Title

Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids

Journal

Nature Medicine, 27(1)

ISSN

1078-8956

Authors

Gu, Wei
Deng, Xianding
Lee, Marco
et al.

Publication Date

2021

DOI

10.1038/s41591-020-1105-z

Peer reviewed

HHS Public Access

Author manuscript

Nat Med. Author manuscript; available in PMC 2022 April 20.

Published in final edited form as:

Nat Med. 2021 January ; 27(1): 115–124. doi:10.1038/s41591-020-1105-z.

Rapid Pathogen Detection by Metagenomic Next-Generation Sequencing of Infected Body Fluids

Wei Gu[#1,2], **Xianding Deng**[#1,2], **Marco Lee**[#3], **Yasemin Sucu**[1,2], **Shaun Arevalo**[1,2], **Doug Stryke**[1,2], **Scot Federman**[1,2], **Allan Gopez**[1,2], **Kevin Reyes**[1,2], **Kelsey Zorn**[4], **Hannah Sample**[4], **Guixia Yu**[1,2], **Gurpreet Ishpuniani**[1,2], **Benjamin Briggs**[1,2], **Eric Chow**[4], **Amy Berger**[6], **Candace Wang**[1,2], **Steve Miller**[1,2], **Joseph DeRisi**[4,7], **Charles Chiu**[1,2,5,§]

[1]Department of Laboratory Medicine, University of California San Francisco, CA 94107, USA

[2]UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA 91407, USA

[3]Leeds Teaching Hospitals NHS Trust, Leeds, UK

[4]Department of Biochemistry and Biophysics, University of California San Francisco, CA 94107, USA

[5]Department of Medicine, Division of Infectious Diseases, University of California San Francisco, CA 94107, USA

[6]Department of Medicine, University of California San Francisco, CA 94107, USA

[7]Chan Zuckerberg Biohub, San Francisco, CA 94107, USA

[#] These authors contributed equally to this work.

Summary

We developed a metagenomic next-generation sequencing (mNGS) test using cell-free DNA to identify pathogens in acutely ill patients. Parallel nanopore and Illumina sequencing analyses of 87 body fluids (including abscess, joint, peritoneal, pleural, cerebrospinal, urine, and bronchoalveolar lavage fluids) using a hybrid barcoding system were performed. Compared to a composite gold standard based on clinical adjudication, culture, and PCR testing, sensitivities/specificities were 81-86%/91-95% for bacterial detection and 63-70%/92-96% for fungal detection. Real-time computational analysis enabled pathogen identification in a median of 50 minutes after starting nanopore sequencing. PCR and mNGS testing were concordant for 6 of 8 culture-negative body fluids, but mNGS detected 2 additional occult pathogens. Using mNGS, a pathogen was also detected in all 5 prospectively collected body fluids from patients with probable infection, but testing negative by all conventional microbiological assays. Rapid mNGS testing is a promising tool for clinical diagnosis of unknown infections from body fluids.

[§]Corresponding author: 185 Berry Street, Box #0134, UCSF China Basin, San Francisco, CA 94107, charles.chiu@ucsf.edu.

Keywords

Metagenomic next-generation sequencing (mNGS); cell-free DNA (cfDNA); infected body fluids; nanopore sequencing; Illumina sequencing; 16S PCR; real-time sequence analysis; SURPIrt (sequence-based ultra-rapid pathogen identification, real-time); digital PCR (dPCR)

Introduction

Early detection of causative microorganisms in patients with severe infections is critical to informing clinical interventions and administering appropriately targeted antibiotics (Messacar et al., 2017). Timely and accurate diagnosis, however, remains highly challenging for many hospitalized patients. As most infectious syndromes present with indistinguishable clinical manifestations, broad-based, multiplexed diagnostic tests are urgently needed but not yet available for the vast majority of potential pathogens. Some microorganisms are unculturable or fastidious (e.g. *Mycoplasma sp.*, *Bartonella sp.*), while others (e.g. mycobacteria and molds) can take weeks to grow and speciate (Rea et al., 2019). Accurate molecular detection by PCR provides an alternative diagnostic approach to culture, but is hypothesis-driven and thus requires *a priori* knowledge of the suspected pathogen(s). Although PCR tests targeting the conserved 16S ribosomal RNA (rRNA) gene ("16S PCR") and internal transcribed spacer (ITS) regions of bacteria and fungi, respectively, have been developed, (Janda and Abbott, 2007; Petti et al., 2005), concerns have been raised regarding detection sensitivity (Reuwer et al., 2019; Wilson et al., 2014, 2019). Failure or delay in diagnosing infections results in extended hospitalizations, readmissions, and increased mortality and morbidity (Glimåker et al., 2015; Kumar et al., 2006; Weiss et al., 2014). In addition, undiagnosed patients nearly always require empiric broad-spectrum therapy, with an increased risk of adverse side effects and antimicrobial drug resistance (Llor and Bjerrum, 2014).

Metagenomic next-generation sequencing (mNGS) has enabled unbiased detection of nearly all pathogens simultaneously from clinical samples (Chiu and Miller, 2019; Gu et al., 2019; Simner et al.). Previous work in this area has focused only on a specific, generally non-purulent body fluid type (Charalampous et al., 2019; Ivy et al., 2018; Langelier et al., 2018; Leo et al., 2017; Schmidt et al., 2017; Simner et al.; Thoendel et al., 2018; Wilson et al., 2018, 2019), and only a few studies to date have demonstrated clinical validation and/or utility (Blauwkamp et al., 2019; Schlaberg et al., 2017a, 2017b; Wilson et al., 2019). Methodology and sample types are also highly variable, making it difficult to evaluate the comparative performance across these studies. In particular, purulent fluids, which often suggest an infectious etiology, are challenging to analyze by mNGS due to high human host DNA background, which can decrease assay sensitivity (Miller et al., 2019). Methods exist to enrich for pathogen-specific reads from metagenomic data, such as differential lysis of human cells (Charalampous et al., 2019; Ivy et al., 2018; Thoendel et al., 2018), but the scope of detection using these approaches is typically restricted to bacteria and/or fungi. A generalizable mNGS test for simultaneous detection of all microbial pathogen types (viruses, bacteria, fungi, and parasites) and that is compatible across a variety of different body fluids, especially purulent body fluids (e.g. abscesses), has hitherto not been reported.

Most metagenomic studies have employed Illumina sequencing platforms, with sequencing run times exceeding 16 hours and overall sample-to-answer turnaround times of 48-72 hours. In contrast, nanopore sequencing (MinION sequencer by Oxford Nanopore Technologies) can rapidly detect microbes within minutes of starting sequencing and with a <6 hour turnaround time (Charalampous et al., 2019; Greninger et al., 2015; Schmidt et al., 2017). This rapid sequencing technology may be essential for time-sensitive diagnostics in acute, life-threatening infectious syndromes such as sepsis, for which mortality increases by 7% for every hour that appropriate treatment is delayed (Kumar et al., 2006). Nanopore sequencing has been extensively used for genomic surveillance of emerging viruses (Faria et al., 2016; Quick et al., 2016), but clinical metagenomic applications of the technology for pathogen detection have been limited to date (Chiu and Miller, 2019). One published study describes the use of a saponin-based differential lysis enrichment method for metagenomic nanopore sequencing-based detection of bacteria in respiratory infections with 96.6% sensitivity but only 41.7% specificity (Charalampous et al., 2019).

Here we describe a rapid, unbiased mNGS method to detect pathogen cell-free DNA (cfDNA) in virtually any body fluid type, ranging from low-cellularity spinal fluid to purulent fluids (e.g. abscesses) with high human host DNA content. A novel hybrid protocol, suitable for either nanopore or Illumina sequencing, is used to evaluate the diagnostic accuracy of mNGS testing on both platforms against traditional culture and 16S PCR based testing. We also present a prospective series of 5 cases with negative clinical testing from body fluids yet judged to be highly probable for infection, for which mNGS testing was able to identify the occult pathogen.

Results

Clinical Body Fluid Sample Collection

A total of 87 body fluid samples from 77 patients, including 19 abscess, 17 joint, 15 pleural, 12 peritoneal, 7 cerebrospinal, and 17 other body fluids (Table S1), were collected as residual samples after routine clinical testing in the microbiology laboratory. These 87 samples were used to evaluate the accuracy of mNGS testing by Illumina sequencing, with all but 2 samples evaluated in parallel by nanopore sequencing (Figure 1). Among these 87 samples, 70 were positive by culture (with pathogen(s) identified to genus/species level), 7 were positive by 16S PCR testing (PCR and sequencing of the 16S rRNA gene for bacterial detection), and 10 were negative controls from patients with alternative non-infectious diagnoses (e.g. cancer, trauma).

To demonstrate the diagnostic utility of mNGS testing for detecting pathogens in unknown cases, we also analyzed body fluids from an additional 5 patients with high clinical suspicion of infection but negative microbiological testing of the body fluid (Supplemental Case Series). These 5 cases were consecutively identified either by physician referral (n=1) or review of microbiology laboratory records (n=4). For 7 of 92 total body fluid samples, paired plasma samples collected within 1 day of the body fluid collection were also available for a head-by-head comparison of diagnostic yield.

Among the 82 total patients in the study, 74 (94%) were hospitalized, of whom 34 (43%) required intensive care unit (ICU) management and 34 (43%) met clinical criteria for sepsis (Raith et al., 2017), 26 (30%) were immunocompromised due to organ transplantation, recent chemotherapy, or drug-induced immunosuppression, and 57 (71%) were on antibiotics at the time of body fluid collection (Table 1). According to usual standard-of-care practices, bacterial cultures were obtained for all 92 body fluids, with 15 (16%) and 24 (26%) having additional cultures done for acid-fast bacilli (AFB) and fungi, respectively, in patients with a high index of suspicion for these infections.

