

Just Identifying Individuals with Forensic Genetic Genealogy

Introduction [00:00:05] Now this is recording RTI International Center for Forensic Science Presents Just Science.

Voiceover [00:00:18] Welcome to Just Science, a podcast for justice professionals and anyone interested in learning more about forensic science, innovative technology, current research, and actionable strategies to improve the criminal justice system. In episode one of our Applications of Forensic Science for Human Identification season, Just Science sat down with Dr. Heather McKiernan, a Research Forensic Scientist at RTI International, and Ashley Rodriguez, a Research Public Health Analyst at RTI International, to discuss the use of forensic genetic genealogy in identifying human remains. Oftentimes unknown human remains are recovered fully or partially skeletonized, with no close biological relatives in CODIS to facilitate identification. Therefore, combining nontraditional DNA testing with genealogical methods and research has become a powerful investigative tool to bring closure to families. Listen, as long as Dr. McKiernan and Ashley discuss how forensic genetic genealogy differs from searching a DNA profile in CODIS, what limitations exist when using DNA technologies to identify unknown remains and the importance of familial reference samples. This episode is funded by the National Institute of Justice's Forensic Technology Center of Excellence. Some content in this podcast may be considered sensitive and may evoke emotional responses or may not be appropriate for younger audiences. Here's your host, Jaclynn McKay.

Jaclynn McKay [00:01:35] Hello and welcome to Just Science. I'm your host, Jaclynn Mackay, with the Forensic Technology Center of Excellence, a program of the National Institute of Justice. Today, we will be discussing the application of forensic genetic genealogy to unidentified remains. Here to guide us in our discussion is Dr. Heather McKiernan and Ashley Rodriguez. Welcome, Heather and Ashley. Thank you for taking the time to talk with us today.

Heather McKiernan [00:01:59] Thanks for having us.

Jaclynn McKay [00:02:00] In order to lay the groundwork for our discussion today, Heather, would you mind explaining what forensic genetic genealogy or FGG is for our listeners?

Dr. Heather McKiernan [00:02:09] Typically, the first thing that comes to mind when somebody says forensic genetic genealogy or investigative genetic genealogy is DNA or DNA analysis. But the definition for forensic genetic genealogy is the application of genealogical methods and research to legal issues. So they use DNA testing results in conjunction with all the same tools that a genealogist would use for traditional family history research. So, family tree construction, census records, vital records like birth, marriage, death certificates, but for a different role or application. In this case, the identification of victims or suspects.

Jaclynn McKay [00:02:53] And how does forensic genetic genealogy differ from the standard practice of searching a DNA profile in CODIS?

Ashley Rodriguez [00:03:02] FGG and searching a profile within CODIS, they're completely different things. One of the questions I get a lot is, I already have DNA in CODIS, can't we just use that for FFG? And the answer is no. The main difference between FFG and CODIS is that they actually require different types of DNA profiles. So

for CODIS, law enforcement is likely familiar with the STR profiles that are developed from crime scene evidence or from human remains, and that are uploaded into the CODIS database in an attempt to identify an unknown perpetrator or a victim. These STR profiles are generally developed by local, state or federal accredited crime labs. On the other hand, for FGG, we use something called SNP, S-N-P, SNP based testing. You may have heard the term targeted SNPs or SNP microarray or whole genome sequencing, and these are all SNP based technologies that yield profiles sufficient for FGG. The easiest way to understand the difference between SNPs and STR profiles is that SNP based testing gives us a much more robust profile and gives us more data to work with and currently crime labs that do the STR profiles for CODIS, they don't have the capabilities to perform the SNP testing and as such, we have to turn to private DNA vendors for this type of analysis. Another difference between the two is the databases used for the actual searching. So STR profiles are obviously uploaded and searched within the CODIS database, but the SNP profiles are uploaded and searched within third party or direct to consumer genealogy databases. Currently GEDMatch Pro, FamilyTreeDNA and DNASolves are the only databases available to law enforcement to upload a forensic sample from an unidentified victim. Lastly, in order to have a hit or an association within CODIS, we need close biological relatives. Ideally, parents or full siblings who have uploaded their DNA also into CODIS, but with FGG the SNP profile actually allows us to identify really distant matches. So think fourth or fifth cousins. So yes, the searching within CODIS and FGG, two very different things. They use different DNA profiles developed by different laboratories, upload it into different databases, and with CODIS, we're typically relying on associations being made to close relatives, whereas FGG can be successful with an initial identification of a very distant relative.

