nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-018-0349-6

Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease

Timothy A. Blauwkamp[1,3*], Simone Thair[2,3], Michael J. Rosen[1], Lily Blair[1], Martin S. Lindner[1], Igor D. Vilfan[1], Trupti Kawli[1], Fred C. Christians[1], Shivkumar Venkatasubrahmanyam[1], Gregory D. Wall[1], Anita Cheung[1], Zoë N. Rogers[1], Galit Meshulam-Simon[1], Liza Huijse[1], Sanjeev Balakrishnan[1], James V. Quinn[2], Desiree Hollemon[1], David K. Hong[1], Marla Lay Vaughn[1], Mickey Kertesz[1], Sivan Bercovici[1], Judith C. Wilber[1,3] and Samuel Yang[2,3]

Thousands of pathogens are known to infect humans, but only a fraction are readily identifiable using current diagnostic methods. Microbial cell-free DNA sequencing offers the potential to non-invasively identify a wide range of infections throughout the body, but the challenges of clinical-grade metagenomic testing must be addressed. Here we describe the analytical and clinical validation of a next-generation sequencing test that identifies and quantifies microbial cell-free DNA in plasma from 1,250 clinically relevant bacteria, DNA viruses, fungi and eukaryotic parasites. Test accuracy, precision, bias and robustness to a number of metagenomics-specific challenges were determined using a panel of 13 microorganisms that model key determinants of performance in 358 contrived plasma samples, as well as 2,625 infections simulated in silico and 580 clinical study samples. The test showed 93.7% agreement with blood culture in a cohort of 350 patients with a sepsis alert and identified an independently adjudicated cause of the sepsis alert more often than all of the microbiological testing combined (169 aetiological determinations versus 132). Among the 166 samples adjudicated to have no sepsis aetiology identified by any of the tested methods, sequencing identified microbial cell-free DNA in 62, likely derived from commensal organisms and incidental findings unrelated to the sepsis alert. Analysis of the first 2,000 patient samples tested in the CLIA laboratory showed that more than 85% of results were delivered the day after sample receipt, with 53.7% of reports identifying one or more microorganisms.

More than 1,000 microorganisms are known to cause human disease, collectively causing approximately five million hospitalizations per year in the United States[1]. Early pathogen identification is key to targeting effective therapy[2–4], but identification of the causative pathogen remains challenging in many clinical scenarios. Many common pathogens are difficult or impossible to culture outside the body[5,6], deep-seated infections often require invasive biopsy of infected tissue to make a diagnosis[7,8] and the administration of broad-spectrum antibiotics before aetiological identification frequently confounds specific diagnoses that would enable more effective and less toxic antimicrobial therapy[2,9,10].

Approaches based on PCR that identify microorganisms directly from the specimen can speed identification but these tests include only a small number of pathogens, requiring a presumptive diagnosis before a test is chosen[6,11–13]. A number of sequencing-based diagnostics that test for many pathogens at once are emerging, but these tests still require samples from infected tissue[14–18] and are limited to infections of the collected tissue. The ability to comprehensively identify pathogens causing infection throughout the body from samples obtained non-invasively remains an unmet clinical need.

Sequencing of cell-free DNA (cfDNA) has recently been shown to enable non-invasive diagnosis of several indications that previously required invasive procedures, including the diagnosis of fetal abnormalities[19–21], detection of transplanted organ rejection[22–26] and characterization or monitoring of cancer[27–30]. We have previously shown that fragments of genomic DNA from pathogens causing infection at various locations in the body are found in purified plasma cfDNA[31,32], thus raising the possibility of non-invasive detection of a wide range of infections by sequencing the microbial cfDNA. Although next-generation sequencing (NGS)-based approaches hold promise for clinical infectious disease testing, the need for quality controls and validation strategies that address the specific challenges for clinical metagenomics must be addressed[14,17,33–38].

Here we describe the analytical and clinical validation of a quantitative microbial cfDNA sequencing test that identifies 1,250 human pathogens based on the fragments of genomic DNA these pathogens leave in blood (the Karius Test). Performance characterization was based on Clinical and Laboratory Standards Institute guidelines and informed by the US Food and Drug Administration draft guidance for characterizing NGS-based infectious disease tests[39].

Results

Analytical validation strategy. An overview of the workflow for microbial cfDNA sequencing is shown in Fig. 1, and a summary of the pathogens detected is presented in Supplementary Table 1. A reference panel of 13 microorganisms was designed to characterize test performance across several potential sources of bias (Fig. 2 and Supplementary Table 2). This reference panel spanned all kingdoms, ranged from 20 to 70% GC-content and had genomes

[1]Karius, Inc., Redwood City, CA, USA. [2]Department of Emergency Medicine, Stanford University, Palo Alto, CA, USA. [3]These authors contributed equally: Timothy A. Blauwkamp, Simone Thair, Judith C. Wilber, Samuel Yang. *e-mail: tim.blauwkamp@kariusdx.com

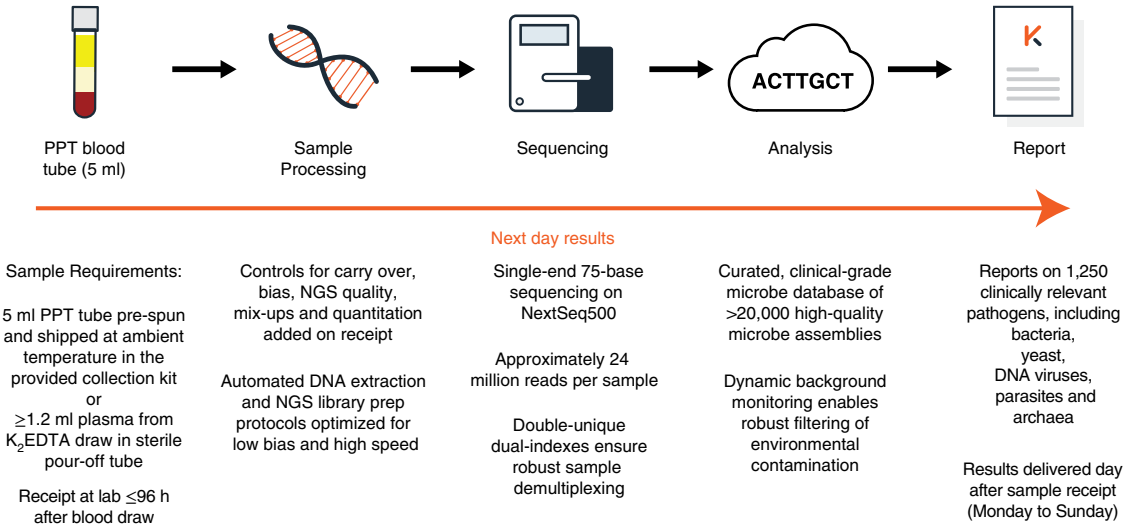

Fig. 1 | The Karius test workflow. Blood plasma from a routine draw is isolated and shipped overnight at ambient temperature to the Karius CLIA/CAP laboratory. Sample-specific controls are added on receipt and an automated liquid-handling platform performs cfDNA extraction and NGS-library preparation. The NGS libraries are multiplexed, inspected for quality and sequenced. A custom-built analysis pipeline uses a clinical-grade database to identify microbial DNA fragments found in plasma. Pathogens with plasma DNA levels that are significantly higher than real-time background thresholds are listed on the patient report, along with the concentration of the microbial cfDNA in plasma.

that ranged from kilobases to megabases in length. It also included both exclusive pathogens and frequent colonizers, as well as microorganisms that are commonly found as high-level environmental contaminants. Two pairs of closely related organisms, *Escherichia coli*/*Shigella flexneri* and *Staphylococcus aureus*/*Staphylococcus epidermidis*, were included to ensure fidelity of species discrimination during coinfections (Supplementary Table 2). Genomic DNA (gDNA) from each reference microbe was sheared to a typical microbial cfDNA fragment length and spiked into human plasma obtained from asymptomatic donors. As the concentration of human cfDNA in plasma can range over 1,000-fold, potentially affecting the sensitivity of sequencing-based assays[40,41], test performance was characterized in three different human plasma matrices representing 'low human', 'medium human' and 'high human' cfDNA samples (Fig. 2e).

In addition to the experiments with reference microorganisms in the laboratory, performance metrics were also assessed in silico across a broader range of microorganisms than is possible in the clinical laboratory. The design of these in silico experiments tested performance in the face of genetic divergence between clinical isolates and reference genomes. This was accomplished primarily by blinding the analysis pipeline to the strain from which the simulated reads were drawn, thereby forcing the analysis to proceed in the absence of the exact strain match in the database. The conclusions drawn from the contrived and in silico samples were further verified using clinical samples where appropriate.

Analytical performance characterization. *Limit of detection.* The limit of detection (LoD) was determined for each of the thirteen reference microorganisms in each of the low-, medium- and high-human-DNA-plasma matrices (Fig. 3). At a typical sequencing depth, the LoD ranged from 33 to 74 molecules of microbe-specific cfDNA per microlitre of plasma (MPM) for most reference organisms in the low human plasma (Supplementary Table 3). The LoD range increased slightly to 39–103 MPM in the high human plasma. The performance of *Pseudomonas aeruginosa* differed from the other reference microorganisms in two ways. First, the LoD in the low human plasma was approximately tenfold higher than most other reference microorganisms, at 415 MPM. Second, the calculated LoD improved as the human cfDNA increased, decreasing to 132 MPM in the high-human-plasma matrix. Both of these aberrant properties of *P. aeruginosa* can be explained by the high levels of contaminating DNA from this species in the environment (Fig. 2d). The sensitivity of sequencing-based tests depends on the number of sequencing reads obtained, so we also sub-sampled every sample down to the minimum number of reads required to pass the quality controls (Supplementary Table 3). Sub-sampling to this level removed approximately 90% of the reads from each sample of typical depth and the LoD showed a corresponding approximately tenfold increase, ranging from 326 to 596 MPM for most organisms and from 4,159 to 1,341 MPM for *P. aeruginosa*, across all plasma backgrounds. The LoD at a variety of intermediate sequencing depths between full and minimum depth is shown in Fig. 3b. Of all of the samples run in the analytical validation, 95% had sequencing coverage of at least 181,000 unique whole-assay internal normalization control (WINC) molecules, above which the median LoD of all reference microorganisms in all human plasmas varied from 34 to 74 MPM.

