Automation of SOPHiA Clinical Exome Solution on Hamilton STARlet

Authors: Pantelis Constantoulakis[1], Jean-François Vanbellinghen[2], Jose Rincon Garriz[2], Maja Garbulinska[2], Lionel Ausset[3], and Federico Sebastiani[3]
[1] BioAnalytica-Genotypos, Ilision 3-5, 115 28, Athens
[2] SOPHiA GENETICS, Rue du Centre 172, 1025 Saint-Sulpice, Switzerland
[3] Hamilton Bonaduz AG, Via Crusch 8, 7402 Bonaduz, Switzerland

Introduction

Next-Generation Sequencing (NGS) is transforming the way experts analyze inherited diseases by enabling the screening of multiple genes with a single assay. However, the complex and cumbersome library preparation has been identified as a significantly challenging workflow. In fact, library preparation protocols usually consist of multistep processes and require costly reagents and substantial hands-on time.

An automated solution minimizes manual intervention, saving technicians' time and in parallel offering high quality libraries for reliable sequencing.

- **Standardized and integrated workflow for advanced analytical performance**
- **Increased productivity and reduced human intervention**
- **Reduced bias due to manual sample preparation**

The BioAnalytica-Genotypos center in Athens chose to couple the Hamilton STARlet robot with the SOPHiA Clinical Exome Solution (CES) for dealing with the increasing workload they were facing. In fact, over the past decade, the number of samples and types of analyses processed in the laboratory increased exponentially, leading to a pressure to develop, optimize, and validate NGS high-throughput assays.

To ensure high quality results on the new workflow and facilitate its implementation, the automated protocol has been extensively tested by SOPHiA GENETICS and Hamilton on a smaller representative panel of 128 genes. The validation study ensured a reliable and flexible automation, which has been efficiently applied on larger gene panels, including CES.

SOPHiA GENETICS™

HAMILTON

Method Description and Protocol

The BioAnalytica-Genotypos center in Athens has automated CES on the Hamilton STARlet robot to manage the increased number of samples to process. 200 ng of DNA extracted from blood was used for library preparation and target capture that were performed over 2 working days.

Pre and post-PCR workflows were performed on the same deck layout. The Pre-PCR workflow was performed on day 1, starting from diluted genomic DNA at the proper working concentration, placed in an input plate. As a result, amplified genomic libraries were obtained. These libraries were used to perform the target-enrichment workflow during the post-PCR phase on day 2.

Extracted DNA was processed on the Hamilton STARlet through an input worklist. The automated workflow processes up to 48 samples per run in 2 days, including PCR amplification on the deck of the platform.

Libraries generated from genomic DNA were sequenced on an Illumina NextSeq® 550 sequencing platform with a Mid Output (2x151bp) flow cell kit. Sequencing output files were then analyzed by the SOPHiA DDM™ platform.

Validation study

The Genomics Laboratory of SOPHiA GENETICS generated 2 sets of data (each with 16 different samples, pooled by 8 in 2 captures) with a mix of probes (with representative AT- and GC-rich regions). One set of data was generated following the manual workflow of the 128 gene solution, and the other one using the STARlet instrument, for which specific scripts were developed to automate the library preparation and capture enrichment protocols. Capture efficiency of the scripts was validated using one mix of probes with a target region of 497kb. The enriched libraries were sequenced on an Illumina MiSeq® instrument using v3 chemistry. Demultiplexing and data analysis were performed and reported by the SOPHiA DDM™ platform.

SOPHiA Clinical Exome Solution

The SOPHiA Clinical Exome Solution is a genomic application that bundles a smart capture-based target enrichment kit with the analytical power and advanced features of the SOPHiA DDM™ platform.

