The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: Gene Polymorphism and Human Pigmentation

Author: Murray H. Brilliant, Ph.D.

Document No.: 223980

Date Received: September 2008

Award Number: 2002-IJ-CX-K010

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Final Report

NIJ Grant# 2002-IJ-CX-K010

Murray H. Brilliant, Ph.D., (PI) University of Arizona

Portions of this report have been submitted for publication to the Journal *Genetics*

Current address: Department of Pediatrics, University of Arizona College of Medicine mhb@peds.arizona.edu

This project was carried out at the University of Arizona. The study involved the joint effort of the following individuals: Robert K. Valenzuela, Miquia S. Henderson, Monica H. Kim, Nanibaa' A. Garrison, Jessica T. Kelch, Orit Cohen-Barak, Drew T. Erickson, F. John Meaney, J. Bruce Walsh, Keith C. Cheng, Shosuke Ito, Kazumasa Wakamatsu, Tony Frudakis, Matthew Thomas.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

1. Abstract

The overall goal of this research project was to determine with a high degree of accuracy, the pigmentation phenotype (hair, eye and skin color) of an individual from a forensic DNA sample. Human pigmentation is programmed genetically. Genetic data suggested that most of the variation in human pigmentation is the result of the actions of only 3 or 4 genes. Candidate genes mediating pigment variation have been identified in previous animal and human genetic studies. These genes include those associated with oculocutaneous albinism (OCA) such as tyrosinase (OCA1), P transporter (OCA2), tyrosinase-related protein 1 (OCA3) and SLC45A2 (OCA4). Additional candidate genes include MC1R and ASIP (associated with mouse coat color and human hair color variation) and SLC24A5 (associated with population differences in pigmentation). In total, we examined 74 polymorphisms in candidate 23 genes. Our hypothesis was that polymorphisms within these specific pigmentation genes result in normal human pigment variation. We set out to correlate these polymorphisms with hair, skin and eye color. The application of this research to forensic science is the development of a finite and specific set human SNPs that can be used as a predictor of hair, eye and skin color.

Our analyses proceeded in three phases: Phase 1 was the determination of the spectrum of polymorphism in these genes in a limited number of individuals of diverse ethnic origins and various pigmentation types. Phase 2 was the correlation of these SNPs with pigmentation types in a representative sample of individuals. Phase 3 was the testing of the models determined in Phase 2 and the adaptation of the analysis to forensic DNA samples. Phases 1 and 2 have been completed and Phase 3 is nearly complete.

Using multiple linear regression (MLR) modeling, six SNPs in five genes were found to account for large proportions of pigmentation variation in hair, skin, and eyes in our across-population analyses: single polymorphisms in *SLC45A2*, *SLC24A5*, and *OCA2* accounted for 77.3% of the variance in total amount of scalp-hair melanin; single polymorphisms in *SLC45A2*, *SLC24A5*, and *MC1R* accounted for 38.3% of the variance in the natural logarithm of the ratio of black-to-red melanin in scalp hair; single polymorphisms in *SLC45A2*, *SLC24A5*, and *ASIP* accounted for 45.7% of the variance in skin pigmentation; and single polymorphisms in *SLC45A2*, *SLC24A5*, and *OCA2* accounted for 52.2% of the variance in eye color. Interactions between *ASIP* and *SLC45A2* were inferred from our models and increased the explanation of skin pigmentation variation to 49.6%. Thus, these models offer useful predictive tools using only three SNPs to determine an individual's pigmentation type, independent of ethnic origin.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table of Contents page

1. Abstract 2

2. Executive Summary 4

3. Introduction and History 20

4. Conclusions 58

5. Appendix 59

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

2. Executive Summary

INTRODUCTION

Variation in human pigmentation results from differences in the amount and type of melanin synthesized in melanosomes and in the size, shape, and export of those melanosomes to the hair and skin. Although the rate of synthesis of melanin is much lower in the developed eye than in the skin and hair, additional background color is contained in the iris, making eye color a more complex trait. Genes previously implicated in mediating pigment variation include the melanocortin-1 receptor (*MC1R*) gene and a gene encoding its inhibitor, agouti signaling protein (*ASIP*); two genes associated with oculocutaneous albinism *P* (*OCA2*) and *SLC45A2* (*OCA4,* formerly named MATP); and most recently *SLC24A5*, the human orthologue of the zebrafish *golden* gene.

Multiple polymorphisms in the *MC1R* gene have been linked to red hair and fair skin (BRANICKI *et al.* 2007; FLANAGAN *et al.* 2000; REES 2004; SULEM *et al.* 2007; VALVERDE *et al.* 1995) and the *ASIP* gene, encoding an inhibitor of the *MC1R* ligand, α-MSH, has been linked to skin pigmentation (BONILLA *et al.* 2005; KANETSKY *et al.* 2002; VOISEY *et al.* 2006). A wide range of OCA phenotypes has been noted for various mutants of *OCA2* (OETTING and KING 1999) and *SLC45A2* alleles (INAGAKI *et al.* 2004; NEWTON *et al.* 2001; RUNDSHAGEN *et al.* 2004). Similarly, a wide range of coat color phenotypes is seen in mice with various mutations in their respective orthologous genes (*p*: (JOHNSON *et al.* 1995; LYON *et al.* 1992) *Slc45a2*, and formerly named *uw*: (SWEET *et al.* 1998)). These observations suggested that variations in these genes may be associated with variation in the normal range of human pigmentation. Indeed, gene(s) associated with brown eyes and brown hair were found to map to chromosome 15q, with the *OCA2* gene as a prime candidate (EIBERG and MOHR 1996). Population studies have shown that specific polymorphisms in the *OCA2* and *SLC45A2* genes (together with polymorphisms in *MC1R*) are strongly associated with variations in normal pigmentation of the hair (EIBERG and MOHR 1996; STURM *et al.* 2001), skin (AKEY *et al.* 2001; DUFFY *et al.* 2004; NAKAYAMA *et al.* 2002; SHRIVER *et al.* 2003), and eyes (DUFFY *et al.* 2007; FRUDAKIS *et al.* 2007; FRUDAKIS *et al.* 2003; REBBECK *et al.* 2002; STURM and FRUDAKIS 2004; ZHU *et al.* 2004). Moreover, the *SLC45A2* polymorphism rs16891982 (L374F) may be a useful marker of population origin (SOEJIMA and KODA 2007; YUASA *et al.* 2006). Another marker of population origin that plays an important role in pigmentation, *SLC24A5* (or NCKX5, the human orthologue of the zebrafish *golden* gene), has been recently identified, with a coding polymorphism divergent between European/Caucasians and other human populations (LAMASON *et al.* 2005; SOEJIMA and KODA 2007).

Alleles of other pigmentation genes such as *tyrosinase-related protein 1* (*TYRP1*) and *dopachrome tautomerase* (*DCT*) have been statistically associated with human iris pigmentation (FRUDAKIS *et al.* 2003). In association with certain alleles of other genes, specific alleles of agouti signaling protein (*ASIP*, an antagonist of α-MSH signaling through *MC1R*) have also been associated with

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

human iris color (FRUDAKIS *et al.* 2003) as well as skin color (BONILLA *et al.* 2005; KANETSKY *et al.* 2002; VOISEY *et al.* 2006).

Previous studies have focused on the effects of a limited number of genes on hair or skin color (FLANAGAN *et al.* 2000; LAMASON *et al.* 2005; NAKAYAMA *et al.* 2002; VALVERDE *et al.* 1995). The most comprehensive studies to date have focused on genome-wide association of SNPs with pigmentation within specific populations (STOKOWSKI *et al.* 2007; SULEM *et al.* 2007).

In this report, we set out to determine markers predictive for human pigmentation, independent of ethnic origin. We assayed 74 polymorphisms in 23 genes that were previously implicated in human or animal pigmentation studies for analysis of single- and multi-locus associations with hair, skin, and eye color in 789 individuals of various ethnic backgrounds. Multiple linear regression (MLR) modeling revealed that a surprisingly small number of markers account for large proportions of pigmentation variation in hair, skin, and eyes in our across-population analyses: single polymorphisms in *SLC45A2*, *SLC24A5*, and *OCA2* accounted for 77.3% of the variance in total amount of scalp-hair melanin; single polymorphisms in *SLC45A2*, *SLC24A5*, and *MC1R* accounted for 38.3% of the variance in the natural logarithm of the ratio of black-to-red melanin in scalp hair; single polymorphisms in *SLC45A2*, *SLC24A5*, and *ASIP* accounted for 45.7% of the variance in skin pigmentation; and single polymorphisms in *SLC45A2*, *SLC24A5*, and *OCA2* accounted for 52.2% of the variance in eye color. Analyses of interactions between *ASIP* and *SLC45A2* increased the explanation of pigmentation variation in skin reflectance to 49.6%.

MATERIALS AND METHODS

Participants

Informed consent was obtained from 791 participants between the ages of 18 and 40 years with no gray hair and at least one inch (measured from the roots) of un-dyed scalp hair. Phenotype data, hair samples, and buccal cell samples were collected from each participant following an IRB approved protocol. Participant hair color and eye color were self-reported and independently scored by an investigator; participants also indicated other relevant information, such as tanning response and ethnicity. Buccal cell samples were collected using Catch-All Sample Collection Swabs (Epicentre, Madison, WI) and processed for DNA according to manufacturer's protocol.

Melanin Analysis

Approximately three hundred scalp hairs (1 cm at the base) were collected from each subject. Hair samples from 186 randomly selected participants were analyzed for both total melanin (combined amount of eumelanin and pheomelanin) content and the two subtypes of melanin, eumelanin and pheomelanin, following a previously published protocol (ITO and WAKAMATSU 1994; WAKAMATSU *et al.* 2002).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Skin Reflectance
Skin reflectance was measured with a portable spectrophotometer (Mercury 1000, Datacolor International, Lawrenceville, NJ) fitted with a 15mm aperture. Three reflectance measurements (measured as CIEL, L (lightness) scale of the International Committee on Illumination) of the inner aspect of the upper arm were recorded and averaged for each participant.

Eye color
Eye color was measured by matching subjects' eye color to the Kolberg Iris Color Chart® (ocularistsupplies.com) and recorded. Measurements were binned into six different color categories based on studies done by Wakamatsu et al. (unpublished data) that correlate pigmentation content to color. Categories were binned one through six (1=blue, 2=yellow brown, 3=green, 4=packets of brown + blue/green, 5=brown, and 6=dark brown/black), where bin one corresponded to the least amount of pigmentation, and six corresponded to the highest pigmentation.

DNA Amplification and Sequencing
Each PCR reaction consisted of the final concentrations/quantities of the following: 1xPCR buffer, 1.5 mM $MgCl_2$, 10pmol of each primer (forward and reverse; Table 1), 0.25 mM dNTPs, 1 U Taq, and ddH_2O to 20 µl. PCR amplification was performed using a PTC-200 Thermal Cycler (MJ Research, Watertown, MA). Thermal cycle program was as follows: 3 min. @ 95C, 34 cycles of 30 sec. at each temperature setting (95C, 55C, and 72C), and a final extension of 5min. at 72C.

Sequencing was performed by the Genomic Analysis and Technology Core at the University of Arizona on a 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA). Additionally, the first 287 DNA samples collected were genotyped at 74 SNPs in 23 genes implicated in melanin biosynthesis by DNA Print Genomics (Sarasota, FL) following a previously published protocol (Frudakis et al. 2003). The remaining 504 individuals were also genotyped by DNA Print Genomics (Sarasota, FL) for 41 of the 74 SNPs that were statistically significant.

Mathematical Modeling
Statistical analysis was performed using SAS (version 9.1) and JMP IN (release 5.1) statistical analysis software (SAS Institute, Cary, NC). The pool of SNPs was reduced from 74 (phase I; Table 2) to 39 (within 15 genes) by choosing SNPs that were statistically significant ($p<0.05$, Table 2) by ANOVA (N=287). Finally, rs1426654 (*SLC24A5*) and rs6058017 (*ASIP*) were genotyped for all individuals (N=791). Two samples were dropped from the study (sample 102 and 352) because of genotyping inconsistencies. In the analysis of the ratio of eumelanin-to-pheomelanin, the natural log of the ratio was used to transform the data.

A total of 39 SNPs (phase II; Table 2) were used to build MLR models of three-SNPs. Initially, forward, reverse, and mixed step-wise regression methods were used to trim the number of SNPs. Different models were obtained based

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

on the method used. To determine SNPs that were most significant in a three-SNP MLR model, we used SAS to generate all possible models of three-SNPs from the pool of 39 significant SNPs. All models were plotted by squared regression coefficient (R^2) in descending order of value and inflections in the resulting curves were noted. To find the basis of these inflections, we plotted histograms using JMP IN that contained all SNPs that comprised all models up to the R^2 inflections (along the steepest initial slopes). In doing so it became obvious which SNPs were predominantly responsible for the inflections. This method was performed in determining SNPs for each trait.

RESULTS

Our primary aim was to develop a forensic DNA test predictive for pigmentation phenotype. Therefore, we analyzed the data across all populations of the study. The pool of SNPs was reduced from 74 to 39 SNPs by ANOVA (Table 2). The models accounted for significant variation in each of the four measured traits: scalp-hair total melanin (77.3%), natural log of the ratio of eumelanin-to-pheomelanin (38.3%), skin reflectance (45.7%), and eye color (52.2%); Table 3.

SNPs from five genes comprised the top three-SNPs in the MLR models for each of the four different traits. Analysis of the three-SNP curves showed that all MLR models of the four traits contained the same two SNPs, rs16891982 (*SLC45A2*) and rs1426654 (*SLC24A5*). However, the third SNP varied in each trait model. Scalp-hair total melanin contained SNP rs1800410 (*OCA2*); natural log of the ratio of eumelanin-to-pheomelanin contained SNP rs1805007 (*MC1R*); skin reflectance contained SNP rs2424984 (*ASIP*); and eye color contained the SNP rs10852218 (*OCA2*).

Scalp-hair total melanin

The three-SNP R^2 curve (Figure 1A) showed SNP rs1800410 (*OCA2*) to be the third most frequent SNP for the top 250 R^2 models (2.34%). The last inflection, the top 1800 models (16.9%), showed rs1724630 (*MYO5A*) to be the third most frequent SNP. SNP rs1800410 (*OCA2*) in a three-SNP MLR model with rs16891982 (*SLC45A2*) and rs1426654 (*SLC24A5*) yielded the highest R^2 value of 77.3% (n=137) as compared to the other candidate third SNPs.

Natural log of the ratio of eumelanin-to-pheomelanin

The third most frequent SNP was rs1805007 (*MC1R*) for all inflections in the three-SNP MLR curve (Figure 1B). A three-SNP MLR model of rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), and rs1805007 (*MC1R*) yielded an R^2 of 38.3% (n=153).

Skin Reflectance

The third most frequent SNP (Figure 1C) was rs2424984 (*ASIP*). A three-SNP MLR yielded an R^2 value of 45.7%. An interaction term of rs2424984 (*ASIP*) and rs16891982 (*SLC45A2*) increased the R^2 value of the model by approximately 4% to 49.6% (n=447).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Eye color

The three-SNP R^2 curve (Figure 1D) showed the third most frequent SNP to be rs10852218 (*OCA2*) at inflections 170 (1.6%), and 385 (3.6%). However, at inflections 1230 (11.5%) and 1455 (13.6%) its frequency was not pronounced compared to all other SNPs. A three-SNP MLR model of rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), and rs10852218 gave the highest R^2 value of 52.2% (n=223).

DISCUSSION

Using MLR, we have shown that a large portion of variation in hair, skin, and eye color across diverse human populations can be accounted for by a small number of SNPs. Starting from an initial candidate group of 74 SNPs, we found by ANOVA that 39 of these SNPs were significantly associated with hair pigmentation, skin reflectance, or eye color phenotype. Initially, we trimmed the number of SNPs by forward, backward, and/or a combination of both step-wise regression methods, however, each method yielded a different model. In an attempt to circumvent this problem we modeled all possible combinations of significant SNPs. The top models R^2 values differed on average by less than 3-10 thousandths of a percent or smaller. The question arose as to which was the best model. To answer this question we chose SNPs that were most frequent in the highest R^2 three-SNP-models. This method ultimately trimmed the SNPs to six SNPs (three coding, and three not coding) in four genes that accounted for most of the variance (77.3% for hair total melanin; 38.3% for hair eumelanin-to-pheomelanin ratio; 45.7% for skin CIEL; and 52.2% for eye color).

Total Hair Melanin and Eumelanin-to-pheomelanin Ratio

Previous studies have examined the overall color of hair in relationship to various genetic markers. Overall hair color is the result of at least two parameters: total melanin and the ratio of eumelanin-to-pheomelanin. These can be measured objectively by chemical analysis (WAKAMATSU and ITO 2002). We found that although both parameters are associated with SNPs from *SLC24A5* and *SLC45A2*, they differ in the third most significant genetic contributor (*OCA2* for total hair melanin, and *MC1R* for the ratio of eumelanin-to-pheomelanin).

A high proportion of phenotypic variance of total hair melanin (77.3%) can be accounted for by three-SNPs: rs1426654 (*SLC24A5*), rs16891982 (*SLC45A2*), and rs1800410 (*OCA2*). While *SLC24A5* is an ancestry informative marker (AIM), it does play a role in pigmentation and so it is likely that the coding SNP rs1426654 is indeed a determinant for normal human pigmentation variation. Similarly, rs16891982 (*SLC45A2*) has been shown to be an AIM and has been shown to be associated with pigmentation variation in mice (SWEET *et al.* 1998) and in humans (GRAF *et al.* 2005; SOEJIMA and KODA 2007). Moreover, mutations in *SLC45A2* cause OCA4 (NEWTON *et al.* 2001) in diverse populations.

For analysis of total hair melanin, SNP rs1800410 (*OCA2*) was the third most significant contributor in a three-SNP model. Mutations in this gene lead to albinism in diverse populations (OETTING and KING 1999). *OCA2* has been associated with hair color by linkage analysis (EIBERG and MOHR 1996).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

For the natural log of the ratio of eumelanin-to-pheomelanin, SNP rs1805007 (*MC1R*) was the third most significant contributor in a three-SNP model. *MC1R* has been shown to be a major determinant in whether eumelanin or pheomelanin is produced (REES 2003; STURM *et al.* 2001). *MC1R* variants that decrease the protein's functionality have been shown to be associated with an increased incidence of red hair color in humans (FLANAGAN *et al.* 2000; VALVERDE *et al.* 1995). Chemically, this translates to increased pheomelanin and decreased eumelanin production.

Different studies have analyzed *MC1R* in different ways. Many analyzed specific populations to determine which SNPs were associated with variation in hair color, other studies focused on specific populations and on red hair. Moreover, the statistical analyses employed were different (i.e., they analyzed different combinations of variants), whereas, we analyzed for significance by 1) one-way ANOVA, and 2) by our criteria for designing a three-SNP model; Table 4.

Although we did not examine all known SNPs of *MC1R*, some of the SNPs we examined have been studied in relationship to red hair color. Valverde et al. (1995) found rs2228479 (V60L) in combination with other non-synonymous *MC1R* SNPs in British and Irish to be associated with red hair color. In contrast, two studies (BRANICKI *et al.* 2007; FLANAGAN *et al.* 2000) did not find an association of rs2228479 with red hair. Looking across populations of all hair colors, we did not find rs2228479 to be significant by ANOVA. Additionally, Flanagan et al. 2000 and Branicki et al. 2007 found that rs1805007 (R151G) in combination with rs1805008 (R160W) to be associated with red hair color. Moreover, Sulem et al. 2007 found both of these SNPs to be associated with hair color in Icelandic and Dutch populations. In our three-SNP MLR model, rs1805007 (R151G) was the third most important genetic contributor for the ratio of pigmentation in hair for the across population data set. We note that our analysis of the ratio of hair pigmentation did not focus on any particular hair color. However, when all red heads are removed from the across population data set (24 self-reported Caucasians and 3 self-reported Hispanics), rs1800407 (*OCA2*) becomes the third most significant contributor for the hair ratio of melanins rather than rs1805007 (R151G).

