC		

RS_35167514	F1	GCAGCCTGCCTCGTCA	T	90 bp
	R	TCCCCACACTTCGATTGCCTG		
RS_628031	F1	TTTTTCGCATCTACCCCATGGCCA	T	190 bp
	F2	TTTTTTTTTTTTTTTTTTTTTTCGCATCTACCCCATGTCCG	C	210 bp
	R	GACAAAGGTAGCACCTCATC		
RS_117474883	F1	ACCAACTTACCAGGTGAGATAAACATC	G	280 bp
	F2	AGCGGACACGCAGGACCGCCAACTTACCAGGTGAGATAAAGATG	C	300 bp
	R	CTTCAGTCTCTGACTCATGCC		
RS_34205214	F1	AGACCTGTTCCGCACGCT	A	400 bp
	F2	TTTTTTTTTTTTTTTTTTTTAGACCTGTTCCGCACGCC	G	420 bp
	R	CCACTGGTGATCTCCCAAAGG		

Preliminary results obtained using each primer set in uniplex proved to be of better quality than the previously designed sets. Rigorous optimization was, however, required for each primer set to work optimally. In order to achieve this, the approach set out by Markoulatos *et al*. (2002) was followed. It prescribes that the amount of primer and reaction constituents be optimized first. Since the Qiagen Multiplex kit PCR buffer being used is sold pre-optimized, only the amount of primer could be further optimized. All primers were diluted to a working concentration of 4 µM. This would then result in a primer concentration of 0.4 µM in each reaction. To begin with, two primer sets were duplexed in a single reaction, each at 0.4 µM final primer concentration. The concentration of each primer set was carefully adjusted initially by decreasing the concentration to 0.2 µM for one primer set. When a consensus was reached, each primer set at 0.2 µM, a third primer set was introduced. The process then continued to incorporate a third and fourth primer set. This approach worked very well until the fourth primer set was added, and some drop out was observed. Some bands were no longer resolved on the gel. This finding was compounded when the fifth primer set was incorporated. Three bands were resolved, two of the bands from the triplex and the expected fifth primer set band. It was thus determined to be a result of secondary structures and dimmers forming between the primers in the multiplex. All secondary structures produced by the primers were re-analysed. Alignments were then drawn for each primer set and between primer sets. It was discovered that several of the primers were indeed forming dimers, each dimer having a greater binding-specificity than they would have with DNA.

Therefore, it was necessary to run a gradient PCR to determine optimum annealing temperatures. Once completed, it became evident that the optimized annealing

temperatures had had no effect on the binding specificities of the primer sets. Persistently, an amplicon drop out was observed. As a result of the random tails attached to each primer, secondary structures and dimers had formed within the multiplexes and this was determined to be the primary factor affecting the binding specificity of the primers.

In publications penned by Chun *et al*. 2007, Dixon *et al*. 2005 and Kenta *et al*. 2008, a similar technique to the PAMSA method is described. The methods described made use of the same allele-specific primer design as described by Gaudet *et al*, 2007, but used a multitude of additions to the primer tails instead of random bases for primer tails. These included poly (I) additions within the primer sequence (Chun *et al*, 2007), universal primer tails (Dixon *et al*, 2005) and dye-labelled allele specific primers (Kenta *et al*, 2008). Since it has been validated and is highly reliable for SNaPshot® genotyping, Poly (T)-tails were determined to be the best approach, as described by Nelson *et al*. (2007). This has prevented the formation of many hairpin loops and other secondary structures, provided that the primer itself did not have any A-runs. Furthermore, the primer sequences and the DNA sequence of interest could not be AT-rich as the primers would bind non-specifically and amplify the incorrect fragment of DNA. This method formed the basis of design for M2. Four of the five primer sets in M2 were designed to have 5' poly-T tails. This resulted in a much more stable and specific interaction between the primers and targets. Optimization for M2 was virtually non-existent, barring the adjustment of the concentration of one primer set with a random base tail. Furthermore, the sizes of the amplified products remained similar, thus allowing primers to be exchanged between the two multiplexes. Therefore, two primer sets were exchanged between the two multiplexes. These

consisted of primer sets 2282143 and 34305973 in M1, and primer sets 34205214 and 35167514 in M2.

A great improvement was obtained in the performance of the re-designed multiplexes. However, some optimization was still required as the annealing temperatures of the new primer sets were above that of the others. Immediately, this presented a new challenge. Non-specific amplification would be prevalent as a result of the difference in annealing temperature. To overcome this challenge, the concentration for those new primer sets was analysed. A concentration spectrum (Figure 2.2) was generated for each primer to determine the lowest possible concentration that could be used without sacrificing resolution quality. It was discovered that a final reaction concentration of 0.2 µM presented the best possible results.

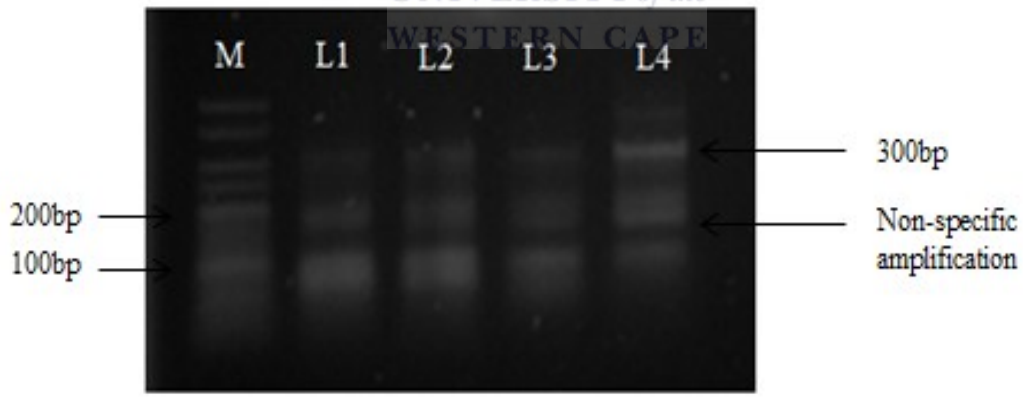


Figure 2.2: Primer concentration gradient for primer set 3610 on 3% agarose gel. Lane M: Marker. Lane 1: 0.5 µM primer. Lane 2: 0.4 µM primer. Lane 3: 0.3 µM primer. Lane 4: 0.2 µM primer concentration. Lane 4 presented the optimal concentration for this primer. Note, the non-specific amplification is a result of the standard, non-optimized, annealing temperature used for amplification.

Moreover, this also allowed any primer set causing a loss of resolution to be exchanged for one from M2 without further optimization required. The resulting multiplexes (Table 2.3) were, therefore, significantly more specific and generated better results. However, in order to reach such results it was through trial and error primer sets were exchanged until a perfect fit was achieved across all multiplexes.

Table 2.2. Summary of final multiplex components

RsID	Polymorphism	Product Size
M1		
34130495	G > A	50 – 70 bp
35167514	A > Del	90 bp
34104736	C > T	100 – 120 bp
36103319	G > T	210 – 230 bp
34205214	A > G	400 – 420 bp
M2		
683369	G > C	70 – 90 bp
628031	T > C	190 – 210 bp
34305973	T > Del	250 bp
117474883	G > C	280 – 300 bp
2282143	C > T	320 – 340 bp

This technique was then ready to be critically tested and evaluated in order to determine whether genotyping would be possible. The initial results showed a good potential. However, in order to properly resolve the target products on agarose gel the correct balance between the concentration of the gel and the size of the fragments needed to be established. Because each SNP site has one, or both, of two possible

fragment sizes, separated by a 10 – 20bp size difference it was necessary to determine at which voltage the fragment would correctly separate through the gel without the smallest fragments running off the gel. It was thus determined, by sequentially increasing the concentration of the gel and voltage of the electrophoresis, that agarose gel of concentration of between 1.5% and 3%, and an electrophoresis voltage of between 60 and 75 volts, yielded the most suitable results. This allowed the smallest fragment to be clearly viewed without running off the gel as well as the larger fragments to separate out correctly. It was determined that an agarose gels concentration of 2% and electrophoresis voltage of 75 volts (7.5V.cm^{-1}) was optimal and set as a standard.

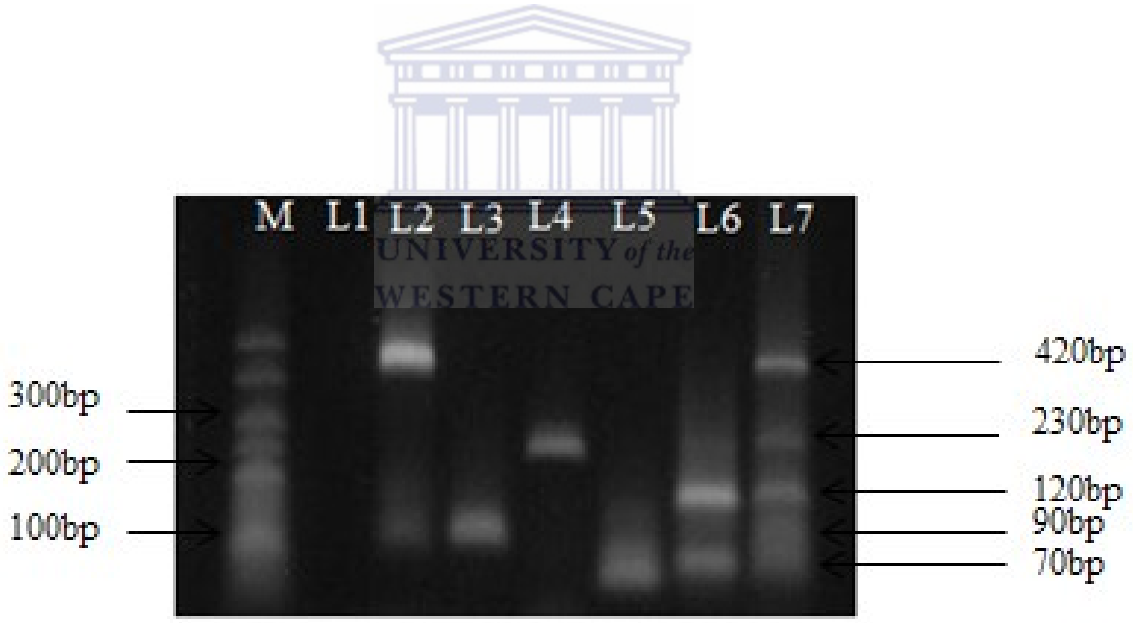


Figure 2.3: Multiplex one genotyping result on 2.5% agarose gel. Lane M: Marker. Lane 1: Negative control. Lane 2: 34205214, genotype GG. Lane 3: 35167514, genotype CC. Lane 4: 36103319, genotype CC. Lane 5: 34130495, genotype AA. Lanes 6: 34104736, genotype AA. Lane 7: Sample 38, genotype GCCAA.

In M1 some non-specific amplification was observed (Figure 2.3). Primer set 34205214 amplified a second band at 100 bp while primer set 34104736 amplified a second band at 80 bp. It is important to note that although this nonspecific amplification may be seen in Figure 2.3, the amplified products had no effect on the

efficacy of genotyping. The bands displayed were separate and settled between the bands of interest. Therefore, it may be said that further optimization will be required to generate sharper genotyping profiles. Multiplex 2 (Figure 2.4) did not display any non-specific amplification.

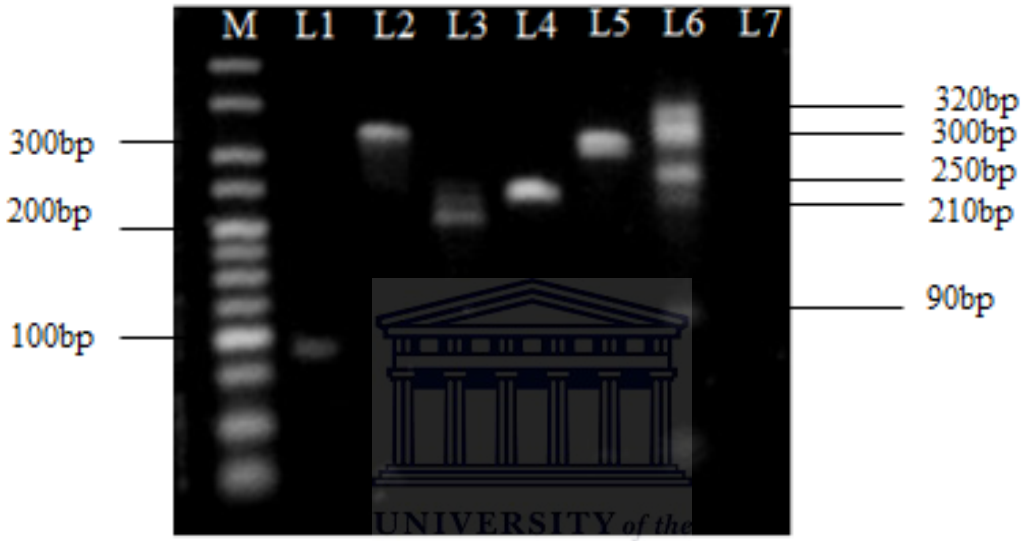


Figure 2.4 Multiplex two genotyping result on 2.5% agarose gel. Lane M: Marker. Lane 1: 683369, genotype GG. Lane 2: 2282143, genotype GG. Lane 3: 628031, genotype CC. Lane 4: 34305973, genotype GG. Lane 5: 117474883, genotype GG. Lane 6: Sample 38, genotype GGCGG. Lane 7: Negative control.

The development of this technique may be said to be of paramount importance not only specifically for its application presented herein but also for future work. This technique would be useful in a large host of applications and will form the basis of genotyping for future students and academics. It has been demonstrated to allow for large sample numbers, and SNP numbers, to be investigated in a relatively short amount of time with great accuracy and precision. Moreover, since it has been

optimized as a multiplex of five SNPs each, the cost incurred by use of this technique is minimal compared with many similarly accurate methods. Developing this technique has therefore become crucial in the expansion and advancement of a laboratory which cannot afford expensive equipment and which has a minor annual budget.

2.3.2. Validation of MAS-PCR Technique

Validation plays a fundamental role in determining the reliability and limitations of novel techniques. It is important to determine if the analytical procedure is adequate for its intended use (Butler *et al*, 2004). The technique as describe above was validated using SNaPshot® mini-sequencing. The genotypes for 10 samples, generated from MAS-PCR (Figure 2.5), were compared with those generated by SNaPshot® mini-sequencing (Figure 2.6). Sample 38 was chosen as reference and genotyped with both MAS-PCR and SNaPshot®. The results obtained were a perfect match.

The SNaPshot® technique, being more specific, did, however, resolve and confirm those genotypes which were harder to determine from the gel (Data not shown).

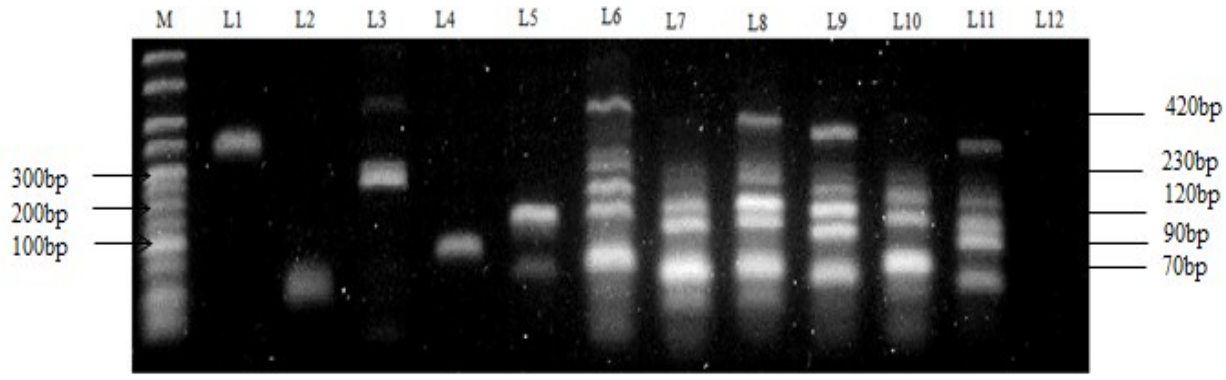


Figure 2.5: Multiplex one genotyping result on 1.8% agarose gel. Lane M: Marker. Lane 1: 34205214, genotype GG. Lane 2: 35167514, genotype CC. Lane 3: 36103319, genotype GG. Lane 4: 34130495, genotype AA. Lane 5: 34104736, genotype AA. Lane 6: Sample 38, genotype

GGCAA. Lane 7: Sample 23 genotype GCCAA. Lane 8: Sample 5, genotype GCCAA. Lane 9: Sample 64, genotype GCCAA. Lane 10: Sample 33, genotype GCCAA. Lane 11: Sample 80, genotype GCCAA. Lane 12: Negative control.

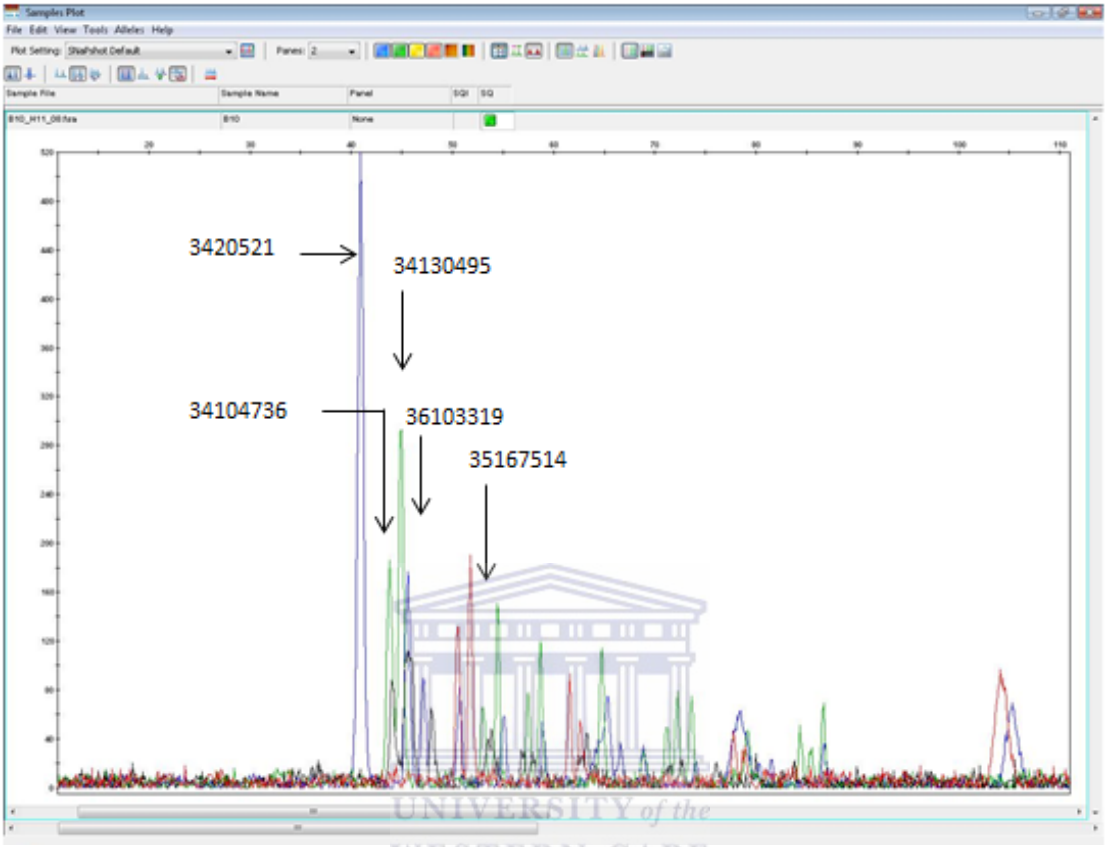


Figure 2.6: SNaPshot® validation of the SNPs genotyped by MAS-PCR using Sample 38 as the reference sample. As a result of the variability of the electroconductivity of SNaPshot® extended sequences, an amount of drift is observed within the samples. Thus, smaller fragments tend to drift by up to 20bp from their expected size.

2.3.3. Blind Tests

To confirm and validate the multiplex system developed, an intralaboratory blind test was performed as determined by Yang *et al*. (2004). The exact PCR conditions and set up was used for the blind test. However, in order to test the efficacy of the multiplex developed, a representative number of samples were chosen at random (10% of the total). These samples were then shuffled so as to further randomise the test. After gel electrophoresis the exact sample had to be identified, from a known

consignment of sample IDs, using only the genotype generated (Figures 2.7 and 2.8 respectively).

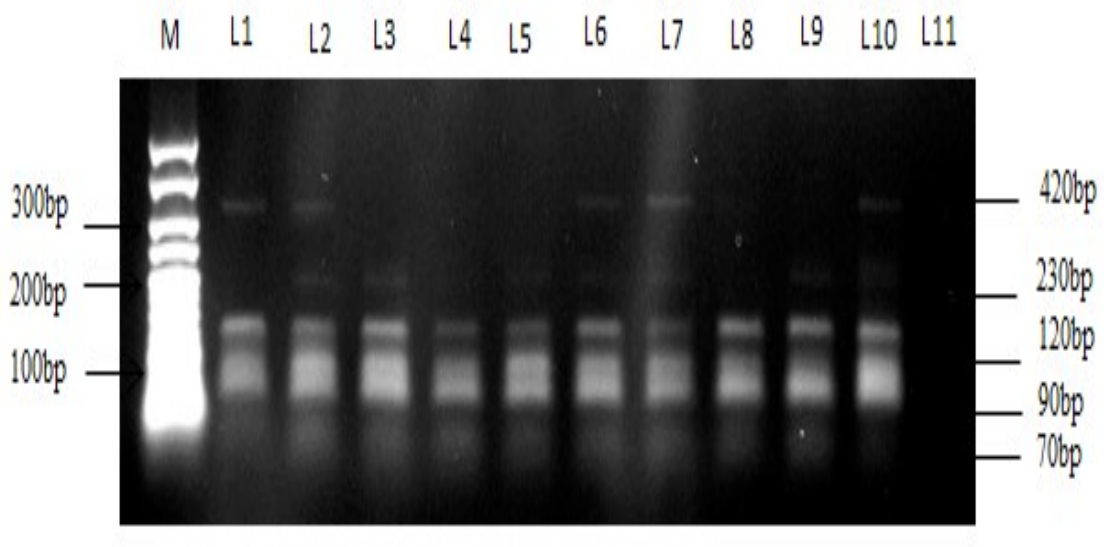


Figure 2.7: Genotyping results for multiplex 1 used for the blind test. A 2% agarose gel was used. Lane M: Marker. Lane 1: Sample 7, genotype G*CAA. Lane 2: Sample 9, genotype GGCAA. Lane 3: Sample 3, genotype *GCAG. Lane 4: Sample 10, genotype **CAA. Lane 5: Sample 1, genotype GGCAG. Lane 6: Sample 8, genotype GGCAG. Lane 7: Sample 2, genotype GGCAG. Lane 8: Sample 6, genotype G*CAA. Lane 9: Sample 5, genotype *GCAA. Lane 10: Sample 4, genotype GGCAA. Lane 11: negative control. (* indicates missing allele)

It is important to note that all SNP loci in M2 are homozygous (Figure 2.8). Therefore some amplicons might appear too close to the next. However, this has no effect on genotyping since only one allele is present. More time for separation would increase the discriminative ability of the gel.

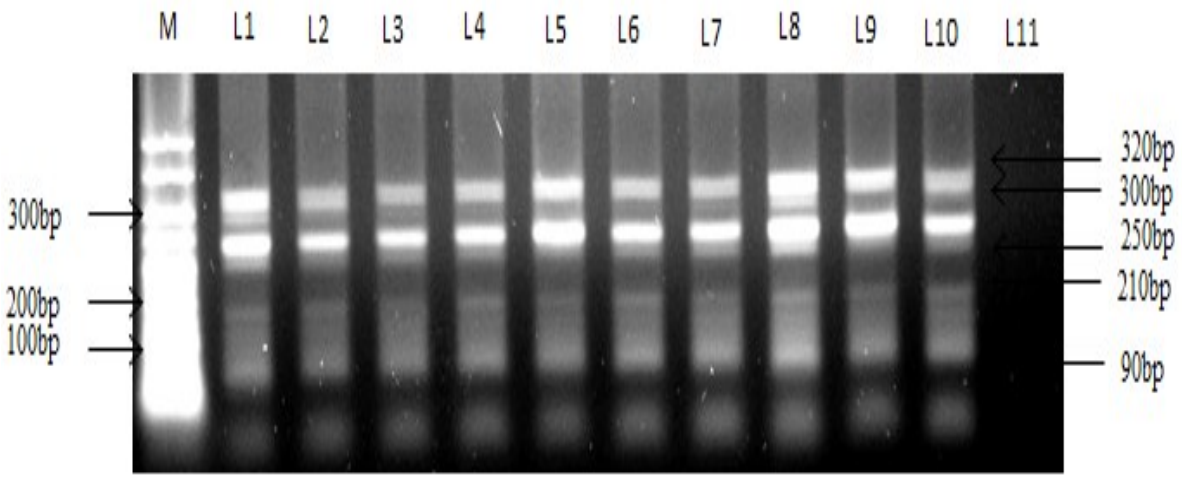


Figure 2.8: Genotyping result for Multiplex 2. A 2% agarose gel was used. Lane M: Marker. Lane 1: Sample 7, genotype CGTCG. Lane 2: Sample 9, genotype CCTCG. Lane 3: Sample 3, genotype CCTCG. Lane 4: Sample 10, genotype CCTCG. Lane 5: Sample 1, genotype CCTCG. Lane 6: Sample 8, genotype CCTCG. Lane 7: Sample 2, genotype CCTCG. Lane 8: Sample 6, genotype CGTCG. Lane 9: Sample 5, genotype CCTCG. Lane 10: Sample 4, genotype CCTCG. Lane 11: negative control.

