

CLAIMS

1. A method of analyzing a biological sample of an organism for deletions or amplifications in one or more chromosomal regions associated with cancer, the biological sample including cell-free nucleic acid molecules originating from non-malignant cells and potentially from tumor cells associated with cancer, the method comprising:
 - receiving sequences obtained from a random sequencing of nucleic acid molecules contained in the biological sample;
 - aligning at least a portion of the sequences to a human genome;
 - determining a first amount of the sequences identified as aligning to a first chromosomal region that is part of a first chromosome, wherein an aberration in the first chromosomal region is associated with cancer;
 - determining a second amount of the sequences identified as aligning to one or more second chromosomal regions;
 - determining a first parameter from the first amount and the second amount, wherein the parameter includes a ratio between the first and second amounts; and
 - comparing the first parameter to one or more cutoff values to determine a classification of whether the first chromosomal regions exhibits a deletion and/or an amplification associated with cancer.
2. The method of claim 1, wherein the classification includes whether the first chromosomal region is amplified, and wherein a classification of amplification is made when the first parameter is greater than a first cutoff value of the one or more cutoff values.
3. The method of claim 1, wherein the classification includes whether the first chromosomal region has a deletion, and wherein a classification of deletion is made when the first parameter is less than a first cutoff value of the one or more cutoff values.
4. The method of claim 1, further comprising:
 - determining a third amount of the sequences identified as aligning to a third chromosomal region, wherein an aberration in the third chromosomal region is associated with cancer;
 - determining a second parameter for the third chromosomal region, wherein the second parameter represents a relative amount between the third and second amounts;

comparing the second parameter to one or more cutoff values to determine a classification of whether the third chromosomal region exhibits a deletion and/or an amplification associated with cancer; and

using the classifications of the first and third chromosomal regions to detect whether the organism has cancer corresponding to the classifications of the first and third chromosomal regions.

5. The method of claim 1, wherein the biological sample is plasma, serum, urine or saliva.

6. The method of claim 1, wherein the biological sample is transcervical lavage fluid.

7. The method of claim 1, wherein the ratio provides a percentage of the sequences identified as aligning to the first chromosomal region among the sequences.

8. The method of claim 1, wherein the ratio is obtained from any one or more of a fractional count of the number of sequenced tags, a fractional number of sequenced nucleotides, and a fractional length of accumulated sequences.

9. The method of claim 1, wherein the sequences that are used to determine the first amount are selected to be less than a specified number of bases.

10. The method of claim 9, wherein the specified number of bases is 300 bases, 200 bases, or 100 bases.

11. The method of claim 1, wherein the nucleic acid molecules of the biological sample have been enriched for sequences originating from the first chromosomal region.

12. The method of claim 1, wherein the nucleic acid molecules of the biological sample have been enriched for sequences less than 300 bases.

13. The method of claim 1, wherein the nucleic acid molecules of the biological sample have been enriched for sequences less than 200 bases.

14. The method of claim 1, wherein the nucleic acid molecules of the biological sample have been amplified using a polymerase chain reaction.

15. The method of claim 1, wherein the aligned portion represents at least a pre-determined fraction of the human genome.

16. The method of claim 1, wherein the fraction represents at least 0.1% of the human genome.

17. The method of claim 1, wherein the fraction represents at least 0.5% of the human genome.

18. The method of claim 1, wherein at least one of the cutoff values is derived from a percentage of tumor cells in the biological sample.

19. The method of claim 1, wherein at least one of the cutoff values is a reference value established in a normal biological sample.

20. The method of claim 19, wherein the at least one cutoff value is determined from a mean of the value of the first parameter in a plurality of normal biological samples.

21. The method of any one of claims 1, wherein the one or more cutoff values are derived from a normal value for the relative amount between the first and second amounts.

22. The method of claim 1, further comprising:
identifying an amount of tumor DNA in the biological sample; and
calculating a number N of sequences to be analyzed based on a desired accuracy.

23. The method of claim 1 to 22, wherein the sequences identified as aligning to the first chromosomal region uniquely align to the first chromosomal region.

24. The method of claim 1 to 23, wherein the sequences identified as aligning to the one or more of the second chromosomal regions uniquely align to only one of the second chromosomal regions.

25. The method of claim 1 to 24, wherein the human genome is repeat-masked.
26. The method of any one of claims 1 to 25, wherein the first amount is determined based on a pool of sequenced tags that align to a plurality of positions of the first chromosomal region.
27. The method of claim 26, wherein which part of the human genome that the sequenced tags are aligned is not pre-determined.
28. The method of claim 27, wherein the part of the human genome corresponding to the sequenced tags aligned to the first chromosomal region is not pre-determined.
29. The method of claim 1, wherein the first parameter is a fractional representation of sequences identified as aligning to the first chromosomal region, the fractional representation being a measure of a proportion of nucleic acid molecules in the biological sample that are from the first chromosomal region.
30. The method of claim 29, wherein the one or more cutoff values take into account a size of the first chromosomal region relative to a size of the one or more second chromosomal regions.
31. The method of claim 1, wherein the sequencing includes paired-end sequencing, and wherein aligning each sequence includes aligning paired-end sequences.
32. The method of claim 1, further comprising:
performing the random sequencing of nucleic acid molecules contained in the biological sample to obtain the sequences.
32. 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 31.

