Functional Genomics and Eye Color Prediction

Gruber M, Bogdon E, Blonde M, and Coyle HM*

Department of Forensic Science, University of New Haven, United States

*Corresponding author: Heather Miller Coyle, Department of Forensic Science, University of New Haven, 300 Boston Post Road, West Haven, CT 06516 United States, Tel: 203479-4595; Email: Hcoyle@newhaven.edu

Abstract

The ability to reconstruct a recognizable face from skeletal remains is a useful investigative tool for human identification. Functional genomics plays a role in facial reconstruction through the identification of the regulation of pigmentation pathways. Functional genomics here refers to the study of how genes and intergenic sequences contribute to metabolic pathways and work together to produce a particular phenotype. This form of genetic information adds and confirms the physical traits that help with a visual identification of a person by allowing for the biological predictions of pigmentation for eye, hair, and skin color in facial reconstruction. The principle of SNP testing for eye color prediction from human tooth DNA using both destructive and nondestructive DNA extraction methods is presented using the IrisPlex eye color prediction software. This IrisPlex eye color prediction software is not one hundred percent accurate, and we have explored the classification issues behind those discrepancies by examining DNA from blue, brown, and intermediate eye color donors as well as those donors exhibiting heterochromia (mixed eye colors). The importance of correct prediction of eye color for facial reconstruction is to aid in correct identification of skeletal remains through forensic phenotyping investigation. Irisplex is useful for the correct prediction of blue and brown eye individuals but is less able to distinguish between the subcategories of grey, green, heterochromia and hazel for the intermediate category.

Keywords: DNA; SNP; Functional Genomics; Eye Color; Forensic Science

Abbreviations: SNP: Single Nucleotide Polymorphisms; DNA: deoxyribonucleic acid; EVC: Externally Visible Characteristics; UV: Ultra Violet; FDP: Forensic DNA Phenotyping.

Introduction

DNA can be collected from the bones and teeth of the skeleton to create an image of the person using a process called forensic DNA phenotyping (FDP). Pigmentation based externally visible characteristics (EVCs) – hair color, skin color, and eye color – are the most studied of the FDP phenotypic characteristics to date and have the most accurate prediction models [1-3]. These models assist in the anthropological reconstruction of human remains, particularly the skull [2].

In 1843, Petrequin documented a phenotypic eye color classification scale that had five colors: grey, blue, hazel, brown and black [4]. In 1845, Cornaz divided iris color into two main color categories: blue and brown. Since then, phenotypic iris color categories have varied significantly to classify eye color correctly. Coon (1939) (light/mixed/dark) and Diaz (2004) (blue/hazel/brown) categorized eye color differently using three categories [5-6]. Wilde (1862) (grey/blue/hazel/brown), Brownlee (1912) (pure blue/grey or pale yellow/yellow/dark brown) and Tocher (1908)

(blue/grey/mixed/brown) used four separate categories [7]. Seddon (1990) used five categories of eye color to classify iris phenotypes [8]. Ridell (1942) (blue/grey/green/yellow/tan/chocolate) and Simionescu (2014) (blue/grey/green/hazel/light brown/dark brown) both used six categories [7]. For even more elaborate classification schemes, Galton (1886) used 8 categories while Mackey (2011) used nine total different categories [7]. Grive and Morant (1946), Martin (1903) and Fraser (2008) used fourteen, sixteen and twenty-four categories, respectively, for their iris color pattern classifications [7]. Clearly, iris color and pattern are a complex polygenic trait that is challenging to classify and even more challenging to predict using biological modeling strategies. Methods vary for classification of iris eye color and include painted glass eye models, photographs, and digital images, spectrophotometry of melanin, hyper spectral iris color analysis and high spatial resolution photographs with software that measures quantitatively the number of blue and brown pixels to provide an iris eye color score [4]. The history of eye color classification illustrates the difficulty in formally classifying individuals by this trait.

As a genetic approach to eye color classification, single nucleotide polymorphisms (SNPs) in eye color regulation genes have been used to categorize and compare to phenotypic information. Though eye color regulation is not yet fully understood, many point mutations have been studied that affect the phenotype. One predictive model, IrisPlex, uses a three-category system to match the genetics to the physical characteristics. IrisPlex is a forensic phenotyping tool that uses six SNPs to predict eye color with the outcomes being blue, brown, or intermediate [9]. These six SNPs have overall AUC values of 0.93 for brown, 0.91 for blue, and 0.72 for intermediate eye colors, respectively [10,11]. AUC, or Area under the ROC Curve, is a measure of performance of a model with the AUC measuring how well the predictions rank.

