

Recreating the Face of a Perpetrator from a Drop of Blood

Researchers are identifying genetic variants that influence facial features, enabling the development of facial imaging from DNA.

Introduction

Our love of a good whodunit can be seen in the TV police dramas we tune into each week. At first, the crime seems unsolvable. There are no eyewitnesses and the evidence left at the scene—a strand of hair, flake of skin, or drop of blood—doesn't fit anyone in the criminal DNA database. But the detectives' skill and a little luck usually means that the guilty party is in handcuffs by the end of the episode.

Reality isn't so tidy. Sometimes even the best inspectors have nothing to go on, no leads to investigate, and the case runs cold. What's more, even when arrests are made, eyewitness testimony can be unreliable, causing cases to fall apart at trial. Although genomic tools such as the MiSeq® FGx System are revolutionizing forensic science, Pennsylvania State University anthropologist Mark Shriver, PhD wants to push the science further. His work has shown that single nucleotide polymorphism (SNP) data can be used to construct an image of a person's face, opening up new avenues of inquiry and investigation for DNA examiners and criminal investigators.

Facial Features and Evolution

Dr. Shriver didn't start out with an interest in forensic science. He wanted to answer basic questions, including 'Which genes affect facial features?' and 'How have these genes evolved over time in different human populations?' "Natural selection shaped fundamental processes, such as how our immune systems respond to pathogens, or whether we can drink milk as an adult," Dr. Shriver said. "Facial features are also the result of natural selection. The genes affecting these more visible traits have evolved faster than the rest of the genome. SNPs can inform us about the process of human evolution and what evolutionary changes have occurred recently. We can use them to extrapolate changes back in time to gain a deeper understanding of evolution as it has shaped our species."

Beyond its value in anthropology, Dr. Shriver soon realized that facial SNP data might someday be useful in creating the image of a perpetrator's face from a small amount of DNA left at a crime scene. Tackling that problem requires identifying all the SNPs involved in the creation of a face. "In the end there's going to be hundreds of genes that are important in determining the shape of someone's face," Dr. Shriver stated. "Yet, specific subsets of genetic markers will likely control certain features in subpopulations. Certain markers will also enable us to estimate biogeographical ancestry, which impacts facial shape."

Facial SNP Data Analysis Challenges

In 2000, Dr. Shriver's team began collecting DNA samples, along with measurements of traits like skin, eye, and hair color, and photographs of Penn State students, as well as attendees at World Science and Twins Day festivals. In 2005, they added another layer of data—3-D facial scans—enabling comparison of physical face shape and human genotypes.

Obtaining and interpreting these data hasn't been an easy process. "The face is so complex that describing it is difficult," Dr. Shriver says. "Each trait, such as eye-to-eye distance, nose height and width, or chin length, can be affected by many genetic variants with a small effect size. That means we have to be creative in how we choose to summarize the face, and must collect and analyze data from very large sample populations."

Before high-throughput SNP genotyping methods were developed, it was difficult to evaluate enough polymorphic loci. The team could only study 10–100 SNPs simultaneously, uncovering only a fraction of the total variation. The arbitrarily selected small SNP "panels" didn't yield much biogeographical ancestry information.



Dr. Mark Shriver is Professor of Anthropology at Pennsylvania State University.

Dr. Shriver's team has been using the personal genome service 23andMe to genotype samples for their research. As an incentive to volunteer for Dr. Shriver's project, participants are provided with access to their 23andMe data. 23andMe uses genotyping microarrays manufactured by Illumina that can analyze over 600,000 SNPs in a single run. Yet, Dr. Shriver believes that all these SNPs are not created equal: He focuses in part on ancestry informative SNPs (aiSNPs), which are loci with alleles that have very different frequencies across populations.

"Most SNPs have similar frequencies regardless which population a person comes from because human populations are not as different genetically as they might appear," Dr. Shriver said. "When you sift out those SNPs that have more ancestry information, you can assay a smaller number of markers, obtaining a highly accurate estimate of someone's individual ancestry."

Dr. Shriver focuses in part on aiSNPs, which are loci with alleles that have very different frequencies across populations.

Dr. Shriver has created a mathematical model that associates the informative genetic variants with variation in human facial features. The 3-D face scans provide a physical representation of the face, which Dr. Shriver uses to create a multidimensional principal component face space. Each facial landmark, whether it is on the cheek bone or the tip of the nose, are given XYZ coordinates that define its location in 3-D space, and information about color and texture. Using more than 7,000 facial landmarks, a landmark mesh is created that can be overlaid on each person's face, providing the raw analytical data. His team then determines how these facial components inform the independent variables they are studying, such as sex, age, body size, genomic ancestry, and various individual SNPs.

"We've seen that we can estimate the age, body size, sex, and biogeographical ancestry of a person quite accurately from just the shape of their face," Dr. Shriver added. "Genetic and environmental factors play a large role in defining someone's facial features."

