3-31-2022

The Utilization of Volatile Organic Compounds and Human Leukocyte Antigen Genes for Ethnic-Specific Differentiation within Target Populations

Alice Breia Boone
Florida International University, aboon005@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

THE UTILIZATION OF VOLATILE ORGANIC COMPOUNDS AND HUMAN LEUKOCYTE ANTIGEN GENES FOR ETHNIC-SPECIFIC DIFFERENTIATION WITHIN TARGET POPULATIONS

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Alice Breia Boone

2022

To: Dean Michael R. Heithaus
College of Arts, Sciences and Education

This dissertation, written by Alice Breia Boone , and entitled The Utilization of Volatile Organic Compounds and Human Leukocyte Antigen Genes for Ethnic- Specific Differentiation within Target Populations , having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Konstantinos Kavallieratos

_____
Francisco Fernandez-Lima

_____
DeEtta Mills

_____
Yong Cai

_____
Kenneth G. Furton, Major Professor

Date of Defense: March 31, 2022

The dissertation of Alice Breia Boone is approved.

_____
Dean Michael R. Heithaus
College of Arts, Sciences and Education

_____
Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2022

DEDICATION

I would like to dedicate this dissertation to my cousins Lamethia McCollum and Mia. At age 28, while 8.5 months pregnant, Lamethia was brutally murdered in her home by the father of her child. This case was cold for 10 years. However, due to forensic science, my family was able to finally receive closure and some peace in knowing that justice was served.

ACKNOWLEDGMENTS

This dissertation could not have been completed without the emotional and financial support of the McKnight Doctoral Fellowship. My cohort has been an asset to my personal and professional development, and I am forever grateful for the opportunity they gave me to pursue my doctorate. Furthermore, I would like to acknowledge the faculty and staff at FIU who have given me guidance through my matriculation through the program. Under the tutelage of my Ph.D. advisor, Dr. Kenneth G. Furton, I was given several opportunities to gain expertise in several sectors of forensic science. I would also like to acknowledge Howard Holness for sharing his scientific knowledge and personal experience to serve as an additional mentor to me. I would like to thank my advisory committee members, Dr. Yong Cai, Dr. Francisco-Fernandez Lima, and Konstantinos Kavallieratos for always supporting me and providing positive feedback to further enhance my research. I would like to give a special thank you to Dr. DeEtta Mills, who also serves on my committee, for taking the time to meet with me in support of my biological studies and for providing the emotional support and encouragement I needed to complete my degree.

The support system I have gained since enrolling in graduate school was a staple of my ability to remain grounded in the doctoral program at FIU. I would like to thank my friends (Christina Burns, Deidra Jordan, Vanquilla Shellman-Francis, and Chantrell Fraizer) who poured so much into me academically and emotionally. Their selflessness with their knowledge, unwavering encouragement, and willingness to guide me through concepts I did not understand well initially eased several frustrating moments in my

graduate career. Furthermore, I would like to thank my friends, Jasmine Bray, Oshea Johnson, and Zashery Mercedes for being my family away from home and providing me with countless memories and a bond that cannot be broken.

Finally, I would like to express endless gratitude to my parents, husband, and brother. The tremendous sacrifice, patience, and love they have exhibited at every stage of my life is a constant reminder of how blessed I truly am. Their active role in ensuring I was able to finish school with two newborns will never be taken for granted.

ABSTRACT OF THE DISSERTATION

THE UTILIZATION OF VOLATILE ORGANIC COMPOUNDS AND HUMAN LEUKOCYTE ANTIGEN GENES FOR ETHNIC-SPECIFIC DIFFERENTIATION WITHIN TARGET POPULATIONS

by

Alice Breia Boone

Florida International University, 2022

Miami, Florida

Professor Kenneth G. Furton, Major Professor

Human scent has been previously defined as a complex mixture of volatile organic compounds (VOCs) detected in the headspace above a scent sample. Humans generate odor from several areas of the body including hair, mouth, hand, axillae, and foot. Due to the novelty of human scent research, human scent evidence has been undervalued in the court of law. However, this type of evidence has significant value when physical evidence is not available at crime scenes. To increase the individualization and differentiation power of human scent evidence, this study aimed to further investigate the identification of chemical signatures within axillae of specific ethnicities (Caucasian, Hispanic, and African American) and determine if ethnic-specific genetic signatures are present among Human Leukocyte Antigen (HLA) genes.

During the study, the axillae of 68 participants were investigated. Upon collection, samples were extracted using Headspace Solid Phase Micro extraction (HS-SPME) and solvent extraction. The samples were analyzed using Gas Chromatography- Mass Spectrometry

(GC-MS). The utilization of SPME immediately followed by solvent extraction complements the extraction of both semi-volatile and non-volatile compounds, thus filling in the gaps of the compounds that could not be recovered using HS-SPME alone.

The samples were evaluated statistically via logistic regression and Receiving Operating Characteristic (ROC) curves to evaluate the performance and prediction power of VOCs for ethnicity inferences. The study concluded that logistic regression served as an efficient model predicting the VOCs capable of class characteristic determination when comparing ethnicities.

The HLA gene complex was evaluated to determine its contribution to human scent and the ability to differentiate between ethnicities. Using buccal swabs extracted from 31 subjects, five genes were successfully amplified using Multiplex Polymerase Chain Reaction (Multiplex PCR). The Multiplex PCR products were analyzed using capillary electrophoresis. The genotype frequencies were observed, and linear discriminant analysis (LDA) was performed to assess the ability of predicting ethnicity using genotype frequencies of individuals. Four of the five genes predicted ethnicity at 80% or greater accuracy, which validates that the HLA genes (D6S2925, D6S2937, D6S2917, and D6S2787), coupled to the VOCs, can be used as a biomarker for class characteristic determination of an individual.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

LIST OF EQUATIONS

EQUATION PAGE

APPENDICIES

APPENDIX PAGE

ABBREVIATIONS AND ACRONYMS

75 um Carboxen/Polydimethylsiloxane	CAR/PDMS-Black-Mid Polar
Divinylbenzene/Carboxen/Polydimethylsiloxane	DVB/CAR/PDMS-Grey-Polar
Area Under the Curve	AUC
Capillary Electrophoresis	CE
Deoxyribose Nucleic Acid	DNA
Gas Chromatography Mass Spectrometry	GC-MS
Genomic DNA	gDNA
Headspace Solid Phase Microextraction	HS-SPME
Human Leukocyte Antigen	HLA
Indicator Species Analysis	ISA
Length Heterozygosity Polymerase Chain Reaction	LH-PCR
Linear Discriminant Analysis	LDA
Multiplex Polymerase Chain Reaction	Multiplex PCR
PC1	First Principal Component
PC2	Second Principal Component
Polymerase Chain Reaction	PCR
Principal Component Analysis	PCA
Receiving Operating Characteristic Curve	ROC Curve
Short Tandem Repeats	STRs
Solid Phase Microextraction	SPME
Volatile Organic Compounds	VOCs

1. INTRODUCTION

1.1. INTRODUCTION TO RESEARCH

The thorough analysis of a crime scene requires the collection of all available evidence. The forensic scientist's obligation is to identify and establish the source of the evidence collected. This entails linking a suspect to a victim or a specific crime scene. Throughout the United States, there are many crime-based television programs that portray the collection of trace evidence from crime scenes. The information transmitted within these programs enables criminals to strategically incorporate their knowledge of forensic investigations to their advantage. A study conducted by Balemba et al. (2014) has revealed that "forensically aware" criminals who committed sexual homicides (accounting for 33% of the population studied)[i] believe that DNA and fingerprints have been generally accepted in the scientific community and within the media as viable tools for the identification of individuals. However, when physical evidence is not present at the crime scene, the likelihood of the case remaining unsolved increases significantly.

In accordance with Locard's Exchange Principle—"*Wherever he steps, whatever he touches, whatever he leaves, even unconsciously, will serve as a silent witness against him*" —even human odor is easily transferred onto virtually anything with which the body comes in contact.[1] Because of its ability to transfer, and the unique attributes of each person's odor profile, the forensic analysis of this type of evidence can be a significant factor in the investigation of a criminal case. The past decade has seen rapid developments in the field of forensic science. However, due to the novelty of human scent research, human scent evidence has been undervalued in the court of law. To

increase the integrity of human scent evidence, knowledge regarding volatile organic compounds (VOCs) profiles must be expanded as this type of evidence has significant value, even when physical evidence is not available at crime scene. To date, though, there is minimal understanding about the efficacy of VOCs as physical evidence.

As a result, this dissertation serves as a bridge between the biological and chemical components of human scent evidence. By determining a correlation between odor profiles and Human Leukocyte Antigen (HLA) genes, chemical signatures were identified that can be used to effectively narrow the potential suspects during criminal investigations. This research specifically focuses on enhancing the profiles of three relatively large populations Caucasian, Hispanic, and African American. Each subject in the study provided axillary (underarm) odor samples and buccal swabs to determine chemical and genetic information, respectively. The VOC profiles were extracted using Solid Phase Micro Extraction (SPME). Semi-volatile and non-volatile compounds were extracted using solvent extraction, all compounds extracted via either SPME or solvent extraction were then analyzed via Gas Chromatography-Mass Spectrometry. Once a complete odor and genetic profile was obtained for each race, unique signatures within each race can be determined using statistical analysis, potentially portraying the correlation hypothesized.

1.1 RESEARCH OBJECTIVES

This dissertation aims to further enhance the VOC profiles of three populations (Caucasian, Hispanic African Americans). The novel incorporation of African American subjects will highlight the significance of investigating and comparing human scent expression among all ethnic groups. The objective of this research is to identify specific biomarkers in individual scent profiles that can be used to differentiate among races. Furthermore, using

the HLA profiles of each individual subject as supplemental data to identify race will generate a complete human profile that can be utilized for class characteristic determination and exclusion purposes in forensic casework. The use of analytical and statistical methods facilitates the successful analysis of human subjects. The tasks performed to meet the dissertation's goals are listed below:

a) Determination of VOCs present in Caucasians, Hispanics, and African Americans.
b) Pilot study: Identification of VOCs present in Caucasians, Hispanics, and African Americans.
c) Optimization of solvent extraction method for identification of fatty acid and sulfur compounds.
d) Optimization of HLA gene acquisition via buccal swabbing.
e) Determination of the correlation between VOCs and HLA.

1.2 SIGNIFICANCE OF STUDY

Human scent evidence has been challenged in U.S. courts of law as the technique's scientific validity has been called into question. To date, for a scent identification to be admitted into court and used as incriminating evidence, it must satisfy the Kelly rule, which highlights the general principles of admissibility in court when novel techniques are utilized[2].

The pursuit of this project's objectives constitutes a significant advance for the forensic science community. Applying synergistic extraction methods (SPME immediately followed by solvent extraction) to human scent profiles complements the extraction of semi-volatile and non-volatile compounds, hence filling in the gaps of the compounds that could not be recovered using HS-SPME alone. This ensures that a more complete VOC

profile is obtained, allowing for improved statistical analysis without requiring any additional sample collection from subjects. The evaluation of the fatty acids provide new data to build on and enhance the information obtained during previous studies of odor profiles and their characteristic features. Moreover, the novel incorporation of African American profiles validates the discriminatory power of human scent evidence regardless of ethnicity. Furthermore, the data analysis facilitates specific quantitative and qualitative observations that can enhance the understanding of the chemical composition of scent for the populations under study as well as the determination of ethnic-specific VOCs and quantitative VOC representation in specific populations.

This dissertation serves as an introduction to the holistic evaluation of chemical and genetic components of human scent. Establishing the correlation between VOCs and HLA genes significantly increases the discriminating power of human scent evidence. The use of extraction methods and analytical techniques also provides insights that can be used for class characteristic determination. This will serve as a groundbreaking technique when other physical evidence such as DNA is not available. Finally, the statistical analyses implemented can be reliably utilized in court cases, further validating the potential of human scent to associate an individual with a crime scene, thus satisfying the Kelly rule.

2. Human Scent Research Background

2.1 BIOLOGICAL SPECIMENS CAPABLE OF IDENTIFYING INDIVIDUALS

Forensic evidence has a significant impact on the criminal justice process and outcomes. Presently, there are several biological specimens capable of identifying individuals (blood, semen, saliva, hair) that can be collected from crime scenes for investigative leads. Every crime involves both a perpetrator and a victim. Particularly in the case of a struggle, biological specimens are often transferred throughout the scene from the victim and/or perpetrator.

The unique characteristics of biological specimens per individual are essential for the conviction of perpetrators and the exoneration of the innocent. In a database study conducted by Peterson et al., biological specimens were collected at 97% of homicide cases,[3] the majority of which (38.3%) are biological samples. However, Figure 1 shows that of the 400 samples used in these cases, only 6% involved scent evidence.[3] Additionally, a total of 44.3% of homicide cases resulted in no arrest. This statistic highlights the significance of the present study. With effective human scent evidence collection, cold cases can be resolved when other, more common, evidence is not available.

Evidence Type	N=	Collected		Submitted		Examined	
		n	%	n	%	n	%
Total	400	388	97.0%	354	88.5%	324	81.0%
Biological		153	38.3%	129	32.3%	102	25.5%
blood		137	34.3%	84	21.0%	60	15.0%
DNA		18	4.5%	14	3.5%	14	3.5%
saliva		13	3.3%	36	9.0%	27	6.8%
vaginal		0	0.0%	6	1.5%	4	1.0%
semen		0	0.0%	8	2.0%	8	2.0%
condom		0	0.0%	1	0.3%	1	0.3%
bone		1	0.3%	0	0.0%	0	0.0%
tissue		2	0.5%	3	0.75%	2	0.5%
sexual assault kit		16	4.00%	6	1.5%	4	1.0%
scents		0	0.0%	24	6.0%	11	2.8%
urine		0	0.0%	3	0.8%	3	0.8%
biological, other		32	8.0%	17	4.3%	12	3.0%

Figure 1: Percentage of evidence collected at crime scene[4]

2.2 ANATOMY AND PHYSIOLOGY OF HUMAN SKIN AND BODILY SECRETIONS

As the largest organ of the body, human skin provides a protective barrier that plays an active role in fighting against physical, chemical, and microbial intruders.[5] The extensive surface area of the skin accounts for 15% of total adult body weight, with about 40,000 skin cells being shed per minute.[6] Moreover, a recent study has suggested that human skin surface area has been miscalculated due to the assumption of human skin being a flat surface; the appendage openings such as sweat ducts and hair follicles significantly increase the epithelial surface area from $2m^2$ to 25 m^2 [7].

Skin is divided into three layers: epidermis, dermis, and subcutaneous tissue.[6,8] The glands within the dermis (apocrine, eccrine, and sebaceous) are responsible for the secretion of sweat present in the axillary region.[9] The increase in apocrine sweat glands within humans also plays a prominent role in the foundation of armpit odor.[10]

Eccrine sweat glands are mostly composed of water, amino acids, electrolytes, and minerals. These glands are essential for body temperature control and sweat's dissipation of heat by evaporation within both primates and humans.[11] Eccrine sweat glands are typically present in the hands and feet.[12] However, the chemical composition of eccrine sweat modifies according to diet, drug intake, and disease.[11]

Apocrine glands, located at the base of hair follicles, secrete a fluid very similar to the eccrine glands from which they derive,[13] but the overall sweat rate of apocrine glands supersedes all other types of glands.[14] The secretions within the protein,, lipid, and steroid-rich apocrine glands have been recognized as the primary source of genital axillary odor.[8] Upon collection in the axilla, eccrine and apocrine sweat glands can be deciphered by fluorescence.[14]

Lastly, subcutaneous fat—composed of lipocytes, large blood vessels, and collagen—stores energy for humans and provides the body with buoyancy.[15] Each of these three layers (epidermis, dermis, and subcutaneous fat) contribute to bodily secretions (see

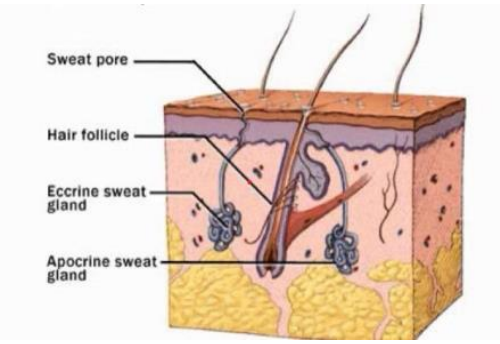


Figure 2).

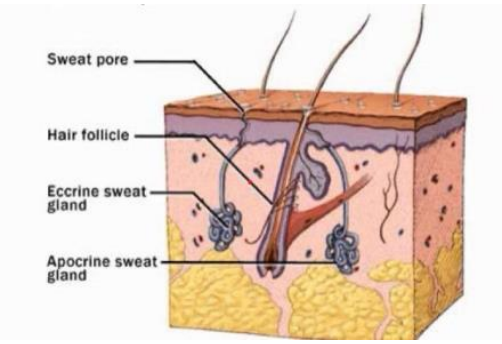


Figure 2: Structure of human skin[15]

2.3 VOC'S CONTRIBUTION TO HUMAN SCENT

Human scent has been previously defined as a complex mixture of VOCs detected in the headspace above a scent sample.[16] These VOC mixtures arise from the impact of different biological, chemical, and environmental components. In the human body, VOCs excreted from skin often reveal important information about external and internal causes of odor.[17] These different factors have been classified into three different types of odor: primary, secondary, and tertiary.[16] Primary odor does not change over time and is only impacted by genetics. Secondary odor emits by the influence of an individual's diet and environmental factors. Tertiary odor exists from external additives onto a subject that add odor to an individual's odor profile (e.g., perfumes, lotions, deodorants). While conducting experimental research, primary odor is investigated, secondary odor is documented, and tertiary odor is preferably controlled.

Human scent has been used as evidence in forensic investigations and also employed in the medical field as a tool to identify biomarkers for specific diseases.[16]The forensic value of the VOCs that emanate from several biological specimens (saliva,

fingernails, hand odor, and hair) have been previously studied and reported.[18] A total of 1,849 VOCs have been discovered as contributors to the volatilome of humans. As depicted in Table 1 below, over 500 volatiles have been reported to have been isolated from skin secretions.[18]

Table 1: VOCs from skin secretions (Filipiak et al.)

Biological Specimen	Number of VOCs Detected
Urine	279
Skin	504
Saliva	353
Blood	130
Feces[19]	381

However, to date, there have been minimal studies on the forensic contribution of VOCs emitted from axillary (underarm) secretions.

2.4 HUMAN SCENT STUDIES

Numerous studies have assessed the role of VOC abundance and chemical nature in human scent. The chemical compounds identified from these sources include fatty acids, steroids, ketones, hydrocarbons, alcohols, and aldehydes.[16,20] The analysis of the chemical

composition of human scent has enabled scientists to describe variations within these factors (age, gender, and ethnicity) to distinguish among individuals. A study by Zeng et has concluded that the volatiles present in men and women are qualitatively similar.[21] Moreover, a study by Prokop-Prigge et al. evaluated the quantitative chemical signatures in odor profiles from Caucasian, East Asian, and African American individuals, and a "marginal effect" was noted in the amounts of compounds across the ethnic groups observed.[22] In addition, research by Bates et al. has highlighted the ability to differentiate ethnicities based on odor. In this study, elephants were used to classify ethnic groups based on their odor alone.

VOC mixtures arise from the impact of different biological, chemical, and environmental components in the human axilla. There are several types of bacteria known to colonize in the human axilla, including staphylococci, aerobic coryneforms, Propionibacterium, micrococci, and Malassezia.[10] Skin bacteria have been identified as a differentiating variable among individuals' body odor. Specifically, bacteria such as Corynebacterium colonize in moist areas like the axilla.[23] Moreover, fatty acids have also been previously reported, by different research studies, as major contributors to axillary malodor. The major metabolic route to short- and medium-chain volatile fatty acids (VFAs) in the axilla was described as the partial degradation of long-chain fatty acids by Corynebacterium.[24] However, in the authors' newest publication, they propose that the major route to short-chain VFAs is the metabolism of branched aliphatic amino acids by staphylococci.[25] Natsch et al. discovered that medium-chain VFAs such as 3M2H and 3-hydorxy-3-methyl-hexanoic acid were derived from N^{a}-acyl-L-glutamine predecessors.

As a result of the conditions within the axilla (high temperature and extreme moisture), bacteria heavily colonize the body region and have the ability to convert aliphatic amino acids into extremely odorous (C4-C5) methyl-branched short-chain volatile fatty acids such as isovaleric acid. [24]Besides fatty acid compounds, sulfur-containing compounds have also been reported in previous studies as contributors to axillary odor. According to Troccaz et al., thiols and other sulfur-containing amino acids are excreted in sweat, causing malodor within the axilla region.[26]

Since the 1950s, malodor has been connected with the biotransformation of natural secretions into volatile molecules.[8] Indeed, bodily secretions play a prominent role in the formation of body odor. Humans generate odor from several areas of the body including the scalp, hair, mouth, hand, axillae, and foot.[27] However, the axilla serves as the primary site where malodor exists.

2.5 GENETIC INFLUENCE ON HUMAN ODOR

Several studies have evaluated the influence that genetics has on VOCs. Recent studies have advocated that the variants in the ABCC11 gene play a significant role in axillary malodor. [28]However, the study by Prokop-Prigge et al. argues that this gene cannot be solely responsible for the VOCs present within the axilla, and other biochemical pathways are also involved in the release of VOCs there.

One of the most informative, diverse genetic coding regions is the major histocompatibility complex (MHC), known as HLA in humans. MHC and HLA are commonly known for their influence on mating preference in mammals; however, these genes also have an effect on body odor.[29] The HLA is a large chromosomal region with

highly polymorphic genes.[18] Its primary function is to activate immune defenses; however, the alleles present within the HLA have a high discriminating power due to the combination of alleles unique to every individual with the exception of twins.[29]

There are several theories as to exactly how the HLA contributes to the odor present in human sweat.[30] The most generally accepted theory is that metabolites of MHC molecules are the carriers of VOCs[30] A specific allele in the HLA complex has been identified to have an impact on metabolic pathways in the human B-cell lines. It has been further demonstrated that there are HLA-associated VOCs in mammalian biological specimens such as urine and blood.[30]

Previous studies have used the Genomic Matching Technique (GMT) to extract HLA genes;[31] however, a method has recently been developed for the STR genotyping of the HLA region.[19] Figure 3 illustrates STR loci correlated to the HLA and its respective primer sequence.[19]

Table 1
Summary of HLA-associated STR loci and primer sequences comprising the genotyping panel. Bold sequences represent primer tails added to optimize PCR efficiency.

STR[a]	Repeat unit	# Alleles observed	Observed size range, bp	Forward primer sequence (5′–3′)	Reverse primer sequence (5′–3′)	Amp mix
D6S105	CA	12	279–303	FAM-TCAGGCTGGTCTCAGACTCC	**GTTTCTT**GCAAAACTCCATATTGAACAGG	1
D6S2917	TTCC	24	247–361	HEX-AGGCCTTTTTCACTTGTTTTTCTA	TCTTCTGCTCTACCCACCA	1
D6S2811	GT	22	222–268	HEX-TGGAAATGTCATTTTGATCCAGGC	**GTTTCTT**CCAGGGCCCAGTAGTAAGCC	2
D6S265	CA	12	302–324	FAM-AGCATTTATTTCTCTGTGTTGGG	**GTTTCTT**AGCATGTTCTCACTCATGTGC	2
D6S2787	GT	15	134–163	FAM-CTCCAGCCTGGATAACAG	ACAAGGGCTTTAGGAGGTCT	2
M138	AAAG	17	235–320	FAM-GAGAGAGCGAGATTCTGC	**GAATGCTTAGAAGATGAAGGTG**	3
D6S2749	GT	9	285–301	HEX-TTGCATACACTCTGAAGCAGC	TCCCTGTGGATGTCAAGAATC	3
D6S2925	TAAA	9	106–141	FAM-TCTAGTGTCTTCTGGCCTTG	**GTTTCTT**ATCAGGGAAAGGTGCTGGT	3
D6S2837	GATA	13	137–199	HEX-CCAAGAAAGAAAGAACCAATAGCA	GAGCCAATTAGCCCAATAAATCAC	3

Figure 3: STR loci and primer sequence used in previous HLA studies[19]

2.6 SOLID-PHASE MICROEXTRACTION

Sample preparation is one of the most critical steps in any analytical technique. This step is essential for the successful analysis of representative samples. The collection and preparation of samples entail extracting a small aliquot of material and preparing it for quantitative analysis. Adequate sample preparation ensures that a sample is suitable for injection into analytical instruments such as GC, NMR, and HPLC and that it will not damage the column in use. The significance of sample preparation can be seen by the variety of fields to which the process is applicable. The environmental, geoscience, and forensic communities in particular, benefit from the utilization of sample preparation techniques. From the forensic perspective, sample preparation is essential for a successful investigation. The quality of the sample preparation increases the validity of the collected evidence. Several techniques such as solvent extraction, Solid Phase Extraction, and Purge and trap are used to analyze trace evidence[32]. However, more recently, the forensic science community has revolutionized its sampling procedures using Solid- Phase Micro Extraction (SPME) as a novel sampling technique for a variety of applications (environmental science, forensic science, food chemistry, etc.).

SPME was invented in 1990 by Dr. Janusz Pawliszyn and his colleagues.[33] The technique employs a syringe-like device composed of a fused silica fiber coated with an adsorbent polymer. The process involves the use of different fiber coatings to extract the analyte (in either the liquid or gas phase). The technique allows for equilibrium to be reached between the analyte and the fiber. During extraction, the fused-silica fiber is exposed to the sample and the analyte adsorbs onto the fiber coating.[34] Solid-phase

microextraction can be performed in two ways: direct immersion and headspace. In direct-immersion SPME, the fiber is directly immersed into the analyte of interest.[35] However, in headspace SPME, the fiber is placed above the headspace of the sample for volatile analysis. The amount of analyte adsorbed onto the fiber depends on the thickness of the coating. After the extraction is complete, the SPME fiber is transferred into an injection port of an analytical instrument such as a GC-MS, where desorption of the analyte occurs. In order to develop the most accurate and precise method, specific parameters for SPME must be optimized, including fiber chemistry, equilibration time, extraction time, and extraction temperature. The type of fiber needed depends on the molecular weight and polarity of the analyte. Increasing the porosity of the fiber tightly retains analytes; varying the pore size increases the analyte sensitivity[33] There are several types of fiber chemistries available from Supelco, which can be divided into two categories: adsorbent and absorbent. Adsorbent fibers physically trap or chemically bond with the analyte of interest. Absorbent fibers absorb by partitioning.[36] Table 2 lists the SPME fibers that have been used to assess specific human scent analyte types in the previous literature.

Table 2:SPME fibers commercially available

Fiber Coating	Polarity	Compounds to be analyzed
Polydimethylsiloxane (PDMS)	Non-Polar	Volatiles
Divinylbenzene/Carboxen PDMS	Bipolar	Odorous and flavors
Carboxen PDMS	Bipolar	Gases and volatiles

However, this only serves as a guideline regarding which fiber to use. Scientists must test these fibers on their specific analyte to determine which fiber produces the best yield with the lowest standard deviation. Once the appropriate fiber is selected, the optimal equilibration and extraction time are selected by testing a range of time periods (one minute, five minutes, thirty minutes, one hour, two hours, etc.). Equilibrium is assumed to be attained when 95% of the analyte is extracted from the sample.[35] Previous human scent studies have determined that the optimal extraction time is 15 hours. The extraction temperature study assesses which temperature extracts the most VOCs. Increasing the temperature of the sample usually increases the vapor density of the analyte.[13]However, when heating a sample, the scientist must ensure that SPME offers several advantages in the analysis of VOCs in comparison to other sampling techniques. The sampling procedure is relatively low in cost and can be used on various types of analytes (gases, solids, and liquids). The design of the SPME device protects the sample, making contamination less likely. The absence of solvents used in SPME makes it a safe, environmentally friendly

procedure. Additionally, the absence of solvents enables faster separation [36]. The method is highly sensitive and simple to use. The small size of the SPME fiber makes it a useful portable field instrument.

As with any scientific method, there are also several disadvantages of using SPME. The SPME device is very fragile, and several parts of the device are easily broken. (Problems can include stripped fibers, bending of the needle, and fiber breakage.) The apparatus consists of a polymer coating on fused silica fiber. HS-SPME has been utilized to examine the chemical composition of drugs, explosives, and human volatiles for forensic identification purposes. The fiber is exposed to the headspace of the sample, and the analyte molecules either absorb into or adsorb onto the polymer coating. It is an equilibrium extraction technique, so it is non-exhaustive. However, it should provide a snapshot of the environment within the sample vial. As depicted in Figure 4, the septum of a headspace vial is pierced, and the fiber is exposed to the headspace of the sample. The analyte molecules either absorb into or adsorb onto the polymer coating, which provides a snapshot of the environment within the sample vial. Upon completion, the fiber is exposed in the GC inlet for desorption of analytes.

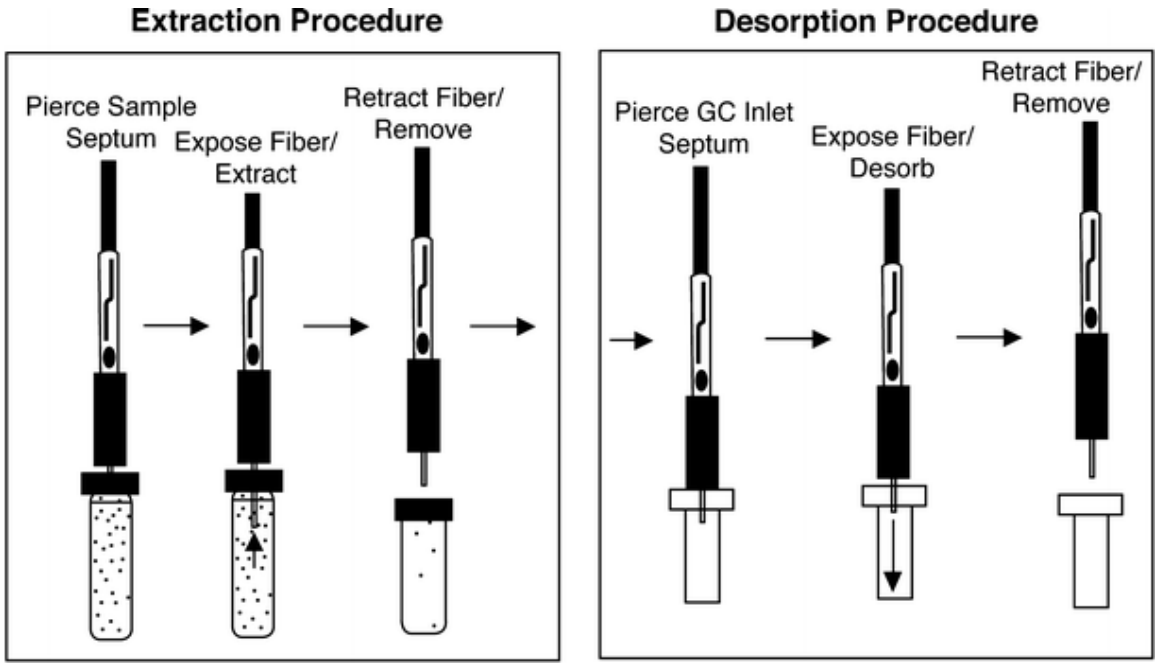


Figure 4:Extraction procedure for HS-SPME-GC-MS[38]

2.7 SOLVENT EXTRACTION

Solvent extraction is a desorption technique commonly used in the analytical chemistry field to separate the compounds of a mixture. This method effectively separates compounds into two immiscible liquids based on their solubility. The immiscible solvents allow the compound to distribute itself between the two solvents. Solvent extraction is based on the use of VOCs as the second organic liquid phase[39] For human scent samples specifically, solvent extraction is used to remove inorganic compounds or deprotonate acids to make them soluble in the aqueous phase. At a specific temperature, the ratio of concentrations of a solute within each solvent reaches an equilibrium[39] This is displayed by the distribution constant (K) in Equation 1.

Equation 1: Distribution Constant[39]

$$K = \frac{S_{org}}{S_{aq}}$$

Equation 1 displays the distribution constant as a ratio between the concentration of analyte within the solvent (S_{org}) and the remaining concentration of the analyte present within the solute (S_{aq}). The partition coefficient is used to assess the efficiency of the extraction. LE is a common sample technique. This technique complements the extraction of semi-volatile and non-volatile compounds, filling in the gaps of the VOCs that cannot be recovered using SPME. One of the most important steps of LE is choosing the appropriate solvent system (two immiscible phases). Most extractions involve water because of its high polarity and compatibility with organic solvents. LE utilizes a large amount of organic solvent to dissolve the analyte of interest. The choice of solvent is influenced by many factors; however, the most significant factors are the selectivity, partition coefficient, volatility, and toxicity of the compound. Once the two layers are identified, the less-dense layer is separated from the remaining layer using a separatory funnel. The product is then isolated from the solvent using rectification, evaporation, crystallization, or further extraction. For human scent extractions, the solvent is evaporated using nitrogen gas for maximum recovery.

2.8 GAS CHROMATOGRAPHY MASS SPECTROMETRY (GC-MS)

Gas Chromatography–Mass Spectrometry (GC-MS) is a technique that conjoins a mass spectrometer and gas chromatography. Chromatography is a separation method in which the partitioning of a sample occurs between the stationary and mobile phase.[40] The variety of chromatographic methods are categorized by the physical state of the mobile phase.[41] This technique has been used to analyze organic and inorganic substances in gases, liquids, and solids;[41] the mobile phase is present in a gaseous state. Gas chromatography is a specific chromatography in which a gas is in the moving phase. This technique is known for its ability to separate and analyze volatile and semi volatile compounds. These compounds typically have higher boiling points. The most significant advantage of gas chromatography is its ability to separate complex mixtures within seconds.[42] Gas chromatography involves a carrier gas, a flow control, an inlet, a column, an oven, and a detector. The inert carrier gas flows constantly through the injection port, the column, and the detector. Mass spectrometry is known for its sensitivity and ability to provide both qualitative and quantitative data on specific analytes.

GC- MS analysis is initiated by placing the sample of interest in a heated injection port. This injection port operates in two different modes (split or split less) which should be selected based on the analyte of interest. Mass spectrometry is an analytical technique utilized for identifying unknown compounds, facilitating the quantitation of known materials, and determining chemical and structural properties of specific molecules. The foundation of mass spectrometry is the production of ions that can be separated according to their mass-to-charge ratio and identified. A mass spectrometer is composed of several elements as presented Figure 5. If a substance can ionize, ions can move in a vacuum within

its respective electric field. Within a mass spectrum, the y-axis represents the abundance, and the x-axis is the mass-to-charge ratio. The chromatogram provides the relationship between the retention time of a substance and the peak intensity. Substances that are eluting at the same time can be detected, and structural information on compounds can be revealed. Two ionization methods are typically used: electron (hard) ionization and chemical (soft) ionization. In GC-MS, electron ionization is the most commonly used ionization method. In electron ionization (EI), the analyte vaporizes into the source and is surrounded by 70-ev electrons.[43] When the energy of the electron beam is transferred to the molecule, a molecular ion is generated. During this process, an electron beam interacts with the molecules that have eluted to form molecular ions.[44] The reaction between the molecular ion and electron is depicted below.

Equation 2: Electron Impact Equation[42]

$$M + e^{-} \rightarrow M^{+\cdot} + 2e$$

Once this process is complete, the ions need to be sorted based on their masses. Once ionization of the molecules occurs, they are transferred into a mass analyzer.[43] There are several mass analyzers, such as the time of flight, quadrupole, orbitrap, and sector. However, when using GC-MS, the quadrupole is most typically used. A schematic of the GC-MS process is depicted in Figure 5.

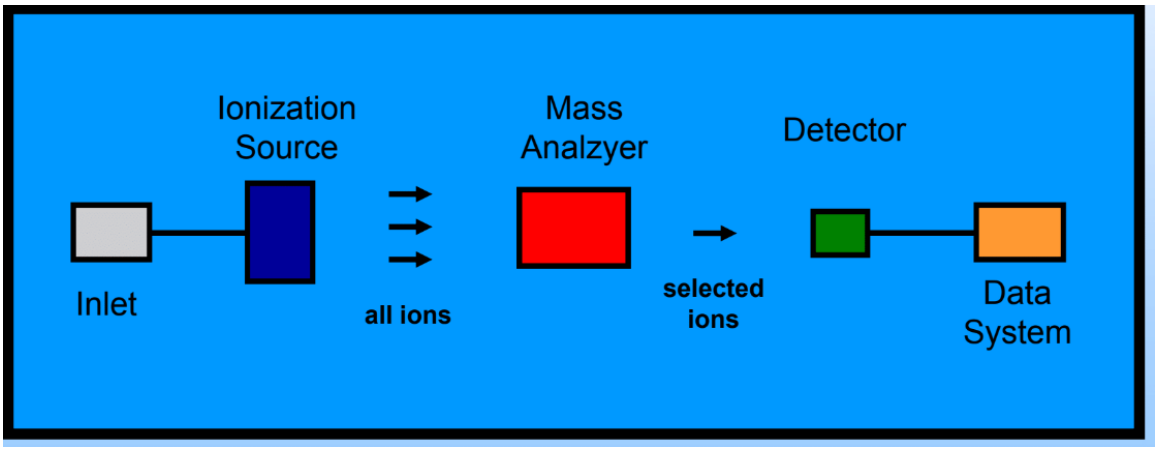


Figure 5:Schematic of GC-MS[42]

GC-MS has been utilized in forensic casework for the identification and quantitation of drugs, explosives, and biological specimens. GC-MS is the analytical instrument of choice when analyzing volatiles of any specimen. Previous studies have highlighted the effectiveness of GC-MS for human scent analysis. Therefore, using two extraction methods simultaneously—GC-MS for VOC identification and quantitation in human scent samples—is hypothesized to provide high-quality results.

3.Materials

3.1 RECRUITMENT OF SUBJECTS

During two separate time periods, authorization was obtained from the FIU Institutional Review Board (IRB). The sole difference between the two sets of samples collected was compensation. During the first sample collection, each subject was rewarded with a $150 Walmart gift card. The second sample collection did not include compensation due to limited funding and less restrictions on participants. A total of 68 subjects between

the ages of 18 and 30 were recruited for the study. Several methods of recruitment were implemented to ensure that all interested participants were obtained. For the first sample collection, participants were recruited via fliers posted on campus (see Appendix), email, and word of mouth. The fliers clearly depicted the commitment necessary for the study and the compensation given for participation. Two different sets of fliers were disbursed to cater to English and Spanish-speaking participants (see Appendix). For the second sample collection, word of mouth and social media networks such as WhatsApp were utilized to identify volunteers. The ethnicities represented in the study included Hispanics, Caucasians, and African Americans. Once each subject was recruited, the protocol was implemented as previously described. Each participant was provided with a handout with detailed instructions on preparation for sampling (see Appendix 1-2). The optimized sampling time, commitment to the discontinuation of routine hygiene products, and depletion of funding caused limitations on subject availability for the project. Overall, the youngest participant in the study was 18 years old and the oldest was 29.

After the recruitment efforts were complete, a total of 68 participants were recruited. Buccal swab samples were obtained from 34 of the 68 subjects for the investigation of human leukocyte antigen genes (HLA). Each subject was asked both their age and which ethnic group they identified with. Out of the total number of subjects sampled, 22 were of Caucasian descent, 31 were of Hispanic descent, and 15 were of African American descent. Because the preliminary study was utilized solely for optimization purposes, the study involved fewer subjects. Three subjects from each ethnicity were assessed.

Table 3:Demographics of populations under study

Race/Ethnicity	Males	Females
Caucasian	10	12
Hispanic	18	13
African Americans	6	9

3.2 SAMPLE COLLECTION MATERIAL

Several textiles have been used to collect volatile organic compounds for human scent evidence including cotton, rayon, polyester, and wool.[27] Cotton has been shown to be the most efficient collection method yielding the greatest number of VOCs and greatest reproducibility.[12] Each subject's odor profile was collected using a 2 x 2 inch 12 ply Dukal gauze pad (Dukal Corporation, Syossett, NY, USA). This sorbent material was stored in a 10 mL glass vial with PTFT/silicone septa (SUPELCO, Bellefonte, PA). After sampling, each gauze pad was transferred using sterilized stainless-steel tweezers (Fisher Scientific, Pittsburgh, PA). Additionally, powder-free nitrile exam gloves were used (one pair per subject) to decrease the likelihood of contaminating the scent profiles with hand odor. Twenty-four hours prior to sample collection, the vial and gauze were conditioned. The vials were rinsed with acetone (HPLC grade, Sigma Aldrich, St. Louis, MO) and heated at 105^{o} C. Additionally, each cotton gauze was treated by spiking 1 mL of methanol directly in the center of the gauze and placing it in clean vial.[27,37]

2.2.3 Development of Chemical Standard Mix

An extensive literature review was conducted to determine which VOCs are significant contributors to human odor. A list of 31 compounds were selected based on their ability to serve as biomarkers for ethnicity when analyzing human scent evidence. Standard reference materials were purchased from Sigma Aldrich (St. Louis, MO). A list of the VOCs previously characterized as odor contributors is depicted in Appendix 4. A mixture of these compounds was utilized for the emulation of human odor. Additionally, this "VOC cocktail" was used to create calibration curves that were used for both qualitative and quantitative purposes. A series dilution of five analyte concentrations (1 ppm, 20 ppm, 40 ppm, and 60 ppm, 80 ppm) was prepared weekly throughout this study. Each tier was placed in a 2 ML glass vial and run using the GC-MS autosampler on a weekly basis. The quantitation of analytes was calculated using the calibration cures constructed on the Enhanced MSD ChemStation (Agilent Technologies, 1989-2006) software. Furthermore, the qualitative assessment of compounds identified in samples by retention time was implemented. Appendix 4 depicts the selected VOCs of interest with their respective retention times.