Skin Reflectance (CIEL)

A high proportion of phenotypic variance of skin reflectance (45.7%) can be accounted for by three-SNPs: rs1426654 (*SLC24A5*), rs16891982 (*SLC45A2*), and rs2424984 (*ASIP*). Previous studies have examined SNP rs1426654 (*SLC24A5*) by statistical analysis in relationship to human ethnicity across populations (LAMASON *et al.* 2005; NORTON *et al.* 2007; SOEJIMA and KODA 2007). Stokowski et al. 2007 showed rs1426654 to be significantly associated with a dichotomously defined skin-reflectance in South Asians. *SLC24A5* has been shown to biologically affect pigmentation in zebrafish, *Danio rerio* (LAMASON *et al.* 2005). Although it is an AIM, it does not clearly distinguish between Europeans and Sri Lankans (SOEJIMA and KODA 2007). In contrast, SNP rs16891982 (*SLC45A2*) does distinguish between Europeans and Sri Lankans (SOEJIMA and

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

KODA 2007). Moreover, rs16891982 (*SLC45A2*) has been associated with skin pigmentation variation across and within populations (GRAF *et al.* 2005; NAKAYAMA *et al.* 2002; STOKOWSKI *et al.* 2007).

Using MLR, we found that both SNPs rs1426654 (*SLC24A5*) and rs16891982 (*SLC45A2*) described much of the variation in skin pigmentation across populations. The third most significant genetic contributor in a three-SNP MLR model was SNP rs2424984 (*ASIP*). *ASIP* has been shown to be associated with skin pigmentation, namely for rs6058017 (BONILLA *et al.* 2005; KANETSKY *et al.* 2002). Although we found rs6058017 to be significant by ANOVA, we did not find it to be a better predictor in skin reflectance than rs2424984. For both a single SNP analysis and a three-SNP model, rs2424984 was a better predictor for skin reflectance.

Other studies have found *MC1R* (LAO *et al.* 2007; REES 2003; SHRIVER *et al.* 2003; STURM *et al.* 2001), *OCA2*, and *DCT* (LAO *et al.* 2007) to be associated with normal skin pigmentation. Similarly, we found SNPs within the above genes to be associated by ANOVA with normal skin pigmentation. However, none were found to be significant contributors using our three-SNP MLR model. This does not imply that these genes are not important in pigmentation, it simply means that our method was unable to detect their significance. This may be due to a variety of factors: sampling error due to small sample size, different populations studied, and/or different methods of measuring skin reflectance.

Eye Color

Previous studies have examined eye color in relationship to various genetic markers. *OCA2* (DUFFY *et al.* 2007; EIBERG and MOHR 1996; FRUDAKIS *et al.* 2003; REBBECK *et al.* 2002; SULEM *et al.* 2007), *SLC45A2* (GRAF *et al.* 2005), and *MC1R* (GRAF *et al.* 2005; KANETSKY *et al.* 2004; SULEM *et al.* 2007) have been statistically associated with variation in eye color. In agreement with these studies, we found that a high proportion of eye color variation (52.2%) can be accounted for by three-SNPs: rs1426654 (*SLC24A5*), rs16891982 (*SLC45A2*), and rs10852218 (*OCA2*).

Duffy et al. (2007) found three-SNPs within intron 1 of *OCA2* that when considered as a haplotype-diplotype explained about 74% of eye color variation. However, there are significant differences between our study and theirs. Among the differences were: populations studied, SNPs genotyped, and binning of eye colors. They analyzed almost completely Northern Europeans, whereas, we analyzed across various ethnic populations and we genotyped additional SNPs within genes other than *OCA2*. SNPs rs1426654 (*SLC24A5*) and rs16891982 are likely contributors to the three-SNP MLR model because they act as AIMs, such that they distinguish Caucasians from non-Caucasians. Most non-Caucasians have brown/dark-brown eyes. Therefore, if *SLC24A5* and *SLC45A2* do not in fact have a role in iridial melanin production, then they are still adequate predictors of eye color when analyzing a sample in the circumstance when ethnic origin is unknown.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Statistical Interactions

Statistical interaction of SNPs rs12910433 (*OCA2*) and rs2228479 (*MC1R*) have been associated with skin reflectance in a Tibetan population (AKEY *et al.* 2001). We did not find this interaction, however, we did find an interaction between SNPs rs16891982 (*SLC45A2*) and rs2424984 (*ASIP*) for the across population skin reflectance MLR model. These differences may reflect our use of an ethnically-diverse sample; thus, the interaction of these two genes might be more predictive of skin reflectance across populations. Other interactions likely exist, but sample size decreased the statistical power to detect them.

We found 6 SNPs in 5 genes that were informative for normal human pigmentation (*SLC45A2*, *SLC24A5*, *MC1R*, *OCA2*, and *ASIP*). Three of these (*SLC45A2*, *SLC24A5* , and *MC1R*) were coding. Clearly, variations within these proteins can result in functional variation that contributed to the observed pigmentation variation. In contrast, the most significant SNPs found in *OCA2* and *ASIP* were not coding. We genotyped most of all known coding SNPs within these genes that showed allele frequency differences across populations, and none were as significantly correlated to pigmentation as the informative SNPs. This suggests that the regulation of the expression of *OCA2* and *ASIP* underlies pigmentation variation. Although our results demonstrate that relatively few SNPs in a relatively few genes control a significant portion of normal human pigmentation variation, it may be that additional polymorphisms in these same genes account for the remaining variation and/or that additional genes are involved. We note that at least one of these SNPs may be rs1667394 located approximately 200 kb upstream of the OCA2 gene in the HERC2 gene (SULEM *et al.* 2007).

Two of the five SNPs (rs1426654 (SLC24A5) and rs16891982 (SLC45A2)) that were found to be significant in our study are AIMs. Because these lead to amino acid changes, it is certainly possible that they reflect functional changes in these proteins known to be important to normal pigmentation. Indeed, both of these SNPs have been shown to be significant contributors to pigmentation differences within a South Asian population (STOKOWSKI *et al.* 2007). However, even if these SNPs function as AIMs, they are very good predictive markers for normal human pigmentation variation across ethnic backgrounds. We note that these models of human constitutive pigmentation phenotype have significant implications for forensic science. These results suggest that assays can be developed to predict hair, skin, and eye color from DNA samples. Such assays would be an extremely useful forensic tool, especially in cases where an individual's DNA is not in the database and no other identifiers are available.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

References for the Executive Summary

AKEY, J. M., H. WANG, M. XIONG, H. WU, W. LIU *et al.*, 2001 Interaction between the melanocortin-1 receptor and P genes contributes to inter-individual variation in skin pigmentation phenotypes in a Tibetan population. Hum Genet **108:** 516-520.

BONILLA, C., L. A. BOXILL, S. A. DONALD, T. WILLIAMS, N. SYLVESTER *et al.*, 2005 The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin color in African Americans. Hum Genet **116:** 402-406.

BRANICKI, W., U. BRUDNIK, T. KUPIEC, P. WOLANSKA-NOWAK and A. WOJAS-PELC, 2007 Determination of phenotype associated SNPs in the MC1R gene. J Forensic Sci **52:** 349-354.

DUFFY, D. L., N. F. BOX, W. CHEN, J. S. PALMER, G. W. MONTGOMERY *et al.*, 2004 Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. Hum Mol Genet **13:** 447-461.

DUFFY, D. L., G. W. MONTGOMERY, W. CHEN, Z. Z. ZHAO, L. LE *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:** 241-252.

EIBERG, H., and J. MOHR, 1996 Assignment of genes coding for brown eye colour (BEY2) and brown hair colour (HCL3) on chromosome 15q. Eur J Hum Genet **4:** 237-241.

FLANAGAN, N., E. HEALY, A. RAY, S. PHILIPS, C. TODD *et al.*, 2000 Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation. Hum Mol Genet **9:** 2531-2537.

FRUDAKIS, T., T. TERRAVAINEN and M. THOMAS, 2007 Multilocus OCA2 genotypes specify human iris colors. Hum Genet.

FRUDAKIS, T., M. THOMAS, Z. GASKIN, K. VENKATESWARLU, K. S. CHANDRA *et al.*, 2003 Sequences associated with human iris pigmentation. Genetics **165:** 2071-2083.

GRAF, J., R. HODGSON and A. VAN DAAL, 2005 Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. Hum Mutat **25:** 278-284.

INAGAKI, K., T. SUZUKI, H. SHIMIZU, N. ISHII, Y. UMEZAWA *et al.*, 2004 Oculocutaneous albinism type 4 is one of the most common types of albinism in Japan. Am J Hum Genet **74:** 466-471.

ITO, S., and K. WAKAMATSU, 1994 An improved modification of permanganate oxidation of eumelanin that gives a constant yield of pyrrole-2,3,5-tricarboxylic acid. Pigment Cell Res **7:** 141-144.

JOHNSON, D. K., L. J. STUBBS, C. T. CULIAT, C. S. MONTGOMERY, L. B. RUSSELL *et al.*, 1995 Molecular analysis of 36 mutations at the mouse pink-eyed dilution (p) locus. Genetics **141:** 1563-1571.

KANETSKY, P. A., F. GE, D. NAJARIAN, J. SWOYER, S. PANOSSIAN *et al.*, 2004 Assessment of polymorphic variants in the melanocortin-1 receptor gene with cutaneous pigmentation using an evolutionary approach. Cancer Epidemiol Biomarkers Prev **13:** 808-819.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

KANETSKY, P. A., J. SWOYER, S. PANOSSIAN, R. HOLMES, D. GUERRY *et al.*, 2002 A polymorphism in the agouti signaling protein gene is associated with human pigmentation. Am J Hum Genet **70:** 770-775.

LAMASON, R. L., M. A. MOHIDEEN, J. R. MEST, A. C. WONG, H. L. NORTON *et al.*, 2005 SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science **310:** 1782-1786.

LAO, O., J. M. DE GRUIJTER, K. VAN DUIJN, A. NAVARRO and M. KAYSER, 2007 Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. Ann Hum Genet **71:** 354-369.

LYON, M. F., T. R. KING, Y. GONDO, J. M. GARDNER, Y. NAKATSU *et al.*, 1992 Genetic and molecular analysis of recessive alleles at the pink-eyed dilution (p) locus of the mouse. Proc Natl Acad Sci U S A **89:** 6968-6972.

NAKAYAMA, K., S. FUKAMACHI, H. KIMURA, Y. KODA, A. SOEMANTRI *et al.*, 2002 Distinctive distribution of AIM1 polymorphism among major human populations with different skin color. J Hum Genet **47:** 92-94.

NEWTON, J. M., O. COHEN-BARAK, N. HAGIWARA, J. M. GARDNER, M. T. DAVISSON *et al.*, 2001 Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4. Am J Hum Genet **69:** 981-988.

NORTON, H. L., R. A. KITTLES, E. PARRA, P. MCKEIGUE, X. MAO *et al.*, 2007 Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Mol Biol Evol **24:** 710-722.

OETTING, W. S., and R. A. KING, 1999 Molecular basis of albinism: mutations and polymorphisms of pigmentation genes associated with albinism. Hum Mutat **13:** 99-115.

REBBECK, T. R., P. A. KANETSKY, A. H. WALKER, R. HOLMES, A. C. HALPERN *et al.*, 2002 P gene as an inherited biomarker of human eye color. Cancer Epidemiol Biomarkers Prev **11:** 782-784.

REES, J. L., 2003 Genetics of hair and skin color. Annu Rev Genet **37:** 67-90.

REES, J. L., 2004 The genetics of sun sensitivity in humans. Am J Hum Genet **75:** 739-751.

RUNDSHAGEN, U., C. ZUHLKE, S. OPITZ, E. SCHWINGER and B. KASMANN-KELLNER, 2004 Mutations in the MATP gene in five German patients affected by oculocutaneous albinism type 4. Hum Mutat **23:** 106-110.

SHRIVER, M. D., E. J. PARRA, S. DIOS, C. BONILLA, H. NORTON *et al.*, 2003 Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet **112:** 387-399.

SOEJIMA, M., and Y. KODA, 2007 Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2. Int J Legal Med **121:** 36-39.

STOKOWSKI, R. P., P. V. KRISHNA PANT, D. TONY, A. FEREDAY, D. A. HINDS *et al.*, 2007 A Genomewide Association Study of Skin Pigmentation in a South Asian Population. Am J Hum Genet **81**.

STURM, R. A., and T. N. FRUDAKIS, 2004 Eye colour: portals into pigmentation genes and ancestry. Trends Genet **20:** 327-332.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

STURM, R. A., R. D. TEASDALE and N. F. BOX, 2001 Human pigmentation genes: identification, structure and consequences of polymorphic variation. Gene **277:** 49-62.

SULEM, P., D. F. GUDBJARTSSON, S. N. STACEY, A. HELGASON, T. RAFNAR *et al.*, 2007 Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet.

SWEET, H. O., M. H. BRILLIANT, S. A. COOK, K. R. JOHNSON and M. T. DAVISSON, 1998 A new allelic series for the underwhite gene on mouse chromosome 15. J Hered **89:** 546-551.

VALVERDE, P., E. HEALY, I. JACKSON, J. L. REES and A. J. THODY, 1995 Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat Genet **11:** 328-330.

VOISEY, J., C. GOMEZ-CABRERA MDEL, D. J. SMIT, J. H. LEONARD, R. A. STURM *et al.*, 2006 A polymorphism in the agouti signalling protein (ASIP) is associated with decreased levels of mRNA. Pigment Cell Res **19:** 226-231.

WAKAMATSU, K., and S. ITO, 2002 Advanced chemical methods in melanin determination. Pigment Cell Res **15:** 174-183.

WAKAMATSU, K., S. ITO and J. L. REES, 2002 The usefulness of 4-amino-3-hydroxyphenylalanine as a specific marker of pheomelanin. Pigment Cell Res **15:** 225-232.

YUASA, I., K. UMETSU, S. HARIHARA, A. KIDO, A. MIYOSHI *et al.*, 2006 Distribution of the F374 allele of the SLC45A2 (MATP) gene and founder-haplotype analysis. Ann Hum Genet **70:** 802-811.

ZHU, G., D. M. EVANS, D. L. DUFFY, G. W. MONTGOMERY, S. E. MEDLAND *et al.*, 2004 A genome scan for eye color in 502 twin families: most variation is due to a QTL on chromosome 15q. Twin Res **7:** 197-210.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Tables – Executive Summary

Table 1. Primers used to amplify *OCA2* exons 9 and 13, and sequences containing SNPs rs1426654 and rs6058017.

OCA2 exon9	Primer Name	
forward	MHB581	(5'-GGGCTGAATTGTTCCATTTG-3')
reverse	MHB582	(5'-TCTCACGGATCTCAAGCCTC-3')
sequencing	MHB583	(5'-TGGCTGATACAGAGGGAGGT-3')
OCA2 exon13		
forward	MHB593	(5'-GGCTCCCTGTTCTTAAAGTC-3')
reverse	MHB594	(5'-TGAGAATGGAACCTGGAGCC-3')
sequencing	MHB595	(5'-AAGGCTGCCTCTGTTCTACG-3')
rs1426654 (*SLC24A5*)		
forward	MHB1755	(5'-TCATAAAGAAGCAAAACATTGGA-3')
reverse	MHB1756	(5'-AGCACAGATGCCAAGGAGAT-3')
sequencing	MHB1757	(5'-TGCCAATATCTCCCTTTGTG-3')
rs6058017 (*ASIP*)		
forward	MHB1836	(5'-GGCTTCGATGAAGAAAGTGG-3')
reverse	MHB1837	(5'-AGGGGATAGCCTCGTTCCTA-3')
sequencing	MHB1838	(5'-ATGGGACTTCAGGGAGACCT-3')

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 2. 74 SNPs (phase I) within 23 genes.

Gene	SNP	Total hair melanin	Natural log of ratio of melanins	Skin Reflectance (CIEL)	Eye Color
AP3D1	rs2238600				
AP3D1	rs4806830				
ASIP	rs2424984	*	*	*	*
ASIP	rs6058017	*	*	*	*
ASIP	rs2296151				
CYP2C8	rs1341164				
CYP2D6	rs1058174				
CYP4B1	rs1572603			*	
DCT	rs1325611	*	*	*	*
DCT	rs1407995	*	*	*	*
GPR143	rs3044	*	*	*	*
GPR143	rs3810741				
GPR143	rs2521667				
HPS1	rs1804689				
HPS1	rs2296430				
HPS1	rs3830015				
HPS3	rs2689234				
HPS4	rs1894704		*	*	
HPS4	rs3752589				*
HPS4	rs3752590				*
HPS4	rs739289		*		
HPS5	rs2305564				
HPS6	rs4917959				
LYST	rs3768051				
MAOA	rs979605				
MC1R	rs1805007		*	*	
MC1R	rs1805008	*	*	*	*
MC1R	rs3212346	*	*		
MC1R	rs3212355			*	
MC1R	rs3212357	*	*	*	*
MC1R	rs3212370				
MC1R	rs3212368				
MC1R	rs3212366				
MC1R	rs3212363				
MC1R	rs3212364				
MC1R	rs2228479				
MC1R	rs1805005				
MC1R	rs3212353				
MC1R	rs3212352				
MC1R	rs3212351				
MLPH	rs2292885	*		*	
MLPH	rs1343768				
MYO18A	rs11080078				*
MYO5A	rs1724630			*	
MYO5A	rs752864			*	
MYO5A	rs2290332	*	*	*	
MYO7A	rs2276289	*		*	
MYO7A	rs3737454	*		*	
MYO7A	rs948962				
MYO7A	rs1320703				
MYO7A	rs2276288				
OCA2	rs1037208			*	
OCA2	rs10852218		*	*	
OCA2	rs11638265	*		*	*
OCA2	rs1800404	*	*	*	*
OCA2	rs1800407	*		*	
OCA2	rs1800410	*	*	*	*
OCA2	rs1800411	*	*	*	*
OCA2	rs1800414	*			
OCA2	rs1900758	*	*	*	*
OCA2	rs749846	*	*		*
OCA2	rs1800402				
OCA2	rs737051				
OCA2	rs1800415				
OCA2	rs2305253				
OCA2	rs1800401				
OCA2	rs2044627				
SLC24A5	rs1426654	*	*	*	*
SLC45A2	rs16891982	*	*	*	*
SLC45A2	rs2287949		*	*	*
SLC45A2	rs26722	*	*	*	*
SLC45A2	rs40132	*	*	*	*
TYRP1	rs2733832	*	*	*	*
TYRP1	rs683				

* Significant SNPs by ANOVA ($p<0.05$) (39 SNPs within 15 genes) with respect to each of the four phenotypes for the first 287 samples.
Note: rs3212363 (*MC1R*) was genotyped for all individuals, it was not significant across all populations.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 3. Across populations R^2 values of individual SNPs, interaction terms, full model, and sample size of each ethnic group (self-reported).

