Forensic Science International: Genetics 18 (2015) 33–48

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes☆

Manfred Kayser *

Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 1 October 2014
Received in revised form 29 January 2015
Accepted 11 February 2015

Keywords:
Appearance
Externally visible traits
Physical traits
Forensic DNA Phenotyping
FDP
DNA intelligence

ABSTRACT

Forensic DNA Phenotyping refers to the prediction of appearance traits of unknown sample donors, or unknown deceased (missing) persons, directly from biological materials found at the scene. "Biological witness" outcomes of Forensic DNA Phenotyping can provide investigative leads to trace unknown persons, who are unidentifiable with current comparative DNA profiling. This intelligence application of DNA marks a substantially different forensic use of genetic material rather than that of current DNA profiling presented in the courtroom. Currently, group-specific pigmentation traits are already predictable from DNA with reasonably high accuracies, while several other externally visible characteristics are under genetic investigation. Until individual-specific appearance becomes accurately predictable from DNA, conventional DNA profiling needs to be performed subsequent to appearance DNA prediction. Notably, and where Forensic DNA Phenotyping shows great promise, this is on a (much) smaller group of potential suspects, who match the appearance characteristics DNA-predicted from the crime scene stain or from the deceased person's remains. Provided sufficient funding being made available, future research to better understand the genetic basis of human appearance will expectedly lead to a substantially more detailed description of an unknown person's appearance from DNA, delivering increased value for police investigations in criminal and missing person cases involving unknowns.

1. Forensic DNA Phenotyping: some general considerations

Forensic DNA analysis, i.e., the identification of persons via short tandem repeat (STR) profile matching of unknown evidence material with reference material from known persons, has been considered the golden standard in forensic sciences [1]. However, one of the major limitations of this comparative approach of DNA identification, likewise applying to STRs and single nucleotide polymorphisms (SNP), is that it typically fails to identify persons whose STR or SNP profile is not already known to the investigators. Persons may be unavailable for comparative DNA profile matching because they have successfully escaped police investigations and thus avoided becoming a known suspect. Although this current approach becomes more effective when forensic DNA (profile) databases are in place [2], cases where the evidence DNA profile does not match that of any known person including all stored in the forensic DNA (profile) database are routinely seen by investigators. In the absence of any other information that provides leads for tracing unknown forensic sample donors, cold cases can wait for various periods of time (sometimes for very long), before the evidence STR profile is matched with a known person subsequently added to the grown forensic DNA database or delivered as suspect by police re-investigation of the given case.

DNA mass screenings can be carried out in cases where no DNA profile match is obtained and no other evidence is available [3]. In such DNA dragnets, larger number of persons (hundreds to thousands), usually those living in the geographic region where the crime occurred, are invited to voluntarily provide a saliva sample for STR profiling. Although the true perpetrator may not participate voluntarily, due to awareness of the provided sample leading to identification, non-participation may raise suspicion and thus directing investigators towards additional leads. If the true perpetrator does not participate but only close relatives do, familial search is able to identify them, which provides investigative leads to find the unknown perpetrator. Using conventional

☆ This article is dedicated to the 10th Anniversary of the Department of Forensic Molecular Biology at Erasmus MC University Medical Center Rotterdam (http://www.erasmusmc.nl/fmb/), and all its current and former members involved in fundamental and applied research on human appearance genetics and DNA prediction, among various other achievements between 2004 and 2014.

* Tel.: +31 10 7038073; fax: +31 10 7044575.
E-mail address: m.kayser@erasmusmc.nl (M. Kayser).

http://dx.doi.org/10.1016/j.fsigen.2015.02.003
1872-4973/

autosomal STR profiling in the DNA dragnet limits the possibilities of familial search to close relatives of the unknown, non-participating perpetrator, which can be overcome by using Y-chromosomal STRs instead (if the evidence DNA originates from a male perpetrator) [4]. Since a Y-STR profile identifies a man together with all his paternal male relatives, close and distant ones, a Y-STR dragnet is more effective than dragnets based on autosomal STRs (or SNPs). For instance, a large Y-STR dragnet involving thousands of local volunteering men finally led to solving a murder case in the Netherlands after 13 years of investigation [5]. Still, in order to be potentially successful, close and/or distant relatives of the true perpetrator (if not the perpetrator himself) have to participate in the DNA mass test. The local presence of relatives may be more likely in rural areas, where relatives are less likely to migrate away, than in urban areas. In general, however, such DNA dragnets without specific cause and evidence to ask volunteers are often seen critically due to ethical concerns, and in some countries are legally forbidden. Furthermore, the economic burden to obtain STR profiles of hundreds or even thousands of individuals in a single case is high. Because of these reasons, DNA dragnets are not applied often [3–5].