4.1 RESEARCH METHODS

4.1.1 OPTIMIZATION OF SPME PARAMETERS

There are several types of SPME fibers available commercially. The chemistry of the fibers' polymer coatings varies, portraying polar, non-polar, and mid-polar characteristics.[20] Thus, a critical step in optimizing an efficient SPME extraction is ensuring that the fiber coating selected has an affinity for the VOCs under study. Previous studies suggest that Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS;

SUPELCO, Bellefonte, PA) SPME fiber produces the highest yield when evaluating trace compounds such as human scent.[16,20,37] To further investigate this parameter and test the reproducibility of previous studies, an optimization study was conducted. Six mimic samples were prepared using the chemical standards listed in Appendix 4. The mimic samples were prepared by spiking 5 µl of a 100-ppm standard mix inside of a 10 mL vial. Then, HS-SPME was performed using two different fiber chemistries that are acknowledged to be efficient fibers for human scent VOC analysis:

(1) 50/30 Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS-Grey-Polar)

(2) 75 um Carboxen/Polydimethylsiloxane (CAR/PDMS-Black-Mid Polar; SUPELCO, Bellefonte, PA).[13,16,20,37]

Triplicate extractions were performed for each fiber type. The fibers were exposed for 15 hours, which has been uniformly documented as the most efficient SPME extraction time for human scent compounds.[16,37] However, an extraction time study was conducted to verify the reproducibility of this method.

Once the mimic scent samples were prepared, they were left to equilibrate inside of a dry heating bath at 50^{o} C for 24 hours.[37] Post equilibration, the 50/30 um DVB/CAR/PDMS and 75 um CAR/PDMS fibers were exposed to the headspace of each vial containing axillary (underarm) odor samples for 15 hours. After the extraction was complete, the SPME fibers were unexposed and removed from the headspace vials. Once analyzed via GC-MS, the fiber chemistry and extraction time that allotted for the maximum VOC recovery were selected.

4.1.2 OPTIMIZATION OF SOLVENT EXTRACTION METHOD

Solvent extraction of human scent VOCs was employed to determine the influence that semi-volatile and non-volatile compounds have on odor profiles. These compounds portray high molecular weights and boiling points, which typically cause them to not evaporate in the headspace when performing SPME. A standard mix of three concentrations (20 ppm, 40 ppm, and 60 ppm) was prepared to further investigate the parameters necessary for an efficient solvent extraction of acids known to be contributors to human odor. These acids included 12 of the 33 compounds listed in Table 4.

Table 4: VOC list of compounds extracted by solvent extraction

Compound	CAS Number
Nonanal	124-19-6
Propanoic acid-2 methyl	554-12-1
Decanoic acid, methyl ester	110-42-9
Butanoic acid	107-92-6
Isovaleric acid	503-74-2
Hexadecanoic acid, methyl ester	112-39-0
Undecanoic acid	112-37-8
Dodecanoic acid	143-07-7
Pentadecanoic acid	1002-84-2
n-Hexadecanoic acid	57-10-3
Heptadecanoic acid	50-12-7
Octadecanoic acid	57-11-4

Several GC-MS methods were tested to determine which column would provide the best separation when investigating the acids (SolGel-WAX column or DB5 column). The GC injector mode was constant among all the methods (splitless). However, several variables (temperature ramp, solvent delay time, and hold time) were manipulated within each method to further enhance the chromatographic separation of volatiles. The quality of each method was assessed and the column and method that effectively identified and quantified the most VOCs with the best separation was selected.

4.2 Pilot Study Methodology

4.2.1 COLLECTION OF AXILLARY ODOR SAMPLES

A total of five males per ethnicity (Caucasian, Hispanic, and African American) were selected for preliminary studies. The protocols were approved by the Florida international University Institutional Review Board for research involving human subjects (IRB# 17-0037). Each subject was provided with unscented soap (Tom's of Main, Kennebunk, ME) and instructed to use the soap for two days prior to sampling. Additionally, to eliminate the influence of tertiary odors conflicting with the sample, each subject was instructed to refrain from using any scented products (soaps perfumes, deodorants, body sprays, etc.). Each sampling period was a total of four days. The process for the sampling period is shown in Figure 6.

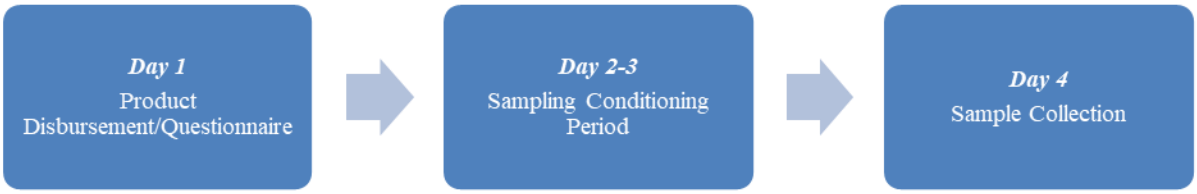


Figure 6:Sampling process for pilot study participants

Human scent samples were obtained by placing a 2 x 2 inch, 12-ply cotton Dukal gauze pad in the center of each underarm as depicted in Figure 6. These gauze pads were placed using disinfected stainless-steel tweezers. Once the gauze pads were placed, each participant was asked to hold their arms down over the gauze pads with limited movement for one hour, Figure 7. A timer was used to ensure that the gauze was removed from each subject's underarm after the same amount of time. The gauze was then transferred into the respective labeled 10 mL vial (which clearly indicated to which armpit the gauze belonged) and was left to equilibrate for 24 hours. The equilibration period facilitated the release of VOCs into the headspace of the vial to promote more efficient extraction with the SPME fiber, thereby providing a snapshot of the VOC profile of each subject.

Figure 7:Gauze placement under human subject axilla (armpit)

4.2.2 EXTRACTION AND ANALYTICAL METHOD FOR THE ANALYSIS OF VOCS IN HUMAN SCENT SAMPLES

Each sample was then subject to exposure utilizing a SPME fiber for 15 hours. After the 15-hour exposure in the headspace of the vial was completed, the SPME fibers were placed inside the injection port of the gas chromatography instrument for 10 minutes to ensure complete thermal desorption of analytes. The gas chromatography (GC) method employed was optimized to begin the separation of volatile organic compounds. The oven ramp was started at an initial temperature of 40° C. This temperature was held for 1.25 minutes. Subsequently, the temperature was raised at a rate of 10° C/minute until it reached 185° C. The temperature was then held at 185° C for a total of 5.42 min. At the same rate, the temperature was increased until it reached 205° C, at which point the temperature was held for five minutes. At the same rate, the temperature was increased until it reached 290° C, at which point the temperature was held for two minutes. The total run time for the GC method was 33 min. The Hewlett Packard 6890 Series GC instrument was equipped with a 30 m x 0.25 mm SGE SolGel-WAX column (30 m x 0.25 mm ID, 0.25 µm phase thickness) and a single quadruple mass analyzer. Helium was utilized as the carrier gas for the system and flowed at a rate of 1.0 mL/min. The ionization method selected for structural elucidation was the electron ionization (EI) mode. The mass spectra were then scanned for a mass range of 45–410 amu and the NIST mass spectra reference library was used to identify the VOCs detected in the samples.

For quality control purposes, after sampling, each SPME fiber was cleaned by exposing it to the GC-MS heated injection port. Furthermore, a blank gauze was run with each sampling set to account for any potential contaminants.

Immediately after the SPME GC-MS analysis was complete, each sample was evaluated using solvent extraction to account for semi-volatile and non-volatile compounds that may have not been recovered using SPME. A protocol adopted from the previous literature was utilized for the extraction parameters.[37] Vials containing gauze pads were filled with 10 mL of methylene chloride and left to equilibrate for 24 hours. After 24 hours, 2 mL of this $MeCl_2$ was transferred into a clean 4 mL vial and evaporated to dryness using nitrogen gas. After this, 300 μl of the solvent (methylene chloride) was reconstituted into the vial and the solution was transferred into a 2 mL ALS vial. The vial was placed into an automated liquid sample (ALS) system and GC-MS was run.

4.2.3 LOGISTIC REGRESSION

Logistic regression is a predictive and explanatory model that provides predictions for a dichotomous outcome.[45] Prediction models apply a statistical model to predict new or future observations.[45] Prediction models are useful for the determination of inclusion criteria and covariate adjustment. Logistic regression models assess the likelihood of future events occurring based on predictor variables. In this dissertation, the predictor variables are the 31 VOCs selected for ethnicity comparison. The logistic function is modeled as follows:

Equation 3: Logistic Function Model

$$p = \frac{e^{\beta 1+\beta_0 x}}{1 + e^{\beta 1+\beta_0 x}}$$

This function graphs an s-shaped curve and is nonlinear. In this equation, p is the probability that a variable will occur and β is the coefficient of the predictor.[46]

In this model, each variable is tested to determine that compound's contribution to the prediction of the outcome. The logistic model was developed by investigating the likelihood of the occurrence of a situation. The estimated odds (x) are the probability that an event will occur divided by the probability that it will not occur.[46] The ratio of these odds is called the odds ratio. The odds ratio is a measure that is directly estimated from the logistic model without the requirement of any special assumption.[47] The odds ratio statistic is calculated using the equation below (see Equation 4).

Equation 4:Odds ratio statistic

$$Odds\ for\ x = \frac{P(x)}{1 - P(x)}$$

Logistic regression has been previously used in the health field to determine the strength of predictors in biological samples such as urine and saliva.[36]

Once a logistic regression is constructed, a Receiving Operating Characteristic (ROC) curve can be developed to assess the true negative rate (specificity) and true positive rate (sensitivity).[48] The accuracy of the test is measured by the area under the curve (AUC). An area of 1 is considered a perfect test. An area of 0.50 is a meaningless test.[48] These rates are calculated using the formulas depicted in Table 5.[48]

Table 5: ROC table calculations

Rate Type	Formula	Variable Explanation
True Positive Rate (Sensitivity)	TP/(TP+FP)	Where: TP = number of true positives FP = number of false positives
True Negative Rate (Specificity)	TN/(TN+FN)	Where: TN = number of true negatives FN = number of false negatives
False Positive Rate	FP/(FP+TN)	Where: FP = number of false positives TN = number of true negatives
False Negative Rate	FN/(FN+TP)	Where: FN = number of false negatives TP = number of true positives

This test can be used to assess the validity of the logistic regression curve. In this research, logistic regression was used to predict the VOCs' contributions to the differences among ethnicities. The statistical software JMP® Student was used to create all the logistic regression plots and ROC curves.

4.2.4 PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) is a multivariate data reduction method used to reduce the variables within a large data set.[49] In recent years, there has been a significant increase in the forensic application of PCA. Principal component analysis has been used to characterize inks, bills, and even voices.[50,51,52] Principal component analysis score plots are

constructed using eigenvectors and eigenvalues. Eigenvectors show where the most variance occurs, and eigenvalues quantify the number of variables within each principal component. Prior to PCA implementation, a scree plot is created to order the eigenvalues in descending order. The scree plot is a graphical representation that reveals how many principal components explain most of the variability in data.[53]

Once a PCA is run, the first principal component (PC1) portrays the largest variance in the dataset, while principal component 2 (PC2) has the second-largest variance in the dataset.[49] The significant variables (VOCs) driving the dataset are established, and score plots are constructed. The score plots developed using PC1 and PC2 show the clustering of each subgroup. In this research, PCA was used to characterize ethnicity using the number of VOCs extracted from each subject's armpit. The statistical software JMP® Student was used to create all the PCA plots.

4.2.5 STUDENT'S *T*-TEST

The student's *t*-test is a parametric statistical analysis method used to validate whether the mean difference between two independent sample populations is statistically significant.

The formula for the student's *t*-test is shown below (see Equation 5). In this formula, x_1 and x_2 are the means of the two sample populations being compared. S_2 is the pooled standard error of both groups. The values n_1 and n_2 represent the total number of observations within each group. The *t* value is utilized to determine whether a null hypothesis will be accepted or rejected. If $t > 0.05$, there is no statistically significant difference between the groups, and the null hypothesis is accepted.[54] On the contrary, if $t < 0.05$, it confirms that the variances between the groups are statistically significant, and the null hypothesis is rejected.[54]

Equation 5:Formula for calculation of student's T-test statistic

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

In this research, the *t*-test was used after logistic regression results were obtained to determine if, when comparing ethnicities, the differences in the average number of VOCs were statistically significant. This technique assisted with the narrowing down of variables prior to PCA implementation.

4.2.6 PRELIMINARY STUDIES

A total of five males per ethnicity were selected for preliminary studies. The protocols were approved by the Florida International University Institutional Review Board for research involving human subjects (IRB# 17-0037). In previous literature, odor has been subdivided into three classes. Primary odor is stable over time and is present due to internal factors such as genetics and bacteria. Secondary odor is present due to diet and other environmental factors. Tertiary odor consists of odors present due to the use of commercial products (soaps, deodorants, perfumes, etc.).[16] Prior to the product disbursement, a protocol was developed to ensure that the experimental design would facilitate the investigation of primary odor, document secondary odor, and control for tertiary odor. Upon product disbursement, each subject was asked to sign a consent form. Additionally, a questionnaire was provided to each subject to provide insight on their lifestyles and ethnic backgrounds. The sampling protocol consisted of a conditioning period in which subjects were instructed to wash solely with the products provided to them. The subjects were provided with Tom's of Maine soap to use for washing for three days.

Once the conditioning period was complete, 2 x 2-inch 12 ply Dukal gauze pads (Dukal Corporation, Syosset, NY, USA) were removed from 10 mL headspace vials with PTFE/Silicone septa (SUPELCO, Bellefonte, PA) using tweezers. Gauze pads were then placed under the armpits of each subject. The subjects were instructed to remain seated for one hour while holding the sampling media under their armpits. At the end of the sampling period, the gauze was obtained from the subjects using tweezers and stored in 10 mL vials. Immediately after SPME GC-MS analysis was complete, each sample was evaluated using solvent extraction to account for semi-volatile and non-volatile compounds that may not have been recovered using SPME. A protocol adopted from previous literature was utilized for the extraction parameters[37] Vials containing gauze pads were filled with 10 mL methylene chloride and left to equilibrate for 24 hours. Then, 2 mL of $MeCl_2$ were transferred into a clean 4 mL vial and evaporated to dryness using nitrogen gas. Next, 300 μl of the solvent (methylene chloride) was reconstituted into the vial, and the solution was transferred into a 2 mL vial. The vial was placed into an automated liquid sample (ALS) system and analyzed via GC-MS.

4.4 RESULTS AND DISCUSSION

4.4.1 DEVELOPMENT OF CHEMICAL STANDARD MIX

An extensive literature review was conducted to determine which VOCs are significant contributors to human odor. Standard reference materials were purchased from Sigma Aldrich (St. Louis, MO). A list of the VOCs previously characterized as odor contributors is depicted in Appendix 4. A mixture of these compounds was utilized for the emulation of human odor. Additionally, this VOC “cocktail” was used to create calibration

curves that were used for both qualitative and quantitative purposes. A series dilution of five analyte concentrations (1 ppm, 20 ppm, 40 ppm, 60 ppm and 80 ppm) was prepared weekly. Figure 1 displays a typical calibration curve obtained from using the *Enhanced MSD Chemstation*© analysis software. Using this curve, the number of VOCs present within each sample was calculated and the identification of compounds within each sample was undertaken.

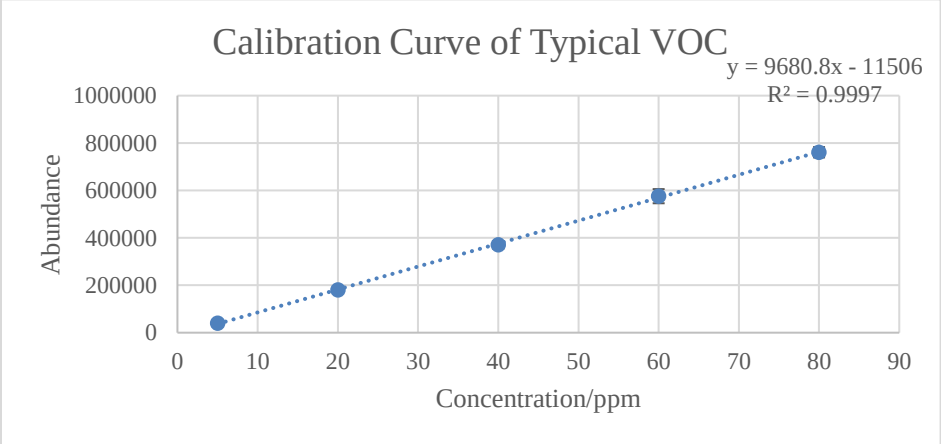


Figure 8:Calibration curve of human scent VOC

4.4.2 OPTIMIZATION OF SPME PARAMETERS

The determination of the best SPME fiber for human scent was evaluated by testing the efficacy of two different fiber chemistries previously studied. As visualized in Figure 8, the DVB/CAR/PDMS-Grey-Polar fiber was shown to be optimal for the mimic samples based on the yield of analyte produced. When a one tailed *t*-test was performed, the *p*-value was 0.0210, meaning that there was a statistically significant difference in the abundance extracted per fiber. Furthermore, the fiber chemistry in the DVB/CAR/PDMS-Grey-Polar fiber extracted six more compounds than its competitor. As seen in Figure 10, 81.25% of

the VOCs of interest were recovered using the grey fiber while only 62.5% of VOCs were recovered using the black fiber. This study validated that the optimization techniques were in fact reproducible and served as the best techniques for the extraction of human scent VOCs.

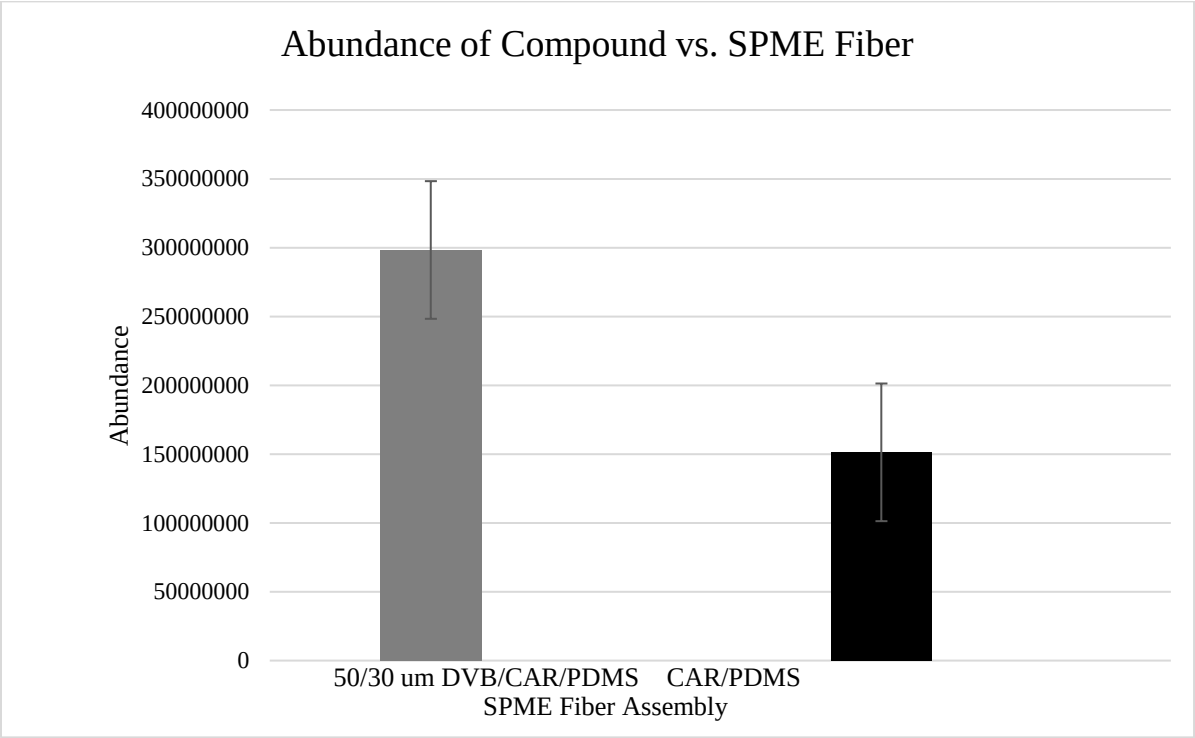


Figure 9:Recovery of chemical reference standard mix using two fiber assemblies

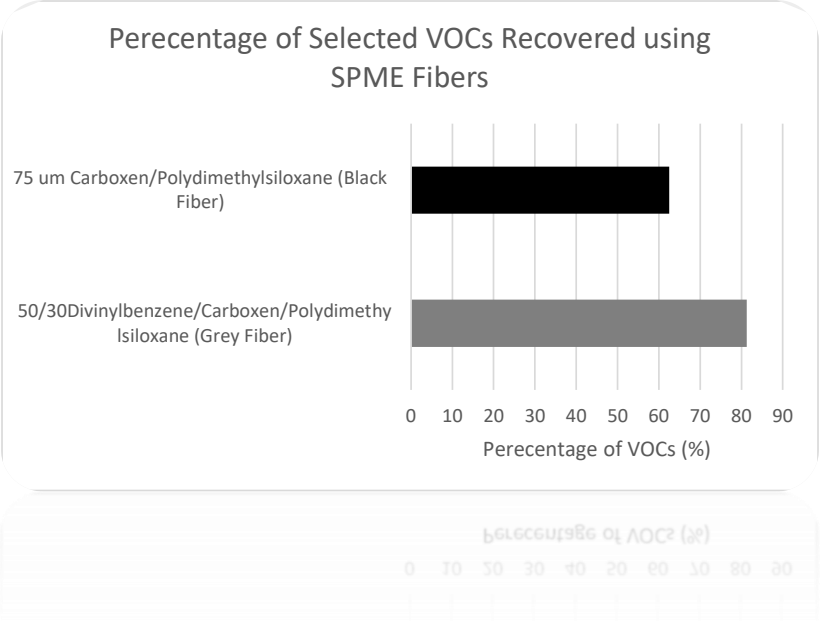


Figure 10:Percentage of VOCs in standard mix recovered using SPME fibers

An extraction study was also implemented to assess the effect of time on VOC extraction. Three samples consisting of 5 ul of the 100-ppm standard mix were spiked and placed into separate headspace vials. Since the 50/30Divinylbenzene/Carboxen/Polydimethylsiloxane fiber extracted the most VOCs with maximum abundance, this fiber was exposed into each respective headspace vial. Different times were selected for extraction (1 hour, 8 hours, and 15 hours). There was a direct correlation between increasing extraction times and the number of VOCs recovered (see Figure 11). These extractions were run in triplicate to assess the reproducibility of the results. The study revealed that the 15-hour extraction time recovered the most VOCs consistently. These findings validated that the 15-hour extraction suggested in previous works was in fact reproducible even when investigating human scent VOC profiles. Further statistical analysis was implemented to ensure that the most reproducible extraction time was selected. The RSD percentages were calculated for each

extraction time using the average total amount of VOC present within each extraction time. The RSDs revealed that the 15-hour extraction time was the most reliable and reproducible (see Figure 12). The results of these study confirmed that a 15-hour extraction using the 50/30 um DVB/CAR/PDMS fiber was the best SPME extraction method for human scent analysis.

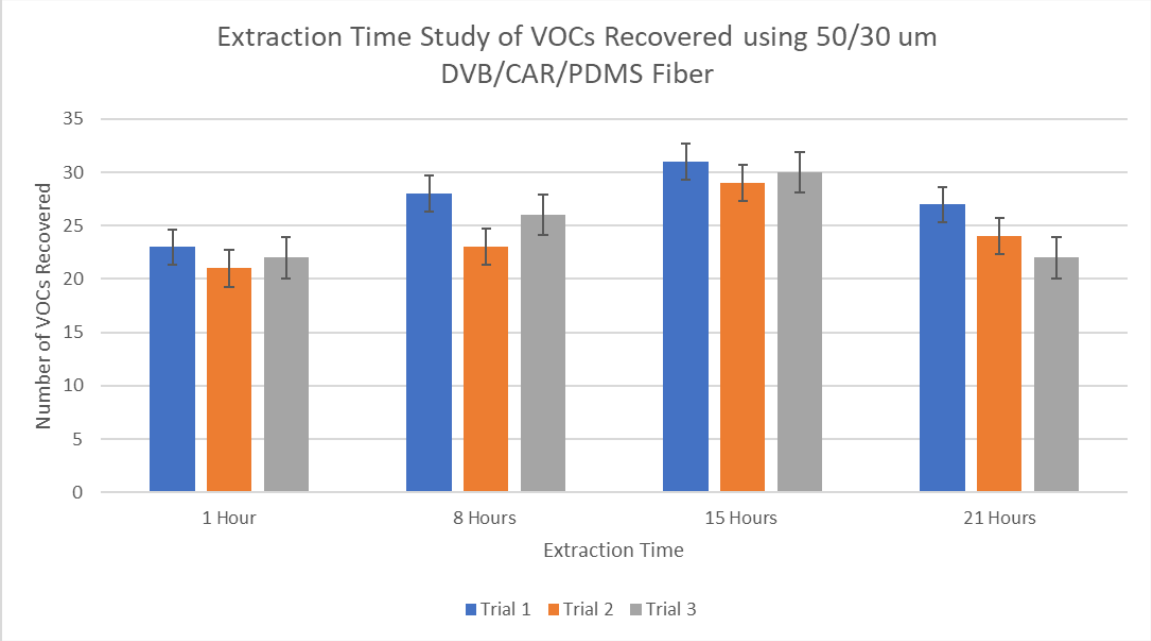


Figure 11: Number of compounds recovered when extraction time is manipulated

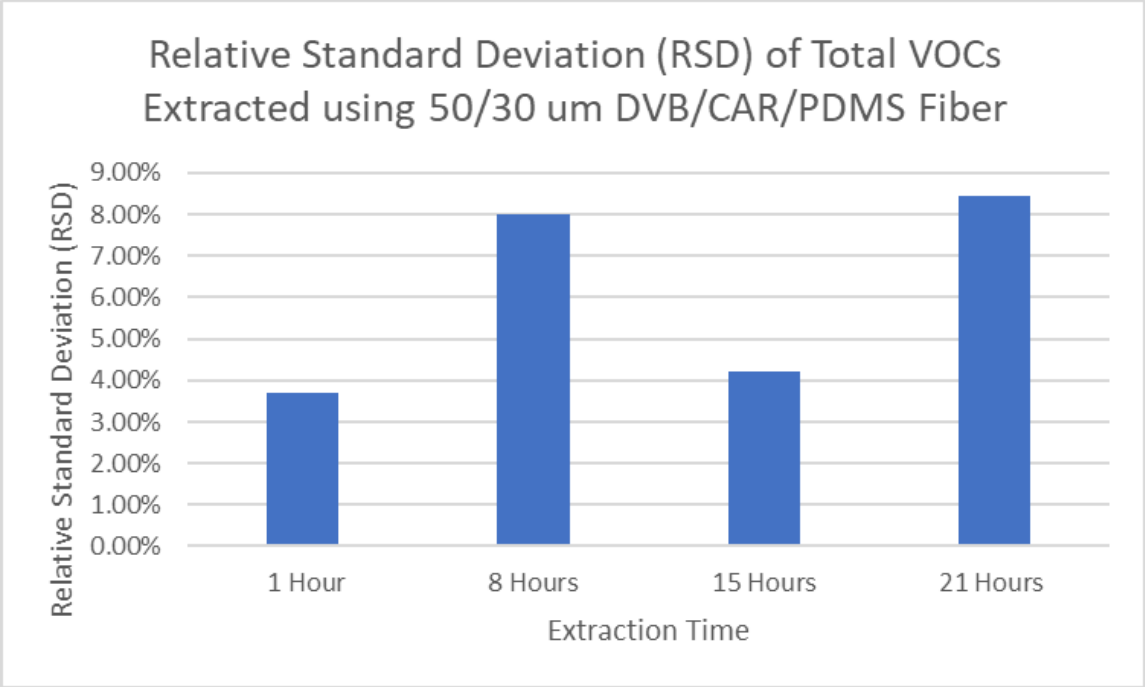


Figure 12:Relative standard deviation of total VOCs extracted using 50/30 um DVB/CAR/PDMS Fiber

4.4.3 OPTIMIZATION OF SOLVENT EXTRACTION PARAMETERS

A gauze saturated with 5 ul of a standard mix of 12 acids was used to analyze the performance of two different instruments or columns.

a) Agilent Technologies Hewlett Packard 6890 Series GC-MS instrument (equipped with a SGE SolGel-WAX column of 30 mm x 0.25 mm ID, 0.25 µm phase thickness).

b) Agilent Technologies Hewlett Packard 6890 Series GC-MS DB5 (equipped with a column of 30 m x 0.25 mm ID, 0.25 µm phase thickness).

Both instruments were equipped with single quadruple mass analyzers. Four analytical methods were developed to determine which method would provide the most sensitive, reproducible results when analyzing semi-volatile and non-volatile VOCs. Each method used the splitless injector mode, which previous studies found revealed the best sensitivity when analyzing human scent VOCs. Furthermore, the oven temperature ramps were manipulated to effectively separate co-eluting compounds with close retention times. The methods selected are depicted in Table 6.

The DB5 column demonstrated poor separation, and the co-elution of compounds that were close in retention time was challenging to resolve. Furthermore, only 19 of the 31 compounds were detected within the mixture. The SolGel-WAX column allowed for the separation and detection of all compounds of interest, and co-elution was resolved by changing the oven temperature. This resulted from the fact that the analyte of interest was polar, and the similar chemical structure of the SolGel-WAX column allowed for optimal separation. Figure 13 shows the average total peak area when comparing the two SolGel-WAX column methods. Method C was selected as the optimal method. The temperature of 205° C allowed for the complete separation of VOCs and resulted in a higher abundance when assessing the 20 ppm, 40 ppm, and 60 ppm standard mixes. The additional time added to the solvent delay prevented the solvent from prematurely eluting from the column.

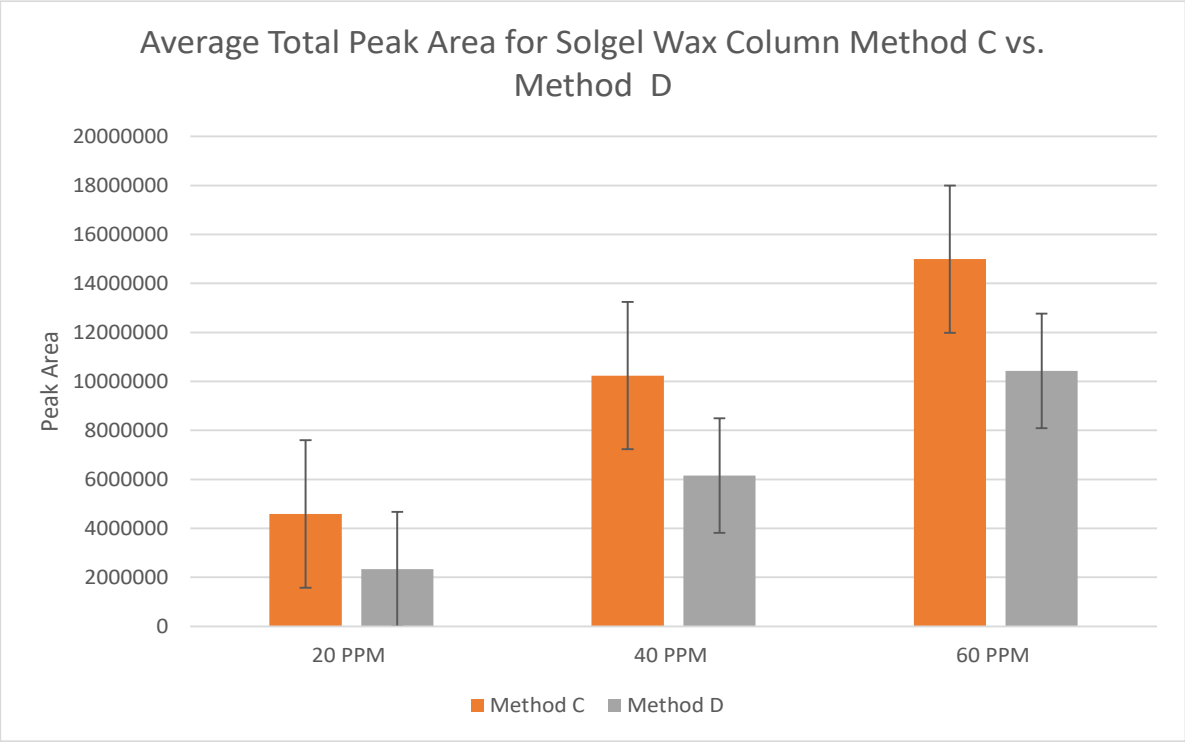


Figure 13:Average total peak area when comparing Method C vs. Method D

Table 6: Parameters of methods for GC-MS analysis

Method	Injector Mode	Solvent Delay	Initial Oven Conditions		Oven Temperature Cycle Sequence		
				Rate of Change (°C/min)	Final Temperature (°C)	Final Time (min)	Total Run Time (min)
Method A (DB5 Column)	Splitless	N/A	40° C for 1.25 min	10	50	0	33.54
				10	185	5.42	
				10	205	5	
				10	218	0	
				10	280	2	
Method B (DB5 Column)	Splitless	N/A	40° C for 1.25 min	10	50	0	44.54
				10	185	5.42	
				10	205	5	
				10	218	11	
				10	280	2	
Method C (Solgel Wax Column)	Splitless	4 min	40° C for 1.25 min	10	50		

				10	140	0	43.78
				10	175	5.42	
				10	200	6.11	
				10	230	3	
				10	260	1	46.42
Method D (Solgel Wax Column)	Splitless	7 min	40° C for 1.25 min	10	40	0	
				10	50	5.42	
				10	185	5	
				10	205	29.47	
				10	218	0	
				0	280	2	
				10	50	0	33.02
				10	185	5.42	
				10	205	5	
				10	290	2	

4.4.3 OPTIMIZATION OF SOLVENT EXTRACTION PARAMETERS

A total of five individuals were selected for the sample to ensure the efficiency of the optimized method before proceeding with the subjects for compensation. A total of 21 compounds were initially investigated. The color-coded stacked columns presented in **Error! Reference source not found.** illustrates the scent profiles extracted from the Hispanic subjects. Each color portrayed in the profile corresponds to a compound extracted

using SPME. The length of each colored bar is directly correlated with the amount of VOC present within the sample. This chart is used as a visual aid in making qualitative inferences. Additionally, statistical evaluations such as PCA were implemented to identify the driving forces behind the data and the potential for grouping by ethnicity (JMP® Student). The first two principal components are plotted in Figure 15; this score plot reveals a clear separation between Caucasian and Hispanic males within the sample. Notably, PC1 and PC2, collectively, were responsible for 62.7% of the variation. This variation enabled the separation of the two ethnicities, which confirms that the compounds depicted in the loading plot were suitable for ethnic group differentiation.

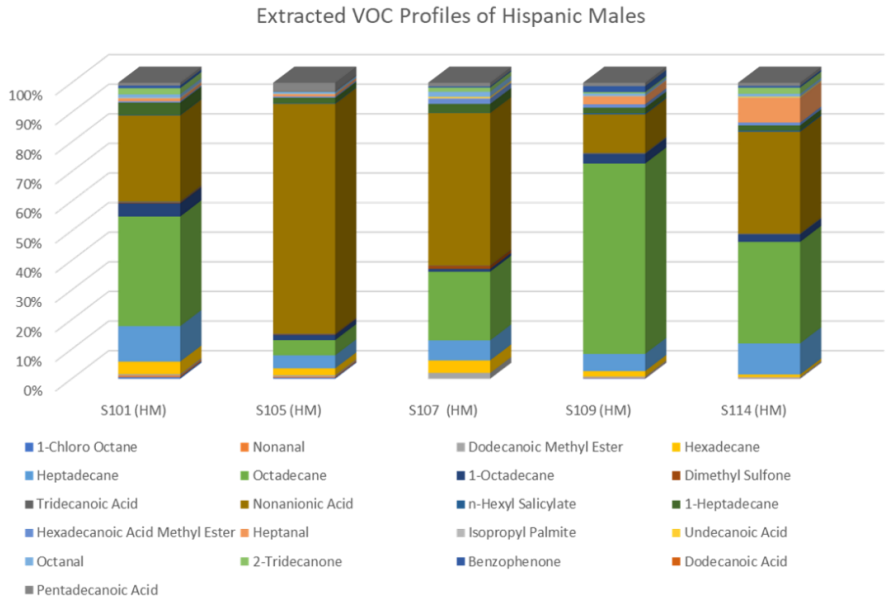


Figure 14:Odor profiles of Hispanic subjects extracted via SPME/LE

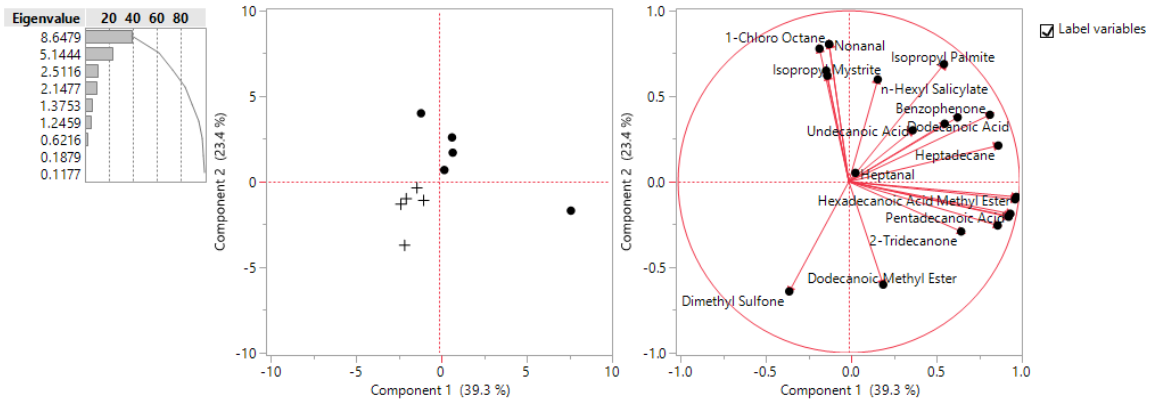


Figure 15:Preliminary PCA of Caucasian vs. Hispanic subjects

4.4.4 PILOT STUDY

4.4.4.1 QUALITATIVE OBSERVATIONS

A total of 136 underarm odor samples were collected from the 68 subjects. Samples from each subject's right armpit were utilized for qualitative and quantitative analysis. A total of 68 subjects between the ages of 18–30 were recruited for the study. The participants were recruited via flyers posted on campus, through email, and by word of mouth. The ethnicities represented in the sample included Hispanics, Caucasians, and African Americans. The subjects were compensated with a $150 Walmart gift card for their completion of the study. Once each subject was recruited, the previously described protocol was implemented.

The VOCs collected from each subject were captured on gauze pads and extracted using HS-SPME and solvent extraction. Figure 16 displays an overlay of a typical chromatogram obtained using the HS-SPME and solvent extraction processes. As the

figure shows, different compounds were detected using these extraction methods. The compounds with higher molecular weight and high boiling points were more readily detected using solvent extraction. The other compounds that were extracted likely included secondary and tertiary odor contributors as well as siloxane peaks from the fiber chemistry. Coupling solvent extraction to SPME ensures that a full VOC profile is obtained. When only HS-SPME is used, compounds with high volatility evaporate. Although several peaks were identified within the samples, only the 33 compounds listed in Appendix 4 were analyzed using standard reference materials. An external calibration curve was composed weekly to quantify the VOCs of interest. In total, 12 compounds were quantified using the solvent extraction method due to their high boiling points and molecular weights. The other 21 compounds were extracted via HS-SPME.

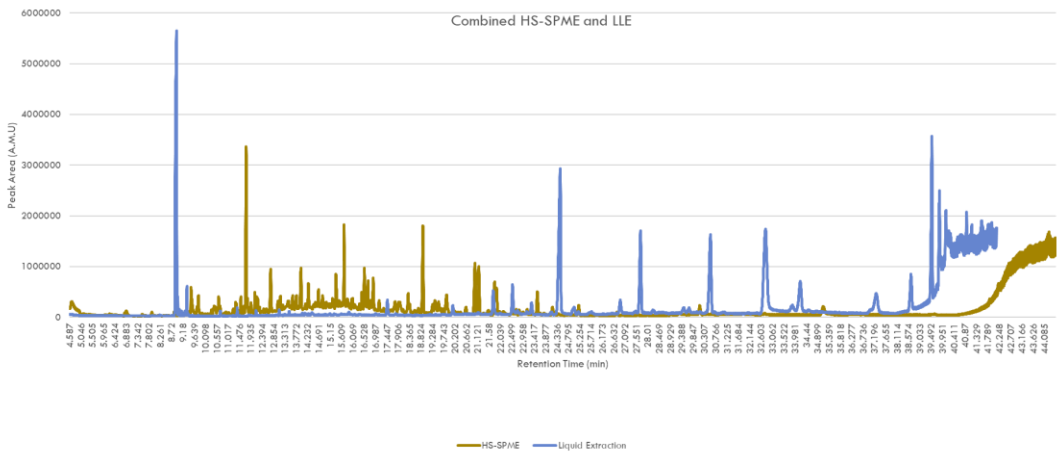


Figure 16: Chromatogram of headspace solid-phase microextraction (HS-SPME) vs. the LE extraction method

The author compiled a list of 33 VOCs, which were selected for the quantitative and qualitative analysis of Caucasian, Hispanic, and African American subjects. The individual VOC profiles were collected and represented by percent on a stacked column

color chart that depicts the peak areas of each VOC extracted from the axillae of each subject. This chart portrays the prevalence of each compound collected from a subject, and each color represents a particular VOC. The trends in the data were qualitatively assessed by indicating any significant observations that could potentially connect an individual to a specific ethnicity.

