Forensic Science International: Genetics 11 (2014) 64–72

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

The common occurrence of epistasis in the determination of human pigmentation and its impact on DNA-based pigmentation phenotype prediction

Ewelina Pośpiech [a,*], Anna Wojas-Pelc [b], Susan Walsh [c,g], Fan Liu [c], Hitoshi Maeda [d], Takaki Ishikawa [d,e], Małgorzata Skowron [b], Manfred Kayser [c], Wojciech Branicki [a,f]

[a] Department of Genetics and Evolution, Institute of Zoology, Jagiellonian University, Kraków, Poland
[b] Department of Dermatology, Collegium Medicum of the Jagiellonian University, Kraków, Poland
[c] Department of Forensic Molecular Biology, Erasmus MC University Medical Centre Rotterdam, Rotterdam, The Netherlands
[d] Department of Legal Medicine, Osaka City University Medical School, Osaka, Japan
[e] Division of Legal Medicine, Faculty of Medicine, Tottori University, 86 Nichicho, Yonago, Japan
[f] Section of Forensic Genetics, Institute of Forensic Research, Kraków, Poland
[g] Yale Molecular Anthropology Laboratory, Department of Anthropology, Yale University, 10 Sachem Street, New Haven, CT 06511, USA

ARTICLE INFO

Article history:
Received 29 November 2013
Received in revised form 30 January 2014
Accepted 31 January 2014

Keywords:
Pigmentation phenotype
Gene-gene interactions
Multifactor dimensionality reduction
Logistic regression
Forensic DNA phenotyping

ABSTRACT

The role of epistatic effects in the determination of complex traits is often underlined but its significance in the prediction of pigmentation phenotypes has not been evaluated so far. The prediction of pigmentation from genetic data can be useful in forensic science to describe the physical appearance of an unknown offender, victim, or missing person who cannot be identified via conventional DNA profiling. Available forensic DNA prediction systems enable the reliable prediction of several eye and hair colour categories. However, there is still space for improvement. Here we verified the association of 38 candidate DNA polymorphisms from 13 genes and explored the extent to which interactions between them may be involved in human pigmentation and their impact on forensic DNA prediction in particular. The model-building set included 718 Polish samples and the model-verification set included 307 independent Polish samples and additional 72 samples from Japan. In total, 29 significant SNP–SNP interactions were found with 5 of them showing an effect on phenotype prediction. For predicting green eye colour, interactions between *HERC2* rs12913832 and *OCA2* rs1800407 as well as *TYRP1* rs1408799 raised the prediction accuracy expressed by AUC from 0.667 to 0.697 and increased the prediction sensitivity by >3%. Interaction between *MC1R* 'R' variants and *VDR* rs731236 increased the sensitivity for light skin by >1% and by almost 3% for dark skin colour prediction. Interactions between *VDR* rs1544410 and *TYR* rs1042602 as well as between *MC1R* 'R' variants and *HERC2* rs12913832 provided an increase in red/non-red hair prediction accuracy from an AUC of 0.902–0.930. Our results thus underline epistasis as a common phenomenon in human pigmentation genetics and demonstrate that considering SNP–SNP interactions in forensic DNA phenotyping has little impact on eye, hair and skin colour prediction.

1. Introduction

The phenomenon of epistasis, also known as gene–gene interactions, is a subject of great interest especially in the field of quantitative genetics. It has been underlined that genetic interactions may play an important role in determining susceptibility to common human diseases and suggested that epistasis can be even more significant than the main independent effect of any susceptibility gene [1,2]. Epistasis was initially described by William Bateson in 1909 as the masking effect of an allele in one gene on an allele in another gene, which prevents the manifestation of its phenotypic effect [3]. The definition proposed soon after by Ronald Fisher [4] is closer to the present understanding of epistasis, which is described as a deviation from additivity in the mathematical model describing the relationship between genetic variants at different loci and the final phenotypic effect [5,6]. Statistical epistasis is however difficult to detect because of the multidimensionality of the data, which may lead to very few data points and in consequence to the occurrence of false positive results [1,7]. Therefore, next to the traditional parametric statistical methods, alternative approaches have been proposed

* Corresponding author. Tel.: +48 12 664 5087.
E-mail addresses: ewelina.pospiech@uj.edu.pl, ewelinaa.pospiech@gmail.com (E. Pośpiech).

