The assessment of molecular markers for skin colour determination in the South African population

Nandi Slabbert

SLBNAN001

Supervisor: Laura Heathfield

Division of Forensic Medicine and Toxicology, Department of Pathology, University of Cape Town

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN

In partial fulfilment of the requirements for the degree

MPhil (Biomedical Forensic Science)

Faculty of Health Sciences, University of Cape Town

Submission: March 2017

Plagiarism declaration

I, **Nandi Slabbert**, 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 nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

Reference Style: Vancouver

Signed by candidate 13/03/2017

Signature Date

Abstract

Conventional forensic deoxyribonucleic acid (DNA) analysis involves a matching principle, which compares DNA profiles from evidential samples to those from reference samples of known origin. In casework, however, the accessibility to a reference sample is not guaranteed, which limits the use of DNA as an investigative tool. This has led to the development of phenotype prediction, which uses single nucleotide polymorphisms (SNPs) analysis to estimate the physical appearance of the sample donor. Physical traits, such as eye, hair and skin colour, have been associated with certain alleles within specific genes involved in the melanogenesis pathways. These genetic markers are also associated with ancestry and their trait prediction ability has mainly been assessed in European and American populations. Therefore, this pilot study aimed to assess two SNPs associated with skin colour (rs1426654 (*SLC24A5*) and rs16891982 (*SLC45A2*)) within a South African (SA) population (n=92), and evaluate any association with (i) SA population groups (Black, Coloured, Indian/Asian, White); (ii) ancestry; and (iii) melanin index (MI) as objectively measured by the DSM II ColorMeter (CyberDerm Inc., USA). These two SNPs were chosen, as they have been consistently associated with MI, even in population groups that exhibit admixture. Results showed that (i) MI was significantly different between some SA population groups; (ii) MI was not significantly different between individuals of Asian and Mixed ancestry; (iii) specific genotypes were significantly different between SA population groups; and (iv) MI was significantly different between individuals exhibiting certain genotype combinations. A general linear model (GLM) was used in an attempt to predict an individual's MI based on genotype. A backward stepwise approach showed that the

best model in this cohort contained limited genotypes and was unable to accurately predict the MI of individuals of African and Mixed ancestry. This could indicate that these SNPs alone do not possess sufficient discriminatory power to be used as a predictive tool in SA; however, verification of results needs to be performed in a larger cohort. Additional SNPs should also be tested to potentially improve the accuracy of the model. This study has nevertheless laid a necessary foundation for additional molecular phenotyping studies in SA. This should be carried out with consideration of the ethical, legal and social implications, which have also been reviewed in this dissertation. Findings propose the use of MI to describe skin colour in a forensic context, rather than SA population groups as a proxy of phenotype prediction.

Acknowledgements

I would like to express my profound gratitude to my supervisor, Laura Heathfield, for her constant support and guidance especially through periods of self-doubt. Her encouragement and belief in me kept me motivated.

I am honoured to have been part of the Division of Forensic Medicine and Toxicology and having the opportunity to develop my knowledge and skills within this prestigious field. I'd like to acknowledge Calvin Mole for his assistance with the statistical analysis of this project and Bronwen Davis for always being available and willing to give advice.

A special thank you to Ruby Bunjun for her insight and support during this emotional process and my fellow classmates, you are the best group of friends I could have wished for.

I want to give my gratitude to Gail Malcomson for all your time spent on proofreading and grammar critique.

Lastly but most important, a big thank you to my parents, without whose emotional and financial support the completion of my thesis would not have been possible.

Table of Contents

SECTION 4. APPENDICES

List of Figures

List of Tables

List of Abbreviations

°C	degrees Celsius
A	adenine
AIC	Akaike information criterion
AIM	ancestry informative marker
bp	base pair
C	cytosine
CSI	Crime Scene Investigation
ddNTP	dideoxynucleotide triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
EVC	external visible characteristic
EXO1	Exonuclease – 1
G	guanine
g	gram
GLM	general linear model
MI	melanin index
MWM	molecular weight marker
NFDD	National Forensic DNA Database

PCR	polymerase chain reaction
RFU	relative fluorescence unit
SA	South Africa
SAP	shrimp alkaline phosphatase
SBE	single base extension
SLC	solute carrier family
SNP	single nucleotide polymorphism
STR	short tandem repeat
Stats SA	Statistics South Africa
T	thymine
UCT	University of Cape Town
US	United States
µl	microliter
µM	micromole

Section 1. **RESEARCH PROTOCOL**

Abstract

Molecular markers associated with skin colour need to be investigated to determine viability and discriminatory power within the SA population. The use of phenotyping as a tool to aid forensic investigation is becoming incorporated internationally and its use within SA needs to be considered. This project proposes to examine the association between skin colour and two markers, which have been identified as potentially problematic within mixed ancestry populations. This will be done through designing and analysing a multiplex SNaPshot genotyping assay and independent pigmentation measurements.

1.1 Introduction

1.1.1 Background

The practice of identifying individuals based on DNA analysis has shown an increase within the international field of forensic science over recent years and its use within the criminal justice system has become a globally accepted norm [1–3]. The method of DNA analysis in these cases are usually based on the principle of matching a DNA profile from an evidentiary sample of unknown origin with a DNA profile from a known reference sample, which can either be that of a suspect or profiles collected within a DNA database [4,5]. Therefore, the usefulness of unknown evidentiary DNA samples for investigation purposes is limited in cases where there is no suspect or when a national DNA database fails to match to any known profiles [5–7]. This has prompted an increased focus regarding research into the prediction of observable physical characteristics of an individual based on the genetic material recovered from a biological sample from a crime scene [5,7].

1.1.2 Forensic DNA profiling

Forensic DNA profiling is a method, which allows the comparison of genetic information, usually between biological evidence collected from the crime scene and potential suspects. This process enables DNA profiles to be used as scientific evidence for the inclusion and exclusion of individuals [3,8,9]. This is based on a select number of specific regions within DNA that have been shown to express a high degree of variation between individuals [8,10]. These regions are known as short tandem repeats (STRs) which fall within non-coding regions of the genome

and thus do not translate to any proteins. STR sequences vary due to the number of times the base pair (bp) motif repeats itself at a locus and can thus be used to discriminate alleles based on size. An STR genotype is therefore the alleles reported according to the number of repeated units observed at a locus and multiple STR genotypes can be combined to form a STR profile that is then used for identification purposes [10].

This DNA analysis process starts with the collection of biological evidentiary samples with an unknown origin. These samples are then sent for laboratory testing where they are subjected to DNA extraction, quantification and amplification. The specific amplified markers are labelled during the incorporation of fluorescently tagged primers to generate a visual representation of the alleles present in the sample. This, as well as statistical analysis to calculate the random match probability of the DNA profile, is carried out with the aid of computer software. This results in a DNA profile that can be entered into a database and compared against other DNA profiles and references [2,3,8–10].

The DNA profiles described above only discriminate between individuals through the variability in the number of STRs, and do not contain any physical or descriptive information other than the sex of the individual. Therefore, if crime scene DNA profiles are not matched to reference samples or profiles from the database they are limited with regards to their use as an investigative tool.

Other DNA targets, which are capable of providing phenotypic characteristics, such as eye, skin and hair colour as well as an inference of ancestry, are now actively being researched and incorporated into standard DNA profiling systems to enable maximum usage of biological evidence gathered at a crime scene [6,7].

1.1.3 SNPs and molecular phenotyping

Phenotypic characteristics are a set of observable physical traits within an organism, which can include its morphology, physiology and other viable traits. These have been shown to be dependent on both environmental and genetic factors [11]. Genetic variations within coding regions of DNA can alter gene expression which can cause phenotypic differentiations. These specific variations in DNA have become markers of interest for forensic identification purposes [10,12].

SNPs are single base sequence variations that occur at particular points in the genome that have been shown to vary between individuals [10]. The mutation rate of forensically relevant SNPs is lower than that compared to forensically relevant STRs and due to the evolution of the human genome, under certain selection pressures; adaptations to the genome on this level has the ability to estimate ancestry and predict phenotypes [13]. This provides the ability to discern between populations through the use of identified ancestry informative markers (AIMs) [14] and individuals with observable different external visible characteristics (EVCs) such as different eye, hair and skin colour [6].

Molecular phenotyping can thus provide the means to predict an individual's visible appearance, which has the potential to aid in forensic cases as an investigation tool [5,6]. This has initiated the research focusing on the development of SNP identity panels [15], such as IrisPlex assays for eye colour prediction [16], and DNAPrint targeting pigmentation genes as AIMs [17].

1.1.4 Molecular phenotyping associated with skin colour

Human skin colour differentiation is mainly dependent on the amount, type and distribution of melanin production by melanocyte cells in skin [18,19]. An objective melanin measurement is able to be made through the use of a derma spectrometer that uses wavelengths to measure the reflectiveness of skin colour and is able to calculate and score the pigmentation on an MI scale [20,21]. The values of human skin colour pigmentation range from 20 to 100 units on the scale representing fair to darker skin tones [22].

The phenotypic trait of skin colour therefore relies on genes involved in pigmentation. Certain polymorphisms occurrences within these genes support the hypothesis of skin lightening being an evolutionary selected mutation [19]. These variations are thus seen as AIMs, enabling the differentiation of possible ancestry [23,24]. Inference of skin colour has been made based on AIMs due to the linkage of ancestry and different pigmentation [19,25]. However, recent research has shown that some SNPs that have been identified as AIMs may in fact be insufficient within admixed populations [22].

1.1.5 SNPs of interest

Two of the identified AIMs discussed above occur within two coding genes, *solute carrier family 24, member 5* (*SLC24A5*) and *solute carrier family 45, member 2* (*SLC45A2*) [22]. The sequence variation rs1426654 (*SLC24A5*) has been shown to influence skin pigmentation and indicate ancestry in the categories of European, African or Asian descent [26]. Rs16891982 (*SLC45A2*) also shows influences regarding skin pigmentation as well as the indication of light skin European ancestry [27]. These SNPs have shown to not provide sufficient discriminatory significance within admixed populations [22] and has thus raised the question as to whether discerning between individuals of diverse populations, as observed in SA, will be possible if these SNPs are incorporated into commercial kits for use in SA.

1.1.6 Issues surrounding molecular phenotyping in SA

During 2014, an amendment within the SA criminal law relating to forensic procedures was passed. This included the formation and regulation of the National Forensic DNA Database of South Africa (NFDD) [28,29], which provides a platform to store DNA profiles of offenders in the database and allow forensic investigators to search the database against samples collected from crime scenes.

The Criminal Law (Forensic Procedures) Amendment Act 37 of 2013 (hereafter referred to as the DNA Act) regulates the information that DNA profiles are allowed to contain. One such regulation is the exclusion of any physical

information (besides the individual's sex) pertaining to the individual. This is to prevent the disclosure and storage of any private or medically sensitive information which could potentially be exploited or used unethically by entities such as insurance or medical aid companies [28]. The imposed legal restriction on what is permitted in a DNA profile is a criticism that has been raised locally as well as internationally in relation to similar international laws. It is seen as a negative factor limiting the use of genetic material as an investigative tool and restricting the growth of its use in the future [1,5,29].

Currently the use of phenotyping DNA for the prediction of an individual's physical appearance is being questioned. Ethical challenges pertaining to the extent to which these phenotypic predictions can be used as well as what possible privacy rights this type of profiling may infringe upon have been brought up in literature [5,6]. The common argument in defence of forensic phenotyping is that characteristics, such as skin, hair and eye colour, are not private information as they are EVCs and are observable to anyone in the public. Therefore, phenotypic predictions provide the same information, if not more than an actual eye witness, due to its objectivity and thus cannot be considered to be revealing private information [6].

An additional challenge that SA faces pertaining to the use of phenotyping, especially regarding the inference of skin colour, is its diverse population. South Africans exhibit a wide variation of ethnical and cultural backgrounds. However, only four categories of classification are used by the Statistics South Africa (Stats

SA) census. These population group categories are Black/African, Indian/Asian, Coloured and White [30]. According to the latest Stats SA census, majority of the population identify as “Black/African” (80.2%), this is the highest proportion in all provinces apart from the Northern Cape and the Western Cape. The “Coloured” population, the highest proportion in the Northern and Western Cape, is comprised nationally of 8.8% while the “White” population makes up 8.4% and Indian/Asian is 2.5% [30].

1.2 Rationale

South Africa consists of a population with a wide range of diversity of skin pigmentation and mixtures of ancestry. Considering that AIMs of an individual may not correctly indicate the individual's EVCs, the viability of these AIMs need to be investigated within the SA context.

With the standardisation of commercial kits for DNA analysis, the possibility of these kits containing markers that are not applicable to the SA context is of great concern. Given the fact that specific markers identified from previous research have displayed insufficient discriminatory power in admixed populations it is of great importance to identify if the same applies to the SA population. For kits to be feasible and valid the markers contained in them must possess sufficient discriminatory power within admixed population, which is commonly the case in SA.

Even though SA law does not allow for this information to be used currently, kits and services predicting phenotypic features are presently being used internationally and changes in law to allow for this are being made [5,7,16]. However, it should be noted that the definition of a DNA profile as stated in the DNA Act [28] refers to the DNA profiles saved on the NFDD and thus does not affect the progression of this research project.

1.3 Aim and objectives

1.3.1 Aim

The aim of this project is to assess the viability and discriminatory power of current molecular phenotyping markers for skin colour within SA.

1.3.2 Objectives

- Design and optimise a multiplex genotype assay for the two identified SNPs (*SLC24A5* rs 1426654 and *SLC45A2* rs 16891982).

- Determine the relationship between the two identified SNPs and the MI taken from an SA population sample.

- Statistically evaluate if there is a significant difference between the different ranges of MI and the polymorphisms present and lastly if this can be applied to self-reported ethnicity and SA population group categories and determine whether MI is a better descriptive measurement than race.

1.4 Proposed methodology

1.4.1 Study paradigm

This project is a cross sectional study with the target group being SA individuals living within the Western Cape. The study is a quantitative analysis measuring the association between the recorded MI (dependent variable) of an individual and the genotype variation (independent) in rs1426654 and rs16891982 in genes *SLC24A5* and *SLC45A2* respectively. These results will also be analysed in relation to the four main nominal categorical groups of the SA population.

1.4.2 Sample size calculations

The following formula was used to calculate the suggested sample size [31].

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

n = sample
p = anticipated proportion (reported SNP frequency)
d = precision (0.05)
Z = statistic for the level of confidence (95%) which corresponds to a value of 1.96

Calculation for rs1426654 in gene *SLC24A5*

SNP frequency = 0.269753 *(taken from SNPedia.com, modified 10 June)*

Therefore, suggested sample size = 302 individuals

Calculation for s16891982 in gene *SLC45A2*

SNP frequency = 0.6917664 *(taken from SNPedia.com, modified 10 June)*

Therefore, suggested sample size = 328 individuals

However, due to the fact that this project is a small pilot study as part of a mini dissertation, the investigators feel that a sample size of 100 individuals is appropriate.

1.4.3 Cohort

The cohort will comprise adult SA individuals (over 18 years old) of all races (n=100) living within the Western Cape. The individuals will be selected based on their MI, together representing a range of values on the MI scale. Therefore, recruitment will be carried out independent of race, but rather on ranging MI scores, n=25 in each quarter (Figure 1.4.3–1). This has been done through the equal division of the MI scale to ensure the collection of samples across the whole range of the scale (broad range of pigmentation tones) for more accurate representation regardless of race.

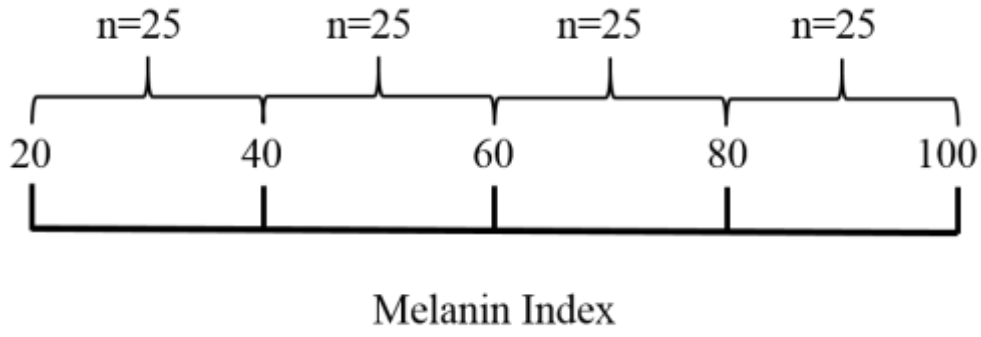

Figure 1.4.3–1 Scale representing MI values and sampling

The investigators will personally approach individuals outside of UCT and ask them if they would like to participate. The study will be explained to volunteers and informed consent will be obtained before samples are taken (Appendix A). Exclusion criteria will consist of individuals with pigment disorders, such as Hyperpigmentation, Melasma and Hypopigmentation (e.g. Albinism), as well as individuals with severe scarring in the areas where MI will be measured [21]. Volunteers reporting the use of any supplements, ointments or treatments designed to alter the skin colour, such as self-tanning or bleaching lotions or tan bedding, will also be excluded [32].

1.4.4 Collection of samples

1.4.4.1 Participant information

Volunteers will be required to fill out a participant questionnaire that will only be linked to a unique participant number (Appendix 1B).

1.4.4.2 MI measurement

Volunteers will have their MI measured using a non-invasive and sensitive derma spectrometer called the DSM II ColorMeter (CyberDerm Inc., USA) [20,21]. This instrument was chosen due to greater sensitivity and feasibility based on a validation study between multiple spectrometers [20].

Two hairless sites on each arm will be measured, namely the inner forearm and the inner arm above the elbow. These sites have been selected based on previous

study protocol and literature indicating it to be a good indication of as basal skin colour [32–34] Measurements will be recorded on the participant questionnaire (Appendix 1B).

1.4.4.3 DNA collection

Buccal cells will be collected from each volunteer, via a saliva sample. Salvia will be obtained through means of a 10ml 0.9% saline solution mouth rinse collected in a 50ml tube (labelled with participant number) and stored on ice until processed. Volunteers have the option to consent to their samples to be stored for future research after the conclusion of this project. These samples will be stored at –80°C within a secure freezer, accessible only to relevant researchers, allocated for long-term sample storage.

1.4.5 Processing of DNA samples

The buccal cells will undergo DNA extraction and total nucleic acid will be quantified using a Nanodrop Spectrophotometer (Thermo Scientific, USA). The two identified SNPs will be amplified using polymerase chain reaction (PCR) with appropriate primers. Primers reported in literature will be verified using bioinformatics tools, such as Primer3 and OligoAnalyser, before ordering. Optimisation during PCR will be looked at in terms of annealing temperature, number of cycles and buffer concentration.

Agarose gel electrophoresis will be run on the post-PCR products along with a molecular weight marker (MWM) for correct amplification confirmation in terms of amplicon size. These results will also be verified through sequencing and alignment to ensure the correct SNP region was amplified. Once this has been confirmed, amplified PCR products will undergo a pre-SNaPshot clean-up for purification. SNaPshot PCR will then be performed via a kit (Applied Biosystems) on the purified post-PCR products in addition with appropriate SNaPshot internal primers.

The next step in the process is the preparation for capillary electrophoresis, which will be run through a genetic analyser. The different emitted fluorescent signals will be interpreted by GeneMapper software to produce an electropherogram for genotyping interpretation. The detection of differences in genotypes will result in certain samples being selected for sequence verification using Sanger sequencing [35]. This technique determines the complete nucleotide base sequence for a specific region of DNA. DNA polymerase elongated the chain through base pair matching. This is initiated from the primer, by incorporating deoxynucleotides (dNTPs) and dideoxynucleosides (ddNTPs), the latter which are chain–terminating versions of the four nucleotides labelled with different coloured dyes [9,35]. Capillary electrophoresis will be performed on a 3130xl Genetic Analyser (Applied Biosystems) and parameters according to those set out by the manufacturer followed. The sequencing verification of correct genome calling will allow selected samples to be incorporated as controls within the multiplex assay.

