The assessment of forensic molecular markers for skin colour in South Africans

Akshay Vanmali

VMLAKS001

Supervisor: Dr Laura Heathfield

Co-supervisor: Calvin Mole

Division of Forensic Medicine and Toxicology, Faculty of Health Sciences

University of Cape Town

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN

In partial fulfilment of the requirement for the degree

MPhil (Biomedical Forensic Science)

Submission: December 2019

Plagiarism declaration

I, **Akshay Vanmali**, hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorise the university to reproduce for the purpose of research either the whole or any portion of the content.

This thesis/dissertation has been submitted to the Turnitin module (or equivalent similarity and originality checking software) and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisors in any manner whatsoever.

Referencing style: Chicago Manual of Style 17th edition (author-date)

Name: Akshay Vanmali **Student number:** VMLAKS001

Signature: Signed by candidate **Date:** 02/12/2019

Abstract

The scientific development of innovative molecular techniques has transformed the approach towards human identification. In forensic casework, the emergence of molecular phenotyping, or phenotypic prediction from DNA, has mitigated some challenges involving the unavailability of references samples for traditional forensic DNA analysis. Molecular phenotyping via SNP analysis can be used as a tool in a forensic setting to predict physical traits, such as hair, skin and eye colour, and provide investigative leads. Several ancestry informative markers (AIMs) have previously been associated with human skin colour in mainly the European and North American population groups, while admixed populations are hardly studied. The present study aims to contribute towards this gap by investigating the relationship between two AIMs (*SLC45A2*, rs16891982 and *SLC24A5*, rs1426654) that are typically involved in molecular phenotyping, and melanin index (MI) in the South African (SA) metapopulation (n = 389). The self-reported ancestry, ethnicity and relevant biographic information for each participant was documented and MI was recorded using a derma-spectrophotometer. DNA was extracted from saliva samples and PCR amplification of target regions was performed. Thereafter, SNaPshot® PCR was used to genotype the variants. Significant differences ($p < 0.0001$) were observed between MI readings and ancestral as well as population census groups. A generalised linear model (GLM) was developed which could accurately predicted the MI readings for each genotype combination within the 95 % confidence interval of the recorded MI readings. Our results suggest that these two markers were consistently associated with MI in the admixed SA population and are thus informative to predict MI in a forensic setting. Finally, this was the first study in a SA context to use SNP analysis for objective MI prediction.

Key words: admixture, ancestry informative markers, melanin index, molecular phenotyping

Acknowledgements

My sincere gratitude and thanks to Laura Heathfield, who played the roles of supervisor and academic support. Thank you for ever so patiently reading through the drafts of my writing and providing efficient and useful feedback. In addition, I would like to acknowledge my co - supervisor, Calvin Mole, for his assistance with the statistical analysis of this project and Andrea Gibbon, who provided me with the necessary technical support during my laboratory work.

A special thanks goes to my colleagues in the Division of Forensic Medicine and Toxicology, I will always treasure the bonds we have created.

To my fellow MPhil students, the support structure we have built as a group has assisted each of us along this painful but joyous 2 years. I wish all the best for your futures and will always treasure our friendship.

And finally, I would like to thank my family members who played the role of emotional support. Thank you for enduring the endless complaints, sharing the excitement of positive results and providing some sort of reassurance about my capabilities and achievements throughout this year. It has been an emotional but memorable year but an enriching experience.

Table of Contents

List of Tables

List of Figures

Abbreviations

°C	degree Celsius
A	adenine
A_{230}	Absorbance at 230 nanometre
A_{260}	Absorbance at 260 nanometre
A_{280}	Absorbance at 280 nanometre
AIC	Akaike Information Criterion
AIM	ancestry informative marker
al	allele
Ala	Alanine
ASIP	Agouti Signalling Protein
BGA	biogeographical ancestry
bp	base pair
C	cytosine
c.	coding region
ddNTP	dideoxynucleotide triphosphate
DEU	Germany
DNA	deoxyribonucleic acid
EVC	external visible characteristic
ExoSAP	*Exonuclease* I and Shrimp Alkaline Phosphatase
G	guanine
GLM	generalised linear model
HERC2	HECT and RLD domain containing E3 ubiquitin protein ligase 2
Hi-Di	Highly deionized

HREC	Human Research Ethics Committee
ht	height
HWE	Hardy Weinberg Equilibrium
ICMP	International Commission on Missing Persons
INDEL	insertion/deletion
IUPAC	International Union of Pure and Applied Chemistry
KITLG	Kit ligand
Leu	Leucine
LLC	limited liability company
MATP	membrane-associated transporter protein
MC1R	Melanocortin 1 receptor
MI	melanin index
mL	millilitre
MSK	molecular phenotype
MWM	molecular weight marker
n	sample size
N	population size
NFDD	National Forensic DNA Database of South Africa
ng	nanogram
NGS	next generation sequencing
NTC	non-template control
OCA	Oculocutaneous albinism
p	p - value
p.	protein reference sequence

PCR	polymerase chain reaction
Phe	Phenylalanine
Q1	1st quartile
Q3	3rd quartile
RFU	relative fluorescent unit
rpm	revolutions per minute
rs	reference SNP
rSAP	recombinant Shrimp Alkaline Phosphatase
SA	South Africa
SAC	South African Coloured
SAPS	South African Police Services
SLC45A2	Solute Carrier Family 45 Member 2
SLC24A4	Solute Carrier Family 24 Member 4
SLC24A5	Solute Carrier Family 24 Member 5
SNP	single nucleotide polymorphism
STR	short tandem repeats
sz	size
T_a	annealing temperature
TBE	Tris-borate-EDTA
Thr	Threonine
TPCN2	Two Pore Segment Channel 2
TYR	Tyrosinase
UCT	University of Cape Town
UK	United Kingdom

µL	microlitre
µM	micromole
USA	United States of America
V	volts
v/v	volume per volume
w/v	weight per volume

Chapter 1: Introduction and literature review

1.1 Introduction

Over recent years, scientific advancements in molecular techniques have contributed to a paradigm shift in forensic identification (Walsh 2005). These advancements in genetic analysis have broadened our knowledge of genetic concepts which provides a basis to convey beneficial information within a forensic context. The application of these molecular techniques has become a focal point in mitigating the issues of human identification (Walsh 2005).

Forensic deoxyribonucleic acid (DNA) analysis has revolutionised the approach towards human identification, as it has been internationally recognised as the golden standard for identifying individuals (Kayser 2015). The use of genetics in forensic science has allowed DNA analysis to (i) identify or exclude a known suspect; (ii) link potential suspects, victims or witnesses to a crime scene; (iii) identify missing individuals/unidentified human remains; and (iv) perform kinship analysis (Kayser and Schneider 2009). Furthermore, the relatively recent discovery of molecular markers that are associated with external physical features and ancestry have enhanced DNA analysis to generate new investigative leads to identify potential suspects and assist with victim identification whereby the individual is beyond visual recognition (Kayser and Schneider 2009).

The estimation of externally visible characteristics (EVCs), or molecular phenotyping, is thought to be of value in a diverse population such as that in South Africa (SA); however, the local relevance of this technique needs to be assessed and placed into the SA context. Thus, it is vital to evaluate the discriminatory power of known molecular markers within the SA population while also addressing the ethical, social and legal implications that overshadow this forensic tool (Slabbert and Heathfield 2018).

1.2 Human identification

1.2.1 Forensic DNA profiling

1.2.1.1 Principle

DNA profiling is globally considered a common forensic tool used for human identification (Butler 2010; Walsh 2005). The principle of this technique involves assessing a DNA profile of a unknown individual compared to a DNA profile from a known origin (Kayser 2015).

This allows for a comparative analysis of genetic data between biological material found at the scene and potential victims or suspects. The biological material which is used for genetic analysis includes, but is not limited to, bodily fluids, skin tissue and buccal cells, hair and bones (Butler 2010).

Comparative genetic analysis can be achieved because DNA is comprised of a number of specific regions that are found abundantly across the genome, called short tandem repeats (STRs), which have a high degree of variability between individuals (Butler 2010). STRs that are associated with forensic DNA profiling are located in the non-functional portion of the DNA that comprises of intronic regions (non-coding) that do not encode protein sequences. The variation in the STR sequences is attributed to the frequency at which the base pair (bp) motif replicates itself consecutively at a particular locus. A STR genotype is based on the number of repeated motif units of the alleles detected at a locus. The combination of several STR genotypes forms a DNA profile, which can assist in human identification (Butler 2010).

1.2.1.2 Limitations

Although forensic DNA profiling is globally considered a common practice for investigative purposes (Singh 2011; Gans and Urbas 2002), the practicality of this technique can be limited by the absence of a reference profile. This may include no matches to a national DNA database of DNA profiles, the absence of suspects or living relatives, or the unavailability of biological references samples from missing individuals (Kayser 2015; Kayser and Schneider 2009).

The challenges associated with forensic DNA profiling has prompted an emphasis on expanding research into predicting externally visible traits of an individual based on genetic analysis of biological material (Kayser and Schneider 2009; Walsh et al. 2014).

1.2.2 Single nucleotide polymorphism (SNP)-based forensic identification

Apart from STR-based DNA analysis, another form of genetic variation, known as SNPs, could play a fundamental role in human identification (Amorim and Pereira 2005). SNPs are single bp variants which occur at specific sites in the genome (Butler 2010). SNPs represent the most common form of genetic variation with an estimate of five million SNPs located in an individual's genome (Sobrino, Brión, and Carracedo 2005). The advantages and disadvantages of using SNPs as a primary identification tool in forensics have been widely debated as previously reviewed by Butler, Coble, and Vallone (2007). Taking into

consideration that the shortcomings of SNPs currently outweigh its benefits, SNPs are not ready to supersede STRs as the mainstay for markers in forensic identification, but it is highly recommended that SNP analysis should supplement STR analysis (Butler, Coble, and Vallone 2007; Gill et al. 2004).

1.2.2.1 Molecular phenotyping

Although SNPs may not replace STRs as the principal forensic molecular markers in human identification applications, they can potentially play a beneficial role in niche applications such as molecular phenotyping (Butler, Coble, and Vallone 2007). SNPs can alter gene expression resulting in phenotypic differentiations which forms the basis of molecular phenotyping. This technique is used to predict observable physical traits of an individual through analysis of informative phenotypic SNPs (Kayser 2015; Kayser and Schneider 2009; Kidd et al. 2006; Budowle and Van Daal 2008). The analysis of informative SNPs can be valuable in a forensic setting to provide an investigative lead to a suspect as well as assist with victim identification (Kayser and Schneider 2009; Jobling and Gill 2004).

According to the 2018 Crime Statistics for South Africa, the total number of unnatural deaths in SA was 522 157, with more than 1 400 missing person cases reported to the South Africa Police Services (SAPS) (Department of Police 2018). In addition, data for an 8-year period at Salt River Mortuary, Cape Town, SA indicated 9 % of all cases were unidentified after initial investigation (Reid, Martin, and Heathfield 2019). These alarming statistics signify the importance of utilising molecular phenotyping in forensic practice whereby victim identification is not possible due to extensive decomposition, burning or skeletonisation.

SNPs influence the EVCs of an individual which comprise of, inter alia, hair, skin and eye colour, dimpling and freckling (MacLean and Lamparello 2014). Forecasting an individual's phenotype is based on inference made by determining the likelihood of certain traits being expressed instead of another (Castel 2014). This concept has prompted the development of various SNP assays such as IrisPlex, HIrisPlex and HIrisPlex-S (Walsh et al. 2011a, 2014; Chaitanya et al. 2018). The IrisPlex assay is a robust and sensitive SNaPshot®-based eye colour prediction tool (blue and brown) which comprises six highly informative eye colour SNPs (Walsh et al. 2011a, 2011b). This prediction tool was successfully utilised with a high degree of accuracy across the Western populations (Purps et al. 2011; Walsh et al. 2011b; Walsh et al. 2012; Dario et al. 2015) and highly admixed populations (Freire-Aradas et al. 2014; Hohl et al. 2018; Prestes et al. 2011). The IrisPlex assay has been validated for use in

criminal cases (Walsh et al. 2011b). For example, Hollard et al. (2017) utilised the IrisPlex system to predict brown eye colour for a decedent with extensive charring of the body as a result of fire with 95 % probability (Hollard et al. 2017).

Additionally, a well-established phenotypic SNP assay known as HIrisPlex was developed which consists of 23 SNPs (including the six IrisPlex SNPs) and an insertion/deletion (INDEL). The HIrisPlex system can predict eye and hair colour and is considered an improvement from the IrisPlex as it demonstrates greater sensitivity (Walsh et al. 2014; Chaitanya et al. 2018). The HIrisPlex assay has successfully been used on ancient DNA samples, wherein it was utilised to assist in victim identification, by ascertaining the EVCs, of Jörg Jenatsch, a freedom fighter assassinated in 1639 (Haeusler et al. 2016). Chaitanya et al. (2017) applied HIrisPlex to retrieve information about EVCs from remains of World War II victims while Draus-Barini et al. (2013) used this assay to verify the hair and eye colour of a World War I general, Władysław Sikorski (Chaitanya et al. 2017; Draus-Barini et al. 2013). Recently, the development of the HIrisPlex assay resulted in the introduction of the HIrisPlex-S assay which included 17 skin colour predictive SNPs (Chaitanya et al. 2018).

Similarly, the advancement in scientific technology has allowed for the development of a comprehensive DNA phenotyping system known as Parabon® SNaPshot® DNA Analysis. This system is utilised by the United States of America (USA) and accurately predicts hair, skin and eye colour; genetic ancestry; freckling and face shape (https://snapshot.parabon-nanolabs.com/#phenotyping; accessed 5th March 2019). Parabon® SNaPshot® has been involved in more than 80 published law enforcement investigations. Some featured cases include the murders of Shaquana Caldwell and Rhonda Blankinship, where molecular phenotyping assisted in the identification of severely decomposed human skeletal remains in the first case as well as provided an investigative lead for a new suspect in the latter case (https://snapshot.parabon-nanolabs.com/snapshot-case-summary--anne-arundel-county-md--shaquana-caldwell-murder.html; accessed 5th March 2019).

1.2.2.2 Biogeographical ancestry (BGA) inference

The controversy surrounding the use of ethnic categories in biomedical research and forensic casework remains heightened, particularly in South Africa (Caulfield et al. 2009; Gannett 2014). The term BGA was coined and is perhaps a more appropriate concept in forensic casework and research. Ancestry can be defined as the genetic inheritance carried by ancestors over generations from population groups that have inhabited similar provenance

(Phillips 2015). BGA inference centres around the analysis of markers with distinctive allele frequencies amongst populations, that can indicate an individual's ancestral origin from a specific geographical region (Phillips 2015). Many forensic scientists emphasise that BGA pertains to the ancestral geographical origin rather than the EVC's of individuals and is dissimilar to race (Kayser and Schneider 2009). However, there is a general misconception amongst the public and non-specialists whereby they translate BGA into a social language of identity or race. BGA was introduced to rather strengthen eyewitness reports by minimising bias. BGA inference has a broad range of applications in a forensic setting which have been carefully reviewed by Phillips (2015).

The primary objective for BGA inference is to assist in generating a 'composite profile' of an individual for investigative purposes and to identify unknown samples in forensic cases (Budowle and Van Daal 2008; Gannett 2014). Although forensic STR markers have been utilised to predict BGA (D'Amato and Kasu 2017), they are less suited for this analysis due to its high degree of allele-sharing among various population groups (Budowle and Van Daal 2008). In contrast, certain SNPs are conserved in populations groups due to its low mutation rate, low heterozygosity and high fixation index allowing for certain variants to be fixed in a population (Budowle and Van Daal 2008; Phillips 2015). SNPs that are conserved in different populations and differ between populations provide a platform to potentially distinguish ancestral origins (Bamshad et al. 2004; Romanini et al. 2015).

SNPs that have the ability to infer BGA are known as ancestry informative markers (AIMs) (Shriver and Kittles 2004). Based on the AIMs' ability to exhibit allele-frequency variations among populations, the likelihood of alleles representing a specific biogeographical population can be determined (Giardina et al. 2008). Owing to this, it can be deduced as to whether DNA is consistently associated with a specific ancestral population group. To date, a vast number of AIM-sets have been generated, through which specific panels have been identified as forensically relevant (Phillips et al. 2007, 2013; Santos et al. 2016; Daca-Roszak et al. 2016; Puente et al. 2016).

The use of AIM panels to infer BGA has been successfully applied in forensic cases. The DNAWitness kit, containing a 73-AIM panel, played a significant role in assisting the Louisiana Homicide Task Force to apprehend a Louisiana serial killer, Derrick Todd Lee (Budowle and Van Daal 2008). The BGA inference (85% African and 15% Native American) along with the EVC prediction contributed towards positively identifying the suspect. Sun et

al. (2017) utilised a 27-AIM panel to infer BGA in order to resolve a forensic case involving skeletal remains located in south-west China. It was theorised that the skeletal remains located within the aircraft debris were supposedly linked to a foreign aviator who crashed in World War II. The results showed that the major ancestral component was European (97.6%) thus, confirming their theory (Sun et al. 2017). Lastly, the SNPforID 34-plex assay was successfully applied for ancestral analysis in the 2004 11-M Madrid train bombings. Phillips et al. (2009) utilised this assay to assign BGA from unmatched DNA case samples retrieved from possible locations in Madrid that were linked to the suspects (Phillips et al. 2009). The BGA for the majority of the samples were of North African and European origin. The BGA prediction of the suspect's ancestral origin assisted law enforcement in the criminal investigation (Phillips et al. 2009).

The discovery of SNPs as potential AIMs in the forensic field needs to be consistently evaluated so as to enhance the prediction models and improve their significance in different populations (Kidd et al. 2006). The analysis of AIMs has been extensively studied in the Western population (Sulem et al. 2007; Pakstis et al. 2008, 2010); however, the use of these AIMs in admixed populations has not always been informative, prompting the need for further research (Freire-Aradas et al. 2014; Halder et al. 2008; Kosoy et al. 2009; Tian et al. 2006; Stokowski et al. 2007; Valle-Silva et al. 2019).

AIMs that have been established through SNP association with EVCs such as skin colour have demonstrated significant discriminatory power in some populations (Han et al. 2008; Lao et al. 2007); however, their practicality within admixed populations remains unclear. Walsh et al. (2017) identified 36 highly informative AIMs for skin pigmentation prediction on a global scale (Walsh et al. 2017). Of the 36 AIMs identified, Cerqueira et al. (2014) found that only two AIMs exhibited a potential association with skin pigmentation for admixed populations in Brazil (Cerqueira et al. 2014). The two AIMs mentioned by Cerqueira et al. (2014) have also been discovered to exhibit significant association with skin pigmentation in admixed Latin American (Adhikari et al. 2019) and African-European populations (Beleza et al. 2013; Hernandez-Pacheco et al. 2017; Llyod-Jones et al. 2017). These two AIMs play a role in skin pigmentation prediction and occur within two solute carrier family genes, namely *SLC45A2* and *SLC24A5* (Maroñas et al. 2015). The characteristics of these AIMs are listed below in Table 1.1.

Table 1.1 The characteristics of the two AIMs that appear to play a role in skin colour prediction in admixed populations. *The data was extrapolated from the Cerqueira et al. (2014) supplementary data.*

Gene	SNP ID	Variation	Probable allele function
SLC45A2	rs16891982	NM_016180.4 c.1122C > G ; p.Phe374Leu	C = darker skin tone G = lighter skin tone
SLC24A5	rs1426654	NM_205850.3 c.331G > A ; p.Ala111Thr	G = increase in melanin A = decrease in melanin

The sequence variation c.1122C > G in *SLC45A2* (rs16891982) plays a role in skin pigmentation and is associated with light skin European descent (Soejima and Koda 2007). This AIM is associated with the membrane-associated transporter protein (MATP) coding region of *SLC45A2* which encodes a transporter protein that regulates melanin synthesis (Cook et al. 2009; Tsetskhladze et al. 2012). The second variant, c.331G > A in *SLC24A5* (rs1426654), also influences skin pigmentation and can infer BGA for European, African or Asian individuals (Crawford et al. 2017; Lamason et al. 2005). It was suggested that the functional role of this AIM affects human melanogenesis through cation-exchange activity (Lamason et al. 2005; Ginger et al. 2008; Sturm 2006).

