

Therefore, the direct detection of one or a small group of cancer-associated mutations in plasma or serum may achieve low analytical sensitivity even among patients with cancers known to be harboring the targeted mutations. Furthermore, it has been shown that there is significant intratumoral heterogeneity in terms of mutations even within a single tumor. The mutations can be found in only a subpopulation of the tumor cells. The difference in the mutational profiles between the primary tumor and the metastatic lesions is even bigger. One example of intratumoral and primary-metastasis heterogeneity involves the *KRAS*, *BRAF* and *PIK3CA* genes in patients suffering from colorectal cancers (Baldus et al. Clin Cancer Research 2010. 16:790-9.).

[0005] In a scenario in which a patient has a primary tumor (carrying a *KRAS* mutation but not a *PIK3CA* mutation) and a concealed metastatic lesion (carrying a *PIK3CA* mutation but not a *KRAS* mutation), if one focused on the detection of the *KRAS* mutation in the primary tumor, the concealed metastatic lesion cannot be detected. However, if one included both mutations in the analysis, both the primary tumor and the concealed metastatic lesion can be detected. Hence, the test involving both mutations would have a higher sensitivity in the detection of residual tumor tissues. Such a simple example becomes more complex when one is screening for cancer, and as one has little or no clue of the types of mutations that might occur. p

[0006] It is therefore desirable to provide new techniques to perform a broad screening, detection, or assessment for cancer

SUMMARY

[0007] Examples can observe a frequency of somatic mutations in a biological sample (e.g., plasma or serum) of a subject undergoing screening or monitoring for cancer, when compared with that in the constitutional DNA of the same subject. Random sequencing can be used to determine these frequencies. A parameter can derived from these frequencies and used to determine a classification of a level of cancer. False positives can be filtered out by requiring any variant locus to have at least a specified number of variant sequence reads (tags), thereby providing a more accurate parameter. The relative frequencies for different variant loci can be analyzed to determine a level of heterogeneity of tumors in a patient.

[0007a] In one aspect, the present disclosure refers to a method for detecting tumor-derived mutations in cell-free DNA molecules, the method comprising: obtaining, by a computer

system, first sequence reads for cell-free DNA molecules from a biological sample of a subject, the first sequence reads comprised of first sequences; obtaining, by the computer system, second sequence reads for DNA molecules from a plurality of blood cells of the subject, the second sequence reads comprised of second sequences; and detecting, by the computer system, the tumor-derived mutations in the cell-free DNA molecules by filtering out a portion of the first sequences of the cell-free DNA molecules, that are also present in the second sequences from the plurality of blood cells of the subject.

[007b] In another aspect, the present disclosure refers to a method of detecting tumor-derived single-nucleotide variants (SNVs) in cell-free DNA molecules from a biological sample, the method comprising: (a) receiving first sequence reads of cell-free DNA molecules from a biological sample of a subject; (b) receiving second sequence reads of DNA molecules from a buffy coat sample of the subject; and (c) determining single-nucleotide variants present in the cell-free DNA molecules, but not in the DNA molecules from the buffy coat sample of the subject, as being tumor-derived by comparing the first sequence reads to the second sequence reads to detect a plurality of loci having single-nucleotide variants, wherein at each of the plurality of loci, a number of the first sequence reads having a single-nucleotide sequence variant relative to the second sequence reads is above a cutoff value, the cutoff value being greater than one.

[007c] In another aspect, the present disclosure refers to a method of identifying one or more tumor-derived mutations in a subject, the method comprising: determining a constitutional genome from a plurality of sequence reads obtained from sequencing DNA molecules from a constitutional sample from the subject; identifying one or more potential sequence variants from a cell-free DNA sample obtained from the subject, wherein identifying the one or more potential sequence variants comprises: obtaining a plurality of sequence reads from the cell-free DNA sample; determining genomic positions for the plurality of sequence reads obtained from the cell-free DNA sample; and detecting the one or more potential sequence variants in one or more sequence reads positioned at one or more loci; and identifying the one or more tumor-derived mutations in the subject based on a comparison of the one or more sequence reads of the one or more potential sequence variants to the constitutional genome at the one or more loci, wherein at each of the one or more loci, a number of the one or more sequence reads having a sequence variant relative to the constitutional genome is above a cutoff value, the cutoff value being greater than one.

[0008] In one example, the parameter can be compared with the same parameter derived from a group of subjects without cancer, or with a low risk of cancer. A significant

difference in the parameter obtained from the test subject and that from the group of subjects without cancer, or with a low risk of cancer, can indicate an increased risk that the test subject has cancer or a premalignant condition or would develop cancer in the future. Thus, in one example, plasma DNA analysis can be conducted without prior genomic information of the tumor. Such an example is thus especially useful for the screening of cancer.

[0009] In another example, examples can also be used for monitoring a cancer patient following treatment and to see if there is residual tumor or if the tumor has relapsed. For example, a patient with residual tumor or in whom the tumor has relapsed would have a higher frequency of somatic mutations than one in whom there is no residual tumor or in whom no tumor relapse is observed. The monitoring can involve obtaining samples from a cancer patient at multiple time points following treatment for ascertaining the temporal variations of tumor-associated genetic aberrations in bodily fluids or other samples with cell-free nucleic acids, e.g. plasma or serum.

