Type of the Paper: Article

Automated capture-based NGS workflow: one thousand patients experience in a clinical routine framework

Elena Tenedini [1], Fabio Celestini [2], Pierluigi Iapicca [3], Marco Marino [1], Sara Castellano [4, 5] Lucia Artuso [1], Fiammetta Biagiarelli [3], Laura Cortesi [6], Angela Toss [6, 7], Elena Barbieri [6], Luca Roncucci [4], Monica Pedroni [4], Rossella Manfredini [8], Mario Luppi [4, 6], Tommaso Trenti [1] and Enrico Tagliafico [1, 4, 9,*]

[1] Department of Laboratory Medicine and Pathology, Diagnostic Hematology and Clinical Genomics Unit, Modena University Hospital, Modena, Italy
[2] Hamilton Italia S.r.l. Agrate Brianza (MB), Italy
[3] SOPHiA GENETICS SA HQ, Saint-Sulpice, Switzerland
[4] Department of Medical and Surgical Sciences, University of Modena and Reggio Emilia, Modena, Italy
[5] PhD program in Clinical and experimental medicine (CEM), University of Modena and Reggio Emilia, Modena, Italy
[6] Department of Oncology and Hematology, Modena University Hospital, Modena, Italy
[7] Department of Surgery, Medicine, Dentistry and Morphological Sciences with Transplant Surgery, Oncology and Regenerative Medicine Relevance, University of Modena and Reggio Emilia, Modena, Italy.
[8] Life Sciences Department University of Modena and Reggio Emilia, Centre for Regenerative Medicine, Modena, Italy;
[9] Center for Genome Research, University of Modena and Reggio Emilia, Modena, Italy.
* Correspondence: enrico.tagliafico@unimore.it; Tel.: +39 059 2055387

Simple Summary: Genomics is increasingly pervading the precision medicine in clinical practice. Tailored cancer prevention strategies, first-line treatment decisions or even surgical options are increasingly based on NGS multigene mutational screenings. However, good laboratory practice for clinical NGS procedures often do not go at the same speed as implementation of informatics analysis. Many companies deliver comprehensive range of kits for genomic testing with a labor-intensive manual sample preparation, even if IVD certified. Capture-based target enrichment protocols have been optimized to generate NGS libraries with very high coverage uniformity. Nevertheless, these workflows are well known to consist of multiple and hands-on demanding steps and so prone to human errors. We performed a validation study in more than 1000 samples demonstrating that the workflow automation standardizes the analytical performances and decreases variability providing more reliable results.

Abstract: (1) Background: the NGS based mutational study of hereditary cancer genes is crucial to design tailored prevention strategies in subjects with different hereditary cancer risk. The ease of amplicon-based NGS library construction protocols contrasts with the greater uniformity of enrichment provided by capture-based protocols and so with greater chances for detecting larger genomic rearrangements and copy-number variations. Capture-based protocols, however, are characterized by a higher level of complexity of sample handling, extremely susceptible to human bias. Robotics platforms may definitely help dealing with these limits, reducing hands-on time, limiting random errors and guaranteeing process standardization. (2) Methods: We implemented and validated the complete automation of the SOPHiA GENETICS' CE-IVD Hereditary Cancer Solution™ (HCS) libraries preparation workflow on the Hamilton's STARlet platform. (3) Results: We demonstrate that this automated workflow, used for more than 1000 samples achieved the same performances of manual setup in terms of coverages and reads uniformity, with extremely lower variability of reads mapping rate onto the regions of interest. (4) Conclusions: This automated solution offers same reliable and affordable NGS data, but with the essential advantages of a flexible, automated and integrated framework, minimizing possible human errors and depicting a laboratory's walk-away scenario.

Keywords: Next Generation Sequencing; Laboratory automation; Hereditary Cancer; Genetic Testing; Clinical Genomics.

1. Introduction

Several omics approaches have been described so far with the potential to lead information and improvement in many aspects of human life, particularly in the healthcare system, for prevention, diagnosis, clinical knowledge and, of course, treatment of diseases. Clinical genomics represents the paradigm of omics and the introduction of NGS technologies has been the foundation for its exponential increase. Advances in sequencing platforms for genomic applications has led to the development of much more informative assays. They range from screening the genetic landscape of a single cell to hundreds of patients together, or they can focus onto few genes' hotspots as well as onto the whole human genome [1-4]. As companies improved their NGS platforms, the interpretation tools grew in number integrating with artificial intelligence programs [5]. Many technical approaches have been introduced to improve NGS working protocols and libraries set-up in particular, to obtain more and more information from data, like copy number variation analysis [6]. So, genomics has put under progressive pressure the traditional clinical diagnostic laboratory to expand its sequencing ability and data analysis skills [7]. The ever-growing workload comes with the same reporting turnaround times and eventually the same lab staff [8]. Thus, genomics concurred pushing the molecular diagnostic laboratory to adopt a laboratory medicine model [9], where a large-scale automation is ubiquitously supporting every analytical component of the total testing process, contributing at the quality of analytical performance and at controlling the high cost of delivering genomics diagnostic services.

