Forensic Technology Center of Excellence

Massively Parallel Sequencing:
Application to Forensics

Guidance Document

December
2016

Principal Investigator:
Jeri Ropero-Miller, PhD, F-ABFT
FTCoE Director
JeriMiller@rti.org

NIJ Contact:
Gerald LaPorte, MSFS
Office of Investigative and Forensic Sciences Director
Gerald.Laporte@usdoj.gov

NIJ | National Institute of Justice
STRENGTHEN SCIENCE. ADVANCE JUSTICE.

RTI INTERNATIONAL

DISCLAIMER

Information provided herein is intended to be objective and is based on data collected during primary and secondary research efforts available at the time this report was written. The information provided herein is intended to provide an overview and guide of massively parallel sequencing (MPS) applications; it is not intended as an exhaustive summary. Experts, stakeholders, and practitioners offered insight related to the application of MPS to forensics and crime laboratories.

NIJ Award Number 2011-DN-BX-K564.

Technical Contacts

Patricia Melton, PhD
pmelton@rti.org

Bruce Budowle, PhD
bbudowle@unt.edu

Jonathan King, PhD
jking@unt.edu

NIJ Award Number 2011-DN-BX-K564

The Forensic Technology Center of Excellence (FTCoE)

The FTCoE is a collaboration of RTI International and the following academic institutions, which are accredited by the Forensic Science Education Programs Accreditation Commission (FEPAC): Duquesne University, Virginia Commonwealth University, and the University of North Texas Health Science Center. In addition to supporting NIJ's research and development (R&D) programs, the FTCoE provides testing, evaluation and technology assistance to forensic laboratories and practitioners in the criminal justice community. NIJ supports the FTCoE to transition forensic science and technology to practice (award number 2011-DN-BX-K564).

FTCoE is led by RTI, a global research institute dedicated to improving the human condition by turning knowledge into practice. With a staff of more than 4,700 providing research and technical services to governments and businesses in more than 58 countries, RTI brings a global perspective. FTCoE builds on RTI's expertise in forensic science, innovation, technology application, economics, data analytics, statistics, program evaluation, public health and information science.

Contents

Appendices

List of Figures

List of Tables

EXECUTIVE SUMMARY

Massively parallel sequencing (MPS), also called next-generation sequencing, is an exciting technology that holds promise for enhancing the capabilities of forensic DNA laboratories. However, several challenges confront the implementation of an MPS system in a crime laboratory. These challenges include laboratory methodology and validation, training and education on the fundamentals of the technologies and chemistries, functionality, genetic marker systems, interpretation guidelines, policy and data procedure developments related to Combined DNA Index System (CODIS) operations, and perceived admissibility and privacy issues. The final outcome of a series of webinars hosted by the Forensic Technology Center of Excellence (FTCoE) is to provide a technical resource document for forensic DNA scientists that covers the fundamentals of the current platforms and chemistries, the capacity and throughput of genetic marker analysis, bioinformatics and validation requirements, potential applications, and potential admissibility issues related to implementing an MPS system in a crime laboratory.

Capillary electrophoresis (CE)–based technology is a fluorescent detection-based platform that allows routine typing of short tandem repeat (STR) markers, which are the primary markers used in human genetic identity testing. In addition, a few laboratories analyze short target sequences of the mitochondrial (mtDNA) genome that provide exquisite sensitivity of detection with challenging samples. Lastly, in some instances, the forensic science community has used single nucleotide polymorphisms (SNPs), which are markers well suited for analyzing degraded samples. The CE, both robust and reliable, has been the standard method for human identity typing applications for more than 15 years. However, as with any technology, limitations exist. The recognized limitations of the CE platform include the methodology's resolution, scalability, and throughput. The extent of these limitations became apparent only with the advent of MPS technology, which, in comparison, excels in performance in these areas.

MPS has the potential to address the previously listed limitations and expand allelic typing success, which therefore improves investigative capabilities with DNA evidence. MPS has a substantially higher throughput, allowing simultaneous typing of larger sets and different types of genetic markers such as autosomal, Y-chromosome, and X-chromosome STRs; identity and ancestry informative and phenotypic SNPs; and even the entire mtDNA genome. MPS technology can provide opportunities to substantially improve current practices and point to opportunities for advancement beyond current DNA typing capabilities. The forensic science community can apply this technology to the characterization of the same wide range of biological evidence as analyzed by CE, and it can expand forensic services to include animal and plant forensic genetics, microbial forensics, and molecular autopsy, which may require direct sequencing of a number of genes to be effective. In addition, MPS may improve the

interpretation of complex mixtures, an application with which the DNA forensic science community has struggled.

To accomplish the task of familiarizing the forensic science community with issues surrounding MPS technology, the National Institute of Justice's (NIJ) TCoE at RTI International, in partnership with the University of North Texas Health Science Center's (UNTHSC) Institute of Applied Genetics, facilitated a webinar series to discuss the general protocols, considerations, and objectives of MPS technology to inform and educate crime laboratories considering early adoption.. Discussions and presentations addressed challenges and potential barriers to implementation such as obtaining a basic understanding of the chemistry and functionality of MPS, designing validation studies for MPS, developing policies/procedures to address the range of markers available with MPS including CODIS considerations for MPS data, and addressing privacy issues. The data generated by MPS systems have the potential to be compatible with current CODIS requirements, and as such, the technology itself is not a major issue for CODIS adoption. However, in order to benefit from the additional capabilities that the MPS systems provide, CODIS indices would need to be modified to accommodate the additional loci attributes, as well as the potential changes in nomenclature.

This document captures the views of international and national scientists involved in developing and validating MPS, the literature to date, and the experience of individuals fluent in both CE and MPS applications. The project team used the webinar series model to collect information, engage open discussion regarding MPS across multiple stakeholders with variant points of view and experiences, obtain the current landscape of this dynamic technology, and address the potential barriers to adoption. This work may be used as a guide to the derivation of policies and procedures that assist with implementation should an agency choose to adopt MPS technology for the analysis of forensic DNA samples.

1. INTRODUCTION

For 30 years, forensic genetics has employed DNA-based molecular biology tools for human identity testing and, more recently, for nonhuman identity applications.[1–18] However, the combination of high sensitivity, specificity, and resolution was not achieved until the technology combined the enrichment of specific genetic targets, or markers (e.g., short tandem repeat [STRs]), by the polymerase chain reaction (PCR) with the subsequent detection of the fluorescently labeled amplicons by capillary electrophoresis (CE). Over the years, the continued improvement of this combination of technologies has substantially advanced forensic DNA analysis. Currently, laboratories can analyze extremely minute quantities of DNA, and, in some cases of multiple-donor DNA evidence samples, laboratories can reduce the number of potential contributors to specific individuals. CE has a well-developed workflow that is suited for relatively high sample throughput in a semiautomated fashion. CE-based typing methods can analyze biological evidence from criminal cases; the biological evidence also has been the platform that has generated the millions of DNA profiles that are housed in forensic DNA databases worldwide.[19–21]

The specific aims of this project are to provide:

- a current informational source on massively parallel sequencing (MPS) technology, procedures, capabilities, and potential applications through a live webinar series, which will then be archived for future viewing; and
- a guidance document that may be used to facilitate laboratories that are considering investing resources and implementing an MPS system.

Although CE-based methods have been the "gold standard" in forensic genetics capabilities for years, advances in science and technology continue to expand the boundaries and produce methodologies that can replace current practices and augment the capability of a forensic DNA laboratory. The most recent and exciting technology development in this regard is massively parallel sequencing (MPS). MPS has the ability to enable higher sample throughput with a larger set of genetic markers for characterization of biological evidence from humans, animals, plants, and microbes, in a semiautomated or an automated fashion, all of which appeals to the forensic science community. This greater capacity will extract more genetic information from a sample than is possible with current CE platforms and opens the door to greater possible applications. The outcome is the potential to provide more investigative leads, to associate individuals with biological evidence, exclude those falsely associated with biological evidence, enhance kinship analyses in missing persons and mass disaster cases, help determine cause and/or manner of death of unexplained or unexpected deaths by autopsy, support microbial forensic investigations, and identify applications related to animal and plant materials.

The CE methodology characterizes STR alleles by measuring the length of the PCR amplicons. The length-based approach determines, in an operationally defined manner, the number of repeats that comprise an STR allele(s). However, mass spectrometry and Sanger sequencing have established that sequence variation exists within the population for some length-based alleles of STRs.[22–24] The forensic science community has used Sanger sequencing on the CE platform to sequence the hypervariable regions of the mitochondrial (mtDNA) genome and is the methodology of choice for this type of forensic analysis; however, this methodology is labor-intensive, time-consuming, and relatively expensive.[15] Furthermore, although the hypervariable regions of the mtDNA genome contain a concentration of genetic variation, significantly more variation exists in the coding region of the genome.[25] Sequencing the entire mtDNA genome or even the coding region is not practical on the CE platform. In addition, single nucleotide polymorphisms (SNPs) are valuable for characterizing biological evidence due to their short amplicon size, which is a feature that makes them well suited for analyzing degraded or compromised samples. The lack of a stutter and high mutation rate, factors that otherwise can complicate STR data interpretation, are not associated with SNPs. Although these features would seem to promote the use of SNPs, the CE platform is unable to effectively distinguish SNPs because SNPs effectively have the same size. Substantial sample preparation methods are required to enable SNP detection on the CE platform, and these approaches are typically not quantitative and are labor-intensive.

In contrast, MPS detects the actual sequence of the targeted DNA marker and thus is able to (1) detect variation within length-based STR alleles in both the repeats and in the flanking regions, (2) sequence the entire mtDNA genome in a relatively straightforward approach requiring no more effort than any other marker analyzed with MPS, and (3) analyze SNPs with the same sample preparation and workflow used for any other marker.[26–34] In addition, the high throughput of MPS makes possible the analysis of a larger battery of genetic markers than that with CE. In CE-based analyses, the amplicons, labeled with the same fluorescent dye molecule at each locus, must vary in size in order to assign alleles to the appropriate locus. In contrast, with MPS analysis, size separation is no longer a requirement. Therefore, the forensic science community can simplify panel designs and substantially reduce the amplicon size of some markers. The MPS platform allows for various types of STRs (e.g., autosomal, Y-chromosome, and X-chromosome) and various types of SNPs (e.g., identity, ancestry, linage, and phenotype) to be analyzed separately or in one panel. An increased number of identity markers and lineage markers would facilitate the characterization of evidence and promote alternate ways to develop investigative leads such as with familial searching.[35]

The most demanding challenge in identity testing is the interpretation of complex mixtures. Forensic crime laboratories struggle with complex mixture interpretation to the extent that sometimes

errors in interpretation occur or data are not used to full capability, potentially hindering investigative leads. Although MPS technology will not rectify all of the issues associated with complex mixture interpretation, the technology can generate informative genetic data that may improve complex mixture interpretation by reducing the number of possible explanations that constitute a particular mixture.

2. MPS INSTRUMENTATION

The sequencing of the first human genome took more than a decade, required more than 40 genome center institutions, and cost billions of dollars.[36] Given such demands, it was inconceivable that large-scale sequencing would become part of the repertoire of the crime laboratory. However, slightly less than a decade ago, benchtop, high-throughput massively parallel sequencers became commercially available.[37–44] These MPS instruments can be characterized as technologies that now make it possible for just about any laboratory to carry out genome-size sequencing in a rapid and cost-effective manner. Massively parallel means that laboratories can perform millions of sequencing reactions simultaneously in a single instrument run. The amount of sequencing that laboratories can perform with MPS in a matter of hours to a few days would take months to years with the CE platform.

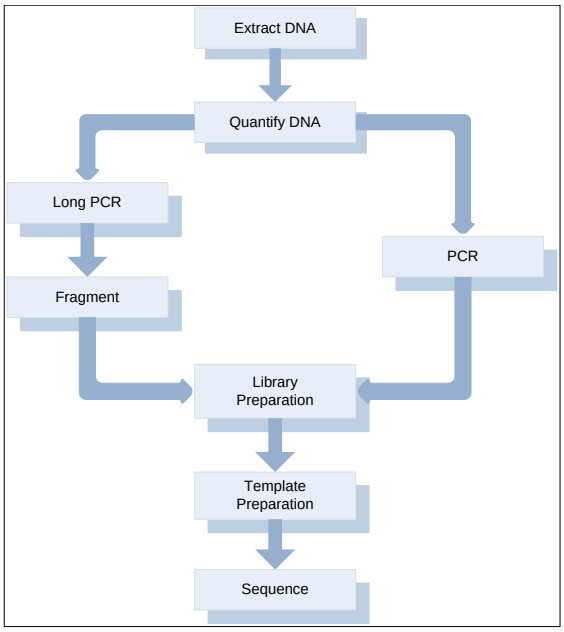
Several MPS platforms are available that provide substantial sequencing capacity in a reduced footprint. The primary benchtop MPS platforms include the MiSeq (Illumina), Ion Torrent Personal Genome Machine® (PGM) and Ion Proton™ System (Thermo Fisher Scientific), and the discontinued GS FLX system (Roche). The Illumina NextSeq 500 system and Ion Torrent S5 are the latest benchtop platforms to become commercially available with increased throughput and simplified formats, all indicating that MPS throughput will continue to increase and cost will continue to decrease.

3. MPS WORKFLOW

The fundamental steps of an MPS workflow share similarities, particularly on the front end, with the current CE platform. Therefore, the expertise gained from current work applies equally to the functionality and interpretation of MPS. ***Figure 1*** shows a general MPS workflow from DNA extraction through data analysis. There is commonality between the MPS and CE workflows at the DNA extraction, DNA quantitation, and PCR steps. In the CE workflow, PCR enriches the target for subsequent analyses; this same enrichment strategy is the primary approach used for MPS. In the MPS workflow, an alternate enrichment strategy—known as capture—can also be employed. The capture process uses probes that target and isolate specific regions of the genome. Typically, relatively large DNA input requirements for capture proved to be a limitation and thus capture would likely only be used for reference sample typing or database applications.. However, recent studies to date show that PCR enrichment and subsequent MPS can provide detection sensitivity that rivals or exceeds that of CE.[32,34]

In most MPS processes, the DNA fragments need to be less than a few hundred bases. Therefore, when appropriate, it may be necessary to fragment the DNA of long amplicons.[45] With long PCR-generated amplicons, such as those generated for typing mtDNA genomes from reference samples, laboratories must perform fragmentation to reduce the size of the DNA molecules to be better suited for sequencing.[46–49] With short PCR-generated amplicons, fragmentation is not required. This PCR step is considered the first enrichment step of the process.

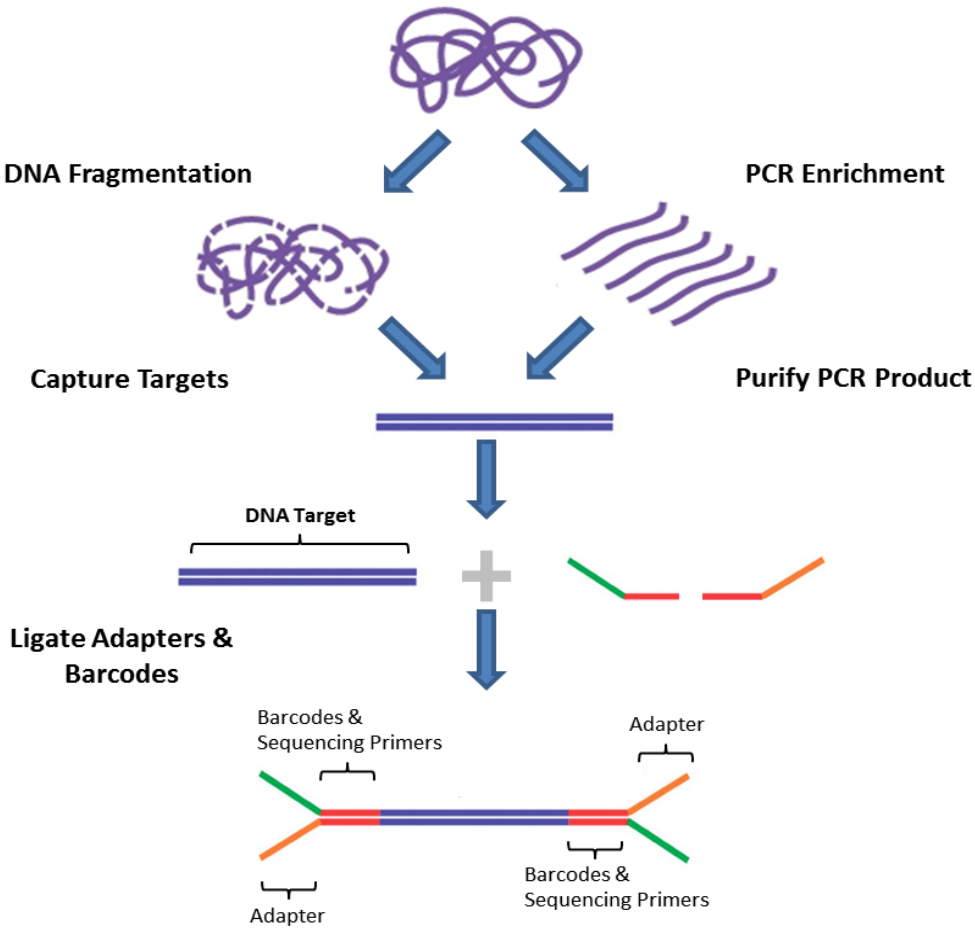
Figure 1. Depiction of the General MPS Workflow from Sample to Result



Note: The initial steps of extracting DNA, determining the quantity of DNA recovered from a sample, and amplifying the DNA through PCR are common to MPS and CE. Thus, the forensic DNA typing community already has experience with a portion of MPS methodology. Library preparation and template preparation are procedures to format the DNA molecules (derived from PCR products) for sequencing. Sequencing determines the bases and their order in DNA fragments. The resultant sequences, and hence the genetic markers, are interpreted in a similar manner as DNA evidence from a CE analysis.

Once analysts have enriched the targets, they conduct library preparation. Library preparation adds specific, short sequences to the ends of the DNA fragments in preparation for sequencing.[50–56] Typically, this process attaches adapters and universal primer sequences to both or either of the 5 and 3 prime ends of the DNA fragments. Adapters are short sequences that allow library-generated fragments to attach to a solid support for cloning of each individual DNA fragment. Priming sequences allow for annealing of primers to initiate sequencing.[37,40,57] Template preparation creates clones of individual target fragments. In a sequence-by-synthesis fashion, the instrument translates signals (with the two most common platforms by change in pH or by fluorescence) of each fragment of the library into sequence data (see ***Figure 2***).

Figure 2. Depiction of a General Library Preparation Workflow



Note: Short DNA fragments are processed (or modified) such that short pieces of DNA—known as adapters, primer binding sites, and barcodes—are added. These tags facilitate sequencing of the DNA molecule.

Multiplexing, the ability to type more than one marker at a time, is extremely important as it reduces consumption of often limited evidence, reduces labor, minimizes the chances of contamination or sample mix-up, and most importantly, provides more genetic information per analysis. The same feature of marker multiplexing on the CE platform is accommodated by MPS. However, due to its high throughput, MPS also can enable multiplexing of *samples*. A process called barcoding allows the interpretation software to distinguish among the different samples sequenced in one analysis. Barcodes are short, specific sequences attached to the ends of the fragments of each sample that tag and individualize the DNA fragments specific to each sample. After sequencing, an individual can parse, or demultiplex, the data bioinformatically based on the distinguishing barcodes. Although it is feasible to generate hundreds to thousands of barcodes, an individual can currently pool anywhere from 12 to 96 different samples using commercially available, application-specific preparation kits.[58,59]

Currently, two MPS platforms—MiSeq and PGM—are the focus of forensic research and validation. Both platforms can yield reliable results.[26–34] The MiSeq uses a flat, solid support—called a flow cell—to perform clonal amplification of each DNA fragment through a patented process known as bridge amplification. The process generates millions of clusters (or clones) of target DNA fragments randomly distributed across the flow cell.[42,44,60,61] The instrument uses sequencing-by-synthesis chemistry to sequence the clusters simultaneously and detects the various bases of the sequence by fluorescence of the four distinctly labeled nucleotides in a similar strategy as that of terminator chemistry employed by Sanger sequencing.[62,63] PGM uses a different strategy known as emulsion PCR (ePCR) for cloning and semiconductor chip technology instead of the fluorescent dye and laser optics system that is used by Illumina and in the CE platform. The ePCR method clones the short fragments attached to beads within a micelle. As a result, ePCR generates millions of beads, or microbeads, containing clones of individual fragments.[40,41,64] Each of these microbeads is placed into individual microwells in a sequencing chip, and the actual sequencing occurs on each microbead contained in these individual microwells of the sequencing chip. Using a sequence-by-synthesis strategy, the instrument detects a specific nucleotide by a change in pH.[41] Although both platforms provide reliable sequencing results, the PGM chemistry has more difficulties in accurately sequencing homopolymers compared with Illumina-based chemistry.[34,37,65,66] Recent advances using the Hi-Q™ polymerase can increase sequencing success through homopolymer stretches.[67] Both platforms can sequence fragments up to 400 bases in length, which should be sufficient for sequencing most forensic STR alleles.

