Eye color and the prediction of complex phenotypes from genotypes

Fan Liu[1], Kate van Duijn[1], Johannes R. Vingerling[2,3], Albert Hofman[2], André G. Uitterlinden[2,4], A. Cecile J.W. Janssens[2] and Manfred Kayser[1,*]

Predicting complex human phenotypes from genotypes has recently gained tremendous interest in the emerging field of consumer genomics, particularly in light of attempting personalized medicine [1,2]. So far, however, this approach has not been shown to be accurate, thus limiting its practical applications [3,4]. Here, we used human eye (iris) color of Europeans as an empirical example to demonstrate that highly accurate genetic prediction of complex human phenotypes is feasible. Moreover, the six DNA markers we identified as major eye color predictors will be valuable in forensic studies.

Facilitated by recent genome-wide studies, single nucleotide polymorphisms (SNPs) in various genes have been identified that are unambiguously associated with human eye color variation in Europeans [5–7], demonstrating that eye color is a genetically complex phenotype. Thus, eye color may be used to exemplify the feasibility of accurate genetic prediction of complex human phenotypes. Recent attempts in predicting eye color have obtained promising results using SNPs in *OCA2* [15], or in combination with *HERC2* [5], or additionally in *SLC24A4* and *TYR* [6]. However, a number of genetic variants with strong eye color association were not used in these previous prediction analyses; most of them were only identified in parallel or later studies [7–10]. To investigate the power of DNA-based eye color prediction, we genotyped 37 SNPs from eight genes [5–15], representing all currently known genetic variants with statistically significant eye color association (Supplemental Data), in a large population sample of 6168 Dutch Europeans from the Rotterdam Study [16]. 67.6% of the population sample had blue eyes, 22.8% brown eyes and 9.6% neither blue nor brown and categorized as intermediate color. We performed prediction analyses with several models and parameters. Population characteristics, phenotype collection, SNP ascertainment, genotyping methods and details of prediction models and parameters are described in the Supplemental Data. All SNPs genotyped were significantly associated ($p < 0.01$) with eye color variation (Supplemental Data), except one in the *ASIP* gene (but, see below).

A prediction model based on multinomial logistic regression constructed in the model-building set ($n = 3804$, 61.7%) using 24 SNPs from eight genes revealed excellent accuracy for predicting blue and brown eye color in the model-verification set ($n = 2364$, 38.3%) based on five parameters (Table 1). 13 SNPs were removed because of strong linkage disequilibrium with other markers in this set (Supplemental Data). Considering the area under the receiver characteristic operating curves (AUC) as an overall measure for prediction accuracy, whereby a completely accurate prediction is obtained at an AUC of 1, we obtained very high values for brown eyes at 0.93 and for blue eyes at 0.91. The prediction of intermediate color was less accurate with an AUC of 0.73. Predicting eye color using four alternative models yielded similar results (Supplemental Data). The lower prediction accuracy for intermediate eye color may be explained by unidentified associated SNPs as well as imprecise phenotype characterization; future investigations with more information on subtle phenotype characterization are warranted.

Furthermore, to assess the contribution of each SNP to the prediction accuracy of eye color, we measured AUC in a step-wise manner by iteratively excluding individual SNPs from the multinomial logistic regression model. Six SNPs from six genes (*HERC2* rs12913832, *OCA2* rs1800407, *SLC24A4* rs12896399, *SLC45A2* rs16891982, *TYR* rs1393350, and *IRF4* rs12203592) were revealed as major genetic predictors of eye colour with an overall AUC of 0.93 for brown, 0.91 for blue, and 0.72 for intermediate colored eyes (Figure 1). Nine additional SNPs (from *TYRP1*, *OCA2*, *HERC2*, and *ASIP*; Supplemental Data) had only minimal additive effects (Figure 1). The remaining nine SNPs had no additive value to the predictive accuracy (Figure 1; Supplemental Data); although they all were significantly associated with eye color in the single-SNP analysis, their effects were most likely being covered by other markers from the same genes included in the set of 15 SNPs. The prediction accuracies presented here were improved considerably compared to our previous attempt using three SNPs in *OCA2* and *HERC2* (e.g. AUC = 0.82 for brown eyes) [5], or compared to another prediction analysis [6] based on four SNPs in *OCA2*, *HERC2*, *SLC24A4*, and *TYR* that applied different methodology but estimating AUC in the Rotterdam Study using these four SNPs gave 0.83 for brown eyes.