Hereditary cancer risk screening with NGS multigene panels, has become the most effective method for programming cancer prevention strategies [10-14]. NGS libraries preparation for these panels are designed either with amplicon-based or with hybridization capture-based target enrichment methods. Amplicon-based approaches are simpler and ask for a very little DNA input, but it has been demonstrated that hybridization-based procedures are less likely to generate false positives and false negative single nucleotide variations (SNVs), and notably they perform better in terms of coverage uniformity, which is essential for correctly predicting large rearrangements and CNVs [15,16]. Among targeted hybridization-based capture approaches, we adopted the SOPHiA Hereditary Cancer Solution (HCS), a CE-IVD certified application able to identify simultaneously SNVs, indels and CNVs in all the 26 tested genes and differentiate pseudogene variants in PMS2 https://www.sophiagenetics.com/hospitals/solutions/solutions/HCS.html). Nevertheless, this workflow is well known to consist of multiple and hands-on demanding steps, possibly prone to human bias. We thought that automation may definitely help dealing with these limitations, especially in a clinical routine framework of a public healthcare system, where diagnostics have to work in a cost-effective manner. We therefore directed our interest towards a liquid handler that could integrate all the devices necessary for the execution of the protocol, such as, thermocyclers, vortexes, magnets, cooling stations, plates sealing and UV decontaminating systems [17]. The platform had to prepare the working mixes directly from the original tubes or plates contained in the kits, as well as contemplate scalable procedures to run the protocol with an intermediate number of samples between the minimum and the maximum that can be sequenced together. With these premises, we implemented the automated SOPHiA HCS library preparation workflow on the Hamilton's STARlet platform and adopted it in the diagnostic routine. The results of the validation work were summarized in a diagnostic application note (https://www.hamiltoncompany.com/press-releases/application-note-automation-of-the-hereditary-cancer-solution-hcs-by-sophia-genetics-on-a-starlet#top).

After processing, sequencing and analysing more than 1000 genomic DNA samples, we compared the NGS results carried out with this automated protocol with the ones carried out manually on 240 samples, collected in the SOPHiA HCS performance evaluation study (https://www.sophiagenetics.com/fileadmin/documents/Solutions/HCS/HCS-bySG_ApplicationNote.pdf) and we get some conclusions about of NGS data robustness and essential benefits of automation.

2. Materials and Methods

Sample collection and DNA isolation

Peripheral blood samples (PB) were collected following the standard procedure for diagnostic testing after written informed consent, in accordance with the current revision of the Helsinki Declaration Genomic DNA was extracted with the DNA Midi Kit via QIASymphony platform (Qiagen); nucleic acids' quantity/quality were checked by Qubit dsDNA High Sensitivity kit and Nanodrop (Thermo Scientific). Samples from SOPHiA GENETICS' performance evaluation study, were obtained from PB as well. Quantity and quality were checked according to HCS protocol's guidelines.

Libraries set-up and Sequencing

Sequencing libraries were prepared using the CE-IVD SOPHiA HCS v1.1 kit, exclusively with the automated procedure implemented on the STARlet platform (Hamilton) as cited before. Individual library quantification was performed via fluorometric quantitation by Qubit dsDNA High Sensitivity kit (Thermo) and quality control analysing the profile of each sample via capillary electrophoresis with Bioanalyzer DNA 1000 (Agilent Technologies). In the routine of our medium-throughput laboratory, the number of samples per preparation was 24 that run onto a 600-cycle format V3 flow-cell, sequenced via Illumina MiSeq DX platform according to Illumina's and SOPHiA GENETICS' protocols. Libraries from SOPHiA GENETICS' performance evaluation study, instead, were manually prepared using both the SOPHiA HCS v1.1 and HCS_M_v1, that did not contain probes for the APC gene. But as well as in our laboratory they were run onto the same flow-cells and sequenced via Illumina MiSeq platforms.

Data analysis and variants interpretation criteria

The SOPHiA HCS allows the enrichment of coding and splicing regions of 26 genes (*APC, ATM, BARD1, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, EPCAM, FAM175A, MLH1, MRE11A, MSH2, MSH6, MUTYH, NBN, PALB2, PIK3CA, PMS2, PTEN, RAD50, RAD51C, RAD51D, STK11, TP53, XRCC2*) and the pseudogene PMS2CL, well-known associated with increased risk for cancer syndromes. The sequencing data were simultaneously processed for single nucleotide variants (SNVs), indels, and copy number variations (CNVs) using the SOPHiA DDM software (DDM) updated al the last available version at the time of sequencing. Data analysis and variant interpretation were limited to virtual panels of actionable genes, in accordance with the informed consent expressed and signed by the patients. For the assessment of breast/ovarian/pancreatic risk *APC, ATM, BARD1, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, EPCAM, MLH1, MSH2, MSH6, MUTYH, NBN, PALB2, PMS2, PTEN, RAD50, RAD51C, RAD51D, STK11, TP53* were analysed; for suspected Lynch syndrome *MLH1, MSH2, MSH6, PMS2, EPCAM, MUTYH*; for suspected familial adenomatous polyposis *APC, MUTYH, PTEN, STK11*; for suspected hereditary gastric cancer *ATM, BARD1, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, FAM175A, MRE11A, NBN, PALB2, PIK3CA, RAD50, RAD51C, RAD51D, TP53, XRCC2*.

