Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Testing the impact of trait prevalence priors in Bayesian-based genetic prediction modeling of human appearance traits

Maria-Alexandra Katsara[a], Wojciech Branicki[b,c], Ewelina Pośpiech[b], Pirro Hysi[d], Susan Walsh[e], Manfred Kayser[f], Michael Nothnagel[a,g,*], on behalf of the VISAGE Consortium

[a] Cologne Center for Genomics, University of Cologne, Cologne, Germany
[b] Malopolska Centre of Biotechnology, Jagiellonian University, Kraków, Poland
[c] Central Forensic Laboratory of the Police, Warsaw, Poland
[d] Department of Twin Research & Genetic Epidemiology, St Thomas Hospital, Campus, Kings College London (KCL), London, UK
[e] Department of Biology, Indiana University Purdue University Indianapolis (IUPUI), Indianapolis, IN, USA
[f] Department of Genetic Identification, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands
[g] University Hospital Cologne, Cologne, Germany

ARTICLE INFO

Keywords:
Appearances
Genetic prediction
Impact of priors
Predictive DNA analysis
Forensic DNA phenotyping
Externally visible characteristics

ABSTRACT

The prediction of appearance traits by use of solely genetic information has become an established approach and a number of statistical prediction models have already been developed for this purpose. However, given limited knowledge on appearance genetics, currently available models are incomplete and do not include all causal genetic variants as predictors. Therefore such prediction models may benefit from the inclusion of additional information that acts as a proxy for this unknown genetic background. Use of priors, possibly informed by trait category prevalence values in biogeographic ancestry groups, in a Bayesian framework may thus improve the prediction accuracy of previously predicted externally visible characteristics, but has not been investigated as of yet. In this study, we assessed the impact of using trait prevalence-informed priors on the prediction performance in Bayesian models for eye, hair and skin color as well as hair structure and freckles in comparison to the respective prior-free models. Those prior-free models were either similarly defined either very close to the already established ones by using a reduced predictive marker set. However, these differences in the number of the predictive markers should not affect significantly our main outcomes. We observed that such priors often had a strong effect on the prediction performance, but to varying degrees between different traits and also different trait categories, with some categories barely showing an effect. While we found potential for improving the prediction accuracy of many of the appearance trait categories tested by using priors, our analyses also showed that misspecification of those prior values often severely diminished the accuracy compared to the respective prior-free approach. This emphasizes the importance of accurate specification of prevalence-informed priors in Bayesian prediction modeling of appearance traits. However, the existing literature knowledge on spatial prevalence is sparse for most appearance traits, including those investigated here. Due to the limitations in appearance trait prevalence knowledge, our results render the use of trait prevalence-informed priors in DNA-based appearance trait prediction currently infeasible.

1. Introduction

Prediction of externally visible characteristics (EVCs) of an individual solely based on genetic information, also referred to as DNA phenotyping or forensic DNA phenotyping (FDP), has become a focus in human genetic research and applications, such as in forensics, ancient DNA analysis and other areas. In forensic cases where conventional DNA-profiling methods, typically based on short tandem repeat (STR) markers, fail to identify the crime scene sample donor, because the evidential DNA-profile does not match the DNA-profile of any of the case suspects or anybody in the criminal offender DNA database, FDP may provide significant leads for police investigations to find unknown

[*] Corresponding author at: Cologne Center for Genomics, Department of Statistical Genetics and Bioinformatics, University of Cologne, Weyertal 115b, 50931 Cologne, Germany.
E-mail address: michael.nothnagel@uni-koeln.de (M. Nothnagel).

https://doi.org/10.1016/j.fsigen.2020.102412
Received 15 April 2020; Received in revised form 9 September 2020; Accepted 25 October 2020
Available online 4 November 2020

perpetrators [1–3]. In such cases, FDP can contribute significantly by narrowing down a potentially large number of putative sample donors to a smaller group of individuals that carry the FDP-derived EVC information on which the police can then focus with further investigation. Groups that do not carry such information can be left out from the police investigation. Thus far, for eye, hair and skin color various underlying genes have been identified, predictive DNA markers have been identified, DNA tests suitable for analyzing such genetic markers in forensic DNA samples and statistical prediction models have been developed [4–10], and some of these DNA test systems have been forensically validated [9,11,12]. For traits such as freckles and hair structure, some associated genetic markers and the first predictive models have already been published, respectively [13–16]; however, no forensically validated tool has been established so far. Prediction models for some other EVCs are currently under investigation [17–20].

Categorical prediction of eye, hair and skin color is often based on multinomial logistic regression (MLR) using established genetic marker panels. For instance, the IrisPlex test and model for eye color prediction consists of a set of 6 single-nucleotide polymorphisms (SNPs) [4,9,21]. Its extension to eye and hair color, the HIrisPlex test and model is based on 24 SNPs in total [11]. The latest extension is the HIrisPlex-S test and model, which consists of 41 SNPs and allows simultaneous prediction of eye, hair and skin color from a DNA sample [12]. All three prediction models are publicly available via https://hirisplex.erasmusmc.nl/. An alternative statistical tool for the prediction of eye, hair and skin color from genotype data is offered by Snipper [8,22,23], which uses pairwise likelihood ratios to present prediction outcomes, while other pigmentation prediction tool models were also developed (see [24] for a review). While some of these models show high prediction accuracies for some pigmentation categories, more research is currently under way in order to improve existing tools, either by including more SNP predictors after they have been identified in large-scale gene mapping studies, or by using alternative prediction methods.

Bayesian classification is a statistical approach that considers the data-independent probability of each category, or class, as well as the data-derived likelihood that a given subject or object belongs to a particular category, and bases the classification decision upon these probabilities. More specifically, the Bayesian approach combines a prior probability distribution on the different categories with the density probabilities obtained from the observed samples, yielding the posterior distribution used to predict category, or class, membership of an individual or object [25]. Prior probabilities for parameters may reflect previous evidence, but also purely subjective assessment or available information on these parameters from the past, before any evidence from the sample set at hand is considered. Incorporation of such prior knowledge in the data analysis may potentially increase the prediction accuracy, namely in situations where the prediction model does not include all causal genetic factors and where the environment contributes significantly to the trait variance via non-genetic factors. In both situations, trait prevalence-informed priors may then act as proxies for the yet unknown causal genetic factors and non-genetic factors in a population, group or region. In the framework of appearance DNA prediction, including FDP, inference of the biogeographic ancestry of an unknown DNA sample from which EVCs are to be predicted, together with the use of the trait class prevalence in such biogeographic ancestry group as prior in the EVC prediction model may improve the prediction accuracy. However, despite the already existing approaches for EVC prediction, the impact of trait prevalence priors on EVC prediction accuracies has not been investigated thus far.

For putting prior-based EVC prediction into practice within the concept of FDP, one would envision to first carry out forensic DNA ancestry testing on the unknown crime scene DNA samples and use the obtained ancestry outcome as guidance for allocating the appropriate trait class prevalence data for the EVC to be predicted, and finally use them as priors in EVC prediction. Based on the DNA-identified geographic region of ancestry of the tested DNA donor, allocated trait class prevalence data for different populations from such region would be averaged (or combined in another suitable way), in order to likely represent continental or sub-continental groups, and would then be used as priors for Bayesian EVC prediction on the same DNA sample previously used for ancestry testing. Alternatively, to avoid population averaging, DNA ancestry testing would need to be specific for a particular population, which not only requires the availability of trait prevalence data for such population but also the ability of forensic DNA ancestry to work on the population level.

Here, we assess the impact of incorporating prior knowledge on EVC trait prevalence in a Bayesian setting on improving the accuracy of DNA-based EVC prediction, but also potential pitfalls caused by misspecification of such prior probabilities. To this end, we consider EVCs such as eye, hair and skin color for which prior-free genetic prediction models have previously been established [9,11,12], but also traits such as hair structure and freckles for which the first prediction models were recently proposed without considering priors [13,15,16]. Given the sparsity or even lack of spatial or population-specific prevalence information available for each of these EVCs [24], we investigated the impact of prevalence-informed priors across a grid in the complete space of all possible values for each trait category, thereby emulating the (mis-) specification of the informative prior values. Prediction modelling was performed by applying previously proposed DNA predictors in datasets from different populations inside and outside of Europe. We report on standard prediction performance measures for each trait category separately and for all model measurements, and then compare prior-informed model-based prediction against prior-free models. Furthermore, we demonstrate the effect of priors on the overall prediction accuracy of the EVCs investigated.

2. Materials and methods

2.1. Data sets

For prediction modelling of eye color, hair color, skin color, hair structure and freckles we used various datasets, most of which were used previously for predicting these EVCs. For *eye, hair and skin color* we applied datasets that were part of the previously used data to establish the IrisPlex model for eye color, the HIrisPlex model for hair color, and the HIrisPlex-S model for skin color prediction, comprising of samples from different continental ancestries [9,11,12]. In particular, we used 1095 samples for eye, 1702 for hair and 1318 for skin color prediction (Table 1). For *hair structure*, we applied data from 2043 samples from different ancestries that were previously used as model testing dataset in the EUROFORGEN study on hair structure prediction [15]. Finally, for *freckles*, we used data from 1801 unrelated samples from the TwinsUK dataset, comprising European individuals from the United Kingdom [26]. For all traits, the available datasets were split into 80 % for model training and 20 % for model validation (Table 1).

