Introduction to the Use of Autosomal DNA Testing to Solve Genealogical Research Questions

Tim Janzen, MD

E-mail: tjanzen@comcast.net

Common Terms Pertaining to Autosomal DNA that are used in Genetic Genealogy

Single Nucleotide Polymorphism (SNP, pronounced snip): a DNA sequence variation occurring when a single nucleotide — A, T, C or G — in the genome (or other shared sequence) differs between members of a biological species or paired chromosomes in an individual. For example, two sequenced DNA fragments from different individuals, AAGCCTA to AAGCTTA, contain a difference in a single nucleotide. In this case we say that there are two alleles: C and T. Almost all common SNPs have only two alleles.

Identical by Descent (IBD): when half-identical regions (HIRs) in two people's autosomal DNA or on the X chromosome match (neglecting rare mutations and testing errors) due to the fact that the two people share a common ancestor and thus share a DNA segment that was passed down to both of them from that common ancestor.

Identical by State (IBS): when half-identical regions (HIRs) in two people's autosomal DNA or on the X chromosome match by coincidence. When two individuals share a half-identical region without being related, those results are identical by state.

Half Identical Region (HIR): a region of two paired chromosomes where at least one of the two alleles from one person's pair of chromosomes matches at least one of the two alleles from a different person's pair of chromosomes throughout the entire region. A half identical region may be either identical by descent (IBD) or identical by state (IBS).

Centimorgan (cM): a unit of measure of genetic recombination frequency; one cM is equal to 1% chance that a marker at one genetic locus will be separated from a marker at another locus due to crossover in a single generation; 1 cM is on average about 1 million base pairs.

Phasing: the process of determining which allele values in an unphased autosomal DNA SNP dataset came from one parent and which came from the other parent. See http://www.isogg.org/wiki/Phasing for more background.

Quantitative analysis. This is the process whereby one compares the amount of autosomal DNA in cMs that two or more people share in common to see if it meets the amount expected. This technique can be safely used going back 4 to 5 generations.

Triangulation. This is process of reviewing the pedigree charts of people who match on the same autosomal DNA segment to see if a common ancestor can be found. This technique can be used going back many generations.

Autosomal DNA Testing Basics
Parents/children share 3552 autosomal cMs
Siblings share 2600 cMs on average
Aunts, uncles/nieces, nephews share 1776 cMs on average
First cousins share 888 cMs on average
Second cousins share 222 cMs on average
Third cousins share 55 cMs on average
Fourth cousins share 14 cMs on average
Fifth cousins share 3.5 cMs on average
Sixth cousins share 0.76 cMs on average

Ranges of total cMs of IBD segments based on family relationship
Parent/child: 3539-3748 cMs
1st cousins: 533-1379 cMs
1st cousins once removed: 115-851 cMs
2nd cousins: 46-515 cMs
2nd cousins once removed: 0-325 cMs
3rd cousins: 0-217 cMs
3rd cousins once removed: 0-173 cMs
4th cousins: 0-127 cMs
5th cousins: 0-94 cMs

Odds of matching someone
1st cousin or closer: 100%
2nd cousin: >99.9%
3rd cousin: ca 90%
4th cousin: ca 50%
5th cousin: ca 15%
6th cousin: ca 1%

Resources for information about Autosomal DNA
https://isogg.org/wiki/Autosomal_DNA_testing_comparison_chart (provides information about the autosomal DNA tests offered by the major companies)
http://isogg.org/wiki/Portal: Autosomal_DNA
http://isogg.org/wiki/Autosomal_DNA_statistics
https://www.familytreedna.com/learn/autosomal-genealogy-matching
https://isogg.org/wiki/Beginners%27_guides_to_genetic_genealogy

Popular blogs covering genetic genealogy
Blaine Bettinger: https://thegeneticgenealogist.com
Roberta Estes: https://dna-explained.com
Jim Barlett: https://segmentology.org

Kitty Cooper: http://blog.kittycooper.com
Debbie Kennett: https://cruwys.blogspot.com
CeCe Moore: http://www.yourgeneticgenealogist.com
Emily Aulicino: http://genealem-geneticgenealogy.blogspot.com
Shannon Christmas: http://throughthetreesblog.tumblr.com

Books about DNA Testing and haplogroups
Genetic Genealogy in Practice by Blaine Bettinger and Debbie Parker Wayne, 2016.
The Family Tree Guide to DNA Testing and Genetic Genealogy by Blaine Bettinger, 2016.
NextGen Genealogy by David Dowell, 2015.
Genetic Genealogy, The Basics and Beyond by Emily Aulicino, 2014.
DNA and Social Networking by Debbie Kennett, 2011.
Trace Your Roots with DNA by Megan Smolenyak and Ann Turner, 2004.
DNA and Family History: How Genetic Testing Can Advance Your Genealogical Research by Chris Pomery, 2004.
Unlocking Your Genetic History by Thomas Shawker, 2004.
The History and Geography of Human Genes by L. Luca Cavalli-Sforza, Paola Menozzi, and Alberto Piazza, 1994.
The Journey of Man by Spencer Wells, 2002.
Deep Ancestry by Spencer Wells, 2006.
The Seven Daughters of Eve by Bryan Sykes, 2001.

