

CLAIMS

~~What is Claimed is:~~

1. A pool of nucleic acid standards of different lengths, wherein the nucleic acid standards comprise a unique molecular identifier (UMI) and:

- a. a 5' universal oligonucleotide, wherein the 5' universal oligonucleotide is the same for all standards;
- b. a 3' universal oligonucleotide, wherein the 3' universal oligonucleotide is the same for all standards; and
- c. at least one region between the UMI and the 5' universal oligonucleotide and/or between the UMI and the 3' universal oligonucleotide;

wherein the length of the at least one region determines the length of the standard.

2. The pool of standards of claim 1, wherein the pool further comprises a further nucleic acid standard that comprises a UMI and:

- a. a 5' universal oligonucleotide, wherein the 5' universal oligonucleotide is the same for all standards; and
- b. a 3' universal oligonucleotide, wherein the 3' universal oligonucleotide is the same for all standards;

wherein the further nucleic acid standard does not comprise at least one region between the UMI and the 5' universal oligonucleotide or between the UMI and the 3' universal oligonucleotide.

3. The pool of standards of claim 1, wherein the at least one region between the UMI and the 5' universal oligonucleotide and/or between the UMI and the 3' universal oligonucleotide comprise 0.2kb-10kb.

4. The pool of standards of any one of claims 1-3, wherein the 5' universal oligonucleotide and/or the 3' universal oligonucleotide each comprise an amplicon amplified from a sequence of interest.

5. The pool of standards of any one of claims 1 or 3-4, wherein the at least one region between the UMI and the 5' universal oligonucleotide and/or between the UMI and the 3' universal oligonucleotide each comprise an amplicon amplified from a sequence of interest.

6. The pool of standards of any one of claims 1 or 3-5, wherein the least one region between the UMI and the 5' universal oligonucleotide and/or between the UMI and the 3' universal oligonucleotide each comprise an arbitrary sequence.
7. A pool of nucleic acid standards of different lengths, wherein the nucleic acid standards comprise a UMI and:
 - a. a 5' partially overlapping oligonucleotide, wherein the 5' partially overlapping oligonucleotide is identical over at least a portion of its sequence for all the standards; and/or
 - b. a 3' partially overlapping oligonucleotide, wherein the 3' partially overlapping oligonucleotide is identical over at least a portion of its sequence for all the standards;wherein the lengths of the 5' partially overlapping oligonucleotide and/or the 3' partially overlapping oligonucleotide determines the length of the standard.
8. The pool of standards of claim 7, wherein:
 - a. the 5' partially overlapping oligonucleotide comprises at least a first portion of a sequence of interest; and
 - b. the 3' partially overlapping oligonucleotide comprises at least a second portion of a sequence of interest.
9. The pool of standards of any one of claims 7-8, wherein the 5' partially overlapping oligonucleotide and/or the 3' partially overlapping oligonucleotide each comprise a sequence that is 20bp-1kb smaller than a sequence of interest.
10. The pool of standards of any one of claims 7-9, wherein the 5' partially overlapping oligonucleotide and/or the 3' partially overlapping oligonucleotide each comprise an amplicon amplified from a sequence of interest.
11. The pool of standards of any one of claims 1-10, wherein the standards are double-stranded.
12. The pool of standards of any one of claims 1-11, wherein the standards comprise double-stranded DNA.
13. The pool of standards of any one of claims 1-12, wherein each standard comprises a different UMI.
14. The pool of standards of any one of claims 1-13, wherein the UMIs comprised in the pool of standards are a random set of sequences comprising 16-20 base pairs.

