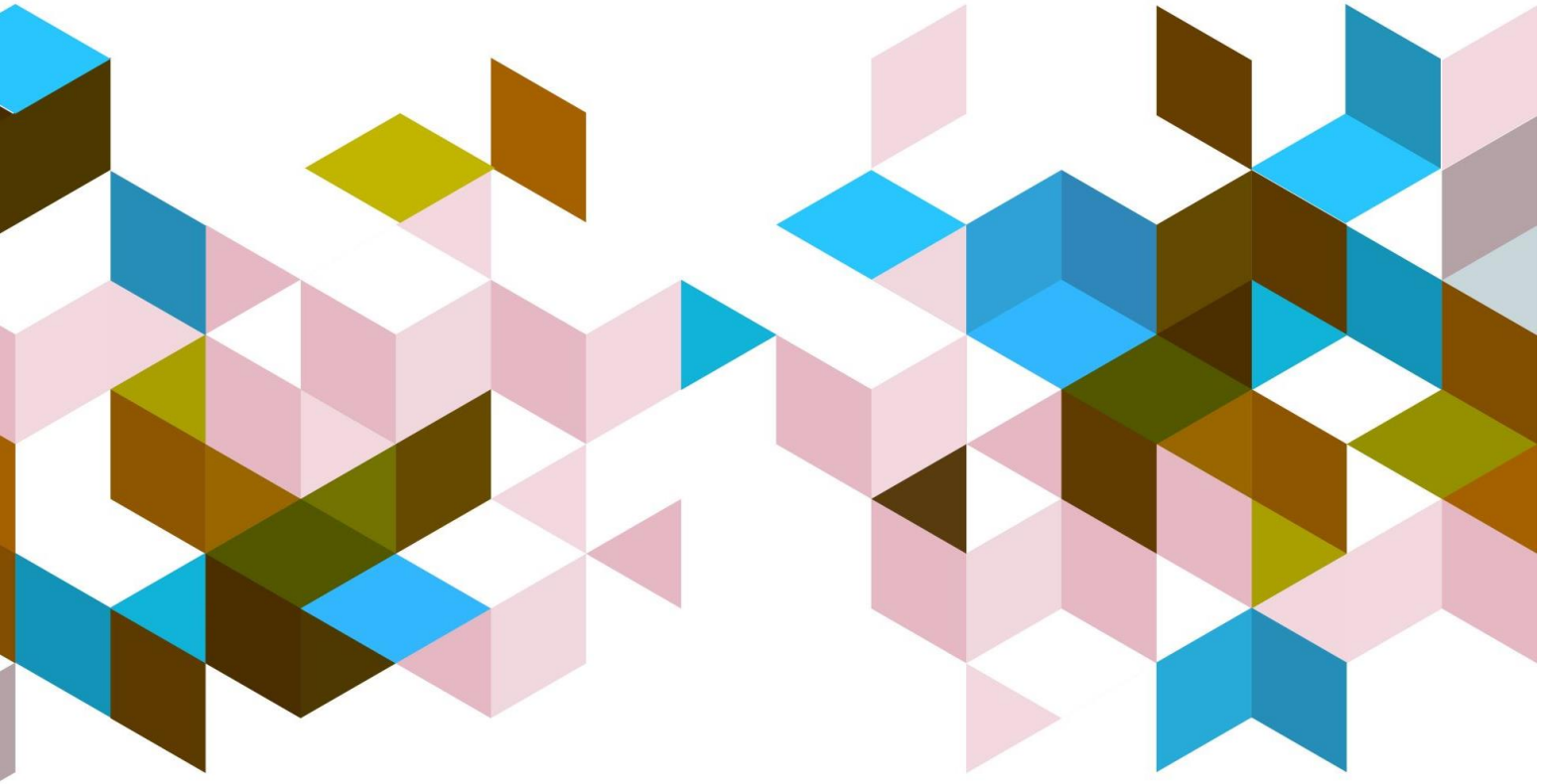


Forensic Investigative Genetic Genealogy

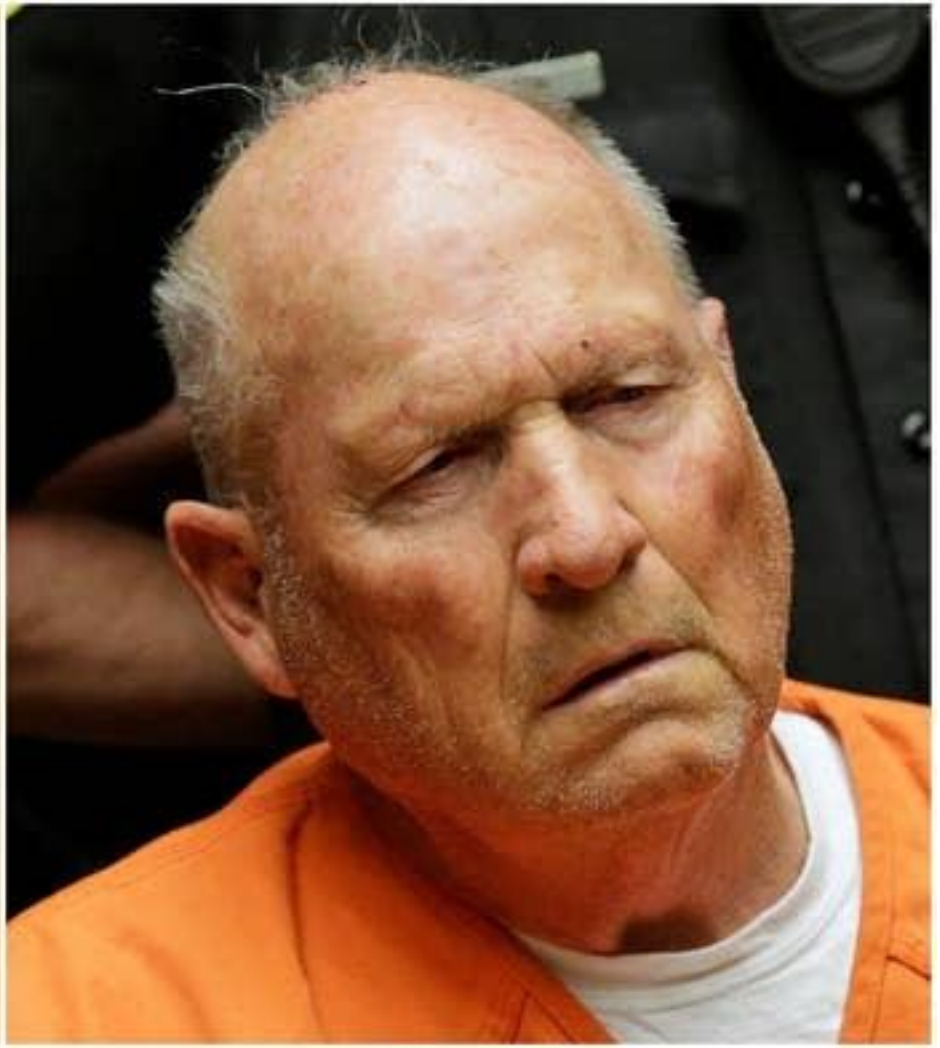
Dr. Wendy Watson

University of North Texas

What Is Forensic Investigative Genetic Genealogy?

