[0005] The immune system mounts a response to various conditions, such as cancer, bacterial infections, viral infections, and fungal infections. Further, in some subjects it may actually produce a deleterious response to the tissues of the body, resulting in, for example, autoimmune disease or rejection of grafts. The types and degrees of these immune responses could, if accurately accessed, potentially be one of the most important and accurate indicators of the presence or absence of a particular disease or of undesirable immune responses.

[0006] Humans, however, are estimated to have as many as 10^{15}-10^{25} different T cells, due to the number of possible VDJ rearrangements, n-additions, and alpha-beta chain combinations. Assuming that, for a particular disease, there are only 10^3 disease-specific CDR3s that may quantitatively (through up or down regulation) or qualitatively (through gain or loss) change, the signal-to-noise ratio in these circumstances is very weak. Therefore, conventional methods are impractical to assess this information for diagnostic purposes.

[0007] What are needed are improved methods for immune response assessment and for the development of diagnostic tests based on such immune response and the resulting immune repertoire.

Summary

[0008] In one aspect the present invention refers to a method for developing a diagnostic test for a particular disease, the method comprising the steps of amplifying and sequencing the CDR3 sequences in a sample collected from each of multiple subjects in a patient group and multiple subjects in a control group, wherein the patient group comprises subjects who have the same disease and the control group comprises subjects who are healthy, and determining the frequency of occurrence of each unique CDR3 sequence present in the samples to identify Public CDR3 sequences that are shared between at least two subjects in each of the patient group and the control group, wherein each subject in the patient group is diagnosed with the same disease and wherein each subject in the control group is healthy and not diagnosed with the disease; ranking the Public CDR3 sequences by order of frequency of occurrence; and identifying Positive Linklets from the Public CDR3 sequences, wherein a disease signature is present when a minimum number of Significant Linklets are identified in a sample, wherein Significant

Linklets are Positive Linklets that are associated to a statistically significant degree of $p<0.05$ with subjects in the patient group relative to subjects in the control group. In another aspect the present invention refers to a method for developing a diagnostic test for a particular disease, the method comprising the steps of amplifying and sequencing the CDR3 sequences in a blood sample, collected from each of multiple subjects in a patient group and multiple subjects a control group, wherein the patient group comprises subjects who have the same disease and the control group comprises subjects who are categorized as healthy, and determining the frequency of occurrence of each unique CDR3 sequence present in the samples to identify at least about 1000 Public CDR3 sequences that are shared between at least two subjects in each of the patient group and the control group, wherein each subject in the patient group is diagnosed with the same disease and wherein each subject in the control group is healthy and not diagnosed with the disease, wherein the patient group and the control group each comprise at least 100 subjects; ranking the Public CDR3 sequences by order of frequency of occurrence; and identifying at least about 106 Positive Linklets from the Public CDR3 sequences, wherein a disease signature is present when a minimum number of Significant Linklets are identified in a sample, wherein Significant Linklets are Positive Linklets that are associated to a statistically significant degree of $p<0.05$ with subjects in the patient group relative to subjects in the control group. The present disclosure also describes a method for developing a diagnostic test using the immune repertoire, the method comprising the steps of: (a) collecting a sample from each of multiple subjects in a patient group and a control group, wherein the patient group comprises subjects who have the same disease and the control group comprises subjects who are categorized as healthy; (b) amplifying and sequencing the immune repertoire of each subject in each of the two groups to identify each unique CDR3 sequence present in the sample and to determine the frequency of occurrence of each unique CDR3 sequence; (c) identifying CDR3 sequences that are shared between at least two subjects in each of the control group and the patient group; (d) ranking the identified CDR3 sequences by order of frequency of occurrence; (e) identifying Linklets from each group; and (f) identifying the Linklets that are associated to a statistically significant degree with the patient group to provide a disease signature.

[0009] In certain examples, the sample is peripheral blood. In other examples, the sample is tissue. In certain examples, fewer than 1,000 CDR3 sequences are identified as shared between at least two subjects. In other examples, at least 1,000 CDR3 sequences are identified as shared between at least two subjects. In certain examples, at least about 10^6 Linklets are identified from each group. In other examples, fewer than 10^6 Linklets are identified from each group.

