What Is This DNA Stuff and Why Should I Care?

Part 5: Autosomal DNA (atDNA) - Cousin Matching

By David R. Neithamer

In the previous two articles (Parts 3 and 4) we discussed mitochondrial DNA (mtDNA) and Y-chromosome DNA (Y-DNA), which are passed down the maternal line and paternal lines respectively. Each one can be very useful for answering specific genealogical questions such as whether two females are maternally related or whether to two males are paternally related. However, neither test is useful for finding random genetic cousins because exact matches in either test may be several generations or hundreds of years in the past. In this article we will discuss autosomal DNA (atDNA).

Humans have 22 pairs of autosomes that have the same morphology, and these pairs are arbitrarily labeled with numbers (1-22) roughly in order of their sizes in base pairs (bp) as shown in **Figure 1**.

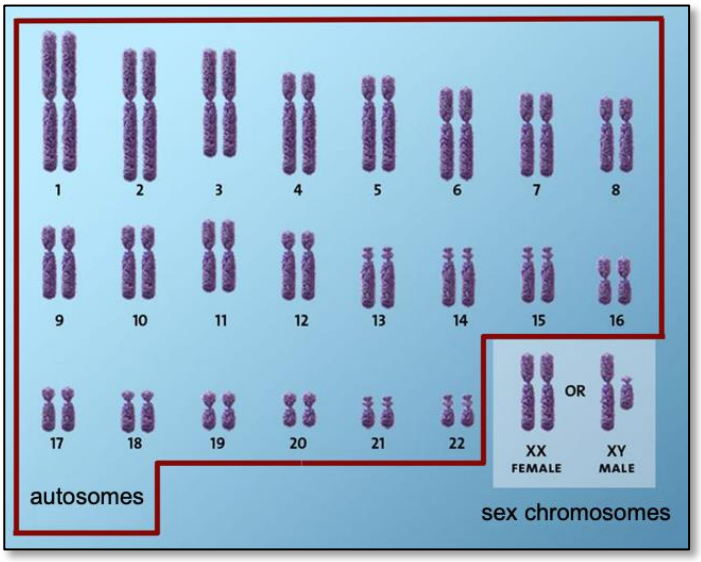

Figure 1. Karotype of 46 Human Nuclear DNA Chromosome
Source: https://medium.com/analytics-vidhya/the-last-explanation-yo will-ever-need-to-answer-this-conditional-probability-question-f39f8e93086c. Analytics Vidhya Article by Gal Gilor. Fair Use.

The autosomal DNA test is a Single Nucleotide Polymorph (SNP) test done on all 22 pairs of chromosomes. Autosomal DNA makes up 94% of all human DNA found in our cells which makes this an ideal test to determine random genetic cousins because the test looks at a much greater amount of DNA than is used for the mtDNA or Y-DNA testing. All four major testing companies also test some X-chromosome SNPs as part of their atDNA test and three of the companies (Ancestry, MyHeritage, and 23&Me) also test Y-chromosome SNPs as part of their atDNA test. Typically, about 630,000 - 700,000 SNPs are tested across the entire human genome, which may seem like a lot, but this only accounts for about 0.02% of our total DNA.

Both males and females can test and compare their autosomal DNA because it is inherited from both parents and this makes it the best test for finding genealogically related cousins. However, the atDNA test does have its limitations, as the results become less reliable the further back in time you go. Theoretically, you have atDNA from all your ancestors. However, some DNA is lost (white regions) at each generation due to chromosomal recombination as shown in **Figure 2**.

