

Article

Direct Nanopore Sequencing of Human Cytomegalovirus Genomes from High-Viral-Load Clinical Samples

Kathy K. Li ^{1,2}, Betty Lau ¹, Nicolás M. Suárez ¹, Salvatore Camiolo ¹, Rory Gunson ³, Andrew J. Davison ¹  and Richard J. Orton ^{1,*} 

¹ Medical Research Council, University of Glasgow Centre for Virus Research, Glasgow G61 1QH, UK; kathy.li@glasgow.ac.uk (K.K.L.); nicolas.suarez@glasgow.ac.uk (N.M.S.); andrew.davison@glasgow.ac.uk (A.J.D.)

² Regional Virus Laboratory, Belfast Health and Social Care Trust, Belfast BT12 6BA, UK

³ West of Scotland Specialist Virology Centre, NHS Greater Glasgow & Clyde, Glasgow G31 2ER, UK; rory.gunson@ggc.scot.nhs.uk

* Correspondence: richard.orton@glasgow.ac.uk

Abstract: Nanopore sequencing is becoming increasingly commonplace in clinical settings, particularly for diagnostic assessments and outbreak investigations, due to its portability, low cost, and ability to operate in near real-time. Although high sequencing error rates initially hampered the wider implementation of this technology, improvements have been made continually with each iteration of the sequencing hardware and base-calling software. Here, we assess the feasibility of using nanopore sequencing to determine the complete genomes of human cytomegalovirus (HCMV) in high-viral-load clinical samples without viral DNA enrichment, PCR amplification, or prior knowledge of the sequences. We utilised a hybrid bioinformatic approach that involved assembling the reads de novo, improving the consensus sequence by aligning reads to the best-matching genome from a collated set of published sequences, and polishing the improved consensus sequence. The final genomes from a urine sample and a lung sample, the former with an HCMV to human DNA load approximately 50 times greater than the latter, achieved 99.97 and 99.93% identity, respectively, to the benchmark genomes obtained independently by Illumina sequencing. Thus, we demonstrated that nanopore sequencing is capable of determining HCMV genomes directly from high-viral-load clinical samples with a high accuracy.

Keywords: human cytomegalovirus; clinical sample; genome; nanopore sequencing; Illumina sequencing



Citation: Li, K.K.; Lau, B.; Suárez, N.M.; Camiolo, S.; Gunson, R.; Davison, A.J.; Orton, R.J. Direct Nanopore Sequencing of Human Cytomegalovirus Genomes from High-Viral-Load Clinical Samples. *Viruses* **2023**, *15*, 1248. <https://doi.org/10.3390/v15061248>

Academic Editor: Robert Kalejta

Received: 19 April 2023

Revised: 18 May 2023

Accepted: 22 May 2023

Published: 26 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Seropositivity to human cytomegalovirus (HCMV; species *Cytomegalovirus human-beta5*) [1] ranges from 40% in developed countries to 100% in low- and middle-income countries (LMICs) [2,3]. Infection is usually asymptomatic [4], but in the absence of an appropriate immune response, may lead to severe morbidity and mortality, as exemplified in the transplant cohort [5] and the congenital setting [6]. Antiviral drugs are available but have myelotoxic and nephrotoxic profiles and are therefore best targeted to those diagnosed as being at highest risk. They also promote the generation of antiviral resistance mutants which are monitored on the basis of DNA sequence data in order to inform clinical management [7]. The identification of HCMV genes targeted by established antivirals and the resistance mutations associated with these genes is one application of HCMV sequence data, and this will continue to be the case as novel antivirals are developed in future. In keeping pace with relevant technologies over four decades, HCMV sequence data have also been pivotal in illuminating the structure, function, and evolution of HCMV genomes. The present ability to sequence HCMV genomes in large numbers also provides a powerful means of investigating the role of viral variation in disease outcome.

HCMV has a linear, double-stranded DNA genome of approximately 236 kbp [8,9]. The overall configuration is $ab-U_L-b'a'c'-U_S-ca$, in which unique regions U_L (193 kbp) and U_S (35 kbp) are flanked by inverted repeats $ab/b'a'$ (1 kbp) and $a'c'/ca$ (3 kbp). Early work involving restriction fragment length polymorphism analysis and Sanger sequencing revealed extensive genome variation among HCMV strains, including the existence of multiple discrete genotypes of a subset of hypervariable genes [10–12], the extensive rearrangement of these genotypes due to recombination during HCMV evolution [13], and the invariable loss of substantial genome segments and the accumulation of other mutations in highly passaged strains [14–16]. These studies were limited by the existing Sanger sequencing technology and typically focused on only one or a few genes. Subsequently, high-throughput, short-read sequencing was developed with a greater accuracy and at a lower cost [17], thus enabling the sequencing of large numbers of HCMV genomes [18–22] and facilitating the genotyping of hypervariable genes directly from read data [21–23]. These advances depended largely on data generated on the Illumina platform, which monitors the incorporation of labelled nucleotides during DNA copying in a hugely parallel fashion and has low error rates (<1% per base sequenced [24]). However, the correct determination of a novel HCMV genome from short-read data remains a challenge [23]. Repeat regions typically hinder the full reconstruction of an HCMV genome solely by de novo read assembly, and recombination, gene loss, and variability among HCMV strains result in suitable sequences usually not being available for reconstruction solely by aligning reads to a reference.

In recent years, there has been an acceleration in the development of high-throughput, long-read sequencing. The Oxford Nanopore Technologies (ONT) platform monitors disruptions in ionic current as DNA molecules pass individually through proteinaceous membrane pores in a highly parallel fashion. Read lengths on this platform can exceed 2 Mb [25,26], thus potentially enabling the entire HCMV genome to be sequenced in a single read. Nanopore sequencing also provides significant benefits for resolving repetitive regions and for investigating the occurrence of recombination [27]. In addition, low capital cost and portability have promoted the use of nanopore sequencing in the field for the rapid monitoring of viral outbreaks in LMICs [28] and in expediting near-real-time molecular epidemiological investigation during the COVID-19 pandemic [29,30]. Early limitations of the technology due to the high read error rate (30%) have been relieved by each iteration of the flow cell pores and base-calling algorithms [25,31–33], and recent upgrades employing a refined motor enzyme, improved flow cells, and enhanced software have further decreased the read error rate to <1% (<https://nanoporetech.com/accuracy>, accessed on 30 March 2023).