The initial blind test results (Table 2.4) showed that five samples in M1 exhibited incomplete genotypes, which was as result of the improper amplification and electrophoresis conditions. However, after performing the blind test again, 100% of the tested samples were correctly identified thus confirming the validity of the multiplex PCR system presented herein.

Table 2.3. Result of the 10 sample blind test

Sample ID	Identities of samples	Pass / Fail
1	Identified sample 7	Pass*
2	Identified sample 9	Pass
3	Identified sample 3	Pass*
4	Identified sample 10	Pass*
5	Identified sample 1	Pass

6	Identified sample 8	Pass
7	Identified sample 2	Pass
8	Identified sample 6	Pass*
9	Identified sample 5	Pass*
10	Identified sample 4	Pass

Where * indicates incomplete genotypes and therefore unidentified samples.

2.4. Summary

To summarise briefly, the technique presented herein does provide a reliable means for genotyping large quantities of samples relatively quickly. Not only is this technique well suited to laboratories with modest budgets, but it is also well designed and has been proven to be highly effective and efficient. The data generated was sufficient for genotyping and proven to be accurate. The assay is, therefore, both cost-effective while remaining reliable and fairly easily implemented. With exceptionally modest effort this technique can be further adapted for more extensive forms of assays and research interests.

Chapter Three

Analysis of Twenty SLC22A1 Genetic Variants within the Cape Coloured Population

3.1. Introduction

The Cape Coloured population of South Africa, also called the Cape Mixed Ancestry population or the Cape admixed population, is a result of contributions from Asian and European immigrants as well as indigenous populations (Abrahams *et al*, 2010). This is advantageous as the mixed ancestries of this populations allows for the identification of genetic risk factors (Smith *et al*, 2001; Tian *et al*, 2007) and may provide the intrinsic variability needed to untangle complex gene-environment interactions which may help explain the population differences in the epidemiology of complex disease (Via *et al*, 2009). The aim of this part of the study was therefore to investigate the genetic diversity of the SLC22A1 gene within the Cape Coloured population. A novel SNaPshot® assay was specifically developed to investigate 20 SNP variations on the SLC22A1 within this population. The SNaPshot® assay, a genotyping technique based on conventional PCR, is a high-throughput method, which requires little more than well-designed primers. It is also less time-consuming and relatively rapid while remaining robust and accurate.

This chapter describes the development and implementation of a SNaPshot® assay as a high-throughput method for genotyping and generating population genetics data for the Organic Cation Transporter 1 gene within the Cape Coloured population. It also presents the statistical analysis of the results including main population genetics

parameters. This was followed by a discussion of possible pharmacogenetic implications of the obtained results for this uniquely admixed population from South Africa.

3.2. Materials and Methods

3.2.1. Sample Collection

Sample collection was performed as previously described in Chapter Two.

3.2.2. DNA Extraction

DNA extractions were performed as previously described in Chapter Two.

3.2.3. SNP Selection

Previously reported SNP variants on the SLC22A1 gene were selected from the Pharmacogenomics Knowledgebase (http://www.pharmgkb.org), the UCSF-PMT (http://www.pharmacogenetics.ucsf.edu/) database and the NCBI-SNP database. Among the 20 SNPs selected for this study, 9 variants were annotated and have been shown to have clinical relevance, and the remaining SNPs are still under investigation.

3.2.5. Primer Design

All primers were designed using the online software BatchPrimer 3 (http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi), Primo (http://www.changbioscience.com/primo) and PrimoSNP (http://www.changbioscience.com/primo/primosnp.html). Primers were designed to genotype the selected SNPs in two multiplexes (MplexI and MplexII), each covering

10 SNPs, according to the method described by Nelson *et al*. (2007). The primers were used for the PCR amplification and the SNaPshot® mini-sequencing reactions.

3.2.6. PCR Amplification

PCR amplification was performed as previously described in Chapter Two with a minor difference in cycling conditions. The primers used for the PCR amplification for the two multiplexes are listed in Table 3.1. PCR amplifications were performed on a GeneAmp 2700/ 2720 (Applied Biosystems) thermal cycler. Thermocycling conditions were as follows; 95°C for 15 min, 94°C for 30 s, 60°C for 90 s, 72°C for 90 s, for a total of 40 cycles, followed by 72°C for 10 min and a final hold at 4°C. PCR products were then verified by agarose gel electrophoresis. An aliquot of 3 µl of each product was run on a 2% agarose gel at $10V.cm^{-1}$. The gel was then stained with 0.1 $\mu g.ml^{-1}$ Ethidium Bromide and visualised under UV light.

Table 3.1. List of PCR amplification and extension primer sequences for Multiplex 1

SNP ID	Primer Sequence	Concentration	Polymorphism	Extension Primer Sequence	Size (bp)
Rs_36103319	F-CTGGTCAGCAAGGGCAAC	0.2 µM	C > A	5-Poly-T-GGGCAACTGGATGGCTG -F	23
	R-GGCCTCATCCCCATGATAA				
Rs_12208357	F-CTGACCACCACTGCCAGAGT	0.2 µM	A > G	10-Poly-T-GGGCTCCAGCCACAGC -F	26
	R-CCTGGCACTGTATAGTTCAGCTC				
Rs_1867351	F-CAGGCCTGGCACTGTTAGTT	0.2 µM	G > A	15-Poly-T-GCTCAGCCACCCCAGG -F	32
	R-TTCACACCTGACCACCACTG				
Rs_34104736	F-TTTGGCCGTAAGCTGTGTCT	0.2 µM	G > A	20-Poly-T-GCCATGAGCACGCCC -R	36
	R-CATGTAGTTGGGCGAGAAGG				
Rs_2282143	F-AAAGCTGAGCCCTTCATTTG	0.2 µM	A > G	25-Poly-T-GACCTGTTCCGCACGC -F	42
	R-CCCCTCACCACAGGTACATC				
Rs_34130495	F-GCCTTCATAGCCCTCATCAC	0.2 µM	C > T	30-Poly-T-CACCATTGACCGCGTG -R	47
	R-CCGCCAACAAATTTGACAT				
Rs_16891138	F-GGGCCACAGGCTGTAGTTTG	0.2 µM	T > G	35-poly-T-AGGTGCCCGAGGGTTC -F	52
	R-AACATCTCTCTCAGGTGCCCG				
Rs_35167514	F-TGGCCATGTCAAATTTGTTG	0.2 µM	T > -	40-Pol-T-GCAGCCTGCCTCGTC -R	55
	R-TGATGAAAGCAGACAACTTACCA				
Rs_2297373	F-ATCTGTGTGGGCATCGTCTT	0.2 µM	G > C	50-Poly-T-GGTCAGGTGTGAAACCCAG -R	69
	R-CAGGACTCTGGCAGTGGTG				
Rs_34305973	F-TGGCCATGTCAAATTTGTTG	0.2 µM	G > -	45-Poly-T-CAACTTACCAGGTGAGATAAAAATC -F	70
	R-TGATGAAAGCAGACAACTTACCA				

All extension primers were used at a final concentration of 0.2 µM

Table 3.2. List of PCR amplification and extension primer sequences for Multiplex 2

SNP ID	Primer Sequence	Concentration	Polymorphism	Extension Primer Sequence	Size (bp)
RS683369	GACTCCTGGAAGCTGGACCT GTAGCCAACACCGAGAGAGC	0.2 µM	G > C	5-poly-T-CAACACCGAGAGAGCCAAA -F	24
RS628031	AGCCCTCATCACCATTGACCG AATCATGACGAGGCAGGCTGC	0.2 µM	T > C	10-Poly-T-CCGCATCTACCCCATGGCC -F	29
MPJ6_O1004	CTCAGCTGTGTAGACCCCCT CGTGTCATACACCCAGCCATC	0.2 µM	G > A	15-Poly-T-CACCAACAGGAGCCACCTGC -F	35
RS34205214	CACCGAAAAGCTGAGCCCTTC CCCCTCACCACAGGTACATC	0.2 µM	A > G	20-Poly-T-ATGAAGGTGCGCTTCCTCAGG -F	41
RS117474883	TGGCCATGTCAAATTTGTTG ATTGCCTGGGAAATGATGAA	0.2 µM	G > C	25-Pol-T-ATTTTTATCTCACCTGGTAAGTTGGTAAG -F	54
RS35191146	GTCAAATTTGTTGGCGGGGGC TCCCCACACTTCGATTGCCTG	0.2 µM	G > -	30-Poly-T-ACCAACTTACCAGGTGAGATAAAAAT -R	56
RS34059508	TCTGAGGTGCTGAGCACTGGA AATGAGGGGCAAGGCTTGCCA	0.2 µM	G > A	35-Poly-T-GGGAGGAACACACCATCACTC -R	61
RS115733275	CTCAGGATCACCTGGATTCCT CTGCTTCATCACGAGAACCAC	0.2 µM	G > T	40-Poly-T-CACAGGGTCCTGGTGGAAC -R	66
RS116824962	CAGTGAGGGAAAGGCAAGACA GGTCTCGAACTCCTGGCTTCA	0.2 µM	C > T	45-Poly-T-CCAGGCATGGTGGCTAA -F	71
RS118060798	TGATTTAGCTAGGCCTCGGGT TTCCAACCGCACCATGGCATT	0.4 µM	A > G	50-Poly-T-GGGTGTGCAGTGAATGGAA -F	79

All extension primers were used at a final concentration of 0.2 µM,

3.2.7. Post PCR Purification

Subsequent to verification, the PCR products were purified using SAP (Amersham) and Exo I (USB, Cleveland, OH, USA). 2 U SAP was added and 21 U Exo I was added to 5 µl PCR product, according to the manufacturer's instructions. Each product was then incubated at 37°C for 60 min followed by 15min inactivation at 75°C.

3.2.8. SNaPshot® Extension

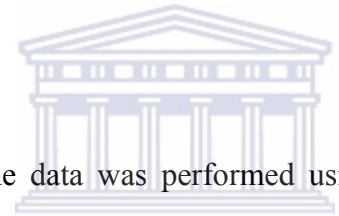
Primers used for single-base extension for Multiplex 1and 2 are listed in table 3.1 and 3.2. All extension primers were synthesised by Integrated DNA Technologies Munich, Germany. The mini-sequencing reactions were performed in a final volume of 10 µl. Each reaction tube contained 2 µl SNaPshot® ready mix (Applied Biosystems), 3 µl purified PCR product, 0.4 µM primer mix and 3.5 µl water was added. A negative and positive control was supplied in the mini-sequencing reaction and was used according to manufacturer's instructions. The samples were then loaded onto a GeneAmp 2700 thermal cycler (Applied Biosystems) for the mini cycle sequencing reaction. Cycling conditions were as follows: 25 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 30 s, followed by a 4°C hold.

3.2.9. Post Extension Purification

Subsequent to extension, the products were purified using SAP (Amersham). To each tube, 2 U SAP was added according to the manufacturer's instructions. Incubation at 37°C for 60 min followed by 15min inactivation at 75°C then followed.

3.2.10. Detection and Data Analysis

Amplicons were analysed on ABI Prism 3500 DNA sequencer. A master mix was set up containing 7.5 µl HiDi Formamide (Applied Biosystems) and 0.5 µl GeneScan Liz 120 (Applied Biosystems) size standard. Reactions were set up on 96-well plate supplied by Applied Biosystems. To each well a volume of 8 µl master mix was added together with 2 µl SAP-treated extension product. The plate was then centrifuged for 1 min at 3500 rpm and incubated at 95°C for 5 min, in a thermal cycler, to initiate DNA denaturation. Thereafter, the plate was loaded into the ABI sequencer (Applied Biosystems). The raw data generated was then analysed using Geneapper v4.1. software (Applied Biosystems).



3.2.10. Statistical Analysis

The statistical analysis of the data was performed using freely available software, including Haploview and Microsoft Excel applications GenAlEx 6.4.1 and SNPTools. GenAlEx, described by Peakall and Smouse (2006), is an in-MS-Excel application allows for various statistical analyses to be calculated and was the primary tool used for all statistical analyses. GeneAlEx was used to calculate allele frequency, the Hard-Weinberg Equilibrium and the Shannon and Fixation indices. GenePop-on-the-web was used to calculate the G Exact Test to compare genotypic differentiation between the Cape Coloured and other global populations. SNPTools was used to convert genotypic data to haploview format. SNPTools (Chen *et al*, 2009) is also an MS-Excel application which allows raw data to be converted into a format suitable for various population genetics software, including Haploview, PLINK, Phase and SNPHAP. Haploview was then used to determine linkage disequilibrium between the 20 SNPs of the SLC22A1 gene within the Cape Coloured population.

3.3. Results and Discussion

3.3.1. SNaPshot® Optimization and Genotyping

The SNaPshot® genotyping assay is one of the most accurate, efficient and rapid methods available for high throughput genotyping. It is highly consistent, reproducible and more flexible than many other systems (Quintans, *et al*, 2004). In addition, this technique also has the advantage of requiring very few optimization steps. For the two multiplexes developed for this study, minor optimization was required for one SNP in the PCR amplification step. The concentration of this primer pair was increased to 0.4 μM as a result poor amplification. Once increased, enough template DNA was amplified for the mini-sequencing reaction. Thereafter, good profiles were generated for each primer set.

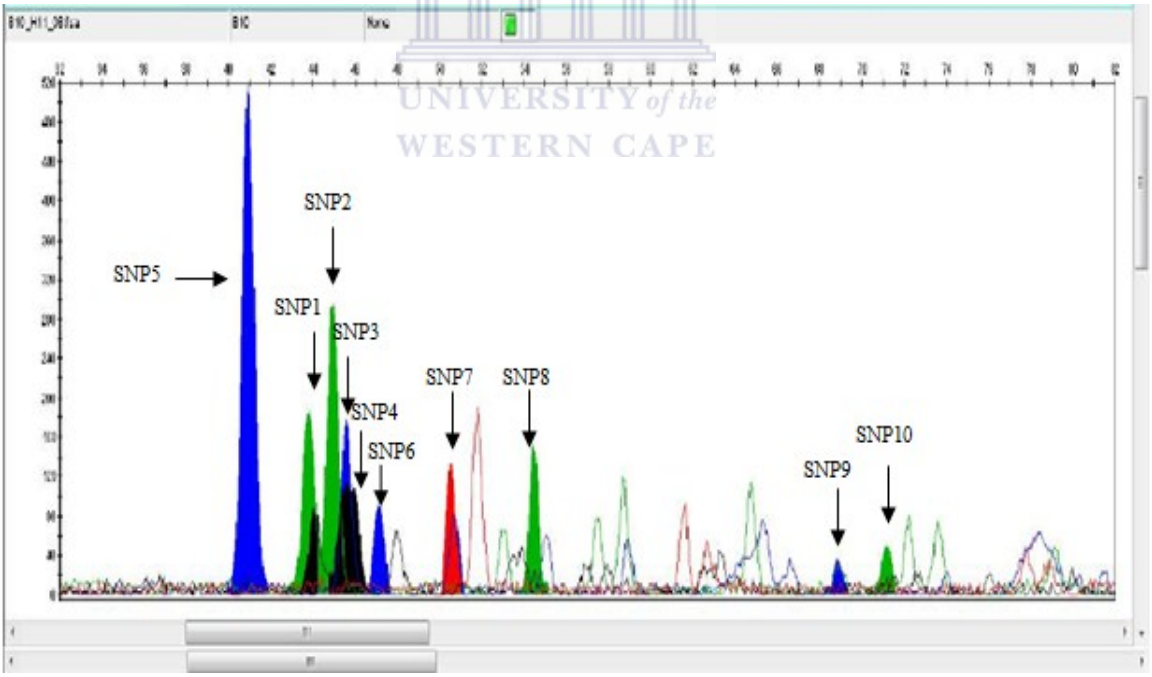


Figure 3.1: Typical Multiplex 1 electropherogram. Note a shift of 12bp and 20bp in SNP1 and SNP3 respectively.

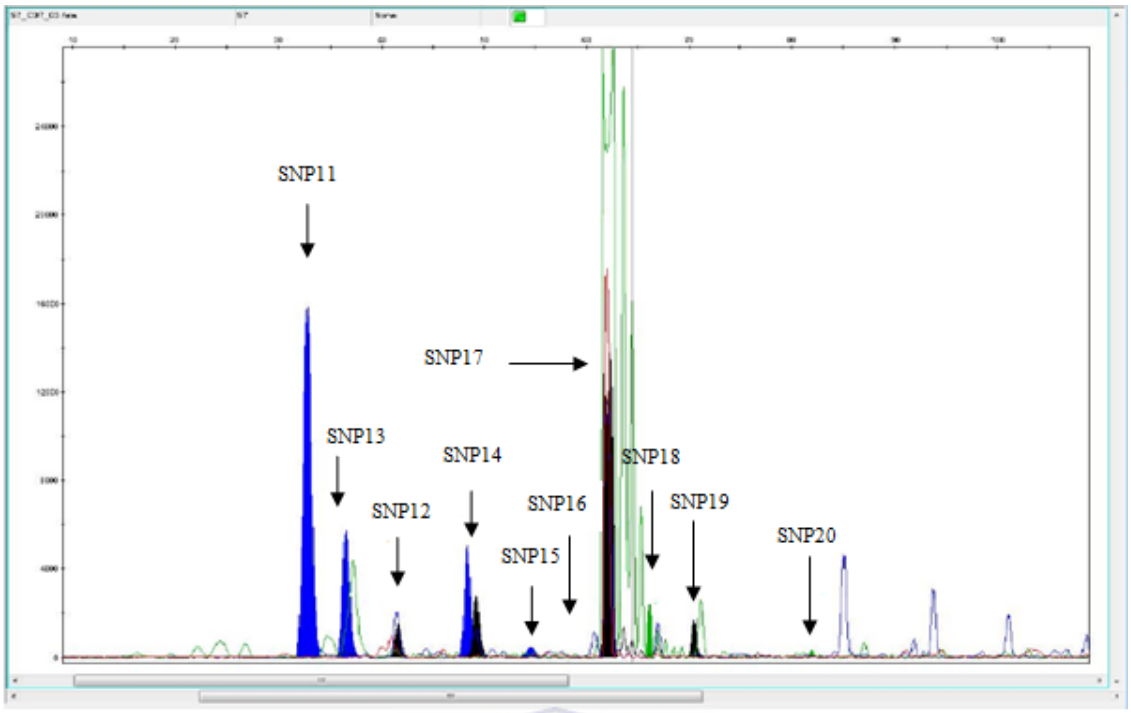


Figure 3.2: Multiplex 2 electropherogram. A large peak at position 58 to 65 for the A base was observed.

A certain amount of drift is common for fragments during capillary electrophoresis. This is as a result of the length, sequence and dyes used to label the ddNTPs. Subsequently, the electrochemical nature of the extension products is slightly altered and the fragment migrates further than its determined size (Quintans, *et al*, 2004). Typically smaller fragments will experience a greater amount of drift than others (Quintans, *et al*, 2004). However, as a result of the fragment drift, and the sensitivity of the ABI 3500 Genetic Analyser, some confusion became evident when attempting to call the genotypes. Some fragment had migrated to form clusters of peaks (Figure 3.1). Therefore, it was decided that the two multiplexes should be split into four by selecting and combining every alternate primer set for a total of 5 SNPs per multiplex. This yielded more accurate results and allowed for more accurate genotyping.

Table 3.3. Minor allele frequencies (MAF) of single nucleotide polymorphisms in Cape Coloured (C), African American (AA), Caucasian American (CA), Asian (A), Mexican (ME), Pacific Islander (P) and Sub-Saharan African (SSA). Minor allele frequencies for populations other than the Cape Coloured population were taken from the Pharmacogenomics Knowledgebase (http://www.pharmgkb.org), the UCSF-PMT (http://www.pharmacogenetics.ucsf.edu/) database and the NCBI-SNP database (http://www.ncbi.nlm.nih.gov)

RS ID	Short ID	Minor Allele	Polymorphism	Genotype	Population Group					
					C	***AA***	***CA***	***A***	***ME***	***P***
RS36103319	**SNP1**	**T**	**G > T**	n	**100**	198	200	60	20	14
				G/G	**96**	98	100	30	10	7
				G/T	**0**	1	0	0	0	0
				T/T	**4**	0	0	0	0	0
					0.020	0.005	0	0	0	0
RS12208357	**SNP2**	**C**	**T > C**	n	**100**	198	194	60	18	14
				T/T	**50**	99	83	30	8	7
				T/C	**0**	0	14	0	1	0
				C/C	**0**	0	0	0	0	0
					0	0	0.072	0	0.056	0
RS1867351	**SNP3**	**C**	**T > C**	n	**100**	198	194	60	18	14
				T/T	**68**	56	61	9	4	4
				T/C	**28**	34	32	16	5	2
				C/C	**4**	9	4	5	0	1
					0.180	0.263	0.206	0.433	0.278	0.286

Table 3.3 Continued

RS ID	Short ID	Minor Allele	Polymorphism	Genotype	Population Group					
					C	***AA***	***CA***	***A***	***ME***	***P***
RS34104736	**SNP4**	**T**	**C > T**	n	**100**	198	200	60	20	14
				C/C	**85**	99	99	30	10	7
				C/T	**1**	0	1	0	0	0
				T/T	**14**	0	0	0	0	0
					0.145	0	0.005	0	0	0
RS2282143	**SNP5**	**T**	**C > T**	n	**100**	196	200	60	20	14
				C/C	**84**	82	100	25	10	7
				C/T	**15**	16	0	5	0	0
				T/T	**1**	0	0	1	0	0
					0.085	0.082	0	0.117	0	0
RS34130495	**SNP6**	**A**	**G >A**	n	**100**	136	184	42	14	14
				G/G	**50**	67	92	21	7	7
				G/A	**0**	1	0	0	0	0
				A/A	**0**	0	0	0	0	0
					0	0.007	0.011	0	0	0

Table 3.3 Continued

RS ID	Short ID	Minor Allele	Polymorphism	Genotype	Population Group					
					C	***AA***	***CA***	***A***	***ME***	***P***
RS16891138	**SNP7**	**C**	**A > C**	n	**100**	200	198	60	20	14
				A/A	**50**	93	99	30	10	7
				A/C	**0**	7	0	0	0	0
				C/C	**0**	0	0	0	0	0
					0	0.035	0	0	0	0
RS35167514	**SNP8**	**Del**	**A > Del**	n	**100**	136	184	42	14	14
				A/A	**50**	64	60	21	5	7
				A/-	**0**	4	30	0	1	0
				-/-	**0**	0	2	0	1	0
					0	0.029	0.185	0	0.214	0
RS34305973	**SNP9**	**Del**	**T > Del**	n	**100**	136	184	42	14	14
				T/T	**74**	64	60	21	5	7
				T/-	**0**	4	30	0	1	0
				-/-	**26**	0	2	0	1	0
					0.130	0.029	0.185	0	0.214	0

Table 3.3 Continued

RS ID	Short ID	Minor Allele	Polymorphism	Genotype	Population Group					
					C	***AA***	***CA***	***A***	***ME***	***P***
RS2297374	**SNP10**	**T**	**C > T**	n	**100**	200	200	60	20	14
				C/C	**73**	71	77	27	9	7
				C/G	**2**	16	16	1	1	0
				G/G	**25**	13	7	2	0	0
					0.260	0.21	0.15	0.083	0.05	0
RS683369	**SNP11**	**G**	**C > G**	n	**100**	200	200	60	20	14
				C/C	**12**	99	88	29	9	7
				C/G	**0**	1	11	1	1	0
				G/G	**88**	0	1	0	0	0
					0.870	0.005	0.065	0.017	0.05	0
RS628031	**SNP12**	**A**	**G > A**	n	**100**	136	184	32	14	14
				G/G	**40**	4	15	2	0	0
				G/A	**49**	28	44	6	3	1
				A/A	**11**	36	33	13	4	6
					0.360	0.735	0.598	0.762	0.786	0.929

Table 3.3 Continued

RS ID	Short ID	Minor Allele	Polymorphism	Genotype	Population Group					
					C	***AA***	***CA***	***A***	***ME***	***P***
MPJ6_OC1004[a]	**SNP13**	**A**	**G > A**	n	**100**	-	-	-	-	-
				G/G	**75**	-	-	-	-	-
				G/A	**24**	-	-	-	-	-
				A/A	**1**	-	-	-	-	-
					0.130	-	-	-	-	-
RS34205214	**SNP14**	**A**	**G > A**	n	**100**	196	200	60	20	14
				G/G	**50**	93	100	30	10	7
				G/A	**0**	4	0	0	0	0
				A/A	**0**	1	0	0	0	0
					0	0.031	0	0	0	0
RS117474883	**SNP15**	**C**	**G > C**	n	**100**	-	-	-	-	-
				G/G	**0**	-	-	-	-	-
				G/C	**0**	-	-	-	-	-
				C/C	**50**	-	-	-	-	-
					1	-	-	-	-	-

Table 3.3 Continued

RS ID	Short ID	Minor Allele	Polymorphism	Genotype	Population Group					
					C	***AA***	***CA***	***A***	***ME***	***P***
RS35191146	**SNP16**	**Del**	**G > Del**	n	**100**	136	184	42	14	14
				G/G	**70**	64	60	21	5	7
				G/-	**30**	4	30	0	1	0
				-/-	**0**	0	2	0	1	0
					0.155	0.029	0.185	0	0.214	0
RS34059508	**SNP17**	**A**	**G > A**	n	**100**	200	200	60	20	14
				G/G	**50**	100	92	30	10	7
				G/A	**0**	0	8	0	0	0
				A/A	**0**	0	0	0	0	0
					0	0	0.04	0	0	0
RS115733275	**SNP18**	**T**	**G > T**	n	**100**	-	-	-	-	-
				G/G	**0**	-	-	-	-	-
				G/T	**0**	-	-	-	-	-
				T/T	**50**	-	-	-	-	-
					1	-	-	-	-	-

Table 3.3 Continued

RS ID	Short ID	Minor Allele	Polymorphism	Genotype	Population Group					
					C	*AA*	*CA*	*A*	*ME*	*P*
RS116824962	SNP19	T	C > T	n	**100**	-	-	-	-	-
				C/C	**50**	-	-	-	-	-
				C/T	**0**	-	-	-	-	-
				T/T	**0**	-	-	-	-	-
					0	-	-	-	-	-
RS118060798	SNP20	G	A > G	n	**100**	-	-	-	-	-
				A/A	**50**	-	-	-	-	-
				A/G	**0**	-	-	-	-	-
				G/G	**0**	-	-	-	-	-
					0	-	-	-	-	-

Zero (0) indicates a value of zero.
Dashes (-) indicate absence of published data.
Minor alleles and minor allele frequencies for other populations were determined from the 1000Genome reference data.
[a]: SNP 13 (MPJ6_OC1004) has been characterised by Itoda *et al.* 2004, only in the Japanese population. Minor allele frequency: 0.004.

