Final Technical Report to the

National Institute of Justice

Forensic DNA Unit Efficiency Improvement Program

June 30, 2013

2010-DN-BX-K181 DNA Unit Efficiency Improvement Program

Funding Period: Oct. 1, 2010 – Mar. 31, 2013

DNA Assay Development and Validation for Pigment-Related Features to Assist in the Identification of Missing Persons and Human Remains

New York City Office of Chief Medical Examiner

PI: Elisa Wurmbach, PhD

ewurmbach@ocme.nyc.gov

June 20, 2013

DNA Assay Development and Validation for Pigment-Related Features to Assist in the Identification of Missing Persons and Human Remains

A. Abstract

An essential component in identifying human remains is documenting the decedent's visible characteristics, such as eye, hair and skin color. However, if a decedent is decomposed or only skeletal remains are found, this critical, visibly identifying information is lost. It is presently not possible to use genetic information to reveal these visible characteristics in detail. The objectives of this project were to design and validate an assay to predict eye and skin color based on DNA to assist in the identification of missing persons and human remains. This assay increases the amount of information available for the identification of unknown decedents.

Genetic differences that account for broad variations in individuals' visible traits can be utilized to identify molecular markers, which then in turn can be used to confirm or predict such traits. The variety of phenotypes is caused by multiple polymorphisms in genes of which some are involved in the pigmentation process. The simplest kind of polymorphism is the single nucleotide polymorphism (SNP), and assessing those can reveal visible characteristics of unidentified decedents. Since there are thousands of SNPs, it is challenging to find those few which are directly responsible for a person's eye, hair and skin color.

Association studies including genome-wide SNP-scans have pointed to a few genes relevant for eye, hair, and skin pigmentation. Recent studies identified candidate-SNPs that correlate with high significance to blue and brown eye color, and more SNPs were found to correlate with light skin coloration of East Asians and Europeans and darker complexion of African-Americans. Validation of these SNPs on over 600 samples from individuals of various populations identified seven SNPs located in or nearby pigmentation genes that can be used to describe the eye and skin color solely based on DNA. Five of the seven SNPs are potentially causative: three cause missense mutations and two SNPs are directly located in predicted transcription factor binding sites. The other two SNPs are located in introns and it is still unclear whether they are causative. Six of the seven SNPs are used to predict the eye color,

which distinguishes among brown, green, not blue (green or brown), and not brown (green or blue). All seven SNPs are used to make predictions of the skin coloration of individuals, distinguishing between not white (light brown or dark) and not dark (light brown or white). Both tests can be applied to various populations including African-American, East Asian, South Asian, European and mixed populations. The error-rates for both predictors are very low: 3% for the eye color predictor and 1% for the skin color predictor.

Just recently, the seven-plex system was improved by adding one SNP (rs12896399, which is located in the 5'-region of *SLC24A4*) and by changing the instructions (Hart, Kimura et al. 2013). More precise predictions were achievable by including blue to the eye colors: brown, green, not blue and not brown, and light to the skin colors: not dark and not light. The number of positive descriptions was also significantly increased, while keeping the error rates low, at approximately 5% (Hart, Kimura et al. 2013).

These eight SNPs were integrated into a forensic kit that is easy, fast and inexpensive. This test has potential for upgrading. A multiplex-SNP-assay fitting these criteria would involve a multiplex PCR followed by a multiplex primer extension reaction leading to fluorescently labeled oligonucleotides of distinct length. Multicolor capillary electrophoresis separates and detects these oligonucleotides. It is possible to apply this assay on degraded DNA (typically found in decomposed human remains) due to the design of small PCR products. This is important for forensic applications. Furthermore, development and validation were cost-efficient since all required equipment including the software was already available in-house.

The utilization of this multiplex-SNP-assay to predict pigment-related features will greatly enhance current efforts of collecting data from human remains to facilitate identification. It may be incorporated into the Office of Chief Medical Examiner DNA Missing Persons Group and the Forensic Anthropology Unit's multidisciplinary effort to reduce the number of unidentified human remains in New York City and New York State. A project, funded by the National Institute of Justice: *Using DNA Technology to Identify the Missing*.

B. Table of contents

C. Abbreviations

ASIP – agouti signaling protein

bp – base pair(s)

bHLH – basic helix-loop-helix

CE – capillary electrophoresis

dNTPs – deoxy-nucleoside triphosphates

GWA – genome-wide association

HERC2 – HECT and RLD domain containing E3 ubiquitin protein ligase 2

MC1R – melanocortin 1 receptor

NIJ – National Institute of Justice

NTC – non-template control

OCME – Office of Chief Medical Examiner

PCR – polymerase chain reaction

QPCR – quantitative polymerase chain reaction

SBPE – single base primer extension

SLC24A5 – solute carrier family 24, member 5

SLC45A2 – solute carrier family 45, member 2

SNP – single nucleotide polymorphism

STRs – short tandem repeats

SWGDAM – Scientific Working Group on DNA Analysis Methods

TF – transcription factor

TYR – tyrosinase

D. Executive Summary

D. 1. Statement of the Problem

The utilization of DNA tests in forensics usually refers to short tandem repeat [STR] profiles that can identify individuals when compared to databases and/or reference samples. However, if there is no data available to compare the profile to then the STR profile is not of any value in terms of identifying a person. In missing persons cases or in mass disaster investigations the utilization of genetic information could reveal phenotypic characteristics to help identify these individuals. However, genetic information is not commonly utilized in forensics, and one reason for this could be that there are no tests available.

Visible characteristics are crucial in identifying human remains. If a decedent is decomposed or only skeletal remains are found, this critical and identifying information is lost. Some of these characteristics such as gender, ethnicity, height, and build can be determined by forensic anthropologists based on (more or less) intact skeletal remains. In the absence of anthropological evidence, using DNA to determine these traits could be very useful. In addition, pigmentation traits such as eye, hair and skin color cannot be determined by anthropological bone analysis.

Compared to the other visible characteristics most is known about the traits in pigmentation, but predictions with 100% accuracy are not feasible. It is estimated that over 100 genes are involved in melanin synthesis. Melanin pigments are found in the iris, hair and skin, but its synthesis may be regulated in a tissue-dependent manner and may involve specific genes. This complexity reveals some of the difficulties in creating a forensic test. Recent advances in molecular genetics led to the selection of a few markers that show significant correlation with pigmentation traits.

We created a predictive test to describe eye and skin color from DNA utilizing seven markers (Spichenok, Budimlija et al. 2011). The awarded funding from NIJ (2010-DN-BX-K181) allowed us then to verify published predictive tests on over 600 samples of various populations to determine error- and call-rates (Pneuman, Budimlija et al. 2012). These are important attributes since they may increase the confidence of a predictive test. Just recently, the 7-plex system was improved by adding one SNP and by changing the instructions. More precise predictions were achievable and the number of positive descriptions was significantly increased, while keeping low error rates (Hart, Kimura et al. 2013).

The developed forensic test included eight markers to predict the eye and skin color of an individual. The method of choice, a multiplex-SNP-assay, was based on techniques available in most forensic laboratories. Reliability, reproducibility, robustness and accuracy of this assay were ensured by its validation.

D. 2. Purpose, Objectives and Goals

The objective of this project was to establish and validate a novel multiplex assay that has the potential to assist in the identification of human remains by predicting eye and skin color of individuals based solely on nuclear DNA.

Single nucleotide polymorphisms [SNPs] that show significant correlation with human eye and/or skin coloration were chosen and tested on over 800 samples from various populations. Seven robust and informative SNPs were selected to be incorporated into a predictive test. The following verification established error- and call-rates thereby increasing the confidence of the predictions. Just recently, the 7-plex system was significantly improved by adding another SNP. An easy-to-use multiplex-SNP-assay was developed based on these findings. The multiplex-SNP-assay makes efficient use of nuclear DNA. It is inexpensive and the technical procedure is known to most forensic laboratories. The method of choice, a multiplex-PCR followed by a multiplex primer extension reaction, also has the advantage of being effective on old and degraded biological material. The validation of this assay demonstrated reliability, reproducibility, robustness and accuracy. This assay represents a novel forensic technique that improves and complements current efforts in identifying human remains and missing persons.

D. 3. Prediction of eye and skin color using seven SNPs (Spichenok, Budimlija et al. 2011)

Most single nucleotide polymorphisms [SNPs] are silent and have no impact on gene function or phenotypes. Some SNPs however, can have functional consequences and may contribute to phenotypes. Association-studies of SNP-genotypes with human traits detected significant SNP-candidates for eye, hair and skin coloration (Sulem, Gudbjartsson et al. 2007; Han, Kraft et al. 2008; Sulem, Gudbjartsson et al. 2008).

Seven SNPs (rs12913832, rs1545397, rs16891982, rs1426654, rs885479, rs6119471, and rs12203592) located within or nearby genes (*HERC2*, *OCA2*, *SLC45A2*, *MC1R*, *ASIP*, and *IRF4*) involved in

pigmentation were tested in samples of various populations. Four of them (rs12913832, rs16891982, rs12203592, and rs1426654) were selected for their significant correlation with the blue or brown eye color (Eiberg, Troelsen et al. 2008; Sturm, Duffy et al. 2008; Liu, van Duijn et al. 2009; Walsh, Liu et al. 2011) and/or their role in reduced melanin content in cultured human melanocytes (Cook, Chen et al. 2009). Five of the seven selected SNPs are potentially causative: three cause missense mutations and two are directly located in predicted transcription factor [TF] binding sites. The other two SNPs are located in introns and it is still unclear whether they can be causative.

The eye color predictor was created based on significant correlations of six SNPs (rs12913832, rs1545397, rs16891982, rs885479, rs6119471, and rs12203592) with eye colors and/or global populations. As described here, a predictor is a standardized test that includes, alongside the selected SNPs, explicit instructions on how to interpret the genotype to reach a prediction. A predictor has the advantage of being independent of the interpreting analyst or the performing laboratory and can be validated.

The eye color prediction relied mainly on rs12913832, as shown in Figure 1, in an exclusive approach: G/G-homozygotes were predicted to have non-brown eyes, meaning the eye color could be in the green or blue range; A/G-heterozygotes and A/A-homozygotes were predicted to have non-blue eyes (green or brown range). Additional SNPs are necessary for a more discriminating prediction. Homozygous genotypes of the remaining five SNPs (rs12913832, rs1545397, rs16891982, rs885479, rs6119471, and rs12203592) were used to further describe the eye color as being either brown or green.

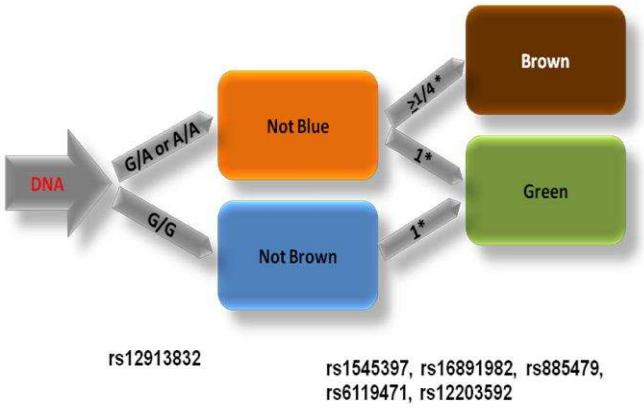

Figure 1: Schematic overview of eye color predictor based on seven SNPs.

The eye color predictor was applied to 554 independent samples from various populations including African-American, South Asian, East Asian, European, and mixed populations. The predicted eye color was then compared to the real eye color, which was grouped into one of the three eye color bins (brown, green, or blue).

The outcome revealed that the eye color predictor can be used on all populations tested including the mixed population, which could be considered as the ultimate testing group. This is an important outcome for forensic analysts because samples of human remains may not allow for the determination of ethnicity. Since people travel and migrate, populations are no longer tied to certain geographical regions.

The determined error-rate for the eye color prediction was very low. Out of the 554 independent samples only four errors occurred, leading to a calculated error-rate of approximately 1%. Each sample, independent of the ethnic origin, led to a prediction, revealing a call-rate of 100%.

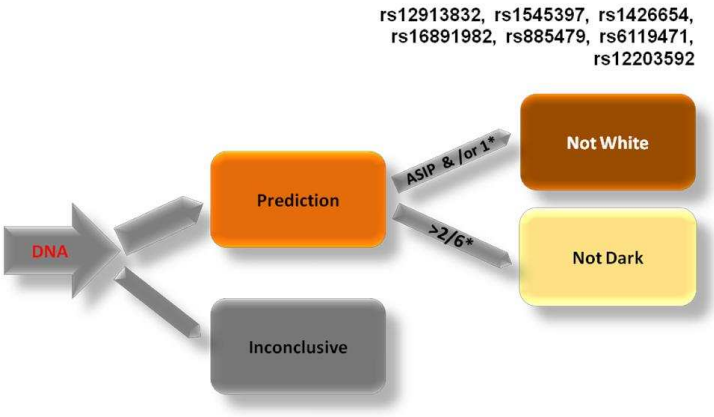

Figure 2: Schematic overview of skin color predictor based on seven SNPs.

The seven selected SNPs were used for the skin color prediction (Figure 2). The skin color was grouped into three bins according to the observed melanin content: light, medium and dark. The skin coloration followed a process of elimination, using only homozygous genotypes. Because of the complexity of this trait at least two SNPs (G/G at rs12913832, T/T at rs1545397, G/G at rs16891982, A/A at rs1426654, A/A at rs885479, or T/T at rs12203592) were used to predict non-dark (light or medium) skin color. One SNP (G/G at rs6119471) was used to predict the skin color of being not light (medium or dark).

As for the eye color prediction, the skin color predictor can be used on all populations tested. However, for several samples the data were insufficient to predict skin coloration: from 554 samples tested 398 predictions could be made (72%). Of these 398 predictions only two errors occurred, revealing a very low error-rate.

D. 4. Validation of eye and skin color predictor (Pneuman, Budimlija et al. 2012)

Published predictors were validated in order to proceed with the one we consider superior: "IrisPlex", an eye color predictor (Walsh, Liu et al. 2011), and the eye and skin color predictor utilizing seven SNPs (Spichenok, Budimlija et al. 2011).

