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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 January 2025 (30.01.2025)

WIPO | PCT

(10) International Publication Number
WO 2025/024111 A1

(51) International Patent Classification:
C12Q 1/689 (2018.01) *C12Q 1/6855* (2018.01)
C12Q 1/6853 (2018.01) *C12Q 1/686* (2018.01)

(21) International Application Number:
PCT/US2024/036971

(22) International Filing Date:
05 July 2024 (05.07.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/515,480 25 July 2023 (25.07.2023) US
63/549,246 02 February 2024 (02.02.2024) US

(71) Applicant: **GUSTO GLOBAL, LLC** [US/US]; 2025 Cassamia Place, Charlotte, North Carolina 28211 (US).

(72) Inventors: **VAN DER LELIE, Daniel**; 2025 Cassamia Place, Charlotte, North Carolina 28211 (US). **OUELLETTE, Lisa**; 2025 Cassamia Place, Charlotte, North Carolina 28211 (US). **TAGHAVI, Safiyh**; 2025 Cassamia Place, Charlotte, North Carolina 28211 (US).

(74) Agent: **HESS, Madeline** et al.; WILSON SONSINI GOODRICH & ROSATI, 650 Page Mill Road, Palo Alto, California 94304 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO,

(54) Title: SINGLE POINT AMPLICON FRAGMENT SEQUENCING AND METHODS FOR DIAGNOSING AND MONITORING DISEASE

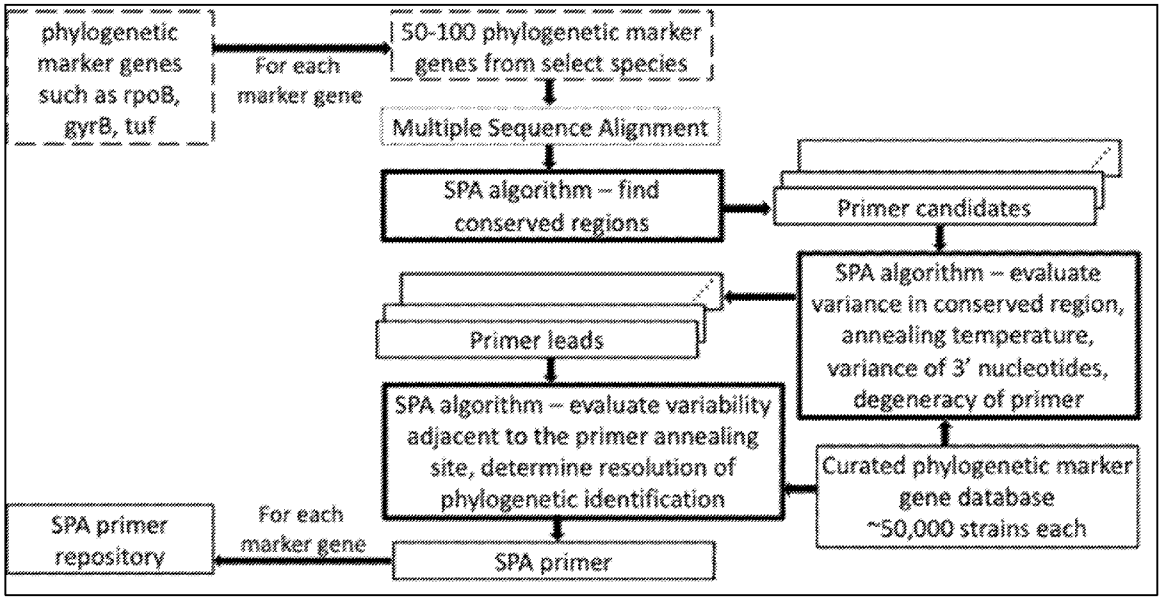

FIG. 4

(57) **Abstract:** Methods are provided for amplifying microbial cell free DNA (mcfDNA) in blood samples. In contrast to amplicon sequencing, in which regions between two conserved sequences are targeted for PCR amplification, the methods herein use a single conserved sequence as the primer annealing site to initiate PCR amplification. Enrichment PCR (E- PCR) is performed using a degenerate primer complementary to a conserved region in a microbial phylogenetic marker gene and an additional primer to the opposite end of the mcfDNA. The conserved region is adjacent to a hypervariable region. Amplification PCR is performed on the enriched mcfDNA fragments using a modified degenerate primer having a 3 '-end extension relative to the degenerate primer, and the additional primer, thereby generating enriched amplified mcfDNA fragments. Sequencing the amplified mcfDNA enables identification of microbial species, which is useful for diagnosing and monitoring inflammatory diseases and cancer, including stratification of cancer by

[Continued on next page]

WO 2025/024111 A1

RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

stage, type, and tumor subtype.

SINGLE POINT AMPLICON FRAGMENT SEQUENCING AND METHODS FOR DIAGNOSING AND MONITORING DISEASE

CROSS REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 63/515,480 filed July 25, 2023, and U.S. Provisional Application No. 63/549,246 filed February 2, 2024, each of which is incorporated herein by reference in its entirety.

INCORPORATION BY REFERENCE

[0002] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. To the extent publications and patents or patent applications incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

TECHNICAL FIELD

[0003] The presently disclosed subject matter relates to a high-throughput, high-resolution and low-cost method for early detection and monitoring of disease including cancer and IBD.

BACKGROUND

[0004] Liquid biopsy based on circulating cell-free DNA (cfDNA) provides a new prospect for the diagnosis, monitoring and risk assessment of a range of diseases. cfDNA molecules circulating in peripheral blood originate from dying human cells as well as from viruses, parasites, and colonizing or invasive microbes that release their nucleic acids into the blood as they die and break down (Jahr et al, 2001). Human-derived cfDNA has evolved into an indispensable biomarker in clinical practice for rapid and noninvasive diagnosis in prenatal screening, organ transplantation, and oncology (Decker and Sholl, 2020; Liang et al, 2019; Sun et al, 2019; Wu et al, 2020).

[0005] Although early studies did not focus on cfDNA of microbial origin (hereinafter referred to as mcfDNA), the development of circulating mcfDNA-based tests for infectious diseases has recently been gaining traction in clinical practice. An increasing number of studies have demonstrated that mcfDNA detection offers the potential to reliably identify a wide variety of infections, such as invasive fungal infection, tuberculosis, sepsis, cystic fibrosis (Rassoulian Barrett et al, 2020) and chorioamnionitis (Witt et al, 2020; for review see Han et al, 2020).

[0006] In addition to their role in infectious diseases, several studies have shown the presence of distinct cultivable bacteria in a range of cancers, including lung (Jin et al, 2019), prostate (Gorelick et al, 1988; Cohen et al, 2005), pancreas (Geller et al, 2017; Riquelme et al, 2019), and colon cancers (Bullman et al, 2019; Castellarin et al, 2012). It was only recently suggested that cancer types outside of the aerodigestive tract, such as breast (Urbaniak et al, 2016) or brain cancer (Venkataramani et al. 2019; Zeng et al, 2019), may also harbor microbiota with distinctive compositions (for review, see Sepich-Poore et al, 2021), including fungi (Narunsky-Haziza et al, 2022). Both Nejman et al. (2020) and Poore et al. (2020) suggested the existence of distinct intratumoral microbiomes among >30 cancer types; these microbiomes also vary in composition at different developmental stages of the tumor, thus providing biomarkers for disease progression and prognosis for patient outcomes. As for other bacteria that are colonizing or infecting the body, the tumor associated bacteria will release distinct mcfDNA in the blood stream, and this let Poore et al (2020) propose the analysis of mcfDNA from the peripheral blood as a tool to gain valuable information regarding the progression of various types of cancers.

[0007] For inflammatory bowel disease (IBD), existing stool and peripheral blood inflammatory biomarkers are not informative to monitor patient disease status, medication performance and treatment outcomes as they fail to address a growing body of evidence that increased intestinal permeability and gut microbiome dysbiosis are correlated to disease progression, symptom severity, and patient outcomes. Currently, the lactulose - mannitol test is the most common and only direct test of a leaky gut. However, this test doesn't provide insights in imbalanced gut microbiome bacteria and a dysfunctional gut epithelial barrier. For example, leakage of microbial material from the gut into the bloodstream can result in an overactive immune response that underlies multiple inflammatory conditions besides IBD, including rheumatoid arthritis, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis and plaque psoriasis, and even inflammatory neurological diseases such as Alzheimer's disease. Therefore, there is an urgent need for new ways of testing to assess the level of microbial material in the bloodstream as a surrogate biomarker in IBD and other inflammatory diseases.

[0008] Conventional amplicon-based sequencing approaches are routinely used to determine microbial community composition in a wide range of biological samples. The most used approach is amplicon sequencing of the 16S rRNA gene based on its variable regions, such as the V1-V2 and V3-V4 regions (Gupta et al, 2019). Shahir et al (2020) applied 16S rRNA gene sequencing to identify region-specific composition and aerotolerance profiles of mucosally adherent bacteria in biopsy samples taken from the colon and ileum of Crohn's disease and non-IBD patients. As an alternative to 16S rRNA gene sequencing, single copy proteins encoding housekeeping genes including the

genes for the DNA gyrase subunit B (*gyrB*) (Poirier et al, 2018), RNA polymerase subunit B (*rpoB*) (Vos at al, 2012; Ogier et al, 2019), the heat shock protein 60 (*hsp60*), the superoxide dismutase A (*sodA*), the TU elongation factor (*tuf*) (Ghebremedhin et al, 2008) and the 60 kDa chaperonin protein (*cpn60*) (Links et al, 2012) have been proposed as phylogenetic marker genes.

[0009] Liquid biopsy samples, especially peripheral blood, represent unique challenges for the analysis of microbial signatures. The majority of mcfDNA fragments in blood were found to be approximately 40 - 100 bp in size (Burnham et al, 2016), as was confirmed by Rassoulian Barrett et al (2020). Due to the small size of mcfDNA fragments conventional amplicon-based sequencing approaches that target DNA fragments of several hundred nucleotides (>400) are not suitable for determining the composition of colonizing or invasive microorganisms using mcfDNA from liquid biopsy samples. For example, the V1-V2 and the V3-V4 regions of the 16S rRNA gene have an average length of 437 and 443 nucleotides, respectively. Furthermore, the concentrations of plasma cfDNA in healthy individuals varies greatly, generally within the range of 0-100 ng per milliliter of plasma, sometimes exceeding 1500 ng per milliliter. Human cfDNA accounts for the vast majority (>90% or even >99%), while mcfDNA accounts for only a small fraction with 0.08%-4.85% from bacteria, 0.00%-0.01% from fungi, and 0.00%-0.16% from viruses/phages. However, it should be noted that elevated levels of mcfDNA can sometimes be observed in certain pathological conditions, including infection, sepsis, trauma, and autoimmune diseases (Han et al, 2020). Because the analysis of mcfDNA requires deep next generation sequencing (NGS) of plasma cfDNA to overcome the limitations of small mcfDNA fragment size and low concentration, this approach is unsuitable for the testing of large patient cohorts or routine health screening.

[0010] For example, although a lot of progress has been made in reducing the cost and increasing the throughput of NGS sequencing, it remains very expensive to analyze mcfDNA on a routine basis for community health screening and disease prognosis / diagnostics, as is routinely performed for many other health related parameters (blood cell panels, metabolic panels, etc.) or non-invasive early detection of diseases in at risk populations, such as the screening for colorectal cancer. Thus, there remains an unmet need for improved methods to accurately determine in a high throughput and cost-efficient way the presence of colonizing and invasive microbes that contribute to mcfDNA present in peripheral blood as part of clinical diagnostics, disease monitoring and community health screening. The presently disclosed subject matter provides such an improved method for high resolution, high-throughput, and low-cost detection of microorganisms as part of disease diagnosing and monitoring.

SUMMARY

[0011] In a first aspect of the invention, a method is provided of amplifying microbial cell free DNA (mcfDNA), comprising: (a) providing a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region; (b) providing an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA; (c) performing, on a sample comprising the mcfDNA, an enrichment polymerase chain reaction (E-PCR) using the degenerate primer and the additional primer, thereby generating enriched mcfDNA fragments; and (d) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer, and (ii) the additional primer, thereby generating enriched amplified mcfDNA fragments.

[0012] In a second aspect of the invention, a method is provided of amplifying microbial cell free DNA (mcfDNA), comprising: (a) performing an enrichment polymerase chain reaction (E-PCR) on a sample comprising the microbial cell-free DNA (mcfDNA), wherein enriched mcfDNA fragments are generated using: (i) a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and wherein the degenerate primer comprises an affinity purification group, and (ii) an additional primer complementary to an end of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region; and (b) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using: (i) a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer, and (ii) the additional primer, thereby generating enriched amplified mcfDNA fragments.

[0013] In a third aspect of the invention, a method is provided of diagnosing or monitoring a disease or disorder, comprising: (a) providing a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region; (b) providing an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA; (c) performing, on a sample comprising the mcfDNA, an enrichment polymerase chain reaction (E-

PCR) using the degenerate primer and the additional primer, thereby generating enriched mcfDNA fragments; (d) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer; and (ii) the additional primer, thereby generating enriched amplified mcfDNA fragments; (e) sequencing or obtaining sequences of the enriched amplified mcfDNA fragments; (f) assigning one of more of the mcfDNA sequences as belonging to a particular microbial species, or as comprising a microbial DNA signature, based on the closest sequence match of the hypervariable region to a reference microbial database; and (g) optionally, calculating a microbial community composition for the sample, based on a relative abundance of the mcfDNA sequences assigned to each microbial species, wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[0014] In a fourth aspect of the invention, a method is provided of diagnosing or monitoring a disease or disorder, comprising: (a) performing, on a sample comprising microbial cell-free DNA (mcfDNA), a polymerase chain reaction (PCR) using: (i) a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and (ii) an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region, thereby generating amplified mcfDNA fragments; (b) sequencing the amplified mcfDNA fragments or obtaining the sequences of the amplified mcfDNA fragments; (c) assigning one or more of the mcfDNA sequences as comprising a microbial DNA signature or as belonging to a particular microbial species, based on the closest sequence match of the hypervariable region to a reference microbial database; and (d) optionally, calculating a microbial community composition for the sample, based on the relative abundance of the mcfDNA fragment sequences assigned to each microbial species, wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[0015] In a fifth aspect of the invention, a set of degenerate oligonucleotides is provided comprising a mixture of oligonucleotides having a V:W:X:Y:Z ratio of 1.2:0.8-1.2:0.8-1.2:0.8-1.2:1.5-3, wherein V is the sequence 5' to 3' CGRTTDCCNARRTGRTCRATRTCATC, wherein

W is the sequence 5' to 3' CGRTTDCCNARRTGRTCRATATCATC, wherein X is the sequence 5' to 3' CGRTTDCCNARGTGRTCRATRTCATC, wherein Y is the sequence 5' to 3' CGRTTDCCNARGTGRTCRATATCATC, and wherein Z is the sequence 5' to 3' CGGTTRCCNARGTGGTCGATRTCGTC, wherein A = adenine, G = guanidine, C = cytosine, T = thymine, R = purine (A or G), D = not C (A, T or G), and N = any nucleotide (A, G, C or T), and wherein D and N can be replaced with inosine.

[0016] In a sixth aspect of the invention, a set of degenerate oligonucleotides is provided comprising a mixture of oligonucleotides having a V:W:X:Y:Z ratio of 1.2:0.8-1.2:0.8-1.2:0.8-1.2:1.5-3, wherein V is the sequence 5' to 3' GCRCGRTTDCCNARRTGRTCRATRTC, wherein W is the sequence 5' to 3' GCRCGRTTDCCNARRTGRTCRATATC, wherein X is the sequence 5' to 3' GCRCGRTTDCCNARGTGRTCRATRTC, wherein Y is the sequence 5' to 3' GCRCGRTTDCCNARGTGRTCRATATC, and wherein Z is the sequence 5' to 3' GCRCGGTTRCCNARGTGGTCGATRTC, wherein A = adenine, G = guanidine, C = cytosine, T = thymine, R = purine (A or G), D = not C (A, T or G), and N = any nucleotide (A, G, C or T), wherein D and N can be replaced with inosine, and wherein the oligonucleotides optionally comprise an affinity purification group at a 5' terminus.

[0017] In a seventh aspect of the invention, a kit is provided for amplifying microbial cell free DNA (mcfDNA), comprising: (a) an adaptor for ligating to ends of mcfDNA; (b) a degenerate primer comprising complementarity to a conserved region, wherein the degenerate primer comprises an affinity purification group, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region on the phylogenetic marker gene, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region; (c) an additional primer complementary to a repaired version of the adaptor; (d) a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer; and (e) instructions for performing: (i) an enrichment polymerase chain reaction (E-PCR) on the mcfDNA having the adaptor-ligated ends with the degenerate primer and the additional primer to generate enriched amplified mcfDNA fragments, (ii) an affinity purification of the enriched mcfDNA fragments, and (iii) an amplification PCR (A-PCR) on the enriched mcfDNA fragments using the modified degenerate primer and the additional primer to generate enriched amplified mcfDNA fragments.

[0018] In an eighth aspect of the invention, a method is provided of diagnosing or monitoring a disease or disorder, comprising: (a) providing mcfDNA sequences, wherein the

mcfDNA sequences are derived from a sample, wherein the mcfDNA sequences comprise sequence corresponding to (i) a conserved region spanning at least 18 nucleotides of a microbial phylogenetic marker gene and (ii) a hypervariable region at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region; (b) assigning one or more of the mcfDNA sequences as comprising a microbial DNA signature or as belonging to a particular microbial species based on the closest sequence match of the hypervariable region to a reference microbial database; and (c) optionally, calculating a microbial community composition for the sample, based on a relative abundance of the mcfDNA sequences assigned to each microbial species, wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[0019] In a ninth aspect of the invention, a method is provided of diagnosing or monitoring a disease or disorder, comprising: (a) sequencing amplified mcfDNA fragments derived from a sample, wherein the amplified mcfDNA fragments comprise sequence corresponding to (i) a conserved region spanning at least 18 nucleotides of a microbial phylogenetic marker gene and (ii) a hypervariable region at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region; (b) assigning one or more of the mcfDNA sequences as comprising a microbial DNA signature or as belonging to a particular microbial species based on the closest sequence match of the hypervariable region to a reference microbial database; and (c) optionally, calculating a microbial community composition for the sample, based on a relative abundance of the mcfDNA sequences assigned to each microbial species, wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[0020] In a tenth aspect of the invention, a method is provided of diagnosing or monitoring a disease or disorder, comprising: (a) performing, on a sample comprising microbial cell-free DNA (mcfDNA), a polymerase chain reaction (PCR) using: (i) a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and (ii) an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region, thereby generating amplified mcfDNA fragments; and (b) sequencing the amplified mcfDNA

fragments or obtaining the sequences of the amplified mcfDNA fragments, wherein the closer the sequence match between the mcfDNA sequences and the nucleic acid sequences in a reference sample for a disease or disorder the more likely the sample is to have the disease or disorder.

[0021] In an eleventh aspect of the invention, a method is provided of detecting infection or inflammation, comprising: (a) performing, on a sample comprising mcfDNA from a subject, an enrichment polymerase chain reaction (E-PCR) using a degenerate primer and an additional primer, thereby generating enriched mcfDNA fragments, wherein: (i) the degenerate primer comprises complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region, and (ii) the additional primer is complementary to a repaired version of an adaptor ligated to ends of the mcfDNA; (b) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer; and (ii) the additional primer, thereby generating enriched amplified mcfDNA fragments; (c) sequencing or obtaining sequences of the enriched amplified mcfDNA fragments; and (d) assigning one of more of the mcfDNA sequences as belonging to one or more pathogenic bacterial species based on the closest sequence match of the hypervariable region to a reference microbial database, wherein the one or more pathogenic bacterial species comprise one or more of Escherichia species, Escherichia coli, Shigella species, Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral, Pseudomonas species, Pseudomonas aeruginosa, Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species, Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium, and wherein a presence of the mcfDNA sequences from one or more of the one or more pathogenic bacterial members indicates that the subject has the infection or inflammation.

[0022] In a twelfth aspect of the invention, a method is provided of detecting infection or inflammation, comprising: (a) performing, on a sample comprising microbial cell-free DNA (mcfDNA) from a subject, a polymerase chain reaction (PCR) using: (i) a degenerate primer

comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and (ii) an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region, thereby generating amplified mcfDNA fragments; (b) sequencing the amplified mcfDNA fragments or obtaining sequences of the amplified mcfDNA fragments; and (c) assigning one or more of the mcfDNA sequences as belonging to one or more pathogenic bacterial species based on the closest sequence match of the hypervariable region to a reference microbial database, wherein the pathogenic bacterial species comprise one or more of Escherichia species, Escherichia coli, Shigella species, Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral, Pseudomonas species, Pseudomonas aeruginosa, Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species, Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium, and wherein a presence of the mcfDNA sequences from one or more of the pathogenic bacterial members indicates that the subject has the infection or inflammation.

[0023] In a thirteenth aspect of the invention, a method is provided of detecting infection or inflammation, comprising: (a) providing mcfDNA sequences, wherein the mcfDNA sequences are derived from a sample, wherein the mcfDNA sequences comprise sequence corresponding to (i) a conserved region spanning at least 18 nucleotides of a microbial phylogenetic marker gene, and (ii) a hypervariable region at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region; and (b) assigning one or more of the mcfDNA sequences as belonging to a pathogenic bacterial species based on the closest sequence match of the hypervariable region to a reference microbial database, wherein the pathogenic bacterial species comprise one or more of Escherichia species, Escherichia coli, Shigella species, Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral, Pseudomonas species, Pseudomonas aeruginosa, Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species,

Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium, and wherein a presence of the mcfDNA sequences from one or more of the pathogenic bacterial members indicates that the subject has the infection or inflammation.

[0024] In a fourteenth aspect of the invention, a system is provided for amplifying microbial cell free DNA (mcfDNA), comprising: (a) a reaction vessel; (b) a reagent dispensing module; and (c) software to execute any of the methods provided herein, wherein the method is executed at least partially robotically.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] Fig. 1 is a schematic of SPA fragment generation. The arrow indicates the position of the SPA primer (5' to 3'). The SPA fragment refers to the mcfDNA fragment region that will be amplified that is shown in dark shading.

[0026] Fig. 2 is a schematic overview of one embodiment of the protocol for generating single point amplification (SPA) fragments for sequencing. The various steps are in order of their successive execution. Once single point amplicon fragments are generated, they are sequenced using the standard protocol for next generation paired-end Illumina sequencing.

[0027] Fig. 3 is a schematic representation of the ligation and PCR steps that can be included in the SPA fragment sequencing protocol. Details of the SPA-linker used in the ligation step, the Linker.v2-amp primer and the 16SV4-R785-Bio primer used in the enrichment (E-PCR) step, the six phased primers R1-linker.v2 -SP0 to SP5 and the 16SV4-R781-Rd2 primer used in the amplification (A-PCR) step, and the indexing primers used in the indexing (I-PCR) step are shown. The sequences of the UMI region (with sequence NNTTNN (SEQ ID NO: 1)) and the non-complementary 3'-end sequence of the shorter SPA-linker fragment (with sequence 3'-ACG) are indicated via shading. The 3'-end extension to the degenerate primer used in the A-PCR reaction compared to the degenerate primer used in the E-PCR reaction is highlighted in ***bold/italics***. The Illumina Read-1 and Read-2 sequences for use in multiplex ILLUMINA NEXTERA sequencing are included in the primers used in the A-PCR and these sequences are shown in **bold**, with the linker sequences being underlined. The following nucleotide codes were used: A: adenine; G: guanidine; C: cytosine; T: thymine; B: not A (T, C, or G); V: not T (A, C, or G); H: not G (A, T or C); D: not C (A, T or G); N: any nucleotide (A, G, C or T).

[0028] Fig. 4 is an overview of an exemplary method used for SPA primer selection.

[0029] Fig. 5 shows a comparison between SPA fragment sequencing, 16S rRNA V3-V4 amplicon sequencing and 16S rRNA V6 amplicon sequencing to determine the microbial

community composition in water after a spike in with 2.5 pg/µl MC1 DNA. The 16S rRNA gene amplicon sequencing results are based on sequencing the 130 bp 5' and 3' ends of the V3-V4 region using the Rd1-16S-V3-F primer and the Rd2-16S-V4-R primer, respectively, and sequencing of the V6 region using the Rd1-16S-V6-F and Rd2-16S-V6-R primers. Only bacteria identified with >50% confidence at the genus level are reported. For comparison, the expected theoretical composition of MC1, after correction for genome size and 16S rRNA gene copy number, is provided.

[0030] Fig. 6 shows a comparison between SPA fragment sequencing, 16S rRNA V3-V4 amplicon sequencing and 16S rRNA V6 amplicon sequencing to determine the microbial community composition in blood plasma after a spike in with 2.5 pg/µl MC1 DNA. The 16S rRNA gene amplicon sequencing results are based on sequencing the 130 bp 5' and 3' ends of the V3-V4 region using the Rd1-16S-V3-F primer and the Rd2-16S-V4-R primer, respectively, and sequencing of the V6 region using the Rd1-16S-V6-F and Rd2-16S-V6-R primers. Only bacteria identified with >50% confidence at the genus level are reported. For comparison, the expected theoretical composition of MC1, after correction for genome size and 16S rRNA gene copy number, is provided.

[0031] Fig. 7 shows a comparison between SPA fragment sequencing and 16S rRNA V3-V4 amplicon sequencing to determine the microbial community composition in blood plasma after a spike in with 5.0 pg/µl MC1 DNA. The 16S rRNA gene amplicon sequencing results are based on sequencing the 3' end of the V3-V4 region using the Rd2-16S-V4-R primer. Only bacteria identified at >50% confidence level at the genus level are reported. For comparison, the expected theoretical composition of MC1, after correction for genome size and 16S rRNA gene copy number, is provided.

[0032] Fig. 8 shows the successful stratification of colon cancer patients and non-small cell lung cancer patients using principal component analysis of their mcfDNA microbial signatures obtained with SPA fragment sequencing of the 16S rRNA gene V4 hypervariable region. The type of cancer as well as the stage of the cancer are indicated for each patient. The light shading represents colon cancer and the dark fill represents non-small cell lung cancer.

[0033] Fig. 9 shows the successful stratification of colon cancer patients and breast cancer patients using principal component analysis of their mcfDNA microbial signatures obtained with SPA fragment sequencing of the 16S rRNA gene V4 hypervariable region. The type of cancer as well as the stage of the cancer are indicated for each patient. The light shading represents breast cancer and the dark fill represents colon cancer.

[0034] Fig. 10 shows the successful stratification of non-small cell lung cancer patients and breast cancer patients using principal component analysis of their mcfDNA microbial signatures obtained with SPA fragment sequencing of the 16S rRNA gene V4 hypervariable region. The type of cancer as well as the stage of the cancer are indicated for each patient. The dark fill represents breast cancer and the light shading represents non-small cell lung cancer.

[0035] Fig. 11 shows stratification of colorectal cancer patients and healthy individuals using principal component analysis of their mcfDNA microbial signatures obtained with SPA fragment sequencing of the 16S rRNA gene V4 hypervariable region. The stage of colorectal cancer is indicated for each patient. A healthy individual is defined as an individual not having a cancer diagnosis and presumably cancer-free. The dark fill represents colorectal cancer and the light shading represents healthy cancer-free; the water control is represented by a double triangle shape.

[0036] Fig. 12 shows nucleotide statistics for the *rpoB* gene region 1327-1355 and degenerate sequence (GAYGAYATYGAYCAYYTNGGHAAYCGHGC (SEQ ID NO: 2)) which from position 1327-1352 is the reverse complement sequence of degenerate primer RpoB1-R1327. The relative abundance of a nucleotide at a specific position was calculated using the nucleotide sequences of 47,505 aligned unique *rpoB* genes from the PATRIC database and used to design the degenerate sequence, which is provided from 5' to 3' using the following nucleotide codes: A: adenine; G: guanidine; C: cytosine; T: thymine; R: purine (A or G); Y: pyrimidine (T or C); H: not G (A, T or C); N: any nucleotide (A, G, C or T); N*: presence of any nucleotide (N) at a specific *rpoB* gene position. The percentages of highly conserved nucleotide sequences used to determine the consensus sequence for the degenerate primer are highlighted. The position of the region is based on the nucleotide sequence of the *Escherichia coli rpoB* gene.

[0037] Fig. 13 shows the machine learning informed diagnosis/classification of patients using their *rpoB* gene-derived mcfDNA signatures. The SciKitLearn (https://scikit-learn.org/stable/) package in python was used for stratification.

[0038] Fig. 14A shows box plots of the blood pathogenicity index versus the disease status for ulcerative colitis. Disease status was determined as being in remission or "not in remission", respectively, with "not in remission" being defined as an active disease flair or medication failure – medication stopped working or serious side effects of the medication were observed. The blood pathogenicity index, determined for the samples used to create Table 7, was calculated as the percentage of mcfDNA fragments coming from a defined set of pro-inflammatory bacteria as a fraction of the total amount of mcfDNA fragments found in a blood plasma sample. The null hypothesis that the distribution of the blood pathogenicity index values of patients in remission is

the same as the distribution of patients not in remission is rejected for ulcerative colitis patients with p-value of 0.0006881 (Mann-Whitney U test). The blood pathogenicity median is shown as a line inside the box extending from the first quartile (Q1) to the third quartile (Q3). The whiskers extend from the box to the farthest data point lying within 1.5x the inter-quartile range (IQR) from the box. Outlier points are those past the end of the whiskers.

[0039] Fig. 14B shows box plots of the blood pathogenicity index versus the disease status for Crohn's disease. Disease status was determined as being in remission or "not in remission", respectively, with "not in remission" being defined as an active disease flair or medication failure – medication stopped working or serious side effects of the medication were observed. The blood pathogenicity index, determined for the samples used to create Table 7, was calculated as the percentage of mcfDNA fragments coming from a defined set of pro-inflammatory bacteria as a fraction of the total amount of mcfDNA fragments found in a blood plasma sample. The null hypothesis that the distribution of the blood pathogenicity index values of patients in remission is the same as the distribution of patients not in remission is rejected for Crohn's disease with p-value of 6.33e-05 (Mann-Whitney U test). The blood pathogenicity median is shown as a line inside the box extending from the first quartile (Q1) to the third quartile (Q3). The whiskers extend from the box to the farthest data point lying within 1.5x the inter-quartile range (IQR) from the box. Outlier points are those past the end of the whiskers.

[0040] Fig. 15A shows box plots of the fecal calprotectin levels versus the disease status for ulcerative colitis. Disease status was determined as being in remission or "not in remission", respectively, with "not in remission" being defined as an active disease flair or medication failure – medication stopped working or serious side effects of the medication were observed. Calprotectin levels were determined as μg/g fecal material for the samples used to generate the data for Table 8. The null hypothesis that the distribution of calprotectin levels of patients in remission is the same as the distribution of patients not in remission is rejected for ulcerative colitis patients with a p-value of 0.0212876 (Mann-Whitney U test). The calprotectin median is shown as a line inside the box extending from the first quartile (Q1) to the third quartile (Q3). The whiskers extend from the box to the farthest data point lying within 1.5x the inter-quartile range (IQR) from the box. Outlier points are those past the end of the whiskers.

[0041] Fig. 15B shows box plots of the fecal calprotectin levels versus the disease status for Crohn's disease. Disease status was determined as being in remission or "not in remission", respectively, with "not in remission" being defined as an active disease flair or medication failure – medication stopped working or serious side effects of the medication were observed. Calprotectin levels were determined as μg/g fecal material for the samples used to generate the data for Table 8.

The null hypothesis that the distribution of calprotectin levels of patients in remission is the same as the distribution of patients not in remission is accepted for Crohn's disease patients with a p-value of 0.8428739 (Mann-Whitney U test). The calprotectin median is shown as a line inside the box extending from the first quartile (Q1) to the third quartile (Q3). The whiskers extend from the box to the farthest data point lying within 1.5x the inter-quartile range (IQR) from the box. Outlier points are those past the end of the whiskers.

[0042] Fig. 16 shows a schematic representation of the journey of Crohn's disease patient CD-4 and how the blood-based pathogenicity index (BPI) can be successfully used to influence a better disease outcome. The patient is in remission (12/2018; 06/2019) on a combination of azathioprine and adalimumab, the latter being discontinued due to side effects (10/2019). Stopping this medication results in full remission (01/2020). In retrospect (10/2019), clinical use of the BPI would have identified inflammatory markers in blood that alter CD-4 patient pathway by changing medication. The side effects from adalimumab were missed by the calprotectin (Calpr) levels and the SCDAI score (SCADI). Data points marked by open circles or grey circles indicate remission or active disease, respectively. Arrows indicate medication is being taken over a certain time period. A STOP symbol indicates that a medication is being discontinued at a given timepoint.

[0043] Fig. 17 shows a schematic representation of the journey of ulcerative colitis patient UC-3 and how the blood-based pathogenicity index (BPI) can be successfully used to influence a better disease outcome. It is assumed that the initial combination of tofacitinib, vedolizumab and mesalamine is working before 11/2018. Issues are reported for mesalamine, which is removed from the medication (02/2019). In retrospect (11/2018), clinical use of BPI would have identified inflammatory markers in blood that alter UC-3 patient pathway by removing mesalamine as medication, resulting in a decrease of the BPI (03/2019) indicative for remission in line with fecal urgency. The side effects from mesalamine were missed by the calprotectin levels (Calpr) and the Mayo-6 score (Mayo-6). Data points marked by open circles or grey dots indicate remission or active disease, respectively. A data point marked by a black dot indicates that no sample to determine the BPI was available. Arrows indicate medication is being taken over a certain time period. A STOP symbol indicates that a medication is being discontinued at a given timepoint.

[0044] Fig. 18 shows a schematic representation of the journey of ulcerative colitis patient UC-9 and how the blood-based pathogenicity index (BPI) can be successfully used to influence a better disease outcome. Initially, mesalamine is working as medication but starts to lose efficiency (11/2017) based on disease symptoms and calprotectin levels. Adalimumab is added as medication (12/2017), resulting in remission (11/2019). In retrospect (11/2017), clinical use of the BPI would have identified inflammatory markers in blood that alter the UC-9 patient pathway by changing

medication to include adalimumab, resulting in a decrease of the BPI (11/2019). Data points marked by open circles or grey dots indicate remission or active disease, respectively. A data point marked by a black dot indicates that no sample to determine the BPI was available. Arrows indicate medication is being taken over a certain time period.

DETAILED DESCRIPTION

[0045] The presently disclosed subject matter now will be described more fully hereinafter. The presently disclosed subject matter may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will satisfy applicable legal requirements. Indeed, many modifications and other embodiments of the presently disclosed subject matter set forth herein will come to mind to one skilled in the art to which the presently disclosed subject matter pertains having the benefit of the teachings presented in the descriptions provided herein. Therefore, it is to be understood that the presently disclosed subject matter is not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims.

[0046] Following long-standing patent law convention, the terms “a,” “an,” and “the” refer to “one or more” when used in this application, including the claims. Thus, for example, reference to “a degenerate primer” includes more than one degenerate primer, unless the context clearly is to the contrary, and so forth.

[0047] Throughout this specification and the claims, the terms “comprise,” “comprises,” and “comprising” are used in a non-exclusive sense, except where the context requires otherwise. Likewise, the terms “having” and “including” and their grammatical variants are intended to be non-limiting, such that recitation of items in a list is not to the exclusion of other like items that can be substituted or added to the listed items.

[0048] For the purposes of this specification and claims, the term “about” when used in connection with one or more numbers or numerical ranges, should be understood to refer to all such numbers, including all numbers in a range and modifies that range by extending the boundaries above and below the numerical values set forth. The recitation of numerical ranges by endpoints includes all numbers, e.g., whole integers, including fractions thereof, subsumed within that range (for example, the recitation of 1 to 5 includes 1, 2, 3, 4, and 5, as well as fractions thereof, e.g., 1.5, 2.25, 3.75, 4.1, and the like) and any range within that range. In addition, as used herein, the term "about", when referring to a value can encompass variations of, in some embodiments +/-20%, in some embodiments +/-10%, in some embodiments +/-5%, in some embodiments +/-1%, in some

embodiments +/-0.5%, and in some embodiments +/-0.1%, from the specified amount, as such variations are appropriate in the disclosed compositions and methods. Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude, preferably within 5-fold, and more preferably within 2-fold, of a value. Where particular values are described in the application and claims, unless otherwise stated the term "about" meaning within an acceptable error range for the particular value should be assumed.

[0049] Throughout this specification and the claims, the term "subject" includes humans and animals and can be used interchangeably with the term "human" and the term "patient".

[0050] The terms "SPA fragment" and "SPA fragment sequence" are herein used interchangeably.

[0051] The term "microbial phylogenetic marker gene" as used herein means any conserved gene from any organism, including but not limited to bacteria, fungi, parasites, and viruses, that is suitable for phylogenetic identification.

[0052] The terms "amplified mcfDNA fragments" and "amplified mcfDNA fragment sequences" are used interchangeably herein for the purposes of the specification and claims.

[0053] There are a wide range of diseases where microbial community analysis, especially of the gut microbiome, provides important information regarding the disease or its treatment options. This includes conditions such as IBD (Ananthakrishnan et al, 2017), metabolic diseases (Boulangé et al, 2016), diseases of the central nervous system (Bhattacharjee and Lukiw, 2013), including Alzheimer's disease (Ferreiro et al, 2023), and cancer where the interaction of the gut microbiome can provide clues regarding the response to specific treatments including immune checkpoint inhibitors (Gopalakrishnan et al, 2020; Sepich-Poore et al, 2021). Deep microbial metagenome sequencing is the most informative approach when it comes to microbial community analysis, as it will provide detailed information regarding community composition as well as the key functions encoded by the community members. Unfortunately, despite major breakthroughs in metagenome sequencing technologies to reduce its costs, it is currently still too expensive for routine screening purposes of human associated microbial communities in large population screenings. Another disadvantage of deep microbial metagenome sequencing is the need for relatively large amounts of high-quality microbial DNA. This has hindered its application to study the microbial communities associated with liquid and solid biopsy samples, where only a small fraction of the total DNA is of microbial origin.

[0054] The amplification and subsequent sequencing of phylogenetic marker genes provides an alternative, cheaper high throughput method for microbial community analysis. For example, in tissue biopsy samples where there is sufficient concentration of DNA having average

fragment length of about 5,000 bp or more, amplification-based sequencing approaches have been successfully applied to identify differences in microbial communities between healthy individuals and patients suffering from a wide range of diseases. Advantages of the amplification and subsequent sequencing method include that it requires significantly less DNA than metagenome sequencing, and because specific DNA primers are used to amplify phylogenetic target genes, there is little contamination with host DNA, making this method suitable to analyze the microbial communities associated with tissue biopsy samples, from which small amounts of high molecular weight DNA can be obtained. However, analysis of microbial signatures in liquid biopsy samples, especially peripheral blood samples, results in additional challenges as compared to tissue biopsy samples, due to the low concentration of mcfDNA having small fragment sizes.

[0055] For example, in plasma, human cfDNA accounts for the vast majority of cfDNA (>90% or even >99%), while mcfDNA accounts for only a small fraction with 0.08%-4.85% from bacteria, 0.00%-0.01% from fungi, and 0.00%-0.16% from viruses/phages (Han et al, 2020). However, the percentage of mcfDNA compared to cfDNA should be placed in the context of the human genome size and the size of an average microbial genome, with sizes of 6.4 billion and approximately 6 million nucleotides, respectively, therefore providing similar coverage. Thus, mcfDNA represents an important signal that is largely being ignored in liquid biopsy testing.

