DNA: Unlocking the Code
Iowa Genealogical Society
Des Moines, IA
April 9, 2016

Instructor:
Bryant McAllister, PhD
Associate Professor of Biology
222 Biology Bldg (BB)
319-335-2604
bryant-mcallister@uiowa.edu

Website: https://wiki.uiowa.edu/display/2360159/G4G

Suggested Reading: *The Invisible History of the Human Race* by Christine Kenneally, 2014, Viking

Workshop Description: With a mouse-click and credit card sophisticated DNA tests for detecting relatedness and predicting ancestry are available directly to consumers at a relatively low cost. Advances in human genetics over the past 30 years have enabled the emergence and growth of this "recreational genetics" marketplace. This course will introduce the science underlying these tests and explore the products delivered by the companies that dominate the US marketplace (e.g., 23andMe, AncestryDNA, and Family Tree DNA). Personal genetic analysis provides a window into the past at different time scales, and this course will examine the discovery of genetic relationships between cousins as a genealogical research tool along with inferences that extend beyond the historical record.

Workshop Objectives

- Evaluate the products and deliverables of direct-to-consumer DNA tests
- Conceptualize the organization of the human genome and properties that contribute to its variability among individuals
- Build proficiency in the methods and interpretation of test results of human ancestry and biological relationships

Schedule	**Topic and Objectives**
9:00 - 10:15	**Navigating the DNA Marketplace** Describe the direct-to-consumer genetic testing marketplace Consider differences between DNA testing options and testing strategies that support your interests Introduce the principles of discovering relationships among individuals
10:30 - 11:45	**Connecting Cousins for Genetic Genealogy** Navigate lists of DNA relatives and individual DNA matches Cluster DNA relatives into family groups having common ancestry Recognize patterns of autosomal identity produced by fundamental principles of inheritance

11:45 - 12:45 **Lunch**

12:45 - 2:00 **Option 1 – Getting More from Your SNP Data (Advanced)**
Focus on advanced applications including data imports into GEDmatch
Apply genetic principles to predict patterns of DNA identity between biological relatives
Compare matching DNA segments to map your genome and inform relationships
Recognize the utility of raw data downloaded from the testing company

Option 2 – Small Group Tutorials* (Beginner)
Focus on 23andMe and AncestryDNA results
Preview and navigate test results
Use features available in the account
Ask questions about the interpretation of test results

2:15 - 3:30 **Ancestry Analysis from DNA**
Consider potential negative consequences of personal genomics
Trace deep genealogical history and further with mtDNA and Y chromosome
Interpret ancestry estimates from genome-wide SNP test

***Led by University of Iowa Students**
Nicholas Stewart
Kyle McElroy
Robb Todd
Dzavid Pandzic
Mary-Kayt Jones

Session 1 – Navigating the DNA Marketplace

Session Objectives

- Describe the direct-to-consumer (DTC) genetic testing marketplace
- Consider differences between DNA testing options and testing strategies that support your interests
- Introduce the principles of discovering relationships among individuals

Major Testing Options (list price)

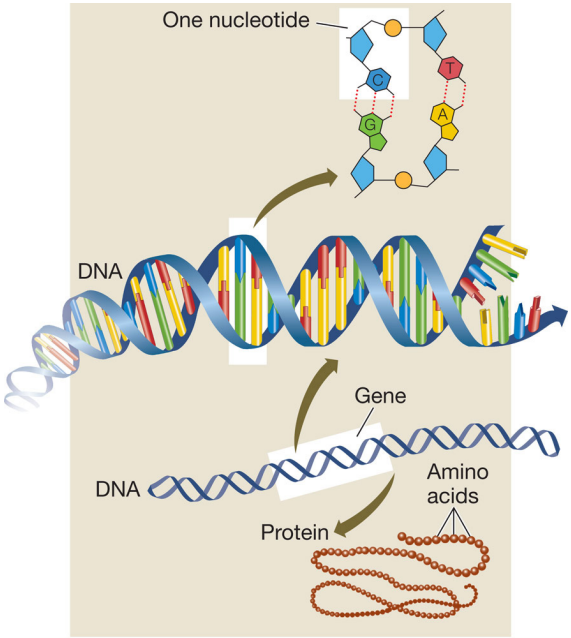
- **23andMe** – Report revised in Fall 2015 ($199). Includes Ancestry Estimates, DNA Matching, Carrier Status, and Trait Prediction reports.
- **AncestryDNA** – Includes Ancestry Estimate and DNA Matching ($99). Integrated with Ancestry.com family tree for an extra subscription fee.
- **Family Tree DNA** – Various tests available including Family Finder SNP test ($99).
- **The Genographic Project** – GENO 2.0 Next Generation ($200) provides ancestry estimates exclusively.

