

**What is claimed is:**

1. A system for labeling nucleic acid molecules in a sample comprising:  
a set of oligonucleotides comprising a plurality of barcodes, each barcode comprising a stretch of contiguous bases comprising:
  - (i) a sample barcode comprising a pre-determined number of sample index positions comprising one or more specific nucleotides, wherein the location of sample index positions varies between samples; and
  - (ii) a molecular barcode comprising molecular index positions comprising a nucleotide that differs from the nucleotides at sample index positions, wherein sample index positions are interspersed among molecular index positions.
2. The system of claim 1, wherein the pre-determined number of sample barcode positions varies among different sample barcodes.
3. The system of claim 1, wherein the barcode comprises about 10 to about 35 nucleotides.
4. The system of claim 1, wherein the barcode comprises about 12 to about 25 nucleotides.
5. The system of claim 1, wherein the sample barcode comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 sample index positions, or a combination thereof.
6. The system of claim 1, wherein the sample barcode comprises about 4 to about 12 sample index positions.
7. The system of claim 1, wherein the molecular barcode comprises about 5 to about 25 molecular index positions.
8. The system of claim 1, wherein the molecular barcode comprises about 5 to about 15 molecular index positions.
9. The system of claim 1, wherein sample index position nucleotides and molecular index position nucleotides are selected from:
  - (A) the sample index position nucleotide is A and the molecular index position nucleotide is C, G, T, or a combination thereof;

(B) the sample index position nucleotide is T and the molecular index position nucleotide is C, G, A, or a combination thereof;

(C) the sample index position nucleotide is C and the molecular index position nucleotide is G, A, T, or a combination thereof;

(D) the sample index position nucleotide is G and the molecular index position nucleotide is C, A, T, or a combination thereof;

(E) the sample index position nucleotide is A, T, or a combination thereof and the molecular index position nucleotide is C, G, or a combination thereof;

(F) the sample index position nucleotide is A, C, or a combination thereof and the molecular index position nucleotide is T, G, or a combination thereof;

(G) the sample index position nucleotide is A, G, or a combination thereof and the molecular index position nucleotide is T, C, or a combination thereof;

(H) the sample index position nucleotide is T, C, or a combination thereof and the molecular index position nucleotide is A, G, or a combination thereof;

(I) the sample index position nucleotide is T, G, or a combination thereof and the molecular index position nucleotide is A, C, or a combination thereof; or

(J) the sample index position nucleotide is G, C, or a combination thereof and the molecular index position nucleotide is A, T, or a combination thereof.

**10.** The system of claim 1, wherein each barcode comprises one or more additional index barcodes comprising index positions.

**11.** The system of claim 10, wherein the one or more additional index barcode is a cellular barcode, a barcode that provides a measure of DNA length of an unrepaired end, or both a cellular barcode and a barcode that provides a measure of DNA length of an unrepaired end.

**12.** The system of claim 1, wherein each oligonucleotide in the set of oligonucleotides further comprises non-barcode positions comprising sites for hybridization, sites for sequence primer binding, sites for amplification, or any combination thereof.

**13.** A set of oligonucleotides for labeling nucleic acid molecules in a sample comprising a plurality of barcodes, each barcode comprising:

(i) a sample barcode comprising a pre-determined number of sample index positions comprising one or more specific nucleotides, wherein the location of sample index positions varies between samples; and

(ii) a molecular barcode comprising molecular index positions comprising a nucleotide that differs from the nucleotides at sample index positions,

wherein sample index positions and molecular index positions are interspersed in a stretch of contiguous bases.

**14.** The set of oligonucleotides of claim 13, wherein the pre-determined number of sample barcode positions varies among different sample barcodes.

**15.** The set of oligonucleotides of claim 13, wherein the barcode comprises about 10 to about 35 nucleotides.

**16.** The set of oligonucleotides of claim 13, wherein the barcode comprises about 12 to about 25 nucleotides.

**17.** The set of oligonucleotides of claim 13, wherein the sample barcode comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 sample index positions, or a combination thereof.

**18.** The set of oligonucleotides of claim 13, wherein the sample barcode comprises about 4 to about 12 sample index positions.

