

## Article

# Towards a Rapid-Turnaround Low-Depth Unbiased Metagenomics Sequencing Workflow on the Illumina Platforms

Winston Lian Chye Koh <sup>1</sup>, Si En Poh <sup>2</sup>, Chun Kiat Lee <sup>3</sup>, Tim Hon Man Chan <sup>3</sup>, Gabriel Yan <sup>4,5</sup>, Kiat Whye Kong <sup>2</sup>, Lalita Lau <sup>2</sup>, Wai Yip Thomas Lee <sup>6</sup>, Clark Cheng <sup>6</sup>, Shawn Hoon <sup>2</sup> and Yiqi Seow <sup>2,7,\*</sup>

- <sup>1</sup> Bioinformatic Institute, A\*STAR (Agency for Science, Technology and Research), Singapore 138632, Singapore; winston\_koh@bii.a-star.edu.sg
- <sup>2</sup> Institute of Molecular and Cell Biology, A\*STAR (Agency for Science, Technology and Research), Singapore 138673, Singapore; poh\_si\_en@imcb.a-star.edu.sg (S.E.P.); kongkw@imcb.a-star.edu.sg (K.W.K.); lalita\_lau@imcb.a-star.edu.sg (L.L.)
- <sup>3</sup> Department of Laboratory Medicine, National University Hospital, Singapore 119228, Singapore; chun\_kiat\_lee@nuhs.edu.sg (C.K.L.); hon\_man\_chan@nuhs.edu.sg (T.H.M.C.)
- <sup>4</sup> Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore; gabriel\_zherong\_yan@nuhs.edu.sg
- <sup>5</sup> Division of Microbiology, Department of Laboratory Medicine, National University Health System, Singapore 119228, Singapore
- <sup>6</sup> Paths Diagnostics Pte Limited, Singapore 349317, Singapore
- <sup>7</sup> Genome Institute of Singapore, A\*STAR (Agency for Science, Technology and Research), Singapore 138672, Singapore
- \* Correspondence: seowy@gis.a-star.edu.sg; Tel.: +65-68-247-108



**Citation:** Koh, W.L.C.; Poh, S.E.; Lee, C.K.; Chan, T.H.M.; Yan, G.; Kong, K.W.; Lau, L.; Lee, W.Y.T.; Cheng, C.; Hoon, S.; et al. Towards a Rapid-Turnaround Low-Depth Unbiased Metagenomics Sequencing Workflow on the Illumina Platforms. *Bioengineering* **2023**, *10*, 520. <https://doi.org/10.3390/bioengineering10050520>

Academic Editors: Teen-Hang Meen, Chun-Yen Chang, Charles Tijus, Po-Lei Lee and Kuei-Shu Hsu

Received: 30 March 2023

Revised: 20 April 2023

Accepted: 21 April 2023

Published: 25 April 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/>).

**Abstract:** Unbiased metagenomic sequencing is conceptually well-suited for first-line diagnosis as all known and unknown infectious entities can be detected, but costs, turnaround time and human background reads in complex biofluids, such as plasma, hinder widespread deployment. Separate preparations of DNA and RNA also increases costs. In this study, we developed a rapid unbiased metagenomics next-generation sequencing (mNGS) workflow with a human background depletion method (HostEL) and a combined DNA/RNA library preparation kit (AmpRE) to address this issue. We enriched and detected bacterial and fungal standards spiked in plasma at physiological levels with low-depth sequencing (<1 million reads) for analytical validation. Clinical validation also showed 93% of plasma samples agreed with the clinical diagnostic test results when the diagnostic qPCR had a Ct < 33. The effect of different sequencing times was evaluated with the 19 h iSeq 100 paired end run, a more clinically palatable simulated iSeq 100 truncated run and the rapid 7 h MiniSeq platform. Our results demonstrate the ability to detect both DNA and RNA pathogens with low-depth sequencing and that iSeq 100 and MiniSeq platforms are compatible with unbiased low-depth metagenomics identification with the HostEL and AmpRE workflow.

**Keywords:** metagenomic sequencing; host depletion; DNA/RNA library preparation; liquid biopsy; infectious disease

## 1. Introduction

Infectious disease is a worldwide health burden, with just the eight major infectious diseases (HIV, malaria, measles, hepatitis, dengue fever, rabies, tuberculosis, and yellow fever) causing more than 156 million life-years lost in 2016 [1]. Thus, early detection and diagnosis is important for timely interventions and treatment. Conventional methods, such as culturing of isolated pathogens and molecular amplification of nucleic acids, are widely used in diagnosis, but these methods require prior knowledge or suspicion of the pathogens [2–5]. In contrast, unbiased (shotgun) metagenomic next-generation sequencing (mNGS), as opposed to a targeted sequencing approach [6], can not only detect multiple pathogens in a single assay but also discover novel or unexpected pathogens, like

SARS-CoV-2 during the recent COVID-19 outbreak [3,7–13]. Conceptually, a dependable, robust metagenomic sequencing solution is an ideal first-line diagnostic if costs are low, time-to-result is fast, and sensitivity is high.

Although metagenomic sequencing technologies are developing rapidly [14–17], there are certain limitations. First, typical protocols require multiple days to complete due to the time-consuming host background depletion, library preparation, and data analysis processes [3,4]. Second, as the samples contain high human background, sequencing depth needs to be high to account for the excess of human reads, which significantly contributes to incurred costs [5,18,19]. Although a few solutions exist for human background depletion, selective depletion followed by centrifugation enriches only bacteria DNA [18,20], which limits utility by omitting DNA or RNA viruses. Third, most library preparation kits either recommends DNA or RNA as input, but viruses can be both RNA and DNA-based; thus, the need to deploy separate workflows for DNA and RNA preparation adds manpower time, and costs to comprehensive, unbiased sequencing.

To address these, we developed a workflow consisting of a HostEL (Host Elimination) kit—a human background depletion strategy that also allows for enrichment of viruses; and Amplification-Restriction Endonuclease fragmentation (AmpRE)—a single tube DNA/RNA library preparation. HostEL uses magnetic bead-immobilized nucleases to deplete human background after selective lysis, replacing enzyme deactivation with magnetic pull-down, to enrich both pathogen DNA and RNA. AmpRE is a purely additive process for DNA and RNA input with only two clean-up steps before the library is ready for sequencing. It can reduce the processing time while maintaining pathogen signal when performing shallow sequencing on iSeq 100, which allows more rapid diagnosis of communicable diseases. With the two strategies combined, our method can not only detect bacteria but also fungi and DNA and RNA viruses. We conducted an analytical validation for the workflow using plasma with spiked-in microbes (ZymoBIOMICS Microbial Community Standard) at various concentration to mimic the physiological infection conditions down to 70 cells/ $\mu\text{L}$  of plasma [21]. The samples were sequenced on two sequencing platforms, iSeq 100 and MiniSeq, with three different sequencing turnaround time to cater to different clinical workflow needs and demonstrated high sensitivity and specificity with each platform. We also conducted a clinical validation with 42 plasma samples, with 93% of the samples agreeing with the qPCR results on the clinical samples.

## 2. Materials and Methods

### 2.1. Analytical Validation

ZymoBIOMICS Microbial Community Standard (Supplementary Table S1), consisting of 8 bacterial and 2 fungal strains at  $1.4 \times 10^7$  total cells/ $\mu\text{L}$ , was diluted 100 $\times$ , 250 $\times$ , 1000 $\times$  and 2000 $\times$  before spiking 5  $\mu\text{L}$  into 500  $\mu\text{L}$  of human plasma to yield at  $1.4 \times 10^6$ ,  $5.6 \times 10^5$ ,  $1.4 \times 10^5$ ,  $7.0 \times 10^4$  total cells/ $\text{mL}$  of plasma, respectively. The spiked-in plasma was either HostEL-depleted (Paths Diagnostics, Singapore) or just bead beaten. For HostEL-treated samples, 30  $\mu\text{L}$  of incubation buffer and 10  $\mu\text{L}$  of nuclease beads were mixed with 500  $\mu\text{L}$  of plasma and incubated for 20 min shaking at 37 °C. The nuclease beads were then attracted with a magnetic rack and supernatant bead-beaten for 30 s before extraction using Zymo Quick DNA/RNA Viral extraction kit (Zymo Research, Irvine, CA, USA), eluting in 30  $\mu\text{L}$ . The control samples were just bead-beaten before extraction. Libraries were prepared with AmpRE kit (Paths Diagnostics) as per manufacturer's instructions. 7  $\mu\text{L}$  of TNA was then used as input into the AmpRE kit. First strand synthesis in 10  $\mu\text{L}$  reaction is followed by the addition of 10  $\mu\text{L}$  second strand reagent for second strand synthesis, then an addition of 30  $\mu\text{L}$  amplification mix was added for amplification with methylated nucleotides. Cleanup with SPRI beads and elution in 10  $\mu\text{L}$  was followed by digestion for 4  $\mu\text{L}$  of eluted DNA in a 5  $\mu\text{L}$  reaction for restriction digest of methylated nucleotides. An amount of 10  $\mu\text{L}$  of ligation buffer was added to the fragmented DNA for ligation and tailing of sequencing arms before amplification and barcoding. The sample was then cleaned up using SPRI beads and eluted in 10  $\mu\text{L}$  and quantified on Agilent 4150 TapeStation Instrument (Agilent

Technologies, Santa Clara, CA, USA) using Agilent D5000 ScreenTape System (Agilent Technologies). Samples were pooled and sequenced on iSeq 100 and MiniSeq (Illumina, San Diego, CA, USA), respectively.