Although it was feasible to measure the LoD directly for the 13 reference microorganisms, our reportable range spans 1,250 taxa with a potentially broader range of sensitivities, due, for example, to variations in the level of background DNA or divergence between clinical strains and our reference database. We designed an in silico experiment to test that the LoDs estimated above were representative of the assay sensitivity across the broader clinical reportable range (CRR). Briefly, we selected 125 microorganisms from the CRR at random and simulated a number of reads that corresponded to the LoD. These reads were added to reads from a sequenced library of human cfDNA and run through our analytical pipeline while blinded to the assembly from which the simulated reads were generated. The simulated organism was correctly identified in 121 of 125 simulations for a sensitivity of 97% and the positive predictive value (PPV) was 99% (121 of 122), consistent with the expected 95% sensitivity at the LoD (Fig. 3c).

Limit of quantitation and linearity. The limit of quantitation (LoQ) for this test was defined as the lowest nominal input concentration of microbial cfDNA that had a precision corresponding to a

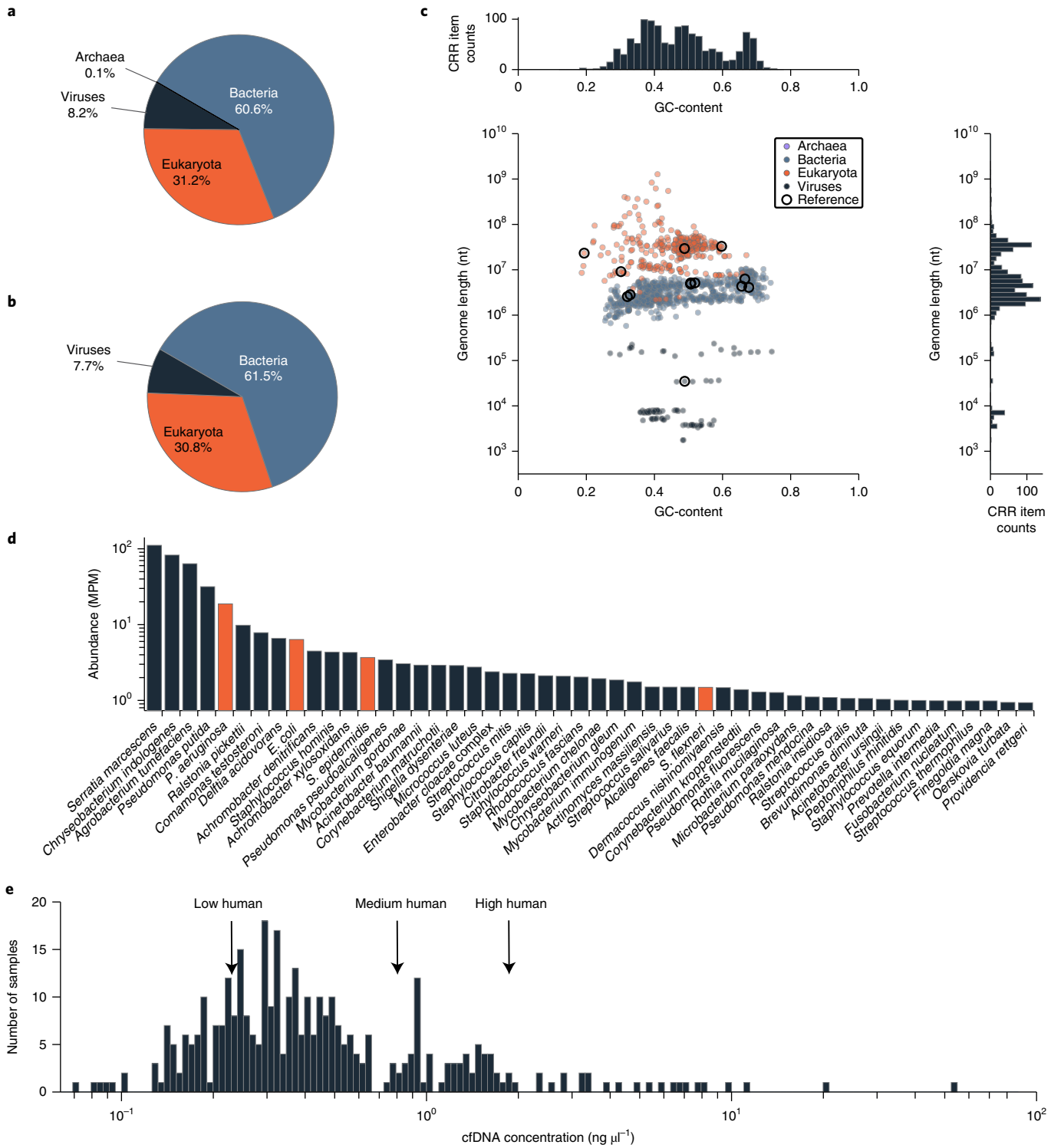

Fig. 2 | A reference panel of 13 microorganisms was designed to reflect the diversity of the 1,250 microorganisms tested by the assay. A full list of the 1,250 microorganisms tested in each sample (CRR) is available from ref. [54]. **a**, Relative number of microorganisms comprising the CRR by superkingdom. **b**, Relative number of microorganisms comprising the reference panel used for analytical validation by superkingdom. **c**, The GC-content and genome length of all microorganisms comprising the CRR are shown, colour coded according to superkingdom. The microorganisms used for reference materials are shown. Histograms above and to the right of the scatter plot show the total number of CRR organisms at each GC-content and genome length. **d**, The 97.5th percentile abundance observed in negative control samples is shown for the 4% of microorganisms found as top environmental contaminants. The orange bars indicate the four microorganisms that were part of the analytical validation reference panel. **e**, Histogram of the cfDNA concentrations post-extraction isolated from 350 SEP-SEQ study samples. The performance characteristics were established in three different human plasma matrices containing different concentrations of human nucleosomal DNA, representing the 23rd (low human), 77th (medium human) and 93rd (high human) percentiles of plasma cfDNA concentration observed in patients with suspected sepsis.

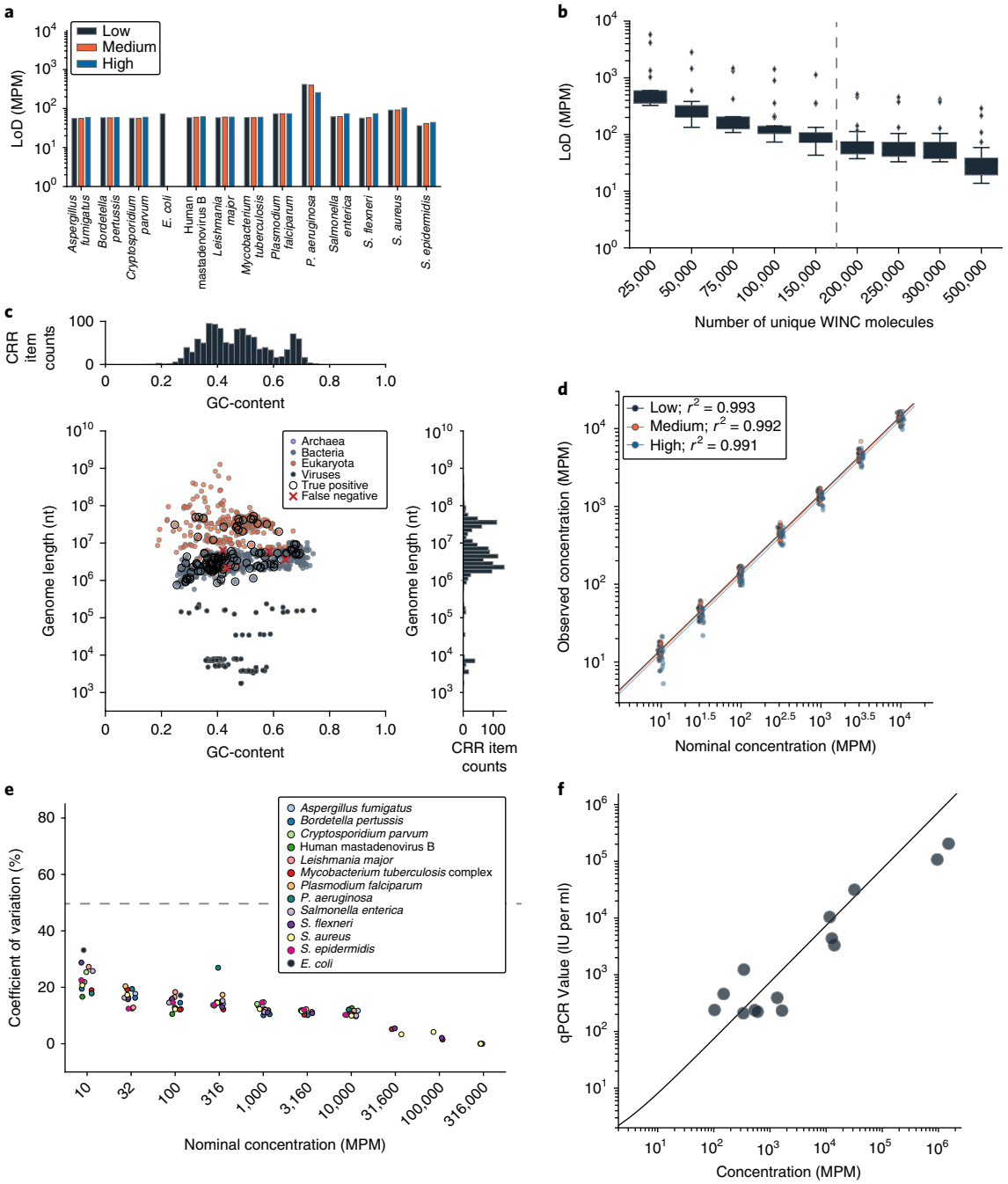

Fig. 3 | Analytical sensitivity. The LoD and LoQ were determined using sheared microbial cfDNA reference materials for each of the 13 representative pathogens mixed together at nominally identical concentrations and spiked into low-, medium- and high-human-plasma matrices over seven 0.5-log serial dilutions ranging from 10,000 to 10 MPM. Twelve replicates of each concentration in each healthy plasma matrix were measured over 12 different days, by different operators, using different instruments. **a**, The LoD was determined for each of the 13 representative microorganisms in each plasma matrix by Probit analysis after down-sampling sequencing depth to levels representative of the depth in clinical batches from the CLIA laboratory (300,000 unique WINC molecules). **b**, The LoD was determined as a function of sequencing depth. Each box plot represents the LoD distribution of the 13 pathogens in high-, medium- and low-human-plasma matrices ($n = 37$ LoD calculations at each dilution). The boxes represent the interquartile range and whiskers extend to 1.5× of that range; 95% of the validation samples have sequencing coverage levels that lie to the right of the dotted line. **c**, The robustness of the LoD was verified in silico across 125 additional microorganisms (10% of CRR) by contriving datasets that represented cfDNA abundances at the lowest concentration above the LoD. Both the microorganisms that were accurately identified ($n = 121$) and those that were not identified by the assay ($n = 4$) are indicated. Histograms above and right of the scatter plot show the total number of CRR organisms at each GC-content and genome length. **d**, The correlation between the nominal input and observed concentrations is shown for *Aspergillus fumigatus*, based on a linear fit of log-transformed values in each of the three human background levels ($n = 12$ replicates at each concentration for each human background). **e**, Each data point represents the coefficient of variation of the 12 replicate concentrations for the indicated microbe at the indicated nominal concentration in the low-human-plasma matrix. The dotted line indicates the 50% coefficient of variation. **f**, The MPM correlated with quantitative CMV-PCR for patients with low, medium and high viral loads. All 15 samples that were within the quantifiable range of the CMV-PCR test were included. IU, International Units.

coefficient of variation lower than 50% while maintaining linearity with higher concentrations. The LoQ was calculated for each of the 13 reference microorganisms in each of the three human plasma backgrounds. Strong linearity was observed for all microorganisms in all human plasma matrices across the entire measured concentration range of 10 to 316,000 MPM (Fig. 3d and Supplementary Figs. 1,2). Linearity was equally strong in three culture-positive clinical samples (one was coinfected with two microorganisms) that were serially diluted with healthy human plasma (Supplementary Fig. 3a). Note that the concentration of microbial cfDNA reported was not affected by the amount of human cfDNA.

