SUMMARY OF THE WORK

The technique of individual identification in modern forensics is DNA typing by short tandem repeats (**STRs**) has brought a standardized, quantitative method with strong statistical underpinnings to the criminal justice system. While the fundamental principles behind STR typing have not changed, newly developed instrumentation and informative biological markers have the potential to address the limitations of current techniques as well as to improve throughput at lower costs. The forensic community is beginning to evaluate massively parallel sequencing (**MPS**) as a means to overcome these problems. Such methods not only add additional sequencing information but have a nearly unlimited capacity for additional STRs as well as SNP markers, thereby enhancing individual identification. These instruments also have the potential for significant improvements in throughput at lower costs.

The goal of this study was to evaluate Thermo Fisher's Precision ID GlobalFiler™ NGS STR Panel on their Ion S5 System for forensic casework. To achieve this goal five Specific Aims were addressed: i) to determine accuracy and sensitivity using single source samples, ii) assess reliability, barcoding, and contamination, iii) test challenging samples, including degraded samples from bones, blood cards and teeth as well as mixed samples in various ratios from two to six persons, iv) establish a high throughput liquid handling system for consistent library preparations, and v) calculate costs and time from preparation through analysis.

GlobalFiler™ NGS STR Panel v2 was tested thoroughly and the following outcomes were obtained: Concordance could be verified for all DNA samples tested to the overlapping loci of PowerPlex Fusion (CE technology) and ForenSeq DNA Signature Prep kit (MPS technology). Some samples with 50 pg input revealed full profiles, repeatedly. Reducing the sample number per chip from 32 to 24 increased read numbers and on samples with limited DNA input decreased the number of drop-outs. Sample specific barcodes prevented cross-contamination between samples, while pooled for the sequencing reaction. Recovery PCRs performed on

libraries had a positive effect on DNA samples with limited DNA input. The GlobalFiler™ NGS STR Panel v2 was not much affected by degraded DNA. All expected alleles were obtained with significant degraded DNA samples, degradation indices were up to 100. Data from two-person mixtures showed drop-outs of the minor contributor. The higher the ratio the more drop-outs were detected. Mixed samples with a ratio of 1:20 could barely be identified as mixed. The total time for two experiments (each consisting of 32 samples, including controls) using the GlobalFiler™ NGS STR Panel v2 is about 60h and the costs for 1 plate (32 samples) were calculated to $5,029.

1. PURPOSE OF THIS PROJECT

Individual identification by DNA analysis has become an essential forensic tool in the criminal justice system. By improving the quality of evidence, based on sound scientific principles, DNA testing has improved the quality of justice. The original DNA method, however, is decades old and advances in chemistry, population genetics and instrumentation now offer new platforms for more efficient and informative testing. The forensic community, including our laboratory, has been investigating these technologies for potential transition to forensic casework.

Massive parallel sequencing (**MPS**) is a method that can sequence hundreds of DNA amplicons (e.g. STRs) from multiple people simultaneously. The key to this technique is DNA barcoding of each person's sample prior to mixing and sequencing. Currently for forensic testing, 32 samples can be uniquely barcoded and multiplexed into a single run giving MPS the potential for increased throughput and efficiency. An advantage of this method is that STRs are not merely sized, but sequenced. Thus, it is possible to distinguish alleles of identical length if a SNP is present in one. In addition, as amplicon length is no longer a limiting constraint, multiple informative loci of similar or identical length can be used. Thus, legacy STR amplicons can be redesigned to the shortest possible length making them more useful for degraded samples.

Prior to implementing any new technology, a rigorous comparison with existing methods is required. This means determining accuracy, sensitivity, repeatability, robustness, time and costs.

The overall goal of this project was to evaluate Thermo Fisher MPS instruments, Ion Chef and Ion S5, and their STR chemistry, the precision ID GlobalFiler, in order to find the appropriate next generation DNA technology for routine casework. To achieve this goal five Specific Aims were addressed: i) to determine accuracy and sensitivity of single source samples, ii) assess reliability, barcoding, and contamination, iii) test challenging samples, including degraded samples from bones, blood cards and teeth as well as mixed samples in various ratios from two persons, iv) establish a high throughput liquid handling system for consistent library preparations, and v) calculate costs and time from preparation through analysis.

