Host-pathogen interactions in urinary tract infection from the structure and diversity of urinary cell-free DNA

Philip Burnham[1], Darshana Dadhania[2], Michael Heyang[1], Fanny Chen[1], Manikkam Suthanthiran[2], John Richard Lee[2,+,*], Iwijn De Vlaminck[1,+,*]

[1]Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853;
[2]Division of Nephrology and Hypertension, Departments of Medicine and Transplantation Medicine, New York Presbyterian-Weill Cornell Medical Center, New York, NY 10065
[+]These authors contributed equally to this work
*To whom correspondence should be addressed: vlaminck@cornell.edu and jrl2002@med.cornell.edu.

Abstract

Infections of the urinary tract are the most common form of infection in the human population. Here, we tested the utility of urinary cell-free DNA to comprehensively monitor host and pathogen dynamics in the scope of bacterial and viral urinary tract infection. We assayed 115 urinary cell-free DNA isolates from a cohort of 67 kidney transplant recipients by unbiased sequencing. We find that urinary cell-free DNA simultaneously informs about the composition of the urinary microbiome and virome, antimicrobial resistance and susceptibility, bacterial growth dynamics, kidney allograft injury, and the host response to infection. These different layers of information are accessible from a single assay and individually agree well with corresponding clinical tests based on quantitative PCR, conventional bacterial culture, and urinalysis. In addition, cell-free DNA reveals the frequent occurrence of clinically relevant pathologies that remain undiagnosed during conventional diagnostic workups. Our work identifies genomic assays of urinary cell-free DNA as a highly informative clinical and research tool to monitor infections of the urinary tract.

Introduction

Urinary tract infection (UTI) is one of the most common medical problems in the general population[1]. Among kidney transplant recipients, UTIs occur at an alarmingly high rate[2]. Bacterial UTI affects at least 20% of kidney transplant recipients in the first year after transplantation[3], and at least 50% in the first three years after transplantion[4]. In addition, complications due to viral infection occur often. An estimated 5-8% of kidney transplant recipients suffer nephropathy from BK polyomavirus infection in the first three years after transplantation[5,6]. Other viral infections in kidney transplant recipients include adenovirus, JC polyomavirus, cytomegalovirus (CMV), and parvovirus. The current gold standard for diagnosis of bacterial UTI is *in vitro* urine culture[7]. Although improved culture methods are being investigated[8], bacterial culture is limited to detection of relatively few cultivable organisms. A recent study reports that almost all women with symptoms of UTI but a negative culture still have an infection[9]. In addition, bacterial culture is unable to inform about commensal microbiota, viral infections, or about bacterial growth dynamics. Last, urinalysis is often required in conjunction with culture to make treatment decisions.

A large number of small fragments of cell-free DNA (cfDNA) are present in plasma and urine[10–12]. These molecules are the debris of cell death across the body and offer opportunities for precision diagnostics based on '*omics* principles, with applications in a wide range of medical settings, including pregnancy, cancer and solid-organ transplantation[12–15]. Here, we have investigated the utility of urinary cfDNA to comprehensively monitor host and pathogen dynamics that arise in the scope of viral and bacterial infections of the urinary tract. We used unbiased sequencing to assay cfDNA isolated from 115 urine samples collected from a cohort of 67 kidney transplant recipients, including patients diagnosed with bacterial UTIs and BK polyomavirus nephropathy. We implemented a single-stranded DNA (ssDNA) sequencing library preparation, optimized for the analysis of ultrashort DNA fragments (average length less than 100 bp), and were able to perform robust analyses of ultrashort cfDNA isolated from just one milliliter of urine supernatant[16,17]. We find that urinary cfDNA detects clinically identified uropathogens, while also uncovering co-infections and frequent viral and bacterial infections that remain undiagnosed in conventional diagnostic workups.

We further investigated cfDNA analysis methodologies that go beyond mere identification of microbial sequences and provide a deeper understanding of ongoing infections. First, we show that the pattern of cfDNA sequencing read coverage across bacterial genomes is highly non-uniform, with an overrepresentation of sequences at the origin of replication. A similar pattern has previously been observed in whole genome sequencing of gut microbiome communities, and the disproportionate genome coverage was shown to reflect the bacterial growth rate, where an overrepresentation of genomic coverage at the

replication origin signals faster growth[18]. Here, we show that quantifying bacterial growth rates from urinary cfDNA can be used to inform diagnosis of UTI. Second, we mined cfDNA for antimicrobial resistance (AR) genes and show that antimicrobial resistome profiles can be used to evaluate the susceptibility to antibacterial therapy. Last, we show that cfDNA informs about the host response to infection on both a cellular and tissue level. Recent reports have demonstrated that cfDNA in plasma comprises the footprints of DNA-binding proteins and nucleosomes. We show, for the first time, that nucleosome structures within transcription regulatory elements are preserved in urinary cfDNA, as was previously described for plasma cfDNA[19]. We find that the occupancy of nucleosomes in gene regions flanking the transcription start site is notably reduced for transcribed genes, consistent with known chromatin alterations in gene promoters during transcription, thus providing a measure of gene expression. We furthermore find that the relative proportion of kidney donor specific cfDNA indicates graft tissue injury in the scope of viral infection and host immune cell activation in the scope of bacterial infection. Finally, we report that the graft is the predominant source of mitochondrial cfDNA in the urine of kidney transplant recipients with BK polyomavirus nephropathy.

Collectively, this study supports the utility of unbiased sequencing of urinary cfDNA as a comprehensive tool for monitoring patient health and discovering novel interactions between pathogens and the host.

Results

Biophysical properties of urinary cell-free DNA

Urinary cfDNA is comprised of chromosomal, mitochondrial, and microbial cfDNA released from host cells and microbes in the urinary tract and of plasma-derived cfDNA that passes from the blood circulation into urine[20]. Urine can be collected non-invasively in large volumes, and therefore represents an attractive target for diagnostic assays. Compared to plasma DNA, relatively few studies have examined the properties and diagnostic potential of urinary cfDNA. The urinary environment degrades nucleic acids more rapidly than plasma resulting in fewer DNA fragments that are far shorter[21]. Consequently, sequence analyses of urinary cfDNA have to date required relatively large (> 10 ml) volumes of urine[22,23]. Here, we applied a single-stranded library preparation technique that employs ssDNA adapters and bead ligation to create diverse sequencing libraries that capture short, highly degraded cfDNA[16,17] (Fig. 1A). We find that single-stranded library preparation enables robust sequence analyses of urinary cfDNA from just one milliliter of cell-free urine supernatant. The sequencing libraries comprised on average 4.07 billion unique molecules. We assayed 115 urine samples collected from kidney transplant recipients, including subjects diagnosed with bacterial UTI and BK

polyomavirus nephropathy (overview of collection dates and categories depicted in Fig. 1B, see Methods for a detailed description of the study cohort). We obtained 38.2 +/- 12.1 million paired-end reads per sample, yielding a per-base human genome coverage of 0.42x +/- 0.14x. Many fragments aligned to microbiota, for example, for patients diagnosed with bacterial UTI, bacterial cfDNA accounted for up to 34.65% of the raw sequencing reads and in cases of BK polyomavirus nephropathy, polyomavirus cfDNA accounted for up to 10.27% of raw sequencing reads. To account for technical variability and sources of environmental contamination during extraction and library preparation, a known-template control sample was included in every sample batch and sequenced (see Methods).