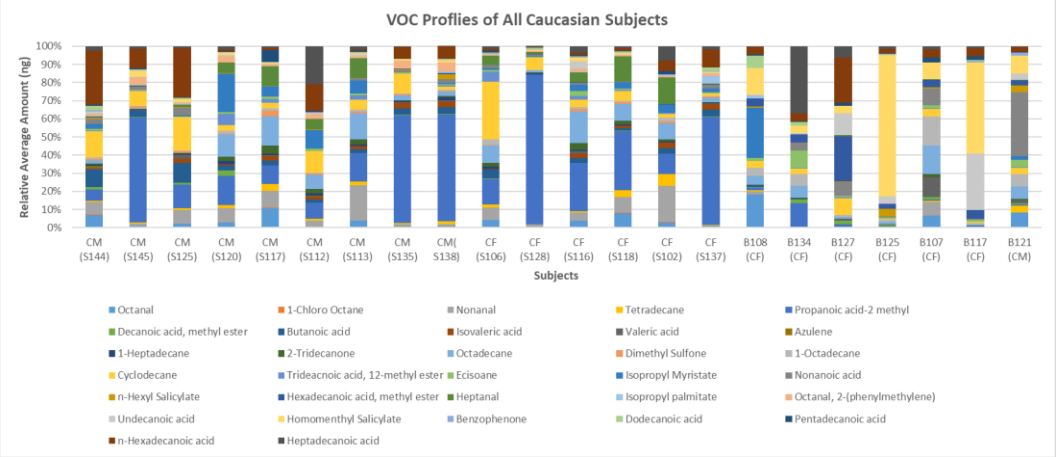


Figure 17: The VOC profiles of all Caucasian subjects

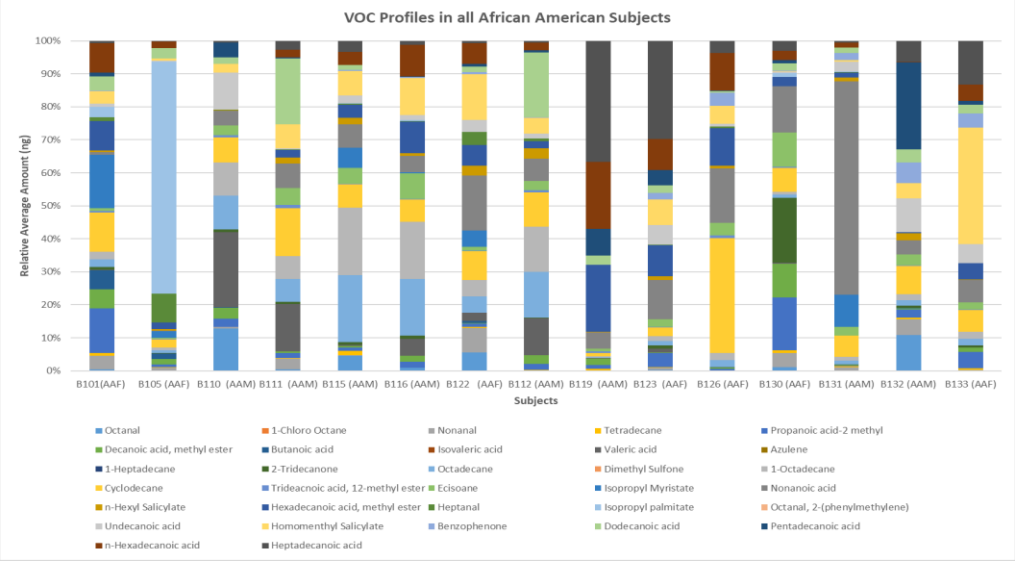


Figure 18: The VOC profiles of all African American subjects

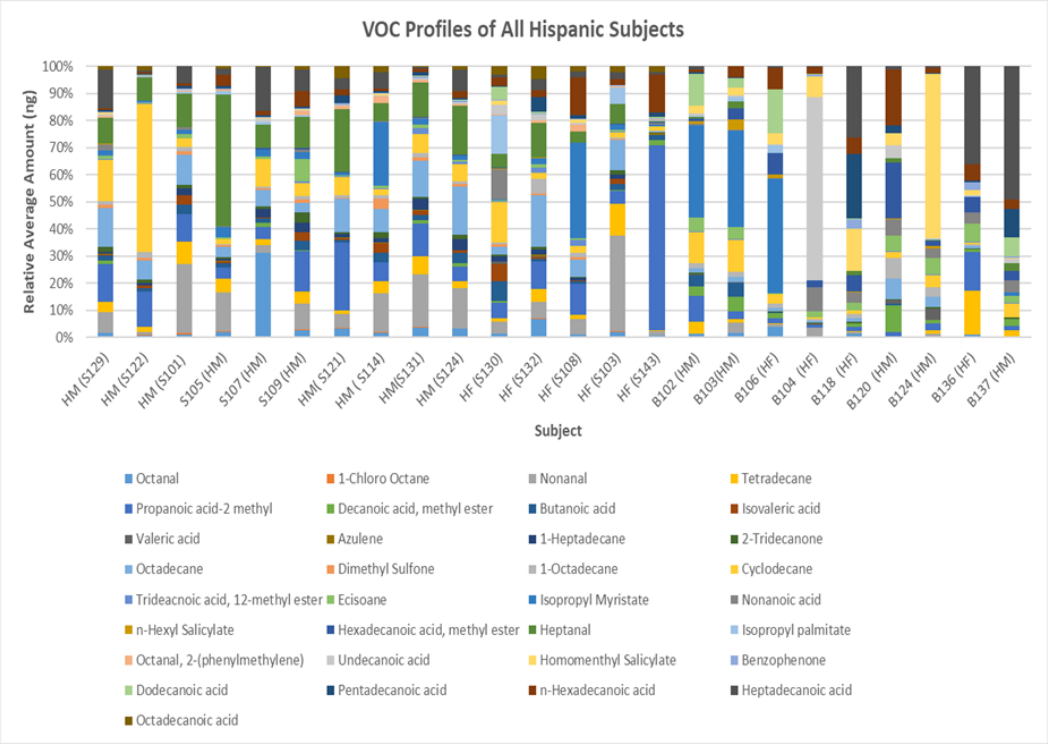


Figure 19: The VOC profiles of all Hispanic subjects

The myriad differences that appear in the above figures are as a result of the different concentrations of each VOC present within the participants' samples. The visual representation highlights several differences between the participants of different ethnicities. One can make several notable observations about the compounds extracted via solvent extraction. Previous studies have revealed that axillary odor is composed of a mixture of several acids (normal, branched, and unsaturated).[55] In individuals from different ethnic backgrounds, one can identify specific trends in the presence of these acids—trends that are sufficient to identify specific ethnic groups. The African Americans within the sample produced larger amounts of short chained acids (butanoic acid, valeric acid, and isovaleric acid). These findings correspond to a study that revealed that African

Americans produce slightly higher levels of these short chained acids than East Asian and Caucasian individuals.[22] Furthermore, the African Americans in the sample also exhibited higher levels of long chained acids (pentadecanoic acid, n-hexadecenoic acid, heptadecanoic acid, octadecanoic acid, and dodecanoic acid), with Hispanic subjects having the next highest levels. In contrast, Caucasians exhibited the highest levels of branched acids (isovaleric acid; propanoic acid, 2-methyl).

The qualitative analysis (frequency of occurrence, see Figure 20) performed for this paper is represented through a histogram. The histogram in Figure 20 displays how frequently all the VOCs selected for this study were extracted via HS-SPME and solvent extraction.

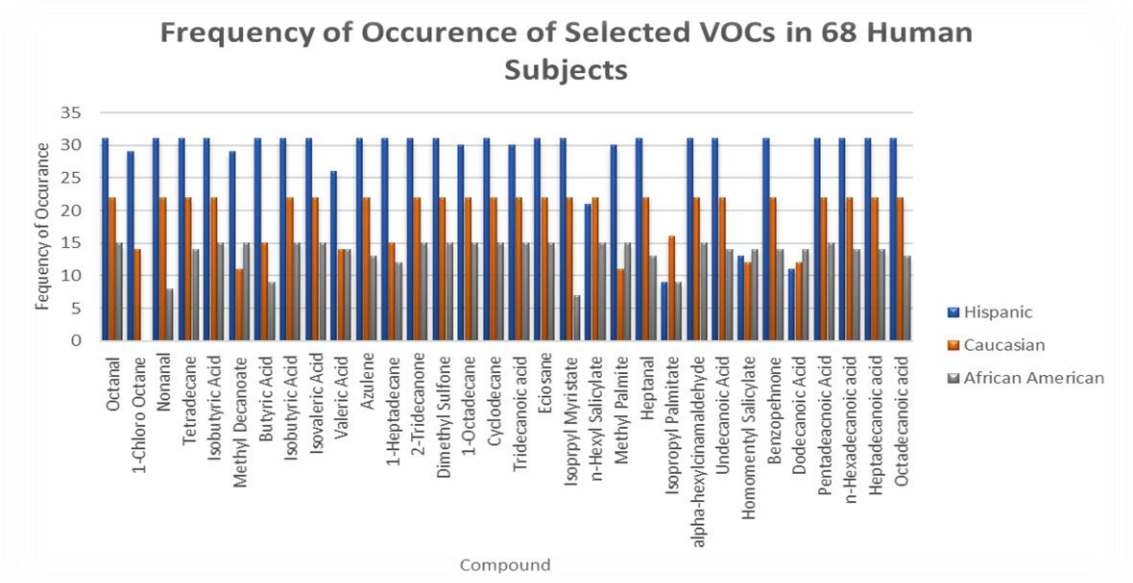


Figure 20: Frequency of occurrence of VOCs according to ethnicity

Among the samples collected from the 31 Hispanic subjects, the author identified the presence of 29 compounds at high frequency (22–31), three at medium frequency (14–

21), and one at low frequency (0–14). Among the 22 Caucasian subjects, there were 29 compounds present at high frequency (15–22), three at medium frequency (7–14), and one at low frequency (0–7). Among the 15 African American subjects, 28 compounds were present at high frequency (11–15), four were present at medium frequency (5–10), and one was present at low frequency (0–5). However, this qualitative analysis also highlighted that two VOCs (isopropyl palmitate and isopropyl myristate) are not ideal for the identification of ethnic groups. This is due to their low frequency of occurrence in African American and Caucasian subjects. The results indicate that the high frequency of occurrence of the VOCs selected for this study further validates the consistency of these markers among all human scent samples (with the exception of isopropyl palmitate and isopropyl myristate). Furthermore, one VOC octanal 2-phenylethylene was not found in any of the African American subjects. This indicates that this VOC can be utilized to exclude African American suspects in a criminal investigation (via scent lineups). Statistical analyses were implemented to further assess each VOC's capability of differentiating between ethnicities. The statistical analysis process is shown in Figure 21.

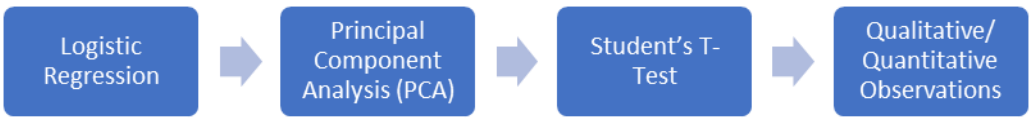


Figure 21: Statistical analysis process for VOC study

Data management and analysis were performed using JMP® Student. The variety of compounds utilized to evaluate human scent profiles made it challenging to identify patterns and relationships. To facilitate the analysis of these complex profiles, a predictive modeling technique (logistic regression) was applied to determine which variables can be used to effectively differentiate between ethnic groups. The benefit of analyzing the profiles using logistic regression is that it considers how each independent variable affects the data.[46] Each logistic plot was investigated between two different ethnicities simultaneously. This method simplifies the identification of unique VOC patterns among populations. To minimize the risk of misclassification, a decision boundary was established in which the relationship between a compound and an ethnicity is dependent on the *p*-value.[36] In cases where $p < 0.05$ for a particular compound, that compound was accepted as a key compound in ethnicity differentiation. Once the variables were assessed, improved models were selected for PCA implementation. Furthermore, a student's *t*-test was run on each selected compound to determine if the quantitative differences were statistically significant. Two sample populations of different ethnicities were assessed at a time. The comparisons are listed in Table 7.

Table 7: Ethnicity comparison breakdown

Populations of Interest for Comparison
Caucasian vs. Hispanic (A)
Caucasian vs. African American (B)
Hispanic vs. African American (C)

4.4.4.2 STATISTICAL ANALYSIS

4.4.4.2.1 CAUCASIAN VS. HISPANIC SUBJECTS RESULTS

As part of the data analysis process, two ethnicities were compared at a time to determine whether logistic regression can effectively predict which compounds cause variation among ethnic groups (particularly Caucasian vs. Hispanics). The odor profiles of all Caucasian and Hispanic subjects were evaluated. Prior to PCA implementation, a logistic plot was constructed to narrow down potential VOCs which can be used to efficiently distinguish between Caucasian and Hispanic subjects. Figure 22 offers a logistic plot that compares Caucasian and Hispanic subjects.

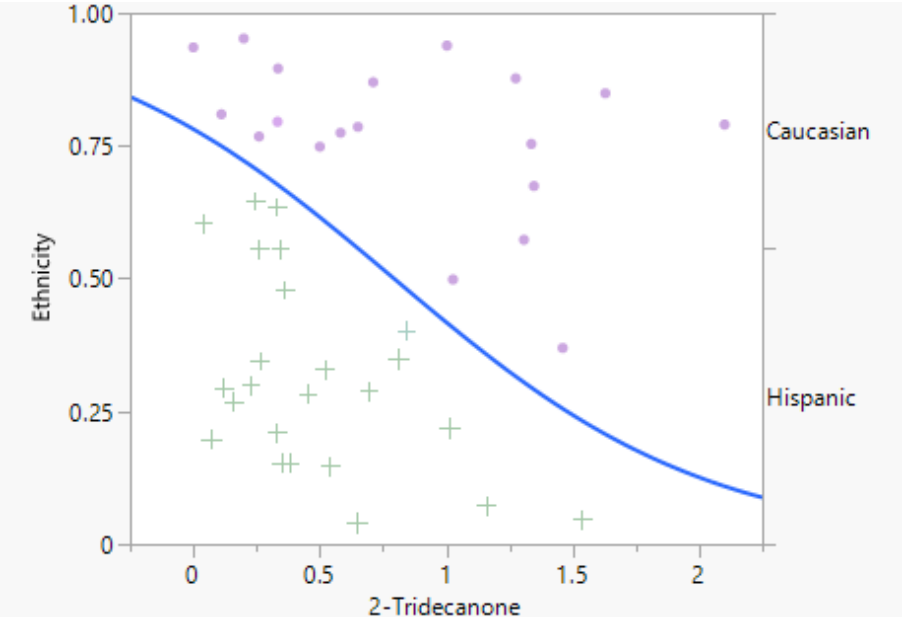


Figure 22: Logistic regression graph of 2-tridecanone

This plot illustrates how the VOC 2-tridecanone can be used to predict ethnicity when comparing Caucasian and Hispanic subjects. Of the 31 compounds selected for analysis, the logistic plot includes three compounds extracted using SPME and two compounds extracted via solvent extraction. This qualitative observation highlights the advantage of coupling SMPE and solvent extraction to obtain a complete profile of human subjects. Once the contributing VOCs were selected, an ROC curve was constructed to provide a graphical representation of the classification effectiveness. The sensitivity and specificity of the results were determined, and the performance of the test was evaluated by analyzing the area under the ROC curve (AUC).

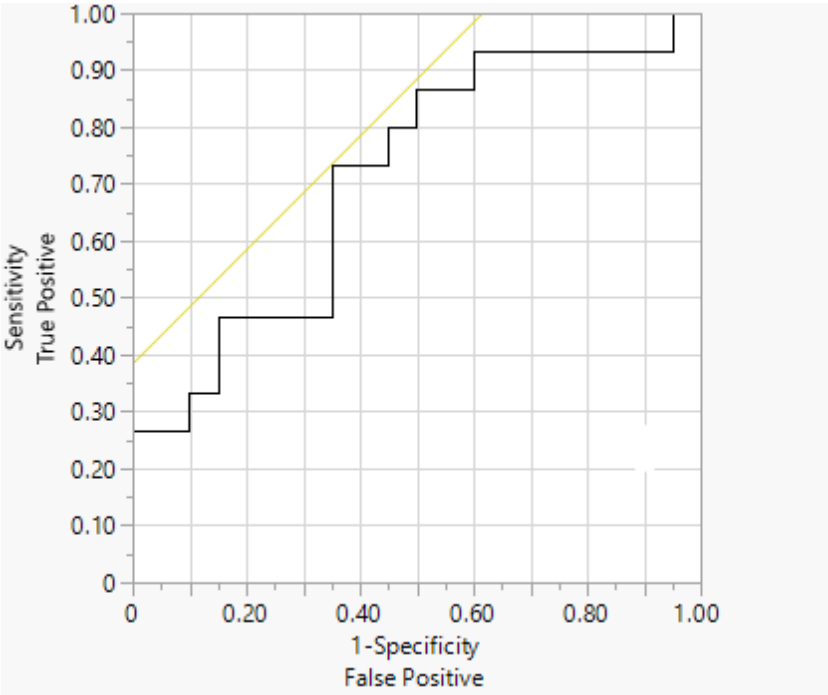


Figure 23: ROC curve of 2-tridecanone

Table 8: Performance of logistic regression model for Caucasian vs. Hispanic classification

Compound	*P*-value	Extraction Method Used	Area Under the Curve (AUC)	Sensitivity (True Positive Rate)	Specificity	False Positive Rate	False Negative Rate
2-tridecanone*****	0.0154	SPME	0.69318	70.83%	39.02%	31.82%	29.17%
Octanal-2 phenylmethylene *****	0.0219	LE	0.81667	69%	62%	6.67%	31.25%
Octadecane	0.0322	SPME	0.67194	69.57%	42.29%	27.27%	30.43%
Tridecanoic acid	0.0336	LE	0.62689	54.17%	26.89%	27.30%	45.83%
Dimethyl sulfone*	0.0237	SPME	0.7133	73.3%	38.30%	35.00%	26.67%

In cases where there is no difference between the two populations (Caucasians and Hispanics), the AUC of the variable (VOC) under study is equal to 0.50. An AUC of 1 would indicate a perfect separation of the populations. Since all the selected compounds' AUCs had an AUC greater than 0.50, they were found to be significant. However, as a general rule of thumb, when constructing an ROC curve, an AUC of 0.50–0.70 is considered poor discrimination, and an AUC of 0.70–0.80 is deemed to be acceptable discrimination.[56] Therefore, only the compounds with an AUC of 0.69 or higher were accepted for the PCA model (2-tridecanone, octanal 2-phenylmethylene, and dimethyl sulfone). All other compounds (octadecane and tridecanoic acid) were excluded due to their poor ability to discriminate between Caucasian and Hispanic subjects.

The information generated from the ROC curve was translated into an ROC table. The ROC table assesses the sensitivity and specificity of each logistic plot. An optimal threshold to determine the point at which the most correct classifications will occur was identified.[56] This was determined by selecting the cutoff that produced the highest sensitivity without having significant adverse effects on specificity. The ROC tables reveal that the VOC dimethyl sulfone showed the greatest sensitivity when differentiating between Caucasians and Hispanics (73.30%). The closer that a sensitivity value is to 100%, the more sensitive the compound for ethnicity-specific differentiation is. However, the VOC that showed the highest level of specificity (true negative rate) at 62% was octanal 2-phenylmethlene. As with the sensitivity value, a specificity value of 100% is perfect.

Unlike true rates, when analyzing false rates (false positives and false negatives), the closer the rate is to zero, the less likely a false positive or negative is present; in other words, a false positive or false negative rate of zero means that no false positives are present, which is ideal in forensic analysis. Octanal 2-phenylmethlene also had the lowest false positive rate (6.67%). Dimethyl sulfone had the lowest false negative rate. Figure 8 shows the false positive and false negative rates when using each compound for classification purposes. The two compounds with the greatest false positive and false negative rates (octadecane and tridecanoic acid, respectively) were excluded from the PCA plot because they did not meet the AUC cutoff. This graph reveals that the lower the AUC, the higher the false positive and false negative rates.

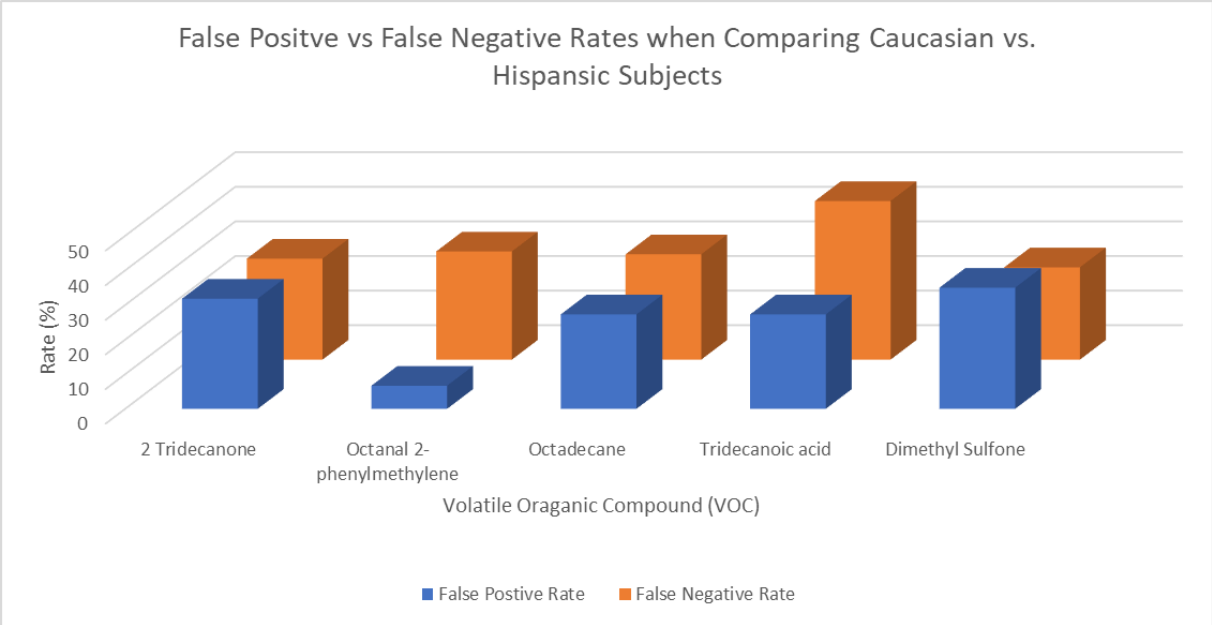


Figure 24: Comparison of false positive and false negative results when comparing Caucasian vs. Hispanic subjects

A PCA plot of the original 31 compounds was constructed to evaluate the abilities of the compounds to discriminate between Caucasian and Hispanic subjects. This initial PCA plot revealed that using all 31 compounds led to poor discrimination. The overlap between Caucasian (.) and Hispanic (+) subjects was evident. In other words, running a PCA with 31 VOCs did not facilitate differentiation between Caucasians and Hispanic subjects. After assessing the low PC1 (10.6%) and PC2 (18%), a scree plot was constructed to identify how many eigenvectors had to be incorporated into the score plot to explain the variance among ethnicities. The scree plot shows that selecting a total of three compounds for the PCA plot accounts for over 80% of the variation between ethnicities.

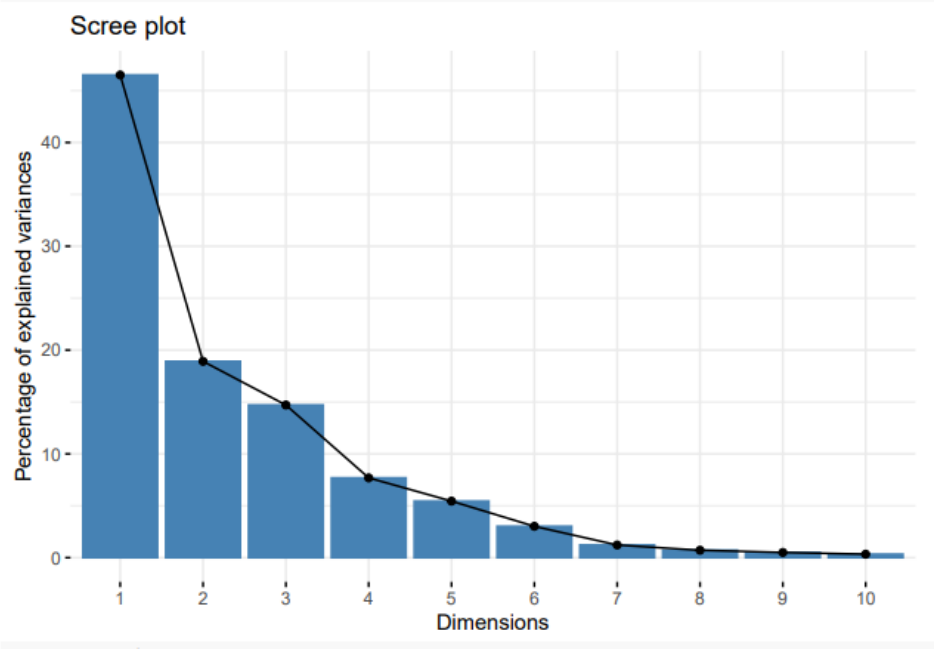


Figure 25:Percentage of explained variances when assessing all VOCs capability of differentiating between ethnicities

A new PCA plot was constructed using compounds that were predicted via logistic regression that passed the Student's *t*-test (2-tridecanone, octanal 2-phenylmethylene, and dimethyl sulfone). Using only the three compounds predicted via logistic regression resulted in a significant increase in variation. This variation increase resulted from the adjustment of the discrimination variables, whereby only those statistically significant compounds predicted via the decision boundary that met the AUC threshold were selected for the logistic regression model. The percentage in PC1 increased from 10.6% to 53.9%, while the percentage in PC2 increased from 18% to 31.8%. However, the variation increase did not promote clustering among Caucasian and Hispanic subjects as expected. The overlap between Caucasian and Hispanic individuals is a result of self-identification bias; diversity within the Hispanic race leads many Hispanics to identify as Caucasian. Although

Hispanic origin and race are two separate concepts, individuals typically use self-identification to determine how they are classified.[57] According to the 2000 U.S. Census, Miami has the greatest number of Caucasian-Hispanic people in the U.S.[57] Since all of this study's samples were collected in Miami, Florida, it is suspected that some subjects of Hispanic origin may have identified as Caucasian, hence causing the overlap that appears in the PCA plot.

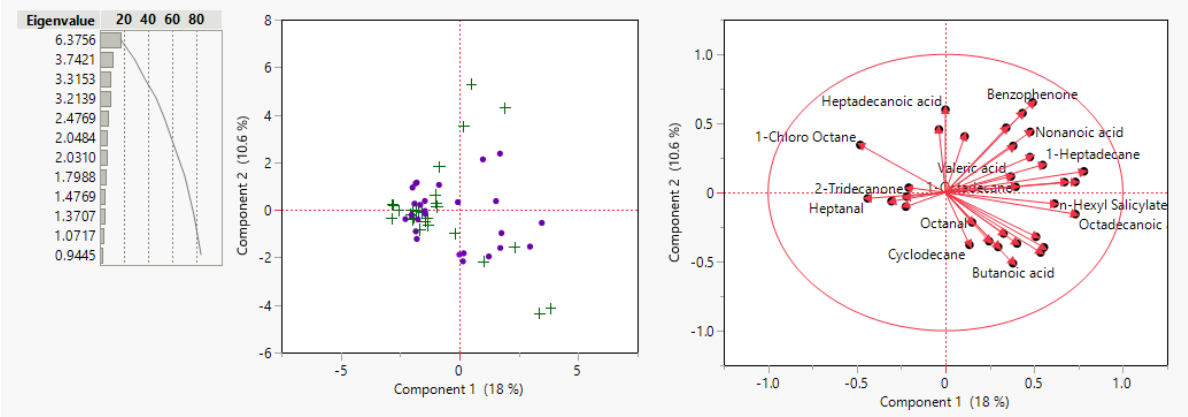


Figure 26: PCA plot of Caucasian vs. Hispanic subjects using all VOCs

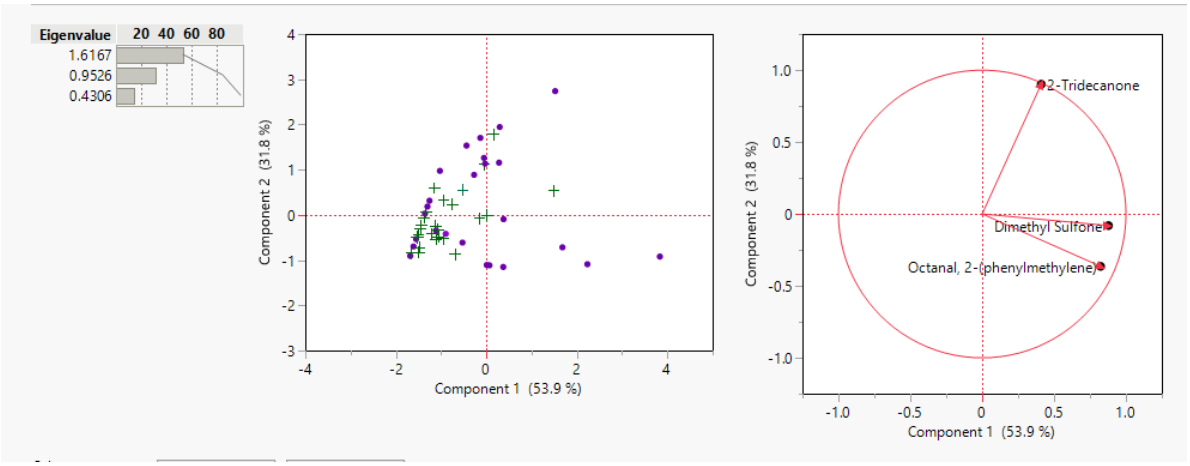


Figure 27: PCA plot of Caucasian vs. Hispanic subjects using VOCs predicted via logistic regression

The average amount of each VOC predicted via logistic regression was evaluated. Figure 26 shows that of the three compounds, Caucasian subjects exhibited a higher average amount of all of the VOCs (dimethyl sulfone, 2-tridecanone, octanal, and 2-phenylmethylene) than Hispanics. The student's *t*-test that was run on these compounds revealed that the differences that were visually observed are statistically significant. The *p*-values obtained are listed in Table 9.

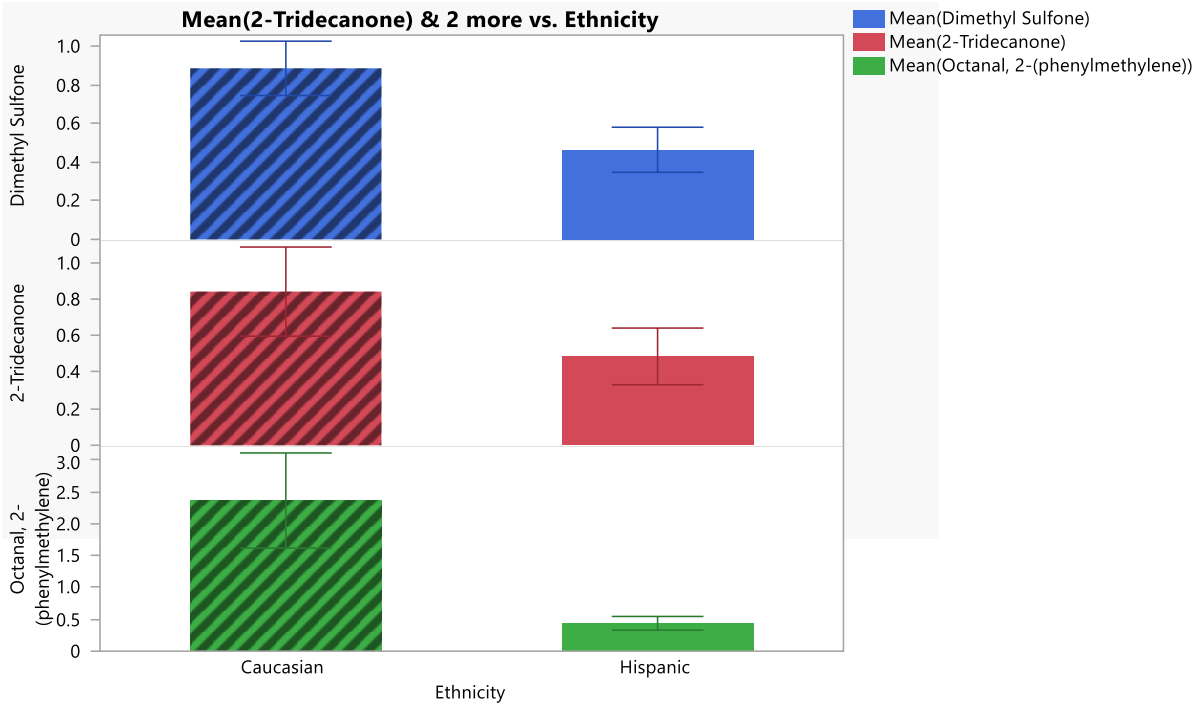


Figure 28: Average quantitated values of VOCS observed within Caucasian and Hispanic subjects

Table 9:Student's t-test p values t 95% confidence interval for VOCs selected for PCA implementation

Compound	*P*-value
Dimethyl sulfone	0.0137
2-tridecanone	0.0076
Octanal 2-phenylmethylene	0.0112

4.4.4.2.2 Caucasian vs. African American Subjects Results

As with the comparison of Caucasian and Hispanic subjects, the VOCs that were extracted from each odor profile were evaluated to determine their usefulness in discriminating between individuals of Caucasian and African American backgrounds. Prior to PCA implementation, a logistic regression analysis was conducted on all 31 VOCs. As in the comparison between Hispanics and Caucasians, only compounds with an AUC of 0.69 were accepted for analysis. The ROC tables reveal that propanoic acid 2-methyl had the highest sensitivity when differentiating between Caucasians and African Americans (81.82%). However, the VOC with the highest specificity (true negative rate) at 62.99% was octanal 2-phenylmethlene.

Figure 29 shows the false positive and false negative rates when using each compound for classification. The false positive rate ranged from 7.00%–35.70% with nonanal exhibiting the lowest false positive rate (7.14%). However, the results of the student's *t*-test reveal that the observed quantitative differences were not statistically significant. The false negative rate ranged from 18%–36.36%, with propanoic acid-2 methyl having the most ideal false negative rate.

Table 10: Performance of individual VOCs for Caucasian vs. African American differentiation

(*) indicates that the compound was statistically significant when the student's t-test was conducted

Compound	*P*-value	Extraction Method Used	Area Under the Curve (AUC)	Sensitivity (True Positive)	Specificity (True Negative Rate)	False Positive Rate	False Negative Rate
Octanal	0.0488	SPME	0.73377	77.27%	55.84%	21.43%	22.73%

Nonanal	0.0044	SPME	0.74838	54.55%	47.40%	7.14%	45.45%
Tetradecane*	0.0028	SPME	0.76786	72.27%	48.70%	28.57%	22.73%
Butanoic acid*	0.0252	LE	0.79058	77.27%	62.99%	14.29%	22.73%
Eicosane	0.0232	SPME	0.71753	77.27%	38.30%	28.57%	22.73%
Hexadecanoic acid, methyl ester	0.0498	LE	0.74351	63.64%	49.35%	14.29%	36.36%
Propanoic acid-2 methyl*	0.0010	LE	0.7629	81.82%	46.10%	35.71%	18.18%

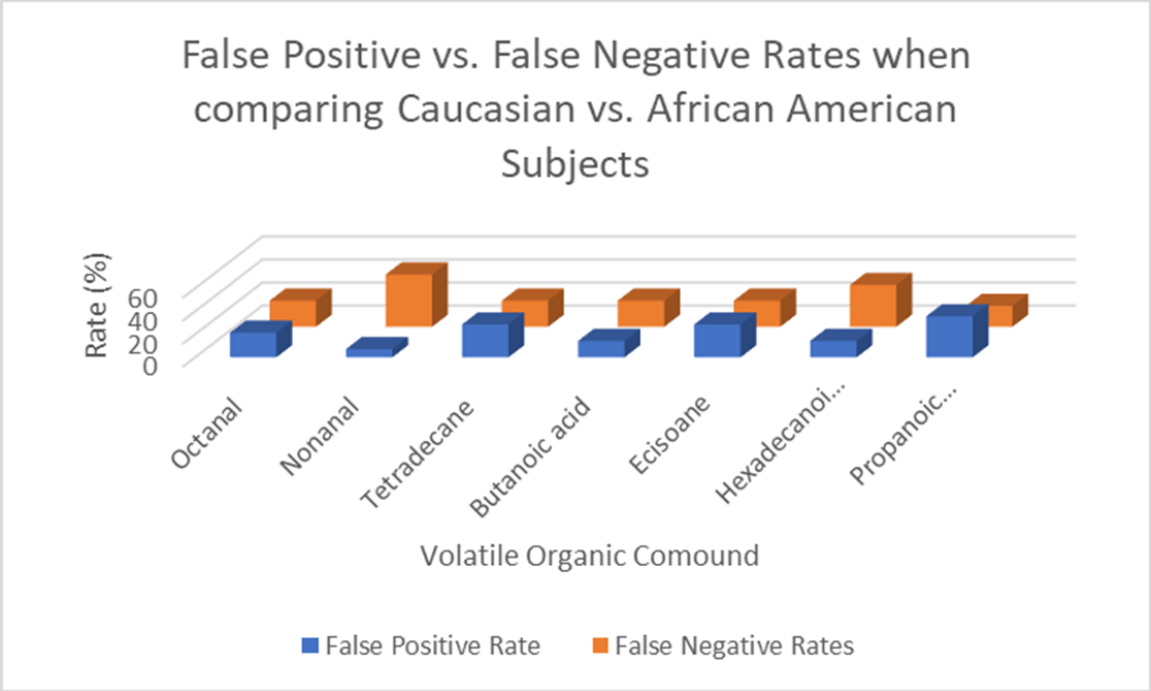


Figure 29:False positive and false negative rates for statistically significant compounds predicted via logistic regression

The Student's *t*-test was conducted at a 95% confidence interval to determine if the differences between the two populations were statistically significant. The Student's *t*-test revealed that three of the seven compounds that passed the AUC cutoff exhibited

statistically significant differences when comparing ethnicities (butanoic acid, propanoic acid, 2-methyl,and hexanoic acid, methyl ester). The *p*-values observed for these compounds are depicted below. These three compounds were selected as the most reliable predictors for the comparison of Caucasian and African American subjects.

Table 11: Student's t-test p-values at the 95% confidence interval for Caucasian vs. Hispanic predictors

Compound	*P*-value
Butanoic acid	0.0143
Propanoic acid, 2-methyl	0.0233
Hexadecanoic acid, methyl ester	0.0374

A PCA plot of the original 31 compounds was constructed to evaluate the ability of those compounds to discriminate between Caucasian and African American subjects. As with the original plot that compares Caucasian and Hispanic individuals, this initial PCA plot reveals that using all 31 compounds to distinguish between individuals of different backgrounds led to poor discrimination. When evaluating all 31 compounds, the extremely low PC1 (17.8%) and PC2 (10.7%), as well as the overlap between Caucasian and African American subjects, highlight the need for data reduction. In assessing these two populations, the logistic regression analysis facilitated the reduction of variables of interest. In only using the three compounds predicted via logistic regression in the PCA plot, PC1 increased from 17.8% to 51.3% while PC2 increased from 10.7% to 30.7%. This indicates that there was a significant increase in separation among Caucasian and African American subjects. When using these compounds to distinguish between individuals from

the two ethnic groups, African American subjects clustered very tightly, with only three Caucasian subjects overlapping with African American profiles. Due to the complexity of human scent samples, a perfect cluster was not expected. The results show that comparing the two PCAs reflected the validity of the logistic regression analysis.

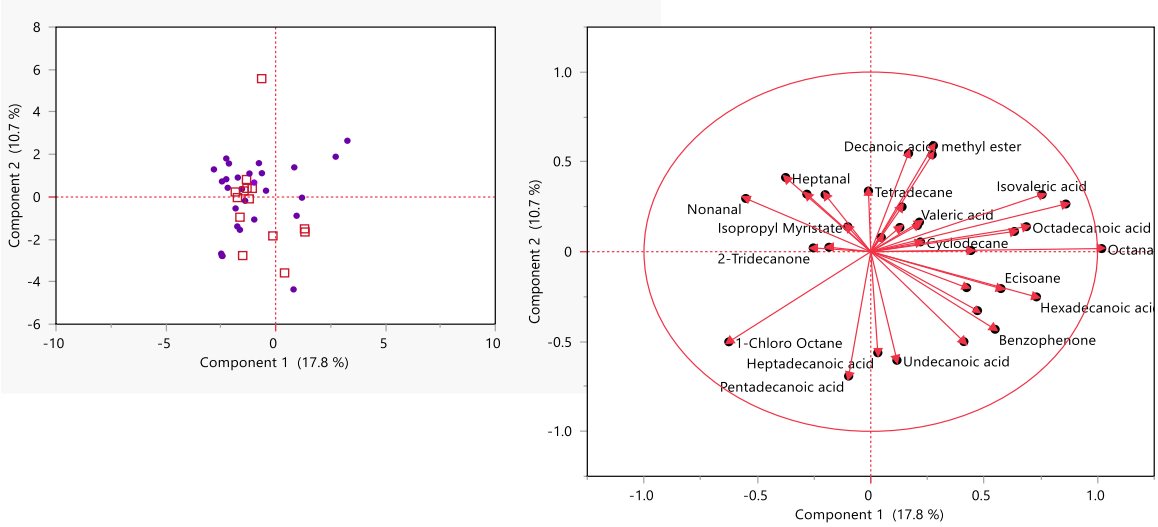


Figure 30: PCA plot of Caucasian vs. African American subjects using all VOCs

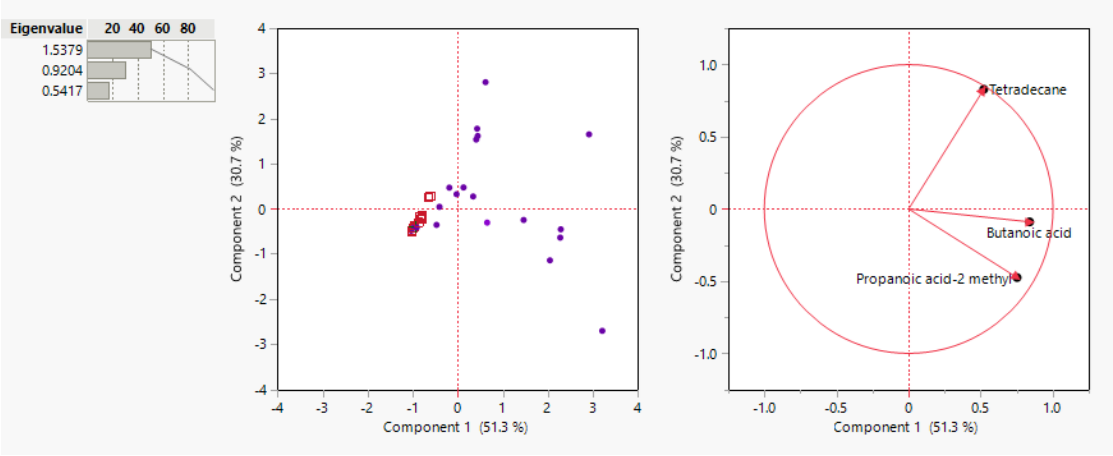


Figure 31: PCA plot of Caucasians vs. African American subjects using only the three VOCs predicted via logistic regression

The average amount of each VOC predicted via the logistic regression was evaluated. Of the three compounds, the samples from Caucasian subjects exhibited higher average amounts of two of the three VOCs (butanoic acid and propanoic acid, 2-methyl). In keeping with the literature, the results show that branched acids like propanoic acid, 2-methyl are more prevalent among those of Caucasian descent.[22] The samples from African American subjects showed higher average amounts of hexanoic acid, methyl ester, which supports evidence from previous studies that African Americans produce greater amounts of straight chained acids than other ethnic groups.[22] In contrast to previous findings, the straight chained acid, butanoic acid, was unexpectedly more prevalent among Caucasian subjects. However, there was only a slightly higher amount of straight chained acids present among African American subjects in previous literature than was observed in this study. The observed increase in butanoic acid among Caucasian subjects may be due to dietary factors. Butanoic acid is more prevalent among obese subjects[58]; if more overweight Caucasian subjects were sampled than overweight African American subjects, the data would be skewed when assessing this compound.