As genetic markers in the prediction modelling, we used previously established DNA predictors for eye, hair, skin color, hair shape and freckles, respectively. In particular, for eye color prediction, we used the 6 SNPs from the previous IrisPlex eye color model [9]; for hair color prediction we used the 22 hair color informative SNPs from the previous HIrisPlex hair color model [11]; for skin color prediction, we used the 36

Table 1
EVC-specific data sets used for prediction model training and testing with and without the use of prevalence-informed priors.

Appearance trait	Training set (80 %)	Test set (20 %)	References
Eye color	876	219	[9,11,12]
Hair color	1361	341	[9,11,12]
Skin color	1054	264	[9,11,12]
Hair Structure	1634	409	[15]
Freckles	1440	361	[26]

skin color informative SNPs from the previous HIrisPlex-S skin color model [12]; for hair shape prediction, we used the 38 SNPs from the previous EUROFORGEN study on hair shape prediction [15]; and for freckles prediction, we used the 13 out of the 22 SNPs recently proposed for this purpose by Kukla-Bartoszek [13]. Not using the remaining 9 previously proposed freckles DNA predictors is explained by data availability and quality control issues (see below). Samples with incomplete genotype information per each EVC were excluded from our analysis.

2.2. Appearance trait categories

We considered the following trait categories:

- Eye color: Blue, Intermediate, Brown
- Hair color: Blond, Brown, Red, Black
- Skin color: Very Pale, Pale, Intermediate, Dark, Dark to Black
- Hair structure: Straight, Wavy, Curly
- Freckles: Freckled, Non-freckled

All traits were treated as categorical variables and were coded as ‘1’, ‘2’, ‘3’ etc. up to the number of considered categories, which ranged between two for the presence or absence of freckles and five for skin color. In the course of our study, five-class problems turned out to be extremely computationally expensive and prohibitive for a comprehensive analysis. To overcome this problem, we reduced the 5-class category problem for skin color into two 4-class problems by either merging the first two categories very pale and pale or the last two dark or dark to black. Predictive DNA markers were considered under an allele-based model and, correspondingly, numerically coded as 0 for homozygosity of the major, i.e. more frequent, allele, 1 for heterozygosity and 2 for homozygosity of the minor, i.e. less frequent, allele. We did not consider interaction terms in the prediction models, as recently proposed for instance for freckles [13], in order to allow a consistent derivation of the posterior probabilities in the Bayesian approach across all EVCs. That means that for all EVCs, the models were defined considering the additive effects of the corresponding genetic markers.

2.3. Data cleaning

All data sets had undergone previously described quality control [9,11,12,15] and could be readily used in the prediction models, except for the TwinsUK data set for freckles prediction. For this reason, we applied standard quality control on the raw Twins UK data in order to be able to use them further in our analysis. For the freckles prediction we considered the markers recently proposed from Kukla-Bartoszek [13]; however, only 14 out of the previously reported 22 markers were available in the TwinsUK dataset we received for this study up on request from the Department of Twin Research, King’s College London, of which 13 passed the quality-control and were thus used for freckles prediction modeling. More specific, we intended to remove markers that showed a strong deviation from Hardy-Weinberg equilibrium ($p < 0.001$), excessive heterozygosity (>0.001) [27], more than two alleles, an imputation info score of less than 0.8 or very low minor allele frequencies ($MAF < 0.01$). One of the markers did not pass this step of quality control, and was thus excluded from our analysis. Out of sample pairs with excessive identity-by-descent (IBD) allele sharing (>0.2), one randomly selected sample was removed in order to assure (approximate) independence. Finally, we performed a principal-components analysis (PCA) on the merged data set of TwinsUK and the complete dataset of the 1000 Genomes population data [28], comprising known ancestry, in order to identify and subsequently remove all samples with large-scale differences in ancestry. The latter were defined by the first two principal components, which were sufficient to cluster the individuals in population groups ($PC1 \geq 0.01$ and $PC2 \leq$ -0.02). From this data set, we extracted those performed using PLINK v1.9 [29] and ‘RStudio’ v 3.4.4 [30].

2.4. Statistical analysis

2.4.1. Prior-free trait prediction

For the prediction of eye, hair and skin color, we used standard multinomial logistic regression (MLR), as established by Liu et al. [4]. We also used MLR for the three-class problem of predicting hair structure [15], whereas standard binomial logistic regression (BLR) was used for predicting freckle presence or absence [13]. Individuals were predicted, or classified, as presenting with a specific trait category according to the highest posterior probability across all categories, with no minimal threshold imposed on this probability, although being explicitly equivalent to a minimum threshold of 1 by the number of trait categories. For all traits included in our study, each of the trait-specific data sets was randomly split into two independent subsets (Table 1), with 80 % being used for model training (training set) and 20 % for model prediction (test set).

2.4.2. Prior-incorporated trait prediction

In the absence of detailed trait prevalence information on virtually all externally visible characteristics (EVCs) considered here for different populations or continental groups, we sought to assess the impact of priors on the prediction performance by exhaustively exploring the space of all possible tupels, i.e. an ordered list with respect to categories, of prior probability values. More specific, we performed Bayesian classification based on either MLR or BLR, again depending on the number of trait classes, by including the prior information in the calculation of the posterior probabilities. For a 3-class trait, the model was formed as follows [4]:

$$\ln\left(\frac{p_2}{p_1}\right) = \alpha_2 + \sum_{j=1}^{k} \beta_2(\pi_2)_j x_j$$

$$\ln\left(\frac{p_3}{p_1}\right) = \alpha_3 + \sum_{j=1}^{k} \beta_3(\pi_3)_j x_j$$

where the p_i $(i = 1, 2,3)$ denote the probabilities of each category and α_i, β_i $(i = 2, 3)$ the respective regression coefficients, with the first category being used as reference, while $(\pi_1, \pi_2, \pi_3)_{\sum_{i=1}^{3} \pi_i = 1}$ forms the tupels of prior values for the three categories and k refers to the number of genetic markers included in the model, e.g. $k = 6$ for the IrisPlex model, whereas j is an index referring to those genetic markers. Estimates for α_i, β_i were obtained by the MLR model from the respective training data sets (Table 1). Analysis was conducted in R version 3.4.3 [31] by using the nnet R package [32]. Following the standard Bayesian prediction framework, posterior probabilities were then obtained as the product between the data-dependent likelihood and the prior information:

$$\frac{\tau_2}{\tau_1} = \frac{\pi_2}{\pi_1} \times \frac{f_2(x)}{f_1(x)}$$

$$\frac{\tau_3}{\tau_1} = \frac{\pi_3}{\pi_1} \times \frac{f_3(x)}{f_1(x)}$$

where τ_i, π_i, f_i $(i = 1, 2, 3)$ denote the posterior probabilities, the prior probabilities and the likelihoods for each of the three categories, respectively. From the above formulas, the posterior probabilities for each trait category were eventually obtained as:

$$\pi_2 = \frac{\exp\left(\ln\left(\frac{p_2}{p_1}\right) + a_2 + \sum_{j=1}^{k} \beta_2(\pi_2)_j x_j\right)}{1 + exp\left(\ln\left(\frac{p_2}{p_1}\right) + a_2 + \sum_{j=1}^{k} \beta_2(\pi_2)_j x_j\right) + exp\left(ln\left(\frac{p_2}{p_1}\right) + a_3 + \sum_{j=1}^{k} \beta_3(\pi_3)_j x_j\right)}$$

