Journal of Forensic Research
ISSN: 2157-7145

Valenzuela et al., J Forensic Res 2011, 2:8
DOI: 10.4172/2157-7145.1000139

Research Article Open Access

Prediction Model Validation: Normal Human Pigmentation Variation

Robert K. Valenzuela[1,2], Shosuke Ito[3], Kazumasa Wakamatsu[3] and Murray H. Brilliant[1,2]*

[1]Department of Pediatrics, College of Medicine, University of Arizona, Tucson, AZ 85724, USA
[2]Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI 54449, USA
[3]Department of Chemistry, Fujita Health University School of Health Sciences, Toyoake, Aichi, Japan

Abstract

In a past study, we developed multiple linear regression (MLR) models that employed three single nucleotide polymorphisms (SNPs) that predicted a significant proportion of variation in pigmentation phenotypes from a large population cohort (n=789, training sample). Multiple linear regression models were developed for skin reflectance, eye color, and two aspects of hair color (log of the ratio of eumelanin-to-pheomelanin and total melanin). In this report, using an independent cohort (n=242 , test sample), we 1) externally cross-validated the prediction models, and 2) tested and refined the algorithm presented in the study by Valenzuela and colleagues, (2010). Relative shrinkage was moderate for skin reflectance (23.4%), eye color (19.4%), and the log of the ratio of eumelanin-to-pheomelanin in hair (37.3%), and largest for total melanin (67%) in hair. Independent construction of predictive models using our algorithm for the test sample set yielded the same or similar models as the training sample set. Two of the three SNPs composing the models were the same, with some variability in the third SNP of the model.

Keywords: Forensic Science; Genetics; Human; Pigmentation; Prediction Models; QTL

Background

According to the Federal Bureau of Investigation (FBI) Laboratory's Combined DNA Index System (CODIS) – National DNA Index System (NDIS) statistics (http://www.fbi.gov/hq/lab/codis/clickmap.htm), there are significantly more unmatched profiles than there are matched profiles. Ancestry informative markers (AIMs) can be helpful in reducing the pool of suspects. However, a more efficient means of reducing a pool of suspects is to predict an unmatched profile's phenotype based on their genetic information. Forensically informative phenotypes include skin, eye, and hair color. The appearances of these traits are largely influenced by pigmentation, which is a quantitative trait controlled by many genetic loci.

In developing prediction models, interpretation of correlated genetic variants can be confounded by population stratification. When population stratification is not accounted for, erroneous inferences of a gene's involvement, and therefore false inferences of the biology of a trait, may be made. Clearly, accounting for population stratification is important in determining the biology of a trait. Correlation of a genetic marker to a trait may result if the marker is the causal variant that presumably affects the expression/function of a gene, if the marker is closely linked to a causal variant, or as a result of population stratification. Confounding genetic associations are markers that co-segregate with a trait that varies between populations, allele frequency differences are haphazardly associated with a trait due to unique evolutionary histories of each population. Therefore, by definition, ancestry informative markers (AIMs) are confounding associations with respect to a given trait in most instances. However, multiple studies have demonstrated that specific AIMs that are associated with melanin pigmentation are functional [1-3].

Melanin is the main pigment responsible for skin, eye, and hair color. Variation in a number of genes, including the melanocortin 1 receptor (MC1R), agouti-signaling protein (ASIP), oculocutaneous albinism 2 (OCA2), solute-carrier transport protein 45A2 (SLC45A2), and solute-carrier transport protein 24A5 (SLC24A5), have been associated with pigmentation. Functional and bioinformatics analyses support the biological role of variants (rs1805007, rs2424984, rs12913832, rs16891982, and rs1426654) associated with these genes.

The melanocortin 1 receptor (MC1R), a seven transmembrane G-protein coupled receptor located in the membrane of epidermal and follicular melanocytes, is a key protein involved in the regulation of melanin production (reviewed in [4]). Ligands of MC1R include the paracrine hormones, alpha-melanocyte stimulating hormone (α-MSH) and adrenocorticotropic hormone; both are produced in the keratinocytes associated with the melanocyte. They are derived from the precursor protein, proopiomelanocortin (POMC) (reviewed in [5]). The binding of α-MSH to MC1R causes a cAMP signal cascade resulting in an increased production of eumelanin. Non-synonymous SNPs, including rs1805007, used in this study, have been associated with red hair and fair skin [4,6]. Functional and bioinformatics studies have demonstrated that SNP rs1805007 alters the function of MC1R [7-9], and hence, melanin production.

The protein antagonist to signaling through MC1R is agouti-signaling protein (ASIP). The antagonistic action of ASIP results in a relative decrease in the production of eumelanin to pheomelanin (reviewed in [10]). Hence, MC1R acts as a switch between the two types of melanin for skin and hair melanoctyes. Several polymorphisms, within *ASIP*, including rs6058017 and rs2424984, have been associated with pigmentation variation. The rs6058017 polymorphism located within the 3' un-translated region (UTR) of the ASIP has been associated with normal human pigmentation variation of the skin [10-12], hair [11], and eyes [11,13]. In particular, the G allele is associated with increased

***Corresponding author:** Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI 54449, USA, Tel: +1 715-207-9493; Fax: +1 715-389-4950; E-mail: Brilliant.Murray@mcrf.mfldclin.edu

Received September 24, 2011; **Accepted** October 28, 2011; **Published** October 29, 2011

Citation: Valenzuela RK, Ito S, Wakamatsu K, Brilliant MH (2011) Prediction Model Validation: Normal Human Pigmentation Variation. J Forensic Res 2:139. doi:10.4172/2157-7145.1000139

eumelanin. The G allele is postulated to decrease the stability of the mRNA transcript, resulting in a decrease in ASIP, and consequently, a decrease in the antagonistic action on α-MSH. Additionally, a less studied polymorphism located within the vicinity of a conserved region of intron 1, rs2424984, was found to be more significantly associated with skin pigmentation variation across various populations compared to rs6058017 [14]. According to the National Center for Biotechnology Information (NCBI) website, there is a large difference (approximately 50%) in allele frequency of variants of rs2424984 between Blacks and non-Blacks. Hence, due to its large allele frequency differences between Blacks and non-Black populations, it may also be considered an AIM.

Many studies have shown that the putative transmembrane proteins OCA2, SLC45A2 (OCA4) (solute carrier transport protein, family 24, member 5; also called NCKX4), and SLC24A5 are essential for melanin production and are likely involved in regulating ion transport. Variation within these genes has been associated with variation in pigmentation. SLC45A2 and OCA2 have been localized to the melanosome surface [15,16], while studies have been conflicting on the locality of SLC24A5 [cf. [3,17]].

The OCA2 gene codes for a putative 12 trans-membrane protein [15]. The OCA2 protein has homology with anion transporters and is thought to be involved in influencing the pH of the melanosome [18-20], and either directly or indirectly involved in the trafficking of internal melanosomal proteins, tyrosinase (TYR) and tyrinosinase-related protein 1 (TYRP1) [21]. Polymorphisms of *OCA2* have been associated with skin, hair, and eye color variation. The strongest genetic variant associated with *OCA2* and eye color variation is rs12913832. This variant lies in an evolutionary conserved region of an intron of an adjacent gene (*HERC2*) located immediately upstream of *OCA2* and it has been hypothesized to be within a promoter region of *OCA2* [22]. In skin melanocytes, Cook et al. [2] found that the non-blue eye color variant of rs12913832 was associated with increased transcript levels of *OCA2*, supporting the hypothesis that rs12913832 is located within a promoter region of *OCA2* [2]. According to the National Center for Biotechnology Information (NCBI) website, the blue eye color variant has a frequency of approximately 80% in Whites, while in non-White populations, it is virtually non-existent. Hence, due to its large allele frequency differences between White and non-White populations, it may be considered an AIM. However, in this case, rs12913832 is also a functional variant.

SLC45A2 (OCA4, formerly known as MATP and AIM1) is a putative melanosomal membrane transport protein that is predicted to have 12 trans-membrane regions [16] localized to the melanosome [17]. SLC45A2 has regions that are similar to sucrose symporters in plants, and because of this it has been hypothesized to regulate osmosis by transporting sugar across the melanosome membrane [16]. In humans, Cook et al. [2] found that there was a greater amount of tyrosinase (TYR) associated with the dark-skin allele (374L) in melanocytes. Interestingly, they also found lower mRNA expression levels of samples homozygous for 374L. Non-synonymous genetic polymorphisms of *SLC45A2* (*OCA4*) that have been associated with pigmentation variation are rs16891982 (F374L) and rs26722 (E272K) [23,24]. The light-skin allele, 374F, decreases in frequency along a cline from northwest Europe to southeast Europe [25-27]. According to NCBI, the "light" allele of rs16891982 is present at very high frequency (i.e., >97%) in White populations. Similarly, in non-White populations, the "dark" allele is present in very high frequencies. Hence, rs16891982 is a functional AIM for melanin pigmentation.

