Connecting the Dots – Introduction to AutoClusters at MyHertiage DNA

Paul Woodbury ● DNA Team Lead at Legacy Tree Genealogists (https://www.legacytree.com/) ● Author at Genealogenes (https://genealogenescards.car.blog/) ● pwoodbury@legacytree.com

The AutoClusters Tool at MyHeritage DNA

In February 2019, MyHeritage DNA in collaboration with Evert-Jan Bloom of Genetic Affairs, introduced a licensed version of Bloom's AutoClusters tool as part of the MyHeritage DNA experience.[1] The AutoClusters tool is one of many genetic network tools that assists users in grouping genetic cousins based on their relationships to each other. These groups can then be analyzed further to determine which ancestral lines they likely represent, and better isolate the genetic cousins whose relationships are most likely pertinent for a research question.

Shortly after the release of the AutoClusters tool at MyHeritage, Ran Snir offered an excellent webinar through Legacy Family Tree Webinars describing how to obtain an AutoClusters report, and how to interpret it.[2] While we will review some of the same elements in this webinar, we will also share new insights from genetic genealogy in practice by reviewing basic genetic network theory, reviewing the elements of AutoClusters tool at MyHeritage, exploring tips and tricks for interpretation, and considering direct applications of how AutoClusters might be used.

Genetic Cousins, Shared Genetic Cousins, Shared Cousin Groups and Genetic Networks

Each individual inherits half of their autosomal DNA from each of their parents. Beyond that, the amount of DNA shared in common is only approximate due to a random process called recombination which shuffles the DNA each generation. Each individual will inherit about 25% from each grandparent and approximately half again from each previous generation.

Different descendants of a particular ancestor inherit different portions of that ancestor's DNA. Therefore, some descendants may be the sole heirs of specific segments, while in other cases, multiple descendants may share inheritance of other segments. When two individuals share a segment of autosomal DNA, they are "genetic cousins" and they share a common ancestor. Determining the generational distance to and the identity of that ancestor requires additional consideration of the size of the segment, the number and size of other shared segments, the genetic cousins who also share DNA with a test subject and the genetic cousin either on the same segment or on different segments and traditional genealogical evidence. When three or more

[1] Esther, "Introducing AutoClusters for DNA Matches," *MyHeritage Blog,* 28 February 2019, https://blog.myheritage.com/2019/02/introducing-autoclusters-for-dna-matches/, accessed July 2020; and, Evert-Jan Bloom, "AutoCluster," https://www.geneticaffairs.com/features-autocluster.html, accessed July 2020.
[2] Ran Snir, "Visualizing Ancestral Lines with DNA AutoClusters," https://familytreewebinars.com/download.php?webinar_id=1052, accessed July 2020.

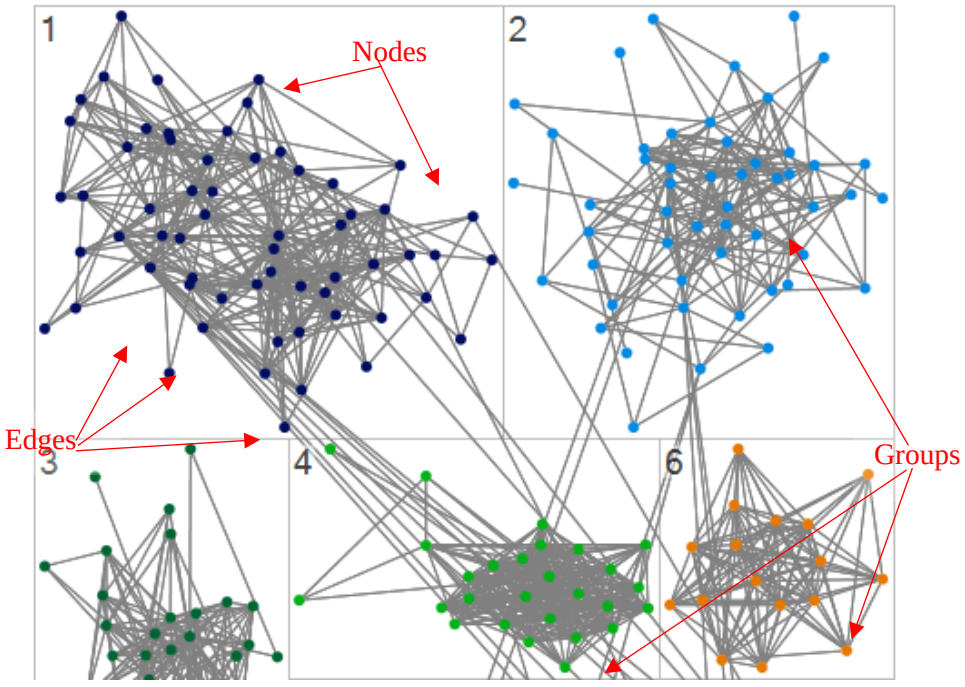
individuals are genetic cousins to each other, then they often (but not always) share at least one common ancestor or ancestral couple. The DNA they share may or may not be triangulated (shared between all three on the same segment of DNA).

