

ORIGINAL ARTICLE

Gene–gene interactions contribute to eye colour variation in humans

Ewelina Pośpiech¹, Jolanta Draus-Barini¹, Tomasz Kupiec¹, Anna Wojas-Pelc² and Wojciech Branicki^{1,3}

Prediction of phenotypes from genetic data is considered to be the first practical application of data gained from association studies, with potential importance for medicine and the forensic sciences. Multiple genes and polymorphisms have been found to be associated with variation in human pigmentation. Their analysis enables prediction of blue and brown eye colour with a reasonably high accuracy. More accurate prediction, especially in the case of intermediate eye colours, may require better understanding of gene–gene interactions affecting this polygenic trait. Using multifactor dimensionality reduction and logistic regression methods, a study of gene–gene interactions was conducted based on variation in 11 known pigmentation genes examined in a cohort of 718 individuals of European descent. The study revealed significant interactions of a redundant character between the *HERC2* and *OCA2* genes affecting determination of hazel eye colour and between *HERC2* and *SLC24A4* affecting determination of blue eye colour. Our research indicates interactive effects of a synergistic character between *HERC2* and *OCA2*, and also provides evidence for a novel strong synergistic interaction between *HERC2* and *TYRP1*, both affecting determination of green eye colour.

Journal of Human Genetics (2011) 56, 447–455; doi:10.1038/jhgc.2011.38; published online 7 April 2011

Keywords: eye colour; gene–gene interactions; multifactor dimensionality reduction; pigmentation genes; *HERC2*; *OCA2*; *TYRP1*; *SLC24A4*

INTRODUCTION

The significance of interactions between genes for their phenotypic effect has been known since Bateson¹ first described the masking effect of an allele in one gene on an allele in another gene. Soon afterwards, epistasis was defined by Fisher² as a deviation from additivity in the effect of variants at various loci with respect to their contribution to a phenotypic trait.³ Presently, epistasis or genetic interaction is understood as a departure from a linear model that describes how various factors predict a final phenotypic effect.⁴ Prediction of phenotypes from genetic data is considered to be important in both medical and forensic examinations. In particular, it can improve medical care by more accurate prognosis and prevention^{5,6} and speed up forensic investigations by providing information concerning externally visible traits in criminal non-suspect cases or cases concerning identification of human remains.^{7,8} Variation in pigmentation phenotype is particularly high in humans, especially in populations of European descent, and thus prediction of this phenotype would seem to provide information of particular importance for forensic examinations. Human pigmentation is considered to be mostly under genetic control. It has been assessed; for example, that hair colour is 92% controlled by the genetic component.⁹ Several genes were associated with various pigment traits long before the genome-wide association study era.^{10–16} However, it is important to note that—unlike in the case of testing for many complex diseases¹⁷—

genome-wide association study has been particularly effective in association testing for human pigmentation and has confirmed already known, and revealed multiple new polymorphisms and loci involved in determination of human pigmentation.^{18–21} This has been particularly striking in the case of variation in eye colour, which is assessed to be 50% explained by known polymorphisms.²² This has led to the first predictive tests of extreme eye colours (blue or brown) in forensic sciences with relatively high accuracies.^{23,24} Importantly, the intermediate iris colours (green or hazel) are still difficult to predict.²⁵ It is possible that detection of nonlinear effects between known predictors and their implementation in prediction models may be important for better prediction of these intermediate states, and may lead to a further increase in prediction accuracy of extreme phenotypes.

Some reports have concluded that interactions between genes may be important in determination of the pigmentation phenotype on the basis of results suggesting epistatic effects between *MC1R* and *HERC2*, *HERC2* and *OCA2*, *HERC2* and *SLC24A4*, *OCA2* and *SLC45A2* and between *HERC2* and *IRF4*.^{22,26} Therefore, we further investigated this issue and using multifactor dimensionality reduction (MDR) and logistic regression methods undertook a study of interaction effects, examining 24 polymorphisms in 11 known genes involved in pigmentation in humans in a population sample of 718 individuals of European descent.

¹Section of Forensic Genetics, Institute of Forensic Research, Kraków, Poland; ²Department of Dermatology, Collegium Medicum of the Jagiellonian University, Kraków, Poland and ³Department of Genetics and Evolution, Institute of Zoology, Jagiellonian University, Kraków, Poland

Correspondence: Dr W Branicki, Section of Forensic Genetics, Institute of Forensic Research, Westerplatte 9, Krakow, 31-033, Poland.

E-mail: wbranicki@ies.krakow.pl

Received 30 December 2010; revised 25 February 2011; accepted 16 March 2011; published online 7 April 2011

MATERIALS AND METHODS

Study subjects

The study was approved by the Ethics Committee of the Jagiellonian University in Krakow (KBET/17/B/2005) and the Commission on Bioethics of the Regional Board of Medical Doctors in Krakow (48 KBL/OIL/2008). A total of 718 specimens were obtained from unrelated males and females (between 18 and 85 years old) of European descent (Polish nationality) by a specialist from the Department of Dermatology of the Jagiellonian University Hospital. Phenotyping was conducted under professional observation. Eye colour was classified into four categories; that is, blue, green, hazel and brown.

DNA extraction and quantification

DNA was extracted with the NucleoSpin Tissue extraction kit (Macherey-Nagel GmbH & KG, Duren, Germany), using a protocol intended for buccal swabs. Briefly, a buccal swab was placed into a 2-ml tube and treated with 400 µl of TE buffer, 400 µl of lysis buffer B3 and 25 µl of proteinase K for 20 min in a water bath at 70 °C. Next, 400 µl of 96% ethanol was added to the lysis mixture and 600 µl of total content was put onto the NucleoSpin Tissue Column, centrifuged for 1 min at 11 000 RCF and washed with BW and B5 buffers. Finally, DNA was eluted into a new 1.5-ml tube with warmed (~70 °C) elution buffer BE. The DNA concentration was measured using a Quant-iT Pico Green dsDNA Assay Kit (Invitrogen, Eugene, OR, USA) and Fluoroskan Ascent FL (LabSystems, Helsinki, Finland).

