[0005] Besides screening for an existence of cancer, current techniques are also lacking for providing a prognosis of a patient with cancer and for monitoring the effects of treatment (e.g. recovery after surgery or chemotherapy or immunotherapy or targeted therapy). Such techniques are often expensive (e.g. imaging techniques), inaccurate, ineffective, insensitive or may subject the patient to the radiation used for imaging techniques.

[0006] Accordingly, it is desirable to provide new techniques for screening, prognosticating, and monitoring a patient for cancer.

BRIEF SUMMARY

[0006a] In one aspect, the present disclosure provides a method of analyzing a biological sample of an organism, the biological sample including nucleic acid molecules originating from normal cells and potentially from cells associated with cancer, wherein at least some of the nucleic acid molecules are cell-free in the biological sample, the method comprising: identifying one or more first chromosomal regions of the organism, wherein the one or more first chromosomal regions are non-overlapping, each first chromosomal region including a first plurality of loci and associated with a copy number aberration in a cancer; enriching the biological sample for nucleic acid molecules originating from the one or more first chromosomal regions, thereby providing an enriched biological sample; for each of a plurality of nucleic acid molecules in the enriched biological sample of the organism: identifying a location of the nucleic acid molecule in a reference genome of the organism; and for each first chromosomal region of the one or more first chromosomal regions: identifying a first group of nucleic acid molecules as being from the first chromosomal region based on the identified locations, the first group including at least one nucleic acid molecule located at each of the first plurality of loci of the first chromosomal region; calculating a first value of the first group of nucleic acid molecules, the first value defining a property of the nucleic acid molecules of the first group, wherein the first value determines a parameter; and comparing the paramater to a reference value to determine a classification of whether the first chromosomal region exhibits a deletion or an amplification in any cells associated with the cancer.

[0006b] In another aspect, the present disclosure provides a computer program product comprising a computer readable medium encoded with a plurality of instructions for controlling a computing system to perform the method of the disclosure.

[0006c] In another aspect, the present disclosure provides a computer system including processors that are programmed to implement the method of the disclosure.

[0007] Examples provide systems, apparatus, and methods for determining genetic aberrations associated with cancer. Biological samples including cell-free DNA fragments are analyzed to identify imbalances in chromosomal regions, e.g., due to deletions and/or amplifications in a tumor. Using a chromosomal region with multiple loci can allow for greater efficiency and/or accuracy. Such imbalances can then be used to diagnose or screen a patient for cancer, as well as prognosticate a patient with cancer. The severity of an imbalance as well as the number of regions exhibiting an imbalance can be used. Additionally, a patient can be tested over time to track severity of each of one or more chromosomal regions and a number of chromosomal regions to enable screening and prognosticating, as well as monitoring of progress (e.g. after treatment).

[0008] According to one example, a method of analyzing a biological sample of an organism for chromosomal deletions or amplifications associated with cancer is provided. The biological sample includes nucleic acid molecules originating from normal cells and potentially from cells associated with cancer. At least some of the nucleic acid molecules are cell-free in the sample. First and second haplotypes are determined for normal cells of the organism at a first chromosomal region. The first chromosomal region including a first plurality of heterozygous loci. Each of a plurality of the nucleic acid molecules in the sample have a location in a reference genome of the organism identified and have a respective allele determined. The locations and determined alleles are used to determine a first group of nucleic acid molecules from the first haplotype and a second group from the second haplotype. A computer system calculates a first value of the first group and a second value of the second group. Each value defined a property of the respective group of nucleic acid molecules (e.g. an average size or number of molecules in the group). The first value is compared to the second value to determine

a classification of whether the first chromosomal region exhibits a deletion or an amplification in any cells associated with cancer.

[0009] According to another example, a method of analyzing a biological sample of an organism is provided. The biological sample includes nucleic acid molecules originating from normal cells and potentially from cells associated with cancer. At least some of the nucleic acid molecules are cell-free in the sample. A plurality of non-overlapping chromosomal regions of the organism are identified. Each chromosomal region includes a plurality of loci. Each of a plurality of the nucleic acid molecules in the sample have a location in a reference genome of the organism identified. For each chromosomal region, a respective group of nucleic acid molecules are identified as being from the chromosomal region based on the identified locations. Each respective group includes at least one nucleic acid molecule located at each of the plurality of loci of the chromosomal region. A computer system calculates a respective value of the respective group, where the respective value defines a property of the nucleic acid molecules of the respective group. The respective value is compared to a reference value to determine a classification of whether the chromosomal region exhibits a deletion or an amplification. An amount of chromosomal regions classified as exhibiting a deletion or amplification is then determined.

