Review

Colorful DNA polymorphisms in humans

Fan Liu a,*, Bei Wen b, Manfred Kayser a,*

a Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands
b Department of Science System Assessment, Rathenau Institute, Den Haag, The Netherlands

A R T I C L E   I N F O

Article history:
Available online 12 April 2013

Keywords:
Human Pigmentation DNA polymorphisms Association Prediction Forensic DNA phenotyping

A B S T R A C T

In this review article we summarize current knowledge on how variation on the DNA level influences human pigmentation including color variation of iris, hair, and skin. We review recent progress in the field of human pigmentation genetics by focusing on the genes and DNA polymorphisms discovered to be involved in determining human pigmentation traits, their association with diseases particularly skin cancers, and their power to predict human eye, hair, and skin colors with potential utilization in forensic investigations.

© 2013 Elsevier Ltd. All rights reserved.

Contents

1. Introduction
   1.1. Setting the scene: human pigmentation genetics – the field and the people .................................................. 563
   1.2. Pigmentation phenotypes: quantification and normalization matters ............................................................... 563
   1.3. Evolution of human pigmentation traits: adaptation and sexual selection ......................................................... 565
   1.4. Skin cancer: color and environment together ....................................................................................................... 565
   1.5. Melanogenesis: various pathways making colors ................................................................................................. 567
2. Human pigmentation genes
   2.1. HERC2 and OCA2 .............................................................................................................................................. 567
   2.2. MC1R and ASIP .................................................................................................................................................. 567
   2.3. TYR, TYRP1, and DCT .................................................................................................................................. 568
   2.4. SLC24A5 ............................................................................................................................................................ 568
   2.5. SLC45A2 ............................................................................................................................................................ 569
   2.6. IRF4 ................................................................................................................................................................... 569
   2.7. Other genes ........................................................................................................................................................ 569
3. Predicting human pigmentation traits from DNA
   3.1. Eye color ............................................................................................................................................................ 569
   3.2. Hair color .......................................................................................................................................................... 570
   3.3. Skin color .......................................................................................................................................................... 571
4. Conclusions ............................................................................................................................................................ 572
Appendix A. Supplementary data .......................................................................................................................... 572
References ............................................................................................................................................................... 572
1. Introduction

1.1. Setting the scene: human pigmentation genetics – the field and the people

Human pigmentation traits, including color variation in skin, eye, and hair, belong to the most visible and differentiating human traits. The genetic basis underlying variation in human pigmentation traits has been the subject of intensive research by investigators in a variety of life science communities. We start out this review article by providing an overview of these scientific communities and acknowledge the key players by examining their co-keyword and co-authorship networks (Fig. 1). Via a keyword searching approach using the Web of Science and the PubMed databases (deadline Nov. 1, 2012, for method details see supplemental materials), we identified 1638 papers in peer-reviewed scientific journals published by 6439 authors in the field of human pigmentation genetics, which contained a total number of 2838 unique author-assigned keywords. A network graph of co-occurrence of these keywords (Fig. 1A) highlights at least 10 major research topics representing about 60% of the field, which, after inspection, are interpreted as: Tyrosinase and melanogenesis (14.8%), skin cancer and melanoma (13.4%), SNP association and prediction (5.1%), population genetics and diversity (4.7%), human evolution and mutation (4.0%), hypopigmentation and skin aging (3.7%), melanosome and related disorders (3.7%), cell biology and gene function (3.7%), animal models (3.4%), and face detection algorithms (3.1%). A co-authorship network graph (Fig. 1B) illustrates the collaboration patterns between 10 top clusters consisting of 1903 authors representing 29.6% of the field. Some influential authors (only one per cluster is listed here due to space limitation) are listed in the following in alphabetical order together with their respective fields of research: BARSH GS (mammalian pigmentation [1,2]), FISHER DE (melanoma [3]), HEARING VJ (melanogenesis [4]), STURM RA (human pigmentation [5-7]), KAYSER M (prediction [8]), PAUS R (hair follicle [9]), REES JL (evolution [10]), SHRIVER MD (evolution [11,12]), SOUFIR N (melanoma [13]), SPRITZ RA (vitiligo [14]). These figures demonstrate the variety of research areas involved in human pigmentation genetics some of which we discuss below. On the other hand, it can be concluded from these figures that more cross-boundary collaborations may be recommended for researchers in different co-authorship clusters with overlapping scientific interests.

1.2. Pigmentation phenotypes: quantification and normalization matters

Human pigmentation is primarily explained by the presence of melanin in epidermis, iris and hair. Phenotypic variation in pigmentation traits is determined by the amount, type, and distribution of melanin within particular cells, the melanocytes, while the number of melanocytes is usually not varying. Pigmentation, particularly skin color variation, belongs to the limited human traits where the phenotypic, and consequently also the underlying genotypic variation, is much larger between than within populations. In contrast, for many human traits the phenotypic and underlying genotypic variance is larger within than between populations. In genetic studies, pigmentation phenotypes have often been measured categorically based on broad consensus, such as blue and brown categories for eye color [15]. Often such categories are accessed via questionnaires either directly by asking which of the given eye color categories a person believes to have, or more indirectly by asking about tanning ability [16] and tendency to burn [17] as measurements related to skin color. Such questionnaire-based phenotypes, however, are likely to be error prone due to individual differences in perceiving colors and classifying them into categories, and the more so the more similar the categories are (such as blue vs. gray eye color, or light blond versus dark blond versus light brown hair color). This can partly be avoided by applying single observer phenotype grading where individual differences in phenotyping are canceled out within a study (but still causes problems in multicenter studies involving different graders in different studies). Questionnaire-based phenotypes are also error-prone for pigmentation traits that can change with age such as hair color. In such cases, and depending on the age of the volunteer at time of interview, the phenotype knowledge needs to rely on the volunteers’ memory, which also applies to artificially changing pigmentation phenotypes such as via hair dyeing.