[0056] The intrinsic properties of cfDNA and mcfDNA, especially its small fragment sizes, make its analysis for disease detection and monitoring challenging. More than 70% of plasma cfDNA is smaller than 300 bp, with an average size of 170 bp (Fernández-Carballo et al, 2019). However, the average size of mcfDNA fragments was found to be significantly smaller, approximately 40-100 bp (Burnham et al, 2016), as was confirmed by Rassoulian Barrett et al (2020). As a result of this size limitation, conventional amplicon-based sequencing approaches including 16S rRNA gene and rpoB gene amplicon sequencing that target DNA fragments of several hundred nucleotides, are not suitable for determining the composition of colonizing or invasive microorganisms using mcfDNA from peripheral blood and other liquid biopsy samples. The small size of mcfDNA makes it nearly impossible to use mcfDNA in amplicon-based sequencing protocols, such as 16S rRNA gene sequencing, leaving no other option than high-cost and low-throughput NGS sequencing.

[0057] To overcome the above-mentioned limitations, the present inventors developed a single point amplification sequencing approach that exploits the combination of a degenerate primer for a conserved region of a marker gene located adjacent to a phylogenetic hypervariable region of the gene for a wide range of microbes. The method is based on the targeted amplification of high-resolution phylogenetic identifier fragments from mcfDNA, which comprises a fraction of

the total cfDNA isolated from, for example, biopsy samples. To generate the phylogenetic identifier fragments, a hypervariable DNA region with high phylogenetic resolution is targeted. The hypervariable region located next to the highly conserved region that functions as a primer annealing site as is illustrated in Figure 1. In the methods disclosed herein, the fragments resulting from specific amplification of the hypervariable DNA regions are referred to as SPA fragments.

[0058] In various embodiments, methods and kits are provided herein for generating the SPA fragments from a sample, such as a blood sample. The methods and kits provided herein can be used to determine the presence of one or more microbial species, microbial DNA signatures, and/or microbial community compositions for the sample. In the methods and kits provided herein, the microbial phylogenetic marker gene can be eukaryotic, fungal, viral, or bacterial, and combinations thereof. In one embodiment, the microbial phylogenetic marker gene is a eubacterial phylogenetic marker gene.

[0059] In the methods of the invention, the length of the SPA fragment is determined by the distance between the end of the mcfDNA fragment and the 3'-end of the primer annealing site. Only mcfDNA fragments that contain the primer annealing site will give SPA fragments, which can be subsequently sequenced and used for high resolution phylogenetic identification of microbial species and/or microbial signatures. Some embodiments include analysis of microbial community composition.

[0060] In contrast to commonly used amplicon sequencing, in which regions between two conserved DNA sequences are targeted for PCR amplification, the methods provided herein use a single conserved DNA sequence as the primer annealing site to initiate PCR amplification. The amplification initiated from this single conserved DNA sequence allows for targeted amplification of the hypervariable region located adjacent to the primer annealing site, independent of the size of the fragment, followed by sequencing of the amplified fragment. This method may be referred to herein as Single Point Amplification (SPA) fragment sequencing.

[0061] In one embodiment, a method is provided of amplifying microbial cell free DNA (mcfDNA), comprising: (a) providing a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region; (b) providing an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA; (c) performing, on a sample comprising the mcfDNA, an enrichment polymerase chain reaction (E-PCR) using the degenerate primer and the additional primer, thereby generating enriched mcfDNA

fragments; and (d) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer, and (ii) the additional primer, thereby generating enriched amplified mcfDNA fragments.

[0062] In another embodiment, a method is provided of amplifying microbial cell free DNA (mcfDNA), comprising: (a) performing an enrichment polymerase chain reaction (E-PCR) on a sample comprising the microbial cell-free DNA (mcfDNA), wherein enriched mcfDNA fragments are generated using: (i) a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and wherein the degenerate primer comprises an affinity purification group, and (ii) an additional primer complementary to an end of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region; and (b) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using: (i) a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer, and (ii) the additional primer, thereby generating enriched amplified mcfDNA fragments.

[0063] Additional advantages of the present invention include that it enables identification of microbial species and/or microbial DNA signatures in a sample, such as a blood sample, that allows for diagnosis and/or stratification of cancer patients by type of cancer, stage of cancer, and specific tumor subtype driven by the location of the tumor.

[0064] Therefore, in some embodiments, methods are provided for diagnosing a disease or disorder in a sample, including a blood sample. The methods include assigning one or more mcfDNA fragment sequences derived from the sample as belonging to a particular microbial species based on the closest sequence match of the hypervariable region to a reference microbial database, or as having a microbial DNA signature. Optionally, the method can include calculating a microbial community composition for the sample, based on the relative abundance of the mcfDNA fragment sequences assigned to each microbial species. Diagnosing the sample as having a disease or disorder includes linking the microbial DNA signature, the microbial community composition, or one or a combination of the presence, the absence, or the relative abundance of the microbial species to the disease or disorder.

[0065] In one embodiment, a method is provided of diagnosing or monitoring a disease or disorder, comprising: (a) providing a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial

phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region; (b) providing an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA; (c) performing, on a sample comprising the mcfDNA, an enrichment polymerase chain reaction (E-PCR) using the degenerate primer and the additional primer, thereby generating enriched mcfDNA fragments; (d) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer; and (ii) the additional primer, thereby generating enriched amplified mcfDNA fragments; (e) sequencing or obtaining sequences of the enriched amplified mcfDNA fragments; (f) assigning one of more of the mcfDNA sequences as belonging to a particular microbial species, or as comprising a microbial DNA signature, based on the closest sequence match of the hypervariable region to a reference microbial database; and (g) optionally, calculating a microbial community composition for the sample, based on a relative abundance of the mcfDNA sequences assigned to each microbial species, wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[0066] In another embodiment, a method is provided of diagnosing or monitoring a disease or disorder, comprising: (a) performing, on a sample comprising microbial cell-free DNA (mcfDNA), a polymerase chain reaction (PCR) using: (i) a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and (ii) an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region, thereby generating amplified mcfDNA fragments; (b) sequencing the amplified mcfDNA fragments or obtaining the sequences of the amplified mcfDNA fragments; (c) assigning one or more of the mcfDNA sequences as comprising a microbial DNA signature or as belonging to a particular microbial species, based on the closest sequence match of the hypervariable region to a reference microbial database; and (d) optionally, calculating a microbial community composition for the sample, based on the relative abundance of the mcfDNA fragment sequences assigned to each microbial species, wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a

combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[0067] In yet another embodiment, a method is provided of diagnosing or monitoring a disease or disorder, comprising: (a) providing mcfDNA sequences, wherein the mcfDNA sequences are derived from a sample, wherein the mcfDNA sequences comprise sequence corresponding to (i) a conserved region spanning at least 18 nucleotides of a microbial phylogenetic marker gene and (ii) a hypervariable region at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region; (b) assigning one or more of the mcfDNA sequences as comprising a microbial DNA signature or as belonging to a particular microbial species based on the closest sequence match of the hypervariable region to a reference microbial database; and (c) optionally, calculating a microbial community composition for the sample, based on a relative abundance of the mcfDNA sequences assigned to each microbial species, wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[0068] In one aspect, a method is provided of diagnosing or monitoring a disease or disorder, comprising: (a) sequencing amplified mcfDNA fragments derived from a sample, wherein the amplified mcfDNA fragments comprise sequence corresponding to (i) a conserved region spanning at least 18 nucleotides of a microbial phylogenetic marker gene and (ii) a hypervariable region at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region; (b) assigning one or more of the mcfDNA sequences as comprising a microbial DNA signature or as belonging to a particular microbial species based on the closest sequence match of the hypervariable region to a reference microbial database; and (c) optionally, calculating a microbial community composition for the sample, based on a relative abundance of the mcfDNA sequences assigned to each microbial species, wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[0069] In another aspect, a method is provided of diagnosing or monitoring a disease or disorder, comprising: (a) performing, on a sample comprising microbial cell-free DNA (mcfDNA), a polymerase chain reaction (PCR) using: (i) a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and (ii) an additional primer complementary to a repaired version of an

adaptor ligated to ends of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region, thereby generating amplified mcfDNA fragments; and (b) sequencing the amplified mcfDNA fragments or obtaining the sequences of the amplified mcfDNA fragments, wherein the closer the sequence match between the mcfDNA sequences and the nucleic acid sequences in a reference sample for a disease or disorder the more likely the sample is to have the disease or disorder.

[0070] In addition to the methods provided above, a blood pathogenicity index (BPI) is provided as an example of a microbial signature that can be used to monitor disease progression, medication efficiency and medication side effects in IBD patients or in patients suffering from other inflammatory conditions. The BPI is calculated as the percentage of mcfDNA fragments coming from pro-inflammatory bacteria as part of the total amount of mcfDNA fragments found in a blood plasma sample.

[0071] More specifically, a method is provided of detecting infection or inflammation, comprising: (a) performing, on a sample comprising mcfDNA from a subject, an enrichment polymerase chain reaction (E-PCR) using a degenerate primer and an additional primer, thereby generating enriched mcfDNA fragments, wherein: (i) the degenerate primer comprises complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region, and (ii) the additional primer is complementary to a repaired version of an adaptor ligated to ends of the mcfDNA; (b) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer; and (ii) the additional primer, thereby generating enriched amplified mcfDNA fragments; (c) sequencing or obtaining sequences of the enriched amplified mcfDNA fragments; and (d) assigning one of more of the mcfDNA sequences as belonging to one or more pathogenic bacterial species based on the closest sequence match of the hypervariable region to a reference microbial database, wherein the one or more pathogenic bacterial species comprise one or more of Escherichia species, Escherichia coli, Shigella species, Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral,

Pseudomonas species, Pseudomonas aeruginosa, Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species, Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium, and wherein a presence of the mcfDNA sequences from one or more of the one or more pathogenic bacterial members indicates that the subject has the infection or inflammation.

[0072] In another embodiment, a method is provided of detecting infection or inflammation, comprising: (a) performing, on a sample comprising microbial cell-free DNA (mcfDNA) from a subject, a polymerase chain reaction (PCR) using: (i) a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and (ii) an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region, thereby generating amplified mcfDNA fragments; (b) sequencing the amplified mcfDNA fragments or obtaining sequences of the amplified mcfDNA fragments; and (c) assigning one or more of the mcfDNA sequences as belonging to one or more pathogenic bacterial species based on the closest sequence match of the hypervariable region to a reference microbial database, wherein the pathogenic bacterial species comprise one or more of Escherichia species, Escherichia coli, Shigella species, Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral, Pseudomonas species, Pseudomonas aeruginosa, Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species, Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium, and wherein a presence of the mcfDNA sequences from one or more of the pathogenic bacterial members indicates that the subject has the infection or inflammation.

[0073] In one aspect, a method is provided of detecting infection or inflammation, comprising: (a) providing mcfDNA sequences, wherein the mcfDNA sequences are derived from a sample, wherein the mcfDNA sequences comprise sequence corresponding to (i) a conserved region spanning at least 18 nucleotides of a microbial phylogenetic marker gene, and (ii) a

hypervariable region at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region; and (b) assigning one or more of the mcfDNA sequences as belonging to a pathogenic bacterial species based on the closest sequence match of the hypervariable region to a reference microbial database, wherein the pathogenic bacterial species comprise one or more of Escherichia species, Escherichia coli, Shigella species, Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral, Pseudomonas species, Pseudomonas aeruginosa, Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species, Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium, and wherein a presence of the mcfDNA sequences from one or more of the pathogenic bacterial members indicates that the subject has the infection or inflammation.

[0074] In the methods provided herein, the degenerate primer can be used in combination with an adaptor, such as, for example, an asymmetric linker cassette which is attached to the ends of the cfDNA fragments in the sample. A PCR amplification reaction is performed using the degenerate primer and a primer complementary to the 5' asymmetrical end of the linker cassette. The degenerate primer is designed to allow for DNA synthesis into the hypervariable region. However, successful PCR amplification of the hypervariable region occurs only when the asymmetric linker cassette is repaired. In a PCR reaction, the asymmetric linker cassette will be repaired only when located downstream from the degenerate primer annealing site, i.e, when the asymmetric linker cassette has been ligated to a mcfDNA fragment that contains the conserved region of the phylogenetic marker gene. In this manner, microbial DNA fragments that originate from the hypervariable region are selectively amplified.

[0075] To allow for enrichment of amplified fragments containing the hypervariable region, an affinity tagged (e.g., a biotinylated) version of the degenerate primer can be used. After a limited number of amplification cycles (e.g. less than 10 PCR cycles), the amplified and biotinylated fragments can be purified using streptavidin coated magnetic beads. This step results in removal of the background cfDNA, especially human cfDNA, that lacks the annealing site for the degenerate primer. This step is referred to as "enrichment PCR (E-PCR)" (see Figure 3).

[0076] The purified biotinylated fragments are subsequently used in a second PCR reaction. The degenerate primer used in this PCR reaction, referred to as the "amplification PCR (A-PCR) reaction, recognizes the same conserved DNA region as the primer used in the E-PCR

reaction, except that its annealing site located in the conserved region is shifted towards the 3'-direction compared to the primer used in the E-PCR reaction (see, e.g., Figure 3). For example, in one embodiment, the degenerate primer used in the A-PCR comprises a 3'-end extension of at least three nucleotides relative to the degenerate primer of the E-PCR. This allows for further enrichment of DNA fragments that contain the conserved DNA region while avoiding the amplification of DNA fragments that originated from aspecific hybridization of the E-PCR primer.

[0077] To further improve the quality of the sequencing data, phased primers recognizing the linker sequence can be used for the A-PCR reaction (see, e.g., Figure 3).

[0078] In one example of the invention, to overcome the above-mentioned limitations for determining microbial profiles from mcfDNA, such as, for example, mcfDNA in liquid biopsy samples, the present inventors experimentally exemplified the methods provided herein by exploiting the phylogenetic resolution of a hypervariable region of the V3-V4 hypervariable region of the 16S rRNA gene, a multi copy gene. In another experimental example, the present inventors exemplified an approach that exploits the phylogenetic resolution of the rpoB gene, a single copy housekeeping gene. See EXAMPLES 1-5 herein below.

[0079] Additional embodiments of the invention include use of a conserved DNA sequence as the primer annealing site for more than one site on a phylogenetic marker gene or for a site on two or more different phylogenetic marker genes in a single amplification reaction. Specifically, the methods provided herein can further include providing one or more additional degenerate primers comprising complementarity to one or more additional conserved regions of the microbial phylogenetic marker gene or to one or more additional genes. The one or more additional genes can be one or more additional microbial phylogenetic marker genes. In another example, the one or more additional genes can encode virulence factors, secondary metabolites, or drug degradation pathways, and combinations thereof. In one instance, the methods provided herein can include a degenerate primer for both the 16S rRNA gene and the *rpoB* gene. The use of two or more degenerate primers for annealing to two or more conserved regions on a single or two different phylogenetic marker genes may be referred to herein as "multi-loci SPA fragment sequencing".

[0080] In the specific EXAMPLES provided herein (i.e., EXAMPLES 1-5 below) the 16S rRNA gene and the RNA polymerase subunit B (*rpoB*) gene were used, but the SPA fragment sequencing method is very broadly applicable to conserved housekeeping genes, including, but not limited to, the prokaryotic genes coding for the DNA gyrase subunit B (*gyrB*), the heat shock protein 60 (*hsp*60), the superoxide dismutase A protein (*sodA*), the TU elongation factor (*tuf*), and the DNA recombinase proteins (including *recA, recE*). The SPA fragment sequencing method can also be applied on the Eukaryotic internal transcribes spacer (ITS) regions ITS1, which is located

between the 18S and 5.8S rRNA genes, and ITS2, which is located between the 5.8S and 28S rRNA genes. The SPA fragment sequencing method can also be applied to genes that are unique to pathogenic fungi including the *trr1* gene that encodes for thioredoxin reductase; the *rim8* gene that encodes for a protein involved in the proteolytic activation of a transcriptional factor in response to alkaline pH; the *kre2* gene that encodes for α-1,2-mannosyltransferase; and the *erg6* gene that encodes for Δ(24)-sterol C-methyltransferase (Abadio et al, 2011); or any conserved gene from any organism, including bacteria, fungi, parasites, and viruses that is suitable for phylogenetic identification. This includes conserved genes from the human DNA-based oncoviruses, more specifically the Epstein-Barr Virus (EBV), Human Papillomavirus (HPV), Hepatitis B virus (HBV), Human Herpesvirus-8 (HHV-8), and Merkel Cell Polyomavirus (MCPyV) (Mui et al, 2017). One or a combination of any conserved housekeeping gene can be used in the presently disclosed methods.

[0081] Advantages of the disclosed SPA fragment sequencing methods include an increase in the diversity of hypervariable regions that can be targeted for amplicon analysis as the method only requires one neighboring conserved region to bind the primer (compared with the two required by dual primer approaches). As such, the SPA fragment sequencing method is more adaptable, flexible, and offers greatly improved resolution over current methods, especially when it comes to the analysis of short DNA fragments such as mcfDNA.

[0082] In addition to the degenerate primer for the conserved region, an adaptor such as, for example, an asymmetric linker cassette, can be used to introduce a DNA sequence that is targeted by an additional primer in the PCR amplification reaction. In one embodiment, to avoid amplification of any DNA fragment flanked by two adaptors, the adaptors are "defective" or in other words "asymmetric". This can be accomplished by designing an adaptor as an asymmetric linker cassette where the strand that serves as the template for primer annealing is missing. Typical asymmetric linker cassette configurations include, but are not limited to:

1. A "Y"-shaped linker cassette, where two single stranded DNA fragments that are only partially complementary are annealed. This results in an asymmetric linker cassette where one end is double stranded, allowing for ligation, but where the other end is comprised of two single stranded non-complementary DNA strands.
2. A "single arm" linker cassette, where a shorter single stranded DNA fragment is annealed to the complementary 3'-end of a longer single stranded DNA fragment. This results in an asymmetric linker cassette with a single stranded the 5'-end and a double stranded 3'-end.

[0083] The single strands of the asymmetric linker cassette can be complementary over a stretch of about at least 16 nucleotides with an annealing temperature of approximately 50°C or

higher, allowing for a linker cassette that is stable at room temperature. The single strand of the asymmetric linker can also contain 4 to 6 random nucleotides that constitute a Unique Molecular Identifier (UMI) to correct PCR induced errors including PCR amplification bias, improve sequencing accuracy, and correct for fragment copy number. To avoid self-ligation, in one example, the asymmetric linker cassette includes a 3'sticky end. The 3'sticky end can be formed by a single nucleotide, such as, for example, thymine. To avoid undesirable repair of the asymmetric linker cassette initiated from the shorter single stranded DNA fragment, the terminal 3' nucleotide can be a dideoxy nucleotide that functions as a chain-elongating inhibitor of DNA polymerase.

[0084] The term Unique Molecular Identifier (UMI) is intended to mean a stretch of random, non-random, or partially random nucleotides that are added as an identifier barcode to a DNA or RNA fragment before the fragment is amplified. The objective is to aid in distinguishing between identical copies of a fragment present in a sample. For example, SPA fragment sequences derived from the same microorganisms are identical except for the length of the sequenced fragment, which varies as a function of the distance between the gene specific primer annealing site (e.g., 16SV4-R781-Rd2 primer) and the end of the mcfDNA fragment, to which the UMI is attached. As such, the uniqueness of a sequenced DNA fragment can be inferred from a combination of its UMI, the length of the SPA fragment, and the sequence of the SPA fragment. As a result of this combination of features, the number of UMIs can be smaller than the number of fragments that cover a specific DNA region from the same microorganism. Given that the pool of UMI's does not have to be in excess of the number of fragments that cover a specific DNA region from the same microorganism, each of the UMI's present on the DNA or RNA fragments are not required to be unique. Therefore, the term "UMI" is not intended to be understood according to the dictionary definition of "unique". In one embodiment, the UMIs can range in length from 4 to 6 nucleotides, encoding for 256 or 4096 unique sequences, respectively. In another specific case, a UMI is used with sequence NNTTNN (SEQ ID NO: 1), where N = any nucleotide (A, G, C or T), and wherein A = adenine, G = guanidine, C = cytosine and T = thymine. The presence of the two thymine nucleotides allows for determination of the frequency of errors that occur in the UMI sequence during PCR amplification.

[0085] In a PCR reaction, the asymmetric linker cassette will only be repaired when located downstream from the degenerate primer annealing site. For purposes of the specification and claims, the term "repaired" when used in the context of the asymmetric linker cassette, means that a new DNA strand is created in the PCR reaction that is complementary at the 5' end of the asymmetric linker cassette. DNA synthesis initiated from the degenerate primer into the asymmetric linker cassette will restore the defective DNA strand complementary to the 5'-end of

the linker and in this manner the asymmetric linker cassette is repaired. In subsequent PCR cycles this strand is used for primer annealing, allowing for the amplification of the hypervariable region.

[0086] To allow for sample multiplexing and sequencing, the resulting amplicons can be further amplified in a third PCR reaction to introduce two Unique Dual Indexes (UDI), one at each end of the amplicons, and, for example, the Illumina sequencing anchors P5 and P7. These indexes provide distinct barcode sequences for each of the i5 and i7 index reads for a specific sequencing library, allowing for multiplexing of sequencing libraries. After sequencing, the barcodes are used to assign each sequenced fragment to its specific library.

[0087] In one embodiment of the invention, the method includes one or more of the following steps as outlined in Figure 2:

1. Blood plasma collection and isolation of cfDNA using standard protocols. Cell-free DNA can be extracted from 0.5 ml blood plasma using the typically yielding 0.1 ng to 10 ng to be used for sequencing. cfDNA can also be isolated from urine, saliva, stool, spinal fluid, cerebral fluid, and other biopsy samples. For this step many commercial kits are available that either use columns or magnetic beads for cfDNA isolation from liquid biopsy samples, including blood and urine. In addition to cfDNA, total DNA that has been fragmented to smaller sizes via mechanical shearing or enzymatic digestion can be used.
2. End repair and 5'-phosphorylation of cfDNA fragments followed by the 3' addition of a deoxy-adenine to create a 3'-sticky end formed by a single adenine nucleotide using standard protocols. A typical protocol to process cfDNA includes end repair (blunting and 5′ phosphorylation), 3′ A-tailing, followed by adaptor ligation. The fragment ends are repaired by blunting and 5′ phosphorylation with a mixture of enzymes, such as T4 polynucleotide kinase (PNK) and T4 DNA polymerase (T4 DNA pol). This end repair step is followed by 3′ A-tailing at 37 °C using a mesophilic polymerase such as Klenow Fragment 3′-5′ exonuclease minus (Head et al, 2014). Many commercial kits are available to perform this step.
3. Ligation of the adaptor, which in this case is an asymmetric linker cassette, using T4 DNA ligase. Many commercial kits are available to perform this step and combine it with the end-repair/A-tailing step, such as the NEBNEXT ULTRA II Ligation kit (NEB) or the IDT XGEN DNA Lib Prep MC kit. Sequences of asymmetric linker cassettes, referred to as the SPA-Linker (e.g. SPA-linker-UMI.v1-Y-SP or SPA-Linker-UMI.v2-Y-SP), are provided in Table 1 along with other primer sequences that can be utilized in the methods and kits provided herein. Annealing of two partially complementary single stranded DNA fragments results in a "Y"-shaped or single arm DNA linker cassette. On one end, the two

strands of the linker cassette are not complementary. On other end, where the two strands are complementary, the linker cassette includes a 3'sticky end formed by a single thymine nucleotide. Due to the sticky ends, the only possible ligation is between cfDNA fragments and asymmetric linker cassettes, while self-ligation of linker cassettes and repaired cfDNA fragments is blocked.

4. Single point linker cassette repair as part of the Enrichment PCR step (E-PCR). The terms "ePCR" and "E-PCR" are herein used interchangeably. PCR is performed on the ligation product using the following primers: (a) the Linker.v2-amp primer that recognizes the repaired 5' asymmetrical end of the SPA-Linker-UMI.v2-Y-SP linker cassette; (b) one or more primers that recognize the primer annealing site specific for the conserved region of the one or more phylogenetic marker genes. DNA amplification initiated from the gene-specific SPA primer will result in the repair of the asymmetric linker cassette but only when this cassette is bound to a cfDNA fragment that contains the primer annealing site on the conserved region. This will be limited to mcfDNA fragments that contain the targeted region of the phylogenetic marker gene such as, for example, position 781 - 805 of the 16S rRNA gene, which is absent in Eukaryotic DNA including human DNA. As such no human cfDNA fragments flanked by asymmetric linker cassettes will be repaired. In EXAMPLE 1 herein below, the 16SV4-R785-bio primer which recognizes the 16S rRNA gene sequence between positions 785-805, was used. In addition, due to the direction of the amplification reaction from the gene specific primer such as, for example, the 16SV4-R785-bio primer, only linker cassettes that are bound to the region downstream of the targeted region of the phylogenetic marker gene will be repaired (e.g., position 785 - 805 of the 16S rRNA gene). Neither primer used in the first step of the enrichment PCR contains the Illumina Read-1/2 extension

5. Once the asymmetric linker cassette has been repaired, the primer (Linker.v2-amp) that recognizes the repaired 5' asymmetrical end of the linker cassette can anneal, and PCR amplification is initiated. In the case of the 16SV4-R785-bio primer this will result in the amplification of DNA sequences located upstream of position 785 of the 16S rRNA gene. Using a 5'-biotinilated version of the one or more gene specific primers (e.g., 16SV4-R785-bio primer), followed by capturing the hybridized primer on magnetic streptavidin-coated beads allows for further enrichment of amplified fragments. Using streptavidin-coated beads in the subsequent purification step allows for the removal of the non-biotinylated DNA fragments. The enrichment of the amplified fragments is referred to herein as performing an "affinity purification" and the biotin is referred to as an affinity purification

group or an affinity tag. Affinity purification can be performed using any combination of steps and types of affinity purification groups known to those of ordinary skill in the art.

6. To further reduce the amplification of non-targeted DNA fragments, an additional PCR step, referred to as the amplification PCR (A-PCR) step, is used to reduce background amplification resulting from nonspecific primer annealing to e.g. human DNA fragments. The terms "aPCR" and "A-PCR" are herein used interchangeably. The A-PCR can use the Linker.v2-amp primer in combination with one primer annealing to the conserved region extended at its 3'-end by 3 or more nucleotides (e.g. primer 16SV4-R781-Rd2) compared to the primer used in the E-PCR step (STEP 4) (e.g. 16SV4-R785-bio; see Figure 3).

Often, sequencing libraries of DNA fragments that share identical sequences, such as linker fragments, results in loss of sequencing quality. To address this issue, in another embodiment, a set of primers can be used that are identical except that their sequences are phased by a single nucleotide, as shown in Figure 3. Therefore, as an alternative for using a single Linker.v2-amp primer, a set of six primers referred to as Linker-aPCR mix, can be used (see Table 1 and Figure 3 for their sequences). Furthermore, in some embodiments, the forward (Linker-aPCR mix) and reverse (e.g. 16SV4-R781-Rd2) primers include an additional 5' sequence corresponding to the Illumina Read-1 and Read-2 sequences, respectively, to allow for indexing as part of sequencing library preparation, and these sequences are shown in bold in Figure 3, with the linker sequence being underlined.

7. In the Indexing PCR reaction (I-PCR), Unique Dual Indexes (UDI) and Illumina sequencing anchors (P5 and P7) are added to the amplified SPA fragments using P5-I5-Rd1 and P7-I7-Rd2 primers (see Table 1). The I-PCR reactions are performed using unique sets of UDI for each sample, subsequently allowing the pooling/multiplexing of the libraries, after which fragments are paired-end sequenced using NGS Illumina sequencing, e.g. on the Illumina NEXTSEQ 1000 (Illumina, Inc., San Diego, CA). This approach will result in sequenced fragments that all share the sequence of the gene specific primer (e.g., 16SV4-R781-Rd2 primer) followed by sequences that vary in length and nucleotide composition. Sequences derived from the same microorganisms will be identical except for the length of the sequenced fragment, which will vary as a function of the distance between the gene specific primer annealing site (e.g., 16SV4-R781-Rd2 primer) and the end of the mcfDNA fragment.

Protocol ID	Primer Name	Pooling ratio	Sequence (5' → 3')
Asymmetric SPA linker cassette-v1 construction, SPA fragment amplification and sequencing			
SPA-Linker	SPA-linker-UMI.v1-Y-SP	Ordered as duplex	GACAGGGATTTGCTGGTCGNNNNNNAATTCAACTAGGCTTAATCCGAC*G*T (SEQ ID NO: 3)
			/5Phos/CGTCGGATTAAGCCTAGTTGAG*C*A (SEQ ID NO: 4)
Linker-ePCR	Linker.v1-amp	n/a	GACAGGGATTTGCTGGTCG (SEQ ID NO: 5)
Linker-aPCR	Rd1-linker.v1	n/a	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGATTTGCTGGTCG (SEQ ID NO: 6)
Asymmetric SPA linker cassette-v2 construction, SPA fragment amplification and sequencing			
SPA-Linker	SPA-linker-UMI.v2-Y-SP	Ordered as duplex	GCAGATACATAGAGTACCACTCNNTTNNAATTCAACTAGGCTTAATCCGAC*G*T (SEQ ID NO: 7)
			/5Phos/CGTCGGATTAAGCCTAGTTGAG*C*A (SEQ ID NO: 8)
Linker-ePCR	Linker.v2-amp	n/a	GCAGATACATAGAGTACCACTC (SEQ ID NO: 9)
Linker-aPCR mix	Rd1-linker-SP5	16.7%	**TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAGGCATAGCAGATACATAGAGTACCACTC (SEQ ID NO: 10)
	Rd1-linker-SP4	16.7%	**TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAGCATAGCAGATACATAGAGTACCACTC (SEQ ID NO: 11)
	Rd1-linker-SP3	16.7%	**TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAGATAGCAGATACATAGAGTACCACTC (SEQ ID NO: 12)
	Rd1-linker-SP2	16.7%	**TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAGTAGCAGATACATAGAGTACCACTC (SEQ ID NO: 13)
	Rd1-linker-SP1	16.7%	**TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAGAGCAGATACATAGAGTACCACTC (SEQ ID NO: 14)

	Rd1-linker-SP0	16.7%	**TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAGGCAGATACATAGAGTACCACTC (SEQ ID NO: 15)
SPA fragment sequencing initiated from the 16S rRNA gene region 781-805			
16SV4-ePCR	16SV4-R785-bio	n/a	/5Biosg/CGTGGACTACHVGGGTATCTA (SEQ ID NO: 16)
16SV4-aPCR	16SV4-R781	n/a	GACTACHVGGGTATCTA***ATCC*** (SEQ ID NO: 17)
16SV4-aPCR	16SV4-R781-Rd2	n/a	**GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAGGACTACHVGGGTATCTA***ATCC*** (SEQ ID NO: 18)
SPA fragment sequencing initiated from the rpoB gene region 1327-1355			
RpoB-ePCR mix	RpoB1-R1330-T3Y6Y15-bio	12.50%	/5biosg/GCRCGRTTDCCNARRTGRTCRATRTC (SEQ ID NO: 19)
	RpoB1-R1330-T3T6Y15-bio	12.50%	/5biosg/GCRCGRTTDCCNARRTGRTCRATATC (SEQ ID NO: 20)
	RpoB1-R1330-T3Y6C15-bio	12.50%	/5biosg/GCRCGRTTDCCNARGTGRTCRATRTC (SEQ ID NO: 21)
	RpoB1-R1330-T3T6C15-bio	12.50%	/5biosg/GCRCGRTTDCCNARGTGRTCRATATC (SEQ ID NO: 22)
	RpoB1-R1330-C3-bio	50%	/5biosg/GCRCGGTTRCCNARGTGGTCGATRTC (SEQ ID NO: 23)
RpoB-aPCR mix	RpoB1-R1327-T3Y6Y15	12.50%	CGRTTDCCNARRTGRTCRATRTC***ATC*** (SEQ ID NO: 24)
	RpoB1-R1327-T3T6Y15	12.50%	CGRTTDCCNARRTGRTCRATATC***ATC*** (SEQ ID NO: 25)
	RpoB1-R1327-T3Y6C15	12.50%	CGRTTDCCNARGTGRTCRATRTC***ATC*** (SEQ ID NO: 26)
	RpoB1-R1327-T3T6C15	12.50%	CGRTTDCCNARGTGRTCRATATC***ATC*** (SEQ ID NO: 27)
	RpoB1-R1327-C3	50%	CGGTTRCCNARGTGGTCGATRTC***GTC*** (SEQ ID NO: 28)
	Rd2-RpoB1-R1327-T3Y6Y15	12.50%	**GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAGCGRTTDCCNARRTGRTCRATRTC***ATC*** (SEQ ID NO: 29)

	Rd2-RpoB1-R1327-T3T6Y15	12.50%	**GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAGCGRTTDCCNARRTGRTCRATATC***A******TC*** (SEQ ID NO: 30)
	Rd2-RpoB1-R1327-T3Y6C15	12.50%	**GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAGCGRTTDCCNARGTGRTCRATRTC***A******TC*** (SEQ ID NO: 31)
	Rd2-RpoB1-R1327-T3T6C15	12.50%	**GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAGCGRTTDCCNARGTGRTCRATATC***A******TC*** (SEQ ID NO: 32)
	Rd2-RpoB1-R1327-C3	50%	**GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAGCGGTTRCCNARGTGGTCGATRTC***G******TC*** (SEQ ID NO: 33)
Illumina sequence library construction			
UDI (001-384)	P5-I5-Rd1	50%	AATGATACGGCGACCACCGAGATCTACAC / Index5 (10nt) / **TCGTCGGCAGCGTC** (SEQ ID NO: 34)
	P7-I7-Rd2	50%	CAAGCAGAAGACGGCATACGAGAT / Index7 (10nt) / **GTCTCGTGGGCTCGG** (SEQ ID NO: 35)
16S rRNA gene amplification and Illumina sequence library construction			
16S V3-V4 region	Rd1-16S-V3-F	50%	**TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG-CCTACGGGNGGCWGCAG (SEQ ID NO: 36)
	Rd2-16S-V4-R	50%	**GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG-GACTACHVGGGTATCTAATCC (SEQ ID NO: 37)
16S V6 region	Rd1-16S-V6-F	50%	**TCGTCGGCAGCGTC**AGATGTGTATAAGAGACAG-CAACGCGAAGAACCTTACC (SEQ ID NO: 38)
	Rd2-16S-V6-R	50%	**GTCTCGTGGGCTCGG**AGATGTGTATAAGAGACAG-CGACAGCCATGCANCACCT (SEQ ID NO: 39)

Table 1: Overview of primer sequences. The following nucleotide codes were used: A: adenine; G: guanidine; C: cytosine; T: thymine; R: purine (A or G); V: not T (A, C, or G); H: not G (A, T or C); D: not C (A, T or G); N: any nucleotide (A, G, C or T). The 3'-end extension to the degenerate primers used in the A-PCR reaction compared to the degenerate primers used in the E-PCR reaction are highlighted in ***bold/italics***. The Rd1 and Rd2 sequences plus the standard linker sequences to connect the original primer sequences (i.e., linker and degenerate primer sequences) to the Rd1 or Rd2 sequences used for multiplex ILLUMINA NEXTERA sequencing are included in A-PCR primers and are shown in bold and underlined, respectively. *: Phosphorothioated modification to protect nucleotide from nuclease degradation; /5Phos/: 5'-end phosphorylation modification; /5Biosg/: 5'-end biotin modification. The pooling ratio indicates the ratio of primers combined in a specific primer mix. n/a: not applicable.

[0088] In some embodiments of the invention, the processing and analysis of the SPA fragment sequences includes one or more of the following steps:

1. Paired-end reads are filtered based on read quality, expected amplicon structure, and length.
2. The UMI (unique molecular identifier) of each read is recorded for downstream deduplication and denoising.
3. Human-derived and singleton reads are removed.
4. Forward and reverse reads are merged if possible and clustered into bins of similar sequence.
5. The longest read in each bin is searched against a database of bacterial 16S rRNA genes or genes corresponding to the microbial phylogenetic marker gene being targeted to aid the assembly into consensus sequences.
6. Community composition is determined by consensus sequence classification, and relative abundances are calculated based on the percent of reads contributing to each consensus sequence, taking into consideration the number of deduplicated reads identified in step 2.

[0089] Additional primers besides those derived from the 16SV4 region can be used for SPA fragment sequencing. EXAMPLE 4 describes the design and use of *rpoB* gene specific primers.

[0090] The utility of the methods of the invention is exemplified in EXAMPLES 1-5 of the present disclosure. For example, in EXAMPLE 1 of the present disclosure the inventors describe an improved SPA fragment sequencing protocol targeting the conserved

region upstream of the hypervariable V4 region of the 16S rRNA gene. Surprisingly, the improved method resulted in the percentage of 16S rRNA gene derived fragments (as identified by Blastn) increasing from less than 0.01% to an average of 35%, a 3,500 fold increase (see Table 2).

[0091] In EXAMPLE 2 of the present disclosure, a comparison between SPA fragment sequencing and amplicon sequencing targeting the V3-V4 or the V6 regions of the 16S rRNA gene is described using samples that were spiked with DNA of defined microbial communities. SPA fragments targeting the 16S rRNA gene V4 were superior over amplicon-based methods, providing a more accurate description of the relative abundances and the phylogenetic identification of the bacteria present in the samples. SPA fragment sequencing was also less sensitive to the presence of contaminating DNA (see Figures 5 to 7).

[0092] In EXAMPLE 3 of the present disclosure, the inventors demonstrate that the mcfDNA in blood plasma of cancer patients can be used to generate SPA fragments targeting the 16S gene V4 region. These SPA fragments enabled distinguishing between cancer-free individuals and those with colorectal cancer as well as between patients with lung, breast and colorectal cancer, including patients with early-stage cancer (see Figures 8 to 11).

[0093] In EXAMPLE 4 of the present disclosure, the inventors demonstrate that the mcfDNA in blood plasma of cancer patients can be used to generate SPA fragments targeting the hypervariable region of the *rpoB* gene located upstream of position 1327. These SPA fragments allowed, with 85 percent accuracy, to distinguish between healthy individuals and patients with lung and breast cancer, including patients with early-stage cancer.

[0094] In EXAMPLE 5 of the present disclosure, the inventors demonstrate that the mcfDNA in blood plasma of IBD patients can be used to monitor disease severity and progression in IBD patients. SPA fragment sequences were used to determine the blood pathogenicity index (BPI), which is calculated as the percentage of mcfDNA fragments coming from pro-inflammatory bacteria as a fraction of the total amount of mcfDNA fragments found in a blood plasma sample. The BPI was successfully used to monitor disease progression and medication efficiency and failure in IBD patients (see Figures 14 to 16).

[0095] With the improved SPA fragment sequencing protocol, in one example, sequencing 300,000 fragments yields approximately 100,000 fragments that can be identified as 16S rRNA gene sequences. Since 100,000 sequencing reads represent the standard depths for amplicon-based sequencing for complex microbial community analysis, the latest Illumina NEXTSEQ instruments allow for an unprecedented number of samples to be sequenced in parallel. For example, the Illumina NEXTSEQ 6000 allows to theoretically

collect 20 billion reads with a single run, which would correspond to approximately 65,000 paired-end sequenced samples. This is in sharp contrast to the scalability of deep sequencing of total cfDNA, where only approximately 200 samples can be processed in a single paired-end sequencing run (Poore et al, 2020).