Validation of testing methods is an essential feature in all scientific endeavors, but it is particularly important in forensics. Due to the sensitive nature of these investigations and the limited content of collected evidence, it is crucial to validate all employed procedures to learn more of their capabilities and limitations before incorporating them as routine methods. Ideally, validations are performed on large sample sets that mimic real cases.

The eye color predictor "IrisPlex" utilizes six SNPs (rs12913832, rs16891982, rs1800407, rs12896399, rs1393350, and rs12203592), of which three are identical with the 7-plex-system, to distinguish between the eye-colors brown, green/intermediate and blue (Walsh, Liu et al. 2011). It calculates probabilities for each of them, and the probability that exceeds 0.7 predicts the eye color. This eye color predictor was validated on 803 independent samples of various populations. It was found that "IrisPlex" can be applied and works with all of them. Since the predictor operates with thresholds, it results in some inconclusive outcomes that do not reach the threshold (26%). The error-rate was calculated to be 31%. Most errors were due to the green eye color that was either predicted as being brown or blue (Pneuman, Budimlija et al. 2012).

The other predictor, for eye and skin color based on the seven selected SNPs, was tested on an additional 251 newly-collected samples. Seven errors were found out of the 251 eye color predictions (call-rate was 100%). The error-rate for this validation was then calculated to approximately 3%, which is within the range of the previous test. Testing the 251 samples led to 203 skin color predictions, out of which 2 errors occurred confirming the low error-rate (1%). The call-rate was approximately 80% and depended on the population tested which confirmed previous data (Spichenok, Budimlija et al. 2011).

This validation was performed on newly collected samples to gain more information and to increase the confidence of the prediction by calculating error-rates (and call-rates).

D. 5. Improved eye- and skin-color prediction based on 8 SNPs (Hart, Kimura et al. 2013)

Just recently, our research improved the 7-plex system. Analysis of an eighth SNP, rs12896399 (*SLC24A4*), on 803 samples showed a statistically significant association with human eye color (p=0.007) but a rather poor strength of agreement (=0.063). Based on these results rs12896399 was added to the 7-plex system to generate an 8-plex system. In addition, the instructions on how to interpret the genotypes unambiguously were also changed: Five of the eight SNPs are used to predict eye color (blue, green, brown, not blue, and not brown) and six to predict skin coloration (light, not dark and not light), whereby three are shared.

The newly-designed 8-plex system was established on 803 samples (training set). The prediction for eye color is a two-step procedure: First, rs12913832 (*HERC2*) distinguishes eye colors as being not blue (i.e. brown or green) or not brown (i.e. green or blue). Further distinctions for predicting eye color as being brown, green or blue are dependent on the genotypes of three additional SNPs: rs12203592, rs16891982, and rs6119471 (Fig. 3). Brown eye color is predicted by the following genotype combinations: A/A or G/A at rs12913832, plus either G/G at rs6119471, or C/C at rs16891982; green eye color is predicted by G/G at rs12913832 plus C/C at rs16891982, or by G/A at rs12913832 plus T/T at rs12203592; and blue eye color is predicted by G/G at rs12913832 plus T/T at rs12203592.

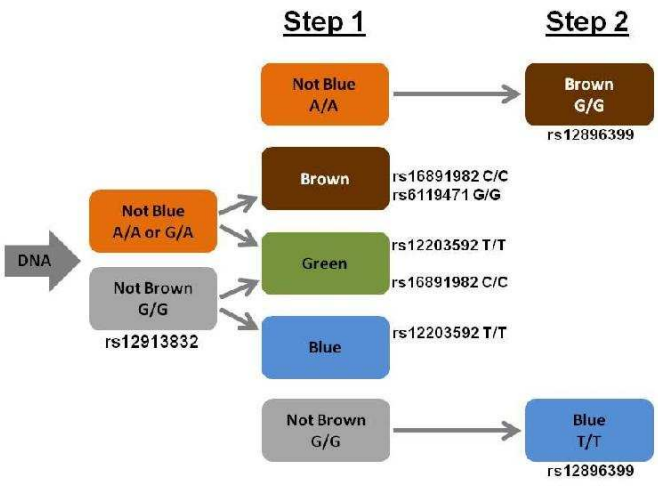

Figure 3: Schematic overview of eye color predictor based on eight SNPs.

The second step proceeds with samples that were not positively described as being brown, green or blue in the first step (Fig. 3). Samples that are homozygous for G/G at rs12913832 (i.e. not brown) plus T/T at rs12896399 are then predicted to have blue eyes, while samples that are homozygous for A/A at rs12913832 (i.e. not blue) plus G/G at rs12896399 are predicted to have brown eyes. This step utilizes the newly-included SNP rs12896399 (*SLC24A4*) and leads to increased numbers of positively described eye colors.

212 samples were collected to test the newly-developed 8-plex system. The eye-color prediction followed a two-step procedure, whereby the first step led to 56 positive descriptions, in which 2 errors occurred (error rate of 1.75%). The second step, including rs12896399, led to 30 further positive descriptions (56+30=86). Three errors were found in the second step (error rate equals 2.73%). All together five errors were counted; two arose from rs12913832, which was recently identified within the 7-plex system as an error source that led to a 3% error rate (Pneuman, Budimlija et al. 2012). The other three errors were related to the newly included rs12896399. All five errors found in the test set (n=212) of the 8-plex system occurred in the European population and led to an error rate of approximately 5% (4.54%).

Skin color prediction of the 8-plex system utilizes six SNPs. In a similar manner to the eye color, the skin color is binned into three groups: light, medium and dark. The skin-color predictor utilizes five SNPs to predict a lighter skin color and one to predict darker shades. Skin coloration is predicted in an elimination process. Non-dark skin color (i.e. light or medium) is predicted by any two of the following alleles: G/G at rs12913832, G/G at rs16891982, A/A at rs1426654, T/T at rs1545397, or A/A at rs885479. Light skin color is predicted by more stringent conditions: G/G at rs12913832, plus G/G at rs16891982, and A/A at rs1426654. Non-light skin color (i.e. medium or dark) is predicted by G/G at rs6119471.

The validation on 212 samples revealed that a prediction could be made for 131 (62%) samples out of 212 samples. Of these, 63 were predicted to have light skin color and all of those were correct.

The newly-developed eye- and skin-color predictor utilizing 8 SNPs is an improved version of the 7-plex system, leading to more precise descriptions including blue-eye and light-skin coloration. In addition, the new 8-plex system increased the number of positive descriptions significantly (p=6.98e-15, z-score: -7.786) while keeping the error rate low.

D. 6. Development of multiplex-SNP-assay

We decided to proceed with the multiplex-SNP-assay that predicts eye and skin color using seven selected SNPs (rs12913832, rs16891982, rs1545397, rs1426654, rs885479, rs6119471, rs12203592, and rs12896399) that are associated with seven pigmentation genes (*HERC2*, *OCA2*, *SLC45A2*, *SLC24A5*, *MC1R*, *ASIP*, *IRF4, and SLC24A4*). The multiplex-SNP-assay includes a multiplex-PCR followed by a multiplex single-base primer extension reaction, which creates fluorescently labeled oligonucleotides of distinct lengths. The detection of the fluorescent signals is achieved by multicolor capillary electrophoresis. This assay has the main advantages of making efficient use of DNA and utilizing methods available in most forensic laboratories.

The first step of the assay development addressed the multiplex-PCR. Primers were designed to flank the selected SNPs, the targets of each reaction. The primers must work under similar conditions because the eight PCR reactions would eventually be pooled in order to create the multiplex-PCR.

Numerous primers targeting each SNP were designed and tested in singleplex PCR reactions with the aim to find at least one PCR-reaction that would show the proper product. The reaction conditions were kept the same. The outcome was verified using agarose gel electrophoresis, allowing for examination of additional products throughout all size-ranges. Singleplex PCR reactions that showed the desired outcome with regard to size and similarity in abundance were combined into multiplex reactions.

Initially, the multiplex-PCR-reactions were not showing the expected results. The PCR products, confirmed by their size, varied greatly in their amount and some of them even vanished when combined with other PCR reactions. Since it was confirmed that no larger products appeared, the multiplex-PCR-reactions were monitored using the Agilent Bioanalyzer (DNA 1000 Kit, Agilent Technologies, Santa Clara, CA), which had a better resolution. However, amplicons of similar sizes, 80bp and 87bp for example, could not be separated. Therefore, only PCR reactions that could be separated based on their product lengths were combined. This led to the consequence of incorporating longer PCR products. Testing a large number of primer combinations solved the problem of undetectable PCR-products and resulted in a multiplex-PCR-reaction that revealed all eight amplicons.

During the single-base primer extension [SBPE] reaction, a primer anneals one bp short of the targeted SNP and is extended using a fluorescently-labeled dideoxy-nucleotide which is complementary to the targeted SNP. Further extension is prevented because the required 3'OH group is missing. All eight SBPE-reactions should be performed in a multiplex reaction; therefore all reactions must work under the same conditions.

Based on genotypes, as well as on flanking sequences of the selected SNPs, several SBPE 8-plex scenarios were designed. A positive control, which is heterozygous at each SNP (allelic ladder), was created from three samples. The various SBPE-scenarios were tested with the newly created allelic ladder, a sample of which the eight genotypes were known, and a non-template control [NTC]. Some of the scenarios revealed peaks in the NTC due to primer-dimer or hairpin formation. Eventually, one of the SBPE-scenarios showed the expected results: two peaks per SNP for the allelic ladder, a full and correct profile for the known sample, and no peaks for the NTC. This SBPE reaction was combined with the multiplex-PCR to form the multiplex-SNP-assay (Figure 4). In the following step, this assay was validated according to the SWGDAM guidelines (SWGDAM 2004).

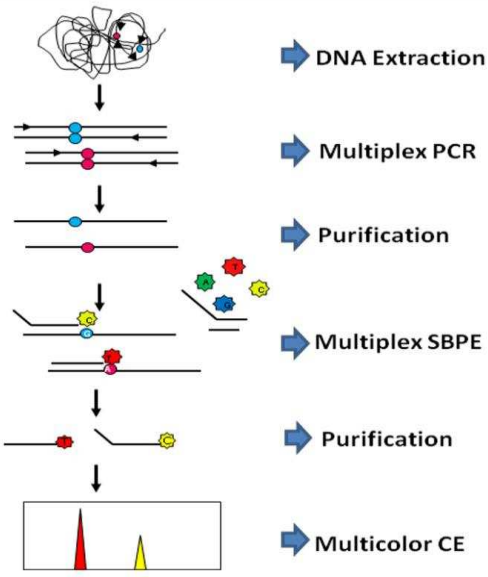

Figure 4: Outline for the multiplex-SNP-assay (shown for two SNPs).

D. 7. Validation of multiplex-SNP-assay (Mushailov, Rodriguez et al.)

The multiplex-SNP-assay, which determines the genotypes of eight SNPs to predict eye and skin color of individuals, consists of a multiplex-PCR followed by a multiplex-SBPE reaction. This incorporates fluorescently-labeled oligonucleoties of distinct length that can be separated and detected using a multicolor capillary electrophoresis (Figure 4). The validation of this method followed the guidelines of SWGDAM, a forum to evaluate forensic biology methods to enhance services. The validation assesses the ability of a procedure to obtain reliable results, determines the conditions under which such results can be obtained, and defines the limitations of this procedure ((SWGDAM) 2004).

The validation of the multiplex-SNP-assay included investigations on:

- *reliability*: the outcome of the assay, the genotypes, was confirmed by another method (QPCR: *TaqMan*-assay);

- *reproducibility*: several technical repeats, performed by different persons on different days, led to the same outcome;

- *various DNA sources*: various DNA extraction methods from blood, skin cells, semen, and saliva led to full profiles to determine eye and skin color, revealing that none of these extraction techniques interferes with the multiplex-SNP-assay;

- *robustness*: testing several instruments (including various PCR machines, which may differ in their ramping times, and multicolor CEs) had no impact on the outcome of the assay. In addition, PCR conditions were varied to reveal a range in which the assay works accurately;

- *sensitivity*: nuclear DNA that is higher in quantity than 100pg led to full and correct profiles in each case;

- *degraded material*: Three models of DNA degradation were tested, including treatment with DNase I, boiling at 95^{o}C, and UV-radiation. Further, extracted DNA from several archival samples (blood stain cards – considered as waste tissues, after obtaining internal OCME approval in compliance with OCME policy manual, January 1, 2010) led to full profiles; and

- *species specificity*: DNA from various (15) species were tested. None of them led to all eight PCR-products; therefore no full profile was obtained. The multiplex-SNP-assay is specific for humans.

The validation demonstrated that the developed multiplex-SNP-assay is reliable, robust, sensitive, and can be used on challenging samples in forensics to determine reliably eye and skin color of individuals.

D. 8. Implications for Policy and Practice

The number of missing persons and unidentified human remains is what experts call a mass disaster over time (Ritter 2007). About 100,000 missing person cases are active in the United States. In June 2007, the Office of Justice Program's Bureau of Justice Statistics confirmed that in a typical year, medical examiners and coroners handled approximately 4,400 unidentified human decedent cases, of which 1,000 remain unidentified (Hickman 2007).

Efforts in identifying human remains and solving missing person cases include the use of forensic DNA, which is supported by the National Institute of Justice [NIJ] in several programs: training on forensic DNA evidence, providing free tests for unidentified human remains and family references, encouraging States to collect DNA samples before unidentified remains are disposed, and supporting the use of Federal databases to solve missing person and unidentified human remain cases (Ritter 2007).

The ability to predict visible pigment-related features based on DNA has the potential to be incorporated into this multidisciplinary effort to identify the missing and unidentified human remains (Budowle and van Daal 2008). Prediction of eye and skin colors by utilizing the multiplex-SNP-assay was the goal of this project. This could provide enormously helpful information which could be entered into NamUs, a national database for missing and unidentified persons.