**19.** The set of oligonucleotides of claim 13, wherein the molecular barcode comprises about 5 to about 25 molecular index positions.

**20.** The set of oligonucleotides of claim 13, wherein the molecular barcode comprises about 5 to about 15 molecular index positions.

**21.** The set of oligonucleotides of claim 13, wherein sample index position nucleotides and molecular index position nucleotides are selected from:

(A) the sample index position nucleotide is A and the molecular index position nucleotide is C, G, T, or a combination thereof;

(B) the sample index position nucleotide is T and the molecular index position nucleotide is C, G, A, or a combination thereof;

(C) the sample index position nucleotide is C and the molecular index position nucleotide is G, A, T, or a combination thereof;

(D) the sample index position nucleotide is G and the molecular index position nucleotide is C, A, T, or a combination thereof;

(E) the sample index position nucleotide is A, T, or a combination thereof and the molecular index position nucleotide is C, G, or a combination thereof;

(F) the sample index position nucleotide is A, C, or a combination thereof and the molecular index position nucleotide is T, G, or a combination thereof;

(G) the sample index position nucleotide is A, G, or a combination thereof and the molecular index position nucleotide is T, C, or a combination thereof;

(H) the sample index position nucleotide is T, C, or a combination thereof and the molecular index position nucleotide is A, G, or a combination thereof;

(I) the sample index position nucleotide is T, G, or a combination thereof and the molecular index position nucleotide is A, C, or a combination thereof; or

(J) the sample index position nucleotide is G, C, or a combination thereof and the molecular index position nucleotide is A, T, or a combination thereof.

**22.** The set of oligonucleotides of claim 13, wherein each barcode comprises one or more additional index barcodes comprising index positions.

**23.** The set of oligonucleotides of claim 22, wherein the one or more additional index barcode is a cellular barcode, a barcode that provides a measure of DNA length of an unrepaired end, or both a cellular barcode and a barcode that provides a measure of DNA length of an unrepaired end.

**24.** The set of oligonucleotides of claim 13, wherein each oligonucleotide in the set of oligonucleotides further comprises non-barcode positions comprising sites for hybridization, sites for sequence primer binding, sites for amplification, or any combination thereof.

**25.** A method for analyzing sequences of nucleic acid molecules in a sample comprising:

(a) attaching a plurality of oligonucleotides to the nucleic acid molecules, wherein each oligonucleotide comprises a barcode comprising:

(i) a sample barcode comprising a pre-determined number of sample index positions comprising one or more specific nucleotides, wherein the location of sample index positions varies between samples; and

(ii) a molecular barcode comprising molecular index positions comprising a nucleotide that differs from the nucleotides at sample index positions, wherein sample index positions and molecular index positions are interspersed in a stretch of contiguous bases; and

(b) sequencing the nucleic acid molecules, wherein sequence reads comprise barcode sequences.

**26.** The method of claim 25, further comprising attaching an oligonucleotide comprising the same sample barcode to each end of a nucleic acid molecule in the sample.

**27.** The method of claim 25, wherein the pre-determined number of sample barcode positions varies among different sample barcodes.

**28.** The method of claim 25, wherein the barcode comprises about 10 to about 35 nucleotides.

**29.** The method of claim 25, wherein the barcode comprises about 12 to about 25 nucleotides.

**30.** The method of claim 25, wherein the sample barcode comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 sample index positions, or a combination thereof.

**31.** The method of claim 25, wherein the sample barcode comprises about 4 to about 12 sample index positions.

**32.** The method of claim 25, wherein the molecular barcode comprises about 5 to about 25 molecular index positions.

**33.** The method of claim 25, wherein the molecular barcode comprises about 5 to about 15 molecular index positions.

**34.** The method of claim 25, wherein sample index position nucleotides and molecular index position nucleotides are selected from:

(A) the sample index position nucleotide is A and the molecular index position nucleotide is C, G, T, or a combination thereof;

(B) the sample index position nucleotide is T and the molecular index position nucleotide is C, G, A, or a combination thereof;