2. PROJECT DESIGN AND METHODS

Following exchange of the Ion Chef and update of the Converge software to the version 2.2, a total of 24 experimental runs were performed, to address the following Specific Aims:

Specific Aim i: Determination of accuracy and sensitivity using single source samples. This aim was addressed with repeated benchmark and various sensitivity runs, where the range of DNA input varied as well as the number of PCR cycles.

Specific Aim ii: Assessment of reliability of GlobalFiler chemistry to define conditions which will result in consistent outcomes, barcoding, and contamination. This included re-running library pools, testing the effect of recovery PCRs, mixing libraries before purification, and the importance of library dilution to 50 pM for loading onto the Ion S5 for sequencing by synthesis.

Specific Aim iii: Test of challenging samples, including purposefully degraded samples by boiling as well as samples from bones, blood cards and teeth, with varying degradation indices. Mixed samples of two contributors were tested in various ratios.

Specific Aim iv: Establishment of a high throughput liquid handling system for consistent library preparations using the Biomek FX liquid handling robot, which was dedicated for this project and was used for other MPS library preparations (2016-DN-BX-0172).

Specific Aim v: Calculation of costs and time from preparation to analysis.

3. DATA ANALYSIS

The software Converge version 2.2 was used for data analysis. Results of samples were compared to their known genotypes in order to indicate alleles, artifacts, and to detect drop-outs. Data analysis is still in progress.

4. PROJECT FINDINGS

Specific Aim i: Concordance could be verified for all DNA samples tested to the overlapping loci of PowerPlex Fusion (CE technology) and ForenSeq DNA Signature Prep kit (MPS technology). No dropouts were detected for the two independent benchmark experiments, with the recommended DNA input of 1ng for each sample. The outcomes of these two benchmark experiments were similar. The number of artifacts detected for the experiments were 10 and 11, respectively. They were mostly found (8x) at D12S391 and were allele drop-ins (**ADI**) of a 1 nt insertion ranging from 9 to 37% to the true allele. Fewer ADIs were detected at D3S4529, a 1 nt deletion up to 33%, and at D12ATA63, a 1 nt insertion around 32%.

The experimental runs testing sensitivity revealed that several samples with an input of 50 pg DNA resulted in full outcomes, repeatedly. However, dropouts were observed, mostly at PentaD and PentaE. Artifacts included, besides the ADIs at D12S391, various stutter products as well as ADIs at other loci. The sensitivity run testing samples with higher DNA inputs, ranging from 500 to 1000 pg DNA, showed no dropouts but some ADIs, mostly at D12S391. For lower DNA inputs more PCR cycles were recommended by the manufacturer, up to 26. The lower the DNA input the more drop-outs and artifacts were observed, however, with low frequency. For 12 samples that had less than 100 pg DNA input were in one run 15 artifacts and 12 drop-outs observed and in an independent repeat 9 artifacts and 1 drop-out. Drop-outs occurred mostly at Penta D and Penta E. Interestingly, the number of artifacts and drop-outs for low input samples (100 and 50 pg DNA) was comparable when performed with only 23 PCR cycles.

Reducing the sample number per chip from 32 to 24 improved outcomes. The read numbers were increased and on samples with limited DNA input the number of drop-outs was decreased.

Specific Aim ii: To address technical variation introduced by the Ion Chef through chip loading and the S5 by the sequencing reaction, a library pool was loaded three times. All three runs showed all alleles at all loci. However, they varied in their ADIs: 10 were found in Expt. 1, while the reloaded pools showed 9 and 8, respectively, most of them were 1 nt insertion at D12S391. It seems, that these artifacts were introduced during the sequencing-by-synthesis reaction on the S5.

