

A photograph of an Illumina HiSeq 2000 DNA sequencer in a laboratory setting. The machine is white with black vertical supports and has two computer monitors attached to it. A green light bar is visible on the right side of the machine. The text "HiSeq 2000" is printed on the top right of the machine's body.

# Autosomal DNA (atDNA)

**Don Worth**  
[worth@ucla.edu](mailto:worth@ucla.edu)

Illumina HiSeq 2000 DNA Sequencers at  
Family Tree DNA Labs  
Courtesy of CeCe Moore, [yourgeneticgenealogist.com](http://yourgeneticgenealogist.com)

# Why Test Autosomal DNA?

- Find relatives who are researching your family
- Prove a relationship between living relatives
- Determine lines of descent
- Authenticate your genealogical research
- Find “Deep Ancestry” (aka Admixture)

*But it won't replace good,  
old-fashioned genealogical research!*

# Privacy Concerns

- Identify theft?
  - Anyone who really wanted your DNA can get it
  - There are easier ways to make a (dishonest) living!
  - Social Media is probably a bigger risk than DNA
  - You can test under a pseudonym
- Health Information?
  - Genetic Information Nondiscrimination Act
    - Prevents unfair treatment due to DNA differences (health)
    - Life insurance, disability insurance, long-term care excepted
  - Family Tree DNA and Ancestry “scrub” medical DNA from results to enhance privacy

# Genetic Genealogy Testing

## Four kinds of tests

- **Y-DNA**

Dad's-dad's-dad... going way back (test males only)

- **mtDNA**

Mom's-mom's-mom... going way back

- **Autosomal DNA**

Any line but best up to 5 or 6 generations

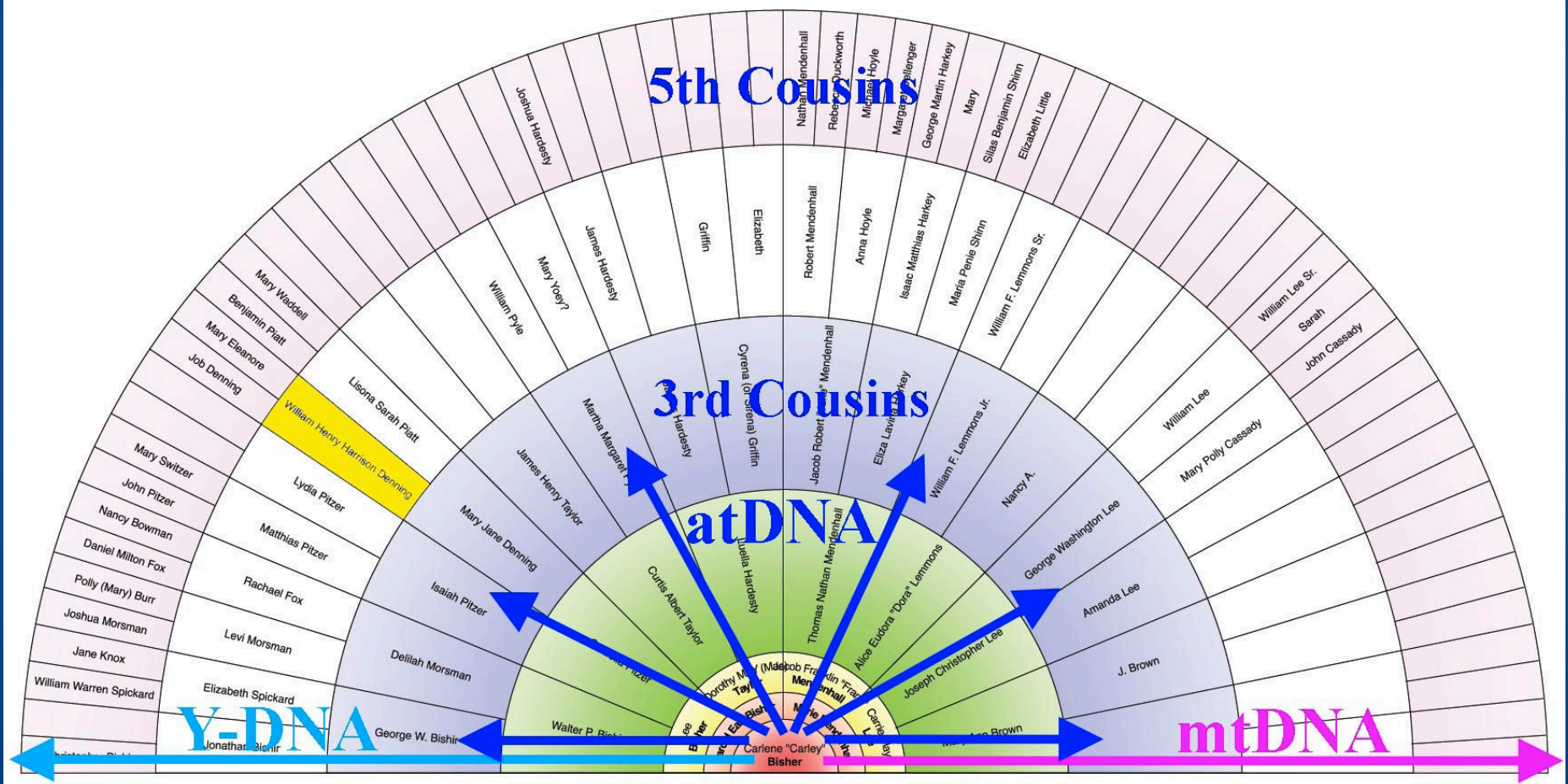
May include admixture (deep ancestry) & medical

