

Furthermore, both approaches are not able to discriminate between commensal microorganisms, contamination, and the true infectious agent that the patient is suffering from. This ultimately leads to many false positive results.

Traditional blood culture testing takes between two to seven days. During this time, before the causative microorganism is known, patients are treated using broad spectra antibiotics, as adjudged by the treating physician following the most recent treatment guidelines. Due to this, microorganisms can become multi-resistant due to the indiscriminate overuse of broad spectra antibiotics, in view of the inferior diagnostic procedures. Thus, in order to provide fast and effective treatment of patients using appropriate anti-infective agents, it is necessary that the infectious agent be identified as fast as possible, and it is of the utmost importance to be able to discriminate between the infectious agent and commensal microorganisms/contamination during the diagnostic procedure.

There are examples in the literature of the sequencing of samples obtained from patients to identify microorganisms contained therein, such as Hasman *et al.*, 2014, Journal of Clinical Microbiology 52:139-146, describing whole genome sequencing on urine samples to identify microorganisms contained therein, which sequence results were compared to the results obtained with conventional culturing and identification. Others include Grumaz *et al.*, 2016, Genome Medicine 8:73, disclosing next-generation sequencing of samples obtained from septic patients; Andersson *et al.*, 2013, Clin Microbiol Infect 19:E405-E408, describing ultra-deep sequencing of DNA derived from a vaginal swab diagnostic specimen; and Turnbaugh *et al.*, 2009, Nature 457:480-484, describing shotgun sequencing of total fecal DNA to identify genes commonly enriched in the obese or lean gut microbiome. These methods simply sequence and compare non-host nucleic acids to databases for identifying any microorganisms in the sample.

However, there remains a need in the art for more efficient processing of the sequence data such that more accurate results are provided and/or allows for the faster identification of the disease causing microorganism such that effective treatment can be started earlier.

SUMMARY

In one aspect, the present disclosure refers to a method for determining the presence of microorganisms in a subject comprising: (a) sequencing nucleic acids present in a biological

sample obtained from the subject to obtain a plurality of nucleic acid sequence reads; (b) comparing sequence reads obtained in step (a) with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species comprised within the one or more databases; (c) determining over time the number of compared sequence reads mapping to a particular microorganism and the number of compared sequence reads mapping to a species; and (d) computing a significance score over time for the probability of finding in the subject a compared sequence read mapping to the particular microorganism based on the number of compared sequence reads mapping to the particular microorganism and the number of compared sequence reads mapping to a species, wherein when the significance score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be present in the subject. In another aspect, the present disclosure refers to a method for determining the presence of microorganisms in a subject comprising: a step of computing over time a significance score for the probability of finding in the subject a sequence read mapping to a particular microorganism based on the number of sequence reads mapping to the particular microorganism and the number of sequence reads mapping to a species, wherein when the significance score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be present in the subject, wherein the sequence reads mapping to the particular microorganism and the sequence reads mapping to a species are obtained by comparing sequence reads with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species comprised within the one or more databases, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject. In yet another aspect, the present disclosure refers to a method for determining the presence of microorganisms in a subject comprising: (a) step of determining over time the number of sequence reads mapping to a particular microorganism and the number of sequence reads mapping to a species, wherein the sequence reads are obtained by comparing sequence reads with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a sequence read maps to a species comprised within the one or more databases, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject; and (b) computing a significance score over time

for the probability of finding in the subject a sequence read mapping to the particular microorganism based on the number of sequence reads mapping to the particular microorganism and the number of sequence reads mapping to a species, wherein when the significance score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be present in the subject. In a further aspect, the present disclosure refers to a method for determining the presence of a disease state in a subject comprising: a step of computing over time a significance score for the probability of finding in the subject a sequence read not mapping to the control subject based on the number of sequence reads not mapping to the control subject and the number of sequence reads mapping to the control subject, wherein when the significance score meets or exceeds a threshold value, the disease state is determined to be present in the subject, wherein the sequence reads mapping to the control subject and the sequence reads not mapping to the control subject are obtained by comparing sequence reads with one or more databases comprising the genetic information from a control subject of the same species to determine whether or not a compared sequence read maps to the control subject, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject. In yet another aspect, the present disclosure refers to a method for diagnosing an infectious disease caused by microorganisms in a subject comprising: (a) sequencing nucleic acids present in a biological sample obtained from the subject to obtain a plurality of nucleic acid sequence reads; (b) comparing sequence reads obtained in step (a) with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species comprised within the one or more databases; (c) determining over time the number of compared sequence reads mapping to a particular microorganism and the number of compared sequence reads mapping to a species; and (d) computing a significance score over time for the probability of finding in the subject a compared sequence read mapping to the particular microorganism based on the number of compared sequence reads mapping to the particular microorganism and the number of compared sequence reads mapping to a species, wherein when the significance score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be causing the infectious disease. In one aspect, the present disclosure refers to a computer-readable storage medium storing program code comprising instructions which when executed by a processor carry out the method disclosed herein. In another aspect, the

