DNA for Genealogy

Jay Sage

Jewish Genealogical Society of Greater Boston
Jay@SageFamily.org

PDF of slides at: DNA.SageFamily.org

Jewish Genealogy Society
of the Conejo Valley and Ventura County

October 11, 2015

Rev. 2015-10-07

Outline

- Introduction
- Basic biology background
- How DNA can inform genealogy
- Interpreting results and examples
- How testing is done

Introduction

- DNA testing is a complex subject
  - Don’t expect to understand it all at first
  - Review this talk several times
  - Listen to other speakers at conferences
  - Read educational materials on testing sites
- This talk is oversimplified
- Its aim is to elucidate the basic principles

Introduction

- Genetic testing can answer some (but only a limited number of) genealogical questions that documentation and family stories cannot
- Science has developed since the discovery of the chemical structure of DNA in 1953
- Made practical by the advances in DNA sequencing technology to the point where individuals can afford it

Genealogical Value of DNA Testing

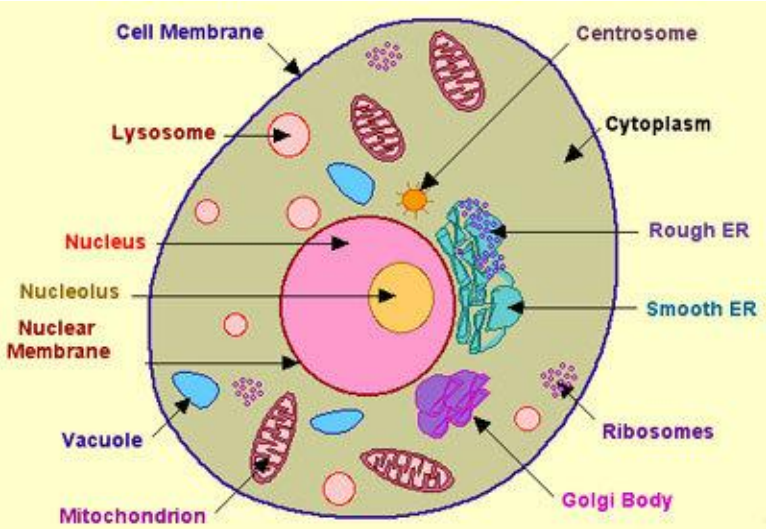
- My opinion in summary: mostly useless and a waste of money
- DNA testing has value when either:
  - You have a specific relationship hypothesis to test
  - You have a relative who is the end of a line and you want to get a sample for future use
  - You have the money to spare and want to have some fun
- Money generally better spent on Ancestry and/or MyHeritage subscriptions

Outline

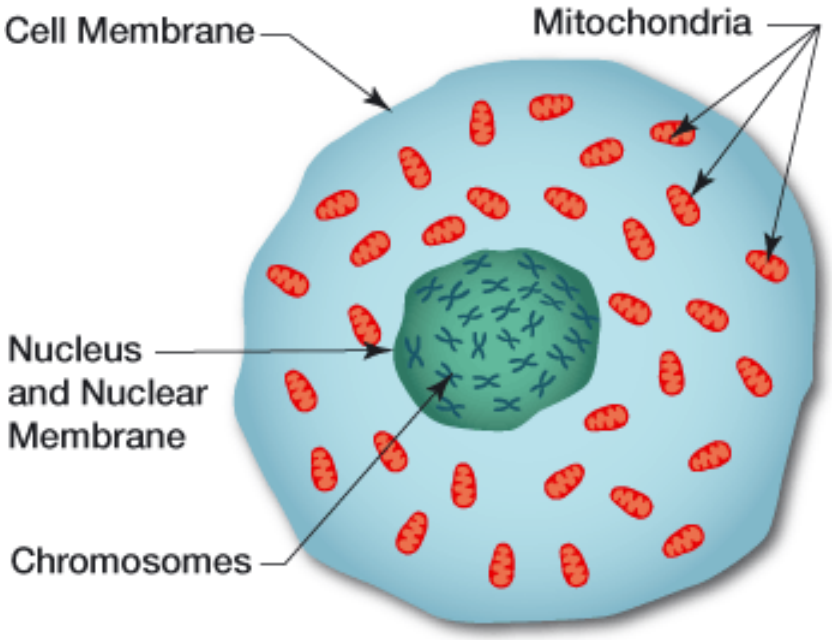
- Basic biology background
  - The cell and its genetic components
  - Nature of DNA
  - Chromosomes
  - Reproductive cells and progeny
  - Mitochondrial DNA
  - Mutations (SNPs and STRs)

Basic Structure of Cells

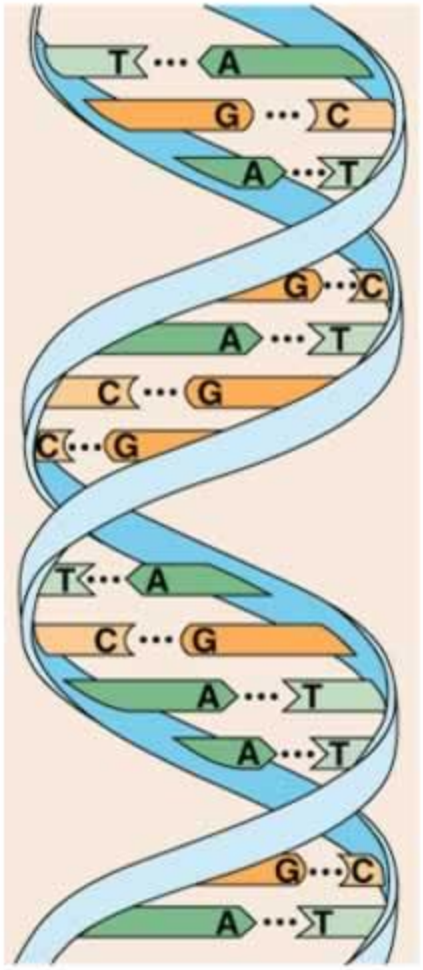
Complete Cell Structure

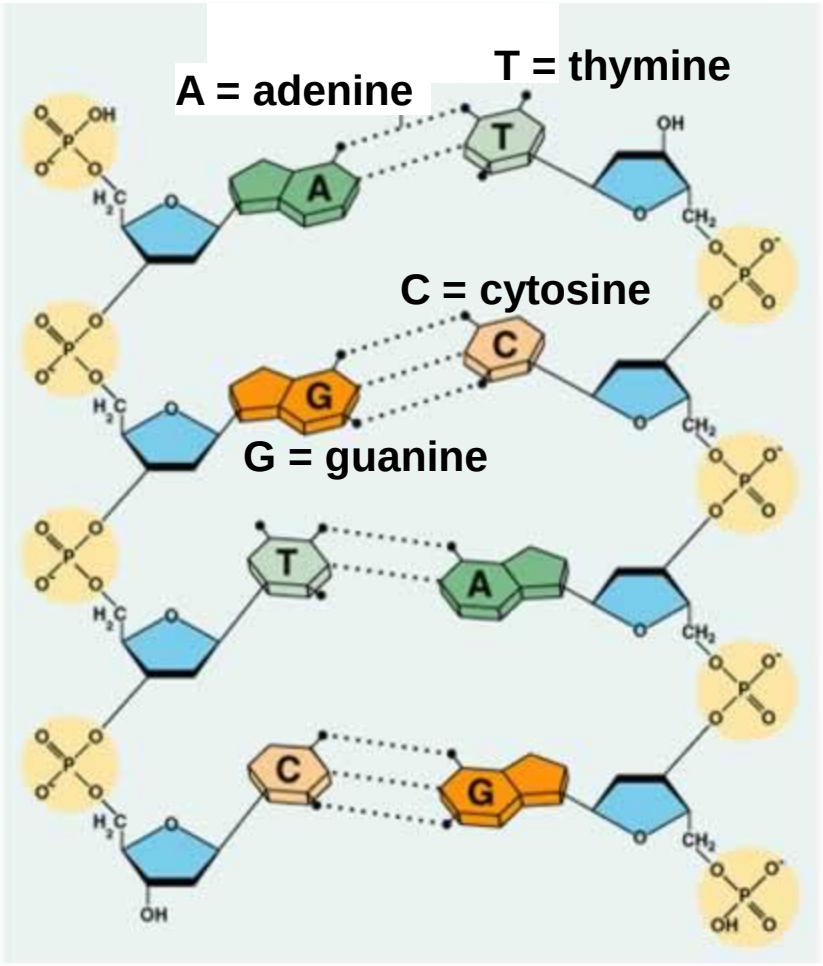


Structures with Genetic Material



Structure of Genetic Material: DNA

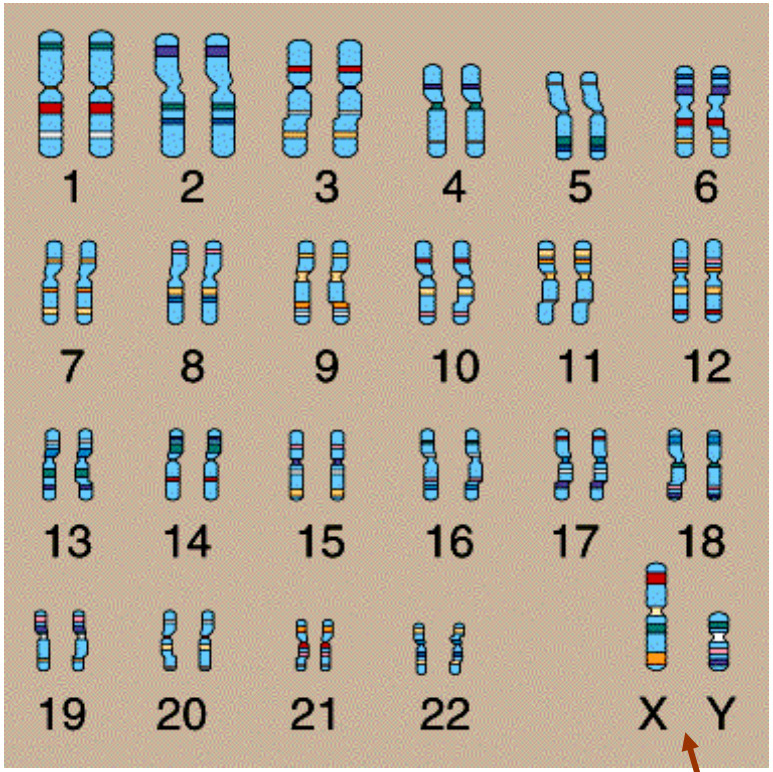


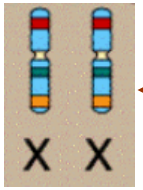


Structure of DNA

- Sequence of only 4 bases (A, T, C, G)
- Total of ~3 billion bases in the human genome (i.e., the DNA in cell nucleus)
  - Complete blueprint for creating the organism
- DNA is 99.9% identical in all humans
  - But that leaves ~3 *million* differences!
  - Thus all but identical twins are genetically unique
- Mitochondrial DNA (mt-DNA) has 16,569 bases

Chromosomes





- DNA is normally dispersed as threads throughout the nucleus (in all ~5' long if stretched out)
- During cell replication, the DNA winds up tightly as shown here
- 46 separate strands of DNA called chromosomes
  - 44 comprise 22 pairs that carry the same genes, <u>one each from father & mother</u>
  - Two are special chromosomes: X and Y
    - <u>Females</u> have two Xs, one from each parent
    - <u>Males</u> have one X (from mother) and one Y (from father)

Cell Division in Normal Tissue Growth

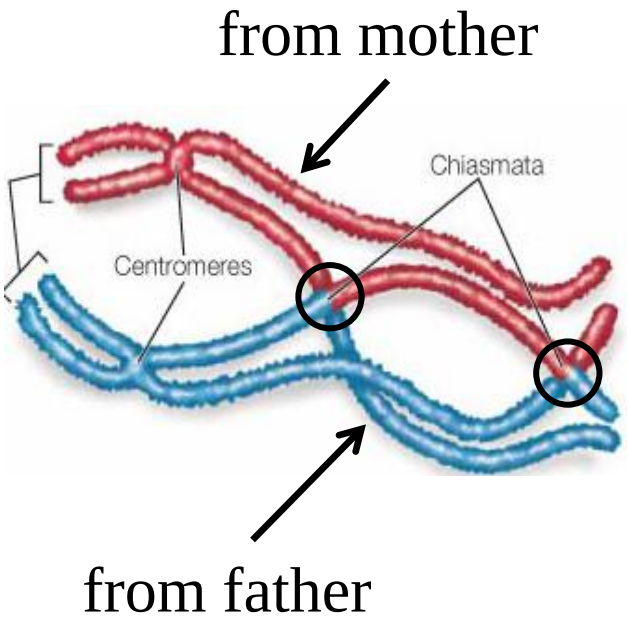
- All 46 chromosomes duplicate (essentially) exactly and separate into the daughter cells
- The result is two cells with DNA identical to that of the original cell
- Consequently, all tissue cells in the body generally have the same DNA (exceptions covered later)