Questions about the use of DNA tests

- What outcome(s) do you want in the test results?
- What does the DNA test assay?
- What interpretations are delivered in the results?
- How would you use these results?
- What is the cost, and how does it compare with competitors?
- Are there additional fees associated with access or use of test results?
- Are other customers satisfied with the service?
- Are the terms of service acceptable to you?

DNA is the cellular recipe for life

- A **Genome** is the entire set of **DNA** molecules present in a cell.
- Each DNA molecule contains a string of A=T pairs and G≡C pairs as a **DNA sequence**.
- DNA sequence contains the cellular recipe for life. Regions of DNA molecules (genes) encode products (mostly proteins) that perform cellular and organismal functions.
- Different cells of an individual contain the same genome sequence, but unless you have an identical twin, your DNA sequence is different from any other individual.

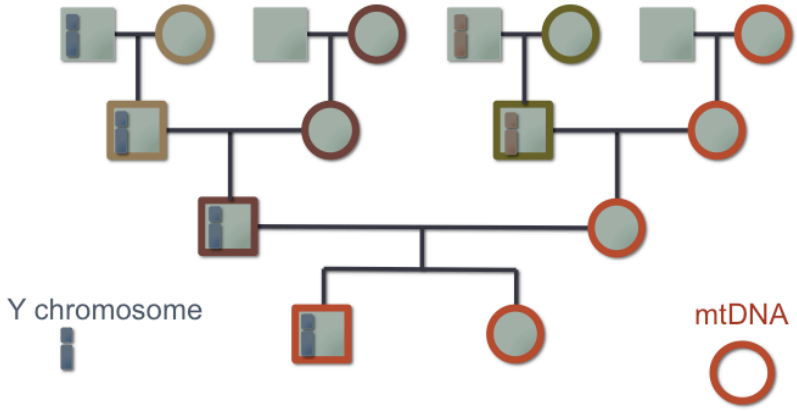


LIFE 10e, Figure 1.5

Methods for placement of an individual within the Human Family Tree

- Shared sequence in DNA of maternally-inherited mitochondrion or male-limited Y chromosome
  Placement on a branch of the human family tree
  Revelation of closest relatives by fewest mismatches
- Genetic similarity with reference populations sampled throughout the globe
  Prevalence of mtDNA and Y chromosome haplogroups
  Estimate of compositional mixture from different ancestry groups
- Segments of sequence identity along the DNA of autosomes and X chromosome
  Number and length of identical segments decreases with relationship distance

Uni-Parental DNA Markers



- Single DNA molecules inherited from one parent, and that trace a single ancestral line.
- **Mitochondrial DNA** (mtDNA) is a small circular molecule present in the mitochondria. Only transmitted from the maternal parent, and thus is inherited through the maternal line.
- The DNA molecule of the **Y chromosome** contains a gene that directs development toward the male pathway, thus it is present only in males. The Y chromosome is transmitted to sons only, and is thus inherited through the paternal line.
- Shared changes in the sequence of the mtDNA or the Y chromosome indicate common ancestry. Acquisition and comparison of DNA sequences from mtDNA or the Y chromosome from many individuals reveal deep relationships of maternal and paternal ancestry, respectively. Branches on reference trees depicting these relationships are designated with a specific **haplogroup** nomenclature that represents the underlying DNA sequence diversity.

Genome-Wide SNP Analysis

- Most of the human genome is contained in 23 pairs of linear DNA molecules, where one member of each chromosome pair is inherited from each parent.
- Autosomes are inherited equally from each parent, whereas inheritance of the pair of sex chromosomes depends on the sex of the parent and child.
- Each company uses a common technology that assesses about 700,000 different known variable sites (SNPs) in the DNA of your genome (including mtDNA in some cases). Most of the same sites are used in the tests of different companies.
- Each **Single Nucleotide Polymorphism** (**SNP**) analyzed has a unique ID and has an identified position along a chromosome in the reference human genome sequence.