Nanopore sequencing has been applied to HCMV in several studies. In one study, nanopore data were generated from a laboratory strain (TB40/E) grown in cell culture and assembled de novo. The analysis was complicated by the existence of many variants in the viral stock used, but the accuracy of the consensus sequence (which was not published) in comparison with previously published sequences for this strain was reportedly high [34]. In a second study, nanopore data were generated from another laboratory strain (Merlin) grown in cell culture using bait-based target enrichment. This approach compromised one of the advantages of nanopore sequencing by significantly reducing DNA fragment size and therefore read length [35]. In a third study, the feasibility of nanopore sequencing was demonstrated in a clinical setting focused on antiviral resistance. This study was based on the nanopore sequencing of PCR amplicons from the two genes (UL54 and UL97) most likely to generate resistance mutations and demonstrated that nanopore data were concordant with those obtained by the usual approach of Sanger sequencing [36].

Here, we assessed the utility of direct nanopore sequencing (i.e., without viral DNA enrichment or PCR amplification) for determining complete HCMV genomes from high-viral-load clinical samples. Our finding that the nanopore sequences were highly accurate, although inferior in quality to the sequences determined on the Illumina platform, encourages a future investigation into the application of nanopore sequencing to research HCMV, including that in the clinical setting.

2. Materials and Methods

2.1. Sample Preparation and Characterisation

Two fully anonymised surplus diagnostic extracts from samples that had been prepared on the NucliSENS easyMAG (bioMérieux, Lyon, France) or Abbott m2000 (Abbott Laboratories, Chicago, IL, USA) platforms were provided by the West of Scotland Specialist Virology Centre. The samples had high ratios of HCMV to human DNA, as determined using qPCR assays for the HCMV UL97 gene (UL97F: ACCGTCTGCGCGAATGTTA; UL97R: TCGCAGATGAGCAGCTTCTG; UL97P: VIC-CACCCTGCTTTCCGAC-MGB (VIC, 2'-chloro-phenyl-1,4-dichloro-6-carboxyfluorescein; MGB, minor groove binder.)) and the human *FOXP2* gene (FOXP2_FWD: TCACTACTAACAATTCCTCCTCGACTAC; FOXP2_REV: GATGAGTTATTGTGGTGATGCTT; FOXP2_PROBE: VIC-TCCTCCAACACTTCG-MGB). The amplification conditions were as follows: an initial step of 50 °C for 2 min, denaturation at 95 °C for 10 min, 40 cycles of 95 °C for 30 s, and annealing and elongation at 60 °C for 1 min. One sample had been extracted from urine and had viral and human DNA loads of 258,199 IU/μL and 323 genome copies/μL, respectively; 1 IU is approximately equivalent to 1 genome copy. The other sample had been extracted from lung tissue and had viral and human DNA loads of 172,259 IU/μL and 9636 genome copies/μL, respectively. The urine sample was therefore approximately 50 times more enriched in viral DNA than the lung sample.

The pipelines used to generate sequence data from the samples and assemble the HCMV genomes are summarised in Figure 1. The key steps in the Illumina pipeline (A1–A7) and the ONT pipeline (B1–B13) are summarised in the legend and referred to in the descriptions below.

2.2. Illumina Sequencing

DNA (0.29 μg and 0.58 μg from the urine and lung sample, respectively) was sheared to an average size of 500 bp using an LE220 sonicator (Covaris, Woburn, MA, USA). HCMV DNA enrichment using Agilent SureSelectXT (Agilent Technologies, Santa Clara, CA, USA) custom biotinylated RNA baits and a KAPA Biosystems sequencing library preparation kit were carried out as described previously [18,21,22] (A1). The sequencing libraries were indexed using ultrapure TruGrade oligonucleotides (Integrated DNA Technologies, Leuven, Belgium) and sequenced on a MiSeq instrument (Illumina, San Diego, CA, USA) (A2). The 4,652,420 and 5,699,896 paired-end reads of 150 nucleotides (nt) obtained from the urine and lung samples, respectively, formed the input Illumina data for analysis (A3).

