

CLAIMS

What is claimed is:

1. A set of oligonucleotides comprising a plurality of subsets of oligonucleotides,
5 wherein:
 - the set of oligonucleotides is configured to identify sources of nucleic acid samples in multiplex massively parallel sequencing, each of the nucleic acid samples comprising a plurality of nucleic acid molecules;
 - the set of oligonucleotides comprises a set of index sequences comprising at least
10 6 different index sequences, each subset of the plurality of subsets of oligonucleotides comprising a plurality of index sequences of the set of index sequences;
 - a Hamming distance between any two index sequences of the set of index sequences is not less than a first criterion value, wherein the first criterion value is at least 2; and
 - 15 the set of index sequences comprises a plurality of pairs of color-balanced index sequences, wherein any two bases at corresponding sequence positions of each pair of color-balanced index sequences include both (i) an adenine (A) base or a cytosine (C) base, and (ii) a guanine (G) base, a thymine (T) base, or a uracil (U) base.
- 20 2. The set of oligonucleotides of claim 1, wherein each subset of the plurality of subsets of oligonucleotides comprises one or more pairs of color-balanced index sequences.
3. The set of oligonucleotides of any of the preceding claims, wherein the set of
25 oligonucleotides comprises (a) double-stranded or Y-shaped sequencing adapters, each strand of each double-stranded or Y-shaped sequencing adapter comprises an index sequence of the set of index sequences or a reverse complement thereof; or (b) pairs of single-stranded oligonucleotides, each pair being provided together in a reagent, each oligonucleotide of a pair comprising an index sequence of the set of index sequences or a
30 reverse complement thereof.

4. The set of oligonucleotides of claim 3, wherein (a) each Y-shaped or double-stranded sequencing adapter comprises a first strand comprising a first index sequence selected from a first subset of the set of index sequences and a second strand comprising a second index sequence selected from a second subset of the set of index sequences (or reverse complements of the second subset); or (b) each pair of oligonucleotides comprises a first oligonucleotide comprising a first index sequence selected from a first subset of the set of index sequences and a second oligonucleotide comprising a second index sequence selected from a second subset of the set of index sequences (or reverse complements of the second subset).

10

5. The set of oligonucleotides of claim 4, wherein the first strand of each Y-shaped sequencing adapter comprises a P5 flow cell amplification primer binding site, and the second strand of each Y-shaped sequencing adapter comprises a P7' flow cell amplification primer binding site.

15

6. The set of oligonucleotides of claim 4, wherein the first subset comprises a subset listed in Table 1 and the second subset comprises a subset listed in Table 2.

7. The set of oligonucleotides of claim 4, wherein the first and the second index sequences respectively are: the n^{th} 10-mer in SEQ ID NO: 10 and n^{th} 10-mer in SEQ ID NO: 11 or a reverse complement thereof; the n^{th} 10-mer in SEQ ID NO: 12 and n^{th} 10-mer in SEQ ID NO: 13 or a reverse complement thereof; the n^{th} 10-mer in SEQ ID NO: 14 and n^{th} 10-mer in SEQ ID NO: 15 or a reverse complement thereof; the n^{th} 10-mer in SEQ ID NO: 16 and n^{th} 10-mer in SEQ ID NO: 17 or a reverse complement thereof; the n^{th} 10-mer in SEQ ID NO: 18 and n^{th} 10-mer in SEQ ID NO: 19 or a reverse complement thereof.

8. The set of oligonucleotides of 4, wherein the first subset and the second subset are the same.

30

9. The set of oligonucleotides of claim 8, wherein the first strand of each Y-shaped sequencing adapter comprises a P5 flow cell amplification primer binding site, and the second strand of each Y-shaped sequencing adapter comprises a P7' flow cell amplification primer binding site.
- 5
10. The set of oligonucleotides of claim 8, wherein the subset is selected from one of the subsets of index sequences in Table 3.
11. The set of oligonucleotides of any of the preceding claims, wherein the set of
10 index sequences comprise 10-mers in SEQ ID NO: 9.
12. The set of oligonucleotides of any of the preceding claims, wherein each subset of the plurality of subsets of oligonucleotides comprises index sequences corresponding to 10-mers in one of the following sequences: SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID
15 NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, or SEQ ID NO: 19.
13. The set of oligonucleotides of any of the preceding claims, wherein each subset of the plurality of subsets of oligonucleotides comprises index sequences corresponding to
20 reverse complements of 10-mers in one of the following sequences: SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, or SEQ ID NO: 19.
14. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences comprises a plurality of non-overlapping subsets of index sequences, a
25 Hamming distance between any two index sequences in any subset is not less than a second criterion value.
15. The set of oligonucleotides of claim 14, wherein the second criterion value is larger than the first criterion value.
- 30