- Using tools for doing genealogical work (both genetic and traditional) to identify the unknown contributor of a DNA sample
- Applications
  - Finding heirs and biological parents (often just considered “genetic genealogy”
  - Identifying John and Jane Does
  - Identifying criminal suspects who left DNA at a crime scene
  - Identifying victims of mass casualty events

Golden State Killer: Joseph D’Angelo

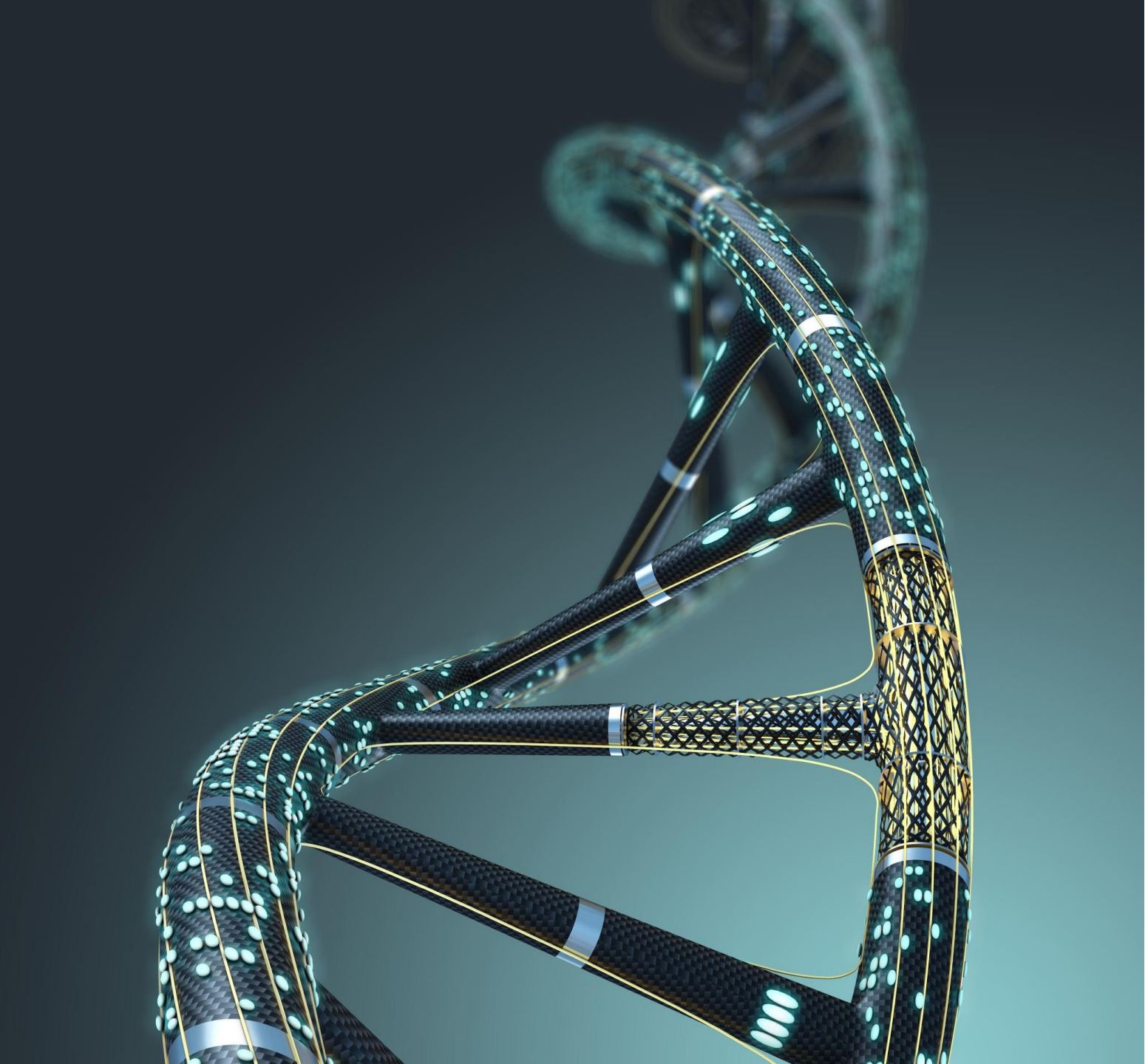


DONALD
HADLAND
JR.

RAMAPO
COLLEGE

Nogales
John Doe
2002

Tulsa Race Massacre: CL Daniel Identified



INTRO TO DNA

Blueprint for Biology

- Double helix structure
  - The “spiral” of the double helix is composed of deoxyribose and phosphate
  - The connection between the helices are like rungs on a ladder
  - Each rung is made up of two base pairs:
    - Adenine (A) and Thyamine (T) connected by two hydrogen bonds
    - Guanine (G) and Cytosine (C) connected by three hydrogen bonds
- In replication, the double helix ”unzips” and we can sequence each strand separately
- Two types of DNA
  - Autosomal (in the cell nucleus)
  - Mitochondrial (in the mitochondria – “powerhouse” – of the cell)

Chromosomes

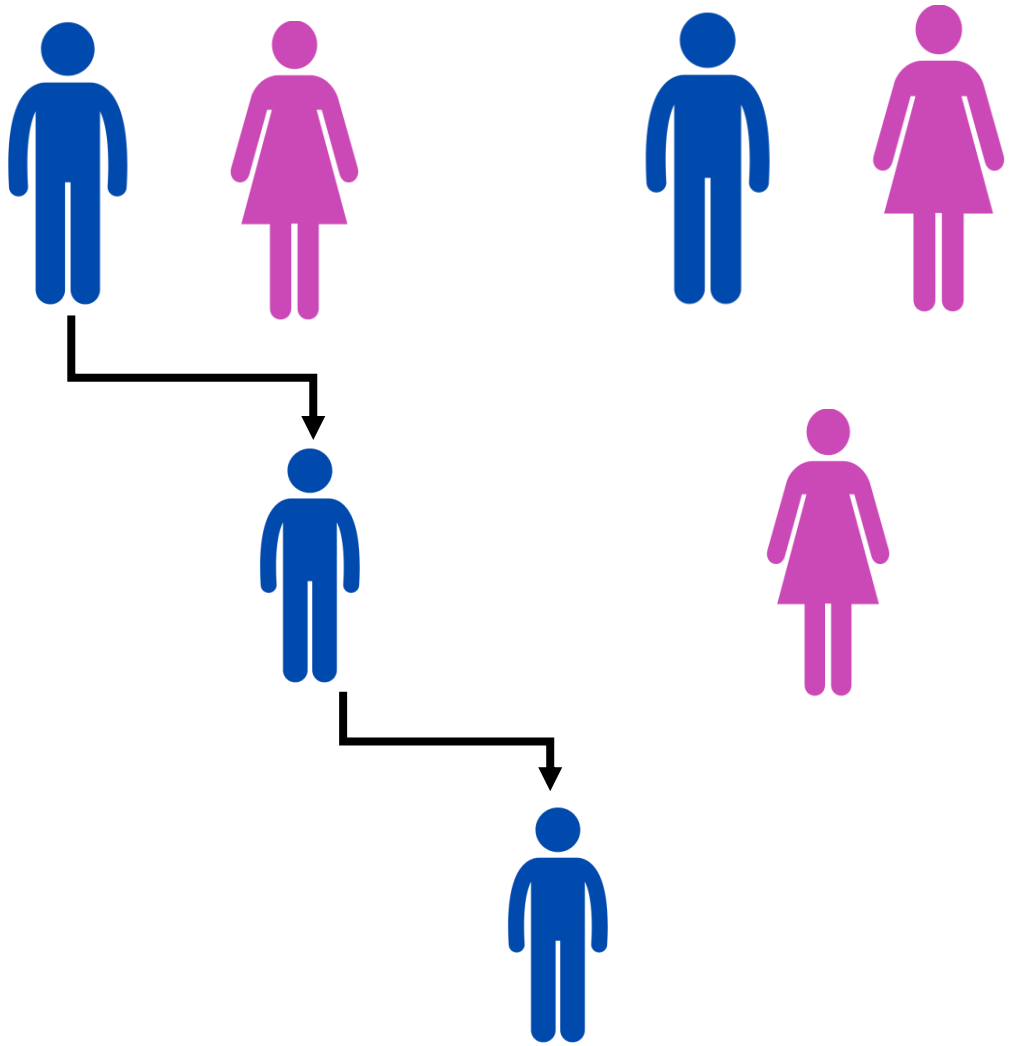
- The double helix structures form 23 chromosomes (1 – 22 + the sex chromosomes, X and Y)
- Each person has two copies of each of the 23 chromosomes
  - One set of chromosomes comes from your mom and one from your dad
  - XX = female (generally)
  - XY = male (generally)

