(12) STANDARD PATENT (11) Application No. **AU 2014346562 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Cell-free nucleic acids for the analysis of the human microbiome and components thereof

(51) International Patent Classification(s)
C12Q 1/68 (2006.01) ***C40B 30/04*** (2006.01)
C12Q 1/70 (2006.01)

(21) Application No: **2014346562** (22) Date of Filing: **2014.11.07**

(87) WIPO No: **WO15/070086**

(30) Priority Data

(31) Number	(32) Date	(33) Country
61/901,857	**2013.11.08**	**US**
61/901,114	**2013.11.07**	**US**

(43) Publication Date: **2015.05.14**
(44) Accepted Journal Date: **2018.11.29**

(71) Applicant(s)
The Board of Trustees of the Leland Stanford Junior University

(72) Inventor(s)
De Vlaminck, Iwijn;Kertesz, Michael;Khush, Kiran K.;Kowarsky, Mark Alec;Martin, Lance;Quake, Stephen R.;Valantine, Hannah

(74) Agent / Attorney
Griffith Hack, GPO Box 1285, Melbourne, VIC, 3001, AU

(56) Related Art
WO 2013159035 A2
US 2002015955 A1
WO 2013156627 A1
HIGHLANDER ET AL., "High throughput sequencing methods for microbiome profiling: application to food animal systems.", ANIM HEALTH RES REV., (201206), vol. 13, no. 1, PAGE 40 - 53
US 2012295810 A1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 May 2015 (14.05.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/070086 A1

(51) International Patent Classification:
C40B 30/04 (2006.01) *C12Q 1/70* (2006.01)
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2014/064669

(22) International Filing Date:
7 November 2014 (07.11.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/901,114 7 November 2013 (07.11.2013) US
61/901,857 8 November 2013 (08.11.2013) US

(71) Applicant: **THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY** [US/US]; 1705 El Camino Real, Palo Alto, California 94306-1106 (US).

(72) Inventors; and
(71) Applicants : **DE VLAMINCK, Iwijn** [—/US]; 318 Campus Drive, Stanford, California 94304 (US). **KERTESZ, Michael** [IL/US]; 328 Trenton Way, Menlo Park, California 94025 (US). **KHUSH, Kiran K.** [—/US]; 300 Pasteur Drive, Falk CVRC 263, Stanford, California 94305 (US). **KOWARSKY, Mark Alec** [AU/AU]; 4/4 Testar Grove, Caulfield North, Victoria 3161 (AU). **MARTIN, Lance** [US/US]; 620 Miramar Drive, Half Moon Bay, California 94019 (US). **QUAKE, Stephen R.** [US/US]; 318 Campus Drive Clark Center E350Q, Stanford, California 94305 (US). **VALANTINE, Hannah** [—/US]; 300 Pasteur Drive, Dean's Office, Room M121, Stanford, California 94305 (US).

(74) Agent: **SHERWOOD, Pamela J.**; Bozicevic, Field & Francis LLP, 1900 University Avenue, Suite 200, East Palo Alto, California 94303 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

(54) Title: CELL-FREE NUCLEIC ACIDS FOR THE ANALYSIS OF THE HUMAN MICROBIOME AND COMPONENTS THEREOF

(57) Abstract: Methods, devices, compositions and kits are provided for analysis of the microbiome or individual components thereof in an individual. The methods find use in a determination of infection, in analysis of the microbiome structure, in determining the immunocompetence of an individual, and the like. In some embodiments of the invention, the individual is treated with a therapeutic regimen, e.g. drugs, diet, radiation therapy, and the like.

WO 2015/070086 A1

2014346562 15 Nov 2018

CELL-FREE NUCLEIC ACIDS FOR THE ANALYSIS OF THE HUMAN MICROBIOME AND COMPONENTS THEREOF

GOVERNMENT RIGHTS

[0001] This invention was made with Government support under grant RC4AI092673 awarded by the National Institutes of Health. The Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

[0002] The human microbiome is now recognized as an important component of human health. Community level analyses have shed light on factors that shape the structure of the bacterial and viral components of the microbiome, such as age, diet, geographical location, antibiotic treatment and disease. For example, an individual microbiome can be altered by infection with a pathogenic organism, such that there is an increased prevalence of that organism systemically, or in an undesirable tissue. The microbiome can also be altered by changes in the immunocompetence of the individual.

[0003] For a variety of purposes it would be desirable to have a method for rapid identification of specific microbiome components, e.g. the presence and prevalence of commensal, mutualistic, parasitic, opportunistic and pathogenic organisms in an individual microbiome; as well as an analysis of the overall microbiome structure. The present invention provides sensitive, rapid, non-invasive methods of monitoring the microbiome composition in clinical samples.

SUMMARY OF THE INVENTION

[0004] The disclosure provides methods, devices, compositions and kits for analysis of the microbiome or individual components thereof in an individual. The methods find use in a determination of infection, in analysis of the microbiome structure, in determining the immunocompetence of an individual, and the like. In some embodiments, the disclosure provides methods of determining the presence and prevalence of microorganisms in an individual, comprising the steps of: (i) providing a sample of cell-free nucleic acids, i.e. DNA and/or RNA from an individual; (ii) performing high-throughput sequencing, for example from about 10^5, and up to about 10^9 or more reads per sample; (iii) performing bioinformatics analysis to subtract host sequences, i.e. human, cat, dog, etc. from the analysis; and (iv) determining the presence and prevalence of microbial sequences, for example by a comparison of the coverage of sequences mapping to a microbial reference sequence to coverage of the host reference sequence.

[0004a] A first aspect provides a method of detecting at least one bacterium, fungus, or parasite in a sample comprising cell-free nucleic acids from a non-microbial host, the method comprising:

(i) providing said sample comprising cell-free nucleic acids from said non-microbial host, wherein said cell-free nucleic acids comprise microbial cell-free nucleic acids;

(ii) performing high-throughput sequencing of said cell-free nucleic acids to generate sequence reads that comprise sequence reads from said microbial cell-free nucleic acids; and

(iii) detecting said at least one bacterium, fungus, or parasite at a strain or species level in said sample.

[0004b] A second aspect provides a computer readable medium comprising:

a set of instructions recorded thereon that when executed cause a computer to perform

(i) receiving high throughput data from cell-free nucleic acids detected in a sample from a non-microbial host, wherein said high throughput data comprises sequence reads; and

(ii) using sequence reads from microbial cell-free nucleic acids to detect at least one bacterium, fungus, or parasite at a strain or species level in said sample comprising cell-free nucleic acids from said non-microbial host.

[0005] The subtraction of host sequences may include the step of identifying a reference host sequence, and masking microbial sequences or microbial mimicking sequences present in the reference host genome. Similarly, determining the presence of a microbial sequence by comparison to a microbial reference sequence may include the step of identifying a reference microbial sequence, and masking host sequences or host mimicking sequences present in the reference microbial genome.

[0006] A feature of the invention is the unbiased analysis of cell-free nucleic acids from an individual. The methods of the invention generally include an unbiased amplification step, for example by performing PCR with universal primers, or by ligation of adapters to the nucleic acid and amplifying with primers specific for the adaptors. The methods of the invention are typically performed in the absence of sequence specific amplification of microbial sequences. A benefit to this approach is that analysis then includes all available microbiome sequences, however it requires bioinformatics analysis to identify sequences of interest in a complex dataset predominated by host sequences.

[0007] A further benefit of the methods of the disclosure is the ability to provide a rapid assessment of an individual microbiome, for example analysis may be completed in less than about 3 days, less than about 2 days, less than one day, e.g. less than about 24 hours, less than about 20 hours, less than about 18 hours, less than about 14 hours, less than about 12 hours, less than about 6 hours, less than about 2 hours, less than about 30 minutes, less than about 15 minutes, less than about 1 minute.

[0008] In some embodiments, the analysis of cell-free nucleic acids is used to compute a pathogenicity score, where the pathogenicity score is a numeric or alphabetic value that summarizes the overall pathogenicity of the organism for ease of interpretation, e.g. by a health practitioner. Different microbes present in the microbiome may be assigned different scores.

[0009] The analysis of the presence and prevalence of microbial sequences can be used to provide a determination on infection, of response to therapy, including anti-microbial treatment such as treatment with antibiotics, anti-viral agents, immunization, passive immunotherapy, and the like; diet; immunosuppression, and the like; of response in clinical trials, etc. The information obtained from the analysis may be used to diagnose a condition, to monitor treatment, to select or modify therapeutic regimens, and to optimize therapy. With this approach, therapeutic and/or diagnostic regimens can be individualized and tailored according to the specificity data obtained at different times over the course of treatment, thereby providing a regimen that is individually appropriate. In addition, patient samples can be obtained at any point during the treatment process; following exposure to a pathogen; course of infection, *etc.* for analysis. The analysis of presence and prevalence of microbial sequences can be provided as a report. The report may be provided to the individual, to a health care professional, etc.

[0010] In some embodiments, the cell-free nucleic acid is obtained from a biological sample selected from the group consisting of blood, serum, cerebrospinal fluid, synovial fluid, urine, and stool. The nucleic acid is extracted from the cell free portion of the sample, e.g. serum or plasma portion of blood may be used. In some embodiments, the nucleic acid is selected from the group consisting of double-stranded DNA, single-stranded DNA, single-stranded DNA hairpins, DNA/RNA hybrids, single-stranded RNA, double-stranded RNA and RNA hairpins. In some embodiments, the nucleic acid is selected from the group consisting of double-stranded DNA, single-stranded DNA and cDNA. In some embodiments, the nucleic acid is mRNA. In some embodiments, the nucleic acid is circulating cell-free DNA.

[0011] In some embodiments, the methods comprise quantitating the one or more nucleic acids to provide a determination of prevalence of a microorganism in the sample. In some embodiments, the amount of the one or more nucleic acids above a predetermined threshold value is indicative of infection or altered prevalence. In some embodiments, there are different predetermined threshold values for different microbial organisms. In some embodiments, temporal differences in the amount of the one or more nucleic acids are indicative of changes in infection, altered prevalence, response to therapy, *etc.*

[0012] Also disclosed is a computer readable mediums comprising: a set of instructions recorded thereon to cause a computer to perform the steps of: (i) receiving high throughput sequencing data from one or more nucleic acids detected in a sample of cell-free nucleic acids from a

subject; (ii) performing bioinformatics analysis to subtract host sequences, i.e. human, cat, dog, etc. from the analysis; and (iii) determining the presence and prevalence of microbial sequences, for example by a comparison of the coverage of sequences mapping to a microbial reference sequence to coverage of the host reference sequence.

[0013] Also disclosed are reagents and kits thereof for practicing one or more of the methods described herein.

[0014] In some embodiments, compositions and methods may be used for the assessment of immunocompetence of an individual, particularly an individual human, by analysis of the microbiome, for example by analysis of the virome. In some embodiments, the individual is treated with an immunosuppressive regimen, e.g. drugs, radiation therapy, and the like. In some embodiments the individual is a graft recipient treated with an immunosuppressive regimen. In some embodiments the individual has an autoimmune disease treated with an immunosuppressive regimen. In other embodiments an individual is assessed for immunocompetence in the absence of an immunosuppressive regimen.

[0015] In some embodiments a measurement from an individual is taken at two or time points, where a change in virus burden is indicative of a change in immunocompetence. The individual may be treated in accordance with the assessment of immunocompetence, e.g. where an indication of undesirable increased immunocompetence in a transplant patient is treated with increased levels of immunosuppressive agents; or where an undesirable decrease in immunocompetence is treated with therapeutic agents, e.g. anti-viral agents, etc.

[0016] Nucleic acid analysis is used to identify and quantify nonhuman cell-free nucleic acids in a sample collected from a patient. The composition of the components of the microbiome is performed as described above. The structure of the viral component of the microbiome (the virome) allows a prediction of immunocompetence. In some embodiments, the methods further comprise establishing a virome profile prior to an immunosuppressive regimen, at the initiation of an immunosuppressive regimen, or during the course of an immunosuppressive regimen, which is used as a reference to changes in the individual virome. In some embodiments the circulating cell-free DNA is annellovirus DNA.

[0017] In particular, the load of viruses of the anelloviridae family is a predictor of immune strength, which is correlated with the probability of organ transplant rejection. While other viruses may also be predictive, it is common for patients to be treated with antivirals that affect the load of such viruses.

[0018] Also disclosed are methods of diagnosing or predicting transplant status or outcome comprising the steps of: (i) providing a sample from a subject who has received a transplant from a donor; (ii) determining the presence or absence of one or more virome nucleic acids; and (iii) diagnosing or predicting transplant status or outcome based on the virome load. In some embodiments, the transplant status or outcome comprises rejection, tolerance, non-

rejection based allograft injury, transplant function, transplant survival, chronic transplant injury, or titer pharmacological immunosuppression. In some embodiments, the amount of the one or more nucleic acids above a predetermined threshold value is indicative of viral load and immunocompetence. In some embodiments, the threshold is a normative value for clinically stable post-transplantation patients with no evidence of transplant rejection or other pathologies. In some embodiments, there are different predetermined threshold values for different transplant outcomes or status. In some embodiments, temporal differences in the amount of the one or more nucleic acids are indicative of immunocompetence.

[0019] In any of the embodiments described herein, the transplant graft maybe any solid organ, bone marrow or skin transplant. In some embodiments, the transplant is selected from the group consisting of kidney transplant, heart transplant, liver transplant, pancreas transplant, lung transplant, intestine transplant and skin transplant.

[0020] Also disclosed are reagents and kits thereof for practicing one or more of the methods described herein.

[0021] All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

[0021a] It is to be understood that if any prior art publication is referred to herein, such reference does not constitute an admission that the publication forms a part of the common general knowledge in the art in Australia or any other country.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0023] Fig. 1. Study design, read statistics and phylogenetic distribution. A. Immunosuppression reduces the risk of rejection in transplantation but increases the risk of infection. B. Design of study. 656 plasma samples were collected, cell-free DNA was purified and sequenced to an average depth of 1.2 Gbp per sample. C. Number of samples collected as function of time for the different patient groups part of the study. D. Treatment protocol for patients in the study cohort, all patients are treated with maintenance immunosuppression (tacrolimus-based (TAC) for adult heart and lung transplant recipients and cyclosporine (CYC) for pediatric patients). CMV positive (donor or recipient, CMV+) transplant cases are treated with anti-CMV prophylaxis, valganciclovir (VAL). Mean level of tacrolimus measured in blood of transplant recipients treated with a TAC-based protocol (dashed line actual, solid line window average

2014346562 15 Nov 2018

filter). E. Fraction of reads that remain after filtering of lower quality and duplicate reads (mean 86%, left) and after removal of human and low complexity reads (mean 2%, right). F. Relative genomic abundance at different levels of taxonomic classification after removal of human reads (average over all samples from all organ transplant recipients (n = 656)).

[0024] Fig. 2. Relative viral genomic abundance as a function of drug dose and comparison to healthy reference. A. Mean virome composition for patients treated with the immunosuppressant tacrolimus (47 patients, 380 samples) as function of antiviral drug dose (valganciclovir) and concentration tacrolimus measured in blood. To account for the delayed effect of the virome composition on drug dose, the data on drug doses were window average filtered (window size 45 days, see fig. 1C). Herpesvirales and caudovirales dominate the virome when patients receive low doses of immunosuppressants and antiviral drugs. Conversely, anelloviridae dominate the virome when patients receive high doses of these drugs. B. Comparison of virome composition corresponding to healthy references (n = 9),

10840330_1 (GHMatters) P103007.AU 15-Nov-18

post-transplant day one samples with low drug exposure, (n=13), and samples corresponding to high drug exposure (tacrolimus ≥ 9 ng/ml, valganciclovir ≥ 600 mg, n = 68). The virome structure for day one samples (1) and the virome structure measured for a set of healthy individuals (H) are distinct from the anellovirus-dominated distribution measured for samples corresponding to high drug doses (D). The piecharts show the mean fractions, p-values in boxplot based on the Mann-Whitney test. C. Bray-Curtis beta diversity for all samples, among patients with the same transplant type (heart or lung), within subjects, for patients treated with a similar drug dosage (tacrolimus level ± 0.5 ng/ml, valganciclovir ± 50 mg), and for samples collected from the same subjects within a one-month timespan.

[0025] Figure 3. Temporal dynamics of the microbiome composition post-transplant. A. Relative abundance of dsDNA and ssDNA viruses for different time periods (average for all samples). The relative abundance of ssDNA viruses increases rapidly after the onset of the post-transplant drug therapy. After 6 months, the opposite trend is observed. B. Viral genome abundance at the family and order level of taxonomic classification for different time periods. The fraction of anelloviridae expands rapidly in the first several months post-transplant. The fraction of herpesvirales, caudovirales and adenoviridae decreases in that same time period. After 6 months, the opposite trends are observed. C. Time-variation in the relative abundance of bacterial phyla. Compared to the viral abundance, the representation of different bacterial phyla is relatively unchanged over the observed post-transplant period. D. Shannon entropy as a measure of the within-sample alpha- diversity for bacterial and viral genera as function of time (data grouped per one month time period).

[0026] Figure 4. Virome composition and total viral burden in the absence and presence of antiviral prophylaxis. A. Absolute viral load as a function of time, measured as viral genome copies per human genome copies detected by sequencing. Box plots are shown for different time periods with centers of the time periods marked on the x-axis. For all patient classes, the total viral load increases in the first weeks post-transplant (black line is sigmoid fit, change in load 7.4 ± 3). B. Viral load and composition for CMV+ cases that are treated with both immunosuppressants and antiviral drugs (78 patients, 543 samples). C. Viral load and composition for CMV -/- cases, only treated with immunosuppressants (12 patients, 75 samples).

[0027] Figure 5. Lower anellovirus burden in patients that suffer from graft rejection. A. Time dependence of the anellovirus load in the subgroup of patients that suffer from a severe rejection episode (biopsy grade ≥ 2R/3A, red data, 20 patients, 177 time points) and in the subgroup of patients that do not suffer from a severe-rejection-free post-transplant course (blue data, 40 patients, 285 time points). Box plots are shown for different time periods with centers of the time periods marked on the x-axis. Solid lines are cubic splines (smoothing parameter 0.75). The inset shows a cartoon of the expected opposite association of the

incidence of rejection and infection with immunocompetence. B. Anellovirus load relative to the average load measured for all samples at the same time point. The time-normalized load for non-rejecting patients (N = 208) is compared to the load measured for patients suffering from a mild rejection event (biopsy grade 1R, N = 102) and patients suffering from a severe rejection episode (biopsy grade ≥ 2R/3A, N = 22). The p-values reflect the probability that the median viral load is higher for the subgroups at greater risk of rejection. The p-values are calculated by random sampling of the population with a greater amount of measurement points. N-fold random sampling, $p = \text{sum}(\text{median}(A_{rej}) > \text{median}(A_{non\text{-}rej}))/N)$, where $N=10^4$ and A_{rej} and $A_{non\text{-}rej}$ are the relative viral loads for the populations at greater and lesser risk of rejection and non-rejecting respectively. C Test of the performance of the relative anellovirus load in classifying patients as non-rejecting vs. severely rejecting, receiver-operating characteristic curve, area under the curve = 0.72.

[0028] Figure 6. Genome Sizes and Hit Statistics, qPCR Assay and Influence of Read Length on Measured Relative Abundance of Species at Different Levels of Taxonomic Classification (A) Distribution of genome sizes in the reference database with, 1401 viral genomes, 32 fungal genomes and 1980 bacterial genomes. (B) Distribution of unique blast hits per million unique molecules sequenced (average number of hits specified in the x axis label). C. Distribution of genome equivalents (infectious agents/diploid human) for viruses, bacteria and fungi (average number of genome equivalents specified in the x axis label). (D) Comparison of sequencing hits found per million of total reads sequenced to the number of viral copies detected using qPCR. For the qPCR assays, DNA was purified from 1 ml of plasma and eluted in a 100 μl volume. (E) Measurements of CMV and parvovirus burdens for selected cases. The highest loads of CMV virus (genome equivalents, viral/human diploid, G.E.) measured for all samples corresponded to two cases of clinically diagnosed disseminated CMV infection (a and b, shaded area denotes time-window of clinical diagnosis, * denotes time of death). (c) shows a time trace of a pediatric patient that suffered from CMV viremia. Parvovirus was detected in one pediatric heart transplant patient immediately post-transplant (d).* (F) Influence of read length on measured relative abundance of species at different levels of taxonomic classification (n = 52). Spearman sample-to-sample correlation, r, and p value, p, (two-sample Mann-Whitney U test) for the abundance of the most abundant node extracted from the 50 and 100 bp data sets: r = 0.80, p = 0.8 (a), r = 0.86, p = 0.4 (b), r = 0.92, p = 0.6 (c), r = 0.84, p = 0.5 (d), r = 0.7, p = 0.28 (e), r = 0.99, p = 1 (f).

[0029] Figure 7. Average Drug Doses and Measured Levels for Adult Heart and Lung Transplant Patients after Transplantation and Influence of Drug Dosage on Virome composition. (A–C) Average dose of valganciclovir and prednisone (A and C) administered and measured level of tacrolimus in blood (B) for the adult heart and lung transplant patients part of this study. (D) Compared to the viral component, the composition of the bacterial

component of the microbiome is relatively insensitive to antivirals and immunosuppressants. (E) Virome composition as function of dose of anti-CMV drug (valganciclovir) and immunosuppressant (prednisone).

[0030] Figure 8. Temporal Dynamics of the Bacterial Component of the Microbiome Posttransplant, (A) Relative abundance of bacterial phyla as function of time. (B) Relative abundance of bacterial genera as function of time.

[0031] Figure 9. Virome Composition and Total Viral Burden for Different Patient Classes, (A and B) Viral load and composition for CMV positive adult heart (A), adult lung (B), and pediatric heart (C) transplant recipients, treated with both immunosuppressants and antivirals.

[0032] Figure 10. CMV-infection induced allograft injury. A. Correlation between clinical report of CMV (Human herpes virus 5, HHV-5) infection from specific body fluids (BAL and serum) with donor organ cfdDNA signal matched to clinical test date (P-values; Mann-Whitney U test). B. P-values for the correlation between clinical diagnosis of infection and cell-free DNA level (dashed line indicates the bonferonni-corrected significance threshold) for infections with greater than one clinical positive test result. C. An ROC curve that tests the performance of CMV derived cell-free DNA level in CMV-positive and CMV-negative patients (AUC=0.91).

[0033] Figure 11. Monitoring the infectome. A. Clinical testing frequency compared to the incidence of viral infections detected in sequencing. B. Time-series data for patients who tested positive (red arrows) for specific infections relative to those who were un-tested. (1) Adenovirus signal in L78 with clinical positives highlighted relative to untested patient (L34). (2) Polyomavirus signal in L69 with one positive test relative to sustained signal in un-tested patient (L57). (3) Three herpesvirus infections (HHV-4, 5, and 8) in L58 with both positive (red) and negative (black) tests for CMV (HHV-5) highlighted. (4) Microsporidia signal in I6, with four positive tests shown, relative to the signal observed in L78, which had symptoms of microsporidiosis, but was un-tested. Data are logged genome equivalents relative to human where zero values were replaced with the detection limit of the assay (the number of genome equivalents consistent with a single sequence read assigned to the target genome).