15. The pool of standards of claim 14, wherein the UMIs comprised in the pool of standards are a random set of sequences comprising 18 base pairs.
16. The pool of standards of any one of claims 1-15, wherein the pool of standards comprises 1×10^{10} or greater, 10×10^{10} or greater, or 100×10^{10} or greater standards, wherein each standard comprises a different UMI.
17. The pool of standards of any one of claims 1-16, wherein the number of standards in the pool is greater than the number of amplicons generated by an amplification reaction.
18. A pool of standards, wherein:
- a. at least a first portion of the standards are from any one of claims 1-6 or 11-17 and wherein
 - b. at least a second portion of the standards are from any one a pool of nucleic acid standards of claims 7-17 different lengths, wherein the nucleic acid standards comprise a UMI and:
 - i. a 5' partially overlapping oligonucleotide, wherein the 5' partially overlapping oligonucleotide is identical over at least a portion of its sequence for all the standards; and/or
 - ii. a 3' partially overlapping oligonucleotide, wherein the 3' partially overlapping oligonucleotide is identical over at least a portion of its sequence for all the standards, wherein the lengths of the 5' partially overlapping oligonucleotide and/or the 3' partially overlapping oligonucleotide determines the length of the standard.
19. A method of generating a pool of nucleic acid standards comprising:
- a. providing multiple copies of at least one sequence of interest comprising nucleic acids;
 - b. providing a collection of oligonucleotides each comprising a UMI;
 - c. providing a collection of insertion oligonucleotides of varying lengths; and
 - d. ligating at least one sequence of interest of (a), at least one oligonucleotide comprising a UMI of (b), and at least one insertion amplicon of (c) to produce multiple nucleic acid standards of the pool of nucleic acid standards.
20. The method of claim 19, wherein the at least one sequence of interest and/or insertion oligonucleotide are prepared by amplification.

21. The method of claim 19 or claim 20, wherein the sequence of interest, the oligonucleotides each comprising a UMI, and/or the insertion oligonucleotides comprise a restriction enzyme cleavage site.
22. The method of claim 21, wherein the restriction enzyme cleavage site is proximal to the 5' and/or 3' end of the sequence of interest, the oligonucleotides each comprising a UMI, and/or the insertion oligonucleotides.
23. The method of claim 21 or claim 22, wherein the method further comprises cleaving the sequence of interest, the oligonucleotides each comprising a UMI, and/or the insertion oligonucleotides with a restriction enzyme before the ligating.
24. The method of claim 23, wherein the cleaving with a restriction enzyme produces sticky ends for the ligating.
- ~~25. A method of generating a pool of nucleic acid standards comprising:~~
- ~~a. providing multiple copies of at least one sequence of interest comprising nucleic acids;~~
- ~~b. providing a collection of oligonucleotides each comprising a UMI; and~~
- ~~c. ligating at least one sequence of interest of (a) and at least one oligonucleotide comprising a UMI of (b).~~
- ~~26. The method of claim 25, wherein the at least one sequence of interest are prepared by amplification.~~
- ~~27. The method of claim 25 or 26, wherein the sequence of interest and/or the oligonucleotides each comprising a UMI comprise a restriction enzyme cleavage site.~~
- ~~28. The method of claim 27, wherein the restriction enzyme cleavage site is proximal to the 5' and/or 3' end of the sequence of interest and/or the oligonucleotides each comprising a UMI.~~
- ~~29. The method of claim 27-28, wherein the method further comprises cleaving the sequence of interest and/or the oligonucleotides each comprising a UMI with a restriction enzyme before the ligating.~~
- ~~30. The method of claim 29, wherein the cleaving with a restriction enzyme produces sticky ends for the ligating.~~
- ~~31. A method of normalizing amplicon size bias comprising:~~
- ~~a. combining a sample comprising a target nucleic acid with a pool of nucleic acid standards of different lengths, wherein each standard comprises a UMI;~~

