

METHODS AND SYSTEMS FOR ENRICHMENT OF BARCODES

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/788,891, filed January 6, 2019, which is entirely incorporated herein by reference.

BACKGROUND

[0002] A sample may be processed for various purposes, such as identification of a type of moiety within the sample. The sample may be a biological sample. Biological samples may be processed, such as for detection of a disease (e.g., cancer) or identification of a particular species. There are various approaches for processing samples, such as polymerase chain reaction (PCR) and sequencing.

[0003] Biological samples may be processed within various reaction environments, such as partitions. Partitions may be wells or droplets. Droplets or wells may be employed to process biological samples in a manner that enables the biological samples to be partitioned and processed separately. For example, such droplets may be fluidically isolated from other droplets, enabling accurate control of respective environments in the droplets.

[0004] Biological samples in partitions may be subjected to various processes, such as chemical processes or physical processes. Samples in partitions may be subjected to heating or cooling, or chemical reactions, such as to yield species that may be qualitatively or quantitatively processed.

SUMMARY

[0005] The present disclosure provides sample preparation techniques that allow the enrichment of analytes from a library. Procedures for barcoding and isolating analytes from single cells are herein discussed. However, there is benefit in enrichment of target barcodes corresponding analytes of interest such that not all analytes from a library are not subject to downstream processes. This enrichment of analytes may allow downstream processes such as sequencing to be more efficient or provide higher quality data on the desired analytes.

[0006] In an aspect, the present disclosure provides a method of nucleic acid analysis, comprising: (a) providing a plurality of barcoded nucleic acid molecules comprising a plurality of different barcode sequences; (b) identifying at least one barcode sequence from said plurality of different barcode sequences, wherein identifying comprises sequencing said plurality of barcoded nucleic acid molecules to generate a plurality of sequencing reads and analyzing said plurality of sequencing reads to identify said at least one barcode sequence; and (c) enriching nucleic acid molecules comprising said at least one barcode sequence. In one example, the present disclosure provides a method of nucleic acid analysis, comprising: (a) providing a plurality of barcoded nucleic

acid molecules comprising a plurality of different barcode sequences; (b) identifying at least one barcode sequence from the plurality of different barcode sequences; and (c) enriching nucleic acid molecules comprising the at least one barcode sequence.

[0007] In some examples, the method comprises in step (c) comprising performing a nucleic acid extension reaction using (i) a nucleic acid molecule comprising the at least one barcode sequence and (ii) a primer comprising a sequence specific for the at least one barcode sequence to generate an enriched plurality of nucleic acid molecules comprising the at least one barcode sequence.

[0008] In some examples, the nucleic acid extension reaction is a polymerase chain reaction (PCR). In some example, the method further comprises performing an additional PCR on the enriched plurality of nucleic acid molecules comprising the at least one barcode sequence. In some examples, the additional PCR comprises use of an additional primer comprising a sequence specific for the enriched plurality of nucleic acid molecules comprising the at least one barcode sequence. In some examples the additional primer comprises one or more functional sequences that facilitate sequencing the enriched plurality of nucleic acid molecules. In some examples, the one or more functional sequences comprise a sequencing primer sequence. In some examples, the one or more functional sequences comprise a sequence configured to attach to a flow cell of a sequencer.

[0009] In some examples, the primer comprises an affinity group and wherein said enriched plurality of nucleic acid molecules comprises said affinity group. In some examples, the affinity group comprises biotin.

[0010] In some examples, the method further comprises, subsequent to the nucleic acid extension reaction, performing a size selection to remove unincorporated primer.

[0011] In some examples, the method comprises in (c) further comprising coupling the enriched plurality of nucleic acid molecules comprising the affinity group to a solid support specific for the affinity group. In some examples, the solid support is a bead.

[0012] In some examples, the affinity group comprises biotin, and the solid support comprises avidin or streptavidin.

[0013] In some examples, the method further comprises performing a polymerase chain reaction (PCR) on said enriched plurality of nucleic acid molecules.

[0014] In some examples, the method further comprises prior to the nucleic acid extension reaction, (i) hybridizing (1) a first nucleic acid molecule complementary to a first portion of the at least one barcode sequence and (2) a second nucleic acid molecule complementary to a second portion of the at least one barcode sequence; and (ii) ligating the first nucleic acid molecule to the second nucleic acid molecule to generate the primer comprising the sequence specific for the at least one barcode sequence.

[0015] In some examples, the identifying of (b) comprises sequencing the plurality of barcoded nucleic acid molecules to generate a plurality of sequencing reads, and analyzing the plurality of sequencing reads to identify the at least one barcode sequence.

[0016] In some examples, each of the plurality of barcoded nucleic acid molecules further comprise one or more functional sequences selected from the group consisting of a sequencing primer sequence and a sequence configured to attach to a flow cell of a sequencer.