Each of the major DNA testing companies provide a list of genetic cousins who are ordered based on the amounts of DNA they share with a research subject. For each genetic cousin, testing companies will also identify which other genetic cousins in the list are "shared matches" or are "In-Common-With" (ICW) the test subject. 23andMe, Family Tree DNA and MyHeritage consider any genetic cousins sharing any amount of DNA with another genetic cousin to be an ICW match, while AncestryDNA has a 20 cM threshold for genetic cousins to be considered as shared matches. When exploring the relationships between genetic cousins as reported in shared matches reports, researchers often uncover groups of genetic cousins who are all shared matches to each other. These shared match groups form an important element of genetic networks.

Genetic networks show the relationships between genetic cousins and can assist in identifying multiple groups of matches who are likely related through common ancestors or particular ancestral lines or regions. When multiple genetic cousins of a test subject share descent from a common ancestor, that individual is likely among the ancestors or collateral relatives of the test subject as well. Genetic networks enable meaningful organization of genetic cousins and prioritization of which genetic cousins might be most helpful for addressing a research problem.

Components of a Genetic Network

Genetic networks are composed of subjects, nodes, edges, and clusters.



Genetic Network of Kathleen Jones (subject) including relationships between 330 estimated fourth cousins and 19 network groups.

Subjects: Most often (and in the case of the MyHeritage DNA AutoClusters tool) the subject of a genetic network is a test subject themselves. The network shows the relationships between all or some of that individuals' genetic cousins. Alternatives include creating a network with the data of multiple subjects or creating a network for a specific ancestor.

The members or *nodes* of a genetic network are the genetic cousins of a test subject or subjects. To maximize the benefit of a network analysis, thresholds are placed to determine which individuals are included in a network. close genetic cousins should be excluded from a network in order to better differentiate between groups. Also, it may be wise to focus a network analysis on genetic cousins sharing DNA over a lower threshold as well in order to form meaningful groups and avoid the background noise and confusion introduced with analysis of more distant genetic cousins. In the case of the MyHeritage DNA AutoClusters tool, the thresholds determining which genetic cousins are included as nodes or members of the network is determined through automatic threshold selection and attempts to limit the overall size of the network to about 100 nodes.

Connections or edges represent genetic relationships between two individuals besides the network subject. These too can be altered with centimorgan thresholds. At MyHeritage, 23andme, and GEDmatch it is possible to determine the strength of relationships between genetic cousins and this can be utilized to manipulate or alter networks based on edge data. In the case of the MyHeritage DNA AutoClusters tool, the threshold for what constitutes a shared relationship is determined through automatic threshold selection.

Criteria can also be applied to group (*cluster*) definitions in order to enhance genetic network analysis. The most basic definition of a cluster is two nodes who share DNA with the subject and with each other. Criteria might be applied to the definition of a cluster based on the number of individuals required for its formation. Alternatively, clustering algorithms might be utilized to automatically group individuals. In the case of the MyHeritage DNA AutoClusters tool, network groups are formed when two or more matches of a test subject are found to share DNA with each other based on the automated thresholds from the clustering tool. When other genetic cousins share DNA with the majority of a cluster's nodes (members), they are added to the same cluster.

Types of Genetic Networks

Three ways to visualize genetic networks include tabular formats, graph formats and matrix formats. While the MyHeritage DNA AutoClusters tool utilizes a combination of the matric and tabular format, several other tools or methodologies use different representations.

Tabular Format

In tabular formats, genetic cousins are listed in a spreadsheet with important information, and at least one other column (but potentially more), indicate the groups to which the match may belong.

One effective method for clustering close autosomal DNA matches in a tabular format is the Leeds Method.[3] In this method, researchers skip the closest autosomal matches and focus on individuals who are estimated 2nd – 3rd cousins: those sharing less than about 400 cMs of DNA. Since the purpose of clustering methodologies is to separate people into unique groups, it is important to exclude those whose presence would cause groups associated with more distant ancestors to be merged into a single mega-group. For example, while one group might be composed of descendants and relatives of a pair of paternal-paternal great grandparents, another group might be composed of relatives of a pair of maternal-maternal great grandparents. Including the seven tested children, and the five tested nieces and nephews of a test subject would merge these distinct relationship groupings into a single group – those who share DNA with the test subject's closest relatives. It is recommended that descendants, parents, siblings, aunts, uncles, first cousins and other close relatives be excluded from network analyses.