DETAILED DESCRIPTION OF THE INVENTION

[0034] Reference will now be made in detail to particularly preferred embodiments of the invention. Examples of the preferred embodiments are illustrated in the following Examples section.

[0035] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. All patents and publications referred to herein are incorporated by reference in their entirety.

[0036] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0037] Certain ranges are presented herein with numerical values being preceded by the term "about." The term "about" is used herein to provide literal support for the exact number that it precedes, as well as a number that is near to or approximately the number that the term precedes. In determining whether a number is near to or approximately a specifically recited number, the near or approximating unrecited number may be a number which, in the context in which it is presented, provides the substantial equivalent of the specifically recited number.

[0037a] In the claims which follow and in the description of the invention, except where the context requires otherwise due to express language or necessary implication, the word “comprise” or variations such as “comprises” or “comprising” is used in an inclusive sense, i.e. to specify the presence of the stated features but not to preclude the presence or addition of further features in various embodiments of the invention.

[0038] The practice of the present invention employs, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

[0039] The invention provides methods, devices, compositions and kits for analysis of the microbiome or individual components thereof in an individual. The methods find use in a determination of infection, in analysis of the microbiome structure, in determining the immunocompetence of an individual, and the like. In some embodiments, the invention provides methods of determining whether a patient or subject is displaying immunocompetence. The term “individual”, “patient” or “subject” as used herein includes humans as well as other mammals.

Definitions.

[0040] As used herein the term “diagnose” or “diagnosis” of a status or outcome includes predicting or diagnosing the status or outcome, determining predisposition to a status or outcome, monitoring treatment of patient, diagnosing a therapeutic response of a patient, and prognosis of status or outcome, progression, and response to particular treatment.

[0041] *Microbiota*. As used herein, the term microbiota refers to the set of microorganisms present within an individual, usually an individual mammal and more usually a human individual. The microbiota may include pathogenic species; species that constitute the normal flora of one tissue, e.g. skin, oral cavity, etc., but are undesirable in other tissues, e.g. blood, lungs, etc.; commensal organisms found in the absence of disease; etc. A subset of the microbiome is the virome, which comprises the viral components of the microbiome.

[0042] The term “microbiome component” as used herein refers to an individual strains or species, The component may be a viral component, a bacterial component, a fungal component, *etc.*

[0043] In a healthy animal, while the internal tissues, e.g. brain, muscle, etc., are normally presumed to be relatively free of bacterial species, the surface tissues, i.e., skin and mucous membranes, are constantly in contact with environmental organisms and become readily colonized by various microbial species. The mixture of organisms known or presumed to be found in humans at any anatomical site is referred to as the "indigenous microbiota", including various components of the indigenous microbiota. In addition to the indigenous microorganisms are various transient components, such as pathogenic or opportunistic infections. Reference sequences of organisms described below are publicly available and known in the for, example at the Genbank database.

[0044] The intestinal microbiota of humans is dominated by species found within two bacterial phyla: members of the Bacteroidetes and Firmicutes make up >90% of the bacterial population. Actinobacteria (*e.g.*, members of the Bifidobacterium genus) and Proteobacteria among several other phyla are less prominently represented. Common species of interest include prominent or less abundant members of this community, and may comprise, without limitation, *Bacteroides thetaiotaomicron; Bacteroides caccae; Bacteroides fragilis*; *Bacteroides melaninogenicus*; *Bacteroides oralis*; *Bacteroides uniformis*; *Lactobacillus*; *Clostridium perfringens*; *Clostridium septicum*; *Clostridium tetani*; *Bifidobacterium bifidum*; *Staphylococcus aureus*; *Enterococcus faecalis*; *Escherichia coli*; *Salmonella enteritidis*; *Klebsiella sp.*; *Enterobacter sp.*; *Proteus mirabilis*; *Pseudomonas aeruginosa*; *Peptostreptococcus sp.*; *Peptococcus sp., Faecalibacterium sp,; Roseburia sp.; Ruminococcus sp.; Dorea sp.; Alistipes sp.; etc*.

[0045] In the skin microbiome most bacteria fall into four different phyla: Actinobacteria, Firmicutes, Bacteroidetes and Proteobacteria. Microorganisms that are generally regarded as skin colonizers include coryneforms of the phylum Actinobacteria (the genera *Corynebacterium, Propionibacterium,* such as *Propionibacterium acnes;* and *Brevibacterium*), the genus *Micrococcus* and *Staphylococcus* spp. The most commonly isolated fungal species are *Malassezia* spp., which are especially prevalent in sebaceous areas. The *Demodex* mites (such as *Demodex folliculorum* and *Demodex brevis*) may also be present. Other types of

fungi that are thought to grow on the skin, include *Debaryomyces* and *Cryptococcus* spp. As non-commensals, burn wounds commonly become infected with *S. pyogenes*, *Enterococcus* spp. or *Pseudomonas aeruginosa*, and can also become infected with fungi and/or viruses. *S. epidermidis* is a very common skin commensal, but it is also the most frequent cause of hospital- acquired infection on in-dwelling medical devices such as catheters or heart valves. For a review, see Nat Rev Microbiol. (2011) Apr;9(4):244-53.

[0046] Pathogenic species may be bacteria, virus, protozoan parasites, fungal species, etc.. Bacteria include *Brucella sp.*, *Treponema sp., Mycobacterium sp., Listeria sp., Legionella sp., Helicobacter sp, Streptococcus sp, Neisseria sp, Clostridium sp, Staphylococcus sp.* or *Bacillus sp.*; including without limitation *Treponema pallidum, Mycobacterium tuberculosis, Mycobacterium leprae, Listeria monocytogenes, Legionella pneumophila, Helicobacter pylori, Streptococcus pneumoniae, Neisseria meningitis, Clostridium novyi, Clostridium botulinum, Staphylococcus aureus*, *Bacillus anthracis*, *etc*.

[0047] Parasite pathogens include *Trichomonas, Toxoplasma, Giardia, Cryptosporidium, Plasmodium, Leishmania, Trypanosoma, Entamoeba, Schistosoma, Filariae, Ascaria*, *Fasciola*; including without limitation *Trichomonas vaginalis, Toxoplasma gondii, Giardia intestinalis, Cryptosporidium parva, Plasmodium falciparum, Trypanosoma cruzi, Entamoeba histolytica, Giardia lamblia, Fasciola hepatica*, *etc*.

[0048] Viruses that infect humans include, for example, Adeno-associated virus; Aichi virus; Australian bat lyssavirus; BK polyomavirus; Banna virus; Barmah forest virus; Bunyamwera virus; Bunyavirus La Crosse; Bunyavirus snowshoe hare; Cercopithecine herpesvirus; Chandipura virus; Chikungunya virus; Cosavirus A; Cowpox virus; Coxsackievirus; Crimean-Congo hemorrhagic fever virus; Dengue virus; Dhori virus; Dugbe virus; Duvenhage virus; Eastern equine encephalitis virus; Ebolavirus; Echovirus; Encephalomyocarditis virus; Epstein-Barr virus; European bat lyssavirus; GB virus C/Hepatitis G virus ; Hantaan virus; Hendra virus; Hepatitis A virus; Hepatitis B virus ; Hepatitis C virus; Hepatitis E virus; Hepatitis delta virus; Horsepox virus; Human adenovirus; Human astrovirus; Human coronavirus; Human cytomegalovirus; Human enterovirus 68, 70; Human herpesvirus 1; Human herpesvirus 2; Human herpesvirus 6; Human herpesvirus 7; Human herpesvirus 8; Human immunodeficiency virus; Human papillomavirus 1; Human papillomavirus 2; Human papillomavirus 16,18; Human parainfluenza; Human parvovirus B19; Human respiratory syncytial virus; Human rhinovirus; Human SARS coronavirus; Human spumaretrovirus; Human T-lymphotropic virus; Human torovirus; Influenza A virus; Influenza B virus; Influenza C virus; Isfahan virus; JC polyomavirus; Japanese encephalitis virus; Junin arenavirus; KI Polyomavirus; Kunjin virus; Lagos bat virus; Lake Victoria marburgvirus; Langat virus; Lassa

virus; Lordsdale virus; Louping ill virus; Lymphocytic choriomeningitis virus; Machupo virus; Mayaro virus; MERS coronavirus; Measles virus; Mengo encephalomyocarditis virus; Merkel cell polyomavirus; Mokola virus; Molluscum contagiosum virus; Monkeypox virus; Mumps virus; Murray valley encephalitis virus; New York virus; Nipah virus; Norwalk virus; O'nyong-nyong virus; Orf virus; Oropouche virus; Pichinde virus; Poliovirus; Punta toro phlebovirus; Puumala virus; Rabies virus; Rift valley fever virus; Rosavirus A; Ross river virus; Rotavirus A; Rotavirus B; Rotavirus C; Rubella virus; Sagiyama virus; Salivirus A; Sandfly fever sicilian virus; Sapporo virus; Semliki forest virus; Seoul virus; Simian foamy virus; Simian virus 5; Sindbis virus; Southampton virus; St. louis encephalitis virus; Tick-borne powassan virus; Torque teno virus; Toscana virus; Uukuniemi virus; Vaccinia virus; Varicella-zoster virus; Variola virus; Venezuelan equine encephalitis virus; Vesicular stomatitis virus; Western equine encephalitis virus; WU polyomavirus; West Nile virus; Yaba monkey tumor virus; Yaba-like disease virus; Yellow fever virus; Zika virus;

[0049] *Anelloviridae*. The *Anelloviridae* family consists of non-enveloped, circular, single-stranded DNA viruses. Three genera of anellovirus are known to infect humans, named TTV, TTMDV, and TTMV.

[0050] Torque Teno Virus (TTV) is a non-enveloped, single-stranded DNA virus with a circular, negative-sense genome. A smaller virus, which was subsequently named Torque Teno-like Mini Virus (TTMV) has also been characterized, and a third virus with a genomic size in between that of TTV and TTMV was discovered and subsequently named Torque Teno-like Midi Virus (TTMDV). Recent changes in nomenclature have classified the three anelloviruses able to infect humans into Alphatorquevirus (TTV), Betatorquevirus (TTMV), and Gammatorquevirus (TTMDV) Genera of the Anelloviridae family of viruses. To date anelloviruses are still considered "orphan" viruses waiting to be linked to human disease.

[0051] The human anelloviruses differ in genome size ranging from 3.8–3.9 kb for TTV, 3.2 kb for TTMDV, and 2.8–2.9 kb for TTMV. A characteristic feature of anelloviruses is the extreme diversity found both within and between anellovirus species; they can exhibit as much as 33%–50% divergence at the nucleotide level. Despite the nucleotide sequence diversity, anelloviruses share conserved genomic organization, transcriptional profiles, a non-coding GC rich region, and sequence motifs resulting in shared virion structure and gene functions.

[0052] Anellovirus infections are highly prevalent in the general population. A study in Japan found that 75–100% of patients tested were infected with at least one of the three human anelloviruses, and many were infected with multiple species. Anelloviruses can infect young children, with the earliest documented infections occurring within the first months of life. These viruses have been found in nearly every body site, fluid, and tissue tested including blood plasma, serum, peripheral blood mononuclear cells (PBMCs), nasopharyngeal aspirates,

bone marrow, saliva, breast milk, feces, as well as various tissues including thyroid gland, lymph node, lung, liver, spleen, pancreas, and kidney. The replication dynamics of anelloviruses are virtually unknown because of the inability to propagate these viruses in culture. Positive-strand TTV DNA, indicative of local viral replication, has been described in hepatocytes, bone marrow cells, and circulating PBMCs.

[0053] Anelloviruses are spread primarily through fecal-oral transmission, although mother-child and respiratory tract transmissions have also been reported. There are conflicting reports regarding the presence of TTV in cord blood specimens.

[0054] Reference sequences for anellovirus may be accessed at Genbank, e.g. Torque teno mini virus 1, Accession: NC_014097.1; Torque teno mini virus 6, Accession:NC_014095.1; Torque teno midi virus 2, Accession:NC_014093.1; Torque teno midi virus 1, Accession:NC_009225.1; Torque teno virus 3, Accession:NC_014081.1; Torque teno virus 19, Accession:NC_014078.1; Torque teno mini virus 8, Accession:NC_014068.1.

[0055] The term "antibiotic" as used herein includes all commonly used bacteristatic and bactericidal antibiotics, usually those administered orally. Antibiotics include aminoglycosides, such as amikacin, gentamicin, kanamycin, neomycin, streptomycin, and tobramycin; cephalosporins, such as cefamandole, cefazolin, cephalexin, cephaloglycin, cephaloridine, cephalothin, cephapirin, and cephradine; macrolides, such as erythromycin and troleandomycin; penicillins, such as penicillin G, amoxicillin, ampicillin, carbenicillin, cloxacillin, dicloxacillin, methicillin, nafcillin, oxacillin, phenethicillin, and ticarcillin; polypeptide antibiotics, such as bacitracin, colistimethate, colistin, polymyxin B; tetracyclines, such as chlortetracycline, demeclocycline, doxycycline, methacycline, minocycline, tetracycline, and oxytetracycline; and miscellaneous antibiotics such as chloramphenicol, clindamycin, cycloserine, lincomycin, rifampin, spectinomycin, vancomycin, and viomycin. Additional antibiotics'are described in "Remington's Pharmaceutical Sciences," 16th Ed., (Mack Pub. Co., 1980), pp. 1121-1178.

[0056] *Antiviral agents*. Individuals may receive antiviral therapy, which will alter the viral load for those viruses affected by the therapy. Examples of viral infections thus treated include HIV, Bowenoid Papulosis, Chickenpox, Childhood HIV Disease, Human Cowpox, Hepatitis C, Dengue, Enteroviral, Epidermodysplasia Verruciformis, Erythema Infectiosum (Fifth Disease), Giant Condylomata Acuminata of Buschke and Lowenstein, Hand-Foot-and-Mouth Disease, Herpes Simplex, Herpes Virus 6, Herpes Zoster, Kaposi Varicelliform Eruption, Rubeola Measles, Milker's Nodules, Molluscum Contagiosum, Monkeypox, Orf, Roseola Infantum, Rubella, Smallpox, Viral Hemorrhagic Fevers, Genital Warts, and Nongenital Warts.

[0057] Antiviral agents include azidouridine, anasmycin, amantadine, bromovinyldeoxusidine, chlorovinyldeoxusidine, cytarbine, didanosine, deoxynojirimycin, dideoxycitidine,

dideoxyinosine, dideoxynucleoside, desciclovir, deoxyacyclovir, edoxuidine, enviroxime, fiacitabine, foscamet, fialuridine, fluorothymidine, floxuridine, hypericin, interferon, interleukin, isethionate, nevirapine, pentamidine, ribavirin, rimantadine, stavirdine, sargramostin, suramin, trichosanthin, tribromothymidine, trichlorothymidine, vidarabine, zidoviridine, zalcitabine and 3-azido-3-deoxythymidine, and analogs, derivatives, pharmaceutically acceptable salts, esters, prodrugs, codrugs, and protected forms thereof.

[0058] Immunosuppression or immunosuppressive regimen, as used herein, refers to the treatment of an individual, for example a graft recipient with agents to diminish the immune responses of the host immune system against autoantigens or graft. Exemplary immunosuppression regimens are described in more detail herein.

[0059] Primary immunosuppressive agents include calcineurin inhibitors, which combine with binding proteins to inhibit calcineurin activity, and which include, for example, tacrolimus, cyclosporine A, etc. Levels of both cyclosporine and tacrolimus must be carefully monitored. Initially, levels can be kept in the range of 10-20 ng/mL, but, after 3 months, levels may be kept lower (5-10 ng/mL) to reduce the risk of nephrotoxicity.

[0060] Adjuvant agents are usually combined with a calcineurin inhibitor and include steroids, azathioprine, mycophenolate mofetil, and sirolimus. Protocols of interest include a calcineurin inhibitor with mycophenolate mofetil. The use of adjuvant agents allows clinicians to achieve adequate immunosuppression while decreasing the dose and toxicity of individual agents. Mycophenolate mofetil in kidney transplant recipients has assumed an important role in immunosuppression after several clinical trials have shown a markedly decreased prevalence of acute cellular rejection compared with azathioprine and a reduction in 1-year treatment failures.

[0061] Antibody-based therapy may use monoclonal (eg, muromonab-CD3) or polyclonal antibodies or anti-CD25 antibodies (eg, basiliximab, daclizumab) and is administered in the early post-transplant period (up to 8 wk). Antibody-based therapy allows for avoidance or dose reduction of calcineurin inhibitors, possibly reducing the risk of nephrotoxicity. The adverse effect profile of the polyclonal and monoclonal antibodies limits their use in some patients.

[0062] The term “nucleic acid” used herein refers to a polynucleotide comprising two or more nucleotides. It may be DNA or RNA. A “variant” nucleic acid is a polynucleotide having a nucleotide sequence identical to that of its original nucleic acid except having at least one nucleotide modified, for example, deleted, inserted, or replaced, respectively. The variant may have a nucleotide sequence at least about 80%, 90%, 95%, or 99%, identity to the nucleotide sequence of the original nucleic acid.

[0063] Circulating, or cell-free, DNA was first detected in human blood plasma in 1948. (Mandel, P. Metais, P., C R Acad. Sci. Paris, 142, 241-243 (1948)) Since then, its connection

to disease has been established in several areas. (Tong, Y.K. Lo, Y.M., Clin Chim Acta, 363, 187-196 (2006)) Studies reveal that much of the circulating nucleic acids in blood arise from necrotic or apoptotic cells (Giacona, M.B., et al., Pancreas, 17, 89-97 (1998)) and greatly elevated levels of nucleic acids from apoptosis is observed in diseases such as cancer. (Giacona, M.B., et al., Pancreas, 17, 89-97 (1998); Fournie, G.J., et al., Cancer Lett, 91, 221-227 (1995)). Particularly for cancer, where the circulating DNA bears hallmark signs of the disease including mutations in oncogenes, microsatellite alterations, and, for certain cancers, viral genomic sequences, DNA or RNA in plasma has become increasingly studied as a potential biomarker for disease. For example, Diehl et al recently demonstrated that a quantitative assay for low levels of circulating tumor DNA in total circulating DNA could serve as a better marker for detecting the relapse of colorectal cancer compared with carcinoembryonic antigen, the standard biomarker used clinically. (Diehl, F., et al., Proc Natl Acad Sci, 102, 16368-16373 (2005); Diehl, F., et al., Nat Med, 14, 985-990 (2008)). Maheswaran et al reported the use of genotyping of circulating cells in plasma to detect activating mutations in epidermal growth factor receptors in lung cancer patients that would affect drug treatment. (Maheswaran, S., et al., N Engl J Med, 359, 366-377 (2008)) These results collectively establish circulating DNA free in plasma as a useful species in cancer detection and treatment. Circulating DNA has also been useful in healthy patients for fetal diagnostics, with fetal DNA circulating in maternal blood serving as a marker for gender, rhesus D status, fetal aneuploidy, and sex-linked disorders. Fan et al recently demonstrated a strategy for detecting fetal aneuploidy by shotgun sequencing of cell-free DNA taken from a maternal blood sample, a methodology that can replace more invasive and risky techniques such as amniocentesis or chorionic villus sampling. (Fan, H.C., Blumenfeld, Y.J., Chitkara, U., Hudgins, L., Quake, S.R., Proc Natl Acad Sci, 105, 16266-16271 (2008)).

[0064] The term “derived from” used herein refers to an origin or source, and may include naturally occurring, recombinant, unpurified or purified molecules. A nucleic acid derived from an original nucleic acid may comprise the original nucleic acid, in part or in whole, and may be a fragment or variant of the original nucleic acid. A nucleic acid derived from a biological sample may be purified from that sample.

[0065] A “target nucleic acid” in the method according to the present invention is a nucleic acid, DNA or RNA, to be detected. A target nucleic acid derived from an organism is a polynucleotide that has a sequence derived from that of the organism and is specific to the organism. A target nucleic acid derived from a pathogen refers to a polynucleotide having a polynucleotide sequence derived from that specific the pathogen.

[0066] In some embodiments, less than 1 pg, 5pg, 10 pg, 20 pg, 30 pg, 40 pg, 50 pg, 100 pg, 200 pg, 500 pg, 1 ng , 5ng, 10 ng, 20 ng, 30 ng, 40 ng, 50 ng, 100 ng, 200 ng, 500 ng, 1 μg, 5 μg, 10 μg, 20 μg, 30 μg, 40 μg, 50 μg, 100 μg, 200 μg, 500 μg or 1 mg of nucleic acids are

obtained from the sample for analysis. In some cases, about 1-5 pg, 5-10 pg, 10- 100 pg, 100 pg- 1 ng, 1 -5 ng, 5 - 10 ng, 10 - 100 ng, 100 ng- 1μg of nucleic acids are obtained from the sample for analysis.

[0067] In some embodiments, the methods described herein are used to detect and/or quantify nucleic acid sequences that correspond to a microbe of interest, or a microbiome of organisms. The methods described herein can analyze at least 1; 2; 3; 4; 5; 10, 20; 50; 100; 200; 500; 1,000; 2,000; 5,000; 10,000, 20,000; 50,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 10^6, $5x10^6$, 10^7, $5x10^7$, 10^8, $5x10^8$, 10^9 or more sequence reads.

[0068] In some embodiments, the methods described herein are used to detect and/or quantify gene expression, e.g. by determining the presence of mRNA from a microorganism in relation to DNA from that microorganism. In some embodiments, the methods described herein provide high discriminative and quantitative analysis of multiple genes. The methods described herein can discriminate and quantitate the expression of at least 1, 2, 3, 4, 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 50,000, 100,000, or more different target nucleic acids.

[0069] A sample containing cell-free nucleic acids is obtained from a subject. Such subject can be a human, a domesticated animal such as a cow, chicken, pig, horse, rabbit, dog, cat, goat, *etc*. In some embodiments, the cells used in the present invention are taken from a patient. Samples include, for example, the acellular fraction of whole blood, sweat, tears, saliva, ear flow, sputum, lymph, bone marrow suspension, lymph, urine, saliva, semen, vaginal flow, cerebrospinal fluid, brain fluid, ascites, milk, secretions of the respiratory, intestinal or genitourinary tracts fluid, a lavage of a tissue or organ (e.g. lung) or tissue which has been removed from organs, such as breast, lung, intestine, skin, cervix, prostate, pancreas, heart, liver and stomach. Such samples can be separated by centrifugation, elutriation, density gradient separation, apheresis, affinity selection, panning, FACS, centrifugation with Hypaque, etc. Once a sample is obtained, it can be used directly, frozen, or maintained in appropriate culture medium for short periods of time.