- b. ~~amplifying the standards and amplicons of a sequence of interest comprised in the target nucleic acid;~~
 - e. ~~sequencing the standards and the amplicons of the sequence of interest to generate sequencing data;~~
 - d. ~~determining a bias profile based on amplicon size using sequencing data from the standards; and~~
 - e. ~~normalizing amplicon size bias using the bias profile.~~
32. ~~The method of claim 31, wherein the standards in the pool of nucleic acid standards range from 0.2kb to 20kb base pairs.~~
33. ~~The method of claim 31 or claim 32, wherein each standard comprised in the pool of nucleic acid standards comprises a different a UMI.~~
34. ~~The method of claim 31-33, wherein the UMIs comprised in the pool of standards are a random set of sequences comprising 16-20 base pairs.~~
35. ~~The method of claim 31-34, wherein the UMIs comprised in the pool of standards are a random set of sequences comprising 18 base pairs.~~
36. ~~The method of any one of claims 31-35, wherein the pool of standards comprises 1×10^{10} or greater, 10×10^{10} or greater, or 100×10^{10} or greater standards, wherein each standard comprises a different UMI.~~
37. ~~The method of any one of claims 31-36, wherein the number of standards in the pool of standards is greater than the number of amplicons generated by the amplifying.~~
38. ~~The method of any one of claims 31-37, wherein the pool of nucleic acid standards comprises the pool of nucleic acid standards of any one of claims 1-18.~~
39. ~~The method of any one of claims 31-37, wherein the pool of nucleic acid standards comprises a first portion comprising the pool of nucleic acid standards of any one of claims 1-6 or 11-17 and a second portion comprising the pool of nucleic acid standards of any one of claims 7-17.~~
40. ~~The method of any one of claims 31-39, wherein the sequence of interest comprises a restriction enzyme cleavage site that is not at or in close proximity to the 5' and/or 3' end of the sequence of interest.~~
41. ~~The method of any one of claims 31-40, wherein the sequence of interest may comprise insertion or deletion mutations.~~

42. — The method of any one of claims 31–41, wherein the sequence of interest has been subjected to gene editing, optionally wherein the sequence of interest comprises a cut site introduced by gene editing.
43. — The method of any one of claims 31–42, wherein amplifying amplicons of the sequence of interest comprises amplifying amplicons from the target nucleic acid with a pair of PCR primers that bind to primer binding sequences at the ends of the sequence of interest.
44. — The method of any one of claims 31–43, wherein the standards comprise the same primer binding sequences as those at the ends of the sequence of interest.
45. — The method of any one of claims 31–44, further comprising generating a library of fragments after the amplifying and before the sequencing.
46. — The method of claim 31–45, wherein the generating a library of fragments is by tagmentation.
47. — The method of any one of claims 31–46, wherein the sequencing data from the standards used to determine the bias profile is the unique molecule count of UMIs comprised in the standards.
48. — A method of determining the presence of DNA damage in a library comprising one or more library molecule, wherein each library molecule comprises a double-stranded DNA insert with a hairpin adapter at each end of the insert, comprising:
- a. — denaturing the first stand and second strand of the double-stranded DNA inserts comprised in library molecules;
 - b. — annealing a forward primer and a reverse primer to library molecules;
 - c. — amplifying to produce library amplicons; and
 - d. — assessing the presence of DNA damage based on the number of library amplicons produced.
49. — The method of claim 48, wherein the forward primer and/or the reverse primer bind to one or more sequences comprised in one or both hairpin adapter.
50. — The method of claim 48 or claim 49, wherein the forward primer binds to a sequence comprised in the hairpin adapter attached to a first end of the double-stranded DNA insert and the reverse primer binds to a sequence comprised in the hairpin adapter attached to a second end of the double-stranded DNA insert.

51. — The method of any one of claims 48-50, wherein the number of library amplicons produced is estimated by measuring a cycle of quantification (Cq) value.
52. — The method of any one of claims 48-51, wherein a higher number of library amplicons results in a lower Cq value.
53. — The method of any one of claims 48-52, wherein a library with a lower Cq value has less DNA damage.
54. — The method of any one of claims 51-53, further comprising determining conditions for analysis of the library based on the Cq value.
55. — The method of claim 54, wherein the analysis is sequencing.
56. — The method of any one of claims 48-55, wherein the amplifying is optimized for amplifying library molecules that are 5kb or greater, 10kb or greater, 15kb or greater, 20kb or greater, 25kb or greater, or 30kb or greater.
57. — The method of any one of claims 48-56, wherein the amplifying is performed with a polymerase optimized for amplification of long amplicons.
58. — The method of claim 57, wherein the polymerase is optimized for amplification of amplicons of 20kb or more or 30kb or more.
59. — The method of claim 57 or claim 58, wherein the polymerase has a higher processivity or extension rate as compared to a wildtype Taq polymerase.
60. — The method of claim 59, wherein the polymerase comprises one or more mutation or fusion that increase processivity or extension rate.
61. — The method of claim 59 or claim 60, wherein the polymerase has an extension rate of 3kb/minute or higher.
62. — The method of any one of claims 48-61, wherein the amplifying is exponential.
63. — The method of any one of claims 48-62, wherein 30 or more or 40 or more cycles of amplifying are performed.
64. — The method of any one of claims 48-63, wherein the DNA damage comprises one or more nicks in a library molecule.
65. — The method of claim 64, wherein the one or more nicks are within the insert.
66. — The method of claim 64 or claim 65, wherein the Cq value is greater when a greater percentage of library molecules in the library comprise one or more nicks.