1.4.6 Analysis of results

Statistical analysis will include tests for deviation from Hardy – Weinberg equilibrium and allele frequency estimated by gene counting. Different frequencies will be compared using the Pearson test and Pearson's product correlation co-efficient will be used to estimate the influence of ancestry in the dataset. ANOVA testing for association of genetic polymorphism with the MI will be performed and a p value of 0.0028 will be considered as statistically significant as reported and used in a previous admixed population study [22]. STATA (StataCorp LP) statistical software, version 11 will be used for calculations.

1.5 Outcome and social value

DNA phenotyping increases the degree to which genetic material can be used in forensic investigations including criminal or missing person identification cases [5]. The accurate prediction of EVCs can help narrow the field of suspects through the positive exclusion of individuals, saving time and financial resources during an investigation. It will also provide a platform to assign identity to unknown victims, unclaimed deceased and human remains beyond recognition at the forensic autopsy stage.

Therefore, the outcome concerning the significance of discriminatory power of these markers within the SA population is important. This will allow for accurate validation of the potential application benefits relating to commercial kits of this nature.

An additional social value that may also result from this study is a new method of skin colour categorisation. The use of MI provides a continuous scale for objective skin colour inference that does not limit skin colour to specific ethnicities or racial groups placed in a set number of population categories. This steers away from stereotypes and political racial discrimination to unprejudiced identification of individuals.

1.6 Ethics and risk assessment

1.6.1 Adverse effects to participants

The level of risk associated with the participating in this study of is minimal since DNA will be collected through the use a non-invasive mouth rinse and standard protocol will be adhered to. The participants will be allowed to take their own sample under supervision. It has been identified that during the reading of the MI, the derma spectrometer, which is a non-invasive method of reading skin colour; has the potential of transferring microbes from one participant to the next [34]. Sterilizing the spectrophotometer lens between each measurement will control this risk.

1.6.2 Risk to researcher

There is minimal risk to the researcher as appropriate laboratory induction has been completed and health and safety guidelines will be adhered to.

1.6.3 Confidentiality

Names of participants will only be recorded on the signed consent page, which will provide a unique participant number; this number will then be used to label all information and samples keeping and protecting the privacy of the participant. Consent forms will be stored in a locked cabinet access only available to supervisor of project. The use of a participant number provides confidentially during the project, but allows for traceability in cases where voluntary participants wish to withdraw from the study. In the case of

withdrawal, which is allowed at any stage during the project, the DNA or data obtained from that individual will immediately be discarded. In the case of publication and final write up no individual identification will be reported or linked to results.

1.7 Budget, project time scale and monitoring

1.7.1 The budget

Table 1.7.1–1 Table indicating the materials and cost involved for the purposes of the project

	Description	**Quantity required**	**Price**	**Price: project**
Sample collection	Sterile buccal swabs	200 swabs	R750 (500 swabs)	R300
Extraction	Extraction kit	120 rxn	R1500 (50 rxn)	R3600
PCR Amplification	Primers	4	R120 (1 primer)	R480
	Master mix	250 rxn	R2000 (500 rxn)	R1000
Gel electrophoresis	Agarose	100g	R3000 (500g)	R600
	TBE buffer	250ml	R710 (1L)	R200
	Molecular weight marker	1x50ug	R1800 (5x50ug)	R360
	Loading dye	0.1ml	R1200 (0.5ml)	R240
SNaPShot PCR	Internal primer	2	R250 for 1	R500
	SNaPshot reaction mix	240	R30000 (5000rxn)	R1500
	SAP	<1 tube	R1000	R1000
	Exo1	<1 tube	R1000	R1000
Sequencing	Sequencing	10	R30 for 1	R300
General consumables	Gloves	2 boxes	R100 for 1	R200
	Pipette tips	1000	R1200 for 1000	R1200
	Filter tips	500	R4500 for 1000	R2250
	PCR tubes	1000	R2600 for 1000	R2600
	Eppendorf tubes	500	R1000 for 1000	R500
Total				R17 830

1.7.2 Project time scale

Gannt chart displaying the estimated time allocated for each stage of the project.

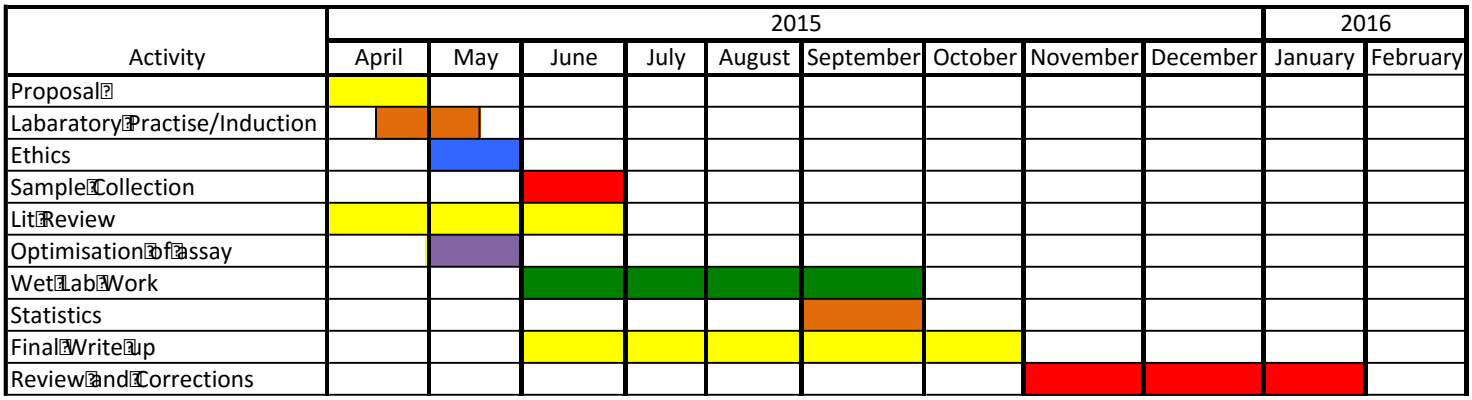

Figure 1.7.2–1 Estimated project time scale

1.7.3 Monitoring

Please refer to the Memorandum of Understanding between supervisor and student for further details with pertaining to meetings, communication and progress report.

1.8 Importance of research and dissemination of knowledge

1.8.1 Importance of research

As previously described, the SA population is very diverse and consists of many mixed ancestry backgrounds. Since many individuals do not fall into neat pre-described categories of population groups, it is difficult to associate skin colour to race or ethnicity. Therefore, an objective method to measure skin colour and pigmentation and further identify the correct phenotypic characteristics from molecular markers is of great scientific and social importance. Assistance with forensic investigations has been seen internationally and taking into consideration that the NFDD is a new development in SA and thus has yet to reach its full potential, the use of different avenues of investigations may hold even greater benefits.

1.8.2 Dissemination of knowledge

This research will be written as a mini dissertation required for the completion of the MPhil in Biomedical Forensic Science. If deemed suitable the associated manuscript will be submitted for publication. A poster summarising results will also be made for display.

1.9 Appendices

Appendix 1A: Research information and participation consent form

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

Affiliation: Division of Forensic Medicine
Department of Clinical Laboratory Sciences
Faculty of Health Sciences
University of Cape Town

Researcher: Nandi Slabbert *MPhil: Biomedical Forensic Science candidate*

Supervisor: Laura Heathfield *BSc (Med) (Hons) MSc*

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 characteristics (EVCs), which are, however, closely and reliably connected to their genes. These EVCs include, but are not limited to eye, skin and hair colour. This knowledge becomes useful when trying to identify individuals within the flied 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 reference 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 look at DNA and its connection with EVCs as possible new investigative tool by means of appearance prediction. This allows for the narrowing of suspect pools and provide visual EVCs to aid in the identification of unknown remains.

The use of this phenotypic prediction from DNA is becoming more common internationally and standardised kits are being made. However, South Africa is known

for its diversity and this may complicate tests if international markers are not applicable 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 two hairless sites on each arm, namely, the inner forearm and the inner arm above the elbow. This will be done using a DSM II ColorMeter (CyberDerm Inc., USA), and will be recorded on the participant's questionnaire.

3. DNA sample – Buccal cells will be collected from each volunteer, via a salvia sample. Salvia will be obtained through means of a 10ml 0.9% saline solution mouth rinse collected in a 50ml 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 with this procedure. The derma spectrometer 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 traceability of your sample and information, and therefore if you wish to withdraw at any 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.

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 (insert email address).

Participant number: XXX

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 the 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

_____

Signature of Witness Date

<u>Appendix 1B:</u> Participant Questionnaire

Participant reference number: XXXX

1. Sex ☐ Male ☐ Female

2. Age ☐ 18 – 39 ☐ 40 – 60 ☐ 60+

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

☐ African Black ☐ Coloured ☐ Indian/Asian ☐ White

4. Race according to anthropological classification you associate most to

☐ Caucasoid ☐ Negroid ☐ Mongoloid ☐ Australoid
☐ Mixed (tick combination above)

5. Ancestral origin

☐ European – N W S E
☐ Africa – N W S E
☐ Asian – N W S E
☐ Middle Eastern
☐ Other
Specify: _____

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

_____

7. Father's ethnicity and/or ancestral origin

_____

8. Mother's ethnicity and/or ancestral origin

_____

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

Grandfather: _____

Grandmother: _____

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

Grandfather: _____

Grandmother: _____

11. Recorded Melanin Index

Right inner fore arm: _____

Right inner arm above elbow _____

Left inner forearm: _____

Left inner arm above elbow: _____

Calculated Average: _____

Appendix 1C: Ethics Approval

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

HREC 317/2015

2006), based on the Association of the British Pharmaceutical Industry Guidelines (ABPI),and Declaration of Helsinki guidelines.
The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

HREC 317/2015

Appendix 1D: Research proposal amendments

The following changes transpired during the study period and edits to the project protocol are listed below

a) Section 1.4.2 Sample size calculation

"Due to the fact that this project is a small pilot study as part of a mini dissertation the investigators deem that a sample size of 100 individuals is appropriate."

The final study sample size comprised 92 individuals due to some DNA being too degraded to amplify during the experimental stage.

b) Section 1.4.3 Cohort

"Recruitment will be carried out independent of race, but rather on ranging MI scores; n=25 in each quarter (Figure 1.4.3–1). This has been done through the equal division of the MI scale to ensure the collection of samples across the whole range of the scale (broad range of pigmentation tones) for more accurate representation regardless of race. "

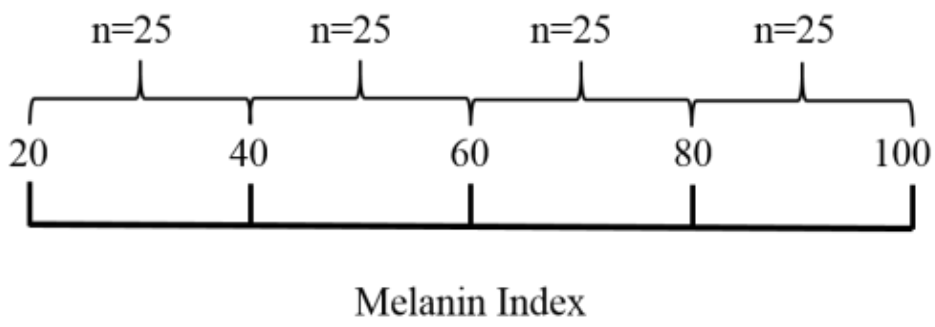

Figure 1.4.3–1 Scale representing MI values and sampling

During recruitment, the investigators found it challenging to locate individuals within the 80 – 100 MI range. Due to the time constraints placed on the recruitment period the use of the aforementioned MI divisions was discontinued.

However, recruitment was still collected independent of race, and volunteers were recruited to comprehensibly distribute the average MI across the human pigmentation scale within the demographic.

c) Section 1.4.6 Analysis of results

"Frequencies will be compared using the Pearson test and Pearson's product correlation co-efficient will be used to estimate the influence of ancestry in the dataset. ANOVA testing for association of genetic polymorphism with the MI will be performed and a p value of 0.0028 will be considered as statistically significant as reported and used in a previous admixed population study [22]. STATA (StataCorp LP) statistical software, version 11 will be used for calculations."

This was amended to the Kruskal-Wallis test due to the non-parametric nature of the data as well as it not being normally distributed. The statistical analysis programme IMB SPSS Statistics (v22.0) was used and the Bonferroni correction was accounted for within the programme. Significant difference was set at a level exceeding of 0.05 significance.

Additionally, it was decided that sufficient discriminatory power of SNPs would be better determined through the means of an MI predictive statistical model. The prediction model was assessed on its MI mean prediction accurateness for each genotype combination within a 95% confidence interval and then compared to the actual recorded MI mean for that genotype.

a) Section 1.7.2 Project time scale

The original time estimated time line was modified following an extended period of statistical analysis and final write up.

Section 2. **LITERATURE REVIEW**

Ethical, legal and social implications of forensic molecular phenotyping in South Africa

Abstract

Conventional forensic DNA analysis involves a matching principle, which compares DNA profiles from evidential samples to those from reference samples of known origin. In casework, however, the accessibility to a reference sample is not guaranteed which limits the use of DNA as an investigative tool. This has led to the development of phenotype prediction, which uses SNP analysis to estimate the physical appearance of the sample donor. Physical traits, such as eye, hair and skin colour, have been associated with certain alleles within specific genes involved in the melanogenesis pathways. These genetic markers are also associated with ancestry and their trait prediction ability has mainly been assessed in European and American populations. This has prompted research investigating the discriminatory power of these markers in other populations, especially those exhibiting admixture. South Africa is well known for its diversity, and the viability of these particular SNPs still needs to be assessed within this population. South African national law currently restricts the use of DNA for molecular phenotyping, and there are also numerous ethical and social considerations, which are discussed.

2.1 Introduction

There have been great advancements in the analysis of the human genome that have expanded our understanding of genetic variation among populations and individuals. Although research of this nature can be considered controversial [14,36,37], it has nevertheless provided beneficial insight into a wide variety of fields, including diseases, behaviour, population migration patterns and forensic science [38–41]. The application of genetics to forensic science has enabled analysis of DNA to link suspects to crime scenes, as well as identify victims. More recently, genetic markers which are linked to physical traits (such as hair, eye and skin colour) [42–45] and ancestry [5,14,46–48] have been analysed to generate new leads in forensic investigations, to narrow suspect pools, as well as assist to in the identification of deceased persons beyond recognition [6,44]. However, before these methods can be integrated into practices in SA, the potential viability of such investigative tools needs to be assessed within our specific context. The relevance and accuracy of previously identified genetic markers needs to be investigated within SA population groups and the ethical, legal and social implications associated with these methods need to be addressed.

2.2 Conventional forensic DNA profiling

2.2.1 The principle of forensic DNA profiling

Forensic DNA profiling is an internationally accepted method used to identify individuals for forensic purposes [1–3]. Genetic evidence is collected in the form of biological material, which includes but is not limited to blood, saliva, buccal cells, epithelial cells, semen and vaginal fluid and can be retrieved from items at crime scenes or directly from individuals [10]. Forensic DNA profiling encompasses the principle of matching a DNA profile from an evidential sample of unknown origin with a DNA profile from a reference sample, which can either be from a known suspect, or a profile stored on a DNA database [3,4]. This process enables DNA profiles to be used as an investigative tool and as scientific evidence in forensic cases [3,8].

Standard forensic DNA profiling involves the use of a select number of specific regions within DNA that have been shown to express a high degree of variation between individuals [10,13]. These regions are known as STRs which fall within non-coding regions of the genome and are thus not translated into proteins [8]. STR sequences vary due to the different number of times a bp motif repeats itself at a locus; this variation is used to separate alleles based on size. Thus, a STR genotype is the allele(s) reported according to the number of repeated units observed at a locus, and multiple STR genotypes are combined to form an STR profile, which can then be used for forensic identification purposes [10].

2.2.2 Limitations of forensic DNA profiling as evidence

Generating STR profiles from biological evidence is the standard practice for forensic DNA profiling globally and is the accepted norm within the criminal justice system [1–3]. However, the use of these profiles relies on the presence of a known reference sample [4,5]. Therefore, in cases where there is no identified suspect, or when there is no match to the NFDD, the unknown evidential DNA sample is limited in its application as an investigative tool [5–7]. This is also the limitation with the identification of body parts of highly decomposed bodies. While STR profiles can still be generated in these situations, linking them to missing persons or mass victim disaster cases is often difficult as comparative reference samples are rarely available.

This challenge has prompted an increased focus regarding research into the prediction of observable physical characteristics based on the genetic material recovered from biological samples [5,7,46,49].

2.2.3 Alternative individual identification techniques

Besides the aforementioned STR regions, another genetic variation that can be used for identification purposes is the single nucleotide polymorphism (SNP). SNPs are single base sequence variations that occur at a particular point in the genome, that have been shown to vary between individuals [10]. Each SNP locus typically possesses only two possible alleles, therefore if used for forensic identification, more markers are required to obtain the same discriminatory power between individuals than compared to STRs [13]. SNPS, which meet

specific criteria, such as displaying globally low allele frequency variation while maintaining high heterozygosity for individual identification, have been selected to form panels for forensic identification purposes such as IISNP and SNPforID. These panels have been designed with sufficient scientific basis to not be rejected in court [15,50].

It is unlikely for SNPs to replace STR profiling as the primary genetic analysis technique due to the existence of well-established forensic DNA databases comprising STR profiles. However, there are some significant benefits of using SNPs for forensic DNA profiling [13,51]; for example, there is usually a higher information recovery rate from degraded DNA samples from analysing SNPs as compared to STRs, due to their smaller target regions [13]. This is better suited for the short DNA fragments present in highly degraded DNA samples commonly encountered in forensics, such as in decomposed bodies and environmentally exposed crime scene evidence [10,51].

2.3 Additional forensic DNA analysis tools

2.3.1 SNP application: Ancestry inference

Certain SNPs located in coding regions of genes can be conserved between individuals and within population groups. Due to the low mutation rate of such SNPs, specific variations have the potential to become "fixed" in population groups [51]. Conservation within different population groups due to different geographical selection pressures provides the potential to discern the ancestral origin of population groups [51–53].

A number of SNPs called Ancestry Informative Markers (AIMs) have been identified and have shown to be able to infer ancestry [14]. Due to exhibited frequency variations of these AIMs among populations [25], statistical calculations can provide the probability of the alleles being representative of specific biographical populations. This enables inferences to be made regarding whether or not the DNA is consistent with a particular ancestral population group [14]. It is important to note that allocating an individual to an "ancestral population" is more descriptive of where their genealogy originated from, rather than depicting their visual appearance. However, due to significant associations between certain visible traits and ancestry, AIMs are commonly used to infer ethnicity and/or race [17,25,54]. For example, lighter skin pigmentation is conserved in Northern Europeans, therefore with appropriate ancestral population databases, a DNA sample inferring that the donor could be of genetic North European ancestry will therefore also infer that the donor is likely to have a light skin colour [51]. However, this is an indirect method of appearance

prediction and it should be noted that the terms “ethnicity” and “race” are widely used interchangeably in studies across different research fields. However, “race” is more commonly used in scientific literature to refer to biogeographic ancestry, whereas “ethnicity” is regarded as a more complex multidimensional construct incorporating cultural and social classifications [55,56]. Nonetheless, even considering this inclination the term “race” used in many study designs, lacks a clear and established definition with a failure to distinguish between self-identified categories and assigned assumed categories [57,58]. In this review, “race” will be defined according to the Stats SA census population groups, while ethnicity refers to self-identified groups encompassing cultural meaning [59].

SNP panels for ancestry inference have been based on the selection of specific markers that have met set criteria, as described previously [54,60]. Currently, 128 SNPs have been identified; not only have they met the conditions, they have also been broadly tested over major population groups for the creation of an assay relevant on a global scale [61]. Commercial kits, such as the DNAWitness Bio-Geographical Ancestry kit (DNAPrint Genomics, Inc.), have been developed and can infer ancestry to sub-Saharan African, European, East Asian or North American population groups as well as determine relative percentages of admixture.

This analysis tool was successfully applied in the forensic field during the investigation of a serial killer in Louisiana, United States (US), in 2013 and received a lot of media attention [37,62,63]. Homicide Task Force Investigation

used proportional ancestral estimations gained from the analysis of a panel consisting of 71 AIMs to infer the ancestry of the donor of an evidential DNA sample. The results inferred that the donor was primarily of African ancestry, which was in complete contradiction to the single eyewitness testimony, which, at that stage, had been one of the main investigative leads. A suspect was later arrested using this phenotype prediction as a new lead and was subsequently confirmed as the perpetrator through the use conventional forensic DNA profiling [62,63].