Jaclynn McKay [00:05:48] Thank you for that robust explanation, Ashley. I really appreciate it. Is there a reason why the traditional crime laboratories use STR testing as opposed to SNP testing?

Dr. Heather McKiernan [00:06:00] So that was the original technology that was proposed with a startup CODIS. So we have a lot of historical data that's contained within the national DNA database or CODIS that is going to search historically 13 different STR data points, now 20+ data points that search there. So to completely replace that with a new technology means losing a lot of historical records and data that we might find. You also have to consider that the applications for these technologies have changed a lot. In the past, what was typically done was comparing samples from a crime scene. So blood, semen sample, things like that directly to the same individual. And so using those STR markers are just, you know, up to 20 some points of data for comparison is sufficient to give you that answer, right? You can use population statistics to say, hey, this person had these 20 data points. The sample at the crime scene also produced the same 20 data points, and the chances of seeing that are really, really small. So it works really well for that type of comparison, and that's still done for a lot of the samples that go into CODIS. And as Ashley had said, when we're making these comparisons, then instead of from the same person to the same person or I guess apples to apples, and instead we want to compare it to family members, it's still sufficient to make those comparisons where you share, expect to share about 50% of your DNA. It's only now in these cases where we're looking to extend that searching capability to much greater genetic distances to individuals that we don't share as much DNA with. Where now that's become kind of insufficient and we're looking at these additional markers that could yield a lot more information because with SNPs, as Ashley indicated, instead of just looking at 20 some markers, we're literally looking at hundreds of thousands of data points of comparison. So significantly more

points of comparison to be able to make those much more distant connections in family members.

Jaclynn McKay [00:08:15] It makes a lot of sense as to why you would use SNPs for finding those distant relatives in order to help pinpoint a suspect or a victim. So can you maybe elaborate a little bit more on how exactly forensic genetic genealogy is used specifically to help identify unidentified remains?

Dr. Heather McKiernan [00:08:35] Sure. I'll start out just by putting that into a little bit of context. So there's over 600,000 individuals that go missing in the United States every year with tens of thousands of those remaining missing for a year or more. On top of that, there's roughly 4,400 unidentified bodies that are recovered every year, with about a thousand of those also remaining unidentified after a year. So if you think about it, it's really incredibly sad that nobody's recognized these individuals after so long, and you have to imagine there's family members out there wondering what happened to their loved ones. So in the context of cold missing persons and unidentified human remains cases, forensic genetic genealogy is an extremely powerful tool. In terms of how it works, you would select kind of the best sample that you have available for submission to a vendor lab, as Ashley had indicated. Usually this is done in talking with a medical examiner, coroners, anthropologists, the crime labs, law enforcement, really any pertinent parties that are involved in that case. Once at the lab, that sample would then undergo DNA extraction and cleanup, and then whole genome sequencing is performed to build out those SNP profiles that Ashley had referenced. You can also use something like a microarray that can traditionally be done with samples that are a little bit higher quantity and quality. Usually when we're dealing with unidentified remains, that is not the case. These are kind of the degraded, lower quality samples. And so for that whole genome, sequencing is preferred. Once those SNP profiles are produced, they can be uploaded into searchable databases like GEDMatch or FamilyTreeDNA, and there you're looking for individuals who share large stretches of DNA in common with that sample from your unidentified human remains. So it's a fundamental principle of genetics that the more closely related someone is to you, the more DNA you share in common with them. So like I said, with forensic genetic genealogy, we're looking at hundreds of thousands of these SNP markers to look at these extended family searching capabilities. So once you have the profile in that open data or direct to consumer database, it's compared against the profiles of individuals who have voluntarily submitted their own biological samples or entered their own genetic profiles for this type of searching. There's then computer algorithms that are used to give you a number, and that number corresponds to how much of the DNA matches somebody else's in the database. And then that's where those traditional genealogy research tools come in to help build out family trees, looking for common ancestors to come up with an inference as to who that person might be. And then that's where ideally you're looking to reconnect those genes, or John or even Baby Doe remains back to some distant family member. And a lot of those cases, when you reach out to the family member, they think that person might have run away or they were never reported missing because they never really knew where to look for them or how to go about looking for them. So this really does help bring closure in these cases and I think it's a great example of how really good detective work and good science can bring closure to these types of challenging cases.