Furthermore, in reality, all human pigmentation traits are of true quantitative nature. In general, the use of categorized information inevitably oversimplifies continuous traits. It therefore has been emphasized to utilize the full quantitative spectrum of coloration in genetic studies of human pigmentation traits [18,19]. Currently, there are several strategies to study the genetics of human pigmentation in a quantitative way based on reflectance, bioimaging, or biochemical technologies. Commonly used technologies include tristimulus colorimetry, specialized narrow-band reflectometry, diffuse reflectance spectroscopy, digital photographs, and high-performance liquid chromatography (for a more systematic review on reflectometry and chromatography see [20]). Quantifying pigmentation phenotypes from digital photographs was proposed recently [21], and so far has been successfully applied to genetic studies of eye [18] and skin colors [19]. This method has obvious advantages in terms of cost, time efficiency, and portability, but – if not considered adequately in the design of the image collection – can also have disadvantages regarding noise variance being introduced by various source of environmental lights [22]. For example, the intensity of camera flashes may not be the same even if they were produced by the same camera, and the distance between flash and volunteer can influence the lightening condition and thus the color information collected. Controlling for such environmental effects during image collection and selecting appropriate color models is thus important to reduce the noise variance. The most commonly used color model is the Commission International d’Eclairage (CIE) L*a*b* system established in 1976 [23], which has been typically employed in genetic studies of pigmentation information collected via spectrophotometers [24-26]. In the CIELab color model, any color can be represented by three variables: L*, the lightness-darkness axis; a*, the red-green axis; and b*, the blue-yellow axis, which can be plotted in three-dimensional space [27]. Another color model is the HSB (or HSV) model, which consists of hue, saturation, and brightness dimensions. In both the HSB and CIELab color spaces, the brightness in HSB or luminance in CIELab is influenced by the reflectance of light, and considered as a separate dimension. The remaining color dimensions are invariant to lightness, which may still provide useful information when the environmental lights cannot be fully controlled. In such cases, the three color dimensions (HSB or Lab) may be reduced into two-dimensions (HS or ab) by removing the brightness dimension. For instance, in a previous genome-wide association study (GWAS) on quantitative eye color [18], as well as a recent candidate gene study of skin color [19], the HS color space was chosen for color quantification since the hue and saturation are intuitively more biologically meaningful than the a* and b* in CIELab, and highlighted already known but also new genetic variants underlying quantitative eye and skin color, respectively.

Obviously, the more accurate and the more detailed pigmentation phenotypes are collected the better the chance to find the underlying genetic variation. This is especially true in the current days of human pigmentation genetics where the “low hanging fruits” regarding human pigmentation genes with strong effects have already been harvested (see below). Future studies further
Fig. 1. Network graphs of co-keyword and co-authorship in human pigmentation genetics. (A) Co-occurrence network graph of author-assigned keywords of publications in the field of human pigmentation genetics. Each node represents an author-assigned keyword with size weighted by its frequency. In total 536 keywords with frequency greater than 0.001 are shown. The top 10 modules are shown in colors (see legends). (B) Co-authorship network graph of authors in the field of human pigmentation genetics. Each node represents an author with his/her size weighted by the number of publications. The presence of an edge between two nodes indicates the two authors had collaborations in one or more of their published works. The top 10 communities consisting of 1903 out of total 6500 authors are shown in different colors. The high resolution version of this graph with all author names superimposed on the graph can be found in the supplementary materials.
exploring the genetic basis of human pigmentation phenotypes, to unveil the currently missing heritability of human eye, hair, and especially skin color, should therefore concentrate on high accuracy and high level of detail in color phenotype description instead of simply increasing the number of samples with phenotype information on roughly measured color categories.

1.3. Evolution of human pigmentation traits: adaptation and sexual selection

Natural variation in human skin color is assumed to be an adaptive trait shaped by recent evolutionary history. The strong correlation observed between the skin color of people and their worldwide place of origin regarding latitudes (but not longitudes) (Fig. 2A) strongly suggests environmentally induced genetic adaptation to play a key role in the evolutionary history of human skin color variation. One of the environmental factors that strongly correlate with latitude is the intensity and duration of solar radiation including its ultraviolet (UV) component (Fig. 2B). UV radiation is stronger in equatorial regions of the world, where dark skin color is usually observed, and less strong in regions distant to the equator, where light skin color is usually found.

The most currently popular evolutionary hypothesis for human skin color variation is the vitamin D/folate hypothesis [28–31]. According to this hypothesis, skin pigmentation would be on the one hand determined by the amount of UV radiation the human body receives via its skin and on the other hand by the vitamin D and folate syntheses. UV is needed for the synthesis of vitamin D in the skin from pro-vitamin D. Vitamin D deficiency results in rickets and increases the risk of miscarriage [32]. However, high amounts of UV radiation tend to degrade folate and thus increase the risk of neonatal malformations including spina bifida [33]. Dark skin pigmentation protects better against the harmful effects of solar radiation than lightly pigmented skin [34]. Consequently, light skin pigmentation would have a biological advantage over UV-hypoprotected dark skin pigmentation for people who live in regions distant to the equator where levels of UV are smaller, while dark skin would have a biological advantage over UV hypoprotected light skin pigmentation in people from equatorial regions where levels of UV are higher [30,31].

Given this, the Out-of-Africa migration history mostly assumed for anatomically modern humans is likely to have provided the basis for human genetic adaptation toward solar radiation via skin lightening in Europeans and Asians from the ancestral dark skin in ancestral Africans. Although a number of alternative views exist regarding some details of this hypothesis [5,35,36], there seems to be general agreement that human skin color variation arose as result of natural selection, most likely via genetic adaptation to solar radiation. Several studies could show evidence of selective sweeps in those genomic regions where known skin color genes are located [37,38], for an overview see [20], providing strong support for the idea of natural selection on human skin color. There also is emerging genetic evidence that skin lightening in Europeans and Asian happened independently from each other [12,37] as would be expected from different routes of human migration into Europe and Asia. While the selective sweeps are typically explained by beneficial mutations with increased fitness to the environment, they can also be caused by positive selection via mate choice or sexual selection. It was already proposed by Charles Darwin in 1871 [39] that sexual preference could have been important in shaping the distribution of human pigmentation. Anecdotal knowledge tells that light-skinned persons are preferred marriage partners in regions were dark skin is very frequent, such as in India. However, this hypothesis, in contrast to the leading vitamin D/folate hypothesis, alone is insufficient to explain the global pattern of skin color correlation with latitudes.

