Universidade de Aveiro Departamento de Biologia
Ano 2018/2019

Inês Cristina Romão Cardoso

DNA phenotyping and drug induce pigmentation: a theoretical approach

Fenotipagem do DNA e pigmentação induzida por drogas: uma abordagem teórica

DECLARAÇÃO

Declaro que este relatório é integralmente da minha autoria, estando devidamente referenciadas as fontes e obras consultadas, bem como identificadas de modo claro as citações dessas obras. Não contém, por isso, qualquer tipo de plágio quer de textos publicados, qualquer que seja o meio dessa publicação, incluindo meios eletrónicos, quer de trabalhos académicos.

Universidade de Aveiro
Ano 2018/2019
Departamento de Biologia

Inês Cristina Romão Cardoso

DNA phenotyping and drug induce pigmentation: a theoretical approach

Fenotipagem do DNA e pigmentação induzida por drogas: uma abordagem teórica

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Aplicada, realizada sob a orientação científica do Doutor Luís Manuel Souto de Miranda, Professor Auxiliar do Departamento de Biologia da Universidade de Aveiro

o júri

presidente

Prof.ª Doutora Etelvina Maria de Almeida Paula Figueira
Professora Auxiliar do Departamento Biologia da Universidade de Aveiro

Prof.ª Doutora Maria de Lourdes Gomes Pereira
Professora Associada com agregação da Universidade de Aveiro

Prof. Doutor Luís Manuel Souto de Miranda
Professor Auxiliar em Regime Laboral da Universidade de Aveiro

Agradecimentos

Agradeço, em primeiro lugar, ao Prof. Doutor Luís Souto, pela oportunidade de desenvolver o trabalho conducente à presente dissertação e pela orientação e acompanhamento prestados.

Agradeço também á Dr. Helena Moreira que esteve sempre presente durante todo o desenvolvimento desta dissertação, cujo apoio e paciência foi fundamental.

Aos meus amigos, por toda a compreensão, paciência e palavras de incentivo. Obrigada por não me deixarem desistir.

Por último um agradecimento aos meus pais e a toda a família por sempre me apoiarem e pelo patrocínio de mais um ano de faculdade. Obrigada pela confiança que depositam em mim e obrigada pela pessoa que sou hoje em dia, porque se sou como sou é graças a vocês!

Palavras-chave Fenotipagem do DNA, características externamente visíveis, pigmentação, marcadores genéticos, genética forense

Resumo

A variação genética humana tem um alto poder discriminativo; portanto, a previsão de fenótipos através do genótipo tem sido um tópico de crescente interesse para estudo, existindo expectativa do seu uso na medicina personalizada e na área da genética forense, sendo esta última de maior interesse na previsão de características visíveis externamente.

Uma das áreas de variação genética humana com mais variantes é a pigmentação, e pela sua natureza multifatorial é de grande interesse estudar o efeito de medicamentos sobre essa característica, ainda mais considerando que medicamentos já identificados como relacionados com a pigmentação induzida, estão presentes no ambiente e, por esse motivo, somos expostos a eles, mesmo sem conhecimento.

A influência de drogas na pigmentação é uma área com quase nenhum estudo, sendo apenas mencionada mas não explorada muito além, principalmente porque aqueles estudos que se concentram nos efeitos das drogas nos organismos normalmente estão mais relacionados a patologias, sendo a pigmentação uma característica descartada facilmente.

Keywords DNA phenotyping, externally visible characteristics, pigmentation, genetic markers, forensic genetics

Abstract

Human genetic variation has a high discriminatory power, so the prediction of phenotypes through genotype has been a topic of growing interest for study, and there is anticipation for its use in personalized medicine and in the area of forensic genetics, the latter having a greater interest in predicting externally visible traits.

One of the areas of human genetic variation with more variants is the pigmentation, and for its multifactorial nature is of great interest to study the effect of pharmaceutical drugs on this characteristic, even more considering that drugs that have already been identified as related to induced pigmentation, are present in the environment and for that reason we are exposed to them even without knowledge.

Drug influence in pigmentation is an area with almost no studies, having only a few that mention pigmentation but do not explore this any further mainly because the ones who focus on drugs effects on organisms normally are more related to pathologies being the pigmentation a characteristic that is easily discarded.

Table of Contents

Figure index

List of abbreviations

AIM - Ancestry Informative Marker

DNA - Deoxyribonucleic Acid

EVC - Externally Visible Characteristics

HLA - Human Leukocyte Antigen

LD - Linkage Disequilibrium

MHC - Major Histocompatibility Complex

mtDNA - Mitochondrial Deoxyribonucleic Acid

NSAIDs - Nonsteroidal anti-inflammatory drugs

RFLP - Restriction Fragment Length Polymorphism

rRNA - Ribosomal Ribonucleic Acid

SNP - Single Nucleotide Polymorphism

STRs - Short Tandem Repeats

tRNA - Transfer Ribonucleic Acid

Introduction

During an investigation several samples are taken, some of them finish in a forensic genetics service for donor identification, when this situation occurs the identification profile is determined, this profile is then compared with other reference sample profiles (known donor samples, or profiles in the national database) for identification purposes. However, there are many cases where identification is not established because there is no match (Kayser & Schneider, 2009).

However, the information contained in the genetic material of biological samples is not limited to identification, many physical (for example eye colour, hair colour, skin colour, height) and even behavioural characteristics are based on genetic factors that when known can be used to profile the donor from a sample that had no result when analysed (Kayser, 2015; Kayser & Schneider, 2009).

Medicines are an emerging class of environmental contaminants widely used in human medicine and veterinary medicine, and after their administration, a significant proportion is eliminated into the environment contaminating groundwater, surface water and soil (Jones *et al.*, 2003; Mompelat *et al.*, 2009). The environmental impact that results from their presence is a growing public health concern worldwide as they have a biological effect and there are more and more studies in this area, but the results are currently insufficient to establish a risk for the human race. Thus, it is possible to verify that even if the human being does not have direct contact with a drug, it comes into contact with it through the environment, and may suffer possible consequences of this interaction.

As it has already been described in the literature there are a variety of medications that have already been associated with changes in the pigmentation, having this symptom the denomination of drug-induced pigmentation, being estimated to represent 10 to 20% of all cases of acquired hyperpigmentation (Dereure, 2001).

Hereupon the main objectives that accompanied the realization of this dissertation were in general to describe the theoretical, bioethical and legal aspects underlying the prediction of ancestry and phenotypes from DNA (Forensic DNA Phenotyping); to describe the theoretical aspects of the influence of drugs on pigmentation and finally to propose a practical work to study the influence of a drug on pigmentation in order to try to answer some questions raised in this document.

1. Forensic Genetics

The discovery of the structure and function of DNA in the twentieth century was one of the greatest revolutions in the field of biology, as it made possible the understanding of the fundamental mechanisms of life. This has led to an increase in curiosity about the genome, and the genomes of several species, including the human being, are already fully sequenced. The study of the genome has made possible and still allows the analysis, discussion and understanding of the phenomena of heredity, inter and intraspecific variability, among other aspects (Pelt-Verkuil *et al.,* 2008).

When looking to the theme "forensic genetics" there is the need to clarify the meaning of the word "forensic". This word comes from forensic (Latin) meaning "before the forum" referring to Roman times when a criminal case was presented to a forum of people, so forensic science is the application of scientific expertise to questions of law in relation to criminal or civil actions, being in this case the area of genetics. In the beginning were used the ABO blood groups once the investigators realized that the group variants could be used to solve paternity testing cases and crimes. In the late 50s the genetic differences detected through protein polymorphisms were discover, representing an advance in this field, since that moment until 1970s the number of polymorphic proteins that can be analysed by electrophoretic methods continuously increased up. In the beginning of the 60s were first mentioned and typed histocompatibility (HLA) antigens, that were more polymorphic than any other genetic marker known until then, however these markers had limitations for example where when it was necessary to analyse minimal or degraded material and the difficulty to analyse biological samples that were not blood, like for example semen in rape cases (Carracedo, 2013). These markers complemented the fingerprints, and in some cases, became the primary forensic evidence in the courts (Trent, 2012). Despite of this, the information that these markers were able to offer in many cases was clearly insufficient and, therefore, the discovery of hypervariable loci in minisatellites represented a milestone in the field and one of the most important discoveries in the history of forensic science, once DNA analysis are more informative and DNA can be obtained from any tissue and on degraded samples, once is more resistant to degradation than proteins (Carracedo, 2013; Jeffreys *et al.*, 1985).

The concept of phenotype concerns observable characteristics in a cell or organism and results of the expression of genes and also the influence of environmental factors, this phenotypic characteristics can be characterized as simple (single gene) or complex (result of several factors), being that most of the physical characteristics of human beings are complex characteristics, which makes it difficult to identify the genetic factors that underlie them (Pulker *et al.*, 2007).

1.1 Genetic marker

To determine the genetic factors related to the phenotypic characteristics is use the study of genetic markers, i.e. segments of DNA whose sequence and position in the genome are known. Genetic markers may be autosomal, mitochondrial, X-chromosome or Y-chromosome, DNA polymorphisms may assume the role of genetic markers, since by definition they are the simultaneous existence of more than one sequential form in a given locus in a given population, when their position is known they may play the role of a genetic marker with great discrimination power (J. M. Butler, 2009).

In the case of autosomal markers, in forensic context, it is mainly used for identification and paternity testing, since autosomes, ie non-sexual chromosomes 1 to 22, have one of the elements in each pair of chromosomes inherited from the father and the other from the mother.

1.1.1 Mitochondrial DNA (mtDNA)

Mitochondrial DNA (mtDNA) appears as a circular double-stranded DNA molecule (**Figure 1**) and is located in the mitochondrial matrix (Chial & Craig, 2008). The mitochondrial genome includes 37 coding regions, which correspond to thirteen proteins involved in the oxidative phosphorylation process, encode two rRNA molecules and twenty-two tRNA molecules, all transcripts and translation products remain in the organelle not being exported to the cytoplasm. The level of variation is not constant throughout the molecule being the rate of change in the control region, non-coding region, 10 times higher than that observed in the coding region, this difference being explained by the fact that the occurrence of mutations in the coding regions results in transcriptional product changes

therefore at the level of the functioning of vital processes, whereas in the control region most of these mutations have no practical consequences on cellular functioning (Hoffmann & Spengler, 2018; van Oven & Kayser, 2009). Since in most cases the mitochondria come from the female gamete, the mtDNA of each individual is inherited from the mother so the analysis of this type of genetic material allows direct access to the individual's maternal lineage. However more recent discoveries as pointed to a biparental inheritance of mtDNA but as it is a rare finding, for now has no applications on forensic sciences (Luo *et al.*, 2018; Mcwilliams & Suomalainen, 2019).

The major advantage of mtDNA is that it exists in large numbers which makes it very advantageous in the forensic context in cases where there is limited amount of autosomal markers, so mtDNA analysis plays an important role in, for example, the identification of human remains from bones, hair or teeth, since by degradation processes, autosomal markers are not always available, in this case the comparison is made with living maternal relatives (Kayser, 2007).

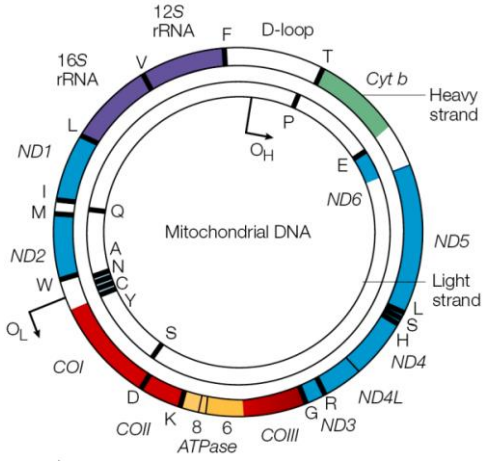

Figure 1 Human mitochondrial DNA (mtDNA). The mtDNA consists of a inner and outer circular DNA strand, they can be designated light and heavy respectively. Promoters on the heavy strand and on the light strand drive mtDNA gene expression. **Retrieved from:** (Chial & Craig, 2008)

1.1.2 X Chromosome

X Chromosome is one of the chromosomes that determine the gender of the individual, female individuals have a pair of X chromosomes, while male individuals have an X chromosome paired with a Y chromosome, that is, female individuals inherit a X

chromosome from the mother and one from the father, while males receive only one X chromosome from the mother which is later transmitted exclusively to female offspring. X chromosome markers in forensic context are very useful, especially in paternity testing under special conditions, for example determining if a man is the father of a girl but only the DNA of a daughter of the alleged father is available for analysis; since the X chromosome is transmitted integrally from father to daughter, if both girls are daughters of the same father inherited the same chromosome (Szibor, 2007).

1.1.3 Y chromosome

As mentioned earlier, the Y chromosome is unique to males and is transmitted from parent to child. When talking about Y chromosome in the context of molecular studies the reference generally refers to the portion of the chromosome that does not undergo any recombination, non-recombinant Y (NRY). In the forensic context Y-chromosome markers are useful in particular for analysing sample retrieved from rape cases and paternity testing determination of boys where no DNA sample is available from the father but genetic material from a male relative of paternal lineage is available (Jobling & Tyler-Smith, 2003).

1.1.4 DNA polymorphisms

DNA polymorphisms were first described in the 1970s/ 1980s, as the existence of at least two variants with respect to gene sequences or chromosome structure; in other words, are the different DNA sequences among individuals, groups, or populations (Teama, 2016; Trent, 2012).

DNA polymorphisms provided a more sophisticated approach to tissue comparisons leading to the concept of DNA fingerprinting, where minisatellites, complex DNA polymorphisms, could be used to produce DNA profiles for individuals. Later with the application of PCR and the use of the polymorphism called microsatellite DNA fingerprinting became more efficient, having the value DNA fingerprint improved with the advances in DNA technology and with the better understanding of the distribution of DNA polymorphisms within populations (Trent, 2012).

There are several forms of DNA polymorphisms such as variations on the number of copies, Single Nucleotide Polymorphisms (SNPs), variable number of tandem repeats (VNTRs, for example mini- and microsatellites), transposable elements and structural alterations, and they can be study using a variety of techniques, for example restriction fragment length polymorphisms (RFLPs) with Southern blots which represents the first method to study genetic variability in order to identify individuals; polymerase chain reactions (PCRs) and genome sequencing. RFLPs study polymorphisms of the VNTR type, having been replaced because later techniques allowed the analysis of degraded samples of limited quantity and in addition the number of markers under study and their characteristics allow a greater power of discrimination (Teama, 2016).

1.1.4.1 Short Tandem Repeats

STRs also called Short Tandem Repeats or Microsatellite, consist of sequences in which one to six nucleotides are repeated and represent about 3% of the human genome being randomly distributed. When the number of nucleotides present in the repetitive units is higher than six we are in the presence of a minisatellite or a DNA satellite if the number of nucleotides exceeds five hundred (Ellegren, 2004).

The occurrence of microsatellites is higher in non-coding regions and it is assumed that this is due to the fact that this regions are not subjected to selective pressures, in its turn in coding regions the natural selection against mutations makes the appearance of STRs difficult, this STRs may be a point mutation that gives rise to two or three repetitive units or insertion of DNA segments that are repeats of the adjacent sequence (Ellegren, 2004).