The identification of SNPs as possible ancestry trait markers especially for forensic science needs to be reviewed regularly in order to improve prediction models as well as increase their relevance to different population groups [15,51,60]. The majority of previous studies that have identified suitable genetic markers, developed SNP assays and prediction models as well as generated reference databases, and were carried out in Europe and Northern America [50,64]. In these studies ancestral origin was categories by major geographical regions or countries [50,60]. This has prompted investigation regarding the applicability of these SNPs in different and smaller population groups, particularly those that exhibit Mixed ancestry [22,32,65,66].

The ability to accurately differentiate between population groups is important especially for accurate application in forensic cases. AIMs that have also been recognised as SNPs associated with skin colour have been shown to have discriminatory significance in some population groups; however, not all may be

useful within admixed populations [22]. In Brazil, eighteen recognised AIMs previously associated with skin colour were evaluated and their forensic significance within their local admixed population was determined [22]. Some of the classical markers tested were found not to be associated with their population and only two SNPs showed significant association with skin colour [22]. This suggested that perhaps some markers are more informative in certain population groups than others, which in turn could limit their use in international commercial kits.

The two polymorphisms that seem to behave as skin colour predictors in admixed populations occur within two coding genes, namely: *solute carrier family 24, member 5* (*SLC24A5*) and *solute carrier family 45, member 2* (*SLC45A2*) [22].

The sequence variation rs1426654 (*SLC24A5*) has been shown to influence skin pigmentation and can be used to elucidate ancestry in individuals of European, African or Asian descent [26]. This *SLC24A5* gene has a 100% fixed variant in pale Europeans, is absent in African and Asian populations and is estimated to make up 30% of the difference in melanin between Europeans and Africans [67]. The other polymorphism, rs16891982 (*SLC45A2*), also influences skin pigmentation and has been linked to light skin European ancestry [27].

The analysis of genetic markers such as these provides an objective result regarding skin colour, which is not affected by a possible individual bias from an

eyewitness. This has prompted research regarding the genetic association of DNA markers with other phenotypic traits of an individual, which include previously established AIMS as well as other genetic markers [10,12]. In forensics, these are limited to EVCs, such as eye, hair and skin colour, that can also be inferred by SNP analysis [6].

2.3.2 SNP application: Molecular phenotyping

2.3.2.1 External visible characteristics

Certain genetic variations within coding regions of DNA can alter gene expression. These variations can result in differentiations within a certain trait characteristic and can be responsible for high variability in characteristics [15,51,68]. For example, specific SNPs within the pigmentation genes associated with eye colour can change the observable colour from light (e.g. blue eye colour) to dark (e.g. brown eye colour) [45,65,69]. Similarly, SNPs can influence hair colour, skin colour and the presence or absence of freckles [70,71]. Such characteristics are known as EVCs and their prediction through molecular DNA analysis is called molecular phenotyping. This provides the means to predict an individual's visible appearance by inferring the probability of certain characteristics being expressed over another, due to its association with specific genetic variation [24].

There is also potential to use these predictions to aid in forensic investigations [5,6]. External visible characteristics are observable details of an individual, which could form an investigative lead for a suspect; therefore, the ability to

objectively predict what someone may have looked like is advantageous, as it is not affected by an eyewitness's subjectivity or individual bias. This has prompted research focusing on the identification of these markers and the development of SNP assays, such as the IrisPlex assay, which uses genetic information to predict eye colour [16]. These assays genotype multiple SNPs simultaneously, as one EVC is not necessarily the result of only one genetic variation [15,72,73].

The combined analysis of different SNP panels for multiple traits can provide enough discerning information to infer different EVCs and thus predict an individual's appearance. This service has been commercialised by an American company called Parabon NanoLabs, Inc. Their analysis can compose an image that represents the individual's predicted face morphology; genomic ancestry, hair, skin and eye colour, as well as predict the presence of freckles. The report includes the probability of an individual's predicted characteristics as well as the probability of the individual not having other characteristics. These traits and probabilities are derived from the application of statistical methods based on their reference database (Parabon Guidelines and Report, 2015). This produces an individual phenotypic profile based on EVCs that could be incorporated as a forensic investigation tool to aid in narrowing suspect pools as well as to help identify human remains beyond recognition.

2.3.2.2 Skin pigmentation pathway

Pigmentation is perhaps one of the most obvious distinguishing human traits due to its high variability [7,19,74]. This colouration is the result of complex

genetic and biochemical pathways and is also affected by environmental factors. More than 120 different genes and 600 proteins are involved in the pigment phenotype observed in humans [43,49].

The pigmentation pathway of an organism is responsible for the synthesis and regulation of the amount and type of melanin pigment expressed, a process known as melanogenesis [75,76]. The process begins with specialised cells called melanocytes, which are responsible for melanin production. Melanin is a biopolymer that provides the majority of pigment seen within all mammals and comes in two forms: eumelanin, which is black, and pheomelanin, which is red, yellow or brown [77]. Widely varying ratio combinations of these two forms of melanin are observed in individuals with different shades of hair, eye and skin colour [78,79].

The melanocyte pattern determined during embryo development is strictly regulated by genetic pathways [80]. The formation occurs through the migration of melanoblasts, the melanocyte precursor cells. The final destination of these cells is located in the basal layer of the epidermis of the skin, the uvea of the eye (a middle layer) and hair follicles [74]. Melanocytes are independently expressed across different locations, despite originating from the same embryonic source. This means that an individual displaying dark hair characteristics does not necessarily have to have dark eyes [78]. Distribution of melanocytes within skin is typically constant, thus the variation seen between one individual to another is a result of the size, number and structural pattern of these cells. Melanocytes

produce and contain melanosomes. These melanosomes mature and migrate to the edge of the cell where they are transferred to the epidermal cells. Darker skin pigmentation in individuals is due to larger, individually arranged melanosomes, with more concentrated melanin compared to paler pigmented individuals (Figure 2.3.2–1) [81]. Pale skinned individuals also have smaller melanosomes containing less melanin grouped together within the membrane structure [74,78].

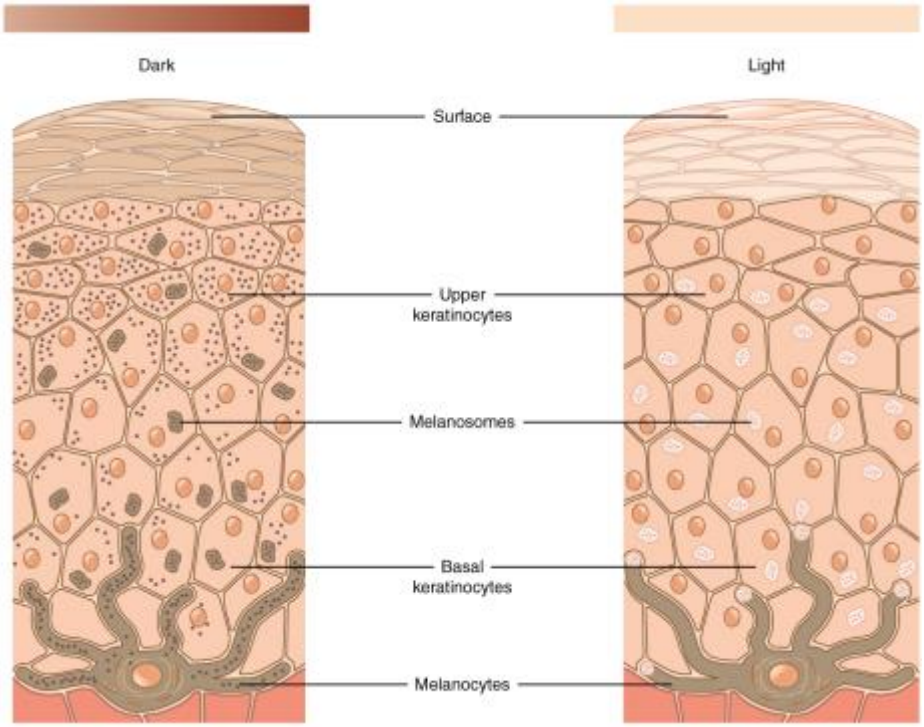

Figure 2.3.2–1 Cross-section of human skin showing the difference between melanocyte arrangement and melanosomes between dark and light skin pigmentation. (Figure taken from OpenStax College, 2013) [81]

2.3.2.3 Measuring skin colour

Natural skin pigmentation is a continuous spectrum with a wide variety of different skin colours between different population groups as well as individuals within the same population group [76]. Methods of skin colour classification vary

between different research fields such anthropology and dermatology [18,82,83].

The Fitzpatrick scale is a numerical skin colour classification system created in 1975 for dermatological use. It has since been used worldwide to estimate the risk of skin cancer and malignant melanoma based on an individual's reaction to sun exposure and predicted reactivity to photo chemotherapy [84]. The scale consists of six possible skin type categories that are assigned according to the results of a self-reported questionnaire. The questions include grading an individual's burning and tanning abilities as well as certain genetic dispositions such as hair and eye colour [83]. The scale shows preference to the identification of light skinned individuals, providing four possible classifications for a "white"-skinned person (skin type I, II, III, IV as seen in Figure 2.3.2–2). Darker pigmented individuals are sorted into either the "brown" skin type, classified as skin type V, or "black" skin, classified as skin type VI.

Skin type	Erthema and tanning reactions to first sun exposure in early summer
I	Allways burn, never tan
II	Usually burn, tan less than average (with difficulty)
III	Sometimes mild burn, tan about average
IV	Rarely burn, tan more than average (with ease)
V	(Brown-skinned persons)
VI	(Black-skinned persons)

Figure 2.3.2–2 Summarized table representing the Fitzpatrick skin type scale. (Figure taken from M.H. Ravnbak, 2008) [83,84]

In contrast, within the anthropological field human skin colour is categorised against a different numerical classification chart known as the von Lushan scale [82]. This chart consists of 36 categories and was used to establish racial classification of populations according to skin colour. This was achieved through the comparison of 36 established colour tiles. A comparison study showed a positive correlation between this colour chart and a spectrometer reading of skin colours; however, due to the highly subjective nature of the colour matching method, different practitioners have assigned inconsistent categories to the same person [85].

It is acknowledged that the interpretation of colour is very subjective and objectivity is of great importance within forensic science [20]. Thus, to avoid bias, it is beneficial to use an objective instrument to record pigmentation in established units of measurement. A derma spectrometer is one such instrument that can measure skin colour pigmentation in MI units. Due to the light absorbance characteristic of human skin, this hand-held reflectance spectrophotometer can provide readings of the erythema and melanin indices. The units increase as the skin becomes more pigmented, thus the MI can be regarded as a parameter influenced by melanin content [86]. The reflectance spectrophotometer is designed to measure specific colours due to the presence of haemoglobin and melanin by emitting light at selected wavelengths. The reflected light from the skin is detected with a photo detector and converted into a digital unit by a built-in microcomputer [20]. A specific example of such an instrument is the DSM II ColorMeter (CyberDerm, Inc., USA) which outperformed

its competitor models and was deemed more accurate [20].

This method of colour recording can give an impartial interpretation of skin pigmentation and is more representative of real life skin colour since the MI units are counted on a continuous scale of pigmentation rather than being limited by generalised numerical and racial categorisation. This method of measuring skin colour has added benefit in SA, where numerous population groups exhibit high skin colour variations.

2.4 Application of molecular phenotyping in South Africa

2.4.1 The SA population

South Africa exhibits a highly diverse population consisting of groups originating of African, Asian and European descent [37]. The complex history of SA combined with its diversity has led to the unique admixed population officially known as the "Coloured" population. Individuals who categorise themselves within this population group vary in pigmentation levels and display a wide variety of different phenotypic traits [88].

The 2014 mid-year population estimation reported that the "Coloured" population group comprised 8.8%, the second largest portion of the total SA population. These individuals predominately reside within the Western Cape province [30]. The "White" population group was estimated to be 8.4% of the total population while the "African" population group was estimated to consist of 80.2% of the population [89].

A study conducted in 2010 that focused on the genomic structure of the Western Cape Coloured population consisted of an analysis of 959 self-identified "Coloured" individuals [87]. Through this study it was estimated that this group has been derived from four to seven ancestral groups with major contributors being from the Khoesan Africans, non-Khoesan Africans (including East African, Buntu and Pygmy populations), European (German, French and Dutch) and Asian groups [87].

A similar concept study was conducted on a self-identified Afrikaans-speaking South African who set out to "deconstruct his genetic heritage" [90]. Results that were expected to be genetically homogenous (as it is assumed that Afrikaners stem from a Dutch European descent) showed that 6% of his genetic heritage was of West African descent providing further evidence to the admixed ancestral nature of the broader SA population. This raises the question as to whether discerning between individuals of admixed populations, such as seen in SA, will be possible if AIMs are analysed for forensic purposes.

2.4.2 Legal considerations

The value of DNA evidence and DNA databases is supported by empirical data and thus developments in the science have had important implications for legislators and the legal system [91]. The recent implementation of the NFDD holds the potential to be a beneficial forensic intelligence tool within the country; however, time is needed for it to reach its full potential. This database is regulated by the DNA Act, which defines a forensic DNA profile as:

"…the results obtained from forensic DNA analysis on bodily samples taken from a person or a crime scene, providing a unique string of alpha numeric characters to provide identity reference: Provided that it does not contain any information on the health or medical condition or any information on the predisposition or physical information of that person other than the sex of that person." [28].

This definition specifies the information that a forensic DNA profile can convey as well as prohibits the inclusion of any physical information except for the individual's sex. The DNA Act provides further definitions such as "forensic DNA analysis":

"…the analysis of sections of the DNA of a bodily sample or crime scene sample to determine the forensic DNA profile: provided that this does not relate to any mental characteristic of the person or to determine any physical information of the person other than the sex of that person." [28]

This definition thereby legally controls the use of genetic material for forensic purposes, restricting it to only include the generation of a forensic DNA profile. This confining definition has been criticised as a negative constraint with regard to any possible future development of DNA analysis for forensic purposes in SA [2,8]. Considering that significant advances have been made in science and technology regarding DNA since its first use as evidence, it is clear that the SA criminal justice system will need to adapt policies and regulations to accommodate new and current developments in the field to remain pertinent [91].

Though the current SA legal definition of a forensic DNA profile prevents the use of molecular phenotyping, this is not the case in all legislation around the world.

An investigative study was conducted in 2008 into the internal regulating laws surrounding forensic DNA phenotyping in eleven countries including SA (before the amendments to the DNA Act were implemented) [92]. The study concluded that "almost all surveyed countries have some form of DNA legislation, but these laws are generally confined to traditional DNA forensics, that is, making DNA profiles or fingerprints from crime-scene material and comparing these to profiles stored in forensic DNA databases" [92]. The researchers found it common that forensic DNA profiling was restricted to non-coding regions of the genome as seen in countries such as Australia, South Africa and Spain. However, the Australia Law Reform Commission has held debates regarding the legislation restriction but has yet to make any formal recommendations [62].

Some countries do allow for the use of molecular phenotyping for forensic purposes whether explicitly stated in their legislation or not. For example, the United Kingdom has implemented the technique in practice without specific legal regulation and used it to provide inferences of race and ethnicity as an investigative aid. The arrest and later conviction of the "Night Stalker" is a known successful British case that involved phenotyping to establish a suspect's racial background. This information aided a 20-year investigation without any previous leads and contradictory accounts of the perpetrator's race [91].

In contrast the Netherlands made a formal amendment to its legislation in 2003 to explicitly allow for DNA to be used in the prediction of physically visible traits (traits that are deemed non-private, or public) of an unknown individual to aid

criminal investigations. The Dutch Code for Criminal Procedure limits these traits to only those that can contribute to the investigation and protects the subject's privacy by stipulating that only characteristics known to the suspect may be typed and must have been present and visible at time of birth [62].

The US currently has no federal legislation on the use of DNA phenotyping in criminal investigations and replies mainly on state law. As previously mentioned, the technique has been used in the past; however, only the state of Texas has legally authorised its use in investigations. Furthermore, the state of Vermont has legislation prohibiting the identification of medical and genetic disorders but does not expressly forbid phenotyping for investigation purposes [92].

The recent formation of the NFDD is a beneficial aid against crime within the country, and is one major step towards increasing criminal intelligence in SA. However, if molecular phenotyping is to be included as a forensic service in SA, not only does the genetic assay need to undergo extensive validation, but the law needs to be amended to include the practical application of molecular phenotyping for investigative purposes – and perhaps not for court presentation, as suggested by literature which will be discussed in the next section [62,92].

It is thus recommended that SA law be amended, either by the exclusion of restrictive definitions that prohibit advancements in the field, or the addition of clauses permitting phenotyping under specified, controlled conditions which include detailed and acceptable storage methods, parties permitted access to the

reports and informed interpretation and application of results in the manner approved for forensic investigations.

2.4.3 Ethical considerations

The observed lack of consensus between the current international laws regulating molecular phenotyping conveys the contentious nature of molecular phenotyping. Besides the arising legal issues surrounding the prediction of EVCs, ethical implications regarding its use in practice have been a highly debatable topic amongst scientist in the field as well as bioethical research groups [5,6,92–94]. It is important that these issues are discussed within the forensic community and the motivations behind their opposing views are understood and addressed for the benefit of the public and legislative bodies.

One of the primary ethical implications that have been brought to attention is the possibility of privacy infringement. Estimating physical traits based on DNA may be perceived as a violation of an individual's privacy, especially if it reveals a possible sensitive phenotype or an underlying trait, which has been intentionally or unintentionally altered. It was argued by Kayser and Schneider that the prediction of EVCs does not lead to the violation of privacy [6] due to the characteristics being *external* and *visible* to others in nature (such as skin, hair and eye colour). Thus, the information gained from this technique cannot be considered as private, since the characteristics are observable to anyone in public and provide the same intelligence, if not more, than eye witness accounts [95]. Molecular phenotyping also provides an objective and unbiased description

of an individual, which is perhaps preferable compared to an eyewitness report, which could be subjective or biased [5,6,95]. However, in a correspondent's reply to one of Kayser's articles stating this argument, the author mentioned that these "biological witnesses" lack the context of a crime committed that could potentially be provided by a real eye witness [96]. Further, any altered or distinguishing characteristic will not be estimated by molecular phenotyping.

Secondly, another major concern included the physical storage of the data and which persons will be granted access to this information during and possibly after the investigation [96]. There is great public unease that saved information could be used inappropriately and either mined for genetic data [97], or used to link individuals to certain medical and behaviour disorders [62,92]. This information could be used to discriminate against individuals based on medical predispositions, or even undermine a person's ability to obtain health insurance [98].

Although it is argued that these issues can be regulated with proper implementation of legislation regarding security and storage procedures as well as restricted practice of molecular phenotyping for cases meeting specific requirements [5,62,92], concerns regarding the "right not to know" have been raised. Although it has been suggested that only cases concerning unknown persons would be deemed appropriate, the effect of gained knowledge of possible conditions of an individual after confirmed identification is still controversial [62,96,98]. Defence motivations include the reasoning that public

interest and the safety and benefit of society outweighs the concern of one possible discrimination in favour of the apprehension of a criminal [5,6]. Conversely, there are claims that the risk of unethical behaviour is too high and could lead to unintended consequences and should be treated with greater sensitivity [92,96,98].

Effects to fundamental rights, such as the right to a fair trial, have also been brought into question [99]. Reservations regarding the weight of DNA evidence in court as well as during the investigation are held due to the phenomenon known as the "CSI Effect" which can possibly cause prejudice in the courtroom [100]. Suggested recommendations have included the regulated use of molecular phenotyping as intelligence only and prohibiting its use in courtroom as evidence [92,101]. The provision of proper training in interpretation of results and clear guidelines for the police is also recommend to avoid over simplification of descriptions and physical profiles [6,37,102]. It is also important to remember that molecular phenotyping will not be used in isolation within an investigation. It should be seen as an additional advantageous tool used alongside and within the context of other evidence relating to the case [6].

2.4.4 Social discrimination and social value

The technology used for DNA analysis for predictive purposes can potentially have serious ramifications if analysis of its application in real-world situations is not extensively carried out [37]. Biomedical ethicists have argued that all genetic variation research should include ethical and social implications related to its

application. This is particularly suggested in research centred on non-medical applications [102].