Metagenomic Sequencing Analysis

We developed a streamlined hybrid protocol for mNGS testing that was cross-compatible on both nanopore and Illumina sequencing platforms, suitable for all body fluids, and automatable on liquid handlers. The amount of input DNA available for metagenomic sequencing varied over 6 logs from approximately 100 pg in low cellularity fluids such as cerebrospinal fluid (CSF) to 100 μg in purulent fluids. A dual-purpose barcode system for library multiplexing was developed for use on either nanopore or Illumina sequencers. The median read depths for Illumina and nanopore sequencing were 6.5M (IQR 2.4-7.8M, range 0.26-35M) and 1.1M (IQR 1.0-1.5M, range 0.29-6.7M), respectively. Metagenomic analysis for pathogen detection from Illumina data was performed using clinical-grade SURPI+ software (Miller et al., 2019; Wilson et al., 2019). Nanopore sequencing yielded 1 million reads per hour on average, with real-time data analysis performed using SURPI*rt* software, a new in-house developed bioinformatics pipeline for pathogen detection from metagenomic nanopore sequence data (Deng, et al., unpublished). After a 5-hour library preparation, nanopore sequencing detected pathogens in a median time of 50 minutes (interquartile range of 23 – 80 minutes) (Figure 1, Table S1), whereas the overall turnaround time for Illumina sequencing was ~24 hr.

The accuracy evaluation focused on the performance of mNGS relative to gold standard culture and/or 16S PCR testing for bacterial pathogen detection (Figure 1). Two reference standards were applied in the evaluation, a clinical gold standard consisting of available culture and 16S PCR-based testing results, followed by a composite gold standard that incorporated additional results from (i) orthogonal clinical testing of other sample types, (ii) confirmatory research-based digital PCR (dPCR) testing, and (iii) blinded clinical adjudication by an infectious disease specialist (CYC) and clinical pathologist (WG). Clinical samples were divided equally into training and validation sets, and receiver operator characteristic (ROC) and precision-recall curves for each set were generated relative to the initial clinical and composite gold standards (Figures 2A-B, S1B-D; Table S2).

Based on optimal thresholds using a normalized reads per million (nRPM) metric derived from training set data, the sensitivity and specificity of mNGS testing for bacterial detection based on the validation set were 81.5% (±5.2% SD) / 95.6% (± 4.4% SD) versus 85.6% (±4.8% SD) / 90.7% (± 3.4% SD) for Illumina and nanopore sequencing, respectively (Figure S2A-B). Nanopore sequencing yielded similar normalized read counts to Illumina sequencing ($p=0.54$) (Figure 2C and 2E). Excluding plasma, the performance of mNGS testing was comparable overall among different body fluid types (Figure 2D), with the

highest accuracy of detection from CSF and pleural fluids. Stratification based on semi-quantitation of culture colonies revealed relatively lower nRPM values for low-titer cultures that required growth in enrichment broth (Figure S1E).

Organisms that were considered false-positives consisted predominantly (4 of 4, 100% for Illumina sequencing and 7 of 9, 78% for nanopore sequencing) of those found at <5% of all sequenced microbial reads relative to high abundance pathogen(s) in co-infected samples (Table S3). These false-positive organisms consisted mainly of human respiratory, gastrointestinal, and/or skin flora, as listed in Table S3. When considering only the predominant organism, the specificity increased from 91-96% to 100% for both Illumina and nanopore sequencing.

Among the false-negative cases, species-specific reads from missed bacterial pathogens were still detected in 71% (10 of 14) and 20% (2 of 10) by Illumina and nanopore sequencing, respectively. False-negative cases were mostly from missed detections of *Staphylococcus aureus* (69.2%, 9 of 13 for Illumina sequencing and 90%, 9 of 10 for nanopore sequencing), of which 79% (11 of 14) were reported as rare or few isolates or cultured from enrichment broth only.

Due to the small number of positive fungal cases available (n=20), mNGS accuracy for fungal detection was assessed separately from bacteria, and the combined training and validation sets were used in the evaluation of accuracy. On average, fungal DNA was at a significantly lower concentration than bacterial DNA (Figure 2F). At the optimal Youden's index (normalized RPM = 0.1), sensitivity / specificity of detection were 70±7% SD / 96±4% SD and 63±9% SD / 100±0% SD for Illumina and nanopore sequencing, respectively (Figure S2C-D; Table S4). Among the 5 false-negative samples by both Illumina and nanopore sequencing, at least 1 read corresponding to the fungal pathogen was detected in 60% (3 of 5), suggesting that sensitivity could be boosted to ≥88% at greater depths of sequencing.

Case Series

To prospectively assess the potential clinical utility of body fluid mNGS for diagnosis of infection, we enrolled 5 unknown cases with a high clinical likelihood of infection even though culture and 16S PCR (when available) was negative. Likely causative pathogens (*Klebsiella aerogenes, Aspergillus fumigatus, Streptococcus pneumoniae, Streptococcus pyogenes*, and anaerobic gut flora) were identified in all five cases using mNGS (Supplemental Material, "Case Series").

Comparison of NGS with 16S PCR

We compared the performance of mNGS relative to bacterial 16S rRNA PCR in all 8 cases for which both tests were performed (Figure 3, Table S5). Clinical 16S PCR testing of body fluids and tissue is routinely ordered at our hospital for culture-negative cases with high clinical suspicion for an infectious etiology. Concordant results between mNGS testing and 16S PCR were obtained in 6 of 8 (75%) cases (Figure 3A).

The first of 2 discordant cases was a case of an immunocompromised child with necrotizing pneumonia (Figure 3B, Supplemental Material, “Case Vignettes”). Clinical 16S PCR testing showed a diagnosis of infection by an organism in the *Streptococcus mitis* group, whereas mNGS testing identified *Klebsiella pneumoniae* within 6 minutes after start of nanopore sequencing (Figure 3B). The finding of *Klebsiella pneumoniae* by mNGS was orthogonally validated as the correct result using 5 approaches: (i) dPCR of the DNA extract, (ii) dPCR of the sequencing library, (iii) Sanger sequencing of PCR clones from the DNA extract, (iv) mNGS (Illumina) sequencing of the contralateral fluid showing *Klebsiella pneumoniae*, and (v) dPCR of the contralateral fluid (Figure S3). In the second case, although culture and 16S PCR testing of the body fluid (CSF) were both negative, a subsequent culture from a neurosurgically removed deep brain stimulator (DBS) was positive for *Klebsiella aerogenes* (Figure 3B, Supplemental Material, “Case Vignettes”). mNGS testing of CSF was also positive for *Klebsiella aerogenes,* with detection within 10 min by nanopore sequencing. In addition to the culture of the DBS, the finding of *Klebsiella aerogenes* by mNGS was orthogonally validated with dPCR (Figure S4A).

To further investigate why these two discordant cases were negative by 16S PCR testing, we analyzed the length distribution of the detected pathogens using paired-end sequencing (Fan et al., 2010) (Supplemental Materials, Figure S5). The mean lengths of species-specific pathogen reads were 77 and 71 bp for these two cases, with nearly all lengths less than 300 bp. This provided a potential explanation for the negative 16S PCR results since 16S PCR typically requires lengths of at least 300 bp for accurate species-specific classification (Klindworth et al., 2013; Salipante et al., 2014; Wang and Qian, 2009).

Comparison of mNGS Pathogen Detection from Body Fluids versus Plasma

Seven patients in our study harboring a total of 9 pathogens had paired body fluid and plasma samples available for mNGS testing and comparison of diagnostic yield (Table S6). We found that the pathogen cfDNA burden based on nRPM was a median 160-fold higher (IQR 34-298) in the local body fluid than in plasma from the same patient (Figure 4, Table S6). Body fluid mNGS detected 8 of 9 (88.9%) pathogens, whereas plasma fluid mNGS only detected 3 of 9 (33.3%) (Figure 4). Among the 5 pathogens detected by body fluid and not plasma mNGS, 2 had been clinically missed by all microbiological testing, including culture, and were thus only retrospectively diagnosed by body fluid mNGS. The first pathogen was *Aspergillus fumigatus* in a respiratory fluid from a patient with a clinically probable fungal infection per international guidelines (Case S90). The second pathogen was *Bartonella henselae* and incidentally detected as an additional pathogen in an abscess from a patient with a cat scratch wound, a finding orthogonally confirmed by serology (Case S87). (Supplementary Material, “Clinical Vignettes”).

Detection of Anaerobic Bacteria and Viruses

Anaerobic bacteria were not included in the accuracy assessment as anaerobic culture was often not performed and detected anaerobic organisms were typically not speciated (e.g. reported as gastrointestinal flora). Only one sample in the accuracy study was reported as culture positive for an anaerobic bacterium (*Finegoldia magna* in S87), a bacterium that was successfully detected by mNGS testing (Table S7). Among the 22 additional anaerobic

bacteria detected in 10 body fluids, 14 were detected at low levels, representing <1% of all sequenced microbial reads. Of the remaining 8 anaerobes detected at higher levels, 7 were from cases of abdominal abscess and/or bowel perforation, in which gastrointestinal anaerobic bacteria would be expected to be abundant, whereas 1 was found in BAL from a patient with pneumonia of unclear etiology (Table S7, Supplementary Materials).

