Cheng HK, Tan SK, Sweeney TE, Jeganathan P, et al. Combined use of metagenomic sequencing and host response profiling for the diagnosis of suspected sepsis.

Supplementary Material

Supplementary Methods

Supplementary References

Supplementary Methods

Healthy Volunteer Subjects

Inclusion Criteria

1) Must be healthy adults between the ages of 18 and 75, and should not be under the care of a physician for any chronic condition.

2) Must be able to read, understand and sign the approved consent form.

3) Must be able and willing to follow study procedures and instructions.

4) Must be willing and able to transport stool samples from their home to the Stanford Center for Translational & Clinical Research.

Exclusion Criteria

1) Have used systemic or intra-oral antibiotics or antifungals within the 6-month period preceding study enrollment.

2) Require antibiotics before dental treatments.

3) Have fewer than 15 teeth.

4) Are pregnant

5) Have a condition known to compromise the immune system, including HIV infection.

Criteria regarding stool samples, antimicrobials, and oral health were included because the healthy volunteer samples in this study were also used for a concurrent study of bacterial translocation of microbial sequences in blood during states of health.

Of our 10 healthy volunteers, 6 were female. The median age was 32 years, with an interquartile range of 26.75 to 35.5 years.

mNGS Plasma Processing and Sequencing

Whole blood samples were obtained from 200 suspected sepsis patients in 6 mL EDTA tubes by nurses and phlebotomists at the Stanford ED, and immediately stored at 4°C. Plasma was prepared from whole blood by centrifuging (1,500 x g for 10 minutes) within 72 hours of collection and stored at -80C. Prior to extraction, plasma samples underwent an additional centrifugation step at 16,000 x g. DNA was extracted from 400 µL of plasma with the QIAGEN Circulating Nucleic Acid extraction kit. (In some patients, only 200-400 µL of plasma was available due to low blood draw volume.) Three negative controls (molecular-grade water drawn into an EDTA tube) were included in every extraction batch of 24 samples, for a total of 36 negative controls across all batches. Extracted DNA was quantified with the Quant-iT dsDNA Assay Kit, high sensitivity (ThermoFisher Scientific), and quality-controlled with the AATI Fragment Analyzer or 2100 Bioanalyzer using the Agilent High Sensitivity D1000 kit.

Libraries were prepared with the KAPA HyperPrep Kit (Roche) at the High-Throughput Sequencing and Genotyping Unit (HTSGU) at the University of Illinois at Urbana-Champaign. The starting amount of plasma DNA ranged from 4.96 ng to 0.669 ng, with one outlying sample at 1.542 g. All libraries were prepared using the Kappa Hyper Kit (Roche) without size selection and sequenced on the HiSeq 4000 (Illumina) with 2x150 nucleotide paired-end reads.

In a pilot experiment, we sequenced DNA from a plasma sample from each of 15 patients with a positive blood culture, at a depth of 40-60M reads/sample, and from 4 negative controls at a depth of 2-6M reads/sample, using unique single-indexed 10-nt barcodes. Because of the

potential impact of barcode hopping[1] on sequencing data, we used unique dual-indexed barcodes for library prep for the remaining plasma samples from the other 185 patients, and obtained 10-52M reads/sample, as well as 3-6M reads/sample for each of 36 negative controls.

Bioinformatics. Reads were de-multiplexed with Illumina software, adapters were removed with SeqPrep (SeqPrep; https://github.com/jstjohn/SeqPrep), low-quality bases were removed with Sickle (Sickle; https://github.com/najoshi/sickle), and human reads were subtracted with bowtie2[2] under default parameters. Kraken[3] was run on all non-human reads using a database of all complete bacterial genomes and viral genomes from RefSeq, and all human, protozoa, archaeal, and fungal genomes (including contigs, scaffolds, and/or chromosomes) from RefSeq downloaded in July 2016. All genomes used in our Kraken database had low-complexity regions masked with DUST[4]. The human genome was added to reduce false-positive eukaryotic pathogen classifications, and a conservative filtering threshold of 0.3 was applied to reduce false-positive classifications during Kraken alignments. Finally, bacterial reads that were classified to the species-level by Kraken were imported into R and further analyzed with phyloseq[5].

We did not further analyze Kraken results for eukaryotic organisms because we found a high number of eukaryote reads across all plasma samples. We believe that the vast majority of these reads do not represent true infections but are most likely human reads that were not detected by bowtie2, or misalignment errors due to Kraken or to the RefSeq database. Additionally, we decided not to further analyze archaea reads because of the paucity of evidence for an archaeal species acting as a human pathogen[6].

In an exploratory analysis to identify potential batch effects, we considered a count table of 1,644 species in 200 plasma samples and 42 negative controls. Then, we reduced the number

to 1,248 species by removing species that were not detected in any of the plasma samples. Next, we ranked species abundance in each sample, where the species with the largest abundance was assigned the largest rank. To reduce the artificially large difference in ranks, species with rank below some threshold (900) were set at one[7]. PCA performed on a truncated-ranking transformation showed possible batch-effects which may have contributed to variation in sample sequence composition (figure S1). Two distinct clusters were detected when samples were grouped into sets based on their extraction batches: Set 1, consisting of 90 plasma samples and 18 negative controls from extraction batches 1, 2, 3, 4, 11, and 12; and Set 2, which consisted of 95 plasma samples and 18 negative controls from extraction batches 5, 6, 7, 8, 9, and 10. The Pilot Set, consisting of 15 plasma samples and 6 negative controls of the pilot experiment, clustered with Set 1.

Rather than use a log scale for visualization (because of its crushing effect on intermediate-abundance species), we used arcsinh. We accounted for the unequal library depths using the median-of-ratios method[8]. Closer examination of species in the negative controls of Sets 1 and 2 revealed distinct contamination signatures, with numerous high-abundance taxa unique to each Set (figure S2). We hypothesized that differences in manufacturing lots of the nucleic acid extraction kits may have caused the variation in sample sequence composition, as Glassing et al. have previously reported[9].

To distinguish blood-associated DNA sequences from contaminant sequences in plasma samples, we developed a Bayesian statistical method that leverages data from negative control samples. A detailed description is provided in appendix 1, and an open-source R package of this method is available at **https://github.com/PratheepaJ/BARBI**. We ran the contaminant removal method within each batch, Set 1, Set 2, and the Pilot Set, with the observed abundance

data. We analyzed the Pilot Set separately, even though it behaved similarly to Set 1, because the two sets were extracted by separate technicians and sequenced using different barcode adapters several months apart.

VirCapSeq-VERT High Throughput Sequencing

Plasma samples (150 µl) were mixed with NucliSens buffer and total nucleic acid extracted on the easyMag instrument (bioMerieux). Ten microliters of extract were subjected to reverse transcription with random hexamer priming (SuperScript III, Thermo Fisher) and second strand DNA synthesis with Klenow fragment polymerase (New England Biolabs). The resulting cDNA/DNA preparation was fragmented by sonication to an average size of 250 bp (E210 sonicator, Covaris), purified (AxyPrep), and up to 50 ng sheared product (Qubit) used for library preparation with KAPA kits (Hyper Library Preparation kit, KAPA Biosystems) and custom dual uniquely indexed barcode adaptors (Integrated DNA Technologies). The libraries were evaluated for quality and quantity by TapeStation (4200 System, Agilent) and pooled at equimolar quantities for hybridization with the ~2 million oligonucleotides comprising the VirCapSeq-VERT biotinylated probe library (47ºC, ON; NimbleGen/Roche). Each pool included a negative control (Salmon nucleic acid) that had been processed alongside the samples in the pool. Sequences hybridized to biotinylated probes were collected by magnetic streptavidin beads (DynaMag-2 magnet; Thermo Fisher), washed, and on-bead amplified by low cycle post-hybridization PCR (SeqCap EZ accessory kit V2; NimbleGen/Roche). Amplification products were purified (Agencourt Ampure beads; Beckman Coulter) and quantitated (TapeStation) for sequencing on HiSeq 2500 sequence analyzer (Illumina).

Bioinformatics analysis. Sequence reads were demultiplexed with Illumina software, Q30-filtered, and further cleaned by PRINSEQ v20.2[10]. Sequence data were then depleted of host background by alignment to human reference sequences downloaded from the NCBI database. Host-depleted reads were *de novo* assembled with MIRA v4.0[11] and the resulting contigs as well as remaining unique singletons subjected to homology search by MegaBlast against the NCBI non-redundant nucleotide database. Sequences that showed poor or no homology at the nucleotide level were screened by BLASTX against the viral protein database, and subsequently the whole database to exclude forced alignments and potentially false positives. Based on BLAST results the best matching NCBI sequence entries were identified and downloaded as scaffolds for mapping the entire data set to recover partial or complete genome sequences (Bowtie2 mapper 2.0.6; http://bowtie-bio.sourceforge.net). SAMtools v0.1.19[12] were used to generate consensus genomes and coverage statistics. Geneious (v10; www.geneious.com) or Tablet[13] were employed to visualize and evaluate read mappings.

Read yields were normalized to 10,000 host-depleted total reads and a positive viral signal was assigned to samples with a normalized read count >0.2 after subtraction of reads occasionally recorded in the negative Salmon nucleic acid control and for which these reads distributed to at least three genome regions.

Host Response Calibration Chart Review

Four physicians with subspecialty training in infectious diseases performed retrospective physician chart reviews to establish likely admission diagnoses while being blinded to sequencing and host response assay results. Each physician was assigned to 100 patients such that two physicians reviewed each patient. If there were discrepancies among the two physicians'

classifications, both physicians met in person to discuss and adjudicate the classification. The following questions were reviewed during this first chart review:

1. Infection status at time of enrollment?
    a. Answer choices: Yes, Possible, No
2. Lab evidence of a clinically significant **bacterial** infection from specimen taken within first 5 days after enrollment? (Select yes if patient w/lab evidence is possibly infected.)
    a. Answer choices: Yes, No.
3. Lab evidence of a clinically significant **viral** infection from specimen taken within first 5 days after enrollment? (Select yes if patient w/lab evidence is possibly infected.)
    a. Answer choices: Yes, No.
4. Lab evidence of a clinically significant **fungal** infection from specimen taken within first 5 days after enrollment? (Select yes if patient w/lab evidence is possibly infected.)
    a. Answer choices: Yes, No.
5. Lab evidence of a clinically significant **parasitic** infection from specimen taken within first 5 days after enrollment? (Select yes if patient w/lab evidence is possibly infected.)
    a. Answer choices: Yes, No.

All questions were in regard to the patient's presentation, i.e. SIRS or sepsis.

Physician Chart Review

After obtaining sequencing and host response assay results, the main retrospective chart review was performed for all 200 patients by three additional physicians with specialty training in

infectious diseases. Information provided to physicians for interpreting host response and sequencing results in this study are provided in appendix S2. The questions in the chart review are provided in appendix S3 and summarized below.

Phase I: First, physicians were provided the patient's medical record (while blinded to mNGS, VirCapSeq-VERT, and host response results), and asked to assess the following:

1. Whether the patient had an infection, and if so, a bacterial infection, viral infection, fungal infection, or parasitic infection.

Phase II: Next, physicians were provided with mNGS and VirCapSeq-VERT results alongside the patient's medical records, and asked to assess the following:

1. Whether the patient had an infection, and if so, a bacterial infection, viral infection, fungal infection, or parasitic infection.
2. Whether each of the patient's positive mNGS and VirCapSeq-VERT organisms, if any, were clinically relevant.

Phase III: Finally, physicians were provided host response results alongside the patient's medical charts, mNGS, and VirCapSeq-VERT results, and asked to assess the following:

1. Whether the patient had an infection, and if so, a bacterial infection, viral infection, fungal infection, or parasitic infection.
2. Whether each of the patient's positive mNGS and VirCapSeq-VERT organisms, if any, were clinically relevant.

For all questions, the physicians were provided with the following answer choices on a 5-point scale: Yes, Probably Yes, Unsure, Probably No, and No. All questions were in regard to the patient's presentation, i.e. SIRS or sepsis. Phase III was not conducted for the 93 patients whose host response scores were used to generate score cutoffs, nor the additional 7 patients with no host response scores due to PCR errors.