3.3.2. Statistical Parameters

Statistical analysis plays an important role in population genetics as it allows population data to be summarized in a clear and concise format (Figure 3.3) (Mailloux and LaBerge, 2010). Population genetics data obtained for the investigated SNPs were analysed using various statistical analysis software to calculate important parameters, and possible pharmacogenetics implications were then discussed. GeneAlEx was used to calculate allele frequency, the Hardy-Weinberg Equilibrium and the Shannon and Fixation indices. GenePop-on-the-web was used to calculate the G Exact Test to compare genotypic differentiation between the Cape Coloured and other global populations. Haploview was used to determine linkage disequilibrium between the 20 SNPs of the SLC22A1 gene within the Cape Coloured population.

In Table 3.4 an assortment of figures and indices are given. Shannon's index (I) is a measure of the differentiation across the loci. However, no values can be calculated for any monomorphic loci. The value calculated in the table specified above, is indicative of the difference in allele frequency, where a large difference in the frequency of each allele at a locus is given by a low value. If the allele frequencies are more evenly distributed, at the specific locus, then the Shannon Index approaches 1. Only two of the variants displayed such extreme difference between the allele frequencies. SNP15 and SNP18 both exhibited allele frequencies of above 95% for the minor allele. This is indicative of a large variation when compared with the expected ancestral allele. It is suggested that the results produced here additionally indicate that the minor allele frequency far outweighs the major frequency.

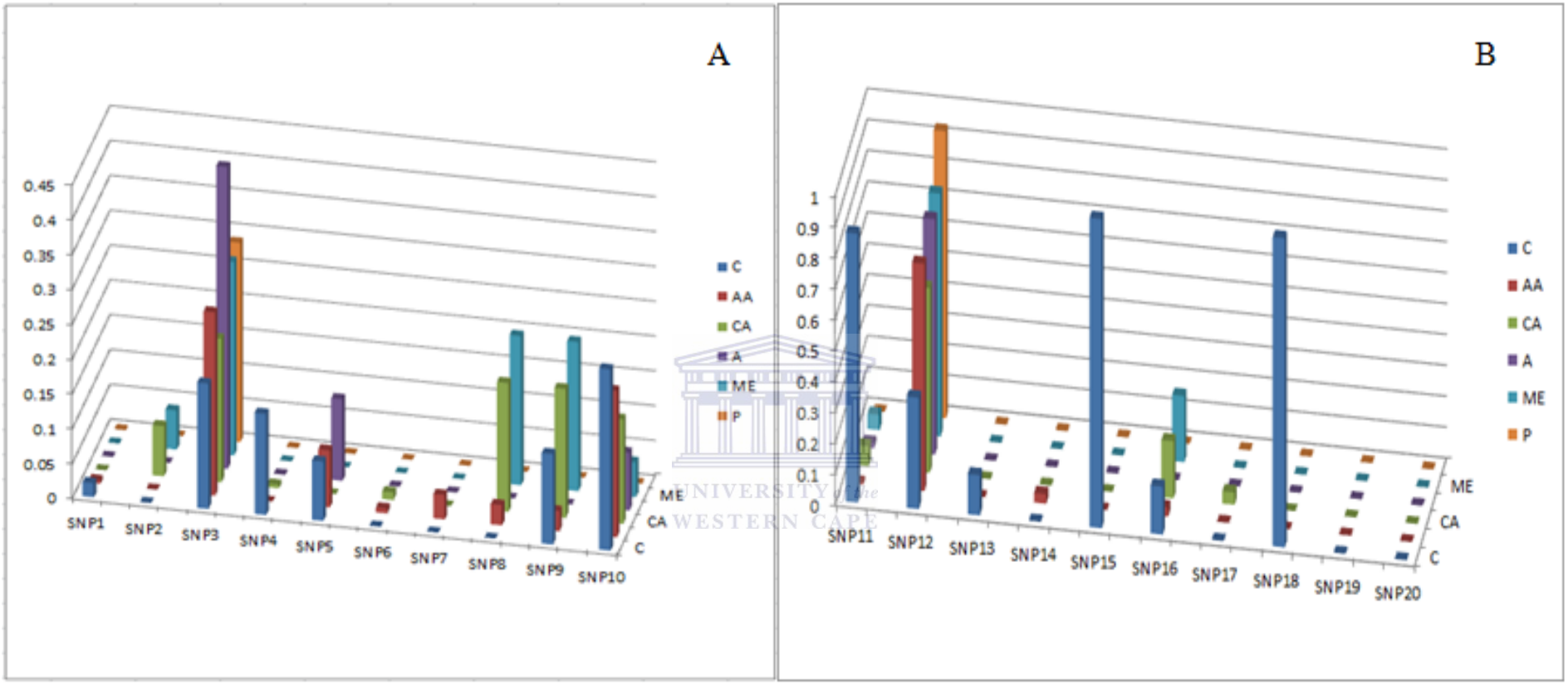


Figure 3.3: Minor allele frequencies across various populations. A: Multiplex 1 data. B: Multiplex 2 data. C – Cape Coloured, AA – African American, CA – Caucasian American, A – Asian, ME – Mexican, P – Pacific Islander, SSA – Sub-Saharan African.

Table 3.4. Summary of Heterozygosity and Fixation Indices generated using GenAlEx

Population	Locus	N	Na	Ne	I	Ho	He	UHe	F
Cape Coloured									
	rs36103319	100	2.000	1.041	0.098	0.040	0.039	0.039	-0.020
	rs12208357	100	1.000	1.000	0.000	0.000	0.000	0.000	#N/A
	rs1867351	100	2.000	1.419	0.471	0.280	0.295	0.297	0.051
	rs34104736	100	2.000	1.330	0.414	0.010	0.248	0.249	0.960
	rs2282143	100	2.000	1.184	0.291	0.150	0.156	0.156	0.036
	rs34130495	100	1.000	1.000	0.000	0.000	0.000	0.000	#N/A
	rs16891138	100	1.000	1.000	0.000	0.000	0.000	0.000	#N/A
	rs35167514	100	1.000	1.000	0.000	0.000	0.000	0.000	#N/A
	rs34305973	100	2.000	1.292	0.386	0.260	0.226	0.227	-0.149
	rs2297374	100	2.000	1.625	0.573	0.020	0.385	0.387	0.948
	rs683369	100	2.000	1.292	0.386	0.020	0.226	0.227	0.912
	rs628031	100	3.000	1.877	0.683	0.500	0.467	0.469	-0.070
	MPJ6_O1004	100	2.000	1.292	0.386	0.240	0.226	0.227	-0.061
	rs34205214	100	1.000	1.000	0.000	0.000	0.000	0.000	#N/A
	rs117474883	100	1.000	1.000	0.000	0.000	0.000	0.000	#N/A
	rs35191146	100	2.000	1.355	0.431	0.290	0.262	0.263	-0.107
	rs34059508	100	1.000	1.000	0.000	0.000	0.000	0.000	#N/A
	rs115733275	100	1.000	1.000	0.000	0.000	0.000	0.000	#N/A
	rs116824962	100	1.000	1.000	0.000	0.000	0.000	0.000	#N/A
	rs118060798	100	1.000	1.000	0.000	0.000	0.000	0.000	#N/A

Population		N	Na	Ne	I	Ho	He	UHe
Cape Coloured	**Mean**	100.000	1.550	1.185	0.206	0.091	0.127	0.127
	SE	0.000	0.135	0.056	0.053	0.032	0.034	0.034

Where: Na = No. of Different Alleles, Ne = No. of Effective Alleles = 1 / (Sum pi^2), I = Shannon's Information Index = -1* Sum (pi * Ln (pi)), Ho = Observed Heterozygosity = No. of Hets / N, He = Expected Heterozygosity = 1 - Sum pi^2, UHe = Unbiased Expected Heterozygosity = (2N / (2N-1)) * He, F = Fixation Index = (He - Ho) / He = 1 - (Ho / He), Where pi is the frequency of the *n*th allele for the population & Sum pi^2 is the sum of the squared population allele frequencies.

The Fixation index, also called the Inbreeding Coefficient, is a function of the heterozygosity of a locus. The indices shown here suggest that for some SNPs, SNP2, SNP6, SNP7, SNP8, SNP14, SNP15, SNP17, SNP18, SNP19 and SNP20, the population is indeed separate from others. However, this may also indicate that these SNPs are always inherited in the same manner in all populations and would, therefore,

not display any sort of genetic polymorphism. These indices also indicate that a large amount of inbreeding has occurred in this population. It may therefore be said that this population rarely interbreeds with other populations.

The expected and observed genotype frequencies are determined according to the Hardy-Weinberg principle (Table 3.5). The difference in observed and expected genotypes is indicative of the amount of external selective pressure applied by mutation, mating and random genetic drift. None of the loci showed an inclination toward the expected results determined by the Hardy-Weinberg principle. According to the HWE principle, a significant amount of mutation has occurred within the study samples. Although the observed genotypes, for most samples, conform to the standard set by the expected genotypes the actual weight for each genotype is far greater than expected. Therefore, with the exception of three SNPs (4, 10 and 11) the majority of SNPs were determined to be in HWE (Appendix II).

However, SNP4, SNP10 and SNP11 displayed very low or no amount of heterozygosity. SNP10 displayed an observed heterozygosity value of 0.010, which equates to 2% of the subject population. For the remaining SNPs (10 and 11), individuals were homozygous for either the minor or major allele. These results may be said to be the cause of them being designated as significant as neither case, very low heterozygosity and homozygosity, fits the Hardy-Weinberg model.

Table 3.5. Hardy-Weinberg Chi-Square test

Locus	DF	ChiSq	Prob	Signif
rs36103319	1	0.042	0.838	ns
rs12208357	Monomorphic			
rs1867351	1	0.265	0.607	ns
rs34104736	1	92.097	0.000	***
rs2282143	1	0.127	0.721	ns
rs34130495	Monomorphic			
rs16891138	Monomorphic			
rs35167514	Monomorphic			
rs34305973	1	2.233	0.135	ns
rs2297374	1	89.875	0.000	***
rs683369	1	83.098	0.000	***
rs628031	2	2.350	0.503	ns
MPJ6_O1004	1	0.372	0.542	ns
rs34205214	Monomorphic			
rs117474883	Monomorphic			
rs35191146	1	1.147	0.284	ns
rs34059508	Monomorphic			
rs115733275	Monomorphic			
rs116824962	Monomorphic			
rs118060798	Monomorphic			

Where DF indicates the amount of the number of alleles other than the ancestral allele and * indicates P<0.001**

Interestingly, SNP11 remained a constant feature in the statistical analyses presented. Under the Shannon and Fixation indices, this SNP was found to have high statistical significance. However, this may be due to the high amount of individuals homozygous for the minor allele. It may be suggested that this SNP will have to be screened against other populations to determine if it is always inherited in the same manner or if this population is completely separate from others, with regard to this SNP. In table 3.6 the SNPs are compared between populations to highlight the similarity between these populations. SNP4 and SNP11 are shown to be completely different from all populations. This would suggest that, in relation to these SNPs, the Cape Coloured population is indeed separate from all other populations screened.

Table 3.6. Population differentiation between Cape Coloured and global populations. Significant data is shown only. All monomorphic or absent data has been removed. Where C – Cape Coloured, CA - Caucasian American, AA – African American, A – Asian, ME – Mexican and P – Pacific Islander

	SNP ID	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10
Population Comparison											
	CA & C	0.36892	0.01281	1.00000	0	0.00005	0.49770	0.01354	N/A	N/A	0.04704
	AA & C	0.12294	N/A	0.05089	0	1.00000	1.00000	N/A	N/A	N/A	0.43133
	A & C	0.12141	N/A	0	0	0.43272	N/A	N/A	N/A	N/A	0.00051
	ME & C	0.12081	1.00000	0	0	0	N/A	N/A	N/A	N/A	0
	P & C	0.11981	N/A	0	0	0	N/A	N/A	N/A	N/A	0

	SNP ID	SNP11	SNP12	SNP13	SNP14	SNP15	SNP16	SNP17	SNP18	SNP19	SNP20
Population Comparison											
	CA & C	0	0.47494	N/A	N/A	N/A	N/A	0.00626	N/A	N/A	N/A
	AA & C	0	0.03446	N/A	0.05774	N/A	N/A	N/A	N/A	N/A	N/A
	A & C	0	0.00625	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	ME & C	0	0.00073	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	P & C	0	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Where N/A indicates no published data present.

SNPTools was used to convert the raw data to a format acceptable for use with Haploview. Although SNPTools incorporates various statistical functions, GenAlEx was found to be more beneficial for complex analyses. Once converted into Haploview format, a Linkage-disequilibrium (LD) map could then be generated. Several approaches for viewing the data may be harnessed using the software. As a default, LD maps are displayed under the standard D' colour scheme. Diagonal lines represent different SNPs while the boxes represent the pairwise comparison between two SNPs (Figure 3.4). This scheme arranges the linkage between SNPs based on a scoring system designed to make use of the Log of Odds Ratio (LOD) and D-prime (D') value for each SNP combination (Barrett *et al*, 2005).

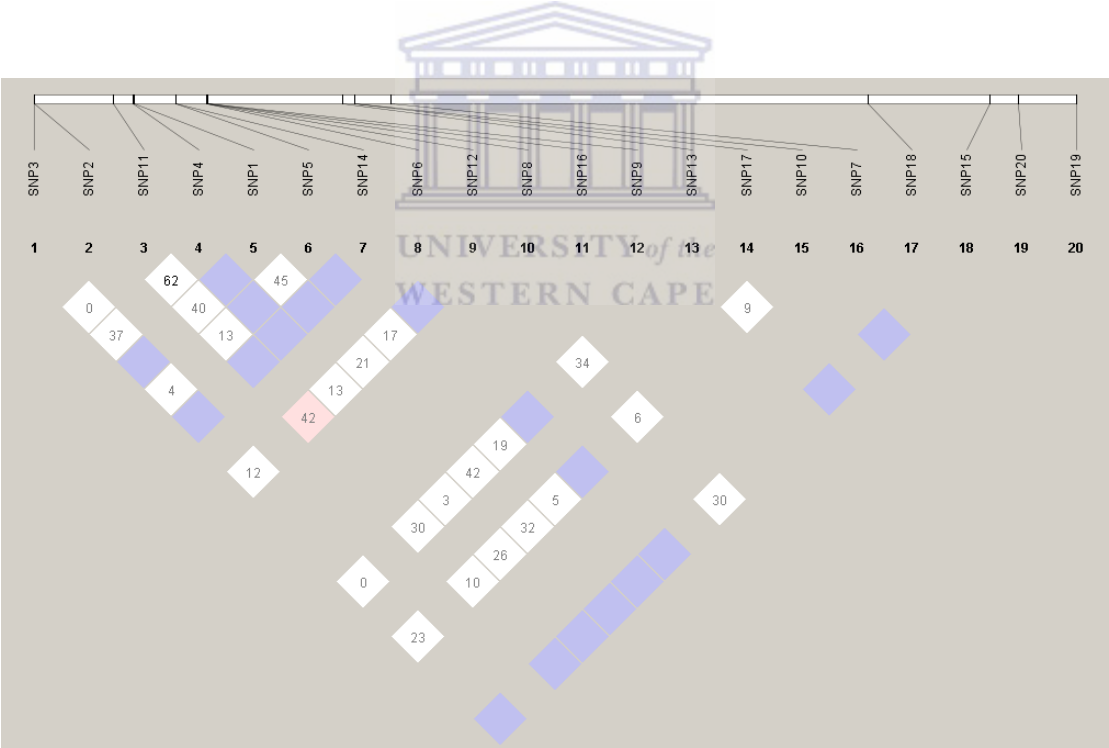


Figure 3.4: LD map of the genotype frequencies for all 20 SNPs. Gaps are seen as a result of the presence of homozygous samples.

Refining the input data to include only those SNPs which are heterozygous, a more accurate depiction of the LD is produced (Figure 3.5).

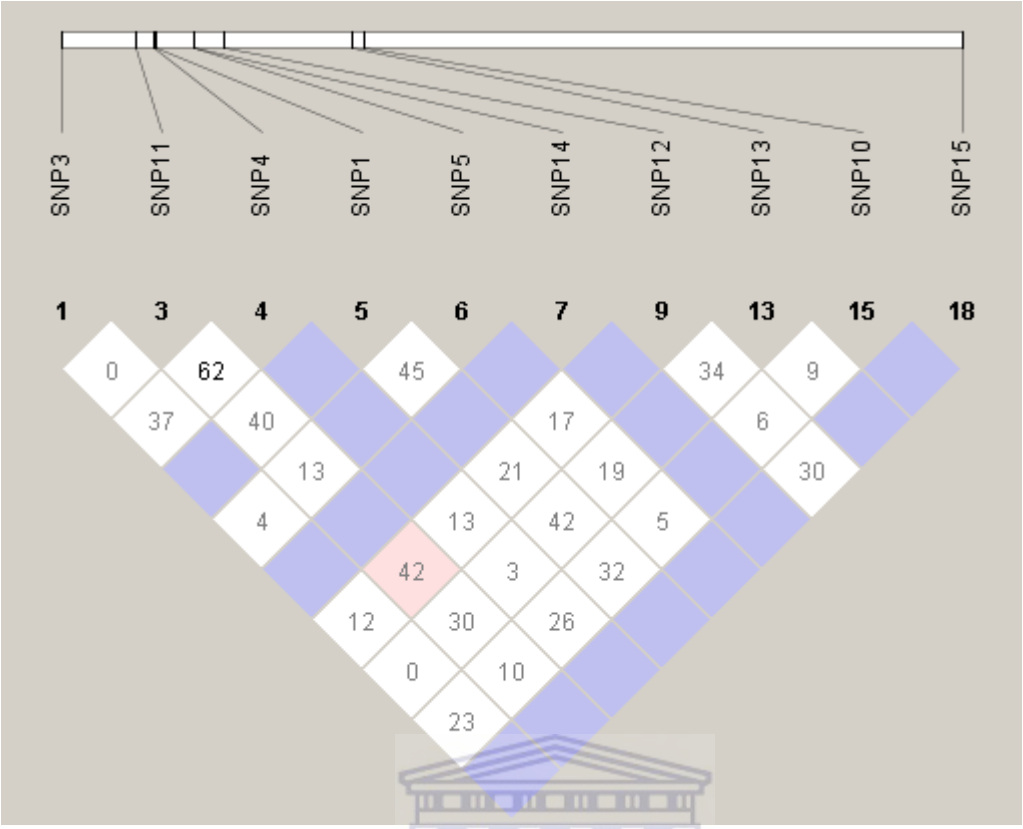


Figure 3.5: LD map of the 10 heterozygous SNP markers.

In Figure 3.5, 2 pairwise comparisons, blocks 62 and 42, display an amount of LD. The block labelled 42, a pairwise comparison between SNP 11 and SNP 12, has a D' value of 0.422 and LOD of 2.53 thus indicating minor LD. The same can be said for block 62, comparison between SNP11 and SNP4. However, because the D' value is 0.62 and the LOD score is 0.76, the LD exhibited by this block is even weaker than the aforementioned block (block 42). Therefore, the degree of LD that existed between these comparisons can be said to be of minor significance as an LOD score of 3 is considered evidence for linkage. As described above, the blue block indicate perfect LD scores, however, they are not statistically significant.

Under a different setting (Figure 3.6), the same data can be analysed in a more simple fashion. The map generated displays the data based on the confidence bounds set by

the user. This method is also reliant on the D' values for each pairwise comparison but does not take the LOD score into account. Dark Grey squares indicate strong evidence of LD while white squares are indicative of a strong existence of recombination. Light grey squares (not seen here) are deemed uninformative. Figure 3.6, below shows that the same marker comparisons exhibit similar LD. This may be due to the program taking only the D' value into account. Thus, all the blue blocks, D' value = 1, are deemed to exhibit strong LD. Interestingly, this method confirmed the amount of recombination that occurs at all other blocks.

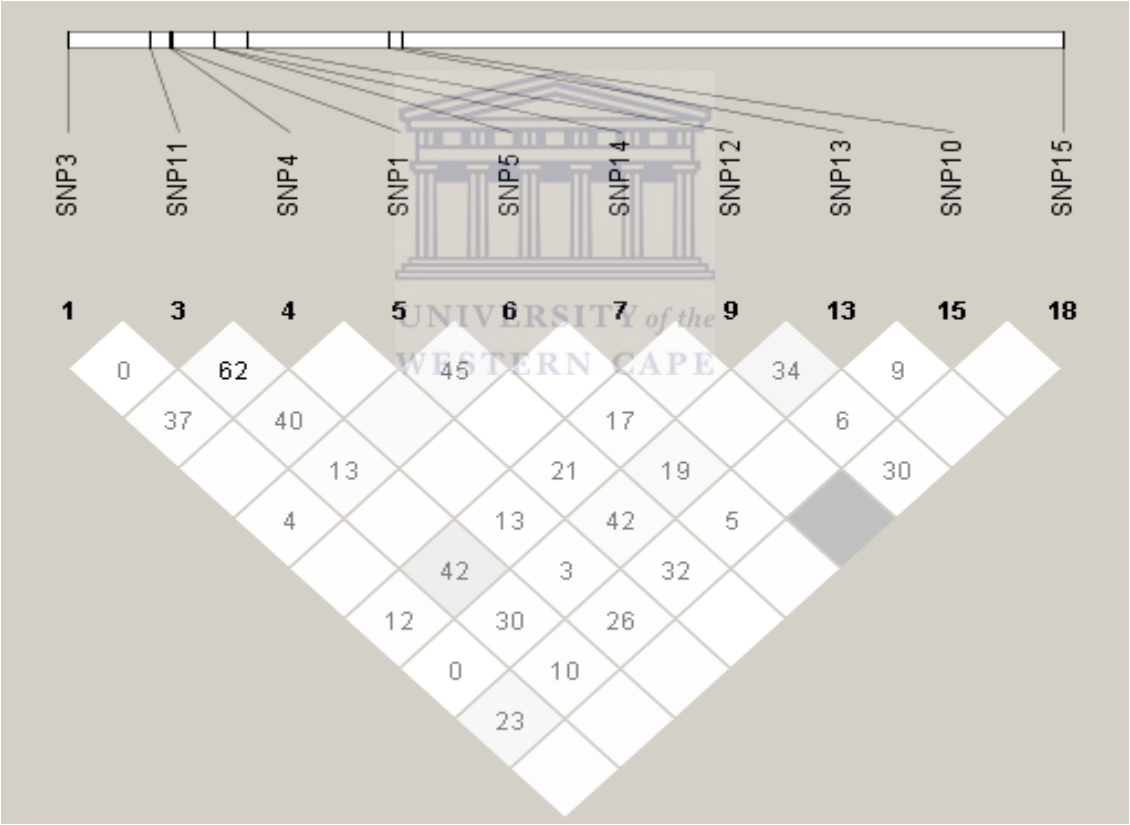


Figure 3.6: LD map of the 13 heterozygous SNP markers under Confidence Bounds setting. A simpler method for determining the meaning of each block.

Interestingly, since the commencement of this study, several SNPs have subsequently been removed from all databases. SNP 15 (RS117474883) and SNP 20 (RS118060798) have been retracted from all known databases. The reason for the

retraction of these SNPs remains unclear. SNP 13 (MPJ6_OC1004), identified in the Japanese population by Itoda *et al*. (2004), and cited by Shikata *et al*. (2007), appears only in the National Institute of Health Science (NIH) Japanese database. Therefore, population data only exists for the Japanese population as this SNP has not been investigated in other populations. It has been demonstrated, herein, that a significant percentage of the Cape Coloured population are homozygous for the minor allele at this locus (Table 3.3). However, the effect that this variation imparts has not yet been determined by any preceding study but it has been established that the amino acid change is non-synonymous.