Outline

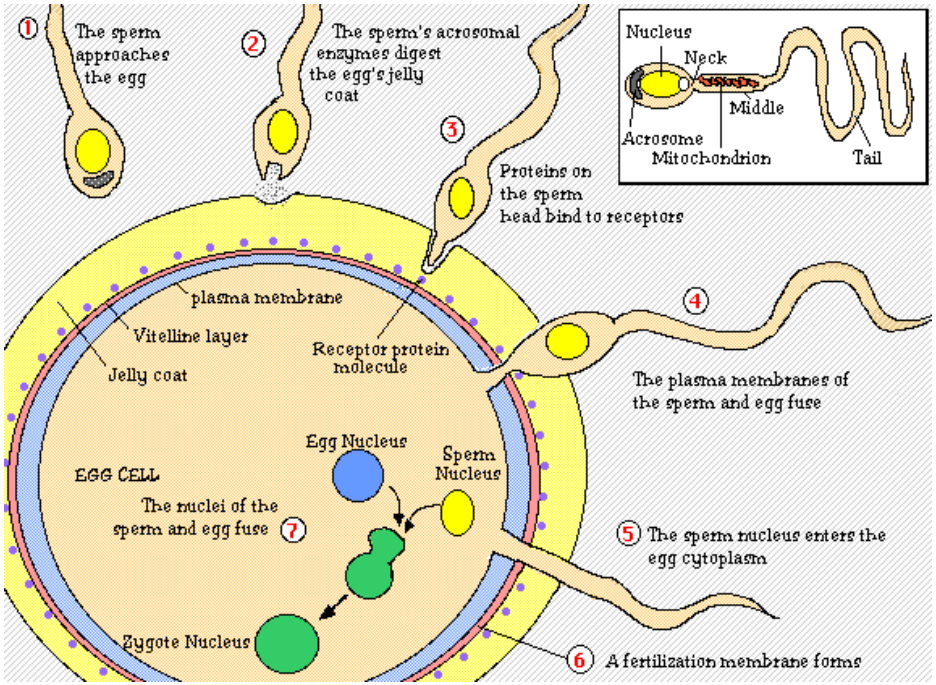
- Basic biology background
  - The cell and its genetic components
  - DNA
  - Chromosomes
  - ➔ Reproductive cells and progeny
  - Mitochondrial DNA
  - Mutations (SNPs and STRs)

Cell Division for Eggs and Sperm

- First, the chromosomes duplicate
- Then chromosome pairs **swap parts**; results thus generally contain DNA from each parent's father **and** mother (and thus from many earlier ancestors)
- Eggs and sperm get one chromosome of each pair
  (22 for sperm, 23 for eggs incl. X)
- The 23rd chromosome of sperm is *either* the X **or** the Y from the **father** (**unmixed**)



Formation of Embryo



- Nucleus of sperm enters egg and combines with egg's nucleus, restoring the count of 46 chromosomes
- If sperm had X, child is a female
- If sperm had Y, child is male

Critical Fact #1
for Genetic Genealogy

- A male receives a Y chromosome **unmixed** from his father, who received it from his father, who received it from his father, etc.
- Thus the Y chromosome is relatively easy to trace (and it follows surnames)
- Testing a (relatively) few locations (markers) on the DNA is sufficient to determine a match or mismatch to another person

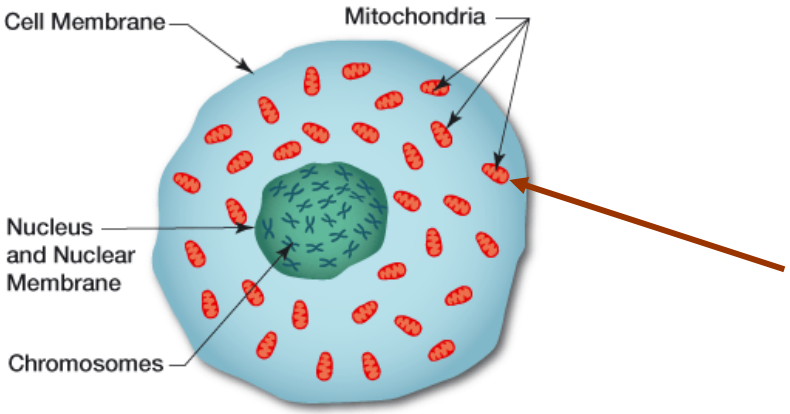
Aside: Y-Adam

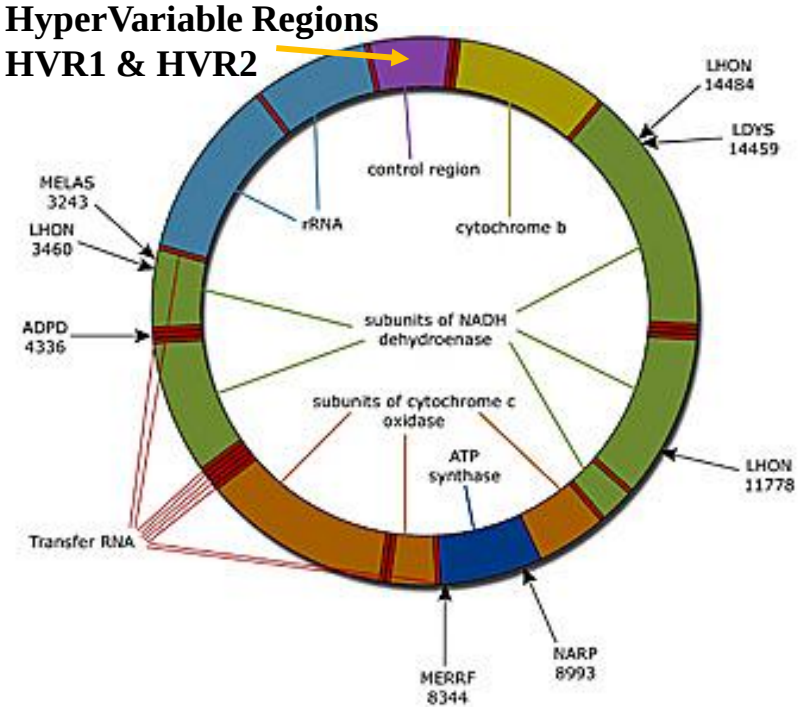
- Y chromosomes tend to disappear
  - If a man has no sons, his Y chromosome is lost
- Example
  - My great-grandfather Moshe-Yossel Rigler has ~130 descendants, but none has his Y chromosome
  - He had six children but only two sons
    - One son (my grandfather) had only three daughters
    - Other son had a son, but that son was adopted

Aside: Y-Adam

- Result: all men living today have the Y chromosome from **one** common ancestor (fancifully named "Y-Adam")
- Today's Y-Adam is estimated to have lived 120,000–340,000 years ago (in Africa)
- But today's Y-Adam is not tomorrow's Y-Adam, since Y chromosomes are constantly both changing and being lost
- While men today have the **Y-chromosome** from only one man (Y-Adam), they have other DNA from many men of that time

Mitochondrial DNA (mt-DNA)





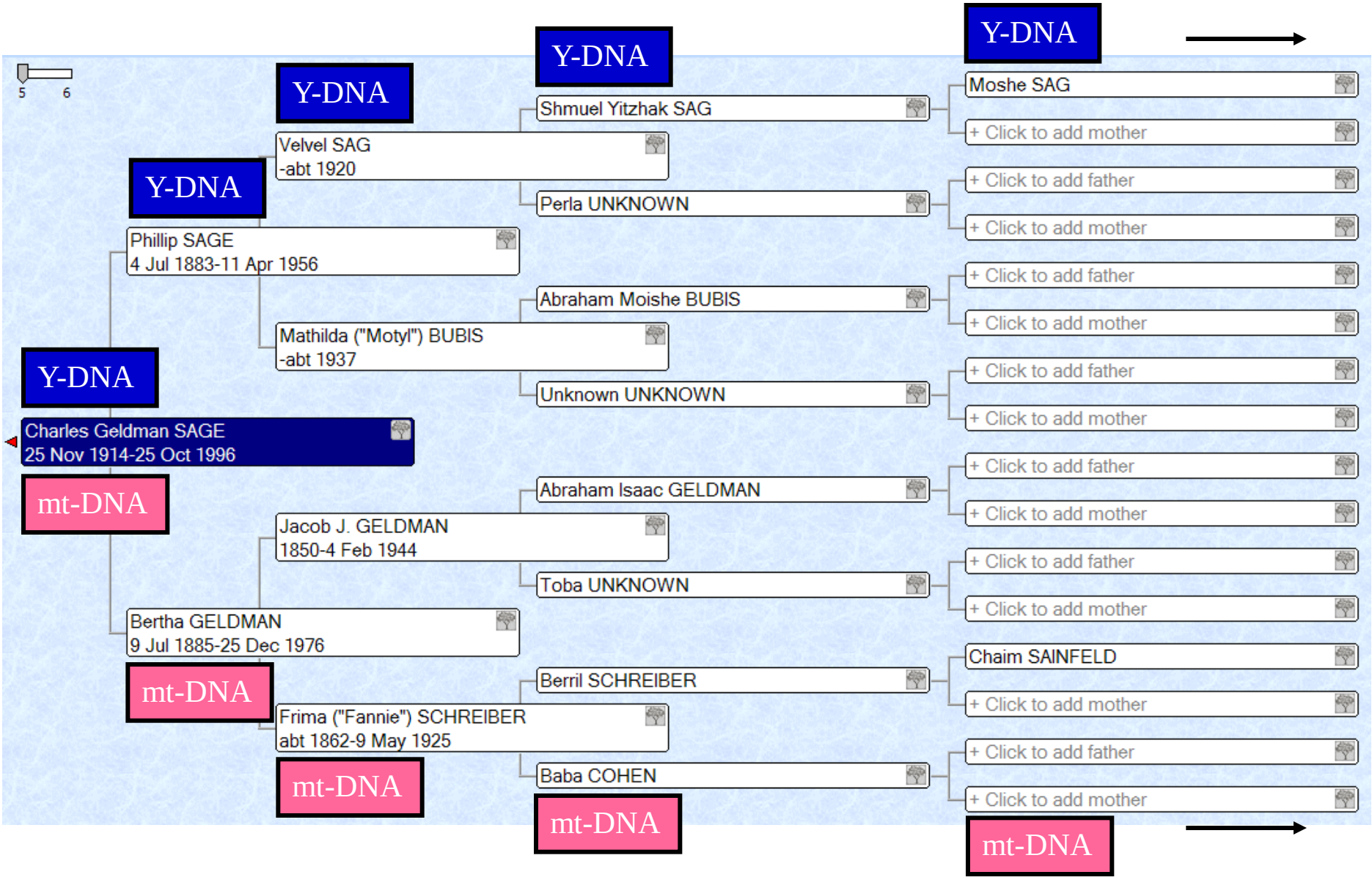
- Cell nucleus is not the only cell organelle with DNA
- Mitochondria also have DNA (they are ancient bacterial “guests”)
- Ring shaped with 16,569 bases (Y has 60,000,000)
- Several rings of DNA per mitochondrion
- 100–1,000,000 mitochondria per cell

Critical Fact #2 for Genetic Genealogy

- Sperm's mt-DNA generally does not enter the egg (or is destroyed by the egg if it does)
- **All children** receive their mt-DNA **unmixed** from their mother, who received it from her mother, who received it from her mother, etc.
- Thus the mt-DNA, like Y-DNA, is relatively straightforward to trace (but doesn't follow surnames)

Aside: mt-Eve

- Mitochondrial DNA tends to disappear
  - If a woman has no daughters, her mt-DNA is lost in the following generation (sons get it but do not pass it on)
- Example
  - My great-grandmother Frima Schreiber has more than 30 descendants, but none has her mt-DNA
  - She had five children but only one daughter
    - That daughter (my grandmother) had only a son (my father)
- Result: all men and women living today have mt-DNA from **one** common ancestor ("mt-Eve") who lived 100,000–200,000 years ago (in Africa)