present disclosure refers to a computer system comprising a processor configured to carry out the method disclosed herein.

The present disclosure is based, at least in part, on the inventors' discovery that the likelihood that a subject has a disease state can be determined in view of the amount of a nucleic acid present in a biological sample obtained from a subject, but which is not normally present in a

healthy subject. For example, by determining the amount of a nucleic acid mapping to microorganism(s) in a biological sample obtained from a subject, the likelihood that the subject is suffering from a disease state, such as an infection, caused by the microorganism(s) can be determined. Also, this discovery allows for the determination of the likelihood that the subject has cancer, and is particularly useful in monitoring cancer treatment. This likelihood is, in one example, determined by computing a significance score for the probability of finding in the subject a nucleic acid sequence mapping to a particular microorganism based on the total number of sequence reads mapped (assigned) to the particular microorganism and the total number of all sequence reads that can be mapped (assigned) to a species, including the number of reads mapped to the same species as the subject and the number of reads mapped to any microorganisms in the sample. This significance score, based essentially on the ratio of the number of sequence reads mapped to the particular microorganism and the total number of sequence reads mapped to a species present in a biological sample obtained from a subject, can be computed over time, *i.e.*, computed in real time, as the total number of mapped reads increases (as more and more sequence reads are obtained and mapped to a species).

In an example, the present disclosure is directed to a method for determining the presence of microorganisms in a subject which in one example comprises determining the number of sequence reads mapping to the genome of a particular microorganism and the number of sequence reads mapping to the genome of a species, including the same species as the subject. The sequence reads, obtained from sequencing nucleic acids present in a biological sample obtained from the subject, can be compared to one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms. It follows that the number of sequence reads mapping to a species, including the host species and any microorganisms, and the number of sequence reads mapping to a particular microorganism can be determined. In an example, the method further comprises calculating a significance score for a particular microorganism, which significance score is based on the number of sequence reads mapping to that particular microorganism and the total number of reads mapping to a species. Since the determining step can be carried out over time, this significance score calculation also can be performed over time as the sequence reads are being obtained and mapped. Also, this calculation can be performed over time as the sequence reads are compared to the genetic information in the one or more databases in examples where the sequence reads already have been obtained but not yet have been compared and mapped to a species.

The present disclosure is directed to a method for determining the presence of microorganisms in a subject comprising the steps of (a) sequencing nucleic acids present in a biological sample obtained from the subject to obtain a plurality of nucleic acid sequence reads; (b) comparing sequence reads obtained in step (a) with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species; and (c) determining over time the number of compared sequence reads mapping to a particular microorganism and the number of compared sequence reads mapping to a species.

The present disclosure is also directed to a method for determining the presence of microorganisms in a subject comprising (a) comparing sequence reads with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species, wherein the sequence reads are obtained by sequencing nucleic acids present in a biological sample obtained from the subject; and (b) determining over time the number of compared sequence reads mapping to a particular microorganism and the number of compared sequence reads mapping to a species.

The present disclosure is also directed to a method for determining the presence of microorganisms in a subject comprising a step of determining over time the number of compared sequence reads mapping to a particular microorganism and the number of compared sequence reads mapping to a species, wherein the compared sequence reads are obtained by comparing generated sequence reads with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject.

In an example of the disclosure, the method further comprises computing a significance score for the probability of finding in the subject a compared sequence read mapping to the particular microorganism based on the number of compared sequence reads mapping to the particular microorganism and the number of compared sequence reads mapping to a species.