The SNPs investigated in this study displayed many similarities to data published for the global population. However, for certain SNPs the statistics indicate significant differences between our population and other populations. SNPs 1, 2, 4, 6, 7, 11, 14 and 17 were homozygous in all populations and statistically identical, with the exception of the Caucasia American population in which numerous heterozygotes occurred across the SNPs. No published population data was available for SNPs 15, 19 and 20. For SNPS 3, 4, 5, 6, 7, 10, 12, 14 and 17 our population displayed a significant amount of similarity to all other populations investigated (Table 3.6) (Appendix II). Conversely, SNP 11 was determined to be different, in our population, to all other populations as 88% of the population displayed homozygosity for the minor allele. Similarly, SNP 18 was shown to be homozygous for the minor allele in 100% of the Cape Coloured population. However, no other published population data exists for this SNP.

The online databases (PMT, PharmGKB, and NCBI) provide a vast magnitude of information on most of the SNPs analysed herein. However, for some SNPs no information is available for any population other than African American. For example, the allele frequencies determined for SNP11 and SNP18 were contrary to what was published in the databases. This may be said to be a result of the unique genetic influence inherited from the indigenous people of the Cape. Currently, many studies are in process analysing the same SNPs, using a multitude of methods, for the same purpose, to generate population data. More data is constantly being produced. This may be a supporting reason for retracting SNPs from the databases if they are determined to be the result of an error in a study.

In the table 3.3., SNP11 was determined to have an allele frequency exactly opposite to that described by the NCBI database for other populations. Only one other population in the database exhibits the same allele frequencies as determined herein for that SNP. The dataset used by the database was generated from the genetics material sampled from Bushmen of the Northern and Southern Kalahari. It is noted that, although they are comprised of two distinct populations, the KhoiSan were indigenous to the Cape and form a large part of the genetic ancestry of most Cape Coloured people. Therefore, it is not surprising that the data generated herein is confirmed by the database specifically for this population and the subject population. Moreover, this indicates that the Cape Coloured population is indeed a separate population from others, with regard to specific SNPs (SNP18 and SNP11), as these variants seem only to be found in this population.

3.3.3. Pharmacogenetic Implications of Results

Cape local communities bear immense genetic diversity. This is a result of the arrival of various settlers sent to colonize the Cape. The interbreeding between foreign settlers and the indigenous people at the Cape gave rise to a highly genetically diverse population over hundreds of years. The resulting population, existing today, thus is comprised of several haplotypes. Therefore, it has become apparent that this genetic diversity would result in an increase in interindividual variability, specifically with regard to drug uptake and metabolism.

When compared to the online databases available it becomes apparent that the Cape Coloured population shares largely the same genetic characteristics as the global population. It is therefore safe to assume that as a result the interbreeding of various ethnicities, a large proportion of the global population is represented in this population. Furthermore, only a small percentage of the results do not correspond to the global statistics. Twelve of the SNPs were determined to have non-synonymous amino acid substitutions. However, published finding for only 7 of these SNPs were available from the databases.

SNP1, SNP2 and SNP6, amino acid changes R61C, G220V and G401S respectively, have been found to fully abolish OCT1 activity and reduce the uptake of both metformin and MPP^+ in transfected cells (Shu, *et al*, 2003; Shu, *et al*, 2007, Tzvetkov *et al*, 2009). SNP4, with amino acid variant S189L and described by Shu, *et al*. (2007), has been determined to reduce the uptake of metformin in transfected cells. It has been determined that SNP8 and SNP9 belong to a cluster of single base deletions where each deletion causes the same amino acid variation, namely M420del. This

cluster has been shown to partially reduce metformin uptake in transfected cells (Shu, *et al*, 2007, Tzvetkov *et al*, 2009). It has been established that SNP5, amino acid variation P341L, described by Shu, *et al*. (2003), results in reduced MPP^+ uptake in oocytes. Interestingly, SNP5 was determined, in this study, to have a minor allele frequency of 1%. Thus, indicating that the polymorphism is present in 1% of the samples analysed. However, since all the samples were collected from male donors, the effect this polymorphism may have on those individuals remains undetermined.

It may therefore be suggested that having a combination of these 7 SNPs may reduce the efficacy of Metformin and increase its toxicity (Shu *et al*, 2007). It may also be suggested that the combination of SNPs may be responsible for individuals to display no response to the drug (Tzvetkov *et al*, 2009).

Interestingly, the individual's genotype also has a significant impact on the efficacy of the drug (Tzvetkov *et al*, 2009). In most instances, different effects will be seen for homozygous and heterozygous individuals. In a study by Tzvetkov *et al*. (2009), the renal clearance of metformin was determined in relation to the genotype displayed by Caucasian individuals. It was determined that 90% of individuals homozygous for the minor, or inactive, allele of each SNP displayed a significantly greater rate of renal clearance, thus implying a loss of OCT1 activity and by extension a loss in metformin uptake. Furthermore, those heterozygous individuals displayed only a slightly increased rate of renal clearance. Conversely, for SNP10, it was determined that heterozygous individuals displayed a greater increase in the rate of renal clearance than those homozygous for the minor allele.

The findings presented herein suggest that for many of the drugs regulated and taken up by OCT1, the customization with regard to individualised medicine will remain constant for many populations across the world. However, it is those minor differences which ultimately lead to improved uptake and metabolism of drugs. Thus, from a pharmacogenetics stand point, the percentage of the results which do differ from published findings will be most crucial to the development of better medications.

Since only two of the SNPs differ significantly from the database, and these two are primarily found in individuals descendant from Bushmen, it may be suggested that these SNPs will ultimately decide the degree of individualisation of specific drugs. Furthermore, SNP11 has been determined to be a non-synonymous SNP where the amino acid residue at position 160 is altered from Phenylalanine to Leucine. The effect that this has on the uptake of any drug, or endogenous compound, has not been determined as of yet. It may therefore be speculated, since no finding to the contrary have been published yet, that this SNP will require more detailed investigation in order to determine its effect. No data has been published for SNP18. Therefore, the effect produced by any variation at this site is currently unknown. However, it is suggested that future studies conducted should include this SNP in order to characterise the effect it causes, if any, since its incidence is relatively high specifically in the Cape Coloured population.

3.3.4. Summary

To summarize, the above mentioned method has proven to be a both rapid and accurate method for high throughput genotyping. Very little optimization was

required but further optimization of this system is possible. The results obtained were of high quality and accurate and showed a great deal of statistical significance. Furthermore, the software available for the analysis of population data played a pivotal role in the completion of this study and produced excellent results. Moreover, the findings presented herein established not only that the Cape Coloured population shares genetic polymorphisms with the global population but also that this population presented its own challenges and polymorphisms. It may be concluded that the importance of population data, from a pharmacogenetic perspective, is critical to the improvement of existing medical practices and the development of new ones.

Chapter Four

Development of High Resolution Melt Analysis Genotyping System

4.1. Introduction

High Resolution Melt (HRM) analysis is a mutation scanning technique which is able to monitor the progressive change in fluorescence caused by the release of intercalating dye during denaturation (Wittwer *et al*, 2003). It is an in-tube method requiring only the addition of a saturating intercalating dye and a high resolution melting step, specified as a stepwise marginal increase in temperature, after PCR amplification (Wittwer *et al*, 2003).

Different PCR products generally have different melting temperatures. This is dependant on their GC content, length and sequence (Ririe *et al*, 1997). Therefore, this method presents an ideal means for SNP genotyping since the above mentioned parameters are known and a theoretical melting temperature can be calculated. Formerly, temperature resolution was limited, making SNP genotyping using this method highly inaccurate (von Ahsen *et al*, 2001). However, with the recent introduction of new techniques, assays and dyes, SNP genotyping has become routine (Wittwer *et al*, 2003; Zhou *et al*, 2005). Most homozygous sequence variants produce a melting temperature (T_m) shift when compared to that of the wild type (Liew *et al*, 2004; Palais, *et al*, 2005). In contrast, heterozygotes are not identified by the T_m, but by the shape of the melting curve profile (Graham *et al*, 2005). A growing interest in HRM has been observed with many new applications being developed.

It is therefore the aim of this study to develop an HRM-based method for SNP genotyping and by doing so confirm the newly reported SNPs on exon 4 of OCT1in the Cape Coloured population.

4.2. Materials and Methods

4.2.1. Sample Collection

Samples were collected as described in Chapter 1.

4.2.2. DNA Extraction and Quantification

DNA was extracted and quantified as described in Chapter 1.

4.2.3. Site Selection

A newly reported 600bp region on exon 4 of the *OCT1* gene was selected for pre-sequencing scanning. This covered two SNP sites, namely RS34163122 and RS35446461.

4.2.4. Primer Design

All DNA samples were diluted to a final concentration of 5 ng.μl^{-1} and used for the HRM analysis. The primers used for the HRM analysis are listed in tables 4.1. All primers were synthesized by Integrated DNA Technologies (IDT) Munich, Germany. Primers were diluted to a working stock of 2 µM.

Table 4.4. HRM primer sequences for newly reported Exon 4 SNPs

SNP Name	Sequence	Amplicon Size
RS_34163122	F-CACAAGAGAGAAGCCTGGGAG R-GGTACATGATCGCCACCGTTC	320bp
Rs_35446461	F-GAACGGTGGCGATCATGTACC R-AGGAAGGGCCTCACCAGTAG	280bp

4.2.5. HRM Analysis

HRM was performed using the Quiagen Type-IT kit in a final volume of 10 µl according to manufacturer's instructions. PCR amplifications were performed using 5 µl of the Qiagen Type-IT, 2 µM Primers and 15 ng DNA adjusted to a final volume of 10 µl using RNA-free water. HRM analysis was performed on a RotorGene Q (WhiteSci) real-time PCR thermocycler. Two-step thermocycling conditions were as follows; 95°C for 10 min, 94°C for 30 s and 55°C for 60 s for a total of 35 cycles, followed by a timed melt beginning at 80°C and sequentially increasing in temperature at a rate of 0.1°C to reach a final temperature of 95°C. Melt profiles were then analysed using the RotorGene data collection software and ScreenClust HRM software (WhiteSci).

4.3. Results and Discussion

Two newly reported, and non-annotated, SNPs were from exon 4 of the OCT1 gene were selected from the UCSF-PMT database. These SNPs were selected as they are within 600 bp of each other. The ideal range for HRM-based polymorphism detection is approximately 350 bp; therefore, each SNP could be amplified and analysed individually at an amplicon size of between 250 and 300 bp.

No optimization of the reaction conditions was required. However, primer design played a significant role in the successful amplification and melting of the targets. A minimum of three primer sets were designed for each SNP. Of the three sets a forward and reverse primer set was chosen which displayed the highest specificity, lowest amount of complementarity and lowest amount of secondary structure formation. A total of 40 samples were genotyped for RS34163122 and 20 samples for RS35446461.

The data collection software provided with the RotorGene Q provided an excellent insight into the potential of the thermocycler. Without any third party software or specifically focused software from WhiteSci, the genotype of each locus could be observed. The sensitivity of the thermocycler allowed for the differentiation between multiple alleles at a specific locus (Figure 4.2). Furthermore, differentiation between homo- and heterozygotes was well observed on the data difference plots (Figure 4.2).

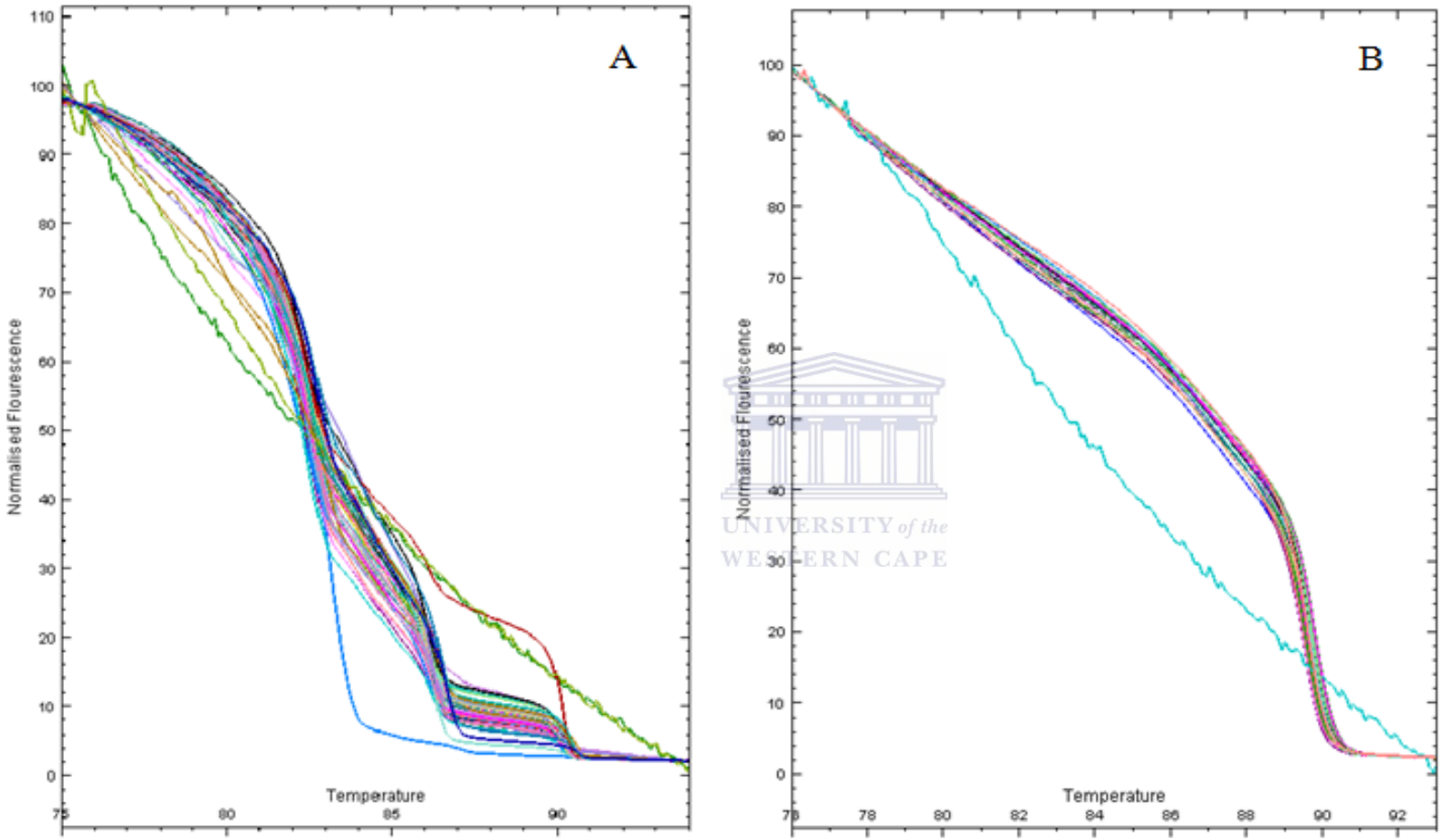


Figure 4.1: Raw fluorescence plots for each of the SNP amplifications. A: RS34163122, 6 reactions displayed late amplification where as one displayed no amplification. B: RS35446461, only one reaction displayed no amplification.

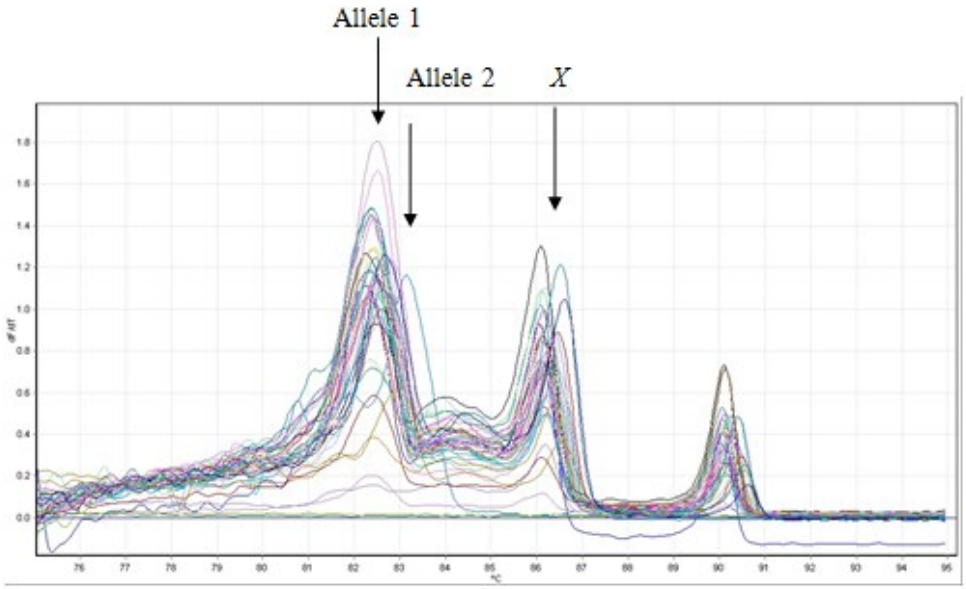


Figure 4.2: RS34163122 HRM mutation screen. Alleles 1 and 2 are indicated at the target site. Note a secondary target marked *X*.

In Figure 4.2 a series of peaks are seen. The target, and largest, peak is labelled by allele 1. Although the position of the peak for the second allele seems relatively close to those of allele 1, it is indicative of the class of the sensitivity of the machine.

However, the difference plot (Figure 4.3) provides a greater detail with reference to the difference between wild type and mutant genotypes. In this instance wild type and mutant refers to allele 1 and allele 2 respectively. Only the target peak is shown in this plot.

Interestingly, a secondary peak was observed (Figure 4.2). This may be indicative of a novel SNP within the Cape Coloured population. Further sequencing will be required to confirm the presence of a novel SNP. The unlabelled third peak was determined to be a non-specific PCR artefact and had no significant detrimental effect on mutation scanning.

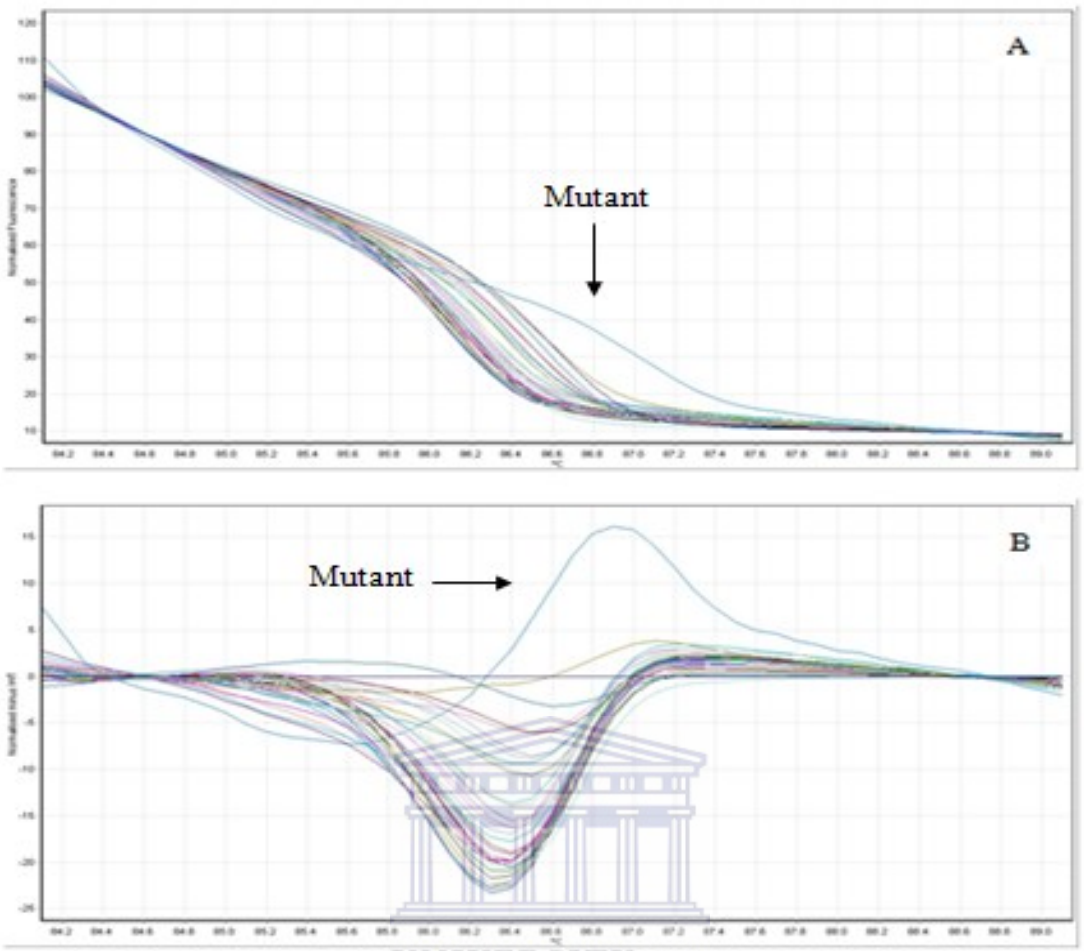


Figure 4.3: Normalised plot (A) and difference plot (B) for RS34163122 using the wild type (WT) genotype as a reference. Note the larger divergence depicted by the mutant (Mut) in blue.

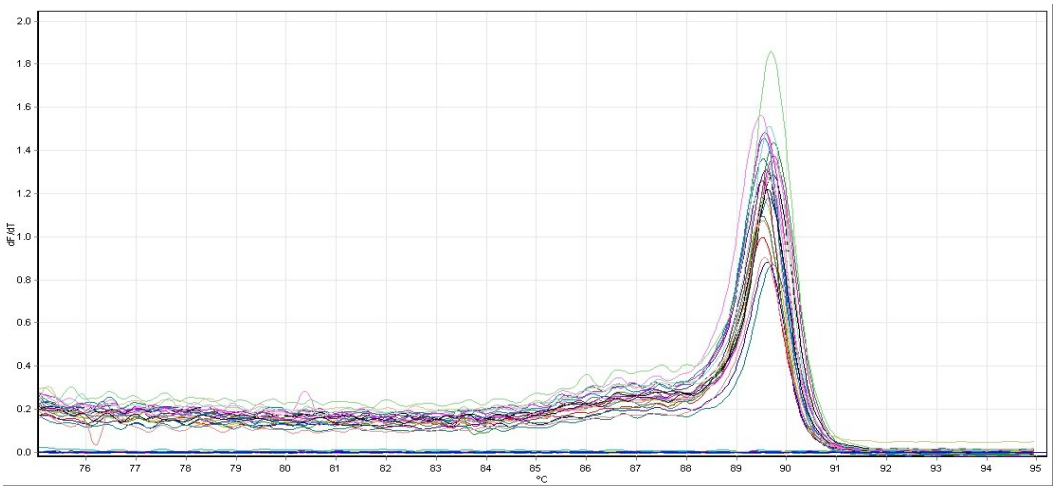


Figure 4.4: RS35446461 HRM mutation scan. A single target peak is observed. This SNP was determined to be homozygous in previous chapters in this study.

No significant mutations were detected for RS3544646 (Figure 4.5). Both the normalised data and the difference plot data suggest that no mutations were identified. Therefore, it is assumed that all samples exhibit the same genotype, homozygosity for allele 1.

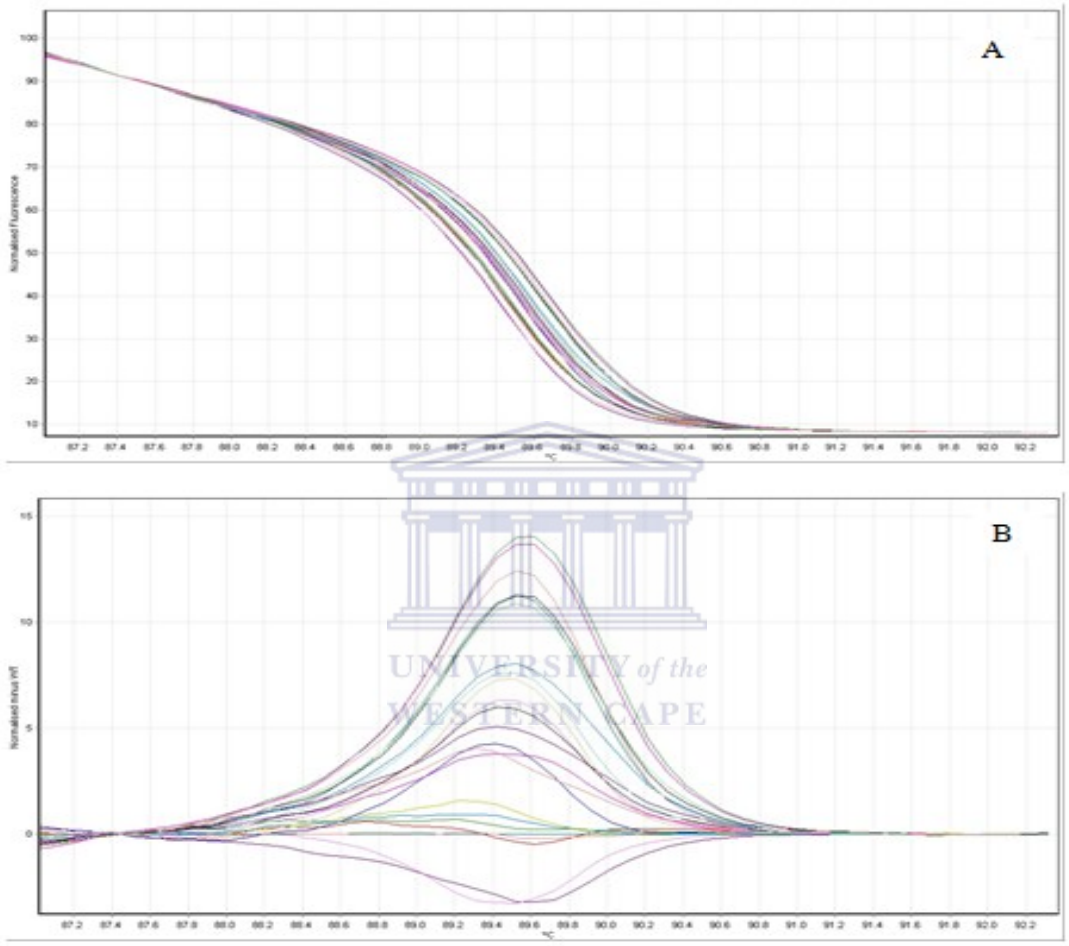


Figure 4.5: Normalised plot (A) and difference plot (B) for RS3544646. All samples exhibit wild type genotype (Alelle 1 homozygous).

