Forensic Science International: Genetics 33 (2018) 38–47

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Genetic determinants of freckle occurrence in the Spanish population: Towards ephelides prediction from human DNA samples

Barbara Hernando[a], Maria Victoria Ibañez[b], Julio Alberto Deserio-Cuesta[a], Raquel Soria-Navarro[a], Inca Vilar-Sastre[a], Conrado Martinez-Cadenas[a,*]

[a] Department of Medicine, Jaume I University of Castellon, Ave. Sos Baynat, 12071, Castellon, Spain
[b] Department of Mathematics, Jaume I University of Castellon, Ave. Sos Baynat, 12071, Castellon, Spain

ARTICLE INFO

Keywords:
Freckles
Ephelides
Externally visible traits
DNA-based prediction
Forensic science

ABSTRACT

Prediction of human pigmentation traits, one of the most differentiable externally visible characteristics among individuals, from biological samples represents a useful tool in the field of forensic DNA phenotyping. In spite of freckling being a relatively common pigmentation characteristic in Europeans, little is known about the genetic basis of this largely genetically determined phenotype in southern European populations. In this work, we explored the predictive capacity of eight freckle and sunlight sensitivity-related genes in 458 individuals (266 non-freckled controls and 192 freckled cases) from Spain. Four loci were associated with freckling (*MC1R*, *IRF4*, *ASIP* and *BNC2*), and female sex was also found to be a predictive factor for having a freckling phenotype in our population. After identifying the most informative genetic variants responsible for human ephelides occurrence in our sample set, we developed a DNA-based freckle prediction model using a multivariate regression approach. Once developed, the capabilities of the prediction model were tested by a repeated 10-fold cross-validation approach. The proportion of correctly predicted individuals using the DNA-based freckle prediction model was 74.13%. The implementation of sex into the DNA-based freckle prediction model slightly improved the overall prediction accuracy by 2.19% (76.32%). Further evaluation of the newly-generated prediction model was performed by assessing the model's performance in a new cohort of 212 Spanish individuals, reaching a classification success rate of 74.61%. Validation of this prediction model may be carried out in larger populations, including samples from different European populations. Further research to validate and improve this newly-generated freckle prediction model will be needed before its forensic application. Together with DNA tests already validated for eye and hair colour prediction, this freckle prediction model may lead to a substantially more detailed physical description of unknown individuals from DNA found at the crime scene.

1. Introduction

Identifying predictive biomarkers of human appearance traits is being systematically investigated by the forensic community with the purpose of individual identification merely from a biological sample [1]. Human pigmentation traits are some of the most differentiable externally visible characteristics among individuals. For this reason, researchers have been focused in the design of genetic prediction tests for eye, skin and hair colour variation [2–7]. However, DNA-based prediction of other human pigmentation traits under a strong genetic control, such as ephelides occurrence, has not been generated yet.

Ephelides (also known as freckles) are small, flat, pale-brown spots commonly observed in fair-skinned and/or red-haired individuals. Ephelides typically appear early in childhood, may increase in size, number and intensity during adolescence and partly disappear during the young adulthood period [8]. Although the development of these hyperpigmented spots may be triggered by exposure to sunlight, the occurrence of ephelides is largely genetically determined [9]. The melanocortin-1 receptor (*MC1R*) gene seems to be the major contributor to the formation of freckles in European-origin individuals, independent of skin type and hair colour [10,11]. From all non-synonymous allelic variants found in the *MC1R* gene, six have been traditionally associated with a more severe phenotype, characterised by fair skin, red hair and freckling (known as the RHC phenotype): D84E, R142H, R151C, I155T, R160W and D294H [11,12]. Functional analyses have demonstrated that these *MC1R* genetic variants severely affect receptor function reducing stimulation of the pigmentation pathway. These six variants as well as other rare alleles that completely hamper MC1R function are known as 'R' alleles. Alternatively, weaker variants with lower penetrance are classified as 'r' alleles, and other rare non-

* Corresponding author.
E-mail address: ccadenas@uji.es (C. Martinez-Cadenas).

https://doi.org/10.1016/j.fsigen.2017.11.013
Received 27 June 2017; Received in revised form 13 November 2017; Accepted 22 November 2017
Available online 23 November 2017

synonymous variants that do not seem to have a noticeable effect in receptor function are defined as pseudoalleles [13].

The presence of alleles with impaired function ultimately results in an increased synthesis of phaeomelanin (instead of eumelanin) in melanocytes [13–15]. Nevertheless, it is thought that other genes must contribute to freckling, since a significant percentage of the individuals with freckles do not harbour mutations in the *MC1R* gene. Accordingly, other genes have also been associated with ephelides occurrence via genome-wide association studies, including *IRF4*, *ASIP*, *TYR* and *BNC2* [16–18].

In this work, we analysed the role of eight genes previously associated with sunlight sensitivity in an ephelides case-control study. As far as we know, this study tackles for the first time the genetic basis of freckles in a southern European population (Spain), where the freckling genotype presents a reduced frequency compared to northern Europe. After performing an association study, we developed a multivariate regression approach where only the most informative loci responsible for ephelides were included, in order to predict human ephelides occurrence. To test the power of the newly-generated freckle prediction model in future forensic applications, we evaluated the model's prediction performance in terms of accuracy, sensitivity and specificity by means of a cross-validation approach as well as an external validation using an independent sample.