In contrast to skin color where a fitness advantage depending on solar radiation can be assumed, it is more difficult to explain eye and hair color variation via genetic adaptation toward environmental factors. Although dark hair is expected to protect the head skin more from solar radiation than light hair does, this effect can be considered as relatively minor for the human organism given the amount of head skin relative to the full body skin. Furthermore, the back side of the iris is black pigmented in irises of all colors on the front side, which provides the protective effect independent of eye color variation. Because of the absence of a plausible fitness advantage for eye and hair color variation, and because of the regional distribution of eye and hair color variation in Europe and surrounding areas but not on a worldwide latitude scale, the hypothesis of sexual selection and mate choice preference is more attractive for explaining eye and hair color differences among humans. Anecdotal knowledge on eye and hair color preferences in human mate choice exists in many societies with a preference to the non-ancestral non-brown such as blue eye and the non-black such as blond hair colors, which usually are rarer than the ancestral brown eyes and black hair phenotypes in these groups. Experimental data on color-based mate choice preference exists for fish [40,41], where evidence for mate choice preference based on rare color types was established, fitting with anecdotal knowledge in humans. Given the absence of eye and hair color variation outside of Europeans and their direct geographic neighbors such as people from the Middle East and certain parts of Western Asia, it appears likely that eye and hair color variation arose in Europe via genetic mutations, and both the mutations and the derived non-brown eye color and the non-black hair color phenotypes increased in frequency due to sexual selection via eye and hair color preferences in human mate choice in Europeans. There currently is no direct data evidence on when about human eye, hair, and skin color variation first arose out of the ancestral pigmentation types being brown eye, black hair and dark skin color. A recent genetic study aiming to shed light on the evolutionary history of human skin color lightening in a more indirect way suggests an origin of European light skin about 11,000–19,000 years ago based on computer simulations of data obtained from candidate pigmentation genes in small samples of Europeans, Africans and East Asians [42]. Future studies focusing on ancient DNA found in pre-history human remains may provide more direct evidence about the timing of the skin-lightening event and narrow down the estimated time interval.

1.4. Skin cancer: color and environment together

Human skin cancer is one of the most common malignancies worldwide, with a pronounced incidence in Australians of European (mostly British) ancestry due to the non-evolutionary combination of light skin and strong solar radiation [43]. Currently known risk factors of skin cancers and melanoma include skin color, sun-exposure, geographic location, and genetic factors [44,45]. The association between genetic variants and skin color variation had been described in earlier investigations [46,47] and confirmed via GWAS [16,17,48]. Many of the human pigmentation traits are correlated with each other, and those that are correlated with light skin such as red and blond hair, and green and blue eyes, including their underlying genetic polymorphisms, are also correlated with skin cancer risk [49–58]. Cutaneous melanoma, a disease of fair skinned individuals with a rapidly increasing incidence over the past decades [59,60], is the most aggressive and treatment-resistant skin cancer. There are four main types of cutaneous malignant melanoma [61]: superficial spreading malignant melanoma (accounts for about 70% of all melanomas), nodular melanoma (15–30%), lentigo maligna melanoma (4–10%), and acral lentiginous melanoma (2–8%). The risk for melanoma is associated
with sun-exposure including the use of sunbath, family history [62], pigmentation characteristics and moles [63], the number of melanocytic nevi [64,65], and iatrogenic immuno-suppression [66]. Rare high penetration mutations in two genes, CDKN2A [67,68] and CDK4 [69], are known to co-segregate with melanoma in families. Polymorphisms in several pigmentation genes were found to confer low-to-moderate risk for melanoma in general populations, including MC1R [70], ASIP [71], TYR [55], SLC45A2/MTAP [55], IRF4 [72], and HERC2/OC2 [73]. For an extensive review on genetic loci that confer melanoma risk, including those involved in pigmentation, see Tsao et al. [3]. DNA polymorphisms in non-pigmentation genes have also been associated with melanoma risk or the development of cutaneous nevi, including VDR [74], MYH7B, PIGU, TP53INP2, NCOA6, GGT13, ACSS2, GSS, E2F1 [71], ATM, MX2, CASP8 [75], ARNT, SETDB1 [75,76], NID1 [77], PLA2G6 [78], CDK10, and FANCA [79]. Many of these non-pigmentation genes are physically close to the pigmentation genes, questioning the exact identities of the causal genes. For example, the 16q24 locus contains MC1R, CDK10 and FANCA; CDKN2A is adjacent to SLC45A2 at 9p21; MYH7B, PIGU, TP53INP2, NCOA6, GGT13, ACSS2, GSS, E2F1 are all close to ASIP at 20q11.22. Further, it is notoriously difficult to separate confounding effects from causal effects in various studies including genetic epidemiological ones. Hence, it is difficult to know if DNA variants determining light skin pigmentation are providing an increased risk for melanoma (and other skin cancers with pigmentation dependence) because light skin is more sensitive for UV radiation, or if they instead or in addition provide a direct cancer effect. A recent paper by Mitra et al. [80] provides a first answer by showing that in mice the pheomelanin pigment pathway produces ultraviolet-radiation-independent carcinogenic contributions to melanomagenesis by a mechanism of oxidative damage. This study concludes that the effects of pigmentation and ultraviolet radiation are likely to work together in determining melanoma risk.
1.5. Melanogenesis: various pathways making color

As this review is on human pigmentation genetics while most of the melanogenesis is rather similar across mammalian species including humans, we will not describe melanogenesis in detail here [for an extensive review on cellular mechanisms regulating human melanogenesis, see, e.g., Park et al. [81]], but summarize it briefly to allow putting the human genetics details explained below in perspective. The most common form of melanin is eumelanin, a brown-black polymer of dihydroxyindole carboxylic acids, and their reduced forms, derived from the amino acid Tyrosine. Another form of melanin is pheomelanin, a cysteine-containing red-brown polymer of benzothiazine units largely responsible for red hair and freckles. Both types of melanin are synthesized through a complex chain of chemical reactions, catalyzed by different enzymes involved in the pigmentation machinery and determine human pigmentation as well as that of other mammals. Melanogenesis, or the procedures involved in the synthesis of melanin, is under complex genetically controlled (see below) regulation. The most important positive regulator is the melanocortin 1 receptor (MC1R) with its ligands melanocortic peptides. MC1R activates the cyclic AMP (cAMP) response-element binding protein (CREB). Increased expression of MITF and its activation by phosphorylation (P) stimulate the transcription of Tyrosinase (TYR), Tyrosinase-related protein 1 (TYRP1), and dopachrome tautomerase (DCT), which produce melanin. Melanin synthesis takes place within melanosomes using the amino acid Tyrosine as the major substrate. TYR is the key enzyme that mediates the first steps in melanin synthesis, which involve the hydroxylation of Tyrosine to dopaquinone. Dopaquinone is a key intermediate compound that undergoes further modification in two alternative pathways which ultimately lead to either brown-black eumelanin or red-yellow pheomelanin.