As has always been the case with forensic DNA typing, the major and critical components of MPS are data analysis and interpretation of results. MPS generates massive amounts of data that have not been encountered previously in forensic identity testing. Therefore, bioinformatics is integral for the analysis of the sequence data. Accuracy and reliability of sequence data rely on a number of factors such as base calling, alignment or flank selection strategies, depth of sequence coverage, heterozygote allele balance (allele coverage ratios), strand balance (or only using one strand), misincorporation rates, read length, size selection method, and other chemistry-specific issues (see ***Table 1***). There are a number of software tools available to analyze sequence data.[27,68,69,70,71] In addition, commercial software products are available that typically support kits developed by the manufacturer. The requirement and parameters of these software and features, or a subset thereof, will need to be determined and assessed by the laboratory during validation studies.

Table 1. Metrics Affecting the Accuracy and Reliability of Sequence Data

Metric	Definition
Alignment	Secondary analysis; mapping quality assigned based on how well an individual read aligns to a reference; typically problematic for insertions and/or deletions (InDels) and STRs
Allele coverage ratio	Heterozygote allele balance (equivalent to peak height ratios in traditional, forensic STR typing)
Base calling	Primary analysis; quality scores assigned to each base of individual reads to assign a probability of base call reliability
Chemistry-specific issues	May exist at platform or assay level, and must be defined during validation
Flanking region selection strategy	Tertiary analysis; InDel and STR data may be called by designation of flanking regions immediately surrounding the expansion/contraction site
Misincorporation rates	Incorrect bases incorporated into, or removed from, the true sequence observed in the combination of individual reads
Read length	The number of bases sequenced in a row along a DNA fragment
Strand balance	Measure of coverage generated for each sequenced strand at a given locus

3.1 Approach

The webinar series served to educate forensic science practitioners, legal professionals, and decision makers on the potential benefits of MPS technology and how it may best fit into the laboratory system. Additionally, the webinar series presented options and potential solutions to the current barriers to adoption. It was essential that the panelists in the webinar series be (1) leaders in the forensic science community who are currently conducting research in MPS that addresses development of the technology, the genetic markers that may be analyzed, and bioinformatics tools; (2) individuals with expertise in validation and implementation issues; or (3) individuals with expertise in the legal arena in admissibility and/or privacy concerns. Therefore, the project team selected panelists for either their current MPS experience or knowledge to discuss and present solutions based on other relevant experiences (e.g., previously validated systems, quality assurance, admissibility of scientific evidence, and/or privacy issues related to genetics and genomics). The webinar series represented an international perspective and engaged with researchers, practitioners, and other experts in both the United States and Europe to elucidate the current landscape of MPS. The panelists were all well-known, respected experts in their respective disciplines and brought varied perspectives on MPS and its potential applications as well as how data generated from MPS could be used and might be constrained.

The project team constructed the webinar series to obtain high engagement from the online participants and provide as much information as possible on practices, procedures, as well as implementation considerations for the production of this document. The project team emphasized the engagement of the leaders in MPS research and application on an international scale. The project team also stressed the participation of the legal perspective and social issues from a U.S. perspective because

using genetic data ultimately impacts the legal system and, more importantly, society. To facilitate participation, the presentations and discussions intentionally allowed time for viewer questions. To ensure fluent dialogue in each webinar, the Web host and a presenter monitored the active chat pod for questions, where presenters were able to answer many questions directly. While the RTI Web host highlighted other questions directed to specific individuals (i.e., one or more of the presenters or panelists) and brought those questions to the attention of that individual. Finally, in order to achieve a focused discussion and to prevent panelists from straying off topic, the project team provided questions to the panelists several days before each live event. These questions were designed to maintain direction of the discussion. The project team encouraged the panelists to address any concerns or questions about the content of the upcoming discussion with the lead presenter; however, no panelists expressed any concerns.

The project team made surveys available to all Web participants immediately following each webinar to obtain data metrics to assess the quality and impact of discussion content, and to gain information on the structure of the Web audience. RTI, the NIJ, and the FTCoE newsletter (29,000 subscribers) were some venues and channels through which the webinar series was advertised and promoted. In addition, the lead presenters reached out to multiple agencies and practitioners.

To achieve the project objectives within the scope of the four-part webinar series, the project team generated the following tasks.

- **Task 1—Assess the current state of MPS as applied to forensic applications.**

  In order to assess the current state of MPS, the project team conducted an extensive literature review (***Appendix A***) and derived specific topics from these documents for discussion during the four webinars and for the final document on MPS. The team ensured that these topics aligned with the project goals, provided sufficient information, and framed the discussion so the content would be engaging, be substantial in depth and scope, and provide some key resources to which the attendees could refer as needed.

- **Task 2—Obtain panelists for the discussion.**

  The project team selected the panelists for each webinar based on their knowledge of the subject matter for the webinar. The team also made every effort to include stakeholders with a variety of perspectives and experiences. There was no set limit on the number of panelists, but rather a focus on creating a dynamic discussion group that could provide as many experiences and perspectives as possible. In that regard, the project team selected a number of panelists from Europe with expertise in MPS research and applications. The project team deemed it necessary to look for international panelists because a large effort by forensic

science researchers and developers to develop MPS strategies is concentrated in Europe, and this expertise was critical to gain insight into the strengths and obstacles to implementing MPS-based technologies.

- **Task 3—Host, webcast, and archive the individual series.**

  The webinar series consisted of one webinar per month for a 4-month block and began in May 2015. The project team constructed each webinar around a 2-hour time slot, with a combination of presentation and discussion to maximize engagement. The project team made available archived versions of the broadcasted webinars after each one of the series was completed. The intent of memorialization was to allow access of the information to individuals who may have missed a particular webinar; in this manner, no viewer was "behind" for any subsequent webinars. Moreover, it was anticipated that viewers would not likely be engaged in MPS and thus may not fully absorb the technology discussions. Therefore, the project team established the archive so the viewers could become more intimate and involved with MPS at a later time.

- **Task 4—Prepare and provide a final report.**

  The project team generated this final report to capture and document the webinar discussions, serve as a resource document, provide the current landscape of MPS, and present discussions for future adoption. The development of MPS technology and the panels of markers that can be analyzed by MPS are dynamic, and, over the next few years, they will likely evolve substantially. Nonetheless, certain issues will be common to all technologies and markers. Therefore, this document provides foundations and concepts and where and how MPS data may be used and not used to guide the forensic science community to make informed decisions as it begins to embrace MPS.

3.2 Project Team

This project was a collaborative effort with the UNTHSC, whose primary consultants and discussion leaders were Dr. Bruce Budowle and Mr. Jonathan King. Dr. Budowle is currently the executive director of UNTHSC's Institute of Applied Genetics and professor in the Department of Molecular and Medical Genetics. He was previously employed by the Federal Bureau of Investigation (FBI) for 26 years and is an expert in forensic genetics. Mr. King directs the research laboratory of the Institute of Applied Genetics and has been conducting research in MPS and bioinformatics for several years.

The RTI team comprised Dr. Patricia Melton and Mr. Shane Hamstra from the Center for Forensic Sciences. Dr. Melton is a senior research forensic scientist and was the project leader responsible for project coordination and logistics. She has nearly a decade of experience as a forensic DNA analyst and has been on the faculty of two universities. Mr. Hamstra is a research training specialist and was responsible for all technical webinar logistics including coordination with subject matter experts, graphic artists, and instructional designers. Biographies of individuals are available in ***Appendix B***.

The panelists were an integral part of the overall goal of describing MPS and defining the issues that the forensic science community should consider. Choosing panelists with backgrounds relevant to the topics presented within each webinar ensured greater impact and more dynamic discussion. Most participants in the first three webinars had a high level of experience in research, development, and application of MPS and/or bioinformatics. The final webinar also included well-known individuals with well-established expertise in legal and social issues. Also, a member of the FBI involved in the operation of the Combined DNA Index System (CODIS) participated to provide insight into what may be considered issues regarding national databases. ***Table 2*** summarizes the subject content and participants for the entire webinar series.

Table 2. Summary of Webinar Series

Broadcast Date	Presenters and Panelists	Affiliation	Subject Content
Session 1: May 20, 2015	Bruce Budowle	Institute of Applied Genetics, UNTHSC	• What is the value of this technology and chemistry as it relates to forensic applications? • Review the current platforms and chemistry. • What are some of the focus areas that need to be evaluated for implementation? • Does this technology represent a paradigm shift?
	Seth Faith	North Carolina State University	
	Ernesto Guzman	Illumina, Inc.	
	Jonathan King	Institute of Applied Genetics, UNTHSC	
	Mike Leliveit	ThermoFisher Scientific	
	Mark Wilson	Western Carolina University	
	Peter Vallone	National Institute of Standards and Technology	

(continued)

Table 2. Summary of Webinar Series (continued)

Broadcast Date	Presenters and Panelists	Affiliation	Subject Content
Session 2: June 17, 2015	Bruce Budowle	Institute of Applied Genetics, UNTHSC	• How does the implementation of MPS technology affect the work flow of a crime laboratory? • What benefits do additional markers bring to forensics? • What considerations need to be made to accommodate national databases?
	Dieter Deforce	Laboratory of Pharmaceutical Biotechnology, Ghent University	
	Manfred Kayser	Department of Forensic Molecular Biology, Erasmus University	
	Jonathan King	Institute of Applied Genetics, UNTHSC	
	Peter de Knijff	Leiden University Medical Center	
	Lilliana Moreno	Federal Bureau of Investigation (FBI)	
	Niels Morling	Department of Forensic Medicine, University of Copenhagen	
	Doug Storts	Promega Corporation	
Session 3: July 29, 2015	Bruce Budowle	Institute of Applied Genetics, UNTHSC	• What is bioinformatics, and what are the basic software tools available for the process? • What are the basic and specialized tools required for analysis of all the marker types? • What considerations need to be made for mixture interpretation?
	Jonathan King	Institute of Applied Genetics, UNTHSC	
	Narasimhan (Narsi) Rajagopalan	ThermoFisher Scientific	
	Christophe Van Neste	Laboratory of Pharmaceutical Biotechnology, Ghent University	
	John Walsh	Illumina, Inc.	
	Brian Young	Battelle	
Session 4: August 19, 2015	Bruce Budowle	Institute of Applied Genetics, UNTHSC	• How do we address and decide on the validation needs of this technology? • What considerations need to be made to accommodate the CODIS database? • What are the legal considerations for MPS data?
	Thomas Callaghan	FBI	
	Rockne Harmon	Senior deputy district attorney, Alameda County (retired)	
	Sara Katsanis	Duke Science and Society, Duke University	
	Jonathan King	Institute of Applied Genetics, UNTHSC	
	Brian Young	Battelle	

4. ASSESSMENT OF THE MPS LANDSCAPE

4.1 Literature Review

The project team conducted a review of peer-reviewed literature, current issues, resources, and policies associated with MPS technology, thus providing an overview of the MPS landscape. Although extensive, the literature review was not intended to be exhaustive; rather, the purpose of the literature review was to identify key resources that address the discussion topics that may assist those considering

whether or not to implement an MPS platform. The key sources designated within the literature review section may assist decision makers in developing policies and procedures related to MPS technology. Currently, this field is very dynamic, and the literature is anticipated to grow rapidly. Regardless, the sources in ***Appendix A*** describe the foundations of MPS and current studies that should be extremely useful for those seeking to gain a foothold in understanding MPS.

4.2 MPS: Understanding the Basic Technology

Session 1

The first webinar's objectives were to provide an overview of massively parallel sequencing (MPS) that accomplished the following:

- Addressed the value of this new technology and chemistry as it relates to forensic applications.
- Reviewed the current platforms and their respective chemistries.
- Addressed the implementation of MPS technology representing a paradigm shift.

To facilitate the discussion of these topics, the project team provided the following questions to the panelists prior to the webinar:

- As laboratories consider these platforms and accompanying chemistries, what areas should be focused on to test and develop the system?
- What value does MPS bring to forensic applications?
- With respect to turnaround time and additional information, how does MPS impact the workflow of a crime laboratory?
- What target enrichment strategies are available?
- What unique chemistry issues arise with the library preparation process?
- What unique chemistry issues arise with sequencing, and how should they be addressed?
- What are the causes of sequencing noise, and how is it distinguished from true sequence?
- What is strand bias, and should it be a concern?
- What are the issues associated with barcoding?
- Does this technology represent a true paradigm shift?

A total of 182 registrants attended the webinar. The majority of them (71%) listed themselves as forensic DNA professionals, with an additional 7% representing crime laboratory managers or directors. Law enforcement and legal representation were each 3%. The remaining attendees listed themselves as forensic professionals (4%) or were from academia (12%). The survey questions and respective metrics are listed in the following table. An event performance sheet (EPS) captured the specific demographic and impact information of this webinar. ***Appendix C*** includes the EPS for this webinar.

Question	Response
How informative was the webinar?	• Highly informative: 17% • Somewhat informative: 80% • Not very informative: 3%
Prior to this webinar, how familiar were you with the concept of massively parallel sequencing (MPS) as applied to forensic DNA applications?	• Very familiar: 15% • Basic understanding: 48% • Not very familiar at all: 37%
How likely are you to share the information presented in this webinar with other policy makers associated with your agency?	• Highly likely: 34% • Somewhat likely: 63% • Not likely: 3%
Choose the option that best describes why you are viewing this webinar.	• My agency is considering implementing an MPS instrument: 25% • I want to know how other agencies are addressing and using MPS technology: 7% • I want to know more about MPS in general: 65% • Other: 3%
Based on the information presented today, do you believe that there are suitable criteria for the application of MPS to forensics?	• Yes: 52% • Possibly: 48% • No: 0%

The project team asked webinar participants the following question: "What was the biggest benefit of attending this webinar?" Some of the feedback included the following responses:

"I learned a great deal about the MPS technology and have a better understanding of the limitations, concerns, and potential of the information and capabilities of the MPS systems."

"Learning about a potential new technology that may be implemented in the future of forensics."

"Hearing more specific information how the two systems work than had previously been made available to me; enjoyed the questions segment at the end."

Summary of Webinar Session 1

The primary value of MPS resides in the massive amounts of data it generates. The forensic science community can combine much larger panels and different types of markers to better characterize evidentiary samples and address particular needs in novel ways for supporting investigative leads. It can also analyze autosomal, Y-chromosome, and X-chromosome STRs simultaneously—and combine SNPs with STRs. Alternatively, the forensic science community can contain each marker system within a single, multiplex panel. Autosomal STRs are informative of identity testing, are well suited for analyzing mixture evidence, and are the mainstay of the forensic genetics community. Y-chromosome STRs allow

for identifying and characterizing male contributor DNA, particularly in mixtures containing large amounts of female DNA. X-chromosome STRs are useful for certain kinship cases, such as incest situations. Due to their smaller amplicon size, SNPs are useful for typing degraded DNA. Although identity SNPs provide similar genetic information as STRs, certain classes of SNPs (e.g., ancestry and phenotype) can be used to generate novel investigative leads that were not readily possible (although feasible) with CE-based approaches. With CE, the forensic science community analyzes these various marker systems separately, each consuming a portion of precious sample. With MPS, the forensic science community can type all of these markers, or subsets thereof, from one portion of a sample. Therefore, the forensic science community can reduce evidence sample consumption and yet gain more information. This sample saving is particularly valuable when there may be enough sample for only a single analysis or when (although not routinely recommended) a sample may have to be divided.

STR polymorphisms (i.e., alleles) manifest as length differences by CE analysis. With MPS, the forensic science community also obtains length-based polymorphisms. Therefore, MPS will provide STR results that are backward compatible with the current millions of STR profiles stored in forensic DNA databases worldwide. Implementation of MPS does not require re-testing of the millions of samples with profiles archived in DNA databases. However, sequencing of STRs—instead of just determining the length of the alleles—provides additional information, which is sequence variation within length-based alleles. Obtaining sequence variants will increase the discrimination power of some STRs for identity testing and kinship analyses. This increased diversity also will enable better deconvolution of some mixtures than what is currently possible.

STR data are best suited for identity testing applications (e.g., direct matching and kinship analyses). There will be situations in which the forensic science community may search an STR profile that is derived, for example, from an evidentiary sample against a database but receives no hits. Currently, under this scenario, DNA evidence provides no investigative lead value. With MPS, the forensic science community can consider two additional approaches that may generate investigative leads even without a direct match from a database search. These approaches are familial searching and ancestry and/or phenotypic information. Familial searching exploits potential kinship relationships (typically parent-offspring or sibling-sibling) between the donor of the evidence and reference profiles in a DNA database (https://www.forensiccoe.org/Our-Impact/Advancing-Technology/Reports/Familial-DNA-Searching-Current-Approaches). In jurisdictions which allow familial testing, MPS could enhance the efficacy of the familial searching process because the increase in the number of markers typed and the ability to simultaneously type lineage markers (e.g., Y-chromosome STRs) can reduce adventitious hits. Moreover, if reference samples are typed both for autosomal STRs and, for example, Y-chromosome STRs, the

amount of labor will be substantially reduced that is currently required for re-testing samples to reduce adventitious hits—which, in turn, will identify highly viable candidates for potential lead value. Ancestry and/or phenotypic SNPs either indirectly or directly, respectively, can provide evidence of the physical appearance of the donor who provided the biological evidence, which can focus investigations and be useful for verifying or refuting eyewitness accounts. Overall, MPS enables the forensic science community to generate more data and novel data that can enhance the power of forensic DNA typing.

All webinar panelists agreed that MPS is of substantial value and will be implemented over the next few years. A number of considerations are necessary to effectively adopt MPS for forensic analyses: validation studies; quality assurance requirements; workflow; throughput; cost, data type, and storage requirements; and analyst training and education. These issues are not new to forensic DNA typing, and the same general approaches that were used to bring to fruition previous DNA typing methods apply to MPS. Clearly, education and training are critical. Workshops and webinars, such as this series and many others, will be essential for providing the information, especially for those analysts who are not engaged in MPS. Analysts should seek formal training, or analysts in training could visit an MPS functional laboratory for hands-on training. When hands-on training opportunities are limited due to budgetary or other constraints, analysts have the opportunity to engage in virtual and simulated environments such as the MPS training tool developed by the FTCoE. Access to this tool along with associated manuals and training materials can be found at https://www.forensiccoe.org/massively_parallel_sequencing_workshop. The turnaround time will increase in comparison with traditional DNA typing methods. There will be demands for some cases to be analyzed expeditiously, and MPS currently cannot provide a DNA result in a similar time frame as CE. However, the sheer increase in data generated may override the time requirement, as more data can develop more leads, and additional markers (e.g., SNPs) can increase the success of typing challenged samples. Until the turnaround time of MPS is comparable with that of CE, laboratories may want to consider maintaining both CE and MPS capabilities in a laboratory.

The MPS chemistry, technology, and jargon may seem foreign to users unfamiliar with them, but the overall process does not represent a paradigm shift in the manner in which DNA evidence is analyzed. The initial process is the same for MPS and CE; that is, one extracts DNA, determines the DNA quantity, and uses PCR to enrich the target markers of each sample. The primary enrichment strategy for forensic DNA typing involves PCR to obtain a sufficient sensitivity of detection. This PCR process is similar, if not identical, to that of traditional CE approaches. Therefore, the experience of amplification and stochastic effects of current DNA typing provide a basis that analysts can rely on when embracing MPS typing strategies. In some situations, one may employ alternate enrichment strategies (e.g., probe-based

capture assays), and these approaches may enable typing of highly degraded samples with marker systems (e.g., mtDNA genome).[72,73]

The differences between MPS and CE lie in the library preparation, cloning, and sequencing. Commercial kits for each of these steps are available, and commercial manufacturers already provide workflows to facilitate the process. There are differences in the various library preparation and sequencing chemistries, and therefore each system will have to be validated. Laboratories that use the same workflows may be able to leverage data and experience among them. In contrast, the results of a different chemistry may not provide support for the validation of a different system, with one exception. As previously stated, the amount of generated data is massive. It would take weeks to months for a CE system to match the output of a single MPS run. Therefore, it will be an unreasonable expectation that current CE-based data will have to be used to validate MPS data. Instead, data from different MPS platforms and respective chemistries may be a better way to validate these systems. Orthogonal testing is not a new concept and should be considered for validation by some laboratories.

The volume of data produced by MPS appears daunting, and the forensic science community has yet to deal with such a magnitude of data. Researchers often use a variety of software tools and combine these tools into a workable pipeline, which can be a very cumbersome process for an operational laboratory, especially if a bioinformatician is not present in the laboratory. However, commercial entities already are providing streamlined pipelines capable of analyzing and managing the data and for providing helpful data outputs of results. The laboratory will need to validate these software packages. Another data issue is that new types of data analysis (e.g., principal component analyses [PCAs] plots for estimating ancestry and reporting probabilities for hair and eye color) may be unfamiliar to many forensic DNA analysts. Training, education, and validation of these approaches will have to occur so that the data will be understood and reported properly. Although the amount of data is substantially greater, the transition to the laboratory likely will be accomplished in a similar fashion as that which occurred for other DNA typing systems.