Scalp-Hair Total Melanin	Individual R^2 values % (sample size; p value)	Caucasian	African-American	Hispanic	South Asian	East Asian	Native American	Admixture	not listed
rs16891982 (*SLC45A2*)	63.7 (n=172, p=7.2x10-38)	102	18	14	4	10	5	19	
rs1426654 (*SLC24A5*)	62.3 (n=166, p=2.76x10-35)	101	18	11	2	8	5	21	
rs1800410 (*OCA2*)	16.2 (n=160, p=9.41x10-7)	100	13	11	4	10	5	17	
Model	77.3 (n=137)	85	13	8	2	8	5	16	
ln(Eumelanin/Pheomelanin)									
rs16891982 (*SLC45A2*)	24.7 (n=172, 3.88x10-11)	102	18	14	4	10	5	19	
rs1426654 (*SLC24A5*)	24.0 (n=166, p=1.97x10-10)	101	18	11	2	8	5	21	
rs1805007 (*MC1R*)	14.0 (n=186, p=1.41x10-7)	113	19	14	4	10	5	21	
Model	38.3 (n=153)	91	17	11	2	8	5	19	
Average Skin Reflectance									
rs16891982 (*SLC45A2*)	32.1 (n=487, p=2.28x10-41)	267	19	83	23	42	14	30	9
rs1426654 (*SLC24A5*)	23.9 (n=528, p=7.09x10-32)	296	20	88	22	41	15	33	13
rs2424984 (*ASIP*)	15.6 (n=538, p=2.18x10-20)	300	20	89	25	44	15	32	13
Model	45.7 (n=447)	244	19	74	20	39	14	28	9
Model+rs16891982*rs2424984	49.6 (n=447)								
Eye Color									
rs16891982 (*SLC45A2*)	38.4 (n=397, p=3.55x10-42)	191	21	56	35	47	13	24	10
rs1426654 (*SLC24A5*)	34.3 (n=490, p=3.64x10-45)	260	21	69	36	46	15	30	13
rs10852218 (*OCA2*)	6.3 (n=270, p=1.75x10-4)	120	12	42	23	38	12	18	5
Model	52.2 (n=223)	92	12	33	18	35	12	16	5

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 4. ANOVA results for *MC1R* SNPs with respect to hair total melanin and natural log of eumelanin-to-pheomelanin. Phase II SNPs, and phase I SNPs.

	Phase I			Phase II		
MC1R SNP	Total melanin	n	ratio	Total melanin	n	ratio
rs3212357	**	141	**	**	219	**
rs3212346	*	138	**	**	191	**
rs1805008	*	141	**	ns	218	**
rs1805007	ns	141	**	**	219	**
rs3212355	ns	137	ns	**	214	ns
rs3212370	nv	141	nv			
rs3212368	nv	141	nv			
rs3212364	nv	141	nv			
rs3212352	nv	141	nv			
rs3212351	nv	141	nv			
rs3212366	ns	137	ns			
rs3212353	ns	136	ns			
rs1805005	ns	124	ns			
rs2228479	ns	122	ns			
rs3212363	ns	49	ns			

nv = no variation

ns = $p>0.05$

* = $p<0.05$

** = $p<0.01$

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

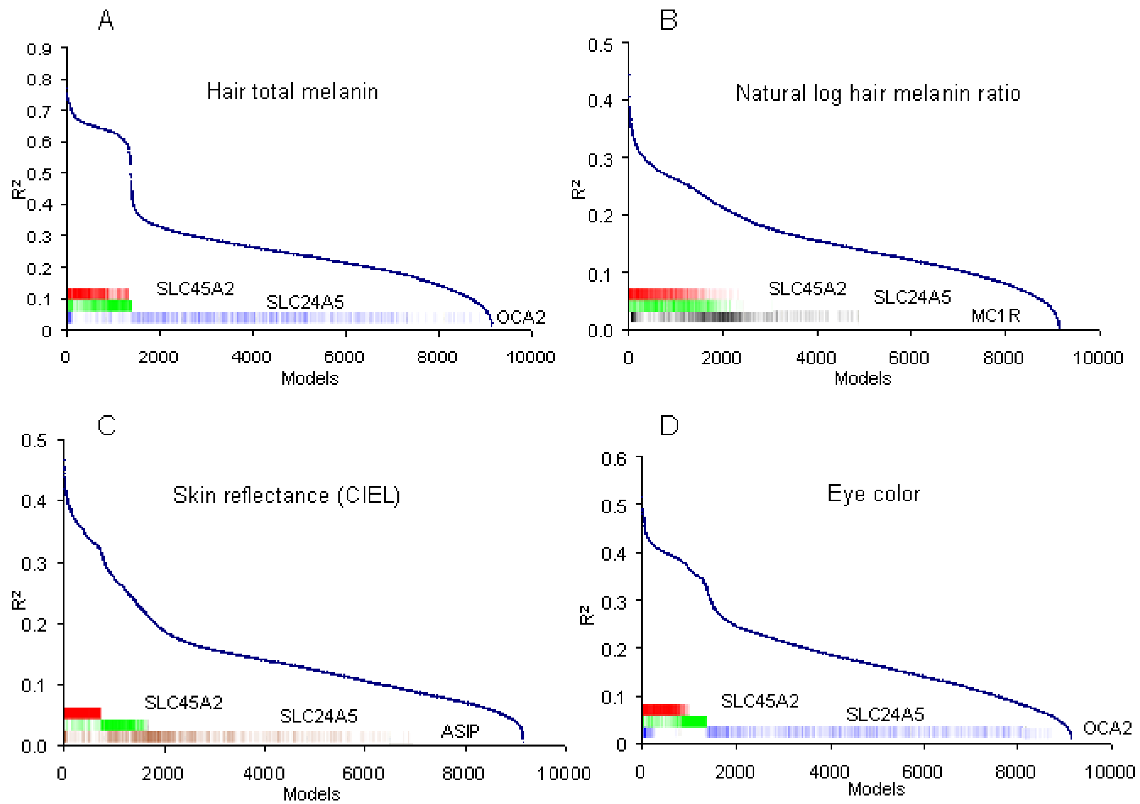

Figure 1. Three-SNP MLR models for four pigmentation traits across populations. In each panel, the horizontal-axis depicts all 9,139 combinations of phase I significant SNPs in a three-SNP MLR model (i.e., 39 choose 3). The vertical-axis is the total R2 value for each model. The three-SNPs that occurred most frequently in the models with the highest R2 values are indicated by the colored bar insets. Each model contained rs16891982 (*SLC45A2*) in red; and rs1426654 (*SLC24A5*) in green. The third SNP of each trait varied. A) Hair total melanin's third SNP was rs1800410 (*OCA2*) in blue, model R2 value of 77.3%. B) Natural log of hair melanin ratio's third SNP was rs1805007 (*MC1R*) in black, model R2 value of 38.3%. C) Skin Reflectance's third SNP was rs2424984 (*ASIP*) in brown, model R2 value of 45.7%. D) Eye color's third SNP was rs10852218 (*OCA2*) in blue, model R2 value of 52.2%.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

3. Introduction and History

Background to the Project

Human pigmentation is genetically controlled. Environmental factors (e.g., sunlight) can modify the constitutive pigmentation (e.g., through tanning), but even the ability to modify constitutive pigmentation is genetically controlled. The genetic control of pigmentation implies that it is possible to predict a person's pigmentation (hair, eye and skin color) if we know the individual genes involved, how their encoded proteins act and interact with each other, and what polymorphic variants lead to specific measures of pigmentation. The ability to predict these indicators of appearance from a DNA sample would be a powerful asset and novel tool for forensic analysis. Until recently we did not know which genes were involved in pigmentation. We now know the key genes involved in pigmentation from the study of mice and people with hypopigmentation disorders such as albinism. We also have an emerging picture of how the encoded proteins function and interact through molecular, cellular and biophysical characterizations. Although mutations and some polymorphisms have been described in these key genes (primarily as a result of disease-related studies), critical research needs to be done before we can translate this knowledge into useful tools for forensic analyses.

A brief overview of pigmentation in mammals. Pigmentation in mammals is manifest in the hair, the eye and the skin. The production of melanin, the major mammalian pigment, is highly controlled and compartmentalized, most likely because melanin is a toxic product (King et al, 1994; Hill et al, 1997). Many genes are known to affect pigmentation in mouse models (Silvers, 1979; Jackson, 1997). These genes affect various processes: some affect the viability, proliferation and migration of melanocytes themselves and as mutations are generally manifested as "spotting" traits (e.g., piebaldism). Other genes are involved in melanin biosynthesis (on or off), with very few variants (alleles) that result in intermediate pigmentation. Other genes are associated with high allelic variation and a wide spectrum of pigmentation types. These genes control the rate of melanin biosynthesis, the sizes and shapes of the melanosome (organelle/granules within the melanocytes), and regulate external cues that signal the melanocyte to make melanin or to make certain types of melanin. Although many pigmentation genes are known through mouse models, normal human (skin) pigment variation may be mediated by only a few (3 or 4) genes (Stern, 1970; Fig. 2). It is likely that normal variation in pigmentation is associated with variations in the size, shape, packaging and melanin content of the melanosome organelle as is seen among various racial types (Szabo et al, 1969; Fig. 3).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

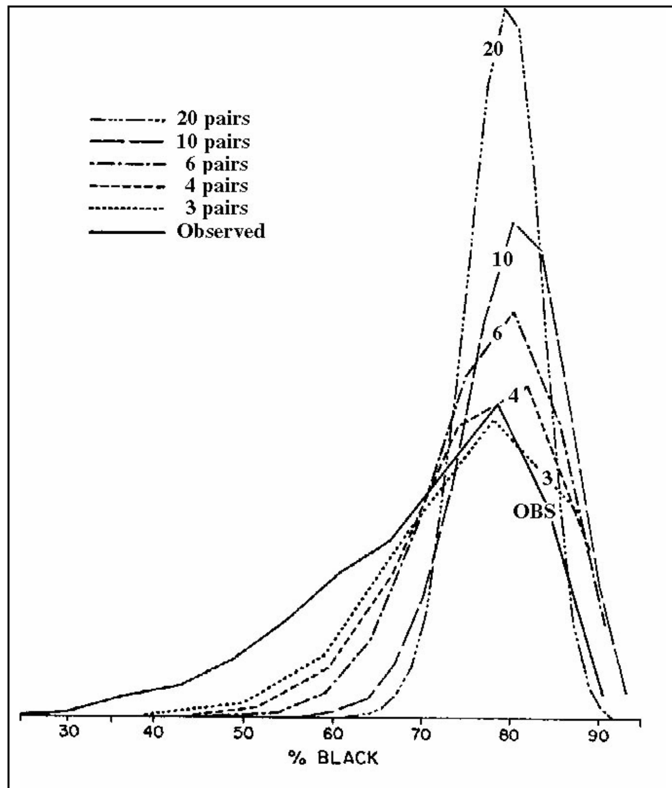

Fig. 2. Estimation of the number of pigmentation genes (taken from Stern, 1970). Model of numbers of genes as the basis of the pigment variation in African Americans given an average (Caucasian) admixture of 20%. The actual observed data (solid line) falls between that predicted for 3 and 4 pairs of genes that exhibit allelic variation between Caucasians and Blacks.

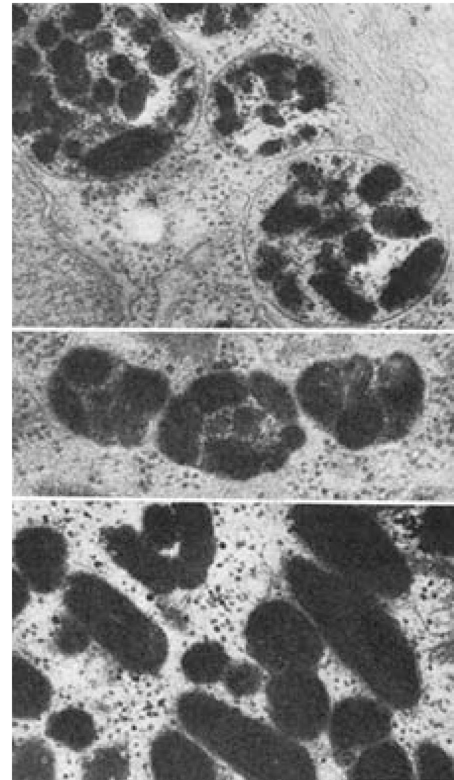

Fig. 3. Melanosome complexes in the skin of three racial types (taken from Szabo et al, 1969). The top panel is from a Caucasian, the middle panel is from an Asian (Chinese), and the bottom panel is from a Black African. The electron micrographs are magnified at approx. 50,000 X. Note the differences in size, shape and melanin content of the melanosomes.

Melanin biosynthesis occurs in melanosomes, an organelle thought to be derived from the fusion of at least two vesicle compartments, including endosome-derived vesicles that share some features with lysosomes (Orlow 1995; Marks & Seabra, 2001). Proteins that are specifically involved in melanin biosynthesis include tyrosinase, the product of the mouse TYR gene (Kwon et al, 1987) and tyrosinase related proteins, TYRP1 (Jackson, 1988) and TYRP2 (Tsukamoto et al, 1992). Tyrosinase is the rate-limiting enzyme in the biosynthesis of melanin and catalyzes the initial step: tyrosine to DOPAquinone and Tyrp1 and Tyrp2 may modify intermediates in melanin synthesis (Hearing and Tsukamoto, 1991; Cooksey et al, 1997). Although tyrosinase and the tyrosinase related proteins are critical to melanin biosynthesis, polymorphic variations in these genes have not been associated with specific normal pigmentation types. Moreover, no genetic linkage studies have demonstrated them to be important factors in normal pigmentation variation, although mutations

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

in the TYR and TYRP1 genes can cause forms of albinism. Indeed, levels of tyrosinase transcripts and tyrosinase protein are essentially the same in Caucasian and Black skin, although enzyme activity *in situ* varies considerably (Fuller et al, 2001). Thus, other genes must encode proteins that regulate tyrosinase function. It is as if tyrosinase functions as an “on/off” switch in melanin biosynthesis, with tyrosinase regulation mediated by other gene products.

Mature (highly pigmented “stage 4”) melanosomes derive from earlier, less pigmented stages. The first recognizable stage is the pre-melanosome that is distinguished by a disorganized lumenal matrix and no melanin. Stage I melanosomes also lack melanin, but are defined by a parallel array of the lumenal matrix. Stages 2-3 are characterized by increased melanin deposition and stage 4 melanosomes contain the most melanin. Although melanosomes are acidic (Devi et al, 1987; Bhatnagar et al, 1993; Ramaiah, 1996), the lumenal pH is probably not static as melanosomes develop.

To make melanin: 1) tyrosinase must be correctly targeted to the melanosome; 2) tyrosine must be transported into the melanosome, and; 3) conditions that favor the enzymatic activity of tyrosinase must exist inside the melanosome. All of these processes are affected by pH (Devi et al, 1987; Bhatnagar et al, 1993; Ramaiah, 1996; Potterf et al, 1998; Ancans & Thody, 2000; Ancans et al, 2001; Fuller et al, 2001; Manga et al, 2001). The product of the mouse *p* (*pink-eyed dilution*) or human P gene is likely an anion transporter (Gardner et al, 1992) that plays a critical role in melanin biosynthesis through the regulation of melanosome pH (Puri et al, 2000). The mouse *underwhite* (*uw*) gene once known as MATP, now known as SLC45A2, also plays a key role in melanin biosynthesis (Sweet et al, 1998; Lehman et al, 2000). SLC45A2 likely mediates the transport of a critical substance across the melanosome membrane that regulates the size and shape of the melanosome (Newton et al, 2001). Mutations in tyrosonase, the P gene and SLC45A lead to albinism (Fig. 4). Thus, it is reasonable to assume that polymorphic variation in the human homologues of these genes may play a role in the normal variation of human hair, skin and eye color.

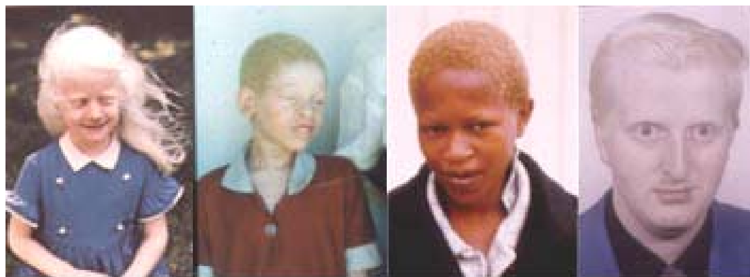

Figure 4. The four known forms of OCA or oculocutaneous albinism . Images from **left to right**: A Caucasian with OCA1, the result of null mutations in the tyrosinase (TYR) gene associated with the complete absence of melanin pigment. An African with OCA2, the result of null mutations in the P gene associated with residual (mainly yellow-red) melanin pigment. An African with OCA3, the result of null mutations in the TYRP1 gene associated with reduced pigmentation (especially of the hair) with residual (mainly yellow-red) melanin pigment. A Turk with OCA4, the result of null mutations in the SLC45A2 gene associated with residual (mainly yellow-red) melanin pigment. In the same racial/ethnic groups, OCA4 and OCA2 have indistinguishable phenotypes.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Other pigmentation genes encode ligands and receptors that signal melanocytes to make specific types of melanin, namely eumelanin (brown/black melanin) or pheomelanin (yellow/red) melanin. Most mammals have banded hairs in which these types of melanin alternate. Several genes are involved in the eumelanin/pheomelanin switch, e.g., MSH, encoded by the POMC gene (reviewed by Eberle, 1988), and its receptor, MC1R (melanocortin receptor 1; Robbins et al, 1993) located on the melanocyte membrane. Several specific variants in the human MC1R gene are associated with red hair (Valverde et al, 1995; reviewed by Rees, 2000; Flanagan et al, 2000; Sturm et al, 2001) and a combination of alleles of P and MC1R was recently found to be associated with normal variation in skin pigmentation in a Tibetan population (Akey et al, 2001). No variations in POMC are known that correlate with normal human pigment variation.

Activation of MC1R by MSH leads to an increase of cAMP and eumelanin production, whereas antagonism by agouti signaling protein (ASP), a paracrine signaling molecule and the product of the mouse *a* (agouti) locus results in pheomelanin production (reviewed by Furumura et al, 1996; Brilliant and Barsh, 1998). No variations in ASP are known that correlate with normal human pigment variation. The downstream targets of the cAMP signaling are not well characterized, but probably include long term effects on transcription and short term effects that may result from phosphorylation of target proteins by cAMP-dependent protein kinase A (PKA). Additional regulation of melanin biosynthesis is demonstrated by the UV (tanning) response that includes signaling through PKA and PKC-mediated protein phosphorylation (Park & Gilchrest, 1999). One potential target of both the PKA and PKC signaling pathways is the SLC45A2 protein. Indeed, we have found genetic evidence for the interaction (in mice) of SLC45A2 with both MC1R and P (Lehman et al, 2000).

Racial differences. Racial differences in pigmentation are correlated with melanosome shape, size and packaging (see Fig. 3). Tyrosinase activity *in situ* varies considerably in different skin types, but no differences in tyrosinase protein sequence, transcription levels, or protein levels underlie racial differences in pigmentation. There is considerable evidence that melanosome pH varies among racial groups (Fuller et al, 2001) and controls tyrosinase activity. As the P protein is a key regulator of melanosome pH (Puri et al, 2000) and the SLC45A2 gene very likely regulates the size and shape of melanosomes (Lehman et al, 2000; Newton et al, 2001), these genes are excellent candidates for regulating constitutive pigmentation.

