Application Note

Automating ForenSeq kits using Aurora Biomed's VERSA 1100 liquid handling workstation

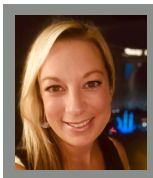
Introduction

Next-generation sequencing (NGS), also referred to as massively parallel sequencing (MPS), is a paradigm-shifting technology that provides higher accuracy, greater throughput, and more applications than historic methods like capillary electrophoresis (CE). To date, high-throughput sequencing technologies for DNA and RNA have been successfully applied to both genetic research and clinical practice. In the field of forensics, NGS delivers progressive advantages such as cost-effectiveness and a growing suite of forensic markers that can be analyzed simultaneously. This enables forensic identity applications such as typing autosomal STRs accepted by national databases, kinship applications such as mitochondrial typing, Y-STRs for familial searching and forensic investigative genetic genealogy, and investigative applications such as body fluid identification, age estimation, and DNA phenotyping.

In forensic practice, library preparation for sequencing has been identified as a significant bottleneck to broader adoption. NGS library preparation protocols are usually a multi-step process and require costly reagents and substantial hands-on time. To ensure a high degree of robustness and reproducibility of results, standardized sample preparation approaches and quality control measures are critical. Automation solutions can perform complex protocols with high reproducibility, leading to both reduced error rates and lower contamination risks associated with human interaction with the reagents and samples. Overall cost per sample can also be lowered by decreasing the hands-on time and labor associated with the workflow.

Implementing an automation platform involves upfront investment in time and resources to develop scripts, test, and validate the instrument within the laboratory workflow. This application note describes three separate internal validation studies leveraging Aurora Biomed's

Figure 1: Automation of the ForenSeq workflow with the ForenSeq® Kintelligence Kit, ForenSeq mtDNA Whole Genome Kit, and ForenSeq® DNA Signature Prep Kit on the Aurora VERSA 1100.



Rachel Oefelein
Chief Scientific Officer
DNA Labs International

> *"Implementation of the three ForenSeq systems for manual and automated processing has been instrumental in identification of unidentified human remains. The future and the possibilities are endless with NGS for cold cases and current casework!*

VERSA 1100 automation platform and the industry-leading forensic NGS library preparation chemistry – the ForenSeq® workflow by Verogen Inc. Scripts were developed for the ForenSeq Kintelligence Kit, the ForenSeq mtDNA Whole Genome Kit, and the ForenSeq DNA Signature Prep Kit (DNA Primer Mix B or DPMB) by Aurora Biomed Inc. (Figure 1). The studies, conducted by Rachel Oefelein, Chief Scientific Officer of DNA Laboratories International, compared libraries generated using manual and automated workflow and sequenced on the MiSeq FGx® Sequencing System. Comparable reproducibility and sensitivity results across manual and automated workflows demonstrate the ability of Aurora's VERSA 1100 liquid handling system to efficiently automate NGS workflows in alignment with forensic standards.

Materials and methods

Sample selection

Across the three internal validations a total of 461 samples were sequenced across 28 sequencing runs including; NIST SRM 2391c, 2800M, HL60, CHR, 9947A, known male and female buccal and blood samples, hair root and hair shaft extracts, semen extracts (sperm and cell fractions), vaginal swabs (sperm and cell fractions), 'touch' samples, bone extracts, and associated reagent blanks.

