

BRIEF SUMMARY

[0004a] In one aspect, the present disclosure provides for a method of determining a classification of whether an individual has a disorder, the method comprising: receiving a first sample comprising a first set of double-stranded nucleic acid molecules derived from cell-free nucleic acid molecules in a biological sample obtained from the individual, wherein: one or more double-stranded nucleic acid molecules of the first set of double-stranded nucleic acid molecules each have one or more defects, and the one or more defects are present in the respective double-stranded nucleic acid molecule at a location at least one nucleotide away from a closest end of the respective double-stranded nucleic acid molecule; receiving a second sample comprising a second set of double-stranded nucleic acid molecules derived from the cell-free nucleic acid molecules in the biological sample; adding a first mixture comprising an enzyme to the first sample; repairing the one or more defects in each of the one or more double-stranded nucleic acid molecules of the first set of double-stranded nucleic acid molecules using the enzyme to produce a repaired first set of double-stranded nucleic acid molecules; determining a value of a parameter characterizing a difference in defects between the repaired first set of double-stranded nucleic acid molecules and the second set of double-stranded nucleic acid molecules; comparing the value of the parameter to a reference value, wherein the reference value is determined using one or more subjects identified to have the disorder or from one or more subjects identified to not have the disorder; and determining the classification of whether the individual has the disorder based on the comparison of the value of the parameter to the reference value.

[0004b] In another aspect, the present disclosure provides for a computer product comprising a computer readable medium storing a plurality of instructions for controlling a system to perform the method as described herein.

[0004c] In another aspect, the present disclosure provides for a system comprising: the computer product as described herein; and one or more processors for executing instructions stored on the computer readable medium.

[0004d] In another aspect, the present disclosure provides for a system comprising means for performing the method as described herein.

[0004e] In another aspect, the present disclosure provides for a system comprising one or more processors configured to perform the method as described herein.

[0004f] In another aspect, the present disclosure provides for a system comprising modules that respectively perform the method of as described herein.

[0005] ~~Examples~~ ~~Embodiments~~ of the present ~~invention~~ disclosure provide for improving the quality of cell-free DNA for analysis. Cell-free DNA may include DNA with defects that do not allow for analysis of those DNA with techniques such as sequencing and targeted capture enrichment. These defects may be defects within the strands of the DNA and not present at the ends of the DNA. ~~Embodiments~~ Examples of the present ~~invention~~ disclosure repair these intrastrand defects in cell-free DNA. The repair of the defects in cell-free DNA may then allow for these repaired cell-free DNA to be analyzed by techniques, including sequencing and targeted capture enrichment.

[0006] Repairing cell-free DNA may increase the amount of DNA available for analysis and/or may allow for longer DNA fragments to be analyzed. Because more DNA without defects would be available after repair, analysis of DNA would be possible in some cases where the amount of DNA without defects was not sufficient for analysis or sufficient to allow for a statistically significant conclusion from the analysis. Additionally, longer DNA fragments may be more prone to defects repairable by techniques described herein. Repairing the defects may then allow for more longer DNA fragments to be analyzed. Long DNA fragments allow for analysis of disorders and conditions with sequence patterns are repeated over a minimum number of loci.

[0007] The pattern or severity of defects in the cell-free DNA may be an indication of a disorder or a condition. A certain disorder, for example, may have more defects in cell-free DNA than defects in cell-free DNA from a subject without the disorder. The number or pattern of defects can then be used to provide a classification of whether an individual has a disorder.

[0008] A better understanding of the nature and advantages of ~~embodiment~~ examples of the present ~~invention~~ disclosure may be gained with reference to the following detailed description and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] **FIGS. 1A, 1B, and 1C** show different defect modes affecting DNA according to embodiments of the present invention.

[0010] **FIGS. 2A and 2B** show repair changes in the size profile of cfDNA in maternal plasma according to embodiments of the present invention.