- **X-DNA**

Many lines but males don't get an X from dad



# “Fan” Pedigree Chart



# Major Testing Companies

- **Family Tree DNA** – “Family Finder”  
More genealogists (75K), best tools, 3<sup>rd</sup> party transfers (\$69)
- **23andMe** – “Relative Finder”  
Largest database (400K), medical results, broader test
- **Ancestry.com** – “AncestryDNA”  
Big database (200K), link to Ancestry trees, poor tools

*All are \$99 now*

*All can export raw data for 3<sup>rd</sup> party tools*

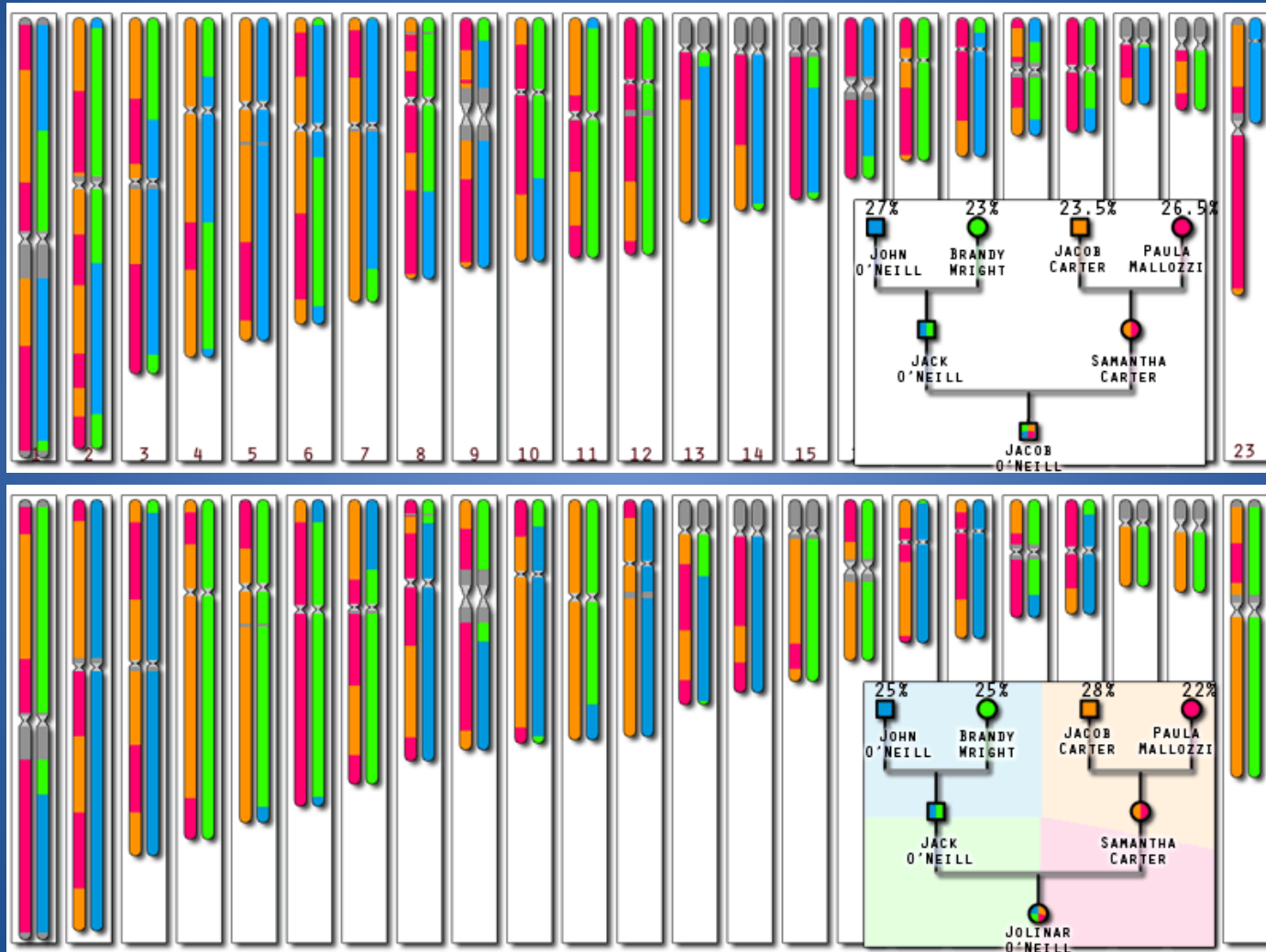
[www.isogg.org/wiki/Autosomal\\_DNA\\_testing\\_comparison\\_chart](http://www.isogg.org/wiki/Autosomal_DNA_testing_comparison_chart)

# atDNA Video

[http://www.smgf.org/education/  
animations/autosomal.jspx](http://www.smgf.org/education/animations/autosomal.jspx)

