Int J Legal Med (2015) 129:963–975
DOI 10.1007/s00414-015-1226-y

CrossMark

ORIGINAL ARTICLE

Exploration of SNP variants affecting hair colour prediction in Europeans

Jens Söchtig[1] · Chris Phillips[1] · Olalla Maroñas[1] · Antonio Gómez-Tato[2] · Raquel Cruz[3] · Jose Alvarez-Dios[2] · María-Ángeles Casares de Cal[2] · Yarimar Ruiz[1,6] · Kristian Reich[4,5] · Manuel Fondevila[1] · Ángel Carracedo[1,3,7] · María V. Lareu[1]

Received: 9 July 2014 / Accepted: 23 June 2015 / Published online: 11 July 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract DNA profiling is a key tool for forensic analysis; however, current methods identify a suspect either by direct comparison or from DNA database searches. In cases with unidentified suspects, prediction of visible physical traits e.g. pigmentation or hair distribution of the DNA donors can provide important probative information. This study aimed to explore single nucleotide polymorphism (SNP) variants for their effect on hair colour prediction. A discovery panel of 63 SNPs consisting of already established hair colour markers from the HIrisPlex hair colour phenotyping assay as well as additional markers for which associations to human pigmentation traits were previously identified was used to develop multiplex assays based on SNaPshot single-base extension technology. A genotyping study was performed on a range of European populations (n=605). Hair colour phenotyping was accomplished by matching donor's hair to a graded colour category system of reference shades and photography. Since multiple SNPs in combination contribute in varying degrees to hair colour predictability in Europeans, we aimed to compile a compact marker set that could provide a reliable hair colour inference from the fewest SNPs. The predictive approach developed uses a naïve Bayes classifier to provide hair colour assignment probabilities for the SNP profiles of the key SNPs and was embedded into the Snipper online SNP classifier (http://mathgene.usc.es/snipper/). Results indicate that red, blond, brown and black hair colours are predictable with informative probabilities in a high proportion of cases. Our study resulted in the identification of 12 most strongly associated SNPs to hair pigmentation variation in six genes.

Keywords Hair colour prediction · Pigmentation · Naïve Bayes classification · DAPC · ROC · SNP · Forensic DNA phenotyping (FDP) · Externally visible characteristics (EVCs)

Electronic supplementary material The online version of this article (doi:10.1007/s00414-015-1226-y) contains supplementary material, which is available to authorized users.

✉ Chris Phillips
c.phillips@mac.com

1 Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, A Coruña, Spain

2 Faculty of Mathematics, University of Santiago de Compostela, A Coruña, Spain

3 CIBERER, Genomic Medicine Group, University of Santiago de Compostela, A Coruña, Spain

4 Dermatologikum Hamburg, Hamburg, Germany

5 Clinic for Dermatology, Allergology and Venerology, University Medical Center Göttingen, Göttingen, Germany

6 Criminalistics Unit Against Violation of Fundamental Rights (UCCVDFAMC), Public Ministry, Caracas, Venezuela

7 Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, Saudi Arabia

Introduction

Forensic genetics now encompass the field of forensic DNA phenotyping (FDP) that aims to reconstruct externally visible characteristics (EVCs) from DNA obtained from the crime scene. FDP differs from traditional DNA typing in many regards and moves away from identification towards guidance of criminal investigations that failed to identify a crime scene sample when no suspect matched the DNA profile or was found in DNA database searches [1–5].

Pigmentation traits are amongst the most variable and conspicuous human phenotypes, making them particularly informative characteristics for the initial introduction of FDP to forensic analysis. Diversity in the colour of skin, hair and eyes

is mainly determined by production of melanin, represented by the two distinct forms of eumelanin (brown to black) and pheomelanin (yellow to reddish-brown) [6]. Variation in hair pigmentation is predominantly confined to Europeans, reaching its maximum phenotypic range in an area centred on the Eastern Baltic extending over the North and East of Europe. Outside Europe, hair colour is black with some notable exceptions such as Near East [7] and Melanesian populations [8]. Within Europe, human hair colour ranges from the darkest black to the lightest white-blond hues with numerous variations generally summarised into broader colour categories: black, brown, blond and red, while sub-categories "brunette", "chestnut brown", "auburn" and "strawberry red" are also recognised. This diversity can be explained by positive selection favouring light or fair pigmentation traits that has been in effect in European populations for over 5000 years [9].

Studies of the genetic basis of human pigmentation diversity indicate a high heritability and few genes playing a key role in determining hair colour [10–13]. Individual-specific differences are largely due to single-nucleotide polymorphisms (SNPs), with the largest proportion showing strong signals of association to human pigmentation variation from genome-wide association studies (GWAS) [14–17]. The first human polymorphisms recognised to have hair pigmentation association were in melanocortin 1 receptor (MC1R) [18], and subsequently, melanocortin receptor protein was identified as a key regulator of melanogenesis [19]. Many of the numerous MC1R variants have strong association with the red hair colour (RHC) phenotype [20]. In contrast to red hair, blond hair has been less extensively studied, but SNPs in TYRP1, TPCN2, KITLG and ASIP differentiate blond from other hair colours [16, 21, 22]. Black hair is commonly recognised as the ancestral phenotype, with ASIP [23] and SLC45A2 [24, 25] SNP associations in Europeans.

Forensic hair colour predictive tests aiming to exploit the above SNP associations began with a test for 12 MC1R variants to predict red hair [26]. Several years later, these were extended by Branicki et al. to analyse the full range of European hair colours [21]. The model developed by Branicki includes 11 SNPs in ten genes, together with two MC1R classes of low-penetrance MC1R-r variants with recessive effect and high-penetrance MC1R-R variants with dominant effect. Branicki's study assessed the scope for hair colour prediction with a multinomial logistic regression model and achieved area under the receiver operator characteristic curve (AUC) values >0.9 for red hair, almost 0.9 for black and ≈0.8 for blond and brown hair. These studies formed the basis for development of the HIrisPlex system [27], a single multiplex of 24 eye and hair colour predictive markers, including 13 of MC1R and incorporating four from IrisPlex [28]. HIrisPlex was recently enhanced with a much larger reference database and an online predictive tool [29] that now provides predictive accuracies of 69.5 % for blond hair, 78.5 % for brown, 80 % for red and 87.5 % for black.

Although HIrisPlex already provides an informative and robust test system for predicting common pigmentation variation, we wished to explore a range of additional SNP variants together with new compilations of established SNP sets identified by previous studies, for their predictive value for hair colour variation in European populations. We previously developed a forensic eye colour predictive test comprising 13 SNPs from analysis of 37 pigmentation-associated SNPs [30]. The 37 SNPs formed two discovery panels termed SHEP 1 and SHEP 2 (skin, hair and eye pigmentation), and in this study, we added recently identified hair colour-associated markers into a new SHEP 4 assay to assess 63 SNPs in total. The SHEP assays were used to genotype 605 subjects from 17 European countries. Statistical analysis of genotype data gauged the predictive power of the analysed SNPs, including amongst others, logistic regression (LR) analysis and discriminant analysis of principal components (DAPC).

As with previous assessments of eye colour predictability from small-scale SNP tests, we centred the main pigmentation phenotype prediction strategy on a naïve Bayes system by uploading profiles to Snipper [http://mathgene.usc.es/snipper/index.php]. Snipper already contains links for the prediction of eye and skin colour, and to extend this functionality, it has been updated with links for hair colour. The predictive value of a reduced set of the most closely associated hair SNPs was assessed with these same tools. Assessing the predictive performance of systematically reduced SNP sets for four hair colours provided a final set of 12 markers most strongly associated with hair colour and the best forensic classification framework for our data. The accuracy of the final set was assessed through receiver operating characteristic (ROC) analysis and calculation of the associated AUC. The main aim of this study was to examine in detail the individual contribution of SNPs known to be associated with hair colour and therefore informative for inference of hair colour in forensic analyses. For this reason, we compiled a compact SNP set that maintains reasonable predictive performance but with differences in SNP components to those of HIrisPlex. It is important to emphasise that the main study objective was not to replace HIrisPlex but to contribute to a better understanding of the SNP variants that underlie hair colour phenotypes by comparing the key predictors of both sets.