67. — The method of any one of claims 64–66, wherein the DNA damage comprises two or more nicks in a library molecule, wherein the nicks are in the same strand of the double-stranded DNA insert.

68. — The method of any one of claims 64–66, wherein the DNA damage comprises two or more nicks in a library molecule, wherein the nicks are in both strands of the double-stranded DNA insert.

69. — The method of any one of claims 48–68, wherein the forward primer and/or the reverse primer cannot generate an amplicon corresponding to the full sequence of the library molecule if the library molecule comprises one or more nicks.

70. — The method of claim 69, wherein an amplicon generated from a library molecule comprising a nick lacks a sequence for binding to the forward and/or reverse primer.

71. — The method of any one of claims 64–70, wherein library molecules comprising a nick generate fewer amplicons during the amplifying as compared to library molecules not comprising a nick.

72. — The method of any one of claims 64–71, further comprising generating a double-stranded break from a nick before annealing the forward primer and the reverse primer.

73. — The method of claim 72, wherein the generating a double-stranded break is performed using an enzymatic reaction.

74. — The method of claim 73, wherein the enzymatic reaction is performed by an endonuclease.

75. — The method of claim 74, wherein the endonuclease is a T7 endonuclease.

76. — The method of any one of claims 72–75, wherein a library molecule comprising a double-stranded break does not generate amplicons corresponding to the full sequence of the library molecule during the amplifying.

77. — The method of claim 72–76, wherein an amplicon generated from a library molecule comprising a double-stranded break lacks a sequence for binding to the forward and/or reverse primer.

78. — A method of quantifying DNA damage in a sample comprising DNA using fluorescence comprising:

a. — combining:

i. — an aliquot of a sample comprising DNA,

- ii. ~~one or more DNA repair enzyme; and~~
 - iii. ~~dNTPs, wherein one or more dNTP is fluorescently labeled;~~
 - b. ~~preparing repaired DNA;~~
 - c. ~~dephosphorylating the phosphates from dNTPs;~~
 - d. ~~binding the repaired DNA to carboxylate or cellulose beads;~~
 - e. ~~eluting the bound repaired DNA from the carboxylate or cellulose beads with a resuspension buffer; and~~
 - f. ~~measuring fluorescence of the repaired DNA to determine the amount of DNA damage.~~
79. ~~The method of claim 78, wherein a greater fluorescence of the repaired DNA indicates greater DNA damage.~~
80. ~~The method of claim 78 or claim 79, wherein the fluorescence of the repaired DNA is linear over a range difference amounts of DNA damage.~~
81. ~~The method of any one of claims 78-80, wherein the assay can assess DNA damage induced by a manipulation of the sample by assessing an aliquot of the same sample before and after the manipulation.~~
82. ~~The method of claim 81, wherein the manipulation is sequencing of a sample.~~
83. ~~The method of claim 81 or claim 82, wherein measuring fluorescence of the repaired DNA comprises preparing a standard curve of dilutions of repaired DNA and measuring the fluorescence of the dilutions of repaired DNA.~~
84. ~~The method of any one of claims 78-83, wherein measuring fluorescence of the repaired DNA comprises comparing the fluorescence of the repaired DNA against a separate standard curve of dilutions of only the one or more dNTP that is fluorescently labeled to determine the number of fluorescent dye molecules comprised in the repaired DNA.~~
85. ~~The method of claim 84, further comprising calculating the normalized number of fluorescent dye molecules comprised in the repaired DNA by dividing the number of fluorescent dye molecules determined by the mass of the repaired DNA.~~
86. ~~The method of any one of claims 78-85, wherein the DNA is genomic DNA, cDNA, or a library comprising fragmented double-stranded DNA.~~
87. ~~The method of claim 86, wherein the DNA is genomic DNA and cDNA and the method further comprising preparing a library after determining the amount of DNA damage.~~