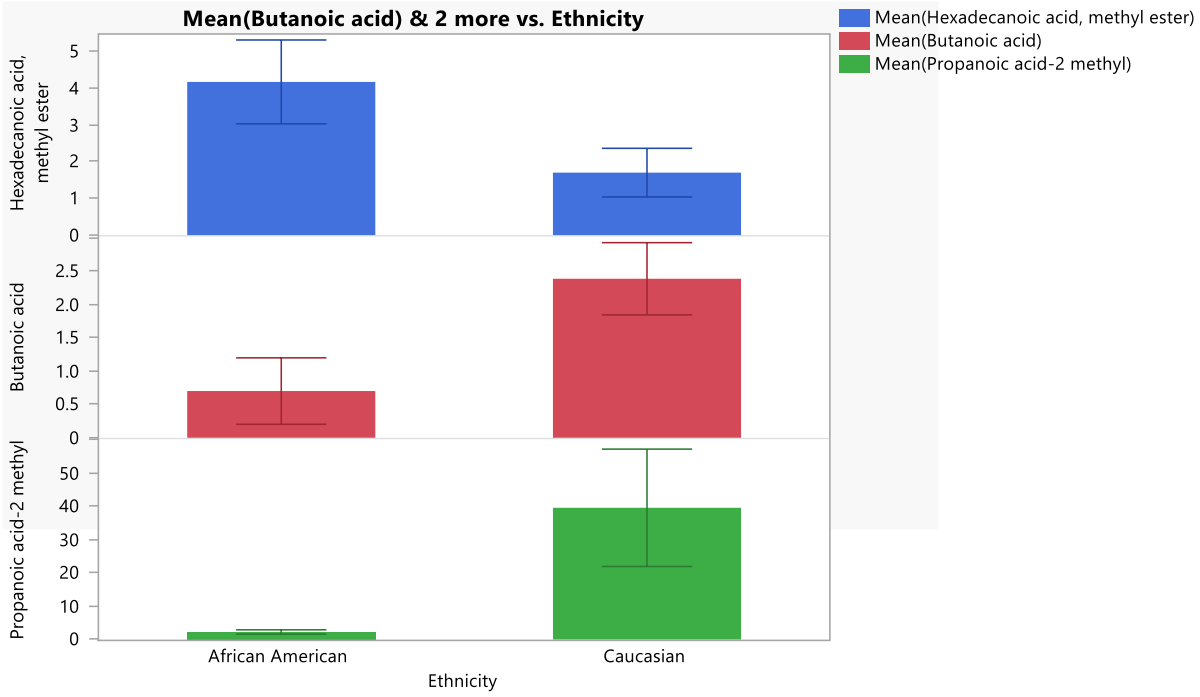


Figure 32:Average quantitated values of VOCs observed within Caucasian and African American subjects

4.4.4.2.4 Hispanic vs. African American Subjects Results

The same statistical method that was utilized to compare previous populations was also used to determine the usefulness of VOCs for classifying individuals according to their ethnicity. For this comparison, the two subject populations under observation were Hispanics and African Americans. Prior to PCA implementation, a logistic regression analysis on all 31 VOCs was conducted. As in the previous study, only compounds with an AUC of 0.69 were accepted for further analysis. The findings are presented in the ROC below. Of the 11 compounds, two compounds (valeric acid and benzophenone) were immediately excluded as their AUCs fell below the threshold (0.60357 and 0.55179, respectively).

Table 12: Performance of individual VOCs for Hispanic vs. African American differentiation

Compound	*P*-value	Extraction Method Used	Area Under the Curve (AUC)	Sensitivity (True Positive Rate)	Specificity	False Positive Rate	False Negative Rate
Nonanal	0.0027	LE	0.76429	55.00%	55.00%	0%	45.00%
Tetradecane*	0.0117	SPME	0.85893	85.00%	63.57%	21.00%	15.00%
Valeric acid*	0.0161	LE	0.60357	60.00%	10.00%	50.00%	40.00%
1-Octadecane*	0.0020	SPME	0.84286	100%	71.43%	29.00%	0%
Ecisoane*	0.0171	SPME	0.76429	75.00%	60.71%	14.00%	25.00%
Nonanoic acid	0.0027	SPME	0.88571	75.00%	67.86%	7.00%	25.00%
n-hexyl salicylate*	0.0395	SPME	0.83929	70.00%	48.57%	21%	30%
Hexadecanoic acid, methyl ester*	0.0006	LE	0.83036	85.00%	63.57%	21.00%	15.00%
Undecanoic acid*	0.0027	LE	0.78036	75.00%	53.57%	21.00%	28.57%
Benzophenone*	0.0154	SPME	0.55179	65.00%	15.00%	50.00%	3.00%
Dodecanoic acid*	0.0167	LE	0.80714	80.00%	72.86%	7.00%	20.00%

The ROC tables reveal that 1-octadecane had the highest sensitivity when differentiating between Hispanic and African American subjects (81.82%). However, the VOC with the highest specificity (true negative rate) was dodecanoic acid at 72.86%. Figure 33 shows the false positive and false negative rates when using each compound for classification. This chart shows that the compounds with AUCs below 0.69 have a

significantly higher false positive rate. The false positive rate ranged from 7% to 45% with with nonanoic acid portraying the most ideal false positive rate. The false negative rate ranged from 0% (which is most ideal) to 45%, with octadecane having a perfect false negative rate.

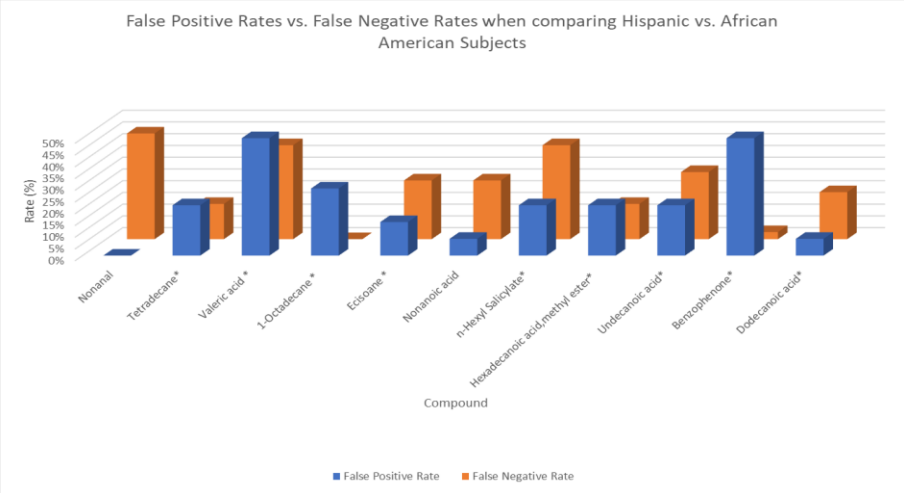


Figure 33:False positive and false negative rates of statistically significant VOCs predicted via logistic regression

Of the 11 compounds predicted via logistic regression for the differentiation of Hispanic and African American individuals, six compounds used the SPME extraction method, and five compounds used solvent extraction. Because there were several compounds with an AUC over 0.69, only compounds with an AUC of 0.80 or higher were selected for PCA analysis (tetradecane, t; 1-octadecane; nonanoic acid; n-hexyl salicylate; hexadecenoic acid, methyl ester; benzophenone; and dodecanoic acid).

A Student's *t*-test was conducted at a 95% confidence interval to determine if the differences observed between the two populations were statistically significant. The student's *t*-test revealed that of the six compounds of interest, five of the six that passed

the AUC cutoff (0.80) showed statistically significant differences when comparing Hispanic and African American subjects (tetradecane; 1-octadecane; n- hexyl salicylate, hexadecenoic acid, methyl ester, benzophenone, and dodecanoic acid). The *p*-values observed for these compounds are depicted in Table 13.

Table 13: Student's t-test p-values at the 95% confidence interval for African Americans vs. Hispanic predictors

Compound	*P*-value
Tetradecane	0.0387
Dodecanoic acid	0.0235
1-Octadecane	0.0039

These three compounds were selected as the most reliable predictors for the comparison of Hispanic and African American subjects. According to the scree plot in Figure 25, a total of three compounds account for over 80% of the variation among ethnicities. When comparing dodecanoic acid and hexadecenoic acid, methyl ester, dodecanoic acid had slightly lower sensitivity, with 10% higher sensitivity. Although the AUC values were relatively similar, hexanoic acid, methyl ester had higher false positive and false negative rates. Therefore, tetradecane, dodecanoic acid, and 1-octadecane were selected as the primary compounds for Hispanic and African American subject comparison.

A PCA plot of the original 31 compounds was constructed to assess the original human scent VOCs that were selected for analysis, based on their capability to distinguish between Hispanic and African American subjects. As with the previously examined

groupings (Caucasian and Hispanic, Caucasian and African American) and the original plots, this initial PCA plot revealed that using all 31 compounds led to poor discriminatory capabilities. The PCA plot yielded low PC1 (19.2%) and PC2 (12.7%) values; this caused significant overlap between the two populations, and no clustering was observed. The overlap confirmed that using all 33 compounds did not facilitate any class characteristic determination mechanisms. Furthermore, this result further validates that when assessing human scent profiles, using 31 compounds to perform the analysis incorporates too many variables.

When the number of compounds for the PCA was reduced to include only those compounds selected from the logistic regression analysis, the variation percentage increased from 19.2% to 49%, and PC2 increased from 12.7% to 28.7%. The increase in PC1 and PC2 provided a clear explanation of the successful separation of Hispanic and African American subjects with little overlap. Once again, the variation percentages of both principal components showed a substantial increase when data reduction was implemented.

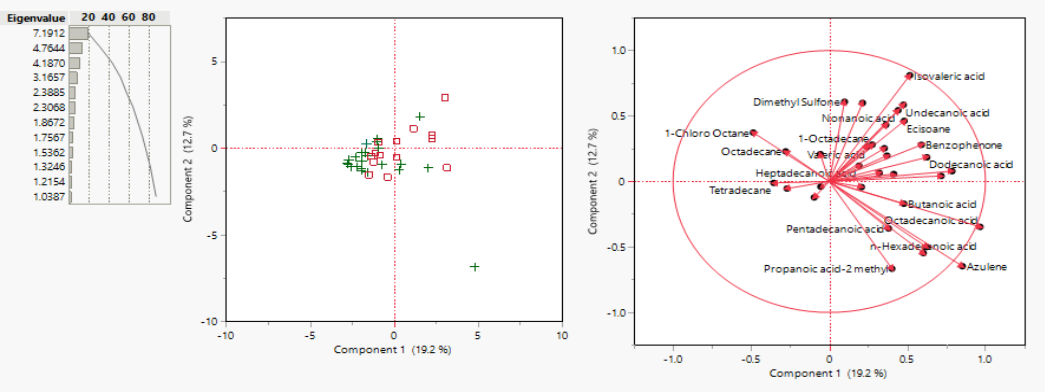


Figure 34:PCA of Hispanics vs. African American subjects using all VOCs

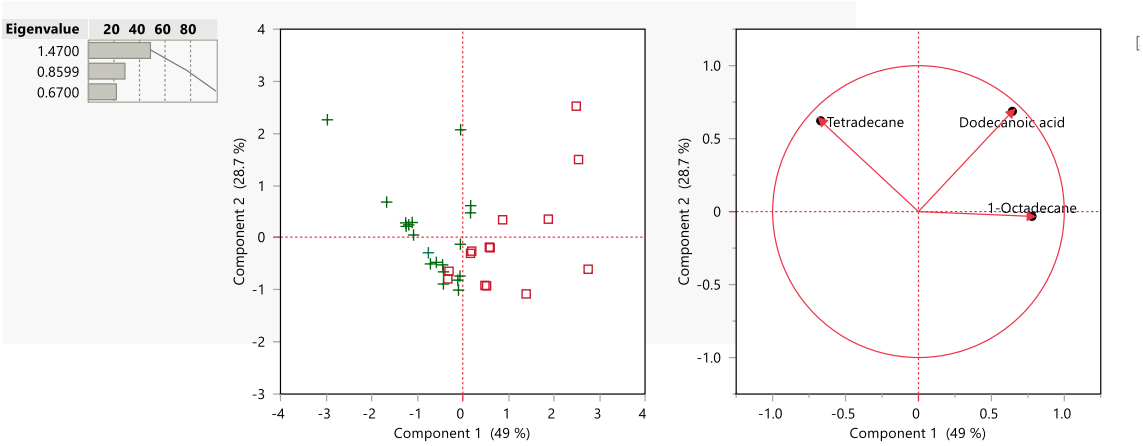


Figure 35:PCA of Hispanic vs. African American subjects using only VOCs predicted via logistic regression

The comparison of the average amount of statistically significant compounds that were predicted using logistic regression was assessed. The proposed model suggests that samples from African American subjects had a higher average amount for two of the three VOCs (1-octadecane, dodecanoic acid). The samples from the Hispanic participants had a higher average concentration of tetradecane. The graph shows that there is a statistically significant difference between the amounts of these substances, which directly corresponds to the student's *t*-test previously performed in this study.

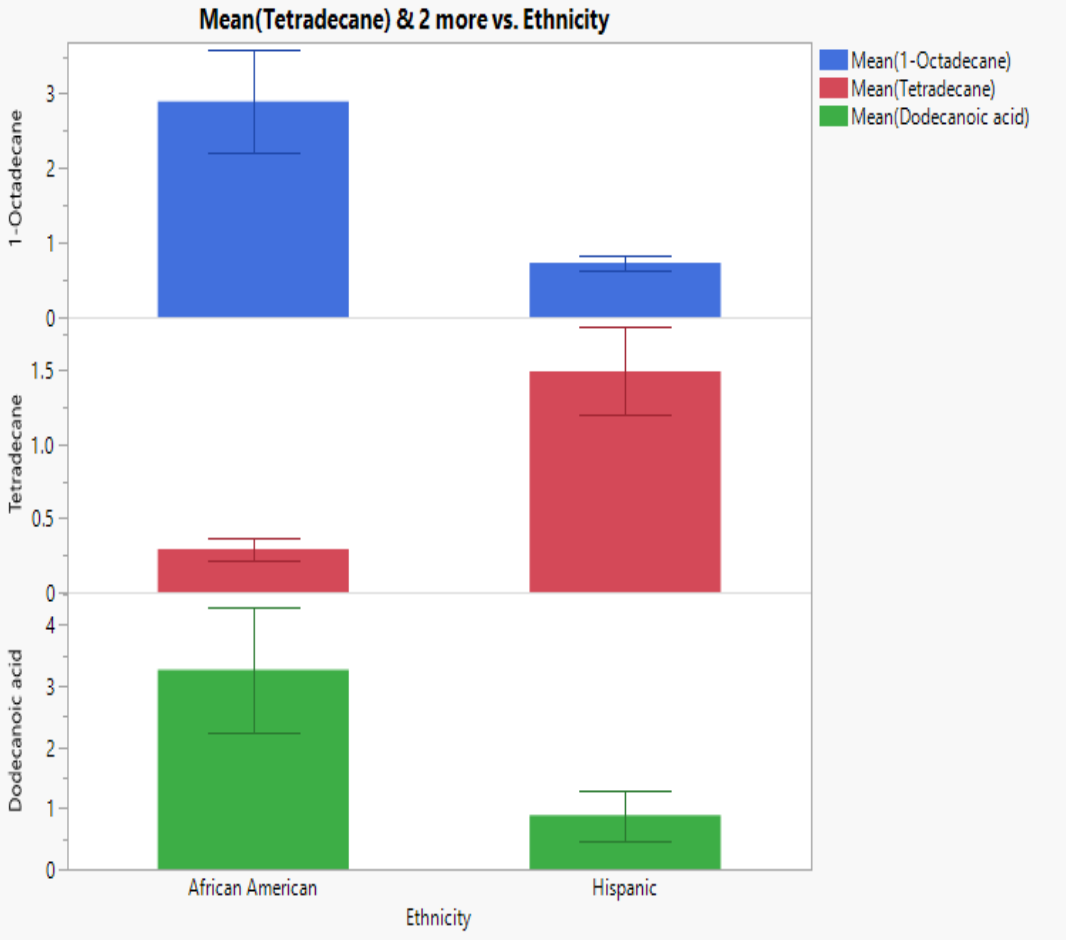


Figure 36:Average quantitated values of VOCs observed within Hispanic and African American subjects

5.HUMAN SCENT STUDY CONCLUSIONS

In this study, several underarm odor samples were collected from participants and analyzed using SPME-GC/MS. Thirty-three VOCs, previously identified as odor contributors, were selected for qualitative and quantitative analysis. The results of the analyses demonstrate the discriminating power of VOCs within human odor analysis.

The comparison of individuals from two ethnic groups facilitated the data analysis of these complex human scent profiles. The study provided insights into how VOCs can be used for class characteristic determination. Previous studies have found that a combination of VOCs is responsible for the differences between ethnic groups.[27,37] The findings in this study further validate this claim. Moreover, SPME-GC-MS has consistently demonstrated its reliability when extracting human scent profiles.

By employing qualitative and quantitative methods of inquiry (relative peak area ratio comparisons, logistic regression, and principal component analysis), it is possible to use VOCs to discriminate between different ethnicities. Of the 31 compounds selected for analysis, 23 of the 31 compounds were predicted to be of importance when attempting to classify between ethnicities. Of these 23 compounds, 10 compounds were deemed to be statistically significant when assessing ethnicity-specific differentiation. The ROC table extracted from the logistic plot allowed for observation of the sensitivities and specificities of each compound. The trends depicted within the ROC tables show that compounds with higher sensitivity had a reduced false negative result rate. Conversely, the compounds with high specificity had a lower false positive rate.

This work represents the first study that has compared the ability of VOCs to differentiate between individuals of different ethnicities using two extraction methods

simultaneously. The novel use of multiple extraction methods (SPME and solvent extraction) allowed for the acquisition of data that are pivotal to the findings. Five of the 10 compounds used for ethnicity-specific differentiation were extracted via solvent extraction, and the remaining five compounds were extracted via SPME. This highlights the advantage of coupling HS-SPME to solvent extraction for maximum recovery of VOCs and enhanced discrimination power.

Furthermore, this project highlights the benefit of using multivariate analysis for data reduction. Coupling logistic regression to PCAs simplified the identification of significant variables. In coherence with the scree plot constructed for the dataset, when three compounds were utilized for ethnicity-specific differentiation, a variation percentage of 80% or greater occurred for each comparison. It is important to note that when more than three compounds were plotted, there was a decrease in observed PC1 and PC2.

Three novel compounds have been identified as the primary contributors of differentiation between Hispanic and Caucasian individuals (dimethyl sulfone; 2-tridecanone; and octanal, 2-phenylmethylene). The high percentage of variation contribution from PC1 and PC2 (85.7%) is promising and suggests that these compounds can be used for class characteristic determination in forensic casework. However, in future work, researchers should ensure that participants clearly state their origin rather than their ethnicity to avoid potential self-identification bias. Furthermore, a cross-validation technique to confirm origin (like an ancestry test) could provide genetically verifiable origin data for each subject.

This work has identified three novel compounds as the primary contributors for VOC differentiation between Caucasian and African American individuals (butanoic acid;

propanoic acid, 2-methyl; and hexanoic acid, methyl ester). The results obtained using the Student's *t*-test are in alignment with findings from Prokop-Prigge et al.'s study; they suggest that African Americans produce higher amounts of straight chained acids than other ethnicities. This study also shows how diet and weight can affect the data that are obtained when analyzing human scent profiles.

Moreover, three novel compounds have been identified as the primary contributors for VOC differentiation between Hispanic and African American individuals (tetradecane, dodecanoic acid, and 1-octadecane). This study is the only one to have portrayed complete separation with little overlap. The high AUC values presented in the ROC table (above 0.80) show the efficiency of these compounds and are directly associated with the tight clustering visualized in the PCA plot.

Although a slight overlap was present within two of the three ethnicity groupings, there is a significant increase in PC1 and PC2 in all three groups when the variables are narrowed. There was an average 33.07% increase in PC1 when using logistic regression and an average 19.06% increase in PC2 when comparing ethnicities. The increase in these principal components is directly correlated to the logistic plot that predicts that these variables act as the key contributors to ethnicity-specific differences. The compounds that portrayed the acceptable AUCs were responsible for the majority of the variation observed among ethnicities.

While this study did not confirm that complete ethnicity-specific determination can occur, it did demonstrate the following:

a) Logistic regression was effective for predicting the compounds that contribute to the differences between two ethnicities. The ROC table successfully highlights the

performance of each logistic plot. Furthermore, the sensitivity and specificity of the compounds provide insights into how reliable each compound is for the prediction of ethnicity.

b) The implementation of the Student's *t*-test assisted with data reduction by identifying the statistically significant compounds and using them for the PCA analysis.

c) Overloading PCA plots resulted in poor data, but they can show significant improvement when the appropriate statistical methods are used.

Future work for this project involves the collection of additional odor profiles from subjects of different ethnicities. The increase in subjects analyzed will facilitate the completion of a human scent database of VOCs that can be used effectively for identification of persons regardless of race or ethnicity. Furthermore, to aim for a higher value of both sensitivity and specificity, more samples should be collected for each respective ethnicity.

Although a larger sample set is necessary to definitively identify which VOCs can be used to distinguish between subjects of different ethnicities, the discovered and confirmed VOC biomarkers for ethnicity-specific differentiation have the potential to assist with class characteristic determinations of human subjects in forensic casework. Much uncertainty still exists regarding the relationship between the VOCs categorized for human scent analysis and HLA genes. The following chapter of this dissertation aims to determine whether HLA genes can be used to differentiate between ethnicities.

6. HLA GENE INTRODUCTION

6.1 INTRODUCTION

The proper collection and analysis of biological evidence has a significant impact on the outcome of police investigations and rightful prosecutions. The analysis of blood, bodily fluids, and tissues for DNA identification has been pivotal in forensic casework. Incriminating DNA evidence has proven to effect jury conviction decisions greatly due to its high discrimination power.

The exploration of the Human Leukocyte Antigen (HLA) was included as a task in this dissertation. DNA typing has proven to be extremely useful for validating genetic markers of individuals in forensic casework. However, when assessing the correlation of a population, HLA markers are more useful due their responsibility in kin recognition and genetic similarity among unrelated individuals.[59]

HLA genes have been used as criteria for genetic markers in paternity testing and disease detection.[42] Furthermore, forensic medicine has utilized the HLA genes to distinguish between subcultures using HLA markers.[60] The observation of allelic sharing and variation among allelic frequencies can facilitate the class characteristic determination of subcultures.

To date, the evaluation of healthy subjects for biomarkers in correspondence to ethnicity has yet to be undertaken. The purpose of this study was to determine if the information provided by the HLA genes can serve as a biomarker when distinguishing between three populations (Caucasian, Hispanic, and African American). The DNA profiles are extracted using Maxwell 16® Extraction. The compounds extracted are then amplified via Multiplex-Polymerase Chain Reaction (Multiplex PCR). Once the specific regions of

interest are amplified, the regions are then analyzed using Capillary Electrophoresis. The allelic frequencies observed show the discrimination power of HLA and its influence on genotype relationships among populations.

6.2 DEOXYRIBOSE NUCLEIC ACID (DNA)

Located inside the nucleus of each cell of living organisms, Deoxyribonucleic acid (DNA) provides genetic information unique to everyone. The basic unit of DNA (nucleotide) is composed of a deoxyribose sugar unit, nitrogen-containing base and a phosphate group.[61] Figure 37 shows how sugar phosphate groups connect to the nitrogen-containing base to form a nucleotide.[62]

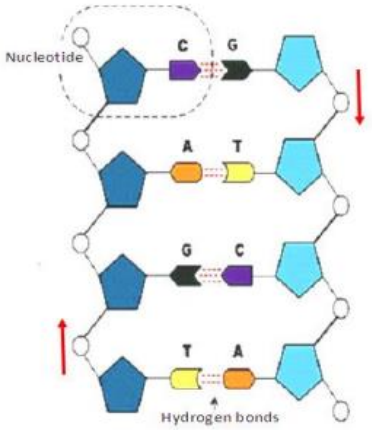


Figure 37:Primary structure of DNA[62]

In 1953, James Watson and Francis Crick discovered the structure of DNA that has been generally accepted as the golden standard for forensic genetic analysis.[63] DNA has a double helix structure with alternating sugar-phosphate backbones present on the outside helix. Four nucleotide bases, Adenine (A), Cytosine (C), Guanine (G), and Thymine (T)

are on opposite strands. These bases pair with their associate via hydrogen bonding. When adenine forms, it is paired with thymine; the same is true for guanine and cystine (C).[61]

In a human, genes are present that are simply segments of a DNA molecule. Their primary responsibility is to describe the genetic makeup and functionality of the human body. A human has 50–100,000 genes within their body.[64] There are no limitations in the sequences in which these bases can occur on a single chain. In humans, 3 billion nucleotides are present within 24 chromosomes,[65] and these components determine what we are composed of and the operation of our cells. Because variation among human genomes is present at approximately 3 million sites, DNA can be used to exclusively identify an individual (with the exception of identical twins).[66] In forensic casework, DNA was first implemented in 1985 in the United States. By 1998, DNA had been accepted as the gold standard for individual identification due to its reliability and validity. DNA can be found in several biological specimens including blood, hair, semen, urine, and saliva.[66] The least invasive biological sample (saliva) was selected for this study. Saliva holds 1,000-10,00 ng/mL of DNA.[66] With the appropriate sample collection methods, saliva can be used to provide useful information for forensic casework.

6.3 HUMAN LEUKOCYTE ANTIGEN GENE COMPLEX

The Human Leukocyte Antigen (also known as the Major Histocompatibility Complex in vertebrates) was discovered by Jean Dausset in 1958.[67] This complex region determines the cells that belong in the body and fights off viruses and bacteria to protect the immune system.[68] The HLA Gene Complex is located on chromosome 6 spanning approximately 3.6 mega base pairs.[60] The HLA Gene Complex is composed of over 100 genes that account for about 50% of known protein coding genes to date.[69] HLA genes are classified into three regions: class I, class II, and class III.[29] As seen in Figure 38, HLA class I consist of HLA-A, HLA-B, and HLA-C. Class II genes consist of HLA-DP, HLA-DQ, and HLA-DR.

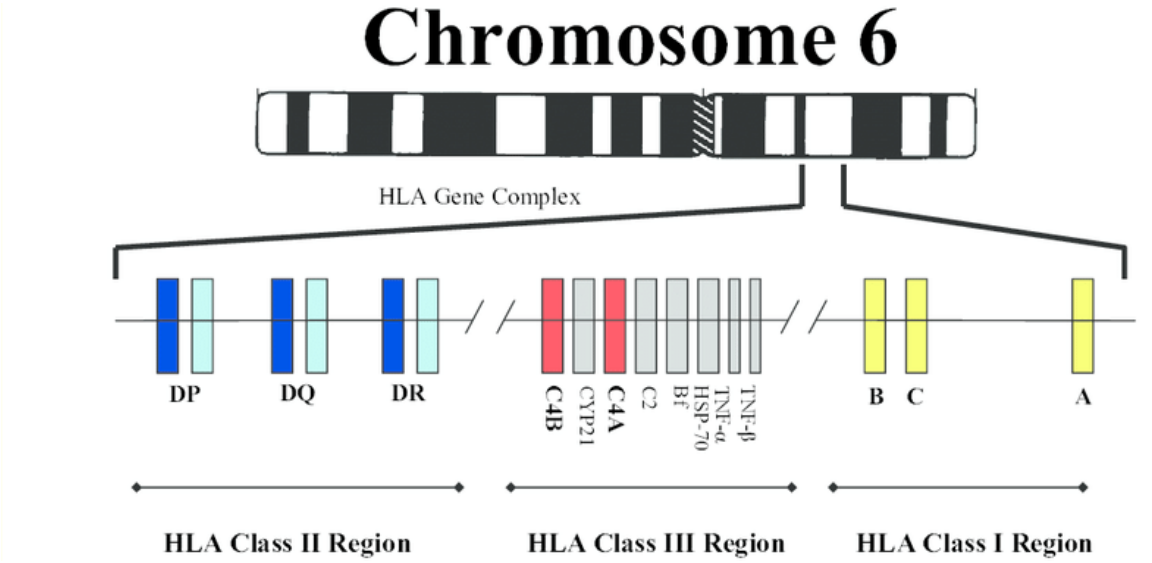


Figure 38:Diagram of HLA region on chromosome 6 [60]

HLA I genes have been used as a biomarker for diseases such as Alzheimer's and lung cancer.[70,68] However, Class I HLA alleles are less dependent on ethnicity than class II.[71]Therefore, the HLA class II region is of interest for this study.

Within the HLA region, prior studies have noted several ancestry informative alleles that are useful when exploring ancestry and evolution. The most common five loci explored are HLA-A, HLA-B, HLA-C, HLA- DRB1, and HLA-DPB1. There are hundreds of alleles within these regions that encode proteins capable of identifying individuals.[60] HLA DQ A1 system has been labeled one of the most informative in forensic analysis. Previous studies have indicated that loci from this region can identify which alleles are contributors to Japanese subcultures by using a principal component analysis model[72]. HLA-DRB1 loci have also been proven to identify ethnic and geographical differences within Caucasian and non-Caucasian individuals.[71] HLA DQ A1 alleles have been used in previous literature to discriminate between two Spanish populations.[72]

The most common methods for determining the origin and evolution of individuals is allelic association studies.[73] The frequency distribution of HLA alleles allow for the genetic diversity of human populations to be explored.[72] Allelic association studies observe whether the frequency of an allele significantly increases or decreases when assessing different populations. The application of PCR-based techniques have led to the development of the HLA's ability to detect significant differences in allele frequency among different populations.[74] The significance of genetic ancestry classification was exemplified in a bone-marrow study that revealed that nearly 20% of subjects self-identified their race/ethnicity inaccurately with respect to their genetic lineage.[75]

6.4 BUCCAL SWABBING

The oral cavity has been utilized as a primary source in biologic studies of genetics, metabolic, and microbiomes.[76] Within the oral cavity, there are several structures and tissues. Each one of these structures and tissues is colonized with bacteria and engulfed in salivary fluids.[77] In comparison to other biological specimens (such as tissues and blood), oral samples are preferred due to their non-invasive nature and low cost for collection. The two most common oral samples are saliva samples and buccal swabs. A buccal swab is a scientific term that refers to a sample collected from the cells inside a subject's cheek. The cells within the cheek contain DNA information in the form of squamous epithelial cells.[76] There are different cell types classified within a cheek's morphology. Table 14 summarizes the percentage of cell types present within cheek samples. The figure clearly depicts the most abundant cell type (ISC), which is the pink non-keratinous superficial squamous cells

Table 14:Proportion of epithelial cell types present within adult cheek samples[76]

Cell Type	% Of Epithelial Cells in Adult Subjects
Pink Non- Keratinous Superficial Squamous Cells	70.5%
Orange Keratinous Superficial Squamous Cells	19.6%
Blue Intermediate Squamous Cells	6.4%

Figure 39 presents a visualization of the cellular morphology acquired when a buccal swab is taken. The three types of buccal swabs can be differentiated by abundance and color via pap staining.

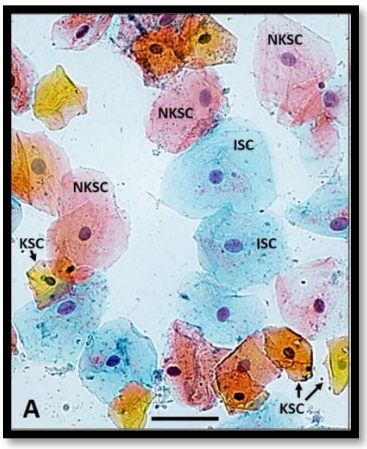


Figure 39:Cellular morphology of cheek sample via pap staining[76]

Cheek cells are often studied due to the simplicity of their acquisition and their ability to produce a high yield of DNA.[78] The cells recovered from sampling the cheek using a cotton swab contain the genetic makeup of the subject's entire body. There are several types of material used for DNA swabs. The most common types of material are cotton, foam, and flock.[79] Cotton swabs are commonly used in DNA sampling. In comparison, the polyester single fibers present within the flocked swabs are designed to produce better cell yield.[79] However, according to Ruiz et al., standard cotton swabs are more cost efficient and yielded a similar amount of DNA in comparison to flocked medical grade swabs.[79] In the forensic field, the buccal swab is the most efficient and common way for

forensic pathologists to determine if there is a link between a suspect and a crime scene. Buccal swabs have also been proven to be an effective sample collection method for the investigation of the most common HLA loci.[80] Sample acquisition is the first step in the forensic workflow of DNA analysis. DNA extraction follows with the sole purpose of isolating DNA from the epithelial cells obtained via buccal swabbing.

6.5 DNA EXTRACTION

The limited amount of biological evidence recovered at crime scenes highlights the significance of recovering DNA from samples. The automation of extracting DNA in forensic casework is essential to the reduction of handling errors and the increase in DNA yield in comparison to manual extraction methods.[81]

The Maxwell® 16 is an automated DNA purification system that extracts 16 or fewer samples simultaneously.[82] The Maxwell® 16 instrument is coupled to the DNA IQ Reference Sample Kit™ for the extraction of DNA in forensic casework. There are several advantages to the Maxwell® 16 instrument including its pre-programmed methods for purification and rapid purification, its small 12 x 15 footprint, and lower instrument and kit costs.[81,83] Additionally, the movements of resin through the cartridges decrease the likelihood of clogged tips and partial liquid transfers.

The extraction method is ideal for the evaluation of blood, semen, tissue, touch, and trace samples.[84] There are two types of extraction kits available on the market: Standard Elution Volume (SEV) and Low Elution Volume (LEV). Low Elution Volume (LEV) is typically preferred in forensic casework due to smaller sample processing and its ability to produce higher DNA yields and concentrations. The DNA IQ Reference Sample Kit™ is equipped with cartridges that are prefilled with chemicals. Figure 40

displays a diagram of the Promega Maxwell 16 DNA IQ Reference Sample Kit.[83] The "contents" on the left side of the figure identify the makeup of a Promega Maxwell 16 DNA IQ Reference Sample Kit cartridge.[83] The sample is placed in #1A. The magnetic rods present within the Maxwell® 16 instrument transports the DNA resin through the purification reagents present in the cartridge.

When utilizing the Maxwell, cell lysis occurs in which DNA binds to magnetic particles in the presence of guanidinium salt.[84] This salt denatures membrane proteins to lyse cells and facilitate the binding of the DNA. Subsequently, a series of washes remove undesirable compounds, and the target DNA is eluted into a buffer.[81] The DNA is released from the resin by heating at 60° C, and the purification of up to 16 samples transpires within 30 minutes. The resulting sample extracts are then prepared for quantitation and amplification.[82]

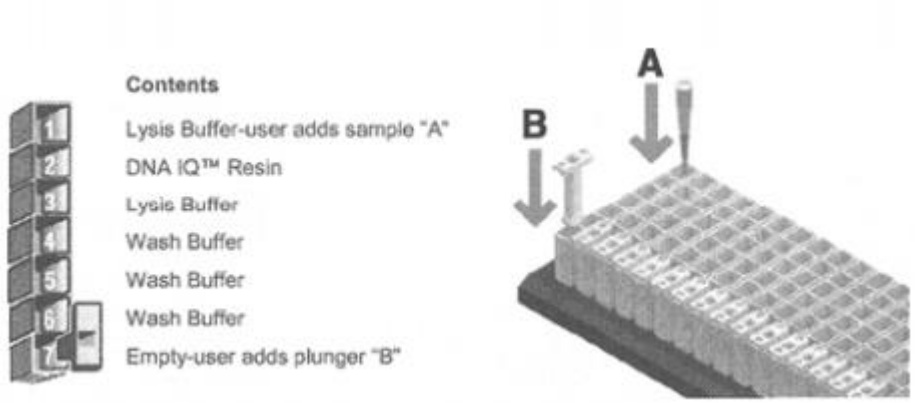


Figure 40: Diagram of DNA IQ Reference Sample Kit[83]

6.6 QUANTIFICATION OF DNA

In order to optimize a PCR reaction, DNA must be quantified so that adjustments to template DNA can be implemented prior to STR analysis.[85] There are three common methods used to quantify DNA: UV spectroscopy, fluorometry, and qPCR. Of the three methods, qPCR is the most ideal method for quantifying DNA. However, qPCR is more expensive and time consuming than UV spectroscopy and fluorometry.[86] UV spectroscopy is incapable of differentiating between DNA and RNA.[87] Fluorometry was selected as the method for this study. Specifically, the Qubit 2.0 Fluorometer was used.

Figure 41:Qubit Fluorometer for quantification of gDNA[88]

Fluorometry is a common technique to quantify DNA with minimal interference from contaminants. Fluorometric measurement of DNA are advantageous due to its ability to quantitate DNA using as little sample as possible with high sensitivity.[89] Fluorometry uses fluorescent dyes to determine the concentration of nucleic acids and protein in a

sample.[87] The Qubit fluorometer is capable of quantitating molecules (DNA, RNA, or protein) by using low fluorescent dyes that bind to their target molecule and illuminate brightly. Fluorescence-based quantification is 1,000 times more sensitive than quantification via UV spectroscopy.[90] Using the Qubit dsDNA kit, DNA is quantitated with high selectivity. The Qubit dsDNA kit is equipped with a working solution and pre-diluted DNA standards. The genomic DNA (gDNA) is mixed with a working solution composed of a dilution buffer and fluorophore to bind the DNA. The Qubit determines the concentration of the gDNA based on the relationship between the two prediluted standards given. Once the quantitation of the DNA is evaluated, a PCR method that allows for sufficient amplification of the loci of interest can be optimized.[91]

6.7 POLYMERASE CHAIN REACTION

Polymerase chain reaction (PCR) is a common DNA amplification technique that produces millions of amplicons of a target sequence quickly.[92] This sensitive technique is capable of converting trace amounts of DNA into several copies for analysis, which makes it an ideal technique for biomedical research and criminal forensic investigation. Figure 42 shows the exponential increase of gene copies that occur within each cycle of polymerase chain reaction.[93] PCR begins with a single molecule of genetic material that can come from biological specimens such as hair, saliva, microbes, or dried blood.[93] To conduct a PCR reaction, template DNA, primers, nucleotides, and DNA polymerase are necessary.[94] The nucleotides present within the reaction (adenine, thymine, cytosine, and guanine) are found in DNA. Primers are responsible for specifying which DNA product

is being amplified. In PCR, the primers act as a complementary sequence of DNA to the target sequence detected and amplified.[94]

To allow primers to bind to each targeted DNA duplex, the strands are heated and then cooled. The reaction is prepared in small reaction tubes (0.2–0.5 mL volumes) and is heated using a thermocycler. A thermocycler successfully programs the temperature needed to successfully amplify the template DNA through each step of the PCR process. These steps have been classified as denaturing, annealing, and extension. The temperatures used and the time applied per cycle are optimized based on the primer conditions and DNA polymerase.[93] During the denaturation step, the reaction mixture is heated from 94–96°. Denaturation serves to disrupt the hydrogen bonds between bases to separate double-stranded DNA to a single strand.[95] During the annealing step, the reaction temperature is lowered between 50–65° C.[95] The primer anneals to the complementary bases and should be 3–5° C below the temperature of the primers used. In the final step (extension), the temperature depends on the type of DNA polymerase utilized. Within each temperature cycle, the DNA copies double as seen in Figure 42; multiple target regions can be copied by implementing Multiplex PCR.

