The Journal of Molecular Diagnostics, Vol. ■, No. ■, ■ 2021

the Journal of Molecular Diagnostics
jmdjournal.org

TECHNICAL ADVANCE

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Genome Sequencing from Post-Mortem Formalin-Fixed, Paraffin-Embedded Lung Tissues

Claude Van Campenhout,* Ricardo De Mendonça,* Barbara Alexiou,* Sarah De Clercq,* Marie-Lucie Racu,* Claire Royer-Chardon,* Stefan Rusu,* Marie Van Eycken,* Maria Artesi,† Keith Durkin,† Patrick Mardulyn,‡ Vincent Bours,†§ Christine Decaestecker,¶‖ Myriam Remmelink,* Isabelle Salmon,*‖ and Nicky D'Haene*

From the Department of Pathology, Erasme Hospital, the Evolutionary Biology and Ecology,‡ and the Laboratory of Image Synthesis and Analysis,¶ Université Libre de Bruxelles, Brussels; the Laboratory of Human Genetics,† GIGA Research Institute, Liège; the Department of Human Genetics,§ University Hospital of Liège, Liège; and DIAPath,‖ Center for Microscopy and Molecular Imaging, Université Libre de Bruxelles, Gosselies, Belgium*

Accepted for publication May 27, 2021.

Address correspondence to Nicky D'Haene, M.D., Ph.D., Department of Pathology, Erasme Hospital, 808 Route de Lennik, Brussels 1070, Belgium. E-mail: nicky.d.haene@erasme.ulb.ac.be.

Implementation of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) testing in the daily practice of pathology laboratories requires procedure adaptation to formalin-fixed, paraffin-embedded (FFPE) samples. So far, one study reported the feasibility of SARS-CoV-2 genome sequencing on FFPE tissues with only one contributory case of two. The present study aimed to optimize SARS-CoV-2 genome sequencing using the Ion AmpliSeq SARS-CoV-2 Panel on 22 FFPE lung tissues from 16 deceased coronavirus disease 2019 (COVID-19) patients. SARS-CoV-2 was detected in all FFPE blocks using a real-time RT-qPCR targeting the *E* gene with crossing point (Cp) values ranging from 16.02 to 34.16. Sequencing was considered as contributory (ie, with a uniformity >55%) for 17 FFPE blocks. Adapting the number of target amplification PCR cycles according to the RT-qPCR Cp values allowed us to optimize the sequencing quality for the contributory blocks (ie, 20 PCR cycles for blocks with a Cp value <28 and 25 PCR cycles for blocks with a Cp value between 28 and 30). Most blocks with a Cp value >30 were noncontributory. Comparison of matched frozen and FFPE tissues revealed discordance for only three FFPE blocks, all with a Cp value >28. Variant identification and clade classification was possible for 13 patients. The present study validates SARS-CoV-2 genome sequencing on FFPE blocks and opens the possibility to explore correlation between virus genotype and histopathologic lesions. *(J Mol Diagn 2021, ■: 1–13; https://doi.org/10.1016/j.jmoldx.2021.05.016)*

The coronavirus disease 2019 (COVID-19) pandemic is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Coronaviruses are a family of enveloped single-strand, positive-sense RNA viruses that cause a wide spectrum of respiratory diseases. Since the initial report on this novel coronavirus in Wuhan, China,[1–3] mortality and morbidity rapidly increased around the globe. Researchers worldwide are contributing to sequencing initiatives to try to understand how the virus is spreading. As of April 2021, up to 1,211,666 SARS-CoV-2 genomes were sequenced and uploaded to the Global Initiative on Sharing All Influenza Data (GISAID; *https://www.gisaid.org*, last accessed April 2021).[4] SARS-CoV-2 genome sequencing allows the

Supported by Fonds Y. Boël (Brussels, Belgium), Fonds Erasme pour la Recherche Médicale (Brussels, Belgium), Appel à projet Spécial COVID-19–ULB (Brussels, Belgium), and Appel Coronavirus–Crédits Urgents de Recherche, F.N.R.S. (Belgian National Fund for Scientific Research). The CMMI is supported by the European Regional Development Fund and the Walloon Region of Belgium (Wallonia-biomed; grant 411132-957270; project CMMI-ULB support the Center for Microscopy and Molecular Imaging and its DIAPath department).

Disclosures: C.D. is a Senior Research Associate with the F.N.R.S. (Belgian National Fund for Scientific Research).

https://doi.org/10.1016/j.jmoldx.2021.05.016

detection of genetic modifications that could have occurred. Most SARS-CoV-2 virus detection and genotyping methods are based on fresh samples from upper or lower respiratory tract, such as nasopharyngeal swab, oropharyngeal swab, sputum, or bronchoalveolar lavage. In the current COVID-19 pandemic, pathology laboratories face the major challenge to implement SARS-CoV-2 testing in their daily practice. In pathology laboratories, most surgical and cytology specimens are formalin fixed, paraffin embedded (FFPE). Post-mortem studies indicated that SARS-CoV-2 could be detected by RT-qPCR on FFPE blocks of lungs and other organs.[5–9] Adapting the SARS-CoV-2 genome sequencing protocols to FFPE blocks may provide valuable diagnostic tools for its detection and genotyping.

Virus sequencing can be achieved by Sanger[10,11] and/or next-generation sequencing (NGS).[10,12] NGS is now well implemented in pathology laboratories for detection of cancer-related molecular alterations, using FFPE tissues.[13,14] The use of targeted NGS panels allows the identification of tumor molecular profiles using small quantities of nucleic acids from FFPE blocks. However, only a few studies have reported the use of NGS to detect pathogens in FFPE blocks.[7,15–17] Sekulic et al[7] showed the feasibility of SARS-CoV-2 sequencing on FFPE blocks, but only one case of two was contributory. The present study aimed to optimize SARS-CoV-2 genome sequencing using NGS on 22 post-mortem FFPE tissues.

Materials and Methods

Clinical Series

Lung samples were collected from the 16 first confirmed COVID-19 (positive RT-qPCR assay on nasopharyngeal swab and/or bronchoalveolar lavage) patients who died in Hôpital Erasme (Brussels, Belgium) since March 13, 2020, and with a positive SARS-CoV-2 *E* gene RT-qPCR on lung FFPE blocks (see below). The study protocol was approved by the local ethics committee (P2020/218). The autopsy procedure, clinical courses, and histopathologic findings have been already described.[5] Briefly, six samples per lung lobe (ie, a total of 30 samples) were collected, formalin fixed, and paraffin embedded (except for two patients who had previously undergone lobectomy for cancer and for whom only 18 samples were taken). One or two blocks were randomly selected for molecular analysis among FFPE blocks showing histopathologic lesions. When two blocks were tested, they included one FFPE block from the left lung and one FFPE block from the right lung, to evaluate the heterogeneity of viral spread. Moreover, one sample was snap frozen for each lung lobe. The material was biobanked by the Biobanque Hôpital Erasme-ULB (BE_BERA1), CUB Hôpital Erasme, BBMRI-ERIC. Q10

Semiquantitative evaluation of hemorrhage on hematoxylin and eosin slides was performed by two senior pathologists (N.D. and M.R.) as follows: negative or <10% (0); between 10% and 20% of lung parenchyma showing intra-alveolar hemorrhage (+); between 20% and 30% of lung parenchyma showing intra-alveolar hemorrhage (++); and >30% of lung parenchyma showing intra-alveolar hemorrhage (+++). Evaluation of necrosis was also performed as follows: negative (0) or positive (+).

Nucleic Acid Extraction and SARS-CoV-2 Detection by RT-qPCR

For FFPE blocks, total nucleic acids were extracted from two unstained slides (10 μm thick) using the Maxwell RDC DNA FFPE kit and the Promega Maxwell extractor following the protocol described by the manufacturer (Promega Corp., Madison, WI) in an elution volume of 50 μL. For frozen tissues, RNAs were extracted using PureLink RNA Mini Kit (ThermoFisher Scientific, Waltham, MA) following manufacturer's instructions. The RNA yield was quantified using a Qubit 2.0 Fluorometer (ThermoFisher Scientific). For FFPE blocks, RNA quality was analyzed with the Agilent RNA 6000 Pico Kit on a Bioanalyzer 2100 (Agilent, Santa Clara, CA). The RNA from the FFPE blocks showed a fragmented profile, with a mean peak height of 130 nucleotides. The mean percentage of RNA fragments >200 nucleotides was of 60%, and no samples showed a percentage of RNA fragments >200 nucleotides of <30% (data not shown).

