MODEL BASED EYE COLOR PREDICTION USING BAYESIAN NETWORKS WITH CONSEQUENT LIKELIHOOD RATIO INTERPRETATION IN CZECH REPUBLIC

Anastassiya Zidkova[1,2], Ales Horinek[1], Vlastimil Stenzl[3], Marie Korabecna[1]
[1]Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague
[2]Institute of Forensic Medicine and Toxicology, First Faculty of Medicine, Charles University in Prague
[3]Institute of Criminalistics Prague, Prague, Czech Republic

Forensic molecular phenotyping allows to narrow subset of possible suspects. Prediction of eye color and ancestry could be useful in case of unknown remnants identification and verification of eye witness testimony. Many studies were conducted to create SNP panel for eye color determination, but only few of them were focused on practical implementation of these findings. Here we would like to present model based approach to create panel of SNP markers for eye color and ancestry prediction with subsequent likelihood-ratio (LR) based interpretation.

Population sample contained 131 unrelated individuals: 100 (76%) Caucasians, 31 (24%) Asian. Samples were binned according to eye color: 47 (36%) light eye color and 84 (64%) dark eye color. Genotyping was performed using TaqMan MGB assays. All samples were analyzed using panel of 22 SNP markers selected from literature. Data were divided into two sets: training (100 Caucasian, 31 Asian) and testing (47 Caucasian, 81 Asian). Caucasian data for testing set were obtained by genotyping, Asian data for testing set were extracted from HapMap. Firstly, SNP data of the training set were filtered using ReliefF, which minimizes risk of SNP false positive association with studied phenotype. Consequently, multifactor dimensionality reduction was performed to uncover possible epistatic regulation between studied loci. For validation of selected markers Bayesian network was constructed. Network contained ancestry determination SNPs (rs1426654, rs16891982) and eye color determination SNPs (rs12913832, rs7495174, rs916977). Sensitivity of the model reached 0.96, specificity was 0.99. According to the constructed Bayesian network, LR was calculated in real-cases. LR value was than transformed to the verbal statement.

The study was supported by the grant No. I/328 of the Ministry of Industry and Trade of the Czech Republic.