5
6
Y-DNA
Y-DNA
Y-DNA
Y-DNA
Charles Geldman SAGE
25 Nov 1914-25 Oct 1996
mt-DNA
mt-DNA
mt-DNA
mt-DNA
Phillip SAGE
4 Jul 1883-11 Apr 1956
Velvel SAG
-abt 1920
Mathilda ("Motyl") BUBIS
-abt 1937
Bertha GELDMAN
9 Jul 1885-25 Dec 1976
Jacob J. GELDMAN
1850-4 Feb 1944
Frima ("Fannie") SCHREIBER
abt 1862-9 May 1925
Shmuel Yitzhak SAG
Perla UNKNOWN
Abraham Moishe BUBIS
Unknown UNKNOWN
Abraham Isaac GELDMAN
Toba UNKNOWN
Berril SCHREIBER
Baba COHEN
Moshe SAG
+ Click to add mother
+ Click to add father
+ Click to add mother
+ Click to add father
+ Click to add mother
+ Click to add father
+ Click to add mother
+ Click to add father
+ Click to add mother
+ Click to add father
+ Click to add mother
Chaim SAINFELD
+ Click to add mother
+ Click to add father
+ Click to add mother

Outline

- Basic biology background
  - The cell and its genetic components
  - DNA
  - Chromosomes
  - Reproductive cells and progeny
  - Mitochondrial DNA
  - ➔ – Mutations (SNPs and STRs)

Mutations: the Key to Identification Using Y-DNA and mt-DNA

- If DNA replication were perfect, we would all have the same DNA
- Mutations create variety—and make genetic genealogy possible
- Possible causes of mutation:
  - Random chemical errors
  - Radiation
  - Environmental toxins

Two of the Types of Mutations

- Single Nucleotide Polymorphism: SNP
  - Called a “snip”
  - Change in a single base (e.g., A→G)
  - Measured for mt-DNA and autosomal DNA tests
- Short Tandem Repeat: STR
  - Sections of chromosomal DNA contain repeating sequences (e.g., CATG CATG CATG CATG)
  - Number of repeats changes
  - Measured for Y-DNA tests

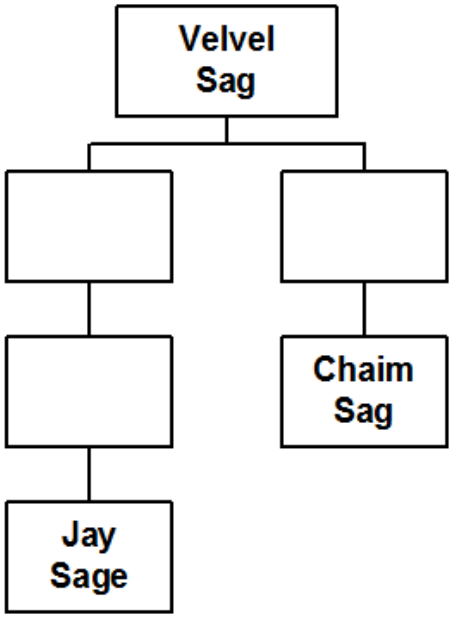
Mutation Frequencies

- Short Tandem Repeat (STR) mutations: fast
  - Once in 250-500 generations per STR
  - Need ~50-100 markers to see mutation in 5 generations
- SNP mutations in mt-DNA: medium
  - Once in ~35,000 generations per SNP
  - Need ~7000 markers to see mutation in 5 generations
- SNP mutations in chromosomal DNA: very slow
  - Once in ~50,000,000 generations (!!!) per SNP
  - With 3 billion bases $\rightarrow$ 60 SNP mutations/generation

Probability Example

- Assume
  - Two sons of a common father
  - Testing of 50 markers with 1-in-500 mutation rate
- Probabilities for each son
  - Y-DNA same as father: 90%
  - One new mutation: 10%
- Probabilities of matches between two sons
  - Both same as father: 90% x 90% = 81%
  - One-marker difference: 2 x 90% x 10% = 18%
  - Two-marker difference: 10% x 10% = 1%

Example Result



- Y-DNA tests on me and a first cousin once-removed
- Five generational steps
- On average: 1 mutation
- Actually: 3 mutations

67 MARKERS - 10 MATCHES

Genetic Distance	Name		Most Distant Ancestor
3	Mr. Chaim Sag	Y-DNA67	Shmuel Yitzhak SAG, b. 1817 and d. 1894
4	Armand Charles Sachs	Y-DNA67 FF	Dr. Henry Sachs
5	Alex Wolf	Y-DNA67 FF	

Autosomal DNA

- Much more complex issue
- Defer until later in talk

Outline

- Introduction
- Basic biology background
- ➔ How DNA can inform genealogy
- Interpreting results and examples
- How testing is done

How DNA Can Inform Genealogy

- Poor Use: can *suggest* a previously unknown relationship to other individuals tested
- Problem: useless matches
  - The more people get tested, the more chance of false matches
  - Many matches are to common ancestors too far back in time to be of genealogical interest

How DNA Can Inform Genealogy

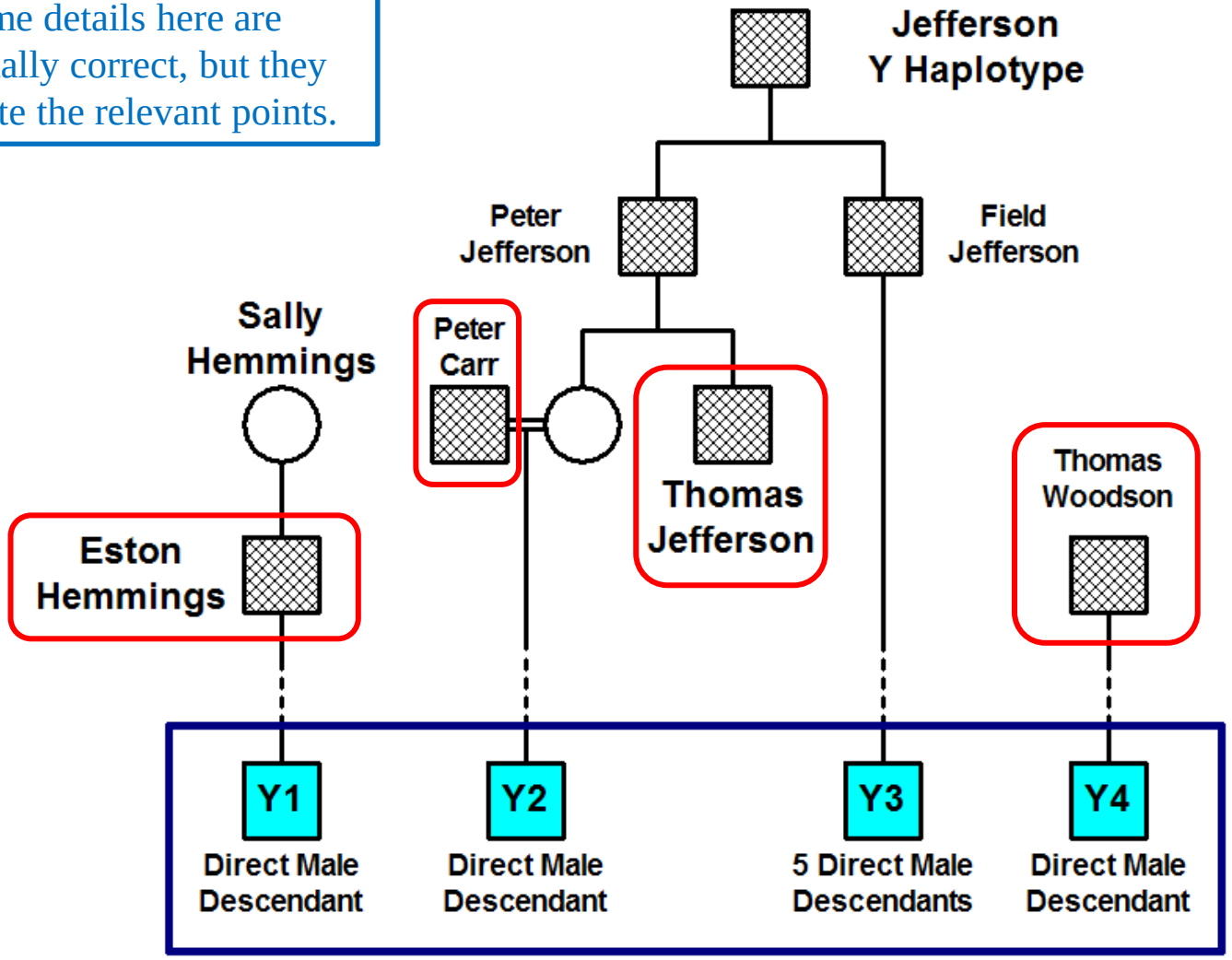
- Good Use: can more-or-less *conclusively* prove a (biological) relationship on a genealogical time scale in a direct-male or direct-female line where a relationship is suspected
- Can less definitely disprove a relationship
- Autosomal test extends use to non-direct lines but with higher uncertainties

Example: Jefferson and Hemmings

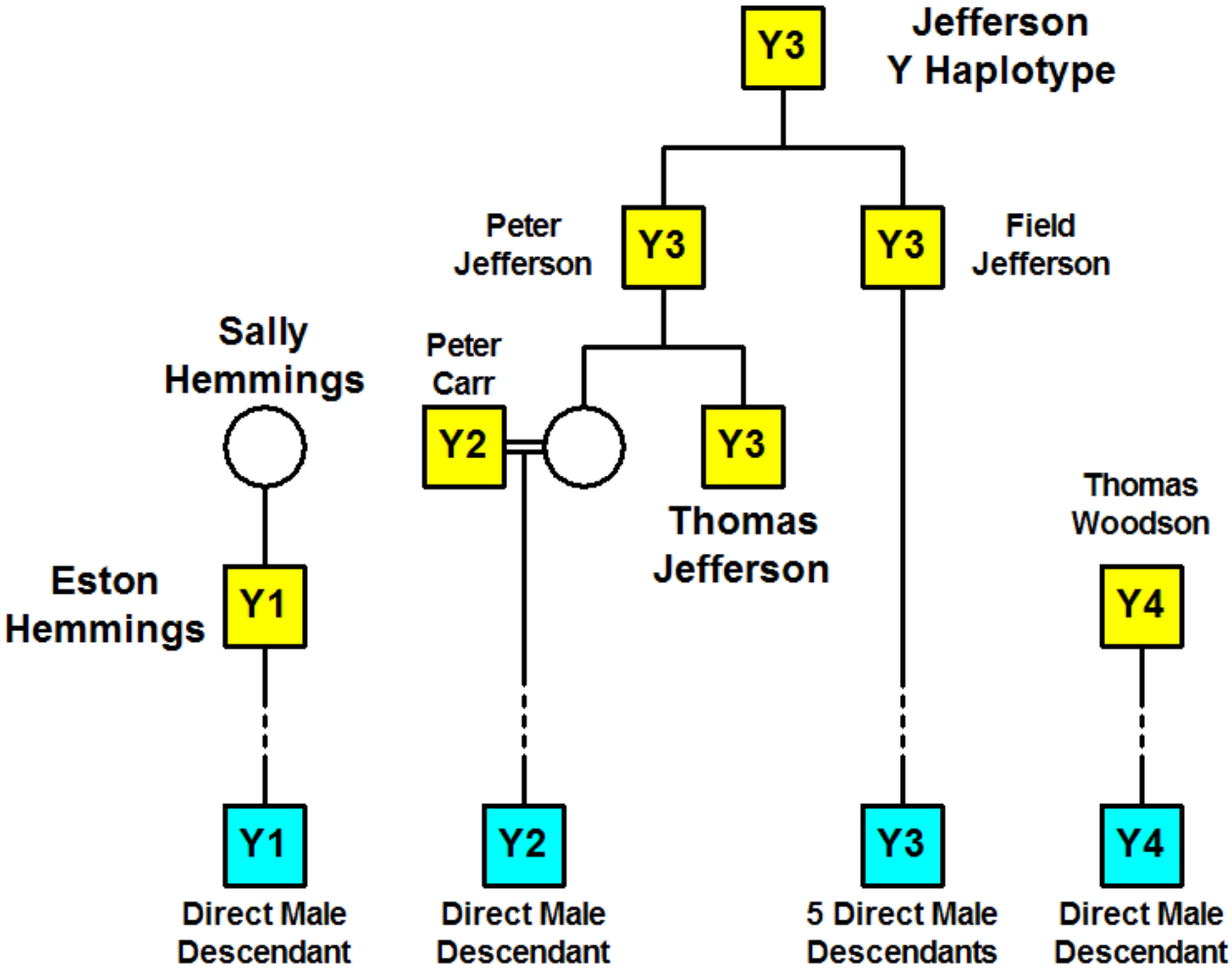
1. Long-time rumors (supported by some historical evidence) asserted that Thomas Jefferson fathered children by his house slave Sally Hemmings, including her youngest son, Eston/Easton Hemmings.
2. Others claimed that Peter Carr, the husband of Jefferson's sister, was the father of Hemmings's children.
3. Thomas Woodson made the claim that his father was Thomas Jefferson.