E. Final Technical Report

E.1. Introduction

While DNA has been used in forensic science for over 25 years, all of its inherent informative power has not yet been exploited. DNA is generally used for individual profile comparisons, whereby several short tandem repeats [STRs] are assessed to generate a profile that is subsequently compared to specific reference samples or a database to make identifications. Recent scientific findings regarding the inherent information of the DNA have found that sequence variations at distinct loci can be used to predict phenotypic features of a person. Phenotypic DNA tests could be used to provide information about physical characteristics aiding in the identification of human remains and missing persons. Some of these characteristics such as gender, ethnicity, height, and build can be determined by forensic anthropologists based on (more or less) intact skeletal remains. In the absence of anthropological evidence, using DNA to determine these traits could be very useful. Further, pigmentation traits, such as eye, hair and skin color, cannot be determined based on bone analysis.

E. 1. 1. Statement of Problem

An essential component in identifying human remains is the documentation of the decedent's visible characteristics such as eye, hair and skin color. If a decedent is decomposed or only skeletal remains are found, much of the critical visibly-identifying information is lost. It would be beneficial to use genetic information to reveal these traits.

SNPs located within or nearby pigmentation genes may correlate strongly with eye, hair or skin color. Due to the complexity of pigmentation traits (e.g. being dependent on several genes and other factors), one SNP may not be sufficient to describe one characteristic even though its correlation may be statistically significant. Eight SNPs were chosen for their expected correlation with pigment-related features to describe the eye and skin color of an individual. A standardized forensic test - a predictive test - requires the selection of SNPs as well as instructions on how to interpret the genotype. This makes the test independent of the interpreting analyst or the performing laboratory. In addition, this allows its validation on various populations, which is important for forensic tests. Since samples from certain human remains do not allow for the determination of ethnicity, even if a population is more prominent in a specific area people travel and migrate. Further, by comparing the prediction with the

corresponding phenotype, call- and error-rates can be determined. These validations add valuable information to the predictor and increase confidence in its predictive power.

A forensic test that utilizes the validated eye- and skin-color predictor would ideally be a multiplex-SNP-assay to save precious material. The multiplex-SNP-assay was based on a multiplex-PCR followed by a multiplex single-base primer extension [SBPE] reaction, which creates fluorescently-labeled oligonucleotides of distinct lengths. These are detected by multicolor capillary electrophoresis. This method was chosen because it is extremely cost-effective, since all equipment was already available in-house. This assay is specifically valuable for severely fragmented template DNA since it relies only on short amplicons and therefore is well suited for forensic samples.

Development, optimization and validation of this genotyping assay resulted in a robust test that consumes very little genomic DNA and can be used routinely to assist in the identification of missing persons and human remains.

E. 1. 2. Literature Review and Citations

Usually a missing person's report contains general descriptive information about an individual. With skeletonized human remains, physical characteristics such as age, ethnicity, gender and height are determined by forensic anthropologists. While these metrics are extremely useful, it is not possible to infer eye, hair, or skin color from skeletal remains. These characteristics are largely genetically determined, and recently, some DNA markers related to these traits have been identified. Melanin is the main pigment of eye, hair and skin color. It is packaged in specialized subcellular compartments, the melanosomes, which are exported to adjacent keratinocytes where most pigment is found (Barsh 2003). Differences in pigmentation arise from variation in number, size, composition, and distribution of melanosomes (Lin and Fisher 2007). During melanogenesis, two types of melanin are produced: the brown-black eumelanin and the red-yellow pheomelanin. Both melanins are found in the iris, hair and skin. There, the amount and ratio of both melanins varies between individuals, resulting in the many shades of eye, hair and skin color which people may have (Sturm and Frudakis 2004).

Only a minimal correlation exists among iris, hair and skin color within the European population, where blue- and brown-eyed individuals can have all shades of natural hair colors. In contrast, in other geographical regions, populations with darker skin tones also tend to have darker eye and hair colors.

Skin color correlates strongly with the ultraviolet (UV) radiation levels and can be explained by varying physiological requirements of photoprotection and vitamin D synthesis (Jablonski and Chaplin 2000; Norton, Kittles et al. 2007). This suggests that genetic determinants for pigmentation in various tissues are distinct, though they have been subject to a common set of evolutionary forces that have shaped their distribution in world populations (Frudakis, Thomas et al. 2003; Chaplin 2004).

Comparison of two randomly selected human genomes demonstrates 99.9% identity. The remaining 0.1% contains variations, including single nucleotide polymorphisms [SNPs]. The human genome contains at least 11 million SNPs, which occur at approximately 1 in 1000 base pairs (bp) (Brookes 1999; Kruglyak and Nickerson 2001).

Most SNPs do not contribute to a phenotype, but others are associated with susceptibility to diseases or visible traits such as height or curly hair (Martin, Boomsma et al. 1997). Some of these traits are multi-genic and could also be influenced by other factors such as age or the environment and are considered as complex traits. Diabetes, schizophrenia, height and, importantly, pigmentation are a few examples of these. This project focuses on human pigmentation. Features like height or build were not included in this project, as it would involve too many genes, even more SNPs, as well as numerous samples to reach statistical significance.

To reduce complexity, previous studies focused on a single trait in one population, or on few genes, such as *MC1R*, *SCL45A2*, *OCA2*, *HERC2*, *ASIP*, and *SLC24A5*. Those gene products are associated with melanosomes, involved in the melanosome transport, the uptake by keratinocytes, or are part of the melanin biogenesis cascade (Valverde, Healy et al. 1995; Stokowski, Pant et al. 2007; Branicki, Brudnik et al. 2008; Shekar, Duffy et al. 2008; Branicki, Brudnik et al. 2009). Genome-wide association [GWA] studies identified new alleles associated with human pigmentation (Sulem, Gudbjartsson et al. 2007; Han, Kraft et al. 2008; Sulem, Gudbjartsson et al. 2008). Much progress was achieved with association studies performed in Europeans, which identified the *OCA2* locus as major contributor for eye color determination (Frudakis, Thomas et al. 2003; Sturm and Frudakis 2004; Duffy, Montgomery et al. 2007; Frudakis, Terravainen et al. 2007). The *OCA2* gene codes for a 12 pass-transmembrane protein with homology to ion permeases. This protein is localized at melanosomes and involved in melanin synthesis (Sitaram, Piccirillo et al. 2009).

Another GWA study that focused solely on eye color identified *HERC2*, a gene directly adjacent to *OCA2*, as the gene specifying human iris color (Kayser, Liu et al. 2008). Fine mapping of the *OCA2*-*HERC2* locus performed in European populations identified two SNPs, rs12913832 and rs1129038, located in the 86th intron and the 3'UTR of *HERC2* that demonstrate strong correlation with blue and brown eye color (Eiberg, Troelsen et al. 2008; Sturm, Duffy et al. 2008). It was found that intron 86 of the *HERC2* gene contains a highly conserved promoter region of *OCA2* and that the SNP rs12913832 is directly located in a predicted transcription factor [TF] binding site (Sturm, Duffy et al. 2008). Studies in human melanocytes detected binding of TFs to this side. The A-allele of rs12913832 led to elevated OCA2 expression, which was reduced in lighter pigmented cells carrying the G-allele (Visser, Kayser et al. 2012).

Less is known about hair and skin color. Lighter eye colors (blue, gray and amber) as well as lighter hair colors (blonde and light red) are more common in the European population, in particular in the North-Eastern part and the East Baltic (Frost 2006). It was shown that light eye and light hair color as well as brown eye and dark hair color are significantly correlated in Scottish and Danish populations (Mengel-From, Wong et al. 2009). Variations in the *MC1R* gene have been associated with red hair due to elevated levels of pheomelanin (Valverde, Healy et al. 1995; Branicki, Brudnik et al. 2007). This gene codes for the melanocortin 1 receptor, a G-protein-coupled receptor controlling melanogenesis.

SLC45A2, *SLC24A5*, and *HERC2* seem to be responsible for the total melanin content in hair color (blonde to black) (Valenzuela, Henderson et al. 2010). A model for hair color prediction based on 385 individuals included, for this reason, two compound markers located in the *MC1R*-gene ("R" and "r", which refer to seven and four variations, respectively), as well as 11 SNPs associated with 10 additional genes. This model distinguishes between blond, brown, red, and black hair (Branicki, Liu et al. 2011). Further research will lead to more useful predictions (Bouakaze, Keyser et al. 2009; Kayser and Schneider 2009).

Skin coloration is closely associated with various populations. Investigations of SNP variations in six genes (*SLC45A2, SLC24A5, OCA2, TYR, ASIP*, and *MC1R*) found evidence that the light skin color of Europeans and East Asians arose independently (Tishkoff and Verrelli 2003; McEvoy, Beleza et al. 2006; Norton, Kittles et al. 2007). Consequently, more than one marker is necessary to predict skin coloration (Brilliant 2008; Valenzuela, Henderson et al. 2010).

This project established, developed and validated a forensic method that is able to predict traits in pigmentation reliably and independent of ethnic or geographic population. Eight meaningful SNPs were selected that show strong correlation with eye and skin color (Spichenok, Budimlija et al. 2011; Pneuman, Budimlija et al. 2012; Hart, Kimura et al. 2013). A multiplex-SNP-assay was developed and validated to be used as a forensic kit. This assay is robust and fast, and has the advantage of being effective on degraded and low quantity DNA (Bouakaze, Keyser et al. 2009).

E. 1. 3. Purpose, Goal and Objectives

The objective of this project was to develop and validate a novel multiplex assay based on nuclear DNA that has the potential to assist in the identification of unidentified human remains by predicting eye and skin color. Identified SNPs have been validated on over 1000 samples from various populations; eight robust and predictive SNPs were selected and incorporated into a multiplex-technique. This includes a multiplex-PCR followed by a multiplex primer extension reaction, which has the advantage of being effective on old and degraded biological material. This assay represents a novel forensic method that improves and complements current methods. In addition, progression to upgraded versions with additional SNPs (e.g. for additional traits) could use the same technique and therefore could easily be integrated into the assay.

E. 2. Materials and Methods

This research project, entitled *Validation of Markers for Human Pigmentation to Assist in the Identification of Unidentified Human Remains,* was approved by the New York City Department of Health and Mental Hygiene Institutional Review Board that serves as IRB for the Office of Chief Medical Examiner [OCME] (IRB# 08-066).

E. 2. 1. Equipment and Materials

All necessary equipment, including NanoDrop 1000 (Thermo Scientific, USA), Agilent Bioanalyzer (Agilent Technologies Inc, USA), gel-electrophoresis (Thermo Fisher Scientific, USA), gel-documentation system (Imgen Technologies, New City, NY, USA), PCR- (thermal cycler, ABI, Foster City, CA) and real-time PCR-machines (RotorGene 6000, Qiagen, Valencia, CA), multicolor capillary electrophoresis detection systems (3130*xl* DNA Genetic Analyzer, ABI, Foster City, CA), and the software GeneMapper (ABI, Foster City, CA) required for data analysis were already available at the OCME.

The following kits, enzymes and chemicals were used: Puregene DNA Procedure kit (Qiagen, Valencia, CA), Puregene Proteinase K (Qiagen, Valencia, CA), AmpliTaq Gold PCR Mastermix (ABI, Foster City, CA), Multiplex-PCR kit (Qiagen, Valencia, CA), GoTaq Hot Start Polymerase PCR kit (Promega, USA), pGEM DNA Marker (Promega, USA), Agilent DNA 1000 kit (Agilent Technologies Inc, USA), SNaPshot Primer Focus kit (ABI, Foster City, CA), ABI PRISM SNaPshot Multiplex Kit (ABI, Foster City, CA), ExoSAP-IT (Affymetrix-USB, USA), SAP (Affymetrix-USB, USA), and TaqMan genotyping assays (ABI, Foster City, CA).

E. 2. 2. Sample collection

Before donating a sample each volunteer read and signed the consent form. For precise data acquisition, each volunteer filled out a questionnaire, which asked for detailed information on eye, hair, and skin coloration. For confirmation and to prevent bias, a picture of the eye was taken. Volunteers were also asked for their association with populations.

E. 2. 3. Sample binning

Collected eye and skin color information was assigned into three bins: blue, green or brown for eye color and light, medium or dark for skin coloration. The populations were distinguished among African-American, South Asian, East Asian, European descendants, and mixed, which included Hispanics (Tishkoff and Verrelli 2003) as well as individuals whose parents were not associated with the same geographic population.

E. 2. 4. DNA extraction

Buccal swabbing was used to collect DNA samples. DNA extractions were performed following instructions of the manufacturer (Gentra Puregene Buccal Cell Kit, Qiagen, Valencia, CA, USA) with slight modifications, which were published elsewhere (Spichenok, Budimlija et al. 2011). Chelex extraction was used for blood and buccal samples. Differential chelex extraction was used to separate sperm cells from epithelial cells. M48 (MagAttract) extraction is an automatic DNA extraction method that uses a BioRobot, whereby magnetic resins are utilized. This extraction was used for blood stains on FTA paper. Extracted DNA was quantified by an in-house human DNA quantification kit based on quantitative real-time PCR (QPCR) (Caragine, Mikulasovich et al. 2009), or the NanoDrop (Thermo Scientific, Wilmington, DE).

E. 2. 5. Singleplex-SNP-assay

The allelic discrimination of the seven SNPs (rs12913832, rs1545397, rs16891982, rs1426654, rs885479, rs6119471, rs12203592, and rs12896399) was performed by PCR-based *TaqMan*-assays in the presence of two differently fluorescently-labeled probes which allow for the detection of both alleles in a single reaction (Applied Biosystems Inc, Foster City, CA, USA). Using optimized PCR-conditions [volume 25ul; using the TaqMan Universal PCR Master Mix as indicated by the manufacturer (Applied Biosystems Inc.), PCR: 10min 95^{O}C, 50 cycles: 60sec 60^{O}C, 15sec 92^{O}C, performed on RotorGene 6000 (Qiagen, Valencia, CA)], reliable and accurate results were obtained with 100pg or more chromosomal DNA. No assay including lower concentrations (down to 5pg) showed a false positive result.