After gathering the results of Phase I, we grouped patients into categories of definite infection status (table 1 and figures 2-4) using the following guidelines:

1. Noninfected: Requires a "No" for infection question by at least two of three physicians.
2. Bacterial: Requires a "Yes" for bacterial infection question by at least two of three physicians.
3. Viral: Requires a "Yes" for viral infection question by at least two of three physicians.
4. Fungal: Requires a "Yes" for fungal infection question by at least two of three physicians.
5. Bacterial-Viral Coinfection: Requires a "Yes" for both bacterial and viral infection questions by at least two of three physicians.
6. Bacterial-Fungal Coinfection: Requires a "Yes" for both bacterial and fungal infection questions by at least two of three physicians.
7. Probable or Uncertain: Requires any choice but "No" for infection status question, and any choice but "Yes" for remaining questions by at least two of three physicians.

Appendix S1. Description of the Contaminant Removal Method

McMurdie and Holmes[7] proposed the use of simple gamma-Poisson mixtures (negative binomial) to model microbiome count data. Following their approach, we modelled the data generating process (factoring out the library depth effect) of species-specific reads in a plasma sample as the sum of two independent Poisson distributions that included 1) **true** reads belonging to the plasma sample, and 2) reads originating from contamination sources. Each of the Poisson distribution intensity parameters (intensity of true reads and contaminant reads) was considered to come from a gamma distribution. If we only made technological replicates with fixed library depths, we would observe a number of reads K_{ij} which is the sum of two independent Poisson random variables; one with $\lambda_{ij}^{(r)}$ as the true intensity parameter and $\lambda_{i}^{(c)}$ as the contaminant intensity parameter for each species i in plasma sample j. In reality, we observed random reads with biological variation and unequal library depth S_j for each plasma sample j. We estimated the effect of library depth using the negative controls and median-of-ratios method that benefits from the scaling property of the gamma distribution. Based on this mixture model and our observed data, we provided a Bayesian method[14,15] for inferring the true intensity of the plasma sample microbial DNA in the presence of microbial DNA contamination using the negative controls. First, we defined a prior density for the contaminant intensity in a plasma sample for each of the species using the negative controls. Next, we found the marginal likelihood and the marginal reference prior for the true intensity in the plasma sample for each of the species. Then, using Bayesian reference analysis, we obtained the marginal posterior for the true intensities up to a constant. Finally, we used the Metropolis-Hasting (MH) Markov Chain Monte Carlo (MCMC) method to sample from the marginal posterior of the true intensity.

Then we used the Bayesian method to estimate the marginal posterior for the true intensity for a given plasma sample.

Table S.1: Count matrix $K \in \mathbb{R}^{m \times N}$

Species	Plasma_1	Plasma_2	$\cdots$	Plasma_{n_1}	$\text{Control}_{(n_1+1)}$	$\text{Control}_{(n_1+2)}$	$\cdots$	Control_N
Species_1	K_{11}	K_{12}	$\cdots$	K_{1n_1}	$K^0_{1(n_1+1)}$	$K^0_{1(n_1+2)}$	$\cdots$	K^0_{1N}
Species_2	K_{21}	K_{22}	$\cdots$	K_{2n_1}	$K^0_{2(n_1+1)}$	$K^0_{2(n_1+2)}$	$\cdots$	K^0_{2N}
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
Species_i	K_{i1}	K_{i2}	$\cdots$	K_{in_1}	$K^0_{i(n_1+1)}$	$K^0_{i(n_1+2)}$	$\cdots$	K^0_{iN}
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
Species_m	K_{m1}	K_{m2}	$\cdots$	K_{mn_1}	$K^0_{m(n_1+1)}$	$K^0_{m(n_1+2)}$	$\cdots$	K^0_{mN}

Table S.1 shows the count matrix of n_1 plasma samples, n_2 negative controls and m species, where K_{ij} is the number of reads of species i in the j-th plasma sample whose true prevalence we suppose to be μ_{ij} and whose dispersion parameter is γ_i, and K^0_{il} is the number of reads of species i in the l-th negative control with prevalence μ^0_{il} and dispersion γ^0_i. In notation, $K_{ij} \sim \text{NB}\left(d_j\mu_{ij}, \gamma_i\right)$ and $K^0_{il} \sim \text{NB}(d^0_l\mu^0_{il}, \gamma^0_i)$, where d_j and d^0_l are the linear scaling factors for plasma sample j and negative control l that account for the library depths S_j and S^0_l.

If there is no contamination, the hierarchical mixture model for a plasma sample gives the number of reads K_{ij} as $\text{Poisson}(\lambda_{ij} d_j)$ and $\lambda_{ij} \sim \text{gamma}\ (\alpha_{ij}, \beta_{ij})$, where λ_{ij} is the true intensity of species i in plasma sample j after factoring out the library depth effect d_j.

In the presence of contamination, we considered the observed reads in a plasma sample to be a mixture of the true and contaminant reads. Thus, we modelled K_{ij} as the sum of two independent Poisson random variables with two different intensities and we write, $K_{ij} = K_{ij}^{(r)} + K_{ij}^{(c)}$: 1) the true intensity parameter is $\lambda_{ij}^{(r)}$ and 2) the contaminant intensity parameter is $\lambda_{ij}^{(c)}$. We assumed $\lambda_{ij}^{(c)}$ follows a gamma $\left(\alpha_{ij}^{(c)}, \beta_{ij}^{(c)}\right)$ distribution, encoding our prior degree of belief of the contaminant intensity, and we wrote the contaminant parameters specific to species i in each plasma sample j: $\alpha_{ij}^{(c)}$ and $\beta_{ij}^{(c)}$. Then, we estimated these parameters using the negative controls as in (14).

Given the plasma sample, $K_j = [K_{1j}, K_{2j}, \cdots, K_{mj}]^T$, where $\sum_{i=1}^{m} \mathbb{E}\left[K_{ij}\right] = S_j$, $\mathbb{E}\left[S_j\right] = d_j \sum_{i=1}^{m} \mu_{ij}$, and S_j is the library depth of the j-th plasma-sample, the model for the number of reads of each species i in plasma sample j is written according to the hierarchical model

$$\begin{aligned} K_{ij} \mid \left(\lambda_{ij}^{(r)} + \lambda_{ij}^{(c)}\right) d_j &\sim \text{Poisson}\left(\left(\lambda_{ij}^{(r)} + \lambda_{ij}^{(c)}\right) d_j\right), \\ \Pi\left(\lambda_{ij}^{(r)}\right) &= \frac{\left|I\left(\lambda_{ij}^{(r)}\right)\right|^{1/2}}{|I(0)|^{1/2}}, \\ \lambda_{ij}^{(c)} &\sim \text{gamma}\left(\alpha_{ij}^{(c)}, \beta_{ij}^{(c)}\right), \end{aligned} \tag{1}$$

where $\Pi\left(\lambda_{ij}^{(r)}\right)$ is a marginal reference prior for the true intensity and $I(\cdot)$ is the Fisher information obtained through their marginal probability density.

Using the negative controls in Table S1, we estimated the prior density of the contaminant intensities $\lambda_{ij}^{(c)} \sim \text{gamma}\left(\alpha_{ij}^{(c)}, \beta_{ij}^{(c)}\right)$. Then, using this prior information, we derived the marginal reference prior for true intensities.

By considering the contaminant intensity as a nuisance parameter and our knowledge that $\lambda_{ij}^{(c)} \sim \text{gamma}\left(\alpha_{ij}^{(c)}, \beta_{ij}^{(c)}\right)$, the marginal model for the true intensity is

$$p\left(k_{ij}|\lambda_{ij}^{(r)} d_j\right) = \int_0^\infty \text{Poisson}\left(k_{ij}|\left(\lambda_{ij}^{(r)} + \lambda_{ij}^{(c)}\right)d_j\right) \text{gamma}\left(\lambda_{ij}^{(c)}|\alpha_{ij}^{(c)}, \beta_{ij}^{(c)}\right) d\lambda_{ij}^{(c)}. \qquad (2)$$

Since we can estimate the contaminant intensities as $\lambda_{ij}^{0(c)} = \frac{\alpha_{ij}^{(c)}}{\beta_{ij}^{(c)}}$ using negative controls, (2) can be simplified using a Dirac delta function

$$\begin{aligned} p\left(k_{ij}|\lambda_{ij}^{(r)} d_j\right) &= \int_0^\infty \text{Poisson}\left(k_{ij}|\left(\lambda_{ij}^{(r)} + \lambda_{ij}^{(c)}\right)d_j\right) \delta\left(\lambda_{ij}^{(c)} - \lambda_{ij}^{0(c)}\right) d\lambda_{ij}^{(c)} \\ &= \text{Poisson}\left(k_{ij}|\left(\lambda_{ij}^{(r)} + \lambda_{ij}^{0(c)}\right)d_j\right). \qquad (3) \end{aligned}$$

Then we used the marginal model in (3) to compute the Fisher information:

$$\left|I\left(\lambda_{ij}^{(r)}\right)\right| = -\mathbb{E}\left[\frac{\partial^2}{\partial\left(\lambda_{ij}^{(r)}\right)^2}\log p\left(k_{ij}|\lambda_{ij}^{(r)}d_j\right)\middle|\lambda_{ij}^{(r)}\right] = \frac{1}{\left(\lambda_{ij}^{(r)}+\lambda_{ij}^{0(c)}\right)d_j}. \quad (4)$$

Thus, the reference prior for the true intensity was

$$\Pi\left(\lambda_{ij}^{(r)}\right) = \frac{\left|I\left(\lambda_{ij}^{(r)}\right)\right|^{\frac{1}{2}}}{|I(0)|^{\frac{1}{2}}} = \sqrt{\frac{\lambda_{ij}^{0(c)}}{\lambda_{ij}^{0(c)}+\lambda_{ij}^{(r)}}}, \quad (5)$$

where $\lambda_{ij}^{0(c)} = \frac{\alpha_{ij}^{(c)}}{\beta_{ij}^{(c)}}$.

By Bayes' theorem, the joint posterior density of $\lambda_{ij}^{(r)}$ and $\lambda_{ij}^{(c)}$ is

$$\begin{aligned} p\left(\lambda_{ij}^{(r)}d_j, \lambda_{ij}^{(c)}d_j|k_{ij}\right) &\propto p\left(k_{ij}|\left(\lambda_{ij}^{(r)}+\lambda_{ij}^{(c)}\right)d_j\right)p\left(\lambda_{ij}^{(r)}, \lambda_{ij}^{(c)}\right) \\ &= p\left(k_{ij}|\left(\lambda_{ij}^{(r)}+\lambda_{ij}^{(c)}\right)d_j\right)\Pi\left(\lambda_{ij}^{(c)}|\lambda_{ij}^{(r)}\right)\Pi\left(\lambda_{ij}^{(r)}\right), \end{aligned} \quad (6)$$

where $\Pi\left(\lambda_{ij}^{(c)}|\lambda_{ij}^{(r)}\right)$ is the conditional prior density for the contaminant intensity and $\Pi\left(\lambda_{ij}^{(r)}\right)$ is the marginal prior density for the true intensity.

With the reference prior in (5), the estimate for the contaminant intensities $\lambda_{ij}^{0(c)}$, and the assumption that $\lambda_{ij}^{(r)}$ and $\lambda_{ij}^{(c)}$ are independent, the joint posterior was

$$p\left(\lambda_{ij}^{(r)}d_j, \lambda_{ij}^{(c)}d_j|k_{ij}\right) \propto p\left(k_{ij}|\left(\lambda_{ij}^{(r)}+\lambda_{ij}^{(c)}\right)d_j\right)\delta\left(\lambda_{ij}^{(c)}-\lambda_{ij}^{0(c)}\right)\sqrt{\frac{\lambda_{ij}^{0(c)}}{\lambda_{ij}^{0(c)}+\lambda_{ij}^{(r)}}}. \quad (7)$$

Hence, the marginal posterior up to a constant for the true intensities was obtained by integrating (7) with respect to $\lambda_{ij}^{(c)}$

$$p\left(\lambda_{ij}^{(r)} d_j \middle| k_{ij}\right) \propto \text{gamma}\left(\left(\lambda_{ij}^{(r)} + \lambda_{ij}^{0(c)}\right) d_j \middle| \left(k_{ij} + .5\right), 1\right). \quad (8)$$

Finally, the marginal posterior for the true intensities is

$$\begin{aligned} p\left(\lambda_{ij}^{(r)} \middle| k_{ij}\right) &\propto \text{gamma}\left(\lambda_{ij}^{(r)} + \frac{\alpha_{ij}^{(c)}}{\beta_{ij}^{(c)}} \middle| \left(k_{ij} + .5\right)/d_j, 1\right) \quad \text{when} \quad k_{ij} \neq 0, \\ p\left(\lambda_{ij}^{(r)} \middle| 0\right) &\propto \text{gamma}\left(\lambda_{ij}^{(r)} \middle| .5/d_j, 1\right) \quad \text{when} \quad k_{ij} = 0, \end{aligned} \quad (9)$$

where $j = 1, \ldots n_1 \; and \;\; i = 1, \ldots, m.$

Next, we estimated $\alpha_{ij}^{(c)}$ and $\beta_{ij}^{(c)}$ in (9) using the negative controls. Then, we could sample from the marginal posterior distributions for the true intensities as in (14) by plugging in the estimates. Now we show how we estimated $\alpha_{ij}^{(c)}$ and $\beta_{ij}^{(c)}$.