DNA viruses were also excluded in the accuracy assessment due to lack of routine clinical testing for viruses from body fluids. Using previously validated clinical mNGS thresholds of 3 non-overlapping reads for viral detection (Miller et al., 2019), 12 viruses were detected from the *Anelloviridae* (n=5), *Herpesviridae* (n=6, 2 herpes simplex virus 1, 2 cytomegalovirus, and 2 human herpesvirus 6B) , and *Adenoviridae* (n=1), families (Table S8). Four of the 5 patients with anellovirus infection were immunocompromised, consistent with the association of anelloviruses as non-pathogenic markers of active inflammation, especially in patients with weakened immune systems (De Vlaminck et al., 2013; Focosi et al., 2016). Among the 7 remaining viruses detected by mNGS, 4 were able to be verified as true positive by either prior clinical viral PCR testing of blood or virus-specific confirmatory PCR testing of available residual body fluid sample (Table S8).

Discussion

We demonstrate a rapid diagnostic method for untargeted metagenomic detection of DNA-based pathogens across a broad range of body fluid types. Notably, the same protocol is compatible with input cfDNA concentrations varying across 6 orders of magnitude (100 pg – 100 μg), which includes acellular fluids such as plasma and CSF as well as highly cellular purulent fluids such as abscesses. The protocol incorporates automated library preparation, a single hybrid barcoding system for Illumina or nanopore sequencing, and customized clinical-grade bioinformatics pipelines for metagenomic analysis. Importantly, we found the sensitivities and specificities for bacterial and fungal detection to be comparable between the two sequencing platforms. The utility of body fluid mNGS is highlighted by the clinically relevant detection of occult pathogens in 5 of 5 prospectively identified infectious cases without a microbiological diagnosis. Our mNGS methodology using pathogen cfDNA thus has the potential to be a clinically useful tool for rapid and accurate diagnosis of body fluid infections.

The metagenomic nanopore sequencing protocol described here can provide sample-to-result turnaround times of <6 hours. Timely and accurate identification of pathogens associated with infected body fluid compartments is essential because empiric antimicrobial treatment in the absence of an established diagnosis is often suboptimal, contributing to increased morbidity and mortality (Costales and Butler-Wu, 2018; Glimåker et al., 2015; Lucas et al., 2016; Mbaeyi et al., 2018; McGill et al., 2016; Singal et al., 2014). Nanopore sequencing, with its rapid turnaround time, may be particularly useful for routine identification of slow-growing pathogens (e.g. *Mycoplasma, Mycobacterium*, molds) as demonstrated in this study, or for critically ill patients necessitating an immediate diagnosis.

Our hybrid Illumina and nanopore sequencing protocol relies on finding pathogen-specific cfDNA sequences in body fluids supernatant. Other groups have used host depletion

methods such as differential lysis to enrich for intact pathogen and their genomic DNA from high host background samples such as respiratory or joint fluid prior to mNGS (Charalampous et al., 2019; Thoendel et al., 2018). However, as the supernatant along with pathogen cfDNA is removed during the differential lysis protocol, this enrichment method may not work well for low cellularity samples such as plasma and CSF. Differential lysis may also hinder detection of other DNA pathogens such as viruses and parasites. Furthermore, these methods involve multiple steps of lysis and centrifugation, thus prolonging turnaround times.

Recent studies have shown the utility of metagenomic sequencing for pathogen detection in sepsis and pneumonia (Blauwkamp et al., 2019; Charalampous et al., 2019). However, the test specificities of 63% and 42.7% may limit broad clinical application, since it may be challenging to evaluate the clinical significance of positive results. In contrast, an overall specificity of 91-96% was achieved using our protocol. Our calculated sensitivities for bacterial (81-86%) and fungal (63-70%) detection are slightly lower than previously reported (Blauwkamp et al., 2019; Charalampous et al., 2019). These other studies focused on patients with acute pneumonia or sepsis, whereas most of the patients in our study were already on broad-spectrum empiric antibiotics (71%) and were not septic (58%) at the time of body fluid collection, suggesting a lower pathogen burden. This is supported by the finding that 11 of 14 (79%) of the bacteria in false-negative cases were rare/few isolates or cultured from broth only, thus either signifying clinical contaminants or requiring higher sequencing depth for detection (Table S3, Figure S1E). Notably, subthreshold reads to the causative pathogen were still detected in 71% of the false-negative cases, and reporting of subthreshold results may still be clinically useful as previously described (Wilson et al., 2019). Finally, our reported sensitivity may be an underestimate because of varying read depths across samples and the use of residual body fluid samples that were left at room temperature for up to 24 hours before centrifugation. Cell lysis and release of human cfDNA in blood have been shown to increase human host background (Chan et al., 2005), which can decrease the sensitivity of metagenomic testing (Miller et al., 2019; Wilson et al., 2019). Future studies and clinical testing would optimally rely on body fluids that are immediately processed (e.g. <6 hours).

Recent studies have investigated the use of pathogen cfDNA analysis from blood for diagnosis of deep-seated infections (Abril et al., 2016; Blauwkamp et al., 2019; Grumaz et al., 2016; Long et al., 2016; Vlaminck et al., 2015). However, bacterial DNA is often present at low levels in blood, with a lower quartile of 5 bacterial genome copies per mL in patients with sepsis (Blauwkamp et al., 2019). Decreased sensitivity of metagenomic detection is especially concerning for patients without sepsis or infected by fungi (Armstrong et al., 2019), which comprised 43% (34 of 80) and 20% (17 of 80), respectively, of patients in the current study. We show here that in paired samples, causative pathogens could be identified for 8 of 9 (88.9%) cases in local body fluids versus only 3 of 9 (33.3%) in plasma, consistent with the observed 160-fold higher pathogen cfDNA burden in body fluids. Similarly, tumor cfDNA is higher in adjacent body fluids than in blood (De Mattos-Arruda et al., 2015; Dudley et al., 2019; Pan et al., 2015; Springer et al., 2018; Wang et al., 2018). The higher levels of pathogen cfDNA in the body can decrease sequencing depths necessary for detection, and thus lower the cost of testing. In addition, identification of a pathogen

from a body fluid can help localize the source of an infection, which is critical to guiding definitive management and treatment.

In comparing mNGS with 16S PCR, occult pathogens were detected by mNGS but missed by 16S PCR in 2 of 8 cases. False-negative 16S PCR results have been previously reported, and are generally attributed to suboptimal primer design or decreased assay sensitivity from background contamination (Wilson et al., 2014, 2019). Length analysis of mapped reads in two of our cases suggests another reason for false-negative 16S PCR results: the short read lengths associated with pathogen cfDNA in clinical body fluids samples (99.9% of the mapped reads were found to have lengths of <300 bp). Notably, size ranges for bacterial 16S PCR amplicons span at least 300-460 bp (Klindworth et al., 2013; Salipante et al., 2014; Wang and Qian, 2009), whereas those for fungal ITS PCR amplicons span 250-650 bp (Hoggard et al., 2018). Decreases in sensitivity due to fragmented cfDNA that are not amenable to long-read amplicon PCR have also been observed for detection of EBV virus in clinical samples (Chan et al., 2003).

Limitations of our study are as follows. First, not all body fluid types were available for metagenomic testing, including pericardial fluid, semen, stool, and saliva. Second, clinical samples had varying depths of sequencing, which may have contributed to the number of false-negative results. Third, orthogonal testing such as bacterial 16S PCR and fungal ITS PCR were not performed on all samples. Finally, further investigation with more samples will be needed to rigorously assess the clinical utility of metagenomic body fluid sequencing in unknown cases.

STAR METHODS

Ethics

Study protocols were reviewed and approved by the UCSF Institutional Review Board (IRB). The majority of residual samples (n=88) were collected in the clinical laboratory with a waiver of consent (IRB #10-01116). A subset of samples (n=4) were obtained from patients after consenting them for enrollment in a metagenomic sequencing study (IRB #15-18425, 17-22051). All experimental methods followed guidelines established by the Helsinki Declaration.

Sample selection and processing

All body fluid samples were obtained from patients at the University of California San Francisco (UCSF) hospitals and clinics. Body fluid samples were directly collected in sterile tubes or occasionally using swabs as part of routine clinical care and include abscess, joint, peritoneal, pleural, cerebrospinal, urine, bronchoalveolar lavage and other fluids (Table 1, S1). Swabs were stored in charcoal gel columns (Swab Transport Media Charcoal 220122, BD) and reconstituted in 0.5 mL of Universal Transport Media (350C, Copan Diagnostics, Murrieta, CA). This study only used residual body fluid samples after standard-of-care laboratory testing was performed. This included cultures for bacteria, fungi, and AFB that was done in-house at UCSF as well as clinical 16S rDNA PCR for bacterial detection performed by a reference laboratory at the University of Washington. Residual samples were

stored at 4°C and tested within 14 days of collection or centrifuged at 16,000 rcf for 10 minutes and the supernatant stored at −80°C until time of extraction.

Plasma samples were obtained by collecting blood from hospitalized patients as part of routine clinical testing into EDTA Plasma Preparation Tubes (BD) or standard EDTA Tubes (BD). The tubes were centrifuged (4000-6000 rcf for 10 minutes) within 6 hours, and plasma was isolated from the buffy coat and red cells. The plasma component was further aliquoted and centrifuged at 16,000 rcf for 10 minutes in microcentrifuge tubes. Plasma samples were stored at −80°C until the time of extraction.

DNA extraction

Samples were processed in a blinded fashion In a CLIA (Clinical Laboratory Improvement Amendments)-certified clinical microbiology laboratory with physically separate pre and post-PCR rooms. Cells were first removed through centrifugation to minimize host background. 400 μL of body fluid supernatant or plasma then underwent total nucleic acid extraction to 60 μL extract using the EZ1 Advanced XL BioRobot and EZ1 Virus Mini Kit v2.0 (QIAGEN) according to the manufacturer's instructions.