We analyzed the fragment length profiles of urinary cfDNA at single nucleotide resolution using paired-end read mapping[10]. This analysis confirmed previous observations of the highly fragmented nature of urinary cfDNA compared to plasma cfDNA[22] (Fig. 1C). We observed a 10.4 bp periodicity in the fragment length profile for chromosomal cfDNA in the urine (Fourier analysis, Fig. 1C, inset), consistent with the periodicity of DNA-histone contacts in nucleosomes. Polyomavirus is known to hijack histones of infected host cells, and to form minichromosomes after infection[24]. The fragment length profiles of BK polyomavirus cfDNA in urine reflect this pathobiology, and indicate a predominant nucleosomal origin of polyomavirus cfDNA (Fig. 1C). A similar nucleosomal footprint is not observed for bacterial and mitochondrial cfDNA, and cfDNA arising from parvovirus, which is expected given the non-nucleosomal compaction of the genomes that contribute these cfDNA types (Fig. 1D). These data demonstrate that analyses of the structure of cfDNA can be used to learn about the pathobiology of uropathogens.

Infectome screening

We assessed the presence of cfDNA from bacterial and viral pathogens reported in clinical diagnostic workups. Here, we used previously described bioinformatics approaches to quantify non-human cfDNA sequences in the datasets[25]. Briefly, human sequences were identified by alignment of the sequences to the human reference and removed. Remaining sequences were BLASTed against a custom database of microbial reference genomes. We estimated the relative genomic representation of different species using GRAMMy[26]. To directly compare the measured microbial abundance across samples and species, we computed the representation of microbial genome copies relative to the representation of the human genome in the datasets, and expressed this quantity as RGE, relative genome equivalents.

We detected an extremely high load of BK polyomavirus cfDNA in all 24 samples collected from 22 patients with BK polyomavirus nephropathy (BKVN) per needle biopsy (mean 1.5

+/- 1.1 x10^5 RGE, Fig. 2A), but not in 10 patients that were BKVN negative per needle biopsy (mean 0.17 RGE). This high load of BK derived DNA is consistent with the pathobiology of BKVN. The BK cfDNA abundance (RGE) correlated with a urine cell pellet BKV VP1 mRNA copy measurement that has been validated as a noninvasive diagnostic biomarker for BKVN[27,28] (Spearman: $\rho = 0.73$, $p = 1.0 \times 10^{-6}$).

We quantified bacterial urinary cfDNA in 27 urine samples from 17 patients who had a corresponding positive conventional bacterial culture on the same day. For 26 out of 27 clinically positive samples (based on conventional bacterial culture), unbiased sequencing of urinary cfDNA detected the clinically reported uropathogen (Fig. 2B), with 25 of the 27 pathogens identified by culture being in the top 10 abundant organisms identified by cfDNA. For a single sample, urinary cfDNA did not detect the uropathogen reported in culture (*Raoultella* was not detected in the corresponding diagnosed sample, see detailed discussion of this discordant readout and its clinical implications in methods). In another sample, we confirmed co-detection of *Enterococcus* and *Staphylococcus*, matching the only clinically reported polymicrobial infection in this study – a co-infection of *Enterococcus faecalis* and coagulase-negative *Staphylococcus*. To test the performance of urinary cfDNA to identify specific genera, we compared the relative abundance measured for patients diagnosed with a specific agent (n = 27) against the relative abundance measured for 40 negative urine cultures (defined as < 10,000 CFU/mL) from 29 patients (Fig. 2C). We find agreement between urinary cfDNA and clinical testing for *Enterococcus* (n=9, AUC 95% CI 0.911-1.000), *E. coli* (n=9, AUC 95% CI 0.904-1.000), *Klebsiella* (n=3, AUC 1.000), and *Pseudomonas* (n=3, AUC 1.000).

In just over half (59%, 16 of the 27 UTI cases) of examined samples, we found that the uropathogen reported by culture was the most prevalent pathogen in the sample (Fig. 2B). Whereas reports of bacterial culture are skewed towards species that are responsive to culture, cfDNA sequence analyses are sensitive to the full spectrum of uropathogens.

Cell-free DNA reveals frequent undiagnosed viral infections

We next screened for the occurrence of viral uropathogens, including those not tested for in the clinical protocol currently in place in our transplant program. Approximately two-thirds of samples (n = 78) had detectable levels of clinically relevant viruses in the whole cohort of 115 urine specimens. Figure 2D highlights the occurrence of different viruses across specific patient groups and reveals the frequent occurrence of infections with JC polyomavirus, Merkel cell polyomavirus, several herpesviruses (Epstein-Barr virus, cytomegalovirus, human herpesvirus 6A and 6B, human herpesvirus 1) and various

known oncoviruses in this cohort. In addition, we detected parvovirus in twelve samples, with a very high abundance in one sample (1.715 x10^4 RGE). Several patients were simultaneously infected with different polyomavirus species (Merkel cell, JC polyomavirus, or BK polyomavirus). These data illustrate the disconnect that exists between the frequency of current clinical infection testing and the incidence of viral pathogens in this cohort of transplant patients.

Quantifying bacterial growth rates

Conventional metagenomic sequencing can provide a snapshot of the microbiome, yet does not inform about microbial life cycles or growth dynamics. In a recent study, Korem et al. reported that the pattern of metagenomic sequencing read coverage across a microbial genome can be used to quantify microbial genome replication rates for microbes in complex communities[18]. Here, we tested whether this concept can be used to estimate bacterial growth dynamics from measurements of cell-free DNA. Figure 3A shows the urinary cfDNA sequence coverage for four bacterial species, *E. coli*, *K. pneumoniae*, *G. vaginalis* and *P. acnes*. For two patients diagnosed with *E. coli* and *K. pneumoniae* UTI (Fig. 3A), the *E. coli* and *K. pneumoniae* genome coverage was highly non-uniform, with an overrepresentation of sequences at the origin of replication and an underrepresentation of sequences at the replication terminus. The shape of the *E. coli* and *K. pneumoniae* genome coverage is a result of bi-directional replication from a single origin of replication. The skew in genome coverage reflects the bacterial growth rate, where a stronger skew signals faster growth[29]. The genome coverage of a typically commensal bacterial species, *Gardnerella vaginalis*, exhibited non-uniform genome coverage (Fig. 3A, *G. vaginalis*), similar to the above uropathogens but less pronounced. *Propionibacterium acnes* has been recognized as a common skin and lab contaminant[30]. The genome coverage for *P. acnes*, was highly uniform, indicative of slow or no growth (aggregate across 81 samples, Fig. 3A).

We asked whether this measure of bacterial growth can be used to inform UTI diagnosis. We calculated an index of replication based on the shape of the sequencing coverage using methods described previously[29]. For 81 samples, we used BLAST to identify abundant bacterial strains and then re-aligned all sequences with BWA to a curated list of bacterial species. Samples for which the genome coverage was too sparse or too inhomogeneous were excluded from this analysis (see Methods). Figure 3B compares the index of replication for bacteria in samples from patients diagnosed with UTI, to bacteria in samples from patients with negative cultures and samples collected from patients before and after UTI diagnosis. Species categorized in the UTI group had markedly greater growth rates, than those in the no UTI and pre-/post-UTI groups (two-tailed Wilcox rank sum test, p = 0.034). A single-snapshot measurement of bacterial

growth from cfDNA may enable identification of virulent microbial strains and evaluation of the response to anti-bacterial drug treatments.

Antimicrobial resistome profiling

For 25 of 27 samples collected from patients with clinically confirmed UTIs, we determined the relative abundance of genes conferring resistance to several classes of antibiotics. Here, we aligned non-human sequences against known antibacterial resistance genes and mutations using blastp again the Comprehensive Antibiotic Resistance Database (CARD)[31] (see Methods). AR gene sequences were aggregated and called against a non-redundant CARD reference that indicates the drug resistance conferred by the given gene.

