Cell

# DNA Phenotyping: Snapshot of a Criminal

## Recent progresses in genomic technologies provide both opportunities and challenges to forensics.

In 1984, a British geneticist studying inherited diseases stared in confusion at the jumbled results of a test he had just run, a test that included DNA from a father, mother, and daughter. Then, with a flash of insight, he realized that the differences and similarities between the three results meant that a DNA profile could be used to distinguish between individuals and to resolve paternity disputes. A decade later he was knighted by Queen Elizabeth for this work.

Today, the method of "DNA fingerprinting" discovered by Sir Alec Jeffreys is used all over the world by forensic scientists. A single drop of blood or a few skin cells left at a crime scene is enough to generate a DNA profile of a suspect, and this profile can be compared to the millions of DNA profiles already collected in databases of convicted offenders. If a match is made, then the name and face of the guilty party is revealed.

This criminal DNA profile is not an individual's entire genome; it is simply a list of numeric values associated with 13 markers (locations) within the genome. These 13 markers were chosen because they show high variability from person to person and because the probability of two persons sharing the same values for all 13 markers is vanishingly small.

These 13 markers each contain a short tandem repeat (STR), a string of nucleotides that are repeated a number of times. For example, an STR for one person could look like *gatagatagatagatagata gata*, where four nucleotides (*gata*) are repeated 6 times, but for another person *gata* could be repeated 15 times. The number of repeats varies from person to person, and it's this repeat number that is included in a person's DNA profile. A DNA profile contains 13 pairs of numbers (a pair represents heterozygous alleles). These 13 pairs of numbers can be combined in trillions of ways to create trillions of unique profiles. Even if a DNA sample recovered from a crime scene is small and only includes a fraction of the 13 STR regions that are tested, the analysis can still be helpful.

DNA profiles were designed to be used for identification purposes only; that is, they were not meant to contain personal genetic information, such as physical traits, tendencies toward diseases, or emotional and intellectual dispositions. This is a genotype-to-genotype comparison. The phenotype of the person (how the genes are expressed to produce physical traits, propensities toward diseases, etc.) is not involved. Think of a driver's license number or a fingerprint—the number itself or the fingerprint ridge pattern reveals no personal information. The real information (a name, date of birth) is disclosed when a match is made with a record in a database. Likewise, a DNA profile is not meant to reveal personal information itself; it is simply a way to make a match with a preexisting DNA profile in a database. (To clarify, the DNA profile was designed to contain no personal information, but recent studies show this is not true. See more on this topic below.)

What happens if there is no match between the DNA profile generated from crime scene evidence and the existing DNA profiles in criminal databases (because the guilty party has no previous record of crime)? If there are no witnesses and no profile-to-profile match, the criminal investigation comes to a standstill.

### Predicting Eye, Hair, and Skin Color

Susan Walsh, a forensic geneticist at Indiana University–Purdue University Indianapolis, would like to move that stalled investigation forward. Working with colleagues in the Netherlands, she helped develop computational models that predict physical traits of a suspect based on DNA left at a crime scene. This process is completely different from generating the standard numerical DNA profile that is used as a reference in databases. Instead Walsh looks at the DNA itself to predict eye color, hair color, and skin color of a suspect.

Walsh started with eye color. She and Manfred Kayser, head of the Department of Genetic Identification at Erasmus University Medical Center in Rotterdam, conducted association studies across the genomes of over 9,000 people and found that using just six single nucleotide polymorphisms (SNPs) associated with six genes led to high categorical prediction values for eye color. (A SNP is a variation in just one nucleotide in the genome at a certain location, and SNPs often act as a marker of a gene responsible for a specific trait or disease.)

Walsh and Kayser then designed a biological assay and developed a statistical model (using multinomial logistic regression) to predict eye color given an unknown DNA sample. This model is accurate up to 95% of the time in predicting blue or brown eyes. More research needs to be done to discover all the genes associated with eye color, especially for colors that are not blue or brown.