The lack of pigmentation results in blue eye color. Full pigmentation is classified as brown. The P protein, encoded by OCA2, is involved in melanocyte maturation, and affects the quantity and quality of melanin stored in the iris. Most individuals have a similar number of melanocytes but the melanosomes within the melanocytes that store the pigment differ in number as well as the amount of melanin. The result is varying shades of eye color pigmentation based on genetic pre-coding of an individual during fetal development. At birth, a child's eye color will continue to darken to the final true color of the eye on exposure to light until the child reaches its first year of life. Brown and blue eye colors are simple to predict based on genetics. However, the intermediate category of eye color that includes gray, green, violet, and hazel is the genetic part still being deciphered. True violet-colored eyes are rare (1% of the population) and occur with albinism due to a lack of pigment and the reflection of light

off red blood vessels. Grey eyes are like blue but have spots of gold and brown in the iris. An estimated 3% of the population has grey eyes because of little to no pigmentation like blue eyes but they have more collagen, and the light scattering effect yields the grey color. Green eyes are present in 2% of the population and have lipochrome but low melanin that yields the distinctive color through the Tyndall Effect. Hazel eye color is present in 5% of the population and is a combination of brown, green, and gold with the inner rim of the iris often being a different color than the outer rim. The mixed color appearance is due to Rayleigh scattering of light.

Much of our current genetic understanding of eye color comes from genome-wide association studies, linkage studies, and candidate gene studies [12]. At first, the OCA2 gene was thought to be the most informative gene for eye color expression since mutations in the OCA2 gene can result in pigmentation disorders [12,13]. OCA2 is clearly related to eye pigmentation, but more recent studies have shown that the HERC2 gene is the most informative gene for eye color [9,12,13]. The promoter region for OCA2 is in an intron in the HERC2 gene, and changes to the rs12913832 enhancer SNP in HERC2 regulate the binding site for OCA2 transcription, effectively controlling the OCA2 gene [13]. Since OCA2 is associated with human P protein production, any changes to the expression of OCA2 will change the amount of melanin in the melanocytes, changing the eye pigmentation. Our focus in this study was on these two main genes that control iris pigmentation in a basic transcription regulation model. Irisplex also has four additional genes that affect the iris color. The 6 different IrisPlex SNPs in total are [rs12913832 (HERC2), rs1800407 (OCA2), rs12896399 (SLC24A4), rs16891982 (SLC45A2) (MATP), rs1393350 (TYR), and rs12203592 (IRF4)].

Theory

Tyndall Effect: The Tyndall Effect is a phenomenon where light is scattered by small, suspended particles in its path. It was first described by John Tyndall, a physicist, in the 1800's. It is caused by reflection from the interior walls of the particles, and refraction and diffraction as the light passes through the particles. This phenomenon is related to green eye color [14].

Rayleigh Scattering: Rayleigh scattering is from the scattering of light off molecules of air and can scatter from small particles with less than 1/10th the wavelength of light as well. Rayleigh scattering of light off of air particles gives the sky its' blue color. The process was described in 1871 by Lord Rayleigh, an English physical scientist who also received the Nobel Prize in 1904 in physics. Rayleigh scattering is related to hazel eye color [15].

AUC Values: AUC, or Area under the ROC Curve, is a measure of performance of a model with the AUC measuring how well

the predictions rank. The AUC ranges from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0. A model that has predictions 100% correct has an AUC of 1.0. This rank is similarly used in IrisPlex for eye color prediction accuracy by SNP analysis. The prediction certainty level is represented under the eye color score p-value Tables 1 & 2. The closer the value is to 1.0, the more certainty that an accurate prediction with IrisPlex has been made. The eye color column with the highest p-value is the more accurate prediction.

Materials and Methods

- **Experiment 1:** For our SNP assay, we chose to use aged baby and wisdom teeth as the DNA source. Teeth are made up of three hard tissue layers- enamel, dentin, and cementum- and a soft tissue layer, pulp. Enamel covers the top of the tooth and covers the dentin. The cementum covers the dentin under the gum line, mostly on the roots of the tooth. Pulp is on the inside of the tooth and is surrounded by the dentin. Dentin, cementum, and pulp all contain DNA. Cementum and pulp are valuable sources of nuclear DNA and dentin is a good source for mitochondrial DNA. We chose to use cementum in a nondestructive enzyme digestion method for the aged baby teeth and a destructive method for the aged wisdom teeth that targeted the pulp for DNA.