Jaclynn McKay [00:12:01] Thank you for that, Heather. So along those same lines in regards to unidentified remains, is FGG more beneficial to helping figure out someone's identity as opposed to trying to search their STR profile in CODIS and is one method preferred over the other?

Ashley Rodriguez [00:12:21] There's certainly pros and cons to both methods. When we are relying on CODIS, if we do upload an STR profile from an unidentified victim, we are hoping the individual has been reported missing and that sufficient family reference samples have also been collected and uploaded into CODIS. But that's not always the case. You know, a lot of the victims, you know, maybe they were never reported missing for whatever reason as Heather kind of alluded to. But with FGG, whether someone has been reported missing or not, it's not an issue. Also, you know, as previously mentioned, FGG does give us the ability to identify distant matches. And we don't have to rely on strict associations to close biological relatives within CODIS. So it kind of expands our pool of potential results and ways to develop investigative leads. On the other hand, CODIS is a readily available resource to law enforcement that uses DNA developed by accredited crime labs as opposed to private DNA vendors. One benefit of using CODIS is that CODIS offers us the ability to identify case connectivity. So, for example, if we're seeking to identify an unknown perpetrator and we upload his or her DNA into CODIS, we may obtain hits to other unresolved cases. For unidentified remains cases, let's just say that we have - we only recovered partial remains, but CODIS will afford us the opportunity to identify multiple sets of remains that may have originated from the same individual, but for whatever reason, if the sets of remains were recovered at different times in different locations, law enforcement may not be able to make the association just by reading reports, but if they upload all of the DNA profiles into CODIS, CODIS can make that association for us. So with FGG, we don't have this option to link cases. So with that being said, the U.S. Department of Justice has issued an interim policy on the application of FGG to criminal investigations. This policy was put into place back in 2019, and it offers suggestions and guidance on how to apply this technology appropriately. And the current guidance from DOJ is that a DNA profile must be searching in CODIS prior to attempting FGG, so that is what we should aim to adhere to.

Jaclynn McKay [00:14:50] With regards to unidentified remains, decedents tend to be in various stages of decomposition and as you both have discussed, in order to use FGG or search a DNA profile in CODIS, you actually have to be able to obtain DNA from the individual. So if you are unable to obtain blood or take buccal swab from the decedent, what other sample types could be utilized in order to get DNA profiles from the decedent?

Dr. Heather McKiernan [00:15:24] Unlike with assailant or crime scene samples where the DNA might be typically more pristine, the DNA with unidentified human remains cases can be much more degraded and or commingled with bacterial DNA and other sources. So we almost never have blood or buccal swabs or things like that available to us for testing. Most of the times these are remains cases where what we have left is skeletonized. So typically the samples that we deal with are going to be human bone samples. Hair has also been used with some success for these types of analysis as well. And then once they go through their whole process, goes to kind of that DNA extraction, there's an additional quality assessment generally that includes some clean up and human enrichment. Again, that's needed in these types of cases to remove any bacterial contamination, and that's done before they move forward with that DNA sequencing and the bioinformatics.

Jaclynn McKay [00:16:33] Heather you stated that most of the unidentified remains are typically skeletonized by the time DNA analysis is performed. Are there any other methods that could be attempted prior to DNA or at what point would DNA and FGG be brought into the mix?