Library preparation and PCR amplification

Library preparation was performed using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), with the use of 25 μL of extracted DNA input and half of the reagent volumes suggested by the manufacturer's protocol. Briefly, extracted DNA was quantified on a NanoDrop spectrophotometer (ThermoFisher) and diluted to 10-100 ng of input as recommended by the manufacturer. Plasma or CSF DNA was not quantified or diluted as it could not be reliably detected using a spectrophotometer. The DNA was then end-repaired, ligated with the NEBNext Adapter (0.6 μM final concentration) to enrich for short-fragment pathogen DNA (100-800 bp) relative to residual human genomic DNA (>1 kb), and cleaned using AMPure beads. In addition to the initial manual preparation of 17 samples, an automated protocol using the epMotion 5075 liquid handler (Eppendorf) was used to process the remaining 75 samples, with 16-48 samples batch-processed per run.

PCR amplification was performed using a 40 μL mix consisting of adapter-ligated DNA, premixed custom index primers (Table S10) at 3 μM final concentration, and a quantitative PCR master mix (KAPA RT-kit, KK2702, Roche). DNA amplification was performed to saturation on a qPCR thermocycler (Lightcycler 480, Roche) using the following PCR conditions: initiation at 98°C x 45 s, then 24 cycles of 98°C x 15 s/63°C x 30 s/72°C x 90 s, and a final extension step of 72°C x 60 s. Ct values were continually monitored until the libraries were fully amplified to saturation of the fluorescent signal. Final DNA libraries were cleaned up using Ampure beads (Beckman) at a 0.9X volumetric ratio and eluted in 30 μl EB buffer (Qiagen).

Hybrid multiplex sequencing on Illumina and nanopore sequencing platforms

Multiplexing barcodes on the Illumina platform typically have the lengths of 8 bp flanking the sequence read on both ends, but they are not ideal for multiplexing samples being sequenced on a nanopore instrument (Oxford Nanopore Technologies) due to the higher

error rate of this platform. We designed a hybrid dual-platform barcode system that contained a distinct 37 nucleotide (nt) barcode on each side of sequencing adaptor (the first 8 nt of which were used for Illumina multiplexing), which enables the multiplexed DNA library to be sequenced on both Illumina and nanopore platforms.

Illumina sequencing

DNA libraries were pooled in equal volumes and the sequencing library pool was quantified using the Qubit fluorometer (ThermoFisher). Illumina sequencing was performed on the MiSeq (2x150 nt paired-end) initially and then HiSeq 1500/2500 instruments (140 nt single or 2x140 nt paired-end), according to the manufacturer's protocol.

Nanopore sequencing

Amplified DNA libraries were prepared for nanopore sequencing using the 1D library preparation kit (Oxford Nanopore Technologies) either manually or on an epMotion 5075 liquid handler biorobot (Eppendorf), with the processing of 8-16 samples per batch. The input DNA ranged from 200-1000 ng. The DNA was then sequenced using either R9.4 or R9.5 flow cells on a MinION or GridION X5 instrument (Oxford Nanopore Technologies). Up to four barcoded samples were sequentially loaded on the nanopore instrument for sequencing. Between each sample, flow cells were washed according to the manufacturer's instructions.

Bioinformatics analysis

Illumina sequencing data were analyzed for pathogens using the clinically validated SURPI+ (sequence based ultra-rapid pathogen identification) computational pipeline (Miller et al., 2019; Wilson et al., 2019). Nanopore sequencing data were analyzed using SURPI*rt* (SURPI "real-time", unpublished) software. Raw fast5 files were basecalled using Guppy software installed on the GridION in the real-time mode without polishing. The basecalled reads were run through in-house developed scripts for sample demultiplexing using a BLASTn E-value for a significance of 10^{-2}. After trimming adapters and removing low-quality and low-complexity sequences, the first 450 nt of the preprocessed read was partitioned into three 150 nt segments, followed by rapid low-stringency identification of candidate pathogen reads using SNAP alignment with editing distance of 50 (Zaharia et al., 2011). Candidate reads were then filtered and taxonomically classified as previously described (Miller et al., 2019). To perform real-time analysis, the SURPIrt pipeline was run every ~100k-200k nanopore reads for pathogen identification.

Computational algorithm for pathogen identification

We developed a pathogen identification algorithm applicable for both Illumina and nanopore datasets that maximized the performance of the training set. An initial gold standard database was manually tabulated based on pathogens detected in body fluids by culture and/or 16S PCR. Briefly, the algorithm (1) excluded artifactual and known background organisms, (2) calculated a normalized reads per million (nRPM) pathogen count, (3) filtered out taxonomically related organisms, and (4) defined criteria for pathogen detection.

(1) Excluding artifactual organisms—From our previous experience with metagenomic informatics, reproducible false-positive artifacts appeared for certain tropical parasitic worms (*Onchocerca flexuosa, Onchocerca ochengi, Enterobius vermicularis, Spirometra erinaceieuropaei, Haemonchus placei, Wuchereria bancrofti*) because DNA reads from these parasites share high-level sequence homology with the human genome. These organisms were thus excluded from the data analysis. In addition, we excluded the top two bacterial background organisms: *Achromobacter xylosoxidans* and *Cutibacterium (Propionibacterium) acnes*.

(2) Calculating a normalized RPM—We calculated a RPM for each organism to standardize the comparison across samples with uneven sequencing depths. For Illumina sequencing, the RPM was the number of pathogen reads divided by the number of preprocessed reads (reads remaining after adapter trimming, low-quality filtering, and low-complexity filtering). For nanopore sequencing, the RPM was the number of pathogen reads divided by the number of basecalled reads.

We then calculated a "normalized RPM" (nRPM) that normalized the RPM with respect to background based on the Ct value (to the nearest 0.5 increments) during the PCR amplification step of library preparation. As the average Ct value across all samples was 7, the nRPM was defined as nRPM = RPM / $2^{(Ct-7)}$ (Figure S2).

(3) Filtering out closely-related organisms—To mitigate cross-species misalignment for closely-related organisms, we penalized (subtracted) the RPM of organisms that share a genus or family designation. A penalty of 10% and 5% was used for genus and family respectively, based on the empirical maximization of specificity from the ROC curve of the training set. For example, if *Escherichia coli* had a normalized RPM of 100 and *Shigella sonnei* (same *Enterobacteriaceae* family) had a normalized RPM of 5, then all normalized RPM of *Shigella sonnei* would be subtracted by 100*5% = 5 RPM and become 0.

(4) Criteria for pathogen detection—We developed 3 criteria for pathogen detection, all of which needed to be fulfilled: (i) ranked within the top 5 organisms by nRPM, (ii) having a minimum number of reads for that organism identified (≥3 for bacteria and ≥1 for fungi), (iii) meeting an optimal nRPM threshold. Optimal nRPM thresholds (nRPM of 2.1 for Illumina sequencing or 0.1 for nanopore sequencing) were set to the maximum Youden's index as determined from the receiver operating characteristic (ROC) curve of the training set.

Statistical methods for Accuracy Assessment

To determine the accuracy, we used two gold standards: (i) an initial gold standard based on culture and 16S PCR results obtained through routine clinical care, and (ii) a composite gold standard based on orthogonal testing (e.g. digital PCR, serology) and clinical adjudication. The specific scoring algorithm is as follows (Table S12): True positives (TP) and false negatives (FN) were scored for each gold standard organism that was detected or not detected by mNGS respectively. True negatives (TN) and false positives (FP) were

scored for each sample that was negative for all other organisms besides the gold standard organism(s). Multiple FP(s) in a sample were counted as one FP overall.

Calculating p-values

We calculated p-values using a two-sided Welch's t-test. Values in the two independent data sets were considered statistically significant if $P < 0.05$.

Confidence intervals for the ROC curves

To evaluate the reliability of the validation set data (or aggregate training and validation set data for fungal organisms), we wrote a custom python script that took a 100% replaceable subset of all the samples at random and reanalyzed for 2000 bootstrap iterations. The resultant distribution was used to produce a 95% confidence interval (CI) for the sensitivity metric.

Orthogonal confirmation of mNGS results

Digital PCR was performed using the Biorad QX200 Droplet Digital PCR System. All primer and probe pairs were synthesized by Integrated DNA Technologies, Inc. and first validated using positive control organisms (Table S9). Genomic DNA from positive control organisms was purchased from ATCC and mechanically sheared (MiniTUBE, Covaris) to an average of 200-300 base pairs. For Sanger sequencing, DNA was first cloned into colonies using a TOPO TA Cloning Kit (ThermoFisher). Sanger sequencing of the clones was then performed at Elim Biopharmaceuticals, Inc. Sequencing traces were analyzed on Geneious software (version 10.2.3) and aligned to the National Center for Technology Information nucleotide (nt) database using BLAST (Altschul et al., 1990). Serology confirmation of the *Bartonella* case was performed by Quest Diagnostics.

Analysis of pathogen and human DNA lengths

For Illumina sequencing data, sequencing (FASTQ) files were trimmed for Illumina adapters with cutadapt (v 1.16). Trimmed files were aligned with BWA (v 0.7.12) to the hg38 human reference genome. For pathogen alignments, unaligned, human-depleted FASTQ reads were extracted using the bamtofastq function in the bedtools software package, followed by alignment to species-specific microbial reference genomes using BWA. An in-house developed Python program and Linux shell scripts were used to extract read lengths from resultant paired-end SAM files. For nanopore sequencing data, read lengths were directly extracted from SAM-formatted pathogen reads outputted from the SURPI*rt* pipeline. Histograms of the read lengths were plotted using python and the package matplotlib.