The detection of the SARS-CoV-2 virus in the nucleic acid extracts was performed by RT-qPCR. One-step RT-qPCR assay specific for the amplification of SARS-CoV-2 *E* gene was adapted from the protocol described by Corman et al[18] and as previously described.[5] Briefly, 100 ng of RNA was amplified in 20-μL reaction mixture containing 5 μL of 4× TaqMan Fast Virus 1-step master mix (ThermoFisher Scientific), 0.4 μmol/L of forward (5′-ACAGGTACGT-TAATAGTTAATAGCGT-3′) and reverse (5′-ATATTG-CAGCAGTACGCACACA-3′) primers, and 0.2 μmol/L of probe (5′-FAM-ACACTAGCCATCCTTACTGCGCTTCG-BBQ-3′). Amplification was performed on the LightCycler 480 type II (F. Hoffmann-La Roche SA, Basel, Switzerland) following the manufacturer's instructions. Amplification condition was 50°C for 10 minutes for reverse transcription, followed by 95°C for 20 seconds and then 45 cycles at 95°C for 3 seconds and 58°C for 30 seconds. Crossing point (Cp) values were calculated using the second derivative maximum method from the Roche LightCycler software. A clinical Q11 sample highly positive for SARS-CoV-2 (with a low Cp), diluted 1:1000, was used as positive control; and a clinical sample obtained from a patient autopsied before the pandemic was used as negative control in each analysis.

Library Preparation and Sequencing

For library construction, 10 ng of RNA (5 and 1 ng for testing robustness) was retrotranscribed with the SuperScript VILO (ThermoFisher Scientific) in accordance with

SCO 5.6.0 DTD ■ JMDI1089_proof ■ 17 July 2021 ■ 1:35 pm ■ EO: JMDI-D-20-00277

the manufacturer's instructions. The Ion AmpliSeq SARS-CoV-2 Research Panel (ThermoFisher Scientific) was used to manually prepare the libraries. The panel consists of two 5× primer pair pools that target 237 amplicons specific to the SARS-CoV-2 coronavirus and 5 human expression controls. The amplicon lengths range from 125 to 275 bp and are designed to provide >99% coverage of the SARS-CoV-2 genome, covering from position 43 to position 29,842 (positions related to reference sequence[3]). Amplification condition was 98°C for 2 minutes for initial denaturation, followed by 20, 25, or 30 cycles (Supplemental Table S1) at 98°C for 15 seconds and 60°C for 4 minutes. Then, the amplicons were digested, barcoded, and purified using AMPure XP Beads (Beckman Coulter, Brea, CA). The libraries were amplified by PCR, and size selection was performed using AMPure XP Beads. The Ion 510, Ion 520, and Ion 530 Kit, Chef and the Ion Chef (ThermoFisher Scientific), were used for template preparation and chip loading. Sequencing was performed using the S5 Gene Studio instrument (ThermoFisher Scientific).

SARS-CoV-2 whole-genome sequencing using Oxford Nanopore technology was performed as previously described.[19]

Data Analysis

The raw sequencing data were analyzed using the torrent suite software version 5.12 (ThermoFisher Scientific). The sequencing metric analysis was performed using the coverage analysis plug-in. For fresh samples, the manufacturer (ThermoFisher Scientific) recommends obtaining 1 mol/L reads per sample and reports that the uniformity is >85%. The following sequencing quality classification was used: optimal if the mapped reads were >1,000,000 and uniformity >90%; suboptimal if the mapped reads were between 1,000,000 and 500,000 and/or uniformity between 80% and 90%. If the mapped reads were <500,000 and/or uniformity between 55% and 80%, the sequencing quality was considered as poor. If the uniformity was <55%, the sequencing was considered as noncontributory.

The sequencing fragments were assembled using Iterative Refinement Meta-Assembler.[20] Alignment to the SARS-CoV-2 genome reference and variant detection were performed using the Variant Caller plug-in COVID19AnnotateSnpEff version 1.0.0.1 (ThermoFisher Scientific). The variants were defined as sequence variations from the reference sequence of the severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 NC_045512.2. Each variant with an allelic frequency (AF) >90% and recurrent variants (Supplemental Table S2) reported in the literature[21–24] were verified in the Integrative Genome Viewer (IGV) from the Broad Institute (*http://www.broadinstitute.org/igv*, last accessed November 9, 2020).[25] Sequences were aligned using the MUSCLE algorithm.[26] Clades were allowed according to GISAID definitions (ie, clade G for patients with *C241T*, *C3037T*, and *A23403G* variants; clade GR for patients with *C241T*, *C3037T*, *A23403G*, and *GGG28881AAAC* variants; and clade GH for patients with

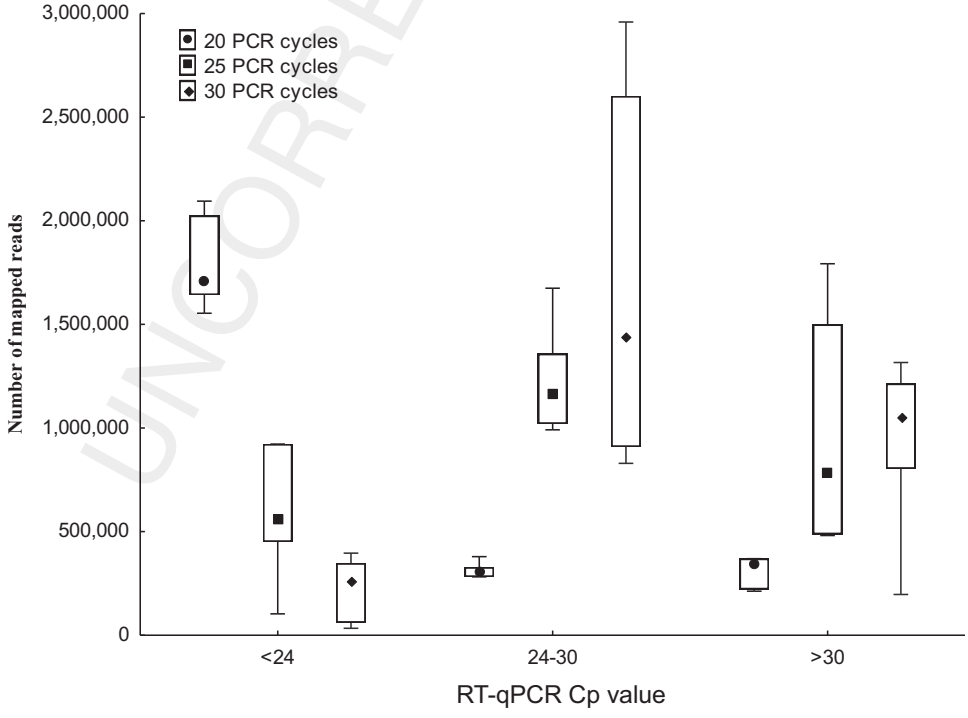

Figure 1 Variation of mapped read numbers according to RT-qPCR crossing point (Cp) values and number of PCR cycles. Data are displayed as medians (circle, 20 PCR cycles; square, 25 PCR cycles; diamond, 30 PCR cycles), 25% to 75% quartiles (box plots), and nonoutliers (bars).

