

regions on chromosome 14 [8]. Jensen et al, on the other hand, only focused their statistical analysis on the deleted region on chromosome 22 [9]. Thus, from the data presented by Peters et al and Jensen et al, it is not clear if the approach would be robust enough for a genomewide survey of microdeletions or microduplications, or indeed for the noninvasive determination of a fetal karyotype.

[0005] Lo et al reported that fetal single nucleotide polymorphisms (SNPs) can be genotyped in a genomewide scale using maternal plasma DNA sequencing [10]. In particular, these investigators have demonstrated that SNP alleles and mutations for single gene disorders that are inherited by a fetus from its mother can be elucidated by a process called relative haplotype dosage analysis [10]. Fan et al confirmed the robustness of relative haplotype dosage analysis and used this approach to detect a ~2.85 Mb deletion inherited by a fetus from its mother [11]. There are two concerns for using this method for the clinical implementation of noninvasive prenatal karyotyping. First, this method requires maternal haplotyping to be performed which would require additional analytical steps [12,13] or pedigree analysis. Second, it is unclear if this method could be used to detect de novo subchromosomal deletion or duplication.

BRIEF SUMMARY

[0006] Disclosed herein are methods, systems, and apparatus for detecting microamplifications or microdeletions in the genome of a fetus. In some embodimentsexamples, the method comprises receiving sequence tags for each of a plurality of DNA fragments in a biological sample; determining genomic positions for the sequence tags; determining whether the density of DNA in each of a plurality of genomic regions is aberrantly high or low; identifying as a microamplification a set of consecutive genomic regions having aberrantly high density; and identifying as a microdeletion a set of consecutive genomic regions having aberrantly low density. The biological sample

may be a blood sample obtained noninvasively from a female subject pregnant with the fetus.

[0006a] In one aspect, disclosed herein is a method of estimating a fetal DNA percentage in a biological sample obtained from a female subject pregnant with a fetus, the biological sample including cell-free DNA from the fetus and from the female subject, 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 one or more genomic regions: determining, with a computer system, a respective amount of DNA fragments within the genomic region from sequence tags having genomic positions within the genomic region; and normalizing the respective amount to obtain a respective density; comparing the respective density to a reference density to identify whether the respective density is statistically different from the reference density; estimating the fetal DNA percentage from the reference density and the respective density of the one or more genomic regions identified to have respective densities statistically different than the reference densities, wherein: the size of each genomic region is less than or equal to 10 Mb.

[0006b] In another aspect, disclosed herein is a method of identifying microamplifications or microdeletions in a genome of a fetus by analyzing a biological sample obtained from a female subject pregnant with the fetus, the biological sample including cell-free DNA from the fetus and from the female subject, the method comprising: determining a minimum number of DNA fragments to be analyzed for a specified diagnostic sensitivity and a size of a microamplification or microdeletion corresponding to a specified resolution; receiving one or more sequence tags for each of a plurality of DNA fragments in the biological sample, the plurality of DNA fragments being at least the minimum number of DNA fragments for the specified diagnostic sensitivity and the specified resolution; determining genomic positions for the sequence tags; for each of a plurality of genomic regions: determining, with a computer system, a respective amount of DNA fragments within the genomic region from sequence tags having genomic positions within the genomic region; normalizing the respective amount to obtain a respective density; comparing the respective density to a reference density to identify whether the respective density is statistically different from the reference density;

when the respective density is statistically higher than the reference density, identifying the genomic regions as a microamplification; and when the respective density is statistically lower than the reference density, identifying the genomic region as a microdeletion.

5

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] FIG 1 is a flowchart showing a method of identifying microamplifications or microdeletions in a genome of a fetus.

WHAT IS CLAIMED IS:

1. A method of estimating a fetal DNA percentage in a biological sample obtained from a female subject pregnant with a fetus, the biological sample including cell-free DNA from the fetus and from the female subject, 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 one or more genomic regions:

determining, with a computer system, a respective amount of DNA fragments within the genomic region from sequence tags having genomic positions within the genomic region; and

normalizing the respective amount to obtain a respective density;

comparing the respective density to a reference density to identify whether the respective density is statistically different from the reference density;

estimating the fetal DNA percentage from the reference density and the respective density of the one or more genomic regions identified to have respective densities statistically different than the reference densities, wherein:

the size of each genomic region is less than or equal to 10 Mb.