We compared clinical profiles of antimicrobial susceptibility testing (see Methods) to the resistance profiles determined by sequencing. For most samples, there was a high diversity in alignments with highly abundant resistance classes including resistance to macrolides, aminoglycosides, and beta-lactams (Fig. 4). We studied vancomycin-resistant *Enterococcus* (VRE) infections, which often lead to complications after transplantation, in depth. For all samples diagnosed with *Enterococcus* UTI (9 samples from 6 patients), resistance to vancomycin was assessed via measurement of minimum inhibitory concentration or bioMérieux Etest. We detected fragments of genes conferring resistance to the glycopeptide antibiotic class, of which vancomycin is a member, for all VRE positive samples (n=3). Moreover, for samples with *Enterococcus* UTI that were identified as vancomycin susceptible (n=6), we did not detect glycopeptide class resistance. When expanding to all samples, we identified vancomycin resistance in one additional sample from a patient who had developed VRE UTI prior to the collection of the urine sample but was untreated (see below). These data indicate significant potential to predict antimicrobial susceptibility from measurements of urinary cfDNA.

Host response to infection

We next examined the host response to viral and bacterial infections. Recent work has identified transplant donor specific cfDNA in plasma as a marker of graft injury in heart, lung, liver and kidney transplantation[14,25,32,33]. Here, we quantified donor specific cfDNA in urine for sex-mismatched donor recipient pairs by counting Y chromosome derived cfDNA (Fig. 5A, Methods). We observed elevated levels of donor cfDNA for samples from patients diagnosed with BKVN (mean proportion of donor DNA 65.6%, n=13) compared to samples from patients who had normal protocol biopsies without diagnosis of BKVN (mean 51.4%, n=4) and samples from patients who did not develop a clinical UTI in the first 3 months of transplantation (mean 23.4%, n=12, samples collected within 5 days

after transplant excluded). The release of donor DNA reflects severe cellular and tissue injury in the graft, a hallmark of BKVN. In contrast to patients diagnosed with polyomavirus infection, patients diagnosed with bacterial UTI had lower proportions of donor DNA as compared to stable individuals. This is likely explained by an elevated number of recipient immune cells in the urinary tract following immune activation. Indeed, comparison to clinical urinalysis indicates that the donor fraction decreases with white blood cell count (WBC) per high power field, HPF (400x magnification, inset Fig. 5A, Spearman: $\rho = -0.424$, $p = 0.01$). Furthermore, clinical cases of pyuria – defined as greater than ten WBC per HPF – had a lower donor fraction than those without (two-tailed Wilcox rank sum test, $p = 0.01$). In addition, we found that the level of donor DNA in the first few days after transplant was elevated, consistent with early graft injury, and in line with earlier observations in heart and lung transplantation[25,34]. We tracked the relative and absolute abundance of donor specific urinary cfDNA in the first few days after transplantation for a small subset of patients. The initial elevated level of donor DNA quickly decayed to a lower baseline level (Fig. 5B), in line with previous observations in heart and lung transplantation[14,35].

Two recent studies have demonstrated that the structure of chromatin in gene promoters is conserved within circulating cfDNA in plasma[19,36]. Ulz et al. employed whole-genome sequencing of plasma DNA to show that nucleosomal occupancy at transcription start sites (TSSs) results in different read depth coverage patterns for expressed and silent genes[19]. Here, we have found for the first time that the footprints of nucleosomes in gene promoters and transcriptional regulatory elements are conserved within urinary cfDNA (Fig. 5C, aggregation and normalization across all samples), and that the extent of nucleosomal protection is proportional to gene expression. Measurements of nucleosomal depletion can serve as a proxy for gene expression, and may be used to investigate host-pathogen interactions in the scope of urinary tract infection in more detail.

Mitochondrial DNA (mtDNA) in the urine was recently identified as a possible biomarker for hypertensive kidney damage[37]. Recent data furthermore indicate a role for extracellular mitochondrial DNA as a powerful damage-associated molecular pattern (DAMP) - elevated levels of mtDNA in plasma have been reported in trauma, sepsis and cancer and recent studies have identified mitochondrial DNA released into the circulation by necrotic cells[38]. For a small subset of patients diagnosed with BKVN (8 samples from 7 patients), we quantified donor and recipient specific mtDNA in urine, using an approach we have previously described[17]. We found that the graft is the predominant source of mitochondrial urinary cfDNA in seven of the eight samples (two-tailed Student t-test, $p \ll 10^{-6}$; see Methods). Molecular techniques to track DAMPs in urine released in the scope of kidney graft injury may provide a non-invasive window into the potential role of these molecules in the pathogenesis of immune-related complications.

Finally, to illustrate the utility of urinary cfDNA monitoring, we conclude with a case study. A kidney transplant recipient clinically presented with an *E. coli* infection on post-transplant day 13 and was placed on a two-week course of ertapenem/meropenem, starting post-operative day 15. On day 20 (cfDNA available), the patient had a urine culture positive for VRE infection while on ertapenem/meropenem. On day 27 (cfDNA available), a urine culture was negative for both *E. coli* and VRE. Urinary cfDNA profiling revealed a high abundance of *Enterococcus* cfDNA at day 20 (2.473 RGE), as well as the presence of glycopeptide class resistance (GCR) genes. A low donor fraction was measured (11.8%), indicative of immune activation. At day 27, the presence of *Enterococcus* cfDNA (1.028 RGE) and GCR genes diminished while the fraction of donor cfDNA nearly doubled (D = 21.2%), indicating successful treatment. However, by day 31, four days after the antibiotic treatment concluded, the abundance of *Enterococcus* cfDNA (49.5 RGE) and GCR genes had increased, and the donor fraction decreased (D = 10.3%), indicating relapse. Conventional culture was not performed at that time. The patient in question later developed *E. coli* UTIs (post-operative day 232, and 337) and VRE UTIs (post-operative day 260 and 337). This case illustrates how the different layers of information acquired from cfDNA - microbiome composition, recipient DNA, and ARG identification – might have predicted the recurrence of infection in this patient.

Discussion

We have presented a strategy to comprehensively identify and assess infections of the urinary tract based on profiling of urinary cfDNA and '*omics* analysis principles. We show that different layers of clinical information are accessible from a single assay that are either inaccessible using current diagnostic protocols, or require parallel implementation of a multitude of different tests. In nearly all samples with clinically reported viral or bacterial infection of the urinary tract, cfDNA identified the causative agent of infection. cfDNA furthermore revealed the frequent occurrence of both viral and bacterial (co-)infections that remain unidentified in current clinical practice. In many samples, including samples from patients regarded as clinically stable, we detected viral infections that may be clinically relevant but not routinely assayed in the screening protocol at our institution. The assay we present therefore has the potential to become a valuable tool for the monitoring of bacterial and/or viral infections in transplant cohorts, and ascertain their potential impact on allograft health.

Cell-free DNA in the blood circulation is, to a large extent, cleared from the blood via the kidneys. Analyses of urinary cfDNA can therefore reveal infections that occur throughout the body, and are not limited to the detection of infections of the urinary tract. Blood collection requires a minimally invasive procedure, whereas urine can be collected

entirely noninvasively, repeatedly, and without need for clinical visitation. In this light, urine may prove an ideal specimen type for whole-body infection analysis based on unbiased sequencing of cfDNA.

Beyond measurements of the abundance of different components of the microbiome, urinary cfDNA provides information about uropathogen phenotypes. We show, for the first time, that analysis of the structure of microbial genomes from cfDNA allows estimating bacterial growth rates, thereby providing dynamic information from a single snapshot. We compared the bacterial growth rates in samples with clinically-diagnosed UTI to those with no confirmed UTI and we observed higher growth rates for both clinically-diagnosed and co-infecting bacteria in patients with infection. The bacterial growth rate measurement may have the potential to measure the effect of antibiotic treatment from a single snapshot.