1.2.3 Skin pigmentation

The high degree of variation in skin pigmentation between individuals of different genetic ancestries renders it as one of the most distinguishable human traits (Sturm 2009). The complex biological pathways and environmental adaptations are the underlying factors that contribute to skin pigment variation (Lin and Fisher 2007). More than 125 genes have been identified to associate with human pigment phenotype; however, only a dozen of them have been successfully characterised (Tully 2007).

The human pigmentary pathway is reliant on the production of a biopolymer known as melanin (Sturm, Teasdale, and Box 2001; Rees 2003). The production of melanin is characterised by a complex biochemical process known as melanogenesis (Branicki 2009). Melanin is produced by melanocytes which are specialised neural crest cells located at the base of the epidermal skin layer (Rees 2003). Melanin can be classified into two principal classes: eumelanin (black or brown pigment) and pheomelanin (yellow or red pigment) (Rees 2003). The various pigmentary phenotypes are defined based on the arrangement of melanocytes and the deposition of melanin. The darkly pigmented skin phenotype comprises

individuals containing the arrangement of large, dispersed melanocytes with a high melanin concentration whereas the opposite is observed in the lightly pigmented skin phenotype (Maroñas et al. 2015; Sturm, Box, and Ramsay 1998; Sturm, Teasdale, and Box 2001).

1.2.3.1 Skin pigmentation measurement

The range of human skin pigmentation does not consist of discrete skin colours and should rather be viewed as a continuous spectrum which differs between diverse population groups (Tully 2007). Several skin type classification systems have been developed of which two, the Fitzpatrick scale and the von Luschan scale, are recognised as the benchmark for skin colour evaluation in dermatology (Roberts 2009). The Fitzpatrick scale is a numerically derived skin classification system that denotes six different skin pigment phenotypes ranging from extremely fair (skin type I) to extremely dark (skin type VI) based on an individual's exposure to the sun (Table 1.2) (Fitzpatrick 1988; Sachdeva 2009). Alternatively, the von Luschan scale classification system is characterised by a chromatic scale colour chart which categorises skin pigment hues against numerical categories (Table 1.2) (Swiatoniowski et al. 2013). This colour chart describes 36 categories which are utilised to compare against an individual's skin colour to establish racial classification (Table 1.2) (Treesirichod, Chansakulporn, and Wattanapan 2014).

Table 1.2 The two skin colour evaluation tools and their associated pigment phenotypes *(Fitzpatrick 1988; Swiatoniowski et al. 2013)*

Fitzpatrick scale	**von Luschan scale**	**Skin pigment phenotype**
Skin type:	**Score:**	
I	0-6	Pale white
II	7-13	White
III	14-20	Light brown
IV	21-27	Moderate brown
V	28-34	Dark brown
VI	35-36	Extremely dark brown to black

Although the aforementioned skin classification systems are well-known on a global scale, the principle behind their skin pigment interpretation is highly subjective which poses a concern within forensics research (Van der Wal et al. 2013). Thus, an objective instrument, such as that obtained from a derma-spectrophotometer, should rather be utilised to record pigmentation in melanin index (MI) measurements. The MI measurements are attainable because the human skin's light absorptive characteristic allows for the derma-spectrophotometer to measure specific colours based on reflectance spectrophotometry (Van der Wal et al. 2013). Several spectrophotometers have been developed; however, based on a

validation experiment between them, the DSM II ColorMeter (CyberDerm Inc., USA) was found to be highly sensitive and accurate (Van der Wal et al. 2013). The development of derma-spectrophotometer for skin colour evaluation has shifted the approach away from the subjective numerical and racial classification towards an objective and representative MI measurement.

1.2.4 Molecular phenotyping in SA

Some traditional genetic markers used for skin pigmentation evaluation in European and North American population groups have previously displayed poor association in admixed populations (Cerqueira et al. 2014). The selection of genetic markers for phenotype prediction in a local context needs to be considered based on population differences, especially in the SA context. SA consists of diverse ancestral groups with a continuum of skin colour and MI (Slabbert and Heathfield 2018). Thus, it is vital to evaluate the discriminatory power of AIMs within the SA context and whether these AIMs can be associated with MI.

1.2.4.1 SA population demographics

The SA population is rich in diversity because it is characterised by groups originating from African, Asian and European descent (de Wit et al. 2010). The complex amalgamation of these continental population groups has contributed to SA's diversity through the establishment of an admixed population group known as the South African Coloureds (SAC) (Adhikari 2005). SA utilises four racial categories for bureaucratic identification – African Black, Coloured, Indian/Asian and White (Brown 2000). The 2018 mid-year population estimates stated that the SA population is estimated to be 57.7 million and comprises 80.9% African Black, 8.8% Coloured (SAC), 2.5% Indian/Asian and 7.8% White (Statistics SA 2018). The SAC comprises the second largest population group in SA with the majority residing in the Western Cape (approximately 50%) (Statistics SA 2018). A study of the genomic structure of the SAC population in the Western Cape revealed the major ancestral components that contribute this population to be Khoesan, African (Bantu-speaking), European and Asian populations (de Wit et al. 2010).

1.2.5 Ethical and social considerations of molecular phenotyping

The use of molecular markers for EVC prediction poses major ethical and social concerns. The ethical considerations need to be reviewed amongst the forensic community and

bioethics research organisations (Kayser and Schneider 2009). By raising these ethical concerns, the conflicting viewpoints for utilising molecular phenotyping will be easily understood to benefit the public and legislative bodies. Thus, it is important that these issues are scientifically debated to provide a balanced view in which the advantages of using these molecular markers are compared against the associated risks to ethical values (Kayser and Schneider 2009).

The primary ethical implication that has been highly debated is the violation of privacy. Privacy issues have been at the forefront of ethical interrogation of forensic genetics since its establishment (Williams and Wienroth 2017). EVC prediction based on DNA may infringe on the confidentiality of an individual, particularly if it discloses a vulnerable phenotype such as a propensity for a genetic disease or an altered physical trait (Slabbert and Heathfield 2018). In contrast, Toom et al. (2016) argue that eyewitness reports are relevant as they provide a valuable context of a crime as opposed to a 'biological witness' such as molecular phenotyping (Toom et al. 2016). In SA, the right to privacy can be limited in certain instances, according to the Bill of Rights in the Constitution of the Republic of South Africa (*Constitution for the Republic of South Africa, No. 108 of 1996* 1996: chap2). The limitations; however, must be justified and reasonable based on human dignity, freedom and equality.

Apart from the ethical concerns, molecular phenotyping within a forensics context requires the inclusion of social implications related to its use. BGA inference and its association with racial classifications and ethnicity poses a concern for potential prejudicial profiling, thus causing an imbalance in the treatment of certain social groups who have become the centre of investigative attention (Williams and Wienroth 2017). As a result, minority groups will be susceptible to suspicion which leads to further discrimination. The risk to these vulnerable social groups for investigative attention will be heightened which will further exacerbate the existing social justice disadvantages that exist within these groups (Wienroth, Morling, and Williams 2014).

The ability for molecular phenotyping to predict EVCs and estimate BGA can potentially assist in alleviating the fear of prejudicial profiling and relieve social tensions within communities (Kayser and Schneider 2009; Ossorio 2006). M'Charek (2008) reported that molecular phenotyping allayed rather than incited ethnic bias. The case involving the murder of a Dutch girl led the town residents to believe that the suspect was from the minority

Middle Eastern ethnic group; however, BGA testing refuted that suspicion, stating the suspect was of European descent (M'Charek 2008).

Molecular phenotyping also carries social value by increasing the extent to which DNA is used in forensic investigations (Kayser 2015). EVC prediction can assist in narrowing down the suspect pool by means of positive exclusion of individuals (Slabbert and Heathfield 2018). It can also assist in assigning an identity to unknown victims and unidentified human remains in addition to providing social justice and closure for families.

1.2.6 Legal implications of molecular phenotyping

The novel techniques and applications for DNA evidence along with the empirical research that supports its value have significant implications for the judicial system and the legislators (Smith and Mann 2015). In SA, no specific DNA legislation existed until 2014 when the Criminal Law (Forensic Procedures) Amendment Act 37 of 2013 (DNA Act) was declared operational. The introduction of the DNA Act addressed the absence of a legal framework around using forensic DNA profiling as an intelligence tool. The key provision of the DNA Act was the establishment and regulation of the National Forensic DNA Database of South Africa (NFDD). Recently, the implementation of the NFDD has potentially advanced criminal intelligence in SA; however, the current repository of DNA profiles is not sufficiently populated therefore limiting its use as a criminal investigative tool (Heathfield 2014).

The DNA Act prohibits the use of molecular phenotyping for forensic casework. The definition of forensic DNA profile and forensic DNA analysis in the DNA Act explicitly states that analysis of DNA and the subsequent results obtained must not encompass any information relating to an individual's medical or mental condition or the propensity or physical information of an individual other than their sex (*Criminal Law (Forensic Procedures) Amendment Act, No. 37 of 2013* 2014). Based on these definitions, the use of genetic material in forensics is lawfully regulated thus, limiting its use to only generate forensic DNA profiles.

Koops and Schellekens (2008) reported that many countries, including SA, have legislation for the use of DNA; however, it is mainly regulated for conventional DNA profile analysis, with no inclusion of molecular phenotyping (Koops and Schellekens 2008). This close-minded concept highlights the negative aspect of this act to expand DNA analysis development for forensic investigations in SA (Heathfield 2014). It is evident that the SA

legal system will need to revise its policies to accommodate novel scientific developments relevant to the forensics field.

Although molecular phenotyping is outlawed in SA, several countries permit its use for forensic purposes regardless of whether it is mentioned in their legislation or not. In the Netherlands, the government has amended the Dutch Code of Criminal Procedure in 2003 to allow for phenotypic information to be derived from DNA found at a crime scene. The Dutch legislation; however, limits these traits to only aid in criminal investigations and to predict traits that are known to an individual and were visibly present at birth (Koops and Schellekens 2008). Furthermore, in Germany, the state of Bavaria has recently amended its regulations to include molecular phenotyping as an investigative tool. In other German states, there are currently ongoing policy discussions relating to the German legislation and its regulation of molecular phenotyping (Samuel and Prainsack 2018). In contrast, the United Kingdom has no legislation that specifically regulates this technique; however, it has been applied in forensic casework on numerous occasions (Koops and Schellekens 2008; Smith and Mann 2015). Similarly, in the United States, there is no federal legislation on molecular phenotyping; however, most states laws regulate its use. Texas is the only state that has legally sanctioned its application in criminal investigations (Koops and Schellekens 2008).

The recent development of the NFDD limits the use of DNA as a criminal investigative tool. It has been recommended that alternative forensic DNA analysis, such as molecular phenotyping, should be considered in forensic casework in SA (Slabbert and Heathfield 2018), particularly given the burden of unidentified remains in forensic mortuaries (Reid, Martin, and Heathfield 2019). Extensive validation in conjunction with amending SA legislation is needed in order to include and regulate molecular phenotyping as a tool in forensic practice in SA (Slabbert and Heathfield 2018).

1.3 Rationale

Although the recent establishment of the NFDD has shown potential to advance criminal intelligence in SA, its infancy has raised concerns regarding providing investigative leads for law enforcement. Alternatively, the use of molecular phenotyping has the ability to aid in forensic cases by predicting EVCs and BGA from evidential biological samples allowing this technique to develop into a useful identification tool to aid in certain investigations.

Traditional molecular phenotyping markers have mainly been developed and researched in European and North American population groups and have been incorporated into commercials kits. The significance and informative value of these markers within the SA context is unknown. Therefore, the relevance of these markers needs to be assessed in SA.

As previously stated, skin colour forms part of the EVCs which plays an important role in victim identification for early-phase investigation. Many conventional markers used for skin pigmentation evaluation; however, were insignificantly associated with admixed populations. It is evident from the above literature review that previous studies have provided the framework to objectively research appropriate SNPs for skin colour assessment in admixed populations; however, the subsequent data is limited particularly in a SA context. This, in conjunction with the limited numbers of forensic studies utilising MI globally, has consequently prompted the need for a study to be performed in SA. The two AIMs selected for analysis were *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654). These AIMs were identified by Cerqueira et al. (2014) to be constantly associated with only MI in the Brazilian admixed populations, which prompted the selection of these AIMs for our analysis. The initial results from a previous pilot study in our research group were promising (Slabbert 2017); however, the sample size was underpowered. Thus, further research is needed to evaluate molecular phenotyping markers as a forensic tool to predict MI in a significantly large cohort within SA.

1.4 Aim and objectives

1.4.1 Aim

The aim of this study is to investigate the relationship between SNPs involved in molecular phenotyping and MI in SA individuals.

1.4.2 Objectives

- Continue recruitment of participants to reach a significantly powered cohort.
- Genotype the two AIMs (*SLC45A2*, rs16891982 and *SLC24A5*, rs1426654) using the SNaPshot® genotype assay for the cohort.
- Statistically assess and model the relationship between the MI of the participants and their genotypes for the two markers.

Chapter 2: Methods and materials

2.1 Study design

A prospective, cross-sectional, experimental study design with a quantitative research paradigm was used to address the aim of this project. The quantitative analysis involved measuring the association between the genetic variations (independent variables) represented by the genotypes of two AIMs and the recorded MI (dependent variable) of an individual.

2.2 Cohort

Based on the sample size calculation (Appendix A), the number of participants required to obtain statistical power was 354 individuals. Prior to this minor dissertation, a total of 289 individuals had previously been recruited with sufficient DNA, of which 92 individuals were successfully genotyped for these two variants. Therefore, a further 65 individuals were required to participate in this study; however, 100 individuals were recruited in case insufficient DNA was obtained for downstream analysis.

Individuals of all ethnicities who were over 18 years old residing in the Western Cape were included in the study. Individuals with pigment disorders or scarring in the region where MI was measured were excluded. Participants reporting the use of supplements or remedies that alter their skin colour were also excluded.

The recruitment of the participants was performed by oral communication and volunteers gave informed consent to participate (Appendix B). This study received ethics approval from the Human Research Ethics Committee (HREC) of the University of Cape Town (UCT HREC 317/2015) (Appendix C).

2.3 Data collection

2.3.1 Participant questionnaire

The participants were required to complete an information questionnaire self-reporting their ethnic background (Appendix D). Each questionnaire was linked to a unique participant number in order to maintain anonymity. For the ancestral origin, the mixed ancestry group comprised of participants with more than one ancestral origin.

2.3.2 MI measurement

The MI of the participants was measured at three hairless sites (inner forearm, the inner arm above the elbow and forehead) in duplicate using a non-invasive derma-spectrophotometer (DSM II ColorMeter, CyberDerm Inc., USA). The measurements were recorded on the participant questionnaire.

2.4 DNA analysis

2.4.1 Sample collection and DNA extraction

The participants vigorously rinsed their mouth with 10 mL of 0.9 % saline solution for 30 seconds. The saline was collected in a 50 mL tube and the cells were pelleted at 3 000 rpm for 5 minutes in an Eppendorf® Centrifuge 5804 (Eppendorf, Hamburg, DEU). The supernatant was discarded without disturbing the cell pellet.

DNA extraction was performed on the pellet according to the manufacturer's protocol from the Quick-DNA™ Miniprep Plus Kit (Zymo Research, Irvine, USA). The protocol was amended to include the addition of 150 µL of BioFluid and Cell Buffer (Red), 150 µL of DNA Elution Buffer and 15 µL of Proteinase K to the pelleted cells. Incubation steps were done using a ThermoMixer F2.0 (Eppendorf, Hamburg, DEU) while centrifugation was done in the Eppendorf Centrifuge 5417C (Eppendorf, Hamburg, DEU). The DNA was eluted into 50 µL of DNA Elution Buffer (Zymo Research, Irvine, USA).

2.4.2. DNA quantification

Following DNA extraction, the samples were quantified by spectrophotometry. Quantification analysis was performed using the NanoDrop™ 2000 Spectrophotometer (NanoDrop Technologies LLC, Wilmington, USA). The instrument was blanked using DNA Elution Buffer (Zymo Research, Irvine, USA). A volume of 2 µL of DNA was added onto the instrument for quantification.

The readings obtained from the quantification analysis included the total genomic DNA concentration of the sample and the A_{260}/A_{280} and $A_{260}/_{230}$ absorbance ratios. The quantification was performed in duplicate to obtain a mean DNA concentration which was used to dilute the DNA samples to a final concentration of 50 ng/µL using AccuGene® Molecular Biology Grade Water (Lonza, Rockland, USA).

2.5 Amplification of target regions

2.5.1 PCR optimisation and amplification

PCR optimisation was performed for both primers sets (Table 2.1) in order to identify the ideal annealing temperature (T_a). Optimisation was carried out again in this study since the pilot study made use of a different DNA polymerase, which has since been discontinued in SA. To this end, temperature gradient PCRs was performed in a T100 Thermal Cycler (Bio-Rad, Hercules, USA). Each standard PCR reaction contained 1X GoTaq® G2 green master mix (Promega, Madison, USA), 10 µM of each primer (Table 2.1) and 50 ng of genomic DNA. AccuGene® Molecular Biology Grade Water (Lonza, Rockland, USA) was added to make up a final reaction volume of 25 µL.

PCR cycling conditions for T_a optimisation were as follows: initial denaturation at 95 °C for 5 minutes; followed by three steps, cycled 30 times, consisting of: (1) denaturation at 95 °C for 30 seconds, (2) annealing at a temperature gradient (ranging from 50 °C to 60 °C) for 30 seconds, and (3) extension at 72 °C for 30 seconds. A final extension step was included at 72 °C for 7 minutes.

Table 2.1 The primers (forward and reverse) designed for each gene to amplify the regions of interest

Gene	**Fragment size (bp)**	**Forward Sequence (5'→3')**	**Reverse Sequence (5'→3')**	**Reference adapted from**
SLC45A2 (rs16891982)	201	TCCAAGTTGTG CTAGACCAGA	CGAAAGAGGAGTCG AGGTTG	(Giardina et al. 2008; Valenzuela et al. 2010; Allwood 2013; Walsh et al. 2012)
SLC24A5 (rs1426654)	199	CCCTTGGATTG TCTCAGGATG	TGAGTAAGCAAGAA GTATAAGGAGCAA	

Once the optimal T_a for both primer sets was identified as 56.1 °C, PCR amplification was performed as described above on the cohort samples (n = 197 from previously recruited cohort + 100 newly recruited individuals). A non-template control (NTC) was included for each PCR reaction.

2.5.2 Agarose gel electrophoresis

Gel electrophoresis was utilised to confirm target region amplification. Gels containing 2 % (w/v) agarose (SeaKem LE Agarose, Walkersville, USA) prepared with 1X TBE buffer (Sigma Aldrich, St. Louis, USA) were used, and subsequent visualisation of PCR products was facilitated by adding 0.01 % (v/v) of SYBR® Safe DNA Gel Stain (ThermoFisher

Scientific, Waltham, USA). A volume of 5 µL of PCR product and 5 µL of molecular weight marker (MWM) (Quick-load® Purple 50 bp DNA ladder, New England BioLabs, Ipswich, USA) was loaded. The MWM was included on each gel for estimation of amplicon size. Electrophoresis was performed for 1 hour at 80 volts (V) and was visualised using the Syngene Bio-imaging System (Syngene, Cambridge, UK). Images were captured using GeneSnap v.7.12 software (Syngene, Cambridge, UK).

2.6 SNaPshot® genotyping

2.6.1 Post-PCR clean-up

The PCR products that were successfully amplified underwent genotyping using a multiplex SNaPshot® PCR method. Post-PCR purification was performed prior to genotyping. The PCR products were purified according to the ExoSAP-IT™ Express PCR Product Clean-up protocol (Thermo Fisher, Waltham, USA). The protocol was amended to include multiplexing the post-PCR reaction product in a 4:1 (*SLC45A2:SLC24A5*) mixture ratio after initial optimisation with various post-PCR product mixture ratios (1:1; 3:2 and 4:1).