Genetic variants annotations were also integrated with data present in literature and open source bioinformatics tools customized and validated in the laboratory (Annovar [18] and Variant Effect Predictor (VEP), [19]) and specific databases, Leiden Open source Variation Database (https://grenada.lumc.nl/LOVD2/mendelian_genes/home.php?), ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/), 1000 Genomes Project (http://www.1000genomes.org/data), ExAC (http://exac.broadinstitute.org/), dbSNP (http: //www.ncbi.nlm.nih .gov / projects / SNP /), The Genome Aggregation Database (http://gnomad.broadinstitute.org/), BRCA Share (http://www.umd.be/BRCA1/ http://www.umd.be/BRCA2/).

Variants were reported using the international standard HGVS nomenclature and classified into 5 categories: pathogenic (P), likely pathogenic (LP), variant of uncertain significance (VUS), likely benign (LB) and benign (B), according to the American College of Medical Genetics and Genomics (ACMG) criteria [20].

NGS variants confirmation

All the gene variants interpreted as pathogenic or likely pathogenic were confirmed by Sanger sequencing [21], performed with predesigned primers and the BigDye Direct Cycle Sequencing Kit. They were sequenced via Applied Biosystems 3500xL Dx Genetic Analyzer platform and results analysed by SeqScape3 software (Thermo Scientific) updated to the latest available version. All the CNVs interpreted as pathogenic or likely pathogenic were confirmed by MLPA (MRC-Holland) and analysed with the Coffalyser.Net software (MRC-Holland) updated to the latest available version

3. Results

Library preparation is a laborious process with many hands-on sessions: samples can be prepared in parallel but each one must be set in a different well before they are pooled for target enrichment. Moreover, paramagnetic bead-based purification steps are several and tricky, and execution speed is critical.

Donut graphs (Fig.1) of the two working days, summarize the multiple steps of the manual protocol and show how this automated procedure definitively impacts the hands-on time, letting the staff to accomplish other lab's activities, even if it not necessarily shortening the overall execution time.

The laboratory needed to plan a medium-throughput workflow, setting the number of routine samples per preparation to 24. Thawing the 48 samples-packaged reagents more than once was therefore a need. Hence, the work of the Hamilton platform was designed for both the two days preparing all the working mixes at the beginning of the running session and keep them at right temperatures in the cooling stations. This allows technicians to put back quickly the reagents' leftover to the freezer/fridge, accomplishing a fast robot's set up and coming back at the end of the working day. The platform was indeed programmed to work autonomously until the libraries need to be quantified and pooled -on day one- or the captured pools have to be checked to be sequenced -on day two- (Fig1).

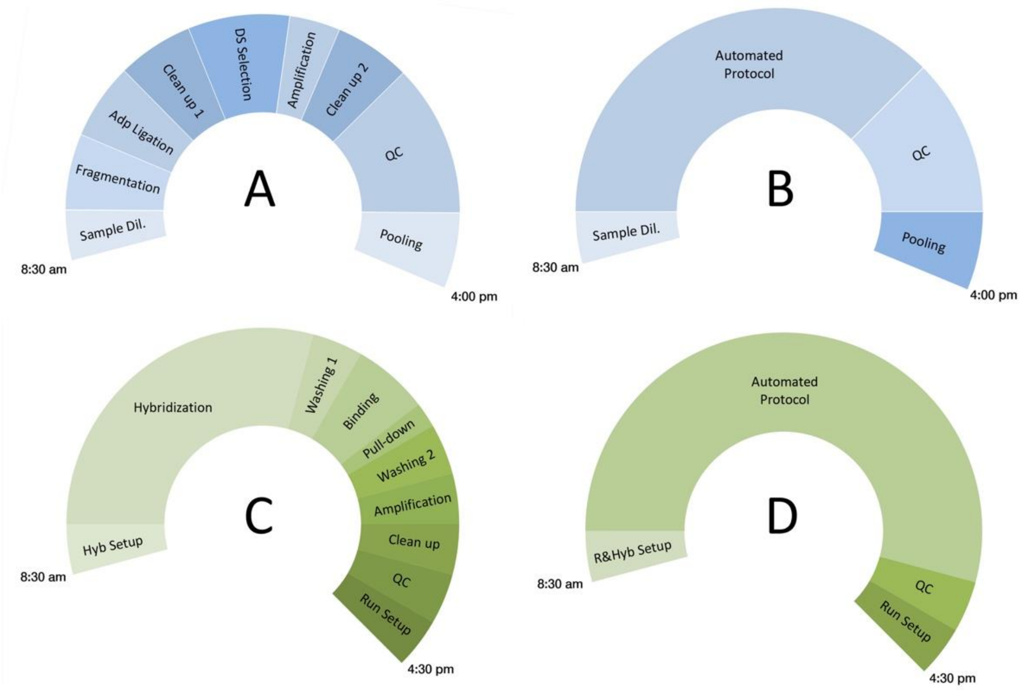

Figure 1. Donuts show the different steps performing in two days: day 1, manual protocol (A), day 1, automated protocol (B), day 2, manual protocol (C); day 2, automated protocol (D). Abbreviations: (Day 1): Sample Dil.: sample dilutions; Fragmentation: mixes setup, enzymatic fragmentation, end repair and A-tailing; Adp Ligation: Adapter ligation; Clean up 1: Post ligation clean up;

