
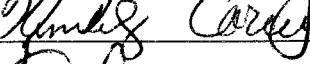
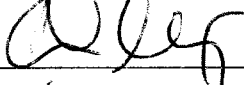





THE USE OF "OMICS" IN NEXT GENERATION SNP ANALYSIS: FUTURE OF  
FORENSIC PHENOTYPING

by

Khatera Noor Harrif  
A Research Project  
Submitted to the  
Graduate Faculty  
of  
George Mason University  
in Partial Fulfillment of  
The Requirements for the Degree  
of  
Master of Science  
Forensic Science

Committee:

	Joseph DiZinno, Research Project Director
	Professor Kimberly Carisi, Committee Member
	Professor Amandeep Gill, Committee Member
	Professor William Whildin, Department Chairperson
	Dr. Timothy L. Born, Associate Dean for Student and Academic Affairs, College of Science
	Dr. Vikas Chandhoka, Dean, College of Science

Date: 5-01-12 Spring Semester 2012  
George Mason University  
Fairfax, VA

The Use of “Omics” in Next Generation SNP Analysis: Future of Forensic Phenotyping

A research project submitted in partial fulfillment of the requirements for the degree of  
Master of Science at George Mason University

by

Khatera Noor Harrif  
Bachelor of Arts  
George Mason University, 2005

Director: Joseph DiZinno, Professor  
Department of Forensic Science

Spring Semester 2012  
George Mason University  
Fairfax, VA



This work is licensed under a [creative commons attribution-noncommercial 3.0 unported license](https://creativecommons.org/licenses/by-nc/3.0/).

## **DEDICATION**

This is dedicated to my youngest brother Shabbier Ahmad Harrif, who passed away on September 5, 2011.

## **ACKNOWLEDGEMENTS**

I would like to thank my loving husband, Mustafa, who has helped me every step of the way. I would also like to thank my family and friends for their support. Finally thanks to Professor DiZinno, Professor William Whildin and Professor Emily Rancourt for their encouragement throughout this program.

## TABLE OF CONTENTS

	Page
List of Tables .....	vii
List of Figures .....	viii
List of Abbreviations or Symbols .....	ix
Abstract .....	x
Chapter One: INTRODUCTION .....	1
1.1 History of Forensic Genetics.....	1
1.2 Purpose of Research.....	3
1.3 Single Nucleotide Polymorphism.....	3
1.4 Advantages vs. Disadvantages of SNPs.....	4
1.5 SNP Categories in Forensics.....	5
1.5.1 Identity testing SNPs (IISNPs).....	6
1.5.2 Lineage or Lineage Informative SNPs (LISNPs).....	6
1.5.3 Ancestry Informative SNPs (AISNPs).....	6
1.5.4 Phenotype Informative SNPs (PISNPs).....	7
1.6 SNP Database.....	7
Chapter Two: FORENSIC PHENOTYPING .....	9
2.1 Indirect Phenotyping.....	9
2.2 Direct Phenotyping.....	10
2.2.1 SNP Panels for Skin Color.....	15
2.2.2 Case Study: The IrisPlex, an Intelligence DNA System.....	18
2.2.3 Case Study: Regression Statistical Model to accurately predict eye color variations in Dutch Individuals.....	18
2.2.4 Case Study: Human Eye Color and HERC2, OCA2 and MATP/SLC45A2 genes.....	19
2.2.5 Case Study: Model Based Prediction of Human Hair Color Variation.....	19
2.2.6 Case Study: MC1R gene indicative of Red Hair.....	20

Chapter THREE: CURRENT AND FUTURE GEN “OMICS” PLATFORMS .....	21
3.1 Current “Omics” Technologies .....	21
3.1.1 MiniSequencing.....	22
3.1.2 Pyrosequencing.....	24
3.2 Future “Omics” Technologies.....	25
3.2.1 Sequencing Principles for Illumina, 454 and SOLiD platforms.....	26
3.2.2 Third Generation Sequencing.....	30
Chapter FOUR: ETHICAL, PRIVACY CHALLENGES and DAUBERT GUIDELINES .....	33
4.1 Ethical and Privacy Issues.....	33
4.2 Acceptance of PISNPs and Daubert Guidelines .....	35
Chapter Five: CONCLUSION .....	37
References.....	40

## LIST OF TABLES

Table	Page
Table 1- Genes associated with pigmentation .....	14
Table 2- Current SNP Typing Platforms .....	21



## LIST OF FIGURES

Figure	Page
Figure 1- Budowle et al., 2008.....	10
Figure 2- Melansome (Brilliant, 2008) .....	12
Figure 3- Minisequencing (Sobrino et al., 2005).....	23
Figure 4- An Example of MALDI-TOF MS (Stanssens et al., 2004).....	24
Figure 5- Pyrosequencing Cascade System retrieved from <a href="http://www.lifetechnologies.com">www.lifetechnologies.com</a> ..	25
Figure 6- Three major NGS chemistries (Berglund, Kiialainen & Syvanen, 2011). .....	28
Figure 7- Helicos System (Berglund et al., 2011) .....	31

## LIST OF ABBREVIATIONS OR SYMBOLS

Agouti Signaling Protein Gene .....	ASIP
Base Pair .....	bp
Deoxyribonucleic Acid .....	DNA
External Visible Characteristics.....	EVCs
Genome Wide Association .....	GWA
Melanocortin 1 Receptor Gene .....	MC1R
Nanogram.....	ng
National Center for Biotechnology .....	NCBI
Next Generation Sequencing .....	NGS
Oculocutaneous Albinism Type 4.....	PCA4
Picogram .....	pg
Polymerase Chain Reaction .....	PCR
Restriction Fragment Length Polymorphism.....	RFLP
Single Nucleotide Polymorphism .....	SNP
Short Tandem Repeat.....	STR
Single Nucleotide Polymorphism .....	SNP
Trypsinase Gene.....	TYR
Scientific Working Groups on DNA Analysis Methods.....	SWGDM

## **ABSTRACT**

### **THE USE OF “OMICS” IN NEXT GENERATION SNP ANALYSIS: FUTURE OF FORENSIC PHENOTYPING**

Khatera Noor Harrif, M.S.

George Mason University, 2012

Research Project Director: Prof. Joseph DiZinno

When a crime occurs, investigators collect samples from the crime scene and interview witnesses to learn the External Visible Characteristics (EVCs) of the suspect. This EVC includes the hair color, eye color and skin color of the suspect and lead investigators to develop a profile.

One type of sample often collected at crime scenes is DNA. Genetic Typing, or DNA profiling of an individual's identity, has proven a powerful tool. DNA profiles generated are uploaded in established databases in hopes of finding a match. However, this type of traditional genetic profiling has several important drawbacks. First, it provides no information about the suspect's EVC or phenotypic profile. Furthermore, it requires a DNA sample from the suspect in order to compare with the DNA sample from the crime scene. If no sample can be obtained or no match is made after testing, this typically leads to a case turning “cold”.

**But** what if scientists were able to generate a phenotypic profile or EVCs of the suspect, using the minute amount of DNA often left at the crime scene? This would constitute a major breakthrough allowing investigators to narrow down the list of potential suspects. This project discusses the use of Phenotype Informative Simple Nucleotide Polymorphisms (PISNPs), SNPs that identify an individual's particular phenotype, such as skin color, hair color, and eye color.

With recent advances in new gen “omics” technologies, such as Next and Third Generation Sequencing, an increasing number of applications in forensic phenotyping are becoming possible. These new powerful technologies require new approaches to quality control and bioinformatics – this will likely be a key aspect in the development of new PISNPs to further expand profiling.

Though the idea and process of forensic phenotyping is still under development, ethical issues have been raised by privacy advocates. Such groups argue that phenotypic profiling is a violation of privacy. Despite such arguments, research in the field of PISNP is very promising and will have a tremendous impact in helping investigators reach accurate conclusions and prevent wrongful convictions.

## **CHAPTER ONE: INTRODUCTION**

### **1.1 History of Forensic Genetics**

Development of forensic genetics began in 1900 when Karl Landsteiner identified the use of ABO blood typing for human identification. The classification of human blood typing was based on the presence or absence of the Agglutinogens antigens, Antigen A or B, which are carried on the surface of red blood cells. An individual can have blood Type A, Type B, Type O, or Type AB. Agglutinogens are glycoproteins and are relatively unstable due to insufficient levels in all tissue types (Ronny 2010). Since ABO blood typing required a large amount of sample to produce reliable results, the system was not very sensitive. A large number of the population would have the known blood types, and since discrimination levels were low by ABO typing, the pool of suspect donors was extremely large and individualization was very difficult (Butler, 2010). Due to such limitations, forensic crime laboratories began looking into nuclear DNA as a potential means of human identification.

