



## FORENSIC APPLICATION OF A RAPID TEST FOR RED HAIR COLOUR PREDICTION AND SEX DETERMINATION

Wojciech BRANICKI<sup>1</sup>, Paulina WOLAŃSKA-NOWAK<sup>1</sup>, Urszula BRUDNIK<sup>2</sup>, Tomasz KUPIEC<sup>1</sup>, Katarzyna SZYMAŃSKA<sup>1</sup>, Anna WOJAS-PELCE<sup>2</sup>

<sup>1</sup> *Institute of Forensic Research, Krakow*

<sup>2</sup> *Collegium Medicum of the Jagiellonian University, Krakow*

### Abstract

Genetic prediction of physical traits is of potential interest to forensic science. It has been shown that the MC1R gene is responsible for most cases of red hair colour in humans. The impact of this gene is very distinct, especially in populations of Northern European origin. Its analysis can be informative for prediction of the so-called RHC phenotype, but any new method implemented in forensic science should be subjected to validation studies. We optimised and validated a test which allows simultaneous analysis of the eleven most important MC1R polymorphisms and additionally one variable site in amelogenin, which is informative for sex typing. The developed simple Bayesian network enables evaluation of the influence of the kind of MC1R genotype as well as other factors on the unknown person's hair colour. The overall probability structure is completely determined by specifying the conditional probability tables for each node, based on the obtained data.

### Key words

Forensic genetics; Red hair colour; MC1R; Validation; Bayesian network.

*Received 16 August 2007; accepted 10 September 2007*

### 1. Introduction

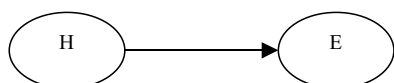
Forensic biological traces collected at crime scenes are normally subjected to human identification which is based on analysis of highly variable STR systems. Obtained crime scene DNA profiles are compared with reference profiles and an identified match is understood as strong evidence supporting the hypothesis of common origin of the analysed samples. If no suspect is involved in a case and thus no reference material is available, obtained crime scene profiles are usually loaded into a national DNA database and if matched with a reference profile present in the database the evidential value is also considered significant. However, if no match is found, the investigation may significantly slow down as other hypotheses concern-

ing the perpetrator must be verified. In some more serious cases, a DNA screen may be applied in order to eliminate individuals from the inquiry. This may involve analysis of hundreds or even thousands of samples and in worst-case scenarios take many years [2]. The crime scene samples may, however, also be used for intelligence purposes and in this way may reduce the number of samples to be screened. Currently available identification tests allow gender determination and thus make it possible for an investigation to focus on male or female suspects. Genetic prediction of physical traits is another emerging area for forensic science, and doubtlessly this may become very useful at the investigation stage [5, 16, 24].

In general, physical traits should be considered as quantitative traits and thus difficult to analyse. For ex-

ample, it is assumed that pigmentation, one of the most distinguishing traits in humans, is controlled by more than 100 genes [18]. Some of them have already been identified and their association with different pigment traits has been shown [17, 21]. Interestingly, inheritance of red hair colour in humans is strongly related with the only gene which encodes the melanocortin 1 receptor (MC1R) (16q24.3, MIM 155555). Numerous studies carried out on variation within the MC1R have shown that polymorphisms of the MC1R gene play a crucial role in determination of the red hair colour (RHC) phenotype in man [e.g. 7, 9, 12]. Our previous study showed that in the Polish population approximately 97.5% individuals with genotypes 151/151, 160/160, 151/160 had red hair colour [6]. These two and a few other polymorphic sites have high predictive value and thus their analysis may be of great importance for forensic phenotype prediction. Our previously developed minisequencing assay allows rapid analysis of five crucial MC1R SNPs in a single reaction [6].

This paper presents an enhanced assay for analysis of eleven SNPs within the MC1R gene. The test also enables genotyping of one SNP located in the first intron of the amelanin gene sequence, which is informative for sex determination. The performed validation of the assay assumed its application to forensic examinations of biological traces. We also address the issue of evaluation of the impact that obtained MC1R results should have on expressivity of the RHC phenotype. To graphically represent the relationships between the observed data we developed a simple causal network enabling inferring as to the probability of individual hair colour on the basis of a known MC1R genotype. A Bayesian network (directed acyclic graph) consists of the following: a set of variables (represented by nodes) and a set of directed edges between variables, each variable has a finite set of mutually exclusive states. To each variable  $A$  with parents  $B_1, \dots, B_n$ , there is attached the conditional probability table  $Pr(A|B_1, \dots, B_n)$  [10]. The vertical line means conditioned on or given [14]. Such a causal network can be used to follow how a change in the level of certainty in one variable (e.g. the kind of MC1R genotype) may change the level of certainty for other variables (e.g. probability of a given hair colour of an unknown donor of a crime sample) [23]. A Bayesian networks operates on conditional probability. Therefore, if the occurrence of evidence  $E$  depends on hypothesis  $H$ , the probability for both  $E$  and  $H$  can be expressed using Bayesian network expression by the scheme:



Using the probabilistic expression the above outline can be expressed by the formula:

$$Pr(H, E) = Pr(H) \cdot Pr(E|H) \quad \{1\}$$

## 2. Methods

### 2.1. Samples

390 samples of buccal swabs were taken from unrelated southern Polish Caucasians.

The validation study included different types of biological material analysed in forensic science: details are given in further sections. DNA was extracted either using the silica based extraction protocol (A & A Biotechnology, Gdansk) or with Biorobot M48 (Qiagen, Hilden, Germany). DNA concentration was measured with Quant-IT PicoGreen dsDNA Quantitation Kit (Invitrogen, Karlsruhe, Germany).

### 2.2. Assay design

Eleven variable DNA sites within the MC1R exon – missense, nonsense or frameshift mutations – were selected on the basis of our own research and literature data [7, 4, 3, 6]. Seven of them are found to strongly predispose to the red hair colour phenotype. These are all polymorphisms studied with multiplex 1 (86.1A, C252A, G425A, C451T, C478T, G880C) and one polymorphism (C456A) included in multiplex 2. These polymorphisms are defined in literature as high-penetrance MC1R alleles “R” (significantly associated with red hair colour and fair skin) [13]. An additional four polymorphic sites included in multiplex 2 (G178T, G274A, T464C, G488A) are considered to be low-penetrance MC1R alleles “r” (weakly associated with the RHC phenotype). The multiplex 2 also includes a single nucleotide polymorphism located in the intron 1 of the amelanin which is used for sex determination (Figure 1).

#### 2.2.1. Multiplex PCR

PCR primers were designed to enable simultaneous amplification of three fragments of the MC1R gene encompassing all eleven SNPs (Table I) [6]. The multiplex PCR reaction was optimised for three MC1R fragments and the intron 1 amelanin fragment (primers described by Sullivan and co-workers) [22]. The PCR reaction mixture consisted of 5  $\mu$ l of Multiplex PCR Master Mix, 1  $\mu$ l Q solution and 1  $\mu$ l of primer mix (Table I). The temperature profile was as

Query	186	TAAAAGCTACCACCTCATCCTGGGCACCCTGGTTATATCAACTTCAGCTATGAGGTAATT	245
Sbjct	186	TAAAAGCTACCACCTCATCCTGGGCACCCTGGTTATATCAACTTCAGCTATGAGGTAATT	245
Query	246	TTTCTCTTTACTAATTTTGATCACTGTTTGATTAGCAGTCCCCTGGGCTCTGTAAAGAA	305
Sbjct	246	TTTCTCTTTACTAATTTTGACCATTGTTGCGTTAACAATGCCCTGGGCTCTGTAAAGAA	305
Query	306	TAGTGGGTGGATTCTTCATCCCAAATAAAGTGGTTTCTCAAGTGGTCCCAATTTTACAGT	365
Sbjct	306	TAGTGTGTGATTCTTTATCCAGAT-----GTTTCTCAAGTGGTCCCTGATTTTACAGT	359
Query	366	TCCTACCATCAGCTTCCCAGTTTAAAGCTCTGATGGTTGGCCTCAAGCCTGTGTTGCTCCA	425
Sbjct	360	TCCTACCACCAGCTTCCCAGTTTAAAGCTCTGATGGTTGGCCTCAAGCCTGTGTCGTCCCA	419
Query	426	GCACCCTCCTGCCTGACCATTGCGATTGACTCTTTCCTCCTAAATATGGCTGTAAGTTTA	485
Sbjct	420	GCAGCCTCCCGCCTGGCCACTCTGACTCAGTCTGTCTCCTAAATATGGCCGTAAGCTTA	479
Query	486	TTCATTGATGAACCACTGCTCAGGAAGGTTCCATGAAAGGGCAAAAAGTCAACTCTGACT	545
Sbjct	480	CCCA-TCATGAACCACTACTCAGGGAGGCTCCATGATAGGGCAAAAAGTAAACT----CT	534
Query	546	GACCAGCTTGGTTCTATCCCATCCGGTAAAATGTAAGATTAGGTAA	592
Sbjct	535	GACCAGCTTGGTTCTAACCCAGCTAGTAAAATGTAAGGATTAGGTAA	581

Fig. 1. GenBank sequences for the amelogenin intron 1 fragments characteristic for the X (Sbjct) and Y (Query) chromosome. The analysed SNP position is marked with an arrow.

