DNA molecules, as the libraries are denatured, diluted and diffused on the two-dimensional surface followed by bridge amplification (Head et al. Biotechniques. 2014;56:61-4).

SUMMARY

[0002a] In one aspect, the present disclosure refers to a method of analyzing a biological sample obtained from a female pregnant with a fetus, the female having a first haplotype and a second haplotype in a first chromosomal region, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

receiving reads corresponding to the plurality of cell-free DNA molecules;

measuring sizes of the plurality of cell-free DNA molecules;

identifying a first set of cell-free DNA molecules from the plurality of cell-free DNA molecules as having sizes greater than or equal to a cutoff value, wherein the cutoff value is at least 500 nt;

determining a sequence of the first haplotype and a sequence of the second haplotype using reads corresponding to the first set of cell-free DNA molecules;

aligning a second set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the first haplotype, the second set of cell-free DNA molecules having sizes less than the cutoff value;

aligning a third set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the second haplotype, the third set of cell-free DNA molecules having sizes less than the cutoff value;

measuring a first value of a parameter using the second set of cell-free DNA molecules;

measuring a second value of the parameter using the third set of cell-free DNA molecules;

comparing the first value to the second value; and

determining a likelihood of the fetus inheriting the first haplotype based on the comparison of the first value to the second value.

[0002b] In another aspect, the present disclosure refers to a computer product comprising a non-transitory computer readable medium storing instructions that, when executed, control a computing system to perform a method of analyzing a biological sample obtained from a female pregnant with a fetus, the female having a first haplotype and a second haplotype in a first chromosomal region, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

receiving reads corresponding to the plurality of cell-free DNA molecules;

measuring sizes of the plurality of cell-free DNA molecules;

identifying a first set of cell-free DNA molecules from the plurality of cell-free DNA molecules as having sizes greater than or equal to a cutoff value, wherein the cutoff value is at least 500 nt;

determining a sequence of the first haplotype and a sequence of the second haplotype using reads corresponding to the first set of cell-free DNA molecules;

aligning a second set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the first haplotype, the second set of cell-free DNA molecules having sizes less than the cutoff value;

aligning a third set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the second haplotype, the third set of cell-free DNA molecules having sizes less than the cutoff value;

measuring a first value of a parameter using the second set of cell-free DNA molecules;

measuring a second value of the parameter using the third set of cell-free DNA molecules;

comparing the first value to the second value; and

determining a likelihood of the fetus inheriting the first haplotype based on the comparison of the first value to the second value.

[0003] Methods and systems described herein involve using long cell-free DNA fragments to analyze a biological sample. Using these long cell-free DNA fragments allows for analysis not contemplated or not possible with shorter cell-free DNA fragments. The status of methylated CpG sites and single nucleotide polymorphisms (SNPs) is often used to analyze DNA fragments of a biological sample. A CpG site and a SNP are typically separated from the nearest CpG site

or SNP by hundreds or thousands of base pairs. The length of most of the cell-free DNA fragments in a biological sample is usually less than 200 bp. As a result, finding two or more consecutive CpG sites or SNPs on most cell-free DNA fragments is improbable or impossible. Cell-free DNA fragments longer than 200 bp, including those longer than 600 bp or 1 kb, may include multiple CpG sites and/or SNPs. The presence of multiple CpG sites and/or SNPs on long cell-free DNA fragments may allow for more efficient and/or accurate analysis than with short cell-free DNA fragments alone. The long cell-free DNA fragments can be used to identify a tissue of origin and/or to provide information on a fetus in a pregnant female. In addition, using long cell-free DNA fragments to accurately analyze samples from pregnant women is surprising as one would expect that such long cell-free DNA fragments are predominantly maternal in origin. One would not expect that long cell-free DNA fragments of fetal origin are present in sufficient amounts to provide information about the fetus.

[0004] Long cell-free DNA fragments with a SNP present may be used to determine the haplotype inherited by a fetus. Long cell-free DNA fragments, by having multiple CpG sites, may have a methylation pattern that indicates a tissue of origin. Additionally, trinucleotide repeats and other repeated sequences may be present on long cell-free DNA fragments. These repeats may be used to determine the likelihood of a genetic disorder in fetus or the paternity of a fetus. The amount of long cell-free DNA fragments may be used to determine gestational age. Similarly, the motifs at the end of long cell-free DNA fragments may also be used to determine gestational age. The long-cell free DNA fragments (including, for example, amounts, length

distribution, genomic locations, methylation status, etc. of such fragments) may be used to determine a pregnancy-associated disorder.

