Department of Medical Biology, Faculty of Health Sciences

Forensic DNA phenotyping

Evaluation of the performance of two prediction models, the IrisPlex and the novel EC12 model, for eye colour predictions in a Norwegian population

Marianne Kjersem

Master thesis in Biomedicine, MBI-3911, May 2020

Acknowledgement

The work for this master thesis was carried out at Centre of Forensic Genetics, The Artic University of Norway, Tromsø (UiT) from August 2019 to June 2020 and in a collaboration with Section of Forensic Genetics, Copenhagen, Denmark, under the primary supervision of Gunn-Hege Olsen.

First and foremost, I would like to thank my supervisors Thomas Berg, Kirstin Janssen, Claus Børsting and Nina Mjølsnes Savo. Thank you, Nina, for making the ROC-curves, calculating the AUC-values and PIE-scores and all your guidance throughout this Masters' thesis. I will also send a special thank you to my primary supervisor Gunn-Hege Olsen. Thank you for your endless patience, your availability and all your time spent proofreading the thesis.

Secondly, I would like to thank Section of Forensic Genetics, Copenhagen Denmark for letting me be able to stay there for seven weeks and learn the SBE method. A special thank you to Olivia Strunge Meyer and Jeppe Dyberg Andersen for making me feel so welcome and always be willing to answer my questions during my stay and also after heading back to Tromsø and Centre of Forensic Genetics.

Thank you, Mikkel Meyer Andersen at Department of Mathematical Sciences, Aalborg University, Denmark for calculating the p-values for the EC12 prediction model used in this study.

Lastly, my gratitude goes to my classmates, and especially Miriam Nilsen, Arja Arnesen Løchen and Lennart Van Ligtenberg for encouraging me through these two years. I am also grateful for the endless support and understanding from my family throughout these years.

Marianne Kjersem
Tromsø, May 2020

Table of contents

Abstract

Biological material obtained from a crime scene is used to generate DNA-profile by typing short tandem repeat (STR) markers. However, if the STR-profile do not match the DNA profile of suspects or from a crime DNA database, the investigation can go towards typing markers that can estimate externally visible characteristics (EVCs). EVCs can function as a "biological witness" and thus aid a police investigation.

In this work the IrisPlex prediction model and a novel prediction model, EC12, were evaluated in 521 samples from the Norwegian population. A PCR-SBE-CE assay amplifying the fourteen SNPs included in the two models was optimised at Section of Forensic Genetics, Copenhagen, Denmark before it was established at Centre of Forensic Genetics, Tromsø, Norway.

IrisPlex showed high prediction accuracy for blue and brown eye colour (AUC-value of 0.84 and 0.94, respectively). However, the model did not perform good in prediction of intermediate eye colour (AUC-value of 0.6), which represented 24% of the Norwegian population and thus all these individuals were incorrectly predicted.

Comparison of EC12 and an adjusted IrisPlex model (IP NO) showed a small increase in correct predictions from 72% to 75%, respectively. A higher prediction accuracy for all eye colours were observed for the EC12 model, with AUC-value of 0.84 (blue), 0.97 (brown) and 0.68 (intermediate), while IP NO obtained AUC-values of 0.81 (blue), 0.93 (brown) and 0.59 (intermediate). This increase may imply that the additionally SNPs included in this model has an improving effect on eye colour prediction. However, the prediction of intermediate eye colour was still not good, indicating the importance of further phenotypic investigation of this category.

Abbreviations

STR - Short tandem repeat

EVC - Externally visible characteristic

FDP - Forensic DNA phenotyping

SNP - Single nucleotide polymorphism

A - Adenine

T - Thymine

C - Cytosine

G - Guanine

UVR – Ultraviolet radiation

EM - Eumelanin

PM - Pheomelanin

ROC – Receiver under the operating

AUC - Area under the curve

P-value – Probability value

PIE - Pixel Index of the iris

SBE - Single base extension

MPS - Massive parallel sequencing

PCR - Polymerase chain reaction

dNTP - Deoxyribonucleotide triphosphate

ddNTP - Dideoxynucleotide triphosphate

SAP - Shrimp Alkaline Phosphate

CE - Capillary electrophoresis

Bp – Base pair

GMID-X - GeneMapper ID-X

RFU - Relative fluorescent unit

qPCR - Quantitative PCR

DIAT - Digital Iris Analysis Tool

LOOCV - Leave one out cross validation

IP NO – IrisPlex Norwegian population

Genes

TYR - Tyrosinase

MITF - Microphtalmia-associated transcription factor

DCT - Dopachrome tautomerase

TYRP1 - Tyrosine related protein

cAMP - Cyclic adenosine mono phosphate

ATP - Adenosine triphosphate

AC - Adenylate cyclase

αMSH - α-melanocyte stimulating hormone

MC1R - Melanocortin 1 receptor

ASIP - Agouti signalling protein

IRF4 - Interferon regulatory factor 4

OCA2 - Oculocutaneous albinism II

HERC2 - Hect domain and RLD2

SLC24A4 - Solute carrier family 24, member 4

SLC24A5 - Solute carrier family 24, member 5

SLC24A2 - Solute carrier family 45, member 2

1. Introduction

1.1 General introduction

After a grown-up man had been a night out in the northern parts of Norway, he was found murdered in his own apartment. There were no eyewitness or obvious suspects. When the police and the technicians searched the crime scene, they found the possible murder weapon, a bloody knife. The Centre of forensic Genetics analysed the biological material on the knife and generated a single source DNA-profile by typing short tandem repeat (STR) markers (1). The STR-profile was compared to profiles in the crime DNA database, but there was no match. Thus, the perpetrator could not be identified from the DNA trace using standard methods. The question is, can the biological material obtained from the crime scene help the police investigation when there are no eyewitnesses nor suspects?

In these kinds of cases, the investigation can turn towards alternative methods and involve DNA typing of loci that may be used to estimate externally visible characteristics (EVCs). This is also called forensic DNA phenotyping (FDP) and can function as a "biological witness" (2). FDP do not only have to involve murderer cases, but may also be helpful in all types of criminal investigations (3). Additionally, FDP can be helpful in cases where there are human eyewitnesses, who may be unreliable (4). Humans can have different perception of phenotypic traits and it may be difficult to remember details about a person's phenotypic characteristics. Also, the circumstances, such as lighting and distance, surrounding the criminal act can make it difficult to see all details.

The pigmentary traits are among the most prominent visible characteristic in humans and often used to describe individuals, groups of individuals and populations. FDP can therefore be a helpful tool in police investigation by narrowing down a pool of suspects (5). A standard STR-profile, used to identify the donor of the biological material, do not contain any phenotypic information (6). Therefore, other markers or regions in the human genome that reveals genetic variation and helps differentiate individuals from one another, must be analysed.

1.2 Genetic variation within the human genome and its influence on the phenotype

Only 1% of the human genome varies between different individuals (7). The genetic variations can include variations that may influence the human health or even be pathogenic. This is often

variations that involve larger parts of the genome, such as losses or expansion of whole chromosomes, referred to as structural changes (8). However, most of the differences do not affect human health, but can affect e.g. human phenotypes. Single nucleotide polymorphisms (SNPs) are one of these genetic variations which may affect a phenotype. Some SNPs are therefore helpful genetic markers in the forensic field, as they can be used to estimate EVCs.

A SNP is a sequence variation between individuals in a specific position in the genome where two, occasionally three, different bases can occur (8,9). However, the variation is only called a SNP if the least common allele has a frequency of at least 1%. If the frequency of the least frequent allele is less than 1%, it is called a rare mutation (10).

SNPs are highly abundant and estimated to occur at 1 out of every 1000 bases in the human genome, which make them account for more than 90% of all differences between unrelated people (9,10). The majority of SNPs are biallelic, meaning that they have two possible alleles and therefore three possible genotypes (Figure 1). For example, if the alleles for a SNP locus is A and B, where both alleles can be either an A (adenine), T (thymine), C (cytosine) or G (guanine) nucleotide, the three possible genotypes would be AA, AB and BB (11).

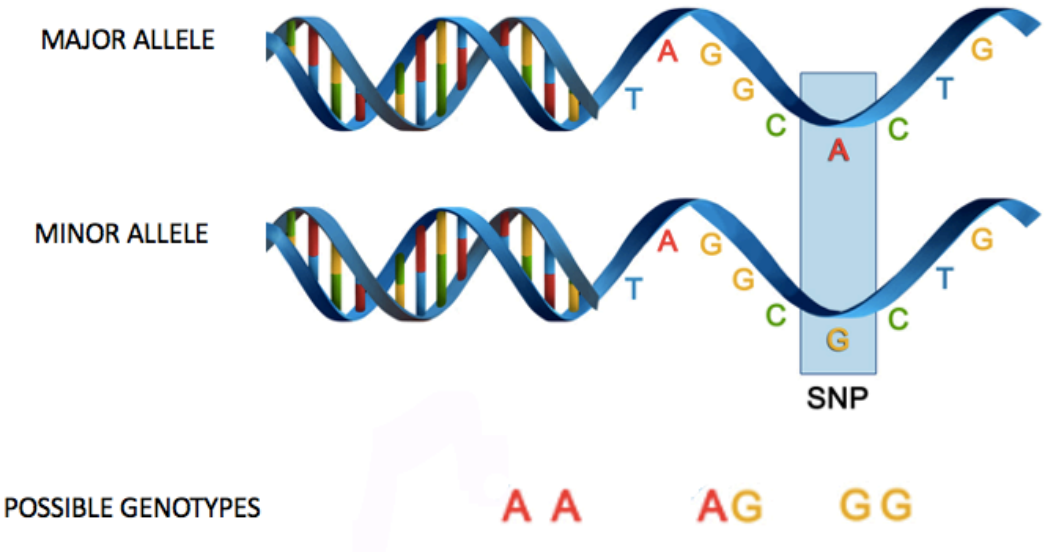


Figure 1: *Illustration of a biallelic SNP. The most frequent allele A and the least frequent allele G make the three possible genotypes AA, AG and GG. Figure modified from* (12).

Depending on where a SNP is located, it may have different consequences at the phenotypic level. If a SNP for instance is found in the noncoding region of the genome, it could be located

in promotors and transcription factor binding sites and thereby affect the phenotypic outcome (13). However, SNPs located in the coding regions can either be synonymous, which do not affect the phenotype, or nonsynonymous. A nonsynonymous SNP will changes the genetic code for an amino acid, known as a codon, (either for a new amino acid or to a nonsense mutation), and thereby change the amino acid sequence incorporated into the polypeptide (13). Nevertheless, most SNPs affecting phenotypes such as eye, hair and skin colour are found in introns without functional evidence for causal trait involvement (3). These SNPs most likely provide information of phenotypes due to physical linkage with other, unknown, SNPs involved in human pigmentation (3). As some SNPs can portray an EVC, they are interesting genetic markers within the forensic field.

1.3 Eye colour distribution throughout the world

Most EVCs are complex traits, meaning that these phenotypic characteristics are influenced by several genetic and environmental factors. This is also the case for the pigmentary traits, although they are genetically considered as the simplest EVCs. The traits involve skin, hair and eye colour and were often considered to be simple mendelian traits inherited in a one-gene to one-trait manner. However, among humans, there are many different eye colours, and the colour varies from the lightest of blue to the darkest of brown (Figure 2). (5,14)

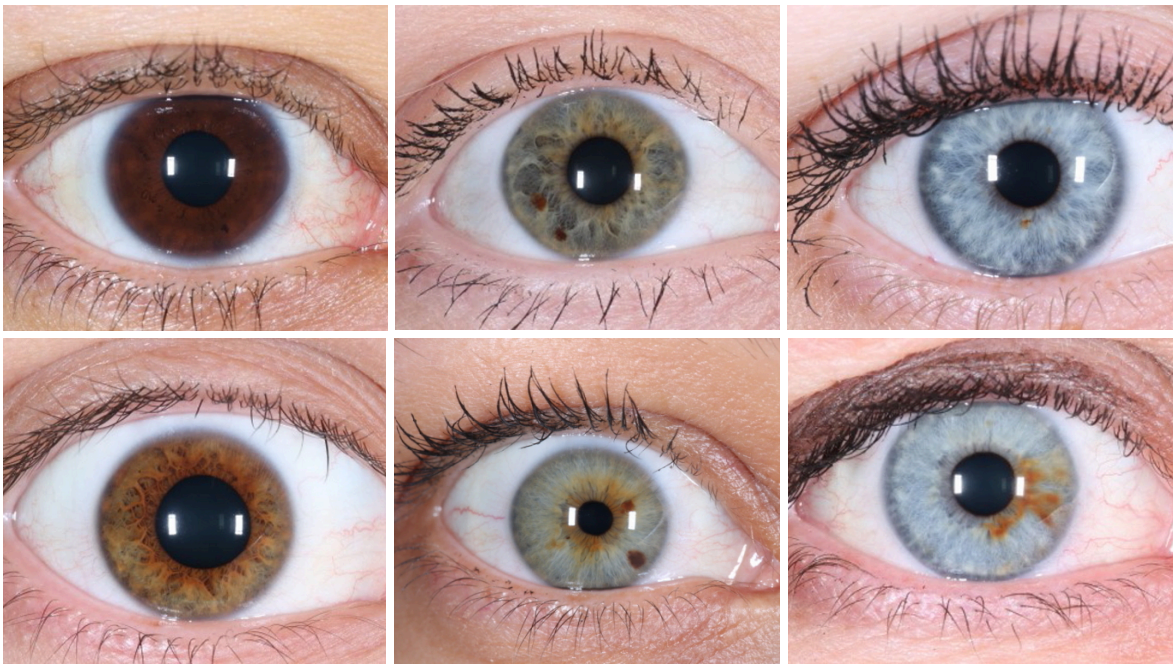


Figure 2: *Representative eye colours ranging from light blue to dark brown. The photos are taken with a Canon EOS 5D camera with a Canon macro lens 100mm 2.8L IS and Macro Twin Lite MT-24 EX blitz (Canon). (Photos from Centre of forensic genetics, UiT).*

The pigmentation of the eye varies all over the world and is considered as a polygenetic trait. Nevertheless, brown iris colour is most widespread worldwide and is assumed to reflect the ancestral human state (3,14,15). This is described by the out-of-Africa hypothesis which implies that human species evolved to its modern form in East Africa, before they emigrated to the rest the world 70.000 years ago (16). However, the eye colour in Europeans are unusually diverse, especially in the northern and eastern parts (Figure 3). At this continent, the frequency of the ancestral human eye colour is lower (3). The non-brown eye colour found here is assumed to be of European origin. Today, this colour is thought to occur as a result of positive selection for rare colour preferences regarding mate choice (3,17). It is presumed that colourful traits or bright colour, stimulate sexual attraction and individuals will select the one that "stands out from the crowd" (17). Another theory for development of light-coloured eyes is the vitamin D hypothesis (18,19). The hypothesis suggests that light eye, skin and hair colour co-evolved as humans migrated to areas where they were less exposed to ultraviolet radiation (UVR). UVR is important for vitamin D synthesis, and individuals with lighter colour will be more exposed to UVR.

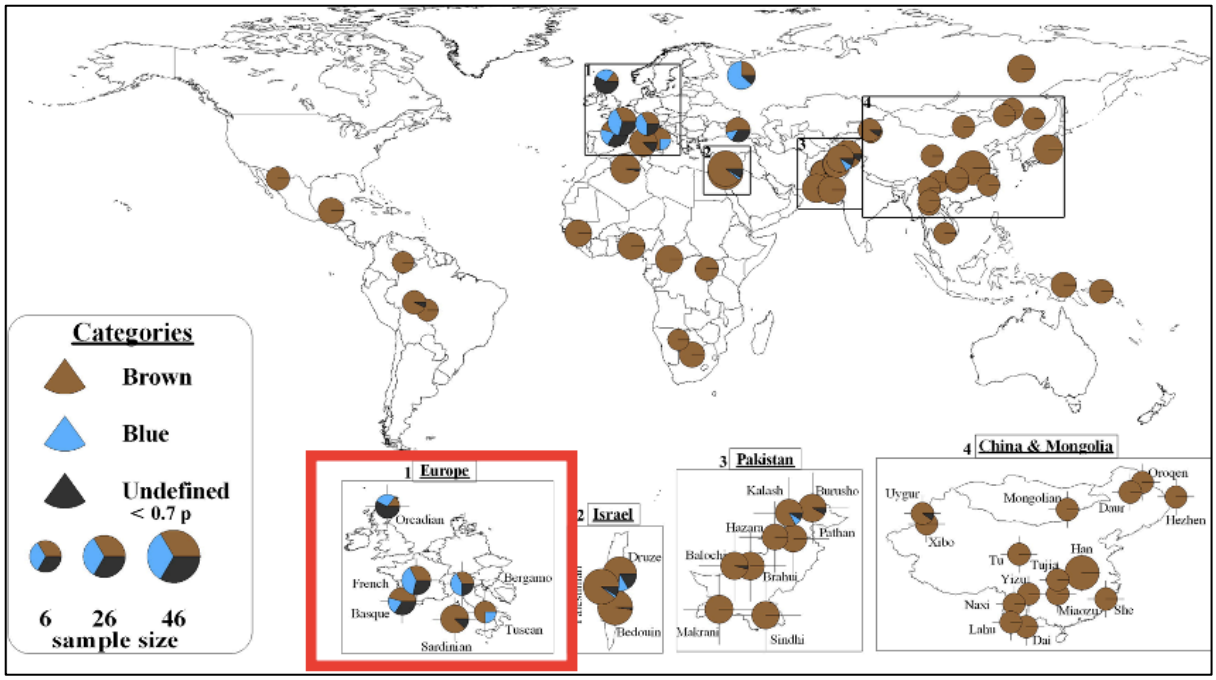


Figure 3: *Worldwide distribution of eye colour. Brown eye colour is seen all over the world, whereas blue and intermediate is mainly found in Europe (red square) and East Asia. Figure modified from (3).*

1.4 Melanin and the melanogenic pathway

The eye colour is mainly affected by the genetics and not that much by environmental factors. Therefore, eye colour prediction tools with great potential can be developed as long as the genetics are understood. Successful prediction of EVCs depends on the genetic markers included in the prediction model. The genetic markers involved, have to carry information that affects the phenotypic outcome. For eye colour, it is mainly the light absorbing biopolymer melanin that affects the colour (20).

Variable quantity, packing and quality of the melanin pigment results in variations in eye colour (21). The production of melanin, the melanogenesis, takes place in intracellular organelles called melanosomes which are found in melanocyte cells. Melanin also affects hair and skin colour as the melanocytes are found in several tissues including hair follicles, the basal layer of epidermis of the skin and in the iris. In mammals, two different types of melanin are synthetized, eumelanin (EM) and pheomelanin (PM). The brownish/black EM is responsible for dark colourisation and is packed in eumelanosomes. It almost absorb the full light spectrum, whereas the reddish yellow PM reflects a much broader light spectrum (15,21,22). EM mainly influence the skin and eye colour, whereas PM influence the hair colour (5). Different levels of PM and EM are affected by the numbers of melanosomes and the activity of melanogenesis

in the various tissues. E.g. an iris with few melanosomes and minimal pigment will be observed as blue eye colour, whereas an iris with a larger number of melanosomes and high level of pigment will be observed as brown eye colour (Figure 4) (20). However, in eyes were the colour appear neither blue nor brown, herein called intermediate eye colour, the amount of EM in the melanocytes vary (5). This means that some areas could be brown while other areas could be blue. As a result, humans will observe this eye colour combination as green or hazel.

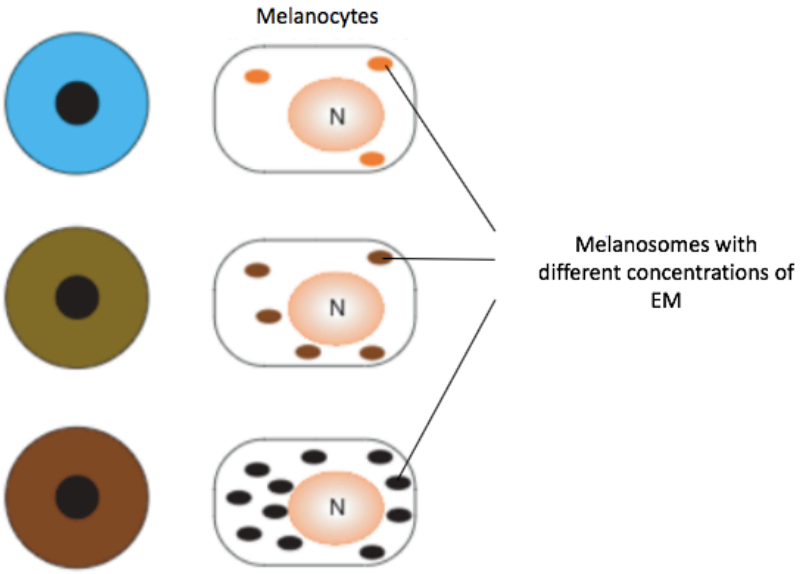


Figure 4*: Three different eye colours and their iridial melanocytes. The blue eye has melanocytes with few melanosomes and minimal levels of EM/pigment. The greenish eye has melanocytes with some melanosomes and a higher level om EM. The brown eye has melanocytes with a high number of melanosomes and a high level of EM. (N= cell nucleus). Figure modified from (20).*

1.4.1 Melanogenic pathway

Melanogenesis is a complex process that is controlled and regulated by many genes (Figure 5). The key protein in the pathway is tyrosinase (*TYR*), which is a membrane protein located in the melanosome membrane (23). *TYR* catalyses the conversion of tyrosine to dopaquinone, the last common step in the production of EM or PM. If cysteine is present, dopaquinone will spontaneously be converted to cysteinyldopas and the melanosomes will produce PM (24). However, if cysteine is depleted, production of EM can occur. As *TYR* is so important in the melanogenic pathway, any mutation found in *TYR* that affects the function and stability of the protein will have a major effect in the pigmentary phenotype (5). E.g. the SNP rs1126809 was

shown to have effect on eye colour and sun sensitivity. Another SNP located in the *TYR* region, rs7120151 had significant association with eye colour (25,26).

Microphtalmia-associated transcription factor (*MITF*) stimulates the transcription of *TYR*, but also transcription of *dopachrome tautomerase* (*DCT*) and the *tyrosine related protein* (*TYRP1*), which are important proteins for production of EM (22). Both *TYRP1* and *DCT* encodes enzymes important for the production of EM. Nevertheless, the function of *DCT* in human pigmentation diversity seems to be minor, while loss of function in *TYRP1* is known to cause oculocutaneous albinism type 3 (5,27). A SNP found in *TYRP1,* rs1408799, had shown significant association with blue versus non-blue eyes (28).

