



Original article

Eye color prediction using single nucleotide polymorphisms in Saudi population

Jahad Alghamdi^a, Manal Amoudi^b, Ahmad Ch. Kassab^c, Mansour Al Mufarrej^d, Saleh Al Ghamdi^{a,*}^a King Abdullah International Medical Research Center/King Saud bin Abdulaziz University for Health Sciences, Medical Genomics Research Department, Ministry of National Guard – Health Affairs, Riyadh, Saudi Arabia^b King Abdullah International Medical Research Center/King Saud bin Abdulaziz University for Health Sciences, Pathology and Laboratory Medicine, Ministry of National Guard – Health Affairs, Riyadh, Saudi Arabia^c Naif Arab University for Security Sciences, Riyadh, Saudi Arabia^d King Abdullah International Medical Research Center/King Saud bin Abdulaziz University for Health Sciences, Ophthalmology Department, Ministry of National Guard – Health Affairs, Riyadh, Saudi Arabia

ARTICLE INFO

Article history:

Received 6 August 2018

Revised 27 September 2018

Accepted 27 September 2018

Available online 28 September 2018

Keywords:

Externally visible characteristics

Eye color prediction

IrisPlex

Forensics DNA phenotyping

ABSTRACT

Background: DNA prediction of eye color represent one application of the externally visible characteristics (EVC), which attained growing interest in the field of DNA forensic phenotyping. This is mainly due to its ability to narrow the pool of suspects without the need to compare any retrieved DNA material from the crime scene to a reference DNA. Several methods and multiplex genetic panel were proposed with variable prediction accuracy between different populations. However, such panel was not previously tested in the Saudi population, nor any populations of the Middle East and North Africa origin.

Method: A panel of eleven single nucleotide polymorphisms (SNPs) was tested for their association with three eye colors (brown, hazel, and intermediate) in 80 volunteer Saudi individuals. SNPs and haplotype association test with eye colors were performed to identify the top significant SNPs with the three eye colors. Also, multinomial logistic regression was used to construct the prediction model using a training set of 60 subjects, and a validation set of 20 subjects. The goodness of fit parameter of the model to correctly predicts each eye color as compared to the other was performed.

Results: Eye color was significantly associated with rs12913832, rs7170852, and rs916977 that are located within *HERC2*. SNP rs12913832 was the top significant SNP (p -value = $1.78E-15$) that accounted for the association in this region, as the other SNPs were not significant after adjusting for rs12913832. A prediction model containing five SNPs showed high prediction accuracy with Area Under the receiver operating characteristic Curves (AUC) equals to 0.95 and 0.83 for brown and intermediate eye colors, respectively. However, the model's performance was very low for predicting the hazel eye color with AUC equals 0.75.

Discussion: Despite the small sample size of our study, we reported very significant SNP associations with eye color. Our model to predict eye colors based on DNA material showed high accuracy for brown and intermediate eye colors. The eye color prediction-model underperformed for the hazel eye colors, suggesting that larger sample size, as well as more comprehensive set of SNPs, could improve the model-prediction accuracy.

© 2018 Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding author. King Abdullah International Medical Research Center, Ministry of National Guard – Health Affairs, P.O. Box 3660, Mail Code 1515, Riyadh 11481, Saudi Arabia.

E-mail address: alghamdismg@gmail.com (S. Al Ghamdi).

Peer review under responsibility of King Saud University.



Production and hosting by Elsevier

1. Introduction

Utilizing genetic information from DNA materials has received a grown interest in DNA-based prediction of the externally visible characteristics (EVCs), especially for forensic genetics (Kayser and Schneider, 2009; Tully, 2007). The application of such practice has been known as Forensic DNA Phenotyping (FDP), which enable a prediction of the appearance of individuals based on their DNA that can be for instance collected from the crime scene. Such

<https://doi.org/10.1016/j.sjbs.2018.09.011>

1319-562X/© 2018 Production and hosting by Elsevier B.V. on behalf of King Saud University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

evidence can be very useful in limiting the number of potential suspects if the perpetrator is unknown, allowing the police investigators to focus the investigations to a specific group of people with the predicted EVCs. The potential pool of suspects can be narrowed as more reliable EVCs are available for predication methods. Importantly, unlike conventional DNA profiling that usually requires comparing the identified DNA to a reference, either a governmental criminal DNA databases or dragnets screening, EVCs can help in tracing unknown suspects initially without the need to compare their DNA to references using tools that were carefully developed through scientific research (Kayser and Schneider, 2009). The group of suspects identified by their EVCs can be then subjected to conventional DNA profiling to specifically identify the perpetrator. This process is deemed to be difficult given the multifactorial nature of the EVCs with several factors that contribute to the phenotypic variability, including gene-gene interaction, and gene-environmental interactions (Pośpiech et al., 2011).