Next, with Wojciech Branicki's research group in Poland, they searched for SNPs associated with hair color and located 22 SNPs associated with 11 genes that gave reasonably high prediction accuracies. They expanded their prediction model to include categories for hair color and have determined that the model is approximately 90% accurate at predicting black hair and 80% accurate at predicting red and brown hair. However, blond hair is more difficult to predict because a number of individuals tend to darken with age, going from blond-haired children to brown- or black-haired adults. At present when tested on over 100 individuals from Europe, the combined model for eye color and hair color—which has been validated by multiple forensic labs—gives an accurate prediction in approximately 75% of cases.

Now Walsh, Kayser, and Branicki are working on a model for skin color based on 36 SNPs from 15 genes. The accuracy of this model, using five skin color categories based on a dermatological scale, is in line with hair color prediction. They will publish the work later this year.

"Right now it's a very simple categorical prediction that we can do," says

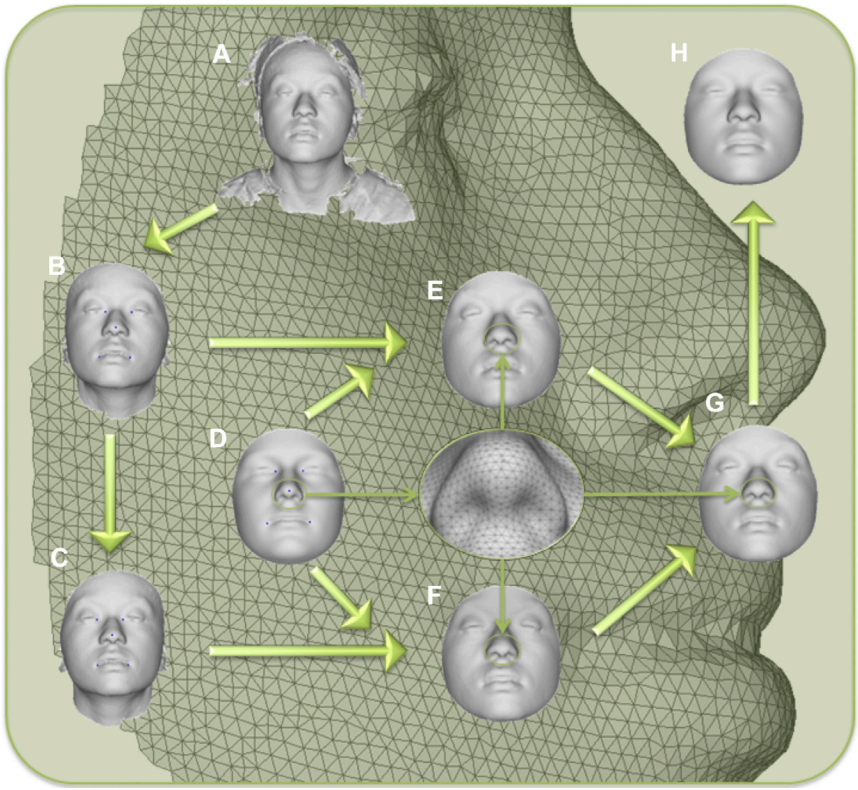


**Figure 1. Mark Shriver's Workflow for a 3D Facial Scan**
(A–H) The original facial surface (A) is trimmed (B) and reflected to make a mirror image (C); then facial landmarks are mapped (D), remapped and reflected (E and F), made symmetrical (G), and reconstructed (H). Reprinted from Claes, P., Liberton, D.K., Daniels, K., et al. (2014) Modeling 3D Facial Shape from DNA. PLOS Genetics *10*, e1004224.

Walsh, "to get a general description: blue eyes, blond hair, pale skin." She was recently awarded $1.1 million from the National Institute of Justice, the research and development branch of the US Department of Justice, to expand the model to predict precise quantitative values of color from a palette, which will improve the efficiency of criminal investigations.

Walsh clarifies that her method isn't for all cases and that it is not intended for identification purposes. The predictions won't stand up in court because eye color and hair color are shared by many people, and hair and skin color can be changed by dyes and by the sun. The purpose of prediction is simply to create a new lead in a cold case, to direct investigators to a new group of suspects. Walsh explains that after suspects are found, "we still need to make a profile-to-profile match" using current DNA analysis. "Perhaps [my model] may lead to a suspect being interrogated, their [DNA] profile received, and a match being found."