Inheritance

- In both egg and sperm production, a person's DNA goes through recombination
  - For example, mom has two versions of Chromosome 1 (one from her mother and one from her father)
  - Those two versions of Chromosome 1 "mix and match" before a single copy of Chromosome 1 (which is a combination of mom and dad) comes to exist in a single egg
  - Every egg is going to get a different "blend" of mom's two Chromosome 1's

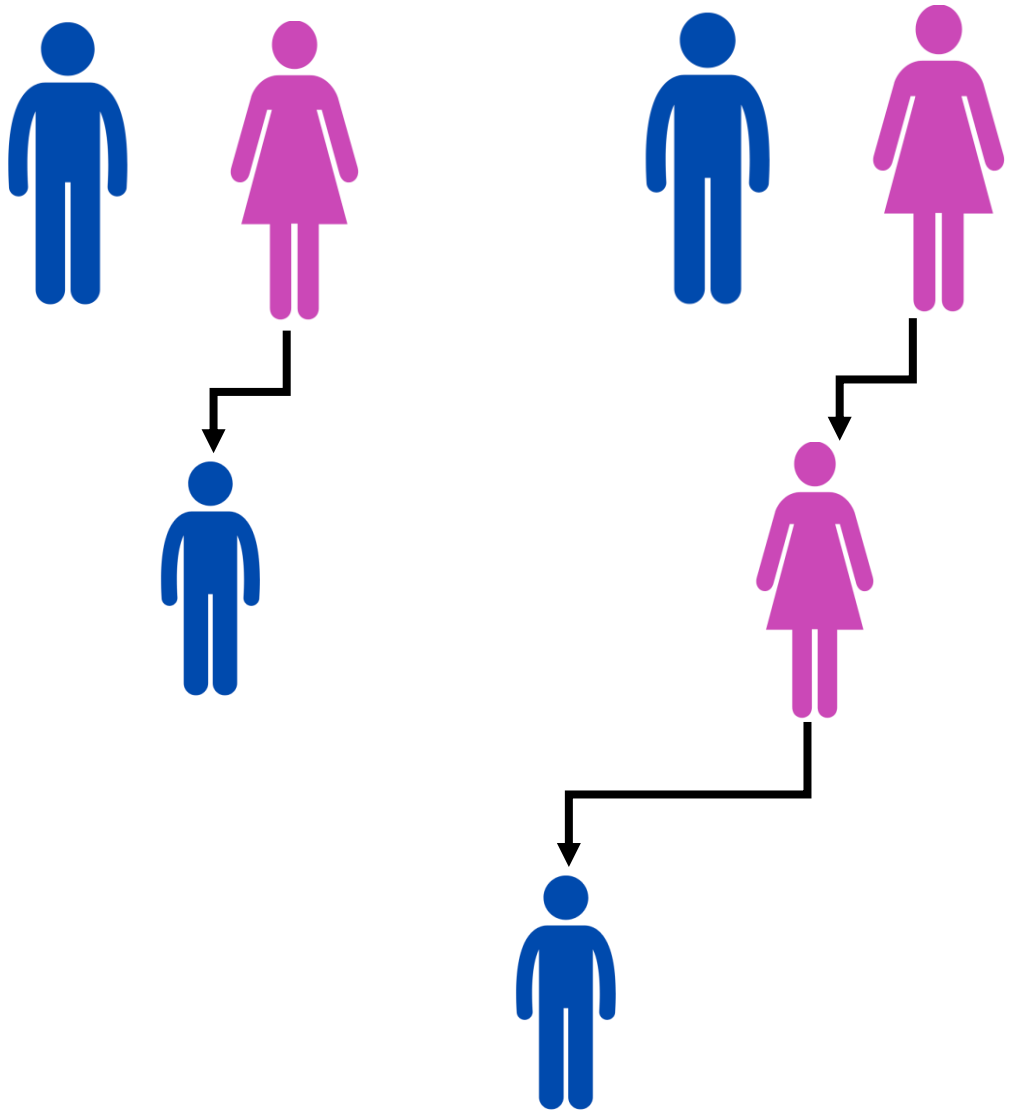
Special Inheritance Patterns: Y DNA

- Y DNA
  - Because a man only has one Y chromosome, it does not recombine
  - Instead, he passes his whole Y chromosome to each son
  - Unless there is a mutation, Y chromosomes can pass down for hundreds of years, dozens of generations, without any change



Special Inheritance Patterns: Mitochondrial DNA

- Mitochondrial DNA
  - Mitochondrial DNA is more simple, and it also does not recombine
  - Mt DNA is passed from mother to children (both sons and daughters), generally unchanged, for many generations



Measuring Genetic Relationship

- Unit of measurement is a “centimorgan” – cM
- One centimorgan is about 1 million base pairs
- The human genome is approximately 3 billion base pairs

- Unless you do “whole genome sequencing,” DNA tests map a tiny fraction of the human genome
- Specifically, Ancestry identifies approximate 700,000 “markers” (specific locations) within your genome

TRADITIONAL FORENSIC DNA

Short Tandem Repeat Sequencing

- Much of our DNA is ”junk”
  - If it does something, we don’t know what that something is
- There are a number of “loci” where we see repeating sequences of nucleotides:
  - For example AGC – AGC – AGC – AGC
- People differ with respect to the number of repeats of the sequence (usually there are few specific numbers of repeats that people might have
- STR sequencing counts the number of repeats at specific locations in your DNA
- This is the type of DNA profile used for traditional law enforcement

CODIS

- CODIS is the Combined DNA Index System
- Federal project that connects databases maintained by individual states
- CODIS is made up primarily of “offender” STR profiles
- If law enforcement finds DNA at a crime scene, they do an STR profile and search CODIS for an exact match
- CODIS uses 20 locations of STRs within the human genome. That’s it!!

Limitations of CODIS

- First, states differ regarding which offenders they add to their database
- Second, if DNA comes from someone who was never arrested before, their DNA will not be in CODIS
  - We are stuck

DIRECT TO CONSUMER DNA TESTING

History

- 23 and Me
  - Created to provide people with HEALTH information
- Later, companies realized that DNA is helpful in genealogical research
- Ancestry.com is the largest database of direct to consumer DNA tests in the world

DTC Tests v. CODIS Tests

- Remember, CODIS is using your “junk” DNA (Short Tandem Repeat segments, or STRs)
- Direct to Consumer Tests, instead, map SNPs
  - Single nucleotide polymorphisms
  - Areas of active genetic importance where the difference between an “A” and a “G” at one location determines something significant
  - Eye color
  - Ability to smell cyanide
  - Tendency to flush from alcohol
  - Tolerance for Brussels sprouts and cilantro