The analysis of phenotypic information within the field of forensic science is one such application. The inference of ancestry and its alignment with racial categories and ethnicity raises the concern of the potential of prejudicial profiling, resulting in minority groups being vulnerable to suspicion and leading to further discrimination [37,62,98]. It is important to recognise the fear that phenotyping could also make law enforcement more dependent on racial identifiers and thus target at risk racial populations [92,102].

Racial and ethnic categories are commonly seen as inconsistent "socially fluid labels", whereby their "definitions" vary in different countries [99,102]. The variations in these identification labels also highlight their subjective nature [103]. This is further exhibited by the findings in a US study where one-third of participants in the cohort changed their own self-identified race or ethnicity during two consecutive years [102].

This can be partly attributed to the fact that ethnic identification is the result of one, or a combination of, several contributing factors, including: physical appearance, upbringing, cultural influences and/or social circumstances. South Africans exhibit a wide variation of ethnic and cultural backgrounds. However, only four population group categories of classification are used by the Stats SA

census, namely, Black/African, Indian/Asian, Coloured and White [30]. These categories were originally politically based and have been carried over from the segregation categories used during the Apartheid regime [104]. This is obviously undesirable.

"Ethnicity", within the SA context, is defined as how individuals self-identify themselves within a population and is representative of both cultural and visible aspects [88]. Therefore, race is not defined by ethnicity and individuals' ethnicity is not restricted by their phenotypic traits [88]. Underlying homogeneity explains why children generally have similar skin pigmentations as their parents. However, due to genetic heterogeneity observed in admixed population groups, children may inherit different combinations of particular variants from their parents, explaining why children with two parents of very different skin pigmentations can inherit a range of skin colour [105]. It follows then that two individuals may have different skin colours but may be closely related and/or will identify with the same ethnic group.

A concern raised by Cho and Sankar is the extent to which AIMs are used to construct a physical profile when applied to forensic science. They ask the question, "but how does one look for someone who is 'largely West African'?" [102]. This major challenge lies predominantly in the prediction of skin colour for admixed individuals since proportional percentages of different ancestry cannot predict a skin colour [6,7]. In SA this problem is amplified due to the variations in skin colour exhibited in Sub-Saharan Africans compared to other

population groups [7]. In this context, the implementation of molecular phenotyping has an additional social value when MI units are used to measure pigmentation and thus eliminate skin colour "categorisation". The use of MI provides a continuous scale for objective skin colour inference that does not limit skin colour to discrete, specific ethnicities or racial groups. This steers investigation away from stereotypes and political racial discrimination by promoting unprejudiced identification of individuals. It also allows for a representation that can be displayed visually for interpretation thereby not relying on verbal outcomes [5].

The objective nature of phenotyping is the argument against the fear of prejudicial profiling since genetic material cannot be biased if it is based on accurate and valid procedures [5,6,92,93]. It has even been argued to mitigate, rather than encourage, prejudice, as demonstrated in a case whereby a previously biased suspicion held against an individual from a minority group was proven incorrect as DNA showed the real perpetrator to be of European descent [47]. However, this does not quell the fear of many who believe that the inference of ancestry and prediction of EVCs will encourage DNA dragnets and allow for the discrimination of suspect populations [62,96,102].

Although molecular phenotyping carries concerns, which are important to acknowledge and address, it also carries many advantages and social value. Molecular phenotyping increases the degree to which genetic material can be used in forensic investigations, including criminal or missing person

identification cases [5]. The accurate prediction of EVCs can help narrow the field of suspects through the positive exclusion of individuals, saving time and financial resources during an investigation. It will also provide a platform to assign identity to unknown victims and unclaimed human remains beyond recognition at the forensic autopsy stage.

2.5 Conclusion

The recent implementation of the South African National Forensic DNA database is a necessary step against crime within the country. However, given that its development is still in its infancy, it will be some time before the database is able to reach its full potential. This limits the number of matches that the database can return on a forensic DNA profile from unknown origin, for example, from a crime scene. In these instances, there is limited investigative lead. Molecular phenotyping has the potential to become a useful tool to aid forensic investigation by enabling the prediction of EVCs and ancestry from evidential biological samples. This tool could provide new leads in the form of predicting what someone may have looked like and could be of immense value in the post-mortem investigation context, in helping to identify human remains beyond recognition.

Before molecular phenotyping can be applied to forensic applications in SA, however, numerous ethical, legal and social aspects of this technique need to be considered and addressed. These include amending SA law to allow for this type of investigation on certain samples, as well as a policy that regulates the use, analysis, storage and reporting of this information. In addition, and of equal importance, the appropriateness of the standard genetic markers needs to be evaluated within the SA population, and the resulting assay would need to be thoroughly validated in a local context before its use. South Africa consists of a population with a wide range of skin pigmentation and mixtures of ancestry, and considering that an individual's AIMs may not correctly indicate the individual's

EVCs, the viability of these markers within the SA context needs to be investigated. Thus, it is imperative that phenotypic markers, which are currently being used, are assessed and their relevance evaluated within the SA population.

Section 3. **RESEARCH STUDY**

IN MANUSCRIPT FORMAT

The assessment of molecular markers for skin colour in South Africa

Abstract

Conventional forensic DNA analysis involves a matching principle, which compares DNA profiles from evidential samples to those from reference samples of known origin. In casework, however, the accessibility to a reference sample is not guaranteed. This limitation has led to the development of phenotype prediction, which uses SNP analysis to estimate the physical appearance of the sample donor. Physical traits, such as eye, hair and skin colour, have been associated with certain SNP variations within specific genes. These identified genetic markers constitute SNP panels, which have been selected due to their trait predication ability primarily in European and North American populations. This pilot project therefore aimed to determine the significant discriminatory value of two of these markers associated with skin pigmentation, within the SA population, namely, rs1426654 (*SLC24A5*) and rs16891982 (*SLC45A2*). The study cohort was comprised of SA residents (n=92) living within the Western Cape province. The population group classification used followed the Stats SA census categories. Self-identified ethnicity was documented and an objective pigmentation reading was recorded using an MI value using the DSM II ColorMeter (CyberDerm Inc., USA). DNA was extracted from saliva and PCR was used to amplify the targeted regions. A SNaPshot PCR assay was then developed and optimised to genotype the variants. A significant difference ($p<0.05$) was observed between different genotypes and MI values. A General linear model

was developed attempt to predict MI based on genotypes. The predictive model was deemed inadequate within this specific cohort, which indicated that these two markers alone do not possess sufficient discriminatory power to be used as a forensic investigation tool. The value of this project, however, given that it was a pilot study with a small cohort (a limiting factor for the development of the model), was that it established baseline methods. These can be utilised in on-going research with a larger cohort and additional genetic markers, with the overall aim of identifying markers, which are informative for molecular phenotyping in the SA population.

Keywords: molecular phenotyping, admixture, ancestry informative markers, melanin index, South Africa

3.1 Introduction

Forensic molecular phenotyping is when informative genetic variations are analysed to predict an individual's EVCs [24]. These genetic variations are responsible for a high variability of certain traits [15,51,68]. For example, specific SNPs within pigmentation genes associated with eye colour can change the observable colour from light (e.g. blue eye colour) to dark (e.g. brown eye colour) [45,65,69]. Similarly, SNPs can influence other EVCs such as hair colour, skin colour and the presence or absence of freckles [10,12,70,71].

Certain genetic markers have also been linked to ancestry and have been acknowledged as AIMs [5,14,46–48]. Together, these variations provide the means to discern between different populations and can be used to estimate an individual's visible appearance through the inference of the probability of certain characteristics being expressed over another [24]. Since biological samples are often encountered on crime scenes, the use of molecular phenotyping can be especially valuable in forensic science applications, such as aiding to narrow suspect pools as well as helping to identify human remains beyond recognition [5,6].

The ability to accurately discern between population groups is important especially for its application in forensic cases. Previous research has indicated that certain AIMs associated with skin colour only have discriminatory significance in some population groups; and may not be as informative in admixed populations [22,69]. In Brazil, 18 recognised AIMs previously

associated with skin colour were evaluated to determine their possible association with skin colour in two local admixed populations. Of the total SNP panel, only two markers consistently showed association with skin colour measurements in both samples, independently of the admixture level of the populations [22]. The two identified polymorphisms occurred within two coding genes: *solute carrier family 24, member 5* (*SLC24A5*) and *solute carrier family 45, member 2* (*SLC45A2*) [22].

The Brazilian study suggested that additional studies are necessary to confirm the significance of these genetic markers for skin pigmentation especially in admixed populations, which are seldom considered [22]. Considering their findings showed some classical genetic markers were not associated with skin pigmentation in their sample it raises concerns regarding the limited use of these AIMs internationally [22].

This is particularly relevant in SA, which consists of multiple populations exhibiting a wide range of diversity and ancestry [87,88]. With the standardisation of commercial kits for DNA analysis it is important to investigate so called "standard" molecular markers to determine their viability and discriminatory power within local population groups. Thus, this study aimed to investigate if two previously identified AIMs, which seemed to behave as skin colour predictors in admixed populations internationally, have significant discerning value within the SA population.

Natural skin pigmentation is a continuous spectrum with a wide variety of different skin colours between different population groups as well as individuals within the same population group [76,106]. It is acknowledged that the interpretation of colour is very subjective and objectivity is of great importance within forensic science [20,107]. Thus, to avoid bias it was decided that the MI, an objective measurement correlating with skin colour pigmentation would be used in this study, alongside self-identified SA population group and ancestry.

3.2 Methods and materials

3.2.1 Cohort

A total of 92 individuals (aged 18 – 60 years) residing in Cape Town, SA, were recruited for participation is the project. Individuals with pigment disorders, such as Hyperpigmentation, Melasma and Hypopigmentation (e.g. Albinism) as well as individuals with severe scarring in the areas where MI measurements were taken, were excluded. Volunteers reporting the use of any supplements, ointments or treatments designed to alter the skin colour, such as self-tanning, bleaching lotions or tan bedding were also excluded.

This study was approved by the University of Cape Town, Faculty of Health Sciences, Human Research Ethics Committee (UCT REC REF: 317/2015). All participants gave informed consent and the study was executed in accordance with the Declaration of Helsinki, as amended in Brazil (2013).

3.2.2 Sample collection

Each participant completed a questionnaire where individuals self-reported their ancestry (Figure 3.2.2-1), ethnicity and population group; the latter according to the STATS SA census categories (African Black (n=29), Coloured (n=14), Indian/Asian (n=8), White (n=38) and Other (n=2)).

MI was objectively measured using the Derma spectrometer (DSM II ColorMeter, CyberDerm Inc., USA). Measurements were collected in duplicate from the participant's inner forearm and the inner arm above the elbow (avoiding hair, moles, or other discoloured areas such as scars). The average MI for each individual was calculated and distribution of MI across the human pigmentation scale can be seen in Figure 3.2.2–2. Material for DNA extraction was collected from individuals in the form of a previously published saliva mouth rinse method [108,109].

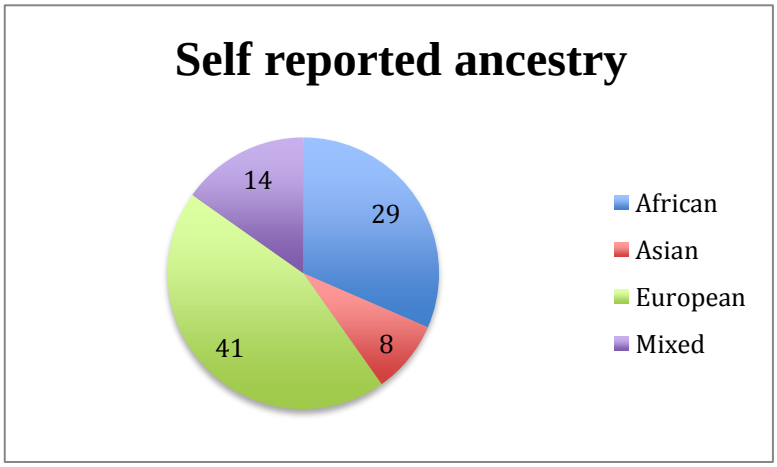

Figure 3.2.2–1 Graph depicting individual self-reported ancestry of volunteers (n=92).

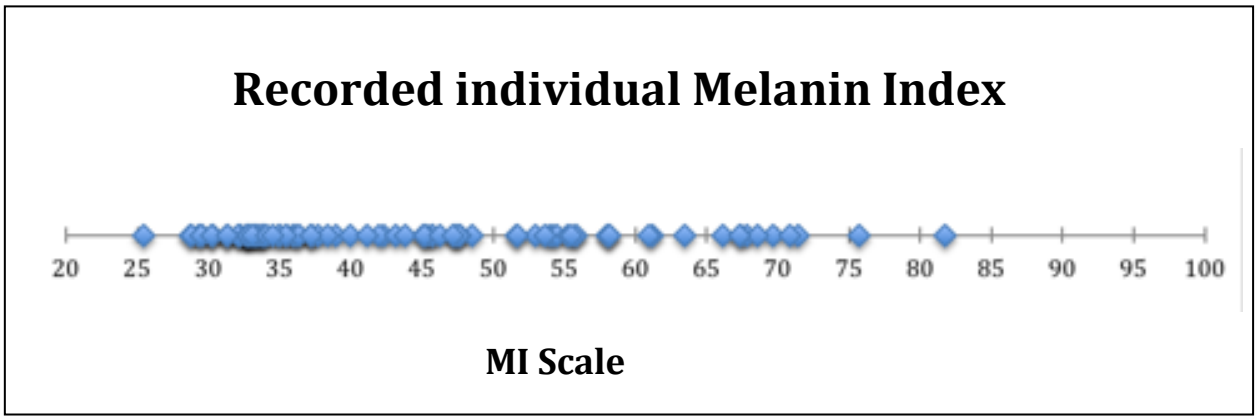

Figure 3.2.2–2 Graph depicting average MI unit per individual (n=92) across human pigmentation scale (20 – 100 units).

3.2.3 Genetic analysis

Genomic DNA was extracted from the saliva using the Quick-gDNA Mini Prep Extraction Kit (Zymo Research, CA, USA), Samples were initially centrifuged at 1500 rmp for 15 minutes or until a pellet formed, using Neofuge 15R (Vacutec). The supernatant was discarded and the DNA pellet was re-suspended in Genomic Lysis buffer (Zymo Research, CA, USA). Thereafter, the manufacturer's protocol was followed, and DNA was eluted into 50 µl of elution buffer.

DNA fragments containing each SNP (*SLC24A5* rs1426654 and *SLC45A2* rs16891982) were amplified individually. Each reaction contained between 80 to 100 ng of DNA template, 0.4 µM each of the forward and reverse primers (Table 3.2.3–1) and 1X 2X KAPA Taq ReadyMix (KAPA Biosystems, SA) in a total volume of 20 µl. Primers were obtained from previous literature [69,97,110] and verified using bioinformatics tools; BLAST, NCBI build 37.5 (https://blast.ncbi.nlm.nih.gov/Blast.cgi), and Oligo Analyser v.3.1 (https://eu.idtdna.com/calc/analyzer). Amplification was carried out on the T100 thermal cycler (BioRad, Berkeley, CA, USA) at the following cycling conditions; initial denaturation, 94°C for 5 minutes, followed by 30 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, elongation at 72°C for 30 seconds and final extension was carried out at 72°C for 7 minutes.

During optimisation, Sanger sequencing was initially used to confirm the amplification of correct target regions, and thereafter, amplification was

confirmed through gel electrophoresis visualised by ethidium bromide staining on a 1.5% agarose gel. A Quick-Load 50bp DNA ladder was added and electrophoresis was performed at 100 V for 1 hour.

Removal of unincorporated primers and dNTPs from the post-PCR amplicons was carried out by incubating 5 µl PCR product with 2 units of *Exonuclease*-1 (EXO1) and 1 unit of shrimp alkaline phosphatase (SAP) at 37°C for 1 hour followed by 15 minutes at 75°C. This was carried out using a T100 thermal cycler (BioRad, Berkeley, CA, USA).

A SNaPshot™ PCR multiplex was developed and optimised. Single base extension (SBE) primers (Table 3.2.3 - 1) were designed and modified with non-binding elongated tails from previously published primers [25,69,97,110]. Primers were tested *in silico* for hairpins and dimer formation using Oligo Analyser v.3.1 (https://eu.idtdna.com/calc/analyzer), and were then synthesised (IDT, Inqaba Biotechnical Industries, Pretoria). A primer mixture ratio of 1:1 was made and 2 µM of the pooled primers was added to the PCR product of both *SLC24A5* rs1426654 and *SLC45A2* rs16891982 in a ratio of 2.33:1. SNaPshot™ PCR was then carried out according to the manufacturer's protocol (Thermo Fisher Scientific, MA, USA).

After SNaPshot™ PCR, a second clean-up was performed to remove remaining fluorescently labelled ddNTPs by adding 1 unit of SAP. The incubation period was the same as for the first purification step. Purified post SNaPshot™ PCR

products (2 µl) were added to 7.7 µl of Hi-DiTM Formamide and 0.3 µl of GeneScan LIZ 120 size standard per well in a MicroAmp plate. A denaturation step for 5 min at 95°C then occurred followed by a snap-freeze step on ice. Capillary electrophoresis was then performed on the ABI 3130*xl* Genetic Analyser (Applied Biosystems, Foster City, CA).

Data were collected using the 3130xl Genetic Analyser Data Collection Software v 3.0 and electropherograms were analysed using GeneMapper v4.1. An analytical threshold was set at 50 relative fluorescence units (RFUs). Samples representing each possible genotype underwent Sanger sequencing for confirmation.

Table 3.2.3–1 Designed primer sequences for both external and SBE primers for *SLC24A5* (rs1426654) and *SLC45A2* (rs16891982)

	PURPOSE	DIRECTION	SEQUENCE OF PRIMER (5' – 3')	REFERENCE ADAPTED FROM
SLC24A5 **rs1426654**	External PCR	Forward	CCC TTG GAT TGT CTC AGG ATG	[25,69,97,110].
		Reverse	TGA GTA AGC AAG TAT AAG GAG CAA	
	Single Base Extension	Reverse	TCT CAG GAT GTT GCA GGC	
SLC45A2 **rs16891982**	External PCR	Forward	TCC AAG TTG TGC TAG ACC AGA	
		Reverse	CGA AAG AGG AGT CGA GGT TG	
	Single Base Extension	Forward	TTA ATT ATG TTA TAT CTT ACA CGG AGT TGA TGC A	

3.2.4 Statistical analysis

Allele and genotype frequencies for each SNP were calculated by counting within each self-identified SA population group. Hardy-Weinberg Equilibrium (HWE) was assessed to investigate the behaviour of the markers within the genetic population.

The Kruskal-Wallis test was performed to determine whether a significant difference of MI mean observed between self-identified SA populations groups as well as between different ancestry and observed genotypes. This test was also used to determine the distribution between observed genotypes within the SA Population groups. The Bonferroni correction was accounted for within the statistical analysis programme IMB SPSS Statistics (v22.0). A confidence interval of 95% was selected deeming p-values less than 0.05 as significantly different.

A mean MI predictive statistical model based on observed genotypes of individuals was designed. The model considered all observed genotypes of each SNP individually as well as combined genotype observations as preliminary covariates. The best model was selected using a backward stepwise approach based on the Akaike Information Criterion (AIC) value. Criteria following a smaller-is-better form, meaning that the combination of covariates producing the lowest AIC value was selected for the model equation. The prediction model was then assessed on its MI mean prediction accuracy for each genotype combination within a 95% confidence interval compared to the actual recorded MI mean.

3.3 Results

3.3.1 Genotype frequencies

The calculated allele and genotype frequencies for each SNP as per SA population groups are listed in Table 3.3.1-1. The genotype and allele frequencies observed for *SLC24A5* (rs1426654) and *SLC45A2* (rs16891982) within the cohort were found to not be in Hardy-Weinberg Equilibrium.