1872-4973/$ – see front matter © 2014 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsigen.2014.01.012

belonging to the group of machine learning methods. One of them is the Multifactor Dimensionality Reduction (MDR), which was developed as a non-parametric and model-free approach to increase the power of detection of epistatic effects in studies dealing with relatively small sample numbers [8–11].

It has been recognized by the forensic community that knowledge on pigmentation genetics finds a practical application in forensic science for the DNA based prediction of externally visible characteristics (EVCs). The idea, also known as forensic DNA phenotyping (FDP) assumes description of the physical appearance of an unknown felon or decomposed body based on genotypic data of highly predictive DNA polymorphisms, usually single nucleotide polymorphisms (SNPs) [12,13]. The identification of the important role of the *HERC2–OCA2* gene complex for eye colour determination [14–18], as well as the *MC1R* gene for hair colour determination [19–23], were found to be particularly significant for prediction [17,23]. Information about the variation in several pigmentation genes allowed the development of the first forensic prediction models and assays [17,23–30]. The HIrisPlex system is such a proposed DNA prediction tool and consists of a genetic assay targeting 24 DNA markers in conjunction with multinomial logistic regression models for the simultaneous prediction of eye and hair colour from DNA [29]. This system provides a useful tool for accurate prediction of blue and brown eye colour [17,24,31,32] as well as red hair colour [23], while the achieved accuracies for non-blue and non-brown eye colour as well as non-red hair colours are somewhat lower. Another tool is the 8-plex system that is used for the prediction of eye and skin colour based on the genetic analysis of 8 SNPs from 8 genes using a decision tree method for genotype interpretation [25,26,30]. SNP-based eye colour prediction models using the classification tree method have also been proposed recently in New Zealand [28]. Snipper, an online Bayesian classifier, which was originally developed for genetic ancestry analysis has recently been modified and used to assign eye colour likelihoods [27]. All the conducted studies indicated that prediction accuracy of some phenotype categories is still unsatisfactory, especially in cases of the intermediate/green eye colours [17,24,25,31–35].

One potential solution to increase the prediction accuracy of pigmentation traits is via gene–gene interactions and their consideration in the prediction modeling. Several previous studies have already reported that epistatic effects may be relevant to pigmentation phenotypes [16,18,36,37]. Our previous study confirmed the statistical significance of epistatic effects between *HERC2* and *OCA2* and between *HERC2* and *SLC24A4* and revealed a new interaction between *HERC2* and *TYRP1* in the determination of eye colour in humans [38]. Here we further investigated this issue and using two alternative methods, MDR and logistic regression, we examined all three pigmentation phenotypes: eye, hair, and skin colour at 38 DNA polymorphisms located in 13 selected genes in a population of 718 individuals from Poland. Moreover, using an additional set of 307 samples from Poland and 72 samples from Japan, we estimated the impact of the revealed interactions on DNA-based eye, hair and skin colour phenotype prediction.

2. Materials and methods

2.1. Population samples and pigmentation phenotyping

Population samples in the form of buccal swabs were collected from 1025 unrelated Poles over 18 years old (41.3% male) in the Department of Dermatology of the Jagiellonian University Hospital. All the participants gave informed consent and were interviewed and assessed for eye, hair and skin colour by a physician specializing in dermatology. Eye colour was classified as blue (52.4%), green (12.4%), hazel (21.3%) or brown (13.9%) whereas hair colour was categorized as blond (comprising blond and dark-blond, 58.8%), red (comprising red and blond-red, 6.6%), brown (23.0%) and black (11.6%). Skin type was assessed according to the Fitzpatrick scale of classification: type I (3.1%), II (45.5%), III (40.3%) and IV (11.1%). The study was approved by the Ethics Committee of the Jagiellonian University in Krakow (KBET/17/B/2005) and the Commission on Bioethics of the Regional Board of Medical Doctors in Krakow (48 KBL/OIL/2008). An additional set of 72 samples in the form of FTA papers from Japanese individuals with uniform black hair colour, brown eye colour and skin colour equivalent to the category dark (III to V in the Fitzpatrick scale for Asians) was collected in the Department of Legal Medicine of the Osaka City University Medical School with written consent and the approval of the Ethics Committee of Osaka City University Graduate School of Medicine.