Library preparation, sequencing, and analysis

Samples were prepared using the ForenSeq Kintelligence Kit, the ForenSeq mtDNA Whole Genome Kit, and the ForenSeq DNA Signature Prep Kit (DNA Primer Mix B). The number of libraries prepared per workflow were in accordance with the manufacturer's protocol. Half of the samples for each workflow were manually prepared while the other half was prepared using the Aurora VERSA 1100 NGS workstation. All libraries were sequenced on the MiSeq FGx Sequencing System using the recommended sequencing reagents. Results were analyzed in the corresponding analysis module in the Universal Analysis Module using settings established during internal validation studies. PCR1 setup for all three ForenSeq systems currently requires manual setup by the user. The Aurora VERSA 1100 automated liquid handling platform was used for the setup of all the post-PCR1 steps: target enrichment, library purification, and library normalization. Target enrichment adds indexed adaptors and PCR2 reaction mix, while library purification and normalization require the addition of their respective purification or normalization beads, followed by ethanol washes and resuspension of the samples. Finally, the Aurora VERSA 1100 NGS platform was also used to pool the libraries prior to the run. The data in this paper represent a sampling of the three internal validations performed and is representative of the total results.

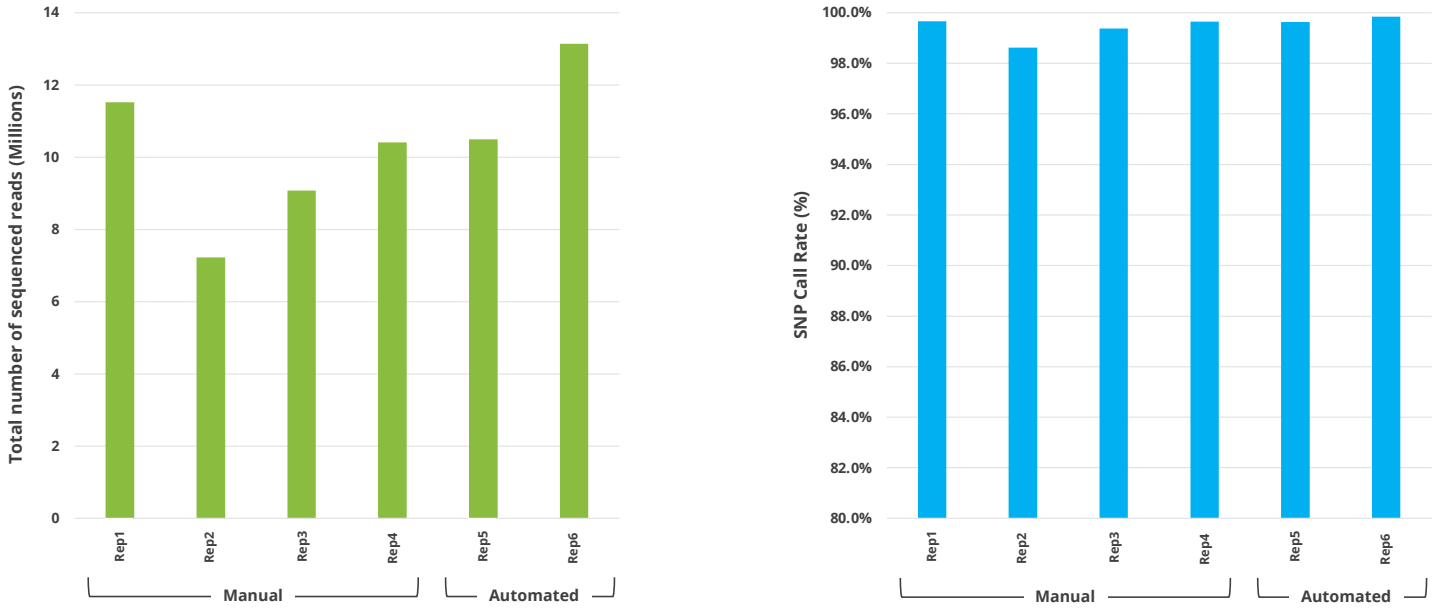


Figure 2: Repeatability and reproducibility studies using the ForenSeq Kintelligence Kit.

Reproducibility and repeatability studies

Two replicates of libraries were processed manually while two replicates of libraries were processed using the Aurora VERSA 1100 automated liquid handling workstation. All samples had an input DNA amount of 1 ng from a male donor. The plates containing the first set of replicates across manual and automated workflows included a no-template control (NTC), while the third replicate from the second plate did not.

