By Digging Deeper into the Genome, Next-Generation Sequencing May Yield More Forensic Clues

Dr. Bruce Budowle at the University of North Texas Health Science Center discusses how next-generation sequencing will transform forensic genomics.

Bruce Budowle, Ph.D., is Executive Director of the Institute of Applied Genetics at the University of North Texas Health Science Center (UNTHSC). Founded in 2009, the institute brings together a number of well-known scientists focused on refining genetic approaches to enhance several fields of study, including forensic DNA. Students enrolled in UNTHSC's forensic science program (1 of only 14 in the United States) also participate in the institute's forensic DNA program, where Dr. Budowle leads a team that balances case work and human identity testing with research and development into new technologies, such as next-generation sequencing (NGS). Their goal is to determine how best to apply and optimize NGS and other methods, and develop protocols that streamline their use in forensic analysis.

We spoke to Dr. Budowle about his thoughts on NGS and its potential impact on forensic genomics.

Q: What forensic DNA R & D efforts are being conducted by your team at the Institute?

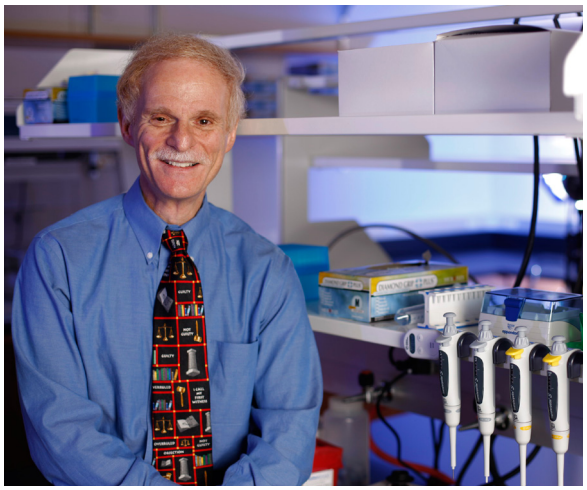
Bruce Budowle (BB): We're developing methods for better DNA recovery and removal of inhibitors from the environment that may affect our analyses. We're also investigating new genetic markers that will allow us to look deeper into forensic and human genetics questions, and identifying new technologies that will expand our capabilities in forensics. We're using a variety of new methods, from synchronous coefficient of drag alteration methods (SCODA) to extract, purify, and concentrate DNA from challenging forensic samples, to NGS technologies that can analyze those samples and help us retrieve the most information possible.

Q: Do you think that NGS will be transformational in forensics?

BB: I think it's going to be another transformational tool that will greatly benefit forensics. We've had a number of them over the years, from restriction enzyme digestion and Southern blotting to the polymerase chain reaction (PCR) that improved molecular biology across the board. Now we have new NGS tools that allow us to look at things in more depth and analyze many more samples in a parallel, multiplex fashion. NGS will also enable us to be more flexible in addressing the forensic challenges we face.

Q: Are there limitations in the current forensic workflow that NGS will alleviate?

BB: Our biggest gap in forensics is getting quality results from the limited information that we can obtain from low quantity, and often poor quality, DNA samples. NGS will enable us to obtain higher quality results from much smaller DNA samples.



Bruce Budowle, Ph.D., is Executive Director of the Institute of Applied Genetics at the University of North Texas Health Science Center (UNTHSC) where he leads a team of researchers in investigating forensic applications, such as next-generation sequencing.

The other limitation is the limits of our core set of markers. There are situations where having more genetic markers would help us answer a forensics question. For instance, when there's a blood stain found on the floor and we're making a comparison between it and a suspect's sample (reference sample), we're directly comparing genetic profiles. We're looking for a 1:1 correspondence of the evidence to the reference sample. In missing persons cases, we're performing kinship analysis, where we're looking at part of a genetic profile of an individual that may be shared with a potential or alleged relative. That means we don't have as much information to compare profiles. It's a little easier to compare first-degree relatives, such as a parent, offspring, a brother, or brother–brother–sister scenario. Identifying half-brothers or cousins, or comparing grandparents to grandchildren, becomes much more complicated. With the power of full genetic marker sets created using NGS, we're going to be able to answer kinship questions we could not even conceive of solving just a few years ago.