The precision at various cfDNA concentrations was determined using the 12 replicates in the LoD experiment. For most microorganisms, the coefficient of variation did not exceed 50% at any concentration in any human plasma background (Fig. 3e and Supplementary Fig. 4). Only *S. flexneri* and *P. aeruginosa* exceeded the 50% threshold at any concentration and both were at cfDNA concentrations below the LoD. Note that the depth of sequencing coverage did not affect the reported concentration of microbial cfDNA (Supplementary Fig. 3b). As all of the microorganisms in all of the backgrounds had a coefficient of variation that was lower than 50% and strong linearity down to concentrations lower than the calculated LoD for that organism, the LoQ equalled the LoD for all 13 microorganisms in all three human backgrounds.

To verify the quantitative accuracy of this test, we compared our results with those of quantitative cytomegalovirus (CMV)-PCR testing on blood samples from 25 immunocompromised patients. All five samples that were negative by CMV-qPCR were also not reported by microbial cfDNA sequencing. Five samples were positive but below the LoQ for CMV-qPCR, all were detected by this test at relatively low levels. Finally, of the 15 samples that were detected and quantified by CMV-qPCR, all were detected by this test and the abundances that were reported by each test were correlated with an r^2 value of 0.88 on log-transformed data (Fig. 3f). Approximately 1,000 cfDNA fragments were detected for every international unit of CMV, although this ratio is expected to vary with several factors including the location of infection, the specific pathogen and the clinical treatment history.

Precision. Within-run (repeatability) and within-laboratory (reproducibility) quantitative precision were established using the low-human-plasma matrix. Samples were spiked with different control mixtures each day to ensure the precision estimates included all possible sources of variation. The coefficient of variation for all microorganisms ranged from 16.7 to 18.9% within a run and 17.9 to 22.2% within the laboratory (Supplementary Fig. 5a). Precision showed no correlation to either GC-content or background. To confirm the precision estimates, a set of 20 samples from patients with blood-culture-confirmed infections were run twice in separate batches on different days with different operators. Duplicate measurements of individual microorganisms showed an average coefficient of variation of 21.1% (with standard deviation of 13.6%), which is comparable to the within-laboratory precision (Supplementary Fig. 5b).

Analytical specificity. The analytical specificity in clinical metagenomic testing can be influenced by a number of factors that are not accounted for in traditional validation guidance. These include the ubiquitous contamination of reagents and consumable surfaces with DNA fragments from a variety of microorganisms[17,33], similarity in the genome sequences of distinct microorganisms, errors in the reference genome sequences and the divergence of clinical isolates away from the reference strains. We addressed analytical specificity in light of each of these factors below.

Because the composition of endogenous microbial cfDNA species in plasma varies considerably from person to person, it is difficult to accurately distinguish environmental contaminant DNA from plasma cfDNA at low levels when using a variety of different clinical samples. Therefore, we first assessed environmental contamination-related analytical specificity using replicate measurements of a single well-characterized asymptomatic human plasma pool. Fifty frozen aliquots were generated from the same plasma pool and subsequently tested in batches of seven or eight samples over nine days (Fig. 4a). Of the 50 replicates tested, unexpected microbial cfDNA was detected in only one replicate, from *Gordonia bronchialis* at 1,000 MPM, yielding an overall specificity of 98% (49 of 50 true negative samples; *Helicobacter pylori* cfDNA was present near LoD). As the test reports on 1,250 microorganisms in each sample and 1,249 microorganisms were expected to be negative in each of the 50 samples, the specificity was 99.998% on a per analyte basis (62,449 of 62,450).

One of the highest risk factors for false positives in bioinformatic analyses is the cross-reactivity between genetically similar microorganisms when a high level of DNA from one of these organisms is present. Such cross-reactivity is likely to occur when genomic regions can be mapped to multiple organisms or when reference genomes in the database are not of clinical grade. We first examined the risk of false positive calls as a result of pathogen cross-reactivity using high levels of simulated reads (60,000 MPM) from each of the 1,250 microorganisms in the reportable range spiked into read sets from asymptomatic plasma samples (Table 1). All of the spiked-in microbial cfDNAs were identified (1,250 of 1,250), with a 99.4% positive predictive value (seven samples had one additional false positive call each, each belonged to the same genus as the true call).

The risk of false positive calls may be higher when the clinical isolate differs from the available reference sequences due to natural biological diversity or change in microbial genomes. To address this, we generated infection read sets for 10% of the CRR (125 taxa) as above but blinded the analysis pipeline to the assembly from which the reads were drawn to simulate differences between clinical isolates and reference genomes in the database (Table 1). Only one false positive call was made in 125 simulations of cfDNA near the LoD, in addition to 121 true positive calls, for a PPV of 99.2% per sample ($n = 121$ of 122) and 99.9994% per analyte ($n = 156{,}124$ of 156,125). To test the effects of clinical isolate diversity even more rigorously, high-level infections were simulated (60,000 MPM) in the absence of any restrictions on divergence between the simulated infecting strain and the database strains. Here, 93.6% (117 of 125) of simulations identified the correct microbe at the species level, with a positive predictive value of 92.1% (117 true calls of 127 total calls). All ten incorrect calls were in the expected genus.

The final aspect of analytical specificity investigated was quantitative interference during coinfection with genetically similar organisms. Pre-analytical mixtures of healthy plasma spiked with DNA from *E. coli* and *S. flexneri* or *S. aureus* and *S. epidermidis* at ratios of 3:1, 1:1 and 1:3 demonstrated no significant difference in microbial DNA concentration during coinfection relative to the concentration expected from a mono-infection (Fig. 4b and Supplementary Fig. 6). In silico experiments with an additional 125 simulated coinfections of related species at ratios of 3:1, 1:1 and 1:3 showed that 99.3% of microorganisms (745 of 750) had an MPM that was higher than 50% of the expected value (Fig. 4c), confirming a very low level of interference from coinfection with closely related species for this test.

Clinical performance characterization. To assess the validity of plasma-derived microbial cfDNA as a biomarker for a broad range of infecting pathogens, we compared the results of this microbial cfDNA sequencing with pathogens identified by standard-of-care testing in a cohort of 350 patients presenting to the emergency department that met the sepsis alert criteria[42] from July 2016 to August 2017 (SEP-SEQ Study, NCT02730468; Supplementary Fig. 7). Of these, 348 had both sequencing test and initial blood culture