Sensitivity and stochastic studies

A range of DNA inputs including 4 ng, 1 ng, 500 pg, 250 pg, 125 pg, 62 pg, 31 pg, 16 pg, and 8 pg were processed using the three ForenSeq kits. Two plates were manually prepared while two plates were prepared using the Aurora VERSA 1100.

Metrics

Total sample read counts and call rates were evaluated across all sequenced libraries. Total sample read counts is defined as the sum of sequencing reads detected across all marker types and all loci in a sample. Sample read count values provide an indication of signal intensity, with higher read counts corresponding to higher amounts of DNA input. As DNA input decreases so should the total reads. SNP call rates (%) were calculated as: (total number of expected alleles called/total number of expected alleles) x 100).

Results

High degree of reproducibility between manual and automated NGS workflows at all DNA inputs was observed.

Across all 3 ForenSeq workflows, samples processed manually and those processed using the Aurora VERSA 1100 generated calls that were highly reproducible, irrespective of the amount of input DNA.

ForenSeq Kintelligence

Repeatability and reproducibility between manual and automated runs was evaluated. The total number of sequenced reads detected across six replicates per sample distributed each across three runs are shown in Figure 2A. All samples processed on the Aurora VERSA 1100 NGS workstation generated similar number of reads to those that were processed manually. The inclusion of a NTC on one plate did not impact the overall number of reads. All samples and replicates generated more than the manufacturer's recommended guidelines for sequencing reads. Similarly, all six replicates typed sufficient calls to meet the criteria for upload to GEDmatch PRO™ as shown in Figure 2B.

Sensitivity studies were conducted to evaluate the performance of the system over a range of DNA inputs to inform profile result expectations and identify interpretation limitations. In addition to assessing ideal