Wendy's 7 ancestral regions

34% Scotland →
3 Stories >

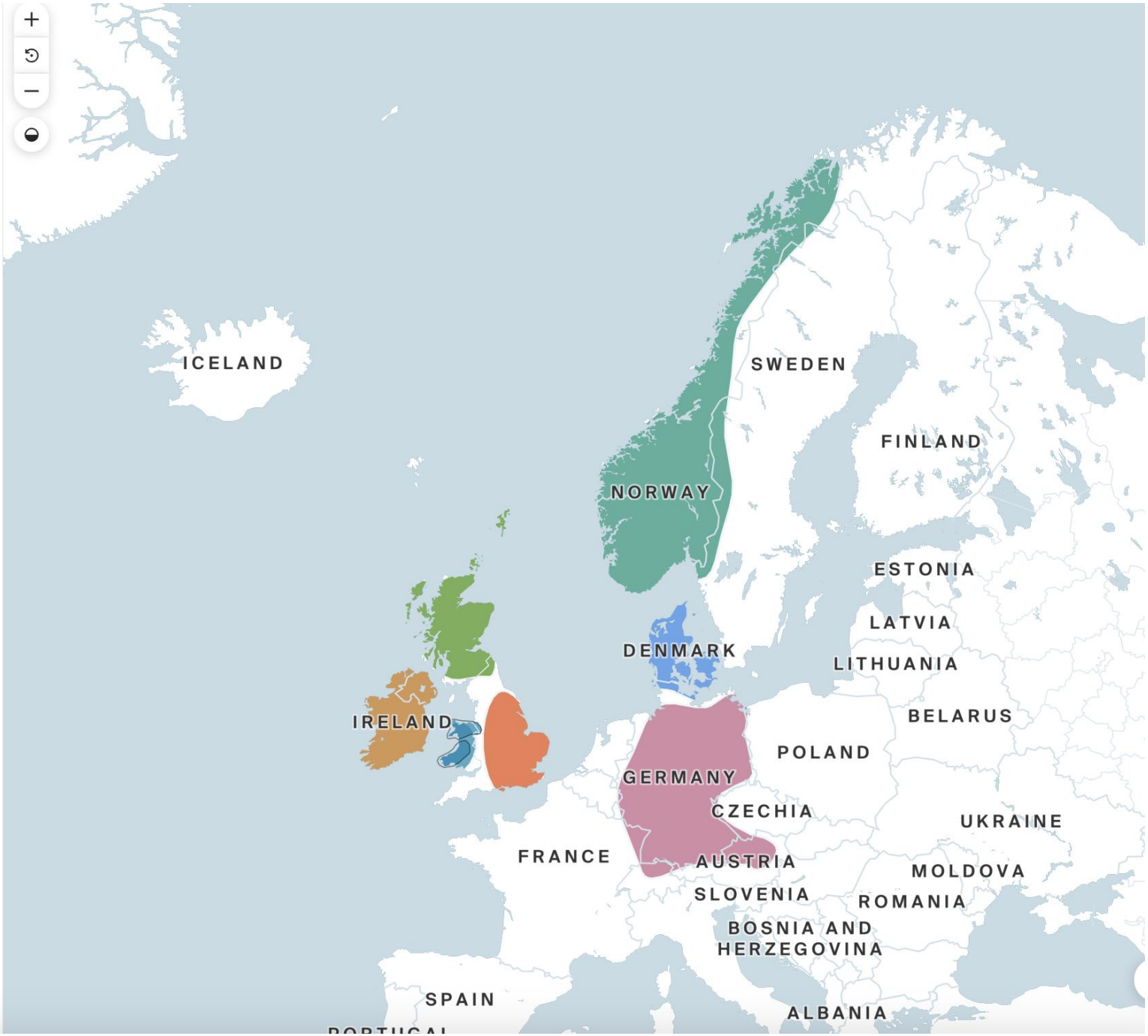
19% Wales →
Northern Wales > New
Western Wales > New
3 Stories >

17% England & Northwestern Europe →
3 Stories >

15% Germanic Europe →
4 Stories >

7% Ireland →
3 Stories >

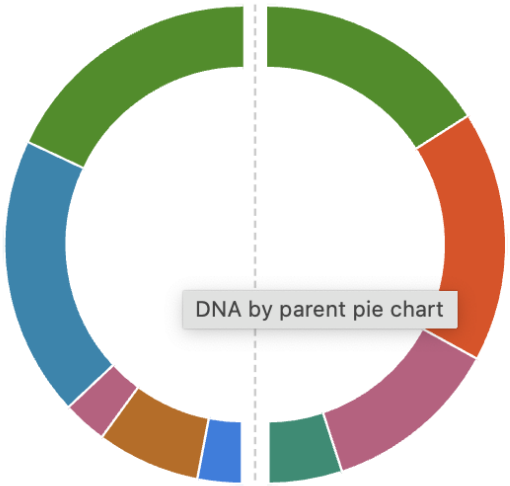
5% Norway →
2 Stories >



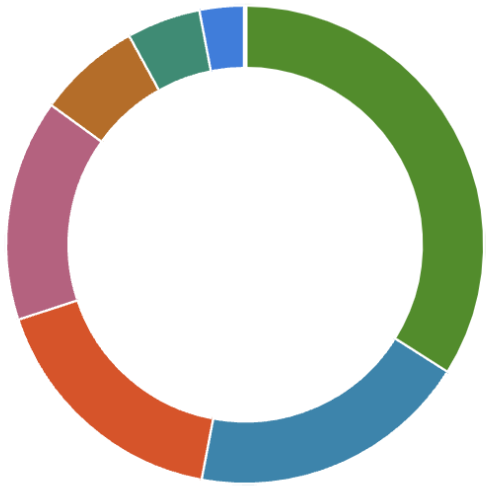
Overview

Share

Paternal Maternal



You



Select one or more regions to highlight.

All | Scotland | Wales | England & Northwestern Europe | Germanic Europe | Ireland | Norway | Denmark