### 2.2. Sequencing on iSeq 100 System and MiniSeq System

Each quantified sample using the Agilent TapeStation was diluted to a concentration of 190 pM with 10 mM Tris HCl, pH 8.5. Up to 5 samples were then pooled together in equal volume, making up to a total of at least 20  $\mu$ L for sequencing on the iSeq100 system and up to 20 samples were pooled for the Miniseq system to ensure that there are sufficient reads per sample across the systems. An amount of 20  $\mu$ L of the pooled sample was loaded onto the iSeq 100 cartridge before loading onto iSeq 100 for sequencing. For MiniSeq, each quantified sample was diluted to a concentration of 10 nM with RSB (10 mM Tris-HCl, pH 8.5 with 0.1% Tween 20) and pooled at equal volume before diluting down to 1 nM with RSB. The pooled library was denatured by mixing 5  $\mu$ L of 1 nM pooled sample and 5  $\mu$ L 0.1 N NaOH and incubating at room temperature for 5 min. 5  $\mu$ L of 200 mM Tris-HCl, pH 7.0, was then added to the denatured library. The denatured library was diluted down to 5 pM by adding 985  $\mu$ L of pre-chilled Hybridization buffer (HT1) to the 15  $\mu$ L of denatured library. The library was further diluted down to loading concentration of 1.6 pM by mixing 160  $\mu$ L of 5 pM denatured libraries with 340  $\mu$ L of prechilled HT1. 500  $\mu$ L of the diluted sample was loaded onto the MiniSeq cartridge before putting onto the system for sequencing.

### 2.3. Data Analysis

Fastq sequence files were exported from the sequencers, and quality control was performed using Prinseq tool, specifically to remove low complexity reads, entropy is set to be 0.4, with quality trimming on both left and right side of the reads up to 10 bases [22]. Reads that are less than 20 bp in length after trimming are removed and resulting high quality reads was then aligned to the human genome using HiSAT2 v2.21 using the GRCh38 reference [23]. On average, ~88.4% of reads were mapped to the human genome. These are considered to be human background reads and removed from downstream analysis. There can be the possibility of reads from non-human sources sharing homology with the human genome being removed at this stage. The remaining unaligned reads can thus be considered as unique reads that do not share homology with human background. These remaining reads were then identified using KRAKEN2 v2.1.2 and BRACKEN v2.6.2 tools. For KRAKEN2, the confidence level is set to 0.05 and was specified to report the minimizer data for manual inspection of assigned reads. As KRAKEN is not an alignment based algorithm, when inspecting and visualizing assigned reads, reads assigned to specific microbes are first aligned to their respective NCBI reference genomes using BWA v0.7.17-r1188, then visualized using IGV v11.0.9.1 and R v4.0.2. The KRACKEN2 database used is a custom database, which is a combination of the standard database with Fungi added on downloaded from (<https://benlangmead.github.io/aws-indexes/k2> (accessed on 2 March 2020)). At least five BRACKEN assigned reads to the species was the cutoff for a positive detection. For the simulated iSeq 100 run, the reads were trimmed down to 100-bases pair-end reads using AWK before processing through the same pipeline. Statistical analysis of sensitivity, specificity and clinical agreement was then performed in R. For the analysis, we defined these statistics as Sensitivity = True positives / (True positives + False negatives) and Specificity = True negatives / (True negatives + False positives); where true positives are the number of spiked-ins that were correctly identified by sequencing as positive. False positives are the number of spike-ins that were incorrectly identified by the sequencing as present in negative plasma. True negatives are defined as the number of spike-ins that were correctly identified by being not present in negative plasma. False negatives are the number of spike ins that are identified by the sequencing test in negative plasma. For the analysis on clinical agreement, we illustrate the numbers in a cross tabulation illustrating

the of occurrence of the different combinations of situations of positive/negative clinical diagnosis vs. sequencing positive/negative.

#### 2.4. Sample Collection and Ethics Statement

The studies involving human participants were reviewed and approved by NHG DSRB Study Reference Number: 2017/00632. The patients/participants provided their written informed consent to participate in this study. An amount of 5 mL of blood was collected in EDTA tubes, and plasma was obtained by centrifugation at  $2000\times g$  for 10 min. 500  $\mu\text{L}$  was aliquoted out and stored at  $-80\text{ }^{\circ}\text{C}$  before HostEL and AmpRE (Paths Diagnostics) processing as per manufacturer's instructions. The samples were prepared as for analytical validation but extracted with EZ1 Virus Mini 2.0 kit (Qiagen, Hilden, Germany) instead, eluting in 60  $\mu\text{L}$  of water.

#### 2.5. qPCR Validation

An amount of 1  $\mu\text{L}$  of extracted samples and 0.3  $\mu\text{M}$  of each primer for pathogen (Supplementary Table S3) was added to each 20  $\mu\text{L}$  reaction. For bacteria, fungi, and DNA viruses, a Maxima SYBR Green/ROX qPCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) was used. For RNA virus, Toyobo Thunderbird One Step qRT-PCR kit (Toyobo, Osaka, Japan) was used with  $1\times$  SyBr Green (Thermo Fisher Scientific). The mixture was amplified on the QuantStudio 1 qPCR system (Thermo Fisher Scientific) using the thermal profile:  $50\text{ }^{\circ}\text{C}$  for 15 min,  $95\text{ }^{\circ}\text{C}$  for 2 min for the initial denaturation, followed by 60 cycles of  $95\text{ }^{\circ}\text{C}$  denaturation (15 s) and  $60\text{ }^{\circ}\text{C}$  annealing and extension (60 s). Melting curve analysis was performed by ramping the temperature up to  $95\text{ }^{\circ}\text{C}$  after qPCR to verify the specificity and identity of each of the PCR product. The list of primers used is shown in Supplementary Table S3.

