

of exposure to environmental mutagens, to monitor endogenous DNA repair, and to study the accumulation of somatic mutations in aging individuals. Additionally, more sensitive methods to detect rare variants can enhance prenatal diagnosis, enabling the characterization of fetal cells present in maternal blood.

SUMMARY OF THE INVENTION

[0005] In view of the foregoing, there is a need for improved methods of detecting rare sequence variants. The compositions and methods of the present disclosure address this need, and provide additional advantages as well. In particular, the various aspects of the disclosure provide for highly sensitive detection of rare or low frequency nucleic acid sequence variants (sometimes referred to as mutations). This includes identification and elucidation of low frequency nucleic acid variations (including substitutions, insertions and deletions) in samples that may contain low amounts of variant sequences in a background of normal sequences, as well as the identification of low frequency variations in a background of sequencing errors.

[0005a] In one aspect, the disclosure provides a method for processing or analyzing a biological sample of a subject, comprising: (a) obtaining said biological sample of said subject; (b) using a ribonucleic acid (RNA) molecule or derivative thereof derived from said biological sample to generate a circular RNA molecule; (c) subjecting said circular RNA molecule or derivative thereof to reverse transcription to generate a complementary deoxyribonucleic acid (cDNA) concatemer; (d) creating an amplicon of a target sequence of said cDNA molecule using linear amplification to amplify said cDNA concatemer or derivative thereof using a primer that comprises (i) a sequence that binds to said target sequence or a sequence that binds to a complement of a junction sequence between a 5' end and a 3' end of said RNA molecule in said circular RNA molecule comprising said target sequence; and (ii) an adaptor sequence; and (e) subjecting said amplicon of said cDNA concatemer or derivative thereof to sequencing to identify a sequence of said RNA molecule.

[0006] In one aspect, the disclosure provides a method of identifying a sequence variant, such as in a nucleic acid sample. In some embodiments, each polynucleotide of the plurality has a 5' end and a 3' end, and the method comprises: (a) circularizing individual polynucleotides of said plurality to form a plurality of circular polynucleotides, each of which having a junction between the 5' end and 3' end; (b) amplifying the circular polynucleotides of (a); (c) sequencing the amplified polynucleotides to produce a plurality of sequencing reads; (d) identifying sequence differences between sequencing reads and a reference sequence; and (e) calling a sequence difference that occurs in at least two circular polynucleotides having different junctions as the sequence variant. In some embodiments, the method comprises identifying sequence differences between sequencing reads and a reference sequence, and calling a sequence difference that occurs in at least two circular polynucleotides having different junctions as the sequence variant, wherein: (a) the sequencing reads correspond to amplification products of the at least two circular polynucleotides; and (b) each of the at least two

circular polynucleotides comprises a different junction formed by ligating a 5' end and 3' end of the respective polynucleotides.

[0007] The plurality of polynucleotides can be single- or double-stranded. In some embodiments, the polynucleotides are single-stranded. In some embodiments, circularizing is effected by subjecting the plurality of polynucleotides to a ligation reaction. In some embodiments, an individual circular polynucleotide has a junction that is unique among the circularized polynucleotides. In some embodiments, the sequence variant is a single nucleotide polymorphism (SNP). In some embodiments, the reference sequence is a consensus sequence formed by aligning the sequence reads with one another. In some embodiments, the reference sequence is a known reference sequence, such as a reference genome or portion thereof. In some embodiments, circularizing comprises the step of joining an adapter polynucleotide to the 5' end, the 3' end, or both the 5' end and the 3' end of a

CLAIMS

WHAT IS CLAIMED IS:

1. ~~———— A method of identifying a sequence variant in a nucleic acid sample comprising a plurality of polynucleotides, each polynucleotide of the plurality having a 5' end and a 3' end, the method comprising:~~

~~———— (a) circularizing individual polynucleotides of said plurality to form a plurality of circular polynucleotides, each of which having a junction between the 5' end and 3' end;~~

~~———— (b) amplifying the circular polynucleotides of (a);~~

~~———— (c) sequencing the amplified polynucleotides to produce a plurality of sequencing reads;~~

~~———— (d) identifying sequence differences between sequencing reads and a reference sequence; and~~

~~———— (e) calling a sequence difference that occurs in at least two circular polynucleotides having different junctions as the sequence variant.~~

2. ~~———— A method of identifying a sequence variant, the method comprising: identifying sequence differences between sequencing reads and a reference sequence, and calling a sequence difference that occurs in at least two circular polynucleotides having different junctions as the sequence variant, wherein:~~

