

SUMMARY

[0005] Although many of these methods are able to reduce or combat the errors introduced by the sample preparation and sequencing processes for all molecules that are converted and sequenced, these methods are not able to infer the counts of molecules that were converted but not sequenced. Since this count of converted by unsequenced molecules can be highly variable from genomic region to region, these counts can dramatically and adversely affect the sensitivity that can be achieved.

[0006] To address this issue, input double-stranded deoxyribonucleic acid (DNA) can be converted by a process that tags both halves of the individual double-stranded molecule, in some cases differently. This can be performed using a variety of techniques, including ligation of hairpin, bubble, or forked adapters or other adaptors having double-stranded and single stranded segments (the unhybridized portion of a bubble, forked or hairpin adapter are deemed single-stranded herein). If tagged correctly, each original Watson and Crick (i.e., strand) side of the input double-stranded DNA molecule can be differently tagged and identified by the sequencer and subsequent bioinformatics. For all molecules in a particular region, counts of molecules where both Watson and Crick sides were recovered ("Pairs") versus those where only one half was recovered ("Singlets") can be recorded. The number of unseen molecules can be estimated based on the number of Pairs and Singlets detected.

[0006a] In one aspect, there is provided a method for determining a quantitative measure indicative of a number of individual double-stranded deoxyribonucleic acid (DNA) molecules in a sample, comprising:

- (a) tagging the individual DNA molecules in the sample with a set of duplex tags, wherein each duplex tag differently tags complementary strands of a double-stranded DNA molecule of the DNA molecules in the sample to provide tagged strands;
- (b) sequencing at least some of the tagged strands to provide sequence reads;
- (c) using the sequence reads to determine a quantitative measure of individual DNA molecules for which both strands are detected;
- (d) using the sequence reads to determine a quantitative measure of individual DNA molecules for which only one of the DNA strands is detected;
- (e) inferring from (c) and (d) a quantitative measure of individual DNA molecules for which neither strand was detected; and

(f) using (c)-(e) to determine the quantitative measure indicative of a number of individual double-stranded DNA molecules in the sample.

[0007] An aspect of the present disclosure provides a method for detecting and/or quantifying rare deoxyribonucleic acid (DNA) in a heterogeneous population of original DNA fragments, comprising tagging the original DNA fragments in a single reaction using a library of a plurality of different tags such that greater than 30% of the fragments are tagged at both ends, wherein each of the tags comprises a molecular barcode. The single reaction can be in a single reaction vessel. Greater than 50% of the fragments can be tagged at both ends. The plurality of different tags can be no more than any of 100, 500, 1000, 10,000 or 100,000 different tags.

[0008] Another aspect provides a set of library adaptors that can be used to tag the molecules of interest (e.g., by ligation, hybridization, etc.). The set of library adaptors can comprise plurality of polynucleotide molecules with molecular barcodes, wherein the plurality of polynucleotide molecules are less than or equal to 80 nucleotide bases in length, wherein the molecular barcodes are at least 4 nucleotide bases in length, and wherein (a) the molecular barcodes are different from one another and have an edit distance of at least 1 between one another; (b) the molecular barcodes are located at least one nucleotide base

CLAIMS

WHAT IS CLAIMED IS:

1. A method for determining a quantitative measure indicative of a number of individual double-stranded deoxyribonucleic acid (DNA) molecules in a sample, comprising:
 - (a) tagging the individual DNA molecules in the sample with a set of duplex tags, wherein each duplex tag differently tags complementary strands of a double-stranded DNA molecule of the DNA molecules in the sample to provide tagged strands;
 - (b) sequencing at least some of the tagged strands to provide sequence reads;
 - (c) using the sequence reads to determine a quantitative measure of individual DNA molecules for which both strands are detected;
 - (d) using the sequence reads to determine a quantitative measure of individual DNA molecules for which only one of the DNA strands is detected;
 - (e) inferring from (c) and (d) a quantitative measure of individual DNA molecules for which neither strand was detected; and
 - (f) using (c)-(e) to determine the quantitative measure indicative of a number of individual double-stranded DNA molecules in the sample.
2. The method of Claim 1, further comprising detecting copy number variation in said sample by determining a normalized quantitative measure determined in step (f) at each of one or more genetic loci and determining copy number variation based on the normalized measure.
3. The method of Claim 1, wherein said sample comprises double-stranded polynucleotide molecules sourced substantially from cell-free nucleic acids.
4. The method of Claim 1, further comprising sorting sequence reads into paired reads and unpaired reads, wherein (i) each paired read corresponds to sequence reads generated from a first tagged strand and a second differently tagged complementary strand derived from a double-stranded polynucleotide molecule in said set, and (ii) each unpaired read represents a first tagged strand having no second differently tag complementary strand derived from a double-stranded polynucleotide molecule represented among said sequence reads in said set of sequence reads.

5. The method of Claim 4, further comprising determining quantitative measures of (i) said paired reads and (ii) said unpaired reads that map to each of one or more genetic loci to determine a quantitative measure of total double-stranded DNA molecules in said sample that map to each of said one or more genetic loci based on said quantitative measure of paired reads and unpaired reads mapping to each locus.
6. The method of claim 1, further comprising reducing and/or tracking redundancy in the set of sequence reads.
7. The method of claim 6, wherein reducing redundancy in the set of sequence reads comprises collapsing sequence reads produced from amplified products of an original polynucleotide molecule of the individual DNA molecules in the sample back to the original polynucleotide.
8. The method of claim 7, further comprising determining a consensus sequence for the original polynucleotide molecule.
9. The method of claim 8, further comprising determining a quantitative measure of paired reads that map to a locus, wherein both strands of the pair comprise a sequence variant.
10. The method of claim 8, further comprising determining a quantitative measure of paired molecules in which only one member of the pair bears a sequence variant and/or determining a quantitative measure of unpaired molecules bearing a sequence variant.
11. The method of claim 1, further comprising attaching one or more double-stranded adaptors to both ends of the double-stranded DNA molecule.
12. The method of claim 1, wherein a quantitative measure is determined by comparing the number of individual DNA molecules in the sample that map to a locus of interest with a number of individual DNA molecules in the sample mapping to a reference sequence.
13. The method of claim 1, wherein in (b), both ends of each of the individual DNA molecules are tagged with the duplex tags to provides tagged strands.
14. The method of claim 1, wherein the set of duplex tags comprises non-unique tags.