

METHODS AND COMPOSITIONS FOR PROXIMITY LIGATION

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/867,463 filed June 27, 2019, U.S. Provisional Patent Application No. 62/931,069 filed November 5, 2019, U.S. Provisional Patent Application No. 63/011,490 filed April 17, 2020, U.S. Provisional Patent Application No. 62/870,297 filed July 3, 2019, and U.S. Provisional Patent Application No. 63/014,422 filed April 23, 2020, each of which is incorporated herein by reference in its entirety.

BACKGROUND

[0002] Obtaining high-quality, contiguous genome sequences is often difficult, especially in cases when limited source material is available for sequence analysis. While obtaining raw sequence data has become faster and available at a lower cost, suitable methods for analyzing and assembling the data efficiently and accurately remains a challenge.

SUMMARY

[0002a] In one aspect, provided herein is a method comprising: (a) contacting a cross-linked biological sample with a non-specific endonuclease, wherein said cross-linked biological sample comprises a nucleic acid molecule complexed to at least one nucleic acid binding protein, thereby cleaving the nucleic acid molecule into a plurality of segments; (b) attaching a first segment and a second segment of the plurality of segments at a junction; and (c) subjecting the plurality of segments to size selection to obtain a plurality of selected segments, thereby obtaining a product, wherein attaching comprises contacting at least the first segment and the second segment to at least one bridge oligonucleotide.

[0003] In one example, there are provided methods of nucleic acid analysis. In some cases, methods may comprise: (a) obtaining a stabilized biological sample comprising a nucleic acid molecule complexed to at least one nucleic acid binding protein; (b) contacting the stabilized biological sample to a non-specific endonuclease to cleave the nucleic acid molecule into a plurality of segments; (c) attaching a first segment and a second segment of the plurality of segments at a junction; and (d) subjecting the plurality of segments to size selection to obtain a plurality of selected segments. In some cases, the plurality of selected segments is about 145 to about 600 bp. In some cases, the plurality of selected segments is about 100 to about 2500 bp. In some cases, the plurality of selected segments is about 100 to about 600 bp. In some cases, the plurality of selected segments is about 600 to about 2500 bp. In some cases, the method further comprises, prior to step (d), preparing a sequencing library from the plurality of segments. In some cases, the method further comprises subjecting the sequencing library to a size selection to obtain a size-selected library. In some cases, the size-selected library is between about 350 bp and 1000 bp in size. In some cases, the size selection is conducted with gel electrophoresis, capillary electrophoresis, size selection beads, or a gel filtration column. In some cases, the method further comprises analyzing the plurality of selected segments to obtain a QC value. In some cases, the QC value is a chromatin digest efficiency (CDE) based on the proportion of segments between 100 and 2500 bp in size prior to step (d). In some cases, the method further comprises selecting a sample for further analysis when the CDE value

is at least 65%. In some cases, the QC value is a chromatin digest index (CDI) based on the ratio of a number of mononucleosome-sized segments to a number of dinucleosome-sized segments prior to step (d). In some cases, the method further comprises selecting a sample for further analysis when the CDI value is greater than -1.5 and less than 1. In some cases, the method further comprises subsequent to the contacting the stabilized biological sample to a non-specific endonuclease, binding the plurality of segments to one or more surfaces. In some cases, the one or more surfaces comprise one or more beads. In some cases, the one or more beads are solid phase reversible immobilization (SPRI) beads. In some

CLAIMS

WHAT IS CLAIMED IS:

1. A method comprising:
 - (a) contacting a cross-linked biological sample with a non-specific endonuclease, wherein said cross-linked biological sample comprises a nucleic acid molecule complexed to at least one nucleic acid binding protein, thereby cleaving the nucleic acid molecule into a plurality of segments;
 - (b) attaching a first segment and a second segment of the plurality of segments at a junction; and
 - (c) subjecting the plurality of segments to size selection to obtain a plurality of selected segments, thereby obtaining a product, wherein attaching comprises contacting at least the first segment and the second segment to at least one bridge oligonucleotide.
2. The method of claim 1, wherein the plurality of selected segments is 100 to 2500 bp.
3. The method of claim 1, further comprising, prior to step (c), preparing a sequencing library from the plurality of segments.
4. The method of claim 3, further comprising subjecting the sequencing library to a size selection to obtain a size-selected library.
5. The method of any one of claims 1 to 4, wherein the method further comprises analyzing the plurality of selected segments to obtain a QC value.
6. The method of claim 5, wherein the QC value is a chromatin digest efficiency (CDE) is calculated based on the proportion of segments from 100 bp to 2500 bp in size prior to step (d) and wherein the method further comprises selecting a sample for further analysis when the CDE value is at least 65%..
7. The method of claim 5, wherein the QC value is a chromatin digest index (CDI) is calculated based on the ratio of a number of mononucleosome-sized segments to a number of dinucleosome-sized segments prior to step (c) and wherein the method further comprises selecting a sample for further analysis when the CDI value is greater than -1.5 and less than 1.
8. The method of claim 1, further comprising, subsequent to the contacting the cross-linked biological sample to a non-specific endonuclease, binding the plurality of segments to one or more surfaces.