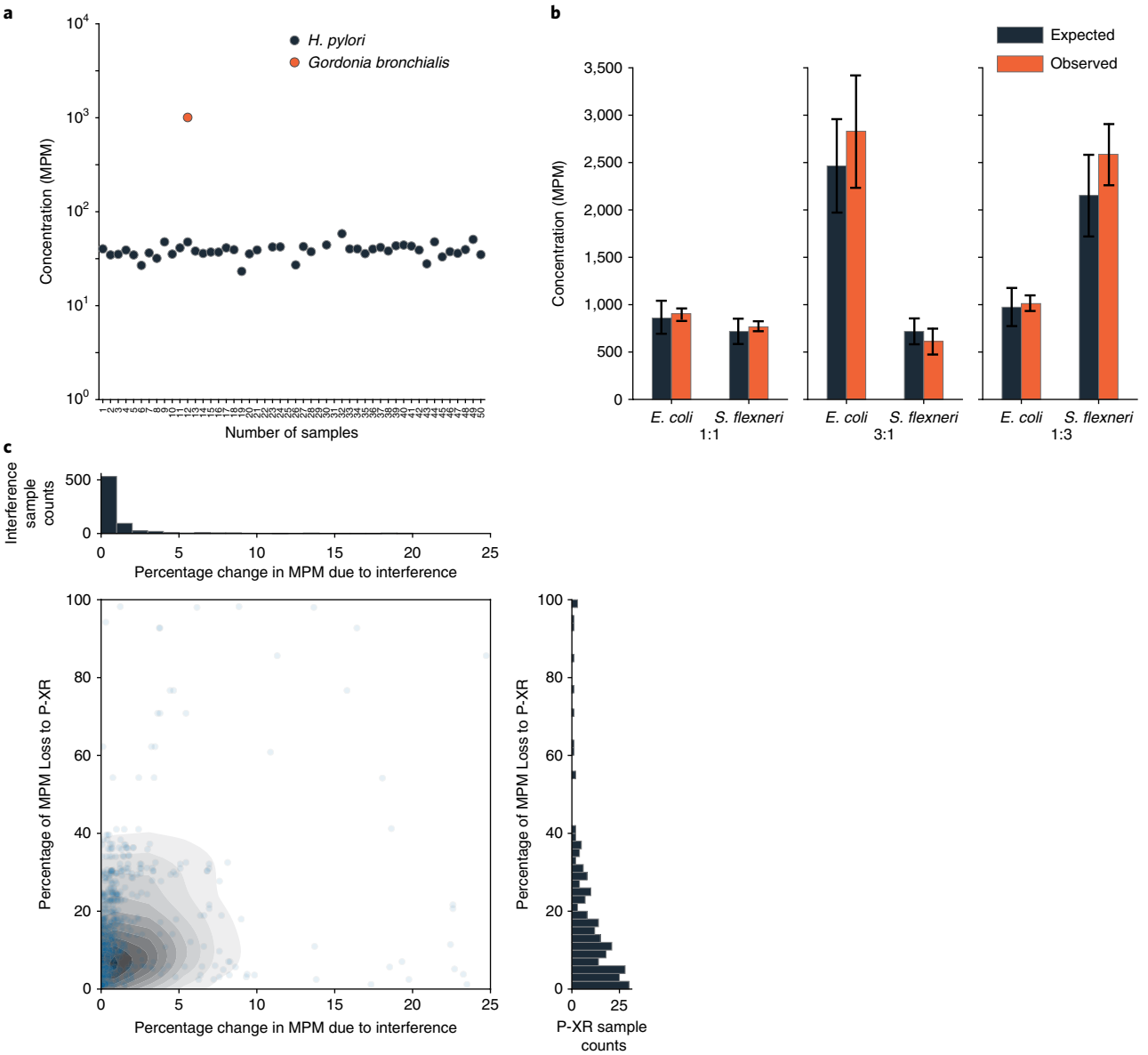

Fig. 4 | Analytical specificity. a, Analytical specificity as a function of environmental contamination was investigated using repeated measurements of a single, thoroughly characterized healthy plasma sample composed of eight different donors. The results from 50 replicate measurements taken across 9 days are shown. *H. pylori* cfDNA was known to be present in this sample, whereas other microorganisms were not. **b**, Potential interference during coinfection was assessed using healthy human plasma samples spiked with various ratios of genetically similar microbial cfDNA reference materials. The reported plasma cfDNA concentrations are shown for *E. coli* and *S. flexneri* when coinfection was contrived at ratios of 1:1, 3:1 and 1:3. The expected plasma cfDNA concentration was calculated from samples where each organism was spiked into a sample alone. The mean ± s.d. from three replicates is shown. $P > 0.05$ for all comparisons ($P = 0.58$, 0.56 and 0.13 for *E. coli* and 0.41, 0.13 and 0.32 for *S. flexneri*, at 1:1, 3:1 and 1:3 ratios, respectively), two-tailed *t*-tests . **c**, In silico assessment of accuracy in coinfected samples. The *y* axis shows the bias in concentration due to pathogen cross-reactivity (P-XR) in mono-infection and the *x* axis shows the additional bias due to interference during coinfection. Histograms show the number of in silico simulated samples with various levels of MPM bias due to pathogen cross reactivity (right) and interference from co-infecting organisms (above).

results available and were included in the analysis. The demographic and clinical characteristics of the enrolled patients are summarized in Supplementary Table 4. More than one-quarter (27.7%; 97 of 350) of the patients received antimicrobial treatment within two weeks preceding presentation. The mean length of patient hospital stay was 4.7 days, 6% required care in the intensive care unit during their stay and four patients died during hospitalization.

Compared with initial blood culture, cfDNA sequencing had a sensitivity of 93.7% (59 of 63; confidence interval (CI) of 84.5–98.2%; Table 2 and Fig. 5a). Discordant positive results included unculturable bacteria, bacteria from patients that were pre-treated with antimicrobials, viruses and eukaryotic pathogens. Additional microbiological testing over the first seven days of admission, including tissue and fluid cultures, serology, nucleic acid testing and subsequent blood cultures, identified 69 additional microbiological causes of the sepsis alert, of which microbial cfDNA sequencing identified 53, yielding an overall sensitivity of 84.9% ($n = 112$ of 132) compared with all microbiological testing. The discordant positive results from this comparison were adjudicated according to the criteria outlined in Methods and Supplementary Fig. 8 to generate a composite reference standard for the aetiology of the sepsis alert. In comparison to this composite reference standard, cfDNA sequencing demonstrated

Table 1 | Diversity robustness

	Bioinformatic cross-reactivity	Diversity robustness near the LoD	Unconstrained diversity robustness
Simulations	1,250	125	125
Diversity robustness	Exact match	No exact match; ≥ 1 assembly; <3% divergence	No exact match; ≥1 assembly; unconstrained divergence
Simulated infection level	High	Near LoD	High
PPV (%)	99.4 (n=1,250 of 1,257)	99.2 (n=121 of 122)	92.1 (n=117 of 127)
Specificity per analyte (%)	99.9995 (1,561,243 true negative calls of 1,561,250 total negative calls)	99.9994 (156,124 true negative calls of 156,125 total negative calls)	99.994 (156,115 true negative calls of 156,125 total negative calls)

Cross-reactivity was assessed using in silico simulations to measure the rate of false positive calls at a variety of simulated infection levels and with varying degrees of genetic distance between the infecting microbe and the genomes present in the database to simulate genetic diversity of clinical isolates. For bioinformatic cross-reactivity experiments, the infecting microbe exactly matched an assembly in the database. For diversity rubustness near the LoD, the infecting microbe did not have an exact assembly match in the database, but did have at least one assembly in the database less than 3% diverged from the infecting microbe. For unconstrained diversity robustness, the infecting microbe did not have an exact assembly match in the database, but did have at least one assembly in the database with the same species label (no limit on divergence).

Table 2 | Positive and negative agreement of microbial cfDNA sequencing versus initial blood culture, all microbiological testing and composite reference standard

Patient characteristics (n=348)	NGS positive	NGS negative	Agreement (%)	95% CI (%)
Positive by initial blood culture	59	4	93.7	84.5–98.2
Negative by initial blood culture	171	114	40.0	34.3–45.9
Positive by all microbiological testing	112	20	84.8	77.6–90.5
Negative by all microbiological testing	112	104	48.2	44.3–55.0
Positive by composite reference standard	169	13	92.9	88.1–96.1
Negative by composite reference standard	62	104	62.7	54.8–70.0

The composite reference standard includes the results from all microbiological tests (including the initial blood culture) performed within seven days of presentation and clinical adjudication. The NGS false negatives compared to initial blood culture included *Listeria monocytogenes*, coagulase-negative *S. aureus*, *Streptococcus agalactiae* and *Stenotrophomonas maltophilia* (this organism was not included in the NGS-test reportable range). NGS agreement with other methods was calculated as described in Supplementary Figures 7 and 8.

a sensitivity of 92.9% (169 of 182; CI of 88.1–96.1%; Table 2 and Supplementary Table 5). Overall, the identification of sepsis alert aetiology was higher for cfDNA sequencing (169 of 348) than for both blood culture (63 of 348) and all microbiological testing combined over seven days (132 of 348). Of the 96 subjects that received antimicrobial treatment within two weeks preceding presentation, cfDNA sequencing identified pathogens that are classified as definite or probable causes of the sepsis alert in 46 (47.9%), whereas blood culture identified pathogens in only 19 (19.6%).

Among the 166 samples for which no definite or probable microbiological cause of the sepsis alert was identified, no significant microbial cfDNA was detected in 104 by cfDNA sequencing, resulting in a specificity for sepsis alert aetiology of 62.7% (104 of 166; CI of 55.2–70.4%; Table 2). The organisms identified by cfDNA adjudicated as possible or unlikely causes of the sepsis alert included a number of reactivated herpesviruses, chronic infections such as *H. pylori* and human papillomavirus, microorganisms likely to be commensals and possible causes of non-sepsis-related acute infection (Supplementary Tables 6 and 7).

For samples with a composite reference positive result, the estimated time to result for cfDNA sequencing was compared with conventional testing (Fig. 5b). For cfDNA sequencing, the time to result was represented using the median shipping and testing times for the last 400 samples run in the Clinical Laboratory Improvement Amendments of 1988 (CLIA)-certified lab. This time to result (53.0 h) was significantly shorter than the median time to positive result for conventional testing, based on the electronic medical records (EMR) (92.4 h; P=0.0004; Fig. 5b). Although cfDNA sequencing did not provide antimicrobial susceptibility testing, this analysis demonstrates that the sensitivity and speed with which species-level identification is provided by cfDNA sequencing may offer significant benefit to patients.

As microbial cfDNA is a new biomarker, additional context was provided by assessing test performance on 167 asymptomatic donors. Of the asymptomatic samples, 77.2% (129 of 167) had no microorganisms reported (Fig. 5c). Among the 22.8% of samples in which microbial cfDNA was reported, a single species was detected in most cases. In general, the concentrations of microbial cfDNA detected from asymptomatic donors were lower than the concentrations of microbial cfDNA detected from patients with confirmed infecting pathogens and similar to the concentration of microorganisms classified as unlikely causes of the sepsis alert (Fig. 5d). The most frequently detected microorganisms in these patients included *H. pylori*, *Klebsiella pneumoniae* and *Haemophilus influenzae*; the latter two are commonly present as human commensals (Supplementary Table 8).

Clinical laboratory experience with the first 2,000 samples. We characterized test performance across the first 2,000 patient plasma samples from across the United States submitted to our CLIA-certified, College of American Pathologists (CAP)-accredited laboratory for microbial cfDNA sequencing (Table 3). Due to collection-site handling errors or excessive shipping delays, 46 samples were not tested. Of the samples tested, 98.1% were reported, with 87.6% of the reports delivered the operating day after on-time sample receipt. Microorganisms were reported in 53.7% of all of the tested samples, covering 318 different species of bacteria, fungi, parasites and viruses. Among the positive reports, 49.6% reported a single organism, with the remainder reporting two or more organisms (Supplementary Fig. 9). Although no patient-specific information was used in determining which microorganisms to report, the analysis of passively collected ICD-10 codes associated with approximately half of the samples suggested that the most common use for the test was in immunocompromised patients, followed by sepsis, endocarditis and complicated pneumonia.