[0010] According to another example, a method is provided for determining a progress of chromosomal aberrations in an organism using biological samples including nucleic acid molecules originating from normal cells and potentially from cells associated with cancer. At least some of the nucleic acid molecules are cell-free in the biological samples. One or more non-overlapping chromosomal regions are identified for a reference genome of the organism. Each chromosomal region includes a plurality of loci. Samples taken from the organism at different times are analyzed to determine the progress. For a sample, each of a plurality of the nucleic acid molecules in the sample have a location in a reference genome of the organism identified. For each chromosomal region, a respective group of nucleic acid molecules are identified as being from the chromosomal region based on the identified locations. The respective group including at least one nucleic acid molecule located at each of the plurality of loci of the chromosomal region. A computer system calculates a respective value of the respective group of nucleic acid molecules. The respective value defines a property of the nucleic acid molecules of the respective group. The respective value is compared to a reference value to determine a classification of whether the first chromosomal region exhibits a deletion or

an amplification. Then, the classifications of each of the chromosomal regions at the plurality of times are used to determine the progress of the chromosomal aberrations in the organism.

[**0011**] Other examples of the disclosure are directed to systems, portable consumer devices, and computer readable media associated with methods described herein.

[0012] 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

[0013] FIG. I illustrates a chromosomal region of a cancer cell exhibiting an aberration of a deletion.

[0014] FIG. 2 illustrates a chromosomal region of a cancer cell exhibiting an aberration of an amplification.

[0015] FIG. 3 shows a table 300 illustrating different types of cancers and associated regions and their corresponding aberrations.

[0016] FIG. 4 illustrates chromosomal regions within a cancer cell that do not exhibit an aberration along with measurements made in plasma according to embodiments of the present invention.

[0017] FIG. 5 illustrates the deletion of chromosomal region 510 within a cancer cell along with measurements made in plasma to determine the deleted region according to embodiments of the present invention.

[0018] FIG. 6 illustrates the amplification of chromosomal region 610 within a cancer cell along with measurements made in plasma to determine the amplified region according to embodiments of the present invention.

[0019] FIG. 7 shows an RHDO analysis of the plasma DNA of an hepatocellular carcinoma (HCC) patient for a segment located at chromosome 1p which showed mono-allelic amplification in the tumor tissue according to embodiments of the present invention.

[0020] FIG. 8 shows the change in size distribution of fragments for two haplotypes of a chromosomal region when a tumor containing a deletion is present according to embodiments of the present invention.

CLAIMS:

1. A method of analyzing a biological sample of an organism, the biological sample including nucleic acid molecules originating from normal cells and potentially from cells associated with cancer, wherein at least some of the nucleic acid molecules are cell-free in the biological sample, the method comprising:

identifying one or more first chromosomal regions of the organism, wherein the one or more first chromosomal regions are non-overlapping, each first chromosomal region including a first plurality of loci and associated with a copy number aberration in a cancer;

enriching the biological sample for nucleic acid molecules originating from the one or more first chromosomal regions, thereby providing an enriched biological sample;

for each of a plurality of nucleic acid molecules in the enriched biological sample of the organism:

identifying a location of the nucleic acid molecule in a reference genome of the organism; and

for each first chromosomal region of the one or more first chromosomal regions:

identifying a first group of nucleic acid molecules as being from the first chromosomal region based on the identified locations, the first group including at least one nucleic acid molecule located at each of the first plurality of loci of the first chromosomal region;

calculating a first value of the first group of nucleic acid molecules, the first value defining a property of the nucleic acid molecules of the first group, wherein the first value determines a parameter; and

comparing the parameter to a reference value to determine a classification of whether the first chromosomal region exhibits a deletion or an amplification in any cells associated with the cancer.

2. The method of claim 1, further comprising:

identifying one or more second chromosomal regions of the organism, each second chromosomal region including a second plurality of loci and not associated with the copy number aberration in the cancer, wherein the enriched biological sample further includes the

biological sample enriched with nucleic acid molecules originating from the one or more second chromosomal regions; and

for each first chromosomal region of the one or more first chromosomal regions:

identifying a second group of nucleic acid molecules as being from a second chromosomal region of the one or more second chromosomal regions based on the identified locations, the second group including at least one nucleic acid molecule located at each of the second plurality of loci of the second chromosomal region;

calculating a second value of the second group of nucleic acid molecules, the second value defining the property of the nucleic acid molecules of the second group; and

determining the parameter using the first value and the second value.

3. The method of claim 2, wherein each locus of the first plurality of loci is at least 500 bases apart from another locus of the first plurality of loci, and wherein each locus of the second plurality of loci is at least 500 bases apart from another locus of the second plurality of loci.

4. The method of any one of claims 2 to 3, wherein the parameter includes a difference or a ratio between the first value and the second value.

5. The method of any one of claims 2 to 4, wherein the first value corresponds to a statistical value of a size distribution of the nucleic acid molecules of the first group, and the second value corresponds to a statistical value of a size distribution of the nucleic acid molecules of the second group.

6. The method of claim 5, wherein the first value is an average size of the nucleic acid molecules of the first group, and the second value is an average size of the nucleic acid molecules of the second group.

7. The method of claim 5, wherein the first value Q_I is a fraction of nucleic acid molecules in the first group that are shorter than a cutoff size, and the second value Q_{II} is a fraction of nucleic acid molecules in the second group that are shorter than the cutoff size.