TABLE I. DETAILS OF PCR PRIMERS USED FOR AMPLIFICATION OF THE MC1R AND AMELOGENIN SNP POSITIONS

DNA fragment	Primer sequence (5'–3')	Encompassed SNPs	Amplicon length	Concentration [ M]
1*	MIF: TGGGCTCCCTCAACTCCACCC MIR: AGCAGGAGGATGACGGCCGTCT	86insA (C-A); G178T; C252A; G274A	268 bp	0.125
2*	MIIF: AGCTCCATGCTGTCCAGCCT MIIR: AGCAGGACGGCCACGTGGT	G425A; C451T; C456A; T464C; C478T; G488A	193 bp	0.125
3*	MIIIF: CTCACACTCATCGTCCTCTG MIIIR: TGCCCAGCACACTTAAAGC	G880C	206 bp	0.125
4**	SXY1: CCCTGGGCTCTGTAAAGAAT SXY2: ATCAGAGCTTAAACTGGGAAGCTG	C/T	106/112 bp	0.25

\* – according to [6], \*\* – according to [22].

follows: {94°C/15 min, (94°C/0.30 min, 63°C/1.30 min, 72°C/1.30 min) 30 or 34, 72°C/10 min} 4°C/. PCR products were separated with agarose gel and purified using the ExoSap IT kit (Amersham Pharmacia, Freiburg, Germany).

### 2.2.2. Minisequencing reactions

Two multiplex primer extension reactions were designed for genotyping of the analysed SNP positions. Details of extension primers are given in Table II. The purified PCR products were subjected to minisequencing reactions using the SNaPshot multiplex PCR kit, following instructions given by the manufacturer.

TABLE II. DETAILS OF EXTENSION PRIMERS USED FOR GENOTYPING OF THE MC1R AND AMELOGENIN POLYMORPHISMS. NEUTRAL SEQUENCE ACCORDING TO [19]

Multiplex 1						
Polymorphism	Neutral 5'	3'	Target 5'	3'	bp	Concentration [ M]
86insA (C-A)	–		TGGGGCTGGCTGCCAA		16/16	0.8
C252A	–		CTGCCCTGGCCTTGTCTCGGA		18/18	0.1
C451T	TCTGACAA		TCTCCATCTTCTACGCACTG		20/28	0.1
G425A	AAAGTCTGACAA		GCCATCGCCGTGGACC		16/28	0.025
C478T	AAAGTCTGACAA		ACAGCATCGTGACCCTGCCG		20/32	0.1
G880C	ACGTCGTGAAAGTCTG	ACAA	CTGCAATGCCATCATC		16/36	0.05
Multiplex 2						
Polymorphism	Neutral 5'	3'	Target 5'	3'	bp	Concentration [ M]
C456A	–		TTCTACGCACTGCGCTA		17/17	0.1
G178T	–		TGGTGGAGAACGCGCTGGTG		20/20	0.05
G488A	–		AGATGGCCGCAACGGCT		17/17	0.8
G274A	AAGTCTGACAA		TGGTGGAGCGGGAGCAAC		17/28	0.2
T464C	CGTGAAAGTCTGACAA		CCGCGGCAGGGTCACG		16/32	0.05
AMEL.C/T	AAAGTCTGACAA		GTTTCTCAAGTGGTCC		16/28	0.1

### 2.3. Validation study

#### 2.3.1. Reliability

Twenty samples of known MC1R genotypes, which were previously determined by cycle sequencing reactions with the method described by Kanetsky and co-workers [12] were subjected to analysis using the optimised SNaPshot assays. The analysed samples were chosen deliberately in order to represent males and females and all possible MC1R variants.

#### 2.3.2. Sensitivity

The tested DNA concentrations were as follows: 0.5, 0.25, 0.125, 0.06, 0.03 and 0.015 ng. DNA standard 007 (Applera) was used in sensitivity testing. Samples were amplified using a total number of 30 and 34 cycles. Results were evaluated with agarose gels and marginal points were subjected to minisequencing reactions.

#### 2.3.3. Reproducibility

Thirty different samples collected from the same individual were analysed. The determined genotype was confirmed by the cycle sequencing method.

#### 2.3.4. Biological traces

Liquid blood was deposited on cotton (repeated three times) and kept in 1) room conditions (24 hours, 7 days, 1 and 3 months; 2) increased humidity and temperature – water bath set at 37°C (24 h, 48 h, 72 h, 7 days). Biological material from used personal items and other specimens frequently examined in forensic laboratories (razor, toothbrush, spoons, bottles, cups, lipsticks, magnetic keys, credit cards, cigarette stubs and semen stains) were collected. All samples were subjected to DNA extraction using Biorobot M48 and analysed using the optimised assays. One bone sample was subjected to DNA extraction using the organic method and Centricon 100 purification.

### 2.4. Statistical evaluation of the influence of the MC1R genotype on phenotypic features

A causal Bayesian network was prepared using Hugin software ver, 6.8. The network consisted of four nodes: “MC1R”, “individual hair colour”, “population stratification” and “other reason” (Figure 2). It was assumed for the sake of simplicity that hair colour is determined by two factors: 1) the MC1R genotype and 2) the action of other factors, which are in turn influ-

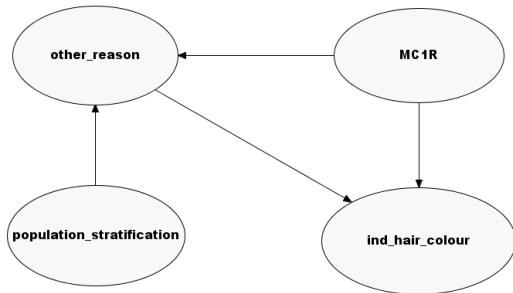


Fig. 2. Causal Bayesian network – representation of the dependencies between particular nodes. Individual hair colour is influenced by the MC1R genotype and other factors, which are in turn dependent on population stratification (the proportion of individuals with different kinds of hair colour in our population).

enced by the kind of individual MC1R SNP panel and population stratification (proportions of dark red, red, dark-blond and black haired individuals in the analysed sample of 390 individuals). The authors are aware that the analysed sample is not representative of the general population of South Poland because of overrepresentation by red-haired individuals. A probability table for the MC1R node contains a priori information on population distribution of the SNP variants. Hence this probability table has 7 mutually exclusive chosen states: 151/151, 151/160, 160/160, 142/142, 151/other and 160/other. The conditional probability table for the node “individual hair colour” represents all probabilities for  $Pr(\text{individual hair colour} | MC1R, \text{other reason})$ , presented in Table III. After setting hard evidence on any of the nodes states, e.g. the crime sample has MC1R genotype – 160/160 and propagating it through the network one can ascertain the probabilities of particular hair colours of the sample donor.

### 3. Results and discussion

#### 3.1. Validation study

Our previous study, performed on 80 red haired individuals and more than 200 controls of both sexes confirmed that in most cases red hair colour can be explained by polymorphisms within the coding segment of the MC1R gene [6]. Particular genotypes are strongly associated with the red hair phenotype. Our results and literature data indicate that the predictive value is especially high in case of high-penetrance – “R” variants – 86.1insA, C252A, G425A, C451T, C456A, C478T, and G880C appearing in the homozygous state or in heterozygous combinations. In this paper we present a SnaPshot test optimised for selective analysis of eleven SNPs involved in red hair colour determination and analysis of single transition within the amelogenin gene, which allows gender determination. Examples of electropherograms (GenScan) for designed minisequencing assays 1 and 2 are presented in Figure 3.