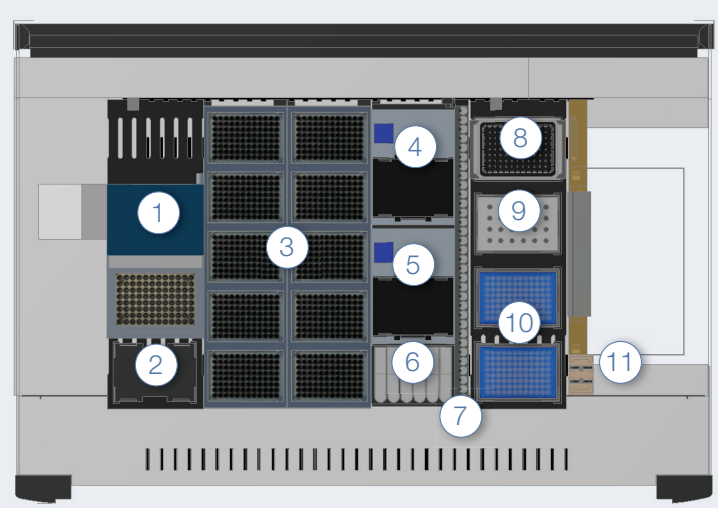
The solution was expertly designed to cover the coding regions (± 5 bp of the intronic regions) of 4,490 genes with known association with various Mendelian diseases, such as autism, inflammatory bowel disease, cardiomyopathies, neuromuscular disorders, epileptic and convulsive disorders, or hereditary cancer, among others. In particular, CES provides efficient detection and characterization of multiple types of variants such as SNVs, Indels, and CNVs in one single experiment. Capture probes are highly optimized to provide highest on-target rates and lowest noise levels, providing optimal input for the data analysis by the SOPHiA DDM™ platform. The results are then displayed on the platform for experts' variant interpretation.

Deck layout

STARlet 8 channels, no autoload; clear cover with UV lamp for deck sterilization; 1x ODTC; 1x PCR plate cooling module; 1x vial cooling module; 1x32 vial carrier; 1x MFX carrier for troughs containing ethanol; 1x shaker carrier with 2x plate modules, 1x Teleshaker and 1x Chemagic magnetic stand; 1x reagent carrier with 3x120 ml troughs as waste (Fig. 1).

Automation technology

The STARlet is an ideal platform for the automation of the SOPHiA Clinical Exome Solution protocol as it is equipped with all the positions and devices required for this NGS workflow. In combination with proven Hamilton technologies like CO-RE, MAD, and cLLD, the process can be run with high robustness.



1. *On-Deck Thermal Cycler*
2. *Lid Park Position*
3. *Tips*
4. *Cooling Module PCR Plate*
5. *Cooling Module Tubes*
6. *Reagent Troughs*
7. *Tube Carrier*
8. *Shaker*
9. *Magnet*
10. *PCR Plate Positions*
11. *CO-RE Gripper*

Figure 1: Deck layout of the system used in this study.

Results

Reducing hands-on time

Hamilton and SOPHiA GENETICS have worked together to optimize each step of the workflow. The automated protocol is carried out over two days, with the first day covering the library preparation and QC, while the second day used for hybridization capture and cleanup.

The automated protocol drastically reduces hands-on time to 1h 30min for analyzing 48 samples, whereas the manual workflow requires 6h35 for analyzing 24 samples (see table 1).

		Automated preparation	Manual preparation
	Step	Hands-on time for 48 reactions (Execution on the Hamilton STARlet)	Hands-on time for 24 reactions
Day 1	Enzymatic fragmentation End repair & A-tailing Adapter ligation Post-ligation cleanup Dual size selection PCR setup	45’ (3h)	3h
	Post-PCR cleanup	10’ (45’)	45’
	Dilutions for quantification	5’ (5’)	5’
Day 2	Buffers dilution Hybridization (4h or O.N.) Streptadividin bead washes Capture washes PCR setup Post-capture PCR cleanup	30’ (2h)	2h45’
	TOTAL HANDS-ON TIME	1h 30’	6h 35’

Table 1: Overview of the two day workflow. Total hands-on time required for the manual process is 6h 35 minutes (24 samples) and 1h 30 minutes for the automated process (48 samples).

Ensuring advanced analytical performance on targeted applications

The validation study showed that the Hamilton STARlet robot generated NGS libraries with yields and sizes comparable with the manual protocol. Libraries obtained with manual and automated workflows have been then used in duplicate to perform the next steps of the protocol and prepare the capture-based enrichment step. All assessed performance metrics, including coverage uniformity, on-target and off-target rates were comparable between the two protocols (see performance metrics comparison in table 2 and figures 2 and 3). In particular, the per sample coverage uniformity of both manual and automated workflow was very high with values above 99,9% and did not differ significantly between the workflows. No bias has been found for AT-rich regions (in red on the plot of figure 2) compare to the GC-rich ones (in blue on the plot of figure 2). On-target and off-target rate values differed by no more than 0.05 percentage points. Per sample median read coverage was similar for both experiments.