Several studies have developed and assessed EVC for eye, skin, and hair in different Europeans populations (Branicki et al., 2011; Dario et al., 2015; Liu et al., 2009; Walsh et al., 2011a). In particular, three multiplex panels that include a various set of Single Nucleotide Polymorphisms (SNPs) have been developed and tested in different populations (Mengel-From et al., 2010; Ruiz et al., 2013; Walsh et al., 2011b). In particular, Mengel-From et al. (2010) reported a high predictive value for rs12913832 (*HERC2*) for dichotomous eye color category (i.e., light/dark). They also reported additional variants that are associated with eye color in *OCA2* and *SLC45A2*. The IrisPlex is another panel developed by Liu et al. (2009) based on six SNPs with a prediction accuracy up to 93% and 91% for brown and blue eye colors. However, their prediction accuracy was very low for the intermediate (73%) with a sensitivity of 1.1% as compared to the other two colors. The panel includes six SNPs that were found to be a major contributor to eye color: rs12913832 (*HERC2*), rs1800407 (*OCA2*), rs12896399 (*SLC24A4*), rs16891982 (*SLC45A2*), rs1393350 (*TYR*), rs12203592 (*IRF4*) (Liu et al., 2009). The IrisPlex system was later validated in three independent laboratories, showing that the model-based prediction probability estimation generated from the IrisPlex system has a high level of sensitivity, specificity, reliability, and consistency, especially for blue and brown eye color (Walsh et al., 2011a). Recently, Snipper system [http://mathgene.usc.es/snipper] was developed by Ruiz et al. (2013) based on 37 SNPs in pigmentation-associated genes to study SNP-genotype based prediction of eye, skin, and hair color variation.

Population-specific factors play major roles in the predicting power of these models, as several studies reported significant differences in allele frequencies between populations (Dario et al., 2015; Kayser and Schneider, 2009; Yun et al., 2014). Importantly, no previous study has tested the genetic factors of eye color in the Saudi population specifically, or even in any population of the Middle East and North Africa (MENA) origin. Our study aims to test the genetic association of four genes with eye color in the Saudi population and to compare the allele frequency of the identified variants with other population. We also aim to develop an eye color prediction model based on the identified significant genetic variants.

2. Materials and methods

2.1. Subjects and phenotyping

A total of 80 Saudi were voluntarily recruited in this study, each of them was randomly chosen and requested to complete the study questionnaire and ethical consent that was approved by the ethics committee at King Abdullah International Medical Research Cen-

ter. Inclusion criteria included generally healthy male or female above 18 years old. Exclusion criteria included pregnancy and any history of diseases or conditions that may affect iris color or pigmentation genes, such as iris transplantation, albinisms, as well as taking any medication to treat glaucoma, chemotherapy, or hormonal therapy. The study participants were divided into “known group” as a training set for building the prediction model that included 60 subjects, and “blind sample” for that included the remaining 20 subjects for testing the model. The eye color was measured according to subject’s self-description in the questionnaire and by a clinically trained ophthalmologist at the Ophthalmology Clinic at National Guard Health Affairs (NGHA) – King Abdulaziz Medical City – Riyadh. A colorful photograph of the subject right eye were taken using Haag-Streit Eye Slit Lamps Machine, which contains: a macro digital camera, under a constant setting, LED light button on 8, background light on the middle, Magnification = 6.3, camera angle 90° over 50°, preset exposure = 80, gain = 25, intensity brightens = 7, contract = 12, and gamma = 1.4. The eye color of the blind samples was only known by the ophthalmologist, and their records were not revealed after constructing the model. Each subject was then graded as having a “brown”, “intermediate” or “hazel” eye color. A 5 ml venous blood sample was then collected in EDTA tube from each subject. Also, a Buccal swab from the inner cheek area was also collected from the blind group only with a dry sterile swab, to compare their genotyping sensitivity to blood samples. This was accomplished by comparing genotyping results obtained from Buccal swabs to those obtained from blood samples in those subjects.

2.2. Genomic DNA extraction, multiplexing, and sequencing

Genomic DNA was extracted from both peripheral blood and Buccal swabs using QIAamp® DNA extraction kits according to manufacturer’s instructions (Qiagen, Hagen, Germany). DNA concentration was measured using Nanodrop spectrophotometer and then was stored at –80 until the time of performing PCR reactions. PCR was performed to amplify ten DNA fragments, containing eleven targeted SNPs that were distributed among four genes located on chromosome 5 and 15 (Table 1). Primer3 online tool was used to design primer pairs flanking each of the selected SNPs in

Table 1
Selected SNPs and PCR primers.

