Maintaining Quality Next Generation Sequencing (NGS) with Manual and Aurora's VERSA 1100 Automated Processing of ForenSeq kits

Introduction

Next generation sequencing (NGS), also reffered 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 at the DNA and RNA level 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 or 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 multistep 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 reduced error rates, while reducing risks associated with contamination 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.

Figure 1: Automation of the ForenSeq™ workflow on the Aurora's VERSA 1100

Aurora
Illuminating Solutions
DNA LABS INTERNATIONAL
ISO/IEC 17025:2017 and FBI QAS DNA Laboratory
VEROGEN

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's VERSA 1100 automation platform and the industry-leading forensic NGS library preparation chemistry – the ForenSeq workflow by Verogen Inc. Scripts were developed for the ForenSeqTM Kintelligence Kit, the ForenSeqTM mtDNA Whole Genome Kit, and the ForenSeqTM Signature Prep Kit (DNA Primer Mix B or DPMB). (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. (Figure 2). Comparable reproducibility and sensitivity results across manual and automated workflows demonstrate the ability of Aurora's VERSA 1100 to efficiently automate NGS workflows in alignment with forensic standards.

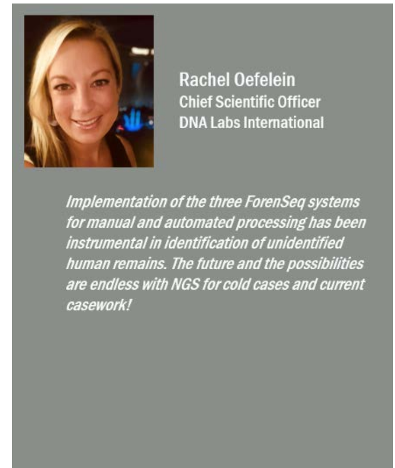


Figure 2: Rachel Oefelein, CSO, DNA Labs International

Materials & 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 samples, known male and female 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 prepped using the ForenSeq^{TM} Kintelligence Kit, the ForenSeq^{TM} mtDNA Whole Genome Kit and the ForenSeq^{TM} DNA Signature Prep Kit (DNA Primer Mix B). The number of libraries prepared per workflow were in accordance with the manufacturer's protocol. Half the samples for each workflow were manually prepared while the other half was prepared using the Aurora's VERSA 1100. 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's VERSA 1100 platform was used for setup involved with all the steps post PCR1; enriching the targets by adding the adaptors and PCR2 reaction mix to the samples, and all steps involved in the library purification and normalization. These steps require the addition of purification and normalization beads, subsequent ethanol washes, and resuspension of the samples. Finally, the Aurora's VERSA 1100 platform was also used to pool the libraries prior to the run. A sampling of these results generated as part of three internal validations is shown.

Reproducibility and Repeatability Studies

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

Sensitivity and Stochastic Studies

A range of input DNA spanning 4ng, 1ng, 500pg, 250pg, 125pg, 62pg, 31pg, 16pg and 8pg's were processed using the three ForenSeq kits. Two plates were manually prepped, while two plates were prepped using Aurora Versa 1100.

Metrics

Total sample read counts and call rates were evaluated were 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

Across all 3 ForenSeq workflows, samples processed manually and those processed using the Aurora Versa 1100 generated calls that were highly reproducibility, 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 3A. All samples processed on the Aurora's VERSA 1100 were 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 manufacturers 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 3B.