Genetic polymorphism of nucleated cells became of great interest to forensics after Alec Jefferys revolutionized forensics with the advent of DNA fingerprinting. Jefferys discovered minisatellites, called hypervariable loci, a highly variable DNA sequence that makes each individual unique (Butler 2011). To distinguish variations within the genome, the technique of Restriction Fragment Length Polymorphism (RFLP)

was developed. RFLP was a process where restriction enzymes cut DNA sequences in different fragments and were visualized by gel electrophoresis according to their length (Butler, 2010).

The discrimination power by RFLP analysis was high but the major drawback of RFLP was that it could not obtain a DNA profile from degraded DNA. Large amounts of samples, up to 500ng of DNA, were needed to produce a reliable profile (Ronny, 2010). RFLP allowed scientists to identify individuals based on their DNA polymorphisms.

In 1983, Kary Mullis transformed the forensic identification platform by discovering the use of a DNA polymerase to target a DNA sequence and make an infinite number of copies with pg of DNA. Together with the introduction of capillary electrophoresis, a sequencing technology, Short Tandem Repeats (STRs) have become the golden standard for DNA profiling today (Butler, Coble & Vallone, 2007). Forensic scientists are able to use genetic evidence derived from any biological material, including blood, semen, bone, hair, teeth, muscle tissue, and saliva for identification purposes. Multiplex STR kits have been developed to amplify up to 17 model loci and the sex-determining marker, Amelogenin, in one PCR reaction. These kits are able to generate accurate profiles from as low as 1ng of DNA (Butler, 2010). Some popular STR kits on the market today include Powerplex 18D by *Promega* and Identifier Plus by *Life Technologies*.

In 1994, the FBI was authorized by federal legislation to establish the Combined DNA Index System (CODIS) in an effort to digitally store DNA profiles of offenders and

missing persons. Currently, the database houses the 13 established STR loci. The FBI maintains the database which enables state and local forensic laboratories to exchange and compare DNA profiles for potential matches. To facilitate greater power of discrimination and missing person investigations, the FBI is considering adding additional loci to the common 13 loci currently present in CODIS.

## **1.2 Purpose of Research**

DNA based prediction of EVC of an unknown person who leaves DNA samples at a crime scene can be extremely useful to law enforcement agencies. Currently, the validated STR loci can identify the gender of the individual from a DNA sample but individual profiles are sometimes challenging to achieve if the samples are degraded or in very low quantities.

This research will discuss forensically relevant SNPs to identify physical traits such as skin, eye and hair color. Forensic scientists are employing different statistical and molecular methods to predict PISNPs. This paper will evaluate ethical and privacy issues that concern PISNPs. PISNP analysis using validated Next and Third Generation sequencing platforms, that can meet court system guidelines, will be explored.

## **1.3 Single Nucleotide Polymorphism**

While STRs can provide a DNA match and distinguish between two individuals, Single Nucleotide Polymorphism, SNP (also called “snip”), provides a reliable way to discriminate between two individual features. SNPs are a single base change in a DNA

sequence that occurs throughout the human genome. SNP variations are not distributed uniformly in the genome and one third of SNPs are within coding regions (Vignal et al., 2002). SNPs occur at least 1% every 100 to 300 bases in the human genome and makes up to 90% of human genetic variations. These sequence variations occur when a single nucleotide (A,T,C,or G) in the genome sequence is altered usually due to insertion or deletion events (Butler, 2011). An example is the following SNP:

Allele A (Individual 1)	ACGTGCA T CCACTAG
Allele A (Individual 2)	ACGTGCA C CCACTAG

Two thirds of all SNPs are called transitions where a base substitution occurs between pyrimidine (C, T) or between purines (A, G). When a base substitution occurs between purine and pyrimidine, it is called a transversion (Vignal et al., 2002). The most common SNPs are bi-allelic SNPs, where there are two nucleotide variation within the region of the DNA. Other common SNPs are tri-allelic and tetra-allelic, where the range of nucleotide variation is three and four nucleotides at a given region. Tri-allelic and Tetra-allelic SNPs are the rarest. SNP analysis is an emerging field of research in forensics. A rich set of SNP analysis platforms can potentially be developed as investigative leads for law enforcement agencies.

#### **1.4 Advantages vs. Disadvantages of SNPs**

Over several decades, efforts have been made to establish SNPs markers for forensic genetic testing. While SNPs may replace the common STR established markers, they have several advantages. The smaller amplicon size produced by polymerase chain



reaction (PCR), usually less than 100bp, is ideal for analyzing degraded DNA (Butler et al., 2007). Since SNPs are abundant in the human genome with low mutation rates, multiple typing platforms can be used to create high throughput assays for forensic applications. Forensic laboratories are also able to predict specific information such as identity, lineage, ancestry or phenotypic traits from SNP analysis. The major disadvantage of SNPs would be its inability to identify and separate mixtures (Butler et al., 2007). During mixture analysis with SNP panels, it is difficult to distinguish between heterozygotes or two homozygotes.

Aside from its inability to decipher mixtures, there are no commercial kits available for SNPs to multiplex degraded mixtures. Since multiple platforms are available, validation and concordance studies become extremely difficult for scientists to decide on a single SNP platform to use. Another disadvantage is that a string of SNP markers are needed to identify a specific trait. An average of 25-45 SNP loci are required to produce a random probability match as it is now achieved with 12 STR loci (Chakraborty et al. 1999). The last major challenge is the lack of SNP loci and a database to contain all the SNP markers. Currently, millions of STR genetic profiles are housed in CODIS but there are no current representations of SNP loci in the national DNA database.

### **1.5 SNP Categories in Forensics**

SNPs can be broken down into four categories for forensic use 1) Identity testing SNPs (IISNPs), 2) Lineage informative SNPs (LISNPs), 3) Ancestry Informative SNPs

(AISNPs), also known as AIMS, and 4) Phenotype Informative SNPs, also known as PISNPs (Budowle & van Daal, 2008).

### **1.5.1 Identity testing SNPs (IISNPs)**

To date, a great deal of study has been done on IISNPs. IISNPs serve the same purpose as the STRs. These SNPs are able to provide genetic information to differentiate between two individuals and exclude those who are not the source of the evidentiary DNA sample at a crime scene. To date, between 21 and 52 IISNP panels have been developed by European forensic laboratories. Additional studies are being conducted to validate identity SNPs and make the analysis platforms available for forensic laboratories for identity testing (Budowle & van Daal, 2008).

### **1.5.2 Lineage or Lineage Informative SNPs (LISNPs)**

LISNPs are sets of allelic markers used to identify relatives. These tightly linked SNPs function as multi-allelic markers and are found on the X-Chromosome, Y-Chromosome and mitochondrial DNA. LISNPs can be used to identify individuals in missing persons cases (Budowle & van Daal, 2008).

### **1.5.3 Ancestry Informative SNPs (AISNPs)**

AISNPs, also known as AIMS, can provide a high probability of an individual's ancestry as having originated from a specific geographical origin. In cases where an STR provides a reliable profile but the perpetrator cannot be found in the DNA database, AISNPs can be used to provide information on the person's biogeographic ancestry and,

in turn, give an indirect physical appearance. The DNA-Witness Kit developed by DNAPrint Genomics, is able to apply AISNPs to identify SNPs from four major geographical areas in the world. Version 2.0 of the DNA-Witness Kit that is based on 73 SNP panels was used to successfully solve several homicide cases in Louisiana and apprehend a serial killer (Budowle & van Daal, 2008).

#### **1.5.4 Phenotype Informative SNPs (PISNPs)**

PISNPs are defined as SNPs that provide a high probability that an individual has a specific phenotypic trait such as a particular skin color, hair color and eye color. PISNPs are of great forensic value to the scientists because of their ability to describe physical traits of individuals. The focus of this research is to identify and review forensically relevant PISNPs (Budowle & van Daal, 2008).

#### **1.6 SNP Database**

For SNPs to be available to forensic scientists for genetic phenotyping, they must be catalogued in a database. Currently no such database exists for forensic scientists. The SNP Consortium (TSC) was established in April 1999 in an effort to characterize a map of the human genome and catalog human genetic variations (Philip, 2005).

