

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 ~~OF THE INVENTION~~

[0003] Accordingly, the present ~~invention~~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 ~~aspect~~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 embodimentexamples, the sequence information from obtaining step (c) includes identification of nucleic acids that are in spatial proximity to each other.

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

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

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

[0009] In further embodimentexamples, 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 embodimentexamples, the nucleic acids are not isolated from the sample prior to the separating step (b).

[0011] In any embodimentexamples, 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 embodimentexamples, 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 aspectexamples, 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 aspectexamples, 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 aspectexamples, 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 embodimentexamples, the sample is a formalin-fixed paraffin sample.

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

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

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

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

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

[0022] In any embodimentexamples, 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 aspectexamples, the present inventiondisclosure 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~~while maintaining structural context, the method~~ comprising:
 - ~~(a) providing a sample comprising nucleic acids, wherein 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;~~
 - 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 thereby analyzing nucleic acids while maintaining structural context~~
 - 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.
- ~~2. The method of claim 1, wherein the sequence information from obtaining step (c) includes identification of nucleic acids that were in spatial proximity to each other in the sample.~~

- ~~3. The method of claims 1-2, wherein the obtaining step (c) provides information on intrachromosomal and/or interchromosomal interactions between genomic loci.~~
- ~~4. The method of claims 1-2, wherein the obtaining step (c) provides information on chromosome conformations.~~
- ~~5. The method of claims 1-4, wherein 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.~~
- ~~6. The method of claims 1-5, wherein the sample is a formalin-fixed paraffin sample.~~
- ~~7. The method of claims 1-6, wherein the nucleic acids are not isolated from the sample prior to the separating step (b).~~
- ~~8. The method of claims 1-7, wherein the discrete partitions comprise beads.~~
- ~~9. The method of claim 8, wherein the beads are gel beads.~~
- ~~10. The method of claims 1-9, wherein prior to the obtaining step (c), the nucleic acids within the discrete partitions are barcoded to form a plurality of barcoded fragments, wherein fragments within a given discrete partition each comprise a common barcode, such that the barcodes identify nucleic acids from a given partition.~~
- ~~11. The method of claims 1-10, wherein 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.~~
- ~~12. The method of claims 1-11, wherein the sample comprises a tumor sample.~~
- ~~13. The method of claims 1-11, wherein the sample comprises a mixture of tumor and normal cells.~~

- ~~14. The method of claims 1-13, wherein the sample comprises a nuclear matrix.~~
- ~~15. The method of claims 1-14, wherein the nucleic acids comprise RNA.~~
- ~~16. A method of analyzing nucleic acids from a sample while maintaining structural context of nucleic acids within the sample, the method comprising:~~
- ~~(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.~~
- ~~17. The method of claim 16, wherein the processing step (b) includes blunt-end ligation under conditions favoring intramolecular ligation, such that the spatially adjacent nucleic acid segments are ligated within the same molecule.~~
- ~~18. The method of claim 17, wherein the conditions favoring intramolecular ligation comprise diluting the sample to reduce concentration of the nucleic acids under 10 ng/ μ L.~~
- ~~19. The method of claims 16-18, wherein the sample is a formalin-fixed paraffin sample.~~
- ~~20. The method of claims 16-18, wherein the sample comprises a nuclear matrix.~~

- ~~21. The method of claims 16-20, wherein the nucleic acids comprise RNA.~~
- ~~22. The method of claims 16-21, wherein the nucleic acids are not isolated from the sample prior to the step (a).~~
- ~~23. The method of claims 16-21, wherein prior to step (a), the nucleic acids are immunoprecipitated such that associated DNA binding proteins remain bound to the nucleic acids.~~
- ~~24. The method of claims 16-23, wherein the amount of nucleic acids in the sample is less than 50 ng/ml.~~
- ~~25. The method of claims 16-24, wherein the partitions comprise beads.~~
- ~~26. The method of claim 25, wherein the beads are gel beads.~~
- ~~27. The method of claims 16-26, wherein the sample comprises a tumor sample.~~
- ~~28. The method of claims 16-27, wherein the sample comprises a mixture of tumor and normal cells.~~
- ~~29. The method of claims 16-28, wherein the processing step includes reversal of the linking subsequent to forming the ligation products.~~
- ~~30. The method of claims 16-29, wherein the obtaining step (e) provides information on intrachromosomal and/or interchromosomal interactions between genomic loci.~~
- ~~31. The method of claims 16-29, wherein the obtaining step (e) provides information on chromosome conformations.~~
- ~~32. The method of claim 31, wherein the chromosome conformations are associated with disease states.~~

- ~~33. The method of claims 16-32, wherein the processing step results in ligation products comprising nucleic acids that were originally in close spatial proximity in the sample.~~
- ~~34. The method of claims 16-33, wherein the obtaining step (e) comprises a sequencing reaction selected from the group consisting of: short read-length sequencing reactions and long read-length sequencing reactions.~~
- ~~35. The method of claim 34, wherein the sequencing reaction is a short read, high accuracy sequencing reaction.~~
- ~~36. The method of claims 16-35, wherein the forming step (a) comprises cross-linking nucleic acids in the sample.~~
- ~~37. The method of claims 16-36, wherein the forming step (a) results in covalent links between spatially adjacent nucleic acid segments.~~
- ~~38. A method of analyzing nucleic acids from a sample while maintaining structural context of nucleic acids within the sample, the method comprising:~~
- ~~(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.~~

