Advanced Autosomal DNA Techniques used in Genetic Genealogy

Tim Janzen, MD

E-mail: tjanzen@comcast.net

Summary of Chromosome Mapping Technique

The following are specific instructions on how to map your chromosomes. If you would like to see a more detailed chromosome map, I suggest that you download my parents' chromosome map at http://tinyurl.com/y7zxt9fy. It is suggested that you review this file before trying to understand the details of chromosome mapping. It is important that you note that in that file column G is the chromosome that I received from my mother and that column H is the chromosome that my mother has that she didn't pass on to me. You can delete the data in those columns and then use this file as the file for which you do your analysis of your own data. You can also use the raw data file from any of the genetic genealogy testing companies for the SNPs that you include in your chromosome map. If you would like to use a simpler format in which there is one row per million base pairs, I suggest that you download my template from http://tinyurl.com/y8l2b7cz.

To start with…

1. You need one or both parents' 23andMe results or Family Finder results, and you need one of their children's results. In addition, you can use matching segment data from GEDmatch which may be found using the Tier 1 option "Matching Segment Search".

2. Then compile a list of all of the known relatives of the parent who have been tested who are either first cousins to the parent or more distantly related to that parent.

If you have done the FTDNA Family Finder test or have transferred your 23andMe, Ancestry.com, or MyHeritage data to Family Finder then follow these steps:

1. Log into your Family Finder account and go to https://my.familytreedna.com/family-finder/chromosome-browser.aspx.
2. Click on the link in the right upper corner of the webpage to "Download All Matches to Excel (CSV Format)".
3. Save the downloaded file on your hard drive.
4. Follow steps 1 to 3 for your parents and your children if they have done Family Finder.
5. Merge the Excel files that contain the matching segment files for two parents and one child into a single Excel file.
6. Open the Excel files for the two parents and the child and add a record index number (RIN) column at the far left in each column with a RIN for each row in the file.
7. Sort each of the files by the cM (centimorgan) column from largest to smallest. Delete all matching segment data for segments (HIRs) that are less than 5 cMs (or 7-10 cMs if you prefer). Then resort each file by the RIN. Then save each file. (If you so desire, you can save versions of these files that contain all of the matching segment data for future reference.)
8. Sort the matching segment data in the Excel files for the two parents and for the child into two broad categories: matching segment data for known relatives and matching segment data for genetic cousins with no known genealogical relationship to the child. You can either save the data for known relatives and genetic cousins in separate files if you would like or you can keep

the data combined. It may simplify things for you if you create separate files before proceeding to step 9.
9. Compare the Excel file that contains the matching segment data for the child to the Excel files for each of the parents that also contain matching segment data. Initially do this only for the data for genetic cousins with no known genealogical relationship to the child. Add a column in each of the parents' files with the heading "Child also has this matching segment". For each row in the parents' files enter "yes" if the child also shares the segment with the person that the parent matches and enter "no" if the child doesn't share the segment with the person the parent matches. You can use Louis Kessler's program Double Match Triangulator at http://www.doublematchtriangulator.com to help sort matches into those that match the parent but don't match the child on a specific segment and those that match both the parent and the child on the same segment or you can do the comparisons by hand. If you use the Double Match Triangulator compare the parent's Family Finder chromosome browser download file to the corresponding file for the child. Put one file in as "File a" and the other file as "File b".
10. Create an Excel file for each parent that includes the information from step 9 in a file format similar to my sample master match list spreadsheet at http://tinyurl.com/y87sttmk. Add further information in the spreadsheet to complete the various columns as the information becomes available to you.
11. If you have also tested at 23andMe then merge the Excel files mentioned in step 10 in the 23andMe section above that contain 23andMe data with the Excel files mentioned in step 10 in this section that contain Family Finder data for each parent.