### 3. Results

#### 3.1. Principles of HostEL and AmpRE

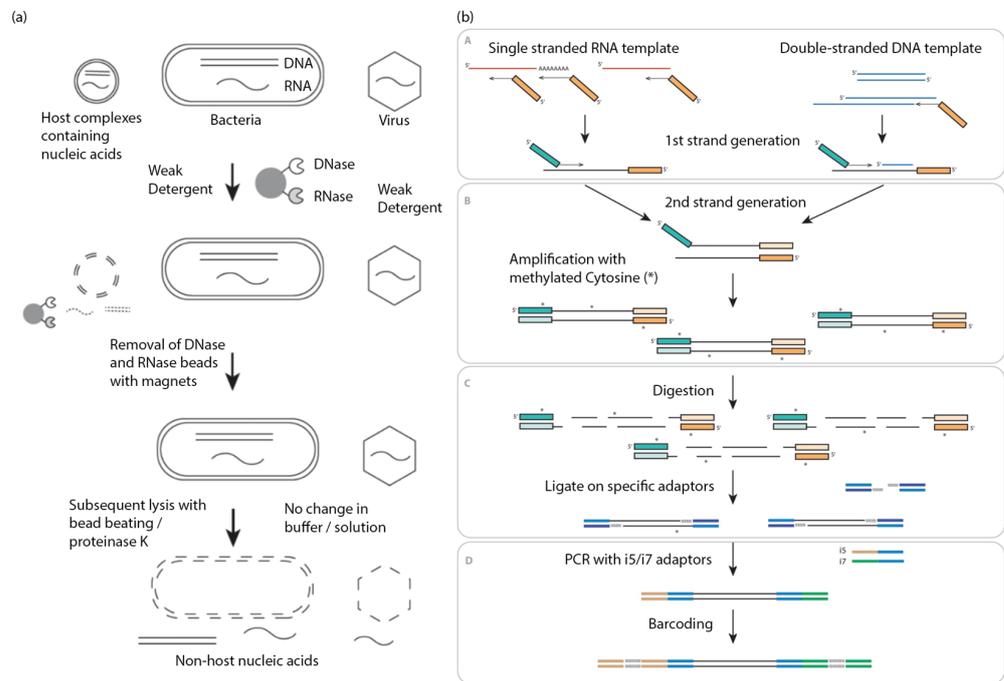
The HostEL kit relies on nucleases were conjugated onto magnetic beads and mixed into 500  $\mu\text{L}$  plasma samples and 30  $\mu\text{L}$  of a Tris-HCl pH 7.5 buffered solution containing 8.25% *w/v* CHAPS hydrate detergent to lyse lipid membranes to release protected nucleic acids, similar to other host depletion methods [24]. The samples were then mixed for 30 min at  $37\text{ }^{\circ}\text{C}$  before the beads were removed. As the nucleases are conjugated onto beads, these are effectively removed prior to extraction and cannot then result in degradation of signal. This is particularly important for RNase as any carryover can be detrimental to metagenomic RNA signal (Figure 1a).

AmpRE is a library preparation kit that uses random primers for both first and second strand synthesis. Once the double stranded library is made, methylated cytosines are incorporated at random through PCR such that the median fragment size derived after cutting with a methylation-dependent restriction endonuclease results in a median of around 400 bp. SPRI purification is performed after PCR, followed by enzymatic fragmentation before universal primers are ligated on and the amplicons amplified by PCR. Barcoding primers are added last before SPRI purification to remove background primers (Figure 1b). Through empirical optimization of the protocol, it can take both DNA and RNA as input and assay both DNA and RNA pathogens.

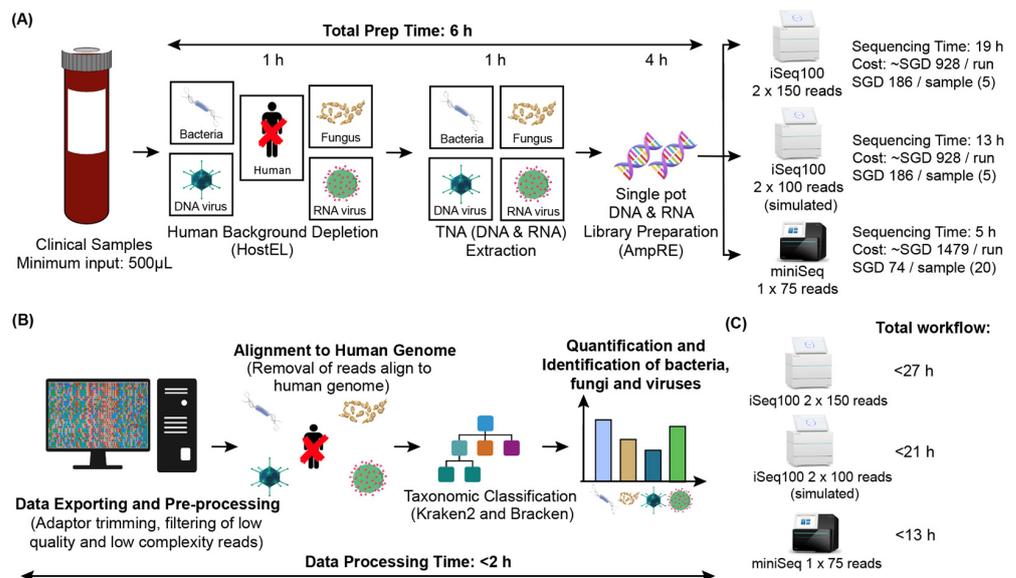
#### 3.2. Low Depth Sequencing Can Detect Physiological Levels of Bacteria

The mNGS workflow utilized includes a host depletion stage with HostEL (Figure 2A), followed by bead-beating and total nucleic acid (TNA) extraction. Sequencing libraries are then generated with a 4 h additive library preparation protocol before sequencing in batches of 5 samples on iSeq 100, a sequencing platform which can produce  $\sim 4$  million reads in 19 h, to achieve low depth  $<1$  million reads per sample. The iSeq 100 was selected for the proof-of-principle as it utilizes a plug-and-play cartridge that is user-friendly for the clinical

laboratory. The reads were aligned to the human genome to discard background before identifying pathogen sequences and quantifying read counts in under 2 h (Figure 2B).



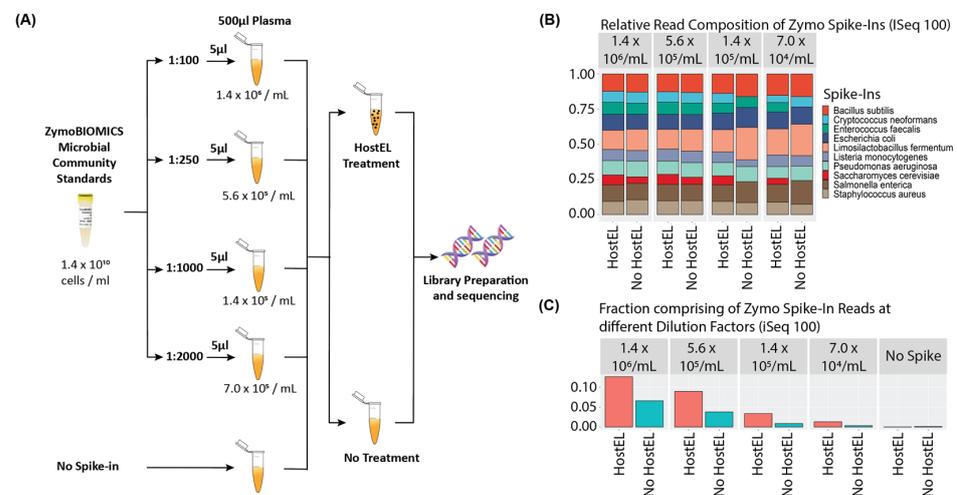
**Figure 1.** HostEL and AmpRE workflows (a) HostEL makes use of selective lysis with nucleases conjugated on microbeads that can be removed magnetically to remove nuclease from downstream reactions without the need for centrifugation, allowing the retention of viral particles; (b) AmpRE allows for simultaneous processing of DNA and RNA in single pot reaction followed by amplification, methylation-based restriction digestion for fragmentation, ligation of sequencing adaptors, and barcoding. \* represents the presence of the methylated cytosine in the DNA library.



**Figure 2.** Schematic diagram of mNGS workflow using HostEL and AmpRE. (A) Workflow of mNGS using HostEL and AmpRE. A minimum of 500 µL of a clinical sample, such as plasma, is treated with HostEL to deplete human background before total nucleic acids (TNA) are extracted. TNA is then prepared into a sequencing library with AmpRE, followed by sequencing on the iSeq 100 (2 × 150; 2 × 100 simulated) or MiniSeq (1 × 100). The time required for each step and the cost of sequencing are indicated. (B) Workflow of data analysis. Sequenced data is exported in Fastq format from the

sequencing platforms, and high quality and compatibility reads are selected. These sequences are then aligned to human genome, where the matched human sequences are removed. The reads of the pathogens (bacteria, fungus, and both DNA and RNA viruses) are then quantified and identified using KRAKEN2 and BRACKEN tools. The attached time for each step is included for reference. (C) Comparison of the different sequencing platforms on the total workflow time.