- ~~39. The method of claim 38, wherein the processing step (c) includes blunt end ligation under conditions favoring intramolecular ligation, such that the spatially adjacent nucleic acid segments are ligated within the same molecule.~~
- ~~40. The method of claims 38-39, wherein the sample is a formalin-fixed paraffin sample.~~
- ~~41. The method of claims 38-39, wherein the sample comprises a nuclear matrix.~~
- ~~42. The method of claims 38-41, wherein the nucleic acids comprise RNA.~~
- ~~43. The method of claims 38-42, wherein the nucleic acids are not isolated from the sample prior to step (a).~~
- ~~44. The method of claims 38-42, wherein prior to step (a), the nucleic acids are immunoprecipitated such that associated DNA binding proteins remain bound to the nucleic acids.~~
- ~~45. The method of claims 38-44, wherein the partitions comprise beads.~~
- ~~46. The method of claim 45, wherein the beads are gel beads.~~
- ~~47. The method of claims 38-46, wherein the sample comprises a tumor sample.~~
- ~~48. The method of claims 38-46, wherein the sample comprises a mixture of tumor and normal cells.~~
- ~~49. The method of claims 38-48, wherein the processing step (c) results in ligation products comprising nucleic acids that were originally in close spatial proximity in the sample.~~

- ~~50. The method of claims 38-49, wherein the obtaining step (e) provides information on intrachromosomal and/or interchromosomal interactions between genomic loci.~~
- ~~51. The method of claims 38-50, wherein the obtaining step (e) comprises a sequencing reaction selected from the group consisting of: short read-length sequencing reactions and long read-length sequencing reactions.~~
- ~~52. The method of claim 51, wherein the sequencing reaction is a short read, high accuracy sequencing reaction.~~
- ~~53. A method of analyzing nucleic acids from a sample while maintaining structural context of nucleic acids within the sample, the method comprising:~~
- ~~(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.~~
- ~~54. The method of claim 53, wherein the processing step (b) includes blunt-end ligation under conditions favoring intramolecular ligation, such that the spatially adjacent nucleic acid segments are ligated within the same molecule.~~
- ~~55. The method of claims 53-54, wherein the sample is a formalin-fixed paraffin sample.~~
- ~~56. The method of claims 53-54, wherein the sample comprises a nuclear matrix.~~

- ~~57. The method of claims 53-56, wherein the nucleic acids comprise RNA.~~
- ~~58. The method of claims 53-57, wherein the nucleic acids are not isolated from the sample prior to the cross-linking step (a).~~
- ~~59. The method of claims 53-57, wherein prior to the cross-linking step (a), the nucleic acids are immunoprecipitated such that associated DNA binding proteins remain bound to the nucleic acids.~~
- ~~60. The method of claims 53-59, wherein prior to the obtaining step (d), the ligation products are associated with a barcode.~~
- ~~61. The method of claim 60, wherein ligation products within the same partition receive common barcodes, such that the barcodes identify ligation products from a given partition.~~
- ~~62. The method of claims 53-61, wherein the obtaining step (d) comprises a sequencing reaction selected from the group consisting of: short read-length sequencing reactions and long read-length sequencing reactions.~~
- ~~63. A method of analyzing nucleic acids while maintaining structural context, the method comprising:~~
- ~~(a) providing a sample comprising nucleic acids;~~
 - ~~(b) applying a library of tags to the sample such that different geographical regions of the sample receive different 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 in the discrete partitions,
thereby analyzing nucleic acids while maintaining structural context.~~
- ~~64. The method of claim 63, wherein the library of tags comprises oligonucleotide tags.~~
- ~~65. The method of claims 63-64, wherein the library of tags comprises m RNA tags.~~
- ~~66. The method of claims 63-65, wherein the applying step (b) is accomplished through diffusion.~~
- ~~67. The method of claim 66, wherein the applying step (b) results in different concentrations of tags in different spatial locations in the sample.~~
- ~~68. The method of claim 67, wherein the identifying step (e) comprises determining concentration of tags in the discrete partitions.~~
- ~~69. The method of claims 63-65, wherein the applying step (b) is accomplished through one or more active processes within the sample.~~
- ~~70. The method of claim 69, wherein presence of identical tags at different spatial locations in the sample indicates that elements of the sample in those different spatial locations are connected through one or more active processes.~~
- ~~71. The method of claims 69-70, wherein the one or more active processes include cellular transport processes.~~
- ~~72. The method of claims 63-71, wherein the applying step (b) comprises randomly distributing different portions of the library of tags to different spatial locations of the sample.~~

~~73. The method of claims 63-71, wherein the applying step (b) comprises adding a predetermined portion or concentration of the library of tags to identified spatial locations of the sample.~~

~~74. The method of claims 63-73, wherein the obtaining step (d) and the identifying step (e) are performed simultaneously.~~

~~75. The method of claims 63-74, the obtaining step (d) comprises a sequencing reaction selected from the group consisting of: short read-length sequencing reactions and long read-length sequencing reactions.~~

~~76. The method of claim 75, wherein the sequencing reaction is a short read, high accuracy sequencing reaction.~~

~~77. The method of claims 63-76, wherein the sample comprises a tumor sample.~~

~~78. The method of claims 63-76, wherein the sample comprises a mixture of tumor and normal cells.~~

~~79. The method of claims 63-76, wherein the sample is a formalin-fixed paraffin sample.~~

~~80. The method of claims 63-76, wherein the sample comprises a nuclear matrix.~~

~~81. The method of claims 63-76, wherein the nucleic acids comprise RNA.~~

~~82. The method of claims 63-82, wherein the partitions comprise beads.~~

~~83. The method of claim 82, wherein the beads are gel beads.~~

~~84. The method of claims 63-83, wherein the amount of nucleic acids in the discrete partitions is less than 10 ng/ml.~~

~~85. The method of claims 63-83, wherein the amount of nucleic acids in the discrete partitions is less than 1 ng/ml.~~