FORENSIC DNA PHENOTYPING AND MASSIVE PARALLEL SEQUENCING

by

Krystal Breslin

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Biological Sciences

Indianapolis, Indiana

December 2017

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Susan Walsh, Chair
Department of Biology
Dr. Kathleen Marrs
Department of Biology
Dr. Benjamin Perrin
Department of Biology

Approved by:

Dr. Stephen Randall
Head of the Graduate Program

Dedicated to my husband and two sons. Thank you for always giving me the courage to follow my dreams, no matter where they take us.

ACKNOWLEDGMENTS

I would like to acknowledge and thank everyone who has helped me to become the person I am today. I would like to begin by thanking Dr. Susan Walsh, my PI and mentor. Your passion for science is contagious and has been instrumental in helping to inspire me to pursue my goals. Without your trust and guidance, I would not have been able to grow into the scientist I am today. You are a true role model and I am eternally grateful to have worked with you for the last three years. I cannot wait to see where the future takes you and your research.

Secondly, I would like to thank Dr. Ben Perrin and Dr. Kathy Marrs for being a valuable part of my advisory committee. You are both very impressive scientists, and I truly value the input you have provided to my thesis.

I would also like to acknowledge and thank each and every member of the Walsh laboratory: Charanya, Ryan, Noah, Morgan, Mirna, Bailey, Stephanie, Lydia, Emma, Clare, Sarah, Megan, Annie, Wesli, Kirsten, Katherine, and Gina. Without each one of you this would not have been possible. Charanya, I honestly don't know what I would have done without you. You have grown to be not only a lab mate, but one of my very best friends. We have spent many late nights and early mornings bouncing ideas off each other and I would never have made it to where I am today without you. I cannot wait to see what the future holds for you, I truly believe you can change the world. I look forward to working with you after you finish your doctorate. Ryan, thank you for all your help, especially when it comes to computers. I will miss our chats, especially those about music. Good luck in pursuing your desire to work in academia, you will be phenomenal.

To my husband, Chris- thank you for putting up with me for all of these years. You have had a ridiculous amount of patience with me as I have pursued my dreams. Without your guidance, support, and love I would not be who I am today. Thank you for standing by me through everything, I could not have picked a better partner for this crazy life, I love you!

To my children, Clayton and Grayson- thank you for always putting what is important into perspective for me. It is a true honor to be your mom and I hope that one day I make you both proud. I cannot wait to see the people you become. I love you both!

To my family- Anita (Mom) and Kyle, I love you both! Thank you for always being there through all this craziness that we have seen. Your support and encouragement has helped me every step of the way. Kyle, you are not only my brother, but also one of my very best friends and I hope that never changes. I hope I make you both proud!

To my grandpa- William Clifton, thank you for being a wonderful grandfather. Without your love and I would not have made it this far. Thank you for all you have done for me!

To my best friends- Sam, Devin, Kyle, and Joe. Thank you for always sticking with me through the good times and bad. Thursdays would not be the same without you all. Sam, you are truly the very best friend anyone could ask for, you might as well be my sister. Thank you for always being someone I can talk to about anything. I enjoy our long talks and I swear sometimes you can read my mind. I would never have made it through undergrad or grad school without your support. Devin, thank you for being there all this time. We have been through a lot together, but I truly am grateful for your friendship. Without your music playlist, this thesis would have never been written! Kyle, you are a true inspiration to me, especially with all you have achieved in your short career. I cannot wait to see what the future holds for you, I know you are going to make a big impact in your field. Thank you for inspiring me and pushing me to work harder. Joe, thank you for always knowing how to lighten the mood and make the day that much better.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Author: Breslin, Krystal, A. MS
Institution: Purdue University
Degree Received: December 2017
Title: Forensic DNA Phenotyping and Massive Parallel Next Generation Sequencing
Committee Chair: Susan Walsh

In the forensic science community, there is an immense need for tools to help assist investigations where conventional DNA profiling methods have been non-informative. Forensic DNA Phenotyping (FDP) aims to bridge that gap and aid investigations by providing physical appearance information when other investigative methods have been exhausted. To create a "biological eye witness", it becomes necessary to constantly improve these methods in order to develop a complete and accurate image of the individual who left the sample.

To add to our previous prediction systems IrisPlex and HIrisPlex, we have developed the HIrisPlex-S system for the all-in-one combined prediction of eye, hair, and skin color from DNA. The skin color prediction model uses 36 variants that were recently proposed for the accurate prediction of categorical skin color on a global scale, and the system is completed by the developmental validation of a 17-plex capillary electrophoresis (CE) genotyping assay that is run in conjunction with the HIrisPlex assay to generate these genotypes. The predicted skin color output includes Very Pale, Pale, Intermediate, Dark and Dark-to-Black categories in addition to categorical eye (Blue, Intermediate, and Brown) and hair (Black, Brown, Blond, and Red) color predictions. We demonstrate that the HIrisPlex-S assay performs in full agreement with guidelines from the Scientific Working Group on DNA Analysis Methods (SWGDAM), achieving high sensitivity levels with a minimum 63pg DNA input.

In addition to adding skin color to complete the pigmentation prediction system termed HIrisPlex-S, we successfully designed a Massively Parallel Sequencing (MPS) assay to complement the system and bring Next Generation Sequencing (NGS) to the forefront of forensic DNA analyses methods. Using Illumina's MiSeq system enables the generation of HIrisPlex-S's 41 variants using sequencing data that has the capacity to

better deconvolute mixtures and perform with even more sensitivity and accuracy. This transition opens the door for a plethora of new ways in which this physical appearance assay can grow as sequencing technology is not limited by variant number; therefore, in essence many more traits have the potential to be included in this one assay design. For now, the HIrisPlex-S design of 41 variants using MPS is being fully assessed according to SWGDAM validated guidelines; therefore, this design paves the way for Forensic DNA Phenotyping to be used in any forensic laboratory. This new and improved HIrisPlex-S system will have a profound impact on casework, missing persons cases, and anthropological cases, as it is relatively inexpensive to run, HIrisPlex-S is easy to use, developmentally validated and one of the largest systems freely available online for physical appearance prediction from DNA using the freely available online web tool found at https://hirisplex.erasmusmc.nl/.

Lastly, moving forward in our aim to include additional traits for prediction from DNA, we contributed to a large-scale research collaboration to unearth variants associated with hair morphology. 1026 samples were successfully sequenced using an in-house MPS design at 91 proposed hair morphological loci. From this reaction, we were able to contribute to the identification of significant correlations between the SNPs rs2219783, rs310642 and rs80293268 with categorical hair morphology: straight, wavy or curly.

CHAPTER 1. INTRODUCTION

1.1 Purpose and Objectives

A forensic DNA profile is what many prosecutors hinge their cases and possible convictions upon. The current standard for DNA profiling in crime laboratories in the United States is STRs, or Short Tandem Repeats. These are used to develop and compare unknown profiles obtained at a crime scene to known suspect profiles obtained by detectives from a suspect individual or through the use of a national database like CODIS, the Combined DNA Indexing System [1]. STRs are an excellent method of determining the perpetrator of the crime when a suspect is known and an unknown sample is retained. This standard, however, is extremely unhelpful when a sample has been obtained from the crime scene, but there are no suspects in the case, no "hits" in national databases like CODIS, or when the DNA is too badly degraded for an STR profile to even be obtained. This is where technologies like Forensic DNA Phenotyping, or FDP, are vital to the progression of forensic science. FDP enables the prediction of an individual's externally visible characteristics from a sample left at a crime scene, therefore acting as a "biological eyewitness" [2].

FDP bypasses the limitations of DNA databases and ancestry testing alone and looks at specific locations in the genome that are correlated either directly or indirectly for normal variation in physical appearance between individuals. Researchers have developed systems like HIrisPlex and Snipper 2.5, that, through the use of Single Nucleotide Polymorphism (SNP) Multiplexing, can predict head hair pigmentation, skin pigmentation, eye pigmentation, and many other possible traits, with the potential for additional traits to be added through continued research [2-5]. These types of assays use SNPs, which are single base changes that occur at a specific spot in the genome or INDELS which are insertions or deletions that occur in a specific region of the genome. Not only do these types of variants offer new leads, but they also allow profiles to be generated from DNA material as low as 60pg/µl, which is significantly less than the published ranges needed for a typical STR profile [2, 6].

In 2014 the HIrisPlex system, a Forensic DNA Phenotyping tool released by Walsh *et al.* [7], and its validation were published for public use. This system is a combination of a SNP- SNaPshot all-in-one multiplex assay and prediction model for two externally visible characteristics; eye and hair pigmentation. Six of the SNPs that were incorporated into HIrisPlex were from a previous interactive model, termed IrisPlex [7]. IrisPlex is an eye pigmentation prediction tool developed by the same research group which classifies an individual's eye pigmentation as being either blue, brown, or intermediate purely based upon their DNA [7]. This assay is fully compatible with SWGDAM guidelines, who govern forensic assays and validations [7, 8]. A measurement of accuracy termed AUC (Area Under the Receiver Operating Curve) was used to illustrate how accurate a prediction model such as this performs, where 0.5 denotes the random toss of a coin and a value of 1 equals a perfect prediction. The AUCs for eye pigmentation were 0.95 for Brown, 0.94 for Blue, and 0.74 for Intermediate [8]. The HIrisPlex system combines eye pigmentation prediction with hair pigmentation prediction for a total of 24 DNA variants. The hair pigmentation portion of the prediction includes the categories of black, blond, brown, and red. It also incorporates light or dark hair shade into the model. This all-in-one multiplex assay was fully developmentally validated per SWGDAM guidelines for use in any forensic laboratory, giving AUC values for this model's hair pigmentation prediction were 0.93 for red, 0.85 for black, 0.81 for blond, 0.75 for brown [2]. The HIrisPlex system has not only been tested on data sets to prove its versatility, but it has also been used in casework and ancient remains to prove its practicality. A trial was conducted on 26 ancient, degraded samples of teeth and bones that are between 1 and 800 years old (post mortem). 23 of these samples yielded full HIrisPlex profiles [9]. These examples illustrate the reliability and necessity of implementing such a system in especially difficult casework where investigators are potentially limited by the lack of possible leads to pursue.

While FDP is incredibly valuable in casework, continued research is required, especially with regards the addition of new externally visible characteristics such as skin pigmentation and hair morphology, for example. In Capillary Electrophoresis (CE) SNP analysis, which is the current technology used for FDP, Taq polymerase orchestrates the synthesis of millions of copies of a template DNA strand while incorporating

fluorescently tagged dideoxyribonucleotide triphosphates. These florescent tags are identified by the genetic analyzer based upon their different wavelengths and therefore result in the fragment being read by the analyzer. The biggest drawback to designing these assays on the CE is that the chemistry of the SNaPshot technology used caps the number of SNPs in an assay at around 25-30 SNPs. This means that adding in SNPs for new traits would mean adding in extra runs, more time, and more materials, which is an immense drawback as a forensic tool where sample and time may be very limited. MPS, also known as Massively Parallel Sequencing, is a technology that is changing the face of genetic research and could solve this problem. This relatively new form of sequencing differs from the typical Sanger method and also whole genome sequencing in that it allows a much higher throughput of samples by choosing many pieces, called amplicons, of DNA to sequence and not the entire genome. Machines like Illumina's Miseq and ThermoFisher's Ion Torrent, allow multiple fragments of the genome to be sequenced simultaneously, therefore resulting in the cost being drastically lower than the previous methods of individually sequencing genomes [10] or Sangers singular fragments. The ability of this technique to produce results from hundreds to thousands of variants in a single run is an ideal tool for forensic researchers as it allows the capacity to expand the number of variants required in FDP (in addition to ancestry estimation) to assist law enforcement investigations.

Based upon the current state of FDP, the goal of this research is three-fold. The first goal is to assist in completing the design and validation of the HIrisPlex-S system and its model. This system uses 41variants to predict not only an individual's eye and hair pigmentation, but their skin pigmentation as well, on a 5-category scale ranging from very pale to dark-to-black. The second goal of this research is the design and execution of a custom MPS Assay that translates the CE based HIrisPlex-S system to the Illumina Miseq FGx allowing FDP to be taken to the next level via this new platform. The final goal of the research is the design and execution of a custom MPS assay on the Illumina Miseq FGx to assess 91 SNPs for their correlation with categorical hair morphology association i.e. straight, wavy and curly.

1.2 The Pigmentation Pathway, Including Skin and Eye Melanin Formation and Distribution

Pigmentation is one of the key distinguishing features between humans especially when looking at eye, hair, and skin pigmentation. This pigmentation is formed through melanin, which originates from neural crest cells that arise during the embryonic development and follow a very specific pathway. Research stemming from animal model studies has given us a better understanding of this pathway. In humans, α-*MSH* stimulates Melanocortin-receptor 1 (*MC1R)* causing the activation of the cAMP pathway. *MC1R* is a single exon and mutations and insertions in this gene tend to be associated with red hair, light skin, and freckles. *MC1R* controls the downstream regulation of the cAMP pathway. Down regulation of this results in low *TYR* and therefore a higher production of pheomelanin [11]. However, if *ASIP*, an antagonist, binds to *MC1R* instead of α-MSH, it shifts the synthesis towards pheomelanin by reducing *TYR*. Without *ASIP* the process tends toward eumelanin which requires further enzymatic activity [10].

MITF is then introduced into the pathway resulting in the expression of *TYR*, *TYRP1* and *DPC*. Tyrosinase catalyzes, and is the limiting reagent to, the oxidation of tyrosine to DOPAquinone [12]. Tyrosinase (*TYR*) is an enzyme that catalyzes steps one and two of melanogenesis and is required for both eumelanin and pheomelanin production. Up regulation of this enzyme (often due to a neutral pH, as opposed to an acidic pH which results in pheomelanin) results in the synthesis of eumelanin. Tyrosinase-related protein (*TYRP1*) follows and is required for eumelanin only [11]. Both types of melanin; eumelanin and pheomelanin, are derivatives of 3,4di-hydroxyphenylalanine [11].

From this point in the pathway, the resultant melanin is based upon the presence or absence of Cysteine. Cysteintyldopa is produced and stagnant until Cysteine is present for it to oxidate to [11]. This therefore initiates small molecule transport, ion transport, and pH regulation. Oculocutaneous Albinism II (*OCA2*) encodes the p protein, which is involved in the transfer of small molecules to initiate melanogenesis. Solute carrier families (*SLC*) transport proteins *SLC24A4* and *SLC24A5* are sodium-calcium transporters that are potassium dependent, the latter being involved in the regulation of calcium within the melanosome. Two-pore segment channel 2 (*TPCN2*) participates in a

similar function as well [11]. In the end, eumelanin is responsible for darker pigmentations such as brown-black pigment where as its counterpart, pheomelanin, is responsible for the lighter pigmentations such as the red-yellow pigment. The difference in the amounts (ratio) of pheomelanin to eumelanin then portrays the visible differentiations in eye, hair, and skin pigmentation [11].

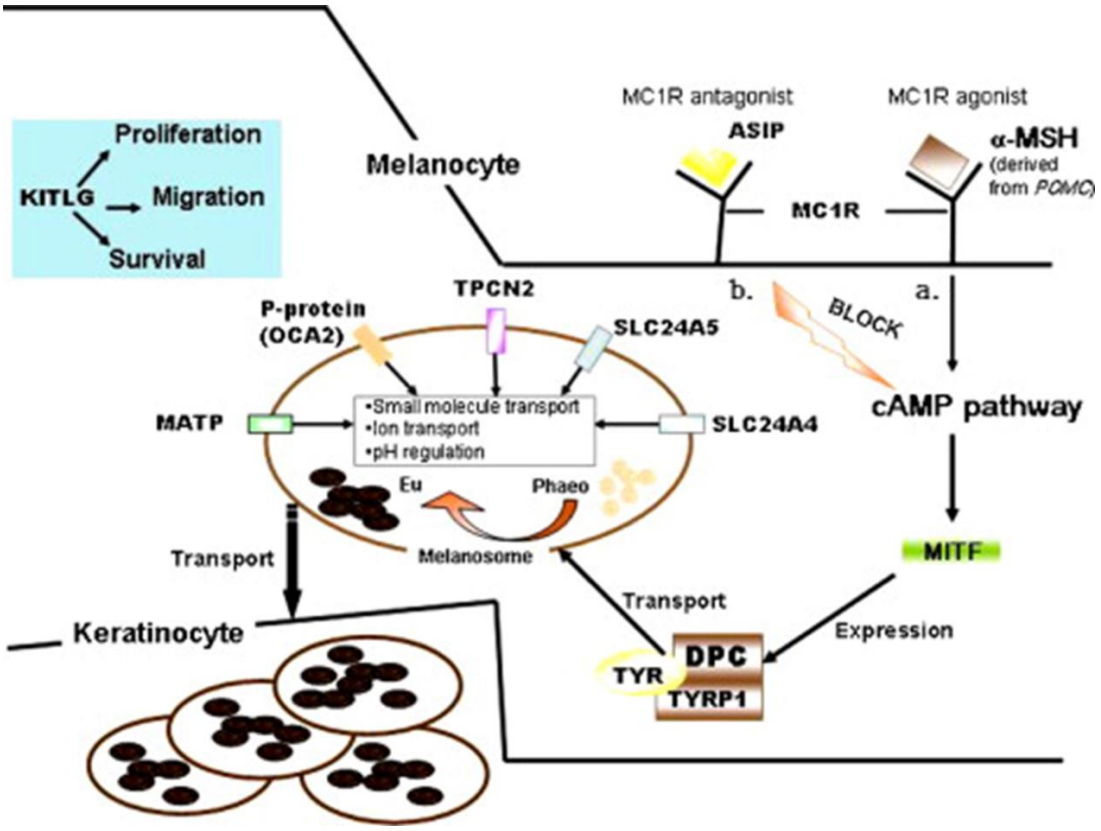


Figure 1 The biochemical pathway and synthesis of pheomelanin and eumelanin

The process by which melanocytes develop is a four-step process. In the beginning, vacuoles arising from the endoplasmic reticulum and bud to form premelanosomes. The premelanosomes then take in enzymes such as *TYR* and structural proteins to begin melanin synthesis. The product of this step is melanosomes. At this stage eumelanosomes require more enzymes along with a fibrillar matrix over pheomelanosomes. At this point, only these melanosomes proceed to steps three and four of the process. These eumelanosomes import *TYRP1* and *DCT* to continue. Melanosomes then become transferred to the keratinocytes near already formed melanocytes. This mode of transportation to the keratinocytes is not certain, but the leading hypothesis

involves the release of these melanosomes in which they are taken in by the keratinocytes by phagocytosis [11].

For melanin formation in the skin, melanocytes in the epidermis reside in the basal layer along with keratinocytes. Melanosomes migrate from the melanocytes to the keratinocytes. The keratinocytes then rise to the upper layers of the skin to produce the visible skin pigmentation that is apparent to the naked eye [12]. The skin pigmentation of an individual depends on several key factors such as size, cellular distribution, number, and the type of melanosome that are present in the epidermis [11]. Figure 2 obtained from Yamaguchi *et al* depicts the development and distribution of melanosomes in human skin [13].

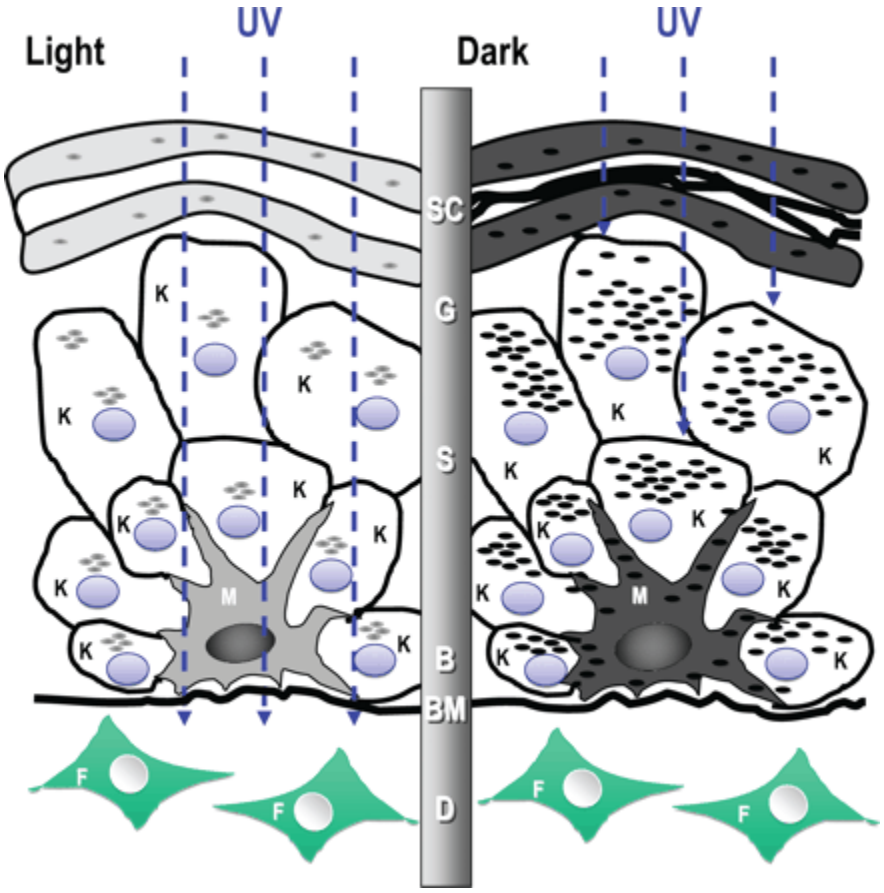


Figure 2 Yamaguchi *et al.* depiction of the development and distribution of melanosomes in human skin. *SC*, stratum corneum; *G*, stratum granulosum; *S*, stratum spinosum; *B*, stratum basale; *BM*, basement membrane; *D*, dermis. Cell types: *K*, keratinocyte; *M*, melanocyte; *F*, fibroblast; *shaded oval*, melanin granule

Hair color is similar in development to skin color in many ways. The melanocytes are located in the proximal bulk of each hair bulb as well as the sebaceous gland. The melanocytes transfer the melanosomes to keratinocytes similarly as in the skin, which then move to the shaft of the hair to produce visible pigmentation [12]. The life cycle of the melanocytes in the hair die at the end of the hair cycle, which only tends to last about 3-8 years as opposed to epidermal melanocytes, which have a much longer life span. The melanogenisis of the hair only takes place during the anagen stage, which is the growing stage of the hair. This process is turned off in the catagen phase and completely absent in the telogen phase [12].

Melanin formation in the eye differs in its formation from that found in hair. In eye pigmentation, blue and brown eyes have a similar number of melanocytes, but the visible pigmentation difference resides in the number of melanosomes that make up the melanin content of these melanocytes in the stroma. This lack of melanin in blue eyes allows for the penetration of light through the stroma, resulting in only a small amount being absorbed and the rest scattered [14]. This therefore gives the eye the appearance of blue, as melanin content increases we begin to see the hazel to intermediate eye color and proceeding to brown eyes.

1.2.1 Some Evolutionary Theories of Pigmentation Lightening

Melanin content is vital in the protection from radiation, especially UVB radiation, from the sun as well as necessary for Vitamin D absorption. It has been proven that both environmental factors like the sun and genetic factors can play a role on this externally visible characteristic's pigmentation along with the distribution and quantity of melanin [15]. Evidence supports the idea that our ancestors originated in Africa about 150,000 years ago [16]. Around 55-60,000 years ago, our ancestors began to migrate out of Africa. Through mitochondrial and Y-chromosomal analysis, they have been traced out of Africa and into the Eurasian continent. As they moved away from the equator their needs for protection from ultra-violate radiation and climate protection changed due to the need to absorb vitamin D amongst others [15]. Based upon this change we can see a variation in the skin pigmentation as latitude changed.

In addition to this theory, evidence in recent research especially by that of Norton *et al*. [16] has supported the idea of certain traits, especially eye, hair, and skin color may also be the product of convergent evolution over thousands of years. Genes such as s *SLC24A5*, *MATP*, and *TYR* have been identified in resulting in lighter pigmentations in individuals of European descent, but not of East Asian descent [17]. This supports the idea that while some genes such as *OCA2* and *ASIP* may play a key role in overall global skin pigmentation distribution, there is also evidence to suggest lighter pigmentation arose independently in certain populations.

Recently in 2017 Crawford *et al.* [17] published a study on African genomes that supported a hypothesis that challenges the idea of lightening as migration occurred from the equator. Their work with the genes *MFSD12*, *DDB1*, *OCA2*, and *HERC2* supports the idea that the ancestral allele of these SNPs actual encodes for lighter skin pigmentation. This then follows the theory that darker pigmentation originated in the last 2 million years after our ancestors lost their protective hair covering leading to the exposure of their lighter skin. However, they also present evidence to support the idea that the darker pigmentation in South East Asian populations may indeed result from a common African ancestor whose migration route led them out of African and into East Asia. Ultimately, it may be possible that both factors of convergent evolution and natural selection are behind the pigmentation of the skin [18].

1.3 Categorical Eye, Hair and Skin Pigmentation Associated Markers

For this research, the previously developed HIrisPlex system is at the core of development for these additional assays. Therefore, it is vital to understand the SNPs that comprise HIrisPlex. In 2014, Walsh *et al.* published the HIrisPlex assay as well as its forensic validation. Each of the 24 SNPs included in this multiplex assay were selected because they provide a high correlation and prediction for eye and/or hair pigmentation of the individual at hand. This assay includes 11 different genes *MC1R, HERC2, OCA2, SLC45A2, KITLG, EXOC2, TYR, SLC24A4, IRF4, ASIP*, and *TYRP1*. 6 of these markers were established in the IrisPlex system [7, 8, 19]. The largest contributor to the prediction of eye pigmentation is rs12913832 in the *HERC2* gene. This gene carries a substantial amount of weight as it has been proposed to be a causal marker for eye color

prediction [21] due to its regulation of *OCA2*, while also affects hair and skin pigmentation [19-21] to a lesser degree. The presence of an ‘A’ allele, Adenine, at this SNP enhances the expression of *OCA2* while a ‘G’, Guanine, has been shown to downregulate its expression [22]. The increase in regulation, the ‘A’, results in an increase in melanin product, which results in darker eye pigmentation (brown). Mutations in *OCA2* have been linked to ocular albinism, which results in the lack of pigmentation in the Iris, therefore indicating the inclusion of rs1800407 [17].

Sturm *et al.* then identified by performing Genome Wide Association Study (GWAS) that *TYRP* also has a large effect on eye pigmentation followed by *SLC45A2* [23]. The three remaining SNPs of rs12896399 (*SLC24A4*), rs1393350 (*TYR*) and rs12203592 (*IRF4*) were then also identified and associated with eye pigmentation by Han *et al.* in 2008 on over 7,000 people from three studies, but also labeled as being associated with hair pigmentation and skin pigmentation [24]. These six SNPs mentioned above were confirmed as the six major genetic contributors for eye pigmentation by a systematic study by Liu *et al.* [19] as well.

The *MC1R* variants of rs1805006, rs11547464, rs1805007, rs1805008, rs1805009, rs1805005, rs2228479, and rs885479 were associated with red hair and pale skin from a study containing family members and twins conducted by Sturm *et al.* [25]. In addition to these *MC1R* SNPs *N29insA* (rs312262906), *Y152OCH* (rs201326893), and rs1110400 were added to a mini-sequencing protocol for a sequencing screen for the red hair phenotype and were found to be influential on this phenotype by Grimes *et al.* [26] in 2001.

Eiberg *et al.* [26] looked at a large Danish family to perform linkage analysis to fine-map the blue eye pigmentation locus. Variants such as rs12913832 and rs1129038 were associated with blue and brown eye pigmentation, which was previously recognized in its location of intron 86 of *HERC2* just upstream of *OCA2* [27]. Rs683 (*TYRP1*), rs1042602 (*TYR*), was identified by Frudakis *et al.* to be an influencing factor in eye pigmentation [28] therefore explaining the addition of some of these SNPs to the HIrisPlex assay.

In 2008 Han *et al.* [23] found in a study of over 10,000 individuals of European ancestry from the United States and Australia, that rs12896399 (*SLC24A4*), rs12821256

(*KITLG*), and rs12203592 (*IRF4*) had a strong association with hair pigmentation. Rs12203592 (*IRF4*) was also associated with eye and skin pigmentation along with skin tanning ability [24].

Mengel-From *et al*. [30] found in a study of 33 candidate genes from a population of Danish and Scottish individuals, rs26722 (*SLC45A2*), rs7170852 (*HERC2*) rs916977 (*HERC2*), rs1129038 (*HERC2*), rs11636232 (*HERC2*) and rs2238289 (*HERC2*), rs1470608 (*OCA2*) to be associated with darker hair pigmentation and rs10777129 (*KITLG*) to be associated with lighter hair pigmentation [30]. Valenzuela *et al*. [19] also found three of these SNPS- rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), and rs12913832 (*HERC2*) to be important to the total hair melanin content [19].