# Recombination: Two Siblings

<http://www.dnainheritance.kahikatea.net/autosomal.html>



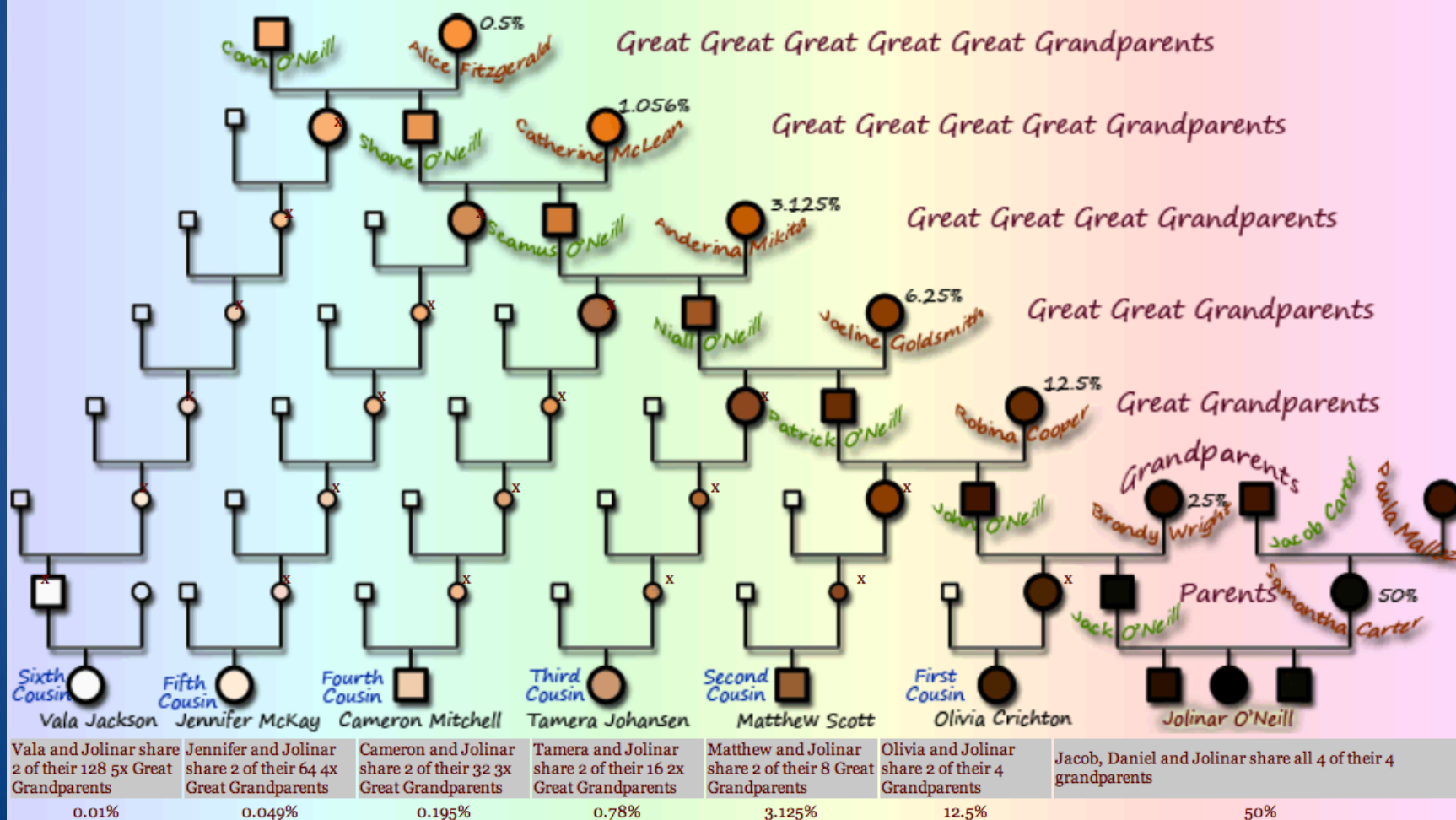
Extra Credit: Which one is a boy? Which one is a girl?