SNaPshot results obtained for 20 samples of different MC1R genotypes were concordant with cycle sequencing reactions in 19 cases. In one case of discordance, the cycle sequencing reaction showed the presence of homozygote 880C genotype, whereas the minisequencing reaction indicated the presence of compound heterozygote 451C/T and 880G/C. Repeated analysis revealed that the cycle sequencing result was incorrect the first time, and the most probable reason for this was allelic dropout (missed allele 451T, 880G). The prepared test allows gender determination. Single T is characteristic for female samples (double X) and heterozygous C/T is characteristic for males (Figure 4). Sex determination using the optimised method was found to be correct for all analysed samples (the total number exceeded 100 samples). Also, in the

TABLE III. TABLE OF CONDITIONAL PROBABILITIES:  $Pr(\text{individual hair colour} | MC1R \text{ genotype})$  AS A RESULT OF PERFORMED POPULATION STUDY

Hair colour	MC1R genotype						
	151/151	151/160	160/160	142/142	160/other	151/other	Other/other
Dark red	0.57	0.46	0.44	0.5	0.1	0.15	0.02
Red	0.43	0.5	0.25	0.5	0.08	0.17	0.03
Auburn	0	0	0.19	0	0.02	0	0.04
Dark-blond	0	0.04	0	0	0.44	0.28	0.46
Blond	0	0	0.06	0	0.18	0.26	0.17
Black	0	0	0	0	0.11	0.06	0.16
Brown	0	0	0.06	0	0.08	0.08	0.12

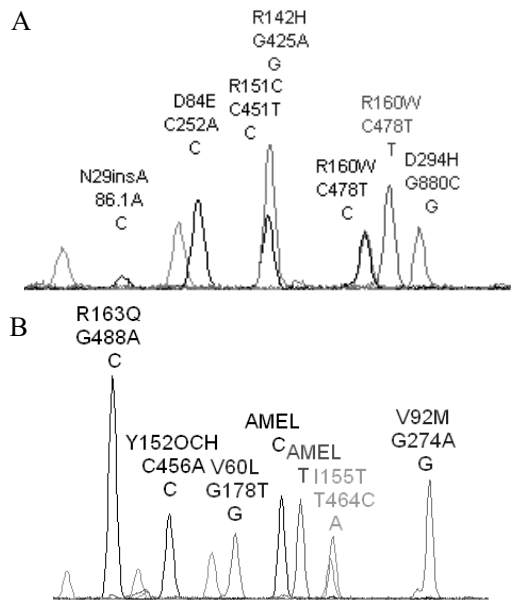


Fig. 3. GeneScan results for minisequencing assay 1 (panel A) and assay 2 (Panel B).

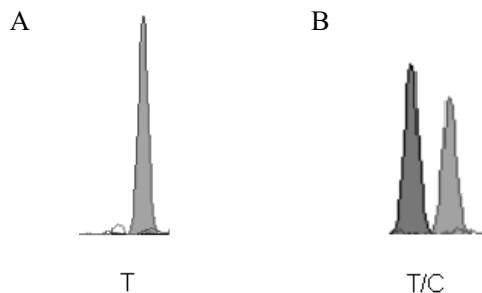


Fig. 4. Sex testing using the amelogenin intron 1 SNP position. Panel A – female DNA. Panel B – male DNA.

case of all 33 different samples collected from one individual, the same correct genotyping results were obtained. These results confirmed the high reliability and reproducibility of the validated minisequencing assays.

Usually, in the case of STR multiplex kits, the recommended DNA template concentration is at least 0.5 ng [1]. Performed experiments confirmed that this concentration is sufficient to obtain correct results using the optimised assays when 30 cycles of PCR amplification were applied. Increased PCR cycles number option (34 cycles) is used by the forensic community when samples containing tiny amounts of DNA, usually classified as LCN – low copy number, (< 100 pg) are to be analysed [8]. As expected, PCR sensitivity of the validated assay was much higher when 34 PCR cycles were applied instead of the regular option of 30 cycles. With this modification, correct genotyping results were obtained with as little as 15 pg of template DNA. The performed experiments

showed, therefore, that the sensitivity of the assay is at the level characteristic for commercial kits used in forensic genetics.

Forensic laboratories frequently examine biological traces which have been exposed to heavy degradation processes. It has been observed that such problematic or low quality specimens subjected to PCR amplification often fail to give positive results when long amplicons are to be generated. Thus amplification of the complete MC1R exon should not be considered as a successful method in forensic examinations of biological traces. The optimised method involves amplification of three fragments of the MC1R gene encompassing all the studied SNPs. Amplicon sizes do not exceed 270 bp. The assays were subjected to testing using typical forensic specimens (Table IV).

TABLE IV. RESULTS OF EXPERIMENTS PERFORMED ON PROBLEMATIC SAMPLES

Specimen	Correct minisequencing results
Room temp. 24 h	3/3
Room temp. 7 days	3/3
Room temp. 1 month	3/3
Room temp. 3 months	3/3
Water bath 24 h	3/3
Water bath 48 h	3/3
Water bath 72 h	3/3
Water bath 7 days	3/3
Semen stains	3/3
Cigarette stubs	4/4
Chewing gum	4/4
Cups, bottles	6/8
Spoons	3/3
Lipstick	3/3
Toothbrush	1/1
Razor	1/1
Jewellery	2/6
Plastic cards	1/4

The experiments showed that the assays worked correctly in the case of 52 out of 61 samples. In six cases, PCR amplification failed, mostly for the most difficult LCN samples (3 rings, 1 earring, 1 plastic card and 1 cup). In three samples, the MC1R genotype was found to be inconsistent with the real donor geno-

type. Such a result was obtained in the case of a swab taken from a cup, and two credit cards. Further investigation of these samples with the SGM Plus kit revealed that incorrect MC1R results must have been caused by mixed DNA. The credit cards could, of course – besides being touched by the cardholders – have been touched by shopkeepers. More surprising was the DNA mixture found on the analysed cup, which, according to the information, should have been used by a single person before swabbing.

The test thus proved to be a very sensitive and efficient tool for analysis of degraded forensic specimens. However, performed experiments also showed that misleading results may be obtained when a DNA mixture is analysed. In some cases, the high imbalance between C/T peaks which refer to amelogenin polymorphism may indicate a male/female mixture (data not present). In most cases, the test results fail to indicate the presence of a mixed genotype and therefore, a sample which is designated for MC1R genotyping should be previously analysed for the possible presence of a DNA mixture and such samples should be excluded from MC1R analysis.

One family study was also performed in a missing person identification case. Standard analysis of STR markers using an Identifiler kit revealed a parent-child relationship between the analysed human remains (femur) and alleged parents. Performed examination using the MC1R/amelogenin assays correctly confirmed the male gender of the analysed human remains and also showed the presence of the MC1R genotype (451C/T), which was possible assuming parental MC1R genotypes.

### 3.2. Bayesian network as a tool for evaluation of the results of MC1R genotyping

Correct evaluation of genotyping results is a fundamental issue in forensic examinations. Homozygous genotypes for so-called R polymorphisms (86.1insA, C252A, G425A, C451T, C456A, C478T, G880C) as well as their heterozygous combinations significantly support the hypothesis that the sample donor has red hair colour and fair skin (RHC phenotype). Even in the case of R/R genotypes, the probability of RHC is not 100% and in order not to mislead the inquiry the result must be properly evaluated. Thus, one of the important aims of this study was to propose an inferring model based on the performed genetic analysis results in the Bayesian network. The proposed model can provide scientific justification for the results or conclusions of MC1R genotypes examinations. In addition, this model can also be used by other forensic practitioners,

who can add information on the influence of other genes on the hair colour phenotype.

Bayesian networks are very flexible in terms of incorporating new variables in order to track and assess additional knowledge so that the reasoning for the conclusions can be evaluated. On the basis of 1) penetrance of R variants into the red hair phenotype, 2) frequency of different kinds of individual hair colour and 3) known population distribution of MC1R genotypes, we inferred about the possibilities of the hair colour of the donor of an unknown crime sample. Figure 5 presents prior information connected with “parent” nodes and resulting conditional probability tables before entering (fixing) the so called “hard” evidence.