DS Selection: dual size selection; Amplification: pre-capture amplification; Clean up 2: post amplification clean up; QC: Libraries quantification and size control; Pooling: libraries pooling and lyophilization; (Day 2): Hyb Setup: hybridization mixes setup, sample denaturation, addition of probes, StB Washing: streptavidin beads washing; Binding: Probe-target duplexes to streptavidin beads binding; Pull Down: DNA-beads complexes pull down; Washing 2: DNA-beads complexes washing; Amplification: post-capture amplification; Clean up: amplification clean up; QC: Libraries quantification and size control.

3.1. Generate good intermediate results and robust NGS data

Intermediate check points, in SOPHiA HCS protocol, gave very good results in individual pre-capture library quantification and quality control. The fluorometric quantification of individual pre-capture libraries resulted in an average concentration equal to 71,6 ng/ul (SD=14,62), that means 1289,8 nanograms total/sample (SD=263,17), almost ten times the 150 nanograms needed quantity to be pooled (12 samples/pool). Besides, individual pre-capture matched the requested size distribution of DNA fragments between 300 and 700bp (data not shown).

The final post-capture libraries pools provided excellent results: they should have a size distribution between 300bp and 700bp and the average size we obtained was as good as 453bp (SD=18,7) and an average concentration equal to 26,82ng/ul (SD=9,4). Therefore, according to SOPHiA HCS working protocol, the average post-capture pool molarity was calculated in 91,5nM (SD=33,2), which is greatly more than the 10pM molarity to be sequenced.

No samples were therefore ever discarded either from pooling or from sequencing, even those with very low concentrated genomic DNA.

NGS data were obtained onto 600-cycle format V3 flow cells with an Illumina MiSeq DX platform. All the runs had a satisfactory cluster density range between 1000 and 1200K/mm2, as Illumina recommended for MiSeqDX, and an average Q(30) score equal to 89,25% (SD=0,021).

Therefore, no run had to be repeated.

FASTQ files were batch uploaded to DDM platform: the quality report results were analysed, and statistics for reads quality, alignment, mapping and coverage evaluated and compared to those obtained from manually prepared samples in SOPHiA HCS performance evaluation study.

The samples we sequenced showed a good total number of reads: an average of 2,287 million reads per sample was the number of all the reads that were used as input for the alignment process and 2,245 million reads that were successfully mapped to the reference genome, that is the 98,16% of reads. The percentage obtained from manually prepared libraries from SOPHiA GENETICS’ study is almost overlapping, 98,65%. Data are collected in Table 1.

Therefore, we assume that the automated protocol implemented on Hamilton’s platform, generates libraries producing a high-quality mapping to reference genome, with no presence of contaminating DNA.

Mapping statistics continued with analysis of what fraction of reads maps exactly on regions of interest (“on-target”), next to the regions of interest ("flankTarget") or outside the region we are interested in (“off-target”): a high number of off-target sequencing reads generally indicate an issue and can decrease the power of calling genetic variants.

The 60% of all mapped base pairs we obtained mapped to the target regions (on-target fraction, light green in Figure 2), while this percentage raised to 66.4% in manual prepared libraries.

The 16,8% of all mapped base pairs mapped in flanking regions, that means they mapped within the fragment length from the target (fragment length is usually defined as the median read length) (blue in figure) and the percentage is similar (15.9%) to manual performance data. The remaining mapped base pairs mapped off-target, accumulating with high coverage in specific places in the genome, for example in pseudogenes in off-target regions, or scattering across the genome with low coverage.

Table 1. NGS performance comparison between automated and manual protocol

	Automated Protocol	Manual Protocol
reads mapped to the reference genome	98,16%	98,65%
reads mapped “on-target”	60%	66.4%
reads mapped "flankTarget"	16,80%	15.9%
reads mapped “off-target” high coverage	13,50%	12,20%
reads mapped “off-target” low coverage	9,70%	5,50%
target regions showing at least 50x coverage	99,99%	99,97%
target regions showing at least 200x coverage	99,95%	99,85%
target region coverage uniformity	99,90%	99,70%

The 13,5% of mapped base pairs we obtained accumulated off-target but in specific genome region (yellow in figure), whereas manually prepared showed an average of 12,2%. Low coverage off-target base pairs accounted, instead, for 9,7% of mapped base pairs in our data (dark green in the figure), while manual protocol stood at 5,5% (Table1).