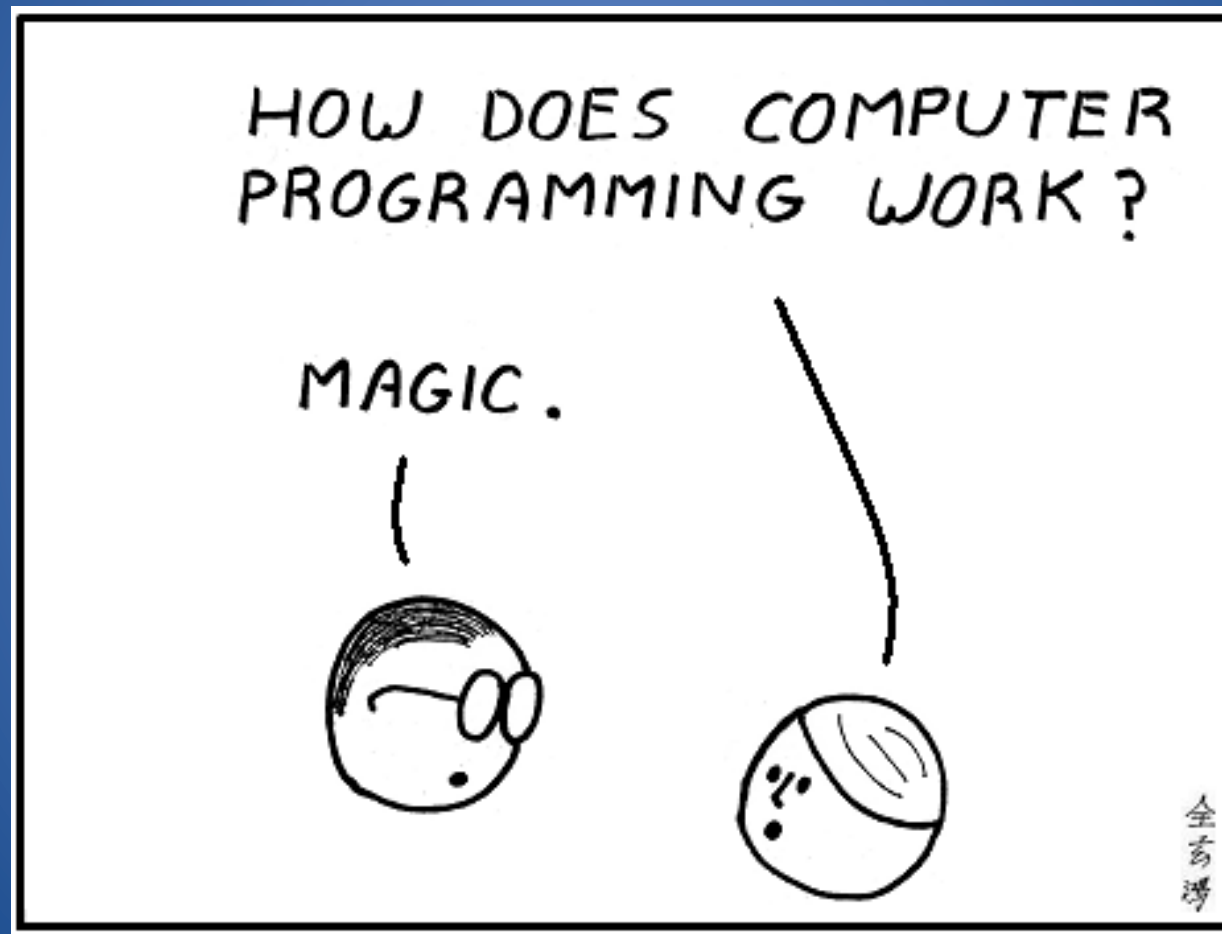
# Recombination: Finding Cousins

<http://www.dnainheritance.kahikatea.net/autosomal.html>

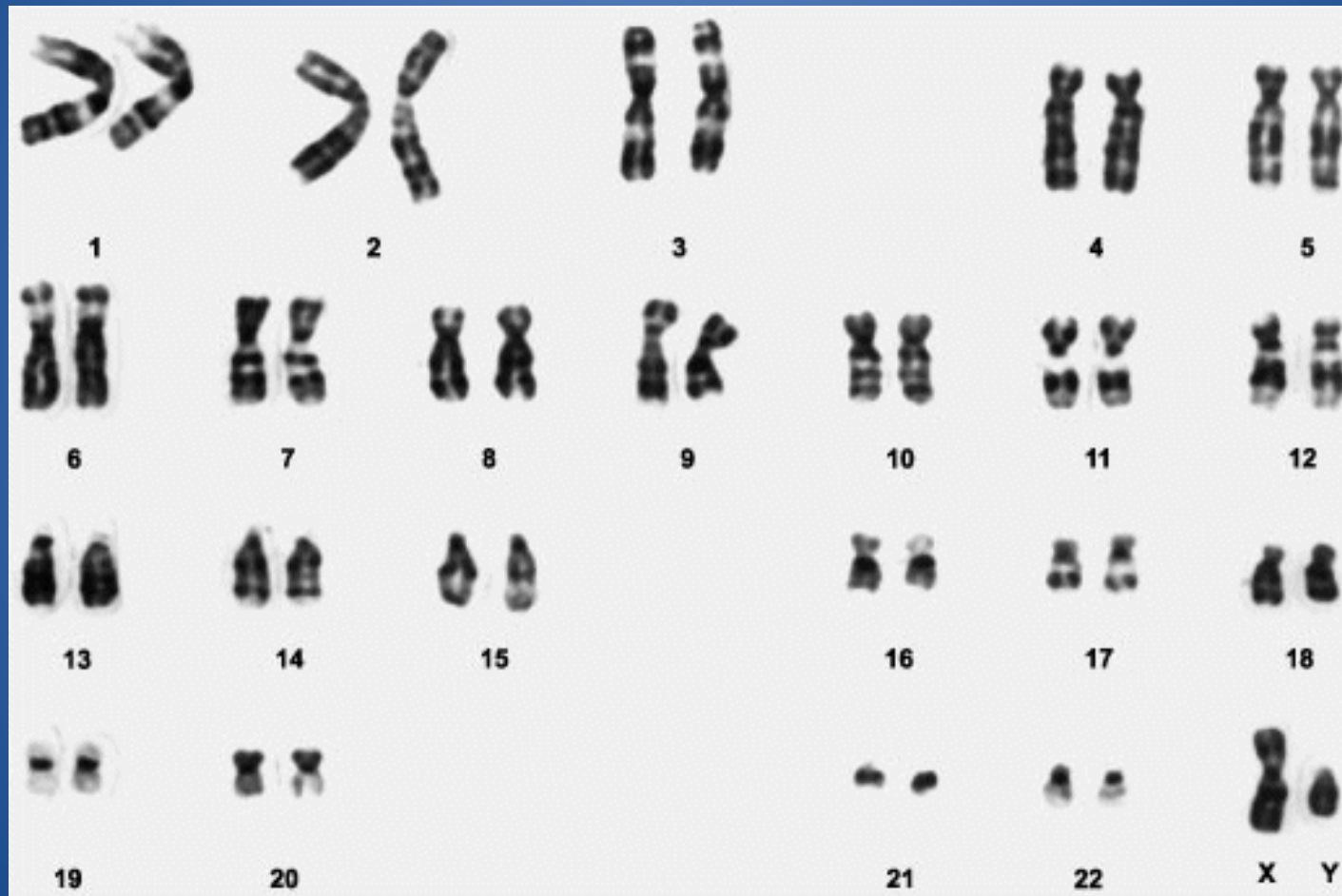
Theoretical %'s shared with different types of cousin