Genetics for Genealogy: Major Testing Options

- **AncestryDNA** ($99) – Fully integrated with active Ancestry.com tree with powerful features; no ability to directly examine matching DNA segment(s)
- **Family Finder Test** at Family Tree DNA ($99) – Broad set of features to examine DNA matches; website navigate is challenging and database is relatively small ($37 to transfer AncestryDNA data into Family Finder database)
- **23andMe** ($199) – Good for broad set of interpretations, entertainment, & validating relationships; most users are private and not motivated by genealogy

  This website contains a details comparison of the major testing companies https://wiki.uiowa.edu/display/2360159/DTC+Genetic+Testing+Companies+Compared

AncestryDNA

- Primary application is family history research through DNA matches, but includes an ancestry composition estimate.
- Largest database of tested customers.
- AncestryDNA test can be linked with family tree in Ancestry.com for fully integrated research.
- Sophisticated DNA matching algorithm that reduces the number of false-positive matches of close relationship.
- Automated family clustering in DNA circles.
- Ability to manage DNA kits of multiple family members in a single account, and share results between accounts.

Family Tree DNA

- Primary application is family history research through DNA matches.
- Broad set of tests and tools for working with DNA matches for family history research, but their utility and integration in the website is not intuitive.
- Relatively small database, but participants are mostly engaged and knowledgeable about family history research.
- Each DNA sample is registered in an account, and different tests on that sample are managed in a single account.
- Participate and/or manage projects that include multiple participants.
- Accepts raw DNA data from AncestryDNA, The Genographic Project, and older 23andMe tests.

23andMe

- Primary application is self-exploration through DNA.
- Broadest set of interpretations from a DNA test including maternal line, paternal line for males, ancestry composition, Neanderthal composition, DNA relatives, carrier status, and traits.
- Choice to maintain privacy or not.
- Interface for browsing raw DNA data.
- Multiple kits can be managed in a single account, and some results can be shared between accounts.

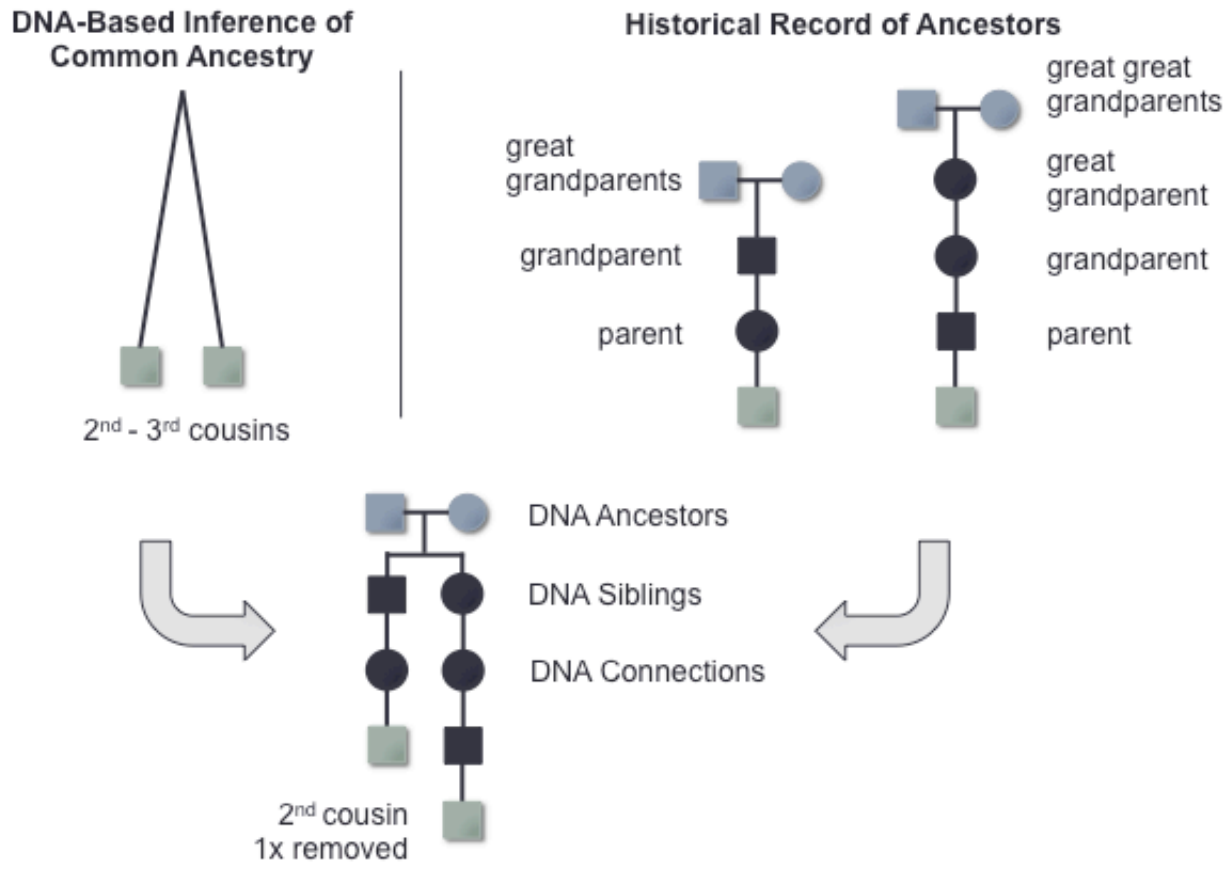
Session 2 – Connecting Cousins for Genetic Genealogy

