

Harnessing NGS Technology to Solve Crimes

Criminalists at the California Department of Justice find that the MiSeq FGx[™] Forensic Genomics System is a significant advancement in DNA profiling technology over traditional CE-based methods.

Introduction

Targeted DNA analysis is an extraordinarily effective tool for law enforcement officials to: identify arrestees; determine links between forensic evidence in criminal cases; solve open, active crimes, as well as "cold cases"; and exonerate innocent suspects. For nearly 2 decades, forensic scientists have used polymerase chain reaction (PCR) and capillary electrophoresis (CE)-based methods to detect fragment length variation in short tandem repeat (STR) markers, creating DNA profiles of suspects, arrestees, and convicted offenders. These profiles serve as national, state, and local databases around the world, and are used by law enforcement to aid in resolving criminal and missing persons cases.

One such database is overseen by the California Department of Justice (Cal DOJ). The CAL-DNA database is the largest working state DNA databank in the United States, housing more than 2.4 million profiles. Most of these profiles were uploaded after California passed the Unsolved Crime and Innocence Protection Act in 2004, which requires that adult detainees provide DNA samples—either via saliva or blood—upon felony arrest.¹

The steady increase in DNA processing of casework samples is limited by the fixed capabilities of CE-based methods. The solution is to harness the massively parallel processing capabilities of nextgeneration sequencing (NGS). The MiSeq FGx Forensic Genomics System uses high-resolution NGS-based genotyping to offer increased efficiency and higher accuracy. In addition to generating more comprehensive genomic information from the least amount of sample, the MiSeq FGx System consolidates numerous individual workflows into 1 streamlined process.

Bill Hudlow, a Senior Criminalist at Cal DOJ, develops and validates methodologies for forensic DNA testing. iCommunity spoke with him about the value of using NGS for DNA profiling and his experience using the MiSeq FGx System.

Q: How many samples do you receive and process each year for testing?

Bill Hudlow (BH): Our laboratory system serves 46 of the 58 counties throughout the state of California, assisting with everything from crime scene response to DNA analysis. The databank section of the lab receives approximately 10,000 samples a month. We perform analysis on about 110 cases a month in our criminal casework section and average approximately 5 samples per case, excluding controls.

Q: How critical is the turnaround time in forensic testing?

BH: Turnaround time of DNA analysis for forensic samples is very critical for several reasons. The first is a matter of public safety. In cases where there's an unknown assailant, the results of forensic testing can be used to confirm their identity. The sooner we can identify that individual, the sooner we can aid the criminal investigation and help get that person off the street. Turnaround time is also critical to the legal process. It is not unusual to have court dates scheduled soon after we receive the cases. We might have to process a case in a short timeframe so that the results are ready for trial.

Q: What is CODIS?

BH: The Combined DNA Index System (CODIS) is an acronym that refers to the US Federal Bureau of Investigation (FBI) program to support criminal justice DNA databases and the software that runs them. CODIS has 3 levels: local (LDIS), state (SDIS) and national (NDIS). The CAL-DNA offender/arrestee database is part of the state level SDIS for California.

Q: What is the value of increasing the number of DNA profiles in CODIS?

BH: In the early days of the CODIS NDIS database, and for the CAL-DNA database in particular, it housed DNA from the 'worst of the worst' criminals, convicted offenders of sexual assaults and homicides. We know that criminals don't always limit their activities to 1 offense category. A person might commit burglary today, and murder



Bill Hudlow is a Senior Criminalist at the California Department of Justice.

or rape tomorrow, or commit these crimes in tandem. By expanding CODIS beyond the DNA profiles of homicide and sexual assault felons, the profiles of offenders who are in a criminal activity escalation phase are available in CODIS to help protect public safety. That increases our chances of associating an unidentified body fluid left at a homicide or assault crime scene with someone who has been convicted previously of a lesser crime.

Q: What kind of genetic markers does Cal DOJ use?

BH: Our primary markers are at the autosomal STR loci that we upload into CODIS and Y-STR loci that we use when we encounter mixtures of male and female DNA. We also use mitochondrial DNA (mtDNA) testing in certain circumstances when the biological material is limited or degraded, such as in missing person cases. Because mtDNA is maternally inherited, an individual's mother, siblings, grandmother, and all other maternally related family members will have the same mtDNA sequences. This enables forensic comparisons using a reference sample from any maternal relative, even distant ones who are separated by several generations. We also use mtDNA testing to analyze hair shafts, where there is no nuclear DNA and STR testing is impossible.

"The MiSeq FGx System is comparable and in some cases, more sensitive than the standard CE technology."

Q: What sample sizes do you typically analyze?

BH: The sample sizes we receive for testing can vary greatly, from nearly microscopic to very large samples. The transfer of DNA from a victim to a suspect, or vice versa, might involve just a small hair with no root shaft. In contrast, homicides can result in large quantities of blood being deposited on large substrates like carpet or bedding.

Q: When did Cal DOJ become interested in using NGS for forensic studies?