[0010] According to one example, a method detects cancer or premalignant change in a subject. A constitutional genome of the subject is obtained. One or more sequence tags are received for each of a plurality of DNA fragments in a biological sample of the subject, where the biological sample includes cell-free DNA. Genomic positions are determined for the sequence tags. The sequence tags are compared to the constitutional genome to determine a first number of first loci. At each first loci, a number of the sequence tags having a sequence variant relative to the constitutional genome is above a cutoff value, where the cutoff value is greater than one. A parameter is determined based on a count of sequence tags having a sequence variant at the first loci. The parameter is compared to a threshold value to determine a classification of a level of cancer in the subject.

[0011] According to another example, a method analyzes a heterogeneity of one or more tumors of a subject. A constitutional genome of the subject is obtained. One or more sequence tags are received for each of a plurality of DNA fragments in a biological sample of the subject, where the biological sample includes cell-free DNA. Genomic positions are determined for the sequence tags. The sequence tags are compared to the constitutional genome to determine a first number of first loci. At each first loci, a number of the sequence tags having a sequence variant relative to the constitutional genome is above a cutoff value,

where the cutoff value is greater than one. A measure of heterogeneity of the one or more tumors is calculated based on the respective first numbers of the set of first genomic locations.

5 [0012] According to another example, a method determines a fractional concentration of tumor DNA in a biological sample including cell-free DNA. One or more sequence tags are received for each of a plurality of DNA fragments in the biological sample. Genomic positions are determined for the sequence tags. For each of a plurality of genomic regions, a respective amount of DNA fragments within the genomic region is determined from sequence tags having a genomic position within the genomic region. The respective amount is normalized to obtain a respective density. The respective density is compared to a reference
10 density to identify whether the genomic region exhibits a 1-copy loss or a 1-copy gain. A first density is calculated from respective densities identified as exhibiting a 1-copy loss or from respective densities identified as exhibiting a 1-copy gain. The fractional concentration is calculated by comparing the first density to another density to obtain a differential, wherein the differential is normalized with the reference density.

15 [0013] Other examples are directed to systems and computer readable media associated with methods described herein.

[0014] A better understanding of the nature and advantages of the present invention may be gained with reference to the following detailed description and the accompanying drawings.

20 BRIEF DESCRIPTION OF THE DRAWINGS

[0015] FIG. 1 is a flowchart of a method 100 for detecting cancer or premalignant change in a subject according to embodiments of the present invention.

[0016] FIG. 2 shows a flowchart of a method comparing the sample genome (SG) directly to the constitutional genome (CG) according to embodiments of the present invention.

25 [0017] FIG. 3 shows a flowchart of a method 300 comparing the sample genome (SG) to the constitutional genome (CG) using the reference genome (RG) according to embodiments of the present invention.

[0018] FIG. 4 is a table 400 showing the number of cancer-associated single nucleotide mutations correctly identified using different number of occurrences as the criterion for
30 classifying a mutation as being present in the sample according to embodiments of the

WHAT IS CLAIMED IS:

1. A method for detecting tumor-derived mutations in cell-free DNA molecules, the method comprising:
 - obtaining, by a computer system, first sequence reads for cell-free DNA molecules from a biological sample of a subject, the first sequence reads comprised of first sequences;
 - obtaining, by the computer system, second sequence reads for DNA molecules from a plurality of blood cells of the subject, the second sequence reads comprised of second sequences; and
 - detecting, by the computer system, the tumor-derived mutations in the cell-free DNA molecules by filtering out a portion of the first sequences of the cell-free DNA molecules, that are also present in the second sequences from the plurality of blood cells of the subject.
2. The method of claim 1, wherein the blood cells comprise white blood cells.
3. The method of any one of claims 1 to 2, wherein the biological sample of the subject comprises blood plasma.
4. The method of any one of claims 1 to 2, wherein the biological sample of the subject comprises one or more of a group consisting of urine, ascetic fluid, peritoneal fluid, saliva, cerebrospinal fluid, and a stool sample.
5. The method of any one of claims 1 to 4, further comprising sequencing the cell-free DNA molecules from the biological sample of the subject to obtain the first sequence reads.
6. The method of claim 5, wherein sequencing the cell-free DNA molecules from the biological sample comprises targeted sequencing.
7. The method of claim 6, wherein the targeted sequencing is performed using a solid-phase capture technique.