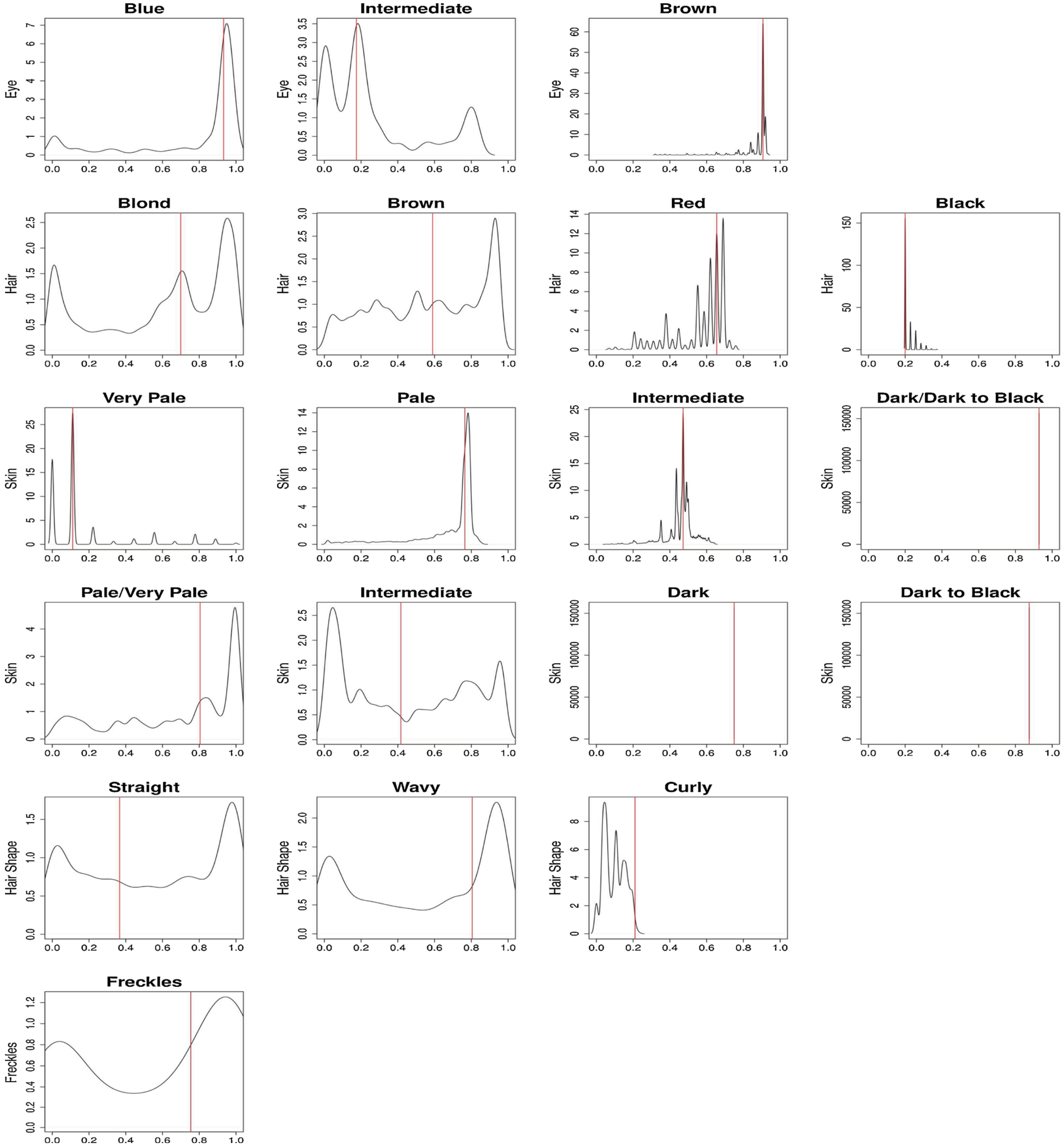

Fig. 1. Impact of the choice of trait prevalence priors on sensitivity in EVC prediction modeling from genetic data. Results are presented for a Bayesian approach using a multinomial logistic regression model for predicting four pigmentation trait categories, namely those of eye color (EC; first line), hair color (HC; second line) and skin color (SC; third line: darkest categories merged; fourth line: palest categories merged), where the vertical line corresponds to a prior-free prediction.

$$\pi_3 = \frac{\exp\left(\ln\left(\frac{p_3}{p_1}\right) + a_3 + \sum_{j=1}^{k}\beta_3(\pi_3)_j x_j\right)}{1 + exp\left(\ln\left(\frac{p_3}{p_1}\right) + a_3 + \sum_{j=1}^{k}\beta_3(\pi_3)_j x_j\right) + exp\left(ln\left(\frac{p_3}{p_1}\right) + a_2 + \sum_{j=1}^{k}\beta_2(\pi_2)_j x_j\right)}$$

$$\pi_1 = 1 - \pi_2 - \pi_3$$

where x_j denotes the number of minor (less frequent) alleles of the jth SNP and the terms α_i and $\beta_i\ (i = 2, 3)$ are the model parameters. As before, indicator j in the sum denotes the sum across all genetic markers. For simplicity, we did not consider interaction terms. This renders our approach only an approximation for the previously published freckles model. Models for the 2- and 4-class problems were defined in a similar fashion. A sample was classified into that category which yielded the maximum posterior probability, again without explicitly applying any minimal threshold.

With lacking trait prevalence information, we exhaustively explored the impact of priors by considering all possible tupels of prior probabilities in order to assess potential prediction improvement but also the risk caused by mis-specifying prior values. To this end, prior probabilities in turn assumed values from 0.01 to 0.99, with step size 0.01, while

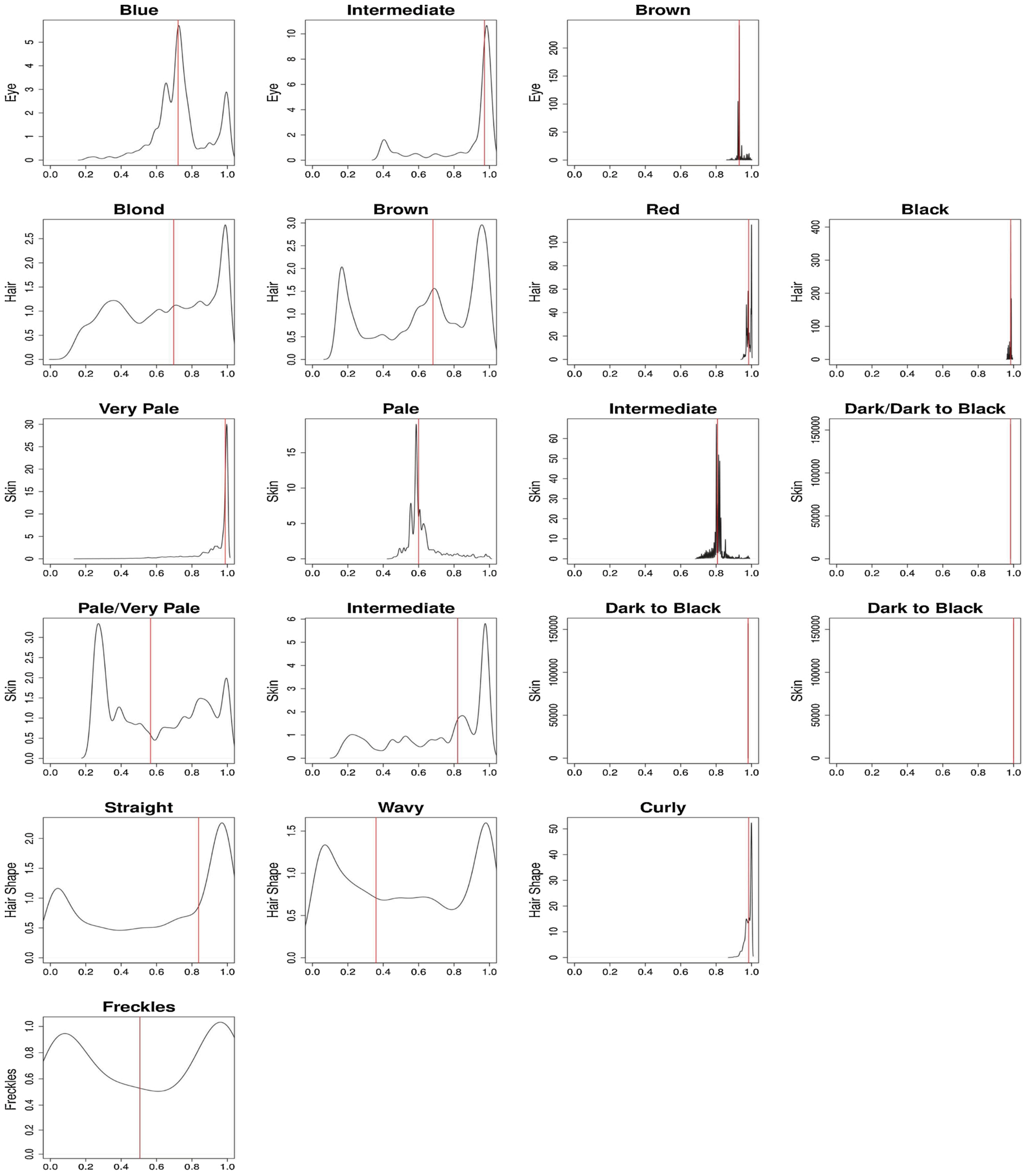

Fig. 2. Impact of the choice of trait prevalence priors on specificity in EVC prediction modeling from genetic data. Results are presented for a Bayesian approach using a multinomial logistic regression model for predicting four pigmentation trait categories, namely those of eye color (EC; first line), hair color (HC; second line) and skin color (SC; third line: darkest categories merged; fourth line: palest categories merged), where the vertical line corresponds to a prior-free prediction.

requiring that those probabilities for all categories sum to unity. Note that prior-free prediction is equivalent to a Bayesian prediction model where the prior probabilities correspond to the relative trait category frequencies in the training set.

