ORIGINAL ARTICLE

Allelic Frequencies of 20 Visible Phenotype Variants in the Korean Population

Ji Eun Lim, Bermseok Oh*

Department of Biomedical Engineering, Kyung Hee University School of Medicine, Seoul 130-701, Korea

The prediction of externally visible characteristics from DNA has been studied for forensic genetics over the last few years. Externally visible characteristics include hair, skin, and eye color, height, and facial morphology, which have high heritability. Recent studies using genome-wide association analysis have identified genes and variations that correlate with human visible phenotypes and developed phenotype prediction programs. However, most prediction models were constructed and validated based on genotype and phenotype information on Europeans. Therefore, we need to validate prediction models in diverse ethnic populations. In this study, we selected potentially useful variations for forensic science that are associated with hair and eye color, iris pattern, and facial morphology, based on previous studies, and analyzed their frequencies in 1,920 Koreans. Among 20 single nucleotide polymorphisms (SNPs), 10 SNPs were polymorphic, 6 SNPs were very rare (minor allele frequency < 0.005), and 4 SNPs were monomorphic in the Korean population. Even though the usability of these SNPs should be verified by an association study in Koreans, this study provides 10 potential SNP markers for forensic science for externally visible characteristics in the Korean population.

Keywords: eye color, gene frequency, hair color, Korean, single nucleotide polymorphism

Introduction

DNA genotyping has been used in forensic science for over 30 years for human identification by using short tandem repeat (STR) markers, which are highly polymorphic variants [1]. Recently, single nucleotide polymorphisms (SNPs) have also been used for forensic marker, because they may provide better profiling from degraded DNA samples than STRs and do not involve repetitive sequences that may make stutter artifacts, which complicates STR profiling interpretation [2, 3]. SNPs for forensic analyses can be used for human identification, kinship analysis, inferring biogeographic ancestry, and estimating appearance traits [4].

In general, forensic DNA profiling is used for identifying persons compared with profiles of known suspects or missing persons included in a DNA database. However, in the case of absent information on felons or missing persons in the DNA profile database, it is helpful to predict unknown persons with DNA markers [5]. The prediction of externally visible characteristics (EVCs) from DNA has been studied for forensic genetics over the last few years [5-7]. The EVCs include hair, skin and eye color, height, and facial morphology, which have high heritability [8-10].

The genome-wide association study (GWAS), a linkage analysis and candidate gene study, has been used to identify genetic variants influencing such EVCs. Variations in the *MC1R* gene have been associated with red hair [11]. The red hair prediction method, based on a combination of non-synonymous SNPs in *MC1R*, was already developed for forensic science more than 10 years ago [12], and its accuracy was 84% in the prediction of red-haired individuals. The SNPs in the *OCA2* locus and the adjacent *HERC2* gene in Europeans showed a strong correlation with blue and brown eye color [13, 14]. Other genes were also identified as contributing to hair, eye, and skin color variations, such as *IRF4, KITLG, SLC24A4, SLC45A2, TYRP1, TYP,* and *ASIP* [15].

Recent studies using genome-wide association analysis have identified genes and variations correlating with human visible phenotypes and developed a phenotype prediction program [16-19]. HIrisPlex is capable of predicting hair and eye color from DNA with 24 SNP variations [19]. Six of the

24 SNPs used for eye color prediction in the HIrisPlex program are rs12913832 (*HERC2*), rs1800407 (*OCA2*), rs12896399 (*SLC24A4*), rs16891982 (*SLC45A2* [*MATP*]), rs1393350 (*TYR*), and rs12203592 (*IRF4*), and 4 of these SNPs within *HERC2, OCA2, SLC45A2,* and *IRF4* are also used for hair color prediction. To predict hair color and hair color shade, HIrisPlex uses 22 SNPs as follows: N29insA, rs11547464, rs885479, rs1805008, rs1805005, rs1805006, rs1805007, rs1805009, Y152OCH, rs2228479, and rs1110400 from the *MC1R* gene, rs28777 (*SLC45A2* [*MATP*]), rs16891982 (*SLC45A2* [*MATP*]), rs12821256 (*KITLG*), rs4959270 (*EXOC2*), rs12203592 (*IRF4*), rs1042602 (*TYR*), rs1800407 (*OCA2*), rs2402130 (*SLC24A4*), rs12913832 (*HERC2*), rs2378249 (*PIGU/ASIP*), and rs683 (*TYRP1*). The results of these 24 SNPs when their minor alleles are input into HIrisPlex are used to predict the color of an individual with the highest probability of the 3 eye color categories (brown, blue, or intermediate) and 4 hair color categories (blond, brown, red, and black). On average, their prediction accuracy was 69.5% for blonde hair, 78.5% for brown, 80% for red, 87.5% for black hair, and over 90% for blue and brown eyes.