The present disclosure is also directed to a method for determining the presence of microorganisms in a subject comprising a step of computing over time a significance score for the probability of finding in the subject a sequence read mapping to a particular microorganism based on the number of sequence reads mapping to the particular microorganism and the number of sequence reads mapping to a species, wherein the sequence reads mapping to the particular microorganism and the sequence reads mapping to a species are obtained by comparing sequence reads with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species comprised within the one or more databases, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject.

The present disclosure is also directed to a method for determining the presence of microorganisms in a subject comprising (a) step of determining over time the number of sequence reads mapping to a particular microorganism and the number of sequence reads mapping to a species, wherein the sequence reads are obtained by comparing sequence reads with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a sequence read maps to a species comprised within the one or more databases, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject; and (b) computing a significance score for the probability of finding in the subject a sequence read mapping to the particular microorganism based on the number of sequence reads mapping to the particular microorganism and the number of sequence reads mapping to a species.

In multiple examples of the present disclosure, the method can be performed wherein the sequencing of the nucleic acids is immediately followed by, *i.e.*, essentially concurrently with, comparing the sequence reads so as to map the reads to a species and calculate a significance score, or the sequencing can take place at any time earlier than the comparing/determining/calculating steps such that the results of the sequencing are stored, and the stored sequencing results then can be used to compare the sequenced reads with the one or more databases, and, *e.g.*, allows for the calculation of the significance score.

In an example, the step of determining over time the number of compared sequence reads mapping to a particular microorganism and the number of compared sequence reads mapping to a species means that the number of compared reads that can be mapped to a particular microorganism is counted and the number of compared reads that can be mapped to a species, *i.e.*, reads not only mapping to the particular microorganism but also reads mapping to the subject, as well as mapping to any other microorganism present in the sample, is counted. Those sequence reads not able to be mapped to a species, possibly due to degradation, too short a length or are from a microorganism not present in the one or more databases, are not used in the present disclosure. Preferably, not all sequence reads are used in the present disclosure, only those that are able to be mapped to a species.

In one example, when the significance score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be present in the subject or when the significance score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be relevant for causing a disease in the subject. In other examples, the more the significance score exceeds the threshold value, the higher the load of the microorganism is in the subject, which can reflect a more severe state of infection. In an example, the threshold value is set in order to minimize the number of false positives and false negatives with regard to the relevance of the particular microorganism for causing a disease in the subject.

In another example, when the significance score for the particular microorganism exceeds a threshold value with few sequence reads mapping to a species, the disease due to the presence of the microorganism can be considered to be severe. In the context of this example, “few” refers to the fact that not all, *i.e.*, a portion of, sequenced reads generated by sequencing the nucleic acids in the sample have been compared and mapped but where the threshold value has already been met or exceeded. The portion of reads compared and mapped can be 1%, 2%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or 95% of all of the compared and mapped reads. Preferably, “few” refers to less than 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, 5%, 2%, 1% of all of the compared and mapped reads. “Few” can also be a fixed number of reads, such as less than 100, 1,000, 10,000, or 100,000 reads.

In an example, the method of the present disclosure can be performed over time until a point at which the information provided, such as the number of reads mapping to the subject/particular microorganism or another parameter, including that described below, allows for the determination with a level of certainty that a subject has or does not have a disease state or is or is not infected with one or more microorganisms, as well as the identity of the one or more microorganisms or of the type of cancer. Once this point has been reached, the method can be stopped since no additional information necessarily needs to be provided in order to determine the presence of microorganisms or a disease state in the subject.

The number of reads mapping to a particular microorganism and the number of reads mapping to a species over time can be used to generate a parameter over time which is useful not only in determining, *e.g.*, whether a particular microorganism is relevant for a disease state in a subject, but also allows for the comparison of disease states (of the same cause) between two or more patients. In other words, where there is the same number of reads mapped to a species between two patients but there is a different (more or less) number of reads mapped to a particular microorganism, this difference can indicate a difference in burden/amount of the particular microorganism between the two patients. For example, if a subject has 1 read of a particular microorganism in 10^6 reads mapped to a species and a second subject has 1 read for the same particular microorganism in 5×10^5 reads mapped to a species, it can be concluded that the microorganism is not only present in the second subject but that the second subject has a higher load/level of infection.