2.4.3. *Prediction performance assessment*

Prediction performance was evaluated in the respective test data sets (Table 1). We calculated commonly used measures of test accuracy, namely sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), area-under-curve (AUC) and overall accuracy for all possible tupels of prior probabilities and subsequently summarized their distribution. In lay terms, sensitivity denotes the proportion of correctly predicted samples among all who manifest the trait category of interest (true-positive rate), whereas specificity denotes the proportion of correctly predicted samples among all that do not manifest the trait category of interest (true-negative rate). On the other hand, a PPV refers to the proportion of correct classifications among all predictions of the trait category of interest, while an NPV refers to the proportion of correct classifications among all predictions other than the trait category of interest. The AUC denotes the area under the receiver-operating characteristic (ROC) curve which is obtained by varying the threshold used for the classification decision and can be interpreted as an indicator for the separability of classes when using the particular classification model. Except for freckles, which comprise only two categories, we performed multiclass ROC analysis, which carries out pairwise comparisons across all categories (one class vs. all other classes). For example, for eye color the following comparisons were conducted: "Blue vs. Non-Blue", "Intermediate vs. Non-Intermediate" and "Brown vs. Non-Brown". Finally, the overall model accuracy refers to the number of correct predictions divided by the total number of predictions made. As pointed out by Caliebe et al. [33], FDP does not operate in a diagnostic-test environment where (bio-)markers are being used to infer the presence of causal factors or conditions and where prediction is done in the opposite direction of causation. Instead, causal genetic markers, or proxies thereof, are used to predict the outcome along the causal direction. Thus, for FDP the most relevant performance measures are the predictive values (PPV and NPV). All analyses were performed in R v3.4.3 [30], using the packages nnet [32] and caret [34] for model building and performance assessment calculation, respectively, and package caTools [35] for multiclass AUC calculation. For visualization of our results, we used the package plot3D [36] and for better interpretability the standard kernel density estimation was used.

3. Results

The use of trait prevalence-informed priors usually had a strong impact on the performance of the prediction model, although the extent differed between EVCs and also between categories of the same EVC (see below). We found that prediction performance of prior-free models could be improved by a substantial proportion of tupels of prior values in the respective models. On the other hand, and perhaps not surprisingly, a substantial proportion of prior tupels led to a deteriorated prediction performance compared to the respective prior-free model.

3.1. Impact of trait prevalence-informed priors on sensitivity and specificity

With few exceptions, sensitivity (Fig. 1) and specificity (Fig. 2) were strongly affected by variation in trait prevalence-informed prior values. A particular choice of priors could shift sensitivity usually in both directions from that of the prior-free model, often even approaching the extreme values of 0 or 1, respectively. All traits showed a strong dependence of their prediction sensitivity on the choice of prior values, most strongly for blue and intermediate eye color, blond and brown hair color, hair structure and freckles. Skin color categories seemed to be less affected by the choice of prior values especially when the darkest categories were merged, but not when the palest categories were merged. Notable exceptions were dark and dark to black skin color, which appeared barely affected by the choice of prior values. In general, specificity of predicting lighter eye and skin color was more strongly impacted by changing prior values than darker tones, as were straight and wavy hair structure categories as well as the presence of freckles. Similarly, blond and brown hair colors were more strongly affected compared to the categories of red and black hair color. Strikingly, dark skin and hair color, but also red hair and curly hair structure appeared almost insensitive in their prediction specificity when it comes to the use of prevalence priors. Interestingly, the probability of a shift away from the prior-free prediction differed between the directions as well as the average extent of this shift for both sensitivity (Table 2) and specificity (Table 3) across all EVCs. In general, we noticed that most of the prior tupels were above or equal to the prior- free value for all EVCs apart from a few exceptions. These exceptions included some skin color categories such as very pale and intermediate, whose sensitivity seemed to be lower than that of the prior-free approach for most prior tupels. The majority of prior tupels for the specificity of blue and intermediate eye color also resulted into lower values than the prior-free approach.

Of note, the distributions of sensitivity and specificity across the space of possible prior values assumed an almost discrete form for skin color when the darkest categories merged, most prominently for the light skin categories. Predicting dark and dark to black skin colors by using prevalence priors does not show any difference from the performance of the prior-free approach, also in the case where these two

Table 2
Shift in sensitivity in EVC prediction modeling from genetic data for the prior-based models compared to the prior-free models.

Trait	Category	Below [%]	Above [%]
Eye color	Blue	35.6	64.4
	Intermediate	38.6	61.4
	Brown	28.9	71.1
Hair color	Blond	51.4	48.6
	Brown	49.8	50.1
	Red	56.6	43.4
	Black	0.0	100.0
Skin color (4/5)	Very Pale	78.9	21.1
	Pale	55.7	44.3
	Intermediate	63.1	36.9
	Dark/Dark to Black	0.0	100.0
Skin color (1/2)	Very Pale/Pale	48.9	51.1
	Intermediate	48.9	51.1
	Dark	0.0	100.0
	Dark to Black	0.0	100.0
Hair structure	Straight	38.3	61.7
	Wavy	59.0	41.0
	Curly	98.9	1.1
Freckles	Freckled/Non-freckled	49.5	50.5

Proportion of prior tupels resulting in sensitivity values below and above the value for the prior-free approach, respectively.
Skin color (4/5) is referring to the skin color prediction when the two darkest categories of dark and dark to black were merged and considered as one single category. Similarly Skin color (1/2) is referring to the case when the two palest categories of very pale and pale were merged and considered as one.

Table 3
Shift in specificity in EVC prediction modeling from genetic data for the prior-based models compared to the prior-free models.

Trait	Category	Below [%]	Above [%]
Eye color	Blue	52.5	47.5
	Intermediate	60.7	39.3
	Brown	27.7	72.3
Hair color	Blond	52.8	47.2
	Brown	49.4	50.6
	Red	43.5	56.5
	Black	24.9	75.1
Skin color (4/5)	Very Pale	52.8	47.2
	Pale	53.8	46.2
	Intermediate	41.2	58.8
	Dark/Dark to Black	100.0	0.0
Skin color (1/2)	Very Pale/Pale	48.9	51.1
	Intermediate	48.9	51.1
	Dark	0.0	100.0
	Dark to Black	0.0	100.0
Hair structure	Straight	59.0	41.0
	Wavy	39.7	60.3
	Curly	41.6	58.4
Freckles	Freckled/Non-freckled	49.5	50.5

Proportion of prior tupels resulting in specificity values below and above the value for the prior-free approach, respectively.
Skin color (4/5) is referring to the skin color prediction when the two darkest categories of dark and dark to black were merged and considered as one single category. Similarly Skin color (1/2) is referring to the case when the two palest categories of very pale and pale were merged and considered as one.

categories were considered as a single one.

3.2. Impact of trait prevalence-informed priors on positive and negative predictive values

Similar to the results for sensitivity and specificity, positive predictive values (PPV; Fig. 3) and negative predictive values (NPV; Fig. 4) were, with few exceptions, strongly affected by the choice of prior values for EVCs such as eye and hair color. Quite similarly, the impact was again strongest for freckles and hair structure. More specifically for the latter, PPV appeared to be quite sensitive for all categories in the change of prior values, while the impact on NPV seems to be larger for all categories apart from curly hair. Regarding skin color, the impact of prevalence priors on PPV and NPV was very small when the darkest categories were merged. When merging the palest categories, the impact of different prior values was very small only for the categories of dark and dark to black.