E. 2. 6. Multiplex-PCR

Multiplex-PCR reactions were usually performed in 25 l using the primers [rs12913832: F 5'-GGC TCT CTG TGT CTG ATC CA, R 5'-GGC CCC TGA TGA TGA TAG C; rs16891982: F 5'-TCC AAG TTG TGC TAG ACC AGA, R 5'-CGA AAG AGG AGT CGA GGT TG; rs1426654: F 5'-CCC TTG GAT TGT CTC AGG ATG, R 5'-TGA GTA AGC AAG AAG TAT AAG GAG CAA; rs1545397: F 5'-TGG AAT TGG ATA CTG ACA ATG G, R 5'-TCA CCG TGG GTA GAA TTA CCA; rs885479: F 5'-GCA CTG CGC TAC CAC AGC AT, R 5'-CCA GCA TAG CCA GGA AGA AG; rs6119471: F 5'-ATG GCG GGT GCC TAC TCT A, R 5'-AGG CTA ACC CGA AGG AAG AG; rs12203592: F 5'-GCT TTG TTT CAT CCA CTT TGG, R 5'-TGG AGT GAA CCC TTC ATC AG; and rs12896399: F 5'-CTG GCG ATC CAA TTC TTT GT, R 5'-CTT AGC CCT GGG TCT TGA TG] and the GoTaq Hot Start Polymerase (Promega, Madison, WI, USA) under standard conditions (buffer 1x, $MgCl_2$ 2mM, dNTPs 0.2mM each, GoTaq 1.25u). Genomic DNA was used at 0.5 to 1ng as template, except otherwise indicated. The PCR conditions were 2min at 95^{O}C, 40 cycles: 15sec at 95^{O}C, 15sec at 59^{O}C, 30sec at 72^{O}C, followed by a final 2min at 72^{O}C on thermal cycler systems 9700 (Applied Biosystems Inc, Foster City, CA, USA). The PCR products were checked on agarose-gel electrophoresis and/or Bioanalyzer DNA 1000 chip separating between 15 and 1500bp (Agilent, Santa Clara, CA, USA), following the instructions of the manufacturer. For validation template DNA and $MgCl_2$ concentrations, as well as PCR annealing temperature, numbers of cycles and reaction volume were varied.

E. 2. 7. PCR clean-up

The clean-up reaction digests primers and dephosphoryaltes dNTPs from the multiplex-PCR to nucleosides and phosphates that cannot interfere with the following single-base primer extension reactions. This was performed by adding ExoSAP IT Product Cleanup (USB Affymetrix, Santa Clara, CA, USA) to the multiplex-PCR products and incubating for 60min at 37^{O}C, following the inactivation of the enzyme by raising the temperature to 80^{O}C for 20min.

E. 2. 8. Single-base primer extension (SBPE)

The single-base primer extension (SBPE) reaction extends a fluorescently labeled dideoxynucleoside-triphosphate to a primer depending on the template (SNP of interest). The SBPE primers [rs12913832: F 5'-CTCTCTCTCTCTCTC GCG AGG CCA GTT TCA TTT GAG CAT TAA; rs16891982: F 5'-TCTCT GTG AGG AAA ACA CGG AGT TGA TGC A; rs1426654: F 5'-GGA TTG TCT CAG GAT GTT GCA GGC; rs1545397: F 5'-CTCTCTCTCTCTCTC GGA TAC TGA CAA TGG TTG TAC AAC TTT GTG AAT ATA CTA AAA TAC; rs885479: R 5'-CAG ATG GCC GCA ACG GCT; rs6119471: R 5'-TCTCTCTCTC CCC GAA GGA AGA GTG AAA ATG CGT AA; rs12203592: F 5'-TCTCTCTCTCTCTCTCTC GTT TCA TCC ACT TTG GTG GGT AAA AGA AGG; and rs12896399: F 5'-CTCTCTCTCTCTCTC CGA TCC AAT TCT TTG TTC TTT AGG TCA GTA TAT TTT GGG] were used at the end-concentration of 0.2μM, mixed with SNaPshot reagent (Applied Biosystems Inc, Foster City, CA, USA) and added to the cleaned multiplex-PCR products. The reaction was performed in 25 cycles of 10sec at 96^{O}C, 5sec at 57^{O}C, and 30sec at 60^{O}C. For the allelic ladder, which leads to two labeled products for each SNP - one for each allele - oligonucleotides [rs12913832: 5'-TGA CA**C** TTA ATG CTC AAA TGA AAC TGG CCT CGC, 5'-TGA CA**T** TTA ATG CTC AAA TGA AAC TGG CCT CGC; rs16891982: 5'-GGC TT**C** TGC ATC AAC TCC GTG TTT TCC TCA C, 5'-GGC TT**G** TGC ATC AAC TCC GTG TTT TCC TCA C; rs1426654: 5'-AGT TG**C** GCC TGC AAC ATC CTG AGA CAA TCC, 5'-AGT TG**T** GCC TGC AAC ATC CTG AGA CAA TCC; rs1545397: 5'-TTC AG**T** GTA TTT TAG TAT ATT CAC AAA GTT GTA CAA CCA TTG TCA GTA TCC, 5'-TTC AG**A** GTA TTT TAG TAT ATT CAC AAA GTT GTA CAA CCA TTG TCA GTA TCC; rs885479: 5'-GCG GC**G** AGC CGT TGC GGC CAT CTG, 5'-GCG GC**A** AGC CGT TGC GGC CAT CTG; rs6119471: 5'-CAG GT**C** TTA CGC ATT TTC ACT CTT CCT TCG GG, 5'-CAG GT**G** TTA CGC ATT TTC ACT CTT CCT TCG GG; rs12203592: 5'-AAT TT**G** CCT TCT TTT ACC CAC CAA AGT GGA TGA AAC, 5'-AAT TT**A** CCT TCT TTT ACC CAC CAA AGT GGA TGA AAC; and rs12896399: 5'-AGA GA**C** CCC AAA ATA TAC TGA CCT AAA GAA CAA AGA ATT GGA TCG, 5'-AGA GA**A** CCC AAA ATA TAC TGA CCT AAA GAA CAA AGA ATT GGA TCG] were used as template.

E. 2. 9. SBPE clean-up

Remaining fluorescently labeled ddNTPs of the SBPE reaction were dephosphorylated using Shrimp Alkaline Phosphatase (SAP) (USB Affymetrix, Santa Clara, CA, USA) for 60min at 37^{O}C. The enzyme was inactivated at 80^{O}C for 20min.

E. 2. 10. Multicolor capillary electrophoresis

Fluorescently labeled oligonucleotides were separated and detected on the 3130*xl* Genetic Analyzer (Applied Biosystems Inc, Foster City, CA, USA). Sample preparation: 1.1µl of the SBPE-clean-up product was added to 8µl deionized HIDI Formamide (Applied Biosystems) and 1µl of Gene-Scan 120 LIZ (Applied Biosystems), following denaturation at 95^{O}C for 5min. Samples were injected for 10sec at 2kV and electrophoresed at 15kV for 600sec in POP7, with a run temperature of 60^{O}C. Data were collected using the 3130*xl* Genetic Analyzer Data Collection Software v3.0.

E. 2. 11. Data analysis

The results from the electrophoresis were analyzed in GeneMapper 4.0. Allele peaks were interpreted when the relative fluorescence units (RFUs) were greater or equal to 100 for C and T, 200 for A, and 500 for G.

E. 2. 12. Models of DNA degradation

Models of DNA degradation included treatment with DNase I, boiling at 95^{O}C, and UV-radiation. Genomic DNA was treated using DNase I (Life Technologies, Invitrogen) at several concentrations (0.31-12.5 mU/µl) for 15min at 37^{O}C, based on the recommendations of the manufacturer. Genomic DNA was incubated at 95^{O}C in a thermal cycler (Applied Biosystems Inc, Foster City, CA, USA) for various times (5-120 min). Genomic DNA was exposed to UV-light (254nm) for different time intervals (referring to 120-9600 mJ) using the Stratalinker UV Crosslinker (Stratagene, La Jolla, CA).

E. 3. Results

The goal of this proposal was to establish a robust and easy-to-use assay to determine genotypes in forensic samples in order to predict pigmentation traits. However, at the time of writing the grant proposal (2010) for this award no predictor for pigmentation traits was published, only candidate SNPs were detected, but instructions on how to interpret the genotypes unambiguously were lacking. We developed a predictor that describes eye and skin color by utilizing seven SNPs, which

marked the beginning of this project. This led to the development of a robust and validated forensic assay. Just recently, the seven-plex system was improved by adding one SNP and by changing the instructions. This added the ability to predict blue eye and light skin color and increased significantly positive descriptions for eye color of being brown, green, or blue.

E. 3. 1. Prediction of Eye and Skin color using seven SNP (Spichenok, Budimlija et al. 2011)

Seven SNPs (rs12913832, rs1545397, rs16891982, rs1426654, rs885479, rs6119471, and rs12203592), located within or nearby genes that are involved in pigmentation (*HERC2*, *OCA2*, *SLC45A2*, *MC1R*, *ASIP*, and *IRF4*), were selected to be validated in various populations in order to make descriptions of individuals based solely on their DNA. Three of the SNPs (rs12913832, rs16891982, and rs12203592) were selected for their significant correlation with the blue or brown eye color (Cook et al 2009). Sequence comparison and analysis of the flanking regions of rs6119471 revealed that this region is conserved within primates and further investigation showed that this SNP is directly located in two predicted transcription factor [TF] binding sites, as shown in Figure 5. Both sites are specific for human TFs: E47 has a basic helix-loop-helix [bHLH] structure and forms homo- or hetero-dimers that bind to E-box motifs (CANNTG). Upon binding, gene expression can be activated or repressed. It is known that these types of gene regulators are involved in pigmentation (Hubbard, Uy et al. 2010). The other TF (HLF) has a basic leucine zipper (bZIP) structure and also forms hetero-dimers. These results support the importance of rs6119471 in pigmentation: according to the SNP data bank (dbSNP on NCBI), the G-allele is the ancestral, which correlates with darker skin color and might be recognized by one of the gene regulators. Five of the seven selected SNPs are potentially causative: three cause missense mutations and two SNPs are directly located in predicted TF binding sites. The other two SNPs are located in introns and it is still unclear whether they can be causative.

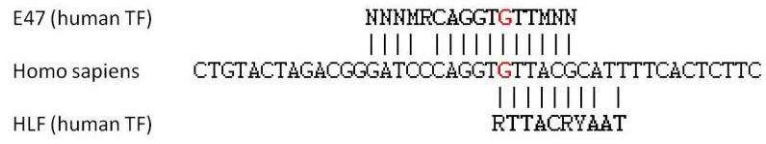

Figure 5: Predicted human transcription factor site at rs6119471 (red).

E. 3. 1. 1. Eye color prediction

Based on the correlation of the selected SNPs with eye pigmentation, the eye color predictor was created. The prediction relies mainly on rs12913832, in an exclusive approach: G/G-homozygotes were predicted to have non-brown eyes (the eye color could be in the green or blue range). A/G-heterozygotes and A/A-homozygotes were predicted to have non-blue eyes (the eye color could be in the green or brown range). For a more discriminating prediction, more SNPs were necessary. Six of the seven SNPs can be used to distinguish among the eye colors brown, green, not blue (green or brown), or not brown (green or blue). Brown eye color was predicted by the following genotypes: A/G or A/A at rs12913832, plus C/C at rs16891982, and/or T/T at rs1545397, and/or A/A at rs885479, and/or G/G at rs6119471. Green eye color was predicted by the following genotypes: G/G at rs12913832 plus C/C at rs16891982. The green eye color was also predicted by A/G at rs12913832 plus T/T at rs12203592. A/G at rs12913832 on its own is associated with non-blue eyes, while T/T at rs12203592 correlates with lighter eye colors. Brown or green eye color (i.e., not blue) was predicted by the combination of A/A at rs12913832 plus T/T at rs12203592.

The eye color predictor was applied to 554 independent samples from various populations including African-American, South Asian, East Asian, European, and mixed populations. Predicted eye color was then compared to the real eye color, which was binned into one of the three eye color bins (brown, green, and blue).

The outcome for the eye color predictions revealed that the error rate was very low. Out of 554 independent samples only four errors occurred. These errors happened within the European population and led to an error rate of ~1% (4/379). No error occurred within the other populations tested. Further, these results confirm that these markers can be used on all individuals independent of their association to a population. It is worth pointing out that this is true for the mixed population, which is the ultimate testing group since it represents the most diverse group and markers that were selected for their correlation with certain global populations should fail in this group. The exceptions would be markers that correlate well with pigmentation traits. For each sample an eye color could be assigned.

E. 3. 1. 2. Skin color prediction

The seven SNPs were used for the skin color prediction. Since pigmentation is a complex trait, at least two SNPs were used to predict lighter skin color. Similar to the eye color prediction, the skin coloration was predicted in an exclusive and conservative approach. Light or medium skin color (i.e., non-dark skin color) was predicted by the presence of any two of the following alleles: G/G (rs12913832), T/T (rs1545397), G/G (rs16891982), A/A (rs1426654), A/A (rs885479), T/T (rs12203592). Medium or dark skin color (i.e., non-white skin color) was predicted by G/G genotype at rs6119471. Non-white skin color was also predicted when G/G at rs6119471 occurred together with any one of the six alleles for "non-dark skin color".

This skin color predictor was applied to independent samples from various populations (n=554) followed by the comparison of the predicted skin color with the skin color bin. The real skin color was also grouped into three bins: light, medium and dark.

The outcome showed that the error-rate was very low: only two errors occurred in 398 predictions. No errors were found in the predictions of the skin color for most populations, revealing that these SNPs are good markers for pigment-related features. However, for several samples the data were insufficient to predict skin coloration: of the 554 samples tested, 398 predictions could be made (72%). A possible explanation for that could be the conservative approach (only homozygous alleles were used for the prediction) and the selection of the seven SNPs; four of them were associated with the European population, two with East Asian and one with the African-American population. Consequently, the calling rate for European samples was very high (91%), while it was below 50% for other populations (including the mixed population). Nevertheless, when calls could be made based on the presence of any of the discriminating genotype combinations, the error-rate was low.

