GENETICS FOR ANCESTRY, GENEALOGY, AND MORE

Bryant McAllister, PhD
Associate Professor of Biology
222/224 Biology Bldg (BB)
bryant-mcallister@uiowa.edu
Course Schedule

• Week 1, Feb. 4    DNA and the Human Genome

• Week 2, Feb. 11   Ancestry Analysis from DNA

• Week 3, Feb. 18   Connecting Cousins for Genetic Genealogy

• Week 4, Feb. 25   Recreational Genetics

https://wiki.uiowa.edu/display/2360159/G4G
More Learning Opportunities

• DNA: Unlocking the Code – Apr 9, 2016
  Iowa Genealogical Society, Des Moines
  http://iowagenealogy.org/?page_id=5265

• DNA Day Iowa City – Apr 24-25, 2016
  Julie Granka, PhD from AncestryDNA
  Public Presentation, 3:00 pm, April 24, 101 BBE

  Hands-on Workshop hosted by Personal Genome Learning Center
  12:00-3:00 pm, April 25, Computer Room, Iowa City Public Library
OBJECTIVES

• Consider differences between DTC DNA testing options and testing strategies that support your interests
• Recognize the utility of raw data downloaded from the testing company
• Evaluate trait predictions from genetic data
Placement within 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 Ychr 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
Shared DNA Sequence in Human Tree
Genetic Similarity with Ancestry Groups
Shared Segments of Autosomal Identity

you

modern human

chimpanzee
Major Testing Options:


- **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.
TESTING STRATEGIES

Who to test?
What do you want to accomplish?
Questions About 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?
The Genographic Project

- Primary application is prediction of deep ancestry.
- Provides a comprehensive summary of findings relevant to geographic history of your ancestors; maternal line from mtDNA, paternal line for the male Y chromosome, and overall composition compared to reference samples.
- **Probably the best dataset for estimating ancestry composition, and potentially the most accurate.**
- **Privacy within The Genographic Project is protected.**
- Option to transfer data to Family Tree DNA.
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.

• Easy 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.

- **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, and traits.
• Choice of maintaining privacy or not.
• Nice interface for browsing raw DNA data.
• Multiple kits can be managed in a single account, and some results can be shared between accounts.
Family Testing Strategy

~50% identical
DNA Support of Common Ancestry

Patrick McAllister
b. 1830 Ireland
d. 1880 Tennessee

Thomas McAllister
b. 1818 Ireland
d. 1875 Ohio

Mary?
b. abt 1789 Ireland
d. after 1860 Ohio

Me!
Circos Image of Shared DNA
TRIANGULATION AND GENOME MAPPING

Compare matching DNA segments to map your genome and inform relationships
Sara Combs  
b. 1824 TN  
d. 1914 MO

William Lane  
b. 1820 TN  
d. 1852 AR

Match A

Match B

Match C

Match D

Match E

Matrix Matches

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match A</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Match B</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Match C</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Match D</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>Match E</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
</tr>
</tbody>
</table>

- This person is identified as a match.
<table>
<thead>
<tr>
<th>★</th>
<th>Name</th>
<th>Strength of Relationship</th>
<th>Sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>name</td>
<td>Second Cousin Once</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Removed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.16% shared, 7 segments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MB</td>
<td>Second Cousin Once</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Removed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.49% shared, 2 segments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>name</td>
<td>Third to Fourth Cousin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.99% shared, 2 segments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anonymous</td>
<td>Third to Fourth Cousin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.82% shared, 3 segments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anonymous</td>
<td>Third to Fourth Cousin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.66% shared, 3 segments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anonymous</td>
<td>Third to Fourth Cousin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.56% shared, 3 segments</td>
<td></td>
</tr>
</tbody>
</table>
- Bryant McAllister and Half 1C
  - Half IBD 450 cM
  - Segments 10

- Bryant McAllister and 4C
  - Half IBD 25 cM
  - Segments 2

- Bryant McAllister and Half 3C 1X
  - Half IBD 8 cM
  - Segments 1

- Bryant McAllister and ?
  - Half IBD 74 cM
  - Segments 2
Matching Triangulation

John Diggins (Deegans)
BIRTH 17 FEB 1836 • Greenup, Greenup, Kentucky, USA
DEATH 29 JUL 1915 • Elk, Missouri, USA
2nd great-grandfather
Non-matching Triangulation
Analyze Your Data

DNA raw data
- 'One-to-many' matches
- 'One-to-one' compare
- X 'One-to-one'
- Admixture (heritage)
- Admixture/Oracle with Population Search
- Phasing
- People who match one or both of 2 kits Updated
- Predict Eye Color
- Are your parents related?
- 3D Chromosome Browser
- Archaic DNA matches
- Multiple Kit Analysis NEW
- DNA File Diagnostic Utility
  Analyze DNA file upload for potential problems.