Q: How do you see the forensic laboratory using NGS?

BB: As I said earlier, one of the real benefits of NGS is that we'll be able to analyze more markers. We'll be able to apply NGS to reference samples where it will give us a lot more latitude on what we may be

able to type for forensic samples. So instead of driving all our assays towards the current core set of repetitive markers [standard tandem repeats (STRs)], we'll have an expansive set of markers that includes single nucleotide polymorphisms (SNPs), giving us the flexibility and versatility to drive analysis to those assays that can best answer the question at hand, whether it's kinship analysis, direct analysis, or something else.

There's nothing wrong with STRs, but SNPs may work better on degraded samples than repetitive elements that require a larger fragment of DNA. If I have a lab that wants to look at very challenged samples, I may opt for SNPs.

Additionally, NGS will give us more markers, allowing us to cover the entire forensic space and enabling laboratories in different countries to communicate better. For instance, if one lab likes working with 13 markers and another lab usually works with a different set of 20, the results can't easily be compared or used to identify a suspect. If we could simultaneously type a large set of markers for a reference sample that was inclusive of each country's chosen marker sets then we could communicate more effectively.

Q: How does NGS compare to Sanger sequencing?

BB: When we perform Sanger sequencing it's usually at 1× or 2× coverage. With NGS, we can reach 1000× coverage and see variation in things that we never saw before. An example would be mitochondrial DNA (mtDNA). Because of the replication fidelity of the mitochondrial genome, there's going to be some variation in our sequences, causing heteroplasmy or more than one mitochondrial species in an individual. Typically with Sanger, we may see just one or two species. We can deal with that level of variation because we've built whole systems around it. With NGS, we're going to see from 10 to 100 mtDNA variants because of its depth of coverage. New issues will arise from this and we'll have to draw the line somewhere on interpretation, but that's not a bad thing. It's just the fact that NGS technology offers superior depth of coverage.

Q: Illumina systems can simultaneously sequence large numbers of loci and multiple samples at once. What would that level of information provide?

BB: We're currently limited in our analyses because we need kits for each marker system and therefore separate analysis runs for each marker set. The nice thing about NGS is that all the markers are treated in a generic way, with the same chemistry used to obtain the results. Yes, you have to design your probes to identify your sites, but the technology is all the same, so it allows us to use a similar chemistry to address all the markers of interest, whether it's STRs, SNPs, Y-STRs, or mtDNA. That's going to be a real boon for forensics, because it means one analysis run will be able to provide a complete set of marker results.

Access to all those markers at once will benefit techniques such as familial searching, which is used to search profiles of evidence against a convicted felon database when the suspect's DNA profile does not generate a direct match. To find the potential source of evidence for investigative leads, we next look to see if there is anybody that might share a close profile with that of the evidence and possibly be a relative of the true source of the evidence. Currently, there are limitations to that type of analysis. Specialized kits are needed to look at Y chromosome and X chromosome markers. In the future, we want to have one technology that everybody can rely on, with all markers based on that technology. That's what NGS offers.

> "NGS will enable us to obtain higher quality results from much smaller DNA samples."

Q: Do you see value for law enforcement in having access to NGS results?

BB: Most crime laboratories provide information to law enforcement for investigative leads to refute or confirm a potential association of a person with the physical evidence. Whenever you can exclude somebody, that's really important. It helps the people who are wrongly associated and enables the case to move forward. Law enforcement now knows this is not the person that's the source of biological evidence and can look into other people that might be involved.

NGS will also benefit missing persons cases, identifying people and helping families find loved ones. Most missing persons or unidentified human remains cases are not just people who happened to walk off into the woods and die. The vast majority of them have been murdered, and finding and identifying the body is the beginning of a homicide investigation. So NGS technology will be beneficial in all sorts of ways. It will help the innocent, enable law enforcement to identify who might have committed crimes, and bring families of missing persons some resolution and clarity as to what might have happened. It also will enable the police to focus their resources on those cases that cannot be solved or directed by DNA.