(C) the sample index position nucleotide is C and the molecular index position nucleotide is G, A, T, or a combination thereof;

(D) the sample index position nucleotide is G and the molecular index position nucleotide is C, A, T, or a combination thereof;

(E) the sample index position nucleotide is A, T, or a combination thereof and the molecular index position nucleotide is C, G, or a combination thereof;

(F) the sample index position nucleotide is A, C, or a combination thereof and the molecular index position nucleotide is T, G, or a combination thereof;

(G) the sample index position nucleotide is A, G, or a combination thereof and the molecular index position nucleotide is T, C, or a combination thereof;

(H) the sample index position nucleotide is T, C, or a combination thereof and the molecular index position nucleotide is A, G, or a combination thereof;

(I) the sample index position nucleotide is T, G, or a combination thereof and the molecular index position nucleotide is A, C, or a combination thereof; or

(J) the sample index position nucleotide is G, C, or a combination thereof and the molecular index position nucleotide is A, T, or a combination thereof.

**35.** The method of claim 25, wherein each barcode comprises one or more additional index barcodes comprising index positions.

**36.** The method of claim 35, wherein the one or more additional index barcode is a cellular barcode, a barcode that provides a measure of DNA length of an unrepaired end, or both a cellular barcode and a barcode that provides a measure of DNA length of an unrepaired end.

**37.** The method of claim 25, further comprising assigning the sequence reads to sample families based on the location of sample index positions.

**38.** The method of claim 25, further comprising assigning the sequence reads to molecular families based on the location of molecular index positions and the nucleotide at each molecular index position.

**39.** The method of claim 25, further comprising correcting for sequencing errors by comparing the number and location of sample index positions in a sequence read to the pre-determined number and location of sample index positions.

**40.** The method of claim 25, further comprising correcting for sequencing errors by comparing sample barcodes at both ends of a sequence read.

41. The method of claim 40, comprising applying a rule to compare non-identical sample barcodes at each end of the sequence read to allowed sample barcodes.
42. The method of claim 25, further comprising applying one or more rules (1) to correct for errors within barcodes, (2) to correct for errors between barcodes at each end of a nucleic acid molecule, (3) for demultiplexing sequence reads into sample families, (4) for assigning sequence reads to molecular families, or any combination thereof.
43. The method of claim 25, wherein each oligonucleotide further comprises non-barcode positions comprising sites for hybridization, sites for sequence primer binding, sites for amplification, or any combination thereof.
44. The method of claim 25, further comprising use of a different genome with each oligonucleotide being tested to sensitively detect sequence read misassignment.
45. A method for labeling nucleic acid molecules in a sample comprising:  
attaching a plurality of oligonucleotides to the nucleic acid molecules comprising a barcode, each barcode comprising:
- (i) a sample barcode comprising a pre-determined number of sample index positions comprising one or more specific nucleotides, wherein the location of sample index positions varies between samples; and
  - (ii) a molecular barcode comprising molecular index positions comprising a nucleotide that differs from the nucleotides at sample index positions, wherein sample index positions and molecular index positions are interspersed in a stretch of contiguous bases.
46. The method of claim 45, further comprising attaching an oligonucleotide comprising the same sample barcode to each end of a nucleic acid molecule.
47. The method of claim 45, wherein the pre-determined number of sample barcode positions varies among different sample barcodes.
48. The method of claim 45, wherein the barcode comprises about 10 to about 35 nucleotides.
49. The method of claim 45, wherein the barcode comprises about 12 to about 25 nucleotides.
50. The method of claim 45, wherein the sample barcode comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 sample index positions.