WHAT IS CLAIMED IS:

1. ~~—A method of improving analysis of a first biological sample, the first biological sample including cell free nucleic acid molecules, the method comprising:~~
~~obtaining a plurality of double stranded nucleic acid molecules from the cell free nucleic acid molecules to produce a second biological sample, wherein:~~
~~one or more double stranded nucleic acid molecules of the plurality of double stranded nucleic acid molecules each have one or more defects, and~~
~~for each of the one or more double stranded nucleic acid molecules having a defect of the one or more defects:~~
~~the defect is present in the respective double stranded nucleic acid molecule at a location at least one nucleotide away from a closest end of the respective double stranded nucleic acid molecule;~~
~~adding a mixture comprising an enzyme to the second biological sample; and~~
~~repairing the one or more defects in each of the one or more double stranded nucleic acid molecules using the enzyme to produce a repaired set of double stranded nucleic acid molecules.~~

2. ~~—The method of claim 1, further comprising:~~
~~producing a sequencing library using the repaired set of double stranded nucleic acid molecules, and~~
~~detecting an aneuploidy or a sequence imbalance using the sequencing library.~~

3. ~~—The method of claim 2, further comprising:~~
~~determining whether a sequence imbalance exists in the second biological sample by:~~
~~obtaining a plurality of sequence reads from the repaired set of double stranded nucleic acid molecules,~~
~~analyzing, by a computer system, the plurality of sequence reads, wherein analyzing a sequence read includes:~~
~~identifying a location of the sequence read in a reference genome by aligning the sequence read to the reference genome,~~

10 determining, by a computer system, an amount of sequence reads from a
11 genomic region using the identified locations;
12 obtaining a value of a normalized parameter for the amount of sequence
13 reads from the genomic region;
14 comparing the value of the normalized parameter to a cutoff value;
15 determining whether the sequence imbalance exists based on the
16 comparison.

1 4. ~~—The method of claim 3, wherein the genomic region is a chromosome, and~~
2 ~~the sequence imbalance is an aneuploidy.~~

1 5. ~~—The method of claim 2, further comprising sequencing the sequencing~~
2 ~~library or performing targeted capture enrichment of the sequencing library to determine a set of~~
3 ~~reads.~~

1 6. ~~—The method of claim 5, wherein the set of reads does not comprise reads~~
2 ~~from any strand of the one or more double stranded nucleic acid molecules having one or more~~
3 ~~defects.~~

1 7. ~~—The method of claim 1, wherein the one or more defects are selected from~~
2 ~~the group consisting of a nick, gap, abasic site, thymidine dimer, oxidized pyrimidine,~~
3 ~~deaminated cytosine, blocked 3' end defect, or a combination thereof.~~

1 8. ~~—The method of claim 1, wherein the repaired set of double stranded~~
2 ~~nucleic acid molecules are free of a nick, gap, abasic site, thymidine dimer, oxidized pyrimidine,~~
3 ~~deaminated cytosine, blocked 3' end defect, or a combination thereof.~~

1 9. ~~—The method of claim 1, wherein the enzyme comprises at least one of a~~
2 ~~polymerase, a ligase, an endonuclease, or a glycosylase.~~

1 10. ~~—The method of claim 1, wherein the mixture comprises at least one of *Taq*~~
2 ~~DNA polymerase, *Bst* DNA polymerase, *Taq* DNA ligase, endonuclease VIII, endonuclease IV,~~
3 ~~T4 endonuclease V (PDG), 8-oxoguanine glycosylase (FPG), or uracil DNA glycosylase (UDG).~~

1 11.—The method of claim 1, wherein the mixture comprises *Taq* DNA
2 polymerase, *Bst* DNA polymerase, *Taq* DNA ligase, endonuclease VIII, endonuclease IV, T4
3 endonuclease V (PDG), 8-oxoguanine glycosylase (FPG), and uracil DNA glycosylase (UDG).

1 12.—The method of claim 1, wherein the first biological sample is maternal
2 plasma from a female subject pregnant with a fetus.

1 13.—The method of claim 1, wherein the one or more double-stranded nucleic
2 acid molecules have lengths in a range from 251 to 600 bp, from 251 to 450 bp, or from 451 bp
3 to 600 bp.

1 14.—The method of claim 1, wherein the one or more double-stranded nucleic
2 acid molecules have lengths associated with dinucleosomal nucleic acid molecules or
3 trinucleosomal nucleic acid molecules.