Dr. Heather McKiernan [00:16:53] We always recommend the lower hanging fruit in terms of biometrics are tackled first. Genealogy is kind of like the shiny new toy and so certainly

with the successes you see in the media, a lot of agencies want to take advantage of this service for their cases that they're trying to clear out. But through kind of triage of what's been done in these cases, a lot of times we can go back and say, hey, fingerprint cards were never looked at or we didn't try to track down dental records. And so when we go back and work with agencies to do some of those things, you can close out some of those cases fairly easily without needing to go through all of the testing and cost and labor that's involved with traditional DNA testing and genealogy. So I would certainly recommend going through kind of those steps and processes first prior to attempting the DNA. We do advise law enforcement, especially in the missing persons side, to collect those family reference samples just as soon as possible, even if they can't submit them yet, just because the family members might not always be around and or willing to provide those samples and so it's important to get them when you can, regardless of kind of what phase of the investigation and or testing they're in.

Jaclynn McKay [00:18:17] So once the forensic genetic genealogy is complete and you believe you found the individual associated with the SNP profile, how is this then confirmed?

Ashley Rodriguez [00:18:29] If we're successful, hopefully FGG will provide us an investigative lead for law enforcement suggesting a possible name for the victim. The FGG analysis and the lead information alone is not enough to officially rule the identification, so it has to be confirmed by traditional STR analysis by an accredited crime laboratory. Once law enforcement has the potential identification and they have conducted appropriate follow up investigation to confirm the suggested individual is likely the unknown victim, we would make arrangements to collect reference samples from family members, ideally from mom and dad or full siblings and then STR profiles would be developed from those reference samples and compared to an STR profile developed from the remains to confirm that identification.

Dr. Heather McKiernan [00:19:18] I would just add on to that, that having the STRs for comparison is kind of the preferred method of confirmation and identification. But in these unidentified human remains cases, unlike with your suspect identification, they're not typically going to court. And so we've certainly seen where we've been unable to collect parents or full siblings to do that confirmation. They're either no longer living or unwilling to collect. We had one case recently where the mom was still living but was of such advanced age and had always convinced herself that her child had kind of just moved away and started a life in another family and they didn't want to kind of soil that imagery she had, so they weren't willing to go and collect the sample from her and so we have to rely on other forms of identification because usually when you're talking about identification of the UHR, this isn't necessarily a question that's going into court itself who the victim was. And so we have used fingerprint records to do that confirmation and or dental records, and sometimes, again, if you don't have those family members that share enough DNA to do the STR analysis on, they have had to rely on comparing the SNP profiles of more distant relatives to that person to issue a report. But it basically comes down to what the coroner or the medical examiner is comfortable with in terms of having sufficient information to make that final identification.

Jaclynn McKay [00:20:54] I think you both alluded to this a little bit already as far as how you can't link cases using FGG and people need to voluntarily submit samples, but are there any other limitations to FGG for criminal justice purposes?

Ashley Rodriguez [00:21:13] Yeah so FGG, as we stated, it's become very popular and it really is a game changer for cold case investigations, but it does have limitations. So for one, FGG is costly. So right now we have to use private DNA vendors that can come with a hefty price tag. And there are also fees associated with uploading the data into GEDMatch and FamilyTreeDNA. And unless the law enforcement agencies are able to conduct the genealogy work in-house, meaning the actual tree building portion of the process in-house, there's likely going to be a need to pay a genealogist to assist. So, you know, in today's world where budgets are tight and law enforcement doesn't have a lot of money to be spending on extra things, FGG that is a big limitation just because of the price tag associated with it. Second, it's really important for us to understand that FGG is not going to resolve every single case. So while the associated DNA technologies are rapidly evolving and improving, there are still some cases in which the DNA is just too minimal or too contaminated to proceed, at least for right now. So that definitely may change as the technologies do continue to improve, but right now, there is some limitation to the technology itself. There are also limitations to the databases used in FGG. Generally speaking, minority populations are somewhat underrepresented in those databases. So if, for example, we are trying to identify a victim who is believed to be Hispanic, we may run into issues getting strong enough matches to proceed with the tree building process. You know, I do think that the genealogists are being able to adapt to some of those issues with the databases and to the populations that are represented in them. It is an issue that we just need to be aware of that it is possible for minority populations that we may just have a more difficult time identifying those good matches to be able to proceed with the process.