E. 3. 2. Verification of the Eye and Skin Color Predictor (Pneuman, Budimlija et al. 2012)

Validation of testing methods is an essential feature in all scientific endeavors, but it is particularly important in forensics. Due to the sensitive nature of these investigations and the limited evidence it is crucial to validate all employed procedures. This includes the implementation of novel forensic phenotypic DNA tests, about which information must be learned before being incorporated as routine methods. Ideally, validations are performed on large sample sets that mimic real cases.

So far, two phenotypic predictors, the eye- and skin-color (7-plex) and another eye color predictor ("IrisPlex") have been published (Spichenok, Budimlija et al. 2011; Walsh, Liu et al. 2011). These predictors are well-defined by a selection of single nucleotide polymorphisms (SNPs) and unambiguous instructions on how to interpret the genotypes. Standardized approaches have the advantages of being applied in diverse laboratories, leading to the same outcome and offering the opportunity for validation. In order for these tests to be used to characterize human remains, they should be validated on various populations to perform reliably without prior knowledge of ethnic origin.

To validate these predictors, over 250 new samples were collected. This allowed validating the "IrisPlex" predictor on over 800 and the eye- and skin-color predictor (7-plex-system) on over 250 samples.

E. 3. 2. 1. Verification of the eye- and skin-color predictor based on the "7-plex-system"

The eye color prediction occurs in a two step procedure utilizing six SNPs. First, the genotype at rs12913832 is used to determine the eye color as being not blue (A/A or G/A) or not brown (G/G). In the second step, eye color is further refined by homozygote genotypes of one or more of the remaining five SNPs, which are used for predicting eye color as being brown or green. The eye color predictor was verified on an additional 251 samples.

Seven errors were found out of 251 eye color predictions, of which five occurred within the European and two within the mixed population. The error-rate for this validation was calculated to be approximately 3%. Taken together, out of the 805 eye color predictions, 11 errors occurred, of which 10 arose from rs12913832. Nine individuals with blue eyes were predicted to have non-blue eyes, and one with brown eyes was predicted to have non-brown eyes. The reason of the eleventh error has yet to be determined. For this person with green eyes the genotypes at rs16891982 and rs6119471 led to the prediction of brown eyes.

Originally, the skin color predictor was tested on 554 samples from unrelated individuals of various populations (Spichenok, Budimlija et al. 2011). Additional, 251 samples were used to further validate this predictor. The skin color predictor utilizes seven SNPs, of which six are used to predict lighter shades and one for darker shades. The outcome of the prediction was either having non-dark (medium or light) or having non-light (medium or dark) skin color.

Testing the 251 samples led to 203 predictions, out of which 2 errors occurred. These two errors plus the two from the first sample set (n=554) add up to four errors that were found in 602 predictions of 805 samples. All four errors occurred for dark-skinned persons that were predicted of being non-dark; they were either associated with mixed populations (n=3) or South Asian (n=1).

Taken together, the error-rate for this predictor is very low (1%). The call rate varied broadly between populations (ranging from 15 to 95%) and was highest for European descendants, confirming the original study (Spichenok, Budimlija et al. 2011).

E. 3. 2. 2. Validation of the eye color predictor "IrisPlex"

The other eye-color predictor that was published is named "IrisPlex". "IrisPlex" utilizes six SNPs, of which three are shared with the 7-plex-system, and uses a multinomial logistic regression model to calculate probabilities for the eye colors brown, blue and intermediate/green, for each sample. An eye color that has a calculated probability greater than 0.7, which is the recommended threshold, is the predicted eye color (Walsh, Liu et al. 2011; Walsh, Wollstein et al. 2011). To validate this eye color predictor it was tested on 803 samples of various populations. The genotypes of the six SNPs [rs12913832 (*HERC2*), rs1800407 (*OCA2*), rs12896399 (*SLC24A4*), rs16891982 (*SLC45A2*), rs1393350 (*TYR*), and rs12203592 (*IRF4*)] were determined and used as input for the *Excel-based* macro, which computes the probabilities for each of the three eye colors (Walsh, Liu et al. 2011; Walsh, Wollstein et al. 2011).

Population	Eye color	Eye color predicted	n ≥0.7	Error	Inclusive ≥0.7
AA (43)	brown (43)	brown	38	0	5
SA (27)	brown (27)	brown	27	0	0
EA (35)	brown (35)	brown	34	0	1
E (555)	blue (191)	blue	149	3	39
		intermediate	0		
		brown	3		
	green (175)	blue	71	116	59
		intermediate	0		
		brown	45		
	brown (189)	blue	1	1	72
		intermediate	0		
		brown	116		
Mixed (143)	blue (11)	blue	8	0	3
		intermediate	0		
		brown	0		
	green (20)	blue	9	14	6
		intermediate	0		
		brown	5		
	brown (112)	blue	1	1	24
		intermediate	0		
		brown	87		
Total	803		594	135	209
%				31%*	26%

Table 1: Verification of "IrisPlex" on 803 samples

Table 1 shows the outcome of this validation. This predictor can be applied on various populations. Using a threshold can cause inconclusive results; this is when none of the calculated probabilities reach the threshold. When using the threshold of ≥0.7 predicted eye color probability, 209 (26%) results were inconclusive out of 803 samples tested. This corroborates the validation of the "IrisPlex"-System on European samples, where the inconclusive-rate was 22.50% (Walsh, Wollstein et al. 2011). Of 594 eye color predictions from the 803 samples tested 135 errors were found. The error-rate was calculated to approximately 31% (120 errors occurred in 385 predictions).

E. 3. 3. Improved eye-and skin-color prediction based on 8 SNPs (Hart, Kimura et al. 2013)

Analysis of an eighth SNP, rs12896399 (*SLC24A4*), showed a statistically significant association with human eye color (p=0.007) but a rather poor strength of agreement (=0.063). Sequence analysis of the flanking regions reveals that this section is highly conserved within primates. The sequence around rs12896399 could contain regulatory regions. In order to improve the 7-plex system this SNP was added and the instruction guidelines on the interpretation of genotypes were changed. The newly developed 8-plex system has the additional ability to predict blue eye and light skin color.

The prediction for eye color is a two-step procedure: First, rs12913832 (*HERC2*) distinguishes eye colors as being not blue (i.e. brown or green) or not brown (i.e. green or blue). Further distinctions for predicting eye color as being brown, green or blue are dependent on the genotypes of three additional SNPs: rs12203592, rs16891982, and rs6119471 (Fig. 3). Brown eye color is predicted by the following genotype combinations: A/A or G/A at rs12913832, plus either G/G at rs6119471, or C/C at rs16891982; green eye color is predicted by G/G at rs12913832 plus C/C at rs16891982, or by G/A at rs12913832 plus T/T at rs12203592; and blue eye color is predicted by G/G at rs12913832 plus T/T at rs12203592. The second step proceeds with samples that were not positively described as being brown, green or blue in the first step (Fig. 3). Samples that are homozygous for G/G at rs12913832 (i.e. not brown) plus T/T at rs12896399 are then predicted to have blue eyes, while samples that are homozygous for A/A at rs12913832 (i.e. not blue) plus G/G at rs12896399 are predicted to have brown eyes. This step utilizes the newly-included SNP rs12896399 (*SLC24A4*) and leads to increased numbers of positively described eye colors.

The newly-designed 8-plex system was established on 803 samples (training set). Thirteen errors were found within the first step of eye-color prediction; eleven of them were in the European population. Ten of the thirteen errors arose from rs12913832. The error rate for the first step was calculated at 2%, which is comparable to the 7-plex system (3%). The second step of the eye color prediction increased the number of positive descriptions by 97 samples (151+97=248). This increase equals approximately 12%. Of these predictions, twelve were erroneous. These errors occurred for green-eyed individuals that were predicted to have either brown or blue eye color. The error rate for the second step was also calculated at 2% for the European population.

212 samples were collected to test the newly-developed 8-plex system (test set). The first step led to 56 positive descriptions, in which 2 errors occurred (error rate of 1.75%). The second step, included rs12896399, led to 30 further positive descriptions (56+30=86). Three errors were found in the second step (error rate equals 2.73%). Five errors were counted: two arose from rs12913832 and three from rs12896399. All five errors in the test set (n=212) of the 8-plex system occurred in the European population and led to an error rate of approximately 5% (4.54%).

Skin-color prediction of the 8-plex system utilizes six SNPs, five to predict a lighter skin color and one to predict darker shades. Skin coloration is predicted in an elimination process. Non-dark skin color (i.e. light or medium) is predicted by any two of the following alleles: G/G at rs12913832, G/G at rs16891982, A/A at rs1426654, T/T at rs1545397, or A/A at rs885479. Light skin color is predicted by more stringent conditions: G/G at rs12913832, plus G/G at rs16891982, and A/A at rs1426654. Non-light skin color (i.e. medium or dark) is predicted by G/G at rs6119471.

Of 803 training samples, 600 (75%) predictions were made, resulting in only 4 errors (1%). The four erroneous predictions described skin color as being non-dark for dark-skinned individuals. The positive predictions for light skin coloration were all correct (255 of 600 predictions). The test set of 212 additional samples led to 131 (62%) predictions. Of these, 63 were predicted to have light skin color and these were correct.

Taken together, the outcomes from training and test sets are in concordance. The newly-developed eye- and skin-color predictor utilizing 8 SNPs is an improved version of the 7-plex system, leading to more precise descriptions including blue-eye and light-skin coloration. In addition, the new 8-

plex system increased the number of positive descriptions significantly (p=6.98e-15, z-score: -7.786) while keeping the error rate low.

E. 3. 4. Development of multiplex-SNP-assay (Mushailov, Rodriguez et al.)

The goal of this proposal was to establish a robust and easy-to-use assay to determine genotypes of forensic samples in order to predict eye and skin color. The multiplex-SNP-assay (eight SNPs) was chosen because it makes efficient use of the nuclear DNA template and is inexpensive. Furthermore, it is possible to extend this test in the future, depending on the detection of novel candidate SNPs. Key elements of this multiplex-SNP-assay include a multiplex-PCR followed by a single-base primer extension, which creates fluorescently-labeled oligonucleotides of distinct lengths. These are detected by multicolor capillary electrophoresis. The assay has the possibility to be applied on severely fragmented templates since it relies on short amplicons and therefore is well suited for forensic samples.

E. 3. 4. 1. Development of multiplex-PCR

The eight SNPs, rs12913832, rs16891982, rs1545397, rs1426654, rs885479, rs6119471, rs12203592, and rs12896399, were incorporated into the multiplex-SNP-assay. The assay, a two-step procedure includes a multiplex-PCR followed by a multiplex-primer extension reaction, creating fluorescently labeled oligonucleotides of distinct lengths, which are then detected by multicolor capillary electrophoresis.

For the first step, multiplex-PCR, primers were designed to flank the SNPs, each of which is the target of a reaction. Since all primers must work under similar conditions, the theoretical melting temperature was set to $60^{O}C \pm 3^{O}C$ and their GC content was kept between 35 - 60%. Sequences with homopolymers containing more than four bases were excluded, and amplicon size was kept small, around 100bp, with a range of 80 – 200bp. When no appropriate primer-pair was found, product size was increased to 500bp. All primers were tested for alignment to undesired sites in the human genome, as well as for primer-dimer and hairpin formations. For each SNP several primers leading to numerous PCR-products were designed.

The various primer combinations were tested in singleplex PCR reactions, with the aim to find at least one PCR-product for each SNP that showed the proper product length without extraneous bands. The reaction conditions were kept the same. Singleplex PCR reactions that showed the desired outcome, regarding size and amount, were combined into multiplex reactions.

Initially, the multiplex reactions were not showing the theoretically expected results. The PCR products, confirmed by their size, varied greatly in their amount and some of them even vanished when combined. The outcome of the PCR reactions was verified using either agarose gels or the Agilent Bioanalyzer. Since the resolution of the Bioanalyzer was better than that of the agarose gels, it was mainly used to verify multiplex PCRs. However, amplicons of similar sizes (e.g. 80bp and 87bp) could not be separated. Only PCR reactions that were dissimilar in their product lengths were combined. This procedure allowed for the detection of each PCR product in the electropherogram of the Bioanalyzer (Fig. 6).

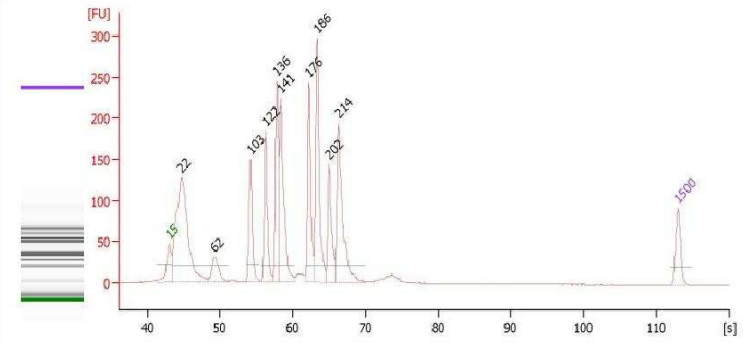

Figure 6: Electropherogram of 8-plex PCR (Bioanalyzer)

E. 3. 4. 2. Multiplex-SBPE

Single base primer extension [SBPE] reactions were also to be performed as a multiplex reaction; therefore all reactions must work under the same conditions. The SBPE-primers must be specific and should have similar theoretical melting temperatures, which was set to 60^{O}C. Since the SBPE-primers target a SNP, their length resulted from the melting temperature. The size of the SBPE-primers for the selected SNPs ranged from 10bp to 60bp for both forward and reverse

primers. Primers were checked for multiple binding sites. After sorting the primers according to their length, some were extended with CT-tails to result in products that differ in their length by 6pb to avoid crowding of peaks and overshooting fluorescent signals in electropherograms of the multicolor capillary electrophoresis. Diverse primer combinations led to nine SBPE scenarios.

