

METHODS AND COMPOSITIONS FOR NUCLEIC ACID ANALYSIS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of United States Provisional Application No. 62/263,532, filed December 4, 2015, which is hereby incorporated by reference in its entirety for all purposes.

BACKGROUND OF THE INVENTION

[0002] Polynucleotide sequencing continues to find increasing use in medical applications such as genetic screening and genotyping of tumors. Many polynucleotide sequencing methods rely on sample processing techniques of the original sample, including random fragmentation of polynucleotides. These processing techniques can provide advantages in terms of throughput and efficiency, but the resultant sequence information obtained from these processed samples can lack important contextual information in terms of the location of particular sequences within the broader linear (two-dimensional) sequence of the original nucleic acid molecule that contained those sequences. Structural context within the three dimensional space of the original sample is also lost with many sample processing and sequencing techniques. There is thus a need for sequencing technologies that retain structural and molecular context of the identified nucleic acid sequences.

SUMMARY

[0003] Accordingly, the present disclosure provides methods, systems and compositions for providing sequence information that retains both molecular and structural context of the originating nucleic acid molecule.

[0004] In some examples, the present disclosure provides methods of analyzing nucleic acids while maintaining structural context. Such methods include the steps of: (a) providing a sample containing nucleic acids, where the nucleic acids comprise three dimensional structures; (b) separating portions of the sample into discrete partitions such that portions of the nucleic acid three dimensional structures are also separated into the discrete partitions; (c) obtaining sequence information from the nucleic acids, thereby analyzing nucleic acids while maintaining structural context.

[0005] In some examples, the sequence information from obtaining step (c) includes identification of nucleic acids that are in spatial proximity to each other.

[0006] In further examples, the sequence information from obtaining step (c) includes identification of nucleic acids that are in spatial proximity to each other.

[0007] In still further examples, the obtaining step (c) provides information on intrachromosomal and/or interchromosomal interactions between genomic loci.

[0008] In yet further examples, the obtaining step (c) provides information on chromosome conformations.

[0009] In further examples, prior to separating step (b), at least some of the three dimensional structures are processed to link different portions of the nucleic acids that are in proximity to each other within the three dimensional structures.

[0010] In any examples, the nucleic acids are not isolated from the sample prior to the separating step (b).

[0011] In any examples, prior to the obtaining step (c), the nucleic acids within the discrete partitions are barcoded to form a plurality of barcoded fragments, where fragments within a given discrete partition each comprise a common barcode, such that the barcodes identify nucleic acids from a given partition.

[0012] In further examples, the obtaining step (c) comprises a sequencing reaction selected from the group consisting of: short read-length sequencing reactions and long read-length sequencing reactions.

[0013] In some examples, the present disclosure provides methods of analyzing nucleic acids while maintaining structural context that include the steps of (a) forming linked nucleic acids within the sample such that spatially adjacent nucleic acid segments are linked; (b) processing the linked nucleic acids to produce a plurality of ligation products, wherein the ligation products contain portions of the spatially adjacent nucleic acid segments; (c) depositing the plurality of ligation products into discrete partitions; (d) barcoding the ligation products within the discrete partitions to form a plurality of barcoded fragments, wherein fragments within a given discrete partition each comprise a common barcode, thereby associating each fragment with the linked nucleic acid from which it is derived; (e) obtaining sequence information from the plurality of barcoded fragments, thereby analyzing nucleic acids from the sample while maintaining structural context.

[0014] In some examples, the present disclosure provides methods of analyzing nucleic acids while maintaining structural context that include the steps of: (a) forming

linked nucleic acids within the sample such that spatially adjacent nucleic acid segments are linked; (b) depositing the linked nucleic acids into discrete partitions; (c) processing the linked nucleic acids to produce a plurality of ligation products, wherein the ligation products contain portions of the spatially adjacent nucleic acid segments; (d) barcoding the ligation products within the discrete partitions to form a plurality of barcoded fragments, wherein fragments within a given discrete partition each comprise a common barcode, thereby associating each fragment with the linked nucleic acid from which it is derived; (e) obtaining sequence information from the plurality of barcoded fragments, thereby analyzing nucleic acids from the sample while maintaining structural context.

