Discovering Common Ancestors:
How (and why) to work-up your DNA matches

GenealogyKC- September 8, 2023
Susan E. Mundt

©2023 Susan E. Mundt

Why do genetic genealogy?

The International Society of Genetic Genealogy (ISOGG by GNU Free Documentation Licence) notes that a DNA match with an individual has the potential to:

- verify existing research

- establish that two surname variants are related

- provide locations for further genealogical research

- help determine the ancestral homeland

- discover living relatives

- confirm or deny suspected connections between families

- prove or disprove theories regarding ancestry

But, we need to start with reasonable expectations:

Genetic Genealogy will not "do" your tree for you. You shouldn't expect to be able to confirm genetic inheritance for relatives beyond your 4G Grandparents/5th cousins with autosomal testing. Try to build and share your tree back to at least 3G grandparents to benefit the most from DNA testing.

How does this work, anyway?

The science of genetic genealogy is based on how much DNA is shared between you and your matches. The number of segments of DNA shared and the size (cM) of the segments is behind relationship prediction.

Your testing company will give you an estimate of how you might be related to a DNA match. I'd suggest going a step further and deepening your understanding by using a free tool online, the shared centimorgan tool.

https://dnapainter.com/tools/sharedcmv4

This project can help you understand the ranges of shared cM associated with various known relationships. A centimorgan, or cM, is the measure for a DNA segment. This measurement will always be given for your DNA matches. I noted the shared cM creators, Blaine Bettinger, Jonny Perl and Leah Larkin, who collected, organized and shared the project data.

After using this simple, but powerful, tool for a bit you'll have a better understanding of how DNA matching works. In brief, the cM you have in-common with your matches only makes biological sense for certain relationships. A 3500 cM match can ONLY be your child or parent. Smaller matches can be due to a broader range of relationships.

The inverse is also true, a first cousin should NOT match you less than 390 cM or more than 1400 cM; if they do you should re-evaluate the trees- something is off!

Labeling:

So, the above was all groundwork to help you use your matching information and start labeling to figure out your matches.

The key to beginning to find *unknown* ancestors is to label your *known* cousins.

Surprisingly you can often (not *always*, but often) get your four grandparental groups sorted *even when you don't know who one or more of the four groups is!*

I outlined the labeling systems available using Ancestry as my example. MyHeritage has both a notes and labeling system, FamilyTree DNA has just a notes system. I admit I know less about 23 and Me..

Ancestry provides three ways to label matches:

1. Relationship- toggle through the "do you recognize them" options
2. Notes
3. Dot system- our main sorting tool!

Dots/Groups:

There is no single "right" way to use the dots, the right way is the one that makes sense to you- so that you will use it!

Some people give each couple a label, an identified match would have dots for all the shared ancestral couples (OR for just the closest shared couple).

Some people label by surname line, each match identified would generally have two dots, one for each partner in the closest shared ancestral couple. This is the system I use.

Give some thought to color selection and plan ahead. I like color families for each line and tend to do a lighter hue for maternal lines. Whatever makes sense to you! Only 24 label choices at Ancestry, MyHeritage has 30.

These labels aren't for filing away and forgetting. They are a tool to *sort* your closer matches and help identify your more distant and/or unknown matches (who may hold the key to that mystery in your tree!).

If you use labels consistently, a match with no tree, or who has a tree where no common ancestor is apparent, can often still be assigned to a general group!

Where do we start? Lead the way (this is a pun, you'll see…). We start by focusing on our matches between 90 cM and 400 cM. This cM range should flag nearly all our second cousins (and some 3C, 4C, 1C1R, etc.).

The Leeds method:

The Leeds Method is a way to take your matches (on *any* DNA matching system) to identify your four grandparent quadrants- paternal grandfather, paternal grandmother, maternal grandfather, and maternal grandmother.

Another way to put this, you are attempting to ID your 2nd and 3rd cousins. This can be an ideal first step to beginning to label your matches.

See Dana Leeds' webpage for more details

https://www.danaleeds.com/the-leeds-method/

The basics of the Leeds method:

You will be sorting your match list with ALL relatives (not just maternal or paternal). Sort from best/closest match to smallest. **You only want to add matches to your list who are between 400 cM and 90 cM.** This should sort out MOST of your relatives more closely related than 2nd cousins and most of those who are more distantly related than 3rd cousin.

For those of us who are over age 50 there may be a handful of first cousins twice or thrice removed in the 400-90 cM group. These are your first cousin's grandkids and great-grandkids. If these matches are included on your list they can lead (no pun intended!) to a messy result. Dana doesn't cover this much in her blog- I'll add some hints later to avoid this complication.

Getting started with Leeds:

We can use any spreadsheet- Excel, Google sheets or plain old lined paper and colored pencils!

From largest to smallest, list your 400 cM to 90 cM matches. You don't need full names, just enough to ID the match when you look back at your results.

If you don't have at least four to six matches in this range you are at a disadvantage. To generate this list: on your DNA matches page select "All Matches". Sort by relationship OR choose shared DNA and set the range to 90 to 400. 23 and Me Users: this is approximately 1.3% to 5.9% of shared DNA- you can get a conversion from percentage to cM for matches on 23 and Me. The conversion isn't perfect but will be close enough to use.

Identify Group 1

Select a color for your first match on the list (use the "tipping paint can" on the tools bar).