SCO 5.6.0 DTD ■ JMDI1089_proof ■ 17 July 2021 ■ 1:35 pm ■ EO: JMDI-D-20-00277

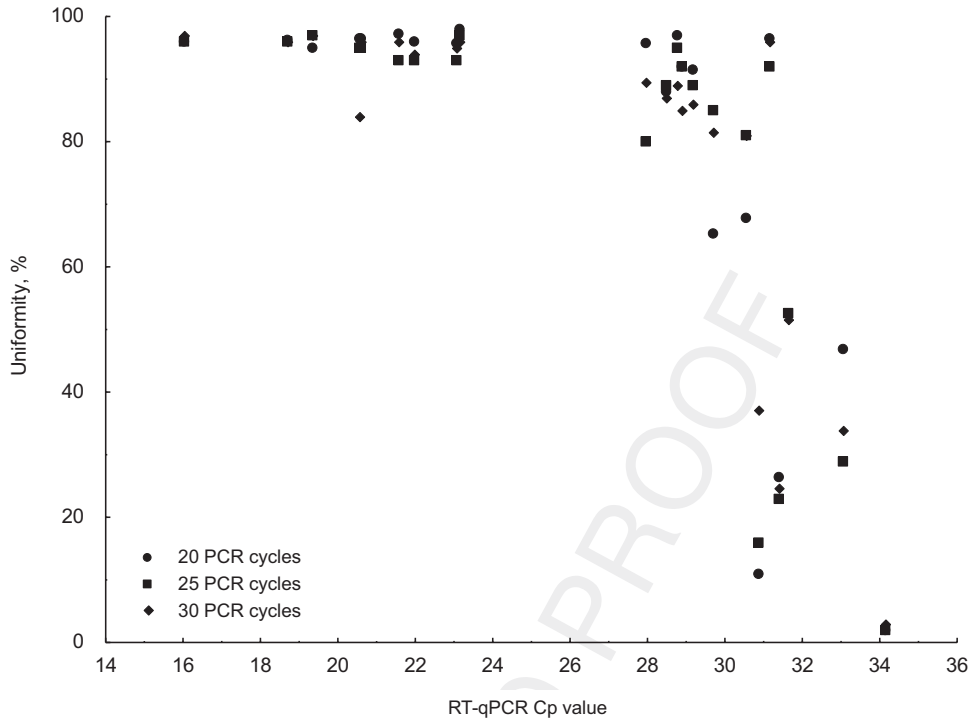

Figure 2 Dot plot of SARS-CoV-2 genome uniformity against RT-qPCR crossing point (Cp) values. For the same block, three different library preparation protocols by varying the number of target amplification PCR cycles were tested. The different conditions are indicated as follows: circle, 20 PCR cycles; square, 25 PCR cycles; diamond, 30 PCR cycles. Spearman correlation between RT-qPCR Cp value and uniformity is −0.63 ($P = 0.002$) at 20 cycles, −0.86 ($P < 10^{-6}$) at 25 cycles, and −0.81 ($P = 0.000006$) at 30 cycles.

C241T, *C3037T*, *A23403G*, and *G25563T* variants). The occurrence of variants was checked on the GISAID (using CoVsurver) and Nextstrain websites to detect new variants. Viral sequences from eight patients with <25 variants in the variant list were deposited in GISAID (*https://www.gisaid.org*, last accessed April 4, 2021).[4] For all contributive sequences, clades were attributed using Nextstrain (*https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses*, last accessed April 4, 2021; *https://github.com/nextstrain/ncov/blob/master/defaults/clades.tsv*, last accessed April 4, 2021)[27,28] and Pangolin[29] classification tools and Pangolin COVID-19 classification according to Rambaut et al[29] (Pangolin, *https://pangolin.cog-uk.io*, last accessed April 4, 2021).

Statistical Analysis

To select the optimal library preparation protocol, uniformities, numbers of mapped reads, and coverages were analyzed for each block and considered as independent. For evaluation of the sequencing performance for the selected PCR condition, the number of variants (total and with an AF >90%) was also analyzed for each block and considered as independent. The *U*-test was applied for the comparison of two independent groups of ranked data. The Friedman test was applied for the comparison of multiple dependent groups. Spearman correlation analysis was used to analyze the relationship between the RT-qPCR Cp values and uniformities. Statistical analyses were performed using Statistica 7.1 (Statsoft, Tulsa, OK).

Results

Sequencing Protocol Optimization

This study included 16 confirmed COVID-19 deceased patients with a positive SARS-CoV-2 *E* gene RT-qPCR on lung FFPE blocks. For six patients, two different lung lobes were tested, leading to 22 FFPE blocks. RT-qPCR Cp values for the different FFPE blocks ranged from 16.02 to 34.16 (Supplemental Table S1). For SARS-CoV-2 genome sequencing, the Ion AmpliSeq SARS-CoV-2 Research Panel was used, which is an amplicon-based library preparation method. Because the 22 FFPE blocks were relatively heterogeneous in terms of RT-qPCR Cp values, three different numbers of target amplification cycles were tested: 20, 25, and 30 PCR cycles for all the blocks. Libraries suitable for sequencing were obtained for all blocks, except for one (block 2-2) for which the library concentration at 20 PCR cycles was too low for sequencing (Supplemental Table S1).

Globally, no significant differences were observed in terms of sequencing metrics (number of mapped reads and coverage) with increased numbers of PCR cycles. Only the uniformity appears higher at 20 PCR cycles (median,

SCO 5.6.0 DTD ■ JMDI1089_proof ■ 17 July 2021 ■ 1:35 pm ■ EO: JMDI-D-20-00277

Table 1 Sequencing Metrics for Matched FFPE and Frozen Tissues

Patient no.	RT-qPCR Cp value	Selected PCR cycles for target amplification	Sequencing quality for FFPE tissue	Sequencing quality for frozen tissue	Variants for FFPE tissue, *n*	Variants for frozen tissue, *n*	Variants with AF >90% for FFPE tissue, *n*	Variants with AF >90% for frozen tissue, *n*
1	18.69	20	Optimal	Optimal	6	7	5	5
2-1	28.76	25	Optimal	Optimal	865	238	5	5
2-2	31.62	/	NC	Optimal	/	189	/	5
3	23.13	20	Optimal	Optimal	12	11	9	9
4	19.32	20	Optimal	Optimal	13	10	6	6
5-1	29.16	25	Suboptimal	Suboptimal	774	184	6	5
5-2	31.41	/	NC	Suboptimal	/	231	/	5
6	34.16	/	NC	NC	/	/	/	/
7	31.14	20	Optimal	Optimal	19	15	9	9
8-1	16.02	20	Suboptimal	Poor	8	10	7	7
8-2	21.57	20	Optimal	Suboptimal	9	7	7	7
9-1	27.96	20	Poor	Suboptimal	896	284	4	4
9-2	30.55	30	Suboptimal	Optimal	340	160	22	4
10	33.03	/	NC	Suboptimal	/	69	/	9
11-1	21.98	20	Optimal	Optimal	18	15	10	10
11-2	23.05	20	Optimal	Optimal	26	15	10	10
12-1	28.46	25	Suboptimal	Suboptimal	1025	293	6	7
12-2	29.69	25	Suboptimal	Optimal	589	210	9	7
13	20.59	20	Optimal	Suboptimal	15	9	7	7
14	30.88	/	NC	Poor	/	162	/	7
15	20.56	20	Optimal	Optimal	18	21	6	6
16	28.87	25	Suboptimal	Poor	707	314	8	5

AF, allelic frequency; Cp, crossing point; FFPE, formalin fixed, paraffin embedded; NC, noncontributory. Q16

95.72%) than at 25 and 30 PCR cycles (medians, 92% and 88%, respectively; Friedman test: $P = 0.04$) (Supplemental Table S1). Next, the analyses were refined according to the [F1] RT-qPCR Cp values (Figure 1). For blocks with low RT-qPCR Cp values (<24), the average number of mapped reads is higher with 20 PCR cycles (Friedman test: $P = 0.002$). In contrast, for blocks with an RT-qPCR Cp value between 24 and 30, the average number of mapped reads is higher with 25 or 30 cycles of PCR (Friedman test: $P = 0.009$). For RT-qPCR Cp values >30, a similar but slighter variation appeared in the number of mapped reads but was not significant (Friedman test: $P = 0.135$). Uniformity clearly decreased with the increase of the RT-qPCR Cp value for the three tested conditions (20, 25, and 30 cycles), as confirmed by the negative Spearman correlations [F2] (Figure 2). In particular, for the seven blocks with an RT-qPCR Cp value >30, five showed a uniformity of <55% for all the tested conditions. These five blocks were considered as noncontributory; 17 blocks were thus considered as contributory. These 17 contributory blocks were coming from 13 patients (including four patients with two blocks tested).