2. Human pigmentation genes

Although human eye, hair, and skin colors are polygenic in nature, they represent rather untypical complex phenotypes compared with many other polygenic human traits, such as body height, where hundreds of genes with small effects are involved [82]. Most pigmentation traits follow a semi-Mendelian type of inheritance, suggesting a major gene effect. SNPs in TYR, TYRP1, OCA2, SLC45A2, SLC24A5, MC1R, ASIP, KITLG, HERC2, SLC24A4, IRF4, TPCN2, LYST, and BNC2 genes have been associated with human pigmentation. TYR, SLC45A2 and SLC24A5 largely account for skin color variation between continents. A cysteine-to-arginine change in TYRP1 is the major determinant of blond hair in Melanesians [83]. Within Europeans, blue-brown eye color is to a large degree explained by a single SNP rs12913832 in HERC2 that has recently been shown to regulate OCA2 expression [84], and also has a large effect on hair and skin colors. Multiple loss-of-function variants in MC1R cause red hair and fair skin. Other genes have more subtle and various effects on different pigmentation traits, but all impact melanosome biogenesis or the melanin biosynthetic pathway, which, taken together explain the color diversity seen within and between human populations.

2.1. HERC2 and OCA2

Two neighboring genes on chromosome 15q11.2-13, HERC2 (HECT and RLD domain containing 3 ubiquitin protein ligase 2; 15q13) and OCA2 (oculocutaneous albinism II; 15q11.2), are of greatest importance in determining eye and hair, and also skin colors in Europeans. OCA2, also known as the P gene, regulates melanosomal pH. Certain mutations in OCA2 cause oculocutaneous albinism type II [85]. The discovery of OCA2 in relation to human eye and skin colors has been systematically reviewed previously [20]. Before 2008, OCA2 was believed to be the dominating gene for eye color determination [17,86,87]. In 2008 and subsequently, several independent studies demonstrated via genome-wide and regional genetic approaches that DNA variants in the neighboring HERC2 gene are associated more significantly with eye color than SNPs in OCA2 or elsewhere in the genome [15,18,49,88]. The most significantly associated SNP rs12913832 was found in intron 86 of the HERC2 gene, 21 kb upstream of the promoter of OCA2. The region surrounding rs12913832 is highly conserved among animal species [49], and this SNP is strongly associated with OCA2 expression levels [89]. Although it has been hypothesized that this SNP may act in a distal regulatory region that either silences or enhances OCA2 expression [49,88,90], until very recently it was unclear if HERC2 should be acknowledged as a pigmentation gene due to a lack of an obvious functional link. In 2012, HERC2 rs12913832 was experimentally shown to function as an enhancer regulating OCA2 transcription by distantly modulating chromatin folding [84]. This study showed via several molecular approaches that the HERC2 rs12913832 enhancer communicates with the OCA2 promoter via a long-range chromatin loop, and that this enhancer activity is mediated by the transcription factors HLF, LEF1, and MITF. The HERC2 rs12913832 T-allele robustly recruits these transcription factors, which, in combination with increased looping to the OCA2 promoter, facilitates OCA2 expression leading to enhanced melanin production and ultimately dark pigmentation. In contrast, chromatin-loop formation, transcription factor recruitment, and OCA2 expression are all reduced in lightly pigmented melanocytes carrying the rs12913832 C-allele. Consequently, HERC2 rs12913832 is now known as the most influential predictor not only for eye [91–93] but also for hair [94] and skin [19,92] colors in Europeans (see below).

The OCA2 gene, independent of HERC2, plays a role in normal variation in eye, hair, and skin colors. Three SNPs within intron 1 of OCA2 showed a very strong association with eye color [86]. When taking HERC2 SNPs into account in a multivariate analysis, OCA2 SNPs displayed a much weaker association but remained statistically significant, pointing to an independent but smaller effect of OCA2 SNPs on eye color relative to HERC2 SNPs [15]. Thus, HERC2 SNPs alone cannot fully explain the association signal observed at 15q11.2-13. Interestingly, when adjusted for HERC2 rs12913832, a missense OCA2 SNP rs1800407 showed a stronger association with eye color, whereas the effect of other OCA2 SNPs was reduced [91]. Statistical interactions between HERC2 rs12913832 and OCA2 rs1800407 have also been reported to affect human pigmentation [18,95,96], which is consistent with the chromatin folding interaction known today [84].

