

Challenges in obtaining high-quality data from a custom-made panel for the next generation sequencing (NGS) using Ion Torrent GeneStudio™ S5 platform

Lana Salihefendić^{1,2}, Adna Ašić^{2, *}, Ivana Čeko¹, Dino Pećar¹, Larisa Bešić², Naida Mulahuseinović¹, Amra Džuhović², Laura-Severina Köhn², Rijad Konjhodžić¹

¹Alea Genetic Center, Bosnia and Herzegovina

²Department of Genetics and Bioengineering, Faculty of Engineering and Natural Sciences, International Burch University, Bosnia and Herzegovina

* Corresponding author: Email: adna.asic@ibu.edu.ba

Received November 4th, 2021

Accepted for publication February 24th, 2022

Online First February 28th, 2022

Keywords: COVID-19, next-generation sequencing, Ion Torrent GeneStudio™ S5, immunological response.

Abstract

The goal of this part of the study was to optimize the sequencing procedure for 16 human genes and their regulatory regions that might be associated with differential immunological response to COVID-19. The study was performed on 60 COVID-19 patients from the General Hospital of Tešanj, Bosnia and Herzegovina, categorized into three groups of mild, moderate, and severe clinical manifestation, based on the diagnosis by the residential physician. Target coding sequences and their regulatory regions were amplified for the following genes: *HLA-A*, *HLA-B*, *HLA-C*, *ACE2*, *IL-6*, *IL-4*, *TMPRSS2*, *IFITM3*, *IL-12*, *RIG-I/DDX58*, *IRF-7*, *IRF-9*, *IL-1B*, *IL-1A*, *CD55*, and *TNF-α*. DNA was isolated from the whole blood samples stored at -20°C for six months using QIAamp® DNA Mini Kit according to manufacturer's instructions. Since NGS analysis of target genomic regions was performed on the Ion Torrent GeneStudio™ S5 platforms, libraries were prepared using Ion AmpliSeq™ Library Kit Plus according to manufacturer's instructions in a protocol optimized for low-quality DNA. Due to dissatisfactory sequencing results, further protocol optimization steps were employed through separating two primer pools, increasing the number of PCR cycles, and decreasing the annealing temperature for the primer pool which showed poorer amplification results. In the end, 36 samples produced optimal results, while the remaining 24 samples will be re-sequenced following repeated sample collection and DNA isolation, accompanied by additional protocol modifications.

Introduction

Considering the global impact of COVID-19 in terms of its consequences on healthcare, economy and social norms of behavior, multidisciplinary studies of issues related to this pandemic have been the focus of the scientific research for the past two years.

Coronaviruses belong to the subfamily *Coronavirinae*, a member of *Coronaviridae* family (Luo et al, 2020). The genome of SARS-CoV-2 is an RNA-based genome that is 29,881 bp in size, encoding 9,860 amino acids (Chen et al, 2020). Symptoms and clinical manifestation of

COVID-19 show high variability. These manifestations include pneumonia with variable severity, acute respiratory distress syndrome (ARDS), as well as a significant fraction of asymptomatic carriers. Most of the patient's experience fever, cough, dyspnea, myalgia, and fatigue. A poorer outcome and prognosis are associated with older age, pre-existing chronic conditions (diabetes, cardiovascular and respiratory diseases) and male sex (Alahmad et al, 2020).

Next-generation sequencing (NGS) or multiparallel sequencing enables obtaining sequence information for

the entire genomes. There is a number of different NGS platforms which can perform parallel sequencing of millions of small DNA fragments. Each of these fragments is sequenced multiple times, thus providing high sequencing depth with the goal of delivering accurate data and providing information on unexpected and/or novel gene variants. There are various potential uses of NGS in clinical practice, including capturing a broader spectrum of mutations than Sanger sequencing, high sensitivity in detection of mosaic mutations, and detection of rare mutations or mutations of low frequency (for example, in tumor tissues and liquid biopsy) (Behjati & Tarpey, 2013).

We hereby report technical issues encountered while sequencing a panel of 16 genes of interest and their regulatory regions for 60 COVID-19 patients from the General Hospital of Tešanj, Bosnia and Herzegovina. This research was performed using semi-conductor sequencer GeneStudio™ S5 (Ion Torrent, Thermo Fisher Scientific, Waltham, MA).