All | By parent | By ancestor | By location

View tree linked to your matches

Filter | + Create group | Sea

Close Family

lylej4662
Managed by jameslyle186

Half aunt
Paternal side
933 cM | 13% shared DNA

Private linked tree
704 people
Common ancestor

+ A

George Watson

1st cousin
Paternal side
717 cM | 10% shared DNA

No trees

+ A

William Lyle
Managed by jameslyle186

Half 1st cousin
Paternal side
512 cM | 7% shared DNA

Public linked tree
7 people
Common ancestor

+ A

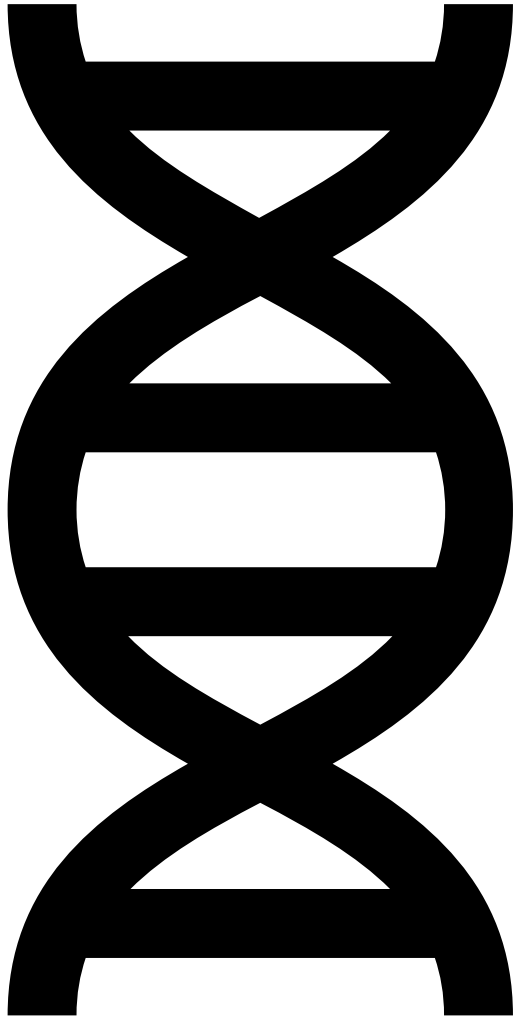
Extended Family

jameslyle186

Half 1st cousin

Private linked tree

+ A

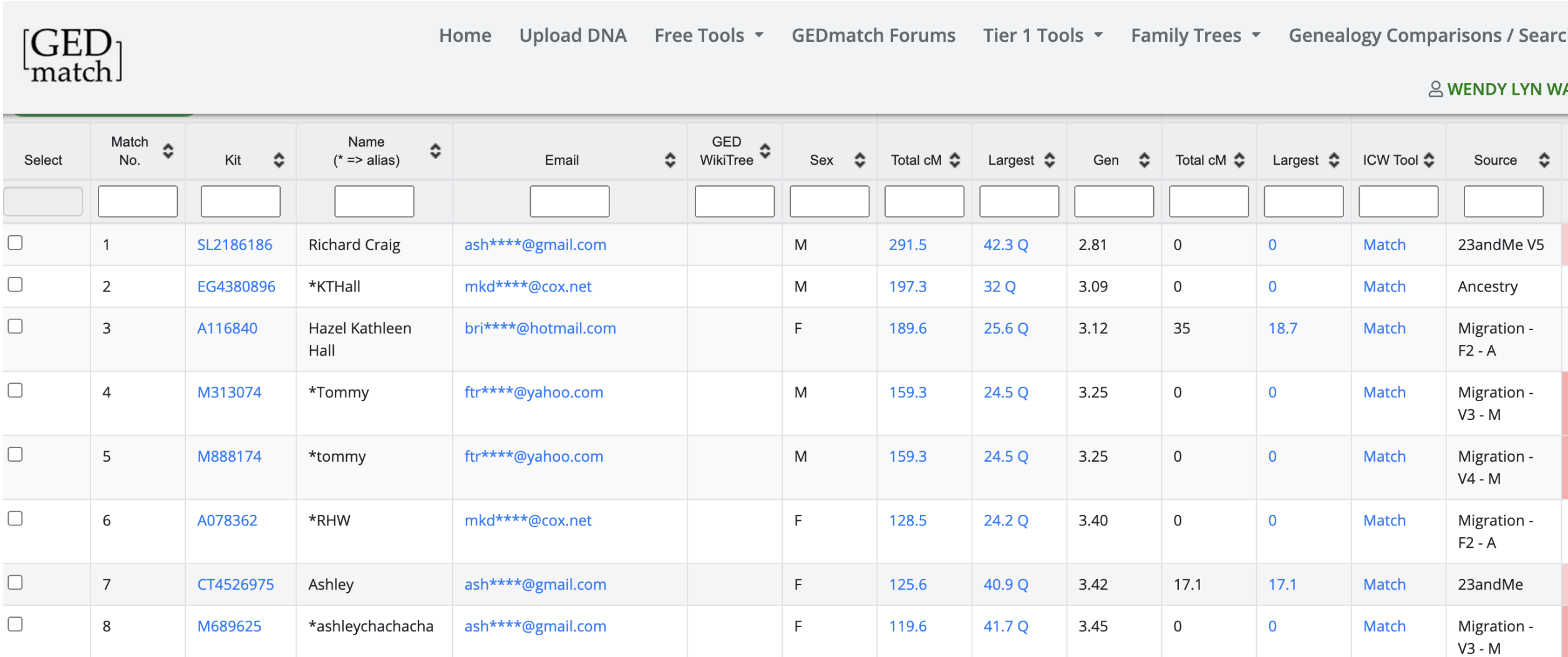


HOW DOES FORENSIC INVESTIGATIVE GENETIC GENEALOGY WORK?

Steps

- Law enforcement obtains a SNP profile of the DNA sample
- That profile is uploaded to one (or more) of three sites:
  - GEDMatch
  - FamilyTree DNA
  - DNA Justice
- FIGG practitioner identifies people who match to the sample
- Build out trees for those people to determine how they are related
- Use math to figure out where, in the big tree, the DNA contributor must be
- Identify candidates
- Law enforcement must verify with a 1:1 sample

GEDMatch



[GED match]

Home | Upload DNA | Free Tools | GEDmatch Forums | Tier 1 Tools | Family Trees | Genealogy Comparisons / Searc

WENDY LYN WA

Select	Match No.	Kit	Name (* => alias)	Email	GED WikiTree	Sex	Total cM	Largest	Gen	Total cM	Largest	ICW Tool	Source
☐	1	SL2186186	Richard Craig	ash****@gmail.com		M	291.5	42.3 Q	2.81	0	0	Match	23andMe V5
☐	2	EG4380896	*KTHall	mkd****@cox.net		M	197.3	32 Q	3.09	0	0	Match	Ancestry
☐	3	A116840	Hazel Kathleen Hall	bri****@hotmail.com		F	189.6	25.6 Q	3.12	35	18.7	Match	Migration - F2 - A
☐	4	M313074	*Tommy	ftr****@yahoo.com		M	159.3	24.5 Q	3.25	0	0	Match	Migration - V3 - M
☐	5	M888174	*tommy	ftr****@yahoo.com		M	159.3	24.5 Q	3.25	0	0	Match	Migration - V4 - M
☐	6	A078362	*RHW	mkd****@cox.net		F	128.5	24.2 Q	3.40	0	0	Match	Migration - F2 - A
☐	7	CT4526975	Ashley	ash****@gmail.com		F	125.6	40.9 Q	3.42	17.1	17.1	Match	23andMe
☐	8	M689625	*ashleychachacha	ash****@gmail.com		F	119.6	41.7 Q	3.45	0	0	Match	Migration - V3 - M

Find Common Ancestors

- These must be the people who contributed the DNA segments shared by the matches and the contributor
- So we know that our contributor must be descended directly from the Most Recent Common Ancestors

What Are the Odds?