2.6.2 SNaPshot® PCR

The SNaPshot® PCR was optimised to also identify the ideal primer mixture ratio. The SNaPshot® PCR was performed using the SNaPshot® Multiplex Kit (Applied Biosystems, Foster City, USA). Each SNaPshot® PCR reaction contained 5 µL of the multiplexed purified post-PCR product, 1 µL of SNaPshot® Ready Reaction mix, 1 µL of primer mixture (Table 2.2) and 3 µL of AccuGene® Molecular Biology Grade Water (Lonza, Rockland, USA). Positive and negative controls were included for each SNaPshot® PCR reaction. The optimisation involved various *SLC45A2:SLC24A5* primer mixture ratios (1:1; 3:2; 4:1 and 6:1). SNaPshot® PCR reactions were incubated in a T100 Thermal Cycler (Bio-Rad, Hercules, USA).

The PCR cycling conditions for the primer mixture ratio optimisation were the following: Three steps, cycled 25 times, consisting of: (1) 96 ºC for 10 seconds, (2) 50 ºC for 5 seconds and (3) 60 ºC for 30 seconds. A final infinite hold step was included at 4 ºC.

Table 2.2 Internal forward primer designed for each gene used for SNaPshot® genotyping

Gene	Primer size (bp)	Expected nucleotide change	Colour change (dye)	Forward Sequence (5'→3')
SLC45A2 (rs16891982)	34	c.1122C > G	Black > Blue (dTAMRA™ > dR110)	TTAATTATGTTATAT CTTACACGGAGTTGATGCA
SLC24A5 (rs1426654)	18	c.331G > A	Blue > Green (dR110 > dR6G)	TCTCAGGATGTTGCAGGC

Once the optimal primer mixture ratio was identified as 4:1 (*SLC45A2:SLC24A5*), SNaPshot® PCR was performed, as described above, on the cohort samples.

2.6.3 SNaPshot® PCR clean-up

A volume of 1 µL of recombinant Shrimp Alkaline Phosphatase (rSAP) (New England BioLabs, Ipswich, USA) was added to the SNaPshot® PCR reaction. The SNaPshot® PCR clean-up was performed in a T100 Thermal Cycler (Bio-Rad, Hercules, USA) with the following conditions: 37 ºC for 1 hour followed by 75 ºC for 15 minutes.

2.6.4 Capillary electrophoresis

In a final reaction volume of 10 µL, 7.9 µL Hi-Di formamide (Thermo Fisher Scientific, Waltham, USA), 0.1 µL GeneScan™ 120 LIZ™ size standard (Thermo Fisher Scientific, Waltham, USA) and 2 µL of cleaned SNaPshot® PCR product was added to a 96-well plate. The samples were denatured using GeneAmp™ PCR System 9700 (ThermoFisher Scientific, Waltham, USA) for 5 minutes followed by storage at -20 ºC on an ice block for 2 minutes.

Capillary electrophoresis was performed in the ABI 3130xl Genetic Analyser (Applied Biosystems, Foster City, USA). The raw data files were analysed using GeneMapper Software v4.1 (Thermo Fisher Scientific, Waltham, USA). The genotyping analysis for both markers utilised the International Union of Pure and Applied Chemistry (IUPAC) nucleotide base codes.

2.6.5 Allele and genotype frequencies

The allele and genotype frequencies were obtained from the 1000Genome Project which is a database containing frequencies from different population groups around the world. The 1000Genome Project (The 1000 Genomes Project Consortium 2015) consisted of samples from 26 different populations grouped into five metapopulations: Africa, East Asia, Europe,

South Asia and America. This genotype data is freely accessible globally rendering it a great asset to research.

2.7 Data analysis

A Shapiro-Wilk test was performed to determine whether the MI measurements followed a normal distribution. The Kruskal-Wallis test was performed to determine whether a significant difference existed between the mean MI observed and the SA population census groups as well as the different ancestral origins. Thereafter, a Wilcoxon sum-rank test was performed to determine the significance of MI within the SA population census groups as well as ancestral groups. The post-hoc test in the form of the Bonferroni correction was accounted for. A 95 % confidence interval was selected which meant p - values less than 0.05 were considered significantly different. Furthermore, a Pearson's Chi-Squared test was performed to evaluate whether the genotype data obtained was in Hardy-Weinberg Equilibrium (HWE). The aforementioned analyses were performed using RStudio v1.2.1335 (RStudio Inc., Boston, USA).

A generalised linear model (GLM) for the mean MI was designed based on the observed genotypes of the individuals (Lloyd-Jones et al. 2017). The model considered all observed alleles and genotypes of each SNP independently in addition to the combined genotype observations. A backward hierarchical approach was selected to generate the best predictive mean MI model based on the Akaike Information Criterion (AIC) value (Bozdogan 1987). The combination of covariates that produced the lowest AIC value was chosen for the model equation. The model was created using a significant level of 0.1 for removal of co-variates from the model. The distribution of the dependent variable (family) used was the Poisson distribution and the link function (link) used was the power function. The MI predictive model was generated using Stata/SE v.13.1 (StataCorp LLC, College Station, USA) and was evaluated on its MI prediction accuracy for each genotype combination using a 95 % confidence interval compared to the recorded mean MI.

Chapter 3: Results

3.1 Cohort

A total of 389 individuals were recruited to participate in this study. Of the 389 individuals, 167 individuals (43 %) were male and 222 (57 %) individuals were female (Figure 3.1A). Overall, the most frequent self-reported population group within the cohort comprised of African Black (37 %; n = 143) and White (29 %; n = 114) participants, followed by the Coloured (SAC) (19 %; n = 75) and Indian/Asian (15 %; n = 57) participants.

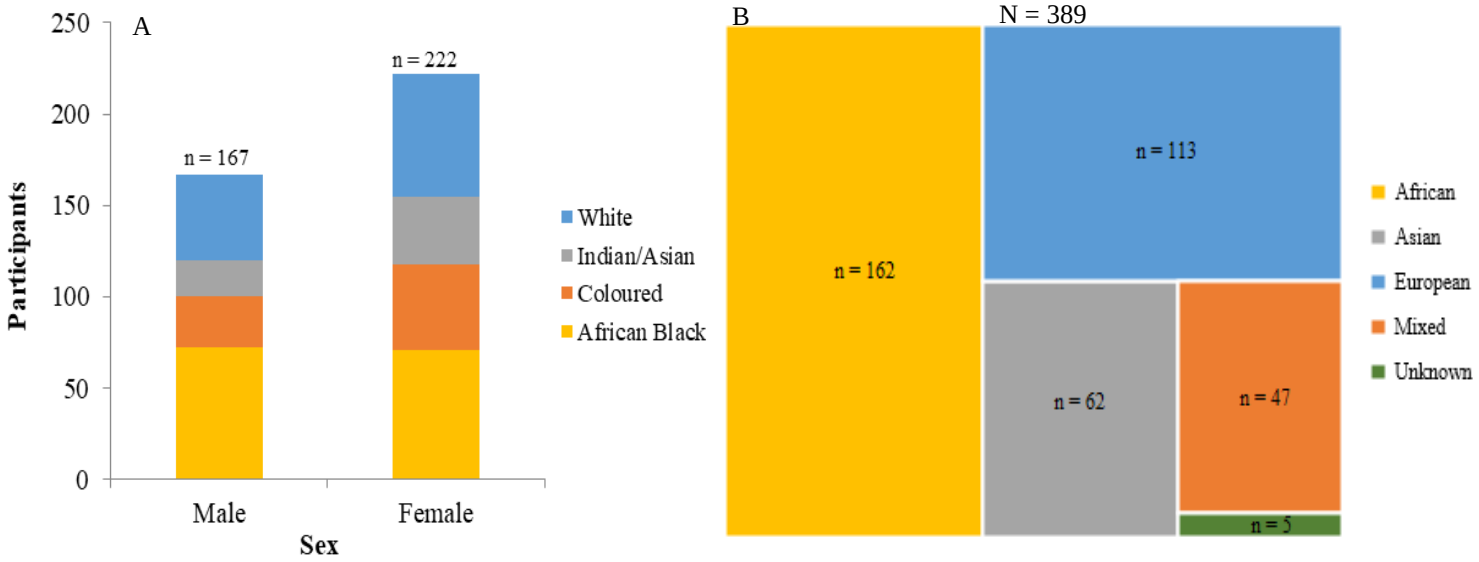

Figure 3.1 (A) Stacked column graph representing the sex of the participants and their self-reported population group according to SA population census categories in the cohort. *The self-reported population group for the male participants consisted of 43 % African Black, 17 % Coloured, 12 % Indian/Asian and 28 % White. For the female participants, the self-reported population group consisted of 32 % African Black, 21 % Coloured, 17 % Indian/Asian and 30 % White*. (**B)** Treemap depicting the various ancestral origins of the participants in the cohort.

The cohort was also represented by participants from various ancestral origins (Figure 3.1B). The major ancestral origin that characterised the cohort was African (42 %; n = 162) (Figure 3.1B). Other ancestral origins of the participants consisted of 29 % European (n = 113), 16 % Asian (n = 62) and 12 % Mixed (n = 47). Of the 389 individuals in the cohort, only 1 % (n = 5) stated their ancestral origin to be unknown. The comparison between the two variables (Figure 3.1A + B) show that some participants identify their SA population census groups, but it does not correspond to their self-reported ancestral group.

3.2 MI measurements

3.2.1 MI of the cohort

The MI of the newly recruited participants were taken and combined with the existing participants in the cohort (Figure 3.2). The majority of the newly recruited participants' (44 %) MI readings ranged between 30 – 35 units. Similarly, this trend was also observed in the MI readings taken from the existing participants in the cohort. Overall, the most frequent MI range (30 - 35) was present in 38 % (n = 147) of the final cohort.

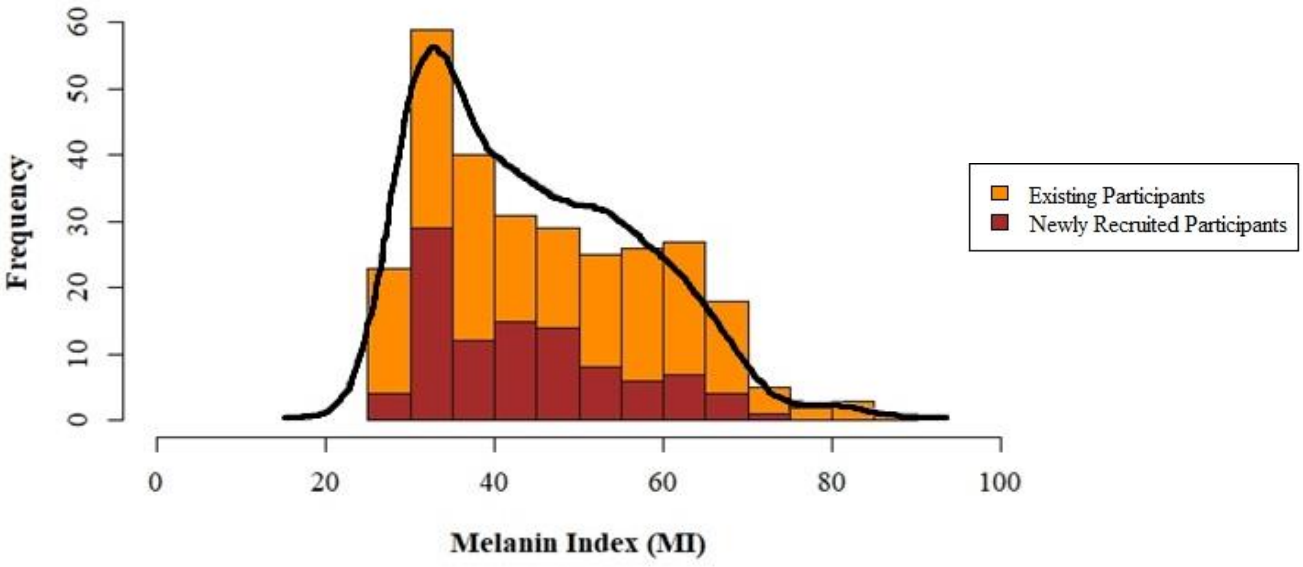

Figure 3.2 Histogram showing the MI frequency of the cohort. *The overlaying line (black) represents the density curve.*

To determine the distribution of the MI readings, a density curve was plotted (overlaying black line). Based on the visual representation of the distribution plot, the density curve suggested that the distribution of the MI readings were positively skewed (right skewed). This observation was confirmed by the Shapiro-Wilk test which indicated that the MI readings were not normally distributed ($p < 0.0001$) and that the MI distribution was asymmetrical. The mean MI of the cohort (45.29) was greater than the median (43.16) and mode (33.29) which further showed the positively skewed distribution of the MI readings.

3.2.2 MI versus self-identified SA population census groups

The MI distribution for the four different SA population census groups is depicted by figure 3.3. The median and range of MI for the following SA population census groups are described in Appendix E (Table E.1). The small range in MI for the White population group

(14.42) indicated less dispersion in the MI readings while a large range for the African Black (49.26), Coloured (30.41) and Indian/Asian (31.23) population groups indicated a greater dispersion. The MI readings observed in each SA population census group overlapped each other.

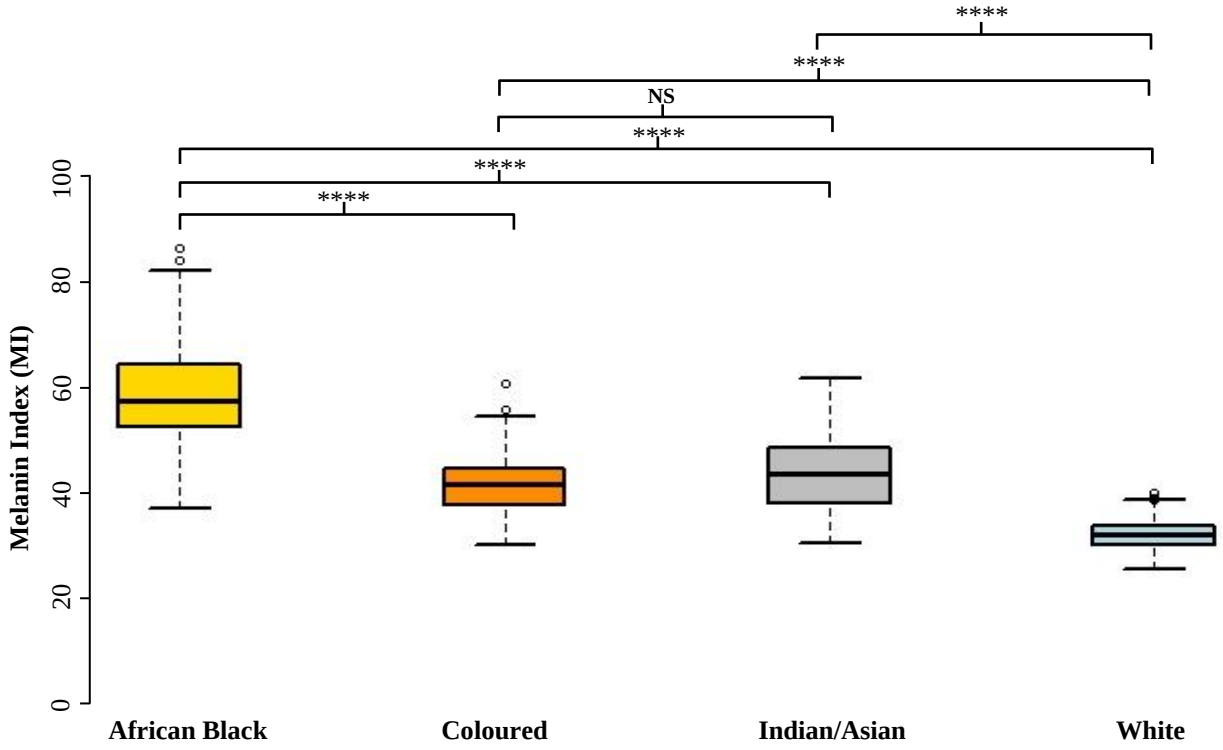

Figure 3.3 Box and whisker plot representing the MI distribution for the self-identified SA population census groups. *The circles represent the outliers. **** p < 0.0001 and NS = not significant (determined by Wilcoxon rank-sum test with Bonferroni correction).*

A Kruskal-Wallis test denoted a significant association of MI across the SA population census groups ($p < 0.0001$) (Appendix F; Table F.1). Therefore, pairwise Wilcoxon rank-sum tests (with a Bonferroni correction) were performed for the MI distribution between each pairwise SA population census group. A statistically significant MI difference was observed between each pairwise population group except between the Coloured and Indian/Asian population groups ($p = 0.25$) (Figure 3.3).

3.2.3 MI versus self-reported ancestral groups

The MI distribution for the self-reported ancestral groups is depicted by figure 3.4. The summary statistics for the MI of the following ancestral groups are described in Appendix E (Table E.2). Similar to the results from the SA population census groups above, there was a small MI range for the European ancestral group (19.14) with larger dispersion ranges for the

African (52.68), Asian (31.23) and Mixed (24.19) ancestral groups (Figure 3.4). The MI readings observed in ancestral groups also overlapped each other.

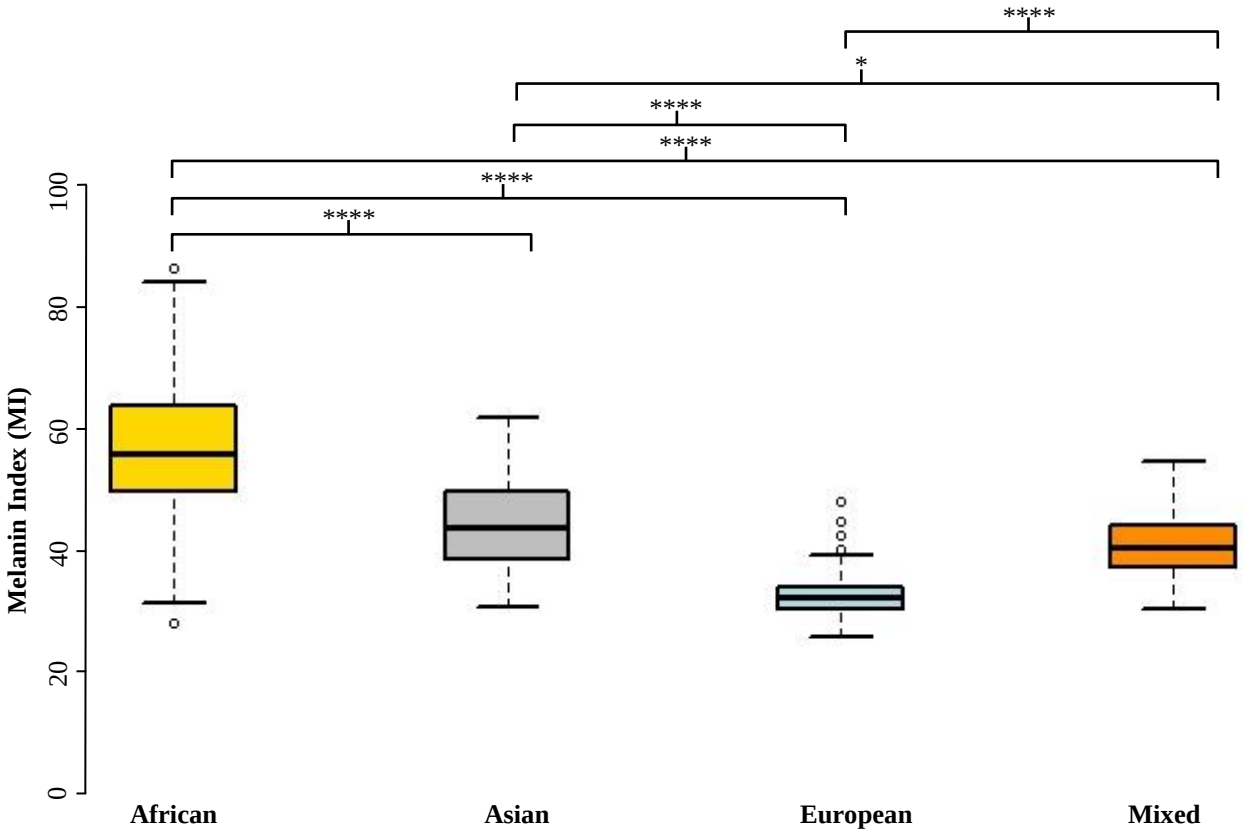

Figure 3.4 Box and whisker plot representing the MI distribution for the self-reported ancestral groups. *The circles represent the outliers. **** $p < 0.0001$ and * $p < 0.05$ (determined by Wilcoxon rank-sum test with Bonferroni correction).*

Again, a Kruskal-Wallis test denoted a significant association of MI across the ancestral groups ($p < 0.0001$) (Appendix F; Table F.2) and Wilcoxon rank-sum tests (with Bonferroni correction) were performed. Each the ancestral group demonstrated a statically significant MI difference between each of the other ancestral groups (Figure 3.4).