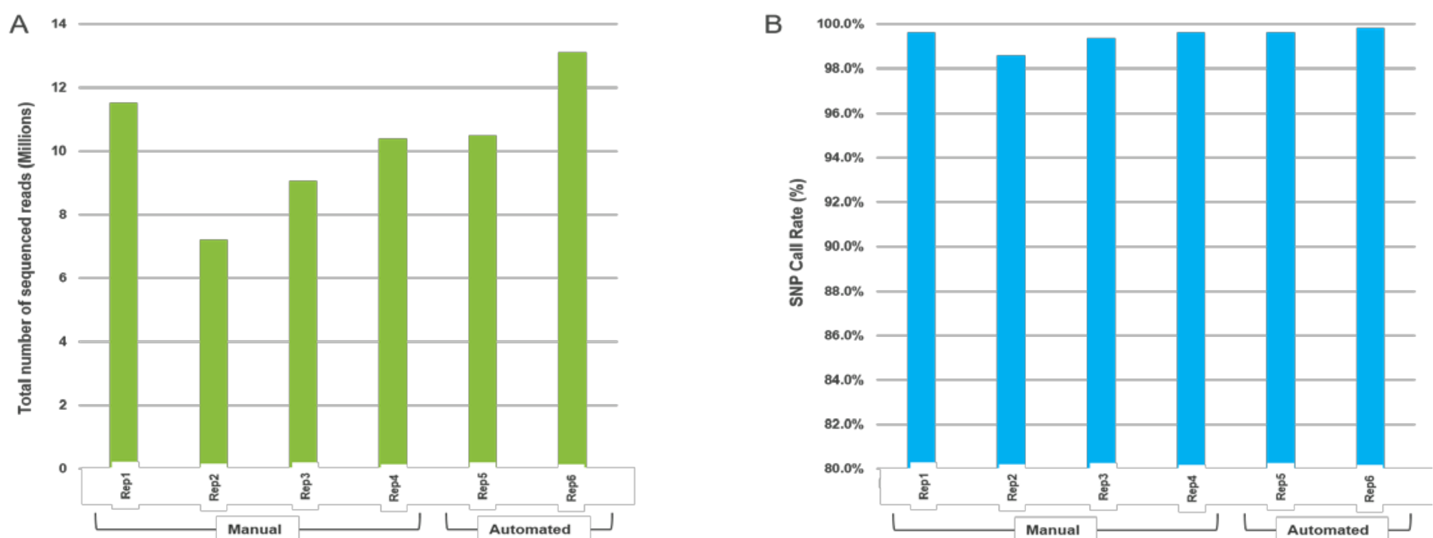


Figure 3: Repeatability and Reproducibility studies using the ForenSeq Kintelligence™ kit

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 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 DLIs internal validation and has not been shown in this application note. The total sample reads for each sample across manual and automated set-ups were plotted against DNA input ranging from 25 picograms to 1000 picograms as shown in Figure 4. Sample input was also plotted against call rates at total inputs exceeding 200 pg, call rates greater than 90.00% were observed while at inputs below 250 pg, call rates of approximately 80% were observed. These samples met the upload criteria for GEDmatch PRO™.