1 15.—The method of claim 1, wherein repairing the one or more defects in each
2 of the one or more double-stranded nucleic acid molecules comprises repairing a greater number
3 of nucleic acid molecules having lengths from 251 to 600 bp than nucleic acid molecules having
4 lengths from 0 to 250 bp.

1 16.—The method of claim 1, wherein:
2 the first biological sample is obtained from a female subject pregnant with a fetus;
3 the one or more double-stranded nucleic acid molecules comprise a plurality of
4 fetal-derived nucleic acid molecules;
5 after repairing the one or more defects, the second biological sample is
6 characterized by a fetal fraction calculated from reads obtained from analyzing the repaired set of
7 double-stranded nucleic acid molecules, and
8 the fetal fraction is greater than 0.05.

1 17.—The method of claim 1, further comprising:
2 performing blunt-end ligation of the plurality of double-stranded nucleic acid
3 molecules or of the repaired set of double-stranded nucleic acid molecules.

1 18.—The method of claim 1, further comprising:

~~haplotyping the HLA gene using the repaired set of double-stranded nucleic acid molecules.~~

~~19. The method of claim 1, further comprising:
haplotyping a monogenic disease by:
obtaining a plurality of sequence reads from the enriched set of double-stranded nucleic acid molecules;
aligning the plurality of sequence reads to a reference genome, and
identifying a mutation at a locus between two proximal SNP using the aligned plurality of sequence reads.~~

~~20. The method of claim 19, wherein the monogenic disease is congenital adrenal hyperplasia.~~

~~21.~~ A method of determining a classification of whether an individual has a disorder, the method comprising:
receiving a first sample comprising a first set of double-stranded nucleic acid molecules derived from cell-free nucleic acid molecules in a biological sample obtained from the individual, wherein:
one or more double-stranded nucleic acid molecules of the first set of double-stranded nucleic acid molecules each have one or more defects, and
the one or more defects are present in the respective double-stranded nucleic acid molecule at a location at least one nucleotide away from a closest end of the respective double-stranded nucleic acid molecule;
receiving a second sample comprising a second set of double-stranded nucleic acid molecules derived from the cell-free nucleic acid molecules in the biological sample;
adding a first mixture comprising an enzyme to the first sample;
repairing the one or more defects in each of the one or more double-stranded nucleic acid molecules of the first set of double-stranded nucleic acid molecules using the enzyme to produce a repaired first set of double-stranded nucleic acid molecules;

determining a value of a parameter characterizing a difference in defects between the repaired first set of double-stranded nucleic acid molecules and the second set of double-stranded nucleic acid molecules;

comparing the value of the parameter to a reference value, wherein the reference value is determined using one or more subjects identified to have the disorder or from one or more subjects identified to not have the disorder; and

determining the classification of whether the individual has the disorder based on the comparison of the value of the parameter to the reference value.

222. The method of claim 211, wherein:
the first mixture comprises a plurality of enzymes.

233. The method of claim 211, wherein the first sample and the second sample are equal in volume.

244. The method of claim 211, further comprising:
adding a second mixture to a third sample comprising a third set of double-stranded nucleic acid molecules derived from the cell-free nucleic acid molecules in the biological sample to produce the second sample comprising the second set of double-stranded nucleic acid molecules, the second mixture excluding the enzyme.

255. The method of claim 244, wherein:
the first mixture comprises a plurality of enzymes, and
the second mixture excludes the plurality of enzymes.

6. The method of claim 1, wherein the enzyme comprises at least one of a polymerase, a ligase, an endonuclease, or a glycosylase.

7. The method of claim 1, wherein the first mixture comprises at least one of *Taq* DNA polymerase, *Bst* DNA polymerase, *Taq* DNA ligase, endonuclease VIII, endonuclease IV, T4 endonuclease V (PDG), 8-oxoguanine glycosylase (FPG), or uracil-DNA glycosylase (UDG).

8. The method of claim 1, wherein the first mixture comprises *Taq* DNA polymerase, *Bst* DNA polymerase, *Taq* DNA ligase, endonuclease VIII, endonuclease IV, T4 endonuclease V (PDG), 8-oxoguanine glycosylase (FPG), and uracil-DNA glycosylase (UDG).