The *pink-eyed dilution* gene. Mice lacking expression of the *pink-eyed dilution*, *p*, gene exhibit severe hypopigmentation (Brilliant, 1992; Lyon et al, 1992). Humans deficient for the P protein have the most common form of albinism, OCA2 (Gardner et al, 1992; Rinchik et al, 1993; Lee et al, 1994a; Lee et al, 1994b; Durham-Pierre et al, 1994; Brilliant et al, 1994; Puri et al, 1997; Lund et al, 1997; Oetting et al, 1998b; Brilliant, 2001). The reduction in eumelanin (black

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

pigment) is greater than the reduction in pheomelanin (yellow/brown pigment) (Prota et al 1995; Ozeki et al, 1995). The action of the P gene is autonomous to melanocytes, as demonstrated by skin transplantation studies (Stephanson and Hornby, 1985). Our analysis of P cDNA predicted a protein with twelve membrane-spanning domains (Gardner et al, 1992). From the predicted protein structure and the phenotype of melanocytes with P mutations, we (Gardner et al, 1992; Rosemblat et al, 1994) and others (Rinchik et al, 1993: Lee et al, 1995) hypothesized that the P gene encodes a transport or pore protein critical to melanocyte function. Consistent with this prediction, we localized the P protein to the melanosome membrane (Rosemblat et al, 1994). Sidman and Pearlstein (1965) suggested that the P protein might play a role in transport of tyrosine based on their observation that retinal melanocytes (in organ culture) from P-deficient mice became pigmented in the presence of very high concentrations of tyrosine. This observation was recently confirmed using cultured mouse skin melanocytes (Rosemblat et al, 1998; Potterf et al, 1998). However, our past results (Gahl et al, 1995) and those of other researchers (Potterf et al, 1998) clearly demonstrate that the P protein does not function as a tyrosine transporter, as both mutant cells (lacking P protein) and wild-type cells (expressing the P protein) exhibit the same properties of tyrosine transport. It is likely, therefore, that the increase in pigmentation detected in melanocytes lacking the P protein when exposed to very high concentrations of tyrosine results from driving the reaction by substantially increasing the substrate. Moreover, the melanin produced from excess tyrosine treatment does not appear to be synthesized in melanosomes, and may result from the activity of mis-localized tyrosinase (Potterf et al, 1998; Manga et al, 2001).

We formulated an alternative hypothesis of P protein function based on sequence similarities between the P protein and certain anion transport channels in bacteria and yeast. We hypothesize that the P protein functions as an anion transporter in the melanosome membrane (Gardner et al, 1992; Rosemblat et al, 1994) that helps to regulate melanosome pH (Puri et al, 2000; Brilliant, 2001; appendix copies). Effects of pH on tyrosinase targeting to the melanosome and tyrosinase enzyme activity have been extensively reported (Devi et al, 1987; Bhatnagar et al, 1993; Ramaiah, 1996; Potterf et al, 1998; Ancans & Thody, 2000; Ancans et al, 2001; Fuller et al, 2001; Manga et al, 2001).

Though distinct, melanosomes and lysosomes share at least partial endosomal origins and both types of organelles are characterized by an acidic lumen (Orlow, 1995; Marks & Seabra, 2001). Melanosomes have been reported to be very acidic, i.e., pH 3.5 (Tripathi et al, 1988; Moellman et al, 1988; Devi et al, 1987; Bhatnagar et al, 1993). Early stage melanosomes may be the most acidic with later stage melanosomes exhibiting a more neutral pH associated with higher tyrosinase activity. The identity of the molecular components that mediate the acidic environment of endosome-derived organelles is not known, but it is thought that an anion channel is essential for the acidification of vacuole compartments by ATP-driven proton pumps present in endosomes, Golgi-derived vesicles and lysosomes (Al-Awqati, 1995). Anion (Cl^-, $SO_4^=$, or HCO_3^-) conductance provides the compensating charge balance to electrogenic proton

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

transport. Thus, without anion conductance, protons cannot accumulate and melanosomes cannot become acidic. Acidification of various intracellular compartments is important for a number of processes including receptor-mediated endocytosis, receptor recycling and membrane trafficking within the cell (Dautry-Varsat, 1983). Whether or not there is a melanosome-specific ATP-driven proton pump remains to be determined. Nevertheless, the melanosome has a unique membrane protein, the P protein, that is very likely an anion transporter that plays a role in the pH regulation of melanosomes (Puri et al, 2000; appendix copy). The regulation of pH is expected to have a dramatic effect of tyrosinase enzyme activity and/or its targeting to the melanosome. Indeed, melanosome pH differences are the major difference between melanosomes from Caucasian and Black individuals (Fuller et al, 2001). Thus, the P protein is can modulate tyrosinase activity and pigment production. Moreover, genetic evidence suggests that it plays a major role in normal human pigmentation variation (Eiberg & Mohr, 1996; Akey et al, 2001).

The human P gene. The size of the human P gene transcript is 3.4 kb and encodes a predicted protein of 838 amino acids (Rinchik et al, 1993). Like the mouse protein, the human P protein is predicted to have 12 membrane spanning domains and sequence similarity to a group of transport proteins that includes anion transporters (Gardner et al, 1992; Rinchik et al, 1993; Rosemblat et al, 1994; Lee et al, 1995). The P gene is encoded by 24 exons (plus one alternate exon that contains an in-frame stop codon corresponding to IR10-1, originally isolated as an anonymous genomic clone) and spans approximately 400 kb of genomic DNA (Lee et al, 1995). The human proximal promoter region contains sequences that might be used as binding sites for transcription factors, including the following motifs: one AP4 site, four discrete and one complex AP2 site, one CF1 site, one GCF site, three SP1 sites, and one TFIID site. No TATA or CCAAT motifs or melanocyte-specific motifs (i.e., M box) were found (Lee et al, 1995). Many diverse mutations in the P gene have been reported in individuals with OCA2 from diverse ethnic backgrounds (Oetting & King, 1999; Oetting et al, 1998b; Spritz et al, 1997). The missense mutations described to date do not seem to cluster in any specific region of the peptide, as observed for tyrosinase, but most mutations described so far are in the carboxy half of the polypeptide that contains the majority of the 12 membrane spanning domains. A significant portion of the P missense mutations are found at amino acids conserved between the mouse and human P genes and a group of homologous transport proteins (Lee et al, 1995).

In addition to OCA2 (see Fig. 4), mutations of the P gene are also associated with two other albinism phenotypes. BOCA (brown oculocutaneous albinism) is associated with heterozygousity for P gene alleles, one of which is null, the other having partial function (Kerr et al, 2000). Additionally at least one case of AROA, or autosomal recessive ocular albinism is the result of P gene mutations (Lee et al, 1994a). Thus, the phenotypic range of P gene mutations is broad.

It is very likely that the P gene is one of the key genes that mediates normal variation in human pigmentation (Eiberg & Mohr, 1996; Akey et al, 2001).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Indeed, a gene associated with brown eyes (total brown iris pigment) and a gene associated with brown hair map to 15q, with the P gene as the prime candidate (Eiberg & Mohr, 1996). Alleles of P directly affect eye pigmentation in mice (Orlow & Brilliant, 1999). Also, constitutive skin pigmentation in a Tibetan population was found to be correlated with specific alleles of P and MC1R (Akey et al, 2001), although the P alleles studied were not associated with amino acid variations of the P protein. We (Brilliant et al, unpublished) have found direct evidence of P gene involvement in normal human hair color (Fig. 5).

Figure 5. Two Indonesian cousins variant for hair color. The girl on the left has the typical pigmentation seen in the ethnic Minahassan people of Indonesia. Her cousin is a carrier for a null allele of the P gene, demonstrating that in this population (of uniformly black hair) the P gene is semidominant. Two normal doses of the Indonesian allele are associated with black hair, whereas one dose of the normal Indonesian allele is associated with brown hair.

A large number of polymorphisms in the P gene were previously described (**Table 5**); some of these may be predictive of skin pigmentation in the context of specific MC1R alleles (Akey et al 2001); other polymorphisms that are yet to be determined (e.g., in the proximal promoter region) may be more directly associated with pigmentation variation. One polymorphism at amino acid 305 (R305W) is especially intriguing. Most Caucasians (80%) have an arginine at this site, while most Blacks (90%) have a tryptophan at this site. Because this is a non-conservative change (difference in charge), there may be a functional difference between these two forms of the protein.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 1. Known P Gene Polymorphisms

Description	Nucleotide Change*	Effect on coding Sequence	PCR exon	♦ 10% polymorphism*	Reference
IVS3 +42	G or T at 326+42	None	3		Oetting et al., 1998
IVS4 -54	T or A at 516-54	None	5		Oetting et al., 1998
IVS5 -39	T or C at 574-39	None	6		Kerr et al., 2000
IVS5 -18	C or T at 574-18	None	6		Lee et al., 1995
IVS7 +23	A or T at 807+23	None	7		Lee et al., 1995
IVS7 +25	G or C at 807+25	None	7		Oetting et al., 1998
D257A	A or C at 770	Asp or Ala at 257	7		Oetting et al., 1998
R305W	C or T at 913	Arg or Trp at 305	9	C, B, I	Lee et al., 1995, Kerr et al., 2000
L308L	C or T at 922	No change Leu308	9		Oetting et al., 1998
321Gln321	G or A at 963	No change Gln321	9		Kerr et al., 2000
A336A	C or T at 1007	No change Ala336	9		Kerr et al., 2000
Y342Y	C or T at 1026	No change Tyr342	9		Lee et al., 1995, Kerr et al., 2000
A355A	G or A at 1065	No change Ala355	10		Lee et al., 1995, Kerr et al., 2000
G371G	C or T at 1113	No change Gly371	10		Lee et al., 1995
L376L	G or T at 1128	No change Leu376	10		Oetting et al., 1998
IVS11 -4	A or G at 1183-4	None	12	C, B, M, I	Lee et al., 1995, Kerr et al., 2000
IVS12 +19	G or A at 1239+19	None	12		Kerr et al., 2000
R419Q	G or A at 1256	Arg or Gln at 419	13	C,I	Lee et al., 1995
L440F	G or C at 1320	Leu or Phe at 440	13	C	Lee et al., 1995
IVS13 +4	C or A at 1364+4	None	13		Lee et al., 1995, Kerr et al., 2000
IVS13 +26	A or G at 1364+26	None	13	C, B, A, M, I	Lee et al., 1995, Kerr et al., 2000
IVS13 -15	C or T at 1365-15	None	14	C, M, I	Lee et al., 1995, Kerr et al., 2000
IVS13 -13	A or T at 1365-13	None	14		Oetting et al., 1998, Kerr et al., 2000
G458G	G or A at 1374	No Change Gly458	14	B	Lee et al., 1995
IVS14 +5	G or A at 1503+5	None	14		Kerr et al., 2000
C517C	C or T at 1551	No change Cys517	15	C, B, A, M, I	Lee et al., 1995, Kerr et al., 2000
IVS16 -49	C or A at 1785-49	None	17		Kerr et al., 2000
IVS16 -47	A or G at 1785-47	None	17	C, M, I	Lee et al., 1995, Kerr et al., 2000
H591H	C or T at 1773	No change His591	16		Oetting et al., 1998
H615R	A or G at 1844	His or Arg at 615	18	A	Lee et al., 1995
V629V	G or T at 1887	No change Val629	18		Lee et al., 1995, Kerr et al., 2000
IVS18 +24	C or G at 1951+24	None	18		Kerr et al., 2000
IVS18 -59	T or - at 1952-59	None	18		Lee et al., 1995
A686A	A or C at 2058	No change Ala686	19	B	Lee et al., 1995, Kerr et al., 2000
IVS21 +22	A or T at 2139+22	None	21		Lee et al., 1995, Kerr et al., 2000
I722T	T or C at 2165	Ile or Thr at 722	22	C, B	Lee et al., 1995, Kerr et al., 2000
S736S	G or A at 2208	No change Ser736	22	C, B	Lee et al., 1995, Kerr et al., 2000
IVS22 +18	A or G at +18	None	22		Lee et al., 1995, Kerr et al., 2000
IVS22 +25	G or C at +25	None	22	C, M, I	Lee et al., 1995, Kerr et al., 2000
A776A	T or C at 2328	No change Ala776	23	C, B, I	Lee et al., 1995, Kerr et al., 2000
G780G	C or T at 2340	No change Gly780	24	C, B, A, M, I	Lee et al., 1995, Kerr et al., 2000
S788S	G or A at 2364	No change Ser788	24	C	Lee et al., 1995, Kerr et al., 2000

* C: Caucasian; B: Black; A: Asian; M: Middle Eastern; I: Indo-Pakistani.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

The mouse *uw* (*underwhite*) gene and its alleles. Among the mouse mutations that lead to severe hypopigmentation, without other phenotypes (i.e., effects restricted to melanocytes), the *uw* (*underwhite*) or SLC45A2 gene is distinguished by an unusual series of alleles (recessive and dominant) with a wide spectrum of pigmentation phenotypes (**Fig. 6**).

Figure 6. Mice variant for alleles of *uw* (SLC45A2). Left to right: $+/uw^d$; uw^d/uw^d; Uw^{dbr}/Uw^{dbr}; $Uw^{dbr}/+$; uw/uw; uw^j/uw^j; and uw^j/uw^j. The genetic background of all mice is non-agouti black and wild-type would have the same color as the uw^d carrier mouse on the far left.

On a non-agouti black (e.g., C57BL/6) background, mice homozygous for the original mutant allele, *uw* (Dickie, 1964), are as severely hypopigmented as *p* mutant mice, whereas mice heterozygous for the semidominant allele, $Uw^{dbr}/+$ are brown. We have recently made significant progress in the characterization of the *uw* phenotype, including a detailed phenotypic description of all known alleles. Mutant *uw* alleles are associated with small, crenated, minimally pigmented melanosomes (Sweet et al, 1998). We have also established that the action of the *uw* gene is autonomous to melanocytes (Lehman et al, 2000) in transplantation assays. Most importantly, we (Newton et al, 2001) have identified the protein encoded by *uw* as MATP (membrane spanning transport protein) also known as AIM-1 (antigen in melanoma 1) and now known as SLC45A2, a gene expressed exclusively in melanocytes and melanoma (Harada et al, 2001; GenBank #AF172849). **N.B.:** There are two other distinct genes called AIM-1 in the literature (Katayama et al, 1998; Teichmann et al, 1998).

SLC45A2 is homologous to known transporters, especially plant H^+/sucrose symporters that often serve as osmotic regulators (Stadler et al, 1999). Previously, we have found that in the absence of a functional SLC45A2 protein, melanosomes are small, irregularly shaped (raisin-like, crenated structures) with reduced pigment (Sweet et al, 1998). Thus, it is possible that the SLC45A2 protein functions in osmotic regulation of melanosomes. Drug and sugar mediated disruption of the osmotic balance of endosomes and lysosomes has effects on protein turnover and lumenal pH (Isaac et al, 2000; Schreiber and Haussinger, 1995; Schreiber et al, 1996) and the transport of amino acids across membranes is affected by osmotic imbalance (Gomez-Angelats et al, 1997) and pH (Potterf et al, 1996). We believe that the SLC45A2 protein is an osmotic regulator of the melanosome, mediating the size and shape of these organelles. Variations in the size and shape of melanosomes are the basis of racial pigmentation differences (Fig. 2). Similar variations in the size and shape of melanosomes are seen in the *uw* series of alleles. Moreover, the hypopigmentation of these mutations is correlated with abnormally shaped (some almost raisin-like) and partially pigmented melanosomes (Fig. 7).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

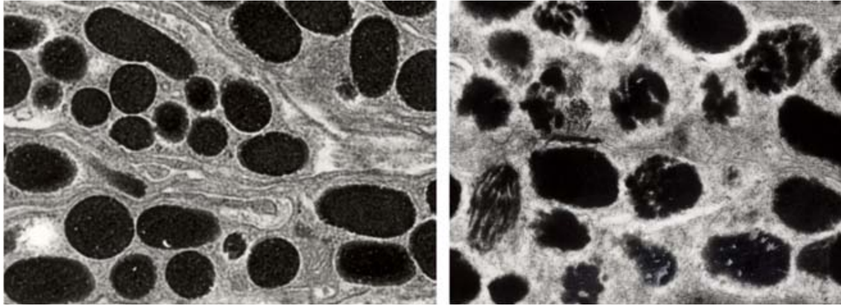

Fig. 7. Electron micrograph of melanosomes within the choroid layer of an adult mouse eye. LEFT: C57BL/6 wild-type showing typically well-pigmented and round melanosomes. RIGHT: *uw*/*uw* mutant showing irregularly shaped melanosomes with uneven distribution of melanin. (Adult uw^d mice have dark eyes.) Original figures of eye melanocytes from these and other *uw* alleles are in Sweet et al, 1998.

To gain some insights as to the position of SLC45A2 in the genetic hierarchy of pigmentation genes, two double mutant mouse lines were generated, $uw^d/uw^d;Mc1r^e/Mc1r^e$ and $uw^d/uw^d;p^{cp}/p^{6H}$ (Lehman et al, 2000, appendix copy). The singly mutant $Mc1r^e/Mc1r^e$ mouse is sooty-yellow; the singly mutant uw^d/uw^d mouse is light brown; while the singly mutant *p*/*p* mouse is light gray. The doubly mutant $uw^d/uw^d;Mc1r^e/Mc1r^e$ mouse has a very light coat with dark eyes. The doubly mutant $uw^d/uw^d;p^{cp}/p^{6H}$ mouse lacks all visible pigment in the coat and eyes. Its white coat and pink eyes are indistinguishable from the *albino* or tyr^c/tyr^c mouse. Thus, the double mutants demonstrated additional loss of pigment, suggesting that the action of the SLC45A2 gene product is at least partially independent of the action of the P and MC1R proteins.

We identified the *uw* gene product. Previously, we published the first mapping of *uw* relative to molecular markers (Sweet et al, 1998) and had subsequently further refined its map position. The refined map allowed us to position the human UW gene on chromosome 5p. We then evaluated human ESTs and known genes in the syntenic region as candidates for UW based on expression patterns and implied function. Only one candidate gene from the homologous region, AIM-1 (antigen in melanoma 1), was found exclusively in libraries derived from melanomas or tissues containing melanocytes. Originally defined only by a GenBank report (Harada et al 2001; GenBank #AF172849), a recent publication on AIM-1 details how it was detected and isolated by T cell activity against a melanoma. The expression of the AIM-1 transcript appears to be restricted to melanocytes and melanoma (Harada et al, 2001). **N.B.:** There are two other distinct genes called AIM-1 in the literature (Katayama et al, 1998; Teichmann et al, 1998). To distinguish this protein from the other "AIM-1" proteins, we decided to rename it. In light of its predicted membrane spanning structure and high homology to transport proteins, we (in consultation with the mouse nomenclature committee) have renamed this AIM-1 protein as MATP (for membrane associated transporter protein). The nomenclature committee has recently renamed it to SLC45A2.