As part of the Leeds Method, a spreadsheet of close genetic cousins is first constructed. The first individual is analyzed and assigned a color. All of this first match's shared matches are assigned the same color. The next match without an assigned color is analyzed and assigned a different color. Their shared matches are assigned the same color. The process is repeated until all of the closest matches are assigned to relationship clusters or groups. This approach, however, assumes independent genealogies and does not work as well for endogamous populations. Because it is done by hand, it is also prone to user error and can be time intensive.

Graphs

In graph formats, each match is represented by a dot and each relationship is represented by a line.

DNAGedcom.com: The automated scans of autosomal DNA data available to subscribers of this website using a companion software enables spreadsheet analysis of DNA test results from all of the testing companies.[4] Manipulation of these spreadsheets can reveal clusters and network groups. Advanced excel analysis using multiple kits can also be used to isolate pertinent genetic cousins for a particular research question.

NodeXL: an Excel template created by the Social Media Research Foundation and can be used for visualization of networks.[5] While there are paid versions of the template, the basic version is sufficient for the needs of most genetic genealogists. In order to create a genetic network using NodeXL, it is necessary to set up the nodes and edges of the network by importing DNAGedcom Client scans. Detailed instructions on how to import data from these spreadsheets into NodeXL

[3] Dana Leeds, "DNA Color Clustering: The Leeds Method for Easily Visualizing Matches," https://www.danaleeds.com/dna-color-clustering-the-leeds-method-for-easily-visualizing-matches/, accessed July 2020.
[4] *DNAGedcom,* https://www.dnagedcom.com/About.aspx, accessed July 2020.
[5] Social Media Research Foundation, *NodeXL,* https://www.smrfoundation.org/nodexl/, accessed July 2020.

is available at the "Twigs of Yore" blog by Shelley Crawford.[6] NodeXL currently only works for PC users. A comparable software for mac users is *Gephi*.[7]

Matrix

In matrix formats, like the format used in the MyHeritage DNA AutoClusters tool, a list of matches is named along the Y-axis of a matrix and is also listed in the same order along the X-axis. Colored boxes in the intersections of each match represent relationships between matches.

Genetic Affairs: While the AutoClusters report available through MyHeritage was created in collaboration with Evert-Jan Bloom, his website, Genetic Affairs offers additional network analysis and cluster analysis tools.[8] Through this website, users can construct relationship matrices for clustered groups based on 23andMe, Family Tree DNA, and GEDmatch data. New features also permit automated tree analysis between group members and advanced rule-based clustering. Until recently, Genetic Affairs could also be utilized with AncestryDNA data, but as of June 2020, this is no longer an option.

Collins-Leeds Method at DNAGedcom.com: This tool available as part of the suite of tools for subscribers of DNAGedcom.com is also presented in a matrix format and can be used for analysis of data at AncestryDNA, 23andMe, Family Tree DNA and GEDMatch.

Interpretation of Genetic Networks

Once constructed, genetic networks can be utilized to identify the most pertinent genetic cousins for a research problem.

The main value of genetic networks is their ability to group genetic cousins based on their genetic relationships to each other. Using this information, it is often possible to identify specific surnames, ancestors, families, communities, localities, and ethnicities associated with different network groups. Rather than comparing family trees and reported surnames between one match and the entire match list, it is possible to focus on those genetic cousins who are most likely related through the same lines of ancestry.

If several members of a genetic network group all descend from a known ancestor or ancestral couple, then it can be assumed that the other members of that group are also likely related through the same ancestral line. If two genetic network groups are tied to each other by many connections between groups, it could indicate that one group is a subgroup of the other, that the two groups are from an endogamous community, or that the two groups are from two branches of a single ancestral line and are connected by descendants of the alliance between both family branches. On the other hand, if there are very few connections between two genetic network groups, it may indicate that there is a genetic connection between two members of the separate

[6] Shelley Crawford, "Visualizing Ancestry DNA Matches," *Twigs of Yore* (blog), https://twigsofyore.blogspot.com/2017/07/visualising-ancestry-dna-matchesindex.html, accessed July 2020.
[7] *Gephi,* https://gephi.org/, accessed July 2020.
[8] Evert-Jan Bloom, "AutoCluster," https://www.geneticaffairs.com/features-autocluster.html, accessed July 2020.

groups by chance, but that genetic connection is independent of their respective relationships to the network subject(s).