In 2011 Branicki *et al.* [28] published their thorough results of a hair color systematic study performed using Polish Europeans in which they assessed all known hair pigmentation markers and found that 13 genetic markers from 11 genes predicted hair pigmentation with an overall accuracy of 0.8 Area Under the receiver-operating Curve (AUC). Of these SNPs, in addition to the *MC1R* SNPs, they recommended the use of rs12913832 (*HERC2*) and rs1800414 (*OCA2*) for eye pigmentation. Rs12913832 (*HERC2*) was also recommended for hair pigmentation in that it showed a dominant effect on darker hair pigmentation [29]. Black hair was most commonly associated with rs28777 (*SLC45A2*), rs12203592 (*IRF4*), and rs4959270 (*EXOC2*). *SLC45A2* was associated with hair pigmentation melanin content on a whole as well. Rs2378249 (*ASIP/PIGU*) was associated with red, dark blond, and red blond hair. Rs4778138 (*OCA2*) was found by this group to be highly associated with brown hair along with Rs1393350 (*TYR*). Rs4904868 (*SLC24A4*) however was largely associated with blond and dark blond hair [29]. Thus, allowing the HIrisPlex hair and eye pigmentation prediction model to be complete.

In addition, the same group in 2014, Pospiech *et al*. [29] substantiated and added to these findings by performing an analysis of 718 Polish sample, which yielded 38 polymorphisms from 13 genes. Several of these SNPs, such as the previously mentioned rs312262906, rs11547464, rs885479, rs1805008, rs1805005, rs1805006, rs1802007, rs2228479, rs203132689, rs1110400, and rs1805009, all of which are part of *MC1R*, were found to be associated with red hair along with rs4911414 in *ASIP* [30].

Rs12913832 (*HERC2*) was identified again as having an association with blue, brown and hazel eyes as well as blond hair. Rs12203592 (*IRF4*), rs16891982 (*SLC45A2*), and rs2402130 (*SLC24A4*) were also found to have a statistically significant effect on blond/red/brown/black hair pigmentation [30].

Conversely, the approach taken to determine the genes and SNPS influencing skin pigmentation required a very different approach. Many of the genes identified relating to skin pigmentation were identified through positive selection. This was necessary because it is advisable to perform GWAS studies on single populations for skin color, to avoid population stratification where as in hair and eye color this was not a concern due to its variation being restricted to a singular population such as European and its immediate surrounding area. To further understand the genetic factors behind this trait and how they were found, in 2007, eight genes were identified to be associated with the melanin pathway by Lao *et al.* [15]. These genes included *SLC45A2*, *OCA2*, *TYRP1*, *DCT*, *KITLG*, *EGFR*, *DRD2*, and *PPARD*. The work from this group helped to support the idea that derived alleles correlate to lighter skin pigmentation while the ancestral alleles correlate to darker skin pigmentation. They found that rs3782974 (*DCT*), rs1800414 (*OCA2*), rs1448484 (*OCA2*), rs2762464 (*TYRP1*) and rs16891982 (*SLC45A2*) were all associated with skin pigmentation as well as a signature of positive selection [16].

Simultaneously, in 2007, a novel GWAS was performed on a South Asian population by Stokowski *et al*. During this GWAS, this group discovered three genes *SLC24A5*, *TYR*, and *SLC45A2*. These were found to have highly significant associations with skin pigmentation in addition to the previous finding by Lao *et al.* [30]. They discovered that SNP rs1426654 (*SLC24A5)* does not change among European populations, but rather is solely related to individuals outside of Europe, such as the South Asians, whereas rs16891982 *SLC45A2* and rs1042602 *TYR* displayed an increase in minor allele frequency with a decrease in latitude [31].

Duffy *et al.* [31] in 2010 added to these findings on skin pigmentation by exhibiting that rs28777, rs35391, and rs16891982 in *SLC45A2* illustrated a minor association with an increased risk of cutaneous malignant melanoma, but not any more so than the previously observed in the *MC1R* mutation [32]. For skin reflectance rs16891982, (*SLC45A2*), rs1426654 (*SLC24A5*), rs2434984 (*ASIP*) were found to be

influential by Valenzuela *et al.* [21], but only with 45% R^2 value, therefore making the addition of more skin pigmentation prediction SNPs essential.

Pospiech *et al.* [29] also found that certain *MC1R* mutations influenced skin pigmentation, along with *TYR* and *SLC45A2, e*specially SNPs rs1042602 (*TYR*), rs1393350 (*TYR*), rs16891982 (*SLC45A2*) and rs28777 (*SLC45A2*). This association was followed shortly by rs4911414 (*ASIP*) and rs1800407 (*OCA2*) [30] .

Jacobs *et al.* [32] took a novel approach to pigmentation research and digitally quantified skin pigmentation in 5,860 Dutch individuals from which they looked at 14,185 SNPs from 281 candidate genes. Two new genes were identified as significant during this study- *UGT1A* and *BNC2,* that had not been previously tied to skin pigmentation. Previously, *UGT1A* has been shown to influence bilirubin and the levels of this in the blood. *BNC2*, specifically SNP rs10756819, was also shown to be significant. This gene was selected as a candidate gene because of its association to coat pigmentation in mice with a mutation resulting in a loss of brown pigment. Similarly this was seen in a disrupted stripe pattern in zebrafish when a mutation is present in this gene as well [33].

In 2017 Crawford *et al*. [17] addressed an essential area of human pigmentation in which more information is needed. Crawford *et al*. looked at 1570 African genomes to identify variants that were associated with skin pigmentation to possibly gain more insight to the origins of human pigmentation. Their research indicated that *SLC24A5* was introduced in to the East Asian population through migration routes out of Africa. Genes *MFSD12*, *DDB1*, *TMEM138*, *OCA2*, and *HERC2* were found to be significant in relation to skin pigmentation [18]. Thus 29% of all pigmentation in Africans was identified to be in relation to these SNPs and possibly increasing the knowledge base as to the rise of lighter versus darker skin pigmentation [18].

1.4 Categorical Skin Color Predictive Markers

In terms of predicting skin color using these associated markers, Hart *et al*. [5] in 2013 developed an 8-plex system, mentioned briefly above, which included rs12896399 (*SLC24A4*), rs12913832 (*HERC2*), rs1545397 (*OCA2*), rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), rs885479 (*MC1R*) and rs6119471 (*ASIP*), and rs12203592 (*IRF4*)

[5]. Six of these SNPs, rs12913832 (*HERC2*), rs1545397 (*OCA2*), rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), rs885479 (*MC1R*) and rs6119471 (*ASIP*), were used to designate skin pigmentation to be light, not dark, or not light. Based upon the European test set used, this assay's prediction was seen as being 94% accurate therefore establishing one of the earlier skin pigmentation prediction tools [5] however their high level of inconclusive predictions were a cause for concern [34, 35].

Liu *et al.* [20] also performed a genome wide association study in 2015 shortly following the previously mentioned publications of prediction models on a total of 17,262 European individuals, which replicated and furthered many of the findings previously mentioned [20]. They then followed this GWAS up with a functional study. This 2015 GWAS identified five regions of the genome that were believed to be associated with skin pigmentation. These areas include the genes *SLC45A2*, *IRF4*, *OCA2*, *HERC2*, *MC1R*, and *ASIP*. From this study, they could identify the top 9 influencing factors on skin pigmentation. These SNPs include: rs183671 (*SLC45A2*), rs12203592 (*IRF4*), rs10756819 (*BNC2*), rs1393350 (*TYR*), rs17128291 (*SLC24A4*), rs12913832 (*HERC2*), rs2924567 (*SLC24A5*), rs4268748 (*MC1R*), and rs6059655 (*RALY/ASIP*). Based upon their functional study using melanocyte cell lines, these individuals could determine that the expression of *ASIP* in the skin occurs in the dermis rather than the epidermis. It is possible that units in the dermis such as fibroblasts secrete *ASIP*, which then interacts with *MC1R*. This supports the idea that not only the epidermis, but also the dermis plays a role in pigmentation [36].

In 2014 Maronas *et al.* [3] published their own forensic skin pigmentation predictive tool. Using 285 samples from unrelated Europeans and non-Europeans, skin pigmentation associated SNPs were tested. These SNPs were identified using patterns of allele frequency along with literature reviews. A total of 59 SNPs were originally isolated and then narrowed down to sets of 10 and 6 SNPs [3]. The ten SNP set has SNPs that are present in 8 pigmentation genes. SNPs rs16891982 (*SLC45A2*) and rs1426654 (*SLC24A5*) were included in this assay based upon their predictive ability in the lighter skin colors. The SNP rs1426654 (*SLC24A5*) was noted as playing a large role in differentiating black and white to the rest. The effect of white versus the rest was also replicated in rs6119471 (*ASIP*). The final set of SNPs used for the reaction were rs10777129, rs13289,

rs1408799, rs1448484R, rs16891982, rs2402130, rs3829241, and rs6058017, in addition to those mentioned above [3].

1.5 Hair Structure Formation

Another aspect of this thesis involves fundamental research on the structure of human head hair. Hair growth occurs in a cyclic pattern that tends to last anywhere from 3 to 8 years [12]. The first phase of the cell cycle is the Anagen phase, which is a phase of growth for the hair follicle and the phase in which melanogenisis occurs. This is the longest phase of the cycle. The length of this phase varies based upon the location of the hair on the body [37]. This phase is followed by the Catagen phase in which separation of the follicle occurs and melanogenisis ends. The Telogen phase then follows and includes a phase of resting without any sign of melanin production. The final phase, the Exogen phase, then occurs in which the hair shaft is shed and then a new anagen phase is initiated. The length of each stage can vary based upon each hair and a pattern has been seen in aging in which the Anagen phase decreases and the Exogen phase increases with age [37].

While pigmentation differences are the most noticeable hair-feature differences between individuals, the morphology of the hair is also a very distinguishing feature. In the determination of whether an individual has straight or curly hair, the shape of the hair shaft seems to be intrinsically programmed in the hair follicle. Curly hairs seems to have a club like shape at their base and this asymmetrical shape tends to extend upward along the inner root sheath creating the curly shape of the hair [38]. The shape of the follicle tends to dictate the morphology of the hair; therefore, a straight follicle means a straight hair and a clubbed follicle indicates a curly hair. In addition to this, a differentiation in thickness on one side versus the other of the outer root shaft also seems indicative of curly hair [37].

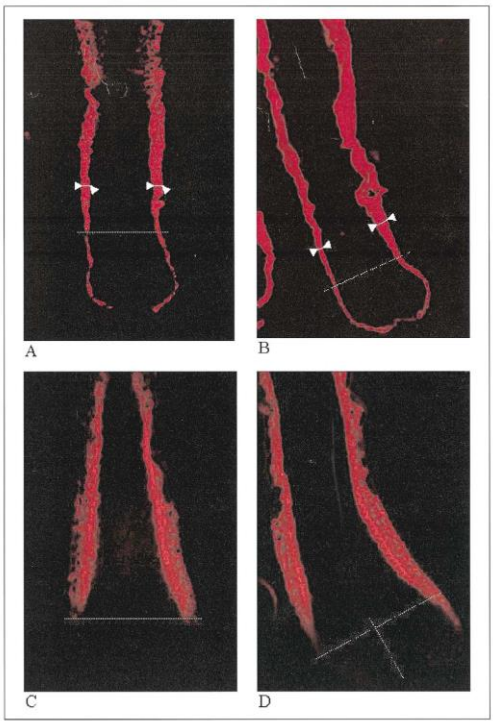


Figure 3 Photos A and C display uniform thickness of both sides of the outer root shaft of a straight hair versus B and D in a curly hair.

1.6 Markers Associated with Categorical Hair Morphology

With regards hair structure, a large study collaboration to identify markers associated with this trait was undertaken. A list of 91 potential markers was provided by collaborators, some previously published, others simply proposed. To generate accurate genotypes to assess these markers, it was necessary to design an MPS hair morphology assay on the Miseq sequencer. Some background on this list of markers is detailed below.

Marcinska *et al.* [38] created a simple 5-SNP assay along with a 20 SNP prediction model from 50 SNPs for Androgenetic Alopecia. These models were based on SNPS from 305 males. These males were European and either over 50 years old and had not started balding or under 50 with significant baldness. The loss of hair makes some of these SNPs interesting as possible indicators of hair structure. While identifying these genes, the 5 SNP model was created using rs5919324 (upstream of *AR*), rs1998076, rs929626 (*EBF1*), rs12565727 (*TARDBP*), and rs756853 (*HDAC9*) [39]. Rs5919324, for example, a ‘T’ allele makes a male 2.1 times more likely to have male pattern baldness as opposed to the ‘C’ allele. Their extended model included SNPs rs1041668, rs5919324

(upstream of *AR*), rs6625163, rs6625150, rs962458, rs12007229, rs1998076, rs2180439, rs913063, rs1160312, rs6113491, rs756853 (*HDAC9*), rs6461387, rs6945541, rs12565727 (*TARDBP*), rs7349332 (*WNT10A*), rs4679955 (*SUCNR1/MBNL1*), rs929626 (*EBF1*), rs9668810, and rs10502861 [39].

Li *et al*. [39], in 2012, performed a genome wide association study on 12,806 individuals who were of European ancestry who suffer from early onset Androgenetic Alopecia as well allowing the indication of several additional SNPs. This GWAS allowed the identification of the SNPs rs12565727 (*TARDBP*), rs9287638 (*HDAC4*), rs2073963 (*HDAC9*), rs6945541 (*AUTS2*), rs12373124, rs10502861 (*SETBP1*), rs6047844 (*PAX1*, *FOXA2*), and rs2497938 (*AR*) [40].

Buket *et al.* [40] just recently released a study in December of 2016 looking to identify the causal variants behind "un-combable hair syndrome". This disorder makes the individual's blond, dry, wiry hair truly impossible to comb. This study revealed a triangular shaped cross-section of the hair as opposed to the typical round shape observed. This study could identify that the genes *PADI3*, *TGM3,* and *TCHH*. Upon knocking out *PADI3* in mice, this group confirmed that this enzyme is essential to prevent this disorder. In mice, its knockout causes alterations in whiskers and hair coat of the animal. Their coat became wavy and whiskers were irregular and twisted, solidifying the functional genetics behind these variants [41].

The *VDR* gene has been proven to be essential in hair follicle integrity, not necessarily hair morphogenesis as previously hypothesized. It has been proven that in mice with this gene knocked out develop alopecia. This gene's interaction with DNA is essential for the hair cycle to continue after embryonic development and therefore makes the SNPs vital to assess [42].

LGR4 in knockout mice was shown to have reduced amounts of *EDAR*, *LEF1*, and *SHH* which is essential for mouse hair follicle development. This suggests that if there is a deficiency in *LGR4* there is a reduction of signal to keratinocytes for hair morphogenesis and therefore may play a role in hair morphology in humans [43].

1649 individuals from 837 families were used by Medland *et al.* [43] to identify SNPs that indicated that the *TCHH* plays a significant role in determining hair structure. Four key SNPs were identified in this investigation rs17646946, rs11803731, rs4845418,

and rs12130862. Rs11803731 was determined to code for methionine at position 790 in *TCHH* protein. While the function of this is not known for sure, it is hypothesized that its function could possibly react with reactive oxygen species and result in a change in the structure of the protein therefore resulting in a modification to the hair follicle [44]

Pośpiech *et al*. [44] evaluated six SNPs reported in the *TCHH*, *WNT10A*, and *SLC45A2* genes in 528 individuals from Poland. During this study, they were looking at identifying the predictive capacity of these genes, especially relating to straight hair in European individuals [45]. The study this group performed confirmed the idea that rs11803731 (*TCHH*), rs73449332 (*WNT10A*), and rs1268789 (*FRAS1*) do, in fact, play a vital role in the morphology of hair. These three SNPs together were found to give greater than 80% predictability to straight hair [45].

Additionally, through use of genetic data obtained from the web based company 23andMe, Inc. (a direct to consumer genetic information company), Ericksson *et al.* [37] was able to identify several novel genetic associations for physical appearance traits such as hair morphology, freckling, photic sneeze reflex and many others [38]. Some of the most relevant findings to this research focused upon the SNPs rs17646946 (near *TCHH*), rs7349332 (near *WNT10A*), and rs1556547 (near *OFCC1*). Rs17646946 (near *TCHH*) minor allele was found to reduce hair curl (therefore associated with straighter hair), this effect was also seen with rs1556547 (*OFCC1*). Rs499697 (*LCE3E*) whose minor allele was found to have the opposite effect in that it was associated with curlier hair, which was also similarly found with rs7349332 (*WNT10A*) [38].

The rest of the SNPs for the hair morphology assay were based upon the function of the genes in which they were in or near and therefore it is necessary to access the function of these genes. *PADI3* enzyme converts *L-arginine* with a positive charge into a neutral Citrulline residue only in the presence of Calcium ions. This is true in the hair shaft proteins. *TCHH*, the most commonly associated gene with hair structure, codes for a structural protein localized in the inner root sheath of the hair follicle as well as the medulla. This gene helps to give the hair its strength by crosslinking with itself and other major proteins on the outer layer of the hair [32]. When this is deaminated by *PADI3* it reduces the charge of the *TCHH* and allows it to be attached to the keratin filaments. *TGM3* then crosslinks these filaments with the *TCHH*. These filaments are stabilized and

further cross-linked. A mutation at any point in this process may result in a compromise in the structural integrity of the hair [41].

Huber et al. [14] performed mouse studies to determine the true function of the repetin gene (*RPTN*), also believed to have a role in hair structure. This gene has been identified as being fused with other cornified cellular proteins such as profilaggrin, trichohyalin, and hornerin, occurring late in differentiation. Through this study, they could hypothesize that *RPTN* also functions with a calcium binding protein. They also identified the expression of this gene in the inner root sheath of hair follicles alongside *TCHH*. It has been shown that *TCHH* is cross-bound to proteins in the hair shaft, including *RPTN*. These support the idea that this gene plays a role in mechanical strength of the hair follicle [14].

Schlake *et al*. [45] determined that Insulin Growth Factor Binding Protein 5 (*IGFBP5*) in mice is a distinguishing factor in hair types. The expression of this protein in the medulla of the hair results in the bending zones of the hair allowing that this gene is a major contributor to mouse hair curving and thinning. Therefore identifying and studying SNPs located in this gene may also indicate if this gene plays a role in determining the type of hair an individual has [46].

LIPH is a vital large triglyceride lipase, which initiates regulating and maintaining the hair growth cycle. Based upon the sequence variations of these genes, they can result in partial or complete hair loss as well as development of wooly hair, which indicates that this gene is a huge component of hair development. This condition of wooly hair, hypertrichosis, is a result of a frame shift mutation or a deletion in this gene [15]. For this disorder, it is also being identified in the autosomal dominant hypertrichosis, the gene keratin, *KRT71*. *KRT71* has been previously identified to cause wavy/curly hair in mice, cats, rats, and dogs [16]. *KRT74* has been identified in causing the wooly hair syndrome in humans. This suggests that regulation of *KRT71* is performed through the activation of a growth factor receptor [17]. *P2RY5* has also been implicated in Hypertrichosis by Pasternack *et al.* [46], from their study of a Saudi Arabian family they were able to determine that *LPA,* a bioactive lipid is a ligand for *P2RY5* and therefore is the link between the cellular membrane and the hair follicle [47].

WNT10A is a member of the *WNT* family, which is a pancreas signaling molecule; these regulate cell organization during embryo development. *WNT10A* has been found to play a role in the formation of odontoblasts, which aid in the process of forming dentin in teeth. A dysfunction in this gene not only results in disorders concerning the mouth and teeth, but also the hair and nails. *EDAR* variants, especially rs3827760, have been isolated to result in tooth and hair health [18, 19]. Based upon their roles in overlapping processes looking at these results could provide insight in to hair structure.

1.7 Current SNP Genotyping Methods

Currently, there are two major methods for the genotyping of SNPs in the forensic field. The current widespread method is completed using capillary electrophoresis on a genetic analyzer. The current methodologies for HIrisplex and HIrisPlex-S revolve around this technology. All the primers used for the HIrisPlex and HIrisPlex-S reactions are part of multiplex reactions using SNaPshot chemistry [48] (Applied Biosystems). These reactions begin by a flanking PCR step where an 80-200bp region around the SNP of interest is isolated. These regions are isolated and copied thousands of times, then cleaned up by Illustra ExoProStar to remove excess dNTPS and primers left from the reaction. Following this step, the single base extension primer hybridizes the area between the flanking primers and lines up to the SNP of interest. This is then extended by a single base, which is labeled by a florescent tag on the dideoxynucleotide triphosphate (ddNTPs) that can be recognized by Capillary Electrophoresis on the genetic analyzer and results in the termination of the PCR after a single base extension. Once this PCR is complete, Shrimp Alkaline Phosphatase (SAP) is used to clean up the excess ddNTPs present in the reaction. In the final step, the sample is placed on the analyzer with formamide for suspension and capillary travel along with LIZ an internal size standard. This will then produce an electropherogram that can be easily binned and genotyped.

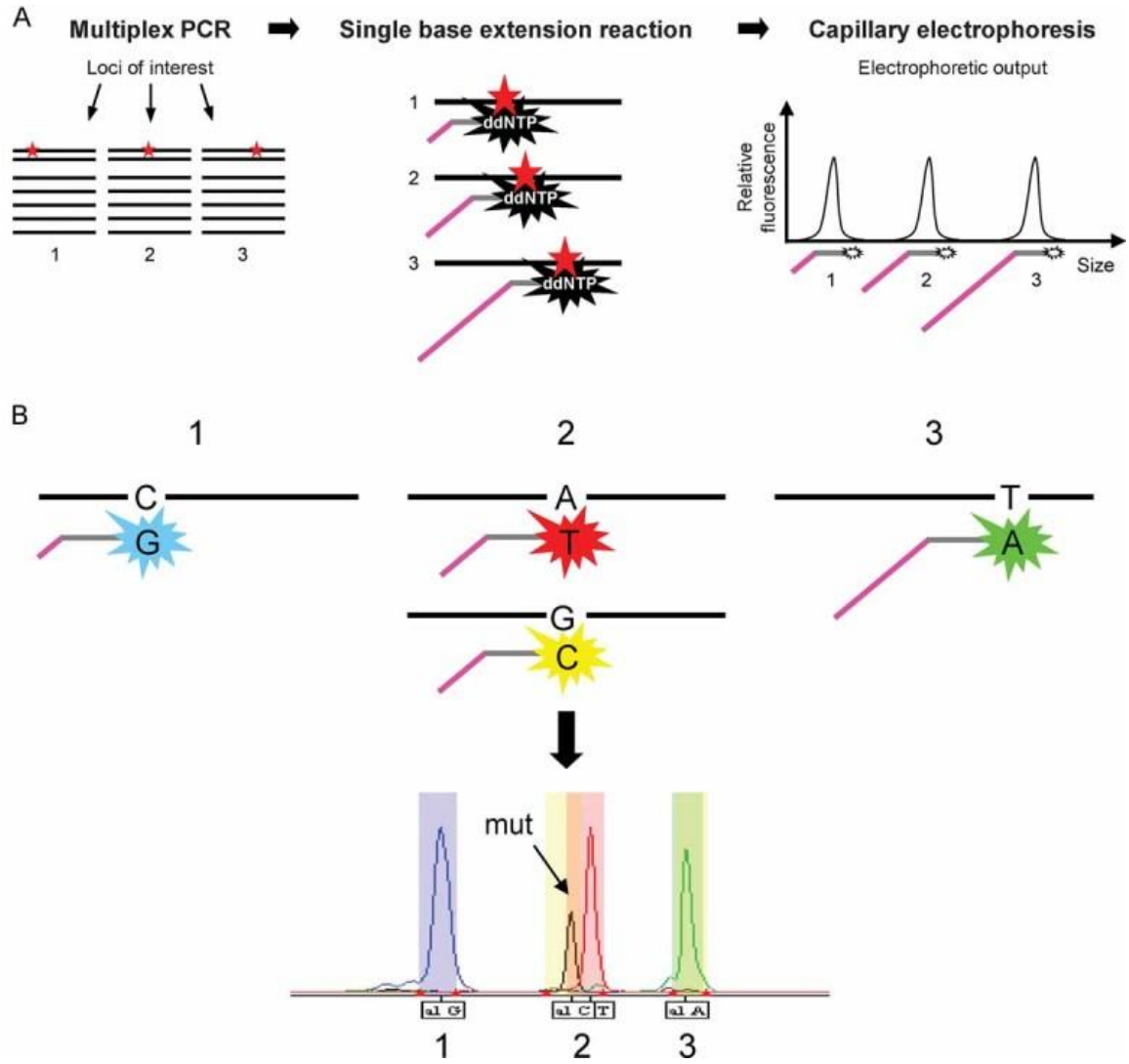


Figure 4 Santagata *et al* depiction of Snapshot Multiplex Reaction

The second, and newest method is Massive Parallel Sequencing (MPS), which is a four-step process that results in the sequencing of millions of fragments of DNA. The two main types of MPS systems are the Ion Torrent, a pH based system, and the Miseq a sequencing by synthesis system. For the intent of this thesis, we will focus all our Massive Parallel Sequencing on the Miseq.

The process for the Miseq begins with library preparation in which DNA is fragmented based upon primers designed to specific regions that contain adapter sequences in addition to the primer sequence. PhiX, which is spiked in during this procedure, operates as a control library for the cluster generation, sequencing, and

alignment, therefore allowing us to determine an accurate error rate within our reaction. These adapters are specific to the technology being used and allow for the fragments to adhere to the flow cell. Once these are cleaned up, the next step is cluster generation. This process is completed by the loading of the library onto the polyacrylamide flow cell that contains a lawn with complementary oligos to the adapters on the library that was assembled. Bridge amplification occurs to form clusters of amplicons. These clusters are then read using a sequence by synthesis method. This method uses four fluorescently labeled nucleotides. A single dNTP is added to the sequencing chain of nucleic acids. This stops polymerization and the dye is then imaged and then an enzymatic reaction is used to cleave this dNTP so that the next nucleotide can be incorporated. This step results in sequencing of these fragments like CE sequencing [27]. This is then imaged on the flow cell and based upon the wavelength and intensity, a base is recorded. The fragments are then de-multiplexed in order for them to be separated into each individual indices that correspond with the individual samples. The final step is to then align to a custom genome and analyze the fragments.

CHAPTER 2. METHODS AND MATERIALS: GLOBAL SKIN COLOR PREDICTION MODEL INCLUDING HIRISPLEX-S DESIGN AND VALIDATION ON CAPILLARY ELECTROPHORESIS

2.1 Sample Collection and Skin Color Phenotyping

Samples that were contributed towards this study were collected in compliance with IRB#1409306349, 329 US samples were used along with 1625 samples previously genotyped from Poland, Ireland, and Greece for the model validation. For the US samples, 5ml of saliva was collected from each individual, after the individual had consented to the study. Each participant filled out a questionnaire and had three hairs removed from the back of their head. Photographs were then taken of each participant's externally visible characteristics with a Nikon D5300 camera with AF-S Micro Nikon 60mm lens and two Nikon Wireless Speedlight SU-800- Remote SB-R200 flashes. Once the photos were complete, pigmentation measurements were taken with a Konica Minolta 700d/600d Spectrophotometer which measures wavelengths from 400-700nm, Melanin Index, and L**ab* values. The spectrometer measurements for L**ab* were as follows: 74.14-60.36 for White, comprising 132 samples; 59.32-40.04 for Intermediate, comprising 43 samples; 39.75-29.99 for Black, comprising 20 samples. 4ml of saliva was then added to 1ml of DNA lysis buffer and stored at -20ºC until DNA extraction took place. 1ml of saliva was also added to 1ml of RNAlater (Fisher Scientific International Inc., Hampton, NH) and stored at -20ºC until RNA extraction took place. The hair samples were stored at -20ºC in RNAlater as well for future use. The photo of the SpyderCHECKR™ calibration card was taken at each collection to help standardize the lighting over multiple collections.

2.2 DNA Extraction, Quantification & SNP Genotyping

To perform the DNA extraction, 1ml of the saliva sample/lysis buffer mixture, 15µl of Proteinase K (New England Biolabs, Ipswich, MA), and 130µl of 10% SDS (Dot Scientific Inc., Burton, MI) were added to a 2ml tube and incubated on an Eppendorf Thermomixer set to 300 rpm at 37°C overnight. Once this incubation was complete,

200µl of 0.5M NaCl (Dot Scientific Inc.) was added, inverted, and then incubated at room temperature for 10 minutes. The samples were then centrifuged for 12 minutes at 10ºC at maximum speed (16200cfg). The supernatant was transferred to a new tube and an equal volume of 100% isopropanol (Fisher Scientific International Inc., Hampton, NH) was added. It was then incubated for 10 minutes at room temperature, and then centrifuged for 15 minutes at 10ºC at maximum speed (16200cfg). The supernatant was discarded and the pellets were washed. The supernatant was discarded and the wash repeated with 500µl 70% ethanol (Decon Laboratories Inc., King of Prussia, PA) after inversion. They were then centrifuged again for 5 minutes. After the second wash, the supernatant was removed and the pellets were dried for 10 minutes. The samples were then re-suspended in 30µl of deionized H_2O and stored at -20°C or -80°C until further use.