The transcription of *MITF* is further induced by *cyclic adenosine mono phosphate (cAMP).* Production of *cAMP* from adenosine triphosphate (ATP) happens via *adenylate cyclase* (AC) when *α-melanocyte stimulating hormone* (*αMSH*) is bound to the *melanocortin 1 receptor* (*MC1R*) at the melanocyte membrane. However, if the *agouti signalling protein* (*ASIP*), which function as a receptor antagonist, block the binding of *αMSH* to *MC1R*, the melanogenesis is inhibited. (5)

In cooperation with *MITF*, the *interferon regulatory factor 4* (*IRF4*), which encodes the transcription factor interferon regulatory factor 4, is most likely activating the expression of *TYR* by binding to the *TYR* promotor region (29). Sequence variants of *IRF4* is found to be associated with human pigmentation, such as eye colour. rs12203592, located in the enhancer element of *IRF4* is the SNP located within this gene that has shown the strongest association with human pigmentation (30,31).

The *Oculocutaneous albinism II* (*OCA2*) locus encodes for the human P-gene transcript which is classified as an integral melanosomal membrane protein. Mutations in *OCA2* result in type II albinism. The P protein is believed to stabilize or traffic *TYR* and *TYRP1* to the developing melansome (5,32). Also, the protein may be important for regulation of melanosomal pH, as it is thought to function as an anion transporter (33). The pH in the melanosomes have an effect on *TYR* activity. Fuller et al. showed that acidic internal environment reduces the catalytic activity of tyrosinase, and thereby influence the production of eumelanin (34).

The *OCA2* gene was for a time believed to be the most informative human eye colour gene. However, three parallel studies in 2008 reported the hect domain and RLD2 (*HERC2*) in the regulatory site of *OCA2* as the most important eye colour gene (21,35,36). The SNP, rs12913832, located in the *HERC2* gene has demonstrated the strongest association with eye colour and is assumed to be located in a functional regulator of *OCA2* (36). Nevertheless, the *OCA2* SNP, rs1800407, which is also known to have a strong association with eye colour, acts as a penetrance modifier rs12913832 (3). In 2016, Andersen et al. demonstrated that the SNPs rs74653330 and rs121918166, both located in *OCA2*, also had a measurable effect on normal pigment variation (37). The *HERC2/OCA2* regions affect the eye colour the most, however other genes play a role in eye pigmentation as well.

The solute carrier family proteins solute carrier family 24, member 4 (*SLC24A4*), solute carrier family 24, member 5 (*SLC24A5*), and solute carrier family 45, member 2 (*SLC24A2*) are all transmembrane proteins (5). These proteins are involved in human pigmentation by regulating transport of small molecules across the melanosomal membrane and thus retrieving optimal pH. *SLC45A2* is associated with normal diversity in hair, skin and eye colour. The rs16891982 SNP located within the gene is especially associated with hair and eye colour variation in Europeans (38). Also, the rs12896399, located in *SLC24A4*, is found to be associated with hair and eye colour variations (38).

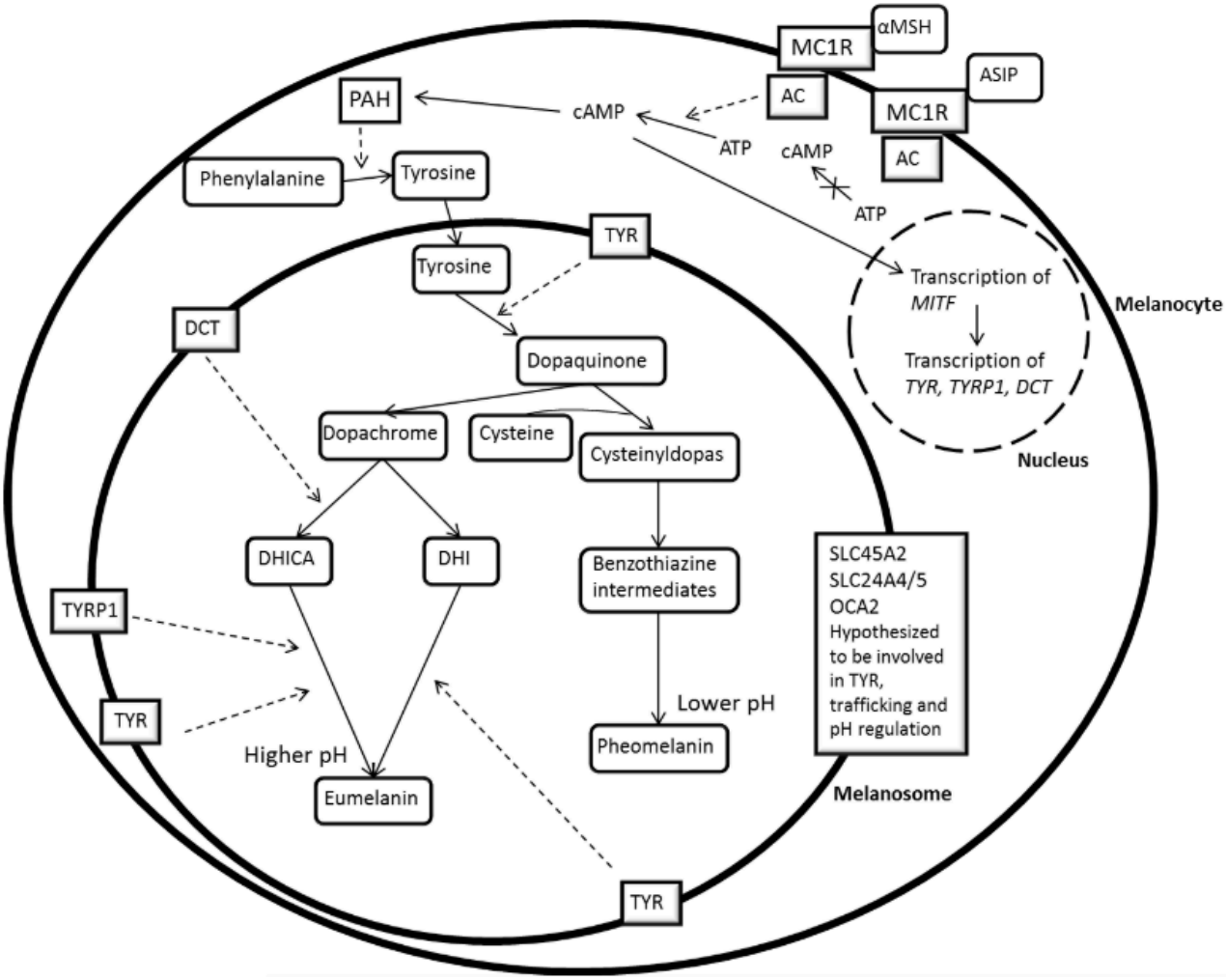


Figure 5: *Production of cAMP is activated when α-melanocyte stimulating hormone (αMSH) is bound to melanocortin 1 receptor (MC1R) and adenylate cyclase (AC). cAMP will then lead to transcription of tyrosinase (TYR). TYR will then convert tyrosine to dopaquinone, which is the last common step for EM and PM synthesis. Depending on presence of cysteine, dopaquinone will go towards production of EM (cysteine depleted) or PM. Figure from (5).*

1.6 Prediction of eye colour

The genetic markers associated with the pigmentary traits are of great interest within the forensic field. They can be used to predict eye, hair and skin colour of a person based on DNA obtained from a crime scene. Prediction of pigmentary traits may reduce the number of suspects and help the police to focus on groups with the predicted phenotypes (5). Today, the eye colour trait is intensively investigated, and several eye colour prediction tools have been published (39–41) e.g. the IrisPlex prediction model (3).

In 2009, Liu et al. genotyped 37 SNPs from eight genes which represented all currently known genetic variants with statistically significant eye colour association (38). Out of the 37 SNPs, 13 were removed due to strong linkage disequilibrium with other markers. The remaining 24 SNPs was used to make a prediction model based on multinomial logistic regression. To evaluate the influence of each SNP to the prediction accuracy of eye colour, receiver under the operating curve (ROC-curve) and the corresponding area under the receiver characteristic operating curve (AUC) value was measured. The AUC-values were measured by excluding one SNP at a time from the model. Finally, they ended up with six SNPs found in six different genes, respectively in *HERC2* (rs12913832), *OCA2* (rs1800407), *TYR* (rs1393550), *IRF4* (rs12209592), *SLC45A2* (rs16891982) and *SLC24A4* (rs12896399).

The prediction accuracy of this model was measured using AUC-values. Perfect prediction refers to an AUC-value equal 1.0 while random prediction refers to an AUC-value equals to 0.5 (42). The model with the six SNPs had a high prediction accuracy for blue (AUC=0.91) and brown (AUC=0.93) eye colour, while the prediction accuracy for the colour between blue and brown (intermediate) was lower (AUC=0.73) (38). However, rs12913832 contributes to the majority of the prediction accuracy with AUC-values 0.88 for blue, 0.90 for brown and 0.66 for intermediate.

Based on the six SNPs reported in 2009, Walsh and co-workers developed a forensic DNA eye colour prediction tool, named the IrisPlex model (3). IrisPlex model is a web-based tool (available from http://hirisplex.erasmusmc.nl/) which predicts eye colour depending on the genotype result for all the six SNPs. The model calculates the probabilities (P-value) of the eye colours blue, brown or intermediate using the same multinomial logistic regression model published by Liu et al (for more details, see (3)).

The year after Walsh et al. published the IrisPlex system model, they also published a validation study for the use in forensic case work (2). Walsh et al. demonstrated that the IrisPlex assay fulfilled all Scientific Working Group on DNA Analysis Methods guidelines in terms of species specificity, sensitivity, reproducibility, and mixture studies, precisions and accuracy, casework samples and population studies (2). Also, over the years the underlying reference data for the model has been enlarged. Today, the database consists of 9466 individuals from eight parts of Europe (Netherlands, Norway, Estonia, UK, France, Italy, Spain, and Greece) and a population

from US (43,44). The IrisPlex prediction model obtained an overall high prediction accuracy for blue (0.94) and brown (0.95) eye colour, while the prediction accuracy for intermediate eye colour were lower (0.74) (43,45) However, prediction of eye colour only reached a correct call rate (sensitivity) of 87.5% for brown and 91.6% for blue eye colour with a probability threshold at ≥ 0.5 (3). Further, the validation of the prediction model concluded that a 100% accuracy prediction for blue or brown eye colour only was achievable with a threshold of ≥ 0.7 (2). When using a threshold of 0.7, the eye colour of many individuals may be inconclusive, and the model can be limited to prediction of eye colour for only a certain number of individuals. Walsh et al. also implied that their prediction model was not well suited for intermediate eye colour (3). Prediction of intermediate eye colours by the IrisPlex model has also showed limitations in other studies (46,47).

As already mentioned, the prediction accuracy is mainly dependent on the *HERC2* SNP, rs12913832. This can be explained by the reduced expression of *OCA2* in lightly pigmented melanocytes with the rs12913832 GG/CC genotype, when compared to darkly pigmented melanocytes with the rs12913832 AA/TT genotype (37). Therefore, the determination of the eye colour is thought to be brown when rs12913832 is AA/TT and to be blue when rs12913832 is GG/CC (Figure 6) (3). Also, because of the dominant hypothesis, where it is thought that brown eye colour is dominating, brown eye colour is the outcome when the rs12913832 genotype is GA/CT (37). However, this is not always the case. Several studies have reported that some individuals with genotype GA/CT for rs12913832 had intermediate or even blue eye colours (37,48,49). In addition, some individuals with the rs12913832 GG/CC genotype had brown eyes.

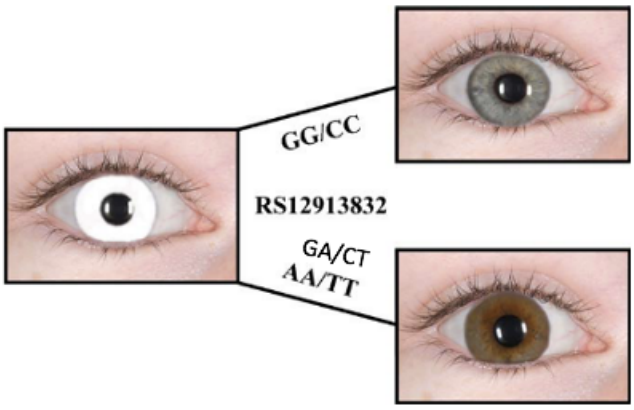


Figure 6: *Hypothetically scenario for genetic determination of brown and blue eye colour caused by rs12913832. Figure modified from (3).*

As the SNP rs12913832, as well as the rest of the IrisPlex markers, do not fully explain the eye colour outcome, and especially not for intermediate eye colour, Section of Forensic Genetics at the University of Copenhagen and Department of Mathematical Sciences, Aalborg University (Denmark) are developing a new prediction model named EC12 ("Eye Colour 12"). The SNPs selected for EC12 are a result of three different studies: the already mentioned study performed by Liu et al. (the six SNPs included in the IrisPlex model), a study performed by Andersen et al. and a study performed by Meyer et al. (38)(37)(Meyer et al., unpublished article).

Andersen et al. sequenced the *OCA2/HERC2* region in 47 individuals with the genotype rs12913832 GA/CT (37). Out of the 47 individuals, 35 did not follow the dominant hypothesis based on the rs12913832 genotype. In this paper, five additional SNPs in *OCA2* that possible explain eye colour variants were identified. These SNPs were included in the development of the EC12 model.

Meyer and co-workers sequenced the *SLC24A4, TYRP1, SLC24A5, IRF4, TYR, SLC45A2* and *OCA2*/*HERC2* genes and surrounding areas in 40 individuals with the genotype rs12913832 GG/CC (Meyer et al., unpublished article). A total of 208 variants in *TYRP1, SLC24A4, IRF4*, and *TYR* with association to brown eye colour in individuals with the genotype rs12913832

GG/CC were identified. Out of these 208 variants, 35 tag-SNPs that captured 100% of the variants were selected (whereas two were already included in the IrisPlex SNP panel).

In total, 44 SNPs with statistically significant association with eye colour were genotyped in 725 individuals (319 Danes, 198 Swedish and 208 Italians). To select the SNPs with the highest additive effect on eye colour variation, different types of models were analysed with continuous and categorical eye colour outcomes. Three different regression models were performed with a continuous outcome (PIE-scores) and the top 12 performing SNPs were selected for the EC12 model (Meyer et al. unpublished article). Four of them were SNPs included in the IrisPlex model (rs12913832 (*HERC2*), rs1800407 (*OCA2*), rs12896399 (*SLC24A4*) and rs16891982 (*SLC45A2*), whereas the remaining eight were SNPs reported by Andersen et al. and Meyer et al. (rs74653330 (*OCA2*), rs121918166 (*OCA2*), rs1408799 (*TYRP1*), rs1126809 (*TYR*), rs7120151 (*TYR*), rs1800401 (*OCA2*), rs10131374 (*SLC24A4*) and rs4904927 (*SLC24A4*)), listed in Table 1.

Table 1: *SNPs included in the prediction model from Denmark and their corresponding gene.*

Gene	SNP
HERC2	rs12913832[1]
OCA2	rs1800407[1]
	rs746553330
	rs121918166
	rs1800401
TYRP1	rs1408799
TYR	rs7120151
	rs1393350[2]
	rs1126809
IRF4	rs12203592[2]
SLC24A4	rs12896399[1]
	rs10131374
	rs4904927
SLC45A2	rs16891982[1]

[1]=Also included in IrisPlex prediction model. [2]=Included in the IrisPlex prediction model, but not in the model from Denmark.

1.7 SNP-genotyping

Genotyping of eye colour SNPs can be performed using different methods such as Single base extension (SBE) (originally termed minisequencing, but often called SNaPshot) and massive parallel sequencing (MPS). Each method has its own weaknesses and strengths (11).

MPS (also termed next generation sequencing (NGS)) is a high-throughput method, which gives complete nucleotide sequence information (50). This method is used on several platforms. The majority of the platforms group the different analysis steps into library preparation, sequencing, imaging and data analysis (51). MPS is a potentially valuable technique in forensic genetics because of its multiplexing capacity and the possibility to analyse different types of markers in the same assay e.g SNPs and STRs, and thereby address all major forensic DNA questions in one analysis (52). The sequencing instrument MiSeq FGx Forensic Genomic System (Illumina) was developed for the forensic community (5). This instrument, together with the ForenSeq^{TM} DNA Signature Prep kit and the Universal Analysis Software package enables complete sample-to-answer analysis (53,54). However, MPS is an expensive method, especially if only a few samples are analysed per run.

SBE is a cheaper alternative to MPS for genotyping of SNPs. This method has been one of the preferred methods for forensic SNP analysis because of its relative simplicity, robustness, sensitivity and multiplexing capacity (52). Another advantage with this method is that it does not require other equipment than already used in forensic genetic laboratories (55). Genotyping of SNPs with SBE starts with a polymerase chain reaction (PCR) amplification to enrich the genomic regions carrying the target SNPs. PCR amplification is followed by an enzymatic clean-up to remove the remaining PCR primers and unincorporated deoxyribonucleotide triphosphates (dNTPs), to prevent interference in subsequent reactions. Next, the cleaned-up PCR products are used as targets in the SBE reaction where the SBE primer is extended with a fluorescent dideoxynucleotide triphosphate (ddNTP). The SBE reaction is followed by a new cleaning step to remove unused ddNTPs. Finally, the cleaned SBE products are separated and detected with capillary electrophoresis (CE) (52). Figure 7 summarise the SBE process. For more detailed explanation of each step, see section 4.

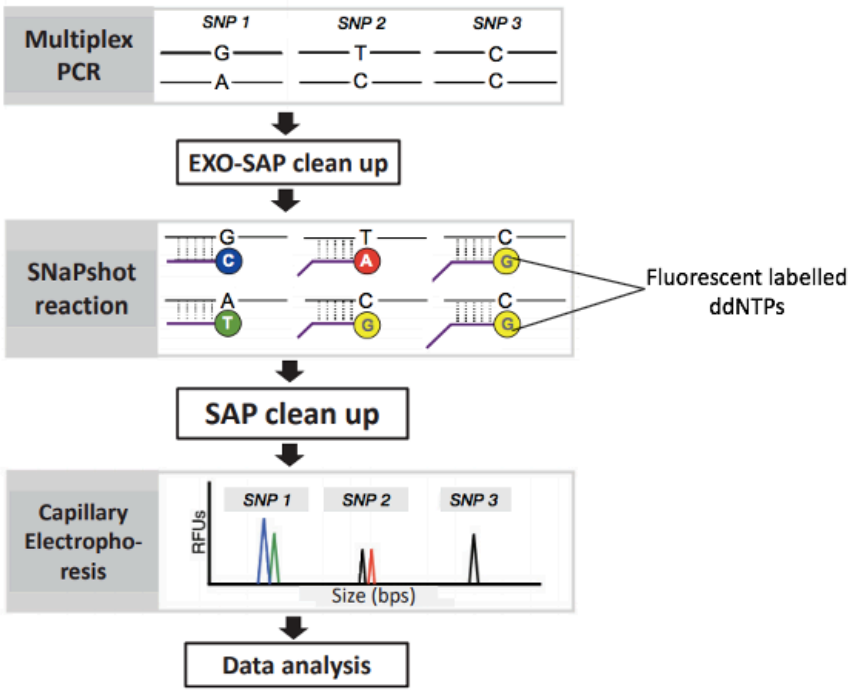


Figure 7: *Schematic overview of the SBE workflow. Figure modified from* (52).

2 Aims of the study

This study consists of several aims, which can be summarised as follows:

- Establishment of a PCR-SBE-CE assay for prediction of eye colour at Section of Forensic Genetics, Copenhagen, Denmark.
- Establishment of the PCR-SBE-CE method at Centre of Forensic Genetics lab in Tromsø.
- Genotype samples from Centre of Forensic Genetics' research biobank with the established PCR-SBE-CE assay.
- Evaluation of the performance of the IrisPlex model in a Norwegian population
- Evaluation of the performance of the novel EC12 prediction model developed by Section of Forensic Genetics at the University of Copenhagen and Department of Mathematical Sciences, Aalborg University, (Denmark).
- Comparison of the EC12 model and an adjusted IrisPlex model (IP NO)

3 Materials

3.1 Reagents

Reagents and kits used in this project are presented in

Table ***2***. More information about PCR and SBE primer-sequences, and the length of the PCR products are presented in appendix 10.1

Table 2: *overview of all reagents and kits used during this project*

Name	Company	Purpose
PCR primers	LGC Biosearch technology	Amplification of target genomic regions
SBE primer	LGC Biosearch technology	Detection of target SNPs
2X Qiagen Multiplex PCR master mix	Qiagen	Amplification of target genomic regions
ExoSAP-IT™	Thermo Fisher Scientific	Clean-up after PCR
SNaPshot™ Multiplex ready reaction mix	Thermo Fisher Scientific	Detection of target SNPs
Shrimp Alkaline Phosphate (SAP)	Thermo Fisher Scientific	Clean-up after SBE
GeneScan™-120 LIZ® Size Standard	Thermo Fisher Scientific	Sizing DNA fragments
HiDi (formamide)	Thermo Fisher Scientific	Denature SBE products before CE
Quantifiler Trio DNA Quantification kit	Thermo Fisher Scientific	Quantitation of extracted DNA samples
POP-4™ Polymer	Thermo Fisher Scientific	Separate DNA fragments in CE

3.2 Instruments

During this project, several instruments were used. All of the instruments are listed in Table 3.

Table 3: *List of instruments and software used during this project.*

Instrument	Company	Laboratory
Mastercycler Nexus gradient	Eppendorf	Copenhagen
Veriti™ 96-Well Thermal Cycler	Thermo Fisher Scientific	Tromsø
3500xL Genetic Analyzer for Human Identification	Thermo Fisher Scientific	Copenhagen and Tromsø
GeneMapper ID-X software version 1.5	Thermo Fisher Scientific	Copenhagen and Tromsø
7500 Real-Time PCR System	Thermo Fisher Scientific	Tromsø

4 Methods

The performance of the laboratory work for this master thesis was divided in two parts. Part one was performed at the Section of Forensic Genetics, Copenhagen, Denmark. This part was originally limited to five weeks but had to be extended to seven. Part two was performed at the Centre of Forensic Genetics, Institute of Medical Biology, Faculty of Health Sciences, UiT The Artic University of Norway.

