

polyclonal barcode packets. Other approaches have involved the incorporation of multiple genomic or transcriptomic targets of interest together into single continuous amplicons, as in the case of overlap extension PCR. These approaches rely on tailored universal primer sets to physically link the targets of interest rather than appending a barcode to each target. Since the primer sets are universal and many copies are introduced into every container, the mechanism by which targets are assigned to their cells of origin does not face the disadvantage of a Poisson-driven distribution. However, these approaches are severely limited in the number of potential genes or transcripts that can be targeted from each cell. Most if not all existing methods are presently limited to two (DeKosky, BJ, Nat Biotechnol 31(2) (2013)).

BRIEF SUMMARY OF THE INVENTION

[0005] Provided herein are methods of identifying the origin of a nucleic acid sample. The methods include forming a reaction mixture comprising a nucleic acid sample comprising nucleic acid molecules from a single cell and a set of barcodes, incorporating the set of barcodes into the nucleic acid molecules of the sample, and identifying the set of barcodes incorporated into the nucleic acid molecules of the single cell thereby identifying the origin of the nucleic acid sample. Also provided are methods of identifying the origin of a nucleic acid sample that include providing a set of 5' and 3' nucleic acid constructs each construct comprising a unique 5' or 3' barcode sequence respectively, providing a nucleic acid sample comprising nucleic acid molecules from a single cell, contacting the nucleic acid sample with the set of 5' and 3' nucleic acid constructs, incorporating into the nucleic acid molecules the 5' and 3' barcode sequences from the 5' and 3' nucleic acid constructs respectively, and identifying the set of 5' and 3' barcode sequences thereby identifying the origin of the nucleic acid sample from the single cell.

[0005a] In one aspect, there is provided a method of identifying the origin of a nucleic acid sample comprising the steps of (a) forming a reaction mixture comprising a nucleic acid sample comprising nucleic acid molecules from a single cell and a set of barcodes comprising one or more 5' nucleic acid constructs each construct comprising a unique 5' barcode and one or more 3' nucleic acid constructs each construct comprising a unique 3' barcode, wherein forming the reaction mixture comprises providing cells in a suspension buffer and contacting the suspension buffer comprising the cells with a barcoding buffer comprising the set of barcodes under conditions to form the reaction mixture comprising the single cell and the set of barcodes; (b) incorporating the set of barcodes into the nucleic acid molecules of the sample; and (c) identifying the set of barcodes incorporated into the nucleic acid molecules of the single cell thereby identifying the origin of the nucleic acid sample, wherein the combination of 5' and 3' barcodes in the nucleic acid sample identifies the origin of the nucleic acid sample.

[0005b] In one aspect, there is provided a method of identifying the origin of a nucleic acid sample comprising the steps of (a) forming a reaction mixture comprising a nucleic acid sample comprising nucleic acid molecules from a single cell and a set of barcodes comprising one or more 5' nucleic acid constructs each construct comprising a unique 5' barcode and one or more 3' nucleic acid constructs each construct comprising a unique 3' barcode, wherein the set of barcodes comprises at least two 5' nucleic acid constructs each construct comprising a unique 5' barcode; (b) incorporating the set of barcodes into the nucleic acid molecules of the sample; and (c) identifying the set of barcodes incorporated into the nucleic acid molecules of the single cell thereby identifying the origin of the nucleic acid sample, wherein the combination of 5' and 3' barcodes in the nucleic acid sample identifies the origin of the nucleic acid sample.

[0005c] In one aspect, there is provided a method of identifying the origin of a nucleic acid sample comprising the steps of (a) forming a reaction mixture comprising a nucleic acid sample comprising nucleic acid molecules from a single cell and a set of barcodes comprising one or more 5' nucleic acid constructs each construct comprising a unique 5' barcode and one or more 3' nucleic acid constructs each construct comprising a unique 3' barcode, wherein the set of barcodes comprises at least two 3' nucleic acid constructs each construct comprising a unique 3' barcode; (b) incorporating the set of barcodes into the nucleic acid molecules of the sample; and (c) identifying the set of barcodes incorporated into the nucleic acid molecules of the single cell thereby identifying the origin of the nucleic acid sample, wherein the combination of 5' and 3' barcodes in the nucleic acid sample identifies the origin of the nucleic acid sample.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] Figure 1 is a schematic drawing of a microfluidic droplet chip with oil input channels in a flow-focusing configuration for droplet formation and the following aqueous input channels: (1) cells in a suspension buffer and (2) barcoded beads in barcoding mix. The suspension buffer and barcoding mix are located different microfluidic channels that converge to merge their components at ratios to comprise the ultimate mix that is desired in the droplets. Barcoded beads and cells are loaded into aqueous droplets as described by a Poisson distribution. The average