DNA That We Can Measure

Some details here are not totally correct, but they illustrate the relevant points.



DNA That We Can Infer

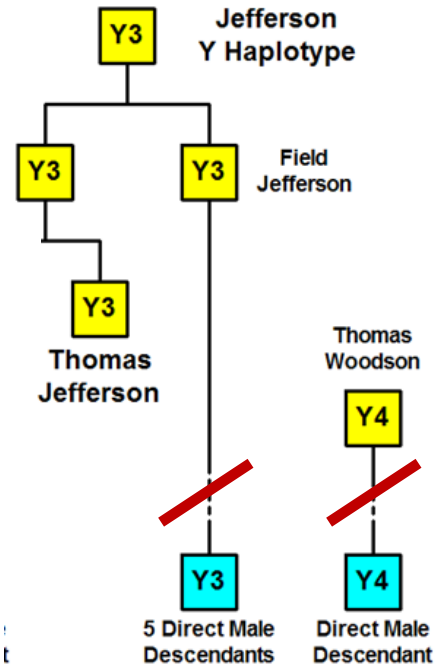


Results and Conclusions

- Y4 (Woodson) ≠ Y3 (Jefferson)
  - Conclusion: Woodson was not Jefferson's son

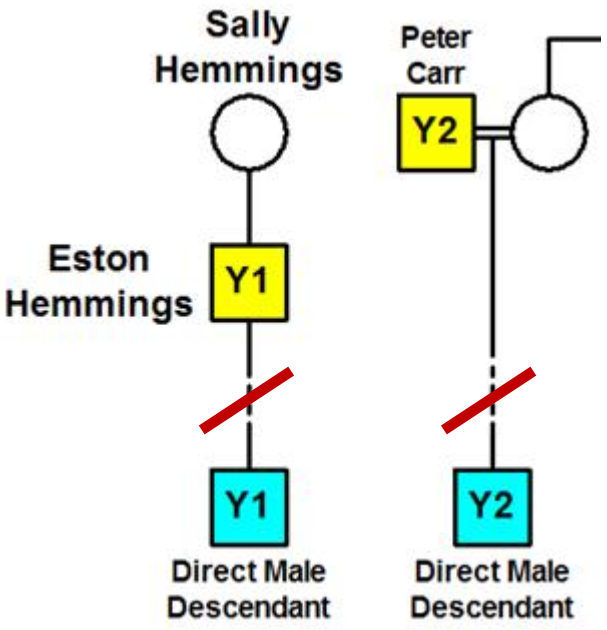
Results and Conclusions

- Y4 (Woodson) ≠ Y3 (Jefferson)
  - Conclusion: Woodson was not Jefferson's son
  - **But wait!** What about a "non-paternity" event in one of those lines?



Results and Conclusions

- Y2 (Carr) ≠ Y1 (Hemmings)
  - Carr was not the father of Eston Hemmings
  - But, again, what about a non-paternity event?

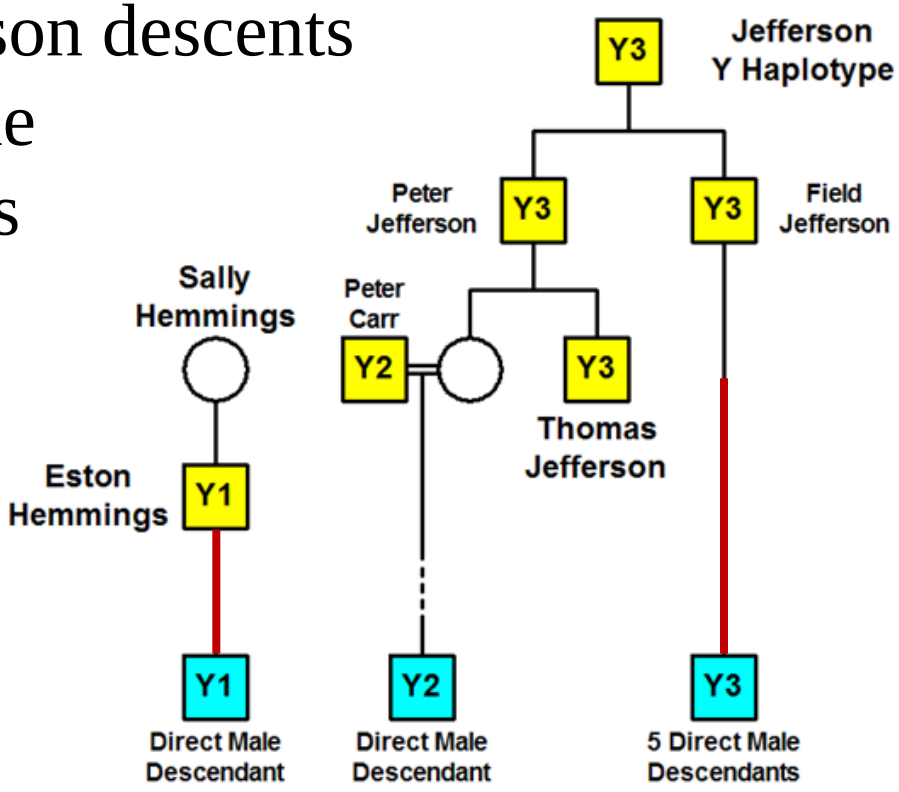


Results and Conclusions

- Y3 (Jefferson) = Y1 (Hemmings)
  - Eston Hemmings was definitely a Jefferson
  - But not necessarily the son of Thomas Jefferson (could be any male Jefferson, since they share the Y chromosome)

Results and Conclusions

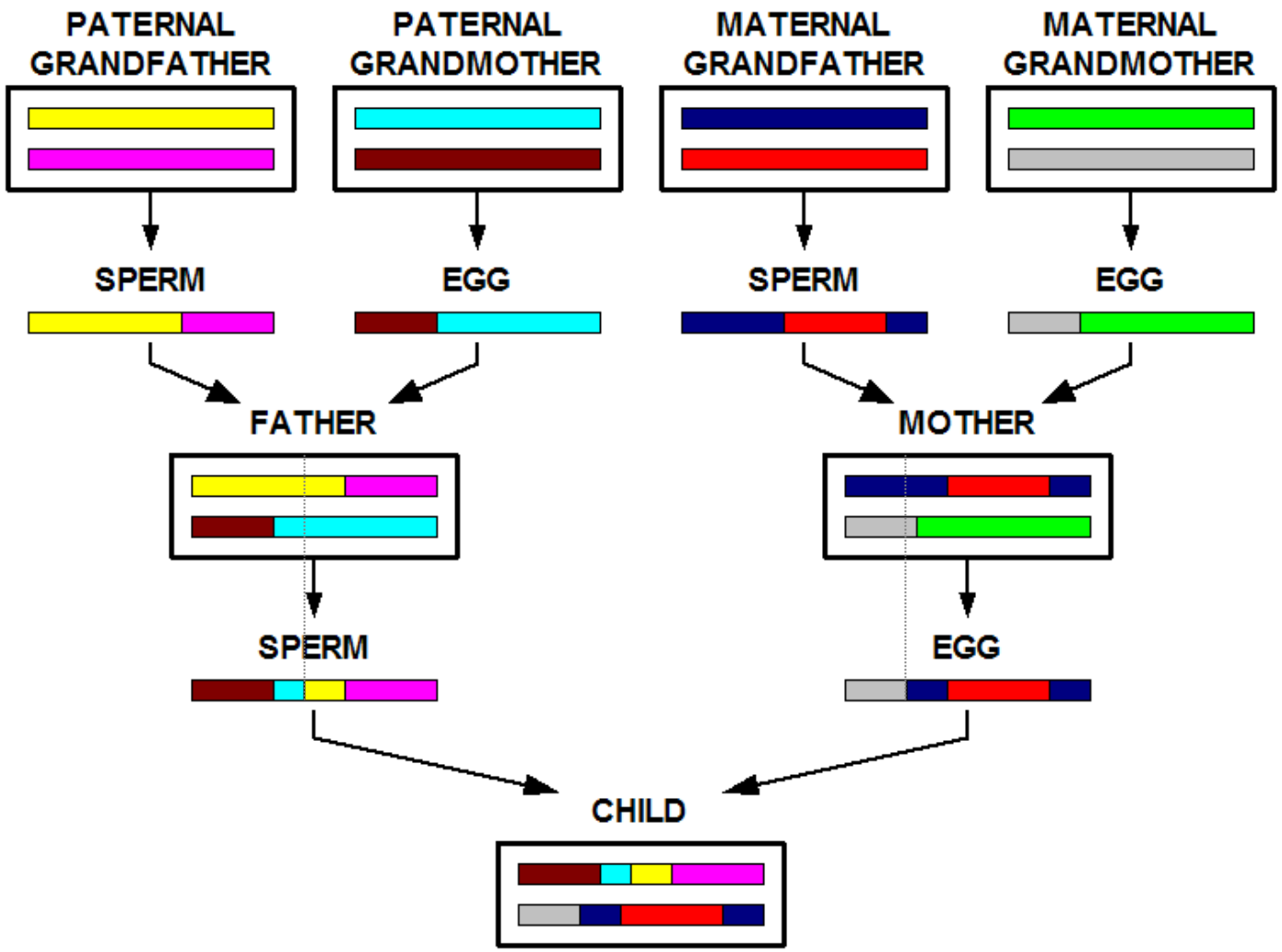
- Y3 (Jefferson) = Y1 (Hemmings)
  - **Bonus**: this confirms the absence of non-paternity events in the Hemmings-Jefferson descents and *strengthens* the earlier conclusions



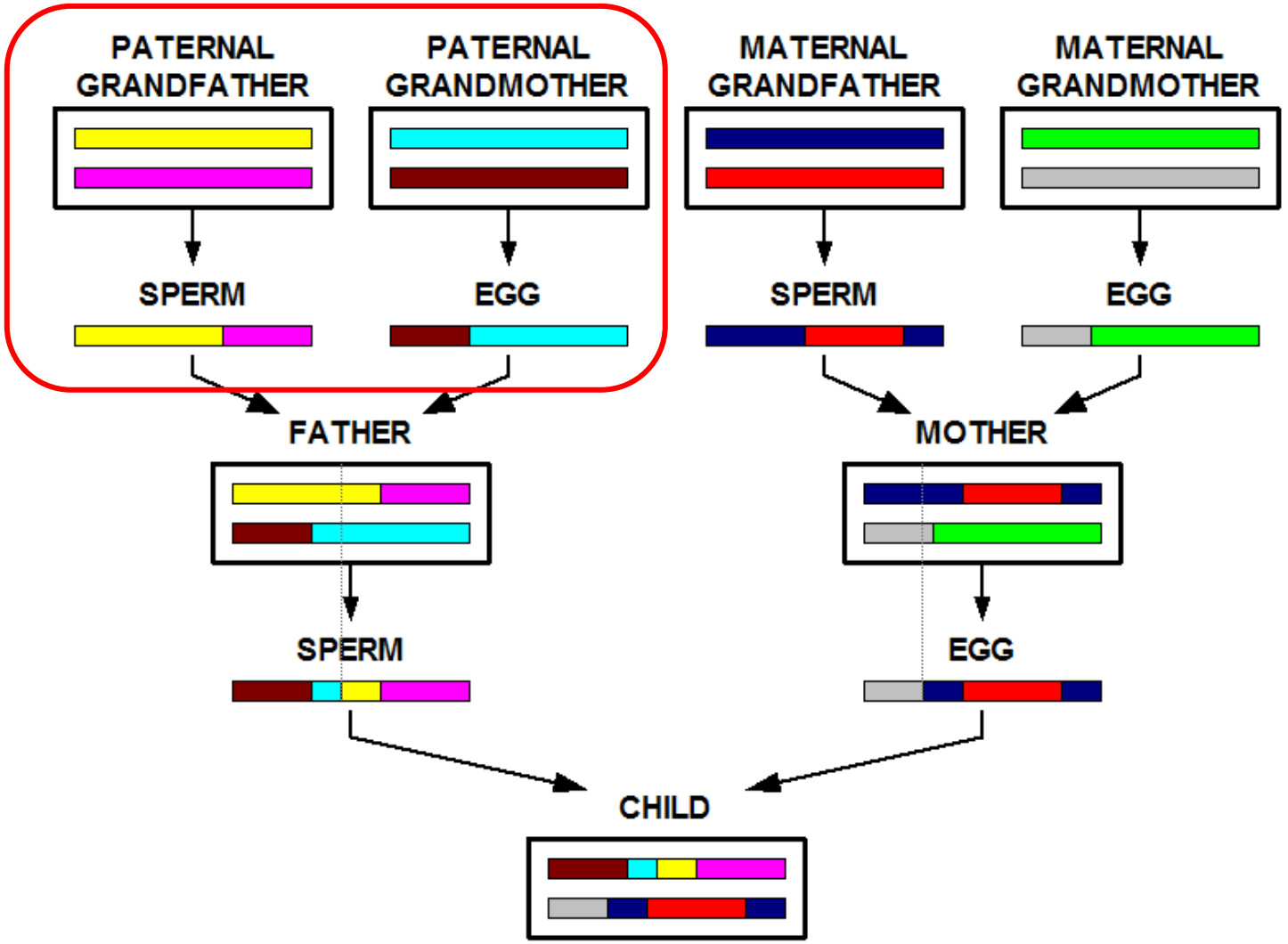
Autosomal DNA

(Stand up, stretch, and take a deep breath)