4.3 MPS: The Genetic Marker Potential

Session 2

The objectives of this webinar were to

- Discuss how the implementation of MPS technology affects the workflow in a crime laboratory.
- Review the types of genetic markers that can be typed with MPS technology.
- Discuss the benefits that these additional markers bring.
- Discuss the considerations to accommodate national databases.

To facilitate the discussion of these topics, the project team provided the following questions to the panelists prior to the webinar:

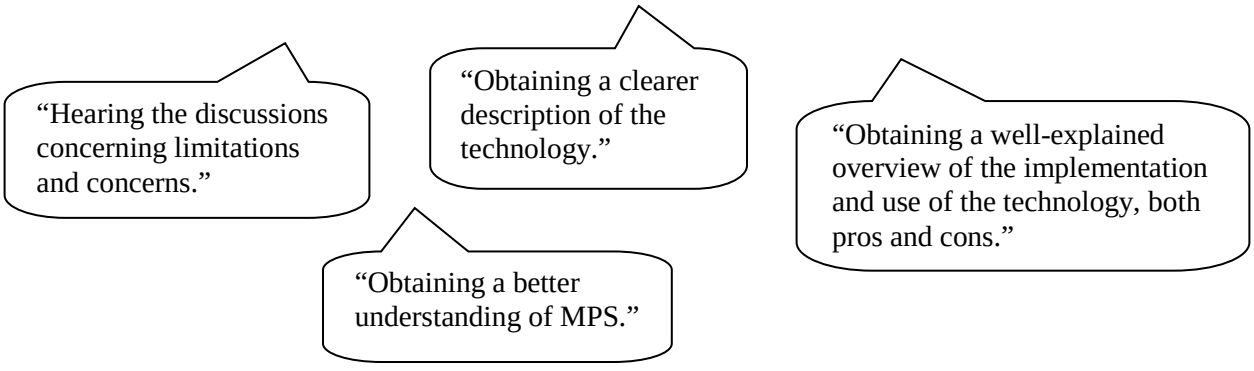
- What types of STRs can be analyzed?
- Should all STR categories be sought?
- What additional information or benefit can be obtained?
- Why seek this additional sequence information?
- Why seek SNPs for forensic applications?
- What types of SNPs can be analyzed?
- What SNPs should we focus on, and why?
- Should we make efforts to expand to whole genome sequencing? If so, why?
- Is the European DNA Profiling Population Database (EMPOP) working toward accommodating the additional information?
- What can we do to support EMPOP in this endeavor?
- What can we do to support Y-chromosome STR databases, such as Y-Chromosome STR Haplotype Reference Database (YHRD), in order to accommodate additional sequence information?
- What sequence information should be uploaded into a national DNA database?
- What are the benefits for each of the two different strategies for using MPS data?
- Is there a maximum number of markers that can be typed?
- What other information/markers would one want to seek (if any)?

Nearly 100 people registered for this webinar and 40% attended. The majority (67%) of participants listed themselves as forensic DNA professionals, with an additional 9% representing crime laboratory managers or directors. Law enforcement and legal representation were each 3%. The remaining attendees listed themselves as forensic professionals (6%), academia (9%), or other (3%). ***Appendix D*** includes the EPS for this webinar.

The project team asked participants the following questions to assess a metric of impact.

Question	Response
Choose the option that best describes why you are viewing this webinar.	• My agency is considering implementing an MPS instrument: 24% • I want to know how other agencies are addressing and using MPS technology: 18% • I want to know more about MPS technology in general: 58%
How informative was the webinar?	• Highly informative: 41% • Somewhat informative: 59% • Not very informative: 0%
Based on the information presented today, do you believe there are suitable criteria for the application of MPS technology to forensics?	• Yes: 94% • Possibly: 6% • No: 0%
Prior to this webinar, how familiar were you with the concept of MPS as applied to forensic applications?	• Very familiar: 19% • Basic understanding: 70% • Not very familiar at all: 11%
After viewing this webinar, do you feel you have a better understanding of how MPS would fit into the workflow of a DNA crime laboratory?	• Yes: 58% • Somewhat: 42% • Not really: 0%
How likely are you to share the information presented in this webinar with other practitioners associated with your agency?	• Highly likely: 75% • Somewhat likely: 25% • Not likely: 0%

The project team asked webinar participants the following question: "What was the biggest benefit of attending this webinar?" Some of the feedback included the following responses:



Summary of Webinar Session 2

It became apparent during the first webinar via participant comments that some attendees were not sufficiently familiar with the general workflow of MPS, which made it difficult to follow some of the discussion. This lack of knowledge is expected because very few in the community are engaged in MPS research or testing, but lack of knowledge reinforces the earlier recommendation for education and

training to promote and expand MPS use. To address this temporary limitation, the project team dedicated a substantial portion of the second webinar for the panelists to provide a basic tutorial to familiarize attendees with the MPS workflow. That portion is not summarized herein because an introduction to the basic technology and chemistry of MPS is located at the beginning of this document. In addition, if the reader desires to become further informed about the details of MPS workflow, the comprehensive presentations are archived at https://forensiccoe.org/Our-Impact/Sharing-Knowledge/Virtual-Education/Massively-Parallel-Sequencing.

The selection and use of genetic markers are critical because the genetic markers impact the types of evidence that the laboratory can analyze and the information that it may translate into viable investigative leads. Although MPS may analyze any type of genetic marker, the CODIS core STR loci should be accommodated by MPS regardless of the potential to expand the MPS STR marker set(s). However, research indicates that a few STRs may not be compatible with an MPS system. Relying on the required core for platform selection may need to be reconsidered as a strategy moving forward. It might be worthwhile to consider relaxing the requirement of all core STR loci that MPS will genotype. For example, if two different MPS kits were required to meet at least 17 out of 20 core STR loci, they would still have, at a minimum, 14 loci in common, although the common number could be higher. Fourteen loci would still provide substantial power far exceeding any foreseeable increase in the size of DNA databases. Moreover, given that MPS can type many more STRs, losing a few markers by design or due to sample quality can easily be regained many fold. The forensic science community should research the value and consequences of this strategy of requiring all core loci in a panel.

More data are needed before selecting additional STR markers. Data to consider include genetic diversity, sequence variation, stutter rates, sequence motif, error rates, noise, and chemistry compatibility. The forensic research community should place a focus on reducing the amplicon size for more mini STRs. MPS allows for reduction in size of STR amplicons, thereby allowing for more challenging samples to be typed.

Considering the wide range in applications (e.g., paternity, missing persons, and forensic casework), a single kit containing all three STR categories may not be the most strategic approach. For example, markers with low mutation rates are better suited for kinship analyses. An example of a highly discriminating marker for identity testing is SE33; however, its high mutation rate would be problematic for kinship testing. X-chromosome STRs can be quite useful for particular kinship cases but are not particularly applicable to mixture analyses. Thus, a more specialized kit may be more suitable. Each manufacturer has addressed this matter by offering multiple primer mixes for different applications. In contrast, for reference samples placed into DNA databases, a comprehensive panel may be sensible. As

the evidence and case context drive the best markers to analyze, one cannot know the type of marker that will provide the best information. If the reference samples contain all or most usable marker systems, then no matter what panel is used to analyze an evidence sample, some database searching may be feasible.

Many laboratories use CE to type Y-chromosome STRs and mtDNA sequence data. These markers are valuable for close relative and distant relative associations, for mixture analyses, and/or for increased sensitivity of detection. The Y-chromosome STR sequence data (as with all STRs) also will require nomenclature for communication and substantial database updates to accommodate the sequence information. In contrast, the EMPOP database already supports mtDNA data and has made great strides to accommodate whole mtDNA genome data.[74] Therefore, MPS mtDNA data likely will be more readily implemented because most of the requirements are in place.

One issue webinar participants raised was the use of rapidly mutating Y-chromosome STRs placed into a multiplex kit. These markers are highly polymorphic and thus desirable for identity testing, but their high mutation rate makes them less suitable for kinship association cases. They, however, are good markers for potentially distinguishing close paternal relatives, such as brothers. All STRs have relatively high mutation rates, so partitioning highly mutating Y-chromosome STRs does not eliminate the need for understanding how to deal with mutation in kinship analyses. Regardless, the forensic science community should consider choosing the appropriate marker for the application for panel and kit design. For characterizing database reference samples, one should consider all selected markers and only search those types for case-specific circumstances. The forensic science community should consider which markers are best suited for each application. More studies are required to establish the global utility of each marker or panel of markers. Ultimately, the consumers (i.e., the forensic science community) will need to be educated on the use and application of the individual markers/panels because the community will be the driver for more marker selection.

Although MPS can sequence any marker, there is some debate on the best strategy to move forward. Some suggest that the entire mtDNA genome be sequenced, and others advocate that a better strategy is combining nuclear markers and the hypervariable region of the mtDNA genome. It is well established that mtDNA typing is the most sensitive approach for typing DNA and can yield results when the nuclear markers fail to do so. Therefore, sequencing the whole genome may be a better strategy for analyzing the most challenging samples. In contrast, small-sized amplicon nuclear markers likely will increase the sensitivity of detection to possibly rival that which is currently enjoyed with mtDNA analyses by CE. Some panelists expressed doubts about the multiplexing efficiency of including markers with notable copy number differences (which is the case for nuclear and mtDNA) in the same multiplex.

X-chromosome STRs have a more limited application than other markers and were not advocated as a priority by the webinar participants. Despite perceived legal issues, the webinar participants placed ancestry and phenotypic markers ahead of X-chromosome STRs for kit development purposes. All panelists agreed that ancestry and phenotype SNPs have investigative value on a case-by-case basis, but these should not be maintained in forensic DNA databases. Such genetic information has little intrinsic value for database searches (i.e., limited discrimination power), but does have value on a per-case, investigation level.

SNPs, which consist of small-sized amplicons, may provide more genetic data from challenged or degraded samples than STRs. One panelist noted that the amplicon sizes of STRs were sufficiently small in commercial MPS STR panels, and therefore SNPs offered nominal improvements. There is no disagreement that SNPs in general can be generated on average from smaller amplicons than that of STRs. Regardless, the panelists agreed that SNPs offered several analytical advantages (e.g., ease of typing) and simplified interpretation. On a marker-by-marker basis, STRs appear to be better for mixture interpretation compared with SNPs because of a greater number of alleles that can facilitate mixture deconvolution. However, it is possible that a very large panel of SNPs may rival STRs for mixture analyses. One does not have to contend with the artifact stutter with SNPs. In addition, microhaplotype SNPs, two or more SNPs within an amplicon, could mimic to some degree a multiallelic STR marker. Multiple panelists indicated that if, 20 years ago, SNPs could be typed as easily as they can be today by MPS, they would have been the marker of choice for identity testing. Validation studies should shed light on whether this size difference has any practical impact in typing success and whether SNPs can rival STRs for mixture deconvolution.

Tissue source identification is an important analysis for some cases. Currently, a variety of protein- or chemical-based tests are used to presumptively determine or confirm the tissue source of a biological sample. Laboratories often cannot run these tests in parallel, which creates a resource burden. A promising approach is to use nucleic acid markers, either messenger RNA (mRNA) or methylation sites in DNA, to determine tissue source. Several researchers are evaluating the possibility of combining both mRNA and DNA markers within the same assay. Panelists suggested that one approach may be parallel library preparations using traditional primer design or bisulfite conversion of epigenetic markers. However, a single assay containing all relevant markers has yet to be described.

Although there was substantial discussion about marker selection and practical implications of marker use, the overwhelming criterion for choice of system was whether the provided kit is sufficiently robust. The ability to type challenged samples appears to override the choice of marker.

Nomenclature is an obstacle that must be addressed, especially for STRs. Sequence data require a new mechanism(s) to describe alleles. The nomenclature experience with sequencing mtDNA will provide guidance for MPS data. STR data may be presented as full text strings, motif-based, or a simplified set of characters for each variant. The forensic science community may embrace all three strategies. Efforts are underway under the auspices of the International Society of Forensic Genetics to develop an overarching nomenclature system; therefore, it is likely that the forensic science community will address nomenclature in the near term, and some organization will take the responsibility to oversee and maintain the integrity of the system. Funding is essential to coordinate and maintain a robust nomenclature infrastructure. To facilitate the process, as many participants as possible should submit high-quality data to identify the genetic variation and generate population frequency databases for statistical calculation regarding DNA evidence. Lastly, the forensic science community will have to modify or enhance databases to accommodate the expanded genetic variation.

Although it is anticipated that a few laboratories may implement MPS as a standalone technology, there was overwhelming agreement that MPS technology initially will most likely be implemented primarily as an adjunct to current CE approaches. The primary reason provided for this strategy was the need for education and training regarding defined use, workflow, validation, implementation, and interpretation of data. Additional reasons given for MPS as an adjunct to CE included other new technologies (e.g., rapid DNA typing) and the associated costs with MPS; however, the latter concern may be a misguided perception. On a per-sample basis, MPS cost is nearly equivalent to CE-based DNA typing after ~32 samples/run (see Figure 1). This cost argument is countered further when considering the amount of data generated from an MPS run. On a per-marker basis, some MPS chemistries can be cheaper than CE-based typing after multiplexing only 10 samples within a run. The forensic research community should perform more cost-benefit analyses to determine the resource needs for MPS implementation.

4.4 MPS: Bioinformatics

Session 3

The objectives of this webinar were as follows:

- Provide an overview of what bioinformatics is and how it pertains to MPS technology.
- Discuss both basic and specialized bioinformatics tools.
- Discuss the considerations required for mixture interpretation.

To facilitate the discussion of these topics, the project team provided the following questions to the panelists prior to the webinar:

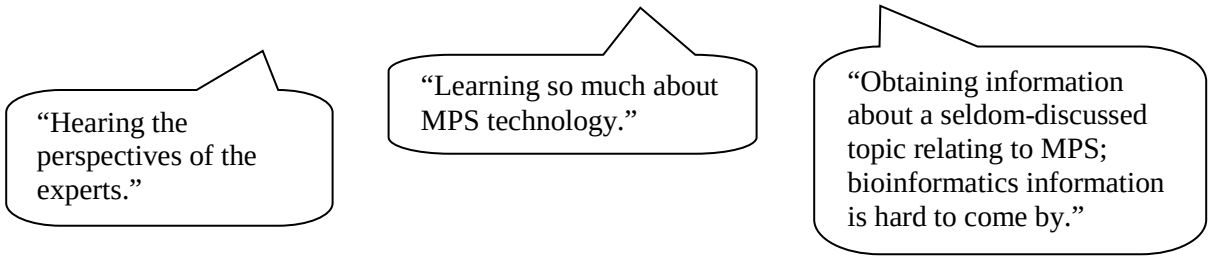
- Does the battery of current software meet the needs for forensic analysis?
- How user-friendly is the current software?
- What improvements are needed to facilitate the transfer and use of software to the community?
- Will there be a need for bioinformatics specialists in the forensic DNA laboratory?
- Given a certain number of markers, how long will it take to generate an output?
- Will data analysis be a bottleneck?
- How do we validate software?
- Predictive modeling is used for some markers, but what do the results entail?
- How do we convey the interpretation of the data?
- What data should be stored or maintained?
- Are there considerations for storage issues?
- Should the DNA forensic science community add bioinformatics training to the training requirements?
- The Open Forensic DNA Analysis Toolbox (OFDAT) may combine strategies or take the best of different approaches. What do you see as the highest priority to tackle bioinformatically?
- How should we address nomenclature?
- Should there be positive or negative strand strategies for consistency?
- Should there be a bioinformatic ranking of the marker difficulty for analysis?
- How should sequencing data and results from different platforms be analyzed or compared?
- What data are important from a bioinformatics tool perspective?

More than 100 people registered for this webinar and 45% attended. The majority (66%) of participants listed themselves as forensic DNA professionals, with an additional 7% representing crime laboratory managers or directors. Law enforcement and legal representation were each 4%. The remaining attendees listed themselves as forensic professionals (7%), academia (4%), medicolegal death investigators (4%), or other (4%). ***Appendix E*** includes the EPS for this webinar.

The project team asked participants the following questions to assess a metric of impact.

Question	Response
Choose the option that best describes why you are viewing this webinar.	• My agency is considering implementing an MPS instrument: 31% • I want to know how other agencies are addressing and using MPS technology: 14% • I want to know more about MPS technology in general: 55%
How informative was the webinar?	• Highly informative: 35% • Somewhat informative: 65% • Not very informative: 0%
Based on the information presented today, do you believe there are suitable criteria for the application of MPS technology to forensics?	• Yes: 100% • Possibly: 0% • No: 0%
Prior to this webinar, how familiar were you with the concept of MPS as applied to forensic applications?	• Very familiar: 21% • Basic understanding: 69% • Not very familiar at all: 10%
How likely are you to share the information presented in this webinar with other practitioners associated with your agency?	• Highly likely: 72% • Somewhat likely: 28% • Not likely: 0%
After viewing this webinar, do you feel you have a better understanding of how MPS would fit into the workflow of a DNA crime laboratory?	• Yes: 58% • Somewhat: 42% • Not really: 0%

The project team asked webinar participants the following question: "What was the biggest benefit of attending this webinar?" Some of the feedback included the following responses:



Summary of Webinar Session 3

Bioinformatics combines computer science, mathematics, statistics, and engineering functionalities to analyze and interpret biodata. The application of bioinformatics is not different in concept for MPS data as it is for CE data with STRs. A large amount of data processing occurs behind the scenes for generating CE output data. The CE output, an electropherogram and supporting data, is a simplified picture resulting from raw data, processing signals, and the creation of an image that is readily interpretable. The same thing occurs with MPS data, except the magnitude of data is on a scale unprecedented in forensic genetics, and the forensic scientist will need to become familiar and appreciate

what goes on behind the scenes of bioinformatics with MPS data to develop and implement effective systems. It is unlikely that each crime laboratory will have a resident bioinformatician, so it will be necessary to develop facile and robust data and output processes.

The forensic science community has performed sequencing of biological evidence for more than 20 years through sequencing of mtDNA by Sanger sequencing chemistry and detection by the CE system. Sanger sequencing chemistry has an associated error rate. Likewise, sequencing error is associated with MPS chemistries. The error rate with Sanger sequencing was nominal as laboratories sequenced both strands of the DNA molecule, and the consistency of sequence results of both strands was required to call the specific bases in a target site. This *two times X* (*X* stands for coverage and is presented as *2X* for Sanger sequencing of both strands) approach reduced Sanger sequencing error to levels considered tolerable in the forensic genetics laboratory. Overall, the sequencing accuracy for current forensic MPS kits has yet to be established; however, manufacturers indicate the accuracies of their systems to be between 99.5% and 99.999%, based on internal studies. The webinar panelists have had similar experiences with high coverage once data processing is performed. The user will need to be able to analyze data effectively, with facile bioinformatics tools, to achieve sequence data accuracies similar to current methods in the crime laboratory. All indications are that such sequencing accuracy can be achieved. Further studies must address base calling accuracy and any locus or motif-specific error rates. The accuracy levels for STRs, in particular, may impact selection of analytical thresholds with sequence-based data. Length-based data (i.e., the backward-compatible data with that of CE) may not be similarly restricted, as sequence error may not affect the length of a fragment. It may be possible, on a case-by-case basis, that when coverage is very low (i.e., only a few fragments of the target are sequenced) and confidence in sequence accuracy is reduced, it still may be possible to type a length-based STR polymorphism with confidence. It is not clear yet whether sequence from both strands will be required for obtaining high accuracy—at least for some loci—or if data from only one strand will suffice. Validation studies will be required to determine whether sequence data from one or both strands will be necessary. Strand balance may be necessary for some markers but unlikely for all markers.

The current “needs” of the community for bioinformatics have yet to be fully realized. The current requirements are relative to those individuals involved in the development and testing of the current software and their experiences may not represent the needs of eventual downstream users. It is anticipated that as the technology transitions from early to mainstream users, the forensic science community will seek additional requirements. Early users tend to use a variety of software tools that are not designed specifically for the demands of the casework laboratory. More tools, user-friendly interfaces, and single pipeline systems will be needed for more widespread use of MPS technology for forensic

analyses. Commercial manufacturers have been developing, and continue to develop, software that include features to address these needs based on early access feedback. If software tools facilitate analyses and provide user-friendly capabilities, laboratories should not need dedicated bioinformatics specialists. Software validation is critical to ensure that tools are sufficiently robust for mainstream application. As previously stated, software validation is critical. For example, during an analysis of a large panel of markers, some markers were not detected, which would lead the researcher to hypothesize that the chemistry failed. However, for a number of technical reasons, the software failed to locate the specific data. The failure was readily resolved but points to the need for understanding the limitations of software, testing software, and producing robust tools for the end user. All are part of the validation process.

The MPS workflow currently requires a number of manipulations, especially on the front-end sample preparation. The process must be able to prepare and analyze a relatively large number of samples simultaneously and minimize chances of contamination within the laboratory setting to realize the benefits of MPS for practical use. Forensic analysts will need to be familiar with the entire process to ensure proper application. Training and educating will be necessary for analysts to understand the pipelines and the data generated. Bioinformatics training should be considered for forensic DNA analysts as soon as possible to prepare them for MPS and the data that will be generated. The training should focus on understanding the potential and applications, as well as to help participants become fluent on the topic.