Single nucleotide polymorphism Genotyping

The 24 (SNPs) located in 11 genes, which have been shown to be significantly associated with human pigmentation^{10,13,14,16,18,19,21–25,27–32} were genotyped in two multiplex minisequencing assays. The list of the studied polymorphisms is given in Supplementary Table 1. Multiplex 1 (M1) included 13 SNP positions from 10 genes: *SLC45A2*, *IRF4*, *TYRP1*, *TPCN2*, *TYR*, *KITLG*, *SLC24A4*, *OCA2*, *HERC2* and *ASIP*, and multiplex 2 (M2) included 11 SNPs, all located in the *MC1R* gene. In the case of M1, 13 DNA fragments were amplified simultaneously using PCR primers given in Supplementary Table 2 and the Qiagen Multiplex PCR kit (Qiagen, Hilden, Germany). The PCR reaction consisted of 2.5 µl Qiagen Multiplex PCR mixture, 0.5 µl Q solution, 0.5 µl primer premix (final concentrations are shown in Supplementary Table 2), and 1.5 µl (~1–10 ng) of template DNA. The amplification reaction was carried out in the GenAmp 9700 thermocycler (Applied Biosystems, Foster City, CA, USA). The applied temperature profile was as follows: 95 °C/15 min, (94 °C/30 s, 58 °C/90 s, 72 °C/90 s) × 32, 72 °C/10 min. In the case of M2, the complete *MC1R* exon was amplified using primers described by Kanetsky *et al.*²⁹ The PCR reaction consisted of 2.5 µl Qiagen Multiplex PCR mixture, 0.5 µl of primer premix (the final concentrations are shown in Supplementary Table 2) and 2 µl of template DNA. The amplification reaction was carried out at the following temperature profile: 95 °C/1 min, (94 °C/10 s, 64 °C/30 s, 72 °C/40 s) × 34, 72 °C/10 min. The PCR products were always purified with a mixture of Exonuclease I (ExoI) and Alkaline Phosphatase (SAP) enzymes (Fermentas, Vilnius, Lithuania) and subjected to minisequencing reactions using a SNaPshot multiplex kit (Applied Biosystems). A single reaction consisted of 0.5 µl SNaPshot mix, 0.5 µl of extension primer premix, 1 µl of purified PCR product and 3 µl of Dnase-free water. Details of extension primer sequences and their final concentrations are shown in Supplementary Table 3. The applied temperature profile was as follows: (96 °C/10 s, 50 °C/5 s, 60 °C/30 s) × 26. The products of extension reactions were purified with SAP enzyme (Fermentas) and finally analysed on an ABI 3100 Avant Genetic Analyser (Applied Biosystems).

Statistical calculations

Population analyses and haplotype evaluation. The genetic data obtained for each SNP were tested for agreement with Hardy–Weinberg expectations and degree of linkage disequilibrium between analysed polymorphic sites. Whole-population analyses were conducted using Arlequin version 3.1 software (<http://cmpg.unibe.ch/software/arlequin3>). Haplotypes were reconstructed and evaluated using the statistical methodology developed in Epstein and Satten³³ for all the SNPs located on the same chromosome except of the *MC1R* polymorphisms, which all may have functional effects. The method relies on a likelihood approach that uses the Expectation-Conditional-Maximization

(ECM) algorithm and was implemented in the Chaplin version 1.2.2 computer software (<http://genetics.emory.edu/labs/epstein/software/chaplin/index.html>). This approach also enables evaluation of haplotype association with particular phenotypes. The association of individual haplotypes (only those with a frequency exceeding 5% were considered) with eye colour (blue vs non-blue test) was estimated with the Wald test, and different genetic models were evaluated using the Akaike Information Criterion (AIC).

Multifactor dimensionality reduction analysis. Gene–gene interactions between analysed SNP positions were evaluated using the multifactor dimensionality reduction approach (MDR; software ver. 2.0 beta 8.1 (www.epistasis.org)). The method relies on reduction of high-dimensional data to one-dimensional variables by pooling genotypes into two groups: ‘high-risk’ and ‘low-risk’ on the basis of the ratio of cases to controls. In the procedure, the analysed data are divided into 10 equal parts to perform 10-fold cross validation (CV). The model is created on 9/10 of the data (training set) and then evaluated for its prediction ability using the remaining 1/10 (testing set). The procedure is repeated for each possible 9/10 of the data (that is, 10 different possible intervals), and then CV consistency is calculated as the number of times that a particular model is chosen as the best one for each of the 10 intervals. The second important measure is balanced accuracy (BA), which shows how many instances are correctly classified using the MDR model and is calculated as (Sensitivity+Specificity)/2. BA is especially useful when the data set is not balanced (that is, different numbers of cases and controls) and gives equal weight to each class.

In this study, 24 SNP positions were tested for potential gene–gene interactions explaining observed variation in eye colour in humans. Various eye pigmentation classifications were tested, in particular separate models were built for blue, green, hazel and brown eye colours. The genotypes for SNP positions from M1 were categorized in an additive manner. In case of the rs12913832 position additionally recessive categorisation was applied and this was chosen for reporting as more appropriate and giving clearer results, which is also supported by our previous study.²⁶ In the case of the *MC1R* gene, polymorphisms were divided into two groups: the high penetrance variants coded as ‘R’ (N29insA, V84E (rs1805006), R151C (rs1805007), R142H (rs11547464), R160W (rs1805008), D294H (rs1805009) and Y152OCH), and low penetrance variants coded as ‘r’ (V60L (rs1805005), V92M (rs2228479), I155T (rs1110400) and R163Q (rs885479)). For high penetrance variants (‘R’), three states were considered, assuming the existence of major function mutations; that is, 0=no ‘R’ variant carriers, 1=one ‘R’ variant carrier and 2=two ‘R’ variant carriers. The same approach was used to categorise low penetrance variants (0=no ‘r’ variant carriers, 1=one ‘r’ variant carrier and 2=two ‘r’ variant carriers).

Before actual MDR analysis, pre-process filtering using a ReliefF filter on the data set was conducted to reduce the chances of false positives. ReliefF statistic is a method that qualifies the relevance of variables and the dependency between them for predicting of phenotypic trait^{34–36} and was applied to limit the number of variables that were considered in further analysis. In this study, the top five attributes were selected. After filtering, exact analyses were conducted for all eye colour classifications. In all cases, one to three attribute combinations were considered and the model with the highest BA and CV consistency was chosen as the best for explaining eye-colour variation in humans. The statistical significance of models was evaluated by a 1000-fold permutation test using MDR Permutation Testing Module 1.0 beta 2 software (www.epistasis.org).

Study of interaction information using the entropy-based approach. To interpret the obtained results, MDR software provides dendrogram graphs to view the presence of interactions between attributes and the strength of those interactions. The dendrograms are built using hierarchical cluster analysis. To explain the nature of particular interactions, the concept of information gain based on entropy measure proposed by Jakulin and Bratko is used.^{37,38} The entropy-based approach is a non-parametric method for measuring ‘uncertainty’ and enables estimation of the benefit in information gain from considering two attributes together over considering them apart. Negative entropy between two attributes suggests redundancy (the interaction provides redundant information), whereas positive entropy indicates synergy between attributes (the interaction between attributes delivers more information than the sum of the individual attributes).

Logistic regression analysis. Binary logistic regression (eye colour classified as blue versus non-blue, green versus non-green, hazel versus non-hazel) was used to calculate odds ratios (OR) with 95% CIs and respective *P*-values for interaction models predicted by MDR to confirm obtained results, as well as for testing potential main effects of particular factors included in the detected interaction models. The same manner of categorisation of genotypes for particular SNPs was used as for MDR analysis. Confirmation of significance of the revealed gene–gene interactions was achieved using multinomial logistic regression where eye colour was defined into four categories; that is, blue, green, hazel and brown. Calculations were carried out using PASW statistics version 17 software (SPSS Inc., Chicago, IL, USA).