[0015] In some examples, the present disclosure provides methods of analyzing nucleic acids while maintaining structural context that include the steps of (a) cross-linking nucleic acids within the sample to form cross-linked nucleic acids, wherein the cross-linking forms covalent links between spatially adjacent nucleic acid segments; (b) depositing the cross-linked nucleic acids into discrete partitions; (c) processing the cross-linked nucleic acids to produce a plurality of ligation products, wherein the ligation products contain portions of the spatially adjacent nucleic acid segments; (d) obtaining sequence information from the plurality of ligation products, thereby analyzing nucleic acids from the sample while maintaining structural context.

[0016] In any examples, the sample is a formalin-fixed paraffin sample.

[0017] In any examples, the discrete partitions comprise beads. In further examples, the beads are gel beads.

[0018] In any examples, the sample comprises a tumor sample.

[0019] In any examples, the sample comprises a mixture of tumor and normal cells.

[0020] In any examples, the sample comprises a nuclear matrix.

[0021] In any examples, the nucleic acids comprise RNA.

[0022] In any examples, the amount of nucleic acids in the sample is less than 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 ng/ml.

[0023] In some examples, the present disclosure provides a method of analyzing nucleic acids while maintaining structural context, in which the method includes the steps of: (a) providing a sample that contains nucleic acids; (b) applying a library of tags to the sample such that different geographical regions of the sample receive different tags or different concentrations of tags; (c) separating portions of the sample into discrete

partitions such that portions of the library of tags and portions of the nucleic acids are also separated into the discrete partitions; (d) obtaining sequence information from the nucleic acids, and (e) identifying tags or concentrations of tags in the discrete partitions, thereby analyzing nucleic acids while maintaining structural context.

[0023a] In one aspect, the present disclosure refers to a method of analyzing a plurality of nucleic acids in a fixed sample, the method comprising:

- (a) providing a fixed sample comprising a plurality of nucleic acids;
- (b) applying a plurality of geographical tags to the fixed sample such that different regions of the sample receive different geographical tags or different concentrations of geographical tags;
- (c) partitioning the fixed sample into discrete partitions, each partition comprising a portion of nucleic acids from the plurality of nucleic acids and a portion of geographical tag from the plurality of geographical tags;
- (d) obtaining sequencing information from the nucleic acids in each of the discrete partitions; and
- (e) identifying a characteristic of the geographical tags in each of the discrete partitions, wherein for each partition, the characteristic of the geographical tag provides geographic context for the portion of nucleic acids in the fixed sample, thereby analyzing the plurality of nucleic acids.

BRIEF DESCRIPTION OF THE DRAWINGS

[0024] FIGURE 1 provides a schematic illustration of molecular context and structural context in accordance with the methods described herein.

[0025] FIGURE 2 provides a schematic illustration of a process described herein.

[0026] FIGURE 3 illustrates a typical workflow for performing an assay to detect sequence information, using the methods and compositions disclosed herein.

[0027] FIGURE 4 provides a schematic illustration of a process for combining a nucleic acid sample with beads and partitioning the nucleic acids and beads into discrete droplets.

[0028] FIGURE 5 provides a schematic illustration of a process for barcoding and amplification of chromosomal nucleic acid fragments.

[0029] FIGURE 6 provides a schematic illustration of the use of barcoding of nucleic acid fragments in attributing sequence data to their originating source nucleic acid molecule.

[0030] FIGURE 7 provides a schematic illustration of an exemplary sample preparation method.

