**Segment Analysis: Handout**

**Some Definitions**:

* Matching Segments = specific segments on specific chromosomes where we have matches with other people; longer segments are more valuable as more solid indicators of correct matching
* Triangulated Groups (TGs) = a situation where at least 3 people have well defined matches on specific segments of specific chromosomes
* Chromosome Browsers = tools used by the DNA companies to define the shared segments

**My Heritage’s Special Segment Analysis Tool**

* Today we’re going to focus on the use of the My Heritage special tool for segment analysis.
* An example was shown of the results of this tool – this was kind of a “teaser” for our discussion.

**Process to Arrange the Good Data Using the My Heritage Tool**

* Go to My Heritage, go to DNA Matches
* Scroll down; click on the ellipse.
* A drop-down menu appears. It will say: Export shared DNA segment data info for all DNA matches.
* My Heritage will respond saying they will send you that data via email.
* The email arrives. Open it. It will be an Excel file. You need to do some sorting to get the data arranged the way you need it. Save this original Excel data file before you do any sorting.
* Also, add some of your own data to the large Excel file that My Heritage sent you. This data will be some key information you have on some of your higher scoring matches.
* Sort the data via Column C, the Chromosome number column. Also, add a few new columns and delete some columns that we don’t need.
* We will then do a final sorting of the Excel data, this time using Column D to arrange the data in numeric sequence in the START column for each chromosome.
* New columns include: a column to note whether a match is on the father’s or mother’s side; a column showing cM scores for the high scoring matches; a column showing known relationships for matches; a column showing known family branches for matches; a column to show the cluster number for each match based on My Heritage data; and, a column to show the major chromosome for each match.
* In addition, set up a few columns for future use. One column will designate the Triangulation Group (TG) for a specific match. One column will designate the donor ancestor. Other columns will be added to show Remarks or Notes.

**Chromosome Size**.

* Chromosome 1 is the longest, with each other chromosome diminishing in size. Thus, in general, our own numbers of matches will usually be greater for the longer chromosomes.
* We will be looking for TGs within each chromosome, perhaps a total of about 200 TGs for each side of the house.

**3 Strategies for Segment Analysis**

* 1. Key on the high scoring matches, especially those for which you have the most data.
* 2. Key on the clusters that were defined by My Heritage.
* 3. Go through each chromosome methodically from beginning to end.

**Strategy 1: Key on the High Scoring Matches**.

* Start the analysis by finding matches that have long segments, for example, segments scoring over 30 cM.
* These segments will offer a great immediate look at other adjacent matches who also have matching segments within the same range as the known high scoring match. Using this strategy you can quickly identify perhaps 20-30 new family branch members for many of the high scoring matches.
* An example was shown where a known match with a cM of 44 had about 25 solid matches within the same range as his 44 point segment.

**Strategy 2: Key on Defined Clusters**.

* My Heritage has a tool where it will give us clusters for our higher scoring matches; in my case, I have 18 clusters.
* If you haven’t already done so, click on the Auto Clusters tool and get data on your clusters.
* You should be able to quickly see which of your clusters goes with which chromosome.
* For an analysis, select one of the cluster groups, and go to that chromosome to find those cluster matches.
* We had an example of a 7-match group within a single chromosome. When we looked at the data for that chromosome, we found that our 7 matches were spread around within that chromosome and formed a good solid segment there. Also within that segment, we identified a “new” group of matches who also had data within that segment. These people were, of course, good family branch members of the same branch as the ones originally defined in the cluster.

**Strategy 3: Key on Each Chromosome, Methodically Defining the Many TGs in Each**

* This is a task for the super-diligent. When you complete this task, you will have constructed your own personal genome.
* If you have total success with this effort, you will identify your specific TGs and also will be able to identify which donor ancestry is providing his DNA for that TG.
* We call this type of analysis segmentology.
* Many of us will prefer to perform just a partial chromosome analysis, focusing on certain family branches and certain chromosomes.
* An example was shown for my own family branches, where I have been able to place different family branch people within specific chromosomes. To do this I used FTDNA, GEDMATCH, and My Heritage. I could also have used 23 and Me.