To directly evaluate SLC45A2 as the *uw* gene, we performed Northern analysis utilizing RNA from the eyes of *uw*/*uw*, Uw^{dbr}/Uw^{dbr}, uw^d/uw^d, and wild-type mice using the SLC45A2 cDNA as a probe. The most severe allele, *uw*/*uw* gave no detectable expression of the SLC45A2 transcript (Newton et al., 2001; presumably the mutant transcript is unstable due to a seven-base pair deletion in

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

the coding sequence of exon 3; Du & Fisher, 2002). SLC45A2 transcripts are expressed from the two other alleles that have intermediate pigmentation, but each mutant allele has a unique point mutation that leads to a non-conservative amino acid substitution (uw^{d}: serine to proline; Uw^{dbr}: aspartic acid to asparagine), consistent with their partial function (Newton et al, 2001). Based on the mouse phenotype, we considered the human SLC45A2 gene as a candidate gene in oculocutaneous albinism (OCA) with a phenotype similar to OCA2, P gene albinism. We sequenced the SLC45A2 gene in several OCA individuals that had some pigment, a few of whom had been previously screened for mutations on the P gene. One OCA individual (with distantly related parents) was found to be homozygous for a splice site mutation of the SLC45A2 gene and did not have a P gene mutation. This individual defined a new form of OCA, OCA4 (Fig. 4, right panel; Newton et al, 2001). We identified several other individuals with OCA and unique amino acid changes in MATP, however, we did not yet know if these are polymorphisms or mutations. Other researchers (Fukamachi et al, 2001) have found that mutations in the Medaka fish paralog result in hypopigmented fish. A publication subsequent to ours confirmed our mouse data and described potential transcriptional regulation of SLC45A2 by the melanocyte transcription factor, MITF (Du & Fisher, 2002). Intriguingly, the SLC45A2 protein may be regulated by MC1R (melanocortin receptor 1). SLC45A2 has two consensus sites for cAMP-dependent protein kinase A (PKA) phosphorylation: Arg-X-X-Ser/Thr (e.g., Glass et al, 1986) and three consensus sites for protein kinase C (PKC) phosphorylation: Ser/Thr-X-Arg/Lys (e.g., Quest et al, 1997). The signaling of MSH through the MC1R results in an increase of cAMP and favors production of tyrosine-based eumelanin in melanocytes. Antagonism of MSH binding to the MC1R by ASP (agouti signaling protein) prevents this increase of cAMP and favors production of cysteine-rich pheomelanin (reviewed by Furumura et al, 1996; Brilliant and Barsh, 1998). Although transcriptional regulation is thought to play a role in the eumelanin/pheomelanin switch, it is possible that modification (phosphorylation) of existing proteins may also be regulated. We note that the hypothesis that regulation of SLC45A2 by cAMP is consistent with the phenotype of mice doubly mutant for *uw* and $Mc1r^{e}$ (Lehman et al, 2000). The PKA and PKC sites are indicated in Fig. 8, next page). It is known that variations in the human MC1R gene are associated with red hair (Valverde et al, 1995). If SLC45A2 is in this pathway, then it is also possible that specific variants in SLC45A2 might be associated with red hair or other normal pigmentation variants.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

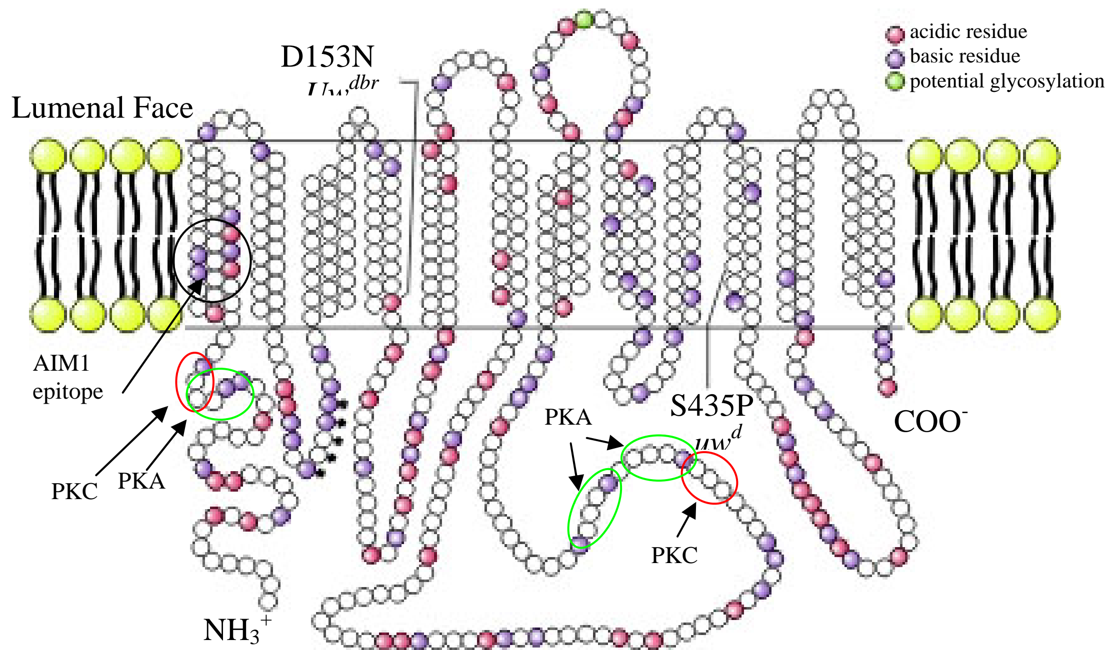

Fig. 8. Mouse SLC45A2 structure with important features/epitopes shown. **The "AIM1" epitope indicates the homologous region of the human protein recognized by a T cell clone in a melanoma patient (Harada et al, 2001). The sites of the point mutations are indicated: D153N, an aspartic acid (conserved in +H/sugar transporters) to asparagine substitution near the cytoplasmic face of transmembrane domain 4 of *Uw*^{dbr}; and S435P, a serine to proline substitution in transmembrane domain 10 of *uw*^{d} (Newton et al, 2001, appendix copy). Also indicated are potential PKA and PKC sites. The starred amino acid residues (between transmembrane domains 2 and 3) indicate the conserved sugar recognition sequence.**

Recently, another critical gene involved in pigmentation was discovered that is responsible for the hypopigmentation mutation, *Golden*, in the zebrafish. The SLC24A5 protein has high homology to known Sodium/Calcium ion exchangers and likely functions as such. Two main variants of this gene, SLC24A5, are found in human populations. The Ala111 (ancestral) allele is found in 93 to 100% of African, Indigenous American, and East Asian population samples and the Thr111 variant allele is found in 98.7 to 100% of several European-American population samples. It is very likely that these coding polymorphisms lead to functional and phenotypic variation in melanocytes.

Combining what we know about the phenotypes of mice, people, and melanocytes with the suspected functions of the critical proteins, we have developed a model of melanosome structure and function (Fig. 9, next page). This model predicts that P and SLC45A2 control melanosome shape, size and the rate of melanin production. Moreover, there is potential regulation of these two gene products via MC1R (through transcriptional control of P gene expression and potential phosphorylation of SLC45A2 via PKA and/or PKC). We know that these genes interact to affect overall pigmentation in the mouse (Lehman et al, 2000) and that P and MC1R interact to affect at least skin pigmentation in humans (Akey et al, 2001). Thus, we now have an emerging picture of the key genes involved in regulating human pigmentation.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

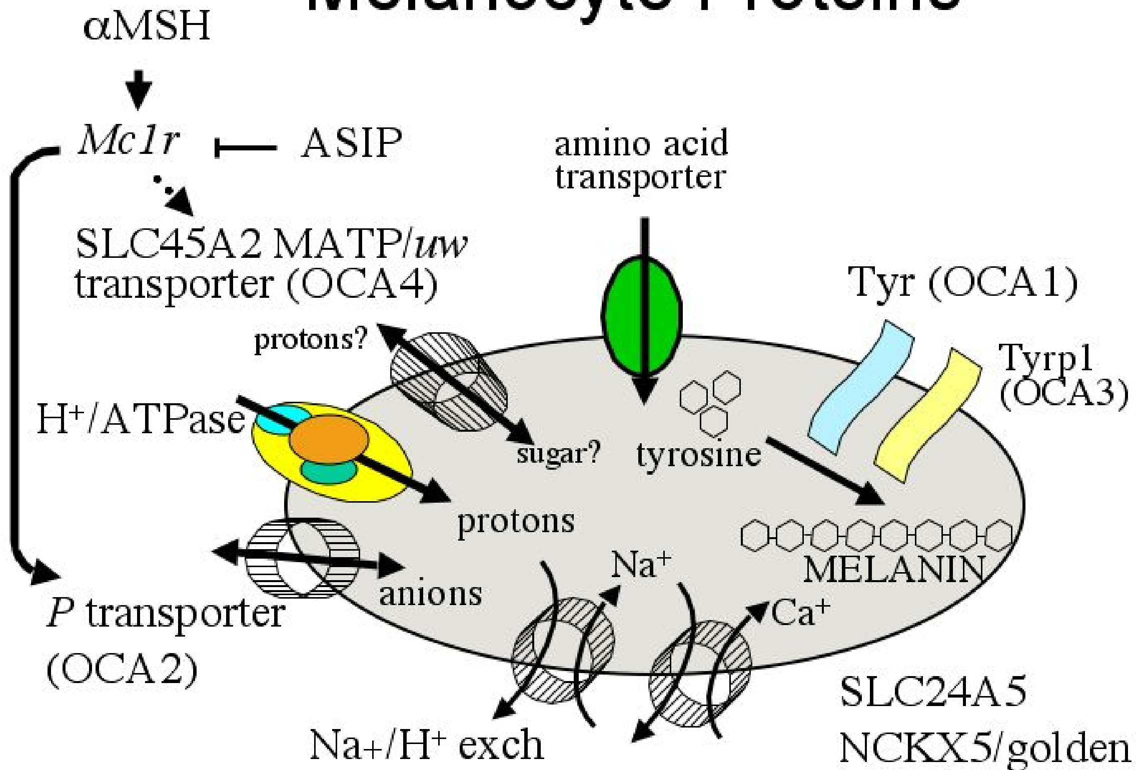

Fig. 9. A simplified melanosome with key proteins is shown. The melanosome organelle is the site of melanin biosynthesis and storage. The vacuolar H^+/ATPase brings in protons and the P protein facilitates anion transport needed to maintain electrogenic neutrality. The lumen of a melanosome is acidic; however, we note that our model allows for a dynamic change in the lumenal pH as the melanosome matures. This is consistent with the relationship between pH and tyrosinase activity *in vitro* (Townsend et al, 1984; Ramaiah, 1996) and the increase in tyrosinase activity in mature melanosomes seen after treatment with agents that disrupt the pH gradient (Ancans and Thody, 2000). Thus, the P protein likely plays a major role in the normal regulation of melanosome size and melanin content. Tyr (tyrosinase), Tyrp1 (tyrosinase-related protein 1), P and SLC45A2 are transmembrane proteins and each underlies a form of oculocutaneous albinism (OCA; see Fig. 3). The barrel structures on the left represent the two transport proteins under study in this application, MATP and P that regulate tyrosinase activity and the size and shape of melanosomes. The melanosome membrane location for the SLC45A2 is based solely on the phenotype of mutant *uw* melanocytes; we do not yet know if SLC45A2 actually resides in the melanosome membrane. We also do not know the identity of the transported solute or the directionality of its transport.. There are several Na+/H+ exchangers that are thought to reside on the melanosome membrane and one or more of these may work in concert with the SLC24A5 protein to regulate ion flow across the melanosomal membrane. Thus, ultimately regulation of pH and Ca+ ion concentration are regulated in the melanosome MC1R signaling likely controls P gene expression and may control SLC45A2 gene expression (Lehman et al, 2000) and/or phosphorylation (Newton et al, 2001).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

In sum, three transporter genes are known to regulate human pigment variation (P, SLC45A2 and SLC24A5). In the past, we have cloned two of these genes (P and SLC45A2) and, together with our colleagues, have described numerous mutations and normal polymorphic variations in these genes (Gardner et al, 1992, Durham-Pierre et al, 1994; Brilliant et al, 1994; Puri et al, 1997; Lund et al, 1997; Oetting et al, 1998b). In addition, it is also likely that the receptor/signal modulation proteins, MC1R and ASIP, play a major role in determining pigmentation type in human beings. Variation in human skin pigmentation may be controlled by as few as 3 or 4 genes (Stern, 1970). It is reasonable to assume that variations in the P, SLC45A2, SLC24A5, MC1R and ASIP genes will be good predictors of human pigmentation, including hair, eye and skin color. This knowledge can be used in forensic studies as an identification tool. If the only information about a suspect is DNA (contained in a blood specimen, semen sample or tooth), a likely description of the donor can be derived: e.g., given specific polymorphisms in these three genes, there is a 60% chance that the donor has black hair, an 50% chance that the donor has brown eyes and a 80 % chance that the donor has a medium complexion. This type of prediction may even be more accurate than an eyewitness account and, as such, can be a powerful tool in suspect identification. Proof that it is a specific person with black hair, brown eyes and medium complexion will still need to come from other assays in current (forensic) use.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

EXPERIMENTAL PLAN

<u>Phase 1</u>**: To determine the spectrum of polymorphisms of genes (MC1R, P and MATP) that are known, or very likely, to be major contributors to normal human pigment variation**. Many polymorphic variants are already known in these genes (especially MC1R), others need to be identified (in the P gene proximal promoter and throughout the MATP gene). Polymorphic variants in a limited number of individuals of diverse ethnic origins and various pigmentation types, will be studied in this phase. We will of course exploit the known polymorphisms of these genes.

<u>Rationale and Strategy</u>: Polymorphic variants in MC1R, P and MATP are the basis for human normal pigmentation variation. We will determine the spectrum of polymorphic variation in the proximal promoter of the P gene and throughout the MATP gene in approximately 300 individuals. These will include approximately 100 African-American samples, 100 Caucasian-American or Caucasian-European individuals and 100 others including Asians and Asian-Americans, Native-Americans, and Hispanic-Americans. These samples are already in hand in our laboratory and some have been used in previous population studies (e.g., Durham-Pierre et al, 1996).

The MC1R gene has been extensively studied and several polymorphism are known that effect pigmentation independent of other genes (Valverde et al, 1995; reviewed by Rees, 2000; Sturm et al, 2001) and in the context of specific P alleles (Akey et al, 2001). No new polymorphisms of the MC1R gene will be sought in this study. Although many polymorphisms of the P gene have been described in people of diverse ethnic backgrounds (Table 1), these are almost exclusively from in, and near, the coding exons. This is because the P gene has been studied in the context of looking for mutations that lead to albinism. (**N.B.:** Past NIH funding for these efforts is the result of NIH's interest in disease-related studies. The types of studies outlined in this proposal are not disease-related and, therefore, do not lend themselves to NIH funding.) To determine if a particular nucleic acid (or amino acid) change results in an inactive product, we typically check to see if it is exclusively found in individuals with albinism or in known carriers of albinism. Those nucleic acid (or amino acid) changes that are found in non-OCA and non-carriers are considered neutral polymorphisms. These are "neutral" polymorphisms with regard to albinism only and may still have a profound affect on pigmentation. We have previously reported that albinism is at one end of the spectrum of pigment variation (Oetting et al, 1996, appendix copy). We will exploit the known polymorphisms of the P gene in Phase 2 of this study. In addition to the known P polymorphisms, we will search for new polymorphisms in the proximal promoter region of P that may modulate P gene expression as part of Phase 1 of this study. Very little is currently known about polymorphism of the MATP gene (Table 2). We will determine the polymorphic variation of MATP in both coding and proximal promoter regions in Phase 1.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Experimental Protocol and Feasibility:
For the P gene, we will determine the sequence of 400 nucleotides of the proximal promoter in all 300 samples. This region contains all of the known regulatory elements of the P gene (Lee et al, 1995). This sequence is very high in G-C content and so we will adjust our sequencing strategy and conditions accordingly (Choi et al, 1999; Buck et al, 1999). The sequences will be determined from PCR products based on a 96-well format. The University of Arizona DNA sequencing facility has the capacity and high throughput for this type of analysis. For the MATP gene, all seven exons will be sequenced as previously described (Newton et al, 2001) in the 300 control individuals. We will also sequence 400 nucleotides of the putative MATP proximal promoter. Intron/exon boundaries for MATP were previously determined by us (Newton et al, 2001) by sequence comparison between the human cDNA sequence (GenBank accession number AF172849) and sequence data deposited in the rough draft of human Chromosome 5p (http://www/ncbi.nlm.nih.gov/genome/guide/human). We identified seven exons spanning a region of nearly 40 Kb and designed PCR primers for amplification of individual exons from genomic DNA as well as internal sequencing primers for direct sequencing of PCR products. MATP exons will be amplified from 10 to 200 ng of genomic DNA using AmpliTaq DNA Polymerase at an annealing $T=55^{o}$ with the primers pairs previously described (Newton et al, 2001). PCR products will then be purified and sequenced directly with the previously identified sequencing primers using ABI 377 and 3700 apparati. The sequence traces are electronically transmitted to our laboratory over the University of Arizona Network. Analysis of those traces is completed using Sequencher and MacVector software. We have extensive experience with both the P and MATP genes (e.g., Gardner et al, 1992, Durham-Pierre et al, 1994; Brilliant et al, 1994; Puri et al, 1997; Lund et al, 1997; Oetting et al, 1998b; Newton et al, 2001) and with all aspects of sequencing large numbers of samples (e.g., Newton et al, 2001). All the necessary samples have are already been collected. We do not anticipate any problems with these analyses.

Phase 2: To correlate polymorphisms in the MC1R, P and MATP genes with pigmentation types in a representative sample of individuals. Although the genetics of human pigmentation is complex, it is likely that the vast majority of human (skin) pigment variation is mediated by only three or four genes (Stern, 1970). The three genes under study are genes that are known, or very likely, to be major contributors to human pigment variation. Statistical analysis will be used to confirm this hypothesis and allow for the prediction of hair, eye and skin pigmentation from a given set of alleles at these loci.

Rationale and Strategy: We will enroll approximately 1,000 people (aged 18 to 40; recruiting mostly college-age students) in this study over a two-year period. We will attempt to obtain a set of individuals that proportionally reflects the population make-up of the United States. Each of these individuals will be characterized phenotypically (pigmentation of the hair, eye and skin) and

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

genotypically (polymorphisms of the MC1R, P and MATP genes). Polymorphic variation will be determined by DNA sequence analysis. The MC1R gene is the most straightforward gene of the three to assay. It has no introns and will be sequenced after PCR amplification using three sets of sequencing primers, revealing all known polymorphic variants linked to pigmentation [e.g., 29insAV60L, D84E, V92M, R142H, R151C, Y152OCH, I155T, R160W, R163Q, 179insC, D294H (Grimmes et al, 2001); Gln163Arg, Val92Met (Akey et al, 2001); and Ile40Thr, Val122Met (Jimenez-Cervantes et al, 2001]. The P gene genotype analysis is more complex, as the gene has 24 exons (Lee et al, 1995). We will examine the two polymorphisms (IVS13-15 and Gly780Gly) previously shown to correlate with skin pigmentation in the context of specific MC1R alleles (Akey et al, 2001). These are contained in PCR products of exons 14 and 24, respectively. Other polymorphisms have been described within the PCR products of exons 14 and 24 (Table 1) and we will certainly record those also. We will also look at exon 9, where coding polymorphisms and IVS polymorphisms have been noted in several ethnic groups, including R305W, where arginine predominates in Caucasians (80%) and tryptophan predominates in Blacks (90%) (Lee et al, 1995). If we find polymorphic variation in the proximal promoter of the P gene in the Phase 1 studies, we will include that region for analysis as well. The Phase 1 data for MATP polymorphism will also be incorporated in our strategy for Phase 2 genotyping. We will select at least three variants for analysis from MATP (Table 2; e.g., L65V, F374L and R206R, or others in the proximal promoter). We will prioritize the ascertained polymorphisms based on the predicted types of changes that they make in the protein or its expression and its prevalence in the various populations.

Experimental Protocol and Feasibility:

Phenotype characterization (Hair, eye and skin color determination)

Data from all subjects will include gender, age, and date of sampling. This will allow us to control for sexual dimorphic effects, age effects and seasonal variation, respectively, in the data analysis.