Discussion

The results presented here show that microbial cfDNA sequencing offers the potential to reliably identify a wide variety of infections

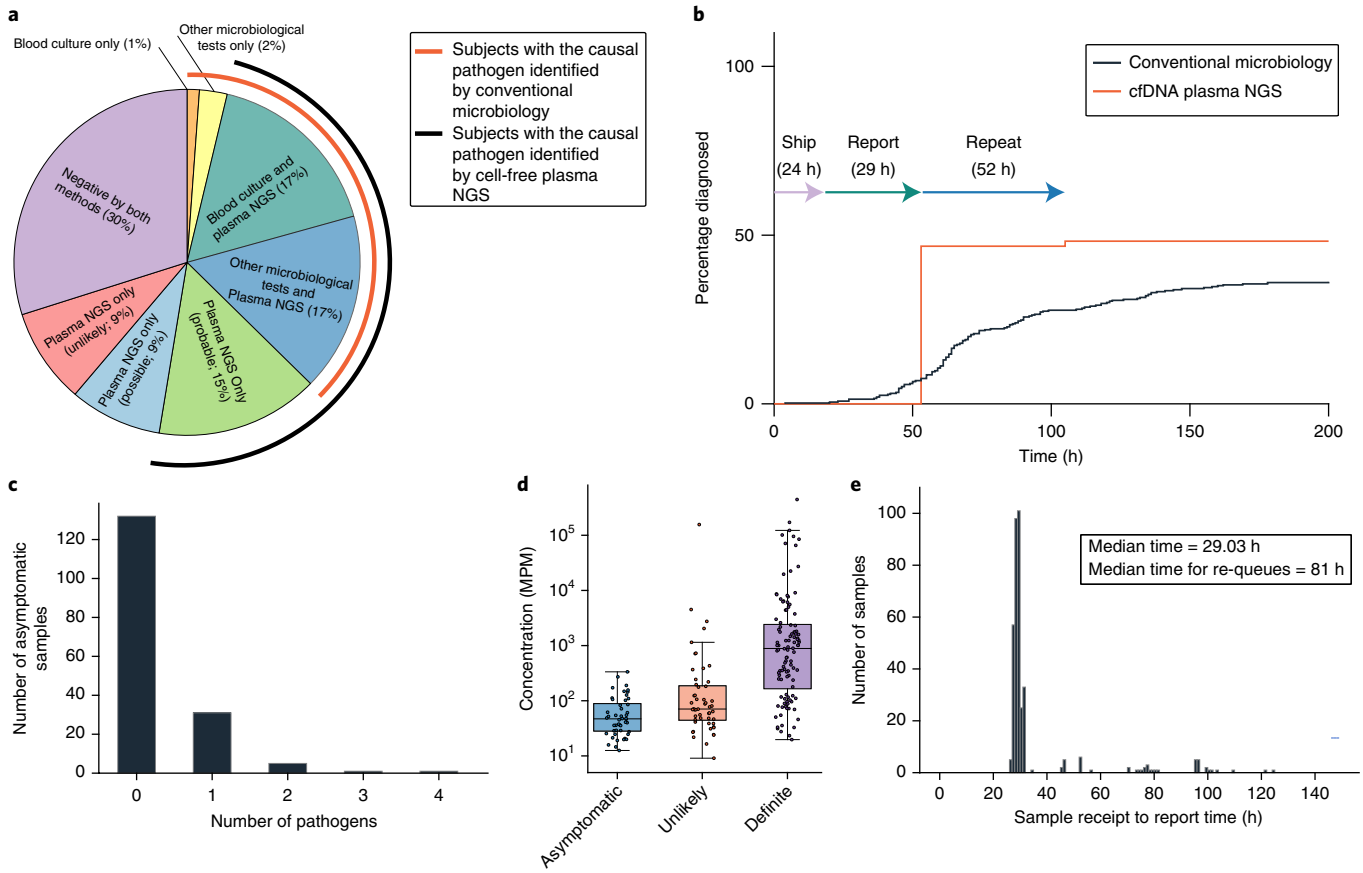

Fig. 5 | Clinical validity. a, The proportion of subjects with pathogens that were identified by the different methods. **b**, A comparison of the time to test results for both test methodologies. For display purposes, the time-to-result for all microbiology tests was truncated to 200 h (range, 4–1,104 h). **c**, The frequency of microbial and poly-microbial detection in 167 asymptomatic donors. **d**, The distribution of the microbial cfDNA abundances detected in asymptomatic individuals ($n=48$), the 'unlikely' calls in the SEP-SEQ cohort ($n=53$) and all microbiologically confirmed 'definite' calls ($n=112$). The boxes represent the interquartile range and whiskers are 1.5× this range. **e**, The time taken from sample receipt to report delivery for the last 400 clinical practice samples run in the clinical lab.

Table 3 | Test performance in production

Sample acceptance rate	97.7%
Total yield	98.1%
First-pass yield	91.0%
Results delivered next operational day	87.6%
Paediatric samples	45.3%

Table of relevant testing metrics for 2,000 clinical practice samples run through the CLIA lab.

throughout the body from a plasma sample in a clinically useful time frame. A number of challenges to the provision of high-quality diagnostic testing for clinical metagenomics applications have been previously identified[17,33,34] and we present additional considerations. Here, we attempt to address all of these challenges through a combination of traditional and metagenomic-specific validation strategies, including the use of 13 representative microorganisms that were chosen to probe the boundaries of analytical performance in 348 contrived samples, thousands of in silico simulations that tested the integrity of the bioinformatics pipeline in the face of clinical-isolate divergence, 580 clinical samples that assessed performance in patient samples and 2,000 samples that were run through our CLIA-certified laboratory and reported in real-time.

Having developed the test for low bias, we did not observe significant differences in analytical performance as a result of differing GC-content, genome size, superkingdom, genetic similarity among coinfecting organisms or differences of up to 3% between the detected strains and reference genome. Elevated levels of human cfDNA background in the sample had only minor effects on sensitivity and precision. The level of environmental contamination did influence test sensitivity, but only for the 5–8 microorganisms with the highest environmental backgrounds among 1,250 microorganisms probed. The sequencing depth also influenced LoD, but the processing methods for this test are designed to provide similar sequencing depth for all samples, such that 95% of the samples tested fell into a range of sequencing depths where sensitivity was consistent.

The positive agreement between this test and blood culture (93.7%) of patients with a sepsis alert is equal to or better than other direct molecular diagnostic methods, including real-time PCR panels and PCR combined with electrospray ionization[25,26]. Adjudication of results from the SEP-SEQ study also showed that this test identified a greater number of aetiological causes of the sepsis alert than standard-of-care testing. However, the sensitivity and breadth of microorganisms detected, combined with the diversity of the microbiome compositions across patients[17,34], makes it challenging to achieve high diagnostic specificity. The analytical validation experiments demonstrated very low levels of falsely reporting of microbial cfDNA that was not in the original plasma sample (Fig. 4), consistent with high reproducibility of cfDNA detection across independent

runs (Supplementary Fig. 5). Therefore, the microbial cfDNA identified in 22.8% of the asymptomatic samples is probably derived from the original donor specimen, representing cfDNA primarily from commensal organisms or subclinical colonization (Supplementary Table 8). Similarly, the cfDNA species reported in the possible or unlikely SEP-SEQ samples (37.3% of negative composite reference samples) is presumed to originate from these sources in addition to incidental findings unrelated to the sepsis alert.

Although it would be convenient if the cfDNA concentration alone were indicative of true infection, both the microbe identity and the location of infection are likely to influence the concentration of microbial cfDNA observed in plasma. The cfDNA concentrations associated with definite calls spanned four orders of magnitude (20–450,000 MPM). The cfDNA concentrations of microorganisms that were detected in asymptomatic donors, however, did not exceed 300 MPM and investigation of the unlikely causes of sepsis alert with concentrations greater than 300 MPM showed that these were primarily reactivated herpesviruses, chronic infections with human papillomavirus and probable causes of acute infection. Therefore, high concentrations of microbial cfDNA were typically associated with true infections, whereas low concentrations were associated with both true infections and commensal/colonizer/contaminant microorganisms of unknown clinical significance. Regardless of the cfDNA concentrations, the entire clinical picture must be considered when determining the clinical significance of a microbe detected by cfDNA sequencing.

The performance characteristics reported here point towards clinical applications in which the benefits of sensitivity, non-invasive sampling and broad testing outweigh the limitations of sequencing cost, turn-around time and identification of microorganisms beyond those related to a specific indication. In light of these trade-offs, use cases that could potentially benefit from this approach include rule-out testing for sepsis when the standard of care was not sensitive enough or too invasive, testing immunocompromised patients presenting with non-specific symptoms of infection or testing before invasive procedures with high cost or morbidity. Future studies that directly examine clinical utility will be important to establish the specific indications for which microbial cfDNA sequencing should be used.

Methods

Clinical-grade microbial cfDNA sequencing for infectious disease. *Sample preparation and sequencing.* All samples were processed with locked and version-controlled protocols, analytical pipelines and reference databases in the Karius CLIA-certified and CAP-accredited laboratory. Following receipt, plasma samples were thawed, if frozen, spiked with a known concentration of synthetic normalization molecule controls (see Methods below) and centrifuged at 16,000g for 10 min to remove residual cells. Cell-free DNA was extracted from 0.25 ml plasma using a modified Mag-Bind cfDNA Kit (Omega Biotek) in Hamilton STAR liquid handling workstations. DNA libraries for sequencing were constructed using customized dual-indexed Ovation Ultralow System V2 library preparation kits (NuGEN) in Hamilton STAR liquid handling workstations. Sequencing libraries were pooled with environmental and assay control samples that were processed alongside the test samples in every batch. Pooled sample libraries were purified and up to 24 libraries per batch were multiplexed and sequenced on Illumina NextSeq500 sequencers using a 75-cycle single-end, dual index sequencing kit. On average, approximately 24 million reads were obtained for each sample.

Sequence data processing and alignment. The primary sequencing output was demultiplexed by bcl2fastq v2.17.1.14 (with default parameters), the reads were quality trimmed and subsequently filtered if shorter than 20 bases by Trimmomatic v0.32 (ref. [43]). Reads that passed these filters were aligned against human and synthetic (including synthetic normalization molecule control and sequencing adapter) references using Bowtie v2.2.4 (ref. [44]). Reads that aligned to either were set aside. Reads potentially representing human satellite DNA were also filtered via a *k*-mer-based method. The remaining reads were aligned with our microorganism reference database using BLAST v2.2.30 (ref. [45]). Reads with alignments that exhibited both high per cent identity and high query coverage were retained, with the exception of reads that aligned with any mitochondrial or plasmid reference sequences. PCR duplicates were removed based on their alignments.

Microorganism abundance estimation. Relative abundances were assigned to each taxon in a sample on the basis of the sequencing reads and their alignments. For each combination of read and taxon, we defined a read sequence probability that accounted for the divergence between the microorganism present in the sample and the reference assemblies in our database. A mixture model was used to assign a likelihood to the complete collection of sequencing reads that included the read sequence probabilities and the (unknown) abundances of each taxon in the sample. An expectation-maximization algorithm was applied to compute the maximum likelihood estimate of each taxon abundance[46]. From these abundances, the number of reads arising from each taxon was aggregated up the taxonomic tree.

Pathogen detection. A set of environmental control samples were processed and sequenced within each batch. The estimated taxon abundances from the environmental control samples within the batch were combined to parameterize a model of read abundance arising from the environment with variations driven by counting noise. Statistical significance values were computed for each estimated taxon abundance in each non-environmental-control sample and those within the CRR at high significance levels comprised our candidate calls. Final calls were made after additional filtering was applied, which accounted for read location uniformity, read percent identity and cross-reactivity originating from higher abundance calls. The microorganism calls that passed these filters were reported along with abundances in MPM, as estimated using the ratio between the unique reads for the taxon and the number of observed unique reads of the WINCs (described below).