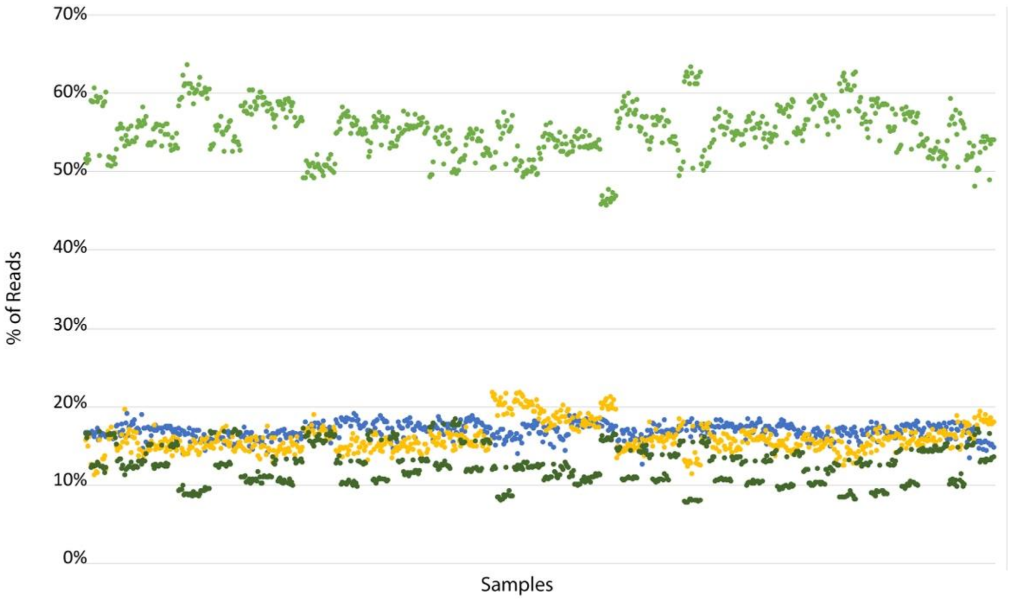

Figure 2. Reads base mapping distribution. On the x axis all the samples processed, on the y axis the percentage of reads base on-target (light green), mapping within a fragment length of target region (blue), mapping off target accumulating in specific region (yellow) and off target scattering across the genome with low coverage (dark green).

So, looking at the reported percentages spanning across all the runs, it would seem that the manual protocol average performances were better than the automatic ones, because the “on-target” fraction was higher (66, 4% vs 60%) and vice versa the low-coverage “off-Target” fraction is lower (5,5% vs 9,7%), but if we observe standard deviations (Table 2) we realize that the variability of the results was much greater in the manual performances if compared to the automated ones. Indeed, the dispersion from the average was very much higher (see Table 2). Actually, as expected, these results confirmed that performing numerous steps manually can introduce variability in the sequencing results, even when performed with a well-equipped and experienced laboratory staff.

Table 2. onTarget, flankTarget, offTarget_HighCov, offTarget_LowCov mapped reads standard deviations

	Automated Protocol	Manual Protocol

onTarget	0,08	6,59
flankTarget	0,01	2,41
offTarget_HighCoverage	0,04	2,12
offTarget_LowCoverage	0,04	4,51

Thus, libraries prepared with the automated protocol showed a higher reproducibility but a lower fraction of reads mapping "on-target" regions. This anyway does not impact the target regions' coverage, which is an important prerequisite for reliable variant calling in NGS data. All sequenced samples, in the format 24 samples/run, showed indeed a high on-target coverage with 99,99% of target regions showing at least 50x coverage and 99,95% have 200x coverage, that is the requested minimum coverage for germline samples to run the SOPHiA's CNV algorithm. Manual prepared libraries, despite the higher fraction of on-target base pairs and the same 24 samples/run, showed overlapping target coverage percentages (50x=99,97%; 200x=99,85%) (Table1). Then, all positions in the target regions were verified for coverage heterogeneity/uniformity. Coverage heterogeneity is calculated as the percentage of base pairs, in the target region, whose coverage is lower than 0.2 of the median target coverage or higher than 5x the median target coverage. The lower is the number, higher is the consistency of the assay. Heterogeneity was definitely very low in our NGS data, with average uniformity equal to 99,9% [SD=0,52] overlapping to Sophia's manually prepared samples (99,7%) [SD=0,43] (Table1). We can conclude that the automated protocol generates good libraries and sequencing data fully comparable with those obtained with manual protocol, but with robustness of automation.

3.2. Checking variants calling accuracy of the data generated with the automated protocol

All the data were processed using the DDM software updated al the last available version at the time of sequencing. The software was set to prefilter genomic data of each sample according to the obtained informed consent. So, genomic variants accuracy was limited to the actionable genes as explained in Material and Methods.

Analyzed samples are summarized in Table 3. The vast majority of samples were obtained from patients with a family history for breast, ovarian or pancreatic cancers (90,5%), some from patients with suspected Lynch Syndrome (5,2%) or with suspected Familial Adenomatous Polyposis (3,7%). A very limited number had a double clinical suspect (0,5%) and just one patient was tested for suspected Hereditary Gastric Cancer.

Table 3. Analyzed samples

	BC/OC/P	Lynch	FAP	Lynch + FAP	Gastric	Total
# of tested samples	915	53	37	5	1	1011
% of tested samples	90,5%	5,2%	3,7%	0,5%	0,1%	100%

Genetic annotations present in variant caller files generated from DDM platform were integrated with open-source bioinformatics tools validated in the laboratory, updating the functional annotations and increasing information for interpreting the biological function of observed variants (see Materials and Methods).