[0096] In addition to monitoring of specific diseases, SPA fragment sequencing can be useful as part of the general health screening. Unlike the stool microbiome, the microbiome of colonizing and infecting bacteria will be relatively stable, with changes occurring when the relation between host and microbes is changing. This includes situations of new invasions by infectious and colonizing microorganisms, such as the formation of stomach ulcers, the formation of intestinal polyps/adenomas and their progression into malignancies, gastrointestinal diseases including Irritable Bowel Disease (IBD), various tumors and their specific microbiomes including pancreatic cancer, lung cancer and cervical cancer, Central Nervous System (CNS) diseases including multiple sclerosis (MS) and Alzheimer's disease, minimal residual disease (MRD) monitoring, and other diseases characterized by dysbiotic and inflammatory microbiomes such as cystic fibrosis or tuberculosis, and general risk monitoring of infections in patient populations with a compromised immune system such as patient taking immune suppressing medications, positioning SPA fragment sequencing as an ideal tool for risk monitoring, early detection, prognostics and evaluation of disease progression. Contrary to PCR based detection methods that monitor for the presence of specific bacteria, SPA fragment sequencing provides an "open" diagnostics approach to detect any bacterium or fungus based on the presence of its mcfDNA in peripheral blood.

[0097] SPA fragment sequencing can provide an important non-invasive method for (early) detection and identification of infectious and colonizing bacteria using mcfDNA from peripheral blood samples, which can subsequently be linked to a broad range of diseases, including: screening for tuberculosis and other diseases caused by *Mycobacterium* species; determining pulmonary infection risks and causes in cystic fibrosis patients; determining the risk and onset of sepsis in patients with compromised immune systems; detection of opportunistic bacterial pathogens originating from the oral cavity that have been linked to Alzheimer's disease, pancreatic cancer and other serious conditions such as endocarditis; women's health issues including *Chlamydia* linked to mucopurulent cervicitis, pelvic inflammatory disease, tubal factor infertility, ectopic pregnancy and cervical cancer; detection and monitoring of progression of cancer; monitoring of minimal residual disease after oncology treatments; detection and monitoring of progression and minimal residual

disease of breast cancer including triple negative breast cancer; detection of esophageal cancer, precancerous colonic polyps and early stage colorectal cancer, and detection and monitoring of progression and minimal residual disease of gastrointestinal cancers in general; detection and monitoring of progression and minimal residual disease in lung cancer; non-invasive analysis of the microbiome in pancreatic cancer patients to propose treatment protocols and prognostics for long-term survival; detection of *Clostridium difficile* infections; post-transplant bloodstream infections and Graft versus Host Disease (GvHD); detection of hospital acquired infections by emerging pathogens of clinical concern; detection of an infection in an immune compromised person; or detection of infection or inflammation of the gastrointestinal track in Irritable Bowel Disease (Crohn's disease, Ulcerative colitis); monitoring the risk of infection of inflammation in patients suffering from rheumatoid arthritis, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis, plaque psoriasis; and combinations thereof. Therefore, SPA fragment sequencing represents a quantum leap forward to apply mcfDNA sequencing as a high-resolution, high-throughput and low-cost routine test in disease detection, patient monitoring, risk assessment and large-scale population screenings using mcfDNA informed biomarkers. For example, the microbial footprint obtained with SPA fragment sequencing combined with the mutational footprint and methylation footprint that are currently being used as biomarkers for the detection, monitoring and prognostics of cancers, can provide a powerful tool for improved early detection and monitoring of progression of various types of cancer. The term "microbial footprint" is used herein interchangeably with the terms "microbial community composition" and "community composition".

[0098] In some embodiments, the microbial footprint is used in combination with alternative tests, such as tests that focus on the analysis of circulating tumor DNA (ctDNA), for the detection of cancer, guiding tumor-specific treatments, monitoring of cancer treatment, and monitoring periods of symptom free remission and identifying potential early recurrence of cancer. Including the microbial footprint can increase the specificity and sensitivity of these screening tests, e.g. tests for the detection of specific cancers, such as early stage adenomas and carcinomas in colorectal cancer; or tests for early multi-cancer detection, such as the GALLERI cancer screening test and the GUARDANT test. Furthermore, once unique SPA fragments have been identified that correlate with the detection of specific diseases and monitoring of their progression, their sequences can be used to develop species-specific PCR-based screening assays as part of diagnostic platforms.

[0099] The microbial footprint can be used to calculate the blood pathogenicity index (BPI). Included as pathogenic or pro-inflammatory bacteria in the BPI are the relative abundances of *Enterobacteriaceae* including *Enterobacter* species, *Escherichia* species (e.g. *E. coli*), *Shigella* species, *Klebsiella* species (e.g. *K. pneumoniae* and *K. oxytoca*), *Kluyvera* species, *Erwinia* species, *Haemophilus* species, *Salmonella* species (e.g. *S. typhi* and *S. enterica*), and *Listeria* species (*L. monocytogenes*); pathogenic *Clostridium* species such as C. *difficile, C. perfringens* and *C. botulinum*; *Veillonella* species such as *V. atypica, V. dispar* and *V. oral*; *Pseudomonas* species like *P. aeruginosa*; *Stenotrophomonas* species like *S. maltophilia; Burkholderia* species like *B. cepacia*; pathogenic *Staphylococcus* species like *S. aureus; Porphyromonas* species; and *Fusobacterium* species like *F. nucleatum* and *F. varium*.

[00100] The results described in EXAMPLE 5 show that SPA fragment sequencing of mcfDNA in blood can be successfully used to determine the BPI of patients suffering from inflammatory diseases such as IBD. The BPI can be used as a surrogate biomarker for both medication performance and assessing the risk for serious or life-threatening complications, such as an increased risk of inflammation.

[00101] The increased risk for serious or life-threatening infections has been well documented for immune suppressing medications and there is a need for biomarkers that allow for early detection and prevention of this complication. The results of the retrospective study, described in Tables 7, 9 and 10 of EXAMPLE 5 show that failure or complications due to immune suppressing medication is accompanied by an increase in the mcfDNA signature of (opportunistic) pathogenic bacteria in the blood of the patients, as determined by the BPI. Therefore, the BPI can be used as a surrogate biomarker for diseases where patients are at risk of serious or life-threatening infections due to treatment with immune repressing drugs, including rheumatoid arthritis, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis, plaque psoriasis, acute Graft versus Host disease.

[00102] Immune suppressing mediations that can benefit from patient monitoring using the BPI to determine the risk of serious or life-threatening infections include, but are not limited to, abatacept (Orencia); adalimumab (Humira); infliximab (Remicade); Tofacitinib (Xeljanz); risankizumab (Skyrizi); anakinra (Kineret); azathioprine (Azasan); certolizumab (Cimzia); cyclosporine (Gengraf, Neoral, Sandimmune); etanercept (Enbrel); golimumab (Simponi); infliximab (Remicade); methotrexate (Otrexup, Rasuvo, Trexall); rituximab (Rituxan); steroids including dexamethasone, methylprednisolone (Medrol), prednisolone

(Orapred ODT, Prelone), and prednisone (Rayos); tacrolimus (Astagraf, Envarsus XR, Prograf); and tocilizumab (Actemra).

[00103] The BPI can be used to monitor the performance and possible side effects, when used for a prolonged period, of immune suppressing medications including corticosteroids, 5-amine-salicylic acid, inflammatory response modulators, calcineurin phosphatase inhibitors, purine metabolism antagonists, and biologicals (antibodies) including α4β7 integrin blockers, antiTNFα, antiIL12/IL23 and JAK-kinase inhibitors.

[00104] In addition, the BPI can be used to monitor the performance and possible side effects of other medications/treatments that can cause complications of inflammation, including the gastrointestinal track, or affect the immune response, such as immune checkpoint inhibitor drugs like Pembrolizumab, Nivolumab, and Cemiplimab as anti-PD-1 antibodies, Ipilimumab as an anti-CTLA-4 antibody, as well as Atezolizumab, Avelumab, and Durvalumab as anti-PD-L1 antibodies.

[00105] The BPI can be used in combination with alternative tests, such as tests that focus on the analysis of inflammatory biomarkers in stool and blood, such as the stool calprotectin levels or C-reactive protein (CRP), erythrocyte sedimentation rate (ESR) and plasma viscosity (PV) in blood. The BPI can also be used in combination with alternative microbial biomarkers, such as lipo-polysaccharide (LPS) levels.

[00106] In addition to using mcfDNA from blood, the SPA fragment sequencing approach provided herein is applicable to analyze microbial DNA compositions in any sample type, especially when in samples having low amounts of small fragment microbial DNA. This includes biopsy samples from solid tumors, skin grafts, and other liquid biopsy samples besides peripheral blood such as spinal and cerebral fluids, urine and saliva, as well as mcfDNA present in stool samples.

[00107] In the SPA fragment sequencing method, a DNA region is identified in a suitable microbial phylogenetic marker gene that has the following characteristics:

1. Presence of a highly conserved DNA region to design a degenerate primer for annealing to the phylogenetic marker gene.
2. Adjacent to the primer annealing site the presence of a highly variable DNA region with high phylogenetic resolution. This region will become part of the single point amplicon (SPA) fragment.

[00108] An overview of an exemplary SPA primer design method is shown in Figure 4. For each phylogenetic marker gene, such as *rpoB, cpn60*, 16S rRNA, *gyrB, tuf* or other phylogenetic marker gene or conserved housekeeping gene including, but not limited

to, those used by CheckM (Parks et al, 2015), 50-100 species are initially selected that cover the prokaryotic diversity, including members of the phylum Proteobacteria (including representative α-, β-, γ-, δ- and ε-Proteobacteria), the phylum Firmicutes (including representatives for the classes *Bacilli, Clostridia, Erysipelotrichia* and *Negativicutes*), and the phyla *Acinetobacteria* and *Fusobacteria*. Marker genes for these species can be aligned using a multiple sequence alignment tool like ClustalW. The SPA algorithm can subsequently be used to identify conserved regions as putative annealing sites for primer candidates by looking for the highest "average sequence variance" scores over 25 nucleotide-long DNA regions among this limited set of sequences. This can be performed as follows:

- Determine the percent of nucleotides for each nucleotide (GATC) at each position.
- Calculate the variance of the percentages at each position.
- Calculate region variance as the average of the variances of each position in the region.

[00109] A completely conserved nucleotide position will have 100% of one nucleotide and 0% for the other three nucleotides, and a variance of 0.25. A completely non-conserved region will have 25% of each nucleotide and a variance of 0. Primer candidates are prioritized based on their "average sequence variance" scores.

[00110] Primer candidates are evaluated for key properties including the level of primer degeneracy and annealing temperature (>50°C). The sequences from the complete curated marker gene database are aligned to these conserved regions to determine their nucleotide compositions. The conservation of their 3' nucleotide (must be >99% conserved among entries) and their "average sequence variance" scores are calculated (highly conserved regions have the highest score) and used to rank and select primer leads, prioritizing primers with the highest score.

[00111] In the next step, a curated marker gene database and an algorithm (referred to as "SPA algorithm" in Figure 4) are used to determine the "average sequence variance" for the regions adjacent to the primer annealing site. Primers with adjacent 25 nucleotide-long and 50 nucleotide-long regions with ideally an average sequence variance of <0.15 and <0.075, respectively, are prioritized based on the lowest score. The algorithm also identifies the resolution of phylogenetic identification for the regions adjacent to each primer lead by determining the number of unique SPA fragments. SPA primers with the highest phylogenetic resolution are added to the SPA primer repository.

[00112] The same approach can be used for other organisms, including fungi and viruses. For fungi, the ITS1 and ITS2 regions can be used as phylogenetic markers.

[00113] Figure 12 shows nucleotide statistics for the *rpoB* gene region 1327-1355 and degenerate sequence (GAYGAYATYGAYCAYYTNGGHAAYCGHGC (SEQ ID NO: 2)) which is the reverse complement sequence recognized by the degenerate primers RpoB1-R1327 and RpoB-R1330. In this specific example, the relative abundance of a nucleotide at a specific position was calculated using the nucleotide sequences of 47,505 aligned unique *rpoB* genes from the PATRIC database and used to design the degenerate sequence, which is provided from 5' to 3' using the following nucleotide codes: A: adenine; G: guanidine; C: cytosine; T: thymine; R: purine (A or G); Y: pyrimidine (T or C); H: not G (A, T or C); N: any nucleotide (A, G, C or T); *: presence of an ambiguous sequence at a specific *rpoB* gene position. The percentages of highly conserved nucleotide sequences used to determine the consensus sequence for the degenerate primer are highlighted. The position of the region is based on the nucleotide sequence of the *Escherichia coli rpoB* gene.

[00114] In the next step, the proposed degenerate primer sequences are matched to the human genome sequence and the number of hits with increased number of allowed mismatches is determined. To minimize annealing to human genomic DNA, a primer should ideally have two or more mismatches with the human genome.

[00115] In one embodiment of the invention, a method is provided for amplifying microbial cell free DNA (mcfDNA), the method including: (a) providing one or more degenerate primers that are complementary to one or more conserved regions, wherein the one or more conserved regions span at least 18 nucleotides of one or more phylogenetic marker genes designated for a set of reference microbes, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the one or more conserved regions comprise a hypervariable region, and wherein the one or more degenerate primers are oriented to prime polymerase extension of the hypervariable region; (b) providing one or more additional primers that are complementary to a repaired version of an adaptor ligated to ends of the mcfDNA; (c) performing, on a sample comprising the mcfDNA, an enrichment polymerase chain reaction (E-PCR) using the one or more degenerate primers and the one or more additional primers, thereby generating one or more enriched mcfDNA fragments; and (d) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer, wherein the modified degenerate primer comprises a 3'-end extension of at least three nucleotides relative to a degenerate primer of the one or more

degenerate primers; and (ii) the one or more additional primers, thereby generating enriched amplified mcfDNA fragments.

[00116] In another embodiment of the invention, a method is provided for amplifying microbial cell free DNA (mcfDNA), that includes: performing an enrichment polymerase chain reaction (E-PCR) on a sample comprising microbial cell-free DNA (mcfDNA) wherein enriched mcfDNA fragments are generated using: (i) one or more degenerate primers comprising complementarity to one or more conserved regions, wherein the one or more conserved regions span at least 18 nucleotides of one or more phylogenetic marker genes designated for a set of reference microbes, and wherein at least one of the one or more degenerate primers comprises an affinity purification group, and (ii) an additional primer comprising complementarity to an end of the mcfDNA. In the method, at least 25 adjacent nucleotides upstream or downstream of an end of the one or more conserved regions comprise a hypervariable region, the one or more degenerate primers are oriented to prime polymerase extension of the hypervariable region, and wherein enriched amplified mcfDNA fragments are generated in the E-PCR. The method includes performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using: (i) a modification of at least one of the one or more degenerate primers of the E-PCR, wherein the modified degenerate primer comprises a 3'-end extension of at least three nucleotides relative to a degenerate primer of the one or more degenerate primers of the E-PCR, and (ii) the additional primer of the E-PCR, thereby generating enriched amplified mcfDNA fragments.

[00117] The methods of the invention may further include sequencing the enriched amplified mcfDNA fragments.

[00118] The methods of the invention may further include performing an affinity purification reaction on the enriched mcfDNA fragments.

[00119] In one aspect, the methods of the invention include, using a computer system, a step: (a) aligning the amplified mcfDNA fragment sequences on a sequence of the one or more degenerate primers and assigning matching sequences from the hypervariable region as representative of the same microbial species; (b) for the microbial species in part (a), searching a database of the one or more phylogenetic marker genes against the amplified mcfDNA fragment sequences and assigning a microbial species based on the closest match; and (c) optionally, for the one or more phylogenetic marker genes, calculating a microbial community composition based on the relative abundance of the amplified mcfDNA fragment sequences assigned to each microbial species.

[00120] In the methods of the invention, the phylogenetic marker gene of the one or more phylogenetic marker genes can be a multicopy gene, in which case the method includes correcting for copy number variation between each species. In some embodiments, there are two or more phylogenetic marker genes, and the method includes determining the microbial community composition by calculating a mathematical mean of the relative abundance of each microbial species for each of the two or more phylogenetic marker genes.

[00121] The microbial community composition can include one or more members of Eukaryotes, bacteria, viruses, or fungi.

[00122] In some instances of the methods of the invention, one or a combination of: the microbial community composition or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species associated with a disease or disorder, indicates that the subject has the disease or disorder.

[00123] In one embodiment, a system is provided for amplifying microbial cell free DNA (mcfDNA), including: a reaction vessel; a reagent dispensing module; and software to execute the method of any of the methods provided herein, wherein the method is executed at least partially robotically.

[00124] In another embodiment, a set of degenerate oligonucleotide primers is provided comprising a mixture of oligonucleotides having a V:W:X:Y:Z ratio of 1.2:0.8-1.2:0.8-1.2:0.8-1.2:1.5-3, wherein V is the sequence 5' to 3' CGRTTDCCNARRTGRTCRATRTCATC (SEQ ID NO: 24), wherein W is the sequence 5' to 3' CGRTTDCCNARRTGRTCRATATCATC (SEQ ID NO: 25), wherein X is the sequence 5' to 3' CGRTTDCCNARGTGRTCRATRTCATC (SEQ ID NO: 26), wherein Y is the sequence 5' to 3' CGRTTDCCNARGTGRTCRATATCATC (SEQ ID NO: 27), and wherein Z is the sequence 5' to 3' CGGTTRCCNARGTGGTCGATRTCGTC (SEQ ID NO: 28), wherein A = adenine, G = guanidine, C = cytosine, T = thymine, R = purine (A or G), D = not C (A, T or G), and N = any nucleotide (A, G, C or T), and wherein D and N can be replaced with inosine.

[00125] In another aspect, a set of degenerate oligonucleotide primers is provided comprising a mixture of oligonucleotides having a V:W:X:Y:Z ratio of 1.2:0.8-1.2:0.8-1.2:0.8-1.2:1.5-3, wherein V is the sequence 5' to 3' GCRCGRTTDCCNARRTGRTCRATRTC (SEQ ID NO: 19), wherein W is the sequence 5' to 3' GCRCGRTTDCCNARRTGRTCRATATC (SEQ ID NO: 20), wherein X is the sequence 5' to 3' GCRCGRTTDCCNARGTGRTCRATRTC (SEQ ID NO: 21), wherein Y is the sequence 5' to 3' GCRCGRTTDCCNARGTGRTCRATATC (SEQ ID NO: 22), and wherein Z is the

sequence 5' to 3' GCRCGGTTRCCNARGTGGTCGATRTC (SEQ ID NO: 23), wherein A = adenine, G = guanidine, C = cytosine, T = thymine, R = purine (A or G), D = not C (A, T or G), and N = any nucleotide (A, G, C or T), wherein D and N can be replaced with inosine, and wherein the primers optionally comprise an affinity purification group at a 5' terminus.

[00126] In the degenerate oligonucleotide primers, one or more nucleotides at positions represented by D or N can be replaced by inosine.

[00127] A kit is provided for amplifying microbial cell free DNA (mcfDNA). The kit includes: (a) an adaptor for ligating to ends of mcfDNA; (b) a degenerate primer comprising complementarity to a conserved region, wherein the degenerate primer comprises an affinity purification group, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region; (c) an additional primer complementary to a repaired version of the adaptor; (d) a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer; and instructions for performing: (i) an enrichment polymerase chain reaction (E-PCR) on the mcfDNA having the adaptor-ligated ends with the degenerate primer and the additional primer to generate enriched mcfDNA fragments, an (ii) affinity purification of the enriched mcfDNA fragments, and (iii) an amplification PCR (A-PCR) on the enriched mcfDNA fragments using the modified degenerate primer and the additional primer to generate enriched amplified mcfDNA fragments.

[00128] The kit can further include one or more additional degenerate primers comprising complementarity to one or more additional conserved regions of the microbial phylogenetic marker gene or to one or more additional genes. The one or more additional genes can be one or more additional microbial phylogenetic marker genes. In another example, the one or more additional genes can encode virulence factors, secondary metabolites, or drug degradation pathways, and combinations thereof.

[00129] A method is provided for diagnosing or monitoring a disease or disorder. The method includes (a) providing one or more degenerate primers that are complementary to one or more conserved regions, wherein the one or more conserved regions span at least 18 nucleotides of one or more phylogenetic marker genes designated for a set of reference microbes, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the one or more conserved regions comprise a hypervariable region, and wherein the one or more degenerate primers are oriented to prime polymerase extension of the

hypervariable region; (b) providing one or more additional primers that are complementary to a repaired version of an adaptor ligated to ends of the mcfDNA; (c) performing, on a sample comprising the mcfDNA, an enrichment polymerase chain reaction (E-PCR) using the one or more degenerate primers and the one or more additional primers, thereby generating one or more enriched mcfDNA fragments; (d) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer, wherein the modified degenerate primer comprises a 3'-end extension of at least three nucleotides relative to a degenerate primer of the one or more degenerate primers;, and (ii) the one or more additional primers, thereby generating enriched amplified mcfDNA fragments; (e) sequencing or obtaining sequences of the enriched amplified mcfDNA fragments; (f) assigning one of more of the amplified mcfDNA fragment sequences as belonging to a particular microbial species, or as comprising a microbial DNA signature, based on the closest sequence match of the hypervariable region to a reference microbial database; and (g) optionally, calculating a microbial community composition for the sample, based on the relative abundance of the mcfDNA fragment sequences assigned to each microbial species, wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[00130] In another embodiment, method is provided of diagnosing or monitoring a disease or disorder, that includes: performing, on a sample comprising microbial cell-free DNA (mcfDNA), a polymerase chain reaction (PCR) using: (i) one or more degenerate primers comprising complementarity to one or more conserved regions, wherein the one or more conserved regions span at least 18 nucleotides of one or more phylogenetic marker genes designated for a set of reference microbes and (ii) an additional primer comprising complementarity to a repaired version of an adaptor ligated to ends of the mcfDNA. In the method, at least 25 adjacent nucleotides upstream or downstream of an end of the one or more conserved regions comprise a hypervariable region, and the one or more degenerate primers are oriented to prime polymerase extension of the hypervariable region, wherein amplified mcfDNA fragments are generated in the PCR. The method includes sequencing the amplified mcfDNA fragments or obtaining the sequences of the amplified mcfDNA fragments and assigning one or more of the amplified mcfDNA fragment sequences as comprising a microbial DNA signature or as belonging to a particular microbial species based on the closest sequence match of the hypervariable region to a reference microbial database.

Optionally, the method includes calculating a microbial community composition for the sample, based on the relative abundance of the mcfDNA fragment sequences assigned to each microbial species. In the method, one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[00131] In one aspect, a method is provided of diagnosing or monitoring a disease or disorder that includes: providing mcfDNA sequences, wherein the mcfDNA sequences are derived from a sample, wherein the mcfDNA sequences comprise sequence corresponding to (i) a conserved region spanning at least 18 nucleotides of at least one microbial phylogenetic marker gene and (ii) a hypervariable region at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region. The method includes assigning one or more of the mcfDNA fragment sequences as comprising a microbial DNA signature or as belonging to a particular microbial species based on the closest sequence match of the hypervariable region to a reference microbial database. Optionally, the method may include calculating a microbial community composition for the sample, based on the relative abundance of the mcfDNA fragment sequences assigned to each microbial species. In the method, one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[00132] In another aspect, a method of diagnosing or monitoring a disease or disorder is provided, the method including sequencing amplified mcfDNA fragments derived from a sample, wherein the mcfDNA sequences comprise sequence corresponding to (i) a conserved region spanning at least 18 nucleotides of at least one microbial phylogenetic marker gene and (ii) a hypervariable region at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region. The method includes assigning one or more of the amplified mcfDNA fragment sequences as comprising a microbial DNA signature or as belonging to a particular microbial species based on the closest sequence match of the hypervariable region to a reference microbial database. The method can optionally include calculating a microbial community composition for the sample, based on the relative abundance of the mcfDNA fragment sequences assigned to each microbial species. In the method, one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a

combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

[00133] In one embodiment, a method is provided of diagnosing or monitoring a disease or disorder, that includes: performing, on a sample comprising microbial cell-free DNA (mcfDNA), a polymerase chain reaction (PCR) using: (i) one or more degenerate primers comprising complementarity to one or more conserved regions, wherein the one or more conserved regions span at least 18 nucleotides of one or more phylogenetic marker genes designated for a set of reference microbes and (ii) an additional primer comprising complementarity to a repaired version of an adaptor ligated to ends of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the one or more conserved regions comprise a hypervariable region, and the one or more degenerate primers are oriented to prime polymerase extension of the hypervariable region, wherein amplified mcfDNA fragments are generated in the PCR. The method includes sequencing the amplified mcfDNA fragments or obtaining the sequences of the amplified mcfDNA fragments. In the method, the closer the sequence match between the amplified mcfDNA fragment sequences and the nucleic acid sequences in a reference sample the more likely the sample is to have the disease or disorder.

[00134] In some instances of the methods provided herein for diagnosing or monitoring a disease or disorder, the disease or disorder comprises detection of infection or inflammation in one or a combination of Irritable Bowel Disease, Crohn's disease, Ulcerative colitis, rheumatoid arthritis, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis, or plaque psoriasis. In these instances of the methods, a presence of the amplified mcfDNA fragments from one or more pathogenic bacterial members of: Escherichia species, Escherichia coli, Shigella species, Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral, Pseudomonas species, Pseudomonas aeruginosa, Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species, Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium, indicates that the subject has the infection or inflammation.

[00135] A method of detecting infection or inflammation is provided, the method comprising: (a) performing, on a sample comprising mcfDNA from a subject, an

enrichment polymerase chain reaction (E-PCR) using one or more degenerate primers and one or more additional primers, thereby generating one or more enriched mcfDNA fragments, wherein: (i) the one or more degenerate primers are complementary to one or more conserved regions, wherein the one or more conserved regions span at least 18 nucleotides of one or more phylogenetic marker genes designated for a set of reference microbes, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the one or more conserved regions comprise a hypervariable region, and wherein the one or more degenerate primers are oriented to prime polymerase extension of the hypervariable region, and (ii) the one or more additional primers are complementary to a repaired version of an adaptor ligated to ends of the mcfDNA; (b) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer, wherein the modified degenerate primer comprises a 3'-end extension of at least three nucleotides relative to a degenerate primer of the one or more degenerate primers; and (ii) the one or more additional primers, thereby generating enriched amplified mcfDNA fragments; (c) sequencing or obtaining sequences of the enriched amplified mcfDNA fragments; and (d) assigning one of more of the amplified mcfDNA fragment sequences as belonging to one or more pathogenic bacterial species based on the closest sequence match of the hypervariable region to a reference microbial database, wherein the one or more pathogenic bacterial species comprise Escherichia species, Escherichia coli, Shigella species, Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral, Pseudomonas species, Pseudomonas aeruginosa, Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species, Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium. In this method, a presence of the enriched amplified mcfDNA fragments from one or more of the one or more pathogenic bacterial members indicates that the subject has the infection or inflammation.

[00136] In another embodiment, a method of detecting infection or inflammation is provided, the method comprising: performing, on a sample comprising microbial cell-free DNA (mcfDNA) from a subject, a polymerase chain reaction (PCR) using: (i) one or more degenerate primers comprising complementarity to one or more conserved regions, wherein the one or more conserved regions span at least 18 nucleotides of

one or more phylogenetic marker genes designated for a set of reference microbes and (ii) an additional primer comprising complementarity to a repaired version of an adaptor ligated to ends of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the one or more conserved regions comprise a hypervariable region, and the one or more degenerate primers are oriented to prime polymerase extension of the hypervariable region, thereby generating amplified mcfDNA fragments; sequencing the amplified mcfDNA fragments or obtaining the sequences of the amplified mcfDNA fragments; and assigning one or more of the amplified mcfDNA fragment sequences as belonging to one or more pathogenic bacterial species based on the closest sequence match of the hypervariable region to a reference microbial database, wherein the one or more pathogenic bacterial species comprise Escherichia species, Escherichia coli, Shigella species, Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral, Pseudomonas species, Pseudomonas aeruginosa, Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species, Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium, In this embodiment, a presence of the enriched amplified mcfDNA fragments from one or more of the one or more pathogenic bacterial members indicates that the subject has the infection or inflammation.

[00137] In the methods provided herein, the disease or disorder can include detection of infection or inflammation in Irritable Bowel Disease and the one or more pathogenic bacterial members can further include one or more of *Rhuminococcus gnavus*, *Enterococcus faecium* or *Enterococcus faecalis*.

[00138] The methods provided herein for detecting infection or inflammation can include calculating a percentage of the amplified mcfDNA fragments coming from the one or more bacterial pathogenic members as a fraction of a total amount of the amplified mcfDNA fragments in the sample, wherein the calculated percentage of less than one percent indicates an absence of infection or inflammation and the calculated percentage of greater than one percent indicates the presence of infection or inflammation.

[00139] In the methods provided herein for detecting infection or inflammation, the presence of infection or inflammation can be in the gastrointestinal tract.

[00140] The methods for detecting infection or inflammation provided herein can be used in combination with determination of a level of lipo-polysaccharide (LPS) in the sample.

[00141] In the methods of diagnosing or monitoring the disease or disorder provided herein, the disease or disorder may include tuberculosis and other diseases caused by *Mycobacterium* species; determining pulmonary infection risks and causes in cystic fibrosis patients; determining the risk and onset of sepsis in patients with compromised immune systems; detection of opportunistic bacterial pathogens originating from the oral cavity that have been linked to Alzheimer's disease, pancreatic cancer and other serious conditions such as endocarditis; women's health issues including Chlamydia linked to mucopurulent cervicitis, pelvic inflammatory disease, tubal factor infertility, ectopic pregnancy and cervical cancer; detection and monitoring of progression of cancer; monitoring of minimal residual disease after oncology treatments; detection and monitoring of progression and minimal residual disease of breast cancer including triple negative breast cancer; detection of esophageal cancer, precancerous colonic polyps and early stage colorectal cancer, and detection and monitoring of progression and minimal residual disease of gastrointestinal cancers in general; detection and monitoring of progression and minimal residual disease in lung cancer; non-invasive analysis of the microbiome in pancreatic cancer patients to propose treatment protocols and prognostics for long-term survival; detection of *Clostridium difficile* infections; post-transplant bloodstream infections and Graft versus Host Disease (GvHD); detection of hospital acquired infections by emerging pathogens of clinical concern; detection of an infection in an immune compromised person; or detection of infection or inflammation of the gastrointestinal track in Irritable Bowel Disease (Crohn's disease, Ulcerative colitis); monitoring the risk of infection of inflammation in patients suffering from rheumatoid arthritis, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis, plaque psoriasis; and combinations thereof.

[00142] In some instances of methods provided herein, the disease or disorder can include colon cancer, breast cancer, or non-small cell lung cancer. The disease or disorder can include a specific type of cancer, an early or later stage cancer, or a specific tumor subtype driven by the location of the tumor, and combinations thereof.

[00143] In another aspect, including one or a combination of circulating tumor DNA (ctDNA) signatures, ctDNA methylation patterns, or gene mutations that provide information on the disease or disorder in the methods provided herein can indicate whether the sample is likely to have the disease or the disorder.

[00144] In the methods provided herein, the disease or disorder can include IBD. The disease or disorder can include Crohn's disease or ulcerative colitis.

[00145] In one aspect, calculating the blood pathogenicity index (BPI) is used to monitor disease progression, medication efficiency and medication side effects in IBD patients.

[00146] In another aspect, the methods provided herein can be used with one or a combination of inflammatory biomarkers in stool and blood, such as the stool calprotectin levels or C-reactive protein (CRP), erythrocyte sedimentation rate (ESR) and plasma viscosity (PV) in blood that provide information on the disease or disorder. In the methods provided herein, the combination with these inflammatory biomarkers can indicate whether the sample is likely to have the disease or the disorder.

[00147] In another aspect, the BPI can be used in combination with alternative microbial biomarkers, such as lipo-polysaccharide (LPS) levels.

[00148] In the methods and kits of the invention, the additional primer in the A-PCR can be used with a set of at least two phased primers, wherein the phased primers are identical to the additional primer except that the phased primers are phased by a single nucleotide. In one instance, the set of phased primers comprises five or more phased primers.

[00149] In the methods and kits provided herein, the conserved region of the microbial phylogenetic marker gene can have an average sequence variance score of greater than 0.175. The hypervariable region can have an average sequence variance score of less than 0.15, 0.1, or 0.075. The conserved region can span 18 to 40 nucleotides, 20 to 30 nucleotides, or 22 to 28 nucleotides of the phylogenetic marker gene.

[00150] In the methods and kits provided herein, the at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region can include less than 150 adjacent nucleotides, less than 75 adjacent nucleotides, or less than 50 adjacent nucleotides.

[00151] In the methods and kits provided herein, the ends of the mcfDNA can include an adaptor and the additional primer is complementary to a repaired version of the adaptor. The adaptor can be a double stranded asymmetric linker cassette comprising a 5' asymmetrical end and a 3' end where the two strands are complementary. The asymmetric linker cassette can be a Y-shaped linker cassette or a single arm linker cassette. In one example, the additional primer is complementary to a repaired 5' end of the asymmetric linker cassette and, in the E-PCR, polymerase extension from the degenerate primer results in repair of the asymmetric linker cassette.

[00152] In the methods and kits provided herein, the mcfDNA can be an average fragment length of less than about 100 bp. The percentage of the mcfDNA in the sample can be less than about 0.05%, less than about 0.1%, less than about 1%, less than about 5%, or less than about 15%.

[00153] In the methods and kits provided herein, the set of reference microbes can be eubacterial microbes. In other instances, the set of reference microbes are eukaryotic, fungal, viral, or bacterial.

[00154] The methods and kits provided herein can include performing one or more reactions to repair the ends of the mcfDNA. The degenerate primer can include one or more sequencing adaptor sequences. In another embodiment, the method can include adding one or more sequencing adaptor sequences to the amplified mcfDNA fragments in a second amplification reaction or to the enriched mcfDNA fragments in the A-PCR.

[00155] In some instances of the methods and kits provided herein, themicrobial phylogenetic marker gene comprises *rpoB*. In another example, the microbial phylogenetic marker gene comprises 16S rRNA. In other instances, the microbial phylogenetic marker gene comprises a combination of two or more of *rpoB*, *cpn60*, *tuf*, or 16S rRNA. The phylogenetic marker gene can include *rpoB* and the conserved region can include nucleotide positions 1327 – 1355 based on the *Escherichia coli rpoB* gene sequence. In another example, the microbial phylogenetic marker gene comprises 16S rRNA and the conserved region comprises nucleotide positions 781-805 based on the *Escherichia coli* 16S rRNA gene sequence.

[00156] In the methods and kits provided herein, the degenerate primer can include one or both of RpoB-ePCR mix and RpoB-aPCR mix. In one instance, the degenerate primer comprises one or both of 16SV4-ePCR and 16SV4-aPCR. In another instance, the microbial phylogenetic marker gene comprises 16S rRNA and the conserved region comprises a V1, V2, V3, V4, V5, or V6 region of the 16S rRNA phylogenetic marker gene. In one example, the degenerate primer can include RpoB-ePCR mix, RpoB-aPCR mix, 16SV4-ePCR, or 16SV4-aPCR, and combinations thereof.

[00157] In the methods and kits provided herein, the microbial phylogenetic marker gene can include DNA gyrase subunit B (*gyrB*), heat shock protein 60 (*hsp60*), superoxide dismutase A protein (*sodA*), TU elongation factor (*tuf*), DNA recombinase proteins (including *recA, recE*), *trr1* gene that encodes for thioredoxin reductase; *rim8* gene that encodes for a protein involved in the proteolytic activation of a transcriptional factor in

response to alkaline pH; *kre2* gene that encodes for α-1,2-mannosyltransferase; or *erg6* gene that encodes for Δ(24)-sterol C-methyltransferase.

[00158] In the methods and kits provided herein, the phylogenetic marker gene can be a fungal phylogenetic marker gene or a human fungal phylogenetic marker gene. In one instance, the human fungal phylogenetic marker gene comprises nuclear ribosomal internal transcribed spacer region 1 (ITS1) or nuclear ribosomal internal transcribed spacer region 2 (ITS2). The amplified mcfDNA fragments or enriched amplified mcfDNA fragments can include mcfDNA from one or a combination of members of the Ascomycota, Basidiomycota and Mucoromycota, including *Alternaria* species, *Aspergillus* species, *Blastomyces* species, *Candida* species, *Capnodiales* species, *Cladosporium* species, *Malassezia* species, *Phaeosphaeria* species, *Pseudozyma* species, *Saccharomyces* species, *Sporobolomyces* species, *Vishniacozyma* species, and *Yarrowia* species.

[00159] The methods provided herein for amplifying mcfDNA can include adding, in the PCR or in both the E-PCR and the A-PCR, a primer to a functional gene to determine the presence of the functional gene in the sample. The functional gene can be a pathogenicity factor, a PKS gene cluster essential for colibactin synthesis, or a choline trimethylaminelyase gene.

[00160] In another example, the methods for amplifying mcfDNA can include adding, in the PCR or in both the E-PCR and the A-PCR a primer to a viral gene to determine the presence of the viral gene in the sample. The viral gene can include a human DNA- or RNA-based oncovirus gene. The oncovirus can be one or a combination of Epstein-Barr Virus (EBV), Human Papillomavirus (HPV), Hepatitis B virus (HBV), Human Herpesvirus-8 (HHV-8), or Merkel Cell Polyomavirus (MCPyV).

[00161] In the methods provided herein for amplifying mcfDNA, the sample can include a bodily fluid, a tissue, or an extracellular bodily substance. The bodily fluid can include whole blood, a blood fraction, serum, plasma, or combinations thereof. In some instances, the sample includes a biopsy sample from a solid tumor, a skin graft, a liquid biopsy sample other than blood, or combinations thereof. In another example, the sample includes a stool sample.