Table 3.3.1–1 Allele and genotype frequencies for *SLC24A5* (rs1426654) and *SLC45A2* (rs16891982) according to self-identified SA population groups

	***SLC24A5 rs*1426654**					***SLC45A2 rs*16891982**				
	Allele		**Genotype**			**Allele**		**Genotype**		
	A	**G**	**AA**	**GA**	**GG**	**G**	**C**	**GG**	**GC**	**CC**
African Black (n=29)	0.10	0.90	0.00	0.21	0.79	0.10	0.90	0.00	0.21	0.79
Coloured (n=14)	0.29	0.71	0.07	0.43	0.50	0.36	0.64	0.07	0.57	0.36
Indian/Asian (n=8)	0.44	0.56	0.25	0.38	0.38	0.25	0.75	0.00	0.50	0.50
White (n=39)	0.96	0.04	0.92	0.08	0.00	0.90	0.10	0.82	0.15	0.03
Other (n=2)	0.50	0.50	0.50	0.00	0.50	0.00	1.00	0.00	0.00	1.00
TOTAL (n=92)	**0.53**	**0.47**	**0.43**	**0.2**	**0.37**	**0.42**	**0.58**	**0.36**	**0.13**	**0.51**
Sub Saharan African HapMap-YR1 (n=113)* (n=57)**	0.01	0.99	0.00	0.03	0.97	0.00	1.00	0.00	0.00	1.00
Asian HapMap –HCB (n=43)* (n =44)**	0.01	0.99	0.00	0.02	0.98	0.01	0.99	0.00	0.02	0.98
European HapMap-CEU (n=58)* (n=58)**	1.00	0.00	1.00	0.00	0.00	0.98	0.02	0.97	0.03	0.00

African AFR (n=661)^	0.07	0.93	n/a	n/a	n/a	0.04	0.96	n/a	n/a	n/a
South Asian SAS (n=489)^	0.69	0.31	n/a	n/a	n/a	0.06	0.94	n/a	n/a	n/a
East Asian EAS (n=504)^	0.01	0.99	n/a	n/a	n/a	0.01	0.99	n/a	n/a	n/a
European EUR (n=503)^	1.00	0.00	n/a	n/a	n/a	0.94	0.06	n/a	n/a	n/a

Frequencies accessed from http://www.ncbi.nlm.nih.gov

1000Genome reference numbers:

***ss43742806 (*SLC24A5* rs1426654) and **ss44669869 (*SLC45A2* rs16891982)**

^ss1353359076 (*SLC24A5* rs1426654) and ss1314583040 (*SLC45A2* rs16891982)

The International HapMap Project and the 1000Genome Project are databases containing allele and genotype frequencies from different population groups in the world. HapMap is a catalogue of SNPs, representing a collaboration of scientists from public and private organisations in six countries and together have assembled data from 11 global ancestry groups [111]. The 1000Genome Project has samples individuals from 26 populations in Africa (AFR), East Asia (EAS), Europe(EUR), South Asia (SAS) and the America (AMR). Individuals were sequenced using both whole genome sequencing and targeted exome sequencing [112]. The information is freely accessible worldwide and is thus a great asset to research.

3.3.2 Hypothesis testing

The average MI per individual grouped according to their observed genotype for both SNPs is displayed in Figure 3.3.2-1. Genotypes observed in *SLC24A5* rs1426654 exhibited a mean MI of 32.99 for genotype AA (n = 44), mean MI of 49.43 for genotype GA (n=18) and a mean MI of 53.78 for genotype GG (n=34). *SLC45A2* rs16891982 exhibited a mean MI of 55.42 for genotype CC (n=47), mean of 38.43 for genotype GC (n=12) and a mean of 32.29 for genotype GG (n =33).

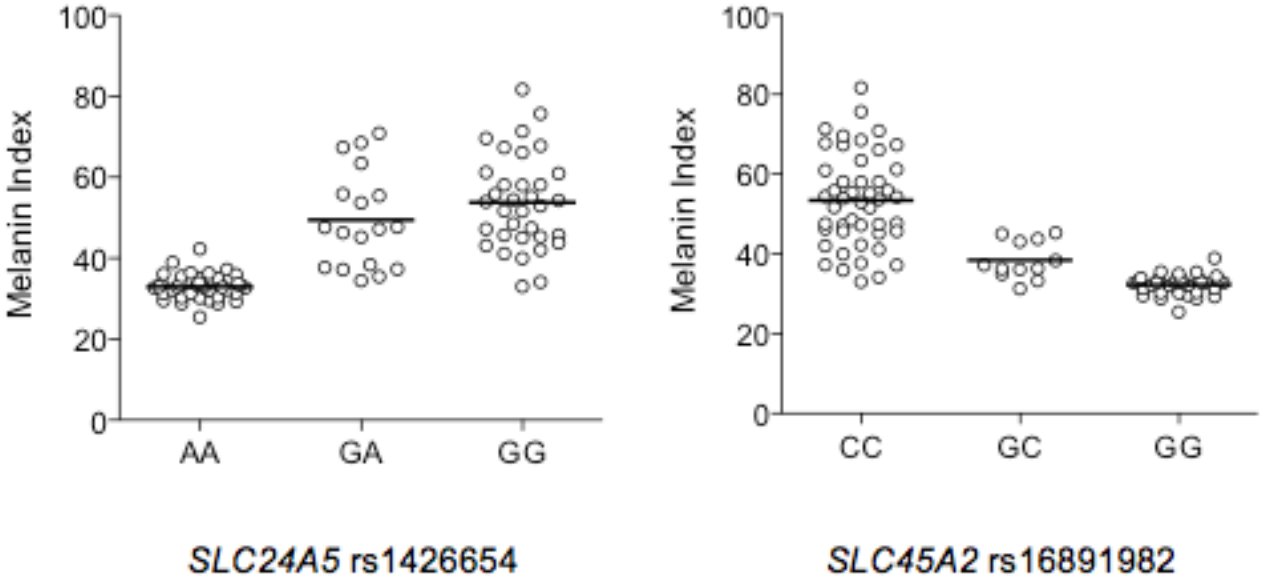

Figure 3.3.2–1 Graphs depicting average MI measurement for each individual (each represented by a circle) grouped according to the genotype observed in each SNP. *SLC24A5* rs1426654 containing alleles A and G; *SLC45A2* rs16891982 containing alleles G and C. The horizontal bar represents the mean MI reading per genotype.

The eight observed genotype combinations from both SNPs and their MI distribution range are shown in Figure 3.3.2–2. MI averages from each participant were used to calculate the mean MI for each observed genotype

combination. The mean MI for each observed combined genotype was: **AA** *(SLC24A5* rs1426654*)* **CC** *(SLC45A2* rs16891982*)* = 38.49 (n=3); **AA** *(SLC24A5* rs1426654*)* **GC** *(SLC45A2* rs16891982*)* = 34.72 (n=6); **AA** *(SLC24A5* rs1426654*)* **GG** *(SLC45A2* rs16891982*)* = 32.12 (n=31); **GA** *(SLC24A5* rs1426654*)* **CC** *(SLC45A2* rs16891982*)* = 53.16 (n=14); **GA** *(SLC24A5* rs1426654*)* **GC** *(SLC45A2* rs16891982*)* = 37.79 (n=2); **GA** *(SLC24A5* rs1426654*)* **GG** *(SLC45A2* rs16891982*)* = 34.97 (n=2); **GG** *(SLC24A5* rs1426654*)* **CC** *(SLC45A2* rs16891982*)* = 55.04 (n=30) and **GG** *(SLC24A5* rs1426654*)* **GC** *(SLC45A2* rs16891982*)* = 44.32 (n=4). The combination of **GG** *(SLC24A5* rs1426654*)* **GG** *(SLC45A2* rs16891982*)* was not observed in the cohort.

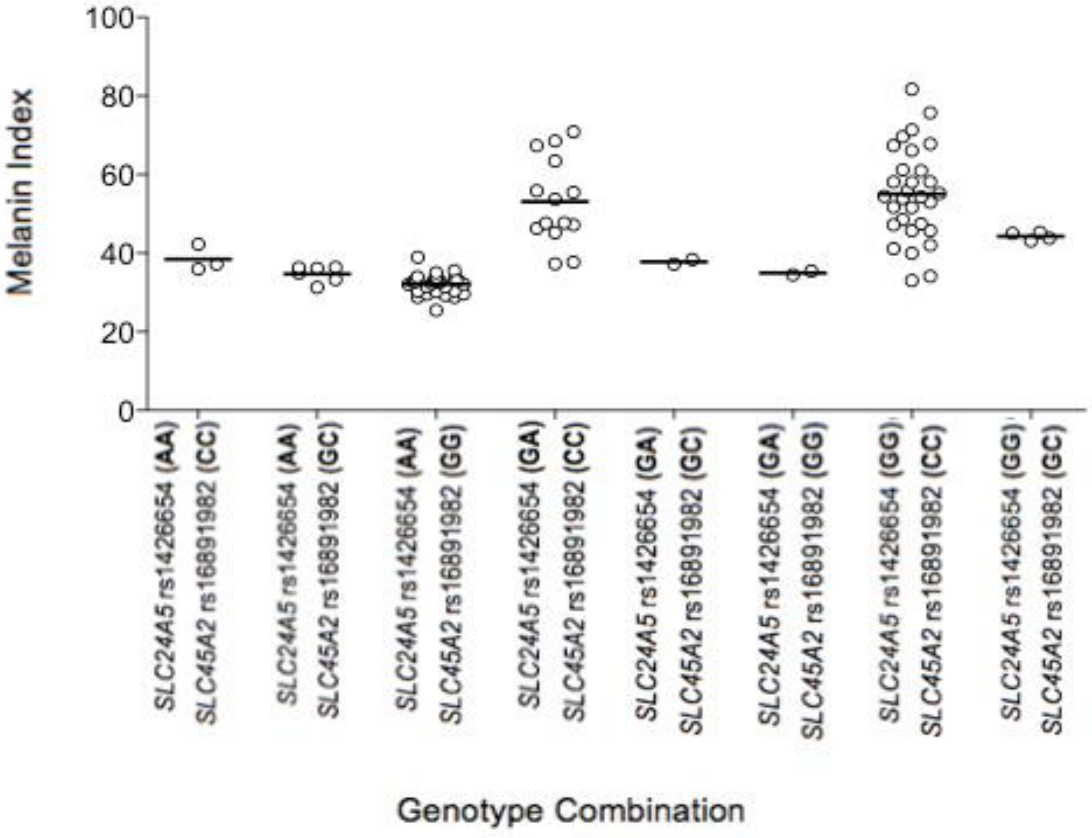

Figure 3.3.2–2 Graph depicting average MI measurement for each individual (represented by a circle) grouped according to the observed combined genotype for both SNPs. *SLC24A5* rs1426654 containing alleles A and G; *SLC45A2* rs16891982 containing alleles G and C. The horizontal bar represents the mean MI reading per combined genotype.

The data was not normally distributed and thus the Kruskal-Wallis test, a non-parametric statistical test; was used to determine if there was a significant difference between the observed combined genotypes for *SLC24A5* rs1426654 and *SLC45A2* rs16891982 amongst the different self-identified SA population census groups. The individual comparison results can be seen in Table 3.3.2-1. Only three group comparisons showed a statically significant difference, Indian/Asian compared to Black African ($p = 0.04$); White compared to Black African ($p < 0.0001$) and White compared Coloured ($p < 0.0001$).

Table 3.3.2–1 Comparison of observed combined genotypes between the different self-identified SA population census groups

SA vs Genotype	African Black	Coloured	Indian/ Asian	White	Other
African Black	-				
Coloured	p=0.740 SE= 0.473	-			
Indian/ Asian	**p=0.040 SE=0.580**	p=0.474 SE=0.644	-		
White	**p<0.0001 SE=0.356**	**p<0.0001 SE=0.453**	p=0.169 SE=0.564	-	
Other	p= 0.294 SE=1.062	p=0.657 SE=1.098	p=0.996 SE=1.148	p=0.936 SE=1.053	-

***Shaded blocks indicate significant difference (p<0.05) results **SE = standard error**

The average MI distribution of the cohort is represented in Figure 3.3.2 – 3 and is separated into two parts. Part A shows the MI average of the individual grouped according to self-identified ancestral group, and the mean MI for each ancestral group was calculated and results were as follows: African = 60.04 (range: 41.12 – 81.74); Asian = 41.51 (range: 33.08 – 47.71); European = 33.44 (range: 25.46 – 47,60) and Mixed = 42.4 (range: 32.79 – 55.85). Part B shows the MI average of the individual grouped according to self-identified SA population census groups and the mean MI for each group was calculated and results are as follows: African Black = 60.11 (range: 41.12 – 81.74); Coloured = 42.84 (range: 32.79 – 53.68); Indian/ Asian = 41.51 (range: 33.08 – 47.71); White =32.82 (range: 25.46 – 38.96) and Other = 41.6 (range 35.96 – 47.24). It is important to note that the SA population group 'Other' only included two individuals.

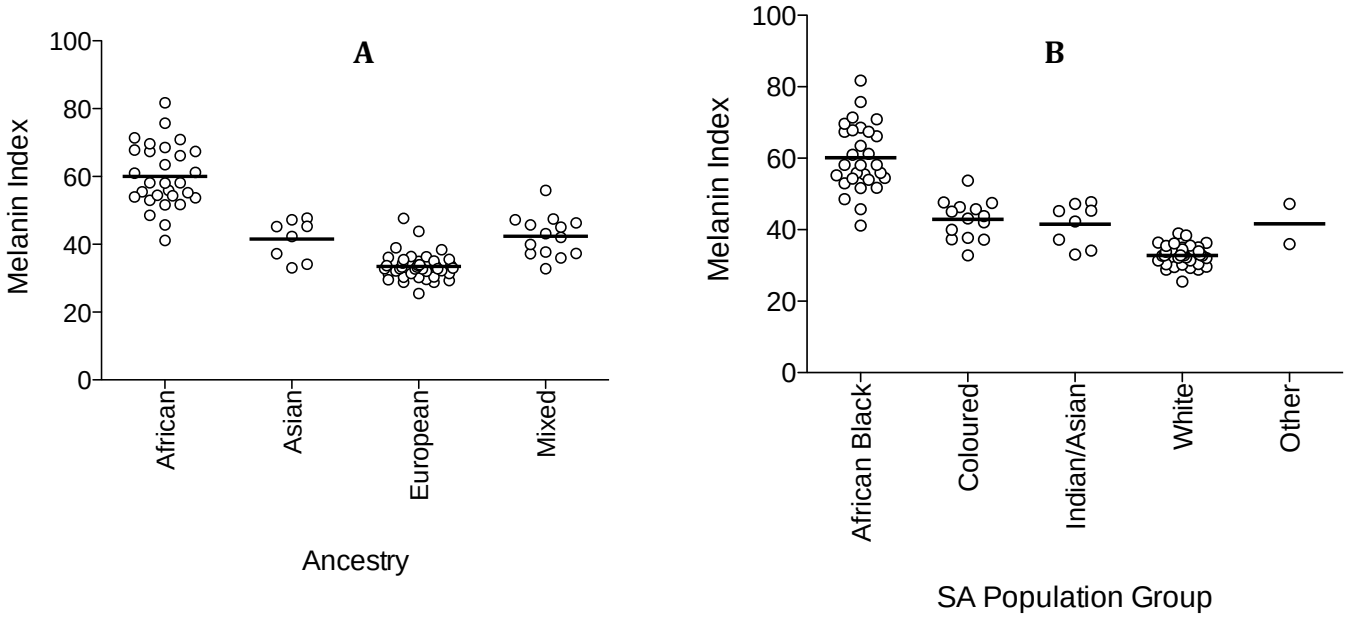

Figure 3.3.2–3 Graphs depicting average MI measurement for each individual (represented by a circle) grouped according to self-identified Ancestry in Part A, which includes African (n=29), Asian (n=8), European (n=41) and Mixed (n=14). Part B is grouped according to self-identified SA population group according to the census and included African Black (n=29), Coloured (n=14), Indian/Asian (n=8), White (n=39) and Other (n=2). The horizontal bar in each part represents the mean MI reading per ancestry and population group, respectively.

The Kruskal-Wallis test (with the application of the Bonferroni correction) was used to calculate any significant differences between the mean MI grouped according to different ancestral groups as well as between different self-identified South African population groups. Pairwise results of each group can be seen in Table 3.3.2 -2 and Table 3.3.2 -3 respectively. It was observed that all ancestral groups displayed a mean MI significantly different between each group except for Mixed ancestry compared to Asian ancestry (p = 0.992). Self-identified South African populations also showed significant differences between the mean MI calculated for most groups, excluding the group identifying as 'Other' which was only significantly different to Black African (p = 0.004), however it must be taken into consideration that the group 'Other' only consisted of two individuals. The group Indian/Asian was also not significantly different from the Coloured group (p =0.994).

Table 3.3.2–2 Mean MI comparisons between different ancestral groups

MI vs. Ancestry	African	Asian	European	Mixed
African	-			
Asian	**p<0.0001 SE= 2.616**	-		
European	**p<0.0001 SE=1.589**	**p=0.012 SE=2.531**	-	
Mixed	**p<0.0001 SE=2.903**	p=0.992 SE=2.903	**p<0.0001 SE=2.027**	-

***Shaded blocks indicate significant difference (p<0.05) results** ****SE = standard error**

Table 3.3.2–3 Mean MI comparisons between the different SA population census groups

SA vs. MI	African Black	Coloured	Indian/ Asian	White	Other
African Black	-				
Coloured	**p<0.0001 SE= 2.035**	-			
Indian/ Asian	**p<0.0001 SE=2.497**	p=0.994 SE=2.771	-		
White	**p<0.0001 SE=1.533**	**p<0.0001 SE=1.948**	**p=0.016 SE=2.426**	-	
Other	**p= 0.004 SE=4.571**	p=0.999 SE=4.726	p=1.00 SE=4.942	p=0.445 SE=4.533	-

***Shaded blocks indicate significant difference (p<0.05) results** ****SE = standard error**

3.3.3 MI predictive statistical model

The generalised linear model which best fit the cohort consisted of the intercept value of 0.029 ($p < 0.0001$) and covariates were selected based on their effect on the AIC value, shown in Table 3.3.3 - 1. The final model equation is shown below (Equation 1) and is portrayed to identify the specific individual SNP genotype and specific genotype combination selected as covariates. The coefficient values to be used in the equation for each covariant is recorded in Table 3.3.3 – 1.

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

Mean MI = [0.029** + (SLC24A5 rs1426554: genotype AA) + (SLC45A2 rs16891982: genotype CC) + (SLC45A2 rs16891982: Genotype GC) + (Combined genotype AAGG) + (Combined genotype GAGC)**] $^{-1}$

Table 3.3.3–1 General linear model variables and values

Covariate	Coefficient	Standard error	P-value
Intercept	0.029	0.003	p<0.0001
AA (*SLC24A5* rs1426654) alleles A/G	0.007	0.002	p<0.0001
CC (*SLC45A2* rs16891982) alleles G/C	-0.010	0.003	p=0.001
GC (*SLC45A2* rs16891982*)* alleles G/C	-0.006	0.003	p=0.062
Combination **AA** (*SLC24A5* rs1426654) **GG** (*SLC45A2* rs16891982)	-0.004	0.004	p=0.221
Combination **GA** (*SLC24A5* rs1426654) **GC** (*SLC45A2* rs16891982)	0.004	0.003	p=0.187

Model evaluation was performed through the comparison of the model predicted MI mean according to genotype combination and the actual MI mean observed for each genotype combination. This is illustrated in Figure 3.3.3–1.