Gene-sequence variants were reported into a 5-tier system according to the ACMG classification rubric. This is summarized in Figure 3.

To check the accuracy of the obtained results, internal and external quality assessment were adopted: first, confirming Likely Pathogenic and Pathogenic variants via

doi:10.20944/preprints202103.0103.v1

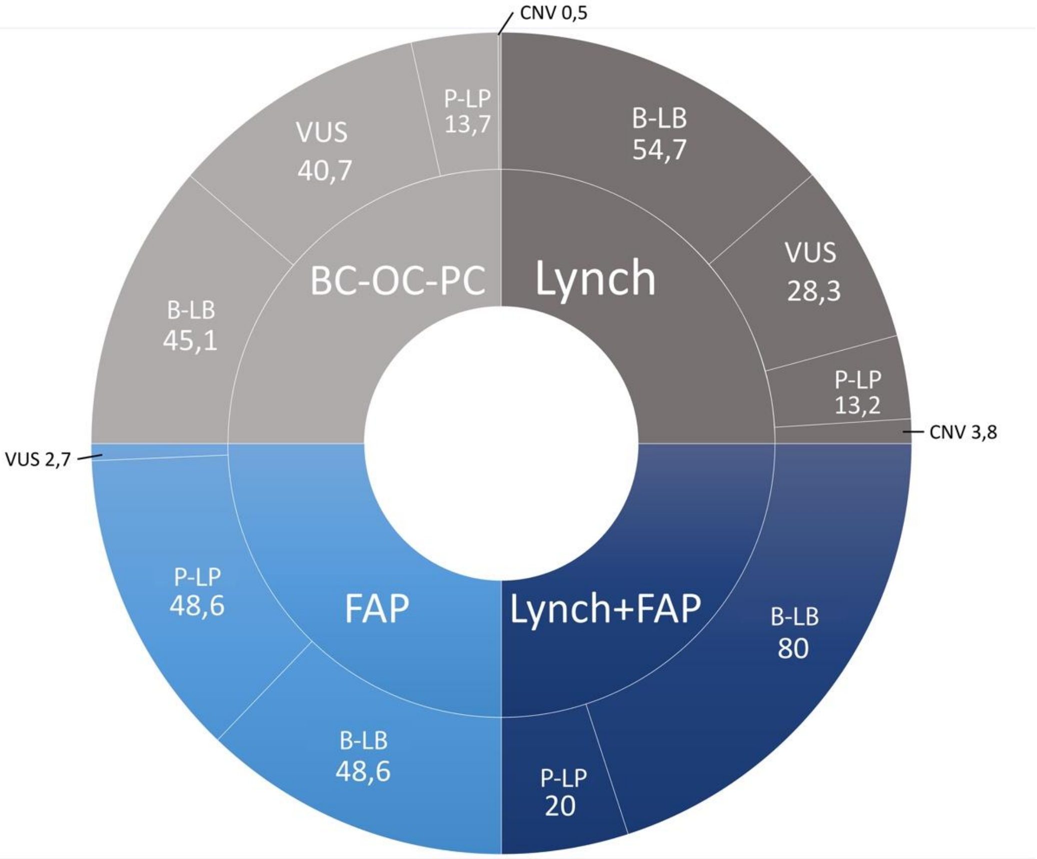

Figure 3. Classification of variants. Inner circle represents the four tested groups of patients: oncologic patients with family history for breast, ovarian or pancreatic cancers (BC-OC-PC, twenty-two genes analysed), patients with suspected Lynch Syndrome (Lynch, six genes analysed); suspected Familial adenomatous polyposis (FAP, four genes analysed) and patients with a double clinical suspect (Lynch+FAP, ten genes analysed). The very only patient with suspected hereditary gastric cancer is not shown. Outer circle sections represent variants annotation classes and percentages: polymorphisms, benign and likely benign variants (P-B-LB), variants with uncertain significance (VUS), likely pathogenic (LP) and pathogenic variants (P), copy number variations (CNVs).

Sanger sequencing and validating copy number variations by MLPA; secondly, re-evaluating previously *BRCA1-BRCA2* tested patients and finally performing this automated protocol on samples from an international laboratory quality assessment program.

Among the entirety of about one thousand patients we tested, 14,7% carried one variant classified as pathogenic or probably pathogenic (149/1011) and 0.89% (9/1011) carried two Likely Pathogenic and/or Pathogenic variants: compound heterozygosis (3/9) or double heterozygosity (6/9). Pathogenic or likely pathogenic variants were found in 18 different genes (*APC, ATM, BARD1, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, MLH1, MSH2, MUTYH, NBN, PALB2, PMS2, RAD50, RAD51C, RAD51D, TP53*), accounting for 167 variants that were confirmed by Sanger sequencing. A 100% concordance was achieved, none turned out to be a false positive. Confirmation of CNVs, instead, has been realized by MLPA in 16 samples that showed deletions in one of the following genes: *ATM, EPCAM, MLH1, MSH2 or PALB2*. We were able to confirm all of them. In one patient we couldn't check the "undetermined" copy number of exon2 in *BARD1* gene, because the MLPA kit is not available.