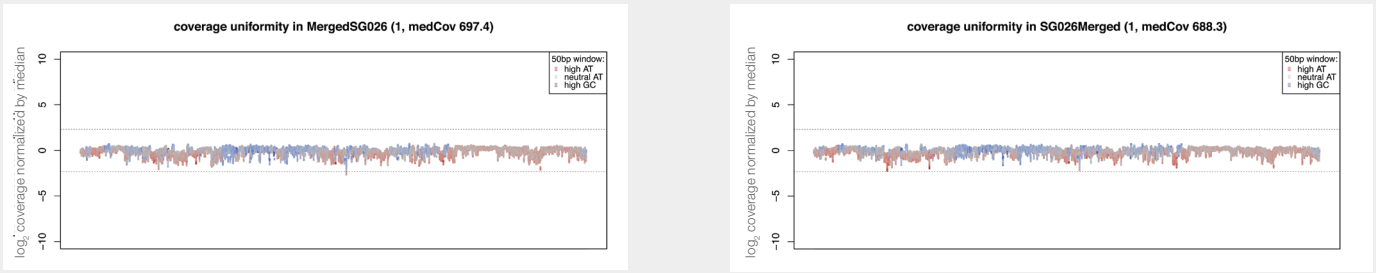


Figure 2: Coverage uniformity plot of one exemplary sample obtained with the automated (left) and manual protocol (right). The X-axis shows the genes included in the application, and the Y-axis the log_2 normalized coverage by the median. The closer the dots are to the 0 line, the more homogenous the reads are covering each target. Dashed lines represent 20% (lower line) and 500% (upper line) of the median coverage. The great majority of dots fall between the 2 dotted lines, meaning that the coverage is uniform including AT- and GC-rich regions.

Performance metrics per sample	Automated	Manual
Median coverage uniformity	99.99%	100%
Median on-target rate	76.51%	76.56%
Median off-target rate	3.18%	3.21%
Median read coverage	630x	656x
Average read number	3,832,267	3,952,151

Table 2: Performance metrics of the evaluation study comparing the automated vs manual protocol.

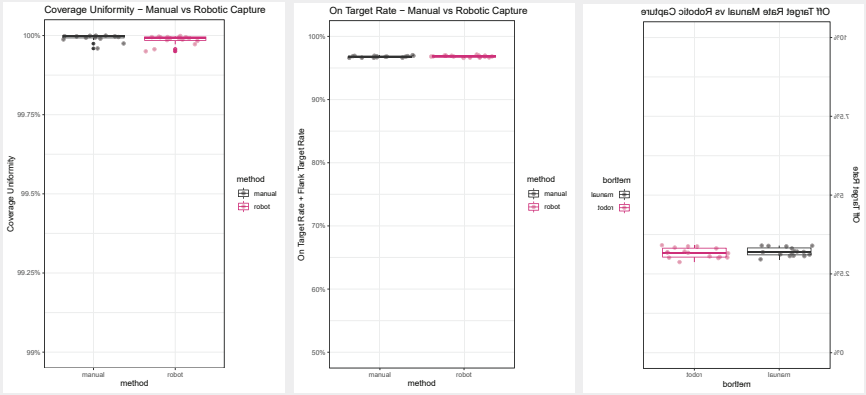
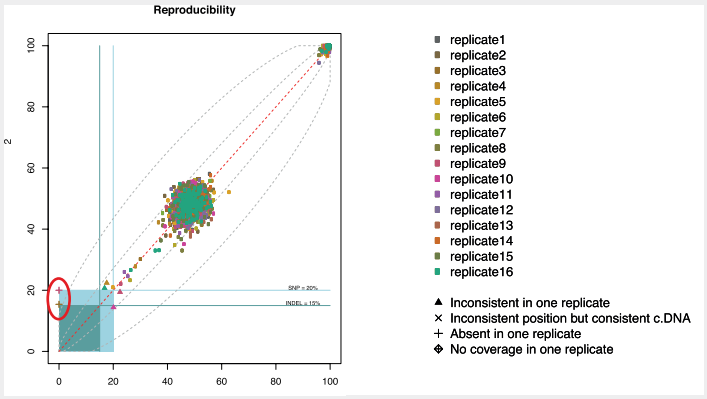


Figure 3: Boxplots of coverage uniformity, on-target rate, and off-target rate for the automated vs the manual protocol. The distribution of the points and the medians are comparable between the two protocols (in grey the automated and in pink the manual) for all the three measured performance metrics.