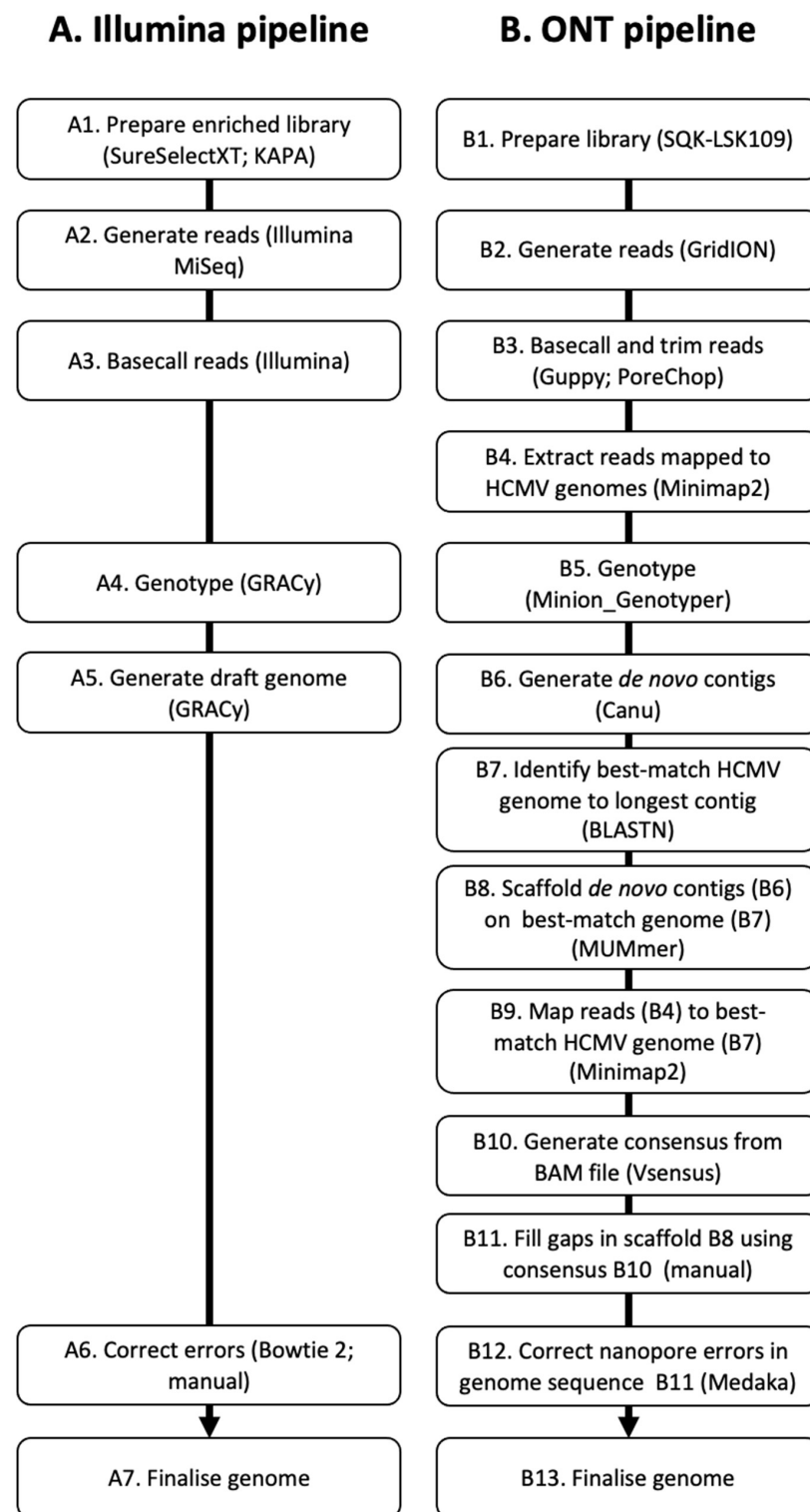


Figure 1. Sequencing and bioinformatics pipelines. The steps in (A) the Illumina pipeline and (B) the ONT pipeline consist of preparation of sequencing libraries (A1 and B1), generation of sequence reads (A2 and B2), extraction of read data (A3 and B3–B4), identification of hypervariable gene genotypes (A4 and B5), generation of draft genomes (A5 and B6), improvement of draft genomes (A6 and B7–B12), and finalisation of genomes (A7 and B13). The bioinformatic means by which the steps were achieved are given in parentheses. The steps are explained further in the text.

2.3. Nanopore Sequencing

DNA (2.5 µg and 7.9 µg from the urine and lung samples, respectively) was processed into a sequencing library using an SQK-LSK109 ligation sequencing kit (ONT, Oxford, UK; B1). This involved treatment with an NEBNext FFPE DNA repair mix and an NEBNext Ultra II end repair/dA-tailing module (New England Biolabs, Ipswich, MA, USA), purification using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA), ligation of adapters using an NEBNext Quick ligation module (New England Biolabs), enrichment of large (>3 kb) fragments using long fragment buffer, and final purification in elution buffer. The libraries were loaded onto primed R9.4.1 flow cells (ONT) at concentrations of 17.7 and 37.4 ng/µL for the urine and lung samples, respectively, and sequencing was carried out for 72 h on a GridION instrument (ONT; B2). The reads were acquired using MinKNOW v4.3.11 (ONT; <https://nanoporetech.com>, accessed on 16 November 2021), and FAST5-formatted files were base-called using Guppy v4.0 (ONT; <https://nanoporetech.com>, accessed on 16 November 2021) in high accuracy mode with a minimum quality score of 7. The reads were assessed using MinIONQC v1.3.5 [37], and adapters were trimmed and chimeric reads with internal adapters removed using PoreChop v0.2.3 (<https://github.com/rrwick/Porechop>, accessed on 16 November 2021) (B3). The 10,616,519 and 8,731,192 reads obtained from the urine and lung samples, respectively, were then mapped to a collated set of 265 HCMV genomes representing all HCMV strains that had been sequenced at the time of analysis (April 2021; Table S1) using Minimap2 v2.17 [38], and mapped reads were extracted using Samtools v1.9 [39]. The 37,334 and 11,841 HCMV-enriched reads obtained from the urine and lung samples, respectively, formed the input nanopore data for analysis (B4). Longer reads were obtained from the urine sample: for the urine and lung samples, respectively, mean read lengths were 1431 and 735 nt and maximum read lengths were 169,079 and 92,239 nt.

2.4. Genotyping Using Illumina and Nanopore Data

The genotypes of 13 hypervariable HCMV genes were determined from the input Illumina data using the genotyping module in the HCMV genome assembly program GRACy v0.4.4 [23] (A4) and from the input nanopore data using Minion_Genotyper v1.0 (https://github.com/salvocamiolo/minion_Genotyper/, accessed on 16 November 2021) (B5). Both programs enumerate reads containing conserved genotype-specific kmers for each hypervariable gene. These kmers were initially defined at one per genotype for 12 genes [21,22] and were then increased to include several per genotype for 13 genes [23].

2.5. Genome Determination Using Illumina Data

The Illumina input data were assembled using GRACy (A5). The draft genomes were assessed by trimming and quality-filtering the reads using Trim Galore v0.4.0 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/, accessed on 10 November 2020) under the parameters `-illumina -paired`, aligning the filtered reads to the sequences using Bowtie 2 v2.3.1 [40] and Samtools v1.3 [38], inspecting the alignments visually using Tablet v1.21.02.08 [41] and making any necessary amendments (A6). Coverage statistics were generated using Tablet. These steps resulted in properly configured HCMV genomes (A7).

2.6. Genome Determination Using Nanopore Data

Draft genomes were determined using two parallel approaches, one involving de novo read assembly into contigs, and the other involving read alignment to a reference.

In the first approach (B6–B8), the nanopore input data for each sample were assembled into contigs using Canu v1.9 [42] under the parameters `genomeSize = 240,000 min ReadLength = 500 min OverlapLength = 50 -nanopore-raw` (B6). The genome in the collated set of 265 genomes that best matched the longest contig was identified using BLASTN v2.4.0 (B7). The GenBank accession numbers of these genomes are KY490081.1 and KP745641.1 for the urine and lung genomes, respectively. The contigs were compared with the best-matching genome using MUMmer v4.0 and Mummerplot v3.5 [43] and rear-

ranged manually via inspection of an alignment in AliView v1.27 [44] (B8). To accomplish this, a text editor was used to order and divide the contigs to ensure that the genome termini were at the ends of the sequences, to locate any gaps, and to represent the unique regions (U_L and U_S) and inverted repeats ($ab/b'a'$ and $a'c'/ac$) appropriately (unlike the situation with the Illumina genomes, the inverted repeats were not assumed to be identical). This approach generated an incomplete consensus genome.

