

DNA apps promise deeper insights for consumers – but at what cost?

May 21 2018, by Sarah Catherine Nelson



Credit: AI-generated image ([disclaimer](#))

Last month, law enforcement in California [apprehended the alleged Golden State Killer](#), a serial murderer and rapist who terrorized the state during the 70s and 80s. Investigators triangulated to a suspect using a combination of DNA evidence left at crime scenes and publicly available genetic genealogy websites, [most notably GEDMatch](#).

Quite provocative in its own right, the incident has sparked [much discussion and debate](#) about genetic privacy, ownership of genetic data and appropriate uses of DNA evidence, and database searching by law enforcement. [As a researcher studying](#) personal access to "raw" or uninterpreted genetic data and the ways people use it on third-party interpretation websites such as GEDMatch, I'm glad to see the public focusing on a few key distinctions in the world of personal genetic testing. There are notable but often underappreciated differences between the more well-known consumer testing companies and the third-party sites that promise users further interpretation of their data.

After you get your DNA data ...

Direct-to-consumer (DTC) genetic testing has been relatively mainstream since around 2007. The two largest companies, 23andMe and AncestryDNA, have collectively amassed [over 10 million customers](#). For a [cost of around US\\$100](#), DTC companies have you spit into a tube from which they extract your DNA. They focus on a predefined set of between half a million and a million sites – what geneticists call base pairs (the coupled As, Cs, Ts and Gs) – in your genome. This subset is less than 0.05 percent of the whole genome; [few companies offer to sequence the entire thing](#).

Depending on the [company](#), this genetic information is then used to generate reports about genetic ancestry, physical traits, disease predispositions or relatedness to other customers in the company's database. Most DTC companies also allow customers to download a file of their "raw" or uninterpreted genetic data. Essentially this is a long text file that lists all the sites genotyped by the company and the customer's DNA sequence at those sites.

So now DTC customers have raw data files sitting on their hard drives that they may want to further leverage but aren't sure how. Enter third-

party interpretation websites run by other companies, entrepreneurs and citizen scientists. These online tools enable users to pursue further interpretation and analysis of their raw data files – perhaps also contributing to additional research. There are currently dozens of such tools, and the list seems to be growing.

If the past decade of consumer genomics has been a music festival, the large DTC companies have been playing the main stage, while third-party tools have been off performing smaller venues. They might have adamant fans and a loyal following, but they just haven't been on the general public's radar to the same extent as DTC testing. Similarly, there's a [small amount](#) of [research on third-party tools](#) compared to how much academic attention has been paid to DTC testing.



Third-party tools take the raw DNA further than the first two parties: the individual who provided the DNA and the direct-to-consumer genetic testing company. Credit: Montage by Sarah Catherine Nelson

This group of apps and sites is also a very disparate genre. While you can parse them according to the primary categories of information DTC companies return – health, ancestry and relatives – they're more varied in scale, modes of operation and other functionalities, like whether they're connected with citizen science research initiatives.

Spit versus genetic data

















One key difference between DTC companies and third-party websites is what you submit to them and how. Most DTC companies collect customer DNA samples via mail-in spit kits. Producing a milliliter or two of spit is surprisingly difficult – it's definitely not something you're going to do without noticing. Indeed companies settled on this method of sample collection in part [to avoid surreptitious testing](#) – meaning it's unlikely someone can send in your sample without your knowledge.

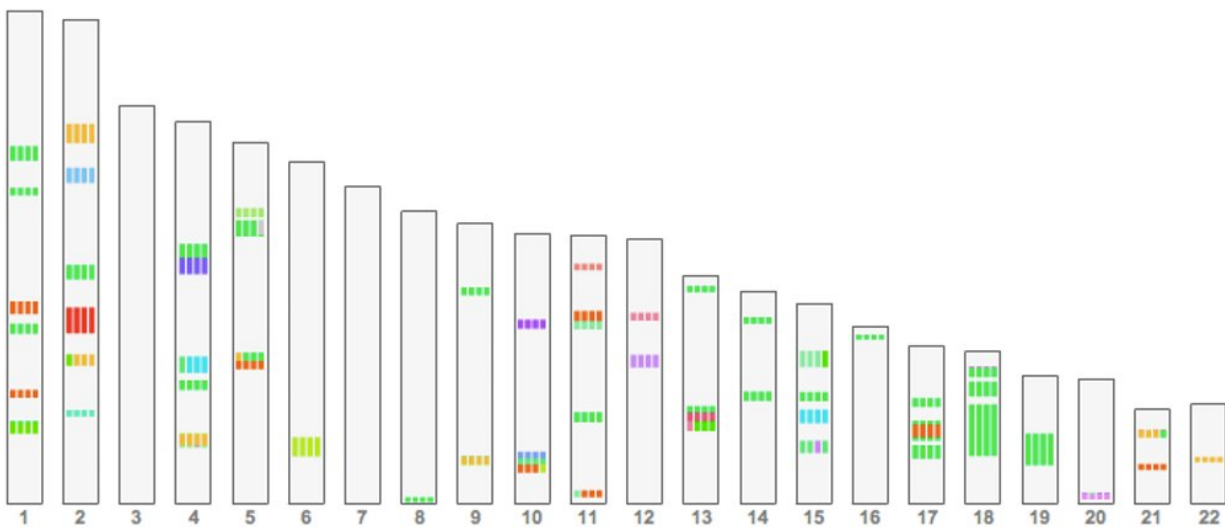
The spit is the physical material from which the DTC companies extract DNA molecules that they then genotype to produce a file with the customer's raw genetic data. Once encapsulated in a file, genetic data is quite portable and more difficult to control. You can email it, upload it or copy it onto thumb drives and ship to your 10 closest friends. Compared to spitting in a tube, you may not know if someone obtains and uploads your data file to a third-party site.

