

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.

[0012a] In one aspect, the present disclosure refers to a method of analyzing nucleic acids obtained from a formalin-fixed paraffin embedded (FFPE) tissue sample while retaining spatial context, comprising: a) partitioning the nucleic acids obtained from the FFPE tissue sample into a plurality of wells; wherein nucleic acids in spatial proximity to each other in the FFPE tissue sample are introduced into the same well; b) barcoding partitioned nucleic acids to form a plurality of barcoded nucleic acids, wherein barcoded nucleic acids within a given discrete well each comprise a common partition-specific barcode sequence, such that the barcode sequences identify nucleic acids from a given well; c) obtaining sequence information from the plurality of barcoded nucleic acids, wherein the sequence information from the plurality of barcoded nucleic acids comprises sequence information of the partition-specific barcode sequence; and d) attributing the plurality of barcoded nucleic acids to a region of spatial proximity, wherein barcoded nucleic acids derived from a region of spatial proximity in the FFPE tissue sample comprise the same partition-specific barcode sequence.

[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

CLAIMS

What is claimed:

1. A method of analyzing nucleic acids obtained from a formalin-fixed paraffin embedded (FFPE) tissue sample while retaining spatial context, comprising:
 - a) partitioning the nucleic acids obtained from the FFPE tissue sample into a plurality of wells; wherein nucleic acids in spatial proximity to each other in the FFPE tissue sample are introduced into the same well;
 - b) barcoding partitioned nucleic acids to form a plurality of barcoded nucleic acids, wherein barcoded nucleic acids within a given discrete well each comprise a common partition-specific barcode sequence, such that the barcode sequences identify nucleic acids from a given well;
 - c) obtaining sequence information from the plurality of barcoded nucleic acids, wherein the sequence information from the plurality of barcoded nucleic acids comprises sequence information of the partition-specific barcode sequence; and
 - d) attributing the plurality of barcoded nucleic acids to a region of spatial proximity, wherein barcoded nucleic acids derived from a region of spatial proximity in the FFPE tissue sample comprise the same partition-specific barcode sequence.
2. The method of claim 1, wherein the barcoding comprises amplifying with a primer comprising a barcode sequence.
3. The method of claim 1, wherein at least two of the nucleic acids partitioned into the same well in the partitioning a) step comprise different sequences.
4. The method of claim 1, further comprising, prior to (a), a prior step of imaging the FFPE tissue sample.
5. The method of claim 1, wherein the FFPE tissue sample is a cancer tissue sample.
6. The method of claim 1, wherein the barcoded nucleic acids in different wells are pooled prior to obtaining sequence information.

7. The method of claim 1, wherein the sequence information further comprises information relating to a nucleic acid obtained from the FFPE tissue sample.
8. The method of claim 1, further comprising, prior to a), a prior step of releasing the nucleic acids from the FFPE tissue sample.
9. The method of claim 1, wherein the nucleic acids obtained from the FFPE tissue sample comprise nucleic acid tags previously applied to the sample.
10. The method of claim 1, wherein the step of obtaining sequence information comprises high throughput sequencing of the plurality of barcoded nucleic acids.
11. The method of claim 1, wherein the partition-specific barcode sequences comprise two or more separate subsequences.
12. The method of claim 4, wherein said imaging comprises imaging tags with optical properties, such that particular tags are associated with particular regions of the imaged sample.
13. The method of claim 12, further comprising correlating barcoded nucleic acids derived from a region of spatial proximity in the FFPE tissue sample with the imaged tags.