Gene	SNPs	#Chr	SNP	Primer
<i>OCA2</i>	rs1800407	15	A/G	R:GAAGAGGAGCATGGTGGTAA F:CTCTGTTCTACGACCTGCT
<i>OCA2</i>	rs1375164	15	T/C	R:CTACCCCTGTGTGAAGAGCA F:ACAAGGGTCTTCTCTGGGAGG
<i>OCA2</i>	rs7495174	15	G/A	R:GCGGACCTTAAACTCACTCC F:CTGAACCTTCCAACTCCCC
<i>HERC2</i>	rs12913832 rs7183877	15	G/A C/A	R:TCAGAGGCCAGTGACAATT F:GCGAGGCCAGTTTCATTGTA
<i>HERC2</i>	rs7170852	15	T/A	R:AGCAAGGTGGAAGGAGCAA F:ACGATGATACCCAGGCCTT
<i>HERC2</i>	rs916977	15	T/C	R:ACTGCTGACTTTTGTCTCTGT F:TAAACAGGATGGAGCCGAGA
<i>SLC45A2</i>	rs16891982	5	G/C	R:AGAGGTGGAGAAGCAGAGTG F:ACCCACTGATTCGAAGAGCA
<i>SLC45A2</i>	rs26722	5	C/T	R:TGGGTACAGAATTCAGGTCA F:ACAACAACAACAAGAGCAAGA
<i>SLC24A5</i>	rs1834640	15	G/A	R:GTCAGATGCAGGTTGAAGCT F:TTCGCGTTGTGTCATCTTG
<i>SLC24A5</i>	rs1426654	15	G/A	R:CCTATGTGCAAGATAGTACCAGC F:AGAAGCAAAACATTGGACTCTTT

addition to a universal M13 primer extension to help in the co-amplification of all fragments. A multiplex PCR with the ten amplicons was designed to amplify the loci with the eleven selected SNPs. In which, a total of 1 μ l of 20 ng/ μ l genomic DNA was amplified in 10 μ l PCR reaction with; 1.5 μ l PCR primer, five μ l BigDye[®] Direct PCR Master Mix, and up to 2.5 μ l Deionized water. PCR Thermal cycling was performed using Veriti[®] Thermal Cycler with the following program: (1) 95 °C for 10 min, (2) 35 cycles of 96 °C for 3 s, 62 °C for 15 s, and 68 °C for 30 s, (3) 2 min at 72 °C. All the PCR products were then analyzed by running them on 2% Agarose Gel Electrophoresis and stained by fluorescent dye Ethidium Bromide to visualize the DNA fragments.

Sanger sequencing was performed on ice following the protocol of Applied BioSystems BigDye[®] Direct Cycle Sequencing Kit. For each reaction, 1.0 μ l of the PCR product was added to 2.0ml of BigDye[®] Direct Sequencing Master Mix and 1.0 ml of BigDye[®] Direct M13 Forward Primer and run using the following Cycle Sequencing thermal conditions using Veriti[®] Thermal Cycler: (1) 37° for 15 min, (2) 80° for 2 min, (3) 96° for 1 min, (4) 25 cycles of 96° for 10 s, 50° for 5 s, and 60° for 7 s. The sequencing product was then purified using BigDye XTerminator Purification Kit. Finally, Capillary gel electrophoresis was performed using Applied Biosystems[®] 3730xl sequencer, and the Variant Analysis module implemented in the ThermoFisher Cloud [www.thermofisher.com/it/en/home/cloud.html] was used to recall the variants genotype for each subject.

2.3. Statistical analyses

Single SNP association was carried out using a linear model where intermediate, hazel and brown eye colors were coded as 1, 2, and 3 quantitatively, and genetic variants were coded as 0, 1, and 2 minor alleles. SNPs descriptive analysis and association test were performed using PLINK v1.07 (Purcell et al., 2007). Also, a multimarker test was performed by constructing haplotype and performing logistic haplotype associations using Haploview 4.2 (Barrett et al., 2005), which was also used to estimate the linkage disequilibrium (LD) and define the LD blocks. For the multimarker test, the eye color was dichotomized into light color (i.e., intermediate/hazel) or dark color (i.e., brown). The eye color prediction model was constructed using a Multinomial Logistic Regression Model, as described by Liu et al. (2009), and implemented in the eye color prediction model of the IrisPlex system (Walsh et al., 2011b). Furthermore, a prediction model based on multinomial logistic regression was constructed to assess the accurate prediction of eye color, and sensitivity analysis of the model was performed by reporting sensitivity, specificity, negative predictive

value (NPV), positive predictive value (PPV) and the area under the receiver characteristics operating curves (AUC), which ranges from 0.5 reflecting that the model provides no better prediction ability than chance, up to 1.0 reflecting a model with a perfect prediction ability. The analyses were carried out by “mlogit” package for the freely available statistical software “R”, version 3.43. For each subject in the validation set, the prediction probability value of having hazel, intermediate, or brown eye color was generated based on the parameters generated from the training set, and the formula provided by Liu et al. (2009). The predicted color is then informed from the eye color category that had the highest probability. Finally, the sensitivity of the Buccal swab DNA sample to the DNA blood samples was compared among the 20 samples by reporting the sensitivity parameters, considering the blood DNA sample as the gold standard method.