3.3 DNA quantification

The quality and quantity of DNA extracted from the newly recruited participants were measured spectrophotometrically. Overall, the mean DNA concentration obtained from the newly recruited participants was 143.78 ng/µL, A_{260}/A_{280} ratio was 1.83 and A_{260}/A_{230} ratio was 1.57 (Table 3.1).

Table 3.1 The five-number summary, mean DNA concentration and mean A_{260}/A_{280} and A_{230}/A_{260} ratios obtained from the newly recruited participants.

Five-number summary (ng/µL)				
Min	**1st Quartile (Q1)**	**Median**	**3rd Quartile (Q3)**	**Max**
25.45	95.99	124.53	168.91	414
Mean DNA concentration (ng/µL)				
143.78				
Mean A_{260}/A_{280}		**Mean A_{260}/A_{230}**		
1.83		1.57		

3.4 PCR optimisation and amplification

Optimal conditions were determined for PCR amplification of the desired fragments by performing temperature gradient PCRs. The temperature gradients for both *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654) primers indicated a range of optimal T_a, where bands were bright, and no non-specific binding was present. For ease, 56.1 °C was chosen for both PCRs (Figure 3.5A and Figure 3.5B) and subsequent PCR amplification at T_a = 56.1 °C for both *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654) resulted in specific amplification of the target fragments: 201 bp (*SLC45A2*, rs16891982) (Figure 3.5C) and 199 bp (*SLC24A5*, rs1426654) (Figure 3.5D) in the newly recruited (n = 100) participants and previously recruited but not yet genotyped (n = 197) participants' DNA samples. The NTC was clear showing no contamination and the MWM (Appendix G) was utilised to identify and confirm the size of the target fragments. The gels (Figure 3.5C – D) are a representation of the successful amplification of both target regions that was achieved from the DNA samples of the remaining cohort (n = 297).

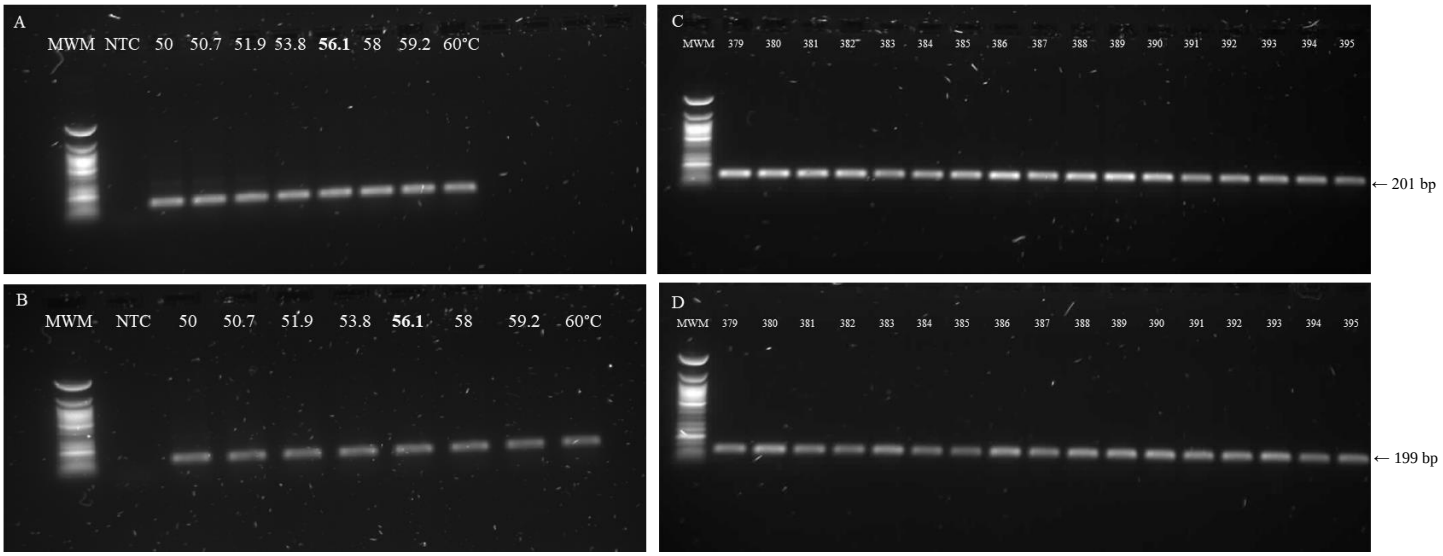

Figure 3.5 Images of PCR products visualised with SYBR® Safe on a 2 % agarose gel with 5 µL of PCR product, NTC and 5 µL of 50 bp MWM loaded. The agarose gel(s) were electrophoresed for 1 hour at 80 V. **(A + B)** A temperature gradient ranging from 50 – 60 °C for identification of the optimal T_a for the **(A)** *SLC45A2* (rs16891982) fragment of 201 bp and **(B)** *SLC24A5* (rs1426654) fragment of 199 bp. **(C + D)** PCR amplification of DNA samples from some of the newly recruited participants in the cohort (MSK 379 - 395) at the optimum T_a for **(C)** *SLC45A2* (rs16891982) and **(D)** *SLC24A5* (rs1426654) which was identified as 56.1 °C.

3.5 SNaPshot® genotyping

SNaPshot® genotyping was performed on all amplified PCR products to identify the presence of both variants (rs16891982 and rs1426654) in their respective genes (*SLC45A2* and *SLC24A5*) (Figure 3.6). In an electropherogram, peak pairs at the same location represent heterozygosity while a single, with an often heightened peak, represents homozygosity.

The electropherograms below represent different genotypes for the two markers for a selection of individuals. As an example, samples MSK 129 and MSK 257 were found to be homozygous for both variants as represented by the single monochromatic peaks (Figure 3.6A + B). However, sample MSK 414 was heterozygous for both variants as denoted by the two dichromatic peaks at each of the SNP loci (Figure 3.6C).

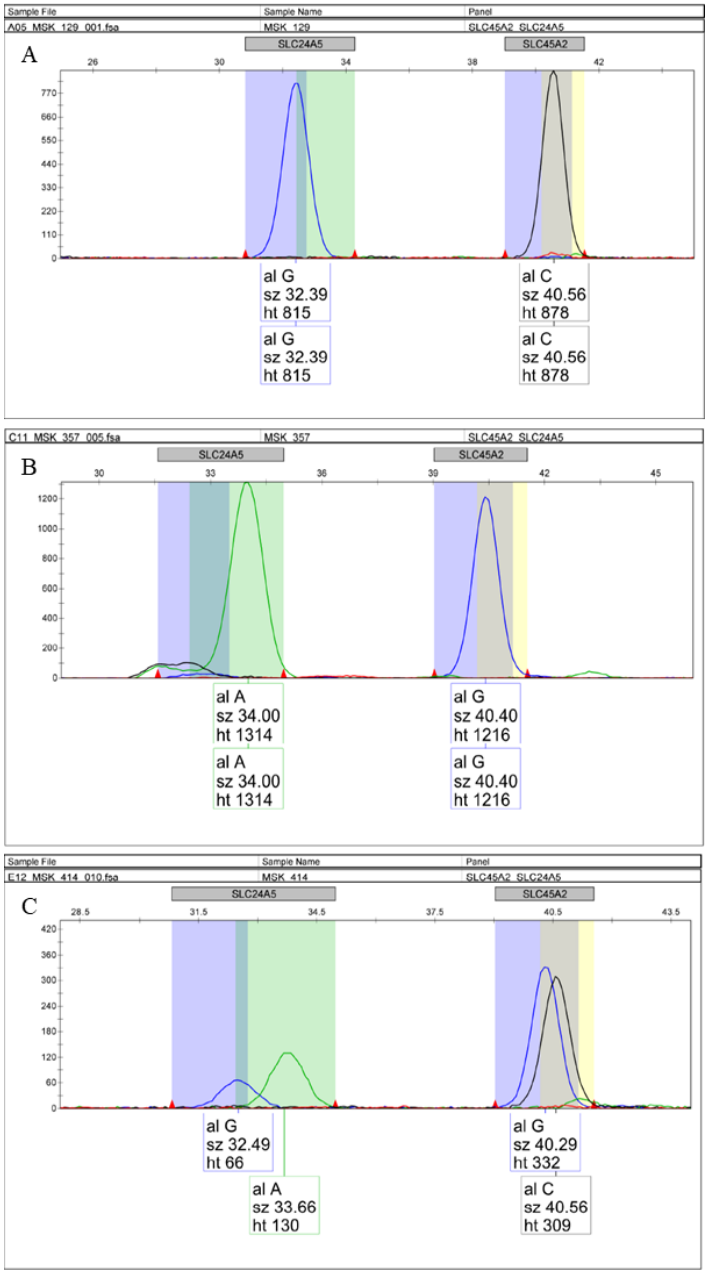

Figure 3.6 SNaPshot® electropherograms showing the presence of the *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654) variants. **(A)** The homozygous variant c.1122C > G (allele C) identified in *SLC45A2* (rs16891982) and the homozygous variant c.331G > A (allele G) identified in *SLC24A5* (rs1426654) for sample MSK 129. **(B)** The homozygous variant c.1122C > G (allele G) identified in *SLC45A2* and the homozygous variant c.331G > A (allele A) identified in *SLC24A5* for sample MSK 357. **(C)** The heterozygous variant c.1122C > G identified in *SLC45A2* and the heterozygous variant c.331G > A identified in *SLC24A5* for sample MSK 414. *The y-axis represents the relative fluorescent units (RFU) and the x-axis represents size in base pairs (bp). The panel above the peaks represent the gene of interest. The dye mix probe colours represent the following ddNTPs: Adenine (A) - green; Guanine (G) – blue and Cytosine (C) – black. Each peak label contains the allele (al), size of fragment (sz) and height of peak (ht).*

3.6 Allele and genotype frequencies

The calculated allele and genotype frequencies for *SLC45A2* (rs16891982) and *SLC24A5* (rs1426554) according to the SA population census groups are listed in Table 3.2. There were eight observed genotype combinations from both SNPs that were identified within the cohort. The genotype combination **GG** (*SLC45A2*, rs16891982) **GG** (*SLC24A5*, rs1426554) was not observed in the cohort. Based on the Pearson's Chi-Squared test, the allele data observed for each SNP within the cohort were not in HWE ($p < 0.0001$; Appendix H).

Table 3.2 Allele and genotype frequencies for *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654) according to the SA population census groups.

	***SLC45A2* (rs16891982)**					***SLC24A5* (rs1426654)**				
	Allele		**Genotype**			**Allele**		**Genotype**		
	G	**C**	**GG**	**GC**	**CC**	**A**	**G**	**AA**	**GA**	**GG**
African Black (n = 143)	0.01	0.99	0.01	0.01	0.99	0.10	0.90	0.01	0.19	0.80
Coloured (n = 75)	0.31	0.69	0.15	0.32	0.53	0.49	0.51	0.27	0.44	0.29
Indian/Asian (n = 57)	0.05	0.95	0.00	0.11	0.89	0.57	0.43	0.4	0.33	0.26
White (n = 114)	0.93	0.07	0.86	0.13	0.01	0.96	0.04	0.91	0.09	0.00
TOTAL (n = 389)	**0.34**	**0.66**	**0.28**	**0.12**	**0.60**	**0.49**	**0.51**	**0.38**	**0.23**	**0.39**
African* *(n = 661)*	*0.04*	*0.96*	*0.01*	*0.06*	*0.93*	*0.07*	*0.93*	*0.01*	*0.13*	*0.86*
East Asian* *(n = 504)*	*0.01*	*0.99*	*0.00*	*0.01*	*0.99*	*0.01*	*0.99*	*0.00*	*0.02*	*0.98*
South Asian* *(n = 489)*	*0.06*	*0.94*	*0.01*	*0.11*	*0.88*	*0.69*	*0.31*	*0.51*	*0.36*	*0.13*
European* *(n = 503)*	*0.94*	*0.06*	*0.89*	*0.10*	*0.01*	*0.99*	*0.01*	*0.99*	*0.01*	*0.00*

*Allele and genotype frequencies obtained from:

https://www.ensembl.org/Homo_sapiens/Info/Index (accessed 19th September 2019)

*Allele and genotype frequencies according to the **1000 Genomes Project Phase 3** for *SLC45A2* (rs16891982) and *SLC24A5* (rs1426554)

3.7 MI predictive statistical model

A GLM was then developed to assess if MI could be predicted from the genotypes from the two SNPs. The GLM that best suited the cohort had an AIC value of 6.76. There were four genotypes that were comprised the model: **CC** (*SLC45A2*, rs16891982), **GC** (*SLC45A2*, rs16891982), **AA** (*SLC24A5*, rs1426554) and **GA** (*SLC24A5*, rs1426554).

The intercept value of the GLM was 0.026 ($p < 0.0001$) and the covariates were chosen based on their influence on the backward hierarchical selection (Table 3.3). The MI predictive model equation (Equation 1) was used to identify the specific individual SNP genotype which was selected as covariates for the model. The coefficient values for the GLM were used in the equation for each covariant (Table 3.3).

Equation 1: Mean MI prediction model equation for *SLC45A2* (rs16891982) and *SLC24A5* (rs1426554)

Mean MI = [0.026 – 0.008 (*SLC45A2*, rs16891982: genotype CC) **– 0.003** (*SLC45A2*, rs16891982: genotype GC) **+ 0.006** (*SLC24A5*, rs1426554: genotype AA) **+ 0.003** (*SLC24A5*, rs1426554: genotype GA)**]$^{-1}$**

Table 3.3 GLM covariates and their respective values

Covariate	Coefficient	Standard Error	P-value
Intercept	0.026	0.0007	$p < 0.0001$
CC (*SLC45A2*, rs16891982)	-0.008	0.0007	$p < 0.0001$
GC (*SLC45A2*, rs16891982)	-0.003	0.0008	$p = 0.001$
AA (*SLC24A5*, rs1426554)	0.006	0.0006	$p < 0.0001$
GA (*SLC24A5*, rs1426554)	0.003	0.0004	$p < 0.0001$

The model was assessed through comparing the mean predicted MI based on the genotype combination to the mean recorded MI observed for each genotype combination. The predicted MI reading was generated using the above equation (Equation 1) while the recorded MI reading was taken using the DSM II ColorMeter. The figure below illustrates the genotype combinations against the mean MI (Figure 3.7).

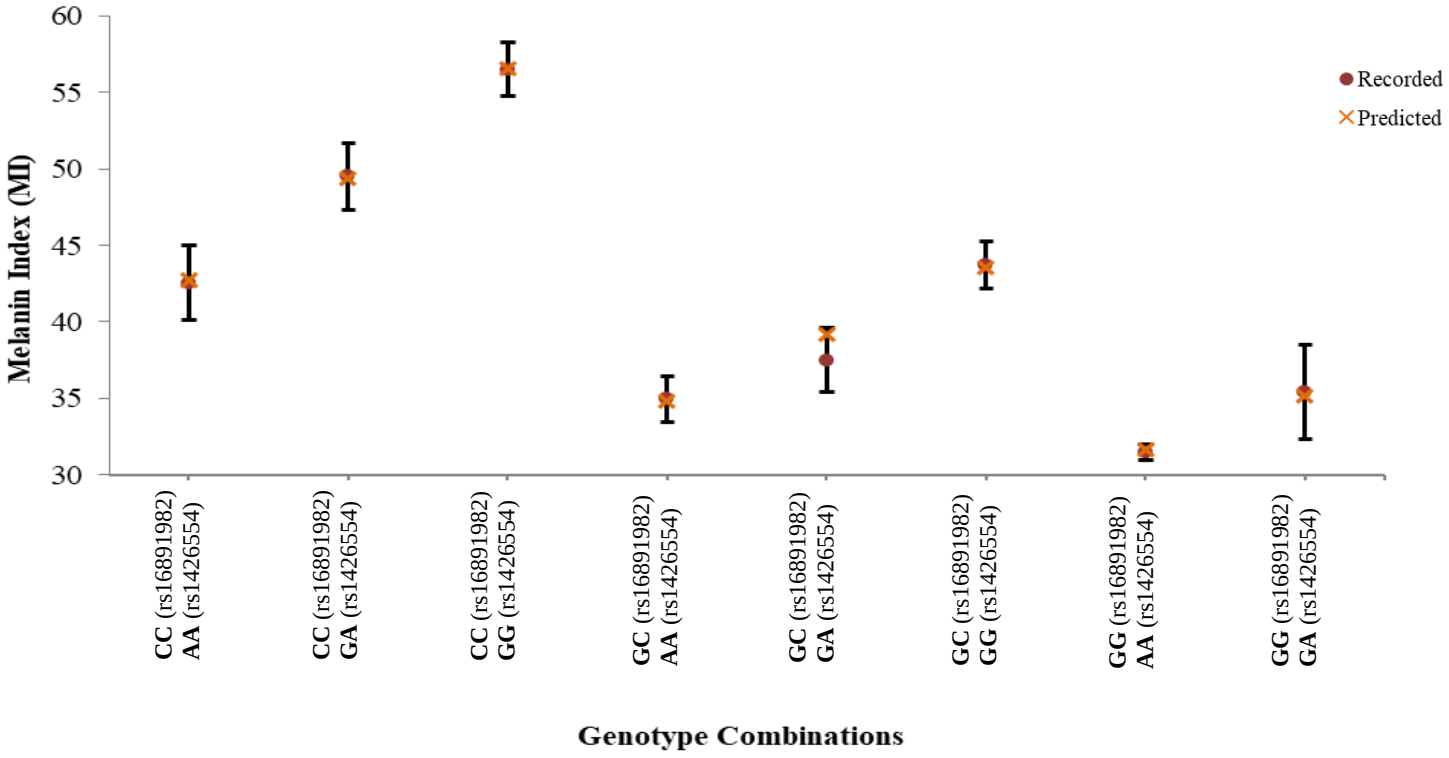

Figure 3.7 Scatter plot depicting the predicted and recorded MI for each combined genotype with a 95 % confidence bands for each combination. *Circles represent recorded MI while crosses represent predicted MI. The error bars (black) represent the 95 % confidence interval.*

The predicted and recorded mean MI for each observed combined genotype is mentioned in Table 3.4 below. All the predicted mean MI readings for each combined genotype fell within the 95 % confidence interval range of the recorded mean MI readings (Figure 3.7 + Table 3.4). Thus, for all genotype combinations, an association between the predicted MI and the recorded MI was detected.

Table 3.4 The predicted and recorded mean MI readings for the genotype combinations within the cohort.