WHAT IS CLAIMED IS:

1. A method of identifying the origin of a nucleic acid sample comprising the steps of
 - (a) forming a reaction mixture comprising a nucleic acid sample comprising nucleic acid molecules from a single cell and a set of barcodes comprising one or more 5' nucleic acid constructs each construct comprising a unique 5' barcode and one or more 3' nucleic acid constructs each construct comprising a unique 3' barcode, wherein forming the reaction mixture comprises providing cells in a suspension buffer and contacting the suspension buffer comprising the cells with a barcoding buffer comprising the set of barcodes under conditions to form the reaction mixture comprising the single cell and the set of barcodes;
 - (b) incorporating the set of barcodes into the nucleic acid molecules of the sample; and
 - (c) identifying the set of barcodes incorporated into the nucleic acid molecules of the single cell thereby identifying the origin of the nucleic acid sample, wherein the combination of 5' and 3' barcodes in the nucleic acid sample identifies the origin of the nucleic acid sample.
2. The method of claim 1, wherein forming the reaction mixture comprises contacting the single cell with the set of barcodes and lysing the cell to provide the nucleic acid sample from the single cell.
3. The method of any one of claims 1 or 2, wherein forming the reaction mixture comprises providing a cells in a suspension buffer, and contacting the suspension buffer with a barcoding buffer comprising the 5' and 3' nucleic acid constructs under conditions to form the reaction mixture comprising the single cell and at least one 5' nucleic acid construct and at least one 3' nucleic acid construct.
4. The method of any one of claim 1 or 2, wherein forming the reaction mixture comprises providing a cells in a suspension buffer, and contacting the suspension buffer with a first barcoding buffer comprising the 5' nucleic acid constructs and a second barcoding buffer comprising the 3' nucleic acid constructs under conditions to form the reaction mixture comprising the single cell and at least one 5' nucleic acid construct and at least one 3' nucleic acid construct.

5. The method of any one of claims 3 or 4, wherein the constructs further comprise a unique molecular identifier (UMI) sequence.
6. The method of any one of claims 3-5, wherein each 5' nucleic acid construct is provided in multiple copies attached to a solid surface.
7. The method of any one of claims 3-6, wherein each 3' nucleic acid construct is provided in multiple copies attached to a solid surface
8. The method of claim 6 or 7, wherein the solid surface is a bead, microbead, microparticle, microsphere, nanoparticle, nanobead, or hydrogel.
9. The method of any one of claims 3-8, wherein the reaction mixture comprises the nucleic acid sample from the single cell, at least one bead comprising multiple copies of the 5' nucleic acid constructs, and at least one bead comprising multiple copies of the 3' nucleic acid construct.
10. The method of any one of claims 3-9, wherein each 5' and 3' nucleic acid construct further comprises a photocleavable linker or a restriction enzyme site.
11. The method of any one of claims 3-9, wherein each 5' nucleic acid construct further comprises a photocleavable linker and each 3' nucleic acid construct further comprises a restriction enzyme site.
12. The method of any one of claims 3-9, wherein each 5' nucleic acid construct further comprises a restriction enzyme site and each 3' nucleic acid construct further comprises a photocleavable linker.
13. The method of claim 11 or 12, wherein the method further comprises contacting the reaction mixture with a restriction enzyme and exposing the reaction mixture to UV light to release the 5' and 3' nucleic acid constructs from the solid surface.

14. The method of claim 10, wherein the 5' and 3' nucleic acid constructs further comprise a restriction enzyme site and the method further comprises contacting the reaction mixture with a restriction enzyme to release the 5' and 3' nucleic acid constructs from the solid surface.
15. The method of claim 10, wherein the 5' and 3' nucleic acid constructs further comprise a photocleavable linker and the method further comprises exposing the reaction mixture to UV light to release the 5' and 3' nucleic acid constructs from the solid surface.
16. The method of any one of claims 3-15, wherein the 5' nucleic acid construct is a double stranded nucleic acid construct comprising a promoter and wherein the 5' nucleic acid construct serves as a template to produce multiple copies of the 5' nucleic acid construct.
17. The method of any one of claims 1-16, wherein the set of barcodes comprises 2, 3, 4, 5, 6, 7, or 8 5' nucleic acid constructs and 2, 3, 4, 5, 6, 7, or 8 3' nucleic acid constructs.
18. The method of any one of claims 1-17, wherein the 3' nucleic acid constructs further comprise a binding sequence.
19. The method of claim 18, wherein the binding sequence comprises a poly-T sequence.
20. The method of claim 19, wherein the poly-T sequence binds to the poly-A sequences on the nucleic acid molecules of the sample.
21. The method of any one of claims 1-20, wherein the barcodes on the 3' nucleic acid constructs are incorporated by reverse transcription.
22. The method of claim 21, wherein the reverse transcription produces nucleic acid molecules comprising the 3' barcode sequence.

