



The science of DNA facial reconstruction is advancing rapidly. Composite from Parabon and PNAS

DNA facial prediction could make protecting your privacy more difficult

May 2, 2018 3.05am EDT

*Technologies for amplifying, sequencing and matching DNA have created new opportunities in genomic science. In this series *When DNA Talks* we look at the ethical and social implications.*

Authors



Caitlin Curtis

Research fellow, Centre for Policy Futures (Genomics), The University of Queensland



James Hereward

Research fellow, The University of Queensland

Everywhere we go we leave behind bits of DNA.

We can already use this DNA to predict some traits, such as eye, skin and hair colour. Soon it may be possible to accurately reconstruct your whole face from these traces.

This is the world of “DNA phenotyping” – reconstructing physical features from genetic data. Research studies and companies like 23andMe sometimes share genetic data that has been “anonymised” by removing names. But can we ensure its privacy if we can predict the face of its owner?

Here’s where the science is now, and where it could go in the future.

Read more: Is your genome really your own? The public and forensic value of DNA

Predicting hair, eye and skin colour

DNA phenotyping has been an active area of research by academics for several years now. Forensic biology researchers Manfred Kayser and Susan Walsh, among others, have pioneered several DNA phenotyping methods for forensics.

In 2010, they developed the IrisPlex system, which uses six DNA markers to determine whether someone has blue or brown eyes. In 2012, additional markers were included to predict hair colour. Last year the group added skin colour. These tests have been made available via a website and anyone who has access to their genetic data can try it out.

Trait predictions are being used to address a number of questions. Recently, for example, they were used to suggest that the “Cheddar Man” (the UK’s oldest complete human skeleton) may have had dark or dark to black skin and blue/green eyes. The predictive models are mostly built on modern European populations, so caution may be required when applying the tests to other (especially ancient) populations.

The full picture

Research on DNA phenotyping has advanced rapidly in the last year with the application of machine learning approaches, but the extent of our current capabilities is still hotly debated.

Last year, researchers from American geneticist Craig Venter’s company Human Longevity, made detailed measurements of the physical attributes of around 1,000 people. Whole genomes (our complete genetic code) were sequenced and the data combined to make models that predict 3D facial structure, voice, biological age, height, weight, body mass index, eye colour and skin colour.

Read more: How cops used a public genealogy database in the Golden State Killer case

The study received strong backlash from a number of prominent scientists, including Yaniv Erlich, aka the “genome hacker”. The study seemed to predict average faces based on sex and ancestry, rather than specific faces of individuals. The method of judging the predictions on small ethnically mixed cohorts was also criticised.

Even with accurate facial predictions, Erlich noted that for this approach to identify someone in the real world:

an adversary ... would have to create [a] population scale database that includes height, face morphology, digital voice signatures and demographic data of every person they want to identify.

Because without a detailed biometric database you can't get from the physical predictions to a name.

A database to match?

It turns out that the Australian government is in the process of building such a database. “The Capability” is a proposed biometric and facial recognition system that will match CCTV footage to information from passports and driving licences. Initially billed as a counter-terrorism measure, there are already reports the service may be provided for a fee to corporations.

At the same time, the Australian Tax Office has just initiated a voice recognition service. It's easy to imagine how this kind of system could be integrated with “The Capability”.

And it's not only Australia establishing the capability to become a biometric, face-recognising surveillance state. India is deploying the Aadhar system, and China leads the world in facial recognition.



The Australian Government is building a facial recognition system called The Capability that will match CCTV footage to information from passports and driving licences. Queensland Government

DNA mugshots

At present, most forensic DNA profiling techniques rely on “anonymous” markers that match identity to a database, but reveal little else about a suspect. With advances in genomic technology, forensic genetics is moving toward tests that can tell us much more about someone.

There are a number of companies that offer DNA phenotyping services for a fee. One company, Parabon NanoLabs, claims to be able to accurately predict the physical appearance of an unknown person from DNA. Police forces already use their services, including the Queensland police in a recent case of a serial rapist on the Gold Coast.

The Parabon system is also based on a predictive model. This was developed by applying machine learning tools to their genetic/trait reference database. The company predicts skin colour, eye colour,

hair colour, freckles, ancestry, and face shape from a DNA sample. These predictions, the confidence around them, and a reconstruction made by a forensic artist are used to make a “Snapshot” profile.

Read more: New cryptocurrencies could let you control and sell access to your DNA data

There is scepticism about the capabilities of Parabon. It is difficult to assess Parabon’s system because the computer code is not open, and the methodology has not been published with peer-review scrutiny.

As with any type of DNA evidence, there is a risk of miscarriages of justice, especially if the evidence is used in isolation. The utility of DNA phenotyping at this point may be more in its exclusionary power than its predictive power. Parabon does state that Snapshot predictions are intended to be used in conjunction with other investigative information to narrow the list of possible suspects.

Where will this all end up?

We only need to look at identical twins to see how much of our face is in our DNA. The question is how many of the connections between DNA and our physical features will we be able to unlock in the future, and how long will it take us to get there?

Some features are relatively easy to predict. For instance, eye colour can be inferred from relatively few genetic variants. Other traits will be more complicated because they are “polygenic”, meaning that many gene variants work together to produce the feature.

A recent study of hair colour genetics, for example, examined 300,000 people with European ancestry. They found 110 new genetic markers linked to hair colour, but the prediction of some colours (black or red) is more reliable than others (blonde and brown).



Twins can show us how much of our face is in our DNA. Clard/Pixabay

The way that DNA codes our physical features might be different in people from different ancestral groups. Currently, our ability to predict modern Europeans will be better than other groups – because our genetic databases are dominated by subjects with European ancestry.

As we employ increasingly sophisticated machine learning approaches on bigger (and more ethnically representative) databases, our ability to predict appearance from DNA is likely to improve dramatically.

Parabon's services come with a disclaimer that the reconstructions should not be used with facial recognition systems. The integration of these technologies is not impossible in the future, however, and raises questions about scope creep.

What does this mean for genetic privacy?

Despite the controversy around what we can do now, the science of DNA phenotyping is only going to get better.

What the rapidly developing field of DNA phenotyping shows us is how much personal information is in our genetic data. If you can reconstruct a mugshot from genetic data, then removing the owner's name won't prevent re-identification.

Protecting the privacy of our genetic data in the future may mean that we have to come up with innovative ways of masking it – for example genome cloaking, genome spiking, or encryption and blockchain-based platforms.

The more we understand about our genetic code the more difficult it will become to protect the privacy of our genetic data.

Genetics Privacy DNA Forensics DNA data DNA profiling When DNA Talks

Before you go...

The 2020 presidential campaign is expected to feature unprecedented amounts of disinformation. At The Conversation, we'll be doing our best to produce journalism written by experts, based on their research. Your donation of \$5, \$50 or \$500 will help ensure that this journalism is distributed, free of charge, where it is needed.

[Donate now](#)

Naomi Schalit

Senior Politics + Society Editor