~~———— (a) the sequencing reads correspond to amplification products of the at least two circular polynucleotides; and~~

~~———— (b) each of the at least two circular polynucleotides comprises a different junction formed by ligating a 5' end and 3' end of the respective polynucleotides.~~

3. ~~———— The method of claim 1 or 2, wherein the plurality of polynucleotides are single-stranded.~~

4. ~~———— The method of claim 1 or 2, wherein circularizing is effected by subjecting the plurality of polynucleotides to a ligation reaction.~~

5. ~~———— The method of claim 1 or 2, wherein an individual circular polynucleotide has a junction that is unique among the circularized polynucleotides.~~

6. ~~———— The method of claim 1 or 2, wherein the sequence variant is a single nucleotide polymorphism.~~

7. ~~———— The method of claim 1 or 2, wherein the reference sequence is a consensus sequence formed by aligning the sequencing reads with one another.~~

8. ~~———— The method of claim 1 or 2, wherein the reference sequence is a known reference sequence.~~

~~9. The method of claim 1, wherein circularizing comprises the step of joining an adapter polynucleotide to the 5' end, the 3' end, or both the 5' end and the 3' end of a polynucleotide in the plurality of polynucleotides.~~

~~10. The method of claim 1, wherein amplifying is effected by using a polymerase having strand displacement activity.~~

~~11. The method of claim 1, wherein amplifying comprises subjecting the circular polynucleotides to an amplification reaction mixture comprising random primers.~~

~~12. The method of claim 1, wherein amplifying comprises subjecting the circular polynucleotides to an amplification reaction mixture comprising one or more primers, each of which specifically hybridizes to a different target sequence via sequence complementarity.~~

~~13. The method of claim 11 or 12, wherein the amplified polynucleotides are subjected to the sequencing step without enrichment.~~

~~14. The method of claim 11 or 12, further comprising enriching one or more target polynucleotides among the amplified polynucleotides by performing an enrichment step prior to sequencing.~~

~~15. The method of claim 14, wherein the enrichment step comprises hybridizing amplified polynucleotides to a plurality of probes attached to a substrate.~~

~~16. The method of claim 14, wherein the enrichment step comprises amplifying a target sequence comprising sequence A and sequence B oriented in a 5' to 3' direction in an amplification reaction mixture comprising:~~

~~—— (a) the amplified polynucleotides;~~

~~—— (b) a first primer comprising sequence A', wherein the first primer specifically hybridizes to sequence A of the target sequence via sequence complementarity between sequence A and sequence A';~~

~~—— (c) a second primer comprising sequence B, wherein the second primer specifically hybridizes to sequence B' present in a complementary polynucleotide comprising a complement of the target sequence via sequence complementarity between B and B'; and~~

~~—— (d) a polymerase that extends the first primer and the second primer to produce amplified polynucleotides;~~

~~—— wherein the distance between the 5' end of sequence A and the 3' end of sequence B of the target sequence is 75nt or less.~~

~~17. The method of claim 1 or 2, wherein a microbial contaminant is identified based on the calling step.~~

~~18. — A method of identifying a sequence variant in a nucleic acid sample comprising less than 50 ng of polynucleotides, each polynucleotide having a 5' end and a 3' end, the method comprising:~~

- ~~—— (a) circularizing with a ligase individual polynucleotides in said sample to form a plurality of circular polynucleotides;~~
- ~~—— (b) upon separating said ligase from said circular polynucleotides, amplifying the circular polynucleotides to form concatemers;~~
- ~~—— (c) sequencing the concatemers to produce a plurality of sequencing reads;~~
- ~~—— (d) identifying sequence differences between the plurality of sequencing reads and a reference sequence; and~~
- ~~—— (e) calling a sequence difference that occurs with a frequency of 0.05% or higher in said plurality of reads from said nucleic acid sample of less than 50ng polynucleotides as the sequence variant.~~

~~19. — The method of claim 18, wherein the polynucleotides of the sample are single-stranded.~~

~~20. — The method of claim 18, wherein an individual circular polynucleotide has a junction that is unique among the circularized polynucleotides.~~

~~21. — The method of claim 18, wherein the sequence variant is a single nucleotide polymorphism.~~

~~22. — The method of claim 18, wherein the reference sequence is a consensus sequence formed by aligning the sequencing reads with one another.~~

~~23. — The method of claim 18, wherein the reference sequence is a known reference sequence.~~

~~24. — The method of claim 18, wherein amplifying is effected by using a polymerase having strand displacement activity.~~