4.1 Part one

During part one, a PCR-SBE-CE assay for prediction of eye colour was established. The research group at Section of Forensic Genetics had already chosen the 12 genetic markers to be included in the prediction model and designed the primers for PCR and SBE. Four of the 12 markers were overlapping with the IrisPlex markers. In order to have genotype results from all six IrisPlex markers, the remaining two were also included in the assay for this master's thesis. In total, 14 markers were multiplexed. The DNA samples for testing and optimisation of the assay were already extracted and sequenced (to assess the concordance).

4.1.1 Polymerase chain reaction (PCR) for amplification of target genomic regions and ExoSAP-IT clean-up

PCR is a method that allows amplification of billions of copies (amplicons) of a specific DNA fragment from a complex pool of DNA (56–58). Every PCR reaction requires presence of template DNA, forward and reverse primers, thermostable DNA polymerase and deoxyribonucleotide triphosphates (dATP, dCTP, dGTP and dTTP). Each reaction consists of several repetitive cycles, which are divided in three steps; denaturation, annealing and elongation (Figure 8). First, the double stranded DNA is denaturised and separated from each other by heating up the DNA. Next, the primers anneal to the target DNA in a sequence-specific manner. Then the thermostable DNA polymerase (Taq Polymerase) extends the primers by adding dNTPs to the 3' end. A PCR can either amplify one target sequence (singleplex) or several sequences at the same time in the same tube (multiplex). When designing a multiplex-PCR it is important that all primers have approximately same optimal annealing temperature, and that the primers do not align to each other (56–58).

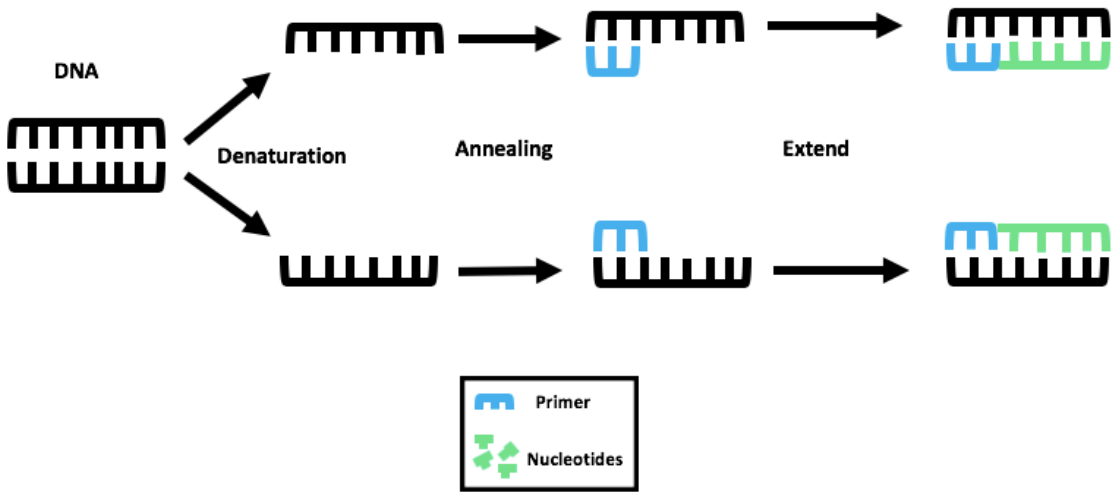


Figure 8: *Schematic overview of the three steps PCR reaction is divided into: Denaturation, annealing and extension.*

During this study, all the primer pairs targeted genomic regions carrying SNPs that potentially have an effect on pigmentation of the eye. First, all primer pairs were tested in singleplex reactions to see if each primer pair successfully amplified the target genomic region. Next, all primers were tested together in multiplex PCR reactions.

In the present study, the 2X Qiagen Multiplex PCR Master Mix kit was used for PCR. This kit consists of dNTPs, HotStarTaq DNA polymerase and a PCR buffer. PCR-products were treated with ExoSAP-IT^{TM} PCR product cleanup reagent. This is a mix of Exonuclease I and SAP (59). Exonuclease I remove excess primers and other extraneous single-stranded DNA product, while SAP remove the remaining dNTPs. This step is important to prevent these substances to interfere with the SBE reactions(59).

Protocol

Before setting up the PCR reactions, all primer pairs were diluted to a 100 µM stock solution. For singleplex experiments, the 100 µM stock solution for each primer pairs were diluted to a 1 µM working solution. For multiplex experiments, a multiplex mix of all 14 primer pairs was made and the final concentration of each primer pair was 1 µM.

PCR set-up:

- All reagents were vortexed and centrifuged before use.
- A mastermix was made by mixing primers, nuclease-free water and 2X Qiagen Multiplex PCR Master Mix according to this:

Component	Volume per well (µL)
2X Qiagen Multiplex PCR Master Mix	5
PCR primer/multiplex mix	2
Nuclease-free water	2
Total	9

- 9 µL was added to each well in an Eppendorf twin.tec PCR 96-well plate.
- 1 µL DNA (approximately 1 ng) was then added to each sample. In the negative control the DNA was replaced by nuclease-free water.
- The PCR 96-well plate was covered with adhesive PCR plate seals and centrifuged briefly before the samples were amplified with the following PCR-program on the Mastercycler Nexus gradient:

Step		Temperature (°C)	Time (min)
Activate enzyme		95	15
35 cycles	Denature	94	0,5
	Annealing	58	0,5
	Extend	72	0,5
Final extension		72	10
Hold		4	Hold

Clean-up step:

- After the PCR, the plate was centrifuged briefly.
- 5 µL of each PCR-product was transferred to respective wells in a new PCR 96-well plate before 2 µL ExoSAP-IT^{TM} was added.
- The new PCR 96-well plate was covered with adhesive PCR plate seals and centrifuged.
- The plate was incubated at 37 °C for 60 min, 75 °C for 15 min, and hold at 4 °C.

4.1.2 Single base extension (SBE) to label target SNPs and SAP clean-up

An SBE reaction is performed on either the forward or the reverse strand of the PCR product, and the SBE primers are therefore designed to hybridise to one of the strands. The SBE primers hybridize directly adjacent to the nucleotide base to be identified (Figure 9)(52). The DNA

polymerase extends the primer with a fluorescent labelled ddNTP, where each of the four ddNTPs (ddTTP, ddATP, ddGTP, ddCTP) are labelled with its own fluorescence dye. ddNTPs are used instead of dNTPs as they lack the 3' OH group and ensures that no more nucleotides extend the SBE primers than the one ddNTP complementary to the base in the SNP position (34).

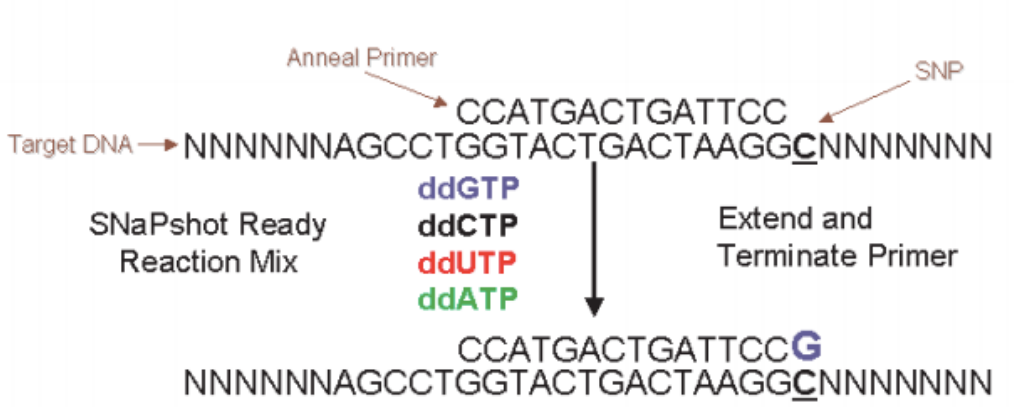


Figure 9: *Schematic overview of the SBE reaction. SBE primer hybridized immediately upstream of the target SNP, extension of the SBE primer by fluorescent-labelled ddNTPs. Figure modified from (34).*

In this study, the SBE primers were first tested in a singleplex reaction on cleaned singleplex PCR products. Next, the SBE primers were tested in a multiplex reaction on cleaned multiplex PCR products, for the same reasons as for the PCR primers.

During this project, the SNaPshot™ Multiplex ready reaction mix was used for the SBE-reaction. This reagent/kit consist of fluorescently labelled ddNPTs and enzymes (34). After the SBE reaction, the SBE product was treated with SAP, which removes unused dye-linked ddNTPs.

Protocol

Prior to the SBE reaction, all 14 primers were diluted to a 100 µM stock solution. For singleplex experiments the primers was further diluted to a 1 µM working solution. For multiplex experiments, the 100 µM stock solution for each primer was diluted to a 10 µM working solution. Then a multiplex solution of all 14 primers was made according to Table 4.

Table 4: *Overview of amount of each SBE primers added to make the multiplex mix for multiplex experiment.*

SNP ID	Adjustment	Amount of 10 µM stock to add (µL)	Concentration of each primer in mix (µM)	Final concentration (µM)
rs10131374	4x	8	1.14	0.23
rs1126809	5x	10	1.43	0.29
rs12896399	3x	6	0.86	0.17
rs1408799	1x	2	0.29	0.06
rs16891982	2x	4	0.57	0.11
rs7120151	1x	2	0.29	0.06
rs74653330	3x	6	0.86	0.17
rs1393350	2x	4	0.57	0.1
rs1800407	0,5x	1	0.14	0.03
rs4904927	1x	2	0.29	0.06
rs121918166	4x	8	1.14	0.23
rs12913832	1x	2	0.29	0.06
rs1800401	1x	2	0.29	0.06
rs12203592	2,5x	5	0.7	0.14
Total		62		
Water to add		8		

SBE set-up:

- All reagents and primers were vortexed and centrifuged briefly before usage.
- 1 µL of each ExoSAP-IT™ treated PCR-product was pipetted into a new Eppendorf twin.tec PCR 96-well plate.
- SNaPshot™ Multiplex ready reaction mix, nuclease-free water and SBE primers were added to each sample according to this:

Component	Volume per well (µL)
PCR-product (ExoSAP-IT™ treated)	1
SNaPshot™ Multiplex ready reaction mix	2
Nuclease-free water	1
SBE primer/multiplex mix	1

- The 96-well plate was covered with adhesive PCR plate seal and centrifuged briefly.
- The thermal cycling conditions for the SBE reaction were as follow:

Step		Temperature (°C)	Time (s)
30 cycles	Denature	96	10
	Annealing	55	5
	Extend	60	30
Hold		4	Hold

Clean-up step:

- 1 µL SAP was added directly to each sample.
- The 96 well plate was centrifuged before incubated at 37 °C for 60 min, 75 °C for 15 min, and hold at 4 °C.

4.1.3 Capillary electrophoresis to detect SBE product

After the SBE reaction, the cleaned SBE products were separated and detected by capillary electrophoresis (CE). CE is a method composed of capillaries filled with viscous polymer, which acts like a gel (11). Within the capillary, short DNA fragments will migrate faster through the polymer than the longer fragments. The extended SBE primers with the fluorescent labelled ddNTPs are detected when passing through the detection window. In the detection window there is a laser which causes the dyes on the fragment to fluorescence at different emission levels. The detector measures the different fluorescence intensities and record the time when the molecule passed the window. To find the lengths of the DNA molecules, an internal size standard is mixed with the samples. The internal size standard has molecules with known lengths measured in nucleotides and its own, distinctive fluorophore. By adding this to the samples, the known length of the molecules in the size standard can be used to make a standard curve as a function of time. Finally, the unknown lengths of the extended SBE primers can be decided based on the standard curve (11).

The internal size standard used in the present study (the GeneScan™-120 LIZ® Size Standard) can identify the size of DNA molecules in the 15-120 nucleotide range and provides nine single-stranded labelled fragments (60).

Protocol

- Hi-Di Formamide and GeneScan™-120 LIZ® Size Standard was mixed in a 200:1 ratio. The total volume of each component in the mastermix depended on the number of

injections in each set-up (for one injection: 477.6 µL Hi-Di Formamide and 2.4 µL GeneScan™-120 LIZ® Size Standard).

- The mix was vortexed and centrifuged briefly.
- 20 µL of the formamide/size standard mix was added in each well in a new 96-well plate.
- 1 µL of SAP treated SBE-product was added to each sample.
- The 96 well plate was covered with septa and the plate was centrifuged.
- The plate was heat-shocked for 3 minutes (96 °C) followed by 3 minutes on ice.
- The plate assembly was prepared before placing the 96-well plate to the autosampler.
- The plate record was completed and linked to the 96-well plate.
- Following settings were used: Analysis method: fragment analysis, capillary length: 36 cm, polymer: POP4 and dye set: E5.
- Capillary electrophoresis run was as follows:

Parameter	Control Module GS POP-4 E5
Injection time	24 seconds
Run voltage	15 kV
Run time	1210 seconds
Injection voltage	1.2 kV

4.1.4 Data analysis using GeneMapper

After CE, the genotype results were analysed in GeneMapper ID-X software version 1.5 (GMID-X). The genotypes are presented in electropherograms and the four possible SNP alleles are reported using different fluorochromes, one for each ddNTPs (G=blue, A=green, C=yellow and T=red). These dyes have different emission levels, which make their signal strength vary, even when the quantity of extended SBE primers is identical. This may cause artefacts, which makes it important to analyse the results and remove unreal peaks and noise (52). Artefacts can also be related to PCR and CE. There are several types of artefacts which can appear in an electropherogram, and the most common ones observed in this project are presented in Table 5.

All samples are analysed using a predefined analytic threshold (50 relative fluorescent units (RFU)). Only peaks above the analytical threshold will be analysed and considered.

Table 5: *Most common artefacts observed in electropherograms during this project.*

Name of artefact	Appearance	Reason for occurrence
Pull up or "bleed through"	Often seen as a low, allele-like peak with a regularly shape in another colour channel in the same position.	Spectral overlap between the dyes makes the colour "bleed" into another colour channel (11).
Spikes	Sharper than the allele peaks and often observed in the same position in several of the colour channels.	This artefact is caused by small air bubbles, urea crystals or voltage spikes.
Split peak	Seen as a peak with double peaks in it.	Signal is too high because too much template was used in the SBE reaction and too much SBE product was created.

Protocol

Before any data could be analysed in GMID-X, the panel and allelic bins had to be defined for the 14 markers and their alleles used in this assay. Within the panel, there is a collection of bins (an allele definition; a size (bp) and dye colour) called bin sets (61). By using the observed lengths and dye colours from the singleplex experiments, individual bins for each allele were defined.

Data analysis:

- A new project was made, and sample files from CE were added.
- Analysis parameters for the project were set: Analysis method, panel and size standard. And the samples were analysed.
- The electropherogram for each sample was inspected and all artefacts and peaks considered as noise were excluded/removed from the electropherograms.
- Unlabelled peaks considered as true alleles were labelled manually.
- The conferred SNP-genotype data were exported from GMID-X, and a table of all samples was made in Excel.

4.1.5 Optimisation

During this first part of the project, several optimising steps were performed:

- Several concentrations of the SBE primers in the SBE multiplex mix were used before ending up with the concentrations listed in Table 4.
- Three SBE primers (rs408799, rs7120151 and rs74653330) were redesigned (done by Section of Forensic Genetics) because the SBE product overlapped with other SBE products, that made it impossible to interpret which marker the observed peak belonged to.
- Some SBE products did not appear in the electropherogram at all or had very low peak heights. To test if the amplification was suboptimal in the multiplex PCR, multiplex PCR products was tested on singleplex SBE primers.
- For one particular marker, rs7120151, several other optimising steps were tested:
    - Several concentrations of the PCR primers in the PCR multiplex mix were tested.
    - Several annealing temperatures were tested (53 °C, 54 °C and 56 °C), unsuccessfully.
    - PCR primers were redesigned (done by section of forensic genetics).

Even though my stay in Copenhagen was extended two weeks, the performance of the assay was not optimal and especially not for rs7120151. Consequently, an alternative workflow was decided in order to genotype this SNP. The assay was split into a 13-plex PCR multiplex mix (without rs7120151) and a singleplex PCR for rs7120151. After ExoSap-IT treatment, the multiplex and singleplex PCR products were mixed in a 1:1 ratio before the SBE reaction.

4.2 Part two – performed at Centre of Forensic Genetics, Tromsø, Norway

In this part of the project, the PCR-SBE-CE assay was first established at the Centre of Forensic Genetics, Department of Medical Biology, Faculty of Health Sciences, UiT-The Artic University of Norway. Next, 543 DNA samples from the centres' research biobank were genotyped. These samples were collected with informed consent from individuals residing in Tromsø and Bodø, Norway from 2015-2017. High resolution photographs of the participants eyes were taken upon sampling. The participants self-reported their birthplace and the ancestry of their four grandparents. Out of the 543 individuals, 449 of them had four Scandinavian grandparents. The DNA extraction and quantitation had already been carried out prior this

project. Sequencing of the samples was carried out in previous project using the ForenSeq DNA Signature Prep kit (Verogen) on the MiSeq FGx instrument (Illumina) (6).

4.2.1 Quantitative PCR - Real time PCR (qPCR)

Before genotyping the samples with the PCR-SBE-CE assay, 14 randomly chosen samples were re-quantitated to check if the DNA quantitation were approximately the same after stored in either a fridge or a freezer over a longer time.

The samples were quantified using real time quantitative polymerase chain reaction (qPCR), also called real time PCR (RT-PCR). The key feature in qPCR is that amplification of DNA is detected in real time by the use of a fluorescent reporter (TaqMan probe). The fluorescent reporter signal is directly proportional to the number of amplified DNA molecules (62). The method works by sequence-specific probes that binds to the region between the forward and reverse primer. The probe consists of a minor groove binder, a reporter and a non-fluorescence quencher (11). The reporter normally emits fluorescent light, but the non-fluorescent quencher will absorb the light as long as the probe is intact. The Taq DNA polymerase, with endonuclease activity, will cleave the probe during extension of the primers and the reporter will then emit light. The light is then detected and converted to readable signals. To quantify the DNA in the samples, the emitted light from the DNA sample is compared to a standard curve. The standard curve is made from a dilution series with known DNA concentrations (62).

In this study the Quantifiler Trio DNA Quantification kit from Thermo Fisher Scientific was used. This kit only amplifies human DNA. It contains four different TaqMan probes: One for targeting small autosomal DNA, one for large autosomal DNA, one for targeting Y chromosome DNA (target male DNA) and one for an internal PCR control that detects PCR-inhibitors. By having probes that target both small and large fragments, it is possible to measure the degradation of the sample by analysing the ratio between the small and large DNA fragments.

Protocol:

- First, a mastermix with THP PCR reaction mix and primer mix was made. The volume of each component needed was calculated dependent on the number of samples (including

quantification standards, negative control and calibrator) to be quantified: 10 µl THP PCR reaction mix and 8 µl Trio primer mix per rx.

- The mastermix was vortexed and centrifuged briefly.
- 18 µl of mastermix was dispensed to each well on a 96-well plate.
- 2 µL of DNA sample, DNA quantification standard, calibrator or control was added to the applicable wells.

  Notably: The DNA quantification standards had 5 concentration points (0.005 ng/µL, 0.050 ng/µL, 0.500 ng/µL, 5.000 ng/ µL, and 50.000 ng/µL) and were added in duplicates.
- The reaction plate was sealed with Optical Adhesive Cover before the plate was centrifuged.
- The plate was put on the 7500 Fast DX Real-Time PCR Instrument.

The DNA concertation for all the fourteen DNA samples were approximately the same as previously determined. Consequently, the rest of the DNA samples was not re-quantified.

4.2.2 Establishing of SBE at Centre of Forensic Genetics

The first step when establishing the method at Centre of Forensic Genetics was to test different input DNA concentrations in the PCR reaction to see if the concentrations affected the amplification and thereby the allele peak heights. As the DNA samples in the Centre of Forensic Genetics research biobank had DNA concentrations ranging from 1-13 ng, 2 ng and 20 ng were tested.

4.2.3 PCR-SBE-CE assay at Centre of Forensic Genetics

The PCR amplification of the markers was performed as one multiplex PCR (13 SNPs) and one singleplex PCR (rs7120151). The PCR products from multiplex and singleplex PCR were mixed before a single SBE reaction (as described in section 4.1.1 in part one).

Protocol for PCR-SBE-CE assay in Tromsø:

PCR and SAP/EXO treatment:

- Diluted primer pairs for singleplex PCR (rs7120151) and 13-plex PCR multiplex mix were vortexed and centrifuged briefly.
- Next, two different mastermixes were made, one with 2X Qiagen Multiplex PCR Master Mix, nuclease-free water and diluted rs7120151 primer pairs, and the other with X Qiagen

Multiplex PCR Master Mix, nuclease-free water and 13-plex PCR multiplex. The ratio between each component were as follow:

Component	**Volume per well (uL)**
2X Qiagen Multiplex PCR Master Mix	5
Primer pairs rs7120151/13-plex PCR multiplex mix	2
Nuclease-free water	2
DNA approx. 1 ng	1
Total	10

- The thermal cycler conditions and ExoSAP-IT™ treatment was as performed as described in section 4.1.1.

SBE and SAP treatment:

- ExoSAP-IT™ treated PCR-products from multiplex and singleplex were mixed in a 1:1 ratio.
- 1 μL of the mixed ExoSAP-IT™ treated PCR-product was transferred to the respective wells on the 96-well plate.
- Mixing of the SBE mastermix, SBE thermal cycling conditions and SAP treatment were as described in section 4.1.2.

CE:

- The procedure for capillary electrophoresis was performed as described in section 4.1.3.

Data analysis:

- The electropherograms were analysed as described in section 4.1.4.
- For quality control, two people independently analysed all samples, and the results from both analyses were compared. Samples where the analyses were not in concordance were genotyped again.
- Also, the PCR-SBE-CE results for the six IrisPlex genetic markers were compared to previous results obtained by sequencing.

4.2.4 Categorisation of eye colour

The prediction models predict three eye colour categories, blue, intermediate and brown. Therefore, the eye photos in our dataset were categorised with an eye colour. Thereby, the predicted eye colour results could be compared to the categorised eye colour from the photo.