[0005] These and other examples of the disclosure are described in detail below. For example, other examples are directed to systems, devices, and computer readable media associated with methods described herein.

[0006] A better understanding of the nature and advantages of examples of the present disclosure may be gained with reference to the following detailed description and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] FIGS. 1A and 1B show the size distribution of cell-free DNA determined according to embodiments of the present invention. (A) 0–20 kb on a linear scale, (B) 0–20 kb on a logarithmic scale.

[0008] FIGS. 2A and 2B show the size distribution of cell-free DNA determined according to embodiments of the present invention. (A) 0–5 kb on a linear scale for the y-axis. (B) 0–5 kb on a logarithmic scale for the y-axis.

[0009] FIGS. 3A and 3B show the size distribution of cell-free DNA determined according to embodiments of the present invention. (A) 0–400 bp on a linear scale for the y-axis. (B) 0–400 bp on a logarithmic scale for the y-axis.

[0010] FIGS. 4A and 4B show the size distribution of cell-free DNA between fragments carrying shared alleles (Shared) and fetal-specific alleles (Fetal-specific) determined according to embodiments of the present invention. (A) 0–20 kb bp on a linear scale for the y-axis. (B) 0–20 kb on a logarithmic scale for the y-axis. The blue line indicates the fragments carrying shared alleles (predominant of maternal origin) and the red line indicates the fragments carrying fetal-specific alleles (of placental origin).

[0011] FIGS. 5A and 5B show the size distribution of cell-free DNA between fragments carrying shared alleles (Shared) and fetal-specific alleles (Fetal-specific) determined according to embodiments of the present invention. (A) 0–5 kb bp on a linear scale for the y-axis. (B) 0–5 kb

WHAT IS CLAIMED IS:

1. A method of analyzing a biological sample obtained from a female pregnant with a fetus, the female having a first haplotype and a second haplotype in a first chromosomal region, the biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

receiving reads corresponding to the plurality of cell-free DNA molecules;

measuring sizes of the plurality of cell-free DNA molecules;

identifying a first set of cell-free DNA molecules from the plurality of cell-free DNA molecules as having sizes greater than or equal to a cutoff value, wherein the cutoff value is at least 500 nt;

determining a sequence of the first haplotype and a sequence of the second haplotype using reads corresponding to the first set of cell-free DNA molecules;

aligning a second set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the first haplotype, the second set of cell-free DNA molecules having sizes less than the cutoff value;

aligning a third set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the second haplotype, the third set of cell-free DNA molecules having sizes less than the cutoff value;

measuring a first value of a parameter using the second set of cell-free DNA molecules;

measuring a second value of the parameter using the third set of cell-free DNA molecules;

comparing the first value to the second value; and

determining a likelihood of the fetus inheriting the first haplotype based on the comparison of the first value to the second value.

2. The method of claim 1, wherein the cutoff value is 600 nt.

3. The method of claim 1, wherein the cutoff value is 1 knt.

4. The method of any one of claims 1 to 3, wherein determining the sequence of the first haplotype and the sequence of the second haplotype from the reads corresponding to the first set of cell-free DNA molecules comprises:

aligning reads corresponding to the first set of cell-free DNA molecules to a reference genome.

5. The method of claim 1, wherein determining the sequence of the first haplotype and the sequence of the second haplotype from the reads corresponding to the first set of cell-free DNA molecules comprises:

aligning a first subset of the reads to a second subset of the reads to identify a different allele at a locus in the reads,

determining that the first subset of the reads have a first allele at the locus,

determining that the second subset of the reads have a second allele at the locus,

determining that the first subset of the reads corresponds to the first haplotype, and

determining that the second subset of the reads corresponds to the second haplotype.

6. The method of any one of claims 1 to 5, wherein the parameter is a count of cell-free DNA molecules, a size profile of cell-free DNA molecules, or a methylation level of cell-free DNA molecules.

7. The method of claim 6, wherein:

the parameter is the count of cell-free DNA molecules, and

the method further comprises:

determining that the fetus has a higher likelihood of inheriting the first haplotype than the second haplotype when the first value is greater than the second value.