DETAILED DESCRIPTION OF THE INVENTION

[0031] The practice of the present invention may employ, unless otherwise indicated, conventional techniques and descriptions of organic chemistry, polymer technology, molecular biology (including recombinant techniques), cell biology, biochemistry, and immunology, which are within the skill of the art. Such conventional techniques include polymer array synthesis, hybridization, ligation, phage display, and detection of hybridization using a label. Specific illustrations of suitable techniques can be had by reference to the example herein below. However, other equivalent conventional procedures can, of course, also be used. Such conventional techniques and descriptions can be found in standard laboratory manuals such as *Genome Analysis: A Laboratory Manual Series* (Vols. I-IV), *Using Antibodies: A Laboratory Manual*, *Cells: A Laboratory Manual*, *PCR Primer: A Laboratory Manual*, and *Molecular Cloning: A*

CLAIMS

What is claimed:

1. A method of analyzing a plurality of nucleic acids in a fixed sample, the method comprising:
 - (a) providing a fixed sample comprising a plurality of nucleic acids;
 - (b) applying a plurality of geographical tags to the fixed sample such that different regions of the sample receive different geographical tags or different concentrations of geographical tags;
 - (c) partitioning the fixed sample into discrete partitions, each partition comprising a portion of nucleic acids from the plurality of nucleic acids and a portion of geographical tag from the plurality of geographical tags;
 - (d) obtaining sequencing information from the nucleic acids in each of the discrete partitions;and
 - (e) identifying a characteristic of the geographical tags in each of the discrete partitions, wherein for each partition, the characteristic of the geographical tag provides geographic context for the portion of nucleic acids in the fixed sample,thereby analyzing the plurality of nucleic acids.
2. The method of claim 1, wherein a plurality of tagged nucleic acid fragments are produced in each discrete partition prior to the (d) obtaining.
3. The method of claim 2, wherein each of the tagged nucleic acid fragments in each discrete partition comprises a copy of a fragment of the portion of nucleic acids in the partition and a partition-specific tag.
4. The method of claim 3, wherein the partition-specific tag is an oligonucleotide barcode.
5. The method of claim 2, wherein each of the tagged nucleic acid fragments in a discrete partition comprise the same partition-specific tag.
6. The method of claim 2, wherein the sequence information obtained in (d) is obtained by sequencing the plurality of tagged nucleic acid fragments in each partition.
7. The method of claim 1, wherein each partition further comprises a tagged particle, the tagged particle comprising a plurality of partition-specific tags releaseably attached to a particle.

8. The method of claim 7, wherein the partition-specific tags each comprise an oligonucleotide barcode tag and a random n-mer oligonucleotide.
9. The method of claim 7, wherein the partition-specific tags in a discrete partition are identical and the partition-specific tags between discrete partitions are different.
10. The method of claim 8, wherein the partition-specific tag in each discrete partition comprises the same oligonucleotide barcode and a different random n-mer oligonucleotide.
11. The method of claim 7, wherein the particle is a bead.
12. The method of claim 1, wherein the geographical tags are applied to the sample in step (b) such that different regions of the sample tag receive different concentrations of geographical tags.
13. The method of claim 12, wherein the characteristic identified in (e) is the concentration of tags in each of the discrete partitions.
14. The method of claim 1, wherein the geographical tags are applied to the sample in step (b) such that different regions of the sample tag receive different geographical tags.
15. The method of claim 12, wherein the characteristic identified in (e) is the sequence of the geographical tags in each of the discrete partitions.
16. The method of claim 1, wherein the geographical tags comprise oligonucleotide barcodes.
17. The method of claim 1, wherein the fixed sample is a formalin-fixed paraffin sample.
18. The method of claim 17, wherein the fixed sample is deparaffinized prior to the (d) obtaining.
19. The method of claim 1, wherein the amount of nucleic acid in each discrete partition is less than 10 ng/ml.

20. The method of claim 1, wherein the fixed sample is a fixed tumor sample.
21. The method of claim 1, wherein the geographical tags are attached to particles.
22. The method of claim 1, wherein the obtaining (d) comprises a sequencing reaction selected from the group consisting of: short read-length sequencing reactions and long read-length sequencing reactions.