3. Results

3.1. SNPs association test

The eye color distribution across subjects included in the training set (n = 60) consisted of 25 (41.7%) Brown, 17 (28.3%) Hazel and 18 (30%) Intermediate eye color. The SNPs association test is presented in Table 2, which shows that rs12913832 located within the *HERC2* gene is the most significantly associated SNPs with eye color (P-value = 1.78E-15; Table 2). The minor allele (G) is associated with the largest effects of having non-brown eyes. All the individuals with Brown eye color were homozygous for the A allele, except one heterozygous individual only. Similarly, only one subject with Intermediate eye color was homozygous for the A allele, and the vast majority of remaining Intermediate eye color were homozygous for the G allele. Another two SNPs that are located within the same LD blocks (rs12913832 and rs916977) were also significantly associated with eye color (Fig. 1). Two SNPs within OCA2 were significantly associated with eye color at the nominal P-value (<0.05). Also, rs1800407 (OCA2) was excluded from all the analysis as it was not polymorphic in this population with all subjects carrying the C allele. Although the LD analysis showed that the top three SNPs were not in the same LD blocks. None of the significant SNPs remained significantly associated with eye color after performing a conditional analysis on rs12913832. The haplotype analysis showed two haplotypes that were statistically significant with the dichotomous eye colors (Table 3). For instance, the haplotype AG of rs7495174 and rs12913832 was more common among subjects with light eye color (54.3%) as compared to dark colors (2%) (P-value = 1.65E-9). A total of 446 genotypes were available for comparison between buccal swab and

Table 2
SNPs association test with eye color in Saudi population.

Chr	SNP ^a	A ¹ /A ²	A ¹ F	Genotypic frequency ^a			Beta ^b	P-val	P-val ^c
				Brown	Hazel	Intermediate			
5	rs16891982	G/C	0.16	0.08, 0.12, 0.8	0.06, 0.12, 0.82	0.11, 0.17, 0.72	-0.09	0.61	0.25
15	rs1375164	T/C	0.46	0.24, 0.36, 0.4	0.29, 0.53, 0.18	0.17, 0.28, 0.56	0.32	0.02	0.62
15	rs7495174	G/A	0.17	0.08, 0.36, 0.56	0, 0.24, 0.76	0, 0.11, 0.89	0.52	0.01	0.71
15	rs12913832	G/A	0.33	0, 0.04, 0.96	0, 0.53, 0.47	0.67, 0.28, 0.06	-0.86	1.78E-15	-
15	rs7183877	A/C	0.12	0, 0.16, 0.84	0.06, 0.41, 0.53	0, 0.17, 0.83	-0.06	0.78	0.03
15	rs7170852	T/A	0.43	0.12, 0.48, 0.4	0.35, 0.41, 0.24	0, 0.28, 0.72	0.60	6.62E-06	0.87
15	rs916977	T/C	0.47	0.08, 0.48, 0.44	0.24, 0.53, 0.24	0.06, 0.22, 0.72	0.61	3.83E-06	0.57
15	rs1834640	G/A	0.07	0.04, 0.16, 0.8	0, 0.18, 0.82	0, 0.06, 0.94	0.37	0.16	0.61
15	rs1426654	G/A	0.06	0.04, 0.12, 0.84	0, 0.12, 0.88	0, 0.06, 0.94	0.34	0.23	0.84

SNPs in bolds are statistically significant at the multiple-testing corrected values.

^a The frequency are presented for A¹/A¹, A¹/A², and A²/A², respectively, per each eye color within the training set (n = 60).

^b Effect size for each addition copy of A¹.

^c SNPs association test conditional on rs12913832.

^{*} rs1800407 was not polymorphic in this population with all subjects having the allele C.