[00162] In the methods provided herein, the amplified mcfDNA fragments or the enriched amplified mcfDNA fragments can include mcfDNA from one or more members of: *Flavobacterium sp., Staphylococcus auricularis, Pseudomonas toyotomiensis, Rheinheimera sediminis, Finegoldia magna, Parvularcula sp., Pseudomonas stutzeri, Pseudomonas soyae, Pseudomonas saponiphila, Pseudomonas sp., Peptoniphilus harei,*

Quisquiliibacterium sp., Azoarcus sp., Sphingopyxis terrae, uncultured Clostridiales bacterium strain UMGS460, Staphylococcus schweitzeri, Flavobacterium erciyesense, Rhodococcus yananensis, Dietzia massiliensis, Cutibacterium acnes subsp. elongatum, Angustibacter aerolatus, Aerococcus urinae, Klebsiella quasivariicola, Comamonas fluminis, Mycobacterium tuberculosis, Mycobacterium abscessus, Mycobacterium avium, Mycobacterium chimaera, Mycobacterium leprae, Mycobacterium xenopi, Mycobacterium (para)intracellulare, Mycobacterium kansasii, Mycobacterium gilvum, Mycolicibacterium gen. nov. ("fortuitum-vaccae" clade), Mycobacterium gen. ("tuberculosis-simiae" clade), Staphylococcus aureus, Staphylococcus argenteus, Staphylococcus schweitzeri, Pseudomonas aeruginosa, Burkholderia cepacia complex, Burkholderia ubonensis, Burkholderia species Nov., Burkholderia multivorans, Burkholderia pseudomultivorans, Burkholderia pseudomallei, Burkholderia mallei, Trinickia species, *Burkholderia thailandensis, Haemophilus* species, *Haemophilus influenzae, Haemophilus parainfluenzae, Streptococcus* species at the various group and species levels, *Streptococcus dysgalactiae, Streptococcus pyogenes, Streptococcus mutans, Streptococcus suis, Streptococcus mitis, Streptococcus pneumoniae, Streptococcus agalactiae, Streptococcus anginosus, Streptococcus intermedius, Streptococcus constellatus, Streptococcus equi subsp. zooepidemicus, Streptococcus oralis, Streptococcus gordonii, Streptococcus uberis, Streptococcus parasanguinis, Streptococcus sanguinis Streptococcus parauberis, Streptococcus infantarius, Streptococcus iniae, Streptococcus salivarius, Streptococcus thermophilus, Streptococcus vestibularis, Streptococcus bovis, Streptococcus gallolyticus subsp. gallolyticus, Streptococcus gallolyticus subsp. macedonicus, Streptococcus gallolyticus subsp. pasteurianus, Streptococcus equinus, Enterococcus faecalis, Enterococcus faecium, Porphyromonas gingivalis, Porphyromonas cangingivalis, Porphyromonas uenonis, Porphyromonas endodontalis, Propionibacterium acidifaciens, Porphyromonas asaccharolytica, Porphyromonas macacae, Prevotella pallens, Prevotella histicola, Prevotella melaninogenica, Prevotella copri, Prevotella intermedia, Prevotella oral, Prevotella nanceiensis, Prevotella salivae, Prevotella nigrescens, Prevotella denticola, Prevotella buccae, Prevotella stercorea, Prevotella oris, Prevotella disiens, Prevotella bryantii, Prevotella shahii, Tannerella forsythia, Bacteroides fragilis, Helicobacter pylori, Chlamydia trachomatis, Neisseria meningitidis, Neisseria gonorrhoeae, Neisseria subflava, Neisseria perflava, Neisseria flavescens, Neisseria cinerea, Neisseria lactamica, Neisseria weaver, Neisseria zoodegmatis, Neisseria brasiliensis, Neisseria mucosa, Neisseria animaloris, Aggregatibacter actinomycetemcomitans, Aggregatibacter aphrophilus,*

Aggregatibacter segnis, Saccharopolyspora species, *Bacillus clausii,* members of the genera *Pseudoxanthomonas* and *Streptomyces, Fusobacterium nucleatum subsp. polymorphum, Fusobacterium hwasookii, Fusobacterium canifelinum, Fusobacterium nucleatum subsp. animalis, Fusobacterium periodonticum, Fusobacterium necrophorum subsp. funduliforme, Fusobacterium mortiferum, Fusobacterium varium, Fusobacterium nucleatum subsp. nucleatum, Fusobacterium ulcerans, Fusobacterium nucleatum subsp. vincentii, Fusobacterium equinum, Fusobacterium gonidiaformans, Fusobacterium necrogenes, Fusobacterium naviforme, Peptostreptococcus stomatis, Pseudonocardia asaccharolytica, Parvimonas species* including *Parvimonas oral* and *Parvimonas micra, Gemella* species including *Gemella morbillorum, Gemella haemolysans, Gemella palaticanis* and *Gemella sanguinis, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Acinetobacter baumannii, Acinetobacter lactucae, Acinetobacter pittii, Acinetobacter calcoaceticus, Acinetobacter oleivorans, Acinetobacter nosocomialis, Acinetobacter radioresistens, Acinetobacter variabilis, Acinetobacter courvalinii, Acinetobacter ursingii, Enterobacteriaceae*, *Enterobacter* species, *Escherichia* species, *Escherichia coli,* or *Klebsiella* species, *Klebsiella pneumoniae, Klebsiella oxytoca*, *Shigella* species, *Kluyvera* species, *Erwinia* species, *Salmonella* species, *Salmonella typhi, Salmonella enterica, Listeria* species, *Listeria monocytogenes, Veillonella* species, *Veillonella atypica, Veillonella dispar, Veillonella oral, Stenotrophomonas* species, *Stenotrophomonas maltophilia*, and combinations thereof.

[00163] Various modifications and variations of the disclosed methods, compositions, and uses of the invention will be apparent to the skilled person without departing from the scope and spirit of the invention. Although the invention has been disclosed in connection with specific preferred aspects or embodiments, the invention as claimed should not be unduly limited to such specific aspects or embodiments.

[00164] The present invention may be implemented using hardware, software, or a combination thereof and may be implemented in one or more computer systems or other processing systems. In one aspect, the invention is directed toward one or more computer systems capable of carrying out the functionality described herein.

EXAMPLES

[00165] The following Examples have been included to provide guidance to one of ordinary skill in the art for practicing representative embodiments of the presently disclosed subject matter. In light of the present disclosure and the general level of skill in the art, those

of skill can appreciate that the following Examples are intended to be exemplary only and that numerous changes, modifications, and alterations can be employed without departing from the scope of the presently disclosed subject matter.

EXAMPLE 1

SPA fragment sequencing protocol using the V4 region of the 16S rRNA gene as phylogenetic marker

[00166] The protocol for SPA fragment amplification and sequencing was targeting the V4 hypervariable region of the 16S rRNA gene. Several iterations were used for protocol improvement; for each iteration, blood plasma from the same healthy donor was used. Figure 2 shows the order of the steps for the SPA fragments sequencing protocol.

[00167] cfDNA isolation was performed using the Qiagen QIAamp ccfDNA/RNA Kit on 1.0 ml blood plasma from a healthy volunteer. To confirm the presence of mcfDNA in the blood sample, total cfDNA was isolated on 1.0 ml blood plasma and deep sequencing was used to determine the percentage of mcfDNA. In the case of this healthy donor, the percentage of mcfDNA was approximately 0.1% to 1.0% of the total cfDNA (data not shown). This is considerably lower than typically found in blood samples from cancer patients, where this ranged between approximately 1% to 4% (Poore et al, 2020).

[00168] Subsequently, the SPA fragment sequencing protocol was followed as detailed below. In the protocol critical steps that resulted in major improvements are indicated, with the effect of the improved step summarized in Table 2:

1. **End-repair and A-tailing, using the IDT: XGEN DNA Lib Prep kit (8 samples)**
    - Prepare the Reaction according to the table below. First prepare a Premix containing all reagents in a 1.5 mL LoBind tube on ice and aliquot 12 µl into 8 PCR tubes. Transfer 18 µl of each cfDNA sample into the tubes and mix by pipetting 5 times.

IDT xGen Reagent	**volume per reaction (µl)**	**Premix volume for 8.8 reactions (µl)**
Low EDTA TE	8.5	74.8
Buffer E1	2.35	20.7
Enzyme E2	0.5	4.4
Enzyme E3	0.5	4.4
Enzyme E4	0.15	1.3
cfDNA	18	

TOTAL	30	105.6

- Place the reactions in the thermocycler and run the following program: heated lid, 30 min at 20ºC, 30 min at 65ºC, hold at 4ºC.

2. SPA-linker ligation (8 samples)

- The SPA-linker is ordered as duplexed DNA fragments. In Table 1, examples of two SPA-linkers are provided, SPA-linker-UMI.v1-Y-SP, which was used in protocol iterations 1 to 7, and SPA-linker-UMI.v1-Y-SP, which was used for protocol iteration 8.
- Prepare the Reaction according to the table below. First prepare a Premix containing all reagents in a 1.5 mL LOBIND tube on ice. Add 15 µl of the Premix to each tube containing 30 µl of repaired/A-tailed DNA (from step 1).

IDT xGen Reagent	**volume per reaction (µl)**	**Premix volume for 8.8 reactions (µl)**
Buffer L1	9	79.2
Enzyme L2	3	26.4
SPA-linker (1 nM)	1	8.8
Low EDTA TE	2	17.6
End-repaired DNA	30	
TOTAL	45	132

- Place the reactions in the thermocycler and run the following program: heated lid off, 15 min at 20ºC.

3. AMPure bead cleanup at 2X (8 samples)

- Let the AMPure beads equilibrate at RT and vortex to resuspend the beads.
- Add **90 µl of beads** into each tube from step 2, mix by pipetting 10 times and incubate at RT for 5 min. The **90 µl of beads** is twice the reaction volume present in the tube.
- Place the tubes on a magnetic stand for 1 min. Remove and discard the supernatant.
- Keep the tubes on the magnetic stand and add 200µL of 80% ethanol (prepared in nuclease-free water). Incubate at RT for 30 seconds. Remove and discard the ethanol supernatant. Repeat step for total two washes.
- Remove the tubes from the magnetic stand. Add 13 µL of nuclease-free water, mix by pipetting 10 times and incubate at RT for 1 min.
- Place the tubes on the magnetic stand for 1 min and transfer 11.5 µl of the eluates into new 0.2 mL PCR tubes.

4. Enrichment PCR (8 samples)

The E-PCR reaction was introduced during protocol iteration 3. Together with the A-PCR reaction it results in enrichment of target fragments and reduces the level of fragments amplified as the result of non-specific primer hybridization. The E-PCR reaction was further improved during protocol iteration 7 by using the biotinylated 16S-V4-R785-Bio primer, which targets position 785 - 805 upstream of the V4 region of the 16S rRNA gene. The linker-ePCR primer is not biotinylated. Non-specific primer hybridization was further reduced by increasing the E-PCR annealing temperature from 55°C to 60°C (protocol iteration 4).

- Prepare the Reaction according to the table below. First prepare a Premix containing all reagents in a 1.5 mL LOBIND tube on ice. Add 13.5 µl of the Premix to each tube containing 11.5 µl purified ligation product from step 3.

Reagents	**volume per reaction (µl)**	**Premix volume for 8.8 reactions (µl)**
2x KAPA HiFi HOTSTART READYMIX	12.5	110
Linker-ePCR primer (10 µM): linker.v1-amp or linker.v2-amp	0.5	4.4
16SV4-ePCR primer (10 µM) Or RpoB-ePCR primer (10µM)	0.5	4.4
Ligation Product	11.5	
TOTAL	25	118.8

- Place the reactions in the thermocycler and run the following program: heated lid on, 95°C for 10 min, [98°C for 20 sec, 60°C (for 16SV4) or 58°C (for RpoB) for 30 sec, 72°C for 15 sec] for 13 cycles, 72°C for 1 min, 4°C on hold.

5. AMPURE bead cleanup at 2X (8 samples)

This step is essential to remove unincorporated biotinylated primer from the E-PCR reaction before streptavidin bead purification.

- Let the AMPURE beads equilibrate at RT and vortex to resuspend the beads.
- Add **50 µl of beads** into each tube (containing 25 µL E-PCR reaction from step 4), mix by pipetting 10 times and incubate at RT for 5 min.
- Place the tubes on a magnetic stand for 1 min. Remove and discard the supernatant.
- Keep the tubes on the magnetic stand and add 200µL of 80% ethanol. Incubate at RT for 30 seconds. Remove and discard the ethanol supernatant. Repeat step for total of two washes.
- Remove the tubes from the magnetic stand. Add 22 µL of nuclease-free water, mix by pipetting 10 times and incubate at RT for 1 min.
- Place the tubes on the magnetic stand for 1 min and transfer 20 µl of the eluates into new 0.2 mL PCR tubes.

6. **MYONE Streptavidin C1 beads purification (8 samples)**

This step removes non-biotinylated DNA fragments, especially human cfDNA, not amplified during the E-PCR reaction and is essential to reduce non-specific PCR amplification during the A-PCR and I-PCR steps (protocol iteration 7).

- Vortex the Streptavidin C1 beads and add 88 µl into a LOBIND.
- Add 200 µl of 1X B&W Buffer and mix by pipetting 10 times. Place the tube on a magnet for 1 min and discard the supernatant.
- Remove the tube from the magnet and resuspend the beads in 88 µl of 1X B&W Buffer by pipetting 10 times. Place the tube on a magnet for 1 min and discard the supernatant. Repeat 2 more times, for a total of 3 washes.
- Resuspend washed beads in 176 µl of 2X B&W Buffer
- Aliquot 20 µl of beads into 0.2 mL PCR tubes (8 total)
- Add 20 µl of purified E-PCR DNA solution (from step 5)
- Incubate for 15 min at room temperature using gentle rotation
- Separate the biotinylated DNA coated beads with a magnet for 2 min.
- Wash the coated beads 3 times with 50 µl of 1X B&W Buffer
- Resuspend the beads in 20 µl of nuclease-free water. The bead suspension is subsequently used as template in the A-PCR reaction (step 7).

7. Amplification PCR (8 samples)

To overcome sequencing quality issues as the result of long stretches of identical sequences shared between sequencing reads that cover the SPA-linker region, a set of six phased primers referred to as the Linker-aPCR primer mix can be used in combination with SPA-linker-UMI.v2-Y-SP for the A-PCR reaction (protocol iteration 8, used in EXAMPLE 4), instead of a single primer in combination with SPA-linker-UMI.v1-Y-SP (protocol iterations 1 to 7, used in EXAMPLES 2, 3 and 5). A schematic overview of the ligation, E-PCR, A-PCR and I-PCR steps is presented in Figure 3.

Prepare the Reaction according to the table below. First prepare a Premix containing all reagents in a 1.5 mL LOBIND tube on ice. Add the 30 µl of the Premix to each tube containing 20 µl of purified E-PCR product (bead solution from step 6).

Reagents	**volume per reaction (µl)**	**Premix volume for 8.8 reactions (µl)**
2x KAPA HiFi HOTSTART READYMIX	25	220
Nuclease-Free Water	3	26.4
Rd1-linker.v1 primer (10 µM) or Linker-aPCR primer mix (10 µM)	1	8.8
16SV4-aPCR primer (10 µM)	1	8.8
E-PCR purified product	20	
TOTAL	50	264

- Place the reactions in the thermocycler and run the following program: heated lid on, 95°C for 10 min, [98°C for 20 sec, 60°C (for 16SV4) or 58°C (for RpoB) for 30 sec, 72°C for 15 sec] for 20 cycles, 72°C for 1 min, 4°C on hold.

8. AMPURE bead cleanup at 2X (8 samples)

- Let the AMPURE beads equilibrate at RT and vortex to resuspend the beads.
- Add 100 µl of beads into each tube (containing 50 µL A-PCR reaction, with streptavidin C1 beads, from step 7), mix by pipetting 10 times and incubate at RT for 5 min.
- Place the tubes on a magnetic stand for 1 min. Remove and discard the supernatant.
- Keep the tubes on the magnetic stand and add 200 µL of 80% ethanol. Incubate at RT for 30 seconds. Remove and discard the ethanol supernatant. Repeat once for a total of two washes.
- Remove the tubes from the magnetic stand. Add 18 µL of nuclease-free water, mix by pipetting 10 times and incubate at RT for 1 min.
- Place the tubes on the magnetic stand for 1 min and transfer 16-17 µl of the eluates into new 0.2 mL PCR tubes.

9. QC: Qubit (24 samples)

- Use 1 µl of each A-PCR sample for quantification.

10. Indexing PCR (24 samples)

This step allows for subsequent sequencing and multiplexing of the SPA libraries. To avoid over amplification of SPA fragments, the total number of amplification cycles for the E-PCR, A-PCR and I-PCR reaction should not exceed 40 to 45 (protocol iteration 5).

- Prepare the Reaction on ice according to the table below, using a unique set of NEXTERA UDI primers for each sample from step 8.
- Place the reactions in the thermocycler and run the following program: heated lid on, 95ºC for 10 min, [95ºC for 30 sec, 55ºC for 30 sec, 72ºC for 30 sec] for 7 cycles, 72ºC for 5 min, 4ºC on hold.

Reagents	volume per reaction (µl)
2x KAPA HiFi HOTSTART READYMIX	25
UDI primers (12.5 µM)	10
A-PCR Product	15
TOTAL	50

11. AMPURE bead size selection at 0.7X, 0.5X, 2X (24 samples)

To improve the quality of the sequencing libraries, it is important to include exclusion of low molecular weight and high molecular weight DNA fragments (protocol iteration 6). During the first incubation, 0.7x volume of beads is used to capture the DNA fragments above ~200 bp which are then eluted from the beads. A second incubation is then performed on this eluate using 0.5x beads volume to capture and remove larger DNA fragments, above ~600 bp, and keep the supernatant containing DNA fragments of ~200-600 bp. A third and final incubation is performed using 2x beads volume to concentrate the DNA in nuclease-free water.

- Let the AMPURE beads equilibrate at RT and vortex to resuspend the beads.
- Add **35 µl of beads** into each tube (containing 50 µL I-PCR reaction from step 10), mix by pipetting 10 times and incubate at RT for 5 min. DNA molecules >200 bp bind to the beads and <200 bp will stay in solution.
- Place the tubes on a magnetic stand for 1 min. Remove and discard the supernatant (DNA <200 bp).
- Keep the tubes on the magnetic stand and add 200µL of 80% ethanol. Incubate at RT for 30 seconds. Remove and discard the ethanol supernatant. Repeat once for total two washes.
- Remove the tubes from the magnetic stand. Add 42 µL of nuclease-free water, mix by pipetting 10 times and incubate at RT for 1 min.
- Place the tubes on the magnetic stand for 1 min and transfer 40 µl of the eluates into new 0.2 mL PCR tubes.
- Add 20 µl of beads into each tube, mix by pipetting 10 times and incubate at RT for 5 min. DNA molecules >600 bp bind to the beads and <600 bp will stay in solution.
- Place the tubes on a magnetic stand for 1 min and transfer 60 µl of supernatants, enriched in DNA molecules of 200–600 bp in size, into new 0.2 mL PCR tubes.
- Add **120 µl of beads** into each tube, mix by pipetting 10 times and incubate at RT for 5 min. This last bead purification step will allow to concentrate the DNA in nuclease-free water.
- Place the tubes on a magnetic stand for 1 min. Remove and discard the supernatant.

- Keep the tubes on the magnetic stand and add 200 µL of 80% ethanol. Incubate at RT for 30 seconds. Remove and discard the ethanol supernatant. Repeat once for a total of two washes.
- Remove the tubes from the magnetic stand. Add 25 µL of nuclease-free water, mix by pipetting 10 times and incubate at RT for 1 min.
- Place the tubes on the magnetic stand for 1 min and transfer 24-23 µl of the eluates into new 0.2 mL PCR tubes.
- Use 1 µl to analyze on the Bioanalyzer and 1 µl to quantify the libraries by qPCR (see QC protocol section).
- Keep the remaining 21-22 µl of the library at 4°C.

12. Library pooling (24 samples) and sequencing

- Using the library concentration from the qPCR results, dilute each library in Dilution Buffer (10 mM Tris-HCl, pH 8.0 + 0.05% Tween 20) to pool at equimolar concentration of 1 nM.

To improve the quality of the sequencing reads, using 20% to 30% PhiX DNA as Spike-IN instead of 5% was introduced during protocol iteration 2.

- Dilute the library pool to 150 pM including 5-30% PhiX DNA as Spike-IN, using Dilution Buffer (10 mM Tris-HCl, pH 8.0 + 0.05% Tween 20) in 100 µl final volume.

For the results described in EXAMPLES 2 to 5 herein below, the library pool was subsequently sequenced following the 150bp paired-end read sequencing protocol from Illumina using the Illumina iSEQ100 i1 Reagent v2 (300-cycle) kit on an Illumina iSEQ100 instrument according to the manufacturer's instructions.

[00169] The analysis of the SPA fragment sequences included the following steps:

1. Adaptors and the primer (16SV4-R781) are trimmed from the sequences.
2. Sequences are mapped to the human genome (GRCh38.p14) using Bowtie2 (Langmead et al, 2012). All sequences that map to the human genome are removed.
3. Using DADA2, an open-source software used for fast and accurate sample inference from amplicon data with single-nucleotide resolution (Callahan et al, 2016), reads are filtered based on read quality.
4. The remaining reads of different lengths are deduplicated, while recording the number of duplicates by sequence for retaining read abundance.

5. The deduplicated reads are aligned to the16S gene database using the basic local alignment search tool (BLAST, Altschul et al, 1990). Only Blast hits where the read starts aligning at the first nucleotide, and where the read aligned at 90% identity over at least 90% of its length are retained along with the taxonomy of the best hit.

[00170] Table 2 summarizes the improvement steps evaluated and links them to two important criteria: the percentage of reads that can be identified, based on sequence homology using Blastn, as 16S rRNA gene derived SPA sequences; and the number of unique 16S rRNA gene derived SPA sequences.

Protocol iteration	**Description of step to improve protocol**	**Percentage of 16S rRNA reads**	**Number of unique reads (16S rRNA)**
1	Original protocol	0.1%	203
2	Increase PhiX DNA as spike-IN from 5% to 20%-30% during sequencing	0.1%	61
3	Include the E-PCR step	3.3%	336
4	Increase PCR annealing temperature in the E-PCR and A-PCR steps to 60°C to reduce background annealing	10.6%	1637
5	Decrease total PCR cycles (E-PCR, A-PCR, I-PCR reactions combined) from 55 to 40	16.9%	3069
6	Include low and high molecular weight DNA exclusion using bead purification after the A-PCR	38%	6724
7	Biotinylate the 16SV4-R785 primer for affinity purification of biotinylated DNA amplicon fragments after the E-PCR	35%	11727
8	During the A-PCR, replace a single primer recognizing the SPA-linker sequence by the phased Linker-aPCR primer mix	This iteration was tested in EXAMPLE 3 in combination with the RpoB-aPCR primer mix.	

Table 2: Description of the various changes implemented in chronological order during subsequent iterations of the SPA fragment sequencing protocol. The effect of each change on improving the ability to generate unique SPA sequencing fragments was evaluated by determining the number of unique reads representing SPA fragment sequences obtained from the 16S rRNA V4 hypervariable region.

[00171] Surprisingly, the improved method resulted in the percentage of 16S rRNA gene derived fragments (as identified by Blastn) increasing from less than 0.01% to an average of 35%, a 3,500-fold increase (see Table 2). Compared to the original protocol, major improvements were made, especially with the introduction of the enrichment (E-PCR) step (protocol iteration 3). Protocol improvements included increasing the PCR annealing temperature (protocol iteration 4), decreasing the total number of PCR cycles to avoid over amplification of fragments (protocol iteration 5), and use of the biotinylated primer for affinity purification (protocol iteration 7) as part of the E-PCR step. Although the use of the biotinylated primer in the E-PCR reaction did not increase the percentage of reads that identified as derived from the 16S rRNA V4 gene region, it significantly increased the number of unique reads, thereby improving the sensitivity of the SPA fragment sequencing method. As described in EXAMPLE 3, the improved SPA fragment sequencing protocol was successfully used for the stratification of cancer patients.

[00172] The improved SPA fragment sequencing protocol can be fully automated in a 96-well format, or multiples thereof, starting with the DNA extraction step to the library quantification step. To facilitate automation the cfDNA isolation step, which us currently done using the column based Qiagen QIAamp ccfDNA/RNA Kit, should preferably be performed using a magnetic bead-based extraction kit, such as the Maxwell® HT ccfDNA Kit (Promega) or the APOSTLE MINIMAX system (Beckman Coulter).

[00173] For full automation, several systems are commercially available such as the Biomek i7 Hybrid NGS Workstation (Beckman Coulter) or the Fluent® platform from Tecan. It was determined that automation should allow to reduce the protocol's time to approximately 8 hours including all steps from cfDNA isolation to having quantified sequencing libraries ready to be loaded on the DNA sequencer.

EXAMPLE 2

Community composition analysis using SPA fragment sequencing.

[00174] The current standard for community composition analysis is 16S rRNA gene-based sequencing. However, this method has several shortcomings, including PCR-based biases resulting in over and/or underrepresentation of species, or the formation of chimeric sequences, resulting in misrepresentation of community composition. This can have major effects when analyzing microbial communities linked to disease, including mcfDNA

signatures in blood originating from tumor microbiomes or indicative for a leaky gut in the case of IBD patients, among others.

[00175] To compare the performance of SPA fragment sequencing with 16S rRNA gene sequencing on low molecular weight DNA fragments with an average size ranging between 80 bp to 250 bp, such as those represented by mcfDNA, water and plasma samples were spiked with ZymoBIOMICS Microbial Community DNA standard MC1 (https://www.zymoresearch.com/collections/zymobiomics-microbial-community-standards). Before spiking, the DNA present in standard MC1 was sheared to an average fragment size of approximately 150 bp.

[00176] Dilution series (10-fold dilution steps) were made of sheared MC1 DNA to a final concentration of 250 pg/µl. Subsequently, 10 µl was added to 0.5 ml blood plasma or 0.5 ml water. DNA from the spiked samples was extracted, resuspended in 30 µl water, and SPA fragment sequencing libraries were prepared using 18 µl DNA solution according to the protocol described in EXAMPLE 1. In addition, 2 µl DNA solution was used for DNA amplification and sequencing library construction using the V3-V4 16S rRNA primer pair or the V6 rRNA primer pair. The primer sequences are shown in Table 1. Libraries were sequenced on an iSEQ100 instrument using the 150 bp paired-end sequencing protocol from Illumina as mentioned in EXAMPLE 1.

[00177] The analysis of the sequences included the following steps:

1. Adaptors and the primers are trimmed from the sequences.
2. Sequences are mapped to the human genome (GRCh38.p14) using Bowtie2 (Langmead et al, 2012). All sequences that map to the human genome are removed.
3. For the libraries constructed using the V3-V4 16S rRNA primer pair or the V6 rRNA primer pair, DADA2 (Callahan et al, 2016) is used to filter reads and create amplicon sequence variants (ASVs) with read counts for community composition.
4. For the SPA fragment sequencing libraries, consensus sequences and community composition are computed:
    a. The UMI (unique molecular identifier) of each read is recorded for downstream deduplication and denoising.
    b. Forward and reverse reads are merged if possible and clustered into bins of similar sequence.
    c. The longest read in each bin is searched against a database of bacterial 16S rRNA genes to aid the assembly into consensus sequences.

d. Consensus sequence abundances are calculated based on the number of reads contributing to each consensus sequence, taking into consideration the number of deduplicated reads identified in step 4a.

5. Amplicon sequence variants (ASV) and consensus sequence genera are assigned with at least 50% confidence.

[00178] It should be noted that for V3-V4 16S amplicon sequencing, after removal of the sequencing adaptors, fragments of approximately 130 bp are obtained that all cover the same positions of the V3 and V4 regions, thereby not allowing to in silico build ASVs beyond 130 bp. In addition, due to their short length, the large majority of mcfDNA fragments derived from the V3-V4 region will either contain the V3 or the V4 16S rRNA gene primer annealing site, but not both. During the PCR amplification step to create the V-V4 amplicon libraries, incomplete single stranded DNA fragments will be synthesized from these short mcfDNA fragments that in subsequent PCR cycles can hybridize if they share a conserved sequence, thereby creating chimeric fragments or fragments with length shorter than the expected 460 bp, as was reported for 20 to 50% of the V3-V4 amplicons (results not shown). The elevated presence of such fragments significantly impacts the accuracy of the community composition.

[00179] The results for the three different methods on water or plasma spiked with 5.0 pg/µl MC1 DNA are shown in Figure 5 and Figure 6. Compared to 16S rRNA gene amplicon sequencing, SPA fragment sequencing provided a significantly better representation of the species present in MC1 when spiked into water or blood plasma. For instance, when MC1 was spiked into blood plasma, bacteria belonging to the genera Enterococcus, Salmonella, Listeria, Bacillus and Staphylococcus were either underrepresented or could not be identified at the genus level based on V3 and V4 16S rRNA gene sequences, thus showing that SPA fragment sequencing has significantly better phylogenetic resolution and quantification of community composition. In addition, fewer contaminating bacterial genera, and if present at significantly lower level, were observed for SPA fragment sequencing compared to V3-V4 amplicon sequencing, showing that SPA fragment sequencing is less sensitive to signal loss resulting from contaminating DNA fragments. This observation was made when comparing the SPA fragment sequencing results with those obtained for both the V3 and the V4 16S rRNA gene sequences. This is likely due to the larger sizes of the contaminating DNA fragments compared to the MC1 spike in DNA in either water or plasma, resulting in their preferred amplification and sequencing when targeting the ~460 bp

V3-V4 region. A comparison that highlights the observations between the results obtained with SPA fragment sequencing or 130 bp long V4 16S rRNA gene sequences on plasma spiked with 5.0 pg/µl MC1 DNA is shown in Figure 7.

[00180] It should also be noted that SPA fragment sequencing showed superior community composition and species identification on blood plasma instead of water. Similar results were obtained for water or plasma spiked with 0.5 pg/µl MC1 DNA.

[00181] To address the effect of small DNA fragment sizes, which disadvantages V3-V4 amplicon sequencing that requires the amplification of a ~460 bp fragment, we compared fragment sequencing with V6 amplicon sequencing, which targets a 97 bp fragment of the 16S rRNA gene V6 region, which is similar in size to the average 120 bp size of the SPA fragments we observed for the MC1 spike in. The results are shown in Figures 5 and 6. Surprisingly, SPA fragment sequencing also significantly outperformed V6 amplicon sequencing, despite both methods targeting DNA fragments with similar sizes. Compared to the other methods, V6 16S rRNA gene amplicon sequencing gave the highest number of sequence reads that could not be assigned to a bacterial genus, making it the least suitable method for the analysis of mcfDNA. It also gave a significant overrepresentation of Escherichia/Shigella but failed to identify the closely related genera Enterococcus and Salmonella.

[00182] Despite its superior performance over 16S rRNA gene amplicon sequencing, a few issues were observed for SPA fragment sequencing. For instance, like for V3-V4 amplicon sequencing, it underrepresented the percentage of *Limosilactobacillus*. This was likely due to a one nucleotide mismatch between the V4R primer sequence and the target sequence. Interestingly, when MC1 was spiked into water, all methods including SPA fragment sequencing strongly underrepresented or failed to identify *Salmonella* (Figures 5 and 6). This is because the sample with MC1 spiked into water contained shorter DNA fragments, allowing for the consensus sequence representing *Salmonella* that was obtained using SPA fragment sequencing to be assigned with only 39% confidence to the genus *Salmonella*. On the other hand, the MC1 sample spiked into blood plasma contained some longer DNA fragments, allowing to *in silico* construct a larger consensus sequence representing Salmonella, which could be assigned with 99% confidence to the genus *Salmonella*. Since the sequence of the V3-V4 16S rRNA gene region of *Salmonella* is very similar to that of other *Enterobacteriaceae*, only SPA fragment sequencing that allows to generate longer consensus sequences that go beyond the 130 bp sizes of the V3 and V4

amplicon sequences can distinguish between Enterobacteriaceae and successfully identify *Salmonella* at levels similar to the MC1 spike (see Figure 7).

EXAMPLE 3

SPA fragment sequencing using the 16S rRNA gene as phylogenetic markers for cancer detection.

[00183] Despite its superior Several studies have successfully shown the existence of distinct cancer-specific intra-tumoral microbiomes (Nejman et al. (2020); Poore et al. (2020)), including lung, prostate, pancreatic, breast and colorectal cancer. The presence of these tumor microbiomes results in unique mcfDNA signatures in the blood, as was shown by Poore et al (2020) using deep sequencing of total cfDNA. This prompted us to determine if SPA fragment sequencing can be used as an alternative method to successfully distinguish between cancer patients based on the location of the tumor. As target gene the multi-copy 16S rRNA gene was chosen for SPA fragment sequencing.

[00184] Using the protocol from EXAMPLE 1 (protocol iteration 7) and targeting the 16S rRNA gene V4 hypervariable region, SPA fragment sequencing was performed on cfDNA isolated from the blood plasma of 8 colorectal cancer (CRC) patients, 12 breast cancer patients, and 12 non-small cell lung cancer patients. All 32 blood plasma samples were obtained from the same hospital. Based on their histology the CRC and non-small cell lung cancers were identified as adenocarcinomas. Metadata for all 32 patients are provided in Table 3.

[00185] Processing of the SPA fragment sequences was performed as described in EXAMPLE 1. Principal component analysis was subsequently used to pairwise compare CRC, breast and lung cancer patients based on their unique SPA fragment sequencing signatures. The results are presented in Figure 8, Figure 9 and Figure 10 and show that mcfDNA signatures obtained with SPA fragment sequencing of the 16S rRNA gene V4 hypervariable region can be successfully used to distinguish between colorectal cancer patients, breast cancer patients and non-small cell lung cancer patients. In pair wise comparisons, patients could be stratified with approximately 90% accuracy as having either CRC or non-small cell lung cancer (18 out of 20; $p<0.05$), or as having either CRC or breast cancer (18 out of 20; $p<0.05$).

[00186] The type of cancer as well as the stage of the cancer are indicated in Figures 8 to 10. The figures illustrate that CRC, breast and non-small cell lung cancer patients with either early or later stage cancer could be correctly stratified. Furthermore, one of the CRC

patients, although correctly diagnosed based on the PCA analysis, was found to be an outlier (datapoint located right lower quadrant of Figure 8). It turned out that the location of the tumor was rectal instead of in the colon, as was the case for the other CRC patients. Thus, the mcfDNA signatures obtained by SPA fragment sequencing are specific enough to distinguish tumor subtypes, e.g. driven by the location of the tumor.

[00187] Analysis of the microbial signatures that drive the PCA-based patient stratification can be used to identify the microbes that provide the signatures underlying patient stratification and used to generate further insights into their role in carcinogenesis and possible treatment strategies. For instance, it was noticed that based on their 16S rRNA gene derived mcfDNA signatures, CRC patients were more distinct than patients suffering from breast and non-small cell lung cancer (see Figures 8 and 9 compared to Figure 10). This is indicative of a distinct tumor microbiome in CRC, as compared to the microbiomes found in breast and non-small cell lung cancer and can be explained by the preferred colonization of CRC tumors by gut microbiome associated bacteria.

[00188] Targeting the 16S rRNA gene V4 hypervariable region, SPA fragment sequencing was performed on cfDNA isolated from the blood plasma of 18 colorectal cancer (CRC) patients and 26 cancer-free healthy individuals. The results, shown in Figure 11, show that based on PCA analysis of their mcfDNA signature, cancer free healthy individuals cluster separately from the CRC patients.

Sample ID	SPA identified	Gender	Ethnicity	Race	Cancer diagnosis	Histology	Spec Considered	Tumor site	Stage
106447	Yes	Female	Non-Spanish; Non-Hispanic	White	Colon Cancer	Adenocarcinoma	Metastatic	Ovary	IVC
116452	Yes	Female	Non-Spanish; Non-Hispanic	Black, African American	Colon Cancer	Adenocarcinoma	Primary	Rectal	IIA
125695	Yes	Male	Non-Spanish; Non-Hispanic	White	Colon Cancer	Adenocarcinoma	Primary	Colon	Not Documented
125788	Yes	Female	Non-Spanish; Non-Hispanic	White	Colon Cancer	Adenocarcinoma	Primary	Colon	IIA
125856	Yes	Female	Non-Spanish; Non-Hispanic	White	Colon Cancer	Adenocarcinoma	Primary	Colon	IIA
125900	Yes	Female	Non-Spanish; Non-Hispanic	Black, African American	Colon Cancer	Adenocarcinoma	Primary	Colon	I
125962	Yes	Female	Non-Spanish; Non-Hispanic	White	Colon Cancer	Adenocarcinoma	Primary	Colon	IIA
126101	No	Male	Non-Spanish; Non-Hispanic	Black, African American	Colon Cancer	Adenocarcinoma	Primary	Colon	IIA
104821	Yes	Female	Non-Spanish; Non-Hispanic	Black, African American	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	IA1
106456	Yes	Male	Non-Spanish; Non-Hispanic	White	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	IIA
106509	Yes	Female	Non-Spanish; Non-Hispanic	White	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	IIB

125584	Yes	Female	Non-Spanish; Non-Hispanic	Black, African American	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	IA3
116471	Yes	Male	Non-Spanish; Non-Hispanic	White	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	IIB
125725	Yes	Female	Non-Spanish; Non-Hispanic	White	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	IA2
125733	Yes	Female	Non-Spanish; Non-Hispanic	Multi-Racial	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	IB
125767	Yes	Female	Not Documented	Unknown	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	IA3
125796	Yes	Female	Non-Spanish; Non-Hispanic	White	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	IIB
125994	Yes	Male	Non-Spanish; Non-Hispanic	White	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	I
126276	Yes	Male	Non-Spanish; Non-Hispanic	White	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	Not Documented
126308	No	Female	Non-Spanish; Non-Hispanic	White	Non-Small Cell Lung Cancer	Adenocarcinoma	Primary	Lung	Not Documented

126292	Yes	Female	Non-Spanish; Non-Hispanic	White	Breast Cancer	Micropapillary	Primary	Breast	IIB
126291	No	Female	Non-Spanish; Non-Hispanic	White	Breast Cancer	DCIS	Primary	Breast	0
126282	Yes	Female	Non-Spanish; Non-Hispanic	White	Breast Cancer	Ductal	Primary	Breast	IIA
126274	Yes	Female	Non-Spanish; Non-Hispanic	White	Breast Cancer	Ductal	Primary	Breast	IB
126273	Yes	Female	Non-Spanish; Non-Hispanic	Black, African American	Breast Cancer	Ductal	Primary	Breast	IB
126162	Yes	Female	Non-Spanish; Non-Hispanic	White	Breast Cancer	Ductal	Primary	Breast	IIA
126154	Yes	Female	Spanish; Hispanic	Latino	Breast Cancer	Ductal	Primary	Breast	IIA
126120	Yes	Female	Non-Spanish; Non-Hispanic	White	Breast Cancer	Ductal	Primary	Breast	IB
126080	Yes	Female	Non-Spanish; Non-Hispanic	White	Breast Cancer	Ductal	Primary	Breast	IB
126033	Yes	Female	Non-Spanish; Non-Hispanic	White	Breast Cancer	Lobular	Primary	Breast	IIB
125570	Yes	Male	Non-Spanish; Non-Hispanic	White	Breast Cancer	Ductal	Primary	Breast	IIA
105609	Yes	Female	Non-Spanish; Non-Hispanic	Black, African American	Breast Cancer	Lobular	Primary	Breast	IIA

Table 3: Metadata for the 12 lung cancer patients, 12 breast cancer patients and 8 colorectal cancer patients used in this study. Information on the gender, ethnicity, race, cancer diagnosis, results of the tumor histology, tumor site and development stage are provided. SPA identified indicates if a patient was successfully stratified based on its mcfDNA pattern using SPA fragment sequencing of the 16SrRNA gene V4 hypervariable region followed by paired wise PCA analysis using the combinations (a) Colon Cancer versus Non-Small Cell Lung Cancer, and (b) Colon Cancer versus Breast Cancer, of which the results are shown in Figure 8 and Figure 9, respectively.

EXAMPLE 4

SPA fragment sequencing using the *rpoB* gene as phylogenetic markers for cancer patient stratification.

[00189] As a representative example for a single copy housekeeping gene, the SPA sequencing approach was successfully demonstrated for the *rpoB* gene.

[00190] To design a *rpoB* gene specific SPA primer, the process presented in Figure 4 was followed. A total of 50,569 unique *rpoB* gene sequences were downloaded from the PATRIC database (Wattam et al, 2014) using the version available in January 2021. *RpoB* gene sequences were identified based on their annotation as "DNA-directed RNA polymerase beta subunit (EC 2.7.7.6)". A subset of 50 *rpoB* gene sequences, representative for a broad range of phylogenetically distinct eubacterial reference microbes, were initially aligned by clustalW to identify conserved nucleotide regions of the *rpoB* gene, resulting in the identification of several conserved regions as primer candidates. This included the *rpoB* gene regions 1327-1352, 1528-1550, 1690-1709, 3766-3788 and 3808-3830, as well as the two regions identified by Ogier et al (2019), region 1630-1652 and region 2039-2063. The positions of the regions are based on the nucleotide sequence of the *Escherichia coli rpoB* gene.

