BioTechniques
The International Journal of Life Science Methods

# THE FORENSIC GENOMICS TOOLBOX IS EXPANDING

The human genome project (HGP), a decade-long multibillion dollar effort, provided the first-draft human genome sequence in 2001 [1,2]. This draft genome was the beginning of a revolution in the biological sciences, with a focus in medical prognostics and diagnostics and drug therapies, as well as studies of human evolution. While many of the health-related predictions that should come from the HGP are yet to be realized, the value of being able to sequence an entire human genome has outstripped expectations in the field of forensic human identification. In the subsequent two decades since that first human genome sequence, sequencing technology has become democratized with the advent of desktop massively parallel sequencing (MPS) instruments [3]. MPS has become increasingly user friendly and robust, with sequencing times and costs plummeting dramatically such that the seemingly elusive goal of sequencing an entire human genome for $1000 in a day is now within reach. With these features, sequencing an entire genome from biological evidence found at a crime scene is now a viable enterprise for solving crimes and developing investigative leads.

## EXPANDING DNA TECHNOLOGY FOR FORENSIC HUMAN IDENTIFICATION

Human identification (HID) via DNA typing primarily uses 20–30 polymorphic short tandem repeat (STR) markers [4] for analysis of biological samples from, for example, crime scenes, mass disasters, parentage testing, and missing persons and human trafficking cases. Attributing the possible source of biological evidence in criminal cases largely involves direct comparison of an evidence sample DNA profile (from blood, semen, saliva, etc.) with a reference sample DNA profile taken from a person of interest. Missing persons identification makes use of the same STR markers, but the comparison typically is indirect (i.e., kinship analysis), based on the observed genotypes of the remains and those of family member reference samples to determine support for a proposed family relationship. Kinship analyses in forensic genomics have been used almost exclusively for parentage testing, missing persons and human remains cases, and typically rely on comparisons with first-order relatives (i.e., parent, child, siblings). Family reference samples from second-order and beyond relatives are not sufficiently informative with the current limited number of STRs (~20–30) used in routine casework. An additional challenge to DNA typing is that evidence (both crime scene and human remains) is exposed to environmental insults and thus can contain limited quality and quantities of DNA. The DNA can be damaged to such a degree that only a partial genetic profile is obtained (or no profile at all), making identification difficult even with highly sensitive and robust DNA analysis methods.

Investigative leads have relied on direct comparisons and searching of forensic DNA databases, such as the Combined DNA Index System (CODIS) [5]. Kinship approaches do not require the actual donor of an evidence sample to be in the DNA database and thus can extend the reference pool to assist in identification. If a relative(s) of the donor of the evidence sample is in the database, it is potentially possible to identify the source of the evidence by subsequently searching public records.

The primary ways to improve the utility and accuracy of kinship analyses are: 1) obtain the most informative reference family ▶

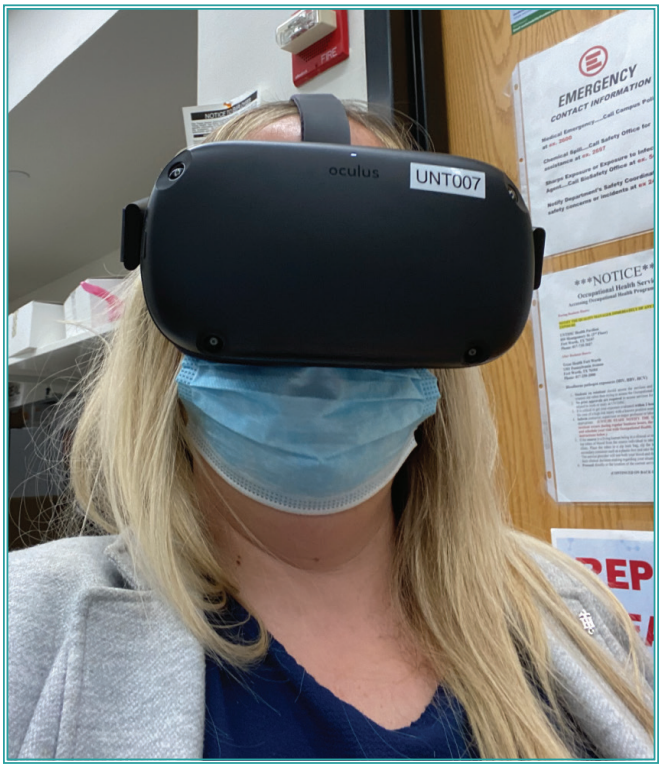


**Figure 1. A member of the Center for Human Identification faculty wearing a virtual reality headset for forensic training purposes.**