The genetic prediction values obtained here for blue and brown eyes in Europeans represent the highest accuracies revealed so far in genetic prediction of human complex phenotypes. We thus demonstrated

Table 1. DNA-based prediction of human eye (iris) color based on multinomial logistic regression using 24 eye-color associated single nucleotide polymorphisms in Dutch Europeans of the Rotterdam Study.

	Blue	Intermediate	Brown
AUC	0.91	0.73	0.93
Sensitivity[1]	93.4	1.1	88.4
Specificity[1]	77.1	99.6	88.0
PPV[1]	89.8	25.0	67.1
NPV[1]	84.4	90.0	96.5

[1]Calculated from three two-by-two contingency tables of predicted and observed color types, where the predicted eye color type was obtained as the eye color with the highest predicted probability based on the multinomial logistic regression model. (AUC, area under the receiver operating characteristic (ROC) curves; PPV, positive predictive value; NPV, negative predictive value.)

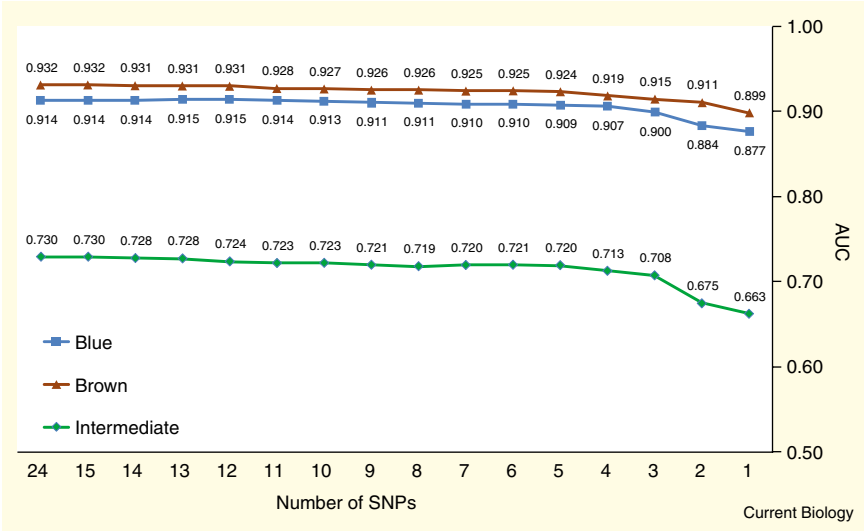

Figure 1. Contribution of 24 SNPs to the prediction accuracy of human eye (iris) color in Dutch Europeans of the Rotterdam Study.

Prediction performance measured by AUC for the model based on multinomial logistic regression (Y-axis) was plotted against the number of SNPs included in the model (X-axis). For each step, the lowest contributor in the model-building set (n = 3804) was excluded from the model; the model was rebuilt and used to predict eye color in the model-verification set (n = 2364).

that highly accurate DNA-based prediction of complex human phenotypes is feasible if strong genetic variants are implicated. Our findings of statistically significant eye color association of several genes, together with the high predictive value of SNPs therein, underline the importance of these genes in determining human iris color variation. Additionally, we provide a small set of DNA markers that are expected to serve as reliable biological evidence in suspectless forensic cases potentially allowing the police to concentrate investigations for tracing unknown persons of European descent according to DNA-predicted eye color. However, predicting with high accuracy the European descent of an unknown person using ancestry-sensitive DNA markers, as a prerequisite for a meaningful interpretation of the proposed forensic eye color prediction test, remains a challenging task [17,18].

Supplemental Data

Supplemental data are available at http://www.current-biology.com/supplemental/S0960-9822(09)00597-1.