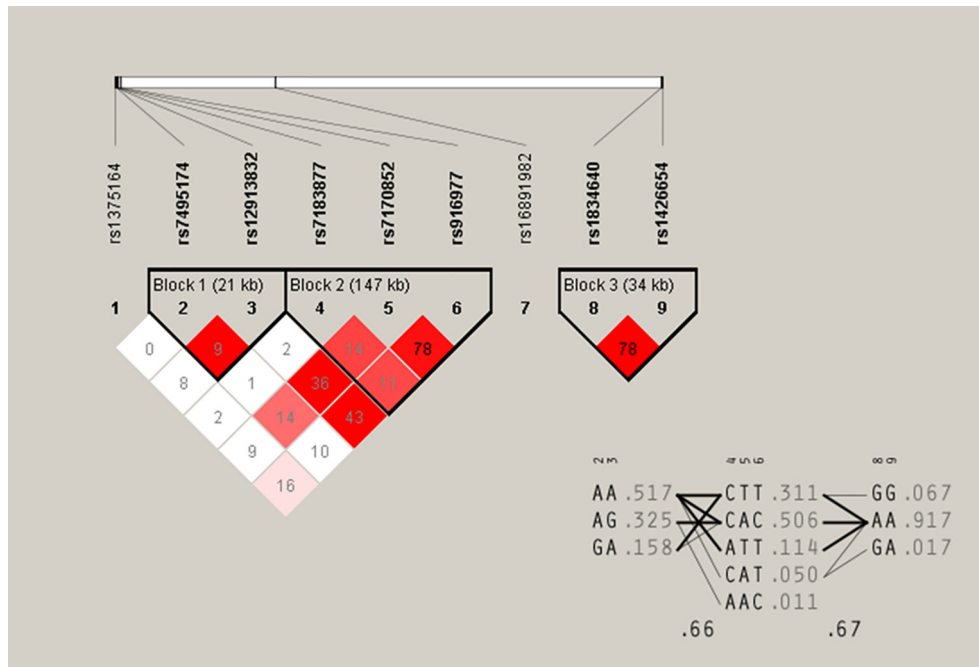


Fig. 1. LD blocks and haplotype frequency.

Table 3
Haplotype association tests.

Block	Haplotype	Freq.	Light, Dark eye Ratio Counts	Light, Dark eye Frequencies	Chi Square	P Value
Block 1	AA	0.517	26.0:44.0, 36.0:14.0	0.371, 0.720	14.191	2.00E-04
	AG	0.325	38.0:32.0, 1.0:49.0	0.543, 0.020	36.347	1.65E-09
	GA	0.158	6.0:64.0, 13.0:37.0	0.086, 0.260	6.648	0.0099
Block 2	CAC	0.506	45.7:24.3, 15.0:35.0	0.653, 0.300	14.526	1.00E-04
	CTT	0.311	9.3:60.7, 28.0:22.0	0.132, 0.560	24.909	6.01E-07
	ATT	0.114	10.7:59.3, 3.0:47.0	0.153, 0.060	2.485	0.1149
	CAT	0.05	3.0:67.0, 3.0:47.0	0.043, 0.060	0.169	0.681
	AAC	0.011	1.3:68.7, 0.0:50.0	0.018, 0.000	0.914	0.3389
Block 3	AA	0.917	66.0:4.0, 44.0:6.0	0.943, 0.880	1.509	0.2194
	GG	0.067	3.0:67.0, 5.0:45.0	0.043, 0.100	1.531	0.216
	GA	0.017	1.0:69.0, 1.0:49.0	0.014, 0.020	0.058	0.8095

SNPs included in each haplotype block are specifically defined in Fig. 1. SNPs in bolds are statistically significant at the multiple-testing corrected values.

blood sample. Of them, 432 genotypes were concordant between the two methods, giving a sensitivity of 92.7% for buccal swab.

3.2. Prediction model

The estimated multinomial logistic regression model to predict eye color is shown in Table 4. Only SNPs that were significantly

associated with eye color, as assessed by the likelihood ratio test (P -value < 0.05) were retained in the final model. Hence, the final model included five SNPs only. The ability of the model to predict the eye color was tested on the validation set ($n = 20$) by estimating the probabilities of having each color for each subject. The color with the highest predicted probabilities was then assigned to the patients as the predicted color. Table 5 shows the goodness-of-fit

Table 4
Prediction model parameters for the fitted model to eye color.

SNP	Allele	Hazel			Intermediate			Likelihood ratio test
		beta	se	p	beta	se	p	
rs12913832	G	3.07	1.12	0.006	6.11	1.51	5.30E-05	4.91E-13
rs1375164	T	-0.47	0.41	0.26	-0.96	0.44	0.03	0.07
rs1426654	G	-0.49	0.81	0.55	-1.18	1.08	0.27	0.44
rs16891982	G	-0.14	0.56	0.81	0.27	0.48	0.58	0.75
rs1834640	G	-0.31	0.71	0.66	-1.45	1.09	0.18	0.29
rs7170852	T	-0.80	0.47	0.09	-2.38	0.64	2.03E-04	2.57E-05
rs7183877	A	1.53	0.70	0.03	0.05	0.81	0.95	0.04
rs7495174	G	-0.97	0.64	0.13	-1.78	0.82	0.03	0.03
rs916977	T	-0.78	0.48	0.11	-2.43	0.64	1.49E-04	1.51E-05

Multinomial logistic regression model parameters of hazel and intermediate eye color to the brown reference level. Only significant SNPs (in bold) were included in the final model.