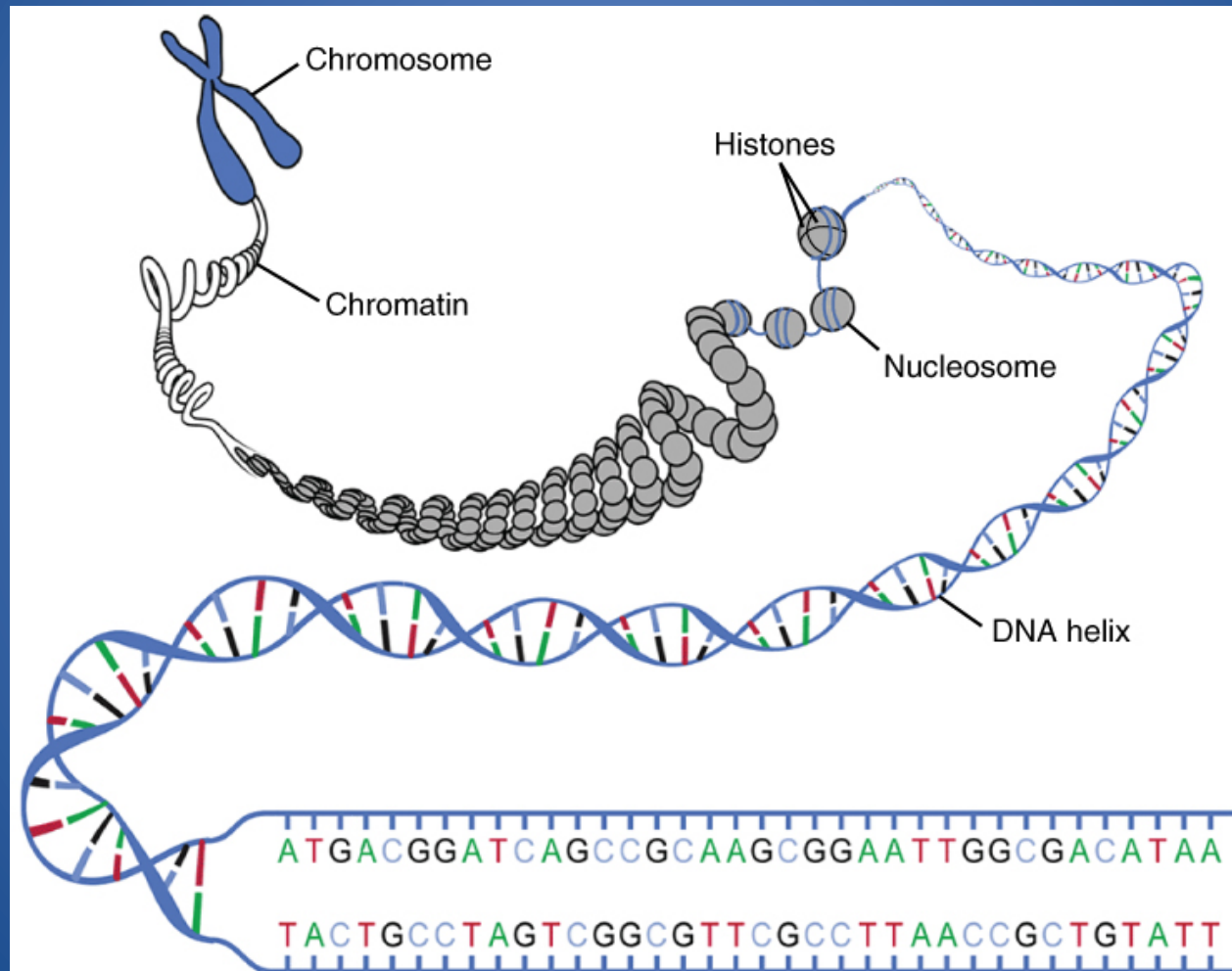
# But how does atDNA work?