2. The method of claim 1, wherein estimating the fetal DNA percentage comprises calculating the formula

$$\frac{(GR_{x-y_{test}} - meanGR_{x-y_{reference}}) \times 2}{meanGR_{x-y_{reference}}} \times 100\%$$

where $GR_{x-y_{test}}$ is the respective density of a genomic region of the one or more genomic regions, and $meanGR_{x-y_{reference}}$ is the reference density for the genomic region.

3. The method of claim 1 or claim 2, further comprising:

5

comparing the fetal DNA percentage to a cutoff value, and

determining, based on the comparison of the fetal DNA percentage to the cutoff value, whether the female subject and the fetus both share a microamplification or a microdeletion.

4. The method of claim 3, further comprising:

10

determining the female subject and the fetus both share the microamplification or the microdeletion when the fetal DNA percentage is greater than the cutoff value.

5. The method of claim 3, further comprising:

15

determining that only the fetus has the microamplification or the microdeletion when the fetal DNA percentage is less than the cutoff value.

6. The method of any one of claims 1 to 4, wherein the biological sample is maternal blood, plasma, serum, urine, or saliva.

7. The method of any one of claims 1 to 6, wherein the sequence tags are obtained by massively parallel sequencing.

20

8. The method of any one of claims 1 to 7, wherein the genomic regions are non-overlapping.

9. The method of any one of claims 1 to 8, wherein the genomic regions are contiguous.

10. The method of any one of claims 1 to 9, wherein the genomic regions are of equal size.

11. The method of claim 10, wherein the size of each genomic region is about 1 Mb.

5 12. The method of any one of claims 1 to 11, wherein the respective density for a genomic region is obtained by dividing the respective amount of DNA fragments for the genomic region by the total amount of DNA fragments for multiple genomic regions.

10 13. The method of any one of claims 1 to 11, wherein the respective density for a genomic region equals the respective amount of DNA fragments for the genomic region.

15 14. The method of any one of claims 1 to 11, wherein the reference density for a genomic region is a mean or median of a plurality of respective densities determined from one or more other biological samples not exhibiting microamplifications or microdeletions in the genomic region.

15 15. The method of any one of claims 1 to 11, wherein the reference density for a genomic region is the mean or median of a plurality of respective densities obtained for other genomic regions.

20 16. The method of any one of claims 1 to 15, further comprising displaying, by the computer system, the estimated fetal DNA percentage.

17. A method of identifying microamplifications or microdeletions in a genome of a fetus by analyzing a biological sample obtained from a female subject pregnant with the fetus, the biological sample including cell-free DNA from the fetus and from the female subject, the method comprising:

25 determining a minimum number of DNA fragments to be analyzed for a specified diagnostic sensitivity and a size of a microamplification or microdeletion corresponding to a specified resolution;

receiving one or more sequence tags for each of a plurality of DNA fragments in the biological sample, the plurality of DNA fragments being at least the minimum number of DNA fragments for the specified diagnostic sensitivity and the specified resolution;

5 determining genomic positions for the sequence tags;

for each of a plurality of genomic regions:

determining, with a computer system, a respective amount of DNA fragments within the genomic region from sequence tags having genomic positions within the genomic region;

10 normalizing the respective amount to obtain a respective density;

~~and~~

comparing the respective density to a reference density to identify whether the respective density is statistically different from the reference density;

~~determining whether any of the genomic regions identified to have a~~
15 ~~respective density statistically different from the reference density is consecutive with~~
~~another genomic region identified to have a respective density statistically different from~~
~~the reference density;~~

~~when the respective density is at least N first genomic regions identified to~~
~~have respective densities statistically higher than the reference density are consecutive,~~
20 identifying the ~~first consecutive~~ genomic regions as a microamplification, ~~N being an~~
~~integer equal to or greater than two; and~~

~~when the respective density is at least N second genomic regions~~
~~identified to have respective densities statistically lower than the reference~~
~~density densities are consecutive,~~ identifying the ~~second consecutive~~ genomic regions as a
25 microdeletion.

18. The method of claim 17, further comprising:

measuring the fetal DNA percentage in the biological sample; and

using the measured fetal DNA percentage to determine the minimum number of DNA fragments.

19. The method of claim 17, further comprising:
assuming a fetal DNA percentage in the biological sample; and
using the assumed fetal DNA percentage to determine the minimum
number of DNA fragments.

5 20. The method of any one of claims 17 to 19, wherein the minimum
number is specified for the entire genome.