Once the extraction was complete, the samples were quantified using the Qubit Fluorometer following the standard protocol for the DNA High Sensitivity Assay Kit (Fisher Scientific International Inc., Hampton, NH). These samples were genotyped using the HIrisPlex, HIrisPlex-S, and Snipper systems [2, 3].

The assays were set up based upon the standard steps for SNaPshot chemistry. For the HIrisPlex-S reaction the initial flanking primers are used to amplify an area surround the SNP of interest by anywhere from 80-200bp. The first 3 MC1R primers were added at a concentration of 0.5µM for both the forward and reverse flanking primers. The remaining 14 primers were added at a concentration of 0.4µM of both the forward and reverse flanking primers. 1µl of 10X PCR gold buffer (Fisher Scientific International Inc., Hampton, NH), 1µL of 25µM $MgCl_2$ (Fisher Scientific International Inc., Hampton, NH), 0.22µL of 10mM dNTPs (Fisher Scientific International Inc., Hampton, NH), 3.64µL of deionized H_20 and 0.3µL of 5U/µL AmpliTaq Gold (Fisher Scientific International Inc., Hampton, NH) were added per sample. 1µL of DNA was then added per reaction as well. PCR amplification was performed to amplify the regions of interest per the following standards at 95°C for 10 minutes, then 33 cycles of 95°C for 30 seconds then 62°C for 30 seconds, and finally 61°C for 5 minutes. The sample was then held at 15°C until the next step was performed. All of this was performed on the Eppendorf Mastercycler Nexus. For HIrisPlex-S, the appropriate amount of primers, 1µl of 10X PCR gold buffer (Fisher

Scientific International Inc., Hampton, NH), 1µL of 25µM $MgCl_2$ (Fisher Scientific International Inc., Hampton, NH), 0.22µL of 10mM dNTPs (Fisher Scientific International Inc., Hampton, NH), 2.38µL of deionized H_20 and 0.3µL of 5U/µL AmpliTaq Gold (Fisher Scientific International Inc., Hampton, NH) were added per sample. The protocol for HIrisPlex and HIrisPlex-S only differ at the annealing/extension step in the PCR cycle, in that instead of 62ºC, HIrisPlex requires those two steps to be at 61ºC to be in the optimal range for annealing of the primers.

From this point Illustra ExoProStar (GE Healthcare, Waukesha, WI) was used to clean up the samples before the single base extension reaction was performed. 2µL of the ExoProStar was added to 5µL of sample and then placed on the thermocycler at 37°C for 45 minutes and then 80°C for 15 minutes. The sample was then held at 15°C until the next step was performed.

The third step in the process is the actual single base extension reaction. Each SNP's single base extension primer was added per the concentrations on the protocol in Tables 2 and 3 in *Appendices B* and *C*. For HIrisPlex, the primer mixture was added at 1.74µL, 0.26µL H_2O, and 1µL of SNaPshot (Fisher Scientific International Inc., Hampton, NH) are added along with 2µL of sample. For HIrisPlex-S, the primer mixture was added at 1.77µL, 0.225µL H_2O, and 1µL of SNaPshot (Fisher Scientific International Inc.) are added along with 2µL of sample. The sample was then placed on the thermocycler at 96°C for 2 minutes, then 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 30 seconds. The sample was then held at 15°C until the next step performed.

The final step in the amplification process was the final clean up step using Shrimp Alkaline Phosphatase (SAP) (Fisher Scientific International Inc.,). 1µl of SAP is added to each sample and then placed back on the thermocycler at 37°C for 45 minutes followed by 75°C for 15 minutes. The sample was then held at 15°C until it was placed on the analyzer.

The purified samples are then processed and visualized by the Applied Biosystems 3500 Genetic Analyzer using POP7 (Fisher Scientific International Inc.,) with a 50cm array. 1µL of sample was added to 8.9µL of Formamide (Fisher Scientific International Inc.,) and 0.1µL of GeneScan Size Standard LIZ (Fisher Scientific

International Inc.,) per plate well. These samples were visualized using Applied Biosystems GeneMapper software.

2.3 Skin Color Phenotyping and Prediction Modeling

The Fitzpatrick scale was applied by a dermatologist to accurately phenotype the skin pigmentation of the individuals used for model validation [49]. The typical Fitzpatrick scale breaks down skin types into six groups. The groups are based upon the ability of the individual to tan or burn in the sun in addition to pigmentation [50]. Narrowing these categories down to both 3 and 5 categories, a statistical model was built that could then be compared to previous skin color prediction models such as Maronas *et al.* [3]. In this paper [49], we genotyped and assessed 1159 individuals from South Poland, 347 from Ireland, 119 from Greece, and 329 from the United States at 77 markers. Along with these groups we also included 71 individuals from the HGDP-CEPH reference set. Due to the requirement for a complete set of genotypes the number of samples was then reduced to 1423. These individuals were phenotyped using the Fitzpatrick scale [49].The skin types in the Fitzpatrick scale range from skin type 1(pale-white skin, no ability to tan), type 2 (white skin, minimal ability to tan, type 3 (light brown skin, tanning ability), type 4 (moderate skin, tanning ability), type 5 skin (dark brown skin, tanning ability, and type 6 (deep pigmentation of the skin, dark brown to black).

329 samples from an in-house FDP database were used to represent samples from the United States (while their parents were from outside the United States). These included individuals who were from Nigeria, Mexico, Argentina, Columbia, India, Bangladesh, Cuba, Palestine, Canada, China, Honduras, Germany, Philippines, Russia, Sudan, Japan, Saudi Arabia, Pakistan, El Salvador, Spain, Haiti, South Korea, and Vietnam. This information was self-reported on their initial questionnaire. In terms of phenotyping their skin pigmentation, everyone's 2D photograph (arm picture) along with tanning information that was recorded on their questionnaire was graded by a dermatologist to determine the appropriate Fitzpatrick rating that their skin should receive.

77 skin pigmentation SNPs were tested in their predictability for skin pigmentation. A partial correlation (correcting for sex and ancestry) resulted in 53 statistically significant SNPS. From this point AUC values were used to determine the best combination of SNPs to gain the most substantial amount of pigmentation information. From this 36 SNPs from 16 genes were selected; rs1426654 (*SLC24A5*), rs12203592 (*IRF4*), rs1805007 (*MC1R*), rs1805008 (*MC1R*), rs11547464 (*MC1R*), rs885479 (*MC1R*), rs228479 (*MC1R*), rs1805006 (*MC1R*), rs1110400 (*MC1R*), rs1126809 (*MC1R*), rs3212355 (*MC1R*), rs1800414 (*OCA2*), rs1800407 (*OCA2*), rs12441727 (*OCA2*), rs1470608 (*OCA2*) , rs1545397 (*OCA2*), rs16891982 (*SLC45A2*), rs28777 (*SLC45A2*), rs1667394 (*HERC2*), rs2238289 (*HERC2*), rs1129038 (*HERC2*), rs12913832 (*HERC2*), rs6497292 (*HERC2*), rs1042602 (*TYR*), rs1393350 (*TYR*), rs6059655 (*RALY*), rs8051733 (*DEF8*), rs2378249 (*PIGU*), rs6119471 (*ASIP*), rs2402130 (*SLC24A4*), rs17128291(*SLC24A4*), rs12896399 (*SLC24A4*), rs683(*TYRP1*), rs12821256 (*KITLG*), rs3114908 (*ANKRD11*), and rs10756819 (*BNC2*). 194 individuals that had not been previously used to generate the model were then used to compare the skin color prediction performance of both the HIrisPlex (36 of the above variants were genotyped) and Snipper models [2,3] (where 10 variants were typed). The Snipper model includes the following 10 SNPs rs10777129 (*KITLG*), rs13289 (*SLC45A2*) and rs16891982 (*SLC45A2*), rs1408799 (*TYRP1*), rs1426654 (*SLC24A5*), rs1448484 (*OCA2*), rs2402130 (*SLC24A4*), rs3829241 *(TPCN2*), rs6058017 (*ASIP*) and rs6119471 (*ASIP*).

2.4 Optimization and Developmental Validation of the HPS Assay

The HIrisPlex-S assay requires two individual SNaPshot reactions to achieve a complete profile for the 36 required skin color associated SNPs: rs1426654 (*SLC24A5*), rs12203592 (*IRF4*), rs1805007 (*MC1R*), rs1805008 (*MC1R*), rs11547464 (*MC1*R), rs885479 *(MC1R*), rs228479 (*MC1R*), rs1805006 (*MC1R*), rs1110400 (*MC1R*), rs1126809 (*MC1R*), rs3212355 (*MC1R*), rs1800414 (*OCA2*), rs1800407 (*OCA*2), rs12441727 (*OCA2*), rs1470608 (*OCA2*) , rs1545397 (*OCA2*), rs16891982 (*SLC45A2*), rs28777 (*SLC45A2*), rs1667394 (*HERC2*), rs2238289 (*HERC2*), rs1129038 (*HERC2*), rs12913832 (*HERC2*), rs6497292 (*HERC2*), rs1042602 (*TYR*), rs1393350 (*TYR*), rs6059655 (*RALY*), rs8051733 (*DEF8*), rs2378249 (*PIGU*), rs6119471 (*ASIP*), rs2402130

(*SLC24A4*), rs17128291(*SLC24A4*), rs12896399 (*SLC24A4*), rs683(*TYRP1*), rs12821256 (*KITLG*), rs3114908 (*ANKRD11*), and rs10756819 (*BNC2*). The first assay is the previously published HIrisPlex assay designed by Walsh et al. generating the 24 DNA variants using 17 primer sets [6]. The second 17-plex consists of the remaining variants. The flanking primers of SNPs rs10756819, rs1126809, rs6497292, rs1545397, rs6059655, and rs12441727 did not undergo successful amplification during the initial design by collaborators and were therefore redesigned at the Walsh laboratory. The single base extension (SBE) primers for rs3114908, rs10756819, rs1126809, and rs6059655 we also redesigned due to unsuccessful amplification. Details of these new designs can be found in Table 2 in *Appendix A*.

2.5 Concordance of SNaPshot Assay Using Casework Samples

A concordance casework set (30 samples) was received from a collaborating site to ensure the assay worked accurately despite the location, individual running the assay, or machine the assay was performed on. The assays were performed on an Applied Biosystems 3130 Genetic Analyzer at the collaborating site, the Netherlands. Therefore, it was essential to run the same samples on a newer platform, Applied Biosystems 3500 Genetic Analyzer to ensure that the assay's efficiency is the same regardless of the change in parameters. These 30 samples were run following the previously described HIrisPlex-S protocol. The profiles for these samples were obtained using Applied Biosystems GeneMapper software. An interpretative threshold of 50 RFU was used to ensure accuracy when identifying the profile

2.5.1 Sensitivity Testing for Biological Assay Concordance

The sensitivity of this assay was assessed on both the 3130 and 3500 Genetic Analyzers. The concentrations in which the samples were evaluated at were 1ng/µL, 500 pg/µL, 250 pg/µL, 125 pg/µL, 60 pg/µL and 30 pg/µL. The control samples created from DNA extracted from collection participants labeled K1 and C1 were diluted to appropriate concentrations by taking triplicate measurements using the Qubit fluorimeter.

2.5.2 Species Specificity Testing

Species specificity testing was performed to ensure that the primers designed were indeed human specific. This system was tested against cat, chimp, dog, mouse, and pig DNA quantified to 1ng/µl. All samples were obtained externally from pet samples and extracted in house apart from the chimp sample which was obtained from Dr. Brenda Bradley of George Washington University and the mouse sample was obtained from Dr. Benjamin Perrin of Indiana University Purdue University Indianapolis.

CHAPTER 3. METHODS AND MATERIALS: HAIR MORPHOLOGY SNP AND GENE ASSOCIATION THROUGH MASSIVE PARALLEL NEXT GENERATION SEQUENCING

3.1 Sample Collection and phenotyping, DNA Extraction, and Quantification

The samples for this study were collected as previously mentioned in section 2.1 as they are from an in-house FDP database. 1026 samples were used for this study. The extraction and quantification of these samples were completed as described in section 2.2. These samples were phenotyped using a self-assessment of their hair structure as to whether they categorize it as wavy, straight, or curly. 387 individuals identified as having Very Straight Hair, 144 individuals identified as having Curly Hair/ Rings of Curls, and 495 individuals identified as having Light Waves.

3.2 Hair Morphology SNP Genotyping

A custom Miseq Assay consisting of 91 primers that can be found in Table 5 in *Appendix D* for the initial PCR was ran for all samples mentioned in section 3.1. Primers were split into 3 groups of around 30 primers, PCR performed, and then combined post cleanup. The thermocycler settings for this reaction followed a touchdown PCR protocol. Protocols can be found in Tables 5,6, and 7 in *Appendices E, F, and G*.

3.3 Bead Clean Up Solution Preparation

A 5M NaCl solution was prepared along with a 40% PEG-8000 solution. 1ml of Sera-Mag Speed Beads were isolated via magnet from their suspension and then washed twice and resuspended in 1ml deionized water. The suspension solution was then prepared by adding 10 ml 5M NaCl, 9ml of nuclease free water and 1ml of the clean beads. 30ml of 40% PEG was then added to the bead mixture and vortexed, allowing for the beads to suspend. The mixture was then stored in a light protected tube in the 4° C refrigerator until use. This protocol developed by Rohland and Reich on high-throughput DNA sequencing in 2017 [51], was adapted and modified to fit our assay requirements.

3.4 Execution of an MPS Assay for Hair Morphology

384 samples were dispersed amongst four plates and then ran for each of the three primer groups to total 12 plates. All DNA had been previously diluted to quantify around 20ng/µL as reported on the Qubit. The primers were added in the concentrations specified on the three protocols attached in Tables 5, 6, and 7 in *Appendices E, F,* and *G*. 1µl of 10X PCR gold buffer, 1µL of 25uM MgCl2, 0.22µL of 10mM dNTPs, 3.64µL of deionized H_20 and 0.3µL of 5U/µL AmpliTaq Gold were added per sample. 1µL of DNA was then added per sample as well. This PCR amplification was used to amplify regions of interest according to the following standards at 95 °C for 10 minutes, then 28 cycles of 95°C for 30 seconds, a touch down starting at 64 °C for 40 seconds decreasing at 0.3 degrees a cycle and 60 °C for 5 minutes. This was then held at 15°C.

Once this reaction was successful in amplifying the regions of interest, 5µL of each product of the 3 primer group plates was added together in the same plate and along with 5µL of deionized water to dilute the samples.

To prepare the reaction for amplicon sequencing it was then necessary to clean up the plates using the magnetic bead suspension prepared previously. 3 times the amount of the sample is the amount of beads (typically 45µL) that were added to each sample and mixed thoroughly to ensure that the DNA properly adheres to the beads. An 8-minute incubation allowed for this binding of the beads to the DNA to take place. Once the mixture was briefly spun down, the samples were placed on a magnetic stand which forced the beads along with the DNA to the bottom and side along the magnet. The supernatant was discarded while on the stand and then washed with 90µL of 80% Ethanol. The ethanol was removed in the same fashion and then the wash repeated. The samples were air dried for 30 seconds and then resuspended in 25µL of purified water after a 5-7-minute incubation.

The second round of PCR was then performed to adhere the index sequences to each sample as a unique identifier in order to demultiplex each individual sample when pooled. For each well 5µL of KAPA master mix (KAPA Biosystems, Wilmington, MA), 1µl of each index (both forward and reverse to total 2 µL), 2 µL of H_20, and 1 µL of DNA were added to each well of the plate. The samples were placed on the Eppendorf

Mastercycler Nexus SX1 98 °C for 2 minutes, then 12 cycles of 98 °C for 30 seconds, 72°C for 30 seconds and 72 °C for 5 minutes and then held at 15ºC.

Following this step, the products were cleaned up adhering to the exact same procedure as mentioned above, but all reagents were cut in half due to the decreased volume of sample. To verify that the products of the PCR reactions were still present after this clean up, the products were ran on the Bioanalyzer to ensure that there were indeed products between 100bp and 300bp.

To successfully sequence these samples, the products were pooled, diluted and quantified to finalize library preparation. 5µL of each sample was pooled and then quantified using the Qubit fluorometer. The table found in Table 9 in *Appendix H* [52]was used then to calculate an accurate dilution to a 2nM concentration. Denaturing the library then required 5µL of 0.2N NaOH to 5µL of the 2nM library. After a brief spin down, this was incubated for 5 minutes at room temperature. It was essential after the denaturation to then dilute the sample again down to 10pM with 990µL of Hybridization Buffer (Illumina, San Diego, CA) provided with the Illumina Nextera XT Version 2 Kit (Illumina, San Diego, CA). This denatured library was then diluted again to 8nM using 480µL of library and 120µL of Hybridization Buffer with a pulse vortexing.

Now that the library was properly diluted, the PhiX control (Illumina, San Diego, CA) provided by Illumina was added to standardize the reaction. The 4nM PhiX library 5µL was added to 5µL of 0.2N NaOH. The control was then vortexed, spun down, and incubated for 5 minutes. The control was then diluted again using 10µL of PhiX Library and 990µL of Hybridization Buffer to a final concentration of 20pM PhiX library. The final dilution to 12.5 pM was then performed using 375µL of the previously diluted PhiX library and 225µL of Hybridization Buffer. The last step before adding the samples to the Miseq cartridge is to add a 15% PhiX spike in to the custom library prepared via 90µL PhiX and 510µL of the library.

Once the library was prepared and then the cartridge is thawed, 600ul of the prepared library was loaded into the ‘Load Samples’ reservoir on the cartridge. The cartridge was then ran on ‘Research Use Only’ Mode through the Nextera XT sequencing.

3.5 File Conversion, Alignment and Genotype Extraction

To evaluate the data that was synthesized as a result of this Miseq run, a custom genome had to be assembled first to align the samples with. All this assembly and conversion was performed on an iMac in a UNIX environment. This custom genome allows for the isolation and alignment of the DNA to areas of interest as opposed to the whole genome, making evaluation of data a much simpler task. For the type of data that was produced referred to as amplicons, aligning these to the entire genome would be error prone and memory intensive. Instead, by assembling a genome appropriate for this assay we can ensure that the amplicons align exactly where intended.

The first step in this conversion process is the conversion of the reference genome file to the ". fai" format or fasta index format allowing the file to be searchable. This is done through a program called SAM tools [53]. A fasta file is then created in order to generate custom sites of interest. After the file is in fasta format it was then necessary to use the program BWA (Burrows-Wheeler Aligner) to index the fasta file [54].The output file from the Miseq is in the form of a ".fastq" file. Two fastq files are produced by the Miseq for each sample and it contains either the forward or reverse sequencing read for that sample. From this file, conversion must first be completed to visualize the data. This then allows for the sequencing files created by the Miseq to be aligned via BWA to produce a sequence alignment map (.sam file) [54]. Once the ".sam" file is produced a program called Picard tools is needed to convert the ".sam" file to a binary format file (.bam file) [55]. Once converted, the bam file must be sorted. In addition to sorting the bam file, it is also necessary to generate a ".dict" file that contain contig name and sizes in order for the Genome Analysis ToolKit (GATK) to effectively recognize the custom genome [56].

Now that the file was converted and sorted as a bam file along with a searchable reference file, it was necessary to convert the Miseq generated files to the final file type of variant call files (.vcf file). To do so, it was necessary to install bcftools and vcflib. SAMtools, was used to convert the file to the ".vcf" format before exporting [53]. Now that the ".vcf" file had been created, it was then essential to install genomeanalysisTK.jar for the GATK program, this allows the compilation and visualization of the SNP genotypes at specific locations [56]. Based upon the visible results at this stage,

Integrative Genomics Viewer (IGV), a genome viewing software, was used to check and evaluate any SNPs in which the program was unable to successfully call a genotype for [57]. This allowed for the output of the genotypes for all 1130 individuals in the assay. All coding done for this portion of this thesis can be found in *Appendix I.*

CHAPTER 4. METHODS AND MATERIALS: MIGRATING HIRISPLEX- S TO THE MASSIVE PARALLEL NEXT GENERATION SEQUENCING PLATFORM

4.1 Sample Collection, Extraction, and Quantification

The samples for this study were collected as previously mentioned in section 2.1 as they are from the same database. The extraction and quantification of these samples were performed as described in section 2.2.

4.2 Development of a Custom Developmental Validation Set

SWGDAM guidelines required this set to address sensitivity, species specificity, mixture assessment, simulated casework, stability studies, population studies, and concordance. To see the distribution of these samples as well as their individual information refer to Table 10 in *Appendix J*. All sample concentrations were determined by qPCR via InnoQuant Human DNA Quantification and Degradation Assessment Kit [58].

4.2.1 Sensitivity

For sensitivity, three samples of control DNA, 9947A, 9948, and 007, were diluted to concentrations of 1pg, 5pg, 10pg, 25pg, 50pg, 100pg, 250pg, 500pg, and 1ng repeated in triplicate.

4.2.2 Species Specificity

This system was tested against cat, chimp, dog, mouse, and pig DNA at 1ng/µl. All samples were obtained externally from pet samples and extracted in house apart from the chimp sample, which was obtained from Dr. Brenda Bradley of George Washington University, and the mouse sample was obtained from Dr. Benjamin Perrin of Indiana University Purdue University Indianapolis.

4.2.3 Mixture Assessment

Mixture assessment was tested at length in ratios of 1:1, 1:2, 1:5, and two sets of 1:10. Two three person mixtures were also included in 1:1:2 and 2:1:2 ratios. The individuals used in these mixtures represent a variety of population set such as European, African, Chinese, Mexican, and Iranian. They also represent a variety of hair pigmentations ranging from blond to dark brown and eye pigmentations ranging from blue to brown. In addition to the physical traits of hair and eye pigmentation, their skin pigmentations also varied from type 1 to type 5 on the HIrisPlex-S skin pigmentation scale.

4.2.4 Simulated Casework

For the simulated casework, samples were manufactured to mimic dried and degraded blood, wet blood, DNA with heme contamination, dried and degraded saliva, wet saliva, touch DNA, hair, vaginal swab, and vaginal swab mixture with semen.

4.2.5 Stability Samples

To address the ability of the assay to perform on degraded samples two control samples were degraded using DNase at time points of 0 minutes, 5 minutes, 10 minutes, 20 minutes, and 40 minutes. The degradation of these samples was ensured through analysis of these samples using HIrisPlex on the genetic analyzer before use on the Miseq.

4.2.6 Population Studies

For the final aspect of assessing the assay's performance on various populations a variety of samples were included at a 500pg concentration from various populations such as European, Vietnamese, Mexican, African, Chinese, Panamanian, and Iranian. These individuals also represent the same widespread distribution of physical attributes as previously mentioned.

4.2.7 Concordance

The final assessment needed under SWGDAM guidelines is concordance. To address concordance for an assay such as this one, with MPS it is necessary to ensure that this assay performs well on both the Illumina Miseq and the Ion Torrent. To do this,

concordance plates were sent to three European collaborators with the respective instruments. A complete validation plate was sent along with two randomly selected concordance plates. These concordance plates contained 14 500pg samples, 7 degraded samples, 8 mock casework samples, 3 specifies specificity samples, 1 three-person mixture sample, 3 1:1 two-person mixture samples, 6 1:10 two-person mixture samples, 3 1:2 two-person mixture samples, and 6 1:5 two-person mixture samples.

4.3 HIrisPlex-S SNP Genotyping using MPS

A custom Miseq Assay consisting of 41 primers can be found in Table 12 located in *Appendix L* for the initial PCR which was ran for all samples mentioned in section 4.1. The thermocycler settings for this reaction followed a touchdown PCR protocol.

For the initial optimization run, 96 samples from the previously mentioned validation set were ran. 24 500pg samples were ran in duplicate along with 35 mixture samples ranging from ratios of 1:1, 1:2, 1:5, and 1:10. This run also included 2 three person mixtures and 11 mock casework samples. The distribution, population, eye color, hair color, and skin type for these individuals can all be found in Table 11 located in *Appendix K*. All DNA had been previously diluted to quantified through InnoQuant. The primers were added in the concentrations specified on the protocol attached in Table 13 in *Appendix M*. Included in the primer design were a forward adapter of TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG; and a reverse adapter of GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG. 1µl of 10X PCR gold buffer, 1µL of 25uM MgCl2, 0.22µL of 10mM dNTPs, 2.45µL of deionized H_20 and 0.4µL of 5U/µL AmpliTaq Gold were added per sample. 1µL of DNA was then added per sample as well. This PCR amplification was used to amplify regions of interest according to the following standards at 10 minutes at 95^{O}C, followed by 34 cycles of 30 seconds at 95^{O}C then touchdown at 62°C decreasing by 0.3 for 30 seconds and then 60^{O}C for 10s, finally 5 minutes at 60°C.

To prepare the reaction for amplicon sequencing it was then necessary to clean up the plates using the magnetic bead suspension prepared previously. 1.25 times the amount of the sample is the amount of beads that were added to each sample and mixed thoroughly to ensure that the DNA properly adheres to the beads. An 8-minute incubation

allowed for this binding of the beads to the DNA to take place. Once the mixture was briefly spun down, the samples were placed on a magnetic stand for 5 minutes, this forced the beads along with the DNA to the bottom and side along the magnet. The supernatant was discarded while on the stand and then washed with 200µL of 80% Ethanol. The ethanol was removed in the same fashion and then the wash was repeated after 30s. The samples were air dried for 2-5 minutes and then resuspended in 25µL of purified water after a 5-7-minute incubation.

The second round of PCR was then performed to adhere the index sequences to each sample as a unique identifier to demultiplex each individual sample when pooled. For each well 5µL of KAPA master mix (KAPA Biosystems, Wilmington, MA), 1µl of each index (both forward and reverse to total 2 µL), 2 µL of H_20 and 1 µL of DNA were added to each well of the plate. The samples were placed on the Eppendorf Mastercycler Nexus SX1 98 °C for 2 minutes, then 12 cycles of 98 °C for 30 seconds, 72°C for 30 seconds and 72 °C for 5 minutes and then held at 15ºC.

To successfully sequence these samples, the products were pooled, diluted and quantified to finalize library preparation. 5µL of each sample was pooled and then quantified using the Qubit fluorimeter. The table found in Table 9 located in *Appendix H* [52] was used then to calculate an accurate dilution to a 2nM concentration. Denaturing the library then required 5µL of 0.2N NaOH to 5µL of the 2nM library. After a brief spin down, this was incubated for 5 minutes at room temperature. It was essential after the denaturation to then dilute the sample again down to 10pM with 990µL of Hybridization Buffer provided with the Illumina Nextera XT Version 2 Kit. This denatured library was then diluted again to 8nM using 480µL of library and 120µL of Hybridization Buffer with a pulse vortexing.

Now that the library was properly diluted, the PhiX control (provided by Illumina was added to standardize the reaction. The 4nM PhiX library 5µL was added to 5µL of 0.2N NaOH. The control was then vortexed, spun down, and incubated for 5 minutes. The control was then diluted again using 10µL of PhiX Library and 990µL of Hybridization Buffer to a final concentration of 20pM PhiX library. The final dilution to 12.5pM was then performed using 375µL of the previously diluted PhiX library and 225µL of Hybridization Buffer. The last step before adding the samples to the Miseq

cartridge is to add a 20% PhiX spike in to the custom library prepared via 120µL PhiX and 480µL of the library.

Once the library was prepared and then the cartridge is thawed, 600ul of the prepared library was loaded into the Load Samples reservoir on the cartridge. The cartridge was then ran on ‘Research Use Only’ Mode through the Nextera XT sequencing.