Walsh's next challenge is getting police departments to try these new DNA analysis methods. She says that US law enforcement agencies tend to be conservative in adopting new technology and wait to see it applied elsewhere before trying it themselves. She has begun one pilot study on cold cases with the Indiana State Police, and her tools are being used on a few cases by policing agencies in the Netherlands, Poland, and Australia. "The science has been done; now it's just up to training and spreading the word that these types of methods can be used, can be trusted, and that they have been validated."

Walsh emphasizes that full disclosure of research methods in DNA phenotyping is crucial, including where samples were collected, the number of samples, the populations they were collected from, which markers were typed, and which individuals were used to generate the statistical models. A number of companies claim they can predict facial features and generate, from a DNA sample, a full-color photograph of an individual, complete with real pigment phenotypes and even a specific hair style, but they do not publish scientific studies nor divulge their "proprietary" software and analysis. She worries that companies like this will damage the reputation of DNA phenotyping.

## Predicting Facial Features

A step beyond predicting eye, hair, and skin color is predicting the detailed appearance of the face. Mark Shriver, geneticist and professor in the anthropology department at Pennsylvania State University, has created a computer model to predict facial shape from DNA alone. Like Walsh, he is not generating the standard DNA profile used in law enforcement databases; he is looking directly at the DNA to reveal physical traits.

Working with Peter Claes, biomedical imaging expert at KU Leuven in Belgium, Shriver first collected DNA samples and facial data from volunteers. The facial data accurately describe, in three dimensions, every tiny feature of the face. Using an off-the-shelf 3D surface imaging system with cameras set at different angles, he takes multiple images of a facial pose simultaneously, creating a 3D representation of the face. The computer then marks more than 30,000 vertices on each photograph and, through triangulation, comes up with sets of x, y, z coordinates in three-dimensional space. The software then connects the vertices creating an image that looks like a wire frame. This image gets softened to create a smooth surface, approximating the original face (Figure 1). Claes then designed an algorithm, using dense correspondence analysis, that takes the data output from the imaging system and maps it into a common set of 7,000 facial landmarks, so that advanced statistical analysis can be performed on the dataset.

For their first prediction model, Shriver created a list of nearly 200 genes associated with facial development using the Online Mendelian Inheritance in Man (OMIM) catalog. From this list he chose 50 genes that showed accelerated evolution, which could imply that these genes are responsible for visual facial

differences between populations. Using DNA and facial data collected from a subset of about 600 African/European mixed persons, he analyzed how these 50 genes affect facial features and discovered 24 SNPs in 20 different genes that significantly demonstrated a pattern of effect on facial shape. Then, going backward, he developed a model to *predict* the facial shape from DNA alone.

Shriver's model first determines a score for biological sex and ancestry (geographic population group) and from this creates a "base face," for example, a woman with 85% West African ancestry or a man with 95% European ancestry. Next the model determines the effect of each of the 24 SNPs on this face and layers those effects onto the base face.

Shriver and Claes are now working on the third version of their prediction model, using over 1,000 SNPs from hundreds of genes, which they discovered by doing association studies across the genomes of individuals of European ancestry, a subset of the more than 8,000 genomes they have collected from individuals from around the world. Shriver's biggest challenge is the complexity of the face. For any facial feature, such as the nose, multiple genes affect its shape. In addition, Shriver wants to accurately model each measurement of the nose using continuous data to describe the specific length of the nose, the width of the nostrils, and so on, rather than using broad categories. Furthermore, the face is modular, which means that if a gene affects the shape of the nose, it's probably also affecting another facial feature.

In addition, nongenetic factors—such as age, hormones, and nutrition—contribute significantly to facial shape but cannot be deduced from DNA. (At least not yet; recent studies suggest that DNA methylation markers can be used to predict age.) Shriver estimates that for an adult face 25% of its shape is the effect of biological sex and ancestry, and 15%–20% of its shape is the effect of his 1,000 variants, which means that about 40%–45% of facial features can currently be described by his model. His work is still in the early stages and is not yet ready to be used routinely in law enforcement or intelligence work.

However, some scientists are less optimistic that Shriver's model can predict facial appearance because of the genetics of complex features. Shriver agrees that much more work needs to be done across many fields to improve his model in the hope of applying it to actual criminal cases.