One advantage of genetic networks is that they separate groups of related individuals from each other. If one or more members of a network group share more than 100 cMs of DNA with the subject(s), and if there are several members of the group, then it is much less likely that members of other unassociated and non-allied network groups are related through the same ancestral lines. If they are, they are typically more distantly related. Therefore, if a genetic network group is *not* tied to another genetic network group of known ancestral origin, it can act as evidence that the group is not related through the ancestral line of the known group. Also, when analyzing genetic networks, we recommend that researchers pay attention to close genetic cousins who do not fit into a network group. These genetic isolates may be close relatives who have no other close tested cousins. This is particularly common for individuals with recent immigrant ancestry. As with genetic network groups with known origin, genetic isolates can provide evidence that none of the other genetic network groups with close genetic cousins are related through the same ancestral line as a close genetic isolate.

Focus on those with the closest genetic cousins first and then working through groups composed of more distant relatives. This approach enables efficient identification of groups which are related through ancestral lines not pertinent to the research objective and groups which are potentially pertinent.

Exceptions

In some situations, networks may not be as helpful for exploring a research problem. Networks are based on the genetic relationships between genetic cousins. Therefore, if a test subject or a research subject descends from an endogamous population, their network will frequently consist of a single large group; all of their genetic cousins may be tied to many other cousins and groups through multiple ancestral lines. If endogamy plays a role in a tests subject's ancestry, it will be obvious from their genetic network as there will be multiple connections between most groups or specific groups may be indistinguishable. In these cases, researchers still might benefit from a network analysis at a much higher threshold of shared DNA. For example, researchers might focus only on those individuals who share more than 50, 75, or 100 cMs of DNA. By raising the threshold, it may be possible to obtain some useful information regarding the relationships of close genetic cousins. Alternatively, if a test subject has too few genetic cousins, it may be impossible to build a network if none of their cousins appear to be related to each other. In these cases, it may be useful to lower the threshold to determine if there are *any* genetic cousins in the subject's match list who are identified as ICW relatives.

How to Get an AutoClusters Report at MyHeritage DNA

To obtain an AutoClusters report from MyHeritage DNA:

1. Login into your MyHeritage account.
2. Hover over the DNA tab.
3. Select "DNA Tools" from the pop-up menu.

4. From the resulting page, click on “Explore” for the AutoClusters tool.
5. If you manage multiple kits, select the kit for which you wish to generate an AutoClusters report.
6. Click on “Generate.”
7. Wait for a message to arrive in the email associated with your MyHeritage account.
8. Download the resulting Zip file.
9. Click on the HTML file and find your AutoClusters Report.
10. Click on the csv file and view your results in a spreadsheet.
11. Click on the ReadMe pdf and learn more about AutoCluster analyses and the specific settings used for your AutoClusters report.

Elements of the MyHeritage DNA AutoClusters Report

An excellent overview of the MyHeritage DNA AutoClusters report and the elements it contains is provided in Ran Snir’s webinar from April 2019 dedicated to the topic. Here we review some of the same elements:

1. A key showing the colors, names, and sizes of each cluster.
2. A list of genetic cousins listed along the Y-axis (names replaced with numbers in the accompanying chart)
3. A list of genetic cousins in the exact same order listed along the X-axis.
4. Bolded and colored boxes for the intersections between entries for the same match
5. Colored dots where a genetic match shares DNA over the prescribed threshold with other members of the same cluster. Hovering over these boxes reveals description of the relationship between two matches as well as a list of all members of the cluster
6. Tan boxes where a genetic match does not share DNA over the prescribed threshold with other genetic cousins (even in some cases when they are in the same cluster)
7. Gray boxes where a genetic match shares DNA with members of other clusters. Hovering over the box will describe the relationship.