Length changes in microsatellite DNA are normally associated with replication slippage, temporary dissociation of the replicating DNA strands followed by misaligned reassociation (**Figure 2**), better explaining after the replication initiation, the two strands might dissociate, and if the nascent strand then realigns out of register, continued replication will lead to a different length from the template strand. If the loop is on the nascent strand this will represent an increase in repeat length but if the loop is formed in the template strand than represents a decrease in repeat length. However is important to notice that the repairing system will fix most of this mutations, so only a few percentage will persist (Ellegren, 2004).

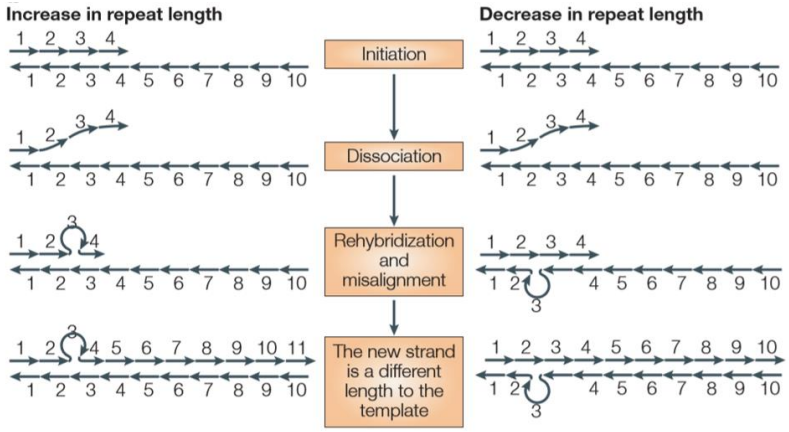

Figure 2 Replication slippage**. Retrieved from:** (Ellegren, 2004)

The high mutation rate, usually directly proportional to the number of repeating units that characterizes microsatellites, makes them highly polymorphic within the population and therefore highly informative, being an STR considered polymorphic when the number of repeating units varies within the population with an incidence greater than 1% (Ellegren, 2004).

In a concise manner the main advantages and the reasons that make STRs the markers of choice in the forensic context are their high informative power, the convenience of their detection, which is based on PCR methods and therefore the fact that it does not require a large amount of DNA, and the ability to transfer marker information based only on PCR primer sequence (Birren *et al.* , 1997).

1.1.4.2 Single Nucleotide Polimorphisms

Single Nucleotide Polymorphisms (SNPs) refers to DNA sequence variations corresponding to differences in a single nucleotide (substitutions, deletions or insertions), however in the literature some authors argue that SNPs correspond only to substitution polymorphisms, with insertions and deletions belonging to a special group called “indels”. That said, in general, studies separate insertions and deletions from substitutions and only the latter are considered SNPs (Budowle & Van Daal, 2008; Collins *et al.*, 1999; Mullaney *et al.*, 2010).

Depending on the location of a base change in the DNA sequence the consequences will also vary, for example if the change is in a non-coding region, which are the most

common changes in eukaryotes, as these regions represent most of their genome, a polymorphism is less likely to cause negative effects on cell function. However, if the change occurs in a coding sequence we can have two possible outcomes, if the change is in an intron, there is usually no change in the gene product, but if the change occurs in an exon, there is a possibility that a change in the gene product will occur. However, this change may also not result in a different product because the genetic code is redundant and one amino acid can be encoded by different triplets and the consequences on cell function will be null, in which case are called silent mutation (Dale & Schantz, 2002), for example in the case of leucine that can be encoded by UUA and AUC (**Figure 3**).

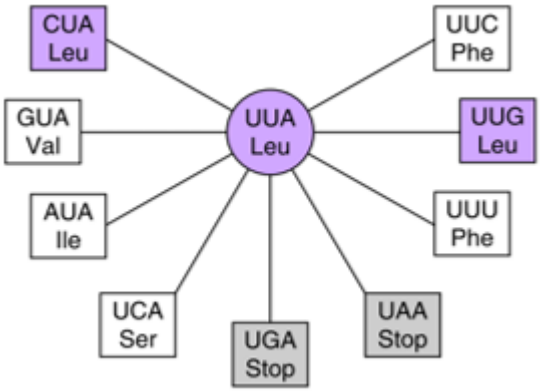

Figure 3 Resulting codons of single-base substitutions from UUA, showing the redundancy of the genetic code. **Retrieved from:** (Dale & Schantz, 2002)

One coding region that exhibits a high level of variability is the human leukocyte antigen (HLA) system, which represents the Major Histocompatibility Complex (MHC) in humans located on the short arm of chromosome 6, whose proteins perform immune system functions, being involved in the presentation of antigens to C cells and molecules involved in the inflammatory process. Studying this variability is very useful not only in clinical terms but also in studying evolutionary history and genetic variation between populations (Buhler & Sanchez-Mazas, 2011). To contradict this high variability zone there are also highly conserved regions of the genome for example, regions that are implicated in RNA processing or transcriptional regulation, where the observed level of variation is extremely low, so it is possible to conclude that the distribution of SNPs in the genome is not homogeneous.

Since SNPs occur very frequently in the human genome, they are good genetic markers for identity testing, mapping of simple or complex traits, genotype-phenotype association studies and reconstruction of human evolution. Because they are physically located close to each other, some markers tend to be block-transmitted, meaning that the presence of an allele at one locus is associated with the presence of a specific allele at another nearby locus, being called haplotypes the set of markers closely linked that are usually inherited together. Thus, loci belonging to a given haplotype are in Linkage disequilibrium (LD) (Kayser, 2015; Wirtz *et al.*, 2018).

Linkage disequilibrium occurs when the presence of one allele at a given locus is associated with the presence of another specific allele at a nearby locus, which means that the probability of occurrence of the two considered alleles is not independent, hereupon LD is considered to be an indirect measure of the distance between loci. LD patterns are affected by many factors such as interlocus physical distance, centromere distance, presence of recombination hotspots, chromosome location, study population, gene conversion and type of markers used, being that over time it decreases as a result of recombination, which causes the LD pattern to reflect several generations of recombination as it is potentially affected by all the recombinations that have occurred throughout history (Jorde, 2001).

In forensic context and following Budowle & Van Daal's, 2008 classification SNPs can be classified into four categories: identification, phenotype informative, ancestry informative and lineage informative. Identification SNPs are used to individualize as the STR identification markers, phenotype informational SNPs are used to establish the likelihood of an individual having a certain phenotypic characteristic, such as hair, skin or eye colour, informational Ancestry SNPs are used in establishing an individual's biogeographic ancestry in order to indirectly infer some phenotypic characteristics that provide clues about the suspect, and finally, lineage informational SNPs that correspond to sets of SNPs that function as haplotype markers to identify missing persons across kinship analysis (Budowle & Van Daal, 2008).

Comparing SNPs with STRs, STRs are more informative, since unlike SNPs that mostly only have two alleles, STRs have several alleles with heterozygote levels of around 70% (Birren *et al.*, 1997). For a discriminating power similar to the one provided by STR multiplex systems there would have to be a battery of fifty to one hundred SNP markers, which would represent the need for a much larger DNA sample than that required for STR

multiplex systems. Another limiting factor of SNPs is the interpretation level in situations involving sample mixing, what lead to a discussion where many authors argue that it is unlikely that SNPs will replace STRs in cases of mixed samples when the objective is identification. It is also important to mention that forensic databases are based on STR markers, which further complicates the possibility of substitution by SNPs as main forensic markers (Budowle & Van Daal, 2008; Frudakis, 2008).

Having said the above, at least for present time STRs are more advantageous than SNPs in the field of identification, but in the field of prediction of biogeographic ancestry and phenotypes SNPs have more advantages, since they constitute the major part of functional and genetic variation, their frequency in coding regions or proximity of them, and since genes directly influence phenotypic characteristics. The low mutation rate associated with SNPs contributes to population substructure effects, making them useful in determining inter-population variation and therefore predicting biogeographic ancestry (Budowle & Van Daal, 2008; Frudakis, 2008).

1.2 Forensic DNA phenotyping

Focusing on DNA analysis in forensic sciences, the introduction of DNA analysis in court was strongly challenged and contested, however nowadays DNA evidence is treated as the "gold standard" of modern forensic techniques. This Forensic DNA analysis consists on the identification of persons via short tandem repeat (STR) profile matching of unknown evidence material with reference material from known persons (Committee on Identifying the Needs of the Forensic Sciences Community; National Research Council, 2009). However, this comparative approach represents one of the main limitations of this technique, failing to identify people whose profile is not yet known to the investigators (**Figure 4**). Although DNA profile databases make this approach more effective, depending on the laws of the countries, these databases may be incomplete, with a low number of individuals, thus reducing the likelihood of a match (Kayser, 2015; Santos *et al.*, 2013).

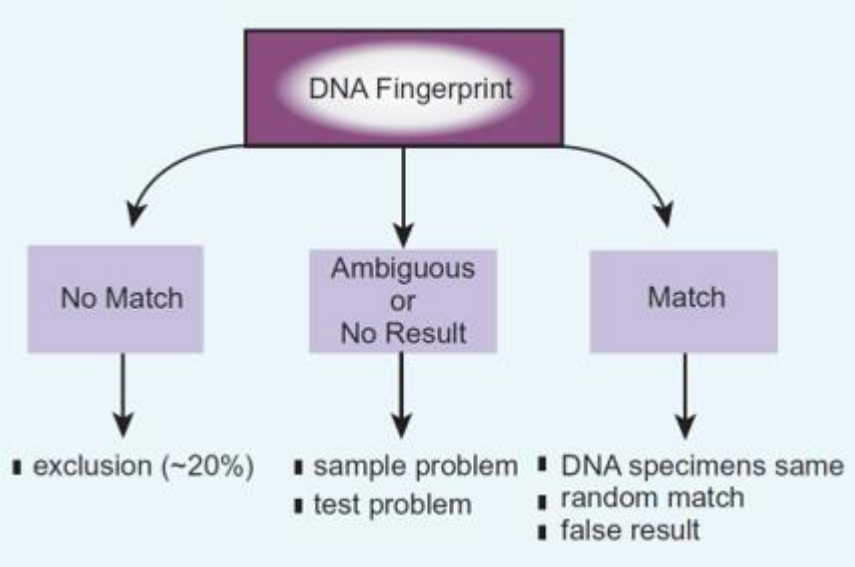

Figure 4 Outcomes from a DNA comparison in the forensic situation. **Retrieved from:** (Trent, 2012)

The comparative nature of DNA analysis led to the relatively recent development of a new method in forensic genetics the Forensic DNA Phenotyping (FDP) whose objective is inferring the visible external characteristics directly from the biological material left at the scene or obtained of unknown bodies. The aim is for the results of the FDP to function as "biological witness" and could potentially provide more accurate information than human eyewitnesses, and may thus assist investigations where individuals are not identifiable by

the conventional comparative DNA profile, which include cases of identification of missing persons whose ante-mortem DNA is not available (Kayser, 2015; Kayser & De Knijff, 2011; Kayser & Schneider, 2009). However, unlike the identification profile, it is not intended to identify the donor of the sample, but to limit the number of suspects to be considered (Kayser & Schneider, 2009).

Determining biogeographic ancestry from DNA is sometimes considered part of the Forensic DNA Phenotyping, however genetic ancestry does not always represent externally visible and accurate characteristics, for example the case of individuals of mixed descent (Kayser & De Knijff, 2011). However, when biogeographic ancestry is used for phenotype prediction, it is considered an indirect prediction. The study of the genetic determinants of phenotypic traits is considered a direct approach, comparing the two methods, the latter is more advantageous in terms of information available, but is rarely applicable because it is expensive and complex to apply (Frudakis, 2008).

Approaching the direct method first, it involves identifying the genetic elements underlying traits in order to make it possible to predict a particular phenotype of an individual basing of the analysis of its DNA. As mentioned earlier, complex traits are determined by a set of variable number genes with particular characteristics which makes their identification difficult, however databases have been created with thousands of markers from different populations (Cavalli-Sforza, 2005).

To establish a relationship between the genetic basis and a phenotype, phenotyping is required, which is the characterization of phenotypic traits, that can become quite complex as is the case with pigmentation-related traits - eye, hair and skin colour, that besides having a complex genetic base are also subject to environmental factors (Frudakis, 2008; Kayser, 2015; Kayser & Schneider, 2009).

In order to proceed with phenotyping there are several types of approaches, for example spectrophotometric methods and digital photography analysis, in order to try to reduce the subjective evaluation as is the case of Frudakis, 2008 which objectively classify phenotypic characteristics such as melanin index rather than "skin tone". However, the Kayser group, known as VISAGE Consortium (VISible Attributes Through GEnomics), which is currently one of the most relevant group in this emerging area, normally uses observational assessment methods, and argue that there is no evidence that quantitative assessment methods are better than methods based on self and heterodescription (Branicki

et al., 2011). After the results of phenotyping these results should always be viewed critically, as mentioned earlier if it is a feature that may be influenced by environmental factors this must be taken into account, besides with the advancement of technology in aesthetics there are many ways to alter physical characteristics voluntarily such as hair dyeing, wearing contact lenses or even through surgical alteration (Frudakis, 2008).

1.2.1 Gender

Determining an individual's gender is one of the oldest approaches in the identification methods, and is achieved through analysis of several markers however the most used and with better results is the Amelogenin gene, whose proteins are expressed in the tooth and have an important role in enamel biosynthesis. Sexual diagnosis is achieved because the Amelogenin gene have differently sized copies on X chromosomes and Y chromosomes (**Figure 5**) (Parker *et al.*, 2019). The most used set of primers target a 6-bp insertion/deletion within an intron of the Amelogenin gene on the X and Y chromosomes and produce 106-bp and 112-bp amplicons for the X and Y chromosomes, respectively (E. Butler & Li, 2014; Codina *et al.*, 2009; Masuyama *et al.*, 2017).

Currently this marker is already included in most forensic identification tests, however it is not error free, being most of these errors due to a deletion of Y chromosome or a point mutation in the X chromosome region to which the reverse Amelogenin gene primer binds in several commercial identification kits (Cadenas *et al.*, 2007; Kao *et al.*, 2007). The best way around these errors is to interpret the results carefully and if possible add other markers whose location is not close to the amelogenin or even Y chromosome STR markers (Kayser & Schneider, 2009).