4.2.4.1 Subjective categorisation of eye colour - human visual identification

523 photographs of eyes were uploaded in a dataset on the platform www.dataturks.com. As an internal control, 20 of the photos was included twice. Nine untrained individuals were invited to annotate photos in a personal dataset. Dataturks shuffles the photos, so each dataset will display the photos in different orders. The participants were asked to categorise the eye colours by evaluating each photo and labelling them as either blue, intermediate-blue, intermediate-brown or brown (appendix 10.2). When analysing the data, the different classes/categories can be merged into a two- or three-category system. In the three-category all photos annotated intermediate-blue and intermediate-brown are merged and classified as intermediate. This results in the three categories blue, intermediate and brown. In the two-category-system, all photos annotated intermediate-blue and blue are merged and classified as blue, while photos classified as intermediate-brown and brown are merged and classified as brown.

Each eye photo was assigned to a final eye colour category based on the annotation made by nine untrained individuals. The category that received the major number of annotations, was assessed as the final eye colour category. For instance, if an eye photo received five annotations for blue and four annotations for intermediate-blue, the final category was assessed as blue.

4.2.4.2 Objective categorisation of eye colour - Digital Iris Analysis Tool (DIAT)

DIAT quantitatively determine the colour of the irises by counting numbers of blue and brown pixels within the iris (63). The software calculates the blue/brown ratio called Pixel Index of the Eye (PIE-score). The PIE-score ranges from -1 to 1, where a PIE-score of 1 is obtained when only blue pixels are represented in the iris, while a PIE-score of -1 is obtained when only brown pixels are represented in the iris. For more details about the software see (63). The calculations of PIE-scores from the eye photos was carried out by Nina Mjølsnes Savo, Centre of Forensic Genetics, UiT.

All the eye photos were classified based on the calculated PIE-score. Photos with PIE-score ≥ 0.8 were classified as blue, photos with PIE-score ≥-0.5 and PIE-score ≤ 0.8 was classified as intermediate and photos with PIE-score ≤ -0.5 was classified as brown, respectively.

4.2.5 Phenotypic estimation

The phenotypic estimation of eye colour was carried out by three different models, the original IrisPlex model (3), an adjusted IrisPlex model generated on a Norwegian population (hereafter referred to as IP NO) and the EC12 model. Each model obtained prediction results with different algorithms and the models were generated using different regression analyses. The two IrisPlex models were generated using different reference populations. The original IrisPlex model was considered for all populations (43,44), while IP NO was based on a Norwegian population using the leave-one-out method. The EC12 model was based on the same reference population as IP NO. In addition, the phenotype evaluations were different. The original IrisPlex model was based on qualitative phenotype evaluations (visual inspection), whereas IP NO and EC12 were based on categorical eye colour determination by PIE-score thresholds.

The sensitivity and specificity (equation 1 and 2), and the AUC-values (section 4.2.6) were calculated for all the three models. The sensitivity is the fraction of individuals with a specific eye colour, that are correctly predicted as such. On the other hand, specificity indicates the fraction of individuals without a specific eye colour, that are correctly predicted as such.

$$1.\, Sensitivity = \frac{True\ positive}{True\ positive + false\ negative}$$

$$2.\, Specificity = \frac{True\ negative}{True\ negative + false\ positive}$$

4.2.5.1 IrisPlex prediction model

Phenotypic estimation by the IrisPlex model was already carried out in a previous project (6). Briefly, a single input file was created for the concurrent prediction of multiples profiles. Each profile consisted of the six SNPs genotypes. The IrisPlex webtool (available from https://hirisplex.erasmusmc.nl/) returns a p-value for blue, intermediate, and brown eye colour (which sums up to one). The overall prediction accuracy, sensitivity, and specificity were calculated using 1) no probability threshold, 2) threshold of p ≥ 0.5, and 3) threshold of p ≥

0.7. The eye colour with the highest p-value were used as the predicted phenotype. With a threshold of $p \geq 0.5$, all samples with p-value under 0.5 was defined as undetermined, while all samples with $p \geq 0.5$ were used in further analysis. The same done with the threshold of $p \geq 0.7$.

4.2.5.2 EC12 and IP NO prediction model

Phenotypic estimation by the EC12 and IP NO prediction model was carried out by Mikkel Meyer Andersen, Department of Mathematical Sciences, Aalborg University, Denmark. Both models were ordinal logit models prediction three categories; blue, intermediate and brown. These models also return a p-value for blue, intermediate and brown eye colour. To obtain these p-values leave one out cross validation (LOOCV) was used. When using LOOCV, one sample is taken out and a model for the remaining sample is made. Next, the prediction model is tested on the sample that was taken out. This was done for all samples (n=521), so the models were trained and tested before a p-value for each sample was obtained. The IP NO model was only used to compare the prediction results from the novel EC12 model including new SNPs with the prediction results from the six IrisPlex SNPs. The prediction accuracy, sensitivity and specificity were calculated as described in section 4.2.5.1.

4.2.6 Statistical Analysis

The statistical calculations for evaluation of the results, boxplots and diagrams were done in Excel 365 version 16.35 for Mac. The boxplots contain upper and lower whiskers, which represents the highest and lowest values. The boxes reflect the values within the first and third quartiles. The mean value is marked with a ×, the median with a line in the boxes and outliers are marked with dots.

All the statistical calculations were carried out by using Real Statistics Resource Pack for Mac 2016/2019/365 (available from www.real.statistics.com). A Shapiro-Wilk test was used to test if the different observed peak height for a genetic marker was normality distributed. Next, to test if the differences between the peak height were statistically significantly different, a T-test for two independent samples was used. To calculate the correlations, the Spearman's correlation coefficient test (alpha=0,05) was used, as the data were not symmetric. This correlation test is a non-parametric method that assesses if a relationship between two variables is linear or not. A perfect correlation (+1 or -1) occurs when the variables is a perfect monotone function of each other. A positive correlation coefficient occurs when observations have a

similar or identical rank between the variables. A negative correlation coefficient occurs when the observations have a dissimilar or fully opposite between the variables. Finally, to test if the differences in PIE-score within each eye colour category and the PIE-score within the three possible genotypes for rs12913832 were statistic significantly different, Mann Whitney for two independent samples was used. This test was used since the data were not normality distributed.

All receiver under the operating curves (ROC-curves) and AUC-values presented in this study was calculated by Nina Mjølsnes Savo at Centre of Forensic Genetics, UiT with assistance from Endre Anderssen, UiT, by using R (version 3.6.1). The ROC-curves illustrates how accurately the prediction model can discriminate between two eye colour states, e.g. blue or non-blue. The different eye colour classes were plotted in the ROC-curve with false positive rate on the x-axis against a false positive rate on the y-axis. A perfect prediction model equals an AUC of 1.0, while an AUC-value of 0.5 indicates a random prediction model. (64)

5 Results

5.1 Establishment of the PCR-SBE-CE assay

5.1.1 Part one - Establishment of assay at Section of Forensic Genetics, Copenhagen

The first step to establish the PCR-SBE-CE assay for prediction of eye colour was to test all primers in singleplex and multiplex reactions using samples with known genotypes. All the 14 SNPs were successfully amplified in singleplex reactions (data not shown). In the first multiplex test, the concentration of PCR primers (1 µM) and SBE primers (0,7 µM) were the same for all the SNPs. This resulted in peak heights that varied from approximately 100 RFU to 28000 RFU (Figure 10). The amplification of some markers was very efficient and caused artefacts such as pull-ups and split peaks in the electropherogram. Also, the amplification of other markers was so insufficient that they did not appear in the electropherogram at all (they were not above analytical threshold of 50 RFU).

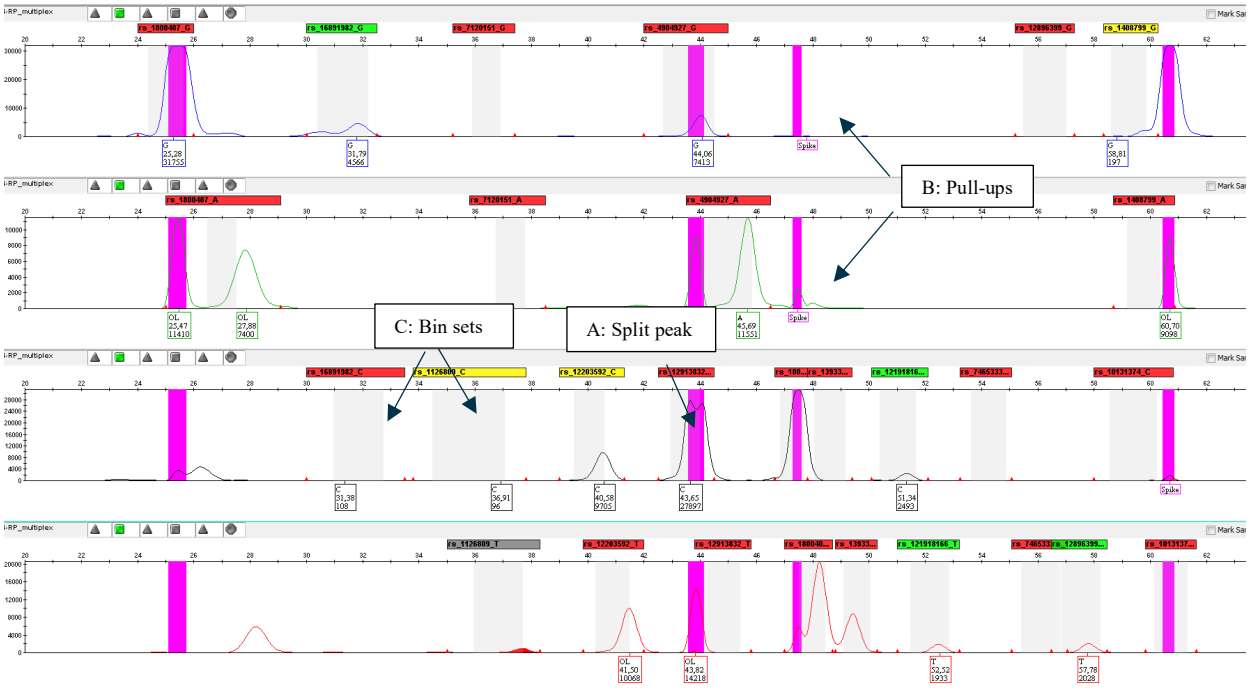


Figure 10: *Electropherogram of the first test of all primers in multiplex mix. Pink areas are areas with intense signal strength. A: Intense signal strength causing a split peak. B: Intense signal strength causing pull-ups. C: Grey areas are bin-sets where allele peaks are supposed to reside.*

Depending on the results from the first multiplex test, the concentration of the SBE primers was adjusted to obtain a more balanced amplification of all the markers. The SBE primer

concentration for the markers that were very efficient amplified was decreased, while the primer concentration for insufficient amplified markers was increased (see Table 4 in section 4.1.2). Furthermore, the separation of rs408799, rs7120151 and rs74653330 were not optimal, as their SBE products had approximately the same as other markers extended SBE products. Therefore, the SBE primers for these markers were redesigned. The optimisation of SBE primer concentrations and primer length resulted in improved peak heights and better separation in the electropherogram (Figure 11).

Although the mentioned optimisation steps improved the results for the multiplex reaction, the amplification of the rs7120151 SNP was not improved (red circle in Figure 11). For this marker, several more adjustments for both PCR and SBE primer concentrations were tested. PCR and SBE primers were redesigned and different annealing temperatures were tried. None of these tests made the amplification for this marker successful. Generally, the peak height for this marker was low, but for heterozygote samples, the A allele was even lower and sometimes it did not appear in the electropherogram at all. This resulted in incorrect homozygous GG genotype calls for the rs7120151 SNP. None of the individuals in the dataset had known genotype AA for rs7120151.

Figure 11: *Electropherogram of the same DNA sample in Figure 10 after adjustment of SBE primer concentrations and after data analysis. All allele markers reside in their bin-sets and 13 out of the 14*

markers shows acceptable peak heights. Red circle indicates the insufficient amplification of the rs7120151 marker, resulting in low peak heights.

Because of the imbalanced peak heights for rs7120151 and the overall insufficient amplification of the A allele, an alternative workflow was tested. First, rs7120151 was amplified in a singleplex PCR reaction, while the remaining markers were amplified in a multiplex PCR reaction. Next, the PCR products from both reactions were cleaned with SAP/EXO. The cleaned PCR products (singleplex and multiplex) were mixed in a 1:1 ratio before the SBE reaction. Finally, the SBE products were treated with SAP before CE and further data analysis (described schematically in Figure 12). The amplification of the rs7120151 was improved with this workflow, where the amplification of the A allele was more efficient. However, the workflow was still not optimal as the peak heights of rs7120151 was still imbalanced (data not shown).

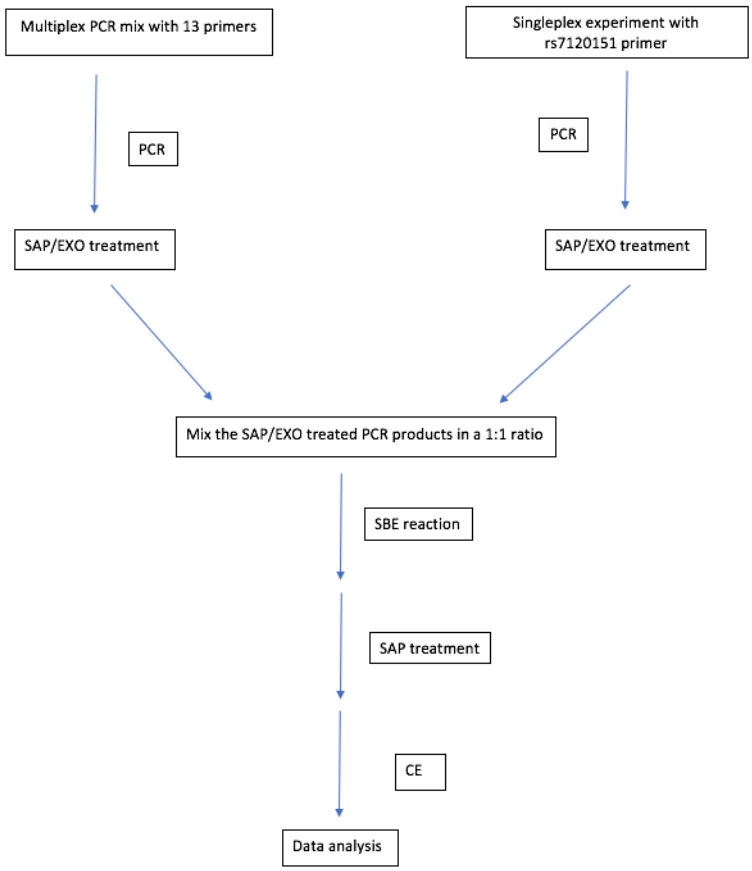


Figure 12: *Flowchart of the alternative workflow for the PCR-SBE-CE assay. First, amplification of the markers was carried out in a singleplex PCR reaction (with rs7120151 primer) and a multiplex PCR reaction (with the remaining 13 primers), followed by a clean-up step with SAP/EXO. Next, the*

SAP/EXO treated PCR product from the multiplex and the singleplex reactions were mixed in a 1:1 ratio. Finally, the mixed SAP/EXO treated PCR products were targets in a common SBE reaction, prior to SAP treatment and CE.

5.1.2 Part two- Establishment of the assay at Centre of Forensic Genetics, Tromsø

Before establishing the PCR-SBE-CE assay at Centre of Forensic Genetics, two different input DNA concentrations in the PCR reaction were tested. This was done to see if the DNA concentration affected the overall quality of the amplification of the different markers. The DNA concentrations of the samples in the Centre of Forensic Genetics research biobank ranged from approximately 1-13 ng/µL, and therefore the performance of the assay was tested on 2 ng and 20 ng. One sample was tested in triplicates for each concentration. When comparing the average peak heights within each marker, no significant difference (T-test: $p > 0.05$) were observed (Figure 13). For heterozygous genotypes, the average peak height of the two alleles was used. Figure 13 implies that the amplification of the markers was similar, regardless of the sample's starting DNA concentration. Consequently, there were no need to dilute the DNA samples to a specific DNA concentration before genotyping the rest of the samples.

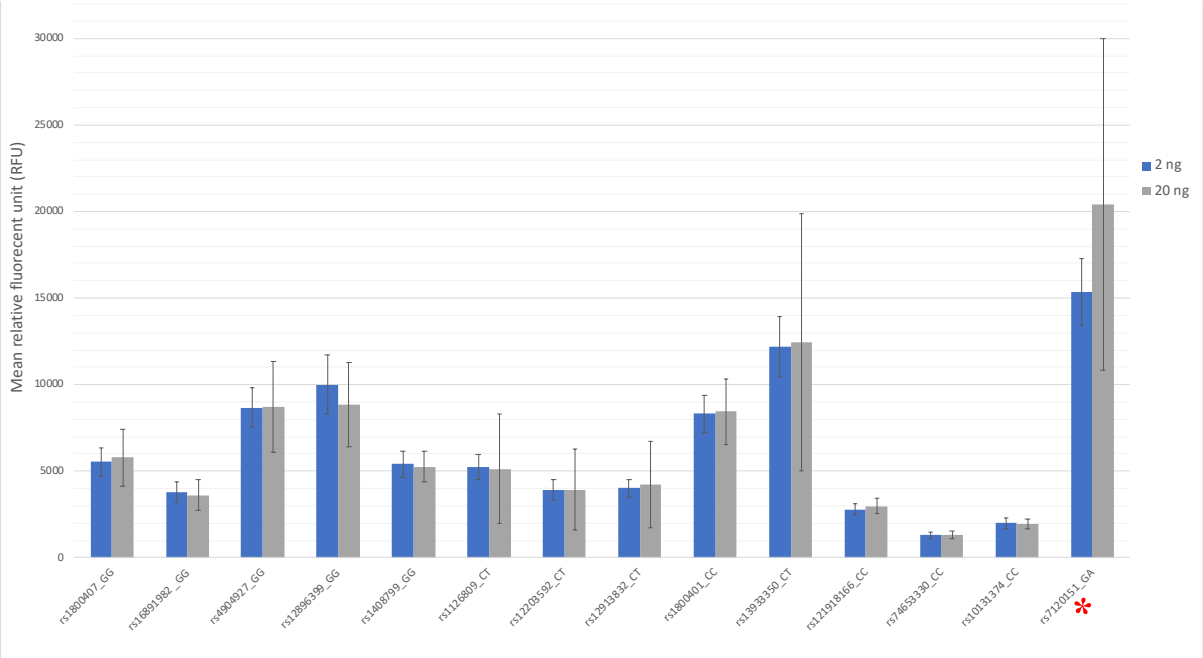


Figure 13: *Comparison of the average peak heights from experiments with two different input DNA concentrations (2 ng DNA (blue bars) and 20 ng DNA (grey bars)). One sample was tested for each concertation. The error bars represent the standard deviations for the three replicates. The average peak height for the two alleles were used for heterozygous genotypes. Homozygous marker=marker name followed by GG or CC. Heterozygous markers= marker name followed by CT or GA.*= marker amplified in singleplex PCR and SBE reaction.*

5.2 Genotyping and data analysis of DNA samples

In total, 543 DNA samples were genotyped with the PCR-SBE-CE assay for prediction of eye colour. Two of the DNA samples were excluded from the study because of poor DNA quality. The electropherograms of the remaining 541 samples were independently analysed by two people in GMID-X, and the results were compared. Seven samples had inconsistencies in the genotype comparison, and consequently re-analysed. After these seven samples were re-analysed, all analysis results were in concordance.

The six SNPs included in the IrisPlex prediction model, were previously been sequenced with the ForenSeq DNA Signature Prep kit (Verogen) on the MiSeq FGx instrument (Illumina). The genotyping results were compared as a quality control. The results were in concordance (data not shown), which confirmed the good quality of the results obtained with PCR-SBE-CE.

5.3 Categorisation of eye colour

The eye photo from each participant was used to categorise the eye colour and calculate the PIE-score. Only individuals with approved genotyping results and PIE-score were used in further analysis, thus 18 additional samples were excluded. The eye photo from the reaming 523 individuals was further used to annotate their eye colours.

Nine untrained individuals annotated the 523 eye photos into one of four categories; blue, intermediate-blue, intermediate-brown and brown. 20 of these photos were included twice as internal controls. Two of the untrained individuals annotated one of the internal controls in two different categories. After these controls were annotated a third time, the final annotation was determined using the category that received the most annotations.

The final eye colour for subjective evaluation of each eye photo was determined using the category that received the majority of the annotations. Two of the samples were excluded as none of the categories received a majority of annotations (four individuals annotated them blue, four individuals annotated them intermediate-blue and one annotated them intermediate-brown). In total, a complete dataset with genotype results, eye photo, PIE-score and annotation results was obtained from 521 individuals. 267 (51%) were observed to have blue eyes, 132 (25%) were observed to have intermediate-blue eyes, 51 (10%) were observed to have

intermediate-brown eyes and 71 (14%) were observed to have brown eyes (Figure 14). Since the IrisPlex model predicts eye colour in three different categories, the four categories were further grouped into three categories. This was done by merging intermediate-blue and intermediate-brown to a final category intermediate, while blue and brown category remained the same (Figure 14). The categories can also be grouped into two categories, which is described in section 4.2.4.1.

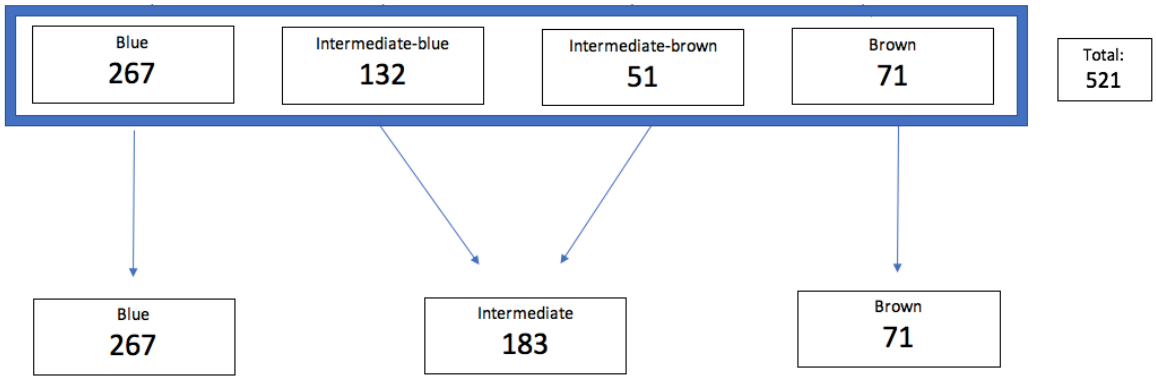


Figure 14: *Illustration of the number of observed eye colour within the four categories (blue square). The eye colours can be grouped into three categories by merging the intermediate-blue and intermediate-brown eye colours into intermediate eye colour.*

For objective evaluation of eye colour, each eye photo was analysed with the DIAT software and a PIE-score was measured. PIE-score of -1 indicates that only brown pixels were present in the iris, while a PIE-score of 1 indicated that only blue pixels were present in the iris. A high correlation (spearman correlation = -0.88, $p < 0.001$) between the eye colour observations and the PIE-score was observed (Figure 15). The highest variation in PIE-score was observed within the intermediate category (median = 0.41), with eye colours ranging from a PIE-score of -1 (only brown pixels) to a PIE-score of 1 (only blue pixels), as illustrated in Figure 15. There was no major variation of PIE-score within the blue and brown category. A PIE-score median of 0.99 for blue category indicates that nearby all pixels in the 267 eye photos observed as blue, are blue, while a median of -0.97 for brown category indicates that almost all pixels in the 71 eye photos observed as brown, are brown. The distribution of PIE-score is clearly defined within the three eye colour categories, and a statistical significance in PIE-score were observed (Mann Whitney: $p < 0.001$). The distribution, together with the correlation coefficient, demonstrates that PIE-score can be used to categories eye colour.