BH: We've been interested in accessing NGS capabilities in forensic testing since the technologies were first introduced. Until recently, there really haven't been any kits that we could purchase and validate internally.

Q: What insights can the MiSeq FGx Forensic Genomics System offer that you can't obtain using CE-based analysis methods? BH: The ForenSeq[™] DNA Signature Prep Kit offers 3 major benefits. First, it gives us additional depth for the loci that we're testing. That includes more autosomal STR loci and Y-STR loci than what's previously been available in a kit, along with sequence variation information, or high-resolution genotyping, that we don't get from CE. Second, it provides us with more forensic information in the form of SNPs. And third, the phenotypic information that we can generate with the ForenSeq Kit Universal Analysis Software can provide investigators with information about the perpetrator's biogeographical ancestry and physical traits, which should allow the investigators to narrow down their list of suspects and not focus on individuals who are not likely to be contributors of the biological evidence.

Q: Cal DOJ has collaborated with Illumina since 2012. What types of forensic studies were you interested in performing in the early stages of your projects?

BH: When we began working with Illumina, the ForenSeq system was under development, so we began collaborating together on SNP typing. That was a valuable experience because it allowed us to become more familiar with the MiSeq platform, and learn about the advantages and limitations of SNP loci. As our collaboration evolved, we became a beta tester and early user of the MiSeq FGx Instrument and the ForenSeq DNA Signature Prep Kit. We're currently performing internal validation studies of these products, including the ForenSeq Universal Analysis Software.

Q: Have you worked with other NGS systems?

BH: No, but several individuals in our laboratory have worked with the Roche 454 GS Junior system that is being discontinued. We are now in the process of phasing in the MiSeq FGx Instrument and phasing out use of the Roche system.

Q: How is the MiSeq FGx Forensic Genomics System performing in terms of reproducibility, sensitivity, and analyzing difficult samples? BH: We've compared our beta test data and the results of testing the same samples using CE technology. The MiSeq FGx System is comparable and in some cases, more sensitive than the standard CE technology. We're in the middle of our validation study and are looking forward to generating more comparison data.

Q: How long did it take before the MiSeq FGx Instrument was up and running in your laboratory?

BH: We had the MiSeq FGx Instrument running almost instantaneously. After the sequencer was installed, it didn't take us any time at all before we were generating the ForenSeq libraries to run on it.

"After the sequencer was installed, it didn't take us any time at all before we were generating the ForenSeq libraries to run on it."

Q: Is the ForenSeq DNA Signature Prep Kit easy to use? BH: From the perspective of an experienced DNA analyst, the ForenSeq Kit is very simple to use and user-friendly. There is some training that is necessary to use it because there are a few steps in the process that we don't have experience performing. For instance, the ForenSeq kit has bead purification and normalization that we don't currently do with a CE-based system. I don't anticipate that will be a problem. It will be a matter of training the individuals and getting them comfortable with the ForenSeq workflow and processes. Today, a few people are using the MiSeq FGx System in our lab. After we've implemented the ForenSeq system, we will train more people on how to operate the System, consisting of the ForenSeq Kit, the MiSeq FGx Instrument, and the ForenSeq Universal Analysis Software.

Bioinformatics is another area in which training will be required, but in many ways, NGS data analysis is simpler than it is with CE. This is because the MiSeq FGx provides a digital output—the sequence reads. With CE-based systems, you have to infer the length of amplicons in question, based on comparison to an internal size standard, which is then compared to an allelic ladder to determine the alleles present within that sample. With CE, there are also interpretation aspects such as dye blobs, pull up or spectral bleed through, and fluorescent spikes that aren't present with NGS.

Q: How does the MiSeq FGx System impact workflow?

BH: The ForenSeq library preparation and forensic testing processes are a little more involved with the MiSeq FGx System, and it has a longer running time. However, it generates far more information, particularly on degraded samples, so it's worth the wait. During a MiSeq FGx Instrument run, criminalists are freed up to perform other laboratory duties.

"I think the benefit of the MiSeq FGx System is that it offers multiple marker sets within 1 typing system."

Q: Are there markers that you're able to process with NGS that you're not able to with CE technology?

BH: The MiSeq FGx System coamplifies and types more than 200 loci simultaneously. With CE technology, you can only look at a few dozen loci at a time, and those are typically one type of marker—either you're looking at autosomal STRs or Y-STRs. The current CE-based megaplexes offer 24–25 autosomal STRs with limited Y-STR information, but that's all they provide. They don't offer comprehensive Y-STR typing at the same time or in the same kit. I think the benefit of the MiSeq FGx System is that it offers multiple marker sets within 1 typing system.

Q: What's the value of SNP marker set testing that the MiSeq FGx System also performs?

BH: The SNP marker set will be valuable, especially for poor quality samples. The size of the SNP loci is significantly smaller than that of an STR. So the SNP markers will provide information that CE-based systems can't deliver for degraded samples. The biogeographical ancestry and phenotypic SNPs will be valuable because they provide information to aid investigations where there's no primary suspect and no hits in forensic databases.

Q: What is the value of the MiSeq FGx System in performing lowlevel mixture detection?