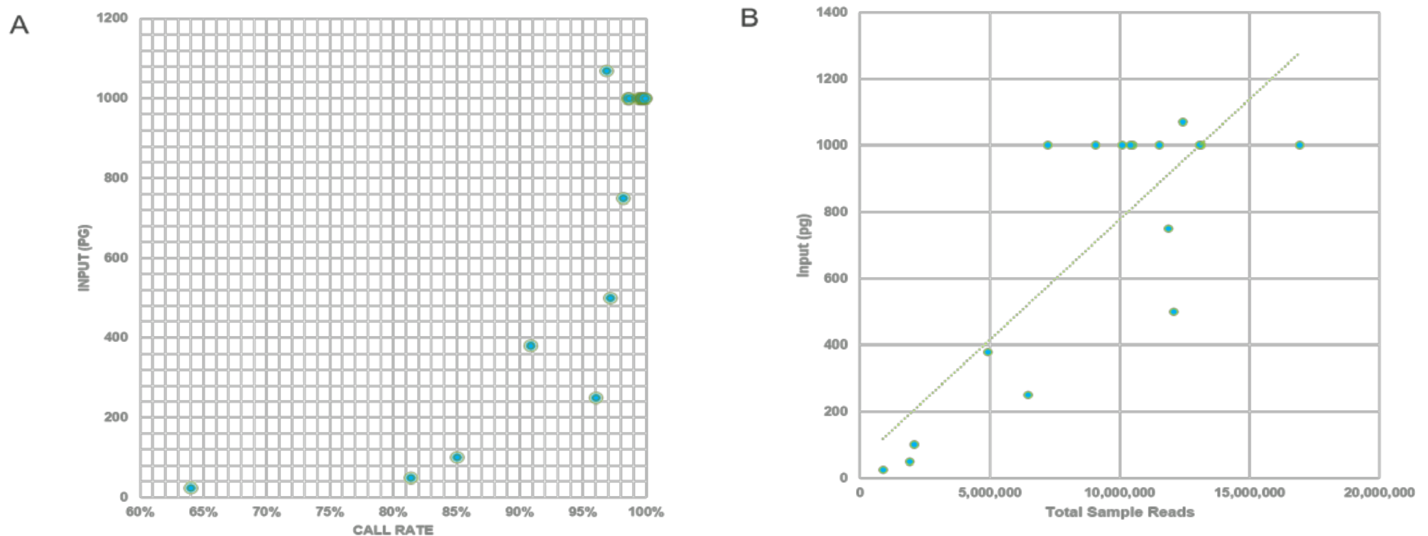
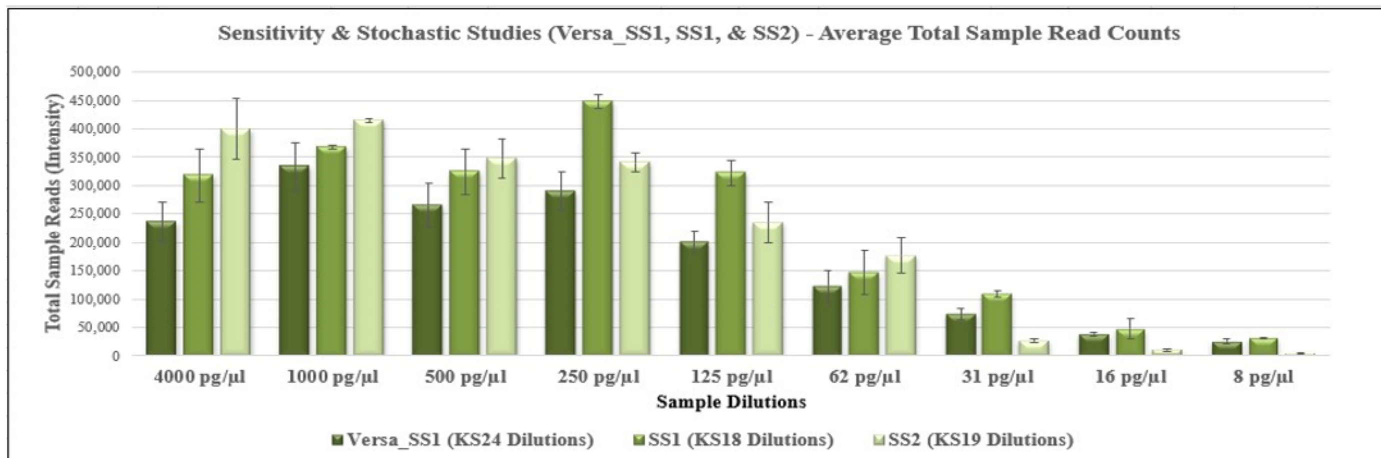


Figure 4: Sensitivity and Stochastic study using the ForenSeq™ Kintelligence kit

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 setup 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 set-ups to obtain full profiles in ranges from 4 ng down to 63 pg, with profiles suitable for comparison still being obtained at inputs as low as 8 pg (Figure 5).