- **Experiment 2:** To investigate the distinction between eye colors categories used in Irisplex genetically, we collected buccal swabs from donors that were visually assessed and photographed for eye color and pattern (Figure 1). These samples were processed for DNA and SNP tested with the HERC2 and OCA2 assays. These buccal swabs also served as known reference samples with recent collection dates and represent non aged samples.

- **Experiment 3:** To investigate the effect of ultra-violet (UV) light exposure on fresh DNA samples, buccal swabs were collected and exposed to UV treatments for a time course study, quantitated and used in a SNP assay to determine the effect on the DNA SNP typing method as a measurement for accuracy of color determination.

Samples

Intact wisdom teeth and baby teeth were used for this study. The exact age of each sample varies. The wisdom teeth (two samples) were greater than 10 years post extraction and stored at -20C. The baby teeth (two samples) were approximately 20 years old and stored at room temperature. Buccal swabs from a variety of blue, brown, and hazel eye color donors were collected manually and stored at -20C before processing. All human subject testing was approved by the University of New Haven Institutional Review Board (IRB) protocol.

Sample Preparation for Cementum DNA Extraction

Prior to sample preparation, a hood was cleaned with DNA-ExitusPlus iodine (AppliChem GmbH, Darmstadt, Germany), and 70% isopropyl alcohol. The hood was recleaned between each sample. Each of the samples were photographed before and after cleaning on a clean surface. Each sample was washed with DNA-ExitusPlus iodine and then 70% isopropyl alcohol and air-dried. A sterilized Dremel was used to sand and clean the outside of each tooth. The Dremel tip was cleaned with DNA-ExitusPlus iodine and 70% isopropyl alcohol. After sanding, each tooth was cleaned using DNA-ExitusPlus iodine and 70% isopropyl alcohol again. Each tooth was then placed in its own sterile 50 mL conical tube. The samples were stored at -20C until extraction.

Sample Preparation for Cementum, Dentin and Pulp DNA Extraction

All work was performed as described in the paragraph above. Then, a SPEX SamplePrep Freezer/Mill (SPEX SamplePrep, Metuchen, NJ) was used with liquid nitrogen as a coolant to grind the teeth into a fine powder. A 1% Tergazyme solution (Alconox Inc, White Plains, NY) and all grinding cylinders, end caps, and impactors were soaked with shaking for 1 hour. After detergent cleansing, a 20% bleach solution was prepared and sprayed on all grinding cylinders, end caps, and impactors. They were then sprayed with 70% ethanol and crosslinked for 30 min with a UV treatment. The vials were assembled, and a tooth added to each cylinder individually. The Freezer/Mill ground each tooth to a powder and the tooth powder was transferred to sterile 1.5 mL tubes and stored at -20°C until extraction.

Sample Treatments for UV Exposure

The buccal swabs were exposed to UV light using a SpectroLinker XL-1500 UV Crosslinker (Spectronics Corp, Melville, NY) for intervals of 0 sec, 1 min, 5 min, 10 min, 15 min, 20 min, and 30 min. The 0 sec exposure is the positive control for each buccal swab donor.

Extraction and Purification

The Purelink Genomic DNA Mini Kit was used for baby teeth extraction. To each tube, 360 uL of PureLink Genomic Digestion Buffer and 40 uL of Proteinase K (20 mg/ml) were added. The tubes were placed on a shaker set to 100 rpm and left for 24 hours. 200 uL of solution was removed from each

tube and added to a sterile 1.5 uL microcentrifuge tube and labelled. The Mammalian Tissue Protocol from the Purelink Genomic DNA Mini Kit was then followed starting at step 5. The QIAamp DNA Investigator kit (QIAGEN, Hilden, Germany) was used to extract DNA from the wisdom teeth following the manufacturer's protocol. The QIAamp DNA Investigator kit (QIAGEN, Hilden, Germany) was used to extract DNA from the buccal swabs for the UV treatment study using the manufacturer's recommended protocol for buccal swabs. DNA extraction was conducted on the Table 2 buccal swabs using the Purelink® 96 Genomic DNA Kit (Thermo Fisher Scientific, Waltham, MA) by utilizing the Purelink® Genomic DNA Kit User Guide for Human Buccal Swabs protocol.