# 23 Chromosome Pairs

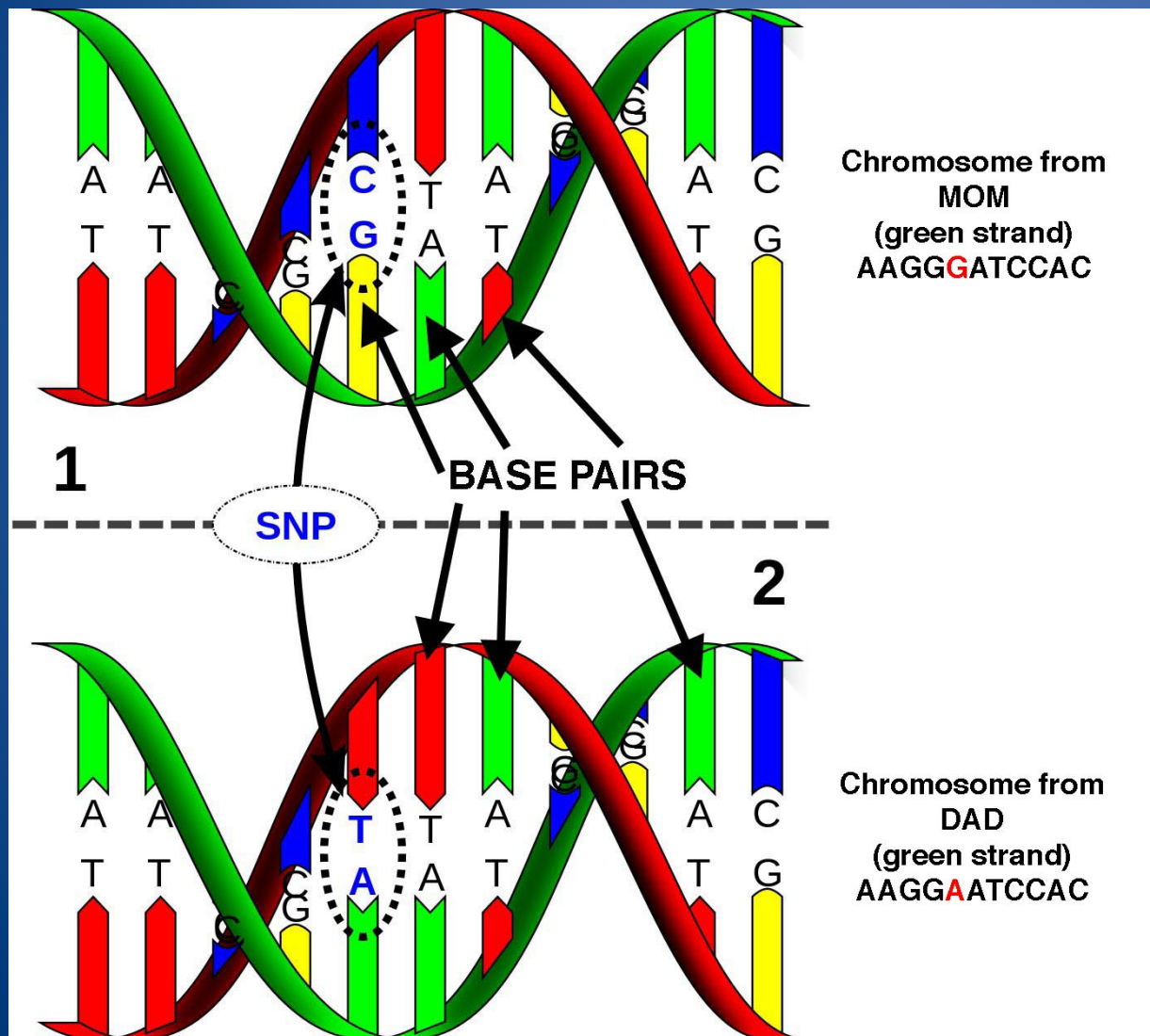


# Inside a Chromosome





# SNP=Single-Nucleotide Polymorphism



A = Adenine  
G = Guanine  
C = Cytosine  
T = Thymine

A always pairs with T  
C always pairs with G

99.9% of base pairs are the same in all humans. The rest can vary and are called SNPs ("Snips")

FTDNA tests about 700,000 out of about 10 million possible SNPs and 3 billion base pairs



# Raw Data from FTDNA

A	B	C	D
RSID	CHROMOSOME	POSITION	RESULT
rs3094315	1	742429	AG
rs3131972	1	742584	AG
rs12562034	1	758311	GG
rs12124819	1	766409	AA
rs11240777	1	788822	AG
rs6681049	1	789870	CC
rs4970383	1	828418	AC
rs4475691	1	836671	TC
rs7537756	1	844113	AG
rs13302982	1	851671	GG
rs1110052	1	863421	GG
rs2272756	1	871896	AA
rs17160698	1	877025	TT
rs3748597	1	878522	CC
rs13303106	1	881808	AA
rs28415373	1	883844	CC
rs13303010	1	884436	AA
rs6696281	1	892967	CC
rs28201282	1	894028	CC

## “Unphased” Alleles

One SNP value (allele) is Mom’s and one is Dad’s - can’t tell which is which

Either allele can match either of someone else’s two alleles at the same location

So you can match two different unrelated people on the same segment of SNPs

# What's in a match?

Comparing DNA segments - it's about the odds

Identical By Descent (**IBD**) vs **Non-IBD** (aka IBS)