2.2. MC1R and ASIP

MC1R (melanocortin 1 receptor; 16q24.3) has been a subject of intensive research already for decades [56,97–99], ever since its variants were first associated with human red hair and fair skin in the mid 1990ies [97]. Variants in MC1R have also been significantly linked to freckles [100], responses to UV radiation [101,102], and skin cancer risk [103], particularly melanoma risk [104]. Interestingly, unlike some other well-known pigmentation genes, MC1R variants have not been found to be associated with eye color. MC1R encodes a G protein-coupled receptor, representing a switch in melanogenesis. Binding of the melanocyte-stimulating hormone (a-MSH) leads to eumelanin production, whereas binding of its antagonist, agouti signaling protein (ASIP), switches to pheomelanin production. MC1R is a very small gene of less than 1 kb in size, which lacks any introns, thus being attractive for direct DNA sequencing to unveil pigmentation associated or causing DNA variants. Non-synonymous variants in
MC1R were found to be nearly absent in Africans but highly enriched in Europeans and Asians [105–108]. Nine DNA variants, V60L (rs1805008), D84E (rs1805006), R151C (rs1805007), 115ST (rs1110440), R160W (rs1805008), R163Q (rs885479), R142H (rs11547464), D294H (rs1805009), and V92M (rs2228479), are common in Europeans, and two DNA variants, A163G and V92M, are common in Asians [106,109]. These, together with more than 90 rare non-synonymous, insertion/deletion frameshift mutations and premature stop codon mutations [46,97], functionally influence the protein expression of MC1R. The common alleles D84E, R151C, R160W and D294H, having the strongest association with red hair/fair skin, were classified as high penetrance alleles [110]. Although with different penetrance [110] and effect size [111], all these causal DNA variants lead to partial or complete loss of function (LOF) of MC1R [112–114], explaining a largely recessive inheritance of red hair in Europeans. Family members with red hair are more often found to carry multiple compound heterozygote (CH) than single homozygote LOF alleles. Studying heterozygote LOF alleles is of methodological interest since it is more challenging to detect CH-like associations than single SNP associations in GWAS. Penetrance of CH is consistent with a multiplicative allelic model [115]. Because the causal alleles with extraordinarily large effects (such as rs1805006, rs1805007, rs1805008, rs1805009) that are not represented on SNP microarrays are tagged by strongly linked array SNPs, significant association between tagging SNPs close to MC1R and red hair could be observed via GWAS [17]. In more common situations LOF variants may have smaller effects and lower frequencies, and so be undetectable. Recently, a method has been proposed to detect CH-like associations in GWAS, which demonstrated an increased statistical power using the relationship between MC1R variants and red hair as an example [116]. While the LOF alleles in MC1R are important, they themselves alone cannot explain all red hair cases since some dizygotic twins carrying identical MC1R haplotypes were found to have discordant hair colors [117].

ASIP(20q11.2-q12) encodes the agouti signaling protein, a MC1R ligand of 132 amino acids that antagonizes the function of the transmembrane receptor and inhibits eumelanogenesis in human melanocytes [118]. Variants and haplotypes at 20q11.2 have been significantly associated with eye color [119], hair color [48,119], skin color [19,120], skin sensitivity to sun [48], freckling [48], melanoma [53,55,71], and basal cell carcinoma [55] in individuals of European descent. The exact causal identities (in terms of both SNPs and genes) for the phenotypic effects associated with this locus remain elusive given that MYH7B, PIGU, TP53INP2, NCOA6, GGTL3, ACSS2, GSS, EZF1 are all close to ASIP at 20q11.22.

2.3. TYR, TYRP1, and DCT

TYR (11q14) encodes Tyrosinase, a key enzyme that catalyzes the initial step of melanogenesis [121]. TYR1 and DCT (also known as TYRP2) encode Tyrosinase-related proteins that modify intermediates [122,123]. For a systematic review on the role of TYR and TYRP2 in the melanin biosynthetic pathway and the related disorders such as albinism and melanoma, see elsewhere [124]. Variants in TYR were associated with eye [17,18], hair [94], and skin [17,19,125,126] color variation in Europeans. Two coding polymorphisms of TYR, rs1042602 S192Y, and rs1126809 A402G, appear at high frequency in Europeans and are largely absent in African populations [126]. The Tyrosinase enzyme shows a reduced activity for the wild-type of both DNA polymorphisms. Recently, polymorphisms in TYR have been associated with susceptibility loci for generalized vitiligo [127].

TYRP1 (9p23) encodes Tyrosinase-related proteins derived from Tyrosinase but showing distinct key residues in the catalytic domain, and constituting an invention of chordate lineage [128].

In Europeans, TYRP1 rs1408799 and other SNPs have been associated with eye color [18,48,129]. This SNP also showed significant association with blond vs. brown hair in Icelandic and Dutch Europeans [48]. A synergistic interaction between HERC2 and TYRP1 variants was shown to affect green eye color [96], although this effect seems to be rather minor. Recently, a cysteine-to-arginine change A93C in TYRP1 was found in Solomon Islanders from the Western Pacific as a major determinant of rarely found blond hair through a recessive mode of inheritance [83]. In contrast to the blond and light-skinned Europeans who mostly have blue eyes, these blond Solomon Islanders have brown eyes and dark skin, while most people in the Pacific region (except those of European ancestry) have ancestral black hair, brown eyes and dark skin. All 12 blond Solomon islanders studied were homozygous for the T allele, whereas dark-haired individuals were either heterozygous or homozygous for the C allele. The T allele has a high frequency (~26%) in the Solomon Islanders, likely a result of genetic drift. About 5–10% of those Islanders have blond hairs, which is consistent with the expected frequency if the T allele has a recessive effect. On the other hand, the A93C variant deviated significantly from Hardy-Weinberg Equilibrium in the studied Solomon Islanders, with a deficit of heterozygotes, likely a result of sexual selection, i.e. blonds tend to marry blonds and produce offspring. This DNA variant is not polymorphic in other world populations; thus it cannot explain the blond-hair phenotype in Europeans. However, further examining TYRP1 variants and haplotypes may provide new insights to blond hair and intermediate eye color phenotypes also in Europeans.

DCT (13q32) encodes the dopachrome tautomerase with a major role in the isomerization of dopachrome to 5,6-dihydroxyindole-2-carboxylic acid. DCT expression has been proposed to affect age-dependent changes in eumelanin composition in hairs between African, European, and Asian populations [130]. Lao et al. [37] showed a haplotype of DCT locus being highly frequent in Han Chinese (75%) but absent in the European and the African populations, suggestive for positive selection on DCT in East Asians as also confirmed by later studies [131,132]. DCT haplotypes have been associated with eye color [129] in Europeans.