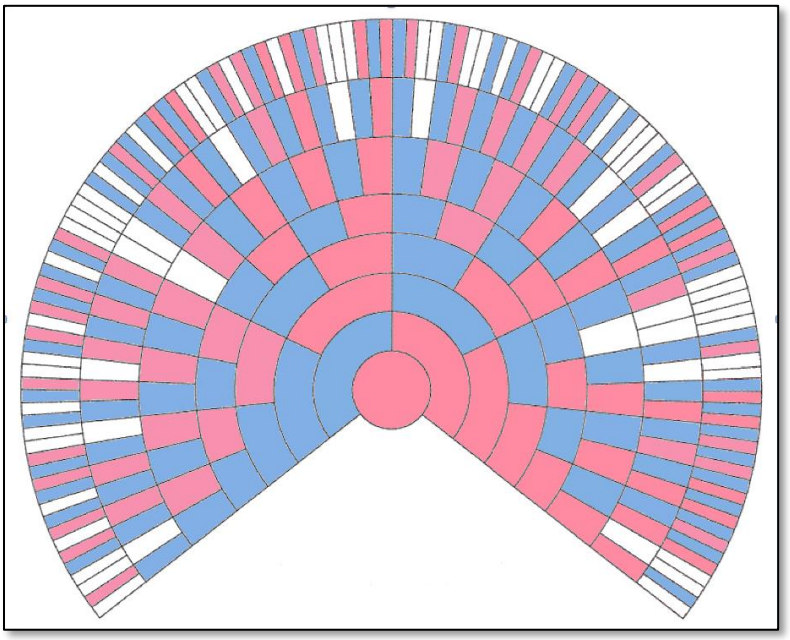

Figure 2. Representative Inheritance Pattern for Autosomal DNA Showing the Loss of Inherited Autosomal DNA Through Subsequent Generations Due to Recombination
Source: Generated by the author from similar fan charts in "Guide to DNA Testing and Genetic Genealogy" by Blaine T. Bettinger, 2016

Recombination occurs when chromosomes entangle during meiosis division and homologous DNA strands crossover or swap with each other. After crossover, the chromosomes separate to form new daughter cells (sperm and ovum). Recombination is what leads to the large variety within humans, even between siblings.

At 4 generations back there is a 96% chance that all 16 gg-grandparents are represented in a person's autosomal DNA. At 5 generations back, this drops to a 54% probability that all 32 ggg-grandparents are represented in one's atDNA. However, at 6 generations back, there is only a 0.01% chance that, all 64 gggg-grandparents are represented. For this reason, individuals with a most recent common ancestor (MRCA) 5-6 generations back in time may not show an atDNA match even with an actual cousin. In other cases, some cousins with an MRCA 7-8 generations back may be identified. The test is therefore only useful for finding cousin matches from an MRCA about 150-200 years in the past.

In order to estimate how closely two individuals are related, geneticists have developed a unit of recombination frequency, called the centimorgan (cM), which is used to measure the amount of shared DNA. It is used to imply distance along a chromosome and takes into account how often recombination occurs in a region. On average, 1 cM corresponds to about 1 million base pairs in humans. Generally, matches less than 7-8 cM are likely random matches (i.e. not related).

Figure 3 provides a chart with the anticipated atDNA relationships between an individual and their family relative, with the average anticipated cM values for each, as well as the possible low and high cM value ranges (at the 99% percentile). Starting with yourself (SELF), a first cousin once removed (1C1R) would have an average cM relationship value of 439 cM, but it could also be within the range of 141-851 cM. Similarly, a fourth cousin (4C) would have an anticipated relationship value of 35 with a range of 0-127 cM. Note, a value of 0 cM indicates no shared atDNA and this would be an example where an actual cousin would share no atDNA with you.