8. The method of claim 6, wherein:

the parameter is the size profile of cell-free DNA molecules, and

the method further comprises:

determining that the fetus has a higher likelihood of inheriting the first haplotype than the second haplotype when the first value is less than the second value, indicating that the second set of cell-free DNA molecules is characterized by a smaller size profile than the third set of cell-free DNA molecules.

9. The method of claim 6, wherein:

the parameter is the methylation level of cell-free DNA molecules, and the method further comprises:

determining that the fetus has a higher likelihood of inheriting the first haplotype than the second haplotype when the first value is less than the second value.

10. The method of any one of claims 1 to 9, wherein the cutoff value is a first cutoff value, the method further comprising:

calculating a separation value using the first value and the second value;

comparing the separation value to a second cutoff value; and

determining a likelihood of a fetal aneuploidy based on the comparison of the separation value to the second cutoff value.

11. The method of claim 10, wherein:

the second cutoff value is determined from reference samples from pregnant females with euploid fetuses,

the second cutoff value is determined from reference samples from pregnant females with aneuploid fetuses, or

the second cutoff value is calculated assuming an aneuploid fetus.

12. The method of any one of claims 1 to 11, further comprising:

identifying a number of repeats of a subsequence in a read of the reads corresponding to the first set of cell-free DNA molecules,

wherein:

determining the sequence of the first haplotype comprises determining the number of repeats of the subsequence.

13. The method of claim 12, wherein:

the number of repeats of the subsequence is associated with a repeat-associated disease, and

the method further comprises determining a likelihood of the fetus inheriting the repeat-associated disease.

14. The method of claim 1, wherein:

the plurality of cell-free DNA molecules is enriched for sizes greater than or equal to the cutoff value relative to the biological sample, wherein over 20% of the cell-free DNA molecules in the biological sample have sizes greater than 200 nt.

15. The method of claim 14, further comprising:

enriching for the plurality of cell-free DNA molecules using electrophoresis.

16. The method of claim 14, further comprising:

enriching for the plurality of cell-free DNA molecules using magnetic beads to selectively bind cell-free DNA molecules based on size.

17. The method of claim 14, further comprising:

enriching for the plurality of cell-free DNA molecules using hybridization, immunoprecipitation, amplification, or CRISPR.

18. The method of claim 16, wherein enriching is for sizes greater than 600 nt, 700 nt, 800 nt, 900 nt, or 1 knt.

19. The method of claim 1, wherein the plurality of cell-free DNA molecules is enriched for a methylation profile relative to the biological sample,

the method further comprising:

enriching for the plurality of cell-free DNA molecules using immunoprecipitation.

20. A computer product comprising a non-transitory computer readable medium storing instructions that, when executed, control a computing system to perform a method of analyzing a biological sample obtained from a female pregnant with a fetus, the female having a first haplotype and a second haplotype in a first chromosomal region, the

biological sample including a plurality of cell-free DNA molecules from the fetus and the female, the method comprising:

receiving reads corresponding to the plurality of cell-free DNA molecules;

measuring sizes of the plurality of cell-free DNA molecules;

identifying a first set of cell-free DNA molecules from the plurality of cell-free DNA molecules as having sizes greater than or equal to a cutoff value, wherein the cutoff value is at least 500 nt;

determining a sequence of the first haplotype and a sequence of the second haplotype using reads corresponding to the first set of cell-free DNA molecules;

aligning a second set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the first haplotype, the second set of cell-free DNA molecules having sizes less than the cutoff value;

aligning a third set of cell-free DNA molecules from the plurality of cell-free DNA molecules to the sequence of the second haplotype, the third set of cell-free DNA molecules having sizes less than the cutoff value;

measuring a first value of a parameter using the second set of cell-free DNA molecules;

measuring a second value of the parameter using the third set of cell-free DNA molecules;

comparing the first value to the second value; and

determining a likelihood of the fetus inheriting the first haplotype based on the comparison of the first value to the second value.

21. The method of claim 1, wherein the reads corresponding to the plurality of cell-free DNA molecules are obtained by single molecule sequencing.

22. The method of claim 21, wherein the single molecule sequencing comprises optical monitoring of a DNA polymerase incorporating a new base into a complementary strand of a cell-free DNA molecule of the plurality of cell-free DNA molecules.