One third of patients (309 subjects) with a family history for breast, ovarian or pancreatic cancers, had already been tested for *BRCA1* and *BRCA2* variants and no pathogenic variations were found. Some of them were tested by Sanger sequencing long ago and others more recently via an amplicon-based ion semiconductor NGS method. After the re-sequencing with this automated protocol, 7 of them turned out to be actually positive for Likely Pathogenic or Pathogenic variants in either *BRCA1* or *BRCA2* gene. Unsuccessful previously performed Sanger sequencing lacked the complete screening of CDS or intron-exon junction regions. Inefficient NGS results with Ion Torrent technology, instead, was imputable to the presence of samples carrying variants in homopolymeric regions with stretches greater than 7/ 8 nucleotides or deletions in problematic regions.

Finally, we could test this automated protocol with samples from the European Molecular Genetics Quality Network assessment program (EMQN, www.emqn.org), in 2019 and 2020 HBOC, FAP and LYNCH schemes. We tested 18 samples, recognizing all the 16 positive samples for SNVs, indels or CNVs and the two negative ones. Likely pathogenic or pathogenic variants were confirmed successfully via Sanger sequencing. No false positive or false negative results. EMQN evaluation assigned a comprehensive fully satisfactory judgment to genotyping, interpreting and clerical job.

4. Discussion

Diagnostic laboratories are commonly part of a wider clinical frame, where the inflow of samples to be tested with the NGS is not generally made of large, constant and fixed samples' cohorts, but of a dynamic and even intermittent rate of smaller number of samples. Clinicians are increasingly using the results of NGS tests (somatic, but even germline) to take first-line treatment decisions or choose among surgical options [22]. Therefore, the laboratory requires to expand its analytical portfolio but also keep dealing the increasing workload, improving standardization and managing rapid turnaround times (TAT).

In our opinion, the automation of capture-based libraries is one of the reasonable answers to reduce NGS error prone procedures and improve standardization during the whole workflow. Comparing the sequencing results of the HCS automated protocol on Hamilton's platform to the ones obtained manually, we proved that an accurate automated design, on a flexible and integrated platform, can free from the complexity of hybridization-based capture libraries procedures. It minimizes the risk of human-introduced errors, standardizing the analytical performances and decreasing sequencing data variability. Variant calling accuracy resulted very good, as far as no false positive likely pathogenic or pathogenic variants were found. Additionally, the tested samples from external quality assessment programs returned fully satisfactory results, i.e. no false negative or false positive results. Finally, this automated hybridization-based capture approach, combined with fluorescent reversible terminator nucleotides' sequencer, confirmed its superior ability to detect genetic variations in critical genomic regions, like homopolymeric stretches [23].

In a more comprehensive laboratory picture, the automation allowed a more efficient working agenda and an improvement of the samples' flow. Besides, using robotics the laboratory scheduled precise reagents and plastics consumes accomplishing a better supply and making NGS more affordable. Moreover, the complete integration of all devices in the platform's deck (thermal cyclers, shakers, cooling stations, magnetic stands, decontamination devices) facilitated the internal and the external maintenance programs. To make the economic investment of automation paid off, we tested the Hamilton Platform with other capture-based commercial or custom SOPHiA GENETICS gene panels, to diagnose hereditary rare diseases or hematological cancers. The automated solution showed broad versatility, with easily scaling number of samples and reagents volumes, as well as adjustments to hybridization and amplification programs to adapt to different working protocols.

The next challenge will be integrating and interfacing this clinical genomics laboratory automation with the Laboratory Information System, using specific middleware. An effort in this direction lies in the power of the automation hardware, as well as in genomic analysis companies. The existing NGS LIMS solutions, currently designed for genomics research laboratories, need to be improved to be suitable also for diagnostic genomics units integrated into a diagnostic clinical environment.

5. Conclusions

This study suggests that automation may definitely help dealing with NGS capture-based target enrichment protocols. We optimized an automated procedure on the Hamilton Starlet that reduced the hands-on time for the preparation of SOPHiA GENETICS HCS libraries, improved reproducibility and reduced the variability in NGS data. The quality of sequencing data we obtained, confirmed accuracy in variant calling, as far as no false positive likely pathogenic or pathogenic variants were found, even in problematic genomic regions like homopolymers.

Author Contributions: Conceptualization, E.Te. and E.Ta.; methodology, E.Te, F.C., M.M., L.A., and F.B.; software, P.I., and S.C.; validation, E.Te., F.C., M.M., L.A.; formal analysis, E.Te. F.C., S.C., L.A., M.M. L.C., A.T., E.B., L.R., and M.P.; investigation, E.Te., R.M., M.L., T.T., and E.Ta.; data curation, E.Te., and S.C.; writing—original draft preparation, E.Te.; writing—review and editing, E.Te., R.M., M.L., T.T. and E.Ta.; supervision, E.Ta. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki. Ethical review and approval were waived for this work since this study reports data obtained during the standard procedures of internal quality controls required for diagnostic laboratories and for which the approval of the ethics committee is not required.