Materials and methods

Population samples

Samples comprised 605 unrelated Europeans (63.8 % females and 36.2 % males) from 17 populations (Supplementary

Fig. S2). Donors were from Spain (284), Germany (228), Sweden (24), Austria (18), Italy (13), Denmark (10), Norway (6), England (6), Finland (4), Portugal (3), Netherlands (2), Poland (2), Bosnia (1), Slovakia (1), Luxembourg (1), Greece (1) and Switzerland (1). All participants gave informed consent, and ethical approval was granted from the clinical investigation ethics committee, Galicia, Spain (CEIC: 2009/246).

Data was collected for participant's grandparental ancestry, with individuals using hair colouring or with grey hair excluded from sampling altogether (not in the 605 collected). Less frequent hair colour phenotypes like "white-blond" and "carrot-red" were not intentionally enriched and reflect corresponding frequencies in sampled populations in Europe [7]. To minimise hair tone variation due to the bleaching effects of sun and saltwater exposure, samples were taken between autumn and winter. In addition to the University of Santiago de Compostela population set, we tested the performance of the predictive approach developed in this work on an independently collected subset of individuals from Göttingen, Germany (n=63). Phenotypes for hair, eye and skin pigmentation were recorded by a dermatologist.

Hair colour phenotyping

The phenotyping regime matched donor's hair to the Fischer-Saller graded colour category system of 30 natural reference shades (Supplementary Fig. S3, GPM Anthropological Instruments, Switzerland). The Fischer-Saller scale is a widely used anthropological system for hair colour assessment [31] and uses letters from A (white-blond) through to Y (black), plus Roman numerals I–VI for red hair shades. The letter or Roman numeral was recorded at time of sample collection by a single scientist (not a dermatologist). For subjects with long hair, the proximal part of the hair shaft, least affected by bleaching effects, was examined. Hair colour was also photographed (12-megapixel reflex Canon EOS 1000D camera). To control photographic colour quality, a colour control patch was used (Kodak, USA), and the patch's white section allowed white balance adjustment using GIMP software v.2.8.10. Hair phenotype descriptions were placed into three categorical divisions of two, four or eight hair colours. The two-category division of light and dark omitted red and dark blond to fair brown colours, similar to the light/dark shade phenotyping regime of HIrisPlex. Red hair individuals were excluded because the RHC phenotype is outside the continuous spectrum of light to dark and depends on a MC1R mutation spectrum. Since we only examined extreme tonalities, intermediate tones were also excluded. The four-category designation comprised red, blond, brown and black, corresponding to the widely used categorisation of hair colour used in Branicki's study [21] and for HIrisPlex [27]. The eight category system differentiated fair and dark blond, light and dark brown and black and placed red hair into carrot-red (orange-copper), auburn (reddish-brown) and blond-red. Hence, we applied one category more than Branicki and HIrisPlex that both used a slightly different fine colour division for intermediate tones and did not consider fair blond as a category. In addition to hair colour, we obtained iris colour by applying the phenotyping approach of Ruiz et al. [30] to apply eye colour as a covariate in the logistic regression (LR) analysis of hair colour.

Training and testing sets

Training sets, forming reference data for the predictive models applied, were established by condensing all samples collected into a subset where hair colours were more clearly differentiated. Four scientists (not dermatologists, two Spanish, two German) independently classified photographs unsupervised, into red, blond, brown and black categories. This photographic review did not refer to Fischer-Saller hair colour information, and samples not classified identically by all reviewers were removed. A 230-sample training set of four hair colours was established from 65 blond, 20 red, 90 brown and 55 black individuals. The remaining 375 samples formed the test set to assess predictive model performance.

SNP selection

The discovery set of 63 pigmentation-associated SNPs was genotyped in three assays, with SHEP1 and SHEP2 run as previously described [30]. However, SHEP1 was extended with skin colour-associated rs10763644 (MPP7) and hair colour-associated rs10777129 (KITLG) and rs1426654 (SLC24A5). SNP rs10777129 was identified as a red/light hair predictor by Mengel-From et al. [32]. SNP rs1426654 showed association with hair colour in general [33] and with the variance of total hair melanin [12]. SHEP2 added SNPs rs1492354 and rs12421746, also identified by Mengel-From (by L*a*b* colorimetry) as red/light hair and blond hair predictors, respectively [32]. The novel assay SHEP4 adds mainly skin and hair predictive markers from searches of the most recent literature. Primer3 [34] and AutoDimer [35] were used to design, test and optimise amplification primers, creating amplicon lengths ranging from 87 to 135 base pairs (bp). Supplementary Table S1 lists primer and locus details for SHEP4 SNPs and cites the published studies that informed their selection as hair colour predictors. Some SNPs were only reported as skin colour associated but were used to develop a skin colour predictive test [36]. The exploratory set used by Branicki analysed 45 SNPs from 12 genes previously associated with hair colour variation [21]. Of those, 10 markers were not included in the SHEP assays: rs9378805, rs2733832, 207 rs2305498, rs1011176, rs1800401, rs16950821, rs11635884, rs8039195,

Y152OCH and N29insA. Except for the last two MC1R loci, none of the 10 markers were implemented into the prediction model developed by Branicki. MC1R InDel N29insA was also tested in all carrot-red haired samples in our study, but no variant alleles were detected, so incorporation of N29insA into the SHEP assays was not pursued. We note that N29insA is amongst the most strongly associated markers with red hair but is extremely rare, providing an example of a highly predictive marker that is not common enough to merit inclusion compared to other more frequent MC1R SNP alleles with less effect in single copy.

SNP genotyping

DNA was extracted with phenol-chloroform methods and SNP genotyping accomplished as previously described [30]. In brief, SHEP assays amplify 1 μL DNA (min. 0.5 ng.) in 6.9 μL reaction volumes of the following: 1× *AmpliTaq Gold* polymerase chain reaction (PCR) buffer (AB, Applied Biosystems. Foster City, USA), 25 mM $MgCl_2$, 10 mM dNTP mix, 3.2 μg/μL BSA, 0.5 U AB *AmpliTaq Gold* polymerase and 1.5 μL of premixed PCR primers at variable concentrations. PCR cycling comprised 10 min at 95 °C, 32–35 cycles of 95 °C for 30 s, 60 °C for 50 s, 65 °C for 40 s, then an elongation at 65 °C for 6 min. Amplifications were cleaned using 1 μL *Exo-SAPit* (USB Products, Affymetrix, Santa Clara, USA) with 2.5 μL of PCR product, incubated at 37 °C for 45 min then inactivated at 85 °C for 15 min. Multiplexed minisequencing reactions used 1.25 μL of SNaPshot *ready reaction mix* plus 0.75 μL of premixed extension primers and 1.25 μL of purified PCR product. Single-base extension (SBE) cycling used 28–30 cycles of 96 °C for 10 s, 55 °C for 5 s and 60 °C for 30 s. Extension reaction products were cleaned with 1 μL of SAP (USB) at 37 °C for 80 min and heat inactivated at 85 °C for 15 min. Then, 1–3 μL of SBE products were added to 9.5 μL AB HiDi™ formamide plus 0.3 μL AB LIZ-120 size standard. CE detection used the ABI 3130xl Genetic Analyser (AB) with POP-4™ polymer and 36 cm capillary arrays (injection voltage 2.0 kV for 22 s, run time of 1000 s at 60 °C). Results were analysed with AB Genemapper ID-X software.

Statistical analyses and classification models

This section outlines the most important components of the statistical analyses made. More detailed information on the application of discriminant analysis of principal components (DAPC), linkage disequilibrium (LD)/haplotype block analysis and analysis of epistasis with multifactor dimensionality reduction (MDR) is given in Supplementary Text S1.

Logistic regression

The 63 pigmentation-related SNPs were analysed for association to hair colour using IBM PASW SPSS Statistical-18 software. Individual SNP associations in the training set were analysed by LR with the additive model, assigning four hair colour categories to the samples and comparing each colour with the others (the alternative colours, herein termed the rest).