The selected SBPE-primers were assessed using the Primer Focus kit. This kit allows rapid evaluation of the selected primers by extending all potential labeled nucleotides and calculating the mobility rate for the resulting allele. Therefore, this kit could help to determine the best combination of SBPE-primers for multiplexing. The expected outcome of this kit was four peaks, one for each nucleotide visible in the electropherogram of the multicolor capillary electrophoresis [CE]. With the help of this kit five of the eight SNPs were resolved. The kit however, revealed the mobility of the labeled SBPE-products: primers extended with a T that was labeled with dROX (red) appeared as the longest product, followed by C and A, which were labeled with dTAMARA (black) and dR6G (green), respectively. The shortest product resulted from the extension of G (blue) that was labeled with dR110.

These SBPE scenarios should ideally be tested on a sample that is heterozygous for all SNPs (allelic ladder), resulting in two peaks per SNP. Since such a sample did not exist in our sample collection of over 800 samples, the allelic ladder, used oligonucleotides, one for each allele as template (Material and Methods).

The nine SBPE-scenarios were then tested with the generated allelic ladder, a sample of a known genotype, and a non-template control [NTC]. Eight of the nine SBPE-scenarios showed additional peaks. These peaks were also found in their according NTCs. Further analysis revealed that two of the SBPE-primers could form primer-dimers under the reaction conditions, and were extended with a fluorescently-labeled nucleotide. The products showed corresponding G- and T-peaks. A third primer was able to form a hairpin that was extended and led to a G-peak.

One of the nine SBPE-scenarios showed the expected results: two peaks per SNP for the provisional allelic ladder, a full profile for the known sample, and no additional peak in the NTC. This scenario was used for further validation of the assay.

E. 3. 5. Validation of multiplex-SNP-assay (Mushailov, Rodriguez et al.)

The multiplex-SNP-assay consists of a multiplex-PCR that amplifies eight regions. The amplification is checked on the Agilent Bioanalyzer. The PCR products are purified by using the ExoSAP-IT kit to prevent remaining primers or dNTPs from interfering with the following step. The SBPE reaction is performed by using the SNaPshot Multiplex kit. The SBPE product is treated with Shrimp Alkaline Phosphatase to dephosphorylate the fluorescently labeled dNTPs. This prevents further reactions of the dNTPs that could lead to extra-peaks in the electropherogram of the multicolor capillary electrophoresis. The products of the multiplex-SNP-assay are separated by color and size and detected using the 3130xl multicolor capillary electrophoresis. Data analysis is performed with GeneMapper.

The validation was performed according to the guidelines of the OCME Department of Forensic Biology Quality Assurance/Quality Control Manual and of the Scientific Working Group on DNA Analysis Methods [SWGDAM]; it demonstrates the accuracy, reliability, and reproducibility of the procedure ((SWGDAM) 2004).

E. 3. 5. 1. Reliability of multiplex-SNP-assay

The outcome of the assay was confirmed by other methods: Ten samples, taken from various populations, including African-American, South-Asian, East-Asian, European and mixed populations, were tested by two methods: multiplex-SNP-assay and corresponding *TaqMan* single-plex assays.

The outcomes for both methods were the same, showing that the assay is reliable and reproducible.

E. 3. 5. 2. Test of various DNA extractions

Various DNA extraction methods, including chelex extraction for blood and buccal samples, differential chelex extraction for epithelial and sperm cells, and a biorobot (M48) that uses magnetic resins for blood stains on FTA paper, that are used at the OCME for casework were tested on the multiplex-SNP-assay. Each extraction method was tested on three samples. All samples led to full profiles that could be used to predict eye and skin color. The prediction could not be verified, since the samples were anonymized. This shows that the multiplex-SNP-assay can be used on various extractions, including the methods used for casework.

E. 3. 5. 3. Validation of multiplex-PCR conditions

The robustness of the assay was tested by using various PCR machines (which may differ in their ramping time) and various multicolor CEs. The various machines had no impact on the outcome of the assay.

Further, PCR conditions for the multiplex-PCR were varied by their $MgCl_2$ concentrations, cycle number, and PCR volume. These conditions were tested on at least three samples of known genotype, three times independently. All these variations had no impact on the outcome of the multiplex-SNP-assay. All results were expected and correct, showing that the multiplex-SNP-assay is robust and performs reliably.

E. 3. 5. 4. Species-specificity of multiplex-SNP-assay

DNA from various species was tested with the multiplex-SNP-assay. The species included were baboon, cat, ferret, dog, lamb, chicken, rabbit, deer, sheep, horse, pony, squirrel, fish, bull, and yeast. The multiplex-PCR revealed the amplification for some PCR products from baboon, ferret, dog, chicken, deer, and yeast. Of these PCR-products only a few had the expected sizes. These were from baboon and ferret. However, none of the species showed a full profile. Therefore, the multiplex-SNP-assay is specific for humans.

E. 3. 5. 5. Sensitivity of multiplex-SNP-assay

Testing the sensitivity defines the range of DNA concentrations required to ensure consistent results. The sensitivity was tested with four known samples, which together were heterozygous at each location. This ensured that each SNP was tested by a heterozygous sample. Test-runs were performed three times independently. Correct and full profiles were received in all cases (100%) by concentrations higher than 100pg DNA.

E. 3. 5. 6. Test of degraded material

DNA was extracted from blood cards (archival waste tissues samples, after obtaining internal OCME approval in compliance with OCME policy manual, issued January 1, 2010). From 13 extractions 12 full profiles were obtained. One of the extractions that showed no result for the multiplex-SNP-assay had a concentration of 0.7pg/ µl and was therefore too low to expect a result.

E. 4. Discussion and Conclusion

The objective of this proposed project was to optimize and validate a multiplex assay based on nuclear DNA that has the potential to assist in the identification of unidentified human remains by predicting pigmentation traits. The proposed method is fast, robust, and, importantly, can be applied to degraded DNA. Available publications and reports on identified genetic markers associated with human eye, hair and skin pigmentation were utilized to select potential markers (Lao, de Gruijter et al. 2007; Lin and Fisher 2007; Parra 2007; Tully 2007; Brilliant 2008; Eiberg, Troelsen et al. 2008; Sturm, Duffy et al. 2008; Liu, van Duijn et al. 2009; Sturm 2009; Valenzuela, Henderson et al. 2010). Eight markers were selected to describe the eye and skin color of an individual (Hart, Kimura et al. 2013). These markers were incorporated into a multiplex-SNP-assay in order to create a standardized forensic test (Pneuman, Budimlija et al. 2012). The utilization of DNA analysis to predict visible pigment-related features of unidentified human remains will enhance current efforts to gather obtainable identifying information from a decedent.

The steps involved in this project included: 1. the selection of useful and reliable SNP-candidates that could be incorporated into the forensic multiplex-assay; 2. definition of instructions on how to interpret the genotype, in order to make the forensic multiplex-assay independent of the interpreting analyst or the performing laboratory; 3. validation of the phenotypic predictor to obtain call and error rates; 4. development of the multiplex-SNP-assay to have a tool that allows for genotyping; and 5. validation of the multiplex-SNP-assay to define accuracy, reliability and limitations of this assay.

E. 4. 1. Discussion of Findings and Procedure

Eight SNP, located in or nearby pigmentation genes were selected for their expected correlation with pigment-related features. Four (rs12913832, rs16891982, rs12203592, and rs12896399) were selected for their correlation with the human blue or brown eye coloration (Eiberg, Troelsen et al. 2008; Sturm, Duffy et al. 2008; Liu, van Duijn et al. 2009; Walsh, Liu et al. 2011; Hart, Kimura et al. 2013), and three (rs12913832, rs16891982, and rs1426654) were selected because they led to reduced melanin content in cultured human melanocytes (Cook, Chen et al. 2009). One of the selected SNPs (rs6119471) was chosen because of its direct location in a predicted human TF binding side in the 5' region of *ASIP* (Spichenok, Budimlija et al. 2011). The seventh and eighth SNPs (rs1545397 and rs885479) were selected for their location in pigmentation genes (*OCA2* and *MC1R*) and their correlation with certain

global populations (Duffy, Box et al. 2004; Ringholm, Klovins et al. 2004; Yuasa, Umetsu et al. 2007). Six of the eight SNPs are potentially causative: three cause missense mutations, two are directly located in predicted TF binding sites, and one in a potentially gene regulatory region. The other two are located in introns and it is unclear whether they are causative. Of the eight SNPs, five were also identified by other groups studying pigmentation (Liu, van Duijn et al. 2009; Mengel-From, Borsting et al. 2010; Valenzuela, Henderson et al. 2010; Walsh, Liu et al. 2011).

In contrast to correlation studies, which detect candidate SNPs based on their significant correlation with pigmentation characteristics, a predictor also includes unambiguous information on how to interpret the genotype of the selected SNPs. Therefore, a predictor has the potential to be independent of the interpreting analyst or the performing laboratory (Pneuman, Budimlija et al. 2012). To select the best available predictor in pigmentation, the "IrisPlex" eye color predictor (Walsh, Liu et al. 2011) was validated on over 800 samples.

The validation showed that the eye and skin color predictor based on the 7-plex as well as the 8-plex-system, and "IrisPlex", can be used in all populations tested including African-American, South Asian (dark), East Asian (light), European and mixed populations. These are important results, particularly for forensic applications since it is possible that the evidence does not reveal visible characteristics. These results confirm previous conclusions for "IrisPlex", based on testing HGDP-CEPH Human Genome Diversity Cell Line Panel (Rosenberg, Pritchard et al. 2002; Rosenberg 2006) of 51 populations which do not provide phenotypic data (Walsh, Liu et al. 2011). The eye color predictor based on the 8-plex-system distinguishes between brown, green, blue, not blue (green or brown) and not brown (blue or green) eye colors (Hart, Kimura et al. 2013). Predictions can be made for all samples (call rate is 100%). The error rate was calculated to be approximately 5% (Hart, Kimura et al. 2013). Further error-assessment pointed to rs12913832 (*HERC2*), which is the primary SNP on which the eye color prediction relies, and rs12896399, the SNP that is used in the second step of eye-color prediction. The skin color predictor showed some inconclusive outcomes, since this predictor utilizes only homozygous genotypes for prediction. The rate of inconclusive outcomes varied broadly between populations, ranging from 5 to 85%, and was lowest for European descendants. This is in agreement with the original study (Spichenok, Budimlija et al. 2011). The reason for this variation may be due to the selection of SNPs; further research may improve the predictor. No errors were found in 212 samples, the test set (Hart, Kimura et al. 2013).

The eye color predictor "IrisPlex" uses thresholds, and therefore leads to inconclusive results. Inconclusive results occurred when none of the calculated probabilities for each eye color (blue, brown and intermediate) reached the threshold. The higher the threshold is set, the higher the rate of inconclusive results (Walsh, Wollstein et al. 2011). When the threshold was set to $\geq$0.7, as recommended by the authors, the rate of inconclusive results was approximately 26%. A comparable but lower rate of inconclusive results (22.5%) was found by the validation recently performed on 3,840 samples of individuals from seven countries across Europe (Walsh, Wollstein et al. 2011). The specific eye color prediction, of blue, brown or intermediate by the "IrisPlex" predictor (Walsh, Liu et al. 2011; Walsh, Wollstein et al. 2011) is more precise but appears less accurate. Most errors of this predictor were related to the intermediate (neither brown nor blue) eye color, a problem which has been acknowledged by the authors and may be elucidated with further research (Walsh, Liu et al. 2011; Walsh, Wollstein et al. 2011). The higher the numbers of individuals with intermediate/green eye colors in a given population, the higher the error-rate. In the validation study 31.5% of the European samples were grouped into the green eye-color bin, which led to an error-rate of 31%, while a different recent study that grouped 7% of the European samples into the intermediate/green eye-color bin, calculated an error-rate of 6.1% (Walsh, Wollstein et al. 2011). By disregarding the intermediate/green eye color, the error-rate was low (approximately 1.5%).

Based on these outcomes, we decided to proceed with 8-plex-system. Eight SNPs were incorporated into a multiplex-SNP-assay. Following its development the assay was validated according to the SWGDAMs guidelines.

The goal of this proposal was to establish a method that is robust and easy-to-use in order to determine genotypes of forensic samples to predict eye and skin colors of individuals. The method of choice was a multiplex-SNP-assay for its efficient use of nuclear DNA template and its inexpensiveness. Key elements include a multiplex-PCR followed by a multiplex single base primer extension that creates fluorescently labeled oligonucleotides of distinct length that are detectable by multicolor capillary electrophoresis [CE]. All necessary equipment was already available at the OCME. Further, this method can easily be implemented in other forensic laboratories, since they use similar equipment for casework. The method is well-known; it was used for comparable tasks, such as describing eye color ("IrisPlex"), or phenotypes of ancient skeletal remains, or for human identification (Sanchez and Endicott 2006;

Bouakaze, Keyser et al. 2009; Walsh, Liu et al. 2011). The method could, if required, be expanded up to 52 SNPs (Sanchez and Endicott 2006) as well.

Subsequent to the development of the multiplex-SNP-assay was its validation according to the guidelines of the OCME Department of Forensic Biology Quality Assurance/ Quality Control Manual and the SWGDAM recommendations ((SWGDAM) 2004). This could demonstrate its accuracy, reliability and reproducibility: The outcome of the multiplex-SNP-assay is reproducible and reliable, which was confirmed by *TaqMan*-QPCR singleplex-assays. The multiplex-SNP-assay is robust and not inhibited by any of the tested DNA extraction methods used at the OCME for casework; this included year-old blood cards. The robustness of the assay was further supported by changing several multiplex-PCR reaction conditions, which led to a full and correct profile. In addition, it was shown that the assay is specific for humans since none of the DNA of tested species, including baboons, resulted in a full profile. These outcomes are in accordance with the goal of this project: to establish a robust and easy-to-use assay to determine genotypes of forensic samples in order to predict pigmentation traits.