In the second approach (B9–B10), the best-matching genome (from B7) was used as the reference for read mapping using Minimap2 (B9). BAM and mpileup files were generated from the alignment using Samtools v1.9 [39], and the mpileup files were parsed to output a consensus genome using VSensus (<https://github.com/rjorton/VSensus>, accessed on 10 November 2020) with the minimum coverage and minimum base quality parameters set at 20 and 0, respectively (B10). Coverage statistics were generated using weeSAM v1.5 (<https://github.com/centre-for-virus-research/weeSAM/blob/master/weeSAM>, accessed on 10 November 2020).

The genomes derived by the two approaches were aligned using MAFFT v7.310, and any gaps in the former were filled with data from the latter (B11). Finally, this draft genome was processed using Medaka v1.0.3 (<https://github.com/nanoporetech/medaka>, accessed on 10 November 2020), which has been trained to correct for common errors in ONT data (B12). These steps resulted in properly configured HCMV genomes that did not depend at any stage on the corresponding Illumina data (B13).

2.7. Analysis of Resistance Mutations

The coding sequences of genes UL97 and UL54 were analysed using a web-based mutation resistance analyser (MRA; <https://www.informatik.uni-ulm.de/ni/mitarbeiter/HKestler/mra/app/index.php?plugin=form>, accessed on 16 June 2021) [45].

3. Results

Initial screening of the Illumina data to determine the number of HCMV strains present in the original samples was carried out using GRACy, and that of the nanopore data involved using a program devised for our study, Minion_Genotyper. Neither tool requires a genome but rather analyses the reads using a kmer-based approach, and each identified the same single genotype for each of the 13 hypervariable genes analysed (Table 1). This analysis confirmed the ability of Minion_Genotyper to genotype HCMV strains from nanopore data accurately and indicated that each sample contained a single strain. Subsequent analysis of the consensus genomes confirmed the genotypes. Consistent with the occurrence of recombination during HCMV evolution, the constellation of genotypes for each genome was unique in comparison with those reported previously [21,22].

Determination of the final Illumina genomes was accomplished using GRACy followed by manual amendments and ended with 86% of reads aligning to the urine genome (234,678 bp) at an average coverage depth of 2554 reads/nt and 66% of reads aligning to the lung genome (235,215 bp) at an average coverage depth of 2368 reads/nt. The final nanopore genomes were determined by a more complex series of steps. De novo read assembly resulted in a single contig (255,416 nt) corresponding to the entire HCMV genome and a partial repeat for the urine sample but in 20 contigs for the lung sample, with the four largest originating from the HCMV genome. The three largest of these contigs (37,585, 40,716, and 156,323 nt) represented the whole genome but for one gap of 1849 nt. Subsequent steps led to an intact nanopore genome for both samples, each with complete read coverage and determined independently of the corresponding Illumina genome. The analysis ended with 37,241 reads aligning to the urine genome at an average coverage depth of 203 reads/nt, and with 11,764 reads aligning to the lung genome at an average coverage depth of 26 reads/nt.

Table 1. Genotypes of 13 hypervariable HCMV genes in the urine and lung samples.

Gene	Urine	Lung
RL5A	G1	G2
RL6	G6	G4
RL12	G7	G1B
RL13	G7	G1
UL1	G7	G1
UL9	G1	G4
UL11	G1	G1
UL20	G6	G6
UL73	G4D	G1
UL74	G5	G1A
UL120	G1A	G4B
UL146	G2	G10
UL139	G1A	G2

Comparisons between the Illumina and nanopore genomes revealed differences at 62 nt (0.03%) and 164 nt (0.07%) for the urine and lung samples, respectively (Table S2). The great majority of differences (97 and 90%, respectively) were due to insertions or deletions (indels) associated with homopolymeric tracts that almost always caused frameshifting in the nanopore genomes when present in coding regions; this is a recognised issue with nanopore sequencing [33]. The Illumina version at the position of each difference was validated by a visual inspection of the read alignment. Moreover, comparisons with the corresponding regions in the collated set of 265 genomes supported the Illumina version more frequently than the nanopore version for the great majority of differences (95 and 96%, respectively). However, although Illumina data are known to represent the lengths of homopolymeric tracts more accurately than nanopore data [46], they tend to become heterogeneous with longer tracts due to natural genome variation, PCR error during sequencing library preparation, and lower coverage depth. Nonetheless, the Illumina genomes were more accurate than the nanopore genomes.

In clinical practice, the sequences of genes UL54 and UL97 are typically examined for resistance mutations when patients fail to respond to antiviral therapy. As a test of the utility of nanopore genomes for such investigations, the coding sequences of these genes were submitted to MRA, which hosts a curated database of published resistance mutations from a range of viruses including HCMV (Figure 2). The UL97 sequences of the Illumina and nanopore genomes from the urine sample were identical, whereas the UL54 sequence contained an erroneous deletion of 1 nt in the nanopore genome from the lung sample (Table S2). MRA successfully registered the frameshift caused by this deletion. Differences associated with strain polymorphisms or mutations not associated with resistance were registered in both genes in both samples, but no mutations associated with drug resistance were found in either gene in either sample. As the clinical samples were fully anonymised, no information was available on whether the patients had received antiviral therapy. However, these findings illustrate the utility of analysing nanopore data using MRA.