We used all negative control samples to estimate the mean prevalence μ_{il}^0, dispersion γ_i^0 and library depth scaling factor d_l^0 for each species i in the l-th negative control using the negative binomial model

$$K_{il}^0 \sim \text{NB}(d_l^0 \mu_{il}^0, \gamma_i^0), \quad \text{where} \quad l = 1, \ldots n_2. \quad (10)$$

The negative binomial model in (10) could be written as a gamma-Poisson mixture

$$\begin{aligned} K_{il}^0 | \lambda_{il}^{(c)} d_l^0 &\sim \text{Poisson}\left(\lambda_{il}^{(c)} d_l^0\right) \\ \lambda_{il}^{(c)} &\sim \text{Gamma}\left(\alpha_{il}^0, \beta_{il}^0\right), \end{aligned}$$

where and α_{il}^0 and β_{il}^0 are shape and rate parameters, respectively, so we know that

$$\alpha_{il}^0 = \frac{1}{\gamma_i^0},$$

and

$$\beta_{il}^0 = \frac{1}{\gamma_i^0 \mu_{il}^0}.$$

That is,

$$\lambda_{il}^{(c)} \sim \text{gamma}\left(\frac{1}{\gamma_i^0}, \frac{1}{\gamma_i^0 \mu_{il}^0}\right). \quad (11)$$

We removed the library depth effect using the median-of-ratios method by computing

$$d_l^0 = \text{median}_{i:\overline{K_\iota} \neq 0} \frac{K_{il}^0}{\overline{K}_\iota}, \quad (12)$$

where $\overline{K}_\iota = \left(\prod_{l=1}^{n_2} K_{il}^0\right)^{1/n_2}$. We estimated γ_i^0 using the three-step procedure in Love et. al. (2014)[16] that depends both on the library depth scaling factor d_l^0 and mean prevalence of species i in negative control l, μ_{il}^0.

To define the prior density for the contaminant intensity in a plasma sample from $\lambda_{il}^{(c)} \sim$ gamma $\left(\frac{1}{\gamma_i^0}, \frac{1}{\gamma_i^0 \mu_{il}^0}\right)$, we assumed that contamination is plasma sample-dependent only through the library depth scaling factor, i.e., $\frac{\lambda_{il}^{(c)}}{d_j} d_l^0 = \lambda_{ij}^{(c)}$.

Using the scaling property of the gamma distribution (that only changes the shape parameter), we obtained the prior density for the contaminant intensity in plasma sample j as

$$\lambda_{ij}^{(c)} \sim \text{gamma}\left(\frac{d_l^0}{d_j}\frac{1}{\gamma_i^0}, \frac{1}{\gamma_i^0 \mu_{il}^0}\right). \quad (13)$$

From (13) we chose l that gives the median of $\frac{d_l^0 \mu_{il}^0}{d_j}$, where $l = 1, \ldots, n_2$.

Now we can write $\alpha_{ij}^{(c)}$ and $\beta_{ij}^{(c)}$

$$\alpha_{ij}^{(c)} = \frac{d_l^0}{d_j}\frac{1}{\gamma_i^0} \quad \text{and} \quad \beta_{ij}^{(c)} = \frac{1}{\gamma_i^0 \mu_{il}^0}. \quad (14)$$

We plugged in $\hat{\alpha}_{ij}^{(c)}$ and $\hat{\beta}_{ij}^{(c)}$ to the formula in (9) and used MCMC to sample from the marginal posterior for the true intensity.

Finally, we could compute the 95% highest posterior density interval for the true intensity $\left(L_{ij}^{(r)}, U_{ij}^{(r)}\right)$ and 95% highest density interval for the contaminant intensity $\left(L_{ij}^{(c)}, U_{ij}^{(c)}\right)$ for each different species in a plasma sample. Species with a lower limit $L_{ij}^{(r)}$ smaller than the upper limit $U_{ij}^{(c)}$ were identified as contaminants. This meant that there was a 95% chance that the species was a contaminant.

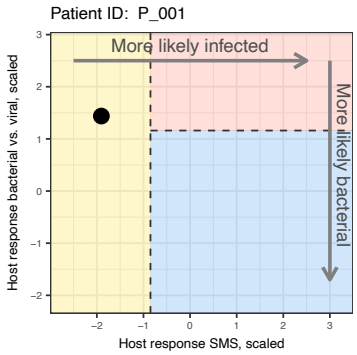
Appendix S2. Host Response, mNGS, and VirCapSeq-VERT Interpretation Information Provided to Physician Chart Reviewers

Host Response Interpretation Information

Background: The Integrated Antibiotic Decision Maker (IADM) is an 18-gene qRT-PCR host response assay that was developed by Sweeney et al.[17]. The assay consists of the 11-gene Sepsis MetaScore, which distinguishes between infection and non-infectious causes of inflammation, as well as the 7 gene Bacterial/Viral metaScore, which distinguishes between bacterial and viral infections.

Sample Type Profiled: Whole blood.

Interpretation: An example of a host response score in a patient is shown below. Two host-response scores are shown: the X axis measures the likelihood that an infection is present (as opposed to non-infectious cause of inflammation) using the SMS; the higher the score, the more likely an infection is present. The Y axis measures whether the infection is more likely bacterial or viral using the bacterial/viral metascore; the lower the score, the more likely the infection is bacterial. Note that 'borderline' bacterial-viral cases may indicate a weak signal, or may indicate a co-infection.



Using data from the host response calibration chart review, the cutoffs for this study were set locally to achieve a 95% sensitivity for bacterial infection when considering all three classes (bacterial, viral, non-infected).

A prior manuscript (Sweeney et al., 2016, *Science Translational Medicine*[17]) showed that in a pooled analysis of publicly-available microarray data consisting of 1,057 samples from 20 cohorts, the IADM had 94.0% sensitivity and 59.8% specificity for bacterial infections; 53.0% sensitivity and 90.6% specificity for viral infections; and 43.0% sensitivity and 97.3% specificity for noninfectious causes of inflammation compared to retrospective chart review adjudication. The manuscript also validated the IADM on 96 pediatric patient samples using the nanoString qRT-PCR platform and showed a 89.7% sensitivity and 70.0% specificity for bacterial infections; 54.5% sensitivity and 96.5% specificity for viral infections; and 61.1% sensitivity and 91.7% specificity for noninfectious SIRS.

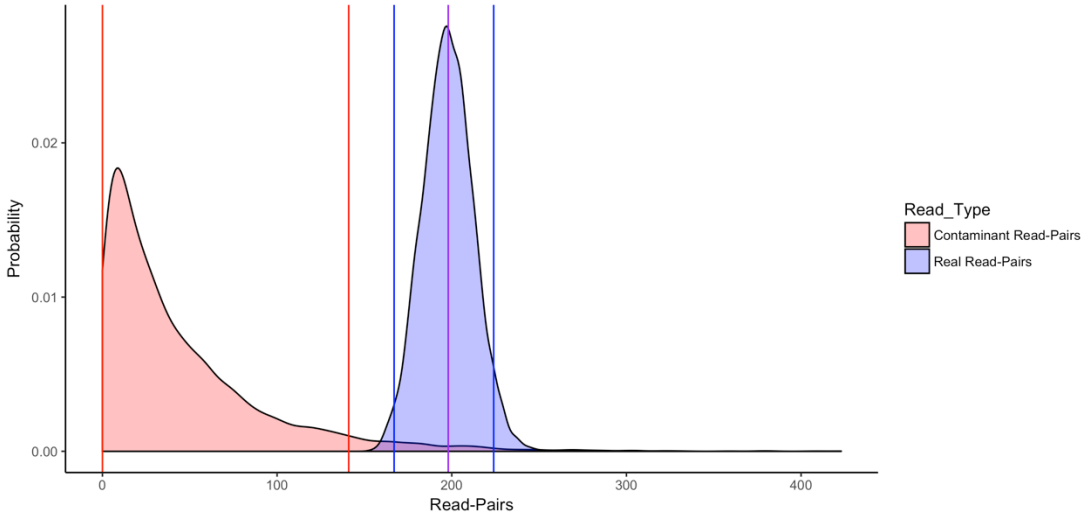
This assay was developed from analysis of transcriptomic data using 922 adult and pediatric patients from 14 cohorts, and has so far been retrospectively validated in 2,452 patients from 38 independent cohorts. It is not yet known how this host response assay performs in complex patient populations, including immunocompromised patients.

mNGS Interpretation Information

Background: We performed metagenomic Next-Generation Sequencing (mNGS) of cell-free DNA from human plasma. DNA was sequenced to a depth of 10-70 million 2x150 nucleotide read-pairs per sample. We processed and sequenced molecular grade water in parallel as negative controls in order to characterize contaminating DNA from the lab environment and reagents. We developed and applied a statistical model to estimate and remove sequencing reads based on negative control data. Only bacterial species are presented in the final results.

Sample Type Profiled: DNA extracted from plasma.

Interpretation: The algorithm relies on the assumption that the number of raw reads of a particular species of a particular sample is equal to the sum of 1) real read-pairs that are truly present in the plasma sample, and 2) contaminating read-pairs introduced. For each species in each sample, the algorithm estimates a distribution for the number of real read-pairs, as well as a distribution for the number of contaminating read-pairs. An example of these two distributions is presented below for *E. coli* in a patient with a positive *E. coli* blood culture. Vertical red and blue bars indicating 95th percentile chance limits of the two estimated distributions, and the purple vertical line indicating the number of raw read-pairs. If the lower-limit of the estimated real read-pairs exceeds the upper-limit of the estimated contaminant read-pairs, the species was considered "positive" in the sample.



You will be provided with a table of only the positive species in each sample, along with a number of values.

An example of the positive results table from the patient mentioned above:

	Species	Raw Reads	Lower Limit Real	Upper Limit Contaminant	In Neg. Cont.
1	s__Escherichia_coli	198	171	141	Yes
2	s__Pantoea_sp_PSNIH1	31	21	4	Yes

1. Raw reads, or the total number of reads of the species that was present in the sample.

2. Lower Limit Real, or the lower limit of the estimated real reads.

3. Upper Limit Contaminant, or the upper-limit of the estimated contaminant reads.

4. In Neg. Cont., or whether the taxon was present at all in at least one of the negative control samples.

While we made concerted efforts to control for microbial contamination, some positive organisms in the results may still represent only contaminants, as we may not have enough appropriate negative controls to gain the statistical power for perfect discrimination. Additionally, misclassifications are known to occur in NGS due to limitations in bioinformatic

tools, particularly at low read counts. Thus, we suggest that you use caution in interpreting taxa if either 1) the number of raw reads, or 2) the difference between the lower-limit of the estimated real reads ("L. Real") and the upper-limit of the estimated contaminant reads ("U. Contam.") columns, are less than 5-10 reads.

Finally, because there was heavy human genome contamination and no enrichment for microbial sequences, sensitivity will be low. We recommend against relying on the negative predictive value of NGS.