We further show that metagenomic analysis of urinary cfDNA can be used to evaluate the susceptibility to antibiotics. We compared sequencing data to a curated database of genes conferring antibacterial resistance, and we found a good agreement between the alignments to AR genes and reports of drug resistance determined via minimum inhibitory concentration measurements. cfDNA resistome profiling may have an added potential over conventional culture as cultures report on the susceptibility testing of one or few cultured colonies. cfDNA profiling can capture AR gene fragments from the entire bacterial population which may be particularly important since cfDNA profiling reveals frequent co-infections within the UTI group. Linking determinants of bacterial resistance to specific strains is challenging for complex infections. UTI offers a good test ground for such analyses, given the low complexity and abundance of normal flora that challenge AR analyses.

More than 15,000 patients receive lifesaving kidney transplants in the US each year[39]. Viral and bacterial infections of the urinary tract occur frequently in this patient group and often lead to serious complications, including graft loss and death. In the general population, UTI is one of the most frequent medical problems that patients present with in medical offices[40]. Unbiased sequencing of cfDNA offers a comprehensive window into the pathobiology of infections of the urinary tract and can be a valuable future diagnostic tool to monitor and diagnose bacterial and viral infections in kidney transplantation as well as in the general population. The assay we have presented can be robustly implemented on low volumes of urine and will benefit from continued technical advances in DNA sequencing that will reduce cost and assay turnaround time in the years to come.

Methods

Study cohort and sample collection. 115 urine samples were collected from kidney transplant recipients who received their clinical care at New York Presbyterian Hospital – Weill Cornell Medical Center. We assayed urine samples from a total of 67 patients. We included 21 subjects who developed bacterial UTIs diagnosed within the first 3 months of transplantation and 14 subjects who never developed urinary tract infections within the first 3 months of transplantation. For the 21 subjects who developed UTIs, we assayed 27 urine samples from 17 subjects who had corresponding, same day positive urine cultures (UTI Group); we assayed 15 urine samples from 15 subjects, collected at least 2 to 25 days (median 7 days) prior to development of the positive urine cultures (Pre-UTI Group), and we assayed 10 urine samples from 8 subjects, collected at least 3 to 31 days (median 12 days) after development of the positive urine cultures (Post-UTI Group) (7 of the 8 subjects were treated with antibiotics). We assayed a total of 29 samples collected within three months after transplantation from 14 subjects who never developed UTI in that time period. Among the 81 samples, 67 (83%) had a corresponding same day urine culture with the associated urine specimens that were assayed. The study additionally included 24 samples from 22 subjects who had a corresponding positive diagnosis of BK virus nephropathy by needle biopsy of the kidney allograft (BKVN positive group) and 10 samples from 10 subjects who had a normal protocol biopsy and was negative for BK virus (BKVN negative group). See also detailed clinical metadata in supplemental datatable “Clinical Data”.

Urine supernatant isolation. Approximately 50 mL of urine was centrifuged at 3,000 g and the supernatant was stored at -80 ˚C in 1 or 4 ml aliquots. Cell-free DNA was extracted from 1 ml (105 samples) or 4 ml (10 samples) of urine (Qiagen Circulating Nucleic Acid Kit, Qiagen, Valencia, CA).

Analysis of discordance against bacterial culture. In a single sample, urinary cfDNA did not identify the uropathogen reported by conventional culture (*Raoultella*). The subject had developed a prior *Escherichia coli* UTI by conventional bacterial culture on post op day 6, was treated with initially aztreonam but switched to cephalexin for 14 day course. The subject developed a *Raoultella* UTI by conventional bacterial culture on post op day 25 in which cfDNA analysis on the same day had high abundance of Escherichia coli UTI but no evidence of *Raoultella* infection. Given the discordant results, it is unclear if the second UTI is a recurrence as suggested by the cfDNA analysis or is a new infection as suggested by the urine culture data.

Negative control. To control for environmental and sample-to-sample contamination, a known-template control sample (IDT-DNA synthetic oligo mix, lengths 25, 40, 55, 70 bp, 0.20 µM eluted in TE buffer) was included in every sample batch and sequenced to approximately 25% of the depth of the cfDNA extracts (~5 million fragments). The five

most abundant genera detected across the 23 microbiome controls consisted of *Propionibacterium* (22.9%), *Salmonella* (10.9%), *Pseudomonas* (7.7%), polyomavirus (5.4%), and *E. coli* (5.2%). The mean representation of each genus in the control was used to filter out genera in samples identified as possible contaminants. Possible sources of contamination in these experiments include: environmental contamination during sample collection in the clinic, nucleic acid contamination in reagents used for DNA isolation and library preparation, sample-to-sample contamination due to Illumina index switching[41].

Library preparation and next generation sequencing. Sequencing libraries were prepared using a single-stranded library preparation optimized for the analysis of ultrashort fragment DNA, described previously[17]. Libraries were characterized using the AATI fragment analyzer. Due to the use of a custom read primer in place of Illumina read sequencing primer 1, digital PCR was not performed. Samples were pooled and sequenced on the Illumina NextSeq platform (paired-end, 2x75 bp). Approximately 50 million paired-end reads were generated per sample.

Analysis – Composition of the urinary microbiome. Low quality bases and Illumina-specific sequences are trimmed (Trimmomatic-0.32[42]). Reads from short fragments were merged and a consensus sequence of the overlapping bases is determined using FLASH-1.2.7. Reads are aligned (Bowtie2, very sensitive mode[43]) against the human reference (UCSC hg19). Unaligned reads are extracted, and the non-redundant human genome coverage is calculated (SAMtools 0.1.19 rmdup[44]). To derive the urinary microbiome, reads are BLASTed (NCBI BLAST 2.2.28+) to a curated list of bacterial and viral reference genomes[45]. Short reads are assigned to specific taxa using a maximum likelihood algorithm that takes into account the ambiguity of read mapping[26], as described in previous work[25,46]. The relative abundance of higher level taxa are determined on the basis of the genomic abundance at the strain or species level.

Bacterial growth dynamics. We determined bacterial genome replication rates using the methods described by Brown et al.[29]. All bacterial strains within a sample were sorted and the GC-skew (the cumulative summation of GC content for sliding 10 bp windows across the genome) was used to identify the origin and terminus of replication (minimum and maximum GC-skew, respectively). We binned the genome by 1 kbp tiles, smoothed the coverage by running mean of 100 nearest neighboring tiles, determined the coverage in each tile and sorted the middle 90% of the tiles by coverage. We performed linear regression between the origin and terminus of replication after further removing the 5% least and most covered bins. The product of the slope of the regression line and the genome length was defined as the growth rate, a metric applied in previous analysis[29]. This analysis was applied for all bacterial strains with genome lengths greater than 0.5

Mbp, R^2 linear regression correlation greater than 0.90, and GINI index coefficient less than 0.2, for which at least 2500 BLAST hits were detected in the sample.

Nucleosome footprints in gene bodies. Paired-end reads are aligned using BWA (mem) to the hg19 human reference genome. Sequence read coverage around a list of TSS to the hg19 build, as described in similar work[19], is determined based on BAM files using the SAMtools depth function. Nucleosomal protection at the TSS of specific genes is determined based on the loss of nucleosomal protection in a 2 kbp window centered on the TSS.