HAIR COLOR: Ascertainment of hair color presents certain challenges: hair color can be changed artificially; hair color can vary with age; hair color likely varies with sex; hair color can vary seasonally; hair color classification can be subjective; and gray hair replaces more pigmented hair as follicular melanocytes die with age. All individuals will be surveyed, noting age, sex and ethnic origin. With regard to hair color we will ask the following questions:

- Is your hair color natural?
- What is the natural color of your hair today?
- What color hair did you have as a newborn/child?
- What does your driver's license say is your hair color?
- Does your hair lighten in the summer/with sun exposure?
- Do you have any gray hair?

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

A trained interviewer will also record their independent assessment of the subject's hair color with regard to standard samples. Subjects with recently dyed hair will not be recruited (i.e., at least the basal 1 cm of hair must be of natural color). However, an objective assay will also be performed: Hair samples will be taken (20-30 mg per donor of the most recent 1 cm growth; approx. 300 hairs, depending on hair thickness) for chemical analyses by our collaborators Dr. Shosuke Ito and Dr. Kazumasa Wakamatsu in the Department of Chemistry of Fujita Health University School of Health Sciences in Aichi, Japan. Drs. Ito and Wakamatsu are the world's leading experts in the chemical analyses of hair melanins. They are able to determine the total melanin content and the relative contributions of pheomelanin (yellow/red melanin) and eumelanin (brown/black) (e.g., Ito, 1993). We believe that this type of chemical analysis provides the most accurate and objective assessment possible of hair color. We have worked with Drs. Ito and Wakamatsu in the past to characterize the effects of mutations in P and MC1R in the context of various mutant and wild-type MATP alleles (Lehman et al., 2000). Letters confirming their role in this study and their CVs are attached to this application.

The effects of aging on hair color (graying) are somewhat problematic. Gray hair results from a reduction in melanin incorporation in the developing hair with the death of follicular melanocytes. The absolute amount of melanin decreases, but the ratio of pheomelanin to eumelanin stays the same. Thus, at least one parameter of the chemical analysis of hair will remain the same. Moreover, we will confine the age of the subjects to be recruited to 18 to 40 years, and so gray hair is not expected to present a significant problem. Additionally, since we will note those subjects with gray hair, we can analyze the data with or without this group.

EYE COLOR: Ascertainment of eye color can also present challenges: eye color can be changed with colored contacts; eye color can change over time; and eye color can be subjective. With regard to eye color we will ask the following questions:

- Is your eye color natural? E.g., do you wear colored contacts?
- What is the natural color of your eyes today?
- What does your driver's license say is your eye color?
- What color eyes did you have as a newborn/child?

A trained interviewer will also record their assessment of the subject's eye color with regard to standard samples. Although image analysis has been employed in attempts to standardize iris color assessments (Takamoto et al, 2001), this technique has not been employed elsewhere and it is relatively expensive and time consuming. We recognize that this will be the most difficult parameter to objectively measure. Nonetheless, we will record these data and determine if we can correlate a measurement of eye color with the genotype data.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

SKIN COLOR: Ascertainment of skin color also presents certain challenges: skin color can be changed artificially, skin color can vary with age, skin color is a sexually dimorphic trait, skin color can vary seasonally, and skin color classification can be subjective. With regard to skin color we will ask the following questions:

- Do you tan, do you sunburn?
- Do you have freckles?
- Are you tan now?
- Do you wear sleeveless blouses (F)? OR Do you wear sleeveless shirt or go shirtless (M)?

A trained interviewer will also record their assessment of the subject's skin color (constitutive pigmentation in non-exposed areas; in this case, the area under the upper arm) with regard to standard classification: Skin Types I-IV. Type I skin is very light, often associated with freckles and does not tan; Type II skin is intermediate with a moderate tanning response; type III skin is darker with a robust tanning response, and type IV skin is very dark in the absence of tanning. However, an objective assay will also be performed: we will employ a spectroscopic measurement of an area of the skin (upper part of the underarm) that is not subjected to tanning using a reflectometer (Photovolt, Inc). This type of analysis and instrument were successfully used in a study that linked alleles of MC1R and P to skin pigment type in a Tibetan population (Akey et al, 2001).

Potential Problems As noted earlier, it has been suggested that only a few genes are involved in (skin) pigmentation and abundant evidence suggests that the MC1R, P and MATP are probably the three most critical genes in pigmentation variation. However, it is possible that we will not find a significant association between these markers and constitutive pigmentation of the hair, eye and skin. At the very least we should be able to determine the contributions of these three genes to constitutive human pigmentation variation. This, in turn, will allow us to know if we need to extend these analyses to other pigmentation genes and these data can be added to and extended. Thus, our set of characterized subjects will still be useful in future efforts to determine the key genetic markers of pigmentation.

Subject enrollment, interview, informed consent, sample acquisition, and confidentiality assurance Over a two-year period, we will enroll 1,000 subjects of different ethnic groups of both sexes, based on the current US Census Bureau demographic data (http://eire.census.gov/popest/archives/national/nation3/intfile3-1.txt). The breakdown of these data are: 82.2% White; 12.8% Black; 0.9% Native American; and 4.1% Asian/Pacific Islander. Hispanics of any race comprise 11.9% of the total (among them 71% are White; 12.2% Black; 0.7% American Indian; 3.8% Asian/Pacific Islander; the remainder classify themselves as "other").

We will concentrate on recruiting students from The University of Arizona. The student population of the University of Arizona is 67% White; 2.7 % African

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

American; 5.3% Asian American; 2.2% Native American; and 13% Hispanic; 7.6% international (all races) and 2.2% unknown (http://daps.arizona.edu/daps/scat/F00/CurrentEnrollment/Enr-5.Html). While the student population does not exactly fit the average US percentages, we do have good population representation and a total of 34,500 students from which to draw from. We believe that payment of $20 per enrolled subject and advertisement of this study in the student newspaper will bring us all the samples that we will need.

Following informed consent, these subjects will be asked to complete a survey including the above 13 questions, plus data on age, sex and ethnic background. Our independent assessment of the subject's pigmentation status will be noted and we will take a buccal cell sample, measure skin reflectance and take hair samples as noted above. Subjects will be recruited from ads in the student newspaper and will be compensated for their participation ($20 per subject). We expect no problems in enrolling the necessary subjects. We will endeavor to recruit subjects so that our final set reflects the demographics of the US population, as detailed above. The subject's data and DNA sample will be coded and the consent forms will be stored in a confidential file administered by The Department of Pediatrics under current Human Subjects Guidelines.

Sequence determination Buccal cell swabs for DNA analysis will be obtained after informed consent is assured. DNA will be extracted from buccal cells by standard procedures as we have done in past studies (Puri et al, 1997; Newton et al, 2001). Sequencing and sequence analysis will be as described in the previous section and in our publications (e.g., Newton et al, 2001).

Statistical Analyses

The following three sections describe the statistical analyses to be completed for the project.

1. Population-genetic description of the polymorphic sites in each candidate locus

Standard chi-square and likelihood-ratio G tests for departures from Hardy-Weinberg (HW) equilibrium (Lynch and Walsh, 1998; Weir, 2001) will be performed on all polymorphic sites of the P, MC1R, and MATP loci. If the sample consists of random draws from multiple subpopulations (each in HW) then a deficiency of homozygotes is expected, with the deficiency a function of the variance in allele frequency differences (Wahlund 1928). Hence, the simple HW test will immediately provide at least some insight into hidden population structure.

Likewise, linkage disequilibrium values for all pair-wise sites (both within a locus and between loci) will be computed. Again, standard tests for linkage equilibrium (Lynch and Walsh, 1998; Weir, 2001) will be employed. Significant disequilibrium, both for sites within a single gene and for sites on two different genes, can be generated by either selection or (more likely) population substructure.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

2. Single-marker association tests.

Associations between measures of pigmentation in eyes, hair, and skin and particular polymorphic sites will be examined. The preliminary analysis will first examine (using a standard general linear model) the effects of age, sex, and measures of sunlight exposure on the quantitative measures of pigmentation. We will treat these three factors as fixed, and allow for interactions. This will provide an initial model correcting for the effects of these environmental factors. Our first analysis of marker-trait associations will look for associations between the marker alleles at a given site and the residuals between the raw observations and the age-sex-exposure predicted value. Allowing for the possibility of marker-specific effects on any of these three environment factors (for example, an age-marker or marker-exposure interaction), we will also fit a full model with age, sex, exposure, and marker genotype as fixed factors.

Given that the three traits to be measured are correlated, we will, in addition to the single trait analysis, also examine associations between the vector of the three traits and individual markers using general linear models with a multivariate response variable. Ronin et al. (1995) found that inclusion of the correlation structure often improves power.

We fully realize that an association between a trait and marker can result as a simple consequence of population structure. Our plan is to first search for any such associations, generating a small set of candidates for future trials. In particular, the ultimate goal will be to employ the transmission-disequilibrium test (TDT), which formally accounts for population structure (Spielman et al. 1993). Given that this test requires collection of data on known sibs, it is prudent to first see if any associations exist before proceeding with the more demanding task of sib recruitment.

3. Multiple marker analysis

We will determine if there are epistatic interactions between sites. The typical problem of detecting epistatic interactions applies here, namely that the number of possible combinations greatly exceeds the number of distinct full genotypes in the sample and hence there is very little (if any) power to examine all combinations. We will thus focus on two-way interactions between individual sites showing significant associations in single-marker analysis. While such an analysis can potentially miss sites that individually have a very small direct effect, but have a very significant interaction, we can partly control for this by using a decreased stringency on the single markers that we test (i.e., including markers with alpha values of 0.1 or less). Again, false-positives are expected given the number of tests that is likely, but again the goal is to present a limited set of interacting markers for future analysis under a full TDT regiment.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

ORGANIZATIONAL AND MANAGEMENT PLAN; TIME TABLE:

Dr. Murray Brilliant will oversee all aspects of the proposed research. Students will perform the experimental analyses outlined in phase 1 of this study that will be completed within six months of the beginning of this grant. Recruitment and processing of samples of the 1,000 subjects (phase 2) is expected to be completed during the first one and one half years. Dr. Brilliant and students will interview and collect buccal cells and hair samples from the recruited subjects. Students will process the DNA samples. Dr. Brilliant will coordinate the shipping and analysis of the hair samples to the laboratory at Fujita University, where our colleagues Drs. Ito and Wakamatsu will perform the chemical analyses (see letters of collaboration in the appendix). Dr. Meaney, Dr. Walsh and Mr. Roberge will develop the databases and analysis programs that will be used in these studies. Ms. Petitt will enter the data into the databases. The final six months of the study will be devoted to data analysis, publication of our findings and writing a final report to the NIJ. Our results will be disseminated to the forensic field through these publications and through interactions at the NIJ grantee meeting and the annual meeting of the American Academy of Forensic Sciences.

Future Studies/Phase 3: To adapt this analysis to forensic DNA samples. The first two phases of this study are expected to take two years, the time-frame of this application. The results of these first two phases of this study will allow us to predict with some accuracy the hair, eye and skin pigmentation of a particular individual from a DNA sample. However, for this type of analysis to be of use in forensic applications, certain improvements will have to be made regarding the standardization of the assays and minimizing the amount of DNA sample needed. To accomplish this, the most robust markers of pigmentation will be selected. Assay conditions will be adjusted to accommodate multiplexing and other more sensitive assays (e.g., use of specific labeled primers). Additionally the assays must be able to work with forensic sample sources, including those of mixed DNAs. We will reapply for an additional year of funding for these studies, taking in to account recent and future progress in the technical aspects of forensic DNA analyses. In this regard, we will develop a working relationship with a state DNA lab or the FBI DNA lab.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

ANTICIPATED CONTRIBUTION TO CRIMINAL JUSTICE POLICY AND PRACTICE

Currently, most forensic DNA analysis is based on several defined highly polymorphic variants that are exceedingly useful for individual identification, but are unrelated to phenotype. By comparing the pattern of alleles of the DNA from an individual to a forensic sample, a "match" can be made or ruled out using statistical methods. However useful the current markers are, they cannot be used to determine what a person may look like. This proposal aims at developing a set of polymorphic markers that will be useful indicators of a person's appearance, i.e., hair, eye and skin color. Thus, our ultimate aim is to be able to accurately predict the hair, eye and skin color of the donor of a forensic DNA sample. The ability to predict the appearance of the donor of a given forensic DNA sample has wide implications for criminal justice policy and practice.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

REFERENCES - Background to the Project

Akey JM, Wang H, Xiong M, Wu H, Liu W, Shriver MD, Jin L. Interaction between the melanocortin-1 receptor and P genes contributes to inter-individual variation in skin pigmentation phenotypes in a Tibetan population. Hum Genet 108:516-520, 2001.

Al-Awqati Q. Chloride channels of intracellular organelles. Curr Opin in Cell Biol 7:504-508, 1995.

Ancans J, Thody AJ. Activation of melanogenesis by vacuolar type H(+)-ATPase inhibitors in amelanotic, tyrosinase positive human and mouse melanoma cells. FEBS Lett 478:57-60, 2000.

Ancans J, Tobin DJ, Hoogduijn MJ, Smit NP, Wakamatsu K, Thody AJ. Melanosomal pH controls rate of melanogenesis, eumelanin/phaeomelanin ratio and melanosome maturation in melanocytes and melanoma cells. Exp Cell Res. 268:26-35, 2001.

Bhatnagar V, Anjaiah S, Puri A, Puri N, Darshanam BNA, Ramaiah A. pH of melanosomes of B16 murine melanoma is acidic: Its physiological importance in the regulation of melanin biosynthesis. Arch Biochem and Biophys 307:183-192, 1993.

Boissy RE, Zhao H, Oetting WS, Austin LM, Wildenberg SC, Boissy YL, Zhao Y, Sturm RA, Hearing VJ, King RA, Nordlund JJ. Mutation in and lack of expression of tyrosinase-related protein-1 (TRP-1) in melanocytes from an individual with brown oculocutaneous albinism: a new subtype of albinism classified as "OCA3". Am J Hum Genet 58:1145-56, 1996.

Brilliant MH. The mouse pink-eyed dilution locus: A model for aspects of Prader-Willi syndrome, Angelman syndrome, and a form of hypomelanosis of Ito. Mammal Genome 3: 187-191, 1992.

Brilliant MH. *Mouse Pigmentation*, in: Encyclopedia of Genetics, Brenner S and Miller J., eds., Academic Press, 2001.

Brilliant MH. The Mouse *p* (*pink-eyed dilution*) and Human P genes, Oculocutaneous Albinism Type 2 (OCA2), and Melanosomal pH. Pigment Cell Res 14:86-93, 2001.

Brilliant MH, and Barsh GS. “Genes Regulating the Type of Melanin Pigment Produced by the Melanocyte” In: The Pigmentary System and Its Disorders, J. J. Nordlund, Editor, Oxford University Press, 1998.

Brilliant MH., King RA, Francke U, Schuffenhauer S, Meitinger T, Gardner JM, Durham-Pierre D, Nakatsu Y. The mouse pink-eyed dilution gene: association with hypopigmenation in Prader-Willi and Angelman syndromes, and with human OCA2. Pigment Cell Res 7:398-402, 1994.

Buck GA, Fox JW, Gunthorpe M, Hager KM, Naeve CW, Pon RT, Adams PS, Rush J. Design strategies and performance of custom DNA sequencing primers. Biotechniques. 27:528-36, 1999.

Choi JS, Kim JS, Joe CO, Kim S, Ha KS, Park YM. Improved cycle sequencing of GC-rich DNA template. Exp Mol Med. 31:20-24, 1999.

Cooksey CJ, Garratt PJ, Land EJ, Pavel S, Ramsden CA, Riley PA, Smit NP. Evidence of the indirect formation of the catecholic intermediate substrate

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

responsible for the autoactivation kinetics of tyrosinase. J Biol Chem 272:26226-35, 1997.
Dautry-Varsat A, Ciechanover A, Lodish HF. pH and the recycling of transferring during receptor-mediated endocytosis. Proc Natl Acad Sci 80:2258-2262, 1983.
Devi CC, Tripathi RK, Ramaiah A. pH-dependent interconvertible allosteric forms of murine melanoma tyrosinase. Physiological implications. Eur J Biochem 166:705-711, 1987.
Dickie MM. Mouse News Lett 30:30, 1964.
Du J, Fisher DE. Identification of Aim-1 as the underwhite Mouse Mutant and Its Transcriptional Regulation by MITF. J Biol Chem. 277:402-406, 2002.
Durham-Pierre D, Gardner JM, Nakatsu Y, King RA, Francke U, Ching A, Aquaron R, del Marmol V, Brilliant MH. African origin of an intragenic deletion of the P gene in tyrosinase positive oculocutaneous albinism. Nature Genet 7:176-179, 1994.
Eberle AN. The melanotropins: Chemistry, physiology, and mechanisms of action (Basel: S. Karger publisher), 1988.
Eiberg H, Mohr J. Assignment of genes coding for brown eye colour (BEY2) and brown hair colour (HCL3) on chromosome 15q. Eur J Hum Genet 4:237-241, 1996.
Flanagan N, Healy E, Ray A, Philips S, Todd C, Jackson IJ, Birch-Machin MA, Rees JL. Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation. Hum Mol Genet. 9:2531-7, 2000.
Fuller BB, Spaulding DT, Smith DR. Regulation of the catalytic activity of preexisting tyrosinase in black and Caucasian human melanocyte cell cultures. Exp Cell Res. 262:197-208, 2001.
Furumura M, Sakai C, Abdel-Malek Z, Barsh GS, Hearing VJ. The interaction of agouti signal protein and melanocyte stimulating hormone to regulate melanin formation in mammals. Pigment Cell Res 9:191-203, 1996.
Gahl WA, Potterf B, Durham-Pierre D, Brilliant MH, Hearing VJ. Melansomal tyrosine transport in normal and pink-eyed dilution murine melanocytes. Pigment Cell Res 8:229-233, 1995.
Gardner JM, Nakatsu Y, Gondo Y, Lee S, Lyon MF, King RA, Brilliant MH. The mouse pink-eyed dilution gene: Association with human Prader-Willi and Angelman Syndromes. Science 257:1121-1124, 1992.
Glass DB, Cheng HC, Kemp BE, Walsh DA. Differential and common recognition of the catalytic sites of the cGMP-dependent and cAMP-dependent protein kinases by inhibitory peptides derived from the heat-stable inhibitor protein. J Biol Chem 261:12166-12171, 1986.
Gomez-Angelats M, Lopez-Fontanals M, Felipe A, Casado FJ, Pastor-Anglada M. Cytoskeletal-dependent activation of system A for neutral amino acid transport in osmotically stressed mammalian cells: a role for system A in the intracellular accumulation of osmolytes. J Cell Physiol 173:343-350, 1997.
Harada, M., Y.F. Li, M. El-Gamil, S.A. Rosenberg, and P.F. Robbins. Use of an *in vitro* immunoselected tumor line to identify shared melanoma antigens

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

recognized by HLA-A*0201-restricted T cells. *Cancer Res*. 61:1089-1094, 2001.

Hearing VJ, Tsukamoto K. Enzymatic control of pigmentation in mammals. FASEB J 5:2902-2909, 1991.

Hill HZ, Li W, Xin P, Mitchell DL. Melanin: a two edged sword? Pigment Cell Res 10:158-61, 1997.

Isaac EL, Karageorgos LE, Brooks DA, Hopwood JJ, Meikle PJ Regulation of the lysosome-associated membrane protein in a sucrose model of lysosomal storage. Exp Cell Res 254:204-209, 2000.