[0070] To obtain a blood sample, any technique known in the art may be used, e.g. a syringe or other vacuum suction device. A blood sample can be optionally pre-treated or processed prior to use. A sample, such as a blood sample, may be analyzed under any of the methods and systems herein within 4 weeks, 2 weeks, 1 week, 6 days, 5 days, 4 days, 3 days, 2 days, 1 day, 12 hrs, 6 hrs, 3 hrs, 2 hrs, or 1 hr from the time the sample is obtained, or longer if frozen. When obtaining a sample from a subject (e.g., blood sample), the amount can vary depending upon subject size and the condition being screened. In some embodiments, at least 10 ml, 5 ml., 1 ml, 0.5 ml, 250, 200, 150, 100, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1

μL of a sample is obtained. In some embodiments, 1-50, 2-40, 3-30, or 4-20 μL of sample is obtained. In some embodiments, more than 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 μL of a sample is obtained.

[0071] The cell-free fraction is preferably blood serum or blood plasma. The term “cell-free fraction” of a biological sample used herein refers to a fraction of the biological sample that is substantially free of cells. The term “substantially free of cells” used herein refers to a preparation from the biological sample comprising fewer than about 20,000 cells per ml, preferably fewer than about 2,000 cells per ml, more preferably fewer than about 200 cells per ml, most preferably fewer than about 20 cells per ml. In contrast to certain prior art methods, genomic DNA is not excluded from the acellular sample, and typically comprises from about 50% to about 90% of the nucleic acids that are present in the sample.

[0072] The method of the present invention may further comprise preparing a cell-free fraction from a biological sample. The cell-free fraction may be prepared using conventional techniques known in the art. For example, a cell-free fraction of a blood sample may be obtained by centrifuging the blood sample for about 3-30 min, preferably about 3-15 min, more preferably about 3-10 min, most preferably about 3-5 min, at a low speed of about 200-20,000 g, preferably about 200-10,000 g, more preferably about 200-5,000 g, most preferably about 350-4,500 g. The biological sample may be obtained by ultrafiltration in order to separate the cells and their fragments from a cell-free fraction comprising soluble DNA or RNA. Conventionally, ultrafiltration is carried out using a 0.22 μm membrane filter.

[0073] The method of the present invention may further comprise concentrating (or enriching) the target nucleic acid in the cell-free fraction of the biological sample. The target nucleic acid may be concentrated using conventional techniques known in the art, such as solid phase absorption in the presence of a high salt concentration, organic extraction by phenol-chloroform followed by precipitation with ethanol or isopropyl alcohol, or direct precipitation in the presence of a high salt concentration or 70-80% ethanol or isopropyl alcohol. The concentrated target nucleic acid may be at least about 2, 5, 10, 20 or 100 times more concentrated than that in the cell-free fraction. The target nucleic acid, whether or not concentrated, may be used for amplification according to the method of the present invention.

[0074] In some embodiments the invention provides methods for diagnosis or prediction of transplant rejection. The term “transplant rejection” encompasses both acute and chronic transplant rejection. "Acute rejection or AR" is the rejection by the immune system of a tissue transplant recipient when the transplanted tissue is immunologically foreign. Acute rejection is characterized by infiltration of the transplanted tissue by immune cells of the recipient, which carry out their effector function and destroy the transplanted tissue. The onset of acute rejection is rapid and generally occurs in humans within a few weeks after transplant surgery.

Generally, acute rejection can be inhibited or suppressed with immunosuppressive drugs such as rapamycin, cyclosporin A, anti-CD40L monoclonal antibody and the like.

[0075] "Chronic transplant rejection or CR" generally occurs in humans within several months to years after engraftment, even in the presence of successful immunosuppression of acute rejection. Fibrosis is a common factor in chronic rejection of all types of organ transplants. Chronic rejection can typically be described by a range of specific disorders that are characteristic of the particular organ. For example, in lung transplants, such disorders include fibroproliferative destruction of the airway (bronchiolitis obliterans); in heart transplants or transplants of cardiac tissue, such as valve replacements, such disorders include fibrotic atherosclerosis; in kidney transplants, such disorders include, obstructive nephropathy, nephrosclerorsis, tubulointerstitial nephropathy; and in liver transplants, such disorders include disappearing bile duct syndrome. Chronic rejection can also be characterized by ischemic insult, denervation of the transplanted tissue, hyperlipidemia and hypertension associated with immunosuppressive drugs.

[0076] In some embodiments, the invention further includes methods for determining the effectiveness of an immunosuppressive regimen for a subject who has received a transplant, e.g., an allograft.

[0077] Certain embodiments of the invention provide methods of predicting transplant survival in a subject that has received a transplant. The invention provides methods of diagnosing or predicting whether a transplant in a transplant patient or subject will survive or be lost. In certain embodiments, the invention provides methods of diagnosing or predicting the presence of long-term graft survival. By "long-term" graft survival is meant graft survival for at least about 5 years beyond current sampling, despite the occurrence of one or more prior episodes of acute rejection. In certain embodiments, transplant survival is determined for patients in which at least one episode of acute rejection has occurred. As such, these embodiments provide methods of determining or predicting transplant survival following acute rejection. Transplant survival is determined or predicted in certain embodiments in the context of transplant therapy, e.g., immunosuppressive therapy, where immunosuppressive therapies are known in the art. In yet other embodiments, methods of determining the class and/or severity of acute rejection (and not just the presence thereof) are provided.

[0078] As is known in the transplantation field, the transplant organ, tissue or cell(s) may be allogeneic or xenogeneic, such that the grafts may be allografts or xenografts. A feature of the graft tolerant phenotype detected or identified by the subject methods is that it is a phenotype which occurs without immunosuppressive therapy, i.e., it is present in a host that is not undergoing immunosuppressive therapy such that immunosuppressive agents are not being administered to the host. The transplant graft maybe any solid organ and skin transplant. Examples of organ transplants that can be analyzed by the methods described

herein include but are not limited to kidney transplant, pancreas transplant, liver transplant, heart transplant, lung transplant, intestine transplant, pancreas after kidney transplant, and simultaneous pancreas-kidney transplant.

Microbiome Detection and Analysis

[0079] The methods of the invention involve high throughput sequencing of a cell-free nucleic acid sample from an individual, followed by bioinformatics analysis to determine the presence and prevalence of microbial sequences, which sequences may be from indigenous organisms, e.g. the normal microbiome of gut, skin, etc., or may be non-indigenous, e.g. opportunistic, pathogenic, etc. infections. Analysis may be performed for the complete microbiome, or for components there, for example the virome, bacterial microbiome, fungal microbiome, protozoan microbiome, etc. Examples of nucleic acids include, but are not limited to double-stranded DNA, single-stranded DNA, single-stranded DNA hairpins, DNA/RNA hybrids, RNA (e.g. mRNA or miRNA) and RNA hairpins. In some embodiments, the nucleic acid is DNA. In some embodiments, the nucleic acid is RNA. For instance, cell-free RNA and DNA are present in human plasma.

[0080] Genotyping microbiome nucleic acids, and/or detection, identification and/or quantitation of the microbiome-specific nucleic acids generally include an initial step of amplification of the sample, although there may be instances where sufficient cell free nucleic acids are available and can be directly sequenced. When the nucleic acid is RNA, the amplification step may be preceded by a reverse transcriptase reaction to convert the RNA into DNA. Preferably the amplification is unbiased, that is the primers for amplification are universal primers, or adaptors are ligated to the nucleic acids being analyzed, and amplification primers are specific for the adaptors. Examples of PCR techniques include, but are not limited, to hot start PCR, nested PCR, in situ polonony PCR, in situ rolling circle amplification (RCA), bridge PCR, picotiter PCR and emulsion PCR. Other suitable amplification methods include the ligase chain reaction (LCR), transcription amplification, self-sustained sequence replication, selective amplification of target polynucleotide sequences, consensus sequence primed polymerase chain reaction (CP-PCR), arbitrarily primed polymerase chain reaction (AP-PCR), degenerate oligonucleotide-primed PCR (DOP-PCR) and nucleic acid based sequence amplification (NABSA). Other amplification methods that may be used to amplify specific polymorphic loci include those described in, U.S. Pat. Nos. 5,242,794, 5,494,810, 4,988,617 and 6,582,938.

[0081] Following amplification, the amplified nucleic acid is sequenced. Sequencing can be accomplished using high-throughput systems some of which allow detection of a sequenced nucleotide immediately after or upon its incorporation into a growing strand, i.e., detection of sequence in red time or substantially real time. In some cases, high throughput sequencing

generates at least 1,000, at least 5,000, at least 10,000, at least 20,000, at least 30,000, at least 40,000, at least 50,000, at least 100,000 or at least 500,000 sequence reads per hour; with each read being at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 120 or at least 150 bases per read. Sequencing can be performed using nucleic acids described herein such as genomic DNA, cDNA derived from RNA transcripts or RNA as a template.

[0082] In some embodiments, high-throughput sequencing involves the use of technology available by Helicos BioSciences Corporation (Cambridge, Massachusetts) such as the Single Molecule Sequencing by Synthesis (SMSS) method. SMSS is unique because it allows for sequencing an entire genome with no pre amplification step needed. Thus, distortion and nonlinearity in the measurement of nucleic acids are reduced. SMSS is described in part in US Publication Application Nos. 2006002471 I; 20060024678; 20060012793; 20060012784; and 20050100932.

[0083] In some embodiments, high-throughput sequencing involves the use of technology available by 454 Lifesciences, Inc. (Branford, Connecticut) such as the Pico Titer Plate device, which includes a fiber optic plate that transmits chemiluminescent signal generated by the sequencing reaction to be recorded by a CCD camera in the instrument. This use of fiber optics allows for the detection of a minimum of 20 million base pairs in 4.5 hours.

[0084] Methods for using bead amplification followed by fiber optics detection are described in Marguiles, M., et al. "Genome sequencing in microfabricated high-density pricolitre reactors", Nature, doi: 10.1038/nature03959; and well as in US Publication Application Nos. 20020012930; 20030058629; 20030100102; 20030148344; 20040248161; 20050079510, 20050124022; and 20060078909.

[0085] In some embodiments, high-throughput sequencing is performed using Clonal Single Molecule Array (Solexa, Inc.) or sequencing-by-synthesis (SBS) utilizing reversible terminator chemistry. These technologies are described in part in US Patent Nos. 6,969,488; 6,897,023; 6,833,246; 6,787,308; and US Publication Application Nos. 200401061 30; 20030064398; 20030022207; and Constans, A,, The Scientist 2003, 17(13):36.

[0086] In some embodiments of this aspect, high-throughput sequencing of RNA or DNA can take place using AnyDot.chips (Genovoxx, Germany), which allows for the monitoring of biological processes (e.g., miRNA expression or allele variability (SNP detection). In particular, the AnyDot-chips allow for 10x - 50x enhancement of nucleotide fluorescence signal detection. AnyDot.chips and methods for using them are described in part in International Publication Application Nos. WO 02088382, WO 03020968, WO 0303 1947, WO 2005044836, PCTEP 05105657, PCMEP 05105655; and German Patent Application Nos. DE 101 49 786, DE 102 14 395, DE 103 56 837, DE 10 2004 009 704, DE 10 2004 025 696, DE

10 2004 025 746, DE 10 2004 025 694, DE 10 2004 025 695, DE 10 2004 025 744, DE 10 2004 025 745, and DE 10 2005 012 301.

[0087] Other high-throughput sequencing systems include those disclosed in Venter, J., et al. Science 16 February 2001; Adams, M. et al, Science 24 March 2000; and M. J, Levene, et al. Science 299:682-686, January 2003; as well as US Publication Application No. 20030044781 and 2006/0078937. Overall such system involve sequencing a target nucleic acid molecule having a plurality of bases by the temporal addition of bases via a polymerization reaction that is measured on a molecule of nucleic acid, i.e., the activity of a nucleic acid polymerizing enzyme on the template nucleic acid molecule to be sequenced is followed in real time. Sequence can then be deduced by identifying which base is being incorporated into the growing complementary strand of the target nucleic acid by the catalytic activity of the nucleic acid polymerizing enzyme at each step in the sequence of base additions. A polymerase on the target nucleic acid molecule complex is provided in a position suitable lo move along the target nucleic acid molecule and extend the oligonucleotide primer at an active site. A plurality of labeled types of nucleotide analogs are provided proximate to the active site, with each distinguishably type of nucleotide analog being complementary to a different nucleotide in the target nucleic acid sequence. The growing nucleic acid strand is extended by using the polymerase to add a nucleotide analog to the nucleic acid strand at the active site, where the nucleotide analog being added is complementary to the nucleotide of the target nucleic acid at the active site. The nucleotide analog added to the oligonucleotide primer as a result of the polymerizing step is identified. The steps of providing labeled nucleotide analogs, polymerizing the growing nucleic acid strand, and identifying the added nucleotide analog are repeated so that the nucleic acid strand is further extended and the sequence of the target nucleic acid is determined.

[0088] In some embodiments, shotgun sequencing is performed. In shotgun sequencing, DNA is broken up randomly into numerous small segments, which are sequenced using the chain termination method to obtain reads. Multiple overlapping reads for the target DNA are obtained by performing several rounds of this fragmentation and sequencing. Computer programs then use the overlapping ends of different reads to assemble them into a continuous sequence.

[0089] In some embodiments, the invention provides methods for detection and quantitation of microbial sequences using sequencing. In this case, one can estimate the sensitivity of detection. There are two components to sensitivity: (i) the number of molecules analyzed (depth of sequencing) and (ii) the error rate of the sequencing process. Regarding the depth of sequencing, a frequent estimate for the variation between individuals is that about one base per thousand differs. Currently, sequencers such as the Illumina Genome Analyzer have read lengths exceeding 36 base pairs. While the fraction of host DNA in the blood may be variable

depending on the status of the individual, one can take 90% as a baseline estimate. At this fraction of donor DNA, approximately one in 10 molecules analyzed will be microbial. On the Genome Analyzer one can obtain about 10 million molecules per analysis channel and there are 8 analysis channels per instrument run. Therefore, if one sample is loaded per channel, one should be able to detect about 10^6 molecules that can be identified as microbial and informative as to the status of the microbiome. Higher sensitivity can be achieved simply by sequencing more molecules, i.e. using more channels.

[0090] The sequencing error rate also affects the sensitivity of this technique. Typical sequencing error rates for base substitutions vary between platforms, but are between 0.5-1.5%. This places a potential limit on sensitivity of 0.16 to 0.50%. However, it is possible to systematically lower the sequencing error rate by resequencing the sample template multiple times, as has been demonstrated by Helicos BioSciences (Harris, T.D., et al., Science, 320, 106-109 (2008)). A single application of resequencing would reduce the expected error rate.

[0091] Following sequencing, the dataset of sequences is uploaded to a data processor for bioinformatics analysis to subtract host sequences, i.e. human, cat, dog, etc. from the analysis; and determine the presence and prevalence of microbial sequences, for example by a comparison of the coverage of sequences mapping to a microbial reference sequence to coverage of the host reference sequence. The subtraction of host sequences may include the step of identifying a reference host sequence, and masking microbial sequences or microbial-mimicking sequences present in the reference host genome. Similarly, determining the presence of a microbial sequence by comparison to a microbial reference sequence may include the step of identifying a reference microbial sequence, and masking host sequences or host-mimicking sequences present in the reference microbial genome.

[0092] The dataset is optionally cleaned to check sequence quality, remove remnants of sequencer specific nucleotides (adapter sequences), and merge paired end reads that overlap to create a higher quality consensus sequence with less read errors. Repetitive sequences are identified as those having identical start sites and length, and duplicates may be removed from the analysis.

[0093] An important feature of the invention is the subtraction of human sequences from the analysis. As the amplification/sequencing steps are unbiased, the preponderance of sequences in a sample will be host sequences. The subtraction process may be optimized in several ways to improve the speed and accuracy of the process, for example by performing multiple subtractions where the initial alignment is set at a coarse filter, i.e. with a fast aligner, and performing additional alignments with a fine filter, i.e. a sensitive aligner.

[0094] The database of reads are initially aligned against a human reference genome, including without limitation Genbank hg19 reference sequences, to bioinformatically subtract the host DNA. Each sequence is aligned with the best fit sequence in the human reference

sequence Sequences positively identified as human are bioinformatically removed from the analysis.

[0095] The reference human sequence can also be optimized by adding in contigs that have a high hit rate, including without limitation highly repetitive sequence present in the genome that are not well represented in reference databases. It has been observed that of the reads that do not align to hg19, a significant amount is eventually identified as human in a later stage of the pipeline, when a database that includes a large set of human sequences is used, for example the entire NCBI NT database. Removing these reads earlier in the analysis can be performed by building an expanded human reference. This reference is created by identifying human contigs in a human sequence database other than the reference, e.g. NCBI NT database, that have high coverage after the initial human read subtraction. Those contigs are added to the human reference to create a more comprehensive reference set. Additionally novel assembled human contigs from cohorts studies can be used as a further mask for human-derived reads.

[0096] Regions of the human genome reference sequence that contain non-human sequences may be masked, e.g. viral and bacterial sequences that are integrated into the genome of the reference sample. For example, Epstein-Barr Virus (EBV) has about 80% of its genome incorporated into hg19.

[0097] Sequence reads identified as non-human are then aligned to a nucleotide database of microbial reference sequences. The database may be selected for those microbial sequences known to be associated with the host, e.g. the set of human commensal and pathogenic microrganisms.

[0098] The microbial database may be optimized to mask or remove contaminating sequences. For example it has been observed that many public database entries include artifactual sequences not derived from the microorganism, e.g., primer sequences, host sequences, and other contaminants. It is desirable to perform an initial alignment or plurality of alignments on a database. Regions that show irregularities in read coverage when multiple samples are aligned can be masked or removed as an artifact. The detection of such irregular coverage can be done by various metrics, such as the ratio between coverage of a specific nucleotide and the average coverage of the entire contig within which this nucleotide is found. In general a sequence that is represented as greater than about 5X, about 10X, about 25X, about 50X, about 100X the average coverage of that reference sequence are artifactual. Alternatively a binomial test can be applied to provide a per-base likelihood of coverage given the overall coverage of the contig. Removal of contaminant sequence from reference databases allows accurate identification of microbes. It is a benefit of the methods of the invention that the databases are improved with alignment of samples, e.g. a database may be

aligned with 1, 10, 20, 50, 100 or more samples to improve the database prior to commercial or clinical use.

[0099] Each high confidence read may align to multiple organisms in the given microbial database. To correctly assign organism abundance based upon this possible mapping redundancy, an algorithm is used to compute the most likely organism an algorithm is selected (for example see Lindner et al. Nucl. Acids Res. (2013) 41 (1): e10). For example GRAMMy or GASiC algorithms can be used to compute the most likely organism that a given read came from. These data provide information regarding the presence of a microbe in the cell-free nucleic acid sample.

[00100] Alignments and assignment to a host sequence or to a microbial sequence may be performed in accordance with art-recognized methods. For example, a read of 50 nt. may be assigned as matching a given genome if there is not more than 1 mismatch, not more than 2 mismatches, not more than 3 mismatches, not more than 4 mismatches, not more than 5 mismatches, etc. over the length of the read. Commercial algorithms are generally used for alignments and identification. A non-limiting example of such an alignment algorithm is the bowtie2 program (Johns Hopkins University). For example, the pre-set options in end-to-end mode may be selected based on the desired speed of alignment,

`--very-fast`	Same as: `-D 5 -R 1 -N 0 -L 22 -i S,0,2.50`
`--fast`	Same as: `-D 10 -R 2 -N 0 -L 22 -i S,0,2.50`
`--sensitive`	Same as: `-D 15 -R 2 -L 22 -i S,1,1.15`
`--very-sensitive`	Same as: `-D 20 -R 3 -N 0 -L 20 -i S,1,0.50`

Comparable setting may be used in other alignment algorithms or software packages.

[00101] These assignments of reads to an organism (i.e. host or microbiome component) are then totaled and used to compute the estimated number of reads assigned to each organism in a given sample, in a determination of the prevalence of the organism in the cell-free nucleic acid sample. The analysis normalizes the counts for the size of the microbial genome to provide a calculation of coverage for the microbe. The normalized coverage for each microbe is compared to the host sequence coverage in the same sample to account for differences in sequencing depth between samples.

[00102] The final determination provides a dataset of microbial organisms represented by sequences in the sample, and the prevalence of those microorganisms. These data are optionally aggregated and displayed for ready visualization, e.g. in the form of a report provided to the individual or health care provider; or written in a browser format with hyper-linked data. The coverage estimation can be aggregated with metadata from the sample, and sorted into tables and figures for each sample, or cohort of samples.

[00103] Optionally the host sequences that are filtered out can be used for other purposes, e.g. in personalized medicine. For example, certain SNPs in the human genome may allow

doctors to identify drug sensitivities for a given patient. The human-derived sequences may reveal integration of viruses into the host's genome (e.g. EBV, HPV, poliomavirus) or be used for synergistic clinical applications (e.g., cell-free tumor DNA may be used to monitor cancer progressing in parallel with infection monitoring in patients that are highly susceptible to infection due to chemotherapy).

[00104] In some embodiments, the analysis of cell-free nucleic acids is used to compute a pathogenicity score, where the pathogenicity score is a numeric or alphabetic value that summarizes the overall pathogenicity of the organism for ease of interpretation, e.g. by a health practitioner. Different microbes present in the microbiome may be assigned different scores. The final "pathogenicity score" is a combination of many different factors, and typically provided as an arbitrary unit, for example ranging from 0-1, 0-10 or 0-100; as a percentile from all observed pathogenicity scores for a microbe of interest, etc. The specific parameters and weights of those parameters may be determined experimentally, e.g. by fitting the function to observed disease severity, or manually by setting the importance of different paramaters and criteria.

[00105] Factors relevant for calculation of a pathogenicity score may include, without limitation, abundance of the microbe, e.g. as computed by number of reads relative to human reads, relative to the abundance of the microbe in a reference subject or group of subjects, e.g. a test population, a known infection, a known un-infected individual, etc. Specific mutations found in the microbe genome, which may be made with reference to a database of toxicity, pathogenicity, antibiotic resistance *etc*. associated with the microbe, and including without limitation SNPs, indels, plasmids etc. The co-incidence of specific microbes, including without limitation specific ratios and groups of organisms. Expression of certain sequences, e.g. be detection of mRNA, can be relevant to the pathogenicity score, e.g. as informative of whether a microbe is actively replicating or is latent; *etc.* Geographic features may also be included, where the geography is indicative of exposure to microbes of interest, e.g. travel history of the host; interactions with infected individuals, and the like.