2.4. SLC24A5

SLC24A5 (15q21.1) encodes the Na+/Ca2+/K+ exchanger 5 (NCX5) protein, an intracellular membrane protein with 2 large hydrophilic loops and 2 sets of multiple transmembrane-spanning segments [133]. A mutation in SLC24A5 was originally found to explain the golden color of zebrafish [133]. The counterpart in humans, a non-synonymous SNP rs14266654, is largely fixed with the ancestral A allele in darkly pigmented Africans, while the derived G allele is fixed in Europeans of light skin and varying eye and hair colors [132]. This SNP also showed association with melanin levels in admixed African Americans and African Caribbean, where carriers of the AG or GG genotypes had lighter skin than AA carriers did [133]. On the other hand, due to the allele fixation, this SNP is incapable to show any genetic association in Europeans, as was seen in studies using exclusively European samples [19,134]. A world-wide allele frequency map shows rs14266654 being highly polymorphic in South Asians (frequency A = 0.52) [135]. Indeed, a GWAS of skin reflectance in South Asians demonstrated a highly significant association at 15q21.1 and strongly suggested rs14266654 as the single source of causality [126]. SLC24A5 rs1426654 showed a relatively large allele-frequency difference (39%) between the high and low skin reflectance groups [126]. Although rs1426654 is important for skin color differences between continents, it unlikely influences human eye and hair colors since eye and hair color variation is
nearly exclusively observed in Europeans where this SNP is largely fixed.

2.5. SLC45A2

SLC45A2 (solute carrier family 45 member 2, 5p13.2), also known as MATP (membrane-associated transporter protein), was initially identified through the positional cloning of a medaka gene [136] and a mouse underwhite (uw) gene [137]. In humans, SLC45A2 mutations are known to cause oculocutaneous albinism type 4 (OCA4) [137–139]. Common SNPs, including two non-synonymous SNPs in SLC45A2, rs16891982 (L374F, c.1122G>C) and rs26722 (E272K, c.814G>A), have been associated with normal variation in human skin, hair, and eye color in Europeans, South Asians, as well as in admixed individuals [126,140–143]. The association of rs26722 may be explained by linkage disequilibrium with rs16891982 [142]. The derived 374F allele is almost fixed in people of European descent, highly frequent in Uyghurs from Urumqi (west China), and quite rare in the other populations examined [141,144].

2.6. IRF4

IRF4 (6p25-p23) encodes the interferon regulatory factor 4, detected in malignant melanoma and normal melanocytes [145], is lymphocyte specific in expression and negatively regulates Toll-like-receptor signaling that is central to the activation of innate and adaptive immune systems [146]. An intergenic SNP close to IRF4 was initially found to be associated with freckling in a GWAS on human pigmentation [17]. Further studies found an intron SNP rs12203902 of IRF4 being strongly associated with hair, eye, skin colors, skin tanning response to sunlight, and adolescents neutus count [16,18,19,72,147,148]. Interestingly, the T allele showed a dominant effect on lower skin color saturation [19]. The SNP is only polymorphic in Europeans and shows a north-south gradient allele frequency distribution across Europe [149,150].

2.7. Other genes

Other pigmentation genes including KITLG, TPCN2, LYST, and BNC2 likely have more subtle effects on normal pigmentation compared to the ones listed above. KITLG (12q22) encodes the c-Kit ligand that influences melanocyte proliferation and activates keratinocytes to produce melanogenic factors [151]. KITLG mutations have been found to cause familial progressive hyper- and hypopigmentation [152,153]. Variants at the KITLG locus have been associated with light hair colors in Europeans [17,147,154]. The associated SNP rs12821256 at KITLG was shown to have a small but independent contribution to categorical hair color prediction in Polish Europeans [94]. A recent simulation study showed that the selective sweep shared by Europeans and East Asians at KITLG occurred much earlier than that for the European-specific alleles at TYRP1, SLC24A5, and SLC45A2 [42]. Two coding variants, rs35264875 and rs3829241, of TPCN2 (encoding two-pore segment channel 2: 11q13.3) were found to be significantly associated with blond versus brown hair color in Icelandic Europeans [48]. LYST (lysosomal trafficking regulator: 1q42.1-q42.2) encodes a large cytosolic protein influencing lysosome-related organelles, including lysosomes, melanosomes, and platelet dense bodies [155]. LYST is known as a pigmentation gene mostly in animal studies such as in mice [156] and cattle [157]. Intronic SNPs of LYST have also been found to have a small but significant effect on quantitatively measured human eye color variation [18]. BNC2 (basonucin 2; 9p22.2) encodes a potential transcriptional regulator in keratinocytes based on its close similarity to basonucin 1 [158]. The pigmentation effect of BNC2 is known from zebrafish [159] and mice [160]. In humans, BNC2 intronic SNPs have been associated with freckles in a multiple trait customer-based genetic association study [147]. Recently, a candidate gene study demonstrated significant association between DNA variants in BNC2 and skin color saturation, which was quantified from high resolution digital photographs [19].

DNA variants in some genes not known to be involved in melanogenesis pathways were also found to be associated with human pigmentation traits, which require further investigations. For instance, rs12896399 in SLC24A4 (solute carrier family 24 number 4, 14q32.12) has been previously associated with eye and hair color, skin sensitivity to sun, and increased risk for cutaneous malignant melanoma [16–18,53,147,161]. A SNP in complete LD with rs12896399 was associated with pigmented rings of the eye [162] suggesting SLC24A4 may influence the iris region where pigmentation is placed, rather than overall homogeneous eye color. However, it has been shown recently that interactions between HERC2 and SLC24A4 affect determination of blue eye color [96]. Variants in UGT1A (UDP glucuronosyltransferase 1 family, polypeptide A complex locus: 2q37) genes have been found to induce yellowness of iris [18] and skin [19] in Dutch Europeans. The UGT1A gene encodes a UDP-glucuronosyltransferase, an enzyme of the glucuronidation pathway that transforms bilirubin into water-soluble metabolites. Variants in this gene influence bilirubin plasma levels [163], and were suggested to cause Gilbert’s syndrome [164], which is the most common syndrome known in humans characterized with mild and harmless jaundice characterized by a yellowish discoloration of the skin.

3. Predicting human pigmentation traits from DNA

Predicting complex human phenotypes from genotypes is the central concept of widely advocated personalized medicine, where phenotype-associated DNA markers can be used for clinical diagnosis, patient stratification, and prognostic or therapeutic categorization [165]. DNA prediction of externally visible characteristics including human pigmentation traits additionally has important forensic implications. Forensic DNA phenotyping (FDP) is a relatively young field within forensic genetics aiming to predict appearance traits of a sample donor from DNA left behind at a crime scene. FDP is expected to be useful during police investigation to trace perpetrators whose conventional DNA profile is not yet known to the investigating authorities and who consequently cannot be identified via conventional DNA profiling. Such DNA intelligence tools are important in cases where the police has no clue who the crime scene sample donor might be and all leads of investigation, including conventional DNA profiling, have not provided useful information to find the unknown perpetrator [8,166]. As with other applications of DNA data in the forensic arena, FDP has to be backed-up by the legislation as for instance currently is the case already for forensic DNA-based eye color prediction in The Netherlands.