CHAPTER 5. RESULTS AND DISCUSSION: GLOBAL SKIN COLOR PREDICTION MODEL INCLUDING HIRISPLEX-S DESIGN AND VALIDATION ON CAPILLARY ELECTROPHORESIS

5.1 Global Skin Color Prediction Model

77 SNPs were assessed for their value in predicting skin pigmentation in 2025 individuals. A partial correlation (correcting for sex and ancestry) resulted in 53 statistically significant SNPs. From this point AIC values were used to determine the best combination of SNPs to gain the most substantial amount of pigmentation information resulting in 36 SNPs from 16 genes being selected; rs1426654 (*SLC24A5*), rs12203592 (*IRF4*), rs1805007 (*MC1R*), rs1805008 (*MC1R*), rs11547464 (*MC1R*), rs885479 (*MC1R*), rs228479 (*MC1R*), rs1805006 (*MC1R*), rs1110400 (*MC1R*), rs1126809 (*MC1R*), rs3212355 (*MC1R*), rs1800414 (*OCA2*), rs1800407 (*OCA2*), rs12441727 (*OCA2*), rs1470608 (*OCA2*) , rs1545397 (*OCA2*), rs16891982 (*SLC45A2*), rs28777 (*SLC45A2*), rs1667394 (*HERC2*), rs2238289 (*HERC2*), rs1129038 (*HERC2*), rs12913832 (*HERC2*), rs6497292 (*HERC2*), rs1042602 (*TYR*), rs1393350 (*TYR*), rs6059655 (*RALY*), rs8051733 (*DEF8*), rs2378249 (*PIGU*), rs6119471 *(ASIP*), rs2402130 (*SLC24A4*), rs17128291*(SLC24A4*), rs12896399 (*SLC24A4*), rs683(*TYRP1*), rs12821256 (*KITLG*), rs3114908 (*ANKRD11*), and rs10756819 (*BNC2*) for model building. 329 individuals were genotyped in our laboratory at the 41 SNPs (for eye, hair, and skin prediction) to help produce a more global model due to their unique allele combination.

After the quality of the samples were evaluated a set of 1423 individuals used to produce the prediction model for the HIrisPlex-S system as a five category based prediction system [49]. The distribution of the samples was Very Pale 98 individuals, Pale 631 individuals, Intermediate 555 individuals, Dark 49 individuals, and Dark-Black 90 individuals. Upon analyzing the results for skin categories, it was most beneficial to combined categories 3 and 4 for our system, reducing it to a five-category system; Very Pale, Pale, Intermediate, Dark, and Dark-Black.

To represent the global population, splitting the light category is advantageous for a more accurate physical prediction rather than maintaining the same 3 category system.

When this is done the AUCs were as follows: 0.97 ±0.03for Dark- Black, 0.87 ±0.1 for Dark, 0.73 ±0.03 Intermediate, 0.72 ±0.03 Pale, 0.74 ±0.05 Very Pale. Based upon this data, we are able to conclude that our prediction model and assay will achieve DNA-based skin pigmentation prediction regardless of biographical ancestry.

5.1 Comparison of HIrisPlex-S to Maroñas 'Snipper' Skin Color Model

An individual set of 194 individuals were genotyped for the 36 SNPs for HIrisPlex-S as well as the 10 predictors SNPs for Maronas *et al.[3]*. On the dataset using the Maronas *et al*. model, AUC values using 194 individuals were generated. This resulted in the AUC values as follows: 0.75 for White, 0.55 for Intermediate, and 0.67 for Black. When comparing these AUCs to those obtained using our modeling system for the 3-category system used in the Maronas model, the AUC increased to 0.79 for White, 0.66 for Intermediate, 0.89 for Black. However, the highest AUC values were achieved when our 36 SNP model employing the Fitzpatrick scale was used with AUC values of 0.9 for Light, 0.91 for Dark, and 0.91 for Dark-Black. Therefore, indicating that our model outperformed and improved the skin color predictions given by the model provided by Maronas *et al*. [3].

5.2 HIrisPlex-S Design and Validation on Capillary Electrophoresis

As for the HIrisPlex-S design, the original design was obtained from previous work with a collaborating site in the Netherlands. While optimizing the initial set up using the SNapShot multiplex assay, it was found that the flanking primers of SNPs rs10756819, rs1126809, rs6497292, rs1545397, rs6059655, and rs12441727 did not undergo successful amplification under singleplex or multiplex conditions and therefore it was necessary to redesign or adjust the concentrations of these primers as required. Following this, the single base extension (SBE) primers for rs3114908, rs10756819, rs1126809, and rs6059655 we also redesigned due to unsuccessful amplification under singleplex and multiplex conditions as well.

Flanking primers for SNPs rs10756819, rs1126809, rs6497292, rs1545397, rs6059655, and rs12441727 were initially unsuccessful during amplification, therefore a

step-by-step process was used to create and test the new primers. In order to design these flanking primers the SNP of interest and the surrounding area were identified on the NCBI SNP database, dbSNP [59]. Once the area was identified, the surrounding SNPs and INDELS were marked and then the sequence was placed into Primer3Plus in order to isolate the specific flanking primers of optimal temperature and size [60]. The primers are then ran through the ePCR program Bisearch to ensure that the specificity of the primer in that it only binds at the specific spot in the genome that it is intending to isolate [61]. For the flanking primers, the entire set was ran together in the program Autodimer to ensure that the primers did not interfere with one another or form hairpins upon themselves [62].

Single base extension primers rs3114908 and rs6059655 were originally forward SBE primers. To allow these primers to work efficiently in the reaction these primers had to be switched to the reverse direction. Primer rs1126809 was flipped to the forward direction from its original reverse direction according to NCBI [59]. Primers rs12441727 and rs10756819 had their "T" tail length adjusted to ensure that the SNPs fell in bins on the genetic analyzer in which they did not compete with other SNPs. See Table 2 in Appendix *A* for table of SNP primers. The HIrisPlex and HIrisPlex-S protocols can be found in Tables 2 and 3 in *Appendices B and C.*

Once the primer designs were finalized, concentrations were adjusted for each primer. This allowed for all peaks in the reaction to be easily visible without preferential amplification occurring. While all PCR-1 primers were kept at 0.4µm, each SBE primer was adjusted individually through singleplex and multiplex designs to find the ideal concentration for a clean profile to be produced at 1ng/µL. All trials were completed with control samples quantified at 1ng/µL to control for sample variation. This was all necessary for model validation and to push forth to publishing a biological assay for widespread use.

To complete the developmental validation of this assay there are several tests that must be performed to state that the assay meets SWGDAM guidelines and is acceptable to be used in a forensic laboratory. These guidelines include Sensitivity Study, Species Specificity, Mixture Assessment, Concordance Testing, Simulated Casework, Stability

Studies, and Population Studies. Of these Sensitivity Testing, Species Specificity, and Concordance Testing were completed in our laboratory.

5.2.1 Sensitivity Testing

Using a sample of control DNA as described in section 2.6, sensitivity testing was performed to determine the sensitivity range for the HIrisPlex-S assay. Based upon the results of this testing, the HIrisPlex-S system produces a full profile down to 60pg/µl of DNA. After 60pg/µl, dropout of SNPs rs12441727, rs3212355, and rs8051733 occur. This is illustrated in Figure 5.

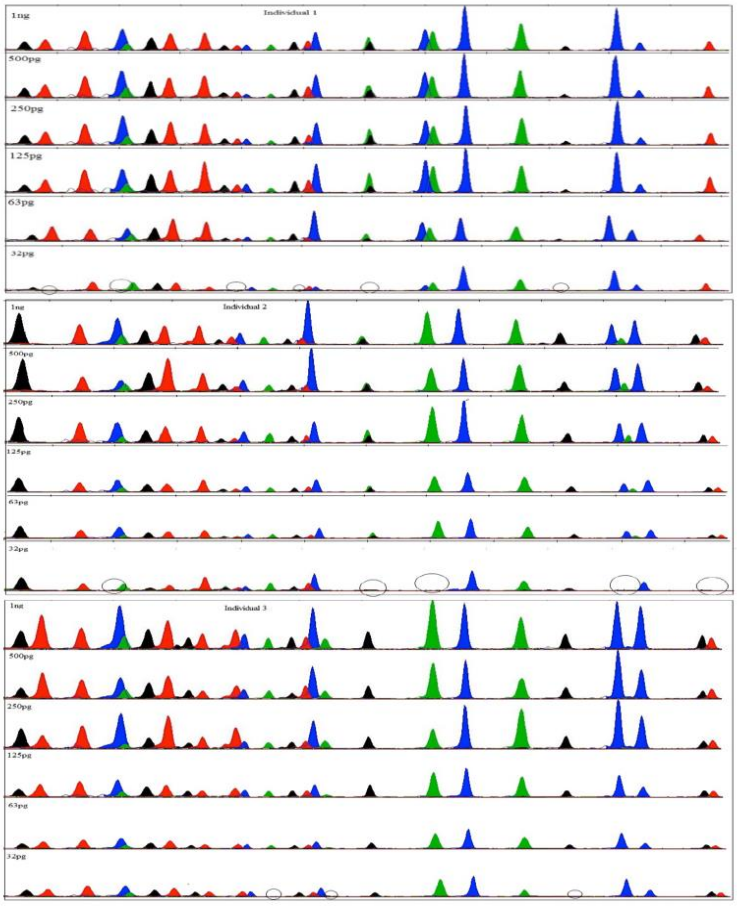


Figure 5 Electropherograms for Sensitivity for the HIrisPlex-S Capillary Electrophoresis Assay from 1ng/µl to 30pg/µl

5.2.2 Species Specificity

After completing the sensitivity testing, species specificity testing was completed to ensure that the assay only produces complete profiles for human samples. Of the non-human samples, there was no successful amplification in any of the electropherograms except for one peak in the chicken, one peak in the bovine, and several peaks in the primate. In the chimp, fourteen of the primers amplified and produced a partial profile. Due to the homology between the chimp and human genomes, this amplification was expected and acceptable.

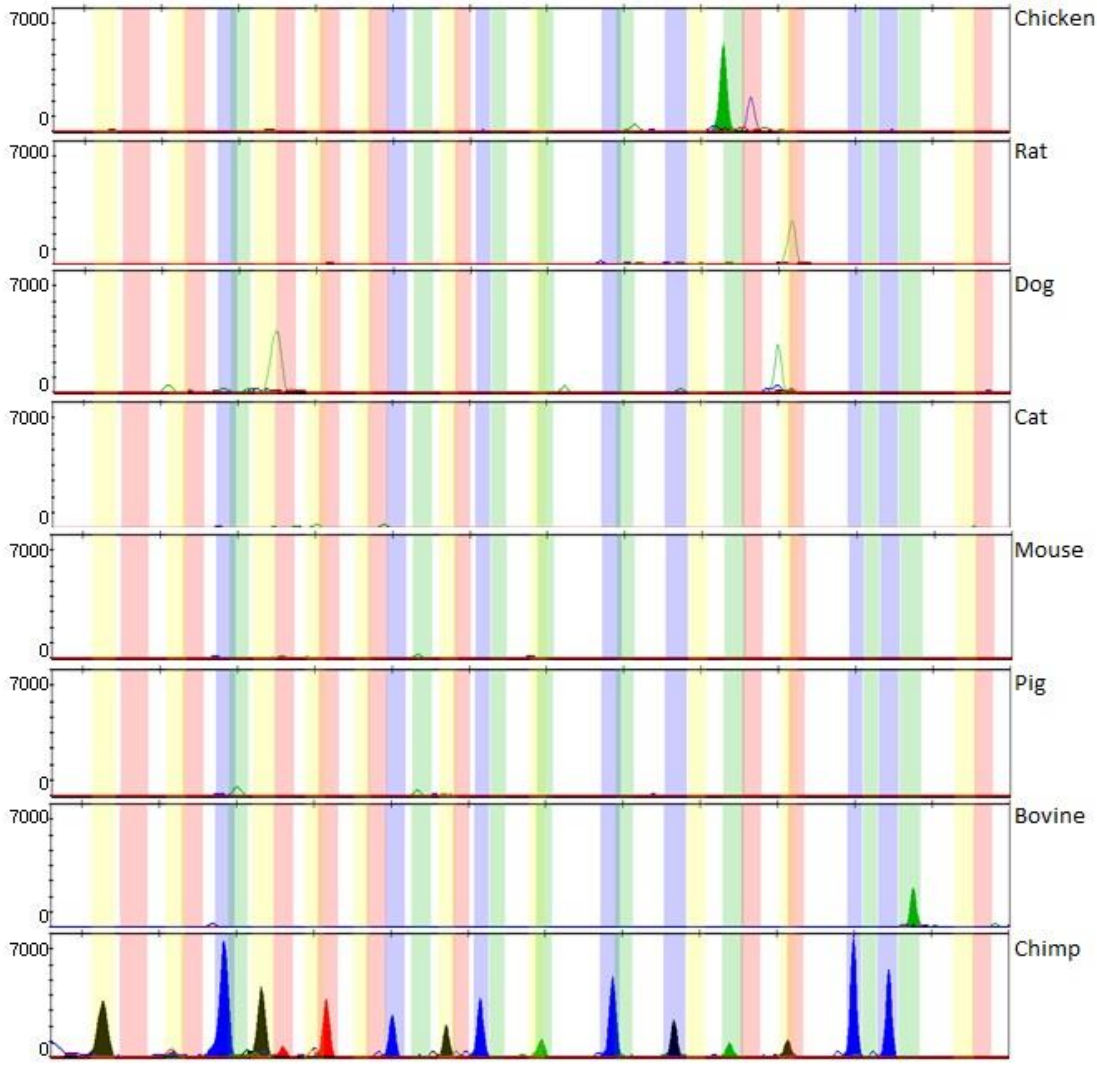


Figure 6 HIrisPlex-S Species Specificity

5.2.3 Concordance Casework Samples

The final results obtained for the validation of the HIrisPlex-S system were the casework concordance results. This assessment ensures that this assay performs the exact same despite use on different machines and having different laboratory personnel setting up the assay. The results for this testing can be found in Table 14 located in *Appendix N*. Single source samples from a variety of sources such as blood, saliva, semen, touch DNA, vaginal secretions, and mucosa all produced clean profiles. Samples from dried blood, dried saliva, and touch DNA all resulted in clean electropherogram, but low peaks compare to the other single source samples. The same result was true for the blood sample with heme inhibition in that it produced a clean electropherogram with low peak heights. The analysis interpretation based upon the profile produced was consistent with the known phenotypic information for those samples. During the analysis of such samples, one touch DNA and one frozen blood sample had peak heights below the standard interpretation threshold and therefore a complete genotypic profile could not be obtained. Based upon these results, the phenotypic profile for these two samples could not be produced. Both samples had concentrations that were at or below the minimum input concentration of 60pg/µl for the assay.

Mixture profiles produced during this concordance set were observed to contain at least two or more individuals. Based upon this conclusion an interpretation of the results could not be made as mixture interpretation of SNP profiles on capillary electrophoresis is not possible since a major and minor contributor cannot be determined with SNPs as opposed to STRs.

Through this concordance testing only one sample produced an incorrect phenotypic profile. This sample was that of a mixture, but because the peak distribution due to degradation looked as though it were a single source sample it resulted in an incorrect prediction. This information along with all other concordance data is in Table 14 in *Appendix N*.

CHAPTER 6. RESULTS AND DISCUSSION: HAIR MORPHOLOGY SNP AND GENE ASSOCIATION THROUGH MASSIVE PARALLEL NEXT GENERATION SEQUENCING

91 SNPS were sequenced on the Illumina Miseq to determine the possible significance they may play in determining the type of hair structure an individual may exhibit. All 91 SNPs were amplified successfully with rs12130862, rs3001978, rs11575161, rs67587000, rs585583, rs2227311, and rs12913832 all being verified by manual reference. All SNPs for this assay can be found in Table 5 located in *Appendix D*.

1026 samples were successfully sequenced at these 91 loci. 743 of these individuals were used in a study of 28,701 subjects from 9 cohorts of multiple geographic origins for multiple Genome Wide Association Studies [manuscript accepted].

6.1 Development of a Custom MPS Assay

The concept of Massive Parallel Sequencing, also known as Next Generation Sequencing, follows the same basic principles as Capillary Electrophoresis (CE) sequencing. In CE sequencing, DNA polymerase orchestrates the synthesis of millions of copies of a template DNA strand while incorporating fluorescently tagged dNTPs, dioxyribonucleotide triphosphates and ddNTPs (for chain termination). These florescent tags are identified by the genetic analyzer based upon their different wavelengths and therefore result in the sequence being read. MPS using cycling reversible termination allows for the overall concept to be the same, but hundreds of samples can be run simultaneously across millions of different DNA fragments instead of one individual sample at a time at one location.

The custom protocol designed for the Illumina Miseq Next Generation Sequencer was based on a successful protocol developed in our laboratory with reference to Bronner *et al*. [27]. Each of the primers were designed to isolate between 100-300bp around the SNP of interest, much like a flanking primer. The same procedure was used as described in section 2.3 to design these Massive Parallel Sequencing primers. The only distinct difference with these types of primers is that the forward and reverse must have specific adapter sequences that allow the fragments or “amplicons” to adhere to the lawn found on

the Miseq flow cell. Owing to the increased number of SNPs for this reaction, it was difficult to have all the primers fall under an isolated temperature range concerning the melting temperature. Therefore, it became necessary to adapt the Miseq PCR amplification protocol to an appropriate type of PCR called a “touchdown PCR”. This type of PCR allows the cycle temperature to slowly decrease by 0.3ºC to ensure that the primers all bind and amplify appropriately based upon their specific temperature range during the annealing phase. In the case of this reaction it was necessary to extend the range from 57ºC to 64ºC.

6.2 Development of a Custom Bead Clean Up Method

Once the protocol for the Massive Parallel Next Generation Sequencing Assay was decided upon and the primers were successfully created, it was essential to develop a protocol for bead clean up that would rival that of the AmPure Bead cleanup but at a fraction of the cost. Bead cleanup is necessary over column cleanup to remove primer-dimers and excess primers. To do this, it was necessary find the correct suspension of the PolyEthylene Glycol (PEG) (Fisher Scientific International Inc.) mixture with the Sera-Mag Speed Beads (GE Healthcare, Waukesha, WI) to ensure that their specificity is only for the fragment sizes that needed removed, not affecting the 100 to 300 bp fragments, but removing the smaller fragments. After several trials at various concentrations of PEG, 24% PEG with 5M NaCl seemed to be the most successful to provide the appropriate binding affinity for the beads to the DNA.

To assess the beads to make sure that they operated proficiently for cleanup, a standard ladder was used as the cleanup sample. A dilution of the Fermentas Ultra Low Range Ladder (Fisher Scientific International Inc., Hampton, NH) to 10ng/µl was prepared. The standard bead cleanup protocol mentioned in section 3.5 was then used to clean up 2-4 duplicate ladder samples. Once the clean-up is achieved, the samples are then run on a 4% agarose gel at 120V for 60 minutes. The agar percentage was increased to 4% to help increase the resolution of the smaller band sizes. The bands should be visible between 100-300bp. A successful bead cleanup validation is displayed in Figure 7.

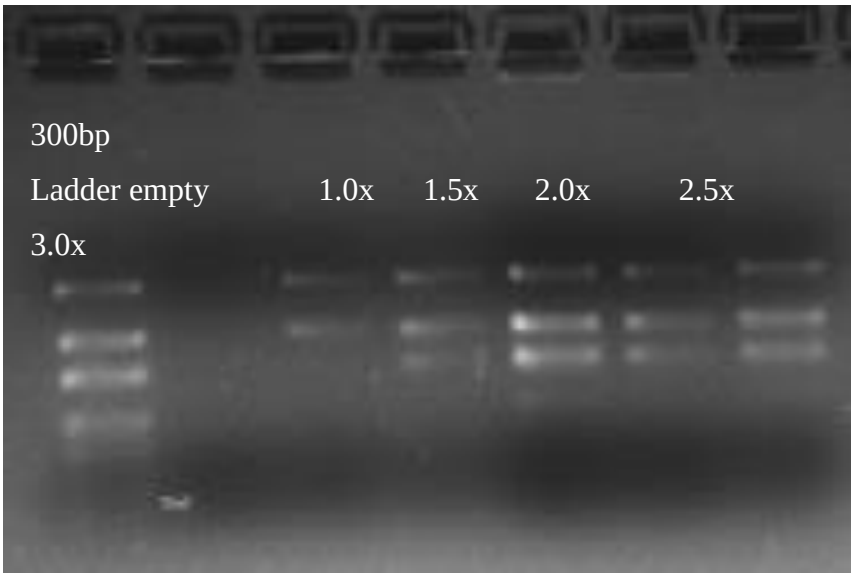


Figure 7 Successful Bead Preparation Validation with ratios of beads to sample of 1.0x to 3.0x.

6.3 Perfecting NGS Primers

For the SNPs identified as essential to evaluating hair morphology, the designed primers were tested through two distinct methods to ensure that they worked successfully before applying them to a Miseq analysis. The first step to ensure the success of the primers, is to perform singleplex reactions, reactions in which only the primers of interest is used while water is substituted for the rest of the primers, and then ran through the standard first PCR protocol mentioned in section 3.5. This was done for all 91 primer sets. They were then ran on 4% agarose gel to see if a band of DNA was formed on the gel indicating successful amplification through the PCR with that specific primer.

For any of the primers that did not successfully form a band (3 primers) on the agarose gel, the 2100 Agilent Bioanalyzer using the Agilent DNA 1000 Kit (Agilent, Santa Clara, CA) was used to also visualize as to whether the primer successfully amplified the DNA as intended or not. Primers rs30077681 and rs670741 did not amplify successfully the first time and were completely redesigned for successful amplification. Primer concentrations were adjusted based upon visibility of the banding on the gel (the lighter the bands mean the concentration needs increased) and on the Bioanalyzer.

Once all of the primers were identified as successfully amplifying the regions of interest, they were multiplexed to one of three groups based upon the Autodimer results [62]. These groups were then verified on the Bioanalyzer as working well without inhibition of one another before being applied to the assay on the Miseq. This can be seen in Figure 8 below.

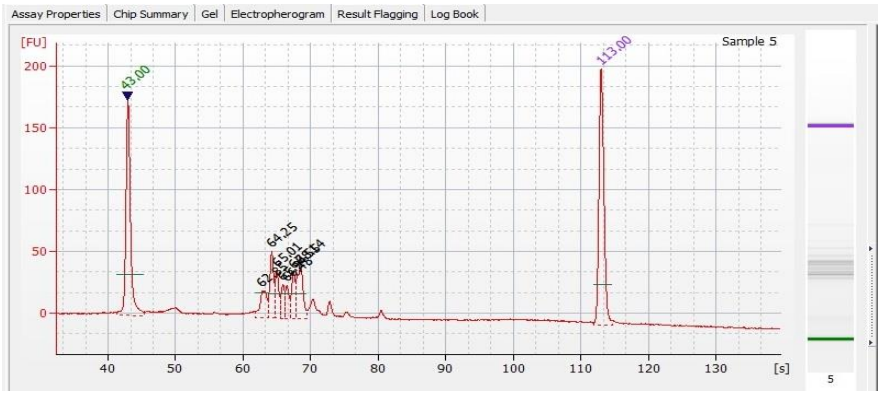


Figure 8 Successful Group Multiplex for Miseq Assay on Bioanalyzer, peaks just after 40 and 110 indicate standards.

6.4 Sequence Alignment and Hair Structure Correlation Results

All this assembly and conversion of the Miseq fastq files was performed on an iMac in a UNIX environment. To evaluate the data that was synthesized as a result of this Miseq run, a custom genome had to be assembled first to align the samples with. This custom genome allows for the isolation and alignment of the DNA to areas of interest as opposed to the whole genome, making evaluation of data a much simpler task. For the type of data that was produced, referred to as amplicons, aligning these to the entire genome would be error prone and memory intensive. Instead, by assembling a genome appropriate for this assay, we can ensure that the amplicons align exactly where intended. In sequencing our SNPs of interest, an area of about 400bp surrounding the SNPs of interest were isolated based upon sequence information provided by NCBI. From this data, a fasta file was created to which the Miseq sequencing data was aligned as described in section 3.5.

In order to access the significance of the genotypic results ascertained from the Miseq hair structure assay, a Pearson's correlation matrix was generated in R using the

Corr and pcorr test from the Psych package [63]. Each SNPs correlation was tested against the hair structure type identified by the participants. This correlation was ran both with and without correcting for age and sex. All SNPs of significance were seen in both correlation, the matrices are located in Tables 13 and 14 in *Appendices O* and *P*. A Bonferroni correction was performed due to the nature of the multiple testing present for the 91 SNPs the p-value was adjusted to 0.000549. 14 SNPs were found to be significant in both the corrected and uncorrected correlation based upon the adjusted p-value. These 14 SNPs were rs1150606, rs11568820, rs11803731, rs12913832, rs17646946, rs310642, rs3827760, rs4672907, rs4845418, rs5919324, rs80293268, rs95365032, rs2219783, and rs9568036.

The correlations for rs1150606, rs11568820, rs11803731, rs12913832, rs17646946, rs310642, rs3827760, rs4672907, rs4845418, rs5919324, rs80293268, and rs95365032 were not mirrored in the Meta-Analysis study as significant. This difference in significance can largely be attributed to the vast difference in sample size as our study only had 1026 individuals (reduced power) while the meta-analysis had 28,701 individuals. The composition of our samples may also contribute to this variation in significance as 743 of these samples were strictly from European populations by design.

6.5 Correlation in Conjunction with Meta-Analysis Collaboration

While twelve of the SNPs found to be significant in the Pearson correlation performed in our assay did not reach significance in the meta-analyses, SNPs rs2219783, rs310642 and rs80293268 were deemed significant in both our correlations and the Meta Analysis [manuscript accepted].

Intronic variant, rs2219783, is located in the *LGR4* gene. *LGR4* triggers the *Wnt* signaling pathway via G-protein couple receptors, this is known to control hair growth and structure. This gene been shown in knockout mice to reduce the effects of *EDAR*, *LEF1*, and *SHH*. These three genes are all required for hair follicle development and thus it would be logical for this SNP to possibly help control the structure of the hair [40].

Intronic variant, rs310642, located in *PTK6* encodes protein tyrosine kinase, which functions in epithelial cells as an intracellular signal transducer [52]. After inspecting the data obtained for all individuals at this SNP, the genotype of homozygous minor allele

(CC) seems to lend an association with a curly haired phenotype. Further investigation needs to be conducted to determine the significance of this SNP in relation to hair morphology.

The SNP rs80293268 was found to be significant and is located intergenic of *ERRFI1* and *SLC45A1*. *ERRFI1*, also known as *RALT*, encodes an inhibitor of epidermal growth factor receptor. This is essential in hair follicle development along with homeostasis of the epidermis [52, 53]. An overexpression of *RALT* has also been identified in mice to result in a wave phenotype, which is consistent with a wavy coat and curly whiskers as well [64]. After the assessment of this SNP for all individuals in our dataset, the genotype of homozygous minor allele (CC) seems to lend an association with a straight-haired phenotype. Further investigation needs to be conducted to determine the significance of this SNP in relation to hair morphology.

CHAPTER 7. RESULTS AND DISCUSSION: MIGRATING HIRISPLEX- S TO THE MASSIVE PARALLEL NEXT GENERATION SEQUENCING PLATFORM

7.1 InnoQuant for Validation Sample Concentration

The InnoQuant Human DNA Quantification and Degradation Assessment Kit was used to assess the quality of the validation samples for the Miseq HIrisPlex-S assay. Before running the actual samples, DNA standards ranging from 0.009 ng/µl to 20ng/µl were ran to insure proper calibration of the system. Based off the calibration, the instrument was set up to run the InnoQuant assay. Five point standards were run along with the validation samples [58]. This is to ensure proper function of the assay. DNA standard at 100ng/µL was provided with the assay and was diluted using InnoQuant Dilution Buffer A to make the five point standards. These standards range from 0.005ng/µL to 20 ng/µL. Prior to running, the instrument was calibrated for the dyes: Cy5, Cy3, and ROX.

The validation samples were diluted to 1ng/µL, based on concentrations previously determined using Qubit dsDNA HS assay, to assess for correct concentration and degradation using InnoQuant assay. To run the validation samples a master mix containing 10µl of Agilent Brilliant Multiplex QPCR Master Mix, 0.3µL Agilent Reference Dye, and 7.7µL InnoQuant Primer Mix was made. 2µl of sample or standard was added to the well initially, followed by 18µl of the master mix. The InnoQuant H-dye HID Run Template_v1.0 was used to set up the appropriate PCR run conditions, dye configurations, and standard curve replicates. The cycling parameters specified were an initial heating time of 10 minutes at 95^{O}C, followed by 32 cycles of 15 seconds at 95^{O}C then 2 minutes at 61^{O}C [58].

Once the run was complete, the R^2 and efficiency values were evaluated to ensure that the run was successful. For a successful run, the reference values for these runs were required to have R2 > 0.98 and the efficiency values have a passing range of greater than 90%, but less than 110%. The validation samples were ran in two separate runs. The first run contained non-casework validation samples. The long read R^2 value was 0.987 and the efficiency percentage was 109.579. The short read for non-casework validation

samples R^2 value was 0.993 and the efficiency percentage was 108.478. For the casework validation samples, the long read R^2 value was 0.999 and the efficiency percentage was 106.54. For the short read R^2 value was 0.998 and the efficiency percentage was 94.119.