Proportion of donor-specific cfDNA in urine. The fraction of donor specific cfDNA in urine and plasma samples was estimated using approaches that modified from Refs.[14,17,25] to allow for donor estimates in the absence of donor genotype information. Sixty-three samples in the study were from sex-mismatched, donor-recipient pairs. Following adapter trimming and quality filtering, reads were aligned to the human genome build hg19. We processed human-aligned reads using HMMcopy[47], binning the genome into windows of 500 bp to adjust for mappability and GC content. We further removed regions of the chrY that had high coverage in female-female donor-recipient pairs. We determined the donor fraction as follows:

Male donor: $D = 2Y/A$,

Female donor: $D = 1 - \left(\frac{2Y}{A}\right)$,

where Y and A are the coverage of the mappability-adjusted Y and autosomal chromosomes, respectively. SNP-genotyping information obtained from a recipient pre-transplantation whole blood sample can be used to distinguish donor and recipient sequences for non-gender matched transplants. We compared the donor fraction estimated from the representation of sex chromosomes and autosomes to the donor-fraction determined from SNP-genotyping for 16 samples available from a previous study and found a strong agreement for the two methods (Pearson's corr. of 0.973, $p << 10^{-6}$).

Mitochondrial donor fraction. We determine the mitochondrial donor fraction using methods previously described[17]. We extracted genomic DNA from pre-transplant donor and recipient blood in parallel and amplified the mitochondrial DNA using the REPLI-g Mitochondrial DNA Kit (Qiagen). Amplified DNA was sheared to 300 bp and prepared and indexed using the NEBNext Ultra kit for Illumina (New England Biolabs, Ipswich, MA). We sequenced the consensus mitochondrial genomes on one lane of the Illumina MiSeq platform (2x150). Raw reads were trimmed and aligned to the human genome using BWA mem (hg19); average sequencing depth across the mitochondrial genome for sixteen samples (eight donor-recipient pairs) ranged from 50x-200x. Subsequent .bam files were analyzed for the nucleotide distribution at each base pair and the most frequent was

chosen as the consensus nucleotide at that position. Post-transplant, cfDNA was also aligned to hg19 to determine mitochondrial-aligned cfDNA (mt-cfDNA) and we identified the frequency of each nucleotide at each base using bam-readcount (version 0.8.0). One sample was removed due to low depth of sequencing. We employed a two-tailed Student t-test under the alternative hypothesis that the mean mitochondrial donor fraction was greater than 50% to determine if the graft was the major contributor to mt-cfDNA in the urinary tract.

Clinical antimicrobial resistance determination. Antimicrobial susceptibility testing was performed on twenty-four matched samples from patients with clinically diagnosed UTIs at New York Presbyterian Hospital – Weill Cornell Medical Center. All tests considered in the analysis used minimal inhibitory concentration, except in one measurement of vancomycin resistance for one sample a bioMérieux Etest was performed, to assess the resistance of the cultured pathogenic microbe to up to twenty-one antibiotics, though we only focus on vancomycin resistance.

Determining antimicrobial resistance gene presence. Nonhuman sequencing reads were aligned to a database of protein sequences of known antimicrobial resistance genes (CARD) using blastx. In many samples, the number of bacterial fragments were not deep enough to statistically identify SNP mutations. We, therefore restrict our identification only to resistance conferred via protein homolog, of which 2,158 genes were provided. We ran blastx such that we required the identity overlap to be 90% and the culling limit to be eight hits. The hits with the highest identity and overlap length were selected for each read and compared to the antimicrobial resistance classes using the CARD ontology. We confirmed the general utility of this approach using a short-read metagenomic sequence alignment software, Shortbred[48]. Briefly, the aforementioned reference genes from CARD were assembled into a proper reference using the shortbred_identify tool. Subsequently, this reference was aligned to our translated data using blastp (cutoff 85% identity). We confirmed the accuracy of this method by using a second, more stringent short-read metagenomic alignment tool, Shortbred[48]. For samples with many alignments to the CARD RGI reference database (>1000), the relative abundances of genes from the various antibiotic classes are strongly correlated between the two methods (mean Pearson correlation 0.851 for six samples).

Statistical analysis. All statistical analysis was performed using R 3.3.2. Unless otherwise noted, distribution comparison was performed using a Mann-Whitney U Test.

Data availability. The sequencing data that support the findings of this study will be made available in the database of Genotypes and Phenotypes (dbGaP).

Acknowledgments. This work was supported by R21 AI 133331 (I.D.V. and J.R.L.), R21AI124237 (I.D.V.), K23 AI 124464 (J.R.L.), R37 AI 051652 (M.S.), and the Robert Noyce Foundation (IDV). P.B. is supported by a NSF GRFP, DGE-1144153. We thank Erin Berthelsen for providing samples for assay development, and Catherine Snopkowski and Carol Li for help with sample handling and qPCR experiments.

Competing financial interests. The authors have no competing financial interests.

Materials & Correspondence. All requests should be submitted to I.D.V. (vlaminck@cornell.edu) and J.R.L. (jrl2002@med.cornell.edu).

Author contributions. P.B. contributed to the study design, experiments and data analysis. D.D. contributed to the study design, and provided clinical expertise. M.H. and F.C. performed experiments. M.S. contributed to the study design, and provided clinical expertise. J.R.L. contributed to the study design, and contributed clinical expertise. I.D.V. contributed to the study design, and data analysis. P.B and I.D.V. wrote the manuscript. All authors provided comments and edits.