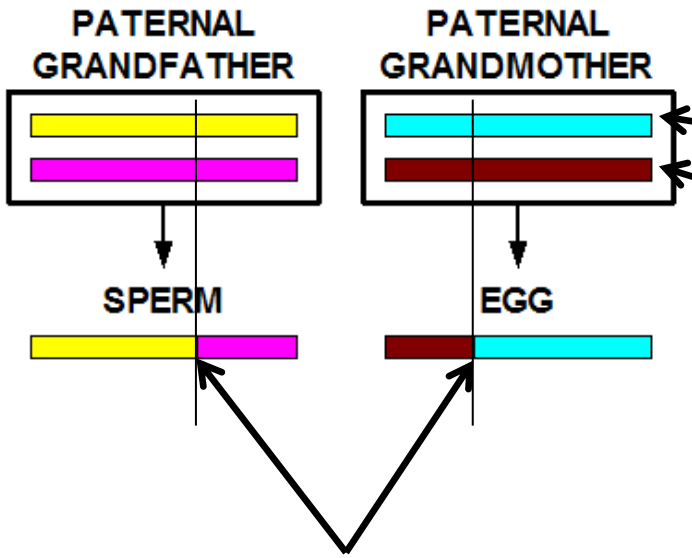
Autosomal Chromosomes



Autosomal Chromosomes



Autosomal Chromosomes



DNA from great-grandparents

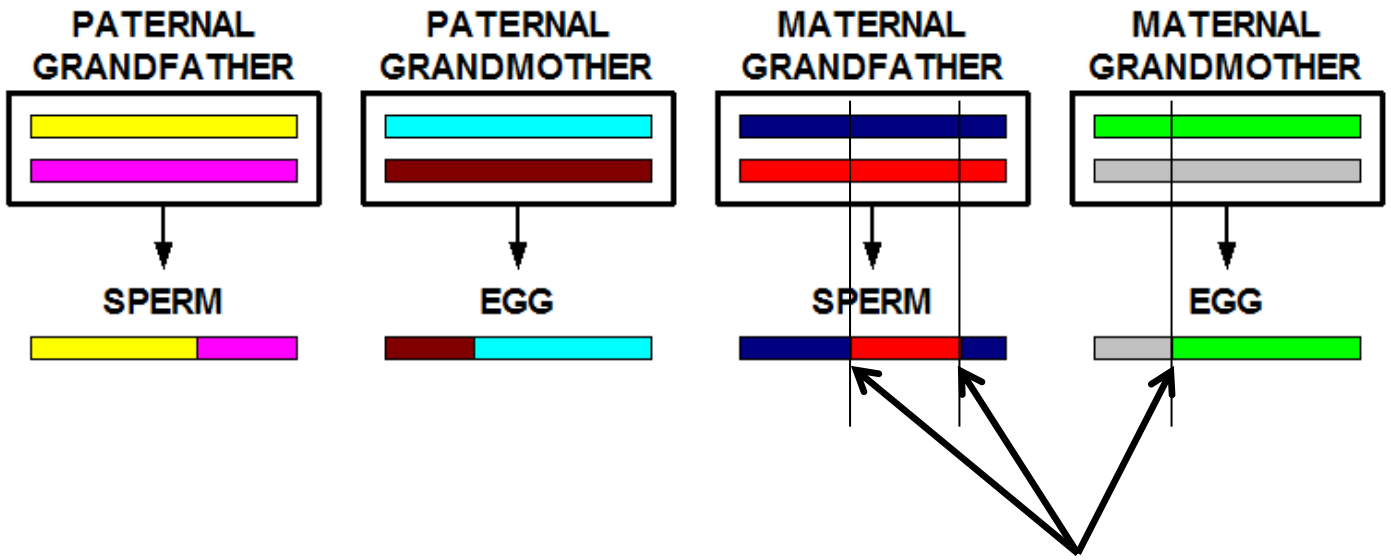
from father's sperm

from mother's egg

random exchange points

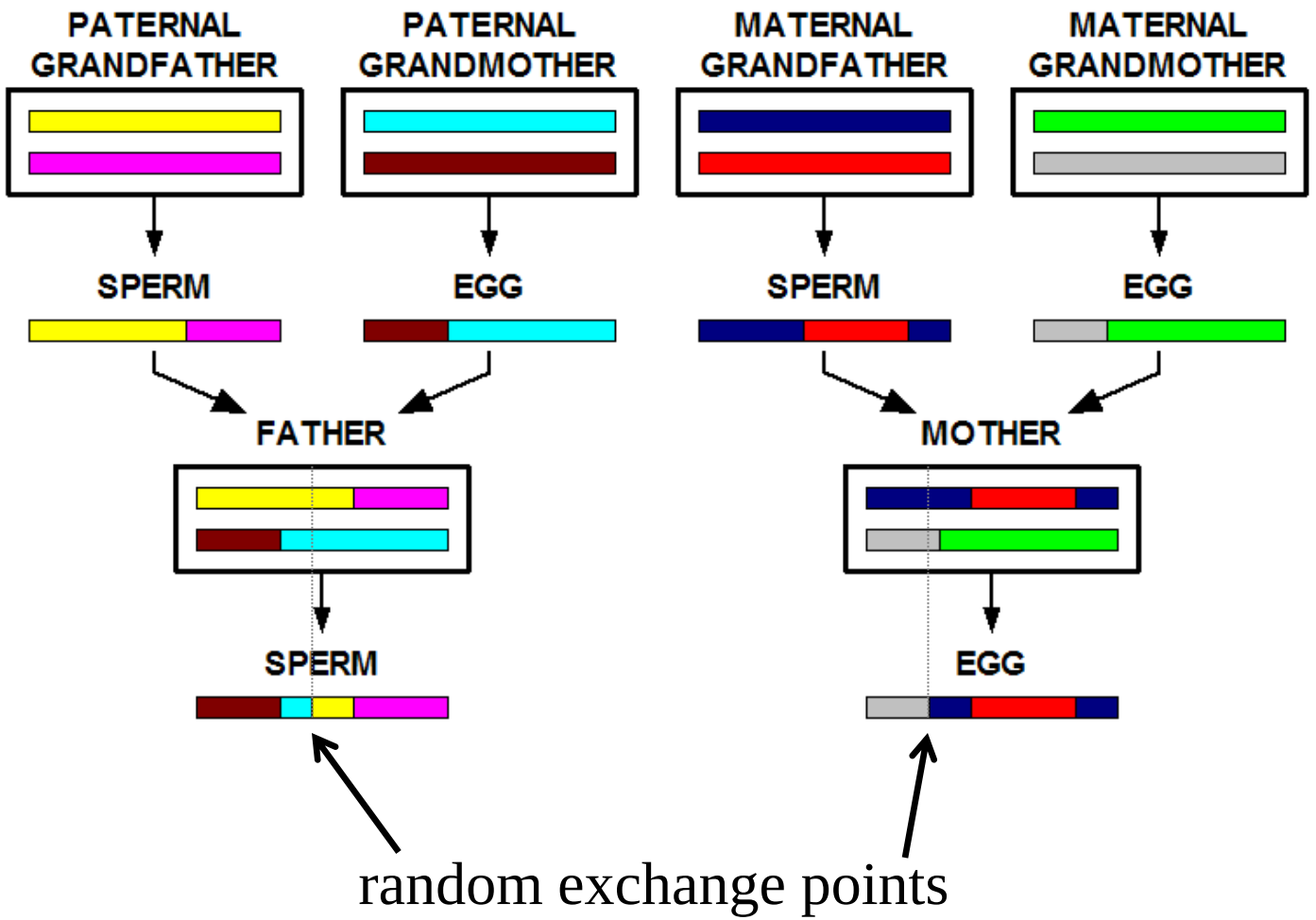
(each sperm and egg is different)