Genealogy
- 1 GEDCOM to all
- 2 GEDCOMs
  Search all GEDCOMs
  Revised
- GEDCOM + DNA matches

[GEDmatch] Tools for DNA & Genealogy Research
RAW DATA
Access to the raw data download
Third party resources that accept raw data
## Genome-wide SNP Analysis

<table>
<thead>
<tr>
<th>rsid</th>
<th>chromosome</th>
<th>position</th>
<th>genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12564807</td>
<td>1</td>
<td>734462</td>
<td>AA</td>
</tr>
<tr>
<td>rs3131972</td>
<td>1</td>
<td>752721</td>
<td>AG</td>
</tr>
<tr>
<td>rs148828841</td>
<td>1</td>
<td>760998</td>
<td>CC</td>
</tr>
<tr>
<td>rs12124819</td>
<td>1</td>
<td>776546</td>
<td>AA</td>
</tr>
<tr>
<td>rs115093905</td>
<td>1</td>
<td>787173</td>
<td>GG</td>
</tr>
<tr>
<td>rs11240777</td>
<td>1</td>
<td>798959</td>
<td>AG</td>
</tr>
<tr>
<td>rs7538305</td>
<td>1</td>
<td>824398</td>
<td>AC</td>
</tr>
<tr>
<td>rs4475691</td>
<td>1</td>
<td>846808</td>
<td>CT</td>
</tr>
<tr>
<td>rs7537756</td>
<td>1</td>
<td>854250</td>
<td>AG</td>
</tr>
<tr>
<td>rs13302982</td>
<td>1</td>
<td>861808</td>
<td>GG</td>
</tr>
<tr>
<td>rs55678698</td>
<td>1</td>
<td>864490</td>
<td>CC</td>
</tr>
<tr>
<td>rs1110052</td>
<td>1</td>
<td>873558</td>
<td>TT</td>
</tr>
<tr>
<td>rs147226614</td>
<td>1</td>
<td>878697</td>
<td>GG</td>
</tr>
<tr>
<td>i6052728</td>
<td>1</td>
<td>878697</td>
<td>GG</td>
</tr>
<tr>
<td>i6019302</td>
<td>1</td>
<td>881843</td>
<td>GG</td>
</tr>
<tr>
<td>rs2272756</td>
<td>1</td>
<td>882033</td>
<td>GG</td>
</tr>
</tbody>
</table>
Eric Mc
This test is shown to matches as E. M.
🔗 LINK TO TREE

**ETHNICITY ESTIMATE**
- 43% Europe West
- 33% Ireland
- 15% Great Britain
- 3 Other regions

See Full Ethnicity Estimate  🔄 Share

**DNA MATCHES**
- 0 Shared Ancestor Hints
- 🌟 Starred matches
- 🔍 391 4th cousins or closer

View All DNA Matches

Download your raw DNA data
To begin the download, you will need to complete a couple of security steps to protect your information.

GET STARTED  What is raw DNA data?
Download Your Raw Data

- The text file consists of lines of your genotype call data (your A's, T's, C's and G's).
- The download is a zipped text file 5 MB to 30 MB in size.
- It can be opened in a text editor like WordPad or Excel.

![Raw Data Text Document Example](image-url)
User Registration

First & Last Name: Please enter your real name here. Your name is needed by administrators for verification purposes. It may also be displayed in some applications that are not associated with individual results. If you do not want your real name shown in those cases, you may also specify a 'screen name' in the 'alias' field to the left.

Optional Alias: 

Email Address: You must be able to receive a confirmation email at this address. If possible, it should match the email address used for uploads prior to the registration requirement on this site.

Enter email address again: 

Password: Case sensitive. At least 6 characters. Include at least one number or symbol.

Enter Password Again: 

Register
**Analyze Your Data**

- 'One-to-many' matches
- 'One-to-one' compare
- X 'One-to-one'
- Admixture (heritage)
- Admixture/Oracle with Population Search
- Phasing
- People who match one or both of 2 kits *Updated*
- Predict Eye Color
- Are your parents related?
- 3D Chromosome Browser
- Archaic DNA matches
- Multiple Kit Analysis *NEW*
- DNA File Diagnostic Utility
  
  *Analyze DNA file upload for potential problems.*

**Genealogy**

- 1 GEDCOM to all
- 2 GEDCOMs
- Search all GEDCOMs
  - Revised
- GEDCOM + DNA matches
KNOW YOUR GENOME
HELP SCIENCE

LOG-IN/REGISTER

0 0 9 7 4 8 genomes and counting!
Interpretome

Explore your genome
- Load your genome file (upper-right corner) and choose some of the analyses above. Currently, only raw data files from 23andme and Lumigenix (unzipped) are supported.
- Sample genotype files (and a description of the individuals) can be found here.
- A detailed description of the website design and some of the modules can be found in our PSB paper as well as in blog posts here and here.