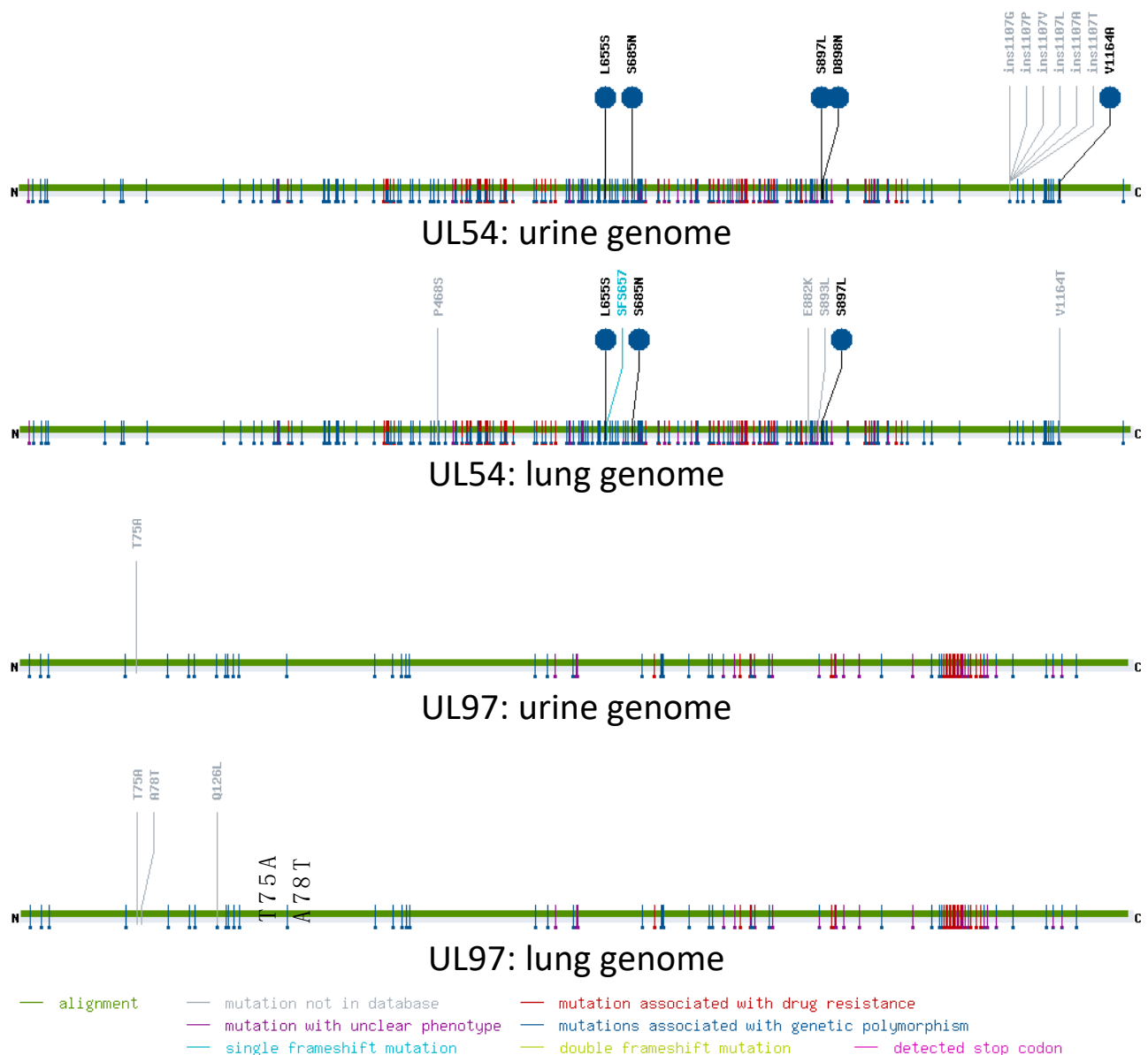


Figure 2. Graphical output from mutation resistance analyser (MRA) for the nanopore sequences of genes UL54 (3.7 kbp) and UL97 (2.1 kbp). In each panel, the sequence alignment is represented by a line extending from the 5' end encoding the N terminus of the protein (N) to the 3' end encoding the C terminus (C). Mutations are shown by the flags coloured according to the key at the foot. Mutations flagged above the line are present in the sequences analysed, and mutations below the line are those in the curated database. The coordinates relate to encoded amino acid residues.

4. Discussion

To our knowledge, our study is the first to use nanopore technology to determine complete HCMV genomes from clinical samples without involving cell culture, bait-based target enrichment, PCR amplification, incorporation of data for the same genome acquired on other platforms, or amendments based on prior knowledge of the target sequence. We sequenced two high-viral-load samples from urine and lung on the Illumina platform with bait-based target enrichment and independently on the ONT platform without enrichment. We used separate genotyping of the Illumina and nanopore data to demonstrate the presence of a single HCMV strain in each sample. The nanopore genomes were determined using a combination of two approaches, one involving de novo read assembly and the other involving read alignment to the best-matching HCMV genome, with the outputs

being incorporated and polished to produce the final sequences. As anticipated from the approximately 50-fold greater ratio of viral to nonviral DNA in the urine sample, the urine genome was assembled more readily and at a greater read depth than the lung genome.

Sequence comparisons suggested that the differences between the Illumina and nanopore genomes were largely or entirely due to errors in the latter, as would be anticipated from the characteristics of the platforms used. Nonetheless, the nanopore genomes were highly accurate (99.97 and 99.93% for the urine and lung genomes, respectively), with the errors distributed thinly (on average, one approximately every 3.8 and 1.5 kbp, respectively) and located mostly in polynucleotide tracts in noncoding regions. Thus, although improvements in the ONT hardware and the software for analysing the data have steadily boosted the base accuracy to >99% (<https://nanoporetech.com/accuracy>, accessed on 30 March 2023), in our study they did not attain the level of accuracy achieved using the Illumina platform. For a completely accurate characterisation of HCMV strains on the ONT platform, supplementary data from another platform would be required. A further implication is that it would be inadvisable to rely on nanopore sequencing alone to determine the sequence of a novel large DNA virus for which no prior sequence information is available.

The main limitation of our study was that the two clinical samples were chosen for sequencing because they contained a high number of HCMV genomes and a relatively high ratio of HCMV to human DNA, thus meeting the requirements of nanopore sequencing. In a clinical setting, the mainstay sample type for post-transplant monitoring of HCMV reactivation or re-infection is blood, but HCMV genomes in blood are highly fragmented, thus obviating one of the chief advantages of nanopore sequencing; blood also contains high levels of host DNA, which can lead to low efficiency enrichment for HCMV genomes [47,48]. In contrast, urine is normally depleted of host cells and was shown in our study to be suitable for sequencing the HCMV genome with a high accuracy on the ONT platform. In instances of post-transplant HCMV disease refractory to treatment, urine is a readily obtained, non-invasive sample type. However, our protocol necessitates high HCMV to human DNA ratios, and for a broader application would require the incorporation of a target enrichment step that is capable of preserving large DNA molecules. The treatment of urine samples with DNase before DNA extraction may be advantageous in this regard if the majority of viral DNA is encapsidated.