23. The method of claim 22, wherein the reverse transcriptase enzyme adds two to five cysteine residues to the end of the nucleic acid molecules in the sample.
24. The method of claim 23, wherein the two to five cysteine residues provides a binding sequence complementary to a binding sequence on the 5' nucleic acid constructs.
25. The method of claim 24, wherein the binding sequence on the 5' nucleic acid constructs comprises a poly-G sequence.
26. The method of claim 25, wherein the binding sequence on the 5' nucleic acid constructs binds to poly-C sequences on the nucleic acid molecules of the sample.
27. The method of any one of claims 1-26, wherein the 5' nucleic acid construct is incorporated into the nucleic acid sample using a reverse transcriptase enzyme.
28. The method of any one of claims 21-27, wherein the reverse transcriptase is an M-MLV enzyme.
29. The method of claim 28, wherein the reverse transcriptase is MMLV H- reverse transcriptase.
30. The method of any one of claims 1-29, wherein the nucleic acid sample is further amplified to produce multiple copies of the plurality of nucleic acid sequences each sequence comprising a 5' barcode and a 3' barcode.
31. The method of any one of claims 1-30, wherein the nucleic acid molecules comprise mRNAs.
32. The method of claim 31, wherein the mRNAs comprise at least 1, 3, 10, 30, 100, 300, 1,000, 3,000, 10,000, 30,000, 100,000, 300,000, or 1,000,000 mRNAs.

33. The method of any one of claims 1-32, wherein the barcodes in the nucleic acid sample are identified by sequencing.
34. A method of identifying the origin of a nucleic acid sample comprising the steps of
- (a) forming a reaction mixture comprising a nucleic acid sample comprising nucleic acid molecules from a single cell and a set of barcodes comprising one or more 5' nucleic acid constructs each construct comprising a unique 5' barcode and one or more 3' nucleic acid constructs each construct comprising a unique 3' barcode, wherein the set of barcodes comprises at least two 5' nucleic acid constructs each construct comprising a unique 5' barcode;
 - (b) incorporating the set of barcodes into the nucleic acid molecules of the sample; and
 - (c) identifying the set of barcodes incorporated into the nucleic acid molecules of the single cell thereby identifying the origin of the nucleic acid sample, wherein the combination of 5' and 3' barcodes in the nucleic acid sample identifies the origin of the nucleic acid sample.
35. A method of identifying the origin of a nucleic acid sample comprising the steps of
- (a) forming a reaction mixture comprising a nucleic acid sample comprising nucleic acid molecules from a single cell and a set of barcodes comprising one or more 5' nucleic acid constructs each construct comprising a unique 5' barcode and one or more 3' nucleic acid constructs each construct comprising a unique 3' barcode, wherein the set of barcodes comprises at least two 3' nucleic acid constructs each construct comprising a unique 3' barcode;
 - (b) incorporating the set of barcodes into the nucleic acid molecules of the sample; and
 - (c) identifying the set of barcodes incorporated into the nucleic acid molecules of the single cell thereby identifying the origin of the nucleic acid sample, wherein the combination of 5' and 3' barcodes in the nucleic acid sample identifies the origin of the nucleic acid sample.
36. The method of claim 1, wherein one or more of the nucleic acid molecules comprises two or more unique barcodes.
37. The method of claim 36, wherein the unique barcodes are located at the 5' end, the '3' end or both the 5' and 3' end of the nucleic acid molecules.

38. The method of any one of claims 1-37, wherein the method comprises forming a plurality of reaction mixtures.
39. The method of claim 38, wherein two or more unique barcodes are present in 30%, 40%, 50%, 60%, 70%, 80%, 90% or 100% of the reaction mixtures.
40. The method of claim 38, wherein forming the plurality of reaction mixtures comprises adding the set of barcodes to the plurality of reaction mixtures at a lambda greater than 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 2.0, 3.0, 4.0 or 5.0.
41. The method of any one of claims 1-37, wherein the reaction mixture is a droplet.