Apart from using an approach that did not take co-variables into account, adjustment was made for rs12913832 in order to detect the additional effect of other SNPs in close physical linkage to this strongly associated HERC2 locus. SNP rs12913832 forms an integral part of IrisPlex [28] and HIrisPlex assays [29], as well as other pigmentation informative sets developed so far [12, 21, 30, 32, 37–40].

The Snipper classifier and iterative naïve Bayes analysis

The Snipper App suite version 2.0 (http://mathgene.usc.es/snipper/) was applied as the standard tool to classify hair colour. Originally developed to handle allele frequencies for SNP-based ancestry analysis [41], Snipper was recently adapted to allow prediction of EVCs. Snipper uses a naïve Bayes classification system for single or multiple SNP profiles by estimating the likelihood of membership to one of several populations (phenotypes or ancestries) defined by their allele frequencies estimated from uploaded or predefined training sets as reference data. Likelihoods are ranked, and Snipper assigns a profile to a population from the ratio of the two largest likelihoods.

Since original adoption of the rs12913832-rs1129038 SNP pair in eye colour tests, their close linkage in HERC2 is now better handled by Snipper. It is still the user's option to treat SNPs as independent or linked, but the latter choice prompts Snipper to convert each 2-SNP allele combination to nucleotide labels. Details of the allele pair re-coding are outlined in Supplementary Table S4, following simple AA→A, AG→C, GA→G and GG→T formats.

We performed several statistical analyses forming part of Snipper to evaluate the robustness of the hair colour reference training set. Cross-validation divided the sample set into subsets followed by construction of the prediction model in several subsets and evaluating the model's performance in the remaining sets. Two types of cross-validation were performed: non-verbose cross-validation with one-out reclassification and bootstrap analysis by random choice of a training set from the full set and then classification of remaining samples with 200 iterations.

Snipper additionally measures the informativeness of each marker from divergence estimates (Jensen and Shannon's divergence [42]). Finally, the predictive value of each SNP can be estimated from the genetic distance algorithm of Snipper

enabling the identification of key SNP genotypes and/or alleles.

Our principal aim was testing established SNPs identified in previous studies, e.g. those of HIrisPlex, together with additional SNPs for their effect on hair colour prediction. To identify the contribution of each SNP to classification success, we developed a new approach, termed iterative naïve Bayes (INB) analysis. Firstly, we ranked the 50 SNPs, suggested to be most associated with hair colour in the current literature, based on their classification power applied to a 230-sample training set using Snipper with one SNP at a time. The best SNP was then fixed in position, and the remaining set of 49 was re-analysed in the same way to find the next most powerful combination of two SNPs. After finding this pairing, the second SNP was fixed and the remaining SNPs were re-analysed again. The process is iterated until all SNPs are placed in ranked order of predictive power. INB was performed for each pairwise phenotype differentiation (e.g. blond vs. non-blond, etc.). One benefit of this approach is the identification of strong classifiers for one pairwise comparison that may be weak for others.

Measuring classification performance

Following the classification approach of Branicki [21], we performed an analysis of the AUC for ROC curves (area under the receiver operating characteristic curve). AUC is the integral of ROC curves that ranges from 0.5 representing total lack of predictive power to 1.0 representing perfect prediction. This technique was applied as an additional assessment to compare the informativeness of two SNP sets: the compact set of 12 markers identified as most strongly associated to hair colour by our study and 22 of the 24 of the HIrisPlex assay. AUC analyses were made on the training and testing set together (605 samples) comprising four and additionally eight hair phenotypes: carrot-red, auburn, blond-red, fair/dark blond, light/dark brown and black. Cross-validation was implemented for all AUC analysis to ensure independence. Calculations were made using the ROCR [43] package in R (ROCR v. 1.0-5, http://rocr.bioinf.mpi-sb.mpg.de/).

Classification performance of hair predictive SNPs was measured with two different testing sets, comprising samples not used in the training set. The first consisted of 375 European samples collected alongside the training set individuals with the same phenotyping regime. Prediction performance was additionally analysed with a test set of 63 Germans, comprising nine with red hair, 22 blond, 30 brown and two black. A 3:1 minimum probability threshold was applied to all classifications (i.e. a ratio for the highest and second highest likelihoods below 3 was treated as not classified) to estimate the classification success.

Results

Prediction modelling

Phenotypes collected in the European samples consisted of 159 blond, 299 brown, 112 black and 35 red hair colour phenotypes. We observed a high frequency of light hair shades (fair blond to light brown) in northern and central European subjects, decreasing towards the south as shown in Supplementary Fig. S2 and in agreement with previous findings [7].

The results of logistic regression (LR) analysis of hair colour association in 63 SNPs are detailed in Supplementary Table S2. In the model which does not consider HERC2 SNP rs12913832 as a co-variable, 24 SNPs gave strong associations with p values below 0.0008 (threshold of significant probability under multi-test correction for 63 SNPs), comprising rs1015362, rs1110400, rs1129038, rs11636232, rs12592730, rs12896399, rs12913832, rs12931267, rs1667394, rs16891982, rs1805005, rs1805007, rs1805009, rs28777, rs35264875, rs4778138, rs4778232, rs4778241, rs4904868, rs7174027, rs7495174, rs8024968, rs885479 and rs916977. By applying rs12913832 as a co-variable, the number of significantly associated SNPs was reduced to eight out of this group. Of the 63 SNPs, 13 did not have hair colour association reported in the literature but did indicate eye and/or skin pigmentation association (marked in grey italics in Supplementary Table S2). Moreover, these 13 SNPs were not found to be significantly associated in LR analysis when adjusting for rs12913832 as a covariate, supporting their lack of a direct relationship with hair colour. However, of these 13 SNPs, three eye colour-associated SNPs rs12592730 (HERC2), rs4778232 and rs8024968 (OCA2) gave significant p values in the model without a co-variable, but we did not pursue the analysis of these SNPs further. For this reason, the 13 SNPs were removed from the marker set, and the remaining 50 SNPs were examined with iterative naïve Bayes (INB) analysis.

INB analysis produced a ranked order of informativeness for each of four hair colour comparisons as shown in Supplementary Fig. S1, with the rising plotline in each graphic corresponding to the contribution to classification success of each new marker added to the existing combination. These plots indicate a subset of 12 markers that keep the maximum proportion of predictive power that can be constructed, based on discernable early (leftmost) inflection points indicated by the arrows on each plot. Although there are multiple inflection points on each hair colour plot, the first strong change in line angle provides a simple system to identify the point where all the best predictors have been assembled. For red vs. rest, classification performance reached 93 % success with six SNPs in ranked order: rs1805007, rs11547464, rs1805008, rs35264875, rs1805009 and rs7495174. Blond classifications

reached 90 % with rs1129038 and rs4778138. Brown classifications reached 61 % success with rs35264875, rs1805006 and rs11547464. Black classifications reached 86 % success with rs12913832, rs28777, rs12931267 and rs1805008. From overlapping SNPs in each category, a final set of 12 common SNPs from five genes comprised rs28777 (SLC45A2), rs35264875 (TPCN2), rs1129038, rs12913832 (HERC2), rs4778138, rs7495174 (OCA2), rs12931267 (FANCA), rs11547464, rs1805006, rs1805007, rs1805008 and rs1805009 (MC1R).

Haploview was used to discount LD between the SNP in the HERC2 gene, applying a correlation threshold of $r^2<0.8$. Haploview results for close SNP pairs on chromosome 15 in the set of 12 SNPs are shown in Supplementary Fig. S4. The r^2 value for the strongest linkage was found between both HERC2 SNPs of rs1129038 and rs12913832, $r^2=0.659$. In INB analysis, HERC2 SNP rs1129038 was the strongest marker for blond hair colour and rs12913832 was the strongest for black hair. Although both markers are in close proximity and previously reported to be in LD [44], the r^2 value for this SNP pair did not reach the correlation threshold in our data set. The 12 SNPs of MC1R were not assessed for LD and treated as independent in Snipper analyses.