Moreover, this parameter can be generated in real time at any time (over time) during the method, not just at the end point where all sequence reads have been compared and where all compared reads have been mapped. Thus, if one subject is seen to have 5 times the number of reads mapped to a particular microorganism relative to the same number of reads mapped to a species as would be seen in a control sample at a time point in which only a fraction of the total reads have been compared and mapped, the method can be stopped at that earlier time point prior to comparing and mapping all sequenced reads since it is clear that the patient with 5 times more reads is likely to have a disease state (infection) due to the particular microorganism.

The ability to generate this parameter over time during the sequencing, comparing and mapping steps, such that the method can be stopped prior to the end of the analysis, *i.e.*, in which all nucleic acids in the sample have been sequenced and all reads have been compared and mapped, advantageously allows for savings in time and resources compared to methodologies that

cannot be so stopped. For example, normally the sequencing, comparing and mapping steps of all of the nucleic acids in a sample can take up to 30 hours or more. However, the present disclosure allows for significantly decreasing this time, for example, in some cases by 10 hours or more, such that 10 hours of sequencing and/or computer time can be saved. Moreover, since the subject can be diagnosed more quickly, appropriate treatment can be started more quickly resulting in a higher likelihood of survival for the subject. This also allows for not wasting pharmaceuticals which are not properly targeted to treating the infection or disease state, *e.g.*, giving an antibiotic for a viral infection or giving an antibiotic for which the microorganism is resistant.

The present disclosure is also directed to a method for determining the presence of a disease state in a subject comprising (a) sequencing nucleic acids present in a biological sample obtained from the subject to obtain a plurality of nucleic acid sequence reads; (b) comparing sequence reads obtained in step (a) with one or more databases comprising the genetic information from a control subject of the same species to determine whether or not a compared sequence read maps to the control subject; and (c) determining over time the number of compared sequence reads mapping and not mapping to the control subject. The present disclosure is also directed to a method for determining the presence of a disease state in a subject comprising (a) comparing sequence reads with one or more databases comprising the genetic information from a control subject of the same species to determine whether or not a compared sequence read maps to the control subject, wherein the sequence reads are obtained by sequencing nucleic acids present in a biological sample obtained from the subject; and (b) determining over time the number of compared sequence reads mapping and not mapping to the control subject. The present disclosure is also directed to a method for determining the presence of a disease state in a subject comprising a step of determining over time the number of compared sequence reads mapping and not mapping to a control subject, wherein the compared sequence reads are obtained by comparing generated sequence reads with one or more databases comprising the genetic information from a control subject of the same species to determine whether or not a compared sequence read maps to the control subject, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject.

In one example, the method further comprises computing a significance score for the probability of finding in the subject a compared sequence read not mapping to the control

subject based on the number of compared sequence reads not mapping to the control subject and the number of compared sequence reads able to be mapped, *e.g.*, mapping to the control subject.

The present disclosure is also directed to a method for determining the presence of a disease state in a subject comprising a step of computing over time a significance score for the probability of finding in the subject a sequence read not mapping to the control subject based on the number of sequence reads not mapping to the control subject and the number of sequence reads mapping to the control subject, wherein the sequence reads mapping to the control subject and the sequence reads not mapping to the control subject are obtained by comparing sequence reads with one or more databases comprising the genetic information from a control subject of the same species to determine whether or not a compared sequence read maps to the control subject, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject.

In an example, when the significance score meets or exceeds a threshold value, the disease state is determined to be present in the subject. As used herein, the term “compared sequence read not mapping to the control subject”, does not always mean that the sequence is not highly similar or is not practically the same as the sequence of the control subject, it often can be. For example, in an example where the disease state is one which is caused by a point mutation in the nucleic acid sequence of the subject, a sequence read having such a point mutation is considered not to map to the control subject, even if all the other nucleotides of the read are identical with the control subject. Moreover, in an example, when comparing the sequence reads, reference can be made to known genomic polymorphisms, *e.g.*, single nucleotide polymorphisms, such that these differences are not considered mutations in the sequenced reads of the subject.

In an example of the disclosure, the disease state is cancer, preferably cancer caused by a genetic abnormality, *e.g.*, a point mutation, a deletion, an insertion or an indel. In another example, the disease state is an infection caused by a microorganism, preferably wherein the microorganism is a virus, a bacterium, a fungus or a parasite.