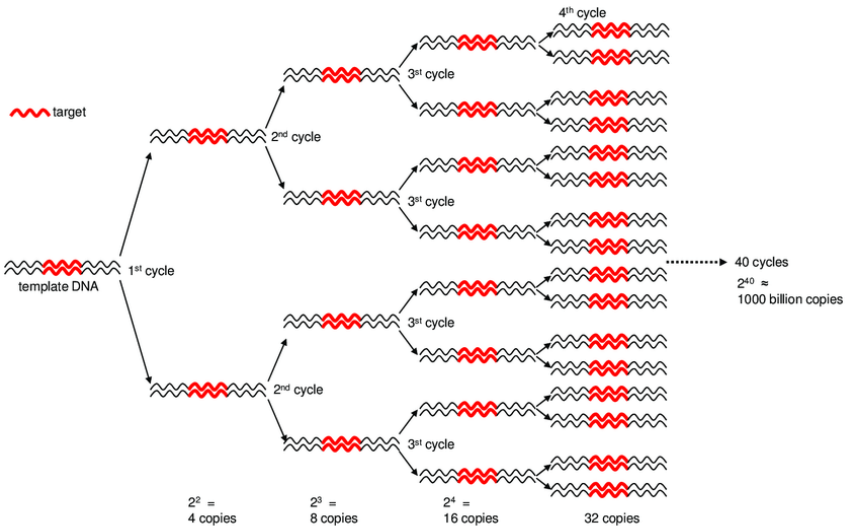


Figure 42:Amplification of gDNA using PCR[96]

6.8 MULTIPLEX POLYMERASE CHAIN REACTION

Multiplex PCR is a technique that uses multiple primers within a single PCR reaction to target and amplify multiple loci.[97] Introduced in 1988 by Chamberlain, Multiplex was introduced in 1988 to improve the efficiency of genotyping.[98] Multiplex PCR has been known for its ability to target more than one sequence within biological specimens. The technique has been used to screen for gene deletions, STR analysis, and SNP analysis.[99] Multiplex PCR has been previously used in genetic disease diagnosis, gender screening, and pathogen identification. There are several advantages to using Multiplex PCR. The quality of the template is more effective in multiplex than single reaction PCR, and less DNA is needed to obtain results. Furthermore, the decrease in expenses and preparation time is less in PCR, which makes the technique ideal for forensic casework.[100] The design and development of a successful Multiplex PCR reaction is contingent upon several

factors. When selecting primers for the reaction, the annealing temperatures must be similar to minimize the risk of cross reactivity. The greatest challenge is optimization of the PCR technique—specifically, the concentration of reagents, annealing temperature, and cycling conditions that are suitable for all primers within the reaction.[97] The effectiveness of Multiplex PCR can be validated by separating products by size via gel electrophoresis.

6.9 GEL ELECTROPHORESIS

Gel electrophoresis is the separation of charged molecules in an electric field. When biomolecules like DNA are placed within this electric field, the charged molecules move in the direction of the electrode of opposite charge.[101] The molecules' mobility varies depending on their net charge, mass/charge ratio, and molecular shape.[102] The gel type assist in the separation of the molecules. The most used gel matrix for the separation of DNA in forensic genetics is the agarose gel.[94] Agarose is a polysaccharide gel matrix, which is composed of galactose and 3,6, anyhdrogalactose.[102]

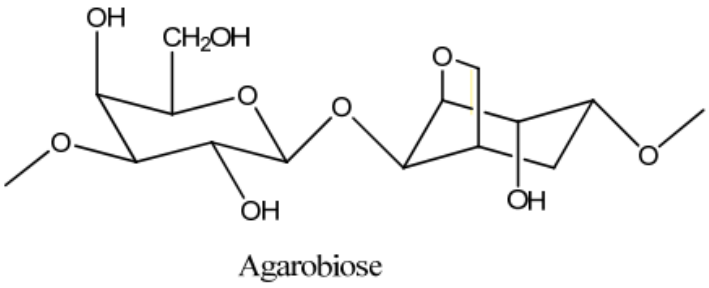


Figure 43: Chemical composition of Agarose gel matrix[102]

By selecting gels with specific concentrations, the resolution of DNA fragments can be manipulated. The higher the agarose concentration, the better separation of small DNA fragment.[103] On the other hand, low agarose concentrations produce higher resolution of large DNA fragments. The PCR products are visualized by staining the product with a chemical dye such as ethidium bromide.[104] Gel electrophoresis is useful for qualitative analysis, specifically to determine if there is presence or absence of a specific DNA product. Furthermore, gel electrophoresis can be used to gain a better understanding of the genotype traits of an individual at a specific locus. The band size and intensity can reveal whether the DNA is heterozygous or homozygous. Gel electrophoresis serves as an effective confirmatory technique to verify the PCR product formed prior to further application. Once it is established that the PCR products are present and are within their respective band size, the fragments are further analyzed using capillary electrophoresis.

6.10 SHORT TANDEM REPEATS

Short tandem Repeats (STRs) are highly polymorphic DNA loci that contain repetitive units used for forensic DNA analysis.[105] The human genome contains one STR for every 200 base pairs.[106] STR profiles have been used as markers for both disease and chromosomal abnormalities due to polymorphic composition.[99,103] The forensic analysis of STR loci have been proven to be a promising method for classifying individuals' biological samples due to the short length of STRS (20–100 bp long). These profiles use low amounts of DNA for amplification which make them ideal for forensic analysis.[107] The application of this phenomenon includes a variety of biological specimens including: hair, blood, and saliva.[108] Results observed by the generated STR profiles can be assessed

based on peak height and allele detection.[109] Human STRs are present approximately once every 2 kb,[110] and the most frequently observed STR is dinucleotide repeats. Previous studies has shown the capability of the Y-STR multiplex system to discriminate between ethnicities using both gene diversity values and allele frequency distribution.[99] STRs have also been used in HLA studies to assess genotypic identification in Caucasian subjects in.[19] The amplified PCR products must be separated in order to distinguish between fragment lengths.[111]

6.11 CAPILLARY ELECTROPHORESIS

Capillary Electrophoresis (CE) is a chemical analysis technique used to separate and identify the ion fragments recovered in PCR by size and charge.[105] CE is known as the primary technique for forensic analysis.[107] In comparison to its predecessor gel electrophoresis, CE increases the charge of the electric field, which results in higher throughput and a faster separation.[111]

During this analytical technique, the extension products of the PCR reaction enter the capillary.

Within a buffer system, the DNA fragments lose an H+ ion from the phosphate group, which facilitates the formation of an anion.[111] Then, the products of the PCR enter the capillary, and a high voltage charge is applied to the sample. This forces the fully automated separation of extension products based on their charge by forcing the negatively charged fragments into capillaries. Prior to reaching the positive electrode, the fluorescently labeled dsDNA fragments move across a laser beam to facilitate the attachment of the dyes to their respective fragment and a CCD camera captures the

fluorescence. The components of a capillary electrophoresis system are depicted below:[112]

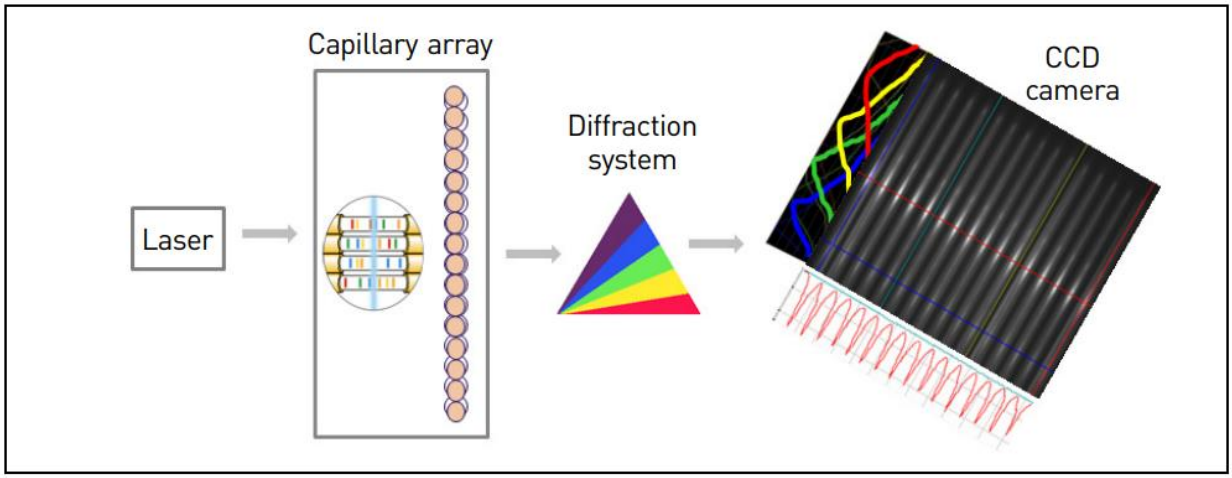


Figure 44:Components of Capillary Electrophoresis system[112]

Each dye used emits light at a variety of emission wavelengths when exposed to the laser. This enables the fluorescently labeled PCR products to distinguish clearly between multi-locus DNA profiles.[100] There are several advantages to using CE for data analysis. The instrument has great sensitivity and resolution.

The SeqStudio Genetic Analyzer (Thermo Fisher) is a four-capillary electrophoresis system equipped with universal polymer and cartridge-based system to conduct the analysis of the samples collected.

7.MATERIALS AND METHODS

7.1 RECRUITMENT OF SUBJECTS

A total of 34 subjects between the ages of 18–30 were recruited by word of mouth for the study. The ethnicities represented in the study were Hispanics, Caucasians, and African Americans.

After all recruitment efforts were complete, a total of 34 subjects were recruited. Each subject was asked both their age and which ethnic group they identify with prior to sample collection. Of the 34 subjects, three samples were used for optimization purposes. Table 15 shows the demographics of the pilot study participants.

Table 15:Demographics of HLA pilot study

Race/Ethnicity	Number of Subjects
Caucasian	8
Hispanic	10
African Americans	13

7.2 MAXWELL EXTRACTION METHODOLOGY

Prior to PCR implementation, buccal swab samples were retrieved from storage for the extraction of genomic DNA using the Maxwell® 16 LEV DNA. Promega developed a sample kit that provides all the components necessary for a successful extraction (lysis buffer, elution buffer, clearing columns, Proteinase K [PK solution], Maxwell® 16 LEV cartridges, elution tubes, and LEV plungers). The protocol for genomic DNA samples was implemented to prepare the buccal swab samples for extraction. The buccal swab

was prepared by creating a master mix consisting of lysis buffer and proteinase K(PK). The master mix was composed of 300 µl of lysis buffer and 30 µl of PK for 1.5x the number of samples. 1.5 mL Clickfit tubes were then labeled to correspond to the samples collected, and a clearing column was placed in each. Each sample was placed into its respective tube and broken off to ensure that only the swab head was in the tube. Of the previously prepared lysis buffer/PK solution, 330µl was added to each swab and the tube was closed tightly. Each tube was vortexed for 10 seconds and the samples were incubated in a heating block at 56° C in an Eppendorf Thermomixer, as seen in Figure 45.

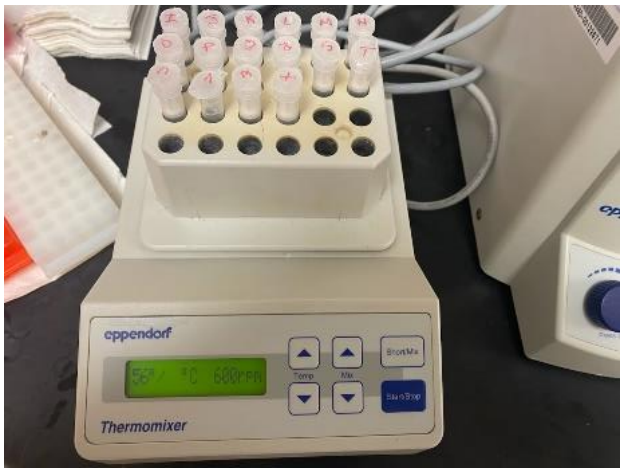


Figure 45:Samples incubating in a heating block using an Eppendorf Thermomixer

One cartridge for each sample was obtained, labeled according to its respective sample, and shaken slowly numerous times to resuspend the magnetic beads. The seal of each

Casework Sample Cartridge was cautiously peeled to avoid mixing or splashing of the reagents. The entirety of the flowthrough was added to well #1 in its respective cartridge. A Low Elution Volume (LEV) plunger was then placed into well #8 as seen in Figure 46

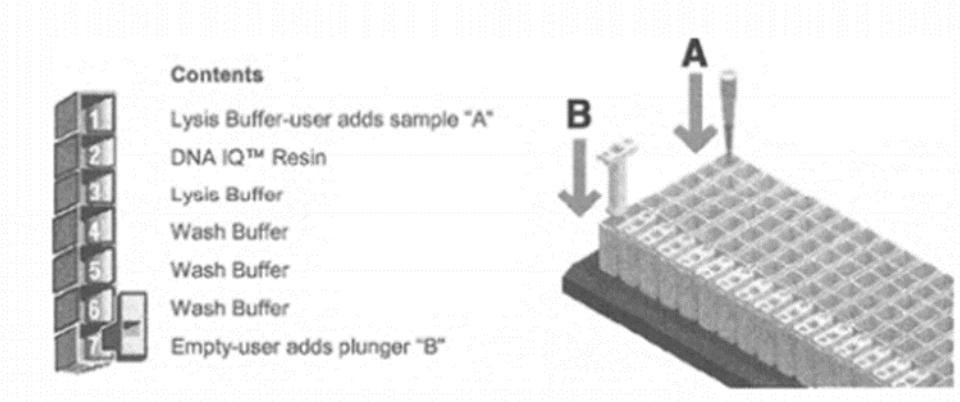


Figure 46:Diagram of Maxwell Extraction Kit[83]

For each sample, an elution tube was loaded with 50 µl of elution buffer and the tube was left open with the lids facing the user. The Maxwell 16 instrument was powered on and the diagnostics began. The "Run" command was pressed on the instrument and DNA -> FFPE/Cell was selected and verified. The loaded cartridges were placed in the platform placeholder and the run was initiated by pressing "Run." After the run was completed, the elution tubes, which now contained the extracted DNA, were sealed and stored at -20° C

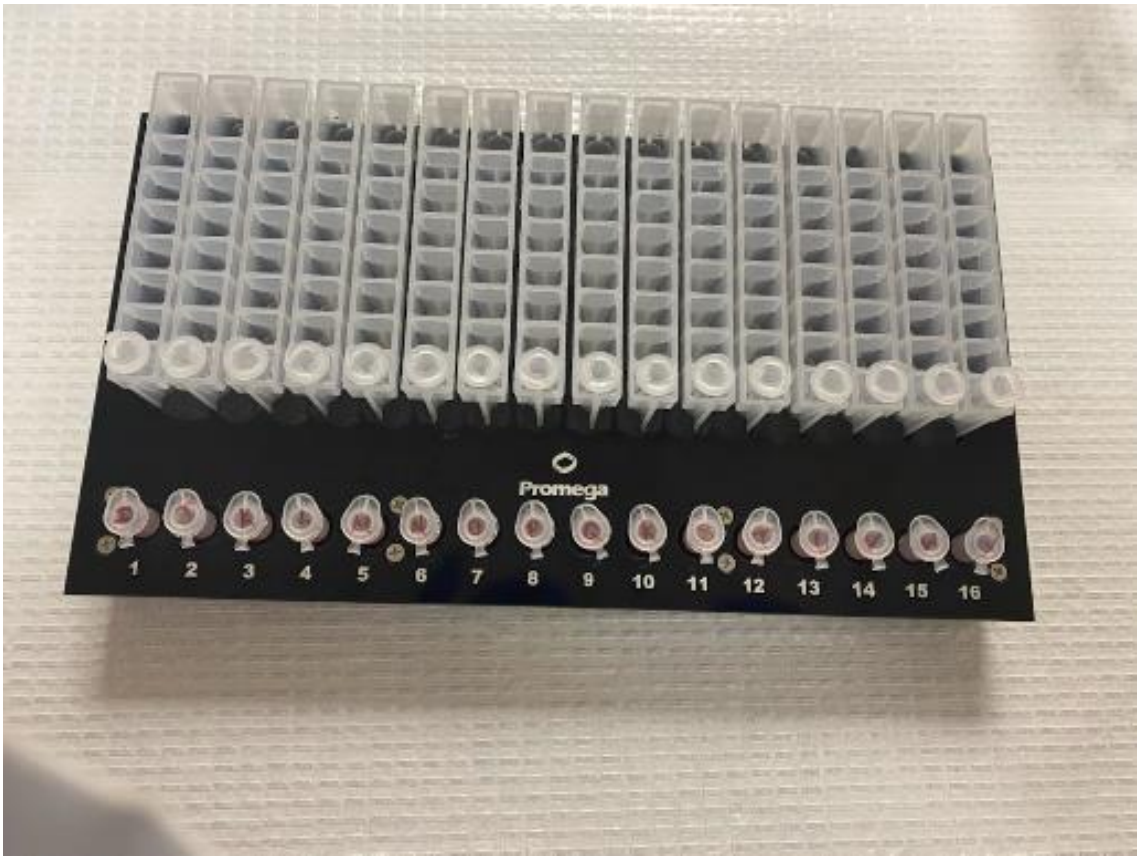


Figure 47:Samples placed in cartridge for Maxwell extraction

7.3 DETERMINATION OF PRIMERS FOR STR ANALYSIS OF HLA GENES

An extensive literature review was conducted to determine which primer sequences will promote the amplification of the HLA regions of interest. Outlined in Table 16 are the primers with their respective primer sequences used for this study.[19,103]

Table 16: STR targets with their respective primer sequence[19,103]

STR	**Forward primer sequence (5'-3')**	**Reverse primer sequence (5'-3')**
D6S2917	HEX-AGGCCTTTTTCACTTGTTTTTCTA	TCTTCTGCTCTACCCACCA
M138	FAM-GAGAGAGCGAGATTCTGC	GAATGCTTAGAAGATGAAGGTG

D6S2837	HEX-CCAAGAAAGAAAGAACCAATAGCA	GTTTCTTATCAGGGAAAGGTGCTGGT GAGCCAATTAGCCCAATAAATCAC
D6S2925	FAM-TCTAGTGTCTTCTGGCCTTG	GTTTCTTATCAGGGAAAGGTGCTGGT
D6S2787	FAM-CTCCAGCCTGGATAACAG	ACAAGGGCTTTAGGAGGTCT

7.4 QUANTIFICATION OF GENOMIC DNA

The genomic DNA was quantified using the Qubit 2.0 kit. The kit came equipped with buffer, fluorophore, and two known standards. The buffer and fluorophore were used to create a working solution that intercalates DNA. The known standards were used to create a calibration curve reflective of gDNA concentration. The amount of 0.5 mL Qubit® assay tubes were collected for standards and samples, and each tube lid was labeled carefully. The working solution was prepared by combining 1 µl of fluorophore and 199 µl of buffer. Each sample received 198 µl of working solution and 2 µl of sample.

The standards received 190 µl of working solution and 10 µl of standards and were vortexed for three seconds. The tubes were incubated at room temperature for two minutes.

7.2.3.2 INSTRUMENTAL ANALYSIS OF STANDARDS AND SAMPLES

On the home screen of the Qubit 2.0 Fluorometer, the following commands were selected:

$$\text{“}DNA\text{'”} \rightarrow \text{"}dsDNA\ High\ Sensitivity\text{"} \rightarrow \text{"}New\ Standards\text{"}$$

The Qubit tube containing standard #1 was placed into the sample chamber and the lid was closed. “Read standard” was then pressed. When the reading concluded, standard #1 was removed and the same process was replicated for sample #2. The calibration curves were constructed based on these standards. Each sample was then placed into the chamber and quantitated based on the calibration curve obtained. Samples were then diluted to 10 ng/µl and stored in the freezer at -20 °C for normalization of gDNA prior to PCR implementation.

STR genotyping was performed by performing PCR to amplify 5 STR loci that, according to the literature, are within the HLA region.[106] Hex and FAM fluorescently labeled STR primers were used. The fragment analysis was then followed using the SeqStudio™ Genetic Analyzer.

7.5 PRIMER OPTIMIZATION

7.5.1 PRIMER DILUTIONS

Once the primers were selected, the primers were purchased through Integrated DNA Technologies (IDT). The labeled (forward and reverse) and unlabeled primers were briefly centrifuged before opening. The oligo yield information was observed on the IDT specification sheet and then used to determine the amount of buffer needed to make a 100 µM solution of each primer. The primers were then reconstituted in PCR grade water under a UV hood to make a 100 µm stock solution. The high concentration of the stock solution, coupled to the possible risk of contaminating the PCR reaction, called for working solutions to be made. A working solution of each primer was created by taking

aliquots of the stock solution. Eight working solutions were then created, and the aliquots were labeled and stored at -20°C.

7.2.4.2 MULTIPLEX PCR MASTER MIX PREPARATION

Multiplex PCR Amplification was required for two master mix preparations. The dilution equation was performed to determine the final primer concentration needed. The master mix contents are depicted in Table 17.

Table 17:Contents of primer mixes for Multiplex PCR amplification

Master mix	**Contents**
Master Mix I	M138 (F/R), D6S2925(F/R), D6S2837 (F/R)
Master Mix II	D6S2917(F/R), D6S2787 (F/R)

Master Mix I stock solution was created by transferring 37.5 µl of each 20 µm primer into a 0.50 µl tube. Next, 37.5µl of hyclone H_20 were added to the mix. Master Mix II stock solution was created by transferring 37.5 of each 20 µm primer into a 0.50 µl tube. After this, 450 µl of hyclone H_20 were then added to the mix. From the calculations below, it was established that 1.25 µm per primer were needed for Master Mix II.

7.5.2 GEL ELECTROPHORESIS

Gel electrophoresis was used to observe the quality of the multiplex PCR amplification. An agarose gel was prepared by mixing agarose powder with TAE buffer (1%). Ethidium bromide was then added to the gel to aid with the visualization of DNA during the

reaction. The solution was then cooled and poured into a casting tray consisting of a sample comb. The system was left to solidify at room temperature. The tray was then transferred into the electrophoresis chamber and covered with buffer. The DNA samples were then pipetted into the sample wells and the DNA began its migration towards the positive electrode. Each individual band was assessed, and their genotypes were identified.

7.6 POLYMERASE CHAIN REACTION (PCR)

HLA class II genotypes were identified using multiplex PCR to amplify the five loci of interest using HEX and FAM fluorescently labeled oligonucleotide primers. Multiplex PCR was applied to 2–3 STR loci per reaction. To conduct the multiplex PCR, Master Mix I and Master Mix II were used for each sample in two separate tubes. A positive and negative control were carried out within each experimental run. For each sample, 10 µl of the 2x Multiplex PCR master mix (Qiagen, Valencia, CA), 2 µl of 5X Q-solution (Quiagen), 4 µl of genomic DNA at 10 ng/µl with 16 µl of primer mix for each tube. A total of 40 cycles were used for the amplification performed using the following cycling parameters: 1 cycle of 34°C for 15 min, 1 cycle of 94° for 30 s, 1 cycle of 65°C for 90 sec, 1 cycle of 72°C for 30 s, 5 cycles at 94 °C for 30 secs, 1 cycle of 60 °C for 90 s, 1 cycle of 72 °C for 30 secs, 30 cycles of 94°C for 30 sec, 1 cycle of 55°C for 90 s, 1 cycle of 72 °C for 2 mins, 1 cycle of 72°C for 45 min, final 4 °C Hold.

Each subject that consented to participating in the study booked an appointment scheduled via email. Subjects were asked to refrain from eating, drinking, brushing teeth,

or smoking 30 minutes prior to sampling. When the subject arrived, they were asked to sit. The protocol was then explained to each subject to ensure they knew what to expect. The HydraFlock 6" Sterile Standard Flock Swab with Polystyrene Handle was used to collect buccal swabs from each subject. Additionally, powder-free nitrile exam gloves were used (one pair per subject) to decrease the likelihood of contaminating the buccal swabs. The buccal swab was removed from the sterile pouch and the specimen was carefully placed in the subject's hand. Each subject was asked to rub their right cheek for a specific time (30 seconds, 60 seconds) as indicated in Table 18.

Table 18:Buccal swab sample time

Sample Number	Time (seconds)
1	30s
2	60s

The swab was then transferred into a sterile tube labeled with the respective subject's information and the amount of time the swab was rubbed against the subject's cheek (30s or 60s). The swabs were then placed in a freezer and stored at -20°C.

8. PILOT STUDY

8.1 PRELIMINARY STUDIES

Prior to extracting and quantifying the pilot study samples, studies were conducted to determine the optimal protocol for buccal swabbing. The goal of this study was to establish a buccal swabbing method that would produce an efficient amount of gDNA. Three subjects of Hispanic and African American descent were selected for a preliminary study. The gDNA was obtained via Maxwell extraction. Two conditions were

manipulated to compare which method would produce the best yield: time and the effect of mouth rinsing. For the study, the following conditions were applied:

Table 19:Optimization of buccal swab conditions

Sample	Conditions applied
T101 (Hispanic Male)	No wash (30 seconds)
T101 (Hispanic Male)	No wash (60 seconds)
T101 (Hispanic Male)	Wash (30 seconds)
T101 (Hispanic Male)	Wash (60 seconds)

The conditions above were repeated for T102 (African American male) and T103 (African American female), respectively. For quality control purposes, a positive control (100 µl of human standard) and negative control (100 µl of PBS) were taken through the process as well.

STR analysis was performed via capillary electrophoresis. The fragments recovered from PCR were identified using the SeqStudio™ Genetic Analyzer (Applied Biosystems, Foster City, CA). The size standard was created by mixing 11.5 µl of Hi-Di ™ with 0.65 µl/rxn of GeneScan™ 500 ROX™ internal standard. Each sample consisted of a 1:3 dilution of each PCR reaction and 1 µl DNA per reaction. The PCR product was then denatured by incubating at 95° C for three minutes and snap cooled for five minutes on ice. Each sample was run on the SeqStudio™ for 42 minutes. The fragments were separated and sized by capillary electrophoresis using polymer POP 1.

The data was then loaded into a DNA fragment sizing software Microsatellite Analysis Software v1.0 (Thermo Fisher Scientific, Waltham MA), which is used to view genetic

profiles and make qualitative inferences using the electropherogram reports produced and quantitative inferences by calculating the size fragments.

The Microsatellite Analysis Software v1.0 were set to the local Southern size calling. Alleles were manually binned based on their expected size. A minimum noise threshold was set at 100 fluorescent units. For amplicons present in PM1 true alleles were called between 106–330 bp. For amplicons present in PM2, true alleles were called between 134–361 bp. Because each STR of interest is a tetranucleotide repeat, any peaks 1 bp apart were deemed an artifact and not accepted as a true value.

8.2 STATISTICAL ANALYSIS

8.21 FRAGMENT ANALYSIS

The electropherograms were created using Thermo Fisher Microsatellite Analysis Software v1.0. The sizing tables were exported to MS Excel (Microsoft, Redmond, WA), and the R software was used for data analysis. The variation in size of alleles within each sample and frequency of their occurrence were analyzed to evaluate if these alleles can serve as unique identifiers when characterizing ethnicity. During fragment analysis, fragment primers labeled with fluorescent dyes from PCR were sized and quantitated for important genetic information about specific targets. In this study, genotype frequency, allele frequency, and estimated size fragments are of interest and will be investigated to determine if the HLA gene complex can identify individuals according to their ethnicity via buccal swabbing.

8.2.2 LENGTH HETEROZYGOSITY POLYMERASE CHAIN REACTION (LH-PCR)/ INDICATOR SPECIES ANALYSIS (ISA)

Using the R software, LH-PCR was used to assess the length variation among the HLA genes under study. The sizing tables of collected human samples were compiled into one large data matrix. The data was then normalized to relative abundance. Relative abundance was calculated by finding the sum of each row within the matrix to obtain the total intensity for the data set. Each row was then divided by total intensity to calculate the relative abundance of each allele. The abundance of each allele was then used as a vector for interspecies analysis. The number "0" was assigned when a peak was absent. The allele was chosen as an indicator if it could predict the diversity of a population. The alpha parameter was set to 0.05, and all ethnic specific allele associations that were not significant at this level were not included.

8.2.2.LINEAR DISCRIMINANT ANALYSIS

Linear Discriminant Analysis (LDA) was performed to determine if the genotype frequencies acquired are reliable biomarkers for ethnicity classification.

Linear Discriminant analysis is a machine learning technique that classifies categorical data based on a covariance matrix. In this study, LDA was used to evaluate the genotype frequency's capability to predict individuals by ethnicity. The genotype frequency was utilized as the continuous variable. The statistical software used for LDA was JMP® Student.

9. Results and Discussion

9.1 Preliminary Studies

Prior to extracting the pilot study samples, the optimal swabbing method was assessed by determining if a subject washing their mouth prior to sampling produced a greater yield. This study is of great importance due to the small sample size acquired. The optimization of methods decreases the likelihood of error and provide samples delegated specifically for troubleshooting purposes. The results revealed that the amount of time the swab was rotated in the mouth of the subject did not make a significant difference in the amount of sample collected. Buccal swabs are known for their ability to acquire DNA with great quality. Seven out of the eight samples collected were out of standard range. As stated in the literature, the amount of DNA recovered is 40% greater in subjects that swab prior to brushing their teeth.[113] The study also validated that the Maxwell 16 was sufficient for DNA acquisition. Therefore, for ease of sample collection, 30-second rotations coupled with no washing of the mouth prior to swabbing was selected as the optimal method.

Table 20: Recovery of DNA based on sample collection method

Sample Name	Wash or No Wash	Recovery
TT101- HM (1&2)	No Wash (30 Secs)	> 600 ng (sample too high)
T101- HM (3&4)	No Wash (1 Min)	> 600 ng (sample too high)

T101- HM (5&6)	Wash (30 sec)	> 600 ng (sample too high)
T101- HM (7&8)	Wash (1 min)	> 600 ng (sample too high)
T102 AAM (1&2)	No Wash (30 Secs)	> 600 ng (sample too high)
T102 AAM (3&4)	No Wash (1 Min)	> 600 ng (sample too high)
T103 AAF (1&2)	Wash (30 sec)	51.0 ng/µl
T103 AAF (3&4)	Wash (1 min)	> 600 ng (sample too high)

9.2 QUANTIFICATION OF GENOMIC DNA

Based on the known standards provided in the Qubit Assay Kit, the DNA recovered via Maxwell extraction was quantified. The concentration of each sample was calculated with high selectivity. The average quantitated DNA extracted by each were then calculated. Figure 48 displays the average amount of gDNA extracted within each population (Caucasian, Hispanic, African American). When comparing the recovery of DNA among ethnicities, African Americans yielded the most DNA (37.94 ng/µl), followed by Hispanic subjects (28.77g/µl), and lastly Caucasian subjects (28.77 ng/µl). The DNA yields obtained using the Maxwell extraction had a small standard deviation among ethnicities (5.08 ng/µl),

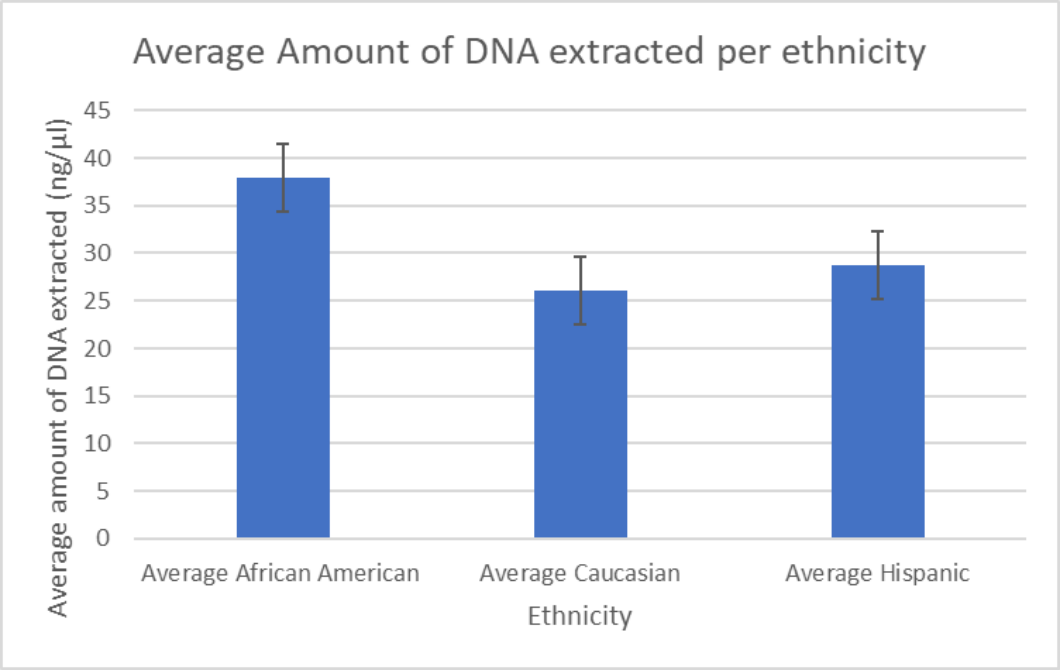


Figure 48: Average amount of DNA extracted per ethnicity

which highlights that the extraction method utilized was precise. Among the African American subjects, the DNA yields ranged from 10.8 ng/µl to 60 ng/µl. The median amount of gDNA extracted was 37.0 ng/µl. Among the Caucasian subjects, the DNA yields ranged from 8.04 ng/µl to 43 ng/µl. The median amount of gDNA was 25.65 ng/µl.

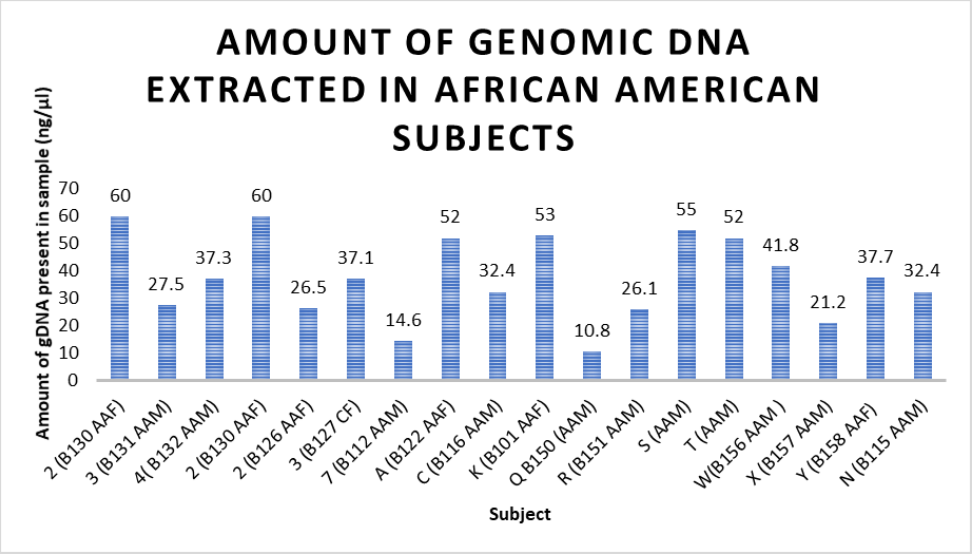


Figure 49: Amount of Genomic DNA extracted from African American Subjects

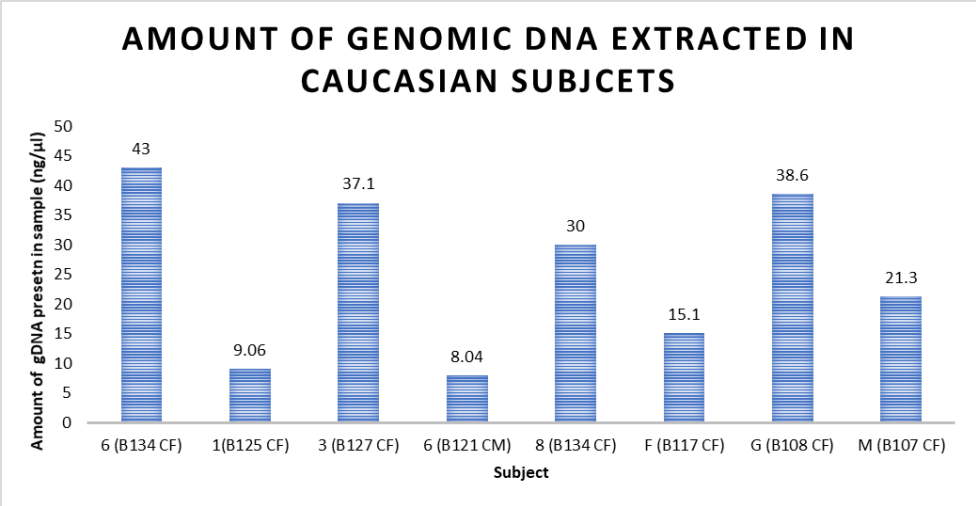


Figure 50:Amount of Genomic DNA extracted from Caucasian subjects

Among Hispanic subjects, the median amount of gDNA was 31.05 ng/µl.

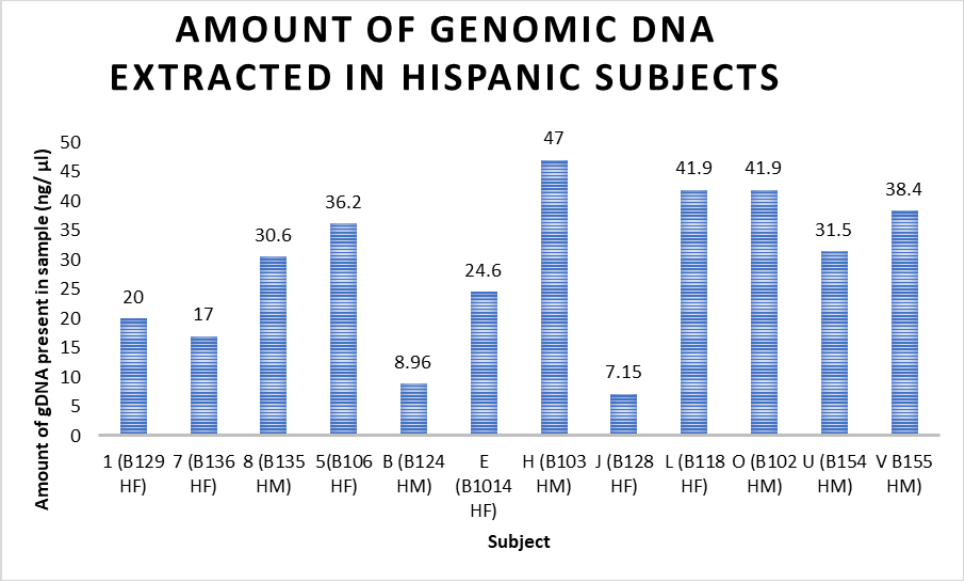


Figure 51: Amount of Genomic DNA extracted from Hispanic subjects

Of the 31 buccal swab samples quantified, 17% of the samples were out of range as a result of the high amount of DNA extracted from subjects. Of the samples, 83% were quantified successfully using the calibration curves constructed using standard 1 and standard 2. For samples greater than 60 ng/µl, the samples were diluted.

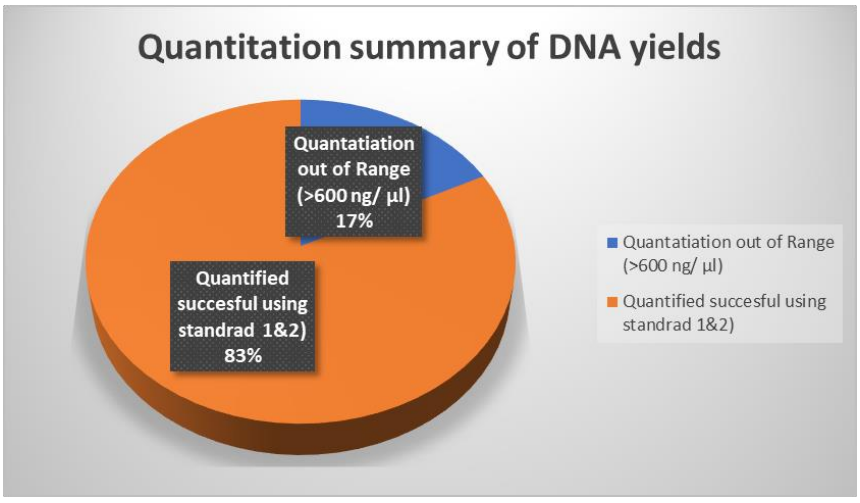


Figure 52: Quantitation summary of DNA yields

9.3 MULTIPLEX PCR OPTIMIZATION

For the seven samples that were not able to be quantified as a result of the high amount of DNA present (> 60 ng/µl), $M_1V_1=M_2V_2$ was used to determine how much H_2O should be used to dilute the samples to 40 ng/µl, where M_1 is equal to the approximate amount of DNA present, V_1 is equal to the volume of DNA left in the tube, M_2 = desired concentration of DNA (40 ng/ µl), and V_2 = volume needed for an efficient dilution (Table 21).

Table 21: Amount of water needed to dilute samples

Sample	**H_20 Needed**
B131 (AAM)	10.0 µl
B133 (AAM)	15.0 µl
B132 (AAM)	7.5 µl
B125 (CF)	17.5 µl
B127 (CF)	17.5 µl
B105 (HF)	20.0 µl
B137 (HM)	17.5µl

The samples were then re-quantified using the Qubit 2.0 Fluorometer. All samples fell within range for quantitation, which exemplifies the importance of diluting high concentration analytes. For normalization purposes, each sample was diluted to 10 ng/µl prior to PCR implementation. The diluted samples were then stored at -20 °C in preparation for PCR.

The gel image in Figure 53 is the result of restriction digestion via gel electrophoresis. Using the PCR product of two subjects, the behavior of the multiplex reactions was studied. Qualitative and semiquantitative observations were made to determine the genotype and to confirm the size range for each primer.