Quality control description. Test results were individually inspected for performance against a number of quality control criteria before being released. The assay was quality controlled using both internal controls spiked into every sample and external controls included in every batch. The internal sample controls included an ID Spike and the WINC molecules. We designed 137 different ID Spike variants, each composed of a unique synthetic DNA sequence not observed in nature. Each sample in a batch received a different ID Spike, which was then used to control for sample mix-ups throughout the entire workflow. In addition, each sample was monitored for the presence of any additional ID-Spike sequences as a marker of sample-to-sample contamination. Any samples showing signs of sample-to-sample contamination at a level that could lead to false positive calls were subject to a one-time requeue.

WINC molecules were used to monitor the yield and quality of the entire workflow. These are synthetic DNA molecules that contain 16 degenerate bases and are spiked into plasma at a specific molar concentration. By counting the number of unique WINC molecules in the sequencing data obtained from each sample, the whole assay yield was monitored and the concentration of each microbial cfDNA species in the original plasma aliquot determined. Samples that failed to achieve a minimum number of unique WINC molecule reads for any reason (for example, poor yield in any workflow step, poor sequencing quality or failure to properly de-multiplex a sample) resulted in a one-time requeue.

Two types of batch controls were run alongside patient samples in every batch. Four replicates of environmental control samples containing buffer instead of plasma were processed in parallel with patient samples all the way from accessioning to report generation. Environmental controls were used as described above to monitor microbial DNA signals arising from the background at the time of batch processing. Two assay control samples, each containing *S. epidermidis*, *E. coli* and *P. aeruginosa* at 1,000 MPM and *A. fumigatus* at 10,000 MPM were also included in every batch. All four spiked microorganisms, and no others, had to be called within a specified MPM range in both assay controls to pass the final quality control inspection.

Reference database and quality control. Reference genomes for *Homo sapiens* and microorganisms (bacteria, viruses, fungi/moulds and other eukaryotic pathogens) were retrieved from the National Center for Biotechnology Information ftp site (NCBI, US National Library of Medicine; https://www.ncbi.nlm.nih.gov/genome/guide/human/ and https://www.ncbi.nlm.nih.gov/genome/microbes/, respectively). Sequence similarities between microorganism references were inspected to identify taxonomic mislabelling and sequence contamination. From the reference genomes passing these quality controls, a subset was selected to maximize sequence diversity. As part of the selection process, NCBI BioSample data (https://www.ncbi.nlm.nih.gov/biosample/) were used to ensure the inclusion of reference genomes from both clinical and non-clinical isolates. The final reference genome dataset included over 21,000 reference genomes, containing over 2.7 million sequences. Selected sequences were collected into a single FASTA file and used to generate our microorganism reference BLAST database.

CRR. The selection of organisms in our CRR was performed as follows. A candidate list was generated by two board-certified infectious disease physicians by including: (1) DNA viruses, culturable bacteria, additional fastidious and unculturable bacteria, mycobacteria, eukaryotic pathogens from the standard text[47] and a number of infectious disease references, (2) organisms in the pathogen database referenced in published case reports and (3) reference genomes sequenced from human clinical isolates (as indicated by the BioSample resource of the

NCBI) with publications supporting pathogenicity. Organisms from the above list that were associated with high-quality reference genomes, as determined by our reference database quality control process (see above), were used to further narrow the range. Finally, organisms at risk of generating common false positive calls because of sporadic environmental contamination were removed. The final list was defined as the CRR of this test (Supplementary Table 1; complete list available at www.kariusdx.com/pathogen-list, v3.1.1).

Analytical validation. *Reference materials.* Genomic DNA from 14 microorganisms was obtained from either the ATCC or NIST. Because the human mastadenovirus B genome was available only in small quantities, larger amounts were produced by seven non-overlapping PCR amplicons of approximately 5 kb each. Enzymatic shearing of each reference microbe genome was accomplished with DNaseI or Fragmentase (New England Biolabs) to create semi-randomly fragmented gDNA. Sheared gDNA was purified using Oligo Clean and Concentrator (Zymo Research) and quantified by fluorometery (Qubit, ThermoFisher Scientific). The fragment length distributions of the sheared gDNAs were evaluated by electrophoresis (TapeStation 2200, Agilent) and optimized to obtain consistent ranges across the 14 genomes (60–90 bp modal length), corresponding to the distribution of pathogen cfDNA found in clinical samples[48]. Quantified sheared gDNA was spiked into plasma pooled from 8–10 healthy donors (ZenBio) to create the low human contrived samples. Two additional healthy human plasma pools were generated by spiking the low-human-plasma pool with purified human mononucleosomes (EpiCypher) to simulate samples with different amounts of human cfDNA, for example, the medium- and high-human-plasma matrices. As measured with QuantIT PicoGreen (ThermoFisher Scientific), the extracted cfDNA levels averaged 0.23 ng μl^{-1} for low- (no added mononucleosomes), 0.81 ng μl^{-1} for medium- and 1.86 ng μl^{-1} for high-human-plasma matrices. The human cfDNA concentration was 100 to 1,000,000 times more abundant than sheared pathogen gDNA in the contrived samples used for this study. One of the microorganisms used for validation, *Clostridium sporogenes*, is not a member of our reportable range at present due to its confounded phylogenetic relationship with *Clostridium botulinum*[49] and was therefore excluded from all analyses. The purified human mononucleosomal DNA (EpiCypher) was found to contain significant levels of contaminating *E. coli* DNA and therefore analyses of *E. coli* in the medium- and high-human plasma were excluded from all analyses.

LoD. An LoD value was estimated for each of the 13 representative pathogens in high-, medium- and low-human-background levels at a variety of sequencing depths. Sheared gDNA from each of the representative pathogens were mixed at nominally equivalent molar concentrations, spiked at 10,000 MPM into healthy human plasma and diluted with healthy human plasma over 7 0.5-log serial dilutions ranging from 10,000 to 10 MPM per microbe. Twelve replicates of each dilution in each of the high-, medium- and low-human-plasma matrices were analysed over 12 different days according to the standard workflow. Probit analysis of 12 replicates at each concentration in each healthy plasma matrix was used to establish the LoD for each reference microbe in each plasma matrix as follows. We assumed that sensitivity (the probability of detection of each replicate) varies with log concentration according to a cumulative normal distribution, that is, a Probit model. This model implied that the number of positives at each concentration followed a binomial distribution and the whole dilution series had a likelihood given by the products of the binomial distributions for each concentration. The parameters of the cumulative normal distribution associating concentration with sensitivity were fitted to maximize the likelihood of the dilution series and the LoD determined by inverting the fitted distribution to determine the log concentration corresponding to a sensitivity of 95%. Note that for this experiment, we did not apply our cross-reactivity call filters due to the special nature of these contrived samples (highly multiplexed with all organisms at nominally equal concentrations).

The variation in LoD as a function of sequencing coverage was determined by sub-sampling to a level where a single unique pathogen read would correspond to MPMs of 12, 6, 4, 3, 2, 1.5, 1.2, 1 and 0.6 (or unique WINC counts of 25,000, 50,000, 75,000, 100,000, 150,000, 200,000, 250,000, 300,000 and 500,000 unique reads). We used the complete fastq in each replicate while performing microbial abundance estimation and then multiplied each taxon abundance by the ratio of the targeted WINC count to the observed WINC count before the determination of statistical significance over background.

LoD robustness. After computing the LoD for each combination of microorganism and human background level, we performed an in silico experiment to estimate the LoD across 125 additional microorganisms. We began by determining which taxa on the CRR corresponded to at least one pair of assemblies that were at most 3% diverged (as estimated from overlaps in *k*-mer content). Then, for each of a random selection of 125 of these taxa, we selected an assembly at random from those that have at least one similar assembly. To determine the number of reads of this assembly to simulate, we selected one of the 13 representative pathogens and a human background level at random to represent the wet-laboratory efficiency of capturing and sequencing reads of the simulated assembly. We found the lowest concentration at which the test had at least 95% sensitivity for this pathogen–background combination at the minimum sequencing depth. We then randomly selected one of the (sub-sampled) technical replicates at this concentration and determined the number of reads of the pathogen. Next, we simulated the same number of reads from the assembly chosen for simulation, selecting for each read a random start location and orientation, and a length drawn from the distribution observed in the wet laboratory representative. These reads were then added to those of a sequenced library of the low human cfDNA matrix without any spiked pathogens and run through our analytical pipeline blinded to the simulated assembly by discarding any alignments to it before abundance estimation. Finally, we determined whether the CRR taxon of the simulated assembly was reported by the analytical pipeline.

Precision. Precision samples were contrived in the low-human-plasma matrix with sheared microbial DNA from all 13 reference microorganisms spiked at a nominal concentration of 1,000 MPM. One-hundred replicate samples were generated and stored frozen at −80 °C until the day of testing. Two aliquots were thawed and processed according to the standard workflow on each day of testing, including independent addition of internal control mixtures on each day. Precision was calculated according to the Clinical and Laboratory Standards Institute EP05-A2, Appendix C guidelines for establishing within-run and within-laboratory precision using one batch a day[50].

Linearity. The best fits for our analysis of linearity were determined as follows. We sought to fit the data on a log scale (appropriate to dilution series) and to ensure that, on a linear scale, the fit was linear and intersected the origin. To achieve both of these objectives, we log-log transformed these data (consisting of nominal and observed concentrations) and then performed a best linear fit with a fixed slope of unity.

Interference. The extent of cross-reactivity between closely related species during coinfection was assessed first using pre-analytical mixtures. Contrived coinfections in healthy human plasma were generated using two pairs of genetically similar and clinically relevant microorganisms, *S. epidermidis* with *S. aureus* or *E. coli* with *S. flexneri*. Each pair of microorganisms was tested at concentration ratios of 1:3, 1:1 and 3:1. Triplicates of each contrived coinfection sample were tested and the MPM values averaged. The change in MPM of each organism, relative to the MPM observed when each organism was present alone in the sample, was calculated.

In silico experiments testing interference were divided into two sections, the first tested for inaccuracy of the observed concentration due to cross-reactivity of mono-infecting microorganisms with a similar taxon in the database and the second evaluated the accuracy of the observed concentration in the presence of closely related coinfecting microorganisms due to interference.

First, 125 pairs of assemblies from different CRR taxa that belonged to the same family were selected (250 assemblies total), where each assembly used for simulation was less than 3% divergent from another assembly in the same taxon (as estimated by *k*-mer distances). Then, we spiked 10,000 simulated reads from each of the 250 assemblies into different healthy human plasma samples down-sampled to the minimum sequencing coverage allowed by the quality control criteria. We then ran these simulated mono-infection samples through our analytical pipeline, blinded to the spiked assemblies to better simulate the diversity expected from clinical isolates.

