[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 invention provides a method of analyzing a biological sample of an organism for chromosomal deletions or amplifications associated with cancer, 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: for each of a plurality of the nucleic acid molecules in the biological sample: identifying a location of the nucleic acid molecule in a reference genome of the organism; and determining a size of the nucleic acid molecule; identifying a first group of nucleic acid molecules as being from a first chromosomal region based on the identified locations; identifying a second group of nucleic acid molecules as being from a reference chromosomal region based on the identified locations; calculating, with a computer system, a first statistical value of a first size distribution of the nucleic acid molecules of the first group; calculating, with the computer system, a second statistical value of a second size distribution of the nucleic acid molecules of the second group; and comparing the first statistical value to the second statistical value to determine a classification of whether the first chromosomal region exhibits a deletion or an amplification in any cells associated with cancer.

[0006b] In another aspect, the present invention provides a computer program comprising a plurality of instructions that when executed control the computing system to perform the method of the present invention.

[0006c] In yet another aspect, the present invention provides a computer readable storage medium comprising the computer program of the present invention.

[0006d] In yet another aspect, the present invention provides a computer system comprising the computer program of the present invention.

[0006e] In yet another aspect, the present invention provides a data packet or packets comprising computer program code that when executed by a computer system implements the method of the present invention.

[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 for chromosomal deletions or amplifications associated with cancer, 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:

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

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

determining a size of the nucleic acid molecule;

identifying a first group of nucleic acid molecules as being from a first chromosomal region based on the identified locations;

identifying a second group of nucleic acid molecules as being from a reference chromosomal region based on the identified locations;;

calculating, with a computer system, a first statistical value of a first size distribution of the nucleic acid molecules of the first group;

calculating, with the computer system, a second statistical value of a second size distribution of the nucleic acid molecules of the second group; and

comparing the first statistical value to the second statistical value to determine a classification ofwhether the first chromosomal region exhibits a deletion or an amplification in any cells associated with cancer.

2. The method of claim 1, wherein the first chromosomal region is classified as exhibiting a deletion or an amplification in cells associated with cancer, and wherein the cells associated with cancer include a malignant tumor and/or a premalignant lesion.

3. The method of claim 1, wherein comparing the first statistical value to the second statistical value includes:

determining a difference between the first statistical value and the second statistical value, and wherein the classification includes comparing the difference to at least one threshold value.

4. The method of claim 3, wherein the at least one threshold value is derived from a healthy organism or from a region not having a deletion or an amplification, wherein the sequential probability ratio testing, the t-test, or the chi-square test is used to determine the at least one threshold value.

5. The method of claim 1, wherein the first statistical value is an average size of the nucleic acid molecules of the first group, and the second statistical value is an average size of the nucleic acid molecules of the second group.

6. The method of claim 1, wherein the first statistical value Q_I is a fraction of nucleic acid molecules in the first group that are shorter than a cutoff size, and the second statistical value Q_{II} is a fraction of nucleic acid molecules in the second group that are shorter than the cutoff size.

7. The method of claim 6, wherein comparing the first statistical value to the second statistical value includes:

determining a difference between the first statistical value and the second statistical value, wherein the difference is $\Delta Q = Q_I - Q_{II}$, and wherein a positive value of ΔQ greater than a threshold value indicates that the first chromosomal region includes an amplification in tumor tissue in the organism.

8. The method of claim 6, wherein comparing the first statistical value to the second statistical value includes:

determining a difference between the first statistical value and the second statistical value, wherein the difference is $\Delta Q = Q_I - Q_{II}$, and wherein a negative value of ΔQ greater than a threshold value indicates that the first chromosomal region includes a deletion in tumor tissue in the organism.

9. The method of claim 6, wherein comparing the first statistical value to the second statistical value includes:

determining a difference between the first statistical value and the second statistical value, wherein the difference is $\Delta Q = Q_I - Q_{II}$, and wherein an approximate value of zero for ΔQ indicates that no deletion or amplification exists in the first chromosomal region.

10. The method of claim 1, wherein the first statistical value F_I and the second statistical value F_{II} are defined for a respective chromosomal region 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.

11. The method of claim 10, wherein comparing the first statistical value to the second statistical value includes:

determining a difference between the first statistical value and the second statistical value, wherein the difference is $\Delta F = F_I - F_{II}$, and wherein a positive value of ΔF greater than a threshold value indicates that the first chromosomal region includes an amplification in tumor tissue in the organism.

4. The method of claim10, wherein comparing the first statistical value to the second statistical value includes:

determining a difference between the first statistical value and the second statistical value, wherein the difference is $\Delta F = FI - FII$, and wherein a negative value of ΔF greater than a threshold value indicates that the first chromosomal region includes a deletion in tumor tissue in the organism.

13. The method of claim 1, further comprising:

for each of a plurality of other chromosomal regions:

repeating the method of claim 1;

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

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

14. The method of claim 13, further comprising:

for each chromosomal region that was identified as exhibiting a deletion or an amplification;

determining an amount of deletion or amplification; and

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

15. The method of claim13, further comprising:

repeating the method of claim 13 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 cancer in the organism.

16. The method of claim15, wherein using the first number at the plurality of times to diagnose, stage, prognosticate, or monitor progress of the level of cancer in the organism includes determining a presence or progress of a premalignant condition.

17. The method of claim 13, wherein each of the chromosomal regions are of predetermined length.

18. The method of claim 13 , wherein the plurality of chromosomal regions span the genome of the organism.

19. The method of claim 1, wherein the reference chromosomal region does not exhibit a deletion or amplification at a time the nucleic acid molecules were obtained.

20. A computer program comprising a plurality of instructions that when executed control the computing system to perform the method of any one of claims 1 to 19.

21. A computer readable storage medium comprising the computer program of claim 20.

22. A computer system comprising the computer program of claim 20.

23. A data packet or packets comprising computer program code that when executed by a computer system implements the method of any one of claims 1 to 19.