[00106] Also provided are reagents and kits thereof for practicing one or more of the above-described methods. The subject reagents and kits thereof may vary greatly. Reagents of interest include reagents specifically designed for use in production of the above-described: (i) profiling of a microbiome and an imdividual; (ii) identification of microbiome profiles; and (ii) detection and/or quantitation of one or more nucleic acids from a microbiome in a sample obtained from an individual. The kits may comprise reagents necessary to perform nucleic acid extraction and/or nucleic acid detection using the methods described herein such as PCR and sequencing. The kit may further comprise a software package for data analysis, which may include reference profiles for comparison with the test profile., and in particular may

include reference databases optimized as described above. The kits may comprise reagents such as buffers, and H_20.

[00107] Such kits may also include information, such as scientific literature references, package insert materials, clinical trial results, and/or summaries of these and the like, which indicate or establish the activities and/or advantages of the composition, and/or which describe dosing, administration, side effects, drug interactions, or other information useful to the health care provider. Such kits may also include instructions to access a database. Such information may be based on the results of various studies, for example, studies using experimental animals involving in vivo models and studies based on human clinical trials. Kits described herein can be provided, marketed and/or promoted to health providers, including physicians, nurses, pharmacists, formulary officials, and the like. Kits may also, in some embodiments, be marketed directly to the consumer.

[00108] Any of the methods above can be performed by a computer program product that comprises a computer executable logic that is recorded on a computer readable medium. For example, the computer program can execute some or all of the following functions: (i) controlling isolation of nucleic acids from a sample, (ii) pre-amplifying nucleic acids from the sample, (iii) amplifying, sequencing or arraying specific regions in the sample, (iv) identifying and quantifying a microbial sequence in the sample, (v) comparing data on a microbe presence or prevalence detected from the sample with a predetermined threshold, (vi) determining infection, microbiome health, immunocompetence status or outcome, (vi) declaring the sample status with respect to infection, microbiome health, immunocompetence, etc.

[00109] The computer executable logic can work in any computer that may be any of a variety of types of general-purpose computers such as a personal computer, network server, workstation, or other computer platform now or later developed. In some embodiments, a computer program product is described comprising a computer usable medium having the computer executable logic (computer software program, including program code) stored therein. The computer executable logic can be executed by a processor, causing the processor to perform functions described herein. In other embodiments, some functions are implemented primarily in hardware using, for example, a hardware state machine. Implementation of the hardware state machine so as to perform the functions described herein will be apparent to those skilled in the relevant arts.

[00110] The program can provide a method of evaluating microbial status in an individual by accessing data that reflects the profiling of the microbiome and the individual, and/or the quantitation of one or more nucleic acids from the microbiome in the circulation of the individual.

[00111] In one embodiment, the computer executing the computer logic of the invention may also include a digital input device such as a scanner. The digital input device can provide information on a nucleic acid, e.g., presence or prevalence

[00112] In some embodiments, the invention provides a computer readable medium comprising a set of instructions recorded thereon to cause a computer to perform the steps of (i) receiving data from one or more nucleic acids detected in a sample; and (ii) diagnosing or predicting a status based on the microbiome quantitation.

[00113] Also provided are databases of microbial reference sequences; and databases of human reference sequences. Such databases will typically comprise optimized datasets as described above.

[00114] In some embodiments the methods of the invention provide an individual's status with respect to infection. In some such embodiments, the microbial infection is a pathogen, where any presence of the pathogen sequence indicates a clinically relevant infection. In other embodiments, the prevalence is indicative of microbial load, where a pre-set level is indicative of clinical relevance. In some such embodiments the individual is treated or considered for treatment with an antimicrobial therapy, e.g. antibiotics, passive or active immunotherapy, antivirals, etc. An individual may be tested before therapy, during therapy, and after therapy.

[00115] A microbial infection may also be indicated by load for a commensal organism, where the level of a commensal in a blood sample is indicative of intestinal health, e.g. gut lumen breakdown.

[00116] A comparison can be made of microbioal RNA, alone or in relation to microbiol DNA, where an excess of RNA for a microbial sequence, e.g. about 5X, 10X, 15X, 20X, 25X the coverage of the microbial DNA is indicative of an active infection. In some embodiments the microbe thus analyzed is one capable of latent infection, e.g. herpesvirus, hepatitis virus, etc.

[00117] In other embodiments, an overall estimate of the microbiome is of interest, where the relative presence of prevalence of classes of microorganisms are of interest. It is known in the art that diet and treatment with drugs, e.g. statins, antibiotics, immunosuppressive agents, etc. can affect the overall health of the microbiome, and it is therefor of interest to determine the composition of the microbiome.

[00118] In some embodiments, temporal differences in the amount of said one or more nucleic acids from the microbiome can be used to monitor effectiveness of anti-microbial treatment or to select an treatment. For instance, the amount of one or more nucleic acids from the microbiome can be determined before and after an treatment. A decrease in the one or more nucleic acids from the microbe after treatment may indicate that the treatment was successful. Additionally, the amount of one or more nucleic acids from the microbiome can be used to choose between treatments, for examples, treatments of different strengths.

[00119] In one aspect the invention provides methods for the diagnosis or prediction of immunocompetence, transplant status or outcome in a subject receiving an immunosuppressive regimen. Following immunosuppression, samples as described above can be drawn from the patient and analyzed for the presence or absence of one or more microbiome, including virome nucleic acids. In some embodiments, the sample is blood, plasma, serum or urine. The proportion and/or amount of microbial nucleic acids can be monitored over time and an increase in this proportion can be used to determine immunocompetence. The quantitation of load may be determined by any suitable method known in the art including those described herein such as sequencing, nucleic acid arrays or PCR.

[00120] In some embodiments, the amount of one or more microbiome nucleic acids in a sample from the immunosuppressed recipient is used to determine the transplant status or outcome. Thus, in some embodiments, the methods of the invention further comprise quantitating the one or more nucleic acids from the microbiome. In some embodiments, the amount of one or more nucleic acids from the donor sample is determined as a percentage of total the nucleic acids in the sample. In some embodiments, the amount of one or more nucleic acids from the donor sample is determined as a ratio of the total nucleic acids in the sample. In some embodiments, the amount of one or more nucleic acids from the donor sample is determined as a ratio or percentage compared to one or more reference nucleic acids in the sample. For instance, the amount of one or more nucleic acids from the microbiome can be determined to be 10% of the total nucleic acids in the sample. Alternatively, the amount of one or more nucleic acids from the microbiome can be at a ratio of 1:10 compared to total nucleic acids in the sample. Further, the amount of one or more nucleic acids from the microbiome can be determined to be 10% or at a ratio of 1:10 of a reference gene such a β-globin. In some embodiments, the amount of one or more nucleic acids from the microbiome can be determined as a concentration. For example, the amount of one or more nucleic acids from the donor sample can be determined to be 1 μg/mL.

[00121] In some embodiments, the amount of one or more nucleic acids from the microbiome above a predetermined threshold value is indicative of a immunocompetence status. For example, the normative values for clinically stable patients with no evidence of graft rejection or other pathologies can be determined. An increase in the amount of one or more nucleic acids from the microbiome below the normative values for clinically stable post-transplantation patients could indicate a stable outcome. On the other hand, an amount of one or more nucleic acids from the microbiome above or at the normative values for clinically stable post-transplantation patients could indicate increased immunocompetence and risk of graft rejection.

[00122] In some embodiments, different predetermined threshold values are indicative of different transplant outcomes or status. For example, as discussed above, an increase in the amount of one or more nucleic acids from the microbiome above the normative values for clinically stable post-transplantation patients could indicate a change in transplant status or outcome such as transplant rejection or transplant injury. However, an increase in the amount of one or more nucleic acids from the microbiome above the normative values for clinically stable post-transplantation patients but below a predetermined threshold level could indicate a less serious condition such as a viral infection rather than transplant rejection. An increase in the amount of one or more nucleic acids from the microbiome above a higher threshold could indicate transplant rejection.

[00123] In some embodiments, temporal differences in the amount of said one or more nucleic acids from the microbiome are indicative of immunocompetence. For instance, a transplant patient can be monitored over time to determine the amount of one or more nucleic acids from the microbiome. A temporary decrease in the amount of one or more nucleic acids from the microbiome, which subsequently return to normal values, might indicate a less serious condition rather than transplant rejection. On the other hand, a sustained decrease in the amount one or more nucleic acids from the microbiome might indicate a serious condition such as lack of effective immunosuppression and graft rejection.

[00124] In some embodiments, temporal differences in the amount of said one or more nucleic acids from the microbiome can be used to monitor effectiveness of an immunosuppressant treatment or to select an immunosuppressant treatment. For instance, the amount of one or more nucleic acids from the microbiome can be determined before and after an immunosuppressant treatment. A decrease in the one or more nucleic acids from the microbiome after treatment may indicate that the treatment was successful in preventing transplant rejection. Additionally, the amount of one or more nucleic acids from the microbiome can be used to choose between immunosuppressant treatments, for examples, immunosuppressant treatments of different strengths. For example, a lower amount in one or more nucleic acids from the microbiome may indicate that there is a need of a very potent immunosuppressant, whereas a higher amount in one or more nucleic acids from the microbiome may indicate that a less potent immunosuppressant may be used.

[00125] The invention provides methods that sensitive and specific. In some embodiments, the methods described herein for diagnosing or predicting transplant status or outcome have at least 50%, 60%, 70%, 80%, 90%, 95% or 100% sensitivity. In some embodiments, the methods described herein have at least 50 % sensitivity. In some embodiments, the methods described herein have at least 78 % sensitivity. In some embodiments, the methods described herein have a specificity of about 70% to about 100%. In some embodiments, the methods described herein have a specificity of about 80% to about 100%. In some

embodiments, the methods described herein have a specificity of about 90% to about 100%. In some embodiments, the methods described herein have a specificity of about 100%.

[00126] The invention provides non-invasive diagnostics for individuals, including individuals that are being treated with immunosuppressive regimens, treated with anti-microbial agents, etc., by monitoring the sequences of cell-free DNA or RNA from non-human sources. For example, individuals carry a number of virus, where the virus load is shown herein to vary with the immunocompetence of the individual. Preferred virus for monitoring immunocompetence are annellovirus, in which the viral burden is shown herein to correlate with immunocompetence of the individual.

[00127] In some embodiments, the invention provides methods, devices, compositions and kits for detection and/or quantitating circulating nucleic acids, usually free in plasma or from viral particles, for the diagnosis, prognosis, detection and/or treatment of an infection, of immunocompetence, transplant status or outcome.

[00128] In some specific embodiments, the invention provides an approach to noninvasive detection of immunocompetence in transplant patients by virome analysis, which circumvents the potential problems of microchimerism from DNA from other foreign sources and is general for all organ recipients without consideration of gender. In some embodiments, a genetic fingerprint is generated for the virome of the individual. This approach allows for a reliable identification of sequences that can be made in a manner that is independent of the genders of donor and recipient.

[00129] Following an immunosuppressive regimen, e.g. in conjunction with transplantation, treatment of autoimmune disease, etc., bodily fluid such as blood can be drawn from the patient and analyzed for markers. Examples of bodily fluids include, but are not limited to, smears, sputum, biopsies, secretions, cerebrospinal fluid, bile, blood, lymph fluid, saliva, and urine. Detection, identification and/or quantitation of the virome sequences can be performed using real-time PCR, chips, high-throughput shotgun sequencing of circulating nucleic acids (e.g. cell-free DNA), as well as other methods known in the art including the methods described herein. The viral load can be monitored over time and an increase in this proportion can be used to determine immunocompetence status or outcome.

[00130] In any of the embodiments described herein, the transplant graft can be any solid organ or skin transplant. Examples of organ transplants that can be analyzed by the methods described herein include but are not limited to kidney transplant, pancreas transplant, liver transplant, heart transplant, lung transplant, intestine transplant, pancreas after kidney transplant, and simultaneous pancreas-kidney transplant.

[00131] In some other embodiments, the methods of the invention are used in determining the efficacy of a therapy for treatment of disease, including infection, either at an individual level, or in the analysis of a group of patients, e.g. in a clinical trial format. Such embodiments typically involve the comparison of two time points for a patient or group of patients. The patient status is expected to differ between the two time points as the result of a therapeutic agent, therapeutic regimen, or disease challenge to a patient undergoing treatment.

[00132] Examples of formats for such embodiments may include, without limitation, analyzing the microbiome at two or more time points, where a first time point is a diagnosed but untreated patient; and a second or additional time point(s) is a patient treated with a candidate therapeutic agent or regimen.

[00133] In another format, a first time point is a diagnosed patient in disease remission, e.g. as ascertained by current clinical criteria, as a result of a candidate therapeutic agent or regimen. A second or additional time point(s) is a patient treated with a candidate therapeutic agent or regimen, and challenged with a disease-inducing agent, for example in a vaccine context.

[00134] In such clinical trial formats, each set of time points may correspond to a single patient, to a patient group, e.g. a cohort group, or to a mixture of individual and group data. Additional control data may also be included in such clinical trial formats, e.g. a placebo group, a disease-free group, and the like, as are known in the art. Formats of interest include crossover studies, randomized, double-blind, placebo-controlled, parallel group trial is also capable of testing drug efficacy, and the like. See, for example, Clinical Trials: A Methodologic Perspective Second Edition, S. Piantadosi, Wiley-Interscience; 2005, ISBN-13: 978-0471727811; and Design and Analysis of Clinical Trials: Concepts and Methodologies, S. Chow and J. Liu, Wiley-Interscience; 2003; ISBN-13: 978-0471249856, each herein specifically incorporated by reference.

EXAMPLES

Temporal Response of the Human Virome to Immunosuppression and Antiviral Therapy

[00135] The viral component of the microbiome, the human virome, remains relatively understudied (Wylie et al. (2012) Transl Res *160*, 283-290) and little is known about the effects of immune modulation and antiviral therapies on virome composition. It was previously shown that the healthy gut virome remains remarkably stable over time (Reyes et al. (2010) Nature 466, 334-338), and that the predominant source of variation is due to differences between subjects, although an association between diet and the virome composition was found (Minot, et al. (2011). Genome Research *21*, 1616-1625).

[00136] Immunosuppressive therapies significantly reduce the risk of graft rejection in organ transplantation but increase the susceptibility of recipients to infections. Infections with viral pathogens, in particular the herpesvirus cytomegalovirus (CMV), occur frequently and

increase the recipient's risk of graft failure. Organ transplant recipients are therefore frequently subjected to antiviral prophylactic or preemptive therapies directed against CMV.

[00137] The inverse relationship between the level of immunosuppression and the risks of infection and rejection leaves only a narrow therapeutic window available for patient treatment. Post-transplant care is further complicated by numerous limitations of the currently available methods for the diagnosis of infection and rejection. Diagnosis of rejection mostly relies on invasive biopsies that suffer from interobserver variability, high cost and patient discomfort. Diagnosis of infections is challenging given the fact that the symptoms of infection are diminished following immunosuppression, and commonly used diagnostic methods, such as antigen-detection and PCR-based molecular tests, rely on a specific target and therefore an a priori hypothesis for the source of the infection.

[00138] As a final complication, patient-to-patient variability in the sensitivity to immunosuppressive drugs can give rise to over- and under immunosuppression, increasing the risk of infection or rejection respectively.

[00139] There are few substantive methods to measure the health of the immune system, and the connection between immunocompetence and the viral component of the microbiome is poorly understood. Organ transplant recipients are treated with a post-transplant therapy that combines immunosuppressive and antiviral drugs, offering a window into the effects of immune modulation on the human virome. We used sequencing of cell-free DNA in plasma to investigate drug-virome interactions in a cohort of organ transplant recipients (656 samples, 96 patients), and find that antivirals and immunosuppressants strongly affect the structure of the virome in plasma. We observe marked virome compositional dynamics at the onset of the therapy and find that the total viral load increases with immunosuppression, whereas the bacterial component of the microbiome remains largely unaffected. The data provide insight into the relationship between the human virome, the state of the immune system, and the effects of pharmacological treatment, and offer a potential application of the virome state to predict immunocompetence.

[00140] In this work, we sequenced cell-free DNA circulating in plasma to investigate drug-microbiome interactions following organ transplantation. We studied the patterns of infection in heart and lung transplant recipients subjected to a combination of immunosuppressants and antiviral prophylaxis. We find that immunosuppressants and antivirals have a strong influence on the structure of the viral component of the microbiome but not the bacterial component. Strong compositional dynamics are observed at the onset of the drug therapy as the virome composition of different individuals converge to a similar, drug-determined state. The total viral load increases markedly in response to the therapy, as viruses, in particular the anelloviruses, take advantage of a reduction of immunocompetence. Finally, we show that measurement of the anellovirus burden enables stratification of rejecting and non-rejecting recipients.

[00141] 656 plasma samples were collected longitudinally from 96 solid organ transplant recipients (41 adult heart, 24 pediatric heart, 31 adult lung). Cell-free DNA was purified from plasma and sequenced. In total, we obtained 820 gigabases (Gbp) of sequencing data, with an average of 1.25 Gbp per sample (Illumina HiSeq, 1x50 bp reads, Fig. 1B). Organ transplant recipients were continuously enrolled in the study over the course of more than 2 years and samples were collected from the recipients at regular time points post transplant, with the highest frequency of sample collection in the first months post transplant. Fig. 1C shows the number of samples analyzed as a function of time post transplant for the different patient classes.

[00142] The patients in the cohort were treated with antiviral prophylaxis and immunosuppression as part of a standardized post-transplant therapy (Fig. 1D). Maintenance immunosuppression was tacrolimus-based for the adult heart and lung transplant recipients and was complemented with mycophenolate mofetil and prednisone. Pediatric patients were treated with a cyclosporine based anti-rejection therapy. CMV positive transplant recipients (prior CMV infection for recipient and/or donor), but not CMV negative recipients, were treated with antiviral prophylaxis. The protocol design entails high doses of immunosuppressants and antiviral drugs in the first few months post transplant, after which the doses are gradually reduced as the risks of rejection and infection diminish. Given the narrow therapeutic window available for immunosuppression and the large patient-to-patient variability in pharmacokinetics of tacrolimus, the concentration of the tacrolimus is directly measured in the blood and the dose is adjusted to maintain a target drug level. Fig. 1D shows the mean level of tacrolimus measured in blood for the tacrolimus-treated patients and illustrates the design of the drug treatment protocol.

[00143] *DNA sequence analysis.* Microbiome-derived sequences were identified after computational subtraction of human-derived sequences. To this end, duplicate and low quality reads were removed and the remaining reads were mapped to the human reference genome, build hg19 (BWA (Li and Durbin, 2009), see methods). Unmapped reads were then collected and low complexity reads were removed. Fig. 1E shows the distribution of the remaining read fraction after applying duplicate and quality filters (average of 86 %) and the distribution of the remaining fraction after subtraction of human reads (average of 2%).

[00144] To identify infectious agents, the remaining, high quality, unique, non-human reads were mapped using BLAST to a reference database of viral (n=1401), bacterial (n=1980) and fungal (n=32) genomes (downloaded from NCBI, Fig. 6A). 0.12% of the uniquely sequenced reads aligned to at least one of the target genomes (Fig. 6B, C). We used a quantitative PCR (qPCR) assay targeted to a subset of sequencing identified targets (herpesviruses 4,5,6 and parvovirus) to validate the positive hits identified by the sequencing-based approach. We

found a quantitative agreement between viral counts as measured by sequencing and qPCR (Fig. 6D).

[00145] We furthermore found that the sensitivity of the sequencing assay for the detection of herpesviruses is on par with qPCR measurements. The larger capture cross-section available to the sequencing assay – the complete target genome versus the PCR amplicon target region – is thus sufficient to overcome the signal loss in sequencing, caused by the finite efficiency of sequencing library preparation and library undersampling. The highest CMV loads measured using sequencing across all samples in the study corresponded to two adult heart transplant patients that suffered from a clinically diagnosed disseminated CMV infection (see Fig. 6E).

[00146] To test for the presence of potential contaminants in the reagents used for DNA extraction and sequencing library preparation, we performed two control experiments. In the first, we prepared 2 samples with a known template (Lambda gDNA, Pacbio Part no: 001-119-535), and purified DNA for sequencing using the above-described workflow (Illumina Miseq, 3.4 and 3.5 million reads). Lambda-derived sequences were removed and the remaining sequences (0.4%) were aligned to the BLAST reference database described above. No evidence was found for the various infectious agents discussed in this work, but we did detect sequences related to the Enterobacteriaceae bacterial family (phylum Proteobacteria), primarily E. coli (> 97%), and enterobacterial phages (<1%), which are likely a remnant of the lambda DNA culture. In a second control, we prepared a sample for sequencing from nuclease-free water. The sample was included in a sequencing run along with a sample unrelated to this work and recruited only a limited number of sequences, 15 in total, which mapped to genomes of two bacterial species. Again, no evidence was found for the infectious agents that are discussed below.

[00147] We studied the microbiome composition in plasma at different levels of taxonomic classification using Grammy, a tool that utilizes the sequence-similarity data obtained with BLAST to perform a maximum likelihood estimation of the relative abundance of species. GRAMMy accounts for differences in target genome size and the ambiguity of read assignments. Note that this approach only allows estimating the abundance of species for which genomic data is available in the reference database. Figure 1F shows the relative abundance of species at different levels of taxonomic classification (average over all samples). We find that viruses (73%) are more abundantly represented than bacteria (25%) and fungi (2%) (Fig. 1F panel a). Among viruses, we find that ssDNA viruses occupy a larger fraction (72%) than dsDNA viruses (28%). Seven distinct viral families are found (abundance > 0.75%), with one dominant family, the *Anelloviridae*, which accounted for 68% of the total population (Fig. 1F, panel b). The anelloviridae fraction is mostly (97%) composed of viruses from the *Alphatorquevirus* genus (Fig. 1F, panel c). The *Alphatorque* genus is the genus of

Torque Teno Viruses (TTVs), and sequences related to 14 different torque teno virotypes were identified (Fig1, panel d). Infections with polyomaviruses are widespread in the human population, and polyomavirus DNAemia is not uncommon in the first year after solid organ transplantation. Polyomavirus-derived sequences were found in 75 samples (11%) corresponding to 36 patients in the present cohort. Evidence for the presence of BK (41%), JC (27%), TS (4%), WU polyomavirus (6%), SV40 (6%) and the recently discovered HPyV6 (13 %) (Schowalter et al., 2010) was found (Fig. 1F, panel e). Among bacteria, Proteobacteria (36%), Firmicutes (50%), Actinobacteria (10%), Bacteroidetes (4%) are the phyla most abundantly represented in the sample (Fig. 1F, panel f).

[00148] To investigate potential incorrect assignments of the relatively short reads available to this study (50 bp), we examined the dependence of the abundance estimates on read length, based on longer, paired-end reads (2x100 bp) collected for a subset of samples (n=55). We found that the abundance estimates based on 50 bp subreads and 100 bp reads are similar for all levels of taxonomic classification reported here (fig. 6F).