Then give the same color to all the shared matches with this "best match" that are on the rest of the list. This is your group 1.

As you mark the list it may become apparent to you which grandparent line this is.

Go to the next best match on your list that did NOT match your group 1 primary person. That starts your group 2. Once again look at shared matches and highlight them.

I like this process when working-up a new tester. It can give you a quick look at what lines may be more difficult because they have few close matches.

If you only get three groups, it may hint at a double relationship or recent cousin marriage muddling the process- or a bad choice of "best match". It is not unusual to get more than four groups- this can happen when your start person for the group is more distant than a second cousin and is related to only "one side" of that line (because they are a 3C, for instance).

Cleaning-up your list before you start the Leeds color system:

You begin your list with your closest match under 400 cM. You want to make sure this person is NOT a first cousin's descendant- because they will match **two** of your grandparents and you are trying to sort into four discrete grandparent groups.

Move any known first cousin's descendants to the bottom of the list (you may later just delete them, I do that as they can't be used to "sort" the rest of your matches). When you are all done with the process generally these folks will match two groups.

If you are picking a main/start person for your group that is unknown to you- consider not using a person who is young enough to be a grandkid/great grandkid of a first cousin. Hints to age (if not given):

If tree or partial tree, you can guess at a match's birthdate- was their grandparent born in 1917 or 1957?

Do they have a picture?

Name (I admit this is less helpful, but Shirley, Dorothy, Harold, etc, imply person is older. Ditto Ashleigh, Brandon, Caitlin, younger).

More on why to take the time to do this:

Some folks go into genetic genealogy with a specific goal- to find that elusive John Anderson born in Ohio, or *which* Susan your ancestor Bill Tiddle married.

In the rush to find that ancestor they may fail to establish the necessary groundwork- both genealogically and genetically.

Taking time to identify and label 2nd cousins gives us the tools to ID 3rd cousins- they are lurking in those 50 to 90 cM matches!!

But, I wanted to get beyond second cousins!

Steps to find third cousins after you have a good Leeds list:

Take one grandparent line to concentrate on. Take their Leeds list and see how many of those 90 to 400 cM matches you can identify- make a note and label for that match if you can ID them. For notes I use the format "relationship to tester, shared ancestral couple surnames"- so "2C, Condit/Winters" or "3C1R Condit/Clark". That tells me right where the tester and I intersect. I also put the match and their linking ancestors into my family tree offline as well (offline since it includes living people).

Next, take the first person on your Leeds list who shares your great-grandparents on that line- this will usually be a 2C but may be a half-1C1R or 2C1R. They need to be related to you by JUST one grandparent. Now look at the shared matches you have with this person (who genealogist Diahan Southard dubs your "Best Known Match- BKM) that have LESS shared DNA with you than the BKM. Why less? Because a match that shares more DNA, or about the same, as the BKM, is likely related to you at the 2C level or more closely- remember we are searching for 3rd cousins. You can take a look at the shared cM tool to remind yourself of the typical cM range we expect.

Some of the 3C should have been identified with Leeds. That leaves us just the 50-90 cM matches to concentrate on (I picked this range, the actual range for 3C is 0-234 cM). This gives you a core group of people who should be your 3C, 3C1R to focus on. If you are lucky you may get matches for both sides of the tree.

Short-cuts (or possible short-cuts, ever gotten lost on one?):

A possible short-cut to labeling: for those who have already identified some of their

close matches, or who are having trouble with their groupings making sense.

If you have a KNOWN second cousin you could use that cousin to make your Group 1, pull up their shared matches and use the pencil/edit tool to add them all to group 1.

If you have identified multiple second cousins on a line, select the one who has the most shared DNA with you.

Label as many groups as you can with known second cousins and then go back to the Leed's approach to finish your labeling (picking the highest cM match who hasn't fit into any group yet).

This gets rid of the concerns with "youngsters" who fit into two groups throwing off your list.

This won't clean up the groupings impacted by double relationships in your tree.

Starting from scratch versus starting over:

This approach can be very successful for those who have done NO grouping. For those who have started grouping (and aren't happy with their system) a possible approach is to, gulp!!, start over by erasing all your current groupings. Diahan Southard recommends that. I'll admit that it makes me nervous…

Other resources:

cM Explainer: Another cM tool (from myheritage but available free as a stand-alone):

https://www.myheritage.com/cm

Diahan Southard's company, Your DNA Guide, has some good resources. Her ebook is inexpensive and is thorough in explaining her approach to DNA matches. Some links for her:

Her You tube video on matching at myheritage (also applicable to other sites):

 https://www.youtube.com/watch?v=xGt6OH-4CzY

Diahan's RootsTech talk (if you don't have a familysearch free account you'll need to make one)

https://www.familysearch.org/rootstech/session/shared-dna-matches-the-only-dna-tool-you-will-ever-need?lang=eng

Dana Leeds' RootsTech talk

https://www.familysearch.org/rootstech/session/the-leeds-method-grouping-dna-matches-to-identify-shared-ancestors?lang=eng

Beth Taylor has lots of RootsTech's talks and they are uniformly really good! You might start with:

https://www.familysearch.org/rootstech/session/grouping-dna-matches

https://www.familysearch.org/rootstech/session/using-dna-to-determine-relationships-in-2023