Prediction performance

Predictive performance was estimated using the success ratio of the 12 SNPs for four hair colour categories, analysed by verbose cross-validation in Snipper. The 12 SNPs gave 85 % classification success for red hair, 92.3 % for blond, 76.7 % for brown and 74.6 % for black. This analysis was also conducted separately for men and women. Females gave 92.9 % for red, 87 % for blond, 85.7 % for brown and 62.5 % for black. Males gave 50 % for red, 89.5 % for blond, 58.8 % for brown and 74.2 % for black. Applying the same approach to two hair colour shade phenotypes (12 SNPs, fair and dark) gave 93.9 % for fair and 94.6% for dark hair for both sexes combined. Applying non-verbose cross-validation produced exactly the same results.

Training set data (both regimes) for the 12 SNPs are available to use in Snipper at: http://mathgene.usc.es/snipper/hairclassifier.html.

DAPC of training set profiles provides further assessment of data structure based on phenotypes and results are shown in Fig. 1. The Fig. 1 plots show genetic clustering of four (Fig. 1a, b) and eight (Fig. 1c, d) different hair colour populations applying 63 SNPs (Fig. 1b, d) and 12 SNPs identified by INB (Fig. 1a, c). A clear differentiation between the four hair colour classes is discernable applying the full set of 63 SNPs (Fig. 1b). The application of just 12 SNPs leads to a loss of separation and increased overlap, most notably between the black and brown clusters. Sub-dividing four into eight hair colours increases this overlap substantially.

The predictive performance of the 12 SNPs was further assessed using two testing sets: (i) 375 European samples recruited during the project and (ii) 63 novel samples from Germany separated from the 605 used for the analyses described so far, as they were collected by a dermatologist applying a subjective assignment of hair colour to the four classes we described. A 3:1 minimum probability threshold was applied to the Snipper classifications, i.e. a ratio for the highest and second highest likelihoods below three denoted no classification.

For the first testing set, 76 (20.3 %) did not reach the minimum threshold ratio and remained unclassified. From the remaining 299, 184 (61.5 %) were correctly classified into the four hair classes: 77.78 % red, 84.71 % blond, 45.45 % brown and 75 % black. For the second testing set, 15 (23.8 %) were unclassified, and of the remaining 48, 37 (77.1 %), were correctly classified: 87.5 % red, 83.33 % blond, 71.43 % brown and 0 % black. A detailed overview of each testing set performance is provided in Supplementary Table S3. The lack of black hair classification success in the northern German sample can be explained by the relatively low sample size for this phenotype with two black hair samples misclassified as red and the other one unclassified.

AUC estimations of the ROC curves were made for each pairwise phenotype comparison by cross-validating genotypes. ROC curve analysis results for two different sets of markers/hair phenotypes are shown in Fig. 2. Results for the best 12 SNPs considering both training and testing set individuals ($n=605$) reached 93.9 % AUC value for red hair, 85 % for blond, 84.2 % for black and 64.2 % for brown (Fig. 2a). Considering the same 12 SNPs and eight hair colour phenotypes, successful predictions were 95.6 % for red-carrot, 85.9 % for black, 85.1 % for red-blond, 83.5 % for fair blond, 78.5 % for dark brown, 78.3 % for red-auburn, 78.0 % for dark blond 78.0 % and 64.0 % for fair brown (Fig. 2b). The application of 22 SNPs from the HIrisPlex system provided AUC values of 92.9 % for red hair, 86.3 % for black, 87.5 % for blond and 64.3 % for brown (Fig. 2c). The overall predictive success for 22 HIrisPlex markers (82.7 %) was slightly higher than with our best 12 SNPs (81.9 %). Similar results were obtained using 50 SNPs suggested to be most associated with hair colour in the current literature and which have been used for our INB analysis (data not shown). The order of the SNPs shown in Fig. 2 follows a ranked divergence order obtained from Snipper.

MDR analysis was applied to the 12 key SNPs identified using training set samples. Positive entropy indicating synergy (the effect of two interacting loci being more than the sum of their individual effects) was found for the brown vs. non-brown comparison. However, negative entropy effects were found, indicating redundancy

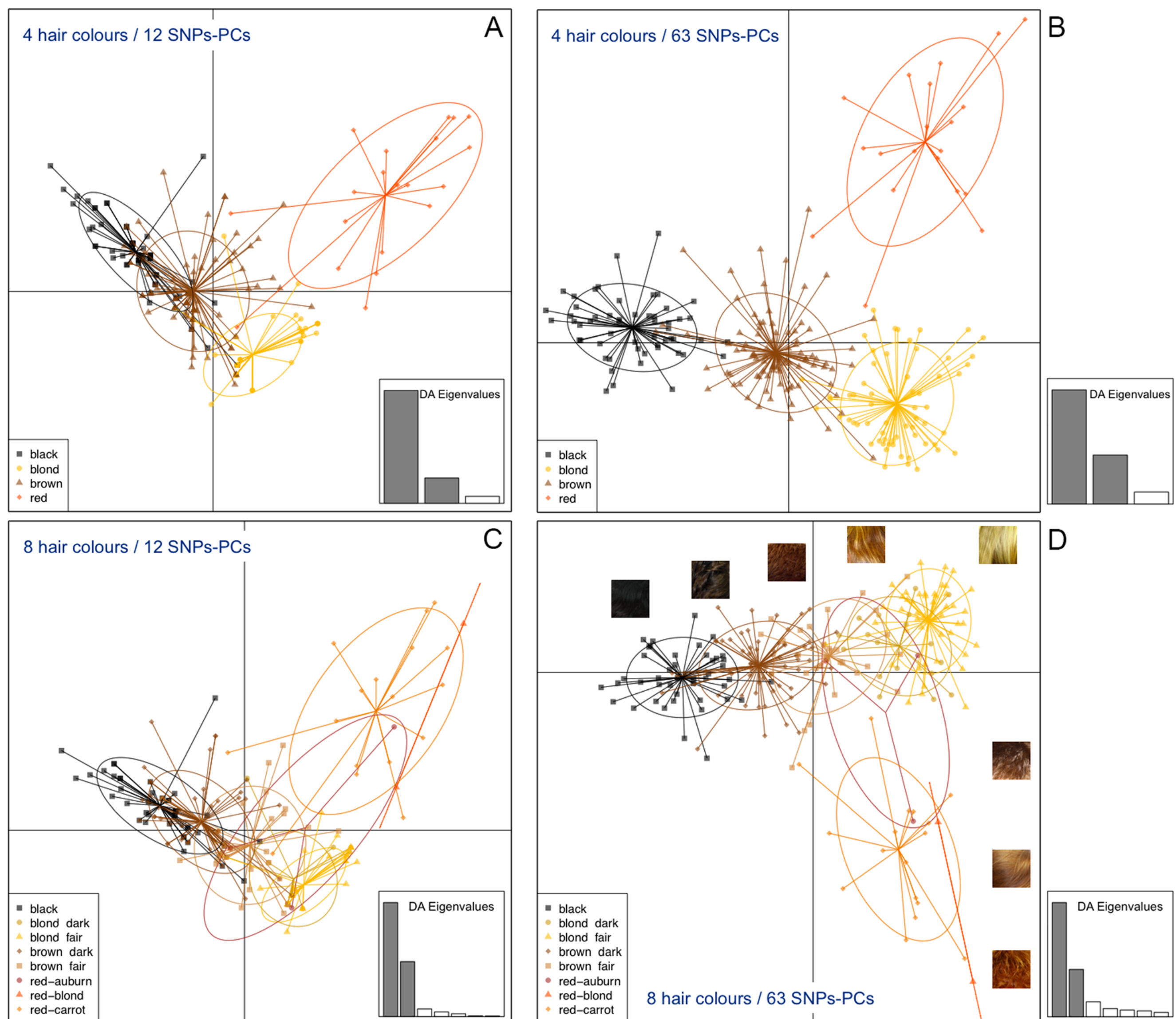

Fig. 1 DAPC of training set profiles in different grouping schemes based on the 12 SNP set (**a**, **c**) and the complete SNP panel (**b**, **d**). Clusters are shown by different colours corresponding to the hair shades and inertia ellipses, while *symbols* represent individuals. Smaller graph insets represent eigenvalues of DA

(two loci with epistatic interactions with an effect less than the sum of individual effects). The nature and the strength of the interactions for brown hair colour are shown in Supplementary Fig. S5.