For the 17 contributory blocks, the aim was to establish the best PCR condition for sequencing performance and variant analyses. As sequencing quality criteria, the uniformity was selected as the most important factor because it is related to the homogeneity of the coverage distribution. The PCR condition with the highest uniformity was thus selected. If there were conditions with similar uniformities (±3%), the condition with the highest number of mapped reads was selected. If there were conditions with similar uniformities (±3%) and number of mapped reads (±20%), the condition with the fewest PCR cycles was preferred (Supplemental Table S1). This allowed us to select 20 PCR cycles for blocks with an RT-qPCR Cp value <28 and 25 PCR cycles for blocks with an RT-qPCR Cp value between 28 and 30. It was not possible to establish rules for blocks with an RT-qPCR Cp value >30, with most of them being noncontributory (Supplemental Table S1).

Sequencing Performances Obtained after Optimization

After adapting the number of target amplification PCR cycles according to the RT-qPCR Cp values for the 17 contributory blocks, the median number of mapped reads and uniformity were 1,642,150 (minimum-maximum: 305,249 to 2,094,563) and 95.9% (minimum-maximum: 81% to 98%), respectively.

The sequencing quality was considered as optimal for 10 blocks (*Materials and Methods*), with a median number of mapped reads of 1,748,009, a median coverage of 10,644, and a median uniformity of 96.4% (Table 1). The RT-qPCR [T1] Cp value of these 10 FFPE blocks varied from 18.69 to 31.14. The sequencing quality was considered as suboptimal

SCO 5.6.0 DTD ■ JMDI1089_proof ■ 17 July 2021 ■ 1:35 pm ■ EO: JMDI-D-20-00277

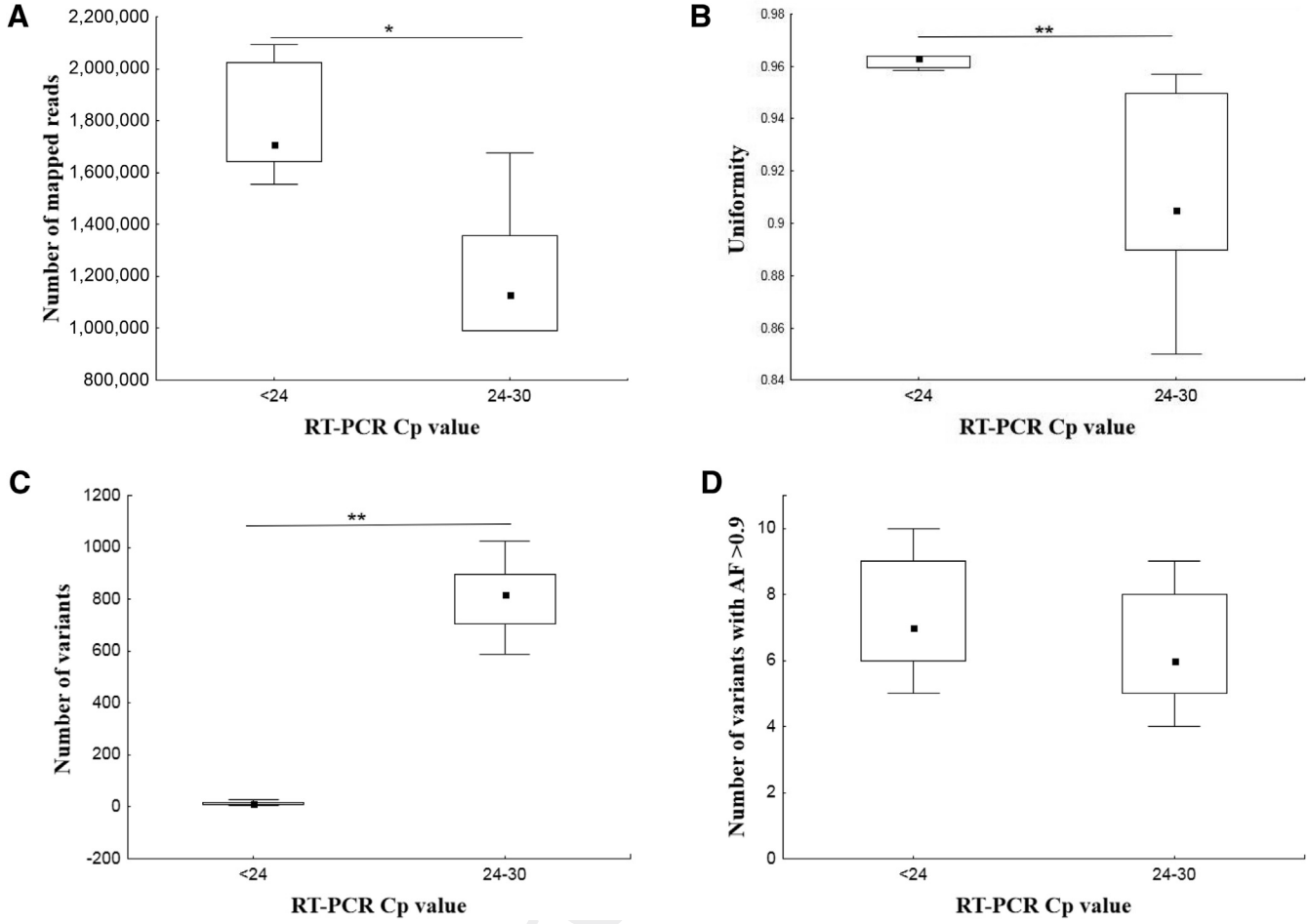

Figure 3 Variation of sequencing performances [number of mapped reads (**A**), uniformity (**B**), number of variants (**C**), and number of variants with an allelic frequency (AF) >0.9 (**D**)] obtained with the selected PCR condition, according to RT-qPCR crossing point (Cp) values for contributory blocks. Data are displayed as medians, 25% to 75% quartiles (box plots), and nonoutliers (bars). The *U*-test was applied. $*P < 0.05$, $**P < 0.01$. Q15

for six blocks, with a median number of mapped reads of 1,128,420, a median coverage of 5385, and a median uniformity of 89%. The RT-qPCR Cp values ranged from 16.02 to 30.55. The sequencing quality of one block was considered as poor (Table 1). According to the RT-qPCR Cp values, significant differences were observed between contributory blocks with an RT-qPCR Cp value <24 and those with an RT-qPCR Cp value between 24 and 30 in terms of number of mapped reads, uniformity, and the number of variants (Figure 3). As it would be easier in daily practice to use the same protocol for each block, a comparison between the sequencing metrics and the data obtained by the Variant Caller plug-in was performed for each block for the three different conditions (20, 25, and 30 cycles) and the selected condition, as proposed above. The adaptation of the number of PCR cycles to the RT-qPCR Cp value (selected condition) allowed obtaining more blocks with an optimal or suboptimal result (Supplemental Table S3). Moreover, increasing the number of PCR cycles lead to a higher number of variants with an AF <0.9, which can reflect sequencing artifacts.

Factors Influencing Sequencing Performances

The presence of hemorrhage and/or necrosis on the 22 FFPE blocks was evaluated to identify if histologic features can affect the sequencing performance and quality (Supplemental Table S4). Hemorrhage was observed for eight blocks, and necrosis was observed for four blocks. The sequencing quality was more often optimal when neither hemorrhage nor necrosis was present (7/11 blocks with optimal sequencing when neither hemorrhage nor lysis was present versus 3/11 blocks with optimal sequencing when hemorrhage and/or lysis was present).

To examine the impact of formalin fixation (a well-known cause of RNA damage-induced changes and sequencing artifacts), the same library preparation (with the adaptation of the PCR amplification cycles to the RT-qPCR Cp values) and sequencing protocols were used on matched frozen tissues. Using this method on the 22 frozen tissues, sequencing quality was considered as optimal for 11 (median number of mapped reads of 1,549,686, median coverage of 9971, and median uniformity of 97%).