Few GWASs on color-related phenotype have been carried out in Asian populations [20, 21]. Therefore, most prediction models for EVCs have been constructed and validated based on genotype and phenotype information on Europeans; these prediction models must be validated in diverse ethnic populations [16-19]. In this study, we selected potentially useful variations for forensic science that are associated with hair and eye color, iris pattern, and facial morphology, based on previous studies, and analyzed their frequencies in the Korean population.

Methods

Population and DNA extraction

The subjects in this study were collected from the Korea Association Resource (KARE) study, which has been described in detail previously [22]. Briefly, KARE consists of 2 community-based cohorts, Ansan (urban community) and Ansung (rural community) in Korea, and includes 10,038 participants aged 40 to 69 years. Genomic DNA of 1,920 male subjects randomly selected from the KARE study was extracted from cell lines immortalized with Epstein Bar virus.

SNP selection

Using the NHGRI GWAS catalog (http://www.genome. gov/gwastudies) [23], 9 articles were identified with the query terms "hair," "eye," "iris," and "facial morphology," and 26 SNPs with association p-values $< 5 \times 10^{-8}$ (GWAS

p-value) were selected from the articles [10, 24-31]. The most significant SNPs were selected from the SNPs that were under pairwise linkage disequilibrium (LD) ($r^2 > 0.8$).

Genotyping and quality control

Genotyping was performed at a multiplex level using the Illumina Golden Gate genotyping system [32]. The genotype quality score for retaining data was set to 0.1. SNPs that could not satisfy the following criteria were excluded: 1) a minimum call rate of 90%, 2) no duplication error, and 3) Hardy-Weinberg equilibrium greater than p > 0.001. All 20 SNPs were successfully genotyped.

Results and Discussion

Using the GWAS catalog, we extracted 26 candidate SNPs that might be useful for forensic science in terms of EVCs. Two SNPs in *MC1R* and *HERC2* were already genotyped in subjects of the KARE study. Among 26 SNPs, 2 pairs of SNPs, rs1847134 and rs1393350 within *TYR* and rs4900109 and rs12896399 within *SLC24A4*, respectively, were under pairwise LD ($r^2 > 0.8$), and 1 SNP in each pair was selected. Two SNPs showing low probe design ability were excluded from the probe selection process. Finally, 20 out of 26 SNPs were selected for genotyping in Koreans.

One SNP, rs7559271 (*PAX3*), was associated with facial morphology [31]; 2 SNPs, rs3739070 (*TRAF3IP1*) and rs10235789 (*SEMA3A*), were identified for the association with iris patterns, such as furrows and crypts, respectively [29]; and the remaining 17 SNPs were associated with hair and/or eye color [10, 24-28, 30]. Among these variants, 8 SNPs were included in the HIrisPlex program the determine hair and iris color.

Allele frequencies and other genetic parameters for the 20 SNPs are provided in Table 1. All analyzed SNPs were in Hardy-Weinberg equilibrium in the Korean population (p > 0.001). Three SNPs that were associated with facial morphology or iris characteristics in Europeans [29, 31] were polymorphic in the Korean population as follows: rs7559271 (*PAX3*), rs3739070 (*TRAF3IP1*), and rs10235789 (*SEMA3A*). Their minor allele frequencies (MAFs) were 0.303, 0.029, and 0.103, respectively. Among 17 SNPs related to hair or eye color, 4 SNPs were monomorphic and 6 SNPs were very rare (MAF < 0.005) in Koreans. The 4 monomorphic SNPs in Koreans were rs12203592 (*IRF4*), rs12821256 (*KITLG*), rs12913832 (*HERC2*), and rs8033165 (intergenic). Their allele frequencies in Europeans (CEU) in the HapMap database were 0.167, 0.146, 0.792, and 0.447, respectively. The 6 rare SNPs in Koreans were rs16891982 (*SLC45A2*), rs1408799 (*TYRP1*), rs35264875 (*TPCN2*), rs1393350 (*TYR*), and rs1805007 (*MC1R*), and their allele