***SLC45A2* (rs16891982)**	***SLC24A5* (rs1426554)**	**n**	**Predicted MI**	**Recorded MI**	**95 % confidence interval**
CC	AA	28	42.72	42.56	2.41
CC	GA	61	49.41	49.52	2.19
CC	GG	144	56.56	56.53	1.72
GC	AA	25	34.84	34.96	1.49
GC	GA	13	39.16	37.48	2.09
GC	GG	8	43.52	43.74	1.56
GG	AA	95	31.65	31.48	0.54
GG	GA	15	35.18	35.44	3.09

Chapter 4: Discussion and conclusion

There is a burden of individuals who remain unidentified both globally and locally (Reid, Martin, and Heathfield 2019). The use of DNA is a well-established method to assist in human identification, but relies on a database of DNA profiles from known individuals for matching purposes (Butler 2010). The NFDD is still in its infancy, as the legal framework to regulate the database was only recently established. While the prediction of phenotypic characteristics is currently prohibited by the DNA Act, its value in generating leads for victim identification prompted this research study. The aim of this study was to develop and assess a prediction model for MI based on the genotypes of two forensic molecular markers for the intended purpose of improving local human identification techniques. Consequently, a SNaPshot® PCR approach (Figure 3.6) was used to successfully genotype *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654) variants in a statistically significant cohort of 389 SA individuals. A GLM (Table 3.3) was developed which was able to predict MI from these variants within the 95 % confidence interval range of the recorded MI readings (Figure 3.7).

4.1 Assessment of the prediction model

In order to predict the MI from an individual's genotype combination of the two aforementioned SNPs, a GLM was developed (Table 3.3). The model was designed based on the genotypes at the two SNP loci. The prediction accuracy was assessed for each genotype combination by comparing the predicted MI against the recorded MI within a 95 % confidence interval. The covariates of the prediction model displayed extremely low standard error values (Table 3.3) which suggested that there was little spread in the MI distribution for each genotype combination (Christensen et al. 2014; Harvill 1991). The low standard error values are possibly attributed to an appropriately large sample size (Harvill 1991) – the cohort from the pilot study conducted by Slabbert (2017) consisted of 92 individuals which was increased to 389 individuals in this study (Figure 3.1) based on the power calculation (Appendix A). The increased cohort size mitigated the challenges associated with the prediction model as discussed by Slabbert (2017). In addition, each covariate also displayed a p – value less than 0.05 (Table 3.3) indicating that there was a statistically significant association between the covariates and MI. It is evident by the results obtained from this study that this model was adequate in predicting MI based on the genotypes for the two SNPs (Figure 3.7).

Although the GLM is best suited to predict MI from the genotypes of the two SNPs, the model only comprises of four genotype covariates – two from each SNP (Table 3.3). The remaining genotypes for each SNP was excluded from the model due to multicollinearity – a situation whereby the predictor variables in a statistical model are correlated as opposed to being independent (Dormann et al. 2013). The genotype combinations obtained from both SNPs were also excluded because their p – values were greater than the standard significant level ($p = 0.1$) for removal from the backward hierarchical estimation model.

Of note, the genotype combination **GG** (*SLC45A2*, rs16891982) and **GG** (*SLC24A5*, rs1426554) was completely excluded from the model as it was not observed within the cohort (Table 3.4). The reason as to why this genotype combination was not observed may be attributed to the contradictory characteristics these SNPs possess (Cerqueira et al. 2014). Several studies have discovered that the probable function for the **G** allele in *SLC45A2* (rs16891982) is associated with increased melanin content (Norton et al. 2007; Spichenok et al. 2011) while the **G** allele function in *SLC24A5* (rs1426554) is related to lighter skin pigmentation (Norton et al. 2007; Cook et al. 2009). This genotype combination has also not been observed in previous literature (Norton et al. 2007; Spichenok et al. 2011; Cook et al. 2009) suggesting that their combined impact may not influence the MI as a result of their opposing SNP characteristics. Another explanation for the exclusion of this genotype combination is that such homozygotes will only occur at extremely low frequencies in both European and African derived populations (The 1000 Genomes Project Consortium 2015) and so will not be captured by the size of the cohort used in the study.

The statistical analysis used for molecular phenotyping has been incorporated through the development of prediction models based on the statistically relevant SNPs' ability to predict phenotypic characteristics (Kastelic and Drobnič 2012). The generation of prediction models are frequently performed with hair and eye colour prediction (Liu et al. 2009; Walsh et al. 2011a). For example, the HIrisPlex system used logistic regression models to predict a phenotype category, where eye colour was categorised as blue, brown and intermediate eye colour (Walsh et al. 2014). Similarly, skin colour prediction has been predominantly dependent on previously established models that predict skin colour using categorically based skin colour charts. These charts are subjective and restrict skin colour categorisation to pale, intermediate, medium or dark skin colour (Hart et al. 2013; Maroñas et al. 2014; Pneuman et al. 2012).

In contrast, this study designed a model (Equation 1 and Table 3.3) that predicts skin colour using MI, a continuous variable, rather than the categorical variables used in previous skin colour prediction models (Walsh et al. 2017; Zaorska, Zawierucha, and Nowicki 2019). The model developed in this study mitigates the limitations associated with utilising the aforementioned categorical skin colour prediction models in both a forensic and SA context. In a SA setting, there is a heightened sensitivity regarding categorising individuals based on skin colour mainly due to the history of Apartheid (Posel 2001). Although this type of prediction modelling is novel in its use to analyse SNPs for MI prediction, its introduction in a local context can assist in steering away from the racial stigmatisation and profiling exhibited based on skin colour in SA.

4.2 MI as a variable of interest

4.2.1 MI versus SA population census groups

Previous studies (Ang et al. 2012; Parra, Kittles, and Shriver 2004; Spichenok et al. 2011) have utilised population groups as a variable for association with genotype data; however, this study utilised MI measurements, which are displayed as a function of the human skin's absorption characteristic and provides an objective melanin assessment for skin pigmentation (Park and Lee 2005; Van der Wal et al. 2013; Wright et al. 2019).

Initially, the MI measurements were compared between the different SA population census groups (Figure 3.3). Although the MI differences were statistically significant between some of the population groups ($p < 0.0001$), there were overlapping MI ranges implying that the SA population census groups were not a suitable representation for MI. For example, an overlapping MI range was observed between the African Black (36.93 – 86.19) and Coloured (30.00 – 60.41) population groups (Figure 3.3). This could be attributed to the extensive variation in skin pigmentation levels in individuals who self-identify with either population group (Brown 2000). The overlapping MI ranges (Figure 3.3) also indicated that SA population census groups were uninformative as a variable for skin colour prediction as they have scientific limitations in addition to its ethical issues as previously discussed by Toom et al. (2016). The results obtained from the MI readings in this study (Figure 3.3) support the notion that MI, a continuous variable, is a more appropriate variable than the SA population census groups, especially since the SA population is highly admixed (Adhikari 2005; de Wit et al. 2010). Furthermore, the population census groups in SA symbolises a history of racism

and racial segregation of individuals based on skin colour (Christopher 2002). This study was hugely important in a local context as it moves away from the use of racial classifications.

Although the application and interpretation of MI as an indicator of skin pigmentation is perhaps understood in the scientific community (Lasisi and Shriver 2018; Majewski et al. 2016; Stokowski et al. 2007; Martin et al. 2017), the same does not necessarily hold true in the current criminal justice system. In a SA context, the law enforcement and justice system adheres to the DNA Act which was introduced to administer and regulate the NFDD (Heathfield 2014). This database is used to store forensic DNA profiles of offenders and its access is strictly limited to the law enforcement for investigative purposes (Dedrickson 2017; de Wet and Visser 2017). However, the NFDD does not store any phenotypic or SNP data because the DNA Act prohibits the use and access of this type of data (*Criminal Law (Forensic Procedures) Amendment Act, No. 37 of 2013* 2014).

The interpretation of MI and SNP data by law enforcement and court of law is not well-established. The law enforcement and justice system unfortunately still utilise racial groups to classify individuals and is currently an important variable in assisting in human identification. For SA law enforcement officials, the SA racial groups act as a proxy for skin pigmentation which provides a basis for criminal investigation purposes such as generating suspect leads, narrowing suspect pools and assisting with victim identification (Dissel and Kollapen 2002). Therefore, questions are raised as to the direct applicability of using MI as a skin pigmentation proxy for the SA state services.

In order to maximise the relevance of the results in this study to SAPS and SA judiciary, one suggestion is to create a visual representation of the MI scale so that the MI readings can be visually interpreted and understood in layman's terms by the officials. The visual representation of the MI scale would be utilised to compare to an assessment of an individual's skin colour by eyewitness or law enforcement records. This scale may also be beneficial for facial reconstruction artists to incorporate skin colour into images, as in SA, facial images are currently produced in grayscale to avoid the assumption of skin colour and race (Russell et al. 2006). However, it is important to be aware of the potential disjoint between visual appearance and measured skin colour, which is a complex aspect for molecular phenotyping. Despite this scientific and social value, the SA judicial system tends to default to racial stigmatisation mainly due to the prejudicial racial history of SA and the formation of classification categories (Braun 2006; Bickford-Smith 1995; Garvett 2017). As

such, the logistics of actively using MI instead of racial groups will need to be debated between legislators and the criminal justice system. Overall, the intention for using MI as a variable for skin pigmentation is to shift the approach *away* from integrating race into forensic genetic research in the hopes of eliminating racial stigmatisation in the SA criminal justice system (Duster 2006; Ossorio 2006).

4.2.2 MI versus ancestral groups

In addition to the MI comparison between the SA population groups, the MI measurements were also compared between the different ancestral groups (Figure 3.4). A similar trend was observed in which the MI differences were significantly different between ancestral groups ($p < 0.0001$) but overlapping MI ranges were also observed (Figure 3.4). These results suggest that the ancestral groups were also not a suitable proxy for MI. Parra, Kittles and Shriver (2004) reported that the correlation between ancestry and skin pigmentation was mutable, emphasising the need to be cautious when utilising skin pigmentation as an ancestral proxy as well as extrapolating genotype-phenotype data from one admixed population to another (Parra, Kittles, and Shriver 2004).

As previously mentioned, the ancestry of an individual refers to either the geographical origin of a population or the line of heritage or descent of a group (Ali-Khan et al. 2011). A similar view of ancestry is BGA, which is used to describe an individual's origin based on the geographical location(s) of their predecessors deduced by comparison with the current populations residing in these locations (Royal et al. 2010; Shriver et al. 2003; Phillips 2015). Based on this definition, an individual's ancestry is not defined by their phenotypic traits; hence, identifying an individual solely based on ancestral origin, particularly in SA, remains questionable (Cho and Sankar 2004). The phenotypic prediction based on ancestry inference is problematic because one has to *assume* an individual's appearance based on a *stereotype* of what individuals generally resemble from a certain region. This approach poses many ethical and social concerns as discussed by Kayser (2015).

The major concern with using ancestry as a variable for skin colour prediction is that admixed populations consists of different ancestry proportions and, thus are unable to predict skin colour for individuals belonging to this population (Valenzuela et al. 2010). Furthermore, another concern is that it might be unclear as to an individual's appearance based on their mixed ancestral background or even if their ancestry proportions are unknown. These challenges are exacerbated in SA, due to the presence of the highly diverse mixed

ancestral populations as well as the extensive skin colour variations exhibited by the SA population census groups (Maroñas et al. 2014). Even within population groups, there are different ethno-linguistic groups, which are often partly geographically isolated. For example, the African Black population group typically represent the indigenous black South Africans, mainly the Bantu-speaking communities, with different ethno-linguistic groups such as Nguni, Shangaan-Tsonga, Sotho-Tswana, and Venda, who are typically clustered in certain provinces within the country (Gradin 2015).

The SAC are a unique admixed population that is native to southern Africa which comprises of mixed ancestries from African (predominantly Khoesan), European and Asian descent (de Wit et al. 2010) rendering it a biologically heterogenous group with variable and complex admixture. Tishkoff et al. (2009) reported that the highest level of mixed ancestry was observed within the SAC population (Tishkoff et al. 2009). Therefore, skin pigmentation variability among African populations is underappreciated and extremely complex with the genetic architecture varying by latitude (Quillen et al. 2019). For example, there are marked differences in skin pigmentation between the Bantu derived and Khoesan populations (non-Bantu) (Lasisi and Shriver 2018). The Khoesan population exhibit a lighter skin pigmentation compared to the majority of other African populations, especially the equatorial African regions, as reported by Martin et al. (2017).

Furthermore, in the SA population, the White South Africans generally represent the European ancestries and are descendants from mainly the British, Dutch, Irish and German settlers who arrived at the Cape during the 17th century (Adams, Van de Vijver, and de Bruin 2012). Lastly, the Indian/Asian population group usually comprise of individuals with ancestral origins from East and South Asia whose descendants migrated to SA in the 19th century (Adams et al. 2014). The natural variation of skin pigmentation in South Asia is extensive. A notable study by Stokowski et al. (2007) identified the AIMs of interest, *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654), to show genome wide significance for association with skin pigmentation. The associations detected account for the majority of the skin pigmentation variation in the South Asian population. The study also reported a clear differentiation of skin-reflectance measurements between South Asian populations (southern and north-western regions) (Stokowski et al. 2007).

Each SA population census group is comprised of diverse ancestral groups and, thus various ancestries could be present in individuals who typically self-identify as one of the SA

population census groups. However, it is important to note that the SA population census and ancestral groups were self-reported by the participants and, thus is entirely subjective. For some participants, their self-identified SA population census group did not align with the expected ancestry of the population group described above (Figure 3.1A + B). For example, some individuals who reported as Coloured also indicated that their ancestry was only African. The term "Coloured" as part of the SA population census groups, is utilised knowing many of the individuals are in fact genetically homozygous to one origin but were classified under this single racial label during Apartheid, which is still used currently in our democracy. This is exemplified by the Khoesan people whom under apartheid, were socially and bureaucratically classified as Coloured but are genetically derived from African descent. This alludes to the fact that the individuals in the cohort who self-reported as Coloured but are of African descent could be Khoesan. This result reiterates that the census groups and self-reported ancestries both have limitations in molecular phenotyping in a SA context and that distinguishing non-Europeans in SA is a forensic challenge.

The use of MI overcomes some of the limitations associated with using ancestry for EVC prediction, because MI no longer relies on assumptions and stereotypes about what someone might look liked based on their ancestry. MI is an objective measure of skin colour represented by in a numerical and quantitative manner. This further motivates that MI should rather be a used as the variable for genotype-phenotype association for skin colour in molecular phenotyping, particularly in SA.

4.3 Next generation sequencing (NGS) for molecular phenotyping

As previously stated, a recent study by Chaitanya et al. (2018) developed the HIrisPlex-S system which was the first forensically validated tool for skin pigmentation prediction and simultaneous hair, eye and skin colour prediction from DNA (Chaitanya et al. 2018). The genetic data was generated from two forensically validated multiplex SNaPshot® assays and the phenotypic prediction was based on statistical models (Chaitanya et al. 2018). A similar approach was undertaken in our study in which SNaPshot® genotyping for only two SNPs namely, *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654) was performed and a skin colour prediction model was generated (Figure 3.6 and Table 3.3).

The limitations associated with SNaPshot® and capillary electrophoresis technology as discussed by Sobiah et al. (2018) has prompted the need to transition towards targeted NGS solutions expanding the knowledge on EVCs and ancestry genetics (Sobiah et al. 2018). A

notable study by Breslin et al. (2019) presented NGS solutions for the HIrisPlex-S system by using NGS platforms to develop NGS-derived HIrisPlex-S assays (Breslin et al. 2019). These NGS assays were found to be extremely sensitive as simulated forensic casework samples, including low quantity touch-DNA and compromised DNA samples, generated good quality data. This study set the benchmark to further expand NGS-based molecular phenotyping by including additional SNPs for phenotypic predictions as well as AIMs for BGA inference (Breslin et al. 2019). Furthermore, the SNPs characteristics discussed by Phillips (2015) have allowed for the development of NGS-derived SNP panels for large scale DNA identification. Parsons et al. (2019) developed NGS strategies through utilising large multiplex SNP panels, containing tri-allelic SNPs and multiple micro-haplotype loci, for missing person's cases at the International Commission on Missing Persons (ICMP) (Parsons et al. 2019).

Essentially, NGS has revolutionised the approach towards the scientific development of molecular techniques for forensic application (Børsting and Morling 2015). Currently, the NGS technology is predominantly utilised in the biomedical fields; however, it has gained traction in the forensics field. NGS has obtained great advances in improving forensic genetic research through its high-throughput capabilities (Alvarez-Cubero et al. 2017). The implementation of NGS in forensics has not only enhanced STR genotyping but also has played an influential role in the analysis of SNP markers for ancestral and phenotypic inferences (Yang, Xie, and Yan 2014; Seo et al. 2013). In forensics, there are two NGS platforms of interest, namely Ion Torrent™ PGM™ and MiSeq® FGx™ (Churchill et al. 2015; Sobiah et al. 2018). Over the last few years, many kits for these NGS platforms have become commercially available especially kits relating to SNP markers for BGA inference and phenotypic prediction (Bruijns, Tiggelaar, and Gardeniers 2018). Previously published articles (Apaga et al. 2017; Ambers et al. 2016; Warshauer et al. 2014) have comprehensively evaluated the use of these NGS-derived kits for forensic SNP analysis.

4.4 Limitations and future work

In this study, the cohort was recruited through convenience sampling which poses various limitations as the information gained from these samples lacked randomness and, thus introduced hidden bias into the cohort. The lack of random sampling of the cohort may not have been fully representative of the population of interest, as the participants were predominately recruited from UCT and its surrounding areas. Although the individuals recruited from these areas originate from diverse backgrounds, they represent individuals

with similar occupations and, thus are exposed to similar environments which could result in a skewed sample dataset.

The recruitment process contained exclusion criteria for the cohort therefore individuals who were exposed to skin colour treatments such as tanning beds or synthetic colouration and individuals with pigment disorders were not included. These variables affect an individual's skin pigmentation through altering the melanin production in the skin which leads to misleading MI readings. These variables were not considered in the development of the GLM and, thus the model cannot be extrapolated to these contexts. The assessment of these markers for skin colour in a wider range of individuals, with diverse occupations and altered skin pigments needs to be explored in future work.

The allele and genotype frequencies obtained from the cohort for each SNP were not in HWE ($p < 0.0001$; Appendix H). Typically, this could be considered as a limitation; however, in this study, this result is expected from the HWE test due to the admixed nature of the population. An admixed population does not demonstrate random mating, given that there was co-ancestry. Consequently, the co-ancestry disrupts linkage disequilibrium as it interferes with heterozygosity which causes the genotype ratios to be disproportional. This can be interpreted as an indication of selection or as an imbalance in the HWE test's ability to constantly detect all genotypes.

Another limitation of this study was that the experimental workflow was optimised using 50 ng of DNA from each participant and this may pose difficulties in direct application to forensic casework. In forensic casework, the DNA yield and quantity varies (Van Oorschot and Jones 1997), but is usually less than 50 ng. While the amount of DNA was deemed appropriate for this study which aimed to develop a model and assess the suitability of DNA markers for MI prediction, it is currently limited in its applications to forensic casework samples. Therefore, future research needs to involve optimising this experimental workflow using compromised samples such as low template DNA samples, degraded DNA samples and mixed DNA samples (Budowle and Van Daal 2008). These assessments could then lead into a validation study, which would be required in a forensic context, to assess the workflow in terms of accuracy, precision, reproducibility and sensitivity (Børsting, Rockenbauer, and Morling 2009).

Lastly, the model could be improved to increase the accuracy of MI prediction by including additional SNPs in the model, particularly SNPs that are relevant to different population and

ancestral groups. The focus should centre around examining SNPs in genes that play a role in the biochemical pathways for melanin synthesis (Branicki 2009). The genes that have been previously studied and should be reviewed include: *OCA* (Oculocutaneous albinism), *MC1R* (Melanocortin 1 receptor), *KITLG* (Kit ligand), *TYR* (Tyrosinase), *HERC2* (HECT and RLD domain containing E3 ubiquitin protein ligase 2), *SLC24A4* (Solute Carrier Family 24 Member 4), *ASIP* (Agouti Signalling Protein) and *TPCN2* (Two Pore Segment Channel 2) (Maroñas et al. 2014; Miller et al. 2007; Stokowski et al. 2007; Valenzuela et al. 2010; Spichenok et al. 2011).