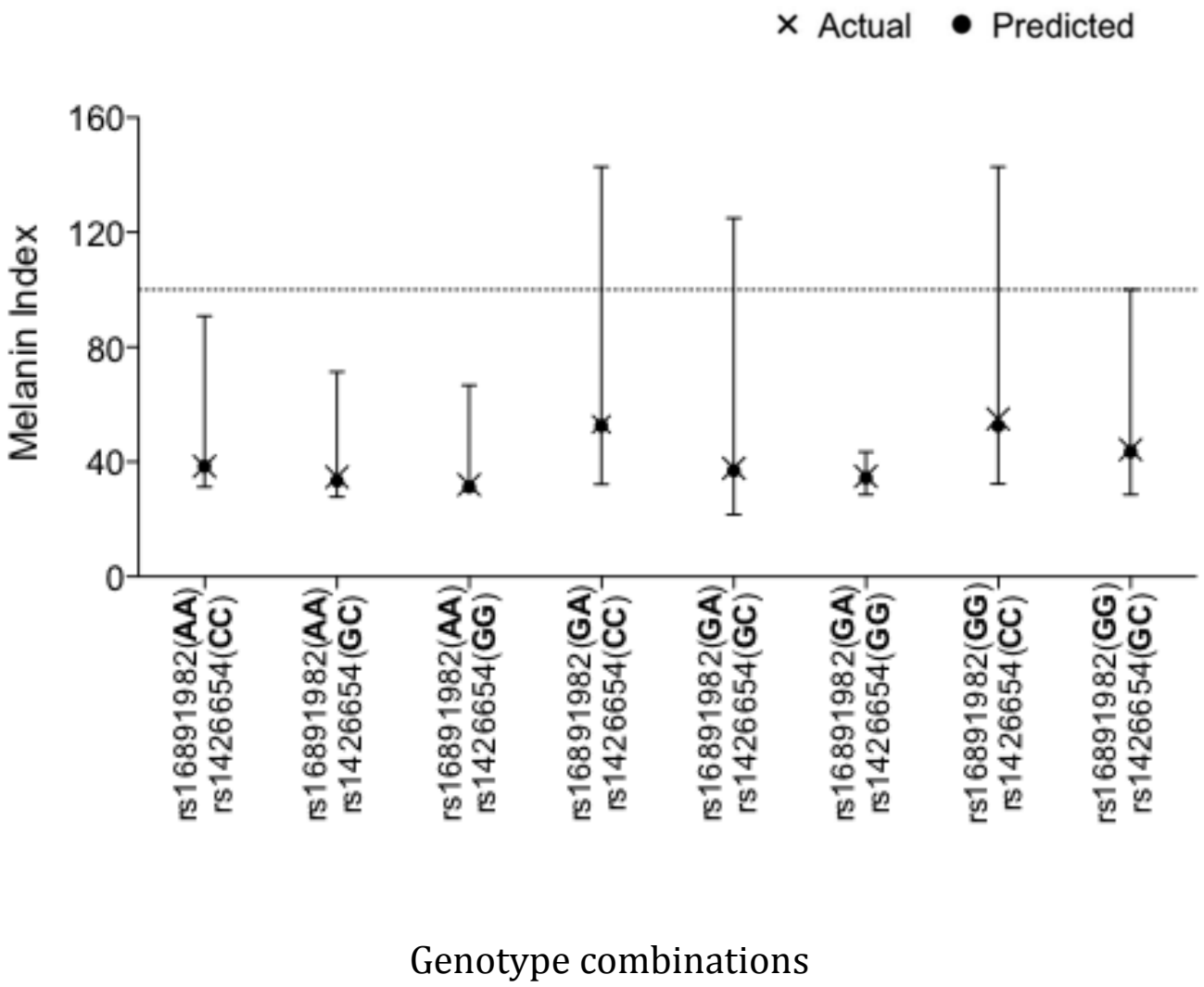

Figure 3.3.3–1 Graph depicting the actual and predicted mean MI for each combined genotype with 95% confidence bands for each combination. Ellipsis line denotes the reported maximum MI for human skin pigmentation (MI=100).

An association between the predicted mean MI and actual mean MI for each genotype group is observed. However, it is seen that the 95% confidence range for certain combinations is very wide and three even exceed the maximum MI measurement units reported for human pigmentation. These include genotype combination **GA** (*SLC24A5* rs1426654) **CC** (*SLC45A2* rs16891982) and genotype combination **GG** (*SLC24A5* rs1426654) **CC** (*SLC45A2* rs16891982); the two combinations with the greatest range of MI measurements as well as genotype combination **GA** (*SLC24A5*rs1426654) **GC** (*SLC45A2* rs16891982), which was only observed in two individuals.

3.4 Discussion

The SA population is considered to be highly diverse due to its multi-racial and admixed population groups, with individuals originating from African, Asian, European and mixed ancestries [37]. With the development of ancestral inference and molecular phenotyping through genetic analysis, the application of this molecular technique within forensic science is being more readily investigated [6,25,42,44,110,113–115]. For effective application in forensic science cases, it is of great importance to discern between individuals of different ancestries both accurately and reliably. As previously discussed, certain SNPs associated with MI and ancestry have shown to vary significantly between different population groups, while others have been less informative, especially in admixed population groups [22,69]. A previous study showed that out of eighteen regularly used AIMs associated with human skin colour, only two were informative in a Brazilian admixed population group [22]. Given the admixture present in the SA population, this study attempted to evaluate the ability of these two SNPs (*SLC24A5* rs1426654 and *SLC45A2* rs16891982) to predict MI of SA individuals of different ancestries.

3.4.1 Prediction model

To investigate this aim, a molecular assay was designed and the two SNPs were genotyped in a cohort of 92 SA individuals who exhibited a range of melanin indices (MI range: 25.46 – 81.74). A prediction model for MI was created based on the genotypes at the two loci under study. The best model, selected using a backward stepwise approach, was assessed on its prediction accuracy for each

genotype combination within a 95% confidence interval compared to the actual mean MI. The resulting model included genotypes and genotype combinations, which, based on the input data, best predicted an individual's MI based on their genotypes at the two loci (Table 3.3.3–1).

However, the small cohort was a limitation of the study, particularly for the statistical model creation. For example, the genotype combination *SLC24A5* rs1426654 **GA** and *SLC45A2* rs16891982 **GG** was only present in two individuals and was not selected for the prediction model; it remains unclear whether this combination was not selected due to small number of observations or whether the combination does not influence MI significantly. Thus, true informative value to predict MI remains unclear. Further, only two genotype combinations were included in the final model, both having a relatively low MI association. This may be due to small number of observations of these combinations, or perhaps because their combined impact on MI is not linked.

The model was then used to predict mean MI values for each observed genotype combination and performance was evaluated through the comparison against actual recorded MI. From Figure 3.3.3–1, it's seen that the predicted MI units overlapped the actual mean MI; however, the 95% confidence interval bands spanned a considerably wide range. Particularly, three of the combined genotypes **GA** (*SLC24A5* rs1426654) **CC** (*SLC45A2* rs16891982); genotypes **GA** (*SLC24A5* rs1426654) **GC** (*SLC45A2* rs16891982) and genotypes **GG** (*SLC24A5* rs1426654) **CC** (*SLC45A2* rs16891982) had upper prediction points exceeding the

human skin pigmentation range, which is certainly not ideal, and renders the model less useful to predict MI for forensic applications. These three genotype combinations were present in almost half of the cohort (n=38), and represented individuals from all four major population groups in SA. As such, two aspects emerged: firstly, these markers alone did not hold sufficient discriminatory power to accurately predict the MI of most individuals in SA, and as such, additional markers, as well as a larger cohort, would be required to improve the model to predict MI in the SA population. This is supported by previous research, which indicated that markers previously associated with pigmentation phenotypes in homogenous population groups were less informative in admixed population groups [69].

Secondly, it emerged that SA population census groups were not a suitable proxy for MI, as individuals from different populations exhibited such wide and overlapping ranges of MI. Individuals that identified as "African Black" had an extensive MI range from 41.12 and 81.74, which overlapped with the MI range of "Coloured" individuals (MI range: 32.79 – 53.68). This was also noted by Brown (2000) who reported that individuals who self-identified with the "Black" and "Coloured" South African population groups varied extensively in pigmentation levels [88]. While many previous studies of similar nature have used population group as the variable to which genotypes were associated [107,116,117], it is suggested here that a continuous MI is a more suitable variable for SA population groups which often exhibit admixture. The continuous MI scale is further

deemed more appropriate within the SA context due to the prejudicial racial history of SA and the classification categories that it produced [104].

Statistical analysis for the prediction of phenotypes from genotypes has often been incorporated through the generation of logistic regression models based on statistically relevant SNPs and their probability to predict characteristics [118]. This has most frequently and accurately been accomplished with hair colour and eye colour [46,73]. Models are evaluated on the number of correct calls/classifications as well as the percentage of error. Most models are categorical-based, in that combinations of genotypes are used to predict a set category of phenotype(s). For example, categories for eye colour were “blue”, “brown” and “not blue” and the model would predict what category was most appropriate based on the individual’s genotype(s). Other previously established models to predict skin colour have also categorised skin colour using colour charts, and resulted in somewhat arbitrary and restricted categorisations of skin colour, such as “light skin” or “dark skin” colour or even “fair skin” as opposed to “olive skin” [7,110,116,119,120]. These categories have limitations in forensic applications due to their subjective nature. This contrasted with the model designed here, which was used to predict the continuous variable, MI, which is deemed a more appropriate measure of skin colour, than pre-defined categories or population groups, especially within a forensic setting.

3.4.2 Allele frequencies

Allele frequencies *SLC24A5* rs1426654 and *SLC45A2* rs16891982 were established within the four major population groups in SA and were compared to data from the HapMap3 (http://www.hapmap.org) and 1000Genomes (http://www.1000genomes.org/data). The 100% frequency of the homozygous A/A genotype in the *SLC24A5* rs1426654 SNP in the European population groups reported on HapMap3 and 1000Genomes supports the use of this marker to infer European ancestry in homozygous A/A individuals. However, while the allele frequencies were similar between the SA "White" population group in this study (n=39) and the European groups reported on HapMap3 (n=58) and 1000Genomes (n=503) (Table 3.3.1–1) the SA "White" population group exhibited heterogeneity in *SLC24A5* rs1426654, despite its comparatively smaller sample size. While only three SA "White" individuals of European ancestry were heterozygous A/G for *SLC24A5* rs1426654, this may have a potential impact on the suitability of this marker to identify SA individuals of European descent – previously, only homozygous A/A genotypes at *SLC24A5* rs1426654 were observed in individuals of European descent (Table 3.3.1–1), which made *SLC24A5* rs1426654 A/A an informative marker and genotype to infer European ancestry. If an A/A was observed, it was likely that the individual was of European descent with light skin [30]. Conversely, if G/G or A/G, were observed, it was likely that the individual was of African ancestry and exhibited darker skin colour (Table 3.3.1–1 and Figure 3.3.2–3) [30].

However, the results here showed that 3/39 (7.69%) SA individuals of European descent (average MI = 36.12) exhibited heterozygosity (A/G) at this marker. Further, several SA individuals of Asian and Mixed ancestries with varying melanin indices also exhibited the traditionally European A/A genotype at this marker (Table 3.3.1-1 and Figure 3.3.2 -3). The observation of A/A genotypes in individuals from ancestries other than European, as well as the observation of SA "White" individuals of European descent with the heterozygous A/G genotype, ultimately lessens the ability which *SLC24A5* rs1426654 can infer about ancestry in SA individuals, particularly if this marker is considered in isolation.

The combination of genotypes at both SNPs and their combined association with skin pigmentation is shown in Figure 3.3.2–2. It was observed that two genotype combinations, namely (i) *SLC24A5* rs1426654 **G/A** and *SLC45A2* rs16891982 **C/C** and (ii) *SLC24A5* rs1426654 **G/G** and *SLC45A2* rs16891982 **C/C**, displayed a wide range of MI units 37.31 – 70.85 and 34.09 – 81.74, respectively, and as such, were considered as potentially uninformative to predict MI in this study. It is important to take into consideration that a total of 44/92 (47.83%) individuals from this study exhibited these two genotypes, and these individuals were from all SA population groups. Individuals with combination (i) were of European, African and Mixed ancestries, while those individuals with combination (ii) were of Asian, African as well as Mixed ancestries. This variation in ancestry within the same observed genotype combination again illustrates the potential difficulty that these two markers alone, will have in inferring the MI of SA individuals.

Considering that these two markers were selected because they were the only two markers previously associated with MI within a Brazilian admixed population [22], it raises the query as to why they were not significantly associated in the study cohort. As previously discussed this may be a consequence of a small cohort and an increase in sample size may show an increased discriminatory power of the two SNPs. It is also possible that due to the complex melanin pathway, the gene signals could be population specific as was demonstrated in a study by Norton et al. (2007) where the same SNPs that influenced the lightening of European skin colour did not affect East Asians [121]. Additionally, this discrepancy can be explained through the different admixture structure of the Brazilian population and the study cohort. Admixed Brazilian populations have varying contributions from ancestral groups such as Native American, European (Portuguese, Italian and Middle East) and Northern African [22,122,123]. The South African population derives from Khoesan Africans, non-Khoesan Africans (including East African, Buntu and Pygmy populations), European (German, French and Dutch) and Asian ancestral groups [87]. As previously discussed, certain SNPs associated with MI and ancestry have shown to vary between different population groups, particularly concerning admixed populations. This points out the relevance of studies investigating appropriate SNPs with significant discriminatory power within specific populations [22,69].

3.4.3 Limitations and future research

A limitation that needs to be taken into consideration when evaluating the results reported by this project is the small study sample size. The impact of this has been discussed above. As such, a larger cohort is required to improve the prediction model. A power calculation was performed, which indicated 362 individuals should form part of this study. However, this number far exceeded the scope of this minor dissertation, and as such, only one quarter of this cohort was included in this pilot study. This was deemed sufficient to demonstrate the appropriate methodology, which could then be reproduced in a larger cohort at a later stage.

It should also be noted that the sample population exhibited an overrepresentation of self-identified "White" SA individuals with only fourteen individuals self-identifying as "Coloured", eight self-categorising themselves within the "Indian/Asian" population group and two within "Other". These demographics were not representative of those of the country's population, which may have influenced the representation of certain MI frequencies included in this study. Again, it is suggested that future research should seek to involve a larger sample size and have better representation of these population groups to verify the results presented here.

To improve the model, additional SNPs need to be included. Future research should perhaps focus on the identification of SNPs with increased relevance in different population groups especially for application within forensic science.

Melanin is the main pigment of eye, hair and skin colour and relies on complex biochemical pathways [43,49]. Several other genes relating to melanogenesis, the synthesis and regulation process of melanin [75,76], have been identified in former studies. These genes include *Melanocortin 1 receptor* (*MC1R*), *Oculocutaneous albinism* (*OCA*), *Hect domain and RCC1-like domain 2* (*HERC2*), and *Tyrosinase-related protein* (*TYRP*) [7,69,75,107]. These need to be reviewed regularly in order to improve prediction models as based on selected and designed SNP panels [15,51,60]. This research can help to optimize, verify and validate panels to potentially be included in kits, which would be locally relevant in the SA context.

3.5 Conclusion

This research is among the first investigations into the use of SNPs to predict skin pigmentation within the SA population for forensic science applications. It explored the use of MI as an objective skin colour measurement, instead of the use of traditional racial categorisation, which carries prejudice in the SA context. Two SNPs, *SLC24A5* rs1426654 and *SLC45A2* rs16891982, were selected due their previous significant association with skin colour in Brazilian admixed populations. A multiplex SNaPshot PCR assay was designed to genotype these SNPs and a statistical model was generated using the genotypes and MI from participants.

The model was used to predict an individual's MI using their combined genotypes. The predicted MI was compared to the actual MI, and although the model was able to predict MI, the standard of error was deemed too large for practical forensic use within SA. This standard of error increased for genotype combinations, which were linked to individuals with higher MI. Some genotypes were excluded from the general linear model (GLM), perhaps due to the small number of observations of that genotype within the cohort. These genotypes may be rare, or this observation may be an anomaly due to the small cohort used in this pilot project. These two SNPs alone do not possess enough discriminatory power to predict MI within the SA population, but the cohort should be suitably expanded in future studies. To improve the model and make it more robust, the methods need to be replicated on a larger scale and include more markers, which are informative in admixed populations.

Despite limitations, this pilot project showed that these two SNPs alone were insufficient to predict MI within the SA population, and laid the foundation for methodology for future studies when cohort and number of markers are expanded. It was also novel in its use of a GLM to predict a *continuous scale* of MI in the SA population. Given the history of SA this method is deemed more appropriate as it steers away from prejudicial racial classifications and the use of a continuous objective variable is more informative for forensic purposes.

References

[1] S.J. Walsh, Legal perceptions of forensic DNA profiling, Forensic Sci. Int. 155 (2005) 51–60.

[2] C. Porter, The Forensic Use of DNA, Aust. J. Forensic Sci. 37 (2005) 5–8.

[3] J. Gans, G. Urbas, DNA identification in the criminal justice system, Aust. Inst. Criminol. 226 (2002) 1–6.

[4] M. Lynch, God's signature: DNA profiling, the new gold standard in forensic science, Endeavour. 27 (2003) 93–97.

[5] M. Kayser, Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes, Forensic Sci. Int. Genet. 18 (2015) 33–48.

[6] M. Kayser, P.M. Schneider, DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations, Forensic Sci. Int. Genet. 3 (2009) 154–161.

[7] O. Maroñas, C. Phillips, J. Söchtig, A. Gomez-Tato, R. Cruz, J. Alvarez-Dios, et al., Development of a forensic skin colour predictive test, Forensic Sci. Int. Genet. 13 (2014) 34–44.

[8] J.M. Butler, Fundamentals of Forensic DNA typing, 2nd ed., Elsevier/Academic Press, San Diego, California, 2010.

[9] J.M. Butler, Forenic DNA Typing, 2nd ed., Elsevier/Academic Press, Burlington, 2005.

[10] J.M. Butler, Advanced Topics in Forensic DNA Typing: Methodology, 1st ed., Elsevier/Academic Press, 2011.

[11] K.R. Birikh, P.L. Bernad, V. V Shmanai, A.D. Malakhov, M.S. Shchepinov, V. a Korshun, SNP detection using trityl mass tags., Methods Mol. Biol. 578 (2009) 345–361.

[12] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis. 27 (2006) 1713–1724.

[13] J.M. Butler, M.D. Coble, P.M. Vallone, STRs vs. SNPs: Thoughts on the future of forensic DNA testing, Forensic Sci. Med. Pathol. 3 (2007) 200–205.

[14] M.D. Shriver, R. a Kittles, Genetic ancestry and the search for personalized genetic histories., Nat. Rev. Genet. 5 (2004) 611–618.

[15] K.K. Kidd, A.J. Pakstis, W.C. Speed, E.L. Grigorenko, S.L.B. Kajuna, N.J. Karoma, et al., Developing a SNP panel for forensic identification of individuals, Forensic Sci. Int. 164 (2006) 20–32.

[16] S. Walsh, A. Lindenbergh, S.B. Zuniga, T. Sijen, P. de Knijff, M. Kayser, et al., Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence, Forensic Sci. Int. Genet. 5 (2011) 464–471.

[17] I. Halder, M. Shriver, M. Thomas, J.R. Fernandez, T. Frudakis, A panel of ancestry informative markers for estimating individual biogeographical

ancestry and admixture from four continents: Utility and applications, Hum. Mutat. 29 (2008) 648–658.

[18] N.G. Jablonski, The Evolution of Human Skin and Skin Color, Annu. Rev. Anthropol. 33 (2004) 585–623.

[19] R.A. Sturm, Molecular genetics of human pigmentation diversity, Hum. Mol. Genet. 18 (2009) 9–17.

[20] M. van der Wal, M. Bloemen, P. Verhaegen, W. Tuinebreijer, H. de Vet, P. van Zuijlen, et al., Objective Color Measurements, J. Burn Care Res. 34 (2013) 187–194.

[21] S. Taylor, W. Westerhof, S. Im, J. Lim, Noninvasive techniques for the evaluation of skin color, J. Am. Acad. Dermatol. 54 (2006) 282–290.

[22] C.C.S. De Cerqueira, T. Hünemeier, J. Gomez-Valdés, V. Ramallo, C.D. Volasko-Krause, A.A.L. Barbosa, et al., Implications of the Admixture Process in Skin Color Molecular Assessment, PLoS One. 9 (2014) e96886.

[23] H. Takiwaki, Y. Miyaoka, H. Kohno, S. Arase, Graphic analysis of the relationship between skin colour change and variations in the amounts of melanin and haemoglobin, Skin Res. Technol. 8 (2002) 78–83.

[24] C.V. Castel, Inference of Biogeographical Ancestry and Pigmentation Phenotype using Single Nucleotide Polymorphisms, University of Technology, Sydney, Australia, 2014.

[25] E. Giardina, I. Pietrangeli, C. Martínez-Labarga, C. Martone, F. de Angelis, A. Spinella, et al., Haplotypes in SLC24A5 Gene as Ancestry Informative Markers in Different Populations, Curr. Genomics. 9 (2008) 110–114.