The values of PPV and NPV differed regarding the direction and also

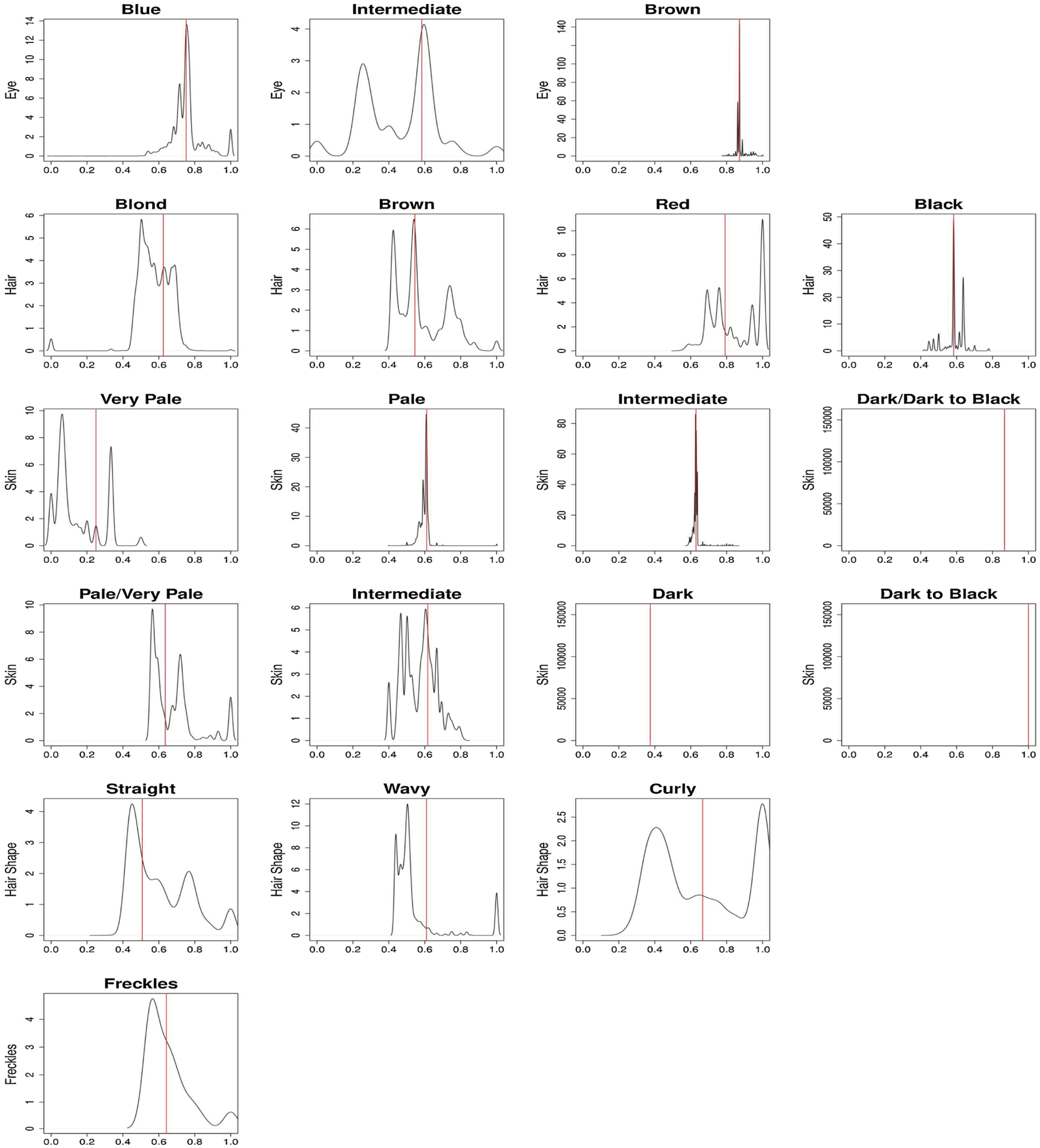

Fig. 3. Impact of the choice of trait prevalence priors on positive predictive values (PPV) in EVC prediction modeling from genetic data. Results are presented for a Bayesian approach using a multinomial logistic regression model for predicting four pigmentation trait categories, namely those of eye color (EC; first line), hair color (HC; second line) and skin color (SC; third line: darkest categories merged; fourth line: palest categories merged), where the vertical line corresponds to a prior-free prediction.

the average extent for each of the considered traits when priors were incorporated in the model (Tables 4 and 5). For example, for freckles we observed that most of the prior tupels incorporated in the model seem to perform better compared to the prior free approach for both PPV and NPV. However, for most of the other traits, we observed that only almost half of the prior tupels showed a better performance when compared to the model without priors, while the other half seemed to show an inferior performance but only for specific categories with respect to both measurements. For example, the brown eye color category showed a high percentage above or equal to the prior-free value for PPV as well as NPV, while for blue and intermediate eye categories the percentage above or equal the prior-free approach is ranging around 50 %.

Red and black hair color appeared to be barely affected by the choice of prior tupels compared to blond and brown especially for NPV, while freckles and eye color showed in general high susceptibility in both measurements, with the only exception of the brown eye color category which seemed less impacted. Generally, we observed small effects for skin color when dark and dark to black categories were considered as

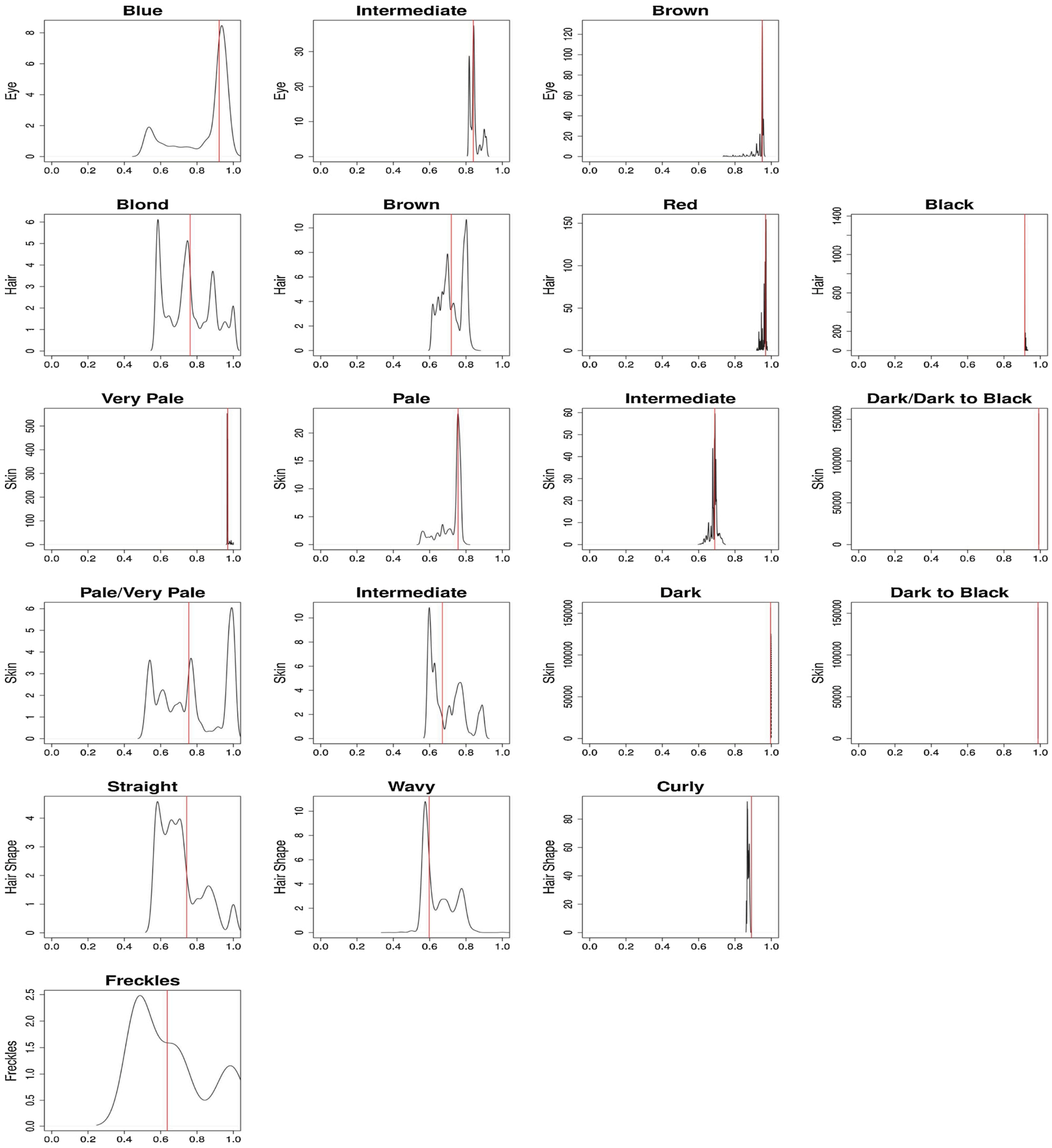

Fig. 4. Impact of the choice of trait prevalence priors on negative predictive values (NPV) in EVC prediction modeling from genetic data. Results are presented for a Bayesian approach using a multinomial logistic regression model for predicting four pigmentation trait categories, namely those of eye color (EC; first line), hair color (HC; second line) and skin color (SC; third line: darkest categories merged; fourth line: palest categories merged), where the vertical line corresponds to a prior-free prediction.

Table 4
Shift in PPV in EVC prediction modeling from genetic data for the prior-based models compared to the prior-free models.

Trait	Category	Below [%]	Above [%]
Eye color	Blue	54.5	41.7
	Intermediate	59.3	24.3
	Brown	37.4	62.6
Hair color	Blond	63.2	30.3
	Brown	49.8	50.2
	Red	45.3	54.7
	Black	60.9	39.1
Skin color (4/5)	Very Pale	60.5	16.6
	Pale	82.1	17.9
	Intermediate	52.8	47.2
	Dark/Dark to Black	0.0	100.0
Skin color (1/2)	Very Pale/Pale	50.8	49.2
	Intermediate	69.7	30.3
	Dark	0.0	100.0
	Dark to Black	0.0	100.0
Hair structure	Straight	37.0	56.1
	Wavy	83.4	10.5
	Curly	50.9	44.1
Freckles	Freckled/Non-freckled	33.3	50.5

Proportion of prior tupels resulting in positive-predictive values (PPV) values below and above the value for the prior-free approach, respectively. In cases where percentages above and below the prior-free approach do not sum to 100 is obtained due to the occurrence of NAs in this model measurements. Thus, those observations were omitted.
Skin color (4/5) is referring to the skin color prediction when the two darkest categories of dark and dark to black were merged and considered as one single category. Similarly Skin color (1/2) is referring to the case when the two palest categories of very pale and pale were merged and considered as one.

one combined category, while merging pale and very pale categories resulted in being more sensitive to the choice of the prior values. For hair structure, the NPV of the curly category was slightly impacted, while the PPV seemed to be very sensitive to prior tupel choice.

3.3. *Impact of trait prevalence-informed priors on AUC and overall accuracy*

Finally, we assessed the overall performance by means of area-under-curve (AUC) and overall accuracy values. We generally observed only a small impact of the choice of prior values on AUC for all EVCs tested (Fig. 5). More specifically, all categories for hair structure and hair color appeared to be barely affected in AUC by the varying prior tupels, with brown, red and black showing a smaller impact compared to blond. Blue and brown eye colors were also barely affected, while intermediate eye color appeared a bit more susceptible. Similar to the aforementioned EVCs, the effect of priors on skin color categories was generally small, either when the palest either when the darkest categories were merged. The category of pale/very pale skin color appeared to perform worse in the model when priors were incorporated compared to the prior-free approach. AUC for freckles showed independence from the choice of prior values since its value remained stable for all possible prior tupels.