16. The set of oligonucleotides of claim 15, wherein the first criterion value is 4, and the second criterion value is 5.
17. The set of oligonucleotides of any of the preceding claims, wherein an
5 oligonucleotide of the set of oligonucleotides comprises a primer having an index sequence on its 3' end and an index sequence on its 5' end.
18. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences is comprised in Y-shaped sequencing adapters, each Y-shaped
10 sequencing adapter comprising an index sequence on only one strand.
19. The set of oligonucleotides of any of the preceding claims, wherein the set of oligonucleotides comprises polymerase chain reaction (PCR) primers or reverse complements thereof.
- 15 20. The set of oligonucleotides of any of the preceding claims, wherein the first criterion value is 3.
21. The set of oligonucleotides of any of the preceding claims, wherein the first
20 criterion value is 4.
22. The set of oligonucleotides of any of the preceding claims, wherein an edit distance between any two index sequences of the set of index sequences is not less than a third criterion value.
- 25 23. The set of oligonucleotides of claim 22, wherein the edit distance is a modified Levenshtein distance where end gaps are assigned no penalty.
24. The set of oligonucleotides of claim 22, wherein the third criterion value is 3.
- 30 25. The set of oligonucleotides of claim 22, wherein the third criterion value is 2.

26. The set of oligonucleotides of claim 22, wherein:
each index sequence of the set of index sequences has 8 bases;
the first criterion value is 3; and
5 the third criterion is 2.
27. The set of oligonucleotides of claim 26, wherein the set of index sequences
comprise sequences listed under Example 2.
- 10 28. The set of oligonucleotides of claim 22, wherein:
each index sequence of the set of index sequences has 10 bases;
the first criterion value is 4; and
the third criterion is 3.
- 15 29. The set of oligonucleotides of any of the preceding claims, wherein the set of
index sequences excludes any subsequence of sequences of adapters or primers in a
sequencing platform, or a reverse complement of the subsequence.
30. The set of oligonucleotides of any of the preceding claims, wherein the sequences
20 of adapters or primers in the sequencing platform comprise SEQ ID NO: 1
(AGATGTGTATAAGAGACAG), SEQ ID NO: 3 (TCGTCGGCAGCGTC), SEQ ID
NO: 5 (CCGAGCCCACGAGAC), SEQ ID NO: 7
(CAAGCAGAAGACGGCATACGAGAT), and SEQ ID NO: 8
(AATGATACGGCGACCACCGAGATCTACAC).
- 25 31. The set of oligonucleotides of any of the preceding claims, wherein each index
sequence of the set of oligonucleotides has 32 or fewer bases.
32. The set of oligonucleotides of any of the preceding claims, wherein each index
30 sequence of the set of oligonucleotides has 16 or fewer bases.

33. The set of oligonucleotides of any of the preceding claims, wherein each index sequence of the set of oligonucleotides has 10 or fewer bases.
34. The set of oligonucleotides of any of the preceding claims, wherein each index
5 sequence of the set of oligonucleotides has 10 bases.
35. The set of oligonucleotides of any of the preceding claims, wherein each index sequence of the set of oligonucleotides has 8 or fewer bases.
- 10 36. The set of oligonucleotides of any of the preceding claims, wherein each index sequence of the set of oligonucleotides has 8 bases.
37. The set of oligonucleotides of any of the preceding claims, wherein each index sequence of the set of oligonucleotides has 7 or fewer bases.
- 15 38. The set of oligonucleotides of any of the preceding claims, wherein each index sequence of the set of oligonucleotides has 6 or fewer bases.
39. The set of oligonucleotides of any of the preceding claims, wherein each index
20 sequence of the set of oligonucleotides has 5 or fewer bases.
40. The set of oligonucleotides of any of the preceding claims, wherein each index sequence of the set of oligonucleotides has 4 or fewer bases.
- 25 41. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences excludes index sequences that were empirically determined to have poor performance of indexing sources of nucleic acid samples in multiplex massively parallel sequencing.
- 30 42. The set of oligonucleotides of claim 41, wherein the excluded index sequences comprise sequences in Table 4.

43. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences comprising at least 12 different index sequences.
- 5 44. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences comprising at least 20 different index sequences.
45. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences comprising at least 24 different index sequences.
- 10 46. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences comprising at least 28 different index sequences.
47. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences comprising at least 48 different index sequences.
- 15 48. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences comprising at least 96 different index sequences.
- 20 49. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences comprising at least 384 different index sequences.
50. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences excludes any homopolymers having four or more consecutive identical bases.
- 25 51. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences excludes index sequences matching or reverse complementing one or more sequencing primer sequences.

30

52. The set of oligonucleotides of claim 51, wherein the sequencing primer sequences are comprised in the sequences of the oligonucleotides.

53. The set of oligonucleotides of any of the preceding claims, wherein the set of
5 index sequences excludes index sequences matching or reverse complementing one or more flow cell amplification primer sequences.

54. The set of oligonucleotides of claim 53, wherein the flow cell amplification primer sequences are comprised in the sequences of the oligonucleotides.
10

55. The set of oligonucleotides of any of the preceding claims, wherein the set of index sequences comprises index sequences having a same number of bases.

56. The set of oligonucleotides of any of the preceding claims, wherein each index
15 sequence of the set of index sequences has a combined number of guanine and cytosine bases between 2 and 6.

57. The set of oligonucleotides of any of the preceding claims, wherein each index sequence of the set of index sequences has a guanine/cytosine (GC) content between 25%
20 and 75%.

58. The set of oligonucleotides of any of the preceding claims, wherein the set of oligonucleotides comprises DNA oligonucleotides or RNA oligonucleotides.

25 59. The set of oligonucleotides of any of the preceding claims, wherein the set of oligonucleotides are provided in a container comprising multiple separate compartments.

60. The set of oligonucleotides of claim 59, wherein the container comprises a multi-well plate.
30

61. The set of oligonucleotides of claim 59, wherein each compartment contains a plurality of oligonucleotides comprising one index sequence of the set of index sequences, the one index sequence being different from index sequences contained in other compartments.

5

62. The set of oligonucleotides of claim 59, wherein each compartment contains (a) a first plurality of oligonucleotides comprising a first index sequence of the set of index sequences and (b) a second plurality of oligonucleotides comprising a second index sequence of the set of index sequences, an ordered combination of (a) and (b) in the compartment being different from ordered combinations of (a) and (b) in any other compartments.

10

63. The set of oligonucleotides of claim 62, wherein the first index sequence in each compartment is different from the first index sequence in any other compartment, and wherein the second index sequence in each compartment is different from the second index sequence in any other compartment.

15

64. The set of oligonucleotides of claim 62, wherein the first and the second index sequences in a compartment respectively are: the n^{th} 10-mer in SEQ ID NO: 10 and the n^{th} 10-mer in SEQ ID NO: 11 (or a reverse complement thereof); the n^{th} 10-mer in SEQ ID NO: 12 and the n^{th} 10-mer in SEQ ID NO: 13 (or a reverse complement thereof); the n^{th} 10-mer in SEQ ID NO: 14 and the n^{th} 10-mer in SEQ ID NO: 15 (or a reverse complement thereof); the n^{th} 10-mer in SEQ ID NO: 16 and the n^{th} 10-mer in SEQ ID NO: 17 (or a reverse complement thereof); the n^{th} 10-mer in SEQ ID NO: 18 and the n^{th} 10-mer in SEQ ID NO: 19 (or a reverse complement thereof).

20

25

65. The set of oligonucleotides of claim 62, wherein the first plurality of oligonucleotides comprises a P5 flow cell amplification primer binding site and the second plurality of oligonucleotides comprises a P7' flow cell amplification primer binding site.