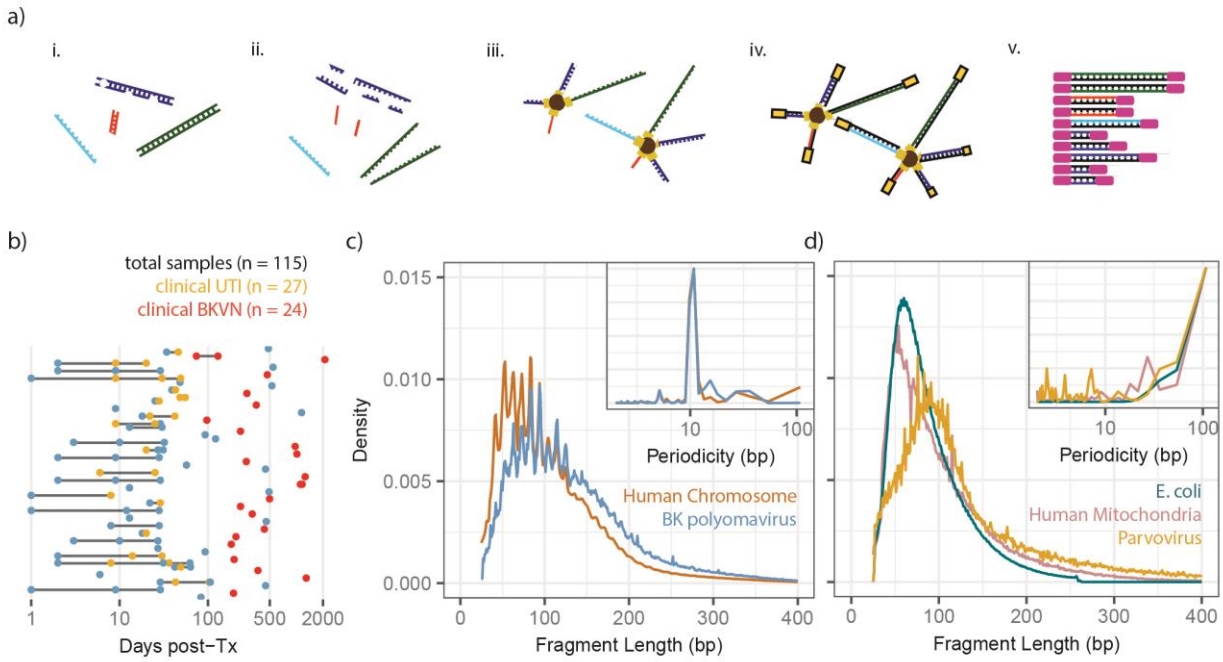

Figure 1. Unbiased sequencing assay and biophysical properties of urinary cfDNA. **(a)** Schematic representation of the ssDNA library preparation protocol used for unbiased sequencing of urinary cfDNA[17]. Key steps include: *i*) cfDNA isolation, *ii*) DNA denaturation, *iii*) ssDNA adapter ligation, *iv*) extension and double-stranded DNA adapter ligation, *v*) and PCR. **(b)** Overview of post-transplant sample collection dates (color indicates pathology, bars connect samples from same patients). **(c-d)** Fragment length density plot measured by paired-end sequencing for different cfDNA types: (**c**) chromosomal and polyomavirus cfDNA from representative samples, and (**d**) *E. coli*, parvovirus, and mitochondrial cfDNA from representative samples. Fourier analysis reveals a 10.4 bp periodicity in the fragment length profiles of chromosomal and BK polyomavirus cfDNA but not in *E. coli,* parvovirus, and mitochondrial cfDNA (insets). See supplemental datatable “Clinical Data”.

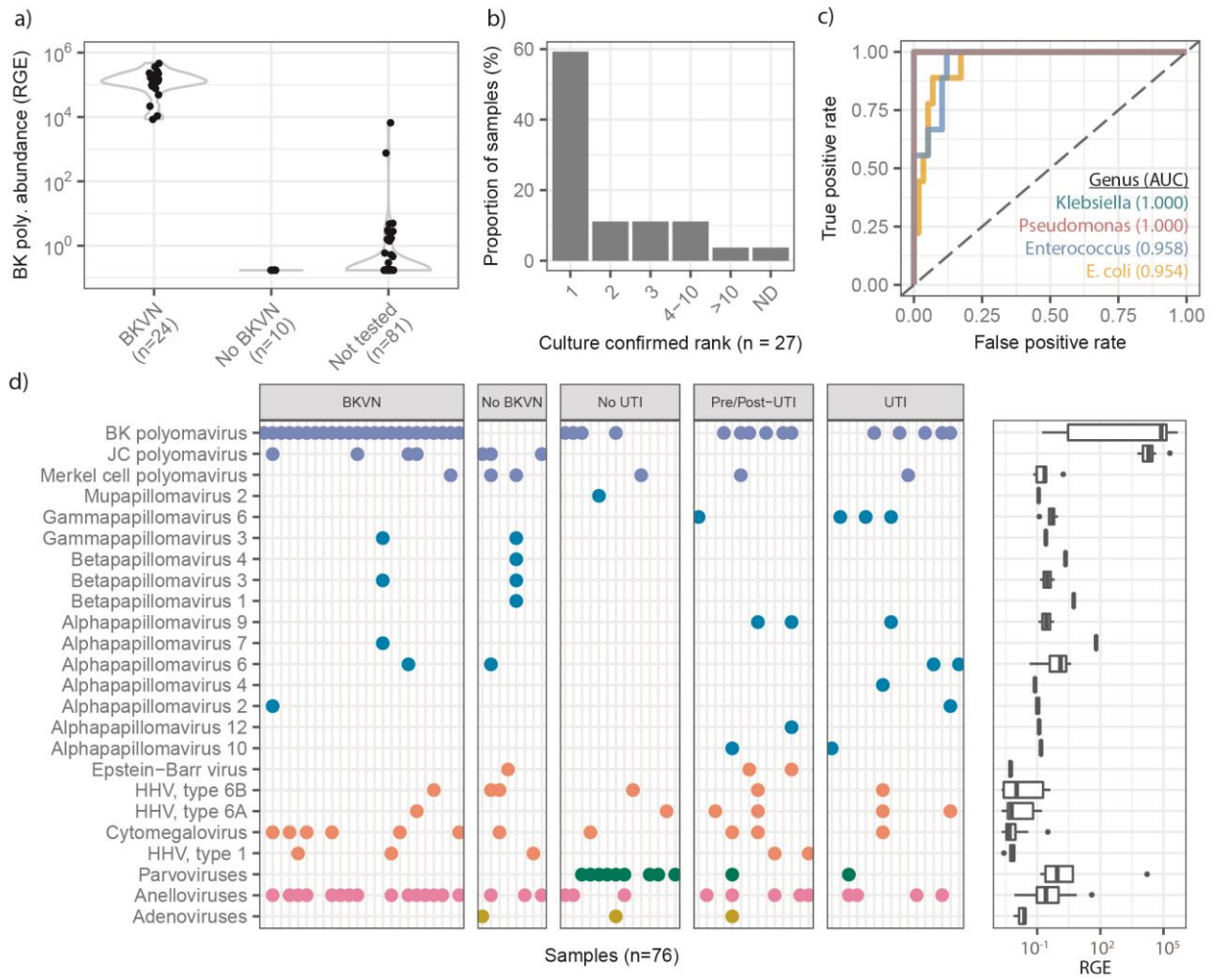

Figure 2. Urinary cell-free DNA confirms clinical infections and identifies co-infecting microbes. (a) Violin plots of BK Polyomavirus cfDNA sequence abundance (relative genome equivalents, RGE) for patients that are BKVN positive per needle biopsy of the kidney allograft, patients that are BKVN negative per biopsy, and untested patients. **(b)** Rank order genomic abundance for clinically reported uropathogens in the urine of patients with bacterial UTI. In approximately half of all samples (59%), the bacteria detected in culture was the most abundant component of the cfDNA urinary microbiome. In one sample, the clinically reported agent was not detected (ND). **(c)** Receiver operating characteristic analysis of the performance of urinary cfDNA in identifying UTIs with common bacterial uropathogens (67 urine samples, *Klebsiella, Pseudomonas, Enterococcus,* and *Escherichia coli* UTI). **(d)** cfDNA reveals frequent undiagnosed viral infections in this cohort (left panel). Boxplot representation of the distribution of RGE (right panel). See associated supplemental datatables.