~~25. — The method of claim 18, wherein amplifying comprises subjecting the circular polynucleotides to an amplification reaction mixture comprising random primers.~~

~~26. — The method of claim 18, wherein amplifying comprises subjecting the circular polynucleotides to an amplification reaction mixture comprising one or more primers, each of which specifically hybridizes to a different target sequence via sequence complementarity.~~

~~27. — A method of amplifying in a reaction mixture a plurality of different concatemers comprising two or more copies of a target sequence, wherein the target sequence comprises sequence A and sequence B oriented in a 5' to 3' direction, the method comprising subjecting the reaction mixture to a nucleic acid amplification reaction, wherein the reaction mixture comprises:~~

———(a) the plurality of concatemers, wherein individual concatemers in the plurality comprise different junctions formed by circularizing individual polynucleotides having a 5' end and a 3' end;

———(b) a first primer comprising sequence A', wherein the first primer specifically hybridizes to sequence A of the target sequence via sequence complementarity between sequence A and sequence A';

———(c) a second primer comprising sequence B, wherein the second primer specifically hybridizes to sequence B' present in a complementary polynucleotide comprising a complement of the target sequence via sequence complementarity between sequence B and B'; and

———(d) a polymerase that extends the first primer and the second primer to produce amplified polynucleotides;

———wherein the distance between the 5' end of sequence A and the 3' end of sequence B of the target sequence is 75nt or less.

28. — A method of amplifying in a reaction mixture a plurality of different circular polynucleotides comprising a target sequence, wherein the target sequence comprises sequence A and sequence B oriented in a 5' to 3' direction, the method comprising subjecting the reaction mixture to a nucleic acid amplification reaction, wherein the reaction mixture comprises:

———(a) the plurality of circular polynucleotides, wherein individual circular polynucleotides in the plurality comprise different junctions formed by circularizing individual polynucleotides having a 5' end and a 3' end;

———(b) a first primer comprising sequence A', wherein the first primer specifically hybridizes to sequence A of the target sequence via sequence complementarity between sequence A and sequence A';

———(c) a second primer comprising sequence B, wherein the second primer specifically hybridizes to sequence B' present in a complementary polynucleotide comprising a complement of the target sequence via sequence complementarity between sequence B and B'; and

———(d) a polymerase that extends the first primer and the second primer to produce amplified polynucleotides;

———wherein sequence A and sequence B are endogenous sequences, and the distance between the 5' end of sequence A and the 3' end of sequence B of the target sequence is 75nt or less.

29. — The method of claim 27, wherein the first primer comprises sequence C 5' with respect to sequence A', the second primer comprises sequence D 5' with respect to sequence B, and neither sequence C nor sequence D hybridize to the plurality of concatemers during a first amplification phase at a first hybridization temperature.

~~30. — The method of claim 29, wherein amplification comprises a first phase and a second phase; the first phase comprises a hybridization step at a first temperature, during which the first and second primer hybridize to the concatemers prior to primer extension; and the second phase comprises a hybridization step at a second temperature that is higher than the first temperature, during which the first and second primers hybridize to amplification products comprising extended first or second primers or complements thereof.~~

~~31. — The method of claim 30, wherein after 5 cycles of hybridization at the second temperature and primer extension, at least 5% of amplified polynucleotides in the reaction mixture comprise two or more copies of the target sequence.~~

~~32. — The method of claim 28, wherein the first primer comprises sequence C 5' with respect to sequence A', the second primer comprises sequence D 5' with respect to sequence B, and neither sequence C nor sequence D hybridize to the plurality of circular polynucleotides during a first amplification phase at a first hybridization temperature.~~

~~33. — The method of claim 32, wherein amplification comprises a first phase and a second phase; the first phase comprises a hybridization step at a first temperature, during which the first and second primer hybridize to the circular polynucleotides or amplification products thereof prior to primer extension; and the second phase comprises a hybridization step at a second temperature that is higher than the first temperature, during which the first and second primers hybridize to amplification products comprising extended first or second primers or complements thereof.~~

~~34. — A reaction mixture for amplifying a plurality of different concatemers comprising two or more copies of a target sequence, wherein the target sequence comprises sequence A and sequence B oriented in a 5' to 3' direction, the reaction mixture comprising:~~

~~—— (a) the plurality of concatemers, wherein individual concatemers in the plurality comprise different junctions formed by circularizing individual polynucleotides having a 5' end and a 3' end;~~

~~—— (b) a first primer comprising sequence A', wherein the first primer specifically hybridizes to sequence A of the target sequence via sequence complementarity between sequence A and sequence A';~~