3.1. Eye color

DNA-based eye color prediction models have been proposed [91] and further developed mostly for forensic applications [92,93,150,167–171]. In one study, 37 eye color associated SNPs from 8 genes have been tested in 6168 Dutch Europeans for their independent predictive effects on three eye color categories [91]. A combined subset of six most eye color predictive SNPs was found to be suitable to predict blue, intermediate, and brown colors with an overall accuracy of 0.91, 0.72, and 0.93, respectively, as measured by Area Under the receiver operation Curves (AUC), a prevalence-adjusted population-based measure for prediction accuracy (Fig. 3A) [91]. These six SNPs were ranked regarding to their individual contribution to the overall capacity in predicting eye color in the
Contribution values descending

Fig. 3. DNA prediction of categorical eye and hair color on the population level. (A) Contribution of 24 SNPs to the prediction accuracy of human eye color in thousands of Dutch Europeans of the Rotterdam Study [91]. (B) Contribution of 13 single or compound markers to the prediction accuracy of hair color in hundreds of Polish Europeans [94]. In both graphs, prediction performance measured by AUC for the model based on multinomial logistic regression (Y-axis) was plotted against the number of SNPs included in the model (X-axis).

descending order: HERC2 rs12913832, OCA2 rs1800407, SLC24A4 rs12896399, SLC45A2 rs16891982, TYR rs1393350, and IRF4 rs12203592 (Fig. 3A). When using single grader eye color phenotypes instead of questionnaire derived eye colors, thus being more accurate in the phenotypic characterization of eye color, the AUC values derived from these 6 SNPs are even higher with 0.96 for blue and brown eye color, respectively, as estimated from thousands of Europeans of seven populations across Europe [150]. A DNA test system, called IrisPlex, was developed that includes a highly sensitive and robust multiplex genotyping assay for analyzing these 6 SNPs as well as a prediction model for deriving eye color probabilities from the combined genotypes [150]. The IrisPlex test system has been further validated for forensic application regarding both the assay [172] and the prediction model [168]. Recently, a further assessment of eye color SNP predictors by Ruiz et al. [93] suggested that adding additional HERC2 SNPs would increase the prediction accuracy: adding rs1129038 improves eye color prediction in general and adding rs1667394 and rs7183877 improves hazel-green prediction in particular. Ruiz et al. provided surprisingly high accuracy estimates with overall AUC values to 0.999 for brown, 0.990 for blue, and 0.816 for green-hazel prediction as estimated from 416 subjects from six European populations, perhaps by using more accurate eye color classifications because the additional HERC2 SNPs are insufficient to explain the accuracy improvement. In both, as well as other studies the HERC2 rs12913832 alone contributed to the vast majority of blue-brown color prediction [91,93]. This major gene effect mostly contributes to the success in practical applications of human eye color DNA prediction. Additional SNPs do provide independent information on eye color prediction and differently so for different colors, but their individual effects are much smaller. This scenario differentiates eye color from many other classical polygenic traits such as adult body height, where such large differences in predictive value between DNA variants are not seen and where instead all significantly associated DNA variants, usually hundreds and expectedly thousands, contribute similarly little (e.g. a few millimeter in height) to the overall predictability of the trait. Although the eye color prediction models were initially developed in European populations, because eye color is a European trait, they appear to work well regardless of the bio-geographic ancestry of the individual tested [150]. Moreover, brown eye color of non-Europeans can usually be predicted with considerably higher brown eye probabilities than brown eyes of Europeans. This is because in regions distant to Europe usually only ancestral alleles of eye color SNPs are found leading to the ancestral brown eye phenotype. In Europe, however, due to the existence of non-brown eye color and underlying DNA variants in the European population, brown eyed people often have derived alleles at some minor DNA predictors lowering the brown eye prediction probability. Nevertheless, the vast majority of brown eyed Europeans are predicted correctly in their brown eyes using IrisPlex system [150,167,172]. With the currently available DNA variants, the non-blue (such as green) and non-brown (such as hazel) eye colors are much less accurately to predict than blue and brown eye colors. In general, the middle positions of the intermediate colors in the distribution limit their predictive accuracies as the confidence interval for predicting a data point in a middle bin is twice as large as that for predicting a data point in a tail bin. Further, intermediate colors are often ascertained with a higher degree of uncertainty relative to the extremes of the continuous eye color spectrum such as blue and brown, which may substantially influence prediction outcomes of such intermediate colors. Also, eyes classified as intermediate color often contain different colors at different regions of the iris such as gray-blue in the inner part and brown in the outer part, which with currently available DNA variants is difficult to predict. It remains to be seen if currently unknown DNA polymorphisms not involved in blue and brown eye colors will be found to determine non-blue and non-brown eye colors, which may be expected for particular intermediate eye colors such as green. Nonlinear effects between known predictors and their implementation in prediction models are also important for a better prediction of the intermediate colors. In particular, genetic interactions between multiple variants at the TYRPI locus may be important in predicting intermediate colors [83], which should be examined in future studies.