Materials and methods

Whole blood samples were obtained from 60 COVID-19-positive patients examined at the General Hospital of Tešanj, Bosnia and Herzegovina. All collected samples were classified into three groups

TABLE 1. Classification of the study participants into three clinical manifestation groups (mild, moderate, and severe) based on observed symptoms (table taken and adapted from Baj et al, 2020).

Clinical manifestations	Imaging findings	Other	Classification (Baj et al, 2020)	Classification (current study)
Absence of typical and atypical clinical manifestations	No changes in CT and X-ray scans	Positive tests for SARS-CoV-2	Asymptomatic	Mild
Typical: fever, dry cough, sore throat, runny nose, sneezing, fatigue, myalgia, tiredness, muscle ache, headache, loss of smell Atypical: nausea, vomiting, diarrhea, abdominal pain	Usually, no changes in CT and X-ray scans	Absence of any manifestations of pneumonia	Mild	
Presence of symptoms of pneumonia, fever (usually persistent and higher than 37.8°C) with a dry cough	Usually, the presence of ground-glass opacities and lung consolidation	Crackles or wheezing during pulmonary auscultation	Moderate	Moderate
Dyspnea, hypoxia, and >50% lung involvement, diarrhea, vomiting, nausea	Ground glass opacities, lung consolidation, pleural effusions, lymphadenopathy	Continually worsening clinical outcome of patients	Severe	Severe
Severe difficulty of breathing and shortness of breath, chest pain, movement impairments, loss of speech	Ground glass opacities (usually bilateral), lung consolidation, pulmonary nodules	Continually worsening clinical outcome of patients Most prevalent complications: acute respiratory distress syndrome or respiratory failure, myocardial injury, arrhythmia or heart failure, acute kidney injury, acute liver injury, encephalopathy, disseminated intravascular coagulation, rhabdomyolysis, septic shock, multiple organ dysfunctions	Critical	

according to the clinical manifestation of COVID-19, namely mild, moderate, and severe clinical manifestation (including five deceased patients). Patient classification into the study groups was performed by residential physicians based on the previously published guidelines (Baj et al, 2020; Table 1). Prior to sample collection, ethical clearance to conduct this research was obtained from the Ethics Committee of the Faculty of Engineering and Natural Sciences, International Burch University (Sarajevo, Bosnia and Herzegovina), as well as from the Joint Ethics Committee of the General Hospital of Tešanj (Tešanj, Bosnia and Herzegovina). All participants signed an informed consent form of voluntary participation in the research.

Samples were collected in November 2020, stored at -20°C and delivered to the laboratory for analysis six months later. Following delivery, samples were immediately de-frosted, and DNA was extracted using QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. Isolated DNA was quantified using Qubit® 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA), according to manufacturer's instructions, with quantification results ranging from 5 ng/μl to 40 ng/μl. Samples with concentration lower than 10 ng/μl were re-extracted and higher concentrations obtained.

NGS panel was custom-made for 16 genes, including *HLA-A*, *HLA-B*, *HLA-C*, *ACE2*, *IL-6*, *IL-4*, *TMPRSS2*, *IFITM3*, *IL-12*, *RIG-I/DDX58*, *IRF-7*, *IRF-9*, *IL-1B*, *IL-1A*, *CD55* and *TNF-α*. Genes of interest were chosen based on their candidate status for modulating patients' response to SARS-CoV-2 infection, either due to their protein products being the members of the host immune system, or encoding host-cell receptors for viral entry into the cell. In addition, previous research on candidate genes for variable response to SARS-CoV or SARS-CoV-2 infection was consulted (Asselta et al, 2020; Chapman & Hill, 2012; de Lang et al, 2006; Lingeswaran et al, 2020; Lipworth et al, 2020; Trowsdale & Knight, 2013). *In silico* coverage was 98% at 100x sequencing depth. Total amplicon number was 185, with one primer pool with 92 and the second pool with 93 amplicons (Table 2).

TABLE 2. The number of primer pairs used for sequencing 16 target genes and their regulatory regions..