The vast amounts of generated data pose a novel problem previously not encountered in the forensic DNA laboratory. The amount of data generated will likely outpace current data storage capacities. Thus, data storage is a critical issue. The forensic science community needs to determine if all data, raw and processed, will be retained. It is anticipated that there will be requests to review raw data, which might be expected during discovery by the legal community, or the final processed data (i.e., FASTA or FASTQ files) that contain all sequence reads used for downstream analysis. Laboratories rarely reanalyze most, if not all, stored raw data. On a practical level, the processed data are much smaller, easier, and cheaper to store than stored raw data, and laboratories typically rely on processed data for genetic interpretation. The forensic science community may have to consider mechanisms for providing large amounts of data. At this time, the forensic science community has not discussed what data to retain. Actual use, quality issues, cost, and admissibility challenges likely will affect policy.

The sensitivity of detection appears to be comparable between CE and MPS systems, and PCR is still the preferred enrichment method; therefore, the same stochastic issues with resultant data will arise. Fortunately, software is available to help the forensic science community interpret forensic DNA data. So-called smart systems, such as probabilistic genotyping, can enable the interpretation of limited and

more complex DNA profiles than had been possible with more binary systems that were used originally. These approaches should translate to MPS data. The forensic science community should seek modeling of MPS data for probabilistic genotyping.

All indications are that with PCR enrichment on the front end of the analytical process, MPS rivals CE for detection sensitivity. Sensitivity will be dynamic and can vary from run to run. When laboratories analyze a few samples per run, coverage will be relatively high compared to a run with a greater number of samples. This aspect of MPS is not a negative or positive observation but rather a reflection of the throughput of the system. Detection sensitivity, an important criterion, will have to be balanced with sample throughput (a criterion for cost benefit). Obviously, detection sensitivity is the more important factor, but for MPS to be a viable approach, the forensic science community will have to consider the cost per marker or cost per sample. One advantage with sequence data is that noise may be distinguished better from true DNA signal. This feature may reduce a detection threshold, which, in turn, may enable greater detection sensitivity. A cautionary note is that reduction in noise, if possible, does not mean that a relaxation in interpretation guidelines will follow. The forensic science community will need to develop guidelines in a manner similar to those in place for current DNA typing systems, but specific to the requirements of MPS.

As is with STR kits from different manufacturers, data compatibility from different MPS kits must occur. Otherwise, the efficacy of searching forensic DNA databases may be somewhat reduced. Studies to date, although limited, demonstrate that the compatibility of MPS allele calls with CE-generated STR data and between MPS platforms are comparable with those that have been observed between commercial STR/CE kits. Thus, data compatibility should not be a hurdle to implementing MPS. However, some conversions of STR allele data may be necessary. With the CE STR, alleles are defined operationally by length in comparison to an allelic ladder. The underlying sequence variability of an allele is unknown. With MPS, laboratories will know the underlying variability and discern the true repeat region within STR alleles. In some cases, the true repeat region may be different than the assumed length-based allele call. This difference is not an error in typing; but it does affect nomenclature. The forensic science community will have to choose to select either the original length-based STR nomenclature or change to one based on sequencing. The decision should be policy-based, may be determined by convenience, and can be facilitated readily bioinformatically.

Unlike STRs, there are no accepted panel of core SNPs. Therefore, manufacturers are developing commercial and in-house SNP panels without direction by the forensic science community. This could possibly result in some incompatibility of marker sets among laboratories. Fortunately, manufacturers have relied on similar resources for SNP selection, and therefore there appears to be substantial overlap in

SNP markers between commercial kits. The forensic science community should develop criteria for SNP selection, define a core set of markers, and/or determine whether only a portion or all of identified candidate markers need to be in a panel for practical use. The International Society for Forensic Genetics (ISFG), the Scientific Working Group on DNA Analysis Methods (SWGDAM), and the research community are considering guidance issues for validation and nomenclature.

4.5 MPS: Validation and Applications

Session 4

The objectives for this webinar were as follows:

- Discuss how to address the validation needs of this technology.
- Discuss what considerations need to made in order to accommodate the CODIS database.
- Discuss the legal considerations for MPS data.

To facilitate the discussion of these topics, the project team provided the following questions to the panelists prior to the webinar:

- How do we manage validation when stability is needed but we are in a phase of technology development where dynamic change is high?
- What areas and steps require validation?
- Are we able to leverage past studies or partner with other laboratories to reduce some of the burden?
- How should software be validated?
- Should there be additional validation criteria for data interpretation?
- Will threshold and noise levels be different for STRs than for SNPs?
- How should sequencing data or results from different platforms be analyzed and compared?
- What reference materials will be needed, and how should they be characterized?
- How should heteroplasmy be addressed?
- Should databases be expanded to accommodate the additional markers available?
- Should database searches accommodate all or a subset of the markers?
- What are the concerns about the admissibility of MPS data?
- Are there potential privacy issues associated with MPS data?

A total of 70 people registered for this webinar and more than 60% attended. The majority (70%) of participants listed themselves as forensic DNA professionals, with an additional 25% representing crime laboratory managers or directors. Law enforcement was 5%. There was not legal, forensic professional, academic, medicolegal death investigator, or other representation. ***Appendix F*** includes the EPS for this webinar.

The project team asked participants the following questions to assess a metric of impact.

Question	Response
Choose the option that best describes why you are viewing this webinar.	• My agency is considering implementing an MPS instrument: 63% • I want to know how other agencies are addressing and using MPS technology: 26% • I want to know more about MPS technology in general: 11%
Based on the information presented today, do you believe there are suitable criteria for the application of MPS technology to forensics?	• Yes: 100% • Possibly: 0% • No: 0%
Prior to this webinar, how familiar were you with the concept of MPS as applied to forensic applications?	• Very familiar: 30% • Basic understanding: 70% • Not very familiar at all: 0%
How likely are you to share the information presented in this webinar with other practitioners associated with your agency?	• Highly likely: 72% • Somewhat likely: 28% • Not likely: 0%
After viewing this webinar, do you feel you have a better understanding of the validation needs for MPS workflows?	• Yes: 40% • Somewhat: 60% • Not really: 0%
After viewing this webinar, do you feel you have a better understanding of the role of MPS technology with respect to the Combined DNA Index System (CODIS)?	• Yes: 60% • Somewhat: 40% • Not really: 0%
After viewing this webinar, do you feel you have a better understanding of the legal considerations for MPS technology?	• Yes: 58% • Somewhat: 42% • Not really: 0%

The project team asked webinar participants the following question: "What was the biggest benefit of attending this webinar?" Some of the feedback included the following responses:

"As a certified latent print examiner and BCI director, I found all of the information valuable."

"The webinar provided sufficient information to enable us to begin our initial assessment of the technology and put a validation plan in place."

Summary of Webinar Session 4

MPS validation has yet to be described extensively for forensic applications. However, a number of studies are underway, and experience from early users suggests that MPS systems will meet requirements for forensic applications and be successfully implemented. Developing specific validation

criteria for guidance within the forensic science community would be helpful. There are some publications on MPS validation in other related disciplines that may facilitate the development of validation criteria.[75–78] Validation initially may be divided into chemistry, instrument, and bioinformatics systems, but the forensic science community must eventually test the entire system. Availability should not be a metric in itself for software reliability. The forensic science community should establish guidelines for software validation. The project team recommends that laboratories conduct performance checks early on to refine the system. One benefit of MPS is that the chemistries and technologies are evolving rapidly, and these advances hold promise to enhance MPS, making it even more desirable for forensic applications. However, this benefit also has a constraint. Substantial resources and time are dedicated to validating a system. There is a risk that by the time a validation study is completed, standard operating procedures (SOPs) are prepared, and analysts are qualified, the existing system may have changed for the better, but the versions of kits and software are no longer available. Thus, the laboratory may have to validate the system again. Some degree of stability is needed to enable implementation and carry out work for a reasonable time frame.

Currently, although the cost per marker can be comparable or even less than that for CE, an overall MPS experiment is rather costly. The community might consider developing a network to share data for certain aspects of development and validation to reduce the cost burden per laboratory. The added value of a network of MPS users is that all users can gain from the experience of others. Leveraging efforts with other laboratories may help move the technology forward in an efficacious manner. To exploit data sharing, the forensic science community should determine the parts of validation that can be shared and the parts that should be performed in each laboratory. For example, internal validation is an important process for determining the functionality of a system within the laboratory, and the process provides invaluable experience for the user. Therefore, data sharing must be balanced with lab-specific needs and cost benefits.

Although validation concentrates on analytical processes and can be accomplished in a straightforward manner, there are other aspects of MPS to be addressed to facilitate technology transfer and make the best use of this powerful system. These include marker selection, privacy concerns, legal admissibility, databases, and data storage. Some individuals have raised concerns that MPS-generated data could pose risks to privacy. The issue of privacy should not be anchored on MPS per se. MPS is just a tool for DNA sequencing. The focus should be on the markers used for forensic typing, regardless of if they were typed by MPS, CE, or another methodology. Marker selection and privacy are not new topics in the forensic DNA typing arena. Sequencing the entire human genome certainly would present an issue regarding potential divulgence of data on genes that could disclose personal information (e.g., health risks

or status of individuals). However, such a concern is likely to be very limited in forensic analyses. Most applications for forensic investigations, except possibly differentiating identical twins, would not involve sequencing the entire genome. The markers selected for identity testing (to date) are targeted for their individualization value and likely pose little or no privacy risks. These markers have little predictive value and essentially have no more privacy risks than current forensic human DNA markers. Moreover, whole genome sequencing currently is not an effective use of MPS technology. Most of the throughput of MPS would be wasted on sequencing regions of little human identity value. Furthermore, whole genome sequencing would not be cost-effective.

MPS-generated STR data are likely to be entered into DNA databases, as these resources already accommodate such data. The forensic science community currently does not enter identity SNP data into DNA databases for developing investigative leads. There is nothing inherently different about identity SNPs and STRs, except that on a per-locus basis, SNPs are less polymorphic. SNPs, however, potentially offer a higher success rate for typing degraded samples. Efforts to upload identity SNP data into DNA databases and implement search tools that can accommodate SNP data would help support the law enforcement community in developing more investigative leads.

Other markers (e.g., ancestry informative SNPs and phenotypic SNPs) require further consideration as to how and when to apply these markers. Clearly, these classes of SNPs can have investigative value in certain situations, most likely when no database search hits (either direct or indirect) have been obtained. Unlike identity markers, the panelists were strongly against entering ancestry and phenotype SNPs into DNA databases. These types of SNPs have limited investigative power, and thus database searches based on these markers are not a good use of resources. Panelists discussed privacy concerns regarding the use of ancestry and phenotype SNPs, and deemed the risk to be low. However, it is possible that an ancestry SNP might have some predictive power regarding personal (i.e., privacy) information. The forensic science community should assess whether any of the ancestry SNPs that are selected have known associations that have reasonably high predictive power. Ancestry SNPs describe the ethnic makeup of an individual. Some individuals may not want that information disclosed. The forensic science community should discuss whether or not ethnic ancestry is a privacy issue. Phenotype has been a bit more confusing. Phenotype has different meanings to the forensic science and general genetics communities. Phenotype, from the forensic science community perspective, is the physical appearance of an individual. In the genetics community, phenotype is used in broader sense as the expression of any gene. These two definitions are sometimes interchanged and result in misunderstandings about privacy risks. The physical appearance of an individual is not a privacy issue, and thus every effort should be

made to clarify what is meant by the term phenotype when discussed for application by the forensic science community.

Primary privacy concerns in the eyes of the public tend not to focus on what data are retained, but rather on who has access to these data. As discussed in the webinar, some individuals may be concerned that the police may misuse the DNA data despite the lack of any such cases to date. However, it seems that methodology is in place to control any misuse of data. After 25 years of no privacy violations, the track record is quite good regarding the proper use of forensic genetic data. Although MPS is different in terms of sophistication, the markers that will be generated are the same or similar to current markers, and existing safeguards will suffice.

As previously stated, the vast amount of MPS data presents an unprecedented challenge. Software tools and effective pipelines address data interpretation, which likely will not be an obstacle for long-term implementation. However, data storage requires further discussion because it can be costly to maintain all sequence data. On a practical level, it does seem reasonable to discard a portion of or all raw data (i.e., data before generating a FASTQ file) from validation studies and population studies. Individuals rarely review the original raw data, which reduces the cost burden. It may be more cost-effective to rerun some samples, if, on a rare occasion, raw data are required. However, from a legal perspective, the retention of raw MPS data may shape policy. Although a researcher or an analyst may never use the raw data, and the data would likely add little value to support the reliability of a result, it is anticipated that some legal professionals will request the raw data. The project team recommends that the forensic science community discuss more about the cost, benefits, and risks of maintaining or discarding raw data.

Including sequence data in a DNA database presents technical challenges that must be addressed. For STRs and the increased allelic information that is obtained by MPS, the forensic science community will have to develop a defined nomenclature system. Allele nomenclature is not a trivial issue. The nomenclature could be in full string (full sequence text), a summary of the motifs and variants, or with some arbitrary letter or number system. Nomenclature systems will require an official stable body to identify and name alleles. The ISFG appears to be considering potential nomenclature strategies for STR alleles. DNA databases and software packages (e.g., tools for population genetic analyses) must be able to accommodate the new data.

As discussed heavily in the webinar, it can be anticipated that MPS-generated data will be met with admissibility challenges through Daubert and/or Frye hearings, as have all other DNA typing methodologies. The defense is likely to vigorously challenge evidence in support of its client(s). As previously stated, MPS use is not considered a paradigm shift. There is nothing inherently different about MPS and CE as far as addressing the criteria for admissibility. MPS technology is a detection tool. New

markers—although most, if not all, are not novel to the forensic arena—may not likely be a basis for an admissibility hearing in most jurisdictions because markers themselves have not been the basis for reliability. To address admissibility challenges, the project team recommends that the forensic science community perform good science, such as validation and peer review.

5. COMPLETE SESSION HIGHLIGHTS AND SUMMARY

More than 470 online participants from around the world attended the live webinar series. The participants represented forensic practitioners and professionals including forensic DNA analysts, technicians and specialists, laboratory directors and managers, legal and law enforcement representatives, academics, and medicolegal death investigators.

This webinar series discussed the considerations and needs addressing MPS technology adoption to the forensic crime laboratory. The participants greatly valued the series, as indicated by the survey metrics. With each webinar, the project team observed an increase in the percentage of individuals who felt that there are suitable criteria for the application of MPS technology to forensics. This strengthens the notion that the forensic science community values MPS technology, and the strengths of this technology are beneficial to forensic applications. Respondents also indicated an increase in the likelihood of sharing the information from this webinar series with colleagues, again demonstrating the importance and comprehensiveness of the information presented in this series. Several guiding observations may be obtained from the discussions and presentations that took place in this series and are listed below:

- The primary value of MPS resides in the massive amounts of data it generates. Much larger panels of markers and different types of markers can be combined to better characterize evidentiary samples and to address particular needs in novel ways for supporting investigative leads. With MPS, the forensic science community can consider two additional approaches—familial searching and ancestry and/or phenotypic information—that may generate investigative leads even without a direct match from a database search.
- The forensic science community may demand for some cases to be analyzed expeditiously, and MPS currently cannot provide a DNA result in a similar time frame as does CE. However, the sheer increase in data generated may override the time requirement, as more data can develop more leads, and additional markers (such as SNPs) can increase the success of typing challenged samples. Until the turnaround time of MPS is comparable with that of CE, it would be advisable to maintain both CE and MPS capabilities in a laboratory.
- The volume of MPS-generated data appears daunting. However, commercial entities are providing streamlined pipelines capable of analyzing and managing the data. The forensic

science community will need to validate these software packages. Another data issue is that many forensic DNA analysts may be unfamiliar with new types of data analysis for estimating ancestry and reporting probabilities for hair and eye color. Training, education, and validation of these approaches will have to occur so that the forensic science community will understand data and report them properly.

- The magnitude of data from MPS is on a scale unprecedented in forensic genetics, and forensic scientists will need to become familiar and appreciate what goes on behind the scenes of bioinformatics with MPS data to develop and implement effective systems. It is unlikely that each crime laboratory will have a resident bioinformatician, so it will be necessary to develop facile and robust data and output processes. The current "needs" of the community for bioinformatics have yet to be fully realized. More tools, user-friendly interfaces, and single pipeline systems will be needed for more widespread MPS technology use for forensic analyses.
- The selection and use of genetic markers are critical because these markers impact the types of evidence that the forensic science community can analyze and the information that it may translate into viable investigative leads. Although MPS can analyze any type of genetic marker, it should accommodate the CODIS core STR loci regardless of the potential to expand the MPS STR marker set(s). More data are needed before selecting additional STR markers, if desired. Data to consider include genetic diversity, sequence variation, stutter rates, sequence motif, error rates, noise, and chemistry compatibility. Focus should be on reducing the amplicon size for more mini STRs. MPS allows for a reduction in STR amplicon size, thereby allowing the forensic science community to type more challenging samples.
- Although it is anticipated that a few laboratories may implement MPS as a standalone technology, there was overwhelming agreement that MPS technology will most likely be implemented primarily as an adjunct to current CE approaches. The primary reason provided for this strategy was the need for education and training regarding defined data use, workflow, validation, implementation, and interpretation. Additional reasons given for MPS as an adjunct to CE included other new technologies (e.g., rapid DNA typing) and MPS costs.
- MPS validation has yet to be described extensively for forensic applications. However, a number of studies are underway, and experience from early users suggests that MPS systems will meet requirements for forensic applications and be successfully implemented. It would be helpful to develop specific validation criteria for guidance within the forensic science community. Validation initially may be divided into chemistry, instrument, and bioinformatics systems, but, eventually, the entire system must be tested.

- There are other aspects of MPS to be addressed to facilitate technology transfer and make the best use of the system's power. These include marker selection, privacy concerns, legal admissibility, databases, and data storage. Some individuals have raised concerns that data generated by MPS could pose risks to privacy. The issue of privacy should not be anchored on MPS. MPS is just a tool for DNA sequencing. The privacy focus should be on the markers used for forensic typing, regardless of if they were typed by MPS, CE, or another methodology. Marker selection and privacy are not new topics in the forensic DNA typing arena. The forensic science community has raised little concern regarding privacy for the identity testing markers.
- MPS-generated STR data are likely to be entered into DNA databases because these resources already accommodate such data. Currently, the forensic science community does not enter identity SNP data into DNA databases for developing investigative leads. SNPs, however, potentially offer a higher success rate for typing degraded samples. Efforts to upload identity SNP data into DNA databases and implementing search tools that can accommodate SNP data would help support the law enforcement community in developing more investigative leads. Other markers such as ancestry informative SNPs and phenotypic SNPs require further consideration as to how and when to apply these markers. Clearly, these classes of SNPs can have investigative value in certain situations, most likely where law enforcement has not obtained any direct or indirect database search hits. Unlike identity markers, the consensus from the discussions is to not enter ancestry and phenotype SNPs into DNA databases. These types of SNPs have limited investigative power, and thus database searches based on these markers are not a good use of resources.

6. CONCLUSION

This webinar series addressed several aspects regarding (1) the MPS technology; (2) the potential applications that can be envisioned with MPS; (3) the logistics surrounding technology transfer into the forensic laboratory; and (4) issues about legal admissibility and privacy, real or perceived. MPS is an exciting technology, and all indications are that it eventually will become an important part of the forensic genetics toolbox. MPS technology enables DNA analyses ranging from traditional human identification approaches, to novel leads based on the phenotype of the unknown sample contributor, to characterization of animals and plants, to identification of microbes that may be used in an act of bioterrorism. This report summarizes the discussions and opinions of the participants and of viewers who provided input. The FTCoE intends that the information herein and the archived webinars will serve as an invaluable information resource to assist the forensic DNA community with embracing this exciting new technology

and/or to understand how it may assist in the missions of developing investigative leads, solving more cases, and exonerating the innocent through biological evidence.

REFERENCES

1. Lazaruk, K., P. S. Walsh, F. Oaks, D. Gilbert, B. B. Rosenblum, S. Menchen, et al. 1998. "Genotyping of Forensic Short Tandem Repeat (STR) Systems Based on Sizing Precision in a Capillary Electrophoresis Instrument." *Electrophoresis* 19:86–93.

2. Krenke, B. E., A. Tereba, S. J. Anderson, E. Buel, S. Culhane, C. J. Finis, C. S. Tomsey, J. M. Zachetti, A. Masibay, D. R. Rabbach, E. A. Amiott, and C. J. Sprecher. 2002. "Validation of a 16-Locus Fluorescent Multiplex System." *Journal of Forensic Sciences* 47(4):773–785.

3. Collins, P. J., L. K. Hennessy, C. S. Leibelt, R. K. Roby, D. J. Reeder, and P. A. Foxall. 2004. "Developmental Validation of a Single-Tube Amplification of the 13 CODIS STR Loci, D2S1338, D19S433, and Amelogenin: The AmpFlSTR Identifiler PCR Amplification Kit." *Journal of Forensic Sciences* 49:1265–1277.