In an example where the disease state is cancer, the methods of the disclosure can also be used to monitor the treatment of the cancer as well as monitor for the reappearance of the

cancer after a round of treatment. For example, a subject who has been diagnosed with cancer is subject to a treatment, such as surgical removal of the tumor. A database of the genetic information of the tumor can be made and nucleic acids obtained from the subject can be sequenced and reads can be compared to one or more databases containing the genetic information of a control subject from the same species and the genetic information of the tumor. The compared reads are then mapped to the control subject or to the tumor database, such that a significance score according to the disclosure is calculated based on the number of reads mapped to the cancer genome and the number of reads mapped to the control genome, thus allowing for the determination of the presence of the cancer, *i.e.*, the reappearance of the cancer. Similarly, samples can be obtained during treatment and the score calculated to determine if the treatment is having an effect.

In an example where the disease state is an infection by a microorganism, the method of the disclosure also can be used to monitor the treatment of the infection and/or monitor for the reappearance of the infection. In such examples, biological samples are obtained from the subject during and/or after treatment and the method as described above is followed such that a significance score is calculated based on the number of reads mapping to a microorganism and the number of reads mapping to a species.

In certain examples, the biological sample can be selected from the group consisting of whole blood, serum, blood plasma, amniotic fluid, synovial fluid, liquor, tissue or cell smear, tissue or cell swab, urine, tissue, sputum, stool, gastrointestinal secretions, lymph fluid, and lavage.

In certain examples, the subject is a vertebrate, preferably a mammal, for example, human, dog, cat, pig, horse, cattle, sheep, goat, mouse, or rat, preferably the subject is human.

In an example, the sequencing is carried out using ultra-deep or high-throughput sequencing methods. In preferred examples of the disclosure, the sequencing is performed by molecular high-throughput sequence analysis, *i.e.*, by next-generation or third generation sequencing, such as by the Illumina/Solexa or the Oxford Nanopore methodology.

In an example of the present disclosure, when the particular microorganism or the disease state is determined to be present in the subject, the method further comprises administering to the subject a pharmaceutically-active compound known to treat the disease caused by the

particular microorganism or the disease state. Moreover, once the microorganism causing the infectious disease has been identified, it can be determined whether or not it is resistant to any type of antibiotics/anti-infectives, such that the treatment will be effective. In an example, the nucleic acids of the subject in the sample can be depleted prior to determining whether or not the microorganism is resistant to any type of antibiotics/anti-infectives.

In one particular example, the method for diagnosing an infectious disease caused by microorganism(s) in a subject comprises computing over time a significance score for the probability of finding in the subject a sequence read mapping to a particular microorganism based on the number of sequence reads mapping to the particular microorganism and the number of compared sequence reads mapping to a species, wherein when the score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be causing the infectious disease, and wherein the sequence reads mapping to the particular microorganism and the sequence reads mapping to a species are obtained by comparing sequence reads with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species comprised within the one or more databases, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject.

In one particular example, the method for diagnosing an infectious disease caused by microorganism(s) in a subject comprises (a) sequencing nucleic acids present in a biological sample obtained from the subject to obtain a plurality of nucleic acid sequence reads; (b) comparing sequence reads obtained in step (a) with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species comprised within the one or more databases; (c) determining over time the number of compared sequence reads mapping to a particular microorganism and the number of compared sequence reads mapping to a species; and (d) computing a significance score for the probability of finding in the subject a compared sequence read mapping to the particular microorganism based on the number of compared sequence reads mapping to the particular microorganism and the number of compared sequence reads mapping to a species, wherein when the score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be causing the infectious disease.

The present disclosure is directed to a method of treating a disease or infection caused by a microorganism in a subject comprising (a) determining the significance score for a particular microorganism in a subject in accordance with any of the foregoing methods for determining the presence of microorganisms in a subject, and (b) when the significance for the particular microorganism meets or exceeds a threshold value, administering to the subject a compound that inhibits the growth of the particular microorganism. The present disclosure is also directed to a method of treating a disease or infection caused by a microorganism in a subject comprising administering to the subject a compound that inhibits the growth of a microorganism whose significance score meets or exceeds a threshold value, wherein the significance score is calculated according to any of the foregoing methods for determining the presence of microorganisms in a subject described herein.

The present disclosure also encompasses a computer-readable storage medium storing program code comprising instructions which when executed by a processor carry out the methods of the disclosure, as well as a computer system comprising a processor, *e.g.*, a field-programmable gate array, configured to carry out the methods of the disclosure.