What third-party interpretation websites actually do with users' genetic data files varies widely and is especially relevant to the Golden State Killer case. Some tools – one is Promethease – delete user data after a fixed and rather short period of time. The site holds onto it just long enough to generate the health-related report. Other tools never require data to leave the user's own computer or device; instead, the analysis happens locally – for instance, GENOtion/Interpretome, David Pike's utilities and the DNA Doctor smartphone app.

For other tools, retaining users' data long term is central to the operation and mission. Genealogy-focused tools such as GEDMatch need to keep data to allow new users to search across the database for relatives – as is also the case with the DTC companies that offer a relative-finding service. Tools such as DNA.Land retain user data in part to conduct research studies. But even these tools that keep user data are not public

or open databases; one can't directly download or access other users' raw genetic data. A notable exception is the [open science project openSNP](#), which explicitly makes all user data freely available for download in the interest of making science more accessible and collaborative.

Country	Color	Percent of Steve Jurtvetson's Genome Covered
 Finland		3.9% – 4.2%
 Estonia		0.5% – 0.7%
 Poland		0.5% – 0.5%
 Norway		0.4% – 0.6%
 United Kingdom		0.2%
 Italy		0.2%
 Russia		0.1% – 0.5%
 Ukraine		0.1% – 0.3%



Once the genetic information has been extracted from a biological sample, the question of ‘ownership’ changes. Credit: Steve Jurtvetson

Physical versus informational property

Another striking feature of genetic data is that once it's been extracted from a physical sample it can be used in many places at once. 23andMe can have my genetic data and use it for research; I may also have it and use it in as many third-party sites as I choose. In a legal sense, the genetic data can be thought of like other informational or intellectual property: Its consumption is "[non-rivalrous](#)," meaning one person's use does not interfere with another's.

These features make genetic data different from physical property, which typically can't be used by many people at once and may be exhausted over time. People often overlook that the bundle of property rights that could be considered for a physical specimen of DNA – a sample of spit or the DNA molecules extracted from it – are different than those of a genetic data file. [Legal precedent holds](#) that individuals do not have property rights in their biospecimens. Potential ownership of personal genetic data, however, is really only now beginning to be [tested in the courts](#).

Regardless of ownership questions, it's difficult to maintain control of how genetic data is used in the distributed system of DTC genetic testing and third-party tools. Joseph DeAngelo didn't submit his sample to a DTC company or upload his data to a third-party site. Instead, law enforcement used abandoned material to generate a genetic data file in the same format that one would download from a DTC company. But some of his more distant relatives apparently did do DTC testing. DeAngelo didn't upload his [genetic data](#) to GEDMatch – investigators did – but some of his relatives uploaded theirs, allowing police to eventually trace their way to him.

Leaping before looking

What genetic and personal information are people exposing when they upload it to third-party sites? There is no blanket answer, as these tools

differ markedly in what they offer and what they do behind the screen.

The bottom line is that third-party interpretation websites are a heterogeneous bunch. Most lack the large legal teams and technical management systems of the major DTC companies. And currently there's not a clear regulatory framework for third-party tools. Indeed, many of the developers [view their activities as simply connecting](#) users to pre-existing genetic annotation sources and therefore not warranting stricter government oversight. The lack of regulatory clarity largely leaves it to individuals to do their due diligence before using any of these tools and services. We as a society are learning the hard way. When we aren't paying (much) for a service, we and our data are the product.

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

Citation: DNA apps promise deeper insights for consumers – but at what cost? (2018, May 21) retrieved 23 April 2024 from <https://medicalxpress.com/news/2018-05-dna-apps-deeper-insights-consumers.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--