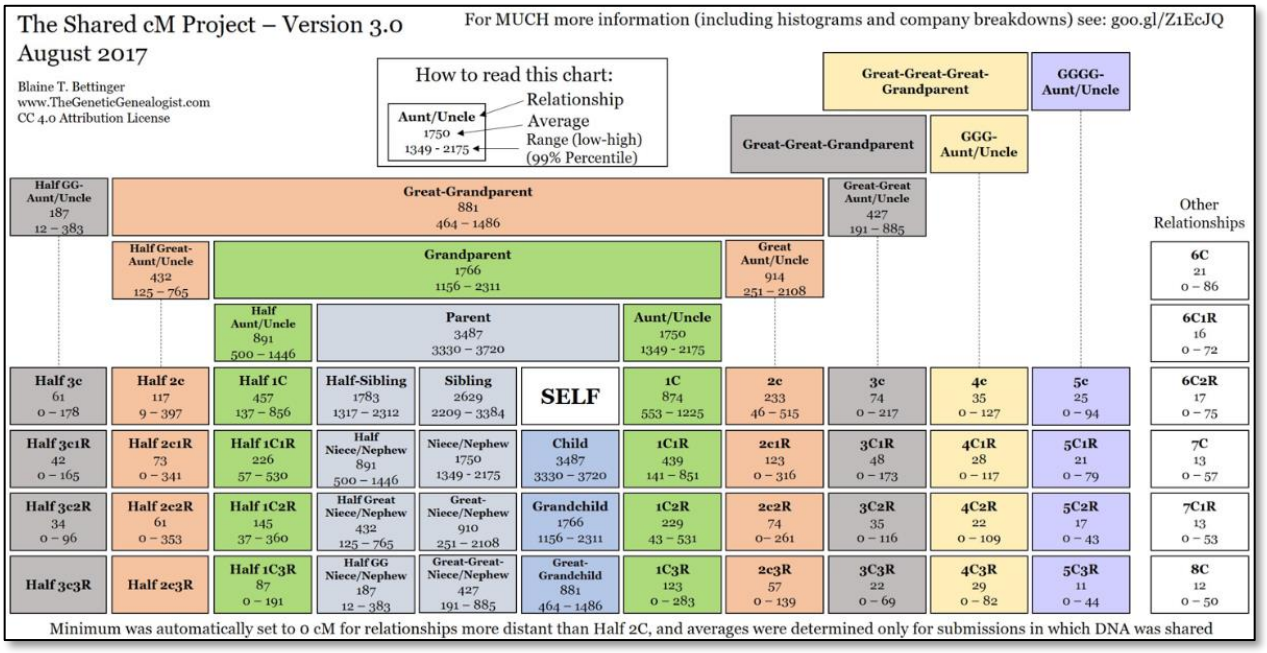

Figure 3. Relationship Between Matching Individuals and Their Anticipated atDNA Centimorgan (cM) Values
Source: TheGeneticGeneologist.com, Blaine T. Bettinger, 2017. CC 4.0 Attribution License
https://thegeneticgenealogist.com/2017/08/26/august-2017-update-to-the-shared-cm-project/

All four of the major testing companies provide atDNA testing. They all provide an ethnicity estimate in an attempt to identify where your ancestors originated, and they all identify your cousin matches based on your shared atDNA. All of them also now provide some information about inherited genetic traits, health information, and potential carrier status although all of these health-related services require additional fees. 23&Me is the best option if you are interested in health-related matters as opposed to genealogical investigations.

The ethnicity estimate compares your DNA with a series of reference populations from around the world to try and determine where your ancestors came from. However, because the reference populations vary from company to company and the algorithms used to calculate these values vary, the ethnicity results can vary widely. For those of European descent, it is difficult to differentiate British Isles, Germanic, and Scandinavian descendancy due to influence of the Norman, Anglo-Saxon, and Norse tribes who invaded England at various times in the past. For those of Middle Eastern, African, and Asian descendancy the results can vary greatly due to fewer individuals testing and to smaller numbers of reference populations. However, this will likely change with time as the companies are continually updating their reference populations. The current reference panels from each company are as follows: Ancestry sub-divides the world into 70 population regions using 44,703 reference DNA samples; FTDNA uses 90 regions from 8053 DNA samples; MyHeritage uses 42 regions from over 5000 samples; and 23&Me uses 45 regions from over 14,000 DNA samples.