[00149] *Sensitivity of virome composition to drug dosage.* The available clinical data on drug dosage was used to analyze drug-microbiome interactions. Here, we examined data for the adult heart and lung transplant patients that were treated with a tacrolimus-based anti-rejection protocol (47 patients and 380 observations), thereby excluding the pediatric patients that were treated with cyclosporine and patients that were switched from tacrolimus to cyclosporine immunosuppression due to drug-intolerance issues. Data on prescription antiviral drug doses (valganciclovir) and the measured levels of tacrolimus in blood were collected from individual patient records and the mean composition for samples corresponding to different drugs levels was extracted. To account for a delayed effect of the microbiome composition on dose changes, the drug level and dose data were sliding window average filtered (see Fig. 1C and Fig.7A-C; window size 45 days).

[00150] We find that the structure of the viral component of the microbiome is a sensitive function of drug dosage (47 patients, 380 samples, Figure 2A). However, the structure of the bacterial component of the microbiome was not significantly altered by the drug therapy, as discussed further below (Fig. 7D). Herpesvirales and caudovirales dominated the virome when patients received a low dose of valganciclovir and tacrolimus. In contrast, a high dose of immunosuppressants and antivirals gave rise to a virome structure that is dominated by anelloviridae (up to 94% occupation at high drug levels). The antiviral prophylaxis is intended to prevent CMV disease, but other herpesviruses are also susceptible to the drug so it is not surprising that a higher dose of valganciclovir gives rise to a lower fraction of viruses from the Herpesvirales order. The observation that anelloviridae take advantage of suppression of the host immune system is consistent with various observations from the literature: it was

previously shown that the incidence of anelloviridae increases with progression towards AIDS in HIV patients, and that the total burden of the anellovirus TTV increases post liver transplantation. Furthermore, an increased prevalence of anelloviridae in pediatric patients with fevers was reported recently.

[00151] We next compare the virome composition measured for organ transplant recipients to the composition observed in healthy individuals, not on immunosupressants or antivirals (n = 9, sequencing data available from a previous study). Here, we compare the healthy composition to the composition measured for organ transplant recipients at the start of the drug therapy (post-operative day one, n = 13), corresponding to a minimal drug exposure, and to the composition measured for transplant recipients exposed to high drug levels (well after the transplant procedure, tacrolimus ≥ 9 ng/ml. valgancicolvir ≥ 600 mg, n = 68). We find a similar composition of the virome for the healthy reference samples and samples corresponding to minimal drug exposure (Figure 2B). However, the compositions of the healthy reference and minimal drug exposure samples are distinct from the anelloviridae-dominated composition measured for high drug exposure samples.

[00152] The tacrolimus-based immunosuppressive therapy is complemented with induction therapy in the first 3 days post transplant (with anti-thymocyte globulin, daclizumab, or basiliximab) and the patients furthermore receive the corticosteroid prednisone throughout the post transplant therapy. The time-dosage profile for prednisone and tacrolimus are similar: high doses at the onset of the therapy followed by a gradual dose reduction (Fig. 7A-C). The data in Fig. 2A thus reflect the combined effect of prednisone and tacrolimus. An analysis of the differential effect of prednisone and valganciclovir on the virome composition (Fig. 7E) shows the same trend observed in Fig. 2A: higher prednisone doses lead to a larger representation of anelloviruses. Lastly, we note that a subset of patients was not treated with antiviral drugs. The data corresponding to this subset of patients allowed us to further disentangle the differential effect of the antiviral drugs and the immunosuppressants on the composition of the virome, as described below.

[00153] Partitioning of microbiome diversity. We studied the diversity of the bacterial and viral components of the microbiome. The within-subject diversity was lower than the between-subject diversity, both for bacteria and viruses (Bray-Curtis beta diversity, bacterial composition at phylum level, family and order level viruses, Fig. 2C). Partitioning the data for patients according to transplant type, heart or lung, or age did not reduce the diversity. Within subjects, the diversity was lower for samples collected within a one month timespan, again both for bacteria and viruses. For viruses but not for bacteria, we find that the diversity is lower when comparing samples collected at a similar drug dosage (tacrolimus level ± 0.5 ng/ml, valganciclovir ± 50 mg). Taken together with the sensitivity of the population averages to drug

dosage in Fig. 2A, we thus find that the composition of the virome for patients that are subject to the same drug therapy converges to a similar state.

[00154] Dynamic response of virome to drug dose changes. A strong temporal response of the virome to changes in drug dosage is observed, consistent with the sensitivity of the virome composition to drug dosage. Figure 3A shows the time dependence of the relative genomic abundance of ssDNA and dsDNA viruses (data from all patient groups and samples, n = 656). The fraction of ssDNA viruses expands rapidly during the first months post transplant followed by the opposite trend after 6 months. Figure 3B shows the time-dependent relative composition of the most abundant viruses grouped at the family and order level and provides more detail on the virome compositional dynamics (data from all patient groups and samples, n = 656). The dsDNA fraction consists of caudovirales, adenoviridae, polyomaviridae and herpesvirales, which together occupy 95% of the virome in the first week(s) post transplant. ssDNA viruses only occupy 5% of the initial virome and mainly consist of members of the anelloviridae family. The fraction occupied by adenoviridae, caudovirales and herpesvirales decreases strongly in the first few months as these virotypes are effectively targeted by the antiviral prophylaxis. In contrast, the relative abundance of anelloviridae increases rapidly as these virotypes largely escape targeting by the antiviral drugs and take advantage of the reduced immunocompetence of the patients (maximum of 84% during months 4.5-6). Six months after the organ transplant procedure, the opposite trends are observed, consistent with the reduction in antiviral and immunosuppressant drugs prescribed by the therapeutic protocol.

[00155] Compared to the viral component, the bacterial component of the microbiome remains relatively stable over time, an observation that is made at the phylum, order and genus taxonomic levels (Fig. 3C, n = 656, and Fig. S3). Figure 3D shows the within-sample alpha diversity for the bacterial and viral genera as function of time (Shannon entropy, one month time periods, 590 bacterial genera, 168 viral genera examined). The diversity of observed viral genera decreases at the onset of the therapy (1.05 ± 0.5 in month 1 to 0.31 ± 0.33 in months 4-5, $p << 10^{-6}$, Mann-Whitney U test), whereas the alpha diversity of bacteria remains relatively unchanged during the course of the post transplant therapy (2.2 ± 1.14 in month 1 to 2.6 ± 0.85 in months 4-5, p = 0.1, Mann-Whitney U test).

[00156] Increase in total viral load at onset of post transplant therapy. To obtain insight into the effect of therapeutic drugs on total viral load, we extracted the absolute genomic abundance of all viruses relative to the number of human genome copies by normalizing the genome coverage of the viral targets to the coverage of the human genome. For all patient groups part of this study an increase in total viral load is observed at the onset of the therapy (Fig. 4A), regardless of transplant type (heart or lung) or age (adult or pediatric) (change in

load, 7.4 ± 3, sigmoid fit, black line). Combined with relative abundance data, the total viral load data reveals a net reduction of the Herpesvirales load and a net increase in anelloviridae load in the first 3 months post-transplant for patients that are simultaneously treated with antivirals and immunosuppressants.

[00157] The data thus show a differential effect of the combination of antivirals and immunosuppressants on different virotypes. The data also show a reduction in total adenoviridae load, indicating that adenoviridae replication is suppressed by valganciclovir, in agreement with previous studies. Figure 4B summarizes data for all transplant types, but the same trends are observed when stratifying according to different transplant types: adult heart transplant recipients (n = 268, Fig. 9A), adult lung transplant recipients (n = 166, Fig. 8B), and the pediatric patients that are treated with cyclosporine as opposed to tacrolimus (n = 99, Fig. 9C).

[00158] Not all patients in the study cohort received both antiviral and immunosuppressant drugs: for transplant cases where both the donor and recipient do not show evidence of a prior CMV infection in a CMV antibody assay, it is judged that the risks of complications due to antiviral prophylaxis outweigh the potential risk of a newly acquired CMV infection, and the patients are accordingly not treated with antiviral prophylaxis. These patients are thus solely treated with immunosuppressants. Figure 9C shows the time dependent viral load and composition of the CMV negative cases (n = 75). The net effect of immunosuppressant-only therapy is an expansion of all virotypes, including Herpesvirales and adenoviridae. Tapering of immunosuppression leads to a reduction of the total viral load.

[00159] Lower anellovirus burden in patients suffering from a graft rejection episode. Given the correlation of the anellovirus burden with the extent of immunosuppression (see figure 2A and Figure 4), and given the association between immunocompetence and the risk of rejection, we asked whether the anellovirus burden can be used for the classification of rejecting and non-rejecting graft recipients. Figure 5A shows the anellovirus load measured for rejecting and non-rejecting patients as function of time post transplant. Here, patients are classified as rejecting in case they suffer from at least one biopsy-determined moderate or severe rejection episode, biopsy grade ≥ 2R/3A (in red; 20 patients, 177 data points). The rejection-free patients correspond to patients that are not diagnosed to suffer from a moderate or severe graft damage throughout their post transplant course (in blue; biopsy grades < 2R/3A, 40 patients, 285 data points).

[00160] Figure 5A shows that the anellovirus burden is significantly lower for the rejecting individuals at almost every time point. We next directly compared the anellovirus burden for patients at rejection with the burden measured for patients in the absence of rejection. To account for the time dependence of the anellovirus load described above (Fig. 5A), we extract

the anellovirus load relative to the mean load measured for all samples at the same time point. Figure 5B shows the time-normalized load for non-rejecting patients (N = 208) compared to the load measured for patients suffering from a mild rejection event (biopsy grade 1R, N = 102) and patients suffering from a severe rejection episode (biopsy grade ≥ 2R/3A, N = 22). The figure shows that the time-normalized loads are significantly lower for the patients at greater risk of rejection. P-values were calculated by random sampling of the population with a greater amount of measurement points, $p = \text{sum}(\text{median}(A_{rej}) > \text{median}(A_{non\text{-}rej}))/N$, where $N = 10^4$ and A_{rej} and $A_{non\text{-}rej}$ are the relative viral loads for the populations at greater and lesser risk of rejection and non-rejecting respectively ($p = 0.011$, $p = 0.0002$ and $p = 0.036$).

[00161] These observations are in line with a view that the risk of rejection and the incidence of infection have an opposite association with the patients' immunocompetence (see inset Figure 5A). The lower viral load observed for rejecting patients is thus indicative of a higher level of immunocompetence in this subgroup of patients, even though these patients are treated with the same immunosuppressive protocol. Patient-to-patient variability in the sensitivity to suppression of immune function is known to occur and the lack of predictability in immunosuppression is an important risk factor in transplantation. A presently-used commercial assay for the measurement of immunocompetence has not been found to be predictive of acute rejection or significant infections. The development of methods for the direct measurement of immunocompetence, that can replace or complement existing assays, will therefore be important. The total anellovirus load recorded in organ transplant recipients could serve as an alternative marker. Figure 5C shows a receiver-operating characteristic and tests the performance of the relative anellovirus load in classifying non-rejecting and rejecting patients (area under the curve = 0.72).

[00162] We have studied drug-microbiome interactions following solid organ transplantation by sequencing cell-free DNA in the recipients' plasma. The data reveal much about the fundamental structure of the human virome in plasma and how it responds to pharmacological perturbation; they also show the relative insensitivity to immunosuppression of the composition of the bacterial component of the microbiome. These data are useful in the design and optimization of post-transplant therapeutic protocols. For example, they show that tapering of antiviral prophylaxis from initial high doses leads to a resurgence of the herpesvirales fraction. CMV DNA load has previously been shown to predict CMV disease relapse and rejection, raising the question of whether patients would benefit from longer-term prophylactic therapy.

[00163] The marked expansion in abundance of anelloviridae upon immunosuppression is also worth further consideration. Anelloviruses are ubiquitous in the human population and, although no pathogenicity has been established, anelloviruses are currently under

investigation as potential cofactors in carcinogenesis. The sensitivity of anelloviridae to immunosuppression makes organ transplantation an ideal setting for the study of the properties of anelloviridae, particularly in the light of the increased incidence of cancer seen in transplant recipients. The observation of a lower-than-average burden of anelloviruses in patients that suffer from a rejection episode is indicative of insufficient immunosuppression in this subgroup of patients, even though these patients were subject to the immunosuppressant levels prescribed per protocol. This suggests that there would be value in designing assays that allow directly gauging the level of a patient's immunocompetence, in addition to measurements of circulating drug levels. The total burden of anelloviruses identified in a transplant recipient's blood may serve as one such marker of the overall state of immunosuppression of the individual patient.

[00164] High throughput DNA sequencing finds use in the hypothesis-free diagnosis of infections. This approach is of particular relevance in the context of transplantation given the fact that infections occur frequently in transplantation and are difficult to diagnose in immunocompromised individuals, and given that sequence analysis can additionally provide information on the graft health through the quantification of donor-derived human DNA circulating in plasma. In other areas of infectious disease, it may be of value to develop subtractive methods to eliminate the human DNA and enrich for DNA of viral and microbial origin.

Experimental procedures

[00165] Clinical sample collection: Patients were enrolled at Stanford University Hospital (SUH) or Lucile Packard Children's Hospital (LPCH), and were excluded if they were recipients of multi-organ transplants. This study was approved by the Stanford University Institutional Review Board (protocol # 17666) and enrollment commenced in March 2010. For details on patient recruitment and post transplant treatment of the patients see the extended experimental procedures section.

[00166] Plasma processing and DNA extraction: Plasma was extracted from whole blood samples within three hours of sample collection, as previously described (Fan et al., 2008), and stored at -80°C. When required for analysis, plasma samples were thawed and circulating DNA was immediately extracted from 0.5-1 ml plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen).

[00167] Sequencing library preparation and sequencing: Sequencing libraries were prepared from the purified patient plasma DNA using the NEBNext DNA Library Prep Master Mix Set for Illumina with standard Illumina indexed adapters (purchased from IDT), or using a

microfluidics-based automated library preparation platform (Mondrian ST, Ovation SP Ultralow library system). Libraries were characterized using the Agilent 2100 Bioanalyzer (High sensitivity DNA kit) and quantified by qPCR. Samples were part of 26 different sequencing runs and were sequenced over the course of 22 months. On average 6 samples were sequenced per lane.

[00168] Posttransplant Monitoring and Clinical Sample Collection. This analysis represents a substudy of a prospective cohort study funded by the National Institutes of Health (RC4 AI092673) to study the clinical utility of a donor-derived cell-free DNA assay for the diagnosis of acute and chronic rejection and allograft failure after thoracic organ transplantation. Patients were enrolled if they received a heart or lung transplant at Stanford University Hospital (SUH) or Lucile Packard Children's Hospital (LPCH), and were excluded if they were recipients of multiorgan transplants or if they were followed at centers other than SUH or LPCH posttransplant. This study was approved by the Stanford University Institutional Review Board (protocol # 17666) and enrollment commenced in March 2010.

[00169] Details of the Posttransplant Therapeutic Protocol, Adult Heart Transplant Recipients. Posttransplant immunosuppression consisted of methylprednisolone 500 mg administered immediately postoperatively followed by 125 mg every 8 hr for three doses. Antithymocyte globulin (rATG) 1 mg/kg was administered on postoperative days 1, 2, and 3. Maintenance immunosuppression consisted of prednisone 20 mg twice daily starting on postoperative day 1 and tapered to < 0.1 mg/kg/day by the 6th postoperative month and tapered further if endomyocardial biopsies showed no evidence of cellular rejection. Tacrolimus was started on postoperative day 1 and dosing was further adjusted to maintain a level of 10–15 ng/ml during months 0–6, 7–10 ng/ml during months 6–12, and 5–10 ng/ml thereafter. Mycophenolatemofetil was started at 1,000 mg twice daily on postoperative day 1 and dose adjustments were made, if required, in response to leukopenia.

[00170] All patients received standard CMV (antiviral) prophylaxis consisting of ganciclovir 5 mg/kg IV, adjusted for renal function, every 12 hr starting on postoperative day 1 unless both donor and recipient were CMV negative. When able to tolerate oral medications, recipients were started on valganciclovir 900 mg twice daily for 2 weeks, then 900 mg daily until 6 months posttransplant, followed by 450 mg daily until 12 months posttransplant, at which point antiviral prophylaxis was discontinued. Valganciclovir dose reductions were made in the setting of leukopenia. CMV- recipients of a CMV+ allograft also received CMV hyperimmune globulin, 150 mg/kg IV, within 72 hr of transplant, 100 mg/kg at posttransplant weeks 2, 4, 6, and 8, and 50 mg/kg at weeks 12 and 16 posttransplant.

[00171] CMV$^-$ recipients of CMV$^-$ allografts were not treated with antiviral prophylaxis until May 2012; subsequently, these recipients were treated with acyclovir 400 mg twice daily for one year. Antifungal prophylaxis consisted of itraconazole 300 mg daily for the first 3 months posttransplant, and prophylaxis against pneumocystis jiroveci infection consisted of trimethoprim/sulfamethoxazole, 80 mg TMP component daily. Prophylaxis against pneumnocystis infection was continued indefinitely, and patients intolerant of TMP-SMX were treated with atovaquone, dapsone, or inhaled pentamidine.

[00172] All heart transplant recipients were monitored for acute cellular rejection by surveillance endomyocardial biopsies performed at scheduled intervals after transplant: weekly during the first month, biweekly until the 3rd month, monthly until the 6th month, and then at months 9, 12, 16, 20, and 24. Biopsies were graded according to the ISHLT 2004 revised grading scale (0, 1R, 2R, 3R) (29). Blood samples were collected from heart transplant recipients at the following time points posttransplant: weeks 2, 4, and 6; months 2, 2.5, 3, 4, 5, 6, 8, 10, 12, 16, 20, and 24. A subset of heart transplant recipients also had blood samples collected on posttransplant day 1. If blood sampling and endomyocardial biopsies were performed on the same day, care was taken to ensure that blood was collected prior to the biopsy procedure.

[00173] Pediatric Heart Transplant Recipients. Induction immunosuppression initially consisted of daclizumab 1 mg/kg IV every 2 weeks for a total of 5 doses, and was switched to basiliximab 10–20 mg IV on postoperative days 0 and 4 beginning in August 2011. Recipients were also treated immediately with pulse methylprednisolone 10 mg/kg IV every 8 hr for 3 doses, followed by prednisone 0.5 mg/kg twice daily for the first 14 days posttransplant; corticosteroids were subsequently tapered off during the first posttransplant year, in the absence of acute rejection.

[00174] Calcineurin inhibition consisted primarily of cyclosporine, with goal levels of 300–350 ng/ml for months 0–3, 275–325 ng/ml for months 4–6, 250–300 ng/ml months 7–12, and 200–250 after month 12 posttransplant. Patients intolerant of cyclosporine were treated with tacrolimus. Protocols for prophylaxis against opportunistic infections and surveillance endomyocardial biopsies were similar to adult heart transplant recipients.

[00175] Lung Transplant Recipients. Posttransplant immunosuppression consisted of methylprednisolone 500–1000 mg administered immediately postoperatively followed by 0.5 mg/kg IV twice daily. Basiliximab, 20 mg IV on days 0 and 4, was given for induction immunosuppression. Maintenance immunosuppression consisted of methylprednisolone 0.5 mg/kg IV twice daily on postoperative days 0–3, followed by prednisone 0.5 mg/kg daily until day 30, and subsequently tapered every 2–3 months to 0.1 mg/kg daily during months 6–12

posttransplant. Tacrolimus was started on postoperative day 0 and dosing was adjusted to maintain a level of 12–15 ng/ml during months 0–6, 10–15 ng/ml during months 6–12, and 5–10 ng/ml thereafter. Mycophenolatemofetil was initiated at 500 mg twice daily on postoperative day 0 and dose adjustments were made, if required, in response to leukopenia. Antiviral, antifungal, and PCP prophylaxis were similar to the adult heart transplant cohort.

[00176] All lung transplant recipients were monitored for acute cellular rejection by protocol transbronchial biopsies performed at months 1.5, 3, 6, 12, 18, and 24 posttransplant. Biopsies were also performed if indicated for clinical reasons, based on symptoms or pulmonary function test results. Blood samples were collected from lung transplant recipients for study purposes at the following intervals: thrice on day 1, twice on day 2, and once on day 3 posttransplant, followed by weeks 1 and 2, and months 1.5, 2, 3, 4.5, 6, 9, 12, 18, and 24. Blood samples were drawn prior to performance of per-protocol and clinically-indicated biopsies.

[00177] Workflow for the Identification of Pathogen-Derived Sequences. Exact duplicates were removed using the C-based utility fastq.cpp. Low-quality reads were removed using the quality filter that is part of the fastx package (fastq_quality_filter -Q33 -q21 - p50). The remaining reads were subsequently aligned using BWA to the human reference genome build hg19 (bwaaln - q25). Unmapped reads were collected using samtools (samtools view -f4) and low-complexity reads were removed using Seqclean (seqclean -l 40 -c 1). Reads were subsequently aligned to a selection of viral, baterial and fungal reference genomes and all references in ncbi_fungi downloaded.

[00178] Figure 6A shows the distribution of the genome sizes. The following parameters were used for the BLAST alignment: reward = 1, penalty = _3, word_size = 12, gapopen = 5, gapextend = 2, e-value = 10_{-4}, perc_identity = 90, culling_limit = 2. Blast hits with alignment length shorter than 45 were removed. For a subset of samples longer reads were available (2 3 100 bp, n = 55). To test the robustness of the genomic abundance estimates, the length dependence of the composition measurement was examined. Here, reads were trimmed to 40, 50, 65, 80 and 100 bp lengths (fastx_trimmer) and analyzed using the above-described workflow. Here the blast hits with alignment lengths shorter than 37, 45, 59, 72, and 80 bp were removed for the 40, 50, 65, 80 and 100 bp reads respectively. Genome Abundance Estimation Relative genome abundance estimation was calculated with GRAMMy. This tool utilizes the BLAST-derived nucleic acid sequence-similarity data to perform a maximum likelihood estimation of the relative abundance of species in the sample. GRAMMy filters hits by BLAST alignment metrics (E-score, alignment length and identity rate) and accounts for the target genome size and the ambiguity of read assignments in assessing the relative abundance of the candidate reference genomes. Grammy was called using following

parameters: python grammy_rdt.py; python grammy_pre.py -q "40,40,1" input set; python grammy_em.py -b 5 -t 0.0001 -n 100 input.mtx; grammy_post.py input.est setinput.btp.

[00179] Custom scripts were used to combine the strain-level abundance estimates to obtain the abundance at higher taxonomic-level abundances. Here, a minimal taxonomy for the reference database was built using Taxtastic.

[00180] Quantification Absolute Viral Load. To quantify the load of infectious agents in the samples the blast hit results were collected and the best hits selected for each read using a custom script (Bioperl). Figure 6B shows the distribution of the number of unique viral, bacterial and fungal blast hits per million unique molecules sequenced. Figure 6C shows the number of viral, bacterial and fungal genome copies relative to the number of human genome copies present in the sample. The coverage of the genome of the infectious agent was normalized with respect to the human genome coverage.