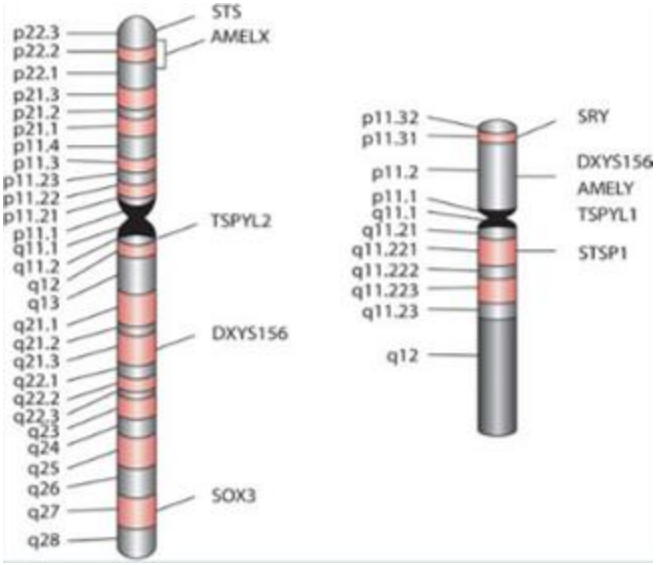

Figure 5 Cytogenetic map of the human X (left) and Y (right) chromosomes with sex typing marker locations, including the Amelogenin gene on the X (AMELX) and Y (AMELY) chromosomes. **Retrieved from:** (Butler & Li, 2014)

Other possible errors are cases where, for example, there have been bone marrow transplants in which the donor and recipient have different genders, which is called chimerism or cases where is observed micro-chimerism, which occurs when a woman is pregnant of a male fetus. Another situation that cannot be overlooked especially today is the difference between biological and legal gender present in the identification documents observed in cases of intersexuality and transsexuality. In all these situations mentioned above the observed phenotype and/or legal records may indicate a gender contrary to the biologically tested which makes the analysis of results more complex (von Wurmb-Schwark *et al.*, 2007).

1.2.2 Age inference

Age inference refers to the analysis of an unknown individual's DNA sample to probabilistically infer that individual's age (Samuel & Prainsack, 2018).

With aging the changes in DNA are for example the accumulation of mtDNA deletions, the decrease in telomere size and the decrease in the number of circular DNA molecules in T cells (Meissner & Ritz-Timme, 2010; Zubakov *et al.*, 2010). The mtDNA deletions occur because they are more susceptible to the harmful action of free radicals released during the oxidative phosphorylation process, these deletions occur randomly and,

as they replicate, promote clonal expansion, this method is seen as a potentially useful method of predicting an individual's age. However, this method has some limitations that do not make it very viable to be applied in forensic sciences, for example the great heterogeneity in terms of abundance of this type of deletion in different tissues and the fact that the method has a margin of error of around 40 years (Meissner & Ritz-Timme, 2010).

Nowadays whilst other genetic tests have been used to infer lifetime age, those based on the analysis of DNA methylation patterns, epigenetic markers, have been shown to be the most reliable. DNA methylation consists on the presence of a specific 'methyl' molecule on the DNA that regulates gene expression, the methylation pattern of an individual changes with age, that characteristic is explore by the investigators (Rhein *et al.*, 2015; Vidaki & Kayser, 2017, 2018).

Using this method is possible to estimate the age of an individual from biological samples of diverse origins and various situations, being from human remains or collected from a crime scene with high accuracy with seven markers (Hong *et al.*, 2017).

1.2.3 Biogeographical ancestry

Biogeographical ancestry ("ancestry") prediction is described as the estimation of the geographical origin of a person's biological ancestors based on DNA analysis. Unable to test ancestral populations, populations currently living in these locations are used as a model (Royal *et al.*, 2010; Samuel & Prainsack, 2018).

Interpopulation phenotypic diversity is also reflected at the molecular level. The inference of physical characteristics from Biogeographic Ancestry is, therefore, a form of indirect inference since the information is obtained from Informative Ancestry Markers (AIM) and not from genes directly responsible for the characteristics (Frudakis, 2008; Phillips, 2015). Ancestry prediction using DNA is based on the understanding that a certain amount of genome variation exists between populations that have originated from different geographical locations around the world and the genetic markers used can be found in autosomal chromosomes as well as in the sex chromosomes being more accurate if using both, these markers can be SNPs, STRs or Indels (Samuel & Prainsack, 2018).

However, information about ancestry cannot be used solely as a criteria for determining the appearance of an individual mainly because the percentage of an individual's ancestral contribution will not necessarily reflect their appearance.

Inside the biogeographical ancestry inference is a subcategory that is lineage ancestry testing, whose mainly objective is to give information about an individual's paternal or maternal lineage. In the case of paternal lineage testing the Y-chromosome is what is tested, which are only present in males being passed down unchanged through a paternal line, especially useful when applied in familial searching, where Y-chromosomes of a range of males can be tested to identify if they are related to the suspect. Talking about maternal lineage testing the mitochondrial DNA is tested, and as this mtDNA is passed from mother to child, analysing genetic markers on this DNA can provide information about maternal lineage ancestry (Samuel & Prainsack, 2018).

1.2.4 Appearance inference

Appearance inference refers to the genetic analysis of an unknown individual's DNA sample to probabilistically predict that individual's appearance, their externally visible characteristics (EVC). Normally tests to predict appearance traits involve SNP testing (Samuel & Prainsack, 2018).

1.2.4.1 Pigmentation related traits

As a trait with a strong hereditary component and easily classifiable, pigmentation related traits have been the preferred target of genetic research over the last decades, being most of these traits mainly influenced by the amount and type of melanin present (Frudakis, 2008; Kayser, 2015).

Melanin is a molecule that has the ability to absorb light and has a high resistance to chemical degradation. There are two types of melanin pigment, eumelanin, brown / black and pheomelanin, yellow / red and the differences between these two types are based on the molecular structure (Sturm, 2009). Pigmentation is influenced by size, number, distribution and rate of formation of melanosomes as well as the type of melanin present in them. (Frudakis, 2008). That said, over the years the phenotypic diversity observed in pigmentation

related traits has been linked to genes whose products intervene in the melanogenesis process (Sturm, 2009).

1.2.4.1.1 Eye Colour

Eye colour can be considered as one of the human traits with the most colour variability, being defined by the amount of melanin and number of melanosomes in the outer layer of the iris, for example blue eyes have less melanin than brown eyes (Chaitanya *et al.*, 2014; Pośpiech *et al.*, 2016; S. Walsh, Liu, *et al.*, 2011). Eyes whose colours are more brownish have a higher concentration of eumelanin, green eyes are characterized by the presence of more pheomelanin and blue tones are determined by the low concentration of both eumelanin and pheomelanin (Frudakis, 2008).

One of the first phenotyping tools developed was the Irisplex consisting of six SNPs distributed among pigmentation genes, HERC2, OCA2, SLC24A4, SLC45A2, TYR, and IRF4. Depending on the eye colour there is a different accuracy percentage but for blue and brown eyes was higher than 90%, being tested in homogeneous and admixed populations, however for Asian populations did not show the same accuracy (Chaitanya *et al.*, 2014; Dembinski & Picard, 2014; Pośpiech *et al.*, 2016; S. Walsh, Lindenbergh, *et al.*, 2011; S. Walsh, Liu, *et al.*, 2011).

As previous mentioned depending on eye colour the accuracy varies being intermediate eye colours considered a problem, however studies have shown a gene–gene interaction between three of the main pigmentation genes, HERC2, OCA2, and TYRP1, related to green eye colour (Pośpiech *et al.*, 2011). To help in these cases of intermediate eye colour, some authors advocate the application of methods such as evaluating the iris melanin index based on the sum of the brightness values or quantifying the continuous colour variation of the eye colour in two axes corresponding to the hue values saturation as an approach that allows for a more objective and comprehensive assessment than direct observation based assessment as the latter allows only to differentiate between blue, brown and intermediate, a classification that does not include much of the phenotypic complexity, apart from the fact that some cases can be very difficult to classify. However, this approach is refute because they claim that the application of eye colour prediction in forensic practice implies the classification of this characteristic based on human interpretation of colours

rather than objective measurements, so the self-description approach becomes more relevant in this context (Frudakis, 2008; Kayser *et al.*, 2008).

One permanent question is if other factor can be associated with eye colour principally gender, mainly because was observed that women tend to have darker eyes than men in some European countries, however, at least for now, there is no scientific proof that justifies this observation (Martinez-Cadenas *et al.*, 2013).

1.2.4.1.2 Hair colour

Individuals with red hair display a relative increase in the amount of pheomelanin compared to eumelanin, whereas in dark hair the amount of eumelanin prevails and in blond hair there is little of each kind. The tone of the hair becomes darker as the size of the melanosomes and the density of the deposited pigment increases (Rees, 2003).

In white hair there is a marked decrease in the number of melanocytes, and persistent melanocytes show an extensive reduction in melanin content (Frudakis, 2008).

Among several genes involved in the melanogenesis process, MC1R was one of the first to demonstrate a strong discriminating power for red hair, fair skin, and freckles, after this was verified, in 2011, Branicki *et al.* made associations with other genes, for example SLC45A2, SLC24A5 and HERC2 and created a predictive model based on 22 SNPs, reaching 81%–93% accuracy for each hair colour. However, in 2013 was developed a new system, which basing itself in the preexisting Irisplex SNPs add 18 hair colour markers, in the genes MC1R, HERC2, OCA2, SLC45A2, KITLG, EXOC2, TYR, SLC24A4, IRF4, ASIP, and TYRP1, and created the HIrisplex System (S. Walsh *et al.*, 2013), despite the fact of this system having less markers than the previous one it can reach similar accuracy values.

Nowadays the bigger challenge to this approach is the accurate prediction of hair colours from individuals who have had hair colour changes throughout life, this represents one of the biggest differences between eye and hair colour, since the last one can change with time (Rees, 2003). The only reason that this is still a challenge comes to the fact that most studies do not contemplate the sampling of younger individuals, or question adult subjects about distinct phenotypes in early childhood. Therefore, prediction models are only elaborated with phenotypic information observed in adults, without taking into account informative markers for age-dependent phenotypes. To show the need to identify new

markers was developed a study with young individuals, and found that HIrisplex model incorrectly predicts hair phenotypes for those individuals who were blond only during early childhood (Kukla-Bartoszek *et al*., 2018).

1.2.4.1.3 Skin Colour

Skin colour represents one of the most complex pigmentation phenotypes studied, mainly because is believed that skin pigmentation variability emerged as an evolutionary response to the intensity of ultraviolet radiation (Jablonski & Chaplin, 2017). This factor makes the genotype/phenotype associations in studies difficult, as well as resulting in correlations that only apply to a specific population group.

Taking into account this evolutionary obstacle, a global prediction model was developed based on 36 markers distributed among 16 pigmentation genes. This model was created taking into account three or five skin tones. For the three skin tones were considered light, dark, dark–black obtaining prediction accuracies ranging from 83%–97%, for the five skin tones were considered very pale, pale, intermediate, dark, dark–black and were obtain accuracy levels of 72%–97%. Some of these associations have previously been described in admixed populations for some of these genes making these seemingly promising for future applications (Chaitanya *et al.*, 2018).

The results from the IrisPlex, HIrisPlex, and HIrisPlex-S systems were compiled into a public and free available interactive tool used to predict eye, hair, and skin colour from DNA data, to use the tool is only need to insert genotype data from the 41 markers available and obtain probabilities for three eye, four hair, and five skin colour categories (*HIRISPLEX-S, HIRISPLEX & IRISPLEX Eye, Hair and Skin colour DNA Phenotyping webtool USER MANUAL*, 2018).

1.2.4.2 Height

Human height have more studies surrounding it after 2008, until that year only a few genes have been described, since then almost 700 markers were described (Wood *et al*., 2014). Most of these genes are involved with growth signalling pathways, such as the fibroblast growth factor, as well as genes expressed in important tissues such as the growth

plate, although many of these markers are not directly involved in human growth pathways (Guo *et al.*, 2018)

Even with significant increase in the number of height-related variants, there are still no significant values for prediction tests, having only reached the max of 75% of accuracy, demonstrating the large number of SNPs still to be discovered and how complex this trait may be. Moreover, human height is influenced by other aspects such as gestational, hormonal, and environmental factors mainly during childhood, for example lack of nutrition can lead to a delay in developmental represented by a lack of growth (Guo *et al.*, 2018; Liu *et al.*, 2014).

Given all these facts it is not foreseeable for the near future the use of any kit for determination of stature in the forensic context (Kayser, 2015; Kayser & Schneider, 2009).

1.2.4.3 Facial features

The facial shape prediction represents one of the major objectives when studying phenotyping, and is studied from the distances between facial landmarks, as nostrils width, lips width, distance between eyes and face height. Some of the genetic markers associated with facial features are first studied because some pathologies, being later correlated to craniofacial development and consequently linked to the normal variation of facial shape (Claes *et al*., 2014).

The two most relevant studies about the genes that determine the morphology of the human face were published in 2012 on by Liu *et al.* and Paternoster *et al.*. In the first mentioned studied were identified 5 candidate genes PAX3, PRDM16, TP63, C5orf50, and COL17A1 with association to different facial distances, being that the first three mentioned genes have been implicated in vertebrate craniofacial development and disease (Liu *et al*., 2012). The study accomplished by Paternoster *et al.*, 2012, who used a few less than 3800 children only identified PAX3 as the only gene with genome-wide significance. These studies demonstrate that facial morphology is a complex trait and there is a large number of DNA variants likely involved.

In 2014 was applied a more complex approach to facial phenotyping, that used SNPs from craniofacial candidate genes with large frequency differences between three

populations, US Americans, Brazilians and Cape Verdeans, this led to the identification of 24 SNPs from 20 genes with association to facial features (Claes *et al.*, 2014).

However there are very few studies available, and the ones available have a small accuracy, so currently is not enough DNA markers are available for practical FDP of the face, if in the future complete facial appearance will be predictable from crime scene DNA with a high-enough accuracy to allow individual identification this will be one of the most important accomplishments in the history of forensic sciences (Kayser, 2015).

1.2.4.4 Hair Morphology

Hair morphology is determined, on the one hand, by the shape of the hair follicle, straight follicles lead to straight hair strands while base-folding follicles lead to curly strands, and on the other hand, asymmetry in thickness the inner sheath of the hair root (Eriksson *et al.*, 2010). It's a highly inherited trait with substantial interpopulation variation, the European population is a prime example of this, as about 45% of individuals have straight hair, 40% wavy hair and 15% curly hair (Loussouarn *et al.*, 2007).

When talking about hair morphology it can't be forgotten the case of baldness where it is empirically known that male pattern of baldness or androgenic alopecia has a strong hereditary factor, displaying a heritability of ~80% (Nyholt *et al.*, 2003). The five SNPs, that have the best association values, having 76.2% accuracy are in the AR/EDA2R, EBF1, TARDBP, and HDAC9 genes reaching 86.4% if other 15 markers are added (rs1041668, rs6625163, rs6625150, rs962458, rs12007229, rs2180439, rs913063, rs1160312, rs6113491, rs6461387, rs6945541, rs7349332, rs4679955, rs9668810, and rs10502861), demonstrating that even low prediction markers can have high accuracy when added to stronger ones (Marcińska *et al.*, 2015).