4.5 Conclusion

This study, in conjunction with the study conducted by Slabbert (2017), was the first study in a SA context that researched the use of SNP analysis to predict skin pigmentation for forensics applications. Although the concepts and techniques for this research were established from previous literature (Slabbert 2017; Cerqueira et al. 2014), this study expanded on them and contributed by; (i) recruiting participants to create a statistically significant cohort; (ii) performing the SNaPshot® assay on a large cohort; and (iii) generating a GLM which was utilised to assess the SNP's ability to accurately predict skin colour. This research investigated the objective assessment of skin colour based on MI as an alternative for the use of established racial categories which conveys prejudice in SA. The intent was to hopefully steer away from prejudicial racial classification and utilise a continuous and more objective variable which was deemed more appropriate and informative for forensic purposes, particular in a SA context.

The aim of this study was achieved as the relationship between the two markers, *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654), and MI were successfully investigated in the SA population. The development of the GLM allowed for MI prediction from the genotype combinations of the two markers. The model was able to predict MI within the 95 % confidence interval of the recorded MI. This model can be improved in the future by including additional markers that are informative in admixed populations, and possibly by using a NGS approach. The use of this model was an innovative technique to predict a continuous MI scale. The results suggested that the aforementioned markers were consistently associated with MI in the highly admixed SA population.

Together with an amendment to the SA legislation, the optimisation of the workflow for casework samples as well as validation needs to occur before this tool can be used as a

forensic tool. However, the results from such a tool should always be used in conjunction with additional evidence such as conventional DNA analysis and anthropological findings (i.e. age, appearance and ancestry). This study was fundamental in laying the foundation for future forensic research in molecular phenotyping within SA.

References

Adams, Byron G., Fons J.R. Van de Vijver, and Gideon P. de Bruin. 2012. “Identity in South Africa: Examining self-descriptions across ethnic groups.” *International Journal of Intercultural Relations* 36 (3): 377–88. https://doi.org/10.1016/j.ijintrel.2011.11.008.

———, Fons J.R. Van de Vijver, and Gideon P. de Bruin and Cynthia B. Torres. 2014. “Identity in descriptions of others across ethnic groups in South Africa.” *Journal of Cross-Cultural Psychology* 45 (9): 1–23. https://doi.org/10.1177/0022022114542466.

Adhikari, Mohamed. 2005. “Contending approaches to coloured identity and the history of the coloured people of South Africa.” *History Compass* 3 (1): 1–16. https://doi.org/10.1111/j.1478-0542.2005.00177.x.

Adhikari, Kaustubh, Javier Mendoza-Revilla, Anood Sohail, Macarena Fuentes-Guajardo, Jodie Lampert, Juan Camilo Chacón-Duque, Malena Hurtado, et al. 2019. “A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia.” *Nature Communications* 10 (1): 1–16. https://doi.org/10.1038/s41467-018-08147-0.

Ali-Khan, Sarah E., Tomasz Krakowski, Rabia Tahir, and Abdallah S. Daar. 2011. “The use of race, ethnicity and ancestry in human genetic research.” *Human Genome Organisation Journal* 5 (1–4): 47–63. https://doi.org/10.1007/s11568-011-9154-5.

Allwood, Julia. 2013. “Using DNA to predict externally visible characteristics in humans for forensic use in New Zealand.” PhD diss., The University of Auckland.

Alvarez-Cubero, Maria Jesus, Maria Saiz, Belén Martínez-García, Sara M. Sayalero, Carmen Entrala, Jose Antonio Lorente, and Luis J. Martinez-Gonzalez. 2017. “Next generation sequencing: An application in forensic sciences?” *Annals of Human Biology* 44 (7): 581–92. https://doi.org/10.1080/03014460.2017.1375155.

Ambers, Angie D., Jennifer D. Churchill, Jonathan L. King, Monika Stoljarova, Harrell Gill-King, Mourad Assidi, Muhammad Abu-Elmagd, Abdelbaset Buhmeida, and Bruce Budowle. 2016. “More comprehensive forensic genetic marker analyses for accurate human remains identification using massively parallel DNA sequencing.” *BioMed Central Genomics* 17 (750): 21–30. https://doi.org/10.1186/s12864-016-3087-2.

Amorim, António, and Luísa Pereira. 2005. “Pros and cons in the use of SNPs in forensic

kinship investigation: A comparative analysis with STRs." *Forensic Science International* 150 (1): 17–21. https://doi.org/10.1016/j.forsciint.2004.06.018.

Ang, Khai C., Mee S. Ngu, Katherine P. Reid, Mei S. Teh, Zamzuraida S. Aida, Danny X.R. Koh, Arthur Berg, et al. 2012. "Skin color variation in Orang Asli tribes of peninsular Malaysia." *PLoS ONE* 7 (8): 1–7. https://doi.org/10.1371/journal.pone.0042752.

Apaga, Dame Loveliness T., Sheila E. Dennis, Jazelyn M. Salvador, Gayvelline C. Calacal, and Maria Corazon A. De Ungria. 2017. "Comparison of two massively parallel sequencing platforms using 83 single nucleotide polymorphisms for human identification." *Scientific Reports* 7 (398): 1–6. https://doi.org/10.1038/s41598-017-00510-3.

Bamshad, Michael, Stephen Wooding, Benjamin A. Salisbury, and J. Claiborne Stephens. 2004. "Deconstructing the relationship between genetics and race." *Nature Reviews Genetics* 5 (8): 598–609. https://doi.org/10.1038/nrg1401.

Beleza, Sandra, Nicholas A. Johnson, Sophie I. Candille, Devin M. Absher, Marc A. Coram, Jailson Lopes, Joana Campos, et al. 2013. "Genetic architecture of skin and eye color in an African-European admixed population." *PLoS Genetics* 9 (3): 1–15. https://doi.org/10.1371/journal.pgen.1003372.

Bickford-Smith, Vivian. 1995. "South African urban history, racial segregation and the unique case of Cape Town?" *Journal of Southern African Studies* 21 (1): 63–78. https://doi.org/10.1080/03057079508708433.

Børsting, Claus, and Niels Morling. 2015. "Next generation sequencing and its applications in forensic genetics." *Forensic Science International: Genetics* 18 (1): 78–89. https://doi.org/10.1016/j.fsigen.2015.02.002.

Bozdogan, Hamparsum. 1987. "Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions." *Psychometrika* 52 (3): 345–70. https://doi.org/10.1007/BF02294361.

———, Eszter Rockenbauer and Niels Morling. 2009. "Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard." *Forensic Science International: Genetics* 4 (1): 34–42. https://doi.org/10.1016/j.fsigen.2009.04.004.

Branicki, Wojciech. 2009. “Studies on predicting pigmentation phenotype for forensics purposes.” *Problems of Forensic Sciences* 77: 29–52.

Braun, Lundy. 2006. “Reifying human difference: The debate on genetics, race, and health.” *International Journal of Health Services* 36 (3): 557–73. https://doi.org/10.2190/8JAF-D8ED-8WPD-J9WH.

Breslin, Krystal, Bailey Wills, Arwin Ralf, Marina Ventayol Garcia, Magdalena Kukla-Bartoszek, Ewelina Pospiech, Ana Freire-Aradas, et al. 2019. “HIrisPlex-S system for eye, hair, and skin color prediction from DNA: Massively parallel sequencing solutions for two common forensically used platforms.” *Forensic Science International: Genetics* 43: 1–13. https://doi.org/10.1016/j.fsigen.2019.102152.

Brown, Kendrick. 2000. “Coloured and black relations in South Africa: The burden of racial hierarchy.” *Macalester International* 9: 198–207.

Bruijns, Brigitte, Roald Tiggelaar, and Han Gardeniers. 2018. “Massively parallel sequencing techniques for forensics: A review.” *Electrophoresis* 39 (21): 2642–54. https://doi.org/10.1002/elps.201800082.

Budowle, Bruce, and Angela Van Daal. 2008. “Forensically relevant SNP classes.” *BioTechniques* 44 (5): 603–10. https://doi.org/10.2144/000112806.

Butler, John M. 2010. *Fundamentals of forensic DNA typing*. *Academic Press*. 1st ed. Burlington, Massachusetts, USA: Elsevier. https://doi.org/10.1016/C2009-0-01945-X.

———, Michael D. Coble, and Peter M. Vallone. 2007. “STRs vs. SNPs: Thoughts on the future of forensic DNA testing.” *Forensic Science,Medicine, and Pathology* 3 (3): 200–205. https://doi.org/10.1007/s12024-007-0018-1.

Castel, Charmain Vanessa. 2014. “Inference of biogeographical ancestry and pigmentation phenotype using single nucleotide polymorphisms.” PhD diss., University of Technology Sydney.

Caulfield, Timothy, Stephanie M. Fullerton, Sarah E. Ali-Khan, Laura Arbour, Esteban G. Burchard, Richard S. Cooper, Billie Jo Hardy, et al. 2009. “Race and ancestry in biomedical research: Exploring the challenges.” *Genome Medicine* 1 (1): 1–8. https://doi.org/10.1186/gm8.

Cerqueira, Caio Cesar Silva de, Tábita Hünemeier, Jorge Gomez-Valdés, Virgínia Ramallo,

Carla Daiana Volasko-Krause, Ana Angélica Leal Barbosa, Pedro Vargas-Pinilla, et al. 2014. “Implications of the admixture process in skin color molecular assessment.” *PLoS ONE* 9 (5): 1–7. https://doi.org/10.1371/journal.pone.0096886.

Chaitanya, Lakshmi, Krystal Breslin, Sofia Zuñiga, Laura Wirken, Ewelina Pośpiech, Magdalena Kukla-Bartoszek, Titia Sijen, et al. 2018. “The HIrisPlex-S System for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation.” *Forensic Science International: Genetics* 35 (October 2017): 123–35. https://doi.org/10.1016/j.fsigen.2018.04.004.

———, Irena Zupanič Pajnič, Susan Walsh, Jože Balažic, Tomaž Zupanc, and Manfred Kayser. 2017. “Bringing colour back after 70 years: Predicting eye and hair colour from skeletal remains of World War II victims using the HIrisPlex System.” *Forensic Science International: Genetics* 26: 48–57. https://doi.org/10.1016/j.fsigen.2016.10.004.

Cho, Mildred K., and Pamela Sankar. 2004. “Forensic genetics and ethical, legal and social implications beyond the clinic.” *Nature Genetics* 36 (11): S8-12. https://doi.org/10.1038/ng1594.

Christensen, Angi M., Christian M. Crowder, Stephen D. Ousley, and Max M. Houck. 2014. “Error and its meaning in forensic science.” *Journal of Forensic Sciences* 59 (1): 123–26. https://doi.org/10.1111/1556-4029.12275.

Christopher, Anthony J. 2002. “‘To define the indefinable’: Population classification and the census in South Africa.” *Area* 34 (4): 401–8. https://doi.org/10.1111/1475-4762.00097.

Churchill, Jennifer D., Joseph Chang, Jianye Ge, Narasimhan Rajagopalan, Sharon C. Wootton, Chien Wei Chang, Robert Lagacé, Wenchi Liao, Jonathan L. King, and Bruce Budowle. 2015. “Blind study evaluation illustrates utility of the ion PGM™ system for use in human identity DNA typing.” *Croatian Medical Journal* 56 (3): 218–29. https://doi.org/10.3325/cmj.2015.56.218.

Constitution for the Republic of South Africa, Act No. 108 of 1996. (1996). Available: https://www.gov.za/sites/default/files/images/a108-96.pdf [2019, March 21]

Cook, Anthony L., Wei Chen, Amy E. Thurber, Darren J. Smit, Aaron G. Smith, Timothy G. Bladen, Darren L. Brown, et al. 2009. “Analysis of cultured human melanocytes based on polymorphisms within the *SLC45A2/MATP, SLC24A5/NCKX5,* and *OCA2/P* loci.”

Journal of Investigative Dermatology 129 (1): 392–405. https://doi.org/10.1038/jid.2008.211.

Crawford, Nicholas G., Derek E. Kelly, Matthew E.B. Hansen, Marcia H. Beltrame, Shaohua Fan, Shanna L. Bowman, Ethan Jewett, et al. 2017. "craws." *Science* 358 (6365): 1–14. https://doi.org/10.1126/science.aan8433.

Criminal Law (Forensic Procedures) Amendment Act, No. 37 of 2013. (2014). Available: https://www.gov.za/sites/default/files/gcis_document/201409/37268act37of2013crimlawamend27jan2014.pdf [2019, March 20]

D'Amato, Maria Eugenia, and Mohaimin Kasu. 2017. "Population analysis of African Y-STR profiles with uniQ TYPER^{TM} Y-10 genotyping system." *Forensic Science International: Genetics Supplement Series* 6: 84–85. https://doi.org/10.1016/j.fsigss.2017.09.048.

Daca-Roszak, P., A. Pfeifer, J. Zebracka-Gala, B. Jarząb, M. Witt, and E. Ziętkiewicz. 2016. "EurEAs-Gplex - A new SNaPshot assay for continental population discrimination and gender identification." *Forensic Science International: Genetics* 20: 89–100. https://doi.org/10.1016/j.fsigen.2015.10.004.

Dario, Paulo, Helena Mouriño, Ana Rita Oliveira, Isabel Lucas, Teresa Ribeiro, Maria João Porto, Jorge Costa Santos, Deodália Dias, and Francisco Corte Real. 2015. "Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population." *International Journal of Legal Medicine* 129 (6): 1191–1200. https://doi.org/10.1007/s00414-015-1248-5.

Dedrickson, Kirsten. 2017. "Universal DNA databases: A way to improve privacy?" *Journal of Law and the Biosciences* 4 (3): 637–47. https://doi.org/10.1093/jlb/lsx041.

Department of Police. 2018. *Annual crime statistics 2017/2018.* Pretoria: Government Printer.

Dissel, Amanda, and Jody Kollapen. 2002. "Racism and discrimination in the South African penal system." *Centre for the Study of Violence and Reconciliation.* Cape Town.

Dormann, Carsten F., Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R.García Marquéz, et al. 2013. "Collinearity: A review of methods to deal with it and a simulation study evaluating their performance." *Ecography* 36 (1): 27–46.

https://doi.org/10.1111/j.1600-0587.2012.07348.x.

Draus-Barini, Jolanta, Susan Walsh, Ewelina Pośpiech, Tomasz Kupiec, Henryk Głab, Wojciech Branicki, and Manfred Kayser. 2013. “Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains.” *Investigative Genetics* 4 (3): 1–15. https://doi.org/10.1186/2041-2223-4-3.

Duster, Troy. 2006. “The molecular reinscription of race: Unanticipated issues in biotechnology and forensic science.” *Patterns of Prejudice* 40 (4): 427–41. https://doi.org/10.1080/00313220601020148.

Fitzpatrick, Thomas B. 1988. “The validity and practicality of sun-reactive skin types I through VI.” *Archives of Dermatology* 124 (1): 869–71. https://doi.org/10.1001/archderm.1988.01670060015008.

Freire-Aradas, A., Y. Ruiz, C. Phillips, O. Maroñas, J. Söchtig, A. Gómez Tato, J. Álvarez Dios, et al. 2014. “Exploring iris colour prediction and ancestry inference in admixed populations of South America.” *Forensic Science International: Genetics* 13: 3–9. https://doi.org/10.1016/j.fsigen.2014.06.007.

Gannett, Lisa. 2014. “Biogeographical ancestry and race.” *Studies in History and Philosophy of Biological and Biomedical Sciences* 47: 173–84. https://doi.org/10.1016/j.shpsc.2014.05.017.

Gans, Jeremy, and Gregor Urbas. 2002. “DNA identification in the criminal justice system.” *Trends and Issues in Crime and Criminal Justice* 226: 1–6.

Garvett, Williem H. 2017. “The myth of objectivity : Implicit racial bias and the law (part 2).” *Potchefstroom Electronic Law Journal* 20 (1): 1–25.

Giardina, Emiliano, Ilenia Pietrangeli, Cristina Martinez-Labarga, Claudia Martone, Flavio Angelis, Aldo Spinella, Gianfranco Stefano, Olga Rickards, and Giuseppe Novelli. 2008. “Haplotypes in *SLC24A5* gene as ancestry informative markers in different populations.” *Current Genomics* 9 (2): 110–14. https://doi.org/10.2174/138920208784139528.

Gill, Peter, David J. Werrett, Bruce Budowle, and Richard Guerrieri. 2004. “An assessment of whether SNPs will replace STRs in national DNA databases - Joint considerations of the DNA working group of the European Network of Forensic Science Institutes

(ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM)." *Science and Justice* 44 (1): 51–53. https://doi.org/10.1016/S1355-0306(04)71685-8.

Ginger, Rebecca S., Sarah E. Askew, Richard M. Ogborne, Stephen Wilson, Dudley Ferdinando, Tony Dadd, Adrian M. Smith, et al. 2008. "*SLC24A5* encodes a trans-golgi network protein with potassium-dependent sodium-calcium exchange activity that regulates human epidermal melanogenesis." *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M707521200.

Gradin, Carlos. 2015. "Poverty and ethnicity among black South Africans." *European Journal of Development Research* 27 (5): 921–42. https://doi.org/10.1057/ejdr.2014.76.

Haeusler, Martin, Cordula Haas, Sandra Lösch, Negahnaz Moghaddam, Igor M. Villa, Susan Walsh, Manfred Kayser, et al. 2016. "Multidisciplinary identification of the controversial freedom fighter Jörg Jenatsch, assassinated 1639 in Chur, Switzerland." *PLoS ONE* 11 (12): 1618–39. https://doi.org/10.1371/journal.pone.0168014.

Halder, Indrani, Mark Shriver, Matt Thomas, Jose R. Fernandez, and Tony Frudakis. 2008. "A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: Utility and applications." *Human Mutation* 29 (5): 648–58. https://doi.org/10.1002/humu.20695.

Han, Jiali, Peter Kraft, Hongmei Nan, Qun Guo, Constance Chen, Abrar Qureshi, Susan E. Hankinson, et al. 2008. "A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation." *PLoS Genetics* 4 (5): 1–11. https://doi.org/10.1371/journal.pgen.1000074.

Hart, Katie L., Shey L. Kimura, Vladimir Mushailov, Zoran M. Budimlija, Mechthild Prinz, and Elisa Wurmbach. 2013. "Improved eye- and skin-color prediction based on 8 SNPs." *Croatian Medical Journal* 54 (3): 248–56. https://doi.org/10.3325/cmj.2013.54.248.

Harvill, Leo M. 1991. "Standard error of measurement." *Educational Measurement: Issues and Practice* 10 (2): 33–41.

Heathfield, Laura J. 2014. "Policy required for entry of DNA profiles onto the national forensic DNA database of South Africa." *South African Journal of Science* 100 (7): 1–3. https://doi.org/10.1590/sajs.2014/20130374.

Hernandez-Pacheco, Natalia, Carlos Flores, Santos Alonso, Celeste Eng, Angel C.Y. Mak, Scott Hunstman, Donglei Hu, et al. 2017. “Identification of a novel locus associated with skin colour in African-admixed populations.” *Scientific Reports* 7 (44548): 1–9. https://doi.org/10.1038/srep44548.

Hohl, Diana María, Brenda Bezus, Julia Ratowiecki, and Cecilia Inés Catanesi. 2018. “Genetic and phenotypic variability of iris color in Buenos Aires population.” *Genetics and Molecular Biology* 41 (1): 50–58. https://doi.org/10.1590/1678-4685-gmb-2017-0175.

Hollard, C., C. Keyser, T. Delabarde, A. Gonzalez, C. Vilela Lamego, V. Zvénigorosky, and B. Ludes. 2017. “Case report on the use of the HID-Ion AmpliSeq^{TM} ancestry panel in a real forensic case.” *International Journal of Legal Medicine* 131 (2): 351–58. https://doi.org/10.1007/s00414-016-1425-1.

Kastelic, Vanja, and Katja Drobnič. 2012. “A single-nucleotide polymorphism (SNP) multiplex system: The association of five SNPs with human eye and hair color in the Slovenian population and comparison using a bayesian network and logistic regression model.” *Croatian Medical Journal* 53 (5): 401–8. https://doi.org/10.3325/cmj.2012.53.401.