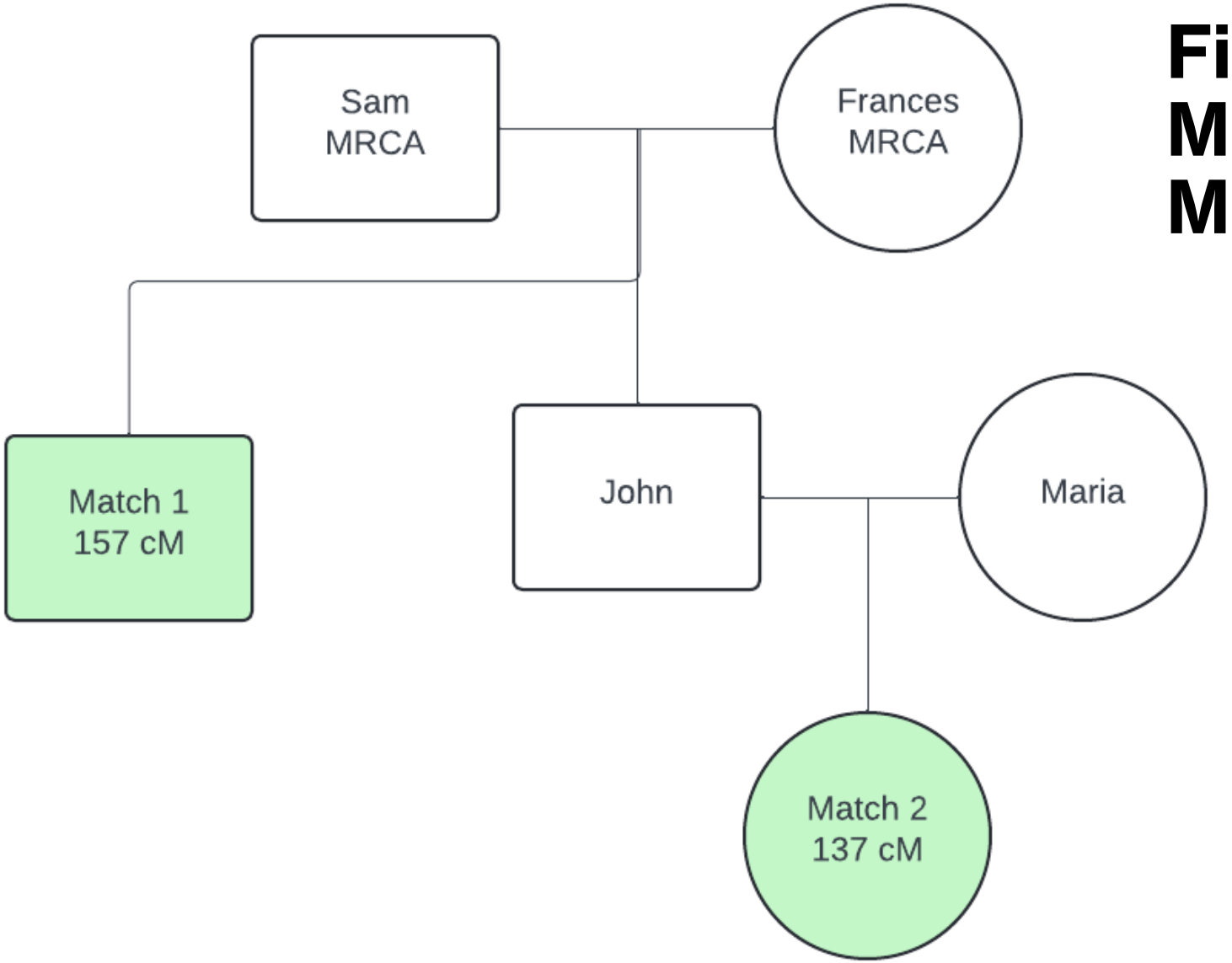
- Where does our unknown DNA contributor fit into the tree?
- Based on probabilities – how can we make each degree of genetic relationship “work” together

EXAMPLE: FINDING FRANK'S GRANDFATHER

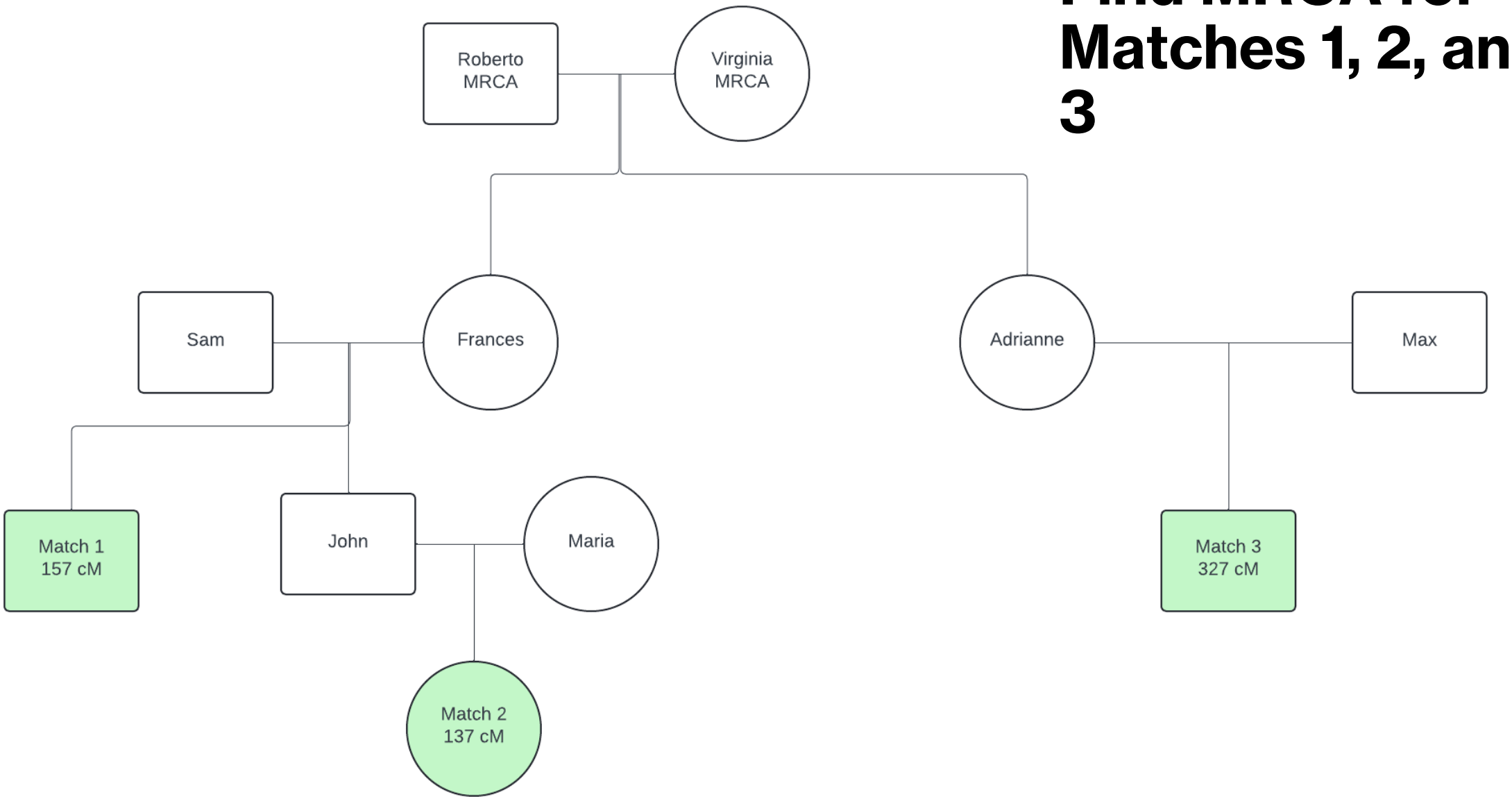
Find Matches to Frank

	Genetic Distance from Frank
Match 1	157
Match 2	137
Match 3	327

Find MRCA for Match 1 and Match 2



Find MRCA for Matches 1, 2, and 3



What Are the Odds?

- What Are the Odds? (WATO) told me that
  - There was a 53% chance that Frank's grandfather was a sibling of Adrianne and Frances
  - There was a 26% chance that Frank's grandfather was a half-sibling of Adrianne and Frances
  - There was a 21% chance that Frank's grandfather was a sibling of Roberto or Virginia