[26] R.L. Lamason, M.-A.P.K. Mohideen, J.R. Mest, A.C. Wong, H.L. Norton, M.C. Aros, et al., SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans, Science. 310 (2005) 1782–1786.

[27] K. Inagaki, T. Suzuki, S. Ito, N. Suzuki, K. Adachi, T. Okuyama, et al., Oculocutaneous albinism type 4: Six novel mutations in the membrane-associated transporter protein gene and their phenotypes, Pigment Cell Res. 19 (2006) 451–453.

[28] Parliament of Republic of South Africa, Government Gazette of the Republic of South Africa, 2014.

[29] L.J. Heathfield, Policy required for entry of DNA profiles onto the National Forensic DNA Database of South Africa, S. Afr. J. Sci. 110 (2014) 8–10.

[30] Statistics South Africa, Statistical release (Revised) Census 2011, (2012) 78. www.statssa.gov.za.

[31] L. Naing, T. Winn, B.N. Rusli, Practical Issues in Calculating the Sample Size for Prevalence Studies, Arch. Orofac. Sci. 1 (2006) 9–14.

[32] R.P. Stokowski, P.V.V.K. Pant, T. Dadd, A. Fereday, D. a. Hinds, C. Jarman, et al., A Genomewide Association Study of Skin Pigmentation in a South Asian Population, Am. J. Hum. Genet. 81 (2007) 1119–1132.

[33] O. Trujillo, P. Vanezis, M. Cermignani, Photometric assessment of skin colour and lightness using a tristimulus colorimeter: Reliability of inter

and intra-investigator observations in healthy adult volunteers, Forensic Sci. Int. 81 (1996) 1–10.

[34] K.B. Gettings, Forensic Ancestry and Phenotype SNP Analysis and Integration with Established Forensic Markers, The George Washinton University, 2013.

[35] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating, 74 (1977) 5463–5467.

[36] T. Caulfield, S.M. Fullerton, S.E. Ali-Khan, L. Arbour, E.G. Burchard, R.S. Cooper, et al., Race and ancestry in biomedical research: exploring the challenges, Genome Med. (2009) 1- 8.

[37] M. Shriver, T. Frudakis, B. Budowle, Getting the science and the ethics right in forensic genetics, Nat. Genet. 37 (2005) 449-450-451.

[38] J. Fujimura, R. Rajagopalan, Different differences: The use of "genetic ancestry' versus race in biolmedical human genetic research, Soc. Study Sci. 41 (2012) 1–22.

[39] M.I. González-Tapia, I. Obsuth, "Bad genes" & criminal responsibility, Int. J. Law Psychiatry, 39 (2015) 60–71.

[40] M. Binstock, F. Hafeez, C. Metchnikoff, S.T. Arron, Single-nucleotide polymorphisms in pigment genes and nonmelanoma skin cancer predisposition: a systematic review, Br. J. Dermatol. 171 (2014) 713–21.

[41] H.J.H. Edgar, Microevolution of African American dental morphology, Am. J. Phys. Anthropol. 132 (2007) 535–544.

[42] H. Pulker, M.V. Lareu, C. Phillips, A. Carracedo, Finding genes that underlie physical traits of forensic interest using genetic tools, Forensic Sci. Int. Genet. 1 (2007) 100–4.

[43] A. Branicki, Wojciech; Brudnik, Urszula: Wojas-Pelc, Genetic Prediction of Pigmentary Traits in Forensic Studies, Probl. Forensic Sci. 64 (2005) 343–357.

[44] K. Butler, M. Peck, J. Hart, M. Schanfield, D. Podini, Molecular "eyewitness": Forensic prediction of phenotype and ancestry, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) 498–499.

[45] S. Walsh, F. Liu, K.N. Ballantyne, M. van Oven, O. Lao, M. Kayser, IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information, Forensic Sci. Int. Genet. 5 (2011) 170–180.

[46] S. Walsh, L. Chaitanya, L. Clarisse, L. Wirken, J. Draus-Barini, L. Kovatsi, et al., Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage, Forensic Sci. Int. Genet. 9 (2014) 150–161.

[47] M. Kayser, P. de Knijff, de K, Improving human forensics through advances in genetics, genomics and molecular biology, Nat. Rev. Genet. 12 (2011) 179–192.

[48] C. Phillips, Forensic genetic analysis of bio-geographical ancestry, Forensic Sci. Int. Genet. 18 (2015) 49–65.

[49] W. Branicki, Studies on Predicting Pigmentation Phenotype, Probl. Forensic Sci. 77 (2009) 29–52.

[50] A.J. Pakstis, W.C. Speed, J.R. Kidd, K.K. Kidd, Candidate SNPs for a universal individual identification panel, Hum. Genet. 121 (2007) 305–317.

[51] B. Budowle, A. van Daal, Forensically relevant SNP classes, Biotechniques, 44 Supplem (2008) 603–610.

[52] C. Romanini, M. Romero, M. Salado, L. Catelli, C. Vullo, Forensic Science International : Genetics Ancestry informative markers : Inference of ancestry in aged bone samples using an autosomal AIM-Indel multiplex, Forensic Sci. Int. Genet. 16 (2015) 58–63.

[53] M. Bamshad, S. Wooding, B.A. Salisbury, J.C. Stephens, Deconstructing the relationship between genetics and race, Nat. Rev. Genet. 5 (2004) 598–609.

[54] J.N. Sampson, K.K. Kidd, J.R. Kidd, H. Zhao, Selecting SNPs to Identify Ancestry, Ann. Hum. Genet. 75 (2011) 539–553.

[55] T.B. Mersha, T. Abebe, Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities, Hum. Genomics. 9 (2015) 1–15.

[56] M.W. Foster, R.R. Sharp, Race, ethnicity, and genomics: Social classifications as proxies of biological heterogeneity, Genome Res. 12 (2002) 844–850.

[57] M. Lillie-Blanton, T. Laveist, Race/ethnicity, the social environment, and health, Soc. Sci. Med. 43 (1996) 83–91.

[58] M. Yudell, D. Roberts, R. DeSalle, S. Tishkoff, Taking race out of human genetics, Science. 351 (2016) 564–565.

[59] S.R. Heathfield LJ, Prins AM, Comparison of Footprint Ridge Density between Two South African Ethnic Groups, J. Forensic Investig. 4 (2016).

[60] K.K. Kidd, W.C. Speed, A.J. Pakstis, M.R. Furtado, R. Fang, A. Madbouly, et al., Progress toward an efficient panel of SNPs for ancestry inference, Forensic Sci. Int. Genet. 10 (2014) 23–32.

[61] J.R. Kidd, F.R. Friedlaender, W.C. Speed, A.J. Pakstis, F.M. De La Vega, K.K. Kidd, Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples, Investig. Genet. 2 (2011) 1- 13.

[62] C.E. MacLean, Creating a Wanted Poster from a Drop of Blood: Using DNA Phenotyping to Generate an Artist's Rendering of an Offender Based Only on DNA Shed at the Crime Scene, Hamline Law Rev. 36 (2013) 357 - 686.

[63] E. Segal, N. Friedman, D. Koller, A. Regev, A module map showing conditional activity of expression modules in cancer, Nat. Genet. 36 (2004) 1090–1098.

[64] K.K. Kidd, Population Genetics of SNPs for Forensic Purposes, 2011.

[65] A. Freire-Aradas, Y. Ruiz, C. Phillips, O. Maroñas, J. Söchtig, a. G. Tato, et al., Exploring iris colour prediction and ancestry inference in admixed

populations of South America, Forensic Sci. Int. Genet. 13 (2014) 3–9.

[66] S. Beleza, N. a. Johnson, S.I. Candille, D.M. Absher, M. a. Coram, J. Lopes, et al., Genetic Architecture of Skin and Eye Color in an African-European Admixed Population, PLoS Genet. 9 (2013).

[67] C. Basu Mallick, F.M. Iliescu, M. Möls, S. Hill, R. Tamang, G. Chaubey, et al., The Light Skin Allele of SLC24A5 in South Asians and Europeans Shares Identity by Descent, PLoS Genet. 9 (2013).

[68] O. Nachtomy, A. Shavit, Z. Yakhini, Gene expression and the concept of the phenotype, Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci. 38 (2007) 238–254.

[69] R.K. Valenzuela, M.S. Henderson, M.H. Walsh, N. a. Garrison, J.T. Kelch, O. Cohen-Barak, et al., Predicting Phenotype from Genotype: Normal Pigmentation, J. Forensic Sci. 55 (2010) 315–322.

[70] D.L. Duffy, G.W. Montgomery, W. Chen, Z.Z. Zhao, L. Le, M.R. James, et al., A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation, Am. J. Hum. Genet. 80 (2007) 241–252.

[71] S.I. Candille, D.M. Absher, S. Beleza, M. Bauchet, B. McEvoy, N. a. Garrison, et al., Genome-Wide Association Studies of Quantitatively Measured Skin, Hair, and Eye Pigmentation in Four European Populations, PLoS One. 7 (2012)1 -12.

[72] K.K. Kidd, J.R. Kidd, W.C. Speed, R. Fang, M.R. Furtado, F.C.L.C.L. Hyland, et al., Expanding data and resources for forensic use of SNPs in individual identification, Forensic Sci. Int. Genet. 6 (2012) 646–652.

[73] F. Liu, K. van Duijn, J.R. Vingerling, A. Hofman, A.G. Uitterlinden, A.C.J.W.W. Janssens, et al., Eye color and the prediction of complex phenotypes from genotypes, Curr. Biol. 19 (2009) 192–193.

[74] R.A. Sturm, N.F. Box, M. Ramsay, Human pigmentation genetics: The difference is only skin deep, BioEssays. 20 (1998) 712–721.

[75] Y. Abe, G. Tamiya, T. Nakamura, Y. Hozumi, T. Suzuki, Association of melanogenesis genes with skin color variation among Japanese females, J. Dermatol. Sci. 69 (2013) 167–172.

[76] G. Tully, Genotype versus phenotype: Human pigmentation, Forensic Sci. Int. Genet. 1 (2007) 105–110.

[77] M.H. Brilliant, The mouse p (pink-eyed dilution) and human P genes, oculocutaneous albinism type 2 (OCA2), and melanosomal pH, Pigment Cell Res. 14 (2001) 86–93.

[78] R.A. Sturm, R.D. Teasdale, N.F. Box, Human pigmentation genes: identification, structure and consequences of polymorphic variation, Gene. 277 (2001) 49–62.

[79] R.A. Sturm, T.N. Frudakis, Eye colour: Portals into pigmentation genes and ancestry, Trends Genet. 20 (2004) 327–332.

[80] V.J. Hearing, K. Tsukamoto, Enzymatic control of pigmentation in mammals, FASEB J. 5 (1991) 2902–2909.

[81] OpenStax College, Anatomy & physiology, OpenStax CNX, Houston, 2013. Retrieved from https://opentextbc.ca/anatomyandphysiology.

[82] A.K. Swiatoniowski, E.E. Quillen, M.D. Shriver, N.G. Jablonski, Technical Note: Comparing von Luschan skin color tiles and modern spectrophotometry for measuring human skin pigmentation, Am. J. Phys. Anthropol. 151 (2013) 325–330.

[83] T.B. Fitzpatrick, The Validity and Practicality of Sun-Reactive Skin Types I Through VI, Arch. Dermatol. 124 (1988) 869.

[84] D. Rouse, Danish Medical Bulletin, J. Clin. Pathol. 43 (1990) 787–787.

[85] A. Treesirichod, S. Chansakulporn, P. Wattanapan, Correlation between skin color evaluation by skin color scale chart and narrowband reflectance spectrophotometer, Indian J. Dermatol. 59 (2014) 339.

[86] K.Y. Roh, D. Kim, S.J. Ha, Y.-J. Ro, J.-W. Kim, H.-J. Lee, Pigmentation in Koreans: study of the differences from caucasians in age, gender and seasonal variations, Br. J. Dermatol. 144 (2001) 94–99.

[87] E. De Wit, W. Delport, C.E. Rugamika, A. Meintjes, M. Möller, P.D. van Helden, et al., Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape, Hum. Genet. 128 (2010) 145–153.

[88] K. Brown, Coloured and black relations in South Africa: The burden of racial hierarchy, Macalester Int. (2000) 198–201.

[89] Statistics South Africa, Statistical Release: Mid-year population estimates 2014, Stats SA. P0301 (2014) A32–A33.

[90] J.M. Greeff, Deconstructing Jaco: Genetic heritage of an Afrikaner, Ann. Hum. Genet. 71 (2007) 674–688.

[91] M. Smith, M. Monique, Recent developments in DNA evidence, Trends & Issues in the Crime and Crimnal Justice. 3 (2015) 1-7.

[92] B.J. Koops, M. Schellekens, Forensic DNA Phenotyping: Regulatory Issues, Sci. Technol. Law Review. 9 (2008) 158–202.

[93] P.N. Ossorio, About Face: Forensic Genetic Testing for Race and Visible Traits, J. Law, Med. Ethics. 34 (2006) 277–292.

[94] M. Kopec, A new use of "race": The Evidence and ethics of forensic DNA ancestry profiling, J. Appl. Philos. 31 (2014) 237–253.

[95] A. M'Charek, Silent witness, articulate collective: DNA evidence and the inference of visible traits, Bioethics. 22 (2008) 519–528.

[96] V. Toom, M. Wienroth, A. M'Charek, B. Prainsack, R. Williams, T. Duster, et al., 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 Sci. Int. Genet. 22 (2016) 1–4.

[97] J. Allwood, Using DNA to predict externally visible characteristics in humans for forensic use in New Zealand, The University of Auckland, 2013.

[98] E. Murphy, Legal and Ethical Issues in Forensic DNA Phenotyping, 2013.

[99] H. Whittall, The forensic use of DNA: Scientific success story, ethical minefield, Biotechnol. J. 3 (2008) 303–305.

[100] J. Wise, Hollywood Profiling: The impact of the CSI Effect on the use Of forensic science in New South Wales, Expert Evid. Conf. (2011) 1-15.

[101] M. Smith, G.F. Urbas, Regulating new forms of forensic DNA profiling under Australian legislation: familial matching and DNA phenotyping, Aust. J. Forensic Sci. 44 (2012) 63–81.

[102] M.K. Cho, P. Sankar, Forensic genetics and ethical, legal and social implications beyond the clinic, Nat. Genet. 36 (2004) 8–12.

[103] L.B. Jorde, S.P. Wooding, Genetic variation, classification and "race," Nat. Genet. 36 (2004) 28–33.

[104] L. Braun, Reifying human difference: the debate on genetics, race, and health, Int. J. Health Serv. 36 (2006) 557–573.

[105] L.N. Borrell, N.D. Crawford, Disparities in Self-Reported Hypertension in Hispanic Subgroups, Non-Hispanic Black and Non-Hispanic White Adults: The National Health Interview Survey, Ann. Epidemiol. 18 (2008) 803–812.

[106] A. Ruiz-Linares, K. Adhikari, V. Acuña-Alonzo, M. Quinto-Sanchez, C. Jaramillo, W. Arias, et al., Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals, PLoS Genet. 10 (2014) 1 -12.

[107] O. Spichenok, Z.M. Budimlija, A. a. Mitchell, A. Jenny, L. Kovacevic, D. Marjanovic, et al., Prediction of eye and skin color in diverse populations using seven SNPs, Forensic Sci. Int. Genet. 5 (2011) 472–478.

[108] ZYMO Research Corp, Instruction Manual: Quick-DNA Universal Kit, (2001) 1–20.

[109] H. V Gudiseva, M. Hansen, L. Gutierrez, D.W. Collins, J. He, L.D. Verkuil, et al., Saliva DNA quality and genotyping efficiency in a predominantly elderly population, BMC Med. Genomics. (2016) 1–8.

[110] S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, et al., The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA, Forensic Sci. Int. Genet. 7 (2013) 98–115.

[111] G.A. Thorisson, A. V Smith, L. Krishnan, L.D. Stein, The International HapMap Project Web site, Genome Res. 15 (2005) 1592–1593.

[112] T. 1000 G.P. Consortium, A global reference for human genetic variation, Nature. 526 (2015) 68–74.

[113] W. Branicki, U. Brudnik, A. Wojas-pelc, Genetic Prediction of Pigmentary Traits in Forensic Studies, Probl. Forensic Sci. 64 (2005) 343–357.

[114] M. Soejima, Y. Koda, Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2, Int. J. Legal Med. 121 (2007) 36–39.

[115] R.K. Valenzuela, S. Ito, K. Wakamatsu, Prediction Model Validation: Normal

Human Pigmentation Variation, J. Forensic Res. 2 (2011) 1000139.

[116] K.C. Ang, M.S. Ngu, K.P. Reid, M.S. Teh, Z.S. Aida, D.X. Koh, et al., Skin Color Variation in Orang Asli Tribes of Peninsular Malaysia, PLoS One. 7 (2012) e42752.

[117] E.J. Parra, R.A. Kittles, M.D. Shriver, Implications of correlations between skin color and genetic ancestry for biomedical research, Nat. Genet. 36 (2004) 54–60.

[118] V. Kastelic, K. Drobnič, K. Drobnic, 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, Croat. Med. J. 53 (2012) 401–408.

[119] K.L. Hart, S.L. Kimura, V. Mushailov, Z.M. Budimlija, M. Prinz, E. Wurmbach, Improved eye- and skin-color prediction based on 8 SNPs, Croat. Med. J. 54 (2013) 248–256.

[120] A. Pneuman, Z.M. Budimlija, T. Caragine, M. Prinz, E. Wurmbach, Verification of eye and skin color predictors in various populations, Leg. Med. 14 (2012) 78–83.

[121] H.L. Norton, R. a. Kittles, E. Parra, P. McKeigue, X. Mao, K. Cheng, et al., Genetic evidence for the convergent evolution of light skin in Europeans and East Asians, Mol. Biol. Evol. 24 (2007) 710–722.

[122] F.S.G. Kehdy, M.H. Gouveia, M. Machado, W.C.S. Magalhães, A.R. Horimoto, Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations, 112 (2015) 8696–8701.

[123] ,F. Saloum, D.N. Manta, R. Pereira, R. Vianna, A.R.B. De, et al., Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels, 8 (2013) 1–11.

Section 4. **APPENDICES**

4.1 Appendix 4A: Saline solution recipe

Saline Solution, 0.9% Sodium Chloride (NaCl)

Makes 100 ml. Store at room temperature.

1. Dissolve 9 g NaCl (mw 58.44) in 700 ml distilled water.
2. Add water to make total volume 1000 ml.
3. Make 10 ml saline solution into 100 sterile 15 ml tubes.

4.2 Appendix 4B: Representation of commercial molecular weight marker (MWM)

Quick-Load® 50 bp DNA Ladder

The digested DNA includes fragments ranging from 50–1350 bps.