The group was led by Arthur L. Holden and included 10 large pharmaceutical companies and the U.K. Wellcome Trust. A total of 1.8 million high quality SNP markers were discovered and soon after the group changed its focus to finding SNPs in world population. The efforts led to the development of The International HapMap Project. The

HapMap project compares genetic sequences of individual populations to identify the locations of genetic variants. Several million characterized SNPs have been entered into the National Institute of Health (NIH) but no such database exists for forensic analysis.

Currently, the Scientific Working Group on DNA Analysis Methods (SWGDM) is evaluating locations of the genes where these SNPs are located and are cataloging PISNPs for forensic analysis.

Utilizing results from the HapMap Project, forensic scientists have discovered genes that can predict EVCs and have developed PISNP panels to predict phenotypic characteristics. Once scientists develop the new PISNP panels, they are ideally examined through Genome Wide Studies (GWS). Currently, the EVCs can be predicted with these PISNP panels for their forensic value (Kayser & Schneider, 2009).

## CHAPTER TWO: FORENSIC PHENOTYPING

### 2.1 Indirect Phenotyping

Biogeographical ancestry and ethnic origin of a person can be tested with the DNA found at the crime scene to indirectly establish a phenotypic profile of the individual (Figure 1). Indirect phenotyping is known to aid in a better understanding of skin color as well as features like hair and eye colors. A well-known commercial kit, DNA Witness – a multiplex assay, has been developed to test for the geographical origin of a person who's DNA has been found on a crime scene with close to 100% accuracy. Frudaski et al's research led to the development of the DNA Witness Kit. His team screened 211 SNPs in the pigmentation and xenobiotic metabolism genes. The study found that there were significant differences in allelic frequencies of 56 SNPs. The differences were in pigmentation genes of unrelated individuals of European, Asian and Africans descents. The group also developed a linear classification method to incorporate SNPs into a classifier model that could accurately identify European, Asian and African individuals with greater than 98% accuracy. This significant advance in forensic genetics has led to investigators solving crimes in Louisiana and California (Walsh, 2004).

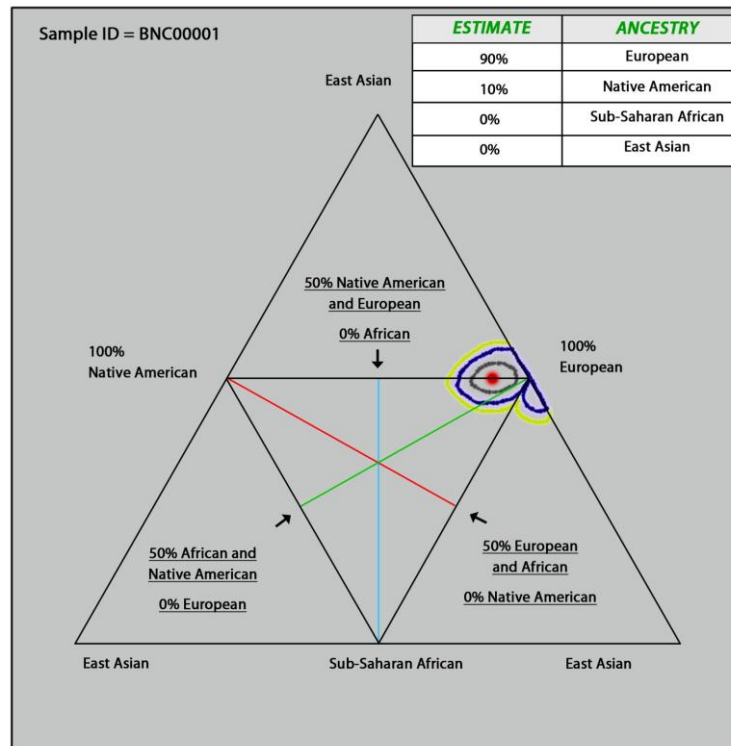


Figure 1- Budowle et al., 2008

## 2.2 Direct Phenotyping

While indirect phenotyping can provide ancestral origin of a person, direct phenotyping assesses genetic factors to determine how someone looks; the most obvious characteristic being the person's skin, hair and eye color. Indirect phenotyping may be useful, however it is limited and unable to distinguish between different characteristics such as hair color, eye color and skin color.

The wide range of pigmentation is known to be due to several genetic factors. Factors like genes associated with pigmentation play a major role in predicting the appearance of a perpetrator for investigative leads.

Forensic scientists are able to study polymorphic variations within the genes that control pigmentation and are able to predict a person's skin color, eye color and hair color. These traits are of great forensic value to the forensic community while analyzing DNA from the crime scenes.

The wide range of variation in skin pigmentation is due to factors like the shapes and sizes of melanocytes and melanin producing cells (Figure 2). The distribution of melanin differs in different regions of the body and is produced in the lower layer of epidermis in the skin, the iris of the eye and in the bulb of the hair. The melanin is produced in larger amounts in skin and hair than it is in the eyes.

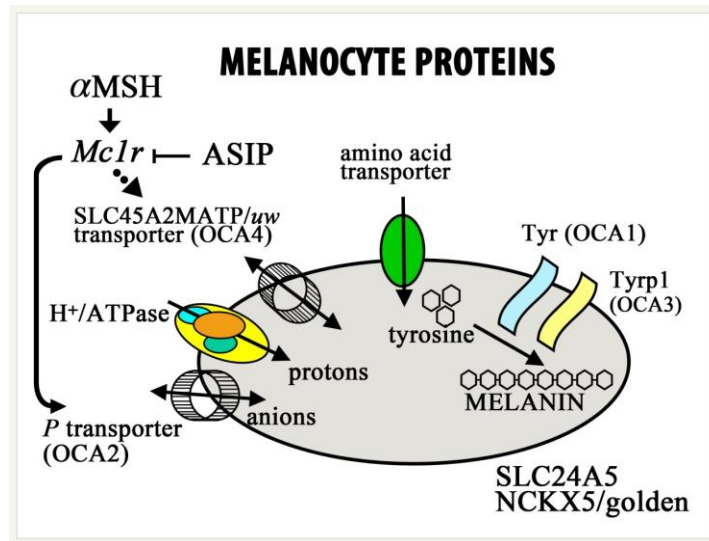


Figure 2- Melansome (Brilliant, 2008)

In the skin, the melanosomes are transported to the upper layers of the epidermis through keratinocytes with the help of dendrites. The different skin tones seen in humans are due to the number of melanocytes, melanin producing cells in the skin. In hair, however, melanosomes are in the follicular unit of the shaft. Human eyes have two tissue layers, iris pigment epithelium (IPE) and the Stroma. The outer layer, Stroma has a heavy distribution of melanosomes (Parra, 2007). The two common forms of melanin are eumelanin and pheomelanin. Eumelanin is responsible for black-brown pigmentation while pheomelanin produces the yellow-reddish pigmentation. Lighter or darker variations of pigmentation in humans are associated with number, size, and density of melanosomes.

Some important genes associated with pigmentation are the Melanocortin 1 Receptor Gene (MC1R), the “golden” gene (SLC24A5), Membrane-Associated



Transporter Protein (MATP), Agouti Signaling Protein gene (ASIP), and Tyrosinase gene (TYRA).

The forensic community has studied the MC1R gene extensively. This gene is associated with red hair color. The MC1R gene, located on chromosome 16, is a 951 bp long exon and plays an important role in pigmentation of skin, eye and hair. The MC1R gene is responsible for switching the production of eumelanin and pheomelanin on and off and ultimately controls the ratio of eumelanin-to-pheomelanin production in melanosomes. The gene has also been associated with light skin color pigmentation and red hair (Kanetsky et al., 2002; Voisey et al., 2006).

SLC24A5 has recently become a great forensic value. Lamason et al. identified the golden gene to be responsible for lighter skin color variations or hypopigmentation. The study was originally done in two types of zebrafish models, one mutant and another wild type. The mutants carried the so-called golden gene which caused a reduction in production of pigmentation. The delay in pigmentation resulted in the zebra fish having yellow stripes. When the system was applied to humans, 25–38% of diversity in skin melanin index between populations of European and African ancestry were explained with the golden gene. (Lamason et al., 2005).

