

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.

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

SUMMARY

[0007] ~~Embedments~~Examples of the present disclosure 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.

[0008] In one ~~embedment~~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 ~~embodiment~~example, plasma DNA analysis can be conducted without prior genomic information of the tumor. Such an ~~embodiment~~example is thus especially useful for the screening of cancer.

[0009] In another ~~embodiment~~example, ~~embodiment~~the present disclosure 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.

[0009a] In one aspect there is provided a method for detecting cancer or premalignant change in a subject, the method comprising: obtaining a constitutional genome of the subject; receiving one or more sequence tags for each of a plurality of DNA fragments in a biological sample of the subject, the biological sample including cell-free DNA; determining genomic positions for the sequence tags; comparing the sequence tags to the constitutional genome to determine a first number of first loci, wherein: at each first loci, a number of the sequence tags having a sequence variant relative to the constitutional genome is above a cutoff value, the cutoff value being greater than one; determining a parameter based on a count of sequence tags having a sequence variant at the first loci; and comparing the parameter to a threshold value to determine a classification of a level of cancer in the subject.

[0010] According to one ~~embodiment~~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.

[0010a] In another aspect, there is provided a method for analyzing a heterogeneity of one or more tumors of a subject, the method comprising: obtaining a constitutional genome of the subject;

receiving one or more sequence tags for each of a plurality of DNA fragments in a biological sample of the subject, the biological sample including cell-free DNA; determining genomic positions for the sequence tags; comparing the sequence tags to the constitutional genome to determine a first number of first loci, wherein: at each first loci, a number of the sequence tags having a sequence variant relative to the constitutional genome is above a cutoff value, the cutoff value being greater than one; and calculating a measure of heterogeneity of the one or more tumors based on the respective first numbers of the set of first genomic locations.

[0011] According to another embodiment 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.

[0011a] In yet another aspect, there is provided a method of determining a fractional concentration of tumor DNA in a biological sample including cell-free DNA, the method comprising: receiving one or more sequence tags for each of a plurality of DNA fragments in the biological sample; determining genomic positions for the sequence tags; for each of a plurality of genomic regions: determining a respective amount of DNA fragments within the genomic region from sequence tags having a genomic position within the genomic region; normalizing the respective amount to obtain a respective density; and comparing the respective density to a reference density to identify whether the genomic region exhibits a 1-copy loss or a 1-copy gain; calculating a first density from one or more respective densities identified as exhibiting a 1-copy loss or from one or more respective densities identified as exhibiting a 1-copy gain; and calculating the fractional concentration by: comparing the first density to another density to obtain a differential, wherein the differential is normalized with the reference density.

[0012] According to another embodiment 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 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.

[0012a] In yet another aspect, there is provided a computer program product comprising a computer readable medium storing a plurality of instructions for controlling a processor to perform an operation for methods described herein, the instructions comprising the steps of any one of the claims above.

[0013] Other embodiment 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.

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.

5 [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.

[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.

10 [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 classifying a mutation as being present in the sample according to embodiments of the

obtaining a constitutional sample of the subject that contains more than 90% constitutional DNA;

performing random sequencing of DNA fragments in the constitutional sample to obtain one or more second sequence tags for each of a plurality of DNA fragments in the constitutional sample;

aligning at least a portion of the second sequence tags to a reference genome, wherein the alignment of a second sequence tag allows for a mismatch between the second sequence tag and the constitutional genome at M or less genomic locations, wherein M is an integer equal to or greater than one; and

constructing the constitutional genome based on the second sequence tags and the aligning.

28. The method of claim 27, wherein the constitutional sample is the biological sample, and wherein constructing the constitutional genome includes:

determining a consensus sequence that includes a determination of a homozygous locus or a heterozygous locus having two alleles; and using the consensus sequence at the constitutional genome.

29. The method of claim 1, further comprising:

receiving the biological sample of the subject; and

performing the random sequencing of DNA fragments in the biological sample to generate the one or more sequence tags for each of a plurality of DNA fragments in the biological sample.

30. The method of claim 1, wherein the one or more sequence tags are generated from a random sequencing of DNA fragments in the biological sample.

~~31. The method of claim 1, wherein the biological sample is urine, pleural fluid, ascitic fluid, peritoneal fluid, saliva, cerebrospinal fluid, or a stool sample.~~

~~3231.~~ The method of claim 1, wherein the parameter is a fractional concentration of tumor-derived DNA.

~~33. A computer product comprising a non-transitory computer readable medium storing a plurality of instructions that when executed control a computer system to perform method 1.~~

3432. A method for analyzing a heterogeneity of one or more tumors of a subject, the method comprising:

obtaining a constitutional genome of the subject;
receiving one or more sequence tags for each of a plurality of DNA fragments in a biological sample of the subject, the biological sample including cell-free DNA;
determining genomic positions for the sequence tags;
comparing the sequence tags to the constitutional genome to determine a first number of first loci, wherein:

at each first loci, a number of the sequence tags having a sequence variant relative to the constitutional genome is above a cutoff value, the cutoff value being greater than one; and

calculating a measure of heterogeneity of the one or more tumors based on the respective first numbers of the set of first genomic locations.

3533. The method of claim 3432, further comprising:
comparing the heterogeneity measure to one or more threshold values to determine a classification of a level of heterogeneity.

3634. The method of claim 3533, wherein the one or more threshold values are determined from one or more other subjects whose tumors have been biopsied and analyzed to determine mutations in the biopsied tumors to determine a level of heterogeneity, and wherein a measure of heterogeneity from a biological sample including cell-free DNA of one or more other subjects is used determine a threshold value.