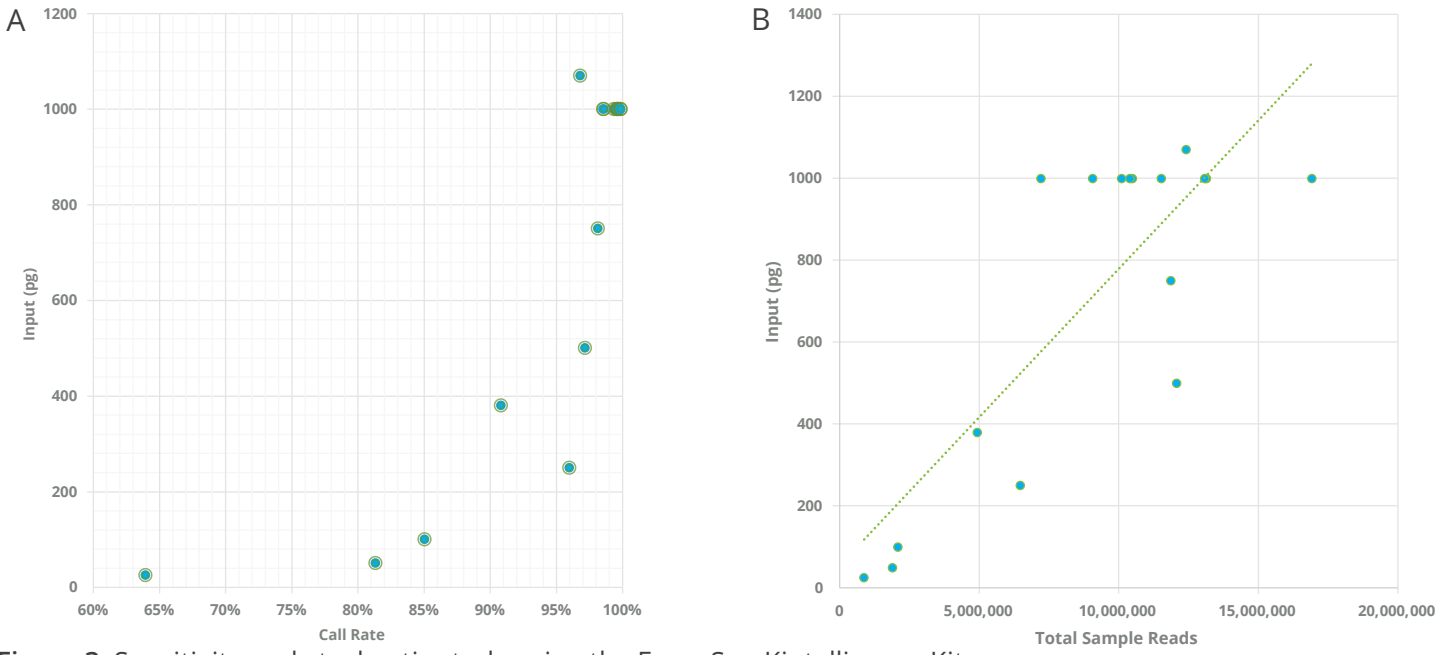


Figure 3: Sensitivity and stochastic study using the ForenSeq Kintelligence Kit.

target input range, this study also provided guidance on sample and marker type read count values, interlocus balance, SNP call rates, the effect of varying number of samples on the flow cells and intensity, and stochastic threshold. This data was part of DNA Lab International's internal validation and has not been shown in this application note. The total sample reads for each sample across manual and automated setups were plotted against DNA input ranging from 25 pg to 1000 pg as shown in Figure 3. Sample input was also plotted against call rates. At total inputs exceeding 200 pg, call rates greater than 90% were observed, while at inputs below 250 pg, call rates of approximately 80% were observed. These samples met the upload criteria for GEDmatch PRO.

ForenSeq DNA Signature Prep Kit - DNA Primer Mix B

Sensitivity runs were conducted to evaluate the performance of the system over a range of DNA quantities to inform profile result expectations and identify interpretation limitations. In addition to assessing ideal target input range, the following types of data were evaluated to inform expectations for interpreting profiles at various inputs: sample and marker type read count values, interlocus balance, allele call rates and drop-out, higher input samples (normalization assessment), intralocus balance, and stochastic threshold (data not shown). A dynamic range of DNA inputs were evaluated (8 pg to 4000 pg). Through this study, it has been demonstrated that the automated protocols tested on the Aurora VERSA 1100 for the ForenSeq DNA Signature Prep produces results comparable to those obtained when the samples were manually set up by an analyst. Both manual and automated workflows demonstrated lower call rates around 31 pg. This data also demonstrates the reproducibility of the system between manual and automated setups to obtain full profiles in ranges from 4 ng down to 62 pg, with profiles suitable for comparison still being obtained at inputs as low as 8 pg. (Figure 4).

ForenSeq Whole Genome mtDNA Kit

A control sample HL-60 was used to demonstrate a portion of reproducibility, repeatability, and sensitivity across two runs. The results inside and outside the