Ito S. High-performance liquid chromatography (HPLC) analysis of eu- and pheomelanin in melanogenesis control. J Invest Dermatol 100:166S-171S, 1993

Jackson IJ. A cDNA encoding tyrosinase-related protein maps to the brown locus in mouse. Proc Natl Acad Sci U S A. 85:4392-4396, 1988.

Jackson IJ. Homologous pigmentation mutations in human, mouse and other model organisms. Hum Mol Genet 6:1613-1624, 1997.

Jimenez-Cervantes C, Germer S, Gonzalez P, Sanchez J, Sanchez CO, Garcia-Borron JC. Thr40 and Met122 are new partial loss-of-function natural mutations of the human melanocortin 1 receptor. FEBS Lett. 508:44-48, 2001.

Katayama H, Ota T, Morita K, Terada Y, Suzuki F, Katoh O, Tatsuka M. Human AIM-1: cDNA cloning and reduced expression during endomitosis in megakaryocyte-lineage cells. Gene. 224:1-7, 1998.

King RA, Hearing VJ, Oetting WS. Chapter 147: Albinism. IN: Scriver CR, Beaudet AL, Sly WS and Valle D (Eds.): *The Metabolic and Molecular Bases of Inherited Disease,* Seventh Edition. McGraw-Hill Publishing Co., New York, pp. 4353-4392, 1994.

Kwon BS, Haq AK, Pomerantz SH, Halaban R. Isolation and sequence of a cDNA locus for human tyrosinase that maps at the mouse c-albino locus. Proc Natl Acad Sci USA 84: 7473-7477, 1987.

Lee S-T, Nicholls RD, Bundey S, Laxova R, Musarella M, Spritz RA. Mutations of the *P* gene in oculocutaneous albinism, ocular albinism, and Prader-Willi Syndrome plus albinism. N Engl J Med 330:529-534, 1994a.

Lee S-T, Nicholls RD, Schnur RE, Guida LC, Lu-Kuo J, Spinner NB, Zackai EH, Spritz RA. Diverse mutations of the P gene among African-Americans with type II (tyrosinase-positive) oculocutaneous albinism (OCA2). Hum Mol Genet 3:2047-2051, 1994b.

Lee S-T, Nicholls RD, Jong MTC, Spritz RA. Organization and sequence of the human P gene and identification of a new family of transport proteins. Genomics 26: 345-363, 1995.

Lehman AL, Silvers WK, Puri N, Wakamatsu K, Ito S and Brilliant MH. The *underwhite* (*uw*) locus acts autonomously and reduces the production of melanin via a unique pathway. J Invest Dermatol, 115:601-606, 2000.

Lund PM, Puri N, Durham-Pierre D, King RA, and Brilliant MH. Oculocutaneous albinism in an isolated Tonga community in Zimbabwe. J Medical Genetics 34:733-735, 1997.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Lynch M, Walsh B. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., Sunderland, MA, 1998.
Lyon MF, King TR, Gondo Y, Gardner JM, Nakatsu Y, Eicher EM, Brilliant MH. Genetic and molecular analysis of recessive alleles at the pink-eyed dilution locus. Proc Natl Acad Sci USA 89: 6968-6972, 1992.
Manga P, Kromberg JG, Box NF, Sturm RA, Jenkins T, Ramsay M. Rufous oculocutaneous albinism in southern African Blacks is caused by mutations in the TYRP1 gene. Am J Hum Genet. 61:1095-101, 1997.
Manga P, Boissy RE, Pifko-Hirst S, Zhou BK, Orlow SJ. Mislocalization of melanosomal proteins in melanocytes from mice with oculocutaneous albinism type 2. Exp Eye Res. 72:695-710, 2001.
Marks MS, Seabra MC. The melanosome: membrane dynamics in black and white. Nat Rev Mol Cell Biol. 2:738-748, 2001.
Moellmann G, Slominski A, Kuklinska E, Lerner AB. Regulation of melanogenesis in melanocytes. Pigment Cell Res 1:79-87, 1988.
Newton JM, Cohen-Barak O, Hagiwara N, Gardner JM, Davisson MT, King RA, Brilliant MH. Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4. Am J Hum Genet 69:981-988, 2001.
Oetting WS, Brilliant MH, King RA. The clinical spectrum of albinism in humans. Mol Med Today. 2:330-335, 1996.
Oetting WS, Fryer JP, King RA. Mutations of the human tyrosinase gene associated with tyrosinase related oculocutaneous albinism (OCA1). Mutations in brief no. 204. Online. Hum Mutat 12:433-4, 1998a.
Oetting WS, Gardner JM, Fryer JP, Ching A, Durham-Pierre D, King RA and Brilliant MH. Mutations of the human P gene associated with Type II Oculocutaneous Albinism (OCA2). Human Mutation 12: 434, 1998b. Online Citation: http://journals.wiley.com/1059-7794/pdf/mutation/205.pdf
Oetting WS, King RA. Molecular basis of albinism: mutations and polymorphisms of pigmentation genes associated with albinism. Hum Mutat 13:99-115, 1999.
Orlow SJ. Melanosomes are specialized members of the lysosomal lineage of organelles. J Invest Dermatol 105: 3-7, 1995
Orlow SJ and Brilliant MH. The *pink-eyed dilution* locus controls the biogenesis of melanosomes and levels of melanosomal proteins in the eye. Exp Eye Res. 68:147-54, 1999.
Ozeki H, Ito S, Wakamatsu K and Hirobe T. Chemical characterization of hair melanins in various coat color mutants of mice. J Invest Dermatol 105:361-366, 1995.
Park HY, Gilchrest BA. Signaling pathways mediating melanogenesis. Cell Mol Biol (Noisy-le-grand). 45:919-930, 1999.
Potterf SB, Furumura M, Sviderskaya EV, Santis C, Bennett DC, Hearing VJ. Normal tyrosine transport and abnormal tyrosinase routing in pink-eyed dilution melanocytes Exp Cell Res 244:319-26, 1998.
Potterf SB, Muller J, Bernardini I, Tietze F, Kobayashi T, Hearing VJ, Gahl WA. Characterization of a melanosomal transport system in murine melanocytes

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

mediating entry of the melanogenic substrate tyrosine. J Biol Chem 271:4002-4008, 1996.

Prota G, Lamoreux, ML, Muller J, Kobayashi T, Napolitano J, Vincensi MR, Sakai C, Hearing VJ. Comparative analysis of melanins and melanosomes produced by various coat color mutants Pigment Cell Res 8:153-163, 1995.

Puri N, Durham-Pierre D, Aquaron R, Lund PM, King RA and Brilliant MH. Type 2 oculocutaneous albinism (OCA2) in Zimbabwe and Cameroon: distribution of the 2.7 kb deletion allele of the P gene. Human Genetics 100:651-656, 1997.

Puri N, Gardner JM and Brilliant MH. Aberrant pH of melanosomes in *pink-eyed dilution* (*p*) mutant melanocytes. J Invest Dermatol, 115:607-613, 2000.

Quest AF, Ghosh S, Xie WQ, Bell RM. DAG second messengers: molecular switches and growth control. Adv Exp Med Biol. 400A:297-303, 1997.

Ramaiah A. Lag kinetics of tyrosinase: its physiological implications. Indian J Biochem Biophys 33:349-356, 1996.

Rees JL. The melanocortin 1 receptor (MC1R): more than just red hair. Pigment Cell Res. 13:135-140, 2000.

Rinchik EM, Bultman SJ, Horsthemke B, Lee S-T, Strunk KM, Spritz RA, Avidano KM, Jong MTC, Nicholls RD. A gene for the mouse pink-eyed dilution locus and for human type II oculocutaneous albinism. Nature 361: 72-76, 1993.

Robbins LS, Nadeau JH, Johnson KR, Kelly MA, Roselli-Rehfuss L, Baack E, Mountjoy KG, Cone RD. Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. Cell 72:827-34, 1993.

Ronin, Y. I., V. M. Kirzhner, and A. B. Korol. Linkage between loci for quantitative traits and marker loci: multi-trait analysis with a single marker. Theor. Appl. Genet. 90:776-786, 1995.

Rosemblat S, Durham-Pierre D, Garner JM, Nakatsu Y, Brilliant MH, Orlow SJ. Identification of a melanosomal membrane protein encoded by the pink-eyed dilution (type II oculocutaneous albinism) gene. Proc Natl Acad Sci USA 91:12071-12075, 1994.

Rosemblat S, Sviderskaya EV, Easty DJ, Wilson A, Kwon BS, Bennett DC, Orlow SJ. Melanosomal defects in melanocytes from mice lacking expression of the pink-eyed dilution gene: correction by culture in the presence of excess tyrosine. Exp Cell Res 239:344-52, 1998.

Schreiber R, Haussinger D. Characterization of the swelling-induced alkalinization of endocytotic vesicles in fluorescein isothiocyanate-dextran-loaded rat hepatocytes. Biochem J 309:19-24, 1995.

Schreiber R, Zhang F, Haussinger D. Regulation of vesicular pH in liver macrophages and parenchymal cells by ammonia and anisotonicity as assessed by fluorescein isothiocyanate-dextran fluorescence. Biochem J 315:385-392, 1996.

Sidman RL, Pearlstein R. Pink-eyed dilution (*p*) gene in rodents: increased pigmenataion in tissue culture. Dev Biol 12:93-116, 1965.

Silvers, WK. *The Coat Colors of Mice*. Springer Verlag, New York, 1979.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Spielman, R. S., R. E. McGinnis, and W. J. Ewens. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. 52:506-516, 1993.

Spritz RA, Lee ST, Fukai K, Brondum-Nielsen K, Chitayat D, Lipson MH, Musarella MA, Rosenmann A, Weleber RG. Novel mutations of the P gene in type II oculocutaneous albinism (OCA2). Hum Mutat. 10:175-177, 1997.

Stadler R, Truernit E, Gahrtz M, Sauer N. The AtSUC1 sucrose carrier may represent the osmotic driving force for anther dehiscence and pollen tube growth in Arabidopsis. Plant J. 19:269-278, 1999.

Stephanson DA, Hornby JE. Gene expression at the pink-eyed dilution (P) locus in the mouse is confirmed to be pigment cell autonomous using recombinant embryonic skin grafts. J. Embryol exp Morph 87:65-73, 1985.

Stern C. Model estimates of the number of gene pairs involved in pigmentation variability of the Negro-American. Human Hered 20:165-168, 1970.

Sturm RA, Teasdale RD, Box NF. Human pigmentation genes: identification, structure and consequences of polymorphic variation. Gene 277:49-62, 2001.

Sweet HO, Brilliant MH, Cook SA, Johnson KR and Davisson MT. A new alleleic series for the underwhite gene on mouse Chromosome 15. J Heredity 89:546-551, 1998.

Szabo G, Gerald AB, Pathak MA, Fitzpatrick TB. Racial differences in the fate of melanosomes in human epidermis. Nature 222:1081-1082, 1969.

Takamoto T, Schwartz B, Cantor LB, Hoop JS, Steffens T. Measurement of iris color using computerized image analysis. Curr Eye Res. 22:412-419, 2001.

Teichmann, U., *et al*. Cloning and tissue expression of the mouse ortholog of AIM1, a betagamma-crystallin superfamily member Mammal. Genome 9:715-720, 1998.

Tomita Y, Takeda A, Okinaga S, Tagami H, Shibahara S. Human oculocutaneous albinism caused by single base insertion in the tyrosinase gene. Biochem Biophys Res Commun 164: 990-996, 1989.

Townsend D, Guillery P, King RA. Optimized assay for mammalian tyrosinase (polyhydroxyl phenyloxidase). Anal Biochem 139:345-352, 1984.

Tripathi RK, Chaya, Devi C, Ramaiah A. pH-dependent interconversion of two forms of tyrosinase in human skin. Biochem J 252:481-487, 1988.

Tsukamoto K, Jackson IJ, Urabe K, Montague PM, Hearing, VJ A second tyrosinase-related protein, TRP2, is a melanogenic enzyme termed DOPAchrome tautomerase EMBO J 11:519-526, 1992.

Valverde P, Healy E, Jackson I, Rees JL, Thody AJ. Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat Genet 11:328-330, 1995.

Wahlund, S. Zuzammensetzung von populationen und korrelationsercheinugen vom standpunkt der vererbungslehre aus betrachtet. Hereditas 11:65-106, 1928.

Weir, BS. *Genetic Data Analysis III: Methods for Discrete Population Genetic Data*. Sinauer Associates, Inc., Sunderland, MA, 2001.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

History of the Project

The studies supported by this grant were officially initiated on July 1,2002. There were delays in the final approval of funds from the DOJ (approved on 9/9/02 and the University of Arizona (funds released on 9/19/02). One of the first items purchased was a skin reflectometer. Our sample preparation protocol was significantly improved (quicker and with higher yields) using a kit from Epicenter Technologies. Because the SNP database was not well established at this time – especially among African-Americans, we began by determining polymorphic variation among human populations. By the end of 2002 we determined the range of polymorphisms among 100 unrelated, normally pigmented, African-American subjects was determined for three exons of the P gene (9, 13, 14 and 24) and exon one of MATP. In this population sample, no polymorphisms were detected in exons 14 or 24 of the P gene. However, potentially informative polymorphisms were detected in exons 9 and 13 of the P gene. Exon 1 of MATP was not polymorphic in the African-American subjects.

By 6/30/03, we recruited more than 750 of the 1,000 subjects of various ethnicities. These subjects (and all others) gave written informed consent and were paid $20 for their participation. Skin reflectance measurements were taken of the inner aspect of the upper arm, a buccal cell sample was taken for DNA isolation, a hair sample was taken for chemical analysis and eye color was determined by comparison to a standard eye color chart. A survey form was also filled out. Although we obtained a fairly proportional ethnic representation of our student population (the focus of our recruitment efforts), our subject population was somewhat under-represented by African-Americans. Survey data was been entered, DNA was been prepared, hair samples were sent to our collaborators in Japan (for chemical analysis), reflectance data was been obtained (although a computer glitch has resulted in the loss of some reflectance data that was not recovered). Analysis of polymorphic variation in and around exons of the P gene was begun. We requested a budget category change (for a sub-contract with DNAprint) that will allow us to expand our studies (survey more polymorphisms in more genes).

By 12/31/03, we had recruited over 800 of the planned 1,000 subjects and concentrated our remaining recruitment efforts on African-Americans, who are somewhat under-represented in our current sample set. We completed the genotyping of nearly 300 individuals for 117 individual polymorphisms (Phase I of our sub-contract with DNAPrint). The polymorphisms include MC1R (26 polymorphisms), P (24 polymorphisms), MATP (14 polymorphisms), ASP (1 polymorphism) and several other genes not originally included in our proposed studies. These include (AP3, 3 polymorphisms; CHS, 3 polymorphisms; MYO5A, 7 polymorphisms; tyrosinase, 4 polymorphisms; OA1, 4 polymorphisms; HPS1-6 genes, 13 polymorphisms; as well as 14 polymorphisms in 5 other genes. Only a few of these SNPs were later found to be uninformative or difficult to interpret.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

By 6/3)/04, initial data analysis was completed for this initial set of 300 samples. We selected 52 polymorphisms in 16 genes that were significantly linked to a pigmentation phenotype. These genes include: MC1R, P, MATP, ASIP, TRP1, DCT, OA1, HPS1, HPS3, HPS4, CYP1, CYP4, MYO5A, MYO7A, AP3D, and MLPH. Intriguingly, single polymorphisms in each of three genes (P, MATP, ASIP) accounted for 66% of the phenotypic variation in skin pigmentation in this sample set, supporting earlier genetic data suggesting that most of the variation in human skin pigmentation is associated with only 3 or 4 genes. Hair color phenotype was analyzed chemically as two independent factors: absolute amount of melanin and the ratio of red to black melanins. In this ample set, single polymorphisms in each of two genes (P and MATP) were found to account for 75% of the variance in absolute amount of melanin; several polymorphisms in two genes (MC1R and ASIP) accounted for 68% of the ratio of red to black melanins (in a non-logarithmic model). Eye color appeared to be more complex.

By 12/31/04 the set of 52 polymorphisms in the remaining 500 samples was completed and analysis was ongoing. The initial data analysis of 800 individuals found that single polymorphisms in each of three genes (P, MATP, ASIP) accounts for 62% of the phenotypic variation in skin pigmentation. Single polymorphisms in each of three genes (P, MATP and ASIP) accounted for 75% of the variance in absolute amount of melanin; several polymorphisms in two genes (MC1R and ASIP) can account for 68% of the ratio of red to black melanins. Eye color analysis remained in progress and we initiated statistical tests to determine the validity and robustness of our models.

By 6/31/05, we continued the data analyses. The results from these analyses were as follows:

Total Hair Melanin.

Hair color is the result of two factors, total melanin pigment and the ration of eumelanin (brown/black) to pheomelanin (yellow/red). Six markers have significant effects on hair total melanin: dbSNPrs16891982 (MATP), dbSNPrs2424984 (ASIP), OCA2E11_263_886895 (P gene), dbSNPrs1800404 (P gene), dbSNPrs1407995 (DCT), and dbSNPrs3212357 (MC1R). When each are considered separately, these markers account for 57.8%, 19.9%, 17.5%, 14.3%, 12.6%, 10.0% (respectively) of the total trait variation. When all six markers are simultaneously considered in a multiple regression, only the effects of dbSNPrs16891982 (MATP) and dbSNPrs1800404 (P gene) are significant, with this model accounting for 62.0% of the total trait variation. The interaction term between these two markers is also highly significant ($p < 0.0001$), with the full model (both single markers plus their interaction) accounting for 67.6% of the total trait variation. Analysis of the log transformation of hair total melanin produced essentially identical results.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Hair Melanin Ratio.

Hair color is the result of two factors, total melanin pigment and the ration of eumelanin (brown/black) to pheomelanin (yellow/red). Four markers showed a significant effect on hair melanin ratio: dbSNPrs16891982 (MATP), dbSNPrs2424984 (ASIP), dbSNPrs3212357 (MC1R), and dbSNPrs1800404 (P gene),. When considered one at a time, these markers (using a standard fixed-effects one way ANOVA) account for 25.3%, 12.9%. 10.8%, and 5.1% of the total trait variance. When all four markers are simultaneously considered in a multiple regression model, the marker dbSNPrs1800404 (P gene) does not make a significant contribution but the regression with the remaining three markers accounts for 32.9% of the total variance. Two of the pair-wise interactions between these markers are also highly significant, dbSNPrs2424984 (ASIP) times dbSNPrs3212357 (MC1R) (p value <0.0001) and dbSNPrs16891982 (MATP) times dbSNPrs2424984 (ASIP) (p value = 0.0026). The full linear model including the three single marker effects and these two interactions accounts for 44.4%of the total trait variance.

Some caution is in order in interpreting these results, however, as there are four strong leverage points (individuals with unusually high hair melanin ratios). The mean ratio for the sample was around 365, but four individuals had values ranging from 2190 to a maximum of 5285.3. When natural log of the hair melanin ratio is considered, the same four markers are still significant, but now account (as single markers) for only 23.1%, 8.4%, 6.9%, and 5.5% of the total variance. Much more interestingly, in a multiple regression only dbSNPrs16891982 (MATP) has a significant effect. Likewise, none of the pairwise interaction terms are significant. Given the relatively small sample of individuals (224) it remains to be seen whether these leverage points will remain biologically significant as more individuals are sampled, and hence whether markers other than dbSNPrs16891982 (MATP) will account for a significant fraction of hair melanin ratio trait variation.