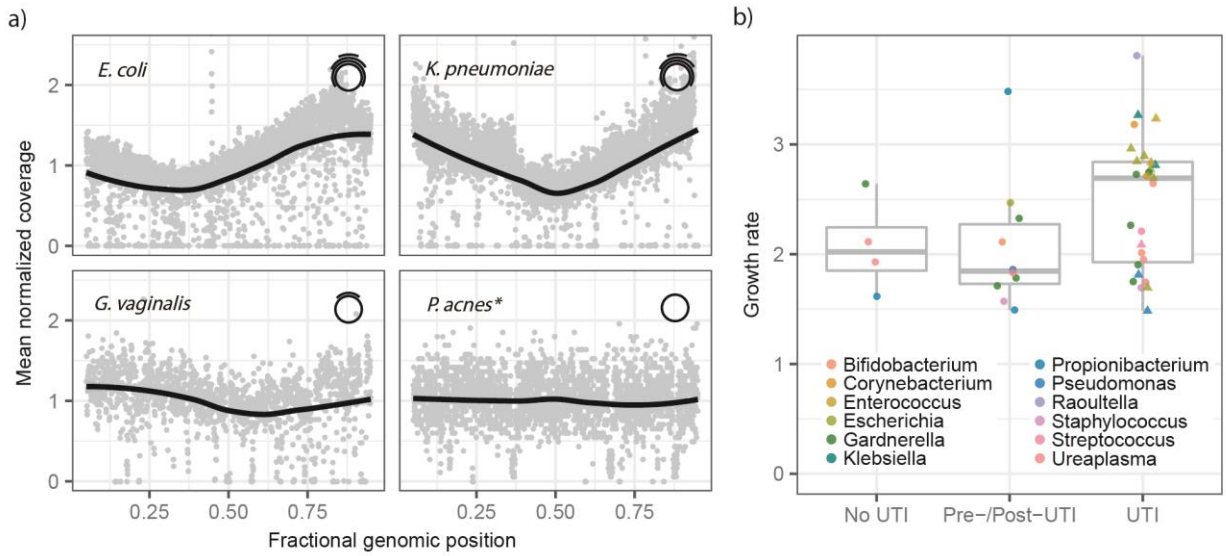

Figure 3. Estimating bacterial growth rates from urinary cfDNA. (a) Normalized bacterial genome coverage for four representative bacterial species. The coverage was binned in 1 kbp tiles and normalized. Each panel represents a single sample (see Supplementary data table), with the exception of *P. acnes* (*) for which the coverage was aggregated across 81 samples. Solid line is a loess filter smoothing curve (span = 0.70). The non-uniform genome coverage for *E. coli* and *K. pneumoniae*, with an overrepresentation of sequences at the origin of replication, is a result of bi-directional replication from a single origin of replication. **(b)** The skew in genome coverage reflects the bacterial growth rate, where a stronger skew signals faster growth[29]. Growth rates for species in twelve genera grouped by patient cohort (at least 2500 alignments, 41 samples). Each point indicates a bacterial species in a sample. Triangles indicate culture-confirmed bacteria by genus. See associated supplemental datatables.

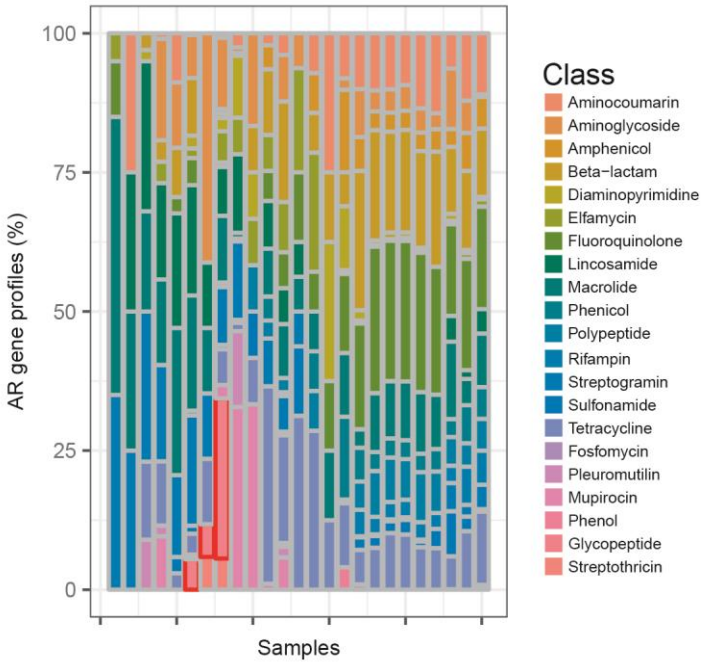

Figure 4. cfDNA based antimicrobial resistome profiling. For 25 samples with clinically confirmed UTIs, AR gene profiling reveals the presence of genes conferring resistance to various drug classes. Genes that confer resistance to glycopeptide class antibiotics (including vancomycin) are highlighted in red outlines. See associated supplemental datatables.

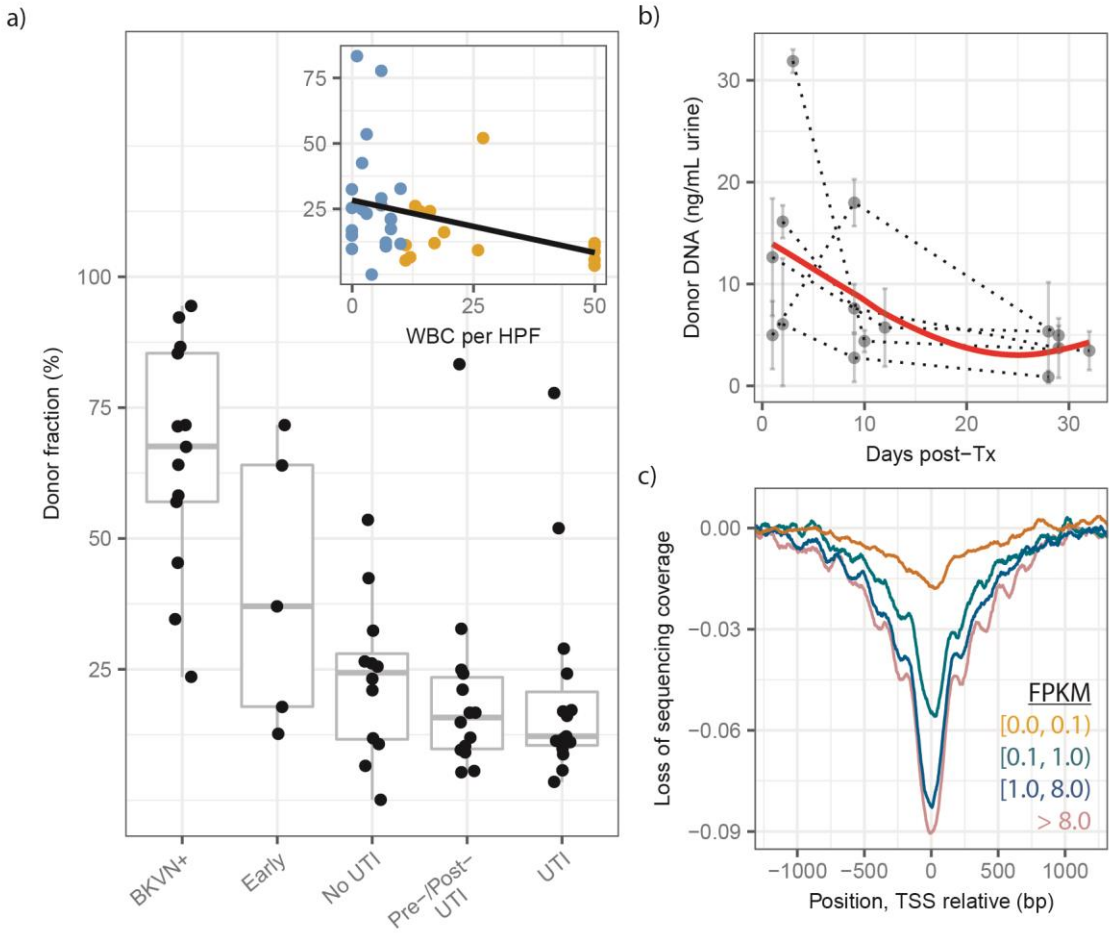

Figure 5 The host response to infection from urinary cell-free DNA. (a) Proportion of donor-specific cfDNA in samples from patients that are BKVN positive per kidney allograft biopsy (BKVN+), in samples collected in the first five days after transplant surgery (Early), samples collected from patients that UTI negative per culture (No UTI), samples collected before and after UTI (Pre-/Post-UTI) and samples collected at the time of UTI diagnosis (UTI). Two outliers in Pre-/Post-UTI and UTI groups correspond to the same patient, who suffered an acute rejection episode in the months prior. Low donor fractions in the Pre-/Post-UTI and UTI groups are likely due to immune cell presence in the lower urinary tract; patients with higher white blood cell counts have lower donor fractions (inset, orange indicates pyuria) **(b)** Absolute abundance of donor cfDNA in the urine of patients not diagnosed with infection in the first month post-transplant. **(c)** Genome coverage at the transcription start site (TSS), binned by gene expression level across all (n=115) samples in the study. FPKM = Fragments per Kilobase of Transcript per Million mapped reads, an RNA-seq measure of gene expression. See associated supplemental datatables.