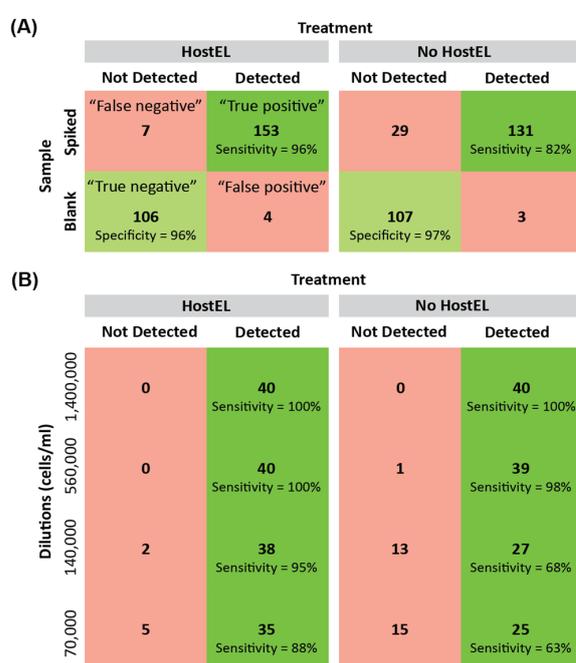
To validate our workflow (Figure 3A), human plasma was spiked with various dilutions of ZymoBIOMICS Microbial Community Standard, which comprises of bacteria and fungi (Supplementary Table S1 and Figure S1), to achieve the total cell counts of 1.4 million, 560,000, 140,000, and 70,000 cells per mL of plasma, respectively. This mimics the physiological concentrations of  $10^5$  copies/mL in the body during active infection [21]. Spike-in samples without HostEL treatment and samples with no spike-in were prepared as negative controls for comparison, which made up a total of 10 samples. Even with a median of 56,498 paired end reads ( $2 \times 150$  bp) per sample across all samples on the iSeq 100, the pipeline identified all 10 species at all dilutions when treated with HostEL while preserving relative distribution of species. In contrast, the untreated samples started losing signal at 140,000 and 70,000 cells per mL (Figure 3B). The enrichment is  $\sim 2$ -fold at 1.4 million cells/mL of plasma of HostEL treated samples and extend to  $\sim 4$ -fold at the lower concentrations on iSeq 100, while the enrichment on MiniSeq is at least 1.5-fold (Supplementary Table S2). In this study, the lowest limit of detection (LOD) was set to 70,000 cells per mL of plasma; the equivalent of a total of 35,000 cells in 500  $\mu$ L of plasma (Figure 3C), demonstrating the depletion of background host reads. The relative distribution of species was preserved at all concentrations, suggesting low-depth ( $<1$  million reads per sample) sequencing allows for detection of microbial species in a high-human-background matrix.



**Figure 3.** (A) Serial dilution of physiologically relevant levels of the ZymoBIOMICS Microbial Community Standard containing 8 bacteria and 2 yeast (Supplementary Table S1) to the estimated total microbial cells/mL concentration in plasma is completed before treatment with or without HostEL followed by TNA extraction and AmpRE library preparation. A total of 5 samples are batched and sequenced on the iSeq 100. (B) Relative contribution of microbial spike-ins towards non-human reads across different dilution factors of HostEL-treated and control samples mapping to the 10 organisms. The relative ratio of the organisms was preserved in the HostEL treated samples at all dilutions, while the control samples lost some organism reads at higher dilutions. (C) Fraction of reads that mapped to spiked-in microbes at different dilutions with or without HostEL treatment. HostEL-treated samples significantly increase the percentage of reads across all different dilution factors as compared to the non-treated samples. Detailed percentages of the reads are shown in Supplementary Table S2.

### 3.3. HostEL Enhances Sensitivity and Specificity of Low Depth Sequencing

To establish analytical sensitivity and specificity of our approach, we prepared and sequenced 11 ‘blank’ plasma samples and 4 samples each at the above-mentioned dilutions with and without HostEL treatment. Using a stringency metric that is based on the minimum number of read counts to call the presence of an organism to reduce the chance of false positive, we evaluated the sensitivity and specificity of the platform at detecting all 10 spiked-in species (10 species per sample). With HostEL treatment, we achieved a sensitivity of 96% and a specificity of 96%, while the values for the control were 82% and 97%, respectively (Figure 4A). Further breakdown of the sensitivity at different dilutions demonstrate the effect HostEL treatment has on the sensitivity of detection, with the lowest concentration registering 88% against 63% of the control sample (Figure 4B). In conclusion, the mNGS workflow shown here enables the detection of bacterial and fungal pathogens with low-depth sequencing.



**Figure 4.** 11 blank plasma and 4 repeats of each dilution of the 10 microbial standards (for a total of 40 expected spike-ins to be detected) were prepared and sequenced on the iSeq 100 with a high stringency cutoff to prevent false positives. Based on the 10 organisms spiked into the plasma, a diagnostic truth table was constructed (A) with and without HostEL treatment and further broken down into (B) detailed sensitivity at different dilutions.

### 3.4. Concordance with Banked Patient Plasma Samples Is >90%

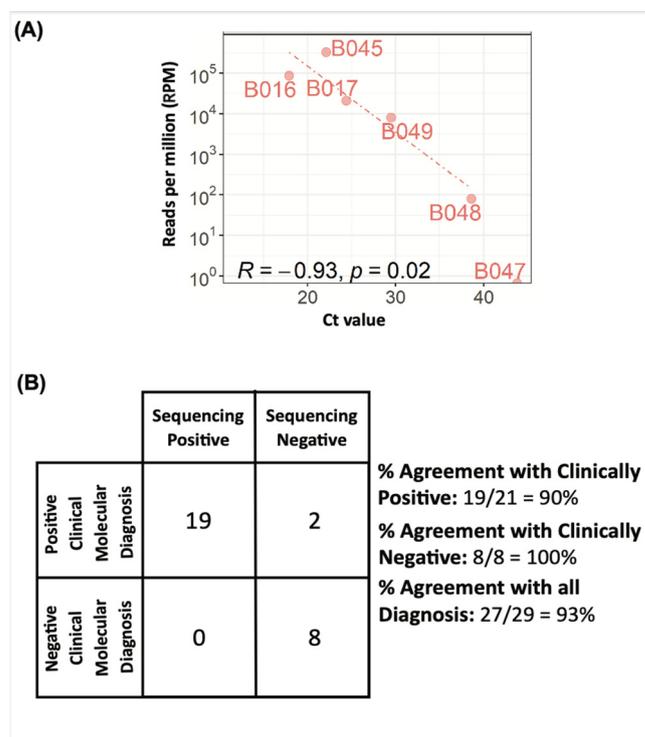
Clinical validation of our mNGS workflow was conducted on 42 patient plasma samples collected in compliance to IRB guidelines. These samples have previously been clinically diagnosed with nucleic acid tests to be positive (34 samples) or negative (8 samples) for pathogen infections (Table 1). To understand the correlation between sequencing workflow and the Ct values, and the LOD of the sequencing platforms, qPCR with targeted primers was conducted on 8 Hepatitis B samples diagnosed clinically with qPCR (Supplementary Table S3). Correlation plots were plotted with Ct values against the sequencing reads, and linear correlation was observed (Figure 5A). As we are targeting 100–500 k reads per sample, we can only confidently detect microbial sequences with an abundance of 1 in every 10 k reads. For samples with Ct~30, the relative abundance is about 1 in 10 k reads; thus, we have high confidence of our sensitivity up to a corresponding Ct value of 30 and below (Supplementary Table S4). When qPCR validation was also completed on all sequence-negative clinical samples as quality controls (Table 1) and some sequencing-positive samples as positive controls. qPCR was unable to detect any pathogen

in 3 out of 13 of these samples, while the rest had a Ct  $\geq$  33. The high Ct values for these samples may be due to the following reasons: (i) pathogens may be present in a low level, and (ii) sample integrity may have been impacted, or samples had degraded due to long-term storage.