4. Flores, S. K., J. Sun, J. King, and B. Budowle. 2014. "Validation of the GlobalFiler™ Express PCR Amplification Kit for the Direct Amplification of Single-Source DNA Samples on a High-Throughput Automated Workflow." *Forensic Science International: Genetics* 10:33–39.

5. Oostdik, K., K. Lenz, J. Nye, K. Schelling, D. Yet, S. Bruski, J. Strong, C. Buchanan, J. Sutton, J. Linner, N. Frazier, H. Young, L. Matthies, A. Sage, J. Hahn, R. Wells, N. Williams, M. Price, J. Koehler, M. Staples, K. L. Swango, C. Hill, K. Oyerly, W. Duke, L. Katzilierakis, M. G. Ensenberger, J. M. Bourdeau, C. J. Sprecher, B. Krenke, and D. R. Storts. 2014. "Developmental Validation of the PowerPlex® Fusion System for Analysis of Casework and Reference Samples: A 24-Locus Multiplex for New Database Standards." *Forensic Science International: Genetics* 12:69–76.

6. Thompson, J. M., M. M. Ewing, W. E. Frank, J. J. Pogemiller, C. A. Nolde, D. J. Koehler, A. M. Shaffer, D. R. Rabbach, P. M. Fulmer, C. J. Sprecher, and D. R. Storts. 2013. "Developmental Validation of the PowerPlex® Y23 System: A Single Multiplex Y-STR Analysis System for Casework and Database Samples." *Forensic Science International: Genetics* 7:240–250.

7. Berger, C, B. Berger, and W. Parson. 2012. "Sequence Analysis of the Canine Mitochondrial DNA Control Region from Shed Hair Samples in Criminal Investigations." *Methods in Molecular Biology* 830:331–348.

8. Budowle, B., and A. J. Eisenberg. 2007. "Forensic Genetics." In E*mery and Rimoin's Principles and Practice of Medical Genetics*, Fifth Edition, Vol. 1, edited by D. L. Rimoin, J. M. Connor, R. E. Pyeritz, and B. R. Korf, pp. 501–517. Philadelphia: Elsevier.

9. Kidd, K. K., A. J. Pakstis, W. C. Speed, E. L. Grigorenko, S. L. Kajuna, N. J. Karoma, S. Kungulilo, J. J. Kim, R. B. Lu, A. Odunsi, F. Okonofua, J. Parnas, L. O. Schulz, O. V. Zhukova, and J. R. Kidd. 2006. "Developing a SNP Panel for Forensic Identification of Individuals." *Forensic Science International* 164(1):20–32.

10. Pakstis, A. J., W. C. Speed, J. R. Kidd, and K. K. Kidd. 2007. "Candidate SNPs for a Universal Individual Identification Panel." *Human Genetics* 121(3–4):305–317.

11. Sanchez, J. J., C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, C. D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P. M. Schneider, A. Carracedo, and N. Morling. 2006. "A Multiplex Assay with 52 Single Nucleotide Polymorphisms for Human Identification." *Electrophoresis* 27(9):1713–1724.

12. Budowle, B., and A. van Daal. 2008. "Forensically Relevant SNP Classes." *Biotechniques* 44:603–610.

13. Gomes, C., M. Magalhães, C. Alves, A. Amorim, N. Pinto, and L. Gusmão. 2012. "Comparative Evaluation of Alternative Batteries of Genetic Markers to Complement Autosomal STRs in Kinship Investigations: Autosomal Indels vs. X-Chromosome STRs." *International Journal of Legal Medicine* 126(6):917–921.

14. Honda, K., L. Roewer, and P. de Knijff. 1999. "Male DNA Typing from 25-Year-Old Vaginal Swabs using Y Chromosomal STR Polymorphisms in a Retrial Request Case." *Journal of Forensic Sciences* 44(4):868–872.

15. Wilson, M. R., J. A. DiZinno, D. Polanskey, J. Replogle, and B. Budowle. 1995. "Validation of Mitochondrial DNA Sequencing for Forensic Casework Analysis." *International Journal of Legal Medicine* 108:68–74.

16. Hsieh, H. M., R. J. Hou, L. C. Tsai, C. S. Wei, S. W. Liu, L. H. Huang, Y. C. Kuo, A. Linacre, and J. C. Lee. 2003. "A Highly Polymorphic STR Locus in Cannabis Sativa." *Forensic Science International* 131(1):53–58.

17. Linacre, A., and S. S. Tobe. 2011. "An Overview to the Investigative Approach to Species Testing in Wildlife Forensic Science." *Investigative Genetics* 2:2.

18. Rasko, D. A., P. L. Worsham, T. G. Abshire, S. T. Stanley, J. D. Bannan, M. R. Wilson, R. J. Langham, R. S. Decker, L. Jiang, T. D. Read, A. M. Phillippy, S. L. Salzberg, M. Pop, M. N. van Ert, L. J. Kenefic, P. S. Keim, C. M. Fraser-Liggett, and J. Ravel. 2011. "Bacillus Anthracis Comparative Genome Analysis in Support of the Amerithrax Investigation." *Proceedings of the National Academy of Sciences of the USA* 108(12):5027–5032.

19. Budowle, B., T. R. Moretti, S. J. NiezgodJ, and B. L. Brown. 1998. "CODIS and PCR-Based Short Tandem Repeat Loci: Law Enforcement Tools." In *Second European Symposium on Human Identification 1998*, pp. 73–88. Madison, Wisconsin: Promega Corporation.

20. Martin, P. D., H. Schmitter, and P. M. Schneider. 2001. "A Brief History of the Formation of DNA Databases in Forensic Science within Europe." *Forensic Science International* 119(2):225–231.

21. CODIS. 2014. http://www.fbi.gov/about-us/lab/biometric-analysis/codis/ndis-statistics.

22. Oberacher, H., F. Pitterl, G. Huber, H. Niederstätter, M. Steinlechner, and W. Parson. 2008. "Increased Forensic Efficiency of DNA Fingerprints through Simultaneous Resolution of Length and Nucleotide Variability by High-Performance Mass Spectrometry." *Human Mutation* 29:427–432.

23. Pitterl, F., K. Schmidt, G. Huber, B, Zimmermann, R. Delport, S. Amory, et al. 2010. "Increasing the Discrimination Power of Forensic STR Testing by Employing High-Performance Mass Spectrometry, as Illustrated in Indigenous South African and Central Asian Populations." *International Journal of Legal Medicine* 124:551–558.

24. Planz, J. V., K. A. Sannes-Lowery, D. D. Duncan, S. Manalili, B. Budowle, R. Chakraborty, et al. 2012. "Automated Analysis of Sequence Polymorphism in STR Alleles by PCR and Direct Electrospray Ionization Mass Spectrometry." *Forensic Science International: Genetics* 6:594–606.

25. King, J. L., B. L. LaRue, N. Novroski, M. Stoljarova, S. B. Seo, X. Zeng, D. Warshauer, C. Davis, W. Parson, A. Sajantila, and B. Budowle. 2014. "High-Quality and High-Throughput Massively Parallel Sequencing of the Human Mitochondrial Genome using the Illumina MiSeq." *Forensic Science International: Genetics* 12:128–135.

26. Warshauer, D. H., J. L. King, and B. Budowle. 2015. "STRait Razor v2.0: The improved STR Allele Identification Tool—Razor" *Forensic Science International: Genetics* 14:182–186.

27. Warshauer, D. H., J. D. Churchill, N. Novroski, J. L. King, and B. Budowle. 2015. "Novel Y-Chromosome Short Tandem Repeat Variants Detected through the Use of Massively Parallel Sequencing." *Genomics, Proteomics, and Bioinformatics* 13(4):250–257.

28. Churchill, J. D., S. E. Schmedes, J. L. King, and B. Budowle. 2015. "Evaluation of the Illumina® Beta Version ForenSeq™ DNA Signature Prep Kit for Use in Genetic Profiling." *Forensic Science International: Genetics* 20:20–29.

29. Fordyce, S. L., M. C. Avila-Arcos, E. Rockenbauer, C. Borsting, R. Frank-Hansen, F. T. Peterson, et al. 2011. "High-Throughput Sequencing of Core STR Loci for Forensic Genetic Investigations using the Roche Genome Sequencer FLX Platform." *Biotechniques* 51:127–133.

30. Fordyce, S. L., H. S. Mogensen, C. Borsting, R. E. Lagacé, C. W. Chang, N. Rajagopalan, et al. 2015. "Second-Generation Sequencing of Forensic STRs using the Ion Torrent™ HID STR 10-Plex and the Ion PGM™." *Forensic Science International: Genetics* 14:132–140.

31. Scheible, M., O. Loreille, R. Just, and J. Irwin. 2014. "Short Tandem Repeat Typing on the 454 Platform: Strategies and Considerations for Targeted Sequencing of Common Forensic Markers." *Forensic Science International: Genetics* 12:107–119.

32. Zeng, X., J. L. King, M. Stoljarova, D. H. Warshauer, B. L. Larue, A. Sajantila, J. Patel, D. R. Storts, and B. Budowle. 2015. "High Sensitivity Multiplex Short Tandem Repeat loci Analyses with Massively Parallel Sequencing." *Forensic Science International: Genetics* 16:38–47.

33. Churchill, J. D., J. Chang, J. Ge, N. Rajagopalan, R. Lagacé, W. Liao, J. L. King, and B. Budowle. 2015. "Blind Study Evaluation Illustrates Utility of the Ion PGM™ System for Use in Human Identity DNA Typing." *Croatian Medical Journal* 56(3):218–229.

34. Seo, S. B., J. King, D. Warshauer, C. Davis, J. Ge, and B. Budowle. 2013. "Single Nucleotide Polymorphism Typing with Massively Parallel Sequencing for Human Identification." *International Journal of Legal Medicine* 127(6):1079–1086.

35. National Institute of Justice, Office of Investigative and Forensic Sciences. 2015. *Familial DNA Searching: Current Approaches, Final Report.* https://rti.connectsolutions.com/p49iz1rzbpi/.

36. Lander, E. S., et al. 2001. "Human Genome." *Nature* 409:860–921.

37. Quail, M. A., M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, A. Bertoni, H. P. Swerdlow, and Y. Gu. 2012. "A Tale of Three Next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq Sequencers." *BMC Genomics* 13:341.

38. Jünemann, S., F. J. Sedlazeck, K. Prior, A. Albersmeier, U. John, J. Kalinowski, A. Mellmann, A. Goesmann, A. von Haeseler, J. Stoye, and D. Harmsen. 2013. "Updating Benchtop Sequencing Performance Comparison." *Nature Biotechnology* 31:294–296.

39. Brenner, S., S. R. Williams, E. H. Vermaas, T. Storck, K. Moon, C. McCollum, J. L. Mao, S. Luo, J. J. Kirchner, S. Eletr, R. B. DuBridge, T. Burcham, and G. Albrecht. 2000. "In Vitro Cloning of Complex Mixtures of DNA on Microbeads: Physical Separation of Differentially Expressed cDNAs." *Proceedings of the National Academy of Sciences of the USA* 97(4):1665–1670.

40. Margulies, M., M. Egholm W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. 2005. "Genome Sequencing in Microfabricated High-Density Picolitre Reactors." *Nature* 437(7057):376–380.

41. Merriman, B., and J. M. Rothberg. 2012. "Progress in Ion Torrent Semiconductor Chip-Based Sequencing." *Electrophoresis* 33(23):3397–3417.

42. Metzker, M. L. 2010. "Sequencing Technologies—The Next Generation." *Nature Reviews Genetics* 11(1):31–46.

43. Quail, M. A., I. Kozarewa, F. Smith, A. Scally, P. J. Stephens, R. Durbin, H. Swerdlow, and D. J. Turner. 2008. "A Large Genome Center's Improvements to the Illumina Sequencing System." *Nature Methods* 5:1005–1010.

44. Shendure, J., and H. Ji. 2008. "Next-Generation DNA Sequencing." *Nature Biotechnology* 26(10):1135–1145.

45. Syed, F., H. Gruenwald, and N. Caruccio. 2009. "Next-Generation Sequencing Library Preparation: Simultaneous Fragmentation and Tagging using *In Vitro* Transposition. *Nature Methods* 6. http://www.nature.com/nmeth/journal/v6/n11/full/nmeth.f.272.html.

46. Fendt, L., B. Zimmermann, M. Daniaux, and W. Parson. 2009. "Sequencing Strategy for the Whole Mitochondrial Genome Resulting in High-Quality Sequences." *BMC Genomics* 10:139.

47. Parson, W., C. Strobl, G. Huber, B. Zimmermann, S. M. Gomes, L. Souto, L. Fendt, R. Delport, R. Langit, S. Wootton, R. Lagacé, and J. Irwin. 2013. "Evaluation of Next Generation mtGenome Sequencing using the Ion Torrent Personal Genome Machine (PGM)." *Forensic Science International: Genetics* 7:543–549.

48. King, J. L., B. L. LaRue, N. Novroski, M. Stoljarova, S. B. Seo, X. Zeng, D. Warshauer, C. Davis, W. Parson, A. Sajantila, and B. Budowle. 2014. "High-Quality and High-Throughput Massively Parallel Sequencing of the Human Mitochondrial Genome using the Illumina MiSeq." *Forensic Science International: Genetics* 12:128–135.

49. Seo, S. B., X. Zeng, J. L. King, B. L. Larue, M. Assidi, M. H. Al-Qahtani, A. Sajantila, and B. Budowle. 2015. "Underlying Data for Sequencing the Mitochondrial Genome with the Massively Parallel Sequencing Platform Ion Torrent™ PGM™." *BMC Genomics* 16(Suppl. 1):S4.

50. Hori, M., H. Fukano, and Y. Suzuki. 2007. "Uniform Amplification of Multiple DNAs by Emulsion PCR." *Biochemical and Biophysical Research Communications* 352:323–328.

51. Hodges, E., Z. Xuan, V. Balija, M. Kramer, M. N. Molla, S. W. Smith, C. M. Middle, M. J. Rodesch, T. J. Albert, G. J. Hannon, and W. R. McCombie. 2007. "Genome-Wide *In Situ* Exon Capture for Selective Resequencing." *Nature Genetics* 39:1522–1527.

52. Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D. B. Jaffe, E. S. Lander, and C. Nusbaum. 2009. "Solution Hybrid Selection with Ultra-Long Oligonucleotides for Massively Parallel Targeted Sequencing." *Nature Biotechnology.* 27:182–189.

53. Kozarewa, I., Z. Ning, M. A. Quail, M. J. Sanders, M. Berriman, and D. J. Turner. 2009. "Amplification-Free Illumina Sequencing—Library Preparation Facilitates Improved Mapping and Assembly of (G+C)-Biased Genomes." *Nature Methods* 6:291–295.

54. Tewhey, R., J. B. Warner, M. Nakano, B. Libby, M. Medkova, P. H. David, S. K. Kotsopoulos, M. L. Samuels, J. B. Hutchison, J. W. Larson, E. J. Topol, M. P. Weiner, O. Harismendy, J. Olson, D. R. Link, and K. A. Frazer. 2009. "Microdroplet-Based PCR Enrichment for Large-Scale Targeted Sequencing." *Nature Biotechnology* 27:1025–1031.

55. Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, J. Shendure, and D. J. Turner. 2010. "Target-Enrichment Strategies for Next-Generation Sequencing." *Nature Methods* 7:111–118.

56. Head, S. R., H. K. Komori, S. A. LaMere, T. Whisenant, F. van Nieuwerburgh, D. R. Salomon, and P. Ordoukhanian. 2014. "Library Construction for Next-Generation Sequencing: Overviews and Challenges." *Biotechniques* 56(2):61–68.

57. Van Tassell, C. P., T. P. Smith, L. K. Matukumalli, J. F. Taylor, R. D. Schnabel, C. T. Lawley, C. D. Haudenschild, S. S. Moore, W. C. Warren, and T. S. Sonstegard. 2008. "SNP Discovery and Allele Frequency Estimation by Deep Sequencing of Reduced Representation Libraries." *Nature Methods* 5:247–252.

58. Knapp, M., M. Stiller, and M. Meyer. 2012. "Generating Barcoded Libraries for Multiplex High-Throughput Sequencing." *Methods in Molecular Biology* 840:155–170.

59. Hamady, M., J. J. Walker, J. K. Harris, N. J. Gold, and R. Knight. 2012. "Error-correcting Barcoded Primers Allow Hundreds of Samples to be Pyrosequenced in Multiplex." *Nature Methods* 5:235–237.

60. Adessi, C., G. Matton, G. Ayala, G. Turcatti, J. J. Mermod, P. Mayer, and E. Kawashima. 2000. "Solid Phase DNA Amplification: Characterisation of Primer Attachment and Amplification Mechanisms." *Nucleic Acids Reearch.* 28(20):E87.

61. Fedurco, M., A. Romieu, S. Williams, I. Lawrence, and G. Turcatti. 2006. "BTA, a Novel Reagent for DNA Attachment on Glass and Efficient Generation of Solid-Phase Amplified DNA Colonies." *Nucleic Acids Research* 34(3):e22.

62. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. "DNA Sequencing with Chain-Terminating Inhibitors." *Proceedings of the National Academy of Sciences of the USA* 74(12):5463–5467.

63. Turcatti, G., A. Romieu, M. Fedurco, and A. P. Tairim. 2008. "A New Class of Cleavable Fluorescent Nucleotides: Synthesis and Optimization as Reversible Terminators for DNA Sequencing by Synthesis." *Nucleic Acids Research* 36(4):e25.

64. Rothberg, J. M., W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth, and J. Bustillo. 2011. "An Integrated Semiconductor Device Enabling Non-Optical Genome Sequencing." *Nature* 475:348–352.

65. Loman, N. J., R. V. Misra, T. J. Dallman, C. Constantinidou, S. E. Gharbia, J. Wain, and M. J. Pallen. 2012 "Performance Comparison of Benchtop High-Throughput Sequencing Platforms." *Nature Biotechnology* 30:434–439.

66. Bragg, L. M., G. Stone, M. K. Butler, P. Hugenholtz, and G. W. Tyson. 2013. "Shining a Light on Dark Sequencing: Characterising Errors in Ion Torrent PGM Data." *PLOS Computational Biology* 9(4):e1003031.

67. Vander Horn, P. B., A. Kraltcheva, G. Luo, M. Landes, S. Chen, K. Heinemann, T. Nikiforov, J. Shirley, E. Tozer, and D. Mazur. 2013. "The Ion PGM™ Hi-Q™ Sequencing Polymerase: Reducing Systematic Error, Increasing Accuracy, and Improving Read-Length." Ion Torrent, Thermo Fisher Scientific, Life Sciences Solutions Group, Carlsbad, CA, USA. http://www.ashg.org/2014meeting/abstracts/fulltext/f140121252.htm.

68. McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data." *Genome Research* 20:1297–1303.

69. Broad Institute. 2013. "Integrative Genomics Viewer." http://www.broadinstitute.org/igv/.

70. Van Neste, C., Y. Gansemans, D. De Coninck, D. Van Hoofstat, W. Van Criekinge, D. Deforce, and F. Van Nieuwerburgh 2015. "Forensic Massively Parallel Sequencing Data Analysis Tool: Implementation of MyFLq as a Standalone Web- and IlluminaBaseSpace®-Application." *Forensic Science International: Genetics* 15:2–7.

71. Carpenter, M. L., J. D. Buenrostro, C. Valdiosera, H. Schroeder, M. E. Allentoft, M. Sikora, M. Rasmussen, S. Gravel, S. Guillén, G. Nekhrizov, K. Leshtakov, D. Dimitrova, N. Theodossiev, D. Pettener, D. Luiselli, K. Sandoval, A. Moreno-Estrada, Y. Li, J. Wang, M. T. Gilbert, E. Willerslev, W. J. Greenleaf, and C. D. Bustamante. 2013. "Pulling Out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries." *American Journal of Human Genetics* 93(5):852–864.

72. Templeton, J. E. L., P. M. Brotherton, B. Llamas, J. Soubrier, W. Haak, A. Cooper, and J. J. Austin. 2013. "DNA Capture and Next-Generation Sequencing can Recover Whole Mitochondrial Genomes from Highly Degraded Samples for Human Identification." *Investigative Genetics* 4:26.

73. Parson, W., and A. Dür. 2007. "EMPOP—A Forensic mtDNA Database." *Forensic Science International: Genetics* 1(2):88–92.

74. Rehm, H. L., S. J. Bale, P. Bayrak-Toydemir, J. S. Berg, K. K. Brown, J. L. Deignan, M. J. Friez, B. H. Funke, M. R. Hegde, E. Ly; Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. 2013. "ACMG Clinical Laboratory Standards for Next-Generation Sequencing. *Genetics in Medicine* 15:733–747.

75. Ellard, S., H. Lindsay, N. Camm, C. Watson, S. Abbs, G. R., Taylor, and R. Charlton. 2012. "Practice Guidelines for Targeted Next Generation Sequencing Analysis and Interpretation." Clinical Molecular Genetics Society. http://www.acgs.uk.com/media/815227/bpg_for_targeted_next_generation_sequencing_2011134.pdf.