Table 5
Prediction model accuracy parameters.

Parameter	Saudi population ^a			Reference ^a		
	Hazel	Intermediate	Brown	Blue ^b	Intermediate	Brown
AUC	0.75	0.83	0.95	0.91	0.73	0.93
Sensitivity (%)	17.0	75.0	100	93.4	1.1	88.4
Specificity (%)	86.0	100	64.0	77.1	99.6	88.0
PPV (%)	33.0	100	55.0	89.8	25.0	67.1
NPV (%)	71.0	86.0	100	84.4	90.0	96.5

^a Our model is constructed based on five SNPs, only rs12913832 is common to our model and the reference IrisPlex model reported by Liu et al (2009).

^b No subject had blue eye color in the Saudi population.

measure of the model, specifically, AUC, sensitivity, specificity, PPV, and NPV. These measures were derived from three 2×2 contingency tables, in which the predicting abilities of predict one color (e.g., brown) is compared to the remaining two colors (e.g., hazel and intermediate). The AUC of the three colors is also shown using ROC in Fig. 2. These results showed that the model prediction ability is high for predicting brown (AUC = 95%) and intermediate (83%) eye color, as compared to hazel eye color (AUC = 75%). Among the 20 samples with both buccal swab and blood samples, a concordant prediction was obtained by the two methods for 19 samples (12 truly predicted, and 7 wrongly predicted), and only one sample was correctly predicted by blood sample but not the buccal swab.

4. Discussion

This study investigated the genetic determinants of eye colors in the Saudi population, which represents a valuable source for genetic and genomics studies due to the high rate of consanguinity (Alkuraya, 2014). In our study, we reported a significant association between rs12913832 (*HERC2*) with eye colors. LD haplotype containing this SNP was also significantly associated with the two categories eye colors (i.e., light/dark), along with another haplotype that included three SNPs within the *HERC2* locus. The developed predictive model based on our SNPs set achieved a 95% prediction accuracy for the brown eye color, and 83% prediction accuracy for the intermediate eye color. Interestingly, all subjects

with Brown eye color were correctly predicted by the prediction model, as well as 75% of subjects with Intermediate eye color. However, the model poorly predicts the hazel eye color with an AUC of 75%, and only truly predicting 17% of subjects with Hazel eye color. In accordance with previous research describing the significant difference in allele frequency of variants associated with eye colors, this was clearly evident in our population by two observations; first, we found that rs1800407 was not polymorphic in our population with the absent of allele T. Second, the G allele of rs12913832 was found to be the minor allele in our population with a frequency of 33%, as compared to being the major allele in the European populations (72%; Walsh et al., 2011b). This is consistent with the differences in the phenotypic differences in eye color between the different populations; particularly that almost half of our population had brown eye color.

Our study replicated the association of *HERC2* gene with eye colors in the Saudi Population. In particular, rs12913832 was the top significantly associated variants with the eye color, similar to several previous studies (Mengel-From et al., 2010; Ulivi et al., 2013). This variant seems to explain all the phenotypic variance in this region, as the other two top significant SNPs lost their statistical significance after accounting for rs12913832. The SNP rs12913832 is located within the intronic region of *HERC2*, 21 kb upstream of the pigment *OCA2* gene. The region surrounding this variant was found to enhance the *OCA2* promoter via a long-range chromatin loop, and enhancer activity is mediated by the other transcription factors (Visser et al., 2012). This study reported that rs12913832 –A allele that is predominant allele in our subjects with Brown eye color is found to robustly recruit these transcription factors to promote *OCA2* genes, and consequently increase the melanin production and the dark iris pigmentation. Although none of the tested SNPs within the *OCA2* gene showed a significant association with eye color, the haplotype analysis showed that rs12913832 was in complete linkage disequilibrium ($D' = 1$) with rs7495174. Interestingly, the AG haplotype that included the two SNPs was significantly associated with light eye color. The rs7495174 was firstly reported by Duffy et al. (2007) to be part of a three SNPs haplotype, located in intron 1 of *OCA2*, that is associated with blue/brown color. This study reported that the haplotype including the A allele is found in 90.5% of subjects with blue/green eyes compared to only 1% of the Brown eyes. Also, this haplotype was more prevalent in subjects with light brown hair, suggesting that this haplotype acts as a recessive modifier of lighter. Although we did not genotype the same set of SNPs within this haplotype, we reported consistent findings of a higher frequency of the A allele in the Intermediate and Hazel eye color as compared to brown eyes carrier.