Another important pigmentation gene, *SLC24A5* (formerly known as *NCKX5*), was found to cause the *golden* phenotype in zebrafish [1]. The main genetic variant of *SLC24A5* associated with pigmentation variation is rs1426654 (A111T) [1]. SLC24A5 was predicted to be a cation exchanger that transports Ca^{2+}/K^{+}, in exchange for Na^{+} [1], and more recently this function has been confirmed [3]. Lamason et al. [1] found that SLC24A5 was located on melanosomes or their precursors and hypothesized that it might function to accumulate Ca^{2+} into the melanosome [1]. Chi et al. [17] isolated SL24A5 in melanosomal fractions and proposed that it functioned on the surface of melanosomes [17]. Whereas, Ginger et al. [3] found SLC24A5 to be associated with the trans-golgi network [3]. They also found that the allele associated with darker skin, 111A, of SNP rs1426654 had a higher ion exchange activity compared to the allele associated with lighter skin, 111T. They hypothesized that SLC24A5 functions in regulating Ca^{2+} concentrations in endosomes, and that this affects delivery of melanosomal proteins (such as PMEL17), and hence, melanosome maturation. SLC24A5 may explain an earlier study demonstrating high Ca^{2+} concentrations in melanosomes [28]. The 111T allele was found to be almost fixed in European populations, while the 111A allele was found to be almost fixed in African and East Asian populations [1]. Hence, rs1426654 is a functional AIM for melanin pigmentation.

In our previous paper, we addressed the problem of constructing forensic models for skin, eye, and hair pigmentation by developing models using an ethnically diverse sample. The prediction models were comprised of SNPs that have been shown either through functional or bioinformatic analyses to be causal variants. To determine the performance of the models we developed [14], we report here the cross-validation of the pigmentation prediction models using an independent and ethnically diverse sample (test sample). We also corroborated the results of this algorithm (i.e., the models determined by the training sample) by applying the previously developed method to the test sample. Finally, we refined the algorithm and present a procedure that allows dynamic analysis of various SNP models and their R^2-curve inflections. This dynamic analysis allows us to observe the individual components (SNPs) of each possible model. This has facilitated the identification of more robust genetic models for describing pigmentation variation across various ethnicities.

Materials and Methods

Phenotype data, hair samples, and buccal cell samples were collected from each participant following Institutional Review Board approval of the protocol. Participants, phenotype measurements, and mathematical modeling have been described elsewhere [14], with the exception of the hair-melanin chemical analysis. The test sample's hair eumelanin chemical analysis was performed using a minor variation (an alkaline peroxide oxidation method rather than the acidic permanganate oxidation method that was used for the training sample) of the chemical analysis described elsewhere [29]. Briefly, sample homogenate (100 µL) was taken in a 10 ml screw-capped conical test tube, to which 375 µL 1 mol/L K_2CO_3 and 25 µL 30% H_2O_2 (final concentration: 1.5%) were added [30], and then mixed vigorously at room temperature for 20 hr. The residual H_2O_2 was decomposed by the addition of 50 µL 10% Na_2SO_3, and the mixture was then acidified with 140 µL 6 mol/L HCl. After vortex-mixing, the reaction mixture was centrifuged at 4,000 g for 1 min, and an aliquot (80 µL) of the supernatant was directly injected into the HPLC system. H_2O_2 oxidation products were analyzed with the HPLC system consisting of a JASCO 880-PU liquid chromatograph

(JASCO Co., Tokyo, Japan), a Shiseido C_{18} column (Shiseido Capcell Pak MG; 4.6 x 250 mm; 5 µm particle size) and a JASCO UV detector. The mobile phase was 0.1 mol/L potassium phosphate buffer (pH 2.1): methanol, 99: 1 (v/v). Analyses were performed at 45˚C at a flow rate of 0.7 mL/min. Absorbance of the eluent was monitored at 269 nm. The results of the two different chemical analysis methods have been shown to be highly correlated [31]. The test sample's data was transformed to match the training sample's data. SNPs were genotyped using the SNPlex™ Genotyping System (Applied Biosystems).

Design parameters of the SNPlex™ Genotyping System did not allow all of the significant SNPs of Valenzuela et al. [14] study to be genotyped, and additional SNPs that have been subsequently shown to be associated with human pigmentation were genotyped, so that the genotyping between the training and test sets was not identical. However, the test set was typed for the most significant SNPs from the training sample. The relationship between the SNPs genotyped for the training and test samples are illustrated in Figure 1.

Briefly, the inflection-point method for choosing the SNPs of the prediction models presented in Valenzuela et al. [14] was performed as described below. SNPs that were significant for a given trait as determined by one-way ANOVA were used to generate all possible combinations of three-SNP MLR models (statistical power considerations limited our models to three SNPs). R^2 values of all models were plotted in descending order and inflections in the resulting R^2 curve were noted. To find the basis of these inflections, we constructed barplots that contained all SNPs that comprised all models beginning with the model that corresponded to the highest R^2 value ending with the model that corresponded to the R^2 inflection. In doing so, it became obvious which SNPs were predominantly responsible for the inflections, as these SNPs corresponded to the most frequent SNPs in the barplot. The three most frequent SNPs were chosen as the final model for a given trait.

The aforementioned method was refined by devising a procedure that enabled visualization of the most frequent SNPs for any R^2 value, independent of barplots. This was accomplished by assigning the highest R^2 value model (all models sorted in descending values of R^2) a value of 1, and each subsequent model was numbered consecutively

$(i=1\ldots,\binom{x}{y})$;

where x=total-number-of-SNPs, and y=number-of-SNPs-in-model). For a given SNP, if it was present in a given model, then it was assigned a value of one, otherwise, it was assigned a value of zero (let presence/absence be called state). A function was chosen that weights the state of a given SNP more heavily for the highest R^2 values as compared to lower R^2 values and such that lower R^2-value models were dependant on higher R^2-value models. The preliminary function was as such,

$$f(i) = \frac{\sum_{i=1}^{i} state_i}{i}$$

When all MLR models were generated all SNPs were equally represented, hence each SNP was present a constant number of times. More precisely, each independent variable was represented a constant $\frac{(x-1)!}{(x-y)!(y-1)!}$ number of times when $\binom{x}{y}$ models were generated. Consequently, all independent variable functions must attain a value of $\frac{y}{x}$ at $i=\binom{x}{y}$. Additionally, a value of $\frac{\text{constant}}{i}$ was attained once an independent variable reached full representation.

In general, an increasing presence of a SNP as a function of i was reflected in its SNP-curve as a positive slope; in contrast, its diminishing presence was reflected in its SNP-curve as a negative slope (Figure 2). As a function of increasing i, the faster a SNP exhausted its representation (i.e., approached its asymptote), then the more important its contribution was to higher R^2 models (this is one way to view "prominent" contributors). Also, the greater the presence of an independent variable at higher R^2 values, then the more "important" it was as a contributor. If a SNP was present at each consecutive i, beginning from i=1, then the slope of its SNP-curve was zero, with a function value of one until its non-presence in a model.

Two parameters were varied in R^2-SNP curve generation/comparison for each trait: sample (training sample/test sample), and model size (3 SNPs/2 SNPs). SNP curves of different R^2 curves were compared by varying one parameter while the other parameter was held fixed.

Statistical analysis

Cross validation was performed by taking the difference in R^2 values of the training and test samples (i.e., shrinkage = $R^2_{training} - R^2_{test}$). Relative shrinkage was calculated by taking the ratio of shrinkage to the training sample's R^2 (i.e., $1-\frac{R^2_{test}}{R^2_{training}}$). All R^2 values were calculated by using the beta estimates of the training sample. Statistical values and models were calculated by using SAS (version 9.1) and JMP (release 8.0) statistical software packages (SAS Institute, Cary, North Carolina). All plots were graphed using R statistical freeware package (version 2.10.1) [32].

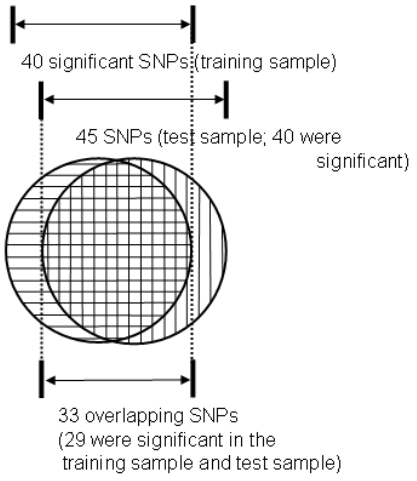

Figure 1: Relationship between SNP pools for the training and the test samples. Note: Significance was determined by ANOVA (alpha < 0.05).

	Caucasian	African-American	Hispanic	South Asian	East Asian	Native American	Admixture	not listed	TOTAL
Average Skin Reflectance	101	9	34	0	17	8	10	7	186
Eye Color	110	12	38	0	17	8	11	8	204
Hair ratio of Eumelanin-to-Pheomelanin	92	6	36	0	14	8	8	6	170
Total hair melanin	90	6	34	0	14	7	7	6	163

Table 1: Test cohort. Sample size of each ethnic group (self-reported) utilized for model validation.