MATP plays an important function in pigment depletion. MATP, also known as SLC45A2 or AIM1, is responsible for Oculocutaneous Albinism type 4 (OCA4). The polymorphism pattern is similar to the SLC24A5 gene and this gene is known to play an important role in pigmentation variations in human skin. Several SNPs within the gene are associated with dark skin and eye color (Valenzuela et al., 2010).

ASIP helps with the synthesis of pheomelanin and acts as an antagonist to the MC1R gene. This gene is related to significant variations in dark hair and brown eyes. SNPs at 3' UTR within the Agouti Signaling Peptide ASIP gene are associated with dark hair (Kanetsky et al., 2002; Voisey et al., 2006).

TYRA encodes for the enzyme called melanogenesis and is known to be responsible for type one albinism (Parra, E.J, 2002).

The following table summarizes the genes involved in different color pigmentations:

**Table 1- Genes associated with pigmentation**

<b>Skin</b>	<b>Eye</b>	<b>Hair</b>
ASIP	ASIP	ASIP
DRD2		
EGFR		
HERC2	HERC2	HERC2
IRF4	IRF4	IRF4
KITLG		KITLG
MATP		MATP
MC1R	MC1R	MC1R
	OCA2	OCA2
SLC24A4	SLC24A4	SLC24A4
TYR	TYR	TYR

As the table illustrates, pigmentation variations are controlled by several genes that are good predictors of human pigmentation. As one may see from the various PISNP

genes discussed above, pigmentation related genes overlap between hair, eye, and skin related variations.

The six genes that have been identified to be associated with variations in eye color are OCA2, MC1R, HERC2, SLC4A4 and TYR (Walsh et al., 2011). Molecular studies pertaining to several genes such as Melanocortin-1 Receptor (MC1R) gene and Agouti Signaling Peptide ASIP have given forensic scientists clues as to what affects these genes have in human skin variations. For example, both MC1R and ASIP are known to be associated with lighter skin tone (Sulem et al., 2007).

Statistical models have been developed by Robert Valenzuela to explain pigmentation SNPs and variations within pigmentation genes. Forensic Scientists are using these genes to create SNP panels to potentially simplify the process to develop a panel of markers allowing for accurate predictions of all three traits. More studies are needed to determine the additional genes associated with intermediate iris colors such as green or hazel eyes.

### **2.2.1 SNP Panels for Skin Color**

Dr. John H. Relethford carried out a study in 1997 in which he showed a strong correlation between skin pigmentation and geographical latitude. In his study, he demonstrated that the farther from the equator an individual lives, the lighter in pigmentation the individual's skin tends to be. In addition, positive selection has been known to have an influence on the distribution of pigmentation according to latitude gradient. Since our human ancestors migrated out of Africa 60,000 years ago, positive

selection has been thought to cause the light skin pigmentation in populations farther away from the equator (Sulem et al., 2007). Individuals close to the equator are known to have dark skin pigmentation and the variations in genes are affixed. The skin color variation has no other advantage only that light skin helps in vitamin D production in northerly latitude regions where the sun exposure is minimal (Parra, 2002).

The wide range of variation in skin pigmentation is due to factors like the distribution of melanosomes, shapes and sizes of melanocytes and melanin producing cells in the skin. Aside from genetic variations, environmental factors like Ultraviolet Radiation and age of the individual also play key roles in variations of pigmentation (Valenzuela et al, 2010). Variation in skin color can be measured by skin reflectance spectrophotometry and molecular assays. The first study, done on mice, showed mutations in 127 genes that were identified to affect pigmentation from which 68 genes are homologous to humans (Sulem et al., 2007). From the 68 genes, 29 cause Hermansky-Pudlak syndrome, a form of albinism in humans. The OCA2 and MC1R genes are known to cause fair skin while MC1R has been reported to cause freckles in humans. Genes like ASIP and TYR cause skin color pigmentation in individuals with mixed ancestry of Europe and Africa. The eSLC45A2 gene has been identified to cause skin pigmentation in African-American and African-Caribbean descents. Several studies have identified natural selection as the cause of different pigmentation in European and East Asian ancestry yet there is no evidence of natural selection in pigmentation variations in African descents (Sulem et al., 2007). Because the factors that affect skin pigmentation can be environmental (UV radiation, geography, etc.), PISNP panels related

to skin color are still being researched but several PISNP panels for hair and iris colors have been developed.

Sulem et al. completed the first GWA study on skin pigmentation. The group genotyped 317,511 SNPs in 2,986 Icelanders. This study aimed to find sequence variations that affect freckles and skin sensitivity in humans. Two variants, rs4959270 and rs1540771 at locus 6p25.3, were identified to be associated with freckles and sensitivity to ultra-violet rays (UVR). The frequency of rs1540771 is 50% in European populations and 30% and 5% in the East Asian populations. This suggests that the variation was caused due to positive selection in European populations. In the previous GWA studies, TYR was reported to cause fair skin, but the study done by Sulem et al. did not find any correlation between the TYR gene and lighter skin pigmentation. Strong association detected that the variation in TYR produced freckles in the population under study.

Forensic scientists are using evolutionary genetics and have found five SNPs that explain 82% of skin color variation in humans (Valenzuela et al, 2010). Though these markers are expected to be useful in individual skin color prediction models, there are no SNP panels to identify human skin pigmentation. There have been statistical models developed by Robert Valenzuela to explain pigmentation SNPs and variations within them. Because the factors that affect skin pigmentation are mostly environmental (UV radiation, geography, etc.), PISNP panels related to skin color are still being researched.

### **2.2.2 Case Study: The IrisPlex, an Intelligence DNA System**

The IrisPlex, also referred to as “DNA intelligence” system, is a sensitive multiplex assay that can predict blue and brown eyes with 90% precision. The assay is able to produce complete SNP genotypes with 31pg of DNA. Along with smaller and degraded amounts of DNA, the assay is able to produce statistical prediction models for eye colors. The IrisPlex was validated using six PISNPs on thousands of European individuals and was also recently tested with a Genome Wide Association Studies (GWAS) group for its reliability, sensitivity and reproducibility. The assay meets the SWGDAM guidelines which makes it ideal for forensic testing (Walsh et al., 2011).

### **2.2.3 Case Study: Regression Statistical Model to accurately predict eye color variations in Dutch Individuals**

Liu et al performed regression statistical model analysis on 37 SNPs from eight genes in a population study consisting of 6,168 Dutch individuals. In this study, prediction performance accuracy was measured by studying the area under the receiver characteristic operating curves also known as AUC. The specification to accurately predict a specific eye color is 1. Amongst the 37 SNPs, 13 SNPs were removed from the study because of their strong linkage and disequilibrium with other SNPs. This statistical study showed high values of AUC for brown and blue eyes at >90%. The study also showed lower AUC values of 0.73 for intermediate colors. Unidentified SNPs for intermediate eye colors could result in lower AUC values in this study. The multinomial logistic regression model was performed as well and yielded the same results of AUC of

>0.90 for blue and brown eyes and 0.73 AUC value for intermediate eye colors. The genetic prediction values from this study have been highly accurate and Kayser et al believes that this statistical data accurately predicts genes associated with blue and brown eye colors and can be of great use in forensic studies (Liu et al., 2009).

#### **2.2.4 Case Study: Human Eye Color and HERC2, OCA2 and MATP/SLC45A2 genes**

2,718 Icelanders and 1,214 Dutch were selected as subjects to carry a study of understanding iris color variations due to SNPs at HERC2, OCA2 and MATP/SLC45A2 genes. The study concluded that 90% of the individuals had blue eyes when the prediction stats for whom brown eye color was <20% and green color was <10% probability. Additionally, individuals who were predicted to have brown eyes between a 45-70% probabilities, in fact did have brown eyes (Kayser & Schneider, 2009).

#### **2.2.5 Case Study: Model Based Prediction of Human Hair Color Variation**

In this study, 25 SNPs from Polish individuals were genotyped to investigate their predictive values as it may be used in forensic cases. Both molecular (mass spectrometry using Sequenome multiplexing, SNaPshot) and statistical assays were used. In this study, prediction performance accuracy was measured by AUC. The specification to accurately predict a specific eye color is 0.5 to 1.0. The study measured AUC values of two sets of hair color categories. The first set comprised of four hair color categories that included Blond, Brown, Red and Black hair. The second set of hair category was comprised of

Blond, D-blond, Brown, Auburn, B-red, Red and Black. The AUC values for the set of four hair categories had values of 0.81 to 0.93. The AUC values for the second set of samples had values of 0.78 to 0.92. Each set predicted red hair with a higher AUC value of 0.93 in the first set and 0.92 in the second set. Both sets of data showed lower AUC values for blond and brown hair colors. The higher prediction accuracies of the red hair could be due to age. Individuals who are born with blond hair often have it change to brown during adolescence. Since only adults were used in this study, individuals with blond genotypes showed lower AUC values for brown hair (Branicki et al, 2011).