21. The method of any one of claims 17 to 19, wherein the minimum
number of DNA fragments to be analyzed is calculated from a number of DNA fragments
for each genomic region of the plurality of genomic regions.

10 22. The method of claim 21, wherein the minimum number of DNA
fragments to be analyzed for a 95% diagnostic sensitivity is at least 42,000 for each
genomic region of the plurality of genomic regions.

23. The method of claim 21, wherein the minimum number of DNA
fragments to be analyzed for a 99% diagnostic sensitivity is at least 53,000 for each
15 genomic region of the plurality of genomic regions.

24. The method of claim 17, wherein the size of each genomic region
is about 1 Mb, about 2 Mb, or about 3 Mb.

25. The method of any one of claims 17 to 24, wherein the genomic
regions are of equal size.

20 26. A computer program product comprising instructions that when
executed control a computer system to perform the method of any one of claims 1 to 25.

27. A computer readable storage medium comprising the computer
program product of claim 26.

28. A computer system comprising the computer program product of
25 claim 26.

29. A data packet or packets comprising computer program code that when executed by a processor implements the method of any one of claims 1 to 25.

2. ~~—The method of claim 1, wherein the biological sample is maternal blood, plasma, serum, urine, or saliva.~~

5 3. ~~—The method of claim 1, wherein the sequence tags are obtained by massively parallel sequencing.~~

4. ~~—The method of claim 1, wherein N is 3.~~

5. ~~—The method of claim 1, wherein the genomic regions are non-overlapping.~~

10 6. ~~—The method of claim 1, wherein the genomic regions are contiguous.~~

7. ~~—The method of claim 1, wherein the genomic regions are of equal size.~~

15 8. ~~—The method of claim 7, wherein the size of each genomic region is about 1 Mb.~~

9. ~~—The method of claim 1, wherein the respective density for a genomic region is obtained by dividing the respective amount of DNA fragments for the genomic region by the total amount of DNA fragments for multiple genomic regions.~~

20 10. ~~—The method of claim 1, wherein the respective density for a genomic region equals the respective amount of DNA fragments for the genomic region.~~

11. ~~—The method of claim 1, wherein the reference density for a genomic region is a mean or median of a plurality of respective densities determined from one or more other biological samples not exhibiting microamplifications or microdeletions in the genomic region.~~

12. — ~~The method of claim 1, wherein the reference density for a genomic region is the mean or median of a plurality of respective densities obtained for other genomic regions.~~

13. — ~~The method of claim 1, further comprising the step of determining whether a microamplification or microdeletion is maternally inherited.~~

14. — ~~The method of claim 13, wherein a microamplification or microdeletion is determined to be maternally inherited if, for each of the consecutive genomic regions corresponding to the microamplification or microdeletion, the difference between the respective density and reference density exceeds a particular cutoff, the particular cutoff being larger than a cutoff used to determine whether the respective density and reference density are statistically different.~~

15. — ~~A computer product comprising a non-transitory computer readable medium storing a plurality of instructions that when executed control a computer system to identify microamplifications or microdeletions in a genome of a fetus by analyzing a biological sample obtained from a female subject pregnant with the fetus, the biological sample including cell-free DNA from the fetus and from the female subject, the method comprising:~~

~~receiving one or more sequence tags for each of a plurality of DNA fragments in a 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 genomic positions 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 respective density is statistically different from the reference density;~~

~~determining whether any of the genomic regions identified to have a respective density statistically different from the reference density is consecutive with~~

~~another genomic region identified to have a respective density statistically different from the reference density;~~

5 ~~when at least N first genomic regions identified to have respective densities statistically higher than the reference density are consecutive, identifying the first consecutive genomic regions as a microamplification, N being an integer equal to or greater than two;~~

~~when at least N second genomic regions identified to have respective densities statistically lower than the reference densities are consecutive, identifying the second consecutive genomic regions as a microdeletion.~~

10 ~~16. —The computer product of claim 15, wherein N is 3.~~

~~17. —The computer product of claim 15, wherein the genomic regions are non-overlapping.~~

~~18. —The computer product of claim 15, wherein the genomic regions are contiguous.~~

15 ~~19. —The computer product of claim 15, wherein the genomic regions are of equal size.~~

~~20. —The computer product of claim 15, wherein the respective density for a genomic region is obtained by dividing the respective amount of DNA fragments for the genomic region by the total amount of DNA fragments for multiple genomic regions.~~