	Training sample		Test sample			
	R^2	sample size	R^2	sample size	Shrinkage	Relative Shrinkage (%)
Skin reflectance	45.7	447	35.0	186	10.7	23.4
Eye color	76.4	353	61.6	204	14.8	19.4
Hair ratio of Eumelanin-to-Pheomelanin	43.2	162	27.1	170	16.1	37.3
Total hair melanin	76.3	143	25.2	163	51.1	67.0

Table 2: Cross validation results. R^2 values of the training sample and the test sample using the training sample's beta estimates.

Skin reflectance		rs16891982			rs1426654			rs2424984		
	intercept	GG	GC	CC	AA	AG	GG	TT	CT	CC
	59.8	3.0	-0.1	-2.9	1.1	-1.7	0.5	3.2	3.0	-6.2
Eye color		rs12913832			rs16891982			rs1426654		
	intercept	AA	GA	GG	GG	GC	CC	AA	AG	GG
	4.4	1.2	0.5	-1.8	-0.4	0.1	0.2	-0.3	0.2	0.1
Hair ratio of Eumelanin-to-Pheomelanin		rs16891982			rs12913832			rs1805007		
	intercept	GG	GC	CC	AA	GA	GG	CC	TC	CC
	4.6	-0.4	0.3	0.2	0.7	0.1	-0.8	0.8	-0.8	0.8
Total hair melanin		rs16891982			rs1426654			rs12913832		
	intercept	GG	GC	CC	AA	AG	GG	AA	GA	GG
	12011.2	-3096.2	163.4	2932.7	-2196.4	953.4	1243.1	2347.4	-115.4	-2232.0

Table 3: Prediction models' beta estimates derived from the training sample.

Results

To determine the predictive ability of the models generated by Valenzuela et al. [14], we externally cross-validated the models. Ethnic composition of the external sample set is listed in Table 1. External cross-validation was performed by taking the difference, or shrinkage, of corresponding R^2 values of each trait for each sample (Table 2). The R^2 values of the test sample were calculated by using the beta estimates derived from the training sample set's prediction models (Table 3).

We also tested the algorithm presented in Valenzuela et al. [14] by applying the algorithm to each sample set and comparing the results for each corresponding trait for each sample set. We generated three- and two-SNP R^2 curves (ie, 29-choose-3 and 29-choose-2, respectively) from which we determined three-SNP prediction models. All possible combinations of models were generated by using a pool of 29 SNPs (Figure 1) that were common to both sample sets and were found to be significant in each sample set by one-way ANOVA ($p < 0.05$; Table 4). We also generated SNP curves (see Materials and Methods) so that we could compare curves of a given trait between sample sets.

External validation

Skin reflectance: The model derived from the training sample set for the average skin reflectance was composed of SNPs rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), and rs2424984 (*ASIP*); together they yielded an R^2 value of 45.7% (n=447). Applying this model's beta estimates to the test sample set yielded an R^2 value of 35.0% (n=186); hence, the shrinkage was 10.7%, with a relative shrinkage of 23.4%.

Applying the algorithm to the training sample set and the test sample set, both the three- (i=395, Figure 3; and i=219 Figure 4) and two-SNP R^2 curves resulted in the same three SNPs: rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), and rs2424984 (*ASIP*). The corresponding SNP curves between the two sample sets were similar. In particular, inflections in the SNP curve of rs16891982 (*SLC45A2*) were often mirrored by inflections in the SNP curve of rs1426654 (*SLC24A5*) indicating that for many of the high R^2 models, either one or the other of the two SNPs was present. The exhaustion of rs16891982 (*SLC45A2*) resulted in a noticeable inflection in all skin reflectance R^2 curves.

Eye color: The model derived from the training sample set for

Gene	SNP	40 significant SNPs (training sample)	45 SNPs (test sample)	40 Significant SNPs (test sample)	33 overlapping SNPs	29 significant SNPs (training and test samples)	Total hair melanin	Natural log of ratio of melanins	Skin reflectance (CIEL)	Eye Color
AMACR	rs13289		+	+					*	
ASIP	rs2424984	+	+	+	+	+			*	
ASIP	rs6058017	+								
CYP4B1	rs1572603	+	+	+	+	+				*
DCT	rs1325611	+	+	+	+	+	*	*	*	*
DCT	rs1407995	+	+	+	+	+	*	*		*
GPR143	rs3044	+	+	+	+	+	*	*	*	*
HERC2	rs1129038		+	+			*	*	*	*
HERC2	rs12913832	+	+	+	+	+	*	*	*	*
HERC2	rs1667394		+	+			*	*	*	*
HERC2	rs916977		+	+			*	*	*	*
HPS3	rs2689234	+	+		+					
HPS4	rs1894704	+								
HPS4	rs3752589	+								
HPS4	rs3752590	+	+		+					
HPS4	rs739289	+	+	+	+	+				*
IRF4	rs12203592		+	+			*		*	*
MC1R	rs1805007	+	+	+	+	+		*		
MC1R	rs1805008	+	+	+	+	+		*		
MC1R	rs3212346	+	+	+	+	+	*		*	
MC1R	rs3212355	+								
MC1R	rs3212357	+								
MLPH	rs2292885	+	+	+	+	+		*	*	
MYO18A	rs11080078	+	+	+	+	+				*
MYO5A	rs1724630	+	+	+	+	+	*	*	*	*
MYO5A	rs2290332	+	+	+	+	+	*	*		*
MYO5A	rs752864	+	+	+	+	+	*			
MYO7A	rs2276289	+	+	+	+	+	*		*	*
MYO7A	rs3737454	+	+	+	+	+			*	*
near ASIP	rs1015362		+	+				*	*	
near KITLG	rs12821256		+	+			*			
near SLC24A4	rs12896399		+	+			*	*	*	
near TYRP1	rs1408799		+	+			*	*	*	*
OCA2	rs1037208	+	+	+	+	+			*	
OCA2	rs10852218	+	+	+	+	+			*	
OCA2	rs11638265	+	+	+	+	+	*	*	*	*
OCA2	rs1375164		+	+			*	*	*	*
OCA2	rs1800404	+	+	+	+	+	*	*	*	*
OCA2	rs1800407	+	+		+					
OCA2	rs1800410	+								
OCA2	rs1800411	+	+	+	+	+	*	*	*	*
OCA2	rs1800414	+								
OCA2	rs1900758	+	+	+	+	+	*	*	*	*

OCA2	rs749846	+	+	+	+	+	*	*		*
SLC24A5	rs1426654	+	+	+	+	+	*	*	*	*
SLC45A2	rs16891982	+	+	+	+	+	*	*	*	*
SLC45A2	rs2287949	+	+	+	+	+	*		*	*
SLC45A2	rs26722	+	+	+	+	+	*	*	*	*
SLC45A2	rs40132	+	+		+					
TPCN2	rs35264875		+							
TYR	rs1393350		+	+			*	*	*	*
TYRP1	rs2733832	+	+	+	+	+	*	*	*	*

* Significant SNPs (test sample) by ANOVA (p<0.05)

Table 4: 52 SNPs within or close to 22 genes.

eye color was composed of SNPs rs12913832 (*HERC2*), rs16891982 (*SLC45A2*), and rs1426654 (*SLC24A5*); together they yielded an R^2 value of 76.4% (n=353). Applying this model's beta estimates to the test sample yielded an R^2 value of 61.6% (n=204); hence, the shrinkage was 14.8%, with a relative shrinkage of 19.4%.

Applying the algorithm to the training sample set and the test sample set, both the three- (i=438, Figure 5; and i=464, Figure 6) and two-SNP R^2 curves resulted in the same three SNPs: rs12913832 (*HERC2*), rs16891982 (*SLC45A2*), and rs1426654 (*SLC24A5*). The corresponding SNP curves between the two sample sets were similar. In both sample sets, the SNP curve of rs12913832 (*HERC2*) was the highest frequency SNP until exhaustion, marked by a major inflection of the R^2 curve. The order that the SNP curves of rs16891982 (*SLC45A2*) and rs1426654 (*SLC24A5*) reached exhaustion varied between sample sets.

Eumelanin-to-pheomelanin ratio: The model derived from the training sample set for the natural logarithm of the ratio of eumelanin-to-pheomelanin was composed of SNPs rs16891982 (*SLC45A2*), rs12913832 (*HERC2*), and rs1805007 (*MC1R*); together yielding an R^2 value of 43.2% (n=162). Applying this model's beta estimates to the test sample yielded an R^2 value of 27.1%; hence, the shrinkage was 16.1%, with a relative shrinkage of 37.3%.