**Table 1.** List of pathogens detected from clinical diagnosis, iSeq 100, and MiniSeq.

Sample	Clinical Diagnosis	iSeq 100 Detection	iSeq 100 Count	MiniSeq Detection	MiniSeq Count	qPCR Ct
B012	Hepatitis C Virus	Hepatitis C Virus	29	Hepatitis C Virus	72	23.8
B013	Hepatitis C Virus	Hepatitis C Virus	32	Hepatitis C Virus	66	27.2
B014	HIV-1	HIV-1	10	HIV-1	26	-
B015	HIV-1	HIV-1	2	HIV-1	3	-
B016	Hepatitis B Virus	Hepatitis B virus	4238	Hepatitis B virus	11,394	17.9
B017	Hepatitis B Virus	Hepatitis B virus	495	Hepatitis B virus	1226	24.4
B018	Hepatitis E Virus	-	0	-	0	34.3
B019	Hepatitis E Virus	-	0	-	0	37.1
B020	Cytomegalovirus	Cytomegalovirus	9	Cytomegalovirus	10	-
B021	Cytomegalovirus	Cytomegalovirus	22	Cytomegalovirus	100	-
B022	Varicella Zoster	-	0	-	0	33.0
B023	BK virus	-	0	-	0	32.0
B024	BK virus	-	0	-	0	33.0
B025	Epstein–Barr virus	Epstein–Barr virus	13	Epstein–Barr virus	2	-
B040	Dengue type 4	-	0	-	0	Neg
B041	Dengue type 2	Dengue virus	2	Dengue virus	3	-
B042	Dengue type 1	-	0	-	0	Neg
B043	Dengue type 3	Dengue virus	7	Dengue virus	18	-
B045	Hepatitis B Virus	Hepatitis B virus	7427	Hepatitis B virus	23,336	22.1
B046	Hepatitis B Virus	-	0	-	0	Neg
B047	Hepatitis B Virus	-	0	-	0	43.8
B048	Hepatitis B Virus	Hepatitis B virus	2	Hepatitis B virus	2	38.6
B049	Hepatitis B Virus	Hepatitis B virus	118	Hepatitis B virus	318	29.5
B050	Hepatitis B Virus	-	0	-	0	Neg
B051	Hepatitis C Virus	Hepatitis C Virus	17	Hepatitis C Virus	128	23.3
B052	Hepatitis C Virus	Hepatitis C Virus	6	Hepatitis C Virus	8	26.7
B053	Hepatitis C Virus	Hepatitis C Virus	2	Hepatitis C Virus	20	24.8
B054	Hepatitis C Virus	-	0	-	0	35.0
B056	Hepatitis C Virus	Hepatitis C Virus	4	-	0	32.9
B057	Hepatitis C Virus	-	0	-	0	35.9
B058	Hepatitis C Virus	-	0	-	0	34.1
B059	Hepatitis C Virus	Hepatitis C Virus	7	Hepatitis C Virus	2	28.8
B060	Hepatitis C Virus	Hepatitis C Virus	8	Hepatitis C Virus	2	27.9
B061	Hepatitis C Virus	-	0	-	0	30.1
B068	No Infection	-	0	-	0	-
B069	No Infection	-	0	-	0	-
B070	No Infection	-	0	-	0	-
B071	No Infection	-	0	-	0	-
B072	No Infection	-	0	-	0	-
B073	No Infection	-	0	-	0	-
B074	No Infection	-	0	-	0	-
B075	No Infection	-	0	-	0	-

Green: Clinically diagnosed pathogen was detected on iSeq 100/MiniSeq; Red: Clinically diagnosed pathogen was not detected on iSeq 100/MiniSeq; Blue: Ct value < 30; Yellow: Ct value > 30 or negative (undetectable).



**Figure 5.** Clinical validation of HostEL and AmpRE workflow. A total of 500  $\mu$ L each of 42 banked patient plasma samples of infected patients or control plasma was put through the workflow as before, with the EZ1 virus mini 2.0 kit rather than a manual extraction kit. A total of 7  $\mu$ L was processed by AmpRE and 5 samples were batched and sequenced on iSeq 100. **(A)** Correlation between sequencing reads and qPCR results for Hepatitis B-positive samples (Supplementary Table S4). We performed qPCR on 8 diagnosed Hepatitis B samples to check the correlation between sequencing reads and pathogen loads. The Ct values were plotted against the sequencing reads. Sequencing reads of both sequencing platforms correlate linearly with the Ct values. For samples with Ct  $\geq$  33, both sequencing platforms can only pick up less than 10 per million reads or have stochastic dropout of these pathogens (Table 1). **(B)** Analysis of clinical validation of HostEL and AmpRE workflow on plasma samples. Our cohort showed concordance results of 100% agreement between the workflow for all the positive clinical molecular diagnosis plasma samples below Ct 33. To ensure there is no false positive, we conducted the workflow on 8 negative clinical molecular diagnosed samples and obtained 100% agreement.

We further compare the percentage agreement of our workflow with the positive clinical molecular diagnosis data and after excluding the 13 samples that were deemed to be low quality (Ct  $\geq$  33 or undetectable) from the statistics (Figure 5B). We obtained 93% agreement with the clinical results. To ensure that there is no false positive in our workflow, 8 control plasma samples (B068–B075) clinical negative for pathogen were also analyzed with the workflow (Table 1). Our workflow shows 93% agreement on iSeq 100 on both positive and negative samples with Ct < 33 (Figure 5B). In rare cases, with Ct  $\geq$  33 where sample quality is low (B048), pathogen detection is still possible, although with a lower sensitivity (Table 1).

### 3.5. Both DNA and RNA Viruses Can Be Detected If Ct Is Less than 33

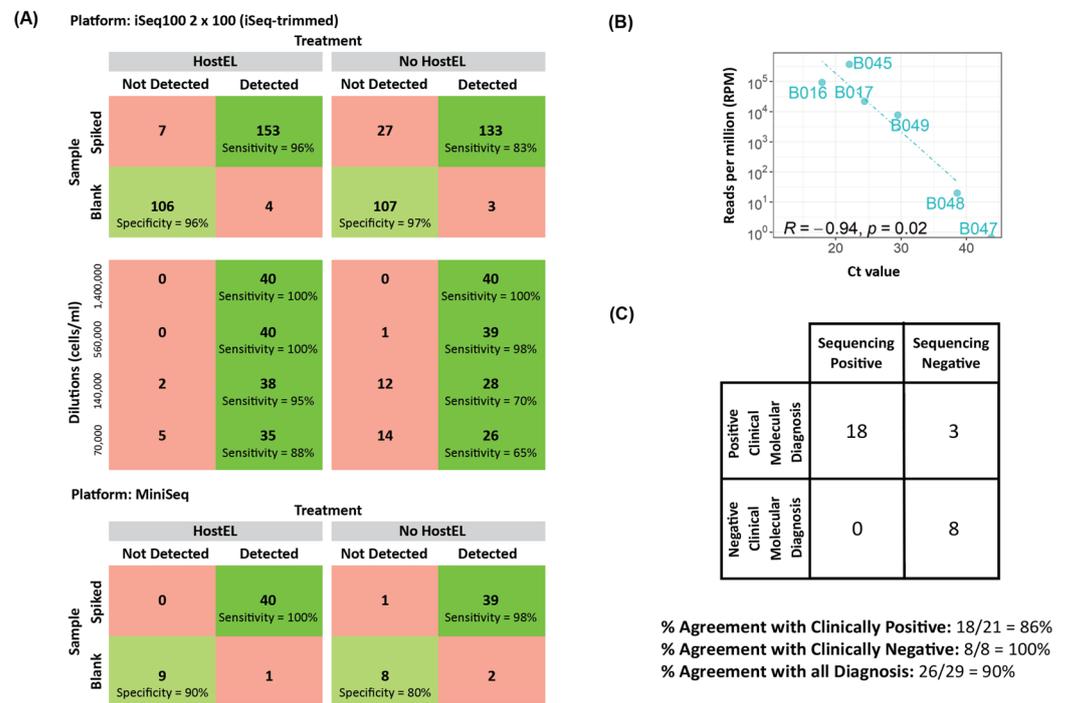
Both DNA viruses (Hepatitis B, CMV, and EBV) and RNA viruses (Hepatitis C, Dengue, HIV) were detected from the clinical samples (Table 1), suggesting that AmpRE library preparation kit enables the detection of both DNA and RNA pathogens. Unlike the analytical validation when the extracted TNA was eluted in 30  $\mu$ L, the clinical samples were eluted in 60  $\mu$ L of water based on clinical SOP. With only 7  $\mu$ L used as input into the library preparation, we were effectively sampling 58  $\mu$ L of plasma (500  $\mu$ L  $\times$  7  $\mu$ L/60  $\mu$ L) to detect

the pathogen. With a more aggressive concentration strategy during nucleic acid extraction, the sensitivity will likely increase.