Our eye-color prediction model that includes five SNPs showed high accuracy for the brown and intermediate eye colors. Compared to the 6 SNPs IrisPlex SNPs suggested by Liu et al. (2009), our model generally had a better goodness-of-fit measures for the brown and intermediate eye colors (Table 3). For instance, all

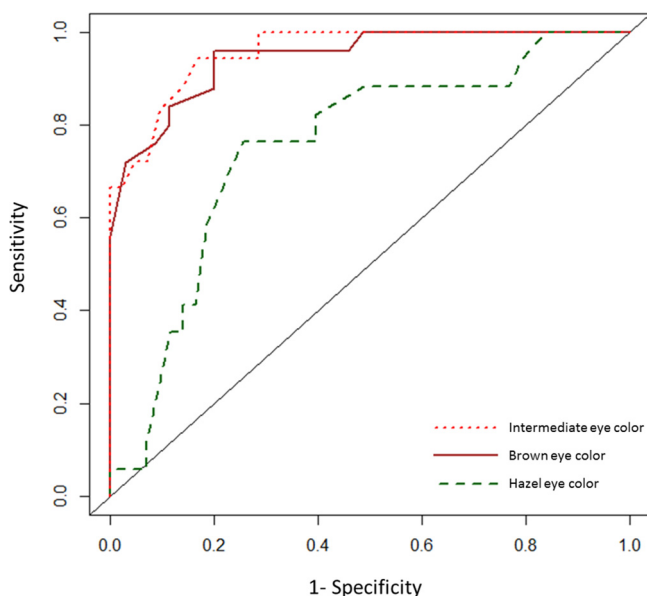


Fig. 2. Receiver operating characteristics curve analysis of the prediction model for the three eye colors.

subjects with brown eyes were correctly predicted by our model yielding AUC of 95% and 83% for the brown and intermediate eye colors, respectively. Both of PPV and NPV were also high in our model, particularly for the intermediate eye color. Comparing to other populations, our model showed better performance than Slovenians (Kastelic et al., 2013), Portuguese (Dario et al., 2015), and South Americans (Freire-Aradas et al., 2014). The higher performance of our model is especially observed for the prediction of brown and intermediate eye colors. Similar to previous studies, our prediction model did not include gender as it was to be found to be insignificantly associated with eye color in our population. Also, including gender into the prediction model had marginal and ambiguous impact on the overall accuracy of prediction model (Pośpiech et al., 2016). This study highlight that further research is still needed to expand the set of SNPs to improve the model predictions for the uncommon eye colors in the Saudi population, including the hazel eye color.

Our study is limited by the sample size that may not be large enough to be truly representative of the Saudi population, as well as to cover all the possible eye colors. However, our randomly selected sample had a fairly similar proportion of the three tested eye colors. Second, our study genotyped a small set of SNPs that do not explain all the phenotypic variability in the eye colors. Also, a direct comparison with the IrisPlex system was not possible due to the small differences in the genotyped SNPs. Hence, we suggest that genotyping more SNPs including those contained within the IrisPlex system may even improve the prediction accuracy in our population. Finally, our prediction model is based on biological samples that were taken under controlled setting; this may not be the case for a forensic sample retrieved from a crime scene. Hence, further replication of the model using forensic sample is needed.

In conclusion, despite the multifactorial nature of EVCs that are under the influence of interactions of genetic and environmental factors, we provided a genetic prediction model that can accurately predict eye color in the Saudi population. Also, our results confirm the previous results of the association of HERC2 (rs12913832) with eye colors. Importantly, the DNA prediction model showed the higher accuracy for predicting brown and intermediate eye colors. However, uncertainty remained high for the hazel eye color. This suggested that increasing the sample size as well as expanding the SNPs set may improve the model prediction accuracy for the uncommon eye color such as hazel eye color.

Acknowledgments

We are grateful to all volunteers who provided DNA sample for this study, and taking the time to visit the Ophthalmology clinic at the Ministry of National Guard Health Affairs. We are additional thankful for the Ophthalmology clinic team. This work was funded by King Abdullah International Medical Research Center (KAIMRC).

Conflict of interest

The authors declare no conflict of interest.