Table 5
Shift in NPV in EVC prediction modeling from genetic data for the prior-based models compared to the prior-free models.

Trait	Category	Below [%]	Above [%]
Eye color	Blue	49.1	50.9
	Intermediate	53.1	46.9
	Brown	41.5	58.5
Hair color	Blond	58.5	41.5
	Brown	50.0	50.0
	Red	70.6	29.4
	Black	44.5	55.5
Skin color (4/5)	Very Pale	64.9	35.1
	Pale	64.3	35.7
	Intermediate	54.3	45.7
	Dark/Dark to Black	100.0	0.0
Skin color (1/2)	Very Pale/Pale	45.8	54.2
	Intermediate	50.8	49.2
	Dark	0.0	100.0
	Dark to Black	0.0	100.0
Hair structure	Straight	71.3	27.7
	Wavy	45.6	54.4
	Curly	100.0	0.0
Freckles	Freckled/Non-freckled	42.4	50.5

Proportion of prior tupels resulting in negative-predictive values (NPV) below and above the value for the prior-free approach, respectively.
Skin color (4/5) is referring to the skin color prediction when the two darkest categories of dark and dark to black were merged and considered as one single category. Similarly Skin color (1/2) is referring to the case when the two palest categories of very pale and pale were merged and considered as one.

Regarding AUC values, most of the categories showed that almost half of the prior tupels performed above or equal to the prior-free approach (Table 6). There were few exceptions, such as hair structure and freckles, where most of the proportions were above the prior-free AUC value. Regarding very pale and pale/very pale, almost all prior prediction values seem to perform worse than the prior-free approach.

In comparison to AUC, overall prediction accuracy (Fig. 6) was much more affected by the choice of prior values. All five EVCs showed substantial susceptibility to the choice of priors reflected in the overall accuracy. Notably, there was some room for improvement for overall prediction accuracy except from hair structure, which seemed to perform worse compared to the prior-free approach. However, the overwhelming majority of prior tupels led to accuracy deterioration (Table 7). We also noticed that misspecification of priors often caused a deterioration in the prediction performance measurements for some traits as well as in the overall accuracy (Fig. 6).

4. Discussion

In the present study, we aimed at assessing the impact of using trait prevalence-informed priors on the prediction accuracy of an expanded set of EVCs, including eye, hair and skin color as well as hair structure and freckles. Our study was motivated by the question if such prior information, possibly representing trait class prevalence in biogeographic ancestry groups, may improve the prediction accuracy of traits over prior-free models. For all EVCs except freckles, we used for our models the same predictive markers as applied in the previously established prediction models [9,11,12,15]. Although due to data availability issues the number of predictors was lower in our freckles prediction modeling than previously [13] this discrepancy shall not affect our main outcomes for freckles significantly, since we applied the same reduced marker set to both the model with and without priors. Regarding the prior information, we surprisingly noticed that there is a limited spatial and population-specific trait prevalence information available for hair, skin and eye color, hair structure [24] and even non-existent for other traits such as freckles. We therefore exhaustively investigated the impact of the choice of prior values for the different trait categories on a fine-grained grid of all possible sets, or tupels, of values to obtain a general picture of the impact of priors on prediction performance. To this end, we trained and tested Bayesian versions of multinomial logistic regression (MLR) and binomial logistic regression (BLR) models, respectively, and compared their performance to the respective prior-free versions, using different trait-specific data sets.

Our results showed that the use of trait prevalence-informed priors can have a strong impact on the performance of the prediction models for the 5 EVCs tested. Such use carries some potential to improve the prediction of most EVCs and some of their categories compared to a prior-free approach, as evidenced by a substantial proportion of prior tupels with better performance statistics. However, we also found large proportions of prior tupels that led to inferior prediction results, indicating the risk that the misspecification of those priors may lead to a gross deterioration in the model performance. This deterioration could be explained by the fact that the true prevalence values are unknown. The prior-free approach is influenced by the proportions of the categories in the data set. Random splitting into separate training (80 %) and test (20 %) datasets, as performed here for all EVCs, resulted in approximately equal proportions for each of the trait categories in these two data sets, respectively. In consequence, the trained model was well adapted to the category proportions in the test data set, possibly leading to some over-fitting of the model. This may have led to a slight over-estimation of the performance of the prior-free models. Accurate trait prevalence specification is of utmost importance to obtain reliable and accurate predictions. However, with the lack of such information, the application of prior-incorporating Bayesian approaches for EVC prediction in forensic cases appears not feasible at this stage.

Given the lack of spatial or population-specific prevalence information for the EVCs considered in this study, which represented a significant obstacle to our analysis, we were not able to compare the performance of prior-incorporating and prior-free approaches against a gold standard. As gold standard we should have had reliable population-representative prior values for all EVCs and their categories, which, however, are not available. Therefore, we explored the impact of priors across the whole space of

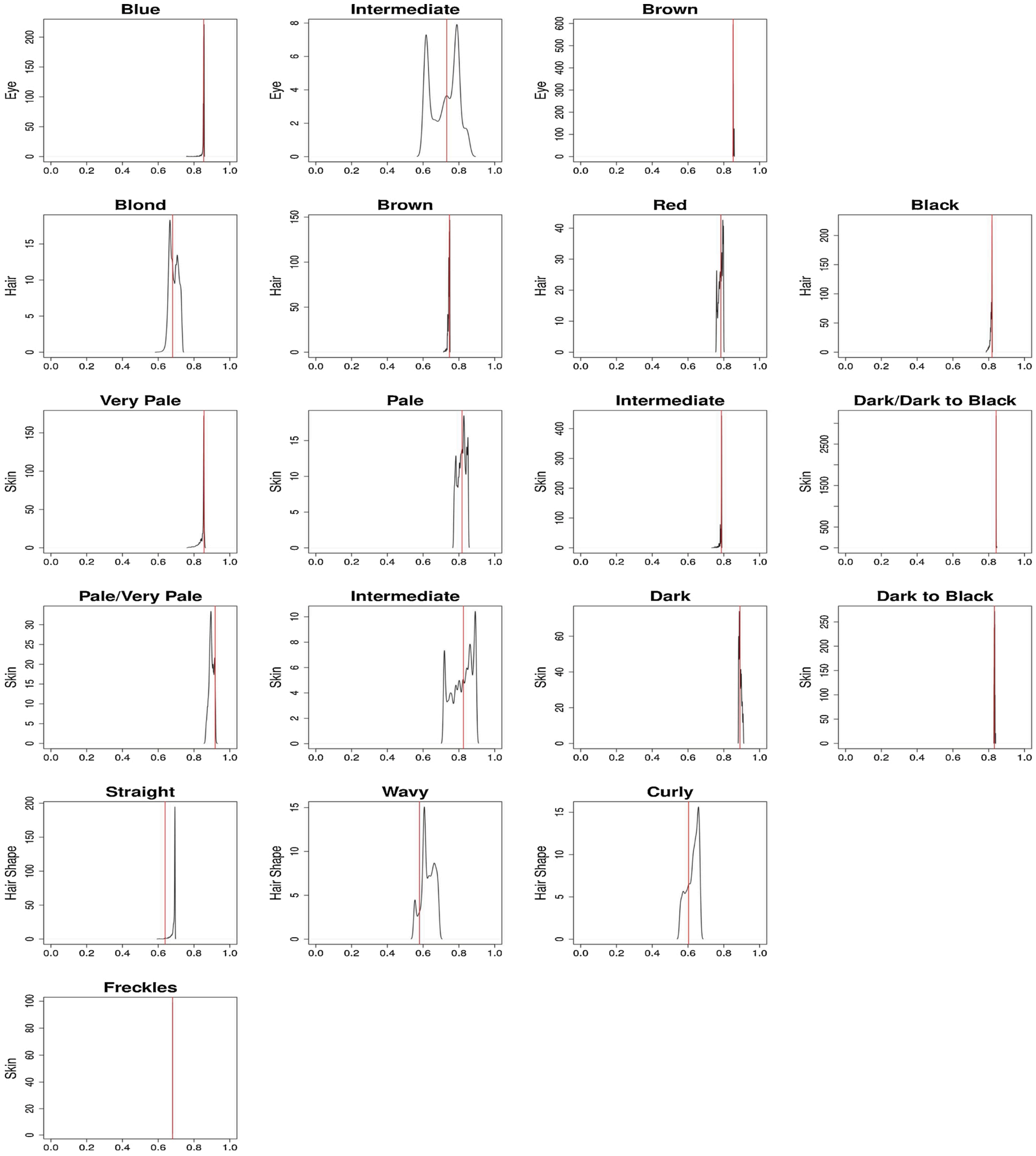

Fig. 5. Impact of the choice of trait prevalence priors on the area-under-curve (AUC) in EVC prediction modeling from genetic data. Results are presented for a Bayesian approach using a multinomial logistic regression model for predicting four pigmentation trait categories, namely those of eye color (EC; first line), hair color (HC; second line) and skin color (SC; third line: darkest categories merged; fourth line: palest categories merged), where the vertical line corresponds to a prior-free prediction.

possible tupels. Another possible interpretation of our approach, given the lack of knowledge about the underlying "truth" regarding the knowledge on trait prevalence over geographic space, is that the priors resemble differential costs for misclassification, which may also be an interesting future approach in forensic applications.