Objectives

- Navigate lists of DNA relatives and individual DNA matches
- Cluster DNA relatives into family groups having common ancestry
- Recognize patterns of autosomal identity produced by fundamental principles of inheritance

Genetic Genealogy

Using genetics in genealogical research concerns the intersection between DNA-based inferences of common ancestry and historical inferences of who these ancestors were. Analysis of an individual's DNA and comparison of those data with others reveals relatives connected through recent common ancestor(s). Comparison of family history with a matching relative may provide evidence on the identity of ancestors potentially responsible for the shared genetic similarity. As the number of genetic relatives increases, each appearing as a DNA match with other descendants of a common identified ancestor(s), the greater the confidence in the identity of the common connection.



Autosomal SNP Matching

- The 22 pairs of autosomes are inherited equally from each parent, and are used by each of the companies to identify DNA relatives. The X chromosome can also be used to identify DNA relatives.
- Each Single Nucleotide Polymorphism (SNP) analyzed has a unique ID and has an identified position along a chromosome in the reference human genome sequence. Each customer's genetic profile is determined at each of the approximately 700,000 SNPs analyzed. The genetic profile may reveal that the pair of chromosomes contain the same base composition at a SNP (e.g., GG), or that the pair of chromosomes contain different bases (e.g., AG).
- Segments of DNA sequence half-identity, where an extended stretch of assessed SNPs indicates a common genetic profile on one member of the chromosome pair, are used to identify DNA relatives.

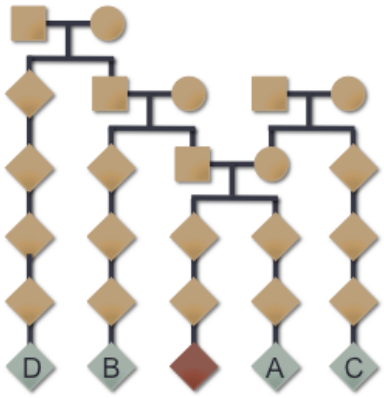
Features of each DNA Relative

- Identifying information (for non-anonymous DNA matches)
- Mechanism for messaging relatives through web interface or email
- Estimate of relationship, and the genetic basis for that inference
- User provided information on family history; surnames, family tree, etc.

Example Organizational Summary of DNA Matches on an Individual Test

Ancestors	DNA Match (relationship)	"in Common With" or "Shared Matches" Match-A	Match-B	Match-C	Match-D
G Grandparents	Match-A (2 C)	Self	Yes	Yes	No
GG Grandparents	Match-B (3C)	Yes	Self	No	Yes
GG Grandparents	Match-C (3C)	Yes	No	Self	No
GGG Grandparents	Match-D (4C)	No	Yes	No	Self

The Corresponding Family Tree:



Mechanisms for Clustering DNA Relatives

"Shared Matches" tab in AncestryDNA

- Shared matches limited to 4th cousin and closer to tester
- Shared matches limited to 4th cousin and closer of 1° match
- Sharing results of a DNA test bypasses this limit