Fig. 5. Representation of probability tables connected with particular nodes before entering the “hard” evidence. MC1R, population\_stratification – parental nodes.

Evidence in the case of a variable is a statement of the certainties of its states. If the variable is instantiated, it is called “hard” evidence; otherwise it is called “soft” evidence [10]. When, for example, we obtain a sample with MC1R genotype 151/151, then the first estimation of the probability that the donor has dark red hair colour is about 57%, whereas that he has red hair equals 43%. This is a result of our finding that in the analysed sample there were no individuals other than red or dark red haired, with the 151/151 genotype. But there were other red haired individuals that had no 151/151 MC1R genotypes. One should differentiate between the two conditional probabilities:  $Pr(\text{red hair} | 151/151 \text{ MC1R genotype})$  and  $Pr(151/151 \text{ MC1R genotype} | \text{red hair})$ . The first probability equals about 43%, but the second 7%. However, the values obtained in our model, after incorporating the hard evidence that the sample e.g. has the 151/151 MC1R genotype, amount to 55.5% probability for having red hair and 44.5% probability for having dark red hair.

This is because they are updated (corrected) by knowledge of the population distribution of the MC1R genotype in our population and the action of other factors (Figure 6). Hence the operational advice to police is fairly certain: “the perpetrator should have dark red or red hair colour”.

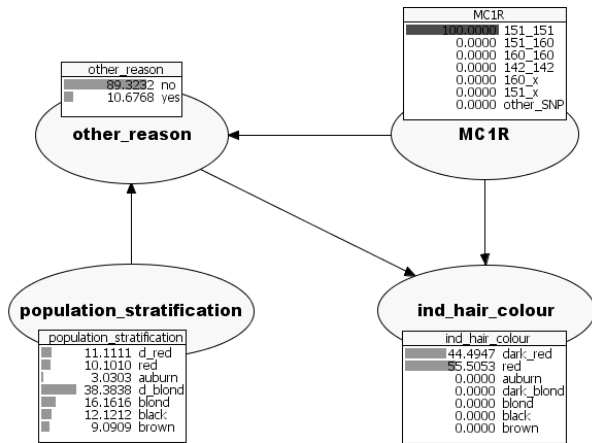


Fig. 6. Representation of probability tables connected with particular nodes after entering the “hard” evidence, that is, e.g. the crime sample has MC1R genotype 151/151. The probability that the donor of the sample has dark red hair is 44.49%, that the hair is red in general is 55.51%.

Other information that can be retrieved from the network, results from the so called diagnostic direction. If one introduces into the network the known information that one person is phenotypically dark red haired, then the probability that he/she has the 151/151 MC1R genotype equals about 1%; that he/she is 151/160 amounts to about 34%; 160/160 – 17% and so on (Figure 7). This direction of reasoning is of course not useful for casework.

The developed network may be easily extended when new information is available. If we have additional knowledge about other genes influencing hair colour it will be utilised for further updating our present knowledge on relevant factors. The established model clearly presents the probabilistic values and the causal relationship between hypotheses and evidence. Through evaluation of known (“hard”) evidence and probabilistic analysis of posterior assessments, the Bayesian network model is able to help the forensic analyst to assess possible states of each piece of evidence. In addition, thanks to the Bayesian network, the evidence that will have the most effect on final findings can be determined.

However, there are some limitations to applying the Bayesian network model despite its advantages. Such a model involves an individual’s subjective

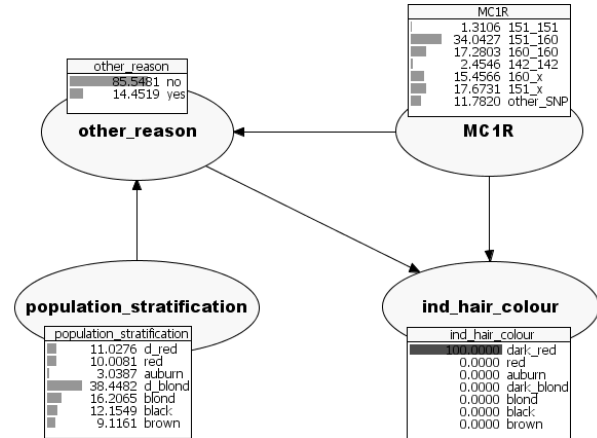


Fig. 7. Representation of probability tables connected with particular nodes after entering the “hard” evidence, that the individual has dark red hair colour. The probability that his/her MC1R genotype is e.g. 151/160 amounts to about 34% and that other factors influence his/her hair colour is assessed at 14.45%.

judgement in determining the causal relationships and preliminary assumptions. So different experts, even having at their disposal the same set of evidence, may find different outcomes, because of a different representation of the same problem. For example, the prior distribution of the MC1R genotype and different hair colour were assessed on the basis of examining 390 individuals; on the other hand, we are aware that this sample may not be representative of the whole population. Hence, further research and analyses are needed to enhance the inference capability and accuracy of the Bayesian network model before it can be applied in routine forensic casework.

The presented Bayesian network calculations were carried out on the basis of results obtained for a population sample from South Poland. Recent MC1R data for a population from Liguria (Italy) revealed 13 compound heterozygotes R/R and surprisingly only 2 of them resulted in red hair colour of the donor. This was explained by the authors by the action of other genes, which may mask the action of R type MC1R alleles and lead to low frequency of the RHC phenotype [15]. This finding has important implications for forensic science, as it proves that RHC prediction is very dependent on the population sample. The procedure should therefore start with definition of the population origin of the donor and then RHC prediction could be accomplished with appropriate certainty. An investigation always has to take into account the fact that the perpetrator may also be deliberately trying to mislead the investigation by dyeing hair.



#### 4. Concluding remarks

The performed experiments showed that the designed test allows rapid and relevant red hair colour prediction based on analysis of biological traces. At the same time, the test enables sex determination. The optimised test may also find application in studies performed in other fields like dermatology. The developed Bayesian network represents in a graphical way the relationships among the set of obtained results and allows us to infer about hair colour of the donor of a crime sample. The created network remains open for prediction of successive traits by genetic examination.

#### Acknowledgments

The study was financially supported by the Ministry of Education and Science in Poland, grants no. 0 T00C 018 29 and 0 T00C 028 29.