1.2.5 Legal and ethical aspects

When seeing the various legal issues surrounding the sample collection from suspects for comparison, an advantage of DNA phenotyping is clearly seen, this is the fact of this approach being more focused on obtaining genetic profiles from crime scene samples, thus not harming dignity or integrity rights, in addition, as the name says, Externally Visible Characteristics are publicly available features and therefore would not involve privacy issues

(Kayser, 2015). However, this would represent that all those individuals who share the characteristics of a facial composite may be interviewed and required to donate samples for comparison, raising questions about harassment that certain groups with a determined physical characteristic could suffer, belonging to a group of suspects solely by their physical appearance. Therefore, prior to any application it should be created new legal and ethical regulations to preserve the integrity and intimacy of people involved in DNA phenotyping-based investigations (S. J. Walsh, 2004).

The classification of the human species into races or subspecies verified in the prediction of ancestry is one of the most controversial issues within the scientific community, and this division is often considered racist and a violation of the standards of scientific objectivity (Lee *et al.*, 2008).

1.2.6 Future Developments

In 2017 the VISible Attributes Through GEnomics (VISAGE) Consortium was created with the aim of establishing new scientific knowledge, to develop, validate, and implement analysis tools that allow the prediction of an individual's appearance from DNA samples for use in forensic routine (Samuel & Prainsack, 2018). Some of the results of this consortium were already evaluated by the European DNA Profiling (EDNAP) Group, whose aim is to test the reliability and consistency of new forensic DNA technologies, and in the case of the IrisPlex System had reproducibility between 21 laboratories, considering it to be successful (Chaitanya *et al.*, 2014).

One area that has been in great interest in the last years is forensic epigenetic, who is studying DNA methylations profile, which is expected in the future help in forensic phenotyping, being already used in age estimation, but with perspectives to help identify some life choices basing in the study of DNA, for example if the person smokes (**Figure 6**) (Vidaki & Kayser, 2017, 2018).

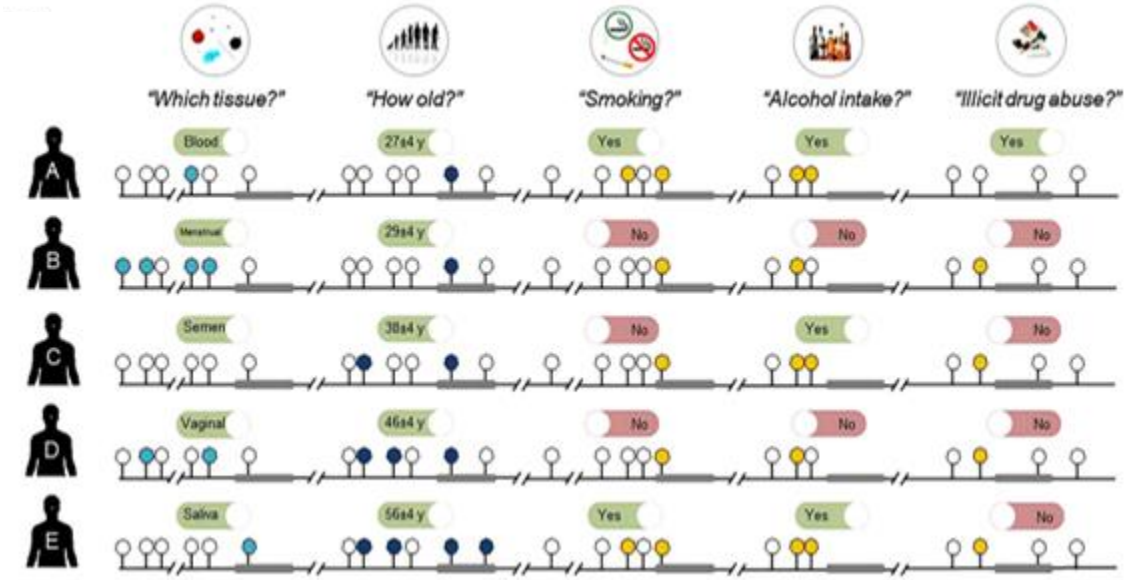

Figure 6 Future concept of epigenetic fingerprinting, where DNA methylation profiling can lead to the simultaneous prediction of a stain's tissue source and an individual's age and lifestyle choices. **Retrieved from:** (Vidaki & Kayser, 2018)

2. Drugs VS Pigmentation

The incidence of drug-induced pigmentation is very variable and depends on the involved medication, normally leading to an increase in pigmentation. Several mechanisms are involved in drug induced changes in pigmentation of the skin, for example, Tyndall effect, that is the optical changes in the refraction and scattering of incident light; alteration of the melanin deposition or combination with haemoglobin to form methaemoglobin (Dereure, 2001; Levantine & Almeyda, 1973).

The incidence of drug-induced pigmentation is difficult to ascertain because of a lack of reported cases and the fact that a lot of patients are on multiple drugs. However, estimates are that about 20% of cases of pigmentation are thought to be drug-induced. No significant differences between gender, age, and racial groups have been noted (Everling *et al.*, 2002).

Some drugs are well known to cause skin pigmentation; however, for other drugs, the appearance of hyperpigmentation is a rare event (Everling *et al.*, 2002).

2.1 Antimalarials

Hyperpigmentation is the most reported and common adverse effects of this category of drugs. This pigmentation tends to range from blueish grey to dark purple, and normally appear on the anterior side of the legs and on the head. Initially lesions occur isolated but with the progress and with the sun exposure they became large lesions.

In this case histology show an increase in quantity of melanin. This pigmentation is reversible after interruption of the treatment, but it takes some time after the interruption of the drug to the pigmentation comes back to normal (Dereure, 2001).

Associated with a specific antimalarial drug that is mainly used in lupus, mepacrine, it can occur even with small daily dosages a lemon-yellow discoloration usually affecting the whole tegument and simulating jaundice, as previous mentioned it resolves within a few months after the interruption of the treatment (Dereure, 2001).

2.2 NSAIDs

Nonsteroidal anti-inflammatory drugs (NSAIDs) is a drug class that reduce pain, decrease fever, prevent blood clots and decrease inflammation, for example aspirin and ibuprofen (Bally *et al.*, 2017). In this document will be consider Paracetamol as a member of this group despite the fact of some authors excluding it of this category because it has only a minor anti-inflammatory activity, however minor is still present and therefore being consider (Braasch *et al.*, 2007; Xia *et al.*, 2017).

The occurrence of pigmented lesions in this category usually result in the fixed drug eruption lesions and is most commonly associated with acetaminophen (paracetamol). Is hypothesized that this happens due to the drug binding to a melanocyte-linked protein and the melanocyte subsequently becoming the target of a specific cytotoxic reaction (Dereure, 2001).

2.3 Psychotropic drugs

The long term use of both phenothiazines or tricyclic antidepressants may induce a pigmentation, usually on sun-exposed areas. In similarity to antimalarials the pigmentation tends to be violet or purple-grey mainly on the face and the extremities, tending to appear progressively after a long period of time. Histological examination displays pigment granules particularly arranged around superficial capillaries and they shown the staining properties of melanin and ultrastructural studies have demonstrated that the pigment was probably complexes of the drug itself, which may bind to melanin, either eumelanin or phemomelanin (Dereure, 2001; Everling *et al*., 2002).

2.4 Anti-arrhythmic drugs

In this category generally the injuries have a distinctive blue-grey or purple discoloration on sun-exposed areas, especially the face with prominent involvement of the nose and ears. Corneal pigmentation may also happen and occur very early in treatment; however the other lesions tend to appear as a long term effect. In similarity to the previous

categories is a temporary effect, however may take up to one year to the lesions disappear (Dereure, 2001).

2.5 Cytotoxic drugs

When talking about cytotoxic drugs one of the most obvious medicines in this category are the chemotherapeutic agents used in cancer treatment, which have been know to cause hyperpigmentation in some cases. It may affect all parts of tegument including hair, nails and mucous membranes, be either diffuse or localized and sometimes have a distinctive pattern. May involve mechanisms as a direct toxic effect on melanocytes with subsequent melanin synthesis stimulation; hypersecretion of corticotropin and melanocyte-stimulating hormone as a result of adrenal toxicity; deficiency in tyrosinase inhibitors; formation of stable drug-melanin complexes.

In some cases, it can occur discoloration after a very variable interval of time following the start of treatment depending on the drug in question and may range from 1 week to several months. The pigmentation usually fades, at least partially, when the inducing agent is stopped but it may persist for a long time after the treatment is discontinued. In rare cases, the lesions are permanent (Dereure, 2001).

2.6 Tetracyclines

Minocycline, one of the drugs that most commonly has pigmentation reported as a side effect, induced hyperpigmentation occur especially in long-duration treatments. Some risk factors have been identified, mainly the duration of treatment, the cumulative dose, the presence of previous skin alterations related to inflammation or excessive sun exposure.

This hyperpigmentation cases have been classified into four basic clinical (i) dark blue-black macules in areas of acne scars or at sites of previous cutaneous inflammation; (ii) localized or diffuse hyperpigmented macules distant from site of inflammation or infection for which the antibacterial has been administered, and affecting mostly the anterior sides of lower legs or other sun-exposed areas; (iii) diffuse brown-grey discoloration known as the ‘muddy skin syndrome’ with a tendency to photo-aggravation; and (iv) hyperpigmentation of the vermilion area of lower lip (Simons & Morales, 1980).

The pathomechanisms are not fully understood, however may involve hyperproduction of melanin, especially on inflammatory or sun-exposed zones, by a direct effect of the antibacterial on the melanocytes (Dereure, 2001).

2.7 Metals

Heavy metal induced hyperpigmentation is a frequent event however with the discontinuation of the prescription of this kind of drugs, there are less reports of their effects being made.

Occupational exposure may also be involved in patients exposed to industrial fumes containing metals. It is of interest that this pigmentation, that may be very similar to gold-induced changes, may also affect mucous membranes, sclera and nails contrasting with drug induce pigmentation where the pigmentation is strictly limited to the skin. The occurrence of this pigmentation is usually slow, insidious after years of treatment.

After discontinuation of the medication, a very slow improvement may be observed but a residual pigmentation is often observed (Dereure, 2001).

2.8 Drugs as environmental contaminants

Pharmaceutical drugs are an emerging class of environmental contaminants widely used in human and veterinary medicine, where after significant administration is eliminated into the environment, contaminating groundwater, surface water and soil. High consumption by the population leads to the continuous entry of residual quantities into the environment, leading aquatic organisms to multigenerational exposure. The environmental impact that results from their presence is a growing public health concern worldwide as they have a biological effect. Its presence in the environment has been known since the 1970s and the presence of drug residues in the environment has been described in the scientific literature worldwide, although the results are still insufficient to establish a risk for the Human being (Jones *et al*., 2003; Mompelat *et al.*, 2009).

Most of the previous mentioned drugs have already been found in ground water mainly paracetamol, diclofenac and ibuprofeno, probably this three are encountered in more

concentrations due to the fact that are the most consumed drug, not needing a prescription so for that reason been easily accessed (Fortunato, 2014).

3. Practical proposal

The rapid growth of populations and technological/ medical advancement has resulted in innumerable pollutants and environmental toxin exposure, which for the bigger part there is no information of consequences of exposure, and no understanding of the biologic, chemical, and genetic mechanisms that underlie possible effects. When it comes to medical drugs the problem remains and even though there is a growing interest in the study of the consequences of pharmaceutical drugs on the environment, there is no certainty of the consequences to the human being, mainly because in normal conditions we don't are exposed to only one but multiple, not to mention that all studies are done using model animals, which will never be a one hundred percent portrait of the human race (Landrigan *et al.*, 2016).

This proposal intends to using a model animal, infer the possible consequences of some pharmaceutical drugs in the pigmentation, and if the changes occur understand if they are temporary or permanent and if they have influence in the genes associated with pigmentation.

Nowadays one of the most use model animal is the zebrafish, *Danio Rerio*, mainly because it provides a unique, in vivo whole vertebrate model, but less expensive than rodents. This species has the benefit of the large population size of offspring, cheap maintenance and easy to manipulate providing the opportunity to study the association of exposures with long-term outcomes in a vertebrate (Bambino & Chu, 2017). Furthermore zebrafish have a high degree of genetic conservation and their morphological and molecular basis of tissue and organ development is either identical or similar to other vertebrates including humans (He *et al.*, 2013). When it comes to pigmentation mechanisms is no difference and the genetic basis of pigment cell development and differentiation is largely conserved between mammals and teleosts, infraclass of bony fishes where zebrafish belongs, having genes such as SLC24A5, TYR and OCA2 have been identified to be involved in the pigmentation in both of human and zebrafish (Braasch *et al.*, 2007). For every fact previous mentioned in this proposal the zebrafish is the model organism chosen.

Pigment cells in zebrafish and other vertebrates are derived from neural crest cells that arise along the dorsal neural tube then disperse along stereotypical pathways throughout the embryo. In the specific case of zebrafish, pigment patterns are a result of the spatial

arrangements of three classes pigment cells: black melanocytes, yellow xanthophores and silver iridophores (Parichy *et al.*, 2000).

The zebrafish generate different pigment patterns during different phases of the life cycle, being the larval and adult pigment patterns different, the larva has a relatively simple pattern, that consists of several stripes of melanocytes and iridophores, as well as xanthophores that are widely distributed over the flank, giving an overall yellow cast to the body; it is possible to distinguish four stripes a dorsal band extending from the head to the tail along the dorsal apex of the myotomes; a lateral band at the level of the horizontal myoseptum; a ventral band, from between the eyes, over the dorsal yolk sac, and to the top of the tail; and a band over the ventral surface of the yolk sac. This pattern persists until 14 days, at which time a metamorphosis begins that ultimately results in the formation of the striped pigment pattern of the adult, by increasing the number of melanocytes and become visible dispersed throughout the skin in regions not previously occupied by these cells. Subsequently, between 21 and 28 days, melanocyte numbers increase gradually, and two melanocyte stripes emerge, dorsal and ventral to the horizontal myoseptum, with a light stripe in between. Later, over the course of weeks and months, additional "secondary" dark and light stripes are added dorsally and ventrally as the fish continue to grow, with 4–5 dark stripes present in the typical adult zebrafish. Dark stripes consist of melanocytes and iridophores, whereas light interstripe regions consist of xanthophores and iridophores. However the mechanisms of adult stripe development remain largely unknown in D. rerio (Parichy *et al.*, 2000; Quigley & Parichy, 2002) (**Figure 7**).

