

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; (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.

[0005b] In one aspect, there is provided a method of identifying the origin of a nucleic acid sample comprising the steps of: (a) providing a set of 3' nucleic acid constructs each construct comprising a unique 3' barcode sequence, a binding sequence, and a barcode association sequence; (b) providing a nucleic acid sample comprising nucleic acid molecules from a single cell; (c) contacting the nucleic acid sample with the set of 3' nucleic acid constructs; (d) separating the 3' barcode sequence and binding sequence from the barcode association sequence; (e) incorporating into the nucleic acid molecules the 3' barcode sequences from the 3' nucleic acid constructs; and (f) identifying the set of 3' barcode sequences using the barcode association sequences thereby identifying the origin of the nucleic acid sample from the single cell.

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