Next…
1. Open the Excel file for one parent mentioned in step 11 above. Sort the file so it contains the matching segment data for only known relatives whose genealogical connection to you is well established.
2. Isolate the matching segment data for a first cousin of the parent first. If no first cousins for the parent have been tested then move to the data for progressively distant relatives (2nd cousins, 3rd cousins, 4th cousins, etc.).
3. Repeat steps 1 and 2 for a child the parent.
4. Merge the matching segment data from the cousin when compared to the parent with the matching segment data from the cousin when compared to the child into a single Excel file.
5. Then sort the data for the known cousin mentioned in step 4 by the chromosome and then by start location so that you can see all of the matching segments for both the parent and the child adjacent to each other in the file.
6. Then delete all segments that are under 5 cMs for which the parent and the child do not both share a corresponding matching segment with the relative. Such segments will only be found in Family Finder data files. It may be reasonable to map some segments in the 3-5 cM range if both the parent and the child share that same segment with the relative, but caution is warranted when mapping segments that don't contain at least 700 or more SNPs because some matching segments could be IBS (identical by state)(false matches) and not IBD (identical by descent).
7. If you are primarily using Family Finder data then open a Family Finder autosomal raw data file in Excel. Add a RIN number column on the far left side of the file for each row for which there is SNP data. Add a RIN number column on the far left side of the file for each row for which there is SNP data. Bear in mind that if you prefer a simpler format for your chromosome

map then you can use my template at http://tinyurl.com/y8l2b7cz in which there is one row for each million base pairs.
8. Add columns to the right of the SNP allele values with these headings: "chromosome mapping of the one chromosome that ____ ____ (the child) inherited from _____ _____ (the father)", "chromosome mapping of the one chromosome that ____ ____ (the child) did not inherit from _____ _____ (the father)", "chromosome mapping of the one chromosome that ____ ____ (the child) inherited from _____ _____ (the mother)", "chromosome mapping of the one chromosome that ____ ____ (the child) did not inherit from _____ _____ (the mother)". These column headings would go in columns F, G, H, and I. Alternatively you can delete the SNP data column from the file if you so desire. If you did this, then the column headings would go in columns E, F, G, and H. If you prefer, you may use an abbreviated chromosome map format such as various sample templates you can obtain from me.
9. Begin the chromosome mapping process as outlined in the next section. Bear in mind that if you are mapping data from 23andMe, the matching segment data is in Build 37 and the start and stop points will be somewhat different from the data from Family Finder and GEDmatch which are in Build 36. The columns being referred to assume that you have deleted the SNP data and are doing a comparison of a cousin to a mother and her child.

In the Chromosome Map Excel file…

Fill in the start point in column G with the name of the ancestor from whom that the parent must have received the segment for all of the segments in which the child also matches the other relative. For example, if a first cousin of the parent is being used for comparison then you would enter into the cell the name of the grandparent of the child who was related to the first cousin in the appropriate cell in the file, the grandparent's year of birth and the name of the relative used for comparison. For instance, if the child's start point was at position 80,000,000 on chromosome 5 then you would fill cell G177752 with the name of the grandparent, their year of birth (YOB) and the name of the relative. Then copy the contents of cell G177752 and drag the contents down to fill in all cells in column G down to the endpoint of that segment. If the end point was at position 104,500,000 then you would fill in the column all the way down to cell G181568.

Simultaneously fill in the contents of cells H177752 to H181568 with the name of the spouse of the grandparent and their YOB. I prefer to add "inf" after the name and the YOB to designate that I have inferred this information from the fact that I have already filled in the corresponding cells in column G. For example, if I have tested a first cousin of my mother's M. Y. on the paternal side, then I knows that the DNA segments that they share came from his mother's father Paul A. Youngman (b. 1907). I would thus fill in G177752 to G181568 with "PAY1907 MY inf". If I didn't inherit the segment that this cousin shares with my mother then I would fill in H177752 to H181568 with the name and YOB of my grandmother, Maude E. McIntire (b. 1907) such as "MEM1907 inf", using her initials and the birth year. Repeat this same procedure for every segment that the child shares with the relative used for comparison.