[00181] qPCR Validation of Sequencing Results for Selected Viral Targets. Standard qPCR kits for the quantification of Human Herpes Virus 4, 5 and 6 and parvovirus (PrimerDesign, genesig) were used to validate the sequencing results for a subset of cell-free DNA samples. qPCR assays were run on cfDNA extracted from ~1 ml of plasma and eluted in a 100 ml Tris buffer (50 mM [pH 8.1–8.2]). The plasma extraction and PCR experiments were performed in different facilities. No-template controls were ran to verify that the PCR reagents were included in every experiment. Figure 6D compares the relative number of blast hits per million reads acquired to the concentration of viral genome copies as determined using qPCR.

[00182] No-Template Control. A no-template control experiment was performed. A sequencing library was generated from nuclease-free water (S01001, Nugen). The library was prepared together with 7 additional sample libraries (cell-free human DNA) to test for possible sample-to-sample crosstalk during library preparation. To ensure formation of clusters with sufficient density on the Illumina flow cell, the sample was sequenced together with a sample unrelated to the study. Whereas the sample unrelated to the study recruited 16 million reads, the no-template control library generated just 15 reads that mapped to two species in the reference database, the methanocalcodoccus janaschii (9 hits) and Bacillus subtillis (5 hits) genomes. No evidence was found for human related sequences, indicating that sample-to-sample contamination was low.

Example 2

Clinical monitoring of the Microbiome

[00183] Using the methods as described in Example, 1 reads that map to the CMV genome were quantified for each sample. An increased CMV abundance was observed in samples that were clinically positive for infection ($p=7.10^{-9}$, Mann-Whitney U test, Fig. 10C); the level of CMV-derived DNA in our samples matched clinical reports of CMV with an AUC of 0.91 (Fig.

10C). This data indicates that CMV surveillance can be performed in parallel with rejection monitoring using the same sequence data, and led us to examine whether other viral infections could be similarly monitored.

[00184] We identified well-characterized pathogenic and onco-viruses (Fig 11A) as well as commensal torque teno viruses (TTVs, *alphatorquevirus* genus), which is consistent with previous observations of a link between immunosuppression and TTV abundance. The frequency of clinical testing for these viruses varied considerably, with frequent surveillance of CMV (Human Herpes Virus 5, HHV-5, n=1082 tests in our cohort) relative to other pathogens (Fig 11A). We evaluated the incidence of infection (number of samples in which a given virus is detected via sequencing) relative to the clinical screening frequency. Although CMV was most frequently screened for (335 samples), its incidence as determined by sequencing (detected in 22 samples) was similar to that of other pathogens that were not routinely screened, including adenovirus and polyomavirus (clinically tested on four occasions and one occasion, respectively, Fig 11A).

[00185] Adenovirus is a community-acquired respiratory infection that can cause graft loss in lung transplant recipients and poses a particularly high risk for paediatric patients. Samples were collected from one paediatric patient (L78, Fig 11B panel 1) that tested positive for adenovirus. This patient also had the highest adenovirus-derived DNA load across the entire cohort. A sustained adenovirus load was furthermore observed in several other adult transplant patients that were not screened clinically (e.g, L34, Fig 11B panel 1), as testing is typically restricted to paediatric lung transplant cases.

[00186] Polyomavirus is the leading cause of allograft rejection after renal transplantation but is not routinely included in post-lung transplant surveillance. We detected polyomavirus in two patients that were not tested for this pathogen (L57 and L15, Fig 11B panel 2). In both cases, the clinical records indicated persistent renal insufficiency, which may have resulted from polyomavirus infection.

[00187] In a last example of the benefit of broad and hypothesis-free screening of infections, we examined a patient that exhibited a high load of human herpes virus (HHV) 8 (Fig 11B panel 3), an oncovirus that can cause complications following solid-organ transplantation. This patient (L58) tested positive for two other herpesviruses (HHV-4a and HHV-5), which have the potential to stimulate HHV-8 re-activation. Though post-transplant monitoring for HHV-8 is only recommended in particular clinical circumstances, use of sequencing enables the identification of the virus in non-suspect cases that would otherwise go undetected.

[00188] *Clinical monitoring of the microbiome*. In addition to viruses measured in serum, we also observed correlation between cell-free measurements and fungal or bacterial infections detected in other body fluids, including *Klebsiella pneumonia* infections detected via urine culture (ROC = 0.98) and fungal infection detected in BAL. Performance on bacterial and

fungal correlations was sensitive both to the infection type and body fluid queried. We observed better performance for body fluids that have tighter coupling to blood and also observed sensitivity to background signal. For example, the most commonly cultured bacterial infection (*Pseudomonas*) was detected in cell-free measurements for over 80% of our patient samples, which was in stark contrast to the most commonly detected viral pathogenic species (*CMV*), detected in only 6% of our patient samples.

[00189] This highlights an important distinction between commensal infections (including *Pseudomonas*), which are part of the normal flora, and non-commensal infections, which are solely pathogenic and have a lower background signal. This distinction may explain differences in sensitivity and specificity measured for commonly cultured commensal infections (e.g., AUC=0.66 and 0.62 for *P. aeruginosa* and *E. coli*, respectively) relative to non-commensals (AUC=0.91 for CMV). In the case of commensal bacteria, the clinical question is not presence or absence but rather presence or absence in inappropriate body sites.

[00190] In our cohort, we also detected cell-free DNA derived from microsporidia, a non-commensal fungus that can cause intestinal infections in immunosuppressed patients. We measured a sustained microsporidia load in L78 (Fig 11B panel 4), a patient that exhibited canonical symptoms of microsporidiosis. An Adenovirus infection (L78, Fig 11B panel 1) was the suspected cause, though endoscopy and sigmoidoscopy results were inconclusive and stool samples tested negative for C. diff as well as Adenovirus. Based upon our sequencing data, the microsporidiosis is the most likely explanation for the patient's symptoms, as the microsporidia signal measured in this patient is similar to that of I6, a patient from an unrelated cohort that tested positive for microsporidia (Fig 11B panel 4).

[00191] With more than 10 billion fragments per ml of plasma, circulating cell-free DNA is an information-rich window into human physiology, with rapidly expanding applications in cancer diagnosis and cancer treatment monitoring, genetic prenatal diagnosis, and monitoring of heart transplant rejection via "genome transplant dynamics" (GTD). In this work, we applied the principle of GTD to lung transplantation--a particularly challenging type of solid organ transplant that is limited by poor survival rates, as well as an inaccurate and invasive test for allograft rejection.

[00192] Because lung transplant recipients with allograft infection and acute rejection may present clinically with similar symptoms, we extended the scope of GTD to infectious disease monitoring. We first demonstrated a strong correlation between clinical test results and cfDNA derived from CMV--a leading cause of post-transplant graft injury. We further showed that hypothesis-free infection monitoring revealed numerous un-tested pathogens, including un-diagnosed cases of adenovirus, polyomavirus, HHV-8, and microsporidia in patients who had

similar microbial cfDNA levels compared to patients with positive clinical test results and associated symptoms. These examples illustrate the benefit of broad, sequencing-based monitoring of infection as opposed to pathogen specific testing. The approach can be of immediate use as tool that can assist in determining the occurrence and source of an infection. This may be of particular relevance in the context of transplantation, where the incidence of infections is high, where rejection and infection can co-occur, and where the symptoms of infection and rejection are difficult to discriminate.

[00193] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

2014346562 15 Nov 2018

WHAT IS CLAIMED IS:

1. A method of detecting at least one bacterium, fungus, or parasite in a sample comprising cell-free nucleic acids from a non-microbial host, the method comprising:

(i) providing said sample comprising cell-free nucleic acids from said non-microbial host, wherein said cell-free nucleic acids comprise microbial cell-free nucleic acids;

(ii) performing high-throughput sequencing of said cell-free nucleic acids to generate sequence reads that comprise sequence reads from said microbial cell-free nucleic acids; and

(iii) detecting said at least one bacterium, fungus, or parasite at a strain or species level in said sample.

2. The method of claim 1, wherein a plurality of bacteria, fungi, or parasites is detected.

3. The method of claim 1 or claim 2, wherein said high throughput sequencing is performed on cell-free nucleic acids that have been amplified by a non-biased method.

4. The method of any one of claims 1 to 3, wherein at least 10^6 sequence reads are performed.

5. The method of any one of claims 1 to 4, wherein (iii) includes comparing coverage of sequences mapping to a microbial reference sequence to coverage of a host reference sequence.

6. The method of any one of claims 1 to 5, further comprising performing bioinformatics analysis on said sequence reads from said microbial cell-free nucleic acids.

7. The method of claim 6, wherein said performing bioinformatics analysis includes identifying a reference host sequence, and masking microbial sequences or microbial mimicking sequences present in said reference host genome.

8. The method of claim 6, wherein said performing bioinformatics analysis includes identifying a reference microbial sequence, and masking host sequences or host mimicking sequences present in said reference microbial genome.

9. The method of any one of claims 1 to 8, wherein said at least one bacterium, fungus, or parasite detected in (iii) is pathogenic.

10. The method of any one of claims 1 to 9, wherein said detecting is performed at two or more time points.

11. The method of any one of claims 1 to 10, wherein an amount of said at least one bacterium, fungus, or parasite is indicative of infection status or treatment outcome.

12. The method of claim 11, wherein an amount of said at least one bacterium, fungus, or parasite above a predetermined threshold value is indicative of an infection status or treatment outcome.

13. The method of any one of claims 1 to 12, wherein said sample is selected from the group consisting of blood, serum, urine, and stool.

14. The method of any one of claims 1 to 13, wherein said microbial cell-free nucleic acids comprise at least one nucleic acid selected from the group consisting of double-stranded DNA, single-stranded DNA, single-stranded DNA hairpins, DNA RNA hybrids, single-stranded RNA, double-stranded RNA and RNA hairpins.

15. The method of any one of claims 1 to 13, wherein said microbial cell-free nucleic acids comprise at least one nucleic acid selected from the group consisting of double-stranded DNA, single-stranded DNA and cDNA.

16. The method of any one of claims 1 to 15, wherein said non-microbial host is human.

17. The method of any one of claims 1 to 16, wherein said non-microbial host is human, and further comprising providing an assessment of said sample to said human non-microbial host.

18. The method of any one of claims 1 to 17, wherein said detecting provides a determination of response to therapy.

19. The method of any one of claims 1 to 17, wherein said detecting provides a measurement of human physiology.

20. The method of any one of claims 1 to 17, wherein said detecting is used to calculate a pathogenicity score for microorganisms present in said non-microbial host.

21. The method of any one of claims 1 to 20, wherein said non-microbial host has received a transplant.

22. The method of claim 21 wherein said transplant is selected from the group consisting of bone marrow transplant, kidney transplant, heart transplant, liver transplant, pancreas transplant, lung transplant, intestine transplant and skin transplant.

23. The method of any one of claims 1 to 17, wherein an amount of said at least one bacterium, fungus, or parasite is indicative of transplant status or outcome.

24. The method of claim 23, wherein an amount of said at least one bacterium, fungus, or parasite above a predetermined threshold value is indicative of a transplant status or outcome.

25. The method of claim 24, wherein said threshold is a normative value for clinically stable post-transplantation patients with no evidence of transplant rejection or other pathologies.

26. The method of claim 24, wherein there are different predetermined threshold values for different transplant outcomes or status.

27. The method of any one of claims 1 to 26, further comprising treating said non-microbial host in accordance with said detection of said at least one bacterium, fungus, or parasite.

28. The method of any one of claims 1 to 27, wherein said-detecting is determined by a method selected from the group consisting of sequencing, nucleic acid array and PCR.

29. The method of any one of claims 1 to 28, further comprising identifying said at least one bacterium at said strain or species level.

30. The method of any one of claims 1 to 28, further comprising identifying said at least one fungus at said strain or species level.

31. The method of any one of claims 1 to 28, further comprising identifying said at least one parasite at said strain or species level.

32. The method of any one of claims 1 to 31, further comprising attaching adapters to said cell-free microbial nucleic acids prior to (ii) to produce adapted cell-free microbial nucleic acids.

33. The method of any one of claims 1 to 32, wherein a second sample is collected from said non-microbial host at a different time point and analyzed.

34. A computer readable medium comprising:

a set of instructions recorded thereon that when executed cause a computer to perform:

(i) receiving high throughput data from cell-free nucleic acids detected in a sample from a non-microbial host, wherein said high throughput data comprises sequence reads; and

(ii) using sequence reads from microbial cell-free nucleic acids to detect at least one bacterium, fungus, or parasite at a strain or species level in said sample comprising cell-free nucleic acids from said non-microbial host.

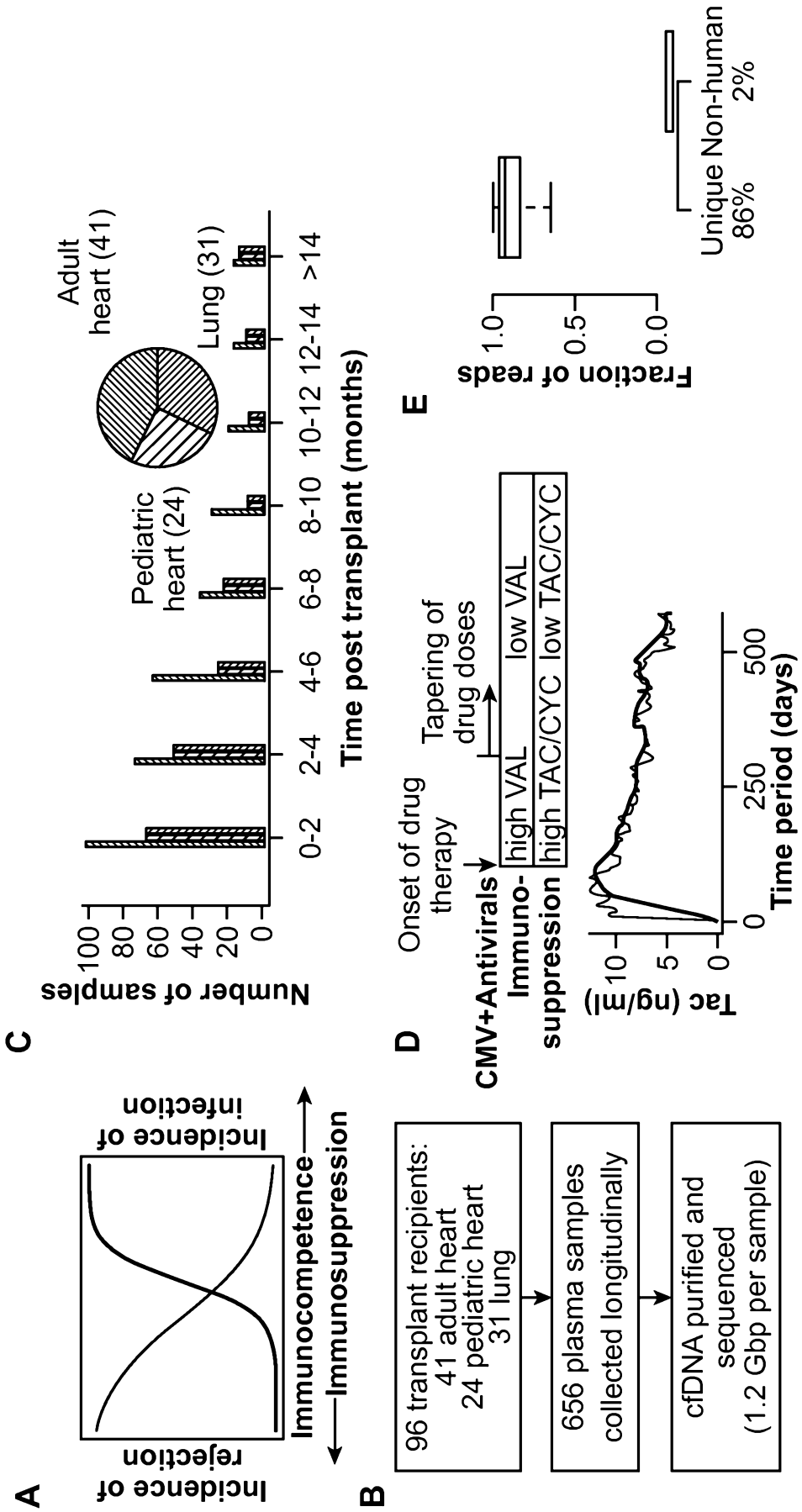
A
Incidence of rejection
Immunocompetence
Immunosuppression
Incidence of infection
B
96 transplant recipients:
41 adult heart
24 pediatric heart
31 lung
656 plasma samples collected longitudinally
cfDNA purified and sequenced
(1.2 Gbp per sample)
C
Number of samples
100
80
60
40
20
0
0-2
2-4
4-6
6-8
8-10
10-12
12-14
>14
Time post transplant (months)
Adult heart (41)
Pediatric heart (24)
Lung (31)
D
Onset of drug therapy
Tapering of drug doses
CMV+Antivirals
Immuno-suppression
high VAL
low VAL
high TAC/CYC
low TAC/CYC
Tac (ng/ml)
10
5
0
0
250
500
Time period (days)
E
Fraction of reads
1.0
0.5
0.0
Unique 86%
Non-human 2%

FIG. 1

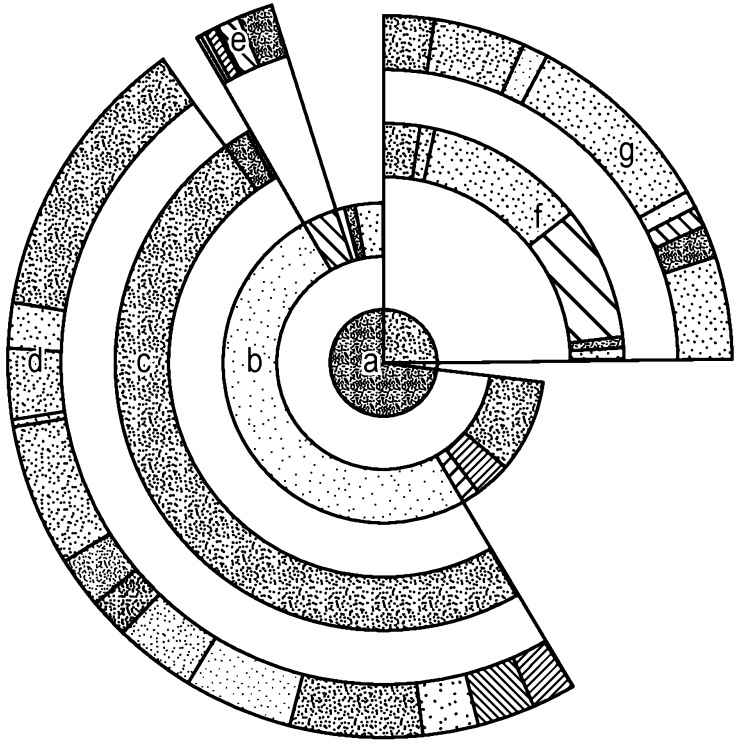

FIG. 1 (Cont.)

F

a: Superkingdom
- Bacteria 25%
- Eukaryota 2%
- Viruses 73%

b: Viruses: order and family
- Herpesvirales 13%
- Caudovirales 5%
- Adenoviridae 2%
- Anelloviridae 68%
- Polyomaviridae 5%
- Poxviridae 1%
- Retroviridae 1%

c: Anelloviridae: genera
- Alphatorquevirus 97%
- Betatorquevirus 3%

d: Alphatorque viruses
- TTV1 3%
- TTV10 5%
- TTV12 5%
- TTV14 <1%
- TTV15 12%
- TTV16 10%
- TTV19 7%
- TTV27 4%
- TTV28 5%
- TTV3 13%
- TTV4 1%
- TTV6 6%
- TTV7 4%
- TT8 26%

e: Polyoma viruses
- WU Polyomavirus 6%
- SV40 6%
- Polyomavirus HPyV6 13%
- TS associated polyomavirus 4%
- JC Polyomavirus 27%
- Bk Polyomavirus 41%

f: Bacteria: phyla
- Actinobacteria 10%
- Bacteroidetes 4%
- Firmicutes 43%
- Proteobacteria 36%
- Tenericutes 3%

g: Bacteria order
- Actinomycetales 10%
- Burkholderiales 17%
- Enterobacteriales 5%
- Lactobacillales 36%
- Mycoplasmatales 3%
- Pseudomonadales 4%
- Selenomonadales 6%
- Other