30

66. The set of oligonucleotides of claim 62, wherein the first plurality of oligonucleotides comprises an i5 index sequence and the second plurality of oligonucleotides comprises an i7 index sequence.

5 67. The set of oligonucleotides of claim 62, wherein each index sequence comprised in the first plurality of oligonucleotides is selected from a first subset of the set of index sequences, and each index sequence comprised in the second plurality of oligonucleotides is selected from a second subset of the set of index sequences, the first subset not overlapping the second subset.

10

68. The set of oligonucleotides of claim 67, wherein a Hamming distance between any two index sequences in the first subset or between any two index sequences in the second subset is not less than a second criterion value.

15 69. The set of oligonucleotides of claim 67, wherein the second criterion value is larger than the first criterion value.

70. The set of oligonucleotides of claim 69, wherein the first criterion value is 4, and the second criterion value is 5.

20

71. The set of oligonucleotides of claim 70, wherein the first subset is a subset selected from Table 1 and the second subset is a subset selected from Table 2.

25 72. The set of oligonucleotides of claim 62, wherein index sequences comprised in the first plurality of oligonucleotides and the second plurality of oligonucleotides are selected from a subset of the set of index sequences.

73. The set of oligonucleotides of claim 72, wherein a Hamming distance between any two index sequences in the subset is not less than a second criterion value.

30

74. The set of oligonucleotides of claim 73, wherein the second criterion value is larger than the first criterion value.
75. The set of oligonucleotides of claim 74, wherein the first criterion value is 4, and
5 the second criterion value is 5.
76. The set of oligonucleotides of claim 75, wherein the subset is selected from a subset in Table 3.
- 10 77. The set of oligonucleotides of claim 62, wherein the multiple separate compartments are arranged in an array of one or more rows of compartments and one or more columns of compartments.
78. The set of oligonucleotides of claim 77, wherein each 1/n row and/or each 1/m
15 column of compartments contain oligonucleotides comprising at least one pair of color-balanced index sequences, wherein n and m are each selected from integers in a range of 1 to 24.
79. The set of oligonucleotides of claim 77, wherein the multiple separate
20 compartments are arranged in an array of 8 rows and 12 columns or in an array of 16 rows and 24 columns.
80. The set of oligonucleotides of claim 79, wherein each $\frac{1}{4}$ row of compartments contain oligonucleotides comprising at least one pair of color-balanced index sequences
25 and wherein each $\frac{1}{4}$ column of compartments contain oligonucleotides comprising at least one pair of color-balanced index sequences.
81. The set of oligonucleotides of claim 80, wherein the multiple separate compartments are arranged in an array of A-H rows and 1-12 columns, and wherein for
30 the n^{th} compartment on the list A1, A2, A3,..., A12, B1, B2,..., B12, ..., H1, H2,..., H12, the first and the second index sequences respectively are: the n^{th} 10-mer in SEQ ID

NO: 10 and n^{th} 10-mer in SEQ ID NO: 11 (or a reverse complement thereof); the n^{th} 10-mer in SEQ ID NO: 12 and n^{th} 10-mer in SEQ ID NO: 13 (or a reverse complement thereof); the n^{th} 10-mer in SEQ ID NO: 14 and n^{th} 10-mer in SEQ ID NO: 15 (or a reverse complement thereof); or the n^{th} 10-mer in SEQ ID NO: 16 and n^{th} 10-mer in SEQ ID NO: 17 (or a reverse complement thereof).

82. The set of oligonucleotides of claim 62, wherein the multiple separate compartments contain oligonucleotides comprising a first plurality of index sequences arranged in a layout shown in Figure 4B and the multiple separate compartments also contain oligonucleotides comprising a second plurality of index sequences arranged in a layout shown in Figure 4C.

83. A set of oligonucleotides consisting essentially of a plurality of subsets of oligonucleotides, wherein:

the set of oligonucleotides is configured to identify sources of nucleic acid samples in multiplex massively parallel sequencing, each of the nucleic acid samples comprising a plurality of nucleic acid molecules;

each subset of the plurality of subsets of oligonucleotides comprises a unique index sequence, index sequences of the plurality of subsets consisting of a set of index sequences;

a Hamming distance between any two index sequences of the set of index sequences is not less than a first criterion value, wherein the first criterion value is at least 2; and

the set of index sequences comprises a plurality of pairs of color-balanced index sequences, wherein any two bases at corresponding sequence positions of each pair of color-balanced index sequences include both (i) an adenine base or a cytosine base, and (ii) a guanine base, a thymine base, or a uracil base.