members [6]; and 2) increase the number of genetic markers used. The infrastructure built and the innovation fostered by the HGP effort have brought to bear novel technologies that can facilitate kinship analyses to identify sources of crime scene evidence and human remains. One such technology, massively parallel sequencing (MPS), offers automated high throughput (millions of genetic markers per sample and multiple samples analyzed simultaneously), robustness and requires no more sample consumption than that of current forensic methods. MPS and its supporting methodologies: 1) permit analysis of a substantially larger number (tens of thousands to millions) of genetic markers per sample, which in turn extends kinship analyses to be able to use more distant family reference samples (up to 4th degree and beyond) to assist in determining the origin of a sample; 2) increase the sensitivity of detection, which enables analysis of lower quantities of template DNA and highly degraded samples; and 3) extend the application of kinship analysis beyond traditional parentage and missing person cases to criminal cases, such as sexual assault.

Some forensic laboratories already have validated and/or implemented whole mitochondrial genome sequencing via MPS [7,8]. For the past 25 years, forensic mitochondrial DNA typing relied on Sanger sequencing to analyze the polymorphic hypervariable (HV1 and HV2) regions. Whole mitochondrial genome sequencing provides a 400% increase in detection of genetic variation, while concomitantly being less costly and less labor intensive [9].

The forensic DNA laboratory often does not have the luxury to increase the number of family members nor select the best, most informative family members for HID purposes. Forensic genetic genealogy combines genome-scale DNA testing and genealogical surveying to determine biological relationships between individuals (and can be used to determine the bioancestry of an individual), which may be used to help investigations (see reviews [10,11]). Indeed, studies have shown that using dense single-nucleotide polymorphism (SNP) data, with hundreds of thousands to millions of markers, enables detection of up to possibly 9th order relatives for HID [12–14]. MPS (and whole-genome SNP arrays) has assisted in solving hundreds of cold cases [15], the most noted being the Golden State Killer [16]. The fruits of the HGP clearly show that genomic advances are impacting directly missing persons and unresolved cases. Further reductions in cost will likely see more widespread adoption of forensic MPS methods as routine HID tools.

## IMPROVING ANALYSIS OF DEGRADED DNA

Forensic genomics innovation, which is built on the foundation of MPS, has improved the ability to analyze low-quality DNA samples. An example of a newly emerging method is reverse complement PCR (RC-PCR) [17]. RC-PCR is a target enrichment, library preparation method performed in a single, closed-tube assay. The benefits are less sample manipulation, fewer chances for contamination, and faster turnaround time for library preparation, all important features for facilitating transfer of MPS tools to the operational forensic laboratory. The other feature of RC-PCR and subsequent MPS is that targeted markers, such as SNPs, can be contained in relatively small amplicons. The smaller an amplicon is, the more likely a genotype can be derived from degraded samples. Indeed, the IDseek® SNP85 panel (Nimagen) contains 85 commonly used identity testing SNPs with amplicon sizes ranging from 48 to 128 base pairs (bp), and 85% of these SNP containing amplicons are <100 bp [18]. RC-PCR can also accommodate STRs and mitochondrial DNA. Thus, the forensic analyst's toolbox is growing, making it likely that results can be obtained from challenged samples that just a few years ago were seemingly untypable.