Application of the Erasmus HIrisPlex DNA Phenotyping Webtool

Genotypes of 605 samples containing training and testing set individuals were introduced into the HIrisPlex DNA Phenotyping Webtool according to user guidelines provided on the web pages. Marker Y152OCH was not applied as it is not part of the SHEP assays; similarly, N29insA, which was only introduced for 18 carrot-red haired individuals, was not used in these analyses as we did not detect the rare variant allele in the carrot-red subjects.

In total, five samples were not classified by the webtool due to missing MC1R genotypes. All showed dark brown hair colour. Regarding the red haired individuals (n=35), 83 % were correctly classified by the HIrisPlex Webtool. The remaining six misclassified subjects were either classified as blond (n=4), ranging from blond/dark blond to dark blond/brown, or as brown/dark brown and dark brown/black (n=2). The group of blond hair samples (n=159) was predicted correctly in 75 % of the cases. More than half of the blond individuals in our set were reported as dark blond when applying eight categories. This was predicted by the enhanced model

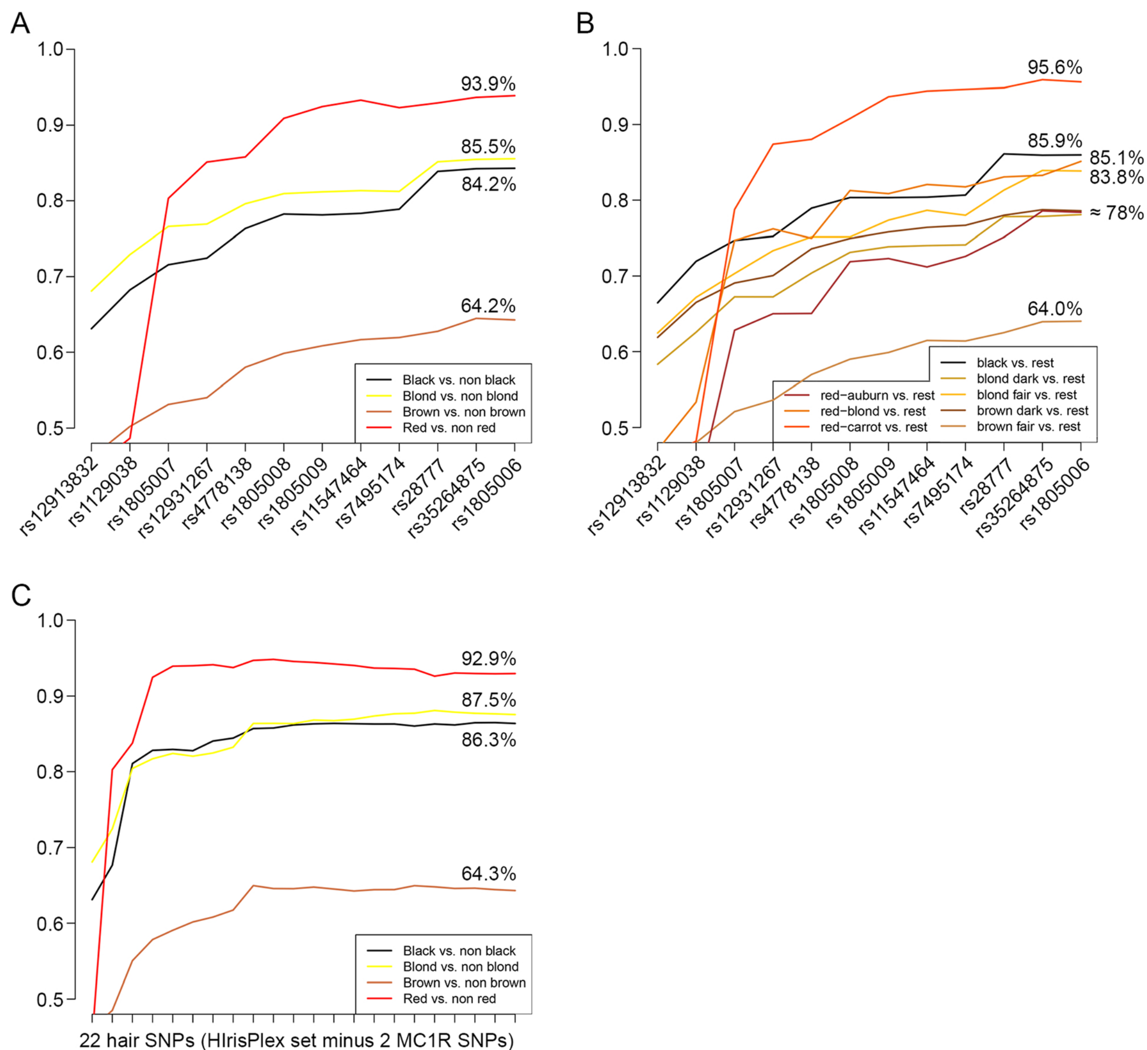

Fig. 2 Evaluation of accuracy of hair colour prediction on the whole sample set (n=605) by means of the area under the ROC curves (AUC), ranging from 0.5 (random) to 1 (perfect) prediction. All AUC calculations were made by applying cross-validation. In Fig. 2a, AUC was plotted against the 12 best SNPs for predicting: red (93.9 %), blond (85.5 %), black (84.2 %) and brown hair colour (64.2 %). Figure 2b assesses AUC predictions from the 12 best SNPs for eight hair colours. The order of the SNPs shown in Fig. 2b follows a ranked divergence order obtained from Snipper of rs12913832 (HERC2), rs1129038 (HERC2), rs1805007 (MC1R), rs12931267 (FANCA), rs4778138 (OCA2), rs1805008 (MC1R), rs1805009 (MC1R), rs11547464 (MC1R), rs7495174 (OCA2), rs28777 (SLC45A2), rs35264875 (TPCN2) and rs1805006 (MC1R). Figure 2c analyses 22 of the 24 HIrisPlex markers (MC1R markers N29insA and Y152OCH not used). For Fig. 2c, the SNP order also corresponds to divergence order obtained from Snipper

of the webtool with a success rate of 64 % as either blond/dark blond or dark blond/brown. The rest of the dark blond group was predicted simply as blond. One single blond individual from the testing set was predicted as black, 18 % as having brown hair (brown/dark brown to dark brown/black) and 7 % as having red hair. Brown hair samples from our set (n=299) were predicted correctly in 63 % of the individuals. The majority of brown hair individuals in our set had dark brown hair (n=184). Of those, 76 % were predicted as either brown/dark brown or dark brown/black by the enhanced model. The black hair colour group (n=112) was inferred as brown/dark brown in 8%, as dark brown/black in 52 % and as solely black in 29 % of the cases. Some black haired individuals were assigned as blond (n=9) or red hair colour (n=4).

In the next step, we analysed the predictive success of HIrisPlex concerning our profiles for the training and testing set separately. Regarding our training set (n=230), HIrisPlex predicted red hair colour (n=20) with 85 % precision, blond (n=65) with 88 %, brown (n=90) with 70 % and black (n=55) with 85 % (dark brown/black to black). The profiles from the testing set (n=375) yielded 80 % prediction accuracy for red (n=15), 66 % for blond (n=94), 60 % for brown (n=209) and 75 % for black (n=57) when considering dark brown/black to black predictions together.