Based upon the concentrations determined through the InnoQuant assay, the samples' original concentrations were determined and the validation samples were diluted to the target concentration. Three of these samples were then spot checked through InnoQuant.

7.2 SBE Product Formation

After visualizing the primer bands, it was necessary to multiplex the primers to determine if any of the primers inhibited one another. It was also decided that a standard ExoProStar clean up would be performed on these multiplexed samples and then, using the already established SBE primers for the HIrisPlex and HIrisPlex-S, a standard SBE reaction and SAP cleanup were performed. The concentrations used for these SBE primers were used as though following the standard HIrisPlex-S assay found in Tables 2 and 3 in *Appendices B and C*. While SBE primers are not at all necessary for NGS, performing this check on the NGS primers allowed for the amplification of the SNP of interest to be observed, therefore ensuring that the primer would anneal and amplify appropriately. All reactions were performed on the same sample at a concentration of 2 ng/µL to take away the variability of concentration change in a variety of samples.

7.3 Singleplex to Standardize Coverage

To continue forth in creating a biological assay that can be used on a widespread scale and to avoid preferential amplification of fragments, the next step was to standardize the amplification coverage of each primer on one control DNA sample as mentioned above to standardize initial concentration. To do this, concentrations were adjusted based upon the band results from section 4.3. The standard PCR protocol used for this assay mentioned in section 4.3 was performed and then rather than running the sample on a gel, the samples were ran on the Qubit to determine the quantity of product

produced prior to clean up. The goal concentration was +/- 5ng/µL of 28 ng/µL for that sample. Several trials of adjustments were made to get these primers into that range.

7.4 Optimizing PCR Conditions and Primer Concentrations

After completing a successful Miseq run to observe the performance of the primers to determine optimal concentrations, it became apparent that it was necessary to adjust our PCR. Based upon the analysis from this first run, it was apparent that the primers were not extending as they should be, possibly due to the PCR settings. It was at this point the cycling parameters were adjusted to an initial heating time of 10 minutes at 95^{O}C, followed by 34 cycles of 30 seconds at 95^{O}C then touchdown at 62°C decreasing by 0.3 for 30 seconds and then 60^{O}C for 10s, finally 5 minutes at 60°C.

Due to the possibility of SNP dropout and preferential amplification, it was necessary to analyze the coverage once the reaction had been deemed successful in that the primers function appropriately. An analytical threshold of 100x coverage was set as our parameter for determining whether a primer's concentration needed to be adjusted. Primers for SNPs rs12821256, rs683, rs1800414, and rs1470608 all had coverage below 100 reads, but above 50 read. This allowed us to determine that the primer had worked successfully, but the concentration needed to be increased to meet our coverage threshold. Primers for SNPs rs6119471 and rs1545397 however did not produce reads anywhere near our necessary minimum and at this point it was necessary to redesign the primers again to allow for the minimum coverage to be achieved. Besides the previously mentioned six primers, all the remaining primers performed at coverages significantly higher than the analytical threshold set forth. Based upon this observation it would be necessary to increase the concentration of the underperforming primers while decreasing the concentration of those performing at a much higher coverage.

The lack of coverage for rs1545397 was remedied by redesigning the forward primer while keeping the previously designed reverse primer. This resolved the lack of coverage at this SNP. The lack of coverage present on rs6119471 was remedied by redesign of the forward primer and reducing the overall fragment size (including adapters) to a much smaller size of 128bp from the previous size of 225bp.

Singleplex runs of each primer were performed at 0.5µm for the MC1R set primers and 0.4 µm for the remaining primers. The appropriate PCR method was determined to be a touchdown PCR with a temperature range of 57˚C to 62˚C with a 0.3ºC interval decrease to optimize the annealing of the primers for product formation. At this point the product was then not cleaned up, but rather ran on a 4% agarose gel to determine if band formation occurred indicating product formation as well as if the product appeared to be the appropriate size.

7.5 Perfecting the Bead Cleanup Protocol

To make the current bead protocol in place work successfully for the Miseq HIrisPlex-S assay it was essential to determine the ratio of beads to sample that successfully remove excess primers so that the sequencing can be focused on just the amplicons we have isolated for sequencing. The ratios and the subsequent size fragments that can be removed can be found in the table below.

Table 1 Sample to Bead Ratio for Bead Cleanup

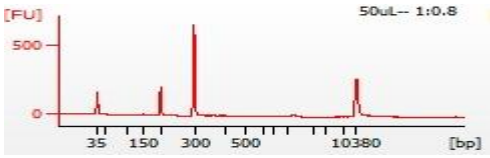
Sample to Beads Ratio	**Example of Sample to Beads Ratio**	**Size Selection Comments**	**Bioanalyzer Electropherogram Profile**	**Bioanalyzer Gel profile**
1: 0.8	20uL: 16uL	Removes everything less than 200bp; Only 200bp and 300bp bands are visible on bioanalyzer with a faint band for 150bp		

Table 1 continued

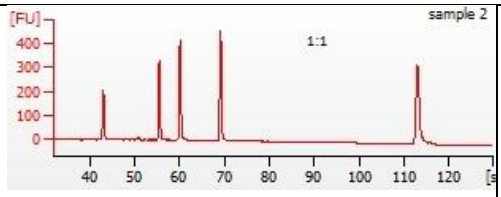
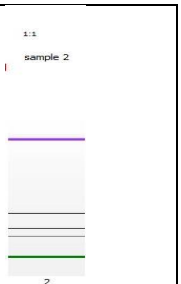
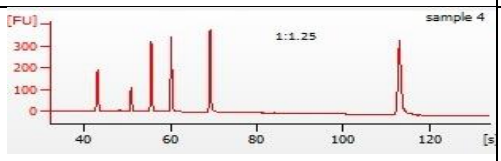
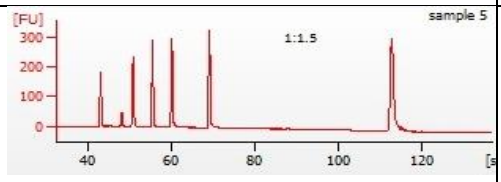
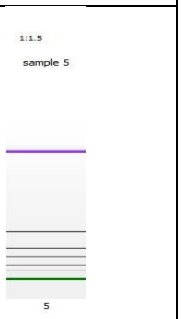
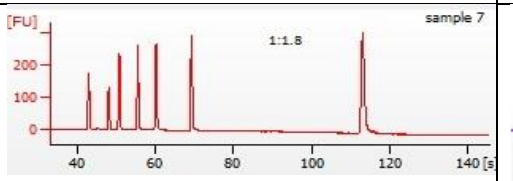
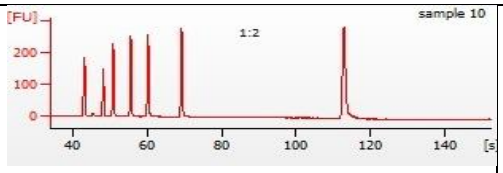
1:1	20uL: 20uL	Removes everything less than 150bp; Only three bands – 150, 200 and 300bp are visible.		
1:1.25	20uL: 25uL	Removes everything less than 100bp; Four bands- 100,150,200 and 300bp are visible		
1:1.5	20uL: 30uL	Removes everything less than 75bp; Five bands are visible- 75,100,150, 200 and 300bp; 75bp is not very strong		
1:1.8	20uL: 36uL	Five strong bands: 75-300bp;50bp starts to show up		

Table 1 continued

1:2	20uL: 40uL	Five strong bands: 75-300bp; 50bp shows up as well		

For this assay fragments under 100bp needed to be removed so a cleanup in the ratio of 1:1.25 is necessary to successfully clean up the reaction to be ran on the Miseq. The standard protocol developed in section 3.3 was then followed as stated.

7.6 Finalized Protocol for HIrisPlex-S Massive Parallel Next Generation Sequencing Assay

The coverage and sensitivity assessment of the optimization run are ongoing but current run results can be found in Table 17 located in *Appendix R*. Through this run we could determine the successful amplification of all areas of interest and perform an assessment of the coverage at each SNP. Based upon this coverage assessment, concentrations will be adjusted and optimized to produce and efficient Forensic DNA Phenotyping tool for the Illumina Miseq. Once final optimization of the concentrations is made validation plates will be ran and assessed.

Depth of coverage was assessed for samples whose concentration was approximately 500pg. Each SNP was individually analyzed for read depth and average coverage across all the 500pg samples were obtained. Based on the analysis, 10 SNPs had coverage < 100x, whereas certain SNPs had greater depth of coverage. The SNPs that had <100x coverage are rs1805009, rs12821256, rs12203592, rs2402130, rs12913832, rs683, rs2238289, rs17128291, rs1470608 and rs1426654. Whereas the SNPs that had greater depth of coverage were rs6119471- 35667x, rs1545397- 4078x, rs10756819- 1254x and rs8051733- 1190x. Further optimization needs to be performed on these SNPs in order to balance the reaction such that there is no preferential amplification during the multiplex

PCR reaction. Ensuring the depth of coverage is uniform on an average reduces the possibility of SNP drop out for low concentration or degraded samples. For example, the primer concentration for each of the low coverage SNPs would be increased and the primer concentration for each of the greater coverage SNPs would be decreased to bring a balance in the PCR amplification process. The standard deviation for each of the SNPs across all the 500pg samples is very high and this could be attributed to the variance in quality of the samples and concentration that comes along with an in-house validation set. This issue can be alleviated by optimizing the coverage using stable control DNAs (9948, 9947A and 007) at various range of concentrations to achieve the desired coverage. These results can be found in Table 17 in *Appendix R*.

7.7 Casework Assessment

Depth of coverage was assessed for two saliva casework samples, one that was fresh and one that was degraded. Each SNP was individually analyzed for read depth across both samples. Based on the analysis, 34 SNPs for the non-degraded saliva casework sample had coverage > 100x, whereas certain SNPs had greater depth of coverage. The SNPs that had >100x coverage are rs312262906, rs11547564, rs885479, rs1805008, rs1805005, rs1805006, rs1805007, rs201326893, rs2228479, rs1110400, rs28777, rs16891982, rs4959270, rs1042602, rs1800407, rs2402130, rs2378249, rs128963999, rs1393350, rs3114908, rs1800414, rs10756819, rs17128291, rs6497292, rs1129038, rs1667394, rs1129038, rs1470608, rs6119471, rs1545397, 6059655, rs12441727, rs3212355, and rs8051733. Whereas the SNPs that had greater depth of coverage were rs6119471- 33315x, rs1545397- 7994x, rs10756819- 1429x and rs8051733- 3962x. However, for the degraded samples, 25 had a coverage >100x, rs312262906, rs11547564, rs885479, rs1805008, rs1805005, rs1805006, rs1805007, rs201326893, rs2228479, rs1110400, rs16891982, rs4959270, rs1042602, rs2378249, rs128963999, rs1393350, rs3114908, rs10756819, rs6497292, rs1667394, rs6119471, rs1545397, 6059655, rs12441727, and rs8051733. SNP rs1805009 dropped out in the degraded analysis. Further optimization needs to be performed on these SNPs to balance the reaction such that there is no preferential amplification during the multiplex PCR reaction. Ensuring the depth of coverage is uniform on an average, reduces the possibility

of SNP drop out for low concentration or degraded samples. For example, the primer concentration for each of the low coverage SNPs would be increased and the primer concentration for each of the greater coverage SNPs would be decreased to bring a balance in the PCR amplification process. While primer adjustment does need to take place, the overall outcome of the casework sample displayed the same decrease in coverage for the degraded sample, but still exhibited coverage over our interpretative threshold. More casework samples will need to be ran and examined during the validation to ensure adequate performance on these types of samples.

7.8 Mixture Assessment

A large problem in the forensic science field today involves the deconvolution of mixtures. Majority of forensic samples that come into laboratories are mixture samples and identifying the contributors could lead to a much higher rate of solving these crimes. Through the design and development of the HIrisPlex-S MPS assay we have identified a possibility that indeed MPS technologies may be able to help solve this problem.

For our validation plate, we ran mixture profiles of two individuals in ratios of 1:1, 1:2, 1:5, and 1:10. The goal behind this was two-fold, in that we wanted to see if we could separate the mixture by individuals with different phenotypes and well as with individuals with like phenotypes. What we observed upon analyzing these profiles is that we may be able to pull apart individuals of different phenotypic characteristics. Below are two examples of mixtures that were analyzed to see if successful separation of individual profiles could be achieved. Both exemplar mixtures below are in 5ng:500pg ratios. In both mixtures, the pigmentation SNP of interest was identified first. Once this SNP was identified, variation in neighboring SNPs to the pigmentation SNP were also identified. We aimed to identify two or more neighboring SNPs to ensure accuracy of observation. A close investigation of SNPs in the adjacent regions of HPS SNPs, led to the identification of unique patterns in neighboring SNPs, allowing for us to identify individual contributors rather than just heterozygous loci. To also rule out heterozygosity, we looked at the coverage of both the pigmentation SNP as well as the neighboring SNPs. In all the 5ng:500pg mixtures, the coverage was in a 1:10 ratio of minor to major contributor therefore reinforcing the identification of individual contributors.

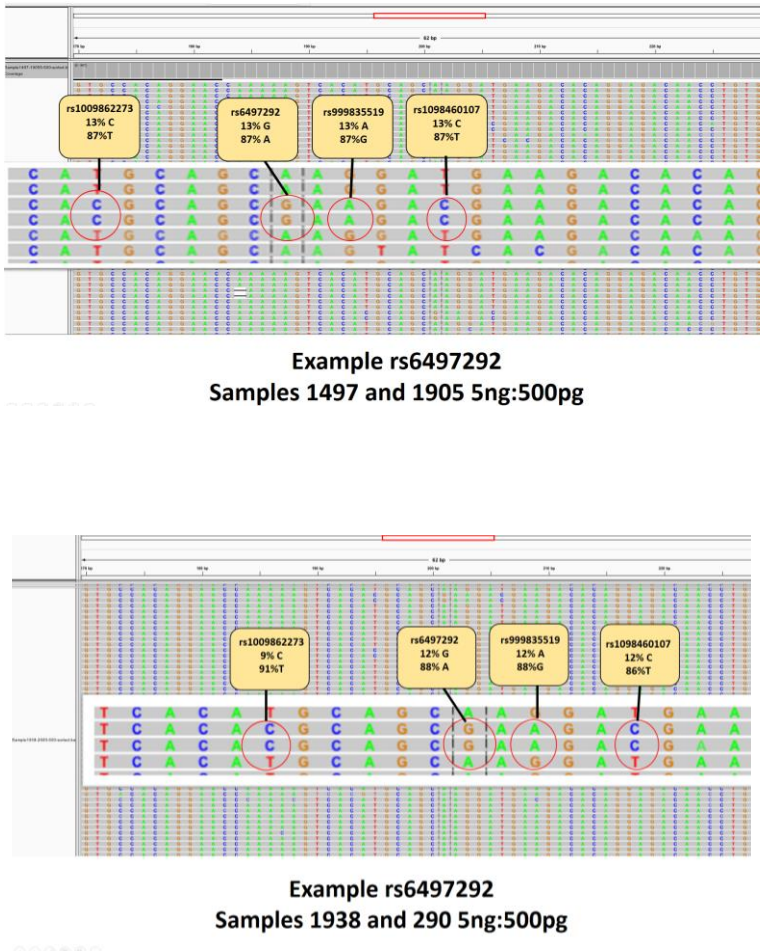


Figure 9 Two 5ng:500pg Mixture Deconvolutions

While this does provide us with an interesting possibility the mixtures can be pulled apart through this method, much more research must be done to determine how to effectively complete this task. To increase the accuracy of this deconvolution, including more neighboring SNPs in the regions of interest may prove useful. This could be successfully achieved through intentionally including more neighboring SNPs in primer design. Having multiple SNPs in one region of interest would potentially increase the likelihood that a pattern in genotypes could be observed, allowing for an individual to be pinpointed for each sequence strand of DNA. Observation at various mixture ratios also needs to be analyzed to determine the success of this technique at all ratios. While we do believe that the separation of these mixtures is possible, more analyses and research is needed to be able to fully assess the success rate of mixture deconvolution.

7.9 Comparison of Capillary Electrophoresis against the Miseq for HPS Genotyping

One of the key motivating factors behind developing this HIrisPlex-S assay on the Next Generation Sequencing platform, in addition to the vast amount of information these assays provide is to reduce the time and cost of performing these predictions to allow these to be more applicable to the forensic science field. To determine the cost efficiency of these assays it was necessary to compare the cost of the two methods when running the same amount of predictions on the same number of samples. The table for 384 samples can be found in Table 18 located in *Appendix S*.

For 384 samples, the cost to run the samples on the Capillary Electrophoresis, the cost came to about $4495; whereas running 384 samples on the Miseq, costs about $2911. A capillary electrophoresis based HIrisPlex-S run for 384 samples takes about 95 hours in total to run from the Flanking PCR step all the way through the genotyping. However, for a Miseq run for 384 samples it only takes about 45 hours. As illustrated, the Miseq reduces not only the cost of the run, but also the run set up time by approximately half. Therefore, this illustrates that running 384 samples on the Miseq is half of the cost and the amount of time that it would take to run the same number of sample predictions on the Genetic Analyzer, thus supporting the idea that the implementation of this in forensic laboratories would not only provide more information to law enforcement, but also save time and money when doing so.

CHAPTER 8. CONCLUSIONS

The first goal of this research was to assist in generating a skin color prediction model and in addition develop and validate the HIrisPlex-S assay for the prediction of eye, hair, and skin color from DNA. 329 individual samples from our laboratory were successfully used in model generation to increase the geographic distribution of the dataset therefore allowing us to obtain a globally accurate skin color prediction tool. 194 model test subjects were used to compare our 41-SNP prediction model to the Snipper model set forth by Maronas *et al*. [3]. This comparison allowed us to demonstrate that our five-category skin pigmentation tool did not only produce more accurate predictions, but also allowed our model to illustrate its ability to predict skin color independently of biogeographical ancestry. We could successfully produce a multiplex assay that targeted 17 skin color predictive SNPs to be used in conjunction with the 24 SNPs identified for the HIrisPlex assay allowing a full profile to be produce through of free online web-based tool. This assay can produce a full 41-SNP profile to with a minimum DNA input value of 63pg. This human-specific assay generates probabilities for 3 eye colors, 4 hair colors, and 5 skin color categories. From this point, research into further physical appearance characteristics is necessary to improve prediction accuracy and add new phenotypic traits. This is especially true when looking at “intermediate” eye color and pale-very pale skin pigmentation as these are areas where more accurate predictions must be worked towards.

The second goal of this research was to develop a MPS assay that migrates the HIrisPlex-S assay from the Capillary Electrophoresis to the Illumina Miseq. This assay develops the 41 SNP profile through sequencing by synthesis. While optimization is ongoing, possible differentiation in sources of mixture profiles has been observed in addition to complete sample profiles. Further optimization including adjusting primer concentrations is necessary to further develop this assay. Once optimization is finalized the developmental validation data must be generated for assessment. This assessment will allow the parameters and success of the assay to be determined. Further research following optimization is necessary for the addition of new physical appearance traits along with the ability of this assay to discern mixture samples.

The final goal of this research was to develop a MPS assay that generated data to assess the correlation of 91 different SNPs with categorical hair morphology. 1026 samples were genotyped in our laboratory and assessed at these 91 SNPs of interest. From our data set we could identify a possible correlation between the SNPs rs2219783, rs310642, and rs80293268. We could observe a possible association between the straight-haired phenotype and the homozygous minor allele (CC) for rs80293268 as well as a possible association between the curly haired phenotype and the homozygous minor allele (CC) for rs310642. Further research and functional animal and human studies into these SNPs, is essential in conclusively identifying the role these genetic variants play in hair morphology.

APPENDIX A

Table 2 SNPs for HIrisPlex-S

Gene	SNP	Chromosome	Location GRCh38.p7	Major Allele	Minor Allele	PCR Primer	Flanking Forward Primer	Flanking Reverse Primer	Product Size(bp)	Reference
MC1R	N29InsA(rs312262906)	16	89919341	A insertion	-	MC1Rset1	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGCAGGGATCCCAGAGAAGAC	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGTCAGAGATGGACACCTCCAG	117	Branicki *et al.* , Pospiech *et al.* , Grimes *et al.*
MC1R	rs2228479	16	89919531	G	A	MC1Rset2	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGTCCAGCCTCTGCTTCCTG	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGAGCGTGCTGAAGACGACAC	147	Branicki *et al.* , Pospiech *et al.* , Grimes *et al.*, Strum *et al.*
MC1R	rs1805005	16	89919435	G	T					Branicki *et al.* , Pospiech *et al.* , Grimes *et al.*, Strum *et al.*
MC1R	rs1805006	16	89919509	C	A					Branicki *et al.* , Pospiech *et al.* , Grimes *et al.*, Strum *et al.*
MC1R	rs11547464	16	89919682	G	A	MC1Rset3	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCTGGTGAGCTTGGTGGAGA	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGTCCAGCAGGAGGATGACG	158	Branicki *et al.* , Pospiech *et al.* , Grimes *et al.*, Strum *et al.*
MC1R	rs885479	16	89919745	G	A					Branicki *et al.* , Pospiech *et al.* , Grimes *et al.*, Strum *et al.*
MC1R	rs1805007	16	89919708	C	T					Branicki *et al.* , Valenzuela *et al.* , Strum *et al.*, Sulem *et al.* , Duffy *et al.*
MC1R	rs1805008	16	89919735	C	T					Branicki *et al.*, Hart *et al.* , Strum *et al.*
MC1R	Y152OCH (rs201326893)	16	89919713	C	A					Pospiech *et al.* , Grimes *et al.*
MC1R	rs1110400	16	89919721	T	C					Branicki *et al.* , Pospiech *et al.* , Grimes *et al.*
MC1R	rs1805009	16	89920137	G	C	MC1Rset4	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCAAGAACTTCAACCTCTTTCTCG	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCACCTCCTTGAGCGTCCTG	106	Branicki *et al.* , Pospiech *et al.* , Grimes *et al.*
SLC45A2	rs28777	5	33958853	C	A		TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTACTCGTGTGGGAGTTCCAT	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGTCTTTGATGTCCCCTTCGAT	150	Han *et al.* , Pospiech *et al.*
SLC45A2	rs16891982	5	33951587	C	G		TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTCCAAGTTGTGCTAGACCAGA	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCGAAAGAGGAGTCGAGGTTG	128	Lui *et al.*, Valenzuela *et al.*, Branicki *et al.* , Pospiech *et al.* , Grimes *et al.*
KITLG	rs12821256	12	88934557	T	C		TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGATGCCCAAAGGATAAGGAAT	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGGAGCCAAGGGCATGTTACT	118	Sulem *et al.* , Branicki *et al.* , Pospiech *et al.* , Han *et al.*
EXOC2	rs4959270	6	457747	C	A		TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTgagaaatctacccccacga	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGTGTTCTTACCCCCTGTGGA	140	Sulem *et al.* , Branicki *et al.* , Han *et al.*
IRF4	rs12203592	6	396320	C	T	rs12203592	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAGGGCAGCTGATCTCTTCAG	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGCTTCGTCATATGGCTAAACCT	126	Sulem *et al.* , Branicki *et al.* , Han *et al.* , Liu *et al.*
TYR	rs1042602	11	89178527	C	A	rs1042602	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCAACACCCATGTTTAACGACA	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGCTTCATGGGCAAAATCAAT	124	Stokowski *et al.* , Pospiech *et al.*
OCA2	rs1800407	15	27985171	G	A	rs1800407	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAAGGCTGCCTCTGTTCTACG	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCGATGAGACAGAGCATGATGA	124	Branicki *et al.* , Pospiech *et al.*, Liu *et al.*, Frudakis *et al.*, Duffy *et al.*
SLC24A4	rs2402130	14	92334858	G	A	rs2402130	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGacctgtctcacagtgctgct	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGTTCACCTCgatgacgatgat	150	Sulem *et al.* , Branicki *et al.* , Pospiech *et al.*
HERC2	rs12913832	15	28120471	A	G	rs12913832	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTCAACATCAGGGTAAAAATCATG T	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGGCCCCTGATGATGATAGC	150	Lui *et al.*, Valenzuela *et al.*, Branicki *et al.* ,Strum *et al.* Sulem *et al.* , Duffy *et al.*
PIGU	rs2378249	20	34630285	A	G	rs2378249	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCGCATAACCCATCCCTCTAA	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCATTGCTTTTCAGCCCACAC	136	Branicki *et al.*
SLC24A4	rs12896399	14	92307318	G	T	rs12896399	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCTGGCGATCCAATTCTTTGT	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGACCCTGTGTGAGACCCAGT	125	Lui *et al.* , Han *et al.* , Sulem *et al.*
TYR	rs1393350	11	89277877	G	A	rs1393350	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTTTCTTTATCCCCCTGATGC	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGGGAAGGTGAATGATAACACG	124	Branicki *et al.* , Han *et al.*, Liu *et al.*
TRYP1	rs683	9	12709304	A	C	rs683	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCACAAAACCACCTGGTTGAA	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGTGAAAGGGTCTTCCCAGCTT	138	Branicki *et al.* Liu *et al.*, Frudakis *et al.*
ANKRD11	rs3114908	16	89317316	G	A	rs3114908	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCAGAACACAGCCACACCCTA	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCATAAAGGGGTCACCAGCAA	99	Law *et al.*
OCA2	rs1800414	15	27951890	A	G	rs1800414	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGCTGCAGGAGTCAGAAGGTT	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGGGACAAACGAATTGAGGAA	197	Edwards *et al.* Lao *et al.*, Soejima *et al.*
BNC2	rs10756819	9	16858085	G	A	rs10756819	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAAAGCAAGCTCATGTTTCCA	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCGTCATGACTAGAAAAACACC AA	145	Liu *et al.*, Visser *et al.*
HERC2	rs2238289	15	28208068	T	C	rs2238289	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGGAACATGAAGATTTCCCAGT	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCTGATTCAGGTCTGCTGTCACT	112	Mengel-form *et al.*
SLC24A4	rs17128291	14	92416481	A	G	rs17128291	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCCAGCACTGCCAAAATAACA	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCTCTTTGGACCCATCACCTC	129	Liu *et al.*
HERC2	rs6497292	15	28251048	G	A	rs6497292	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTCTGCTGTAGAACCAATGTCC	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGAATTGCACCTGTAGCTCCAT	150	Liu *et al.*
HERC2	rs1129038	15	28111712	G	A	rs1129038	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGATGTCGACTCCTTTGCTTCG	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGACACCAGGCAGCCTACAGTC	137	Mengel-form *et al.* , Shekar *et al.* , Liu *et al.*
HERC2	rs1667394	15	28285035	G	A	rs1667394	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGCAGCTGTAGAGAGAGACTTTGAG G	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCACCATTAAGACGCAGCAAT	117	Sulem *et al.* , Eriksson *et al.* Duffy *et al.* , Liu *et al.*
MC1R	rs1126809	16	89919713	G	A	rs1126809	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTGTTTCTTAGTCTGAATAACCTTTT CC	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGGTGCATTGGCTTCTGGATA	100	Branicki *et al.*
OCA2	rs1470608	15	28042974	A	C	rs1470608	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTTTCTTGTGTTAACTGTCCTTACAA A	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGGGAAAATATGTTAGGGTTGATG G	218	Norton *et al.* , Valenzuela *et al.* , Stokowski *et al.*, Branicki *et al.*
SLC24A5	rs1426654	15	48134286	G	A	rs1426654	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTTCAGCCCTTGGATTGTCTC	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGTGAGTAAGCAAGAAGTATAAG GAGCA	212	Hart *et al.* , Maronas *et al.*, Stokowski *et al.*
ASIP	rs6119471	20	34197405	C	G	rs6119471	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGCAGGAGAATTGCTGGAACT	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGAACCCGAAGGAAGAGTGAAAA	174	Frudakis *et al.* , Hart *et al.*
OCA2	rs1545397	15	27942625	A	T	rs1545397	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGGTATAGGATTATTTGGGGAATG A	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGTGGAGATATAGAATTCACACAA CATAAA	144	Edwards *et al.*
RALY	rs6059655	20	34077941	G	A	rs6059655	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGTGAGGAAATCGAGGCTCAG	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGAGGAGAAAGCTGCAGATCCA	112	Jacobs *et al.*
OCA2	rs12441727	15	28026628	G	A	rs12441727	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGGGAAGAGACAGCTCCATGT	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGACAATCCTGGGAGGTACACG	137	Liu *et al.*
MC1R	rs3212355	16	89917969	C	T	rs3212355	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGAGTGAACCCAGGAAGATGC	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCATCAAAGGCAGACCTCTCG	199	Valenzuela *et al.*
DEF8	rs8051733	16	89957797	G	A	rs8051733	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAGGCGGTGGTCTCTCTCTC	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGTTGCAACAGGAGGGTCTAGG	124	Law *et al.*