Interpretome is intended for educational and research purposes only.
No information should be considered diagnostic and as with any genetic testing service, the interpretation is not regulated by the FDA. We assume no responsibility for any injury or damage to persons or property arising out of or related to any use of interpretome annotations or for any errors or omissions: consult your physician with any medical concerns. We retain copyright to the materials herein. By using this website, you agree that you accept these terms and are aware of the risks and benefits of genome interpretation. For more information, please read the full Terms and Conditions.

How are my data kept private?
Your genome will not be sent to any server, it remains on your computer. This website will make requests to a database that only contain "rsid" (without genotypes) and "population" (self-reported in the top-right) information. At no point will any genotypes be sent across the wires (all computation will be done in the browser). Some exercises may have an option to submit personal information, including genotypes or results of analyses, which will be anonymously stored on a secure server.

Compatibility
This website requires an HTML5 compatible browser, including current versions of:
- Google Chrome (≥ 6.0), the preferred browser
GENES THAT INFLUENCE TRAITS

Explore genetic variation that influence YOU
Genes Encode Proteins

Differences in the basepair sequence in DNA within gene regions can affect the functionality of the encoded proteins, resulting in phenotypic differences based on an individual’s genotype.
Why Is This $99 Home DNA Kit Causing Such an Uproar?

23andMe says it can provide you with valuable health information about your genes. The FDA says prove it. What the consumer genetic testing battle means for you.

BY JOSEPH QUINTO

Register your kit at 23andme.com/start
Factors Affecting Risk Assessment

Alzheimer's Disease

Marker Effects

What does this chart show?
The chart shows the approximate effects of the selected person's genotype at the 1 reported marker. Higher, red bars indicate increased risk from the average, while lower, green bars indicate decreased risk from the average. The light gray bars show the maximum possible effects for the possible genotypes at the marker.

Mouse over individual bars to view additional information about each marker. Click on a bar to view detailed information about that marker below. You can read more about all markers in the technical report.
Factors Affecting Risk Assessment

Alzheimer's Disease

Marker Effects

Technical Report

Gene or region: APOE

<table>
<thead>
<tr>
<th></th>
<th>SNPs used</th>
<th>Genotype</th>
<th>Allele</th>
<th>Adjusted Odds Ratio*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lilly Mendel (Mom)</td>
<td>rs7412</td>
<td>CC</td>
<td>ε3/ε4</td>
<td>European: 1.86</td>
</tr>
<tr>
<td></td>
<td>rs429358</td>
<td>CT</td>
<td></td>
<td>Asian: 2.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>African: 1.59</td>
</tr>
<tr>
<td>Greg Mendel (Dad)</td>
<td>rs7412</td>
<td>CC</td>
<td>ε3/ε4</td>
<td>European: 1.86</td>
</tr>
<tr>
<td></td>
<td>rs429358</td>
<td>CT</td>
<td></td>
<td>Asian: 2.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>African: 1.59</td>
</tr>
</tbody>
</table>

*Odds ratios are reported for all available ethnicities.*
Traits Reports
22 Reports

Discover how certain genetic variants can influence your personal characteristics and attributes.

Carrier Status Reports
36 Reports

Learn about variants you may have that may not affect you, but can tell you about potential health risks you might pass on to your children.

Wellness Reports
4 Reports

Understand how your DNA may influence how you respond to certain lifestyle and environmental factors.
Bryant, you are likely to have dark-colored eyes.

65% of customers who are genetically similar to you have dark hazel, light brown, or dark brown eyes.

**Your genetic likelihood**

- Blue: 3%
- Greenish blue: 3%
- Green: 9%
- Light hazel: 20%
- Dark hazel: 25%
- Light brown: 13%
- Dark brown: 27%

**European ancestry customers**

- Blue: 68%
- Greenish blue: 13%
- Green: 13%
- Light hazel: 12%
- Dark hazel: 12%
- Light brown: 6%
- Dark brown: 14%
Cystic Fibrosis

Carrier Status • CFTR gene

Variant not detected

Cystic fibrosis is a rare genetic disorder characterized by impaired lung and digestive function. A person must have two variants in…

View Report
You have no variants detected by this test.