Acknowledgments

We thank the participants of the Rotterdam Study, the local healthcare centres and the municipalities for making this study possible. We thank M. Verbiest for help with Sequenom genotyping as well as K. Ballantyne for useful comments on the manuscript, and acknowledge the work of F. Rivadeneira, M. Moorhouse, P. Arp and M. Jhamai, who all established the Rotterdam Study genome-wide SNP database. A.C.J.W.J. was supported by a VIDI grant of the Netherlands Organization for Scientific Research (NWO). This study was supported by funds from the Netherlands Forensic Institute and received additional support by a grant from the Netherlands Genomics Initiative/ Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands. Genome-wide genotyping of the Rotterdam Study was supported by the Netherlands Organisation for Scientific Research (NWO) (175.010.2005.011, 911-03-012).

References

1. Janssens, A.C., and van Duijn, C.M. (2008). Genome-based prediction of common diseases: advances and prospects. Hum. Mol. Genet. *17*, R166–R173.
2. Brand, A., Brand, H., and Schulte in den Baumen, T. (2008). The impact of genetics and genomics on public health. Eur. J. Hum. Genet. *16*, 5–13.
3. Janssens, A.C., Gwinn, M., Bradley, L.A., Oostra, B.A., van Duijn, C.M., and Khoury, M.J. (2008). A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. Am. J. Hum. Genet. *82*, 593–599.
4. Haga, S.B., Khoury, M.J., and Burke, W. (2003). Genomic profiling to promote a healthy lifestyle: not ready for prime time. Nat. Genet. *34*, 347–350.
5. Kayser, M., Liu, F., Janssens, A.C., Rivadeneira, F., Lao, O., van Duijn, K., Vermeulen, M., Arp, P., Jhamai, M.M., van Ijcken, W.F., *et al.* (2008). Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am. J. Hum. Genet. *82*, 411–423.
6. Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Magnusson, K.P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., *et al.* (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat. Genet. *39*, 1443–1452.
7. Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Jakobsdottir, M., Steinberg, S., Gudjonsson, S.A., Palsson, A., Thorleifsson, G., *et al.* (2008). Two newly identified genetic determinants of pigmentation in Europeans. Nat. Genet. *40*, 835–837.
8. Sturm, R.A., Duffy, D.L., Zhao, Z.Z., Leite, F.P., Stark, M.S., Hayward, N.K., Martin, N.G., and Montgomery, G.W. (2008). A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am. J. Hum. Genet. *82*, 424–431.
9. Eiberg, H., Troelsen, J., Nielsen, M., Mikkelsen, A., Mengel-From, J., Kjaer, K.W., and Hansen, L. (2008). Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Hum. Genet. *123*, 177–187.
10. Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., Hankinson, S.E., Hu, F.B., Duffy, D.L., Zhao, Z.Z., *et al.* (2008). A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet. *4*, e1000074.
11. Frudakis, T., Thomas, M., Gaskin, Z., Venkateswarlu, K., Chandra, K.S., Ginjupalli, S., Gunturi, S., Natrajan, S., Ponnuswamy, V.K., and Ponnuswamy, K.N. (2003). Sequences associated with human iris pigmentation. Genetics *165*, 2071–2083.
12. Graf, J., Hodgson, R., and van Daal, A. (2005). Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. Hum. Mutat. *25*, 278–284.
13. Kanetsky, P.A., Swoyer, J., Panossian, S., Holmes, R., Guerry, D., and Rebbeck, T.R. (2002). A polymorphism in the agouti signaling protein gene is associated with human pigmentation. Am. J. Hum. Genet. *70*, 770–775.
14. Duffy, D.L., Montgomery, G.W., Chen, W., Zhao, Z.Z., Le, L., James, M.R., Hayward, N.K., Martin, N.G., and Sturm, R.A. (2007). A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. Am. J. Hum. Genet. *80*, 241–252.
15. Frudakis, T., Terravainen, T., and Thomas, M. (2007). Multilocus OCA2 genotypes specify human iris colors. Hum. Genet. *122*, 311–326.
16. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R., and Witteman, J.C. (2007). The Rotterdam Study: objectives and design update. Eur. J. Epidemiol. *22*, 819–829.
17. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., *et al.* (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science *319*, 1100–1104.
18. Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R., *et al.* (2008). Genotype, haplotype and copy-number variation in worldwide human populations. Nature *451*, 998–1003.

[1]Department of Forensic Molecular Biology, [2]Department of Epidemiology, [3]Department of Ophthalmology, [4]Genetic Laboratory of the Department of Internal Medicine, Erasmus University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands.
*E-mail: m.kayser@erasmusmc.nl