[0017] In some examples, the plurality of barcoded nucleic acid molecules comprise, from 5' to 3', a first adapter sequence, a sequence derived from a template nucleic acid, and a second adapter sequence. In some examples, the first adapter sequence comprises a first sequence configured to attach to a flow cell of a sequencer, a first sequencing primer sequence, and a barcode sequence. In some examples, the second adapter sequence comprises a second sequencing primer sequence and a second sequence configured to attach to a flow cell of a sequencer. In some examples, the second adapter sequence further comprises an index sequence.

[0018] In some examples, each barcode sequence of said plurality of different barcode sequences identifies a nucleic acid molecule as derived from a single cell. In some examples, each barcode sequence of said plurality of different barcode sequences identifies a nucleic acid molecule as derived from a single partition.

[0019] In some examples, the enriching of (c) results in at least 20-fold enrichment of nucleic acid molecules comprising said at least one barcode sequence. In some examples, the enriching of (c) results in at least 50-fold enrichment of nucleic acid molecules comprising said at least one barcode sequence. In some examples, the enriching of (c) results in at least 100-fold enrichment of nucleic acid molecules comprising said at least one barcode sequence. In some examples, the enriching of (c) results in at least 200-fold enrichment of nucleic acid molecules comprising said at least one barcode sequence. In some examples, the enriching of (c) results in at least 500-fold enrichment of nucleic acid molecules comprising said at least one barcode sequence. In some examples, the enriching of (c) results in at least 1000-fold enrichment of nucleic acid molecules comprising said at least one barcode sequence.

[0020] In some examples, the plurality of barcoded nucleic acid molecules is associated with one or more analytes in a sample, wherein the one or more analytes are selected from DNA, RNA, proteins and lipids, or a combination thereof. In some examples, the plurality of barcoded nucleic acid molecules is associated with one or more mRNA in said sample. In some examples, the plurality of barcoded nucleic acid molecules is associated with one or more proteins in said sample. In some examples, the plurality of barcoded nucleic acid molecules is associated with one or more antibodies in said sample. In some examples, the plurality of barcoded nucleic acid molecules is associated with one or more T-cell receptors in said sample. In some examples, the plurality of barcoded nucleic acid molecules is associated with one or more regions of genomic DNA in said

sample. In some examples, the plurality of barcoded nucleic acid molecules is associated with one or more regions of chromatin in said sample.

[0021] In some examples, the identifying comprises identifying a first barcode sequence and a second barcode sequence, wherein the first barcode sequence is different from the second barcode sequence, and the enriching comprises performing a nucleic acid extension reaction using a first oligonucleotide molecule comprising a sequence specific for the first barcode sequence and a second oligonucleotide molecule comprising a sequence specific for the second barcode sequence to generate an enriched plurality of nucleic acid molecules, wherein nucleic acid molecules of the enriched plurality of nucleic acid molecules comprises the first barcode sequence or the second barcode sequence.

[0022] Another example of the present disclosure provides a non-transitory computer readable medium comprising machine executable code that, upon execution by one or more computer processors, implements any of the methods above or elsewhere herein.

[0023] Another example of the present disclosure provides a system comprising one or more computer processors and computer memory coupled thereto. The computer memory comprises machine executable code that, upon execution by the one or more computer processors, implements any of the methods above or elsewhere herein.

[0024] Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative examples of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different examples, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

INCORPORATION BY REFERENCE

[0025] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. To the extent publications and patents or patent applications incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

BRIEF DESCRIPTION OF THE DRAWINGS

[0026] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be

CLAIMS

WHAT IS CLAIMED IS:

1. A method of nucleic acid analysis, comprising:
 - (a) providing a plurality of barcoded nucleic acid molecules comprising a plurality of different barcode sequences;
 - (b) identifying at least one barcode sequence from said plurality of different barcode sequences, wherein identifying comprises sequencing said plurality of barcoded nucleic acid molecules to generate a plurality of sequencing reads and analyzing said plurality of sequencing reads to identify said at least one barcode sequence; and
 - (c) enriching nucleic acid molecules comprising said at least one barcode sequence.
2. The method of claim 1, wherein (c) comprises performing a nucleic acid extension reaction using (i) a nucleic acid molecule comprising said at least one barcode sequence and (ii) a primer comprising a sequence specific for said at least one barcode sequence to generate an enriched plurality of nucleic acid molecules comprising said at least one barcode sequence.
3. The method of claim 2, wherein said nucleic acid extension reaction is a polymerase chain reaction (PCR), optionally further comprising performing an additional PCR on said enriched plurality of nucleic acid molecules comprising said at least one barcode sequence, optionally wherein said additional PCR comprises use of an additional primer comprising a sequence specific for said enriched plurality of nucleic acid molecules comprising said at least one barcode sequence, optionally wherein said additional primer comprises one or more functional sequences that facilitate sequencing said enriched plurality of nucleic acid molecules, optionally wherein said one or more functional sequences comprise:
 - (i) a sequencing primer sequence; and/or
 - (ii) a sequence configured to attach to a flow cell of a sequencer.
4. The method of claim 2, wherein said primer comprises an affinity group and wherein said enriched plurality of nucleic acid molecules comprises said affinity group, optionally wherein said affinity group comprises biotin.
5. The method of any one of claims 2-4, further comprising, subsequent to said nucleic acid extension reaction, performing a size selection to remove unincorporated primer.