SCO 5.6.0 DTD ■ JMDI1089_proof ■ 17 July 2021 ■ 1:35 pm ■ EO: JMDI-D-20-00277

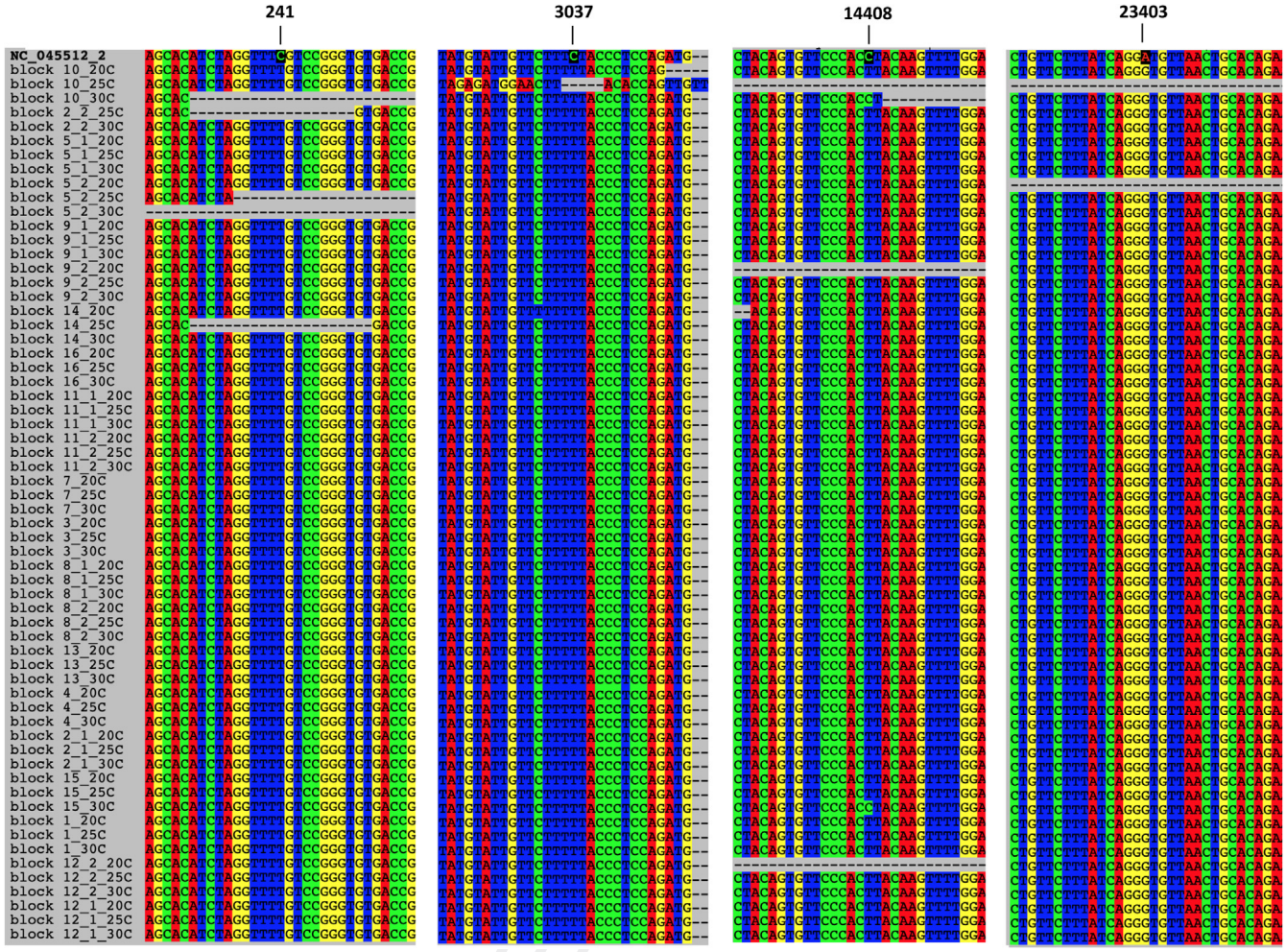

Figure 4 Partial sequence alignments of 21 formalin-fixed, paraffin-embedded blocks with three different numbers of target amplification cycles against the reference sequence NC_0455512.2. Key residue nucleotides for Global Initiative on Sharing All Influenza Data clade classification are indicated. Sequences for block 6 are not included in the alignment as they are much shorter than the others and do not align sufficiently well to the other sequences to give useful information.

Sequencing quality was considered as suboptimal for seven frozen tissues (median number of mapped reads of 1,256,654, median coverage of 5375, and median uniformity of 90%). Finally, for three frozen tissues, the sequencing qualities were considered as poor; and for one frozen tissue, they were considered as noncontributory (Table 1). When considering the six suboptimal FFPE blocks and the matched frozen tissues, optimal quality on frozen tissues was observed for two of them, whereas sequencing remained suboptimal for two and poor for the remaining two. The 9-1 poor sequencing quality from FFPE was suboptimal from frozen tissue. The five FPPE blocks categorized as noncontributory showed various results when frozen tissue was sequenced: one optimal, two suboptimal, one poor, and one noncontributory.

Variant Analysis

Among the 17 contributory FFPE blocks, between 6 and 1025 variants were detected, with an AF varying between 2% and 100%. Between 4 and 22 variants with an AF >90% were detected, with a mean of 8 variants per FFPE block. Each variant with an AF >90% and recurrent variants reported in the literature were verified in the IGV (*Materials and Methods*). Verification using IGV validated all variants with an AF >90%, except for two deletions that were detected by the Variant Caller plug-in but not confirmed (Patients 5 and 12). Moreover, for Patient 2, the variant *G11083T* was detected by the Variant Caller plug-in with an AF of 63% in the FFPE block and with an AF of 73% in the matched frozen tissue, but IGV verification revealed an AF of almost 100% for the two conditions. For Patient 12, the variant *GGG28881AAC* was detected by the Variant Caller plug-in with an AF of 74% and with an AF of 95% in the matched frozen tissue; verification using IGV revealed an AF of almost 100% for the FFPE block. For Patient 16, IGV verification showed the presence of the variants *C241T* and *GGG28881AAC* in both FFPE block and matched frozen tissue. However, the *C27476T* variant identified in the FFPE block with an AF of 95% was not observed in the matched frozen tissue.

SCO 5.6.0 DTD ■ JMDI1089_proof ■ 17 July 2021 ■ 1:35 pm ■ EO: JMDI-D-20-00277

Table 2 Variant Frequencies

Nucleotide variation	Gene	Mutation type	Amino acid change	Frequency
C241T	Upstream (5′UTR)			13/13
	ORF1ab			
C710T	*ORF1ab/NSP1*	Missense	L149F	1/13
C1059T	*ORF1ab/NSP2*	Missense	T265I/T85I	1/12
C2113T	*ORF1ab*	Synonymous	-	1/13
C3037T	*ORF1ab*	Synonymous	-	13/13
C4002T	*ORF1ab/NSP3*	Missense	T1246I/T428I	3/13
C7765T	*ORF1ab*	Synonymous	-	1/13
C8782T	*ORF1ab*	Synonymous	-	0/13
G10097A	*ORF1ab/NSP5*	Missense	G3278S/G15S	3/13
G11083T	*ORF1ab/NSP6*	Missense	L3606F/L37F	1/13
C13536T	*ORF1ab*	Synonymous	-	2/12
C14408T	*ORF1ab/NSP12*	Missense	P4715L/P323L	13/13
C15324T	*ORF1ab*	Synonymous	-	6/12
T15978C	*ORF1ab*	Synonymous	-	1/13
A16166G	*ORF1ab/NSP12*	Missense	N5301S/N909S	1/13
C17690T	*ORF1ab/NSP13*	Missense	S5809L/S485L	1/13
C18060T	*ORF1ab*	Synonymous	-	0/13
C18877T	*ORF1ab*	Synonymous	-	1/13
C21805T	*S*	Synonymous	-	1/13
A23403G	*S*	Missense	D614G	13/13
C23731T	*S*	Synonymous	-	2/12
G24794T	*S*	Missense	A1078S	1/13
G25563T	*ORF3a*	Missense	Q57H	2/12
G26144T	*ORF3a*	Missense	G251V	0/13
T28144C	*ORF8*	Missense	L84S	0/13
G28690T	*N*	Missense	L139F	1/13
A28765G	*N*	Synonymous	-	1/13
GGG28881AAC	*N*	Missense	RG203KR	4/13
G29291A	*N*	Missense	D340N	1/12

UTR, untranslated region.