Gene	Number of primer pairs (amplicons)
<i>HLA-A</i>	11
<i>HLA-B</i>	10
<i>HLA-C</i>	11
<i>ACE2</i>	22
<i>IL-6</i>	7
<i>IL-4</i>	5
<i>TMPRSS2</i>	19
<i>IFITM3</i>	3
<i>IL-12</i>	7
<i>RIG-I/DDX58</i>	23
<i>IRF-7</i>	15
<i>IRF-9</i>	10
<i>IL-1B</i>	10
<i>IL-1A</i>	8
<i>CD55</i>	17
<i>TNF-α</i>	7
Total	185

According to DNA concentrations, libraries were prepared using Ion AmpliSeq™ Library Kit Plus (Thermo Fisher Scientific). Ion Torrent GeneStudio™ S5 (Thermo Fisher Scientific) was used as a sequencing platform, and it requires a minimum 1 ng of DNA. Sixteen libraries were prepared from the mild clinical manifestation group and protocol was immediately adjusted due to the fact that the blood samples used for DNA isolation were not used immediately after collection. The protocol was adjusted for low-quality DNA, meaning that three cycles were added in the amplification PCR and initial DNA input in the reaction was 20 ng. Following amplification, partial digestion of amplicons was achieved using exonucleases and two primer pools were pooled into one. Following the incubation step, barcodes were ligated on the samples and prepared libraries were purified using Agencourt AMPure XP reagent (Beckman Coulter, Brea, CA) and 70% freshly prepared ethanol. Purified libraries were quantified using Ion Library TaqMan™ Quantitation Kit (Thermo Fisher Scientific) on real-time PCR. Library quantification values were all higher than 100 pM and were diluted to 80 pM, according to manufacturer's instructions.

Results and Discussion

Sequencing results for the first group of 16 samples were satisfactory and no further optimization of the protocol presented above was planned at that point. However, when the second set of 16 samples was sequenced, the conflicting results were obtained. While some of the samples showed high coverage of 100x 100%, others showed 1x 50% coverage, even though all libraries had the same starting concentration. Namely, library quantification analysis gave the results ranging from 100 pM to 250 pM. All libraries were therefore diluted to the final concentration of 100 pM to achieve uniform library concentration in the final pool. Since the primer panel was custom-made, subsequent optimization steps, aimed at improving the coverage, included another library preparation step with protocol modification in which two primer pools were prepared separately for each sample in order to test and optimize the primers. After applying the same protocol, library quantification showed large difference between the library concentrations in two primer pools, which partially explained conflicting coverage on NGS (Table 3). Further optimization of the protocol was made by adding two more PCR cycles to the primer pool 2 and, finally, by lowering the annealing temperature by 2°C

for the same pool. Obtained results were better, but still did not reach the desired quality of the output data.

Library preparation of the remaining 28 samples was later initiated to test the quality of samples and whether introduced protocol modifications were satisfactory. Library quantifications were still conflicting, so that out of a total of 60 samples, 36 gave optimal coverage results, while the analyses for the remaining 24 should be repeated.

As presented above, several experimental issues have been encountered while optimizing the NGS protocol for the target gene sequencing in a set of COVID-19 patients. Firstly, whole blood samples were collected from patients and kept frozen at -20°C for six months. While fresh samples are producing better results, it is not unusual to produce high-quality DNA isolated from samples that have been stored for several years (Tagliaferro et al, 2021). Also, QIAamp® DNA Mini Kit that was used in the present study is best valued for its time-effectiveness, lower possibility of sample contamination and ease of use (Chacon-Cortes & Griffiths, 2014). However, previous research has shown that for older samples, other DNA extraction methods, such as modified salting-out method, should be used (Mardan-Nik et al, 2019). Other studies have shown

TABLE 3. Library concentrations for two primer pools in the second set of 16 samples.

Sample no.	DNA concentration (ng/μl)	Library concentration, pool 1 (pM)	Library concentration, pool 2 (pM)
17	14	741	488
18	34	1069	4
19	16	191	1
20	22	100	0
21	15.7	68	9
22	6	1091	138
23	13.9	119	46
24	12	566	0
25	11.5	150	22
26	18	312	45
27	21.9	1102	69
28	9.2	870	15
29	43.6	1035	95
30	17.8	603	100
31	37.4	720	55
32	15.7	102	2

that silica-based DNA isolation using commercial kits, such as QIAamp® DNA Blood Midi Kit (Qiagen) (Mardan-Nik et al, 2019) and DNeasy® Blood & Tissue Kit (Qiagen) (Tagliaferro et al, 2021) are both capable of producing high-quality DNA extracts suitable for subsequent PCR amplification.