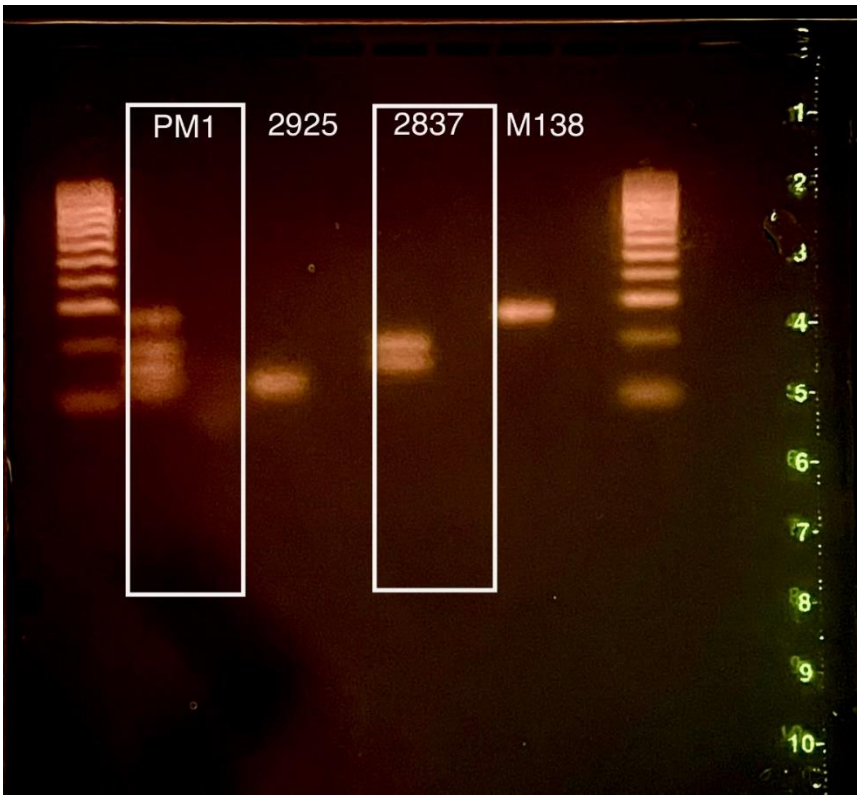


Figure 53: Amplification of three genes using Multiplex PCR reaction (Primer mix I)

For Primer Mix 1(PM1), the Primer mix (PM1) in the first lane was analyzed to easily cross reference and determine if each primer present in the mix would show up. In the second lane (2925), there appears to be only one band present; however, it does not have the intensity of a homozygote. Therefore, it was inferred that the band is heterozygous. In the third lane (2837), there are two clearly distinct bands. This means that this band is

identified as heterozygous. In the lane labeled M138, there is only one band visible that is extremely bright. This band is identified as homozygous. The size allele was also assessed to determine if the size range observed fell within the expected range A summary of the findings from PM1 are shown in Table 21.

Table 22:Multiplex PCR amplified genes with size range and genotype

Primer	Expected size range (bp)	Observed size range (bp)	Genotype
D6S2925	106–141	100–150	Heterozygous
D6S2837	137–199	150–200	Heterozygous
M138	235–320	200–300	Homozygous

Each primer was observed close to the expected size range. For Primer Mix II (PM2), in the lane labeled 2787, there are two alleles that appear to be close together. This is because the repeats are only 2–6 bp apart. The intensity of the band determined STR 28787 to be heterozygous. In the second lane, there is a distinct line between the two bands, which means the band is undoubtedly heterozygous. The qualitative information revealed from the gel electrophoresis confirmed that the multiplex PCR reaction can successfully amplify the regions of interest.

Table 23: Multiplex PCR reaction amplified regions with size range and genotype (Primer Mix II)

Primer	Expected size range (bp)	Observed size range (bp)	Genotype
D6S2787	134–163	200–300	Heterozygous
D6S2917	247–361	100–200	Heterozygous

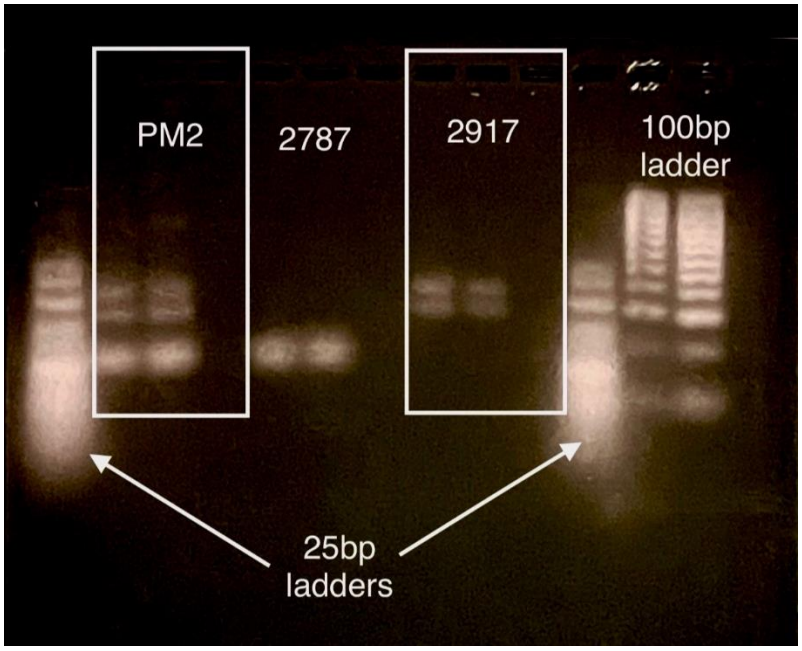


Figure 54: Amplification of two genes (D6S2787 and D6S2917) using Multiplex PCR reaction (Primer Mix II)

9.4 PRIMER MIX I ANALYSIS

9.2.3.1 Primer Mix I indicator species analysis

The frequency of each true allele called within the electropherogram was assessed to determine if there was a significant increase or decrease in specific populations (Caucasian, Hispanic, African American). The frequency of HLA genes has been previously demonstrated to show variation among different ethnic populations.[114] Using the two optimized multiplex PCR reactions on each sample successfully amplified the targets of interest (D6S2925, D6S2837, M138, D6S2787, D6S2917). The size of the amplicons was categorized according to their respective gene, as reported in the literature. The frequency of each gene among each ethnicity was assessed. The high number of peaks observed were due to the genetic variation present among each

tetranucleotide STR loci. A sizing rule was adopted for each gene to ensure consistent rounding and calling of the DNA fragments occurred. Each primer mix was analyzed independently to facilitate the analysis of genes identified in similar size ranges. Primer Mix I amplified the following genes: D6S2925, D6S2837, M138. Preliminary statistical analysis of the data was used to further investigate if there would be any variation in the presence of each gene based on ethnicity. Each gene was identified by their expected size range. Although an overlap between the size ranges is present between D6S2925 and D6S2837, different primers were used to amplify the regions to provide a clear distinction between the two genes.

Table 24: Target regions using Primer Mix I with expected size range

Gene	Expected Size Range (bp)
D6S2925	106–141
D6S2837	137–199
M138	235–320

The bar graph shown in Figure 55 highlights the variation in gene presence among each ethnicity. As visualized, Hispanics carried the D6S2837 gene most frequently, followed by Caucasian and lastly African American. African Americans carried the D6S2925 gene most frequently, followed by Hispanics and lastly Caucasians.

The M138 gene was more frequent in African American subjects, followed by Hispanics and was least observed in Caucasians.

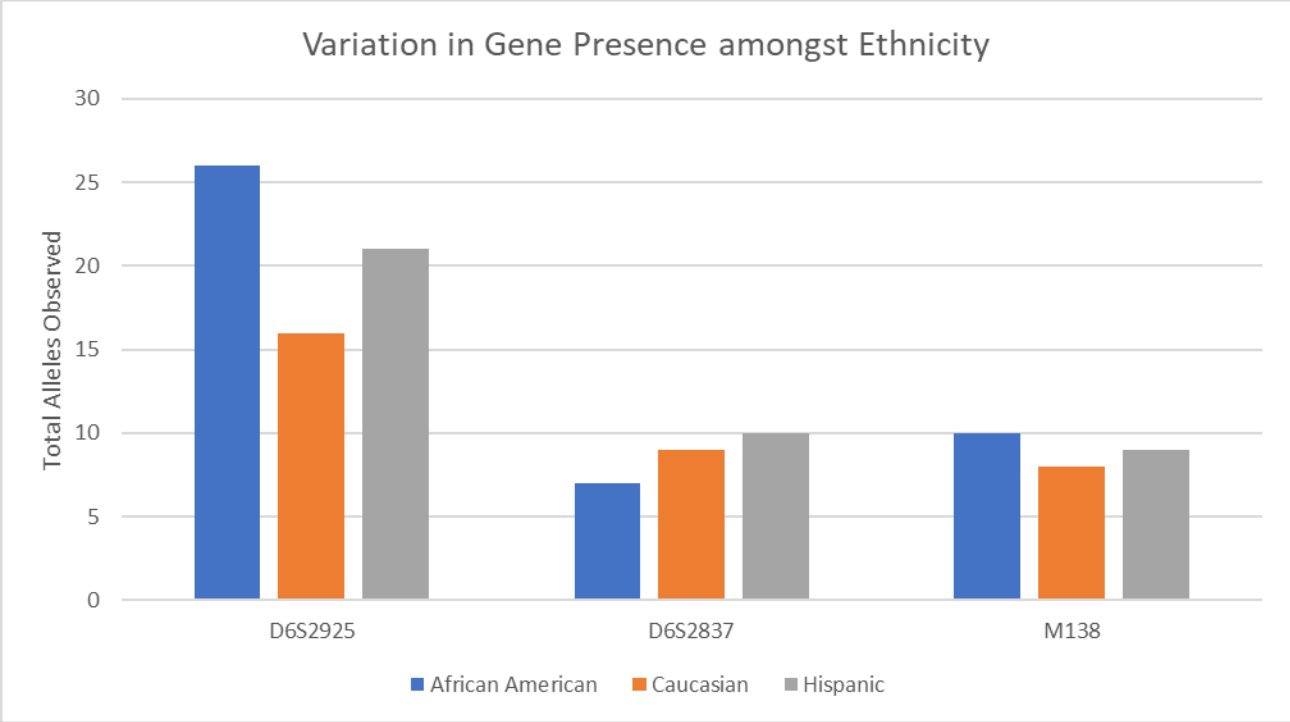


Figure 55: Variation in gene presence among each ethnicity

A chi square test was conducted to determine if the preliminary observation of each gene was statistically significant. The chi square test revealed that the differences between the frequency distributions are statistically significant (p = 0.00293), which means that the presence of the gene within each population is not random and is potentially affected by ethnicity. Further qualitative and quantitative studies were needed to further validate this hypothesis. To further investigate, Length Heterogeneity Polymerase Chain Reaction (LH-PCR) coupled to an indicator species analysis was employed based on the differences in length within each HLA region of human samples from each ethnicity.

Length Heterozygosity PCR (LH-PCR) provided a gateway to quantitatively express the presence of each allele. Indicator species analysis was utilized as a fingerprinting tool to determine which alleles are frequently abundant in each population with significance. The frequency of occurrence of each allele within each population was calculated. The frequently abundant amplicon sizes ranged from 113–178bp. Of the six sizes deemed as a unique associate to a specific ethnicity, only two of the three genes were significantly capable of ethnic specific observations (D6S2925, D6S2837). The indicator species analysis did not associate the M138 gene with a specific ethnicity. Furthermore, the Caucasian population did not have any alleles solely representative of their population. Two sizes (113, 146) were associated with the African American ethnicity. On the other hand, four sizes (175, 176, 177, 178) were associated exclusively with the Hispanic population. The ability of the HLA-B region (D6S2925) to successfully investigate genetic diversity of African Americans supports earlier observations that showed genetic distinctions among two ethnic groups within the African population.[115]

Table 25:Frequently abundant alleles withing gene D6S2925 and D6S2837 detected via indicator species analysis

Size	**Gene**	**Ethnicity Association**	***P* Value**
113	D6S2925	African American	0.0473
146	D6S2837	African American	0.0167
175	D6S2837	Hispanic	0.0205
176	D6S2837	Hispanic	0.0225
177	D6S2837	Hispanic	0.0277
178	D6S2837	Hispanic	0.0015

After the indicator species analysis identified the alleles that are most significant for the populations under study, one-way ANOVA was used to determine whether the differences in the relative abundance between ethnicities were statistically significantly. Only the statistically significant alleles are shown (Size 175–178). Sizes 113 and 146 had a *p* value > 0.05 (0.705 and 0.0734, respectively), which means they were not statistically significant when comparing ethnicities.

Table 26: Statistically significant alleles within D6S2837 gene detected via indicator species analysis

Size	Gene	ANOVA
176	D6S2837	$F = 4.081$ $P = 0.0283$
178	D6S2837	$F = 12.1$ $P = 0.005$

Each of these sizes (176 and 178) are directly correlated to the D6S2837 gene. The Tukey's HSD post hoc test was implemented to compare the mean values for each of these alleles. If the ethnicity had the same letter present within their group, this means they were not statistically different.

Table 27: Tukey's HSD mean comparison of relative abundance among ethnicities (Allele 176)

Ethnicity	Size	Group
Hispanic	0.015207634	a
Caucasian	0.003899443	ab
African American	0	b

The mean values of size 176 ranged from 0–0.0153. This study revealed that the mean values of Hispanic and African Americans were different. All other mean comparisons between ethnicities using size 176 did not show significant differences. The post hoc test for size 178 revealed that the mean values of Hispanics were different from Caucasians and African Americans. All other mean comparisons using size 178 did not show significant differences.

Table 28: Tukey's HSD mean comparison of relative abundance among ethnicities (Allele 178)

Ethnicity	Size	Group
Hispanic	0.0089312405	a
Caucasian	0.003899443	b
African American	0.0011210979	b

The allelic frequency associated with a specific ethnicity was obtained for the data obtained by evaluating each subject's profile for the presence or absence of each allele. The number of occurrences of each allele was tailed and divided by the total number of subjects in the sample pool to obtain a percentage of occurrence for each ethnicity

9.2.3.2 Gene D6S2837 Allele Frequency Study

Within gene D6S2837 (137 bp–199 bp), a total of 12 alleles were observed between the population sampled. Figure 56 shows the number of alleles observed for each ethnicity. The D6S2837 gene was predominant in the Hispanic population with 10 alleles observed within their population, followed by Caucasians (9 alleles) and lastly African Americans (7 alleles).

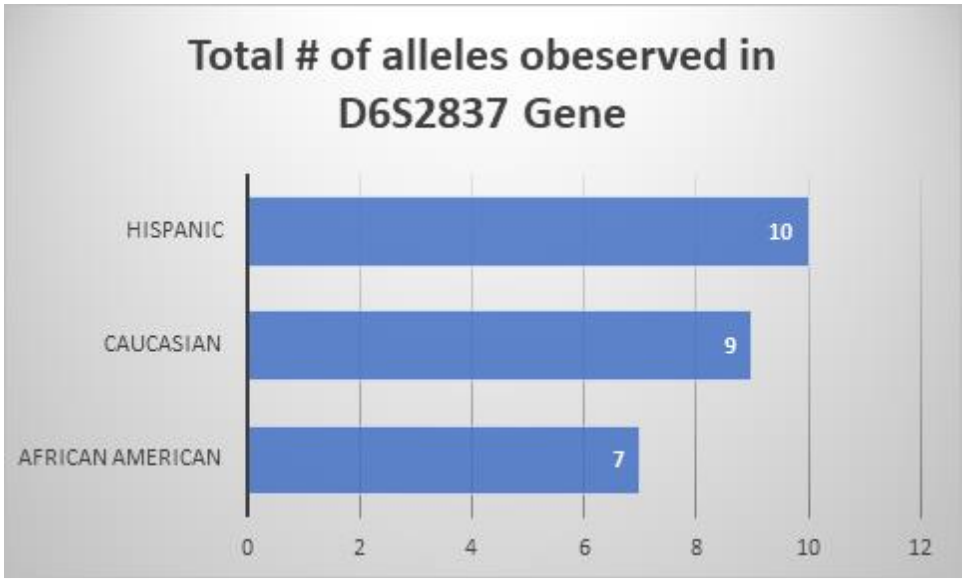


Figure 56: Alleles observed in D6S2837 gene

The number of alleles observed in the Caucasian population supports previous reports of this gene's observation in subjects of European descent.[19] Figure 57 highlights the specific alleles that are most prevalent among each ethnicity.

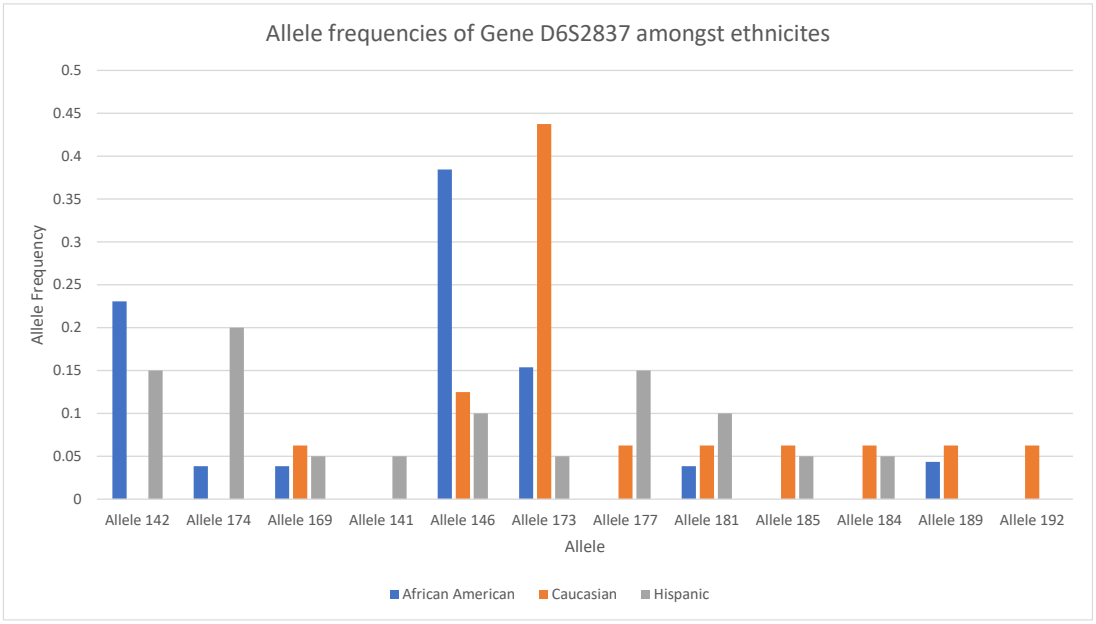


Figure 57: Allele frequencies of Gene D6S2837 among ethnicities

The increased frequency of Allele 146 in African Americans and Allele 177 in Hispanics further validates the findings from the indicator species analysis that deemed these alleles as frequently abundant. The absence of alleles within a population could potentially be used in forensic casework for the exclusion or inclusion of a specific ethnic group. The allele frequency study revealed that Allele 141 was found only in Hispanic subjects. On the other hand, Allele 192 was found only in Caucasian subjects and 184 was only present within African American subjects. This means that if the D6S2837 gene was used to evaluate a suspect, then if the person expresses Allele 184, with the absence of Alleles 141 and 192, the suspect could potentially be identified as African American. A genotype study of each was conducted to further asses the STR profiles of each subject.

9.2.3.3 Gene D6S2837 Genotype Frequency Study

Determination of Genotype Combinations

The HLA genome has been used to identify specific genotype combinations that identify diseases such as multiple sclerosis in populations.[116] The genotype combinations among ethnicities were conducted to determine if any distinct allelic combinations were present within each population. The electropherogram of each subject was examined to determine the genotype of each subject at the specific loci of interest. The combined probability of allele combinations was calculated using the following equation adopted from the literature:[117]

Equation 6: Combined probability of allele combinations

$$P = \frac{m}{n}$$

where m = the number of allele combinations present at a specific loci and n = the total number of samples within the population. The genotype combinations present within each sample were identified in the DS62837, D6S2925, and M138 genes. A summary of the genotype frequency of each subject is depicted in Appendix 4 through Appendix 8. When a subject did not express a specific combination, a value of 0 was assigned.

The 13 genotype combinations observed within the sample set were further assessed to determine if the combination occurred frequently among any specific population. An example of a combination present within two different Hispanic subjects is depicted in Figure 58.

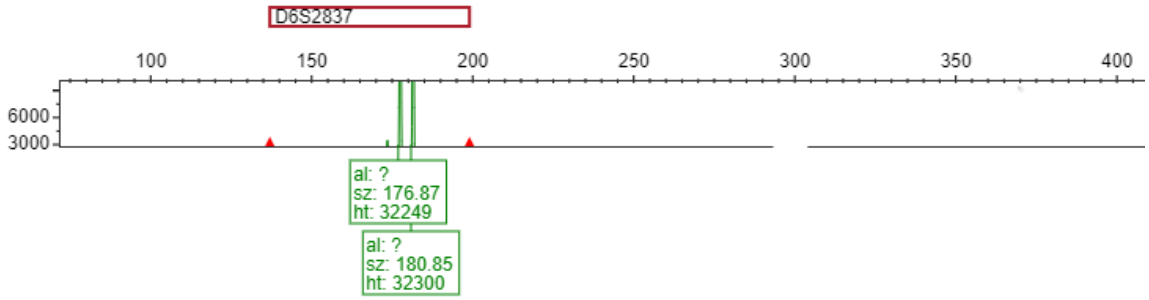


Figure 58:Electropherogram showing genotype of Hispanic subjects

Three allele combinations were present within multiple samples at a different frequency. Of the three combinations, two of the three (142,174 and 177,181) were found exclusively in the Hispanic population. The alleles observed within the Hispanic population as distinct markers highlighted the “extreme polymorphism” of the HLA

region discussed in previous studies.[118] Although these alleles alone can identify Hispanic individuals successfully (according to the interspecies analysis), when each heterozygote was assessed, there were only two unique combinations present. However, 60% of the samples collected had at least one of the sizes predicted using the interspecies analysis present. Furthermore, the Hispanic population was the only population to express 100% heterozygosity at this locus. A linear discriminant analysis was computed to determine if the genotype frequencies acquired are capable of both predicting and separating ethnicity. According to LDA analysis, subjects of different ethnicities are clearly distinguished based on genotype frequencies. With the use of LDA, all 31 subjects were correctly classified at 87.10% accuracy. The prediction rate of the model is illustrated in Table 29.

Table 29:Prediction rate of identifying ethnicity based on genotype frequency using gene D6S2837

Actual	Predicted Rate		
Ethnicity	**African American**	**Caucasian**	**Hispanic**
African American	1.000	0.000	0.000
Caucasian	0.250	0.750	0.000
Hispanic	0.200	0.000	0.800

The 12.90% misclassifications resulted from incorrect classifications of Caucasians and Hispanics. When evaluating the discriminant scores of the eight Hispanic subjects, two were misclassified as African American. Of the ten 10 Hispanics, two were also misclassified as Caucasian. Nonetheless, the true positive rate ranged from 75–100% for the prediction of each ethnicity. The false positive rate ranging from 20–25% could be improved upon by increasing the sample size. Based on the canonical plots and stepwise regression analysis, Table 30 shows the genotype frequencies selected as most influential

for successful ethnicity prediction. For predicting ethnicity, four combinations (173,173/173,185/173/192/173/177) were equally influential. When ethnic variation is considered, Allele 173 is of great interest due to its redundant expression as an influencer.

Table 30: Most influential genotype for ethnicity differentiation using gene D6S2837

Combination	Canon 1	Canon 2
169,189	2.2455914254	0.422465237
173,173	1.8531513237	2.8684746649
173,185	1.8531513237	2.8684746649
173,192	1.8531513237	2.8684746649
173,177	1.8531513237	2.8684746649

9.2.3.4Gene D6S2925 Allele Frequency Study

Within DS62925 gene, a total of 14 alleles were observed among the populations sampled. Figure 59 shows the number of alleles observed for each ethnicity.

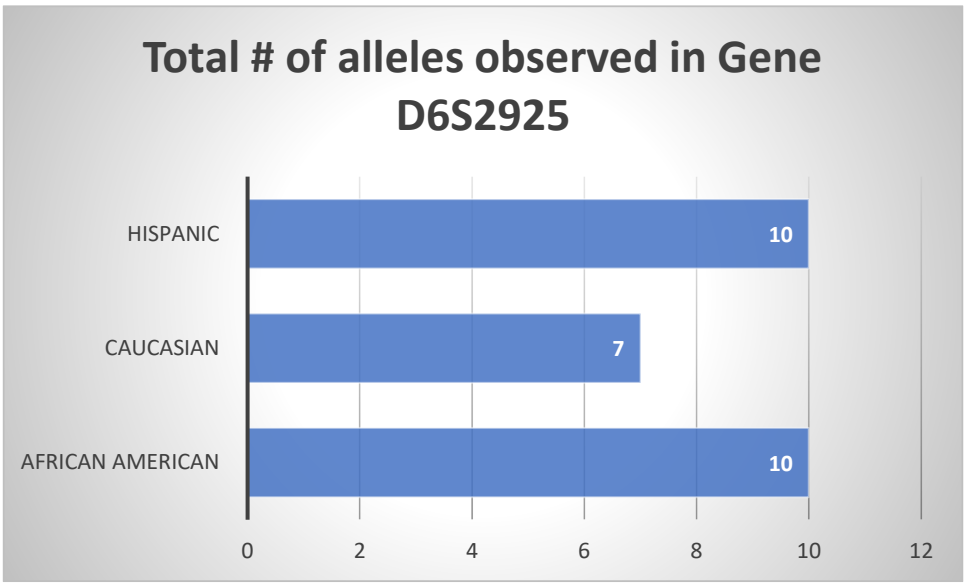


Figure 59: Total number of alleles observed in Gene D6S2925

The D6S2925 gene was equally predominant in the African American and Hispanic population with 10 alleles observed. There were fewer alleles detected in Caucasian subjects (7) than previously reported (9). The bar graph highlights notable differences between the allele frequencies with respect to their ethnicity. Figure 60 displays that African Americans were found to have higher levels of the D6S2925 gene at 7 out of 10 loci. Thus, the ethnicity association of African Americans with this gene supported by the indicator species analysis is further validated. Furthermore, Allele 177 shows a significant increase in frequency among Hispanics, which confirms the association with the Hispanic population inferred by the indicator species analysis. On the other hand, there was an absence of Caucasian subjects at 7 out of the 10 loci observed. Based on these findings, the Hispanic gene D6S925 could be useful in excluding Caucasians from forensic cases. Similarities between allele frequencies were also present. Specifically,

Allele 136 was similar across Caucasian and Hispanics, and Allele 111 was similar across all ethnic groups. However, to further examine the accuracy and reliability of the exclusion rate, a larger sample size should be evaluated.

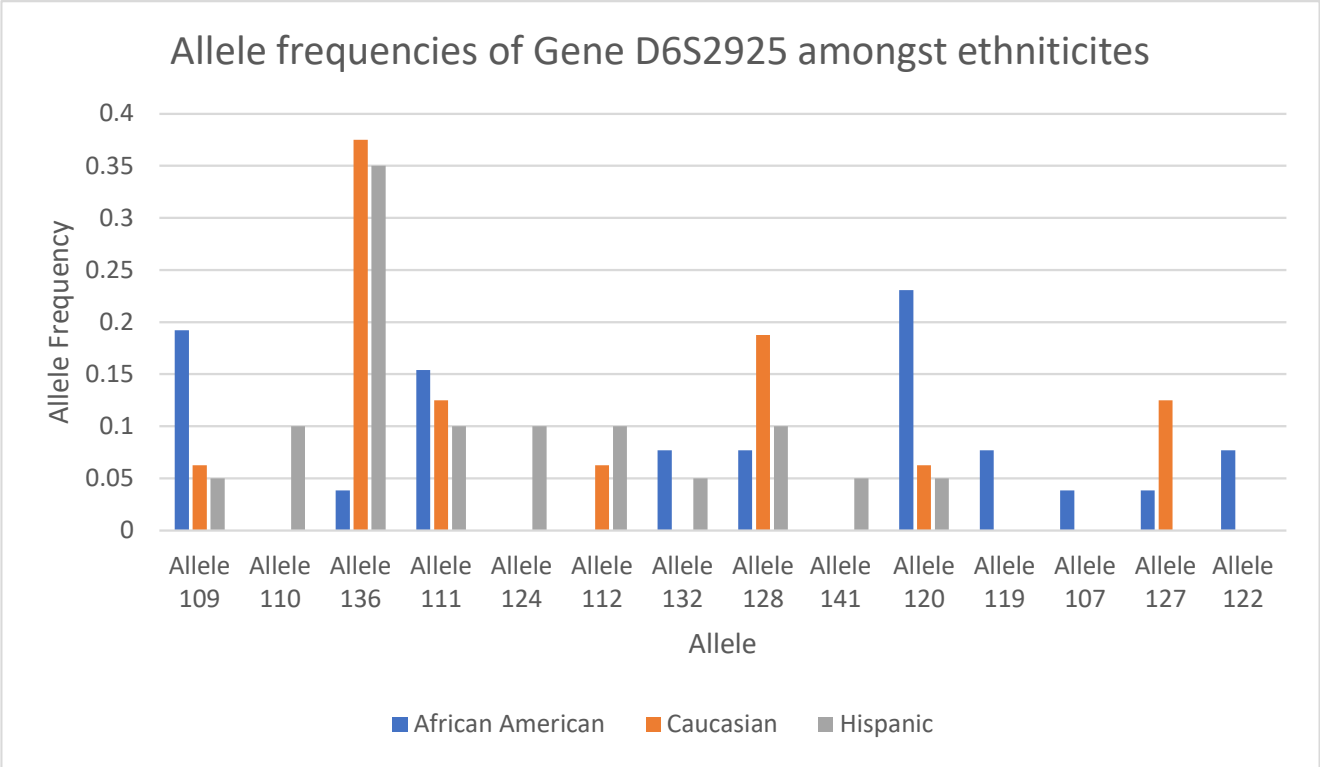


Figure 60:Allele frequencies of Gene D6S2925 amongst ethnicities

9.2.3.5 GENE D6S2925 GENOTYPE FREQUENCY STUDY

To identify whether distinct allelic combinations exist within each ethnicity, various genotype combinations were examined. A summary of the genotype frequencies of each subject is displayed in Appendix 6. Of the 18 loci observed, 14 had genotypes unique to a specific population. However, there was some overlap at four loci. Genotype 109,112 and Genotype 111,127 was present in both Caucasians and African Americans at a different frequency (0.125 and 0.076923, respectively). The homozygous loci 136,136 was present in two Hispanic and two Caucasian subjects at a similar frequency. The haplotype diversity present within the D62925 gene was reflected in the Caucasian sample set. Of the 8 subjects sampled, 6/8 were heterozygous and 2/8 were homozygous. This is in comparison to African American subjects, who had 100% heterozygosity at this specific locus. Homozygosity within the D62925 region may be indicative of Caucasian and Hispanic ethnicity. However, the natural selection influence on HLA alleles across a specific loci requires large sample sizes.[119] Based on the genotype frequencies acquired, it was determined whether ethnicity could be predicted through a linear discriminant analysis. Based on genotype frequencies, different ethnicities can clearly be distinguished by target alleles within the D6S2925 gene. A correct classification of 93.30% was achieved when using the machine-learning technique to predict ethnicity. A summary of the model's prediction rate is given in Table 31.

Table 31: Prediction rate of identifying ethnicity based on genotype frequency using gene D6S2925

Actual	Predicted Rate		
Ethnicity	**African American**	**Caucasian**	**Hispanic**
African American	1.000	0.000	0.000
Caucasian	0.000	1.000	0.000
Hispanic	0.000	0.200	0.800

Actual Ethnicity	Predicted Rate		
	African American	Caucasian	Hispanic

Of the misclassifications, 6.67 % were due to incorrectly classifying Caucasians and Hispanics. When the discriminant scores were evaluated, two Hispanic subjects were incorrectly categorized as Caucasians. African Americans were classified with 100% accuracy. The perfect prediction of this ethnicity supports previous claims that the HLA-DRB1 region (where Gene D6S2925 is located) is a reliable indicator of African American origin.[120] The true positive rate of the analysis ranged from 80–100% for the prediction of each ethnicity. The false positive rates ranged from 0–20%. The canon was evaluated to determine the most influential genotype combinations for ethnicity prediction. The results are shown in Table 32.

Table 32:Most influential genotype for ethnicity differentiation of individuals using D6S2925 Gene

Genotype	Canon 1	Canon 2
128,136	-3.23255	0.546353
112,132	-3.23255	0.546353
112,128	-4.41736	0.291089
111,124	-2.14055	-2.80886
111,120	-2.04775	0.801617

According to the high specificity and sensitivity of this model, the D6S2925 gene could be an effective target for classifying populations based on their genotype frequencies.

9.2.3.6 Gene M138 Allele Frequency Study

Within Gene M138 (235 bp–320 bp), a total of 12 alleles were observed among the population sampled. The M138 gene was predominant in the African American population with 10 alleles observed within their population, followed by Hispanics (9 alleles) and lastly Caucasians (8 alleles). The number of alleles observed in the Caucasian population supports previous reports of this gene's observation in subjects of European descent.[19] Figure 61 highlights the specific alleles that are most prevalent among each ethnicity.

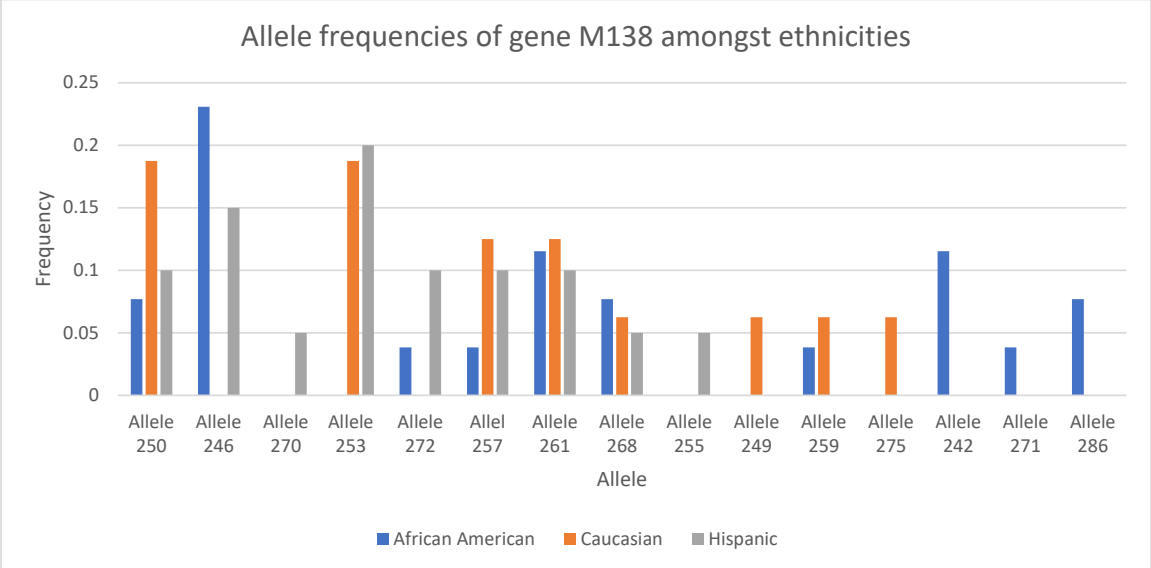


Figure 61: Allele frequencies of gene M138 among ethnicities

Within this gene, each population expressed at least one allele exclusive to their population. A summary of these ethnicity specific alleles is listed in Table 32.

Table 33: Alleles expressed within M138 gene unique to specific ethnicities

Allele	Fingerprint Population
249	Caucasian
255	Hispanic
270	Hispanic
271	African American
275	Caucasian
286	African American

The use of these alleles in forensic investigations could be groundbreaking for identifying the ethnicity of an unknown sample collected at a crime scene. There were also several alleles that portrayed similar frequencies among all populations (Alleles 255, 261, and 238). The presence of alleles at a similar frequency in different regions of the M138 gene may foreshadow challenges with using this gene as a marker for ethnic specific differentiation.

9.2.3.7 Gene M138 Genotype Frequency Study

The genotype of each individual provided more insight on the impact of each allele on the characterization of subjects based on ethnicity.

The genotype combination study revealed that three genotype combinations were frequent among different populations and were not unique to ethnicity. A summary of the genotype frequencies of each subject is displayed in Appendix _

For example, an African American and Caucasian subject both shared the same genotype (235,257) at the same frequency and zygosity (heterozygous). This finding correlates to the allele frequency study that indicated this outcome. However, to offset these similarities, there were two genotype combinations unique to the African American ethnicity (242,268)/ 246,286). As conducted in previous gene studies, the M138 gene was evaluated using LDA to determine this gene's ability to predict ethnicity.

Table 34:Prediction rate of identifying ethnicity based on genotype frequency using gene M138

Actual	**Predicted Rate**		
Ethnicity	**African American**	**Caucasian**	**Hispanic**
African American	0.923	0.000	0.077
Caucasian	0.000	0.875	0.125
Hispanic	0.000	0.100	0.900

The model was able to predict ethnicity at 74.19% accuracy. The true positive rate ranged from 70–77% with African Americans portraying the highest true positive rate (77%) and Hispanics portraying the lowest true positive rate (70%). The 25.8% misclassification rate stems from the overlap of equivalent genotype frequencies. In comparison to the two other genes studied that predicted ethnicity at 80% or greater, this gene had more genotypes that shared the same expression. Therefore, this gene would not

be ideal for forensic analysis. Overall, the combinations observed within each population, amplified using Primer Mix I, were different among each ethnic group. Thus, the specific combinations observed were representative of a particular ethnicity. The possession of these combinations can be used in the future work for class characteristic determination when further validated in a larger sample set. The second primer mix underwent the same analysis to determine if the D6S2787 and D6S2917 genes could discriminate between ethnicities.

9.5 PRIMER MIX II ANALYSIS

9.2.4.1 PRIMER MIX II INDICATOR SPECIES ANALYSIS

Primer Mix I amplified the following genes: D6S2787 and D6S2917. To further examine whether ethnicity would influence gene presence, a preliminary statistical analysis of the data was conducted. Each gene was identified by its expected size range.

Table 35:Genes amplified using Primer Mix II with expected size range

Gene	Expected Size Range (bp)
D6S2787	134–163
D6S2917	247–361

As illustrated in Figure 62, the variations in gene presence across ethnicities were represented by the bar graph.

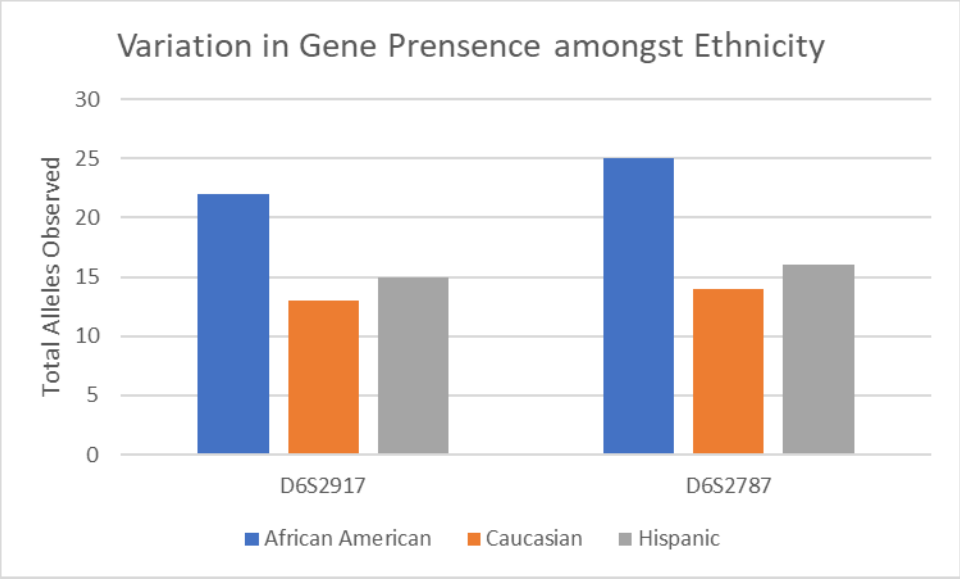


Figure 62:Variation in gene presence among ethnicities

D6S2917 is most prevalent in African Americans, followed by Hispanics and lastly Caucasians. A similar trend is observed in gene D6S2787.The same statistical methods were implemented on the data to determine which alleles were frequently abundant within each population. The data was normalized using LH-PCR, and the frequently abundant alleles were identified using Indicator Species Analysis. To find out whether the relative ratios observed when comparing ethnicity were statistically significant, a one-way ANOVA test was performed on each allele. Statistically significant alleles were also subjected to a Tukey's post hoc test to determine if the differences between the mean values were significant. A summary of the findings from the LH-PCR / Indicator Species analysis is shown in Table 36. The frequently abundant amplicon sizes ranged from136 to 335 base pairs. Nine alleles were identified as strongly associated with a specific ethnic group. Both genes amplified using primer mix II provided useful data for ethnic-specific differentiation. However, the DS62917 gene was the most significant with 78% of

frequently abundant alleles deriving from that region. In this study, gene D6S2917 was more frequently detected than any other gene, confirming previous literature that HLA-B (D6S2787) and HLA-DQB1(D6S2917) loci have greater influence on ancestry and lineage.[71,74] D6S2917 was found to be strongly associated with both Caucasian and African American subjects at several different regions (239 bp, 271 bp, 286 bp, 287 bp, 335 bp, 364 bp, 367 bp).

Table 36:Frequently abundant alleles present within genes D6S2917 and D6S2787 via Indicator Species Analysis

Size	**Gene**	**Ethnicity Association**	***P* Value**
367	D6S2917	Caucasian	0.0152
271	D6S2917	Caucasian	0.0131
364	D6S2917	Caucasian	0.0280
286	D6S2917	African American	0.0060
287	D6S2917	African American	0.0287
136	D6S2787	African American	0.0336
335	D6S2917	African American	0.0103

An ANOVA was used to determine if there were significant differences between ethnic groups in terms of relative abundance. In Table 36, only statistically significant alleles are shown. As seen in the table, the mean values with the same letter in superscript did not differ significantly. Based on the collected data, several quantitative and qualitative observations were made.