Discussion

This study has assessed a wide range of SNP variation affecting hair colour prediction by investigating established SNPs identified in previous studies, notably HIrisPlex [29], together with additional pigmentation-associated SNPs. We found that a subset of 12 SNPs can provide a reasonably good balance between a manageable number of SNPs and reliable hair colour prediction accuracy. It is noteworthy that only seven SNPs of the 12 in our final set match the 24 of HIrisPlex, comprising the following: rs12913832, rs28777 and MC1R-R SNPs: rs11547464, rs1805006, rs1805007, rs1805008 and rs1805009. Our studies confirm rs12913832 as a strong predictor of both eye and hair colour [30]. The regulatory role of rs12913832, located in HERC2 and influencing OCA2 transcription, is well documented [45]. Amongst HIrisPlex SNPs, it has the strongest predictive effect for non-red hair colour together with rs16891982 and rs12203592 [29]. Several other studies identified rs12913832 as the strongest marker for hair colour inference [21, 46, 47] while a recent study by Branicki identified several SNP-SNP interactions for rs12913832 influencing hair colour [48]. Apart from mild synergistic interactions between rs12913832 and rs28777, we did not detect comparable interactions. The SLC45A2 non-synonymous SNP rs28777 is also common to HIrisPlex and the set of 12 predictors. SLC45A2 encodes a transporter protein that mediates melanin synthesis, and rs28777 is amongst the SNPs most strongly associated with pigmentation in humans [24], and other mammals [49]. The remaining five SNPs in common are high-penetrance MC1R-R SNPs with dominant effect: rs11547464, rs1805006, rs1805007, rs1805008 and rs1805009 [50, 21].

The remaining five SNPs of our 12 SNP set that we identified as strongly associated but not part of HIrisPlex were the following: rs7495174, rs4778138 (in OCA2), rs35264875 (TPCN2), rs1129038 (HERC2) and rs12931267 (FANCA). The SNPs rs7495174 and rs4778138, in intron 1 of OCA2, have previously been associated with eye colour [30] and hair colour in general [21, 46, 47, 51], but more specifically to blond vs. brown hair [14, 16] and black vs. blond hair [52]. Additional studies with spectrometrically measured hair colour indicated that rs7495174 was associated with dark hair colour using the CIE L*a*b* metric [53]. Similarly, rs4778138 was associated with dark hair colour applying a difference in the A650t index (reflectance at 650 nm, transformed for normality) for dark vs. light hair [54]. However, the only variant from OCA2 in HIrisPlex is rs1800407, but this SNP did not emerge as a key marker in the INB analysis applied here and was therefore excluded from the set of 12 SNPs.

SNP rs1129038 in exon 93 of HERC2 is already known as an eye colour predictor [30]. However, most eye colour predictive models discard this marker due to its LD with rs12913832 [28, 47]. Its association to spectrometrically measured [54] and observer-reported hair colour has also been published [53]. In our study, rs1129038 was the strongest predictor of blond hair whereas SNP rs12913832 was the strongest for black hair from INB analysis, so both of these HERC2 loci were incorporated into the 12 SNP predictive set, in the same way as previously described for eye colour [30]. Table 1 shows the predictive success for four hair colours obtained with Snipper applying 12 SNPs with and without the rs1129038-12913832 haplotype. The values in Table 1 indicate the effect of accommodating linkage by using SNP allele combinations is negligible. Applying two hair colours, no effect could be detected at all.

Of the SNPs not present in HIrisPlex, rs35264875 and rs12931267 were strongly associated markers in our study. The non-synonymous SNP rs35264875 in TPCN2 was amongst the key markers for brown hair colour prediction from INB analysis. This SNP was found to be significantly associated with blond rather than brown hair in GWAS in northern European population (Iceland and Netherlands) [16]. The rs35264875-T allele occurred at a frequency of 22 %. Branicki reported a similar frequency in a Polish cohort of 385 but did not detect a significant association to any hair colour [21]. However, more recently, the same group found significant association to black hair colour from multivariate binary logistic regression analysis [48]. Moreover, Branicki found synergistic epistatic interactions between rs35264875 and rs1042602 in TYR. SNP rs1042602 is part of HIrisPlex although our study indicated only moderate association from LR analysis but with no discernable contribution to predictive success from INB analysis, so this SNP was excluded. SNP rs12931267 was previously found to be highly associated with freckles and hair colour in a GWAS of Eriksson et al. using self-reported phenotype data from thousands of 23andMe genetic test consumers [46]. Eriksson's results matched a strong association to black hair found in our INB analysis and red hair from logistic regression analysis. This SNP is in the Fanconi anaemia, complementation group A (FANCA) gene near MC1R. The Fanca protein is involved in a cell process known as the Fanconi anaemia pathway, activated when DNA replication is blocked due to damage. However, the link of Fanca with pigmentation variation has not been described to date.

The SNP rs12203592 in IRF4 was found to be associated with eye colour by Liu et al. [44] and was incorporated in IrisPlex [28], although the study of Ruiz et al. did not find strong association to eye colour [30]. This marker also showed association with hair colour and is therefore informative for both eye and hair colour in HIrisPlex. We found only moderate association of rs12203592 with blond and black hair in LR analysis, but it showed minor predictive power in INB analysis and was excluded.

Table 1 Predictive performance (apparent success) for four hair colour categories from Snipper analysis using rs12913832-rs112903 as independent markers (left) or treating them as a paired haplotype (right)

12 independent SNPs	Prediction				11 SNPs (rs1129038-12913832 haplotype)	Prediction			
	Brown (%)	Blond (%)	Red (%)	Black (%)		Brown (%)	Blond (%)	Red (%)	Black (%)
Samples with brown hair	**68.89**	8.89	5.56	16.67	Samples with brown hair	**74.44**	8.89	4.44	12.22
Blond	6.15	**90.77**	3.08	0.00	Blond	6.15	**90.77**	3.08	0.00
Red	5.00	10.00	**85.00**	0.00	Red	5.00	10.00	**85.00**	0.00
Black	20.00	1.82	0.00	**78.18**	Black	21.82	3.64	0.00	**74.55**

Bold values indicate successful predictions of hair colour phenotype

The iterative naïve Bayes analysis we developed for hair and skin colour variation studies [36] provides a feasible tool for identifying the best EVC predictors for a trait from pairwise comparisons of classification success. Using the online tool for hair colour prediction, released alongside this paper, (http://mathgene.usc.es/snipper/hairclassifier.html), likelihood ratios provide assignment to two or four hair colour classes in a straightforward way. Snipper retains the flexibility to handle missing SNP genotypes, common when analysing challenging forensic material such as highly degraded DNA. Therefore, profiles where weakly predictive SNPs are missing are valid and likely to obtain high probabilities for distinct hair colours. However, it is important to consider that likelihood values generated are dependent on which of the 12 SNPs are failing as well as the potential negative effect on likelihoods of such incomplete data on the LRs obtained.

DAPC provided an informative description of the genetic clusters obtained with the SNPs we identified to be most associated with hair colour. Although DAPC demonstrated that hair colour prediction using eight colour categories is not feasible, it is noteworthy that the DAPC cluster arrangement corresponds to the incremental colour patterns ranging from fair (blond) via intermediate tones to dark (black) hair colour. Moreover, the red tones cluster separately from those of the other hair colours, emphasising the different genetic background of the RHC phenotype.

Expressed by AUC values, the 12 SNP set predicted red, blond and black hair colours with 88 % overall accuracy in the complete sample set (training and testing set samples) applying cross-validation (Fig. 2a). However, prediction success for brown hair colour remains well below these levels at 64 % accuracy. These results contrast with those of Branicki's study that achieved balanced AUC values for the four hair colours. While brown hair has the lowest predictive accuracy in our study, Branicki reported the highest error rate for blond hair (19 %). After recent reference database enlargements, the Erasmus HIrisPlex Phenotyping Webtool (http://www.erasmusmc.nl/fmb/resources/Irisplex_HIrisPlex/?laHng=en) gives AUC values of 81 % for blond, 75 % for brown, 92 % for red and 85 % for black hair [29]. When applying 22 of the 24 markers included in the HIrisPlex assay to our complete sample set by cross-validating genotype data, we obtain AUC values for all colours of 93 % for red, 86 % for black, 88 % for blond and 64 % for brown. In general, the overall predictive success reported by HIrisPlex is slightly higher (83.3 %) than with our sample set applying 22 of the 24 HIrisPlex SNPs (82.7 %). Moreover, the overall predictive success of 22 HIrisPlex markers is also higher than with our best 12 markers (81.9 %) when applying our whole sample set. A further extension of these assessments to 50 hair colour-associated SNPs did not result in an improvement of prediction power. This outcome clearly demonstrates that the additional markers investigated in this study do not improve the accuracy of the 12 SNP set or HIrisPlex. However, bias in outcomes may also be due to the differences in the number of study samples, as HIrisPlex is based on a much larger sample size than our study.