Skin Pigmentation.

Five markers showed a significant effect on skin reflectance: dbSNPrs16891982 (MATP), dbSNPrs2424984 (ASIP), OCA2E14_447_886894 (P gene), OCA2E11_263_886895 (P gene), and dbSNPrs1325611 (DCT). When considered separately as single markers, these account for 32.1%, 15.6%, 10.7%, 3.5%, and 2.8% of the total trait variance. When all five markers are simultaneously considered in a multiple regression, four are still significant (dbSNPrs1325611 (DCT) is not), and the resulting four-marker model accounts for 50.8% of the total trait variance. One of the pair-wise interactions between these four markers, OCA2E14_447_886894 (P gene) times dbSNPrs2424984 (ASIP) is highly significant (p < 0.0001), and the resulting full model (four markers plus the single pair-wise interaction) accounts for 56.4% of the total trait variance.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Eye Color.
Eye color was significantly influenced by 5 markers: dbSNPrs1900758 (P gene), OCA2DBSNP_8321_71205 (P gene), dbSNPrs2733832 (TYRP 1), dbSNPrs16891982 (MATP), and dbSNPrs2424984 (ASIP). When considered separately as single markers, these account for 20.0%, 14.7%, 10.3%, 5.7% and 3.4% of the total trait variance. When simultaneously considered in multiple regression, all markers are significant (all with $p < 0.0001$) and the resulting model accounts for 36.1% of the total trait variance. However, when pairwise interactions were considered in a regression model, marker dbSNPrs2424984 (ASIP) was no longer significant ($p = 0.12$) and marker dbSNPrs16891982 (MATP) was just marginally significant ($p = 0.02$), while their interaction term dbSNPrs2424984 (ASIP) x dbSNPrs16891982 (MATP) was highly significant ($p = 0.0076$). The resulting full model (main effects for markers dbSNPrs1900758 (P gene), OCA2DBSNP_8321_71205 (P gene), dbSNPrs2733832 (TYRP 1), and dbSNPrs16891982 (MATP) plus the single pairwise interaction term) accounted for 37.7% of the total trait variation. Thus, there is evidence for a very strong epistatic interaction between term dbSNPrs2424984 (ASIP) and dbSNPrs16891982 (MATP).

By 12/31/05, continued the analyses. These results were:

Total Hair Melanin.
Hair color is the result of two factors, total melanin pigment and the ration of eumelanin (brown/black) to pheomelanin (yellow/red). Six markers have significant effects on hair total melanin: dbSNPrs16891982 (MATP), dbSNPrs2424984 (ASIP), OCA2E11_263_886895 (P gene), dbSNPrs1800404 (P gene), dbSNPrs1407995 (DCT), and dbSNPrs3212357 (MC1R). When each are considered separately, these markers account for 57.8%, 19.9%, 17.5%, 14.3%, 12.6%, 10.0% (respectively) of the total trait variation. When all six markers are simultaneously considered in a multiple regression, only the effects of dbSNPrs16891982 (MATP) and dbSNPrs1800404 (P gene) are significant, with this model accounting for 62.0% of the total trait variation. The interaction term between these two markers is also highly significant ($p < 0.0001$), with the full model (both single markers plus their interaction) accounting for 67.6% of the total trait variation. Analysis of the log transformation of hair total melanin produced essentially identical results.

Hair Melanin Ratio.
Hair color is the result of two factors, total melanin pigment and the ration of eumelanin (brown/black) to pheomelanin (yellow/red). Four markers showed a significant effect on hair melanin ratio: dbSNPrs16891982 (MATP), dbSNPrs2424984 (ASIP), dbSNPrs3212357 (MC1R), and dbSNPrs1800404 (P gene),. When considered one at a time, these markers (using a standard fixed-effects one way ANOVA) account for 25.3%, 12.9%. 10.8%, and 5.1% of the total trait variance. When all four markers are simultaneously considered in a multiple

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

regression model, the marker dbSNPrs1800404 (P gene) does not make a significant contribution but the regression with the remaining three markers accounts for 32.9% of the total variance. Two of the pair-wise interactions between these markers are also highly significant, dbSNPrs2424984 (ASIP) times dbSNPrs3212357 (MC1R) (p value <0.0001) and dbSNPrs16891982 (MATP) times dbSNPrs2424984 (ASIP) (p value = 0.0026). The full linear model including the three single marker effects and these two interactions accounts for 44.4%of the total trait variance.

Some caution is in order in interpreting these results, however, as there are four strong leverage points (individuals with unusually high hair melanin ratios). The mean ratio for the sample was around 365, but four individuals had values ranging from 2190 to a maximum of 5285.3. When natural log of the hair melanin ratio is considered, the same four markers are still significant, but now account (as single markers) for only 23.1%, 8.4%, 6.9%, and 5.5% of the total variance. Much more interestingly, in a multiple regression only dbSNPrs16891982 (MATP) has a significant effect. Likewise, none of the pairwise interaction terms are significant. Given the relatively small sample of individuals (224) it remains to be seen whether these leverage points will remain biologically significant as more individuals are sampled, and hence whether markers other than dbSNPrs16891982 (MATP) will account for a significant fraction of hair melanin ratio trait variation.

Skin Pigmentation.

Five markers showed a significant effect on skin reflectance: dbSNPrs16891982 (MATP), dbSNPrs2424984 (ASIP), OCA2E14_447_886894 (P gene), OCA2E11_263_886895 (P gene), and dbSNPrs1325611 (DCT). When considered separately as single markers, these account for 32.1%, 15.6%, 10.7%, 3.5%, and 2.8% of the total trait variance. When all five markers are simultaneously considered in a multiple regression, four are still significant (dbSNPrs1325611 (DCT) is not), and the resulting four-marker model accounts for 50.8% of the total trait variance. One of the pair-wise interactions between these four markers, OCA2E14_447_886894 (P gene) times dbSNPrs2424984 (ASIP) is highly significant ($p < 0.0001$), and the resulting full model (four markers plus the single pair-wise interaction) accounts for 56.4% of the total trait variance.

Eye Color.

Eye color was significantly influenced by 5 markers: dbSNPrs1900758 (P gene), OCA2DBSNP_8321_71205 (P gene), dbSNPrs2733832 (TYRP 1), dbSNPrs16891982 (MATP), and dbSNPrs2424984 (ASIP). When considered separately as single markers, these account for 20.0%, 14.7%, 10.3%, 5.7% and 3.4% of the total trait variance. When simultaneously considered in multiple regression, all markers are significant (all with $p < 0.0001$) and the resulting model accounts for 36.1% of the total trait variance. However, when pairwise interactions were considered in a regression model, marker dbSNPrs2424984 (ASIP) was no longer significant ($p = 0.12$) and marker dbSNPrs16891982 (MATP) was just marginally significant ($p = 0.02$), while their interaction term

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

dbSNPrs2424984 (ASIP) x dbSNPrs16891982 (MATP) was highly significant (p = 0.0076). The resulting full model (main effects for markers dbSNPrs1900758 (P gene), OCA2DBSNP_8321_71205 (P gene), dbSNPrs2733832 (TYRP 1), and dbSNPrs16891982 (MATP) plus the single pairwise interaction term) accounted for 37.7% of the total trait variation. Thus, there is evidence for a very strong epistatic interaction between term dbSNPrs2424984 (ASIP) and dbSNPrs16891982 (MATP).

At this time, a manuscript was readied for submission detailing these studies and conclusions. However, in November 2005, we became aware of a study *in press* in Science that demonstrates a major effect on human pigmentation by the SLC24A5 gene originally identified as the Zebrafish *golden* gene (Lamason et al., Science 310:1782-1786, 2005). This made the submission of the manuscript highly problematic. Certainly, reviewers would want this gene included in any comprehensive study of the genetics of human pigmentation traits. Therefore, we set out to add the analysis of the SLC24A5 gene to this study.

We added additional aims to this project as follows to include a blind study of approximately 250 additional individuals to evaluate the accuracy of our DNA-based predictions of phenotype.

By the end of 2005, we had recruited over 100 additional subjects and initiated the blind study. We had also planned to develop a multiplexed PCR analytical method (to be performed in collaboration with John Butler and Peter Valone at NIST). Progress towards the development of the multiplex assay was made: PCR primers for 16 of the 17 loci are working; Locus 2 is not amplifying (even with 3 PCR primer re-designs); Primer extension reactions were split into “A” and “B”; Locus 5 is very weak; locus 10 did not giving the proper genotype; and the SLC24A5 SNP needed to be added and we needed to address the technical issues cited above

As of 6/30/06, we were able to analyze and include Golden/SLC24A5 in our model development:

Total Hair Melanin.
Hair color is the result of two factors, total melanin pigment and the ration of eumelanin (brown/black) to pheomelanin (yellow/red). Three markers have significant effects on hair total melanin: SNPrs16891982 (MATP/SLC45A2), SNPrs1800410 (P gene), and SNPrs1426654 (Golden/SLC24A5). When each are considered separately, these markers account for 58.3%, 17.5%, and 58.2% (respectively) of the total trait variation. When simultaneously considered in a multiple regression, this model accounting for 71.3% of the total trait variation. Significant interactions between SNPrs1800410 (P gene) and SNPrs16891982 (MATP/SLC45A2) and SNPrs1426654 (Golden/ SLC24A5) were detected. The resulting full model (three single markers plus two interactions) accounts for 79.5% of the total trait variation.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Hair Melanin Ratio.
Hair color is the result of two factors, total melanin pigment and the ration of eumelanin (brown/black) to pheomelanin (yellow/red). Because this value is a ratio, we used a conservative approach and analyzed the trait as the natural log of the ratio. Six markers in five genes showed a significant effect on hair melanin ratio: SNPrs16891982 (MATP), SNPrs1805007 (MC1R), SNPrs1407995 (DCT), SNPrs10852218 (P gene), SNPrs11638265 (P gene), and SNPrs1426654 (Golden/ SLC24A5). When considered one at a time, these markers account for 23.3%, 15.3%. 9.5%, 4.8%, 4.4% and 21.9% (respectively) of the total trait variance. When simultaneously considered in a multiple regression, this model accounting for 44.5% of the total variance.

Skin Pigmentation (CIEL reflectance scale).
Six markers in five genes showed a significant effect on skin reflectance: SNPrs16891982 (MATP), SNPrs10852218 (P gene), SNPrs1724630 (MYO5A), SNPrs3212363 (MC1R), SNPrs1800410 (P gene), and SNPrs1426654 (Golden/ SLC24A5). When considered separately as single markers, these account for 32.1%, 15.6%, 5.3%, 5.1%, 4.2%, 3.5% and 23.9% of the total trait variance. When all six markers are simultaneously considered in a multiple regression, this model accounts for 48.7% of the total trait variance. Two pair-wise interactions between SNPrs1800410 (P gene) and SNPrs3212363 (MC1R), and SNPrs1426654 (Golden/ SLC24A5) and SNPrs1724630 (MYO5A) were significant. The resulting full model (six markers plus two pair-wise interactions) accounts for 58.7% of the total trait variance.

Eye Color.
Six markers in five genes showed a significant effect on eye color: SNPrs1800410 (P gene), SNPrs2733832 (TRP1), SNPrs1724630 (MYO5A), SNPrs3212363 (MC1R), SNPrs1800414 (P gene) and SNPrs1426654 (Golden/ SLC24A5). When considered separately as single markers, these account for 11.5%, 10.2%, 6.6%, 5.9% and 30.2% of the total trait variance. When all six markers are simultaneously considered in a multiple regression, this model accounts for 47.3% of the total trait variance. Two pair-wise interactions between SNPrs3212363 (MC1R) and SNPrs1426654 (Golden/ SLC24A5) and SNPrs3212363 (MC1R) and SNPrs2733832 (TRP1) were significant. The resulting full model (six markers plus two pair-wise interactions) accounts for 54.3% of the total trait variance.

For the Supplemental Studies, a blind study of approximately 250 additional individuals to evaluate the accuracy of our DNA-based predictions of phenotype was initiated. All subjects were recruited. We have completed typing them for SNPrs1426654 (Golden/SLC24A5), SNPrs3212363 (MC1R), and SNPrs16891982 (MATP). Little progress was made on the multiplex in collaboration with the Butler lab.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

As of 12/31/06, in adition to the analysis (Golden/SLC24A5), we added another SNP in ASP based on the results of recently published reports. We explored multiple ways to analyze the data. In particular, we have used the JMP statistical package to identify those SNPs giving the highest individual R^2 values for each trait. Then the SNPs with the highest individual R^2 values were used to construct models (by both forward and reverse multiple linear regression analyses that together produce the highest cumulative R^2 values. Initially, forward, reverse, and mixed step-wise regression methods were used to trim the number of SNPs. This posed several problems, different models were obtained based on the method used and we had no confidence that the models we generated reflected biological significance, nor did we have any idea how well such a generated model compared to all possible models of 3 SNPs generated from the 41 SNPs. In an attempt to find more reliable models and know where these models ranked compared to all possible models, we used SAS to generate all possible models of 3 SNPs from a pool of 41 SNPs. All models were graphed by R^2 in descending order; this showed inflections in R^2 values. To find the cause in these changes we made histograms that contained all models up to the R^2 inflection. The top 3 SNPs from the histograms were then plotted against the R^2 graph, in doing so, it became obvious which SNPs were predominantly responsible for the inflections in the R^2 curves. This approach has given us more confidence that these models better reflect biological significance as it reduces type I errors.

The R^2 of the multiple linear regression models models composed of the top 3 SNPs, as determined above, for total hair melanin, ln(eumelanin to pheomelanin), skin reflectance, and eye color were 71.3%, 39.4%, 45.7%, and 52.2%, respectively (Table 1).

The multiple linear refression models from all four traits contained the two SNPs rs16891982 (SLC45A2) and rs1426654 (SLC24A5), however, the third SNP varied between all models. Six SNPs from 5 genes comprised the top 3 SNPs for each of the four different traits. Total hair melanin and eye color models contained P gene SNPs rs1800410 and rs2424984, respectively; ln(eumelanin to pheomelanin) contained MC1R SNP rs1805007; and skin reflectance contained ASIP SNP rs2424984.

The three independent variables (SNPs) of the multiple linear regression model developed by the above method for total hair melanin were: rs16891982 (SLC45A2), rs1426654 (SLC24A5), and rs1800410 (P gene). When considered individually, these markers (using a fixed-effects one way ANOVA) accounted for 58.3%, 58.2%, and 17.5% of the variance.

Total melanin and skin reflectance models had interaction terms that increased R^2 values by 4% or more. Total melanin's R^2 increase was 79.5%, interactions were SLC45A2*P gene (rs16891982*rs1800410), and SLC24A5*P gene

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

(rs1426654*rs1800410). Skin reflectance's R^2 increased to 49.6% with interaction term SLC45A2*ASIP (rs16891982*rs2424984).

Total Hair Melanin	**Individual R^2 values (%)**
SLC45A2	58.2
SLC24A5	57.8
OCA2	17.6
Model	71.3
Model+SLC45A2*OCA2	72.2
Model+SLC24A5*OCA2	75.1
Model+both interaction terms	79.5
ln(Eumelanin/Pheomelanin)	
SLC45A2	23.3
SLC24A5	21.9
MC1R	15.3
Model	39.4
Average Skin Reflectance	
SLC45A2	32.1
SLC24A5	23.9
ASIP	15.6
Model	45.7
Model+SLC45A2*ASIP	49.6
Eye Color	
SLC45A2	38.4
SLC24A5	34.3
OCA2	6.3
Model	52.2

Table1. R^2 values of individual SNPs, interaction terms, and full model.

As of 6/30/07, we have used SAS to generate all possible models of 3 SNPs from a pool of 41 SNPs. All models were graphed by R^2 in descending order; this showed inflections in R^2 values. To find the cause in these changes we made histograms that contained all models up to the R^2 inflection. The top 3 SNPs from the histograms were then plotted against the R^2 graph, in doing so, it became obvious which SNPs were predominantly responsible for the inflections in the R^2 curves. This approach has given us more confidence that these models better reflect biological significance as it reduces type I errors. Recent communications with several statisticians have confirmed the validity of this approach. This has led to a manuscript that was submitted to Genetics (see Executive Summary and the Appendix.)

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

4. CONCLUSIONS

This work has led to the development of a forensic test that can predict four phenotypic features of pigmentation in human beings from a DNA sample by typing only six SNPs in five pigmentation genes.

Our primary aim was to develop a forensic DNA test predictive for pigmentation phenotype. Therefore, we analyzed the data across all populations of the study. The pool of SNPs was reduced from 74 to 39 SNPs by ANOVA. The models accounted for significant variation in each of the four measured traits: scalp-hair total melanin (77.3%), natural log of the ratio of eumelanin-to-pheomelanin (38.3%), skin reflectance (45.7%), and eye color (52.2%).

SNPs from five genes comprised the top three-SNPs in the MLR models for each of the four different traits. Analysis of the three-SNP curves showed that all MLR models of the four traits contained the same two SNPs, rs16891982 (*SLC45A2*) and rs1426654 (*SLC24A5*). However, the third SNP varied in each trait model. Scalp-hair total melanin contained SNP rs1800410 (*OCA2*); natural log of the ratio of eumelanin-to-pheomelanin contained SNP rs1805007 (*MC1R*); skin reflectance contained SNP rs2424984 (*ASIP*); and eye color contained the SNP rs10852218 (*OCA2*).

Scalp-hair total melanin
The three-SNP R^2 curve showed SNP rs1800410 (*OCA2*) to be the third most frequent SNP for the top 250 R^2 models (2.34%). The last inflection, the top 1800 models (16.9%), showed rs1724630 (*MYO5A*) to be the third most frequent SNP. SNP rs1800410 (*OCA2*) in a three-SNP MLR model with rs16891982 (*SLC45A2*) and rs1426654 (*SLC24A5*) yielded the highest R^2 value of 77.3% (n=137) as compared to the other candidate third SNPs.

Natural log of the ratio of eumelanin-to-pheomelanin
The third most frequent SNP was rs1805007 (*MC1R*) for all inflections in the three-SNP MLR curve. A three-SNP MLR model of rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), and rs1805007 (*MC1R*) yielded an R^2 of 38.3% (n=153).

Skin Reflectance
The third most frequent SNP was rs2424984 (*ASIP*). A three-SNP MLR yielded an R^2 value of 45.7%. An interaction term of rs2424984 (*ASIP*) and rs16891982 (*SLC45A2*) increased the R^2 value of the model by approximately 4% to 49.6% (n=447).

Eye color
The three-SNP R^2 curve showed the third most frequent SNP to be rs10852218 (*OCA2*) at inflections 170 (1.6%), and 385 (3.6%). However, at inflections 1230 (11.5%) and 1455 (13.6%) its frequency was not pronounced compared to all other SNPs. A three-SNP MLR model of rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), and rs10852218 gave the highest R^2 value of 52.2% (n=223).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

5. APPENDIX

Submitted for Publication in the journal *Genetics* (manuscript attached).

"Genetics of Normal Human Pigmentation Variation Revealed by Multiple Linear Regression Modeling"

Authors: Robert K. Valenzuela, Miquia S. Henderson, Monica H. Kim, Nanibaa' A. Garrison, Jessica T. Kelch, Orit Cohen-Barak, Drew T. Erickson, F. John Meaney, J. Bruce Walsh, Keith C. Cheng, Shosuke Ito, Kazumasa Wakamatsu, Tony Frudakis, Matthew Thomas, and Murray H. Brilliant.

This work has also been presented at NIJ conferences and at the PanAmerican Society for Pigment Cell Research meetings.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.