RESULTS

Population analyses and haplotype evaluation

Among the studied population, 367 individuals (51.1%) had blue eye colour, 96 (13.4%) green, 149 (20.8%) hazel and 106 (14.8%) had brown eye colour. After Bonferroni's correction for multiple testing ($P > 0.0020833$), all the 24 analysed SNP positions were found to be in Hardy–Weinberg equilibrium. As expected, the LD testing revealed linkage disequilibrium (valid after Bonferroni's correction for multiple testing, $P < 0.0020833$) between loci located closely on a single chromosome; that is, rs1800407 in *OCA2* and rs12913832 in *HERC2*, rs49111414 and rs1015362—both in *ASIP*, rs683 and rs1408799—both in *TYRP1*, rs3829241 and rs35264875—both in *TPCN2* and polymorphisms within the *MC1R* gene (data not present). Table 1 shows the haplotypes for which significant associations with eye colour were found based on Wald statistics. Association was tested for eye colour defined as blue versus rest. All the haplotypes reconstructed for the pair of SNPs *OCA2-HERC2* were found to be significantly associated with eye colour. One haplotype; that is, TC for the pair of SNPs located on chromosome 9; that is, rs1408799 and rs683 and three haplotypes; that is, AGA, AAG, TGG for the SNPs located on chromosome 11; that is, rs35264875, rs3829241, rs1393350 also revealed significant association with eye colour.

MDR, interaction entropy and logistic regression analyses

The 24 analysed SNP positions, which had previously been shown to be significantly associated with differences in pigmentation among humans, were tested for potential gene–gene interactions explaining variation in eye colour using the MDR method. Table 2 shows selected by ReliefF filtering procedure the top five SNP positions, for all eye colour classifications, which were further considered in MDR analysis. In Table 3, the best models (consisting of up to three-factors) predicted by MDR are presented together with their CV consistency, BA and *P*-values. Table 4 shows interaction entropy analysis based on MDR to explain the contribution of particular factors to information gain and to explain the nature of epistasis. Binary logistic regression was used for confirmation of the results obtained with MDR (Table 5).

Model—blue versus rest

MDR analysis conducted for eye colour classified as blue versus non-blue gave the two best models with the same values: CVC=10/10, BA=0.8369 and *P*-values < 0.001 ; that is, a one factor model formed exclusively by rs12913832 in *HERC2* and a two-factor model assuming interaction between rs12913832 in *HERC2* and rs12896399 in *SLC24A4*. Analysis of the dendrogram graph (Figure 1a) confirmed the validity of the latter model assuming strong interaction between these two factors; the blue line between them suggests a negative information gain; that is, redundancy. Analysis of entropy reveals that the largest main independent effect is attributed to the rs12913832 position. This means that it removes 37.14% of 'uncertainty' in

Table 1 Haplotypes associated with eye color defined as blue versus rest

Haplotype	Cases	Controls	Wald test (P-value)	AIC	Best genetic model
<i>Chromosome 15: rs12913832, rs1800407</i>					
CG	338	197	14.3365 (0.0000)	1463.92	Multiplicative
CA	10	2	2.6192 (0.0088)	1726.27	Dominant
TG	12	118	12.4241 (0.0000)	1474.93	Multiplicative
TA	5	32	5.0930 (0.0000)	1688.85	Multiplicative
<i>Chromosome 9: rs1408799, rs683</i>					
TC	82	95	2.0059 (0.0449)	24158.18	Recessive
<i>Chromosome 11: rs35264875, rs3829241, rs1393350</i>					
AGA	60	50	4.0102 (0.0003)	3390.56	General
AAG	40	48	8.7213 (0.0000)	3303.61	General
TGG	60	63	2.6455 (0.0082)	3397.50	Recessive

Abbreviation: AIC, Akaike information criterion.

Table 2 The top five attributes selected by ReliefF filtering procedure and considered in MDR analysis

Eye color classification	The top five SNP positions selected by ReliefF filter
Blue versus non-blue	rs12913832, rs12896399, rs1393350, rs49111414, MC1R_r
Green versus non-green	rs12913832, rs1800407, rs12203592, rs1408799, MC1R_R
Hazel versus non-hazel	rs12913832, rs1800407, rs1393350, rs3829241, MC1R_r
Brown versus non-brown	rs12913832, rs1800407, rs1393350, rs12821256, MC1R_r

Abbreviations: MDR, multifactor dimensionality reduction; SNP, single nucleotide polymorphism.

prediction of eye colour defined as blue versus non-blue, whereas rs12896399 removes 1.10% (Table 4). Interestingly, when considering these two factors together, the additional entropy explained by interaction was obtained at a level of -1.10% , which means that 1.10% of 'uncertainty', which is explained by rs12896399 is also explained by rs12913832 (rs12896399 provides redundant information). This suggests a masking effect of rs12913832 on rs12896399. Figure 2a shows that carriers possessing CC, CA or AA genotypes in rs12896399 with CC genotype in rs12913832 are all classified into the group of 'high-risk' of having blue eye colour. Detailed analysis indicates that the chances of having blue eye colour rise with A allele in rs12896399 (which is seen based on the ratio of cases to controls), and the highest chances of having blue eye colour are with AA genotype in rs12896399, but only on condition of possessing CC genotype in rs12913832. The allele T in rs12913832 in *HERC2* masks this effect and, independently of the genotype in rs12896399, individuals possessing the T allele in rs12913832 are classified into the group of 'low-risk' of having blue eye colour (with the ratio of cases to controls on the 0.1–0.2 level for all genotypes in rs12896399). Logistic regression analysis applied to these two factors confirmed a strong interaction effect with OR=7.592 and *P*-value=0.000 (Table 5).

Model—green versus rest

Green versus non-green eye colour is, according to MDR, best predicted by a three-factor model consisting of rs12913832 in

Table 3 Results of the multifactor dimensionality reduction analysis

Eye colour classification	Best candidate model (for up to three factor combinations)	Testing balanced accuracy	Cross validation consistency	Permutation testing P-value
Blue versus non-blue	rs12913832	0.8369	10/10	<0.001
	rs12913832; rs12896399	0.8369	10/10	<0.001
Green versus non-green	rs12913832; rs12896399; rs1393350	0.8245	6/10	<0.001
	rs12913832	0.4621	6/10	0.9510–0.9520
	rs12913832; rs1408799	0.5821	8/10	0.0310–0.0320
Hazel versus non-hazel	rs12913832; rs1408799; rs1800407	0.6184	10/10	0.001–0.002
	rs12913832	0.7842	10/10	<0.001
	rs12913832; rs1800407	0.7860	10/10	<0.001
Brown versus non-brown	rs12913832; rs1800407; rs1393350	0.7851	8/10	<0.001
	rs12913832	0.8391	10/10	<0.001
	rs12913832; rs12821256	0.8391	9/10	<0.001
	rs12913832; rs1393350; MC1R_r	0.8220	5/10	<0.001

Best models are marked with bold.