Autosomal Chromosomes

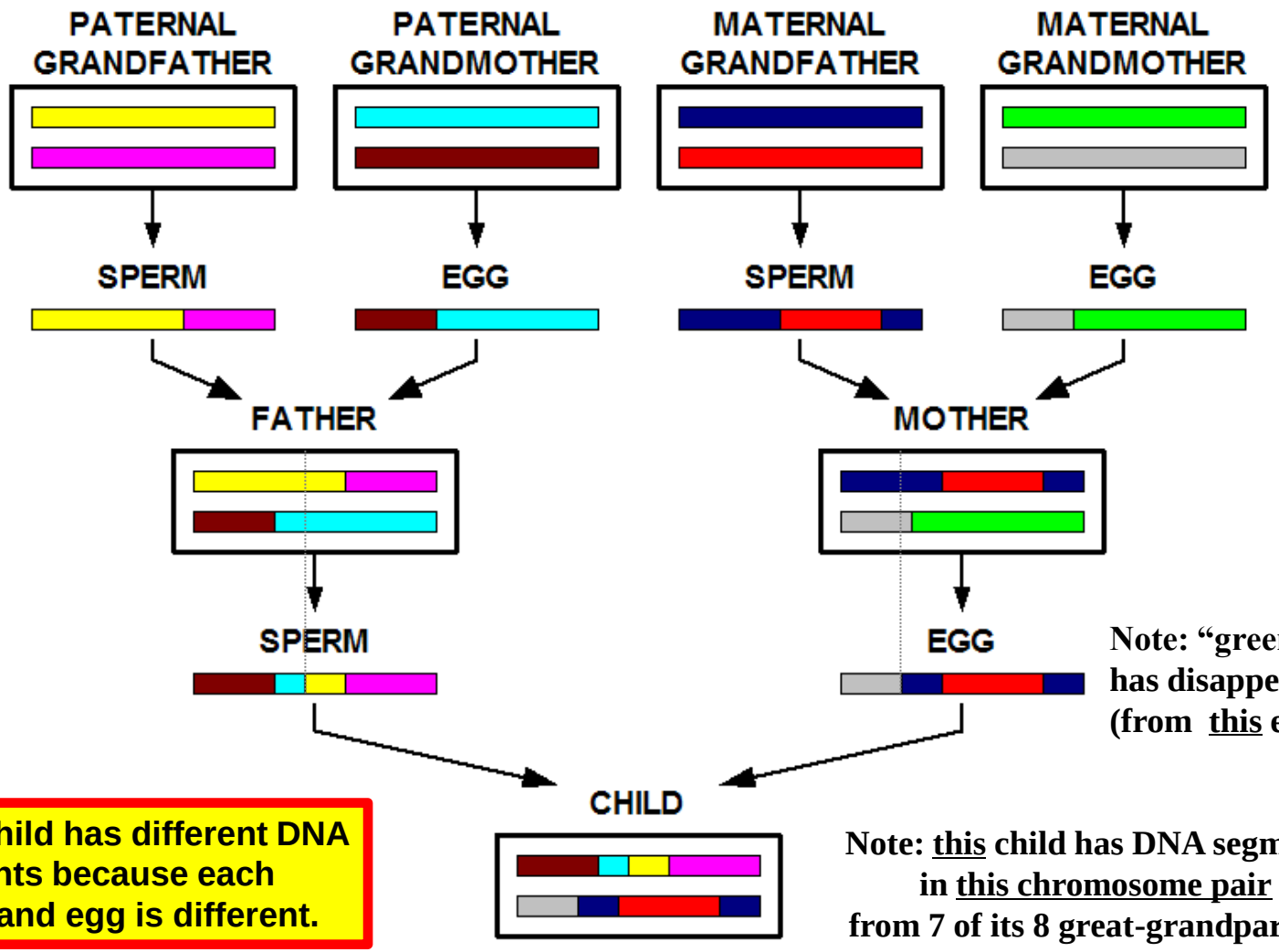


random exchange points

Autosomal Chromosomes



Autosomal Chromosomes

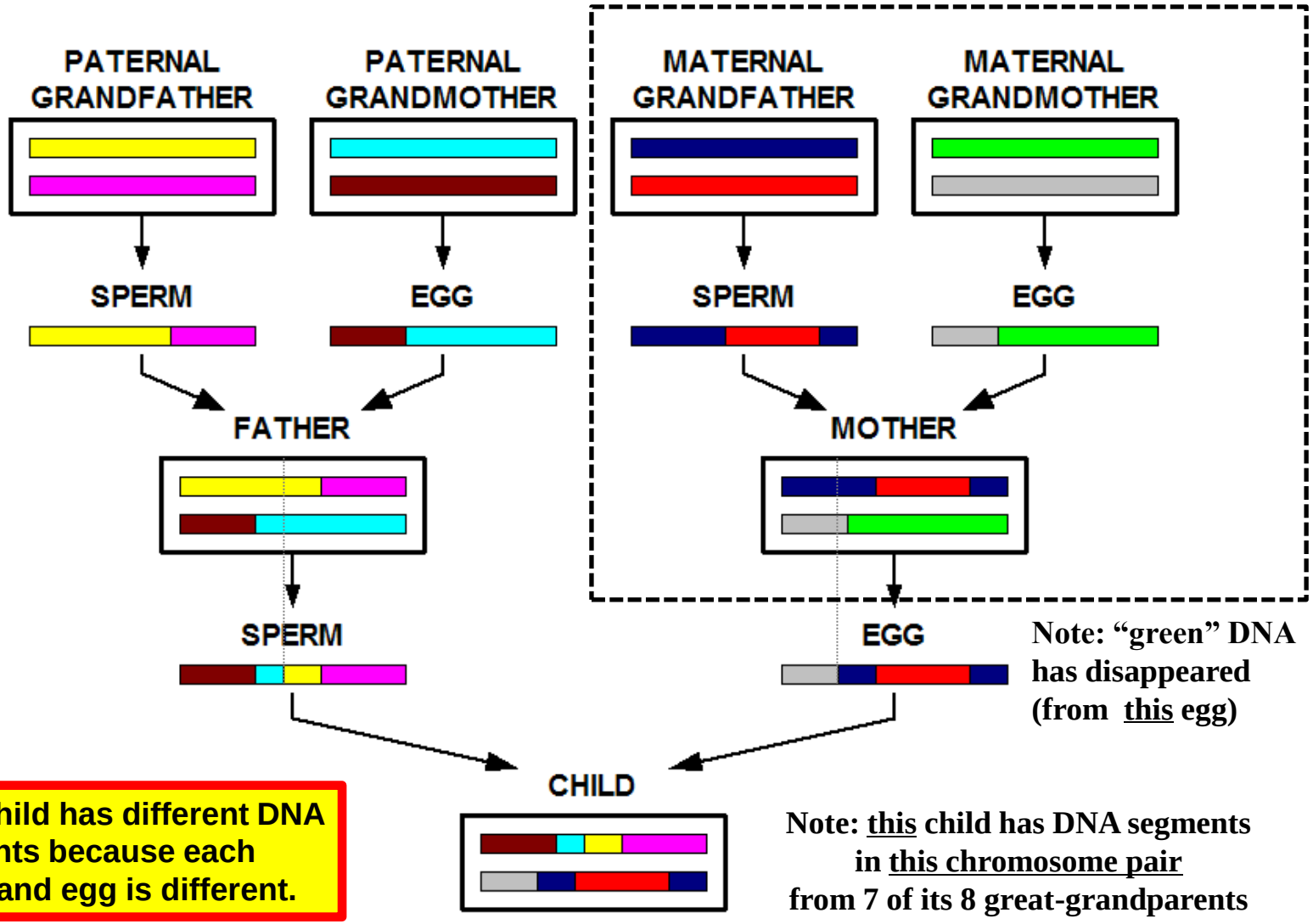


Note: "green" DNA has disappeared (from this egg)

Each child has different DNA segments because each sperm and egg is different.

Note: this child has DNA segments in this chromosome pair from 7 of its 8 great-grandparents

Chromosomes in First Cousins

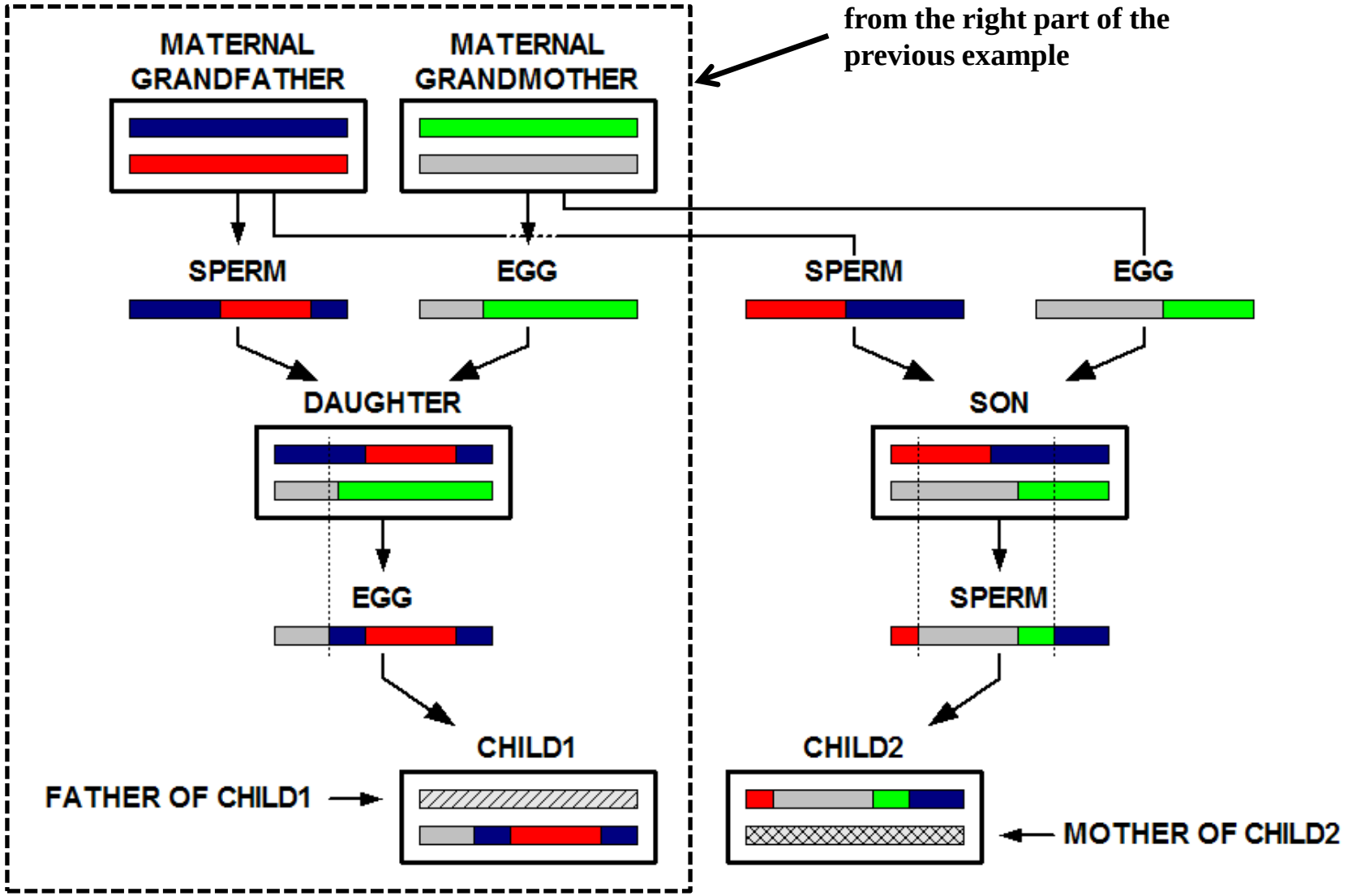


Note: "green" DNA has disappeared (from this egg)

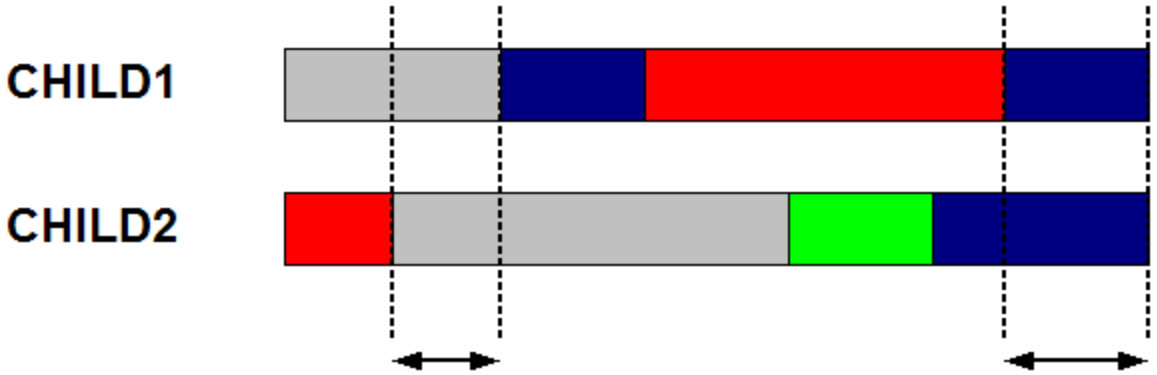
Each child has different DNA segments because each sperm and egg is different.

Note: this child has DNA segments in this chromosome pair from 7 of its 8 great-grandparents

Chromosomes in First Cousins



Common DNA Segments For This Pair of First Cousins



Critical Fact #3 for Genetic Genealogy

- The DNA of paired (autosomal) chromosomes gets scrambled ("recombined") and generally contains segments from many ancestors
- Recombination is variable and unpredictable
- Therefore, it is much more difficult to analyze
- Requires measuring a very large number of markers in order to identify matching segments

Critical Fact #3 for Genetic Genealogy

- Requires measuring a very large number of markers in order to identify matching segments
  - DNA is 99.9% identical in any two individuals
  - The simple example assumed that the DNA of each of the 8 great-grandparents was entirely unique (different solid colors for their chromosomes)
  - Huge number of markers is needed to distinguish the DNA from different ancestors?
  - Imagine the complication of significant intra-marriage?

Contrast Between Y-DNA/mt-DNA and Autosomal DNA

- Y-DNA and mt-DNA remain essentially unchanged from parent to offspring
    - Change occurs only as a result of rarely occurring random mutations
- Autosomal DNA changes drastically from parents to each offspring
    - Only random strings of common DNA match between related individuals
    - Testing multiple family members can help sort things out

Outline

- Introduction
- Basic biology background
- How DNA can inform genealogy
- ➔ Interpreting results and examples
- How testing is done

Classification of DNA Test Results

- Haplotype
  - The set of values obtained in a DNA test
  - Applies to STRs and SNPs
- Haplogroup
  - Larger grouping of populations that share an SNP haplotype
  - Relates to population origins and migrations
  - Not really relevant to genealogy

Example of Y-DNA Test Result

STR MARKER ID

Marker = identifiable location on DNA
Allele = value of marker
Haplotype = set of alleles for a test

