

COMPOSITIONS AND METHODS FOR NUCLEIC ACID ANALYSIS

CROSS-REFERENCE

[0001] This application is related to U.S. Provisional Application No. 63/058,712, filed July 30, 2020, U.S. Provisional Patent Application No. 63/061,093, filed August 4, 2020, U.S. Provisional Patent Application No. 63/105,860, filed October 26, 2020, U.S. Provisional Patent Application No. 63/106,566, filed October 28, 2020, and U.S. Provisional Patent Application No. 63/152,976 filed on February 24, 2021, U.S. Provisional Patent Application No. 63/178,386 filed on April 22, 2021, U.S. Provisional Patent Application No. 63/210,927 filed on June 15, 2021, U.S. Provisional Patent Application No. 63/212,500 filed on June 18, 2021, U.S. Provisional Patent Application No. 63/213,626 filed on June 22, 2021, and U.S. Provisional Patent Application No. 63/215,752 filed on June 28, 2021, each of which applications is herein incorporated by reference for all purposes.

SUMMARY

[0001a] In one aspect, provided herein is a method, comprising: (a) generating a double-stranded polynucleotide comprising: (i) a first base that is a substrate for a deaminase that is unprotected, and (ii) a second base that is a substrate for a deaminase that is protected, (b) contacting the double-stranded polynucleotide with a helicase and said deaminase to yield a nucleic acid molecule comprising a nucleic acid sequence comprising a deaminated base and said second base, wherein a base of said double-stranded polynucleotide is deaminated to yield said deaminated base, wherein said deaminated base is different from said second base; (c) sequencing nucleotides of said nucleic acid sequence or a complement thereof to obtain sequencing data; and (d) processing said sequencing data to identify and distinguish said first base and said second base.

[0002] In various embodiments, a method comprises: providing a forward polynucleotide and a cognate polynucleotide; determining a first identity of a first base at a locus of the forward polynucleotide and a second identity of a second base at or proximal to a corresponding locus of the cognate polynucleotide using sequencing; and using a computer comprising a processor, a memory, and instructions stored thereupon that, when executed, determine a value of a true base at a locus of an original polynucleotide corresponding to the locus of the forward polynucleotide based at least in part on the first identity of the first base and the second identity of the second base. In some cases, the second base is at the corresponding locus of the cognate polynucleotide. In some cases, the second base is proximal to the corresponding locus of the cognate polynucleotide. In some cases, the forward polynucleotide and cognate polynucleotide are linked. In some embodiments of the method, at any

time point prior to sequencing, the forward polynucleotide and the cognate polynucleotide are (i) covalently linked via a hairpin; (ii) linked as a double-stranded polynucleotide via Watson-Crick base pairing; (iii) each coupled to a barcode; or (iv) any combination thereof. In some cases, the forward polynucleotide comprises a complementary deoxyribonucleic acid (cDNA) molecule or an amplicon thereof. In some cases, the method further comprises contacting an RNA nucleotide with a reverse transcriptase, biologically active fragment thereof, or derivative thereof to generate the forward polynucleotide. In some cases, the original polynucleotide comprises a deoxyribonucleic acid (DNA) polynucleotide isolated from a sample obtained from a subject. In some cases, the original polynucleotide comprises a cell free DNA