[00191] Using the SPA algorithm, the 50,569 unique *rpoB* genes sequences were aligned to these conserved regions to determine their nucleotide compositions. The conserved nucleotide sequences of the *rpoB* gene region 1327-1355, which was the most conserved, is provided in Figure 12. In Table 4, the average sequence variances for the primer candidates are shown, with all primer candidates having a similar score, making them all primer leads. Subsequently, the estimate of adjacent region conservation was calculated. For each region, which represents a putative primer annealing site, the variance is shown for 25, 50, 75, 100 or 200 nucleotides (nt) upstream (5') or downstream (3') of the beginning or end of the sequence of the conserved region. The results are summarized in Table 4 and show that the nucleotide sequence upstream of the conserved region 1327-1352 is the most variable, as indicated by the lowest average variance scores of 0.0667 for both the 25 nucleotide-long and 50 nucleotide-long regions. The identification of a hypervariable DNA region in the *rpoB* gene upstream of the conserved region 1327-1352 was unexpected, as it falls outside of the region that has previously been identified and used for *RpoB* gene amplicon sequencing (Ogier et al, 2019).

[00192] To select primers with the least risk for nonspecific annealing to human genomic DNA, the number of putative annealing sites of the proposed degenerate primer sequences to the human genome sequence (Reference: GCF_000001405.40_GRCh38.p14_genomic.fna) with increased number of allowed mismatches

is determined. Results for the degenerate primers 16SV4-R785 and RpoB-R1327 are shown in Table 5. A primer should not have zero or one mismatch, and ideally no more than 10 instances of two mismatches with the human genome. Based on the results from this analysis, both the 16S-V4-R and RpoB-R1327 primers should be suitable for SPA fragment sequencing.

[00193] As the results of two degenerate nucleotides, (H: not G (C, A, T); and V: not T (C, A, G)), both the 16SV4-R785-Bio and 16SV4-R781 primers have a degeneracy of 9 possible combinations. When designing the RpoB-ePCR and RpoB-aPCR primers for the E-PCR and A-PCR steps we noticed that as a results of the degeneracy the region 1327-1355 had a total of 2,304 possible combinations (see consensus sequence in Figure 12). Using primers that included all degeneracies combined with an *E. coli* specific *rpoB* gene primer (RpoB-coli1247-F) failed to amplify a target fragment ranging from position 1220 to 1352 of the *rpoB* gene using *E. coli* DNA as positive control (results not shown).

[00194] To reduce the number of possible combinations in the primer design and inprove primer annealing, nucleotide linkage analysis was performed for this region to detect patterns of nucleotide preference. For instance, if the nucleotide at degenerate position 1329 of the *rpoB* gene is a cytosine, what is the frequency for a specific nucleotide at the next degenerate nucleotide position? This analysis was performed for the sequences of the regions covered by the degenerate primers RpoB1-R1327 and RpoB1-R1330 with either thymine or cytosine in position 1329 or position 1332, respectively. The results of this nucleotide linkage study, shown in Table 6, reveals that *rpoB* gene sequences with a cytosine in degenerate nucleotide position 1329 also has a strong preference for a cytosine in the nucleotide positions 1335 (87%), 1338 (85%), 1341 (82%) and 1351 (91%) of the *rpoB* gene. For *rpoB* gene sequences with a thymine in degenerate nucleotide position 1329, a preference for a thymine in the nucleotide positions 1332 (73%), 1338 (76%), 1341 (74%), 1351 (77%) and 1354 (78%) of the *rpoB* gene was observed. This information was subsequently used to design mixes of degenerate primers to be used in the E-PCR and A-PCR reactions of the SPA fragment sequencing protocol. The individual primer sequences for these mixes, referred to as the RpoB-ePCR mix and the RpoB-aPCR mix, are listed in Table 1. The degeneracy for the primers in these mixes ranged from 64 to 1536 for the RpoB-ePCR mix and 32 to 768 for the RpoB-aPCR mix. Using the RpoB-ePCR mix and the RpoB-aPCR mix in combination with the *E. coli* specific *rpoB* gene primer (RpoB-coli1247-F) on *E. coli* DNA allowed for successful amplification of the target fragment, confirming the importance of reduced primer degeneracy.

	Average of variance on RpoB gene										
	Region upstream of primer					Primer	Region downstream of primer				
Primer name – recognized RpoB gene region	**200 nt before primer**	**100 nt before primer**	**75 nt before primer**	**50 nt before primer**	**25 nt before primer**	**Primer region**	**25 nt after primer**	**50 nt after primer**	**75 nt after primer**	**100 nt after primer**	**200 nt after primer**
RpoB6 Forward - 1630-1652	0.1356	0.1603	0.1548	0.1538	0.1546	0.1810	0.1492	0.1671	0.1633	0.1468	0.1199
RpoB7 Reverse - 2039-2063	0.1035	0.1393	0.1476	0.1576	0.1773	0.1964	0.1142	0.1312	0.1223	0.1077	0.0840
RpoB1 - 1327-1352	0.0495	0.0571	0.0675	0.0667	0.0667	0.1846	0.1390	0.1309	0.1240	0.1159	0.1170
RpoB2 - 1528-1550	0.1123	0.1059	0.1139	0.1184	0.1266	0.1906	0.1368	0.1450	0.1491	0.1520	0.1477
RpoB3 - 1690-1709	0.1525	0.1592	0.1616	0.1643	0.1632	0.1974	0.1160	0.1197	0.1247	0.1095	0.0836

RpoB4 - 3766-3788	0.1277	0.1564	0.1665	0.1651	0.1418	0.1985	0.1846	0.1932	0.1808	0.1786	0.1584
RpoB5 - 3808-3830	0.1513	0.1775	0.1748	0.1862	0.1879	0.2094	0.1614	0.1631	0.1620	0.1538	0.1384

Table 4: Average sequence variance for the primer regions and the regions upstream or downstream of candidate primer annealing regions recognizing conserved *rpoB* gene sequences. For each region adjacent to the primer region, the variance is shown for 25, 50, 75, 100 or 200 nucleotides (nt) upstream (5') or downstream (3') of the beginning or end of the primer annealing sequence. The variance score is calculated as the average of the variance of the percentage of the nucleotides adenine, guanidine, cytosine and thymine at each position of the *rpoB* gene. A lower number is indicative for more variance, while a higher number is indicative for less variance and a more conserved DNA sequence. The maximum theoretical variance score for a region is 0.25 (would represent a 100% conserved DNA region). Regions with a variance score <0.1 are highlighted. The coordinates of the regions recognized by the primers are based on the nucleotide sequence of the *Escherichia coli rpoB* gene.

Primer	**Number of hits with zero mismatch (+ strand; - strand)**	**Number of hits with one mismatch (+ strand; - strand)**	**Number of hits with two mismatches (+ strand; - strand)**	**Number of hits with three mismatches (+ strand; - strand)**
16S-V4-R	0; 0	0; 0	8; 2	67; 47
RpoB1-R1327	0; 0	0; 0	1; 2	30; 26

Table 5: Number of hits for primers to the human genome. For each primer, the number of hits with zero, one or two mismatches are presented. The number of hits was determined based on homology to the nucleotide sequences on both DNA strands (+ and – strand) of the human chromosome (Reference: GCF_000001405.40_GRCh38.p14_genomic.fna).

Position	**1329 (Y)**		**1332 (Y)**		**1335 (Y)**		**1338 (Y)**		**1341 (Y)**		**1342 (Y)**		**1345 (N)**				**1348 (H)**			**1351 (Y)**		**1354 (H)**		
Nucleotide	C	T	C	T	C	T	C	T	C	T	C	T	A	C	G	T	A	C	T	C	T	A	C	T
C - 1329	**100**	**0**	50	50	**87**	13	**85**	15	**82**	18	62	38	12	48	30	9	5	67	27	**91**	9	4	39	57
T - 1329	**0**	**100**	27	73	71	29	76	24	74	26	60	40	29	24	33	15	19	41	40	77	23	9	13	78

Table 6: Linkage between nucleotide preferences for the region 1327 – 1355 of the *rpoB* gene. The nucleotide at position 1329 was either fixed as a C or a T, after which the preferred nucleotides for the other degenerate positions were calculated. This information was subsequently used to design primers with fewer degenerate nucleotides that recognize the region 1327 – 1356 of the *rpoB* gene. Preferred nucleotides with a preference score over 80% are highlighted in grey.

[00195] In EXAMPLE 4, protocol iteration 8 was followed, using a set of six phased primers referred to as the Linker-aPCR primer mix in combination with SPA-linker-UMI.v2-Y-SP for the A-PCR reaction. Using *rpoB* – specific primers to target the *rpoB* gene region 1327 – 1355, SPA fragment sequencing was performed on cfDNA isolated from the blood plasma of 10 breast and 10 lung cancer patients, as well as 4 healthy controls. Twenty blood plasma samples from cancer patients were obtained from the same hospital, while the healthy control samples were obtained locally from healthy volunteers.

[00196] Processing of the SPA fragment sequences was performed as described in EXAMPLE 1. Subsequently, the SciKitLearn (https://scikit-learn.org/stable/) package in python was used for diagnosis/classification of patients based on their *rpoB* gene derived mcfDNA signatures. Specifically, we compiled a feature matrix for each of our samples, where the columns represented "samples" and the rows represented "taxonomical classes at the bacterial genus level". All taxonomical genera were computed for each sample based on the best hit when blasting each *rpoB* gene-derived SPA fragment sequence in the sample against our database of reference genome sequences. For classification, bacterial genus features were binary, equal to "1" if the feature was present in the sample regardless of abundance and "0" otherwise.

[00197] This approach is schematically shown in Figure 13. As can be concluded from Figure 13, *rpoB* gene derived SPA fragments allowed, with 85 percent accuracy, to distinguish between healthy individuals and patients with lung and breast cancer, including patients with early-stage cancer. Of the 10 breast cancer patient samples, 8 samples were correctly classified; and of the 4 control samples, 3 were correctly stratified. Two breast cancer patient samples and 1 control sample ended up with the 10 lung cancer patient samples.

[00198] There are notable differences between SPA fragment sequencing of a single copy housekeeping gene, such as the *rpoB* gene, and SPA fragment sequencing of the 16S rRNA gene. In general, the *rpoB* gene derived SPA fragments provide better phylogenetic resolution than fragments derived from the 16S rRNA gene. However, the number of unique 16S rRNA gene derived SPA fragments obtained per sample is more than 300 times higher than what is observed for single copy housekeeping genes. It is therefore proposed to combine SPA fragment sequencing on the 16S rRNA gene and a single copy housekeeping gene, exemplified by the *rpoB* gene, for improved insights in community composition and phylogenetic identification, respectively. The use of two or more degenerate primers for annealing to two or more conserved regions on a single or two different phylogenetic marker genes may be referred to herein as "multi-loci SPA fragment sequencing".

EXAMPLE 5

SPA fragment sequencing to obtain clinical surrogate endpoint markers to assess remission in IBD

[00199] The therapeutic efficacy of biologics such as antiTNFα antibodies or JAK-kinase inhibiting antibodies, the most used anti-inflammatory drugs to treat severe forms of IBD and other inflammatory diseases like rheumatoid arthritis, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis and plaque psoriasis, has remained around 50% for the last twenty years. Furthermore, the selection of a biologic therapy for these diseases has been largely arbitrary. For IBD, stool and peripheral blood inflammatory biomarkers are not informative to monitor patient disease status, medication performance and treatment outcomes as they fail to address a growing body of evidence that increased intestinal permeability is correlated to disease progression, symptom severity, and patient outcomes. Currently, the lactulose - mannitol test is the most common and only direct test of a leaky gut. However, this test doesn't provide insights in imbalanced gut microbiome bacteria and a dysfunctional gut epithelial barrier, with leakage of microbial material from the gut into the bloodstream resulting in an overactive immune response that underlies multiple inflammatory conditions, including but not limited to IBD. Therefore, a need exists for biomarkers that provide this clinically relevant information.

[00200] SPA fragment sequencing was successfully used to evaluate the levels of mcfDNA in the blood of IBD patients. To further improve it as a diagnostic tool, the Blood Pathogenicity Index (BPI) was created which focuses on a root cause disease target, the presence of pro-inflammatory bacteria in the blood stream. As successfully demonstrated below, the BPI can be integral to clinical decision-making throughout the treatment continuum in IBD patients and patients suffering from other inflammatory conditions, as it provides a surrogate biomarker that can successfully measure the effect of a specific treatment and correlate with a real clinical endpoint: the load of mcfDNA derived from pathogens in blood as an indication for treatment performance and the risk for medical complications from infections.

[00201] The BPI is calculated as the percentage of mcfDNA fragments coming from pro-inflammatory bacteria as a fraction of the total amount of mcfDNA fragments found in a blood plasma sample, and comprises pathogenic members of the following bacterial genera/species: *Enterobacteriaceae* including *Enterobacter* species, *Escherichia* species (e.g. *E. coli*), *Shigella* species, *Klebsiella* species (e.g. *K. pneumoniae* and *K. oxytoca*), *Kluyvera* species, *Erwinia* species, *Haemophilus* species, *Salmonella* species (e.g. *S. typhi* and *S. enterica*), and *Listeria* species (*L. monocytogenes*); pathogenic *Clostridium* species such as C. *difficile, C. perfringens* and *C. botulinum*; *Veillonella* species such as *V. atypica, V. dispar* and

V. oral; *Pseudomonas* species like *P. aeruginosa*; *Stenotrophomonas* species like *S. maltophilia; Burkholderia* species like *B. cepacia*; pathogenic *Staphylococcus* species like *S. aureus; Porphyromonas* species; and *Fusobacterium* species like *F. nucleatum* and *F. varium*. For IBD, additional bacteria including *Rhuminococcus gnavus, Enterococcus faecium* and *Enterococcus faecalis* can be included in the BPI.

[00202] A total of 94 blood plasma samples (0.5 ml aliquots), coming from 18 Crohn's disease patients and 18 ulcerative colitis patients, were provided by the Crohn's and Colitis Foundation IBD Plexus (https://www.crohnscolitisfoundation.org/research/current-research-initiatives/ibd-plexus/about), together with the patient metadata. All patients were using immune response suppressing medication on at least one of the timepoints when blood samples were collected.

[00203] The samples were processed following the protocol described in EXAMPLE 1 using the MAXWELL HT ccfDNA Kit (Promega) for mcfDNA extraction in a 96-well format on the KingFisher (Thermo Fisher) platform. The SPA fragment libraries were sequenced on the iSEQ100 sequencer from Illumina. The analysis of the SPA fragment sequences included the following steps:

1. Paired-end reads are filtered based on read quality, expected amplicon structure, and length.
2. The UMI (unique molecular identifier) of each read is recorded for downstream deduplication and denoising.
3. Human-derived and singleton reads are removed.
4. Forward and reverse reads are merged if possible and clustered into bins of similar sequence.
5. The longest read in each bin is searched against a database of bacterial 16S rRNA genes to aid the assembly into consensus sequences.
6. Community composition is determined by consensus sequence classification, and relative abundances are calculated based on the percent of reads contributing to each consensus sequence, taking into consideration the number of deduplicated reads identified in step 2.

[00204] After SPA fragment sequencing, 63 samples passed the Quality Control criteria that are defined as "at least 1% of the sequencing reads for a sample should align to bacterial 16S rRNA with 90% identity over 90% of the trimmed read length". Having larger sample volumes would have overcome this issue. The BPI was calculated for these 63 samples and

compared to the patients' metadata including disease severity (stool frequency and urgency, disease index, endoscopy results) and historical medication data to evaluate how reliable the BPI was as a biomarker to correctly predict disease activity and clinical events. Clinical events were defined as an active disease flair or medication failure (i.e., medication stopped working or serious side effects of the medication were observed). The results are shown in Figures 14A and 14B for UC and CD, respectively. An overall summary of the BPI results is presented in Table 7. For comparison, a similar evaluation was performed using calprotectin levels as biomarker (see Table 8). Fecal calprotectin levels are currently used as a biomarker for disease activity in IBD patients. However, no US fecal calprotectin test has FDA claims for monitoring IBD disease progression or for identifying flares in IBD patients, though it is often used off label for that purpose. Although fecal calprotectin is a very sensitive marker for inflammation in the gastrointestinal tract, it is not a specific marker for IBD. Increased levels are also seen in gastrointestinal malignancies, infections, polyps and with the use of nonsteroidal anti-inflammatory drugs (Pathirana et al, 2018). Furthermore, patient stool collection and moderate acceptance among GI physicians are the issue for a broader acceptance of the test.

[00205] From Table 7 it can be concluded that the BPI was approximately 87% accurate in predicting disease activity status (55 out of 63 samples). This was considerably higher than ~67% accuracy of calprotectin levels (40 out of 60 samples, see Table 8) using the same criteria to define disease activity status. On this limited dataset the sensitivity of the BPI, calculated as [active/(active + false negative) x 100%], was 84%, while the BPI's specificity, calculated as [remission/(remission + false positive) x 100%], was 92%. For calprotectin as a biomarker the sensitivity calculated as [active/(active + false negative) x 100%] was 67%, while the specificity calculated as [remission/(remission + false positive) x 100%] was 67% on the same dataset, respectively.

[00206] Figures 14A and 14B show that UC and CD patients found to be in remission had very low BPI scores. Similarly, patients with clinical events had significantly elevated BPI scores at the time of these events. These results show that the BPI provides an excellent biomarker to monitor disease activity status.

[00207] Fecal calprotectin levels above 150 μg/g are considered indicative for inflammation of the intestinal epithelial lining. The performance of calprotectin as a biomarker to monitor disease activity status on the same dataset, as presented in Figures 15A and 15B for UC and CD patients, respectively, shows that calprotectin is inferior as a biomarker to assess the activity of CD than that of UC, in line with previous findings (for review, see Sakurai and Saruta, 2023). It can be concluded that the blood-based pathogenicity index is superior to stool-

based fecal calprotectin for predicting an IBD patient's disease activity status, especially for patients suffering from Crohn's disease.

IBD diagnosis	Clinical Event	Remission	False positive	False negative
CD (32)	12	16	2	2
UC (31)	19	8	0	4

Table 7: Prediction of an IBD patient's disease status based on the Blood Pathogenicity Index. The BPIs of 63 IBD patient samples that passed the QC criteria after SPA fragment sequencing, including 32 CD and 31 UC patient samples, were determined, and patient metadata were used to confirm the disease status for each sample. A BPI >1.0 is considered as active disease. A clinical event refers to an active disease flair or medication failure (i.e., medication stopped working or serious side effects of the medication were observed). False positive is defined as an elevated BPI with no recorded clinical event. False negative is defined as a BPI <1.0 while a clinical event has been recorded.

IBD diagnosis	Clinical Event	Remission	False positive	False negative
CD (32)	8	9	7	5
UC (31)	16	7	1	7

Table 8: Prediction of an IBD patient's disease status based on fecal calprotectin levels. The calprotectin levels were determined on stool samples that were taken around the same time as the corresponding 63 blood samples that had passed the QC criteria after SPA fragment sequencing. The samples included 29 CD and 31 UC patient samples where calprotectin results were available, and patient metadata were used to confirm the disease status for each sample. Calprotectin levels >150 μg/g fecal material are considered as active disease. Calprotectin levels were determined on 63 IBD samples that passed QC for SPA fragment sequencing. A clinical event refers to an active disease flair or medication failure (i.e., medication stopped working or serious side effects of the medication were observed). False positive is defined as an elevated BPI with no recorded clinical event. False negative is defined as Calprotectin levels <150 μg/g fecal while a clinical event has been recorded.

[00208] Linking the BPI back to the patient journey showed that for many CD and UC patients an increase in BPI was observed when their medication stopped working, resulting in active disease, or when a patient on the medication reported undesirable side effects and/or disease symptoms, showing that the BPI can be successfully used as a surrogate biomarker to monitor disease severity, treatment efficacy and risk of complications in IBD patients. Importantly, the BPI would have also correctly predicted remission in CD and UC patients. Examples of the correlation between the BPI and the patient's health status are presented in Table 9 and Table 10 for CD or UC patients, respectively.

[00209] Several scenarios from the study described above, that show the usefulness of the BPI to monitor the status and progression of Crohn's disease, are provided in Table 9.

- Patients CD-1 and CD-2, both having a BPI of 0, were in full remission with their current medication.
- Patient CD-3 was treated with a succession of different antiTNFα medications that were likely losing efficacy which coincided with an increase in the BPI.

[00210] Patient CD-4 reported side effects while taking the antiTNFα medication Adalimumab (Humira), which coincided with an increase in the BPI. After the medication was eliminated, the BPI returned to 0. The side effects from adalimumab were missed by the calprotectin levels and the SCDAI score. The journey of patient CD-4 is schematically presented in Figure 16. Specifically, Figure 16 shows a schematic representation of the journey of Crohn's disease patient CD-4 and how the BPI can be successfully used to influence a better disease outcome. The patient is in remission (12/2018; 06/2019) on a combination of azathioprine and adalimumab, the latter being discontinued due to side effects (10/2019). Stopping this medication results in full remission (01/2020). In retrospect (10/2019), clinical use of the BPI would have identified inflammatory markers in blood that alter the CD-4 patient pathway by changing medication. The side effects from adalimumab were missed by the calprotectin (Calpr) levels and the SCDAI score (SCADI). Data points marked by open circles or grey circles indicate remission or active disease, respectively. Arrows indicate medication is being taken over a certain time. A STOP symbol indicates that a medication is being discontinued at a given timepoint.

- Patient CD-5 was diagnosed with CD and put on an antiTNF α medication. This medication was not showing the desired treatment effect, which was confirmed by an elevated BPI.
- Patient CD-6 was initially treated with corticosteroids. With disease progression, the patient was successfully treated with an antiTNFα medication, resulting in remission, which was supported by a low BPI value <1.0.

- Patient CD-7 was initially treated with Vedolizumab, an α4β7 integrin blocker. Once this medication stopped working and was omitted, the BPI went up, indicating the need for changes to the patient's treatment regime.
- Patient CD-8 initially received Azothioprine, a purine metabolism antagonist, as part of its medication. After this medication was omitted, the BPI went up, indicating the need for changes to the patient's treatment regime.
- Patient CD-9 was initially treated with Tacrolimus, a calcineurin phosphatase inhibitor. Once this medication stopped working, and was omitted, the BPI went up, indicating the need for changes to the patient's treatment regime.

Patient	Sample date (year-month)	Calprotectin (µg/g)	SCDAI score	BPI (%)	Medication before or at time of sample	Medication type before at time of sample	Medication usage history and comments
CD-1	2019-03	42.26	72	0.00	Mesalamine, Methylprednisolone	inflammatory response modulators	Mesalamine 2001-01-03 to 2022-07-3 Methylprednisolone 4 MG Oral Table Pack 2015-05-26 to 2019-03-13
CD-1	2020-07	35.74	44	0.00	Mesalamine	inflammatory response modulators	
CD-2	2018-04	326.59	121	0.00	Certolizumab	antiTNFα	Certolizumab Pegol 2017-01-14 to 2018-08-03 OTHER; PRIMARY NON-RESPONSE, Steroids Oral 2018-03-08 to 2018-09-08 Other, PREDNISONE 5 MG PO TABS 2018-09-24 to 2018-11-01
CD-2	2018-11	147.87	N.A.	0.00	Prednisone	Corticosteroid	
CD-2	2019-03	2.87	N.A.	0.00	None	None	
CD-3	2018-02	N.A.	184	0.00	Adalimumab	antiTNFα	Adalimumab (Humira) 2012-02-02 to 2019-12-22 OTHER, Infliximab (Remicade) 2019-07-18 to 2020-11-12 OTHER, Infliximab (Renflexis) 2019-09-12 to 2020-11-12
CD-3	2019-05	156.28	142	**3.33**	**Adalimumab***	antiTNFα	
CD-3	2019-09	175.05	184	**5.86**	**Adalimumab*, Infliximab**	antiTNFα (Multiple)	
CD-4	2018-12	0.00	44	0.00	Azathioprine	purine metabolism antagonist	Adalimumab (Humira) 2019-07-06 to 2019-10-10 SIDE EFFECTS, Azathioprine 1990-03-09 to 2023-02-16
CD-4	2019-06	108.98	128	0.00	Azathioprine	purine metabolism antagonist	
CD-4	2019-10	0.00	128	**6.25**	**Adalimumab*, Azathioprine**	antiTNFα + purine metabolism antagonist	
CD-4	2020-01	5.00	44	0.00	Azathioprine	purine metabolism antagonist	
CD-5	2019-08	1888.50	303	**9.09**	**Steroids Oral***	steroid	Infliximab (Inflectra) 2020-03-04 to 2020-09-15, Infliximab (Remicade) 2019-09-17 to 2020-09-12 OTHER, Ustekinumab 2020-09-05 to 2021-05-19 OTHER, Steroids Oral 2019-07-22 to 2022-03-14
CD-5	2020-03	1888.50	205	**18.49**	**Infliximab*,** Steroids Oral	antiTNFα, Corticosteroid	
CD-6	2018-07	22.47	128	**11.33**	**Budesonide***	Corticosteroid	Budesonide (Entocort) 2004-07-21 to 2018-07-21 OTHER, Adalimumab
CD-6	2019-11	0.00	58	0.00	Adalimumab	antiTNFα	

							(Humira) 2018-10-01 to 2019-05-25 OTHER
CD-7	2018-10	780.57	254	0.00	Azathioprine, Mesalamine, **Vedolizumab*, Steroids Oral**	purine metabolism antagonist + inflammatory response modulators + α4β7 integrin blocker, steroid	Azathioprine 2017-12-20 to 2023-03-01, Mesalamine 2017-01-27 to 2019-06-16 COMPLETED COURSE, Vedolizumab 2014-11-17 to 2019-04-17 STOPPED WORKING, Steroids Oral 1990-07-03 to 2019-07-03 Tapers
CD-7	2019-10	0.00	N.A.	**6.90**	**Azathioprine**	purine metabolism antagonist	
CD-7	2020-01	0.00	N.A.	**14.79**	**Azathioprine**	purine metabolism antagonist	
CD-8	2017-09	1888.50	114	0.00	**Azathioprine***, Infliximab, Methotrexate, Ustekinumab, Prednisone	purine metabolism antagonist + antiIL12/IL23 + antiTNFα, Corticosteroid	Azathioprine 2016-03-10 to 2017-09-2 STOPPED WORKING, Infliximab (Remicade) 2016-04-06 to 2022-01-14, Methotrexate 2017-04-05 to 2022-01-14, Ustekinumab 2017-04-20 to 2022-01-14, prednisone 20 MG Oral Tablet 2016-04-29 to 2022-01-14
CD-8	2018-10	N.A.	N.A.	**8.70**	Infliximab, Methotrexate, Ustekinumab, Prednisone	antiIL12/IL23 + antiTNFα, Corticosteroid	
CD-9	2019-03	57.26	233	0.00	Tacrolimus	calcineurin phosphatase inhibitor	
CD-9	2020-10	104.97	135	**25.00**	**None***	None	Tacrolimus 2017-09-17 to 2019-02-10

Table 9: Examples of the correlation between the Blood Pathogenicity Index (BPI), disease progression and the performance of medication in Crohn's disease (CD) patients. The following information is provided: CD patient number (CD-n); Sample Date; Calprotectin levels (µg/mg stool); Patient's SCDAI score; BPI, indicating the percentage (%) of pathogens present in the mcfDNA signature as determined via SPA fragment sequencing; Medication name and type before or at time of sample; Medication usage history and comments. The BPI was calculated as the percentage of mcfDNA fragments coming from pro-inflammatory bacteria as part of the total amount of mcfDNA fragments found in a blood plasma sample. An increase of BPI is indicative that a medication is no longer working, and/or that a patient on the medication is experiencing an increased level of infection by pathogenic bacteria. BPI values >1.0 are in bold; * indicated that a medication (in bold) is removed or no longer working. The results were obtained in a retrospective study.

Patient	Sample date (year-month)	Calprotectin (µg/g)	Mayo6 score	BPI (%)	Medication before or at time of sample	Medication type before at time of sample	Medication usage history and comments
UC-1	2018-04	926.02	Mild	**5.88**	Mesalamine	inflammatory response modulators	
UC-1	2019-01	352.21	Remission	**11.11**	Mesalamine	inflammatory response modulators	
UC-1	2019-03	621.41	Mild	**13.33**	Mesalamine	inflammatory response modulators	
UC-1	2020-01	24.04	Remission	**29.03**	Mesalamine	inflammatory response modulators	Mesalamine 2014-11-04 to 2023-02-09
UC-2	2018-03	1246.45	Remission	**26.92**	Balsalazide, Mesalamine	5-aminosalicylic acid + inflammatory response modulators	
UC-2	2018-08	364.68	Remission	**52.38**	Balsalazide, Mesalamine	5-aminosalicylic acid + inflammatory response modulators	Balsalazide 2017-12-03 to 2021-05-27, Mesalamine 2014-06-11 to 2019-11-07
UC-3	2018-11	6.29	Remission	**9.52**	**Mesalamine***, Tofacitinib, Vedolizumab, Prednisone, Budesonide	inflammatory response modulator + α4β7 integrin blocker + JAK kinase inhibitor, Corticosteroid	Mesalamine 2016-06-08 to 2019-02-13 SIDE EFFECTS, PARTIAL RESPONSE; Tofacitinib 2018-09-06 to 2023-02-28, Vedolizumab 2018-04-01 to 2020-03-25, PREDNISONE 20 MG PO TABS 2018-05-24 to 2018-11-19, PREDNISONE 5 MG PO TABS 2018-06-08 to 2018-11-19, Budesonide (Uceris MMX) 2018-06-08 to 2019-02-13 Completed course
UC-3	2019-03	305.68	Remission	0.00	Tofacitinib, Vedolizumab, Budesonide	α4β7 integrin blocker + JAK kinase inhibitor, Corticosteroid	
UC-3	2019-09	44.27	Remission	0.00	Tofacitinib, Vedolizumab	α4β7 integrin blocker + JAK kinase inhibitor	
UC-4	2018-10	135.76	Remission	0.00	Vedolizumab	integrin $\alpha_4\beta_7$ blocker	Vedolizumab 2016-08-17 to 2020-09-06 COMPLETE RESPONSE
UC-4	2019-06	60.03	Remission	**15.00**	Vedolizumab	integrin $\alpha_4\beta_7$ blocker	

UC-5	2018-05	59.86	Remission	0.00	Cyclosporine, Mesalamine, **Vedolizumab***	inflammatory response modulators + integrin $\alpha_4\beta_7$ blocker	Cyclosporine 2014-08-16 to 2022-01-10, Mesalamine 2007-03-01 to 2022-03-07, Vedolizumab 2015-09-25 to 2018-06-29 COMPLETE RESPONSE, then dropped as medication
UC-5	2018-11	N.A.	Mild	**30.43**	Cyclosporine, Mesalamine	inflammatory response modulators	
UC-5	2020-02	85.40	Remission	**7.14**	Cyclosporine, Mesalamine	inflammatory response modulators	
UC-6	2019-06	739.72	Mild	**3.85**	**Adalimumab*, Infliximab*,** Prednisone	antiTNFα, Corticosteroid	Adalimumab (Humira) 2018-01-02 to 2019-07-13 PRIMARY NON-RESPONSE Infliximab (Remicade) 2018-03-05 to 2019-07-13, prednisone 20 MG Oral Tablet 2018-11-12 to 2019-07-13
UC-6	2020-10	1727.90	Remission	0.00	None	None	
UC-7	2019-02	0.00	Remission	0.00	Mesalamine, Vedolizumab	inflammatory response modulators + integrin $\alpha_4\beta_7$ blocker	Mesalamine 2014-01-06 to 2021-04-12, Tofacitinib 2019-06-05 to 2022-07-02, Vedolizumab 2016-12-25 to 2022-04-23
UC-7	2019-03	1.28	Remission	**10.00**	**Mesalamine*, Vedolizumab***	inflammatory response modulators + integrin $\alpha_4\beta_7$ blocker	
UC-7	2019-08	6.48	Mild	**4.00**	**Mesalamine*, Vedolizumab*, Tofacitinib**	integrin $\alpha_4\beta_7$ blocker + JAK kinase inhibitor	
UC-8	2019-06	620.77	Remission	0.00	Mesalamine	inflammatory response modulators	Mesalamine 2014-02-08 to 2019-10-22 LOSS OF RESPONSE; Vedolizumab 2019-07-09 to 2021-03-04
UC-8	2019-07	310.86	Mild	**12.50**	**Mesalamine***	inflammatory response modulators	
UC-8	2020-02	342.29	Moderate	0.00	Vedolizumab	binds α4β7 integrin	
UC-9	2017-11	609.14	Severe	**2.33**	**Mesalamine***	inflammatory response modulators	Adalimumab (Humira) 2017-12-23 to 2022-03-17, Mesalamine 2012-12-28 to 2022-10-31
UC-9	2019-11	182.64	Remission	0.00	Mesalamine, Adalimumab	inflammatory response modulators + antiTNFα	
UC-10	2018-06	827.55	Mild	0.00	Vedolizumab	integrin $\alpha_4\beta_7$ blocker	Vedolizumab 2015-12-24 to 2018-08-16 PRIMARY NON-RESPONSE,
UC-10	2019-05	227.60	N.A.	0.00	Adalimumab	antiTNFα	

UC-10	2019-12	1888.50	Severe	**27.78**	**Adalimumab*,**Prednisone	antiTNFα, Corticosteroid	Adalimumab (Humira) 2018-08-16 tc 2020-02-18, prednisone 10 MG Oral Tablet 2019-12-04 to 2020-01-25

Table 10: Examples of the correlation between the Blood Pathogenicity Index (BPI), disease progression and the performance of medication in Ulcerative Colitis (UC) patients. The following information is provided: UC patient number (UC-n); Sample Date; Calprotectin levels (μg/mg stool); Patient's Mayo-6 score; BPI, indicating the percentage (%) of pathogens present in the mcfDNA signature as determined via SPA fragment sequencing; Medication name and type before or at time of sample; Medication usage history and comments. The BPI was calculated as the percentage of mcfDNA fragments coming from pro-inflammatory bacteria as part of the total amount of mcfDNA fragments found in a blood plasma sample. An increase of BPI is indicative that a medication is no longer working, and/or that a patient on the medication is experiencing an increased level of infection by pathogenic bacteria. BPI values >1.0 are in bold; * indicated that a medication (in bold) is removed or no longer working. The results were obtained in a retrospective study.

[00211] Several scenarios from the study that show the usefulness of the BPI to monitor the status and progression of ulcerative colitis are provided in Table 10.

- Patients UC-1 and UC-2 provide examples where patients indicate remission, but where the levels of calprotectin and the BPI are elevated, indicating that the medication is not meeting its primary endpoint of lowering inflammation of the gut epithelium.
- Patient UC-3 was on a combination of medication that included Mesalamine, which started to have side effects and stopped working. This coincided with an increase in the BPI. After the medication was eliminated, the BPI returned to 0. The journey of patient UC-3 is schematically presented in Figure 17. Specifically, Figure 17 shows a schematic representation of the journey of ulcerative colitis patient UC-3 and how the BPI can be successfully used to influence a better disease outcome. It is assumed that the initial combination of tofacitinib, vedolizumab and mesalamine is working before 11/2018. Issues are reported for mesalamine, which is removed from the medication (02/2019). In retrospect (11/2018), clinical use of BPI would have identified inflammatory markers in blood that alter the UC-3 patient pathway by removing mesalamine as medication, resulting in a decrease of the BPI (03/2019) indicative for remission in line with fecal urgency. The side effects from mesalamine were missed by the calprotectin levels (Calpr) and the Mayo-6 score (Mayo-6). Data points marked by open circles or grey dots indicate remission or active disease, respectively. A data point marked by a black dot indicates that no sample to determine the BPI was available. Arrows indicate medication is being taken over a certain time. A STOP symbol indicates that a medication is being discontinued at a given timepoint.
- Patient UC-4 is an example of a false positive. An elevated BPI is observed despite no issues with medication being reported.
- Patient UC-5 was on a combination of medication that included Vedolizumab. Once full remission was obtained, the patient's sample had a BPI of 0. After this medication was omitted, the BPI went up, indicating the need for changes to the patient's treatment regime.
- Patient UC-6 was on a combination of two antiTNFα medications that stopped working, which coincided with an increase in the BPI. After the medications were eliminated, the BPI returned to 0 and the patient reported remission, despite elevated calprotectin levels.
- Patient UC-7 was initially successfully treated with a combination of mesalamine and vedolizumab. Subsequently, resistance was detected which coincided with an increase in BPI. Tofacitinib, a JAK kinase inhibitor, was included in the patient's medication regimen. Inclusion of the JAK kinase inhibitor correlates with a decrease of the BPI.

- Patient UC-8 is initially successfully treated with Mesalamine, an inflammatory response modulator. This medication stopped working, which required an adjustment of the patient's medication. The ceasing of the medication working correlates with an increase in the BPI.
- Patient UC-9 is an example where the initial medication, Mesalamine, did not work. The BPI determined on a blood sample taken from the patient at this time is elevated. Subsequent addition of an antiTNFα medication resulted in remission, which correlates with a decrease in the BPI. The journey of patient UC-9 is schematically presented in Figure 18. Specifically, Figure 18 shows a schematic representation of the journey of ulcerative colitis patient UC-9 and how the BPI can be successfully used to influence a better disease outcome. Initially, mesalamine is working as medication but starts to lose efficiency (11/2017) based on disease symptoms and calprotectin levels. Adalimumab is added as medication (12/2017), resulting in remission (11/2019). In retrospect (11/2017), clinical use of the BPI would have identified inflammatory markers in blood that alter the UC-9 patient pathway by changing medication to include adalimumab, resulting in a decrease of the BPI (11/2019). Data points marked by open circles or grey dots indicate remission or active disease, respectively. A data point marked by a black dot indicates that no sample to determine the BPI was available. Arrows indicate medication is being taken over a certain time.
- Patient UC-10 is initially treated with an integrin $\alpha_4\beta_7$ blocker, which stops working (BPI not available). Subsequently, the patient is treated with an antiTNFα. This medication initially results in remission, which correlates with a BPI of <0.0. However, the antiTNFα medication stopped working, requiring adjustment of the patient's medication. The BPI is increased at the time of the antiTNFα medication not working.

[00212] Overall, the results show that SPA fragment sequencing of mcfDNA in blood can be successfully used to determine the BPI of patients suffering from inflammatory diseases such as IBD. The BPI was successfully used as a surrogate biomarker for both medication performance and assessing the risk for serious or life-threatening complications, such as an increased risk of inflammation.

[00213] The increased risk for serious or life-threatening infections has been well documented for immune suppressing medications and there is a need for biomarkers that allow for early detection and prevention of this complication. The results of the retrospective study, described in Tables 7, 9 and 10, show that failure or complications due to immune suppressing

medication is accompanied by an increase in the mcfDNA signature of (opportunistic) pathogenic bacteria in the blood of the patients, as determined by the BPI. Therefore, in addition to IBD, the BPI can be used as a surrogate biomarker for diseases where patients are at similar risk of serious or life-threatening infections due to treatment with immune repressing drugs, including rheumatoid arthritis, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis, plaque psoriasis, acute Graft versus Host disease. Immune suppressing mediations that can benefit from patient monitoring using the BPI to determine the risk of serious or life-threatening infections include, but are not limited to, abatacept (Orencia); adalimumab (Humira); infliximab (Remicade); Tofacitinib (Xeljanz); risankizumab (Skyrizi); anakinra (Kineret); azathioprine (Azasan); certolizumab (Cimzia); cyclosporine (Gengraf, Neoral, Sandimmune); etanercept (Enbrel); golimumab (Simponi); infliximab (Remicade); methotrexate (Otrexup, Rasuvo, Trexall); rituximab (Rituxan); steroids including dexamethasone, methylprednisolone (Medrol), prednisolone (Orapred ODT, Prelone), and prednisone (Rayos); tacrolimus (Astagraf, Envarsus XR, Prograf); and tocilizumab (Actemra).