#### References

1. AmpFISTR® Identifier™ PCR Amplification Kit User's Manual, Applied Biosystems 2001.
2. Ansell R., Rape case management at the Swedish national laboratory of forensic science, *Problems of Forensic Sciences* 2006, 67, 255–272.
3. Bastiaens M., ter Huurne J., Gruis N. [et al.], The melanocortin-1-receptor gene is the major freckle gene, *Human Molecular Genetics* 2001, 10, 1701–1708.
4. Box N. F., Wyeth J. R., O'Gorman L. E. [et al.], Characterization of melanocyte stimulating hormone receptor variant alleles in twins with red hair, *Human Molecular Genetics* 1997, 6, 1891–1897.
5. Branicki W., Brudnik U., Wojas-Pelc A., Genetic prediction of pigmentary traits in forensic studies, *Problems of Forensic Sciences* 2005, 64, 343–352.
6. Branicki W., Brudnik U., Kupiec T. [et al.], Determination of phenotype associated SNPs in the MC1R gene, *Journal of Forensic Sciences* 2007, 52, 349–54.
7. Flanagan N., Healy E., Ray A. [et al.], Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation, *Human Molecular Genetics* 2000, 9, 2531–2537.
8. Gill P., Whitaker J., Flaxman C. [et al.], An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, *Forensic Science International* 2000, 112, 17–41.
9. Grimes E.A., Noake P. J., Dixon L. [et al.], Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype, *Forensic Science International* 2001, 122, 124–129.
10. Jensen F. V., Bayesian networks and decision Graphs, Springer-Verlag, New York 2001.
11. Jobling M. A., Gill P., Encoded evidence: DNA in forensic analysis, *Nature Reviews of Genetics* 2004, 5, 739–751.
12. Kanetsky P. A., Ge F., Najarian D. [et al.], Assessment of polymorphic variants in the melanocortin-1 receptor gene with cutaneous pigmentation using an evolutionary approach, *Cancer Epidemiology Biomarkers & Prevention* 2004, 13, 808–819.
13. Kanetsky P. A., Rebbeck T. R., Hummer A. J. [et al.], Population-based study of natural variation in the melanocortin-1 receptor gene and melanoma, *Cancer Research* 2006, 66, 9330–9337.
14. Lucy D., Introduction to statistics for forensic scientists, John Wiley and Sons, Chichester 2005.
15. Pastorino L., Cusano R., Bruno W. [et al.], Novel MC1R variants in Ligurian melanoma patients and controls, *Human Mutation* 2004, 24, 103.
16. Pulker H., Lareu M. V., Phillips C. [et al.], Finding genes that underlie physical traits of forensic interest using genetic tools, *Forensic Science International* 2007, 1, 100–104.
17. Rees J. L., Genetics of hair and skin color, *Annual Review of Genetics* 2003, 37, 67–90.
18. Rouzaud F., Kadekaro A. L., Abdel-Malek Z. A. [et al.], MC1R and the response of melanocytes to ultraviolet radiation, *Mutation Research* 2005, 571, 133–152.
19. Sanchez J. J., Borsting C., Hallenberg C. [et al.], Multiplex PCR and minisequencing of SNPs – a model with 35 Y chromosome SNPs, *Forensic Science International* 2003, 137, 74–84.
20. Steffen L., Lauritzen S. L., Sheehan N. A., Graphical models for genetic analysis, *Statistical Science* 2003, 18, 489–514.
21. Sturm R. A., Frudakis T. N., Eye colour: portals into pigmentation genes and ancestry, *Trends in Genetics* 2004, 20, 327–332.
22. Sullivan K. M., Mannucci A., Kimpton C. P. [et al.], A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin, *Biotechniques* 1993, 15, 636–638.
23. Taroni F., Aitken C. G. G., Garbolino P. [et al.], Bayesian networks and probabilistic inference in forensic sciences. Statistics in practice, John Wiley and Sons, Chichester 2006.
24. Tully G., Genotype versus phenotype: Human pigmentation, *Forensic Science International* 2007, 1, 105–110.

---

#### Corresponding author

Wojciech Branicki  
Instytut Ekspertyz Sądowych  
ul. Westerplatte 9  
PL 31-033 Kraków  
e-mail: wbranicki@ies.krakow.pl

---

## ZASTOSOWANIE SZYBKIEGO TESTU DO PRZEWIDYWANIA RUDEGO KOLORU WŁOSÓW I OZNACZANIA PŁCI W BADANIACH GENETYCZNO-SĄDOWYCH

### 1. Wstęp

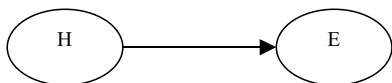
Ślady biologiczne zabezpieczane na miejscach zdarzenia kryminalnego są zazwyczaj poddawane badaniom identyfikacyjnym polegającym na analizie wysoce polimorficznych układów typu STR. Oznaczone profile DNA zidentyfikowane na miejscu przestępstwa są porównywane z profilami referencyjnymi, a stwierdzona zgodność jest rozumiana jako silny dowód przemawiający na korzyść hipotezy o wspólnym pochodzeniu analizowanych próbek. Jeśli w sprawie nie wytypowano żadnych podejrzanych i brak jest materiału porównawczego, oznaczone profile DNA z miejsca przestępstwa są zazwyczaj umieszczane w narodowej bazie danych DNA. Jeśli przeszukanie bazy profili porównawczych doprowadzi do odnalezienia profilu zgodnego z profilem DNA z miejsca przestępstwa, dowodowa wartość stwierdzonej zgodności jest również oceniana jako wysoka. W przypadku braku wykazania zgodności dochodzenie może ulec znacznemu spowolnieniu w związku z koniecznością zweryfikowania alternatywnych hipotez dotyczących osoby odpowiedzialnej za popełnienie przestępstwa. W niektórych poważniejszych sprawach stosowane są przesiewowe badania DNA, które pozwalają na wyeliminowanie kolejnych osób z kręgu podejrzanych. Takie rozwiązanie może prowadzić do konieczności analizy setek lub nawet tysięcy próbek i w skrajnych przypadkach trwać wiele lat [2]. Próbkę ujawnioną na miejscu zdarzenia kryminalnego mogą być jednak również wykorzystane do celów wywiadowczych, co może przyczynić się do ograniczenia liczby próbek podlegających badaniom przesiewowym. Współczesne testy identyfikacyjne pozwalają na oznaczenie płci, co umożliwia organom śledczym wysunąć hipotezę na temat płci osoby podejrzanej. W genetyce sądowej pojawiają się jednak nowe możliwości – genetyczne przewidywanie cech fizycznych. Bez wątplenia mogą one stać się bardzo użyteczne na etapie prowadzonego dochodzenia [5, 16, 24].

Cechy fizyczne powinny być traktowane jako cechy ilościowe, a więc trudne do analizy. Przyjmuje się na przykład, że pigmentacja, jedna z najbardziej różnicujących cech ludzkich, znajduje się pod kontrolą ponad 100 genów [18]. Niektóre z nich zostały już zidentyfikowane. Wykazano również ich związki z dziedziczeniem różnych cech pigmentacyjnych [17, 21]. Co ciekawe, dziedziczenie rudego koloru włosów u człowieka jest silnie związane z jednym genem, który koduje receptor melanokortyny 1 (MC1R) (16q24.3, MIM 155555). Liczne

badania przeprowadzone nad zmiennością genu MC1R wykazały, że polimorfizm ten odgrywa decydującą rolę w determinacji fenotypu przejawiającego się rudym kolorem włosów (RHC) u ludzi [np. 7, 9, 12]. Wcześniejsze badania autorów tego artykułu wykazały, że w populacji polskiej około 97,5% osobników charakteryzujących się genotypami 151/151, 160/160, 151/160 posiadało rudy kolor włosów [6]. Te dwa miejsca polimorficzne oraz kilka innych pozycji nukleotydowych w genie MC1R mają duże znaczenie predyktywne, a w związku z tym mogą mieć istotne znaczenie dla przewidywania fenotypu w badaniach sądowych. Zaproponowany wcześniej przez autorów tego artykułu test pozwala na szybką analizę pięciu pozycji typu SNP w genie MC1R o kluczowym znaczeniu w pojedynczej reakcji typu multipleks [6].

Niniejsza praca przedstawia rozszerzony test umożliwiający analizę jedenastu pozycji typu SNP w genie MC1R. Test pozwala również na analizę jednej pozycji typu SNP zlokalizowanej w pierwszym intronie genu amelogeniny, którego badanie pozwala na oznaczenie płci. Przeprowadzone doświadczenia walidacyjne miały na celu umożliwienie zastosowania testu w sądowych badaniach śladów biologicznych. W pracy również rozważany jest problem oceny wpływu, jaki mają uzyskane wyniki analizy genu MC1R na występowanie fenotypu RHC. W celu graficznego przedstawienia związków pomiędzy obserwowanymi danymi stworzono prostą sieć przyczynową pozwalającą na wnioskowanie na temat prawdopodobieństwa wystąpienia danego koloru włosów na podstawie znanego genotypu MC1R. Sieć bayesowska (ukierunkowany wykres acykliczny) składa się z następujących elementów: zestawu zmiennych przedstawionych w formie węzłów oraz zestawu strzałek łączących węzły, a ukazujących zależności między zmiennymi; każda zmienna posiada skończony zestaw wzajemnie wykluczających się stanów, do każdej zmiennej  $A$  zależnej od zmiennych  $B_1, \dots, B_n$ , dołączona jest tabela z prawdopodobieństwami warunkowymi  $Pr(A|B_1, \dots, B_n)$  [10]. Tak skonstruowana sieć przyczynowa może być stosowana w celu oceny, jak zmiana poziomu pewności w obszarze jednej zmiennej (np. rodzaju genotypu MC1R) może zmieniać poziom pewności w obszarze pozostałych zmiennych (np. prawdopodobieństwo posiadania określonego koloru włosów przez nieznanego dawcę próbki) [23]. Sieci Bayesowskie działają na zasadzie prawdopodobieństw warunkowych. Dlatego więc, jeśli występowanie dowodu  $E$  zależy od hipotezy  $H$ , to zależności

między nimi mogą zostać przedstawione poprzez sieć bayesowską jak na poniższym schemacie:



Powyższy schemat może zostać przedstawiony poprzez wzór:

$$Pr H, E \quad Pr H \quad Pr E|H \quad . \quad \{1\}$$

## 2. Metody

### 2.1. Próbkki

Od 390 niespokrewnionych osób rasy kaukaskiej zamieszkujących Polskę południową pobrano próbki w formie wymazów z jamy ustnej. Badania walidacyjne objęły różne rodzaje materiału biologicznego analizowanego w badaniach sądowych; szczegółowe dane na ten temat podano poniżej. Ekstrakcję DNA prowadzono albo z zastosowaniem manualnej metody krzemionkowej (A & A Bitechology, Gdańsk), albo z zastosowaniem aparatu Biorobot M48 (Qiagen, Hilden, Niemcy). Stężenie DNA mierzono przy zastosowaniu zestawu Quant-IT PicoGreen dsDNA Quantitation Kit (Invitrogen, Karlsruhe, Niemcy).