Applying the algorithm to the training sample set, the three-SNP R^2 curve of the training sample set resulted in the three SNPs (i=162, Figure 7): rs12913832 (*HERC2*), rs16891982 (SLC45A2), and rs1805007 (*MC1R*). The analogous inflection (e.g., a pronounced example of an analogous inflection between sample sets that can be seen by comparing the R^2 curves for eye color; Figures 5 and 6) of the three-SNP R^2 curve of the test sample set resulted in the three SNPs (i=149; Figure 8): rs12913832 (*HERC2*), rs16891982 (SLC45A2), and rs1426654 (*SLC24A5*). A more detailed inspection of the test sample's three-SNP R^2 curve revealed an inflection at i=35 (Figure 9). The three highest frequency SNPs at i=35 inflection were the same as the training sample set's at i=162: rs12913832 (*HERC2*), rs16891982 (SLC45A2), and rs1805007 (*MC1R*). Inflections in the SNP curve of rs12913832

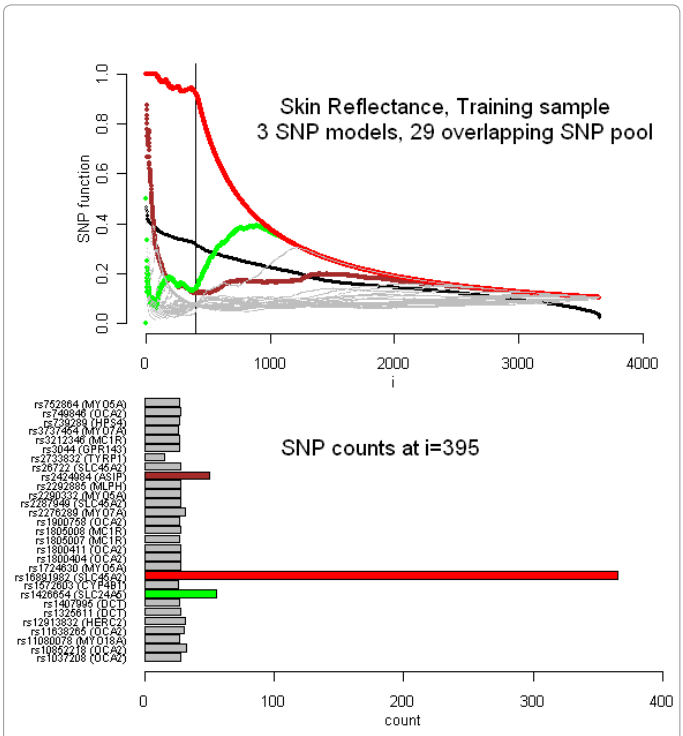

Figure 3: (**A**) Three-SNP multiple linear regression (MLR) models for skin reflectance across populations (training sample set). The horizontal-axis depicts all 3654 combinations (i.e., 29-choose-3) of significant SNPs in a three-SNP MLR model. The vertical-axis is the R^2 value for each model (black curve or three-SNP R2 curve) and also the SNP function value for each SNP curve. The R^2 curve inflection (i=395) is indicated by a vertical black line. The SNP curves of the three highest frequency SNPs at i=395 are indicated by colors (rs16891982 (*SLC45A2*), red; rs1426654 (*SLC24A5*), green; rs2424984 (*ASIP*), brown). (**B**) Bar plot of the SNPs that were present in all models from i=1 to i=395.

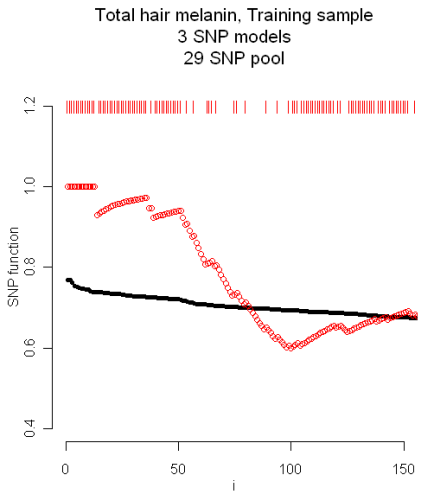

Figure 2: SNP-function (red curve) behavior in relationship to presence/absence (red tics) of SNP rs16891982 (*SLC45A2*) in a three SNP model. The R^2 curve is black.

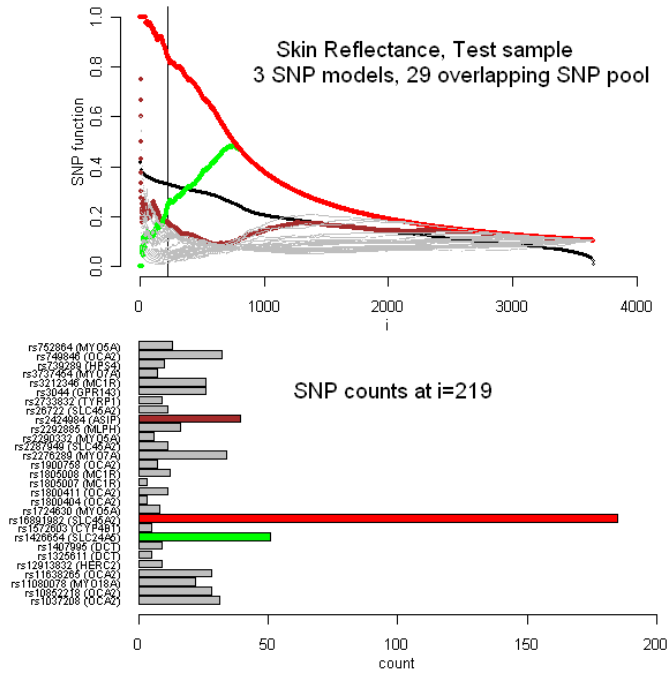

Figure 4: (**A**) Three-SNP multiple linear regression (MLR) models for skin reflectance across populations (test sample set). The horizontal-axis depicts all 3654 combinations (i.e., 29-choose-3) of significant SNPs in a three-SNP MLR model. The vertical-axis is the R^2 value for each model (black curve or three-SNP R^2 curve) and also the SNP function value for each SNP curve. The R^2 curve inflection (i=219) is indicated by a vertical black line. The SNP curves of the three highest frequency SNPs at i=219 are indicated by colors (rs16891982 (*SLC45A2*), red; rs1426654 (*SLC24A5*), green; rs2424984 (*ASIP*), brown). (**B**) Bar plot of the SNPs that were present in all models from i=1 to i=219.

(*HERC2*) were often mirrored by inflections in the SNP curve of rs16891982 (*SLC45A2*) for all R^2 curves; however, the SNP curves varied substantially between sample sets.

Total hair melanin: The model derived from the training sample set for hair total melanin was composed of SNPs rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), and rs12913832 (*HERC2*); together yielding an R^2 value of 76.3% (n=143). Applying this model's beta estimates to the test sample yielded an R^2 value of 25.2% (n=164); hence, the shrinkage was 51.1% with a relative shrinkage of 67.0%.

Applying the algorithm to the training sample set, both the three- and two-SNP R^2 curves resulted in the same three SNPs (i=180, Figure 10): rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), and rs12913832 (*HERC2*). However, applying the algorithm to the test sample set resulted in four SNPs (i=398, Figure 11) rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), rs12913832, and rs1800404 (OCA2); the latter two SNPs were of equal frequency. The test sample set's two-SNP R^2 curve resulted in SNP rs16891982 (SLC45A2); all other SNPs were of equal frequency. The SNP curves of rs16891982 (*SLC45A2*) and rs1426654 (*SLC24A5*), and consequently, corresponding R^2 curves, varied considerably between the samples. In the training sample set, rs16891982 (*SLC45A2*) was present in fewer high-R^2 models as compared to the test sample. However, their inflections were mirror images of each other in both sample sets.

Discussion

In this report, using an independent test sample (n=242) we externally cross-validated the pigmentation prediction models derived from the training sample (n=789) that we presented in Valenzuela et al. [14]. The relative shrinkage was modest for skin reflectance (23.4%), eye color (19.4%), and the ratio of eumelanin-to-pheomelanin of hair (37.3%), but was largest for hair total melanin (67.0%). We also refined the model building algorithm we presented in Valenzuela et al. [14] by adding SNP curves (see Materials and Methods) and tested the model building algorithm by applying it to both the training and test samples. The SNP curves gave us a better understanding of the behavior of the most prominent SNPs with respect to the R^2 curve inflections and in relationship to each other. We determined three-SNP models as we did in Valenzuela et al. [14], from three- and two-SNP model R^2 curves. In doing so, we found that the same third most prominent SNP, as was determined in Valenzuela et al. [14], could often be determined from the two-SNP model R^2 curves. Applying the algorithm to each sample set resulted in the same two SNPs, with variability in the third SNP, when comparing between sample sets for a given trait (total melanin, eumelanin-to-pheomelanin ratio, skin reflectance, and eye color).