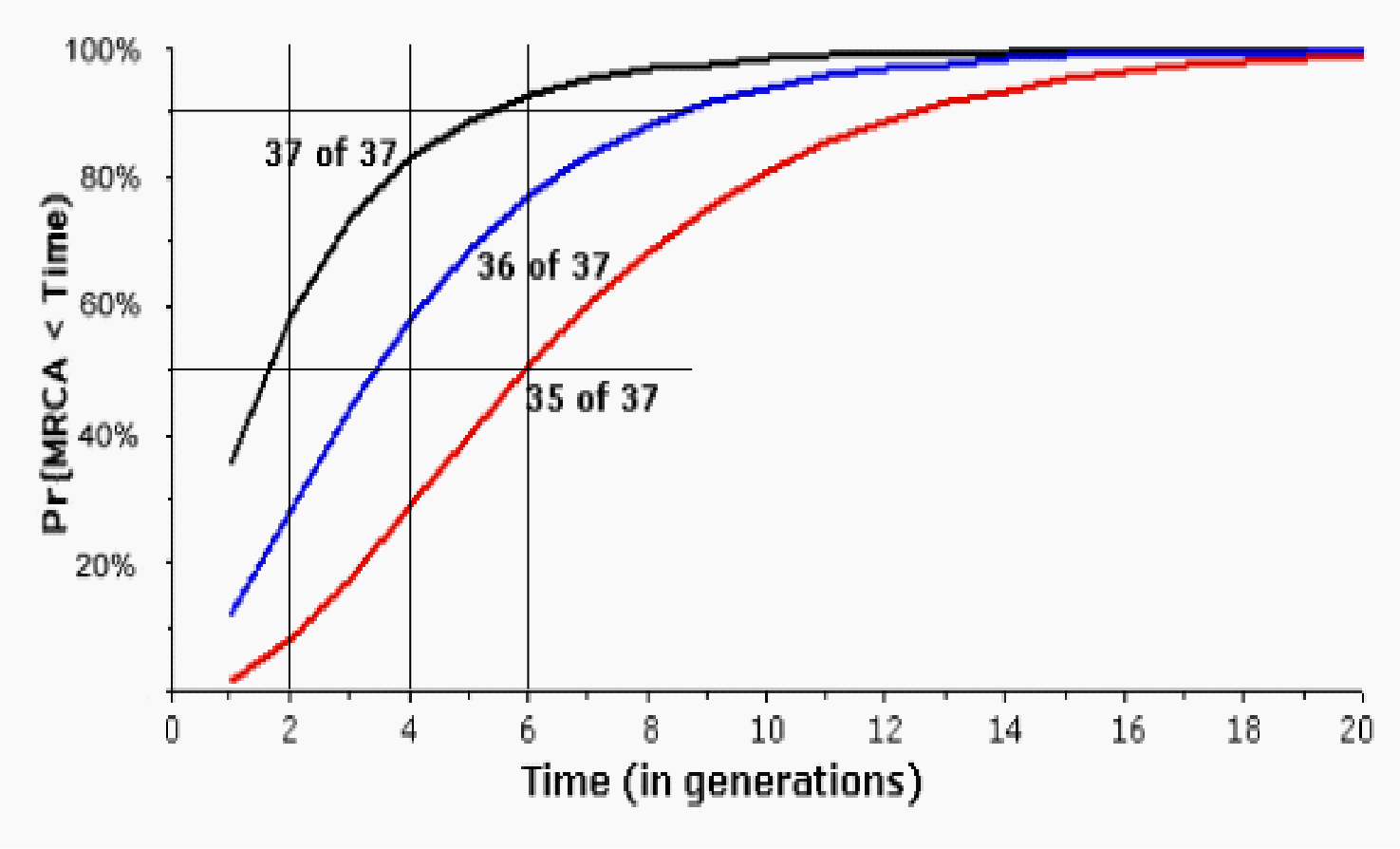
NUMBER OF REPEATS

PANEL 1 (1-12)												
Locus	1	2	3	4	5	6	7	8	9	10	11	12
DYS#	393	390	19*	391	385a	385b	426	388	439	389-1	392	389-2
Alleles	13	25	16	10	11	14	12	12	11	13	11	30

PANEL 2 (13-25)													
Locus	13	14	15	16	17	18	19	20	21	22	23	24	25
DYS#	458	459a	459b	455	454	447	437	448	449	464a**	464b**	464c**	464d**
Alleles	14	9	10	11	11	24	14	20	30	12	13	15	15

PANEL 3 (26-37)												
Locus	26	27	28	29	30	31	32	33	34	35	36	37
DYS#	460	GATA H4	YCA II a	YCA II b	456	607	576	570	CDY a	CDY b	442	438
Alleles	11	11	19	22	14	16	20	21	35	37	14	12

Probabilities Regarding Most Recent Common Ancestor



Y-DNA Match Report

contact

test type(s) and level(s)

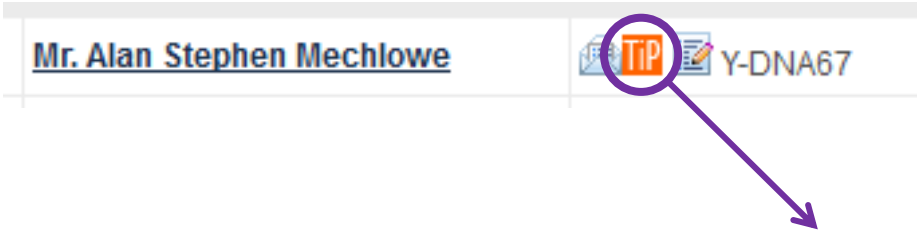
67 MARKERS - 84 MATCHES

Page: 1 2 3 4 of 4

Genetic Distance	Name		Most Distant Ancestor	Y-DNA Haplogroup	Terminal SNP	Match Date
0	Mr. Alan Stephen Mechlowe	Y-DNA67		E-M34		4/2/2013
1	Barry Helfand	Y-DNA111 FF	Zalman Gelfand d prior to 1905	E-M34	M34	6/11/2013
1	Leonid Neiman	Y-DNA67	Benjamin Neiman, Belarus, Bobruysk,1796	E-M34		12/8/2008
1	Ben Kalmuk	Y-DNA67		E-M34		9/28/2007
1	Jerry Chiprin	Y-DNA111	Gedalia Chiprin, b.ca 1830 Kiev Ukraine	E-M34	M34	9/28/2007
2	Dr. Jeffrey Abrams	Y-DNA67 FF	Abramovich from Kovna Gubernia, Lithuania	E-M34		2/15/2014
2	Aleksandr Georgievich Pogrebeckij	Y-DNA67	Republic Dagestan, Odessa	E-M34		4/14/2010
3	Philip Kalson	Y-DNA67	Yonas (Yona) KIELSON, b. circa 1798 d. 1852	E-M34		9/6/2012
3	Mr. Scadron		Yechiel Mechiel Shkadron	E-M34		8/23/2011
3	Jonathan Fluger	Y-DNA67		E-M34		11/1/2010
3	Mark Lewis Belafsky M.D.	Y-DNA111 FF	Uri Belyavsky 1790-?	E-M34	M34	9/28/2007
4	Steve Baker	Y-DNA67 FF	EVERETT PRICE BAKER	E-M34		12/30/2013
4	Mr. Warren Weinzoff	Y-DNA67 FF	Abraham Weinzoff b.1850	E-M34		1/29/2013
4	Dr. Ronald David Doctor	Y-DNA67 FF	Aharon Vurer, before 1734	E-M34		9/29/2008

display family tree

FTDNATiP (Time Predictor)



Detailed calculation of probability of a common ancestor

COMPARISON CHART	
Generations	Percentage
4	89.79%
8	98.96%
12	99.89%
16	99.99%
20	100.00%
24	100.00%

FTDNATiP (Time Predictor)

COMPARISON CHART		
Generations	Percentage	Percentage
4	89.79%	43.47%
8	98.96%	94.23%
12	99.89%	99.41%
16	99.99%	99.94%
20	100.00%	99.99%
24	100.00%	100.00%

From paper trail:
pretty sure NO common ancestor within 4 generations

Example #1: Y-DNA

- A member of our society has Frankel ancestors from a town in Europe
- He found another Frankel family from the same town but could not establish a relationship through documents
- Each family found a living direct male descendant
- Each took a Y-DNA test, and they matched
- Conclusion: the families are related, probably one generation farther back

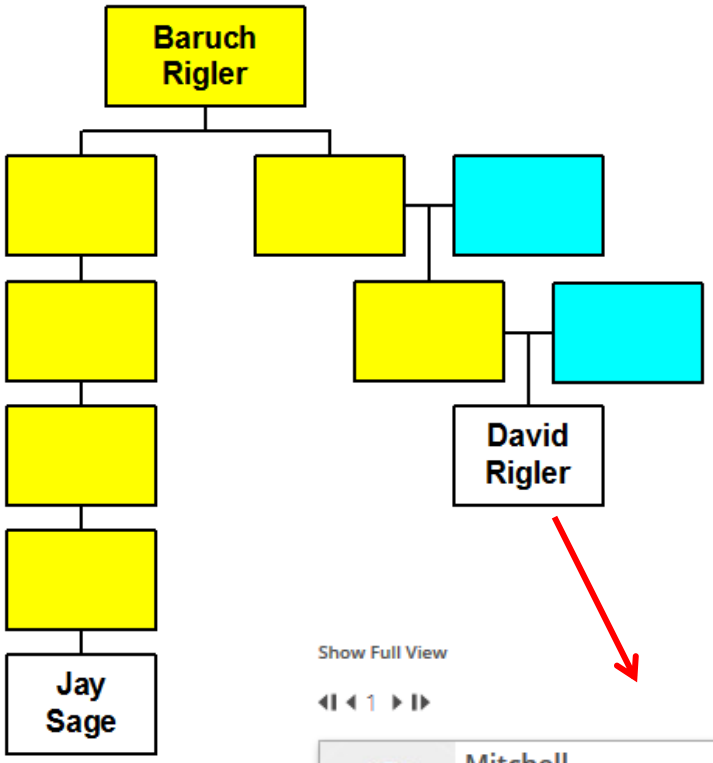
Example #2: Y-DNA

- I made contact with a man with the surname Geldman in England
- Family from “Poland”, not Romania
- But his tree had all the given names in mine
- He and my cousin Steve Geldman took Y-DNA tests, and they did not match
- Conclusion: either the families are not related, or there was a non-paternity event

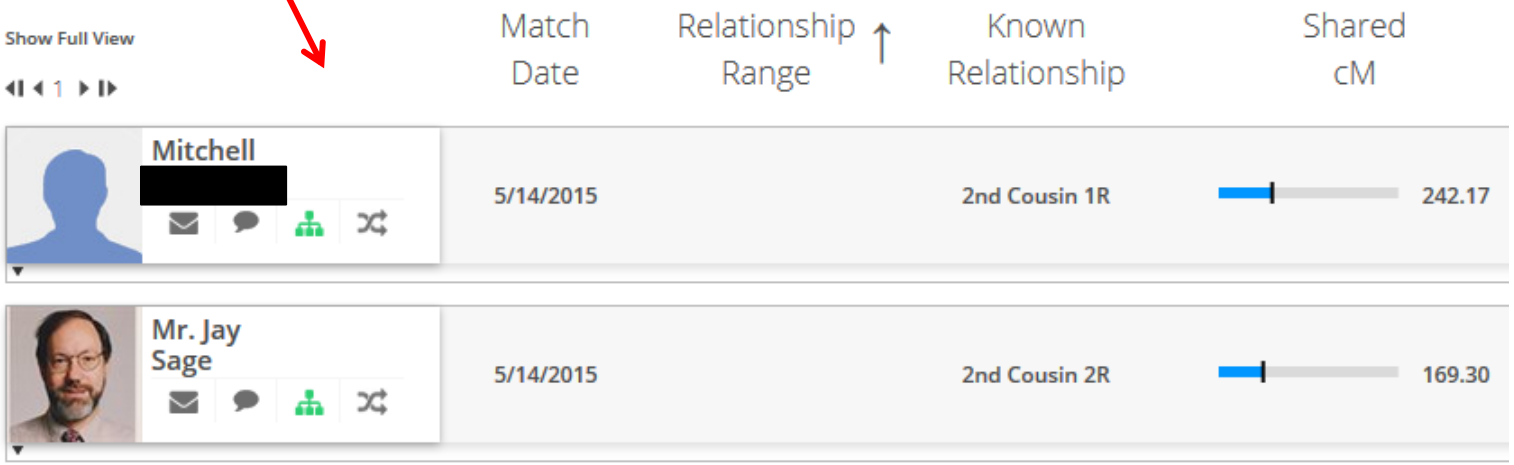
Example #3: Y-DNA

- A friend (Shirley) tested a close male relative
- My wife, Daphnah, later tested the last male relative in one branch of her family
  - Shirley received report of a match!!
- Still later, I tested one of the few male relatives in one branch of my family
  - I noticed Shirley's relative on my list of matches!!!
- Further evidence: they are all Levi!
- We are related, but too far back to ever prove