88. — The method of claim 87, wherein a library is prepared if the amount of DNA damage is 5% or less, 4% or less, 3% or less, 2% or less, or 1% or less of total nucleotides.

89. — The method of any one of claims 78-88, wherein a library is not prepared if the amount of DNA damage is 5% or greater, 4% or greater, 3% or greater, 2% or greater, or 1% or greater of total nucleotides.

90. — The method of any one of claims 78-89, wherein more than one round of binding the repaired DNA to carboxylate or cellulose beads and eluting is performed before measuring the fluorescence.

91. — The method of claim 90, wherein two rounds of binding the repaired DNA to carboxylate or cellulose beads and eluting is performed before measuring the fluorescence.

92. — The method of any one of claims 78-91, wherein the carboxylate or cellulose beads are magnetic.

93. — The method of any one of claims 78-92, wherein the preparing repaired DNA is performed at 37°C.

94. — The method of any one of claims 78-93, wherein the preparing repaired DNA is performed for 10 minutes or more, 20 minutes or more, 30 minutes or more, 45 minutes or more, or 60 minutes or more.

95. — The method of claim 78-94, wherein dephosphorylating the phosphates from dNTPs is performed with an enzyme.

96. — The method of claim 78-95, wherein the enzyme for dephosphorylating the phosphates from dNTPs is shrimp alkaline phosphatase (SAP) or calf intestinal alkaline phosphatase (CIP).

97. — The method of any one of claims 78-96, wherein the one or more DNA repair enzyme comprises a DNA polymerase.

98. — The method of claim 97, wherein the DNA polymerase has 5'–3' polymerase activity but lacks 5'–3' exonuclease activity.

99. — The method of claim 97, wherein the DNA polymerase is *Bst* DNA polymerase, large fragment.

100. — The method of any of claims 78-99, wherein the one or more DNA repair enzyme comprises a ligase.

101. — The method of claim 100, wherein the ligase is *Taq* ligase.

102.—The method of any one of claims 78–101, wherein the DNA damage comprises a nick in double-stranded DNA.

103.—The method of any one of claims 78–102, wherein the one or more DNA repair enzyme comprises T4 pyrimidine dimer glycosylase (PDG).

104.—The method of any one of claims 78–103, wherein the DNA damage comprises a thymine dimer.

105.—The method of claim 104, wherein the thymine dimer was induced by ultraviolet irradiation.

106.—The method of any of claims 78–105, wherein the one or more DNA repair enzyme comprises uracil DNA glycosylase (UDG) and an apurinic or apyrimidinic site lyase.

107.—The method of any one of claims 78–106, wherein the DNA damage comprises a uracil.

108.—The method of any of claims 78–107, wherein the one or more DNA repair enzyme comprises formamidopyrimidine DNA glycosylase (FPG) and an apurinic or apyrimidinic site lyase.

109.—The method of claim 78–108, wherein the DNA damage comprises an oxidized base.

110.—The method of any one of claims 78–109, wherein the dNTPs comprise dATP, dGTP, dCTP, and dTTP or dUTP.

111.—The method of any one of claims 78–110, wherein all the dNTPs are fluorescently labeled.

112.—The method of claim 78–111, wherein dUTP and dCTP are fluorescently labeled.

113.—The method of claim 112, wherein the fluorescent label is Alexa Fluor 488, Alexa Fluor 546, Alexa Fluor 555, Alexa Fluor 633, fluorescein isothiocyanate (FITC), or tetramethylrhodamine 5-(and 6)-isothiocyanate (TRITC).