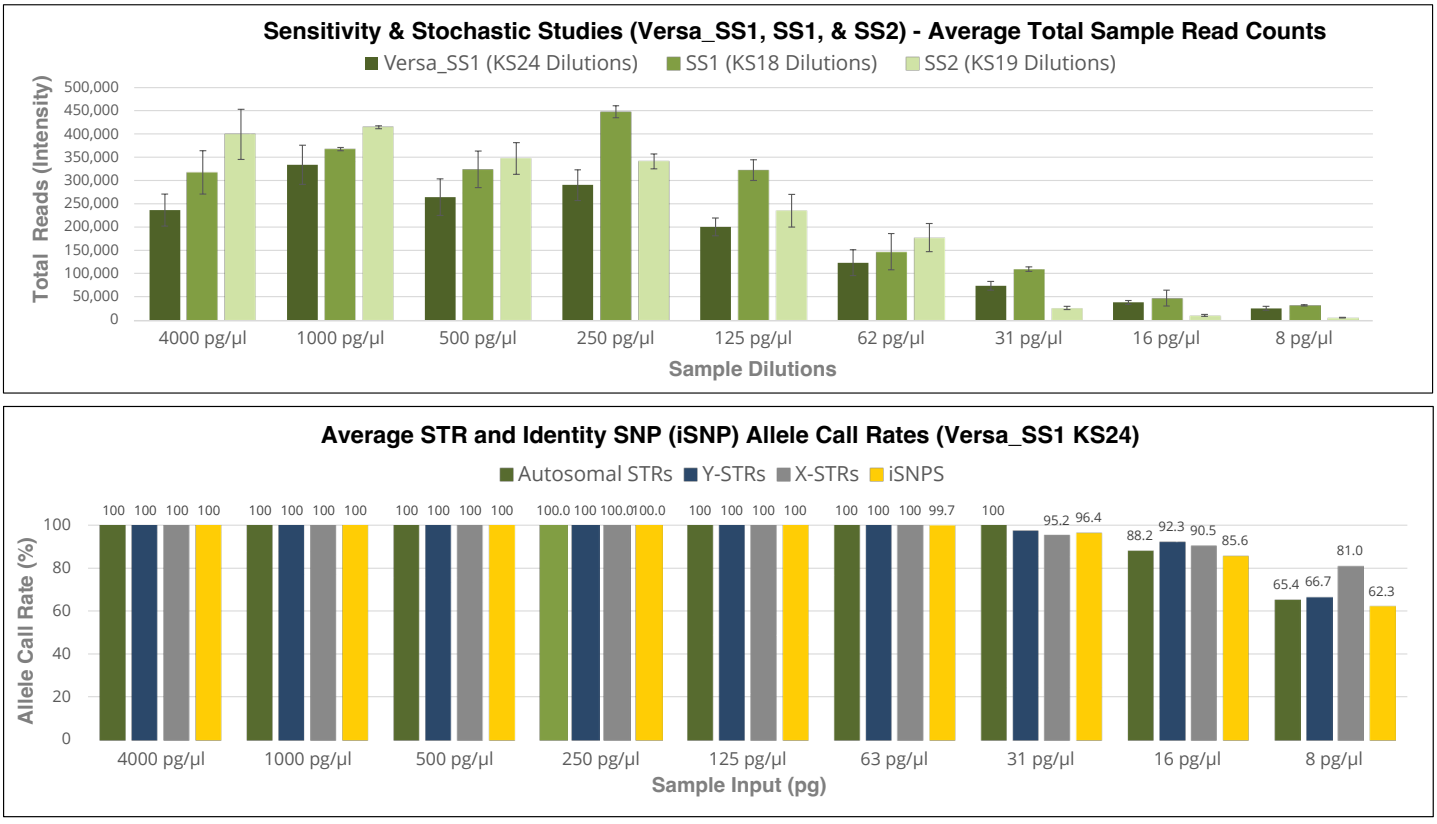


Figure 4: Total sample read counts and calls rates for sensitivity and stochastic studies using the ForenSeq DNA Signature Prep Kit.