Table 1. Allele frequencies of 20 visible phenotype SNPs

SNP ID	Reported gene	Chr:Position	Phenotype	Korean				HapMap[a]				Reference
				Allele (A/B)	MAF	HWE	No. of genotypes (AA/AB/BB)	CEU	HCB	JPT	YRI	
rs7559271	*PAX3*	2:223068286	Facial morphology	G/A	0.30	0.53	932/792/180	0.63	0.35	0.30	0.43	[31]
rs3739070	*TRAF3IP1*	2:239306268	Iris characteristics	A/C	0.03	0.78	1,795/108/2	0.10	0.09	0.02	0.03	[29]
rs10235789	*SEMA3A*	7:83656800	Iris characteristics	T/C	0.10	0.01	1,543/331/31	0.53	0.14	0.13	0.23	[29]
rs28777	*SLC45A2 (MATP)*	5:33958959	Hair	C/A	0.09	0.84	1,579/309/16	0.98	0.11	0.18	0.16	[25]
rs13289810	*TYRP1*	9:12396731	Hair	A/G	0.07	0.18	1,663/229/12	0.36	0.05	0.07	0.25	[30]
rs12896399	*SLC24A4*	14:92773663	Hair, eye	G/T	0.35	0.21	809/845/250	0.57	0.21	0.47	0.00	[10, 25, 27]
rs1667394	*OCA2*	15:28530182	Hair, eye	G/A	0.24	0.46	1,104/684/116	0.87	0.20	0.14	0.05	[10]
rs9894429	*NPLOC4, HGS*	17:79596811	Eye	T/C	0.27	0.01	985/798/120	0.61	0.26	0.30	0.27	[28]
rs1015362	*ASIP*	20:32738612	Hair	G/A	0.22	0.48	1,172/637/95	0.27	0.13	0.30	0.83	[24]
rs1003719	*TTC3, DSCR9*	21:38491095	Eye	A/G	0.39	0.21	712/883/309	0.60	0.28	0.38	0.31	[28]
rs12203592	*IRF4*	6:396321	Hair, eye	C/T	0.00	1.00	1,905/0/0	0.17	0.00	0.00	0.00	[25, 27]
rs12821256	*KITLG*	12:89328335	Hair	T/C	0.00	1.00	1,903/0/0	0.15	0.00	0.00	0.00	[10]
rs12913832	*HERC2 (OCA2)*	15:28365618	Hair, eye	A/G	0.00	1.00	1,904/0/0	0.79	0.00	0.00	0.00	[25, 27, 28]
rs8033165	Intergenic	15:29006093	Hair	C/T	0.00	1.00	1,905/0/0	0.45	0.00	0.00	0.08	[25]
rs16891982	*SLC45A2*	5:33951693	Hair, eye	C/G	0.00	0.96	1,901/4/0	0.98	0.01	0.00	0.00	[27]
rs1408799	*TYRP1*	9:12672097	Eye	T/C	0.00	0.86	1,890/15/0	0.70	0.01	0.01	0.19	[24]
rs35264875	*TPCN2*	11:68846399	Hair	A/T	0.00	0.95	1,900/5/0	0.18	0.00	0.00	0.00	[24]
rs1393350	*TYR*	11:89011046	Eye	G/A	0.00	0.99	1,904/1/0	0.23	0.00	0.00	0.00	[10, 27]
rs12931267	*MC1R*	16:89818732	Hair	C/G	0.00	0.99	1,902/1/0	NA	NA	NA	NA	[27]
rs1805007	*MC1R*	16:89986117	Hair	C/T	0.00	0.94	1,898/7/0	0.12	0.00	0.00	0.00	[10]

SNP, single nucleotide polymorphism; Chr, chromosome; position, NCBI Build 37; A, major allele; B, minor allele; MAF, minor allele frequency; HWE, Hardy-Weinberg equilibrium; CEU, European; HCB, Han Chinese; JPT, Japanese; YRI, Yoruban; NA, not available data in HapMap database.
[a]Allele frequencies of HapMap were based on minor alleles of Koreans.

frequencies in Europeans were 0.983, 0.695, 0.175, 0.226, and 0.121, respectively. The HapMap database did not include the allele frequency of rs12931267 (*MC1R*).

Since frequency information alone does not explain phenotype variability, association analysis is required to confirm the genetic effect of the SNPs between appearance phenotype and SNPs. Stokowski *et al.* [20] reported that SNPs in *TYR*, *SLC45A2*, and *SLC24A5* were associated with skin pigmentation and largely accounted for differences between those with the darkest and lightest skin in a South Asian sample. A previous study analyzed skin and hair color in Koreans, categorizing hair color into 3 types and skin color into 4 types [33]. The 7 polymorphic SNPs in the Korean population are likely to contribute to color variation within the Korean population. In addition, the 4 monomorphic SNPs could be used for distinguishing East Asians from Europeans.

In this study, we provided the allele frequency of 20 EVCs SNPs by genotyping a large number of Koreans (1,920 individuals) in comparison to the 45 Chinese and 45 Japanese recruited for the International HapMap Project. Even though the usability of these SNPs should be verified by association study in Koreans, this study might provide 10 potential SNP markers for forensic science for EVCs in the Korean population.