Figure 4 shows the ethnicity estimates from my four autosomal DNA tests taken at each of the four major testing companies. As you can see, they vary considerably. Ancestry shows 89% ancestral origins from the British Isles compared to 96% from MyHeritage, 32% from FTDNA, and 38% from 23&Me. Similarly, Ancestry shows only 9% Western European compared to 46% from FTDNA, 61% from 23&Me, and surprisingly, 0% from MyHeritage. Furthermore, FTDNA showed a relatively large

amount of Scandinavian and Iberian (Spanish) influences whereas the other companies showed only minor amounts if any. The main point of this is to take their ethnicity estimates with a grain of salt.

<table>
<tr><th></th><th>Ancestry</th><th>MyHeritage</th><th>FTDNA</th><th>23&Me</th></tr>
<tr><td>England</td><td>43%</td><td>72%</td><td rowspan="3">21%</td><td rowspan="4">38%</td></tr>
<tr><td>Scotland</td><td>31%</td><td rowspan="3">24%</td></tr>
<tr><td>Wales</td><td>14%</td></tr>
<tr><td>Ireland</td><td>1%</td><td>11%</td></tr>
<tr><td>Germanic Europe</td><td>9%</td><td>0%</td><td>46%</td><td rowspan="3">61%</td></tr>
<tr><td>Switzerland</td><td>0%</td><td>0%</td><td>0%</td></tr>
<tr><td>France</td><td>0%</td><td>0%</td><td>0%</td></tr>
<tr><td>Scandinavia</td><td>2%</td><td>0%</td><td>6%</td><td>0%</td></tr>
<tr><td>Iberian</td><td>0%</td><td>4%</td><td>9%</td><td>0%</td></tr>
<tr><td>Italian</td><td>0%</td><td>0%</td><td>7%</td><td>1%</td></tr>
</table>

Figure 4. Comparison of Autosomal DNA Ethnicity Estimates
Source: Results of the Author's Autosomal DNA Tests at the Four Major Testing Companies

Cousin matching is the most interesting and beneficial feature of the DNA tests, at least from a genealogical standpoint. All of the testing companies have developed a number of tools to help identify related cousin matches but not all of them have all of the same tools available. We will talk more about this in Part 7, which will compare the four testing companies. However, a brief mention of the more popular tools available to identify cousin matches via atDNA is appropriate.

All four companies have the ability to see cousin matches ranked by predicted relationship as determined by the shared centimorgan (Cm) values (Ancestry, FTDNA, MyHeritage) or % Shared DNA (23&Me). Ancestry and FTDNA have the ability to compare matches and then see how they compare to a third shared match. This is a way to manually cluster shared matches and see which matches are related to each other. MyHeritage has improved on this by developing an automated clustering tool. FTDNA, MyHeritage, and 23&Me all have a chromosome browser which is very useful to visually compare shared DNA regions between matches and triangulate cousin matches. We will discuss triangulation more in Part 8 when we discuss GedMatch.

Ancestry, MyHeritage, and FTDNA allow importing of a family tree to link to their DNA data results; 23&Me does not, but they do allow you to link their DNA to an external family tree. Ancestry and MyHeritage have also developed useful tools that automatically evaluate both the DNA results and uploaded family trees in an attempt to identify how matches may be related; the tools are called ThruLines and SmartMatch respectively. All of the companies can export their raw atDNA data files, but only MyHeritage and FTDNA can upload the other companies DNA files to their website which allows test takers to look for additional cousin matches without having to pay for new tests. However, GedMatch.com is a website to which you can upload your raw DNA files from any of the testing companies and can look for additional cousins who have also uploaded their testing results. GedMatch has a free set of tools which are useful in identifying cousins but for $10 per month you can access additional tools which can make researching your cousin matches much easier and more informative. I will talk about GedMatch in depth in the final article (Part 8) of this series.