Furthermore, the custom-made primer panel that was produced for the purpose of this research has proved itself to be the major experimental obstacle in producing satisfactory data following NGS analysis of the samples. Inconsistent data produced for two primer pools, as well as conflicting results for different samples despite comparable starting DNA concentrations point out to this conclusion. Furthermore, the primers were designed *in silico* for the research-use-only (RUO) application and with the goal of primer panel validation. In order to improve the primer pool performance, protocol modifications were included according to the pre-established practice of increasing the number of PCR cycles to achieve target region amplification. Although the multiparallel sequencing platform was used and 16 genes and their regulatory regions are being analyzed at once, sequencing fragments of around 200 bp in length were still used in this research to achieve better sequence coverage and, therefore, higher result accuracy. Since the sequencing issues were encountered following these modifications, it is suggested that the primer pairs for different genes might differ in their annealing efficiency, as well as that excessive primer-dimer formation was encountered due to sequence complementarity.

Previous studies dealing with NGS data accuracy, sensitivity, specificity, and precision have highlighted the importance of using proper bioinformatics pipeline for the variant calling with the purpose of producing optimum results (Shin et al, 2017), no matter which NGS platform is used as a method of choice with an aim of replacing the Sanger sequencing method (Sandmann et al, 2017). In addition, the importance of proper library preparation is often emphasized as a crucial step in any successful NGS analysis (Forth & Hoepfer, 2019), including adding the proper amount of DNA and conversion of that DNA sample into a functional library that can be successfully sequenced. Previously published detailed technical notes emphasize the

importance of several critical steps in generation of high-quality NGS data, such as obtaining adequate amount and quality of DNA sample to be sequenced, decontamination of pre-amplification area, primer design, as well as gel test to check for the amplification of expected PCR products, amplicon purification and potential primer dimer formation (Wohlhieter et al, 2021).

Conclusion

Since it is generally accepted that the time elapsed from sample collection to sample analysis does not significantly influence the quality of isolated DNA, and based on the DNA quantification results, we can conclude that the sample quality was not a significant limitation in sequencing challenges encountered during this research. It is reasonable to assume that the custom-made primer panel failed to perform as expected in the NGS protocol. Following the sequencing protocol optimization and consulting literature sources explaining optimization of other custom-made sequencing panels, it was clear that accurate results cannot be obtained for 24 samples using hereby established protocol.

When it comes to future steps aimed at solving the issue of unsatisfactory library quantification results, the first option is to replace those samples with freshly collected whole blood samples and repeat the analyses. Furthermore, if this approach does not contribute towards solving the issue, primer pool separation will be employed in order to obtain satisfactory library quantification data for the amplicons contained in primer pool 1. Finally, missing amplicons will be generated using TruSight One Sequencing Panel (Illumina, San Diego, CA) for the clinical exome sequencing of more than 4,800 genes associated with human disease, including the genes of interest in the present research.

Acknowledgements

This research is a part of the project titled “Personalized approach to COVID-19 infection through analysis of molecular genetic predisposition of the patients for a differential immune response”, that is co-financed by

the Ministry of Science, Education, and Youth of the Sarajevo Canton (decision no. 11/05-34-12880-8/20).

The authors have no conflicts of interest to disclose.