Skin reflectance

Our skin reflectance model had a relatively low R^2 value (45.7%, training sample set) and a relative shrinkage of 23.4% when applied to the test sample set. The shrinkage was modest; hence, suggesting our model has forensic utility. The algorithm yielded the same three SNPs in both sample sets: rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), and rs2424984 (*ASIP*). The mirror-like behavior of rs16891982

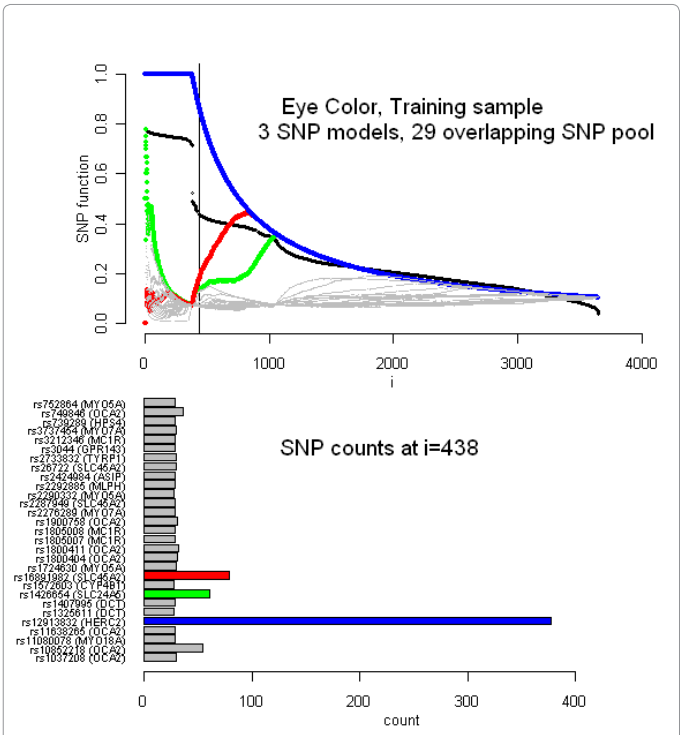

Figure 5: (**A**) Three-SNP multiple linear regression (MLR) models for eye color across populations (training sample set). The horizontal-axis depicts all 3654 combinations (i.e., 29-choose-3) of significant SNPs in a three-SNP MLR model. The vertical-axis is the R^2 value for each model (black curve or three-SNP R^2 curve) and also the SNP function value for each SNP curve. The R^2 curve inflection (i=438) is indicated by a vertical black line. The SNP curves of the three highest frequency SNPs at i=438 are indicated by colors (rs12913832 (*HERC2*), blue; rs16891982 (*SLC45A2*), red; rs1426654 (*SLC24A5*), green). (**B**) Bar plot of the SNPs that were present in all models from i=1 to i=438.

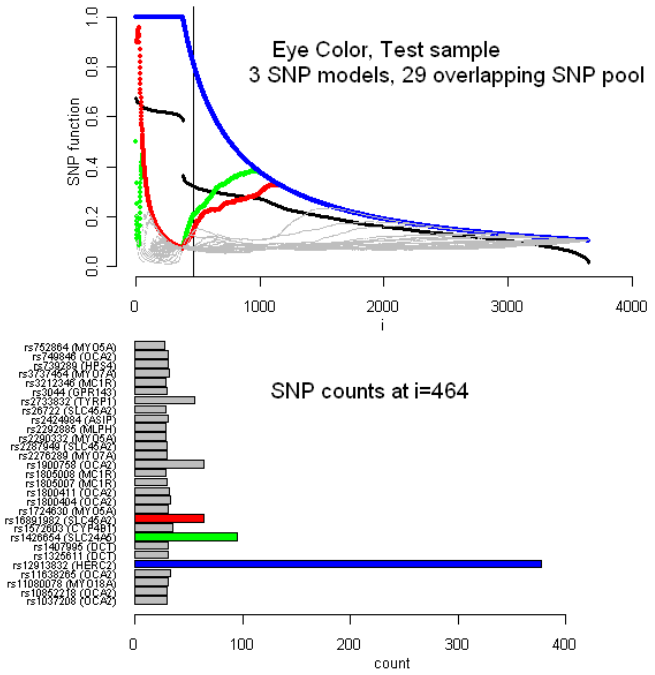

Figure 6: (**A**) Three-SNP multiple linear regression (MLR) models for eye color across populations (test sample set). The horizontal-axis depicts all 3654 combinations (i.e., 29-choose-3) of significant SNPs in a three-SNP MLR model. The vertical-axis is the R^2 value for each model (black curve or three-SNP R^2 curve) and also the SNP function value for each SNP curve. The R^2 curve inflection (i=464) is indicated by a vertical black line. The SNP curves of the three highest frequency SNPs at i=464 are indicated by colors (rs12913832 (*HERC2*), blue; rs16891982 (*SLC45A2*), red; rs1426654 (*SLC24A5*), green). (**B**) Bar plot of the SNPs that were present in all models from i=1 to i=464.

(*SLC45A2*) and rs1426654 (*SLC24A5*) was likely a result of their correlation (chi-square test; df=4; $\chi^2_{training}$=244.733; χ^2_{test}=115.302). Additional SNPs, not present in our pool of SNPs, likely will account for additional phenotypic variability. We have chosen SNPs that have been previously associated with macroscopic measurements of mouse/human pigmentation. To determine additional genetic associations of the macroscopic measurement (skin reflectance), microscopic (and perhaps chemical analysis) measurements are likely necessary. For example, Szabo et al. [33] showed that morphological differences exist in melanosome structure for various ethnicities. Conceivably, microscopic differences of pigment granules, or other differences, exist within ethnicities as well. Our measurements did not take into account these microscopic measurements, nor has any other study of which we are aware. Microscopic resolution may be necessary to determine SNPs that account for additional variation in skin reflectance. In other words, statistically significant genetic signals may be lost by grouping objects of similar macroscopic measurement that differ microscopically. Accounting for the genetic variations associated with these microscopic differences may enable development of models with increased predictive capabilities with relatively few SNPs.

Eye color

Similarly, for eye color, we took macroscopic measurements. However, in contrast to skin color, our model had a relatively high R^2 value (76.4%, training sample) and a relative shrinkage of 19.4% when applied to the test sample. The shrinkage was modest, thus forensically useful, suggesting that much of the variation in eye color is determined by relatively few SNPs, and that the SNPs from our SNP pool captured that variation. The algorithm yielded the same three SNPs in both sample sets: rs12913832 (*HERC2*), rs16891982 (*SLC45A2*), and rs1426654 (*SLC24A5*). The SNP curves were similar in behavior between samples. However, in the training sample rs16891982 (SLC45A2) reached exhaustion before rs1426654 (SLC24A5) did; whereas in the test sample, rs1426654 (SLC24A5) reached exhaustion first. The variability in SNP curves may be the result of experimental error in measurement, as we used an eye chart to record eye color, and/or it could be due to sampling error. More precise measurements [34,35] of intermediate eye colors are necessary in order to determine associated genetic signals, and therefore, to develop a prediction model that accurately describes intermediate eye color.

Eumelanin-to-pheomelanin ratio

Our hair melanin models are based on a sub-phenotype (melanin) of hair color. Our ratio of eumelanin-to-pheomelanin model had a relatively low R^2 value (43.2%, training sample set) and a relative shrinkage of 37.3% when applied to the test sample set. In contrast, our total hair melanin model had a relatively high R^2 value (76.3%, training sample set) and a relative shrinkage of 67.0% when applied to the test sample set. The large shrinkage in total melanin model was likely due to the different chemical analyses. Although the chemical analysis methods were highly correlated, the variance in correlation increased

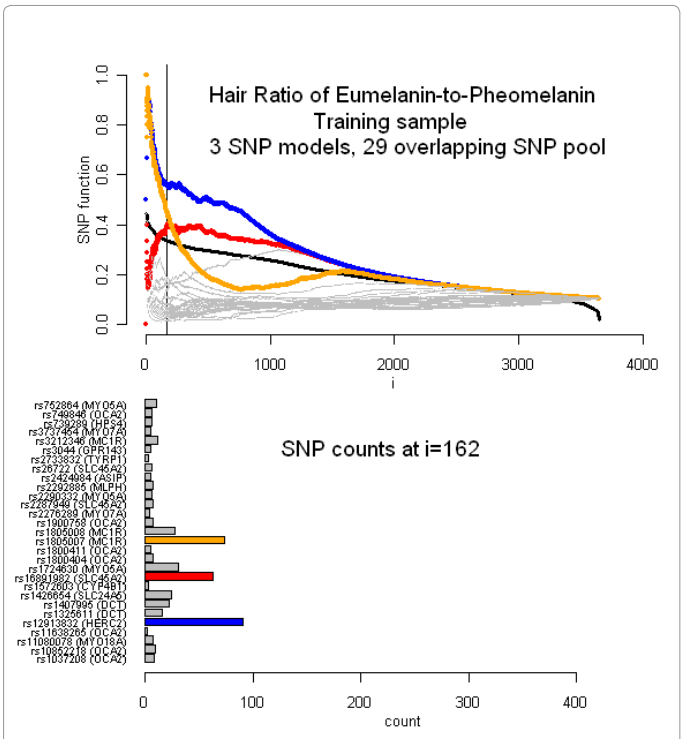

Figure 7: (**A**) Three-SNP multiple linear regression (MLR) models for natural log of hair melanin ratio's across populations (training sample set). The horizontal-axis depicts all 3654 combinations (i.e., 29-choose-3) of significant SNPs in a three-SNP MLR model. The vertical-axis is the R^2 value for each model (black curve or three-SNP R^2 curve) and also the SNP function value for each SNP curve. The R^2 curve inflection (i=162) is indicated by a vertical black line. The SNP curves of the three highest frequency SNPs at i=162 are indicated by colors (rs12913832 (*HERC2*), blue; rs1805007 (*MC1R*), orange ; rs16891982 (*SLC45a2*), red). (**B**) Bar plot of the SNPs that were present in all models from i=1 to i=162.