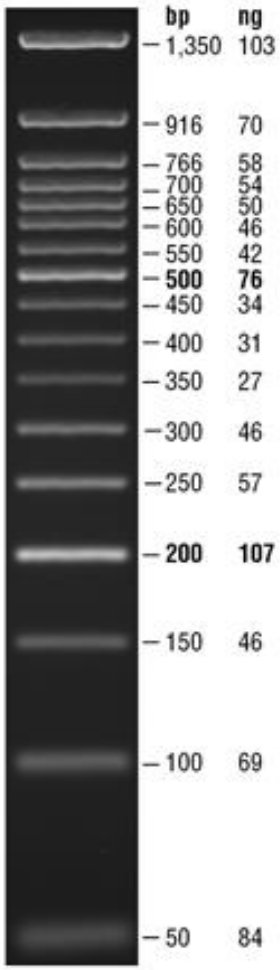

Image source: https://www.neb.com/products/n0473-quick-load-50-bp-dna-ladder

4.3 Appendix 4C: Representation of verified amplifed targets using MWM 4.2

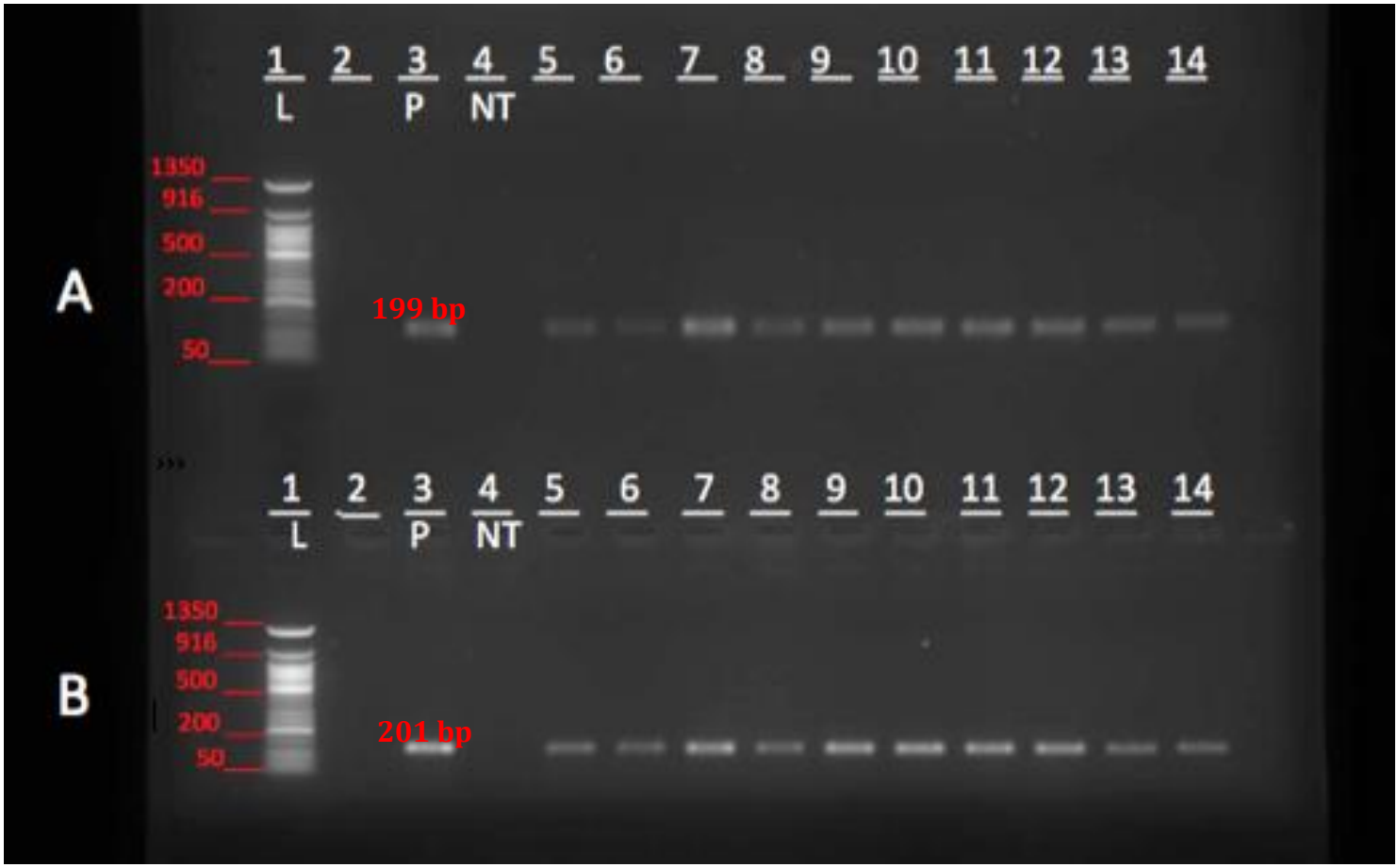

Figure 4C- 1 Image of 1.5% Agarose gel showing amplification of *SLC24A5* rs1426654 in part A and *SLC45A2* rs16891982in Part B. Electrophoresis was performed at 100 V for one hour

L = ladder, P = positive control and NTC = no template control. Lanes 5-14 corresponds to collected samples.

4.4 Appendix 4D: Representative electropherogram results

Multiplex SNaPshot reaction

This electropherogram displays peak intensity in RFU on the Y axis, fragment size (bp) along x-axis (in box along bottom), SNP name (along top) and * LIZ Size Standard. Probe mix dye colours green, blue, black and red refer to the following ddNTPs Adenine (A), Guanine (G), Cytosine (C) and Thymine (T), respectively.

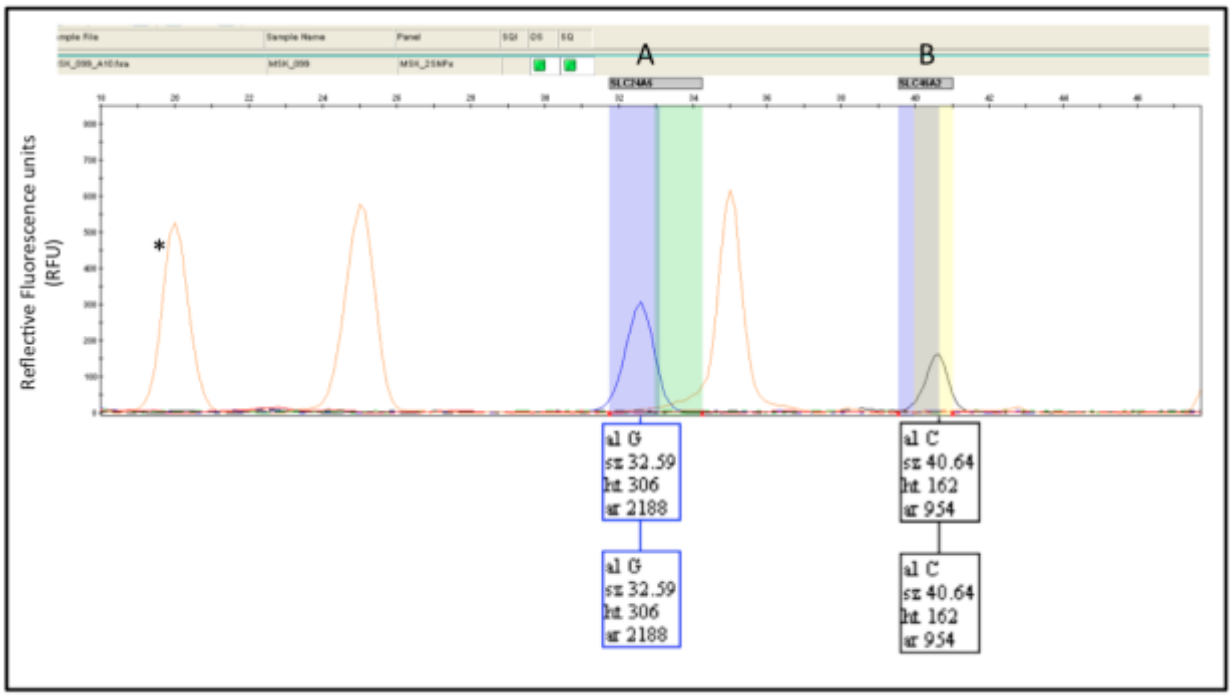

Figure 4D- 1 SNaPshot results obtained from the analysis of SNP SLC24A5 rs1426654, which is homozygous for allele G, as represented by a single blue peak, and SNP SLC45A2 rs16891982, which is homozygous for allele C, as represented by a single black peak

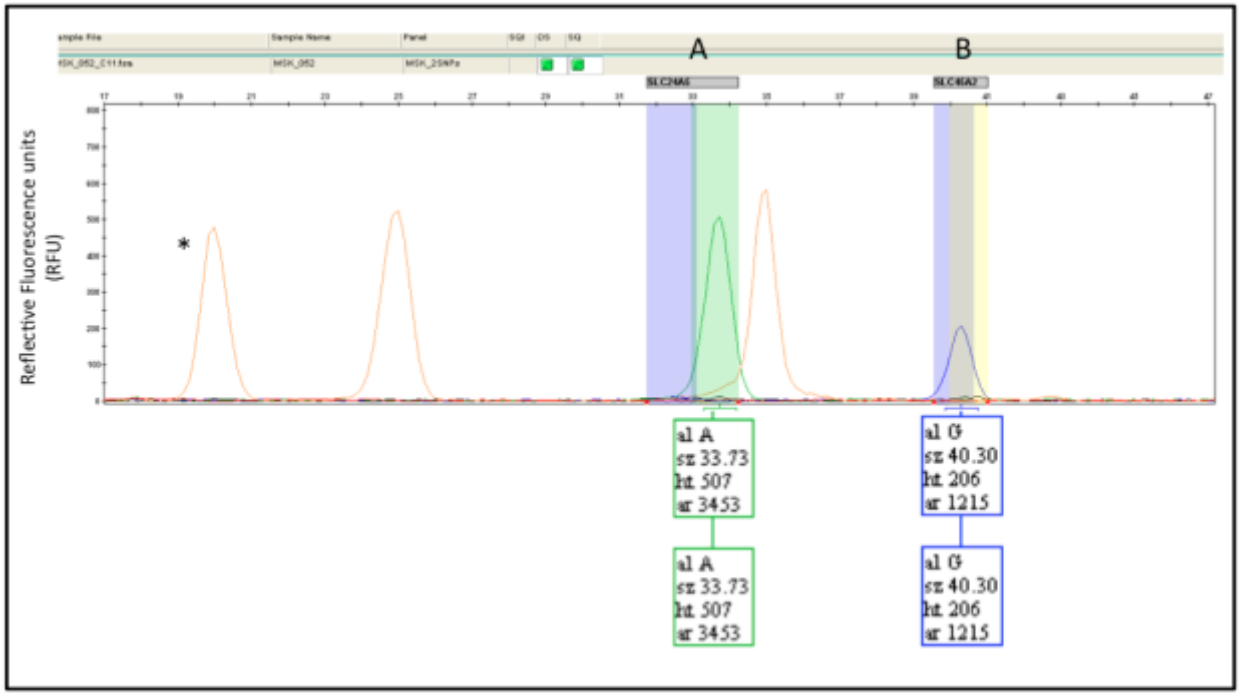

Figure 4D- 2 SNaPshot results obtained from the analysis of SNP SLC24A5 rs1426654, which is homozygous for allele A, as represented by a single green peak, and SNP SLC45A2 rs16891982, which is homozygous for allele G, as represented by a single blue peak

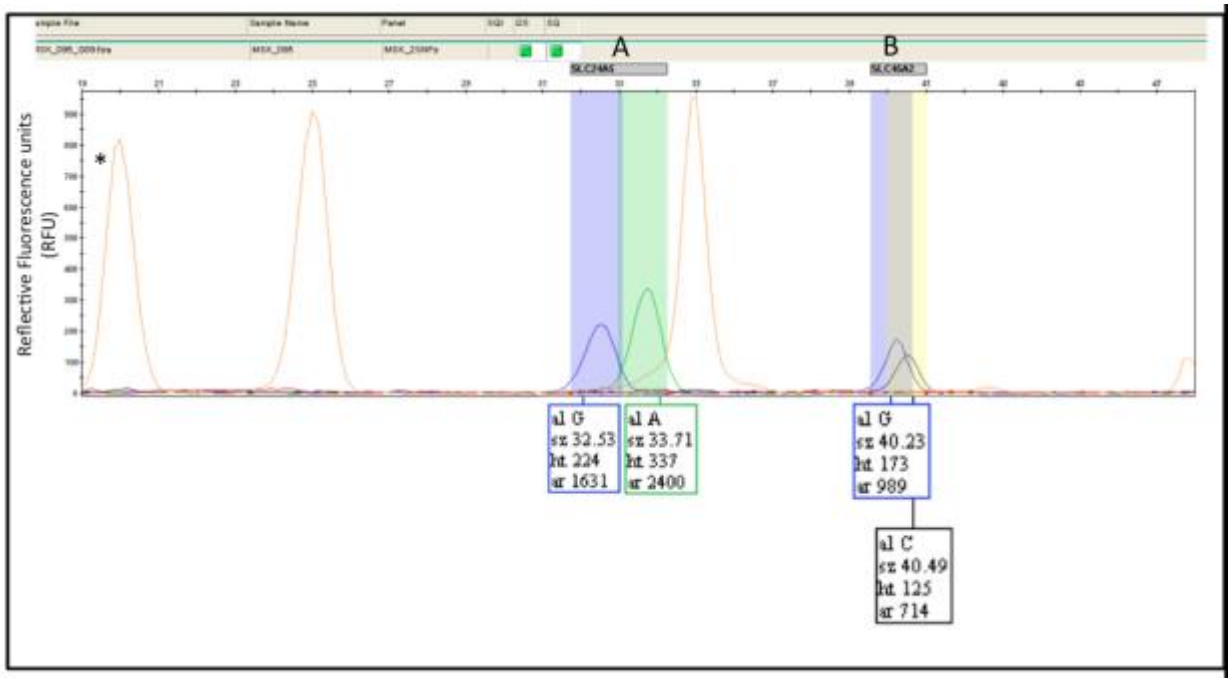

Figure 4D- 3 SNaPshot results obtained from the analysis of SNP SLC24A5 rs1426654, which is heterozygous for alleles G and A, as represented by both blue and green peaks, and SNP SLC45A2 rs16891982, which is heterozygous for alleles G and C, as represented by both blue and black peaks.

4.5 Appendix 4E: Summary of results

Table of Results

Summarized data collected from self-reported questionnaires. Results categorized by participant number and include calculated average MI and genotype for each SNP. Final study sample size comprised 92 individuals, missing participation numbers were samples not included due to DNA being too degraded to amplify during the experimental stage.

Table 4E- 1 Table representing collected data

SAMPLE	SEX	SA POPULATION	ANCESTRY	GENOTYPE		MI AVERAGE
				SLC24A5	*SLC45A2*	
				rs1426654	rs16891982	
01	Female	White	European	AA	GG	32.14
02	Female	White	European	AA	GG	32.06
03	Female	Indian/Asian	Asian	AA	CC	37.21
04	Female	African Black	African	GG	CC	55.20
05	Female	White	European	AA	GG	32.78
06	Female	White	European	AA	GG	32.76
07	Female	White	European	AA	GG	33.78
08	Male	African Black	African	GA	CC	70.85

09	Male	Coloured	Mixed	GA	CC	46.27
10	Male	White	European	AA	GG	32.73
11	Female	White	European	AA	GG	35.54
12	Female	Coloured	Mixed	GA	CC	37.74
13	Female	Other	Mixed	GG	CC	47.24
14	Male	African Black	African	GG	CC	66.12
15	Female	White	European	GA	GG	34.49
16	Female	Indian/Asian	Asian	GA	CC	47.18
17	Male	White	European	AA	GC	36.34
18	Male	White	European	AA	GG	32.23
19	Female	White	European	GA	GC	38.42
20	Female	Coloured	Mixed	GG	CC	47.47
21	Male	White	European	GA	GG	35.45
22	Female	Coloured	Mixed	GG	GC	45.05
23	Female	Coloured	Mixed	GG	CC	39.95
24	Female	African Black	African	GG	CC	48.55
25	Female	Coloured	Mixed	GG	CC	42.00

26	Male	White	European	AA	GG	33.26
27	Female	White	European	AA	GG	29.29
28	Female	White	European	AA	GG	29.52
29	Female	White	European	AA	GG	30.26
30	Male	White	European	AA	GG	33.54
31	Female	African Black	African	GG	CC	67.40
32	Male	African Black	African	GG	CC	41.12
33	Female	African Black	African	GG	CC	51.62
34	Male	Coloured	Mixed	GG	CC	45.74
35	Female	White	European	AA	GC	33.31
36	Male	White	European	AA	GC	36.33
37	Male	White	European	AA	GG	30.33
38	Male	White	European	AA	GG	28.76
39	Male	White	European	AA	GC	36.11
40	Male	White	European	AA	GC	34.96
41	Male	White	European	AA	GG	25.46
42	Male	White	European	AA	GG	33.93

43	Male	Other	Mixed	AA	CC	35.96
44	Male	White	European	AA	GG	28.78
45	Female	Coloured	European	GG	GC	43.81
46	Female	White	European	AA	GG	31.34
47	Female	Indian/Asian	Asian	GG	GC	45.31
48	Male	Indian/Asian	Asian	GA	CC	45.19
49	Female	White	European	AA	GG	32.77
50	Male	Coloured	African	GA	CC	53.68
51	Male	Coloured	European	GA	CC	47.60
52	Male	African Black	African	GG	CC	54.26
53	Female	African Black	African	GG	CC	58.05
54	Male	African Black	African	GG	CC	54.47
55	Female	Coloured	Mixed	GG	GC	43.12
56	Female	White	European	AA	GG	31.34
57	Female	African Black	African	GG	CC	60.92
58	Female	White	European	AA	GG	32.93
59	Female	Coloured	Mixed	AA	GG	32.79

60	Female	African Black	African	GG	CC	52.94
61	Female	African Black	African	GG	CC	45.69
62	Female	Indian/Asian	Asian	GG	CC	34.09
63	Male	African Black	African	GG	CC	58.15
64	Male	African Black	Mixed	GG	CC	55.85
65	Male	African Black	African	GG	CC	69.63
66	Male	African Black	African	GG	CC	75.73
67	Male	African Black	African	GG	CC	67.79
68	Male	African Black	African	GG	CC	71.39
69	Female	African Black	African	GG	CC	58.15
70	Male	African Black	African	GG	CC	81.74
71	Male	Indian/Asian	Asian	AA	CC	42.32
72	Female	African Black	African	GA	CC	63.47
73	Female	Indian/Asian	Asian	GA	CC	47.71
74	Male	White	European	AA	GG	33.73
75	Male	White	European	AA	GG	33.74
76	Male	White	European	AA	GG	38.96

77	Female	Coloured	Mixed	GA	CC	37.32
78	Female	White	European	AA	GC	31.29
79	Male	African Black	African	GA	CC	68.55
80	Male	African Black	African	GA	CC	67.34
81	Male	White	European	AA	GG	29.61
82	Male	African Black	African	GA	CC	55.85
83	Male	Indian/Asian	Asian	GG	CC	33.08
84	Female	African Black	African	GG	CC	53.94
85	Male	White	European	AA	GG	35.03
86	Male	African Black	African	GG	CC	61.21
87	Female	Coloured	Mixed	GA	GC	37.17
88	Male	White	European	AA	GG	30.22
89	Female	African Black	African	GG	CC	51.74
90	Female	White	European	AA	GG	33.29
91	Female	African Black	African	GA	CC	55.48
92	Female	White	European	AA	GG	32.70

4.6 Appendix 4F: Author guidelines for selected journal format

FORENSIC SCIENCE INTERNATIONAL

ISSN: 0379-0738

The following text extracts have been taken from the "Author Information Pack" accessed from www.elsevier.com/locate/forsciint

Description

Forensic Science International publishes original contributions in the many different scientific disciplines pertaining to the forensic sciences. Fields include forensic pathology and histochemistry, chemistry, biochemistry and toxicology (including drugs, alcohol, etc.), biology (including the identification of hairs and fibres), serology, odontology, psychiatry, anthropology, the physical sciences, firearms, and document examination, as well as investigations of value to public health in its broadest sense, and the important marginal area where science and medicine interact with the law.

Forensic Science International publishes:Original Research Papers Review Articles Preliminary Communications Letters to the Editor Book Reviews Case ReportsThe journal covers all legal aspects of the general disciplines listed above, as well as specialist topics of forensic interest that are included in, or are related to, these disciplines, e.g.: Biochemical and chemical analyses, and the forensic application of advanced analytical, physical, chemical and instrumental techniques Bitemark evidenceBattered child syndrome Questioned documents Ballistics, projectiles and wounds Fingerprints and identification Tool marks Contact traces Poisoning Breath analysers Accident investigation and mass disasters

Language

Please write your text in good English (American or British usage is accepted, but not a mixture of these).

References

There are no strict requirements on reference formatting at submission. References can be in any style or format as long as the style is consistent. Where applicable, author(s) name(s), journal title/book title, chapter title/article title, year of publication, volume number/book chapter and the pagination must be present. Use of DOI is highly encouraged.

Formatting requirements

There are no strict formatting requirements but all manuscripts must contain the essential elements needed to convey your manuscript, for example Abstract, Keywords, Introduction, Materials and Methods, Results, Conclusions, and Tables with Captions. Figures and Tables embedded in text

Keywords

Immediately after the abstract, provide a maximum of 6 keywords, using American spelling and avoiding general and plural terms and multiple concepts (avoid, for example, 'and', 'of'). Be sparing with abbreviations: only abbreviations firmly established in the field may be eligible.