Using Forensic Genetic Genealogy – A Practical Guide for Law Enforcement

Terminology you should know

Why you should keep this in your back pocket

These are terminology you should know so you can have effective conversation with your laboratory/investigative partner and make the most out of any interpretation of evidence and data related to FGG and steer your case in the right direction.

Background

The terms Investigative Genetic Genealogy (IGG) and Forensic Genetic Genealogy (FGG) are used synonymously and interchangeably by the forensic community. The Scientific Working Group on DNA Analysis Methods (SWGDAM) & the Department of Justice (DOJ) have used IGG in their working group updates and draft guidelines, respectively. Similarly, the University of New Haven's Henry C. Lee College of Criminal Justice and Forensic Science uses the term FGG as part of its certification course. Verogen currently uses the term FGG in its literature.

Regardless of terminology, FGG can only be used once traditional searching against national law enforcement databases like CODIS has been exhausted. Per the draft guidelines published by the Department of Justice, FGG can be used as an investigative lead generation technique for missing persons cases (unidentified human remains), and violent crime, such as sexual assault and murder. FGG currently cannot be used to investigate, for example, property crimes.

Biological terminology:

Chromosome

A chromosome is a long DNA molecule with part or all of the genetic material of an organism. Humans have 23 pairs of chromosomes. The first pair of 22 chromosomes as referred to as autosomes and the last pair is referred to as sex chromosomes.

X & Y DNA:

X DNA or Y DNA are DNA from sex chromosomes that one inherits from either his/her mother or father. Every biologically female person has two X chromosomes (XX) and every biologically male person has one X chromosome and one Y chromosome (XY). Y DNA can provide information about a person's paternal lineage.

Autosomal DNA:

Autosomal DNA is the kind of DNA one inherits from both parents via the autosomal chromosomes. As opposed to X or Y DNA, that determines your biological sex, Autosomal DNA is 99.9% identical across humans. The 0.1% of variation in this DNA includes information about a person's genealogy / ethnic ancestry.

Mitochondrial DNA

Mitochondrial DNA differs from Autosomal DNA and X & Y DNA in a number of ways, most notably where it's found in the human body. It is found in the mitochondria instead of nucleus of cells. This special kind of DNA is inherited only from one's mother and is relatively unchanged. Mitochondria DNA can provide information about a person's maternal lineage

Single nucleotide polymorphism (SNP)
Single nucleotide polymorphisms, pronounced "snips", are the most common type of genetic variation among people. Each SNP represents a difference in a single DNA building block. SNPs occur normally throughout a person's DNA. There are approximately 10 million SNPs in a person's genetic code. These SNPs contains information about genetic predisposition to certain diseases, and can predict features like hair and eye color, as well as ancestry

SNPs can be used to generate long range pedigrees.

Short tandem repeats (STR)
Short tandem repeats are short repeated sequences of DNA that account for approximately 3% of the human genome. The number of repeat units is highly variable among individuals, as a result of which they have been used for identification purposes. National forensic databases like CODIS/NDIS include the STR profiles of casework forensic samples.

STRs can be used for short range (1st degree) pedigrees.

Whole genome sequencing
Whole genome sequencing (WGS) is the process of determining the entirety, or nearly the entirety, of the DNA sequence of a genome at a single time. This includes typing 3 billion datapoints of DNA

It can generate profiles with low quantities of low-quality DNA samples.

Genotyping by microarrays
A technology that can detect thousands of pre-determined SNPs at once. This is the technology that is used by direct-to-consumer companies like AncestryDNA and 23andMe.

It requires high quantities of high-quality DNA samples.

Targeted sequencing
This is a cost-effective alternative to WGS that generates data on pre-determined set of SNPs that have been selected for a particular application, such as forensic genetic genealogy.

It can generate profiles with low quantities of low-quality DNA samples.

Genealogical terminology:

Centimorgans
Centimorgans measure the overlapping amount of DNA that is shared between two persons (or two DNA profiles). It's a unit of measurement that indicates how related one person (one DNA profile) is to another person (another DNA profile).

Kinship coefficient
Kinship coefficient is a simple statistical measure of relatedness. It is the probability that a piece of autosomal DNA randomly selected from an individual, and the same DNA on the autosomal position from another individual are identical and from the same ancestor.

Kinship coefficients have been used by geneticists for over 30 years to understand relatedness.

Imputation
Imputation is the process of replacing missing DNA information to the DNA profile using statistical models.

Imputation is frequently used for WGS and microarray data because FGG samples tend to be degraded and of low quality. Imputation makes the DNA profile more likely to match more false relations during the genealogical process.

Investigative terminology:

Familial searching
Familial searching or familial DNA searching (FDS) is a deliberate search in a State DNA database using specialized software (separate from CODIS) to find a ranked a list of potential candidates in the DNA database who may be first order biological relatives (e.g. parent, child, sibling) to the unknown individual whose DNA is collected as evidence.

Forensic genealogy
Forensic Genealogy refers to traditional genealogical work associated with civil suits. This includes probate work, landmark records, heir searches etc.

Forensic genetic genealogy
Forensic Genetic Genealogy (FGG) involves uploading a DNA (SNP) profile to a genealogical database so as to identify distant or long-range relatives for the purpose of solving missing persons cases or violent crime.

Investigative lead
Leads that help with triaging a case. In traditional forensics this includes leads identified via phone tip lines, witness statements etc. For FGG this includes potential matches (relatives) to uploaded SNP profile. This list of matches needs to be refined using target testing and additional genealogical information such as census records, marriage license information, predicted biological sex, age etc

Target testing
A method used to filter the number of matches or investigative leads generated by FGG. This involves requesting a subset of matches to take a direct-to-consumer DNA test to help include or exclude certain branches of the family tree for consideration

Confirmatory testing
This is STR based testing of the most likely investigative lead to confirm identify by matching against the original casework sample.

This is the part of the workflow that is court admissible and relies of established forensic practices.

1.833.837.6436 toll-free (North America) • +1.858.285.4101 tel • +44 (0) 208 054 8706 (United Kingdom) • info@www.verogen.com