Informed Consent Statement: Informed consent for genomic testing was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Harbeck, N.; Penault-Llorca, F.; Cortes, J.; Gnant, M.; Houssami, N.; Poortmans, P.; Ruddy, K.; Tsang, J.; Cardoso, F. Breast cancer. *Nat Rev Dis Primers* **2019**, *5*, 66, doi:10.1038/s41572-019-0111-2.
2. Neben, C.L.; Zimmer, A.D.; Stedden, W.; van den Akker, J.; O'Connor, R.; Chan, R.C.; Chen, E.; Tan, Z.; Leon, A.; Ji, J., et al. Multi-Gene Panel Testing of 23,179 Individuals for Hereditary Cancer Risk Identifies Pathogenic Variant Carriers Missed by Current Genetic Testing Guidelines. *J Mol Diagn* **2019**, *21*, 646-657, doi:10.1016/j.jmoldx.2019.03.001.
3. Susswein, L.R.; Marshall, M.L.; Nusbaum, R.; Vogel Postula, K.J.; Weissman, S.M.; Yackowski, L.; Vaccari, E.M.; Bissonnette, J.; Booker, J.K.; Cremona, M.L., et al. Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genetics in medicine : official journal of the American College of Medical Genetics* **2016**, *18*, 823-832, doi:10.1038/gim.2015.166.
4. Crawford, B.; Adams, S.B.; Sittler, T.; van den Akker, J.; Chan, S.; Leitner, O.; Ryan, L.; Gil, E.; van 't Veer, L. Multi-gene panel testing for hereditary cancer predisposition in unsolved high-risk breast and ovarian cancer patients. *Breast Cancer Res Treat* **2017**, *163*, 383-390, doi:10.1007/s10549-017-4181-0.
5. Cobain, E.F.; Milliron, K.J.; Merajver, S.D. Updates on breast cancer genetics: Clinical implications of detecting syndromes of inherited increased susceptibility to breast cancer. *Semin Oncol* **2016**, *43*, 528-535, doi:10.1053/j.seminoncol.2016.10.001.
6. Choi, M.; Kipps, T.; Kurzrock, R. ATM Mutations in Cancer: Therapeutic Implications. *Mol Cancer Ther* **2016**, *15*, 1781-1791, doi:10.1158/1535-7163.MCT-15-0945.
7. Renault, A.L.; Mebirouk, N.; Fuhrmann, L.; Bataillon, G.; Cavaciuti, E.; Le Gal, D.; Girard, E.; Popova, T.; La Rosa, P.; Beauvallet, J., et al. Morphology and genomic hallmarks of breast tumours developed by ATM deleterious variant carriers. *Breast Cancer Res* **2018**, *20*, 28, doi:10.1186/s13058-018-0951-9.
8. Jerzak, K.J.; Mancuso, T.; Eisen, A. Ataxia-telangiectasia gene (ATM) mutation heterozygosity in breast cancer: a narrative review. *Curr Oncol* **2018**, *25*, e176-e180, doi:10.3747/co.25.3707.
9. Marabelli, M.; Cheng, S.C.; Parmigiani, G. Penetrance of ATM Gene Mutations in Breast Cancer: A Meta-Analysis of Different Measures of Risk. *Genet Epidemiol* **2016**, *40*, 425-431, doi:10.1002/gepi.21971.

10. Vignudelli, T.; Selmi, T.; Martello, A.; Parenti, S.; Grande, A.; Gemelli, C.; Zanocco-Marani, T.; Ferrari, S. ZFP36L1 negatively regulates erythroid differentiation of CD34+ hematopoietic stem cells by interfering with the Stat5b pathway. *Mol Biol Cell* **2010**, *21*, 3340-3351, doi:10.1091/mbc.E10-01-0040.

11. Marino, M.; Rabacchi, C.; Simone, M.L.; Medici, V.; Cortesi, L.; Calandra, S. A novel deletion of BRCA1 gene that eliminates the ATG initiation codon without affecting the promoter region. *Clin Chim Acta* **2009**, *403*, 249-253, doi:10.1016/j.cca.2009.02.020.

12. Shafman, T.; Khanna, K.K.; Kedar, P.; Spring, K.; Kozlov, S.; Yen, T.; Hobson, K.; Gatei, M.; Zhang, N.; Watters, D., et al. Interaction between ATM protein and c-Abl in response to DNA damage. *Nature* **1997**, *387*, 520-523, doi:10.1038/387520a0.

13. Weber, A.M.; Drobnitzky, N.; Devery, A.M.; Bokobza, S.M.; Adams, R.A.; Maughan, T.S.; Ryan, A.J. Phenotypic consequences of somatic mutations in the ataxia-telangiectasia mutated gene in non-small cell lung cancer. *Oncotarget* **2016**, *7*, 60807-60822, doi:10.18632/oncotarget.11845.

14. Kozlov, S.V.; Graham, M.E.; Jakob, B.; Tobias, F.; Kijas, A.W.; Tanuji, M.; Chen, P.; Robinson, P.J.; Taucher-Scholz, G.; Suzuki, K., et al. Autophosphorylation and ATM activation: additional sites add to the complexity. *J Biol Chem* **2011**, *286*, 9107-9119, doi:10.1074/jbc.M110.204065.