"DNA Circles" and "New Ancestor Discoveries" in AncestryDNA

- DNA kit linked with shared family tree
- Dependent on both DNA and Ancestor Information matches

"In Common With" sorting in Family Finder (FTDNA)

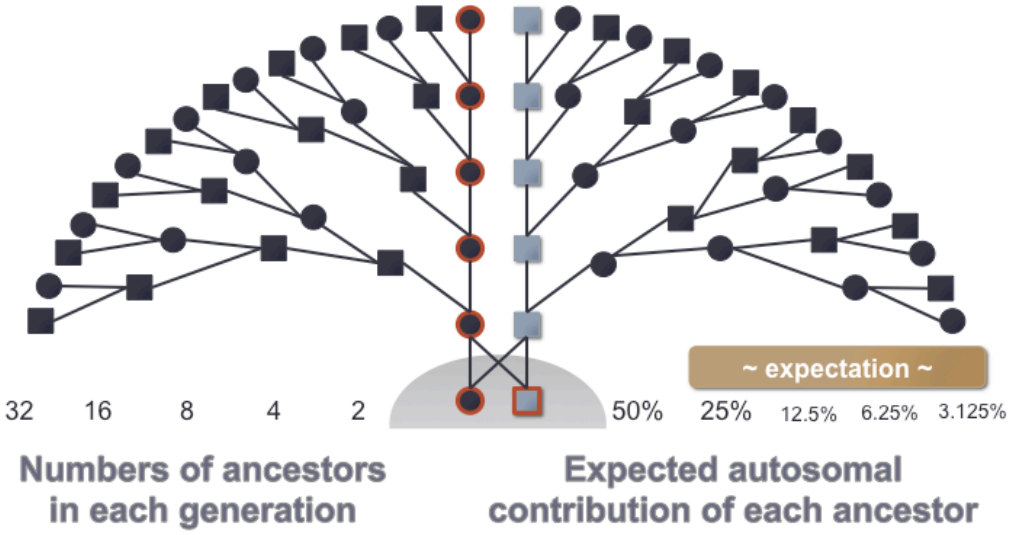
- Use "Family Finder Matrix" to compare multiple DNA relatives

No mechanism for clustering match lists in 23andMe

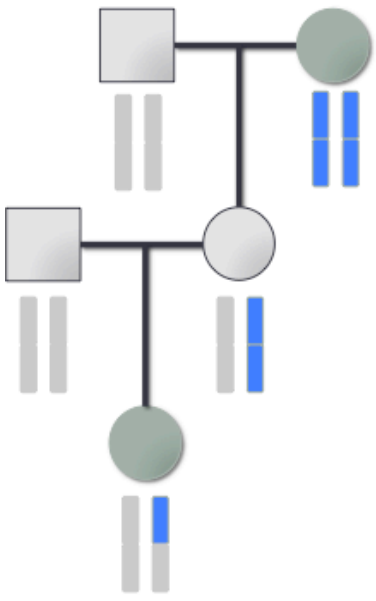
- Greatest limitation of 23andMe platform for genealogical research