Table 4 Results of interaction entropy analysis

Interaction model	The percentage of entropy explained by a single attribute	Additional entropy removed by interaction	Sum of entropy of the individual attributes	Sum of entropy comprising interaction	Nature of epistasis
<i>Blue versus non-blue</i>					
rs12913832	37.14%				
rs12896399	1.10%	−1.10%	38.24%	37.14%	Redundant
<i>Green versus non-green</i>					
rs12913832	0.28%				
rs1408799	0.23%	1.23%	0.51%	1.74%	Synergistic
rs12913832	0.28%				
rs1800407	0.47%	0.82%	0.75%	1.57%	Synergistic
<i>Hazel versus non-hazel</i>					
rs12913832	16.58%				
rs1800407	1.09%	−0.88%	17.67%	16.79%	Redundant

Table 5 Results of logistic regression analysis for interaction models predicted by MDR

Eye colour classification	Interaction model	OR (95% CI) for independent attribute	P-value for independent attribute	OR (95% CI) for interaction model	P-value for interaction model
Blue versus non-blue	rs12913832	32.129 (20.882–49.434)	0.000		
	rs12896399	1.385 (1.120–1.711)	0.003	7.592 (5.450–10.574)	0.000
Green versus non-green	rs12913832	—	0.095		
	rs1408799	—	0.148	1.862 (1.345–2.578)	0.000
	rs12913832	—	0.095		
	rs1800407	—	0.057	1.989 (1.141–3.465)	0.015
Hazel versus non-hazel	rs12913832	15.662 (9.372–26.172)	0.000		
	rs1800407	1.959 (1.250–3.070)	0.003	2.790 (1.729–4.504)	0.000

Abbreviations: MDR, multifactor dimensionality reduction; OR, odds ratio. Significant results are marked with bold.

HERC2, rs1408799 in *TYRP1* and rs1800407 in *OCA2* (CVC=10/10, BA=0.6184, $P=0.0001$ – 0.002). Interestingly, dendrogram graph analysis (Figure 1b) revealed positive interaction effects among these three factors, with the strongest effect between rs12913832 and rs1408799. The entropy-based approach points to two interaction pairs; that is, rs12913832 and rs1408799 and also rs12913832 and rs1800407. The rs12913832 position removes 0.28% of ‘uncertainty’ in green eye

colour prediction, rs1408799 removes 0.23% and the interaction between them eliminates an additional 1.23%, which suggests the benefit of considering these two factors together over treating them separately. This result is supported by logistic regression analysis (Table 5), which revealed that rs12913832 and rs1408799 as independent factors are not significantly associated with green eye colour; that is, have no main effect and only considering them together as an

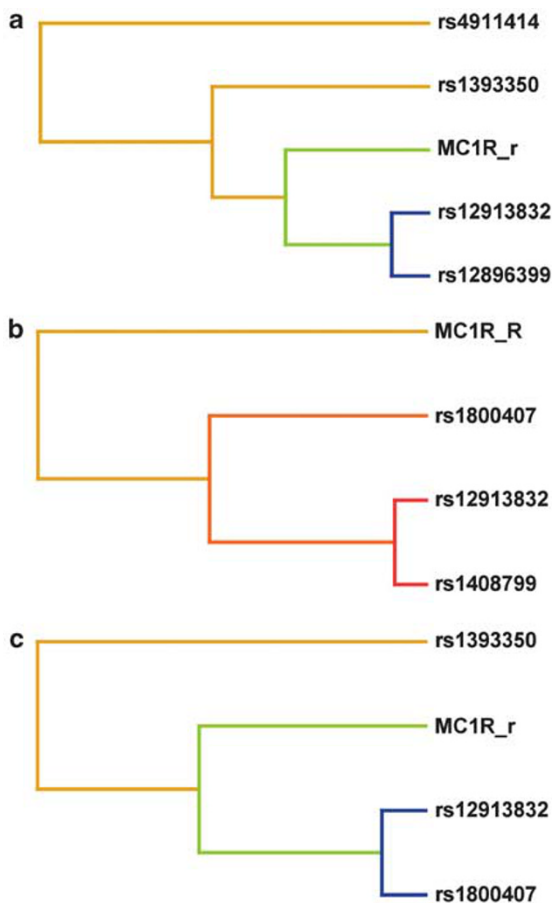


Figure 1 Interaction dendrograms for blue versus non-blue (a), green versus non-green (b) and hazel versus non-hazel (c) eye colour classification. A red line represents a high degree of synergy, orange—lesser degree of synergy, gold—independence or additivity, blue and green—the redundancy with the highest degree represented by blue line. The more the line connecting two attributes is moved to the right side the stronger is the interaction.

interaction model is statistically significant with $OR=1.862$ and $P\text{-value}=0.000$. The second detected positive interaction effect between rs12913832 and rs1800407 is weaker than the previous effect, on the basis of dendrogram graph analysis (Figure 1b). The rs1800407 position explains 0.47% of the entropy, the interaction brings an additional 0.82% and the sum of the entropy comprising interaction explains 1.57% of ‘uncertainty’ in predicting green eye colour. Notably, this interaction effect is also confirmed by logistic regression analysis with $OR=1.989$ and $P\text{-value}=0.015$ (Table 5). At the next stage, we carried out detailed analysis of distributions of genotypes for both interaction models. In the case of the model including rs12913832 and rs1408799, the chances of having green eye colour increase in a nonlinear manner when possessing at least one T allele in rs1408799, but only in the presence of the CC genotype in rs12913832, or when possessing the CC genotype in rs1408799, but only in the presence of the CT or TT genotype in rs12913832 (Figure 2b). In the case of the second interaction effect, the chances of having green eye colour increase when the GG genotype is present in rs1800407, provided that the CC genotype is present in rs12913832, or when possessing at least one T allele in rs12913832, but only on condition of having the GA genotype in rs1800407. As the AA genotype was not observed in rs1800407 in the studied population sample, it is impossible to comment on the effect of this state (Figure 2c).

Model—hazel versus rest

A strong interaction effect between rs12913832 and rs1800407, but of a different; that is, redundant, nature was also detected when eye colour was defined as hazel versus non-hazel (Figure 1c), with the following parameters of the model: $CVC=10/10$, $BA=0.7860$ and $P<0.001$ (Table 3). Entropy analysis showed that rs12913832 has, similarly to the blue versus non-blue model, the largest main independent effect with the ability to remove 16.58% of entropy, whereas the second factor with a smaller independent effect; that is, rs1800407, removes 1.09%. Additional entropy explained by interaction amounts to -0.88% . Genotype combinations associated with hazel and non-hazel eye colours (Figure 2d) suggest the masking effect of rs12913832 on 1800407. Individuals with GG and GA (the AA genotype was not observed) in rs1800407 are classified into the group of ‘low-risk’ of possessing hazel eye colour, but only in the case of having the CC genotype in rs12913832. Possessing at least one T allele in the rs12913832 position masks this effect and all individuals (independently of the genotype in rs1800407) are then classified into the group of ‘high-risk’ of having hazel eye colour. Logistic regression confirms the significance of this interaction with $OR=2.790$, $P\text{-value}=0.000$ (Table 5).