Acknowledgments

References

1. Jobling MA, Gill P. Encoded evidence: DNA in forensic analysis. *Nat Rev Genet* 2004;5:739-751.
2. Dixon LA, Murray CM, Archer EJ, Dobbins AE, Koumi P, Gill P. Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. *Forensic Sci Int* 2005;154:62-77.
3. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, *et al.* A multiplex assay with 52 single nucleotide

polymorphisms for human identification. *Electrophoresis* 2006; 27:1713-1724.

4. Budowle B, van Daal A. Forensically relevant SNP classes. *Biotechniques* 2008;44:603-608, 610.

5. Kayser M, de Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet* 2011;12:179-192.

6. Tully G. Genotype versus phenotype: human pigmentation. *Forensic Sci Int Genet* 2007;1:105-110.

7. Kayser M, Schneider PM. DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. *Forensic Sci Int Genet* 2009;3:154-161.

8. Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, *et al*. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res* 2003;6:399-408.

9. Clark P, Stark AE, Walsh RJ, Jardine R, Martin NG. A twin study of skin reflectance. *Ann Hum Biol* 1981;8:529-541.

10. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, *et al*. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* 2007;39:1443-1452.

11. Branicki W, Brudnik U, Kupiec T, Wolańska-Nowak P, Wojas-Pelc A. Determination of phenotype associated SNPs in the *MC1R* gene. *J Forensic Sci* 2007;52:349-354.

12. Grimes EA, Noake PJ, Dixon L, Urquhart A. Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Sci Int* 2001; 122:124-129.

13. Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, *et al*. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. *Hum Genet* 2008;123:177-187.

14. Sturm RA, Duffy DL, Zhao ZZ, Leite FP, Stark MS, Hayward NK, *et al*. 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* 2008;82:424-431.

15. Sturm RA. Molecular genetics of human pigmentation diversity. *Hum Mol Genet* 2009;18:R9-R17.

16. Spichenok O, Budimlija ZM, Mitchell AA, Jenny A, Kovacevic L, Marjanovic D, *et al*. Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Sci Int Genet* 2011;5:472-478.

17. Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci Int Genet* 2011;5:170-180.

18. Pneuman A, Budimlija ZM, Caragine T, Prinz M, Wurmbach E. Verification of eye and skin color predictors in various populations. *Leg Med (Tokyo)* 2012;14:78-83.

19. Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, *et al*. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet* 2013;7:98-115.

20. Stokowski RP, Pant PV, Dadd T, Fereday A, Hinds DA, Jarman C, *et al*. A genome-wide association study of skin pigmentation in a South Asian population. *Am J Hum Genet* 2007;81:1119-1132.

21. Edwards M, Bigham A, Tan J, Li S, Gozdzik A, Ross K, *et al*. Association of the *OCA2* polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. *PLoS Genet* 2010; 6:e1000867.

22. Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban HJ, *et al*. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet* 2009;41:527-534.

23. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009;106:9362-9367.

24. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, *et al*. Two newly identified genetic determinants of pigmentation in Europeans. *Nat Genet* 2008;40: 835-837.

25. Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, *et al*. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet* 2008;4:e1000074.

26. Kayser M, Liu F, Janssens AC, Rivadeneira F, Lao O, van Duijn K, *et al*. Three genome-wide association studies and a linkage analysis identify *HERC2* as a human iris color gene. *Am J Hum Genet* 2008;82:411-423.

27. Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, *et al*. Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet* 2010;6:e1000993.

28. Liu F, Wollstein A, Hysi PG, Ankra-Badu GA, Spector TD, Park D, *et al*. Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genet* 2010; 6:e1000934.

29. Larsson M, Duffy DL, Zhu G, Liu JZ, Macgregor S, McRae AF, *et al*. GWAS findings for human iris patterns: associations with variants in genes that influence normal neuronal pattern development. *Am J Hum Genet* 2011;89:334-343.

30. Kenny EE, Timpson NJ, Sikora M, Yee MC, Moreno-Estrada A, Eng C, *et al*. Melanesian blond hair is caused by an amino acid change in TYRP1. *Science* 2012;336:554.

31. Paternoster L, Zhurov AI, Toma AM, Kemp JP, St Pourcain B, Timpson NJ, *et al*. Genome-wide association study of three-dimensional facial morphology identifies a variant in *PAX3* associated with nasion position. *Am J Hum Genet* 2012;90:478-485.

32. Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* 2002;Suppl:56-58, 60-61.

33. Kim JH, Jin HY. The study of chromaticity analysis for body color Korea women in 20's. *J Korean Soc Cosmetol* 2011;17: 40-48.