84. A method for making a plurality of oligonucleotides for multiplex massively parallel sequencing, the method comprising:

(a) selecting a set of index sequences from a pool of different index sequences, wherein

the set of index sequences comprise at least 6 different sequences;

5 a Hamming distance between any two index sequences in the set of index sequences is not less than a first criterion value, wherein the first criterion value is at least 2, and

the set of index sequences comprises a plurality of pairs of color-balanced index sequences, wherein any two bases at corresponding sequence positions of each pair of color-balanced index sequences include both (i) an adenine base or a cytosine base, and (ii) a guanine base, a thymine base, or a uracil base; and

10 (b) synthesizing the plurality of oligonucleotides comprising the set of index sequences.

85. The method of any of claims 84-85, wherein the plurality of oligonucleotides comprises double-stranded or Y-shaped sequencing adapters, each strand of each sequencing adapter comprises an index sequence of the set of index sequences.

86. The method of claim 85, wherein each Y-shaped sequencing adapter comprises a first strand comprising an index sequence selected from a first subset of the set of index sequences and a second strand comprising an index sequence selected from a second subset of the set of index sequences, the first subset not overlapping the second subset.

87. The method of claim 86, wherein the first strand of each Y-shaped sequencing adapter comprises a P5 flow cell amplification primer binding site, and the second strand of each Y-shaped sequencing adapter comprises a P7' flow cell amplification primer binding site.

88. The method of claim 86, wherein the one or more first subsets comprises index sequences listed in Table 1 and the one or more second subset comprises index sequences listed in Table 2.

89. The method of 85, wherein each double-stranded or Y-shaped sequencing adapter comprises a first strand comprising an index sequence selected from a subset of multiple subsets of the set of index sequences and a second strand comprising an index sequence selected from said subset.

5

90. The method of claim 89, wherein the first strand of each Y-shaped sequencing adapter comprises a P5 flow cell amplification primer binding site, and the second strand of each Y-shaped sequencing adapter comprises a P7' flow cell amplification primer binding site.

10

91. The method of claim 89, wherein the set of index sequences comprises index sequences in Table 3.

92. The method of any of claims 84-91, wherein an oligonucleotide of the plurality of oligonucleotides comprises an index sequence on its 3' end and an index sequence on its 5' end.

15

93. The method of any of claims 84-92, wherein the set of index sequences is comprised in double-stranded or Y-shaped sequencing adapters, each sequencing adapter comprising an index sequence in only one strand.

20

94. The method of any of claims 84-93, wherein the plurality of oligonucleotides comprises polymerase chain reaction (PCR) primers or reverse complements thereof.

95. The method of any of claims 84-94, wherein the set of index sequences excludes any subsequence of sequences of adapters or primers in a sequencing platform, or a reverse complement of the subsequence.

25

96. The method of any of claims 84-95, wherein the sequences of adapters or primers in the sequencing platform comprise SEQ ID NO: 1 (AGATGTGTATAAGAGACAG), SEQ ID NO: 3 (TCGTCGGCAGCGTC), SEQ ID NO: 5 (CCGAGCCCACGAGAC),

30

SEQ ID NO: 7 (CAAGCAGAAGACGGCATACGAGAT), and SEQ ID NO: 8 (AATGATACGGCGACCACCGAGATCTACAC).