APPENDIX B

Table 3 HIrisPlex Protocol for HIrisPlex-S

No.	SNP	Primer set	PCR Ps d STOCK(50um)	35	Dir	SNPs	Conc p	of 10um(ul)	of 20um(ul)	of 50um(ul)	35
1	N29insA_(DU)snF	**MC1Rset1** (1F&R)	**0.1 each F&R =0.5um**	**3.5**	1F	C/A	1.3um			0.13	4.55
2	rs11547464_(H)snF	**MC1Rset3** (3F&R)	**0.1 each F&R =0.5um**	**3.5**	2F	G/A	0.1 um	0.05			1.75
3	rs885479_(BD)snR	MC1Rset3			3R	C/T	1.25um			0.125	4.375
4	rs1805008_(FU)snF	MC1Rset3			4F	C/T	0.375um			0.0375	1.3125
5	rs1805005_(FU)snF	**MC1Rset2** (2F&R)	**0.1 each F&R =0.5um**	**3.5**	5F	G/T	1um			0.1	3.5
6	rs1805006_(FU)snF	MC1Rset2			6F	C/A	0.8um			0.08	2.8
7	rs1805007_(T)snF	MC1Rset3			7F	C/T	1.1um			0.11	3.85
8	rs1805009_(T)snF	**MC1Rset4** (4F&R)	**0.08 each F&R =0.4um**	**2.8**	8F	G/C	0.12um		0.03		1.05
9	Y152OCH_(FU)snF	MC1Rset3			9F	C/A	0.5um			0.05	1.75
10	rs2228479_(T)snF	MC1Rset2			10F	G/A	0.375um			0.0375	1.3125
11	rs1110400_(T)snF	MC1Rset3			11F	T/C	0.1um			0.01	0.35
12	rs28777_(july)snF	rs28777 - 5F+R	0.08 each F&R =0.4um	2.8	12F	A/C	1.2um			0.12	4.2
13	Rs16891982_(T)snF	Rs16891982 - 6F+R	0.08 each F&R =0.4um	2.8	13F	G/C	0.9um			0.09	3.15
14	rs12821256_(T)snR	rs12821256 - 7F+R	0.08 each F&R =0.4um	2.8	14R	A/G	0.12um		0.03		1.05
15	rs4959270_(T)snF	rs4959270 - 8F+R	0.08 each F&R =0.4um	2.8	15F	C/A	0.3um			0.03	1.05
16	rs12203592_(T)snF	rs12203592 - 9F+R	0.08 each F&R =0.4um	2.8	16F	C/T	0.2um			0.02	0.7
17	rs1042602_(T)snR	rs1042602 - 10F+R	0.08 each F&R =0.4um	2.8	17R	G/T	1.25um			0.125	4.375
18	rs1800407_(T)snF	rs1800407 - 11F+R	0.08 each F&R =0.4um	2.8	18F	G/A	0.1um			0.01	0.35
19	rs2402130_(T)snF	rs2402130 - 12F+R	0.08 each F&R =0.4um	2.8	19F	A/G	0.75um			0.075	2.625
20	rs12913832_(T)snR	rs12913832 - 13F+R	0.08 each F&R =0.4um	2.8	20R	C/T	1um		0.25		8.75
21	rs2378249_(T)snR	rs2378249 - 14F+R	0.08 each F&R =0.4um	2.8	21R	T/C	0.1um			0.01	0.35
22	Rs12896399_(T)snF	Rs12896399 - 15F+R	0.08 each F&R =0.4um	2.8	22F	G/T	1um			0.1	3.5
23	Rs1393350_(DU)snR	Rs1393350 - 16F+R	0.08 each F&R =0.4um	2.8	23R	C/T	1.1um			0.11	3.85
24	rs683_(H)snR	rs683 - 17F+R	0.08 each F&R =0.4um	2.8	24R	T/G	0.3um			0.03	1.05
			total 2,84ul	49.7					**total**	**1.76**	61.6

8/18/2015

			no.samples			no. of samples
Stocks	**PCR 1st set up (ul)**	Final conc	**35**	**SBE reaction**	**(ul)**	**35**
primers	**2.84**		99.4	primers	**1.76**	61.6
10x PCR gold buffer (no mg)	**1**	1X	35	snapshot rxn mix	**1**	35
Mgcl2 (25mM)	**1**	2.5mM	35	H20	**0.24**	8.4
dNtps (10mM each,combo 40mM)	**0.22**	220 uM	7.7	1st PCR product	**2**	
taq gold (5U/ul)	**0.3**	1.75U	10.5		5ul/rxn	105 total
H2O	**3.64**		127.4	**Thermo conditions: SBE rxn**		
Total reagents	**9**		315	96 °C for 2 min and 25 cycles of 96 °C for 10 s, 50 °C for 5 s and 60 °C for 30 s		
DNA	**1**					
	10ul total			**Clean up after SBE**		
Thermo conditions: 1st PCR				Sap(1U/ul)	1ul used /sample	(1)37°C - 45min, (2) 75°C - 15min
(1) 95 °C for 10 min, (2) 33 cycles of 95 °C for 30 s and 61 °C for 30 s, (3) 5 min at 61 °C.						
				CE run		
Clean up after 1st PCR				POP-7 on a 50 cm capillary length array.		
ExoProStar	2ul used per 5ul product of each sample	(1)37°C - 45min, (2) 80°C - 15min		Run parameters: injection voltage of 2.5 kV for 10 s, and run time of 500 s at 60 °C.		

new made up XOSAP	x no. of samples	35
1 EXO!		35
1 SAP		35
2 total	total for all samples	70

APPENDIX C

Table 4 HIrisPlex-S Protocol

No.	SNP	Primer set (all 50um stock)	PCR Ps d STOCK(50um)	35	Dir	SNPs tbc		Conc p	10um(ul)	20um(ul)	of 50um(ul)	35
1	rs3114908	F&R	0.1 each F&R =0.4um	3.5	1R	C/T	ct	0.08		0.07		2.45
2	rs1800414	F&R	0.1 each F&R =0.4um	3.5	2R	T/C/A	t	0.30			0.02	0.70
3	rs10756819	F&R	0.1 each F&R =0.4um	3.5	3F- new tail	A/G	ga	1.80			0.17	5.95
4	rs2238289	F&R	0.1 each F&R =0.4um	3.5	4F	C/T	ct	1.50			0.05	1.75
5	rs17128291	F&R	0.1 each F&R =0.4um	3.5	5R	C/T	ct	1.50			0.1	3.50
6	rs6497292	F&R	0.1 each F&R =0.4um	3.5	6R	C/T	ct	0.30		0.05		1.75
7	rs1129038	F&R	0.1 each F&R =0.4um	3.5	7F	G/A	ga	1.00			0.04	1.40
8	rs1667394	F&R	0.1 each F&R =0.4um	3.5	8R	C/T	ct	0.10			0.02	0.70
9	rs1126809	F&R	0.1 each F&R =0.4um	3.5	9F	C/T	ct	0.02	0.025			0.88
10	rs1470608	F&R	0.1 each F&R =0.4um	3.5	10R	C/A	ca	2.00			0.18	6.30
11	rs1426654	F&R	0.1 each F&R =0.4um	3.5	11F	G/A	ga	0.08		0.09		3.15
12	rs6119471	F&R	0.1 each F&R =0.4um	3.5	12R	G/C	gc	0.15			0.1	3.50
13	rs1545397	F&R	0.1 each F&R =0.4um	3.5	13F	A/T	at	1.00			0.08	2.80
14	rs6059655	F&R	0.1 each F&R =0.4um	3.5	14R	C/T	ct	1.50			0.4	14.00
15	rs12441727	F&R	0.1 each F&R =0.4um	3.5	15F	C/T	ga	0.10			0.07	2.45
16	rs3212355	F&R	0.1 each F&R =0.4um	3.5	16R	G/A	ga	1.20			0.15	5.25
17	rs8051733	F&R	0.1 each F&R =0.4um	3.5	17R	C/T	ct	1.60			0.16	5.60
		Total primers (for both F&R)	119	59.5				total			1.78	62.125
				no.samples					no. of samples			
Stocks		PCR 1st set up (ul)	Final conc	35		SBE reaction		(ul)	35			
primers		3.4		119		primers		1.775	62.125			
10x PCR gold buffer (no mg)		1	1X	35		snapshot rxn mix		1	35			
Mgcl2 (25mM)		1	2.5mM	35		H20		0.225	7.875			
dNtps (10mM each,combo 40mM)		0.22	220 uM	7.7		1st PCR product		2				
taq gold (5U/ul)		0.3	1.75U	10.5				5ul/rxn	105 total			
H2O		3.08		107.8		Thermo conditions: SBE rxn						
Total reagents		9		315		96 °C for 2 min and 25 cycles of 96 °C for 10 s, 50 °C for 5 s and 60 °C for 30 s						
DNA		1										
		10ul total				Clean up after SBE						
Thermo conditions: 1st PCR						Sap(1U/ul)	1ul used /sample		(1)37°C - 45min, (2) 75°C - 15min			
(1) 95 °C for 10 min, (2) 33 cycles of 95 °C for 30 s and 61 °C for 30 s, (3) 5 min at 61 °C.												
						CE run						
Clean up after 1st PCR						POP-7 on a 50 cm capillary length array.						
ExoProStar	2ul used per 5ul product of each sample as per des		(1)37°C - 15min, (2) 80°C - 15min			Run parameters: injection voltage of 2.5 kV for 10 s, and run time of 500 s at 60 °C.						
new made up XOSAP into eosap-Ix no. of samples			35									
1 EXOI			35									
1 SAP			35									
Trying this conc July 27th												
2 total		total for all samples	70									

APPENDIX D

Table 5 Hair Morphology SNP Information

Gene	SNP	Chromosome	Location HG-19	Forward Miseq Primer	Reverse Miseq Primer	Product Size (with indexes and adapters)	Reference
P2RY5	198607	13	48979167	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGAAATGCAAGGACTCAGA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGATTTCCCTTTTGTGGATTCT	218	
PTK6	310642	20	62161998	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGGGCCAGTGAACAAGG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTGTGGAGCTGGGGACTC	281	
FRAS1	436034	4	79256036	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGcctggaactcctgggctct	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGttcaggacaagcctgataacaa	250	
CRCT1; near LCE3E	499697	1	152493154	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAATCTGGGCAGAAGGGAAC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCAATGGTTTAAAGATGCCAAA	225	Eriksson *et al.*
LIPH	505569	3	185236115	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTGTTTGTTTGTTTGCTTTTTGA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCTGGGAGACGGAGGATAC	211	
LIPH	551936	3	185263467	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCCCTTGACGTGAGGAGTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACCTAGGTCTGACTCCAAAACTC	302	
KRT71	585583	12	52929370	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTAAAACAAAGGGAAGAAATGTT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCCACAAAGGTGATTTCAAG	312	
KRT74	632205	12	52959307	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAAAGCCTGTGTCCAAACT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCAACTCAGATGAGGCTGCT	328	
KRT74	670741	12	52965173	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTATCTCCCTGGAAGAGTCG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTCAAGGTGGAGCTTCAGG	265	
KRT74	673449	12	52965761	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCCTCCTGGGCTGAAGT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGGCCAGCAGGAAGAGA	314	
PRKAG3	692243	2	219695487	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTTATGGAGGGACCTGAGGTAGA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCAGCTGTGACCAGCA	316	
VDR	731236	12	48238757	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACGTCTGCAGTGTGTTGGAC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTGGTGGGGATTGAGCAGTG	249	
IGFBP5	741384	2	217551954	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGGAAGGTCTGCCAGAAAG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGTTAGGCAGAGGAAGGAC	252	
HDAC9	756853	7	18890000	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCCAACGATCTGAATGTGG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCCTCCCCTACGCTTTCAT	257	Marcinska *et al.*
EBF1	929626	5	158310631	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACCAGGAACTATGCTGAGGA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGCCAAAGCCTTAACTTCC	217	Marcinska *et al.*
TCHH	1131471	1	152079989	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGACACTGGCAAACTGATGA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCAGCAGAAGGAGGAACAGA	274	Buket *et al.*
FRAS1	1268789	4	79280693	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCTTGATTTAATAAAAATCAGTCAAC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGCTCCATCATATGCAATCC	204	Pospiech *et al.*
near OFCC1	1556547	6	10270377	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCATGTACTGCCTATTAACC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGATTGAAGACTGCTTGCTCTAT	224	Eriksson *et al.*
IGFBP5	1978346	2	217561467	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGAAGGTGGTGGTTTTAGCTT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAGGTGTAACAGCCCACACC	229	
	1998076	20	21880045	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACCAGGCAAATCTCAGG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCCTGTGGAAATGGTAGC	297	Marcinska *et al.*
	1999874	10	8353101	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTTAAACCTTGCCAGTGGA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCACCAAATTCCAAAATTAGAGA	222	
FLG-AS1	2050631	1	152312597	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGttgaatggatacagagtttcagc	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTCTTTCAACATTGATTCCACA	308	
	2146114	1	152390621	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATGTTTGGCAAAGCCTCTAT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGGGAACCCAAAGAAAGC	238	
LGR4	2219783	11	27411298	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTGATGAATTCAGGTCCAAG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAAATGAATCAATGGGAATGAGTT	239	
P2RY5	2227311	13	48987032	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTCCAAAGTTAATGTTTCTTTATGC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGGACAAATTGGGAATGTT	223	
VDR	2228570	12	48272895	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCAGCCTTCACAGGTCATA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCCTGGCACTGACTCTG	292	
VDR	2238136	12	48277713	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGAGACCTCTGTCTTCCAG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGCCAGCCCAGCTTAG	300	
	2489250	10	8274867	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGGCTGAGGCAGAAAACTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTTTGTTGTTGTTCTTTTTGTT	257	
TRAF2	2784081	9	139791891	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCTCTGAAGGCCATGTGTA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTTCCCTCCAGCCCCTA	294	
RPTN	3001978	1	152126467	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCCACAGGAGCCAAAGC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCTCCCTAAAGTTCAGAAACAGAAA	233	
NBPF18P; LOC101927949	3007671	1	151999347	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGACTTCCAGTAGGGTGGA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAGCCACTGTCAAACCACA	230	
	3007681	1	152019357	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGcgcccagcttatttctttg	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGagagtggagacccccactcaa	272	
	3827760	2	109513601	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGAGAAGACTAGCCGAATGC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTAAGAAGGTAGAGGCTGAGCAC	255	
IGFBP5	4442975	2	217920769	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATGTGTTTATATTAAGTGCAGGTCA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTGCTCCTTGTTCATTTGC	256	
IGFBP5	4480966	2	217560985	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGTGACCAGTTGCCTTCAG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTGCAACTCAGCTCTCTAA	256	
VDR	4516035	12	48398826	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTCACAAATCATCCAGCA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCAATTTCCCACCCTTTCA	216	
LOC39148	4672907	2	219821169	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGAGTTGAAAGAGGCTTAGG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCTCTCTGGGACTCGGTTT	258	
IGFBP5	4674107	2	217536354	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTACAAGGCAGTGGGAGGAG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTCAGCACCTGTTCCCCTA	292	
SUCNR1/ MBNL1	4679955	3	151653368	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTGGCTTAGCAGAGGGTTTT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCAAGTATTTATTATGCAAGAGTTTCA	214	Marcinska *et al.*
near TCHH	4845418	1	152136230	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTGCACAAATAAGTTCCCATTTA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCACTGACTCCCACAGATGTT	294	Medland *et al.*, Pospiech *et al.*
CRNN; LCE5A	4845779	1	152479176	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtcctggcttctgtcacctt	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGctggattatgcaggtggactc	258	
up of AR	5919324	X	66430640	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATAATTTCCAAATAACATGAGTTCAA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATGTGTTTCATGAAGGATGAGA	270	Marcinska *et al.*
PEX14	6658216	1	10561604	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGCCCGGCTTTTTGTATTTT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAATAAACTAATAGGGGCAGGG	261	
WNT10A	7349332	2	219756388	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACATCCCAGCCCATACCAAC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACGTGGGAACCAGACCAAT	208	Medland *et al.*, Eriksson *et al.*, Pospiech *et al.*
P2RY5	9535032	13	49012265	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGgcttggcactgGAATTTTATTT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGagtgctagaagggcacctg	297	
P2RY5	9568036	13	48971936	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTAGAGTCCTGATTCATTGACAAA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCCTTTTGTGTTTGCTTCC	213	
	9989836	2	70342727	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGCCAGTTGAAAACTAAGATTC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAAGCAAACCAAGCCTAACAA	244	
KRT71	10783518	12	52938497	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGATGCAGTTGCTGCTGTTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCTCTGCCTGAGTGGGTGT	241	
LOC10537 3470	10788819	1	151895093	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAAGACAGGAATTCCAATAGAGAAA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAAATTTTGGGGGAGACTGA	214	
	10788826	1	152161735	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGGCCTAGGTCTTTCGTATC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGAAGCAGCTTTTCCTACTT	257	
	11150606	16	31099011	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTCCTTGCCTCCTCACACTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGTGAGGTCAGGACAGGTA	256	
	11170678	12	54154174	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGTGAGAATTGGGAGATGA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTCCAGATAAAGAACCTCCA	228	
PADI3	11203346	1	17600822	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAACTGCAGGAAGGCTGTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGCCTTATCACGGCAAGAC	261	Buket *et al.*
LOC10013 2111	11204897	1	151809066	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtgggaggaataaatgGCTTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAGCCCTAGGCCAGAGAACT	252	
	11204925	1	152073120	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCTCGGTCCCCTGACCT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGTTGGAGGTGCCTTTGAG	296	
VDR	11568820	12	48302545	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAAGCAAACCAAGGGGTCTT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATTGTAGAACATCTTTTGTATCAGGA	253	
IGFBP5	11575161	2	217549963	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGACTGAGGAGCTGAATTT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAAGGAGGTTGGATCACAAG	430	
IGFBP5	11575194	2	217543728	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGGTCCTTCTTCACTGCTTC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGAGACTCCCGTGAGCA	238	
RPTN	11582331	1	152134136	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCTTTCTCTTGGCGTGT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTAACTGATTGGCTGATCAACAT	303	
PADI3	11585118	1	17605774	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGACACAGAGGCCAGAG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGttGGGAATAAAGGTGGGTGA	239	Buket *et al.*
TCHH	11803731	1	152083325	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCTCCTCCGGGAAAC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTTCACATGGCAGTGG	254	Medland *et al.*, Buket *et al.*, Pospiech *et al.*
	12116760	1	152437883	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTCAGTGTTAGCGAGCTGTT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGGAAGAAAAGAGGGAATCA	230	
	12123907	1	152467751	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGGGAGAGGTGGAGCATC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTCCAGGATTTACTGCCATTT	208	
near TCHH	12130862	1	152027015	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTTCTGAGGGAAGTTTGTTCT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGAAGAGGCAAAATGGTCA	242	Medland *et al.*, Buket *et al.*, Pospiech *et al.*
TARDBP	12565727	1	11033082	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTGTTTACCTGGGGCATTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTAGTCTTGAGCCCCTCCA	224	Li *et al.*, Marcinska *et al.*
HERC2	12913832	15	28365618	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTCTGTGTCTGATCCAAGAGG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACAATTAATTAAAAACAAAGAGAAGC	227	
	12997742	2	70786598	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCAGGCAAAAATAAACCACA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTAGGATTACAGTTGCCCACAA	301	
IGFBP5	13015993	2	217625523	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGACACCAAGGAGGGAGGAT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGACATCCGTCTTTGTCAGG	225	
TCHHL1	17646946	1	152062767	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCTCAGCATCCTCCCAAAG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTCCAGAGCAGGAAGATATAA	282	Medland *et al.*, Eriksson *et al.*, Pospiech *et al.*
KRT71	17662023	12	52934349	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGGGATGATGAAGGAGCTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTGCACTTCAAACCCAGAG	218	
KRT71	17730088	12	52946336	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGTAGAAATTGACCTGCCTCTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTTCTTGGAACATACGATTTTGT	279	
LIPH	55854644	3	185226492	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCTATTTCTGACTTGCCCTA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTGAAAGCTTCACCTCACCT	273	
	55883933	1	151889447	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTGATGTGGGAGGATGACT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTCTAGCTCTGTCACCCAGA	209	
FLG-AS1	61816764	1	152308971	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCACTTCCCTGGTGATGAA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTGTGTTTATGACCCCCAAA	243	
IGFBP5	67587000	2	217561648-50	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACATTGTGCTTCCTCCCTTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTCTTGAAAGCAGTCCTCCA	237	
TCHH	72696935	1	152085951	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAAAGAATAACACAAGTTCGCTTT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGGCTGCTTCAGACAGAAA	223	Buket *et al.*
RPTN	72696940	1	152128121	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTCTGGCTGACCATAGTGGAA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTACAGTCAGCCAGACAAACAA	270	
KRT74	73107581	12	52967117	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGCAGGAAAACTGCTCAGG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGACCCTGAGATCCAGAAG	253	
WNT10A	74333950	2	219746292	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGTGGGGTCAGTGAGAGAC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATGGGCACACATACAAACTACTGTT	240	
HRNR	74868796	1	152191051	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAAGACCACCCTGAGCCAGA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGAGCGAACAACATGGATC	307	
	75203436	1	151842358	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCAGCCAATGAAGTTTTGT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGCTTGCAAATTAGAAAGGAA	264	
RPTN	75957773	1	152127455	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTGTCCCACATGGACCTT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTAGTGAGCAGGGCCACAG	246	
LINC0149	77157375	2	219779911	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGAGAAGCAGGACAAAACA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTCAGGGCAGGTTCCAAGT	222	
RPTN	78544048	1	152129087	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAAAGCTGAAATCCTTGTCTTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGCAGAACTCCCACCACA	305	
	80293268	1	8207579	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAAGCAAGCAGCTCACCT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAAGTGAGTGGGCAGCATC	229	
	114410520	1	151972609	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTGTGCTTGCTGTGCTAAC	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTGCAGCTCACCTGTAGGA	247	
LINC0152 7	115813648	1	152921586	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCCACACTTCACACTTGCAG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGATGCATGTCAGGAAAGATGA	258	
	140371183	1	152098428	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTGTTGCCTGCATGGAGAT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGCATAGGTTTGTTCTAACCACA	242	
KRTAP2-3	143290289	17	39216977	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAAAGCTGCTAGACTGATTTTATT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGAAAGCTATTTTTCTTTCCTTTCC	253	
LIPH	144288709	3	185236964	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAATCAGAGAGGAAACTGCTG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCAGGATTATAGGAATGGCAAG	233	
	151069963	1	152004241	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGGAATCAGAGGTTTTCTGGA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCAGGGTTAAAGATGACAGATT	237	

APPENDIX E

Table 6 Group 1 Hair Structure Miseq Assay Protocol

No.	SNP	Primer set	PCR Ps d STOCK(50um)	205
1	1268789	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
2	7349332	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
3	55883933	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
4	505569	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
5	10788819	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
6	198607	forward and reverse	0.07 Forward & 0.07 Reverse	14.35
7	1999874	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
8	12565727	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
9	499697	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
10	13015993	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
11	11170678	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
12	1978346	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
13	12116760	forward and reverse	0.07 Forward & 0.07 Reverse	14.35
14	144288709	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
15	67587000	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
16	11575194	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
17	114410520	forward and reverse	0.07 Forward & 0.07 Reverse	14.35
18	3827760	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
19	4442975	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
20	11150606	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
21	756853	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
22	2489250	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
23	10788826	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
24	5919324	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
25	3007681	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
26	55854644	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
27	17730088	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
28	9535032	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
29	551936	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
30	11582331	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
31	74868796	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
32	673449	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
			1086.5	543.25

			no.samples
Stocks	PCR 1st set up (ul)	Final conc	205
primers	5.3		1086.50
10x PCR gold buffer (no mg)	1	1X	205
Mgcl2 (25mM)	1	2.5mM	205
dNtps (10mM each,combo 40mM	0.22	220 uM	45.1
taq gold (5U/ul)	0.4	1.75U	82
H2O	1.08		221.4
Total reagents	9		1845
DNA	1		
	10ul total		

Thermo conditions: 1st PCR

(1) 95 °C for 10 min, (2) 28 cycles of 95 °C for 30 s and touchdown at 64 °C decreasing by 0.3 for 40 s

APPENDIX F

Table 7 Group 2 Hair Structure Miseq Assay Protocol

No.	SNP	Primer set	PCR Ps d STOCK(50um)	205
1	12123907	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
2	9568036	forward and reverse	0.07 Forward & 0.07 Reverse	14.35
3	4679955	forward and reverse	0.07 Forward & 0.07 Reverse	14.35
4	929626	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
5	77157375	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
6	72696935	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
7	1556547	forward and reverse	0.07 Forward & 0.07 Reverse	14.35
8	80293268	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
9	151069963	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
10	2219783	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
11	10783518	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
12	140371183	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
13	61816764	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
14	9989836	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
15	731236	forward and reverse	0.1 Forward & 0.1 Reverse	20.5
16	436034	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
17	143290289	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
18	4480966	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
19	4672907	forward and reverse	0.07 Forward & 0.07 Reverse	14.35
20	115813648	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
21	6658216	forward and reverse	0.2 Forward & 0.2 Reverse	41
22	11203346	forward and reverse	0.1 Forward & 0.1 Reverse	20.5
23	1131471	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
24	17646946	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
25	4674107	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
26	12997742	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
27	2050631	forward and reverse	0.1 Forward & 0.1 Reverse	20.5
28	632205	forward and reverse	0.08 Forward & 0.08 Reverse	16.4
			984	492.00
				no.samples

Stocks	PCR 1st set up (ul)	Final conc	205
primers	4.8		984.00
10x PCR gold buffer (no m	1	1X	205
Mgcl2 (25mM)	1	2.5mM	205
dNtps (10mM each,combo	0.22	220 uM	45.1
taq gold (5U/ul)	0.4	1.75U	82
H2O	1.58		323.9
Total reagents	9		1845
DNA	1		
	10ul total		

Thermo conditions: 1st PCR

(1) 95 °C for 10 min, (2) 33 cycles of 95 °C for 30 s and 61 °C for 30 s, (3) 5 min at 61 °C.

APPENDIX G

Table 8 Group 3 Hair Structure Miseq Assay Protocol

No.	SNP	Primer set	PCR Ps d STOCK(50um)	205
1	1268789	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
2	7349332	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
3	55883933	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
4	505569	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
5	10788819	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
6	198607	forward and reverse	0.07 Forward & 0.07 Reverse	14.35
7	1999874	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
8	12565727	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
9	499697	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
10	13015993	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
11	11170678	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
12	1978346	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
13	12116760	forward and reverse	0.07 Forward & 0.07 Reverse	14.35
14	144288709	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
15	67587000	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
16	11575194	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
17	114410520	forward and reverse	0.07 Forward & 0.07 Reverse	14.35
18	3827760	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
19	4442975	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
20	11150606	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
21	756853	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
22	2489250	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
23	10788826	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
24	5919324	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
25	3007681	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
26	55854644	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
27	17730088	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
28	9535032	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
29	551936	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
30	11582331	forward and reverse	0.1 Forward & 0.1 Reverse	20.50
31	74868796	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
32	673449	forward and reverse	0.08 Forward & 0.08 Reverse	16.40
			1086.5	543.25

			no.samples
Stocks	PCR 1st set up (ul)	Final conc	205
primers	5.3		1086.50
10x PCR gold buffer (no mg)	1	1X	205
Mgcl2 (25mM)	1	2.5mM	205
dNtps (10mM each,combo 40mM	0.22	220 uM	45.1
taq gold (5U/ul)	0.4	1.75U	82
H2O	1.08		221.4
Total reagents	9		1845
DNA	1		
	10ul total		
Thermo conditions: 1st PCR			
(1) 95 °C for 10 min, (2) 28 cycles of 95 °C for 30 s and touchdown at 64 °C decreasing by 0.3 for 4(			

APPENDIX H

Table 9 Miseq Dilution Calculator [50]

MISEQ LIBRARY DILUTION CALCULATOR

Steve Doyle s.doyle@latrobe.edu.au

May 2014

This calculator has been made to help (i) determine the amount of DNA needed, (ii) determine the concentration in your tube, and (iii) normalise the DNA in preparation for sequencing
You should only need to adjust the values in the ORANGE boxes - leave the rest alone if you want it to work properly
The values in GREEN are the results you want
Feedback is welcome!

STEP 1 - How much DNA do you need?

Input 1	What DNA concentration do you need?	4000	pM
Input 2	What is the mean DNA fragment size?	200	bp
Result	You need at least	2.64	in 5 ul
	which is a concentration of	0.528	ng/ul

NOTE: for the MiSeq, you should try and aim for at least 4 nM (=4000 pM), but 2 nM will be fine. Check the MiSeq Preparation of DNA for Sequencing protocol for more information
NOTE: this is the average DNA fragment size of your size selection - if you size selected 400-600 bp, then this valie is 500

STEP 2 - how much DNA do you actually have?