Model—brown versus rest

Brown eye colour according to MDR is mostly under the control of rs12913832 ($CVC=10/10$, $BA=0.8391$, $P<0.001$) and analysis of the dendrogram graph, as well as examination of entropy, did not reveal any significant interactions explaining brown eye colour (data not present).

DISCUSSION

Eye colour is a polygenic trait with multiple genes involved, but a major role is attributed to two genes, *HERC2* and *OCA2*, located on chromosome 15.^{16,26,32,39} Analysis of the *HERC2-OCA2* region enables prediction of extreme (blue or brown) eye colours from genetic data with a reasonably high accuracy and has already been applied in real cases concerning identification of human remains.^{40,41} Several other genes have been found to contribute to a lesser degree to the continuous variation in iris colour from light blue through intermediate (green and hazel) to dark brown, which is observed in humans.^{13,14,18,19,25} As the effect of multiple genes on determination of a phenotype may not be limited to their additive function,^{42,43} it is obvious that epistatic effects should also be taken into account when studying the genetics of a polygenic trait such as eye colour. Several methods have been proposed to study nonlinear effects of multiple genes on a phenotypic trait.⁴ It has been pointed out that traditional parametric methods, such as logistic regression, are problematic in detecting gene–gene interactions in a situation of data deficiency in high dimensions.⁴⁴ The MDR method constitutes an alternative to such methods, and is a non-parametric and model-free data mining approach developed to improve the power of detecting multilocus effects in epidemiological studies.^{44–50} In comparison with parametric methods, MDR is able to identify interactions in the absence of detectable main effects and in relatively small sample sizes.^{44,45} The next advantage of the MDR method is that it enables us to explain the nature of interactions through implemented entropy-based analysis. The examples of application of the MDR method include studies on such complex diseases as: sporadic breast cancer,⁴⁵ essential hypertension,⁵¹ asthma⁵² or systemic sclerosis.⁵³

In this study, by using MDR and logistic regression methods, interactions between known pigmentation genes, *SLC45A2*, *IRF4*, *TYRP1*, *TPCN2*, *TYR*, *KITLG*, *SLC24A4*, *OCA2*, *HERC2*, *ASIP* and

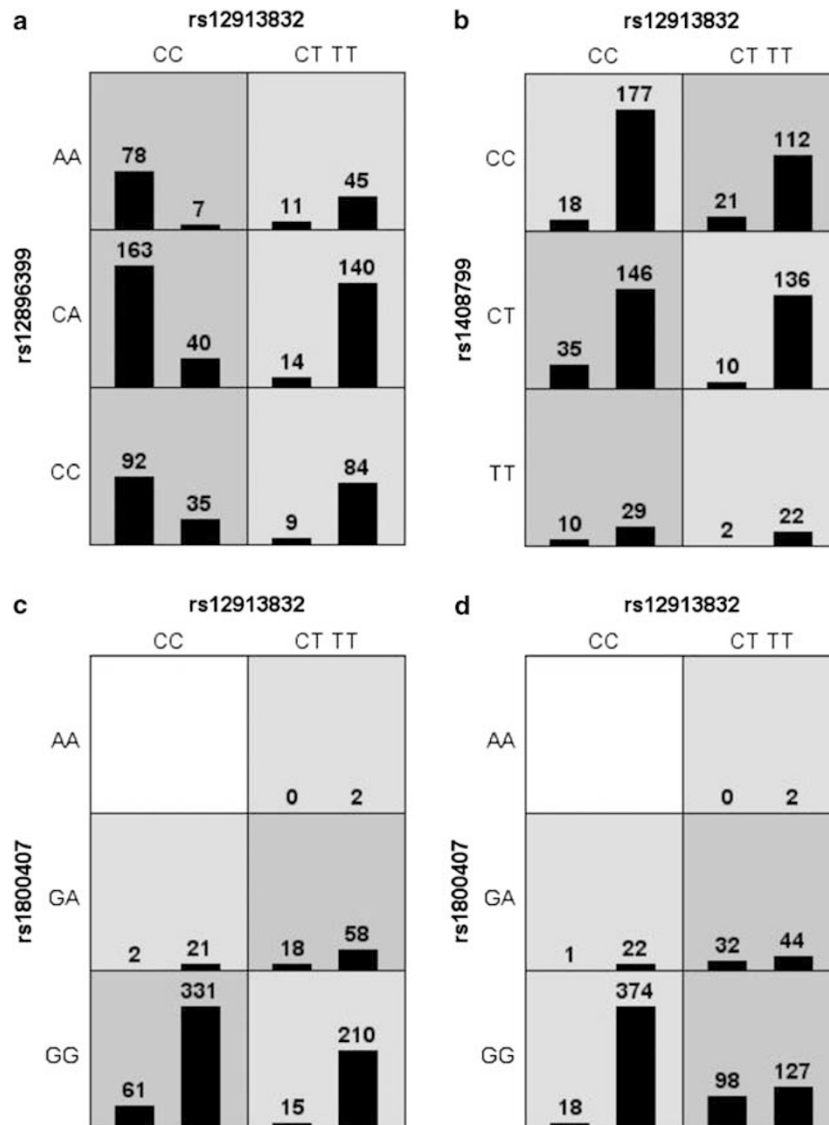


Figure 2 Summary of two-locus genotype combinations among attributes considered in interaction models for blue versus non-blue (a), green versus non-green (b, c) and hazel versus non-hazel (d) eye colour classification. The dark-grey cells represent the genotype combinations associated with 'high-risk' of having particular eye colour, whereas the genotype combinations in the light-grey cells are linked with 'low-risk'. The white cell means lack of data. Each cell is a representation of the number of instances of cases (left bar) and controls (right bar).