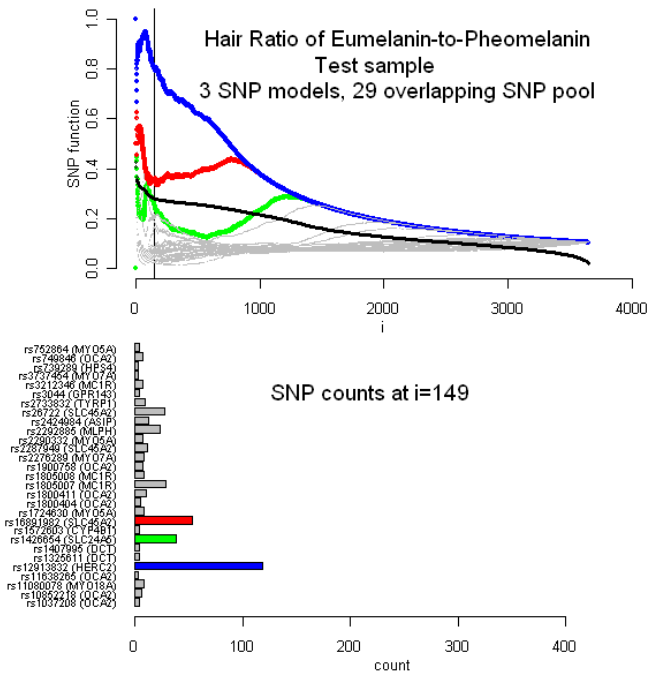

Figure 8: (**A**) Three-SNP multiple linear regression (MLR) models for natural log of the hair ratio's melanins across populations (test sample set). The horizontal-axis depicts all 3654 combinations (i.e., 29-choose-3) of significant SNPs in a three-SNP MLR model. The vertical-axis is the R^2 value for each model (black curve or three-SNP R^2 curve) and also the SNP function value for each SNP curve. The R^2 curve inflection (i=149) is indicated by a vertical black line. The SNP curves of the three highest frequency SNPs at i=149 are indicated by colors (rs12913832 (*HERC2*), blue; rs16891982 (*SLC45A2*), red; rs1426654 (*SLC24A5*), green). (**B**) Bar plot of the SNPs that were present in all models from i=1 to i=149.

at higher total melanin values [31], hence the decrease in correlation may explain the increase in relative shrinkage of our hair melanin models. Less shrinkage was likely observed in the ratio of eumelanin-to-pheomelanin model because of the natural log transformation of the data. Because of the different chemical methodologies employed, we cannot determine whether our hair models are forensically useful or not. Clearly, the efficacy of the model is highly contingent upon the method of measurement. Although hair color is largely influenced by melanin content, other (sub-quantitative) traits, such as hair thickness [36] and rate of growth, likely contribute to hair color. Hence, additional measurements may be necessary to accurately predict hair color.

The algorithm yielded the same three common SNPs in both sample sets (for both the three- and two-SNP R^2 curves): rs12913832 (*HERC2*), rs1805007 (*MC1R*), and rs16891982 (*SLC45A2*). However, in the test sample, SNPs rs16891982 (*SLC45A2*), and rs1426654 (*SLC24A5*) were of equal frequency for the "third" SNP (two-SNP R^2 curve). In comparing the SNP curves of rs16891982 (*SLC45A2*) and rs1426654 (*SLC24A5*) in the test sample, rs16891982 (*SLC45A2*) was clearly more prominent than rs1426654 (*SLC24A5*) after the major inflection. The mirror-like behavior of rs12913832 (*HERC2*) and rs16891982 (*SLC45A2*) was likely a result of their correlation (Pearson's chi-square test; df=4; $\chi^2_{training}$=74.8; χ^2_{test}=66.8).

Our initial choice of SNPs may be one of the reasons for the low R^2 value (43.2%) of the hair ratio of melanins (training data set). We chose SNPs from genes that have previously been associated with pigmentation; however, the ratio of melanins may be governed by other genes that are not detectable when eumelanin and pheomelanin are measured as a sum, but may be detectable when measured as a ratio. This is not surprising as the ratio of melanin, to our knowledge, has not been investigated at this level of detail and associated with genetic variants on a genome-wide scale. However, our choice of SNPs did enable development of the total melanin model that had a relatively high R^2 value.

Total hair melanin

The shrinkage result for total hair melanin was 51.1% with a relative shrinkage of 67.0%. This was likely the result of using different chemical analysis methods for the training sample and the test sample. The algorithm yielded the same three SNPs in both sample sets: rs16891982 (*SLC45A2*), rs1426654 (*SLC24A5*), and rs12913832 (*HERC2*). However, there was a marked difference in the behavior of SNP curve rs16891982 (*SLC45A2*) between samples. In the training sample, SNP curve rs16891982 (*SLC45A2*) slope varied between positive and negative, indicating that it was present in many, but not all, of the highest R^2 models; whereas in the test sample, the SNP curve of rs16891982 (*SLC45A2*) had a constant slope of zero, indicating that it was present in all of the highest R^2 models. The variance in SNP curve rs16891982 (*SLC45A2*) between samples was also reflected in the R^2 curves. The mirror-like behavior of rs16891982 (*SLC45A2*) and rs1426654 (*SLC24A5*) was likely a result of their correlation (Pearson's

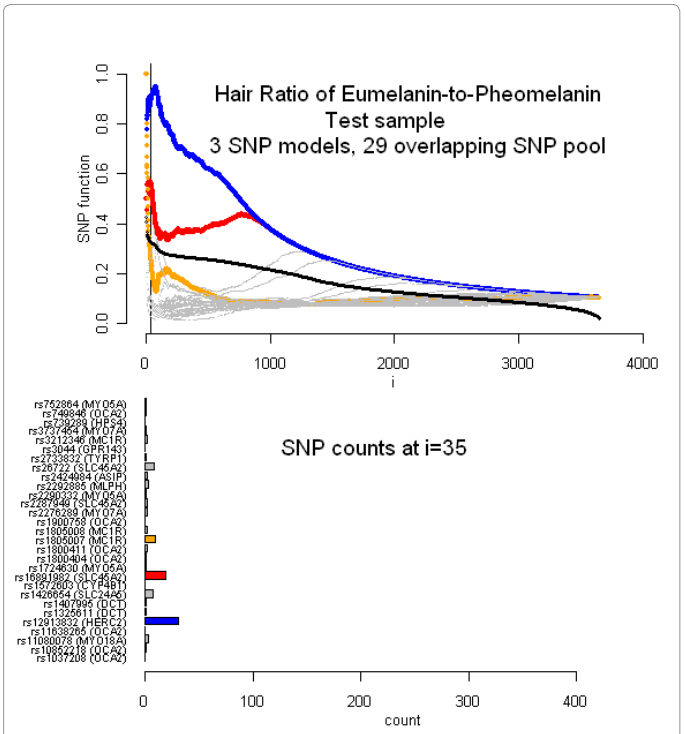

Figure 9: (**A**) Three-SNP multiple linear regression (MLR) models for natural log of the hair ratio's melanins across populations (test sample set). The horizontal-axis depicts all 3654 combinations (i.e., 29-choose-3) of significant SNPs in a three-SNP MLR model. The vertical-axis is the R^2 value for each model (black curve or three-SNP R^2 curve) and also the SNP function value for each SNP curve. The R^2 curve inflection (i=35) is indicated by a vertical black line. The SNP curves of the three highest frequency SNPs at i=35 are indicated by colors (rs12913832 (*HERC2*), blue; rs16891982 (*SLC45A2*), red; rs1805007 (*MC1R*), orange). (**B**) Bar plot of the SNPs that were present in all models from i=1 to i=35.