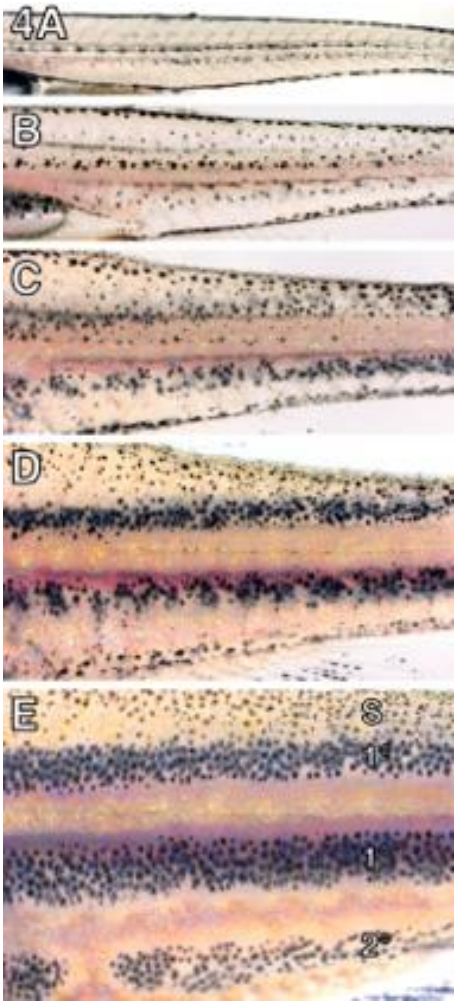

Figure 7 Metamorphosis of the zebrafish pigment pattern. **A**: The early larval pigment pattern persists for about 14 days. **B:** Begin of metamorphosis, melanophores arise dispersed over the flank. **C:** During a second phase of metamorphosis, additional melanophores begin to appear, giving the first illusion of the adult pattern. **D:** 1 month of development, two primary adult melanophore stripes border a lighter interstripe region. **E**: With continuous growth and development, primary (1°) melanophore stripes become more regular and are added secondary (2°) stripes. **Retrieved from:** (Quigley & Parichy, 2002).

Diclofenac, ibuprofeno and paracetamol are pharmaceutical drugs largely used by the word population, being in the case of portuguese population paracetamol the most used one (Fortunato, 2014). So in this proposal it would be suggested because of the relevance the use of paracetamol. In addition, previous work with zebrafish has found that at higher concentrations there is an appearance of individuals lacking pigmentation, which the authors mention, however, and similarly to other studies, was not studied further (David & Pancharatna, 2009).

For the concentrations they should be ecological relevant, concentrations already found in the nature and one higher as a possible prevision of the future.

In order to understand if the changes are temporary or not is suggested that after the exposure some of the individuals, are put in non-contaminant water so if the changes stop being visible it can be assumed that the changes are temporary.

As previous mentioned the pigmentation genes are highly conserved and for that reason the genetic study should centre around genes that have a high homology with human such as HERC2, OCA2 and TYR, that tend to be associated with pigmentation disorders as albinism (Beirl *et al.*, 2014; Wang *et al.*, 2007). If possible, to study any genomic change it should be studied gene expression and in case of doubt take samples to be sequenced, to have a more widely notion of the changes induced.

DNA extraction should be done using protocols specific to the life stage of zebrafish, in case of use of larvae is advisable the use of the protocol developed by Dupret *et al*., 2018 that was developed with the intent of combining genotypic with phenotypic analyses, being only use the tail of the fish to the DNA extraction, however this protocol it was only applied in larval stages so if intending to use an adult zebrafish, once the body has a higher index of fat is probable that the DNA extracted has a low purity index. As the genes of interest are involved in protein synthesis it his convenient to do a transcriptomic analysis, the RNA extraction can be accomplished using the protocol of Dupret *et al.*, 2018 , however this protocol is also applied in the larval stage so once more if intending to apply this approach in adult zebrafish prior testing and consequent optimization is needed.

To do the study of the pigmentation each individual should be photographed throughout the experiment so it's possible to have a notion of the development of pigmentation in each individual, towards the end of the experiment each individual should be carefully identified so it can be done an association between the phenotype and the genotype.

In general, with this approach, is intended, using zebrafish, to do a study to determine possible effects on pigmentation of pharmaceutical drugs present in the environment, suggesting paracetamol once is one of the most use drugs in Portugal. This approach will give the possibility to infer if that changes are temporary or permanent and if so if they have an effect on pigmentation genes, giving new insight into what constant exposure to these environmental pollutants can cause.

Conclusion

After this research done in the field of Forensic DNA phenotyping it is clear that obtaining a set of genetic markers that accurately predict most of human EVCs for forensic use is closer than ever and is not too far to actually be used in forensic routine investigations. The data still should be verified and applied in diverse global populations and check if none of the associations found are due to ancestry or other populational background. Despite the legal and ethical issues is still a viable approach for practical use in forensic routine, mainly because this issues are common to most approaches that involve DNA information.

When analysing the drugs that affect the pigmentation and the ones present on the environment is clear that we are exposed to a variety of chemicals even without knowing, and there is no certain idea of what might be the consequences of that exposure to pigmentation and to other aspects, even more considering that we are not only exposed to one compound but a variety of them. Currently are more and more investigators focusing on the possible effects of pharmaceutical drugs presents in the environment not only the consequences to the human being but also to other animals, mostly aquatic creatures, once they are the ones directly affected.

The practical proposal proposed in this document can help answer some relevant questions about this topic however, statistical studies in the human being are necessary to have an idea of what alterations we might already have suffer. Is of an interest to do more research into multi-gyrational and multifactorial effects, the last one being special relevant once it represents what actually happens in the natural context.

In conclusion to this work it’s important to mention that are very few studies on this area mainly because the ones who focus on drugs effects on organisms normally are more related to pathologies being the pigmentation a characteristic that is easily discarded. For that reason, more studies on the area are needed.

The present work, by its eminently descriptive, comparative and critical nature, is intended as a bibliographic reference for future research and / or academic works related.

References

Bally, M., Dendukuri, N., Rich, B., Nadeau, L., Helin-Salmivaara, A., Garbe, E., & Brophy, J. M. (2017). Risk of acute myocardial infarction with NSAIDs in real world use: Bayesian meta-analysis of individual patient data. *BMJ (Online)*, *357*(Cox 2). https://doi.org/10.1136/bmj.j1909

Bambino, K., & Chu, J. (2017). Zebrafish in Toxicology and Environmental Health. In *Current Topics in Developmental Biology* (1st ed., Vol. 124). https://doi.org/10.1016/bs.ctdb.2016.10.007

Beirl, A. J., Linbo, T. H., Cobb, M. J., & Cooper, C. D. (2014). oca2 regulation of chromatophore differentiation and number is cell type specific in zebrafish. https://doi.org/10.1111/pcmr.12205

Birren, B., Green, E. D., Klapholz, S., Myers, R. M., & Roskams, J. (1997). Genome Analysis – A Laboratory Manual. Cold Spring Harbor Laboratory Press.

Braasch, I., Schartl, M., & Volff, J. N. (2007). Evolution of pigment synthesis pathways by gene and genome duplication in fish. *BMC Evolutionary Biology*, *7*, 1–18. https://doi.org/10.1186/1471-2148-7-74

Branicki, W., Liu, F., Van Duijn, K., Draus-Barini, J., Pośpiech, E., Walsh, S., Kayser, M. (2011). Model-based prediction of human hair color using DNA variants. *Human Genetics*, *129*(4), 443–454. https://doi.org/10.1007/s00439-010-0939-8

Budowle, B., & Van Daal, A. (2008). Forensically relevant SNP classes. *BioTechniques*, *44*(5), 603–610. https://doi.org/10.2144/000112806

Buhler, S., & Sanchez-Mazas, A. (2011). HLA DNA sequence variation among human populations: Molecular signatures of demographic and selective events. *PLoS ONE*, *6*(2). https://doi.org/10.1371/journal.pone.0014643

Butler, E., & Li, R. (2014). Genetic Markers for Sex Identification in Forensic DNA Analysis. *Journal of Forensic Investigation*, *02*(03). https://doi.org/10.13188/2330-0396.1000013

Butler, J. M. (2009). Fundamentals of Forensic DNA Typing. In *Journal of Chemical Information and Modeling* (Vol. 53). https://doi.org/10.1017/CBO9781107415324.004

Cadenas, A. M., Regueiro, M., Gayden, T., Singh, N., Zhivotovsky, L. A., Underhill, P. A., & Herrera, R. J. (2007). Male amelogenin dropouts: phylogenetic context, origins and

implications. *Forensic Science International*, *166*(2–3), 155–163. https://doi.org/10.1016/j.forsciint.2006.05.002

Carracedo, A. (2013). Forensic Genetics: History. In *Encyclopedia of Forensic Sciences: Second Edition* (2nd ed.). https://doi.org/10.1016/B978-0-12-382165-2.00037-4

Cavalli-Sforza, L. L. (2005). The human genome diversity project: Past, present and future. *Nature Reviews Genetics*, *6*(4), 333–340. https://doi.org/10.1038/nrg1596

Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., Walsh, S. (2018). The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. In *Forensic Science International: Genetics* (Vol. 35). https://doi.org/10.1016/j.fsigen.2018.04.004

Chaitanya, L., Walsh, S., Andersen, J. D., Ansell, R., Ballantyne, K., Ballard, D., … Kayser, M. (2014). Collaborative EDNAP exercise on the IrisPlex system for DNA-based prediction of human eye colour. In *Forensic Science International: Genetics* (Vol. 11). https://doi.org/10.1016/j.fsigen.2014.04.006

Chial, H., & Craig, J. (2008). MtDNA and mitochondrial diseases. *Nature Education*.

Claes, P., Hill, H., & Shriver, M. D. (2014). Toward DNA-based facial composites: Preliminary results and validation. *Forensic Science International: Genetics*, *13*, 208–216. https://doi.org/10.1016/j.fsigen.2014.08.008

Codina, A. E., Niederstätter, H., & Parson, W. (2009). GenderPlex a PCR multiplex for reliable gender determination of degraded human DNA samples and complex gender constellations. *International Journal of Legal Medicine*, *123*(6), 459–464. https://doi.org/10.1007/s00414-008-0301-z

Collins, F. S., Brooks, L. D., & Chakravarti, A. (1999). A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research*, *9*(2), 210.

Committee on Identifying the Needs of the Forensic Sciences Community; National Research Council. (2009). Strengthening Forensic Science in the United States: A Path Forward.

Dale, J. W., & Schantz, M. von. (2002). From genes to genome: Concepts and applications of DNA Technology.

David, A., & Pancharatna, K. (2009). Effects of acetaminophen (paracetamol) in the embryonic development of zebrafish, Danio rerio. https://doi.org/10.1002/jat.1446

Dembinski, G. M., & Picard, C. J. (2014). Evaluation of the IrisPlex DNA-based eye color

prediction assay in a United States population. *Forensic Science International: Genetics*, *9*(1), 111–117. https://doi.org/10.1016/j.fsigen.2013.12.003

Dereure, O. (2001). Drug-Induced Skin Pigmentation Epidemiology , Diagnosis and Treatment. *2*(4), 253–262.

Dupret, B., Völkel, P., Follet, P., Le Bourhis, X., & Angrand, P. O. (2018). Combining genotypic and phenotypic analyses on single mutant zebrafish larvae. *MethodsX*, *5*, 244–256. https://doi.org/10.1016/j.mex.2018.03.002

Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics*, *5*(6), 435–445. https://doi.org/10.1038/nrg1348

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., … Mountain, J. (2010). Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genetics*, *6*(6), 1–20. https://doi.org/10.1371/journal.pgen.1000993

Everling, S., Tinsley, C. J., Gaffan, D., & Duncan, J. (2002). Drug-induced skin pigmentation not rare, but difficult to diagnose. *Drugs and Therapy Perspectives*, *18*(12), 10–13. https://doi.org/10.2165/00042310-200218120-00004

Fortunato, A. (2014). Ocorrência de Diclofenac , Ibuprofeno e Paracetamol em águas residuais em Portugal.

Frudakis, A. (2008). *MOLECULAR PHOTOFITTING: PREDICTING ANCESTRY AND PHENOTYPE USING DNA*.

Guo, M. H., Hirschhorn, J. N., & Dauber, A. (2018). Insights and Implications of Genome-Wide Association Studies of Height. *Journal of Clinical Endocrinology and Metabolism*, *103*(9), 3155–3168. https://doi.org/10.1210/jc.2018-01126

He, J. H., Guo, S. Y., Zhu, F., Zhu, J. J., Chen, Y. X., Huang, C. J., … Li, C. Q. (2013). A zebrafish phenotypic assay for assessing drug-induced hepatotoxicity. *Journal of Pharmacological and Toxicological Methods*, *67*(1), 25–32. https://doi.org/10.1016/j.vascn.2012.10.003

HIRISPLEX-S, HIRISPLEX & IRISPLEX Eye, Hair and Skin colour DNA Phenotyping webtool USER MANUAL. (2018). *0*, 0–22.

Hoffmann, A., & Spengler, D. (2018). The mitochondrion as potential interface in early-life stress brain programming. *Frontiers in Behavioral Neuroscience*, *12*(December), 1–19. https://doi.org/10.3389/fnbeh.2018.00306

Hong, S. R., Jung, S. E., Lee, E. H., Shin, K. J., Yang, W. I., & Lee, H. Y. (2017). DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers. *Forensic Science International: Genetics*, *29*, 118–125. https://doi.org/10.1016/j.fsigen.2017.04.006

Jablonski, N. G., & Chaplin, G. (2017). The colours of humanity: The evolution of pigmentation in the human lineage. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1724). https://doi.org/10.1098/rstb.2016.0349

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Hypervariable "minisatellite" regions in human DNA.

Jobling, M. A., & Tyler-Smith, C. (2003). The human Y chromosome: An evolutionary marker comes of age. *Nature Reviews Genetics*, *4*(8), 598–612. https://doi.org/10.1038/nrg1124

Jones, O. A. H., Voulvoulis, N., & Lester, J. N. (2003). Potential impact of pharmaceuticals on environmental health. *Bulletin of the World Health Organization*, *81*(10), 768–769.

Jorde, L. B. (2001). Population genomics: a bridge from evolutionary history to genetic medicine. *Human Molecular Genetics*, *10*(20), 2199–2207. https://doi.org/10.1093/hmg/10.20.2199

Kao, L., Tsai, L., Lee, J. C., & Hsieh, H. (2007). Controversial cases of human gender identification by amelogenin test. *Forensic Science Journalcience Journal*, *6*(2), 69–71.