<table>
<thead>
<tr>
<th>Marker Tested</th>
<th>Your Genotype*</th>
<th>Additional Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DeltaF508</strong></td>
<td><strong>CTT</strong> Typical copy from one of your parents</td>
<td>› Biological explanation</td>
</tr>
<tr>
<td><strong>Gene:</strong> CFTR</td>
<td><strong>CTT</strong> Typical copy from your other parent</td>
<td>› Typical vs. variant DNA sequence(s)</td>
</tr>
<tr>
<td><strong>Marker:</strong> i30000001</td>
<td></td>
<td>› Percent of 23andMe customers with variant</td>
</tr>
</tbody>
</table>
|                       |                                                     | › References [5] | [ClinVar](https://clinvar.ncbi.nlm.nih.gov) }
Lactose Intolerance

Dairy products like milk, yogurt, and cheese contain the sugar lactose. An enzyme called lactase breaks down this sugar. If you don’t produce enough lactase, gut bacteria can convert lactose into gas, causing indigestion.

Bryant, you likely produce the lactase enzyme.

Likely not lactose intolerant
Bryant, you likely produce the lactase enzyme.

**Likely not lactose intolerant**

<table>
<thead>
<tr>
<th>Marker Tested</th>
<th>Your Genotype*</th>
<th>Additional Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C/T-13910</strong></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td><em>Gene: Near LCT</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Marker: rs4988235</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Typical copy from one of your parents</td>
<td></td>
<td>Biological explanation</td>
</tr>
<tr>
<td>Typical copy from your other parent</td>
<td></td>
<td>Typical vs. variant DNA sequence(s)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percent of 23andMe customers with variant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>References [1, 3, 6]</td>
</tr>
</tbody>
</table>
Lactase intolerance, *LCT* gene

Infantile expression of lactase, with a cessation of production in adulthood – lactose intolerance

Infantile expression of lactase, with continued expression into adulthood
Lactase intolerance, 
*LCT* gene

Infantile expression of lactase, with a cessation of production in adulthood – **lactose intolerance**

Infantile expression of lactase, with continued expression into adulthood
### Genes

<table>
<thead>
<tr>
<th>Genes</th>
<th>Marker (SNP)</th>
<th>Genomic Position</th>
<th>Variants</th>
<th>Your Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCM6</td>
<td>rs182549</td>
<td>136616754</td>
<td>C or T</td>
<td>T / T</td>
</tr>
</tbody>
</table>
Promethease

The easiest way to make a report is to visit promethease.com

This takes about 10 minutes and costs $5 (all major currencies and credit cards accepted).

How-to videos

- Making a Promethease report
- Reading a Promethease report

An example report based on either

- 23andMe v1 (2008)
- 23andMe v3 (2011)
- 23andMe v4 (2014)

What is Promethease? [edit]

Promethease is a literature retrieval system that builds a personal DNA report based on the scientific literature cited in SNPedia and a file of genotype (DNA) data. Customers of DNA testing services (23andMe, FamilyTreeDNA, Ancestry.com, Complete Genomics, ...) can use it to retrieve published data about their DNA completely independent of whichever company produced the data.

Read about the features of Promethease.
Before you may use Promethease to retrieve information about the human genome, you must read and agree to the following statements. Please read each statement and check the box next to each one and then click 'I Agree'.

☐ I understand that the information provided in my Promethease report is based on SNPedia.com and that my report is for educational and research purposes only.

☐ I understand that my report is deleted after 45 days but that I can download it before it is deleted.

☐ I realize that most published reports about DNA variations explain only a small part of the heritability of a trait, and they also don't take into account how different variants might interact. In addition, published reports typically ignore environmental, dietary, microbial, medical history and lifestyle factors, any or all of which may well affect my true risk for any trait or disease.

☐ I am aware that I am strongly encouraged to discuss my Promethease report with a doctor, genetic counselor or other health-care provider prior to making any medical or reproductive decisions. I also acknowledge that I am advised to confirm any significant finding discovered in part through the use of Promethease by an independent, clinically validated test for use in connection with the medical trait in question.

☐ I have read and understand the Privacy Policy and the Legal Terms and Conditions of this website. I agree to these conditions.
REVELATIONS OF THE PERSONAL GENOME

Be careful out there in genome land!
More Learning Opportunities

• DNA: Unlocking the Code – Apr 9, 2016
  Iowa Genealogical Society, Des Moines
  http://iowagenealogy.org/?page_id=5265

• DNA Day Iowa City – Apr 24-25, 2016
  Julie Granka, PhD from AncestryDNA
  Public Presentation, 3:00 pm, April 24, 101 BBE

  Hands-on Workshop hosted by Personal Genome Learning Center
  12:00-3:00 pm, April 25, Computer Room, Iowa City Public Library