2. Materials and methods

2.1. Study subjects and data collection

2.1.1. Original population

A total of 458 individuals (266 non-freckled controls and 192 freckled cases) were included in this study. Initially, unrelated participants were randomly selected, the percentage of the freckling phenotype being 21.86% in our population. These participants were recruited among the students and staff of the Jaume I University of Castellon. With the aim of performing a case-control association study, additional consecutive freckled volunteers were included in our study population. All individuals were Europeans of Spanish origin. Written informed consent was provided by all participants, and the study was approved by the Ethics Committee of the Jaume I University of Castellon (Castellon, Spain).

Under the supervision of a professional, each participant completed a standardised questionnaire to collect information on sex, age, pigmentation traits, history of childhood sunburns, Fitzpatrick's skin type classification, and sun exposure habits. Details of ephelides occurrence both during the infancy or adolescence periods and in adulthood were collected in the questionnaire (an illustration is shown in Fig. S1 in Supplementary material in online version at DOI: http://dx.doi.org/10.1016/j.fsigen.2017.11.013).

2.1.2. Independent validation population

A second phase of the study was composed of 212 unrelated individuals (109 non-freckled controls and 103 freckled cases) of Spanish origin, also recruited among the students and staff of the Jaume I University of Castellon. All individuals gave a written informed consent and completed the standardised questionnaire used to collect phenotypic information.

2.2. DNA extraction

Genomic DNA was obtained from buccal swabs that were stored at −20 °C until sample processing. DNA was isolated using QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol. After DNA extraction, all samples were diluted to a concentration of 5 ng/μl in order to prepare them for PCR amplification.

2.3. SNP selection and genotyping

Previous literature was used to select our candidate gene list. We opted for genes previously associated with sensitivity to sunlight and/or freckling [8,11,16–20]. Eventually, eight SNPs located in seven pigmentation-related genes were selected and genotyped: rs4911442 located in the *ASIP* gene [8,16,20], rs2153271 in the *BNC2* gene [16], rs12896399 in *SLC24A4* [18], rs16891982 in *SLC45A2* [19], rs1393350 and rs1042602 in the *TYR* gene [18], rs12203592 in *IRF4* [18–20], and rs12821256 in the *KITLG* gene [18]. Other candidate SNPs mentioned in the cited literature were excluded due to SNP redundancy or to extremely low frequencies in the Spanish population. SNP codes, locations, ancestral and derived alleles and their frequencies were obtained from the Ensembl Variation database (http://www.ensembl.org/info/genome/variation/index.html).

Genotyping assays were performed by using KASP Genotyping Chemistry (LGC, Hoddesdon, United Kingdom). For *SLC45A2* rs16891982, TaqMan technology was used (Applied Biosystems, Foster City, USA). Genotyping analyses were carried out in a StepOnePlusTM Real-Time PCR System, with varying PCR conditions depending on the requirements of each probe. The genotype of each sample was determined by measuring allele-specific fluorescence, using SDS v2.3 software for allelic discrimination (Applied Biosystems, Foster City, USA). For quality control, we included a negative control and a trio of samples with known genotype (major allele homozygous, heterozygous and rare allele homozygous) in each 96-well plate.

2.4. Sequencing of MC1R coding region

All DNA samples were analysed for the coding sequence of the *MC1R* gene by direct automated DNA sequencing, as previously described [21]. The primers used to amplify the *MC1R* coding region were: MC1R-F (5′-CAGCACCATGAACTAAGCAGGACACCTG-3′) and MC1R-R (5′-AAGGGTCCGCGCTTCAACACTTTCAGAG-3′). Amplification was carried out by using Type-it™ Microsatellite PCR Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. PCR products were purified with EnzSAP (EdgeBio, Gaithersburg, USA) and subsequently sequenced by direct gene sequencing with the Sanger method. A sample with known *MC1R* genotype per 96-well plate was added for quality control. All sequencing results were analysed using SeqScape v2.5 software to align and detect all nucleotide changes. All detected sequence changes were confirmed manually.

Non-synonymous *MC1R* mutations were then defined as 'R', 'r' or 'p' (pseudoallele) alleles according to their impact on protein function (Table 1). R alleles included genetic variants that have been associated with the red hair colour (RHC) phenotype [22,23], and have been shown to cause a significant impairment of receptor function in previous functional *in vitro* or *in silico* analysis [21,24–29]. Genetic variants that have not been associated with the RHC phenotype [23,30,31], and have been shown to display partial loss of function or present a possibly damaging effect based on prediction analysis [25,27,32,33], were defined as r alleles. Variants in which receptor function is similar to the wild-type form were catalogued as p alleles. Individuals were classified based on the number of R and/or r alleles carried.

2.5. Association analysis

Association analyses were performed using the R software (http://www.R-project.org). All analyses performed were two-sided, and a significance level of 0.05 was considered for rejection of the null-hypothesis. The conservative Bonferroni correction was used to adjust the significance level for multiple testing (P-value $< 4.54 \times 10^{-3} = 0.05/11$). Unknown and missing values were excluded at each specific analysis.

For each polymorphism studied, Fisher's exact test was used to check for deviations from Hardy-Weinberg equilibrium (HWE) among