Kayser, M. (2007). Uni-parental markers in human identity testing including forensic DNA analysis. *BioTechniques*, *43*(6), Sxv-Sxxi. https://doi.org/10.2144/000112580

Kayser, M. (2015). Forensic DNA Phenotyping : Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics*, *18*, 33–48. https://doi.org/10.1016/j.fsigen.2015.02.003

Kayser, M., & De Knijff, P. (2011). Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics*, *12*(3), 179–192. https://doi.org/10.1038/nrg2952

Kayser, M., Liu, F., Janssens, A. C. J. W., Rivadeneira, F., Lao, O., van Duijn, K., … van Duijn, C. M. (2008). Three Genome-wide Association Studies and a Linkage Analysis Identify HERC2 as a Human Iris Color Gene. *American Journal of Human Genetics*, *82*(2), 411–423. https://doi.org/10.1016/j.ajhg.2007.10.003

Kayser, M., & Schneider, P. M. (2009). DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. *Forensic Science International: Genetics*, *3*(3), 154–161. https://doi.org/10.1016/j.fsigen.2009.01.012

Kukla-Bartoszek, M., Pośpiech, E., Spólnicka, M., Karłowska-Pik, J., Strapagiel, D., Żądzińska, E., … Branicki, W. (2018). Investigating the impact of age-depended hair colour darkening during childhood on DNA-based hair colour prediction with the HIrisPlex system. *Forensic Science International: Genetics*, *36*, 26–33. https://doi.org/10.1016/j.fsigen.2018.06.007

Landrigan, P. J., Sly, J. L., Ruchirawat, M., Silva, E. R., Huo, X., Diaz-Barriga, F., … Sly, P. D. (2016). Health Consequences of Environmental Exposures: Changing Global Patterns of Exposure and Disease. *Annals of Global Health*, *82*(1), 10–19. https://doi.org/10.1016/j.aogh.2016.01.005

Lee, S. S. J., Mountain, J., Koenig, B., Altman, R., Brown, M., Camarillo, A., … Underhill, P. (2008). The ethics of characterizing difference: Guiding principles on using racial categories in human genetics. *Genome Biology*, *9*(7), 7–10. https://doi.org/10.1186/gb-2008-9-7-404

Levantine, A., & Almeyda, J. (1973). Drug induced changes in pigmentation. *British Journal of Dermatology*, *89*(1), 105–112. https://doi.org/10.1111/j.1365-2133.1973.tb01932.x

Liu, F., Hendriks, A. E. J., Ralf, A., Boot, A. M., Benyi, E., Sävendahl, L., … Kayser, M. (2014). Common DNA variants predict tall stature in Europeans. *Human Genetics*, *133*(5), 587–597. https://doi.org/10.1007/s00439-013-1394-0

Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G., … Kayser, M. (2012). A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans. *PLoS Genetics*, *8*(9). https://doi.org/10.1371/journal.pgen.1002932

Loussouarn, G., Garcel, A. L., Lozano, I., Collaudin, C., Porter, C., Panhard, S., … de La Mettrie, R. (2007). Worldwide diversity of hair curliness: A new method of assessment. *International Journal of Dermatology*, *46*(SUPPL. 1), 2–6. https://doi.org/10.1111/j.1365-4632.2007.03453.x

Luo, S., Valencia, C. A., Zhang, J., Lee, N. C., Slone, J., Gui, B., … Huang, T. (2018). Biparental inheritance of mitochondrial DNA in humans. *Proceedings of the National*

Academy of Sciences of the United States of America, *115*(51), 13039–13044. https://doi.org/10.1073/pnas.1810946115

Marcińska, M., Pośpiech, E., Abidi, S., Andersen, J. D., Van Den Berge, M., Carracedo, Á., … Branicki, W. (2015). Evaluation of DNA variants associated with androgenetic alopecia and their potential to predict male pattern baldness. *PLoS ONE*, *10*(5), 1–18. https://doi.org/10.1371/journal.pone.0127852

Martinez-Cadenas, C., Penãa-Chilet, M., Ibarrola-Villava, M., & Ribas, G. (2013). Gender is a major factor explaining discrepancies in eye colour prediction based on HERC2/OCA2 genotype and the IrisPlex model. *Forensic Science International: Genetics*, *7*(4), 453–460. https://doi.org/10.1016/j.fsigen.2013.03.007

Masuyama, K., Shojo, H., Nakanishi, H., Inokuchi, S., & Adachi, N. (2017). Sex determination from fragmented and degenerated DNA by amplified product-length polymorphism bidirectional SNP analysis of amelogenin and SRY genes. *PLoS ONE*, *12*(1), 1–17. https://doi.org/10.1371/journal.pone.0169348

Mcwilliams, T. G., & Suomalainen, A. (2019). Fate of a father's mitochondria. *Nature*, *565*, 296–297. https://doi.org/10.1038/d41586-019-00093-1

Meissner, C., & Ritz-Timme, S. (2010). Molecular pathology and age estimation. *Forensic Science International*, *203*(1–3), 34–43. https://doi.org/10.1016/j.forsciint.2010.07.010

Mompelat, S., Le Bot, B., & Thomas, O. (2009). Occurrence and fate of pharmaceutical products and by-products, from resource to drinking water. *Environment International*, *35*(5), 803–814. https://doi.org/10.1016/j.envint.2008.10.008

Mullaney, J. M., Mills, R. E., Stephen Pittard, W., & Devine, S. E. (2010). Small insertions and deletions (INDELs) in human genomes. *Human Molecular Genetics*, *19*(R2), 131–136. https://doi.org/10.1093/hmg/ddq400

Nyholt, D. R., Gillespie, N. A., Heath, A. C., & Martin, N. G. (2003). Genetic Basis of Male Pattern Baldness. *Journal of Investigative Dermatology*, *121*(6), 1561–1564. https://doi.org/10.1111/j.1523-1747.2003.12615.x

Parichy, D. M., Ransom, D. G., Paw, B., Zon, L. I., & Johnson, S. L. (2000). An orthologue of the kit-related gene fms is required for development of neural crest-derived xanthophores and a subpopulation of adult melanocytes in the zebrafish, Danio rerio. *Development*, *127*(14), 3031–3044.

Parker, G. J., Yip, J. M., Eerkens, J. W., Salemi, M., Durbin-Johnson, B., Kiesow, C., …

Phinney, B. S. (2019). Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel. *Journal of Archaeological Science*, *101*(May 2018), 169–180. https://doi.org/10.1016/j.jas.2018.08.011

Paternoster, L., Zhurov, A. I., Toma, A. M., Kemp, J. P., St. Pourcain, B., Timpson, N. J., … Evans, D. M. (2012). Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *American Journal of Human Genetics*, *90*(3), 478–485. https://doi.org/10.1016/j.ajhg.2011.12.021

Pelt-Verkuil, E., Belkum, A. van, & Hays, J. P. (2008). *Principles and Technical Aspects of PCR Amplification*.

Phillips, C. (2015). Forensic genetic analysis of bio-geographical ancestry. *Forensic Science International: Genetics*, *18*, 49–65. https://doi.org/10.1016/j.fsigen.2015.05.012

Pośpiech, E., Draus-Barini, J., Kupiec, T., Wojas-Pelc, A., & Branicki, W. (2011). Gene-gene interactions contribute to eye colour variation in humans. *Journal of Human Genetics*, *56*(6), 447–455. https://doi.org/10.1038/jhg.2011.38

Pośpiech, E., Karłowska-Pik, J., Ziemkiewicz, B., Kukla, M., Skowron, M., Wojas-Pelc, A., & Branicki, W. (2016). Further evidence for population specific differences in the effect of DNA markers and gender on eye colour prediction in forensics. *International Journal of Legal Medicine*, *130*(4), 923–934. https://doi.org/10.1007/s00414-016-1388-2

Pulker, H., Lareu, M. V., Phillips, C., & Carracedo, A. (2007). Finding genes that underlie physical traits of forensic interest using genetic tools. *Forensic Science International: Genetics*, *1*(2), 100–104. https://doi.org/10.1016/j.fsigen.2007.02.009

Quigley, I. K., & Parichy, D. M. (2002). Pigment Pattern Formation in Zebrafish : A Model for Developmental Genetics and the Evolution of Form. *455*(February), 442–455. https://doi.org/10.1002/jemt.10162

Rees, J. L. (2003). Genetics of Hair and Skin Color. *Annual Review of Genetics*, *37*(1), 67–90. https://doi.org/10.1146/annurev.genet.37.110801.143233

Rhein, M., Hagemeier, L., Klintschar, M., Muschler, M., Bleich, S., & Frieling, H. (2015). DNA methylation results depend on DNA integrity-role of post mortem interval. *Frontiers in Genetics*, *6*(MAY), 1–7. https://doi.org/10.3389/fgene.2015.00182

Royal, C. D., Novembre, J., Fullerton, S. M., Goldstein, D. B., Long, J. C., Bamshad, M. J.,

& Clark, A. G. (2010). Inferring Genetic Ancestry: Opportunities, Challenges, and Implications. *American Journal of Human Genetics*, *86*(5), 661–673. https://doi.org/10.1016/j.ajhg.2010.03.011

Samuel, G., & Prainsack, B. (2018). The regulatory landscape of forensic DNA phenotyping in Europe. 119. Retrieved from http://www.visage-h2020.eu/Report_regulatory_landscape_FDP_in_Europe2.pdf

Santos, F., Machado, H., & Silva, S. (2013). Forensic DNA databases in European countries : is size linked to performance ? (Cutter 2006), 1–13.

Simons, J. J., & Morales, A. (1980). Minocycline and generalized cutaneous pigmentation. *Journal of the American Academy of Dermatology*, *3*(3), 244–247. https://doi.org/10.1016/S0190-9622(80)80186-1

Sturm, R. A. (2009). Molecular genetics of human pigmentation diversity. *Human Molecular Genetics*, *18*(R1), 9–17. https://doi.org/10.1093/hmg/ddp003

Szibor, R. (2007). X-chromosomal markers: Past, present and future. *Forensic Science International: Genetics*, *1*(2), 93–99. https://doi.org/10.1016/j.fsigen.2007.03.003

Teama, S. (2016). DNA Polymorphisms: DNA-Based Molecular Markers and Their Application in Medicine. *Intech*, (tourism), 13. https://doi.org/http://dx.doi.org/10.5772/57353

Trent, R. J. (2012). MOLECULAR MEDICINE Genomics to Personalized Healthcare. https://doi.org/10.1016/B978-0-12-381451-7.00011-6

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, *30*(2), 386–394. https://doi.org/10.1002/humu.20921

Vidaki, A., & Kayser, M. (2017). From forensic epigenetics to forensic epigenomics: broadening DNA investigative intelligence. 1–13. https://doi.org/10.1186/s13059-017-1373-1

Vidaki, A., & Kayser, M. (2018). Recent progress, methods and perspectives in forensic epigenetics. *Forensic Science International: Genetics*, *37*(July), 180–195. https://doi.org/10.1016/j.fsigen.2018.08.008

von Wurmb-Schwark, N., Bosinski, H., & Ritz-Timme, S. (2007). What do the X and Y chromosomes tell us about sex and gender in forensic case analysis? *Journal of Forensic and Legal Medicine*, *14*(1), 27–30.

https://doi.org/10.1016/j.jcfm.2005.09.003

Walsh, S. J. (2004). Recent advances in forensic genetics. *Expert Review of Molecular Diagnostics*, *4*(1), 31–40. https://doi.org/10.1586/14737159.4.1.31

Walsh, S., Lindenbergh, A., Zuniga, S. B., Sijen, T., De Knijff, P., Kayser, M., & Ballantyne, K. N. (2011). Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics*, *5*(5), 464–471. https://doi.org/10.1016/j.fsigen.2010.09.008

Walsh, S., Liu, F., Ballantyne, K. N., Van Oven, M., Lao, O., & Kayser, M. (2011). IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics*, *5*(3), 170–180. https://doi.org/10.1016/j.fsigen.2010.02.004

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-kamysz, A., … Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics*, *7*(1), 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005

Wang, J., Hou, L., Zhang, R., Zhao, X., Jiang, L., Sun, W., … Li, X. (2007). The tyrosinase gene family and albinism in fish. *Chinese Journal of Oceanology and Limnology*, *25*(2), 191–198. https://doi.org/10.1007/s00343-007-0191-9

Wirtz, J., Rauscher, M., & Wiehe, T. (2018). Topological linkage disequilibrium calculated from coalescent genealogies. *Theoretical Population Biology*, *124*(xxxx), 41–50. https://doi.org/10.1016/j.tpb.2018.09.001

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., … Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, *46*(11), 1173–1186. https://doi.org/10.1038/ng.3097

Xia, L., Zheng, L., & Zhou, J. L. (2017). Effects of ibuprofen , diclofenac and paracetamol on hatch and motor behavior in developing zebra fish (Danio rerio). *Chemosphere*, *182*, 416–425. https://doi.org/10.1016/j.chemosphere.2017.05.054

Zubakov, D., Liu, F., Van Zelm, M. C., Vermeulen, J., Oostra, B. A., Van Duijn, C. M., … Langerak, A. W. (2010). Estimating human age from T-cell DNA rearrangements. *Current Biology*, *20*(22), R970–R971. https://doi.org/10.1016/j.cub.2010.10.022

Appendix - article proposal

Forensic DNA phenotyping: current applications

Inês Cardoso[a]; Luís Souto[a]; Helena Moreira[a]

[a] University of Aveiro

Abstract: DNA analysis in forensic investigations is able to identify the donor of a biological by direct comparison of short tandem repeats genetic profile, however one of the major limitations of this approach is the need for a reference sample for comparison. This limitation led to the increase of studies seeking to understand the relation between certain polymorphisms and certain phenotypic characteristics, this process of inferring externally visible characteristics from biological samples in the forensic contexto is known as forensic DNA phenotyping. With the development of this technology some ethical and legal aspects must be taken into consideration. Despite this, this technology can help to orientate investigations where is need to identify both suspects and victims, without needing a profile to compare it with.

Keywords: DNA phenotyping, externally visible characteristics, pigmentation, genetic markers, forensic genetics

Introduction

During an investigation several samples are taken, some of them finish in a forensic genetics service for donor identification, then the profile is determined and compared with other reference sample profiles (known donor samples, or profiles in the national database) for identification purposes, however, there are many cases where identification is not established because there is no match. This genetic profile obtained from DNA polymorphisms is based on STRs, also known as Short Tandem Repeats and consist of sequences in which one to six nucleotides are repeated and represent about 3% of the human genome being randomly distributed (Ellegren, 2004; Kayser & Schneider, 2009).

The major advantage of STR marker is their high allele diversity which make them highly informative, once that except for identical twins each individual on the world may have a unique genetic profile, however this analysis are of a comparative nature, requiring reference samples to be compared, being this the major disadvantage of the use of this

markers (Ellegren, 2004; Butler, 2009). Although DNA profile databases make this approach more effective, depending on the laws of the countries, these databases may be incomplete, with a low number of individuals, thus reducing the likelihood of a match (Kayser, 2015; Santos *et al.*, 2013).

The comparative nature of DNA analysis led to the relatively recent development of a new method in forensic genetics the Forensic DNA Phenotyping (FDP) whose objective is inferring the visible external characteristics directly from the biological material left at the scene or obtained of unknown bodies (Kayser, 2015).

This new methodology is based on the study of SNPs, Single Nucleotide Polymorphisms that is DNA sequence variations corresponding to differences in a single nucleotide, in this case this variations are located in DNA coding or regulatory regions can lead to amino acid substitutions, altering the functional properties of the translated protein and consequently being expressed in distinct phenotypes, some of them being the visible characteristics of the individual (Dale & Schantz, 2002; Budowle & Van Daal, 2008).

The aim of Forensic DNA Phenotyping is to function as “biological witness” and potentially provide more accurate information than human eyewitnesses, and assist investigations where individuals are not identifiable by the conventional comparative DNA profile, which include cases of identification of missing persons whose ante-mortem DNA is not available (Kayser, 2015; Kayser & De Knijff, 2011; Kayser & Schneider, 2009). However, unlike the identification profile, it is not intended to identify the donor of the sample, but to limit the number of suspects to be considered (Kayser & Schneider, 2009).

Determining biogeographic ancestry from DNA is sometimes considered part of the Forensic DNA Phenotyping, however genetic ancestry does not always represent externally visible and accurate characteristics, for example the case of individuals of mixed descent (Kayser & De Knijff, 2011), for this reason when using biogeographic ancestry in phenotype prediction, it is considered an indirect prediction.