E. 4. 2. Future Prospects

Should novel markers be identified to predict visible characteristics of an individual, they can be evaluated on over 800 samples. 250 of them were collected during the funded period of this project. If novel candidate markers would improve the current validated multiple-SNP-assay, they could be integrated and would lead to an updated version of this assay. The upgrading procedure would be cost-effective, because no changes in equipment and software are required. Upgrades could be implemented quickly and would require little, if any, additional training time for analysts. These advantages demonstrate the flexibility of the multiple-SNP-assay. Meanwhile, “IrisPlex” was further developed into “HIrisPlex”, which utilizes 24 markers to predict eye and hair color of individuals (Walsh, Liu et al. 2013). It would be interesting to validate this system and determine call and error rates. In the future, it may be conceivable to combine these two predicting system, the 8-plex system and the “HIrisPlex sytem”, to describe all pigmentation traits (eye, hair, and skin coloration) using just one assay.

Companies are interested in the outcome of these studies. Identitas Inc. is a phenotypic characterization company providing DNA based assays amongst others to test for physical appearance including eye, hair and skin coloration. Recently, the Identitas v1 Forensic Chip was described (Keating,

Bansal et al. 2013). Illumina is also working on a forensic platform that will include phenotypic characteristics.

E. 4. 3. Implications for Policy and Practice

The number of missing persons and unidentified human remains is what experts call a "mass disaster over time" (Ritter 2007). According to the 2004 Bureau of Justice Statistics Census of Medical Examiner's and Coroner's Offices, nearly 13,500 decedents remain unidentified nationwide (Hickman 2007). The Bureau of Justice Statistics further found that approximately 4,400 unidentified human remains are reported each year, with roughly 1,000 decedents still unidentified after one year. New York ranks as the second highest state in the nation of reported unidentified human remains (as reported between 2000-2004), and accounts for nearly 25% of all unidentified cases nationwide (Hughes 2007).

In recognition of the number of unidentified decedents in New York City, the OCME has initiated a multi-disciplinary approach to enhance the probability of identification. With funding awarded by the National Institute of Justice (#2008-DN-BX-K156 and #2009-DN-BX-K038), the OCME has been able to assemble a uniquely qualified team of forensic anthropologists and DNA scientists. The DNA Missing Person Group performs nuclear and mitochondrial DNA testing on postmortem samples, as well as on familial reference samples, and uploads results into CODIS 6.0. OCME's Forensic Anthropology Unit analyzes all decomposed, disassembled or buried remains discovered within the City of New York. OCME's anthropologists perform critical skeletal analyses to assist in the identification of a decedent, and attempt to determine the estimated age, race, sex, and stature of human remains.

The ability to utilize DNA analysis to predict visible pigment-related features of unidentified human remains has the potential to be incorporated into this multidisciplinary effort to identify the missing and unidentified human remains (Budowle and van Daal 2008). Prediction of visible traits of skeletal remains, such as eye and skin colors, by using the developed multiplex-SNP-assay of this project are considered enormously helpful; the prediction could be entered into NamUs and the National Crime Index System.

F. Dissemination

Outcomes and results from this project have been or will be published and presented at scientific meetings with specific emphasis on medical examiner/coroner offices, as well as law enforcement agencies who could benefit from it.

In more detail, one publication introduced the selection of seven SNPs to predict eye and skin color of individuals: Spichenok et al. (2011) *FSI: Genetics*: In November 2010, one month after we received funding for this project, our publication of "Prediction of eye and skin color in diverse populations using seven SNPs" became electronically available. In this study, seven single nucleotide polymorphisms (SNPs), located in and nearby pigmentation genes, were tested on 554 samples from non-related individuals of various populations. Six SNPs were used to predict eye color, and all seven to describe skin coloration of individuals. Results showed that these markers can be applied to all populations with very low error rates. However, the call-rate to determine the skin coloration varied between populations demonstrating the complexity of this trait. Overall, these results demonstrate the importance of these seven SNPs, which is useful when establishing a forensic pigmentation test.

This publication was included in the top 25 "hottest articles," ranking at number 5 of all publications in 2011 in *Forensic Science International: Genetics*.

Our second publication of this project was in part supported by this award from NIJ (No. 2010-DN-BX-K181): Pneuman et al. (2012) *Legal Medicine:* "Verification of eye and skin color predictors in various populations." Recently, three phenotypic predictors, the eye and skin color predictor and another eye color predictor, have been published (Spichenok *et al.* 2011, Walsh *et al.* 2011). It could be confirmed that all predictors can be used in various and mixed populations. The error and call rates, however, varied. The error-rate was low (3%) for the predictors that describe the eye and skin color exclusively (non-brown or non-blue and non-white or non-dark, respectively) and higher (31%) for the predictor that describes individual eye colors (blue, brown, and intermediate/green) because of uncertainties with the green eye color prediction. The call rate for the seven-plex system to predict the eye color was 100% and for "IrisPlex" was 74%. The error rate for the skin color predictor was low (1%) and the call rated varied between populations (5 to 85%).

A third manuscript also supported in part from NIJ (No. 2010-DN-BX-K181), by Hart et al. (2013) was just recently accepted for publication in the *Croatian Medical Journal*. The title is "Improved eye- and skin-color prediction based on 8 SNPs." Analysis of an eighth SNP, rs12896399 (*SLC24A4*), showed a statistically significant association with human eye color (p=0.007) but a rather poor strength of agreement (₸0.063). This SNP was added to the 7-plex system (rs12913832 at *HERC2*, rs1545397 at *OCA2*, rs16891982 at *SLC45A2*, rs1426654 at *SLC24A5*, rs885479 at *MC1R*, rs6119471 at *ASIP*, and rs12203592 at *IRF4*). Further, the instruction guidelines on the interpretation of genotypes were changed to create a new 8-plex system. This was based on the analysis of an 803-sample training set from various populations. The newly developed 8-plex system can predict the eye colors brown, green and blue, and skin colors light, not dark, and not light. It is superior to the 7-plex system with its additional ability to predict blue eye and light skin color. The 8-plex system was tested on an additional 212 samples, the test set. Analysis showed that the number of positive descriptions for eye colors as being brown, green or blue increased significantly (p=6.98e-15, z-score: -7.786). The error rate for eye-color prediction is low, at approximately 5%, while the skin color prediction showed no error in the test set (1% in training set).

The manuscript "Development and validation of the 8-plex assay" is in preparation to be submitted soon. This study describes the development of a multiplex-SNP-assay for these eight SNPs and its validation. The 8-plex assay was tested of its reproducibility, reliability, and sensitivity. Further validation included tests with different models for DNA degradation, and species specificity. It can be concluded that the assay based on the 8-plex system is robust and can be used on challenging samples in forensics to determine reliably eye and skin color of individuals.

Furthermore, this project was introduced at three forensic meetings: A presentation, entitled "Validation of SNPs to Predict Pigment-Related Features in Diverse Populations," was given by Dr. Elisa Wurmbach at the American Academy of Forensic Science at their 63rd annual meeting held on February 21-26, 2011 in Chicago, IL. A second presentation followed on June 20-22, 2011 at the NIJ conference in Arlington, VA, entitled "Prediction of eye and skin color in diverse populations using seven SNPs." Dr. Zoran Budimlija, Co-PI on this project, presented at an international Workshop on DNA typing of Bone Samples in Prague, Czech Republic on October 3-6, 2012. This presentation was entitled "Novel forensic phenotypic DNA tests: eye and skin color predictors."

At the International Society for Applied Biological Science (ISABS) conference in Split (Croatia), on June 24-28, 2013, Elisa Wurmbach, invited speaker, will present “Improved eye- and skin-color prediction based on 8 SNPs.”

G. References

(SWGDAM), S. W. G. o. D. A. M. (2004). "Revised Validation Guidelines." Forensic Science Communications 6(3).

Barsh, G. S. (2003). "What controls variation in human skin color?" PLoS Biol 1(1): E27.

Bouakaze, C., C. Keyser, et al. (2009). "Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis." Int J Legal Med 123(4): 315-325.

Branicki, W., U. Brudnik, et al. (2008). "Association of polymorphic sites in the OCA2 gene with eye colour using the tree scanning method." Ann Hum Genet 72(Pt 2): 184-192.

Branicki, W., U. Brudnik, et al. (2007). "Determination of phenotype associated SNPs in the MC1R gene." J Forensic Sci 52(2): 349-354.

Branicki, W., U. Brudnik, et al. (2009). "Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype." Ann Hum Genet 73(2): 160-170.

Branicki, W., F. Liu, et al. (2011). "Model-based prediction of human hair color using DNA variants." Hum Genet 129(4): 443-454.

Brilliant, M. H. (2008). Gene Polymorphism and Human Pigmentation, NIJ.

Brookes, A. J. (1999). "The essence of SNPs." Gene 234(2): 177-186.

Budowle, B. and A. van Daal (2008). "Forensically relevant SNP classes." Biotechniques 44(5): 603-608, 610.

Caragine, T., R. Mikulasovich, et al. (2009). "Validation of testing and interpretation protocols for low template DNA samples using AmpFlSTR Identifiler." Croat Med J 50(3): 250-267.

Chaplin, G. (2004). "Geographic distribution of environmental factors influencing human skin coloration." Am J Phys Anthropol 125(3): 292-302.

Cook, A. L., W. Chen, et al. (2009). "Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci." J Invest Dermatol 129(2): 392-405.

Duffy, D. L., N. F. Box, et al. (2004). "Interactive effects of MC1R and OCA2 on melanoma risk phenotypes." Hum Mol Genet 13(4): 447-461.

Duffy, D. L., G. W. Montgomery, et al. (2007). "A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation." Am J Hum Genet 80(2): 241-252.

Eiberg, H., J. Troelsen, et al. (2008). "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 123(2): 177-187.

Frost, P. (2006). "European hair and eye color

A case of frequency-dependent sexual selection?" Evolution and Human Behavior 27: 85-103. Frudakis, T., T. Terravainen, et al. (2007). "Multilocus OCA2 genotypes specify human iris colors." Hum Genet 122(3-4): 311-326.

Frudakis, T., M. Thomas, et al. (2003). "Sequences associated with human iris pigmentation." Genetics 165(4): 2071-2083.

Han, J., P. Kraft, et al. (2008). "A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation." PLoS Genet 4(5): e1000074.

Hart, K. L., S. L. Kimura, et al. (2013). "Improved eye- and skin-color prediction based on 8 SNPs." Croatian Medical Journal in press.

Hickman, M. J., Hughes, K.A., Strom, K.J., and Ropero-Miller, J.D. (2007). "Medical Examiners and Coroners' Offfices, 2004." Bureau of Justice and Statistics.

Hubbard, J. K., J. A. Uy, et al. (2010). "Vertebrate pigmentation: from underlying genes to adaptive function." Trends Genet 26(5): 231-239.

Hughes, K. (2007). "Unidentified Human Remains in the United States, 1080-2004." Bureau of Justice Statistics.

Jablonski, N. G. and G. Chaplin (2000). "The evolution of human skin coloration." J Hum Evol 39(1): 57-106.

Kayser, M., F. Liu, et al. (2008). "Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene." Am J Hum Genet 82(2): 411-423.

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

Keating, B., A. T. Bansal, et al. (2013). "First all-in-one diagnostic tool for DNA intelligence: genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the Identitas v1 Forensic Chip." Int J Legal Med 127(3): 559-572.

Kruglyak, L. and D. A. Nickerson (2001). "Variation is the spice of life." Nat Genet 27(3): 234-236.

Lao, O., J. M. de Gruijter, et al. (2007). "Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms." Ann Hum Genet 71(Pt 3): 354-369.

Lin, J. Y. and D. E. Fisher (2007). "Melanocyte biology and skin pigmentation." Nature 445(7130): 843-850.

Liu, F., K. van Duijn, et al. (2009). "Eye color and the prediction of complex phenotypes from genotypes." Curr Biol 19(5): R192-193.

Martin, N., D. Boomsma, et al. (1997). "A twin-pronged attack on complex traits." Nat Genet 17(4): 387-392.

McEvoy, B., S. Beleza, et al. (2006). "The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model." Hum Mol Genet 15 Spec No 2: R176-181.

Mengel-From, J., C. Borsting, et al. (2010). "Human eye colour and HERC2, OCA2 and MATP." Forensic Sci Int Genet 4(5): 323-328.

Mengel-From, J., T. H. Wong, et al. (2009). "Genetic determinants of hair and eye colours in the Scottish and Danish populations." BMC Genet 10: 88.

Mushailov, V., S. A. Rodriguez, et al. ( ). "Development and validation of an 8-plex assay to predict eye and skin coloration." in preparation ( ).

Norton, H. L., R. A. Kittles, et al. (2007). "Genetic evidence for the convergent evolution of light skin in Europeans and East Asians." Mol Biol Evol 24(3): 710-722.

Parra, E. J. (2007). "Human pigmentation variation: evolution, genetic basis, and implications for public health." Am J Phys Anthropol Suppl 45: 85-105.

Pneuman, A., Z. M. Budimlija, et al. (2012). "Verification of eye and skin color predictors in various populations." Leg Med (Tokyo).

Ringholm, A., J. Klovins, et al. (2004). "Pharmacological characterization of loss of function mutations of the human melanocortin 1 receptor that are associated with red hair." J Invest Dermatol 123(5): 917-923.

Ritter, N. (2007). "Missing Persons and Unidentified Remains: The Nation's Silent Mass Disaster." NIJ Journal(256).

Rosenberg, N. A. (2006). "Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives." Ann Hum Genet 70(Pt 6): 841-847.

Rosenberg, N. A., J. K. Pritchard, et al. (2002). "Genetic structure of human populations." Science 298(5602): 2381-2385.

Sanchez, J. J. and P. Endicott (2006). "Developing multiplexed SNP assays with special reference to degraded DNA templates." Nat Protoc 1(3): 1370-1378.

Shekar, S. N., D. L. Duffy, et al. (2008). "Linkage and association analysis of spectrophotometrically quantified hair color in Australian adolescents: the effect of OCA2 and HERC2." J Invest Dermatol 128(12): 2807-2814.