## EDUCATION/TRAINING: VIRTUAL REALITY

Significant effort and resources are dedicated to developing and validating the methods in the field of forensic genomics. Often ignored is the equally important need for training and education. Forensic scientists must have knowledge of the theory and procedures of new and existing technologies to properly implement them and more so to properly interpret the data generated. After all, the consequences of forensic results can impact a person's life and liberty, victims, victim's families, communities, and the government. The rapidly developing use of molecular genomics in the forensic field places a demand on suitable education and training for current employees as well as students who aspire to be forensic scientists. Traditionally education and training in forensic genomics rely heavily on wet lab analyses, which is labor intensive, time consuming and costly. The instrumentation and consumables required are expensive and involve specialized handling and storage. Hands-on training using these reagents and equipment often necessitates repetitive in-house instruction, which can be difficult when the instruments in some laboratories are fully committed to operational use. Most trainers in forensic laboratories have other duties as well,

which limits their time to instruct or creates backlogs when taken away from casework for training purposes. Since most forensic laboratory training is in house and not standardized, there can be significant variation in the quality and amount of training experiences across the country and the world. In order to improve training, enhance the overall quality and defray some of the demands on time and budget, other avenues for education and training must be explored.

Almost all industries, including education, have been altered substantially by technology. Desktop and laptop computers, tablets and smartphones have transformed how people learn and gather information. Virtual reality (VR), a technology that offers exciting and novel opportunities, can be the next tool to improve forensic genomics education and accommodate the burgeoning demand to gain expertise in molecular biology, genetics, bioinformatics and laboratory workflow. The decrease in cost of software development as well as equipment (either headset or computer/tablet) now makes VR a viable option for the education sector, much like the decrease in cost made sequencing accessible to not just research institutes but also service laboratories.

VR (herein generically refers to the range of 'reality' tools) creates an artificial or simulated environment in which the user is immersed in an experience, and the physical world is excluded [19]. Learning experiences can be made easily accessible to students with repeated opportunities to practice the task(s) so they can hone skills before venturing onto delicate instrumentation and consuming costly reagents. Although the cost of $100,000–$250,000 may seem high to develop a module, the cost benefit long term is immense. Modules can be developed which can be accessed by the thousands of forensic analysts worldwide, which makes for a very cost-effective approach to education and training, particularly in fields in which hands-on experiential training is required. Moreover, VR modules can overcome the variance of in-house training, since all students can receive standardized, quality training as well as improve training of the trainers (another overlooked need for technology education and transfer) [20]. VR can reduce the amount of time these individuals are required to train. Thus, they are less encumbered and can meet their other duties as assigned. Additionally, VR with its collaborative features, active learning environment, peer interaction and peer feedback can and will enhance proficiency-based training.

There are standard practices to which forensic laboratories are held accountable, and each of the procedures used has documented knowledge and practical requirements. Even seemingly simple practices, such as proper pipetting, are critical to obtaining reproducible quality results, and training in all aspects is requisite to ensure a skill set(s). The repetitive, standardized practicing of a skill in a VR environment will encourage innate muscle memory. Indeed, our Center is developing VR modules for procedures such as DNA extraction, PCR amplification, DNA sequencing, etc. for a more effective approach to molecular biology training. These VR modules will place students into a virtual reality laboratory to experience each step of the workflow, repeatedly, by utilizing virtual laboratory equipment, reagents and samples. The training modules can be utilized on a standalone headset designed for the VR experience or on PC computers offering flexibility and various modalities to meet different training environments and situations.

## CONCLUSION

Technology spawned from the HGP is being embraced in some forensic science sectors. The analysis of DNA, from whole-genome sequencing to targeted highly compromised fragments, will likely be a significant part of the future forensic genetics toolbox. However, societal and policy issues still need to be addressed – from collection to dissemination of information. The concerns of privacy, data availability, data security, equity, when it is appropriate to use comprehensive genetic analyses, surreptitious sample collection and collecting samples from relatives to reduce the labor demand on the genealogy portion of an investigation are among issues that need to be considered. Already the state of Maryland has passed legislation defining how forensic genetic genealogy may be applied in criminal investigations [21].

The success of identifying human remains and solving heinous, serious crime suggests that advanced technologies and comprehensive genomic analyses are here to stay. The possibilities are almost endless and offer a very exciting future for forensic genomics and the expansion of knowledge and training through efficient and innovative technologies such as VR.