[00214] The BPI can also be used to assess the side effects for other medications/treatments that can cause complications of inflammation, including the gastrointestinal track, or affect the immune response, such as immune checkpoint inhibitor drugs like Pembrolizumab, Nivolumab, and Cemiplimab as anti-PD-1 antibodies, Ipilimumab as an anti-CTLA-4 antibody, as well as Atezolizumab, Avelumab, and Durvalumab as anti-PD-L1 antibodies.

REFERENCES

Abadio A. K. R., et al. (2011). Comparative genomics allowed the identification of drug targets against human fungal pathogens. BMC Genomics 12:75.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990), Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Ananthakrishnan A. N., et al. (2017). Gut Microbiome Function Predicts Response to Antiintegrin Biologic Therapy in Inflammatory Bowel Diseases. Cell Host & Microbe 21: 603–610.

Bhattacharjee S., Lukiw W. J. (2013). Alzheimer's disease and the microbiome. Frontiers in Cellular Neuroscience 7: Article 153.

Boulangé C. L., et al. (2016). Impact of the gut microbiota on inflammation, obesity, and metabolic disease. Genome Medicine 8:42

Bullman S., et al. (2017). Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. Science 358: 1443–1448.

Burnham P., Kim M. S., Agbor-Enoh S., Luikart H., Valantine H. A., Khush K. K., et al. (2016). Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. Sci Rep-Uk. 6.

Callahan, B. J., et al. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods 13: 581-583.

Castellarin M., et al. (2012). *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. Genome Res. 22: 299–306.

Cohen R. J., Shannon B. A., McNeal J. E., Shannon T., Garrett K. L. (2005). *Propionibacterium acnes* associated with inflammation in radical prostatectomy specimens: A possible link to cancer evolution? J. Urol. 173: 1969–1974.

Decker B., Sholl L. M. (2020). Cell-Free DNA Testing. In: Genomic Medicine. Springer: 41-54.

Fernández-Carballo B. L., Broger T., Wyss R., Banaei N., Denkinger C. M. (2019). Toward the development of a circulating free DNA-based in vitro diagnostic test for infectious diseases: a review of evidence for tuberculosis. J Clin Microbiol. 57: e01234-18.

Ferreira P. C. L., et al. (2023). Plasma biomarkers identify older adults at risk of Alzheimer's disease and related dementias in a real-world population-based cohort. Alzheimer's & Dementia. DOI:10.1002/alz.12986

Geller L. T., et al. (2017). Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. Science 357: 1156–1160.

Ghebremedhin B., Layer F., Konig W., Konig B. (2008). Genetic classification and distinguishing of Staphylococcus species based on different partial gap, 16S rRNA, hsp60, rpoB, sodA, and tuf gene sequences. J Clin Microbiol. 46: 1019–1025.

Gopalakrishnan V., Weiner B., Ford C. B., Sellman B. R., Hammond S. A., Freeman D. J., Dennis P., Soria J-C., Wortman J. R., Henn M. R. (2020). Intervention strategies for microbial therapeutics in cancer immunotherapy. Immuno-Oncology Technology 6: 9-17.

Gorelick J. I., Senterfit L. B., Vaughan E. D. Jr. (1988). Quantitative bacterial tissue cultures from 209 prostatectomy specimens: Findings and implications. J. Urol. 139: 57–60.

Gupta S., et al. (2019). Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. Communications Biology 2: 291.

Han D., Li R., Shi J., Tan P., Zhang R., Li J. (2020). Liquid biopsy for infectious diseases: a focus on microbial cell-free DNA sequencing. Theranostics 10: 5501-5513.

Head S. R., et al. (2014). Library construction for next-generation sequencing: overviews and challenges. BioTechniques 56, 61–64, 66, 68, passim.

Jahr S., Hentze H., Englisch S., Hardt D., Fackelmayer F. O., Hesch R. D., et al. (2001). DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res. 61: 1659-1665.

Jin C., et al. (2019). Commensal microbiota promote lung cancer development via γδT cells. Cell 176: 998–1013.

Langmead B, Salzberg S. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods. 9: 357-359.

Liang W., Zhao Y., Huang W., Gao Y., Xu W., Tao J., et al. (2019). Non-invasive diagnosis of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA). Theranostics 9: 2056-2070.

Links M. G., Dumonceaux T. J., Hemmingsen S. M., Hill J. E. (2012). The chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. PLoS One 7: e49755.

Mui U. N., Haley C. T., Tyring S. K. (2017). Viral Oncology: Molecular Biology and Pathogenesis. J. Clin. Med. 6 (111). doi:10.3390/jcm6120111.

Narunsky-Haziza L., et al. (2022). Pan-cancer analyses reveal cancer-type-specific fungal ecologies and bacteriome interactions. Cell 185: 3789–3806.

Nejman D., et al. (2020). The human tumor microbiome is composed of tumor type-specific intracellular bacteria. Science 368: 973–980.

Ogier J.-C., Pages S., Galan M., Barret M., and Gaudriault S. (2019). *rpoB*, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. BMC Microbiology 19: 171.

Parks D. H., Imelfort M., Skennerton C. T., Hugenholtz P., Tyson G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research 25: 1043–1055.

Pathirana, W. G. W., et al. (2018). Faecal Calprotectin. Clin Biochem Rev. 39(3): 77-90.

Poirier S., et al. (2018). Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using gyrB amplicon sequencing. A comparative analysis with 16S rDNA V3-V4 amplicon sequencing. PLoS One. 13: e0204629.

Poore G. D., et al. (2020). Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature 579: 567–574.

Rassoulian Barrett S. L., Holmes E. A., Long D. R., Shean R. C., Bautista G. E., Ravishankar S., Peddu V., Cookson B. T., Singh P. K., Greninger A. L., Salipante S. J. (2020). Cell free DNA from respiratory pathogens is detectable in the blood plasma of Cystic Fibrosis patients. Scientific Reports 10, Article number: 6903.

Riquelme E., et al. (2019). Tumor microbiome diversity and composition influence pancreatic cancer outcomes. Cell 178: 795–806.

Sakurai, T., Saruta, M. (2023). Positioning and usefulness of biomarkers in inflammatory bowel disease. Digestion 104: 30–41.

Sepich-Poore G. D., Zitvogel L., Straussman R., Hasty J., Wargo J. A., Knight R. (2021). The microbiome and human cancer. Science 371: 1331-1343.

Shahir N. M., et al. (2020). Crohn's disease differentially affects region-specific composition and aerotolerance profiles of mucosally adherent bacteria. Inflammatory Bowel Disease 26: 1843-1855.

Sun Y., An K., Yang C. (2019). Circulating cell-free DNA. In: Liquid Biopsy. Intech Open.

Urbaniak C., et al. (2016). The microbiota of breast tissue and its association with breast cancer. Appl. Environ. Microbiol. 82: 5039–5048

Venkataramani V., et al. (2019). Glutamatergic synaptic input to glioma cells drives brain tumour progression. Nature 573: 532–538.

Vos M., Quince C., Pijl A. S., de Hollander M., Kowalchuk G. A. (2012). A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. PLoS One 7 :e30600.

Wattam A. R., et al. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 42, D581-591, doi:10.1093/nar/gkt1099.

Witt R. G., Blair L., Frascoli M., Rosen M. J., Nguyen Q.-H., Bercovici S., et al. (2020). Detection of microbial cell-free DNA in maternal and umbilical cord plasma in patients with chorioamnionitis using next generation sequencing. PLoS ONE 15 (4): e0231239.

Wu J., Hu S., Zhang L., Xin J., Sun C., Wang L., et al. (2020). Tumor circulome in the liquid biopsies for cancer diagnosis and prognosis. Theranostics 10: 4544-4556.

Zeng Q., et al. (2019). Synaptic proximity enables NMDAR signaling to promote brain metastasis. Nature 573: 526–531.

WHAT IS CLAIMED IS:

1. A method of amplifying microbial cell free DNA (mcfDNA), comprising:

(a) providing a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprise a hypervariable region, and wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region;

(b) providing an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA;

(c) performing, on a sample comprising the mcfDNA, an enrichment polymerase chain reaction (E-PCR) using the degenerate primer and the additional primer, thereby generating one or more enriched mcfDNA fragments; and

(d) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer, wherein the modified degenerate primer comprises a 3'-end extension of at least three nucleotides relative to the degenerate primer; and (ii) the additional primer, thereby generating enriched amplified mcfDNA fragments.

2. A method of amplifying microbial cell free DNA (mcfDNA), comprising:

performing an enrichment polymerase chain reaction (E-PCR) on a sample comprising microbial cell-free DNA (mcfDNA), wherein enriched mcfDNA fragments are generated using:

i) a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and wherein the degenerate primer comprises an affinity purification group, and
ii) an additional primer complementary to an end of the mcfDNA,

wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprise a hypervariable region, wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region; and

performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using: (i) a modification of the degenerate primer of the E-PCR, wherein the modified degenerate primer comprises a 3'-end extension of at least three nucleotides relative to the degenerate primer of the E-PCR, and (ii) the additional primer of the E-PCR, thereby generating enriched amplified mcfDNA fragments.

3. The method of claim 1 or 2, further comprising providing one or more additional degenerate primers comprising complementarity to one or more additional conserved regions of the microbial phylogenetic marker gene or to one or more additional genes.

4. The method of claim 3, wherein the one or more additional genes comprises one or more additional microbial phylogenetic marker genes.

5. The method of claim 3, wherein the one or more additional genes encodes virulence factors, secondary metabolites, or drug degradation pathways, and combinations thereof.

6. The method of claim 1 or 2, further comprising sequencing the enriched amplified mcfDNA fragments.

7. The method of claim 6, further comprising, using a computer system:

    a. aligning the amplified mcfDNA fragment sequences on a sequence of the degenerate primer and assigning matching sequences from the hypervariable region as representative of the same microbial species;

    b. for the microbial species in part (a), searching a database of the phylogenetic marker gene against the amplified mcfDNA fragment sequences and assigning a microbial species based on the closest match; and

    c. optionally, calculating a microbial community composition based on a relative abundance of the amplified mcfDNA fragment sequences assigned to each microbial species.

8. The method of claim 7, wherein the phylogenetic marker gene is a multicopy gene, and further comprising correcting for copy number variation between each microbial species.

9. The method of claim 7, further comprising providing a second degenerate primer complementary to a second conserved region of a second phylogenetic marker gene, and further comprising determining the microbial community composition by calculating a mathematical mean of the relative abundance of each microbial species for each of the phylogenetic marker gene and the second phylogenetic marker gene.

10. The method of claim 7, wherein the microbial community composition comprises one or more members of Eukaryotes, bacteria, viruses, or fungi.

11. The method of claim 7, wherein one or a combination of: the microbial community composition or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species associated with a disease or disorder, indicates that the subject has the disease or disorder.

12. A method of diagnosing or monitoring a disease or disorder, comprising:

(a) providing a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region;

(b) providing an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA;

(c) performing, on a sample comprising the mcfDNA, an enrichment polymerase chain reaction (E-PCR) using the degenerate primer and the additional primer, thereby generating one or more enriched mcfDNA fragments;

(d) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer, wherein the modified degenerate primer comprises a 3'-end extension of at least three nucleotides relative to the degenerate primer; and (ii) the additional primer, thereby generating enriched amplified mcfDNA fragments;

(e) sequencing or obtaining sequences of the enriched amplified mcfDNA fragments;

(f) assigning one of more of the amplified mcfDNA fragment sequences as belonging to a particular microbial species, or as comprising a microbial DNA signature,

based on the closest sequence match of the hypervariable region to a reference microbial database; and

(g) optionally, calculating a microbial community composition for the sample, based on a relative abundance of the mcfDNA fragment sequences assigned to each microbial species,

wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

13. The method of any one of claims 1, 2, or 12, further comprising performing an affinity purification of the enriched mcfDNA fragments.

14. A method of diagnosing or monitoring a disease or disorder, comprising:

performing, on a sample comprising microbial cell-free DNA (mcfDNA), a polymerase chain reaction (PCR) using: (i) a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and (ii) an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprise a hypervariable region, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region, thereby generating amplified mcfDNA fragments;

sequencing the amplified mcfDNA fragments or obtaining the sequences of the amplified mcfDNA fragments;

assigning one or more of the amplified mcfDNA fragment sequences as comprising a microbial DNA signature or as belonging to a particular microbial species, based on the closest sequence match of the hypervariable region to a reference microbial database; and

optionally, calculating a microbial community composition for the sample, based on the relative abundance of the mcfDNA fragment sequences assigned to each microbial species,

wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

15. The method of any one of claims 11, 12, or 14, where the disease or disorder comprises: tuberculosis and other diseases caused by *Mycobacterium* species; determining pulmonary infection risks and causes in cystic fibrosis patients; determining the risk and onset of sepsis in patients with compromised immune systems; detection of opportunistic bacterial pathogens originating from the oral cavity that have been linked to Alzheimer's disease, pancreatic cancer and other serious conditions such as endocarditis; women's health issues including Chlamydia linked to mucopurulent cervicitis, pelvic inflammatory disease, tubal factor infertility, ectopic pregnancy and cervical cancer; detection and monitoring of progression of cancer; monitoring of minimal residual disease after oncology treatments; detection and monitoring of progression and minimal residual disease of breast cancer including triple negative breast cancer; detection of esophageal cancer, precancerous colonic polyps and early stage colorectal cancer, and detection and monitoring of progression and minimal residual disease of gastrointestinal cancers in general; detection and monitoring of progression and minimal residual disease in lung cancer; non-invasive analysis of the microbiome in pancreatic cancer patients to propose treatment protocols and prognostics for long-term survival; detection of *Clostridium difficile* infections; post-transplant bloodstream infections and Graft versus Host Disease (GvHD); detection of hospital acquired infections by emerging pathogens of clinical concern; detection of an infection in an immune compromised person; or detection of infection or inflammation of the gastrointestinal track in Irritable Bowel Disease (Crohn's disease, Ulcerative colitis); monitoring the risk of infection of inflammation in patients suffering from rheumatoid arthritis, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis, plaque psoriasis; and combinations thereof.

16. The method of any one of claims 11, 12, or 14, wherein the disease or disorder comprises colon cancer, breast cancer, or non-small cell lung cancer.

17. The method of any one of claims 11, 12, or 14, wherein the disease or disorder comprises a specific type of cancer, a specific early or later stage of cancer, or a

specific tumor subtype driven by the location of the tumor, and combinations thereof.

18. The method of any one of claims 1, 2, or 12, wherein the additional primer in the A-PCR is used in combination with a set of at least two phased primers, wherein the at least two phased primers are identical to the additional primer except for being phased by a single nucleotide.

19. The method of claim 18, wherein the set of phased primers comprises five phased primers.

20. The method of any one of claims 1, 2, 12, or 14 wherein the conserved region has an average sequence variance score of greater than 0.175.

21. The method of any one of claims 1, 2, 12, or 14 wherein the hypervariable region has an average sequence variance score of less than 0.15.

22. The method of any one of claims 1, 2, 12, or 14 wherein the hypervariable region has an average sequence variance score of less than 0.1.

23. The method of any one of claims 1, 2, 12, or 14 wherein the hypervariable region has an average sequence variance score of less than 0.075.

24. The method of any one of claims 1, 2, 12, or 14 wherein the conserved region spans 18 to 40 nucleotides, 20 to 30 nucleotides, or 22 to 28 nucleotides of the phylogenetic marker gene.

25. The method of claim 2, wherein the ends of the mcfDNA comprise an adaptor and the additional primer is complementary to a repaired version of the adaptor.

26. The method of any one of claims 1, 12, or 14, wherein the adaptor is a double stranded asymmetric linker cassette comprising a 5' asymmetrical end and a 3' end where the two strands are complementary.

27. The method of claim 26, wherein the asymmetric linker cassette is a Y-shaped linker cassette or a single arm linker cassette.

28. The method of claim 26, wherein the additional primer is complementary to a repaired 5' end of the asymmetric linker cassette, and wherein in the E-PCR,

polymerase extension from the degenerate primer results in repair of the asymmetric linker cassette.

29. The method of any one of claims 1, 2, 12, or 14, comprising performing one or more reactions to repair the ends of the mcfDNA.

30. The method of any one of claims 1, 2, 12, or 14, wherein the degenerate primer comprises one or more sequencing adaptor sequences.

31. The method of any one of claims 1, 2, 12, or 14, further comprising adding one or more sequencing adaptor sequences to the amplified mcfDNA fragments in an additional amplification reaction or to the enriched mcfDNA fragments in the A-PCR.

32. The method of any one of claims 1, 2, 12, or 14, wherein the phylogenetic marker gene comprises *rpoB*.

33. The method of any one of claims 1, 2, 12, or 14, wherein the phylogenetic marker gene comprises 16S rRNA.

34. The method of any one of claims 1, 2, 12, or 14, wherein the phylogenetic marker gene comprises a combination of two or more of *rpoB*, *cpn60*, *tuf*, or 16S rRNA.

35. The method of any one of claims 1, 2, 12, or 14, wherein the phylogenetic marker gene comprises *rpoB* and the conserved region comprises nucleotide positions 1327 – 1355 based on the *Escherichia coli rpoB* gene sequence.

36. The method of any one of claims 1, 2, 12, or 14, wherein the phylogenetic marker gene comprises 16S rRNA and the conserved region comprises nucleotide positions 781-805 based on the *Escherichia coli* 16S rRNA gene sequence.

37. The method of any one of claims 1, 2, 12, or 14, wherein the degenerate primer comprises a set of degenerate oligonucleotide primers comprising a mixture of oligonucleotides having a V:W:X:Y:Z ratio of 1.2:0.8-1.2:0.8-1.2:0.8-1.2:1.5-3, wherein V is the sequence 5’ to 3’ GCRCGRTTDCCNARRTGRTCRATRTC, wherein W is the sequence 5’ to 3’ GCRCGRTTDCCNARRTGRTCRATATC, wherein X is the sequence 5’ to 3’ GCRCGRTTDCCNARGTGRTCRATRTC, wherein Y is the sequence 5’ to 3’ GCRCGRTTDCCNARGTGRTCRATATC, and wherein Z is the sequence 5’ to 3’

GCRCGGTTRCCNARGTGGTCGATRTC, wherein the set of degenerate oligonucleotide primers comprises an affinity purification group at a 5' terminus, wherein A = adenine, G = guanidine, C = cytosine, T = thymine, R = purine (A or G), D = not C (A, T or G), and N = any nucleotide (A, G, C or T), and wherein D and N can be replaced with inosine.

38. The method of any one of claims 1, 2, 12, or 14, wherein the one or more degenerate primers comprises one or both of 16SV4-ePCR and 16SV4-aPCR.

39. The method of any one of claims 1, 2, 12, or 14, wherein the phylogenetic marker gene comprises 16S rRNA and the conserved region comprises a V1, V2, V3, V4, V5, or V6 region of the 16S rRNA phylogenetic marker gene.

40. The method of any one of claims 1, 2, 12, or 14, wherein the one or more degenerate primers comprises RpoB-ePCR mix, RpoB-aPCR mix, 16SV4-ePCR, or 16SV4-aPCR, and combinations thereof.

41. The method of any one of claims 1, 2, 12, or 14, wherein the phylogenetic marker gene comprises DNA gyrase subunit B (*gyrB*), heat shock protein 60 (*hsp60*), superoxide dismutase A protein (*sodA*), TU elongation factor (*tuf*), DNA recombinase proteins (including *recA, recE*), *trr1* gene that encodes for thioredoxin reductase; *rim8* gene that encodes for a protein involved in the proteolytic activation of a transcriptional factor in response to alkaline pH; *kre2* gene that encodes for α-1,2-mannosyltransferase; or *erg6* gene that encodes for Δ(24)-sterol C-methyltransferase.

42. The method of any one of claims 1, 2, 12, or 14, wherein the microbial phylogenetic marker gene comprises a human fungal phylogenetic marker gene

43. The method of claim 42, wherein the human fungal phylogenetic marker gene comprises nuclear ribosomal internal transcribed spacer region 1 (ITS1) or nuclear ribosomal internal transcribed spacer region 2 (ITS2).

44. The method of claim 42, wherein the amplified mcfDNA fragments comprise mcfDNA from one or a combination of members of the Ascomycota, Basidiomycota and Mucoromycota, including *Alternaria* species, *Aspergillus* species, *Blastomyces* species, *Candida* species, *Capnodiales* species, *Cladosporium* species, *Malassezia* species, *Phaeosphaeria* species, *Pseudozyma*

species, *Saccharomyces* species, *Sporobolomyces* species, *Vishniacozyma* species, and *Yarrowia* species.

45. The method of any one of claims 1, 2, 12, or 14, further comprising including in the PCR or the E-PCR and the A-PCR a primer to a functional gene to determine a presence of the functional gene in the sample.

46. The method of claim 45, where the functional gene is a virulence factor, a PKS gene cluster essential for colibactin synthesis, or a choline trimethylaminelyase gene.

47. The method of any one of claims 1, 2, 12, or 14, further comprising including in the PCR or the E-PCR and the A-PCR a primer to a viral gene to determine the presence of the viral gene in the sample.

48. The method of claim 47, wherein the viral gene comprises a human DNA- or RNA-based oncovirus gene.

49. The method of claim 48, wherein the oncovirus is one or a combination of Epstein-Barr Virus (EBV), Human Papillomavirus (HPV), Hepatitis B virus (HBV), Human Herpesvirus-8 (HHV-8), or Merkel Cell Polyomavirus (MCPyV).

50. The method of any one of claims 1, 2, 12, or 14, wherein the sample comprises a bodily fluid, a tissue, or an extracellular bodily substance.

51. The method of claim 50, wherein the bodily fluid comprises whole blood, a blood fraction, serum, plasma, or combinations thereof.

52. The method of claim 50, wherein the sample comprises a biopsy sample from a solid tumor, a skin graft, a liquid biopsy sample other than blood, or combinations thereof.

53. The method of claim 50, wherein the sample comprises a stool sample.

54. The method of any one of claims 1, 2, 12, or 14, wherein the mcfDNA comprises an average fragment length of less than about 100 bp.

55. The method of any one of claims 1, 2, 12, or 14, wherein a percentage of the mcfDNA in the sample is less than about 0.05%, less than about 0.1%, less than about 1%, less than about 5%, or less than about 15%.

56. The method of any one of claims 1, 2, 12, or 14, wherein the at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprising the hypervariable region comprises less than 150 adjacent nucleotides.

57. The method of any one of claims 1, 2, 12, or 14, wherein the at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprising the hypervariable region comprises less than 75 adjacent nucleotides.

58. The method of any one of claims 1, 2, 12, or 14, wherein the at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprising the hypervariable region comprises less than 50 adjacent nucleotides.

59. The method of any one of claims 1, 2, 12, or 14, wherein the microbial phylogenetic gene is a eubacterial phylogenetic gene.

60. The method of any one of claims 1, 2, 12, or 14, wherein the microbial phylogenetic gene is a eukaryotic, fungal, or bacterial phylogenetic gene.

61. The method of any of claims 7-10, wherein the calculated microbial community composition is a screening for one or more of: tuberculosis and other diseases caused by *Mycobacterium* species; pulmonary infection risks and causes in cystic fibrosis patients; the risk and onset of sepsis in patients with compromised immune systems; detection of opportunistic bacterial pathogens originating from the oral cavity that have been linked to Alzheimer's disease, pancreatic cancer and other conditions such as endocarditis; women's health issues including Chlamydia linked to mucopurulent cervicitis, pelvic inflammatory disease, tubal factor infertility, ectopic pregnancy and cervical cancer; detection and monitoring of progression in cancer; monitoring of minimal residual disease after oncology treatments; detection and monitoring of progression and minimal residual disease of breast cancer including triple negative breast cancer; detection of esophageal cancer, precancerous colonic polyps and early stage colorectal cancer, and detection and monitoring of progression and minimal residual disease of gastrointestinal cancers in general; detection and monitoring of progression and minimal residual disease in lung cancer; non-invasive analysis of the microbiome in pancreatic cancer patients to propose treatment protocols and prognostics for long-term survival; detection of *Clostridium difficile* infections; post-transplant bloodstream infections and Graft versus Host Disease (GvHD); detection of

hospital acquired infections by emerging pathogens of clinical concern; detection of an infection in an immune compromised person; or detection of infection or inflammation of the gastrointestinal track in Irritable Bowel Disease (Crohn's disease, Ulcerative colitis).

62. The method of any one of claims 1, 2, 12, or 14, wherein the enriched amplified mcfDNA fragments or the amplified mcfDNA fragments comprise mcfDNA from one or more members of: *Flavobacterium sp., Staphylococcus auricularis, Pseudomonas toyotomiensis, Rheinheimera sediminis, Finegoldia magna, Parvularcula sp., Pseudomonas stutzeri, Pseudomonas soyae, Pseudomonas saponiphila, Pseudomonas sp., Peptoniphilus harei, Quisquiliibacterium sp., Azoarcus sp., Sphingopyxis terrae, uncultured Clostridiales bacterium strain UMGS460, Staphylococcus schweitzeri, Flavobacterium erciyesense, Rhodococcus yananensis, Dietzia massiliensis, Cutibacterium acnes subsp. elongatum, Angustibacter aerolatus, Aerococcus urinae, Klebsiella quasivariicola, Comamonas fluminis, Mycobacterium tuberculosis, Mycobacterium abscessus, Mycobacterium avium, Mycobacterium chimaera, Mycobacterium leprae, Mycobacterium xenopi, Mycobacterium (para)intracellulare, Mycobacterium kansasii, Mycobacterium gilvum, Mycolicibacterium gen. nov. ("fortuitum-vaccae" clade), Mycobacterium gen. ("tuberculosis-simiae" clade), Staphylococcus aureus, Staphylococcus argenteus, Staphylococcus schweitzeri, Pseudomonas aeruginosa, Burkholderia cepacia complex, Burkholderia ubonensis, Burkholderia species Nov., Burkholderia multivorans, Burkholderia pseudomultivorans, Burkholderia pseudomallei, Burkholderia mallei, Trinickia* species, *Burkholderia thailandensis, Haemophilus* species, *Haemophilus influenzae, Haemophilus parainfluenzae, Streptococcus* species at the various group and species levels, *Streptococcus dysgalactiae, Streptococcus pyogenes, Streptococcus mutans, Streptococcus suis, Streptococcus mitis, Streptococcus pneumoniae, Streptococcus agalactiae, Streptococcus anginosus, Streptococcus intermedius, Streptococcus constellatus, Streptococcus equi subsp. zooepidemicus, Streptococcus oralis, Streptococcus gordonii, Streptococcus uberis, Streptococcus parasanguinis, Streptococcus sanguinis Streptococcus parauberis, Streptococcus infantarius, Streptococcus iniae, Streptococcus salivarius, Streptococcus thermophilus, Streptococcus vestibularis, Streptococcus bovis, Streptococcus gallolyticus subsp. gallolyticus, Streptococcus*

gallolyticus subsp. macedonicus, Streptococcus gallolyticus subsp. pasteurianus, Streptococcus equinus, Enterococcus faecalis, Enterococcus faecium, Porphyromonas gingivalis, Porphyromonas cangingivalis, Porphyromonas uenonis, Porphyromonas endodontalis, Propionibacterium acidifaciens, Porphyromonas asaccharolytica, Porphyromonas macacae, Prevotella pallens, Prevotella histicola, Prevotella melaninogenica, Prevotella copri, Prevotella intermedia, Prevotella oral, Prevotella nanceiensis, Prevotella salivae, Prevotella nigrescens, Prevotella denticola, Prevotella buccae, Prevotella stercorea, Prevotella oris, Prevotella disiens, Prevotella bryantii, Prevotella shahii, Tannerella forsythia, Bacteroides fragilis, Helicobacter pylori, Chlamydia trachomatis, Neisseria meningitidis, Neisseria gonorrhoeae, Neisseria subflava, Neisseria perflava, Neisseria flavescens, Neisseria cinerea, Neisseria lactamica, Neisseria weaver, Neisseria zoodegmatis, Neisseria brasiliensis, Neisseria mucosa, Neisseria animaloris, Aggregatibacter actinomycetemcomitans, Aggregatibacter aphrophilus, Aggregatibacter segnis, Saccharopolyspora species, *Bacillus clausii,* members of the genera *Pseudoxanthomonas* and *Streptomyces, Fusobacterium nucleatum subsp. polymorphum, Fusobacterium hwasookii, Fusobacterium canifelinum, Fusobacterium nucleatum subsp. animalis, Fusobacterium periodonticum, Fusobacterium necrophorum subsp. funduliforme, Fusobacterium mortiferum, Fusobacterium varium, Fusobacterium nucleatum subsp. nucleatum, Fusobacterium ulcerans, Fusobacterium nucleatum subsp. vincentii, Fusobacterium equinum, Fusobacterium gonidiaformans, Fusobacterium necrogenes, Fusobacterium naviforme, Peptostreptococcus stomatis, Pseudonocardia asaccharolytica, Parvimonas species* including *Parvimonas oral* and *Parvimonas micra, Gemella* species including *Gemella morbillorum, Gemella haemolysans, Gemella palaticanis* and *Gemella sanguinis, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Acinetobacter baumannii, Acinetobacter lactucae, Acinetobacter pittii, Acinetobacter calcoaceticus, Acinetobacter oleivorans, Acinetobacter nosocomialis, Acinetobacter radioresistens, Acinetobacter variabilis, Acinetobacter courvalinii, Acinetobacter ursingii,* Enterobacteriaceae, *Enterobacter* species, *Escherichia species*, *Escherichia coli,* or *Klebsiella* species, *Klebsiella pneumoniae, Klebsiella oxytoca*, *Shigella* species, *Kluyvera* species, *Erwinia* species, *Salmonella* species, *Salmonella typhi, Salmonella enterica,*

Listeria species, *Listeria monocytogenes, Veillonella* species, *Veillonella atypica, Veillonella dispar, Veillonella oral, Stenotrophomonas* species, *Stenotrophomonas maltophilia*, and combinations thereof.

63. A system for amplifying microbial cell free DNA (mcfDNA), comprising:

    a. a reaction vessel;

    b. a reagent dispensing module; and

    c. software to execute the method of any of the foregoing claims,

wherein the method is executed at least partially robotically.

64. A set of degenerate oligonucleotides comprising a mixture of oligonucleotides having a V:W:X:Y:Z ratio of 1.2:0.8-1.2:0.8-1.2:0.8-1.2:1.5-3, wherein V is the sequence 5' to 3' CGRTTDCCNARRTGRTCRATRTCATC, wherein W is the sequence 5' to 3' CGRTTDCCNARRTGRTCRATATCATC, wherein X is the sequence 5' to 3' CGRTTDCCNARGTGRTCRATRTCATC, wherein Y is the sequence 5' to 3' CGRTTDCCNARGTGRTCRATATCATC, and wherein Z is the sequence 5' to 3' CGGTTRCCNARGTGGTCGATRTCGTC,

    wherein A = adenine, G = guanidine, C = cytosine, T = thymine, R = purine (A or G), D = not C (A, T or G), and N = any nucleotide (A, G, C or T), and wherein D and N can be replaced with inosine.

65. A set of degenerate oligonucleotides comprising a mixture of oligonucleotides having a V:W:X:Y:Z ratio of 1.2:0.8-1.2:0.8-1.2:0.8-1.2:1.5-3, wherein V is the sequence 5' to 3' GCRCGRTTDCCNARRTGRTCRATRTC, wherein W is the sequence 5' to 3' GCRCGRTTDCCNARRTGRTCRATATC, wherein X is the sequence 5' to 3' GCRCGRTTDCCNARGTGRTCRATRTC, wherein Y is the sequence 5' to 3' GCRCGRTTDCCNARGTGRTCRATATC, and wherein Z is the sequence 5' to 3' GCRCGGTTRCCNARGTGGTCGATRTC,

    wherein A = adenine, G = guanidine, C = cytosine, T = thymine, R = purine (A or G), D = not C (A, T or G), and N = any nucleotide (A, G, C or T), wherein D and N can be replaced with inosine, and wherein the oligonucleotides optionally comprise an affinity purification group at a 5' terminus.

66. The set of degenerate oligonucleotides of claim 64 or 65, wherein one or more nucleotides at positions represented by D or N are replaced by inosine.

67. A kit for amplifying microbial cell free DNA (mcfDNA), comprising:

    a. an adaptor for ligating to ends of mcfDNA;

    b. a degenerate primer comprising complementarity to a conserved region, wherein the degenerate primer comprises an affinity purification group, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region on the microbial phylogenetic marker gene, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region;

    c. an additional primer complementary to a repaired version of the adaptor;

    d. a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer; and

    e. instructions for performing: (i) an enrichment polymerase chain reaction (E-PCR) on the mcfDNA having the adaptor-ligated ends with the degenerate primer and the additional primer to generate enriched amplified mcfDNA fragments, an (ii) affinity purification of the enriched mcfDNA fragments, and (iii) an amplification PCR (A-PCR) on the enriched mcfDNA fragments using the modified degenerate primer and the additional primer to generate enriched amplified mcfDNA fragments.

68. The kit of claim 67, further comprising one or more additional degenerate primers comprising complementarity to one or more additional conserved regions of the microbial phylogenetic marker gene or to one or more additional genes.

69. The kit of claim 68, wherein the one or more additional genes comprises one or more additional microbial phylogenetic marker genes.

70. The kit of claim 68, wherein the one or more additional genes encodes virulence factors, secondary metabolites, or drug degradation pathways, and combinations thereof.

71. The kit of claim 67, further comprising a set of at least two phased primers, wherein the set of at least two phased primers are identical to the additional primer except for being phased by a single nucleotide, and wherein the instructions comprise instructions to use the additional primer in combination with the set of at least two phased primers in the A-PCR.

72. The kit of claim 71, wherein the set of phased primers comprises five phased primers.

73. The kit of claim 67, wherein the adaptor is an asymmetric linker cassette.

74. The kit of claim 67, wherein the degenerate primer comprises the oligonucleotide sequence 5' to 3' CGTGGACTACHVGGGTATCTA and has an affinity purification group at the 5' terminus, and wherein the modified degenerate primer comprises the oligonucleotide sequence 5' to 3' GACTACHVGGGTATCTA***ATCC***,

wherein A = adenine, G = guanidine, C = cytosine, T = thymine, H = not G (A, T or C), and V = not T (A, C, or G), and wherein H and V can be replaced with inosine

75. The kit of claim 67, wherein the degenerate primer comprises a set of degenerate oligonucleotide primers comprising a mixture of oligonucleotides having a V:W:X:Y:Z ratio of 1.2:0.8-1.2:0.8-1.2:0.8-1.2:1.5-3, wherein V is the sequence 5' to 3' GCRCGRTTDCCNARRTGRTCRATRTC, wherein W is the sequence 5' to 3' GCRCGRTTDCCNARRTGRTCRATATC, wherein X is the sequence 5' to 3' GCRCGRTTDCCNARGTGRTCRATRTC, wherein Y is the sequence 5' to 3' GCRCGRTTDCCNARGTGRTCRATATC, and wherein Z is the sequence 5' to 3' GCRCGGTTRCCNARGTGGTCGATRTC, wherein the set of degenerate oligonucleotide primers comprises an affinity purification group at a 5' terminus; and

wherein the modified degenerate primer comprises a set of modified degenerate oligonucleotides comprising of a mixture of oligonucleotides having a V:W:X:Y:Z ratio of 1.2:0.8-1.2:0.8-1.2:0.8-1.2:1.5-3, wherein V is the sequence 5' to 3' CGRTTDCCNARRTGRTCRATRTCATC, wherein W is the sequence 5' to 3' CGRTTDCCNARRTGRTCRATATCATC, wherein X is the sequence 5' to 3' CGRTTDCCNARGTGRTCRATRTCATC, wherein Y is the sequence 5' to

3' CGRTTDCCNARGTGRTCRATATCATC, and wherein Z is the sequence 5' to 3' CGGTTRCCNARGTGGTCGATRTCGTC,

wherein A = adenine, G = guanidine, C = cytosine, T = thymine, R = purine (A or G), D = not C (A, T or G), and N = any nucleotide (A, G, C or T), and wherein D and N can be replaced with inosine.

76. The kit of claim 67, wherein the microbial phylogenetic marker gene comprises *rpoB*.
77. The kit of claim 67, wherein the microbial phylogenetic marker gene comprises 16S rRNA.
78. The kit of claim 67, wherein the microbial phylogenetic marker gene comprises one or a combination of *RpoB* or *16S rRNA* genes.
79. The kit of claim 67, wherein the microbial phylogenetic marker gene comprises 16S rRNA gene and the conserved region comprises a V1, V2, V3, V4, V5, or V6 region of the 16S rRNA gene.
80. The kit of claim 67, wherein the microbial phylogenetic marker gene comprises *rpoB* and the conserved region comprises nucleotide positions 1327 – 1358 based on *Escherichia coli rpoB* gene sequence.
81. The kit of claim 67, wherein the microbial phylogenetic marker gene comprises *16S rRNA* and the conserved region comprises nucleotide positions 781-805 based on *Escherichia coli 16S rRNA* gene sequence.
82. The kit of claim 67, further comprising a primer for a functional gene to determine a presence of the functional gene in the sample.
83. The kit of claim 82, wherein the functional gene is a virulence factor, a PKS gene cluster essential for colibactin synthesis, or a choline trimethylaminelyase gene, and combinations thereof.
84. The kit of claim 67, further comprising a primer for a viral gene to determine a presence of the viral gene in the sample.
85. The kit of claim 84, wherein the viral gene comprises a human DNA- or RNA-based oncovirus gene.

86. The kit of claim 85, wherein the oncovirus gene is from one or a combination of Epstein-Barr Virus (EBV), Human Papillomavirus (HPV), Hepatitis B virus (HBV), Human Herpesvirus-8 (HHV-8), or Merkel Cell Polyomavirus (MCPyV).

87. The kit of claim 67, wherein the microbial phylogenetic marker gene is a fungal phylogenetic marker gene.

88. The kit of claim 87, wherein the fungal phylogenetic marker gene comprises internal transcribed spacer region 1 (ITS1) or internal transcribed spacer region 2 (ITS2).

89. The kit of claim 87, wherein the amplified mcfDNA fragments comprise mcfDNA from one or a combination of members of the Ascomycota, Basidiomycota and Mucoromycota, including *Alternaria* species, *Aspergillus* species, *Blastomyces* species, *Candida* species, *Capnodiales* species, *Cladosporium* species, *Malassezia* species, *Phaeosphaeria* species, *Pseudozyma* species, *Saccharomyces* species, *Sporobolomyces* species, *Vishniacozyma* species, and *Yarrowia* species.

90. A method of diagnosing or monitoring a disease or disorder, comprising:

providing mcfDNA sequences, wherein the mcfDNA sequences are derived from a sample, wherein the mcfDNA sequences comprise sequence corresponding to (i) a conserved region spanning at least 18 nucleotides of a microbial phylogenetic marker gene and (ii) a hypervariable region at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region;

assigning one or more of the mcfDNA sequences as comprising a microbial DNA signature or as belonging to a particular microbial species based on the closest sequence match of the hypervariable region to a reference microbial database; and

optionally, calculating a microbial community composition for the sample, based on a relative abundance of the mcfDNA sequences assigned to each microbial species,

wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

91. A method of diagnosing or monitoring a disease or disorder, comprising:

sequencing amplified mcfDNA fragments derived from a sample, wherein the amplified mcfDNA fragments comprise sequence corresponding to (i) a conserved region spanning at least 18 nucleotides of a microbial phylogenetic marker gene and (ii) a hypervariable region at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region;

assigning one or more of the mcfDNA sequences as comprising a microbial DNA signature or as belonging to a particular microbial species based on the closest sequence match of the hypervariable region to a reference microbial database; and

optionally, calculating a microbial community composition for the sample, based on a relative abundance of the mcfDNA sequences assigned to each particular microbial species,

wherein one or a combination of: the microbial DNA signature; the microbial community composition; or (i) a presence, (ii) an absence, or (iii) a relative abundance of one or a combination of the microbial species, associated with a disease or disorder, indicates that the subject has the disease or disorder.