Data Availability

Metagenomic sequencing data (FASTQ files) after removal of human genomic reads have been deposited into the NCBI Sequence Read Archive (SRA) (accession numbers pending).

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to Dr. Charles Chiu (charles.chiu@ucsf.edu).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Helen Reyes, Agnes Chan, Danielle Ingebrigtsen, Lydia Dang, Walter Lorizio, Jessica Streithorst, Becky Fung, and the UCSF clinical laboratory staff for assistance and help with orthogonal testing. We thank Dr. Rong Da for assistance with sequencing runs. We thank Dr. Stephen Shiboski for biostatistics feedback. We thank members of the Chiu, Miller, and DeRisi lab for feedback and support.

References

Abril MK, Barnett AS, Wegermann K, Fountain E, Strand A, Heyman BM, Blough BA, Swaminathan AC, Sharma-Kuinkel B, Ruffin F, et al. (2016). Diagnosis of Capnocytophaga canimorsus Sepsis by Whole-Genome Next-Generation Sequencing. Open Forum Infect. Dis 3.

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol 215, 403–410. [PubMed: 2231712]

Armstrong AE, Rossoff J, Hollemon D, Hong DK, Muller WJ, and Chaudhury S (2019). Cell-free DNA next-generation sequencing successfully detects infectious pathogens in pediatric oncology and hematopoietic stem cell transplant patients at risk for invasive fungal disease. Pediatr. Blood Cancer 66, e27734. [PubMed: 30941906]

Blauwkamp TA, Thair S, Rosen MJ, Blair L, Lindner MS, Vilfan ID, Kawli T, Christians FC, Venkatasubrahmanyam S, Wall GD, et al. (2019). Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. Nat. Microbiol 4, 663. [PubMed: 30742071]

Burnham P, Dadhania D, Heyang M, Chen F, Westblade LF, Suthanthiran M, Lee JR, and Vlaminck ID (2018). Urinary cell-free DNA is a versatile analyte for monitoring infections of the urinary tract. Nat. Commun 9, 2412. [PubMed: 29925834]

Chan KCA, Zhang J, Chan ATC, Lei KIK, Leung S-F, Chan LYS, Chow KCK, and Lo YMD (2003). Molecular Characterization of Circulating EBV DNA in the Plasma of Nasopharyngeal Carcinoma and Lymphoma Patients. Cancer Res. 63, 2028–2032. [PubMed: 12727814]

Chan KCA, Yeung S-W, Lui W-B, Rainer TH, and Lo YMD (2005). Effects of Preanalytical Factors on the Molecular Size of Cell-Free DNA in Blood. Clin. Chem 51, 781–784. [PubMed: 15708950]

Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, Rae D, Grundy S, Turner DJ, Wain J, et al. (2019). Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat. Biotechnol 37, 783. [PubMed: 31235920]

Chiu CY, and Miller SA (2019). Clinical metagenomics. Nat. Rev. Genet 20, 341. [PubMed: 30918369]

Costales C, and Butler-Wu SM (2018). A Real Pain: Diagnostic Quandaries and Septic Arthritis. J. Clin. Microbiol 56, e01358–17. [PubMed: 29187561]

De Mattos-Arruda L, Mayor R, Ng CKY, Weigelt B, Martínez-Ricarte F, Torrejon D, Oliveira M, Arias A, Raventos C, Tang J, et al. (2015). Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. Nat. Commun 6, 8839. [PubMed: 26554728]

De Pauw B, Walsh TJ, Donnelly JP, Stevens DA, Edwards JE, Calandra T, Pappas PG, Maertens J, Lortholary O, Kauffman CA, et al. (2008). Revised Definitions of Invasive Fungal Disease from the European Organization for Research and Treatment of Cancer/Invasive Fungal Infections

Cooperative Group and the National Institute of Allergy and Infectious Diseases Mycoses Study Group (EORTC/MSG) Consensus Group. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am 46, 1813–1821.

De Vlaminck I, Khush KK, Strehl C, Kohli B, Luikart H, Neff NF, Okamoto J, Snyder TM, Cornfield DN, Nicolls MR, et al. (2013). Temporal Response of the Human Virome to Immunosuppression and Antiviral Therapy. Cell 155, 1178–1187. [PubMed: 24267896]

Dudley JC, Schroers-Martin J, Lazzareschi DV, Shi WY, Chen SB, Esfahani MS, Trivedi D, Chabon JJ, Chaudhuri AA, Stehr H, et al. (2019). Detection and surveillance of bladder cancer using urine tumor DNA. Cancer Discov. 9, 500–509. [PubMed: 30578357]

Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, and Quake SR (2010). Analysis of the Size Distributions of Fetal and Maternal Cell-Free DNA by Paired-End Sequencing. Clin. Chem 56, 1279–1286. [PubMed: 20558635]

Faria NR, Sabino EC, Nunes MRT, Alcantara LCJ, Loman NJ, and Pybus OG (2016). Mobile real-time surveillance of Zika virus in Brazil. Genome Med 8, 97. [PubMed: 27683027]

Focosi D, Antonelli G, Pistello M, and Maggi F (2016). Torquetenovirus: the human virome from bench to bedside. Clin. Microbiol. Infect 22, 589–593. [PubMed: 27093875]

Glimåker M, Johansson B, Grindborg Ö, Bottai M, Lindquist L, and Sjölin J (2015). Adult Bacterial Meningitis: Earlier Treatment and Improved Outcome Following Guideline Revision Promoting Prompt Lumbar Puncture. Clin. Infect. Dis 60, 1162–1169. [PubMed: 25663160]

Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, Stryke D, Bouquet J, Somasekar S, Linnen JM, et al. (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 7, 99. [PubMed: 26416663]

Grumaz S, Stevens P, Grumaz C, Decker SO, Weigand MA, Hofer S, Brenner T, von Haeseler A, and Sohn K (2016). Next-generation sequencing diagnostics of bacteremia in septic patients. Genome Med. 8, 73. [PubMed: 27368373]

Gu W, Miller S, and Chiu CY (2019). Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. Annu. Rev. Pathol. Mech. Dis 14.

Hoggard M, Vesty A, Wong G, Montgomery JM, Fourie C, Douglas RG, Biswas K, and Taylor MW (2018). Characterizing the Human Mycobiota: A Comparison of Small Subunit rRNA, ITS1, ITS2, and Large Subunit rRNA Genomic Targets. Front. Microbiol 9.

Ivy MI, Thoendel MJ, Jeraldo PR, Greenwood-Quaintance KE, Hanssen AD, Abdel MP, Chia N, Yao JZ, Tande AJ, Mandrekar JN, et al. (2018). Direct Detection and Identification of Prosthetic Joint Infection Pathogens in Synovial Fluid by Metagenomic Shotgun Sequencing. J. Clin. Microbiol 56.

Janda JM, and Abbott SL (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J. Clin. Microbiol 45, 2761–2764. [PubMed: 17626177]

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, and Glöckner FO (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41, e1. [PubMed: 22933715]

Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, Suppes R, Feinstein D, Zanotti S, Taiberg L, et al. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit. Care Med 34, 1589–1596. [PubMed: 16625125]

Langelier C, Kalantar KL, Moazed F, Wilson MR, Crawford ED, Deiss T, Belzer A, Bolourchi S, Caldera S, Fung M, et al. (2018). Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. Proc. Natl. Acad. Sci 115, E12353–E12362. [PubMed: 30482864]

Leo S, Gaïa N, Ruppé E, Emonet S, Girard M, Lazarevic V, and Schrenzel J (2017). Detection of Bacterial Pathogens from Broncho-Alveolar Lavage by Next-Generation Sequencing. Int. J. Mol. Sci 18.

Llor C, and Bjerrum L (2014). Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. Ther. Adv. Drug Saf 5, 229–241. [PubMed: 25436105]

Long Y, Zhang Y, Gong Y, Sun R, Su L, Lin X, Shen A, Zhou J, Caiji Z, Wang X, et al. (2016). Diagnosis of Sepsis with Cell-free DNA by Next-Generation Sequencing Technology in ICU Patients. Arch. Med. Res 47, 365–371. [PubMed: 27751370]

Lucas MJ, Brouwer MC, and van de Beek D (2016). Neurological sequelae of bacterial meningitis. J. Infect 73, 18–27. [PubMed: 27105658]

Mbaeyi SA, Joseph SJ, Blain A, Wang X, Hariri S, and MacNeil JR (2018). Meningococcal Disease Among College-Aged Young Adults: 2014–2016. Pediatrics e20182130.

McGill F, Heyderman RS, Panagiotou S, Tunkel AR, and Solomon T (2016). Acute bacterial meningitis in adults. Lancet Lond. Engl 388, 3036–3047.

Messacar K, Parker SK, Todd JK, and Dominguez SR (2017). Implementation of Rapid Molecular Infectious Disease Diagnostics: the Role of Diagnostic and Antimicrobial Stewardship. J. Clin. Microbiol 55, 715–723. [PubMed: 28031432]

Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, Stryke D, Pham E, Fung B, Bolosky WJ, et al. (2019). Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. Genome Res.

Moftakhar B, Patel YS, Poblete JE, and Myers JP (2018). Primary Peritonitis Due to Streptococcus pyogenes: Report of 2 Cases and Review of the 21st-Century Literature. Infect. Dis. Clin. Pract 26, 66.