~~—— (c) a second primer comprising sequence B, wherein the second primer specifically hybridizes to sequence B' present in a complementary polynucleotide comprising a complement of the target sequence via sequence complementarity between B and B'; and~~

~~—— (d) a polymerase that extends the first primer and the second primer to produce amplified polynucleotides;~~

~~—— wherein the distance between the 5' end of sequence A and the 3' end of sequence B of the target sequence is 75nt or less.~~

~~35. — The reaction mixture of claim 34, wherein the first primer comprises sequence C 5' with respect to sequence A', the second primer comprises sequence D 5' with respect to sequence B, and neither sequence C nor sequence D hybridizes to the two or more concatemers during a first amplification step in an amplification reaction.~~

~~36. A composition comprising a plurality of circularized polynucleotides that are single-stranded and substantially free of ligase.~~

~~37. A composition comprising a plurality of concatemers, wherein the plurality of concatemers correspond to a group of 10000 or fewer target polynucleotides, and further wherein individual concatemers in the plurality are characterized in that:~~

~~(a) they comprise two or more copies of a sequence repeat, wherein all of said copies correspond to the same target polynucleotide; and~~

~~(b) the junction between the two or more copies of the sequence repeat of one individual concatemer is different from that of another individual concatemer in said composition.~~

~~38. — A system for detecting a sequence variant, the system comprising:~~

~~(a) a computer configured to receive a user request to perform a detection reaction on a sample;~~

~~(b) an amplification system that performs a nucleic acid amplification reaction on the sample or a portion thereof in response to the user request, wherein the amplification reaction comprises the steps of (i) circularizing individual polynucleotides to form a plurality of circular polynucleotides, each of which having a junction between the 5' end and 3' end; and (ii) amplifying the circular polynucleotides;~~

~~(c) a sequencing system that generates sequencing reads for polynucleotides amplified by the amplification system, identifies sequence differences between sequencing reads and a reference sequence, and calls a sequence difference that occurs in at least two circular polynucleotides having different junctions as the sequence variant; and~~

~~(d) a report generator that sends a report to a recipient, wherein the report contains results for detection of the sequence variant.~~

~~39. — The system of claim 38, wherein the recipient is the user.~~

~~40. — A computer-readable medium comprising codes that, upon execution by one or more processors, implement a method of detecting a sequence variant, the method comprising:~~

~~(a) receiving a customer request to perform a detection reaction on a sample;~~

~~(b) performing a nucleic acid amplification reaction on the sample or a portion thereof in response to the customer request, wherein the amplification reaction comprises the~~

steps of (i) circularizing individual polynucleotides to form a plurality of circular polynucleotides, each of which having a junction between the 5' end and 3' end; and (ii) amplifying the circular polynucleotides;

(c) performing a sequencing analysis comprising the steps of (i) generating sequencing reads for polynucleotides amplified in the amplification reaction; (ii) identifying sequence differences between sequencing reads and a reference sequence; and (iii) calling a sequence difference that occurs in at least two circular polynucleotides having different junctions as the sequence variant; and

(d) generating a report that contains results for detection of the sequence variant.

41. — The method of claim 1 or 2, wherein the sample is a sample from a subject.

42. — The method of claim 41, wherein the sample is urine, stool, blood, saliva, tissue, or bodily fluid.

43. — The method of claim 41, wherein the sample comprises tumor cells.

44. — The method of claim 41, wherein the sample is a formalin-fixed paraffin-embedded (FFPE) sample.

45. — The method of claim 41, further comprising diagnosing said subject based on the calling step.

46. — The method of claim 1 or 2, wherein the sequence variant is a causal genetic variant.

47. — The method of claim 1 or 2, wherein the sequence variant is associated with a type or stage of cancer.

48. — The method of claim 1 or 2, wherein the plurality of polynucleotides comprise cell-free polynucleotides.

49. — The method of claim 48, wherein the cell-free polynucleotides comprise circulating tumor DNA.

50. — The method of claim 18, wherein the polynucleotides of the sample comprise cell-free polynucleotides.

51. — The method of claim 18, wherein the sequence variant is a causal genetic variant.

52. — The composition of claim 36, wherein the polynucleotides are cell-free DNA.