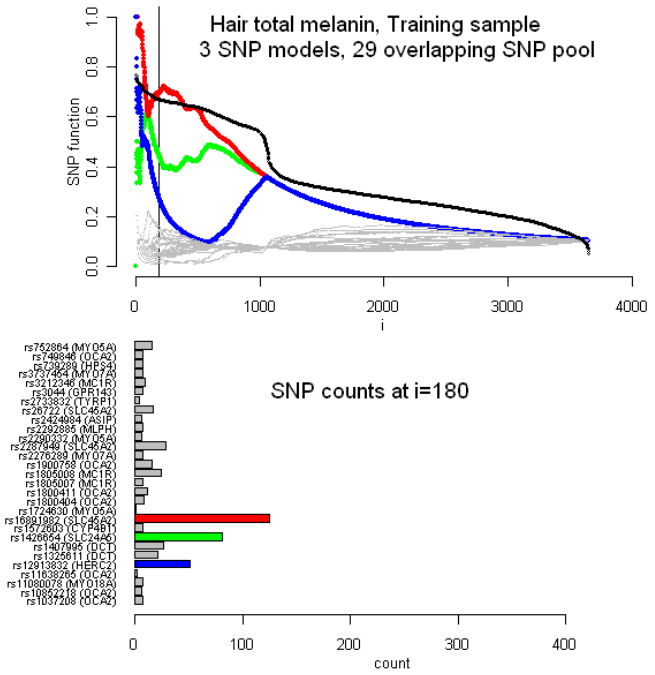

Figure 10: (**A**) Three-SNP multiple linear regression (MLR) models for hair total melanin across populations (training sample set). The horizontal-axis depicts all 3654 combinations (i.e., 29-choose-3) of significant SNPs in a three-SNP MLR model. The vertical-axis is the R^2 value for each model (black curve or three-SNP R^2 curve) and also the SNP function value for each SNP curve. The R^2 curve inflection (i=180) is indicated by a vertical black line. The SNP curves of the three highest frequency SNPs at i=180 are indicated by colors (rs12913832 (*HERC2*), blue; rs16891982 (*SLC45A2*), red; rs1426654 (*SLC24A5*), green). (**B**) Bar plot of the SNPs that were present in all models from i=1 to i=180.

chi-square test; df=4; $\chi^2_{training}$=124.288; χ^2_{test}=114.991). However, although the prominent SNP curves varied between samples, their relationship within a sample set remained unchanged, consequently supporting the argument that differences in algorithm results were due to the different chemical analysis methods.

Extending the comparison of determining the three most prominent SNPs from the two-SNP R^2, we also compared the SNPs determined by our algorithm to the three most significant SNPs, as determined by one-way ANOVA. We found that the third most prominent SNP, as determined from either two- or three-SNP R^2-curves, were not always the same as the three most significant SNPs as determined by one-way ANOVA. In particular, the third most prominent SNP of the skin reflectance model, as determined by the algorithm, was rs2424984 (*ASIP*). However, as determined by one-way ANOVA, it was the fourth most statistically significant SNP, while rs12913832 (*HERC2*) was the third most statistically significant SNP (training sample). Similarly, the third most prominent SNP of the natural log of the ratio of eumelanin-to-pheomelanin model, as determined by the algorithm, was rs1805007 (*MC1R*). However, as determined by one-way ANOVA, rs1805007 (*MC1R*) was the fourth most prominent SNP, while rs1426654 (*SLC24A5*) was the third most prominent SNP (training sample).

To determine if differential-missing SNP data could be attributed to the non-correspondence of the third SNP between the algorithm and ranking by one-way ANOVA, we selected the 10 most significant SNPs as determined by one-way ANOVA and removed all individuals with missing genotype information, such that all SNPs contributed the same amount of genetic information in all models. Applying the algorithm yielded the same prominent SNPs for skin reflectance (training sample). Interestingly, however, one-way ANOVA of the non-missing data set yielded a different ranking of the SNPs, such that rs2424984 (*ASIP*) was the seventh most significant SNP rather than the fourth most significant SNP, as was the case in the missing genotype data set.

Conclusion

The results demonstrate the utility of our algorithm for consistently selecting the same independent variables of a given trait for building prediction models. Additionally, the refinement of our algorithm, by adding curves of each independent variable (SNP curves), allowed us to determine the most frequent SNPs at any given inflection point. Whereas, before refinement, SNPs were selected by choosing an arbitrary inflection point and determining the most frequent SNPs from a barplot. Hence, the SNP curves condensed the barplot information from any point on the R^2-curve into one graph. The SNP curves also gave us insight into the behavior of prominent SNPs in relationship to each other (namely, covariance/co-inheritance), and between samples sets. Moreover, by comparing the algorithm results of two- and three-SNP R^2-curves, we found that the third most prominent SNP, as determined by the two-SNP R^2-curve, was often the same third SNP as determined by the three-SNP R^2-curve. Our results suggest that the third most prominent SNP may be inferred from the two-SNP R^2-curve. We note that a weakness to our algorithm [14] is that SNPs not significant by one-way ANOVA are excluded from the analysis; therefore, significant genetic interactions of non-significant single

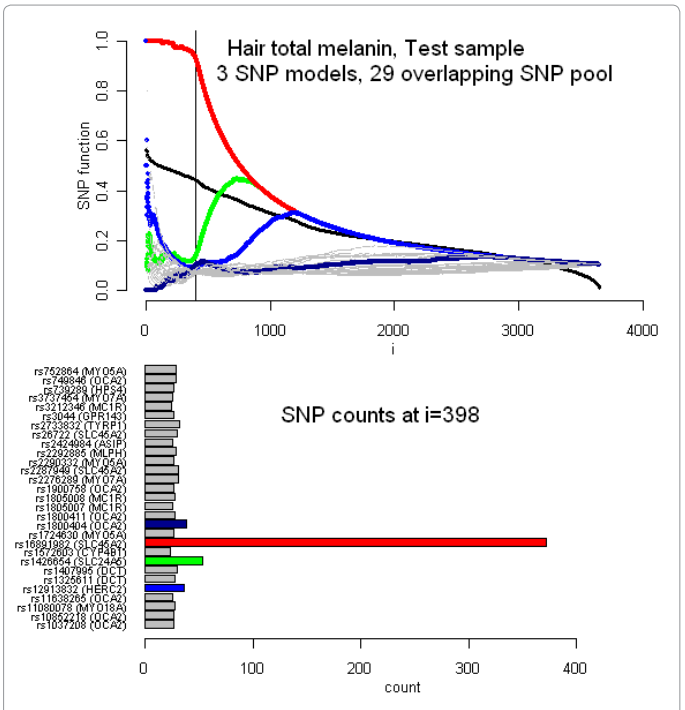

Figure 11: (**A**) Three-SNP multiple linear regression (MLR) models for hair total melanin across populations (test sample set). The horizontal-axis depicts all 3654 combinations (i.e., 29-choose-3) of significant SNPs in a three-SNP MLR model. The vertical-axis is the R^2 value for each model (black curve or three-SNP R^2 curve) and also the SNP function value for each SNP curve. The R^2 curve inflection (i=398) is indicated by a vertical black line. The SNP curves of the four highest frequency SNPs at i=398 are indicated by colors (rs16891982 (*SLC45A2*), red; rs1426654 (*SLC24A5*), green; rs12913832 (*HERC2*), blue; rs1800404 (*OCA2*), dark blue). (**B**) Bar plot of the SNPs that were present in all models from i=1 to i=180.

SNPs, as detected by the method presented by Akey et al. [37] may not be detected.

Our model building method, as with any model building method, strives to develop robust prediction models. These models are merely a starting place to *predict* normal human pigmentation variation, independent of ethnic origin. Other studies have developed prediction models for eye, skin, and hair color [38-43]. However, with the exception of the study by Spichenok et al., these studies trained their models utilizing a population of exclusively European descent. Not surprisingly, their models are lacking a major melanin associated SNP, rs1426654 (SLC24A5).

Garrison et al. [in preparation] utilized the software program structure [44] and 44 AIMs to distinguish ethnicities of a subset of the training sample set reported in Valenzuela et al. [14] We used self-described ethnicity of the training sample as a nominal predictor for skin reflectance, this resulted in an R^2 value of 0.56. Our skin reflectance model utilized three markers, resulting in an R^2 value of 0.45. Hence, although we may indirectly account for ethnicity through the utilization of AIMs to increase the predictive capability of our skin reflectance model, the cost of utilizing 44 markers likely will result in a loss of statistical power, not to mention additional costs.

We acknowledge the importance of controlling for population stratification for the purpose of making inferences about the biology of a trait. However, the purpose of this study was to validate models that are *predictive* for skin reflectance, eye color, and hair melanin pigmentation. Although we have developed and selected models that are comprised of genetic variants that have previously been functionally associated with pigmentation, we do not propose to have elucidated the biology of melanin pigmentation. However, in support of our models having biological relevance to pigmentation, studies suggest that the variants comprising our models are indeed functional [1-3,7-9,17]. We presented these models as an investigative tool in Valenzuela et al. [14] to predict externally visible pigmentation traits of an unidentified DNA donor [14]. In this study we validated our models on an independent data set, our results suggests that our skin reflectance and eye color models are predictive.

Competing Interest

The authors have no actual or potential competing interest to declare.