The ScreenClust software, supplied by WhiteSci, provided further tools for analysis. Using this software, samples can be clustered into specific genotypes (Figure 4.4 and 4.5). Furthermore, the software also clustered those samples which displayed incomplete amplification or no amplification at all.

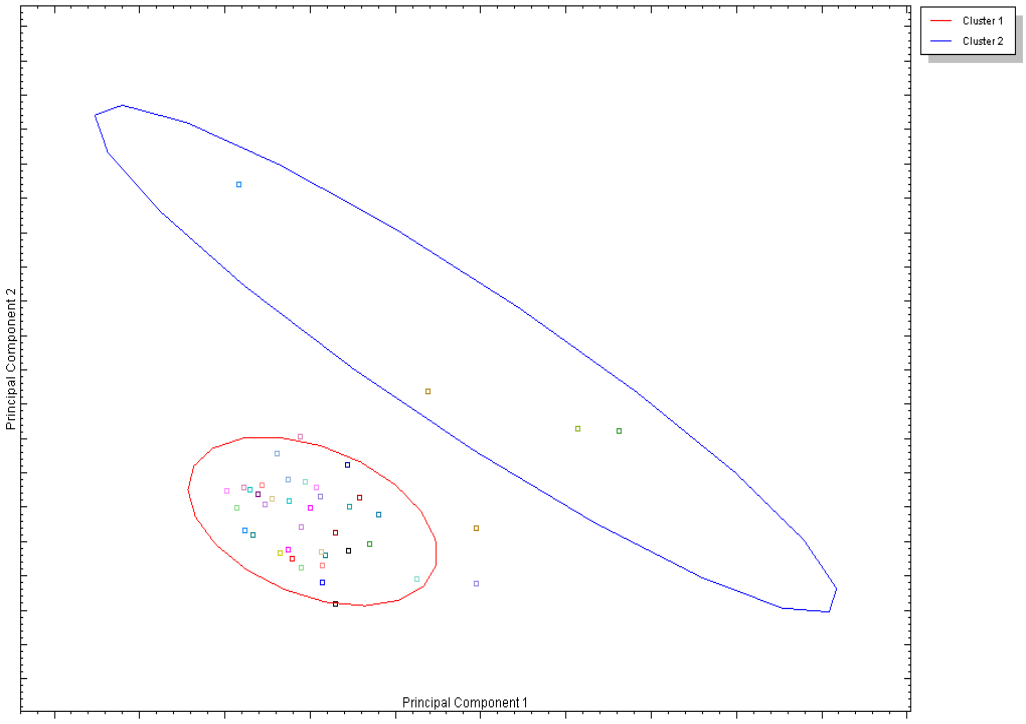


Figure 4.6: Clustered data for RS54163122 using ScreenClust. Note the samples falling inside of the blue cluster. These samples underwent incomplete amplification resulting in the formation of non-specific or incomplete products.

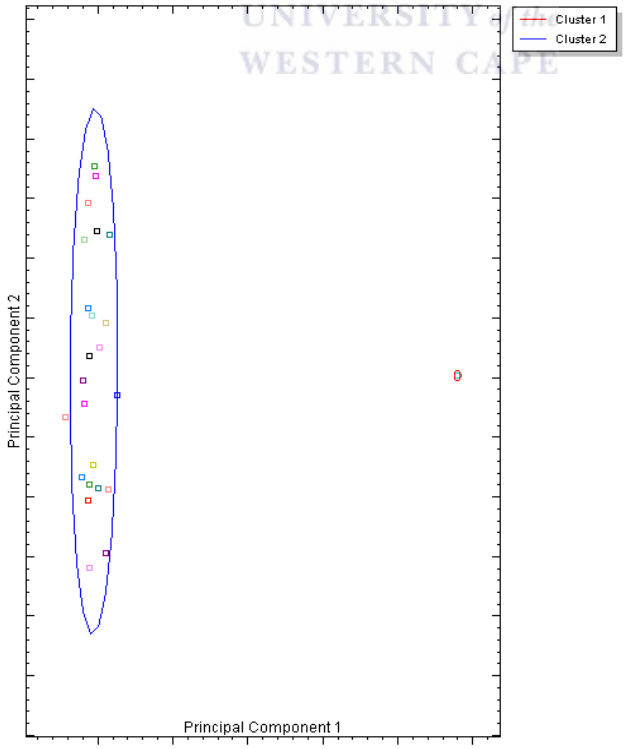


Figure 4.7: Clustered data for RS35446461 using ScreenClust. Note only one sample appears in the red cluster as it did not amplify at all.

In Figure 4.6 two samples were found outside of the cluster zones. These samples displayed alleles other than the ancestral for RS34163122. Therefore, these samples appear at the same position as the red cluster but occur at different melt temperature. A similar conclusion may be drawn for the single outlying sample in Figure 4.7. However, in this instance a deficiency of initiation of amplification resulted in no amplification occurring for this sample.

The purpose of this study was to introduce the precursory design and development of an HRM-based genotyping system. From the results obtained it is evident that the method presented herein would require optimization in order to compensate for the minute differences in in peak position. However, this study serves only to introduce the method for HRM genotyping. It is noted that this technique will successfully generate SNP data in a low-cost, rapid and proficient manner. In future work the addition of multiplex HRM will be introduced in order to develop and provide an HRM-based high-throughput genotyping assay.

4.4. Summary

The method presented herein serves only to introduce the capabilities and possibilities of HRM analysis. It is suggested that in future more complex techniques will incorporated into this one to develop a more comprehensive genotyping assay. However, as an introductory section, the method presented herein is considered to be successful and it will be beneficial to further develop the efficacy and sensitivity of the assay.

Chapter Five

Conclusion

Genomic diversity within sub-Saharan Africa, and for that matter the entire African continent, is relatively under-studied, despite being home to significant portion of human genomic diversity (Hardy *et al*., 2008). South Africa in particular contains a wealth of different population groups. The country is indeed home to the indigenous Khoisian, Xhosa, Zulu, Venda, Sotho and Pedi groups, the Afrikaners and the Cape Coloured, the latter being a uniquely admixed population of immigrant Europeans, Asians and the indigenous populations (Hardy *et al*., 2008). Despite this fact, it is well recognized that African populations have been underrepresented in global pharmacogenomics research.

The study of the genetic diversity of the solute carrier transporter genes and its pharmacogenetic implications within the South African and Sub-Saharan Africa will contribute in filling the gap of missing important pharmacogenetics data from African populations (Benjeddou, 2010). This includes establishing baseline frequency distribution of previously reported alleles for the SLC22A1, SLC22A2, SLC22A3, SLC47A1 and SLC47A2, SLC22A4 and SLC22A5 genes, as well as the discovery of new genetic variants within the investigated populations for these members of the super-family of solute carrier transporters.

For the present study, described in this thesis, important population pharmacogenetic data was generated for 20 SNP variants from the SLC22A1 gene within the Cape Coloured population of South Africa. Among others, allelic and genotypic

frequencies, as well as linkage disequilibrium were determined and compared with world populations.

A MAS-PCR genotyping system was successfully designed and optimized for the genotyping of 10 SNPs from the SLC22A1gene. The system is made of 2 multiplexes each covering 5 SNPs. The system is currently being used to genotype samples from the Cape Coloured population as well as other populations. This MAS-PCR system could undergo further evaluation, and may be additional optimization and even re-designing will be needed.

The SNaPshot® Multiplex System was used as the standard genotyping technique for this study. However, it is recommended to continue developing MAS-PCR genotyping systems to include more SNPs that are investigated in the project. MAS-PCR method should be used whenever possible to reduce the cost of genotyping. MAS-PCR genotyping systems are well suited for many smaller and less equipped laboratories with modest budgets, and give them the opportunity to generate population data in a rapid and cost effective manner.

It is also recommended to develop HRM genotyping systems, which could reduce the cost and increase the throughput rate of genotyping. In addition to genotyping, HRM analysis can be used to scan large numbers of samples for novel genetic variations.

Novel genetic variants' discovery study should be initiated for the SLC22A1 and other SLC members within immigrant and indigenous populations, using a combination of HRM scanning and direct sequencing methods. HRM analysis is used to scan large numbers of samples for genetic variation. This mutation scanning strategy is used to scan DNA samples from many individuals for minor genetic variations to identify candidates with genetic variations for full sequencing analysis (Vossen *et al*., 2009).

References

Arndt, P., Volk, C., Gorboulev, V., Budiman, T., Popp, C., Ulzheimer- Teuber, I., Akhoundova, A., Koppatz, S., Bamberg, E., Nagel, G. and Koepsell, H. (2001) Interaction of cations, anions, and weak base quinine with rat renal cation transporter rOCT2 compared with rOCT1. Am. J. Physiol. Renal Physiol., 281(3): F454-F468

Asaka, J., Terada, T., Ogasawara, K., Katsura, T. and Inui, K. (2007). Characterization of the Basal promoter element of human organic cation transporter 2 gene. J. Pharmacol. Exp. Ther.,321(2): 684–689

Avery, P. and Mousa, S. S. (2009). Pharmacogenomics in type II diabetes mellitus management: Steps toward personalized medicine. *Pharmacogenomics and Personalized Medicine* 2: 79-91

Barton NH and Slatkin M, (1986). A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. Heredity 56:409-415

Barrett JC, Fry B, Maller J, Daly MJ. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. [PubMed ID: 15297300]

Becker, M. L., Visser, L. E., van Schaik, R. H. N., Hofman, A., Uitterlinden, A. G. and Stricker, B. H. C. (2009). Genetic variation in the organic cation transporter 1 is associated with metformin response in patients with diabetes mellitus: Genetic variation in OCT1 and metformin. *The Pharmacogenomics Journal*. 9: 242-247

Benjeddou, M. (2010). Solute carrier transporter: Pharmacogenomics research opportunities in Africa. *Afr. J. Biotechnol*. 9(54): 9191-9195

Biermann, J., Lang, D., Gorboulev, V., Koepsell, H., Sindic, A., Schroter, R., Zvirbliene, A., Pavenstadt, H., Schlatter, E. and Ciarimboli, G. (2006). Characterization of regulatory mechanisms and states of human organic cation transporter 2. Am. J. Physiol. Cell Physiol., 290(6): C1521–C1531

Brockmoller, J., and Tzvetkov, M. V. (2008) Pharmacogentics: data, concepts and toolsto improve drug discovery and drug treatment. *Eur. J. Clin. Pharmacol*. 64: 133-157

Butler, J. M. (2005). Forensic DNA Typing: biology, technology, and genetics of STR markers, Elsevier Academic Press, London.

CardioGenomics (2005). About Haploview figures. (Cited 22 February 2012) (Online) Available from: http://cardiogenomics.med.harvard.edu/groups/comp08/pages/about-haploview.html

Caulfield, M. J. (2001). Genes for common diseases, Br. J. Clin. Pharmacol. 51: 1-3

Cetinkaya, I., Ciarimboli, G., Yalcinkaya, G., Mehrens, T., Velic, A., Hirsch, J.R., Gorboulev, V., Koepsell, H. and Schlatter, E. (2003). Regulation of human organic cation transporter hOCT2 by PKA, PI3K, and calmodulin-dependent kinases. Am. J. Physiol. Renal Physiol., 284(2): F293–F302

Chen, J.J., Li, Z., Pan, H., Murphy, D.L., Tamir, H., Koepsell, H. and Gershon, M.D. (2001). Maintenance of serotonin in the intestinal mucosa and ganglia of mice that lack the high-affinity serotonin transporter: abnormal intestinal motility and the expression of cation transporters. J. Neurosci. 21:6348–6361

Choi, M. and Song, I. (2008). Organic Cation Transporters and their Pharmacokinetic and Pharmacodynamic Consequences. Drug Metab. Pharmacokinet. 23 (4): 243–253

Ciarimboli, G. and Schlatter, E. (2005). Regulation of organic cation transport. Pflugers Arch., 449(5): 423–441

Ciarimboli, G., Koepsell, H., Iordanova, M., Gorboulev, V., Durner, B., Lang, D., Edemir, B., Schroter, R., Van Le, T. and Schlatter, E. (2005). Individual PKC-phosphorylation sites in organic cation transporter 1 determine substrate selectivity and transport regulation. J. Am. Soc. Nephrol., 16(6): 1562–1570

Denk, G. U., Soroka, C. J., Mennone, A., Koepsell, H., Beuers, U. and Boyer, J. L. (2004). Down-regulation of the organic cation transporter 1 of rat liver in obstructive cholestasis. Hepatology, 39(5): 1382–1389

Echols, M.R. and Yancy, C.W. (2006). Isosorbide dinitrate-hydralazine combination therapy in African Americans with heart failure. Vascular Health and Risk Management 2(4) 423–431

Edward, M. C. and Gibbs, R. A. (1995) Multiplex PCR, in: Dieffenbach, C. W. and Dveksler, G. S (Eds.), PCR Primer: A laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 157-171

Eichelbaum, M., Ingelman-Sundberg, M. and Evans, W. E. (2006). Pharmacogenomics and individualized drug therapy. Annu. Rev. Med., 57: 119–137

Evans, W. E. and McLeod, H. L. (2003). Pharmacogenomics–drug disposition, drug targets, and side effects. N. Engl. J. Med., 348(6): 538–49

Evans,W. E. and Relling, M. V. (2004). Moving towards individualized medicine with pharmacogenomics. Nature, 429(6990): 464–468

Fahrmayr, C., Fromm, M.F. and Koning, J. (2010). Hepatic OATP and OCT uptake transporters:their role for drug-drug interactions and pharmacogenetic aspects. Drug Metabolism Reviews 42(3): 380-401

Fujita, T., Urban, T. J., Leabman, M. K., Fujita, K. and Giacomini, K. M. (2006). Transport of drugs in the kidney by the human organic cation transporter, OCT2 and its genetic variants. J. Pharm. Sci., 95(1): 25–36

Gaudet M, Fara A, Sabatti M, Kuzminsky E & Mugnozza G (2007) "Single-reaction for SNP Genotyping on Agarose Gelby Allele-specific PCR in Black Poplar (Populus nigra L.)." Plant Mol Biol Rep 25: 1–9

Gorboulev, V., Shatskaya, N., Volk, C. and Koepsell, H. (2005). Subtype-specific affinity for corticosterone of rat organic cation transporters rOCT1 and rOCT2 depends on three amino acids within the substrate binding region. Mol. Pharmacol., 67(5): 1612–1619

Gorboulev, V., Ulzheimer, J. C., Akhoundova, A., Ulzheimer- Teuber, I., Karbach, U., Quester, S., Baumann, C., Lang, F., Busch, A. E. and Koepsell, H. (1997) Cloning and characterization of two human polyspecific organic cation transporters. DNA Cell Biol., 16(7): 871–881

Gorboulev, V., Volk, C., Arndt, P., Akhoundova, A. and Koepsell, H. (1999). Selectivity of the polyspecific cation transporter rOCT1 is changed by mutation of aspartate 475 to glutamate. Mol. Pharmacol., 56(6): 1254–1261

Graham, R., Liew, M., Meadows, C., Lyon, E. and Wittwer, C.T. (2005). Distinguishing different DNA heterozygotes by high-resolution melting. Clin. Chem. 51: 1295-1298

Green, R. M., Lo, K., Sterritt, C. and Beier, D. R. (1999). Cloning and functional expression of a mouse liver organic cation transporter. Hepatology, 29(5): 1556–1562

Griffiths AJ, F., Miller JH, Suzuki DT, Lewontin RC, and. Gelbart WM. (2000) *An Introduction to Genetic Analysis*, 7th edition. New York: W. H. Freeman & Co.

Griffiths AJF, Gelbart WM, Miller JH, and Lewontin RC. (1999) *Modern Genetic Analysis*. New York: W. H. Freeman & Co.

Grundemann, D., Gorboulev, V., Gambaryan, S., Veyhl, M. and Koepsell, H. (1994). Drug excretion mediated by a new prototype of polyspecific transporter. Nature, 372(6506): 549–552

Grundemann, D., Schechinger, B., Rappold, G. A. and Schomig, E. (1998). Molecular identification of the corticosterone-sensitive extraneuronal catecholamine transporter. Nat. Neurosci., 1(5): 349–351

Hansson B and Kawabe A. (2005) A simple method to score single nucleotide polymorphisms based on allelespecific PCR and primer-induced fragment-length variation. Mol Ecol Notes 5:692–6

Hardy, B.J., Seguin, B., Goodsaid, F., Jimenez-Sanchez, G., Singer, P.A. and Daar, A.S. (2008). The next steps for genomic medicine: challenges and opportunities for the developing world. Nat. Rev. Genet. 9(1): 23-27

Hediger, M. A., Romero, M. F., Peng, J. B., Rolfs, A., Takanaga, H. and Bruford, E. A. (2004). The ABCs of solute carriers: physiological, pathological and therpeuticimplications of human membrane transport proteins. Introduction. Pflugers Arch. 447:465-468

Ishiguro A, Kubota T, Soya Y, Sasaki H, Yagyu O and Takarada Y. (2005). High-throughput detection of multiple genetic polymorphisms influencing drug metabolism with mismatch primers in allele-specific polymerase chain reaction. Anal Biochem 337:256–61

Itoda, M., Saito, Y., Maekawa, K., Komamura, K., Kamakura, S., Kitakaze, M., Tomoike, H., Ueno, K., Ozawa, S. and Sawada, J. (2004). Seven novel single nucleotide polymorphisms in the human SLC22A1 gene encoding organic cation transporter 1 (OCT1). *Drug*. *Metab*. *Pharmacokinet.*, 19(3): 239-244

Jonker, J. W. and Schinkel, A. H. (2004). Pharmacological and physiological functions of the polyspecific organic cation transporters: OCT1, 2, and 3 (SLC22A1–3). J. Pharmacol. Exp. Ther., 308(1): 2–9

Jonker, J. W., Wagenaar, E., Van Eijl, S. and Schinkel, A. H. (2003). Deficiency in the organic cation transporters 1 and 2 (Oct1/Oct2 [Slc22a1/Slc22a2]) in mice abolishes renal secretion of organic cations. Mol. Cell Biol., 23(21): 7902–7908

Kalinowski, L., Dobrucki, I.T., Tomasian, D. (2004). Race-specific differences in endothelial function: Predisposition of African Americans to vascular disease. Circulation, 109:2511–2517

Kang, H. J., Lee, S. S., Lee, C. H., Shim, J. C., Shin, H. J., Liu, K. H., Yoo, M. A. and Shin, J. G. (2006) Neurotoxic pyridinium metabolites of haloperidol are substrates of human organic cation transporters. Drug Metab. Dispos., 34(7): 1145–1151

Kang, H. J., Song, I. S., Shin, H. J., Kim, W. Y., Lee, C. H., Shim, J. C., Zhou, H. H., Lee, S. S. and Shin, J. G. (2007). Identification and functional characterization

of genetic variants of human organic cation transporters in a Korean population. Drug Metab. Dispos., 35(4): 667–675

Kekuda, R., Prasad, P. D., Wu, X., Wang, H., Fei, Y. J., Leibach, F. H. and Ganapathy, V. (1998). Cloning and functional characterization of a potential-sensitive, polyspecific organic cation transporter (OCT3) most abundantly expressed in placenta. J. Biol. Chem., 273(26): 15971–15979

Kerb, R., Brinkmann, U., Chatskaia, N., Gorbunov, D., Gorboulev, V.,Mornhinweg, E., Keil A, Eichelbaum,M. and Koepsell, H. (2002) Identification of genetic variations of the human organic cation transporter hOCT1 and their functional consequences. Pharmacogenetics, 12(8): 591–595

Kerb, R., Brinkmann, U., Chatskaia, N., Gorbunov, D., Gorboulev, V., Mornhinweg, E., Kell, A., Eichelbaum, M. and Koepsell, H. (2002). Identification of genetic variations of the human organic cation transporter hOCT1 and their functional consequences. Pharmacogenetics, 12(18): 591-595

Koepsell H (1998) Organic cation transporters in intestine, kidney, liver and brain. Annu Rev Physiol 60: 243-266

Koepsell, H., Gorboulev, V. and Arndt, P. (1999) Molecular pharmacology of organic cation transporters in kidney. *J. Membr. Biol.*, 167(2): 103–17

Koepsell, H., Lips, K. and Volk, C. (2007) Polyspecific organic cation transporters: structure, function, physiological roles, and biopharmaceutical implications. Pharm. Res., 24(7): 1227–51

Kroetz, D.L., Yee, S.W. and Giacomini, K.M. (2010). The pharmacogenetics of membrane Transporters Project: Research at the Interface of Genomics and Transporter Phamacology. Clin. Pharm. Ther. 87(1): 109-116

Lazar, A., Grundemann, D., Berkels, R., Taubert, D., Zimmermann, T. and Schomig, E. (2003). Genetic variability of the ex252 traneuronal monoamine transporter EMT (SLC22A3). J. Hum. Genet., 48(5): 226–30

Leabman M. K., Huang,C.C., DeYoung, J., Carlson, E. J., Taylor, T. R., De la Cruz, M., Johns, S. J., Stryke, D., Kawamoto, M., Urban T. J., Kroetz, D. L., Ferrin, T. E., Clark, A. G., Risch, N., Herskowitz, I. and Giocomini, K. M. (2003). Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. *Proc*. *Natl*. *Acad*. *Sci*. USA 100: 5896-5901

Leabman, M. K., Huang, C. C., Kawamoto, M., Johns, S. J., Stryke, D., Ferrin, T. E., DeYoung, J., Taylor, T., Clark, A. G., Herskowitz, I. and Giacomini, K. M. (2002). Polymorphisms in a human kidney xenobiotic transporter, OCT2, exhibit altered function. Pharmacogenetics, 12(5): 395–405

Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E. and Wittwer, C.T. (2004). Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. Clin. Chem. 50: 1156-1164

Liu Q, Thorland EC, Heit JA and Sommer SS. (1997). Overlapping PCR for bidirectional PCR amplification of specific alleles: a rapid one-tube method for simultaneously differentiating homozygotes and heterozygotes. Genome Res 7:389–98

Mao, X., Bigham, A.W., Mei, R., Gutierrez, G., Weiss, K.M., Brutsaert, T.D., Leon-Velarde, F., Moore, L.G., Vargas, E., McKeigue, P.M., *et al*. (2007). A genomewide admixture mapping panel for Hispanic/Latino populations. Am. J. Hum. Genet. 80, 1171–1178

Markoulatos, P., Siafakas, N. and Moncany, M. (2002). Multiplex polymerase chain reaction: A practical approach. J. Clin. Lab. Anal. 16: 47-51

Meier, Y., Eloranta, J. J., Darimont, J., Ismair, M. G., Hiller, C., Fried, M., Kullak-Ublick, G. A. and Vavricka, S. R. (2007). Regional distribution of solute carrier mRNA expression along the human intestinal tract. *Drug Metab. Dispos*. 35: 590-594

Mooslehner, K. A. and Allen, N. D. (1999). Cloning of the mouse organic cation transporter 2 gene, Slc22a2, from an enhancertrap transgene integration locus. Mamm. Genome, 10(3): 218–224

Mountain, A. (2003). The First People of the Cape (Claremont, South Africa: David Philips Publishers).

Nakanishi, T. and Tamai, I. (2011) Solute Carrier Transporters as Targets for drug delivery and Pharmacological intervention for chemotherapy. J. Pharm. Sci. 100(9): 3731-3750

Nei, M. and Kumar, S. (2000). Molecular Evolution and phylogenetics, Oxford University Press, United States of America.

Nurse, G.T., Weiner, S., and Jenkins, T. (1985). The Peoples of Southern Africa and Their Affinities (Oxford: Clarendon Press).

Nussbaum RL, McInnes RR, and Willard HF (2007). Genetics in Medicine. 7th ed. Philadelphia: Saunders.