6. The method of claim 4 or claim 5, wherein (c) further comprises coupling said enriched plurality of nucleic acid molecules comprising said affinity group to a solid support specific for said affinity group, optionally:

- (i) wherein said solid support is a bead;
- (ii) wherein said affinity group comprises biotin and wherein said solid support comprises avidin or streptavidin; and/or
- (iii) further comprising performing a polymerase chain reaction (PCR) on said enriched plurality of nucleic acid molecules.

7. The method of any one of claims 2-6, further comprising, prior to said nucleic acid extension reaction, (i) hybridizing (1) a first nucleic acid molecule complementary to a first portion of said at least one barcode sequence and (2) a second nucleic acid molecule complementary to a second portion of said at least one barcode sequence; and (ii) ligating said first nucleic acid molecule to said second nucleic acid molecule to generate said primer comprising said sequence specific for said at least one barcode sequence.

8. The method of any one of claims 1-7, wherein each of said plurality of barcoded nucleic acid molecules further comprise one or more functional sequences selected from the group consisting of a sequencing primer sequence and a sequence configured to attach to a flow cell of a sequencer.

9. The method of any one of claims 1-8, wherein said plurality of barcoded nucleic acid molecules comprise, from 5' to 3', a first adapter sequence, a sequence derived from a template nucleic acid, and a second adapter sequence, optionally wherein said first adapter sequence comprises a first sequence configured to attach to a flow cell of a sequencer, a first sequencing primer sequence, and a barcode sequence, optionally wherein said second adapter sequence comprises a second sequencing primer sequence and a second sequence configured to attach to a flow cell of a sequencer, optionally wherein said second adapter sequence further comprises an index sequence.

10. The method of any one of claims 1-9, wherein each barcode sequence of said plurality of different barcode sequences identifies a nucleic acid molecule as derived from (i) a single cell; and/or (ii) a single partition.

11. The method of any one of claims 1-10, wherein said enriching of (c) results in at least 10-fold enrichment, at least 20- fold enrichment, at least 50- fold enrichment, at least 100-fold enrichment, at least 200-fold enrichment, at least 500-fold enrichment, or at least 1,000-fold enrichment of nucleic acid molecules comprising said at least one barcode sequence.

12. The method of any preceding claim, wherein the plurality of barcoded nucleic acid molecules corresponds to one or more analytes in a sample, wherein the one or more analytes are selected from DNA, RNA, proteins, a region of chromatin, and lipids, or a combination thereof, optionally wherein the plurality of barcoded nucleic acid molecules corresponds to:

- (i) one or more mRNA in said sample;
- (ii) one or more proteins in said sample, optionally one or more antibodies in said sample or one or more T-cell receptors in said sample;
- (iii) one or more regions of genomic DNA in said sample; or
- (iv) one or more regions of chromatin in said sample.

13. The method of any of the preceding claims, wherein (b) comprises identifying a first barcode sequence and a second barcode sequence, wherein said first barcode sequence is different from said second barcode sequence, and wherein (c) comprises performing a nucleic acid extension reaction using a first oligonucleotide molecule comprising a sequence specific for said first barcode sequence and a second oligonucleotide molecule comprising a sequence specific for said second barcode sequence to generate an enriched plurality of nucleic acid molecules, wherein nucleic acid molecules of said enriched plurality of nucleic acid molecules comprises said first barcode sequence or said second barcode sequence.

14. The method of any one of the preceding claims, wherein each barcoded nucleic acid molecule of the plurality of barcoded nucleic acid molecules is associated with an analyte from a sample.

15. The method of any one of the preceding claims, wherein the plurality of barcoded nucleic acid molecules comprises a plurality of subsets of barcoded nucleic acid molecules, wherein a first subset of said plurality of subsets of barcoded nucleic acid molecules comprises a first common barcode sequence, optionally wherein:

- (i) a second subset of the plurality of subsets of barcoded nucleic acid molecules comprises a second common barcode sequence that is different from the first common

barcode sequence, optionally wherein the second common barcode sequence corresponds to analytes from a second single cell; or

(ii) the first common barcode sequence corresponds to analytes from a first single cell.