Adoptee resources
http://www.dnaadoption.com
Richard Hill's website: http://www.dna-testing-adviser.com

Genetic Genealogy Discussion Groups

RootsWeb Genealogy-DNA list http://archiver.rootsweb.com/th/index/GENEALOGY-DNA
Androgenica: http://www.anthrogenica.com/activity.php
FTDNA http://www.familytreedna.com/forum
ISOGG Newbies: http://groups.yahoo.com/group/DNA-NEWBIE
ISOGG list: http://groups.yahoo.com/group/ISOGG
Genetic Genealogy Tips and Techniques on Facebook:
https://www.facebook.com/groups/geneticgenealogytipsandtechniques
ISOGG on Facebook: https://www.facebook.com/groups/isogg
DNA Detectives on Facebook: https://www.facebook.com/groups/DNADetectives

Creating your master match list spreadsheet
1. Download your family's matching segment data (HIR data) from the companies that provide chromosome browsers and from GEDmatch using the Tier 1 utility Matching Segment Search.
2. Integrate all of this data into one Excel file.
3. Sort your Excel file first by chromosome, then by whether or not one of your children shares the same segment as your match, and then by start position.
4. Merge duplicate data from the same people.

Keeping track of all pieces of information that you would like to monitor and record regarding the people who share autosomal or X chromosome DNA segments with you is a challenging task. I prefer to keep all of my relatives' matches' data in Excel files. I find that this offers me maximum flexibility in terms of sorting and organization. I create a matching segment data file for each relative who is in the oldest generation. If data is redundant, then I omit the data from some Excel files. For instance, I maintain matching segment data files for my mother and her two brothers. If one of my mother's brothers shares a DNA segment with her then I don't record the people who share that segment with my mother in her brother's matching segment file. I have a priority ranking system in this regard. If my mom's oldest brother has a DNA segment that my mom doesn't have then any matches on that segment are included in his matching segment file. If both of my mom's brothers share a segment that my mother doesn't have, then I record the information about the matches only in the matching segment file for my mom's oldest brother. My mom's youngest brother's matching segment file only contains information about matches on DNA segments that he has but neither his brother nor my mother possess. This approach reduces redundancy in the matching segment files. In theory, I could combine the matching segment files for my mom and her two brothers, but I have decided to keep them all separate at this time. As you organize your matching segment data you may want to take a similar approach.

In addition to including the matching segment data and contact information in my master match list spreadsheets, I like to include columns for the following information:
1. Ancestor the shared DNA segment was inherited from is on based on chromosome mapping. I put the surname of the ancestral line or the specific ancestor to which I have mapped the HIR in question to in this column.
2. Profile. This would be the URL for their profile at the company where they tested.
3. GEDmatch number, if they have uploaded to GEDmatch.
4. Pedigree. I put a "yes" in this column if I have a pedigree on file for the match in question.
5. Pedigree chart URL. If the person has a pedigree on the internet somewhere I include that URL in this column.
6. Genealogical connection. If I have found a genealogical connection to the person in question I put this information here. Otherwise add any shared surnames or ancestral locations.
7. Message fields. I add multiple columns on the right side of the spreadsheet include the messages I have exchanged with the match. All important messages are given their own cell. I also record the date of the message in the same field.

The McCleary Case Example

A. K. was tested and transferred his data to GEDmatch. When I ran a comparison between A. K. and 9 other great-great-grandchildren of Moses McCleary (b. 1815, d. 1865), A. K. shares an average of 47.4 cMs with those 9 people. Third cousins should share about 55 cMs on average with each other and that 3rd cousins once removed should share about 27.6 cMs with each other on average, suggesting a 3rd cousin or 3rd cousin once removed relationship between A. K. and the McCleary cousins. A. K. had no information suggesting that he was a descendent of Moses McCleary, but subsequent additional genealogical research established that A. K. was a 3rd cousin once removed to Betty Janzen.