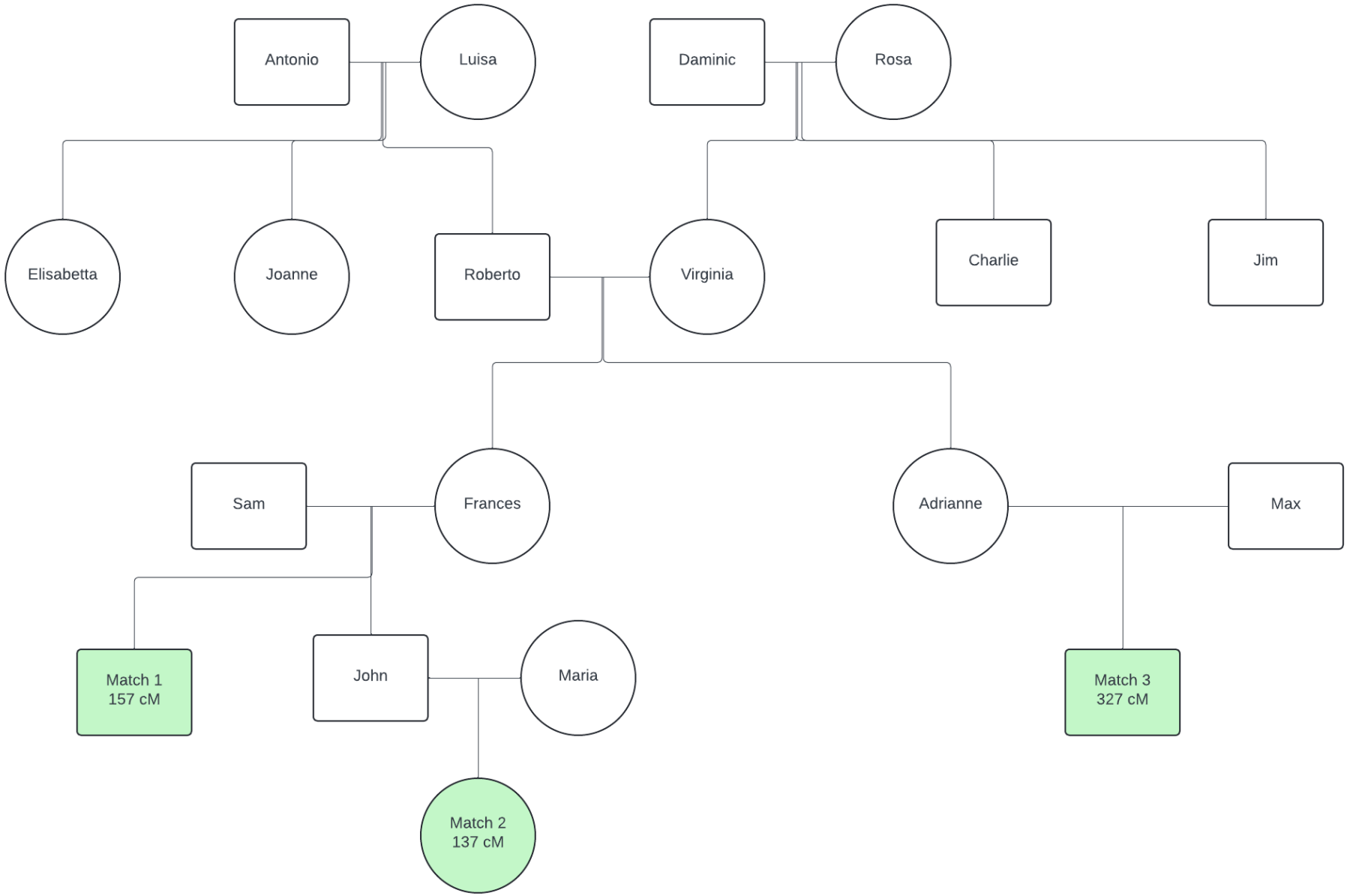
Narrow Down the Possibilities

- Based on when Frank's father was born (1948), I could exclude siblings or half-siblings of Adrianne and Frances
- How?
  - Because Roberto and Virginia were born in 1925
  - The EARLIEST a sibling or half-sibling of Adrianne and Frances could be born was 1925 + 14 (or 1939)
  - If Frank's father was born in 1939, Frank's father would have been under 10 years old when Frank was born
- That means Frank's father must be a sibling of Roberto or Virginia

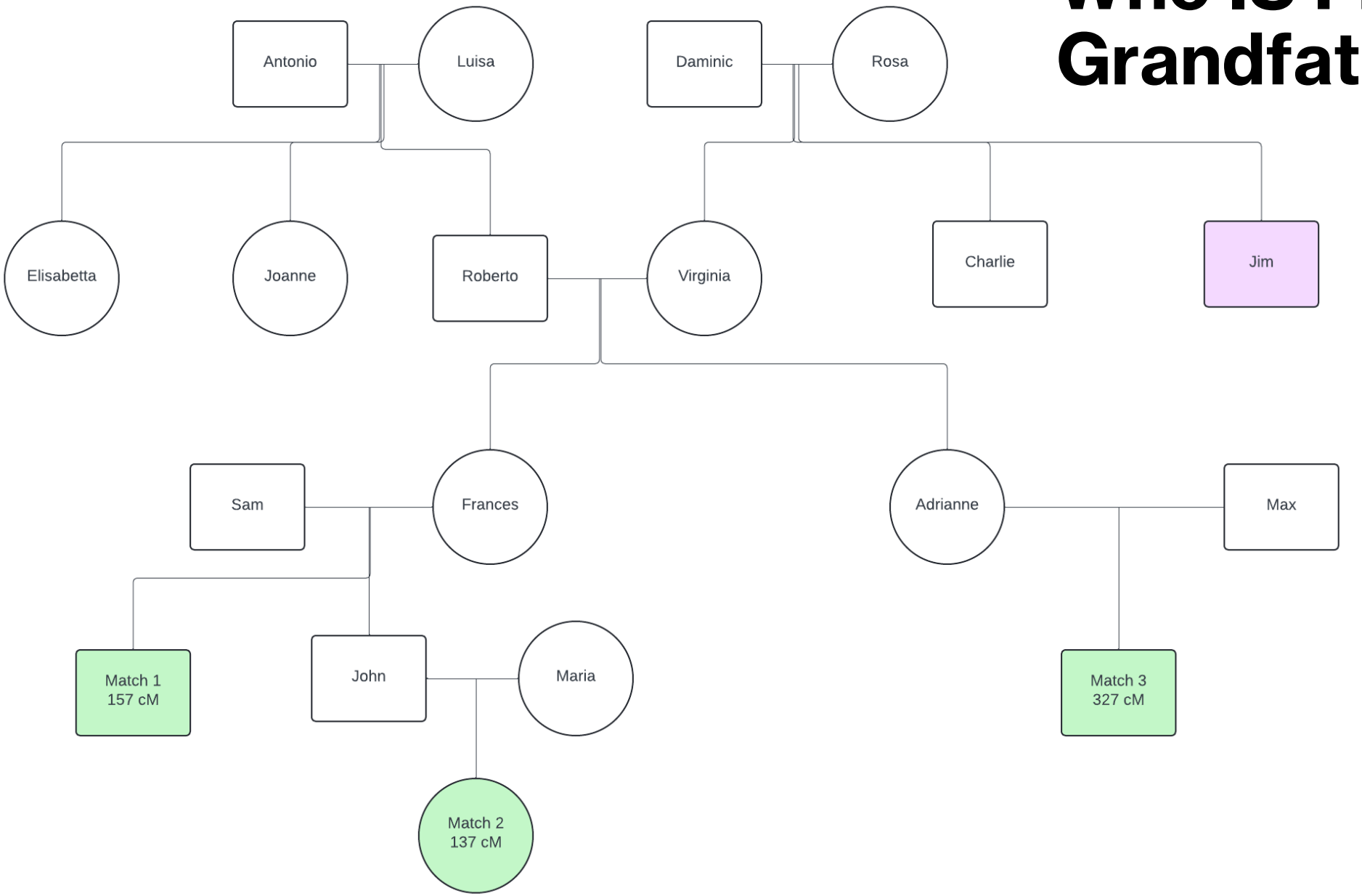
Next Steps

- Build Roberto and Virginia’s trees back a generation and identify the siblings of Roberto and Virginia
- Which of those siblings is most likely to be Frank’s father?

Who Could Be Frank's Grandfather?