Table 37:Tukey's HSD mean comparison among ethnicities using gene D6S2917

Size	Gene	ANOVA	Caucasian	Hispanic	African American
367	D6S2917	$F = 6.68$ $P = 0.0072$	0.0007049225^{a}	0^{b}	0^{b}
271	D6S2917	$F = 5.62$ $P = 0.0134$	0.0003516266^{a}	0^{b}	0^{ab}
364	D6S2917	$F = 4.144$ $P = 0.0342$	0.0009160669^{a}	0.0001474911^{b}	0.0001472000^{b}
286	D6S2917	$F = 7.93$ $P = 0.00389$	0.001315590^{a}	0.003050132^{b}	0.022945513^{b}
287	D6S2917	$F = 4.64$ $P = 0.0246$	0.005385375^{b}	0.009390577^{b}	0.061988773^{a}
136	D6S2787	$F = 3.95$ $P = 0.039$	0.002218232^{ab}	0.001362258^{b}	0.013016488^{a}

9.2.4.2 Gene D6S2917 Allele Frequency Study

Within gene D6S2917 (247 bp–361 bp), a total of 29 alleles were observed in the population sampled. Figure 63 shows the number of alleles observed for each ethnicity. The D6S2917 gene was predominant in the African American population with 22 alleles observed within their population, followed by Hispanics (15 alleles) and lastly Caucasians (7 alleles). Although African Americans had the most variation in alleles, the frequency of each allele within each population was higher in Caucasian and Hispanic subjects. Previous studies highlighted the observation of this trend. The prevalence of HLA-B genes (where D6S2917 is located) varies by ethnicity. Caucasians and Hispanics are more highly represented in this region in comparison to African Americans.[121]

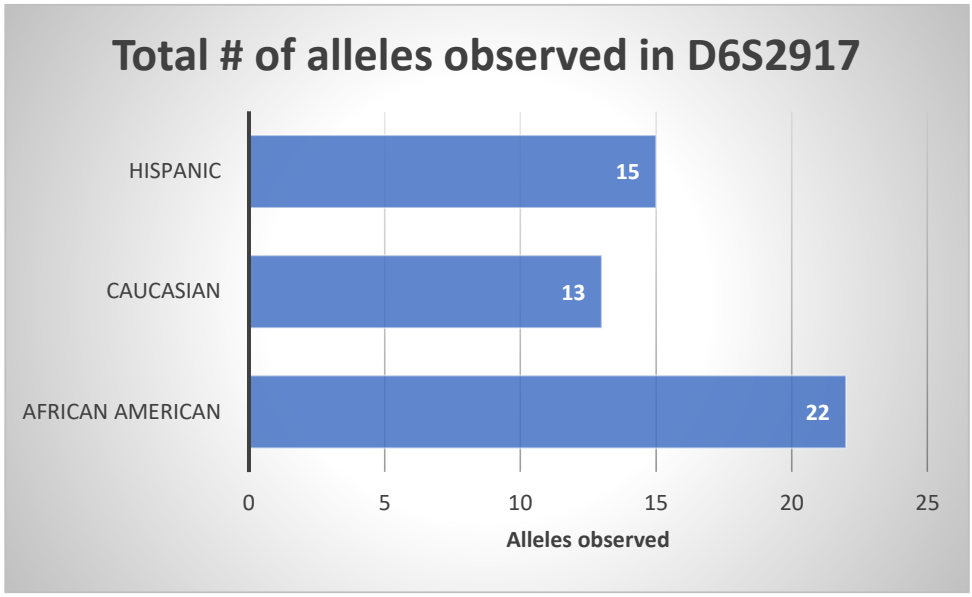


Figure 63: Total number of alleles observed in D6S2917

Among the alleles observed, there was patterns observed within each population. Each population had several regions in which their lack of presence makes it the region of interest when conducting forensic investigations. For example, Allele 302 was found in Caucasian and Hispanic subjects. However, this allele was not found in any Caucasian subjects. Gene D6S217 expressed several alleles linked solely to a specific ethnicity. Caucasians had three alleles unique to their population (Allele 279, Allele 271, Allele 327). The Hispanic population also expressed a unique allele that could potentially be used to classify their ethnicity (Allele 319). African Americans possessed the most alleles unique to their ethnicity (Allele 278, Allele 282, Allele 350, Allele 252). When comparing the results of the indicator species analysis to the allele frequencies observed, there was a correlation with relative abundance and allele frequency for the African American population. For example, Allele 287 demonstrated the highest frequency and relative abundance. The same pattern was observed at Allele 327 for Caucasian subjects. The differences within each observed population validate that if it were further enhanced, gene D6S2917 could serve as a biomarker for ethnicity-related studies. A genotype study was conducted to further assess the reliability of this gene as a biomarker for ethnicity.

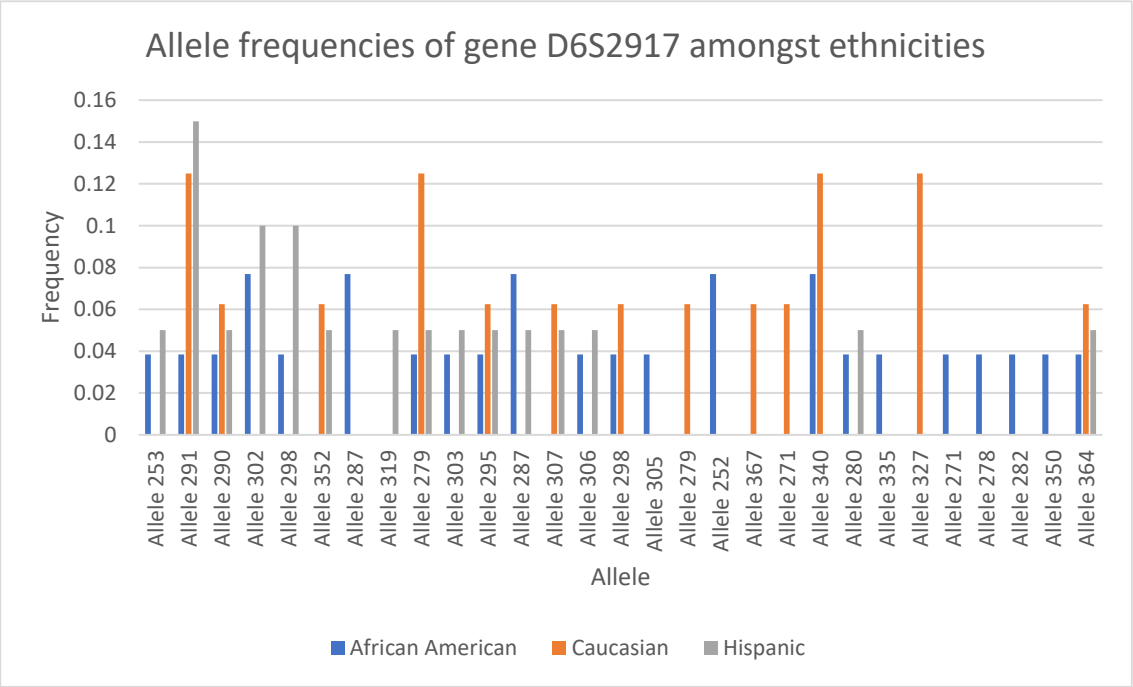


Figure 64: Allele frequencies of gene D6S2917 among ethnicities

9.2.4.3 Gene D6S2917 Genotype Frequency Study

The 26 genotype combinations observed within the sample set were further assessed to determine if the combination occurred frequently among any specific population. The zygosity of all subjects was observed. Of the 31 subjects, 30 subjects were heterozygous and only one subject was homozygous (302,302). Three allele combinations (252,287/278,302/290, 290/291) were present within multiple samples at a different frequency (see Appendix 8). Two genotypes were found exclusively in the African American population (252,287/ 278/302), appearing in 15% of the African American subjects sampled. One genotype was found in 20% of Hispanic subjects sampled (290,291). The recurrence of these alleles within specific ethnic groups with no overlap strongly indicate their reliability as a biomarker for class-characteristic determination of

individuals. A linear discriminant analysis was computed to further validate this claim. With the use of LDA, all 31 subjects were correctly classified at 90.32% accuracy. The prediction rate of the model is illustrated in Table 38.

Table 38: Predicted rate of identifying ethnicity based on genotype frequency using gene D6S2917

Actual	Predicted Rate		
Ethnicity	**African American**	**Caucasian**	**Hispanic**
African American	0.923	0.000	0.077
Caucasian	0.000	0.875	0.125
Hispanic	0.000	0.100	0.900

The true positive rate ranged from 87.5% to 92.30%. The false positive rate ranged from 0% to 13%. The high true positive and low false positive rates support previous findings that report a sensitivity greater than 90% when using HLA-B genes as a biomarker for disease.[122] The 9.68% misclassification rate resulted primarily from incorrect classifications of Caucasians and Hispanics. When evaluating the discriminant scores, a Caucasian subject was incorrectly identified as Hispanic. A Hispanic individual was also incorrectly identified as Caucasian. This finding highlights the misinterpretation challenges that occur when subjects self-identify their ethnicity. Based on the canonical plots and stepwise regression analysis, Table 39 shows the genotype frequencies selected as most influential for successful ethnicity prediction. Their influence was determined by assessing their canon 1 and canon 2 values. The higher canon 1, the greater the influence.

Table 39: Most influential genotype for ethnicity differentiation of individuals using Gene D6S2917

Genotype	Canon 1	Canon 2
280,335	4.176087	0.560593
271,303	4.176087	0.560593
306,340	4.176087	0.560593
287,340	4.176087	0.560593
287,302	4.176087	0.560593

9.2.4.4 Gene D6S2787 Allele Frequency Study

Within gene D6S2787 (134 bp–163 bp), a total of 7 alleles were observed within the population sampled. Caucasians expressed the most alleles (7), followed by African Americans (4) and lastly Hispanics (3). The gene expressed several notable patterns observed. The Caucasian population was represented in every allele expressed within this gene. Furthermore, the Caucasian population also had three alleles that were unique to their ethnicity (Allele 137, Allele 149, Allele 152). As seen in the Indicator Species Analysis, Allele 136 was frequently abundant in African American subjects. The LHPC ratio observed for this allele was directly correlated to the genotype frequency observed. Allele 138 showed a high frequency for all ethnicities. The expression of this allele in all populations signifies challenges with prediction of ethnicity using this allele alone. However, the variation in genotype frequencies may highlight differences between individuals. On the other hand, Allele 140 shows great potential for ethnic-specific differentiation based on the presence of this allele at different frequencies. The genotype frequencies of each subject were further observed to determine whether genotype combinations could reliably be used to differentiate among ethnicities.

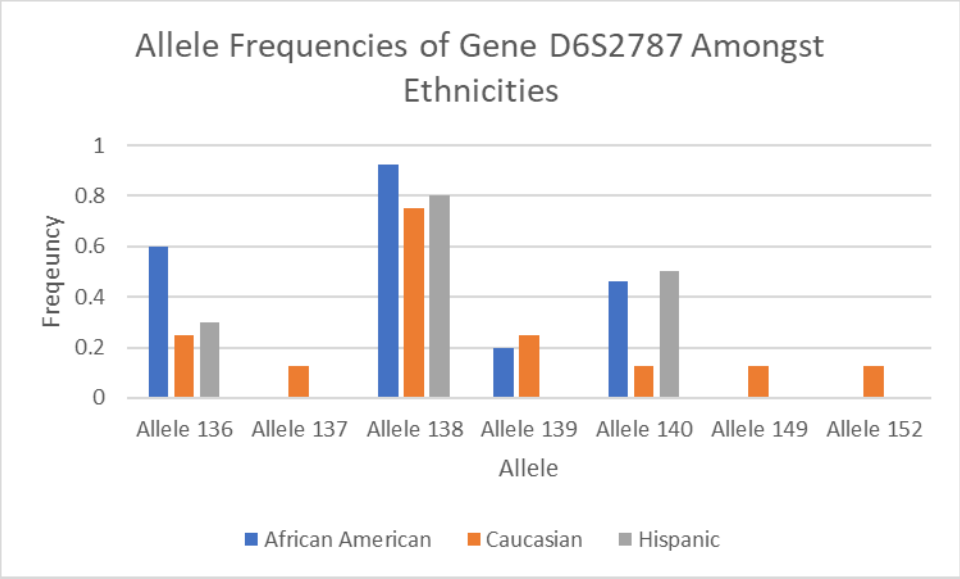


Figure 65:Allele Frequencies of gene 2787 among ethnicities

9.2.4.5 Gene D6S2787 Genotype Frequency Study

The genotype combinations of each subject were examined to identify trends present within the populations. Combination 138, 140 was present in approximately 50% of the populations sampled. This genotype was the most frequent in Caucasians (62.5%), followed by African Americans (30.8%); it was least observed in Hispanics (30%). Genotype 136,138 was present within both Hispanic and African American subjects. The overlap in genotype expression at similar frequency in Hispanic and African American subjects foreshadow a challenge with deciphering between the two populations. LDAS was implemented on the dataset to identify the predictive ability of the alleles expressed within the D6S2925 gene. With the use of LDA, subjects were accurately classified at 80% and misclassified at 20%. The prediction rate of the model is illustrated in Table 39.

Table 40:Prediction rate of identifying ethnicity based of genotype frequency using gene D6S2787

Actual Ethnicity	Predicted Rate African American	Caucasian	Hispanic
African American	0.769	0.000	0.231
Caucasian	0.000	1.000	0.000
Hispanic	0.300	0.000	0.700

The true positive rate for both the African American population and Hispanic population were relatively low (70–77%). This is due to the similarities observed within both populations in both the allele frequency study and genotype study. These similarities accounted for all misclassifications observed in the analysis. Therefore, this could be useful for the classification of African American and Hispanic subjects. However, supplemental data would be needed to further validate the characterization of an individual. By contrast, there was a 100% true positive rate for Caucasian subjects, which highlights the capability of Gene 2787 to predict Caucasian ethnicity. The false positive rate ranged from 0% to 30%. The most influential genotypes for the accurate predictions observed were genotype 137,139 and genotype 150,152. Both genotypes were present in only Caucasian subjects at a low frequency (0.125). This further validates allele 137 and allele 152's usage for exclusion of the Caucasian population. Overall, Gene D6S2787 provided useful information that can be used to assist with ethnic specific observations.

Table 41:Most influential genotypes for ethnicity differentiation of individuals using gene D6S2787

Genotype	Canon 1	Canon 2
137,139	3.212617	2.111731
150,152	3.212617	2.111731
138,140	1.626492	-0.369995

138,140	1.626492	-0.369995
136,138	0.995345	-0.31572

10.HLA GENE STUDY CONCLUSIONS

Because the HLA complex is a highly polymorphic region, the genetic variations that contribute to ethnic specific differences were observed. These variations were evaluated for statistical significance using several statistical methods for validation. The novel approach of utilizing LHPCR coupled to Indicator Species Analysis for the HLA region successfully highlighted alleles that were frequently abundant in almost each population. This robust method, typically used in microbial community analysis, can identify amplicons in humans that originate from different ethnicities based on their natural variations prior to the traditional STR genotyping analysis. The results found in this study confirm that individuals have a distinct HLA region that is composed of regions that express ethnic specific combinations that could potentially be used for class-characteristic determination when a larger sample size is evaluated. The study assessed the ability of the HLA gene complex to differentiate among three large populations (Caucasian, Hispanics, African Americans). The study included 31 subjects who were separated into groups according to their self-identified ethnicity. Using five genes, 20 alleles were identified as unique contributors to the population due to their ability to predict ethnicity with 80% or greater accuracy. Of these ten alleles, 10 were detected using Primer Mix I and 10 alleles were detected using Primer Mix II. The significance of utilizing both primer mixes to amplify specific regions is reflected in the data observed. Multiplex PCR allowed for the simultaneous investigation of several regions that facilitated the

identification of alleles unique to each population. However, this study revealed that genes located in the HLA-DPB1 (M138) was not as relevant in ethnic specific studies. The findings of this study are in direct alignment with previous observations of the HLA region. The genotype study of frequent combinations revealed that African Americans have the lowest genetic diversity, which is why the chance of seeing two African Americans that share the same genotype is lower than with Caucasians.[118] These finding also validate the claim in the literature that the HLA-B (D6S2925) and HLA-DQRB1(D6S2917) loci have greater influence when assessing ancestry and lineage.[72] In accordance with Nelson v. State: "Without the necessary statistical calculations, the evidence is deemed meaningless to the jury and thus, inadmissible." In the absence of an expected value that could be used to calculate Hardy Weinberg Equilibrium, Linear Discriminant Analysis serves as an efficient statistical method to predict ethnicity with great sensitivity and specificity. In future studies, a larger sample set should be used to further evaluate the exclusion rates and likelihood ratios of these five genes for forensic investigations.

11.VOC AND HLA CORRELATION STUDY

A correlation analysis was implemented on JMP® Student to determine if there was a correlation between the most influential loci and VOCs detected in this study. A total of 18 subjects who completed both VOC sampling and buccal swab sampling were utilized in this study. The demographics are shown in Table 42.

Table 42:Demographics of VOC and HLA correlation study

Ethnicity	Number of Participants
Caucasian	7
Hispanic	5
African American	6

The Kendall Tau Correlation Test was used for the dataset to further assess the relationship between VOCs and genotype frequency. In comparison to the most used correlation testing (Spearman's rank), Kendall Tau was selected due to its more precise *p* value for small sample size.[123] The study revealed that there was minimal correlation between the VOCs and genotype combinations. However, of the nine genotypes, one genotype showed a correlation to VOCs (142,173). This genotype was found within all three populations (Caucasian, Hispanic, and African American). A summary of the correlation is presented in Table 43.

Table 43:VOC and HLA genotype correlations with corresponding p *values*

Genotype	VOCs Correlated to Genotype	*P* Value
142,173	Tetradecane	< 0.05
142,173	Propanoic acid, 2-methyl	< 0.05
142,173	Butanoic Acid	< 0.05
142,173	2-Tridecanone	< 0.05
142,173	1- Octadecane	< 0.05

142,173	Dodecanoic Acid	< 0.05

The negative correlation of -1.0 shown in navy blue in Figure 66 between genotype 142, 173 and the five compounds (Propanoic acid, 2-methyl, Butanoic acid, 2-tridecanone, 1-Octadecane, Dodecanoic acid) showed that an increase in the amount of tetradecane present within a subject led to a decrease in the genotype frequency observed.

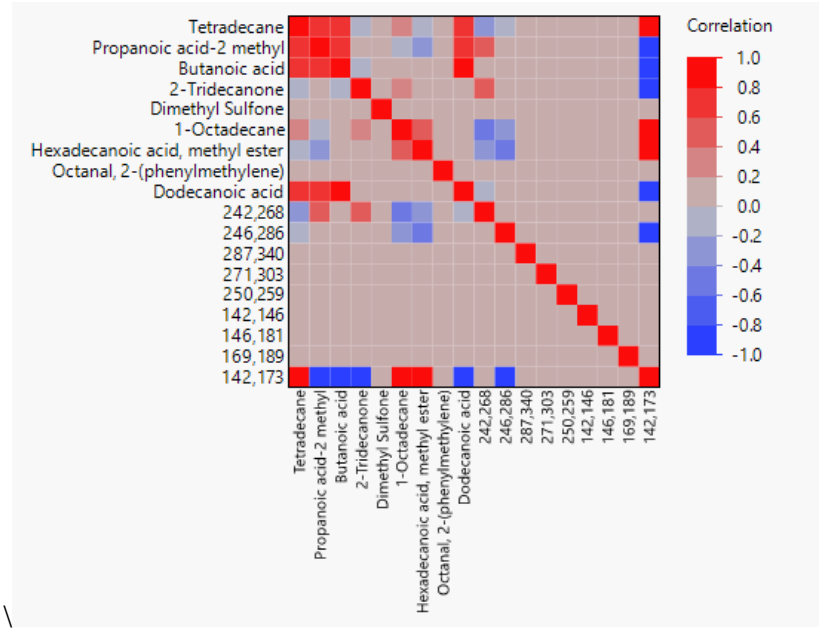


Figure 66:Heat map of correlation between VOCs and HLA genotypes

On the other hand, tetradecane displayed a 1.0 correlation, which means that when the amount of tetradecane increased, an increase in the genotype frequency of 142,173 also occurred. Previous studies have indicated that distinctive odor signatures are created by expression of individual alleles.[30] The findings of this study support those findings and show that genotype combinations expressed also have an influence on odor profiles.

Linear Discriminant Analysis was employed to determine if using the most influential VOCs in odor profiles and genotypes present within ethnicity can effectively predict ethnicity with higher performance than individually assessing the two sample sets. The test revealed that coupling the VOC findings to the HLA allowed for the investigation of all three ethnicities simultaneously. Furthermore, the accuracy of the test was 94.4%. The true positive rate ranged from 80–100%, with two of three ethnicities (African American and Caucasian) having a perfect true positive rate. The false positive rate ranged from 0% to 20%. The LDA performed showed the highest sensitivity and specificity in comparison to the isolated VOC and HLA studies. This highlights the advantage of utilizing both VOC and HLA data to class characterize individuals successfully. Future studies that would complement the findings should investigate the VOC profiles of each buccal swab prior to gDNA extraction. The trends in VOC prevalence within the saliva can be compared to the HLA loci to determine if the correlation between VOCs and genotype can be further enhanced.

Table 44:Predicted rate of identifying ethnicities based on genotype frequency using VOCs and HLA

Actual	**Predicted Rate**		
Ethnicity	**African American**	**Caucasian**	**Hispanic**
African American	1.000	0.000	0.000
Caucasian	0.000	1.000	0.000
Hispanic	0.200	0.000	0.800

12. FINAL CONCLUSIONS

When a criminal investigation is ongoing, evidence collection is pivotal in linking a potential suspect to a crime scene. When physical evidence is not readily

available, trace evidence of odorous compounds that emit from humans can be utilized to provide investigative leads. Human scent evidence has made significant advancements in the last decade. The novel incorporation of using two extraction methods (HS-SPME/ solvent extraction) has added value to the discrimination power of human scent evidence. The statistical methods implemented in this dissertation show that individuals can be characterized by ethnicity based on their odor profile by using nine VOCs. The narrowing of the number of VOCs significant for ethic specific observations will assist in simplifying data analysis for future human scent studies. Furthermore, the utilization of previously optimized HS-SPME methods have confirmed that the acquisition of human scent evidence has been effectively standardized and facilitates the production of reliable data that can be used to discriminate between individuals.

Furthermore, this dissertation validates why genomic DNA is the golden standard in forensic science. The HLA study serves as the foundation for demonstrating the reliability of the HLA region, not only from an immunological and ancestral standpoint but also for valuable forensic observations. This dissertation also highlights the cross-disciplinary application of a statistical method typically used in microbiology for preliminary screening of frequently abundant alleles. Four of the five genes amplified via Multiplex PCR predicted ethnicity with 80% or greater accuracy which was significantly higher and more consistent than VOCs which ranged from 69% to 100% accuracy. Further validation testing and statistical measures should be pursued in future work to assess the capability of the HLA region to differentiate among ethnicities by addressing the limitations of the study (i.e., the sample size).

Coupling VOCs to HLA genes successfully provided a full profile of human subjects. Although previous studies have emphasized how HLA influences odor, prior to this research, there was minimal understanding of the correlation between the VOCs emitted from humans and their genotype. This study serves as a foundation for the utilization of HLA genes for the class-characteristic determination of individuals. This dissertation confirms that the HLA plays a role in the VOCs emitted. However, it is important to note that correlation is not causation.[124] Therefore, future research should investigate the metabolic contribution of VOCs to determine the cause of quantitative differences in primary odor observed within ethnicities. The statistical methods implemented to analyze both VOCs and HLA genes show how machine-learning techniques (ROC curves and Linear Discriminant analysis) can be used to predict ethnicity based on their odor profiles and genotypes. Both data can be individually or synergistically used to make inferences regarding ethnicity. The increased sensitivity and specificity of the prediction rate when using both VOC and HLA genotype frequencies highlights the value each type of biological specimen brings to this research.

Overall, this research has exhibited how the use of both HLA genes and VOCs provide forensic analysts with two reliable forms of evidence that are left at every crime scene. The analytical techniques and statistical methods provided robust and clear results that could be expanded upon for the development of a future database of VOCs and HLA genes useful for the class-characteristic determination of individuals.

REFERENCES

(1) Loeve, A. J.; Bilo, R. A. C.; Emirdag, E.; Sharify, M.; Jansen, F. W.; Dankelman, J. In Vitro Validation of Vaginal Sampling in Rape Victims: The Problem of Locard's Principle. *Forensic Sci. Med. Pathol.* 2013, *9* (2), 154–162. https://doi.org/10.1007/s12024-013-9426-6.

(2) Weissman, S. H.; Busch, K. G.; Schouten, R. The Evolution of Terrorism from 1914 to 2014. *Behav. Sci. Law* 2010, *28* (2), 211–223. https://doi.org/10.1002/bsl.

(3) Peterson, J.; Sommers, I.; Baskin, D.; Johnson, D.; Senior, C.-P.; Associates, R. The Role and Impact of Forensic Evidence in the Criminal Justice Process. *J. Forensic Sci.* 2010, *58*, S78–S90.

(4) Balemba, S.; Beauregard, E.; Martineau, M. Getting Away with Murder: A Thematic Approach to Solved and Unsolved Sexual Homicides Using Crime Scene Factors. *Police Pract. Res.* 2014, *15* (3), 221–233. https://doi.org/10.1080/15614263.2013.846548.

(5) Bragazzi, N. L.; Sellami, M.; Salem, I.; Conic, R.; Kimak, M.; Pigatto, P. D. M.; Damiani, G. Fasting and Its Impact on Skin Anatomy, Physiology, and Physiopathology: A Comprehensive Review of the Literature. *Nutrients* 2019, *11* (2). https://doi.org/10.3390/nu11020249.

(6) Dehdashtian, A.; Stringer, T. P.; Warren, A. J.; Mu, E. W.; Amirlak, B.; Shahabi, L. Anatomy and Physiology of the Skin. *Melanoma A Mod. Multidiscip. Approach* 2018, 15–26. https://doi.org/10.1007/978-3-319-78310-9_2.

(7) Gallo, R. L. Human Skin Is the Largest Epithelial Surface for Interaction with Microbes. *J. Invest. Dermatol.* 2017, *137* (6), 1213–1214. https://doi.org/10.1016/j.jid.2016.11.045.

(8) Wilke, K.; Martin, A.; Terstegen, L.; Biel, S. S. A Short History of Sweat Gland Biology. *Int. J. Cosmet. Sci.* 2007, *29* (3), 169–179. https://doi.org/10.1111/j.1467-2494.2007.00387.x.

(9) McGibbon, D. Rook's Textbook of Dermatology, 7th Edition. *Clin. Exp. Dermatol.* 2006, *31* (1), 178–179. https://doi.org/10.1111/j.1365-2230.2005.02034.x.

(10) Troccaz, M.; Starkenmann, C.; Niclass, Y.; Van De Waal, M.; Clark, A. J. 3-Methyl-3-Sulfanylhexan-1-Ol As a Major Descriptor for the Human Axilla-Sweat Odour Profile. *Chem. Biodivers.* 2004, *1* (7), 1022–1035. https://doi.org/10.1002/cbdv.200490077.

(11) Saga, K. Structure and Function of Human Sweat Glands Studied with Histochemistry and Cytochemistry. *Prog. Histochem. Cytochem.* 2002, *37* (4), 323–386. https://doi.org/10.1016/S0079-6336(02)80005-5.

(12) Dormont, L.; Bessière, J. M.; Cohuet, A. Human Skin Volatiles: A Review. *J. Chem. Ecol.* 2013, *39* (5), 569–578. https://doi.org/10.1007/s10886-013-0286-z.

(13) Zhang, Z.; Yang, M. J.; Pawliszyn, J. Solid-Phase Microextraction: A Solvent-Free Alternative for Sample Preparation. *Anal. Chem.* 1994, *66* (17), 844A-853A. https://doi.org/10.1021/ac00089a001.

(14) Shelly, W. B.; Hurley, H. J. The Physiology of the Human Axillary Apocrine Sweat Gland. *J. Invest. Dermatol.* 1953, *20* (4), 285–297. https://doi.org/10.1038/jid.1953.35.

(15) James, A. G.; Austin, C. J.; Cox, D. S.; Taylor, D.; Calvert, R. Microbiological and Biochemical Origins of Human Axillary Odour. *FEMS Microbiol. Ecol.* 2013, *83* (3), 527–540. https://doi.org/10.1111/1574-6941.12054.

(16) Curran, A. M.; Rabin, S. I.; Prada, P. A.; Furton, K. G. Comparison of the Volatile Organic Compounds Present in Human Odor Using SPME-GC/MS. *J. Chem. Ecol.* 2005, *31* (7), 1607–1619. https://doi.org/10.1007/s10886-005-5801-4.

(17) Gallagher, M.; Wysocki, C. J.; Leyden, J. J.; Spielman, A. I.; Sun, X.; Preti, G. Analyses of Volatile Organic Compounds from Human Skin. *Br. J. Dermatol.* 2008, *159* (4), 780–791. https://doi.org/10.1111/j.1365-2133.2008.08748.x.

(18) Filipiak, W.; Mochalski, P.; Filipiak, A.; Ager, C.; Cumeras, R.; E. Davis, C.; Agapiou, A.; Unterkofler, K.; Troppmair, J. A Compendium of Volatile Organic

Compounds (VOCs) Released By Human Cell Lines. *Curr. Med. Chem.* 2016, *23* (20), 2112–2131. https://doi.org/10.2174/0929867323666160510122913.

(19) Schiller, J. J.; Hopp, K. A.; Pietz, B. C.; Bick, D. P.; Lau, E. C.; Ellis, T. M. A Simplified Method for Screening Siblings for HLA Identity Using Short Tandem Repeat (STR) Polymorphisms. *Hum. Immunol.* 2013, *74* (5), 562–566. https://doi.org/10.1016/j.humimm.2012.12.015.

(20) Cuzuel, V.; Cognon, G.; Rivals, I.; Sauleau, C.; Heulard, F.; Thiébaut, D.; Vial, J. Origin, Analytical Characterization, and Use of Human Odor in Forensics. *J. Forensic Sci.* 2017, *62* (2), 330–350. https://doi.org/10.1111/1556-4029.13394.

(21) T, G. P.; Zeng, C.; Spielmantt, A. I.; Vowels, B. R.; Leyden, J. J.; Bie, K. A Human Axillary Odorant Is Carried by Apolipoprotein D Author (s): Chenhui Zeng , Andrew I . Spielman , Benjamin R . Vowels , James J . Leyden , Klaus Biemann and George Preti Source : Proceedings of the National Academy of Sciences of the United State. 1996, *93* (13), 6626–6630.

(22) Prokop-Prigge, K. A.; Greene, K.; Varallo, L.; Wysocki, C. J.; Preti, G. The Effect of Ethnicity on Human Axillary Odorant Production. *J. Chem. Ecol.* 2016, *42* (1), 33–39. https://doi.org/10.1007/s10886-015-0657-8.

(23) Barzantny, H.; Schröder, J.; Strotmeier, J.; Fredrich, E.; Brune, I.; Tauch, A. The Transcriptional Regulatory Network of Corynebacterium Jeikeium K411 and Its Interaction with Metabolic Routes Contributing to Human Body Odor Formation. *J. Biotechnol.* 2012, *159* (3), 235–248. https://doi.org/10.1016/j.jbiotec.2012.01.021.

(24) James, A. G.; Hyliands, D.; Johnston, H. Generation of Volatile Fatty Acids by Axillary Bacteria. *Int. J. Cosmet. Sci.* 2004, *26* (3), 149–156. https://doi.org/10.1111/j.1467-2494.2004.00214.x.

(25) Starkenmann, C.; Mayenzet, F.; Brauchli, R.; Troccaz, M. 5 a -Androst-16-En-3 a -Ol b - d -Glucuronide , Precursor of 5 a -Androst-16-En-3 a - Ol in Human Sweat. 2013, *10*, 2197–2208.

(26) Troccaz, M.; Borchard, G.; Vuilleumier, C.; Raviot-Derrien, S.; Niclass, Y.; Beccucci, S.; Starkenmann, C. Gender-Specific Differences between the Concentrations of Nonvolatile (R)/(S)-3-Methyl-3-Sulfanylhexan-1-Ol and

(R)/(S)-3-Hydroxy-3-Methyl-Hexanoic Acid Odor Precursors in Axillary Secretions. *Chem. Senses* 2009, *34* (3), 203–210. https://doi.org/10.1093/chemse/bjn076.

(27) Brown, J. S.; Prada, P. A.; Curran, A. M.; Furton, K. G. Applicability of Emanating Volatile Organic Compounds from Various Forensic Specimens for Individual Differentiation. *Forensic Sci. Int.* 2013, *226* (1–3), 173–182. https://doi.org/10.1016/j.forsciint.2013.01.008.

(28) Yoshiura, K. I.; Kinoshita, A.; Ishida, T.; Ninokata, A.; Ishikawa, T.; Kaname, T.; Bannai, M.; Tokunaga, K. A SNP in the ABCC11 Gene Is the Determinant of Human Earwax Type. *Nat. Genet.* 2006, *38* (3), 324–330. https://doi.org/10.1038/ng1733.

(29) Thornhill, R.; Gangestad, S. W.; Miller, R.; Scheyd, G.; McCollough, J. K.; Franklin, M. Major Histocompatibility Complex Genes, Symmetry, and Body Scent Attractiveness in Men and Women. *Behav. Ecol.* 2003, *14* (5), 668–678. https://doi.org/10.1093/beheco/arg043.

(30) Aksenov, A. A.; Gojova, A.; Zhao, W.; Morgan, J. T.; Sankaran, S.; Sandrock, C. E.; Davis, C. E. Characterization of Volatile Organic Compounds in Human Leukocyte Antigen Heterologous Expression Systems: A Cell's "Chemical Odor Fingerprint." *ChemBioChem* 2012, *13* (7), 1053–1059. https://doi.org/10.1002/cbic.201200011.

(31) Laird, R.; Dawkins, R. L.; Gaudieri, S. Use of the Genomic Matching Technique to Complement Multiplex STR Profiling Reduces DNA Profiling Costs in High Volume Crimes and Intelligence Led Screens. *Forensic Sci. Int.* 2005, *151* (2–3), 249–257. https://doi.org/10.1016/j.forsciint.2005.02.018.

(32) Kazantzi, V.; Kabir, A.; Furton, K. G.; Anthemidis, A. Fabric Fiber Sorbent Extraction for On-Line Toxic Metal Determination by Atomic Absorption Spectrometry: Determination of Lead and Cadmium in Energy and Soft Drinks. *Microchem. J.* 2018, *137*, 285–291. https://doi.org/10.1016/j.microc.2017.11.006.

(33) Vas, G.; Vékey, K. Solid-Phase Microextraction: A Powerful Sample Preparation Tool Prior to Mass Spectrometric Analysis. *J. Mass Spectrom.* 2004, *39* (3), 233–254. https://doi.org/10.1002/jms.606.

(34) Bojko, B.; Cudjoe, E.; Gómez-Ríos, G. A.; Gorynski, K.; Jiang, R.; Reyes-Garcés, N.; Risticevic, S.; Silva, É. A. S.; Togunde, O.; Vuckovic, D.; Pawliszyn, J. SPME - Quo Vadis? *Anal. Chim. Acta* 2012, *750*, 132–151. https://doi.org/10.1016/j.aca.2012.06.052.

(35) PAWLISZYN, J.; PAWLISZYN, B.; PAWLISZYN, M. Solid Phase Microextraction (SPME). *Chem. Educ.* 1997, *2* (4), 1–7. https://doi.org/10.1007/s00897970137a.

(36) Zhang, L.; Xiao, H.; Zhou, H.; Santiago, S.; Lee, J. M.; Garon, E. B.; Yang, J.; Brinkmann, O.; Yan, X.; Akin, D.; Chia, D.; Elashoff, D.; Park, N. H.; Wong, D. T. W. Development of Transcriptomic Biomarker Signature in Human Saliva to Detect Lung Cancer. *Cell. Mol. Life Sci.* 2012, *69* (19), 3341–3350. https://doi.org/10.1007/s00018-012-1027-0.

(37) Colón-Crespo, L. J.; Herrera-Hernández, D.; Holness, H.; Furton, K. G. Determination of VOC Marker Combinations for the Classification of Individuals by Gender and Race/Ethnicity. *Forensic Sci. Int.* 2017, *270*, 193–199. https://doi.org/10.1016/j.forsciint.2016.09.011.

(38) Schmidt, K.; Podmore, I. Current Challenges in Volatile Organic Compounds Analysis as Potential Biomarkers of Cancer. *J. Biomarkers* 2015, *2015* (April), 1–16. https://doi.org/10.1155/2015/981458.

(39) Atanassova, M. Solvent Extraction Of Metallic Species In Ionic Liquids: An Overview Of S-, P- And D-Elements. *J. Chem. Technol. Metall.* 2021, *56* (3), 443–466.

(40) Mcnair, Harold M., James M. Miller, and N. H. S. *Basic GasBasic Gas Chromatography*, 2019th ed.; John Wiley & Sons.

(41) Kolomnikov, I. G.; Efremov, A. M.; Tikhomirova, T. I.; Sorokina, N. M.; Zolotov, Y. A. Early Stages in the History of Gas Chromatography. *J. Chromatogr. A* 2018, *1537*, 109–117. https://doi.org/10.1016/j.chroma.2018.01.006.

(42) Sparkman, O. David, et al. *Gas Chromatography and Mass Spectrometry: A Practical Guide*, 2011th ed.; Technologoy, E. S., Ed.; ProQuest Ebook Central, 2011.

(43) Mellon, F. A. MASS SPECTROMETRY | Principles and Instrumentation. In *Encyclopedia of Food Sciences and Nutrition*; Elsevier, 2003; pp 3739–3749. https://doi.org/10.1016/b0-12-227055-x/00746-x.

(44) Essentials, T. H. E. The Essential Guide to Electron Ionization in GC – MS. 2019, No. April.

(45) Shmueli, G. To Explain or to Predict? *Stat. Sci.* 2010, *25* (3), 289–310. https://doi.org/10.1214/10-STS330.

(46) Sperandei, S. Understanding Logistic Regression Analysis. *Biochem. Medica* 2014, *24* (1), 12–18. https://doi.org/10.11613/BM.2014.003.

(47) Mary L. McHugh. The Odds Ratio: Calculation, Usage, and Interpretation. *Biochem. Medica* 2009, *19* (2), 120–126.

(48) Kumar, R.; Indrayan, A. Kumar-2011-Receiver Operating Characteristic 1. *Indian Pediatr.* 2011, *48*, 277–287.

(49) Lever, J.; Krzywinski, M.; Altman, N. Points of Significance: Principal Component Analysis. *Nat. Methods* 2017, *14* (7), 641–642. https://doi.org/10.1038/nmeth.4346.

(50) Appoloni, C. R.; Melquiades, F. L. Portable XRF and Principal Component Analysis for Bill Characterization in Forensic Science. *Appl. Radiat. Isot.* 2014, *85*, 92–95. https://doi.org/10.1016/j.apradiso.2013.12.004.

(51) Gál, L.; Oravec, M.; Gemeiner, P.; Čeppan, M. Principal Component Analysis for the Forensic Discrimination of Black Inkjet Inks Based on the Vis-NIR Fibre Optics Reflection Spectra. *Forensic Sci. Int.* 2015, *257*, 285–292. https://doi.org/10.1016/j.forsciint.2015.09.011.

(52) Senior, S.; Hamed, E.; Masoud, M.; Shehata, E. Characterization and Dating of Blue Ballpoint Pen Inks Using Principal Component Analysis of UV-Vis Absorption Spectra, IR Spectroscopy, and HPTLC. *J. Forensic Sci.* 2012, *57* (4), 1087–1093. https://doi.org/10.1111/j.1556-4029.2012.02091.x.

(53) Zhu, M.; Ghodsi, A. Automatic Dimensionality Selection from the Scree Plot via the Use of Profile Likelihood. *Comput. Stat. Data Anal.* 2006, *51* (2), 918–930. https://doi.org/10.1016/j.csda.2005.09.010.

(54) Mishra, P.; Singh, U.; Pandey, C. M. Application of Student ' s t - Test , Analysis of Variance , and Covariance. 2019, 407–411. https://doi.org/10.4103/aca.ACA.

(55) Alves-Leon, S. V.; Papais-Alvarenga, R.; Magalhães, M.; Alvarenga, M.; Thuler, L. C. S.; Fernández Y Fernandez, O. Ethnicity-Dependent Association of HLA DRB1-DQA1-DQB1 Alleles in Brazilian Multiple Sclerosis Patients. *Acta Neurol. Scand.* 2007, *115* (5), 306–311. https://doi.org/10.1111/j.1600-0404.2006.00750.x.

(56) Scott, A. J.; Hosmer, D. W.; Lemeshow, S. Applied Logistic Regression. *Biometrics* 1991, *47* (4), 1632. https://doi.org/10.2307/2532419.

(57) Logan, J. R. How Race Counts for Hispanic Americans. *SAGE Race Relations Abstr.* 2004, *29* (1), 7–19. https://doi.org/10.1177/0307920104040799.

(58) Raman, M.; Ahmed, I.; Gillevet, P. M.; Probert, C. S.; Ratcliffe, N. M.; Smith, S.; Greenwood, R.; Sikaroodi, M.; Lam, V.; Crotty, P.; Bailey, J.; Myers, R. P.; Rioux, K. P. Fecal Microbiome and Volatile Organic Compound Metabolome in Obese Humans with Nonalcoholic Fatty Liver Disease. *Clin. Gastroenterol. Hepatol.* 2013, *11* (7), 868-875.e3. https://doi.org/10.1016/j.cgh.2013.02.015.

(59) Torosin, N. S.; Ward, J.; Bell, A. V.; Knapp, L. A. Genetic Similarity and Facial Cues for Kin Recognition in Humans. *bioRxiv* 2020, 1–12. https://doi.org/10.1101/2020.04.06.028134.

(60) B., J.; L., T.; Benson, M.; Bray-Ward, P.; R., A. Immune Dysfunction in Autism Spectrum Disorder. *Autism - A Neurodev. Journey from Genes to Behav.* 2011, No. August 2011. https://doi.org/10.5772/22318.

(61) Travers, A.; Muskhelishvili, G. DNA Structure and Function. *FEBS J.* 2015, *282* (12), 2279–2295. https://doi.org/10.1111/febs.13307.

(62) Chaabani, H. The Miraculous Dynamism of the Qur'an: An Example of a Modern Reading Reveals a DNA Designation. *Int. J. Mod. Anthropol.* 2019, *2* (12), 15. https://doi.org/10.4314/ijma.v12i2.1.

(63) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Clin. Orthop. Relat. Res.* 2007, *462* (1), 3–5. https://doi.org/10.1097/BLO.0b013e31814b9304.

(64) Hardison, R. Book: {Working} with {Molecular} {Genetics} ({Hardison}). *Biol. Libr.* 2013.

(65) Minchin, S.; Lodge, J. Understanding Biochemistry: Structure and Function of Nucleic Acids. *Essays Biochem.* 2019, *63* (4), 433–456. https://doi.org/10.1042/EBC20180038.

(66) Press, N. A. *DNA Technology in Forensic Science*; 1992.