8. The method of claim 7, wherein the parameter includes a difference between the first value and the second value, wherein the difference is $\Delta Q = Q_I - Q_{II}$, wherein a value of ΔQ greater than a first threshold value indicates that the first chromosomal region includes the amplification in tumor tissue in the organism, wherein a value of ΔQ less than a second threshold value indicates that the first chromosomal region includes the deletion in the tumor tissue in the organism, and wherein an approximate value of zero for ΔQ indicates that no deletion or amplification exists in the first chromosomal region.

9. The method of claim 5, wherein the first value F_I and the second value F_{II} are defined for a respective group as $F = \sum^{w}\text{length} / \sum^{N}\text{length}$, where $\sum^{w}\text{length}$ represents a sum of lengths of the nucleic acid molecules of the corresponding group with a length equal to or less than a cutoff size w; and $\sum^{N}\text{length}$ represents a sum of lengths of the nucleic acid molecules of the corresponding group with a length equal to or less than N bases, where N is greater than w.

10. The method of claim 9, wherein the parameter includes a difference between the first value and the second value, and wherein the difference is $\Delta F = F_I - F_{II}$, wherein a value of ΔF greater than a first threshold value indicates that the first chromosomal region includes the amplification in tumor tissue in the organism, wherein a value of ΔF less than a second threshold value indicates that the first chromosomal region includes the deletion in the tumor tissue in the organism, and wherein an approximate value of zero for ΔF indicates that no deletion or amplification exists in the first chromosomal region.

11. The method of any one of claims 2 to 4, wherein the first value of the first group corresponds to a number of nucleic acid molecules located at the first plurality of loci of the first chromosomal region, and the second value of the second group corresponds to a number of nucleic acid molecules located at the second plurality of loci of the second chromosomal region.

12. The method of any one of claims 1 to 11, wherein the one or more first chromosomal regions includes a plurality of first chromosomal regions, the method further comprising:

determining a first number of the plurality of first chromosomal regions that exhibit the deletion or the amplification; and

comparing the first number to one or more threshold values to determine a level of the cancer in the organism.

13. The method of claim 12, further comprising:

determining an amount of deletion or amplification for each first chromosomal region that was identified as exhibiting the deletion or the amplification; and

comparing the amount to one or more threshold values to determine the level of the cancer in the organism.

14. The method of claim 12, further comprising:

repeating the method of claim 12 at a plurality of times; and

using the first number at the plurality of times to diagnose, stage, prognosticate, or monitor progress of the level of the cancer in the organism.

15. The method of claim 14, wherein using the first number at the plurality of times to diagnose, stage, prognosticate, or monitor progress of the level of the cancer in the organism includes determining a presence or progress of a premalignant condition.

16. The method of claim 12, wherein each of the plurality of first chromosomal regions are of predetermined length.

17. The method of any of one of claims 1 to 11, wherein the one or more first chromosomal regions includes a plurality of first chromosomal regions, the method further comprising:

identifying which of the plurality of first chromosomal regions exhibit the copy number aberration associated with the cancer; and

determining whether the organism has the cancer based on which of the plurality of first chromosomal regions exhibit the copy number aberration associated with the cancer.

18. The method of claim 17, wherein the cancer is thyroid cancer, gastric cancer, prostate cancer, small cell lung cancer, non-small cell lung cancer, nasopharyngeal carcinoma, bladder cancer, colorectal cancer, head and neck cancer, melanoma, or lymphoma.

19. The method of any one of claims 1 to 18, wherein the reference value is derived from a healthy organism or another organism known to have the cancer.

20. The method of any one of claims 1 to 19, wherein the one or more first chromosomal regions are classified as exhibiting the deletion or the amplification in cells associated with the cancer, and wherein the cells associated with the cancer include a malignant tumor and/or from a premalignant lesion.

21. A computer program product comprising a computer readable medium encoded with a plurality of instructions for controlling a computing system to perform the method of any one of claims 1 to 20.

22. A computer system including processors that are programmed to implement any of the method of any one of claims 1 to 20.

ABSTRACT

DETECTION OF GENETIC OR MOLECULAR ABERRATIONS ASSOCIATED WITH CANCER

Systems, apparatus, and methods are provided for determining genetic or molecular aberrations in a biological sample from an organism. Biological samples including cell-free DNA fragments are analyzed to identify imbalances in chromosomal regions, e.g., due to deletions and/or amplifications in a tumor. Multiple loci are used for each chromosomal region. Such imbalances can then be used to diagnose (screen) a patient for cancer, as well as prognosticate a patient with cancer, or to detect the presence or to monitor the progress of a premalignant condition in a patient. The severity of an imbalance as well as the number of regions exhibiting an imbalance can be used. A systematic analysis of non-overlapping segments of a genome can provide a general screening tool for a sample. Additionally, a patient can be tested over time to track severity of each of one or more chromosomal regions and a number of chromosomal regions to enable screening and prognosticating, as well as monitoring of progress (e.g. after treatment).

[Figure 10]