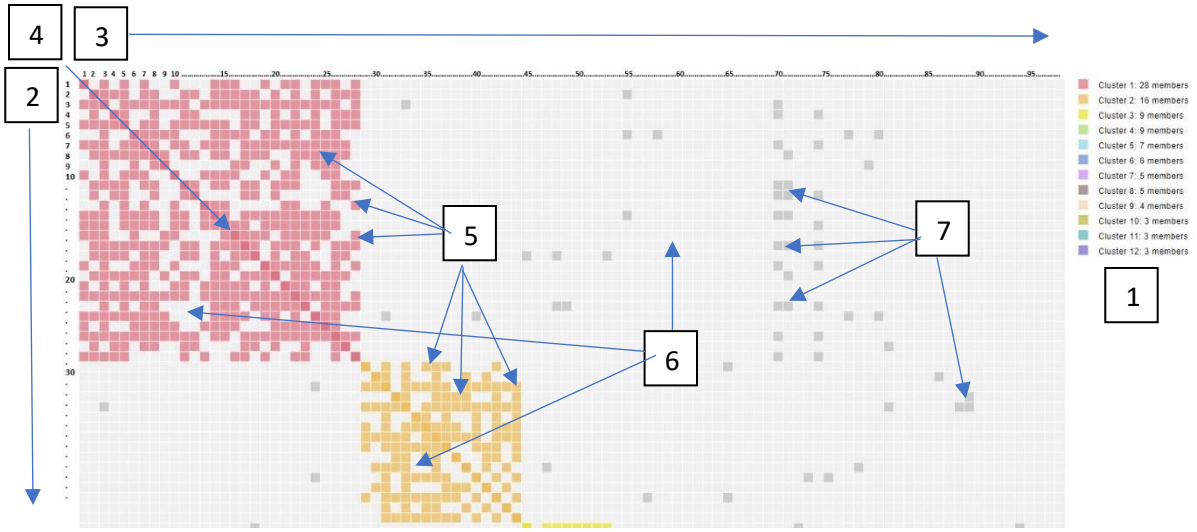


Figure 1: Elements of a MyHeritage DNA AutoCluster Report Part I

8. Colored clusters of matches ordered by the size of the cluster (the number of matches), and then by the shared cM with the closest match in the cluster.
9. A description of the thresholds used for your particular report. MyHeritage automatically selects these thresholds when analyzing your data. They are selected by running several iterations with varying thresholds until an optimal AutoCluster chart is formed. Note that while you cannot adjust these thresholds in the MyHeritage AutoCluster report, adjustment of thresholds is permitted on similar reports run on data from other companies at Genetic Affairs.
    a. The total number of DNA matches included in the report.
    b. The minimum cM threshold that a DNA match had to share with the subject in order to be included in the report.
    c. The maximum number of cMs that a match could share in order to be included in the report (no close relatives were included in the report).
    d. The threshold that DNA matches had to share with each other in order to be considered shared matches.

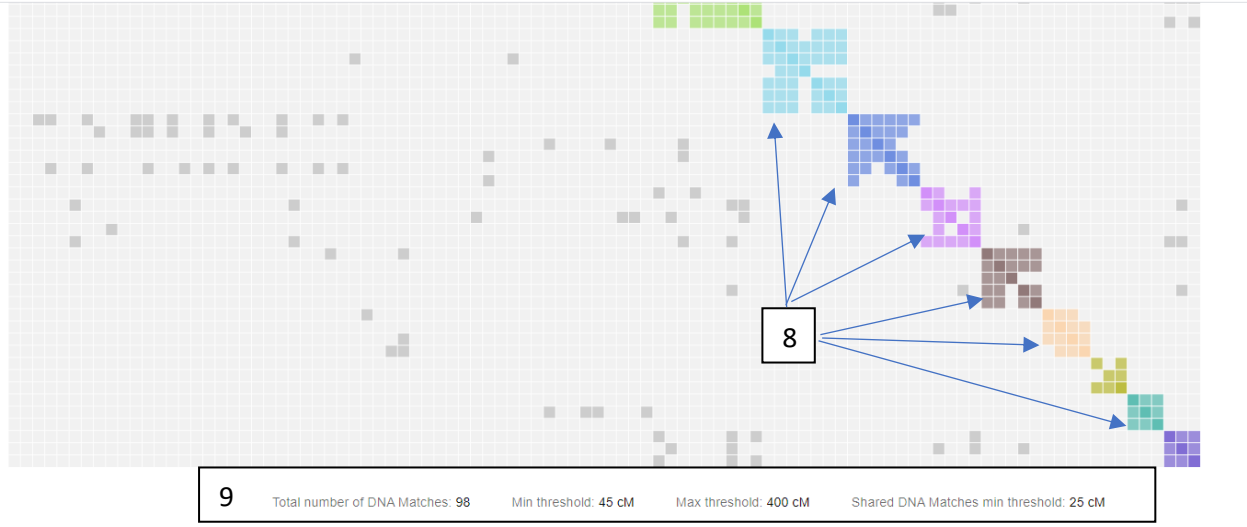


Figure 2: Elements of a MyHeritage DNA AutoClusters Report Part II

10. A table of matches which by default is organized by cluster. The table can be searched or filtered by each column heading. The table contains the following information:
    a. The names of matches (changed to numbers in the accompanying picture), which link to the corresponding match's DNA profile at MyHeritage. Individual names can be searched with the search bar at the top of the table.
    b. The number of cM shared with each match.
    c. The largest cM shared with each match
    d. The number of segments shared with each match
    e. The number of in-common-with relationships in the AutoCluster report.
    f. The cluster to which the match has been assigned.

g. The number of individuals in the match's tree, which link to the match's tree.
h. Any notes that you have added to the match's profile at MyHeritage DNA.