97. The method of any of claims 84-96, wherein step (a) comprises:
- 5 (i) selecting a candidate set of index sequences from the pool of index sequences;
- (ii) separating the selected candidate set into a plurality of groups of color-balanced pairs of index sequences; and
- (iii) partitioning each group into two subgroups of color-balanced pairs using a bipartite graph matching algorithm, wherein each color-balanced pair is a node in the
- 10 bipartite graph.
98. The method of claim 97, wherein two nodes are connected if the Hamming distance between index sequences of the two nodes is less than a second criterion value, wherein the second criterion value is larger than the first criterion value.
- 15 99. The method of claim 98, wherein the first criterion value is 4 and the second criterion value is 5.
100. The method of claim 97, wherein step (i) comprises:
- 20 (1) adding to the candidate set a randomly chosen pair of color-balanced index sequences from the pool of index sequences, wherein the pool includes all possible n-mers;
- (2) sorting index sequences remaining in the pool of index sequences based on minimum Hamming distance to members in the candidate set;
- 25 (3) removing any remaining index sequence whose minimum Hamming distance to the members in the candidate set is less than the first criterion value or minimum edit distance to the members in the candidate set is less than a third criterion; and
- repeating steps (1)-(3) to maximize a size of the candidate set.
- 30 101. The method of claim 100, wherein the first criterion value is 4 and the third criterion value is 3.

102. The method of claim 100, wherein the n-mers comprise 8-mers, 9-mers, or 10-mers.
- 5 103. The method of claim 100, wherein the first criterion value is 3, the third criterion value is 2, and the n-mers comprise 8-mers.
104. The method of claim 97, wherein step (ii) comprises:
(1) randomly selecting a seed for each of the plurality of groups; and
10 (2) greedily expanding each of the plurality of groups.
105. The method of claim 97, further comprising, before step (i), removing a subset of index sequences from the pool of index sequences.
- 15 106. The method of claim 105, wherein the removed subset of index sequences comprises index sequences having four or more consecutive identical bases.
107. The method of claim 105, wherein the removed subset of index sequences comprises index sequences having a combined number of guanine and cytosine bases
20 smaller than 2 or larger than 6.
108. The method of claim 105, wherein the removed subset of index sequences comprises index sequences having a G/C content smaller than 25% or larger than 75%.
- 25 109. The method of claim 105, wherein the removed subset of index sequences comprises index sequences having a sequence matching or reverse complementing one or more sequencing primer sequences.
110. The method of claim 109, wherein the sequencing primer sequences are
30 comprised in the sequences of the index oligonucleotides.

111. The method of claim 105, wherein the removed subset of index sequences comprises index sequences having a sequence matching or reverse complementing one or more flow cell amplification primer sequences.

- 5 112. The method of claim 111, wherein the flow cell amplification primer sequences are comprised in the sequences of the index oligonucleotides.

113. The method of claim 105, wherein the removed subset of index sequences comprises index sequences that were empirically determined to have poor performance in
10 indexing sources of nucleic acid samples in multiplex massively parallel sequencing.

114. The method of claim 113, wherein the removed subset of index sequences comprises sequences in Table 4.

- 15 115. A computer program product comprising a non-transitory machine readable medium storing program code that, when executed by one or more processors of a computer system, causes the computer system to implement a method for selecting a set of index sequences to be incorporated in a set of oligonucleotides configured to be used to in multiplex massively parallel sequencing, said program code comprising:

- 20 (a) code for adding to a candidate set of index sequences a randomly chosen pair of color-balanced index sequences from a pool of different index sequences, wherein any two bases at corresponding sequence positions of each pair of color-balanced index sequences include both (i) an adenine base or a cytosine base, and (ii) a guanine base, a thymine base, or a uracil base;

- 25 (b) code for sorting index sequences remaining in the pool of index sequences based on minimum Hamming distance to members in the candidate set;

- (c) code for removing any remaining index sequence whose minimum Hamming distance to the members in the candidate set is less than a first criterion value or minimum edit distance to the members in the candidate set is less than a second criterion
30 value;

- (d) code for repeating (a)-(c) to maximize a size of the candidate set; and

(e) code for selecting from the candidate set the set of index sequences to be incorporated into the set of oligonucleotides configured to be used in multiplex massively parallel sequencing.

- 5 116. The computer program product of claim 115, said program code further comprising:

code for separating the candidate set into a plurality of groups; and

code for partitioning each group into two subgroups using a bipartite graph matching algorithm, wherein each index sequence is a node.

10

117. The computer program product of any of claims 115-116, said program code further comprising code for controlling a nucleic acid synthesizer to synthesize the set of oligonucleotides configured to be used in multiplex massively parallel sequencing.