References

- Alahmad, B., Al-Shammari, A. A., Bennakhi, A., Al-Mulla, F. & Ali, H. (2020). Fasting Blood Glucose and COVID-19 Severity: Nonlinearity Matters. *Diabetes Care*, 43(12), 3113-3116.
- Asselta, R., Paraboschi, E. M., Mantovani, A., & Duga, S. (2020). ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. *Aging*, 12(11), 10087–10098. doi:<https://doi.org/10.18632/aging.103415>.
- Baj, J., Karakuła-Juchnowicz, H., Teresiński, G., Buszewicz, G., Ciesielka, M., Sitarz, E., Forma, A., Karakuła, K., Flieger, W., Portincasa, P., & Maciejewski, R. (2020). COVID-19: Specific and Non-Specific Clinical Manifestations and Symptoms: The Current State of Knowledge. *Journal of Clinical Medicine*, 9(6), 1753. doi:<https://doi.org/10.3390/jcm9061753>.
- Behjati, S., & Tarpey, P. S. (2013). What is Next Generation Sequencing? *Archives of Disease in Childhood: Education and Practice Edition*, 98(6), 236-238. doi:10.1136/archdischild-2013-304340.
- Chacon-Cortes, D. F., & Griffiths, L. (2014). Methods for extracting genomic DNA from whole blood samples: current perspectives. *Journal of Biorepository Science for Applied Medicine*, 2014(2), 1-9. doi:<https://doi.org/10.2147/BSAM.S46573>.
- Chan J. F., Yuan S., Kok K. H., To K. K., Chu H., Yang J., Xing, F., Liu, J., Yip, C. C.Y., Poon, R. W. S., Tsoi, H. W., Lo, S. K. F., Chan, K. H., Poon, V. K. M., Chan, W. M., Ip, J. D., Cai, J. P., Cheng, V. C. C., Chen, H., Hui, C. K. M. & Yuen, K. Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, 395, 514–523. doi:[https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9).
- Chapman, S. J., & Hill, A. V. (2012). Human genetic susceptibility to infectious disease. *Nature Reviews: Genetics*, 13(3), 175–188. doi:<https://doi.org/10.1038/nrg3114>.
- de Lang, A., Osterhaus, A. D., & Haagmans, B. L. (2006). Interferon-gamma and interleukin-4 downregulate expression of the SARS coronavirus receptor ACE2 in Vero E6 cells. *Virology*, 353(2), 474–481. doi:<https://doi.org/10.1016/j.virol.2006.06.011>.
- Forth, L. F., & Höper, D. (2019). Highly efficient library preparation for Ion Torrent sequencing using Y-adapters. *BioTechniques*, 67(5), 229–237. doi:<https://doi.org/10.2144/btn-2019-0035>.
- Lingeswaran, M., Goyal, T., Ghosh, R., Suri, S., Mitra, P., Misra, S., & Sharma, P. (2020). Inflammation, Immunity and Immunogenetics in COVID-19: A Narrative Review. *Indian Journal of Clinical Biochemistry: IJCB*, 35(3), 260–273. doi:<https://doi.org/10.1007/s12291-020-00897-3>.
- Lipworth, B., Chan, R., & Kuo, C. R. (2020). Predicting Severe Outcomes in COVID-19. *The Journal of Allergy and Clinical Immunology in Practice*, 8(8), 2582–2584. doi:<https://doi.org/10.1016/j.jaip.2020.06.039>.
- Luo, H., Tang, Q. L., Shang, Y. X., Liang, S. B., Yang, M., Robinson, N. & Liu, J. P. (2020). Can Chinese Medicine Be Used for Prevention of Corona Virus Disease 2019 (COVID-19)? A Review of Historical Classics, Research Evidence and Current Prevention Programs. *Chinese Journal of Integrative Medicine*, 26(4), 243–250. doi:10.1007/s11655-020-3192-6.
- Mardan-Nik, M., Saffar Soflaei, S., Biabangard-Zak, A., Asghari, M., Saljoughian, S., Tajbakhsh, A., Meshkat, Z., Ferns, G.A., Pasdar, A., & Ghayour-Mobarhan, M. (2019). A method for improving the efficiency of DNA extraction from clotted blood samples. *Journal of Clinical Laboratory Analysis*, 33(6), e22892. doi:<https://doi.org/10.1002/jcla.22892>.
- Sandmann, S., de Graaf, A. O., van der Reijden, B. A., Jansen, J. H., & Dugas, M. (2017). GLM-based optimization of NGS data analysis: A case study of Roche 454, Ion Torrent PGM and Illumina NextSeq sequencing data. *PLoS one*, 12(2), e0171983. doi:<https://doi.org/10.1371/journal.pone.0171983>.
- Shin, S., Kim, Y., Chul Oh, S., Yu, N., Lee, S. T., Rak Choi, J., & Lee, K. A. (2017). Validation and optimization of the Ion Torrent S5 XL sequencer and OncoPrint workflow for BRCA1 and BRCA2 genetic testing. *Oncotarget*, 8(21), 34858–34866. doi:<https://doi.org/10.18632/oncotarget.16799>.
- Tagliaferro, S. S., Zejnelagic, A., Farrugia, R., & Wettinger, S. B. (2021). Comparison of DNA extraction methods for samples from old blood collections. *BioTechniques*, 70(5), 243-250. doi:<https://doi.org/10.2144/btn-2020-0113>.
- Trowsdale, J., & Knight, J. C. (2013). Major histocompatibility complex genomics and human disease. *Annual Review of Genomics and Human Genetics*, 14, 301–323. doi:<https://doi.org/10.1146/annurev-genom-091212-153455>.
- Wohlhieter, C. A., Uddin, F., Quintanal-Villalonga, À., Poirier, J. T., Sen, T., & Rudin, C. M. (2021). An optimized NGS sample preparation protocol for in vitro CRISPR screens. *STAR Protocols*, 2(2), 100390.