269. The method of claim 241, further comprising:
 sequencing or performing target capture enrichment of the repaired first set of double-stranded nucleic acid molecules to determine a first set of reads, and
 sequencing or performing targeted capture enrichment of the second set of double-stranded nucleic acid molecules to determine a second set of reads, wherein:
 the first set of reads does not comprise reads from double-stranded nucleic acid molecules having one or more defects,
 the second set of reads does not comprise reads from double-stranded nucleic acid molecules having one or more defects as a result of the sequencing or the performing target capture enrichment being unable to analyze the double-stranded nucleic acid molecules having one or more defects,
 the first set of reads has a first amount of reads,
 the second set of reads has a second amount of reads, and
 the value of the parameter is determined using the first amount of reads and the second amount of reads.

2710. The method of claim 269, further comprising:
 calculating a first statistical value characterizing the sizes of the first set of reads,
 calculating a second statistical value characterizing the sizes of the second set of reads,
 wherein:
 the value of the parameter is determined using the first statistical value and the second statistical value.

2811. The method of claim 269, wherein:
 the value of the parameter is determined using a difference between the first amount of reads and the second amount of reads, or

the value of the parameter is determined using a ratio of the first amount of reads and the second amount of reads.

2912. The method of claim 269, wherein:
the first amount of reads is a normalized amount of reads, and
the second amount of reads is a normalized amount of reads.

3013. The method of claim 241, wherein:
the parameter is multidimensional,
a first dimension of the parameter is based on a size of a nucleic acid molecule,
and
a second dimension of the parameter comprises an amount of the nucleic acid molecules.

3414. The method of claim 3013, wherein the first dimension is a range of sizes or a range of ratios of the size of the nucleic acid molecule to a reference size.

3215. The method of claim 3013, wherein:
the reference value is multidimensional, and
the reference value specifies amounts of nucleic acid molecules at different sizes of the nucleic acid molecules.

3316. The method of claim 241, wherein:
the parameter is multidimensional,
a first dimension of the parameter comprises a location in a reference genome of a nucleic acid molecule, and
a second dimension of the parameter comprises an amount of the nucleic acid molecule.

3417. The method of claim 241, wherein the reference value is determined using one or more subjects identified to have the disorder ~~or from one or more subjects identified to not have the disorder~~.

1 ~~35~~18. The method of claim ~~34~~17, wherein comparing the value of the parameter
2 to the reference value comprises determining whether the value of the parameter exceeds the
3 reference value.

~~36~~19. The method of claim ~~24~~1, wherein the disorder is systemic lupus
erythematosus, congenital adrenal hyperplasia (CAH), Fragile-X Syndrome, graft-versus-host
(GVHD) disease, a pregnancy-related disorder, an aneuploidy, or a sequence imbalance.

20. The method of claim 1, wherein the disorder is associated with cfDNA
damage.

21. The method of claim 1, wherein the disorder is an autoimmune disorder or
a sequence imbalance.

1
1 ~~37~~22. The method of claim ~~35~~18, wherein:
2 determining the classification is based on a difference between the value of the
3 parameter and the reference value, and
4 the classification comprises a severity of the disorder, the classification indicating
5 a more severe disorder when the difference is greater than a cutoff value.

1 ~~38~~23. A computer product comprising a computer readable medium storing a
2 plurality of instructions for controlling a system to perform any one of claims ~~24~~1 to ~~37~~22.

1 ~~39~~24. A system comprising:
2 the computer product of claim ~~38~~23; and
3 one or more processors for executing instructions stored on the computer readable
4 medium.

1 ~~40~~25. A system comprising means for performing any ~~of the above methods~~one
2 of claims 1 to 22.

1 ~~41~~26. A system comprising one or more processors configured to perform any ~~of~~
2 ~~the above methods~~one of claims 1 to 22.

- 1 ~~4227~~. A system comprising modules that respectively perform the ~~steps~~method
- 2 of any ~~of the above methods~~one of claims 1 to 22.