Pan W, Gu W, Nagpal S, Gephart MH, and Quake SR (2015). Brain Tumor Mutations Detected in Cerebral Spinal Fluid. Clin. Chem 61, 514–522. [PubMed: 25605683]

Petti CA, Polage CR, and Schreckenberger P (2005). The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods. J. Clin. Microbiol 43, 6123–6125. [PubMed: 16333109]

Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, et al. (2016). Real-time, portable genome sequencing for Ebola surveillance. Nature 530, 228–232. [PubMed: 26840485]

Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, Pilcher DV, and Australian and New Zealand Intensive Care Society (ANZICS) Centre for Outcomes and Resource Evaluation (CORE) (2017). Prognostic Accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for In-Hospital Mortality Among Adults With Suspected Infection Admitted to the Intensive Care Unit. JAMA 317, 290–300. [PubMed: 28114553]

Rea B, Maisel JR, Glaser L, and Alby K (2019). Identification of Clinically Relevant Mycobacterial Species After Extended Incubation Times in the BACTEC MGIT System. Am. J. Clin. Pathol 151, 63–67. [PubMed: 30169764]

Reuwer AQ, van den Bijllaardt W, Murk JL, Buiting AGM, and Verweij JJ (2019). Added diagnostic value of broad-range 16S PCR on periprosthetic tissue and clinical specimens from other normally sterile body sites. J. Appl. Microbiol 126, 661–666. [PubMed: 30431696]

Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings LA, Sengupta DJ, Harkins TT, Cookson BT, and Hoffman NG (2014). Performance Comparison of Illumina and Ion Torrent Next-Generation Sequencing Platforms for 16S rRNA-Based Bacterial Community Profiling. Appl. Environ. Microbiol 80, 7583–7591. [PubMed: 25261520]

Schlaberg R, Chiu CY, Miller S, Procop GW, and Weinstock G (2017a). Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. Arch. Pathol. Lab. Med 141, 776–786. [PubMed: 28169558]

Schlaberg R, Queen K, Simmon K, Tardif K, Stockmann C, Flygare S, Kennedy B, Voelkerding K, Bramley A, Zhang J, et al. (2017b). Viral Pathogen Detection by Metagenomics and Pan-Viral Group Polymerase Chain Reaction in Children With Pneumonia Lacking Identifiable Etiology. J. Infect. Dis 215, 1407–1415. [PubMed: 28368491]

Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, Khan AM, Woodford N, Saunders NJ, Wain J, et al. (2017). Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. J. Antimicrob. Chemother 72, 104–114. [PubMed: 27667325]

Simner PJ, Miller S, and Carroll KC Understanding the Promises and Hurdles of Metagenomic Next-Generation Sequencing as a Diagnostic Tool for Infectious Diseases. Clin. Infect. Dis

Singal AK, Salameh H, and Kamath PS (2014). Prevalence and in-hospital mortality trends of infections among patients with cirrhosis: a nationwide study of hospitalised patients in the United States. Aliment. Pharmacol. Ther 40, 105–112. [PubMed: 24832591]

Springer SU, Chen C-H, Rodriguez Pena MDC, Li L, Douville C, Wang Y, Cohen JD, Taheri D, Silliman N, Schaefer J, et al. (2018). Non-invasive detection of urothelial cancer through the analysis of driver gene mutations and aneuploidy. ELife 7, e32143. [PubMed: 29557778]

Thoendel MJ, Jeraldo PR, Greenwood-Quaintance KE, Yao JZ, Chia N, Hanssen AD, Abdel MP, and Patel R (2018). Identification of Prosthetic Joint Infection Pathogens Using a Shotgun Metagenomics Approach. Clin. Infect. Dis 67, 1333–1338. [PubMed: 29648630]

Vlaminck ID, Martin L, Kertesz M, Patel K, Kowarsky M, Strehl C, Cohen G, Luikart H, Neff NF, Okamoto J, et al. (2015). Noninvasive monitoring of infection and rejection after lung transplantation. Proc. Natl. Acad. Sci 112, 13336–13341. [PubMed: 26460048]

Wang Y, and Qian P-Y (2009). Conservative Fragments in Bacterial 16S rRNA Genes and Primer Design for 16S Ribosomal DNA Amplicons in Metagenomic Studies. PLOS ONE 4, e7401. [PubMed: 19816594]

Wang Y, Li L, Douville C, Cohen JD, Yen T-T, Kinde I, Sundfelt K, Kjær SK, Hruban RH, Shih I-M, et al. (2018). Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers. Sci. Transl. Med 10, eaap8793. [PubMed: 29563323]

Weiss S, Fitzgerald J, Balamuth F, Alpern E, Lavelle J, Chilutti M, Grundmeier R, Nadkarni V, and Thomas N (2014). Delayed Antimicrobial Therapy Increases Mortality and Organ Dysfunction Duration in Pediatric Sepsis*. Crit. Care Med 42, 2409–2417. [PubMed: 25148597]

Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S, Federman S, Miller S, et al. (2014). Actionable Diagnosis of Neuroleptospirosis by Next-Generation Sequencing. N. Engl. J. Med 370, 2408–2417. [PubMed: 24896819]

Wilson MR, O'Donovan BD, Gelfand JM, Sample HA, Chow FC, Betjemann JP, Shah MP, Richie MB, Gorman MP, Hajj-Ali RA, et al. (2018). Chronic Meningitis Investigated via Metagenomic Next-Generation Sequencing. JAMA Neurol.

Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, Federman S, Stryke D, Briggs B, Langelier C, et al. (2019). Clinical Metagenomic Sequencing for Diagnosis of Meningitis and Encephalitis. N. Engl. J. Med 380, 2327–2340. [PubMed: 31189036]

Zaharia M, Bolosky WJ, Curtis K, Fox A, Patterson D, Shenker S, Stoica I, Karp RM, and Sittler T (2011). Faster and More Accurate Sequence Alignment with SNAP. ArXiv11115572 Cs Q-Bio.

Zinter MS, Mayday MY, Ryckman KK, Jelliffe-Pawlowski LL, and DeRisi JL (2019). Towards precision quantification of contamination in metagenomic sequencing experiments. Microbiome 7.

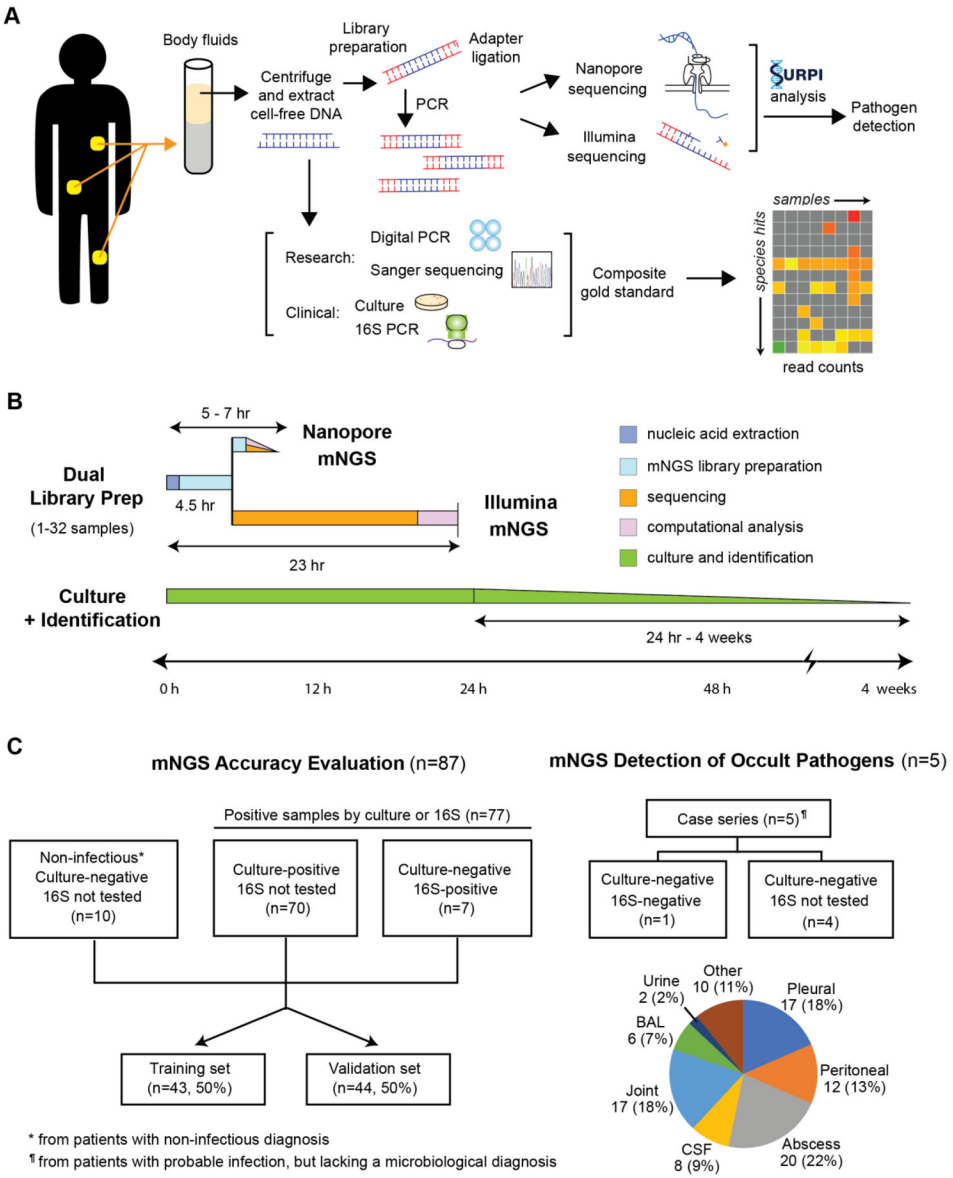

Figure 1.