Okimoto R. and Dodgson J.B. (1996). Improved PCR amplification of multiple specific alleles (PAMSA) using internally mismatched primers. BioTechniques 21:20–26

Okuda, M., Saito, H., Urakami, Y., Takano, M. and Inui, K. (1996). cDNA cloning and functional expression of a novel rat kidney organic cation transporter, OCT2. Biochem. Biophys. Res. Commun., 224(2): 500–507

Palais, R.A., Liew, M.A. and Wittwer, C.T. (2005). Quantitative heteroduplex analysis for single nucleotide polymorphism genotyping. Anal. Biochem. 346: 167-175

Peakall, R. and Smouse, P. (2006). Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6(1): 288-295

Peakall, R. and Smouse, P. (2009). Tutorial Part 1 - An Introduction to Frequency-Based Population Genetic Analysis. (Online)(Cited 15 February 2012) Available from: www.anu.edu.au/BoZo/GenAlEx/GenAlEx%20WWW/GenAlEx%206.3/GenAlEx%20Tut1.zip

Pelis, R. M., Zhang, X., Danqprapai, Y. and Wright S. H. (2006). Cysteine accessibility in the hydrophilic cleft of the human organic cation transporter 2. J. Biol. Chem. 281:35272–35280

Popp, C., Gorboulev, V., Muller, T. D., Gorbunov, D., Shatskaya, N. and Koepsell, H. (2005). Amino acids critical for substrate affinity of rat organic cation transporter 1 line the substrate binding region in a model derived from the tertiary structure of lactose permease. Mol. Pharmacol., 67(5): 1600–1611

Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, van Helden P, Hoal E and Behar D (2010). Strong maternal Khoisan contribution to the South African coloured population: A case of gender-biased admixture. Am. J. Hum. Genet 86: 611-620

Raymond M. and Rousset F, (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Heredity, 86:248-249

Ririe, K.M., Rasmussen, R.P. and Wittwer, C.T. (1997). Product differentiation by analysis of DNA melting curves during polymerase chain reaction. Anal. Biochem. 245: 154-160

Rousset, R. (2008). Genepop'007: a complete re-implementation of the GenePop software for Windows and Linux. Mol. Ecol. Res. 8: 103-106

Rudin, N., Inman, K. (2002). An introduction to forensic DNA analysis–second edition, CRC Press, Washington D. C.

Saborowski, M., Kullak-Ublick, G. A. and Eloranta, J. J. (2006). The human organic cation transporter-1 gene is transactivated by hepatocyte nuclear factor-4alpha. J. Pharmacol. Exp. Ther., 317(2): 778–785

Sakata, T., Anzai, N., Shin, H. J., Noshiro, R., Hirata, T., Yokoyama, H., Kanai, Y. and Endou, H. (2004). Novel single nucleotide polymorphisms of organic cation transporter 1 (SLC22A1) affecting transport functions. Biochem. Biophys. Res. Commun., 313(3): 789–793

Sasvari-Szekely M, Gerstner A, Ronai Z, Staub M and Guttman A. (2000). Rapid genotyping of factor V Leiden mutation using single-tube bidirectional allele-specific amplification and automated ultrathin-layer agarose gel electrophoresis. Electrophoresis 21:816–21

Shastry, B. S. (2007). SNPs in disease gene mapping, medical drug development and evolution, J. Hum. Genet. 52: 871-880

Shell, R. (1994). Children of bondage (Johannesburg, South Africa: Witwatersrand University Press).

Shu Y, Sheardown, S.A., Brown, C., Owen, R.P., Zhang, S., Castro, R.A., Ianculescu, A.G., Yue, L., Lo, J.C., Burchard, E.G., Brett, C.M. and Giacomini, K.M..

(2007). Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. J Clin Invest.117:1422–1431

Shu, Y., Brown, C., Castro, R.A., Shi, R.J., Lin, E.T., Owen, R.P., Sheardown, S.A., Yue, L., Burchard, E.G., Brett, C.M. and Giacomini, K.M. (2008). Effect of Genetic Variation in the Organic Cation Transporter 1, OCT1, on Metformin Pharmacokinetics. Clin. Pharmacol. Ther. 83(2): 273-280

Shu, Y., Leabman, M. K., Feng, B., Mangravite, L. M., Huang, C. C., Stryke, D., Kawamoto, M., Johns, S. J., DeYoung, J., Carlson, E., Ferrin, T. E., Herskowitz, I. and Giacomini, K. M. (2003). Evolutionary conservation predicts function of variants of the human organic cation transporter, OCT1. Proc. Natl. Acad. Sci. USA. 100(10): 5902–5907

Smith K., (2002) Genetic Polymorphism and SNPs: Genotyping, Haplotype Assembly Problem, Haplotype Map, Functional Genomics and Proteomics

Smith, M.W., and O'Brien, S.J. (2005). Mapping by admixture linkage disequilibrium: Advances, limitations and guidelines. Nat. Rev. Genet. 6, 623–632

Smith, M.W., Lautenberger, J.A., Shin, H.D., Chretien, J.P., Shrestha, S., Gilbert, D.A., and O'Brien, S.J. (2001). Markers for mapping by admixture linkage disequilibrium in African American and Hispanic populations. Am. J. Hum. Genet. 69, 1080–1094

Song, I., Shin,H., Shim, E., Jung, I., Kim, W., Shon, J. and Shin, J. G. (2008). Genetic Variants of the Organic Cation Transporter 2 Influence the Disposition of Metformin. Clin. Pharmacol. Ther. 84: 559 –562

Strachan, T. and Read, A. P. (2004). Human Molecular genetics 3–Third Edition, Garland Publishing, New Delhi, India.

Sweet, D.H., Miller, D.S. and Pritchard, J.B. (2001). Ventricular choline transport: a role for organic cation transporter 2 expressed in choroid plexus. J. Biol. Chem. 276:41611–41619

Syvanen, A.C. (1999). From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. Hum Mutat. 13(1): 1-10

Takeuchi, A., Motohashi, H., Okuda, M. and Inui, K. (2003). Decreased function of genetic variants, Pro283Leu and Arg287Gly, in human organic cation transporter hOCT1. Drug Metab. Pharmacokinet., 18(6): 409–412

Tate, S.K. and Goldstein, D.B. (2004). Will tomorrow's medication work for everyone? Nature Genetics Supplement 36(11): s34-42

Terashita, S., Dresser, M. J., Zhang, L., Gray, A. T., Yost, S. C. and Giacomini, K. M. (1998). Molecular cloning and functional expression of a rabbit renal organic cation transporter. Biochim. Biophys. Acta., 1369(1): 1–6

Tian, C., Hinds, D.A., Shigeta, R., Adler, S.G., Lee, A., Pahl, .V., Silva, G., Belmont, J.W., Hanson, R.L., Knowler, W.C., *et al*. (2007). A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. Am. J. Hum. Genet. 80, 1014–1023

Tian, C., Hinds, D.A., Shigeta, R., Kittles, R., Ballinger, D.G., and Seldin, M.F. (2006). A genomewide single-nucleotide polymorphism panel with high ancestry information for African American admixture mapping. Am. J. Hum. Genet. 79, 640–649

Tzvetkov, M. V., Vormfelde, S. V., Balen, D., Meineke, I., Schmidt, T., Sehrt, D., Sabolic, I., Koepsell, H. and Brockmoller, J. (2009). The effects of genetic

polymorphisms in the organic cation transporters OCT1, OCT2 and OCT3 on the renal clearance of metformin. *Clin*. *Pharmacol*. *Ther*. 86: 299-306

Von Ahsen, N., Oellerich, M. and Schultz, E. (2001). Limitations of genotyping based on amplicon melting temperature. Clin. Chem. 47: 1331-1332

Vossen, R.H.A.M., Aten, E., Roos, A. and Dunnen, J.T. (2009). High-Resolution Melting Analysis (HRMA) – more than just sequence variant screening. Hum. Mut. 30: 1-7

Waterfall C.M and Cobb B.D. (2002). SNP genotyping using single-tube fluorescent bidirectional PCR. BioTechniques 33:80–90

Waterfall C.M. and Cobb B.D. (2001). Single tube genotyping of sickle cell anaemia using PCR-based SNP analysis. Nucl Acids Res 29:e119

Weinshilboum, R. (2003). Inheritance and drug response. N. Engl. J. Med., 348(6): 529–537.

Wittwer, C.T., Red, G.H., Gundry, C.N., Vandersteen, J.G. and Pryor, R.J. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. Clin. Chem. 49: 853-860

Wojtal, K. A., Eloranta, J. J., Hruz, P., Gutmann, H., Drewe, J., Staumann, A., Beglinger, C., Fried, M., Kullak-Ublick, G. A., and Vavricka, S. R. (2009). Changes in mRNA expression levels of solute carrier transporters in inflammatory bowel disease patients. *Clin*. *Pharmacol*. *Ther*. 37: 1871-1877

Wright, S. H., and Dantzler, W. H., (2004). Molecular and Cellular Physiology of Renal Organic Cation and Anion Transport. Physiol. Rev. 84: 987-1049

Wu, X., Huang, W., Ganapathy, M. E., Wang, H., Kekuda, R., Conway, S. J., Leibach, F. H. and Ganapathy, V. (2000). Structure, function, and regional

distribution of the organic cation transporter OCT3 in the kidney. Am. J. Physiol. Renal. Physiol., 279(3): F449-F458

Yang, I., Kim, Y.,Byun, J. and ark, S. (2005). Use of multiplex polymerase chain reaction to indicate the accuracy of the annealing temperature of thermal cycling. An. Bio. 338: 192-200

You, F. M., Huo, N., Gu, Y. Q., Luo, M., Ma, Y., Hane, D., Lazo, G. R., Dvorak, J. and Anderson, O. D. (2008) BatchPrimer3: A high throughput web application for PCR and sequencing primer design. BMC Bioinformatics, 9: 253

Zhang, L., Dresser, M. J., Gray, A. T., Yost, S. C., Terashita, S. and Giacomini, K. M. (1997). Cloning and functional expression of a human liver organic cation transporter. Mol. Pharmacol., 51(6): 913–921

Zhou, L., Wang, L., Palais, R., Pryor, R. and Wittwer, C.T. (2005). High-resolution melting analysis for simultaneous mutation scanning and genotyping in solution. Clin. Chem. 51: 1770-1777

Zwart, R., Verhaagh, S., Buitelaar, M., Popp-Snijders, C. and Barlow, D. P. (2001). Impaired activity of the extraneuronal monoamine transporter system known as uptake-2 in Orct3/Slc22a3- deficient mice. Mol. Cell Biol., 21(13): 4188–4196

Appendix I

Protocols

1. Composition of Buffer and Solutions

Lysis Buffer (stock)

10 ml 2 M NaCl

500 µl 1 M Tris-HCL pH = 8

200 µl 2 M EDTA

1% SDS

All dissolved in a volume of 40 ml SABAX water and made up to a final volume of 50 ml.

Agarose Gel (50 ml)

1.5g Agarose

Added to 48.5 ml 1X TBE Buffer

10X TBE Buffer (1L)

108g Tris Base EDTA

~55g Boric Acid

Add Tris Base, EDTA and Boric Acid to 800 ml DI H_2O and mix to dissolve. Adjust to 1 L with additional DI H_2O once dissolved.

1X TBE Buffer (1L)

100 ml 10X TBE Buffer

Add TBE to 900 ml DI H_2O.

20 mgml^{-1} Proteinase K (5ml)

100 g Proteinase K

Add Proteinase K to 5 ml DI H_2O.

1X EDTA Buffer

10 ml 10X EDTA 3730 Running buffer

Added to 40 ml DI H_2O

2. DNA Extraction

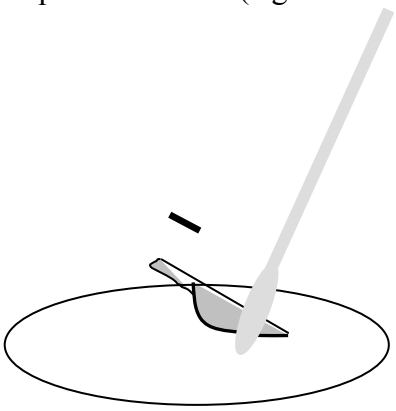
2.1. Preparation of the lysis buffer

Add all reagents in 80% of their final volume in bidestilled water. Add the SDS and leave in the oven at 60ºC until it is fully dissolved. Transfer the volume to a volumetric flask and add the necessary SABAX water to get 100 % desired volume.

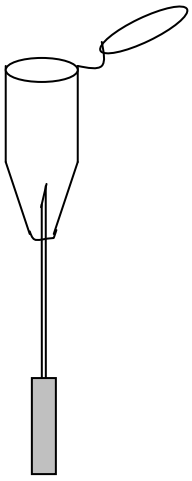
2.2. Extraction Procedure

- Cut off the surface of the swab with a clean scalpel surgery blade. Work on a clean surface (sterile Petri dish or small plastic bag) on top of a tray.

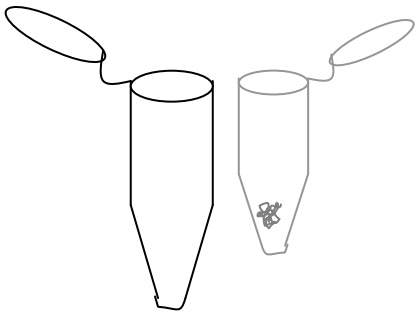
- Prepare 1.5 or 2ml Eppendorf tubes, and add a volume of lysis buffer and add the proteinase K. (e.g add 3 µl PK 20 mg/ml to 600 µl lysis buffer).



- Add the little pieces of excised swab to the Eppendorf tubes and vortex for 30 seconds. Incubate at 56 ºC and leave it overnight.
- Transfer all the volume to a clean tube.
- The pieces of swab still have trapped lysis solution with biological material. To recover this volume, proceed as follows:
- Perforate the end of a 0.5 ml tube with a needle (21-22 gauge)



- Place this tube inside a 1.5 ml Eppendorf tube and spin for 1 min in a microcentrifuge. Add the collected volume to the previously separated lysis material.



- Precipitation was done by adding 1/3 volume of 6M NaCl and shaking the tube (s) vigoursly for 15 seconds.
- Centrifuge for 15min at 5000rpm and transfer supernatant with DNA to another tube.
- Add **equal** volume of cold isopropanol. Leave at -80ºC for 30 mins or overnight at -20ºC.
- Pellet the DNA by centrifugation: 14,000 rpm 30 minutes.
- Wash the pellet with 100 ul of 70% ethanol to remove the salts. Centrifugate @ 14,000 rpm for 8 minutes to 30 minutes.
- Dry pellet shortly in SpeedyVac or at 65ºC, prevent that the DNA becomes too dry since it will hardly dissolve then.
- Dissolve DNA in 30 ul of SABAX water. Store at -20ºC.

3. MAS-PCR

3.1. MAS PCR Primer mix

Primer mix of Multiplex 1 and 2 markers was set up according to the table below

- Firstly the F_1, F_2 and R primers synthesized at Whitehead Scientific was made up to a 100 µM of each individual primer from this 3 µl of each the forward

and reverse primers is aliquot into a tube labelled mix and resuspended in a final volume of 50 µl. This yielded a stock concentration of 33 µM.

- Then from the 33 µM stock of these primers a working stock was prepared as listed below.
- All the primers must be at the same concentration which in this case is 0.2 µM
- To ensure that the concentration remains constant in 10 µl PCR reaction a 10 X mix is prepared.
- However in the PCR reaction 0.7 is required making the final concentration 0.7 µM instead of 0.5 µM
- It is always best to vortex all primers then centrifuging it prior to making up a primer mix
-

3.3. Agarose Gel Electrophoresis

- Amplified products are viewed on a 3% agarose gel (made-up in a final volume of 50ml) to ensure that the correct amplicon size was obtained.
- Flat capped 0.2ml PCR tubes may be used to prepare samples and PCR marker.
- PCR Marker (New England Biolabs) is prepared by adding 2µl of loading dye + 3 µl water +1.2 µl PCR marker.
- Either the 50bp maker or 100bp marker may be used from (NEB) however the 50bp marker is preferable.(most of the bands fall between 50 -100bp)
- Samples are prepared by loading 1.5 µl of PCR product and 2 µl loading dye onto gel.

4. SNaPshot® Protocol

4.1. Primer mix

Primer mix of Multiplex 1 and 2 markers was set up according to the table below

- Firstly the F and R primers synthesized at Whitehead Scientific was made up to a 100 µM of each individual primer from this 10 µl of both the forward and reverse of a primer is aliquoted into a tube labelled forward and reverse mix

Forward primer 10 µl (100µM)

Reverse primer 10 µl

---------- ----------

Mix 20 µl F +R final concentration is now 50 µM

- Then from the 50 µM stock of these primers a working stock was prepared as listed below.
- All the primers must be at the same concentration which in this case is 0.5 µM
- To ensure that the concentration remains constant in 10 µl PCR reaction a 10 X mix is prepared.
- However in the PCR reaction 0.7 is required making the final concentration 0.7 µM instead of 0.5 µM
- It is always best to vortex all primers then centrifuging it prior to making up a primer mix

4.2. Parameters used

- New primer mix was setup for reactions
- Vortex reagents prior to making up mastermix once all reagents have been added, vortex master mix then centrifuge.
- Then add 7 µl of master mix to the correctly labelled tube
- Before adding DNA vortex it.(briefly)
- Once master mix and DNA has been added to all reaction tubes centrifuge the tubes then place in the old 96 well PCR machine

4.3. Agarose Gel Electrophoresis

- Amplified products are viewed on a 3% agarose gel (made-up in a final volume of 50ml) to ensure that the correct amplicon size was obtained.
- Flat capped 0.2ml PCR tubes may be used to prepare samples and PCR marker.

- PCR Marker (New England Biolabs) is prepared by adding 2µl of loading dye + 3 µl water +1.2 µl PCR marker.
- The Hyperladder V marker was used from (Bioline)
- Samples are prepared by loading 1.5 µl of PCR product and 2 µl loading dye onto gel.

4.3. PCR cleanup

- **NB : All reactions at this point and thereafter should be done on ice**

- Always work on ice as the enzymes Sap and Exo are activated at 37˚C
- PCR products amplified with Multiplex 1 and 2 are cleaned in the following manner:
- 2.05 µl (1unit/µl) SAP (Amersham) and 2.05 µl (10 units/ µl) of EXO I (Amersham) are used to treat 5 µl PCR products for Multiplex I and 2.
- The samples should then be incubated @ 37˚ C for 1hr and the enzyme inactivated @ 75˚ C for 15 min and kept on hold @ 4 ˚ C. .

4.4. SNaPshot® reactions

4.4.1. Extension Treatment of multiplex I and II

The Multiplex PCR products which had undergone cleanup treatment with SAP + Exo *I*, are then amplified with extension primers.

- All extension primers were synthesized by Whitehead Scientific
- These were first made-up to a 100 µM stock thereafter to 5 µM working stock
- A primer mix may be set up containing all extension primers for the respective Multiplex (1or 2) in a 0.5ml or 1.5ml tube depending on how many reactions are performed.
- SNaPshot® Multiplex mix (Applied Bio systems), purified PCR product and water are added individually to each correctly labelled tube after the primer mix.
- The primer mix consist of the following extension primers for Multiplex 1 and 2

4.4.2. Purification of Extension products of Multiplex 1 and 2

- A volume of 2.05 µl (1unit) of SAP is added to the extension products, which is then incubated @ 37 ºC for 1hr, inactivation of enzyme is @ 75 ºC for 15 min and kept on hold @ 4 ˚ C.
- After purification samples can be prepared to be run on the 3500 GENETIC ANALYZER.

4.4.3. Preparation for the Electrophoresis of extension product on 3500 Genetic Analyzer

- Extension samples which had undergone SAP treatment are prepared to be run on the on 3500 Genetic Analyzer.
- Samples are loaded into a 96 well plate according to the sample sheet after a master mix has been setup.
- The master mix contains only the size standard (Liz 120) and Hi di formamide.
- The LIZ 120 and Hi di fomamide do not have to be vortexed prior to use.
- However once the mix has been setup briefly vortex and centrifuge the master mix.
- Each well will receive x µl of the master mix (depending on the amount of extension product used Hi di formamide volume will change) thereafter the purified extension product is added.
- The amount of purified extension product may vary for electrophoresis.

- Once the plate has been prepared place a cover over.

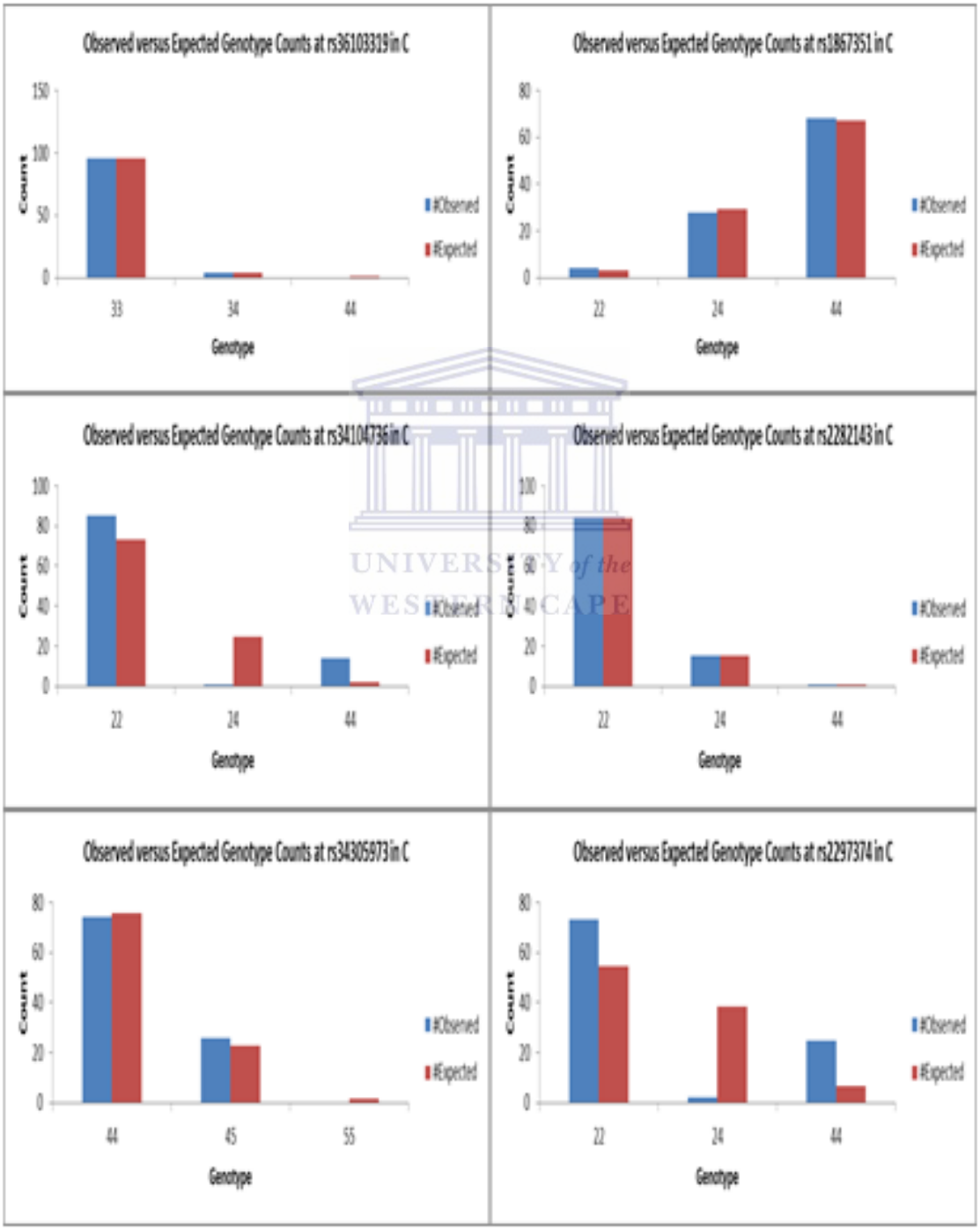
5. GeneMapper 4.1 Software

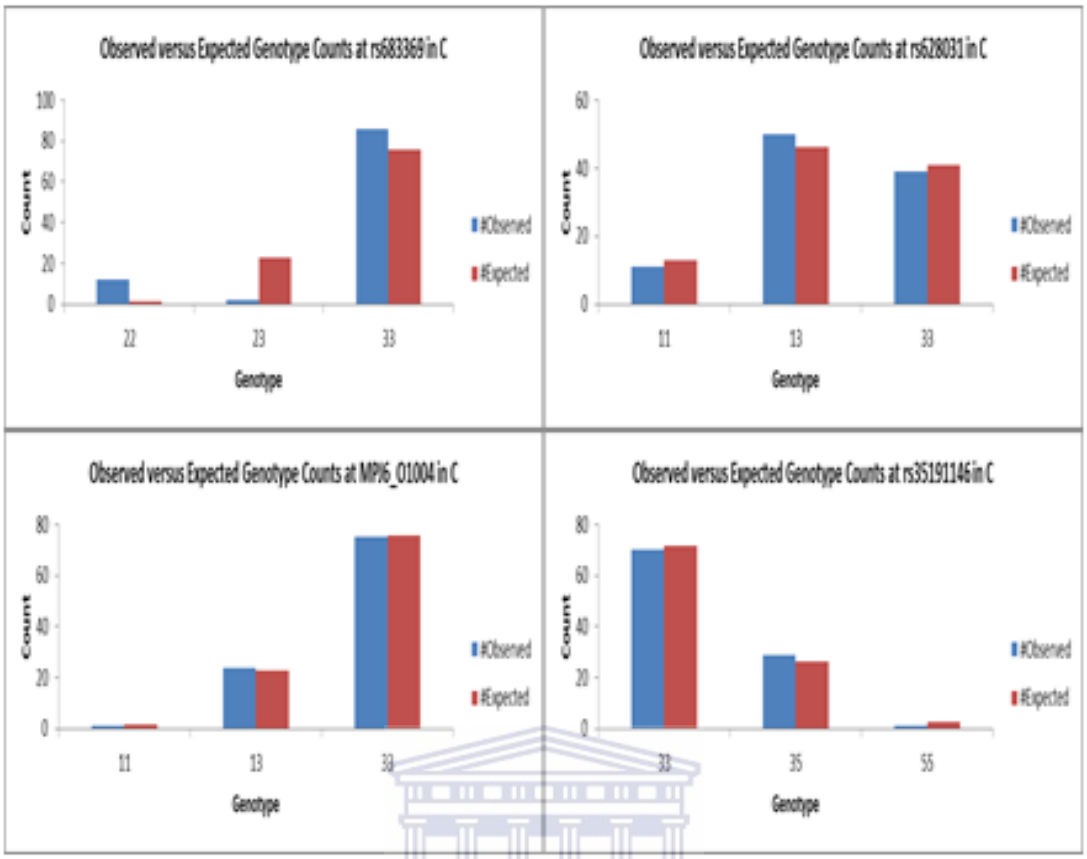
- Open GeneMapper and start a new project.
- Import the relevant run files.
- Select an appropriate size standard and analysis procedure (SNaPshot®).
- Click the analyse tab to analyse the samples
- Once the processing has run to completion, each sample can now be checked individually for the performance of the mini-sequencing reaction.
- Visually inspect each sample to call genotypes.
- Save the project