51. The method of claim 45, wherein the sample barcode comprises about 4 to about 12 sample index positions.
52. The method of claim 45, wherein the molecular barcode comprises about 5 to about 25 molecular index positions.
53. The method of claim 45, wherein the molecular barcode comprises about 5 to about 15 molecular index positions.
54. The method of claim 45, wherein sample index position nucleotides and molecular index position nucleotides are selected from:
- (A) the sample index position nucleotide is A and the molecular index position nucleotide is C, G, T, or a combination thereof;
  - (B) the sample index position nucleotide is T and the molecular index position nucleotide is C, G, A, or a combination thereof;
  - (C) the sample index position nucleotide is C and the molecular index position nucleotide is G, A, T, or a combination thereof;
  - (D) the sample index position nucleotide is G and the molecular index position nucleotide is C, A, T, or a combination thereof;
  - (E) the sample index position nucleotide is A, T, or a combination thereof and the molecular index position nucleotide is C, G, or a combination thereof;
  - (F) the sample index position nucleotide is A, C, or a combination thereof and the molecular index position nucleotide is T, G, or a combination thereof;
  - (G) the sample index position nucleotide is A, G, or a combination thereof and the molecular index position nucleotide is T, C, or a combination thereof;
  - (H) the sample index position nucleotide is T, C, or a combination thereof and the molecular index position nucleotide is A, G, or a combination thereof;
  - (I) the sample index position nucleotide is T, G, or a combination thereof and the molecular index position nucleotide is A, C, or a combination thereof; or
  - (J) the sample index position nucleotide is G, C, or a combination thereof and the molecular index position nucleotide is A, T, or a combination thereof.
55. The method of claim 45, wherein each barcode comprises one or more additional index barcodes comprising index positions.

- 56.** The method of claim 55, wherein the one or more additional barcode is a cellular barcode, a barcode that provides a measure of DNA length of an unrepaired end, or both a cellular barcode and a barcode that provides a measure of DNA length of an unrepaired end.
- 57.** The method of claim 45, wherein each oligonucleotide further comprises non-barcode positions comprising sites for hybridization, sites for sequence primer binding, sites for amplification, or any combination thereof.
- 58.** The method of any one of claims 25-44, further comprising storing nucleic acid sequence data without demultiplexing.
- 59.** The method of claim 58, wherein storing nucleic acid sequence data without demultiplexing prevents use of sequence data in the absence of a demultiplexing key and prevents unauthorized use of the data.
- 60.** A method for identifying erroneous sequence reads comprising:
- (a) attaching a plurality of oligonucleotides to the nucleic acid molecules of the sample, wherein each oligonucleotide comprises a barcode comprising:
- (i) a sample barcode comprising a pre-determined number of sample index positions comprising one or more specific nucleotides, wherein the location of sample index positions varies between samples, and wherein a same sample barcode is attached to each end of a nucleic acid molecule in the sample; and
  - (ii) a molecular barcode comprising molecular index positions comprising a nucleotide that differs from the nucleotides at sample index positions, wherein sample index positions and molecular index positions are interspersed in a stretch of contiguous bases; and
- (b) sequencing the nucleic acid molecules, wherein sequence reads comprise barcode sequences, thereby identifying erroneous sequence reads.
- 61.** The method of claim 60, wherein identifying erroneous sequence reads comprises identifying nucleic acid molecules with discrepant sample barcodes.
- 62.** The method of claim 60, further comprising correcting for sequencing errors by comparing sample barcodes at both ends of a sequence read.
- 63.** The method of claim 61, further comprising removing the nucleic acid molecules with discrepant sample barcodes from the sequence reads and/or from molecular families.

- 64.** The method of claim 61, wherein identifying nucleic acid molecules with discrepant sample barcodes comprises identifying misprimed nucleic acid molecules.
- 65.** The method of claim 64, wherein misprimed nucleic acid molecules are corrected with proper barcodes and used for improving sequence quality.
- 66.** The method of claim 65, wherein nucleic acid molecules with corrected barcodes are assigned to corrected read families.
- 67.** The method of claim 66, wherein corrected read families are used to accurately determine distinct coverage.
- 68.** The method of claim 67, wherein distinct coverage determination is used to evaluate libraries of nucleic acid molecules.
- 69.** The method of claim 60, further comprising assigning the sequence reads to molecular families based on the location of molecular index positions and the nucleotide at each molecular index position.
- 70.** The method of claim 69, wherein identifying erroneous sequence reads comprises identifying nucleic acid molecules assigned to multiple molecular families.
- 71.** The method of claim 70, further comprising removing the nucleic acid molecules assigned to multiple molecular families from the sequence reads and/or from molecular families.