Note: The above "mNGS Interpretation Information" section was what was provided to physicians to interpret mNGS results. While writing our manuscript, we made small terminology changes:

1. *"Estimated real/contaminant reads" are instead referred to as "true/contaminant intensity" in the rest of the manuscript.*
2. *"Distribution for the number of real read-pairs" is instead referred to as "posterior distribution for the true intensities" in the rest of the manuscript.*
3. *Read-pairs are instead referred to as "reads" in the rest of the manuscript.*

VirCapSeq-VERT Information

Background: VirCapSeq-VERT, or viral capture sequencing of vertebrate viruses, is a highly sensitive viral sequencing assay first introduced in 2015 by the Lipkin Lab at Columbia University (Briese et al., 2015, *mBio*[18]). Oligonucleotide probes were used to enrich for DNA and RNA of full genomes from 207 viruses known to infect vertebrates, including humans, and

enriched nucleic acids were sequenced to roughly 10 million single-ended 100 bp reads per sample.

Sample Type Profiled: DNA and RNA extracted from plasma.

Interpretation: VirCapSeq-VERT was shown in a previous study to have a 1,000 to 10,000-fold enrichment over conventional viral sequencing techniques, based on experiments with human lung tissues spiked with three respiratory viruses, and blood spiked with five viruses (Briese et al., 2015, *mBio*[18]). Additionally, when tested on blood samples spiked with enterovirus D68, VirCapSeq-VERT showed sensitivity comparable to agent-specific PCR. Contamination is not a large problem with VirCapSeq-VERT, as the vast majority of microbial contamination is bacterial. Because of the enrichment for viral reads, we expect fewer false-positive classifications due to bioinformatic alignment errors, although they still may be possible.

Two values will be presented alongside the positive viral taxa: the raw reads, and the number of normalized reads per 10,000 host-subtracted reads. We will not present data on viruses of the family *Anelloviridae*, GB virus C, and GB virus B, all of which are not known to be pathogenic in humans.

Appendix S3. Questions in Main Physician Chart Review

Note: The following is an example of a chart review for a hypothetical patient.

Introduction

Patient ID: Pt_999

MRN: 999999999

Date and Time of Enrollment: 9/9/2016 09:09

Birth Year: 1991

Positive hospital microbial test results within the first 5 days after enrollment: Urine Culture (D1): >100,000 CFU/mL E. coli

Chart Review Phase I:

Do NOT open up the NGS, VirCapSeq-VERT, and host response results at this time.

What is the infection status at the time of enrollment?

O Yes **O** Probably Yes **O** Unsure **O** Probably No **O** No

Is there a clinically significant bacterial/viral/fungal/parasitic infection at the time of enrollment that is the etiology of the patient's presentation?

	Yes	Probably Yes	Unsure	Probably No	No
Bacterial	**O**	**O**	**O**	**O**	**O**
Viral	**O**	**O**	**O**	**O**	**O**

Fungal	**O**	**O**	**O**	**O**	**O**
Parasitic	**O**	**O**	**O**	**O**	**O**

Chart Review Phase II

Please open up the NGS and VirCapSeq-VERT results. Do NOT open up the host response results at this time.

For each positive VirCapSeq-VERT organism, is the organism the etiology of the patient's presentation?

If the organism has already been identified by a hospital test in the first 5 days after enrollment, select "Already Identified by Hospital Tests."

Classify each organism in the order presented in VirCapSeq-VERT results. Leave all unnecessary fields blank. For example, if there are only two organisms, classify just VirCapSeq-VERT Organism #1 and VirCapSeq-VERT Organism #2, and leave all other fields below blank.

As a reminder, the patient's positive hospital test results (within the first five days after enrollment) are:

Urine Culture (D1): >100,000 CFU/mL E. coli

	Already Identified by Hospital Tests	Yes	Probably Yes	Unsure	Probably No	No

VirCapSeq-VERT Organism #1	**O**	**O**	**O**	**O**	**O**	**O**
VirCapSeq-VERT Organism #2	**O**	**O**	**O**	**O**	**O**	**O**
VirCapSeq-VERT Organism #3	**O**	**O**	**O**	**O**	**O**	**O**
VirCapSeq-VERT Organism #4	**O**	**O**	**O**	**O**	**O**	**O**
VirCapSeq-VERT Organism #5	**O**	**O**	**O**	**O**	**O**	**O**

For each positive NGS organism, is the organism the etiology of the patient's presentation?

If the organism has already been identified by a hospital test in the first 5 days after enrollment, select "Already Identified by Hospital Tests."

Classify each organism in the order presented in NGS results. Leave all unnecessary fields blank. For example, if there are only two organisms, classify just NGS Organism #1 and NGS Organism #2, and leave all other fields below blank.

As a reminder, the patient's positive hospital test results (within the first five days after enrollment) are:

Urine Culture (D1): >100,000 CFU/mL E. coli

	Already Identified by Hospital Tests	Yes	Probably Yes	Unsure	Probably No	No
NGS Organism #1	**O**	**O**	**O**	**O**	**O**	**O**
NGS Organism #2	**O**	**O**	**O**	**O**	**O**	**O**
NGS Organism #3	**O**	**O**	**O**	**O**	**O**	**O**
NGS Organism #4	**O**	**O**	**O**	**O**	**O**	**O**
NGS Organism #5	**O**	**O**	**O**	**O**	**O**	**O**

(Table expands for up to 25 organisms, depending on how many organisms is present in patient NGS data.)

With the addition of NGS and VirCapSeq-VERT results, what is the infection status at the time of enrollment?

O Yes **O** Probably Yes **O** Unsure **O** Probably No **O** No

With the addition of NGS and VirCapSeq-VERT results, is there a clinically significant bacterial/viral/fungal/parasitic infection at the time of enrollment that is the etiology of the patient's presentation?

	Yes	Probably Yes	Unsure	Probably No	No
Bacterial	**O**	**O**	**O**	**O**	**O**
Viral	**O**	**O**	**O**	**O**	**O**
Fungal	**O**	**O**	**O**	**O**	**O**
Parasitic	**O**	**O**	**O**	**O**	**O**

Chart Review Phase II

Please open up host response results. (Skip this section if there are no host response results.)

With the addition of host response results: For each positive VirCapSeq-VERT organism, is the organism the etiology of the patient's presentation?

If the organism has already been identified by a hospital test in the first 5 days after enrollment, select "Already Identified by Hospital Tests."

Classify each organism in the order presented in the VirCapSeq-VERT results. Leave all unnecessary fields blank. For example, if there are only two organisms, classify just VirCapSeq-VERT Organism #1 and VirCapSeq-VERT Organism #2, and leave all other fields below blank.

	Already Identified by Hospital Tests	Yes	Probably Yes	Unsure	Probably No	No
VirCapSeq-VERT Organism #1	**O**	**O**	**O**	**O**	**O**	**O**
VirCapSeq-VERT Organism #2	**O**	**O**	**O**	**O**	**O**	**O**
VirCapSeq-VERT Organism #3	**O**	**O**	**O**	**O**	**O**	**O**
VirCapSeq-VERT Organism #4	**O**	**O**	**O**	**O**	**O**	**O**
VirCapSeq-VERT Organism #5	**O**	**O**	**O**	**O**	**O**	**O**

With the addition of host response results: For each positive NGS organism, is the organism the etiology of the patient's presentation?

If the organism has already been identified by a hospital test in the first 5 days after enrollment, select "Already Identified by Hospital Tests."

Classify each organism in the order presented in the NGS results. Leave all unnecessary fields blank. For example, if there are only two organisms, classify just NGS Organism #1 and NGS Organism #2, and leave all other fields below blank.

	Already Identified by Hospital Tests	Yes	Probably Yes	Unsure	Probably No	No
NGS Organism #1	**O**	**O**	**O**	**O**	**O**	**O**
NGS Organism #2	**O**	**O**	**O**	**O**	**O**	**O**
NGS Organism #3	**O**	**O**	**O**	**O**	**O**	**O**
NGS Organism #4	**O**	**O**	**O**	**O**	**O**	**O**
NGS Organism #5	**O**	**O**	**O**	**O**	**O**	**O**

(Table expands for up to 25 organisms, depending on how many organisms is present in patient NGS data.)

With the addition of NGS, VirCapSeq-VERT, and host response results, what is the infection status at the time of enrollment?

O Yes **O** Probably Yes **O** Unsure **O** Probably No **O** No

With the addition of NGS, VirCapSeq-VERT, and host response results, is there a clinically significant bacterial/viral/fungal/parasitic infection at the time of enrollment that is the etiology of the patient's presentation?

	Yes	Probably Yes	Unsure	Probably No	No
Bacterial	**O**	**O**	**O**	**O**	**O**
Viral	**O**	**O**	**O**	**O**	**O**
Fungal	**O**	**O**	**O**	**O**	**O**
Parasitic	**O**	**O**	**O**	**O**	**O**

Table S1. Additional bacteria identified after sequencing seven mNGS plasma samples to higher depth

	Raw Reads in Original Library that Revealed Bacterial Species (at 0.3 Kraken Threshold/at 0 Kraken Threshold)*	**Raw Reads in Re-sequenced Library that Revealed Bacterial Species (at 0.3 Kraken Threshold/at 0 Kraken Threshold)***	**Sequencing Depth in Original Library (Reads)**	**Sequencing Depth in Re-Sequenced Library (Reads)**
Pt_091 (Blood Cx grew *Salmonella enterica*)			17,495,799	90,179,007
Salmonellla enterica	0/18	5/129		
Pt_022 (Blood Cx grew *Enterobacter cloacae* complex)			17,528,199	65,103,185
Enterobacter cloacae complex spp.	0/0	3 /14		

Pt_047 (Blood Cx grew CoNS, not contaminant)			21,837,057	174,777,289
CoNS spp.	0/0	20/26		
Pt_073 (Blood Cx grew *Streptococcus mitis* group)			14,867,751	74,667,179
Streptococcus mitis group *spp.*	0/0	0/0		
Pt_098 (Wound Cx grew *Fusobacterium*, *Porphyromonas*; Urine Cx grew *A. urinae*)			19,727,410	172,614,577
Fusobacterium spp.	0/0	0/7		
Porphyromonas spp.	2/6	12/57		
Aerococcus urinae	0/0	0/2		
Pt_152 (Blood Cx grew CoNS, likely contaminant; Urine Cx grew *Klebsiella pneumoniae*)			17,866,699	103,288,138
CoNS, likely contaminant	1/1	0/1		

Klebsiella pneumoniae	0/0	0/11		
Pt_118 (Bronchoalveolar Cx grew *Staphylococcus aureus)*			30,271,177	262,289,640
Staphylococcus aureus	0/0	7/7		

*Reads were processed using two alignment thresholds specified in Kraken: our original conservative threshold of 0.3, and a liberal threshold of 0.