2.2. DNA samples and SNP genotyping

DNA from the study sample was extracted using the NucleoSpin® Tissue extraction kit (Macherey-Nagel GmbH & Co. KG, Germany) as previously described [38]. The study involved 37 SNP positions and one insertion–deletion polymorphism located in 13 selected genes including *SLC45A2*, *IRF4*, *EXOC2, TYRP1*, *TPCN2*, *TYR*, *KITLG*, *SLC24A4*, *OCA2*, *HERC2*, *MC1R*, *ASIP* and *VDR*, which have been associated with pigmentation characteristics in many previous studies [14–23,25,38–49]. The *VDR* gene was the only non-pigmentation gene included in the study due to its hypothesized possible interaction with pigmentation genes [50]. The list and details for all the studied polymorphisms are given in Supplementary Table 1. A subset of 13 SNPs in 10 genes (rs16891982 in *SLC45A2*, rs12203592 in *IRF4*, rs1408799 and rs683 in *TYRP1*, rs35264875 and rs3829241 in *TPCN2*, rs1393350 in *TYR*, rs12821256 in *KITLG*, rs12896399 in *SLC24A4*, rs1800407 in *OCA2*, rs12913832 in *HERC2*, rs4911414 and rs1015362 in *ASIP*) and 11 polymorphisms in the *MC1R* gene were analyzed by amplification and cycle sequencing in two multiplex assays using a procedure described elsewhere [38]. Polymorphisms in the *VDR* gene were genotyped in an additional SBE assay. Nine DNA fragments of *VDR* gene were simultaneously amplified using PCR primers given in Supplementary Table 2 and the Qiagen Multiplex PCR kit (Qiagen, Hilden, Germany). The PCR reaction consisted of 2.5 μl Qiagen Multiplex PCR mixture, 0.5 μl Q solution, 0.5 μl primer premix and 1.5 μl (approximately 1–10 ng) of template DNA. The temperature profile was as follows: 95 °C/15 min, [94 °C/30 s, 58 °C/90 s, 72 °C/90 s] × 32, 72 °C/10 min. The PCR products were next purified with Exonuclease I and Alkaline Phosphatase enzymes (Fermentas, Vilnius, Lithuania) and subjected to minisequencing reactions using the SNaPshot multiplex kit (Applied Biosystems, Foster City, CA). A single minisequencing reaction consisted of 0.5 μl SNaPshot mix, 0.5 μl of extension primer (SBE) premix, 1 μl of purified PCR product and 3 μl of DNase free water. Details of SBE primer sequences and their final concentrations are shown in Supplementary Table 3. The temperature profile was as follows: [96 °C/10 s, 50 °C/5 s, 60 °C/30 s] × 26. The products of SBE reactions were purified with SAP enzyme and finally analyzed on an ABI 3100 Avant Genetic Analyser (Applied Biosystems, Foster City, CA). The remaining 5 SNPs within 5 genes (rs28777 in *SLC45A2*, rs4959270 in *EXOC2*, rs1042602 in *TYR*, rs2402130 in *SLC24A4*, rs2378249 in *ASIP*) were genotyped in the Department of Forensic Biology of the Erasmus Medical Centre in Rotterdam using the protocol described in Ref. [23].

2.3. Population analyses

The obtained genetic data were tested for agreement with Hardy–Weinberg equilibrium using Arlequin version 3.1 software