MC1R, were evaluated for their significance in eye colour determination in a cohort of 718 unrelated individuals from Poland. Except for rs3829241 and rs35264875—both in *TPCN2*, rs1015362 and rs4911414—both in *ASIP* and variation in *MC1R*, all the polymorphisms studied in this research have been reported to affect eye colour in humans.^{13,14,16,18,19,25,28,31,32} As variables without a main effect can still be involved in significant interactions determining polygenic traits, we also considered (in our study) pigmentation-related polymorphisms that have not been associated with eye colour, but have been correlated with other pigmentation traits; that is, skin and hair colour.^{18,19,21,54} Analysis conducted revealed three pairs of SNPs with significant interaction effects contributing to eye colour variation in the studied population. These loci are rs12913832 in *HERC2* and rs1800407 in *OCA2* (for hazel versus non-hazel and green versus non-green eye colour categorisation), rs12913832 in *HERC2* and rs12896399 in *SLC24A4* (for blue versus non-blue model),

rs12913832 in *HERC2* and rs1408799 in *TYRP1* (for the green versus non-green model). The obtained results confirmed the predominant role in eye colour inheritance of two genes located on chromosome 15; that is, *HERC2* and *OCA2*. The *OCA2* gene encodes a protein that is an integral part of the melanosomal membrane and is responsible for regulation of pH inside the melanosome,^{55–57} which, in consequence, has an influence on the activity of the enzyme tyrosinase, which has a crucial role in the synthesis of the pigment melanin. The important role of *OCA2* in eye colour determination was confirmed in many studies,^{11,13,28,31,58–61} and further investigations revealed that regulation of *OCA2* expression through the neighbouring *HERC2* gene might have a crucial role.^{16,20} A functional effect was assigned to the rs12913832 position located in a conservative segment of the intron 86 of the *HERC2* gene, which contains transcription factor-binding sites.¹⁶ Functional studies by Eiberg *et al.*³² indicated that two alleles in rs12913832 have different affinities and bind with different strength

to transcription factors, so a sequence around position rs12913832 may have the role of a transcription silencer. Further studies confirmed the highest significance of rs12913832 in eye colour determination and also showed that in spite of strong linkage between *HERC2* and *OCA2*, one position in the *OCA2* gene; that is, rs1800407, remains independently associated with eye colour and has a modulatory role for rs12913832 in *HERC2*.^{16,31} Our analyses showed that these two SNP positions were in strong LD and all the reconstructed haplotypes were found to be significantly associated with eye colour (Table 1). Moreover, both the SNPs were selected by ReliefF filtering in case of three from four eye colour models (Table 1). The ReliefF filter is capable of detecting dependencies between variables. It enables to select most important predictors considering not to remove the pairs of SNPs that are dependent,³⁶ so it doesn't have to remove SNPs that stays in LD because they can still reveal interaction effects. Analysis of results summarized in Table 2 shows that in our study the pair of SNPs in *HERC2* and *OCA2* was the only pair of polymorphisms that stay in LD and were selected by ReliefF filtering procedure and considered further in MDR analysis. It is also worth noting that in our study, the rs12913832 position in *HERC2* appeared in all tested eye colour classifications as the best one-factor model explaining variation in eye colour and, moreover, is a component of all revealed interaction models (Table 3). The epistatic effect between rs12913832 in *HERC2* and rs1800407 in *OCA2* was detected for green and hazel eye colour models. In the case of hazel versus non-hazel classification, the interaction effect had a redundant character; that is, the interaction between those variables with main independent effects results in a loss of information, which is caused by the masking effect of rs12913832 on 1800407. In the case of green versus non-green classification, rs12913832 and rs1800407 revealed a significant positive (synergistic) interaction effect in the absence of a main effect, which suggests that even though none of these factors are significantly associated with green eye colour as independent factors, considering them together can be highly relevant in green eye-colour prediction. We found that the chances of having green eye colour increase non-additively with the A allele in rs1800407 in combination with the T allele in rs12913832. This confirms results obtained by Sturm *et al.*,¹⁶ who also reported a connection between a combination of genotypes in rs12913832 and rs1800407 and its association with intermediate (including green) eye colour. They observed increasing penetrance for intermediate eyes with the A allele in rs1800407 and a particularly striking effect was seen with the TT genotype in rs12913832.

The next factor that revealed an interaction effect with rs12913832 in *HERC2* was the rs12896399 position in the *SLC24A4* gene, which encodes a calcium ion transporter. Position rs12896399 in *SLC24A4* was initially associated with eye colour by Sulem *et al.*¹⁸ and that finding was further supported by Liu *et al.*²⁵ In this study the masking effect of the position rs12913832 in *HERC2* on rs12896399 in determining blue versus non-blue eye colour was revealed. The strong interaction between these two factors with main effects brings redundant information. The observed masking effect of *HERC2* on *SLC24A4* is supported by a study carried out by Liu *et al.*,²² who tested pairwise interactions between 64 SNPs from genes: *HERC2*, *OCA2*, *SLC24A4*, *TYR*, *TYRP1*, *SLC45A2* and *IRF4*, and loci 1q42.3, 17q25.3 and 21q22.13 by comparing two models with and without the interaction term using the F-test.

The study carried out enabled us to reveal a novel synergistic interaction between rs12913832 in *HERC2* and rs1408799 in *TYRP1*, which is significant for explanation of green versus non-green eye colour. The product of the *TYRP1* gene, tyrosinase-related protein 1, is a melanosomal enzyme that is involved in the eumelanin synthesis

pathway. Rare mutations in *TYRP1* are responsible for oculocutaneous albinism type 3 in humans. The *TYRP1* gene was initially correlated with natural distribution of eye colours by Frudakis *et al.*,¹³ whereas the significance of the rs1408799 position in inheritance of eye colour was identified through a genome-wide association study carried out by Sulem *et al.*¹⁹ In our study, a strong interaction between rs12913832 and rs1408799 was revealed in a model built for green eye colour. It was further worked out that the character of this interaction is synergistic; that is, the interaction between these two factors delivers more information than the simple sum of the individual factors. Interestingly, neither factor revealed significant association with green eye colour; that is, they have no main effects, whereas interaction between them is highly significant, which means that they can only explain green eye colour when an interaction between them is assumed.

Using the MDR method supported by binary logistic regression analysis, we were able to show four different gene–gene interactions affecting variation in human eye colour. Because both methods used in our study consider only dichotomous dependent variables; that is, studied variable must be presented in a binary way (blue versus non-blue etc.) and the analyses had to be conducted for four individual colours separately, we additionally used one more method and conducted analysis using multinomial logistic regression where all eye colours were treated together (variation in eye colour was defined in four categories; that is, blue, green, hazel and brown). This analysis also confirmed the significance of all four revealed interactions in determining human eye colour (data not present). Undoubtedly, according to all analyses conducted, the best eye colour predictor is position rs12913832 in *HERC2*, which is consistent with our previous gene–gene interaction study.²⁶ The polymorphism in *HERC2* is also a component of all the revealed epistatic effects. Interactions between rs12913832 in *HERC2* and rs1800407 in *OCA2* obtained for the hazel eye model and between rs12913832 in *HERC2* and rs12896399 in *SLC24A4* obtained for the blue eye model have a redundant character and rely on the masking effect of the position in *HERC2* on polymorphisms in *OCA2* and *SLC24A4*. The most valuable interaction effects are ones detected for the green eye model; that is, interaction between rs12913832 in *HERC2* and rs1800407 in *OCA2* and the novel strong interaction between rs12913832 in *HERC2* and rs1408799 in *TYRP1*. In our study, both interaction models for green eye colour include factors with no main effect on this eye colour (as evaluated with binary logistic regression), and only assumption of epistasis makes them significant in prediction of this eye colouration. It needs to be pointed out that the detected epistatic effects remove only 2.05% (1.23% due to *HERC2* and *OCA2* interaction plus 0.82% due to *HERC2* and *TYRP1* interaction) of entropy in green eye-colour determination, and thus further studies are necessary to explain the genetic basis of green and other intermediate eye colours. The statistical findings reported here constitute a first step on the way to better understanding of the mechanisms that are responsible for determination of human eye colour. Inferring the biological nature of the detected interactions is evidently not an easy task and needs to be supported by functional studies. However, the gained information can be used directly in models being developed for the purpose of eye colour prediction as should to some degree be beneficial for prediction of intermediate eye colours.