Cx = Culture, CoNS = Coagulase Negative Staphylococcus

Table S2. Clinical relevance of all organisms found by mNGS which were not detected by hospital tests

Patient	VirCapSeq-VERT and mNGS Organism*	Organism Clinically Relevant?†			mNGS oganism likely contaminant and/or misalignment?‡	Standard-of-Care Microbiology, 0-5d After Presentation	Host Response	Final Diagnosis
		R1	R2	R3				
Pt_006	*Escherichia coli* (37, 26, 12, Yes)	5	5	5	No	All negative.	Bacterial	Bacteremia – Source: Line
Pt_013	*Janthinobacterium* sp 1 2014MBL MicDiv (229, 200, 33, Yes)	1	1	1	Likely Contaminant and/or Misalignment	Blood Culture (D1): Escherichia coli (performed 2h after blood draw for study); Urine Culture (D1): 60,000 CFU/mL Streptococcus agalactiae (Group B).	Bacterial (Derivation Cohort)§	Bacteremia – Source: Urine

Pt_020	*Klebsiella pneumoniae* (85, 69, 4, Yes)	3	5	5	No	Blood Culture (D1): Escherichia coli; Abdominal Wound Culture (D1): 2+ Corynebacterium striatum	PCR Error	Bacteremia – Source: Unclear
	Escherichia coli (45, 33, 12, Yes)	In SOC Microbiology			Already identified by SOC microbiology.			
Pt_021	*Thermus scotoductus* (23, 14, 2, Yes)	1	1	1	Likely Contaminant and/or Misalignment	Blood Culture (D1): Streptococcus agalactiae (Group B); Urine Culture (D1): 30,000 CFU/mL Escherichia coli	Bacterial (Derivation Cohort)§	Bacteremia - Unclear Etiology
Pt_023	Hepatitis C Virus (333,057 / 1,652.06)	1	1	1	N/A	All negative.	Noninfected (Derivation Cohort)§	Malignancy - Metastatic Lung
	Gardnerella vaginalis (12, 7, 3, Yes)	1	2	1	No			
Pt_025	Coxsackievirus B5 (30,324 / 181.31)	3	5	5	N/A	All negative.	Bacterial	Viral Syndrome

	Corynebacterium maris (27, 18, 5, Yes)	3	1	2	Likely Contaminant and/or Misalignment	All negative.	Bacterial	Viral Syndrome
Pt_037	*Prevotella denticola* (104, 86, 2, Yes)	2	5	4	No	Blood Culture (D1): Escherichia coli, Streptococcus anginosus group; Perianal Abscess Culture (D1): 3+ Streptococcus anginosus group; Blood Culture (D2): Streptococcus anginosus group; Perianal Abcsess Fluid Culture (D3): 2+ Bacteroides fragilis group; Perianal Abscess Fluid Culture (D3): 3+ Escherichia coli, 4+	Bacterial (Derivation Cohort)§	Bacteremia – Source: Intra-Abdominal
	Porphyromonas asaccharolytica (73, 58, 2, Yes)	2	5	4	No			
	Fusobacterium nucleatum (61, 47, 3, Yes)	3	5	4	No			
	Dialister pneumosintes (8, 3, 0, No)	2	2	2	No			

						Streptococcus anginosus group		
Pt_041	*Fusobacterium nucleatum* (22, 14, 5, Yes)	2	5	4	No	All negative.	Bacterial	Intra-Abdominal Abscess
Pt_057	*Escherichia coli* (119, 100, 5, Yes)	3	4	4	No	Blood Culture (D1): Coagulase Negative Staphylococcus spp.	Noninfected (Derivation Cohort)§	Allograft Rejection
Pt_058	*Erwinia billingiae* (7, 3, 2, Yes)	1	3	1	Likely Contaminant and/or Misalignment	Urine Culture (D1): 30,000 CFU/mL Yeast	Bacterial	Post-Operative Fever vs. UTI
Pt_061	*Thermus scotoductus* (4, 1, 0, Yes)	1	2	1	Likely Contaminant and/or Misalignment	All negative.	Noninfected (Derivation Cohort)§	Febrile Neutropenia - Unclear Etiology
Pt_066	*Propionibacterium* sp oral taxon 193 (8, 4, 2, Yes)	1	1	1	Likely Contaminant and/or Misalignment			Pyelonephritis

	Erwinia billingiae (7, 3, 2, Yes)	1	1	1	Likely Contaminant and/or Misalignment	Urine Culture (D1): >100,000 CFU/mL Escherichia coli	Bacterial (Derivation Cohort)§	
Pt_067	*Propionibacterium acnes* (8, 3, 2, Yes)	1	1	2	Likely Contaminant and/or Misalignment	All negative.	Viral	Pneumonia
Pt_070	*Staphylococcus warneri* (5440, 5307, 16, Yes)	2	3	2	No	All negative.	Noninfected (Derivation Cohort)§	Pneumonia vs. Drug Reaction
	Lactococcus lactis (870, 812, 437, Yes)	2	3	2	No			
	Actinomyces oris (514, 472, 44, Yes)	2	3	1	No			
	Streptococcus gordonii (484, 443, 22, Yes)	2	3	2	No			
	Rothia dentocariosa (396, 357, 76, Yes)	2	3	2	No			

	Veillonella parvula (335, 301, 14, Yes)	2	3	1	No			
	Streptococcus sanguinis (265, 237, 12, Yes)	2	3	2	No			
	Streptococcus mutans (239, 212, 8, Yes)	2	3	2	No			
	Streptococcus intermedius (145, 125, 6, Yes)	2	3	2	No			
	Streptococcus oralis (143, 122, 12, Yes)	2	3	2	No			
	Staphylococcus pasteuri (119, 99, 7, Yes)	2	3	2	No			
	Fusobacterium nucleatum (117, 97, 12, Yes)	2	3	2	No			
	Prevotella dentalis (78, 63, 5, Yes)	2	3	1	No			

	Staphylococcus aureus (53, 39, 10, Yes)	2	3	3	No			
	Selenomonas sp oral taxon 920 (44, 32, 0, No)	2	3	1	No			
	Gardnerella vaginalis (39, 27, 25, Yes)	2	3	1	No			
	Campylobacter gracilis (38, 27, 5, Yes)	2	3	2	No			
	Selenomonas sputigena (37, 26, 6, Yes)	2	3	1	No			
	Capnocytophaga sp oral taxon 323 (32, 22, 5, Yes)	2	3	2	No			
	Leptotrichia sp oral taxon 212 (22, 14, 5, Yes)	2	3	1	No			
	Tannerella sp oral taxon HOT-286 (19, 11, 5, Yes)	2	3	1	No			

	Campylobacter concisus (17, 9, 5, Yes)	2	3	1	No			
	Olsenella sp oral taxon 807 (16, 8, 6, Yes)	2	3	1	No			
	Leptotrichia sp oral taxon 847 (14, 8, 5, Yes)	2	3	1	No			
Pt_071	*Enterobacter cloacae* (11, 6, 2, Yes)	1	2	3	Possible Contaminant and/or Misalignment	All negative.	Noninfected	Coccidioides Meningitis
Pt_073	*Thermus scotoductus* (43, 31, 30, Yes)	2	2	1	Likely Contaminant and/or Misalignment	Blood Culture (D1): Streptococcus mitis group; Mouth Would Culture (D1): Acinetobacter baumannii; Blood Enzyme Immunoassay (D1): Aspergillus (Galactomannan) Antigen;	Noninfected (Derivation Cohort)§	Bacteremia – Source: Unclear

						Lesion PCR (D2): Herpes Simplex Virus 1		
Pt_077	*Thermus scotoductus* (11, 5, 0, Yes)	1	1	1	Likely Contaminant and/or Misalignment	Urine Culture (D5): >100,000 CFU/mL Escherichia coli	Bacterial (Derivation Cohort)§	Necrotizing Pancreatitis
Pt_083	*Borrelia hermsii* (306, 273, 0, No)	4	5	5	No	Urine Culture (D1): >100,000 CFU/mL Coagulase Negative Staphylococcus spp. (not Staphylococcus saprophyticus)	Bacterial	Tick Borne Relapsing Fever
Pt_084	*Delftia acidovorans* (31, 21, 19, Yes)	1	1	2	Likely Contaminant and/or Misalignment	Abdominal Wound Culture (D1): 4+ Staphylococcus aureus	Noninfected (Derivation Cohort)§	Intra-Abdominal Abscess
Pt_086	*Pseudomonas aeruginosa* (67, 52, 29, Yes)	2	5	3	No	Nasopharyngeal Swab PCR (D1): Rhinovirus.	Viral	Cystic Fibrosis Exacerbation

Pt_092	*Streptococcus agalactiae* (21, 13, 0, No)	In SOC Microbiology			Already identified by SOC microbiology.	Blood Culture (D1): Streptococcus agalactiae (Group B), Escherichia coli; Urine Culture (D1): >100,000 CFU/mL Lactobacillus species, 20,000 CFU/mL Escherichia coli	Bacterial (Derivation Cohort)§	Bacteremia – Source: Skin
	Streptococcus anginosus (9, 4, 2, Yes)	4	5	5	No			
Pt_095	*Propionibacterium acnes* (27, 18, 14, Yes)	1	2	2	Likely Contaminant and/or Misalignment	All negative.	Bacterial	Chemotherapy-Associated Fever
Pt_101	*Prevotella intermedia* (10, 5, 3, Yes)	1	4	3	Possible Contaminant and/or Misalignment	Blood Culture (D1): Viridans group Streptococci	Noninfected	Cholangitis
Pt_103	*Enterobacter cloacae* (200, 175, 6, Yes)	In SOC Microbiology			Already identified by SOC microbiology.		PCR Error	

	Chroococcidiopsis thermalis (51, 38, 3, Yes)	2	3	1	Likely Contaminant and/or Misalignment	Blood Culture (D1): Enterobacter cloacae complex		Bacteremia – Source: Unclear
	Enterobacter ludwigii (15, 8, 0, No)	In SOC Microbiology			Already identified by SOC microbiology.			
Pt_104	*Pseudomonas* sp L1010 (120, 100, 5, Yes)	1	2	2	No	All negative.	Bacterial (Derivation Cohort)§	Gastrostomy Tube Dysfunction
	Pseudomonas fragi (90, 72, 14, Yes)	1	2	2	No			
	Acinetobacter baumannii (72, 56, 31, Yes)	1	2	2	No			
	Leuconostoc citreum (63, 49, 12, Yes)	1	2	2	No			
	Psychrobacter alimentarius (30, 20, 3, Yes)	1	2	2	No			
	Cronobacter sakazakii (18, 11, 0, No)	1	2	2	No			

	Xanthomonas campestris (17, 10, 3, Yes)	1	2	1	No			
Pt_113	*Klebsiella pneumoniae* (240, 213, 4, Yes)	2	4	4	No	All negative.	Bacterial (Derivation Cohort)§	Ulcerative Colitis Flair
Pt_115	*Cupriavidus metallidurans* (5, 2, 0, Yes)	1	3	2	Likely Contaminant and/or Misalignment	All negative.	Bacterial (Derivation Cohort)§	Malignancy - Leukemia
Pt_126	*Morganella morganii* (29, 20, 2, Yes)	4	5	5	No	All negative.	Bacterial (Derivation Cohort)§	Bacteremia – Source: Prostate
Pt_133	*Moraxella osloensis* (25, 16, 14, Yes)	3	4	2	Possible Contaminant and/or Misalignment	All negative.	Bacterial	Post-Operative Surgical Site Infection
Pt_136	*Kocuria palustris* (18, 11, 6, Yes)	2	3	1	No	All negative.	Noninfected	Febrile Neutropenia -

	Brevibacterium linens (11, 6, 2, Yes)	2	3	1	No			Unclear Etiology
	Propionibacterium sp oral taxon 193 (10, 5, 2, Yes)	2	3	1	No			
Pt_137	*Escherichia coli* (207, 181, 12, Yes)	In SOC Microbiology			Already identified by SOC microbiology.	Blood Culture (D1): Escherichia coli, Klebsiella oxytoca	Bacterial (Derivation Cohort)§	Bacteremia – Source: Intra-Abdominal
	Clostridium perfringens (101, 83, 2, Yes)	4	5	5	No			
Pt_145	*Enterobacter hormaechei* (123, 103, 2, Yes)	In SOC Microbiology			Already identified by SOC microbiology.	Blood Culture (D1): Enterobacter cloacae complex, Streptococcus anginosus group	Bacterial (Derivation Cohort)§	Bacteremia – Source: Intra-Abdominal
	Klebsiella pneumoniae (26, 17, 4, Yes)	4	4	5	No			
	Enterobacter cloacae (16, 9, 2, Yes)	In SOC Microbiology			Already identified by SOC microbiology.			
	Leclercia adecarboxylata (8, 4, 2, Yes)	2	2	2	Possible Contaminant and/or Misalignment			

Pt_153	Hepatitis C Virus (423 / 2.57)	1	1	3	N/A	Blood Culture (D1): Klebsiella pneumoniae; Serology (D2): Hepatitis B Surface Antibody	Bacterial (Derivation Cohort)§	Bacteremia - Source Line
	Klebsiella pneumoniae (213, 186, 4, Yes)	In SOC Microbiology			Already identified by SOC microbiology.			
	Lactococcus lactis (57, 44, 8, Yes)	1	1	3	Possible Contaminant and/or Misalignment			
Pt_154	Streptococcus *mitis* (19, 12, 6, Yes)	3	4	4	No	Blood Culture (D1): Coagulase Negative Staphylococcus spp.	PCR Error	Febrile Neutropenia - Unclear Etiology
Pt_163	*Leptospira interrogans* (214, 186, 0, No)	4	5	4	No	All negative.	Bacterial	Leptospirosis
	Xanthomonas campestris (14, 7, 4, Yes)	1	1	2	Likely Contaminant and/or Misalignment			
	Anabaena sp wa102 (13, 7, 3, Yes)	1	1	2	Likely Contaminant and/or Misalignment			

Pt_166	*Lactobacillus mucosae* (16, 9, 5, Yes)	3	2	2	No	All negative.	Bacterial	Balanitis
Pt_171	*Helicobacter pylori* (11, 5, 0, No)	1	2	4	No	All negative.	Bacterial	Cholangitis
Pt_186	*Escherichia* coli (*198*, 171, 141, Yes)	In SOC Microbiology			Already identified by SOC microbiology.	Blood Culture (D1): Escherichia coli; Urine Culture (D1): Escherichia coli	Bacterial (Derivation Cohort)§	Bacteremia – Source: Urine
	Pantoea sp PSNIH1 (31, 21, 4, Yes)	1	2	2	Likely Contaminant and/or Misalignment			
Pt_194	*Haemophilus influenzae* (35, 24, 8, Yes)	3	4	4	No	All negative.	Bacterial	Pneumonia vs. Radiation Pneumonitis
Pt_197	*Escherichia coli* (64, 49, 4, Yes)	1	3	3	No	All negative.	Noninfected (Derivation Cohort)§	Abdominal Pain - Possible Calciphylaxis
	Methyloversatilis sp RAC08 (17, 10, 7, Yes)	1	1	1	Likely Contaminant and/or Misalignment			

*mNGS (bacterial species) numbers represent (raw reads, estimated lower limit for the intensity of blood-associated reads, estimated upper limit for the intensity of contaminant reads, presence in negative controls). VirCapSeq-VERT (viruses) numbers represent (Raw Reads / Reads per 10,000 Host Subtracted Reads).