To answer the question, if the barcodes can be used for separating samples or may lead to cross-contamination, two sample libraries, one from a male and one from a female, were mixed directly after PCR, before purification, and two additional samples were mixed after the purification step. Sequencing revealed, no cross-contamination had occurred between the mixed samples.

Before the libraries were loaded onto the chip, they should be diluted to 50 pM, and if the concentration is lower, they should be used as is. The effect of a certain range was tested. Concentrations of 50 pM and 30 pM of the same sample sets were loaded on the same chip with and showed no difference. No drop-outs occurred. Therefore, it can be concluded that the sequencing reaction on the S5 is robust.

After quantification of the library, recovery PCR was tested with up to 10 cycles for samples with less than 50 pM, on two sensitivity and one degradation runs. All experimental runs were improved, but the effect was more pronounced for the sensitivity runs compared to the degradation run.

All positive controls revealed full profiles. Only one showed a N-1 stutter at D2S1338. Out of the 24 experimental runs, five negative controls had some reads (10 to 128) at certain loci. Four runs showed only one ADI, while one showed three ADIs at different loci.

Specific Aim iii: Two independent experimental runs revealed that the GlobalFiler™ NGS STR Panel v2 is not much affected by degraded DNA. Degradation indices (**DI**s) were obtained from Quantifiler Trio kit with <1 indicating no degradation, 1-10 slight to moderate degradation, and >10 significant degradation. With a DNA input of 1 ng all expected alleles were obtained with DIs ranging from 1 to 100. This could be confirmed with DNA from challenging tissues, such as bones, teeth, and blood. However, if the input was less than the recommended 1 ng, more alleles dropped out.

The data from two-person mixtures showed drop-outs of the minor contributor. The higher the ratio the more drop-outs were detected. In 1:20 mixtures, the minor was hardly to detect. Recovery PCR did not improve the outcome. Data are still analyzed in more detail.

Specific Aim iv: Establishment of a high throughput liquid handling system for consistent library preparations: Unfortunately, the Biomek FX liquid handling robot has broken down several times

during the last years. This year, 2021, the Biomek FX stopped working due to errors of the Z-axis motor, the Peltier element, and the Pod. The technician from Beckman Coulter decided that this instrument is unusable. It is 13 years old, some parts are no longer available, and it will be decommissioned. Consequently, this aim cannot be completed.

Specific Aim v: Usually one plate contained 32 samples in this evaluation of the GlobalFiler™ NGS STR Panel v2, including one positive and one negative control. The number of samples is the same for another MPS STR kit, the ForenSeq™ Signature Prep kit. Time and costs will be compared for those kits:

Library preparation for the GlobalFiler™ NGS STR Panel v2 took 10h for one plate.

The time on the Ion Chef for clonal amplification and chip loading was about 13h and 30min and is the same for one or two plates.

Sequencing on the S5 took 2h per chip (32 samples).

Data analysis for two chips was about 20h.

The total time for two experiments using the GlobalFiler™ NGS STR Panel v2 is about 60h and the costs for 1 plate were calculated to $5,029.

Library preparation for one plate (32 samples) using the ForenSeq™ Signature Prep kit took about 10h and 40min. Sequencing and data analysis on the MiSeq was about 28h and 15min.

The total time for one experiment using the ForenSeq™ Signature Prep kit is about 39h with the calculated costs of $2,799.

5. RESULTS OF THE FUNDED PROJECT

A peer reviewed publication is planned after completion of data analysis.

6. IMPLICATIONS FOR CRIMINAL JUSTICE POLICY AND PRACTICE IN THE UNITED STATES

Forensic testing laboratories are on the cusp of a transition between the 30 plus year old CE DNA testing paradigm to MPS technology which is only now being fielded for routine casework. The information gained from this study is vital to other public laboratories for their evaluation of MPS platforms as they work on this transition. Our dissemination of this work in a peer reviewed journal will be important in that effort. Additionally, the information we garner about ideal target ranges, limits of detection, precision, accuracy, and the robustness of genetic markers will help MPS manufacturers hone and improve their products for the forensic community.