76. Gargis, A. S., L. Kalman, M. W. Berry, D, P. Bick, D. P. Dimmock, T. Hambuch, F. Lu, E. Lyon, K. V. Voelkerding, B. A. Zehnbauer, R. Agarwala, S. F. Bennett, B. Chen, E. L. Chin, J. G. Compton, S. Das, D. H. Farkas, M. J. Ferber, B. H. Funke, M. R. Furtado, L. M. Ganova-Raeva, U. Geigenmüller, S. J. Gunselman, M. R. Hegde, P. L. Johnson, A. Kasarskis, S. Kulkarni, T. Lenk, C. S. Liu, M. Manion, T. A. Manolio, E. R. Mardis, J. D. Merker, M. S. Rajeevan, M. G. Reese, H. L. Rehm, B. B. Simen, J. M. Yeakley, J. M. Zook, and I. M. Lubin. 2012. "Assuring the Quality of Next-Generation Sequencing in Clinical Laboratory Practice." *Nature Biotechnology.* 30:1033–1036.

77. Pont-Kingdon, G., F. Gedge, W. Wooderchak-Donahue, I. Schrijver, K. E. Weck, J. A. Kant, D. Oglesbee, P. Bayrak-Toydemir, E. Lyon, Biochemical and Molecular Genetic Resource Committee of the College of American Pathologists. 2012. "Design and Analytical Validation of Clinical DNA Sequencing Assays." *Archives of Pathology and Laboratory Medicine* 136:41–46.

78. Budowle, B., N. D. Connell, A. Bielecka-Oder, R. R. Colwell, C. R. Corbett, J. Fletcher, M. Forsman, D. R. Kadavy, A. Markotic, S. A. Morse, R. S. Murch, A. Sajantila, S. E. Schmedes, K. L. Ternus, S. D. Turner, and S. Minot. 2014. "Validation of High Throughput Sequencing and Microbial Forensics Applications." *Investigative Genetics* 5:9.

Appendix A: Literature Review

Short Tandem Repeat (STR) Analysis

Bornman, D. M., M. E. Hester, J. M. Schuetter, M. D. Kasoji, A. Minard-Smith, C. A. Barden, S. C. Nelson, G. D. Godbold, C. H. Baker, B. Yang, J. E. Walther, I. E. Tornes, P. S. Yan, B. Rodriguez, R. Bundschuh, M. L. Dickens, B. A. Young, and S. A. Faith. 2012. "Short-Read, High-Throughput Sequencing Technology for STR Genotyping." *Biotech Rapid Dispatches* Apr:1–6. doi: 10.2144/000113857.

Churchill, J. D., S. E. Schmedes, J. L. King, and B. Budowle. 2016. "Evaluation of the Illumina® Beta Version ForenSeq™ DNA Signature Prep Kit for Use in Genetic Profiling." *Forensic Science International: Genetics* 20:20–29.

Divne, A. M., H. Edlund, and M. Allen. 2010. "Forensic Analysis of Autosomal STR Markers using Pyrosequencing." *Forensic Science International: Genetics* 4:122–129.

Fordyce, S. L., H. S. Mogensen, C. Børsting, R. E. Lagacé, C.-W. Chang, N. Rajagopalan, and N. Morling. 2015. "Second-Generation Sequencing of Forensic STRs using the Ion Torrent™ HID STR 10-Plex and the Ion PGM™." *Forensic Science International: Genetics* 14:132–140.

Fordyce, S. L., M. C. Avila-Arcos, E. Rockenbauer, C. Borsting, R. Frank-Hansen, F. T. Petersen, E. Willerslev, A. J. Hansen,N. Morling, and M. T. Gilbert. 2011. "High-Throughput Sequencing of Core STR Loci for Forensic Genetic Investigations using the Roche Genome Sequencer FLX Platform." *BioTechniques* 51:127–133.

Gelardi, C., E. Rockenbauer, S. Dalsgaard, C. Børsting, and N. Morling. 2014. Second Generation Sequencing of Three STRs D3S1358, D12S391 and D21S11 in Danes and a New Nomenclature for Sequenced STR Alleles. *Forensic Science International: Genetics* 12:38–41.

Kim, E. H., H. Y. Lee, I. S. Yang, S. E. Jung, W. I. Yang, and K. J. Shin. 2016. "Massively Parallel Sequencing of 17 Commonly Used Forensic Autosomal Strs and Amelogenin with Small Amplicons." *Forensic Science International: Genetics* 22:1–7.

Scheible, M., O. Loreille, R. Just, and J. Irwin. 2014. "Short Tandem Repeat Typing on the 454 Platform: Strategies and Considerations for Targeted Sequencing of Common Forensic Markers." *Forensic Science International: Genetics* 12:107–119.

Warshauer, D. H., J. D. Churchill, N. M. M. Novroski, J. L. King, and B. Budowle. 2015. "Novel Y-Chromosome Short Tandem Repeat Variants Detected Through the Use of Massively Parallel Sequencing." *Genomics, Proteomics & Bioinformatics* 13(4):250–257.

Zeng, X., J. L. King, M. Stoljarova, D. H. Warshauer, B. L. LaRue, A. Sajantila, J. Patel, D. R. Storts, and B. Budowle. 2015. "High Sensitivity Multiplex Short Tandem Repeat Loci Analyses with Massively Parallel Sequencing." *Forensic Science International: Genetics* 16:38–47.

Zeng, X., J. King, S. Hermanson, J. Patel, D. R. Storts, and B. Budowle. 2015. "An Evaluation of the PowerSeq™ Auto system: A Multiplex Short Tandem Repeat Marker Kit Compatible with Massively Parallel Sequencing." *Forensic Science International: Genetics* 19:172–179.

Single Nucleotide Polymorphisms (SNPs)

Kayser, M., and P. de Knijff. 2011. "Improving Human Forensics through Advances in Genetics, Genomics and Molecular Biology." *Nature Reviews Genetics* 12:179–192.

Human Identification (HID)

Børsting, C., S. L. Fordyce, J. Olofsson, H. S. Mogensen, and N. Morling. 2014. "Evaluation of the Ion Torrent™ HID SNP 169-Plex; A SNP Typing Assay Developed for Human Identification by Second Generation Sequencing." *Forensic Science International: Genetics* 12:144–154.

Budowle, B., and A. Van Daal. 2008. "Forensically Relevant SNP Classes." *BioTechniques 25th Anniversary* 44:603–610.

Eduardoff, M., C. Santos, M. de la Puente, T. Gross, M. Fondevila, C. Strobl, B. Sobrino, D. Ballard, PJ. M. Schneider, A. Carracedo, M. V. Lareu, W. Parson, and C. Phillips. 2015. "Inter-Laboratory Evaluation of SNP-Based Forensic Identification by Massively Parallel Sequencing using the Ion PGM™." *Forensic Science International: Genetics* 17:110–121.

Gettings, K. B., K. M. Kiesler, and P. M. Vallone. 2015. "Performance of a Next Generation Sequencing SNP Assay on Degraded DNA." *Forensic Science International: Genetics* 19:1–9.

Pakstis, A. J., W. C. Speed, R. Fang, F. C. Hyland, M. R. Furtado, J. R. Kidd, and K. K. Kidd. 2010. "SNPs for a Universal Individual Identification Panel." *Human Genetics* 127:315–324.

Seo, S. B., J. L. King, D. H. Warshauer, C. P. Davis, J. Ge, and B. Budowle. 2013. "Single Nucleotide Polymorphism Typing with Massively Parallel Sequencing for Human Identification." *International Journal of Legal Medicine* 1–8.

Sobrino, B., M. Brión, and A. Carracedo. 2005. "SNPs in Forensic Genetics: A Review on SNP Typing Methodologies." *Forensic Science International* 154:181–194.

Warshauer, D., C. Davis, C. Holt, Y. Han, P. Walichiewicz, T. Richardson, et al. 2014. "Massively Parallel Sequencing of Forensically Relevant Single Nucleotide Polymorphisms using TruSeq™ Forensic Amplicon." *International Journal of Legal Medicine* 1–6.

Xue, Y., and C. Tyler-Smith. 2010. "The Hare and the Tortoise: One Small Step for Four SNPs, One Giant Leap for SNP-Kind." *Forensic Science International: Genetics* 4:59–61.

Xue, Y., Q. Wang, Q. Long, B. L. Ng, H. Swerdlow, J. Burton, C. Skuce, R. Taylor, Z. Abdellah, Y. Zhao, D. G. MacArthur, M. A. Quail, N. P. Carter, H. Yang, and C. Tyler-Smith. 2009. "Human Y Chromosome Base-Substitution Mutation Rate Measured by Direct Sequencing in a Deep-Rooting Pedigree." *Current Biology* 19:1453–1457.

Phenotypic Determination

Kayser, M., and P. M. Schneider. 2009. "DNA-Based Prediction of Human Externally Visible Characteristics in Forensics: Motivations, Scientific Challenges, and Ethical Considerations." *Forensic Science International: Genetics* 3:154–161.

Keating, B., A. T. Bansal, S. Walsh, J. Millman, J. Newman, K. Kidd, B. Budowle, A. Eisenberg, J. Donfack, P. Gasparini, Z. Budimlija, A. K. Henders, H. Chandrupatla, D. L. Duffy, S. D. Gordon, P. Hysi, F. Liu, S. E. Medland, L. Rubin, N. G. Martin, T. D. Spector, M. Kayser; and International Visible Trait Genetics (VisiGen) Consortium. 2013. "First All-in-One Diagnostic Tool for DNA Intelligence: Genome-Wide Inference of Biogeographic Ancestry, Appearance, Relatedness, and Sex with the Identitas v1 Forensic Chip." *International Journal of Legal Medicine* 127:559–572.

Pulker, H., M. V. Lareu, C. Phillips, and A. Carracedo. 2007. "Finding Genes that Underlie Physical Traits of Forensic Interest Using Genetic Tools." *Forensic Science International: Genetics* 1:100–104.

Ruiz, Y., C. Phillips, A. Gomez-Tato, J. Alvarez-Dios, M. C. de Cal, R. Cruz, et al. 2013. "Further Development of Forensic Eye Color Predictive Tests." *Forensic Science International: Genetics* 7:28–40.

Walsh, S., L. Chaitanya, L. Clarisse, L. Wirken, J. Draus-Barini, L. Kovatsi, H. Maeda, T. Ishikawa, T. Sijen, PJ. De Knijff, W. Branicki, F. Liu, and M. Kayser. 2014. "Developmental Validation of the HIrisPlex System: DNA-Based Eye and Hair Colour Prediction for Forensic and Anthropological Usage." *Forensic Science International: Genetics* 9:150–161.

Ancestry

Fondevila, M., C. Phillips, C. Santos, A. F. Aradas, P. Vallone, J. Butler, M. V. Lareu, and A. Carracedo. 2013. "Revision of the SNPforID 34-Plex Forensic Ancestry Test: Assay Enhancements, Standard Reference Sample Genotypes and Extended Population Studies." *Forensic Science International: Genetics* 7:63–74.

Kidd, K. K., W. C. Speed, A. J. Pakstis, M. R. Furtado, R. Fang, A. Madbouly, M. Maiers, M. Middha, F. R. Friedlaender, and J. R. Kidd. 2014. "Progress Toward an Efficient Panel of SNPs for Ancestry Inference." *Forensic Science International: Genetics* 10:23–32.

Nassir, R., R. Kosoy, C. Tian, P. A. White, L. M. Butler, G. Silva, R. Kittles, M. E. Alarcon-Riquelme, P. K. Gregersen, J. W. Belmont, F. M. De La Vega, and M. F. Seldin. 2009. "An Ancestry Informative Marker Set for Determining Continental Origin: Validation and Extension using Human Genome Diversity Panels." *BMC Genetics* 10:39.

Yang, N., H. Li, L. A. Criswell, P. K. Gregersen, M. E. Alarcon-Riquelme, R. Kittles, R. Shigeta, G. Silva, P. I. Patel, J. W. Belmont, and M. F. Seldin. 2005. "Examination of Ancestry and Ethnic Affiliation using Highly Informative Diallelic DNA Markers: Application to Diverse and Admixed Populations and Implications for Clinical Epidemiology and Forensic Medicine." *Human Genetics* 118:382–392.

Zeng, X., R. Chakraborty, J. King, B. LaRue, R. Moura-Neto, and B. Budowle. 2015. "Selection of Highly Informative SNP Markers for Population Affiliation of Major U.S. Populations." *International Journal of Legal Medicine* 1–12.

Mitochondrial DNA (mtDNA)

Bintz, B. J., G. B. Dixon, and M. R. Wilson. 2014. "Simultaneous Detection of Human Mitochondrial DNA and Nuclear-Inserted Mitochondrial-Origin Sequences (NumtS) using Forensic mtDNA Amplification Strategies and Pyrosequencing Technology." *Journal of Forensic Sciences* 59(4):1064–1073.

Bodner, M., A. Iuvaro, C. Strobl, S. Nagl, G. Huber, S. Pelotti, D. Pettener, D. Luiselli, and W. Parson. 2015. "Helena, the Hidden Beauty: Resolving the Most Common West Eurasian mtDNA Control Region Haplotype by Massively Parallel Sequencing an Italian Population Sample." *Forensic Science International: Genetics* 15:21–26.

Davis, C., D. Peters, D. Warshauer, J. King, and B. Budowle. 2014. "Sequencing the Hypervariable Regions of Human Mitochondrial DNA using Massively Parallel Sequencing: Enhanced Data Acquisition for DNA Samples Encountered in Forensic Testing." *Legal Medicine* 17(2):123–127.

Gunnarsdottir, E. D., M. Li, M. Bauchet, K. Finstermeier, and M. Stoneking. 2011. "High Through-Put Sequencing of Complete Human mtDNA Genomes from the Philippines." *Genome Research* 21:1–11.

King, J. L., B. L. LaRue, N. M. Novroski, M. Stoljarova, S. B. Seo, X. Zeng, D. H. Warshauer, C. P. Davis, W. Parson, A. Sajantila, and B. Budowle. 2014. "High-Quality and High-Throughput Massively Parallel Sequencing of the Human Mitochondrial Genome using the Illumina MiSeq." *Forensic Science International: Genetics* 12:128–135.

McElhoe, J. A., M. M. Holland, K. D. Makova, M. S. Su, I. M. Paul, C. H. Baker, S. A. Faith, and B. Young. 2014. "Development and Assessment of an Optimized Next-Generation DNA Sequencing Approach for the mtGenome using the Illumina MiSeq." *Forensic Science International: Genetics* 13:20–29.

Mikkelsen, M., R. F. Hansen, A. J. Hansen, and N. Morling. 2014. "Massively Parallel Pyrosequencing 454 Methodology of the Mitochondrial Genome in Forensic Genetics." *Forensic Science International: Genetics* 12:30–37.

Parson, W., C. Strobl, G. Huber, B. Zimmermann, S. M. Gomes, L. Souto, L. Fendt, . Delport, R. Langit, S. Wootton, R. Lagacé, and J. Irwin. 2013. "Evaluation of Next Generation mtGenome Sequencing using the Ion Torrent Personal Genome Machine (PGM)." *Forensic Science International: Genetics* 7:543–549.

Parson, W., G. Huber, L. Moreno, M.-B. Madel, M. D. Brandhagen, S. Nagl, C. Xavier, M. Eduardoff, T. C. Callaghan, and J. A. Irwin. 2015. "Massively Parallel Sequencing of Complete Mitochondrial Genomes from Hair Shaft Samples." *Forensic Science International: Genetics* 15:8–15.

Seo, S. B., X. Zeng, J. L. King, B. L. Larue, M. Assidi, M. H. Al-Qahtani, A. Sajantila, and B. Budowle. 2015. "Underlying Data for Sequencing the Mitochondrial Genome with the Massively Parallel Sequencing Platform Ion Torrent™ PGM™." *BMC Genomics* S4.

Yishu Zhou, Fei Guo, Jiao Yu, Feng Liu, Jinling Zhao, Hongying Shen, Bin Zhao, Fei Jia, Zhu Sun, He Song, and Xianhua Jiang. 2016. "Strategies for Complete Mitochondrial Genome Sequencing on Ion Torrent PGM™ Platform in Forensic Sciences." *Forensic Science International: Genetics* 22:11–21.

Bioinformatics

King, J. L., A. Sajantila, and B. Budowle. 2014. "mitoSAVE: Mitochondrial Sequence Analysis of Variants in Excel." *Forensic Science International: Genetics* 12:122–125.

Warshauer, D. H., D. Lin, K. Hari, R. Jain, C. Davis, B. LaRue, J. King, and B. Budowle. 2013. "STRait Razor: A Length-Based Forensic STR Allele-Calling Tool for Use with Second Generation Sequencing Data." *Forensic Science International: Genetics* 7:409–417.

Warshauer, D. H., J. L. King, and B. Budowle. 2014. "STRait Razor v2. 0: The Improved STR Allele Identification Tool–Razor." *Forensic Science International: Genetics* 14:182–186.

Other Marker Types: Molecular Autopsy, Metagenomics, and Microhaplotypes

Apellaniz-Ruiz, M., C. Gallego, S. Ruiz-Pinto, A. Carracedo, and C. Rodríguez-Antona. 2015. "Human Genetics: International Projects and Personalized Medicine." *Drug Metabolism and Personalized Therapy* 31(1):3–8.

Budowle, B., N. D. Connell, A. Bielecka-Oder, R. R. Colwell, C. R. Corbett, J. Fletcher, M. Forsman, D. R. Kadavy, A. Markotic, S. A. Morse, R. S. Murch, A. Sajantila, S. E. Schmedes, K. L. Ternus, S. D. Turner, and S. Minot. 2014. "Validation of High Throughput Sequencing and Microbial Forensics Applications." *BMC Investigative Genetics* 5:9.

Cummings, C.A., Bormann-Chung, C.A., Fang, R., Barker, M., Brzoska, P., Williamson, P.C., Beaudry, J., Matthews, M., Schupp, J., Wagner, D.M., Birdsell, D., Vogler, A.J., Furtado, M.R., Keim P., and Budowle, B. 2010. "Accurate, Rapid, and High-Throughput Detection of Strain-Specific Polymorphisms in Bacillus Anthracis and Yersinia Pestis by Next-Generation Sequencing. *BMC Investigative Genetics* 1:5.

Franzosa, E. A., K. Huang, J. F. Meadow, D. Gevers, K. P. Lemon, B. J. M. Bohannan, and C. Huttenhower. 2015. "Identifying Personal Microbiomes using Metagenomic Codes." Proceedings of the National Academy of Sciences.

Hertz, C., S. Christiansen, L. Ferrero-Miliani, M. Dahl, P. Weeke, G. Ottesen, R. Frank-Hansen, H. Bundgaard, and N. Morling. 2016. "Next-Generation Sequencing of 100 Candidate Genes in Young Victims of Suspected Sudden Cardiac Death with Structural Abnormalities of the Heart." *International Journal of Legal Medicine* 130:91–102.

Hertz, C. L., S. L. Christiansen, L. Ferrero-Miliani, S. L. Fordyce, M. Dahl, A. G. Holst, et al., 2014. "Next-Generation Sequencing of 34 Genes in Sudden Unexplained Death Victims in Forensics and in Patients with Channelopathic Cardiac Diseases." *International Journal of Legal Medicine* 1–8.

Kidd, K. K., A. J. Pakstis, W. C. Speed, R. Lagacé, J. Chang, S. Wootton, E. Haigh, and J. R. Kidd. 2014. "Current Sequencing Technology makes Microhaplotypes a Powerful New Type of Genetic Marker for Forensics." *Forensic Science International: Genetics* 12:215–224.

Kidd, K. K., and W. C. Speed. 2015. "Criteria for Selecting Microhaplotypes: Mixture Detection and Deconvolution." *Investigative Genetics* 6:1.

Koski, A., I. Ojanperä, J. Sistonen, E. Vuori, and A. Sajantila. 2007. "A Fatal Doxepin Poisoning Associated with a Defective CYP2D6 Genotype." *The American Journal of Forensic Medicine and Pathology* 28:259–261.

Lunetta, P., A. Levo, A. Männikkö, A. Penttilä, and A. Sajantila. 2002. "Death in Bathtub Revisited with Molecular Genetics: A Victim with Suicidal Traits and a LQTS Gene Mutation." *Forensic Science International* 130:122–124.

Santori, M., A. Blanco-Verea, R. Gil, J. Cortis, K. Becker, P. M. Schneider, et al. 2015. "Broad-Based Molecular Autopsy: A Potential Tool to Investigate the Involvement of Subtle Cardiac Conditions in Sudden Unexpected Death in Infancy and Early Childhood." *Archives of Disease in Childhood* 100(10):952–956.

Weber-Lehmann, J., E. Schilling, G. Gradl, D. Richter, J. Wiehler, and B. Rolf. 2014. "Finding the Needle in the Haystack: Differentiating "Identical" Twins in Paternity Testing and Forensics by Ultra-Deep Next Generation Sequencing." *Forensic Science International: Genetics* 9:42–6. doi: 10.1016/j.fsigen.2013.10.015. Epub 2013 Nov 8.