92. A method of diagnosing or monitoring a disease or disorder, comprising:

performing, on a sample comprising microbial cell-free DNA (mcfDNA), a polymerase chain reaction (PCR) using: (i) a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and (ii) an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprise a hypervariable region, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region, thereby generating amplified mcfDNA fragments; and

sequencing the amplified mcfDNA fragments or obtaining the sequences of the amplified mcfDNA fragments;

wherein the closer the sequence match between the amplified mcfDNA fragment sequences and the nucleic acid sequences in a reference sample for a disease or disorder the more likely the sample is to have the disease or disorder.

93. The method of any one of claims 90, 91, or 92, wherein the disease or disorder comprises: tuberculosis and other diseases caused by *Mycobacterium* species; determining pulmonary infection risks and causes in cystic fibrosis patients; determining the risk and onset of sepsis in patients with compromised immune systems; detection of opportunistic bacterial pathogens originating from the oral cavity that have been linked to Alzheimer's disease, pancreatic cancer and other serious conditions such as endocarditis; women's health issues including Chlamydia linked to mucopurulent cervicitis, pelvic inflammatory disease, tubal factor infertility, ectopic pregnancy and cervical cancer; detection and monitoring of progression of cancer; monitoring of minimal residual disease after oncology treatments; detection and monitoring of progression and minimal residual disease of breast cancer including triple negative breast cancer; detection of esophageal cancer, precancerous colonic polyps and early stage colorectal cancer, and detection and monitoring of progression and minimal residual disease of gastrointestinal cancers in general; detection and monitoring of progression and minimal residual disease in lung cancer; non-invasive analysis of the microbiome in pancreatic cancer patients to propose treatment protocols and prognostics for long-term survival; detection of *Clostridium difficile* infections; post-transplant bloodstream infections and Graft versus Host Disease (GvHD); detection of hospital acquired infections by emerging pathogens of clinical concern; detection of an infection in an immune compromised person; or detection of infection or inflammation of the gastrointestinal track in Irritable Bowel Disease (Crohn's disease, Ulcerative colitis); and combinations thereof.

94. The method of any one of claims 90, 91, or 92, wherein the disease or disorder comprises colon cancer, breast cancer, or non-small cell lung cancer.

95. The method of claim 94, wherein the disease or disorder comprises a specific type of cancer, an early or later stage cancer, or a specific tumor subtype driven by the location of the tumor, and combinations thereof.

96. The method of claim 95, wherein including one or a combination of circulating tumor DNA (ctDNA) signatures, ctDNA methylation patterns, or ctDNA gene

mutations that provide information on the disease or disorder indicates whether the sample is likely to have the disease or the disorder.

97. The method of any one of claims 12, 14, 90, 91, or 92, wherein the disease or disorder comprises detection of infection or inflammation in one or a combination of Irritable Bowel Disease, Crohn's disease, Ulcerative colitis, rheumatoid arthritis, juvenile idiopathic arthritis, psoriatic arthritis, ankylosing spondylitis, or plaque psoriasis, and wherein a presence of the amplified mcfDNA fragments from one or more pathogenic bacterial members of: Escherichia species, Escherichia coli, Shigella species, Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral, Pseudomonas species, Pseudomonas aeruginosa, Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species, Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium, indicates that the subject has the infection or inflammation.

98. The method of claim 97, wherein the disease or disorder comprises detection of infection or inflammation in Irritable Bowel Disease and the one or more pathogenic bacterial members further comprise one or more of *Rhuminococcus gnavus*, *Enterococcus faecium* or *Enterococcus faecalis*.

99. The method of claim 97 or 98, comprising calculating a percentage of the amplified mcfDNA fragments coming from the one or more bacterial pathogenic members as a fraction of a total amount of the amplified mcfDNA fragments in the sample, wherein the calculated percentage of less than one percent indicates an absence of infection or inflammation and the calculated percentage of greater than one percent indicates the presence of infection or inflammation.

100. The method of any one of claims 97-99, wherein the sample is a blood or plasma sample.

101. The method of any one of claims 97-100, wherein the presence of infection or inflammation is in the gastrointestinal tract.

102. The method of any one of claims 97-101, wherein the method is used in combination with determination of a level of lipo-polysaccharide (LPS) in the sample.

103. A method of detecting infection or inflammation, comprising:

(a) performing, on a sample comprising mcfDNA from a subject, an enrichment polymerase chain reaction (E-PCR) using a degenerate primer and an additional primer, thereby generating one or more enriched mcfDNA fragments, wherein:

i. the degenerate primer comprises complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and wherein the degenerate primer is oriented to prime polymerase extension of the hypervariable region, and

ii. the additional primers is complementary to a repaired version of an adaptor ligated to ends of the mcfDNA;

(b) performing an amplification PCR (A-PCR) on the enriched mcfDNA fragments using (i) a modified degenerate primer comprising a 3'-end extension of at least three nucleotides relative to the degenerate primer, and (ii) the additional primer, thereby generating enriched amplified mcfDNA fragments;

(c) sequencing or obtaining sequences of the enriched amplified mcfDNA fragments; and

(d) assigning one of more of the mcfDNA sequences as belonging to a pathogenic bacterial species based on the closest sequence match of the hypervariable region to a reference microbial database, wherein the pathogenic bacterial species comprise one or more of Escherichia species, Escherichia coli, Shigella species,

Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral, Pseudomonas species, Pseudomonas aeruginosa, Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species, Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium,

wherein a presence of the mcfDNA sequences from one or more of the pathogenic bacterial members indicates that the subject has the infection or inflammation.

104. A method of detecting infection or inflammation, comprising:

performing, on a sample comprising microbial cell-free DNA (mcfDNA) from a subject, a polymerase chain reaction (PCR) using: (i) a degenerate primer comprising complementarity to a conserved region, wherein the conserved region spans at least 18 nucleotides of a microbial phylogenetic marker gene, and (ii) an additional primer complementary to a repaired version of an adaptor ligated to ends of the mcfDNA, wherein at least 25 adjacent nucleotides upstream or downstream of an end of the conserved region comprises a hypervariable region, and the degenerate primer is oriented to prime polymerase extension of the hypervariable region, thereby generating amplified mcfDNA fragments;

sequencing the amplified mcfDNA fragments or obtaining the sequences of the amplified mcfDNA fragments; and

assigning one or more of the mcfDNA sequences as belonging to a pathogenic bacterial species based on the closest sequence match of the hypervariable region to a reference microbial database, wherein the pathogenic bacterial species comprise one or more of Escherichia species, Escherichia coli, Shigella species, Klebsiella species, Klebsiella pneumoniae, Klebsiella oxytoca, Kluyvera species, Erwinia species, Haemophilus species, Salmonella species, Salmonella typhi, Salmonella enterica, Listeria species, Listeria monocytogenes, Clostridium species, Clostridium difficile, Clostridium perfringens, Clostridium botulinum, Veillonella species, Veillonella atypica, Veillonella dispar, Veillonella oral, Pseudomonas species, Pseudomonas aeruginosa,

Stenotrophomonas species, Stenotrophomonas maltophilia, Burkholderia species, Burkholderia cepacia, Staphylococcus species, Staphylococcus aureus, Porphyromonas species, Fusobacterium species, Fusobacterium nucleatum, or Fusobacterium varium,

wherein a presence of the mcfDNA sequences from one or more of the pathogenic bacterial members indicates that the subject has the infection or inflammation.

105. The method of claim 103 or 104, wherein the pathogenic bacterial members further comprise one or more of Rhuminococcus gnavus, Enterococcus faecium, or Enterococcus faecalis.

106. The method of any one or claims 103-105, comprising calculating a percentage of the mcfDNA sequences coming from the one or more pathogenic bacterial members as a fraction of a total amount of the mcfDNA sequences in the sample, wherein the calculated percentage of less than one percent indicates an absence of infection or inflammation and the calculated percentage of greater than one percent indicates the presence of infection or inflammation.

107. The method of any one of claims 103-106, wherein the sample is a blood or plasma sample.

108. The method of any one of claims 103-107, wherein the presence of infection or inflammation is in the gastrointestinal tract.

109. The method of any one of claims 103-108, wherein the method is used in combination with determination of a level of lipo-polysaccharide (LPS) in the sample.

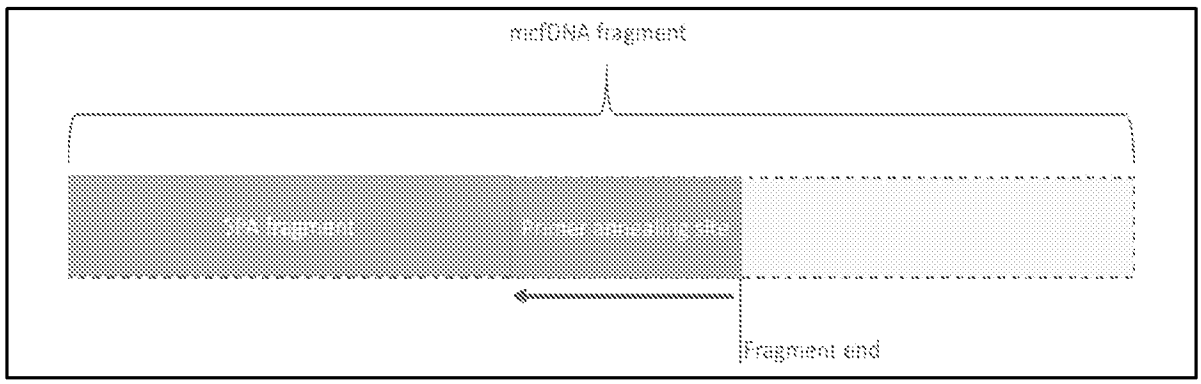

FIG. 1

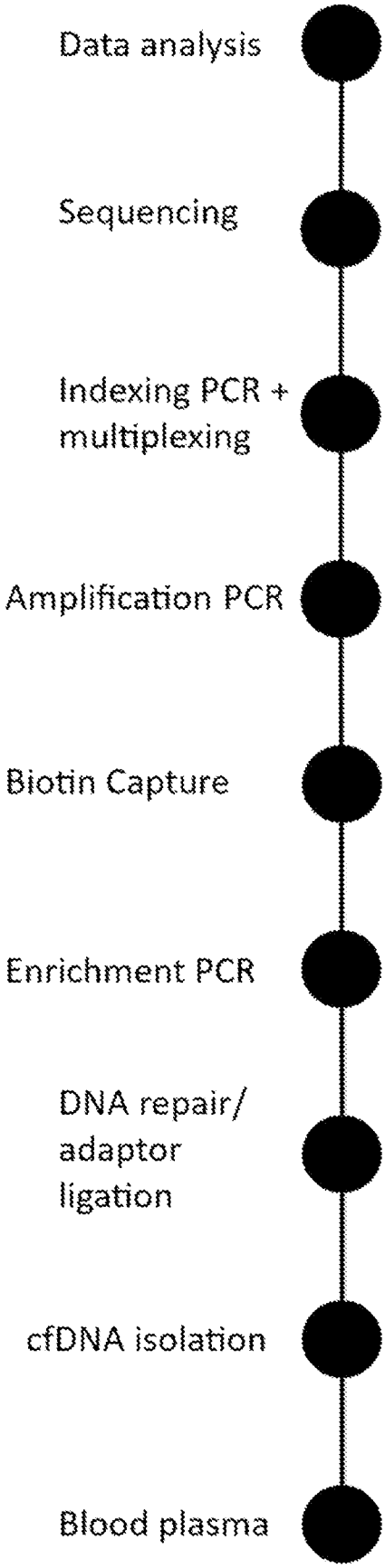
Data analysis
Sequencing
Indexing PCR + multiplexing
Amplification PCR
Biotin Capture
Enrichment PCR
DNA repair/ adaptor ligation
cfDNA isolation
Blood plasma

FIG. 2

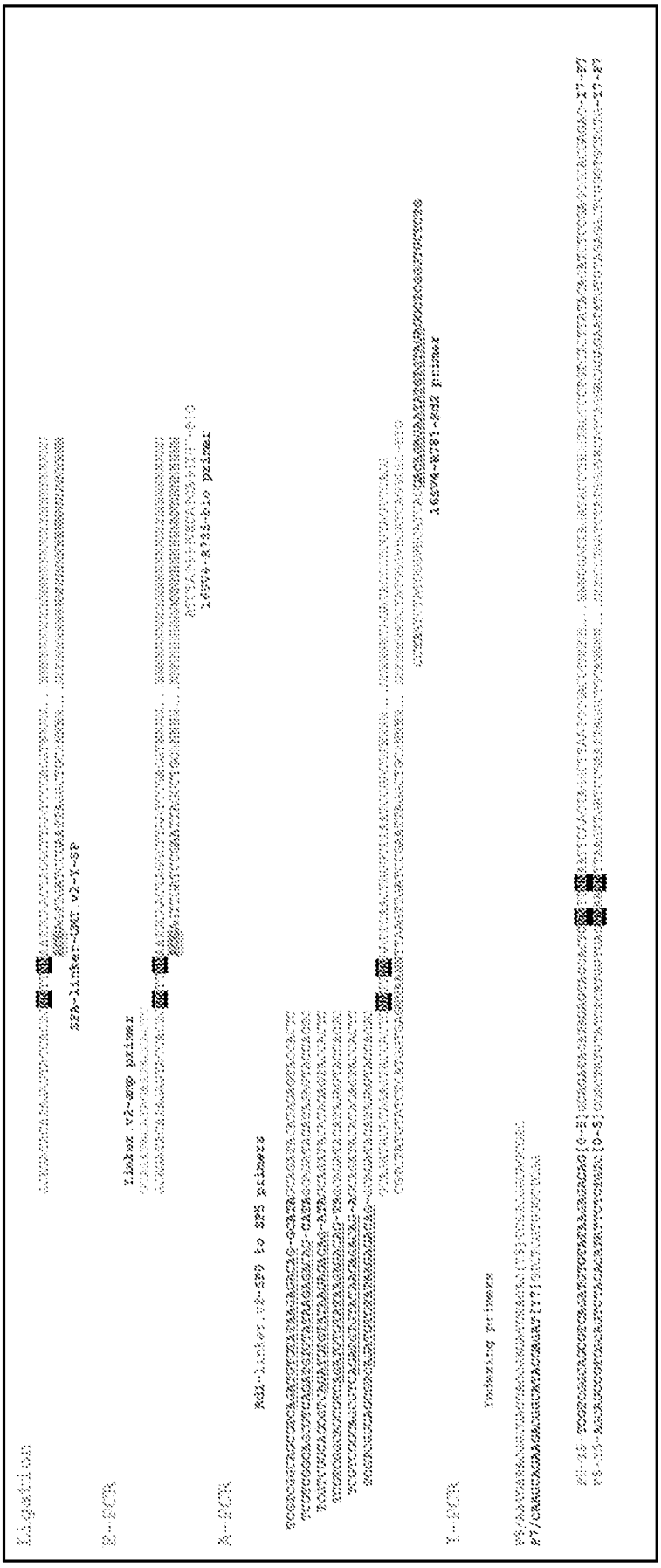

FIG. 3

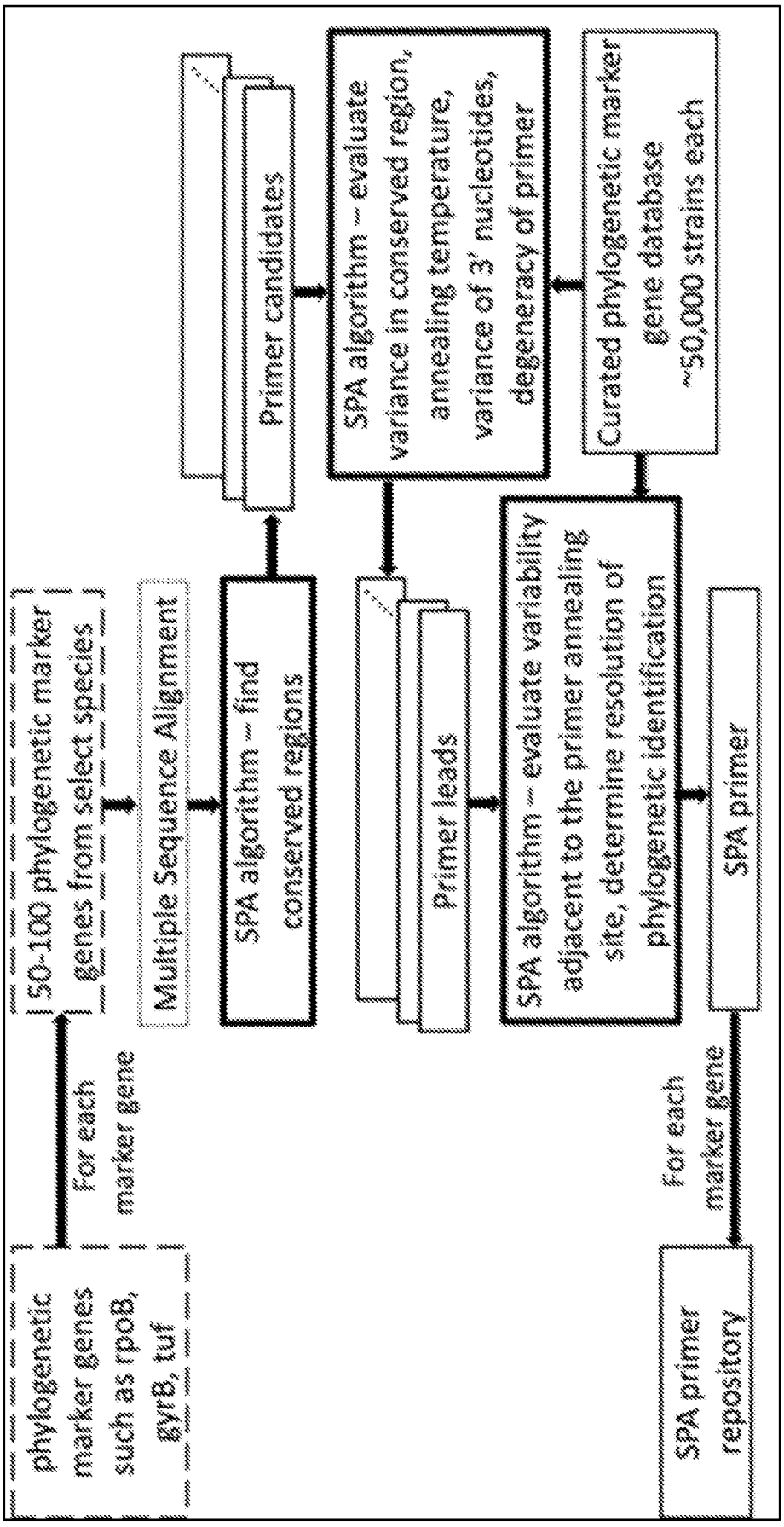
phylogenetic marker genes such as rpoB, gyrB, tuf
For each marker gene
50-100 phylogenetic marker genes from select species
Multiple Sequence Alignment
SPA algorithm – find conserved regions
Primer candidates
SPA algorithm – evaluate variance in conserved region, annealing temperature, variance of 3' nucleotides, degeneracy of primer
Curated phylogenetic marker gene database ~50,000 strains each
Primer leads
SPA algorithm – evaluate variability adjacent to the primer annealing site, determine resolution of phylogenetic identification
SPA primer
For each marker gene
SPA primer repository

FIG. 4

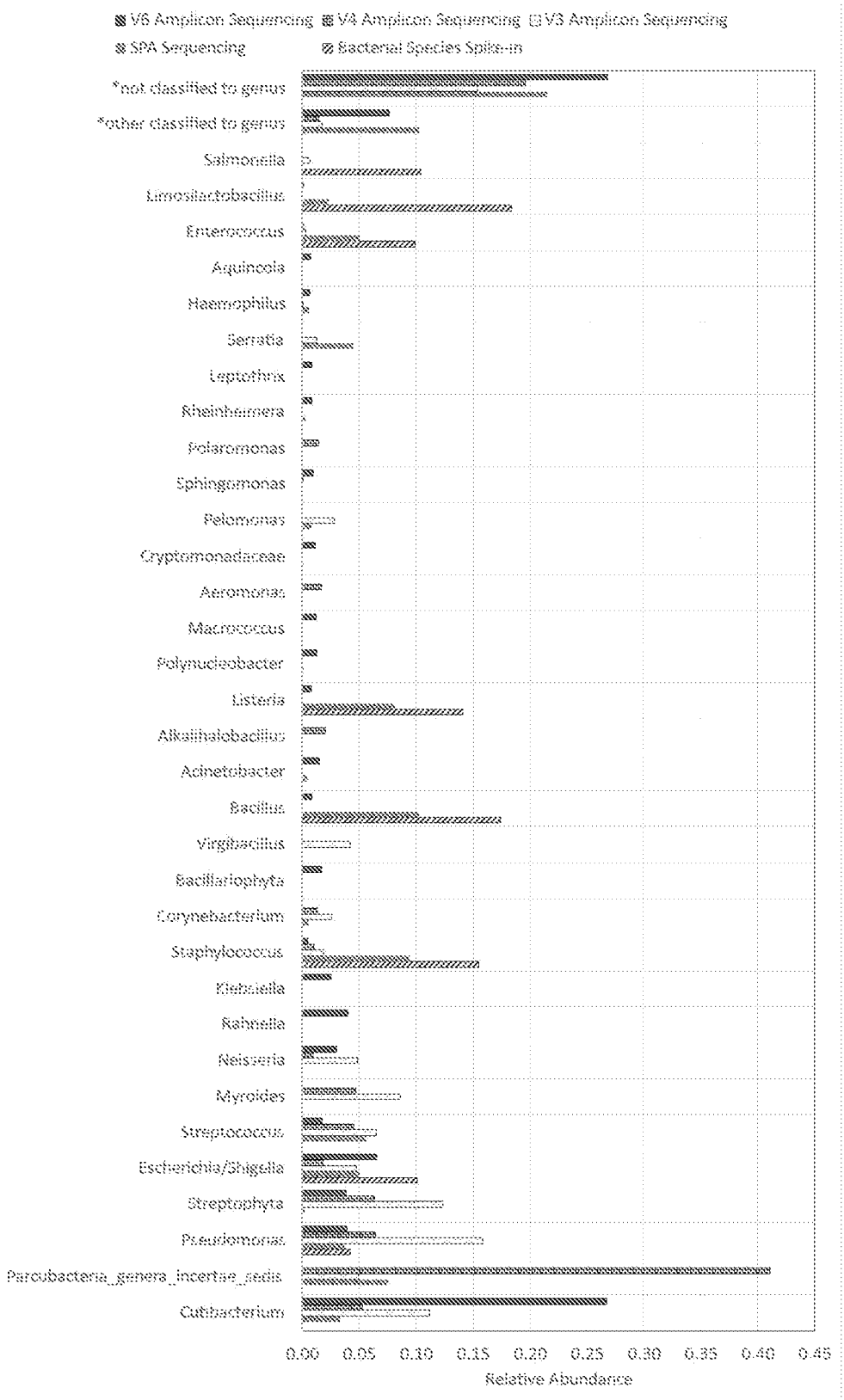
V6 Amplicon Sequencing
V4 Amplicon Sequencing
V3 Amplicon Sequencing
SPA Sequencing
Bacterial Species Spike-in
*not classified to genus
*other classified to genus
Salmonella
Limosilactobacillus
Enterococcus
Aquincola
Haemophilus
Serratia
Leptothrix
Rheinheimera
Polaromonas
Sphingomonas
Pelomonas
Cryptomonadaceae
Aeromonas
Macrococcus
Polynucleobacter
Listeria
Alkalihalobacillus
Acinetobacter
Bacillus
Virgibacillus
Bacillariophyta
Corynebacterium
Staphylococcus
Klebsiella
Rahnella
Neisseria
Myroides
Streptococcus
Escherichia/Shigella
Streptophyta
Pseudomonas
Parcubacteria_genera_incertae_sedis
Cutibacterium
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
Relative Abundance

FIG. 5

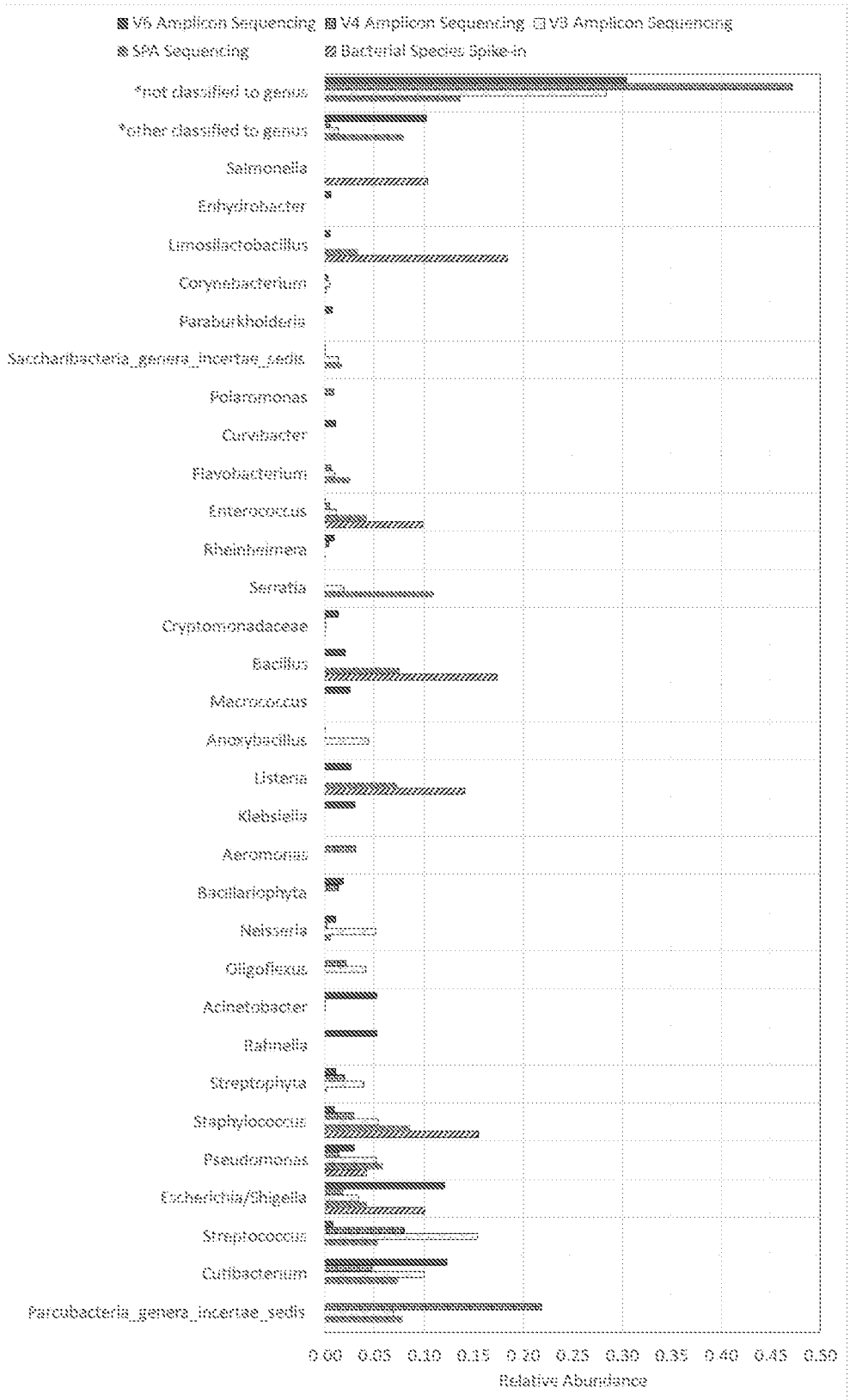
V6 Amplicon Sequencing
V4 Amplicon Sequencing
V3 Amplicon Sequencing
SPA Sequencing
Bacterial Species Spike-in
*not classified to genus
*other classified to genus
Salmonella
Enhydrobacter
Limosilactobacillus
Corynebacterium
Paraburkholderia
Saccharibacteria_genera_incertae_sedis
Polaromonas
Curvibacter
Flavobacterium
Enterococcus
Rheinheimera
Serratia
Cryptomonadaceae
Bacillus
Macrococcus
Anoxybacillus
Listeria
Klebsiella
Aeromonas
Bacillariophyta
Neisseria
Oligoflexus
Acinetobacter
Rahnella
Streptophyta
Staphylococcus
Pseudomonas
Escherichia/Shigella
Streptococcus
Cutibacterium
Parcubacteria_genera_incertae_sedis
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50
Relative Abundance

FIG. 6

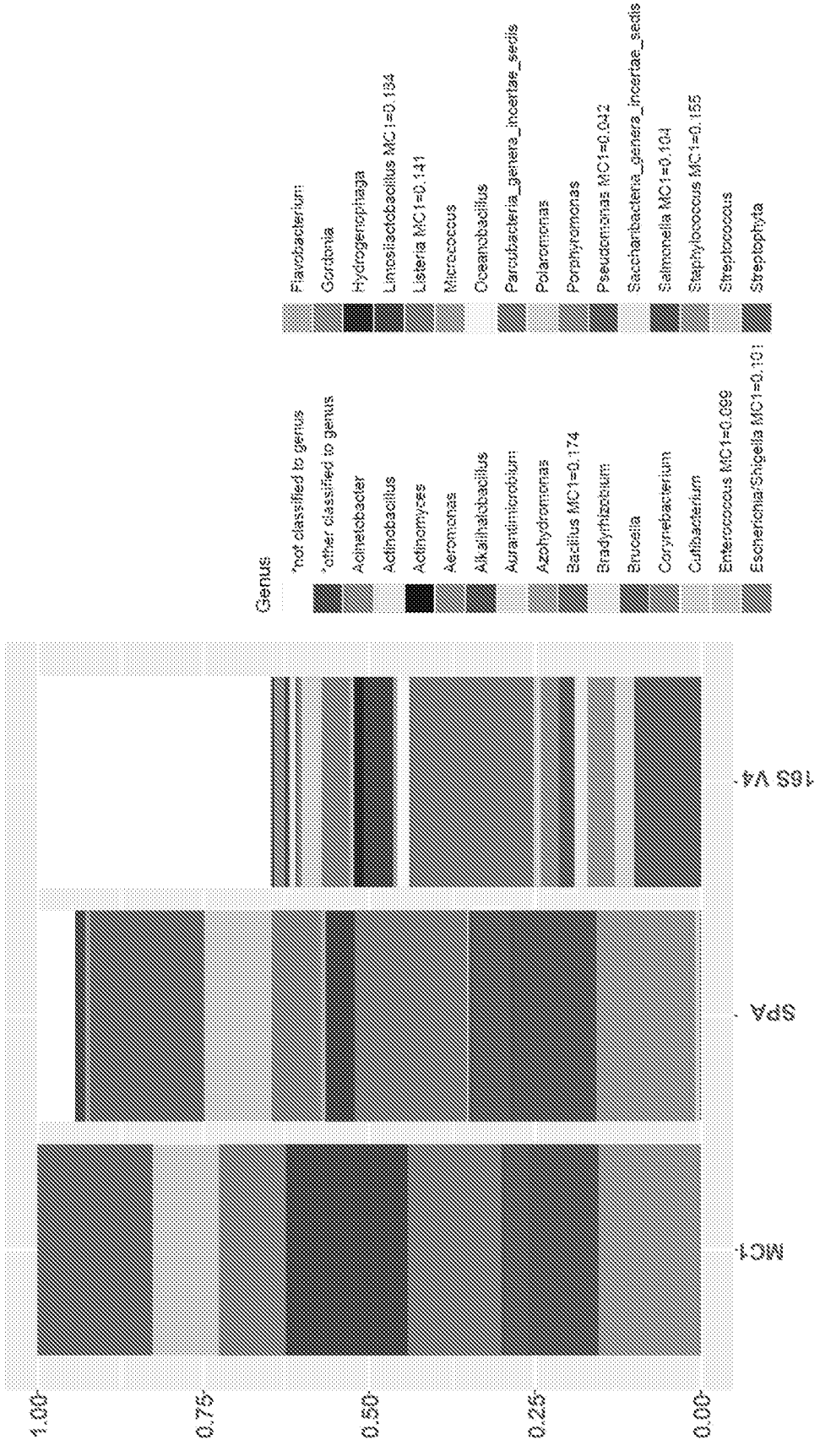
Genus
*not classified to genus
*other classified to genus
Acinetobacter
Actinobacillus
Actinomyces
Aeromonas
Alkalihalobacillus
Aurantimicrobium
Azohydromonas
Bacillus MC1=0.174
Bradyrhizobium
Brucella
Corynebacterium
Cutibacterium
Enterococcus MC1=0.069
Escherichia/Shigella MC1=0.191
Flavobacterium
Gordonia
Hydrogenophaga
Limosilactobacillus MC1=0.134
Listeria MC1=0.141
Micrococcus
Oceanobacillus
Parcubacteria_genera_incertae_sedis
Polaromonas
Porphyromonas
Pseudomonas MC1=0.042
Saccharibacteria_genera_incertae_sedis
Salmonella MC1=0.104
Staphylococcus MC1=0.155
Streptococcus
Streptophyta
1.00
0.75
0.50
0.25
0.00
MC1
QPA
16S V4

FIG. 7

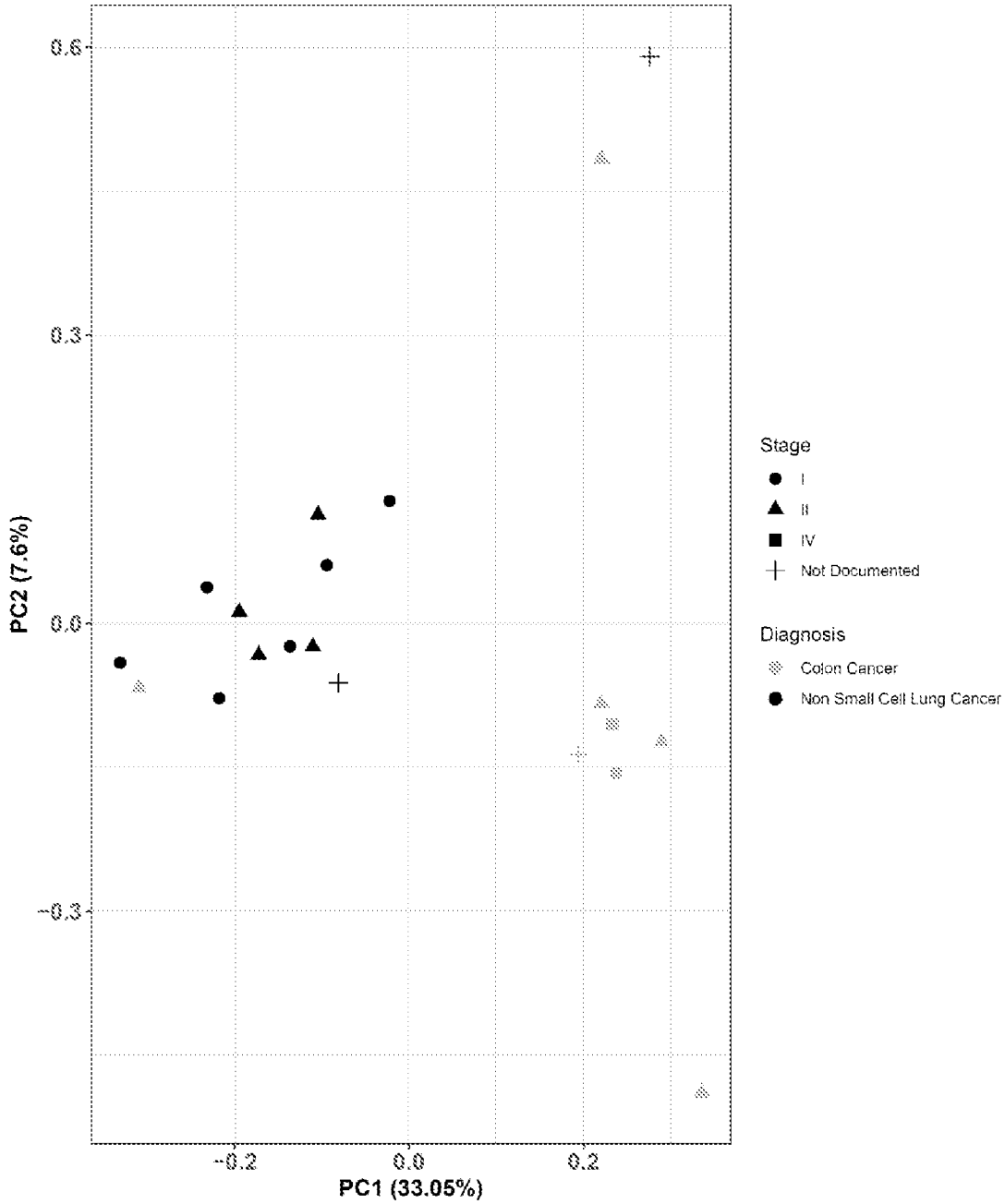
0.6
0.3
0.0
−0.3
PC2 (7.6%)
−0.2
0.0
0.2
PC1 (33.05%)
Stage
I
II
IV
Not Documented
Diagnosis
Colon Cancer
Non Small Cell Lung Cancer