Since autosomal DNA is the preferred method to identify genetically random cousins, it has also been used to find unknown family members, parents of adopted children and, in forensics, to identify, arrest,

and convict wanted criminals. One of the most famous recent examples of the use of autosomal DNA is the identification, arrest, and conviction of the Golden State Killer who was responsible for at least 13 murders, 50 rapes and 120 burglaries in California from 1973-1986 but was never caught. As DNA testing technology improved, the cold case was reopened, and in 2018 a frozen DNA sample from a past crime scene was tested using autosomal DNA and the results were uploaded to GedMatch to look for related cousins. Approximately 10 to 20 distant cousin matches of the Golden State Killer were identified and traced to the same great-great-great grandparents (MRCAs). For 4 months, a team of five investigators constructed a large family tree (>1000 names) of all the descendants from these two individuals.

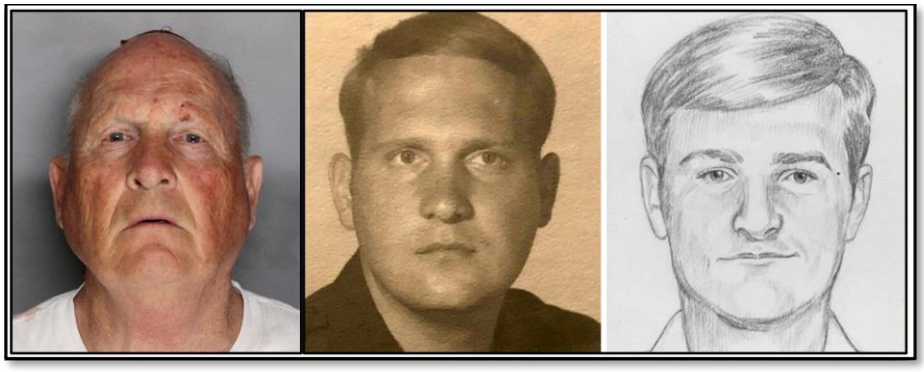

Figure 5. Golden State Killer Suspect Joseph James DeAngelo Identified by Autosomal DNA
Source: Associated Press

The detectives identified two potential suspects in the case, one of whom was ruled out by a relative's DNA test, leaving only Joseph James DeAngelo (**Figure 5**), a former police officer, as the main suspect. A DNA sample was surreptitiously collected from the door handle of DeAngelo's car which showed a consistent match. A second sample was then collected from a tissue found in DeAngelo's curbside garbage. Standard STR Y-DNA testing was then used to compare the DNA from 2 archived rape kits to DeAngelo's Y-DNA. Both samples matched and conclusively proved that Joseph James DeAngelo was the Golden State Killer and he was arrested and ultimately convicted.

In summary, autosomal DNA (atDNA) refers to the 22 pairs of chromosomes, called autosomes, found in the nucleus of all human cells. A child inherits approximately 50% of their atDNA from their father and 50% from their mother. Autosomal DNA testing analyzes between 630,00-700,000 SNPs along the 22 pairs of autosomes as well as some SNPs along the X and Y sex chromosomes depending on the testing company. Due to recombination (swapping of genetic material during meiosis division) not all genetic cousins share their atDNA. Beyond 5th-6th cousins, shared DNA drops dramatically and eventually is lost. All of the testing companies calculate an ethnicity estimate; however, the results will vary dramatically since the companies analyze their SNP data using different algorithms, methodology, and reference populations. atDNA results are best used to fish for random genetic cousins and approximate how matches are related based on the number of shared DNA fragments as measured in centimorgans (cM). atDNA can also be very useful for identifying a MRCAs as long as the individuals link their testing results to a well-documented family tree. Unfortunately many do not. Finally, autosomal DNA is the preferred testing method to locate unknown family members, parents of adopted children, and in forensics to identify and convict criminals.

This concludes our discussion of autosomal DNA. The next article will discuss X-chromosome DNA (X-DNA), which has a unique inheritance pattern that limits its ability to be used as a stand-alone genealogy test; but X-chromosome SNP testing is done as part of the autosomal test and I will discuss it for completeness. Until then, happy genealogical hunting!