BH: I think there is value in using the MiSeq FGx System over CE in mixture interpretation. It appears to offer more sensitivity and provides high-resolution allele sequence data that helps distinguish individuals. There's also the added benefit of having multiple marker types—autosomal STRs, Y-STRs, X-STRs, and SNPs—within the ForenSeq Kit so they can be assayed simultaneously.

Q: How easy is it to use the ForenSeq Universal Analysis software? BH: We have been using the 1.0 version of the ForenSeq Universal Analysis software for about 3 months and it's simple and easy to interpret the results. The software quickly and automatically analyzes the data at the conclusion of a run, so when the run is completed, so is the analysis. You can even change some of the parameters and thresholds, and reanalyze the data if you want. As we become more familiar with the software, we'll be able to use it to the fullest of its abilities. I think the bioinformatics side of the process is something that we're all going to learn a lot more about in the next year.

Q: How does the ForenSeq Universal Analysis Software present data?

BH: We can see the allele calls for the autosomal STRs, X-STRs, and Y-STRs. We can also look at sequence data and see that those alleles have sequence variants, which is something that we're not used to seeing, and might be quite valuable in resolving DNA mixtures. For example, we know that STR alleles of the same length can actually be composed of different sequences, and we cannot detect those with CE typing. Now we have all of that data with the ForenSeq software. We also see phenotypic estimates for hair and eye color, something that we've never had routine access to before in the forensic laboratory.

Q: Cal DOJ is one of several forensic labs that are part of a 19-month study of advanced forensic tools under a National Institute of Justice (NIJ) grant. Will you be using the MiSeq FGx System to analyze DNA samples from the National Institute of Standards and Technologies (NIST)?

BH: We'll be using the MiSeq FGx System for the NIJ grant studies. We'll be preparing the samples with the ForenSeq kit, running those libraries on the MiSeq FGx Instrument, and analyzing data with the ForenSeq Universal Analysis Software. We'll then compare our allele calling results to the NIST allele calls completed by CE-based typing, and confirm that the results are concordant between the MiSeq FGx System and other methods.

"It's amazing that we can use the ForenSeq Kit to amplify more than 200 loci at a single time."

Q: What are the next steps for deploying the MiSeq FGx System in your lab?

BH: After we complete our internal validation, we need to generate protocols and train our personnel. Then, it's likely we will use the MiSeq FGx System in our missing persons DNA program. We have some cases that have samples that yielded limited information, but not enough to make an identity statement. I suspect that the MiSeq FGx System will be very valuable there. I think it will also be a valuable resource on the criminal casework side, particularly in cold cases.

Q: You've been working in forensic science for 20 years. Did you ever think that there would be a time when technology would allow you to learn so much from a sample?

BH: When I began my career, forensic DNA testing was based on restriction fragment length polymorphism (RFLP) analyses that you'd perform by Southern blots. It was a very discriminating system, but yielded limited information. It took about 8 weeks or more to generate a complete profile. When STRs were introduced, we were

pleasantly surprised by the information they could provide. We've now gone so far beyond that. With the MiSeq FGx System, we can obtain information on multiple types of STRs, as well as SNPs and the phenotypic information they provide. I never envisioned that we would be able to do so much with such small samples and provide so much information to law enforcement agencies and to the criminal justice system.

"I never envisioned that we would be able to do so much with such small samples and provide so much information to law enforcement agencies and to the criminal justice system."

Q: How is NGS transforming forensic testing?

BH: In forensic testing, sometimes autosomal results will give you the best information and be the most discriminating, but sometimes not. If a sample is a mixture of two or more individuals, then that's where the Y-STR results can become valuable. You don't know all that before you start testing and it becomes difficult when you have a small sample and have to choose which tests to run.

The ForenSeq DNA Signature Prep Kit gives us the ability to obtain many kinds of information from just one test. It's amazing that we can use the ForenSeq Kit to amplify more than 200 loci at a single time. Instead of running multiple tests, we simply run a sample on the MiSeq FGx Instrument, and we get data on autosomal STRs, Y-STRs, X-STRs, and SNPs. With NGS, you don't have to decide prior to testing. You simply perform the test and use the best information to resolve a case, based on those comprehensive results.

References

 Proposition 69 (DNA). State of California Department of Justice Office of the Attorney General. oag.ca.gov/bfs/prop69. Accessed August 28, 2015

Learn more about the Illumina products and systems mentioned in this article:

- MiSeq FGx Instrument, www.illumina.com/systems/miseq-fgx.html
- ForenSeq Universal Analysis Software, www.illumina.com/systems/ miseq-fgx/software.html
- ForenSeq DNA Signature Prep Kit, www.illumina.com/products/ forenseq-dna-signature-kit.html

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

© 2015 Illumina, Inc. All rights reserved. Illumina, ForenSeq, MiSeqFGx, and the pumpkin orange color are trademarks of Illumina, Inc. and/ or its affiliate(s) in the U.S. and/or other countries. Pub. No. 1470-2015-004 Current as of 05 November 2015