For all optimal (10 of 10) FFPE blocks, the same variants with an AF >90% were detected in the matched frozen tissues. For the six suboptimal FFPE blocks, comparison of the variant caller plug-in results between FFPE and matched frozen tissue revealed additional variants for four FFPE blocks (5-1, 9-2, 12-2, and 16), a missing variant for one FFPE block (12-1), and the same profile for one FFPE block (8-1) (Table 1). However, IGV verification showed that the profile was concordant between FFPE and matched frozen tissue for blocks 5-1 and 12-1. In summary, the comparison of matched frozen and FFPE tissues identified three blocks presenting discordance (9-2, 12-2, and 16), with additional variants in the FFPE blocks that were absent in the matched frozen tissue. All of the three FFPE blocks were characterized by a suboptimal sequencing and by an RT-qPCR Cp value >28.

Regarding the four patients with two different lung lobes tested, two presented the same variant profile (Patients 8 and 11). Discordances were observed between lobes for Patients 9 and 12, but comparison with matched frozen tissues revealed that additional variants observed in one lobe were related to sequencing artifacts.

Regarding recurrent variants reported in the literature, all patients harbored the *C241T*, *C3037T*, *C14408T*, and *A23403G* nucleotide variants (Figure 4 and Supplemental Appendix S1). Distinct variant profiles have been identified across the patients (Tables 2 and 3). According to the GISAID definitions (*Materials and Methods*), clade G was assigned for seven patients, clade GR was assigned for four patients, and clade GH was assigned for two patients. For four patients (Patients 9, 11, 12, and 16), some genomic positions cannot be assessed because of an AF of around 40% to 60%. Using Nextstrain classification, eight patients were classified as clade 20A (because of the *C14408T* and *A23403G* variants), one patient was classified as clade 20B (because of the *GGG28881AAC* variant), one patient was classified as clade 20C (because of the *C1059T* and *G25563T* variants), and three patients were classified as clade 20D (because of the *C4002T*, *G10097A*, *C13536T*, and *C23731T* variants) (Figure 5). According to Pangolin COVID-19 classification from Rambaut et al,[29] 11 patients were classified as B.1 and two patients were classified as C.11 (alias of B.1.1.1.11). Variants were checked on the GISAID (using CoVsurver) and Nextstrain websites, and

SCO 5.6.0 DTD ■ JMDI1089_proof ■ 17 July 2021 ■ 1:35 pm ■ EO: JMDI-D-20-00277

Table 3 Variant Profile per Patient

Patient no.	Profile	GISAID clade	GISAID ID*	Nextstrain clade	Pangolin COVID-19 classification[29]
1	*C241T-C3037T-C14408T-C15324T-A23403G*	G	SARS-CoV-2/human/Brussels/1/2020_EPI_ISL_451935	20A	B.1
2	*C241T-C3037T-G11083T-C14408T-C15324T-A23403G*	G	-	20A	B.1
3	*C241T-C2113T-C3037T-C7765T-C14408T-C17690T-C18877T-A23403G-G25563T*	GH	SARS-CoV-2/human/Brussels/3/2020_EPI_ISL_452142	20A	B.1.9
4	*C241T-C3037T-C14408T-C15324T-A23403G-A28765G*	G	SARS-CoV-2/human/Brussels/4/2020_EPI_ISL_452148	20A	B.1.83
5	*C241T-C3037T-C14408T-C15324T-A23403G*	G	-	20A	B.1
7	*C241T-C3037T-C4002T-G10097A-C13536T-C14408T-A23403G-C23731T-GGG28881AAC*	GR	SARS-CoV-2/human/Brussels/7/2020_EPI_ISL_452140	20D	C11
8	*C241T-C1059T-C3037T-C14408T-A23403G-G25563T-G29291A*	GH	SARS-CoV-2/human/Brussels/8/2020_EPI_ISL_452149	20C	B.1.321
9	*C241T-C3037T-C14408T-A23403G*	G	-	20A	B.1.6
11	*C241T-C3037T-C4002T-G10097A-C13536T-C14408T-A16166G-A23403G-C23731T-GGG28881AAC*	GR	SARS-CoV-2/human/Brussels/11/2020_EPI_ISL_452150	20D	B.1.1.1
12	*C241T-C3037T-C14408T-T15978C-A23403G-G24794T-GGG28881AAC*	GR	-	20B	B.1.1
13	*C241T-C710T-C3037T-C14408T-C15324T-A23403G-G28690T*	G	SARS-CoV-2/human/Brussels/13/2020_EPI_ISL_452151	20A	B.1
15	*C241T-C3037T-C14408T-C15324T-C21805T-A23403G*	G	SARS-CoV-2/human/Brussels/15/2020_EPI_ISL_452152	20A	B.1
16	*C241T-C3037T-C4002T-G10097A-C14408T-A23403G-GGG28881AAC*	GR	-	20D	C.11

*GISAID (*https://www.gisaid.org*, last accessed May 7, 2021).
COVID-19, coronavirus disease 2019; GISAID, Global Initiative on Sharing All Influenza Data; ID, identifier. Q18

three variants that have never been described before were detected [ie, *A16166G* (AF of 99.8% for both tested lobes) for Patient 11, *C710T* (AF of 100%) for Patient 13, and *C21805T* (AF of 99.7%) for Patient 15]. Interestingly, those three variants were also detected in the matched frozen tissues.

To confirm the variants identified using the Ion Torrent sequencing platform, the SARS-CoV-2 genome from the frozen tissues matching the 17 contributory FFPE blocks was also sequenced using Oxford Nanopore technology. Sequences were obtained for all tissues, except one (block 2). Variants reported in Supplemental Table S2 could be confirmed using this third-generation sequencing platform.

Robustness Analysis

To evaluate the robustness of the SARS-CoV-2 genotyping on FFPE blocks, the technique was challenged by lowering the amount of RNA used in the reverse transcription reaction. Instead of 10 ng, 5 or 1 ng was used as input to prepare the libraries for five different FFPE blocks: two blocks with an RT-qPCR Cp value <24 (1 and 4), two blocks with an RT-qPCR Cp value between 24 and 30 (2-1 and 16), and one block with an RT-qPCR Cp value >30 (7). For three of the five blocks, genotyping results (variants with an AF >90%) remained identical regardless of the amount of input RNA. For two blocks (both with an RT-qPCR Cp value >24), the decrease of viral input was associated with discordant results in the number of identified variants (data not shown).

Discussion

Currently, many questions remain about the origin, evolution, and spreading of the SARS-CoV-2. The SARS in

SCO 5.6.0 DTD ■ JMDI1089_proof ■ 17 July 2021 ■ 1:35 pm ■ EO: JMDI-D-20-00277

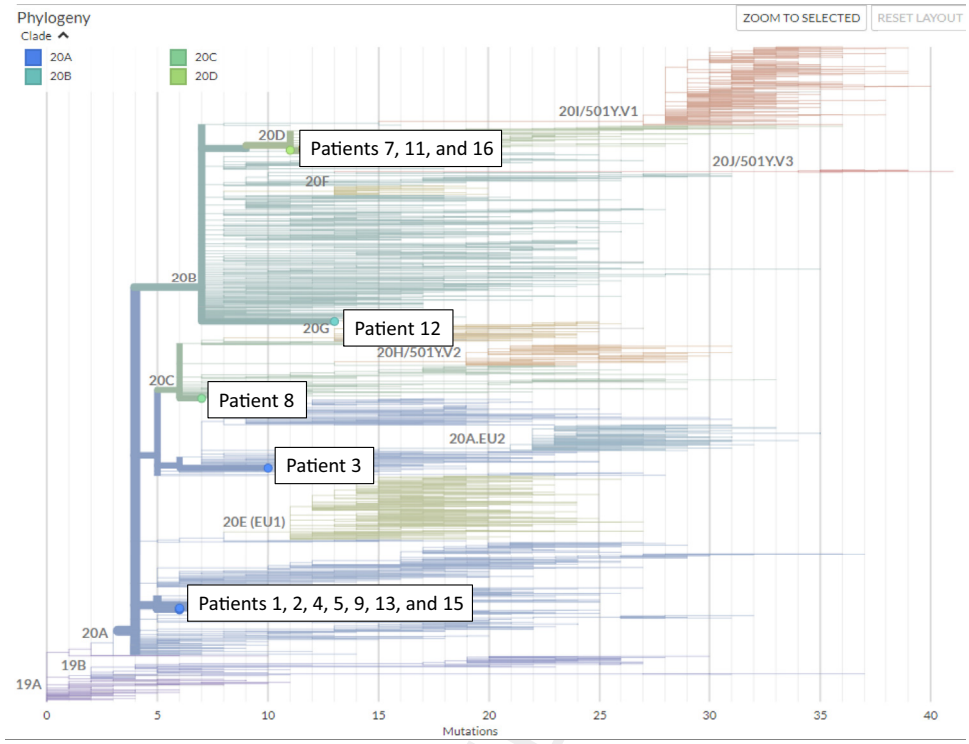

Figure 5 Nextstrain classification for 13 patients.