Author's Contributions

RKV devised the algorithm presented in manuscript, statistically analyzed data, genotyped, participated in phenotyping, and wrote the manuscript. KW and SI chemically analyzed hair samples. MHB conceived the study, participated in its design, coordination, and helped draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. JB Walsh (Department of Ecology & Evolutionary Biology University of Arizona, Tucson, Arizona, USA), Dr. LJ Baier (Diabetes Molecular Genetics Section Phoenix Epidemiology and Clinical Research Branch (NIDDK NIH), Phoenix, Arizona, USA), Dr. YL Muller (Diabetes Molecular Genetics Section Phoenix Epidemiology and Clinical Research Branch (NIDDK NIH) Phoenix, Arizona, USA), Dr. O. Cohen-Barak (Pharmacology Unit, Teva Pharmaceutical Industries Ltd., Netanya, Israel), Dr. DT Erickson (Ernest Gallo Clinic and Research Center, UCSF, San Francisco, California, USA), Christine Klassen (College of Medicine, University of Arizona, Tucson, Arizona, USA), Rongji Chen (University of Arizona, Tucson, Arizona, USA), Jason Fabian (University of Arizona, Tucson, Arizona, USA), and Justin Garrison (University of Arizona, Tucson, Arizona, USA) for their technical help with this project. We also would like to thank Dr. JB Walsh (University of Arizona, Tucson, Arizona, USA) Dr. RP Erickson (University of Arizona, Tucson, Arizona, USA) and Dr. SJ Shrodi (Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, Wisconsin, USA) for reviewing this manuscript prior to submission. We also thank the Marshfield Clinic Research Foundation's Office of Scientific Writing and Publication for editorial assistance of this manuscript.

Funding

The work was supported by a grant from the National Institute of Justice (2002-1J-CX-K010).

References

1. Lamason RL, Mohideen M-APK, Mest JR, Wong AC, Norton HL, et al. (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science 310: 1782-1786.
2. Cook AL, Chen W, Thurber AE, Smit DJ, Smith AG, et al. (2009) Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. J Invest Dermatol 129: 392-405.
3. Ginger RS, Askew SE, Ogborne RM, Wilson S, Ferdinando D, et al. (2008) SLC24A5 encodes a trans-Golgi network protein with potassium-dependent sodium-calcium exchange activity that regulates human epidermal melanogenesis. J Biol Chem 283: 5486-5495.
4. Flanagan N, Healy E, Ray A, Philips S, Todd C, et al. (2000) Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation. Hum Mol Genet 9: 2531-2537.
5. Dores RM (2009) Adrenocorticotropic hormone, melanocyte-stimulating hormone, and the melanocortin receptors: revisiting the work of Robert Schwyzer: a thirty-year retrospective. Ann N Y Acad Sci 1163: 93-100.
6. Branicki W, Brudnik U, Kupiec T, Wolañska-Nowak P, Wojas-Pelc A (2007) Determination of phenotype associated SNPs in the MC1R gene. J Forensic Sci 52: 349-354.
7. Frändberg PA, Doufexis M, Kapas S, Chhájlani V (1998) Human pigmentation phenotype: a point mutation generates nonfunctional MSH receptor. Biochem Biophys Res Commun 245: 490-492.
8. Xue D, Yin J, Tan M, Yue J, Wang Y, et al. (2008) Prediction of functional nonsynonymous single nucleotide polymorphisms in human G-protein-coupled receptors. J Hum Genet 53: 379-389.
9. Zhang C-S, Geng L-Y, Liu Z-Z, Fu Z-X, Gong Y-F, et al. (2011) A Comprehensive in silico Analysis of Functional and Structural Impact SNPS in the MC1R Gene. Journal of Animal and Veterinary Advances 10: 928-931.
10. Voisey J, Gomez-Cabrera MDC, Smit DJ, Leonard JH, Sturm RA, et al. (2006) A polymorphism in the agouti signalling protein (ASIP) is associated with decreased levels of mRNA. Pigment Cell Res 19: 226-231.
11. Kanetsky PA, Swoyer J, Panossian S, Holmes R, Guerry D, et al. (2002) A polymorphism in the agouti signaling protein gene is associated with human pigmentation. Am J Hum Genet 70: 770-775.
12. Bonilla C, Boxill L-A, Donald SAM, Williams T, Sylvester N, 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.
13. Frudakis T, Thomas M, Gaskin Z, Venkateswarlu K, Chandra KS, et al. (2003) Sequences associated with human iris pigmentation. Genetics 165: 2071-2083.
14. Valenzuela RK, Henderson MS, Walsh MH, Garrison NA, Kelch JT, et al. (2010) Predicting phenotype from genotype: normal pigmentation. J Forensic Sci 55: 315-322.
15. Gardner JM, Nakatsu Y, Gondo Y, Lee S, Lyon MF, et al. (1992) The mouse pink-eyed dilution gene: association with human Prader-Willi and Angelman syndromes. Science 257: 1121-1124.
16. Newton JM, Cohen-Barak O, Hagiwara N, Gardner JM, Davisson MT, 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.
17. Chi A, Valencia JC, Hu Z-Z, Watabe H, Yamaguchi H, et al. (2006) Proteomic

and bioinformatic characterization of the biogenesis and function of melanosomes. J Proteome Res 5: 3135-3144.

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

19. Chen K, Manga P, Orlow SJ (2002) Pink-eyed dilution protein controls the processing of tyrosinase. Mol Biol Cell 13: 1953-1964.

20. Cheli Y, Luciani F, Khaled M, Beuret L, Bille K, et al. (2009) {alpha}MSH and Cyclic AMP elevating agents control melanosome pH through a protein kinase A-independent mechanism. J Biol Chem 284: 18699-18706.

21. Manga P, Boissy RE, Pifko-Hirst S, Zhou BK, Orlow SJ (2001) Mislocalization of melanosomal proteins in melanocytes from mice with oculocutaneous albinism type 2. Exp Eye Res 72: 695-710.

22. Sturm RA, Duffy DL, Zhao ZZ, Leite FPN, Stark MS, et al. (2008) A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am J Hum Genet 82: 424-431.

23. Graf J, Hodgson R, van Daal A (2005) Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. Hum Mutat 25: 278-284.

24. Branicki W, Brudnik U, Draus-Barini J, Kupiec T, Wojas-Pelc A (2008) Association of the SLC45A2 gene with physiological human hair colour variation. J Hum Genet 53: 966-971.

25. Nakayama K, Fukamachi S, Kimura H, Koda Y, Soemantri A, et al. (2002) Distinctive distribution of AIM1 polymorphism among major human populations with different skin color. J Hum Genet 47: 92-94.

26. Yuasa I, Umetsu K, Harihara S, Kido A, Miyoshi A, et al. (2006) Distribution of the F374 allele of the SLC45A2 (MATP) gene and founder-haplotype analysis. Ann Hum Genet 70: 802-811.

27. Lucotte G, Mercier G, Diéterlen F, Yuasa I (2010) A decreasing gradient of 374F allele frequencies in the skin pigmentation gene SLC45A2, from the north of West Europe to North Africa. Biochem Genet 48: 26-33.

28. Dräger UC (1985) Calcium binding in pigmented and albino eyes. Proc Natl Acad Sci USA 82: 6716-6720.

29. Wakamatsu K, Fujikawa K, Zucca FA, Zecca L, Ito S (2003) The structure of neuromelanin as studied by chemical degradative methods. J Neurochem 86: 1015-1023.

30. Wakamatsu K, Ohtara K, Ito S (2009) Chemical analysis of late stages of pheomelanogenesis: conversion of dihydrobenzothiazine to a benzothiazole structure. Pigment Cell Melanoma Res 22: 474-486.

31. Ito S, Nakanishi Y, Valenzuela RK, Brilliant MH, Kolbe L, et al. (2011) Usefulness of alkaline hydrogen peroxide oxidation to analyze eumelanin and pheomelanin in various tissue samples: application to chemical analysis of human hair melanins. Pigment Cell Melanoma Res 24: 603-615.

32. R: A Language and Environment for Statistical Computing (2009). Vienna, Austria: R Foundation for Statistical Computing.

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

34. Sturm RA, Larsson M (2009) Genetics of human iris colour and patterns. Pigment Cell Melanoma Res 22: 544-562.

35. Liu F, Wollstein A, Hysi PG, Ankra-Badu GA, Spector TD, et al. (2010) Digital Quantification of Human Eye Color Highlights Genetic Association of Three New Loci. PLoS Genet 6: e1000934.

36. Vaughn MR, Brooks E, van Oorschot RAH, Baindur-Hudson S (2009) A comparison of macroscopic and microscopic hair color measurements and a quantification of the relationship between hair color and thickness. Microsc Microanal 15: 189-193.

37. Akey JM, Wang H, Xiong M, Wu H, Liu W, 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.

38. Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, 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.

39. Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, et al. (2009) Eye color and the prediction of complex phenotypes from genotypes. Current Biology 19: R192-R193.

40. Walsh S, Lindenbergh A, Zuniga SB, Sijen T, de Knijff P, et al. (2010) Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int Genet 5: 464-471.

41. Mengel-From J, Wong TH, Morling N, Rees JL, Jackson IJ (2010) Genetic determinants of hair and eye colours in the Scottish and Danish populations. BMC Genet 10: 88.

42. Spichenok O, Budimlija ZM, Mitchell AA, Jenny A, Kovacevic L, et al. (2010) Prediction of eye and skin color in diverse populations using seven SNPs. Forensic Sci Int Genet 5: 472-478.

43. Branicki W, Liu F, van Duijn K, Draus-Barini J, Pośpiech E, et al. (2011) Model-based prediction of human hair color using DNA variants. Human Genetics 129: 443-454.

44. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.