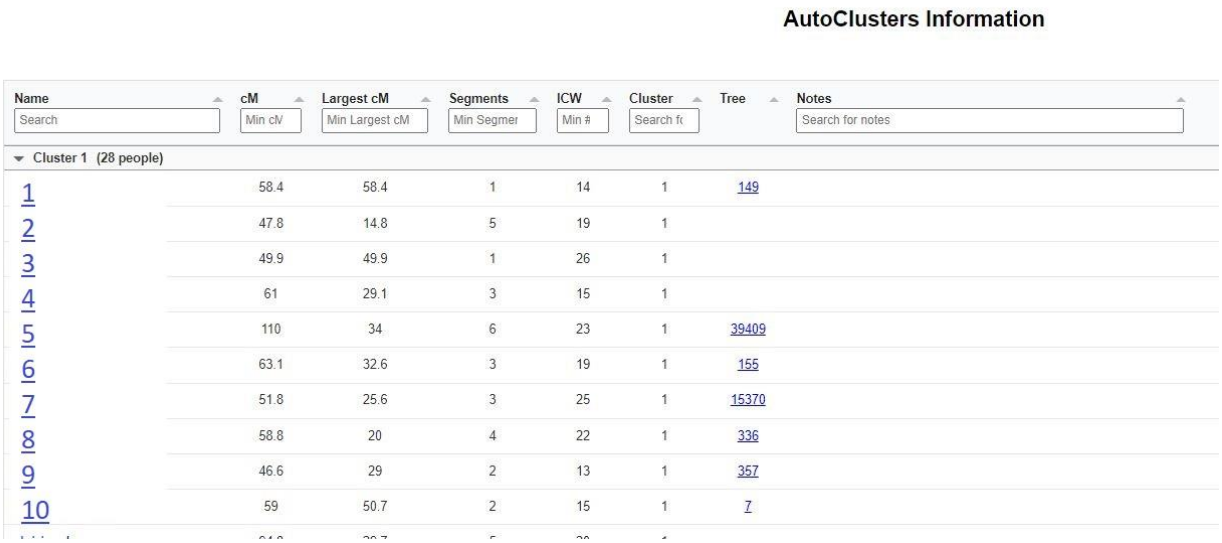


Figure 3: Elements of a MyHeritage DNA AutoClusters Report Part III

The csv file accompanying the download includes a spreadsheet representation of the matrix as well as a list of genetic cousins who met the matching criteria with the subject to be included in the network, but did not share sufficient (or any) DNA with other matches in the analysis and therefore were assigned to their own groups.

What Does a Cluster Represent?

In the past, many have suggested that clusters or groups in genetic networks represent descendants of a single pair of common ancestors. While this is often the case, exploration of relationships between members of the same cluster reveals that there are other possibilities:

1. Members of a cluster could all descend from a single shared common ancestor or ancestral couple.
2. The majority of the members in a cluster could descend from a single common ancestral couple, but some of the members of the cluster might descend from collateral relatives of either the male or the female individual in the couple.
3. The members of a cluster might all claim ancestry from the same isolated or endogamous population.

Who is or who is not included in one cluster or group versus another cluster or group often depends on the specific thresholds and settings used to define who is included in a network analysis, what amount of shared DNA constitutes a relationship between two matches, and how many members of a cluster a genetic cousin must match in order to be considered a member of

the same cluster. Changing these settings can drastically change the groupings and clusters that constitute a network.

Though there is some variability of who is included in a network group and under what settings or circumstances, network groups can still be helpful for the purposes of organizing and prioritizing analysis of genetic cousin relationships in order to answer a research question or achieve a research objective. The perspectives and insights offered by genetic network analyses, including the MyHeritage DNA AutoCluster tool, help in the identification of genetic cousins who are most likely related through a research subject of interest.

Why Do Some Matches Not Match All Other Members of a Cluster?