The examination of our profiles with the HIrisPlex Phenotyping Webtool shows a similar picture of predictive success. The prediction of red hair colour in our total sample set provides the highest success rate (83 %), followed by black (81 %), blond (75 %) and brown hair (63 %).

On the whole, the prediction systems for HIrisPlex and Snipper both show optimum predictive performance for red hair. HIrisPlex predicts black hair better than blond, while we found black hair had lower prediction success than blond. The enhanced black hair colour prediction by the HIrisPlex Webtool could be explained in part by the inclusion of Japanese samples ($n=50$) representing a non-European black hair cohort that improves black hair prediction overall [29]. However, inference of brown hair is more error-prone in both predictive systems. To some extent, this could be due to phenotyping inconsistency as brown hair is an intermediate colour covering the largest range from light brown (nearly blond) to dark brown (nearly black). Classification of a continuous variable covering such a value range by applying training set data with discontinuous definitions will diminish the observed power of statistical testing. This effect handicaps accurate colour discrimination, and the results from different reported hair colour studies are therefore very difficult to compare. The Fischer-Saller hair colour chart used for this study covers a

large range of brown tonalities (Supplementary Fig. S3, letters P to W), but these do not represent the full range of brown shades observed during this study. Furthermore, use of the Fischer-Saller chart seems to be more appropriate for phenotyping populations of northern Europe compared to the south since the fair hair colour range (fair/dark blond) is over-represented in contrast to the limited dark hair sample range (dark brown/black). A lower brown hair colour predictive power could be further attributed to a lack of full marker coverage, epigenetic effects [48] and/or environmental influences. It is well known that hair colour often changes during an individual's lifetime, and such shifts include the darkening of blond hair colour in particular [19]. This effect, as well as its implication for hair colour inference in forensic analysis, was described previously by Walsh [27] and Branicki [48].

Finally, it is important to highlight several limitations to the present study. Although we investigated an exploratory panel of 63 SNPs consisting of established hair colour markers from the HIrisPlex assay as well as additional markers for which associations to human pigmentation traits were previously identified, we did not improve the predictive accuracy of the HIrisPlex assay. It is also evident that the classification success of our 12 SNP set remains below the performance of HIrisPlex. Moreover, our assessments have less statistical power since they are based on a smaller sample set than the studies by Walsh et al. HIrisPlex can predict eye and hair colour simultaneously while the 12 SNP set presented here is restricted to hair colour alone. The rs1129038-12913832 haplotype are the only markers in common with the 13 SNP set we published for eye colour inference [30]. In conclusion, the 12 SNP set presented is not intended to be an alternative forensic test to HIrisPlex, but the additional SNPs we identified to be associated with hair colour and those in HIrisPlex that were not included in the 12 certainly merit further studies of the relative effect of different pigmentation trait predictors.

Acknowledgments The authors are very grateful to all the sample donors who participated in the project. JS would like to acknowledge the support of the German FAZIT foundation. MVL was supported by funding from Xunta de Galicia, Incite 09208163 PR.

Conflict of interest The authors declare that they have no competing interests.

References

1. Kayser M, Schneider PM (2009) DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. Forensic Sci Int Genet 3(3): 154–161
2. Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 12(3):179–192
3. Kayser M (2013) Forensic DNA phenotyping: DNA testing for externally visible characteristics. In: Siegel JA, Saukko PJ, Houck MM (eds) Encyclopedia of forensic sciences. Academic, Waltham, pp 369–374
4. Kayser M (2015) Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genet. doi:10.1016/j.fsigen.2015.02.003
5. Maronas O, Sochtig J, Ruiz Y, Phillips C, Carracedo A, Lareu MV (2015) The genetics of skins, hair, and eye color variation and its relevance to forensic pigmentation predictive tests. Forensic Sci Rev 27(1):13–40
6. Slominski A, Tobin DJ, Shibahara S, Wortsman J (2004) Melanin pigmentation in mammalian skin and its hormonal regulation. Physiol Rev 84(4):1155–1228
7. Beals RL, Hoijer H (1965) An introduction to anthropology, 3rd edn. Macmillan, New York
8. Kenny EE, Timpson NJ, Sikora M, Yee MC, Moreno-Estrada A, Eng C, Huntsman S, Burchard EG, Stoneking M, Bustamante CD, Myles S (2012) Melanesian blond hair is caused by an amino acid change in TYRP1. Science 336(6081):554
9. Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, Unterlander M, Hollfelder N, Potekhina ID, Schier W, Thomas MG, Burger J (2014) Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. Proc Natl Acad Sci U S A 111(13):4832–4837
10. Cook AL, Chen W, Thurber AE, Smit DJ, Smith AG, Bladen TG, Brown DL, Duffy DL, Pastorino L, Bianchi-Scarra G, Leonard JH, Stow JL, Sturm RA (2009) Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. J Invest Dermatol 129(2):392–405
11. Zhang M, Song F, Liang L, Nan H, Zhang J, Liu H, Wang LE, Wei Q, Lee JE, Amos CI, Kraft P, Qureshi AA, Han J (2013) Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. Hum Mol Genet 22(14):2948–2959
12. Valenzuela RK, Henderson MS, Walsh MH, Garrison NA, Kelch JT, Cohen-Barak O, Erickson DT, John Meaney F, Bruce Walsh J, Cheng KC, Ito S, Wakamatsu K, Frudakis T, Thomas M, Brilliant MH (2010) Predicting phenotype from genotype: normal pigmentation. J Forensic Sci 55(2):315–322
13. Gerstenblith MR, Shi J, Landi MT (2010) Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis. Pigment Cell Melanoma Res 23(5):587–606
14. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, Manolescu A, Karason A, Palsson A, Thorleifsson G, Jakobsdottir M, Steinberg S, Palsson S, Jonasson F, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Aben KK, Kiemeney LA, Olafsson JH, Gulcher J, Kong A, Thorsteinsdottir U, Stefansson K (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet 39(12):1443–1452
15. Stokowski RP, Pant PV, Dadd T, Fereday A, Hinds DA, Jarman C, Filsell W, Ginger RS, Green MR, van der Ouderaa FJ, Cox DR (2007) A genomewide association study of skin pigmentation in a South Asian population. Am J Hum Genet 81(6):1119–1132
16. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, Steinberg S, Gudjonsson SA, Palsson A, Thorleifsson G, Palsson S, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Aben KK, Vermeulen SH, Goldstein AM, Tucker MA, Kiemeney LA, Olafsson JH, Gulcher J, Kong A, Thorsteinsdottir U, Stefansson K (2008) Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet 40(7):835–837
17. Kayser M, Liu F, Janssens AC, Rivadeneira F, Lao O, van Duijn K, Vermeulen M, Arp P, Jhamai MM, van Ijcken WF, den Dunnen JT, Heath S, Zelenika D, Despriet DD, Klaver CC, Vingerling JR, de Jong PT, Hofman A, Aulchenko YS, Uitterlinden AG, Oostra BA,

van Duijn CM (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am J Hum Genet 82(2):411–423