1. A method for processing or analyzing a biological sample of a subject, comprising:

(a) obtaining said biological sample of said subject;

(b) using a ribonucleic acid (RNA) molecule or derivative thereof derived from said biological sample to generate a circular RNA molecule;

(c) subjecting said circular RNA molecule or derivative thereof to reverse transcription to generate a complementary deoxyribonucleic acid (cDNA) concatemer;

(d) creating an amplicon of a target sequence of said cDNA molecule using linear amplification to amplify said cDNA concatemer or derivative thereof using a primer that

comprises (i) a sequence that binds to said target sequence or a sequence that binds to a complement of a junction sequence between a 5' end and a 3' end of said RNA molecule in said circular RNA molecule comprising said target sequence; and (ii) an adaptor sequence; and

(e) subjecting said amplicon of said cDNA concatemer or derivative thereof to sequencing to identify a sequence of said RNA molecule.

2. The method of claim 1, wherein said biological sample is a cell-free biological sample.
3. The method of claim 1, further comprising, subsequent to (c), capturing said cDNA concatemer or derivative thereof using a capture probe.
4. The method of claim 3, wherein said capture probe comprises one or more copies of a target sequence of said cDNA concatemer.
5. The method of claim 1, wherein said adaptor sequence comprises a molecular barcode sequence.
6. The method of claim 5, wherein said primer further comprises a first sequencing adaptor sequence.
7. The method of claim 5, further comprising subjecting said amplicon to a primer extension reaction using a first primer comprising a first sequencing adaptor sequence and a second primer comprising a second sequencing adaptor sequence.
8. The method of claim 1, wherein said reverse transcription of (c) is performed with a primer comprising: (i) a sequence that binds to a target sequence of said circular RNA molecule; and (ii) a molecular barcode sequence.
9. The method of claim 1, wherein said reverse transcription of (c) is performed with a primer comprising (i) a random sequence that binds to said circular RNA molecule; and (ii) a molecular barcode sequence.
10. The method of claim 1, wherein said reverse transcription of (c) is performed with a primer comprising (i) a sequence that binds to a junction sequence between a 5' end and a 3' end of said RNA molecule in said circular RNA molecule; and (ii) a molecular barcode sequence.

11. The method of claim 1, wherein said RNA molecule is a micro-RNA (miRNA).
12. The method of claim 1, wherein said sequence comprises a junction sequence in which a 5' end and a 3' end of said RNA molecule are joined.
13. The method of claim 1, wherein said sequence identifies said 5' end or said 3' end of said RNA molecule.
14. The method of claim 1, wherein (b) comprises ligating ends of said RNA molecule or derivative thereof to one another.
15. The method of claim 1, wherein said sequencing comprises (i) bringing said amplicon of said cDNA concatemer or derivative thereof in contact with a plurality of nucleotides in presence of a polymerase to incorporate one or more nucleotides of said plurality of nucleotides into a growing strand complementary to a strand of said amplicon of said cDNA concatemer or derivative thereof, and (ii) detecting one or more signals indicative of incorporation of said one or more nucleotides into said growing strand.
16. The method of claim 1, wherein said RNA molecule is from a tumor.
17. The method of claim 1, further comprising processing said sequence against a reference.
18. The method of claim 1, further comprising adding a molecular barcode sequence to (i) said RNA molecule by ligation prior to circularization; (ii) said cDNA concatemer during reverse transcription; or (iii) a reverse complement strand of said cDNA concatemer during synthesis of a strand complementary to said cDNA concatemer.
19. The method of claim 18, further comprising processing said sequence to identify a consensus sequence when said sequence occurs in (i) a majority of repeats or at least two repeats of said concatemer; and/or (ii) at least two circular RNA molecules or derivatives thereof having different molecular barcode sequences.
20. The method of claim 18, further comprising quantifying said RNA molecule by counting sequencing reads from circular RNA molecules having different molecular barcodes.

21. The method of claim 1, further comprising processing said sequence to identify a consensus sequence when said sequence occurs in (i) at least two copies of said RNA molecule in said cDNA concatemer and/or (ii) a majority of copies of said RNA molecule in said cDNA concatemer.
22. The method of claim 1, further comprising comparing a consensus sequence of said RNA molecule to a reference RNA sequence to identify a sequence difference.
23. The method of claim 1, further comprising processing said sequence to identify a disease in said subject.
24. The method of claim 23, wherein said disease is cancer.
25. The method of claim 24, wherein said sequence is computer processed against a reference to identify one or more genetic variants, and wherein said one or more genetic variants identify said cancer.
26. The method of claim 1, further comprising processing said sequence to monitor a progression of a disease in said subject in response to treatment.
27. The method of claim 1, wherein said biological sample comprises a bodily fluid.
28. The method of claim 27, wherein said bodily fluid is urine, saliva, blood, serum, or plasma.