- Long segments are passed down from a recent ancestor (IBD)
- Small but ubiquitous segments (eg. left-handed) are older and more humans have them (non-IBD)
- Population “bottle-necks” or inbred communities may contribute more & longer non-IBD segments

# What's in a match?

## Centimorgans (cM) and SNPs

- Can think of cM as segment “length” (kinda)
- Based on probability to recombine “breakability”
- One cM has 1% chance to recombine in one generation
  - Averages one million base pairs
  - Usually 100 or more SNPs
- Matching segments with higher cM and more SNPs indicate more recent common ancestor

# What's in a match?

- Family Tree DNA: Proprietary algorithm
  - Longest segment at least 5.5 cM and 500 SNP
  - And additional matching segments(?)
  - De facto 7 cM
- Companies use different thresholds
  - 23andme's is 7 cM and 500 SNPs
  - Some people use "7 & 7", some use 10 cM
- May be able to identify smaller IBD segments by using other accompanying factors

# What's in a match?

- To avoid false negatives, more false positives:

Matching Segment	% IBD	% IBS	Matching Segment	% IBD	% IBS
11 cM	>99%	<1%	7 cM	30%	70%
10 cM	99%	1%	6 cM	20%	80%
9 cM	80%	20%	5 cM	5%	95%
8 cM	50%	50%	4 cM	1%	99%

- 5<sup>th</sup> or 6<sup>th</sup> cousins not too likely to match:  
(but you will probably find a few anyway)

Relationship	Match Probability
2nd cousins or closer	> 99%
3rd cousin	> 90%
4th cousin	> 50%
5th cousin	> 10%
6th cousin and more distant	Remote (typically less than 2%)

But take heart, you can improve your odds!



# “Family Finder” Demo

[www.familytreeDNA.com](http://www.familytreeDNA.com)

- Match Lists, Ancestral Surnames, and GEDs
- Triangulation
  - Eliminates unwanted matches from unphased data
  - Limits pool of matches to most likely possibilities
  - “Zooms in” to limit matches to one area of your tree
- Chromosome Browser
- Population Finder
  - The autosomal equivalent of Haplogroup for Y-DNA
  - Multiple methodologies and more databases needed
  - 23andMe is better at this than FTDNA or Ancestry

# 3<sup>rd</sup> Party Tools

## GEDMATCH.COM – Free DNA tools web site

- LARGE common database across all three testing companies
  - ✓ We have 1000 matches in GEDMATCH vs 500 on FTDNA
- One to Many (matching)
  - ✓ Can set your own matching thresholds
- One to One (with "chromosome browser")
- X Chromosome tools
- Admixture (4 variants)
- Find People Who Match on a Specific Segment
- Software Phasing (needs child & at least one parent)
- Predict Eye Color
- Are your parents related?
- (Triangulation and Genealogies currently unavailable)

# 3<sup>rd</sup> Party Tools

## GEDmatch.Com Autosomal Comparison

Base Pairs with Full Match =	
Base Pairs with Half Match =	
Match with Phased data =	
Base Pairs with No Match =	
Base Pairs not included in comparison =	
Matching segments greater than 3 centiMorgans =	

Comparing Kit F242768 (Don Dyson Worth) and F199791 (Carley Bisher Worth)

Minimum threshold size to be included in total = 500 SNPs

Mismatch-bunching Limit = 250 SNPs

Noise Reduction Threshold = 0.80

Minimum segment cM to be included in total = 3.0 cM

Chr 1

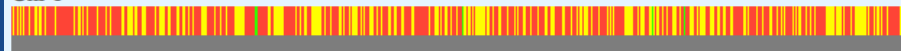


Image size reduction: 1/58

Chr 2



Image size reduction: 1/57

Chr 3



Image size reduction: 1/47

Chr	Start Location	End Location	Centimorgans (cM)	SNPs
4	47,103,595	55,379,333	4.0	845

Chr 4

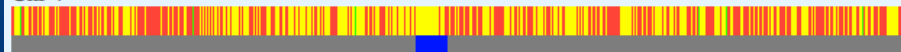


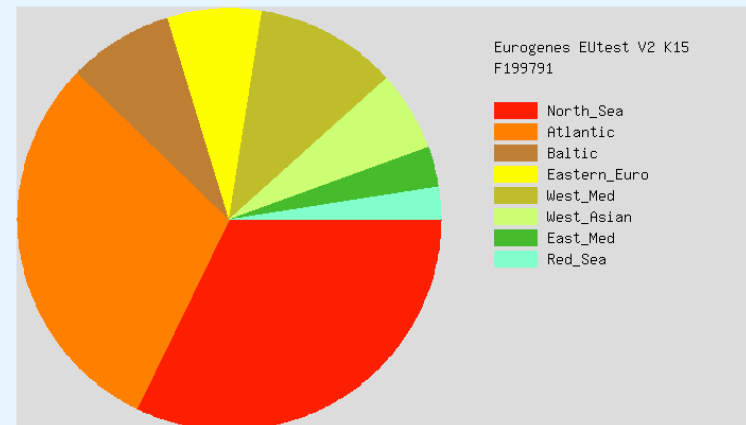
Image size reduction: 1/40

## Eurogenes EUtest V2 K15 Admixture Proportions

This utility uses the Eurogenes EUtest V2 K15 model, created by Davidski (Polako). Questions and comments about this model should be directed to him at his [Project Blog](#).