Quantification and PCR Amplification

The Quantifiler Trio kit (Thermo Fisher Scientific, Waltham, MA) and a QuantStudio 5 Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA) were used to quantify DNA in the tooth samples and to calculate the degradation index. The quantity estimate for the buccal swabs was made using a NanoDrop UV-Vis spectrophotometer. SNP genotyping TaqMan assay primer sets (Thermo Fisher Scientific, Waltham, MA) were used for the two main IrisPlex

SNPs [rs12913832 (HERC2) and rs1800407 (OCA2) with TaqMan genotyping master mix (Thermo Fisher Scientific, Waltham, MA). The SNP alleles were determined using QuantStudio Design and Analysis v1.5.1 Software. The SNP allele calls were manually coded into the IrisPlex prediction model to generate eye color predictions. The prediction and AUC output was compared to donor eye color images to assess accuracy.

Results and Discussion

IrisPlex includes six SNPs that predict eye color with the prediction outcomes being blue, brown, or intermediate [4]. The development and validation of IrisPlex showed that full profiles could be generated with as little as 31 pg of DNA, the ability to be used with degraded samples, and compliance with SWGDAM guidelines [4,9,12]. The prediction accuracy is greater than 90% for blue and brown eye colors but is significantly lower for intermediate eye colors at about 73%. The intermediate eye color category includes all non-blue and non-brown eye colors. The lower prediction accuracy for this category can likely be explained by the fact that eye color is a complex trait, and the genetic basis is still not yet fully understood or simple to classify.

Sample	Blue Eye P-Value	Intermed P-Value	Brown Eye P-Value	Eye-Color Prediction	Degradation Index
T1.1[a]	0.018898	0.04039	0.940712	Brown	1.73
T1.2	0.018898	0.04039	0.940712	Brown	1.34
T2.1[b]	0.919735	0.061282	0.018983	Blue	9.97
T2.2	0.919735	0.061282	0.018983	Blue	6.54
B2[c]	0.823761	0.105558	0.070681	Blue	3.83
B7	0.7156	0.152631	0.131769	Blue	1.73
PC1[d]	0.126337	0.248366	0.625297	Intermediate	1

[a] T1.1 and T1.2 are replicate samples of wisdom tooth 1 with original eye color unknown.
[b] T2.1 and T2.2 are replicate samples of wisdom tooth 2 with original eye color unknown.
[c] B2 and B7 are baby tooth samples with original eye color unknown.
[d] PC1 is a positive control DNA from a buccal swab from a hazel brown-green eye individual.
Table 1: IrisPlex Results for Baby and Wisdom Tooth DNA Extractions.

In our study, the two main IrisPlex SNPs performed well on DNA recovered from aged teeth using two different extraction methods: one for cementum (enzymatic digestion method) and one for cementum, dentin plus pulp (traditional mechanical grinding method) (Table 1). This is valuable as both are common sources of DNA in human skeletal remains and this method yields pigmentation information for forensic phenotyping to aid in the generation of a composite facial

profile to identify a missing person. In this study we noted some variability in the degradation index (DI) per aged tooth sample and performed a purposeful laboratory experiment to assess the effect of UV deterioration of DNA samples to correlate to the DI values and determine the potential effect on the SNP performance. The data from Table 3 illustrates the level of variability noted for the UV treatments and any effects on SNP phenotype determination.

Uv Exposure	Donor	Quantity (ng)	Degradation Index (Di)	Average Di Value	Eye-Color Prediction
0 Sec	1	19.02005	0.776469	0.792536	Blue
	1	25.30495	0.808602		
	3	6.676608	0.894237	0.917384	Brown
	3	19.76738	0.94053		
1 Min	1	22.53483	1.316243	1.233223	Blue
	1	40.3483	1.150202		Brown [A]
	3	8.921627	2.03511	1.813373	Brown
	3	12.82466	1.591635		
5 Min	1	9.752433	2.199978	2.458807	Blue
	1	6.442257	2.717636		
	3	4.400023	2.512962	2.933789	Brown
	3	3.611472	3.354616		
10 Min	1	33.16529	1.378784	1.676228	Blue
	1	12.07662	1.973671		
	3	1.12562	14.43388	9.411898	Brown
	3	2.047285	4.389915		

[A]Donor 1 SNP mistyped as brown for one sample replicate possibly due to damage to the SNP site or due to contamination.
**Table 3:** DNA Quantity and Degradation Index of the UV Exposure Samples.