#### **2.2.6 Case Study: MC1R gene indicative of Red Hair**

The multiplex solid-phase minisequence assay called SNaPshot assay was developed to detect SNPs in MC1R gene which is associated with red hair color. Forensic Science Service (FSS) in England conducted this study. The agency tested SNPs within the melanocyte MC1R gene of 76 red-haired and 108 non-red-haired individuals. The R151C and R160W polymorphism was apparent in 97.9% of the individuals who had red hair. The analysis of MC1R gene can be very helpful in forensic cases; however, if a certain population has a large number of red-haired individuals, this system cannot be used by itself. For an individual to have red hair, they must be the heterozygote for the two allele variations. Since there is a strong difference in red hair color frequency, it mostly depends on the population it is applied to; therefore it may not be practical to do red color prediction models in areas where red-haired individuals are minimal (Grimes et al., 2001)



## CHAPTER THREE: CURRENT AND FUTURE GEN “OMICS” PLATFORMS

### 3.1 Current “Omics” Technologies

Over the last decade, several platforms have been developed to predict PISNPs.

The following table summarizes some of the important SNP platforms available for forensic analysis today (Sobrino, Brion & Carracedo, 2004).

Table 2- Current SNP Typing Platforms

SNP Typing Methodologies	Description
Allele Specific hybridization	Probes anneal specifically to certain alleles with the use of different color beads. Methods include Homogenous hybridization using FRET, TaqMan Assay, LightCycler and Molecular Beacons.
Primer Extension	DNA polymerase incorporates specific nucleotides complimentary to the DNA template. Methods include Minisquencing and Pyrosequencing.
Allele specific oligonucleotide ligation (OLA)	Distinguishes SNPs by using a set of three oligonucleotides in combination with a thermostable Taq and DNA ligase enzymes.
Invasive Cleavage	Two Oligonucleotides called the invader oligonucleotide and probe anneal to a specific DNA region.

### **3.1.1 MiniSequencing**

In Minisequencing, unlabeled oligonucleotide primers are used to anneal nucleotides to a target region of the DNA with the help of polymerase (Figure 3). The sequence reaction requires an amplification step where the region of interest is amplified. The PCR product is then cleaned to rid it of all the dNTPs and primers. Minisequencing is capable of multiplex reactions which make it useful to the forensic laboratories. In multiplex reactions, multiple primer combinations are used in a single tube reaction format. The product is analyzed in a single electrophoresis run which is ideal for investigating SNP sites. Currently, electrophoresis-based technologies, like SNaPshot, and fluorescence detection technologies, like Matrix Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF MS), are being utilized to investigate SNPs.

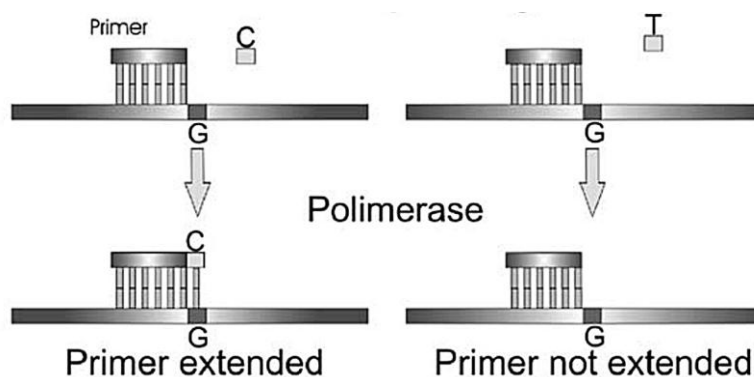


Figure 3- Minisequencing (Sobrino et al., 2005)

SNaPshot assay is extremely popular in forensic laboratories (Sobrino, Brioín & Carracedo, 2005). SNaPshot assay is based on primer extension assay that utilizes fluorescently labeled dinucleotides. The assay can identify multiple SNP sites that can be easily multiplexed. SNaPshot is based on a minisequencing technology, which uses the same genetic analyzer as that for STR assay. Capillary Electrophoresis units like the ABI 3130xl or the new 3500 by Life Technologies is standard among forensic scientists due to its STR capabilities. In addition to the capillary electrophoresis units, data can be analyzed through the GeneMapper ID or Gene Mapper IDX by Life Technologies, (Butler, 2011).

MALDI-TOF MS, has the potential to become the gold standard in SNP genotyping amongst forensic scientists (Figure 4). The MALDI-TOF MS is the most

direct detection assay because a SNP site can be determined by separating DNA fragments based on their molecular weight. The major drawback of the technique is difficulties in reproducibility of data and issues in validation studies (Sobrinho et al., 2005). The technology is also very laborious and time consuming and may not be suitable for high throughput laboratories.

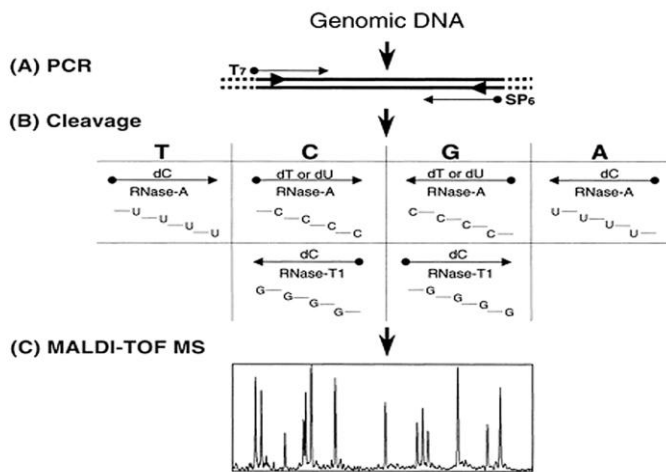


Figure 4- An Example of MALDI-TOF MS (Stanssens et al., 2004)

### 3.1.2 Pyrosequencing

Pyrosequencing uses an enzyme cascade system to produce fluorescence when a nucleotide is added to the complementary strand SNP of interest. This system is also called sequence-by-synthesis (Figure 5). Many Next Generation Sequencing technologies have also been established based on a sequence-by-synthesis system which will be discussed later in this research.

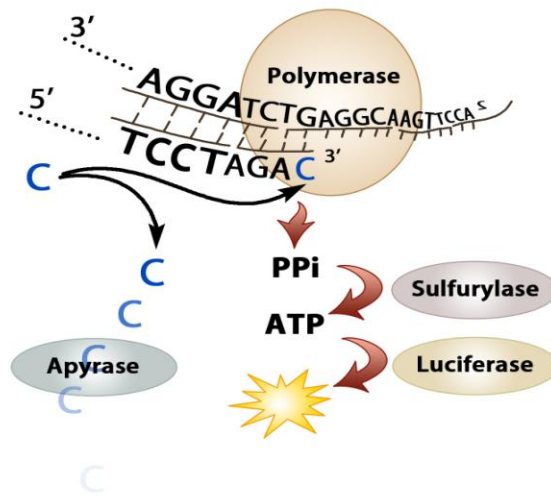


Figure 5- Pyrosequencing Cascade System retrieved from [www.lifetechnologies.com](http://www.lifetechnologies.com)

Pyrosequencing is capable of providing PISNPs in real-time but it requires many preparation stages including PCR cleanup steps that are arduous and time consuming. Pyrosequencing is difficult to validate and it does not have multiplex capabilities. Several steps must be performed before data is detected (Sobrino et al., 2005).

### 3.2 Future “Omics” Technologies

Next gen “omics” platforms, such as Illumina Genome Analyzer, Roche 454, and ABI SOLiD technologies, though expensive at this time, can revolutionize forensic phenotyping. These platforms take the place of the Sanger sequencing methods which are employed in forensic laboratories to test the hypervariable region of mitochondria also known as mtDNA analysis. The technique is used to distinguish between two

samples from low quantity DNA and differentiate between maternally-related individuals (Berglund, Kiialainen & Syvanen, 2011).