Jaclynn McKay [00:23:13] For law enforcement agencies who would like to look into the possibility of using FGG on one of their cases, what steps should they take to get started and how would they know if their case is even eligible?

Dr. Heather McKiernan [00:23:26] So Ashley and I both work with programs that are federally funded, which means that we follow those guidelines set forth by the Department of Justice interim policy for forensic genetic genealogical DNA analysis and searching, long title, that Ashley had reference before. So this essentially requires that the unidentified human remains case has been entered into the NAMUS database, as well as the Violent Criminal Apprehension Program national database, or ViCAP, that the unidentified human remains are believed to be those of a homicide victim. So we're looking in the triage there for obvious signs of foul play. Unfortunately, that does mean we can't currently use federal funding to help identify unidentified human remains in cases of natural, accidental or suicide cases. We're also then looking that traditional DNA testing has been attempted and that a resulting at least STR profile has been uploaded to CODIS and that subsequent CODIS searches have failed to produce a probative and confirmed DNA match. And then finally, that all relevant investigative leads have already been pursued. In addition to that, we have some internal triage mechanisms that we look at for cases, and that's just because there are more requests for forensic genetic genealogical testing for unidentified human remains cases right now than our programs could fund and so some of those internal triage mechanisms look at whether or not mitochondrial DNA analysis has also been attempted and searched in CODIS, how old the cases is, what the quality and quantity of samples that remain are, whether or not those remains are available or if exhumations have to be conducted. So those types of factors we look at in determining which cases we could move forward on. And those would be the same things I would advise law enforcement to consider. Again, they probably have a much larger caseload than they have funding for, for forensic genetic genealogy. So they want to go through the full investigative process and make sure that they've kind of exhausted all of

those other biometrics and then just utilize forensic genetic genealogy for those cases that really do need it.

Ashley Rodriguez [00:26:01] For law enforcement agencies who want to start this process, you know, I can't emphasize enough how important it is to work with your local crime lab and even the vendor labs. If they have already gone through the process of identifying a vendor lab that they would like to work with, just really making sure that the communication is open between law enforcement, the crime lab, the vendor lab to be able to evaluate the cases appropriately. You know, Heather talked about quantity of DNA, quality of DNA, all of those factors that really play a role in how successful FGG will be. And you know, law enforcement, it's to their benefit to be able to work closely with their crime lab to to evaluate the evidence and to make sure that that we're not just submitting evidence just because this is a new fancy tool, but that we're submitting evidence that's going to give us the highest likelihood of producing results without consuming, you know, minimal evidence that may be left in some of these cold case investigations.

Jaclynn McKay [00:27:01] Thank you both for that insight. As we wrap up, do you have any final thoughts you would like to leave our listeners with?

Ashley Rodriguez [00:27:09] You know, FGG has just offered so much hope to families and to law enforcement investigators who are involved with these cold case investigations, and not necessarily cold case investigations, just criminal investigations in general and we certainly don't want to do anything within the field that's going to jeopardize this resource, so I really feel that it's the joint responsibility of all involved criminal justice professionals to make sure that we are doing our best to follow the DOJ interim policy and to adhere to the terms of service that have been set forth by the companies managing the genealogy databases to ensure that this tool can continue to be used as a technique to develop investigative leads and to hopefully continue serving justice.

Dr. Heather McKiernan [00:27:56] I think that's well said, Ashley. I think that's a good point to end on.

Jaclynn McKay [00:27:59] Well, Heather and Ashley, thank you so much for your time and for sharing your insight with us. I've really enjoyed this discussion.

Dr. Heather McKiernan [00:28:06] Thank you for having us.

Ashley Rodriguez [00:28:07] Yes, thank you very much.

Jaclynn McKay [00:28:09] If you enjoyed today's episode, be sure to like and follow Just Science on your platform of choice. For more information on today's topic and resources in the forensics field, visit ForensicCOE.org. I'm Jaclynn McKay and this has been another episode of Just Science.

Voiceover [00:28:28] Next week just sits down with Brian Johnson to discuss postmortem print identification. Opinions or points of views expressed in this podcast represent a consensus of the authors and do not necessarily represent the official position or policies of its funding.