CLAIMS

1. A method of analyzing a biological sample of an organism for deletions or amplifications in one or more chromosomal regions associated with cancer in the organism, the biological sample being obtained from the organism and including cell-free nucleic acid molecules originating from non-malignant cells and potentially from tumor cells associated with cancer, the method comprising:

receiving sequences obtained from a random sequencing of nucleic acid molecules contained in the biological sample;

aligning at least a portion of the sequences to a human genome;

determining a first amount of the sequences obtained from the random sequencing of nucleic acid molecules contained in the biological sample that are identified as aligning to a first chromosomal region that is part of a first chromosome, wherein an aberration in the first chromosomal region is associated with cancer;

determining a second amount of the sequences obtained from the random sequencing of nucleic acid molecules contained in the biological sample that are identified as aligning to one or more second chromosomal regions;

determining a first parameter from the first amount and the second amount, wherein the first parameter includes a ratio between the first and second amounts; and

comparing the first parameter to one or more cutoff values to determine a classification of whether the first chromosomal region exhibits a deletion and/or an amplification associated with cancer in the organism.

2. The method of claim 1, wherein the classification includes whether the first chromosomal region is amplified, and wherein a classification of amplification is made when the first parameter is greater than a first cutoff value of the one or more cutoff values.

3. The method of claim 1, wherein the classification includes whether the first chromosomal region has a deletion, and wherein a classification of deletion is made when the first parameter is less than a first cutoff value of the one or more cutoff values.

4. The method of claim 1, further comprising:

determining a third amount of the sequences identified as aligning to a third chromosomal region, wherein an aberration in the third chromosomal region is associated with cancer;

determining a second parameter for the third chromosomal region, wherein the second parameter represents a relative amount between the third and second amounts;

comparing the second parameter to the one or more cutoff values to determine a classification of whether the third chromosomal region exhibits a deletion and/or an amplification associated with cancer; and

using the classifications of the first and third chromosomal regions to detect whether the organism has cancer corresponding to the classifications of the first and third chromosomal regions.

5. The method of claim 1, wherein the biological sample is plasma, serum, urine or saliva.

6. The method of claim 1, wherein the biological sample is transcervical lavage fluid.

7. The method of claim 1, wherein the ratio provides a percentage of the sequences identified as aligning to the first chromosomal region among the sequences.

8. The method of claim 1, wherein the ratio is obtained from any one or more of a fractional count of the number of aligned sequences, a fractional number of sequenced nucleotides, and a fractional length of accumulated sequences.

9. The method of claim 1, wherein the sequences that are used to determine the first amount are selected to be less than a specified number of bases.

10. The method of claim 9, wherein the specified number of bases is 300 bases, 200 bases, or 100 bases.

11. The method of claim 1, wherein the nucleic acid molecules of the biological sample have been enriched for sequences originating from the first chromosomal region.

12. The method of claim 1, wherein the nucleic acid molecules of the biological sample have been enriched for sequences less than 300 bases.

13. The method of claim 1, wherein the nucleic acid molecules of the biological sample have been enriched for sequences less than 200 bases.

14. The method of claim 1, wherein the nucleic acid molecules of the biological sample have been amplified using a polymerase chain reaction.

15. The method of claim 1, wherein the aligned portion represents at least a pre-determined fraction of the human genome.

16. The method of claim 1, wherein the pre-determined fraction represents at least 0.1% of the human genome.

17. The method of claim 1, wherein the pre-determined fraction represents at least 0.5% of the human genome.

18. The method of claim 1, wherein at least one of the one or more cutoff values is derived from a percentage of tumor cells in the biological sample.

19. The method of claim 1, wherein at least one of the one or more cutoff values is a reference value established in a normal biological sample.

20. The method of claim 19, wherein the at least one cutoff value is determined from a mean of the value of the first parameter in a plurality of normal biological samples.

21. The method of claim 1, wherein the one or more cutoff values are derived from a normal value for the ratio between the first and second amounts.

22. The method of claim 1, further comprising:
identifying an amount of tumor DNA in the biological sample; and
calculating a number N of sequences to be analyzed based on a desired accuracy.

23. The method of any one of claims 1 to 22, wherein the sequences identified as aligning to the first chromosomal region uniquely align to the first chromosomal region.

24. The method of any one of claims 1 to 23, wherein the sequences identified as aligning to the one or more of the second chromosomal regions uniquely align to only one of the second chromosomal regions.

25. The method of any one of claims 1 to 24, wherein the human genome is repeat-masked.

26. The method of any one of claims 1 to 25, wherein the first amount is determined based on a pool of sequences that align to a plurality of positions of the first chromosomal region.

27. The method of claim 26, wherein which part of the human genome that the sequences are aligned is not pre-determined.

28. The method of claim 27, wherein the part of the human genome corresponding to the sequences aligned to the first chromosomal region is not pre-determined.

29. The method of claim 1, wherein the first parameter is a fractional representation of sequences identified as aligning to the first chromosomal region, the fractional representation being a measure of a proportion of nucleic acid molecules in the biological sample that are from the first chromosomal region.

30. The method of claim 29, wherein the one or more cutoff values take into account a size of the first chromosomal region relative to a size of the one or more second chromosomal regions.

31. The method of claim 1, wherein the sequencing includes paired-end sequencing, and wherein aligning each sequence includes aligning paired-end sequences.

32. The method of claim 1, wherein the first parameter includes a ratio between the first and second amounts by including a simple ratio of the first amount divided by a term including the second amount or by including a ratio of separate functions of the first amount and the second amount.

33. The method of claim 1, further comprising:
performing the random sequencing of nucleic acid molecules contained in the biological sample to obtain the sequences.

34. 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 32.