In order to proceed with phenotyping there are several types of approaches, for example spectrophotometric methods and digital photography analysis, in order to try to reduce the subjective evaluation as is the case of Frudakis, 2008 which objectively classify phenotypic characteristics such as melanin index rather than "skin tone". However, the Kayser group, known as VISAGE Consortium (VISible Attributes Through GEnomics), which is currently one of the most relevant group in this emerging area, normally uses

observational assessment methods, and argue that there is no evidence that quantitative assessment methods are better than methods based on self and heterodescription (Branicki *et al.*, 2011). After the results of phenotyping these results should always be viewed critically, once if it is a feature that can be influenced by environmental factors this must be taken into account, besides with the advancement of technology in aesthetics there are many ways to alter physical characteristics voluntarily such as hair dyeing, wearing contact lenses or even through surgical alteration (Frudakis, 2008).

1. DNA Phenotyping

1.1 Gender

Determining an individual's gender is one of the oldest approaches in the identification methods, and is achieved through analysis of several markers however the most used and with better results is the Amelogenin gene, whose proteins are expressed in the tooth and have an important role in enamel biosynthesis. Sexual diagnosis is achieved because the Amelogenin gene have differently sized copies on X chromosomes and Y chromosomes (Parker *et al.*, 2019). The most used set of primers target a 6-bp insertion/deletion within an intron of the Amelogenin gene on the X and Y chromosomes and produce 106-bp and 112-bp amplicons for the X and Y chromosomes, respectively (E. Butler & Li, 2014; Codina *et al.*, 2009; Masuyama *et al.*, 2017).

Currently this marker is already included in most forensic identification tests, however it is not error free, being most of these errors due to a deletion of Y chromosome or a point mutation in the X chromosome region to which the reverse Amelogenin gene primer binds in several commercial identification kits (Cadenas *et al.*, 2007; Kao *et al.*, 2007). The best way around these errors is to interpret the results carefully and if possible add other markers whose location is not close to the amelogenin or even Y chromosome STR markers (Kayser & Schneider, 2009).

There are other special circumstances that can also lead to error, for example, in cases of bone marrow transplants in which the donor and recipient have different genders, also called chimerism, other example is micro-chimerism, which occurs when a woman is pregnant of a male fetus. Another topic that cannot be overlooked especially nowadays is the difference between biological and legal gender present in the identification documents observed in cases of intersexuality and transsexuality. In all these situations mentioned above the observed phenotype and/or legal records may indicate a gender contrary to the

biologically tested which makes the analysis of results more complex (von Wurmb-Schwark *et al*., 2007).

1.2 Age inference

Age inference refers to the analysis of an unknown individual's DNA sample to probabilistically infer that individual's age (Samuel & Prainsack, 2018).

With aging the changes in DNA are for example the accumulation of mtDNA deletions, the decrease in telomere size and the decrease in the number of circular DNA molecules in T cells (Meissner & Ritz-Timme, 2010; Zubakov *et al.*, 2010). The mtDNA deletions occur because they are more susceptible to the harmful action of free radicals released during the oxidative phosphorylation process, these deletions occur randomly and, as they replicate, promote clonal expansion, this method is seen as a potentially useful method of predicting an individual's age. However, this method has some limitations that do not make it very viable to be applied in forensic sciences, for example the great heterogeneity in terms of abundance of this type of deletion in different tissues and the fact that the method has a margin of error of around 40 years (Meissner & Ritz-Timme, 2010).

Nowadays whilst other genetic tests have been used to infer lifetime age, those based on the analysis of DNA methylation patterns, epigenetic markers, have been shown to be the most reliable. DNA methylation consists on the presence of a specific 'methyl' molecule on the DNA that regulates gene expression, the methylation pattern of an individual changes with age, that characteristic is explore by the investigators (Rhein *et al.*, 2015; Vidaki & Kayser, 2017, 2018).

Using this method is possible to estimate the age of an individual from biological samples of diverse origins and various situations, being from human remains or collected from a crime scene with high accuracy with seven markers (Hong *et al.*, 2017).

1.3 Biogeographical ancestry

Biogeographical ancestry ("ancestry") prediction is described as the estimation of the geographical origin of a person's biological ancestors based on DNA analysis. Unable to test ancestral populations, populations currently living in these locations are used as a model (Royal *et al.*, 2010; Samuel & Prainsack, 2018).

Interpopulation phenotypic diversity is also reflected at the molecular level. The inference of physical characteristics from Biogeographic Ancestry is, therefore, a form of indirect inference since the information is obtained from Informative Ancestry Markers

(AIM) and not from genes directly responsible for the characteristics (Frudakis, 2008; Phillips, 2015). Ancestry prediction using DNA is based on the understanding that a certain amount of genome variation exists between populations that have originated from different geographical locations around the world and the genetic markers used can be found in autosomal chromosomes as well as in the sex chromosomes being more accurate if using both, these markers can be SNPs, STRs or Indels (Samuel & Prainsack, 2018).

However, information about ancestry cannot be used solely as a criteria for determining the appearance of an individual mainly because the percentage of an individual's ancestral contribution will not necessarily reflect their appearance.

Inside the biogeographical ancestry inference is a subcategory that is lineage ancestry testing, whose mainly objective is to give information about an individual's paternal or maternal lineage. In the case of paternal lineage testing the Y-chromosome is what is tested, which are only present in males being passed down unchanged through a paternal line, especially useful when applied in familial searching, where Y-chromosomes of a range of males can be tested to identify if they are related to the suspect. Talking about maternal lineage testing the mitochondrial DNA is tested, and as this mtDNA is passed from mother to child, analysing genetic markers on this DNA can provide information about maternal lineage ancestry (Samuel & Prainsack, 2018).

1.4 Pigmentation related traits

As a trait with a strong hereditary component and easily classifiable, pigmentation related traits have been the preferred target of genetic research over the last decades, being most of these traits mainly influenced by the amount and type of melanin present (Frudakis, 2008; Kayser, 2015).

Melanin is a molecule that has the ability to absorb light and has a high resistance to chemical degradation. There are two types of melanin pigment, eumelanin, brown / black and pheomelanin, yellow / red and the differences between these two types are based on the molecular structure (Sturm, 2009). Pigmentation is influenced by size, number, distribution and rate of formation of melanosomes as well as the type of melanin present in them. (Frudakis, 2008). That said, over the years the phenotypic diversity observed in pigmentation related traits has been linked to genes whose products intervene in the melanogenesis process (Sturm, 2009).

1.4.1 Eye Colour

Eye colour can be considered as one of the human traits with the most colour variability, being defined by the amount of melanin and number of melanosomes in the outer layer of the iris, for example blue eyes have less melanin than brown eyes (Chaitanya *et al.*, 2014; Pośpiech *et al.*, 2016; S. Walsh, Liu, *et al.*, 2011). Eyes whose colours are more brownish have a higher concentration of eumelanin, green eyes are characterized by the presence of more pheomelanin and blue tones are determined by the low concentration of both eumelanin and pheomelanin (Frudakis, 2008).

One of the first phenotyping tools developed was the Irisplex consisting of six SNPs distributed among pigmentation genes, HERC2, OCA2, SLC24A4, SLC45A2, TYR, and IRF4. Depending on the eye colour there is a different accuracy percentage but for blue and brown eyes was higher than 90%, being tested in homogeneous and admixed populations, however for Asian populations did not show the same accuracy (Chaitanya *et al*., 2014; Dembinski & Picard, 2014; Pośpiech *et al*., 2016; S. Walsh, Lindenbergh, *et al*., 2011; S. Walsh, Liu, *et al.*, 2011).

As previous mentioned depending on eye colour the accuracy varies being intermediate eye colours considered a problem, however studies have shown a gene–gene interaction between three of the main pigmentation genes, HERC2, OCA2, and TYRP1, related to green eye colour (Pośpiech *et al.*, 2011). To help in these cases of intermediate eye colour, some authors advocate the application of methods such as evaluating the iris melanin index based on the sum of the brightness values or quantifying the continuous colour variation of the eye colour in two axes corresponding to the hue values saturation as an approach that allows for a more objective and comprehensive assessment than direct observation based assessment as the latter allows only to differentiate between blue, brown and intermediate, a classification that does not include much of the phenotypic complexity, apart from the fact that some cases can be very difficult to classify. However, this approach is refute because they claim that the application of eye colour prediction in forensic practice implies the classification of this characteristic based on human interpretation of colours rather than objective measurements, so the self-description approach becomes more relevant in this context (Frudakis, 2008; Kayser *et al.*, 2008).

One permanent question is if other factor can be associated with eye colour principally gender, mainly because was observed that women tend to have darker eyes than

men in some European countries, however, at least for now, there is no scientific proof that justifies this observation (Martinez-Cadenas *et al.*, 2013).

1.4.2 Hair colour

Individuals with red hair display a relative increase in the amount of pheomelanin compared to eumelanin, whereas in dark hair the amount of eumelanin prevails and in blond hair there is little of each kind. The tone of the hair becomes darker as the size of the melanosomes and the density of the deposited pigment increases (Rees, 2003).

In white hair there is a marked decrease in the number of melanocytes, and persistent melanocytes show an extensive reduction in melanin content (Frudakis, 2008).

Among several genes involved in the melanogenesis process, MC1R was one of the first to demonstrate a strong discriminating power for red hair, fair skin, and freckles, after this was verified, in 2011, Branicki *et al.* made associations with other genes, for example SLC45A2, SLC24A5 and HERC2 and created a predictive model based on 22 SNPs, reaching 81%–93% accuracy for each hair colour. However, in 2013 was developed a new system, which basing itself in the preexisting Irisplex SNPs add 18 hair colour markers, in the genes MC1R, HERC2, OCA2, SLC45A2, KITLG, EXOC2, TYR, SLC24A4, IRF4, ASIP, and TYRP1, and created the HIrisplex System (S. Walsh et al., 2013), despite the fact of this system having less markers than the previous one it can reach similar accuracy values.

Nowadays the bigger challenge to this approach is the accurate prediction of hair colours from individuals who have had hair colour changes throughout life, this represents one of the biggest differences between eye and hair colour, since the last one can change with time (Rees, 2003). The only reason that this is still a challenge comes to the fact that most studies do not contemplate the sampling of younger individuals, or question adult subjects about distinct phenotypes in early childhood. Therefore, prediction models are only elaborated with phenotypic information observed in adults, without taking into account informative markers for age-dependent phenotypes. To show the need to identify new markers was developed a study with young individuals, and found that HIrisplex model incorrectly predicts hair phenotypes for those individuals who were blond only during early childhood (Kukla-Bartoszek *et al*., 2018).

1.4.3 Skin Colour

Skin colour represents one of the most complex pigmentation phenotypes studied, mainly because is believed that skin pigmentation variability emerged as an evolutionary

response to the intensity of ultraviolet radiation (Jablonski & Chaplin, 2017). This factor makes the genotype/phenotype associations in studies difficult, as well as resulting in correlations that only apply to a specific population group.

Taking into account this evolutionary obstacle, a global prediction model was developed based on 36 markers distributed among 16 pigmentation genes. This model was created taking into account three or five skin tones. For the three skin tones were considered light, dark, dark–black obtaining prediction accuracies ranging from 83%–97%, for the five skin tones were considered very pale, pale, intermediate, dark, dark–black and were obtain accuracy levels of 72%–97%. Some of these associations have previously been described in admixed populations for some of these genes making these seemingly promising for future applications (Chaitanya *et al.*, 2018).

The results from the IrisPlex, HIrisPlex, and HIrisPlex-S systems were compiled into a public and free available interactive tool used to predict eye, hair, and skin colour from DNA data, to use the tool is only need to insert genotype data from the 41 markers available and obtain probabilities for three eye, four hair, and five skin colour categories (*HIRISPLEX-S, HIRISPLEX & IRISPLEX Eye, Hair and Skin colour DNA Phenotyping webtool USER MANUAL*, 2018).

1.5 Height

Human height have more studies surrounding it after 2008, until that year only a few genes have been described, since then almost 700 markers were described (Wood *et al*., 2014). Most of these genes are involved with growth signalling pathways, such as the fibroblast growth factor, as well as genes expressed in important tissues such as the growth plate, although many of these markers are not directly involved in human growth pathways (Guo *et al.*, 2018)

Even with significant increase in the number of height-related variants, there are still no significant values for prediction tests, having only reached the max of 75% of accuracy, demonstrating the large number of SNPs still to be discovered and how complex this trait may be. Moreover, human height is influenced by other aspects such as gestational, hormonal, and environmental factors mainly during childhood, for example lack of nutrition can lead to a delay in developmental represented by a lack of growth (Guo *et al.*, 2018; Liu *et al.*, 2014).

Given all these facts it is not foreseeable for the near future the use of any kit for determination of stature in the forensic context (Kayser, 2015; Kayser & Schneider, 2009).

1.6 Facial features

The facial shape prediction represents one of the major objectives when studying phenotyping, and is studied from the distances between facial landmarks, as nostrils width, lips width, distance between eyes and face height. Some of the genetic markers associated with facial features are first studied because some pathologies, being later correlated to craniofacial development and consequently linked to the normal variation of facial shape (Claes *et al*., 2014).

The two most relevant studies about the genes that determine the morphology of the human face were published in 2012 on by Liu *et al.* and Paternoster *et al.*. In the first mentioned studied were identified 5 candidate genes PAX3, PRDM16, TP63, C5orf50, and COL17A1 with association to different facial distances, being that the first three mentioned genes have been implicated in vertebrate craniofacial development and disease (Liu *et al*., 2012). The study accomplished by Paternoster *et al.*, 2012, who used a few less than 3800 children only identified PAX3 as the only gene with genome-wide significance. These studies demonstrate that facial morphology is a complex trait and there is a large number of DNA variants likely involved.

In 2014 was applied a more complex approach to facial phenotyping, that used SNPs from craniofacial candidate genes with large frequency differences between three populations, US Americans, Brazilians and Cape Verdeans, this led to the identification of 24 SNPs from 20 genes with association to facial features (Claes *et al.*, 2014).

However there are very few studies available, and the ones available have a small accuracy, so currently is not enough DNA markers are available for practical FDP of the face, if in the future complete facial appearance will be predictable from crime scene DNA with a high-enough accuracy to allow individual identification this will be one of the most important accomplishments in the history of forensic sciences (Kayser, 2015).

1.7 Hair Morphology

Hair morphology is determined, on the one hand, by the shape of the hair follicle, straight follicles lead to straight hair strands while base-folding follicles lead to curly strands, and on the other hand, asymmetry in thickness the inner sheath of the hair root (Eriksson *et al.*, 2010). It's a highly inherited trait with substantial interpopulation variation, the

European population is a prime example of this, as about 45% of individuals have straight hair, 40% wavy hair and 15% curly hair (Loussouarn *et al.*, 2007).