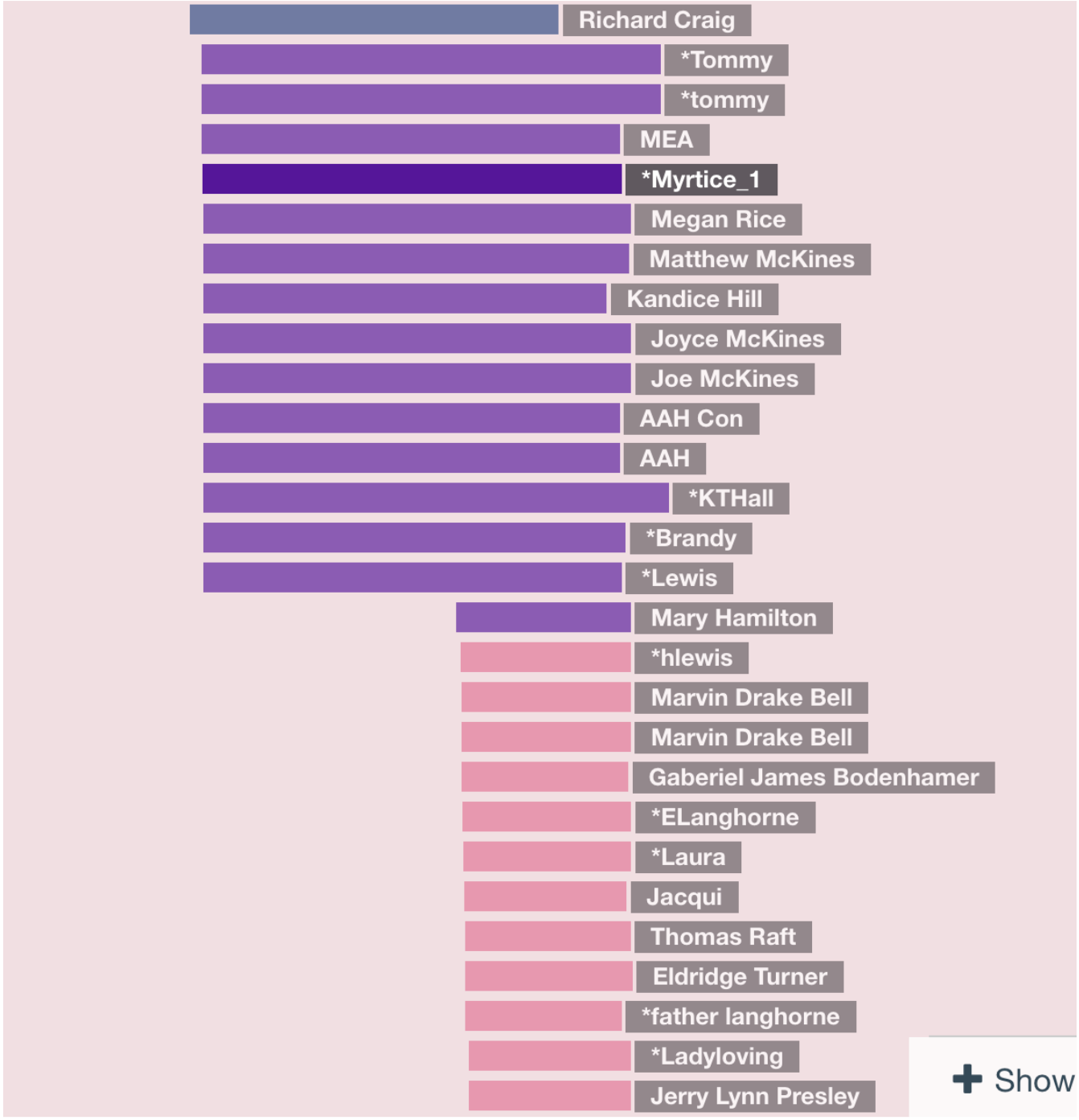


Who IS Frank's Grandfather?

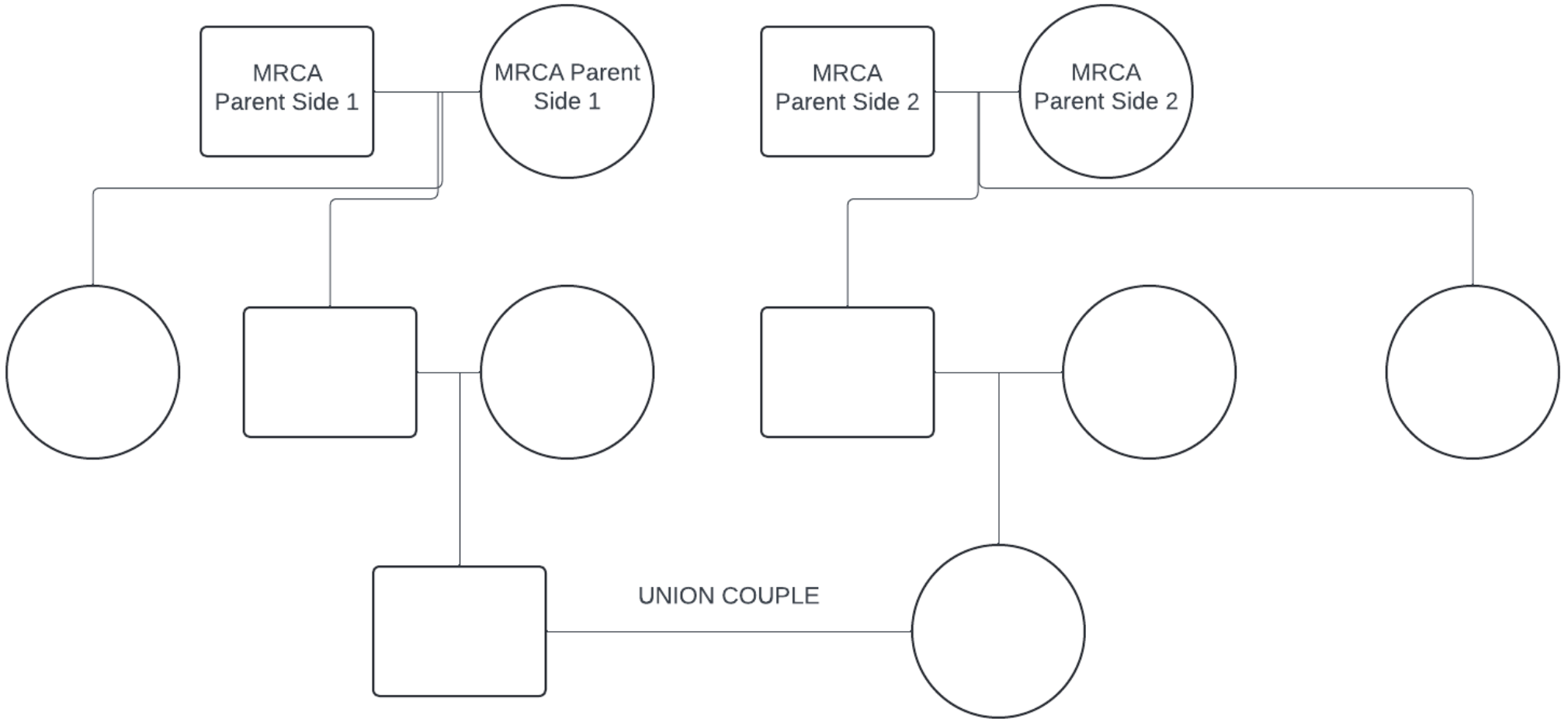


Multiple Clusters

- With a truly unknown DNA contributors we generally find ”clusters” of matches
- One handy thing to do is “paint chromosomes”



Building Down from Maternal and Paternal Common Ancestors



How Can You Help?

- Build your own Ancestry Tree, make it public, and include your name and location in your profile
- Download your DNA from your testing site, and UPLOAD it to GEDmatch and DNAJustice
- Donate to
  - DNA Doe Project
  - Ramapo Investigative Genetic Genealogy Center
  - Moxxy Forensics
  - Intermountain Forensics