Trying to predict which genes are associated with which facial features is a complex task. Over 5 years ago, Dr. Shriver and his group teamed up with Peter Claes, PhD, an engineer by training working who is a researcher at the Catholic University of Leuven in Belgium. Dr. Claes had been working on a statistical analysis of facial variation throughout his scientific career. "We used his methods and my data together, but instead of starting with the traits and looking for the SNPs that encode them, we flipped the approach," Dr. Shriver stated "We focused on answering the question 'Does this SNP affect any part of the facial structure?' I think that approach is what helped us move this research forward successfully."

Real World Testing of the Facial SNP Analysis Method

Dr. Shriver first tested the method in a study of admixed West African populations from 3 countries, Brazil, the United States, and Cape Verde islands.¹ Throughout history, populations of humans have been

splitting apart and merging, creating a genetic mosaic in our DNA. Although some of this admixture is ancient (think of recent genetic evidence of interbreeding with Neanderthal populations),² other examples are more recent. "Different Hispanic groups have ancestry from the New World, the indigenous American populations who were here before the European Colonial Period, and European migration, as well as Africans who were forced to immigrate as part of the transatlantic slave trade," Dr. Shriver said.

"You can use molecular photofitting to analyze unknown DNA and generate investigative leads."

In this study of 600 people, Dr. Shriver's team identified 20 genes and 24 SNPs that affected facial features. However, they didn't know if the data could be used to construct a facial image of the subject successfully. In a subsequent study, Drs. Shriver, Claes, and Harold Hill at the University of Wollongong in Australia showed that there are methods to quantify to what extent a predicted face matches the actual facial image of a person.³ Dr. Shriver termed the process 'molecular photofitting,' further refining the term as 'indirect molecular photofitting' when only ancestry alleles are used, and 'direct molecular photofitting' when functional SNPs (SNPs that have causative effects or trait values) and ancestry alleles are used. Dr. Shriver sees important applications for both methods in forensic genomics.

"Often there is no suspect in a case, and you can't find a hit in a criminal database for the STRs you identify in a DNA evidence sample," Dr. Shriver stated. "When there's no hit and the case is probably going to go cold, what do you do then? You can use molecular photofitting to analyze unknown DNA and generate investigative leads. There have been at least 5 cases that I know about that were assisted by indirect molecular photofitting. Two of the cases involved serial killers in Louisiana and California."⁴

"The ForenSeq assay is already optimized for 1 ng of human DNA samples and interrogates hundreds of SNP markers simultaneously. Those are important first steps."

While the science that Dr. Shriver uses has advanced tremendously, it's premature for routine use in forensic investigations. "There will need to be peer-reviewed publications that describe the value and the limitations of these methods. A particular area of focus can and should be human perception studies on the usefulness of predicted faces in recognition tasks." Dr. Shriver said. "Also, I hope to see a common group of participants assembled so that various methods can be compared. I think that will come to pass as more labs around the world get involved in assessing molecular photofitting methods.

Then we'll bring the evidence to the forensic science community for evaluation, where they can look at the data, and assess the results and the overall method, as well as the mathematical formulas and practical value."

According to Dr. Shriver, the next step will be to add the most informative and confirmed facial markers into a next-generation sequencing (NGS) workflow like the Illumina MiSeq FGx System, and begin offering trait value estimates based on those markers. "Another crucial step will be making sure that these systems are optimized for very small amounts of DNA," Dr. Shriver stated. "In a typical forensic scenario, you're working with 1 ng of total material. The ForenSeq[™] assay is already optimized for 1 ng of human DNA samples and interrogates hundreds of SNP markers simultaneously. Those are important first steps."

"We've embarked on a long and interesting path of research," Dr. Shriver added. "I'm convinced based on our initial published and unpublished results that we will soon be able to make useful estimations of facial features from DNA samples."

References

- 1. Claes P, Liberton DK, Daniels K, et.al. Modeling 3D facial shape from DNA. *Plos Genet.* 2014; 10(3):e1004224.
- Fu Q, Hajdinjak M, Moldovan OT, et.al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature*. 2015; 10.1038/ nature14558.
- Claes P, Hill H, Shriver MD. Toward DNA-based facial composites: preliminary results and validation. *Forensic Sci Int Genet*. 2014; 13:208–216.
- Voss, K. DNA as a Sketch Artist: How Forensic Science Benefits from Physical Predictions. *Forensic Magazine*. 2015; May 15, 2015. Accessed August 18, 2015.

Learn more about the Illumina products and systems mentioned in this article:

- MiSeqFGx System
 www.illumina.com/systems/miseq-fgx.html
- ForenSeq DNA Signature Prep Kit
 www.illumina.com/products/forenseq-dna-signature-kit.html
- ForenSeq Universal Analysis Software www.illumina.com/informatics/research/sequencing-data-analysismanagement/forenseq-universal-analysis-software.html

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

© 2015 Illumina, Inc. All rights reserved. Illumina, ForenSeq, MiSeq FGx, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. Pub. No. 1470-2015-001 Current as of 09 November 2015