Appendix II

Genotypic Data

1. Hardy-Weinberg Equilibrium





Where genotypes are represented numerically (1=A, 2=C, 3=G, 4=T)

2. GenePop Genotypic differentiation (exact G test)

```
File: 191436 (Title line:"Total 20 SNPs")

Number of populations detected : 6
Number of loci detected        : 20

Markov chain parameters
        Dememorisation       : 1000
        Batches              : 100
        Iterations per batch : 1000


Locus: rs36103319
=================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          3   3
          3   4     Total

C         96  4       100
AA        100 0       100
CA        99  1       100
A         100 0       100
```

```
ME        100 0      100
P         100 0      100

Total:    595 5      600

P-value = 0.02276    S.E. = 0.00163021 (45744 switches)

Locus: rs12208357
=================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          2   4
          4   4     Total

C         0   100    100
AA        0   100    100
CA        7   93     100
A         0   100    100
ME        1   99     100
P         0   100    100

Total:    8   592    600

P-value = 0.00024    S.E. = 0.000174147 (55004 switches)

Locus: rs1867351
=================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          2   2   4
          2   4   4     Total

C         4   28  68     100
AA        10  34  56     100
CA        0   35  65     100
A         30  54  16     100
ME        44  56  0      100
P         71  29  0      100

Total:    159 236 205    600

P-value = 0    S.E. = 0 (86330 switches)

Locus: rs34104736
=================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          2   2   4
          2   4   4     Total

C         85  1   14     100
AA        100 0   0      100
```

```
CA         99  1    0       100
A          100 0    0       100
ME         100 0    0       100
P          100 0    0       100

Total:     584 2    14      600

P-value = 0    S.E. = 0 (37095 switches)

Locus: rs2282143
==================================
Pop        Genotypes:
           ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
----------------------------------------
           2   2   4
           2   4   4     Total

C          84  15  1       100
AA         83  17  0       100
CA         100 0   0       100
A          80  17  3       100
ME         100 0   0       100
P          100 0   0       100

Total:     547 49  4       600

P-value = 0    S.E. = 0 (52291 switches)

Locus: rs34130495
==================================
Pop        Genotypes:
           ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
----------------------------------------
           1   3
           3   3     Total

C          0   100    100
AA         1   99     100
CA         2   98     100
A          0   100    100
ME         0   100    100
P          0   100    100

Total:     3   597    600

P-value = 0.43967    S.E. = 0.0043579 (35761 switches)

Locus: rs16891138
==================================
Pop        Genotypes:
           ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
----------------------------------------
           1   1
           1   2     Total
```

```
C          100 0      100
AA         93  7      100
CA         100 0      100
A          100 0      100
ME         100 0      100
P          100 0      100

Total:     593 7      600

P-value = 0    S.E. = 0 (52685 switches)

Locus: rs35167514
==================================
Pop        Genotypes:
           ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
           1
           1     Total

C          100    100
AA         94     94
CA         67     67
A          100    100
ME         71     71
P          100    100

Total:     532    532

No table

Locus: rs34305973
==================================
Pop        Genotypes:
           ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
           4
           4     Total

C          74     74
AA         94     94
CA         67     67
A          100    100
ME         71     71
P          100    100

Total:     506    506

No table

Locus: rs2297374
==================================
Pop        Genotypes:
           ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
           2   2   4
```

```
          2   4   4     Total

C         73  2   25     100
AA        71  16  13     100
CA        77  16  7      100
A         90  4   6      100
ME        90  10  0      100
P         100 0   0      100

Total:    501 48  51     600

P-value = 0    S.E. = 0 (77279 switches)

Locus: rs683369
================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          2   2   3
          2   3   3     Total

C         12  2   86     100
AA        99  1   0      100
CA        88  11  1      100
A         96  4   0      100
ME        90  10  0      100
P         100 0   0      100

Total:    485 28  87     600

P-value = 0    S.E. = 0 (76211 switches)

Locus: rs628031
================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          1   1   1   3
          1   2   3   3     Total

C         11  1   49  39     100
AA        6   0   41  53     100
CA        16  0   48  36     100
A         9   0   27  64     100
ME        0   0   43  57     100
P         0   0   14  86     100

Total:    42  1   222 335    600

P-value = 0    S.E. = 0 (47691 switches)

Locus: MPJ6_O1004
================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
```

```
---------------------------------------------------------------------
-----------------------------------------
          1   1   3
          1   3   3     Total

C         1   24  75     100
AA        0   0   0      0
CA        0   0   0      0
A         0   0   0      0
ME        0   0   0      0
P         0   0   0      0

Total:    1   24  75     100

No information

Locus: rs34205214
=================================
Pop       Genotypes:
          ---------------------------------------------------------
---------------------------------------------------------------------
---------------------------------------------------------------------
-----------------------------------------
          1   1   3
          1   3   3     Total

C         0   0   100    100
AA        1   4   95     100
CA        0   0   100    100
A         0   0   100    100
ME        0   0   100    100
P         0   0   100    100

Total:    1   4   595    600

P-value = 0.00075    S.E. = 0.000358272 (21666 switches)

Locus: rs117474883
=================================
Pop       Genotypes:
          ---------------------------------------------------------
---------------------------------------------------------------------
---------------------------------------------------------------------
-----------------------------------------
          2
          2     Total

C         100    100
AA        0      0
CA        0      0
A         0      0
ME        0      0
P         0      0

Total:    100    100

No table

Locus: rs35191146
=================================
Pop       Genotypes:
```

```
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
----------------------------------------
          3
          3     Total

C         70     70
AA        94     94
CA        67     67
A         100    100
ME        71     71
P         100    100

Total:    502    502

No table

Locus: rs34059508
=================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
----------------------------------------
          1   3
          3   3     Total

C         0   100    100
AA        0   100    100
CA        8   92     100
A         0   100    100
ME        0   100    100
P         0   100    100

Total:    8   592    600

P-value = 0    S.E. = 0 (55092 switches)

Locus: rs115733275
=================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
----------------------------------------
          3   4
          3   4     Total

C         0   100    100
AA        98  2      100
CA        0   0      0
A         0   0      0
ME        0   0      0
P         0   0      0

Total:    98  102    200

P-value = 0    S.E. = 0 (14852 switches)

Locus: rs116824962
```

```
==================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          2   4
          2   4     Total

C         100 0      100
AA        98  2      100
CA        0   0      0
A         0   0      0
ME        0   0      0
P         0   0      0

Total:    198 2      200

P-value = 0.50935    S.E. = 0.00548779 (8237 switches)

Locus: rs118060798
==================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          1
          1     Total

C         100    100
AA        0      0
CA        0      0
A         0      0
ME        0      0
P         0      0

Total:    100    100

No table

==================================

P-value across all loci
(Fisher's method)
-----------------------------------------------------
Locus          P-Value
-------------  --------
rs36103319     0.02276
rs12208357     0.00024
rs1867351      0
rs34104736     0
rs2282143      0
rs34130495     0.43967
rs16891138     0
rs2297374      0
rs683369       0
rs628031       0
rs34205214     0.00075
rs34059508     0
rs115733275    0
```

```
rs116824962    0.50935

All: Chi2= Infinity (df= 28), highly significant

Normal ending.
```

3. Genotypic differentiation for each population pair (exact G test)

File: 181934 (Title line:"Total 20 SNPs")

Number of populations detected : 6
Number of loci detected : 20

Markov chain parameters
Dememorisation : 1000
Batches : 100
Iterations per batch : 1000

Locus: rs36103319
==========

Pop	Genotypes: 3 3	3 4	Total
C	96	4	100
AA	100	0	100
CA	99	1	100
A	100	0	100
ME	100	0	100
P	100	0	100
Total:	595	5	600

Locus	Population pair	P-Value	S.E.	Switches
rs36103319	AA & C	0.12294	0.00166	61897
rs36103319	CA & C	0.36892	0.00254	68480
rs36103319	CA & AA	1.00000	0.00000	49745
rs36103319	A & C	0.12141	0.00168	62129
rs36103319	A & AA	No table		
rs36103319	A & CA	1.00000	0.00000	50107
rs36103319	ME & C	0.12081	0.00158	62038
rs36103319	ME & AA	No table		
rs36103319	ME & CA	1.00000	0.00000	50164
rs36103319	ME & A	No table		
rs36103319	P & C	0.11981	0.00173	62141
rs36103319	P & AA	No table		
rs36103319	P & CA	1.00000	0.00000	49960
rs36103319	P & A	No table		
rs36103319	P & ME	No table		

Locus: rs12208357
==========

Pop	Genotypes: 2 4	4 4	Total

C	0	100	100
AA	0	100	100
CA	7	93	100
A	0	100	100
ME	1	99	100
P	0	100	100
Total:	8	592	600

Locus	Population pair		P-Value	S.E.	Switches
rs12208357	AA	& C	No table		
rs12208357	CA	& C	0.01281	0.00065	72229
rs12208357	CA	& AA	0.01441	0.00067	72155
rs12208357	A	& C	No table		
rs12208357	A	& AA	No table		
rs12208357	A	& CA	0.01332	0.00063	71979
rs12208357	ME	& C	1.00000	0.00000	49801
rs12208357	ME	& AA	1.00000	0.00000	50358
rs12208357	ME	& CA	0.06408	0.00159	71954
rs12208357	ME	& A	1.00000	0.00000	50236
rs12208357	P	& C	No table		
rs12208357	P	& AA	No table		
rs12208357	P	& CA	0.01297	0.00062	72488
rs12208357	P	& A	No table		
rs12208357	P	& ME	1.00000	0.00000	49877

Locus: rs1867351

Pop	Genotypes:			
	2	2	4	
	2	4	4	Total
C	4	28	68	100
AA	10	34	56	100
CA	0	35	65	100
A	30	54	16	100
ME	44	56	0	100
P	71	29	0	100
Total:	159	236	205	600

Locus	Population pair		P-Value	S.E.	Switches
rs1867351	AA	& C	0.05089	0.00355	81105
rs1867351	CA	& C	1.00000	0.00000	70699
rs1867351	CA	& AA	0.02770	0.00257	78786
rs1867351	A	& C	0.00000	0.00000	85490
rs1867351	A	& AA	0.00000	0.00000	85588
rs1867351	A	& CA	0.00000	0.00000	84738
rs1867351	ME	& C	0.00000	0.00000	85950
rs1867351	ME	& AA	0.00000	0.00000	85812
rs1867351	ME	& CA	0.00000	0.00000	85653
rs1867351	ME	& A	0.00045	0.00017	81950
rs1867351	P	& C	0.00000	0.00000	86054
rs1867351	P	& AA	0.00000	0.00000	86242

rs1867351	P	& CA	0.00000	0.00000	85982
rs1867351	P	& A	0.00000	0.00000	82106
rs1867351	P	& ME	0.00007	0.00005	88566

Locus: rs34104736

Pop	Genotypes:			
	2 2	2 4	4 4	Total
C	85	1	14	100
AA	100	0	0	100
CA	99	1	0	100
A	100	0	0	100
ME	100	0	0	100
P	100	0	0	100
Total:	584	2	14	600

Locus	Population pair		P-Value	S.E.	Switches
rs34104736	AA	& C	0.00002	0.00002	55687
rs34104736	CA	& C	0.00000	0.00000	58982
rs34104736	CA	& AA	1.00000	0.00000	49703
rs34104736	A	& C	0.00000	0.00000	55842
rs34104736	A	& AA	No table		
rs34104736	A	& CA	1.00000	0.00000	49872
rs34104736	ME	& C	0.00000	0.00000	55832
rs34104736	ME	& AA	No table		
rs34104736	ME	& CA	1.00000	0.00000	49898
rs34104736	ME	& A	No table		
rs34104736	P	& C	0.00008	0.00006	55955
rs34104736	P	& AA	No table		
rs34104736	P	& CA	1.00000	0.00000	49833
rs34104736	P	& A	No table		
rs34104736	P	& ME	No table		

Locus: rs2282143

Pop	Genotypes:			
	2 2	2 4	4 4	Total
C	84	15	1	100
AA	83	17	0	100
CA	100	0	0	100
A	80	17	3	100
ME	100	0	0	100
P	100	0	0	100
Total:	547	49	4	600

Locus	Population pair		P-Value	S.E.	Switches

Locus	Population pair		P-Value	S.E.	Switches
rs2282143	AA	& C	1.00000	0.00000	59096
rs2282143	CA	& C	0.00005	0.00005	56442
rs2282143	CA	& AA	0.00005	0.00005	80716
rs2282143	A	& C	0.43272	0.00536	69334
rs2282143	A	& AA	0.42341	0.00614	67891
rs2282143	A	& CA	0.00000	0.00000	65942
rs2282143	ME	& C	0.00000	0.00000	56169
rs2282143	ME	& AA	0.00000	0.00000	80722
rs2282143	ME	& CA	No table		
rs2282143	ME	& A	0.00000	0.00000	66024
rs2282143	P	& C	0.00000	0.00000	56452
rs2282143	P	& AA	0.00000	0.00000	80631
rs2282143	P	& CA	No table		
rs2282143	P	& A	0.00000	0.00000	65974
rs2282143	P	& ME	No table		

Locus: rs34130495

Pop	Genotypes: 1 3	3 3	Total
C	0	100	100
AA	1	99	100
CA	2	98	100
A	0	100	100
ME	0	100	100
P	0	100	100
Total:	3	597	600

Locus	Population pair		P-Value	S.E.	Switches
rs34130495	AA	& C	1.00000	0.00000	50117
rs34130495	CA	& C	0.49770	0.00154	49783
rs34130495	CA	& AA	1.00000	0.00000	62487
rs34130495	A	& C	No table		
rs34130495	A	& AA	1.00000	0.00000	49989
rs34130495	A	& CA	0.49775	0.00164	49849
rs34130495	ME	& C	No table		
rs34130495	ME	& AA	1.00000	0.00000	49813
rs34130495	ME	& CA	0.49868	0.00152	49703
rs34130495	ME	& A	No table		
rs34130495	P	& C	No table		
rs34130495	P	& AA	1.00000	0.00000	50150
rs34130495	P	& CA	0.49670	0.00161	49771
rs34130495	P	& A	No table		
rs34130495	P	& ME	No table		

Locus: rs16891138

Pop Genotypes:

```
          1   1
          1   2     Total

C         100 0       100
AA        93  7       100
CA        100 0       100
A         100 0       100
ME        100 0       100
P         100 0       100

Total:    593 7       600
```

Locus	Population pair	P-Value	S.E.	Switches
rs16891138	AA & C	0.01354	0.00061	72442
rs16891138	CA & C	No table		
rs16891138	CA & AA	0.01475	0.00068	72436
rs16891138	A & C	No table		
rs16891138	A & AA	0.01396	0.00071	72123
rs16891138	A & CA	No table		
rs16891138	ME & C	No table		
rs16891138	ME & AA	0.01408	0.00077	72446
rs16891138	ME & CA	No table		
rs16891138	ME & A	No table		
rs16891138	P & C	No table		
rs16891138	P & AA	0.01299	0.00065	72346
rs16891138	P & CA	No table		
rs16891138	P & A	No table		
rs16891138	P & ME	No table		

Locus: rs35167514

```
==================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          1
          1     Total

C         100    100
AA        94      94
CA        67      67
A         100    100
ME        71      71
P         100    100

Total:    532    532
```

Locus	Population pair	P-Value	S.E.	Switches
rs35167514	AA & C	No table		
rs35167514	CA & C	No table		
rs35167514	CA & AA	No table		
rs35167514	A & C	No table		
rs35167514	A & AA	No table		
rs35167514	A & CA	No table		
rs35167514	ME & C	No table		
rs35167514	ME & AA	No table		
rs35167514	ME & CA	No table		
rs35167514	ME & A	No table		

```
rs35167514   P          & C          No table
rs35167514   P          & AA         No table
rs35167514   P          & CA         No table
rs35167514   P          & A          No table
rs35167514   P          & ME         No table

Locus: rs34305973
==================================
Pop        Genotypes:
           ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
           4
           4     Total

C          74     74
AA         94     94
CA         67     67
A          100    100
ME         71     71
P          100    100

Total:     506    506

Locus         Population pair          P-Value  S.E.     Switches
-----------   ----------------------   -------  -------  --------
rs34305973    AA         & C           No table
rs34305973    CA         & C           No table
rs34305973    CA         & AA          No table
rs34305973    A          & C           No table
rs34305973    A          & AA          No table
rs34305973    A          & CA          No table
rs34305973    ME         & C           No table
rs34305973    ME         & AA          No table
rs34305973    ME         & CA          No table
rs34305973    ME         & A           No table
rs34305973    P          & C           No table
rs34305973    P          & AA          No table
rs34305973    P          & CA          No table
rs34305973    P          & A           No table
rs34305973    P          & ME          No table

Locus: rs2297374
==================================
Pop        Genotypes:
           ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
           2   2   4
           2   4   4     Total

C          73  2   25      100
AA         71  16  13      100
CA         77  16  7       100
A          90  4   6       100
ME         90  10  0       100
P          100 0   0       100

Total:     501 48  51      600
```

Locus	Population pair	P-Value	S.E.	Switches
rs2297374	AA & C	0.43133	0.00830	81315
rs2297374	CA & C	0.04704	0.00337	80655
rs2297374	CA & AA	0.25582	0.00583	81015
rs2297374	A & C	0.00051	0.00015	73214
rs2297374	A & AA	0.00438	0.00068	79239
rs2297374	A & CA	0.09481	0.00366	77061
rs2297374	ME & C	0.00000	0.00000	77626
rs2297374	ME & AA	0.00006	0.00006	77966
rs2297374	ME & CA	0.00386	0.00059	74279
rs2297374	ME & A	0.40115	0.00395	70127
rs2297374	P & C	0.00000	0.00000	60280
rs2297374	P & AA	0.00000	0.00000	76237
rs2297374	P & CA	0.00000	0.00000	72123
rs2297374	P & A	0.00197	0.00032	61033
rs2297374	P & ME	0.00163	0.00027	74721

Locus: rs683369

Pop	Genotypes: 2 2 3 / 2 3 3	Total
C	12 2 86	100
AA	99 1 0	100
CA	88 11 1	100
A	96 4 0	100
ME	90 10 0	100
P	100 0 0	100
Total:	485 28 87	600

Locus	Population pair	P-Value	S.E.	Switches
rs683369	AA & C	0.00000	0.00000	69499
rs683369	CA & C	0.00000	0.00000	81230
rs683369	CA & AA	0.00252	0.00045	55263
rs683369	A & C	0.00000	0.00000	74987
rs683369	A & AA	0.36991	0.00250	68624
rs683369	A & CA	0.05615	0.00261	56595
rs683369	ME & C	0.00000	0.00000	80220
rs683369	ME & AA	0.01047	0.00065	76960
rs683369	ME & CA	0.67670	0.00471	57685
rs683369	ME & A	0.16287	0.00302	78611
rs683369	P & C	0.00000	0.00000	62775
rs683369	P & AA	1.00000	0.00000	50050
rs683369	P & CA	0.00054	0.00028	55017
rs683369	P & A	0.11895	0.00175	62417
rs683369	P & ME	0.00165	0.00021	74767

Locus: rs628031

Pop Genotypes:

```
---------------------------------------------------------------------
-----------------------------------------
          1   1   1   3
          1   2   3   3     Total

C         11  1   49  39     100
AA        6   0   41  53     100
CA        16  0   48  36     100
A         9   0   27  64     100
ME        0   0   43  57     100
P         0   0   14  86     100

Total:    42  1   222 335    600

Locus        Population pair        P-Value  S.E.     Switches
-----------  ---------------------  -------  -------  --------
rs628031     AA        & C          0.03446  0.00310     63542
rs628031     CA        & C          0.47494  0.00742     64773
rs628031     CA        & AA         0.00523  0.00072     83564
rs628031     A         & C          0.00625  0.00122     63910
rs628031     A         & AA         0.43979  0.00666     81358
rs628031     A         & CA         0.00009  0.00005     84236
rs628031     ME        & C          0.00073  0.00030     62119
rs628031     ME        & AA         0.24174  0.00672     75208
rs628031     ME        & CA         0.00031  0.00021     82216
rs628031     ME        & A          0.90755  0.00180     78317
rs628031     P         & C          0.00000  0.00000     62188
rs628031     P         & AA         0.00000  0.00000     74442
rs628031     P         & CA         0.00000  0.00000     81565
rs628031     P         & A          0.00001  0.00001     77449
rs628031     P         & ME         0.00000  0.00000     87605

Locus: MPJ6_O1004
=================================
Pop       Genotypes:
          ---------------------------------------------------------------
---------------------------------------------------------------------
---------------------------------------------------------------------
-----------------------------------------
          1   1   3
          1   3   3     Total

C         1   24  75     100
AA        0   0   0      0
CA        0   0   0      0
A         0   0   0      0
ME        0   0   0      0
P         0   0   0      0

Total:    1   24  75     100

Locus        Population pair        P-Value  S.E.     Switches
-----------  ---------------------  -------  -------  --------
MPJ6_O1004   AA        & C          No table
MPJ6_O1004   CA        & C          No table
MPJ6_O1004   CA        & AA         No table
MPJ6_O1004   A         & C          No table
MPJ6_O1004   A         & AA         No table
MPJ6_O1004   A         & CA         No table
MPJ6_O1004   ME        & C          No table
MPJ6_O1004   ME        & AA         No table
```

```
MPJ6_O1004   ME        & CA         No table
MPJ6_O1004   ME        & A          No table
MPJ6_O1004   P         & C          No table
MPJ6_O1004   P         & AA         No table
MPJ6_O1004   P         & CA         No table
MPJ6_O1004   P         & A          No table
MPJ6_O1004   P         & ME         No table

Locus: rs34205214
=================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          1   1   3
          1   3   3     Total

C         0   0   100     100
AA        1   4   95      100
CA        0   0   100     100
A         0   0   100     100
ME        0   0   100     100
P         0   0   100     100

Total:    1   4   595     600

Locus         Population pair        P-Value  S.E.     Switches
-----------   --------------------   -------  -------  --------
rs34205214    AA        & C          0.05774  0.00160     47958
rs34205214    CA        & C          No table
rs34205214    CA        & AA         0.05885  0.00164     47743
rs34205214    A         & C          No table
rs34205214    A         & AA         0.06126  0.00182     47729
rs34205214    A         & CA         No table
rs34205214    ME        & C          No table
rs34205214    ME        & AA         0.05614  0.00182     47836
rs34205214    ME        & CA         No table
rs34205214    ME        & A          No table
rs34205214    P         & C          No table
rs34205214    P         & AA         0.05965  0.00194     47774
rs34205214    P         & CA         No table
rs34205214    P         & A          No table
rs34205214    P         & ME         No table

Locus: rs117474883
=================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          2
          2     Total

C         100     100
AA        0       0
CA        0       0
A         0       0
ME        0       0
P         0       0
```

```
Total:    100    100

Locus        Population pair        P-Value  S.E.     Switches
-----------  ---------------------  -------  -------  --------
rs117474883  AA        & C          No table
rs117474883  CA        & C          No table
rs117474883  CA        & AA         No table
rs117474883  A         & C          No table
rs117474883  A         & AA         No table
rs117474883  A         & CA         No table
rs117474883  ME        & C          No table
rs117474883  ME        & AA         No table
rs117474883  ME        & CA         No table
rs117474883  ME        & A          No table
rs117474883  P         & C          No table
rs117474883  P         & AA         No table
rs117474883  P         & CA         No table
rs117474883  P         & A          No table
rs117474883  P         & ME         No table

Locus: rs35191146
==================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
----------------------------------------
          3
          3     Total

C         70     70
AA        94     94
CA        67     67
A         100    100
ME        71     71
P         100    100

Total:    502    502

Locus        Population pair        P-Value  S.E.     Switches
-----------  ---------------------  -------  -------  --------
rs35191146   AA        & C          No table
rs35191146   CA        & C          No table
rs35191146   CA        & AA         No table
rs35191146   A         & C          No table
rs35191146   A         & AA         No table
rs35191146   A         & CA         No table
rs35191146   ME        & C          No table
rs35191146   ME        & AA         No table
rs35191146   ME        & CA         No table
rs35191146   ME        & A          No table
rs35191146   P         & C          No table
rs35191146   P         & AA         No table
rs35191146   P         & CA         No table
rs35191146   P         & A          No table
rs35191146   P         & ME         No table

Locus: rs34059508
==================================
Pop       Genotypes:
```

```
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
----------------------------------------
          1   3
          3   3     Total

C         0   100    100
AA        0   100    100
CA        8   92     100
A         0   100    100
ME        0   100    100
P         0   100    100

Total:    8   592    600

Locus         Population pair         P-Value  S.E.     Switches
-----------   ---------------------   -------  -------  --------
rs34059508    AA        & C           No table
rs34059508    CA        & C           0.00626  0.00039     72084
rs34059508    CA        & AA          0.00676  0.00048     71967
rs34059508    A         & C           No table
rs34059508    A         & AA          No table
rs34059508    A         & CA          0.00741  0.00049     72238
rs34059508    ME        & C           No table
rs34059508    ME        & AA          No table
rs34059508    ME        & CA          0.00607  0.00050     72173
rs34059508    ME        & A           No table
rs34059508    P         & C           No table
rs34059508    P         & AA          No table
rs34059508    P         & CA          0.00670  0.00047     72293
rs34059508    P         & A           No table
rs34059508    P         & ME          No table

Locus: rs115733275
==================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
----------------------------------------
          3   4
          3   4     Total

C         0   100    100
AA        98  2      100
CA        0   0      0
A         0   0      0
ME        0   0      0
P         0   0      0

Total:    98  102    200

Locus         Population pair         P-Value  S.E.     Switches
-----------   ---------------------   -------  -------  --------
rs115733275   AA        & C           0.00000  0.00000     88436
rs115733275   CA        & C           No table
rs115733275   CA        & AA          No table
rs115733275   A         & C           No table
rs115733275   A         & AA          No table
rs115733275   A         & CA          No table
```

```
rs115733275  ME        & C          No table
rs115733275  ME        & AA         No table
rs115733275  ME        & CA         No table
rs115733275  ME        & A          No table
rs115733275  P         & C          No table
rs115733275  P         & AA         No table
rs115733275  P         & CA         No table
rs115733275  P         & A          No table
rs115733275  P         & ME         No table

Locus: rs116824962
==================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          2   4
          2   4     Total

C         100 0      100
AA        98  2      100
CA        0   0      0
A         0   0      0
ME        0   0      0
P         0   0      0

Total:    198 2      200

Locus        Population pair       P-Value  S.E.     Switches
-----------  --------------------  -------  -------  --------
rs116824962  AA        & C         0.49813  0.00160     49856
rs116824962  CA        & C         No table
rs116824962  CA        & AA        No table
rs116824962  A         & C         No table
rs116824962  A         & AA        No table
rs116824962  A         & CA        No table
rs116824962  ME        & C         No table
rs116824962  ME        & AA        No table
rs116824962  ME        & CA        No table
rs116824962  ME        & A         No table
rs116824962  P         & C         No table
rs116824962  P         & AA        No table
rs116824962  P         & CA        No table
rs116824962  P         & A         No table
rs116824962  P         & ME        No table

Locus: rs118060798
==================================
Pop       Genotypes:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          1
          1     Total

C         100    100
AA        0      0
CA        0      0
A         0      0
```

```
ME        0      0
P         0      0

Total:    100    100

Locus        Population pair        P-Value  S.E.     Switches
-----------  ---------------------  -------  -------  --------
rs118060798  AA        & C          No table
rs118060798  CA        & C          No table
rs118060798  CA        & AA         No table
rs118060798  A         & C          No table
rs118060798  A         & AA         No table
rs118060798  A         & CA         No table
rs118060798  ME        & C          No table
rs118060798  ME        & AA         No table
rs118060798  ME        & CA         No table
rs118060798  ME        & A          No table
rs118060798  P         & C          No table
rs118060798  P         & AA         No table
rs118060798  P         & CA         No table
rs118060798  P         & A          No table
rs118060798  P         & ME         No table

==================================

P-value for each population pair across all loci
(Fisher's method)
-----------------------------------------------------
Population pair             Chi2      df   P-Value
--------------------        --------  ---  --------
C             & AA           Infinity 24   Highly sign.
C             & CA           Infinity 20   Highly sign.
AA            & CA          84.75155  24   0.000000
C             & A            Infinity 14   Highly sign.
AA            & A            Infinity 16   Highly sign.
CA            & A            Infinity 20   Highly sign.
C             & ME           Infinity 16   Highly sign.
AA            & ME           Infinity 18   Highly sign.
CA            & ME           Infinity 18   Highly sign.
A             & ME           Infinity 12   Highly sign.
C             & P            Infinity 14   Highly sign.
AA            & P            Infinity 16   Highly sign.
CA            & P            Infinity 18   Highly sign.
A             & P            Infinity 10   Highly sign.
ME            & P            Infinity 10   Highly sign.

Normal ending.
```