3.2. Hair color

It has also been demonstrated that human hair color categories are predictable from DNA variants. In one study, the independently predictive values of 45 SNPs from 12 genes were investigated in hundreds of Polish Europeans [94]. A prediction model based on a subset of 11 SNPs and two compound MCHR markers from in total 11 genes showed high prediction accuracy for red (AUC = 0.93) and black (AUC = 0.87) hair colors, while predicting blond and brown color was less accurate (0.81 < AUC < 0.82). The contribution of each pigmentation gene to the overall prediction accuracy was ranked as follows: MCHR multiple variants, HERC2 rs12913832, IRF4 rs12203592 and rs4959270, TYR rs1042602, SLC45A2 rs287777 and rs16891982, TYRPI rs683, OCA2 rs1800407, SLC24A4 rs2402130, KITLG rs12821256, and ASIP rs2378249 (Fig. 3B). Using 24 DNA variants as individual predictors (instead of compound MCHR markers in a combined model), a further hair color prediction study using 1551 European subjects from Poland, the Republic of Ireland and Greece showed consistent accuracy estimates and suggested that the obtained hair color prediction is reliable independent of
bio-geographic ancestry [167]. This recently developed DNA test system, called HiirisPlex, for simultaneous hair and eye color prediction includes a highly sensitive and robust multiplex genotyping assay for analyzing 24 hair and eye color predictive SNPs (with all 6 Irisplex SNPs), as well as two prediction models for deriving color probabilities from the combined genotypes, one for hair color and one (the initial IrisPlex model) for eye color [167] (Fig. 4). Although forensic developmental validation of the Hiirisplex assay is still underway, this DNA test system had recently been applied to human remains (teeth and bones) from contemporary times back to medieval times and provided a high success rate for eye and hair color prediction of deceased persons from their remains [173]. Notably, the golden gene SLC24A5 [133] rs1426654 did not contribute to hair color prediction in these studies because the derived allele Thr111 was largely fixed in Europeans. The effect of this DNA variant on hair color should be further investigated in studies including admixed individuals. Yet, the most difficult part for hair color prediction is to differentiate blond from brown colors. This may in part be due to the age-depended hair color change during adolescence, which mostly concerns a change from blond during early childhood to dark blond or brown (rarely black) in late childhood and adulthood [174]. It can be speculated that age-depended hair color change is under biological control, but the responsible molecular mechanism or predicting bio-marker has yet to be identified. We also know relatively little about age-depended loss of hair color, i.e. hair graying and whitening so that molecular prediction of gray/white hair is yet impossible. However, preliminary studies in mice, horses, and men [175–177] suggest that hair graying/whitening, at least in its early onset form [178], is mainly under genetic control. Future studies may identify molecular predictors for gray/white hair, which in combination with molecular age predictors [179], may be useful for practical FDP applications.

### 3.3. Skin color

The genetics of skin color is currently less well understood than that of eye and hair color. This is partly due to the difficulty in gene mapping approaches to distinguish between truly skin color associated alleles from the many other allelic differences that exist between Europeans, Africans and Asians independent from skin color variation. Obviously, there is a strong overlap between the genes influencing eye, hair and skin color but the three pigmentation phenotypes are not completely correlated. Although light skin is present in Europeans and Asians, Europeans have striking eye and hair color variation, whereas Asians usually do not. TYR rs1426654 and SLC45A2 rs16891982 are likely major predictors for skin color differences observed between continents, which still needs to be systematically examined in skin color prediction studies using worldwide samples. A combination of seven SNPs (HERC2 rs12913832, OCA2 rs1545397, SLC45A2 rs16891982, SLC24A5 rs1426654, MC1R rs885479, ASIP rs6119471, and IRF4 rs12203592) has been shown to be a classifier for predicting “not dark” and “not white” in a study of 554 subjects from diverse populations [92]. This finding has been further confirmed in 251 newly collected subjects [171]. The proposed model gives partially overlapping prediction outcomes, which are not commonly accepted in the field of (medical) diagnosis. Although such non-color type predictions have limitations over color-type predictions, they are potentially informative at least as long as color type predictors are not fully available. Furthermore, these studies showed that the SLC24A5 rs1426654 is indeed a very important skin color predictor at the global level. Nevertheless, the current knowledge for skin color prediction is still limited, particularly for the more subtle color variations observed within European, within Asian, and within African populations. Clearly, more skin color genes and associated DNA variants have to be identified, which is complicated by

<table>
<thead>
<tr>
<th>Hair &amp; Eye colour phenotype</th>
<th>Prediction result</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>Hair Dark</td>
</tr>
<tr>
<td></td>
<td>Eye Blue</td>
</tr>
</tbody>
</table>

Fig. 4. DNA prediction of categorical eye and hair color on the individual level. Examples of applying the HiirisPlex system to four European individuals (A–D): [167]. The actual eye and hair colors are displayed on the left side by photographs. The HiirisPlex prediction results, in terms of the probabilities belonging to certain color categories, are shown on the right side, where the color categories with the highest probabilities are highlighted.
the worldwide distribution of the trait as outlined above, in order to have more suitable DNA markers for skin color prediction. For instance, Valenzuela et al. [170] demonstrated that only about half of skin color variance in a multi-ethnic sample could be explained by DNA predictors relative to eye and hair color variation in the same subjects. Measuring skin color using quantitative approaches (see above) will likely lead to the discovery of new genetic variants explaining more skin color variation, as recently demonstrated in a candidate gene study of continuous skin color variation in Europeans [19].

4. Conclusions

Studying the genetics of human pigmentation is important for understanding human biology and evolution, as well as skin cancer biology, and DNA predictors of eye, hair and skin color are useful in practical applications such as in forensics. Boosted by high-throughput genotyping technologies combined with new analytical approaches, the human pigmentation genetic community has made significant progress over the last decade in identifying new genes and highlighting various predicting DNA variants. We now know a number of genes determining human pigmentation traits, and some more are expected to be identified in future studies. We now can tell with high confidence some person’s eye, and hair color (and soon expected perhaps also skin color) based on a small number of his/her DNA polymorphisms. Colorful DNA polymorphisms serve as the exact molecular identities underlying “how” and “why” a striking diversity of colors exists between and within human populations. While these achievements are exciting and have practical potential, lots of work is still ahead of us to find the missing heritability of human pigmentation traits. We hope that the human pigmentation genetic community will continue to be productive and innovative in their mission to harvest all DNA variation underlying eye, hair, and skin colorations in modern humans as well as their archaic counterparts.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.smedcb.2013.03.013.

References


pigmentation
morphism
measures
dran
in
status.
Kanetsky
Liu
Schioth
Harding
the
Struchalin
clone
A,
DL,
agouti
Ollmann
of
HR,
J,
kineti cs
Box
loss-of-function
cellular
inhibits
from
and
of
at
S,
color
oculocutaneous
JS,
Perez
JL,
Montgomery
receptor
of
NG,
P,
Sambuughin
JL,
the
Pigment
dermal
influence
on
the
dermal
on
pigmentary
origins.
Nature
Genetics