### 2.2. Konstrukcja testu

Na podstawie wyników własnych badań oraz danych z literatury wybrano jedenaście zmiennych pozycji DNA (mutacji typu *missense*, *nonsense* oraz mutacji zmiany ramki odczytu) zlokalizowanych w eksonie genu MC1R [7, 4, 3, 6]. Jak wykazano, siedem z tych mutacji silnie predysponuje do posiadania fenotypu RHC. Sześć z nich to miejsca polimorficzne wchodzące w skład reakcji multipleks 1 (86.1A, C252A, G425A, C451T, C478T, G880C), natomiast jeden zawarty jest w reakcji multipleks 2 (C456A). Te miejsca polimorficzne są określane w literaturze jako allele genu MC1R o wysokiej penetracji – „R” (są silnie skorelowane z rudym kolorem włosów i jasną karnacją) [13]. Dodatkowe cztery pozycje polimorficzne zawarte w reakcji multipleks 2 (G178T, G274A, T464C, G488A) są traktowane jako allele o niskiej penetracji – „r” (są słabiej skorelowane z fenotypem typu RHC). Reakcja multipleks 2 umożliwia również analizę polimorfizmu pojedynczego nukleotydu zlokalizowanego w pierwszym intronie genu amelogeniny, która pozwala na oznaczenie płci (rycina 1).

### 2.2.1. Reakcja PCR typu multipleks

Startery reakcji PCR zaprojektowano tak, aby umożliwić jednoczesną amplifikację trzech fragmentów genu MC1R zawierających komplet jedenastu pozycji typu SNP (tabela I) [6]. Reakcję PCR zoptymalizowano dla trzech fragmentów genu MC1R oraz pierwszego intronu amelogeniny (startery opisane przez Sullivan i współpracowników) [22]. Mieszanina reakcyjna PCR składała się z 5  $\mu$ l mieszaniny Multiplex PCR Master Mix, 1  $\mu$ l roztworu Q oraz 1  $\mu$ l mieszaniny starterów (tabela I). Reakcję PCR prowadzono, stosując następujący profil temperaturowy: {94°C/15 min, (94°C/0,30 min, 63°C/1,30 min, 72°C/1,30 min) 30 lub 34, 72°C/10 min} 4°C/ . 5 l produktu PCR rozdzielano na żelu agarozowym, a pozostałą objętość oczyszczano z zastosowaniem zestawu ExoSap IT kit (Amersham Pharmacia, Freiburg, Niemcy).

### 2.2.2. Reakcje minisekwencjonowania

Genotypowanie badanych pozycji typu SNP prowadzono z zastosowaniem dwóch reakcji typu multipleks polegających na wykorzystaniu metody wydłużania startera (ang. primer extension). Szczegóły dotyczące starterów wydłużających przedstawia tabela II. Oczyszczone produkty PCR poddawano reakcjom minisekwencjonowania z zastosowaniem zestawu SNaPshot multiplex PCR kit, stosując się do zaleceń producenta.

## 2.3. Badania walidacyjne

### 2.3.1. Niezawodność

Dwadzieścia próbek o znanych genotypach MC1R, które zostały uprzednio oznaczone z zastosowaniem reakcji cyklicznego sekwencjonowania zgodnie z procedurą opisaną przez Kanatsky i współpracowników [12], poddano analizie z wykorzystaniem zoptymalizowanych testów. Badane próbki wybrano celowo tak, aby reprezentowali byli mężczyźni, kobiety i wszystkie możliwe warianty MC1R.

### 2.3.2. Czułość

Testom poddano następujące stężenia DNA: 0,5, 0,25, 0,125, 0,06, 0,03, oraz 0,015 ng. W badaniach nad czułością testu wykorzystano standard 007 (Applera). Próbkki amplifikowano, stosując dwie opcje: 30 oraz 34 cykli reakcji PCR. Wydajność reakcji PCR oceniano z zastosowaniem żeli agarozowych, a następnie skrajne punkty poddawano reakcjom minisekwencjonowania.

### 2.3.3. Powtarzalność

Analizie poddano trzydzieści różnych próbek pobranych od tego samego osobnika. Oznaczony genotyp MC1R potwierdzono za pomocą reakcji cyklicznego sekwencjonowania.

### 2.3.4. Ślady biologiczne

Płynną krew naniesiono na bawełnę w trzech powtórzeniach i przechowywano w warunkach pokojowych (24 godziny, 7 dni, 1 i 3 miesiące) oraz w warunkach podwyższonej wilgotności i temperatury – łaźni wodnej z zadaną temperaturą 37°C (24 godziny, 48 godzin, 72 godziny, 7 dni). Zebrano również materiał biologiczny z używanych rzeczy osobistych oraz inne próbki regularnie badane w laboratoriach sądowych (maszynka do golenia, szczoteczka do zębów, łyżki, butelki, kubki, szminki, klucze magnetyczne, karty kredytowe, niedopałki papierosów, płamy nasienia). Wszystkie próbki poddano ekstrakcji DNA z zastosowaniem aparatu Biorobot M48 i przeanalizowano, wykorzystując zoptymalizowane testy. Badaniu poddano również pojedynczą próbkę – w tym przypadku ekstrakcję DNA prowadzono z zastosowaniem metody organicznej i z oczyszczaniem próbek na kolumnach Centricon 100.

### 2.4. Statystyczna ocena wpływu genotypu MC1R na cechy fenotypowe

Przyczynowa sieć bayesowska została przygotowana za pomocą programu Hugin, wersja 6.8. Sieć złożona jest z czterech węzłów: „MC1R”, „kolor włosów osobnika”, „stratyfikacja populacji” i „inna przyczyna” (rycina 2). Dla uproszczenia założono, że kolor włosów jest determinowany przez dwa czynniki: 1) genotyp MC1R i 2) działanie innych czynników, które z kolei pozostają pod wpływem poszczególnych wariantów genu MC1R oraz zróżnicowania populacji (odsetek osobników o kolorze włosów ciemno-rudym, rudym, szatyn, blond, czarnym w badanej populacji 390 osób). Autorzy zdają sobie sprawę, że analizowana próba populacyjna nie jest reprezentatywna dla populacji generalnej Polski południowej w związku z tendencją wynikającą z nadreprezentacji osób o rudym kolorze włosów. Tabela prawdopodobieństwa dla węzła MC1R zawiera prawdopodobieństwa *a priori* na temat rozkładu wariantów SNP w populacji. Stąd tabela prawdopodobieństw zawiera 7 wzajemnie wykluczających się stanów: 151/151, 151/160, 160/160, 142/142, 151/inny i 160/inny. Tabela warunkowych prawdopodobieństw dla węzła „kolor włosów osobnika” zawiera wszystkie prawdopodobieństwa warunkowe dla  $Pr(\text{kolor włosów osobnika} \mid MC1R, \text{inna przyczyna})$ , co przedstawia tabela III. Po ustaleniu mocnego dowodu w którymkolwiek z węzłów (np. w próbce z miejsca zda-

rzenia kryminalnego stwierdzono obecność genotypu 160/160) i rozprzestrzenieniu się jego wpływu na całą sieć, można stwierdzić prawdopodobieństwa wystąpienia poszczególnych kolorów włosów u osoby będącej źródłem próbki.