### 3.6. Rapid Turnaround Can Be Achieved with Different Sequencing Platforms

Our workflow was tested on the iSeq 100 2 × 150 sequencing platform due to its ease of use, which would be a big factor in clinical deployment of a metagenomic sequencing solution. However, if implemented in its entirety, the whole workflow will deliver the results to the infectious disease specialist in 27 h (Figure 2C). In theory, with a sample submission cutoff at 11 am and samples loading at 5 pm, the report will be out at 2 pm the next day with our workflow. If the sequencing time was reduced to 13 h, then it is possible to deliver results by 8 am the next morning forward rounds. To simulate such a workflow, we trimmed the reads to 2 × 100 paired end to see if that affected the sensitivity and specificity of such a low-depth (<1 million reads per sample) approach to metagenomic sequencing ('iSeq-trimmed'). In addition, we evaluated the MiniSeq, which can deliver a single end 100 nucleotide read in 5 h, potentially shortening the entire workflow to 13 h, albeit with the trade-off of greater user involvement in setting up the MiniSeq run. To compare the low-depth sequencing, the read count was kept at 1 million reads per sample (MiniSeq: 20 samples per sequencing run; iSeq100: 5 samples per sequencing run).

The analytical validation using 4 (iSeq-trimmed) or 1 sample (MiniSeq) at each dilution level with negative controls obtained comparable sensitivity with the 19 h protocol for both the trimmed and MiniSeq run (Figure 6A). The MiniSeq, in particular, resulted in a higher sensitivity without HostEL treatment than the iSeq 100, albeit with a much smaller sample size. Although the sample size was small, it suggests the specificity is comparable for all 3 protocols (Supplementary Table S2). Thus, we can conclude that reducing the run time with shorter protocols seems to result in no loss of sensitivity.



**Figure 6.** Clinical options for turnaround time. We computationally trimmed the reads of the iSeq 100 to 2 × 100 to simulate a 13 h run time and also ran the same libraries batched with 20 samples per run on the MiniSeq. (A) Sensitivity and specificity performance on the spiked-in samples for the trimmed iSeq 100 and MiniSeq runs were similar to the full iSeq 100 run. Breakdown of the different dilution levels for iSeq-trimmed is also included. (B) Correlation between sequencing reads and qPCR results for Hepatitis B-positive samples (Supplementary Table S4) (C) MiniSeq diagnostic truth table for samples with Ct < 33.

Correlation plots of Ct values against the sequencing reads were also plotted for the MiniSeq, and linear correlation was also observed (Figure 6B, Supplementary Table S4). With the patient samples and a Ct 33 cutoff, MiniSeq performed equivalently to the iSeq 100 run (Figure 6C). Our results suggest that the HostEL and AmpRE workflow is compatible with all three sequencing setup and can deliver rapid unbiased metagenomic sequencing results with shallow sequencing.

#### 4. Discussion

Metagenomic next-generation sequencing (mNGS) has been used extensively in the detection of infectious diseases and is found to have certain advantages over conventional methods. It is an unbiased approach that can detect both known and unknown pathogens. Huge amount of effort has been made in key steps of mNGS, such as human background depletion, non-human sequence enrichment, library preparation, etc., to improve the diagnosis efficiency [8,25–32]. Nevertheless, challenges, such as cost and simplicity of mNGS method, still exist [3]. In this study, we addressed these challenges by establishing a workflow that uses HostEL and AmpRE kits for human background depletion and library preparation.

There are a few challenges involving in human background depletion. First, the soluble nucleases are hard to remove from the reactions. Second, it is hard to fine tune the preferential lysis. We developed the strategy, HostEL, for preferential lysis of human cells followed by enzymatic digestion of human background using enzyme-immobilized magnetic beads, leaving intact pathogens for total nucleic acids (TNA) extraction and sequencing. As only a single step of magnetic beads removal is needed to remove the enzymes without the need for enzyme deactivation, it can be applied to biofluids of as low as 500  $\mu$ L. This strategy allows easy integration into any clinical system as it is a simpler and shorter process. We validated HostEL in our analytical validation using ZymoBIOMICS Microbial Community Standard, consisting of 8 bacteria and 2 yeasts, as a spike-in control to mimic the physiological infection condition. As illustrated in our study, HostEL is able to deplete human nucleic acid without compromising on the composition of the pathogens. At 1.4 million cells/mL of plasma, HostEL is able to increase non-human reads by ~4-fold. When the microbial contribution was diluted to 70,000 cells/mL of plasma, a similar level to the physiological level of  $10^5$  copies/mL as reported in Darton et al., 2009 (30), the composition of the pathogens remained consistent. This appears to be superior to other depletion methodologies available in terms of maintaining relative composition [33]. In a separate study, non-human fraction of the sequencing reads in mNGS was found to be as low as less than 0.1% of the total reads [34]. Hence, with the use of HostEL, we are able to save cost on mNGS as the same amount of data is obtainable from less sequencing data. There is a question on the limit of detection and sensitivity of this approach as opposed to more sensitive qPCR, targeted sequencing approaches or even higher depth unbiased metagenomic sequencing, especially with the Ct33 detection cutoff for the banked clinical samples. While health economics calculations are not beyond the scope of this article, the availability of a first-line diagnostic for febrile neutropenia or fever of unknown origins that can detect an active infection of viruses, bacteria, or fungi at a comparatively low price point can save the patient weeks of diagnostic workup and provide timely targeted therapeutic intervention. This has the potential to reduce hospitalization duration and can be beneficial for antibiotic stewardship by only targeting the right organisms, potentially at the expense of missing residual infections that are not detectable in the blood.

Typical strategy for sequencing library construction for DNA and RNA is to have two separate library preparations. The AmpRE kit is to combine both library preparations into a single pot, which reduces the cost since the same reagents are used. As the kit is fast (under 4 h) and can be used for input as low as 0.7 ng of TNA, we are able to obtain adequate signal from 500  $\mu$ L of biofluids after treatment with HostEL. Sequencing of AmpRE generated libraries can be completed on demand in small batches on iSeq 100, allowing rapid detection of both DNA and RNA viruses, bacteria, and fungus if sufficient

samples are obtained daily. At 5 samples per run, we believe the sample load in a tertiary hospital should enable daily sequencing runs with the corresponding rapid speed of result delivery. Currently, AmpRE is optimized to produce libraries of certain sizes that are better for short-read sequencing. However, the strategy can be tuned for other sequencing platforms. Unlike most library preparation kits, AmpRE is also a pure additive process with minimal clean-up steps, thus reducing the time and effort needed. Nonetheless, there is risk of having background amplicons in the samples, and hence our workflow is optimized to prevent such occurrence.

In our clinical validation cohort of 42 plasma samples, which consist of DNA and RNA viruses and bacteria (Table 1), HostEL and AmpRE workflows are able to achieve high agreement with the clinical molecular diagnosis data. We performed the sequencing on two separate sequencing platforms, iSeq 100 and MiniSeq, and a simulated short iSeq 100 run and achieved concordance results of 90% agreement with the positive clinical molecular diagnosis samples. Our workflow also showed 100% agreement with the negative clinical molecular diagnosis samples, ruling out the high possibility of having false positive result. We further demonstrated that sequencing reads scale linearly with the pathogen load by conducting qPCR on clinically diagnosed Hepatitis B samples. Correlation plots of the Ct value against sequencing reads show that sequencing reads increase linearly with the decreasing Ct values, but the limit of detection of our sequencing platforms dropped when  $Ct > 30$ . Our initial results of our workflow are promising; however, they are limited to small diversity of pathogens and biofluid due to logistical limitation and timeline. We intend to extend our workflow to larger diversity of pathogens with different specimen types. We acknowledge that there are limitations to our workflow as it is not applicable to parasite infections, as parasites will not survive under selective lysis assay.

We also validated this workflow on iSeq 100 and MiniSeq to demonstrate the portability of the assay between different Illumina platforms at different sequencing depths and different run times. While sequencing workflows can theoretically be performed in under 24 h in ideal conditions, there are real-world limitations, such as operational hours of diagnostic laboratories, requirement for batching to minimize sequencing costs, analytical pipeline timing is frequently omitted from the calculations and cost considerations (3). To enable unbiased metagenomic sequencing to be a first-line diagnostic option, there is a need to deliver cost-competitive solutions that fit clinical laboratory setup. In the 3 workflows presented, the relatively short sample preparation time (6 h) coupled with competitive sequencing costs (Figure 2A), the option of different workflows and rapid sample analysis (2 h) enable hospitals to offer unbiased metagenomic sequencing as a diagnostic tool with next-day turnaround routinely. The option of the simplified iSeq 100 workflow can work well with labs with less skilled technicians at the cost of longer turnaround, while the MiniSeq offers speed at the cost of setup complexity. All three protocols have similar detection capabilities and perform well with clinical samples. Thus, we believe the workflow presented in this publication can potentially benefit the clinical community.