9. The method of any one of claims 1 to 8, wherein the cross-linked biological sample comprises a cross-linked cell lysate, a cross-linked intact cell, or a cross-linked intact nucleus.
10. The method of claim 9, wherein step (a) is conducted prior to lysis of the intact cell or the intact nucleus.
11. The method of claim 1, further comprising, prior to step (b), lysing cells and/or nuclei in the cross-linked biological sample.
12. The method of any one of claims 1 to 11, wherein the cross-linked biological sample comprises fewer than 3,000,000 cells, fewer than 1,000,000 cells, or fewer than 100,000 cells.
13. The method of any one of claims 1 to 12, wherein the cross-linked biological sample comprises less than 10 μ g DNA or less than 1 μ g DNA.
14. The method of any one of claims 1 to 13, wherein the non-specific endonuclease is selected from one or more of DNase I, DNase II, and micrococcal nuclease.
15. The method of any one of claims 1 to 14, wherein the cross-linked biological sample has been treated with a crosslinking agent comprising a chemical fixative or ultraviolet light.
16. The method of claim 15, wherein the chemical fixative comprises formaldehyde, psoralen, disuccinimidyl glutarate (DSG), or ethylene glycol bis(succinimidyl succinate) (EGS).
17. The method of any one of claims 1 to 16, further comprising contacting the plurality of selected segments to an antibody and conducting immunoprecipitation on the plurality of segments.
18. The method of claim 17, wherein the immunoprecipitation is conducted subsequent to the attaching.
19. The method of any one of claims 1 to 18, wherein attaching comprises filling in sticky ends using biotin tagged nucleotides, filling in sticky ends using untagged nucleotides, ligating blunt ends, adding overhangs, or contacting at least the first segment and the second segment to a barcode.
20. The method of claim 1, wherein the bridge oligonucleotide comprises a barcode sequence or an affinity tag.
21. The method of claim 1, wherein attaching comprises contacting at least the first segment and the second segment to multiple bridge oligonucleotides in series.

22. The method of claim 21, wherein the attaching results in samples, cells, nuclei, chromosomes, or nucleic acid molecules of the cross-linked biological sample receiving a unique sequence of bridge oligonucleotides.

23. The method of any one of claims 1 to 22, wherein the at least one bridge oligonucleotide is coupled to an immunoglobulin binding protein.

24. The method of claim 1, further comprising:

(e) reversing crosslinks of the product; and

(f) obtaining at least some sequence on each side of the junction to generate a first read pair.

25. The method of claim 24, further comprising:

(g) mapping the first read pair to a set of contigs; and

(h) determining a path through the set of contigs that represents an order and/or orientation to a genome.

26. The method of claim 24, further comprising:

(g) mapping the first read pair to a set of contigs; and

(h) determining, from the set of contigs, a presence of a structural variant or loss of heterozygosity in the cross-linked biological sample.

27. The method of claim 24, further comprising:

(g) mapping the first read pair to a set of contigs; and

(h) assigning a variant in the set of contigs to a phase.

28. The method of claim 27, wherein the variant is a human leukocyte antigen (HLA) variant or a killer-cell immunoglobulin-like receptor (KIR) variant.

29. The method of claim 24, further comprising:

(g) mapping the first read pair to a set of contigs;

(h) determining, from the set of contigs, a presence of a variant in the set of contigs; and

(i) conducting a step selected from one or more of: (1) identifying a disease stage, a prognosis, or a course of treatment for the cross-linked biological sample; (2) selecting a drug based on the presence of the variant; or (3) identifying a drug efficacy for the cross-linked biological sample.

30. The method of any one of claims 1 to 29, wherein the non-specific endonuclease is coupled or fused to an immunoglobulin binding protein.

31. The method of any one of claims 1 to 30, wherein the bridge oligonucleotide is at least 10 nucleotides in length.