†Indicates whether physicians classified organism as clinically relevant to the patient's presentation while blinded to host response results. 1 = No, 2 = Probably No, 3 = Unsure, 4 = Probably Yes, 5 = Yes. R= Reviewer.

‡A fourth unblinded physician assessed whether mNGS organisms were likely to be contaminants and/or misalignments after the completion of our main chart review. Answer choices included No, Possible Contaminant and/or Misalignment, and Likely Contaminant and/or Misalignment. This physician considered medical charts, mNGS results, VirCapSeq-VERT results, host response results, and all classifications made by the three physician chart reviewers in the main chart review.

Comments were made by a fourth unblinded physician who reviewed patient charts after chart reviews were completed.

§Indicates patients from derivation cohort who had their host response results used to re-establish cutoffs for host response scores (see figure 4B).

Table S3. Clinical relevance of organisms found by VirCapSeq-VERT which were not detected by hospital tests: Summary Table

VirCapSeq-VERT Organisms Not Detected by Standard-of-Care Microbiology*	**Likely Clinically Relevant†**	**Uncertain Clinical Relevance**	**Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant**
Human herpesvirus 6	0	2	2
Epstein-Barr virus‡	1	0	7
Hepatitis C virus	0	2	6
Hepatitis B virus	0	0	1
BK virus‡	0	0	2
Trichodysplasia spinulosa-associated polyomavirus	0	0	1
Coxsackievirus B4	0	1	0
Coxsackievirus B5	1	0	0
Coxsackievirus A6	1	0	0
Human parvovirus B19	0	1	0

*Includes tests performed within 5 days after presentation.

†Patients in the "Likely Clinically Relevant" column had VirCapSeq-VERT organisms which were classified as clinically relevant or probably clinically relevant to the patient's presentation

by physician consensus while blinded to host response results. Patients were classified into the remaining two columns by a fourth unblinded physician after the completion of our main chart review. This physician considered medical charts, mNGS results, VirCapSeq-VERT results, host response results, and all classifications made by the three physician chart reviewers in the main chart review.

‡One patient had both viral reactivation with Epstein-Barr virus and possible chronic infection with BK virus. There were 27 patients total.

Clinical details for all patients in this table are presented in table S4.

Table S4. Clinical relevance of all organisms found by VirCapSeq-VERT which were not detected by hospital tests: Clinical Details

Patient	VirCapSeq-VERT and mNGS Organism*	Organism Clinically Relevant?†			Comments on Clinical Relevance‡	Pos. Standard-of-Care Microbiology, 0-5d After Presentation	Host Response	Final Diagnosis
		R1	R2	R3				
Pt_003	Coxsackievirus A6 (33,309,209 / 8,185.44)	3	4	5	Likely Clinically Relevant	All negative.	Bacterial	Viral Syndrome
Pt_004	Human herpesvirus 6 (4,519 / 9.68)	2	4	3	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	All negative.	PCR Error	Febrile Neutropenia - Unclear Etiology
Pt_015	Human herpesvirus 6 (3,550 / 20.53)	1	4	2	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Serology (D4): Hepatitis A virus IgG	Bacterial	Febrile Neutropenia,

								Relapsed AML
Pt_019	Hepatitis C virus (2,785,501 / 3,423.27)	1	2	2	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Urine Culture (D1, D2): >100,000 CFU/mL *Citrobacter freundii* complex, >100,000 CFU/mL *Enterococcus* species	Bacterial (Derivation Cohort)§	Pyelonephritis
	Citrobacter freundii (115, 97, 7, Yes)	In SOC Microbiology			N/A			
Pt_023	Hepatitis C virus (333,057 / 1,652.06)	1	1	1	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	All negative.	Noninfected (Derivation Cohort)§	Malignancy - Metastatic Lung
	Gardnerella vaginalis (12, 7, 3, Yes)	1	2	1	N/A			
Pt_025	Coxsackievirus B5 (30,324 / 181.31)	3	5	5	Likely Clinically Relevant	All negative.	Bacterial	Viral Syndrome

	Corynebacterium maris (27, 18, 5, Yes)	3	1	2	N/A			
Pt_039	Epstein-Barr virus (239,712 / 1,338.79)	2	1	3	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	All negative.	Bacterial	Malignancy - Lymphoma
Pt_040	Epstein-Barr virus (336 / 1.46)	1	2	2	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Plasma PCR (D1, D4): Cytomegalovirus; Serology (D1): Cytomegalovirus IgG, Cytomegalovirus IgM, Epstein-Barr virus VCA IgG, Epstein-Barr virus EBNA IgG; Serology (D3*): Coxiella burnetti* (Q Fever IGG Phase).	Bacterial (Derivation Cohort)§	Infectious Mononucleosis
	Cytomegalovirus (2,936/12.77)	In SOC Microbiology			N/A			
	Cupriavidus metallidurans (8, 3, 1, Yes)	1	1	2	N/A			

Pt_052	Epstein-Barr virus (1,673 / 6.92)	1	2	2	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Blood Culture (*D1): Streptococcus anginosus, Haemophilus influenzae*; Bile Fluid Culture (D4): 4+ *Enterococcus faecalis*, 2+ *Streptococcus mitis*	Bacterial (Derivation Cohort)§	Cholecysitits, Hepatic Abscess
	BK Virus (2,384 / 9.86)	1	2	1	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant			
	Haemophilus influenzae (190, 165, 2, Yes)	In SOC Microbiology			N/A			
Pt_076	Epstein-Barr virus (11,422 / 140.30)	2	4	4	Likely clinically Relevant. (However, possibility for viral reactivation remains.)	All negative. Note: The following test was also identified: Urine Culture (D -3): > 100,000 CFU/mL Enterococcus Species	Bacterial (Derivation Cohort)§	UTI vs. Malignancy - Lymphoma

Pt_081	Hepatitis C virus (88,841 / 389.13)	1	1	1	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Urine Culture (D1): >100,000 CFU/mL Escherichia coli	PCR Error	Pyelonephritis
Pt_089	BK virus (541/ 1.1)	1	1	1	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Blood Culture (D1): *Escherichia coli*; Urine Culture (D1): 50,000 CFU/mL *Escherichia coli*.	Bacterial	Bacteremia – Source: Urine
Pt_096	Coxsackievirus B4 (3,097 / 19.99)	2	3	3	Uncertain: Presentation fits with UTI	Urine Culture (D1): 100,000 CFU/mL *Escherichia coli*	Bacterial (Derivation Cohort)§	Pyelonephritis
Pt_107	Human parvovirus B19 (378/ 0.4)	1	2	2	Uncertain: Presentation fits with Norovirus infection. However, patient exposed to young children.	Stool PCR (D1): Norovirus	Bacterial (Derivation Cohort)§	Diarrhea - Infectious
Pt_112	Epstein-Barr virus (91 / 1.88)	1	3	1	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Urine Culture (D4): >100,000 CFU/mL *Escherichia coli*	Bacterial	Pyelonephritis

Pt_114	Hepatitis C virus (34,254 / 241.11)	1	2	1	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Blood Culture (D1): Probable Coagulase Negative *Staphylococcus* spp.	Bacterial (Derivation Cohort)§	Malignancy - Metastatic Prostate
Pt_130	Hepatitis C virus (3,584,985/2,600.8)	1	1	2	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant. Documented HCV chronic infection.	Blood Culture (D1): Coagulase Negative *Staphylococcus* spp.	Bacterial	Allograft Rejection
Pt_146	Trichodysplasia spinulosa-associated polyomavirus (698 / 2.79)	1	1	2	Likely Viral Reactivation and/or Chronic Infection. Possibly due to steroids and TNF-alpha suppression. No documented skin lesions.	All negative.	Bacterial (Derivation Cohort)§	Crohn's Flair vs. Small Bowel Obstruction
Pt_148	Epstein-Barr virus (275 / 1.62)	1	3	2	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Blood Culture (D2, D4): *Pseudomonas aeruginosa*;	Bacterial (Derivation Cohort)§	Bacteremia – Source: Respiratory

	Pseudomonas aeruginosa (2067, 1979, 46, Yes)	In SOC Microbiology			N/A	Respiratory Culture (D1): *Pseudomonas aeruginosa.*		
Pt_149	Hepatitis C virus (14,080,841/7,060.5)	1	1	2	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Nasopharyngeal Swab PCR (D1): Influenza A 2009 H1N1	Viral (Derivation Cohort)§	URI
Pt_150	Hepatitis B virus (49,462,447 / 9,210.54)	1	1	2	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Urine Culture (D1): 50,000 CFU/mL *Escherichia coli*	Bacterial (Derivation Cohort)§	Pyelonephritis
Pt_152	Epstein-Barr virus (119 / 1.00)	1	2	2	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Blood Culture (D1): Coagulase Negative *Staphylococcus* spp.; Urine Culture (D1): >100,000 CFU/mL *Klebsiella pneumoniae*	Bacterial (Derivation Cohort)§	Pyelonephritis

Pt_153	Hepatitis C virus (423 / 2.57)	1	1	3	Uncertain: Clinical history does not align with chronic HCV infection.	Blood Culture (D1): *Klebsiella pneumoniae;* Serology (D2): Hepatitis B Surface Antibody	Bacterial (Derivation Cohort)§	Bacteremia – Source: Line
	Klebsiella pneumoniae (213, 186, 4, Yes)	In SOC Microbiology			N/A			
	Lactococcus lactis (57, 44, 8, Yes)	1	1	3	N/A			
Pt_156	Epstein-Barr virus (1811/2.7)	In SOC Microbiology			Likely Viral Reactivation	Serology (D1): Epstein-Barr virus Monospot Antibody Test; Plasma PCR (D2): Cytomegalovirus, Epstein-Barr virus	Bacterial	Post-Operative Fever
	Human herpesvirus 6 (21,003/30.8)	2	4	2	Uncertain: Clinically Relevant or Viral Reactivation	Serology (D1): Epstein-Barr virus Monospot Antibody Test; Plasma	Bacterial	Post-Operative Fever

						PCR (D2): Cytomegalovirus, Epstein-Barr virus		
Pt_164	Hepatitis C virus (6,097/22.5)	1	2	2	Uncertain: Clinical history does not align with chronic HCV infection.	Plasma PCR (D1): Cytomegalovirus	Bacterial (Derivation Cohort)§	Infectious Mononucleosis
	Cytomegalovirus (4,502/16.6)	In SOC Microbiology			Already identified by SOC microbiology.	Plasma PCR (D1): Cytomegalovirus	Bacterial (Derivation Cohort)§	Infectious Mononucleosis
Pt_165	Human herpesvirus 6 (288 / 1.42)	2	4	2	Uncertain: Clinically Relevant or Viral Reactivation	All negative.	Bacterial	Pneumonia vs. Drug Reaction
Pt_187	Hepatitis C virus (27,407,996 / 3,987.89)	1	1	2	Likely Viral Reactivation and/or Chronic Infection, Not Clinically Relevant	Serology (D1): Hepatitis B Surface Antigen	Viral (Derivation Cohort)§	Pneumonia vs Aspiration Pneumonitis

*mNGS (bacterial species) numbers represent (raw reads, estimated lower limit for the intensity of blood-associated reads, estimated upper limit for the intensity of contaminant reads, presence in negative controls). VirCapSeq-VERT (viruses) numbers represent (Raw Reads / Reads per 10,000 Host Subtracted Reads).