These limitations of comparative DNA profiling stimulated a relatively new development within forensic genetics, i.e., Forensic DNA Phenotyping (FDP) [6,7]. FDP aims to infer the unknown stain or sample donor's externally visible characteristics (EVCs) from DNA (or other molecular biomarkers) directly from the biological material left behind at the scene of crime, or obtained from unknown bodies. In essence, FDP outcomes can serve as "biological witness", and may potentially provide even more accurate information than human eyewitnesses do, who are known to be unreliable [8]. As such, FDP is expected to provide investigative leads allowing to trace unknown perpetrators, who are not identifiable via conventional comparative DNA profiling. FDP is also expected to be useful for missing persons identification, i.e., in cases where reference DNA profile from putative ante-mortem samples, or from putative relatives are unavailable. The DNA inference of bio-geographic ancestry (see Philipps in this issue) is sometimes considered part of FDP [7]; however, genetic ancestry does not always portray an externally visible characteristic, particularly in individuals of mixed genetic ancestry.

Appearance prediction from DNA for forensic usage started in the early 2000s and first progressed very slowly. The main reason for the relatively late introduction of forensic appearance prediction from DNA is the (still) limited knowledge about the genetics of most human EVCs. Even though it takes the same technological equipment and statistical methods to identify disease genes as to find EVC genes, our knowledge about inherited diseases is currently more advanced [9] than on how we look. One of the reasons for limited appearance genetic knowledge till today might be related to research funding strategies that typically focus more on disease-related variation than on normal human variation and its genetic exploration. Of all EVCs, those that involve pigmentation i.e., variation in the coloration of the human iris, head hair, and (less so) skin, are the best and currently the only examples of practical FDP (see below). Although all EVCs are considered complex traits, where several to many genes are contributing to the phenotype together with environmental factors, human pigmentation traits in general seem the least genetically complex of all EVCs, with a few handful of genes providing most of the phenotypic information, at least on a broad categorical level. Therefore, understanding the genetic basis of pigmentation traits is currently more advanced than for any other EVC, and thus is DNA-based pigmentation prediction. All other EVCs are, based on current knowledge or expectations developed from current knowledge, genetically much more complex with dozens to expectedly thousands of genes contributing, which complicates the identification of responsible genes and predictive DNA markers.

The problem with highly complex genetic traits, as realized for many common diseases, is that every individual gene contributes only a small proportion of the phenotypic variance, and only the combination of a large number of genetic factors may explain the overall inherited component [10]. Moreover, the larger the environmental component, the less can be explained by DNA, and – of course – any non-genetic contribution can never be explained by a DNA test. According to anecdotal knowledge, and based on previous findings from twin heritability studies [11], human EVCs typically carry a large genetic component, but environmental impacts also exist, for some EVCs more so than others. If however a gene only has a small individual effect on the phenotypic trait, it is difficult to be identified with the current toolbox used by genetic epidemiologists, because the measurable statistical signal is minutely small. Therefore it requires the use of large sets of individuals to identify such small genetic effects with the needed level of statistical significance. Since the genomic tools used for finding genes, such as SNP microarrays, are still expensive (i.e., approx. 250 EUR per individual sample, and exome or whole genome sequencing are by magnitudes more expensive), carrying out genome-wide association studies (GWASs) on large numbers of individuals (i.e., tens of thousands) with large numbers of single nucleotide polymorphisms (SNPs) (i.e., hundreds of thousands and more) quickly becomes unaffordable for the average single laboratory. The formation of large international consortia, has demonstrated to be highly successful in finding complex trait genes, mostly common complex diseases, by combining impressively large numbers of samples (up to hundreds of thousands) [9]. Consequently, given the complex genetic nature of EVCs, only large collaborative efforts will allow unveiling their genetic basis as a prerequisite for developing predictive DNA markers and tools for practical FDP.

2. DNA phenotyping of pigmentation traits: the first FDP success story

In the following three sub-chapters I summarize the current knowledge on DNA-based prediction of eye, hair, and skin color, respectively. Due to space constrains, and because it is the predictive value of a SNP that is relevant for FDP purposes, I mostly leave out association and linkage studies on human pigmentation traits. Table 1 lists all SNPs previously applied for eye and/or hair and/or skin color prediction from DNA.

2.1. Eye color

The first two studies that performed DNA-based iris (eye) color prediction were published in 2007. Frudakis et al. [12] used 33 SNPs from the *OCA2* gene, which allowed them to classify 8% of the eye colors observed among >1000 samples. Sulem et al. [13], embedded in the first GWAS on human pigmentation traits, used 9 SNPs from 6 genomic regions (*SLC24A4, KITLG, 6p25.3, TYR, OCA2–HERC2,* and *MC1R*) which they identified with significant eye color association among several thousand Europeans, for categorical eye color prediction. Of the individuals DNA-predicted with <0.2 probability for brown and <0.1 probability for green, about 90% were indeed blue eyed, and of the individuals DNA-predicted to be brown with >0.5 probability, about 60% were indeed brown eyed. In 2008, three parallel studies [14–16]