"I have a lot of confidence that even in just a few years we'll be in a position to do this well… and to report what we're doing in a way that makes it a useful general tool for a lot of cases," says Shriver.

## Racial Profiling?

One concern about analyzing crime scene DNA to predict a suspect's physical coloring and genomic ancestry is that this information could be used for racial profiling. Mark Shriver says that he has come under greater scrutiny than usual for his work and that people have refused to participate as volunteer subjects because they believe his work might add fuel to an already racially biased law enforcement system.

Shriver responds: "Some people would say it's racial profiling to take evidence from a crime scene and make deductions about … what population that person might belong to or come from. If you're using evidence like that, it's not racial profiling in the narrow sense. [Racial profiling] is the *misuse* of patterns or the *misuse* of somebody's racial information to dragnet or pull them in for any crime they might have committed, to focus unfairly on a certain population. If there's evidence from the crime scene, then [a search based on DNA analysis] is not an unfounded focus."

But racial profiling concerns are not just an issue with these newer technologies that predict appearance; these concerns also apply to the current method of creating DNA profiles (the "DNA fingerprint" first discovered by Jeffreys in 1984). DNA profiles were originally designed to contain only "noninteresting" genetic information for identification purposes only, similar to a driver's license number or fingerprint ridge pattern, which can be matched with a profile in a reference database. Department of Justice regulations specify that a profile generated from a DNA sample must not contain any personal information about physical traits, diseases, and dispositions.

However, a study by Bridget Algee-Hewitt, research fellow at Stanford University and soon-to-be professor in the Department of Anthropology at Florida State University, shows that the 13 markers of a traditional DNA profile do in fact contain ancestry (or population) information. Looking at the patterns in the genetic makeup of individuals in a population, Algee-Hewitt and colleagues showed that the highly polymorphic genetic markers in a DNA profile (which are used to provide unique individual identifiability) inherently include ancestry information. This means that, despite the intention of federal regulations, ancestry information *can* be obtained from the markers. (In 2017 the set of markers will be expanded from 13 to 20; this will create a better identification system but will also provide greater detail related to ancestry.)

"If you allow ancestry information to be extracted from the [DNA profile], … then you can start identifying individuals based on features that you might think are associated, or know statistically to be associated, with a particular ancestry group," says Algee-Hewitt. She explains that a test result stating a person "is likely of African origin" may be translated by nonspecialists into the social language of identity or race, such as "African American" or "black." This is a result of the practical needs of law enforcement in areas such as eyewitness descriptions, missing person reports, government forms, and census questionnaires, which all use racial terms. Therefore, if ancestry information is pulled from a DNA profile, then law enforcement might make decisions based on predisposed expectations about the link between ancestry and racial/social identity. Algee-Hewitt says this could become "racial profiling at its worst." She adds: "At its best, however, it could be a way to help exonerate someone."

## Science in Service to Criminal Justice

Science is simply a tool, and it need not be used in support of a racially biased system. It can be used to advance justice and protect the innocent. A 2009 report by the National Academy of Sciences endorsed only one forensic method as rigorous and accurate, DNA analysis. The traditional DNA profile has been consistently and successfully used by

the Innocence Project to exonerate individuals falsely incarcerated. Now these newer types of DNA analyses can be employed at the *beginning* of an investigation to protect the innocent and to reverse inaccurate or racially motivated accusations before a certain group of people is wrongly suspected and questioned.

Indeed, the first time Alec Jeffreys was asked to use his new ''DNA fingerprinting'' technique for a legal case, the test protected an innocent person from a false accusation. It was 1985 and English authorities had detained a 13-year-old boy from Ghana attempting to enter the country to be reunited with his mother. The authorities wanted proof that the boy really was her son and not her nephew. Jeffreys' test concluded, by an overwhelming probability, that the boy was in fact her son.

Forensic tools need not be considered the problem; rather they can be used to help create a more just and hopeful system. Jeffreys understood that his DNA technique advanced justice in a very personal way: ''I was actually there when the mother was told … [and] the look in that lady's eyes was magic,'' said Jeffreys in an interview decades later. ''Of all the various moments in the DNA fingerprinting story, that has to be my favorite.''

**Susan Matheson**
Cambridge, MA
http://dx.doi.org/10.1016/j.cell.2016.08.016