References

1. Foxman, B. Urinary tract infection syndromes. Occurrence, recurrence, bacteriology, risk factors, and disease burden. *Infectious Disease Clinics of North America* **28,** 1–13 (2014).
2. Abbott, K. C. *et al.* Late urinary tract infection after renal transplantation in the United States1 . *Am. J. Kidney Dis.* **44,** 353–362 (2004).
3. Ariza-Heredia, E. J. *et al.* Impact of urinary tract infection on allograft function after kidney transplantation. *Clin. Transplant.* **28,** 683–690 (2014).
4. Chuang, P., Parikh, C. R. & Langone, A. Urinary tract infections after renal transplantation: a retrospective review at two US transplant centers. *Clin. Transplant.* **19,** 230–235 (2005).
5. Hirsch, H. H. *et al.* Polyomavirus-associated nephropathy in renal transplantation: interdisciplinary analyses and recommendations. *Transplantation* **79,** 1277–1286 (2005).
6. Dadhania, D. *et al.* Epidemiology of BK virus in renal allograft recipients: independent risk factors for BK virus replication. *Transplantation* **86,** 521–528 (2008).
7. Schmiemann, G., Kniehl, E., Gebhardt, K., Matejczyk, M. M. & Hummers-Pradier, E. The Diagnosis of Urinary Tract Infection: A Systematic Review. *Deutsches Ärzteblatt International* **107,** 361–367 (2010).
8. Price, T. K. *et al.* The Clinical Urine Culture: Enhanced Techniques Improve Detection of Clinically Relevant Microorganisms. *J. Clin. Microbiol.* **54,** 1216–22 (2016).
9. Heytens, S. *et al.* Women with symptoms of a urinary tract infection but a negative urine culture: PCR-based quantification of Escherichia coli suggests infection in most cases. *Clin. Microbiol. Infect.* (2017). doi:10.1016/j.cmi.2017.04.004
10. Quake, S. Sizing up cell-free DNA. *Clin. Chem.* **58,** 489–490 (2012).
11. Fan, H. C. *et al.* Non-invasive prenatal measurement of the fetal genome. *Nature* **487,** 320–4 (2012).
12. Dennis Lo, Y. M. *et al.* Presence of fetal DNA in maternal plasma and serum. *Lancet* **350,** 485–487 (1997).
13. 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. U. S. A.* **105,** 16266–16271 (2008).
14. De Vlaminck, I. *et al.* Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. *Sci. Transl. Med.* **6,** 241ra77 (2014).
15. Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6,** 224ra24 (2014).
16. Gansauge, M.-T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8,** 737–48 (2013).
17. 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).

18. Korem, T. *et al.* Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349,** 1101–1106 (2015).
19. Ulz, P. *et al.* Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* **48,** 1273–1278 (2016).
20. Botezatu, I. *et al.* Genetic analysis of DNA excreted in urine: a new approach for detecting specific genomic DNA sequences from cells dying in an organism. *Clin. Chem.* **46,** 1078–1084 (2000).
21. Zhang, J. *et al.* Presence of donor- and recipient-derived DNA in cell-free urine samples of renal transplantation recipients: urinary DNA chimerism. *Clin. Chem.* **45,** 1741–1746 (1999).
22. Su, Y.-H. *et al.* Human urine contains small, 150 to 250 nucleotide-sized, soluble DNA derived from the circulation and may be useful in the detection of colorectal cancer. *J. Mol. Diagn.* **6,** 101–7 (2004).
23. Tsui, N. B. Y. *et al.* High Resolution Size Analysis of Fetal DNA in the Urine of Pregnant Women by Paired-End Massively Parallel Sequencing. *PLoS One* **7,** 1–7 (2012).
24. Varshavsky, A. J., Bakayev, V. V, Chumackov, P. M. & Georgiev, G. P. Minichromosome of simian virus 40: presence of histone HI. *Nucleic Acids Research* **3,** 2101–2113 (1976).
25. De Vlaminck, I. *et al.* Noninvasive monitoring of infection and rejection after lung transplantation. *Proc. Natl. Acad. Sci.* **112,** 13336–13341 (2015).
26. 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).
27. Ding, R. *et al.* CD103 mRNA levels in urinary cells predict acute rejection of renal allografts. *Transplantation* **75,** 1307–1312 (2003).
28. Dadhania, D. *et al.* Validation of Noninvasive Diagnosis of BK Virus Nephropathy and Identification of Prognostic Biomarkers. *Transplantation* **90,** 189–197 (2010).
29. Brown, C. T., Olm, M. R., Thomas, B. C. & Banfield, J. F. Measurement of bacterial replication rates in microbial communities. *Nat Biotech* **34,** 1256–1263 (2016).
30. Mollerup, S. *et al.* Propionibacterium acnes: Disease-Causing Agent or Common Contaminant? Detection in Diverse Patient Samples by Next-Generation Sequencing. *J. Clin. Microbiol.* **54,** 980–987 (2016).
31. McArthur, A. G. *et al.* The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* **57,** 3348–3357 (2013).
32. Grskovic, M. *et al.* Validation of a Clinical-Grade Assay to Measure Donor-Derived Cell-Free DNA in Solid Organ Transplant Recipients. *J. Mol. Diagn.* **18,** 890–902 (2016).
33. Lo, Y. D. *et al.* Presence of donor-specific DNA in plasma of kidney and liver-transplant recipients. *Lancet* **351,** 1329–1330 (1998).
34. De Vlaminck, I. *et al.* Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. *Sci. Transl. Med.* **6,** 241ra77 (2014).
35. Beck, J. *et al.* Digital droplet PCR for rapid quantification of donor DNA in the circulation of transplant recipients as a potential universal biomarker of graft

injury. *Clin. Chem.* **59,** 1732–1741 (2013).
36. Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164,** 57–68 (2016).
37. Eirin, A. *et al.* Urinary Mitochondrial DNA Copy Number Identifies Chronic Renal Injury in Hypertensive Patients. *Hypertens. (Dallas, Tex. 1979)* **68,** 401–410 (2016).
38. Lood, C. *et al.* Neutrophil extracellular traps enriched in oxidized mitochondrial DNA are interferogenic and contribute to lupus-like disease. *Nat. Med.* **22,** 146–153 (2016).
39. Hart, A. *et al.* OPTN/SRTR 2015 Annual Data Report: Kidney. *Am. J. Transplant.* **17,** 21–116 (2017).
40. Foxman, B., Barlow, R., D'Arcy, H., Gillespie, B. & Sobel, J. D. Urinary Tract Infection: Self-Reported Incidence and Associated Costs. *Ann. Epidemiol.* **10,** 509–515 (2000).
41. Sinha, R. *et al.* Index Switching Causes 'Spreading-Of-Signal' Among Multiplexed Samples In Illumina HiSeq 4000 DNA Sequencing. *bioRxiv* (2017). at <http://biorxiv.org/content/early/2017/04/09/125724.abstract>
42. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
43. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9,** 357–359 (2012).
44. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
45. Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res* **12,** (2002).
46. De Vlaminck, I. *et al.* Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell* **155,** 1178–87 (2013).
47. Ha, G. *et al.* Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Res.* **22,** 1995–2007 (2012).
48. Kaminski, J. *et al.* High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED. *PLoS Computational Biology* **11,** (2015).