Input 1	The total volume of my library is	5	ul
	If (A) the Qubit tells me I have	1.38	ng/ul
	OR		
	(B) I have in total		ng
Result	Then I have a concentration of (A)	14242.424	pM
	(B)		pM

NOTE: Use (A) or (B) depending on what you know about your sample

STEP 3 - Adjusting the DNA concentration

In STEP 1, you decided you were aiming for	4000	pM
but in STEP 2, you determined your concentration was	14242.424	pM
1. Take 5 ul and transfer to a new tube		
2. add	12.38	ul H2O
3. mix by tube swirl and pulse centrifuge		
4. transfer 5 ul to a new tube, ready for sequencing		

NOTE: this final 5 ul, which is now at the desired concentration, is all you need to be ready to sequence.

APPENDIX I

Coding Used for Conversion and Alignment of Miseq Fastq File

SAMtools command to convert fai file to fasta file

faidx /Users/username/file_location/folder_name/test.fasta

Alignment of sample files to create .sam file

bwa mem -t 8 test.fasta /
Users/username/file_location/folder_name/subfolder_1/subfolder_2/subfolder_3/sample1
_L001_R1_001.fastq/Users/username/file_location/folder_name/subfolder_1/subfolder_2
/subfolder_3/sample1 _L001_R2_001.fastq >/Users/username/file_location/
folder_name/subfolder_1/subfolder_2/test.sam

Use of Picard tools to convert .sam to .bam file

picard AddOrReplaceReadGroups
I=/Users/username/file_location/folder_name/subfolder_1/subfolder_2/subfolder_3/test.s
am O=
Users/username/file_location/folder_name/subfolder_1/subfolder_2/subfolder_3/test.bam
RGID=4 RGLB=lib1 RGPL=illumina RGPU=unit1 RGSM=20

Sort the .bam File

SAMtools sort /
Users/username/file_location/folder_name/subfolder_1/subfolder_2/test.bam > /
Users/username/file_location/folder_name/subfolder_1/subfolder_2 /test-sorted.bam

Create Searchable Reference File

picard CreateSequenceDictionary Reference=/
Users/username/file_location/folder_name/test.fasta OUTPUT=/
Users/username/file_location/folder_name/ test.dict

SAMtools used to create bcf and vcf files

SAMtools mpileup -f / Users/username/file_location/folder_name/test.fasta -g /
Users/username/file_location/folder_name/subfolder_1/subfolder_2/subfolder_3/test-
sorted.bam | bcftools call -m - > /
Users/username/file_location/folder_name/subfolder_1/subfolder_2/subfolder_3/calls.vcf

Use of GATK for Compilation and Visualization

java -jar GenomeAnalysisTK.jar -T VariantsToTable -R /
Users/username/file_location/folder_name /ref.fasta -V /
Users/username/file_location/folder_name/ testcalls.vcf -F CHROM -F POS -F ID -F
REF -F ALT -F QUAL -F FILTER -F INFO --allowMissingData -o results.table

APPENDIX J

Table 10 Plates for Miseq Validation Assay

Plate 1												
	1	2	3	4	5	6	7	8	9	10	11	12
A	1905 500pg	648 500pg	1854 500pg	1539 500pg	1497 500pg	1947 500pg	1938:1092 1ng:1ng	1497:1905 1ng:500pg	1155:1004 100pg:500pg	1938:290 1ng:100pg	1188:1155:1938 1ng:1ng:500pg	1938:290 1ng:1ng
B	1757 500pg	1080 500pg	1547 500pg	1201 500pg	1550 500pg	2156 500pg	1497:1905 1ng:1ng	1155:1004 1ng:500pg	1938:290 100pg:500pg	1550:429 5ng:500pg	1092:1550:1004 500pg:1ng:500pg	1550:429 1ng:500pg
C	1775 500pg	401 500pg	1938 500pg	1092 500pg	346 500pg	1004 500pg	1155:1004 1ng:1ng	1938:290 1ng:500pg	1550:429 1ng:100pg	1188:429 5ng:500pg	1550:429 1ng:1ng	1188:429 1ng:500pg
D	1539 500pg	1497 500pg	1947 500pg	1188 500pg	625 500pg	1992 500pg	1938:290 1ng:1ng	1550:429 100pg:500pg	1188:429 1ng:100pg	1992:1201 5ng:500pg	1188:429 1ng:1ng	1992:1201 1ng:500pg
E	1201 500pg	1550 500pg	2156 500pg	429 500pg	1573 500pg	290 500pg	1550:429 1ng:500pg	1188:429 100pg:500pg	1992:1201 1ng:100pg	1938:1092 5ng:500pg	1992:1201 1ng:1ng	1938:1092 1ng:500pg
F	1092 500pg	346 500pg	1905 500pg	648 500pg	1854 500pg	1550:429 1ng:1ng	1188:429 1ng:500pg	1992:1201 100pg:500pg	1938:1092 1ng:100pg	1497:1905 5ng:500pg	1938:1092 1ng:1ng	1497:1905 1ng:500pg
G	1188 500pg	625 500pg	1757 500pg	1080 500pg	1547 500pg	1188:429 1ng:1ng	1992:1201 1ng:500pg	1938:1092 100pg:500pg	1497:1905 1ng:100pg	1155:1004 5ng:500pg	1497:1905 1ng:1ng	1155:1004 1ng:500pg
H	429 500pg	1573 500pg	1775 500pg	401 500pg	1938 500pg	1992:1201 1ng:1ng	1938:1092 1ng:500pg	1497:1905 100pg:500pg	1155:1004 1ng:100pg	1938:290 5ng:500pg	1155:1004 1ng:1ng	1938:290 1ng:500pg

Table 10 Continued

	1	2	3	4	5	6	7	8	9	10	11	12
A	1550:429 100pg:500pg	1188:429 1ng:100pg	1992:1201 5ng:500pg	1155-5 min 500pg	1550-20 min 500pg	9947A 50pg	9947A 50pg	007 50pg	007 50pg	Wet saliva Charanya 1	Cat 1ng	Semen/female freezer/Walsh 2
B	1188:429 100pg:500pg	1992:1201 1ng:100pg	1938:1092 5ng:500pg	1155-10 min 500pg	1550-30 min 500pg	9947A 25pg	9947A 25pg	007 25pg	007 25pg	Dried&Degraded saliva Charanya	Primate 1ng	Wet saliva Charanya 2
C	1992:1201 100pg:500pg	1938:1092 1ng:100pg	1497:1905 5ng:500pg	1155-20 min 500pg	1550-40 min 500pg	9947A 10pg	9947A 10pg	007 10pg	007 10pg	Touch DNA fresh Wesli 1	Pig 1ng	Dried&degraded saliva Charanya 2
D	1938:1092 100pg:500pg	1497:1905 1ng:100pg	1155:1004 5ng:500pg	1155-30 min 500pg	9947A 5ng	9947A 5ng	007 5ng	007 5ng	Stephanie dried blood	Hair - DNA Ryan 1	Mouse 1ng	Touch DNA fresh Wesli 2
E	1497:1905 100pg:500pg	1155:1004 1ng:100pg	1938:290 5ng:500pg	1155-40 min 500pg	9947A 1ng	9947A 1ng	007 1ng	007 1ng	Stephanie wet blood	Touch DNA degraded Wesli 1	Stephanie dried blood 2	Hair - DNA Ryan 2
F	1155:1004 100pg:500pg	1938:290 1ng:100pg	1188:1155:1938 1ng:1ng:500pg	1550-0 min 500pg	9947A 500pg	9947A 500pg	007 500pg	007 500pg	...Noah's DNA with Charanya's heme cont (conc-1ul)	Vaginal Swab Krystal 1	Stephanie wet blood 2	Touch DNA degraded Wesli 2
G	1938:290 100pg:500pg	1550:429 5ng:500pg	1092:1550:1004 500pg:1ng:500pg	1550-5 min 500pg	9947A 250pg	9947A 250pg	007 250pg	007 250pg	Wet semen frozen sample 1	Vaginal Swab/Semen - Krystal/freezer 1	Noah's DNA with Charanya's heme cont (diluted-0.1uL)	Vaginal Swab Krystal 2
H	1550:429 1ng:100pg	1188:429 5ng:500pg	1155-0 min 500pg	1550-10 min 500pg	9947A 100pg	9947A 100pg	007 100pg	007 100pg	Semen/female freezer/Walsh 1	Dog 1ng	Wet semen frozen sample 2	Vaginal Swab/Semen - Krystal/freezer 2

APPENDIX K

Table 11 Distribution of Samples during Optimization Run

Optimazation Run				
500pg Samples (Ran in Duplicate)	**Population**	**Eye Color**	**Hair Color**	**Skin Color**
Sample1905	Mexico	Dark Brown	Dark Brown	Type 4
Sample1757	Vietnam	Dark Brown	Black	Type 4
Sample1775	European	Blue/Blue Gray	Blond	Type 4
Sample1539	African- Eur	Dark Brown	Dark Brown	Type 5
Sample1201	African	Dark Brown	Dark Brown	Type 5
Sample1092	China	Dark Brown	Black	Type 3
Sample1188	European	Blue/Green Yellow	Light Red/Strawberry Blond	Type 1
Sample429	European	Blue/Green Yellow	Dark Brown	Type 3
Sample648	European	Blue/Green Yellow	Red Brown/Auburn	Type 2
Sample1080	Panama	Dark Brown	Dark Brown	Type 4
Sample401	European	Blue/Green Yellow	Light Red/Strawberry Blond	Type 2
Sample1497	European	Blue/Blue Gray	Blond	Type 2
Sample1550	European	Blue/Blue Gray	Light Brown	Type 2
Sample346	European	Hazel/Light Brown	Light Brown	Type 2
Sample625	Vietnam	Dark Brown	Black	Type 4
Sample1573	European	Blue/Blue Gray	Dark Brown	Type 2
Sample1854	European	Blue/Blue Gray	Blond	Type 2
Sample1547	European	Hazel/Light Brown	Red Brown/Auburn	Type 2
Sample1938	European	Hazel/Light Brown	Light Brown	Type 3
Sample1947	European	Hazel/Light Brown	Red Brown/Auburn	Type 2
Sample 2156	European	Blue/Green Yellow	Blond	Type 2
Sample1004	African	Dark Brown	Black	Type 5
Sample290	Iran	Dark Brown	Dark Brown	Type 3
Sample 1992	European	Hazel/Light Brown	Light Brown	Type 1
Mixture Samples				
1ng:1ng				
Sample1550:429				
Sample1188:429				
Sample1992:690				
Sample1938:1092				
Sample1497:1905				
Sample1155:1004				
Sample1938:290				
1ng:500pg				
Sample1550:429				
Sample1188:429				
Sample1992:690				
Sample1938:1092				
Sample1497:1905				
Sample1155:1004				
Sample1938:290				
100pg:500pg				
Sample1550:429				
Sample1188:429				
Sample1992:690				
Sample1938:1092				
Sample1497:1905				
Sample1155:1004				
Sample1938:290				
1ng :100pg				
Sample1550:429				
Sample1188:429				
Sample1992:690				
Sample1938:1092				
Sample1497:1905				
Sample1155:1004				
Sample1938:290				
5ng:500pg				
Sample1550:429				
Sample1188:429				
Sample1992:690				
Sample1938:1092				
Sample1497:1905				
Sample1155:1004				
Sample1938:290				
Three Person Mixture				
1ng:500pg:500pg				
Sample1188:1155:1938				
Sample1092:1550:1004				
Casework Samples				
Dried Blood				
Wet Blood				
Heme Contaminated Sample				
Seman-Saliva (male-female) Mixture				
Saliva				
Degraded Saliva				
Touch DNA				
Hair				
Degraded Touch DNA				
Vaginal Swab				
VaginalSwab with Semen				

APPENDIX L

Table 12 HIrisPlex-S MPS Primer Sequences

		Forward Adapter	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG								
		Reverse Adapter	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG								
Major Allele	Minor Allele	PCR Primer	Flanking Forward Primer	Flanking Reverse Primer	Product Size(bp)	Final Length (Adapters+primer)	Eye Color	Hair Color	Skin Color	Reference
A insertion		MC1Rset1	GCAGGGATCCCAGAGAAGAC	TCAGAGATGGACACCTCCAG	117	184		X	X	
G	T							X	X	
C	A	MC1Rset2	CTGGTGAGCTTGGTGGAGA	TCCAGCAGGAGGATGACG	158	225	X	X	X	
G	A						X	X	X	
G	A							X	X	
G	A							X	X	
C	T						X	X	X	
C	T	MC1Rset3	GTCCAGCCTCTGCTTCCTG	AGCGTGCTGAAGACGACAC	147	214	X	X	X	
C	A							X	X	
T	C								X	
G	C	MC1Rset4	CAAGAACTTCAACCTCTTTCTCG	CACCTCCTTGAGCGTCCTG	106	173		X	X	
C	A		CTTTCAAAAGGCTTCCACTCA	TCCTTTGATGTCCCCTTCGAT	128	195		X	X	
C	G	rs28777	TCCAAGTTGTCTAGACCAGA	CGAAAGAGGAGTCGAGGTTG	128	195	X	X	X	
T	C	rs16891982	ATGCCCAAAGGATAAGGAAT	GGAGCCAAGGGCATGTTACT	118	185	X	X		
C	A	rs12821256	Tgagaaatctacccccagga	GTGTCTCTTACCCCCTGTGGA	140	207		X		
C	T	rs4959270	AGGGGCAGCTGATCTCTCAG	GCTTCGTCATATGGCTAAACCT	126	193	X	X	X	
C	A	rs12203592	CAACACCCATGTTTAACGACA	GCTTCATGGGCAAAATCAAT	124	191	X	X	X	
G	A	rs1042602	AAGGCTGCCTCTGTTCTACG	CGATGAGACAGAGCATGATGA	124	191	X	X	X	
G	A	rs1800407	actgtctcacagtgctgct	TTCACCTCgatgagatgat	150	217		X	X	
A	G	rs2402130	TGTTCTTCATGGCTCTCTGTG	GGCCCCTGATGATGATAGC	96	163	X	X	X	
A	G	rs12913832	CGCATAACCATCCCTCTAA	CATTGCTTTTCAGCCCACAC	136	203		X	X	
G	T	rs2378249	CTGGCGATCCAATTCTTTGT	GACCCTGTGTGAGACCCAGT	125	192	X	X		
G	A	rs12896399	TTTCTTTATCCCCCTGATGC	GGGAAGGTGAATGATAACACG	124	191	X	X	X	
A	C	rs1393350	CACAAAACCACCTGGTTGAA	TCCCAGCTTTGAAAGTATGC	127	194	X	X	X	
G	A	rs683	CAGAACACAGCCACACCCTA	ACAGGAATGGCAGCTTTGAG	118	185				
A	G	rs3114908	GCTGCAGGAGTCAGAAGGTT	GGGACAAACGAATTGAGGAA	145	212				
G	A	rs1800414	GACCAGTTATTTTGGGTTTGGA	CGTCATGACTAGAAAAACACCAA	76	143				
T	C	rs10756819	GGAACATGAAGATTTCCCAGT	CTGATTCAAGTCTGCTGTCACT	112	179				
A	G	rs2238289	CCAGCACTGCCAAAATAACA	CTCTTTGGACCCATCACCTC	129	196				
G	A	rs17128291	TCTGCTGTAGAACCAATGTCC	GAATTGCACCTGTAGCTCCAT	150	217				
G	A	rs6497292	ATGTCGACTCCTTGCTTCG	ACACCAGGCAGCCTACAGTC	137	204	X			
G	A	rs1129038	CAGCTGTAGAGAGACTTTGAGG	CAGCAATTCAAAACGTGCAT	130	197	X	X	X	
G	A	rs1667394	TGTTCTTAGTCTGAATAACCTTTCC	GGTGCATTGGCTTCTGGATA	100	167				
A	C	rs11226809	TTTCTTGTGTTAACTGTCCTTACAAA	GGAAAATATGTTTAGGGTTGATGG	145	212				
G	A	rs1470608	TTCAGCCCTTGGATTGTCTC	TGAGTAAGCAAGAAGTATAAGGACA	123	190	X	X	X	
C	G	rs1426654	AAAAGAAGTAGCTGTACTAGACGGGAT	AACCCGAAGGAAGAGTGAAAA	63	130	X	X	X	
A	T	rs6119471	AAAGTGTTCTGGAATTGGATACTGACAA	AAATGGAGATATAGAATTCACACAACA	121	188	X			
G	A	rs1545397	GTGAGGAAATCGAGGCTCAG	AGGAGAAAGCTGCAGATCCA	112	179				
G	A	rs6059655	GGGAAGAGACAGCTCCATGT	ACAATCCTGGGAGGTACACG	137	204				
C	T	rs12441727	TTCCACCCTTCAGCACAGA	CATCAAAGGCAGACCTCTCG	144	211				
G	A	rs3212355	AGGCGGTGGTCTCTCTCTC	TTGCAACAGGAGGGTCTAGG	124	191				
		rs8051733								

APPENDIX M

Table 13 HIrisPlex-S MPS Protocol

No.	SNP	Primer set	PCR Ps d STOCK(50um)	1.000
1	MC1Rset1	forward and reverse	0.08 Forward & 0.08 Reverse	0.080
2	MC1Rset2	forward and reverse	0.11 Forward & 0.11 Reverse	0.110
3	MC1Rset3	forward and reverse	0.06 Forward & 0.06 Reverse	0.060
4	MC1Rset4	forward and reverse	0.03 Forward & 0.03 Reverse	0.030
5	rs28777	forward and reverse	0.08 Forward & 0.08 Reverse	0.080
6	rs16891982	forward and reverse	0.06 Forward & 0.06 Reverse	0.060
7	rs12821256	forward and reverse	0.10 Forward & 0.10 Reverse	0.100
8	rs4959270	forward and reverse	0.04 Forward & 0.04 Reverse	0.040
9	rs12203592	forward and reverse	0.04 Forward & 0.04 Reverse	0.040
10	rs1042602	forward and reverse	0.07 Forward & 0.07 Reverse	0.070
11	rs1800407	forward and reverse	0.035 Forward & 0.035 Reverse	0.035
12	rs2402130	forward and reverse	0.03 Forward & 0.03 Reverse	0.030
13	rs12913832	forward and reverse	0.02 Forward & 0.02 Reverse	0.020
14	rs2378249	forward and reverse	0.045 Forward & 0.045Reverse	0.045
15	rs12896399	forward and reverse	0.075 Forward & 0.075 Reverse	0.075
16	rs1393350	forward and reverse	0.06 Forward & 0.06 Reverse	0.060
17	rs683	forward and reverse	0.09 Forward & 0.09 Reverse	0.090
18	rs3114908	forward and reverse	0.02 Forward & 0.02 Reverse	0.020
19	rs1800414	forward and reverse	0.09 Forward & 0.09 Reverse	0.090
20	rs10756819	forward and reverse	0.02 Forward & 0.02 Reverse	0.020
21	rs2238289	forward and reverse	0.04 Forward & 0.04 Reverse	0.040
22	rs17128291	forward and reverse	0.045 Forward & 0.045 Reverse	0.045
23	rs6497292	forward and reverse	0.05 Forward & 0.05 Reverse	0.050
24	rs1129038	forward and reverse	0.04 Forward & 0.04 Reverse	0.040
25	rs1667394	forward and reverse	0.045 Forward & 0.045 Reverse	0.045
26	rs1126809	forward and reverse	0.04 Forward & 0.04 Reverse	0.040
27	rs1470608	forward and reverse	0.10 Forward & 0.10 Reverse	0.100
28	rs1426654	forward and reverse	0.055 Forward & 0.055 Reverse	0.055
29	rs6119471	forward and reverse	0.12 Forward & 0.12 Reverse	0.120
30	rs1545397	forward and reverse	0.10 Forward & 0.10 Reverse	0.100
31	rs6059655	forward and reverse	0.04 Forward & 0.04 Reverse	0.040
32	rs12441727	forward and reverse	0.05 Forward & 0.05 Reverse	0.050
33	rs3212355	forward and reverse	0.075 Forward & 0.075 Reverse	0.075
34	rs8051733	forward and reverse	0.05 Forward & 0.05 Reverse	0.050

Stocks	PCR 1st set up (ul)	Final conc	110.000
primers	3.93		432.300
10x PCR gold buffer (no mg)	1	1X	110.000
Mgcl2 (25mM)	1	2.5mM	110.000
dNtps (10mM each,combo 40mM)	0.22	220 uM	24.200
taq gold (5U/ul)	0.4	1.75U	44.000
H2O	2.45		269.500
Total reagents	9		990.000
DNA	1		
	10ul total		

Thermo conditions: 1st PCR

(1) 95 °C for 10 min, (2) 34 cycles of 95 °C for 30 s and touchdown at 62 °C decreasing by 0.3 for 40 followed by 60 °C for 10s (3) 5 min at 60 °C.

APPENDIX N

Table 14 Concordance Testing Results for HIrisPlex-S

Sample	Type	Simulated casework scenario	DNA quantity (pg)	Profile quality	Brown	Red	Black	Blond	Light	Dark	Blue	Int.	Brown	Very Pale	Pale	Int.	Dark	Dark Black	Analyst interpretation	Correct/Incorrect interpretation?
1	Dried blood	Dried on glass: single source	280	Full profile but peaks	0.153	0.000	0.843	0.004	0.004	0.996	0.000	0.000	0.998	0.000	0.001	0.040	0.941	0.018	Black hair, brown eyes dark skin color	Correct
2	Saliva	Single source	3070	good profile	0.626	0.003	0.221	0.150	0.396	0.604	0.455	0.378	0.167	0.214	0.638	0.128	0.000	0.000	Dark-brown-Black hair, blue to intermediate eyes, pale skin color	Correct
3	Frozen blood	Intentional heme inhibition (2%)	70	very low peaks but full profile	0.348	0.004	0.650	0.498	0.891	0.109	0.888	0.081	0.061	0.017	0.486	0.497	0.001	0.000	Blond hair blue eyes pale to intermediate skin colour	Correct
4	Saliva	Single source	10530	good profile	0.193	0.003	0.032	0.772	0.962	0.038	0.917	0.052	0.031	0.013	0.935	0.482	0.001	0.000	Blond hair blue eyes pale to intermediate skin colour	Correct
5	Frozen blood	Single source	3070	good profile	0.193	0.003	0.032	0.772	0.962	0.038	0.917	0.052	0.031	0.013	0.935	0.482	0.001	0.000	Blond hair blue eyes pale to intermediate skin colour	Correct
6	Semen & blood	Mixture semen & blood	8910	unbalanced peaks;possibly mixture																
7	Saliva	Dried on glass: single source	720	Full profile but low peaks	0.390	0.007	0.025	0.678	0.969	0.041	0.917	0.052	0.031	0.021	0.445	0.533	0.022	0.000	could not merged the possibly mixture sample	-
8	Touched item	Arm swab: single source	140	very low peaks	0.216	0.000	0.778	0.006	0.006	0.994	0.000	0.002	0.998	0.000	0.001	0.038	0.940	0.002	Blond hair, blue eyes pale skin colour	Correct
9	Vaginal secretions & semen	Mixture: vaginal swab & semen	14150	unbalanced peaks;possibly mixture															could not merged the possibly mixture sample	-
10	Vaginal swab	Single source	1840	good profile	0.404	0.003	0.593	0.400	0.602	0.348	0.143	0.155	0.702	0.054	0.621	0.325	0.000	0.000	Brown hair, brown eyes, pale to intermediate skin color	Correct
11	Saliva	Single source	21040	good profile	0.932	0.000	0.068	0.075	0.125	0.875	0.014	0.051	0.935	0.000	0.000	0.178	0.489	0.329	Dark-brown-Black hair, brown eyes, dark skin color	Correct
12	Saliva	Single source	6930	good profile	0.168	0.000	0.838	0.004	0.006	0.994	0.000	0.004	0.996	0.000	0.000	0.311	0.689	0.000	Black hair, brown eyes, dark-black skin color	Correct
13	Touched item	Trace DNA: single source	30	Ful profile, but very low peaks	0.097	0.000	0.900	0.000	0.003	0.997	0.000	0.002	0.998	0.000	0.000	0.748	0.252	0.000	Black hair, brown eyes intermediate skin color	Correct
14	Touched item	Trace DNA: single source	110	good profile	0.464	0.001	0.148	0.367	0.648	0.352	0.052	0.116	0.832	0.004	0.160	0.693	0.001	0.001	Dark-brown-Black hair, brown eyes intermediate skin color	Correct
15	Saliva & semen	Mixture: saliva & semen	11780	unbalanced peaks;possibly mixture	0.135	0.860	0.005	0.010	0.723	0.277	0.052	0.116	0.832	0.004	0.160	0.693	0.001	0.001	Dark-brown-Black hair, brown eyes intermediate skin color	-
16	Frozen blood	Single source	5980	good profile	0.193	0.000	0.843	0.004	0.004	0.996	0.000	0.002	0.998	0.000	0.001	0.040	0.941	0.018	could not merged the possibly mixture sample	Correct
17	Saliva & blood	Mixture: saliva & blood	5800	unbalanced peaks;possibly mixture															Black hair, brown eyes dark skin colour	-
18	Mucus & saliva	Mixture: mucus & saliva	10600	possible mixture															could not merged the possibly mixture sample	-
19	Semen	Single source	4320	good profile	0.348	0.004	0.650	0.698	0.931	0.069	0.888	0.081	0.061	0.017	0.486	0.497	0.001	0.000	could not merged the possibly mixture sample	Correct
20	Touched item	Trace DNA: single source	90	Ful profile, but very low peaks	0.466	0.000	0.419	0.114	0.215	0.785	0.001	0.018	0.981	0.001	0.054	0.572	0.378	0.015	Dark-brown-Black hair, brown eyes intermediate to dark skin color	Correct
21	Touched item	Trace DNA: multiple source	20	very low peaks but single source	0.477	0.003	0.130	0.391	0.746	0.254	0.052	0.116	0.832	0.007	0.408	0.580	0.004	0.000	Dark-brown-Black hair, brown eyes pale to intermediate skin color	Incorrect, as single source and hence the phenotype interpretation
22	Vaginal secretions	Single source	11790	good profile	0.784	0.033	0.070	0.113	0.627	0.373	0.339	0.335	0.606	0.004	0.769	0.147	0.000	0.000	Dark-brown hair, brown eyes, pale skin color	Correct
23	Mucus	Single source	4160	good profile	0.489	0.001	0.417	0.094	0.185	0.815	0.028	0.064	0.909	0.007	0.168	0.764	0.036	0.035	Dark-brown-Black hair, brown eyes intermediate skin color	Correct
24	Touched item	Trace DNA: single source	60	too many drop-outs, very low peaks & partial profile															could not merged due to many drop-outs and hardly any profile	-
25	Frozen blood	Intentional heme inhibition (10%)	30	too many drop-outs, very low peaks & partial profile															could not merged due to many drop-outs and hardly any profile	-

APPENDIX O

Coding Used for Correlation and Partial Correlation for Hair Structure Assay

Used for Correlation- No Control for Variables

```
dat <- hstestcorr[,c(1,2:95)]
for(i in 2:92){
  x <- dat[i]
  subs <- cbind.data.frame(x,dat[,c(93:95)])
  subs_rm <- na.omit(subs)
  test[[i]] <- cor.test(subs_rm[,1],subs_rm$Hair.Type, method = "pearson")
  head(test[[i]])
  corr <- cbind(test[[i]]$estimate, test[[i]]$p.value)
  write.table(corr, file="C:/Users/krcli/Desktop/test_corr_trial3.txt", sep=" ",
row.names=TRUE, col.names = TRUE, append = TRUE)
}
```

Used for Partial Correlation- Controling for Age and Sex

```
dat <- hstestcorr[,c(1,2:95)]
for(i in 2:92){
  x <- dat[i]
  subs <- cbind.data.frame(x,dat[,c(93:95)])
  subs_rm <- na.omit(subs)
  test[[i]] <- pcor.test(subs_rm[,1], subs_rm$Hair.Type, subs_rm[,c(2:3)], method =
"pearson")
  write.table(test[[i]], file="C:/Users/krcli/Desktop/test_pc_trial3.txt", sep=" ",
row.names=TRUE, col.names = TRUE, append = TRUE)
  }
```