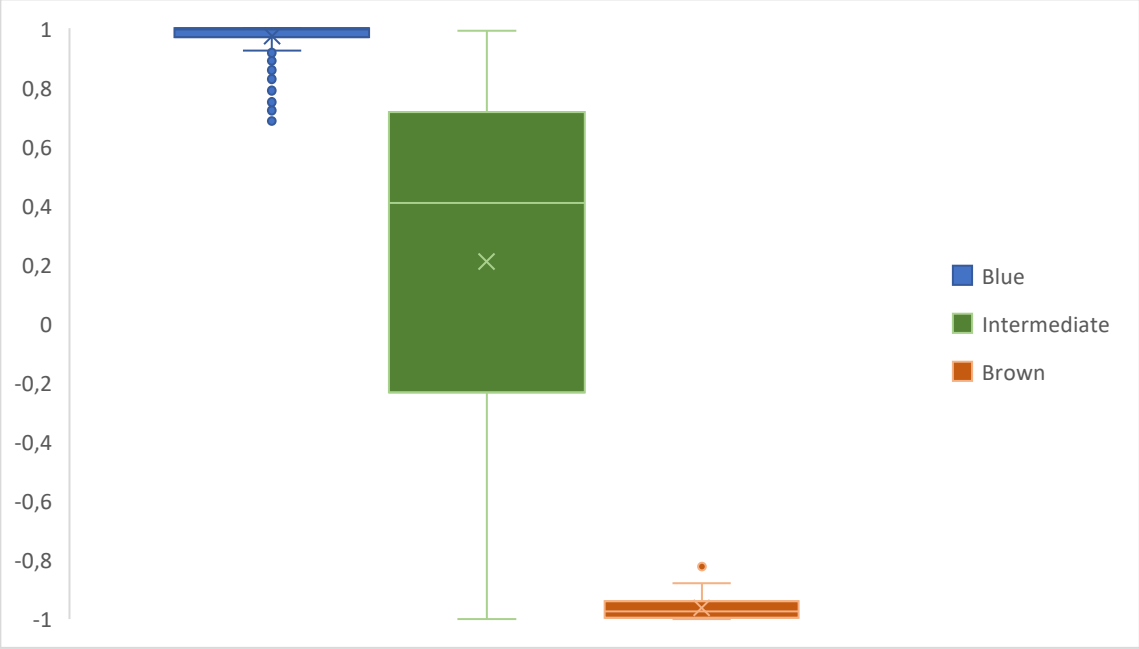


Figure 15: *PIE-score distribution compared to the observed eye colour when using three categories. The boxplots include PIE-scores and annotations of eye colour for 521 individuals categorised as blue (n=267), intermediate (n=183) and brown (n=71).*

During the rest of the data analysis, PIE-score was used to categorise eye colour and thresholds were set for blue (PIE ≥ 0.8), intermediate (-0.5 ≤ PIE ≤ 0.8) and brown (PIE-score ≤ -0.5). These thresholds are the same that have been used during the development of the IP NO and EC12 prediction models. Based on this, 291 (56%) individuals were categorised with blue eye colour, 123 (24%) individuals were categorised with intermediate eye colour and 107 (20%) individuals were categorised with brown eye colour, respectively.

5.4 Estimation of eye colour with prediction models

In this section, the performance and the prediction accuracies of the two prediction models IrisPlex and EC12 were evaluated. Additionally, the prediction results from EC12 were compared with the prediction results obtained by the IP NO model.

5.4.1 Eye colour prediction and prediction accuracy for the IrisPlex model

The IrisPlex model calculate a p-value for the three possible eye colour estimations (blue, intermediate and brown). The highest p-value is considered the estimated phenotype. For the 521 samples, a total of 368 (71%) individuals were predicted to have blue eye colour, none were predicted to have intermediate eye colour and 153 (29%) individuals were predicted to have brown eye colour. When comparing the categorised eye colour by PIE-score, with the

prediction results, 377 (72%) of the predictions were correct, leaving a total number of 144 (28%) predictions incorrect (Figure 16 A). Of the 144 incorrect predictions, 16 individuals were categorised as blue-eyed, but predicted to be brown, whereas five individuals were categorised with brown eye colour, but predicted to have blue eyes (Figure 16 A). As none of the 123 of the individuals with intermediate eye colour were predicted as so, the intermediate eyes were either predicted as blue (88 (72%)) or brown (35 (28%)) and accounted for the vast majority of the incorrect predictions (85%). This showed that the IrisPlex model, as previously reported, had difficulties with intermediate eye colour predictions.

When using the probability threshold of 0.5 and 0.7 recommended by Walsh et al.(3), the proportion of correct predictions decreased to 71% and 64%, respectively, and the incorrect predictions to 26% and 21%, respectively (Figure 16 B and C). Furthermore, 16 (3%) and 77 (15%) individuals had an inconclusive prediction with the thresholds of 0.5 and 0.7.

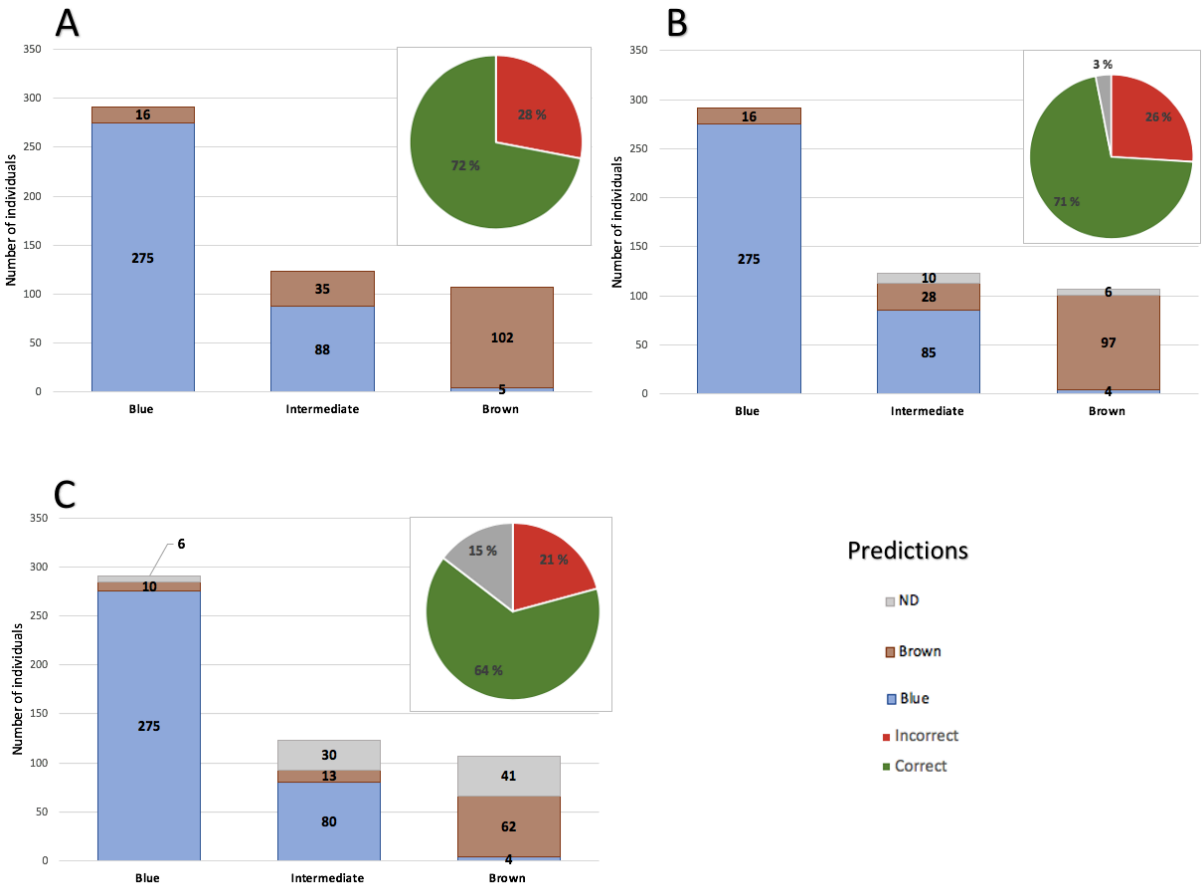


Figure 16: *Eye colour predictions obtained for the 521 DNA samples by the IrisPlex prediction model. The colours within the bars represent the predicted eye colour (blue, brown and inconclusive in grey),*

while each bar represent the categorised eye colour by PIE-score, blue (n=291), intermediate (n=123) and brown (n=107). The circle diagrams represent the correct (green), incorrect (red), and inconclusive (grey) predictions. A: prediction results without any probability threshold. B: prediction results with probability threshold ≥ 0.5. C: prediction results with probability threshold 0.7. ND=not defined and indicates eyes with an undefined prediction (p-value below the threshold).

To evaluate the prediction accuracy for the IrisPlex model, sensitivity, specificity and AUC were calculated. In this dataset, both blue and brown eye colour achieved the highest sensitivity of 95%, whereas brown eye colour had higher specificity than blue eye colour (88% versus 60%) (Table 6). This means that 275 of 291 individuals with blue eye colour and 102 of 107 individuals with brown eye colour were predicted correctly (Figure 16 A). Since none of the 123 individuals with intermediate eye colours were correctly predicted, the sensitivity for this category was 0%. Consequently, all individuals with none-intermediate eye colours were correctly predicted to have non-intermediate eye colour, resulting in a specificity of 100%. Sensitivity and specificity were also calculated with the probability thresholds 0.5 and 0.7 (Table 6). For the blue category, the sensitivity was approximately the same, while the specificity decreased as the threshold increased. This indicates that the model predicts fewer true negatives for blue eye colour when the threshold for conclusive predictions increases. Moreover, the number of false positives does not decrease a lot with more stringent thresholds. For the brown category, the sensitivities were also similar, whereas the specificities increased with increasing threshold, mainly because of fewer false positives. This indicates that more true negatives for brown eye colour are correctly predicted with more stringent thresholds (Table 6).

Table 6: *Sensitivity and specificity for each possible eye colour with no probability threshold, p≥0.5 and p≥0.7 obtained with the IrisPlex model.*

	Blue			**Intermediate**			**Brown**		
Threshold	None	p≥0.5	p≥0.7	None	p≥0.5	p≥0.7	None	p≥0.5	p≥0.7
Sensitivity (%)	95	95	96	0	0	0	95	96	94
Specificity (%)	60	58	47	100	100	100	88	89	94

The overall prediction accuracy was obtained by computing ROC-curves, and the corresponding AUC-values were calculated. The AUC-value determines the inherent ability of the prediction model to discriminate between correct and incorrect predictions. An AUC-value varies from 0.5 to 1.0, whereas a value equal 0.5 corresponds to random prediction and 1.0 corresponds to perfect prediction. The ROC-curve computed from the entire dataset obtained the highest AUC-value for brown eye colour (0.94) (Figure 17). The graph for brown eye colour is closest to the upper left corner, which represents a high true positive rate together with a low false negative rate. The obtained AUC-value for blue eye colour was slightly lower (0.84, still considered good), whereas the lowest AUC-value was seen for the intermediate-eyes (0.60) (Figure 17).

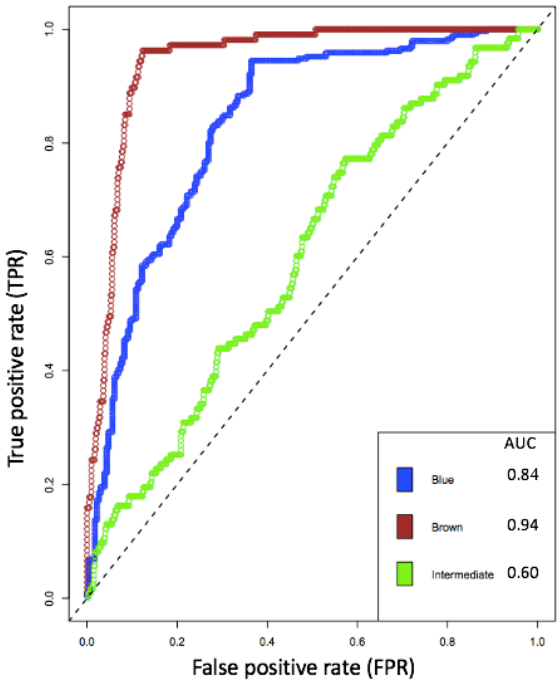
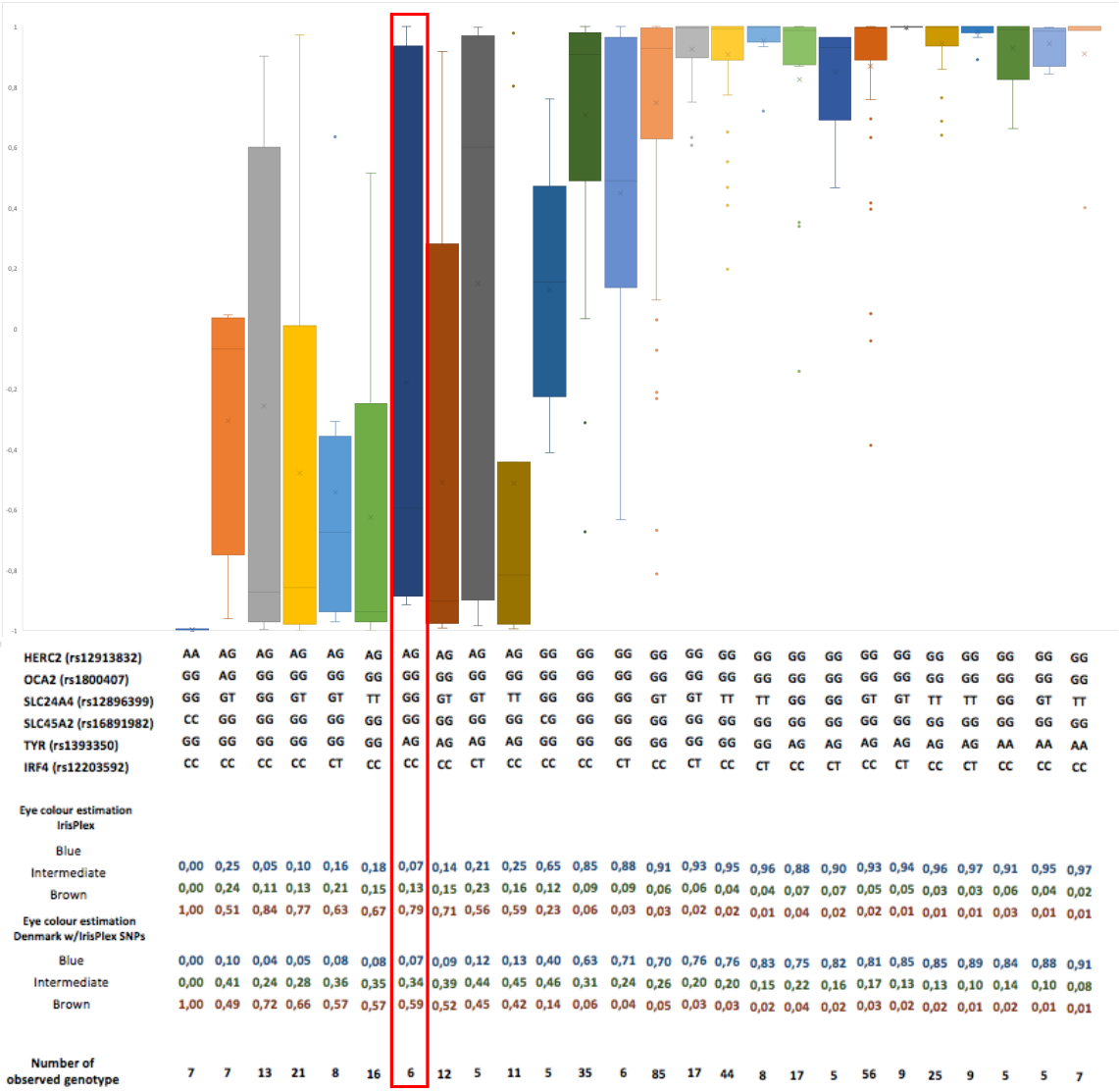


Figure 17: *ROC-curve analysis plotted as true positive rate versus false positive rate for the IrisPlex predictions for the 521 samples (without threshold). The dotted line demonstrates AUC=0.5.*

5.4.1.1 Genotypes and their eye colour predictions

When further analysing the IrisPlex profile for all 521 samples, a total of 75 different genotypes were identified, 26 of which were found in more than five individuals. The PIE-score distribution and the IrisPlex eye colour predictions for these 26 profiles are shown in Figure 18. The six SNPs included in the IrisPlex model have different prediction ranking. The

prediction accuracy is mainly dependent on the rs12913832 SNP. Therefore, this SNP is listed first in Figure 18. Only one IrisPlex profile with rs12913832 AA/TT was observed in more than five individuals (Figure 18). All seven individuals with this profile had low variation in the PIE-score, and the PIE-score indicated mostly brown pixels in the eye (median -1).This is also in concordance with the prediction of eye colour where all individuals obtained high probabilities for brown eye colour (p=1). The highest variations in PIE-score were observed in individuals with genotype AG/TC for rs12913832 with PIE-score ranging from -1 to 1 (from only brown pixels to only blue pixels in the iris). Although the PIE-score varied, all the individuals presented in Figure 18 with this genotype were predicted to have brown eyes with p-values higher than 0.5. This means that some individuals will be incorrectly predicted. The PIE-score did also vary between individuals genotyped GG/CC, but the majority of the IrisPlex profiles obtained a PIE-score corresponding to blue eye colour (PIE $\geq$ 0.8). The figure also illustrates that individuals with the same genotype can have different eye colour e.g. the IrisPlex profile highlighted in Figure 18 (rs12913832 AG/TC, rs1800407 GG/CC, rs12896399 GG/CC, rs16891982 GG, rs1393350 AG/CT and rs12203592 CC/GG). This particular genotype profile results in PIE-scores ranging from approximately -1, were all pixels are brown, to 1, were all pixels are blue. All the individuals with this IrisPlex profile are predicted to have brown eyes with p-value 0.79. This indicates that there are other SNPs or factors that affects the eye colour beside the six SNPs included in the IrisPlex prediction model.



HERC2 (rs12913832)	AA	AG	AG	AG	AG	AG	AG	AG	AG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
OCA2 (rs1800407)	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
SLC24A4 (rs12896399)	GG	GT	GG	GT	GT	TT	GG	GT	GT	TT	GG	GG	GG	GT	GT	TT	TT	GG	GG	GT	GT	TT	TT	GG	GT	TT
SLC45A2 (rs16891982)	CC	GG	GG	GG	GG	GG	GG	GG	GG	GG	CG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
TYR (rs1393350)	GG	GG	GG	GG	GG	GG	AG	AG	AG	AG	GG	GG	GG	GG	GG	GG	GG	GG	AG	AG	AG	AG	AG	AG	AA	AA	AA
IRF4 (rs12203592)	CC	CC	CC	CC	CT	CC	CC	CC	CT	CC	CC	CC	CT	CC	CT	CC	CT	CC	CT	CC	CT	CC	CT	CC	CC	CC
Eye colour estimation IrisPlex																										
Blue	0,00	0,25	0,05	0,10	0,16	0,18	0,07	0,14	0,21	0,25	0,65	0,85	0,88	0,91	0,93	0,95	0,96	0,88	0,90	0,93	0,94	0,96	0,97	0,91	0,95	0,97
Intermediate	0,00	0,24	0,11	0,13	0,21	0,15	0,13	0,15	0,23	0,16	0,12	0,09	0,09	0,06	0,06	0,04	0,04	0,07	0,07	0,05	0,05	0,03	0,03	0,06	0,04	0,02
Brown	1,00	0,51	0,84	0,77	0,63	0,67	0,79	0,71	0,56	0,59	0,23	0,06	0,03	0,03	0,02	0,02	0,01	0,04	0,02	0,02	0,01	0,01	0,01	0,03	0,01	0,01
Eye colour estimation Denmark w/IrisPlex SNPs																										
Blue	0,00	0,10	0,04	0,05	0,08	0,08	0,07	0,09	0,12	0,13	0,40	0,63	0,71	0,70	0,76	0,76	0,83	0,75	0,82	0,81	0,85	0,85	0,89	0,84	0,88	0,91
Intermediate	0,00	0,41	0,24	0,28	0,36	0,35	0,34	0,39	0,44	0,45	0,46	0,31	0,24	0,26	0,20	0,20	0,15	0,22	0,16	0,17	0,13	0,13	0,10	0,14	0,10	0,08
Brown	1,00	0,49	0,72	0,66	0,57	0,57	0,59	0,52	0,45	0,42	0,14	0,06	0,04	0,05	0,03	0,03	0,02	0,04	0,02	0,03	0,02	0,02	0,01	0,02	0,01	0,01
Number of observed genotype	7	7	13	21	8	16	6	12	5	11	5	35	6	85	17	44	8	17	5	56	9	25	9	5	5	7

Figure 18: *Correlation between PIE-scores, eye colour estimations by IrisPlex and IP NO for the IrisPlex profiles with more than five observations in our population sample. The SNPs are listed according their ranking in the IrisPlex prediction model*

The association between PIE-score and the three possible genotypes for the rs12913832 SNP for all 521 individuals were investigated in more details (Figure 19). Within the dataset, 365 (70%) individuals were genotyped rs12913832 GG/CC, 132 (25%) individuals were genotyped AG/TC and only 24 (5%) individuals were genotyped rs12913832 AA/TT. Of the 24 individuals genotyped as rs12913832 AA/TT, only seven had four Norwegian grandparents. As expected, a strong association between PIE-score and the rs12913832 genotypes was observed (Figure 19). In addition, a significant difference in PIE-score were observed between all the three genotype results (Mann Whitney: $p < 0.0001$). The highest variation in PIE-score was observed in individuals with the rs12913832 AG/TC genotype, whereas the AA/TT

individuals showed the least variation. The majority of the individuals with the genotype GG/CC showed a PIE-score corresponding to blue eye colour, but several individuals (n=90) did also have a PIE-score that would categorise them with intermediate (n=85) or brown (n=5) eye colour. Although the highest variation in PIE-score was observed in individuals with rs12913832 AG/TC genotype, it was the rs12913832 GG/CC genotype that accounted for the majority of incorrect predictions. The total number of incorrect predictions for rs12913832 GG/CC genotype were 89 (62%), 51 (35%) for rs12913832 GA/CT individuals and 4 (3%) for the rs12913832 AA/TT individuals.