**Written by Bruce Budowle[1,2], Magdalena M Bus[1,2], Lynley Dungan[1] & Angela van Daal[1]**

[1]Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA

[2]Department of Microbiology, Immunology, and Genetics, Graduate School of Biomedical Sciences, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA

## REFERENCES

1. Lander ES, Linton LM, Birren B *et al.* Initial sequencing and analysis of the human genome. *Nature* 409(6822), 860–921 (2001).
2. Venter JC, Adams MD, Myers EW *et al.* The sequence of the human genome. *Science* 291(5507), 1304–1351 (2001).
3. Quail MA, Smith M, Coupland P *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13, 341 (2012)
4. Hares DR. Expanding the CODIS core loci in the United States. *Forens. Sci. Int. Genet.* 6(1), e52–e54 (2012).
5. CODIS. www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet
6. Ge J, Budowle B, Chakraborty R. Choosing relatives for DNA identification of missing person identification. *J. Forens. Sci.* 56(Suppl. 1), S23–S28 (2011).
7. Marshall C, Sturk-Andreaggi K, Daniels-Higgenbotham J *et al.* Performance evaluation of a mitogenome capture and Illumina sequencing protocol using non-probative, case-type skeletal samples: Implications for the use of a positive control in a next-generation sequencing procedure. *Forens. Sci. Int. Genet.* 31, 198–206 (2017).
8. Cuenca D, Battaglia J, Halsing M, Sheehan S. Mitochondrial sequencing of missing persons DNA casework by implementing Thermo Fisher's Precision ID mtDNA Whole Genome Assay. *Genes* 11(11), 1303 (2020).
9. King JL, LaRue BL, Novroski N *et al.* High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forens. Sci. Int. Genet.* 12, 128–135 (2014).
10. King D, Phillips C Kennett D, Tillmar A. Investigative genetic genealogy: Current methods, knowledge and practice. Forens. Sci. Int. Genet. 52, 102474 (2021).
11. Ge J, Budowle B. Forensic investigation approaches of searching relatives in DNA databases. *J. Forens. Sci.* 66(2), 430–443 (2021).
12. Henn BM, Hon L, Macpherson JM *et al.* Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS ONE* 7(4), e34267 (2012).
13. Ramstetter MD, Shenoy SA, Dyer TD *et al.* Inferring identical-by-descent sharing of sample ancestors promotes high-resolution relative detection. *Amer. J. Hum. Genet.* 103(1), 30–44 (2018).

14. Staples J, Witherspoon DJ, Jorde LB *et al.* PADRE: Pedigree-Aware Distant-Relationship Estimation. *Amer. J. Hum. Genet.* 99(1), 154–162 (2016).
15. Greytak EM, Moore C, Armentrout SL. Genetic genealogy for cold case and active investigations. *Forens. Sci. Int.* 299, 103–113 (2019).
16. Kaiser J. We will find you: DNA search used to nab Golden State Killer can home in on about 60% of white Americans. Science (2018). https://doi.org/10.1126/scien ce.aav7021.
17. Kieser RE, Buś MM, King JL *et al.* Reverse complement PCR: a novel one-step PCR system for typing highly degraded DNA for human identification. *Forens. Sci. Int. Genet.* 44, 102201 (2020).
18. Buś MM, de Jong E, King JL *et al.* RC-PCR, an innovative and effective method for multiplexing forensically relevant SNP marker systems. *BioTechniques* 71(3), 484–489 (2021).
19. Piovesan SD, Passerino LM, Pereira, AS. Virtual reality as a tool in the education. Presented at: IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2012). Sampson DG, Spector JM, Ifenthaler D, Isaias P (Eds). 295–298 (2012). http://www.iadisportal.org/digital-library/virtual-reality-as-a-tool-in-the-education
20. Pantelidis VS. Reasons to use virtual reality in education and training courses and a model to determine when to use virtual reality. In: Themes in Science and Technology Education. Klidarithmos Computer Books, Special Issue, 59–70 (2010).
21. Criminal Procedure – Forensic Genetic Genealogical DNA Analysis, Searching, Regulation, and Oversight, HB0240, Aricle II, Section 17(c) of the Maryland Constitution – Chapter 681, effective date(s): 1 October 2021. https://mgaleg.maryland.gov/mgawebsite/Legislation/Details/hb0240