Despite this limitation, nanopore technology potentially offers much to HCMV testing in clinical practice in the context of transplantation, particularly as the focus is usually on one or two genes, at which scale nanopore sequencing is almost as accurate as Illumina sequencing. In the UK, HCMV antiviral resistance testing is referred to specialist reference laboratories in which analysis is based on Sanger sequencing of the PCR-amplified UL97 and UL54 genes applied on a timescale that is not conducive to prompt clinical action. A nanopore protocol similar to our approach could provide a means of resistance testing on a shorter timescale. Recently, an Illumina sequencing pipeline was established that enables whole genome HCMV analysis for resistance mutations, which could facilitate resistance testing against novel drug targets [48]. Nanopore technology can also sequence whole HCMV genomes and, in addition, has a modest capital outlay and a small footprint. The continued development of online bioinformatic tools is also making nanopore technology increasingly accessible outside of the research context [36]. For lower viral load samples, PCR products covering antiviral genes could be sequenced using nanopore technology, potentially enabling a more rapid analysis. Furthermore, the application of nanopore technology could be envisaged for investigating congenital infection, where viral loads are typically high, in the order of that of the urine sample used in this study. As the management of congenitally infected infants is nuanced, with treatment reserved for a proportion of the 10% who are symptomatic with severe disease at birth [49], a more widely accessible sequencing system may prove advantageous, particularly if it transpires that there are links between certain HCMV strains, or mixtures of strains, and disease outcome. However, further work to establish thresholds for the detection of low-level variants as applied to Illumina sequence data will be required [47].

In conclusion, we were able to determine two HCMV genomes by direct nanopore sequencing from high-viral-load clinical samples at an accuracy that was close enough to that of Illumina sequencing to encourage further investigation. We await with interest additional improvements in the hardware and software used in this technology, as well as appropriate target enrichment techniques.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v15061248/s1>, Table S1: Collated HCMV genomes; Table S2: Differences between HCMV genomes determined on the Illumina (I) and ONT (N) platforms.

Author Contributions: Conceptualization, K.K.L., A.J.D. and R.J.O.; methodology, K.K.L., B.L., S.C., N.M.S., A.J.D. and R.J.O.; software, S.C. and R.J.O.; validation, K.K.L., S.C., A.J.D. and R.J.O.; formal analysis, K.K.L., A.J.D. and R.J.O.; investigation, K.K.L., A.J.D. and R.J.O.; resources, R.G. and A.J.D.; data curation, K.K.L., A.J.D. and R.J.O.; writing—original draft preparation, K.K.L., A.J.D. and R.J.O.; writing—review and editing, K.K.L., B.L., S.C., N.M.S., R.G., A.J.D. and R.J.O.; supervision, B.L., N.M.S., A.J.D. and R.J.O.; funding acquisition, A.J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Medical Research Council (grant numbers MC_UU_12014/12 and MC_UU_12014/3) and the Wellcome Trust (grant number 204870/Z/16/Z).

Institutional Review Board Statement: Ethical approval was not required for the use of extracts from fully anonymised surplus diagnostic samples in validating sequencing protocols.

Informed Consent Statement: Patient consent was waived for the use of extracts from fully anonymised surplus diagnostic samples in validating sequencing protocols.

Data Availability Statement: The Illumina genomes for the lung and urine samples have been deposited in GenBank (accessions OQ466311 and OQ466312, respectively). The nanopore genomes for these samples are available on Zenodo (<https://doi.org/10.5281/zenodo.7744466> accessed on 30 March 2023). The HCMV reads supporting these genomes have been deposited in the NCBI Sequence Read Archive (SRA) [50] under BioProject number PRJNA945420.

Acknowledgments: We greatly acknowledge the support and infrastructure provided by the Genomics and Bioinformatics groups at the MRC-University of Glasgow Centre for Virus Research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Gatherer, D.; Depledge, D.P.; Hartley, C.A.; Szpara, M.L.; Vaz, P.K.; Benkő, M.; Brandt, C.R.; Bryant, N.A.; Dastjerdi, A.; Doszpoly, A.; et al. ICTV Virus Taxonomy Profile: *Herpesviridae* 2021. *J. Gen. Virol.* **2021**, *102*, 001673. [CrossRef]
2. Dollard, S.C.; Grosse, S.D.; Ross, D.S. New estimates of the prevalence of neurological and sensory sequelae and mortality associated with congenital cytomegalovirus infection. *Rev. Med. Virol.* **2007**, *17*, 355–363. [CrossRef] [PubMed]
3. Zuhair, M.; Smit, G.S.A.; Wallis, G.; Jabbar, F.; Smith, C.; Devleeschauwer, B.; Griffiths, P. Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis. *Rev. Med. Virol.* **2019**, *29*, e2034. [CrossRef]
4. Sinclair, J.; Sissons, P. Latency and reactivation of human cytomegalovirus. *J. Gen. Virol.* **2006**, *87*, 1763–1779. [CrossRef]
5. Kotton, C.N.; Kumar, D.; Caliendo, A.M.; Asberg, A.; Chou, S.; Danziger-Isakov, L.; Humar, A.; Transplantation Society International CMV Consensus Group. Updated international consensus guidelines on the management of cytomegalovirus in solid-organ transplantation. *Transplantation* **2013**, *96*, 333–360. [CrossRef]
6. Manicklal, S.; Emery, V.C.; Lazzarotto, T.; Boppana, S.B.; Gupta, R.K. The “silent” global burden of congenital cytomegalovirus. *Clin. Microbiol. Rev.* **2013**, *26*, 86–102. [CrossRef]
7. Britt, W.J.; Prichard, M.N. New therapies for human cytomegalovirus infections. *Antivir. Res.* **2018**, *159*, 153–174. [CrossRef]
8. Dolan, A.; Cunningham, C.; Hector, R.D.; Hassan-Walker, A.F.; Lee, L.; Addison, C.; Dargan, D.J.; McGeoch, D.J.; Gatherer, D.; Emery, V.C.; et al. Genetic content of wild-type human cytomegalovirus. *J. Gen. Virol.* **2004**, *85*, 1301–1312. [CrossRef]
9. Davison, A.J.; Holton, M.; Dolan, A.; Dargan, D.J.; Gatherer, D.; Hayward, G.S. Comparative genomics of primate cytomegaloviruses. In *Cytomegaloviruses: From Molecular Pathogenesis to Intervention*; Reddehase, M.J., Ed.; Caister Academic Press: Poole, UK, 2013; Volume 1, pp. 1–22.