ACKNOWLEDGEMENTS

We wish to thank all sample donors who participated in this project. The study was supported by a grant from the Ministry of Science and Higher Education in Poland no ON301115136 (science fund for years 2009–2012).

- 1 Bateson, W. *Mendel's principles of heredity*, Cambridge University Press: Cambridge, 1909.
- 2 Fisher, R. A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edin.* **52**, 399–433 (1918).
- 3 Cordell, H. J. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **11**, 2463–2468 (2002).
- 4 Cordell, H. J. Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.* **10**, 392–404 (2009).
- 5 Brand, A., Brand, H. & Schulte in den Bäumen, T. The impact of genetics and genomics on public health. *Eur. J. Hum. Genet.* **16**, 5–13 (2008).
- 6 Janssens, A. C. & van Duijn, C. M. Genome-based prediction of common diseases: advances and prospects. *Hum. Mol. Genet.* **17**, 166–173 (2008).
- 7 Kayser, M. & Schneider, P. M. DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. *Forensic Sci. Int. Genet.* **3**, 154–161 (2009).
- 8 Branicki, W. Studies on predicting pigmentation phenotype for forensic purposes. *Probl. Forensic Sci.* **77**, 29–52 (2009).
- 9 Shekar, S. N., Duffy, D. L., Frudakis, T., Montgomery, G. W., James, M. R., Sturm, R. A. *et al.* Spectrophotometric methods for quantifying pigmentation in human hair-influence of MC1R genotype and environment. *Photochem. Photobiol.* **84**, 719–726 (2008).
- 10 Valverde, P., Healy, E., Jackson, I., Rees, J. L. & Thody, A. J. Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat. Genet.* **11**, 328–330 (1995).
- 11 Eiberg, H. & Mohr, J. Assignment of genes coding for brown eye colour (BEY2) and brown hair colour (HCL3) on chromosome 15q. *Eur. J. Hum. Genet.* **4**, 237–241 (1996).
- 12 Kanetsky, P. A., Swoyer, P. J., Panossian, S., Holmes, R., Guerry, D. & Rebbeck, T. R. A polymorphism in the agouti signaling protein gene is associated with human pigmentation. *Am. J. Hum. Genet.* **70**, 770–775 (2002).
- 13 Frudakis, T., Thomas, M., Gaskin, Z., Venkateswarlu, K., Chandra, K. S., Ginjupalli, S. *et al.* Sequences associated with human iris pigmentation. *Genetics* **165**, 2071–2083 (2003).
- 14 Graf, J., Hodgson, R. & van Daal, A. Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. *Hum. Mutat.* **25**, 278–284 (2005).
- 15 Lamason, R. L., Mohideen, M. A., Mest, J. R., Wong, A. C., Norton, H. L. & Aros, M. C. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–1786 (2005).
- 16 Sturm, R. A., Duffy, D. L., Zhao, Z. Z., Leite, F. P. N., Stark, M. S. & Hayward, N. K. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye colour. *Am. J. Hum. Genet.* **82**, 424–431 (2008).
- 17 Ku, C. S., Loy, E. Y., Pawitan, Y. & Chia, K. S. The pursuit of genome-wide association studies: where are we now? *J. Hum. Genet.* **55**, 195–206 (2010).
- 18 Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P. *et al.* Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39**, 1443–1452 (2007).
- 19 Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Jakobsdottir, M. *et al.* Two newly identified genetic determinants of pigmentation in Europeans. *Nat. Genet.* **40**, 835–837 (2008).
- 20 Kayser, M., Liu, F., Janssens, A. C., Rivadeneira, F., Lao, O., van Duijn, K. *et al.* Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris colour gene. *Am. J. Hum. Genet.* **82**, 411–423 (2008).
- 21 Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A. *et al.* A genome-wide association study identifies novel alleles associated with hair colour and skin pigmentation. *PLoS Genet.* **4**, e1000074 (2008).
- 22 Liu, F., Wollstein, A., Hysi, P. G., Ankra-Badu, G. A., Spector, T. D., Park, D. *et al.* Digital quantification of human eye colour highlights genetic association of three new loci. *PLoS Genet.* **6**, e1000934 (2010).
- 23 Walsh, S., Lindenbergh, A., Zuniga, S. B., Sijen, T., deKnijff, P. & Kayser, M. Developmental validation of the IrisPlex System: determination of blue and brown iris colour for forensic intelligence. *Forensic Sci. Int. Genet.* (2010) doi:10.1016/j.fsigen.2010.09.008 (in press).
- 24 Walsh, S., Liu, F., Ballantyne, K. N., van Oven, M., Lao, O. & Kayser, M. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci. Int. Genet.* (2010) doi:10.1016/j.fsigen.2010.02.004 (in press).
- 25 Liu, F., van Duijn, K., Vingerling, J. R., Hofman, A., Uitterlinden, A. G., Janssens, A. C. *et al.* Eye colour and the prediction of complex phenotypes from genotypes. *Curr. Biol.* **19**, 192–193 (2009).
- 26 Branicki, W., Brudnik, U. & Wojas-Pelc, A. Interactions between *HERC2*, *OCA2* and *MC1R* may influence human pigmentation phenotype. *Ann. Hum. Genet.* **73**, 160–170 (2009).
- 27 Box, N. F., Wyeth, J. R., O'Gorman, L. E., Martin, N. G. & Sturm, R. A. Characterization of melanocyte stimulating hormone receptor variant alleles in twins with red hair. *Hum. Mol. Genet.* **6**, 1891–1897 (1997).
- 28 Rebbeck, T. R., Kanetsky, P. A., Walker, A. H., Holmes, R., Halpern, A. C., Schuchter, L. M. *et al.* P gene as an inherited biomarker of human eye colour. *Cancer Epidemiol. Biomarkers Prev.* **11**, 782–784 (2002).
- 29 Kanetsky, P. A., Ge, F., Najarian, D., Swoyer, J., Panossian, S., Schuchter, L. *et al.* Assessment of polymorphic variants in the melanocortin-1 receptor gene with cutaneous pigmentation using an evolutionary approach. *Cancer Epidemiol. Biomarkers Prev.* **13**, 808–819 (2004).
- 30 Branicki, W., Brudnik, U., Kupiec, T., Wolańska-Nowak, P. & Wojas-Pelc, A. Determination of phenotype associated SNPs in the *MC1R* gene. *J. Forensic Sci.* **52**, 349–354 (2007).