Not all genetic cousins in a particular genetic network group or cluster may be genetic cousins to each other. There are several reasons why this might be the case including the following:

1. Even if two genetic cousins are both descended from the same ancestral couple, they may not share DNA with each other. Because of the nature of autosomal DNA inheritance, though two individuals may inherit DNA from a common ancestor, they may inherit different portions of that ancestor's DNA and they may share no DNA with each other. Most of us share at least some DNA with each of our full second cousins, but for more distant relationships, it is entirely possible for two descendants to inherit different portions of a common ancestor's DNA and not share DNA with each other. For example, a cluster composed of the descendants of John and Mary may include two third great grandchildren. Though these individuals share DNA with other members of the cluster who also descend from John and Mary, suggesting that they both inherited at least some DNA from John and/or Mary. They are fourth cousins to each other, but they did not inherit the same DNA and therefore do not share DNA with each other.
2. In situations where clusters are composed primarily of descendants of a specific ancestral couple, but also include descendants of collateral relatives of each member of the couple, it is possible that members of the network group do not match other members of the network group because they do not share any common ancestors. If a genetic cousin descends from a collateral relative of the common male ancestor to the majority of the group and another genetic cousin descends from a collateral relative of the common female ancestor to the majority of the group, those two genetic cousins may not share DNA with each other. For example, a cluster of composed primarily of descendant of John and Mary but which also includes descendants of siblings and first cousins of John and Mary may include a descendant of John's brother and a descendant of Mary's first cousin who have no relationship to each other and yet share DNA with many of the descendants of John and Mary who are also members of the cluster.
3. Another reason why two members of genetic network group may not be identified as matches to each other centers around the definition of the threshold for what constitutes a shared match. Two individuals might be descended from the same ancestral couple, and

may even share a small amount of DNA with each other, but if the amount of DNA they share falls under the threshold necessary to be considered a shared match, they may not be identified as such. For example, a cluster composed primarily of the descendants of John and Mary may include two third-great-grandchildren who are fourth cousins to each other. These two descendants may share a single 10 cM segment with each other, but if the threshold for what constitutes a shared match is set at 20 cM, they will not appear as shared matches to each other and the network subject.

4. In situations where clusters are composed of individuals with ancestry from the same isolated or endogamous population, it is possible that two members of the cluster may not share DNA with each other because they share no common ancestors with each other even if they in turn are related to other members of the network through independent sets of common ancestors, multiple common ancestors, or the same set of common ancestors multiple times.

Why Do Some Matches Share DNA with Members of Other Clusters?

Sometimes members of a genetic cluster will share DNA with members of another cluster. These alliances between clusters could be strong as in the case where many members of a cluster match many members of another cluster. These alliances could be moderate as in the case where one or two members of one cluster match all or many of the members of another cluster. These alliances could be quite weak as in the case where one member of one cluster matches a single member of another cluster. All of these connections between clusters might be due to the way in which the network, clusters, and relationships between matches are defined.

What constitutes a cluster is often determined by the settings and thresholds utilized to define inclusion in a network analysis, inclusion in a cluster, and definition of a shared match relationship. Depending on how these settings are adjusted, a single group in one analysis might be split into two or more groups in a follow-up analysis. Even if these groups or clusters are split from each other, there may still be many shared match connections between their respective members. Alternatively, if the definition of what constitutes a shared match is adjusted, or if the threshold of which matches are included in a network analysis is changed, a cluster from one analysis might be split entirely in a subsequent analysis with few or no connections remaining between the two resultant clusters. Because the number, prevalence, and strength of connections between clusters depends on the settings used to define networks, clusters and relationships, there are many reasons why some DNA matches share DNA with members of other clusters including the following:

1. If a network group is composed primarily of descendants of a common ancestral couple, and another network group is composed primarily of descendants of another common ancestral couple, and if one of the network groups includes one or more descendants of both couples, then the individual may share DNA with the majority of the members of the other cluster. For example, if a cluster is composed primarily of the descendants of John

and Mary and another cluster is composed primarily of the descendants of Joseph and Elisabeth, and if the first cluster includes a descendant of John and Mary's son Peter who married a daughter of Joseph and Elisabeth named Jeanne, then that descendant may have multiple genetic connections to the second cluster.