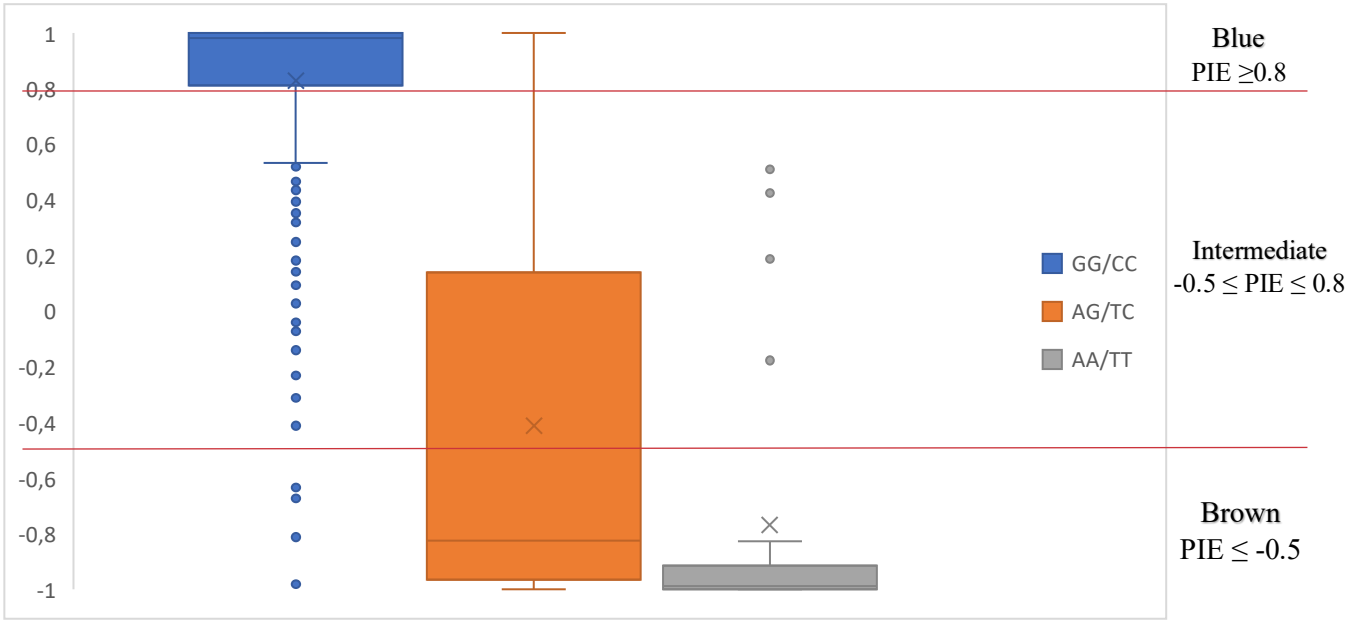


Figure 19: *PIE-score distribution compared to the three possible genotypes for HERC2 rs12913832 (GG/CC (n=365), AG/TC (n=132) and AA/TT (n=24)). The red lines indicate the selected PIE-score range for each eye colour category; blue, intermediate and brown.*

The eye colour to the individuals with PIE-score outside the boxes in Figure 19 were further investigated. Several outliers were observed in individuals genotyped as GG/CC for the rs12913832 SNP. Five of these individuals had a PIE-score below -0.5 and therefore classified with brown eyes (Figure 19). Four of these individuals were, however, incorrectly predicted using the IrisPlex model with high prediction probabilities for blue eye colour (≥ 0.85) Figure 20 A). The corresponding eye photos are in concordance with the PIE-score, clearly displaying a non-blue eye colour (Figure **20** A). The fifth individual with PIE-score below -0.5 was correctly predicted to have brown eyes (described in more detail in section 5.4.2.1).

The four outliers observed for individuals genotyped as rs12913832 AA/TT had a PIE-score within the intermediate range. They were predicted to have brown eye colour with a high p-value for brown eyes (≥ 0.96) (Figure 19 and Figure 20 B). However, these four individuals do not display a brown eye colour (Figure 20 B). No outliers were observed for the individuals typed as AG/TC for the rs12913832 SNP. However, 16 individuals displayed a blue eye colour with a PIE-score ≥ 0.8 Figure 20 C). Not in concordance with the observed eye colour, they were predicted with brown eye colour (p≥0.56) (Figure 20 C).

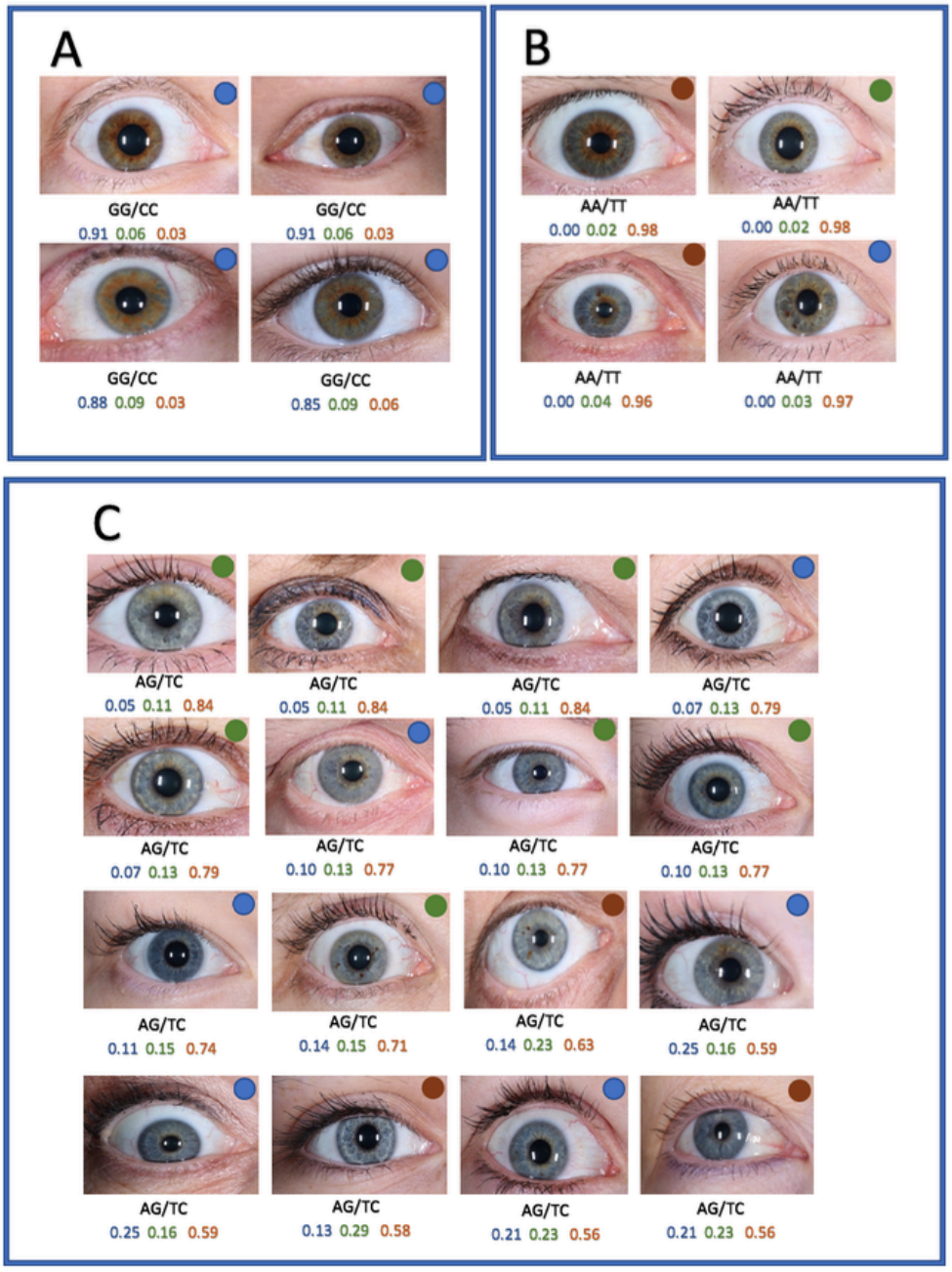


Figure 20: *Selection of individuals with incorrect predictions by the IrisPlex model. A: Individuals with brown eyes predicted to have blue eyes (PIE-score ≤ -0.5). B: Individuals with intermediate eyes*

predicted to have brown eyes (-0.5 ≤ PIE ≤ 0.8).C: Individuals with blue eyes predicted to have brown eyes (PIE-score ≥ 0.8). The p-value for each of the three categories shown in blue (blue category), green (intermediate category) and brown (brown category) numbers below the photos. Blue dot=predicted as blue eye colour with the EC12 prediction model. Brown dot= predicted as brown eye colour with the EC12 prediction model. Green dot= predicted as intermediate eye colour with the EC12 prediction model.

5.4.2 Eye colour prediction and prediction accuracy for the prediction models from Denmark- EC12 and IP NO

The IP NO and the EC12 prediction models both calculates a p-value for each of the three possible eye colour categories; blue, intermediate and brown. IP NO predicts eye colour based on the same SNPs included in the IrisPlex model, while EC12 predicts eye colour based on 11 of the 12 SNPs originally included in this model (Table 1). The 12th SNP (rs7120151) was excluded from the data analysis due to imbalanced alleles and the insufficient amplification of the A allele. In contrast to the IrisPlex model, these two models were generated and tested on the same population (the Norwegian population). Consequently, the prediction results from the novel EC12 model could be compared to the IP NO results, where the differences in the results would be due to the SNPs included in the models and the mathematical algorithms used to make the predictions.

Of the 521 samples, the EC12 model predicted a total of 370 (71%) individuals to have blue eye colour, 36 (7%) individuals to have intermediate eye colour and 115 (22%) individuals to have brown eye colour. Whereas IP NO predicted 359 (69%) individuals to have blue eyes, 35 (7%) individuals to have intermediate eyes and 127 (24%) individuals to have brown eyes. The number of correct predictions without any thresholds applied was 375 (72%) to 393 (75%) with the IP NO and EC12 models, respectively (Figure 21 A and D). The number of incorrect predictions was 146 (28%) and 128 (25%), respectively. Note that these numbers are overall prediction results, meaning that not all correct predicted individuals by the IP NO model were the same correct predicted individuals with the EC12 model.

The majority of the correct predictions for EC12 were found within the blue eye colour category, were 278 (96%) individuals where correctly predicted as blue-eyed. This is three more than the 275 (95%) correctly predicted individuals with IP NO (Figure 16 A and Figure

21 D). An increase of correct predictions were also observed within the brown eye colour category, where 96 (90%) individuals were correctly predicted by EC12 while IP NO correctly predicted 84 (79%) individuals (Figure 16 A and Figure 21 D). Unlike the IrisPlex model, where no individuals were predicted to have intermediate eye colours, both IP NO and EC12 predicted individuals to have intermediate eyes. EC12 correctly predicted 19 (16%) with intermediate eye colour, which was an increase of three compared to the 16 (13%) obtained by the IP NO model (Figure 16 A and Figure 21 D). To sum up, a slightly increase of correct predictions were observed in all categories with the novel EC12 predicted model.

When using the probability thresholds of 0.5 and 0.7, the proportion of inconclusive profiles were 2% and 32%, respectively, with the EC12 model (Figure 21 E and F). With the IP NO model, a higher number of inconclusive profiles were observed with the thresholds 0.5 (10%) and 0.7 (43%) than what was observed for the EC12 model (2% and 32%, respectively) (Figure 21 B and C). Even though the total of inconclusive profiles were lower with the EC12 model, this model obtained a higher number of correct predictions using either of the two thresholds, than IP NO. (Figure 21 B, C, E and F). This means that the EC12 model correctly predict more individuals with a larger dataset, than the IP NO model does with a smaller one.

IP NO EC12

Threshold

None

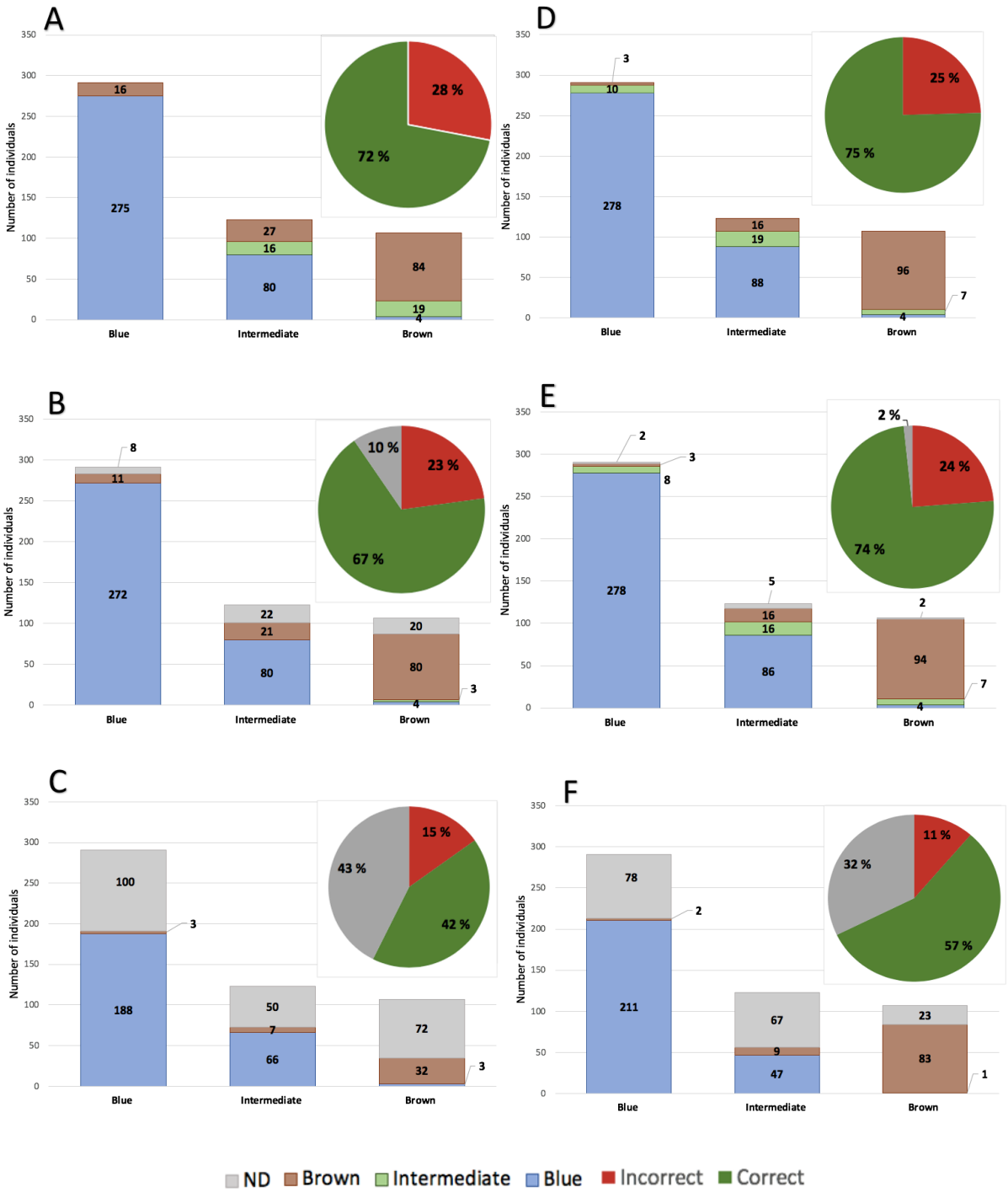


Figure 21: Phenotypic predictions obtained by IP NO (A, B and C) and EC12 (D, E and F) prediction models for the 521 samples. The colours within the bars represent the predicted eye colour (blue, brown and inconclusive in grey), while each bar represent the categorised eye colour by PIE-score, blue (n=291), intermediate (n=123) and brown (n=107). The circle diagrams represent the correct (green), incorrect (red), and inconclusive (grey) predictions. A and D: prediction results without probability threshold. Band E: prediction results with probability threshold ≥ 0.5. C and F: prediction results with probability threshold ≥ 0.7. ND=not defined and indicates eyes with an undefined prediction (p-value below the threshold).

The sensitivity, specificity and AUC were also calculated for evaluating the prediction accuracy of the IP NO and the EC12 prediction models.

For the EC12 model, blue eye colour achieved the highest sensitivity (96%), but also the lowest specificity (60%) of the three eye colour categories. Approximately the same values were obtained by the IP NO and IrisPlex models (Table 7 and Table 6). This means that the three models predict blue eye coloured individuals with high accuracy, while the probability for the models for correctly predicting non-blue individuals is lower. The sensitivity and specificity for intermediate eye colour were also nearly the same for both models, with a very low probability to correctly predict individuals with intermediate eyes (13% with IP NO and 15 % with EC12), and a high probability of correctly predicting non-intermediate individuals (95% with IP NO and 96% with EC12). For the brown eye colour category, there was a substantial increase in sensitivity from IP NO with a sensitivity of 79% to EC12 with 90%, while the specificity increased from 90% to 95% (Table 7). This indicates that the probability for correct prediction of brown eye coloured and none-brown coloured individuals are higher with the EC12 model than it is for the IP NO model.

By using the probability thresholds of 0.5 and 0.7, the EC12 model had a sensitivity of 90% and 99% for brown eye colour, respectively. However, the specificity was similar as without thresholds (96%) (Table 7). This means that with threshold 0.7, the sensitivity and specificity for brown eye colour were close to 100%, meaning that nearly all brown eye coloured and all non-brown eye coloured individuals were correctly predicted. For the IP NO model, the sensitivity was slightly lower with threshold 0.7 (91%), while the specificity were the same as EC12 (96%) (Table 7). Within the blue eye colour category, EC12 had a sensitivity of 96%

with threshold 0.5 and 99% with threshold 0.7. This is approximately the same obtained by IP NO which had a sensitivity of 96% with 0.5 threshold and 98% with 0.7 threshold. However, the specificity for blue increased with increasing threshold (60% and 66%) for the EC12 model, while the opposite was observed for the IP NO model (55% and 36%) (Table **7**). This is similar to what was observed for the IrisPlex model (Table 6). This indicates that the probability to correctly predict true negatives increases with increasing thresholds with EC12, but it decreases with IP NO and IrisPlex.

Table 7: *Sensitivity and specificity for each possible eye colour with none probability threshold, ≥0.5 and ≥0.7 for the IP NO and the EC12 prediction models.*

		Blue			**Intermediate**			**Brown**		
	Threshold	None	p≥0.5	p≥0.7	None	p≥0.5	p≥0.7	None	p≥0.5	p≥0.7
IP NO	Sensitivity (%)	95	96	98	13	0	0	79	92	91
	Specificity (%)	63	55	36	95	99	100	90	92	96
EC12	Sensitivity (%)	96	96	99	15	14	0	90	90	99
	Specificity (%)	60	60	66	96	96	100	95	95	96

The overall prediction accuracy for both IP NO and EC12 were obtained by computing ROC-curves and calculating their corresponding AUC-values. All the AUC-values obtained by the EC12 model were higher than the AUC-values obtained from the IP NO model. Both models obtained low AUC-values for intermediate eye colour, with AUC-value of 0.68 (EC12) and 0.59 (IP NO) (Figure 22 A and B). For blue and brown eye colour, both models obtained a high prediction accuracy expressed with high AUC-values. EC12 obtained a high prediction accuracy for brown eye colour (AUC 0.97), which was an increase from the IP NO (AUC 0.93) (Figure 22 A and B). For blue eye colour, the AUC-values were slightly lower with 0.84 for EC12 and 0.81 for IP NO. However, these values were still considered good.