Little susceptibility of the prediction outcome to the choice of prior values, represented by likelihood ratio values of large magnitude compared to those of the priors, likely reflects a large extent of genetic determination of a trait or a particular trait category and that a large proportion of the causal genetic variants determining this trait, or at least their strongly correlated proxies, are already included in the prediction model [5,37–39]. This agrees with the statement of Caliebe et al. [33] that trait prevalence values provide no (or little) additional information if all (or almost all) genetic trait-determining variants are included as predictors in the model, i.e. that the prediction is independent of the population. From all EVCs and their categories investigated here, red hair color prediction comes closest to this, as red hair is determined by only one gene, *MC1R*, from which multiple DNA variants, most of them being non-synonymous DNA variants that are likely causal, are included in the hair color prediction model based on the HIrisPlex markers for hair color prediction used here. For complex traits or trait categories, however, dozens or even hundreds of genetic factors will contribute to the trait and usually only a fraction of them is known and included in the prediction model. It is assumed that all EVCs and EVC categories, including those tested here besides red hair are complex traits or trait categories determined by large numbers of genes, respectively. This was already demonstrated for hair and skin color based on large-scale genome-wide association studies (GWAS) [40,41], and therefore is also expected for eye color for which such a large-scale GWAS is currently pending. For hair shape and freckles the previous GWAS were not yet on such large scale, but those multiple genes that were successfully identified showed mostly small effect sizes and explained only a fraction of the estimated heritability [42,43]; only large-size GWAS will be able to increase the explained heritably in the future. For complex phenotypes, use of prevalence values may actually increase prediction accuracy if specified correctly, because they contain information on, and can act as proxies for, those variants that also contribute but are not included in the model.

The strong dependency of prediction performance on priors for most traits and categories further reflects that many, if not most, predictions are made based on only moderately different posterior probabilities and, in turn, likelihoods do not differ strongly between the categories, because not all causal factors are yet known and could therefore be included in the prediction models. Use of priors may then easily shift classification decisions, thereby simply facilitating a trade-off between sensitivity and specificity as well as PPV and NPV in the absence of information on true trait prevalence values. Interestingly, the AUC appeared to be largely unaffected by changing prior tupels.

Both observations, the potential for prediction improvement by use of priors as well as the risk of inferior performance when those priors are mis-specified, motivate future studies. An important and preferable way would be to identify more causal genetic factors involved in EVC etiology, thereby obliterating the need for proxies of those causal factors. However, given their likely small and at most moderate effects, this would require very large data sets for future studies to identify such genetic variants. For instance, a recent GWAS on hair color tested more than 290,000 individuals in an European discovery dataset that led to the identification of 124 associated independent genetic loci at genome-wide significance, of which 111 were novel [40]. However, most of these DNA variants will not be causal themselves, because of the focus of commonly used SNP microarrays on markers that allow for good imputation of other, common markers ('imputation backbone'), while providing only limited numbers of SNPs centered on gene regions or selected phenotypic relevance ('contents enrichment').

Another area for future research is to collect, for as many populations from as many geographic regions that are relevant based on the phenotypic variation of the EVCs to be predicted, trait prevalence data on the same or higher level of detail (e.g. categories) as achievable by DNA-based EVC prediction. However, even when such data are available, the use of forensic ancestry DNA testing to identify the geographic region for which EVC trait prevalence data are to be allocated for use as priors in EVC prediction will only be applicable, in case the prevalence values for different populations within such DNA-identified region do not show much variation, and if the regional geographic ancestry can be inferred with high confidence from the crime scene DNA sample. While collection of prevalence data may be achievable in the future, provided such studies are carried out with suitable geographic coverage and EVC phenotypic details, and given that regional such as continental ancestry inference based on enough DNA markers already is possible [44], the trait variation within DNA-identifiable geographic regions remains as problem. For instance, within Europe, which as continental region is identifiable with forensic DNA ancestry testing [44], eye and hair color prevalence values largely vary between populations from different parts of Europe. Thus, averaging such population prevalence values, if available, will not result in suitable priors for any person originating from any European population. This could only be solved by increasing the level of detail of DNA-based ancestry testing to the sub-regional or even population level, which currently, however, is not achievable and also is not expected to be achievable in the near future. Identifying genetic geographic population substructure within continents, such as within Europe [45], requires thousands of autosomal SNPs – a number that currently cannot be achieved given available technologies that are suitable for forensic DNA analysis. The simultaneous and targeted analysis of many thousands of SNPs in low-quantity and low-quality DNA typically available from crime scene stains requires the development of new DNA technology in the future.

In summary, our results provide a first assessment of the impact of trait prevalence-informed priors on the prediction model performance for several EVCs. Incorporation of priors, possibly informed by trait class prevalence values in biogeographic ancestry groups, can improve the performance of predicting appearance traits, but a correct specification of those priors appears mandatory to protect against a deteriorated performance. Future work is needed to obtain unbiased estimates of trait prevalence for EVCs to be predicted in a large variety of populations, when mostly non-causal genetic

Table 6
Shift in AUC in EVC prediction modeling from genetic data for the prior-based models compared to the prior-free models.

Trait	Category	Below [%]	Above [%]
Eye color	Blue	50.5	49.5
	Intermediate	50.5	49.5
	Brown	43.8	56.2
Hair color	Blond	43.9	56.1
	Brown	70.9	29.1
	Red	49.2	50.8
	Black	84.9	15.1
Skin color (4/5)	Very Pale	90.1	9.90
	Pale	49.2	50.8
	Intermediate	46.2	53.8
	Dark/Dark to Black	53.4	46.6
Skin color (1/2)	Very Pale/Pale	96.4	3.64
	Intermediate	50.8	49.2
	Dark	50.8	49.2
	Dark to Black	50.8	49.2
Hair structure	Straight	1.32	98.7
	Wavy	11.6	88.4
	Curly	27.8	72.3
Freckles	Freckled/Non-freckled	0.0	100.0

Proportion of prior tupels resulting in area-under-curve (AUC) values below and above the value for the prior-free approach, respectively.
Skin color (4/5) is referring to the skin color prediction when the two darkest categories of dark and dark to black were merged and considered as one single category. Similarly Skin color (1/2) is referring to the case when the two palest categories of very pale and pale were merged and considered as one.

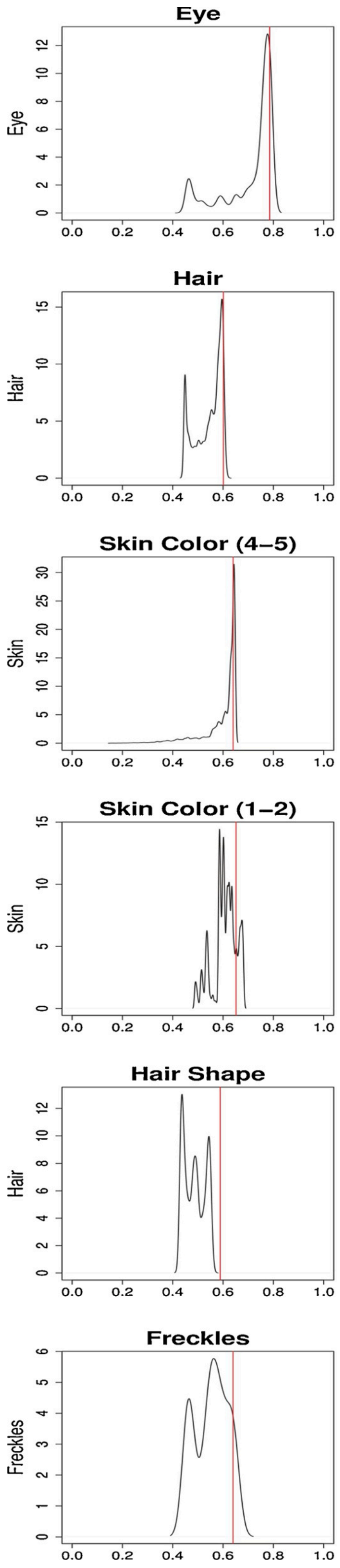

Fig. 6. Impact of the choice of trait prevalence priors on the overall accuracy in EVC prediction modeling from genetic data. Results are presented for a Bayesian approach using a multinomial logistic regression model for predicting four pigmentation trait categories, namely those of eye color (EC; first line), hair color (HC; second line) and skin color (SC; third line: darkest categories merged; fourth line: palest categories merged), where the vertical line corresponds to a prior-free prediction.

Table 7
Shift in overall accuracy in EVC prediction modeling from genetic data for the prior-based models compared to the prior-free models.

Trait	Below [%]	Above [%]
Eye color	75.24	24.75
Hair color	90.97	9.027
Skin color (4/5)	72.96	27.03
Skin color (1/2)	80.92	19.07
Hair structure	100.0	0.0
Freckles	87.87	12.12

Proportion of prior tupels resulting in overall accuracy values below and above the value for the prior-free approach, respectively.
Skin color (4/5) is referring to the skin color prediction when the two darkest categories of dark and dark to black were merged and considered as one single category. Similarly Skin color (1/2) is referring to the case when the two palest categories of very pale and pale were merged and considered as one.

markers are continued to being used for trait prediction. This need will be reinforced by future GWAS whose larger sample sizes will allow the detection of genetic markers with even smaller effect sizes, yet most of them likely being non-causal. Finally, appearance trait research has to overcome the assembly of ever more associated, yet non-causal genetic markers and, via experimental evidence, has to arrive at the identification of the actual causal genetic factors for EVCs. If successful, this will allow to achieve accurate EVC prediction in a population-independent way, eventually rendering the use of trait prevalence priors obsolete in the future.