**Sample Chromosome Analysis**.

* I selected Chromosome 13 for this analysis. It is one of the more modest sized chromosomes, I had already defined my Ellifritz family as being big on C13, and one of my Ellifritz matches had a long segment of 51 there, so this looked like a promising chromosome to study.
* Elijah Ellifritz is my GGGF, so my people who match me at that spot are my 3rd cousins. One of these matches is Kristi, who has a long segment of 51 on C13 and who is my 3rd C 2R.
* When we look at the Excel data for C13, we find there are 142 matches. Our goal will be to go through the entire chromosome, from Start Location to End Location, defining finite TGs along the way. It will turn out that there are 7 TGs here.
* We locate Kristi and her 51 point segment a few cM from the beginning, but not at the actual beginning.
* Our initial task is to identify the first TG in the chromosome, so we start by analyzing the matches who have the lowest number Start locations. We find 4 matches who have an identical early Start location, and they have similar End locations. We select the match with the longest segment as our Base person in this group. He matches each of the others in the group. Further, the matches in this TG do not match other people who are located further down in the list. These 4 matches are identified as TG1. We have limited information about each of the matches in this group, so for now we need to mark it as Unknown.
* We next move on to define TG2. We advance to look at the next Start locations in C13. Again, this task turns out to be straightforward because we readily find 3 matches who have similar Start locations as well as similar End locations, and while they match one another they don’t match other people in C13. Of some interest, each of these 3 matches lives in Europe: one in Germany, another in Austria, and one in the Netherlands. But there is not enough data to define their family group, so I mark them as Unknown.
* We are now ready to define TG3, which we quickly recognize as a solid Ellifritz group. A couple of matches have slightly (1 cM) lower/earlier Start locations than Kristi, but they are also Ellifritz family matches. Because of her long segment, we designate Kristi as the Base person for this TG. We start to look at her shared matches within this group and find a sizable number. We think that perhaps every match within her large 51 point segment will be an Ellifritz. But, to be sure, we need to verify that each person not only has scores within the range but that they are shared matches.
* While analyzing Kristi’s shared matches, I run across an anomaly. One person, Tebbe, has a 17-point segment within Kristi’s segment, but they don’t match. Instead, surprisingly, it turns out that Tebbe is a match for EW, who is on the Baker (my father’s) side and who is a known member of the Wessling family. Very few of my matches are on my father’s side, so this was a huge surprise.
* Looking carefully at Tebbe’s profile, we find important data. He has posted a great tree which shows his people living only about 10 miles away from EW’s Wessling family people. EW, our known Wessling, had not been listed on C13, but upon further investigation (looking at the original Excel file) we see that EW scored 9.8 on C13, slightly below the 10 cM threshold.
* Now, re-thinking about Kristi and her TG3, we go down the list and see which ones are shared matches as well as sharing data in the Kristi segment. There are 44 matches in TG3.
* Moving on with the analysis, we tap Tebbe to be the Base person for TG4. He has shared matches with 4 other people and they don’t match the others in C13, so they become TG4.
* Again moving on to increasingly higher Start locations, we find a match, Adams, who scores 31 but doesn’t match Kristi. Instead, he matches Tebbe, but his score takes him much further into C13, and he has a lot of other matches. We might call this group an extension of the original Tebbe group, but it includes many people who match Adams but not Tebbe. So, we define this newer group as the Adams TG5 group.
* By the way, EW, my strong Wessling family match, is also a match for Adams. Thus, it looks like the Adams group is also a Wessling group. The TG5 group are co-existing in the same segment space as the Kristi TG3 group, but of course the TG5 people are on my father’s side. Of some interest, I found that 20 of the 52 people in the TG5 group live in Europe, so apparently the Wessling group that lived in NW Germany have a lot of connecting matches who continue to live in Europe.
* As a note on Shared Matches, it was important to see that while the TG3 and TG5 people shared the same segment space and had similar Start location numbers, the TG3 people had shared matches only with their other TG3 people – and likewise for the TG5 people.
* After the Start/End Locations for the TG3 and TG5 people, there were a few “new” matches to analyze. Thus, we moved on to define the people in TG6. These people matched one another but didn’t match either the TG3 or TG5 people. We defined the highest scoring member of TG6 as the Base person and then identified all of those others who either matched him or matched other TG6 people.
* At the very end of the Excel list for C13, there were still 3 final matches who had not matched the TG6 people. Therefore, I defined these last people as the TG7 group.

**Benefits of Chromosome Analysis**.

* In our C13 example, we found seven distinct TGs. One large TG was composed of only Elllifritz people. Two other TGs were composed of Wessling people. And, the other groups are still unknown.
* One major surprise occurred in this analysis, namely, the fact that a large group of distant Wessling ancestors was found. Without the segment analysis, I doubt if I would ever have found and identified this group.

**Conclusions**:

* Segment analysis is valuable in fully defining the composition of specific chromosomes, identifying segments belonging to family branches
* My Heritage has an excellent tool that is helpful in this effort
* We can make the best use of the My Heritage tool by re-arranging the Excel formatted data and then analyzing that data
* One strategy that we discussed was focusing on our known matches that have long segments, so that we readily define matching segments within that same range
* Another strategy was discussed where effective use is made of the cluster groupings already provided by My Heritage, building on that data
* The value of analyzing specific chromosomes was discussed, with examples showing that surprising results may occur
* The concept of Triangulated Groups was emphasized, and methods were discussed to show how to best group our matches into those groups