Technology

Brenner, S., S. R. Williams, E. H .Vermaas, T. Storck, K. Moon, C. McCollum, J. I. Mao, S. Luo, J. J. Kirchner, S. Eletr, R. B. DuBridge, T. Burcham, and G. Albrecht. 2000. "In Vitro Cloning of Complex Mixtures of DNA on Microbeads: Physical Separation of Differentially Expressed cDNAs." *Proceedings of the National Academy of Sciences* 97(4):1665–1670.

E. C. Berglund, A. Kiialainen, and A. C. Syvänen. 2011. "Next-Generation Sequencing Technologies and Applications for Human Genetic History and Forensics." *Investigative Genetics* 2:23.

Hamady, M., J. J. Walker, J. K. Harris, N. J. Gold, and R. Knight. 2012. "Error-Correcting Barcoded Primers Allow Hundreds of Samples to be Pyrosequenced in Multiplex." *Nature Methods* 5:235–237.

Head, S. R., H. K. Komori, S. A. LaMere, T. Whisenant, F. van Nieuwerburgh, D. R. Salomon, and P. Ordoukhanian. 2014. "Library Construction for Next-Generation Sequencing: Overviews and Challenges." *Biotechniques* 56(2):61–68.

Hori, M., H. Fukano, and Y. Suzuki. 2007. "Uniform Amplification of Multiple DNAs by Emulsion PCR." *Biochemical and Biophysical Research Communications* 352:323–328.

Jünemann, S., F. J. Sedlazeck, K. Prior, A. Albersmeier, U. John, J. Kalinowski, A. Mellmann, A. Goesmann, A. von Haeseler, J. Stoye, and D. Harmsen. 2013. "Updating Benchtop Sequencing Performance Comparison." *Nature Biotechnology.* 31:294–296.

Knapp, M., M. Stiller, and M. Meyer. 2012. "Generating Barcoded Libraries for Multiplex High-Throughput Sequencing." *Methods in Molecular Biology* 840:155–170.

Loman, N. J., R. V. Misra, T. J. Dallman, C. Constantinidou, S. E. Gharbia, J. Wain, and M. J. Pallen. 2012. "Performance Comparison of Benchtop High-Throughput Sequencing Platforms." *Nature Biotechnology* 30:434–439.

Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, J. Shendure, and D. J. Turner. 2010. "Target-Enrichment Strategies for Next-Generation Sequencing." *Nature Methods* 7:111–118.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F.

Begley, and J. M. Rothberg. 2005. “Genome Sequencing in Microfabricated High-Density Picolitre Reactors.” *Nature* 437(7057):376–380.

Merriman, B., and J. M. Rothberg. 2012. “Progress in Ion Torrent Semiconductor Chip Based Sequencing.” *Electrophoresis* 33(23):3397–3417.

Metzker, M. L. 2010. “Sequencing Technologies—The Next Generation.” *Nature Reviews Genetics* 11(1):31–46.

Quail, M. A., M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, A. Bertoni, H. P. Swerdlow, and Y. Gu. 2012. “A Tale of Three Next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq seQuencers.” *BMC Genomics* 13:341.

Rothberg, J. M., W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth, and J. Bustillo. 2011. “An Integrated Semiconductor Device Enabling Non-Optical Genome Sequencing.” *Nature* 475:348–352.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. “DNA sEquencing with Chain-Terminating Inhibitors.” *Proceedings of the National Academy of Sciences* 74(12):5463–5467.

Shendure, J., and H. Ji. 2008. “Next-Generation DNA Sequencing.” *Nature Biotechnology* 26(10):1135–1145.

Wetterstrand, K. S. 2013. “DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program (GSP).” https://www.genome.gov/sequencingcosts/.

Appendix B: Biographies

Bruce Budowle

Dr. Bruce Budowle received a Ph.D. in Genetics in 1979 from Virginia Polytechnic Institute and State University. From 1979-1982, Dr. Budowle was a postdoctoral fellow at the University of Alabama at Birmingham. Working under a National Cancer Institute fellowship, he carried out research predominately on genetic risk factors for such diseases as insulin dependent diabetes mellitus, melanoma, and acute lymphocytic leukemia.

In 1983, Dr. Budowle joined the research unit at the FBI Laboratory Division to carry out research, development, and validation of methods for forensic biological analyses. The positions he has held at the FBI include: Research Chemist, Program Manager for DNA Research, Chief of the Forensic Science Research Unit, and the Senior Scientist for the Laboratory Division of the FBI. Dr. Budowle has contributed to the fundamental sciences as they apply to forensics in analytical development, population genetics, statistical interpretation of evidence, and in quality assurance. Some of his technical efforts have been: 1) development of analytical assays for typing myriad protein genetic marker systems, 2) designing electrophoretic instrumentation, 3) developing molecular biology analytical systems to include RFLP typing of VNTR loci and PCR-based SNP assays, VNTR and STR assays, and direct sequencing methods for mitochondrial DNA, 4) new technologies; and 5) designing image analysis systems. Dr. Budowle has worked on laying some of the foundations for the current statistical analyses in forensic biology and defining the parameters of relevant population groups. He has published approximately 550 articles, made more than 700 presentations (many of which were as an invited speaker at national and international meetings), and testified in well over 250 criminal cases in the areas of molecular biology, population genetics, statistics, quality assurance, forensic genetics and forensic biology. In addition, he has authored or co-authored books on molecular biology techniques, electrophoresis, protein detection, microbial forensics, and forensic genetics. Dr. Budowle has been involved directly in developing quality assurance (QA) standards for the forensic DNA field. He has been a chair and member of the Scientific Working Group on DNA Methods, Chair of the DNA Commission of the International Society of Forensic Genetics, and a member of the DNA Advisory Board. He was one of the initial architects of the CODIS National DNA database, which maintains DNA profiles from convicted felons, from evidence in unsolved cases, and from missing persons.

Some of Dr. Budowle's efforts over the last 15 years also are in counter terrorism, including identification of victims from mass disasters and in efforts involving microbial forensics and bioterrorism. Dr. Budowle was an advisor to New York State in the effort to identify the victims from the WTC attack. In the area of microbial forensics, Dr. Budowle has been the chair of the Scientific Working Group on

Microbial Genetics and Forensics, whose mission was to set QA guidelines, develop criteria for biologic and user databases, set criteria for a National Repository, and develop forensic genomic applications. He also has served on the Steering Committee for the Colloquium on Microbial Forensics sponsored by American Society of Microbiology, an organizer of four Microbial Forensics Meetings held at The Banbury Center in the Cold Spring Harbor Laboratory, and on steering committees for NAS sponsored meetings.

In 2009 Dr. Budowle became Executive Director of the Institute of Applied Genetics and Professor in the Department of Molecular and Medical Genetics at the University of North Texas Health Science Center at Fort Worth, Texas. His current efforts focus on the areas of next generation sequencing, human forensic identification, microbial forensics, pharmacogenetics, and emerging infectious disease.

Jonathan King

Jonathan King received his MS in Molecular Biology from Tarleton State University in 2009 with a research focus in capturing novel polymorphic InDels for strain-level identification of agricultural pathogens. After graduation, Mr. King joined the University of North Texas as a forensic technologist. He has been the laboratory manager for the research and development lab since March 2011. During his relatively short tenure, Mr. King has co-authored over 40 publications on topics ranging from improved low-template strategies, next generation sequencing, population genetics, DNA repair, small amplicon markers, mitochondrial sequencing, novel DNA extraction techniques, and DNA transfer. His current areas of research include rapid DNA technologies, evaluation of DNA recovery tools, markers for ancestry determination, development of bioinformatics tools and workflows for interpretation of sequencing data, microbial forensics, next generation sequencing, and molecular medicine.

Patricia Melton

Dr. Patricia Melton is currently a Senior Research Forensic Scientist in the Center for Forensic Sciences at RTI International. In this capacity, she implements and procures educational courses to facilitate the knowledge transfer of current forensic DNA technology to law enforcement and judicial practitioners. She also serves as a project team member for the knowledge transfer and best practices development within the forensic community for responses to sexual assaults. Dr. Melton possesses the following specialized skills in forensic sciences: serological screening for biological fluids; nuclear DNA extraction of swabs; bloodstains, tissue, bone, hair roots, and teeth; and nuclear DNA extraction from “touch” DNA samples. In addition, she has specialized skills in short tandem repeat (STR) analysis. Dr. Melton has experience with providing courtroom testimonies and exceeds the education requirements for

a DNA Forensic Casework Analyst as established by the FBI Quality Assurance Standards. Dr. Melton has been on the faculty of two universities and actively participates in the certification program for crime laboratories under the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLAD/LAB) regulations and requirements (legacy and International Organization for Standardization [ISO] programs).

Shane Hamstra

Mr. Shane T. Hamstra has more than 10 years of professional experience in media production and training development. He is currently a Training Specialist in RTI's Center for Forensic Sciences (CFS). He serves as the online training production manager to oversee project development and delivery through on-demand, live online, and/or on-site workshop channels. Mr. Hamstra manages production schedules and directives to meet critical deadlines and ensure the completion of deliverables for the successful release of each training opportunity. Before joining, RTI, he managed productions for national cable networks such as the *National Geographic Channel* and the *Discovery Channel*, independent films, and educational training and corporate videos. In his previous and current work, Mr. Hamstra coordinates with subject matter experts, graphic artists, instructional designers, Web developers, voice talent, recording engineers, and editorial resources to bring projects to fruition.

RTI INTERNATIONAL

www.rti.org

Massively Parallel Sequencing: Understanding the Basic Technology

NIJ Live Online Workshop

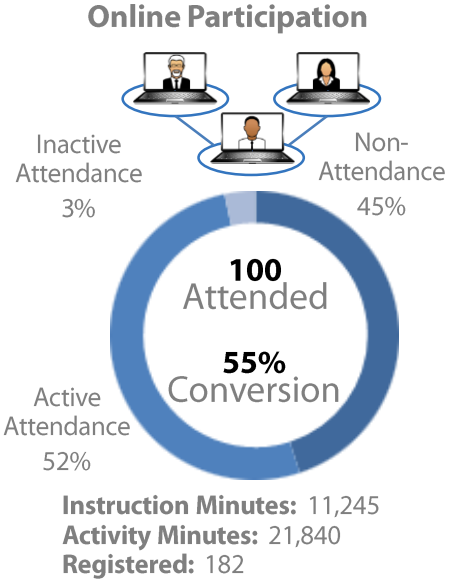
MAY 20 2015 | 1 PM ET

Duration: 120 minutes

Format: Live Online

Massively parallel sequencing (MPS) is an exciting technology that holds promise for enhancing the capabilities of the forensic DNA analyst. This first webinar, Understanding the Basic Technology, will be on the latest cutting edge information from the representatives, stakeholders and practitioners currently bringing this technology to fruition in the forensic science arena. The presentations and panel discussion will include an overview of platforms - the Ion PGM™ and MiSeq™ systems- and their respective chemistries, including sample and library preparation. The discussions will focus primarily on salient features of the platforms/systems so that forensic scientists can be better informed as they decide to move forward with this promising new technology.

Attendee Interactivity Rate



Instruction Minutes: 11,245
Activity Minutes: 21,840
Registered: 182

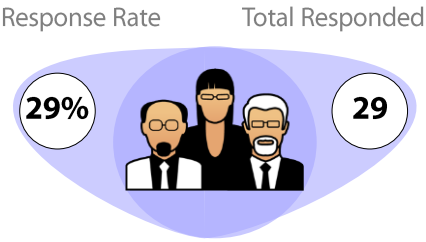
Each event is tracked by the following:
Non-attendance rate: Those who registered but did not attend divided by registration (an indication of conversion from registration to attendance).

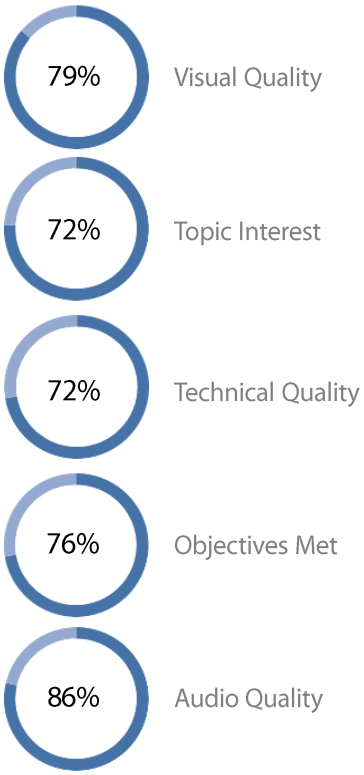
Active attendance rate: Rate at which registrants attend and interact consistently throughout the event. For a day-long event, we expect this rate to be lower because attendees will attend sessions of interest, but not necessarily the entire date. For perspective, we see inactive attendance rates for purely online, 90-minute events of ~5%.

Inactive attendance rate: Rate at which registrants attend but do not stay active for the entire event. We do not have the ability to estimate this interaction of on-site attendees.

Satisfaction

Our standard survey consists of 17 questions. The questions reflecting the overall performance are shown below.





Appendix C: Event Performance Sheet for Webinar 1 (May 2015)

Massively Parallel Sequencing: Understanding the Basic Technology

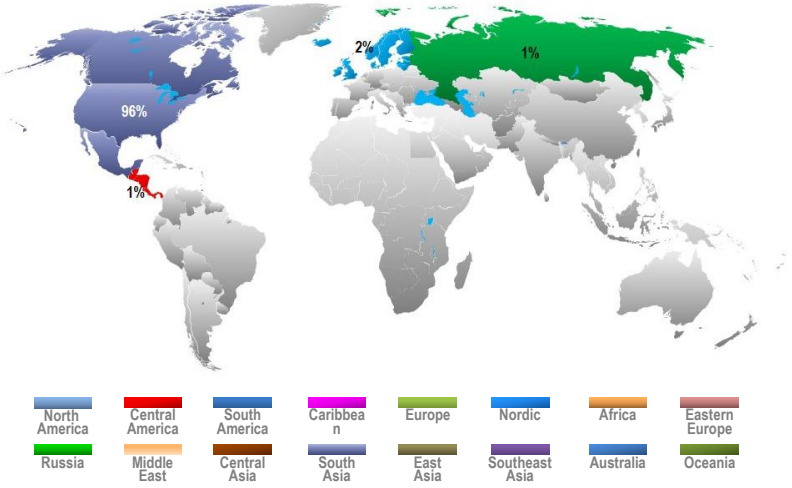
NIJ - Sponsored Event

Event Performance Statement

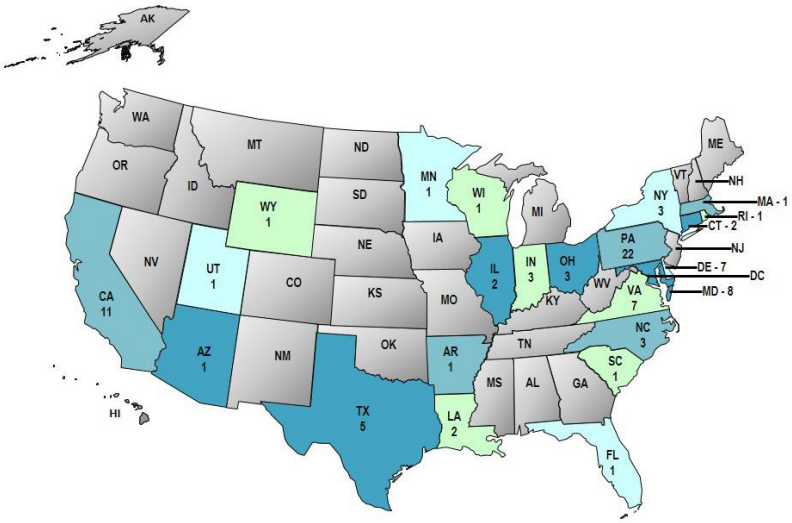
Attendance

Values are a percentage of the attendees throughout the day.

International Online Attendance

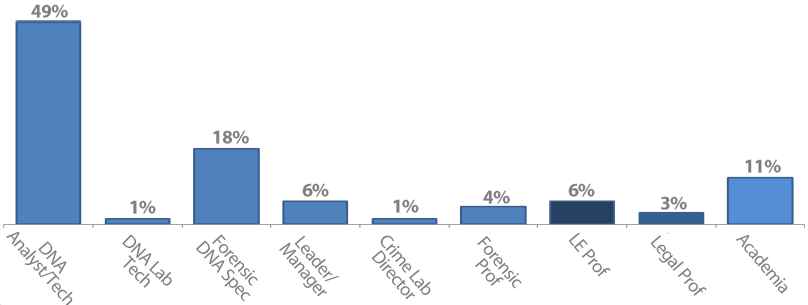


U.S. Online Attendance



Which Best Describes You?

At the beginning of each event attendees are polled "Which best describes you?"



"I learned a great deal about the MPS technology and have a better understanding the limitations, concerns, and potential of the information and capabilities of the new systems."

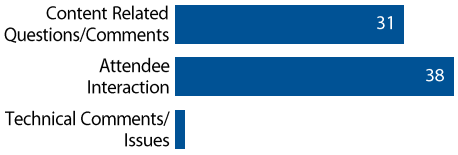
—Online Attendee Response

"Hearing more specific information on how the two systems work than had previously been made available to me; enjoyed the questions segment at the end."

—Online Attendee Response

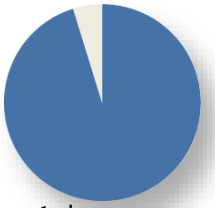
Chat Interactions

An open chat is used in each event. The host and ghost host encourage interaction from attendees to the subject matter expert. New conversation topics brought up by attendees will be extracted from the chat and further discussed.



Chatter

Based on 120 minutes of content delivered



1 chat every 1.71 minutes

Contact

Forensic Technology Center of Excellence
www.forensiccoe.org
866.252.8415
forensiccoe@rti.org

RTI International
3040 E. Cornwallis Road, PO Box 12194
Research Triangle Park, NC
27709-2194 USA

RTI INTERNATIONAL

www.rti.org

Massively Parallel Sequencing: The Genetic Marker Potential

NIJ Live Online Workshop

JUN 17
2015

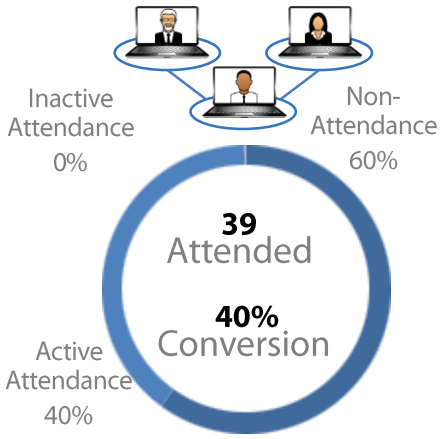
1 PM ET
Duration: 120 minutes

Format: Live Online

Massively parallel sequencing (MPS) is an exciting technology that holds promise for enhancing the capabilities of the forensic DNA analyst. In this second webinar, The Genetic Marker Potential, we will discuss the different genetic markers which can be typed, including the future roles and benefits of these markers, as well as considerations for CODIS compatibility and expansion. In addition, we will discuss the pros and cons of the different sized multiplexes. These discussion will clearly identify the potential benefits this promising technology can bring to forensic science.

Attendee Interactivity Rate

Online Participation



Instruction Minutes: 11,245
Activity Minutes: 21,840
Registered: 97

Each event is tracked by the following:
Non-attendance rate: Those who registered but did not attend divided by registration (an indication of conversion from registration to attendance).

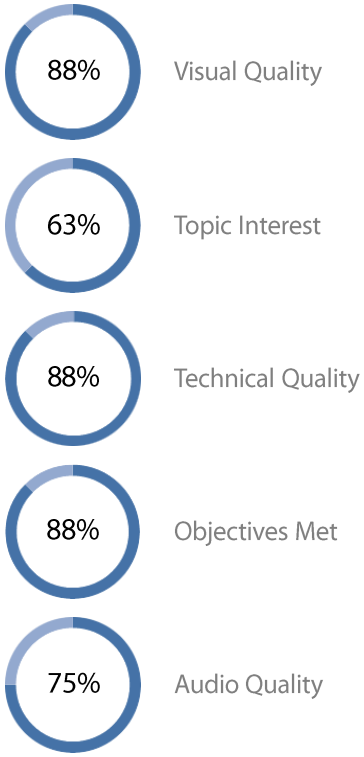
Active attendance rate: Rate at which registrants attend and interact consistently throughout the event. For a day-long event, we expect this rate to be lower because attendees will attend sessions of interest, but not necessarily the entire date. For perspective, we see inactive attendance rates for purely online, 90-minute events of ~5%.

Inactive attendance rate: Rate at which registrants attend but do not stay active for the entire event. We do not have the ability to estimate this interaction of on-site attendees.

Satisfaction

Our standard survey consists of 17 questions. The questions reflecting the overall performance are shown below.