3731. The method of claim 3634, wherein comparing the heterogeneity measure to one or more threshold values includes:

inputting the heterogeneity measure to a calibration function that outputs a heterogeneity level based on the heterogeneity measure.

3836. The method of claim 3432, wherein the one or more measures include a total number of first genomic locations at which more than one DNA fragment aligned to the constitutional DNA while having a sequence variant at that genomic location.

3937. The method of claim 3432, wherein a plurality of heterogeneity measures are calculated, wherein calculating the heterogeneity measures includes:

for each first loci, calculating a proportion of sequence tags having a sequence variation;

creating a histogram of a number of first loci for values of the proportions; and identifying a plurality of peaks in the histogram.

4038. The method of claim 3937, wherein a heterogeneity measure corresponds to the number of identified peaks.

4139. The method of claim 3937, wherein a heterogeneity measure includes a ratio of heights of two peaks.

4240. The method of claim 3937, wherein the proportions each represent a fractional concentration of tumor DNA as measured at a particular first locus.

4341. The method of claim 3432, wherein the heterogeneity measure corresponds to a ratio of a first proportion of first loci having a first specified amount of sequence tags with sequence variations and a second proportion of first loci having a second specified amount of sequence tags with sequence variations.

4442. The method of claim 4341, wherein the first specified amount is less than the second specified amount.

4543. The method of claim 4442, wherein the first specified amount is a first range and the second specified amount is a second range, the first range being below the second range.

4644. The method of claim 4341, wherein the first specified amount and the second specified amount correspond to a fractional concentration or to an absolute number of sequence tags with a sequence variation.

4745. The method of claim 3432, wherein the heterogeneity measure is determined from a histogram of a number of first loci corresponding to each of the respective first numbers.

4846. The method of claim 4745, wherein the set of first identified genomic locations include a first subset and a second subset, wherein the one or more measures includes a first histogram of the respective first numbers corresponding to the first subset and a second histogram of the respective second numbers corresponding to the second subset.

4947. The method of claim 3432, wherein the heterogeneity measure includes a proportion of first loci having corresponding first numbers above a specified value.

5048. A method of determining a fractional concentration of tumor DNA in a biological sample including cell-free DNA, the method comprising:

receiving one or more sequence tags for each of a plurality of DNA fragments in the biological sample;

determining genomic positions for the sequence tags;

for each of a plurality of genomic regions:

determining a respective amount of DNA fragments within the genomic region from sequence tags having a genomic position within the genomic region;

normalizing the respective amount to obtain a respective density; and

comparing the respective density to a reference density to identify whether the genomic region exhibits a 1-copy loss or a 1-copy gain;

calculating a first density from one or more respective densities identified as exhibiting a 1-copy loss or from one or more respective densities identified as exhibiting a 1-copy gain; and

calculating the fractional concentration by:

comparing the first density to another density to obtain a differential, wherein the differential is normalized with the reference density.

5149. The method of claim 5048, further comprising:

creating a histogram of a number of genomic regions having various respective densities;

identifying a plurality of peaks in the histogram; and

calculating a heterogeneity measure from a ratio of the number of genomic regions of one or more first peaks to the number of genomic regions of one or more second peaks.

5 5250. The method of claim 5149, wherein the histogram uses values of fractional concentrations determined separately for each genomic region in creating the histogram.

5351. The method of claim 5149, wherein the first peaks have respective densities of a first specified amount, and the second peaks have respective densities of a second specified amount.

10 5452. The method of claim 5351, wherein the first specified amount is a first range and the second specified amount is a second range, and wherein the first range is less than the second range.

5553. The method of claim 5048, wherein comparing the respective density to the reference density to identify whether the genomic region exhibits a 1-copy loss or a 1-copy gain includes:

15 computing a difference between the respective density and the reference density; and

 comparing the difference to a cutoff value.

20 5654. The method of claim 5048, wherein the differential is normalized with the reference density by:

 dividing the differential by the reference density.

5755. The method of claim 5048, wherein the another density is the reference density.

25 5856. The method of claim 5755, wherein calculating the fractional concentration further includes multiplying the differential by two.

5957. The method of claim 5048, wherein the first density is calculated using respective densities identified as exhibiting a 1-copy gain, and wherein the another density is a second density calculated from respective densities identified as exhibiting a 1-copy loss.

6058. The method of claim 5957, wherein the differential is normalized with the reference density by:

computing a first ratio of the first density and the reference density;

computing a second ratio of the second density and the reference density,

5 wherein the differential is between the first ratio and the second ratio.

6159. The method of claim 5957, wherein comparing the respective density to the reference density to identify whether the genomic region exhibits a 1-copy loss or a 1-copy gain includes:

fitting peaks to a distribution curve of a histogram of the respective densities,

10 wherein the first density corresponds to a first peak and the second density corresponds to a second peak.

6260. The method of claim 5048, wherein all genomic regions determined to exhibit a statistically significant gain in the respective density relative to the reference density are identified as exhibiting a 1-copy gain.

15 6361. The method of claim 5048, wherein normalizing the respective amount to obtain a respective density includes using a same total number of aligned reference tags to determine the respective density and the reference density.

6462. The method of claim 5048, wherein normalizing the respective amount to obtain a respective density includes dividing the respective amount by a total number of aligned reference tags.

6563. The method of claim 5048, wherein the plurality of genomic regions each have a same length.

6664. The method of claim 5048, wherein the genomic regions are non-overlapping.

25 65. The method of any one of the claims above, wherein the biological sample is urine, pleural fluid, ascetic fluid, peritoneal fluid, saliva, cerebrospinal fluid, or a stool sample.

66. A computer program product comprising a computer readable medium storing a plurality of instructions for controlling a processor to perform an operation for methods described herein, the instructions comprising the steps of any one of the claims above.