## 5. Conclusions

In conclusion, we established a mNGS workflow using HostEL and AmpRE, which considerably reduces time, cost, and effort by effectively depleting human background and amplifying both DNA and RNA into sequencing library in a single reaction, as illustrated in our analytical and clinical validations. We plan to explore fungal infection and also infections in a diversity of biofluids and anticipate to scale up the number of clinical samples in our validation to further support the efficiency of our strategies in different clinical settings.

## 6. Patents

Patents have been filed for the HostEL and AmpRE workflows.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/bioengineering10050520/s1>. Figure S1: Zymbiomics theoretical spike-in composition based on 16S rRNA amplicon sequencing, Table S1: ZymoBIOMICS Microbial Community Standard, Table S2: Detailed breakdown of percentage of reads of HostEL treated vs. no treatment enrichment on iSeq 100 and MiniSeq sequencing platforms, Table S3: List of targeted primers used in qPCR validation of clinical samples, Table S4: qPCR and iSeq 100/MiniSeq sequencing results of Hepatitis B plasma samples. References [35–41] are cited in the supplementary materials.

**Author Contributions:** Conceptualization, W.L.C.K., S.H. and Y.S.; methodology, W.L.C.K., S.H., Y.S., G.Y. and C.K.L.; sequencing, S.E.P., W.L.C.K. and Y.S.; HostEL treatment and TNA extraction, C.K.L., T.H.M.C.; qPCR, L.L. and S.E.P.; formal analysis, G.Y., Y.S. and W.L.C.K.; writing—original draft preparation, W.L.C.K. and K.W.K.; writing—review and editing, Y.S., W.Y.T.L., C.C., W.L.C.K. and K.W.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Agency for Science, Technology, and Research (Singapore) through Accelerate Venture Creation Pte Ltd. and A\*STAR spin-off Paths Diagnostics Pte Ltd. under a research collaboration agreement.

**Institutional Review Board Statement:** The studies involving human participants were reviewed and approved by NHG DSRB Study Reference Number: 2017/00632.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The analytical validation sequencing dataset for this study can be found in the figshare repository <https://doi.org/10.6084/m9.figshare.21694184>. The clinical validation dataset contains patient genomic data and is available for download upon request, pending ethics approval for each individual request.

**Acknowledgments:** The authors thank the patients from NUHS for contributing their samples, NUHS for collecting and processing the plasma samples, and Illumina, Inc. for providing access to the MiniSeq system for this study.

**Conflicts of Interest:** Y.S., W.K.L.C., S.H., W.Y.T.L. and C.C. are shareholders of Paths Diagnostics Pte Limited, which has licensed the HostEL and AmpRE IP from A\*STAR. Paths Diagnostics Pte Limited partially funded the entire study.

## References

1. Armitage, C. The High Burden of Infectious Disease. *Nature* **2021**, *598*, S9. [[CrossRef](#)]
2. Yang, S.; Rothman, R.E. PCR-Based Diagnostics for Infectious Diseases: Uses, Limitations, and Future Applications in Acute-Care Settings. *Lancet Infect. Dis.* **2004**, *4*, 337–348. [[CrossRef](#)]
3. Jia, X.; Hu, L.; Wu, M.; Ling, Y.; Wang, W.; Lu, H.; Yuan, Z.; Yi, Z.; Zhang, X. A Streamlined Clinical Metagenomic Sequencing Protocol for Rapid Pathogen Identification. *Sci. Rep.* **2021**, *11*, 4405. [[CrossRef](#)] [[PubMed](#)]
4. Miller, S.; Chiu, C. The Role of Metagenomics and Next-Generation Sequencing in Infectious Disease Diagnosis. *Clin. Chem.* **2022**, *68*, 115–124. [[CrossRef](#)] [[PubMed](#)]
5. Chiu, C.Y.; Miller, S.A. Clinical Metagenomics. *Nat. Rev. Genet.* **2019**, *20*, 341–355. [[CrossRef](#)]
6. Hong, D.K.; Blauwkamp, T.A.; Kertesz, M.; Bercovici, S.; Truong, C.; Banaei, N. Liquid Biopsy for Infectious Diseases: Sequencing of Cell-Free Plasma to Detect Pathogen DNA in Patients with Invasive Fungal Disease. *Diagn. Microbiol. Infect. Dis.* **2018**, *92*, 210–213. [[CrossRef](#)]
7. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; et al. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, *579*, 265–269. [[CrossRef](#)]
8. Miller, S.; Naccache, S.N.; Samayoa, E.; Messacar, K.; Arevalo, S.; Federman, S.; Stryke, D.; Pham, E.; Fung, B.; Bolosky, W.J.; et al. Laboratory Validation of a Clinical Metagenomic Sequencing Assay for Pathogen Detection in Cerebrospinal Fluid. *Genome Res.* **2019**, *29*, 831–842. [[CrossRef](#)]
9. Matranga, C.B.; Andersen, K.G.; Winnicki, S.; Busby, M.; Gladden, A.D.; Tewhey, R.; Stremlau, M.; Berlin, A.; Gire, S.K.; England, E.; et al. Enhanced Methods for Unbiased Deep Sequencing of Lassa and Ebola RNA Viruses from Clinical and Biological Samples. *Genome Biol.* **2014**, *15*, 519. [[CrossRef](#)]
10. Simner, P.J.; Miller, S.; Carroll, K.C. Understanding the Promises and Hurdles of Metagenomic Next-Generation Sequencing as a Diagnostic Tool for Infectious Diseases. *Clin. Infect. Dis.* **2018**, *66*, 778–788. [[CrossRef](#)]
11. Morales, M. The Next Big Thing? Next-Generation Sequencing of Microbial Cell-Free DNA Using the Karius Test. *Clin. Microbiol. Newsl.* **2021**, *43*, 69–79. [[CrossRef](#)]
12. Zaki, A.M.; van Boheemen, S.; Bestebroer, T.M.; Osterhaus, A.D.M.E.; Fouchier, R.A.M. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* **2012**, *367*, 1814–1820. [[CrossRef](#)]