18. Valverde P, Healy E, Jackson I, Rees JL, Thody AJ (1995) Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat Genet 11(3):328–330
19. Rees JL (2003) Genetics of hair and skin color. Annu Rev Genet 37: 67–90
20. Garcia-Borron JC, Abdel-Malek Z, Jimenez-Cervantes C (2014) MC1R, the cAMP pathway, and the response to solar UV: extending the horizon beyond pigmentation. Pigment Cell Melanoma Res
21. Branicki W, Liu F, van Duijn K, Draus-Barini J, Pospiech E, Walsh S, Kupiec T, Wojas-Pelc A, Kayser M (2011) Model-based prediction of human hair color using DNA variants. Hum Genet 129(4): 443–454
22. Guenther CA, Tasic B, Luo L, Bedell MA, Kingsley DM (2014) A molecular basis for classic blond hair color in Europeans. Nat Genet 46(7):748–752
23. Kanetsky PA, Swoyer J, Panossian S, Holmes R, Guerry D, Rebbeck TR (2002) A polymorphism in the agouti signaling protein gene is associated with human pigmentation. Am J Hum Genet 70(3):770–775
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(11-12):966–971
25. 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(3):278–284
26. Grimes EA, Noake PJ, Dixon L, Urquhart A (2001) Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. Forensic Sci Int 122(2-3): 124–129
27. Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, Branicki W, Kayser M (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet 7(1):98–115
28. Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M (2011) 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(3):170–180
29. Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L, Maeda H, Ishikawa T, Sijen T, de Knijff P, Branicki W, Liu F, Kayser M (2014) Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet 9:150–161
30. Ruiz Y, Phillips C, Gomez-Tato A, Alvarez-Dios J, Casares de Cal M, Cruz R, Maronas O, Sochtig J, Fondevila M, Rodriguez-Cid MJ, Carracedo A, Lareu MV (2013) Further development of forensic eye color predictive tests. Forensic Sci Int Genet 7(1):28–40
31. Matheny AP Jr, Dolan AB (1975) Sex and genetic differences in hair color changes during early childhood. Am J Phys Anthropol 42(1):53–56
32. Mengel-From J, Borsting C, Sanchez JJ, Eiberg H, Morling N (2010) Human eye colour and HERC2, OCA2 and MATP. Forensic Sci Int Genet 4(5):323–328
33. Dimisianos G, Stefanaki I, Nicolaou V, Sypsa V, Antoniou C, Poulou M, Papadopoulos O, Gogas H, Kanavakis E, Nicolaidou E, Katsambas AD, Stratigos AJ (2009) A study of a single variant allele (rs1426654) of the pigmentation-related gene SLC24A5 in Greek subjects. Exp Dermatol 18(2):175–177
34. Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23(10):1289–1291
35. Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques 37(2):226–231
36. Maronas O, Phillips C, Sochtig J, Gomez-Tato A, Cruz R, Alvarez-Dios J, de Cal MC, Ruiz Y, Fondevila M, Carracedo A, Lareu MV (2014) Development of a forensic skin colour predictive test. Forensic Sci Int Genet 13:34–44
37. Hart KL, Kimura SL, Mushailov V, Budimlija ZM, Prinz M, Wurmbach E (2013) Improved eye- and skin-color prediction based on 8 SNPs. Croat Med J 54(3):248–256
38. Keating B, Bansal AT, Walsh S, Millman J, Newman J, Kidd K, Budowle B, Eisenberg A, Donfack J, Gasparini P, Budimlija Z, Henders AK, Chandrupatla H, Duffy DL, Gordon SD, Hysi P, Liu F, Medland SE, Rubin L, Martin NG, Spector TD, Kayser M, International Visible Trait Genetics C (2013) First all-in-one diagnostic tool for DNA intelligence: genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the Identitas v1 Forensic Chip. Int J Legal Med 127(3):559–572
39. Allwood JS, Harbison S (2013) SNP model development for the prediction of eye colour in New Zealand. Forensic Sci Int Genet 7(4):444–452
40. Gettings KB, Lai R, Johnson JL, Peck MA, Hart JA, Gordish-Dressman H, Schanfield MS, Podini DS (2014) A 50-SNP assay for biogeographic ancestry and phenotype prediction in the U.S. population. Forensic Sci Int Genet 8(1):101–108
41. Phillips C, Salas A, Sanchez JJ, Fondevila M, Gomez-Tato A, Alvarez-Dios J, Calaza M, de Cal MC, Ballard D, Lareu MV, Carracedo A (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci Int Genet 1(3-4):273–280
42. Grosse I, Bernaola-Galvan P, Carpena P, Roman-Roldan R, Oliver J, Stanley HE (2002) Analysis of symbolic sequences using the Jensen-Shannon divergence. Phys Rev E Stat Nonlinear Soft Matter Phys 65(4 Pt 1):041905
43. Sing T, Sander O, Beerenwinkel N, Lengauer T (2005) ROCR: visualizing classifier performance in R. Bioinformatics 21(20): 3940–3941
44. Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens AC, Kayser M (2009) Eye color and the prediction of complex phenotypes from genotypes. Curr Biol 19(5): R192–R193
45. Visser M, Kayser M, Palstra RJ (2012) HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome Res 22(3):446–455
46. Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, Avey L, Wojcicki A, Pe'er I, Mountain J (2010) Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS Genet 6(6): e1000993
47. Kastelic V, Drobnic K (2012) A single-nucleotide polymorphism (SNP) multiplex system: the association of five SNPs with human eye and hair color in the Slovenian population and comparison using a Bayesian network and logistic regression model. Croat Med J 53(5):401–408
48. Pospiech E, Wojas-Pelc A, Walsh S, Liu F, Maeda H, Ishikawa T, Skowron M, Kayser M, Branicki W (2014) The common occurrence of epistasis in the determination of human pigmentation and its impact on DNA-based pigmentation phenotype prediction. Forensic Sci Int Genet 11:64–72
49. Xu X, Dong G-X, Hu X-S, Miao L, Zhang X-L, Zhang D-L, Yang H-D, Zhang T-Y, Zou Z-T, Zhang T-T, Zhuang Y, Bhak J, Cho Yun S, Dai W-T, Jiang T-J, Xie C, Li R, Luo S-J (2013) The genetic basis of white tigers. Curr Biol 23(11):1031–1035
50. Sturm RA, Duffy DL, Box NF, Newton RA, Shepherd AG, Chen W, Marks LH, Leonard JH, Martin NG (2003) Genetic association

and cellular function of MC1R variant alleles in human pigmentation. Ann N Y Acad Sci 994:348–358

51. Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR, Hayward NK, Martin NG, Sturm RA (2007) A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. Am J Hum Genet 80(2):241–252
52. Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Hu FB, Duffy DL, Zhao ZZ, Martin NG, Montgomery GW, Hayward NK, Thomas G, Hoover RN, Chanock S, Hunter DJ (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet 4(5):e1000074
53. Mengel-From J, Wong TH, Morling N, Rees JL, Jackson IJ (2009) Genetic determinants of hair and eye colours in the Scottish and Danish populations. BMC Genet 10:88
54. Shekar SN, Duffy DL, Frudakis T, Sturm RA, Zhao ZZ, Montgomery GW, Martin NG (2008) Linkage and association analysis of spectrophotometrically quantified hair color in Australian adolescents: the effect of OCA2 and HERC2. J Invest Dermatol 128(12):2807–2814
55. Sturm RA, Duffy DL (2012) Human pigmentation genes under environmental selection. Genome Biol 13(9):248
56. Brown KM, Macgregor S, Montgomery GW, Craig DW, Zhao ZZ, Iyadurai K, Henders AK, Homer N, Campbell MJ, Stark M, Thomas S, Schmid H, Holland EA, Gillanders EM, Duffy DL, Maskiell JA, Jetann J, Ferguson M, Stephan DA, Cust AE, Whiteman D, Green A, Olsson H, Puig S, Ghiorzo P, Hansson J, Demenais F, Goldstein AM, Gruis NA, Elder DE, Bishop JN, Kefford RF, Giles GG, Armstrong BK, Aitken JF, Hopper JL, Martin NG, Trent JM, Mann GJ, Hayward NK (2008) Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat Genet 40(7):838–840
57. Lao O, Vallone PM, Coble MD, Diegoli TM, van Oven M, van der Gaag KJ, Pijpe J, de Knijff P, Kayser M (2010) Evaluating self-declared ancestry of U.S. Americans with autosomal, Y-chromosomal and mitochondrial DNA. Hum Mutat 31 (12): E1875–1893
58. Lao O, de Gruijter JM, van Duijn K, Navarro A, Kayser M (2007) Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. Ann Hum Genet 71(Pt 3):354–369
59. Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulos G, Jin L, Hoggart CJ, McKeigue PM, Kittles RA (2003) Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 112(4):387–399
60. Edwards M, Bigham A, Tan J, Li S, Gozdzik A, Ross K, Jin L, Parra EJ (2010) Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. PLoS Genet 6(3):e1000867