Second, we assessed the fraction of reads that were lost to interference from a coinfection with a related organism. Reads were drawn randomly for each assembly in three ratios (1:3, 1:1 and 3:1) and then spiked into a healthy human plasma sample that was sub-sampled to the unique WINC count minimum. These read sets were run through the analytical pipeline blinded to both assemblies. Assemblies (125 pairs) were run, each with 1:1, 3:1 and 1:3 ratios. The fraction of reads lost due to interference during coinfection is the difference between the concentration observed during coinfection versus mono-infection, divided by the concentration observed during mono-infection.

Assessing MPM versus sequencing depth. To determine whether MPM is dependent on sequencing depth, we sub-sampled sequencing reads from a sample containing all 13 reference microorganisms in a medium human background and a unique WINC molecule count of 600,581. We then sampled (without replacement) the fraction of reads that resulted, on average, in unique WINC counts of 25,000, 50,000, 100,000, 250,000 and 500,000. We performed this sub-sampling ten times for each of the unique WINC target counts and re-executed the analytical portion of the test on these 50 collections of sampled reads. The results are shown in Supplementary Fig. 3b where we average the MPMs of the representative pathogens across the ten replicates at each sampling level.

Clinical samples for CMV-PCR comparison. Samples that were utilized in the quantitative evaluation of the assay included plasma from 25 CMV positive individuals obtained from a biorepository of de-identified remnant blood samples (Discovery Life Sciences) annotated with quantitative PCR test results (Roche cobas CMV, Roche Molecular Systems). All samples were obtained through Institutional Review Board (IRB)-approved protocols with appropriate informed consent provided for sample use (Schulman IRB).

Clinical validation. The SEP-SEQ study (NCT02730468) was a prospective observational study including patients presenting to the Emergency Department at the Stanford University Medical Center identified by a sepsis alert triage model designed for the early identification and intervention in patients with sepsis[51]. The study protocol and patient informed consent were reviewed and approved by the University's IRB (Stanford) and the study was conducted in compliance with the International Conference on Harmonization and Good Clinical Practice Guidelines. After obtaining consent, patients 18 years or older with a sepsis alert, a temperature of >38 °C or <36 °C and at least one of the following systemic inflammatory response syndrome criteria: heart rate >90 beats per minute; respiratory rate >20 breaths per minute or a partial pressure of carbon dioxide <32 mm Hg; white blood cell count of either > 12,000 cells μl^{-1} or <4,000 cells μl^{-1} and >10% bands were enrolled in the study. The clinical data collected included demographics, medical history, baseline characteristics and results from microbiological testing performed as standard care during the first seven days of the hospital stay.

SEP-SEQ sample size calculation and statistical analysis. Sample size was calculated to determine the sensitivity of the test compared with blood culture. A total of 59 culture-positive subjects were required to demonstrate a lower estimate of sensitivity of 75% (90% CI). The expected rate of positive blood cultures in this cohort was anticipated to be approximately 15% and we aimed to enrol 400 subjects. However, the observed blood-culture-positive rate was 18%, thus enabling a final enrolment of 350 subjects, 63 of which had a positive initial blood culture. The primary analysis for the assessment of test performance included all subjects with an available result for both the initial blood culture and a cfDNA sequencing blood sample collected at enrolment. The comparison of quantitative pathogen values between adjudicated groups was performed using the Wilcoxon rank-sum test.

Patient specimens. At enrolment, blood samples were collected in 2–6 ml BD Vacutainer K_2EDTA blood collection tubes (Becton Dickinson and Company) via peripheral blood draw or from an indwelling venous catheter in the Emergency Department. These samples were collected before in-hospital antibiotic administration, at the same time as routine blood cultures that were processed using the BD BACTEC FX blood culture system (Becton Dickinson and Company). Samples were assigned a de-identified code linked to the unique identifier number of the patient. Whole-blood samples were stored at 4 °C, processed to plasma by centrifugation (1,500 r.c.f. for 10 min) within 72 h of collection and stored at −80 °C until shipment to the Karius, Inc. laboratory for processing. Study samples were processed in batches, in the Karius clinical laboratory, using the same protocol used for commercial samples run in the same laboratory.

Assessment of performance. Test performance was assessed by comparing the results of microbial cfDNA sequencing with initial blood cultures, all microbiological testing was performed within seven days of enrolment (including tissue/fluid cultures, serology and nucleic acid testing) and a composite reference standard consisting of all microbiological test results plus clinical adjudication. Patients with negative results for both cfDNA sequencing and microbiological testing were considered negative and patients with ≥1 pathogen identified by cfDNA sequencing concordant with microbiological testing were considered to be definite infections. Single-bottle positive blood cultures with Gram-negative organisms (for example, *E. coli*) were considered positive and single-bottle positive blood cultures with common Gram-positive contaminants (for example, viridans group *Streptococci* and coagulase-negative *Staphylococcus*) were considered negative. The following microbiology tests were not considered when comparing results to microbial cfDNA sequencing: tests for RNA viruses (this assay detects only DNA organisms); stool *Clostridium difficile* toxin PCR test (*C. difficile* colitis is mediated by the *C. difficile* toxin and is non-invasive)[52] and Torque teno virus detected by cfDNA sequencing (considered to be non-pathogenic)[53]. The time to positive result for cfDNA sequencing was calculated using empirical observations of >400 samples processed through the CLIA laboratory in June and July of 2018 (median of 24 h for shipping plus a median of 29 h for testing or an additional median 54 h testing for samples that required repeat testing with a second aliquot derived from the original sample).

Clinical adjudication. A committee composed of three independent board-certified infectious disease physicians performed clinical adjudication (Supplementary Fig. 8). Following a standardized algorithm with a review of all microbiological testing results, radiological testing results and a summary of the hospital stay of the patient, the committee classified discordant cases according to the likelihood that the pathogen identified by cfDNA sequencing was driving the symptoms leading to the sepsis alert. Cases were classified as definite, probable, possible or unlikely using the following standardized criteria: (1) definite: the cfDNA sequencing pathogen result was concordant with a microbiological test performed within seven days of sample collection for cfDNA sequencing; (2) probable: the cfDNA sequencing pathogen result was considered the probable cause of the sepsis alert based on clinical, radiological or laboratory findings; (3) possible: the cfDNA sequencing pathogen result had potential for pathogenicity consistent with clinical presentation but an alternative explanation for the symptoms was more likely and (4) unlikely: the cfDNA sequencing pathogen result had potential for pathogenicity but findings were inconsistent with clinical presentation. The committee chair had final discretion for case classification.

Clinical samples for reference interval. The samples utilized in the determination of the reference range were collected from 167 healthy asymptomatic donors in five geographically diverse areas of the United States who were 18 to 65 years of age and had been screened for common health conditions including infectious diseases through a questionnaire and standard blood donor screening assays (Serologix and StemExpress). All samples were obtained through IRB-approved protocols with the appropriate informed consent provided for sample use (WIRB and Chesapeake IRB). Samples were collected in plasma preparation tubes (Becton, Dickinson and Company) via a peripheral blood draw. All samples were spun to plasma, shipped frozen on dry ice to the Karius laboratory and stored at −80 °C until analysis.

The 97.5th percentile of abundance (in MPM) was estimated for each taxon in the reportable range across the 167 asymptomatic donor samples. The MPMs for all taxa across all samples were used in the estimation of the reference intervals regardless of whether or not they would have been reported. The 97.5th percentile MPM value is provided with each call made in a clinical report.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability

The core software used as part of the Karius test is described in the Clinical-grade microbial cfDNA sequencing for infectious disease section in Methods, under the sub-sections Sequence data processing and alignment, Microorganism abundance estimation and Pathogen detection. The open source software includes the following external tools: bcl2fastq v2.17.1.14, Trimmomatic v0.32, Bowtie v2.2.4 and BLAST v2.2.30. A description of all open source code is included in Methods and further details are available on request. The proprietary portions of the code are not available.

Data availability

The data that support the findings of this study are available from the corresponding author on request. Sequencing data that support the finding of this study (with human reads removed) have been deposited in NCBI SRA and can be accessed with the BioProject identifier PRJNA507824.

Received: 15 May 2018; Accepted: 11 December 2018;
Published online: 11 February 2019

References

1. Christensen, K. L. et al. Infectious disease hospitalizations in the United States. *Clin. Infect. Dis.* **49**, 1025–1035 (2009).
2. Barlam, T. F. et al. Implementing an antibiotic stewardship program: guidelines by the Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America. *Clin. Infect. Dis.* **62**, e51–e77 (2016).
3. Liesenfeld, O., Lehman, L., Hunfeld, K. P. & Kost, G. Molecular diagnosis of sepsis: new aspects and recent developments. *Eur. J. Microbiol. Immunol.* **4**, 1–25 (2014).
4. Kumar, A. et al. Initiation of inappropriate antimicrobial therapy results in a fivefold reduction of survival in human septic shock. *Chest* **136**, 1237–1248 (2009).
5. Fenollar, F. & Raoult, D. Molecular diagnosis of bloodstream infections caused by non-cultivable bacteria. *Int. J. Antimicrob. Agents* **30**, S7–S15 (2007).
6. Mancini, N. et al. The era of molecular and other non-culture-based methods in diagnosis of sepsis. *Clin. Microbiol. Rev.* **23**, 235–251 (2010).
7. Fishman, J. A. Infection in solid-organ transplant recipients. *N. Engl. J. Med.* **357**, 2601–2614 (2007).
8. Tomblyn, M. et al. Guidelines for preventing infectious complications among hematopoietic cell transplantation recipients: a global perspective. *Biol. Blood Marrow Transplant.* **15**, 1143–1238 (2009).
9. Paul, M. et al. Systematic review and meta-analysis of the efficacy of appropriate empiric antibiotic therapy for sepsis. *Antimicrob. Agents Chemother.* **54**, 4851–4863 (2010).
10. Kumar, A. An alternate pathophysiologic paradigm of sepsis and septic shock: implications for optimizing antimicrobial therapy. *Virulence* **5**, 80–97 (2014).
11. Ramanan, P., Bryson, A. L., Binnicker, M. J., Pritt, B. S. & Patel, R. Syndromic panel-based testing in clinical microbiology. *Clin. Microbiol. Rev.* **31**, 1–28 (2018).
12. Schreckenberger, P. C. & McAdam, A. J. Point–counterpoint: large multiplex PCR panels should be first-line tests for detection of respiratory and intestinal pathogens. *J. Clin. Microbiol.* **53**, 3110–3115 (2015).