Funding

This study received support from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 740580 within the framework of the *Visible Attributes through Genomics* (VISAGE) Project and Consortium. TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd and the National Institute for Health Research (NIHR) funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. The funding organization had no influence on the design, conduct or conclusions of the study.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Appendix

Centres and investigators of the VISible Attributes through GEnomics (VISAGE) Consortium

Website: http://www.visage-h2020.eu/Erasmus **University Medical Center Rotterdam (Netherlands):** Manfred Kayser, Vivian Kalamara, Arwin Ralf, Athina Vidaki.

Jagiellonian University (Poland): Wojciech Branicki, Ewelina Pośpiech, Aleksandra Pisarek.

Universidade de Santiago de Compostela (Spain): Ángel Carracedo, Maria Victoria Lareu, Christopher Phillips, Ana Freire-Aradas, Ana Mosquera-Miguel, María de la Puente.

Medizinische Universität Innsbruck (Austria): Walther Parson, Catarina Xavier, Antonia Heidegger, Harald Niederstätter.

Universität zu Köln (Germany): Michael Nothnagel, Maria-Alexandra Katsara, Tarek Khellaf.

King's College London (United Kingdom): Barbara Prainsack, Gabrielle Samuel.

Klinikum der Universität zu Köln (Germany): Peter M. Schneider,

Theresa E. Gross, Jan Fleckhaus.

Bundeskriminalamt (Germany): Ingo Bastisch, Nathalie Schury, Jens Teodoridis, Martina Unterländer.

Institut National De Police Scientifique (France): François-Xavier Laurent, Caroline Bouakaze, Yann Chantrel, Anna Delest, Clémence Hollard, Ayhan Ulus, Julien Vannier.

Netherlands Forensic Institute (Netherlands): Titia Sijen, Kris van der Gaag, Marina Ventayol-Garcia.

National Forensic Centre, Swedish Police Authority (Sweden): Johannes Hedman, Klara Junker, Maja Sidstedt.

Metropolitan Police Service, London (United Kingdom): Shazia Khan, Carole E. Ames, Andrew Revoir.

Centralne Laboratorium Kryminalistyczne Policji (Poland): Magdalena Spólnicka, Ewa Kartasińska, Anna Woźniak.

References

[1] M. Kayser, Forensic DNA Phenotyping: predicting human appearance from crime scene material for investigative purposes, Forensic Sci. Int. Genet. 18 (2015) 33–48.
[2] M. Kayser, d.K. P, Improving human forensics through advances in genetics, genomics and molecular biology, Nat. Rev. Genet. 12 (2011) 179–192.
[3] M. Kayser, P.M. Schneider, DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations, Forensic Sci. Int. Genet. 3 (3) (2009) 154–161.
[4] F. Liu, et al., Eye color and the prediction of complex phenotypes from genotypes, Curr. Biol. (2009) 19.
[5] W. Branicki et al. Model-based prediction of human hair color using DNA variants. 129(4) (2011): p. 443-454.
[6] S. Walsh, et al., Global skin colour prediction from DNA, Hum. Genet. 136 (7) (2017) 847–863.
[7] J. Alghamdi, et al., Eye color prediction using single nucleotide polymorphisms in Saudi population, Saudi J. Biol. Sci. (2018).
[8] Y. Ruiz, et al., Further development of forensic eye color predictive tests, Forensic Sci. Int. Genet. 7 (2013) 28–40.
[9] S. Walsh, et al., 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 (2011) 170–180.
[10] E. Pospiech, et al., 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 (2014) 64–72.
[11] S. Walsh, et al., The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA, Forensic Sci. Int. Genet. 7 (1) (2013) 98–115.
[12] L. Chaitanya, et al., The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation, 35, 2018, pp. 123–135.
[13] M. Kukla-Bartoszek, et al., DNA-based predictive models for the presence of freckles, Forensic Sci. Int. Genet. 42 (2019) 252–259.
[14] F. Liu, et al., Prediction of male-pattern baldness from genotypes, Eur. J. Hum. Genet. 24 (6) (2016) 895–902.
[15] E. Pospiech, et al. Towards broadening Forensic DNA Phenotyping beyond pigmentation: Improving the prediction of head hair shape from DNA (2018). 37: p. 241–251.
[16] B. Hernando, et al., Genetic determinants of freckle occurrence in the Spanish population: towards ephelides prediction from human DNA samples, Forensic Sci. Int. Genet. 33 (2018) 38–47.
[17] S.P. Hagenaars, et al., Genetic prediction of male pattern baldness, PLoS Genet. 13 (2) (2017).
[18] M. Marcinska, et al., Evaluation of DNA variants associated with Androgenetic Alopecia and their potential to predict male pattern baldness, PLoS Genet. 10 (5) (2015) e0127852.
[19] F. Peng et al. Genome-Wide Association Studies Identify Multiple Genetic Loci Influencing Eyebrow Color Variation in Europeans (2019). 139: p. 1601–1605.
[20] F. Liu, et al., Common DNA variants predict tall stature in Europeans, Hum 133 (5) (2013) 587–597.
[21] S. Walsh, et al., DNA-based eye colour prediction across Europe with the IrisPlex system, Forensic Sci. Int. Genet. 6 (3) (2012) 330–340.
[22] O. Maroñas, et al., Development of a forensic skin colour predictive test, Forensic Sci. Int. Genet. 13 (2014) 34–44.
[23] J. Söchtig, et al., Exploration of SNP variants affecting hair colour predictionin Europeans, Int. J. Legal Med. 129 (5) (2015).
[24] M.A. Katsara, M. Nothnagel, True colors: a literature review on the spatial distribution of eye and hairpigmentation, Forensic Sci. Int. Genet. 39 (2019) 109–118.
[25] G. McLachlan, Discriminant Analysis and Statistical Pattern Recognition, John Wiley & Sons, Inc., New Jersey, 2004.
[26] A. Moayyeri, et al., The UK adult twin registry (TwinsUK resource), Twin Res. Hum. Genet. 16 (1) (2012) 144–149.
[27] C.A. Anderson, et al., Data quality control in genetic case-control association studies, Nat. Protoc. 5 (9) (2010) 1564–1573.
[28] T.G.P. Consortium, A global reference for human genetic variation, Nat. Int. J. Sci 526 (2015) 68–74.
[29] C.C. Chang, et al., Second-generation PLINK: rising to the challenge of larger and richer datasets, GigaScience 4 (1) (2015).
[30] R. Team, Integrated Development Environment for R, RStudio, 2016.
[31] R Development Core Team: the R Project for Statistical Computing, 2018. Available from: https://www.r-project.org/.
[32] Venables, W.N. and B.D. Ripley, Modern Applied Statistics with S, 2002, New York: Springer.
[33] A. Caliebe, et al., Likelihood ratio and posterior odds in forensic genetics: two sides of the same coin, Forensic Sci. Int. Genet. 28 (2017) 203–210.
[34] Core, M.K.C.f.J.W.a.S.W.a.A.W.a.C.K.a.A.E.a.T.C.a.Z.M.a.B.K.a.t.R., caret: Classification and Regression Training (2019).
[35] J. Tuszynski, caTools: Tools: Moving Window Statistics, GIF, Base64, ROC AUC, 2019 etc.
[36] K. Soetaert, plot3D: Plotting Multi-Dimensional Data, 2017.
[37] J.V. Schaffer, J.L. Bolognia, The melanocortin-1 receptor: red hair and beyond, Arch. Dermatol. 137 (11) (2001) 1477–1485.
[38] C.J. Binkley, et al., Genetic variations associated with red hair color and fear of dental pain, anxiety regarding dental care and avoidance of dental care, J. Am. Dent. Assoc. 140 (7) (2009) 896–905.
[39] A. Siewierska-Gorska et al. Association of five SNPs with human hair colour in the Polish population 68(2) (2017): p. 134–144.
[40] P.G. Hysi, et al., Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability, Nat. Genet. 50 (2018) 652–656.
[41] A. Visconti, et al., Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure, Nat. Commun. (2018) 9.
[42] F. Liu, et al., Meta-analysis of genome-wide association studies identifies 8 novel loci involved in shape variation of human head hair, Human Molecular Gentics 27 (3) (2018) 559–575.
[43] P. Sulem, et al., Two newly identified genetic determinants of pigmentation in Europeans, Nat. Genet. 40 (2008) 835–837.
[44] P.M. Schneider, et al., The use of forensic DNA phenotyping in predicting appearance and biogeographic ancestry, Dtsch. Arztebl. Int. 51-52 (51-52) (2019) 873–880.
[45] O. Lao, et al., Correlation between genetic and geographic structure in Europe, Curr. Biol. 18 (16) (2008) 1241–1248.