References

- Alkuraya, F.S., 2014. Genetics and genomic medicine in Saudi Arabia. *Mol. Genet. Genom. Med.* 2, 369–378. <https://doi.org/10.1002/mgg3.97>.
- Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. <https://doi.org/10.1093/bioinformatics/bth457>.
- Branicki, W., Liu, F., Van Duijn, K., Draus-Barini, J., Pośpiech, E., Walsh, S., Kupiec, T., Wojas-Pelc, A., Kayser, M., 2011. Model-based prediction of human hair color using DNA variants. *Hum. Genet.* 129, 443–454. <https://doi.org/10.1007/s00439-010-0939-8>.
- Dario, P., Mouriño, H., Oliveira, A.R., Lucas, I., Ribeiro, T., Porto, M.J., Costa Santos, J., Dias, D., Corte Real, F., 2015. Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population. *Int. J. Legal Med.* 129, 1191–1200. <https://doi.org/10.1007/s00414-015-1248-5>.
- Duffy, D.L., Montgomery, G.W., Chen, W., Zhen Zhao, Z., Le, L., James, M.R., Hayward, N.K., Martin, N.G., Sturm, R.A., 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. <https://doi.org/10.1086/510885>.
- Freire-Aradas, A., Ruiz, Y., Phillips, C., Maroñas, O., Söchtig, J., Tato, A.G., Dios, J.Á., de Cal, M.C., Silbiger, V.N., Luchessi, A.D., Luchessi, A.D., Chiurillo, M.A., Carracedo, Á., Lareu, M.V., 2014. Exploring iris colour prediction and ancestry inference in admixed populations of South America. *Forensic Sci. Int. Genet.* 13, 3–9. <https://doi.org/10.1016/j.fsigen.2014.06.007>.
- Kastelic, V., Pośpiech, E., Draus-Barini, J., Branicki, W., Drobnič, K., 2013. Prediction of eye color in the Slovenian population using the IrisPlex SNPs. *Croat. Med. J.* 54, 381–386. <https://doi.org/10.3325/cmj.2013.54.381>.
- Kayser, M., Schneider, P.M., 2009. DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. *Forensic Sci. Int. Genet.* 3, 154–161. <https://doi.org/10.1016/j.fsigen.2009.01.012>.
- Liu, F., van Duijn, K., Vingerling, J.R., Hofman, A., Uitterlinden, A.G., Janssens, A.C.J., W., Kayser, M., 2009. Eye color and the prediction of complex phenotypes from genotypes. *Curr. Biol.* 19, 192–193. <https://doi.org/10.1016/j.cub.2009.01.027>.
- Mengel-From, J., Børsting, C., Sanchez, J.J., Eiberg, H., Morling, N., 2010. Human eye colour and HERC2, OCA2 and MATP. *Forensic Sci. Int. Genet.* 4, 323–328. <https://doi.org/10.1016/j.fsigen.2009.12.004>.
- Pośpiech, E., Draus-Barini, J., Kupiec, T., Wojas-Pelc, A., Branicki, W., 2011. Gene-gene interactions contribute to eye colour variation in humans. *J. Hum. Genet.* 56, 447–455. <https://doi.org/10.1038/jhg.2011.38>.
- Pośpiech, E., Karłowska-Pik, J., Ziemkiewicz, B., Kukla, M., Skowron, M., Wojas-Pelc, A., Branicki, W., 2016. Further evidence for population specific differences in the effect of DNA markers and gender on eye colour prediction in forensics. *Int. J. Legal Med.* 130, 923–934. <https://doi.org/10.1007/s00414-016-1388-2>.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. <https://doi.org/10.1086/519795>.
- Ruiz, Y., Phillips, C., Gomez-Tato, A., Alvarez-Dios, J., Casares De Cal, M., Cruz, R., Maroñas, O., Söchtig, J., Fondevila, M., Rodriguez-Cid, M.J., Carracedo, Á., Lareu, M.V., 2013. Further development of forensic eye color predictive tests. *Forensic Sci. Int. Genet.* 7, 28–40. <https://doi.org/10.1016/j.fsigen.2012.05.009>.
- Tully, G., 2007. Genotype versus phenotype: human pigmentation. *Forensic Sci. Int. Genet.* 1, 105–110. <https://doi.org/10.1016/j.fsigen.2007.01.005>.
- Uliivi, S., Mezzavilla, M., Gasparini, P., 2013. Genetics of eye colours in different rural populations on the silk road. *Eur. J. Hum. Genet.* 21, 1320–1323. <https://doi.org/10.1038/ejhg.2013.41>.
- Visser, M., Kayser, M., Palstra, R.J., 2012. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res.* 22, 446–455. <https://doi.org/10.1101/gr.128652.111>.
- Walsh, S., Lindenbergh, A., Zuniga, S.B., Sijen, T., De Knijff, P., Kayser, M., Ballantyne, K.N., 2011a. Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence. *Forensic Sci. Int. Genet.* 5, 464–471. <https://doi.org/10.1016/j.fsigen.2010.09.008>.
- Walsh, S., Liu, F., Ballantyne, K.N., Van Oven, M., Lao, O., Kayser, M., 2011b. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci. Int. Genet.* 5, 170–180. <https://doi.org/10.1016/j.fsigen.2010.02.004>.
- Yun, L., Gu, Y., Rajeevan, H., Kidd, K.K., 2014. Application of six IrisPlex SNPs and comparison of two eye color prediction systems in diverse Eurasia populations. *Int. J. Legal Med.* 128, 447–453. <https://doi.org/10.1007/s00414-013-0953-1>.