Response Rate | Total Responded



Appendix D: Event Performance Sheet for Webinar 2 (June 2015)

Massively Parallel Sequencing: The Genetic Marker Potential

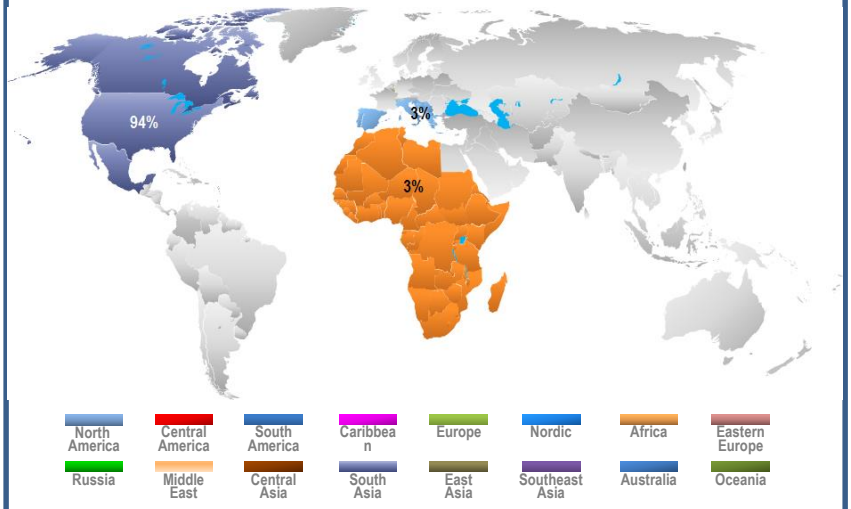
NIJ - Sponsored Event

Event Performance Statement

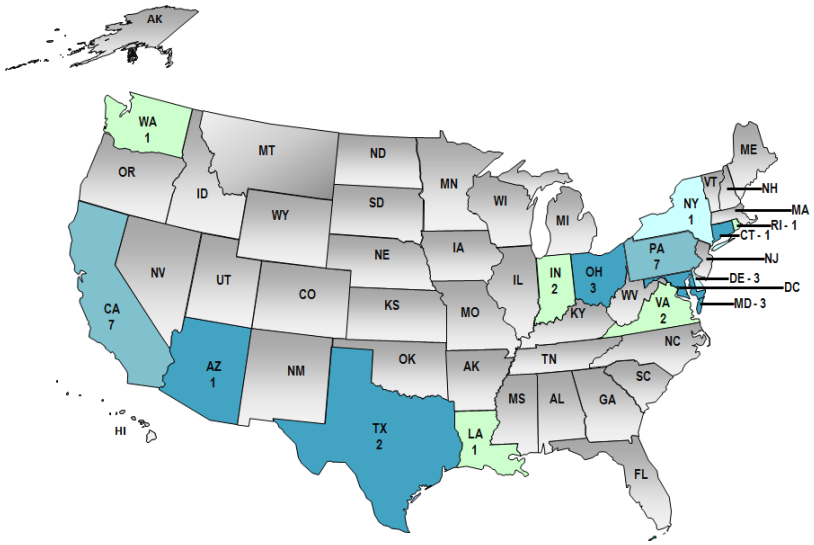
Attendance

Values are a percentage of the attendees throughout the day.

International Online Attendance

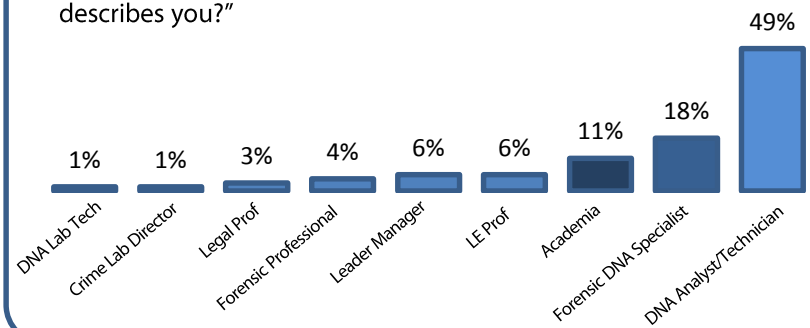


U.S. Online Attendance



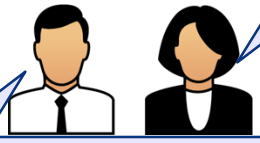
Which Best Describes You?

At the beginning of each event attendees are polled "Which best describes you?"



"Got a well-explained overview of the implementation and use of the technology - both pro and con."

—Online Attendee Response

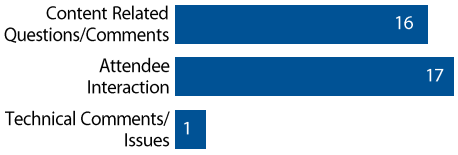


"Learned more about X-STRs that I knew before."

—Online Attendee Response

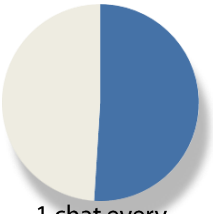
Chat Interactions

An open chat is used in each event. The host and ghost host encourage interaction from attendees to the subject matter expert. New conversation topics brought up by attendees will be extracted from the chat and further discussed.



Chatter

Based on 120 minutes of content delivered



1 chat every 3.53 minutes

Contact

Forensic Technology Center of Excellence
www.forensiccoe.org
866.252.8415
forensiccoe@rti.org

RTI International
3040 E. Cornwallis Road, PO Box 12194
Research Triangle Park, NC
27709-2194 USA

RTI INTERNATIONAL

www.rti.org

Massively Parallel Sequencing: Bioinformatics

NIJ Live Online Workshop

JULY 29
2015

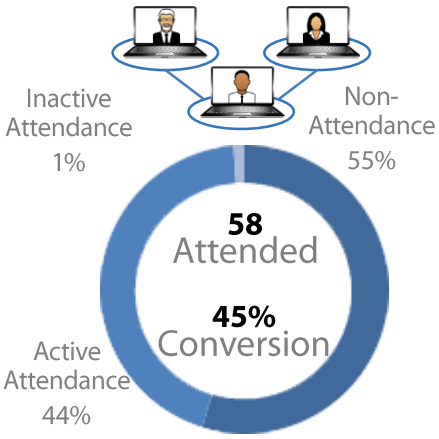
1 PM ET
Duration: 120 minutes

Format: Live Online

Massively parallel sequencing (MPS) is an exciting technology that holds promise for enhancing the capabilities of the forensic DNA analyst. In this third webinar, Bioinformatics, we will discuss the different software tools available for data processing from raw data to variant identification, as well as specialized tools specifically designed for STR, mtDNA and SNP analysis. Furthermore, we will discuss the potential increased ability for mixture deconvolution and interpretation using data from the MPS platforms. These discussion will clearly identify the potential benefits this promising technology can bring to forensic science.

Attendee Interactivity Rate

Online Participation



Instruction Minutes: 6,794
Activity Minutes: 15,360
Registered: 120

Each event is tracked by the following:
Non-attendance rate: Those who registered but did not attend divided by registration (an indication of conversion from registration to attendance).

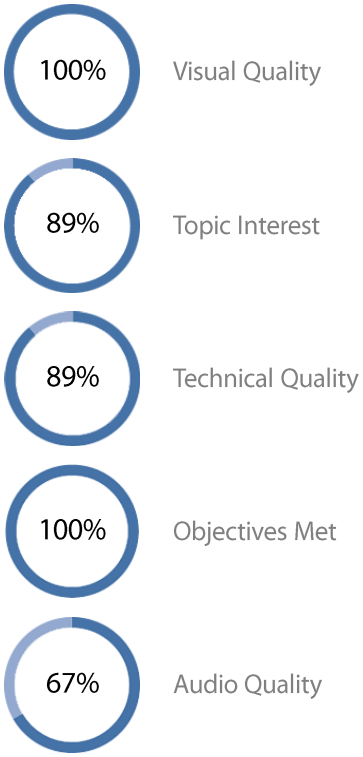
Active attendance rate: Rate at which registrants attend and interact consistently throughout the event. For a day-long event, we expect this rate to be lower because attendees will attend sessions of interest, but not necessarily the entire date. For perspective, we see inactive attendance rates for purely online, 90-minute events of ~5%.

Inactive attendance rate: Rate at which registrants attend but do not stay active for the entire event. We do not have the ability to estimate this interaction of on-site attendees.

Satisfaction

Our standard survey consists of 17 questions. The questions reflecting the overall performance are shown below.

Response Rate | Total Responded



Appendix E: Event Performance Sheet for Webinar 3 (July 2015)

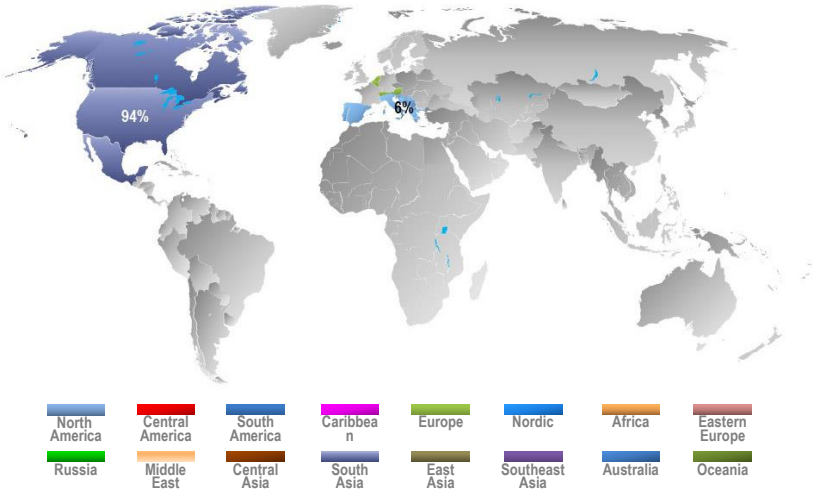
Massively Parallel Sequencing: Bioinformatics

NIJ - Sponsored Event

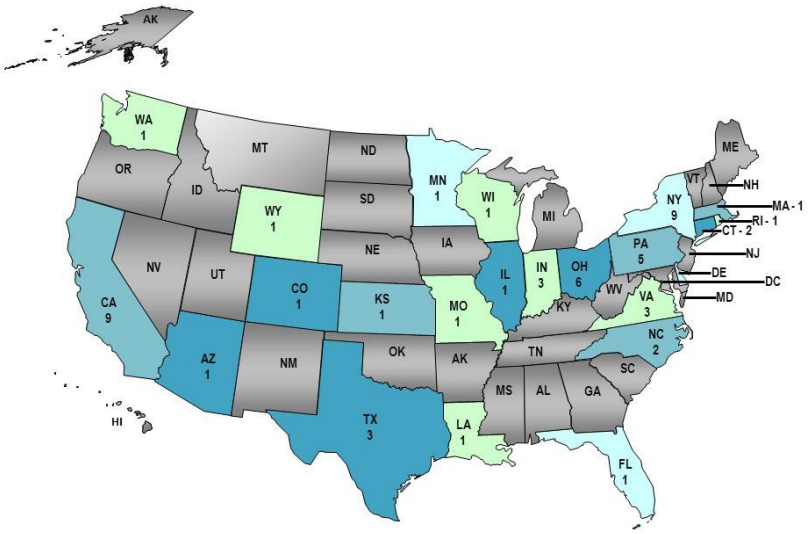
Attendance

Values are a percentage of the attendees throughout the day.

International Online Attendance

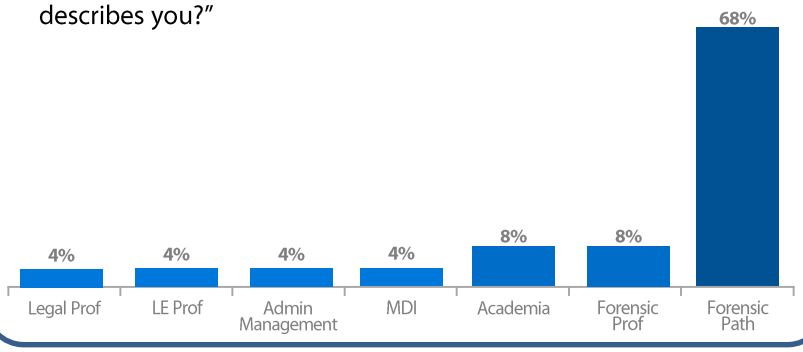


U.S. Online Attendance



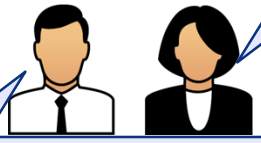
Which Best Describes You?

At the beginning of each event attendees are polled "Which best describes you?"



"The biggest benefit was learning about a seldom-discussed topic relating to MPS; bioinformatics information is hard to come by."

—Online Attendee Response

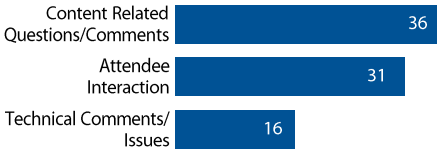


"The biggest benefit was getting perspectives from experts."

—Online Attendee Response

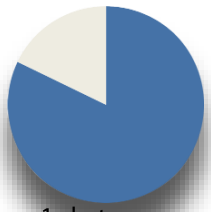
Chat Interactions

An open chat is used in each event. The host and ghost host encourage interaction from attendees to the subject matter expert. New conversation topics brought up by attendees will be extracted from the chat and further discussed.



Chatter

Based on 120 minutes of content delivered



Contact

Forensic Technology Center of Excellence
www.forensiccoe.org
866.252.8415
forensiccoe@rti.org

RTI International
3040 E. Cornwallis Road, PO Box 12194
Research Triangle Park, NC
27709-2194 USA

Event Performance Statement

RTI INTERNATIONAL

www.rti.org

Massively Parallel Sequencing: Validation and Applications

NIJ Live Online Workshop

AUGUST 29 2015 | 1 PM ET

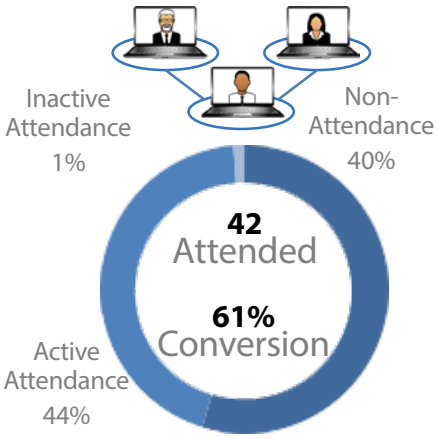
Duration: 120 minutes

Format: Live Online

Massively parallel sequencing (MPS) is an exciting technology that holds promise for enhancing the capabilities of the forensic DNA analyst. In this fourth webinar, Validation and Applications, we will discuss potential validation requirements to derive resolutions to pending issues of MPS implementation, including legality and CODIS considerations. We will look to successful examples of MPS application and lessons learned from previous technology platforms to determine next steps for forensic adoption.

Attendee Interactivity Rate

Online Participation



Instruction Minutes: 8,400
Activity Minutes: 5,640
Registered: 70

Each event is tracked by the following:
Non-attendance rate: Those who registered but did not attend divided by registration (an indication of conversion from registration to attendance).

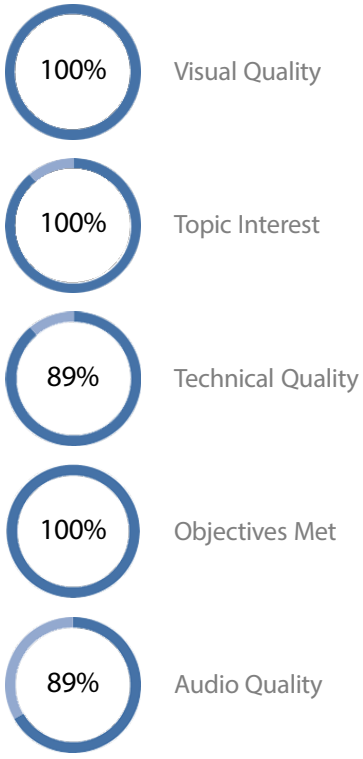
Active attendance rate: Rate at which registrants attend and interact consistently throughout the event. For a day-long event, we expect this rate to be lower because attendees will attend sessions of interest, but not necessarily the entire date. For perspective, we see inactive attendance rates for purely online, 90-minute events of ~5%.

Inactive attendance rate: Rate at which registrants attend but do not stay active for the entire event. We do not have the ability to estimate this interaction of on-site attendees.

Satisfaction

Our standard survey consists of 17 questions. The questions reflecting the overall performance are shown below.

Response Rate | Total Responded



Appendix F: Event Performance Sheet for Webinar 4 (August 2015)

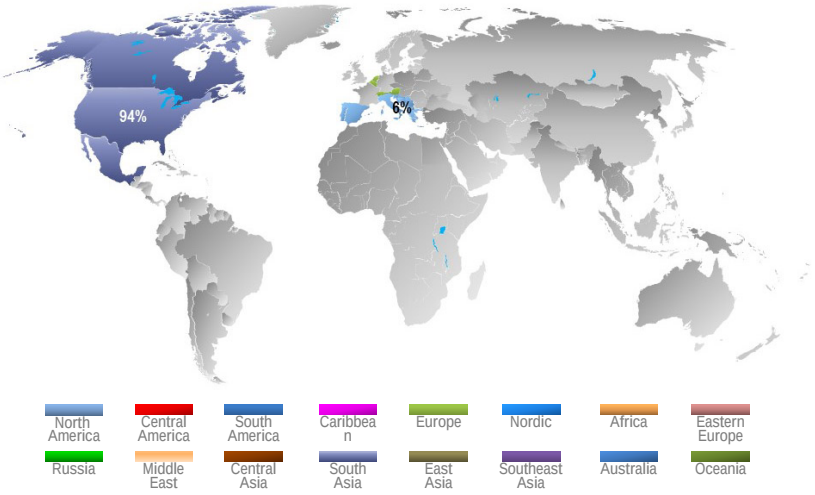
Massively Parallel Sequencing: Validation and Applications

NIJ - Sponsored Event

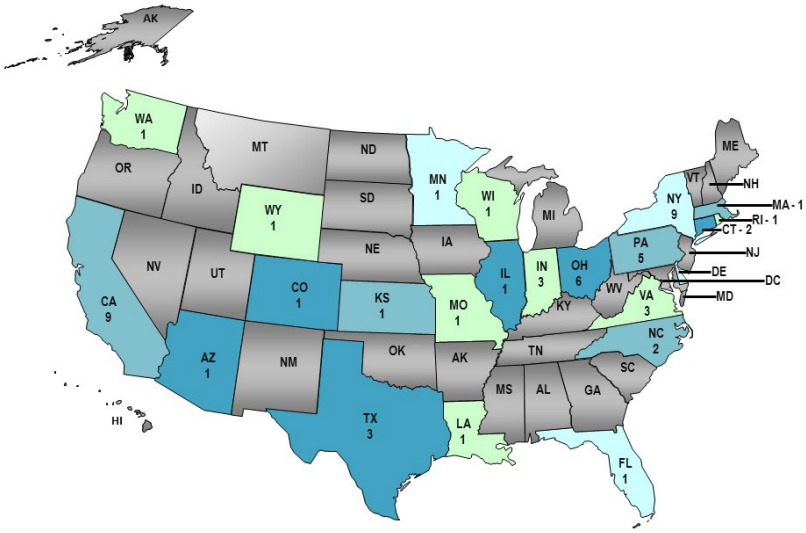
Attendance

Values are a percentage of the attendees throughout the day.

International Online Attendance

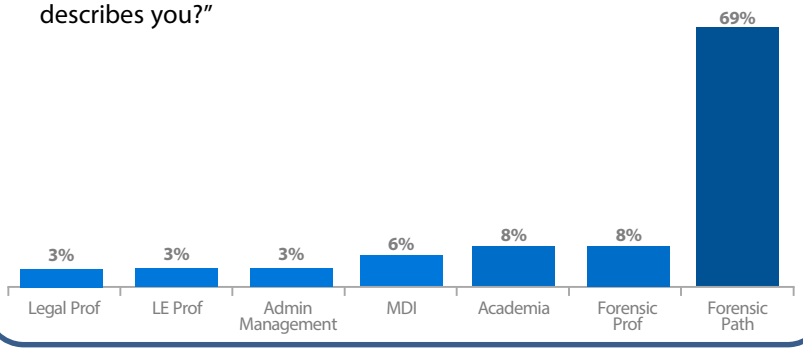


U.S. Online Attendance



Which Best Describes You?

At the beginning of each event attendees are polled "Which best describes you?"

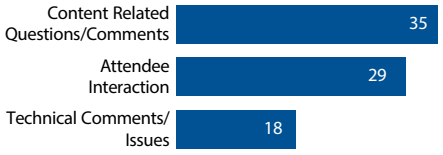


"The webinar provided sufficient information to enable us to begin our initial assessment of the technology and put a validation plan in place." —Online Attendee Response

"As a certified latent print examiner and BCI Director, I found all was of value." —Online Attendee Response

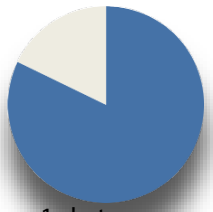
Chat Interactions

An open chat is used in each event. The host and ghost host encourage interaction from attendees to the subject matter expert. New conversation topics brought up by attendees will be extracted from the chat and further discussed.



Chatter

Based on 120 minutes of content delivered



Contact

Forensic Technology Center of Excellence
www.forensiccoe.org
866.252.8415
forensiccoe@rti.org

RTI International
3040 E. Cornwallis Road, PO Box 12194
Research Triangle Park, NC
27709-2194 USA

Event Performance Statement