†Indicates whether physicians classified virus as clinically relevant to the patient's presentation while blinded to host response results. 1 = No, 2 = Probably No, 3 = Unsure, 4 = Probably Yes, 5 = Yes. R= Reviewer.

‡Comments were made by a fourth unblinded physician as to whether viruses were clinically relevant, chronic infections and/or viral reactivations that were not clinically relevant, or of uncertain clinical relevance. This physician considered medical charts, mNGS results, VirCapSeq-VERT results, host response results, and all classifications made by the three physician chart reviewers in the main chart review.

§Indicates patients from derivation cohort who had their host response results used to re-establish cutoffs for host response scores (see figure 4B).

Table S5. Physician and host response interpretation of patients originally classified as probably noninfected or noninfected, and found with bacterial sequences in plasma mNGS

Patient *	mNGS Organisms‡	mNGS Organism Clinically Relevant? ¶			Clinical Comments	Antibiotics\\	Improved?**	Host Response
		R1	R2	R3				
Pt_023	*Gardnerella. vaginalis* (12, 7, 3)	1	2	1	Female with COPD exacerbation, lymphangitic carcinomatosis.	N	Y	Not Infected (Derivation Cohort)††
Pt_057	*Escherichia coli* (119, 100, 5)	3	4	4	Renal transplant rejection.	N	Y	Not Infected (Derivation Cohort)††
Pt_070	*Staphylococcus warneri* (5440, 5307, 16)	2	3	2	Likely Nivolumab-associated pneumonitis. +Hemoptysis, severe gingivitis.	Y	Y	Not Infected (Derivation Cohort)††
	Lactococcus lactis (870, 812, 437)	2	3	2				
	Actinomyces oris (514, 472, 44)	2	3	1				
	Streptococcus gordonii (484, 443, 22)	2	3	2				
	Rothia dentocariosa (396, 357, 76)	2	3	2				

	+19 Additional Organisms§							
Pt_104	*Pseudomonas* sp. L1010 (120, 100, 5)	1	2	2	Clogged gastronomy tube, extensive bowel resection. +Total parenteral nutrition.	N	Y	Bacterial (Derivation Cohort)††
	Pseudomonas fragi (90, 72, 14)	1	2	2				
	Acinetobacter baumannii (72, 56, 31)	1	2	2				
	Leuconostoc citreum (63, 49, 12)	1	2	2				
	Psychrobacter alimentarius (30, 20, 3)	1	2	2				
	Cronobacter sakazakii (18, 11, 0)	1	2	2				
	Xanthomonas campestris (17, 10, 3)	1	2	1				
Pt_113	*Klebsiella pneumoniae* (240, 213, 4)	2	4	4	Ulcerative Colitis flare. +Recent *C. difficiles* colitis on Vancomycin taper.	N	Y	Bacterial (Derivation Cohort)††
Pt_136	*Kocuria palustris* (18, 11, 6)	2	3	1	Neutropenic fever following chemotherapy. +Mucositis, oral ulcers.	Y	Y	Not Infected
	Brevibacterium linens (11, 6, 2)	2	3	1				
	Propionibacterium sp. oral taxon 193 (10, 5, 2)	2	3	1				

Pt_194	*Haemophilus influenzae* (35, 24, 8)	3	4	4	Pneumonia or radiation pneumonitis	N	Y	Bacterial
Pt_197	*Escherichia coli* (64, 49, 4)	1	3	3	Abdominal pain and chronic wounds likely due to calciphylaxis.	N	Y	Not Infected (Derivation Cohort)††
	Methyloversatilis sp. RAC08 (17, 10, 7)[b]	1	1	1				
Pt_058 †	*Erwinia billingiae* (7, 3, 2)†	1	3	1	Urethral stent malfunction.	Y	Y	Bacterial
Pt_095 †	*Propionibacterium acnes* (27, 18, 14)†	1	2	2	Neutropenic fever following chemotherapy.	Y	Y	Bacterial
Pt_115 †	*Cupriavidus metallidurans* (5, 2, 0)†	1	3	2	Myeloid neoplasm.	Y	Y	Bacterial (Derivation Cohort)††

No	Probably No	Unsure	Probably Yes	Yes
1	2	3	4	5

*Patients were identified by the following criteria: 1) patient had consensus classification as either noninfected or probably noninfected by physicians while blinded to mNGS, VirCapSeq-VERT, and host response data; and 2) the patient had a positive mNGS result for an organism not detected by standard-of-care microbiology within 5 days after presentation.

†These organisms were believed by clinicians to be contaminants that remained despite application of the contaminant removal method.

‡The three numbers in parentheses indicate (raw reads, estimated lower limit for the intensity of blood-associated reads, estimated upper limit for the intensity of contaminant reads). All patients in this table had negative VirCapSeq-VERT results, except for Pt_023, who had a positive result for hepatitis C Virus. Additionally, all patients did not have any positive, clinically relevant standard-of-care microbiology results.

§Patient Pt_070 had 19 additional oral-related organisms identified by mNGS. Full results are provided in supplementary attachment 1.

¶Clinical relevance determined by three physician chart physicians (R1-3) who examined mNGS and VirCapSeq-VERT results in the context of the entire medical chart, while blinded to host response results.

\\Indicates whether antibiotics were prescribed to the patient for the sepsis-like illness that prompted their ED visit and/or admission, as documented by medical chart data.

**Indicates whether the patient improved, as documented by medical chart data.

††Indicates patients from derivation cohort who had their host response results used to re-establish cutoffs for host response scores (see figure 4B).

COPD = Chronic obstructive pulmonary disease.

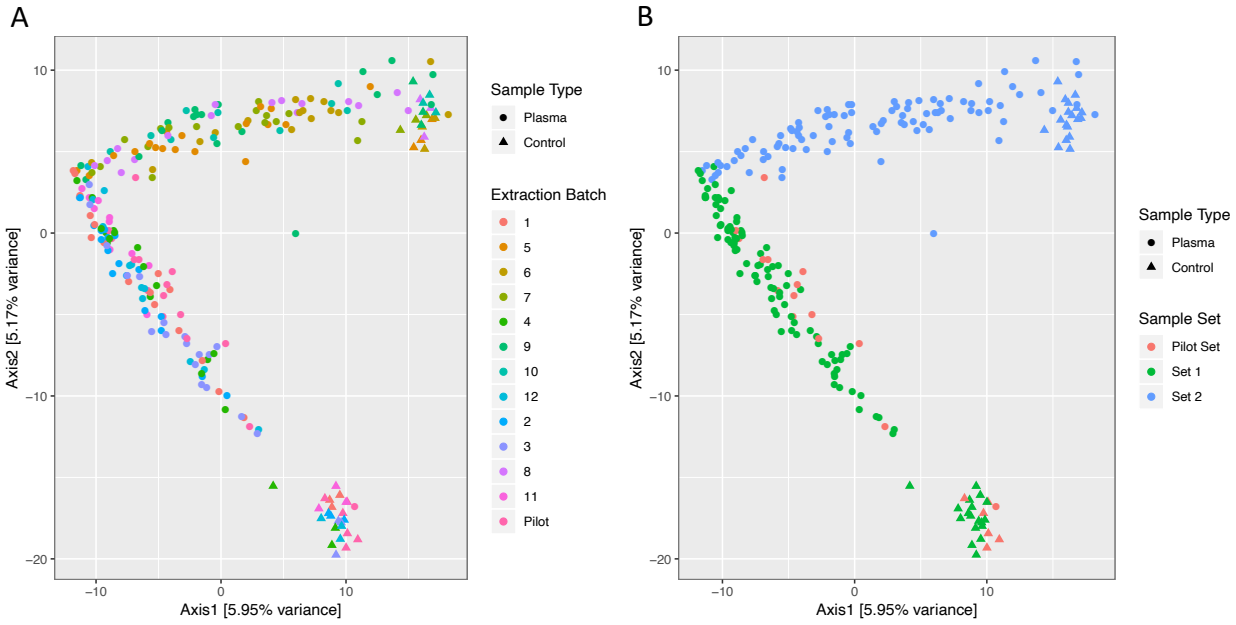


Figure S1. Possible batch effects on sample composition. Principal component analysis of truncated rank-transformed plasma and negative control sample reads. (A) Individual extraction batches do not form distinct clusters. However, two distinct clusters can be visualized (B) when samples are grouped into sets based on their extraction batch. Set 1 (red) corresponds to samples in the second sequencing batch but extracted in extraction batches 1, 2, 3, 4, 11 and 12. Set 2 (green) corresponds to samples in the second sequencing batch but extracted in extraction batches 5, 6, 7, 8, 9 and 10. The Pilot Set (red) refers to all samples extracted and sequenced in the pilot sequencing batch.