When talking about hair morphology it can't be forgotten the case of baldness where it is empirically known that male pattern of baldness or androgenic alopecia has a strong hereditary factor, displaying a heritability of ~80% (Nyholt *et al*., 2003). The five SNPs, that have the best association values, having 76.2% accuracy are in the AR/EDA2R, EBF1, TARDBP, and HDAC9 genes reaching 86.4% if other 15 markers are added (rs1041668, rs6625163, rs6625150, rs962458, rs12007229, rs2180439, rs913063, rs1160312, rs6113491, rs6461387, rs6945541, rs7349332, rs4679955, rs9668810, and rs10502861), demonstrating that even low prediction markers can have high accuracy when added to stronger ones (Marcińska *et al.*, 2015).

2. Legal and ethical aspects

When seeing the various legal issues surrounding the sample collection from suspects for comparison, an advantage of DNA phenotyping is clearly seen, this is the fact of this approach being more focused on obtaining genetic profiles from crime scene samples, thus not harming dignity or integrity rights, in addition, as the name says, Externally Visible Characteristics are publicly available features and therefore would not involve privacy issues (Kayser, 2015). However, this would represent that all those individuals who share the characteristics of a facial composite may be interviewed and required to donate samples for comparison, raising questions about harassment that certain groups with a determined physical characteristic could suffer, belonging to a group of suspects solely by their physical appearance. Therefore, prior to any application it should be created new legal and ethical regulations to preserve the integrity and intimacy of people involved in DNA phenotyping-based investigations (S. J. Walsh, 2004).

The classification of the human species into races or subspecies verified in the prediction of ancestry is one of the most controversial issues within the scientific community, and this division is often considered racist and a violation of the standards of scientific objectivity (Lee *et al.*, 2008).

Conclusion

In the field of phenotype–genotype association there are several studies that have shown significant associations; however, each target population has its own genetic background, and prior to any extrapolation each set of genetic markers associated to a phenotype should be carefully evaluated in additional populations.

In conclusion and having into consideration has the field of Forensic DNA phenotyping has evolve it is safe to say that obtaining a set of genetic markers that accurately predict most of human EVCs for forensic use is closer than ever and is not too far to actually be used in forensic routine investigations. The data still should be verified and applied in diverse global populations and check if none of the associations found are due to ancestry or other populational background. Despite the legal and ethical issues is still a viable approach for practical use in forensic routine, mainly because this issues are common to most approaches that involve DNA information.

Bibliography

Branicki, W., Liu, F., Van Duijn, K., Draus-Barini, J., Pośpiech, E., Walsh, S., … Kayser, M. (2011). Model-based prediction of human hair color using DNA variants. *Human Genetics*, *129*(4), 443–454. https://doi.org/10.1007/s00439-010-0939-8

Budowle, B., & Van Daal, A. (2008). Forensically relevant SNP classes. *BioTechniques*, *44*(5), 603–610. https://doi.org/10.2144/000112806

Butler, E., & Li, R. (2014). Genetic Markers for Sex Identification in Forensic DNA Analysis. *Journal of Forensic Investigation*, *02*(03). https://doi.org/10.13188/2330-0396.1000013

Butler, J. M. (2009). Fundamentals of Forensic DNA Typing. In *Journal of Chemical Information and Modeling* (Vol. 53). https://doi.org/10.1017/CBO9781107415324.004

Cadenas, A. M., Regueiro, M., Gayden, T., Singh, N., Zhivotovsky, L. A., Underhill, P. A., & Herrera, R. J. (2007). Male amelogenin dropouts: phylogenetic context, origins and implications. *Forensic Science International*, *166*(2–3), 155–163. https://doi.org/10.1016/j.forsciint.2006.05.002

Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., … Walsh, S. (2018). The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. In *Forensic Science International: Genetics* (Vol. 35). https://doi.org/10.1016/j.fsigen.2018.04.004

Chaitanya, L., Walsh, S., Andersen, J. D., Ansell, R., Ballantyne, K., Ballard, D., … Kayser, M. (2014). Collaborative EDNAP exercise on the IrisPlex system for DNA-based prediction of human eye colour. In *Forensic Science International: Genetics* (Vol. 11).

https://doi.org/10.1016/j.fsigen.2014.04.006

Claes, P., Hill, H., & Shriver, M. D. (2014). Toward DNA-based facial composites: Preliminary results and validation. *Forensic Science International: Genetics*, *13*, 208–216. https://doi.org/10.1016/j.fsigen.2014.08.008

Codina, A. E., Niederstätter, H., & Parson, W. (2009). GenderPlex a PCR multiplex for reliable gender determination of degraded human DNA samples and complex gender constellations. *International Journal of Legal Medicine*, *123*(6), 459–464. https://doi.org/10.1007/s00414-008-0301-z

Dale, J. W., & Schantz, M. von. (2002). *From genes to genome: Concepts and applications of DNA Technology*.

Dembinski, G. M., & Picard, C. J. (2014). Evaluation of the IrisPlex DNA-based eye color prediction assay in a United States population. *Forensic Science International: Genetics*, *9*(1), 111–117. https://doi.org/10.1016/j.fsigen.2013.12.003

Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics*, *5*(6), 435–445. https://doi.org/10.1038/nrg1348

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., … Mountain, J. (2010). Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genetics*, *6*(6), 1–20. https://doi.org/10.1371/journal.pgen.1000993

Frudakis, A. (2008). *MOLECULAR PHOTOFITTING: PREDICTING ANCESTRY AND PHENOTYPE USING DNA*.

Guo, M. H., Hirschhorn, J. N., & Dauber, A. (2018). Insights and Implications of Genome-Wide Association Studies of Height. *Journal of Clinical Endocrinology and Metabolism*, *103*(9), 3155–3168. https://doi.org/10.1210/jc.2018-01126

HIRISPLEX-S, HIRISPLEX & IRISPLEX Eye, Hair and Skin colour DNA Phenotyping webtool USER MANUAL. (2018). *0*, 0–22.

Hong, S. R., Jung, S. E., Lee, E. H., Shin, K. J., Yang, W. I., & Lee, H. Y. (2017). DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers. *Forensic Science International: Genetics*, *29*, 118–125. https://doi.org/10.1016/j.fsigen.2017.04.006

Jablonski, N. G., & Chaplin, G. (2017). The colours of humanity: The evolution of pigmentation in the human lineage. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1724). https://doi.org/10.1098/rstb.2016.0349

Kao, L., Tsai, L., Lee, J. C., & Hsieh, H. (2007). Controversial cases of human gender identification by amelogenin test. *Forensic Science Journalcience Journal*, *6*(2), 69–71.

Kayser, M. (2015). Forensic DNA Phenotyping : Predicting human appearance from crime

scene material for investigative purposes. *Forensic Science International: Genetics*, *18*, 33–48. https://doi.org/10.1016/j.fsigen.2015.02.003

Kayser, M., & De Knijff, P. (2011). Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics*, *12*(3), 179–192. https://doi.org/10.1038/nrg2952

Kayser, M., Liu, F., Janssens, A. C. J. W., Rivadeneira, F., Lao, O., van Duijn, K., … van Duijn, C. M. (2008). Three Genome-wide Association Studies and a Linkage Analysis Identify HERC2 as a Human Iris Color Gene. *American Journal of Human Genetics*, *82*(2), 411–423. https://doi.org/10.1016/j.ajhg.2007.10.003

Kayser, M., & Schneider, P. M. (2009). DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. *Forensic Science International: Genetics*, *3*(3), 154–161. https://doi.org/10.1016/j.fsigen.2009.01.012

Kukla-Bartoszek, M., Pośpiech, E., Spólnicka, M., Karłowska-Pik, J., Strapagiel, D., Żądzińska, E., … Branicki, W. (2018). Investigating the impact of age-depended hair colour darkening during childhood on DNA-based hair colour prediction with the HIrisPlex system. *Forensic Science International: Genetics*, *36*, 26–33. https://doi.org/10.1016/j.fsigen.2018.06.007

Lee, S. S. J., Mountain, J., Koenig, B., Altman, R., Brown, M., Camarillo, A., … Underhill, P. (2008). The ethics of characterizing difference: Guiding principles on using racial categories in human genetics. *Genome Biology*, *9*(7), 7–10. https://doi.org/10.1186/gb-2008-9-7-404

Liu, F., Hendriks, A. E. J., Ralf, A., Boot, A. M., Benyi, E., Sävendahl, L., … Kayser, M. (2014). Common DNA variants predict tall stature in Europeans. *Human Genetics*, *133*(5), 587–597. https://doi.org/10.1007/s00439-013-1394-0

Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G., … Kayser, M. (2012). A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans. *PLoS Genetics*, *8*(9). https://doi.org/10.1371/journal.pgen.1002932

Loussouarn, G., Garcel, A. L., Lozano, I., Collaudin, C., Porter, C., Panhard, S., … de La Mettrie, R. (2007). Worldwide diversity of hair curliness: A new method of assessment. *International Journal of Dermatology*, *46*(SUPPL. 1), 2–6. https://doi.org/10.1111/j.1365-4632.2007.03453.x

Marcińska, M., Pośpiech, E., Abidi, S., Andersen, J. D., Van Den Berge, M., Carracedo, Á., … Branicki, W. (2015). Evaluation of DNA variants associated with androgenetic alopecia and their potential to predict male pattern baldness. *PLoS ONE*, *10*(5), 1–18. https://doi.org/10.1371/journal.pone.0127852

Martinez-Cadenas, C., Penãa-Chilet, M., Ibarrola-Villava, M., & Ribas, G. (2013). Gender is a major factor explaining discrepancies in eye colour prediction based on HERC2/OCA2 genotype and the IrisPlex model. *Forensic Science International: Genetics*, *7*(4), 453–460.

https://doi.org/10.1016/j.fsigen.2013.03.007

Masuyama, K., Shojo, H., Nakanishi, H., Inokuchi, S., & Adachi, N. (2017). Sex determination from fragmented and degenerated DNA by amplified product-length polymorphism bidirectional SNP analysis of amelogenin and SRY genes. *PLoS ONE*, *12*(1), 1–17. https://doi.org/10.1371/journal.pone.0169348

Meissner, C., & Ritz-Timme, S. (2010). Molecular pathology and age estimation. *Forensic Science International*, *203*(1–3), 34–43. https://doi.org/10.1016/j.forsciint.2010.07.010

Nyholt, D. R., Gillespie, N. A., Heath, A. C., & Martin, N. G. (2003). Genetic Basis of Male Pattern Baldness. *Journal of Investigative Dermatology*, *121*(6), 1561–1564. https://doi.org/10.1111/j.1523-1747.2003.12615.x

Parker, G. J., Yip, J. M., Eerkens, J. W., Salemi, M., Durbin-Johnson, B., Kiesow, C., … Phinney, B. S. (2019). Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel. *Journal of Archaeological Science*, *101*(May 2018), 169–180. https://doi.org/10.1016/j.jas.2018.08.011

Paternoster, L., Zhurov, A. I., Toma, A. M., Kemp, J. P., St. Pourcain, B., Timpson, N. J., … Evans, D. M. (2012). Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *American Journal of Human Genetics*, *90*(3), 478–485. https://doi.org/10.1016/j.ajhg.2011.12.021

Phillips, C. (2015). Forensic genetic analysis of bio-geographical ancestry. *Forensic Science International: Genetics*, *18*, 49–65. https://doi.org/10.1016/j.fsigen.2015.05.012

Pośpiech, E., Draus-Barini, J., Kupiec, T., Wojas-Pelc, A., & Branicki, W. (2011). Gene-gene interactions contribute to eye colour variation in humans. *Journal of Human Genetics*, *56*(6), 447–455. https://doi.org/10.1038/jhg.2011.38

Pośpiech, E., Karłowska-Pik, J., Ziemkiewicz, B., Kukla, M., Skowron, M., Wojas-Pelc, A., & Branicki, W. (2016). Further evidence for population specific differences in the effect of DNA markers and gender on eye colour prediction in forensics. *International Journal of Legal Medicine*, *130*(4), 923–934. https://doi.org/10.1007/s00414-016-1388-2

Rees, J. L. (2003). Genetics of Hair and Skin Color. *Annual Review of Genetics*, *37*(1), 67–90. https://doi.org/10.1146/annurev.genet.37.110801.143233

Rhein, M., Hagemeier, L., Klintschar, M., Muschler, M., Bleich, S., & Frieling, H. (2015). DNA methylation results depend on DNA integrity-role of post mortem interval. *Frontiers in Genetics*, *6*(MAY), 1–7. https://doi.org/10.3389/fgene.2015.00182

Royal, C. D., Novembre, J., Fullerton, S. M., Goldstein, D. B., Long, J. C., Bamshad, M. J., & Clark, A. G. (2010). Inferring Genetic Ancestry: Opportunities, Challenges, and Implications. *American Journal of Human Genetics*, *86*(5), 661–673. https://doi.org/10.1016/j.ajhg.2010.03.011

Samuel, G., & Prainsack, B. (2018). *The regulatory landscape of forensic DNA phenotyping in Europe*. 119. Retrieved from http://www.visage-h2020.eu/Report_regulatory_landscape_FDP_in_Europe2.pdf

Santos, F., Machado, H., & Silva, S. (2013). *Forensic DNA databases in European countries : is size linked to performance ?* (Cutter 2006), 1–13.

Sturm, R. A. (2009). Molecular genetics of human pigmentation diversity. *Human Molecular Genetics*, *18*(R1), 9–17. https://doi.org/10.1093/hmg/ddp003

Vidaki, A., & Kayser, M. (2017). *From forensic epigenetics to forensic epigenomics: broadening DNA investigative intelligence*. 1–13. https://doi.org/10.1186/s13059-017-1373-1

Vidaki, A., & Kayser, M. (2018). Recent progress, methods and perspectives in forensic epigenetics. *Forensic Science International: Genetics*, *37*(July), 180–195. https://doi.org/10.1016/j.fsigen.2018.08.008

von Wurmb-Schwark, N., Bosinski, H., & Ritz-Timme, S. (2007). What do the X and Y chromosomes tell us about sex and gender in forensic case analysis? *Journal of Forensic and Legal Medicine*, *14*(1), 27–30. https://doi.org/10.1016/j.jcfm.2005.09.003

Walsh, S. J. (2004). Recent advances in forensic genetics. *Expert Review of Molecular Diagnostics*, *4*(1), 31–40. https://doi.org/10.1586/14737159.4.1.31

Walsh, S., Lindenbergh, A., Zuniga, S. B., Sijen, T., De Knijff, P., Kayser, M., & Ballantyne, K. N. (2011). Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics*, *5*(5), 464–471. https://doi.org/10.1016/j.fsigen.2010.09.008

Walsh, S., Liu, F., Ballantyne, K. N., Van Oven, M., Lao, O., & Kayser, M. (2011). IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics*, *5*(3), 170–180. https://doi.org/10.1016/j.fsigen.2010.02.004

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-kamysz, A., … Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics*, *7*(1), 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., … Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, *46*(11), 1173–1186. https://doi.org/10.1038/ng.3097

Zubakov, D., Liu, F., Van Zelm, M. C., Vermeulen, J., Oostra, B. A., Van Duijn, C. M., … Langerak, A. W. (2010). Estimating human age from T-cell DNA rearrangements. *Current Biology*, *20*(22), R970–R971. https://doi.org/10.1016/j.cub.2010.10.022