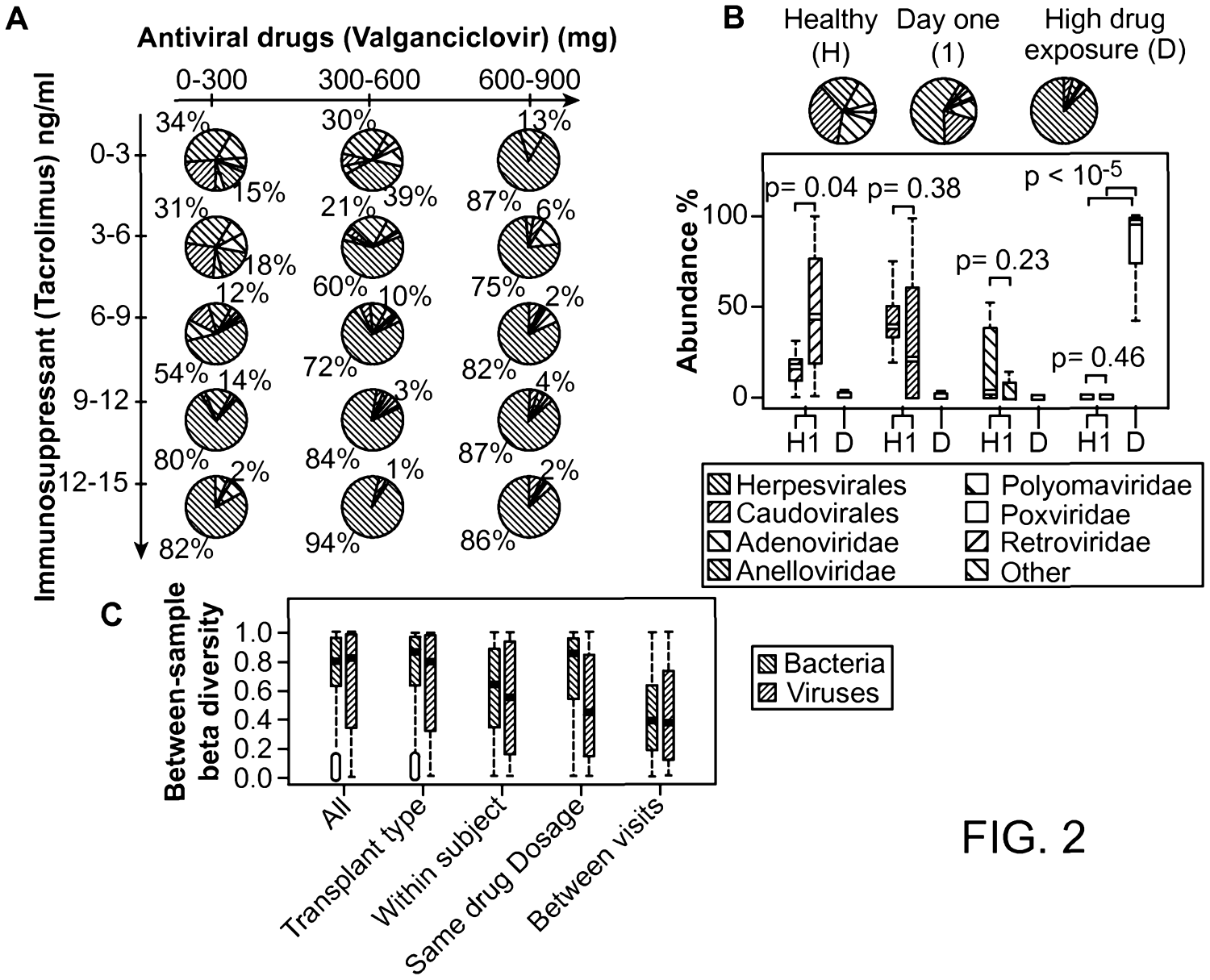
A
Antiviral drugs (Valganciclovir) (mg)
0-300
300-600
600-900
Immunosuppressant (Tacrolimus) ng/ml
0-3
3-6
6-9
9-12
12-15
34%
15%
30%
39%
13%
87%
31%
18%
21%
6%
12%
60%
10%
75%
2%
54%
14%
72%
3%
82%
4%
80%
2%
84%
1%
87%
2%
82%
94%
86%
B
Healthy (H)
Day one (1)
High drug exposure (D)
Abundance %
100
50
0
p= 0.04
p= 0.38
p < 10-5
p= 0.23
p= 0.46
H1 D H1 D H1 D H1 D
Herpesvirales
Caudovirales
Adenoviridae
Anelloviridae
Polyomaviridae
Poxviridae
Retroviridae
Other
C
Between-sample beta diversity
1.0
0.8
0.6
0.4
0.2
0.0
All
Transplant type
Within subject
Same drug Dosage
Between visits
Bacteria
Viruses

FIG. 2

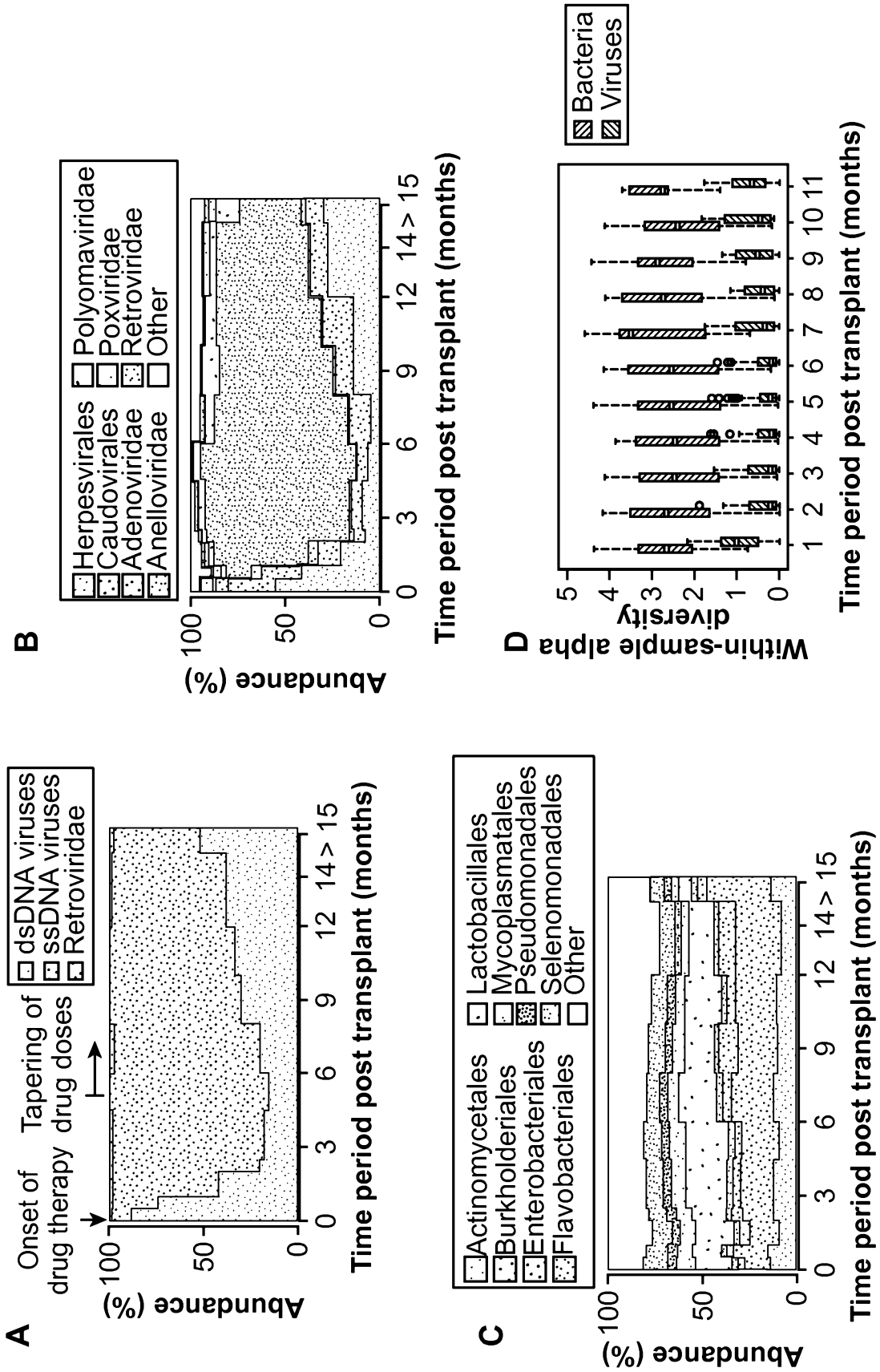
A
Onset of drug therapy
Tapering of drug doses
dsDNA viruses
ssDNA viruses
Retroviridae
Abundance (%)
100
50
0
0 3 6 9 12 14 > 15
Time period post transplant (months)
B
Herpesvirales
Caudovirales
Adenoviridae
Anelloviridae
Polyomaviridae
Poxviridae
Retroviridae
Other
Abundance (%)
100
50
0
0 3 6 9 12 14 > 15
Time period post transplant (months)
C
Actinomycetales
Burkholderiales
Enterobacteriales
Flavobacteriales
Lactobacillales
Mycoplasmatales
Pseudomonadales
Selenomonadales
Other
Abundance (%)
100
50
0
0 3 6 9 12 14 > 15
Time period post transplant (months)
D
Bacteria
Viruses
Within-sample alpha diversity
5 4 3 2 1 0
1 2 3 4 5 6 7 8 9 10 11
Time period post transplant (months)

FIG. 3

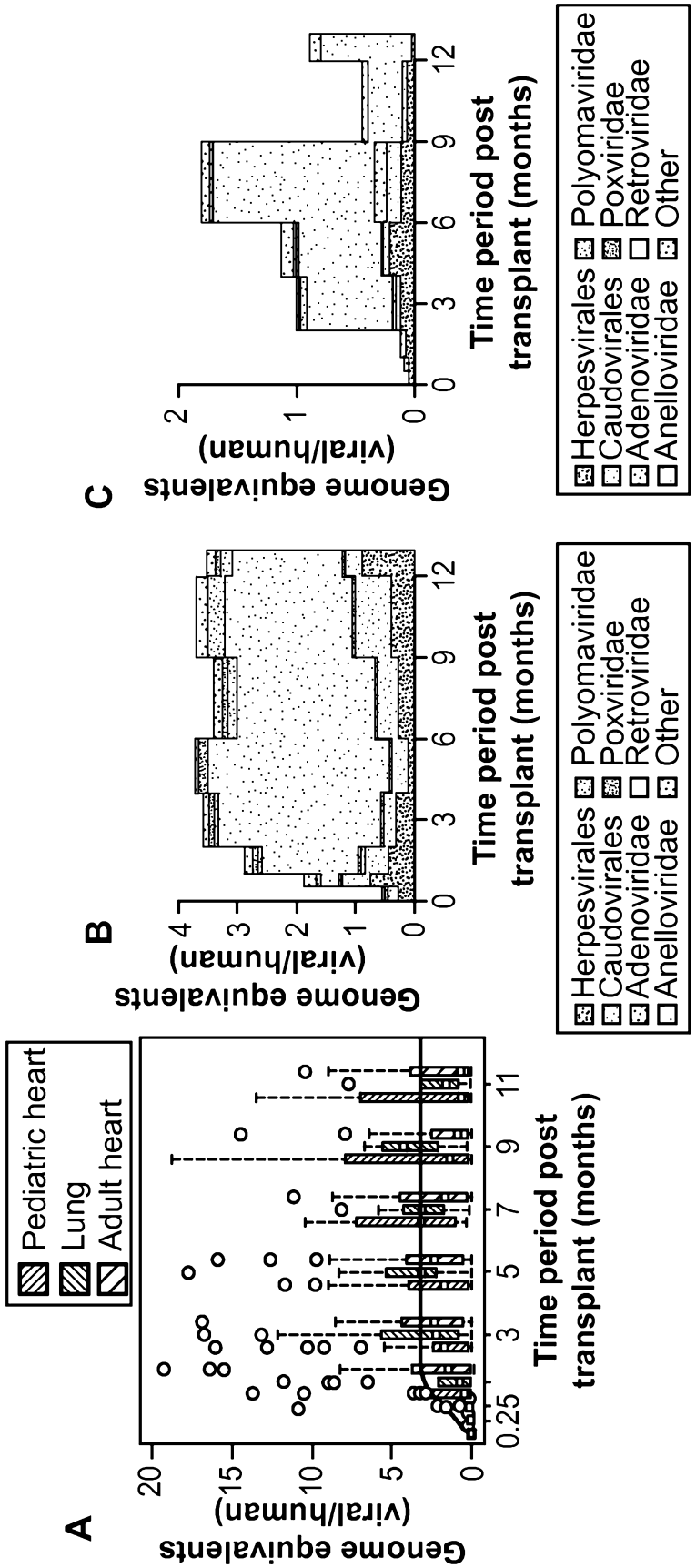
A
Genome equivalents (viral/human)
20
15
10
5
0
0.25
3
5
7
9
11
Time period post transplant (months)
Pediatric heart
Lung
Adult heart
B
Genome equivalents (viral/human)
4
3
2
1
0
0
3
6
9
12
Time period post transplant (months)
Herpesvirales
Polyomaviridae
Caudovirales
Poxviridae
Adenoviridae
Retroviridae
Anelloviridae
Other
C
Genome equivalents (viral/human)
2
1
0
0
3
6
9
12
Time period post transplant (months)
Herpesvirales
Polyomaviridae
Caudovirales
Poxviridae
Adenoviridae
Retroviridae
Anelloviridae
Other

FIG. 4

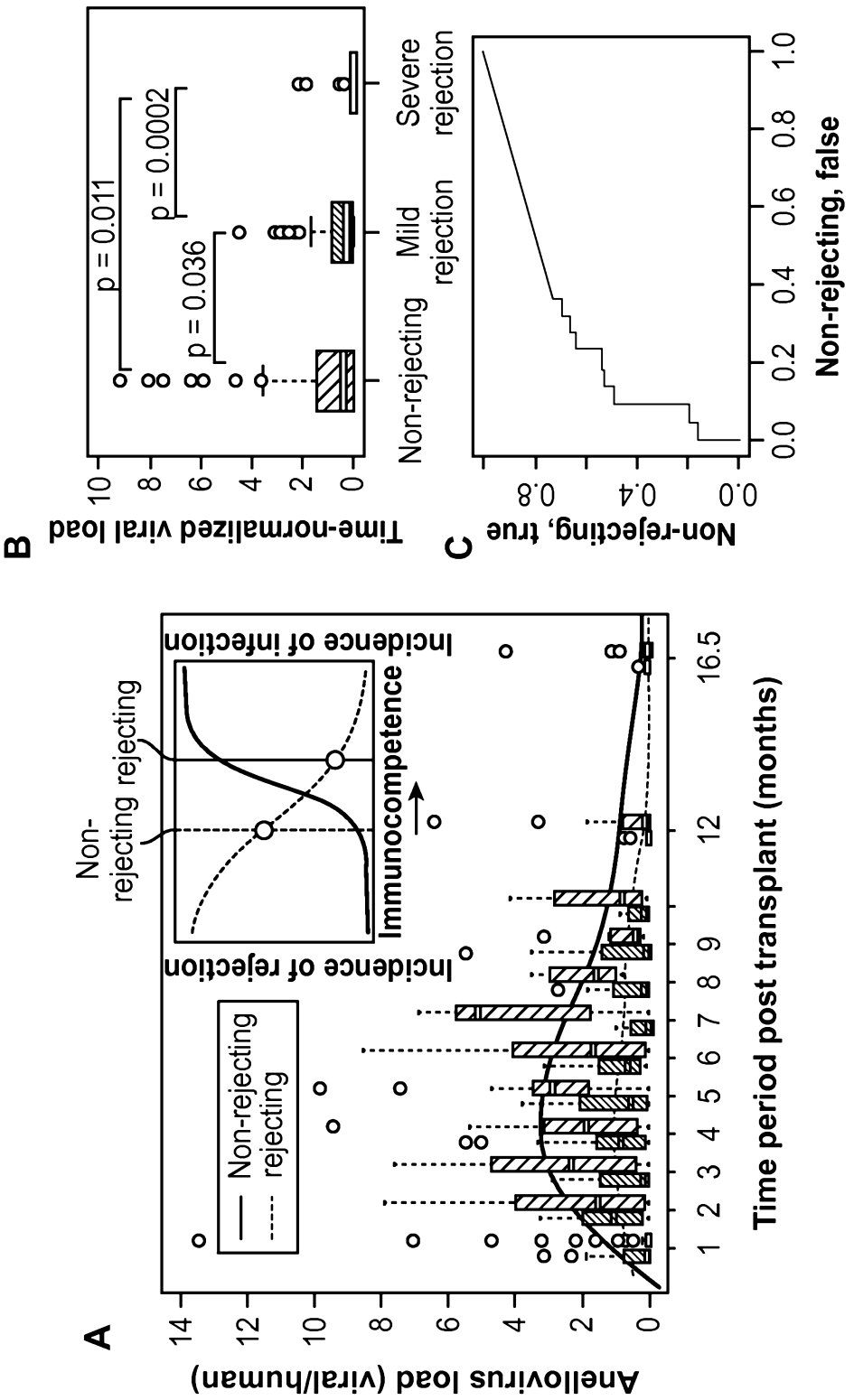
A
Anellovirus load (viral/human)
14
12
10
8
6
4
2
0
Non-rejecting
rejecting
Non-rejecting rejecting
Incidence of rejection
Immunocompetence
Incidence of infection
1 2 3 4 5 6 7 8 9 12 16.5
Time period post transplant (months)
B
Time-normalized viral load
10
8
6
4
2
0
p = 0.011
p = 0.036
p = 0.0002
Non-rejecting
Mild rejection
Severe rejection
C
Non-rejecting, true
0.0
0.4
0.8
0.0 0.2 0.4 0.6 0.8 1.0
Non-rejecting, false

FIG. 5

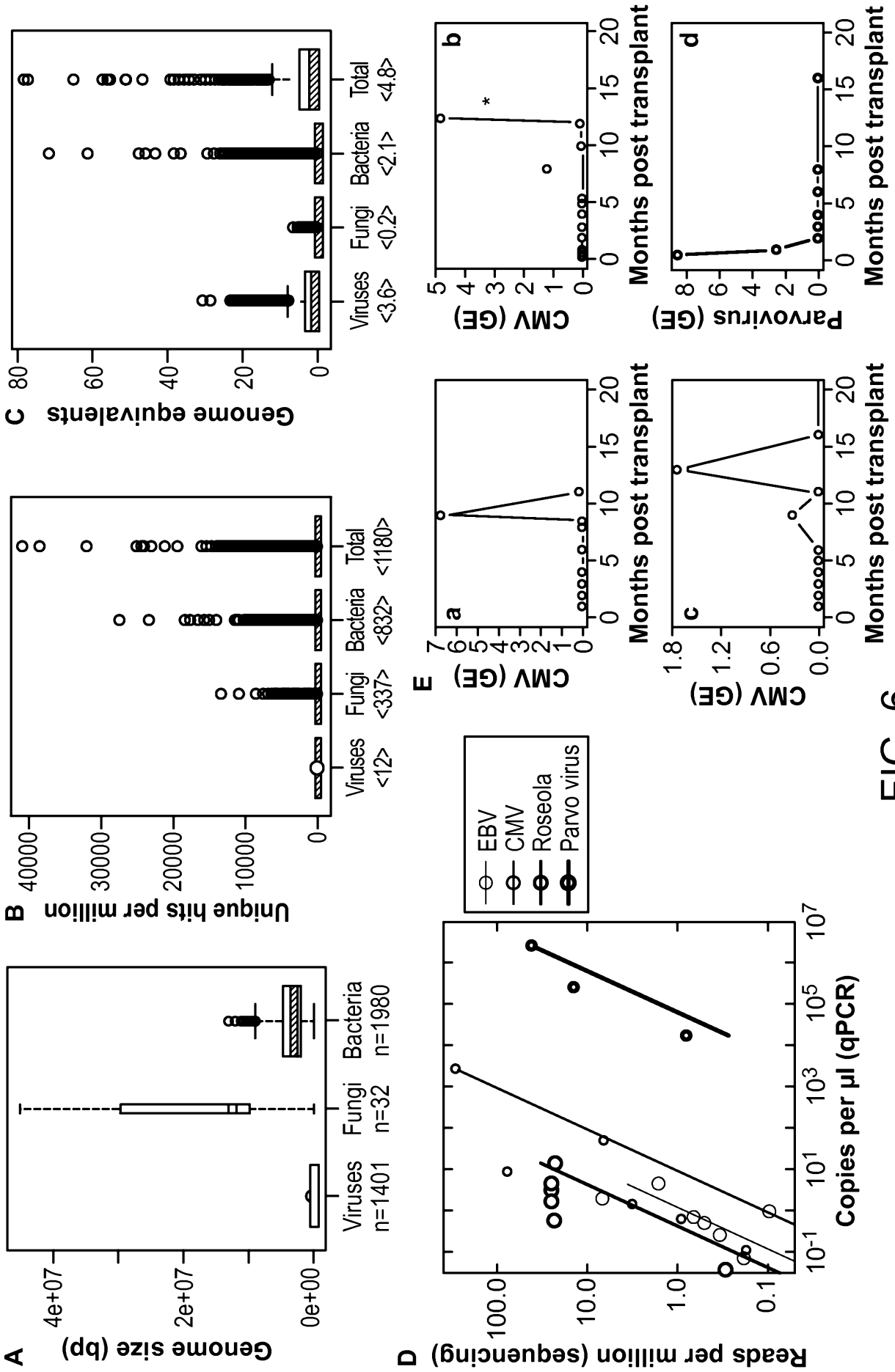
A
Genome size (bp)
4e+07
2e+07
0e+00
Viruses n=1401
Fungi n=32
Bacteria n=1980
B
Unique hits per million
40000
30000
20000
10000
0
Viruses <12>
Fungi <337>
Bacteria <832>
Total <1180>
C
Genome equivalents
80
60
40
20
0
Viruses <3.6>
Fungi <0.2>
Bacteria <2.1>
Total <4.8>
D
Reads per million (sequencing)
100.0
10.0
1.0
0.1
Copies per µl (qPCR)
EBV
CMV
Roseola
Parvo virus
E
a
b
c
d
CMV (GE)
Parvovirus (GE)
Months post transplant

FIG. 6

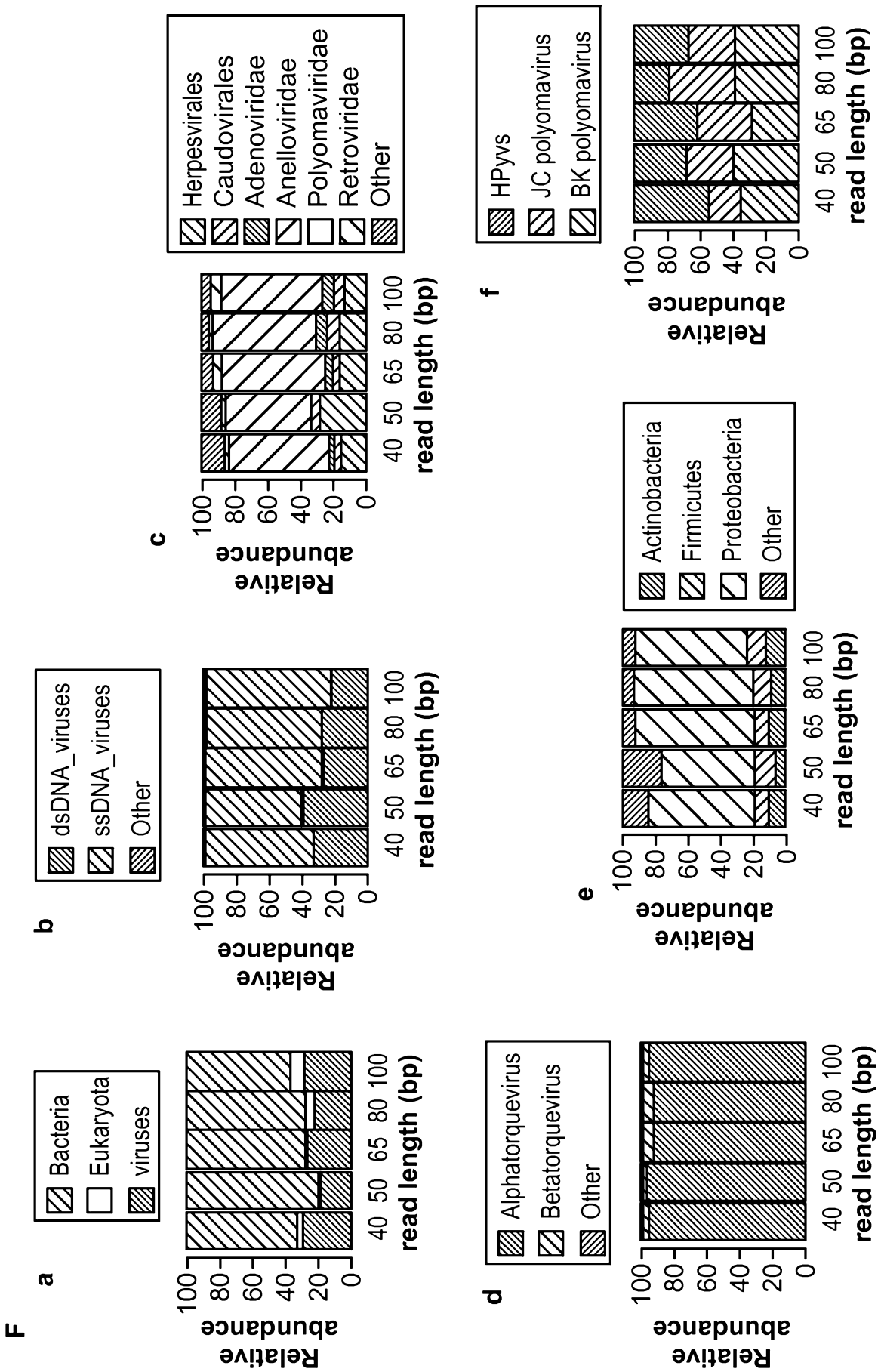
F
a
Bacteria
Eukaryota
viruses
Relative abundance
read length (bp)
b
dsDNA_viruses
ssDNA_viruses
Other
c
Herpesvirales
Caudovirales
Adenoviridae
Anelloviridae
Polyomaviridae
Retroviridae
Other
d
Alphatorquevirus
Betatorquevirus
Other
e
Actinobacteria
Firmicutes
Proteobacteria
Other
f
HPyvs
JC polyomavirus
BK polyomavirus
0 20 40 60 80 100
40 50 65 80 100

FIG. 6 (Cont.)