Q: Do you see NGS enabling new capabilities in forensic genomics?

BB: There are a number of areas where NGS will be beneficial—areas that go beyond what we think of today as forensics. One that we're working on right now is combining pharmacogenomics with molecular autopsies. We know that a number of us have genes that can be beneficial, can cause problems, or pose risks in the presence of environmental insults or stresses.

There are a number of situations where it might be important to look at the genetic constitution of an individual to help determine the cause or manner of death. For example, there was a situation where a 13-day–old newborn died of morphine poisoning. Typically, when that occurs the police are going to investigate because it's an unusual death. They're going to first look into the parents, or anyone else who might have had the opportunity to do something to the child. They also might look at the hospital for negligence. In this particular case, the mother was given codeine after an episiotomy during birth. It was determined that she was an ultrametabolizer of

codeine through sequencing, where it was identified that she had a double dose of a particular gene. Since codeine metabolizes into morphine, the child was receiving morphine through the mother's milk. That information took the investigation in a totally different direction, because we had strong evidence that the manner of death was an accident and not an intentional homicide. With NGS, we can sequence a large number of genes at one time and look objectively at all the variants that may be there, providing us with information that could help determine the cause of death in challenging forensics cases.

Q: The Combined DNA Index System (CODIS) contains DNA profiles contributed by federal, state, and local forensic laboratories in the United States. How do you see CODIS evolving in the near term?

BB: CODIS needs to start preparing its system to better accommodate NGS data. CODIS is currently fixed on its core markers and everything outside that core is not well addressed in its database. With NGS, we may see in the near future 30, 40, 50, or 200 markers added that will increase our ability to search for portions of those markers against a reference set. Most people are driven by the ability to upload things into CODIS. So in the future, if data from NGS systems can't be uploaded into CODIS, it will slow down the database's evolution. That's why it's so imperative for CODIS to start thinking about how to accommodate the markers NGS can identify and put it into their builds for the next version.

> "In the future we'd have one technology that everybody could rely on, with all markers based on that technology. That's what NGS offers."

Q: How can NGS be used for bioancestry?

BB: There are markers that have a higher frequency in one population as opposed to another population. If you see one variant you might be able to infer something about the bioancestry of that individual. With enough markers, population affinity can reach high probabilities. There is tremendous value for us to find more of those markers for our missing persons work. Currently, when law enforcement finds human remains and a skull is located, we give that to an artist who reconstructs the skull and provides us with a drawing of the deceased. Sometimes the drawing is amazingly identical to the individual and other times it's not even close.

In the phenotype area, where we're trying to figure out what a person looks like, a lot of work has been done in determining pigmentation–hair color, eye color, skin color. We now want to focus on markers associated with more specific attributes, such as the shape of the nose, the ears, the lips, the soft tissues, maybe curly versus straight hair. These features might be enough to trigger someone's memory in recognizing the reconstructed individual and would enable us to move forward with a traditional DNA kinship analysis. I think that's the real value of NGS in bioancestry marker discovery.

One other area that might be of interest is microbial forensics. We have a history of microbial forensics dating back to the days of the anthrax letter attack and we are continuing that biosafety and biosecurity work. I think the tremendous depth of coverage that NGS can provide will benefit metagenomics studies that can be performed on the human microbiome, on challenging samples, and on potential threats. It's the ideal technology for that.

Q: What other benefits will NGS provide?

BB: NGS will enable a lot more of our analyses to be automated and done with software as opposed to an individual looking at data, verifying, and a second person looking and verifying. We'll rely more on the semi-expert software system capability to work out what markers are there with quality scores and other bits of information to give us confidence in the results. Because that will be done behind the scenes in the computer, it will improve our throughput capabilities.

Learn more about Forensic Genomics at www.illumina.com/forensics

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

FOR RESEARCH USE ONLY

Pub. No. 773-2012-031 Current as of 09 October 2012

illumina®