The HERC2 and OCA2 analysis of donors with "blue" eyes identified and distinguished between individuals that had blue irises only and those with a blue iris but also a brown ring of pigmentation around the pupil (Table 2). Those individuals with blue irises and brown pigmentation around the pupil were correctly classified as intermediate and appear to have a form of heterochromia called central heterochromia. Central heterochromia is defined as the inner ring of the iris having a different color than the outer ring. Complete heterochromia is when one iris is different than the other iris. Partial heterochromia occurs when a segment of the iris is a different color. Our research shows individuals with blue eye central heterochromia are correctly classified by IrisPlex as having brown pigment in the iris (genetic intermediate type) but likely would be visually described as having the blue eye color trait phenotypically by visual examination (Figure 1).

This is the challenging aspect of phenotype prediction based on genetic models and the genetic control of this iris pattern needs further elucidation. Other published studies agree that heterochromia or eye pigmentation at the edge of each color category for classification are one of the contributing factors for loss of accuracy, only 73%, in biological prediction of the eye color. It is important to understand how Irisplex classifies the eye colors along with the predictive accuracy as it may ultimately affect the ability to correctly assign the eye color in a facial reconstruction. Blue and brown eye colors are assigned with high accuracy (90%) but an intermediate eye color designation could be many possibilities such as grey, violet, green, hazel and those exhibiting heterochromia so multiple images may need to be generated to illustrate the possible feature in facial reconstruction efforts to aid in the identification of the missing person case.

Sample	Blue Eye P-Value	Intermed P-Value	Brown Eye P-Value	Eye-Color Prediction
D3	0.911522	0.057036	0.031442	Blue
D6	0.911522	0.057036	0.031442	Blue
D7	0.911522	0.057036	0.031442	Blue
D9	0.911522	0.057036	0.031442	Blue
D13	0.152655	0.161644	0.6857	Intermediate (Blue With Brown Ring)
D17	0.152655	0.161644	0.6857	Intermediate (Blue With Brown Ring)
PC1[a]	0.152655	0.161644	0.6857	Intermediate (Hazel-Brown)

PC1 is a positive control DNA from a buccal swab from a hazel-brown eye individual.
Table 2: IrisPlex Results for Blue Eye Donors.

Figure 1: intermediate eye color.
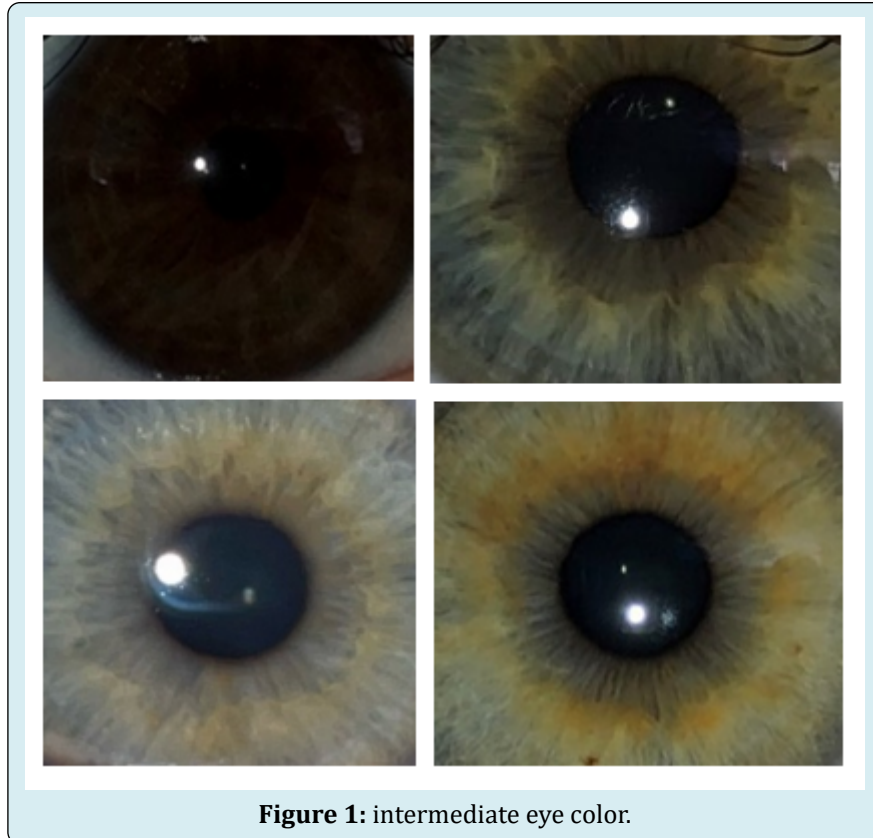
Conclusion