Fig. 8

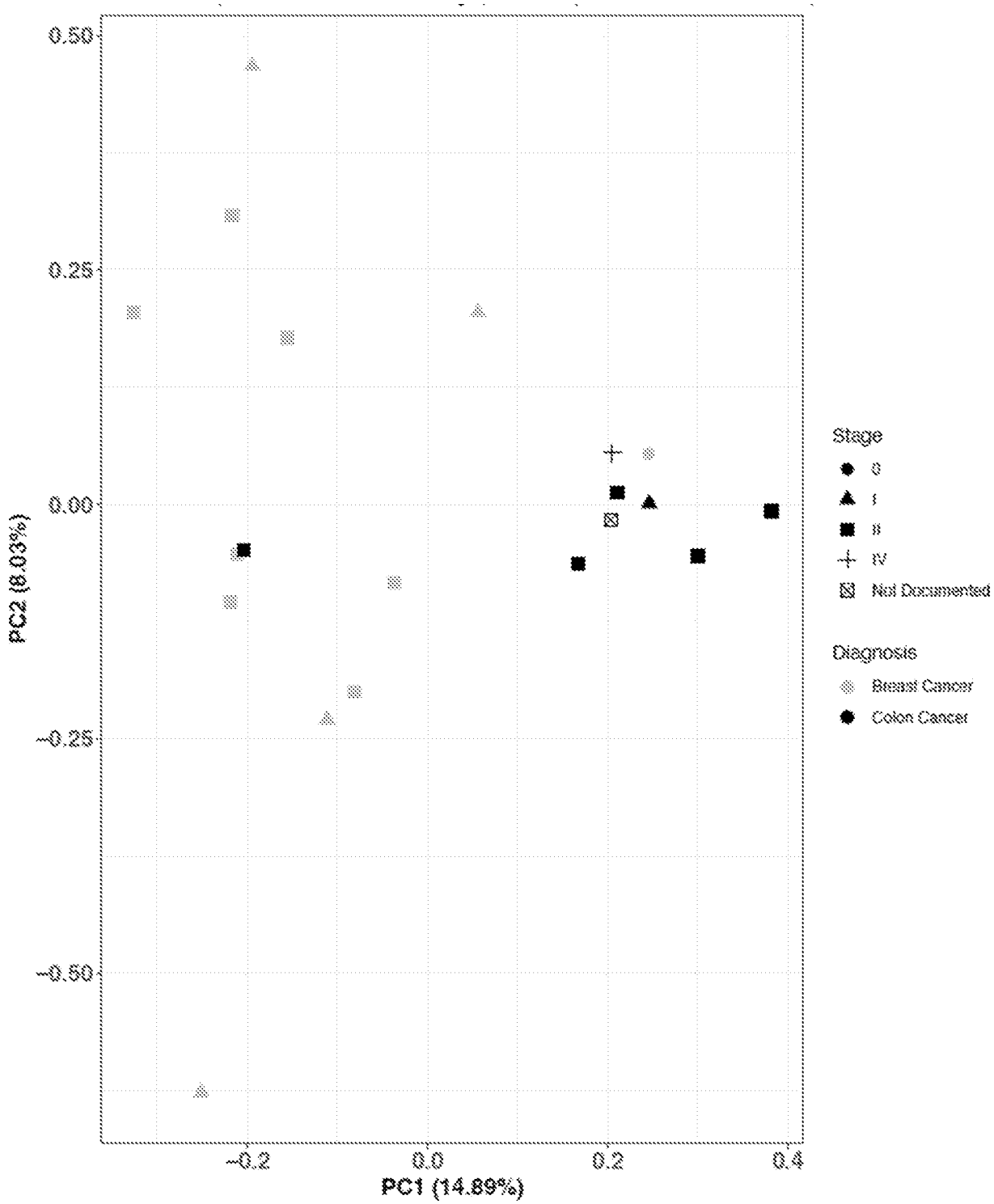
0.50
0.25
0.00
−0.25
−0.50
PC2 (8.03%)
−0.2
0.0
0.2
0.4
PC1 (14.89%)
Stage
0
I
II
IV
Not Documented
Diagnosis
Breast Cancer
Colon Cancer

Fig. 9

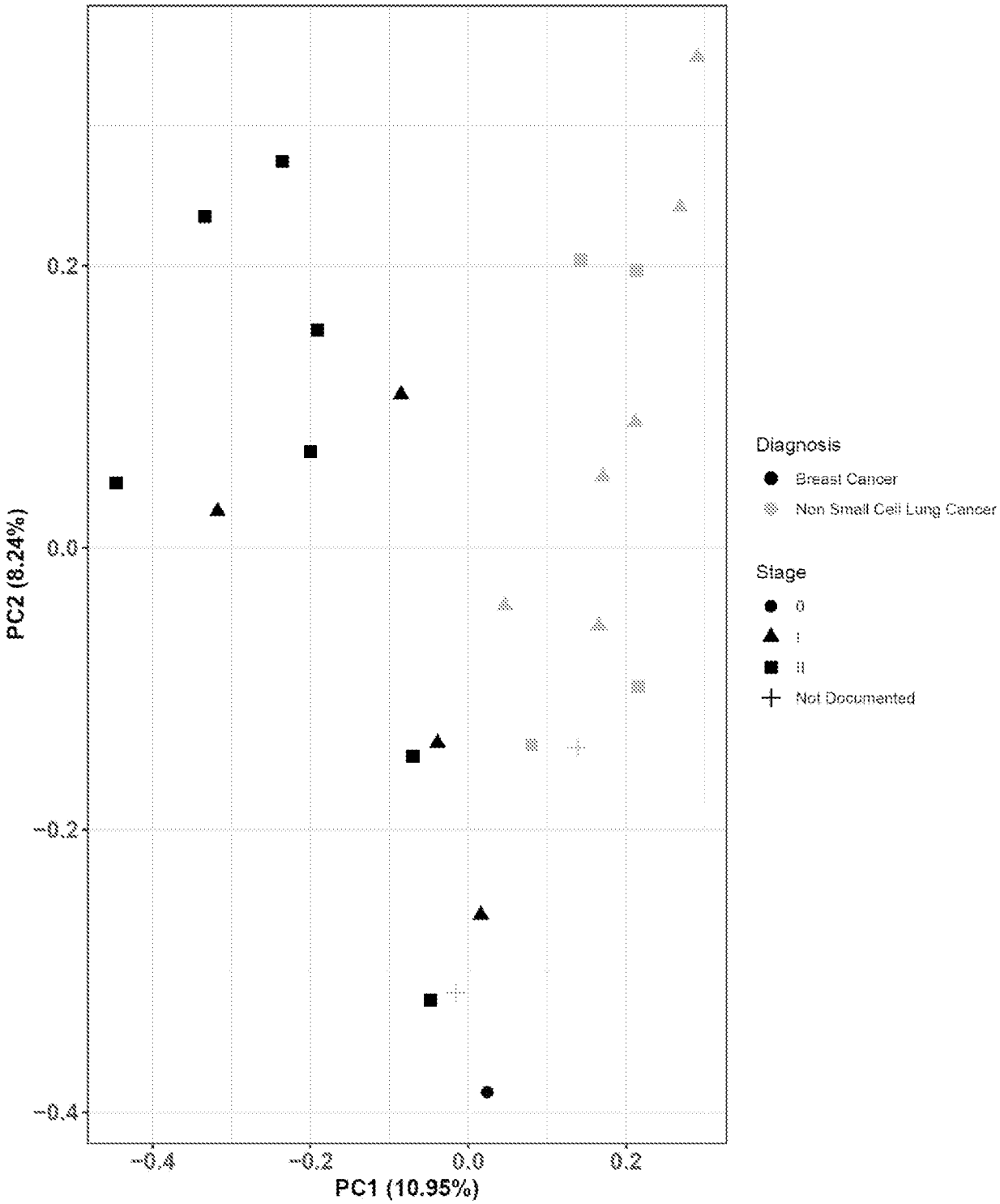
PC2 (8.24%)
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
PC1 (10.95%)
Diagnosis
Breast Cancer
Non Small Cell Lung Cancer
Stage
0
I
II
Not Documented

Fig. 10

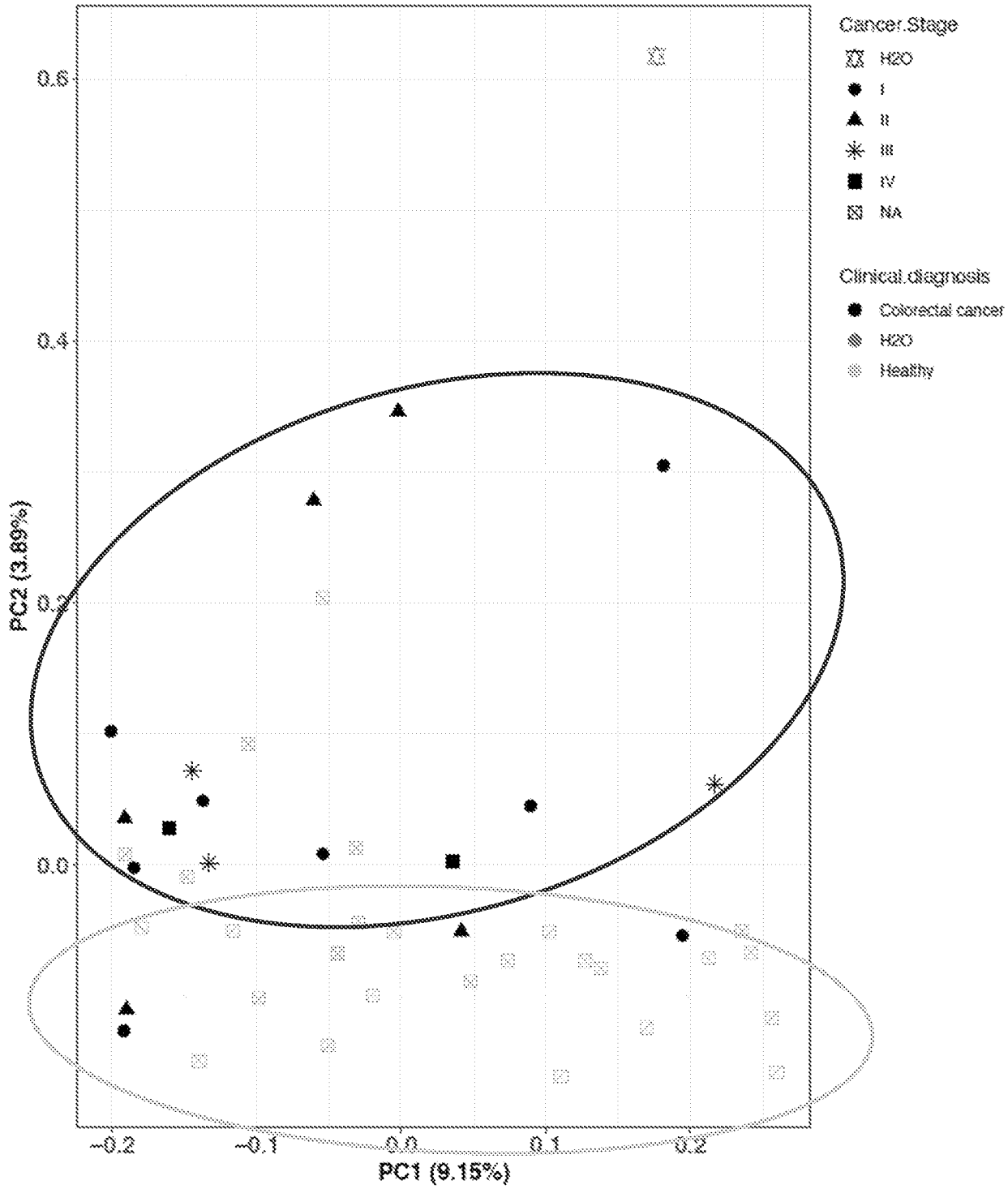
Cancer.Stage
H2O
I
II
III
IV
NA
Clinical.diagnosis
Colorectal cancer
H2O
Healthy
0.6
0.4
0.2
0.0
PC2 (3.89%)
−0.2
−0.1
0.0
0.1
0.2
PC1 (9.15%)

Fig. 11

E. coli RpoB gene position	*Absolute count of nucleotides per position*					*Relative count of nucleotides per position (percentage)*					Consensus sequence 5' --> 3'
	A	T	G	C	N*	A	T	G	C	N*	
1327	23	5	47471	5	1	0%	0%	100%	0%	0%	**G**
1328	47432	40	15	17	1	100%	0%	0%	0%	0%	**A**
1329	29	24804	30	22633	9	0%	52%	0%	48%	0%	**Y**
1330	106	8	47367	23	1	0%	0%	100%	0%	0%	**G**
1331	47436	43	18	8	0	100%	0%	0%	0%	0%	**A**
1332	97	29376	68	17957	7	0%	62%	0%	38%	0%	**Y**
1333	47092	31	271	111	0	99%	0%	1%	0%	0%	**A**
1334	39	46741	680	45	0	0%	98%	1%	0%	0%	**T**
1335	2450	11027	130	33882	16	5%	23%	0%	71%	0%	**Y**
1336	28	45	47411	21	0	0%	0%	100%	0%	0%	**G**
1337	47434	21	12	38	0	100%	0%	0%	0%	0%	**A**
1338	13	11944	18	35522	8	0%	25%	0%	75%	0%	**Y**
1339	3309	57	109	44029	1	7%	0%	0%	93%	0%	**C**
1340	47208	82	81	134	0	99%	0%	0%	0%	0%	**A**
1341	78	11706	154	35556	11	0%	25%	0%	75%	0%	**Y**
1342	56	19645	28	27771	5	0%	41%	0%	58%	0%	**Y**
1343	22	47451	22	10	0	0%	100%	0%	0%	0%	**T**
1344	10491	6115	15012	15873	14	22%	13%	32%	33%	0%	**N**
1345	2270	2327	42835	73	0	5%	5%	90%	0%	0%	**G**
1346	139	71	44391	2903	1	0%	0%	93%	6%	0%	**G**
1347	7203	15207	1649	23431	15	15%	32%	3%	49%	0%	**H**
1348	47286	114	71	34	0	100%	0%	0%	0%	0%	**A**

FIG. 12

1349	46840	39	475	151	0	99%	0%	1%	0%	0%	**A**
1350	89	9265	92	38054	5	0%	20%	0%	80%	0%	**Y**
1351	5226	45	66	42168	0	11%	0%	0%	89%	0%	**C**
1352	187	37	47276	5	0	0%	0%	100%	0%	0%	**G**
1353	5303	27876	2643	11671	0	6%	11%	25%	59%	0%	**H**
1354	4014	22	58	43408	0	0%	8%	91%	0%	0%	**G**
1355	80	34	47323	68	0	100%	0%	0%	0%	0%	**C**

FIG. 12 (continued)

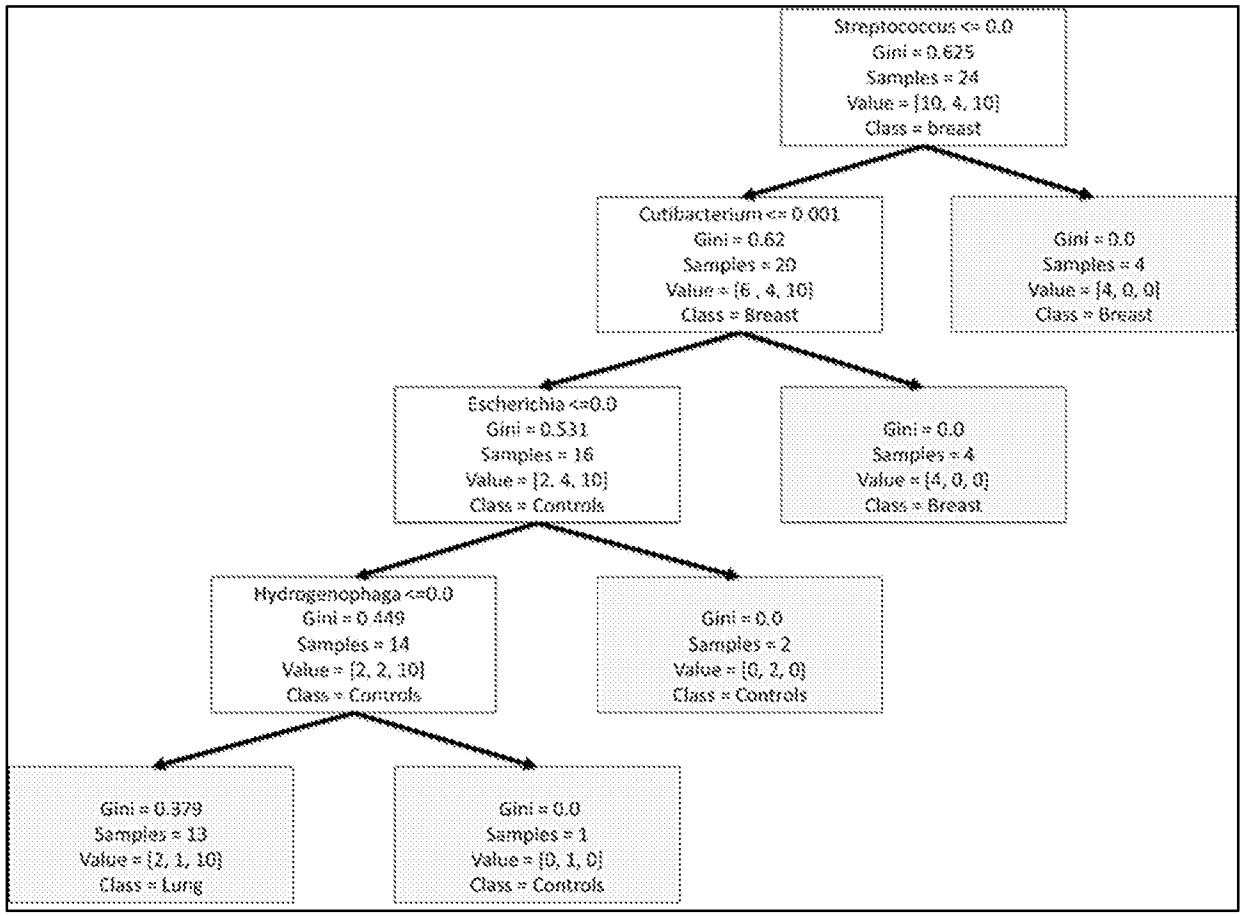
Streptococcus <= 0.0
Gini = 0.625
Samples = 24
Value = [10, 4, 10]
Class = breast
Cutibacterium <= 0.001
Gini = 0.62
Samples = 20
Value = [6 , 4, 10]
Class = Breast
Gini = 0.0
Samples = 4
Value = [4, 0, 0]
Class = Breast
Escherichia <=0.0
Gini = 0.531
Samples = 16
Value = [2, 4, 10]
Class = Controls
Gini = 0.0
Samples = 4
Value = [4, 0, 0]
Class = Breast
Hydrogenophaga <=0.0
Gini = 0.449
Samples = 14
Value = [2, 2, 10]
Class = Controls
Gini = 0.0
Samples = 2
Value = [0, 2, 0]
Class = Controls
Gini = 0.379
Samples = 13
Value = [2, 1, 10]
Class = Lung
Gini = 0.0
Samples = 1
Value = [0, 1, 0]
Class = Controls

FIG. 13

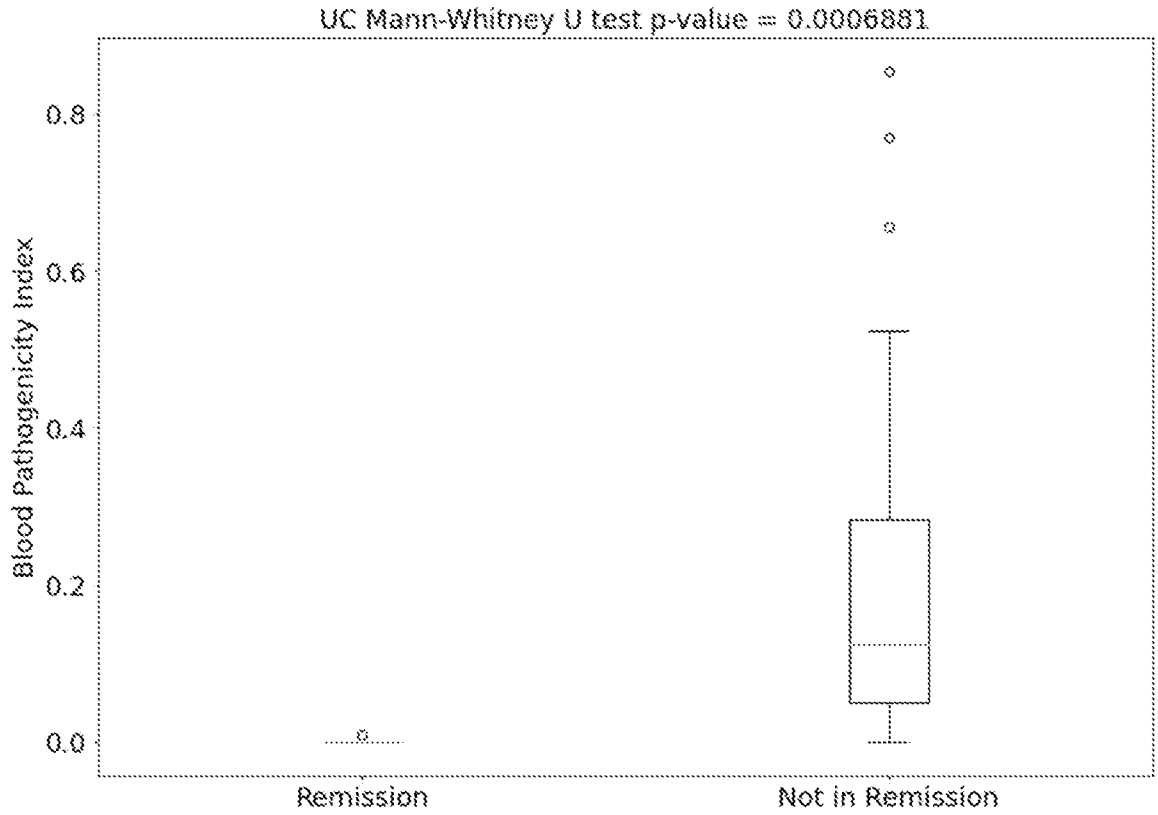
UC Mann-Whitney U test p-value = 0.0006881
Blood Pathogenicity Index
0.0
0.2
0.4
0.6
0.8
Remission
Not in Remission

FIG. 14A

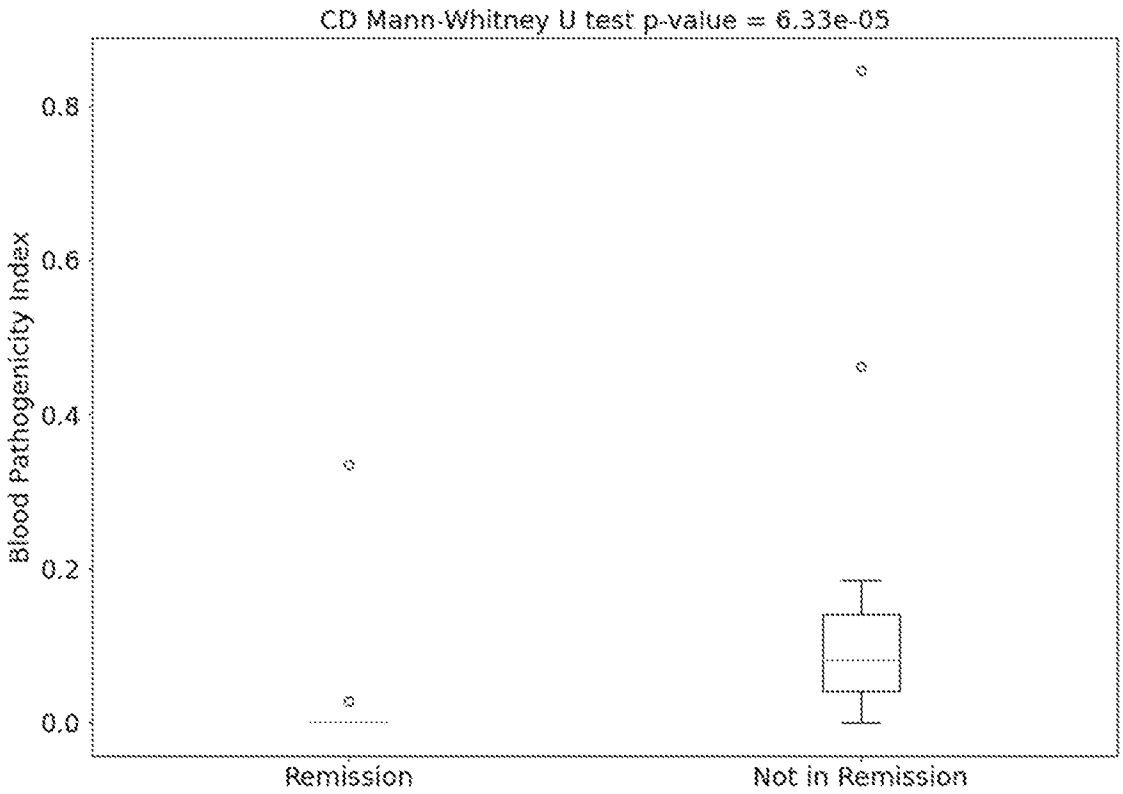
CD Mann-Whitney U test p-value = 6.33e-05
Blood Pathogenicity Index
0.0
0.2
0.4
0.6
0.8
Remission
Not in Remission

FIG. 14B

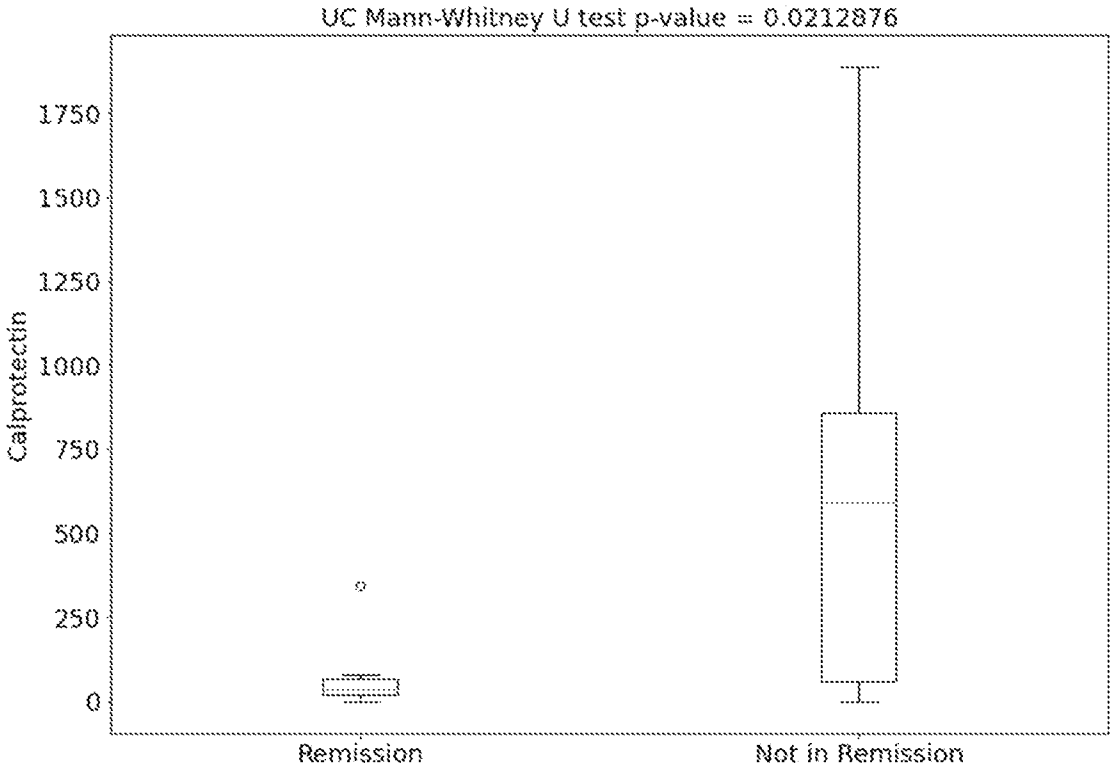
UC Mann-Whitney U test p-value = 0.0212876
Calprotectin
1750
1500
1250
1000
750
500
250
0
Remission
Not in Remission

FIG. 15A

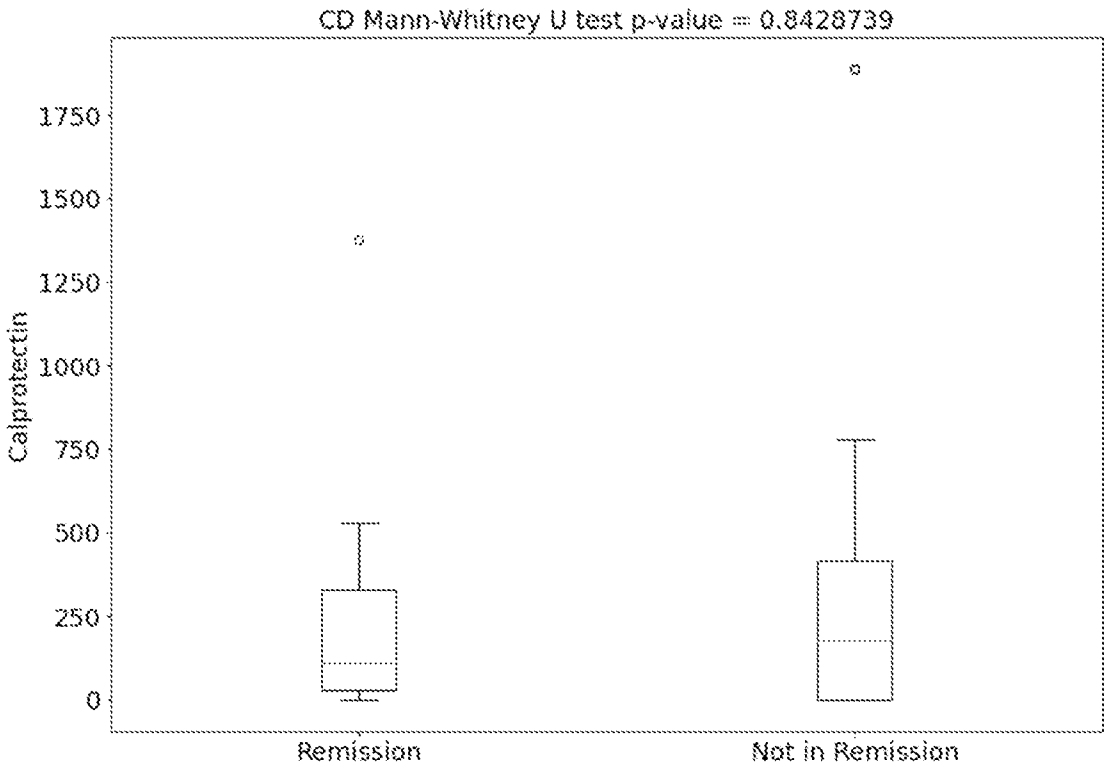
CD Mann-Whitney U test p-value = 0.8428739
Calprotectin
1750
1500
1250
1000
750
500
250
0
Remission
Not in Remission

FIG. 15B

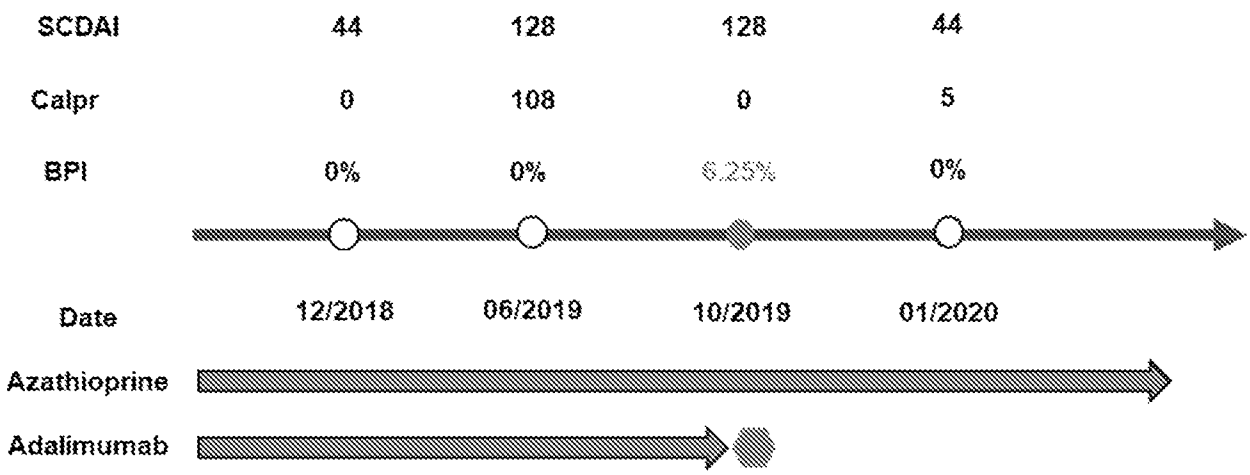
SCDAI
44
128
128
44
Calpr
0
108
0
5
BPI
0%
0%
6.25%
0%
Date
12/2018
06/2019
10/2019
01/2020
Azathioprine
Adalimumab

FIG. 16

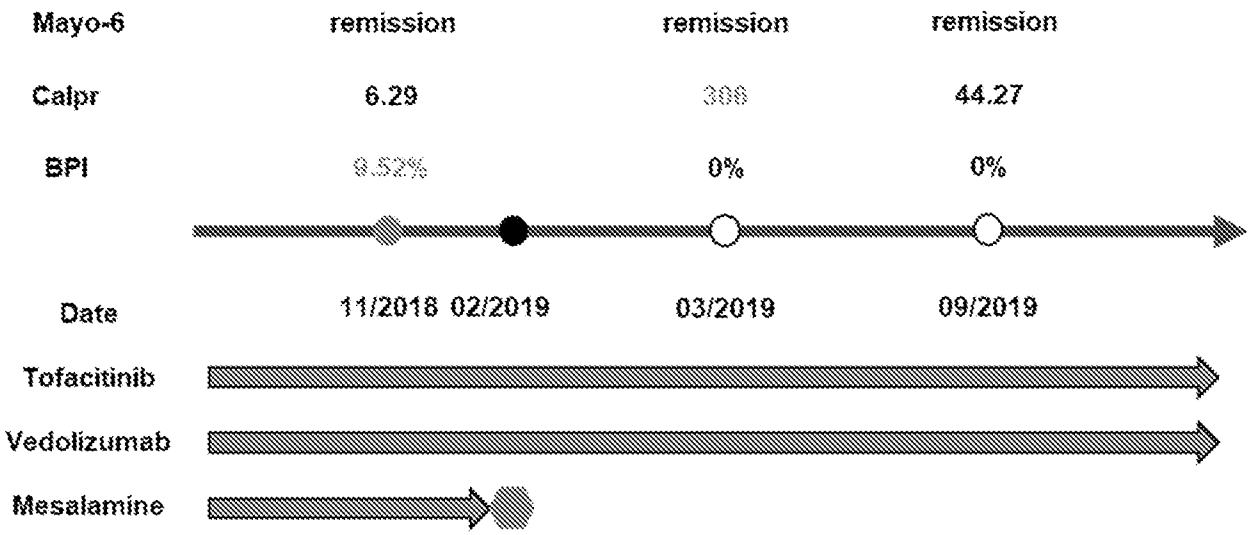
Mayo-6
remission
remission
remission
Calpr
6.29
44.27
BPI
0%
0%
Date
11/2018 02/2019
03/2019
09/2019
Tofacitinib
Vedolizumab
Mesalamine

FIG. 17

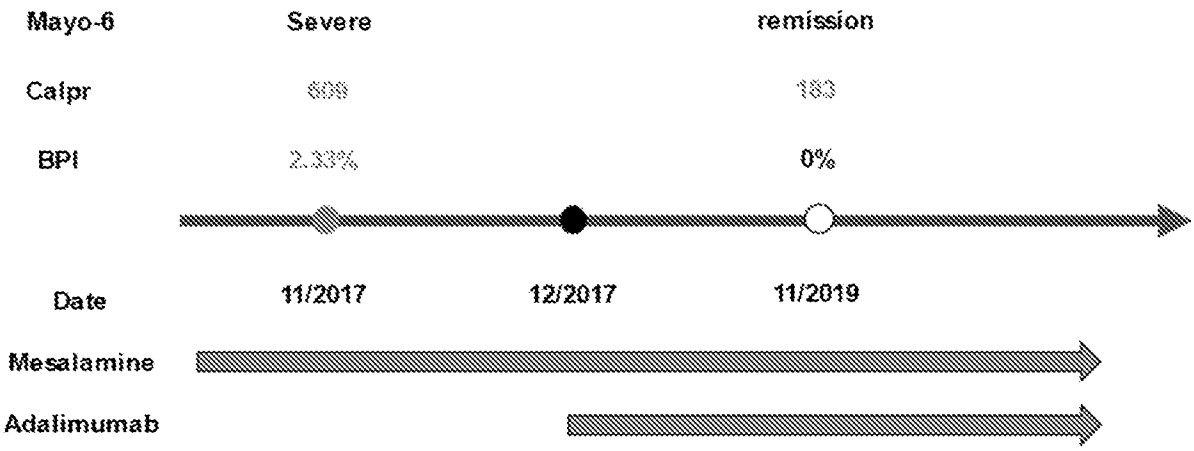
Mayo-6
Severe
remission
Calpr
608
183
BPI
2.33%
0%
Date
11/2017
12/2017
11/2019
Mesalamine
Adalimumab

Fig. 18

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/036971

A. CLASSIFICATION OF SUBJECT MATTER

IPC: ***C12Q 1/689*** (2024.01); ***C12Q 1/6853*** (2024.01); ***C12Q 1/6855*** (2024.01); ***C12Q 1/686*** (2024.01)

CPC: ***C12Q 1/689; C12Q 1/6853; C12Q 1/6855; C12Q 1/686;*** *C12Q 2531/113; C12Q 2600/156; C12Q 2600/158; C12Q 2600/16*

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2021/072439 A1 (LIFE TECHNOLOGIES CORPORATION) 15 April 2021 (15.04.2021) entire document	14, 20-24, 26-36, 38-60, 62, 90-98, 104, 105
Y	US 2021/0310085 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 07 October 2021 (07.10.2021) entire document	14, 20-24, 26-36, 38-60, 62, 90-98, 104, 105
Y	US 2022/0010376 A1 (UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION) 13 January 2022 (13.01.2022) entire document	26-28
Y	US 2016/0244817 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 25 August 2016 (25.08.2016) entire document	30

☑ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 October 2024 (11.10.2024)	**02 December 2024 (02.12.2024)**

Name and mailing address of the ISA/US	Authorized officer
COMMISSIONER FOR PATENTS **MAIL STOP PCT, ATTN: ISA/US** **P.O. Box 1450** **Alexandria, VA 22313-1450** **UNITED STATES OF AMERICA**	**TAINA MATOS**
Facsimile No. **571-273-8300**	Telephone No. **571-272-4300**

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2019/224548 A2 (LUCITE INTERNATIONAL UK LIMITED) 28 November 2019 (28.11.2019) entire document	32, 34, 35
Y	US 2007/0072206 A1 (HILL et al.) 29 March 2007 (29.03.2007) entire document	41
Y	WO 2020/055887 A1 (T2 BIOSYSTEMS INC.) 19 March 2020 (19.03.2020) entire document	42-44
Y	WO 2021/110833 A1 (CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS et al.) 10 June 2021 (10.06.2021) entire document	45, 46
Y	US 2022/0195496 A1 (KARIUS INC.) 23 June 2022 (23.06.2022) entire document	48, 49
Y	US 2020/0131582 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA et al.) 30 April 2020 (30.04.2020) entire document	94-96
P, X	WO 2023/141347 A2 (GUSTO GLOBAL LLC) 27 July 2023 (27.07.2023) entire document	1-14, 18-60, 62, 67-98, 103-105

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/036971

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. ☐ forming part of the international application as filed.
   b. ☑ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)),
      ☑ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/036971

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☑ Claims Nos.: **15-17, 61, 63, 99-102, 106-109**
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I: claims 1-14, 18-60, 62, 67-98, and 103-106 are drawn to methods of amplifying microbial cell free DNA (mcfDNA), methods of diagnosing or monitoring a disease or disorder, and kits.

Group II: claims 64-66 are drawn to sets of degenerate oligonucleotides.

The inventions listed in Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1, because under PCT Rule 13.2 they lack the same or corresponding special technical features for the following reasons:

The special technical features of Group I, methods of amplifying microbial cell free DNA (mcfDNA), methods of diagnosing or monitoring a disease or disorder, and kits, are not present in Group II; and the special technical features of Group II, sets of degenerate oligonucleotides, are not present in Group I.

Additionally, even if Groups I and II were considered to share the technical features of degenerate oligonucleotides, these shared technical features do not represent a contribution over the prior art.

Specifically, WO 2021/072439 A1 to Life Technologies Corporation teaches degenerate oligonucleotides (a mixture of primers can be degenerate primers. Degenerate primers are primers having similar sequences but that differ at one or more nucleotide positions such that one primer may have an A at the position, another may have a G at the same position, another may have a T at the same position and a fourth primer may have a C at the same position, Para. [0050]).

The inventions listed in Groups I and II therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/036971

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☑ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: **1-14, 18-60, 62, 67-98, 103-105**

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.