8. The method of claim 6, wherein the targeted sequencing comprises a solution hybridization technique.
9. The method of claim 6, wherein the targeted sequencing comprises amplification via PCR.
10. The method of claim 6, wherein the targeted sequencing comprises target-enriching a whole exome.
11. The method of claim 6, wherein the targeted sequencing comprises target-enriching exons.
12. The method of claim 6, wherein the targeted sequencing enriches 5' and 3' untranslated regions.
13. The method of claim 6, wherein the targeted sequencing targets specific tumor-associated mutations.
14. The method of claim 5, wherein sequencing the cell-free DNA molecules from the biological sample of the subject comprises paired-end sequencing.
15. The method of claim 5, wherein sequencing the cell-free DNA molecules from the biological sample of the subject comprises random sequencing.
16. The method of claim 5, wherein the sequencing the cell-free DNA molecules from the biological sample of the subject comprises massively parallel sequencing.
17. The method of any one of claims 1 to 16, further comprising sequencing the DNA molecules from the plurality of blood cells of the subject to obtain the second sequence reads.
18. The method of claim 17, wherein sequencing the DNA molecules from the plurality of blood cells comprises paired-end sequencing.
19. The method of claim 18, wherein the sequencing the DNA molecules from the plurality of blood cells comprises random sequencing.

20. The method of claim 17, wherein the sequencing the DNA molecules from the plurality of blood cells comprises massively parallel sequencing.

21. A method of detecting tumor-derived single-nucleotide variants (SNVs) in cell-free DNA molecules from a biological sample, the method comprising:

(a) receiving first sequence reads of cell-free DNA molecules from a biological sample of a subject;

(b) receiving second sequence reads of DNA molecules from a buffy coat sample of the subject; and

(c) determining single-nucleotide variants present in the cell-free DNA molecules, but not in the DNA molecules from the buffy coat sample of the subject, as being tumor-derived by comparing the first sequence reads to the second sequence reads to detect a plurality of loci having single-nucleotide variants, wherein at each of the plurality of loci, a number of the first sequence reads having a single-nucleotide sequence variant relative to the second sequence reads is above a cutoff value, the cutoff value being greater than one.

22. The method of claim 21, further comprising identifying the subject as having a cancer based on the tumor-derived single-nucleotide variants present in the cell-free DNA molecules.

23. The method of any one of claims 21 to 22, further comprising identifying a level of cancer in the subject based, at least in part, on the tumor-derived single-nucleotide variants present in the cell-free DNA molecules.

24. The method of any one of claims 21 to 23, further comprising determining a fractional concentration of tumor DNA based, at least in part, on the tumor-derived single-nucleotide variants present in the cell-free DNA molecules.

25. A method of identifying one or more tumor-derived mutations in a subject, the method comprising:

determining a constitutional genome from a plurality of sequence reads obtained from sequencing DNA molecules from a constitutional sample from the subject;

identifying one or more potential sequence variants from a cell-free DNA sample obtained from the subject, wherein identifying the one or more potential sequence variants comprises:

obtaining a plurality of sequence reads from the cell-free DNA sample;
determining genomic positions for the plurality of sequence reads obtained from the cell-free DNA sample; and

detecting the one or more potential sequence variants in one or more sequence reads positioned at one or more loci; and

identifying the one or more tumor-derived mutations in the subject based on a comparison of the one or more sequence reads of the one or more potential sequence variants to the constitutional genome at the one or more loci, wherein at each of the one or more loci, a number of the one or more sequence reads having a sequence variant relative to the constitutional genome is above a cutoff value, the cutoff value being greater than one.

26. The method of claim 25, wherein the one or more potential sequence variants at the one or more loci are detected in the one or more sequence reads relative to a reference genome.

27. The method of claim 26, wherein the one or more potential sequence variants comprise single-nucleotide variants.

28. The method of any one of claims 25 to 27, wherein the constitutional sample comprises blood cells, white blood cells, buccal cells, hair follicles, or any combination thereof.

29. The method of any one of claims 25 to 28, wherein the constitutional genome comprises one or more germline mutations, de novo variants, and non-cancer derived variants.

30. The method according to claim 29, wherein the one or more non-cancer derived variants are derived from different hematopoietic stem cells.

31. The method of any one of claims 25 to 30, further comprising detecting one or more constitutional variants, wherein identifying the one or more constitutional variants comprises:

determining genomic positions for the plurality of sequence reads obtained from the constitutional sample; and

detecting one or more constitutional variants at one or more loci relative to a reference genome.

32. The method of any one of claims 25 to 31, further comprising identifying the one or more tumor-derived mutations as associated with a non-synonymous change.

33. The method of any one of claims 25 to 32, wherein the constitutional genome is only a part of a genome of the subject.

34. The method of any one of claims 25 to 33, wherein the constitutional genome is part of a human genome.

35. The method of any one of claims 25 to 34, wherein the cell-free DNA sample is a cell-free DNA sample from blood plasma.

36. The method of any one of claims 25 to 35, wherein identifying the one or more tumor-derived mutations in the subject based on the comparison of the one or more potential sequence variants to the constitutional genome comprises subtracting constitutional variants of the constitutional genome from the one or more potential sequence variants.

37. The method of any one of claims 25 to 36, wherein the constitutional sample from the subject contains more than 90% constitutional DNA.

38. The method of any one of claims 25 to 37, wherein determining the constitutional genome from the plurality of sequence reads obtained from sequencing DNA molecules from the constitutional sample of the subject comprises determining a consensus sequence that includes a determination of a homozygous locus or a heterozygous locus having two alleles.