For the application of sequencing in forensic cases, the forensic community has discussed de novo sequencing and re-sequencing projects. De novo sequencing applies when the genome of an organism is sequenced for the first time. In re-sequencing projects, a reference sequence of the genome already exists and a comparison is done by evaluating the available sequences to the one in question (Berglund et al., 2011). For phenotyping, re-sequencing will be more appropriate since reference genome will already be established in a database. Three major next generation sequencing technologies, Illumina, 454 and SOLiD platforms exist in the market today and are competing aggressively to capture the attention of the forensic community. Two major technologies, Heliscope Single Molecule Sequencer System and Single Molecule Real Time Sequencing (SMRT), are among third generation sequencing technologies preferred by forensic scientists.

### **3.2.1 Sequencing Principles for Illumina, 454 and SOLiD platforms**

The Illumina and 454 platforms use the sequence-by-synthesis (SBS) technology in their platforms (Figure 6). In Illumina, DNA polymerase extends primers using fluorescently reversibly terminating nucleotides to the complementary strand of DNA. When the light signal is detected, the base is registered and additional nucleotides continue being added. Unique fluorescently labeled nucleotides enable each base to be incorporated which results in true base-by-base sequencing. This platform has been

widely adopted and allows for the most accurate data collection today (Berglund et al., 2011).

The 454 sequencers are based in pyrosequencing technologies that allow non-terminating deoxynucleotides to be added sequentially to the complementary DNA strand. Pyrophosphate is also released as the nucleotide incorporation releases energy in the form of ATP. Luciferase then uses the ATP to produce light and the light is registered to a charge-coupled device (CCD) camera (Berglund, Kiiialainen & Syvanen, 2011). The platform produces a longer read with high quality scores since the nucleotide substitution errors are lower. While the technology produces a longer read, it can be used to perform de novo sequencing to create reference materials in forensic applications. Since SNPs are caused by insertion-deletion of nucleotides, the lower the insertion-deletion errors, the more accurate the data output. The 454 sequencer can be excellent in screening and detecting of phenotypic SNPs for forensic applications (Berglund et al., 2011).

The SOLiD system (Figure 6) uses a sequence-by-ligation method where fluorescently labeled probes consisting of eight bases are added to an oligonucleotide chain. Five of these bases are specific to the complementary DNA template and three universal bases are used to maintain hybridization with the template. During sequencing reactions, primers, DNA ligase and fluorescently labeled probes are added. The complementary probe ligates to the DNA template and is registered at four different wavelengths. Once the universal bases are cleaved off, new probes are added to repeat the cycle continuously until all the bases in the template are sequenced twice. Since each

nucleotide is sequenced twice to reduce rates, the sequence-by-ligation system such as SOLiD can be a potential standardized platform for forensic laboratories. However, the system produces smaller read lengths, usually 75 nucleotides per read, and analysis can be potentially laborious and time consuming.

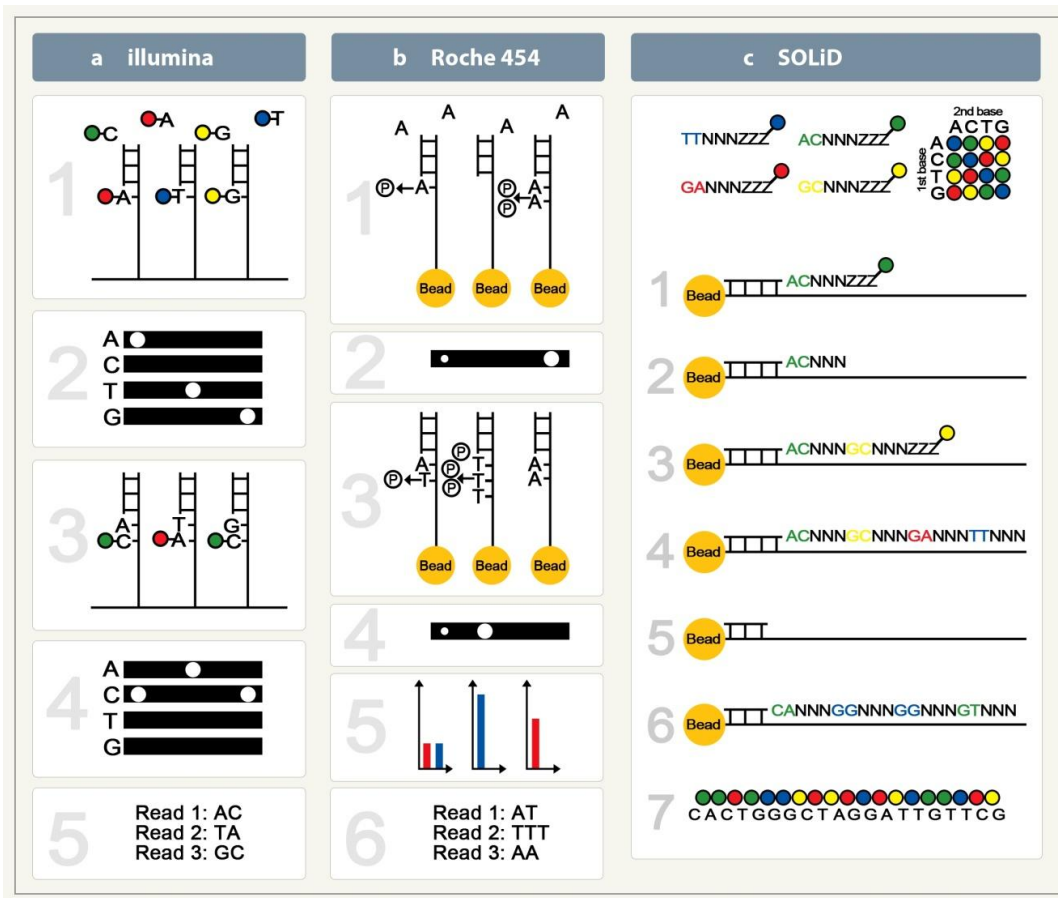


Figure 6- Three major NGS chemistries (Berglund, Kiiialainen & Syvanen, 2011).

In Illumina 454 systems, light intensities from sequencing reactions are converted into bases. The SOLiD system uses a coloring system from the sequencing reaction to



convert to bases. The data is also produced with quality scores of each base. Quality scores are important in forensics because it estimates the true presence of a certain nucleotide at a certain position. All sequences must include control DNA and control beads to ensure that there are no errors in the system. Sequences are also trimmed at 3' and 5' during data analysis to remove low quality data. Once the data cleanup and filtering is complete, the sequence data must match with that of the reference genome.

Since SOLiD systems use color calling, some filtration systems or control DNA cannot be used. The system uses color intensity to make nucleotide calls; therefore, high quality data may not be recovered. Uncertainties in Quality Control (QC) checks can leave room for errors and the SOLiD system may not be useful to forensic laboratories since quality parameters of crime laboratories are extremely stringent.

Illumina has announced their sequence service cost for as low as \$20,000 and their competitors, like Life Technologies, are pushing for the price to come down to \$1,000 (Ledergerber & Dessimoz, 2011). If this platform is widely available due to affordability, forensic laboratories will be able to produce the complete genome sequence of the suspect from very small amounts of DNA and create a comprehensive phenotypic profile of the suspect including the skin, eye, and hair colors. Some downside to forensic phenotyping using next generation platforms will be increasing backlog if SNP phenotyping becomes part of normal forensic testing. Another issue is cost. Even if one sample cost is as low as \$1,000 an STR profile is much less expensive and will be more likely chosen due to budgetary constraints.

### 3.2.2 Third Generation Sequencing

Third-generation sequencing platforms use single molecular sequencing with no amplification step or template purification steps prior to sequencing. The technologies allow direct sequencing of the site of interest with improved accuracy. The sequence reads are much longer and are useful in determining SNP variants within several genes of interest. Since the data produced is free of PCR induced biases and errors, there is also a very limited need for statistical model predictions, which are being employed in PISNP analysis today (Pushkarev, Neff, & Quake, 2009). Currently, two major technologies exist in the market; Heliscope Single Molecule Sequencer system from Helicos Biosciences and Single Molecule Real Time Sequencing (SMRT) by Pacific Biosciences.

The Heliscope Single Molecule Sequencer system (Figure 7) employs sequence-by-synthesis from next generation platforms where DNA samples are fragmented first and the DNA is melted into a single strand. A proprietary surface is used to capture the DNA molecules. Fluorescently labeled nucleotides A, C, G or T are added to the complementary DNA with the help of polymerase. The incorporated nucleotides are captured by the laser as they emit light. Each run generates 35 gigabytes (GB) of data, and the read length is 35 nucleotides on average (Pushkarev et al., 2009).