Sitaram, A., R. Piccirillo, et al. (2009). "Localization to mature melanosomes by virtue of cytoplasmic dileucine motifs is required for human OCA2 function." Mol Biol Cell 20(5): 1464-1477.

Spichenok, O., Z. M. Budimlija, et al. (2011). "Prediction of eye and skin color in diverse populations using seven SNPs." Forensic Sci Int Genet 5(5): 472-478.

Stokowski, R. P., P. V. Pant, et al. (2007). "A genomewide association study of skin pigmentation in a South Asian population." Am J Hum Genet 81(6): 1119-1132.

Sturm, R. A. (2009). "Molecular genetics of human pigmentation diversity." Hum Mol Genet 18(R1): R9-17.

Sturm, R. A., D. L. Duffy, et al. (2008). "A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color." Am J Hum Genet 82(2): 424-431.

Sturm, R. A. and T. N. Frudakis (2004). "Eye colour: portals into pigmentation genes and ancestry." Trends Genet 20(8): 327-332.

Sulem, P., D. F. Gudbjartsson, et al. (2008). "Two newly identified genetic determinants of pigmentation in Europeans." Nat Genet 40(7): 835-837.

Sulem, P., D. F. Gudbjartsson, et al. (2007). "Genetic determinants of hair, eye and skin pigmentation in Europeans." Nat Genet 39(12): 1443-1452.

Tishkoff, S. A. and B. C. Verrelli (2003). "Patterns of human genetic diversity: implications for human evolutionary history and disease." Annu Rev Genomics Hum Genet 4: 293-340.

Tully, G. (2007). "Genotype versus phenotype: human pigmentation." Forensic Sci Int Genet 1(2): 105-110.

Valenzuela, R. K., M. S. Henderson, et al. (2010). "Predicting phenotype from genotype: normal pigmentation." J Forensic Sci 55(2): 315-322.

Valverde, P., E. Healy, et al. (1995). "Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans." Nat Genet 11(3): 328-330.

Visser, M., M. Kayser, et al. (2012). "HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter." Genome Res 22(3): 446-455.

Walsh, S., F. Liu, et al. (2011). "IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information." Forensic Sci Int Genet 5(3): 170-180.

Walsh, S., F. Liu, et al. (2013). "The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA." Forensic Sci Int Genet 7(1): 98-115.

Walsh, S., A. Wollstein, et al. (2011). "DNA-based eye colour prediction across Europe with the IrisPlex system." Forensic Sci Int Genet.

Yuasa, I., K. Umetsu, et al. (2007). "Distribution of two Asian-related coding SNPs in the MC1R and OCA2 genes." Biochem Genet 45(7-8): 535-542.

Protocol of the developed 8-plex assay to predict eye and skin coloration

Summary:

The objective of this DNA, assay based on the validation of recently published tests, is predicting phenotypic characteristics in pigmentation, such as eye and skin color. This assay assesses several single-nucleotide-polymorphisms (SNPs) that are located nearby or within genes involved in pigmentation [rs12913832 at *HERC2*, rs1545397 at *OCA2*, rs16891982 at *SLC45A2*, rs1426654 at *SLC24A5*, rs885479 at *MC1R*, rs6119471 at *ASIP*, rs12203592 at *IRF4*, and rs12896399 at *SLC24A4*]. The 8-plex assay was validated according to the guidelines by the OCME Department of Forensic Biology Quality Assurance/Quality Control Manual and the Scientific Working Group on DNA Analysis Methods. The proposed method includes a multiplex-PCR followed by a single-base primer extension reaction that creates fluorescently labeled oligonucleotides of distinct length, which are detected by multicolor capillary electrophoresis (Fig. 1). This method is fast, robust, and importantly, can be applied on degraded DNA. The utilization of this assay will enhance current efforts to gather obtainable identifying information to assist in the identification of missing persons and human remains.

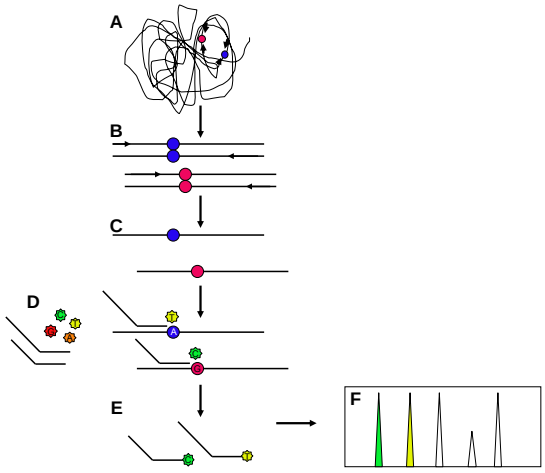

Figure 1: Outline of the multiplex-SNP-assay (shown for two SNPs): **A** Multiplex-PCR performed on genomic DNA with several primer-pairs (two SNPs are shown, in blue and red circles). **B** PCR-amplicons, with primers (arrows) and SNPs (colored circles). **C** Purified and denatured (95^{o}C) PCR-products. **D** Annealing of unlabeled primers and primer-extensions with fluorescently labeled ddNTPs. **E** purification, and **F** detection of fluorescently labeled oligonucleotides by multicolor capillary electrophoresis.

Protocol:

1. Multiplex-PCR

Multiplex-PCR reactions is usually performed in 25μl using the primers [rs12913832: F 5'-GGC TCT CTG TGT CTG ATC CA, R 5'-GGC CCC TGA TGA TGA TAG C; rs16891982: F 5'-TCC AAG TTG TGC TAG ACC AGA, R 5'-CGA AAG AGG AGT CGA GGT TG; rs1426654: F 5'-CCC TTG GAT TGT CTC AGG ATG, R 5'-TGA GTA AGC AAG AAG TAT AAG GAG CAA; rs1545397: F 5'-TGG AAT TGG ATA CTG ACA ATG G, R 5'-TCA CCG TGG GTA GAA TTA CCA; rs885479: F 5'-GCA CTG CGC TAC CAC AGC AT, R 5'-CCA GCA TAG CCA GGA AGA AG; rs6119471: F 5'-ATG GCG GGT GCC TAC TCT A, R 5'-AGG CTA ACC CGA AGG AAG AG; rs12203592: F 5'-GCT TTG TTT CAT CCA CTT TGG, R 5'-TGG AGT GAA CCC TTC ATC AG; and rs12896399: F 5'-CTG GCG ATC CAA TTC TTT GT, R 5'-CTT AGC CCT GGG TCT TGA TG] and the GoTaq Hot Start Polymerase (Promega, Madison, WI, USA) under standard conditions (buffer 1x, $MgCl_2$ 2mM, dNTPs 0.2mM each, GoTaq 1.25u). Genomic DNA is used at 0.1 to 2ng as template. The PCR conditions are 2min at 95°C, 40 cycles: 15sec at 95°C, 15sec at 59°C, 30sec at 72°C, followed by a final 2min at 72°C. Following the PCR, the products can be checked on Bioanalyzer DNA 1000 chips separating 15 and 1500bp (Agilent, Santa Clara, CA, USA), following the instructions of the manufacturer.

2. PCR clean-up

The clean-up reaction digests primers and dephosphoryaltes dNTPs from the multiplex-PCR to nucleosides and phosphates that cannot interfere with the following single-base primer extension reactions. This was performed by adding ExoSAP IT Product Cleanup (USB Affymetrix, Santa Clara, CA, USA) to the multiplex-PCR products and incubating for 60min at 37°C, following the inactivation of the enzyme by raising the temperature to 80°C for 20min.

3. Single-base primer extension (SBPE)

The single-base primer extension (SBPE) reaction extends a fluorescently labeled dideoxynucleoside-triphosphate to a primer depending on the template (SNP of interest). The SBPE primers [rs12913832: F 5'-CTCTCTCTCTCTCTC GCG AGG CCA GTT TCA TTT GAG CAT TAA; rs16891982: F 5'-TCTCT GTG AGG AAA ACA CGG AGT TGA TGC A; rs1426654: F 5'-GGA TTG TCT CAG GAT GTT GCA GGC; rs1545397: F 5'-CTCTCTCTCTCTCTC GGA TAC TGA CAA TGG TTG TAC AAC TTT GTG AAT ATA CTA AAA TAC; rs885479: R 5'-CAG ATG GCC GCA ACG GCT; rs6119471: R 5'-TCTCTCTCTC CCC GAA GGA AGA GTG AAA ATG CGT AA; rs12203592: F 5'-TCTCTCTCTCTCTCTCTC GTT TCA TCC ACT TTG GTG GGT AAA AGA AGG; and rs12896399: F 5'-CTCTCTCTCTCTCTC CGA TCC AAT TCT TTG TTC TTT AGG TCA GTA TAT TTT GGG] were used at the end-concentration of 0.2□M, mixed with SNaPshot reagent (Applied Biosystems Inc, Foster City, CA, USA) and added to the cleaned multiplex-PCR products. The reaction was performed in 25 cycles of 10sec at 96°C, 5sec at 57°C, and 30sec at 60°C.

For the allelic ladder, which leads to two labeled products for each SNP - one for each allele - oligonucleotides [rs12913832: 5'-TGA CA**C** TTA ATG CTC AAA TGA AAC TGG CCT CGC, 5'-TGA CA**T** TTA ATG CTC AAA TGA AAC TGG CCT CGC; rs16891982: 5'-GGC TT**C** TGC ATC AAC TCC GTG TTT TCC TCA C, 5'-GGC TT**G** TGC ATC AAC TCC GTG TTT TCC TCA C; rs1426654: 5'-AGT TG**C** GCC TGC AAC ATC CTG AGA CAA TCC, 5'-AGT TG**T** GCC TGC AAC ATC CTG AGA CAA TCC; rs1545397: 5'-TTC AG**T** GTA TTT TAG TAT ATT CAC AAA GTT GTA CAA CCA TTG TCA GTA TCC, 5'-TTC AG**A** GTA TTT TAG TAT ATT CAC AAA GTT GTA CAA CCA TTG TCA GTA TCC; rs885479: 5'-GCG GC**G** AGC CGT TGC GGC CAT CTG, 5'-GCG GC**A** AGC CGT TGC GGC CAT CTG; rs6119471: 5'-CAG GT**C** TTA CGC ATT TTC ACT CTT CCT TCG GG, 5'-CAG GT**G** TTA CGC ATT TTC ACT CTT CCT TCG GG; rs12203592: 5'-AAT TT**G** CCT TCT TTT ACC CAC CAA AGT GGA TGA AAC, 5'-AAT TT**A** CCT TCT TTT ACC CAC CAA AGT GGA TGA AAC; and rs12896399: 5'-AGA GA**C** CCC AAA ATA TAC TGA CCT AAA GAA CAA AGA ATT GGA TCG, 5'-AGA GA**A** CCC AAA ATA TAC TGA CCT AAA GAA CAA AGA ATT GGA TCG] were used as template.

4. SBPE clean-up

Remaining fluorescently labeled ddNTPs of the SBPE reaction were dephosphorylated using Shrimp Alkaline Phosphatase (SAP) (USB Affymetrix, Santa Clara, CA, USA) for 60min at 37^{o}C. The enzyme was inactivated at 80^{o}C for 20min.

5. Multicolor capillary electrophoresis

Fluorescently labeled oligonucleotides were separated and detected on the 3130*xl* Genetic Analyzer (Applied Biosystems Inc, Foster City, CA, USA). Sample preparation: 1.1μl of the SBPE-clean-up product was added to 8μl deionized HIDI Formamide (Applied Biosystems) and 1μl of Gene-Scan 120 LIZ (Applied Biosystems), following denaturation at 95^{o}C for 5min. Samples were injected for 10sec at 2kV and electrophoresed at 15kV for 600sec in POP7, with a run temperature of 60^{o}C. Data were collected using the 3130*xl* Genetic Analyzer Data Collection Software v3.0.

6. Data analysis

The results from the electrophoresis were analyzed in GeneMapper 4.0. Allele peaks were interpreted when the relative fluorescence units (RFUs) were greater or equal to 100 for C and T, 200 for A, and 500 for G.

7. Eye- and skin-color prediction

Eight SNPs were utilized for eye and skin color predictions: Eye-color prediction utilizes five SNPs and followed a two-step procedure: First, rs12913832 (*HERC2*) distinguishes eye colors as being not blue (i.e. brown or green) or not brown (i.e. green or blue). Further distinctions are made with additional SNPs: brown eye color is predicted by the following genotype combinations: A/A or G/A at rs12913832 (*HERC2*), plus either G/G at rs6119471 (*ASIP*), or C/C at rs16891982 (*SLC45A2*); green eye color is predicted by G/G at rs12913832

(*HERC2*) plus C/C at rs16891982 (*SLC45A2*), or by G/A at rs12913832 (*HERC2*) plus T/T at rs12203592 (*IRF4*); and blue eye color is predicted by G/G at rs12913832 (*HERC2*) plus T/T at rs12203592 (*IRF4*). The second step proceeds only with samples that were not positively described as being brown, green or blue during the first step. Samples that are homozygous for G/G at rs12913832 (*HERC2*) (i.e. not brown) plus T/T at rs12896399 (*SLC24A4*) are then predicted to have blue eyes, while samples that are homozygous for A/A at rs12913832 (*HERC2*) (i.e. not blue) plus G/G at rs12896399 (*SLC24A4*) are predicted to have brown eyes. The error rate is approximately 5% (4.54%).

The skin-color predictor utilizes six SNPs in an elimination process. Five are used to predict lighter skin color and one for darker shades. Non-dark skin color (i.e. light or medium) is predicted by any two of the following alleles: G/G at rs12913832 (*HERC2*), G/G at rs16891982 (*SLC45A2*), A/A at rs1426654 (*SLC24A5*), T/T at rs1545397 (*OCA2*), or A/A at rs885479 (*MC1R*). Light skin color is predicted by more stringent conditions: G/G at rs12913832 (*HERC2*), plus G/G at rs16891982 (*SLC45A2*), and A/A at rs1426654 (*SLC24A5*). Non-light skin color (i.e. medium or dark) is predicted by G/G at rs6119471 (*ASIP*). The error rate is approximately 1%.