Kayser, Manfred. 2015. “Forensic DNA phenotyping: Predicting human appearance from crime scene material for investigative purposes.” *Forensic Science International: Genetics* 18 (1): 33–48. https://doi.org/10.1016/j.fsigen.2015.02.003.

———, and Peter M. Schneider. 2009. “DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations.” *Forensic Science International: Genetics* 3 (3): 154–61. https://doi.org/10.1016/j.fsigen.2009.01.012.

Kidd, Kenneth K., Andrew J. Pakstis, William C. Speed, Elena L. Grigorenko, Sylvester L.B. Kajuna, Nganyirwa J. Karoma, Selemani Kungulilo, et al. 2006. “Developing a SNP panel for forensic identification of individuals.” *Forensic Science International* 164: 20–32. https://doi.org/10.1016/j.forsciint.2005.11.017.

Koops, Bert-Jaap, and Maurice H.M. Schellekens. 2008. “Forensic DNA phenotyping: Regulatory issues.” *Science and Technology Law Review* 9: 158–202. https://doi.org/10.2139/ssrn.975032.

Kosoy, Roman, Rami Nassir, Chao Tian, Phoebe A. White, Lesley M. Butler, Gabriel Silva, Rick Kittles, et al. 2009. “Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America.” *Human Mutation* 30 (1): 69–78. https://doi.org/10.1002/humu.20822.

Lamason, Rebecca L., Manzoor Ali P.K. Mohideen, Jason R. Mest, Andrew C. Wong, Heather L. Norton, Michele C. Aros, Michael J. Jurynec, et al. 2005. “Genetics: *SLC24A5*, a putative cation exchanger, affects pigmentation in zebrafish and humans.” *Science* 310 (5755): 1782–86. https://doi.org/10.1126/science.1116238.

Lasisi, Tina, and Mark D. Shriver. 2018. “Focus on African diversity confirms complexity of skin pigmentation genetics.” *Genome Biology* 19 (13): 1–3. https://doi.org/10.1186/s13059-018-1395-3.

Lao, O., J. M. De Gruijter, K. Van Duijn, A. Navarro, and M. Kayser. 2007. “Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms.” *Annals of Human Genetics* 71 (3): 354–69. https://doi.org/10.1111/j.1469-1809.2006.00341.x.

Lin, Jennifer Y., and David E. Fisher. 2007. “Melanocyte biology and skin pigmentation.” *Nature* 445 (7130): 843–50. https://doi.org/10.1038/nature05660.

Liu, Fan, Kate van Duijn, Johannes R. Vingerling, Albert Hofman, André G. Uitterlinden, A. Cecile J.W. Janssens, and Manfred Kayser. 2009. “Eye color and the prediction of complex phenotypes from genotypes.” *Current Biology* 19 (5): 192–93. https://doi.org/10.1016/j.cub.2009.01.027.

Lloyd-Jones, Luke R., Matthew R. Robinson, Gerhard Moser, Jian Zeng, Sandra Beleza, Gregory S. Barsh, Hua Tang, and Peter M. Visscher. 2017. “Inference on the genetic basis of eye and skin color in an admixed population via bayesian linear mixed models.” *Genetics* 206: 1113–26. https://doi.org/10.1534/genetics.116.193383.

M’Charek, Amade. 2008. “Silent witness, articulate collective: DNA evidence and the inference of visible traits.” *Bioethics* 22 (9): 519–28. https://doi.org/10.1111/j.1467-8519.2008.00699.x.

MacLean, Charles, and Adam Lamparello. 2014. “Forensic DNA phenotyping in criminal investigations and criminal courts: Assessing and mitigating the dilemmas inherent in

the science." *Recent Advances in DNA and Gene Sequences* 8 (2): 104–12. https://doi.org/10.2139/ssrn.2560157.

Majewski, Sara, Chantelle Carneiro, Erin Ibler, Peter Boor, Gary Tran, Mary C Martini, Salvatore Di Loro, Alfred W. Rademaker, Dennis P. West, and Beatrice Nardone. 2016. "Digital dermoscopy to determine skin melanin index as an objective indicator of skin pigmentation." *Journal of Surgical Dermatology* 1 (1): 37–42. https://doi.org/10.18282/jsd.v1.i1.15.

Maroñas, Olalla, Chris Phillips, Jens Söchtig, Antonio Gomez-Tato, Raquel Cruz, José Alvarez-Dios, María Casares De Cal, et al. 2014. "Development of a forensic skin colour predictive test." *Forensic Science International: Genetics* 13: 34–44. https://doi.org/10.1016/j.fsigen.2014.06.017.

———, J. Söchtig, Y. Ruiz, C. Phillips, Carracedo, and M. V. Lareu. 2015. "The genetics of skin, hair, and eye color variation and its relevance to forensic pigmentation predictive tests." *Forensic Science Review* 27 (1): 13–40.

Martin, Alicia R, Meng Lin, Julie M Granka, Justin W Myrick, Xiaomin Liu, Alexandra Sockell, Elizabeth G Atkinson, et al. 2017. "An unexpectedly complex architecture for skin pigmentation in Africans." *Cell* 171 (6): 1340–53. https://doi.org/10.1016/j.cell.2017.11.015.An.

Miller, Craig T, Sandra Beleza, Alex A Pollen, Dolph Schluter, Rick A Kittles, D Mark, and David M Kingsley. 2007. "Evolution of pigmentation in sticklebacks and humans." *Cell* 131 (6): 1179–89. https://doi.org/10.1016/j.cell.2007.10.055.Cis.

Naing, L, T Winn, and B.N Rusli. 2006. "Practical issues in calculating the sample size for prevalence studies." *Archives of Orofacial Sciences* 1: 9–14.

Norton, Heather L., Rick A. Kittles, Esteban Parra, Paul McKeigue, Xianyun Mao, Keith Cheng, Victor A. Canfield, Daniel G. Bradley, Brian McEvoy, and Mark D. Shriver. 2007. "Genetic evidence for the convergent evolution of light skin in Europeans and East Asians." *Molecular Biology and Evolution* 24 (3): 710–22. https://doi.org/10.1093/molbev/msl203.

Ossorio, Pilar N. 2006. "About face: Forensic genetic testing for race and visible traits." *The Journal of Law, Medicine and Ethics* 34 (2): 277–92. https://doi.org/doi:

10.1111/j.1748-720x.2006.00033.x.

Pakstis, Andrew. J., William C. Speed, Judith R. Kidd, and Kenneth K. Kidd. 2008. “SNPs for individual identification.” *Forensic Science International: Genetics Supplement Series* 1 (1): 479–81. https://doi.org/10.1016/j.fsigss.2007.10.200.

———, Rixun Fang, Fiona C.L. Hyland and Manohar R. Furtado.2010. “SNPs for a universal individual identification panel.” *Human Genetics* 127 (3): 315–24. https://doi.org/10.1007/s00439-009-0771-1.

Park, Jung Hun, and Mu Hyoung Lee. 2005. “A study of skin color by melanin index according to site, gestational age, birth weight and season of birth in Korean neonates.” *Journal of Korean Medical Science* 20 (1): 105–8. https://doi.org/10.3346/jkms.2005.20.1.105.

Parra, E. J., R. A. Kittles, and M. D. Shriver. 2004. “Implications of correlations between skin color and genetic ancestry for biomedical research.” *Nature Genetics* 36 (11): 54–60. https://doi.org/10.1038/ng1440.

Parsons, Thomas J., Rene M.L. Huel, Zlatan Bajunović, and Adnan Rizvić. 2019. “Large scale DNA identification: The ICMP experience.” *Forensic Science International: Genetics* 38: 236–44. https://doi.org/10.1016/j.fsigen.2018.11.008.

Phillips, Christopher. 2015. “Forensic genetic analysis of bio-geographical ancestry.” *Forensic Science International: Genetics* 18: 49–65. https://doi.org/10.1016/j.fsigen.2015.05.012.

———, A. Freire Aradas, A. K. Kriegel, M. Fondevila, O. Bulbul, C. Santos, F. Serrulla Rech, et al. 2013. “Eurasiaplex: A forensic SNP assay for differentiating European and South Asian ancestries.” *Forensic Science International: Genetics* 7 (3): 359–66. https://doi.org/10.1016/j.fsigen.2013.02.010.

———, A. Salas, J. J. Sánchez, D. Ballard, A. Gómez-Tato, J. Álvarez-Dios, M. Calaza, et al. 2007. “Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs.” *Forensic Science International: Genetics* 1 (3): 273–80. https://doi.org/10.1016/j.fsigen.2007.06.008.

———, Lourdes Prieto, Marta Montesino, Antonio Salas, Antonio Gómez-Tato, José

Álvarez-Dios, Antonio Alonso, et al. 2009. “Ancestry analysis in the 11-M Madrid bomb attack investigation.” *PLoS ONE* 4 (8): 1–10. https://doi.org/10.1371/journal.pone.0006583.

Pneuman, Amanda, Zoran M. Budimlija, Theresa Caragine, Mechthild Prinz, and Elisa Wurmbach. 2012. “Verification of eye and skin color predictors in various populations.” *Legal Medicine* 14 (2): 78–83. https://doi.org/10.1016/j.legalmed.2011.12.005.

Posel, Deborah. 2001. “What’s in a name ? Racial categorisations under Apartheid and their afterlife.” *Transformation*, 59–82.

Prestes, P. R., R. J. Mitchell, R. Daniel, K. N. Ballantyne, and R.A.H. van Oorschot. 2011. “Evaluation of the IrisPlex System in admixed individuals.” *Forensic Science International: Genetics Supplement Series* 3 (1): e283–84. https://doi.org/10.1016/j.fsigss.2011.08.137.

Puente, M. de la, C. Santos, M. Fondevila, L. Manzo, The EUROFORGEN-NoE Consortium, M. V. Lareu, and C. Phillips. 2016. “The global AIMs nano set: A 31-Plex SNaPshot assay of ancestry-informative SNPs.” *Forensic Science International: Genetics* 22: 81–88. https://doi.org/10.1016/j.fsigen.2016.01.015.

Purps, J., M. Geppert, M. Nagy, and L. Roewer. 2011. “Evaluation of the IrisPlex eye colour prediction tool in a German population sample.” *Forensic Science International: Genetics Supplement Series* 3 (1): e202–3. https://doi.org/10.1016/j.fsigss.2011.08.101.

Quillen, Ellen E., Heather L. Norton, Esteban J. Parra, Frida Lona-Durazo, Khai C. Ang, Florin Mircea Illiescu, Laurel N. Pearson, et al. 2019. “Shades of complexity: New perspectives on the evolution and genetic architecture of human skin.” *American Journal of Physical Anthropology* 168 (S67): 4–26. https://doi.org/10.1002/ajpa.23737.

Rees, Jonathan L. 2003. “Genetics of hair and skin color.” *Annual Review of Genetics* 37 (1): 67–90. https://doi.org/10.1146/annurev.genet.37.110801.143233.

Reid, Kate M., Lorna J. Martin, and Laura J. Heathfield. 2019. “Evaluation of DNA profiles obtained from deceased individuals at Salt River mortuary (South Africa).” *Australian Journal of Forensic Sciences* 51 (1): 48–51. https://doi.org/10.1080/00450618.2019.1569149.

Roberts, Wendy E. 2009. “Skin type classification systems old and new.” *Dermatologic*

Clinics 27 (4): 529–33. https://doi.org/10.1016/j.det.2009.08.006.

Romanini, Carola, Magdalena Romero, Mercedes Salado Puerto, Laura Catelli, Christopher Phillips, Rui Pereira, Leonor Gusmão, and Carlos Vullo. 2015. "Ancestry informative markers: Inference of ancestry in aged bone samples using an autosomal AIM-indel multiplex." *Forensic Science International: Genetics* 16: 58–63. https://doi.org/10.1016/j.fsigen.2014.11.025.

Royal, Charmaine D., John Novembre, Stephanie M. Fullerton, David B. Goldstein, Jeffrey C. Long, Michael J. Bamshad, and Andrew G. Clark. 2010. "Inferring genetic ancestry: Opportunities, challenges, and implications." *American Journal of Human Genetics* 86 (5): 661–73. https://doi.org/10.1016/j.ajhg.2010.03.011.

Russell, Richard, Pawan Sinha, Irving Biederman, and Marissa Nederhouser. 2006. "Is pigmentation important for face recognition? Evidence from contrast negation." *Perception* 35 (6): 749–59. https://doi.org/10.1068/p5490.

Sachdeva, Silonie. 2009. "Fitzpatrick skin typing: Applications in dermatology." *Indian Journal of Dermatology, Venereology and Leprology* 75 (1): 93–96.

Samuel, Gabrielle, and Barbara Prainsack. 2018. "Forensic DNA phenotyping in Europe: Views 'on the ground' from those who have a professional stake in the technology." *New Genetics and Society* 37: 1–23. https://doi.org/10.1080/14636778.2018.1549984.

Santos, Carla, Christopher Phillips, Manuel Fondevila, Runa Daniel, Roland A.H. Van Oorschot, Esteban G. Burchard, Moses S. Schanfield, et al. 2016. "Pacifiplex: An ancestry-informative SNP panel centred on Australia and the Pacific Region." *Forensic Science International: Genetics* 20: 71–80. https://doi.org/10.1016/j.fsigen.2015.10.003.

Seo, Seung Bum, Jonathan L. King, David H. Warshauer, Carey P. Davis, Jianye Ge, and Bruce Budowle. 2013. "Single nucleotide polymorphism typing with massively parallel sequencing for human identification." *International Journal of Legal Medicine* 127 (6): 1079–86. https://doi.org/10.1007/s00414-013-0879-7.

Shriver, Mark D., and Rick A. Kittles. 2004. "Genetic ancestry and the search for personalized genetic histories." *Nature Reviews Genetics* 5 (8): 611–18. https://doi.org/10.1038/nrg1405.

———, Esteban J. Parra, Sonia Dios, Carolina Bonilla, Heather Norton, Celina Jovel, Carrie

Pfaff, et al. 2003. “Skin pigmentation, biogeographical ancestry and admixture mapping.” *Human Genetics* 112 (4): 387–99. https://doi.org/10.1007/s00439-002-0896-y.

Singh, Subhash Chandra. 2011. “DNA profiling and the forensic use of DNA evidence in criminal proceedings.” *Journal of Indian Law Institute* 53 (2): 195–226.

Slabbert, Nandi. 2017. “The assessment of molecular markers for skin colour determination in the South African population.” MPhil diss., University of Cape Town.

———, and Laura Jane Heathfield. 2018. “Ethical, legal and social implications of forensic molecular phenotyping in South Africa.” *Developing World Bioethics* 18 (2): 171–81. https://doi.org/10.1111/dewb.12194.

Smith, Marcus, and Monique Mann. 2015. “Recent developments in DNA evidence.” *Trends and Issues in Crime and Criminal Justice* 1 (506): 1–7.

Sobiah, R, RH Syeda, E Zunaira, Z Nageen, K Maria, AZ Syeda, SM Shahana, M Akifa, J Abdul, and RK Muhammad. 2018. “Implications of targeted next generation sequencing in forensic science.” *Journal of Forensic Research* 09 (02): 1–8. https://doi.org/10.4172/2157-7145.1000416.

Sobrino, Beatriz, María Brión, and Angel Carracedo. 2005. “SNPs in forensic genetics: A review on SNP typing methodologies.” *Forensic Science International* 154 (2–3): 181–94. https://doi.org/10.1016/j.forsciint.2004.10.020.

Soejima, Mikiko, and Yoshiro Koda. 2007. “Population differences of two coding SNPs in pigmentation-related genes *SLC24A5* and *SLC45A2*.” *International Journal of Legal Medicine* 121 (1): 36–39. https://doi.org/10.1007/s00414-006-0112-z.

Spichenok, Olga, Zoran M. Budimlija, Adele A. Mitchell, Andreas Jenny, Lejla Kovacevic, Damir Marjanovic, Theresa Caragine, Mechthild Prinz, and Elisa Wurmbach. 2011. “Prediction of eye and skin color in diverse populations using seven SNPs.” *Forensic Science International: Genetics* 5 (5): 472–78. https://doi.org/10.1016/j.fsigen.2010.10.005.

Statistics South Africa. 2018 *Statistical release: Mid-year population estimates*. Pretoria: Government Printer. P0302:1–2.

Stokowski, Renee P., P.V. Krishna Pant, Tony Dadd, Amelia Fereday, David A. Hinds, Carl

Jarman, Wendy Filsell, et al. 2007. “A Genomewide association study of skin pigmentation in a South Asian population.” *The American Journal of Human Genetics* 81 (6): 1119–32. https://doi.org/10.1086/522235.

Sturm, Richard A. 2006. “A golden age of human pigmentation genetics.” *Trends in Genetics* 22 (9): 464–68. https://doi.org/10.1016/j.tig.2006.06.010.

———. 2009. “Molecular genetics of human pigmentation diversity.” *Human Molecular Genetics* 18 (R1): 9–17. https://doi.org/10.1093/hmg/ddp003.

———, Neil F. Box, and Michele Ramsay. 1998. “Human pigmentation genetics: The difference is only skin deep.” *BioEssays* 20 (9): 712–21. https://doi.org/10.1002/(SICI)1521-1878(199809)20:9<712::AID-BIES4>3.0.CO;2-I.

———, Rohan D. Teasdale, and Neil F. Box. 2001. “Human pigmentation genes: Identification, structure and consequences of polymorphic variation.” *Gene* 277 (1–2): 49–62. https://doi.org/10.1016/S0378-1119(01)00694-1.

Sulem, Patrick, Daniel F. Gudbjartsson, Simon N. Stacey, Agnar Helgason, Thorunn Rafnar, Kristinn P. Magnusson, Andrei Manolescu, et al. 2007. “Genetic determinants of hair, eye and skin pigmentation in Europeans.” *Nature Genetics* 39 (12): 1443–52. https://doi.org/10.1038/ng.2007.13.

Sun, Qifan, Li Jiang, Guangfeng Zhang, Jing Liu, Lei Zhao, Wenting Zhao, and Caixia Li. 2017. “Twenty-seven continental ancestry-informative SNP analysis of bone remains to resolve a forensic case.” *Forensic Sciences Research* 1790: 1–3. https://doi.org/10.1080/20961790.2017.1306431.

Swiatoniowski, Anna K., Ellen E. Quillen, Mark D. Shriver, and Nina G. Jablonski. 2013. “Technical note: Comparing von Luschan skin color tiles and modern spectrophotometry for measuring human skin pigmentation.” *American Journal of Physical Anthropology* 151 (2): 325–30. https://doi.org/10.1002/ajpa.22274.

The 1000 Genomes Project Consortium. 2015. “A global reference for human genetic variation.” *Nature* 526 (7571): 68–74. https://doi.org/10.1038/nature15393.

Tian, Chao, David A. Hinds, Russell Shigeta, Rick Kittles, Dennis G. Ballinger, and Michael F. Seldin. 2006. “A genomewide single-nucleotide–polymorphism panel with high ancestry information for African American admixture mapping.” *The American Journal*

of Human Genetics 79 (4): 640–49. https://doi.org/10.1086/507954.

Tishkoff, Sarah A., Floyd A. Reed, Françoise R. Friedlaender, Christopher Ehret, Alessia Ranciaro, Alain Froment, Jibril B. Hirbo, et al. 2009. "The genetic structure and history of Africans and African Americans." *Science* 324 (5930): 1035–44. https://doi.org/10.1126/science.1172257.

Toom, Victor, Matthias Wienroth, Amade M'Charek, Barbara Prainsack, Robin Williams, Troy Duster, Torsten Heinemann, Corinna Kruse, Helena Machado, and Erin Murphy. 2016. "Approaching ethical, legal and social issues of emerging forensic DNA phenotyping (FDP) technologies comprehensively: Reply to 'forensic DNA phenotyping: Predicting human appearance from crime scene material for investigative purposes' by Manfred Kayser." *Forensic Science International: Genetics* 22: e1–4. https://doi.org/10.1016/j.fsigen.2016.01.010.