- 31 Branicki, W., Brudnik, U., Kupiec, T., Wolańska-Nowak, P., Szczerbińska, A. & Wojas-Pelc, A. Association of polymorphic sites in the *OCA2* gene with eye colour using the tree scanning method. *Ann. Hum. Genet.* **72**, 184–192 (2008).
- 32 Eiberg, H., Troelsen, J., Nielsen, M., Mikkelsen, A., Mengel-From, J., Kjaer, K. W. *et al.* Blue eye colour 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 (2008).
- 33 Epstein, M. P. & Satten, G. A. Inference on haplotype effects in case-control studies using unphased genotype data. *Am. J. Hum. Genet.* **73**, 1316–1329 (2003).
- 34 Kira, K. & Rendell, L. A. A practical approach to feature selection. *Proc. Ninth Int. Workshop Machine Learning* 249–256 (1992).
- 35 Konoenko, I. Estimating attributes: analysis and extensions of Relief. *Proc. Eur. Conf. Machine Learning* 171–182 (1994).
- 36 Robnik-Sikonja, M. & Konoenko, I. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning* **53**, 23–69 (2003).
- 37 Jakulin, A. & Bratko, I. Analyzing attribute interactions. *Lect. Notes Artif. Intell.* **2838**, 229–240 (2003).
- 38 Jakulin, A., Bratko, I., Smrke, D., Demska, J. & Zupan, B. Attribute interactions in medical data analysis. *Lect. Notes Artif. Intell.* **2780**, 229–238 (2003).
- 39 Iida, R., Ueki, M., Takeshita, H., Fujihara, J., Nakajima, T. & Kominato, Y. Genotyping of five single nucleotide polymorphisms in the *OCA2* and *HERC2* genes associated with blue-brown eye colour in the Japanese population. *Cell. Biochem. Funct.* **27**, 323–327 (2009).
- 40 Bogdanowicz, W., Allen, M., Branicki, W., Lembring, M., Gajewska, M. & Kupiec, T. Genetic identification of putative remains of the famous astronomer Nicolaus Copernicus. *Proc. Natl Acad. Sci. USA* **106**, 12279–12282 (2009).
- 41 Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
- 42 Moore, J. H. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum. Hered.* **56**, 73–82 (2003).
- 43 Carlborg, O. & Haley, C. S. Epistasis: too often neglected in complex trait studies? *Nat. Rev. Genet.* **5**, 618–625 (2004).
- 44 Hahn, L. W., Ritchie, M. D. & Moore, J. H. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics* **19**, 376–382 (2003).
- 45 Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl, F. F. *et al.* Multifactor dimensionality reduction reveals high-order interactions among estrogen metabolism genes in sporadic breast cancer. *Am. J. Hum. Genet.* **69**, 138–147 (2001).
- 46 Ritchie, M. D., Hahn, L. W. & Moore, J. H. Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, phenocopy, and genetic heterogeneity. *Genet. Epidemiol.* **24**, 150–157 (2003).
- 47 Ritchie, M. D., White, B. C., Parker, J. S., Hahn, L. W. & Moore, J. H. Optimization of neural network architecture using genetic programming improves the detection and modeling of gene-gene interactions in studies of human diseases. *BMC Bioinform.* **4**, 28 (2003).
- 48 Hahn, L. W. & Moore, J. H. Ideal discrimination of discrete clinical endpoints using multilocus genotypes. *In Silico Biol.* **4**, 183–194 (2004).
- 49 Moore, J. H. Computational analysis of gene-gene interactions in common human diseases using multifactor dimensionality reduction. *Expert. Rev. Mol. Diagn.* **4**, 795–803 (2004).
- 50 Moore, J. H., Gilbert, J. C., Tsai, C.-T., Chiang, F.-T., Holden, T., Barney, N. *et al.* A flexible computational framework for detecting, characterizing and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *J. Theor. Biol.* **241**, 252–261 (2006).
- 51 Moore, J. H. & Williams, S. M. New strategies for identifying gene-gene interactions in hypertension. *Ann. Med.* **34**, 88–95 (2002).
- 52 Lee, J. H., Moore, J. H., Park, S. W., Jang, A. S., Uh, S. T., Kim, H. Y. *et al.* Genetic interactions model among *Eotaxin* gene polymorphisms in asthma. *J. Hum. Genet.* **53**, 867–875 (2008).
- 53 Beretta, L., Cappiello, F., Moore, J. H. & Scorza, R. Interleukin-1 gene complex single nucleotide polymorphisms in systemic sclerosis: a further step ahead. *Hum. Immunol.* **69**, 187–192 (2008).
- 54 Makova, K. & Norton, H. Worldwide polymorphism at the *MC1R* locus and normal pigmentation variation in humans. *Peptides* **26**, 1901–1908 (2005).
- 55 Puri, N., Gardner, J. M. & Brilliant, M. H. Aberrant pH of melanosomes in pink-eyed dilution (P) mutant melanocytes. *J. Invest. Dermatol.* **115**, 607–613 (2000).
- 56 Ancans, J., Tobin, D. J., Hoogduijn, M. J., Smit, N. P., Wakamatsu, K. & Thody, A. J. Melanosomal pH controls rate of melanogenesis, eumelanin/phaeomelanin ratio and melanosome maturation in melanocytes and melanoma cells. *Exp. Cell Res.* **268**, 26–35 (2001).
- 57 Lamoreux, M. L., Zhou, B. K., Rosemblat, S. & Orlow, S. J. The pinkeyed-dilution protein and the eumelanin/pheomelanin switch: in support of a unifying hypothesis. *Pigment Cell Res.* **8**, 263–270 (1995).

- 58 Sturm, R. A., Teasdale, R. D. & Box, N. F. Human pigmentation genes: identification, structure and consequences of polymorphic variation. *Gene* **277**, 49–62 (2001).
- 59 Zhu, G., Evans, D. M., Duffy, D. L., Montgomery, G. W., Medland, S. E., Gillespie, N. A. *et al*. A genome scan for eye colour in 502 twin families: most variation is due to a QTL on chromosome 15q. *Twin Res.* **7**, 197–210 (2004).
- 60 Posthuma, D., Visscher, P. M., Willemsen, G., Zhu, G., Martin, N. G., Slagboom, P. E. *et al*. Replicated linkage for eye colour on 15q using comparative ratings of sibling pairs. *Behav. Genet.* **36**, 12–17 (2006).
- 61 Duffy, D. L., Montgomery, G. W., Chen, W., Zhao, Z. Z., Le, L., James, M. R. *et al*. A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-colour variation. *Am. J. Hum. Genet.* **80**, 241–252 (2007).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)