2003,[30] Middle East respiratory syndrome in 2014,[31] and the current COVID-19 pandemic highlight the need for coronavirus genome characterization. Laboratories worldwide are using their sequencing infrastructure and expertise to deliver and characterize SARS-CoV-2 genome sequences. Most of these sequences are generated from fresh samples; therefore, library preparation and sequencing protocols are not adapted to FFPE blocks.

The present study aimed to optimize SARS-CoV-2 genotyping on post-mortem FFPE lung tissues using the Ion AmpliSeq SARS-CoV-2 Research Panel. According to the manufacturer, the number of target amplification cycles should be adapted to the viral load. Even if the RT-qPCR Cp value can be affected by batch effect[32] and cannot be used as a precise quantitative measure of viral load, the RT-qPCR Cp value can indirectly reflect the viral load. Because the RT-qPCR Cp values were heterogeneous across the FFPE blocks, different numbers of target amplification cycles were tested to optimize the sequencing. The different numbers of amplification cycles were selected to avoid overamplification of smaller fragments, leading to lower uniformity. Low template input and biased amplification of biological material by PCR are also a source of distortion and can potentially affect the accuracy of variant detection.[33–35] The present data highlight the importance of the RT-qPCR Cp value in the sequencing optimization. Indeed, an increase of the number of target amplification PCR cycles is required for blocks with a higher RT-qPCR Cp value (>28). Nevertheless, most FFPE blocks with an RT-qPCR Cp value >30 were noncontributory, even if the number target amplification PCR cycles was increased.

SARS-CoV-2 sequences available from databases (National Center for Biotechnology Information and GISAID) are generated with different sequencing platforms and methods, and quality criteria are not well defined. In the present study, sequencing quality was categorized as optimal, suboptimal, or poor based on the number of mapped reads and the uniformity. Using this classification, all the variants identified in the optimal FFPE blocks were confirmed on matched frozen tissue. Discordances were observed only for blocks with suboptimal or poor sequencing. These data suggest that the proposed sequencing quality evaluation allows the identification of FFPE blocks with reliable results when the sequencing quality is optimal. If the sequencing quality is suboptimal or poor, new variants should be analyzed with caution, especially if a high number of variants was identified.

The present study aimed also to identify factors that can impact the sequencing quality. SARS-CoV-2 genotyping results are influenced by several factors, such as the presence of hemorrhage and/or necrosis in the tissues, RT-qPCR Cp values, and formalin fixation. The sequencing quality was more often optimal when neither hemorrhage nor necrosis was present. Among the five noncontributory blocks,

SCO 5.6.0 DTD ■ JMDI1089_proof ■ 17 July 2021 ■ 1:35 pm ■ EO: JMDI-D-20-00277

three presented hemorrhage and/or necrosis. Regarding RT-qPCR Cp values, after optimization, a significant difference in terms of sequencing metrics was still observed between FFPE blocks with low or high RT-qPCR Cp values. Because formalin fixation leads to cross-linking and fragmentation of nucleic acids, extractions from FFPE are typically fragmented into pieces <300 bp long.[16] The Ion AmpliSeq SARS-CoV-2 Research Panel amplicon lengths range from 125 to 275 bp, with an average length of 202 bp. This relatively short amplicon length can explain the success of sequencing on FFPE blocks. Indeed, the comparison of matched frozen and FFPE tissues revealed discordance only for three blocks, all with a suboptimal sequencing result and with an RT-qPCR Cp value >28. Moreover, three variants never described before have been detected using FFPE blocks and confirmed on the matched frozen tissue. These data confirmed that FFPE material is suitable for SARS-CoV-2 genotyping.

The present study has some limitations: i) The sample size was relatively small. ii) The autopsies were performed from 72 to 96 hours after death. This delay can alter the quality of the nucleic acids. iii) Total nucleic acids were used as starting input. The impact of viral enrichment strategies should be investigated. iv) RT-qPCR Cp value was used to determine the number of PCR cycles. However, RT-qPCR Cp values can vary and should be validated in each laboratory. v) The amount of available material was relatively large as it was autopsy tissue. However, in the daily practice of the pathology laboratories, molecular testing should be adapted to small biopsies and low quantities of nucleic acids. To investigate the robustness of the test, the amount of starting RNA was decreased, with concordant results for FFPE blocks with low RT-qPCR Cp values. These data have to be confirmed in a larger study using biopsies. vi) This study is limited to lung tissues, and other organs were not investigated. vii) No comparison was possible with the premortem sample.

Several publications have shown that third-generation sequencing methods (Oxford Nanopore sequencing and PacBio Sequel) can be used to genotype viral pathogens, such as SARS-CoV-2.[19,36–38] Direct RNA sequencing using nanopores allows virus identification without the amplification biases linked to other sequencing technologies. The third-generation sequencing also offers near to real-time genome sequencing and consequently short turnaround time (hours compared with days with Ion Torrent and Illumina). In the context of a new emerging infectious disease, these methods provide a powerful tool to rapidly identify pathogens. Nevertheless, third-generation NGS platforms are less compatible with FFPE than second-generation platforms.[39] For this reason, the most commonly used NGS platforms in pathology laboratories still belong to the second generation.[40]

The data obtained in the present study allowed us to classify SARS-CoV-2 genomes using the clade nomenclature from GISAID for all contributory sequences as well as Nextstrain and Pangolin COVID-19 classification tools.[27–29] The variant profile was used only for the purpose of classification, as the functional and clinical impacts of these mutations remain unknown. According to the clade classification, most of the 13 patients (8/13) are classified as clade 20A, 1 is classified as clade 20B, 1 is classified as clade 20C, and 3 are classified as clade 20D (Nextstrain, *https://nextstrain.org/blog/2021-01-06-updated-sars-cov-2-clade-naming*, last accessed April 23, 2021).

In conclusion, the present study proposes to adapt the number of target amplification PCR cycles according to RT-qPCR Cp value to optimize and to obtain reliable SARS-CoV-2 genome sequencing on FFPE samples. This opens the possibility to explore correlation between virus genotype and histopathologic lesions.

Acknowledgments

We thank Nathalie Lijsen, Christophe Valleys, Dominique Penninck, and Nicole Haye for technical and logistic supports; Andrea Schiavo for critical reading of the manuscript; and Dr. Marie-Paule Van Craynest for trainees' supervision.

Supplemental Data

Supplemental material for this article can be found at *http://doi.org/10.1016/j.jmoldx.2021.05.016*.