13. Kothari, A., Morgan, M. & Haake, D. A. Emerging technologies for rapid identification of bloodstream pathogens. *Clin. Infect. Dis.* **59**, 272–278 (2014).
14. 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.* **66**, 778–788 (2018).
15. Greninger, A. L. et al. Clinical metagenomic identification of *Balamuthia mandrillaris* encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. *Genome Med.* **7**, 113 (2015).
16. Wilson, M. R. et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* **370**, 2408–2417 (2014).
17. Schlaberg, R. et al. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Arch. Pathol. Lab. Med.* **141**, 776–786 (2017).
18. Naccache, S. N. et al. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. *Clin. Infect. Dis.* **60**, 919–923 (2015).
19. Stokowski, R. et al. Clinical performance of non-invasive prenatal testing (NIPT) using targeted cell-free DNA analysis in maternal plasma with microarrays or next generation sequencing (NGS) is consistent across multiple controlled clinical studies. *Prenat. Diagn.* **35**, 1243–1246 (2015).
20. Song, K., Musci, T. J. & Caughey, A. B. Clinical utility and cost of non-invasive prenatal testing with cfDNA analysis in high-risk women based on a US population. *J. Matern. Fetal Neonat. Med.* **26**, 1180–1185 (2013).
21. Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L. & Quake, S. R. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc. Natl Acad. Sci. USA* **105**, 16266–16271 (2008).
22. Schutz, E. et al. Graft-derived cell-free DNA, a noninvasive early rejection and graft damage marker in liver transplantation: a prospective, observational, multicenter cohort study. *PLoS Med.* **14**, e1002286 (2017).
23. Bloom, R. D. et al. Cell-free DNA and active rejection in kidney allografts. *J. Am. Soc. Nephrol.* **28**, 2221–2232 (2017).
24. De Vlaminck, I. et al. Noninvasive monitoring of infection and rejection after lung transplantation. *Proc. Natl Acad. Sci. USA* **112**, 13336–13341 (2015).
25. De Vlaminck, I. et al. Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. *Sci. Transl. Med.* **6**, 241ra277 (2014).
26. Snyder, T. M., Khush, K. K., Valantine, H. A. & Quake, S. R. Universal noninvasive detection of solid organ transplant rejection. *Proc. Natl Acad. Sci. USA* **108**, 6229–6234 (2011).
27. Aravanis, A. M., Lee, M. & Klausner, R. D. Next-generation sequencing of circulating tumor DNA for early cancer detection. *Cell* **168**, 571–574 (2017).
28. Lanman, R. B. et al. Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. *PLoS ONE* **10**, e0140712 (2015).
29. Bettegowda, C. et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra224 (2014).
30. Dawson, S. J. et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* **368**, 1199–1209 (2013).
31. Abril, M. K. et al. Diagnosis of *Capnocytophaga canimorsus* sepsis by whole-genome next-generation sequencing. *Open Forum Infect. Dis.* **3**, ofw144 (2016).
32. Hong, D. K. et al. Liquid biopsy for infectious diseases: sequencing of cell-free plasma to detect pathogen DNA in patients with invasive fungal disease. *Diagn. Microbiol. Infect. Dis.* **92**, 210–213 (2018).
33. Lefterova, M. I., Suarez, C. J., Banaei, N. & Pinsky, B. A. Next-generation sequencing for infectious disease diagnosis and management: a report of the association for molecular pathology. *J. Mol. Diagn.* **17**, 623–634 (2015).
34. Dunne, W. M. Jr, Westblade, L. F. & Ford, B. Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory. *Eur. J. Clin. Microbiol. Infect. Dis.* **31**, 1719–1726 (2012).
35. Kim, D. et al. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome* **5**, 52 (2017).
36. Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
37. Weiss, S. et al. Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol.* **15**, 564 (2014).
38. Naccache, S. N. et al. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *J. Virol.* **87**, 11966–11977 (2013).
39. *Infectious Disease Next Generation Sequencing Based Diagnostic Devices: Microbial Identification and Detection of Antimicrobial Resistance and Virulence Markers. Draft Guidance for Industry and Food and Drug Administration Staff* (Food and Drug Administration, 2016).
40. Chang, C. P. et al. Elevated cell-free serum DNA detected in patients with myocardial infarction. *Clin. Chim. Acta* **327**, 95–101 (2003).
41. Lo, Y. M., Rainer, T. H., Chan, L. Y., Hjelm, N. M. & Cocks, R. A. Plasma DNA as a prognostic marker in trauma patients. *Clin. Chem.* **46**, 319–323 (2000).
42. Vincent, J. L., Martinez, E. O. & Silva, E. Evolving concepts in sepsis definitions. *Crit. Care Nurs. Clin. North. Am.* **23**, 29–39 (2011).
43. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
45. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
46. Xia, L. C., Cram, J. A., Chen, T., Fuhrman, J. A. & Sun, F. Accurate genome relative abundance estimation based on shotgun metagenomic reads. *PLoS ONE* **6**, e27992 (2011).
47. Bennett J. E., Dolin, R. & Blaser, M. J. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases* 8th edn (Saunders, Philadelphia, 2015).
48. Burnham, P. et al. Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. *Sci. Rep.* **6**, 27859 (2016).
49. Kalia, V. C. et al. Analysis of the unexplored features of *rrs* (16S rDNA) of the Genus *Clostridium*. *BMC Genomics* **12**, 18 (2011).
50. *Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition. NCCLS document EP5-A2* (NCCLS, 2004).
51. Dellinger, R. P. et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med.* **39**, 165–228 (2013).
52. Bagdasarian, N., Rao, K. & Malani, P. N. Diagnosis and treatment of *Clostridium difficile* in adults: a systematic review. *JAMA* **313**, 398–408 (2015).
53. Focosi, D., Antonelli, G., Pistello, M. & Maggi, F. Torquetenovirus: the human virome from bench to bedside. *Clin. Microbiol. Infect.* **22**, 589–593 (2016).
54. Karius, Inc. *Pathogen List* https://www.kariusdx.com/pathogen-list/3.1.1 (2018).

Acknowledgements

The authors would like to thank S. Sinha for assistance with the preparation of the manuscript, as well as H. Quach, R. Davila, S. Madan, V. Baichwal, C. Ho, H. Seng, R. Aquino, A. Parham, R. Mann, I. Brown, P. Callagy, A. Visweswaran, C. Keller, C. Bucsit and A. Araya for their contributions to these validation studies.

Author contributions

S.T., M.J.R., L.B., M.S.L., I.D.V., T.K., F.C.C., S.V., G.D.W., A.C., Z.N.R., G.M.-S., L.H., S.Balakrishnan, J.V.Q., D.H., D.K.H. and M.L.V. designed and carried out experiments, analysed data and summarized the results. T.A.B., S.T., M.L.V., M.K., S.Bercovici, J.C.W. and S.Y. analysed data and supervised the work. T.A.B., S.Bercovici, S.T., D.H. and D.K.H. wrote the paper.

Competing interests

This study was funded by Karius, Inc. and describes the validation of a product developed by Karius, Inc. All authors (excepting S.T., J.V.Q. and S.Y.) are current or former employees and/or share -holders of Karius, Inc. This does not alter our adherence to Nature Microbiology policies on sharing data and materials.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-018-0349-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.A.B.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Timothy A. Blauwkamp

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	The core software used as part of the Karius test is described in the "Clinical-grade Microbial cfDNA Sequencing for Infectious Disease" portion of the methods section, under sub-sections "Sequence data processing and alignment", "Microorganism abundances estimation", and "Pathogen detection". The open source software includes the following external tools: bcl2fastq v2.17.1.14, Trimmomatic v 0.32, Bowtie v2.2.4, and BLAST v2.2.30. A description of all open source code is included in the methods, and further details are available upon request. The proprietary portions of the code are not available.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon request. Sequencing data that support the finding of this study (with human reads removed) have been deposited in NCBI SRA and can be accessed with [DOIs to be provided when complete].

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For all experiments, the sample size was informed by published CLSI guidelines for analytical validation experiments, peer-reviewed publications in related fields, and availability of clinical study samples.
Data exclusions	Criteria for data exclusion were pre-established and based on several quality control indicators. Any sample failing to meet quality control specifications for sequencing depth, sample-carryover limits, sample mix-ups or any sample contained within a batch where the positive or negative batch controls failed to meet the requirements for quality and accuracy were repeated. Any sample failing to meet quality control requirements upon repeat testing was excluded.
Replication	Indications of reproducibility are included with each experiment. The only experiments in which results were measured only once are those involving clinical study specimens. Even here, an indication of the reproducibility in testing of clinical samples was determined on twenty of these samples, as described in the precision testing section. The ability to replicate results obtained on other clinical samples is limited by the volume of sample available for repeated testing.
Randomization	Collection of plasma from self-reported healthy volunteer for the reference interval study was performed by a third party based on the criteria outlined in the materials and methods. Collection of clinical samples for comparison of testing to CMV-PCR was performed by a third-party with instructions to collect samples containing a full range of CMV infection levels, from very low to very high, as determined by CMV-PCR. The inclusion criteria for the 350 prospectively enrolled patients with clinical suspicion of sepsis are described in the methods section.
Blinding	The testing process is automated from sample preparation (performed on high throughput liquid handlers) through data analysis (performed by locked and version-controlled analysis pipelines), as are the criteria by which a test result is either accepted into the dataset or not. There is no human judgment involved in generating or accepting a test result. Nonetheless, all experimentation was performed while blinded to the results of blood culture and CMV-PCR testing for all clinical sample testing.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All materials used in this study are available from the standard commercial sources indicated in the manuscript, except for the clinical study samples. Due to extremely limited sample volumes for these clinical study samples, we are unable to make these available.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Samples utilized in the determination of the reference range were collected from 167 healthy asymptomatic donors in geographically diverse areas of the US, aged 18 to 65 years of age, who had been screened for common health conditions including infectious diseases via a questionnaire and standard blood donor screening assays. Samples utilized in the quantitative evaluation of the assay included plasma from 25 CMV positive individuals obtained from a biorepository of de-identified remnant blood samples for which age and sex of the donors was collected. Samples from patients within the SEP-SEQ cohort were aged 18 years or older with suspected sepsis evidenced by a temperature of >38℃ or <36℃ and at least one additional Systemic Inflammatory Response Syndrome (SIRS) criteria. The first 1500 samples (~50% pediatric) tested sequentially by the Karius commercial laboratory were most commonly from immunocompromised patients, followed by patient suspected of sepsis, endocarditis, and complicated pneumonia.
Recruitment	Samples utilized for the reference range were obtained from specimen procurement companies who recruited screened healthy asymptomatic adult volunteers presenting to geographically diverse regional U.S. blood centers for routine plasma donation. Patients for SEP-SEQ were recruited as part of a prospective, observational study including patients presenting to an academic Emergency Department (Stanford University Hospital, Stanford, CA) with a sepsis alert. The first 1500 clinical patient samples analyzed by Karius were received sequentially by the Karius commercial laboratory up to July 31, 2018.