Brief Description of the Figures

[0010] The disclosure can be better understood with reference to the following figures.

[0011] Fig. 1 is a cartoon illustrating the basic concept of the disclosed method, where polynucleotide sequence data from immune system cells is processed using software designed to sort and count the sequences, rank them by frequency numbers, generate p values, and other criteria, to produce a diagnostic signature from the Linklets designated as "Significant Linklets."

[0012] Fig. 2 is a table illustrating the ranking of CDR3s by number of clones present in a sample. Typically, a sequencing result from one sample will generate as many as 400,000 CDR3s. Each CDR3 is associated with a read count; due to the semi-quantitative nature of the amplification method (arm-PCR), the read count also reflects the relative abundance of the clone. Software analysis removes the errors and ranks the CDR3s to produce an output file as shown in the table.

[0013] Fig. 3 is a table illustrating how Linklets are detected in the sequences obtained from blood samples. CDR3 sequences were tallied to provide a list of the CDR3 sequences that are present in the highest numbers in a blood sample. Among those CDR3 sequences, Linklets represent pairs of CDR3s that are present within the same sample – at a level that is higher than a designated cutoff level.

[0014] Fig. 4 is a representative list of some of the Linklets identified during a breast cancer study (comparison between Linklets detected in two study groups – a first group of subjects who had been diagnosed with breast cancer and a second group of subjects who were designated as healthy controls).

[0015] Fig. 5 lists Public Linklets detected in the breast cancer study, based on their p values. The top ranked Linklet, for example, is the CDR3 pair 'ASSYSRGEEF' and

CLAIMS

Now, therefore, the following is claimed:

1. A method for developing a diagnostic test for a particular disease, the method comprising the steps of:

amplifying and sequencing the CDR3 sequences in a sample collected from each of multiple subjects in a patient group and multiple subjects in a control group, wherein the patient group comprises subjects who have the same disease and the control group comprises subjects who are healthy, and

determining the frequency of occurrence of each unique CDR3 sequence present in the samples to identify Public CDR3 sequences that are shared between at least two subjects in each of the patient group and the control group, wherein each subject in the patient group is diagnosed with the same disease and wherein each subject in the control group is healthy and not diagnosed with the disease;

ranking the Public CDR3 sequences by order of frequency of occurrence; and

identifying Positive Linklets from the Public CDR3 sequences,

wherein a disease signature is present when a minimum number of Significant Linklets are identified in a sample,

wherein Significant Linklets are Positive Linklets that are associated to a statistically significant degree of $p<0.05$ with subjects in the patient group relative to subjects in the control group.

2. The method of claim 1, wherein the sample is blood.

3. The method of claim 1, wherein the sample is tissue.

4. The method of claim 1, wherein at least about 1,000 CDR3 sequences are identified that are shared between at least two subjects in each of the control group and the patient group.

5. The method of claim 1, wherein at least about 10^6 Positive Linklets from each of the patient group and control group are identified.

6. The method of claim 1, wherein the disease is breast cancer.

7. The method of claim 6, wherein the minimum number of Significant Linklets is at least about 500.

8. A method for developing a diagnostic test for a particular disease, the method comprising the steps of:

amplifying and sequencing the CDR3 sequences in a blood sample, collected from each of multiple subjects in a patient group and multiple subjects a control group, wherein the patient group comprises subjects who have the same disease and the control group comprises subjects who are categorized as healthy, and

determining the frequency of occurrence of each unique CDR3 sequence present in the samples to identify at least about 1000 Public CDR3 sequences that are shared between at least two subjects in each of the patient group and the control group, wherein each subject in the patient group is diagnosed with the same disease and wherein each subject in the control group is healthy and not diagnosed with the disease, wherein the patient group and the control group each comprise at least 100 subjects;

ranking the Public CDR3 sequences by order of frequency of occurrence; and

identifying at least about 10^6 Positive Linklets from the Public CDR3 sequences,

wherein a disease signature is present when a minimum number of Significant Linklets are identified in a sample,

wherein Significant Linklets are Positive Linklets that are associated to a statistically significant degree of $p<0.05$ with subjects in the patient group relative to subjects in the control group.