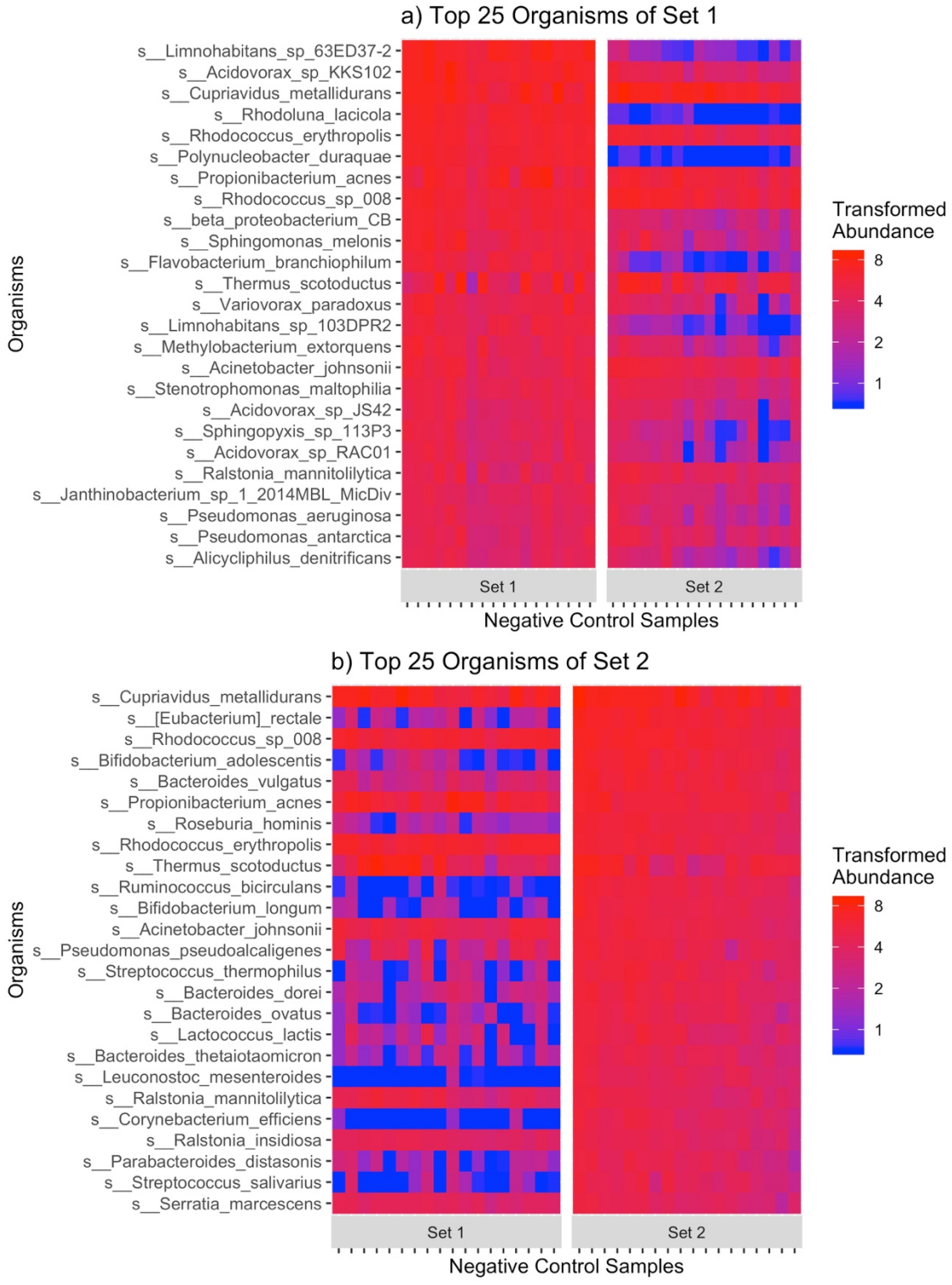

a) Top 25 Organisms of Set 1
s__Limnohabitans_sp_63ED37-2
s__Acidovorax_sp_KKS102
s__Cupriavidus_metallidurans
s__Rhodoluna_lacicola
s__Rhodococcus_erythropolis
s__Polynucleobacter_duraquae
s__Propionibacterium_acnes
s__Rhodococcus_sp_008
s__beta_proteobacterium_CB
s__Sphingomonas_melonis
s__Flavobacterium_branchiophilum
s__Thermus_scotoductus
s__Variovorax_paradoxus
s__Limnohabitans_sp_103DPR2
s__Methylobacterium_extorquens
s__Acinetobacter_johnsonii
s__Stenotrophomonas_maltophilia
s__Acidovorax_sp_JS42
s__Sphingopyxis_sp_113P3
s__Acidovorax_sp_RAC01
s__Ralstonia_mannitolilytica
s__Janthinobacterium_sp_1_2014MBL_MicDiv
s__Pseudomonas_aeruginosa
s__Pseudomonas_antarctica
s__Alicycliphilus_denitrificans
Organisms
Set 1
Set 2
Negative Control Samples
Transformed Abundance
8
4
2
1
b) Top 25 Organisms of Set 2
s__Cupriavidus_metallidurans
s__[Eubacterium]_rectale
s__Rhodococcus_sp_008
s__Bifidobacterium_adolescentis
s__Bacteroides_vulgatus
s__Propionibacterium_acnes
s__Roseburia_hominis
s__Rhodococcus_erythropolis
s__Thermus_scotoductus
s__Ruminococcus_bicirculans
s__Bifidobacterium_longum
s__Acinetobacter_johnsonii
s__Pseudomonas_pseudoalcaligenes
s__Streptococcus_thermophilus
s__Bacteroides_dorei
s__Bacteroides_ovatus
s__Lactococcus_lactis
s__Bacteroides_thetaiotaomicron
s__Leuconostoc_mesenteroides
s__Ralstonia_mannitolilytica
s__Corynebacterium_efficiens
s__Ralstonia_insidiosa
s__Parabacteroides_distasonis
s__Streptococcus_salivarius
s__Serratia_marcescens
Organisms
Set 1
Set 2
Negative Control Samples
Transformed Abundance
8
4
2
1

Figure S2. Two distinct reagent contamination profiles in negative controls. Many of the most highly abundant taxa of Set 1 negative controls were not present in Set 2 negative controls, and vice versa. Heatmaps illustrate the arcsinh-transformed abundances of selected organisms across all Set 1 and Set 2 negative control samples prior to contaminant removal. These selected organisms consisted of a) the top 25 organisms in the negative controls of Set 1, and b) the top 25 organisms in the negative controls of Set 2, as ranked by average arcsinh-transformed abundance in negative controls of each set. A number of human commensal organisms are present as contaminants, particularly in Set 2, such as *Bifidobacterium spp.*, *Bacteroides spp.*, and *P. aeruginosa.* Set 1 and Set 2 correspond to samples in the second sequencing batch but extracted in extraction batches 1, 2, 3, 4, 11 and 12; and second sequencing batch but extracted in extraction batches 5, 6, 7, 8, 9 and 10, respectively.

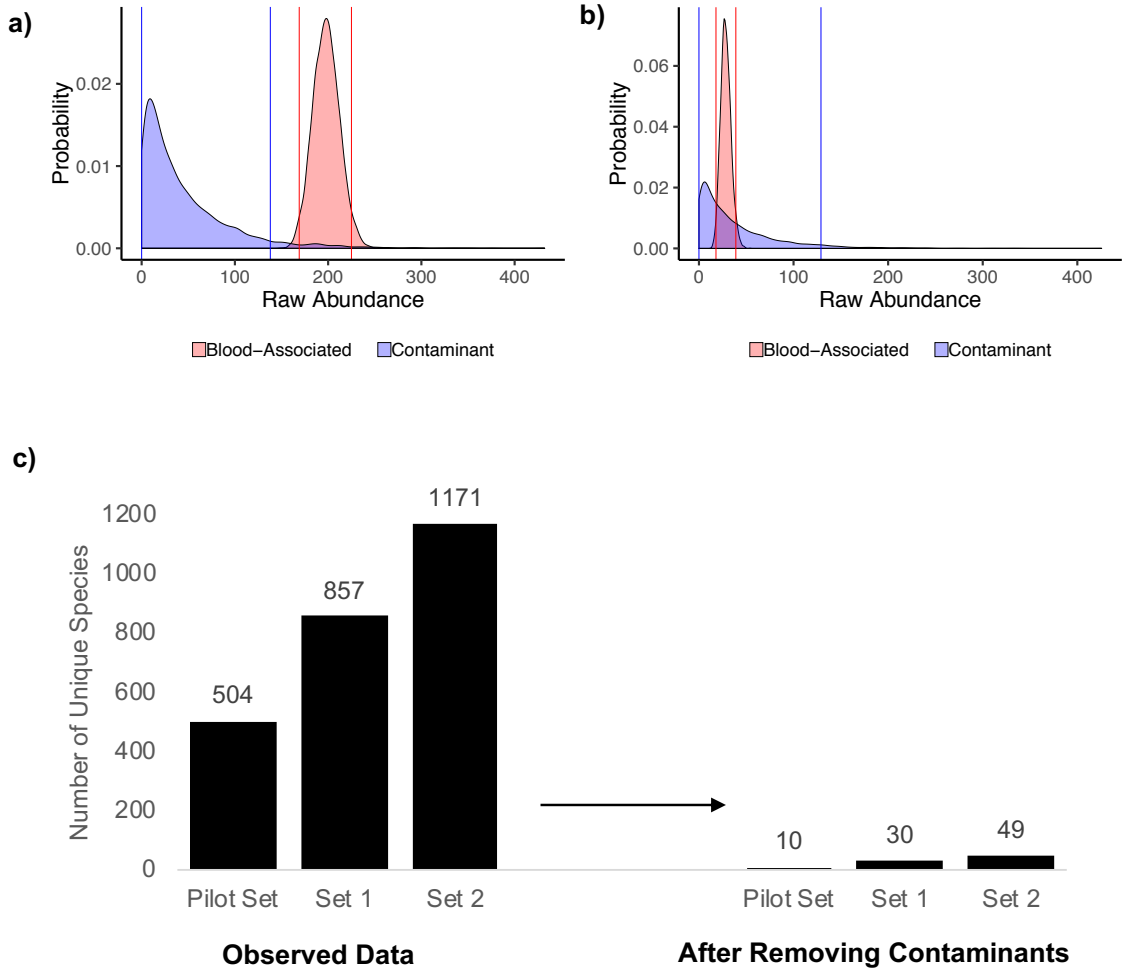


Figure S3. Application of Bayesian inference for distinguishing blood-associated DNA sequences from contaminating DNA sequences in mNGS data. Density plots provide the posterior distribution for true intensity and probability distribution for contaminant intensity (from negative control samples) for blood-associated *Escherichia coli* (red) and contaminating *E coli* (blue) in two patients. (A) A patient who had a positive blood culture for *E. coli* (Pt_186), and (B) A patient who did not have a positive culture for *E. coli* (Pt_043). Vertical lines indicate the lower and upper limits of the 95% highest posterior density interval for true intensity (red)

and highest density interval for the contaminant intensity (blue) A sample was considered to have a positive result for a particular species with 95% chance when the lower limit for the true intensity exceeded the upper limit for the contaminant intensity; otherwise, the species was eliminated from the dataset. (C) The vast majority of unique species were eliminated from samples in our three sets of plasma samples.

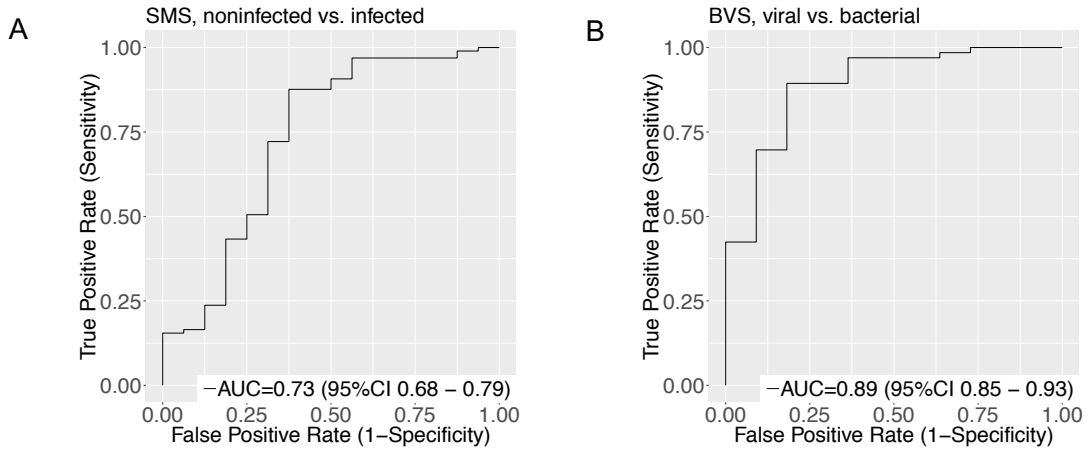


Figure S4: Receiver operating characteristic curves for host response score. Receiver operating characteristic (ROC) curves for the Sepsis MetaScore (SMS, for noninfected SIRS vs. sepsis) and Bacterial/Viral metaScore (BVS).

Supplementary References

1. van der Valk T, Vezzi F, Ormestad M, Dalén L, Guschanski K. Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. *Mol Ecol Resour*. 2019; doi:10.1111/1755-0998.13009.
2. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; 9(4):357-359.
3. Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. 2014; 15(3):R46.
4. Morgulis A, Gertz EM, Schäffer AA, Agarwala R. A Fast and Symmetric DUST Implementation to Mask Low-Complexity DNA Sequences. *J Comput Biol*. 2006; 13(5):1028-1040.
5. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 2013; 8(4):e61217.
6. Eckburg PB, Lepp PW, Relman DA. Archaea and Their Potential Role in Human Disease. *Infect Immun*. 2003; 71(2):591-596.
7. Callahan BJ, Sankaran K, Fukuyama JA, McMurdie PJ, Holmes SP. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Research*. 2016; 5(2):1492.
8. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010; 11(10):R106.
9. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of

microbiota in low bacterial biomass samples. *Gut Pathog*. 2016; 8:24.

10. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 2011; 27(6):863-864.
11. Chevreux B, Wetter T, Suhai S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. *German Conference on Bioinformatics* 1999; 99:45-56.
12. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25(16):2078-2079.
13. Milne I, Bayer M, Cardle L, et al. Tablet--next generation sequence assembly visualization. *Bioinformatics* 2010; 26(3):401-402.
14. Casadei D. Reference analysis of the signal + background model in counting experiments. *J Instrum*. 2012; 7(01):P01012. arXiv:1108.4270v5
15. Berger J. The Case for Objective Bayesian Analysis. *Bayesian Anal*. 2006; 1(3):385-402.
16. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15(12):550.
17. Sweeney TE, Wong HR, Khatri P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Sci Transl Med*. 2016; 8(346):346ra91.
18. Briese T, Kapoor A, Mishra N, et al. Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *MBio* 2015; 6(5):e01491-15.