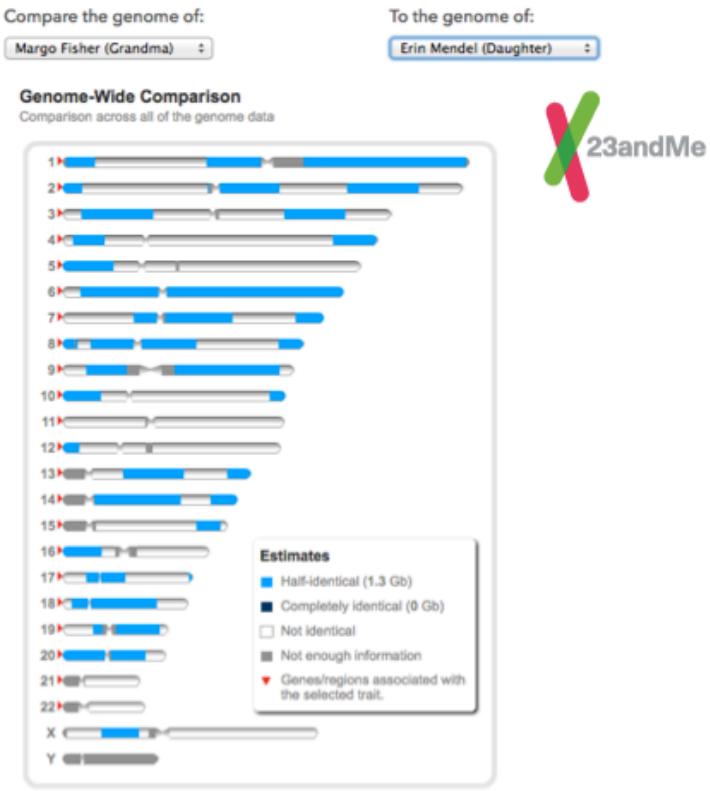
Principles of Inheritance

- **Segregation** – Chromosome pair separate and are transmitted individually and equally to offspring. This 1:1 segregation each generation produces 50% autosomal identity between parent and child, which reduces approximately by ½ each generation further removed.



- **Independent Assortment** – Different pairs of chromosomes are transmitted independently to offspring
- **Recombination** – Independent assortment and exchange between chromosome pairs forms new genetic combinations





Measurements of Matching Chromosome Segments

- Physical Size – basepairs of DNA in the reference human genome
  ~3,000,000,000 basepairs (3 Gb) = half the genome – size of the reference sequence
  ~6,000,000,000 basepairs (6 Gb) = whole genome – two copies of each chromosome
- Genetic Size – centimorgans in the reference linkage map of the human genome
  3,590 cM = all autosomes
  100 cM = 1 Morgan – represents the distance between points along a chromosome pair in which 1 recombination is expected each generation

Genetics for Genealogy

- Results vary dramatically based on individual ancestry and database composition – *test the oldest generation.*
- Endogamy increases the background level of matching among individuals.
- Lists of DNA matches contain two types of errors:
  The genetic identity shared with the relative was not inherited from your most common ancestor(s), but rather a more distant ancestor.
  Relatives descended from a known common ancestor(s) do not appear in the list of DNA relatives, which is an increasingly likely outcome for 4th cousins and beyond.
- DNA relatives may be connected through multiple pathways of common ancestry.

Session 3 – Getting More from Your SNP Data

Objectives

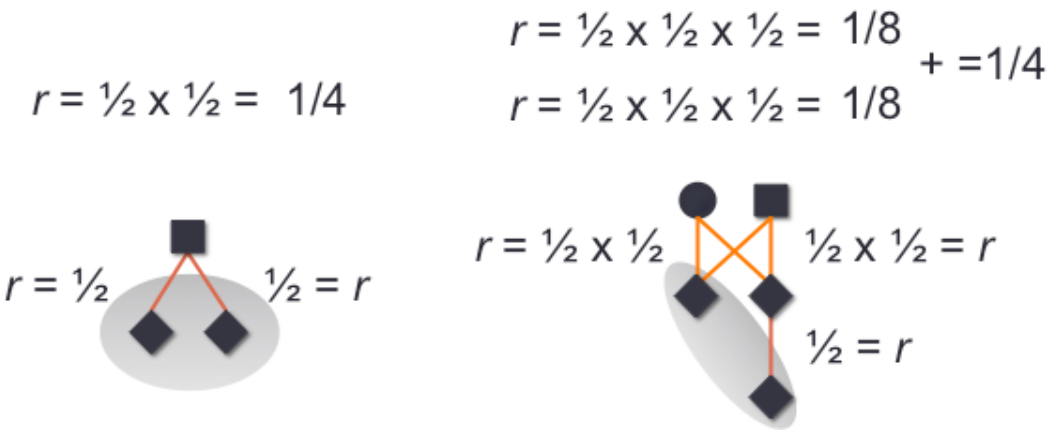
- Apply genetic principles to predict patterns of DNA identity between biological relatives
- Compare matching DNA segments to map your genome and inform relationships
- Recognize the utility of raw data downloaded from the testing company

DNA Testing Strategies for Family History Research

- Full siblings share about 50% autosomal DNA segment identity, thus provide the ability to confirm DNA matches and to expand (primarily 4th cousin and beyond) match lists.
- Close relatives (parents, 1st and 2nd cousins) with known relationships enable methods for positioning matches within the branches of your family tree.
- DNA matching with known half relatives provides support for common ancestry from a single individual, whereas DNA matching with full relatives beyond 1st cousins only provides support for common ancestry from one person of the common ancestral couple.
- Tabular organization of historical and genetic relationships within family groups can be used to summarize the support of the historical record.