(67) Carosella, E. D. From MAC to HLA: Professor Jean Dausset, the Pioneer. *Hum. Immunol.* 2009, *70* (9), 661–662. https://doi.org/10.1016/j.humimm.2009.07.010.

(68) Ferreiro-Iglesias, A.; Lesseur, C.; McKay, J.; Hung, R. J. Fine Mapping of MHC Region in Lung Cancer Highlights Independent Susceptibility Loci by Ethnicity. *Nat. Commun.* 2018, *9* (1), 1–12. https://doi.org/10.1038/s41467-018-05890-2.

(69) Xie, M.; Li, J.; Jiang, T. Accurate HLA Type Inference Using a Weighted Similarity Graph. *BMC Bioinformatics* 2010, *11* (SUPPL. 11), S10. https://doi.org/10.1186/1471-2105-11-S11-S10.

(70) Lehmann, D. J.; Wiebusch, H.; Marshall, S. E.; Johnston, C.; Warden, D. R.; Morgan, K.; Schappert, K.; Poirier, J.; Xuereb, J.; Kalsheker, N.; Welsh, K. I.; Smith, A. D. HLA Class I, II & III Genes in Confirmed Late-Onset Alzheimer's Disease. *Neurobiol. Aging* 2001, *22* (1), 71–77. https://doi.org/10.1016/S0197-4580(00)00180-9.

(71) Wang, J. H.; Zheng, X.; Ke, X.; Dorak, M. T.; Shen, J.; Boodram, B.; O'Gorman, M.; Beaman, K.; Cotler, S. J.; Hershow, R.; Rong, L. Ethnic and Geographical Differences in HLA Associations with the Outcome of Hepatitis C Virus Infection. *Virol. J.* 2009, *6*, 1–8. https://doi.org/10.1186/1743-422X-6-46.

(72) Nakaoka, H.; Mitsunaga, S.; Hosomichi, K.; Shyh-Yuh, L.; Sawamoto, T.; Fujiwara, T.; Tsutsui, N.; Suematsu, K.; Shinagawa, A.; Inoko, H.; Inoue, I. Detection of Ancestry Informative HLA Alleles Confirms the Admixed Origins of Japanese Population. *PLoS One* 2013, *8* (4). https://doi.org/10.1371/journal.pone.0060793.

(73) Chen, N.; Juric, I.; Cosgrove, E. J.; Bowman, R.; Fitzpatrick, J. W.; Schoech, S. J.; Clark, A. G.; Coop, G. Allele Frequency Dynamics in a Pedigreed Natural Population. *Proc. Natl. Acad. Sci. U. S. A.* 2019, *116* (6), 2158–2164. https://doi.org/10.1073/pnas.1813852116.

(74) Tokić, S.; Žižkova, V.; Štefanić, M.; Glavaš-Obrovac, L.; Marczi, S.; Samardžija, M.; Sikorova, K.; Petrek, M. HLA-A, -B, -C, -DRB1, -DQA1, and -DQB1 Allele and Haplotype Frequencies Defined by next Generation Sequencing in a Population of East Croatia Blood Donors. *Sci. Rep.* 2020, *10* (1), 1–13. https://doi.org/10.1038/s41598-020-62175-9.

(75) Hollenbach, J. A.; Saperstein, A.; Albrecht, M.; Vierra-Green, C.; Parham, P.; Norman, P. J.; Maiers, M. Race, Ethnicity and Ancestry in Unrelated Transplant Matching for the National Marrow Donor Program: A Comparison of Multiple Forms of Self-Identification with Genetics. *PLoS One* 2015, *10* (8), 1–15. https://doi.org/10.1371/journal.pone.0135960.

(76) Theda, C.; Hwang, S. H.; Czajko, A.; Loke, Y. J.; Leong, P.; Craig, J. M. Quantitation of the Cellular Content of Saliva and Buccal Swab Samples. *Sci. Rep.* 2018, *8* (1), 4–11. https://doi.org/10.1038/s41598-018-25311-0.

(77) Yoshizawa, J. M.; Schafer, C. A.; Schafer, J. J.; Farrell, J. J.; Paster, B. J.; Wong, D. T. W. Salivary Biomarkers: Toward Future Clinical and Diagnostic Utilities. *Clin. Microbiol. Rev.* 2013, *26* (4), 781–791. https://doi.org/10.1128/CMR.00021-13.

(78) Juárez-Cedillo, T.; Sánchez-García, S.; Mould-Quevedo, J. F.; García-Peña, C.; Gallo, J. J.; Wagner, F. A.; Vargas-Alarcón, G. Cost-Effective Analysis of Genotyping Using Oral Cells in the Geriatric Population. *Genet. Mol. Res.* 2010, *9* (3), 1886–1895. https://doi.org/10.4238/vol9-3gmr939.

(79) Ruiz, C. A.; Chaney, M. E.; Tosi, A. J. Medical-Grade Buccal Swabs versus Drugstore Cotton Swabs: No Difference in DNA Yield. *MethodsX* 2018, *5*, 39–42. https://doi.org/10.1016/j.mex.2018.01.006.

(80) Montgomery, M. C.; Petraroia, R.; Weimer, E. T. Buccal Swab Genomic DNA Fragmentation Predicts Likelihood of Successful HLA Genotyping by Next-Generation Sequencing. *Hum. Immunol.* 2017, *78* (10), 634–641. https://doi.org/10.1016/j.humimm.2017.08.003.

(81) Davis, C. P.; King, J. L.; Budowle, B.; Eisenberg, A. J.; Turnbough, M. A. Extraction Platform Evaluations: A Comparison of Automate Express™, EZ1® Advanced XL, and Maxwell® 16 Bench-Top DNA Extraction Systems. *Leg. Med.* 2012, *14* (1), 36–39. https://doi.org/10.1016/j.legalmed.2011.09.005.

(82) Karija Vlahović, M.; Kubat, M. DNA Extraction Method from Bones Using Maxwell® 16. *Leg. Med.* 2012, *14* (5), 272–275. https://doi.org/10.1016/j.legalmed.2012.04.004.

(83) Endris, S. Gene Conserve - Articles - Articles - - Articles - - Articles - Gene Conserve - Articles - Articles - Volume 7 - Issue 30 - October / December , 2008 . *Production* 2008, *7* (27), 2008–2010. https://doi.org/10.1007/s.

(84) Greenspoon, S. A.; Ban, J. D.; Sykes, K.; Ballard, E. J.; Edler, S. S.; Baisden, M.; Covington, B. L. Application of the BioMek® 2000 Laboratory Automation Workstation and the DNA IQ™ System to the Extraction of Forensic Casework Samples. *J. Forensic Sci.* 2004, *49* (1), 1–11. https://doi.org/10.1520/jfs2003179.

(85) Nielsen, K.; Mogensen, H. S.; Hedman, J.; Niederstätter, H.; Parson, W.; Morling, N. Comparison of Five DNA Quantification Methods. *Forensic Sci. Int. Genet.* 2008, *2* (3), 226–230. https://doi.org/10.1016/j.fsigen.2008.02.008.

(86) Nakayama, Y.; Yamaguchi, H.; Einaga, N.; Esumi, M. Pitfalls of DNA Quantification Using Dnabinding Fluorescent Dyes and Suggested Solutions. *PLoS One* 2016, *11* (3), 1–12. https://doi.org/10.1371/journal.pone.0150528.

(87) Kumar, D.; Panigrahi, M. K.; Suryavanshi, M.; Mehta, A.; Saikia, K. K. Quantification of DNA Extracted from Formalin Fixed Paraffin-Embeded Tissue Comparison of Three Techniques: Effect on PCR Efficiency. *J. Clin. Diagnostic Res.* 2016, *10* (9), 3–5. https://doi.org/10.7860/JCDR/2016/19383.8407.

(88) Fluorometer, T. H. E. Q. The next Generation in Nucleic Acid and Protein Quantitation. 2011, No. April, 12–13.

(89) Rengarajan, K.; Cristol, S. M.; Mehta, M.; Nickerson, J. M. Quantifying DNA Concentrations Using Fluorometry: A Comparison of Fluorophores. *Mol. Vis.* 2002, *8* (December), 416–421.

(90) Armbrecht, M.; Ag, E. Economic DNA Determination in the Eppendorf BioSpectrometer ® Fluorescence Using Qubit ™ Assay Kits. No. 402.

(91) Timken, M. D.; Swango, K. L. The Author (s) Shown below Used Federal Funds Provided by the U . S . Department of Justice and Prepared the Following Final Report : Document Title : Quantitation of DNA for Forensic DNA Typing by QPCR Orrego , Mavis Date Chong , Martin R . Buoncristia. *Office* 2005.

(92) Chang, D.; Tram, K.; Li, B.; Feng, Q.; Shen, Z.; Lee, C. H.; Salena, B. J.; Li, Y. Detection of DNA Amplicons of Polymerase Chain Reaction Using Litmus Test. *Sci. Rep.* 2017, *7* (1), 1–8. https://doi.org/10.1038/s41598-017-03009-z.

(93) Mullis, K. B. The Unusual Origin of the Polymerase Chain Reaction. *Sci. Am.* 1990, *262* (4), 56–65. https://doi.org/10.1038/scientificamerican0490-56.

(94) Carriço, J. A.; Sabat, A. J.; Friedrich, A. W.; Ramirez, M. Bioinformatics in Bacterial Molecular Epidemiology and Public Health: Databases, Tools and the next-Generation Sequencing Revolution. *Eurosurveillance* 2013, *18* (4), 1–8. https://doi.org/10.2807/ese.18.04.20382-en.

(95) Porta, A. R.; Enners, E. Determining Annealing Temperatures for Polymerase Chain Reaction. *Am. Biol. Teach.* 2012, *74* (4), 256–260. https://doi.org/10.1525/abt.2012.74.4.9.

(96) Rodríguez-Lázaro, D.; Cook, N.; Hernández, M. Real-Time PCR in Food Science: PCR Diagnostics. *Curr. Issues Mol. Biol.* 2013, *15* (2), 39–44. https://doi.org/10.21775/cimb.015.039.

(97) Edwards, M. C.; Gibbs, R. A. Multiplex PCR: Advantages, Development, and Applications. *Genome Res.* 1994, *3* (4). https://doi.org/10.1101/gr.3.4.S65.

(98) Hayden, M. J.; Nguyen, T. M.; Waterman, A.; Chalmers, K. J. Multiplex-Ready PCR: A New Method for Multiplexed SSR and SNP Genotyping. *BMC Genomics* 2008, *9*, 1–12. https://doi.org/10.1186/1471-2164-9-80.

(99) Hanson, E. K.; Berdos, P. N.; Ballantyne, J. Testing and Evaluation of 43 "Noncore" Y Chromosome Markers for Forensic Casework Applications. *J. Forensic Sci.* 2006, *51* (6), 1298–1314. https://doi.org/10.1111/j.1556-4029.2006.00263.x.

(100) Moretti, T. R.; Baumstark, A. L.; Defenbaugh, D. A.; Keys, K. M.; Smerick, J. B.; Budowle, B. Validation of Short Tandem Repeats (STRs) for Forensic Usage: Performance Testing of Fluorescent Multiplex STR Systems and Analysis of Authentic and Simulated Forensic Samples. *J. Forensic Sci.* 2001, *46* (3), 15018J. https://doi.org/10.1520/jfs15018j.

(101) Pojmanová, P.; Ladislavová, N.; Škeříková, V.; Kania, P.; Urban, Š. Human Scent Samples for Chemical Analysis. *Chem. Pap.* 2020, *74* (5), 1383–1393. https://doi.org/10.1007/s11696-019-00989-2.

(102) Rabindra, P.; Raju, N. Gel-Electrophoresis and Its Applications. *Gel Electrophor. - Princ. Basics* 2012, No. May. https://doi.org/10.5772/38479.

(103) Bick, S. L.; Bick, D. P.; Wells, B. E.; Roesler, M. R.; Strawn, E. Y.; Lau, E. C. Preimplantation HLA Haplotyping Using Tri-, Tetra-, and Pentanucleotide Short Tandem Repeats for HLA Matching. *J. Assist. Reprod. Genet.* 2008, *25* (7), 323–331. https://doi.org/10.1007/s10815-008-9233-2.

(104) Id, K. M. Development of Highly Sensitive and Low-Cost DNA Agarose Gel Electrophoresis Detection Systems , and Evaluation of Non-Mutagenic and Loading Dye-Type DNA-Staining Reagents. *PLoS One* 2019, 4–7.

(105) McCord, B. R.; Gauthier, Q.; Cho, S.; Roig, M. N.; Gibson-Daw, G. C.; Young, B.; Taglia, F.; Zapico, S. C.; Mariot, R. F.; Lee, S. B.; Duncan, G. Forensic DNA Analysis. *Anal. Chem.* 2019, *91* (1), 673–688. https://doi.org/10.1021/acs.analchem.8b05318.

(106) Cullen, M.; Malasky, M.; Harding, A.; Carrington, M. High-Density Map of Short Tandem Repeats across the Human Major Histocompatibility Complex. *Immunogenetics* 2003, *54* (12), 900–910. https://doi.org/10.1007/s00251-002-0535-8.

(107) Jordan, D.; Mills, D. E. Past, Present, and Future of DNA Typing for Analyzing Human and Non-Human Forensic Samples. *Front. Ecol. Evol.* 2021, *9* (March). https://doi.org/10.3389/fevo.2021.646130.

(108) Schneider, P. M.; Bender, K.; Mayr, W. R.; Parson, W.; Hoste, B.; Decorte, R.; Cordonnier, J. STR Analysis of Artificially Degraded DNA - Results of a Collaborative European Exercise. *Forensic Sci. Int.* 2004, *139* (2–3), 123–134. https://doi.org/10.1016/j.forsciint.2003.10.002.

(109) Jeffreys, A. J.; Wilson, V.; Theint, S. L.; Lel, L. Fingerprints of Dna Alex Jeffreys. 1985, 6–9.

(110) Lareu, M. *Short Tandem Repeats*, 2nd ed.; Elsevier Ltd., 2013. https://doi.org/10.1016/B978-0-12-382165-2.00040-4.

(111) Butler, J. M. Fundamentals of Forensic DNA Typing. *Fundam. Forensic DNA Typing* 2010. https://doi.org/10.1016/C2009-0-01945-X.

(112) Thermo Fisher, S. Fragment-Analysis-Chemistry-Guide. 2014, No. 4474504.

(113) Van Wieren-De Wijer, D. B. M. A.; Maitland-Van Der Zee, A. H.; De Boer, A.; Belitser, S. V.; Kroon, A. A.; De Leeuw, P. W.; Schiffers, P.; Janssen, R. G. J. H.; Van Duijn, C. M.; Stricker, B. H. C. H.; Klungel, O. H. Determinants of DNA Yield and Purity Collected with Buccal Cell Samples. *Eur. J. Epidemiol.* 2009, *24* (11), 677–682. https://doi.org/10.1007/s10654-009-9388-x.

(114) Erlichster, M.; Goudey, B.; Skafidas, E.; Kwan, P. Cross-Ethnicity Tagging SNPs for HLA Alleles Associated with Adverse Drug Reaction. *Pharmacogenomics J.* 2019, *19* (3), 230–239. https://doi.org/10.1038/s41397-018-0039-z.

(115) Tian, W.; Boggs, D. A.; Uko, G.; Essiet, A.; Inyama, M.; Banjoko, B.; Adewole, T.; Ding, W. Z.; Mohseni, M.; Fritz, R.; Chen, D. F.; Palmer, L. J.; Fraser, P. A. MICA, HLA-B Haplotypic Variation in Five Population Groups of Sub-Saharan African Ancestry. *Genes Immun.* 2003, *4* (7), 500–505. https://doi.org/10.1038/sj.gene.6364017.

(116) Favorova, O. O.; Favorov, A. V.; Boiko, A. N.; Andreewski, T. V.; Sudomoina, M. A.; Alekseenkov, A. D.; Kulakova, O. G.; Gusev, E. I.; Parmigiani, G.; Ochs, M. F. Three Allele Combinations Associated with Multiple Sclerosis. *BMC Med. Genet.* 2006, *7* (2), 1–9. https://doi.org/10.1186/1471-2350-7-63.

(117) Gai, L. ping; Liu, H.; Cui, J. hui; Ji, N.; Ding, X. dong; Sun, C.; Yu, L. shui. Distributions of Allele Combination in Single and Cross Loci among Patients with Several Kinds of Chronic Diseases and the Normal Population. *Genomics* 2015, *105* (3), 168–174. https://doi.org/10.1016/j.ygeno.2014.12.008.

(118) Nunes, K.; Aguiar, V. R. C.; Silva, M.; Sena, A. C.; de Oliveira, D. C. M.; Dinardo, C. L.; Kehdy, F. S. G.; Tarazona-Santos, E.; Rocha, V. G.; Carneiro-Proietti, A. B. F.; Loureiro, P.; Flor-Park, M. V.; Maximo, C.; Kelly, S.; Custer, B.; Weir, B. S.; Sabino, E. C.; Porto, L. C.; Meyer, D. How Ancestry Influences the Chances of Finding Unrelated Donors: An Investigation in Admixed Brazilians. *Front. Immunol.* 2020, *11* (November), 1–14. https://doi.org/10.3389/fimmu.2020.584950.

(119) Alter, I.; Gragert, L.; Fingerson, S.; Maiers, M.; Louzoun, Y. HLA Class I Haplotype Diversity Is Consistent with Selection for Frequent Existing Haplotypes. *PLoS Comput. Biol.* 2017, *13* (8), e1005693. https://doi.org/10.1371/journal.pcbi.1005693.

(120) Zumelzu, C.; Le Roux-Villet, C.; Loiseau, P.; Busson, M.; Heller, M.; Aucouturier, F.; Pendaries, V.; Lièvre, N.; Pascal, F.; Brette, M. D.; Doan, S.; Charron, D.; Caux, F.; Laroche, L.; Petit, A.; Prost-Squarcioni, C. Black Patients of African Descent and HLA-DRB115:03 Frequency Overrepresented in Epidermolysis Bullosa Acquisita. *J. Invest. Dermatol.* 2011, *131* (12), 2386–2393. https://doi.org/10.1038/jid.2011.231.

(121) Initial, G.; For, U. T.; In, G.; States, U. Guide Initial Urate-Lowering Therapy for Gout in The. 2018, *46* (5), 594–600. https://doi.org/10.1016/j.semarthrit.2016.10.009.THE.

(122) Liu, X.; Sun, J.; Yu, H.; Chen, H.; Wang, J.; Zou, H.; Lu, D.; Xu, J.; Zheng, S. L. Tag SNPs for HLA-B Alleles That Are Associated with Drug Response and Disease Risk in the Chinese Han Population. *Pharmacogenomics J.* 2015, *15* (5), 467–472. https://doi.org/10.1038/tpj.2015.7.

(123) Hamed, K. H. La Distribution Du Tau de Kendall Pour Tester La Significativité de La Corrélation Croisée Dans Des Données Persistantes. *Hydrol. Sci. J.* 2011, *56* (5), 841–853. https://doi.org/10.1080/02626667.2011.586948.

(124) Rohrer, J. M. Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Adv. Methods Pract. Psychol. Sci.* 2018, *1* (1), 27–42. https://doi.org/10.1177/2515245917745629.

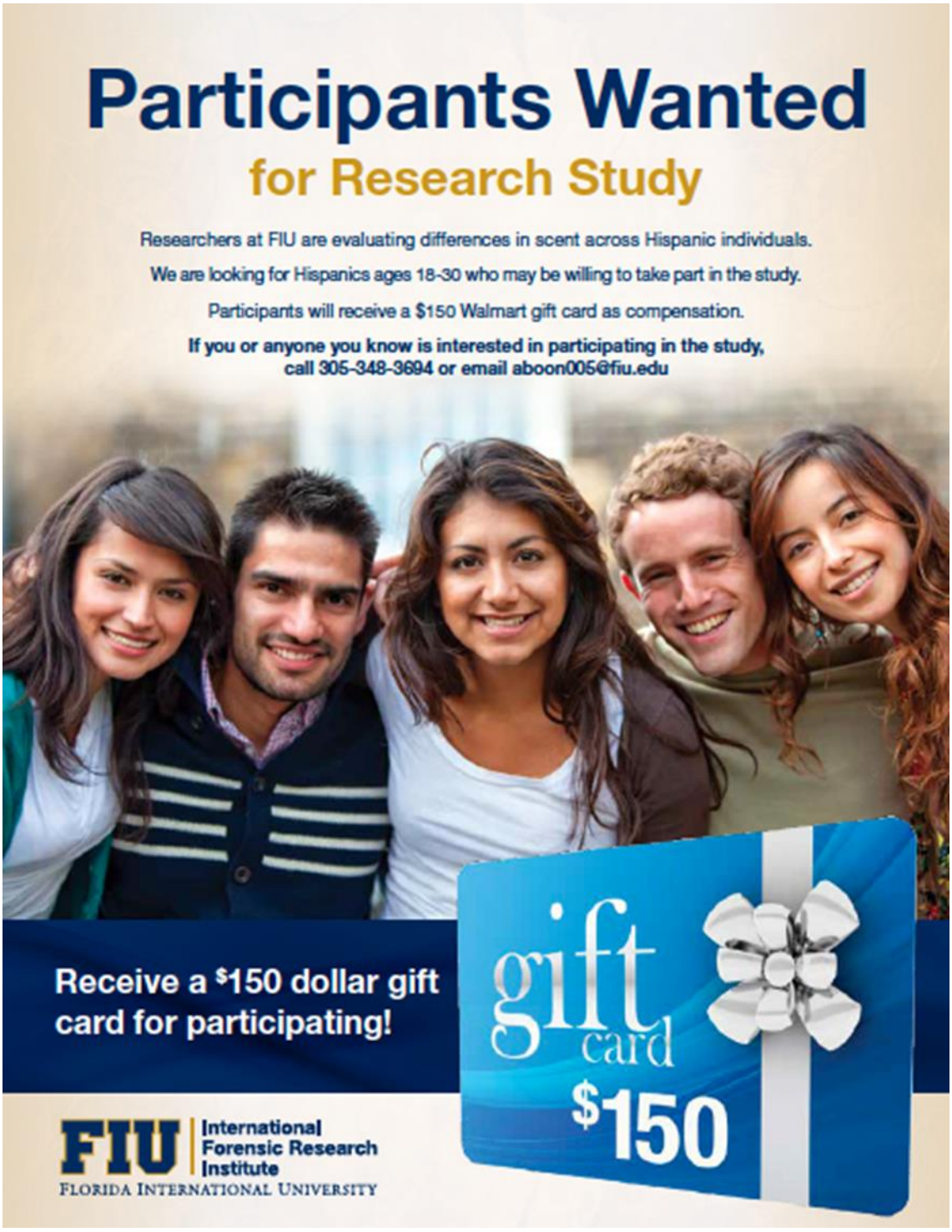
(125) Doležal, P.; Kyjaková, P.; Valterová, I.; Urban, Š. Qualitative Analyses of Less-Volatile Organic Molecules from Female Skin Scents by Comprehensive Two Dimensional Gas Chromatography–Time of Flight Mass Spectrometry. *J. Chromatogr. A* 2017, *1505*, 77–86. https://doi.org/10.1016/j.chroma.2017.04.062.

(126) Caraballo, N. I.; Mendel, J.; Holness, H.; La Salvia, J.; Moroose, T.; Eckenrode, B.; Stockham, R.; Furton, K.; Mills, D. E. An Investigation into the Concurrent Collection of Human Scent and Epithelial Skin Cells Using a Non-Contact Sampling Device. *Forensic Sci. Int.* 2016, *266*, 148–159. https://doi.org/10.1016/j.forsciint.2016.05.019.

(127) Filetti, V.; Di Mizio, G.; Rendine, M.; Fortarezza, P.; Ricci, P.; Pomara, C.; Messina, G.; Riezzo, I.; Zammit, C.; Messina, A.; Salerno, M.; Sessa, F. Volatile Organic Compounds: Instrumental and Canine Detections Link an Individual to the Crime Scene. *Egypt. J. Forensic Sci.* 2019, *9* (1), 1–11. https://doi.org/10.1186/s41935-019-0139-1.

APPENDICES

Appendix 1: Flyer for human scent study recruitment (English)



Appendix 2:Flyer for human scent study recruitment (Spanish)

Appendix 3:Human scent project description and instructions

FIU | FLORIDA INTERNATIONAL UNIVERSITY

Purpose of the Study

In collaboration with Colgate, FIU will be investigating the odor profiles of Hispanics and Caucasians. Due to the novelty of human scent research, human scent evidence has been undervalued in the court of law. However, this type of evidence has significant value when physical evidence is not available at crime scenes. In order to increase the integrity of human scent evidence, the knowledge of the VOC profiles must be expanded upon. To date, there is minimal understanding about the cellular origin of VOCs. Hence, this research focuses on investigating the relationship between the Human Leukocyte Antigen (HLA) and VOCs within human subjects.

What do we need from you?
If you agree to participate in this study, we need you to be dedicated to following the specific instructions that will be given.

- Unscented hygiene products will be provided to you. The subjects must use these products for two days prior to each sampling date.
- During the sampling period, refrain from using any scented products (soaps, perfumes, deodorants, body sprays etc.)
- During the sampling period, refrain from alcohol consumption
- Questionnaire and Consent form must be completed

What will you do on the sampling date?
Once scheduled for a sampling date, the subjects will be asked to sit in a chair. A gauze pad will be placed under their arm for 1 hour.
How long does the study take?
The total time we will need from each subject is 11 days.
Here is a breakdown of the sampling period:

Will I be compensated?
Yes! As advertised on the flyer, A $150 gift card will be provided to you for **completing** the study.

Appendix 4:VOCs evlauated fror human scent study

Compound	**CAS #**	**Retention Time**	**Reference**
Octanal	124-13-0	7.574	[20]
Nonanal	124-19-6	9.007	[20,21]
Tetradecane	629-59-4	9.133	[16]
Propanoic acid, 2-methyl	967-62-1	9.472	[16]
Decanoic acid, methyl	110-42-9	11.268	[16]
Butanoic acid	107-92-6	12.155	[27,125]
Isovaleric acid	503-74-2	12.155	[27]
Valeric acid	109-52-4	13.609	[27]
Naphthalene	91-20-3	10.928	[16,20]
1-Heptadecane	6765-39-5	13.655	[18]
2-Tridecanone	593-08-8	14.429	[16,18]
Octadecane	593-45-3	15.578	[22]
Cyclodecane	203-96-9	15.828	[27]
Tridecanoic acid	638-53-9	16.844	[37]
Eicosane	112-95-8	17.008	[37]
Isopropyl Myristate	110-27-0	17.247	[12,37]
Nonanoic acid	112-05-0	18.114	[27,125]
n-Hexyl salicylate	6259-76-3	19.241	[12,37]
Hexanoic acid, methyl ester	106-70-7	20.254	[126]
Tridecane	629-50-5	20.258	
Heptanal	111-71-7	19.886	[18,20]
Isopropyl Palmitate	142-91-6	20.881	[37]
Octanal	124-13-0	22.279	[126]
Undecanoic acid	112-37-8	22.451	[12,37]
Homomenthyl salicylate	118-56-9	27.474	[37]

Benzophenone	119-61-9	23.809	[37]
Dodecanoic acid	143-07-7	24.831	[16,18]
Pentadecanoic acid	1002-84-2	30.973	[20]
n-Hexadecanoic acid	57-10-3	31.737	[20]
Heptadecanoic acid	506-12-7	32.54	[127]
Octadecanoic acid	57-11-4	33.262	[20,127]

Appendix 5:Genotype frequency present within subjects (Gene D6S2837)

Sample	Ethnicity	142,146	146,181	169,189	142,173	146,173	142,188	146,169	181,184	173,185	173,173	173,177	173,192	142,174	169,174	141,184	142,177
B150	African American	0.230769	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B126	African American	0	0.153846	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B157	African American	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0
B132	African American	0	0	0	0.153846	0	0	0	0	0	0	0	0	0	0	0	0
B151	African American	0	0	0	0	0.153846	0	0	0	0	0	0	0	0	0	0	0
B153	African American	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B115	African American	0.230769	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B156	African American	0	0	0	0.153846	0	0	0	0	0	0	0	0	0	0	0	0
B131	African American	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0
B101	African American	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0
B152	African American	0	0.153846	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B158	African American	0	0	0	0	0.153846	0	0	0	0	0	0	0	0	0	0	0
B130	African American	0.230769	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0	0
B134	Caucasian	0	0	0.125	0	0	0	0	0	0	0	0	0	0	0	0	0
B125	Caucasian	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0
B107	Caucasian	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0
B134	Caucasian	0	0	0	0	0.25	0	0	0	0	0	0.125	0	0	0	0	0
B127	Caucasian	0	0	0	0.125	0	0	0	0	0	0	0	0	0	0	0	0
B117	Caucasian	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0
B121	Caucasian	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0
B108	Caucasian	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B103	Hispanic	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0	0
B154	Hispanic	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0
B102	Hispanic	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0	0
B155	Hispanic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0
B118	Hispanic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B104	Hispanic	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0.1
B135	Hispanic	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
B124	Hispanic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B128	Hispanic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B129	Hispanic	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0

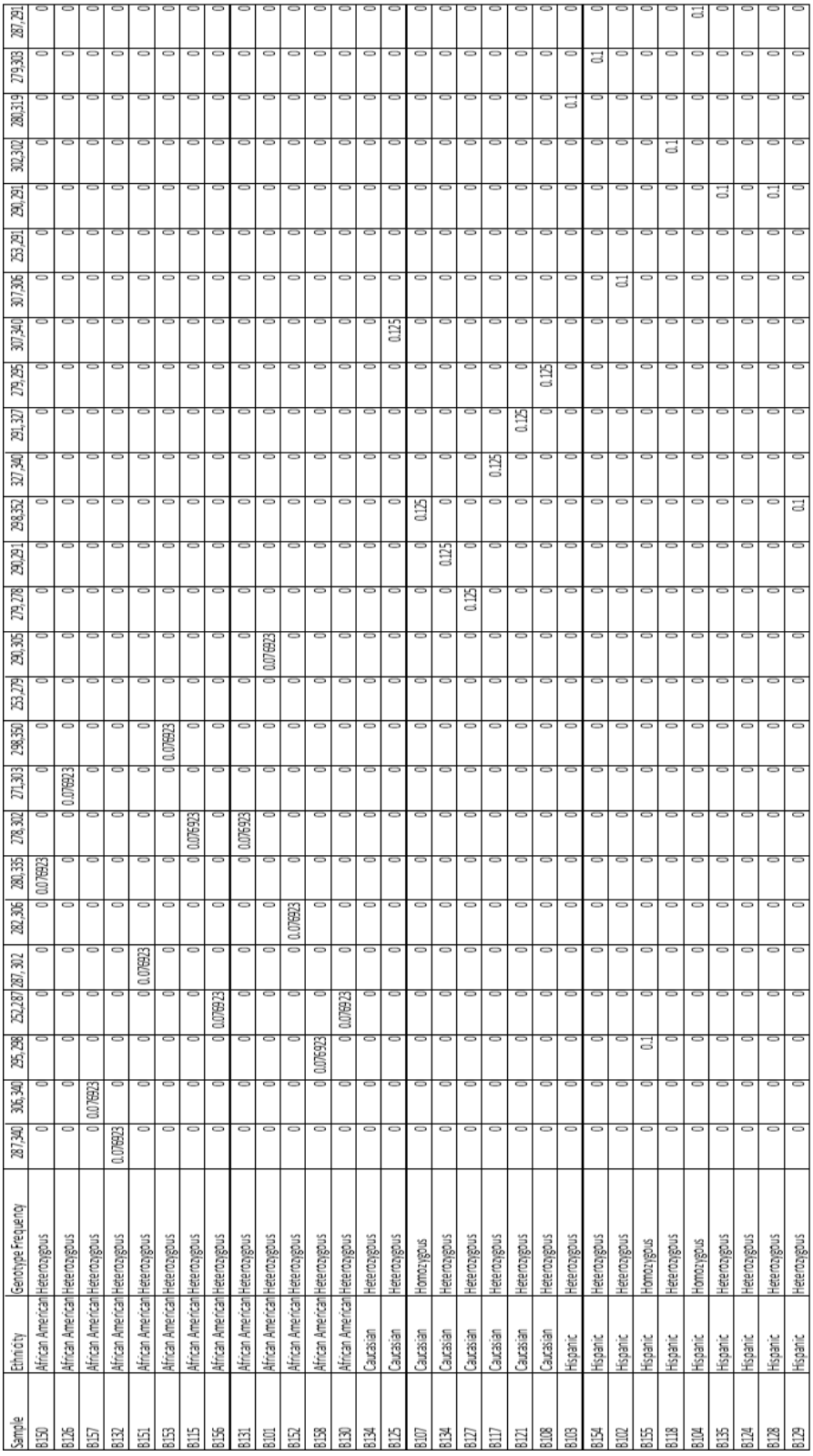
Appendix 6:Genotype frequency present within subjects (D6S2925 gene)

Sample	Ethnicity	Genotype	109,112	109,111	107,109	119,119	120132	111127	120128	111120	120136	128127	136136	111128	109110	111124	112132	128136	112128	136141
B150	African American	Heterozygous	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B126	African American	Heterozygous	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0
B157	African American	Heterozygous	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0
B132	African American	Heterozygous	0.076923	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0
B151	African American	Heterozygous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B153	African American	Homozygous	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B115	African American	Heterozygous	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B156	African American	Heterozygous	0	0	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0
B131	African American	Heterozygous	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B101	African American	Heterozygous	0	0	0	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0
B152	African American	Heterozygous	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0
B158	African American	Heterozygous	0	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0
B130	African American	Heterozygous	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B125	Caucasian	Heterozygous	0	0	0	0	0	0.125	0	0	0	0	0	0	0	0	0	0	0	0
B107	Caucasian	Heterozygous	0.125	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B134	Caucasian	Heterozygous	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0	0
B127	Caucasian	Heterozygous	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0
B117	Caucasian	Heterozygous	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0	0	0	0	0
B121	Caucasian	Homozygous	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0
B108	Caucasian	Heterozygous	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0
B103	Hispanic	Homozygous	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0
B154	Hispanic	Heterozygous	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
B102	Hispanic	Homozygous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1
B155	Hispanic	Heterozygous	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
B118	Hispanic	Heterozygous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0
B104	Hispanic	Heterozygous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0
B135	Hispanic	Heterozygous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0
B124	Hispanic	Heterozygous	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0
B128	Hispanic	Heterozygous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B129	Hispanic	Heterozygous	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0.1	0	0	0	0

Appendix 7:Genotype frequency of each subject (M138 gene)

Sample	Ethnicity	Genotype	253,257	246,271	257,261	250,259	246,286	246,261	242,250	246,272	242,268	246,294	250,250	253,275	261,268	246,270	250,253	253,272	246,250
B150	African Ar	Heterozyg	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B126	African Ar	Heterozyg	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B157	African Ar	Heterozyg	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0
B132	African Ar	Heterozyg	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0
B151	African Ar	Heterozyg	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
B153	African Ar	Heterozyg	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0
B115	African Ar	Heterozyg	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0
B156	African Ar	Heterozyg	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0
B131	African Ar	Heterozyg	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0
B101	African Ar	Heterozyg	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0
B152	African Ar	Heterozyg	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B158	African Ar	Heterozyg	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
B130	African Ar	Heterozyg	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B134	Caucasian	Heterozyg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B125	Caucasian	Heterozyg	0	0	0	0.125	0	0	0	0	0	0	0	0	0	0	0	0	0
B107	Caucasian	Homozygo	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0
B134	Caucasian	Heterozyg	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0
B127	Caucasian	Heterozyg	0.125	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0
B117	Caucasian	Heterozyg	0	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0
B121	Caucasian	Heterozyg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B108	Caucasian	Heterozyg	0	0	0.125	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B103	Hispanic	Heterozyg	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0
B154	Hispanic	Heterozyg	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B102	Hispanic	Heterozyg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B155	Hispanic	Heterozyg	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B118	Hispanic	Heterozyg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0
B104	Hispanic	Heterozyg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B135	Hispanic	Heterozyg	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0
B124	Hispanic	Heterozyg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0
B128	Hispanic	Heterozyg	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0
B129	Hispanic	Heterozyg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1

Appendix 8:Genotype frequencies of subjects (D6S2917 gene)



Sample	Ethnicity	Genotype Frequency	287,290	306,340	295,298	252,287	287,302	282,306	280,335	278,302	271,303	298,350	253,279	290,305	279,278	290,291	298,352	327,340	291,327	279,285	307,340	307,306	253,291	290,291	302,302	280,319	279,303	287,291
B150	African American	Heterozygotus	0	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B126	African American	Heterozygotus	0	0	0	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B157	African American	Heterozygotus	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B132	African American	Heterozygotus	0.076923	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B151	African American	Heterozygotus	0	0	0	0	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B153	African American	Heterozygotus	0	0	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B115	African American	Heterozygotus	0	0	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0
B156	African American	Heterozygotus	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B131	African American	Heterozygotus	0	0	0	0	0	0	0	0.076923	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B101	African American	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B152	African American	Heterozygotus	0	0	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B158	African American	Heterozygotus	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B130	African American	Heterozygotus	0	0	0	0.076923	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B134	Caucasian	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0
B125	Caucasian	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0
B107	Caucasian	Homozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0	0	0	0	0
B134	Caucasian	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0	0	0	0	0	0
B127	Caucasian	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0	0	0	0	0	0	0
B117	Caucasian	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0	0	0	0
B121	Caucasian	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0	0	0
B108	Caucasian	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.125	0	0	0	0	0	0	0	0
B103	Hispanic	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0
B154	Hispanic	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0
B102	Hispanic	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0
B155	Hispanic	Homozygotus	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B118	Hispanic	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0
B104	Hispanic	Homozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1
B135	Hispanic	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0
B124	Hispanic	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B128	Hispanic	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0
B129	Hispanic	Heterozygotus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0

Appendix 9: Genotype frequencies of subjects (D6S2787 Gene)

Sample	Ethnicity	Genotype	138,140	136,138	138,144	140,142	137,144	140,144	137,139	150,152
B150	African American	Heterozygous	0	0	0	0	0	0.076923	0	0
B126	African American	Heterozygous	0	0	0	0	0.076923	0	0	0
B157	African American	Heterozygous	0	0.076923	0	0	0	0	0	0
B132	African American	Heterozygous	0.076923	0.076923	0	0	0	0	0	0
B151	African American	Heterozygous	0	0.076923	0	0	0	0	0	0
B153	African American	Homozygous	0	0	0	0	0	0	0	0
B115	African American	Heterozygous	0	0.076923	0	0	0	0	0	0
B156	African American	Heterozygous	0	0	0	0	0	0	0	0
B131	African American	Heterozygous	0	0	0	0	0	0	0	0
B101	African American	Heterozygous	0.076923	0	0.076923	0	0	0	0	0
B152	African American	Heterozygous	0.076923	0	0	0	0	0	0	0
B158	African American	Heterozygous	0	0	0	0	0	0	0	0
B130	African American	Heterozygous	0.076923	0.076923	0	0	0	0	0	0
B125	Caucasian	Heterozygous	0.125	0	0	0	0	0	0	0
B107	Caucasian	Heterozygous	0	0	0	0	0	0	0.125	0
B134	Caucasian	Heterozygous	0	0	0	0	0	0	0	0.125
B127	Caucasian	Heterozygous	0.125	0	0	0	0	0	0	0
B117	Caucasian	Heterozygous	0.125	0	0	0	0	0	0	0
B121	Caucasian	Homozygous	0.125	0	0	0	0	0	0	0
B108	Caucasian	Heterozygous	0.125	0	0	0	0	0	0	0
B103	Hispanic	Homozygous	0.1	0	0	0	0	0	0	0
B154	Hispanic	Heterozygous	0.1	0	0	0	0	0	0	0
B102	Hispanic	Homozygous	0.1	0	0	0	0	0	0	0
B155	Hispanic	Heterozygous	0	0.1	0	0	0	0	0	0
B118	Hispanic	Heterozygous	0.1	0	0	0	0	0	0	0
B104	Hispanic	Heterozygous	0	0	0	0.1	0	0	0	0
B135	Hispanic	Heterozygous	0	0	0	0	0	0	0	0
B124	Hispanic	Heterozygous	0.1	0	0	0	0	0	0	0
B128	Hispanic	Heterozygous	0.1	0	0	0	0	0	0	0
B129	Hispanic	Heterozygous	0	0.1	0	0	0	0	0	0

VITA

ALICE BREIA BOONE

Born, Brooklyn New York

2011-2015	B.A., Biochemistry Claflin University Orangeburg, South Carolina
2015-2018	M.S., Forensic Science Florida International University Miami, Florida
2015	Gradate Minority Opportunity Program Scholarship Florida International University
2016	McKnight Doctoral Fellowship Recipient
2018-2022	Doctoral Candidate Florida International University Miami Florida
2018-2021	Teaching Assistant Florida International University Miami, Florida
2021	IFRI Development Award

PRESENTATIONS

The Enhancement of Human Scent Profiles as Forensic Evidence. Presented at the IFRI Symposium; Miami, Florida; May 2017

The Utilization of Controlled Odor Mimic Devices to Improve Law Enforcement Canine Training. Presented at the International Working Dog Conference (IWDC); Banff, Canada; March 2017

Utilization of Controlled Odor Mimic Devices for Law Enforcement Canines. Presented at the American Academy of Forensic Science (AAFS); New Orleans, Louisiana; February 2017

The Enhancement of Human Scent Profiles as Forensic Evidence. Presented at the Northwestern Black Graduate Student Association Conference "Black Scholars Matter: PHinDing the Strength to be PHinisheD"; Chicago, FL; April 2018

The Enhancement of Human Scent Profiles as Forensic Evidence. Presented at the American Academy of Forensic Science (AAFS); Seattle, Washington; February,2019

The Enhancement of Human Scent Profiles as Forensic Evidence. Presented at the McKnight Doctoral Fellowship Mid-Year Meeting; Tampa, FL; February 2019

The Enhancement of Human Scent Profiles as Forensic Evidence. Presented at the American Academy of Forensic Science (AAFS); Anaheim, California; February 2020

The Enhancement of Human Scent Profiles as Forensic Evidence.

Presented at the PitConn Conference; Philadelphia, Pennsylvania; March 2020