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020, 395: 497–506
2. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS1, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL: A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020, 579:270–273
3. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ: A new coronavirus associated with human respiratory disease in China. Nature 2020, 579: 265–269
4. Elbe S, Buckland-Merrett G: Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall 2017, 1:33–46
5. Remmelink M, De Mendonça R, D'Haene N, De Clercq S, Verocq C, Lebrun L, Lavis P, Racu ML, Trépant AL, Maris C, Rorive S, Goffard JC, Dewitte O, Peluso L, Vincent JL, Decaestecker C, Taccone FS, Salmon I: Unspecific post-mortem findings despite multiorgan viral spread in COVID-19 patients. Crit Care 2020, 24: 495
6. Tian S, Xiong Y, Liu H, Niu L, Guo J, Liao M, Xiao SY: Pathological study of the 2019 novel coronavirus disease (COVID-19) through postmortem core biopsies. Mod Pathol 2020, 33:1007–1014
7. Sekulic M, Harper H, Nezami BG, Shen DL, Sekulic SP, Koeth AT, Harding CV, Gilmore H, Sadri N: Molecular detection

SCO 5.6.0 DTD ■ JMDI1089_proof ■ 17 July 2021 ■ 1:35 pm ■ EO: JMDI-D-20-00277

of SARS-CoV-2 infection in FFPE samples and histopathologic findings in fatal SARS-CoV-2 cases. Am J Clin Pathol 2020, 154: 190–200

8. Wichmann D, Sperhake JP, Lütgehetmann M, Steurer S, Edler C, Heinemann A, Heinrich F, Mushumba H, Kniep I, Schröder AS, Burdelski C, de Heer G, Nierhaus A, Frings D, Pfefferle S, Becker H, Bredereke-Wiedling H, de Weerth A, Paschen HR, Sheikhzadeh-Eggers S, Stang A, Schmiedel S, Bokemeyer C, Addo MM, Aepfelbacher M, Püschel K, Kluge S: Autopsy findings and venous thromboembolism in patients with COVID-19. Ann Intern Med 2020, 173:268–277
9. Menter T, Haslbauer JD, Nienhold R, Savic S, Hopfer H, Deigendesch N, Frank S, Turek D, Willi N, Pargger H, Bassetti S, Leuppi JD, Cathomas G, Tolnay M, Mertz KD, Tzankov A: Post-mortem examination of COVID19 patients reveals diffuse alveolar damage with severe capillary congestion and variegated findings of lungs and other organs suggesting vascular dysfunction. Histopathology 2020, 77:198–209
10. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W: Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020, 395:565–574
11. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon RW, Tsoi HW, Lo SK, Chan KH, Poon VK, Chan WM, Ip JD, Cai JP, Cheng VC, Chen H, Hui CK, Yuen KY: A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020, 395:514–523
12. Licastro D, Rajasekharan S, Dal Monego S, Segat L, D'Agaro P, Marcello A: Isolation and full-length genome characterization of SARS-CoV-2 from COVID-19 cases in Northern Italy. J Virol 2020, 94:e00543-20
13. D'Haene N, Le Mercier M, De Nève N, Blanchard O, Delaunoy M, El Housni H, Dessars B, Heimann P, Remmelink M, Demetter P, Tejpar S, Salmon I: Clinical validation of targeted next generation sequencing for colon and lung cancers. PLoS One 2015, 10: e0138245
14. D'Haene N, Meléndez B, Blanchard O, De Nève N, Lebrun L, Van Campenhout C, Salmon I: Design and validation of a gene-targeted, next-generation sequencing panel for routine diagnosis in gliomas. Cancers (Basel) 2019, 11:773
15. Bodewes R, van Run PR, Schürch AC, Koopmans MP, Osterhaus AD, Baumgärtner W, Kuiken T, Smits SL: Virus characterization and discovery in formalin-fixed paraffin-embedded tissues. J Virol Methods 2015, 214:54–59
16. Duncavage EJ, Magrini V, Becker N, Armstrong JR, Demeter RT, Wylie T, Abel HJ, Pfeifer JD: Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. J Mol Diagn 2011, 13:325–333
17. Xiao YL, Kash JC, Beres SB, Sheng ZM, Musser JM, Taubenberger JK: High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. J Pathol 2013, 229:535–545
18. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brünink S, Schneider J, Schmidt ML, Mulders DG, Haagmans BL, van der Veer B, van den Brink S, Wijsman L, Goderski G, Romette JL, Ellis J, Zambon M, Peiris M, Goossens H, Reusken C, Koopmans MP, Drosten C: Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill 2020, 25:2000045
19. Artesi M, Bontems S, Göbbels P, Franckh M, Maes P, Boreux R, Meex C, Melin P, Hayette MP, Bours V, Durkin K: A recurrent mutation at position 26340 of SARS-CoV-2 is associated with failure of the E gene quantitative reverse transcription-PCR utilized in a commercial dual-target diagnostic assay. J Clin Microbiol 2020, 58: e01598-20
20. Shepard SS, Meno S, Bahl J, Wilson MM, Barnes J, Neuhaus E: Viral deep sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler. BMC Genomics 2016, 17: 708
21. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, Masciovecchio C, Angeletti S, Ciccozzi M, Gallo RC, Zella D, Ippodrino R: Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant: version 2. J Transl Med 2020, 18:179
22. Yin C: Genotyping coronavirus SARS-CoV-2: methods and implications. Genomics 2020, 112:3588–3596
23. Stefanelli P, Faggioni G, Lo Presti A, Fiore S, Marchi A, Benedetti E, Fabiani C, Anselmo A, Ciammaruconi A, Fortunato A, De Santis R, Fillo S, Capobianchi MR, Gismondo MR, Ciervo A, Rezza G, Castrucci MR, Lista F; on Behalf of Iss Covid-Study Group: Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. Euro Surveill 2020, 25:2000305
24. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, Zhang Z: The establishment of reference sequence for SARS-CoV-2 and variation analysis. J Med Virol 2020, 92:667–674
25. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: Integrative genomics viewer. Nat Biotechnol 2011, 29:24–26
26. Edgar RC: MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 2004, 5: 113
27. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA: Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 2018, 34:4121–4123
28. Bedford T, Neher R, Hadfield J, Hodcroft E, Sibley T, Huddleston J, Lee J, Fay K, Bell S, Megill C, Potter B, Sagulenko P, Callender C, Ilcisin M, Moncla L, Black A, Brito A, Grubaugh N: Nextstrain build for novel coronavirus SARS-CoV-2. Q14
29. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus OG: A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 2020, 5:1403–1407
30. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JS, Poon LL: Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China. Science 2003, 302: 276–278
31. Drosten C, Kellam P, Memish ZA: Evidence for camel-to-human transmission of MERS coronavirus. N Engl J Med 2014, 371: 1359–1360
32. Han MS, Byun JH, Cho Y, Rim JH: RT-PCR for SARS-CoV-2: quantitative versus qualitative. Lancet Infect Dis 2021, 21:165
33. Zanini F, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, Neher RA: Population genomics of intrapatient HIV-1 evolution. Elife 2015, 4: e11282
34. Kebschull JM, Zador AM: Sources of PCR-induced distortions in high-throughput sequencing data sets. Nucleic Acids Res 2015, 43: e143
35. Dyrdak R, Mastafa M, Hodcroft EB, Neher RA, Albert J: Intra- and interpatient evolution of enterovirus D68 analyzed by whole-genome deep sequencing. Virus Evol 2019, 5:vez007
36. Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG, Mohr I, Wilson AC: Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. Nat Commun 2019, 10:754
37. Viehweger A, Krautwurst S, Lamkiewicz K, Madhugiri R, Ziebuhr J, Hölzer M, Marz M: Direct RNA nanopore sequencing of full-length

coronavirus genomes provides novel insights into structural variants and enables modification analysis. Genome Res 2019, 29:1545−1554

38. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole Á, Haverkate M, Mollers M, Kamga SK, Schapendonk C, Pronk M, Lexmond P, van der Linden A, Bestebroer T, Chestakova I, Overmars RJ, van Nieuwkoop S, Molenkamp R, van der Eijk AA, GeurtsvanKessel C, Vennema H, Meijer A, Rambaut A, van Dissel J, Sikkema RS, Timen A, Koopmans M: Dutch-Covid-19 response team. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. Nat Med 2020, 26:1405−1410

39. Euskirchen P, Bielle F, Labreche K, Kloosterman WP, Rosenberg S, Daniau M, Schmitt C, Masliah-Planchon J, Bourdeaut F, Dehais C, Marie Y, Delattre JY, Idbaih A: Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing. Acta Neuropathol 2017, 134:691−703

40. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, Temple-Smolkin RL, Voelkerding KV, Nikiforova MN: Guidelines for validation of next-generation sequencing-based oncology panels: a joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. J Mol Diagn 2017, 19:341−365

SCO 5.6.0 DTD ■ JMDI1089_proof ■ 17 July 2021 ■ 1:35 pm ■ EO: JMDI-D-20-00277