13. Chiu, C.Y. Viral Pathogen Discovery. *Curr. Opin. Microbiol.* **2013**, *16*, 468–478. [[CrossRef](#)]
14. Salzberg, S.L.; Breitwieser, F.P.; Kumar, A.; Hao, H.; Burger, P.; Rodriguez, F.J.; Lim, M.; Quiñones-Hinojosa, A.; Gallia, G.L.; Tornheim, J.A.; et al. Next-Generation Sequencing in Neuropathologic Diagnosis of Infections of the Nervous System. *Neurol. Neuroimmunol. Neuroinflamm.* **2016**, *3*, e251. [[CrossRef](#)]
15. Cazanave, C.; Greenwood-Quaintance, K.E.; Hanssen, A.D.; Karau, M.J.; Schmidt, S.M.; Gomez Urena, E.O.; Mandrekar, J.N.; Osmon, D.R.; Lough, L.E.; Pritt, B.S.; et al. Rapid Molecular Microbiologic Diagnosis of Prosthetic Joint Infection. *J. Clin. Microbiol.* **2013**, *51*, 2280–2287. [[CrossRef](#)]
16. Naccache, S.N.; Peggs, K.S.; Mattes, F.M.; Phadke, R.; Garson, J.A.; Grant, P.; Samayoa, E.; Federman, S.; Miller, S.; Lunn, M.P.; et al. Diagnosis of Neuroinvasive Astrovirus Infection in an Immunocompromised Adult with Encephalitis by Unbiased Next-Generation Sequencing. *Clin. Infect. Dis.* **2015**, *60*, 919–923. [[CrossRef](#)]
17. Wilson, M.R.; Naccache, S.N.; Samayoa, E.; Biagtan, M.; Bashir, H.; Yu, G.; Salamat, S.M.; Somasekar, S.; Federman, S.; Miller, S.; et al. Actionable Diagnosis of Neuroleptospirosis by Next-Generation Sequencing. *N. Engl. J. Med.* **2014**, *370*, 2408–2417. [[CrossRef](#)]
18. Heravi, F.S.; Zakrzewski, M.; Vickery, K.; Hu, H. Host DNA Depletion Efficiency of Microbiome DNA Enrichment Methods in Infected Tissue Samples. *J. Microbiol. Methods* **2020**, *170*, 105856. [[CrossRef](#)]
19. Gu, W.; Deng, X.; Lee, M.; Sucu, Y.D.; Arevalo, S.; Stryke, D.; Federman, S.; Gopez, A.; Reyes, K.; Zorn, K.; et al. Rapid Pathogen Detection by Metagenomic Next-Generation Sequencing of Infected Body Fluids. *Nat. Med.* **2021**, *27*, 115–124. [[CrossRef](#)]
20. Thoendel, M.; Jeraldo, P.R.; Greenwood-Quaintance, K.E.; Yao, J.Z.; Chia, N.; Hanssen, A.D.; Abdel, M.P.; Patel, R. Comparison of Microbial DNA Enrichment Tools for Metagenomic Whole Genome Sequencing. *J. Microbiol. Methods* **2016**, *127*, 141–145. [[CrossRef](#)]
21. Darton, T.; Guiver, M.; Naylor, S.; Jack, D.L.; Kaczmarski, E.B.; Borrow, R.; Read, R.C. Severity of Meningococcal Disease Associated with Genomic Bacterial Load. *Clin. Infect. Dis.* **2009**, *48*, 587–594. [[CrossRef](#)] [[PubMed](#)]
22. Schmieder, R.; Edwards, R. Quality Control and Preprocessing of Metagenomic Datasets. *Bioinformatics* **2011**, *27*, 863–864. [[CrossRef](#)] [[PubMed](#)]
23. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [[CrossRef](#)] [[PubMed](#)]
24. Hasan, M.R.; Rawat, A.; Tang, P.; Jithesh, P.V.; Thomas, E.; Tan, R.; Tilley, P. Depletion of Human DNA in Spiked Clinical Specimens for Improvement of Sensitivity of Pathogen Detection by Next-Generation Sequencing. *J. Clin. Microbiol.* **2016**, *54*, 919–927. [[CrossRef](#)]
25. van Boheemen, S.; van Rijn, A.L.; Pappas, N.; Carbo, E.C.; Vorderman, R.H.; Sidorov, I.; vant Hof, P.J.; Mei, H.; Claas, E.C.; Kroes, A.C.; et al. Retrospective Validation of a Metagenomic Sequencing Protocol for Combined Detection of RNA and DNA Viruses Using Respiratory Samples from Pediatric Patients. *J. Mol. Diagn.* **2020**, *22*, 196–207. [[CrossRef](#)]
26. Lewandowski, K.; Xu, Y.; Pullan, S.T.; Lumley, S.F.; Foster, D.; Sanderson, N.; Vaughan, A.; Morgan, M.; Bright, N.; Kavanagh, J.; et al. Metagenomic Nanopore Sequencing of Influenza Virus Direct from Clinical Respiratory Samples. *J. Clin. Microbiol.* **2019**, *58*, e00963-19. [[CrossRef](#)]
27. Zinter, M.S.; Dvorak, C.C.; Mayday, M.Y.; Iwanaga, K.; Ly, N.P.; McGarry, M.E.; Church, G.D.; Faricy, L.E.; Rowan, C.M.; Hume, J.R.; et al. Pulmonary Metagenomic Sequencing Suggests Missed Infections in Immunocompromised Children. *Clin. Infect. Dis.* **2019**, *68*, 1847–1855. [[CrossRef](#)]
28. Allicock, O.M.; Guo, C.; Uhlemann, A.-C.; Whittier, S.; Chauhan, L.V.; Garcia, J.; Price, A.; Morse, S.S.; Mishra, N.; Briese, T.; et al. BacCapSeq: A Platform for Diagnosis and Characterization of Bacterial Infections. *mBio* **2018**, *9*, e02007-18. [[CrossRef](#)]
29. Wilson, M.R.; Sample, H.A.; Zorn, K.C.; Arevalo, S.; Yu, G.; Neuhaus, J.; Federman, S.; Stryke, D.; Briggs, B.; Langelier, C.; et al. Clinical Metagenomic Sequencing for Diagnosis of Meningitis and Encephalitis. *N. Engl. J. Med.* **2019**, *380*, 2327–2340. [[CrossRef](#)]
30. Deng, X.; Achari, A.; Federman, S.; Yu, G.; Somasekar, S.; Bartolo, I.; Yagi, S.; Mbala-Kingebeni, P.; Kapetshi, J.; Ahuka-Mundeke, S.; et al. Metagenomic Sequencing with Spiked Primer Enrichment for Viral Diagnostics and Genomic Surveillance. *Nat. Microbiol.* **2020**, *5*, 443–454. [[CrossRef](#)]
31. Petty, T.J.; Cordey, S.; Padioleau, I.; Docquier, M.; Turin, L.; Preynat-Seauve, O.; Zdobnov, E.M.; Kaiser, L. Comprehensive Human Virus Screening Using High-Throughput Sequencing with a User-Friendly Representation of Bioinformatics Analysis: A Pilot Study. *J. Clin. Microbiol.* **2014**, *52*, 3351–3361. [[CrossRef](#)]
32. Metsky, H.C.; Siddle, K.J.; Gladden-Young, A.; Qu, J.; Yang, D.K.; Brehio, P.; Goldfarb, A.; Piantadosi, A.; Wohl, S.; Carter, A.; et al. Capturing Sequence Diversity in Metagenomes with Comprehensive and Scalable Probe Design. *Nat. Biotechnol.* **2019**, *37*, 160–168. [[CrossRef](#)]
33. Nelson, M.T.; Pope, C.E.; Marsh, R.L.; Wolter, D.J.; Weiss, E.J.; Hager, K.R.; Vo, A.T.; Brittnacher, M.J.; Radey, M.C.; Hayden, H.S.; et al. Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. *Cell Rep.* **2019**, *26*, 2227–2240.e5. [[CrossRef](#)]
34. Metagenomic Sequencing for Infectious Diseases Diagnostics with Charles Chiu. Available online: <https://asm.org:443/Podcasts/MTM/Episodes/Metagenomic-Sequencing-for-Infectious-Diseases-Dia> (accessed on 17 March 2023).
35. ZymoBIOMICS™ Microbial Community Standard (D6300). Available online: [https://files.zymoresearch.com/protocols/\\_d6300\\_zymobiomics\\_microbial\\_community\\_standard.pdf](https://files.zymoresearch.com/protocols/_d6300_zymobiomics_microbial_community_standard.pdf) (accessed on 29 March 2022).

36. Pinto, G.G.; Poloni, J.A.T.; Paskulin, D.D.; Spuldaro, F.; Paris, F.; Barth, A.L.; Manfro, R.C.; Keitel, E.; Pasqualotto, A.C. Quantitative detection of BK virus in kidney transplant recipients: A prospective validation study. *J. Bras. Nefrol.* **2018**, *40*, 59–65. [[CrossRef](#)]
37. Alm, E.; Lesko, B.; Lindegren, G.; Ahlm, C.; Söderholm, S.; Falk, K.I.; Lagerqvist, N. Universal single-probe RT-PCR assay for diagnosis of dengue virus infections. *PLoS Negl. Trop. Dis.* **2014**, *8*, e3416. [[CrossRef](#)]
38. Liu, C.; Chang, L.; Jia, T.; Guo, F.; Zhang, L.; Ji, H.; Zhao, J.; Wang, L. Real-time PCR assays for hepatitis B virus DNA quantification may require two different targets. *Viol. J.* **2017**, *14*, 94. [[CrossRef](#)]
39. Warkad, S.D.; Nimse, S.B.; Song, K.S.; Kim, T. Development of a Method for Screening and Genotyping of HCV 1a, 1b, 2, 3, 4, and 6 Genotypes. *ACS Omega* **2020**, *5*, 10794–10799. [[CrossRef](#)]
40. Germer, J.J.; Ankoudinova, I.; Belousov, Y.S.; Mahoney, W.; Dong, C.; Meng, J.; Mandrekar, J.N.; Yao, J.D. Hepatitis E Virus (HEV) Detection and Quantification by a Real-Time Reverse Transcription-PCR Assay Calibrated to the World Health Organization Standard for HEV RNA. *J. Clin. Microbiol.* **2017**, *55*, 1478–1487. [[CrossRef](#)]
41. Harbecke, R.; Oxman, M.N.; Arnold, B.A.; Ip, C.; Johnson, G.R.; Levin, M.J.; Gelb, L.D.; Schmader, K.E.; Straus, S.E.; Wang, H.; et al. A real-time PCR assay to identify and discriminate among wild-type and vaccine strains of varicella-zoster virus and herpes simplex virus in clinical specimens, and comparison with the clinical diagnoses. *J. Med. Virol.* **2009**, *81*, 1310–1322. [[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.