(A) Schematic of mNGS body fluid analysis workflow. The composite gold standard consisted of aggregated available results from bacterial and fungal cultures, 16S PCR, confirmatory digital PCR and Sanger sequencing, and clinical adjudication. **(B) Timing for mNGS testing relative to culture.** Whereas culture-based pathogen identification can take days to weeks, mNGS testing has a 5-24 hr overall turnaround time. (**C) Analysis workflow for the 92 total body fluid samples in the study.** 87 total samples were included in the accuracy assessment and 5 samples collected from patients with a clinical diagnosis of infection but negative microbiological testing was included for prospective mNGS analysis. The pie chart shows the sample type of all 92 body fluids. **Other body fluids are: Vitreous Fluid, Perihepatic Fluid, Surgical Swab, Subgaleal Fluid, Heel Fluid Swab, Peri-graft Fluid Swab, Anterior Mediastinal Fluid, Chest Fluid, Chest Mass Fluid, Wound Swab.

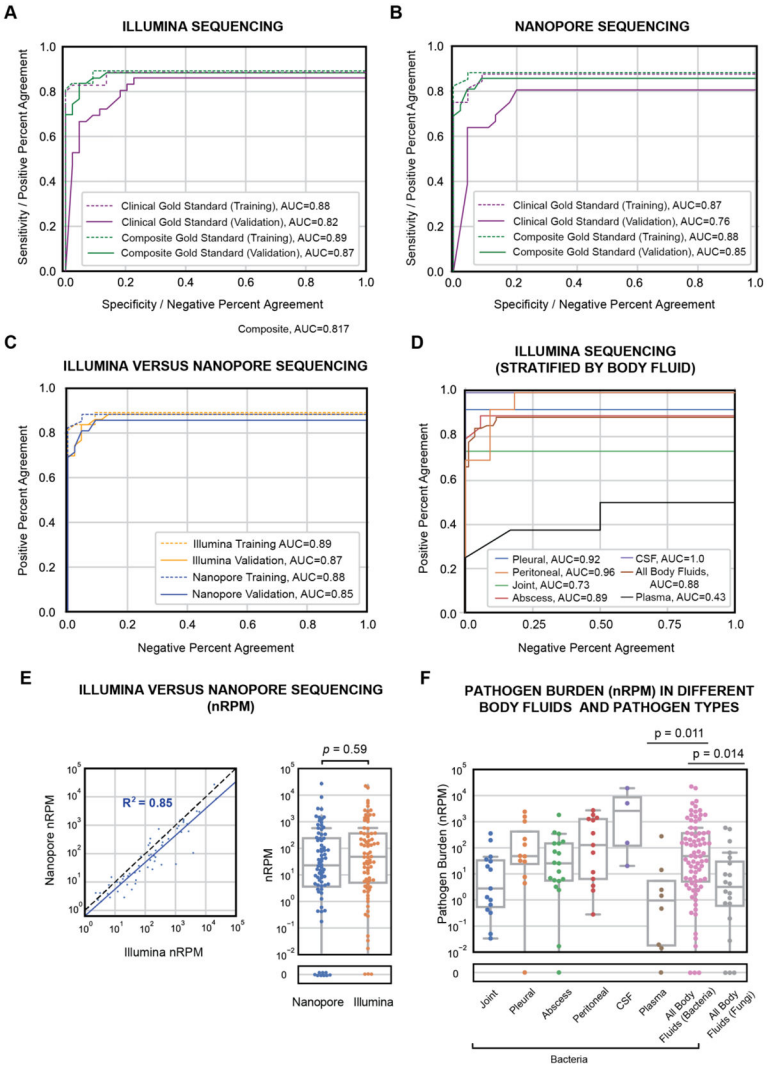

Figure 2. Accuracy of mNGS testing accuracy and relative pathogen burden in body fluid samples.

(A) ROC curves of Illumina mNGS performance. Plotted are mNGS test sensitivities and specificities at normalized RPM (nRPM) threshold values ranging from 0.1 to 100. ROC curves of Illumina training and validation dataset performance are plotted relative to clinical and composite gold standards. In all panels, the displayed n represents counts used for the contingency table as described in the methods section and shown in Table S11. **(B) ROC curves of nanopore mNGS performance. (C) ROC curves comparing Illumina and nanopore performance.** ROC curves of training and validation dataset performance are plotted relative to composite gold standard testing alone. **(D) ROC curves stratified by body fluid type.** Plotted is the performance of the combined Illumina training and validation datasets relative to composite gold standard testing. Plasma is not counted as a body fluid in panels D and F, but is plotted as a separate set. **(E) Comparison of Illumina with nanopore sequencing.** The yield of pathogen-specific reads based on a normalized RPM (nRPM) metric is linearly correlated and comparable between nanopore and Illumina sequencing. **(F) Relative pathogen burden in body fluids, stratified by body fluid and organism**

type. The burden of pathogen cfDNA in body fluid samples is estimated using calculated nRPM values. Based on Illumina data, bacterial cfDNA in plasma was significantly lower on average than in local body fluids (p=0.011), and pathogen cfDNA in body fluids was significantly higher for bacteria than fungi (p=0.014).

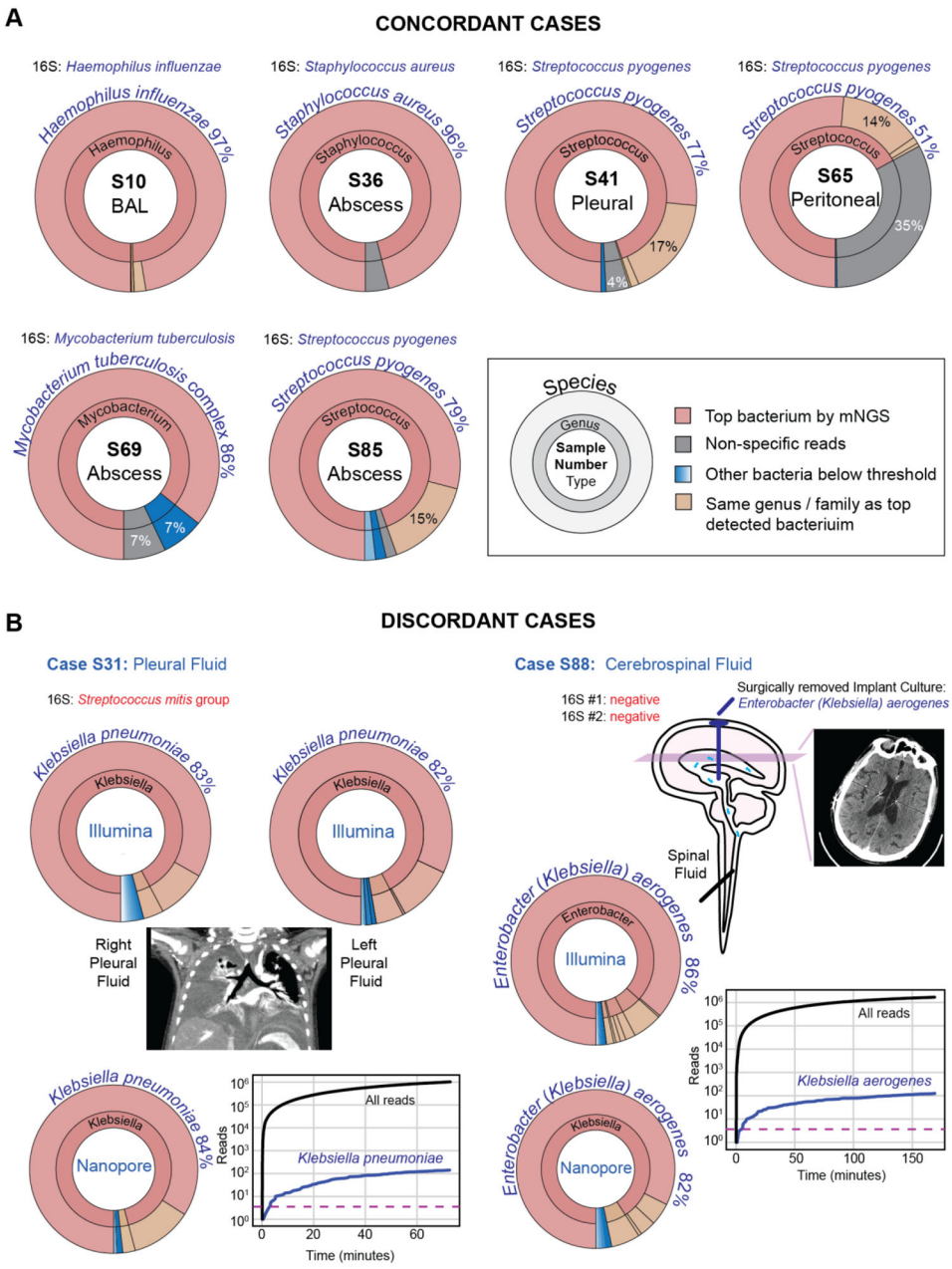

Figure 3. Comparison of mNGS with 16S PCR for 8 culture-negative body fluids.