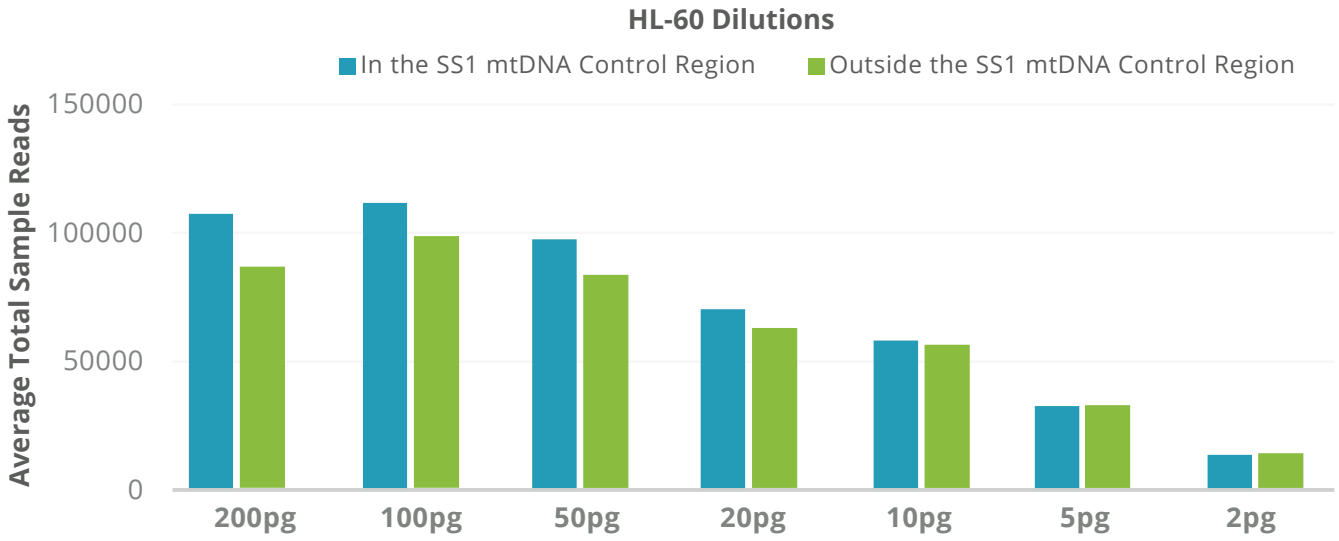


Figure 5: Total sample read counts inside and outside of SS1 mtDNA Control Region.

control region are shown in Figure 5 with comparable total number of reads being generated across the manual and automated workflows. The lowest input sample of 10 pg generated reads that were lower than the manufacturer recommended guideline. All replicates at the optimal input of 100 pg of DNA had no dropout regardless of manual or automation processes. As anticipated, decreasing DNA input resulted in a corresponding decrease in overall sample and individual read count values, regardless of the use of automated or manual processing. Despite decreased reads, when optimal DNA input is not available, significant amounts of mtDNA data were achieved at low-level inputs, with full profiles detected with as little as 20 pg of DNA. Allele dropout was observed in the 2 – 10 pg input range, though actionable data may still be achieved even at these lowest evaluated inputs. It is noted that ForenSeq is still a PCR-based assay and is subject to similar types of stochastic effects observed in CE-based PCR kits, which are exacerbated with decreasing inputs into PCR.

Conclusions

Data from these internal validation studies were evaluated to determine performance and utility of the Aurora VERSA 1100 automated liquid handling workstation within a forensic laboratory. Utilizing ForenSeq kits from Verogen, limitations of the end-to-end system were determined in order to support the development of interpretation guidelines. The MiSeq FGx Sequencing System at DNA Labs International is used in conjunction with the ForenSeq Kintelligence Kit, ForenSeq mtDNA Whole Genome Kit and the ForenSeq DNA Signature Prep Kit (DPMB) and analyzed using the ForenSeq Universal Analysis Software. The sensitivity, reliability, repeatability, concordance, ability to operate with minimal risk to contamination, reproducibility, and accuracy were demonstrated using an automated workflow. These internal validation studies support the use of Aurora's VERSA 1100 NGS system when used with Verogen products for sequencing forensic samples.

Aurora Biomed
www.aurorabiomed.com
(800) 883-2918

DNA Labs International
dnalabsinternational.com
(954) 426-5163

Product documentation is available for download at **www.verogen.com/support**

Acknowledgements

Special thanks to Cristina Servidio, Jayson Burkhardt, Lisbeth Colon, and Melissa Kotkin for their contributions to these internal validations.