APPENDIX P

Table 15 Pearson's Correlation with No Corrections

Number	Estimate	p.value	N, after removing NA	Overall Mega analysis	US pop-MEGA
rs10783518-C	-0.07196	0.018298	1077		
rs10788819-T	-0.05064	0.098121	1070		
rs10788826-T	-0.02115	0.492296	1058		
rs11150606-C	-0.13352	1.19E-05	1071		
rs11170678-G	-0.04052	0.185773	1070		
rs11203346-G	-0.00098	0.974461	1080		
rs11204897- G	-0.06719	0.039868	938		
rs11204925-G	0.012714	0.734852	714		
rs1131471-T	-0.07217	0.018065	1075		
rs114410520-C	-0.05841	0.057893	1057		
rs11568820-T	0.176335	6.41E-08	930		
rs11575161-T	0.010358	0.789914	666	SIGNIFICANT	
rs11575194-A	-0.04187	0.171484	1070		
rs115813648-C	-0.00154	0.959913	1064		
rs11582331-T	0.023025	0.459111	1038		
rs11585118-A	-0.01744	0.58218	1000		
rs11803731-T	-0.1619	3.03E-07	992		
rs12116760-T	-0.05478	0.074485	1063		
rs12123907-A	-0.02596	0.394852	1078		
rs12130862-T	-0.10065	0.003518	841		
rs12565727-G	0.094649	0.001828	1084		
rs1268789-T	0.037382	0.219639	1082		
rs12913832-G	-0.16526	1.72E-06	831		
rs12997742-C	-0.10389	0.00085	1030		
rs13015993-G	0.033714	0.268966	1079		
rs140371183-G	0.034028	0.26542	1075	SIGNIFICANT	
rs143290289-A	-0.09478	0.002099	1053		
rs144288709-A	0.027873	0.360358	1081		
rs151069963-C	-0.02824	0.357164	1067		
rs1556547-A	0.031951	0.29527	1077		
rs17646946-A	-0.13181	2.12E-05	1036		
rs17662023-A	-0.0423	0.166593	1073		
rs17730088-G	-0.01493	0.628358	1055		
rs1978346-A	0.002175	0.943356	1072		
rs198607-C	-0.02933	0.343145	1049		
rs1998076-A	-0.01781	0.58372	951		
rs1999874-A	0.056195	0.069259	1048		
rs2050631-G	-0.09201	0.003426	1012		
rs2146114-C	0.046945	0.12629	1064		
rs2219783-G	0.157161	2.54E-07	1067	SIGNIFICANT	SIGNIFICANT
rs2227311-G	-0.02394	0.48561	853		
rs2228570-A	-0.05372	0.094816	970		
rs2238136-T	0.002389	0.942632	912		
rs2489250-T	-0.05601	0.065789	1082	SIGNIFICANT	
rs2784081-C	-0.02849	0.352512	1069		
rs3001978-T	-0.0623	0.064266	885		
rs3007671-T	-0.01214	0.692493	1066		
rs3007681-C	-0.0357	0.249994	1042		
rs310642-C	0.18404	5.94E-09	987		
rs3827760-G	-0.15983	1.47E-07	1072		
rs436034-T	-0.04142	0.177773	1062		
rs4442975-G	0.10475	0.000632	1063		
rs4480966-C	-0.01233	0.687783	1066		
rs4516035-C	-0.09635	0.002382	994		
rs4672907-A	0.190473	3.70E-10	1067		
rs4674107-C	-0.10147	0.000939	1062		
rs4679955-T	-0.0387	0.211476	1046		
rs4845418-C	-0.13944	2.17E-05	923		
rs4845779-T	0.042728	0.170605	1032		
rs499697-G	-0.02964	0.333403	1069		
rs505569-G	0.00497	0.870827	1075		
rs551936-C	-0.00755	0.804531	1078		
rs55854644-T	-0.0435	0.158362	1055		
rs55883933-C	-0.06627	0.029121	1086		
rs585583-C	-0.06187	0.303133	281		
rs5919324-G	0.148322	1.96E-06	1022		
rs61816764-T	-0.0015	0.960888	1068		
rs632205-T	-0.07769	0.013291	1017	SIGNIFICANT	SIGNIFICANT
rs6658216-C	0.073238	0.017945	1046	SIGNIFICANT	
rs670741-T	-0.0968	0.001637	1058		
rs673449-C	-0.05045	0.11521	978		
rs67587000- no insertion	0.018299	0.552138	1060		
rs692243-C	-0.0723	0.024484	970		
rs72696935-G	-0.04852	0.112895	1071		
rs72696940-A	-0.00121	0.969134	1027	SIGNIFICANT	SIGNIFICANT
rs73107581-T	-0.0061	0.845641	1023		
rs731236-G	0.030183	0.323721	1073		
rs7349332-T	0.081653	0.007478	1074	SIGNIFICANT	
rs741384-C	0.020888	0.498558	1054		
rs74333950-G	0.056624	0.066767	1051		
rs74868796-A	-0.05581	0.074826	1022		
rs75203436-G	-0.06876	0.046591	840		
rs756853-A	-0.01481	0.629966	1062		
rs75957773-C	-0.02426	0.433657	1046		
rs77157375-A	0.007138	0.81543	1074	SIGNIFICANT	
rs78544048-C	2.61E-05	0.999346	991		
rs80293268-C	-0.11956	8.58E-05	1076		
rs929626-G	-0.01348	0.658456	1079		
rs9535032-G	0.171171	3.78E-08	1022		
rs9568036-A	-0.17393	1.04E-08	1071	SIGNIFICANT	SIGNIFICANT
rs9989836-A	0.082187	0.007285	1067		

APPENDIX Q

Table 16 Pearson's Correlation with Correction for Age and Sex

Number	Estimate	p.value	N, after removing NA	Overall Mega analysis	US pop-MEGA
rs10783518-C	-0.07196	0.018298	1077		
rs10788819-T	-0.05064	0.098121	1070		
rs10788826-T	-0.02115	0.492296	1058		
rs11150606-C	-0.13352	1.19E-05	1071		
rs11170678-G	-0.04052	0.185773	1070		
rs11203346-G	-0.00098	0.974461	1080		
rs11204897- G	-0.06719	0.039868	938		
rs11204925-G	0.012714	0.734852	714		
rs1131471-T	-0.07217	0.018065	1075		
rs114410520-C	-0.05841	0.057893	1057		
rs11568820-T	0.176335	6.41E-08	930		
rs11575161-T	0.010358	0.789914	666	SIGNIFICANT	
rs11575194-A	-0.04187	0.171484	1070		
rs115813648-C	-0.00154	0.959913	1064		
rs11582331-T	0.023025	0.459111	1038		
rs11585118-A	-0.01744	0.58218	1000		
rs11803731-T	-0.1619	3.03E-07	992		
rs12116760-T	-0.05478	0.074485	1063		
rs12123907-A	-0.02596	0.394852	1078		
rs12130862-T	-0.10065	0.003518	841		
rs12565727-G	0.094649	0.001828	1084		
rs1268789-T	0.037382	0.219639	1082		
rs12913832-G	-0.16526	1.72E-06	831		
rs12997742-C	-0.10389	0.00085	1030		
rs13015993-G	0.033714	0.268966	1079		
rs140371183-G	0.034028	0.26542	1075	SIGNIFICANT	
rs143290289-A	-0.09478	0.002099	1053		
rs144288709-A	0.027873	0.360358	1081		
rs151069963-C	-0.02824	0.357164	1067		
rs1556547-A	0.031951	0.29527	1077		
rs17646946-A	-0.13181	2.12E-05	1036		
rs17662023-A	-0.0423	0.166593	1073		
rs17730088-G	-0.01493	0.628358	1055		
rs1978346-A	0.002175	0.943356	1072		
rs198607-C	-0.02933	0.343145	1049		
rs1998076-A	-0.01781	0.58372	951		
rs1999874-A	0.056195	0.069259	1048		
rs2050631-G	-0.09201	0.003426	1012		
rs2146114-C	0.046945	0.12629	1064		
rs2219783-G	0.157161	2.54E-07	1067	SIGNIFICANT	SIGNIFICANT
rs2227311-G	-0.02394	0.48561	853		
rs2228570-A	-0.05372	0.094816	970		
rs2238136-T	0.002389	0.942632	912		
rs2489250-T	-0.05601	0.065789	1082	SIGNIFICANT	
rs2784081-C	-0.02849	0.352512	1069		
rs3001978-T	-0.0623	0.064266	885		
rs3007671-T	-0.01214	0.692493	1066		
rs3007681-C	-0.0357	0.249994	1042		
rs310642-C	0.18404	5.94E-09	987		
rs3827760-G	-0.15983	1.47E-07	1072		
rs436034-T	-0.04142	0.177773	1062		
rs4442975-G	0.10475	0.000632	1063		
rs4480966-C	-0.01233	0.687783	1066		
rs4516035-C	-0.09635	0.002382	994		
rs4672907-A	0.190473	3.70E-10	1067		
rs4674107-C	-0.10147	0.000939	1062		
rs4679955-T	-0.0387	0.211476	1046		
rs4845418-C	-0.13944	2.17E-05	923		
rs4845779-T	0.042728	0.170605	1032		
rs499697-G	-0.02964	0.333403	1069		
rs505569-G	0.00497	0.870827	1075		
rs551936-C	-0.00755	0.804531	1078		
rs55854644-T	-0.0435	0.158362	1055		
rs55883933-C	-0.06627	0.029121	1086		
rs585583-C	-0.06187	0.303133	281		
rs5919324-G	0.148322	1.96E-06	1022		
rs61816764-T	-0.0015	0.960888	1068		
rs632205-T	-0.07769	0.013291	1017	SIGNIFICANT	SIGNIFICANT
rs6658216-C	0.073238	0.017945	1046	SIGNIFICANT	
rs670741-T	-0.0968	0.001637	1058		
rs673449-C	-0.05045	0.11521	978		
rs67587000- no insertion	0.018299	0.552138	1060		
rs692243-C	-0.0723	0.024484	970		
rs72696935-G	-0.04852	0.112895	1071		
rs72696940-A	-0.00121	0.969134	1027	SIGNIFICANT	SIGNIFICANT
rs73107581-T	-0.0061	0.845641	1023		
rs731236-G	0.030183	0.323721	1073		
rs7349332-T	0.081653	0.007478	1074	SIGNIFICANT	
rs741384-C	0.020888	0.498558	1054		
rs74333950-G	0.056624	0.066767	1051		
rs74868796-A	-0.05581	0.074826	1022		
rs75203436-G	-0.06876	0.046591	840		
rs756853-A	-0.01481	0.629966	1062		
rs75957773-C	-0.02426	0.433657	1046		
rs77157375-A	0.007138	0.81543	1074	SIGNIFICANT	
rs78544048-C	2.61E-05	0.999346	991		
rs80293268-C	-0.11956	8.58E-05	1076		
rs929626-G	-0.01348	0.658456	1079		
rs9535032-G	0.171171	3.78E-08	1022		
rs9568036-A	-0.17393	1.04E-08	1071	SIGNIFICANT	SIGNIFICANT
rs9989836-A	0.082187	0.007285	1067		

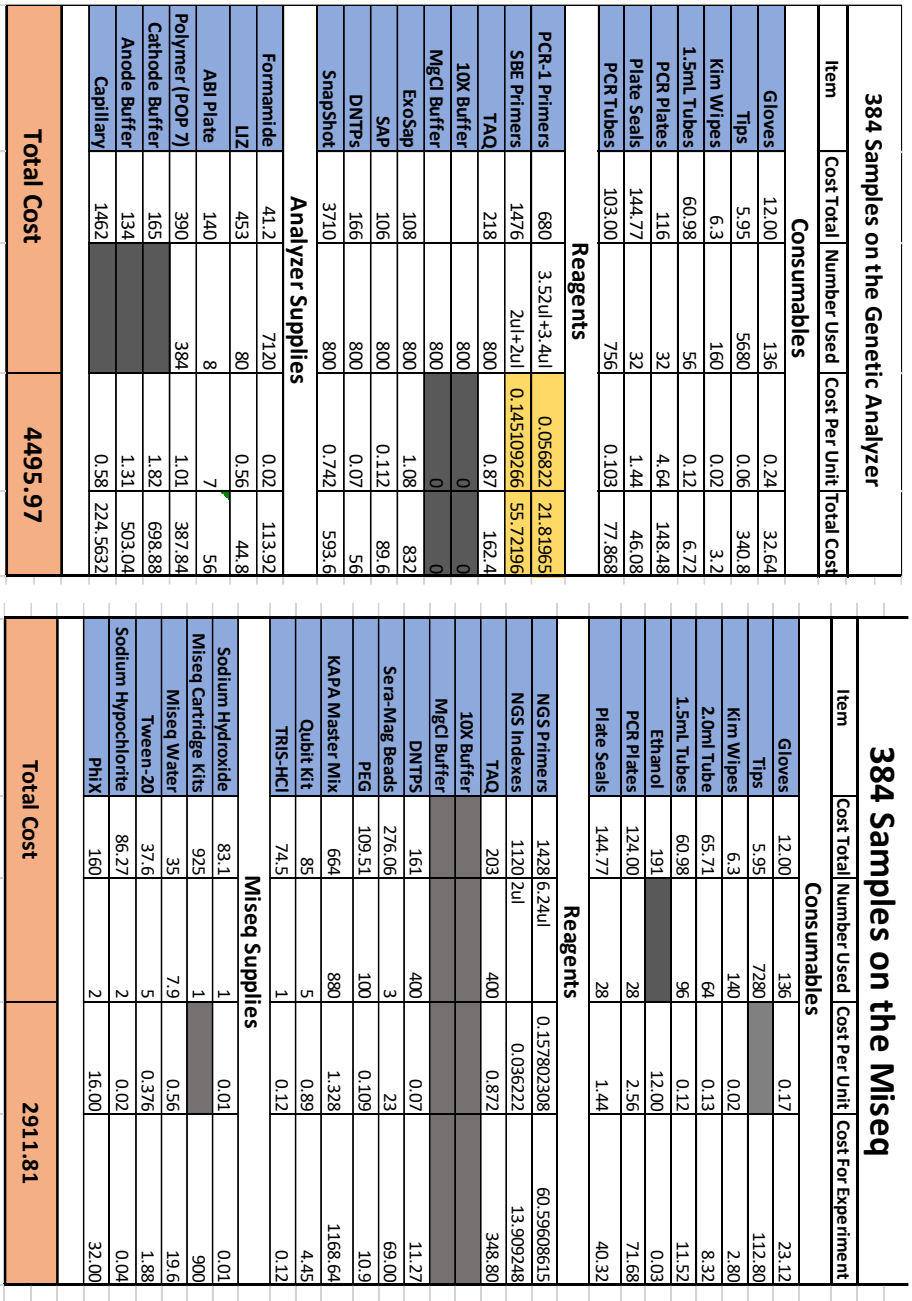
APPENDIX R

Table 17 Analysis of Optimization Run

HIrisPlex-S SNP	AVERAGE COVERAGE	STANDARD DEVIATION
rs312262906	548	458
rs11547464	506	416
rs885479	506	416
rs1805008	506	416
rs1805005	453	361
rs1805006	456	361
rs1805007	506	415
rs1805009	12	9
rs201326893	506	415
rs2228479	453	361
rs1110400	506	415
Rs28777	146	125
Rs16891982	571	525
Rs12821256	57	140
Rs4959270	214	191
Rs12203592	23	15
Rs1042602	481	379
Rs1800407	124	93
Rs2402130	40	32
rs12913832	33	22
rs2378249	268	234
rs12896399	308	245
rs1393350	582	426
rs683	26	20
rs3114908	400	316
rs1800414	137	101
rs10756819	1254	942
rs2238289	46	34
rs17128291	91	73
rs6497292	248	204
rs1129038	183	130
rs1667394	182	139
rs1126809	203	144
rs1470608	70	49
rs1426654	12	10
rs6119471	35667	24878
rs1545397	4078	3083
rs6059655	419	333
rs12441727	294	237
rs3212355	137	102
rs8051733	1190	841

APPENDIX S

Table 18 Cost Analysis for 384 Samples on the Capillary Electrophoresis versus the Miseq



384 Samples on the Genetic Analyzer				
Item	Cost Total	Number Used	Cost Per Unit	Total Cost
Consumables				
Gloves	12.00	136	0.24	32.64
Tips	5.95	5680	0.06	340.8
Kim Wipes	6.3	160	0.02	3.2
1.5mL Tubes	60.98	56	0.12	6.72
PCR Plates	116	32	4.64	148.48
Plate Seals	144.77	32	1.44	46.08
PCR Tubes	103.00	756	0.103	77.868
Reagents				
PCR-1 Primers	680	3.52ul+3.4ul	0.056822	21.81965
SBE Primers	1476	2ul+2ul	0.145109266	55.72196
TAQ	218	800	0.87	162.4
10X Buffer		800	0	0
MgCl Buffer		800	0	0
ExoSap	108	800	1.08	832
SAP	106	800	0.112	89.6
DNTPs	166	800	0.07	56
SnapShot	3710	800	0.742	593.6
Analyzer Supplies				
Formamide	41.2	7120	0.02	113.92
LIZ	453	80	0.56	44.8
ABI Plate	140	8	7	56
Polymer (POP 7)	390	384	1.01	387.84
Cathode Buffer	165		1.82	698.88
Anode Buffer	134		1.31	503.04
Capillary	1462		0.58	224.5632
Total Cost	**4495.97**			

384 Samples on the Miseq				
Item	Cost Total	Number Used	Cost Per Unit	Cost For Experiment
Consumables				
Gloves	12.00	136	0.17	23.12
Tips	5.95	7280		112.80
Kim Wipes	6.3	140	0.02	2.80
2.0ml Tube	65.71	64	0.13	8.32
1.5mL Tubes	60.98	96	0.12	11.52
Ethanol	191		12.00	0.03
PCR Plates	124.00	28	2.56	71.68
Plate Seals	144.77	28	1.44	40.32
Reagents				
NGS Primers	1428	6.24ul	0.157802308	60.59608615
NGS Indexes	1120	2ul	0.036222	13.909248
TAQ	203	400	0.872	348.80
10X Buffer				
MgCl Buffer				
DNTPS	161	400	0.07	11.27
Sera-Mag Beads	276.06	3	23	69.00
PEG	109.51	100	0.109	10.9
KAPA Master Mix	664	880	1.328	1168.64
Qubit Kit	85	5	0.89	4.45
TRIS-HCl	74.5	1	0.12	0.12
Miseq Supplies				
Sodium Hydroxide	83.1	1	0.01	0.01
Miseq Cartridge Kits	925	1		900
Miseq Water	35	7.9	0.56	19.6
Tween-20	37.6	5	0.376	1.88
Sodium Hypochlorite	86.27	2	0.02	0.04
PhiX	160	2	16.00	32.00
Total Cost				**2911.81**

REFERENCES

1. Kayser, M., *Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes.* Forensic Sci Int Genet, 2015. **18**: p. 33-48.
2. Walsh, S., et al., *Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage.* Forensic Sci Int Genet, 2014. **9**: p. 150-61.
3. Maronas, O., et al., *Development of a forensic skin colour predictive test.* Forensic Sci Int Genet, 2014. **13**: p. 34-44.
4. Greytak, E.M. *Using Genome-wide SNPs for DNA Phenotyping and Kinship Inference*. 2015 [cited 2015; Available from: http://docs.parabon.com/pub/Parabon_Snapshot_Scientific_Poster.pdf.
5. Hart, K.L., et al., *Improved eye- and skin-color prediction based on 8 SNPs.* Croat Med J, 2013. **54**(3): p. 248-56.
6. Walsh, S., et al., *The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA.* Forensic Sci Int Genet, 2013. **7**(1): p. 98-115.
7. Walsh, S., et al., *IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information.* Forensic Sci Int Genet, 2011. **5**(3): p. 170-80.
8. Walsh, S., et al., *Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence.* Forensic Sci Int Genet, 2011. **5**(5): p. 464-71.
9. Draus-Barini, J., et al., *Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains.* Investig Genet, 2013. **4**(1): p. 3.
10. Liu, L., et al., *Comparison of next-generation sequencing systems.* J Biomed Biotechnol, 2012. **2012**: p. 251364.
11. Scherer, D. and R. Kumar, *Genetics of pigmentation in skin cancer--a review.* Mutat Res, 2010. **705**(2): p. 141-53.
12. Cichorek, M., et al., *Skin melanocytes: biology and development.* Postepy Dermatol Alergol, 2013. **30**(1): p. 30-41.
13. Yamaguchi, Y., M. Brenner, and V.J. Hearing, *The regulation of skin pigmentation.* J Biol Chem, 2007. **282**(38): p. 27557-61.
14. Rennie, I.G., *Don't it make my blue eyes brown: heterochromia and other abnormalities of the iris.* Eye (Lond), 2012. **26**(1): p. 29-50.
15. Jablonski, N.G. and G. Chaplin, *The evolution of human skin coloration.* J Hum Evol, 2000. **39**(1): p. 57-106.
16. Lao, O., et al., *Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms.* Ann Hum Genet, 2007. **71**(Pt 3): p. 354-69.
17. Norton, H.L., et al., *Genetic evidence for the convergent evolution of light skin in Europeans and East Asians.* Mol Biol Evol, 2007. **24**(3): p. 710-22.

18. Crawford and e.a. N.G, *Loci associated with skin pigmentation identified in African populations.* Science, 2017.
19. Liu, F., et al., *Eye color and the prediction of complex phenotypes from genotypes.* Curr Biol, 2009. **19**(5): p. R192-3.
20. Kayser, M., et al., *Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene.* Am J Hum Genet, 2008. **82**(2): p. 411-23.
21. Valenzuela, R.K., et al., *Predicting phenotype from genotype: normal pigmentation.* J Forensic Sci, 2010. **55**(2): p. 315-22.
22. Visser, M., M. Kayser, and R.J. Palstra, *HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter.* Genome Res, 2012. **22**(3): p. 446-55.
23. Sturm, R.A. and T.N. Frudakis, *Eye colour: portals into pigmentation genes and ancestry.* Trends Genet, 2004. **20**(8): p. 327-32.
24. Han, J., et al., *A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation.* PLoS Genet, 2008. **4**(5): p. e1000074.
25. Sturm, R.A., R.D. Teasdale, and N.F. Box, *Human pigmentation genes: identification, structure and consequences of polymorphic variation.* Gene, 2001. **277**(1-2): p. 49-62.
26. Grimes, E.A., et al., *Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype.* Forensic Sci Int, 2001. **122**(2-3): p. 124-9.
27. Eiberg, H., et al., *Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression.* Hum Genet, 2008. **123**(2): p. 177-87.
28. Frudakis, T., et al., *Sequences associated with human iris pigmentation.* Genetics, 2003. **165**(4): p. 2071-83.
29. Branicki, W., et al., *Model-based prediction of human hair color using DNA variants.* Hum Genet, 2011. **129**(4): p. 443-54.
30. Pospiech, E., et al., *The common occurrence of epistasis in the determination of human pigmentation and its impact on DNA-based pigmentation phenotype prediction.* Forensic Sci Int Genet, 2014. **11**: p. 64-72.
31. Stokowski, R.P., et al., *A genomewide association study of skin pigmentation in a South Asian population.* Am J Hum Genet, 2007. **81**(6): p. 1119-32.
32. Duffy, D.L., et al., *Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma.* J Invest Dermatol, 2010. **130**(2): p. 520-8.
33. Jacobs, L.C., et al., *Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans.* Hum Genet, 2013. **132**(2): p. 147-58.
34. Mushailov, V., et al., *Assay Development and Validation of an 8-SNP Multiplex Test to Predict Eye and Skin Coloration.* J Forensic Sci, 2015. **60**(4): p. 990-1000.
35. Pneuman, A., et al., *Verification of eye and skin color predictors in various populations.* Leg Med (Tokyo), 2012. **14**(2): p. 78-83.

36. Liu, F., et al., *Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up.* Hum Genet, 2015. **134**(8): p. 823-35.
37. Bernard, B.A., *Hair shape of curly hair.* J Am Acad Dermatol, 2003. **48**(6 Suppl): p. S120-6.
38. Eriksson, N., et al., *Web-based, participant-driven studies yield novel genetic associations for common traits.* PLoS Genet, 2010. **6**(6): p. e1000993.
39. Marcinska, M., et al., *Evaluation of DNA variants associated with androgenetic alopecia and their potential to predict male pattern baldness.* PLoS One, 2015. **10**(5): p. e0127852.
40. Li, R., et al., *Six novel susceptibility Loci for early-onset androgenetic alopecia and their unexpected association with common diseases.* PLoS Genet, 2012. **8**(5): p. e1002746.
41. FB, U.B., et al., *Mutations in Three Genes Encoding Proteins Involved in Hair Shaft Formation Cause Uncombable Hair Syndrome.* Am J Hum Genet, 2016. **99**(6): p. 1292-1304.
42. Demay, M.B., et al., *Role of the vitamin D receptor in hair follicle biology.* J Steroid Biochem Mol Biol, 2007. **103**(3-5): p. 344-6.
43. Mohri, Y., et al., *Impaired hair placode formation with reduced expression of hair follicle-related genes in mice lacking Lgr4.* Dev Dyn, 2008. **237**(8): p. 2235-42.
44. Medland, S.E., et al., *Common variants in the trichohyalin gene are associated with straight hair in Europeans.* Am J Hum Genet, 2009. **85**(5): p. 750-5.
45. Pospiech, E., et al., *Evaluation of the predictive capacity of DNA variants associated with straight hair in Europeans.* Forensic Sci Int Genet, 2015. **19**: p. 280-8.
46. Schlake, T., *Segmental Igfbp5 expression is specifically associated with the bent structure of zigzag hairs.* Mech Dev, 2005. **122**(9): p. 988-97.
47. Pasternack, S.M., et al., *G protein-coupled receptor P2Y5 and its ligand LPA are involved in maintenance of human hair growth.* Nat Genet, 2008. **40**(3): p. 329-34.
48. Dias-Santagata, D., et al., *Rapid targeted mutational analysis of human tumours: a clinical platform to guide personalized cancer medicine.* EMBO Mol Med, 2010. **2**(5): p. 146-58.
49. Walsh, S., et al., *Global skin colour prediction from DNA.* Hum Genet, 2017. **136**(7): p. 847-863.
50. Fitzpatrick, T.B., *The validity and practicality of sun-reactive skin types I through VI.* Arch Dermatol, 1988. **124**(6): p. 869-71.
51. Rohland, N. and D. Reich, *Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture.* Genome Res, 2012. **22**(5): p. 939-46.
52. Doyle, S. *Miseq Library Dilution Calculator*. 2014; Available from: https://nematodegenetics.files.wordpress.com/2014/05/miseq-library-concentration-dilution-calculator-aug-2014.xlsx.
53. Li, H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.* Bioinformatics, 2011. **27**(21): p. 2987-93.

54. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
55. Institute, B. *Picard*. Available from: http://broadinstitute.github.io/picard/.
56. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.
57. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.* Brief Bioinform, 2013. **14**(2): p. 178-92.
58. Pineda, G.M., et al., *Development and validation of InnoQuant™, a sensitive human DNA quantitation and degradation assessment method for forensic samples using high copy number mobile elements Alu and SVA.* Forensic Sci Int Genet, 2014. **13**: p. 224-35.
59. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. **29**(1): p. 308-11.
60. Untergasser, A., et al., *Primer3Plus, an enhanced web interface to Primer3.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W71-4.
61. Aranyi, T., et al., *The BiSearch web server.* BMC Bioinformatics, 2006. **7**: p. 431.
62. Vallone, P.M. and J.M. Butler, *AutoDimer: a screening tool for primer-dimer and hairpin structures.* Biotechniques, 2004. **37**(2): p. 226-31.
63. Revelle, W. *psych: Procedures for Personality and Psychological Research*. 2017; Available from: https://CRAN.R-project.org/package=psych.
64. Ballarò, C., et al., *Targeted expression of RALT in mouse skin inhibits epidermal growth factor receptor signalling and generates a Waved-like phenotype.* EMBO Rep, 2005. **6**(8): p. 755-61.

PUBLICATIONS

Walsh, S., Chaitanya, L., **Breslin, K**., Muralidharan, C., Bronikowska, A., Pospiech, E., Koller, J., Kovatsi, L., Wollstein, A., Branicki, W., Liu F., Kayser, M. (2017). Global skin pigmentation prediction from DNA. Human Genetics, 136(7), 847–863.

Chaitanya, L., **Breslin, K,** Clarisse, L., Wirken, L., Pospiech, E Kukla-Bartoszek, M., Sijen, T., Knijff, P., Liu, F., Branicki, W., Kayser, M., Walsh, S. The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. Forensic Sci. Int. Genet. In prep.

Liu, F., Chen, Y., Zhu, G., Hysi, P., Wu, S., Adhikari, K., **Breslin, K.**, Pośpiech, E… Kayser, M. (2017). Meta-analysis of genome-wide association studies identifies 8 novel loci involved in shape variation of human head hair. Human Molecular Genetics. Accepted.

Publications In Prep

Hair Structure Prediction Model Publication

Viking Lineage Ancestry Assay Development

MPS HIrisPlex-S Assay Design and Developmental Validation