Kit Number: F199791 Iteration: 1000 Delta-Q: 5.280226e-04 Elapsed Time: 83.26 seconds

Population	
North_Sea	32.16%
Atlantic	30.06%
Baltic	8.00%
Eastern_Euro	7.42%
West_Med	10.82%
West_Asian	5.92%
East_Med	3.19%
Red_Sea	1.80%
South_Asian	-
Southeast_Asian	-
Siberian	0.25%
Amerindian	-
Oceanian	0.39%
Northeast_African	-
Sub-Saharan	-



Oracle

Oracle-4

☐ Pct. Calc. Option 1

☒ Pct. Calc. Option 2

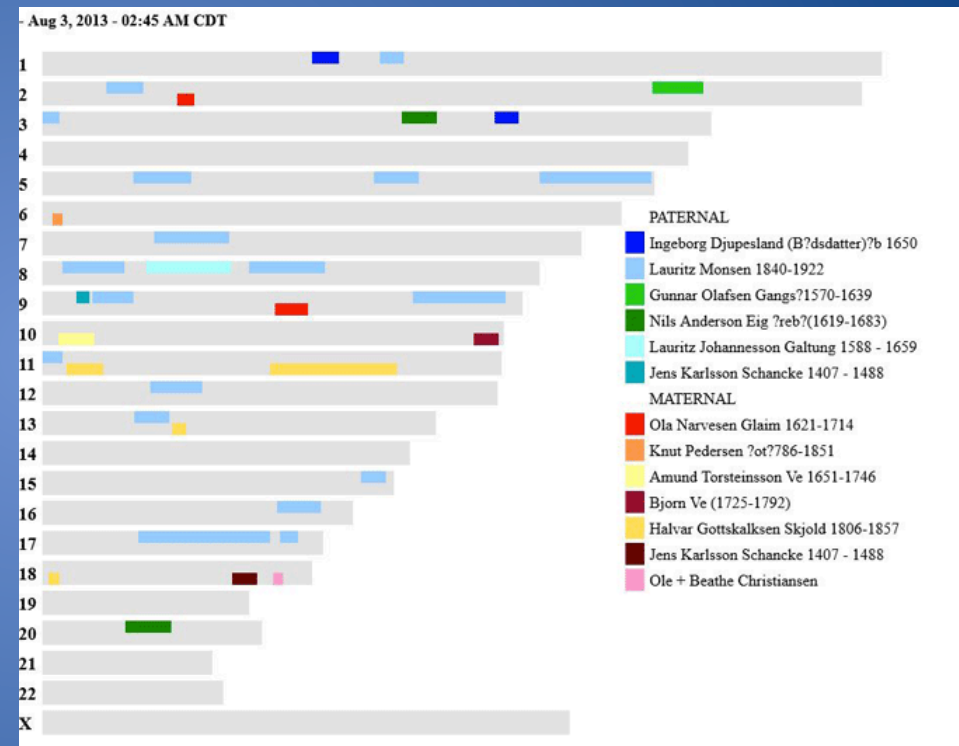
Oracle-x

Web site and contents ©Copyright 2011-2013 by [GEDmatch, Inc.](#)  
Genealogy and DNA data remains the property of the submitter.  
Each Admixture Proportions 'calculator' model remains the property of its developer.

# 3<sup>rd</sup> Party Tools

## Kitty Cooper's Chromosome Graphs

- Segment Mapper  
Graph & compare up to 40 people
- Chromosome Mapper  
Map your genome



[blog.kittycooper.com](http://blog.kittycooper.com)

# Suggestions

- Upload your GED and surnames!!!
- Put your data on GEDMATCH so even more people will find you
- Analyze your matches and contact them to compare trees



# More Suggestions

- Test, test, test, test... (aka \$, \$, \$, \$...)
  - Test one or both of your parents or their siblings to help with triangulation & phasing
  - Test siblings to find matches you don't have
  - Test as many known cousins as you can to use with triangulation and labeling segments
  - Test older generations for more “reach”
- Map your genome - spreadsheet

# Read All About It

- **Kelly Wheaton's Beginners Guide to Genetic Genealogy:**  
Great and thorough overview with lots of wonderful links.  
<http://sites.google.com/site/wheatonsurname/beginners-guide-to-genetic-genealogy>
- **International Society of Genetic Genealogy WIKI:**  
Nice comparisons of testing companies, definitions, etc.  
[http://www.isogg.org/wiki/Wiki\\_Welcome\\_Page](http://www.isogg.org/wiki/Wiki_Welcome_Page)
- **Excellent BLOGs with tutorials, news, tips, and tidbits:**  
Roberta Estes: <http://dna-explained.com>  
Ce Ce Moore: <http://www.yourgeneticgenealogist.com>  
Kitty Cooper: <http://blog.kittycooper.com>  
Blaine Bettinger: <http://www.thegeneticgenealogist.com>
- **DNA-NEWBIE Yahoo Group (must join ISOGG first):**  
<http://groups.yahoo.com/neo/groups/DNA-NEWBIE/info>

Questions?