10. Meyer-König, U.; Vogelberg, C.; Bongarts, A.; Kampa, D.; Delbrück, R.; Wolff-Vorbeck, G.; Kirste, G.; Haberland, M.; Hufert, F.T.; von Laer, D. Glycoprotein B genotype correlates with cell tropism in vivo of human cytomegalovirus infection. *J. Med. Virol.* **1998**, *55*, 75–81. [\[CrossRef\]](#)
11. Barbi, M.; Binda, S.; Caroppo, S.; Primache, V.; Didò, P.; Guidotti, P.; Corbetta, C.; Melotti, D. CMV gB genotypes and outcome of vertical transmission: Study on dried blood spots of congenitally infected babies. *J. Clin. Virol.* **2001**, *21*, 75–79. [\[CrossRef\]](#)
12. Puchhammer-Stöckl, E.; Görzer, I.; Zoufaly, A.; Jaksch, P.; Bauer, C.C.; Klepetko, W.; Popow-Kraupp, T. Emergence of multiple cytomegalovirus strains in blood and lung of lung transplant recipients. *Transplantation* **2006**, *81*, 187–194. [\[CrossRef\]](#)
13. Rasmussen, L.; Geissler, A.; Winters, M. Inter- and intragenic variations complicate the molecular epidemiology of human cytomegalovirus. *J. Infect. Dis.* **2003**, *187*, 809–819. [\[CrossRef\]](#)
14. Cha, T.A.; Tom, E.; Kemble, G.W.; Duke, G.M.; Mocarski, E.S.; Spaete, R.R. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J. Virol.* **1996**, *70*, 78–83. [\[CrossRef\]](#)
15. Bradley, A.J.; Lurain, N.S.; Ghazal, P.; Trivedi, U.; Cunningham, C.; Baluchova, K.; Gatherer, D.; Wilkinson, G.W.G.; Dargan, D.J.; Davison, A.J. High-throughput sequence analysis of variants of human cytomegalovirus strains Towne and AD169. *J. Gen. Virol.* **2009**, *90*, 2375–2380. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Dargan, D.J.; Douglas, E.; Cunningham, C.; Jamieson, F.; Stanton, R.J.; Baluchova, K.; McSharry, B.P.; Tomasec, P.; Emery, V.C.; Percivalle, E.; et al. Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. *J. Gen. Virol.* **2010**, *91*, 1535–1546. [\[CrossRef\]](#)
17. Morey, M.; Fernández-Marmiesse, A.; Castiñeiras, D.; Fraga, J.M.; Couce, M.L.; Cocho, J.A. A glimpse into past, present, and future DNA sequencing. *Mol. Genet. Metab.* **2013**, *110*, 3–24. [\[CrossRef\]](#)
18. Hage, E.; Wilkie, G.S.; Linnenweber-Held, S.; Dhingra, A.; Suárez, N.M.; Schmidt, J.J.; Kay-Fedorov, P.C.; Mischak-Weissinger, E.; Heim, A.; Schwarz, A.; et al. Characterization of human cytomegalovirus genome diversity in immunocompromised hosts by whole-genome sequencing directly from clinical specimens. *J. Infect. Dis.* **2017**, *215*, 1673–1683. [\[CrossRef\]](#)
19. Sijmons, S.; Thys, K.; Mbong Ngwese, M.; Van Damme, E.; Dvorak, J.; Van Loock, M.; Li, G.; Tachezy, R.; Busson, L.; Aerssens, J.; et al. High-throughput analysis of human cytomegalovirus genome diversity highlights the widespread occurrence of gene-disrupting mutations and pervasive recombination. *J. Virol.* **2015**, *89*, 7673–7695. [\[CrossRef\]](#)
20. Lassalle, F.; Depledge, D.P.; Reeves, M.B.; Brown, A.C.; Christiansen, M.T.; Tutill, H.J.; Williams, R.J.; Einer-Jensen, K.; Holdstock, J.; Atkinson, C.; et al. Islands of linkage in an ocean of pervasive recombination reveals two-speed evolution of human cytomegalovirus genomes. *Virus Evol.* **2016**, *2*, vew017. [\[CrossRef\]](#)
21. Suárez, N.M.; Wilkie, G.S.; Hage, E.; Camiolo, S.; Holton, M.; Hughes, J.; Maabar, M.; Vattipally, S.B.; Dhingra, A.; Gompels, U.A.; et al. Human cytomegalovirus genomes sequenced directly from clinical material: Variation, multiple-strain infection, recombination, and gene loss. *J. Infect. Dis.* **2019**, *220*, 781–791. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Suárez, N.M.; Musonda, K.G.; Escrava, E.; Njenga, M.; Agbueze, A.; Camiolo, S.; Davison, A.J.; Gompels, U.A. Multiple-strain infections of human cytomegalovirus with high genomic diversity are common in breast milk from human immunodeficiency virus-infected women in Zambia. *J. Infect. Dis.* **2019**, *220*, 792–801. [\[CrossRef\]](#)
23. Camiolo, S.; Suárez, N.M.; Chalka, A.; Venturini, C.; Breuer, J.; Davison, A.J. GRACy: A tool for analysing human cytomegalovirus sequence data. *Virus Evol.* **2020**, *7*, veaa099. [\[CrossRef\]](#)
24. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. [\[CrossRef\]](#)
25. Jain, M.; Tyson, J.R.; Loose, M.; Ip, C.L.C.; Eccles, D.A.; O’Grady, J.; Malla, S.; Leggett, R.M.; Wallerman, O.; Jansen, H.J.; et al. MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9.0 chemistry. *F1000Research* **2017**, *6*, 760. [\[CrossRef\]](#)
26. Payne, A.; Holmes, N.; Rakyar, V.; Loose, M. BulkVis: A graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* **2019**, *35*, 2193–2198. [\[CrossRef\]](#)
27. Chin, C.S.; Peluso, P.; Sedlazeck, F.J.; Nattestad, M.; Concepcion, G.T.; Clum, A.; Dunn, C.; O’Malley, R.; Figueroa-Balderas, R.; Morales-Cruz, A.; et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **2016**, *13*, 1050–1054. [\[CrossRef\]](#)
28. Quick, J.; Loman, N.J.; Duraffour, S.; Simpson, J.T.; Severi, E.; Cowley, L.; Bore, J.A.; Koundouno, R.; Dudas, G.; Mikhail, A.; et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **2016**, *530*, 228–232. [\[CrossRef\]](#)
29. da Silva Filipe, A.; Shepherd, J.G.; Williams, T.; Hughes, J.; Aranday-Cortes, E.; Asamaphan, P.; Ashraf, S.; Balcazar, C.; Bruncker, K.; Campbell, A.; et al. Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from mainland Europe into Scotland. *Nat. Microbiol.* **2021**, *6*, 112–122. [\[CrossRef\]](#)
30. Meredith, L.W.; Hamilton, W.L.; Warne, B.; Houldcroft, C.J.; Hosmillo, M.; Jahun, A.S.; Curran, M.D.; Parmar, S.; Caller, L.G.; Caddy, S.L.; et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: A prospective genomic surveillance study. *Lancet Infect. Dis.* **2020**, *20*, 1263–1272. [\[CrossRef\]](#)
31. Loman, N.J.; Quinlan, A.R. Poretools: A toolkit for analyzing nanopore sequence data. *Bioinformatics* **2014**, *30*, 3399–3401. [\[CrossRef\]](#)
32. Jain, M.; Fiddes, I.T.; Miga, K.H.; Olsen, H.E.; Paten, B.; Akeson, M. Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* **2015**, *12*, 351–356. [\[CrossRef\]](#)