Degree of Autosomal Relatedness

- Relatedness between each parent-child is 50%, or a probability of identity of 0.5.
- The probability of identity due to common ancestry is the probability of parent-child identity raised to the power of the number (n) of parent-child transmission event connecting the relatives, i.e., $(0.5)^n$ or $(½)^n$
- If relatives are connected through multiple common ancestors, the sum of the probability of identity through each common relative gives the overall expected probability of identity.
- The two examples below each produce relatives (half-sibs and Uncle/Aunt-Niece/Nephew, respectively) expected to have genomes that are 25% identical.



Expectations for the Inheritance of Identical Matching Autosomal Segments

Relationship	Genome Identity	Number of Segments	Avg. Segment Length	*p*(no shared segments)*
1st Cousins	12.50%	41.4	21.7 cM	0.00
2nd Cousins	3.13%	14.8	15.1 cM	0.00
3rd Cousins	0.78%	4.8	11.6 cM	0.02
4th Cousins	0.20%	1.5	9.4 cM	0.31
5th Cousins	0.05%	0.4	7.9 cM	0.70

*The probability that no identical segments are shared assumes no error in the detection of an identical segment, and is based on the model of Donnelly (1983) *Theoretical Population Biology* 23:34-63

Triangulation and Genome Mapping

- Autosomal triangulation commonly refers to the comparison of matching chromosomal regions among multiple DNA relatives to identify potential relationships.
- Requires the reporting of coordinates for matching segments among all of the individuals being compared, and currently this is only available with shared genomes in a 23andMe account and with data transferred into GEDmatch.
- Extended segments of identity commonly shared among multiple individuals likely trace their ancestry to a single common ancestor. Genome mapping is the identification of the ancestral source for each chromosome region in each generation of ancestors.
- DNA relatives that match in the same chromosome region, but do not match each other in this region, likely represent relatives connected through maternal and paternal lineages.

Session 4 – Ancestry Analysis from DNA

Objectives

- Consider potential negative consequences of personal genomics
- Trace deep genealogical history and further with mtDNA and Y chromosome
- Estimate ancestry composition from genome-wide SNP test

Potential Consequences from the Results of a Genetic Test

- Surprising findings about family relationships
  - *New relatives discovered by genetic matching*
  - *Discovery of genetic relationship between your parents*
  - *Greater distance than expected with other tested family members*
- Surprising prediction of ancestry composition
  - *Individuals contain a genetic mixture representative of their ancestry through human history*
- Surprising discovery of genetic susceptibility to disease
  - *You and/or your children may have an increased chance of developing an incurable genetic disease*

Uses of mtDNA Analysis

- Haplogroup Determination
  - *Placement within the mtDNA tree of humans*
  - *Matriline ancestry likely traces its origins through geographic region with high haplogroup representation*
- Matching with Distant Relatives
  - *Matching between relatives extends past historical records*
  - *Join relevant groups with a project administrator*
- Maternal Ethnicity
  - *Haplogroup K present among about 50% of Ashkenazi Jews*

Uses of Y Chromosome Analysis

- Haplogroup Determination
  - *Placement within the Ychr tree of humans*
  - *Patriline ancestry likely traces its origins through geographic region with high haplogroup representation*
- Matching with Distant Relatives
  - *Matching between relatives extends past historical records*
  - *Join relevant groups with a project administrator*
  - *Surname Study*
  - *Identify surname of closely matching Patriline relatives*
  - *Subdivide genetic families within surname*

Potential Outcomes of Y Chromosome Investigation of Surname

- Support for Patriline
  Close relationship with other men that trace their ancestry back through common patriline
- Association with a Different Surname
  Close relationship with other men that share a common surname
- No Clarity in Ancestral Surname
  Close matches exhibit a variety of different surnames

Ancestry Composition Estimates from Genome-Wide SNP Data

- *It's an estimate!!!*
  Each analysis is based on a set of computer algorithms for handling data from a broad diversity of individuals.
  The set of reference samples used in the analysis, and how the analysis is (or isn't) tailored to self-reported ancestry, differs between companies.
- Current political boundaries have little relevance
  Broadest categories capture genetic similarities present among peoples in different countries.
- Small contributions can represent a variety of things
  Application of an admixture model using many reference populations over-emphasizes effect of rare variants.