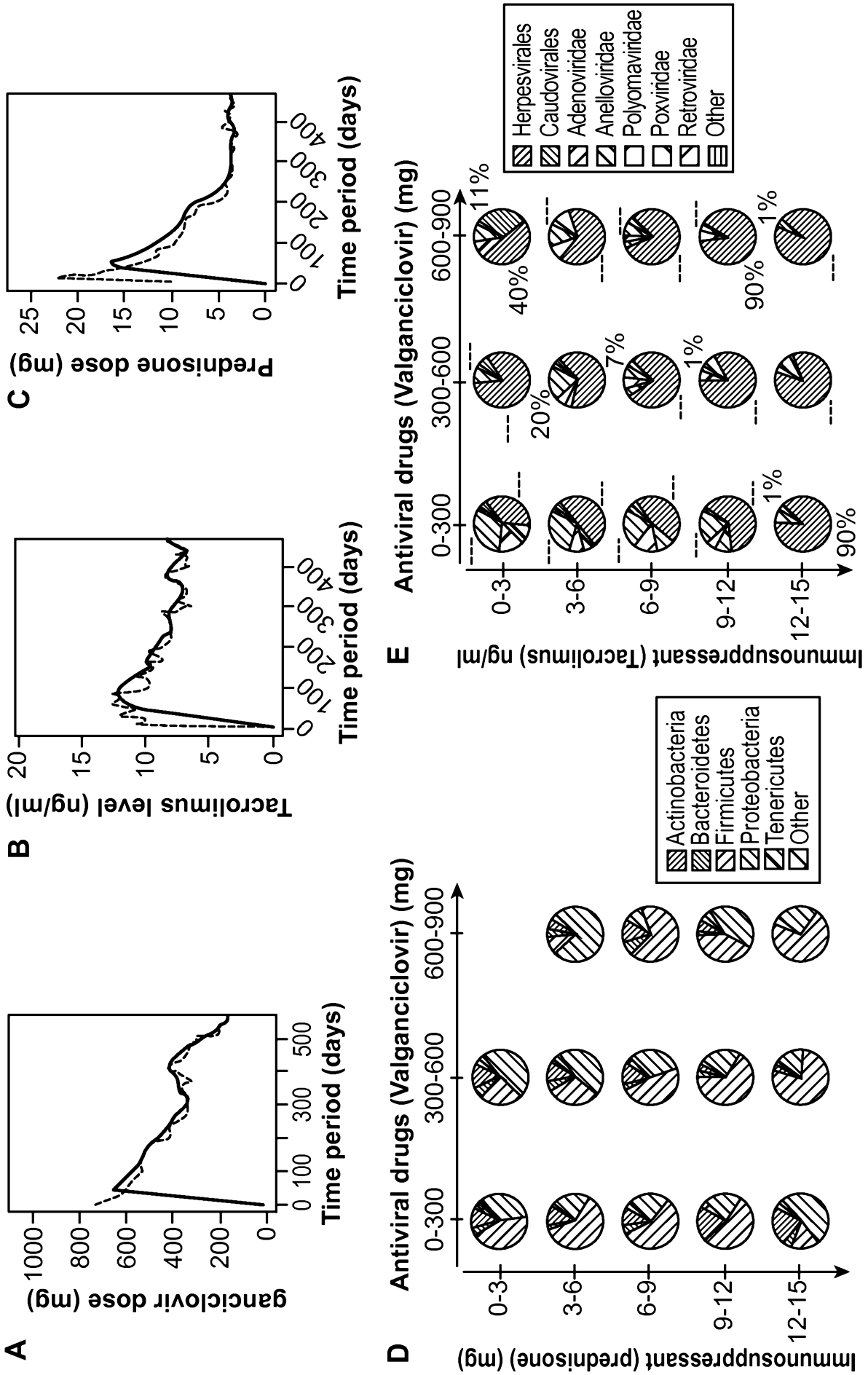
A
ganciclovir dose (mg)
1000
800
600
400
200
0
0 100 300 500
Time period (days)
B
Tacrolimus level (ng/ml)
20
15
10
5
0
0 100 200 300 400
Time period (days)
C
Prednisone dose (mg)
25
20
15
10
5
0
0 100 200 300 400
Time period (days)
D
Antiviral drugs (Valganciclovir) (mg)
0-300
300-600
600-900
Immunosuppressant (prednisone) (mg)
0-3
3-6
6-9
9-12
12-15
Actinobacteria
Bacteroidetes
Firmicutes
Proteobacteria
Tenericutes
Other
E
Antiviral drugs (Valganciclovir) (mg)
0-300
300-600
600-900
Immunosuppressant (Tacrolimus) ng/ml
0-3
3-6
6-9
9-12
12-15
40%
11%
20%
7%
1%
90%
1%
90%
1%
Herpesvirales
Caudovirales
Adenoviridae
Anelloviridae
Polyomaviridae
Poxviridae
Retroviridae
Other

FIG. 7

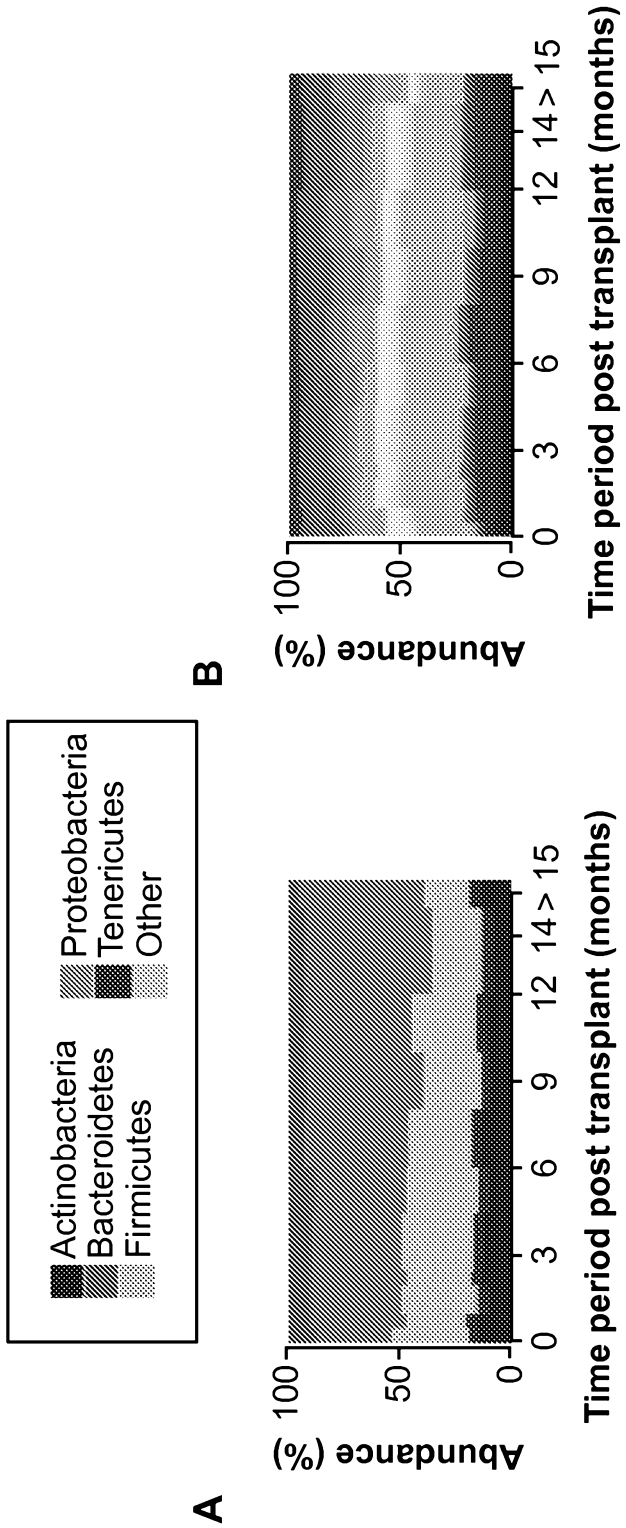
Actinobacteria
Bacteroidetes
Firmicutes
Proteobacteria
Tenericutes
Other
A
B
Abundance (%)
100
50
0
0
3
6
9
12
14 > 15
Time period post transplant (months)

FIG. 8

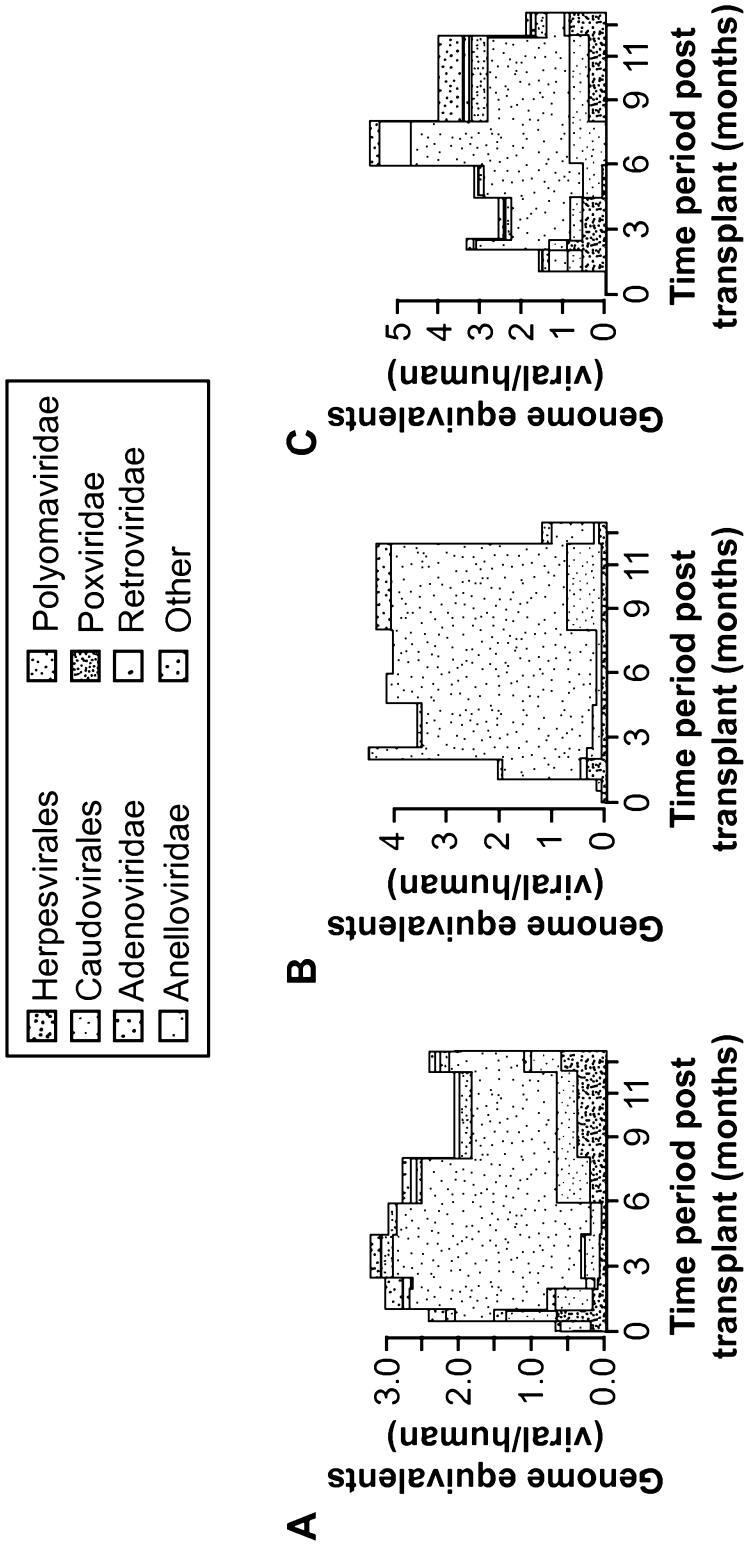
Herpesvirales
Caudovirales
Adenoviridae
Anelloviridae
Polyomaviridae
Poxviridae
Retroviridae
Other
A
B
C
Genome equivalents (viral/human)
Time period post transplant (months)
0
3
6
9
11
0.0
1.0
2.0
3.0
1
2
3
4
5

FIG. 9

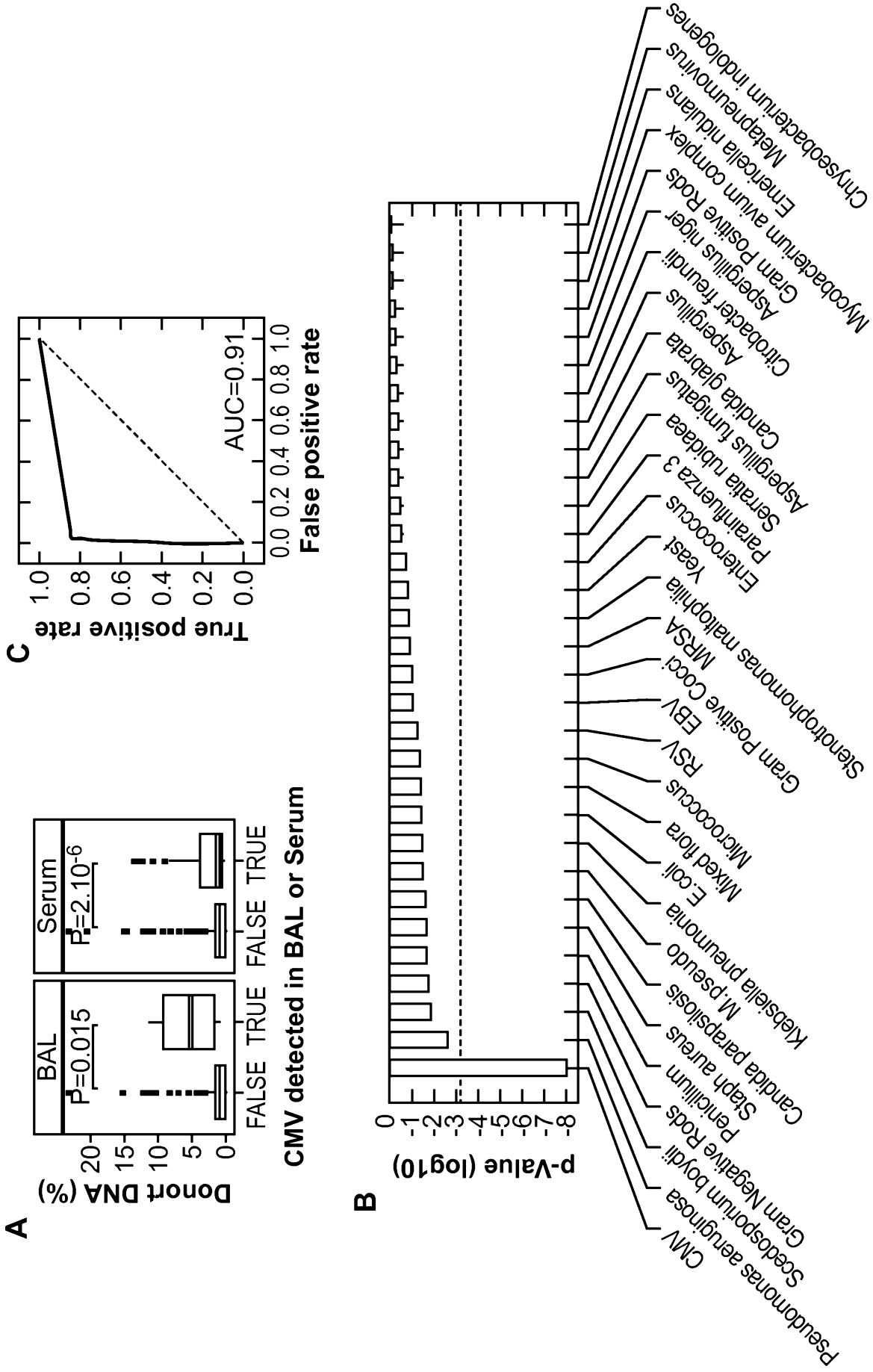
A
Donort DNA (%)
BAL
Serum
P=0.015
P=2.10-6
FALSE TRUE FALSE TRUE
CMV detected in BAL or Serum
B
p-Value (log10)
CMV
Pseudomonas aeruginosa
Scedosporium boydii
Gram Negative Rods
Penicillium
Staph aureus
Candida parapsilosis
M.pseudo
Klebsiella pneumonia
E.coli
Mixed flora
Micrococcus
RSV
EBV
Gram Positive Cocci
MRSA
Stenotrophomonas maltophilia
Yeast
Enterococcus
Paraintluenza 3
Serratia rubidaea
Aspergillus fumigatus
Candida glabrata
Aspergillus
Citrobacter freundii
Aspergillus niger
Gram Positive Rods
Emenicella nidulans
Metapneumovirus
Mycobacterium avium complex
Chrysseobacterium indologenes
C
True positive rate
AUC=0.91
False positive rate

FIG. 10

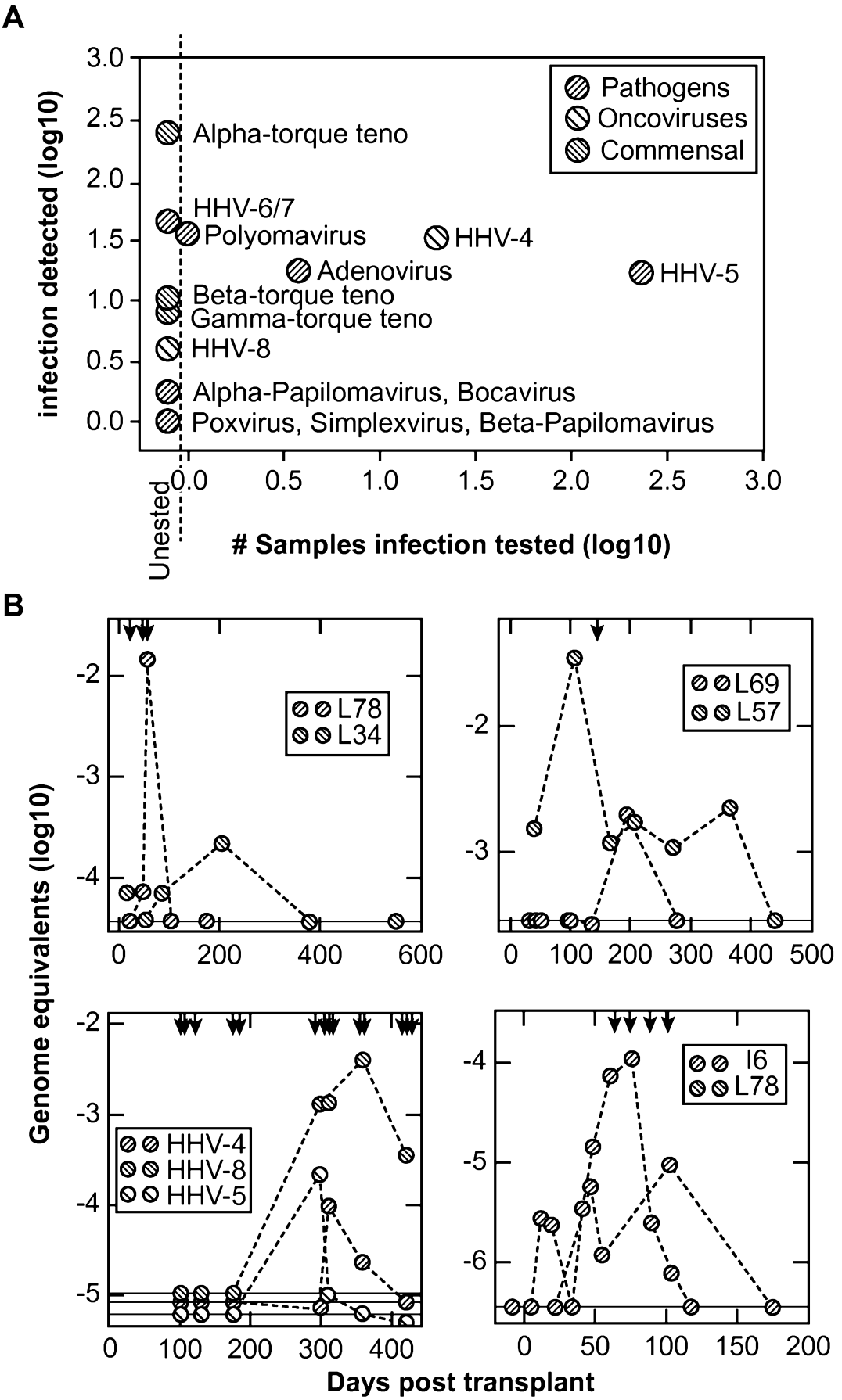
A
infection detected (log10)
Samples infection tested (log10)
Unested
Pathogens
Oncoviruses
Commensal
Alpha-torque teno
HHV-6/7
Polyomavirus
HHV-4
Adenovirus
HHV-5
Beta-torque teno
Gamma-torque teno
HHV-8
Alpha-Papilomavirus, Bocavirus
Poxvirus, Simplexvirus, Beta-Papilomavirus
B
Genome equivalents (log10)
Days post transplant
L78
L34
L69
L57
HHV-4
HHV-8
HHV-5
I6
L78

FIG. 11