Once you have done the above then review the segments that the parent shares with the other relative that the child does not share. For instance, let's say that the parent's matching segment on chromosome 5 starts at 65,000,000 and continues to 110,000,000. You would then go to cell H174981 and enter the name of the grandparent of the child who was related to the first cousin, their YOB, and the name of the relative used for comparison. I also add "inf" after this information to indicate that I inferred the information from the fact that the child doesn't

share that matching segment. You would then fill the contents of cell H174981 down to cell H177751 with the same information as is in cell H174981. You would next fill in cells H181569 to H182595 with the same information. This would cover the segment between position 104,500,000 and 110,000,000. I then simultaneously fill in the contents of cells G174981 to H177751 and G181569 to G182595 with the name of the spouse of the grandparent, their YOB, and "inf" after that. In similar fashion continue filling in all the appropriate cells in column H for the segments that the parent shares with the other relative but the child does not and continue adding the corresponding information for those segments in column G.

Once you have recorded the information as recommended for the first relative selected for comparison in the paragraph above then continue in similar fashion for all known relatives who are related no more closely than at the first cousin level of relationship to the parent. If there are segments that multiple relatives share in common with the parent, then modify the contents of the corresponding cells to include the name or the initials of each relative who shares that same segment. You may also add data from GEDmatch if you are a Tier 1 member.

Summary of the Visual Phasing Technique

1. Run comparisons for at least 2 full siblings (preferably 3 or more) using David Pike's utility at http://www.math.mun.ca/~dapike/FF23utils/pair-comp.php and/or at GEDmatch using the one-to-one comparison tool. If using GEDmatch check the box "Full Resolution" to see the positions of the FIRs more accurately. If you have phased files at GEDmatch, use those for the comparisons since it will be easier to see the locations of the crossovers if you use phased files.
2. Compile the graphic displays and the matching segment data. If using GEDmatch, you must determine the start and stop positions of all FIRs from the graphics display unless you are doing comparisons in the Genesis Project, where the start and stop positions of all FIRs are shown.
3. Determine the person in whom the crossover occurred with for each change from FIR to HIR, from HIR to FIR, from HIR to no matching segment, and from no matching segment to HIR.
4. Assign a generic grandparent code to each segment that came from a specific grandparent. I like to use P1, P2, M1, and M2. The two P codes stand for either the two paternal grandparents or the two maternal grandparents and the two M codes stand for the opposite grandparents. You don't know initially whether grandparents with the P and M codes are paternal or maternal.
5. Start by labeling segments where two siblings are fully identical or where they don't share any DNA whatsoever. It is best to start the mapping process at one end of the chromosome if at all possible.
6. Using logic, assign grandparent codes to the adjacent segments, one segment at a time until you have completed the assignment of codes throughout the entire chromosome. You may only make one arbitrary "leap of faith" with the assignment of codes at the beginning of the process. If crossovers occurred very close to each other, this could throw you off. Larger chromosomes are more complicated and it may be challenging to completely assign codes for all regions of the chromosome. Leave segments blank if there is more than one possible assignment.
7. Review your chromosome maps for the siblings to see if you have any regions of the chromosome mapped to specific parents or grandparents. If you have segments attributed to grandparents, assign those segments to the appropriate grandparent first. Then assign the other segments using logic as completely as possible. Your chromosome maps can help you complete

section 6 above if you get stuck completing the assignment of specific segments to specific grandparent codes.

8. Update the chromosome maps for each sibling using the mapping information you have learned from visual phasing. I suggest using a designation that the information was derived from visual phasing when you enter it on your chromosome map. For example, you could use "John Smith (b. 1876) inferred from visual phasing" for each segment that you have assigned to your grandfather John Smith using the visual phasing technique.