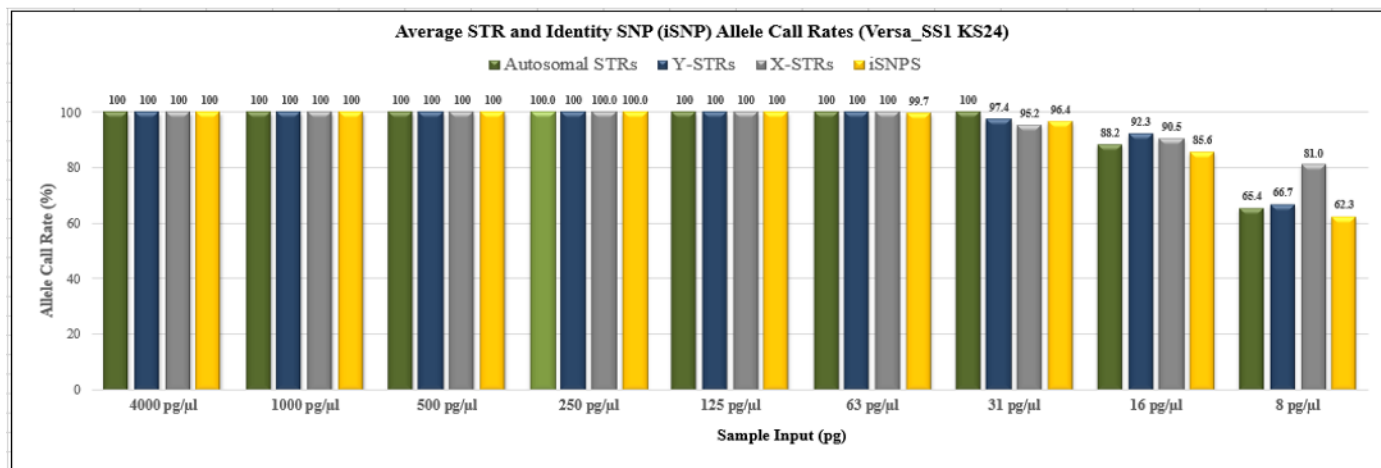


Figure 5: Total sample read counts and call rates for sensitivity and stochastic studies using the ForenSeq™ DNA Signature Prep kit

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 control region are shown in Figure 6 with comparable total number of reads being generated across the manual and automated workflows. The lowest input sample of 10pg 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 as low as 20 pg of DNA. Allele drop-out was observed in the 2 – 10 pg inputs, however, 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 as observed in CE-based PCR kits, which are exacerbated with decreasing inputs to PCR (Figure 6).

Aurora Illuminating Solutions
DNA LABS INTERNATIONAL
ISO/IEC 17025:2017 and FBI QAS DNA Laboratory
VEROGEN

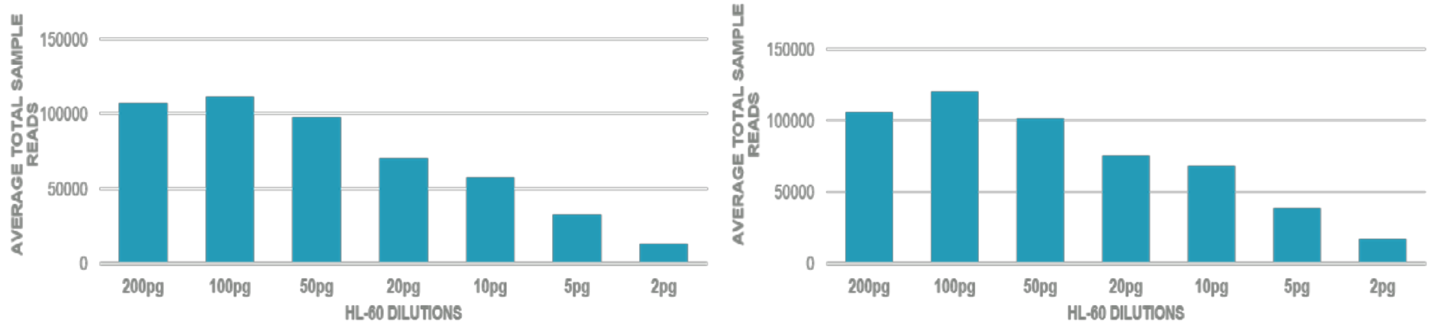


Figure 6: Total sample read counts for SS1 mtDNA control region. Total sample read counts for SS1 outside mtDNA control region

Conclusions

Data from these internal validation studies were evaluated to determine performance and utility of the Aurora's VERSA 1100 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 when used with Verogen products for sequencing of forensic samples

Aurora
Illuminating Solutions
DNA LABS INTERNATIONAL
ISO/IEC 17025:2017 and FBI QAS DNA Laboratory
VEROGEN