## 3. Wyniki i dyskusja

### 3.1. Badania walidacyjne

Poprzednie badania przeprowadzone przez autorów tej pracy na grupie 80 osób o rudym kolorze włosów i populacji kontrolnej ponad 200 osób obu płci potwierdziły, że w większości przypadków przyczyna występowania rudego koloru włosów może być wyjaśniona poprzez warianty regionu kodującego genu MC1R [6]. Niektóre genotypy są silnie skorelowane z rudym kolorem włosów. Badania autorów oraz dane z literatury wskazują, że wartość predykcyjna jest szczególnie silna w przypadku wariantów klasyfikowanych jako „R” (o wysokiej penetracji) występujących jako homozygoty lub kombinacje heterozygotyczne, do których należą 86.1insA, C252A, G425A, C451T, C456A, C478T i G880C. Niniejsza praca przedstawia test wykorzystujący zestaw SNaPshot, który zoptymalizowano w celu wybiórczej analizy jedenastu pozycji typu SNP związanych z dziedziczeniem rudego koloru włosów oraz pojedynczej tranzycji w genie amelogeniny, co pozwala na oznaczanie płci. Przykłady elektroferogramów przedstawiających wyniki analiz przeprowadzonych z zastosowaniem testu minisekwencyjnego 1 i 2 (GeneScan) przedstawia rycina 3.

Wyniki reakcji SNaPshot uzyskane dla 20 próbek reprezentujących różne genotypy MC1R były zgodne z wynikami cyklicznego sekwencjonowania w 19 przypadkach. W jednym przypadku odnotowano brak zgodności – cykliczne sekwencjonowanie wykazało obecność homozygoty 880C, podczas gdy minisekwencjonowanie wykazało obecność złożonej heterozygoty 451C/T i 880G/C. Powtórna analiza wykazała, że uzyskany za pierwszym razem wynik cyklicznego sekwencjonowania był nieprawidłowy. Najbardziej prawdopodobną przyczyną zaistnienia problemu było wypadnięcie jednego z alleli (*allelic dropout*; brakujący allel 451T, 880G).

Przygotowany test pozwala na oznaczenie płci. Homozygota T charakterystyczna jest dla próbek pochodzących od kobiet (podwójny X), podczas gdy heterozygota C/T jest charakterystyczna dla próbek pochodzących od mężczyzn (rycina 4). Badania wykazały prawidłowe oznaczenie płci z zastosowaniem zoptymalizowanej metody we wszystkich analizowanych próbkach, których całkowita liczba przekroczyła 100.

Również w przypadku wszystkich 33 próbek pobranych od jednego osobnika uzyskano ten sam prawidłowy wynik genotypowania. Uzyskane wyniki potwierdziły

wysoką niezawodność i powtarzalność walidowanej metody minisekwencjonowania.

Standardowo rekomendowane stężenie matrycowego DNA w przypadku zestawów do multipleksowej analizy układów typu STR wynosi co najmniej 0,5 ng [1]. Przeprowadzone eksperymenty potwierdziły, że takie stężenie jest wystarczające do uzyskania prawidłowych wyników działania zoptymalizowanych testów przy 30 cyklach reakcji PCR. Modyfikacja polegająca na podniesieniu liczby cykli reakcji PCR (34 cykle) jest stosowana w genetyce sądowej w celu analizy próbek zawierających minimalną ilość DNA klasyfikowanych zgodnie z przyjętymi zasadami jako LCN – (niska liczba kopii; 100 pg) [8]. Zgodnie z oczekiwaniami, czułość walidowanego testu była znacznie wyższa, gdy zamiast 30 cykli zastosowano 34 cykle reakcji PCR. Po zastosowaniu tej modyfikacji prawidłowe wyniki badań uzyskiwano z zaledwie 15 pg matrycowego DNA. Przeprowadzone doświadczenia wykazały, że czułość testu jest na poziomie charakterystycznym dla zestawów dostępnych w handlu i stosowanych w genetyce sądowej.

W laboratoriach sądowych często analizowane są ślady biologiczne, które są narażone na działanie silnych procesów gnilnych. Wielokrotnie obserwowano, że amplifikacja tego typu próbek o niskiej jakości daje negatywne wyniki, gdy celem jest namnożenie długich fragmentów DNA. W związku z tym amplifikacja całego eksonu MC1R nie powinna być brana pod uwagę jako skuteczna metoda analizy w przypadku sądowej analizy śladów biologicznych. Zoptymalizowana przez autorów niniejszej pracy metoda polega na amplifikacji trzech fragmentów genu MC1R zawierających wszystkie badane pozycje typu SNP. Długości ampikonów nie przekraczają 270 pz. Metoda została poddana testom przy zastosowaniu typowych próbek o charakterze sądowym (tabela IV). Eksperymenty wykazały prawidłowe działanie metody w przypadku 52 z całkowitej liczby 61 próbek. W sześciu przypadkach amplifikacja dała wynik negatywny. Dotyczyło to głównie najbardziej problematycznych próbek typu LCN (3 wymazów z pierścionków, 1 wymazu z kolczyków, 1 wymazu z karty kredytowej i 1 wymazu z kubka). W trzech przypadkach oznaczony genotyp MC1R był niezgodny z rzeczywistym genotypem dawcy. Taki wynik uzyskano w przypadku wymazu pobranego z kubka oraz wymazów z kart kredytowych. Szczegółowa analiza tych próbek z zastosowaniem zestawu SGM Plus wykazała, że nieprawidłowe wyniki analizy MC1R musiały wynikać z obecności mieszanin DNA. Karty kredytowe oczywiście mogły być dotykane, poza ich właścicielami, przez sprzedawców w sklepach. Bardziej zaskakujące było uzyskanie mieszaniny DNA na analizowanym kubku, który zgodnie z informacjami, powinien być używany przez jedną osobę przed pobraniem wymazu.

Przeprowadzone doświadczenia wykazały, że metoda jest wystarczająco czułym i wydajnym narzędziem do analizy zdegradowanych próbek sądowych. Pokazały jednak również, że analiza mieszanin DNA może prowadzić do uzyskania nieprawidłowych wyników analizy. W niektórych przypadkach na mieszaninę DNA pochodzącego od mężczyzny i kobiety może wskazywać silna dysproporcja w sile sygnału pomiędzy pikami dla nukleotydów C i T, które odnoszą się do analizowanej pozycji SNP w genie amelogenu (dane nieprezentowane). W większości przypadków wyniki testu nie są w stanie ujawnić obecności ewentualnej mieszaniny DNA. W związku z tym próbka, która przeznaczona jest do analizy genotypu MC1R, powinna zostać uprzednio zbada na obecność mieszaniny DNA, a próbki, które zawierają taką mieszaninę, powinny zostać wykluczone z analizy MC1R. Badaniom poddano również jedną sprawę dotyczącą identyfikacji osoby zaginionej. Standardowa analiza markerów typu STR wchodzących w skład zestawu Identifiler wykazała pokrewieństwo typu rodzice-dziecko pomiędzy analizowanymi szczątkami ludzkimi (kość udowa) a domniemanymi rodzicami. Przeprowadzone badania z zastosowaniem zoptymalizowanych testów umożliwiających analizę MC1R oraz amelogenu prawidłowo oznaczyły, że badane szczątki pochodzą od mężczyzny oraz wykazały obecność genotypu 451C/T genu MC1R, który był prawidłowy w świetle genotypów MC1R charakterystycznych dla domniemanych rodziców zaginionego.

### 3.2. Sieć bayesowska jako narzędzie do oceny wyników genotypowania genu MC1R

Prawidłowa ocena wyników genotypowania jest podstawowym zagadnieniem z punktu widzenia badań sądowych. Homozygotyczne genotypy dla polimorficznych pozycji sklasyfikowanych jako R (86.1insA, C252A, G425A, C451T, C456A, C478T, G880C), jak również heterozygotyczne kombinacje tych wariantów, silnie przemawiają na korzyść hipotezy, że dawca próbki ma rudy kolor włosów i jasną skórę (fenotyp RHC). Nawet jednak w przypadku genotypów R/R prawdopodobieństwo wystąpienia fenotypu RHC nie jest równe 100%, w związku z czym wynik musi być poddany rzetelnej ocenie tak, aby nie wprowadzić w błąd organu dochodzeniowego. Jednym z istotnych celów niniejszej pracy było przedstawienie modelu wnioskowania polegającego na uwzględnieniu wyników analizy genetycznej przez zastosowanie sieci bayesowskiej. Zaproponowany model może dostarczyć naukowego uzasadnienia dla wyników lub wniosków analizy genotypu genu MC1R. Co więcej, stworzony model może zostać wykorzystany przez innych genetyków sądowych, którzy mają możliwość rozbudowania go o informacje na temat wpływu innych genów na cechę fenotypową, jaką jest kolor włosów.