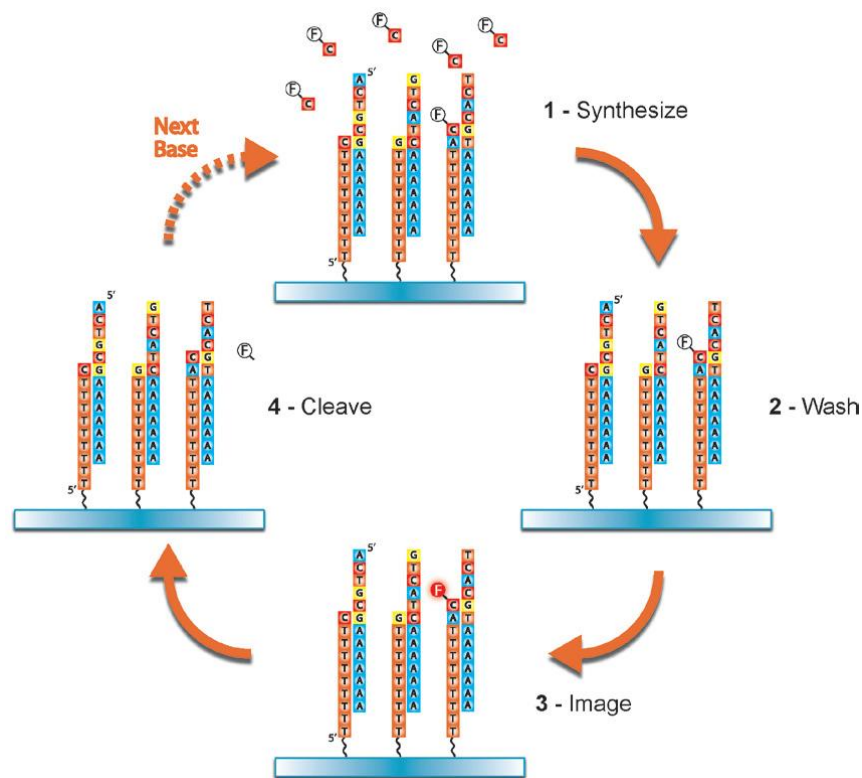


Figure 7- Helicos System (Berglund et al., 2011)

The Pacific Biosciences system is called Single Molecule Real Time (SMRT) sequencing and monitors the incorporation of fluorescently labeled nucleotides in real time. The technology is based on sequence-by-synthesis where nucleotides are fluorescently labeled and are incorporated with the help of DNA polymerase on a proprietary glass surface containing polymerase. The system is able to generate a thousand bases per run but has a higher error rate than the Heliscope Single Molecule Sequencer.

Several additional single molecular sequencing systems are under development. These new systems can produce long high quality reads, which also minimize error rates.

These particular characteristics of NGS methods are attractive to crime laboratories because they allow direct determination of PISNPs. Although standard STR profiling provides discrimination power for individual identification, the use of next or third generation sequencing provides multiple advantages to forensic laboratories.

## **CHAPTER FOUR: ETHICAL, PRIVACY CHALLENGES AND DAUBERT GUIDELINES**

### **4.1 Ethical and Privacy Issues**

DNA based EVC prediction is increasingly becoming prevalent having already been accepted in countries like United Kingdom and the Netherlands (Kayser & Schneider, 2009) despite prohibitions in Belgium. The United States has no federal legislation on forensic phenotyping. However, states such as Wyoming, Rhode Island and Indiana prohibit using DNA to obtain any physical information. The laws in all three states collectively indicate that, “DNA samples and DNA records collected shall never be used for the purpose of obtaining information about physical characteristics, traits or predispositions for disease” (Koops & Schellekens, 2008). While Vermont does not prohibit phenotyping in forensic investigations, its laws strongly prohibit using DNA to gain information on genes that could cause genetic disorders.

The United Kingdom’s General Legislations allow using forensic phenotyping in criminal cases but there is no general supervision or legislative framework. The country is considered as one of the most forward thinkers in forensic phenotyping after the impressive effort by the UK’s Forensic Service (FSS) team. The FSS offered services such as the red hair test, which examined the MC1R gene to detect up to 84% probability if a particular individual has red hair (Koops & Schellekens, 2008).

Aside from the red hair tests, the FSS developed software and additional statistical tests that can predict a suspect's ethnicity. This service was used in 2004 when the UK investigators revealed that they were looking for an individual from African-Caribbean descent (Koops & Schellekens, 2008).

The Netherlands is also very proactive in using DNA for phenotypic purposes. The Dutch Criminal Code Procedure was amended in 2003 to add a clause which states that DNA obtained from the crime scene can be used to gain further information on the phenotypic characteristics of the suspect. However, the law limits the use of the information to external traits only.

Germany's article 18e allows DNA to be used for gender identification, parentage and to identify if the DNA sample is from the victim or the suspect. The laws prohibit using DNA for any other information such as phenotyping (Koops & Schellekens, 2008).

The world's government units are uncertain about using the information from DNA for anything other than gender identification, distinguishing DNA from victim to that of the suspects, and paternity. In the United States, three states prohibit phenotyping but the Netherlands and the UK are proactive and continually developing new PISNP panels.

Privacy advocates argue against profiling EVC from DNA left at a crime scene. Since SNPs are able to provide phenotypic information, privacy advocate groups argue that the EVC traits are against privacy rights and could possibly discriminate against an individual due to physical traits. However, according to the scientific group led by

Manfred Kayser, samples from the crime scene belong to an unknown suspect and therefore have no basis to violate any privacy laws (Butler, 2011).

## **4.2 Acceptance of PISNPs and Daubert Guidelines**

Agencies like the National Institute of Justice (NIJ) and the National Institute of Science and Technology (NIST) are discussing the impacts of rapidly advancing next-generation DNA sequencing technologies and PISNP panels for forensic applications. Government agencies such as the FBI have planned to invest time and money to discuss using these novel technologies in forensic identification. The goal is to understand NGS applications for human identification in forensics and test the platforms for their reproducibility and robustness. There are ongoing efforts by NIST to develop multiplex panels to differentiate pigmentations. If these technologies were introduced into laboratories they must pass stringent accredited crime laboratory guidelines. Once the technologies are identified, they must meet Daubert requirements. According to Daubert, certain requirements must be satisfied if evidence is to be classified as scientific knowledge and admissible in court (Kiely, 2006). There are five requirements as follows:

- Is there any proof of the method being tested based on underlying technique the process was developed,
- Has the method been peer reviewed and published,
- What are known or potential error rates,
- The existence maintenance of standards,

- Has this method been generally accepted by the forensic community.

These requirements must be fulfilled in order to consider PISNPs for identification and for next or third generation technologies to be introduced in criminal investigations. STR loci have been the standard for the last decade. Extensive research and validation is required to introduce these novel technologies in government crime laboratories and to admit evidence of this nature in courts.



## **CHAPTER FIVE: CONCLUSION**

In recent years, GWA studies have identified PISNPs that can help predict phenotypic traits like skin, eye and hair color. This has great value to forensics scientists, since such information can help in identifying a criminal without the need of a reference sample. Phenotypic traits, including skin, eye and hair color, can be accurately predicted. A future option can be to replace STR based testing with high throughput sequencing assays. The major complication for such a system will be replacing the data in current forensic databases like CODIS. Several million profiles are housed in the current databases and a large effort will be required to create a new database for SNP classes such as PISNPs. The current sequencing technologies are expensive and there are few PISNP SNP panels to identify human traits. Thus, replacing STR with new SNP panels will require extensive research and validation.

Third generation sequencing technologies like single molecule sequencing help facilitate the detection of multiple PISNP loci. The possibility of detecting multiple donors and contamination can help scientists provide testimony confidently in courts.

While many technologies have been developed over the past several decades, researchers must come to a consensus over which SNP technology to employ. Sensitivity, reproducibility, and accuracy of the assays must be determined first before making it available to the forensic community. In addition, the phenotypic profiles generated and

model-based prediction statistics must be catalogued in a database. At this time, no such database exists. Technical factors as well as Daubert Guidelines need to be considered if PISNPs are to be included in forensic identification.