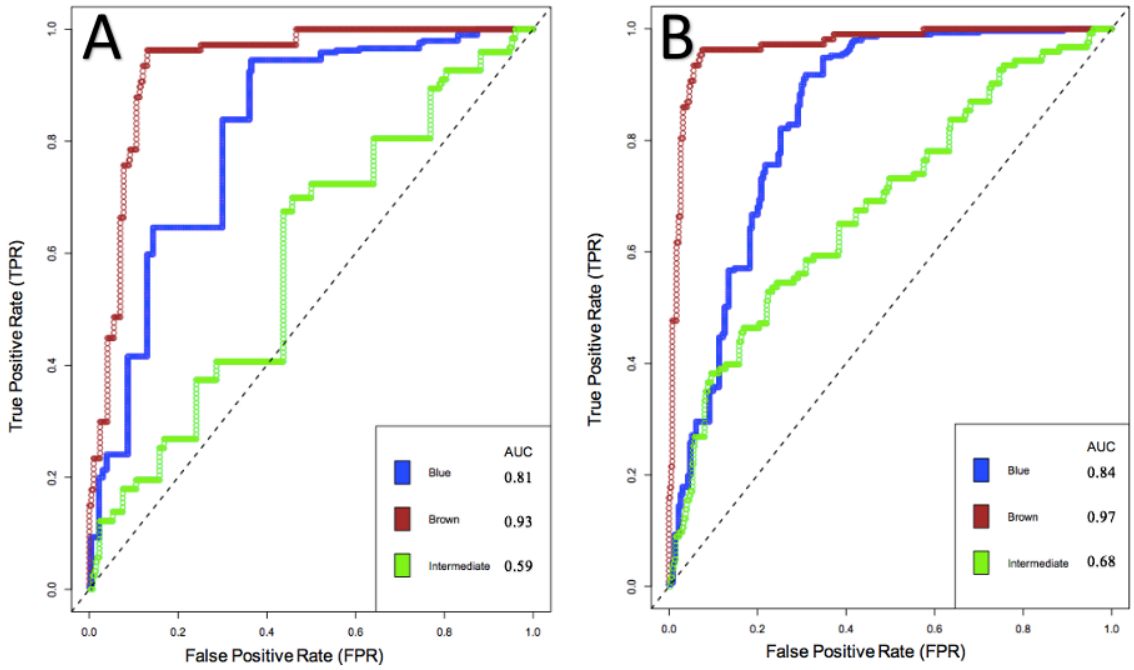


Figure 22: *ROC-curves plotted as true positive rate versus false positive rate for the dataset with 521 samples (without threshold). A: ROC-curve obtained from IP NO model. B: ROC-curve obtained from EC12 model. Brown eye colour achieved the highest prediction accuracy represented as AUC-value for both models. This graph is closest to the upper left corner, which represents a high true positive rate together with a low false negative rate. The dotted line demonstrates AUC=0.5.*

5.4.2.1 Genotypes and their eye colour predictions

The three possible rs12913832 genotypes GG/CC, AG/TC and AA/TT and their association with PIE-score (Figure 19), were further compared with the prediction results obtained by IP NO and EC12. All the three models incorrectly predicted four out of the five individuals genotyped as GG/CC rs12913832 with brown eyes (PIE ≤ -0.5), to have blue eye colour (Figure 19). These four individuals clearly displaying a non-blue eye colour (Figure **20** A). The fifth individual were correctly predicted to have brown eye colour by the IrisPlex model (Figure 23).Visual observation of this eye demonstrates an obvious brown eye colour (Figure 23), however, both EC12 and IP NO predicted this individual incorrectly to have intermediate eyes.

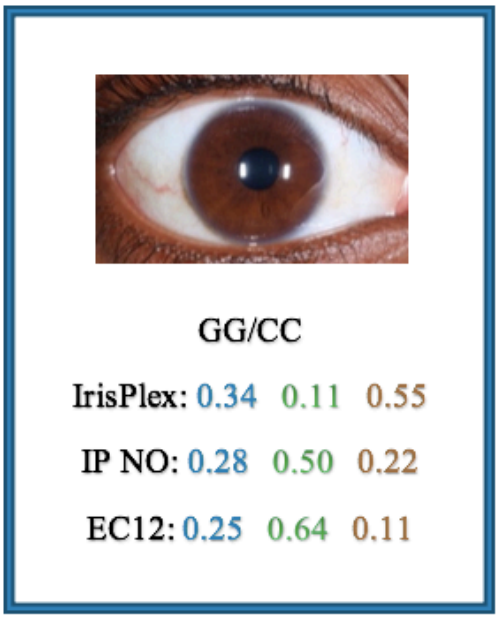


Figure 23: One individual with genotype GG/CC for rs12913832 and PIE-score under -0.5. The IrisPlex model correctly predicted this individual to have brown eyes, whereas IP NO and EC12 both predicted this individual to have intermediate eyes. The p-value for each of the three categories are represented as blue (blue category), green (intermediate category) and brown (brown category).

IP NO (and IrisPlex) incorrectly predicted the four outliers with the genotype rs12913832 AA/TT and a PIE-score within the intermediate eye colour range to have brown eye colour (Figure 19 and Figure 20 B). The EC12 model correctly predicted one of these individuals to have intermediate eye colour (illustrated with a green dot in Figure 20 B).

The 16 individuals with the genotype rs12913832 AG/TC and blue eye colour, were all incorrectly predicted to have brown eye colour by IP NO (and IrisPlex) Figure 20 C). The EC12 performed better in predicting the eye colours of these individuals, and correctly predicted blue eyes for six of them (blue dots in Figure 20 C). The remaining individuals were predicted to have intermediate eye colour (n=7) (green dots in Figure 20 C) or brown eye colour (n=3) (brown dots in Figure 20 C).

6 Discussion

6.1 Establishment of PCR-SBE-CE assay

Prior to this project, the Centre of Forensic Genetics in Tromsø sequenced SNPs relevant in forensic genetics on the MPS machine MiSeq FGx (Illumina) with the ForenSeq DNA signature Prep kit (Verogen). This kit sequence over 200 markers, including the six IrisPlex SNPs for predicting of eye colour. MPS is, however, an expensive method, especially if only a few samples and/or SNPs are analysed per run. As a cheaper alternative to MPS, SBE is a preferred method within the forensic field (52). This method does not require additional equipment for what is already used in forensic laboratories, but this method was not established at the Centre of Forensic Genetics' laboratory. Seven weeks of this project were spent at Section of Forensic Genetics, Copenhagen, Denmark to establish a PCR-SBE-CE assay. The assay amplifying the 12 SNPs with eye colour association included in the EC12 ("EyeColour12") model, where eight of them are new SNPs and four are IrisPlex SNPs. Additionally, the two remaining IrisPlex SNPs were included in the assay so all SNPs for both models could be genotyped in the same assay.

The main problem during the optimisation of the SBE-assay was the weak amplification of the rs7120151 alleles, and especially the A allele. None of the optimisation steps that were tested, improved the amplification of the alleles in multiplex reactions. The amplification of the marker was better in singleplex reactions, with clearly defined peaks in the electropherogram, but, still, the allele peak heights were imbalanced. The reason for the unsuccessful amplification of this marker is unknown. One possible explanation for the weak amplification of rs7120151 alleles, could be that the primer binds better to strand carrying the G allele compared to the strand carrying the A allele. However, there was not time to further investigate this or to perform more optimising of the assay during my stay in Copenhagen. Therefore, the SNP was amplified in an alternative workflow consisting of a singleplex PCR reaction prior to a multiplex SBE reaction with the remaining SNPs in the assay. However, the amplification of this SNP was still considered unreliable and were therefore excluded from the subsequent LOOCV data analysis in the Norwegian population.

Since the optimisation of the assay were not completed, the sensitivity study was not performed. In forensic genetics, evidence samples are often in poor condition, and it is important to know the assays limitations (65). Although the assay demonstrated great potential

with no significant difference in peak heights for samples with high amounts of DNA (between 2 and 20 ng), it is important to also know the assays minimum input DNA concentration.

6.2 Prediction of eye colour

Predicting externally visible characteristic (EVC) from biological material obtained from a crime scene could potentially be a helpful tool for criminal investigation. Especially in cases without suspects and eyewitnesses can EVC function as a "biological witness" and thereby aid the investigation. In this project, the performance of the already well-known IrisPlex prediction model was evaluated (3). Additionally, the performance of a novel prediction model, EC12, based and tested on the Norwegian population, was evaluated.

Prediction of eye colour with the IrisPlex model resulted in 368 (71%) individuals predicted with blue eye colour, none individuals predicted with intermediate eye colour and 153 (29%) individuals predicted with brown eye colour. Overall, 377 (72%) of the prediction were correct and 144 (28%) of the predictions were incorrect (Figure 16 A). The intermediate eye colour category accounted for the vast majority of incorrect predictions (85%), demonstrating the model's limitation to predict this eye colour. This result is also observed in other studies (46,47), and acknowledged in publications of the model (3,38). In our population, intermediate eye colour is the second most prominent observed eye colour (24%). Thereby will the model's limitations in predicting intermediate eye colour have impact in its performance in our population.

The prediction accuracy of the IrisPlex model was expressed by calculating sensitivity, specificity and AUC-values. The model showed a good performance for blue and brown eye colour with high AUC-values of 0.84 and 0.94, respectively. Higher AUC-values have been reported for blue eye colour. In a study by Kastelic et al. they evaluated the IrisPlex model in a Slovenian population. They reported an AUC-value of 0.97 for blue eye colour (47). A possible explanation for this could be that they only categorised individuals to have blue eye colour if the colour was homogenous. This means that eye colour perceived as blue, but had brown spots, were categorised as intermediate instead of blue. Or, it could be due to differences in the allele frequency in the populations. Moreover, both blue and brown eye colour had a high sensitivity of 95%, but brown eye colour had a higher specificity than blue (88% versus 60%) (Table 6). This means that the model predicts true positives with a high accuracy, but the accuracy for prediction of true negatives is lower, especially for blue eye colour.

Although the IrisPlex showed a good performance for blue and brown eye colour, the performance for predicting intermediate eye colour was not considered great. None of the 123 individuals with intermediate eye colour were predicted to have intermediate eyes, and consequently, the sensitivity of the models was 0%. Although the specificity was 100%, it was only because all individuals with non-intermediate eye colour were correctly predicted to have non-intermediate eyes. The low performance of the assay was also reflected in a low AUC-value of 0.6, which is close to a random prediction The same results were also observed in an Italian population with a high proportion of individuals with intermediate eye colour (25%), which is similar to the proportion in our population (24%) (46). In the same Italian study, performed by Salvoro et al., they also investigated whether the absence of intermediate predictions was due to the absence of the respective IrisPlex profiles predicting intermediate eye colour in their population. They simulated all possible IrisPlex profiles and calculated the corresponding prediction values and demonstrated thereby that the intermediate phenotype is rarely the one with the highest probability for any of the possible IrisPlex profiles (Fig. 2 in (46)). Some studies have reported predicted intermediate eye colours when using a re-calibrated IrisPlex model with population specific allele frequencies (46,49,66) This corresponds to the IP NO model in this project, which was trained and tested in the Norwegian population using LOOCV. Using this adjusted model, 35 individuals was predicted with intermediate eye colour. However, this only increased the sensitivity from 0% (IrisPlex) to 13% (IP NO).

When the data was analysed using probability thresholds of 0.5 and 0.7, the proportion of incorrect and correct predictions decreased (Figure 16 B and C). Also, more individuals became inconclusive, meaning that prediction of EVC for individuals with certain IrisPlex profiles will not be achievable. While the sensitivity remained approximately the same for blue and brown eye colour (95%), the specificity decreased and increased, respectively. For blue eye colour, the specificity decreases from 60 % (no threshold) to 47% (0.7), indicating that several non-blue individuals have been predicted to have blue eyes. For instance, 88 individuals with intermediate eye colour were incorrectly predicted to have blue eyes (no threshold), and only eight of these was undefined with a threshold of 0.7 (Figure 16 C). This means that the probability values for blue eyes are high for the remaining 80 individuals and they are still incorrectly predicted. Moreover, maybe thresholds for blue eye colour should not be considered in our population, especially not threshold of 0.7. For brown eye colour, on the other hand, the

specificity increased with increasing thresholds, and with a threshold of 0.7 the specificity was 94%. A high sensitivity together with a high specificity indicates a high probability for correctly predicting individuals to have brown eye colour and individuals to have non-brown eyes. In a study performed by Walsh et al., a sensitivity of only 56% were achieved for brown eye colour with threshold 0.7 (3). A low sensitivity for brown eye colour were also obtained in the Slovenian population with 58% (47). On the other hand, a high sensitivity of 99% and 100% was obtained in two Italian populations (46,67). A possible explanation for some of the variation reported in sensitivities, and also specificity, may be variation in the evaluation of the eye colours. Some studies used qualitative method for eye colour categorisation (visual inspection) (3,47), whereas other studies used quantitative categorisation (different software) (67).

In this study, the categorisation of eye colour was performed with both a subjective (qualitative) and objective (quantitative) method. With the subjective eye colour categorisation, nine untrained individuals living in northern Norway categorised the eye photos in the study population to be either blue, intermediate-blue, intermediate-brown or brown. For this project intermediate-blue and intermediate-brown eye colour were merged into one intermediate eye colour category. The most prominent eye colour in our dataset were blue, which was observed in 267 (51%) individuals, intermediate eye colour was observed in 183 (35%) individuals and brown eye colour in 71 (14%) individuals. When analysing all photos quantitatively using the objective DIAT software and calculating PIE-score, 291 (56%) individuals were categorised with blue eye colour, 123 (24%) individuals with intermediate eye colour and 107 (20%) individuals with brown eye colour.

Categorisation of eye colour by visual observation is difficult since humans may perceive eye colours differently. In a study performed by Meyer and co-workers, 30 Danes evaluated 442 eye photos. A disagreement of 33% were observed when the participants categorised eye colours in three categories; blue, intermediate and brown (68). The perception of eye colour will especially vary throughout the world. Individuals living in Norway are relatively accustomed to see blue eye coloured individuals and are thereby probably biased to categorise a less blue eye into the intermediate category. Furthermore, if the categorisation of eye colours in this study were performed by individuals from another population were blue eye colour is not so frequently observed (e.g. Africa or Asia), some of the individuals with this less blue eye colour would probably been annotated as blue-eyed (3). The human bias during categorisation

of eye colours makes it difficult to reproduce the findings and also compare results between studies (48). This bias can be limited by using objective eye colour categorisation. An objective way to assess eye colour will also make it easier to compare results from different studies. The DIAT software, developed by Andersen et. al (48) calculate the number of blue pixels versus brown pixels in the iris (a PIE-score). A PIE-score range from 1, indicating that all pixels are blue, to -1, which indicate that all pixels are brown. Eye photos with a mix of blue and brown pixels will generate a PIE-score in between 1 and -1. In this project the PIE-score correlated well with the human perception of eye colour (spearman correlation = -0.88) (Figure 15). The PIE-score distribution was clearly defined within the three categories (blue, intermediate and brown), where a statistical significance in PIE-score were observed (Mann Whitney: $p < 0.001$). As already mentioned, the proportion of individuals categorised when using PIE-score or observation from photos are not similar. A possible explanation may be that there is a bias when merging two categories into three (intermediate-blue and intermediate-brown into intermediate). If for instance the same participants were asked to annotate the same photos a second time, but with only three categories; blue, intermediate and brown, the proportion of individuals with intermediate eye colour individuals would truly been decreased, while blue and brown eye colour would increase. This is also what have been observed in our dataset (data not shown). Another reason may be the PIE-score thresholds used when categorising the eye colours in blue, intermediate and brown. From Figure 15, the distribution of PIE-score demonstrates whiskers that overlap between intermediate and blue category, and intermediate and brown category. This means that individuals observed with a particular eye colour were categorised with another eye colour when using the PIE-score thresholds. For instance, individuals observed with intermediate eye colour show a PIE-score ranging from 1 (upper whisker) to -1 (lower whisker). With the PIE-score threshold of PIE $\leq$ - 0.5, several of the individuals observed with intermediate eye colour are categorised with brown eyes. If this threshold were lowered, the proportion of observed intermediate eyes categorised with brown eyes would decrease. Nevertheless, this just proves the difficulties regarding categorisation of eye colour and that it is advantages and disadvantages with both subjective and objective methods.

To investigate the effect different IrisPlex profiles have on eye colour, the profiles were compared to their corresponding PIE-score and probability value (Figure 18). Here, the IrisPlex profiles with rs12913832 AG/CT genotype demonstrated a high variation in PIE-score, ranging from 1 to -1. The same variation was also observed in individuals with the same IrisPlex profile,

indicating that they have eye colours ranging from clearly blue to clearly brown. Andersen et al. did reported the same results (63), indicating that the six SNPs included in eye colour prediction by the IrisPlex model do not completely explain all eye colour variations. In forensic genetics, this is a huge limitation. In such kind of cases, the IrisPlex prediction tool may in fact mislead the police and will not be considered reliable.

The prediction of eye colour is mainly driven by rs12913832, where the GG/CC genotype drive the prediction towards blue eyes and AA/TT and AG/CT towards brown. Other studies have reported that most of the incorrect or inconclusive results comprise individuals heterozygous for rs12913832 (46,49). In our population, this is not the case, and the majority of incorrect predictions were found in individuals homozygous GG/CC. The total number of incorrect predictions were 89 (62%) for this genotype, while the number of incorrect predictions for rs12913832 GA/CT were 51 (35%). A possible explanation may be the high proportion of GG/CC individuals in our population (70%), while the proportion of heterozygous individuals are 25% (genotype frequency for all SNPs are presented in Appendix 10.3). Also, another possible explanation can be the already discussed PIE-score threshold of ≥ 0.8 for blue eye colour. This threshold is a result of observation from north-European people. Because of the already mentioned bias of visual perception of eye colour in populations, the threshold may have been different if other populations had categorised the eye colours. In Figure 19, the majority of the rs12913832 GG/CC individuals have a PIE-score above 0.8, but there is also several individuals with PIE-score below this threshold, where the lower whisker is approximately at 0.6. These individuals have been observed to have blue eye colour but are categorised with intermediate because of the PIE-score threshold.

Furthermore, only 24 (5%) individuals with the rs12913832 AA/TT genotype were observed in our population. In this study, individuals genotyped with AA/TT for rs12913832 showed the least variation in PIE-score with a PIE-score median -0.99, corresponding to brown eye colour (Figure 19). This is similar to what was reported by Andersen et al. with a PIE-score median of -0.87 for this genotype (63). All the 24 individuals with this particular genotype were predicted to have brown eye colour (data not shown). Out of these, only seven had four Norwegian grandparents. This is not unexpected since it has been reported a low frequency of this genotype in Europe (3). However, four out of these seven individuals were categorised with intermediate eye colour. These four individuals are presented in Figure 20 B and are clearly displaying non-brown eye colours, but were incorrectly predicted to have brown eyes

with $p \geq 0.96$. If this result is representative of the Norwegian population cannot be determined by small samples size and have to be further investigated further. All the remaining individuals, without four Norwegian grandparents, were correctly predicted with brown eye colour.

The evaluation of the IrisPlex model has, among other things, shown that several individuals with the same IrisPlex profile results in various PIE-score, and thus various eye colours. It is therefore likely that other genetic markers could be involved in determination of eye colour, and hence predict eye colour more accurate.

The EC12 model included four of the IrisPlex SNPs in addition to eight other SNPs with eye colour association. This model differs from the IrisPlex model in several ways: 1) the model was generated and tested on our Norwegian population using the leave-one-out-method 2) the model made the predictions with other mathematical algorithms than IrisPlex 3) phenotype evaluation was based on a quantitative method using PIE-score. With directly comparison of the IrisPlex and EC12 model, the mentioned differences may be factors that can explain the differences in the prediction results, and not the different SNPs included. Therefore, the IP NO model, including the six IrisPlex SNPs, was generated using the same reference population as EC12 and the same method for phenotype evaluation. Thereby, the prediction results from EC12 could be compared to the results from IP NO.

In this study, the prediction of eye colour with the EC12 model have shown to be slightly more accurate than the IP NO model. Overall, the EC12 model correctly predicted 393 (75%) individuals, which is 18 more correct predictions than the 375 (72%) individuals IP NO obtained (Figure 21 A and D). An increase of correct predictions with EC12 were observed in all three eye colour categories, which indicate a more accurate prediction of all the eye colours with this model.

When calculating the prediction accuracy, expressed by sensitivity, specificity and AUC-values for both models, EC12 was the model that performed best. Within the brown eye colour category, both the sensitivity and specificity increased. The sensitivity increased considerably from 79% (IP NO) to 90% (EC12), while the specificity increased from 90% (IP NO) to 95% (EC12) (Table 7). This indicates that the EC12 model predicts both brown eye coloured individuals and non-brown eye coloured individuals with a higher accuracy than IP NO, and also evident by the increase in AUC-values. IP NO obtained an AUC-value of 0.93 for brown

eye colour, while EC12 obtained a value of 0.97 (Figure 22 A and B). Both models also showed a good performance for blue eye colour, where EC12 obtained a slightly higher AUC-value than IP NO (0.84 and 0.81, respectively). Dembinski et al. reported an even higher AUC value of 0.97 for blue eye colour. They adjusted the IrisPlex model based on the allele frequency in their sample data from a United State population (49).

Unlike the IrisPlex model, where none individuals were predicted to have intermediate eyes, both EC12 and IP NO predicted individuals with intermediate eye colour. However, the performance for this eye colour was not considered good with a low AUC-value of 0.68 with EC12 and an even lower with IP NO (0.59). Only 19 of the 36 individuals predicted with intermediate eye colour was correctly predicted with the EC12 model, resulting in a sensitivity of only 15% and a specificity of 96%. This was similar to what IP NO obtained. This indicates that both models predict intermediate-eyed individuals and non-intermediate eyed individuals with the same, low accuracy. Salvoro et al. suggested that other genetic mechanisms (e.g. factors that causes presence of dark spots in light coloured eyes) than the simple blue/brown proportion could be involved in determination of intermediate eye colour (46). Additionally, epistatic effects may also be studied when investigating the genetics of eye colour, which may explain the limitations of the models which mainly accounting for additive effects (69).

Since none of the models in this study shows great potential for prediction of intermediate eye colour, maybe a prediction model with only two eye colour categories should be considered. As already discussed, the intermediate eye colour is difficult to categorise. In Figure 18, the individuals with rs12913832 GG/CC genotype demonstrate a variation in PIE-score, with some individuals showing PIE-score below the threshold of 0.8. Still, the majority of these individuals' PIE-score are centred in the upper range of the figure. With a two-category prediction model, the PIE-score thresholds would be different (e.g. PIE-score 0), consequently these individuals would be categorised as blue, and not intermediate.

With the use of probability thresholds of 0.5 and 0.7, several profiles were inconclusive with both models. However, EC12 had a lower proportion of inconclusive profiles than IP NO (Figure 21 B, C, E and F), still, this model obtained a higher proportion of correct predictions. When calculating the sensitivity and specificity with probability thresholds, the sensitivity and specificity for brown eye colour increased in both models. With a threshold of 0.7 EC12 obtained a sensitivity of 99%, and a specificity of 96%, which indicates a high prediction

accuracy to predict brown and non-brown eye colour. IP NO obtained the same specificity, but the sensitivity was lower (91%). This implies that the EC12 model correctly predicts more true positives with increasing thresholds, than IP NO does. For blue eye colour, a small increase in sensitivity was observed for EC12 (96% without threshold and 99% with threshold 0.7) and IP NO (95% without threshold and 98% with threshold 0.7). However, the specificity decreases in IP NO, which is similar to what was observed with IrisPlex (Table 6 and Table 7). For the EC12 model, the specificity increased, which implies that this model correctly predicts more true negatives with increasing thresholds, than IP NO does. However, with a threshold of 0.7, 32% of the profiles was inconclusive with EC12, and thus the eye colour to these individuals cannot be predicted. This is a disadvantageous since the model then will be limited to only predict eye colour for a certain number of individuals.