Sieci bayesowskie są bardzo plastyczne pod względem wprowadzania dodatkowych zmiennych służących do oceny i wykorzystania dodatkowej wiedzy dla potrzeb oceny wniosków. Na podstawie: 1) penetracji wariantów R wśród osobników o rudym kolorze włosów, 2) częstości populacyjnych dla poszczególnych kolorów włosów oraz 3) znanego rozkładu częstości genotypów MC1R w populacji, przeprowadzono wnioskowanie na temat możliwości wystąpienia danego koloru włosów u osoby stanowiącej źródło nieznaną próbkę zabezpieczonej na miejscu zdarzenia kryminalnego. Rycina 5 przedstawia informację *a priori* związaną z węzłami „rodzicami” i wynikające z ich zastosowania prawdopodobieństwa warunkowe w pozostałych węzłach przed wprowadzeniem tzw. „pewnego” dowodu.

Dowód w przypadku zmiennej jest określeniem poziomu wiarygodności dla poszczególnych stanów. Jeśli zmienna jest skonkretyzowana (stan zmiennej jest stanem pewnym, tj. jego prawdopodobieństwo jest równe 100%), wówczas uzyskuje miano „pewnego” dowodu; w przeciwnym razie nosi miano dowodu „miękkiego” [10]. Na przykład, kiedy rozpatrujemy próbkę o genotypie MC1R 151/151, wówczas pierwsze przybliżenie prawdopodobieństwa, że jej dawca ma ciemnorudy kolor włosów, będzie równe 57%, natomiast, że ma rudy kolor włosów, wyniesie 43%. Wynika to z uzyskanych przez autorów niniejszej pracy danych, które wskazują, że w badanej próbie populacyjnej nie wystąpiły osoby, które posiadałyby genotyp MC1R 151/151 i kolor włosów inny niż rudy lub ciemnorudy. W populacji występowały jednak osoby rude, które nie posiadały genotypu MC1R 151/151. Należy rozróżnić bowiem dwa prawdopodobieństwa warunkowe:  $Pr(\text{rude włosy} \mid \text{genotyp MC1R 151/151})$  i  $Pr(\text{genotyp MC1R 151/151} \mid \text{rude włosy})$ . Pionowa linia oznacza „pod warunkiem, że” [14]. Wynikiem pierwszego prawdopodobieństwa jest wartość 43%, natomiast drugiego 7%. Jednak wartości prawdopodobieństw uzyskane w rozpatrywanym modelu po wprowadzeniu „pewnego” dowodu 151/151 na temat genotypu MC1R są równe 55,5%, gdy osoba posiada rudy kolor włosów i 44,5%, gdy osoba posiada ciemnorudy kolor włosów. Wynika to z faktu, że prawdopodobieństwa te są skorygowane poprzez wiedzę na temat rozkładu częstości genotypów w badanej populacji oraz poprzez działanie innych przyczyn (rycina 6). Wobec powyższego wiadomość dla policji o charakterze operacyjnym jest w przypadku genotypu 151/151 raczej stanowcza: „napastnik powinien mieć ciemnorudy lub rudy kolor włosów”.

Inne informacje, które można uzyskać dzięki analizie sieci bayesowskiej, wynikają z tak zwanego kierunku diagnostycznego. Jeżeli do sieci wprowadzimy informację, że dana osoba ma ciemnorudy kolor włosów, wówczas prawdopodobieństwo, że posiada ona genotyp MC1R 151/151 jest równe 1%; że posiada ona genotyp

151/160 jest równe około 34%; 160/160 – 17% itd. (rycina 7). Oczywiście ten kierunek wnioskowania jest nieużyteczny w praktyce sądowej.

Skonstruowana sieć może zostać bez trudu rozbudowana, jeśli dostępne będą dodatkowe informacje. Na przykład, gdy uzyskamy wiedzę na temat wpływu innych genów na kolor włosów, zostanie to wykorzystane do rozszerzenia dotychczasowej wiedzy na temat czynników istotnych w tej kwestii. Utworzony model jasno przedstawia probabilistyczne wartości oraz przyczynowe zależności pomiędzy hipotezami a materiałem dowodowym. Model sieci bayesowskich jest w stanie pomóc biegłemu w ocenie najbardziej prawdopodobnych stanów każdego z dowodów po uwzględnieniu znanego („pewnego”) dowodu i pozwala na przeprowadzenie probabilistycznej analizy następujących potem oszacowań. Co więcej, dzięki sieci bayesowskiej można określić ten dowód, który będzie miał największy wpływ na ostateczne ustalenia.

Naturalnie, mimo istotnych zalet, istnieją również pewne ograniczenia w stosowaniu modelu sieci bayesowskich. Taki model wymaga subiektywnej oceny przy wyznaczaniu związków przyczynowych oraz przy formułowaniu wstępnych założeń. Zatem dwaj eksperci dysponujący tym samym materiałem dowodowym mogą znaleźć różne rozwiązania, co spowodowane będzie ich różnym subiektywnym podejściem do tego samego problemu. Na przykład ocena częstości genotypów oraz różnych kolorów włosów została przeprowadzona na podstawie analizy materiału pobranego od 390 osób, ale należy pamiętać, że próba ta może nie być reprezentatywna dla całej populacji. Dlatego więc, zanim stworzony model sieci bayesowskiej zostanie zastosowany do rutynowych badań sądowych, należy poprzez dalsze badania podnieść siłę wnioskowania oraz dokładność stworzonego modelu sieci.

Przedstawione obliczenia prowadzone w ramach stworzonej sieci bayesowskiej prowadzono w oparciu o wyniki uzyskane dla próby populacyjnej z regionu Polski południowej. Uzyskane niedawno dane na temat zmienności genu MC1R w populacji z terenu Ligurii (Włochy) wykazały obecność 13 heterozygot złożonych R/R i nieoczekiwanie zaledwie w dwóch przypadkach taki genotyp związany był z rudym kolorem włosów. Autorzy tłumaczą to zjawisko działaniem innych genów, które mogą maskować wpływ alleli R genu MC1R i w ten sposób prowadzić do niskiej częstości fenotypu RHC w tej populacji [15]. Opublikowane dane mają jednak istotne znaczenie dla genetyki sądowej, gdyż wykazują, że stopień pewności przewidywania fenotypu RHC jest bardzo zależny od próby populacyjnej. Cała procedura analizy nieznaną próbkę powinna zatem zaczynać się od zdefiniowania jej populacji źródłowej. Po takich ustaleniach przewidywanie fenotypu RHC mogłoby zostać przeprowadzone z uzyskaniem odpowiedniej pewności.

Należy wreszcie brać pod uwagę fakt, że napastnik może chcieć umyślnie wprowadzić w błąd organy prowadzące dochodzenie i zafarbować włosy.

#### **4. Uwagi końcowe**

Przeprowadzone doświadczenia wykazały, że zaprojektowana metoda pozwala na szybkie i skuteczne przewidywanie występowania rudego koloru włosów na podstawie analizy śladów biologicznych. Jednocześnie test ten pozwala na oznaczanie płci. Zoptymalizowana metoda może poza genetyką sądową znaleźć również zastosowanie w innych dziedzinach wiedzy, na przykład w dermatologii. Opracowana sieć bayesowska przedstawia w formie graficznej związki pomiędzy uzyskanymi wynikami i pozwala na wnioskowanie co do koloru włosów osoby stanowiącej źródło próbki pobranej z miejsca przestępstwa. Stworzona sieć pozostaje otwarta na przewidywanie kolejnych cech poprzez badania genetyczne.

#### **Podziękowanie**

Projekt był finansowany ze środków przyznanych przez Ministerstwo Edukacji i Nauki w Polsce, granty o numerach 0 T00C 018 29 i 0 T00C 028 29.