2. If a network group is composed primarily of descendants of a common ancestral couple, and another network group is composed of more distant descendants of the parents or grandparents of either the male or female individual in the ancestral couple, then there may be multiple connections between members of both clusters. However, they may be considered as separate clusters, because there are stronger relationships between the more closely related individuals than there are to the more distantly related individuals and some of the descendants of the more recent shared ancestral couple may not share enough DNA or any DNA with the more distant relatives such that two clusters are created. For example, a genetic network group composed primarily of the descendants of John and Mary may have several descendants who share DNA with another cluster composed of descendants of John's paternal grandparents. Because the second cluster is composed of more distant relatives, not all of whom share DNA with the descendants of John and Mary in the first cluster, they are considered to be a separate cluster.
3. If a single match from one cluster shares DNA with a single match or a few matches from another cluster, it is possible that those two genetic cousins share DNA with each other independent of their respective relationships to the network subject. For example, a network subject's paternal cluster composed primarily of descendants of John and Mary has a single member who also shares DNA with one member of a distinct maternal cluster composed primarily of the descendants of Paul and Rachel. These two matches of the network subject share common ancestors Mark and Ruth who are not shared ancestors to the network subject.
4. If a network subject descends from ancestors belonging to an endogamous population, or who hailed from the same geographically isolated area, it is possible that there may be multiple connections between members of two or more network clusters because both clusters include relatives from the same broader population. For example, a network subject's cluster composed primarily of descendants of John and Mary may have several connections to members of a network group composed of descendants of David and Deborah, if John, Mary, David and Deborah all came from the same small community in Appalachia.

What Next?

Once you have generated an AutoClusters report, what next steps should be taken to use that information in your family history research? Here are some ideas:

1. Even though the closest relatives in your match list may not be included in clustering charts, they can still help in interpretation of those charts. Use close tested relatives to quickly identify which clusters may be of interest to a particular research question by

determining which clusters are associated with shared matches of your closest known relatives.
2. Start with the match sharing the most DNA in each cluster.
3. Analyze their family tree. If necessary, extend their family tree through additional research
4. Determine if the genetic cousin shares a common ancestor or ancestral couple with you or with the network subject.
5. Repeat the process for other matches in the same cluster proceeding in order from most shared cM to fewest shared cM.
6. Determine if other matches from the same cluster also descend from the same ancestral couple, or from ancestors of the same ancestral couple.
7. If you cannot identify common ancestors between yourself and the members of the cluster, see if there are any shared ancestors between the members of the cluster.

If three or more individuals in a cluster descend from a common ancestor or ancestral couple (either with you or with each other), and if those cluster members descend from unique and independent descent lines, then it is likely that other members of the cluster also descend from the same common ancestor (or ancestral couple) or are collateral relatives of either the male or female in the shared ancestral couple. Once common ancestors are found between members of a cluster, here are some additional ideas of next steps to help in your genealogical research:

8. If you find that a particular cluster includes descendants of known ancestral lines that are not of interest to you in the context of your current research objective, ignore the members of that network group and proceed with analysis on another network group.
9. If you find that a particular cluster includes descendants of a known ancestral line that is of interest to you given the context of your current research objective, work to identify the relationships between *all* members of the group as it is possible that the cluster may include collateral relatives of a research subject whose genealogies may provide clues regarding the ancestry of your own research subject.
10. If you find that a particular cluster includes descendants of a known ancestral line that is of a known ancestral line that is of interest to you given the context of your current research objective, commence research on clusters which are possibly allied to the pertinent cluster through shared match relationships. Allied clusters may include relatives of collateral relatives of a research subject whose genealogies may provide clues regarding the ancestry of your own research subject.
11. Since the MyHeritage AutoClusters tool is typically limited to the first 100 or so matches in a subject's match list, there are often many more distant genetic cousins who are not included in the analysis who may still be pertinent to a research objective. Once a pertinent cluster has been identified, consider going to the respective shared match lists for the members of the cluster to identify more distant matches who are also pertinent.

Extend the family trees of these individuals to learn more about the collateral relatives and by extension the ancestry of a research subject.

12. Though some genetic cousins may share a sufficient amount of DNA with the network subject to be included in an analysis, they may share too little DNA with shared matches to be assigned to a network group. Exploration of these shared matches (listed at the bottom of the csv file), particularly those sharing a large amount of DNA with the network subject, can also be helpful in determining which matches may or may not be pertinent to a research objective.

Conclusion

As one of many tools to aid in genetic network construction, the MyHeritage DNA AutoClusters tool is a fantastic resource to aid researchers in achieving a birds-eye view of their match lists. Using this tool, it is possible to quickly identify which clusters of shared matches are likely to be pertinent to a research question and which might be ignored. Once pertinent clusters have been prioritized, the relationships between cluster members, and their shared matches not included in the network analysis can be used to break through genealogy brick walls, identify unknown ancestors, and extend ancestral lines.