4. Genic differentiation for each population pair (exact G test)

File: 182531 (Title line:"Total 20 SNPs")

Number of populations detected : 6
Number of loci detected : 20

Markov chain parameters
Dememorisation : 1000
Batches : 100
Iterations per batch : 1000

Locus: rs36103319
==========

Pop	Alleles:		
	3	4	Total
C	196	4	200
AA	200	0	200
CA	199	1	200
A	200	0	200
ME	200	0	200
P	200	0	200
Total:	11955		1200

Locus	Population pair	P-Value	S.E.	Switches
rs36103319	AA & C	0.12522	0.00177	62083
rs36103319	CA & C	0.37183	0.00259	68588
rs36103319	CA & AA	1.00000	0.00000	49764
rs36103319	A & C	0.12345	0.00163	62297
rs36103319	A & AA	No table		
rs36103319	A & CA	1.00000	0.00000	50131
rs36103319	ME & C	0.12244	0.00165	62201
rs36103319	ME & AA	No table		
rs36103319	ME & CA	1.00000	0.00000	50164
rs36103319	ME & A	No table		
rs36103319	P & C	0.12202	0.00170	62396
rs36103319	P & AA	No table		
rs36103319	P & CA	1.00000	0.00000	49962
rs36103319	P & A	No table		
rs36103319	P & ME	No table		

Locus: rs12208357
==========

Pop	Alleles:		
	2	4	Total
C	0	200	200
AA	0	200	200

CA	7	193	200
A	0	200	200
ME	1	199	200
P	0	200	200
Total:	8	1192	1200

Locus	Population pair	P-Value	S.E.	Switches
rs12208357	AA & C	No table		
rs12208357	CA & C	0.01382	0.00066	72433
rs12208357	CA & AA	0.01498	0.00070	72365
rs12208357	A & C	No table		
rs12208357	A & AA	No table		
rs12208357	A & CA	0.01405	0.00063	72216
rs12208357	ME & C	1.00000	0.00000	49890
rs12208357	ME & AA	1.00000	0.00000	50035
rs12208357	ME & CA	0.06622	0.00165	72305
rs12208357	ME & A	1.00000	0.00000	50037
rs12208357	P & C	No table		
rs12208357	P & AA	No table		
rs12208357	P & CA	0.01351	0.00060	72704
rs12208357	P & A	No table		
rs12208357	P & ME	1.00000	0.00000	50438

Locus: rs1867351

Pop	Alleles: 2	4	Total
C	36	164	200
AA	54	146	200
CA	35	165	200
A	114	86	200
ME	144	56	200
P	171	29	200
Total:	554	646	1200

Locus	Population pair	P-Value	S.E.	Switches
rs1867351	AA & C	0.04557	0.00324	90261
rs1867351	CA & C	1.00000	0.00000	89524
rs1867351	CA & AA	0.03034	0.00259	90623
rs1867351	A & C	0.00000	0.00000	91989
rs1867351	A & AA	0.00000	0.00000	92061
rs1867351	A & CA	0.00000	0.00000	92009
rs1867351	ME & C	0.00000	0.00000	91901
rs1867351	ME & AA	0.00000	0.00000	91953
rs1867351	ME & CA	0.00000	0.00000	91833
rs1867351	ME & A	0.00291	0.00063	91518
rs1867351	P & C	0.00000	0.00000	91937
rs1867351	P & AA	0.00000	0.00000	92119
rs1867351	P & CA	0.00000	0.00000	92152
rs1867351	P & A	0.00000	0.00000	91305
rs1867351	P & ME	0.00211	0.00073	90230

Locus: rs34104736

Pop	Alleles: 2	4	Total
C	171	29	200
AA	200	0	200
CA	199	1	200
A	200	0	200
ME	200	0	200
P	200	0	200
Total:	1170	30	1200

Locus	Population pair	P-Value	S.E.	Switches
rs34104736	AA & C	0.00000	0.00000	85353
rs34104736	CA & C	0.00000	0.00000	84726
rs34104736	CA & AA	1.00000	0.00000	49991
rs34104736	A & C	0.00000	0.00000	85048
rs34104736	A & AA	No table		
rs34104736	A & CA	1.00000	0.00000	50002
rs34104736	ME & C	0.00000	0.00000	85103
rs34104736	ME & AA	No table		
rs34104736	ME & CA	1.00000	0.00000	49986
rs34104736	ME & A	No table		
rs34104736	P & C	0.00000	0.00000	84800
rs34104736	P & AA	No table		
rs34104736	P & CA	1.00000	0.00000	49774
rs34104736	P & A	No table		
rs34104736	P & ME	No table		

Locus: rs2282143

Pop	Alleles: 2	4	Total
C	183	17	200
AA	183	17	200
CA	200	0	200
A	177	23	200
ME	200	0	200
P	200	0	200
Total:	1143	57	1200

Locus	Population pair	P-Value	S.E.	Switches
rs2282143	AA & C	1.00000	0.00000	85848
rs2282143	CA & C	0.00000	0.00000	81000
rs2282143	CA & AA	0.00003	0.00002	81096
rs2282143	A & C	0.40702	0.00524	86616
rs2282143	A & AA	0.40945	0.00518	86696

rs2282143	A	& CA	0.00000	0.00000	83345
rs2282143	ME	& C	0.00000	0.00000	81007
rs2282143	ME	& AA	0.00000	0.00000	81299
rs2282143	ME	& CA	No table		
rs2282143	ME	& A	0.00000	0.00000	83476
rs2282143	P	& C	0.00000	0.00000	81109
rs2282143	P	& AA	0.00001	0.00001	81173
rs2282143	P	& CA	No table		
rs2282143	P	& A	0.00000	0.00000	83436
rs2282143	P	& ME	No table		

Locus: rs34130495

Pop	Alleles:		
	1	3	Total
C	0	200	200
AA	1	199	200
CA	2	198	200
A	0	200	200
ME	0	200	200
P	0	200	200
Total:	3	1197	1200

Locus	Population pair		P-Value	S.E.	Switches
rs34130495	AA	& C	1.00000	0.00000	50006
rs34130495	CA	& C	0.49850	0.00163	49791
rs34130495	CA	& AA	1.00000	0.00000	62210
rs34130495	A	& C	No table		
rs34130495	A	& AA	1.00000	0.00000	49867
rs34130495	A	& CA	0.50206	0.00177	49820
rs34130495	ME	& C	No table		
rs34130495	ME	& AA	1.00000	0.00000	50027
rs34130495	ME	& CA	0.49876	0.00143	50072
rs34130495	ME	& A	No table		
rs34130495	P	& C	No table		
rs34130495	P	& AA	1.00000	0.00000	50169
rs34130495	P	& CA	0.49787	0.00161	49979
rs34130495	P	& A	No table		
rs34130495	P	& ME	No table		

Locus: rs16891138

Pop	Alleles:		
	1	2	Total
C	200	0	200
AA	193	7	200
CA	200	0	200
A	200	0	200
ME	200	0	200

```
P         200 0      200

Total:    11937      1200

Locus        Population pair        P-Value  S.E.     Switches
-----------  ---------------------  -------  -------  --------
rs16891138   AA        & C          0.01489  0.00071     72506
rs16891138   CA        & C          No table
rs16891138   CA        & AA         0.01485  0.00073     72511
rs16891138   A         & C          No table
rs16891138   A         & AA         0.01332  0.00065     72472
rs16891138   A         & CA         No table
rs16891138   ME        & C          No table
rs16891138   ME        & AA         0.01571  0.00067     72259
rs16891138   ME        & CA         No table
rs16891138   ME        & A          No table
rs16891138   P         & C          No table
rs16891138   P         & AA         0.01412  0.00071     72515
rs16891138   P         & CA         No table
rs16891138   P         & A          No table
rs16891138   P         & ME         No table

Locus: rs35167514
==================================
Pop       Alleles:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
----------------------------------------
          1     Total

C         200    200
AA        194    194
CA        165    165
A         200    200
ME        157    157
P         200    200

Total:    1116   1116

Locus        Population pair        P-Value  S.E.     Switches
-----------  ---------------------  -------  -------  --------
rs35167514   AA        & C          No table
rs35167514   CA        & C          No table
rs35167514   CA        & AA         No table
rs35167514   A         & C          No table
rs35167514   A         & AA         No table
rs35167514   A         & CA         No table
rs35167514   ME        & C          No table
rs35167514   ME        & AA         No table
rs35167514   ME        & CA         No table
rs35167514   ME        & A          No table
rs35167514   P         & C          No table
rs35167514   P         & AA         No table
rs35167514   P         & CA         No table
rs35167514   P         & A          No table
rs35167514   P         & ME         No table

Locus: rs34305973
==================================
Pop       Alleles:
```

```
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          4     Total

C         174    174
AA        194    194
CA        165    165
A         200    200
ME        157    157
P         200    200

Total:    1090   1090

Locus         Population pair         P-Value  S.E.     Switches
-----------  ---------------------  -------  -------  --------
rs34305973   AA         & C          No table
rs34305973   CA         & C          No table
rs34305973   CA         & AA         No table
rs34305973   A          & C          No table
rs34305973   A          & AA         No table
rs34305973   A          & CA         No table
rs34305973   ME         & C          No table
rs34305973   ME         & AA         No table
rs34305973   ME         & CA         No table
rs34305973   ME         & A          No table
rs34305973   P          & C          No table
rs34305973   P          & AA         No table
rs34305973   P          & CA         No table
rs34305973   P          & A          No table
rs34305973   P          & ME         No table

Locus: rs2297374
================================
Pop       Alleles:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          2   4     Total

C         148 52     200
AA        158 42     200
CA        170 30     200
A         184 16     200
ME        190 10     200
P         200 0      200

Total:    1050150    1200

Locus         Population pair         P-Value  S.E.     Switches
-----------  ---------------------  -------  -------  --------
rs2297374    AA         & C          0.28067  0.00740     90733
rs2297374    CA         & C          0.00960  0.00138     89892
rs2297374    CA         & AA         0.16478  0.00559     89409
rs2297374    A          & C          0.00000  0.00000     89247
rs2297374    A          & AA         0.00056  0.00025     88796
rs2297374    A          & CA         0.04433  0.00260     87261
rs2297374    ME         & C          0.00000  0.00000     89051
rs2297374    ME         & AA         0.00000  0.00000     88094
```

rs2297374	ME	& CA	0.00075	0.00021	87129
rs2297374	ME	& A	0.31630	0.00488	83817
rs2297374	P	& C	0.00000	0.00000	88136
rs2297374	P	& AA	0.00000	0.00000	87098
rs2297374	P	& CA	0.00000	0.00000	85029
rs2297374	P	& A	0.00001	0.00001	79755
rs2297374	P	& ME	0.00199	0.00028	75083

Locus: rs683369

Pop	Alleles:		
	2	3	Total
C	26	174	200
AA	199	1	200
CA	187	13	200
A	196	4	200
ME	190	10	200
P	200	0	200
Total:	998	202	1200

Locus	Population pair		P-Value	S.E.	Switches
rs683369	AA	& C	0.00000	0.00000	92233
rs683369	CA	& C	0.00000	0.00000	92007
rs683369	CA	& AA	0.00146	0.00023	78805
rs683369	A	& C	0.00000	0.00000	92098
rs683369	A	& AA	0.37015	0.00217	68680
rs683369	A	& CA	0.04415	0.00173	81316
rs683369	ME	& C	0.00000	0.00000	92228
rs683369	ME	& AA	0.01018	0.00076	77061
rs683369	ME	& CA	0.66415	0.00338	83528
rs683369	ME	& A	0.17146	0.00264	78698
rs683369	P	& C	0.00000	0.00000	91896
rs683369	P	& AA	1.00000	0.00000	50011
rs683369	P	& CA	0.00024	0.00008	78631
rs683369	P	& A	0.12106	0.00163	62188
rs683369	P	& ME	0.00162	0.00023	75047

Locus: rs628031

Pop	Alleles:			
	1	2	3	Total
C	72	1	127	200
AA	53	0	147	200
CA	80	0	120	200
A	45	0	155	200
ME	43	0	157	200
P	14	0	186	200
Total:	307	1	892	1200

Locus	Population pair	P-Value	S.E.	Switches
rs628031	AA & C	0.03915	0.00558	62324
rs628031	CA & C	0.48299	0.01367	62541
rs628031	CA & AA	0.00625	0.00114	91683
rs628031	A & C	0.00403	0.00155	62197
rs628031	A & AA	0.41825	0.00723	90746
rs628031	A & CA	0.00008	0.00008	91554
rs628031	ME & C	0.00107	0.00048	62011
rs628031	ME & AA	0.27639	0.00836	90830
rs628031	ME & CA	0.00008	0.00005	91500
rs628031	ME & A	0.90381	0.00189	90258
rs628031	P & C	0.00000	0.00000	62068
rs628031	P & AA	0.00000	0.00000	89478
rs628031	P & CA	0.00000	0.00000	90710
rs628031	P & A	0.00000	0.00000	88996
rs628031	P & ME	0.00001	0.00001	88887

Locus: MPJ6_O1004

Pop Alleles:

	1	3	Total
C	26	174	200
AA	0	0	0
CA	0	0	0
A	0	0	0
ME	0	0	0
P	0	0	0
Total:	26	174	200

Locus	Population pair	P-Value	S.E.	Switches
MPJ6_O1004	AA & C	No table		
MPJ6_O1004	CA & C	No table		
MPJ6_O1004	CA & AA	No table		
MPJ6_O1004	A & C	No table		
MPJ6_O1004	A & AA	No table		
MPJ6_O1004	A & CA	No table		
MPJ6_O1004	ME & C	No table		
MPJ6_O1004	ME & AA	No table		
MPJ6_O1004	ME & CA	No table		
MPJ6_O1004	ME & A	No table		
MPJ6_O1004	P & C	No table		
MPJ6_O1004	P & AA	No table		
MPJ6_O1004	P & CA	No table		
MPJ6_O1004	P & A	No table		
MPJ6_O1004	P & ME	No table		

Locus: rs34205214

Pop Alleles:

```
------------------------------------------------------------------------
----------------------------------------
          1   3     Total

C         0   200    200
AA        6   194    200
CA        0   200    200
A         0   200    200
ME        0   200    200
P         0   200    200

Total:    6   1194   1200

Locus         Population pair         P-Value  S.E.     Switches
-----------   ---------------------   -------  -------  --------
rs34205214    AA        & C           0.03003  0.00099     68537
rs34205214    CA        & C           No table
rs34205214    CA        & AA          0.03039  0.00109     68539
rs34205214    A         & C           No table
rs34205214    A         & AA          0.03018  0.00101     68757
rs34205214    A         & CA          No table
rs34205214    ME        & C           No table
rs34205214    ME        & AA          0.02986  0.00099     68697
rs34205214    ME        & CA          No table
rs34205214    ME        & A           No table
rs34205214    P         & C           No table
rs34205214    P         & AA          0.02982  0.00094     68773
rs34205214    P         & CA          No table
rs34205214    P         & A           No table
rs34205214    P         & ME          No table

Locus: rs117474883
==================================
Pop       Alleles:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
----------------------------------------
          2     Total

C         200    200
AA        0      0
CA        0      0
A         0      0
ME        0      0
P         0      0

Total:    200    200

Locus         Population pair         P-Value  S.E.     Switches
-----------   ---------------------   -------  -------  --------
rs117474883   AA        & C           No table
rs117474883   CA        & C           No table
rs117474883   CA        & AA          No table
rs117474883   A         & C           No table
rs117474883   A         & AA          No table
rs117474883   A         & CA          No table
rs117474883   ME        & C           No table
rs117474883   ME        & AA          No table
rs117474883   ME        & CA          No table
rs117474883   ME        & A           No table
```

```
rs117474883  P          & C          No table
rs117474883  P          & AA         No table
rs117474883  P          & CA         No table
rs117474883  P          & A          No table
rs117474883  P          & ME         No table

Locus: rs35191146
==================================
Pop       Alleles:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          3     Total

C         169    169
AA        194    194
CA        165    165
A         200    200
ME        157    157
P         200    200

Total:    1085   1085

Locus         Population pair         P-Value  S.E.     Switches
-----------   ---------------------   -------  -------  --------
rs35191146    AA         & C          No table
rs35191146    CA         & C          No table
rs35191146    CA         & AA         No table
rs35191146    A          & C          No table
rs35191146    A          & AA         No table
rs35191146    A          & CA         No table
rs35191146    ME         & C          No table
rs35191146    ME         & AA         No table
rs35191146    ME         & CA         No table
rs35191146    ME         & A          No table
rs35191146    P          & C          No table
rs35191146    P          & AA         No table
rs35191146    P          & CA         No table
rs35191146    P          & A          No table
rs35191146    P          & ME         No table

Locus: rs34059508
==================================
Pop       Alleles:
          ------------------------------------------------------------
------------------------------------------------------------------------
------------------------------------------------------------------------
-----------------------------------------
          1   3     Total

C         0   200    200
AA        0   200    200
CA        8   192    200
A         0   200    200
ME        0   200    200
P         0   200    200

Total:    8   1192   1200

Locus         Population pair         P-Value  S.E.     Switches
```

Locus	Population pair		P-Value	S.E.	Switches
rs34059508	AA	& C	No table		
rs34059508	CA	& C	0.00788	0.00052	72145
rs34059508	CA	& AA	0.00784	0.00051	72516
rs34059508	A	& C	No table		
rs34059508	A	& AA	No table		
rs34059508	A	& CA	0.00833	0.00061	72260
rs34059508	ME	& C	No table		
rs34059508	ME	& AA	No table		
rs34059508	ME	& CA	0.00780	0.00051	72173
rs34059508	ME	& A	No table		
rs34059508	P	& C	No table		
rs34059508	P	& AA	No table		
rs34059508	P	& CA	0.00725	0.00053	72453
rs34059508	P	& A	No table		
rs34059508	P	& ME	No table		

Locus: rs115733275

Pop	Alleles:		
	3	4	Total
C	0	200	200
AA	196	4	200
CA	0	0	0
A	0	0	0
ME	0	0	0
P	0	0	0
Total:	196	204	400

Locus	Population pair		P-Value	S.E.	Switches
rs115733275	AA	& C	0.00000	0.00000	92215
rs115733275	CA	& C	No table		
rs115733275	CA	& AA	No table		
rs115733275	A	& C	No table		
rs115733275	A	& AA	No table		
rs115733275	A	& CA	No table		
rs115733275	ME	& C	No table		
rs115733275	ME	& AA	No table		
rs115733275	ME	& CA	No table		
rs115733275	ME	& A	No table		
rs115733275	P	& C	No table		
rs115733275	P	& AA	No table		
rs115733275	P	& CA	No table		
rs115733275	P	& A	No table		
rs115733275	P	& ME	No table		

Locus: rs116824962

Pop	Alleles:		
	2	4	Total

C	200	0	200
AA	196	4	200
CA	0	0	0
A	0	0	0
ME	0	0	0
P	0	0	0
Total:	396	4	400

Locus	Population pair	P-Value	S.E.	Switches
rs116824962	AA & C	0.12076	0.00154	62329
rs116824962	CA & C	No table		
rs116824962	CA & AA	No table		
rs116824962	A & C	No table		
rs116824962	A & AA	No table		
rs116824962	A & CA	No table		
rs116824962	ME & C	No table		
rs116824962	ME & AA	No table		
rs116824962	ME & CA	No table		
rs116824962	ME & A	No table		
rs116824962	P & C	No table		
rs116824962	P & AA	No table		
rs116824962	P & CA	No table		
rs116824962	P & A	No table		
rs116824962	P & ME	No table		

Locus: rs118060798

Pop Alleles:

	1	Total
C	200	200
AA	0	0
CA	0	0
A	0	0
ME	0	0
P	0	0
Total:	200	200

Locus	Population pair	P-Value	S.E.	Switches
rs118060798	AA & C	No table		
rs118060798	CA & C	No table		
rs118060798	CA & AA	No table		
rs118060798	A & C	No table		
rs118060798	A & AA	No table		
rs118060798	A & CA	No table		
rs118060798	ME & C	No table		
rs118060798	ME & AA	No table		
rs118060798	ME & CA	No table		
rs118060798	ME & A	No table		
rs118060798	P & C	No table		
rs118060798	P & AA	No table		
rs118060798	P & CA	No table		

```
rs118060798  P          & A           No table
rs118060798  P          & ME          No table

================================

P-value for each population pair across all loci
(Fisher's method)
--------------------------------------------------------
Population pair                Chi2       df   P-Value
--------------------           --------   ---  --------
C              & AA             Infinity 24   Highly sign.
C              & CA             Infinity 20   Highly sign.
AA             & CA            88.14038  24   0.000000
C              & A              Infinity 14   Highly sign.
AA             & A              Infinity 16   Highly sign.
CA             & A              Infinity 20   Highly sign.
C              & ME             Infinity 16   Highly sign.
AA             & ME             Infinity 18   Highly sign.
CA             & ME             Infinity 18   Highly sign.
A              & ME             Infinity 12   Highly sign.
C              & P              Infinity 14   Highly sign.
AA             & P              Infinity 16   Highly sign.
CA             & P              Infinity 18   Highly sign.
A              & P              Infinity 10   Highly sign.
ME             & P             60.637885 10   0.000000

Normal ending.
```