Furthermore, a head-to-head comparison between the reported variant fractions shows a reproducibility R squared of 0,99 for the manual vs the automated workflows. This value has been decreased by two inconsistent variants that appeared in the run, corresponding to the manual capture preparation workflow. Based on previous data, we concluded that those variants were detected due to the noise in this particular run, rather than the bias related to the method of capture preparation. Both variants were detected in a noisy homopolymer region where such inconsistency is expected (see figure 4).

Figure 4: Reproducibility plot comparing variant fraction between replicates obtained with the automated vs. the manual protocol. The variant fractions, depicted by the colored dots, are typically 0.5 (heterozygous) or 1.0 (homozygous), as expected for germline variants. The grey, dotted lines represent the 5% and 10% deviation from identity (diagonal = red dashed line). The blue squares represent the low variant fraction cut-off (SNP=20%, Indel=15%). Note that the majority of dots are within the 5% deviation from identity limits, which shows high reproducibility between the two methods. The two inconsistent variants identified in the automated manual protocol are circled in red.



Reaching high-quality results on exome analyses

BioAnalytica-Genotypos laboratory obtained optimal results by automating the CES workflow. The automated solution showed a very uniform coverage even in GC-rich regions, as in the first exon (see figure 5). Moreover, multiple types of genomic variants were accurately detected in a single experiment, including CNVs (see example in figure 6).

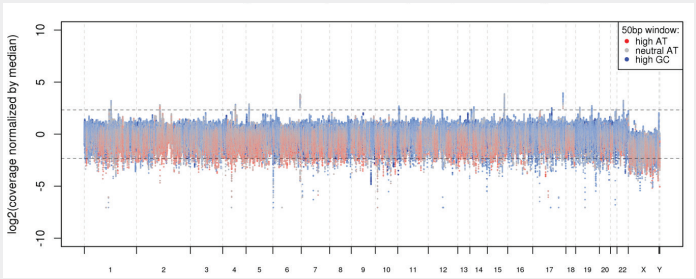


Figure 5: Coverage uniformity plot of one exemplary sample analyzed by the Hamilton STARlet automated CES protocol. The X-axis shows the chromosomes covered by the application and the Y-axis the $\log_2$ normalized coverage by the median. The closer the dots are to the 0 line, the more homogenous the reads are covering each target. Dashed lines represent 20% (lower line) and 500% (upper line) of the median coverage. The great majority of dots fall between the 2 dotted lines, meaning that the coverage is uniform including AT- and GC-rich regions.

Figure 6: CNV detection with the automated CES protocol in a real case linked to cardiac sudden death. Blue dots correspond to target regions without CNVs, red dots indicate the presence of a heterozygous deletion, covering exons 14-15 of KCNH2 gene in chromosome 7. Solid dots represent high-confidence CNV predictions.

SOPHiA GENETICS and Hamilton

The use of Hamilton STARlet instrument in combination with the SOPHiA GENETICS solutions provides a standardized and flexible workflow with advanced performance on targeted and exome applications.

The automation of the SOPHiA Clinical Exome Solution on the Hamilton Starlet robot has allowed BioAnalytica-Genotypos center to increase sample throughput and offer high quality results.

SOPHiA GENETICS products are for Research Use Only and Not for Use in Diagnostic Procedures.

Lit. No. AN-2103-03 — 03/2021

HAMILTON®

www.hamiltoncompany.com
infoservice@hamiltonrobotics.com

United States
Tel: +1-775-858-3000
United Kingdom & Ireland
Tel: +44 (0) 121 272 92 80
Spain, Portugal
Tel: +34 930 186 262
China
Tel: +86 21 6164 6567
France
Tel: +33 184 008 420
Italy
Tel: +39 (039) 930 06 06
Denmark, Norway, Sweden, Finland
Tel: +46 8403 09 090
Germany, Switzerland, Austria, Benelux
Tel: +49 89 248 804 804

To find a representative in your area, please visit hamiltoncompany.com/contacts.