DETAILED DESCRIPTION OF THE INVENTION

Although the present invention is described in detail below, it is to be understood that this invention is not limited to the particular methodologies, protocols and reagents described herein as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art.

In the following, the elements of the present invention will be described. These elements are listed with specific embodiments, however, it should be understood that they may be combined in any manner and in any number to create additional embodiments. The variously described examples and preferred embodiments should not be construed to limit the present invention to

What is claimed:

1. A method for determining the presence of microorganisms in a subject comprising:

(a) sequencing nucleic acids present in a biological sample obtained from the subject to obtain a plurality of nucleic acid sequence reads;

(b) comparing sequence reads obtained in step (a) with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species comprised within the one or more databases;

(c) determining over time the number of compared sequence reads mapping to a particular microorganism and the number of compared sequence reads mapping to a species; and

(d) computing a significance score over time for the probability of finding in the subject a compared sequence read mapping to the particular microorganism based on the number of compared sequence reads mapping to the particular microorganism and the number of compared sequence reads mapping to a species, wherein when the significance score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be present in the subject.

2. A method for determining the presence of microorganisms in a subject comprising:

a step of computing over time a significance score for the probability of finding in the subject a sequence read mapping to a particular microorganism based on the number of sequence reads mapping to the particular microorganism and the number of sequence reads mapping to a species, wherein when the significance score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be present in the subject,

wherein the sequence reads mapping to the particular microorganism and the sequence reads mapping to a species are obtained by comparing sequence reads with one or more

databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species comprised within the one or more databases, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject.

3. A method for determining the presence of microorganisms in a subject comprising:

(a) step of determining over time the number of sequence reads mapping to a particular microorganism and the number of sequence reads mapping to a species, wherein the sequence reads are obtained by comparing sequence reads with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a sequence read maps to a species comprised within the one or more databases, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject; and

(b) computing a significance score over time for the probability of finding in the subject a sequence read mapping to the particular microorganism based on the number of sequence reads mapping to the particular microorganism and the number of sequence reads mapping to a species, wherein when the significance score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be present in the subject.

4. The method according to any one of claims 1 to 3, wherein when the significance score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be relevant for causing a disease in the subject.

5. The method according to claim 4, wherein when the significance score for the particular microorganism exceeds a threshold value with few sequence reads, the disease due to the presence of the microorganism is considered to be severe.

6. A method for determining the presence of a disease state in a subject comprising:

a step of computing over time a significance score for the probability of finding in the subject a sequence read not mapping to the control subject based on the number of sequence reads not mapping to the control subject and the number of sequence reads mapping to the control subject, wherein when the significance score meets or exceeds a threshold value, the disease state is determined to be present in the subject,

wherein the sequence reads mapping to the control subject and the sequence reads not mapping to the control subject are obtained by comparing sequence reads with one or more databases comprising the genetic information from a control subject of the same species to determine whether or not a compared sequence read maps to the control subject, and wherein the sequence reads are generated by sequencing nucleic acids present in a biological sample obtained from the subject.

7. The method according to claim 6, wherein the disease state is cancer.

8. The method according to claim 6, wherein the disease state is an infection caused by a microorganism.

9. A method for diagnosing an infectious disease caused by microorganisms in a subject comprising:

(a) sequencing nucleic acids present in a biological sample obtained from the subject to obtain a plurality of nucleic acid sequence reads;

(b) comparing sequence reads obtained in step (a) with one or more databases comprising the genetic information from a control subject of the same species and the genetic information from a plurality of microorganisms to determine whether or not a compared sequence read maps to a species comprised within the one or more databases;

(c) determining over time the number of compared sequence reads mapping to a particular microorganism and the number of compared sequence reads mapping to a species; and

(d) computing a significance score over time for the probability of finding in the subject a compared sequence read mapping to the particular microorganism based on the number of compared sequence reads mapping to the particular microorganism and the number of compared sequence reads mapping to a species,

wherein when the significance score for the particular microorganism meets or exceeds a threshold value, the particular microorganism is determined to be causing the infectious disease.

10. A computer-readable storage medium storing program code comprising instructions which when executed by a processor carry out the method according to any one of claims 1 to 9.

11. A computer system comprising a processor configured to carry out the method according to any one of claims 1 to 9.