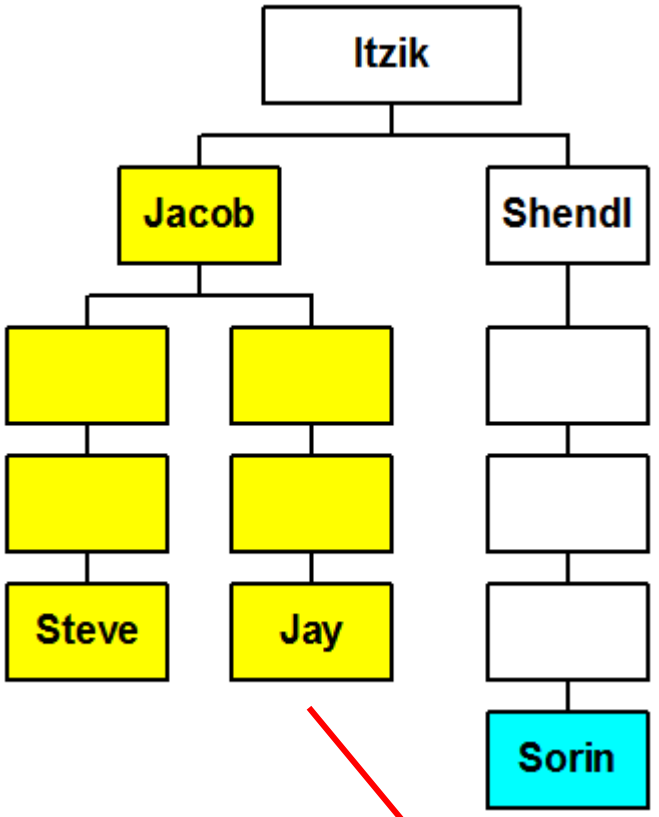
Example #4: Autosomal



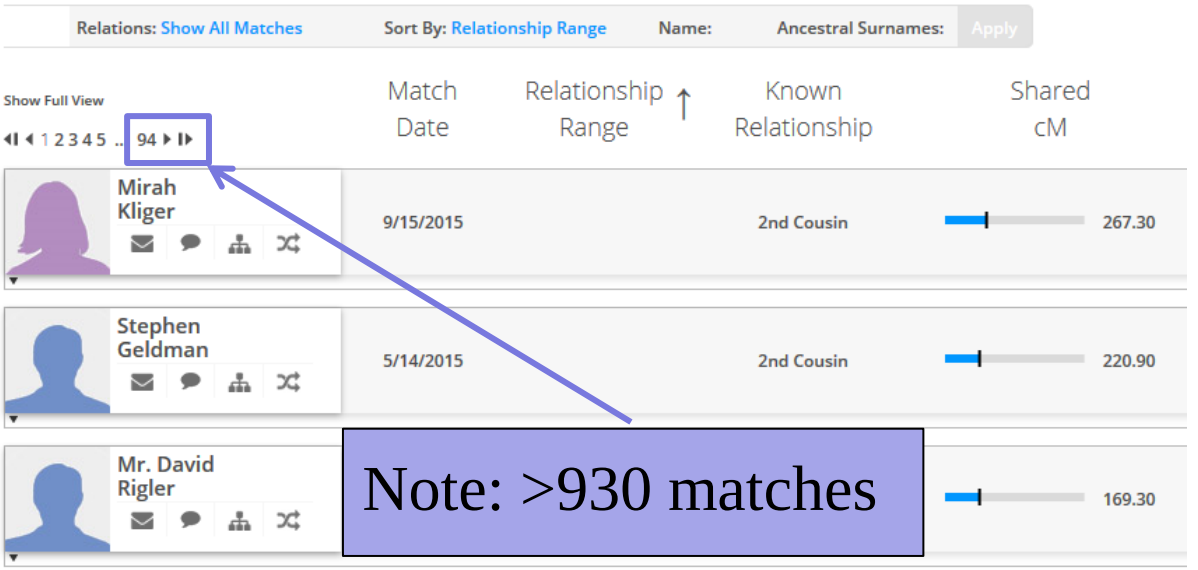
- David closely matches Jay and Mitchell
- But Jay and Mitchell do not match each other at all
- Therefore, Mitchell's relationship is via cyan boxes



Example #5: Autosomal NPE



- Jay and Steve: match (6 steps)
- Sorin: no match (9 steps)
- Conclusion: probably a non-paternity event somewhere between Jacob and Sorin



Outline

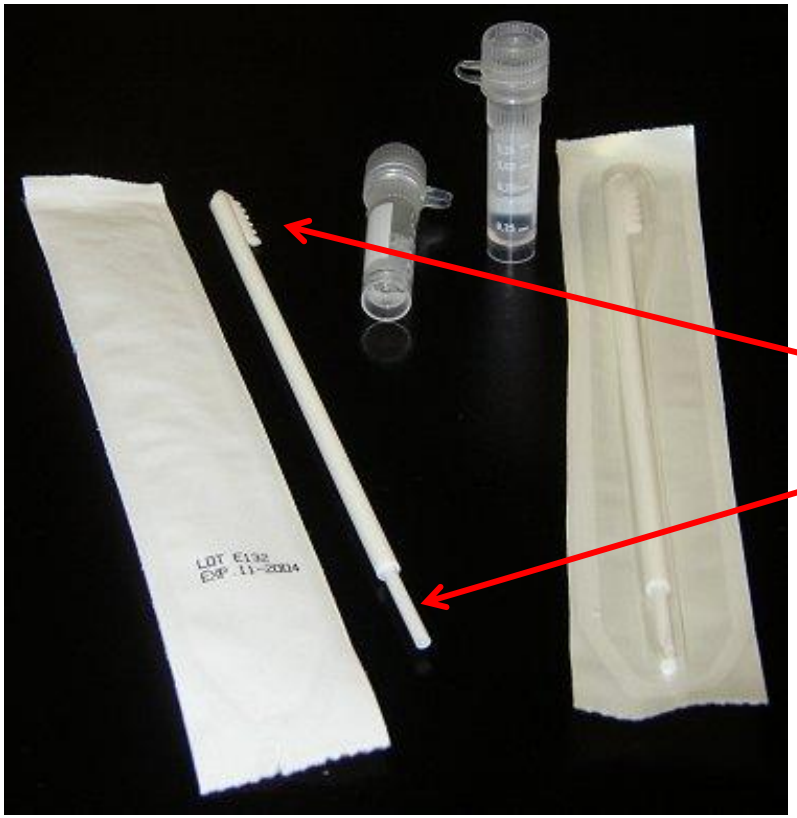
- Introduction
- Basic biology background
- How DNA can inform genealogy
- Interpreting results and examples
- ➔ How testing is done

DNA Testing Companies

- Family Tree DNA (ftdna.com)
- 23andMe (23andMe.com)
- Ancestry (dna.ancestry.com)

For a long list of companies, see
www.isogg.org/wiki/List_of_DNA_testing_companies
(tinyurl.com/dna-companies)

Test Kit



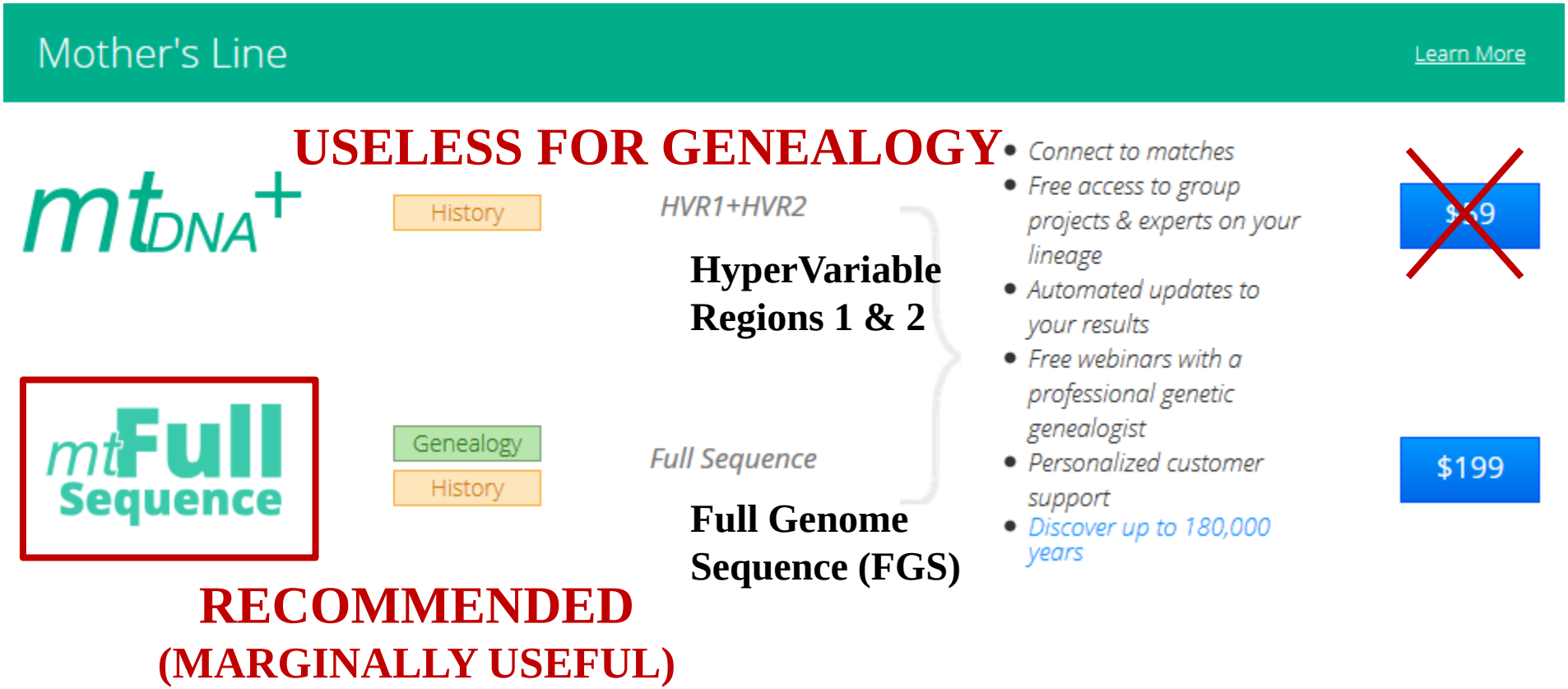
- Order test
  - e.g., Family Tree DNA www.ftdna.com
- Receive kit in mail
- Scrape inside of cheek
- Expel scraper tips into vials of liquid
- Mail vials to testing lab
- Results posted on website in a few weeks

Types of Tests and Costs (FTDNA)

Y-DNA STR

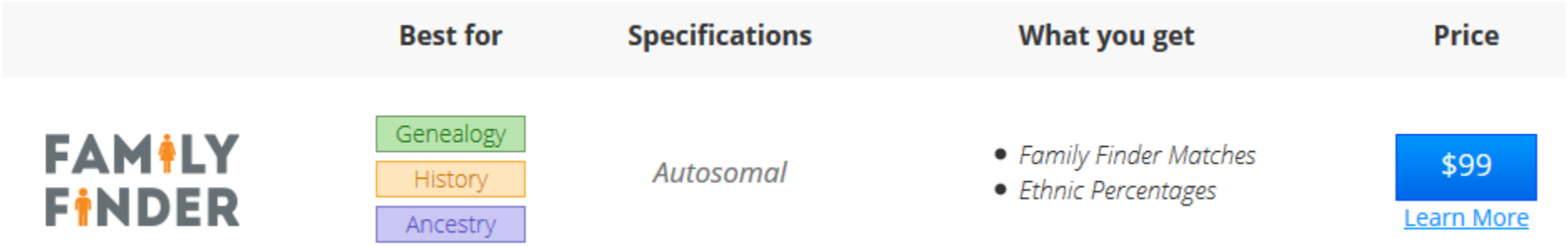
Types of Tests and Costs (FTDNA)

mt-DNA SNP



Types of Tests and Costs (FTDNA)

Autosomal DNA



	Best for	Specifications	What you get	Price
FAMILY FINDER	Genealogy History Ancestry	*Autosomal*	• *Family Finder Matches* • *Ethnic Percentages*	$99 Learn More

Recommended only for fun or when you have a reason to believe that there is a relationship between two people

Types of Tests and Costs (FTDNA)

- Combinations of tests are available at a small saving in cost
- Tests can be upgraded
  - Test of a given type can be upgraded (e.g., Y-DNA from 37 to 67 markers) at a differential cost
  - Other test types can be ordered on existing DNA samples (e.g., from Y-DNA to mt-DNA)
  - Inexpensive test good for getting a sample on file for later use
- There are frequent sales

Warning: Non-Paternity Events!

Do not do DNA testing unless you are prepared for the possibility of learning something uncomfortable.

Summary

- STR testing of Y chromosome is the most useful genetic test for genealogy
  - Works only on direct male lines
  - Can fairly definitively prove a suspected biological relationship
  - Can likely rule out a biological relationship
- SNP testing of mt-DNA is less useful
  - Works only on direct female lines (but male siblings can be used to obtain samples)
  - Only sequencing of the complete genome gives information relevant to relationships within a genealogical time frame
  - Can fairly definitively prove a suspected biological relationship

Summary

- Autosomal testing
  - Allows matching across mixed male/female lines
  - Now available at a very reasonable cost
  - Interpretation open to much doubt (I consider it largely useless)
  - Proper analysis is complex and tricky
- In addition to providing genealogical information, DNA testing can indicate migration patterns and give general clues to ethnicities
- DNA testing in no way supplants traditional genealogical research