Krona plots depict genus and species levels of all sequence-matched bacterial reads. **(A) Concordant cases (n=6). (B) Discordant cases (n=2).** In Case S31 (left), pleural fluids from a pediatric, immunosuppressed patient with necrotizing pneumonia (chest x-ray) were negative by culture. 16S PCR was positive for *Streptococcus mitis* group, an organism that would rarely be the cause of a necrotizing pneumonia. In contrast, mNGS sequencing revealed 83% of the bacterial reads to be specific for *Klebsiella pneumoniae* and no reads detected from *Streptococcus mitis*, a result that was orthogonally confirmed. In case S88 (right), CSF obtained before removal of an infected deep brain stimulator (DBS) was negative by culture and 16S PCR. However, culture of the surgically extracted DBS was

positive for *Klebsiella aerogenes*, as was mNGS sequencing of the CSF (86% of the sequence-matched bacterial reads were specific for *Klebsiella aerogenes*). CT imaging of the brain revealed that the infected DBS was located upstream of the lumbar puncture site (drawing and axial slice), explaining the presence of trace pathogen cfDNA in CSF that was detectable by mNGS. For both cases S31 and S88, nanopore sequencing was able to detect *Klebsiella pneumoniae* and *Klebsiella aerogenes* within 3 minutes after start of sequencing, respectively (xy scatter plots with the dotted line showing the 3-read threshold). Abbreviations: BAL, bronchoalveolar lavage fluid; CT, computed tomography).

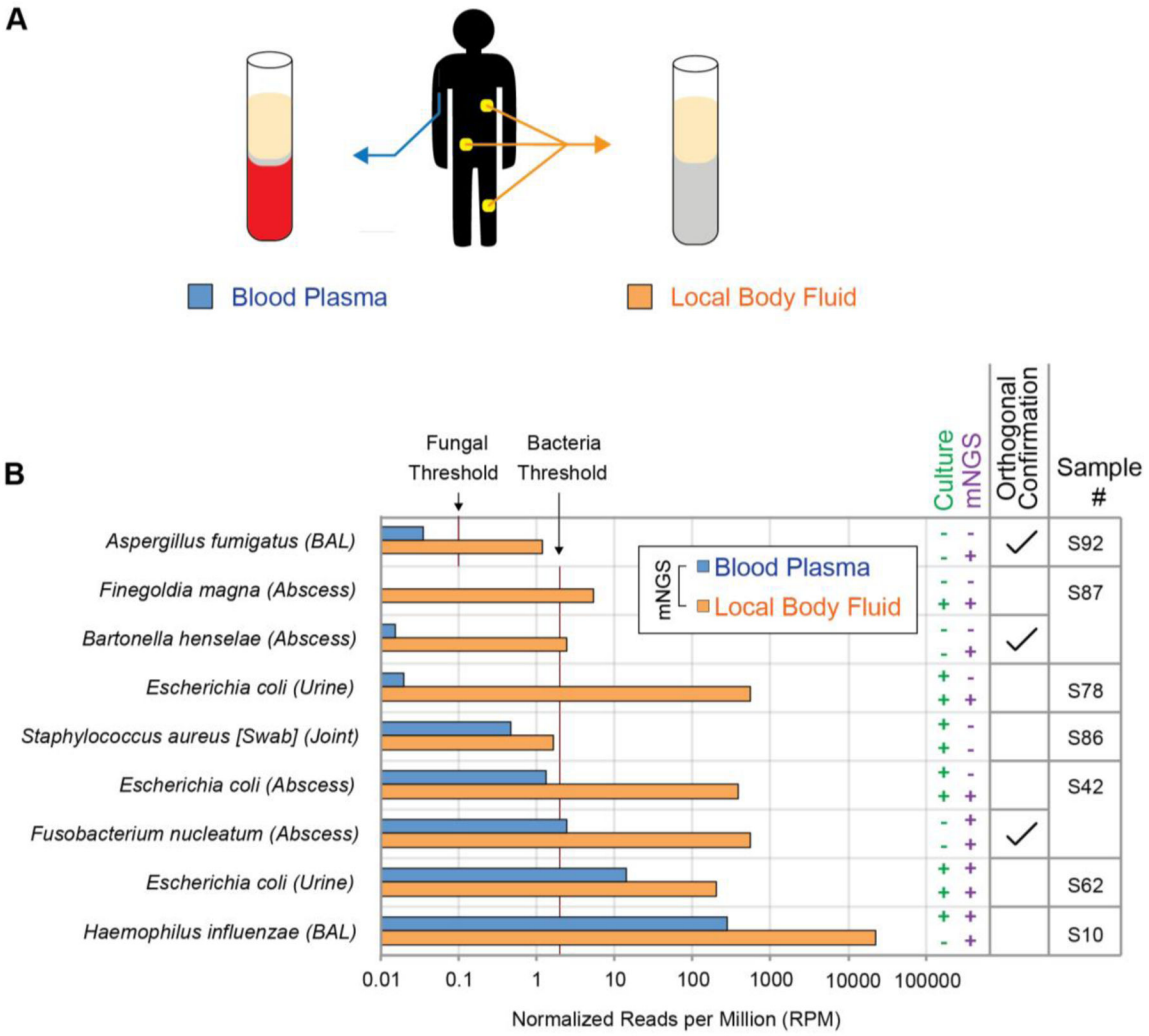

Figure 4. Comparison of relative pathogen burden in paired body fluid and plasma samples.

(A) Schematic showing concurrent collection of blood plasma and body fluid samples from the same patient. (B) Bar plot of the normalized RPM corresponding to 9 organisms in paired body fluid and plasma samples from 7 patients. The vertical lines show the thresholds used for a positive bacterial or fungal detection. The checkboxes denote organisms that were not identified by conventional microbiological testing (culture and/or 16S PCR) but that were orthogonally confirmed by dPCR, serology, and/or clinical adjudication (Supplemental Materials, “Case Vignettes”).

Table 1:

Patient and Sample Characteristics

Patient Demographics (n = 80) *	
Age - years	
Median (interquartile range)	53 (30-62)
Range	(1-89)
Gender	
Female – no. (%)	35 (44%)
Male – no. (%)	45 (56%)
Hospitalization	
Patients Total – no. (%)	79 (100%)
In hospital	74 (94%)
In intensive care unit	34 (43%)
Days Hospitalized – no. (IQR)	15 (7-28)
30 day mortality – no. (%)	4 (5%)
Immunocompromised – no. (%)	26 (30%)
Presumed Illness – no. (%)	
Septic arthritis	14 (18%)
Respiratory infection	21 (26%)
Gastrointestinal abscess	12 (15%)
Soft Tissue abscess	11 (14%)
Peritonitis	10 (13%)
CNS infection	9 (11%)
Urinary Tract Infection	2 (2%)
Eye infection	1 (1%)
Sample Characteristics (n = 92)	
Sample Type – no. (%)	
Abscess	20 (22%)
Cerebrospinal Fluid	8 (9%)
Joint Fluid	17 (18%)
Peritoneal Fluid	12 (13%)
Pleural Fluid	17 (18%)
Bronchoalveolar Lavage	6 (7%)
Urine	2 (2%)
Other **	10 (11%)
WBC count of body fluid – median 10^6/L (interquartile range)	8475 (686-64,625)
Range – 10^6/L	5-328,000

Time to Final Culture Result – median days (interquartile range)	4.8 (3.8-13.9)
Range – days	1.4-31.0
Organism cultured – no. (%)	
Staphylococcus aureus	27 (26%)
Streptococcus spp	11 (11%)
Enterococcus spp	7 (7%)
Gram Negative Rods	22 (21%)
Fungi	19 (18%)
Other	7 (7%)
Negative	11 (11%)

* There was a total of 82 patients in the study and 2 patients were deidentified prior to obtaining patient and sample characteristics.

** Vitreous Fluid, Perihepatic Fluid, Surgical Swab, Subgaleal Fluid, Heel Fluid Swab, Peri-graft Fluid Swab, Anterior Mediastinal Fluid, Chest Fluid, Chest Mass Fluid, Wound Swab.

Table 2:

Case series: Cases that are culture (+/− 16S) negative, but with a clinical diagnosis of infection or highly suspected of infection. Further details are in the clinical vignettes of the supplementary materials.

Case	Sample Type	Presentation	Species detected on mNGS	Culture Results	16S PCR Results	Confirmatory Testing
S88	CSF	Encephalopathy without known cause; has a brain implant	*Klebsiella aerogenes*	negative	negative	Same organism grown in culture from surgically removed brain implant. Concordant dPCR. Matches clinical context.
S89	Retrouterine fluid	Abdominal fluid collection and elevated white blood cell count; a history of abdominal surgery	(multiple GI anaerobes) *	negative	n/a	n/a (determined by clinical context)
S90	Pleural	Fever, cough, bacteremia, loculated pleural effusion	*Streptococcus pneumoniae*	negative	n/a	Concordant positive blood culture matches the clinical context
S91	Pleural	Fever, bacteremia, pneumonia, pleural effusion	*Streptococcus pyogenes*	negative	n/a	Concordant positive blood culture matches the clinical context
S92	BAL	Pulmonary nodules post-chemotherapy	*Aspergillus fumigatus*	negative	n/a	Probable invasive aspergillosis per international guidelines matches clinical context. Beta-D-glucan and Galactomannan positive.

* The top 5 anaerobes were *Faecalibacterium prausnitzii*, *Eubacterium rectale*, *Akkermansia muciniphila*, *Acidaminococcus intestini*, and *Bifidobacterium adolescentis*.

Abbreviations:

dPCR: digital PCR

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
EZ1 Virus Mini Kit v2.0	Qiagen	955134
EZ1 Advanced XL System	Qiagen	9001874
NEBNext Ultra II DNA Library Prep Kit	New England BioLabs	E7645L
NEB USER Enzyme	New England BioLabs	M5505L
Agencourt AMPure XP Beads	Beckman Coulter	A63881
KAPA Real-time Library Amplification Kit	Roche	KK2702
SYBR Green I nucleic acid gel stain	Invitrogen	S7563
Qubit dsDNA HS Assay kit	Thermo Fisher	Q32854
Nanopore 1D kit	Oxford Nanopore	SQK-LSK309