Treesirichod, Arucha, Somboon Chansakulporn, and Pattra Wattanapan. 2014. "Correlation between skin color evaluation by skin color scale chart and narrowband reflectance spectrophotometer." *Indian Journal of Dermatology* 59 (4): 339–42. https://doi.org/10.4103/0019-5154.135476.

Tsetskhladze, Zurab R., Victor A. Canfield, Khai C. Ang, Steven M. Wentzel, Katherine P. Reid, Arthur S. Berg, Stephen L. Johnson, Koichi Kawakami, and Keith C. Cheng. 2012. "Functional assessment of human coding mutations affecting skin pigmentation using zebrafish." *PLoS ONE*. https://doi.org/10.1371/journal.pone.0047398.

Tully, Gillian. 2007. "Genotype versus phenotype: Human pigmentation." *Forensic Science International: Genetics* 1 (2): 105–10. https://doi.org/10.1016/j.fsigen.2007.01.005.

Valenzuela, Robert K., Miquia S. Henderson, Monica H. Walsh, Nanibaa A. Garrison, Jessica T. Kelch, Orit Cohen-Barak, Drew T. Erickson, et al. 2010. "Predicting phenotype from genotype: Normal pigmentation." *Journal of Forensic Sciences* 55 (2): 315–22. https://doi.org/10.1111/j.1556-4029.2009.01317.x.

Valle-Silva, Guilherme do, Flávia Djenane Nunes de Souza, Letícia Marcorin, Alison Luis Eburneo Pereira, Thássia Mayra Telles Carratto, Guilherme Debortoli, Maria Luiza Guimarães de Oliveira, et al. 2019. "Applicability of the SNPforID 52-plex panel for human identification and ancestry evaluation in a Brazilian population sample by next-generation sequencing." *Forensic Science International: Genetics* 40 (1): 201–9.

https://doi.org/10.1016/j.fsigen.2019.03.003.

Van der Wal, Martijn, Monica Bloemen, Pauline Verhaegen, Wim Tuinebreijer, Henrica De Vet, Paul Van Zuijlen, and Esther Middelkoop. 2013. “Objective color measurements: Clinimetric performance of three devices on normal skin and scar tissue.” *Journal of Burn Care and Research* 34 (3): 187–94. https://doi.org/10.1097/BCR.0b013e318264bf7d.

Van Oorschot, Ronald. A.H., and Maxwell K. Jones. 1997. “DNA fingerprints from fingerprints.” *Nature* 387: 767.

Walsh, Simon J. 2005. “Legal perceptions of forensic DNA profiling.” *Forensic Science International* 155 (1): 51–60. https://doi.org/10.1016/j.forsciint.2004.11.001.

Walsh, Susan, Manfred Kayser, Fan Liu, Kaye N. Ballantyne, Mannis Van Oven, and Oscar Lao. 2011a. “IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information.” *Forensic Science International: Genetics* 5 (3): 170–80. https://doi.org/10.1016/j.fsigen.2010.02.004.

———, Alexander Lindenbergh, Sofia B. Zuniga, Titia Sijen, Peter De Knijff, Manfred Kayser, and Kaye N. Ballantyne. 2011b. “Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence.” *Forensic Science International: Genetics* 5 (5): 464–71. https://doi.org/10.1016/j.fsigen.2010.09.008.

———, Agnieszka Kosiniak-Kamysz, Manfred Kayser, Wojciech Branicki, Leda Kovatsi, Arwin Ralf, Andreas Wollstein, and Fan Liu. 2012. “The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA.” *Forensic Science International: Genetics* 7 (1): 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005.

———, Lakshmi Chaitanya, Lindy Clarisse, Laura Wirken, Jolanta Draus-Barini, Leda Kovatsi, Hitoshi Maeda, et al. 2014. “Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage.” *Forensic Science International: Genetics* 9 (1): 150–61. https://doi.org/10.1016/j.fsigen.2013.12.006.

———, Lakshmi Chaitanya, Krystal Breslin, Charanya Muralidharan, Agnieszka Bronikowska, Ewelina Pospiech, Julia Koller, et al. 2017. “Global skin colour prediction

from DNA." *Human Genetics* 136 (7): 847–63. https://doi.org/10.1007/s00439-017-1808-5.

Warshauer, David H., Carey P. Davis, Cydne Holt, Yonmee Han, Paulina Walichiewicz, Tom Richardson, Kathryn Stephens, Anne Jager, Jonathan King, and Bruce Budowle. 2014. "Massively parallel sequencing of forensically relevant single nucleotide polymorphisms using TruSeq™ forensic amplicon." *International Journal of Legal Medicine* 129 (1): 31–36. https://doi.org/10.1007/s00414-014-1108-8.

Wet, S de, and J Visser. 2017. "DNA profiling and the law in South Africa." *Potchefstroom Electronic Law Journal* 14 (4): 170–206. https://doi.org/10.17159/1727-3781/2011/v14i4a2587.

Wienroth, Matthias, Niels Morling, and Robin Williams. 2014. "Technological innovations in forensic genetics: Social, legal and ethical aspects." *Recent Advances in DNA and Gene Sequences* 8 (2): 98–103. https://doi.org/10.2174/2352092209666150328010557.

Williams, R., and M. Wienroth. 2017. "Social and ethical aspects of forensic genetics: A critical review." *Forensic Science Review* 29 (2): 145–69.

Wit, Erika de, Wayne Delport, Chimusa E. Rugamika, Ayton Meintjes, Marlo Möller, Paul D. Van Helden, Cathal Seoighe, and Eileen G. Hoal. 2010. "Genome-wide analysis of the structure of the South African Coloured population in the Western Cape." *Human Genetics* 128 (2): 145–53. https://doi.org/10.1007/s00439-010-0836-1.

Wright, C. Y., R. M. Lucas, T. Kapwata, Z. Kunene, and J. L. du Plessis. 2019. "Towards a reliable, non-invasive melanin assessment for pigmented skin." *Skin Research and Technology* 25: 100–102. https://doi.org/10.1111/srt.12592.

Yang, Yaran, Bingbing Xie, and Jiangwei Yan. 2014. "Application of next-generation sequencing technology in forensic science." *Genomics, Proteomics and Bioinformatics* 12 (5): 190–97. https://doi.org/10.1016/j.gpb.2014.09.001.

Zaorska, Katarzyna, Piotr Zawierucha, and Michał Nowicki. 2019. "Prediction of skin color, tanning and freckling from DNA in Polish population: Linear regression, random forest and neural network approaches." *Human Genetics* 138 (6): 635–47. https://doi.org/10.1007/s00439-019-02012-w.

Appendices

Appendix A: Sample estimation formula for *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654)

The sample size required for this study was calculated using the following sample size estimation formula (Naing, Winn, and Rusli 2006).

$$n = \frac{p(1-p)Z^2}{d^2}$$

n = sample

p = anticipated proportion (reported SNP frequency)

d = precision (0.05)

Z = statistic for confidence level (95 %)

***SLC45A2* (rs16891982):**

The reported SNP frequency was 0.6403 (SNPedia.com, modified 5 December 2018). The suggested sample size required is 354 individuals.

***SLC24A5* (rs1426654):**

The reported SNP frequency was 0.2283 (SNPedia.com, modified 6 December 2018). The suggested sample size required is 271 individuals.

Therefore, 354 individuals need to partake in this study in order to account for the sample size estimation in both variants.

Appendix B: Participant consent form for the use of their DNA

Participant consent form

Participant number: _____

1. I confirm that I am South African and 18 years or older
2. I confirm have read and understand the research information form for the above study and have had the opportunity to ask questions.
3. I agree to participate in this study and understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.
4. I consent to:

   (Tick the appropriate box)

☐ The use of my DNA and information to be used for this study AS WELL AS my samples to be stored and used for ANY future research only if approved by the HREC.

☐ The use of my DNA and information to be used for this study AND stored for the ONLY purpose of possible future continuation of this specific research only if approved by HREC.

☐ The use of my DNA and information to be used ONLY for this study and to be destroyed after conclusion of this project.

_____

Printed Name of Subject

_____ _____

Signature of Subject Date

_____ _____

Appendix C: Ethical approval and amendment letter from UCT HREC

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room E52-24 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6338 • **Facsimile** [021] 406 6411
Email: shuretta.thomas@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

08 July 2015

HREC REF: 317/2015

Ms L Heathfield
Division of Forensic Medicine
Level 5, entrance 2
Falmouth Building

Dear Ms Heathfield

PROJECT TITLE: THE ASSESSMENT OF MOLECULAR MARKERS FOR SKIN COLOUR IN SOUTH AFRICANS (MPhil-candidate-N Slabbert)

Thank you for a very considered response to the Faculty of Health Sciences Human Research Ethics Committee dated 1 July 2015.

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study.

Approval is granted for one year until the 30th July 2016.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

Please quote the HREC REF in all your correspondence.

We acknowledge that the student, Nandi Slabbert will also be involved in this study.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938
This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP), South African Good Clinical Practice Guidelines (DoH

Appendix C (continued)

UNIVERSITY OF CAPE TOWN

HUMAN RESEARCH ETHICS COMMITTEE
25 FEB 2019
HEALTH SCIENCES FACULTY
UNIVERSITY OF CAPE TOWN

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

Form FHS007: Amendment – study staff

HREC office use only (FWA00001637; IRB00001938)			
☐ Approved			
This serves as notification that all changes to the study staff and documentation described below are approved.			
Chairperson of the HREC signature	signature removed	Date	26/2/19

Principal Investigator to complete the following:

1. Protocol information

Date (when submitting this form)	25 February 2019	
HREC REF Number	317/2015	
Protocol title	The assessment of molecular markers for skin colour in South Africans	
Protocol number (if applicable)		
Principal Investigator	Laura Heathfield	
Department / Office Internal Mail Address	Reception, Division of Forensic Medicine and Toxicology, Falmouth Building (entrance 3, level 1), Faculty of Health Science, UCT, Anzio Road, Observatory	
1.1 Does this protocol receive US Federal funding?	☐ Yes	☐✓ No

2.1 Staff changes (tick ✓)

Are new personnel being added to this research?	☐✓ Yes	☐ No
Are current personnel being removed from this research?	☐✓ Yes	☐ No
Is the principal investigator for this research being changed? If yes, please attach revised conflict of interest and PI declaration statements. (Refer: sections 7 and 8.3 in the New Protocol Application Form - FHS013)	☐ Yes	☐✓ No
Do the consent and assent forms need modification to reflect these staff changes? If yes, please attach copies of the revised forms, with all changes highlighted or tracked and listed in the documents for approval.	☐✓ Yes	☐ No

Appendix C (continued)

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

2.2 Amended study staff details

Title, first name, surname	Department/Division	E-mail	Role of new staff member
Akshay Vanmali	Pathology | Forensic Medicine and Toxicology	vmlaks001@myuct.ac.za	Co-investigator
Calvin Mole	Pathology | Forensic Medicine and Toxicology	Calvin.mole@uct.ac.za	Co-investigator

3. List of documentation for approval

Please list below all staff documentation such as CVs, declarations, GCP certificates and revised consent forms which need approval. This information must correspond to all 'yes' answers in 2.1 above. This form will be signed and returned to the PI as notification of approval. Please add extra pages if necessary.

Addition of staff:
CV from Akshay attached
CV from Calvin Mole has been provided previously

Removal of staff:
Dr Karen Shires – she is no longer involved in the project. She withdrew from this project.

Revised consent form attached.

4. Signature

My signature certifies that I will maintain the anonymity and/ or confidentiality of information collected in this research. If at any time I want to share or re-use the information for purposes other than those disclosed in the original approval, I will seek further approval from the HREC.

Signature of PI	signature removed	Date	22/02/2019

Appendix D: Research information and participant questionnaire form

Research information and participant questionnaire

Project Title: The assessment of molecular markers for skin colour in South Africans

Affiliation: Division of Forensic Medicine and Toxicology
Faculty of Health Sciences
University of Cape Town

Researcher: Akshay Vanmali (MPhil: Biomedical Forensic Science candidate)

Supervisor: Dr. Laura Heathfield

Co-supervisor: Calvin Mole

What is this research about?

This project aims to investigate if certain DNA markers that have been associated with skin pigmentation internationally have significant discerning value within the South African population. This will be achieved through the evaluation of molecular genotypes and Melanin Index (MI), which is an objective measurement correlating with skin colour, as well as determining if there is any correlation between the results and to self-reported ethnicity.

Where does this research fit in?

We know that very little of the genetic understand of race is meaningful or true. Someone's self-identified racial category and ethnicity may have little to do with their external visible features. External characteristics; however, are closely and reliably connected to an individual's genes. These characteristics include but are not limited to eye, skin and hair colour. This knowledge becomes useful when trying to identify individuals within the field of forensic science, whether it is an offender, victim or missing person.

Current DNA identification practices rely on the comparison of DNA and, thus its use, as evidence is limited if there are no references samples available or no matches contained on the DNA database. This is where many criminal investigations may reach a dead end due to lack of witnesses or leads to investigate potential suspects. Even unidentified remains that can no longer be recognised become problematic when there is no identifiable way of connecting them with a missing individual. These circumstances have given rise to use DNA and its connection with external visible features as possible new investigative tool by means of appearance prediction. This allows for the narrowing of suspect pools and provides visual characteristics that aid facial reconstruction of unknown remains.

The use of this phenotypic prediction from DNA is becoming more common internationally and standardised kits are being made. However, South African is known for its diversity and this may complicate tests if international markers are not significant to our population. Since it is not guaranteed that what works overseas will apply here, South African specific research needs to be done.

What do we require from you as a participant and how will we collect it?

1. Questionnaire – We require participants to fill out an information questionnaire self-reporting their ethnic background.
2. Melanin Index –MI will be measured at three hairless sites namely, the inner forearm, the inner arm above the elbow and the forehead. This will be done using a DSM II ColorMeter (CyberDerm Inc., USA), and will be recorded on participant questionnaire.
3. DNA sample - Buccal cells will be collected from each volunteer, via a saliva sample. Saliva will be obtained through means of a 10 mL 0.9 % saline solution mouth rinse collected in a 50 mL tube (labelled with participant number) and stored on ice until processed. Volunteers will take these samples themselves under supervision.

Are there any risks?

Minimal risk is associated to this procedure; however, the derma-spectrophotometer is a non-invasive method of reading skin colour and lens will be sterilized between each measurement.

Will information be confidential?

Yes, your personal information will be kept secure and private. Your consent form provides a unique participation number, which will be used to identify all your samples and data from that point forward and therefore no individual name will be linked to specific samples.

Voluntary participation and withdrawal issues

Participation is completely voluntary, and your allocated participation number will ensure tractability of your sample and information, and therefore if you wish to withdraw at any time stage of the project you may do so without providing a reason and your sample and any data relating to it will located and discarded. A total of ~400 volunteers are required (300 volunteers already recruited) for participation and recruitment will be carried out independent of race, but rather on ranging MI scores to allow a representation of the entire spectrum.

Any questions?

If you have any question or queries relating to the research and/or questionnaire and procedures, you are more than welcome ask or contact me at vmlaks001@myuct.ac.za.

Participant questionnaire

Participant reference number: _____

1. Sex ☐ Male ☐ Female

2. Age ☐ 18 – 39 ☐ 40 – 60 ☐ 60+

3. Self-reported population group according to South African population census groups

☐African Black ☐Coloured ☐Indian/Asian ☐White

4. Ancestral origin (if known)

☐ European – N W S E

☐Africa – N W S E

☐Asian – N W S E

☐ Middle Eastern

☐ Other

Specify: _____

5. Your self-reported ethnicity (In reference to cultural self-identification)

_____

6. Father's ethnicity and/or ancestral origin

_____

7. Mother's ethnicity and/or ancestral origin

_____

8. Parental (father's side) grandparent's ethnicity and/or ancestral origin

Grandfather: _____

Grandmother: _____

9. Maternal (mother's side) grandparent's ethnicity and/or ancestral origin

Grandfather: _____

Grandmother: _____

10. Recorded Melanin Index

Right inner forearm: _____

Right inner arm above elbow _____

Left inner forearm: _____

Left inner arm above elbow: _____

Calculated Mean: _____

Forehead reading**:** _____

Appendix E: Summary statistics of the MI readings for SA population census and ancestral groups

Table E.1 Summary statistics of the MI readings taken from the four SA population census groups

SA population census group	MI readings			
	1st Quartile	Median	3rd Quartile	Range
African Black	52.48	57.41	64.27	36.93 – 86.19
Coloured	37.45	41.60	44.40	30.00 – 60.41
Indian/ Asian	37.82	43.62	48.39	30.30 – 61.53
White	30.15	31.96	33.52	25.46 – 39.88

Table E.2 Summary statistics of the MI readings taken from the ancestral groups

Ancestral groups	MI readings			
	1st Quartile	Median	3rd Quartile	Range
African	49.61	55.76	63.49	31.10 – 83.79
Asian	38.27	43.45	49.14	30.30 – 61.53
European	30.16	32.12	33.76	25.46 – 44.60
Mixed	37.01	40.26	43.91	30.00 – 54.19

Appendix F: Statistical analysis of the MI distribution in SA population census and ancestral groups

Table F.1 Kruskal-Wallis and Wilcoxon rank-sum test(s) performed on the MI distribution for the SA population census groups. *The p-values are adjusted according to the Bonferroni correction.*

Kruskal-Wallis one-way analysis of variance test				
Kruskal-Wallis chi-squared = 299.86				
df = 3	p-value < 2.2e-16			
Wilcoxon rank-sum test				
	African Black	**Coloured**	**Indian/Asian**	**White**
African Black	-			
Coloured	p < 1.32e-15 W=10118	-		
Indian/ Asian	p < 1.32e-15 W = 7254	p = 0.25662 W = 1696	-	
White	p < 1.32e-15 W = 16289	p < 1.32e-15 W = 8118.5	p < 1.32e-15 W = 6213	-

Appendix F (continued)

Table F.2 Kruskal-Wallis and Wilcoxon rank-sum test(s) performed on the MI distribution for the ancestral groups. *The p-values are adjusted according to the Bonferroni correction.*

Kruskal-Wallis one-way analysis of variance test				
Kruskal-Wallis chi-squared = 260.9				
df = 4		p-value < 2.2e-16		
Wilcoxon rank-sum test				
	African	**Asian**	**European**	**Mixed**
African	-			
Asian	p < 1.1322e-12 W = 8215	-		
European	p < 1.32e-15 W = 17962	p < 1.32e-15 W = 6658	-	
Mixed	p < 1.32e-15 W = 6888	p < 0.03708 W = 1905	p < 2.844e-15 W = 488	-

Appendix G: Molecular weight marker (Quick-load® Purple 50 bp DNA ladder)

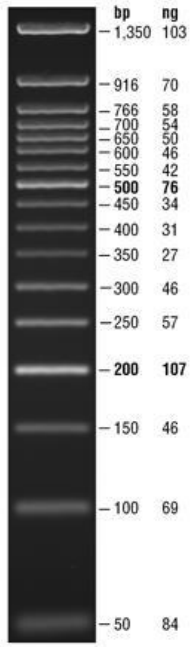

Figure G.1 Image of the MWM (Quick-load® Purple 50 bp DNA ladder, New England BioLabs, Ipswich, USA) that was used on the agarose gels as a reference with its different sizes (bp). *Image source: https://www.neb.com/products/n0473-quick-load-50-bp-dna-ladder*

Appendix H: Genotype combinations for *SLC45A2* (rs16891982) and *SLC24A5* (rs1426554) within the cohort and Pearson's Chi-Squared test for HWE

Table H.1 Genotype combinations for *SLC45A2* (rs16891982) and *SLC24A5* (rs1426554) within the cohort and the Chi-Squared test for HWE

***SLC45A2* (rs16891982)**	***SLC24A5* (rs1426554)**			**Total**
	AA	**GA**	**GG**	
CC	28	61	144	233
GC	25	13	8	46
GG	95	15	0	110
Total	148	89	152	389
Pearson's Chi-Squared Test				
$X^2 = 197.32$	df = 4		p-value < 2.2e-16	