33. Rang, F.J.; Kloosterman, W.P.; de Ridder, J. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* **2018**, *19*, 90. [\[CrossRef\]](#)
34. Karamitros, T.; van Wilgenburg, B.; Wills, M.; Klennerman, P.; Magiorkinis, G. Nanopore sequencing and full genome de novo assembly of human cytomegalovirus TB40/E reveals clonal diversity and structural variations. *BMC Genom.* **2018**, *19*, 577. [\[CrossRef\]](#)
35. Eckert, S.E.; Chan, J.Z.; Houniet, D.; The Pathseek Consortium; Breuer, J.; Speight, G. Enrichment by hybridisation of long DNA fragments for Nanopore sequencing. *Microb. Genom.* **2016**, *2*, e000087. [\[CrossRef\]](#)
36. Chorlton, S.D.; Ritchie, G.; Lawson, T.; McLachlan, E.; Romney, M.G.; Matic, N.; Lowe, C.F. Next-generation sequencing for cytomegalovirus antiviral resistance genotyping in a clinical virology laboratory. *Antivir. Res.* **2021**, *192*, 105123. [\[CrossRef\]](#)
37. Lanfear, R.; Schalamun, M.; Kainer, D.; Wang, W.; Schwessinger, B. MinIONQC: Fast and simple quality control for MinION sequencing data. *Bioinformatics* **2019**, *35*, 523–525. [\[CrossRef\]](#)
38. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [\[CrossRef\]](#)
39. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [\[CrossRef\]](#)
40. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [\[CrossRef\]](#)
41. Milne, I.; Bayer, M.; Stephen, G.; Cardle, L.; Marshall, D. Tablet: Visualizing next-generation sequence assemblies and mappings. *Methods Mol. Biol.* **2016**, *1374*, 253–268. [\[CrossRef\]](#)
42. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **2017**, *27*, 722–736. [\[CrossRef\]](#)
43. Marçais, G.; Delcher, A.L.; Phillippy, A.M.; Coston, R.; Salzberg, S.L.; Zimin, A. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **2018**, *14*, e1005944. [\[CrossRef\]](#)
44. Larsson, A. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **2014**, *30*, 3276–3278. [\[CrossRef\]](#)
45. Chevillotte, M.; von Einem, J.; Meier, B.M.; Lin, F.M.; Kestler, H.A.; Mertens, T. A new tool linking human cytomegalovirus drug resistance mutations to resistance phenotypes. *Antivir. Res.* **2010**, *85*, 318–327. [\[CrossRef\]](#)
46. Dohm, J.C.; Peters, P.; Stralis-Pavese, N.; Himmelbauer, H. Benchmarking of long-read correction methods. *NAR Genom. Bioinform.* **2020**, *2*, lqaa037. [\[CrossRef\]](#)
47. Boom, R.; Sol, C.J.; Schuurman, T.; Van Breda, A.; Weel, J.F.; Beld, M.; Ten Berge, I.J.; Wertheim-Van Dillen, P.M.; De Jong, M.D. Human cytomegalovirus DNA in plasma and serum specimens of renal transplant recipients is highly fragmented. *J. Clin. Microbiol.* **2002**, *40*, 4105–4113. [\[CrossRef\]](#)
48. Suárez, N.M.; Blyth, E.; Li, K.; Ganzenmueller, T.; Camiolo, S.; Avdic, S.; Withers, B.; Linnenweber-Held, S.; Gwinner, W.; Dhingra, A.; et al. Whole-genome approach to assessing human cytomegalovirus dynamics in transplant patients undergoing antiviral therapy. *Front. Cell. Infect. Microbiol.* **2020**, *10*, 267. [\[CrossRef\]](#)
49. Rawlinson, W.D.; Boppana, S.B.; Fowler, K.B.; Kimberlin, D.W.; Lazzarotto, T.; Alain, S.; Daly, K.; Doutré, S.; Gibson, L.; Giles, M.L.; et al. Congenital cytomegalovirus infection in pregnancy and the neonate: Consensus recommendations for prevention, diagnosis, and therapy. *Lancet Infect. Dis.* **2017**, *17*, e177–e188. [\[CrossRef\]](#)
50. Leinonen, R.; Sugawara, H.; Shumway, M.; International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res.* **2011**, *39*, D19–D21. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.