Forensic DNA phenotyping is still a new area of applied biology where datamining of the human genome has identified functional differences between individuals based on genetic variation. Many of these point mutations (SNPs) are related to physical traits (e.g., eye color) or diseases states (e.g., sickle cell anemia) and how they affect the production of proteins is being carefully studied. Innovations in human genome sequencing have led to the identification of important phenotyping SNPs and now forensic science has a tool that can be used for eye color prediction along with many other facial reconstruction pigmentation SNPs for hair color, skin color and freckles. The presence of brown and blue pigmentation can be predicted more accurately with the two SNPs we used in IrisPlex, however, the intermediate eye color category has additional developmental patterns that still need identification markers to be included in the IrisPlex prediction model. The identification of more collagen and lipochrome SNP markers related to perception and scoring of eye color would be a welcome addition to the SNP panel. This would help to increase the resolving power of the intermediate category into the colors of gray and green. The use of forensic DNA phenotyping methods is primarily still in the private sector but for our purposes in academia, we found the DNA methods and analysis to be simple to perform and analyze with the Irisplex statistical model as an excellent teaching and research tool for human remains DNA phenotyping for the eye color trait.

Conflicts of Interest

The authors have declared no conflict of interest.

Acknowledgements

Funding

References

1. Butler JM, Willis S (2020) Interpol review of forensic biology and forensic DNA typing 2016-2019. Forensic Sci Int Synerg 2: 352-367.

2. Budowle B, Van Daal A (2008) Forensically relevant SNP classes. Biotechniques 44(5): 603-610.

3. Schneider PM, Prainsack B, Kayser M (2019) The Use of Forensic DNA Phenotyping in Predicting Appearance and Biogeographic Ancestry. Dtsch Arztebl Int 116(51-52): 873-880.

4. Petrequin J (1843) Sur les diverses couleurs de l'iris at leurs proportions dans nos climats. Ann d'oculistique 10: 120-125.

5. Coon CS (1939) The Races of Europe. New York, The McMillan Company.

6. Diaz MY, Saornil MA, Almaraz A, Munoz Moreno MF, García C, et al. (2004) Iris color: validation of a new classification and distribution in a Spanish population-based sample. Eur J Ophthalmol 19(4): 686-689.

7. Grigore M, Avram A (2015) Iris colour classification scales-then and now. Rom J Ophthalmol 59(1): 29-33.

8. Seddon JM, Sahagian CR, Glynn RJ, Sperduto RD, Gragoudas ES (1990) Evaluation of an irisi color classification system. The Eye Disorders Case-Control Study Group. Invest Ophthalmol Vis Sci 31(8): 1592-1598.

9. Walsh S, Lindenbergh A, Zuniga SB, Sijen T, de Knijff P, et al. (2011) Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int Genet 5(5): 464-71.

10. Kayser M (2015) Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genet 18: 33-48.

11. Liu F, Van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, et al. (2009) Eye color and the prediction of complex phenotypes from genotypes. Curr Biol 19(5): R192-R193.

12. Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, et al. (2011) IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 5(3): 170-180.

13. White D, Rabago Smith M (2011) Genotype–phenotype associations and human eye color. J Hum Genet 56(1): 5-7.

14. (2023) Tyndall effect Accessed 11/07/2023. Website: (https://en.m.wikipedia.org/wiki/Tyndall_effect).

15. (2023) Rayleigh scattering Accessed 11/07/2023. Website: (https://en.m.wikipedia.org/wiki/Rayleigh_scattering).

Coyle HM, et al. Functional Genomics and Eye Color Prediction. Int J Forens Sci 2023, 8(4): 000334.

Copyright© Coyle HM, et al.