Currently, the use of “omics” technologies allows scientists to predict red hair, blue and brown eyes. Research in finding predictive DNA markers for other EVCs is ongoing.

## REFERENCES

## REFERENCES

- Baker, M. (2010). Next-generation sequencing: Adjusting to data overload. *Nature Methods* 7, 495-499.
- Bastiaens, M., Bastiaens, M., ter Huurn, J., Gruis, N., Bergman, N., Westendorp, W., et al. (2001). The melanocortin-1 receptor gene is the major freckle gene. *Human Molecular Genetics*, 10, 1701-1708.
- Berglund, E.C., Kiialainen, A. & Syvanen, A.C. (2011) Next generation sequencing technologies and applications for human Genetic History and Forensics. *Investigative Genetics*, (2) 23.
- Branicki, W., Brudnik, U., Kupiec, T., Wolańska-Nowak, P., Wojas-Pelc, A. (2007). Determination of phenotype associated SNPs in the MC1R gene. *Journal of Forensic Science*, 52, 349-354.
- Branicki, W., Liu, F., van Duijn, K., Draus-Barini, J., Pośpiech, E., Walsh, S., Kupiec, T., Wojas-Pelc, A., Kayser, M. (2011). Model-based prediction of human hair color using DNA variants. *Human Genetics*, 129(4), 443- 54.
- Brilliant, M. (2008) Gene Polymorphism and human pigmentation. Retrieved from <https://www.ncjrs.gov/pdffiles1/nij/grants/223980.pdf>
- Butler, J.M. (2011). *Advanced topics in forensic DNA typing: Methodology*. (First ed). San Diego; USA, Elsevier Academic Press.
- Butler, J.M., Coble, M.D., & Vallone, P.M. (2007). STR vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Science Medicine Pathology*, 3, 200-2005.
- Budowle, B., Bieber, & Van Daa, A. (2008). Forensically relevant SNP classes. *BioTechniques*, 44 Supplement, 603-610.
- Budowle, B. (2004). SNP typing strategies. *Forensic Science International*, 146S, S139-S142.

- Frudakis, T., Thomas M., Gaskin Z., Venkateswarlu, K., Chandra, K., Ginjupalli, S., et al. (2003). Sequence associated with human iris pigmentation. *Genetics*, 165, 2071-2083.
- Forensic DNA ethics. Retrieved from <http://forensicdnaethics.org>
- Grimes, E. A., Noake P. J., Dixon, L., Urquhart, A. (2001). Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Science International*, 122, 124-129.
- Haff, L. A., & Smirnov, I. P. (1997). Single-nucleotide polymorphism identification assays using thermostable DNA polymerase and delayed extraction MALD-TOF mass spectrometry. *Genome Research*, 7, 379-388.
- Helicos Biosciences. Retrieved from <http://www.helicosbio.com>
- International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, 437, 1229-320.
- Irwin, J., Just, R., Scheible, M., Loreille, O. (2011). Assessing the potential of next generation sequencing technologies for missing persons identification efforts. *Forensic Science International, Genetics Supplement Series*, 3, 447-448.
- Kayser, M., & Schneider, P.M. (2009). DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical consideration. *Forensic Science International: Genetics*, 3(3), 154-161.
- Kayser, M. & De Knijff, P. (2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nature reviews. Genetics*, 12, 179-192.
- Kidd, K.K., Pakstis, A.J, Speed, WC., Grigorenkoe EL., Kajuna SL., Karoma NJ., et al. (2006) Developing a SNP panel for forensic identification of individuals. *Forensic Science International*, 164, 20-32.
- Kiely, T.F. (2006) *Forensic Evidence: Science and the Criminal Law* (Second ed.). Boca Raton, FL. Taylor and Francis.
- Koops, B. J., & Schellekens, M. (2008) Forensic DNA phenotyping: Regulatory issues. *The Columbia Science and Technology law Review*, 9, 158-202.
- Lamason, R. L., Mohideen, M., Mest, J.R., Wong, A.C., Norton, H.L., Aros, M.C., et al. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*, 310, 1782-1786.

- LifeTechnologies: Single Stranded Template for Pyrosequencing retrieved from <http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/DNA-RNA-Purification-Analysis/napamisc/Capture-of-Biotinylated-Targets/Single-Stranded-Templates-for-Pyrosequencing.html>
- Liu, F, Duijn, K., Vingerling, R. J., Hofman, A., Uitterlinden, G., Janssens. A. J.W.C., Kayser, M., (2009). Eye color and the prediction of complex phenotypes from genotypes. *Current Biology*, 19(5). R192–R193.
- Lao, O., de Gruijter, JM., van Duijn, K., Navarro, A., Kayser M. (2007). Signature of positive selection in genes associated with human skin pigmentation as revealed from analysis of single nucleotide polymorphism. *Annals of Human Genetics*, 71, 354-369.
- Mengel-From J., Børsting, C., Juan J. S., Eiberg H., Morling N. (2010). Human eye colour and HERC2, OCA2 and MATP. *Forensic Science International Genetics*, 203, 323-328.
- Parra, E., J. (2007). Human Pigmentation Variation: Evolution, Genetic Basis, and Implications for Public Health. *YearBook of Physical Anthropology*, 50, 85-105.
- Perkel, J. (2008). SNP genotyping: Six technologies that keyed a revolution. *Nature Methods*, 5, 447-453.
- Philips, C. (2005). Using online Databases for Developing SNP Markers of Forensic Interest. *Methods in Molecular Biology: Forensic DNA Typing Protocols (Vol: 297, pp.85-95)*. Humana Press.
- Pushkarev, D., Neff, N., F., Quake, S., R. (2009) Single-molecule sequencing of an individual human genome. *Nature Biotechnology*, 27, 847-852.
- Relethford, J.H. (1997). Hemispheric difference in human skin color. *American Journal of Physical Anthropology*, 104 (4), 449–457.
- Ronny, D. (2010) Genetic identification in the 21st century- Current status and future developments. *Forensic Science International*, 201 (1-3), 160-164.
- Roche: Roche 454. Retrieved from <http://www.my454.com>
- Sobrino, B., Brión, M., Carracedo, A. (2005) SNPs in Forensic genetics: A review on SNP Typing methodologies. *Forensic Science International*, 154, 181-194.

- Stanssens, P., Zabeau, M., Meersseman, G., Remes, G., Gansemans, Y., Strom, N., et al. (2004). High-Throughput MALDI-TOF Discovery of Genomics Sequence Polymorphism. *Genome Research*, 14, 126-133.
- Sulem, P., Gudbjartsson, F. D., Simon, N. S., Helgason, A., Thorunn, R., Magnusson, P., et al. (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genetics*, 39, 1443 – 1452.
- Sundquist, T. (2010) Forensic Phenotyping: What DNA can (and cannot) tell us about a criminal's appearance. Promga Connections. Retrieved from <http://promega.wordpress.com>
- The SNP Consortium (TSC) and now International HapMap Data. Retrieved from <http://snp.cshl.org>
- Thorisson, G. A., & Stein, L.D. (2003). The SNP Consortium website: past, present and future. *Nucleic Acid Research*, 31, 124-127.
- Tully, G. (2007). Genotype versus phenotype: Human pigmentation. *Journal of Forensic Sciences*, 1, 105-110.
- Valenzuela, K, R., Henderson, S, M., Walsh, H, M., Garrison, N, A., Kelch, J, T., Orit-Barak, C., et al. (2010). Predicting Phenotype from Genotype: Normal Pigmentation. *Journal of Forensic Sciences*, 55 (2), 315–322.
- Vignal, A., Milan, D., SanCristobal, M., Eggen, A. (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34(3), 275-305.
- Walsh S., Liu F., Ballantyne, N., Kaye O., van M., Lao, O., Kayser, M., (2011). IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics*, 5(3), 170-180.
- Walsh S., Lindenbergh, A., Zuniga, S, B., Sijen, T., de Knijff, P., Kayser M., Ballantyne K., N., (2011). Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics*, 5, 464-471.
- Walsh, S.J. (2004). Recent advances in forensic genetics. *Expert Review of Molecular Diagnostics*, 4(1), 31-40.

## **CURRICULUM VITAE**

Khatera Noor Harrif graduated from Herndon High School, Herndon, Virginia, in 2001. She received her Bachelor of Science degree in Biology from George Mason University in 2005. She is currently employed at American Type Culture Collection (ATCC) as a Senior Molecular QC Biologist.