Since there is many possible genotype profiles for the SNPs included in EC12, the different profiles could not be compared to their corresponding PIE scores and probability values like the IrisPlex profiles were (Figure 18). Thus, the IP NO and EC12 prediction results for the individuals with PIE-score outside the boxes in Figure 19 were further investigated and compared to the IrisPlex prediction results.

The 16 individuals with the rs12913832 AG/CT genotype and with a blue eye colour (PIE above 0.8), were all incorrectly predicted to have brown eye colour by IP NO while EC12 correctly predicted seven of these individuals (blue dots in Figure 20 C). Furthermore, EC12 predicted six of the individuals to have intermediate eye colour (green dots in Figure 20 C), and only three with brown eye colour (brown dots in Figure 20 C). The four individuals with rs12913832 AA/TT genotype in Figure 20 B, were all incorrectly predicted to have brown eye colour by IP NO. However, EC12 correctly predicted one of these individuals to have intermediate eyes (green dot in Figure 20 B). One of the individuals were predicted to have blue eyes (blue dot in Figure 20 B) and two were still predicted to have brown eye colour (brown dot in Figure 20 B).

The four individuals presented in Figure 20 A, with rs12913832 GG/CC genotype, were all incorrectly predicted by all three models. A fifth individual with this genotype and a PIE-score below -0.5 were correctly predicted with brown eyes by the IrisPlex model (Figure 23). However, IP NO and EC12 incorrectly predicted this individual to have intermediate eyes. The rs12913832 GG/CC genotype has shown to reduce the pigmentation level in the iris (37), but,

the individual in Figure 23 is clearly displaying a brown eye colour. rs12913832 GG/CC genotyped individuals with light eye colour have also been reported by other studies (37,63). Peitroni et al. evaluated all possible genotype combinations of the IrisPlex assay and their corresponding probability values. Here, they reported that one genotype combination with rs12913832 GG/CC and rs16891982 CC/GG resulted in prediction of brown eye colour (67). This corresponds with the genotype result for the individual in Figure 23. Nevertheless, the rs16891982 is an ancestry informative marker, were the CC/GG genotype is found in high frequency in African individuals (70). This was also the ancestry of the individual presented in Figure 23. It is therefore questionable if rs16891982 can alter the eye colour determined by rs12891982 GG/CC, or if the phenotypic outcome is due to their ancestry (67).

7 Conclusion

One of the aims in this study was to establish a PCR-SBE-CE assay for prediction of eye colour at Section of Forensic Genetics. Because of limited time spent at Section of Forensic Genetics, this was not completed, mainly because of the rs7120151 SNP which was not successfully amplified in multiplex PCR reaction. Furthermore, the assay was successfully established at Centre of Forensic Genetics laboratory and their DNA samples from their research biobank were genotyped.

The evaluation of the IrisPlex model has shown that the model predicts blue and brown eye colour with high accuracy. However, the prediction results have also shown that some individuals with clearly blue eye colour have been predicted to have brown eyes and some individuals with clearly brown eye colour have been predicted to have blue eyes. Thus, the IrisPlex model may mislead the police during an investigation. Nevertheless, the major issue with this model was the prediction of intermediate eye colour, where none out of the 123 individuals with intermediate eye colour in our population were correctly predicted.

IP NO and EC12 are both generated on the same underlying reference population and the same phenotype evaluation. Like IrisPlex, both models have shown great potential to successfully predict blue and brown eye colours, where EC12 demonstrated to be slightly better than IP NO. However, the performance of prediction of intermediate eye colour was limited. Intermediate eye colours are difficult to categorise, and a two-category prediction tool may be considered in future prediction models.

Overall, the EC12 prediction model has shown an increase in correct predictions compared to the IP NO model as well as the IrisPlex model. This increase may imply that the additionally SNPs included in this model has an improving effect on eye colour prediction.

8 Future perspective

In the future, the optimisation of the PCR-SBE-CE assay for EC12 must be completed. Eventually, could the rs7120151 SNP that did not work in multiplex PCR be replaced with another SNP. Additionally, a sensitivity study must be performed to know the assays limitations with low DNA concentrations. The assay should also be validated for use in forensic case work and fulfil the guidelines published by the Scientific Working Group on DNA Analysis Methods (SWGDAM).

Studies regarding the retrieval of intermediate eye colours, including this one, describes a low AUC-value for predicting intermediate eye colour. This strongly suggests that the SNPs used today are insufficient regarding estimating intermediate eye colours. More genetic information regarding intermediate eye colour are thus necessary to achieve a more accurate prediction of the intermediate eye colours

Eventually, since all three prediction models have demonstrated that prediction of intermediate eye colour is challenging, and these results correlates with what other studies have reported, a two-category prediction model should be tested. The EC12 model are developing, and it has already been generated a two-category prediction model in the Norwegian population, but these results have not been analysed yet.

9 References

1. Lu C, Verbridge SS. Microfluidic methods for molecular biology. Microfluid Methods Mol Biol. 2016;1–376.
2. Walsh S, Lindenbergh A, Zuniga SB, Sijen T, Knijff P De, Kayser M, et al. Forensic Science International : Genetics Developmental validation of the IrisPlex system : Determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int Genet. 2011;5(5):464–71.
3. Walsh S, Liu F, Ballantyne KN, Oven M Van, Lao O, Kayser M. Forensic Science International : Genetics IrisPlex : A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet. 2011;5(3):170–80.
4. Kayser M. Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genet. 2015;
5. Andersen JD. Genetic investigations of eye and skin colours. 2015.
6. Kirsebom MK. Forensic DNA phenotyping. 2017.
7. Daiger SP, Sullivan LS, Bowne SJ. Genetic Mechanisms of Retinal Disease. Retina [Internet]. 2013 Jan 1 [cited 2020 Feb 18];624–34. Available from: https://www.sciencedirect.com/science/article/pii/B978145570737900031X
8. Strachan T, Read A. Human Molecular Genetics [Internet]. CRC Press; 2018. Available from: https://books.google.no/books?id=vviADwAAQBAJ
9. Syvänen AC. Accessing genetic variation: Genotyping single nucleotide polymorphisms. Nat Rev Genet. 2001;2(12):930–42.
10. Twyman RM. Single-Nucleotide Polymorphism (SNP) Analysis. Encycl Neurosci. 2009;871–5.
11. Butler JM. Fundamentals of forensic DNA typing. Burlington, Mass: Academic Press; 2010.
12. Single nucleotide polymorphisms as genomic markers for high-throughput pharmacogenomic studies – Atlas of Science [Internet]. [cited 2020 Apr 2]. Available from: http://atlasofscience.org/single-nucleotide-polymorphisms-as-genomic-markers-for-high-throughput-pharmacogenomic-studies/
13. Hunt R, Sauna ZE, Ambudkar S V, Gottesman MM, Kimchi-Sarfaty C. Silent (Synonymous) SNPs: Should We Care About Them? In: Komar AA, editor. Single Nucleotide Polymorphisms: Methods and Protocols [Internet]. Totowa, NJ: Humana Press; 2009. p. 23–39. Available from: https://doi.org/10.1007/978-1-60327-411-1_2
14. Kayser M, Liu F, Janssens ACJW, Rivadeneira F, Lao O, van Duijn K, et al. Three Genome-wide Association Studies and a Linkage Analysis Identify HERC2 as a Human Iris Color Gene. Am J Hum Genet. 2008;82(2):411–23.
15. Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, Unterländer M. Direct evidence for positive selection of skin , hair , and eye pigmentation in Europeans during the last 5 , 000 y. 2014;111:2–7.
16. Ashraf Q, Galor O. The “Out of Africa” hypothesis, human genetic diversity, and comparative economic development. Am Econ Rev. 2013;103(1):1–46.
17. Frost P. European hair and eye color A case of frequency-dependent sexual selection ? 2006;27:85–103.
18. Sturm RA, Duffy DL. Human pigmentation genes under environmental selection. Genome Biol. 2012;13(9).
19. Why did humans evolve blue eyes? 2015.
20. Sturm RA, Frudakis TN. Eye colour: Portals into pigmentation genes and ancestry. Trends Genet. 2004;20(8):327–32.
21. Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, et al. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Hum Genet. 2008;123(2):177–87.

22. D'Mello SAN, Finlay GJ, Baguley BC, Askarian-Amiri ME. Signaling pathways in melanogenesis. Int J Mol Sci. 2016;17(7).

23. Yamaguchi Y, Brenner M, Hearing VJ. The regulation of skin pigmentation. Vol. 282, Journal of Biological Chemistry. American Society for Biochemistry and Molecular Biology; 2007. p. 27557–61.

24. Schallreuter KU, Kothari S, Chavan B, Spencer JD. Regulation of melanogenesis - controversies and new concepts. Exp Dermatol [Internet]. 2008 May [cited 2020 Apr 15];17(5):395–404. Available from: http://doi.wiley.com/10.1111/j.1600-0625.2007.00675.x

25. Gudbjartsson DF, Sulem P, Stacey SN, Goldstein AM, Rafnar T, Sigurgeirsson B, et al. ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. Nat Genet. 2008;40(7):886–91.

26. Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, et al. Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS Genet. 2010;6(6):1–20.

27. Manga P, Kromberg JGR, Box NF, Sturm RA, Jenkins T, Ramsay M. Rufous oculocutaneous albinism in Southern African blacks is caused by mutations in the TYRP1 gene. Am J Hum Genet. 1997;61(5):1095–101.

28. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, et al. Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet. 2008;40(7):835–7.

29. Praetorius C, Grill C, Stacey SN, Metcalf AM, Gorkin DU, Robinson KC, et al. XA polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. Cell [Internet]. 2013;155(5):1022. Available from: http://dx.doi.org/10.1016/j.cell.2013.10.022

30. Duffy DL, Iles MM, Glass D, Zhu G, Barrett JH, Höiom V, et al. IRF4 variants have age-specific effects on nevus count and predispose to melanoma. Am J Hum Genet. 2010;87(1):6–16.

31. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet. 2007;39(12):1443–52.

32. Rebbeck TR, Kanetsky PA, Walker AH, Holmes R, Halpern AC, Schuchter LM, et al. P gene as an inherited biomarker of human eye color. Cancer Epidemiol Biomarkers Prev. 2002;11(8):782–4.

33. Raposo G, Marks MS. Melanosomes - Dark organelles enlighten endosomal membrane transport. Nat Rev Mol Cell Biol. 2007;8(10):786–97.

34. Fuller BB, Spaulding DT, Smith DR. Regulation of the catalytic activity of preexisting tyrosinase in black and caucasian human melanocyte cell cultures. Exp Cell Res. 2001;262(2):197–208.

35. Kayser M, Liu F, Janssens ACJW, Rivadeneira F, Lao O, van Duijn K, et al. Three Genome-wide Association Studies and a Linkage Analysis Identify HERC2 as a Human Iris Color Gene. Am J Hum Genet. 2008;

36. Sturm RA, Duffy DL, Zhao ZZ, Leite FPN, Stark MS, Hayward NKK, et al. A Single SNP in an Evolutionary Conserved Region within Intron 86 of the HERC2 Gene Determines Human Blue-Brown Eye Color. Am J Hum Genet. 2008;

37. Andersen JD, Pietroni C, Johansen P, Andersen MM, Pereira V, Børsting C, et al. Importance of nonsynonymous OCA2 variants in human eye color prediction. 2016;

38. Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens ACJW, et al. Eye color and the prediction of complex phenotypes from genotypes. Curr Biol. 2009;19(5):192–3.

39. Ruiz Y, Phillips C, Gomez-Tato A, Alvarez-Dios J, Casares De Cal M, Cruz R, et al. Further development of forensic eye color predictive tests. Forensic Sci Int Genet. 2013 Jan 1;7(1):28–40.

40. Allwood JS, Harbison SA. SNP model development for the prediction of eye colour in New Zealand. Forensic Sci Int Genet. 2013 Jul 1;7(4):444–52.

41. Hart KL, Kimura SL, Mushailov V, Budimlija ZM, Prinz M, Wurmbach E. Improved eye- and skin-color prediction based on 8 SNPs. Croat Med J. 2013;54(3):248–56.

42. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36.

43. Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L, et al. Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet. 2014 Mar 1;9(1):150–61.

44. Walsh S, Chaitanya L, Breslin K, Muralidharan C, Bronikowska A, Pospiech E, et al. Global skin colour prediction from DNA. Hum Genet. 2017 Jul 1;136(7):847–63.

45. Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L, et al. Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet. 2014;

46. Salvoro C, Faccinetto C, Zucchelli L, Porto M, Marino A, Occhi G, et al. Performance of four models for eye color prediction in an Italian population sample. Forensic Sci Int Genet. 2019 May 1;40:192–200.

47. Kastelic V, Pośpiech E, Draus-Barini J, Branicki W, Drobnič K. Prediction of eye color in the slovenian population using the irisPlex SNPs. Croat Med J. 2013;54(4):381–6.

48. Andersen JD, Johansen P, Harder S, Christoffersen SR, Delgado MC, Henriksen ST, et al. Genetic analyses of the human eye colours using a novel objective method for eye colour classification. Forensic Sci Int Genet [Internet]. 2013;7(5):508–15. Available from: http://dx.doi.org/10.1016/j.fsigen.2013.05.003

49. Dembinski GM, Picard CJ. Evaluation of the IrisPlex DNA-based eye color prediction assay in a United States population. Forensic Sci Int Genet. 2014 Mar 1;9(1):111–7.

50. Definition of massively parallel sequencing - NCI Dictionary of Genetics Terms - National Cancer Institute [Internet]. [cited 2020 Apr 15]. Available from: https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/massively-parallel-sequencing

51. Metzker ML. Sequencing technologies the next generation. Nat Rev Genet. 2010;11(1):31–46.

52. Fondevila M, Børsting C, Phillips C, de la Puente M, Consortium E-N-E, Carracedo A, et al. Forensic SNP genotyping with SNaPshot: Technical considerations for the development and optimization of multiplexed SNP assays. Forensic Sci Rev. 2017 Jan 1;29:57–76.

53. Verogen. Verogen Forensic Genomics Solution [Internet]. [cited 2020 Apr 6]. Available from: www.verogen.com/products

54. Verogen. ForenSeq DNA Signature Prep Kit.

55. Mehta B, Daniel R, Phillips C, McNevin D. Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review. Vol. 131, International Journal of Legal Medicine. Springer Verlag; 2017. p. 21–37.

56. Garibyan L, Avashia N. Polymerase chain reaction. J Invest Dermatol. 2013;133(3):1–4.

57. Polymerase Chain Reaction (PCR) [Internet]. [cited 2019 Oct 23]. Available from: https://www.ncbi.nlm.nih.gov/probe/docs/techpcr/

58. Butz H, Patócs A. Brief Summary of the Most Important Molecular Genetic Methods (PCR, qPCR, Microarray, Next-Generation Sequencing, etc.). In: Igaz P, Patócs A, editors. Genetics of Endocrine Diseases and Syndromes. Cham: Springer International Publishing; 2019. p. 33–52.

59. ExoSAP-IT™ PCR Product Cleanup Reagent [Internet]. [cited 2019 Oct 23]. Available from: https://www.thermofisher.com/order/catalog/product/78201.1.ML#/78201.1.ML

60. GeneScan™ 120 LIZ™ dye Size Standard [Internet]. [cited 2020 Apr 24]. Available from: https://www.thermofisher.com/order/catalog/product/4324287#/4324287

61. GeneMapper Software Version 4.0: SNaPshot® Kit Analysis: Getting Started Guide (English) [Internet]. [cited 2020 Apr 11]. Available from: https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-

Assets%2FLSG%2Fmanuals%2Fcms_042037.pdf&title=R2VuZU1hcHBlciBTb2Z0d2FyZSBWZXJzaW9uIDQuMDogU05hUHNob3QmcmVnOyBLaXQgQW5hbHlzaXM6IEdldHRpbmcgU3RhcnRlZCBHdWlkZSAoRW5nbGlzaCAp

62. Jia Y. Chapter 3 – Real-Time PCR. Vol. 112, Methods in Cell Biology. Elsevier; 2012. 55–68 p.

63. Andersen JD, Johansen P, Harder S, Christoffersen SR, Delgado MC, Henriksen ST, et al. Genetic analyses of the human eye colours using a novel objective method for eye colour classification. Forensic Sci Int Genet. 2013 Sep 1;7(5):508–15.

64. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Vol. 4, Caspian Journal of Internal Medicine. Babol University of Medical Sciences; 2013. p. 627–35.

65. Dixon LA, Murray CM, Archer EJ, Dobbins AE, Koumi P, Gill P. Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. Forensic Sci Int. 2005 Nov 10;154(1):62–77.

66. Dario P, Mouriño H, Oliveira AR, Lucas I, Ribeiro T, Porto MJ, et al. Assessment of IrisPlex-based multiplex for eye and skin color prediction with application to a Portuguese population. Int J Legal Med. 2015 Nov 1;129(6):1191–200.

67. Pietroni C, Andersen JD, Johansen P, Andersen MM, Harder S, Paulsen R, et al. The effect of gender on eye colour variation in European populations and an evaluation of the IrisPlex prediction model. Forensic Sci Int Genet. 2014;11(1):1–6.

68. Meyer OS, Børsting C, Andersen JD. Perception of blue and brown eye colours for forensic DNA phenotyping. Forensic Sci Int Genet Suppl Ser. 2019 Dec 1;7(1):476–7.

69. Pośpiech E, Draus-Barini J, Kupiec T, Wojas-Pelc A, Branicki W. Gene-gene interactions contribute to eye colour variation in humans. J Hum Genet. 2011 Jun 7;56(6):447–55.

70. Phillips C, Prieto L, Fondevila M, Salas A, Gómez-Tato A, Álvarez-Dios J, et al. Ancestry Analysis in the 11-M Madrid Bomb Attack Investigation. Quintana-Murci L, editor. PLoS One [Internet]. 2009 Aug 11 [cited 2020 Jun 10];4(8):e6583. Available from: https://dx.plos.org/10.1371/journal.pone.0006583

10 Appendix

10.1 Table of PCR and SBE primer sequences for the PCR-SBE-CE assay.

SNP_ID	SNP	Forward primer	Reverse primer	SBE primer (incl. tail)	SBE primer length	Direction (SBE primer)
rs10131374	[C/T]	GATTTCCTTACCGTTACGGG	AATCTTGTGAGCCTGCTGAG	ctctctctctctctctctctctctctctctctctctctGGAGAGAGTCCTGTTGG	57	R
rs1126809	[C/T]	AATGGGTGCATTGGCTTCTG	GCTCAAAAATACTGCAGAGG	ctctctctctctctctctAGAGGACGGTGCCTT	33	R
rs121918166	[C/T]	TTATGGTCACAGGCGTGAAG	CATCATGCTCTGTCTCATCG	ctctctctctctctctctctctctctctctctcGGAGCATGGTGGTGA	48	F
rs12896399	[G/T]	CTGGCGATCCAATTCTTTGT	GACCCTGTGTGAGACCCAGT	ctctctctctctctctctctctctctctctctTCTTTAGGTCAGTATATTTTGGG	53	F
rs12913832	[T/C]	TCAACATCAGGGTAAAAATC	GGCCCCTGATGATGATAGC	ctctctctctctctctctctctctGCGTGCAGAACTTGACA	41	R
rs1408799	[A/G]	ATCAAAACTGGTTCATCCAC	CTTAGCACATTGTCTGGCTC	ctctctctctctctctctctctctctctctctctctctctctTCGGAGCACATGGTCA	58	R
rs16891982	[C/G]	TCCAAGTTGTGCTAGAC	CGAAAGAGGAGTCGAGGTTG	ctctctctctcAAACACGGAGTTGATGCA	29	F
rs1800401	[C/T]	AGGTCGTTGTTTCGTTCTGC	TGAGCCATCAAAAGAGGGAC	ctctctctctctctctctctctctctctcGGTGTCCATCAGCATC	45	R
rs1800407	[C/T]	TGAAAGGCTGCCTCTGTTCT	CGATGAGACAGAGCATGATGA	ctctctcGCATACCGGCTCTCCC	23	F
rs4904927	[A/G]	AAAGTTCACATGGTGGCAGG	GTTGTTTCCTCTGAACCCTG	ctctctctctctctctctctTGAGAAAGGCTGCCTTCTG	41	F
rs7120151	*[A/G]*	*GTGTGCCTATCCCACCCTTA*	*CTGGTTGGAAGGATGCAAAT*	*ctctctctctctctctctACTTTTTAGCACACA*	*35*	F
rs74653330	[C/T]	CTCAGCTCTTGGTTGGAAAC	AAGTCCTGATTGCAGAAGTG	ctctctctctctctctctctctctctctctctctctCCCCGATGGCAGTGG	53	F
rs1393350*	[C/T]	TTCCTCAGTCCCTTCTCTGC	GGGAAGGTGAATGATAACACG	ctctctctctctctctctctctTTTGTAAAAGACCACACAGATTT	47	R
rs12203592*	[C/T]	ACAGGGCAGCTGATCTCTTC	GCTAAACCTGGCACCAAAAG	ctctctctctctctctTTTGGTGGGTAAAAGAAGG	35	F

All primers were designed by Section of Forensic Genetics, Copenhagen Denmark, expect for those marked with *. *=Original IrisPlex (primers in house)

10.2 instruction for eye colour annotation in Dataturks

Instructions for eye colour annotation

Thank you for accepting this invitation to participate in our study on perception of eye colour.

Perception of eye colour is a subjective measure and we are interested in **your** perception.

To be able to participate you need to sign up at www.dataturks.com. Please use the same email you have provided us. You will receive a verification email from "Mohan" to verify your email address upon signing up.

After signing up you will find a project in your profile with your name on it. This is your personal project and it is important that only you annotate the photos. Click on the project and "start tagging".

For every photo, there are four choices: **Blue**, **Intermediate-blue**, **Brown** and **Intermediate-Brown**. Intermediate-blue is something in between blue and brown, but more blue than brown. Intermediate-brown is also something in between blue and brown, but more brown than blue. Remember there is none incorrect answers.

Select a category by clicking on the label next to the photo, or use the short cuts (ctrl+1/2/3/4). Annotate each photo with only **ONE** label and click on "Move to done". Please be aware if you by a mistake select the wrong label, you need to remove that label when selecting a new label.

There are about 640 photos for annotation. It takes approximately one hour to annotate all photos. You can pause, and close down the program, at all-time and continue where you left off at any time.

Please let us know if you have any questions.

Thank you for your contribution,

Marianne and Nina

10.3 genotype frequencies for the 14 SNPs included in the PCR-SBE-CE assay