

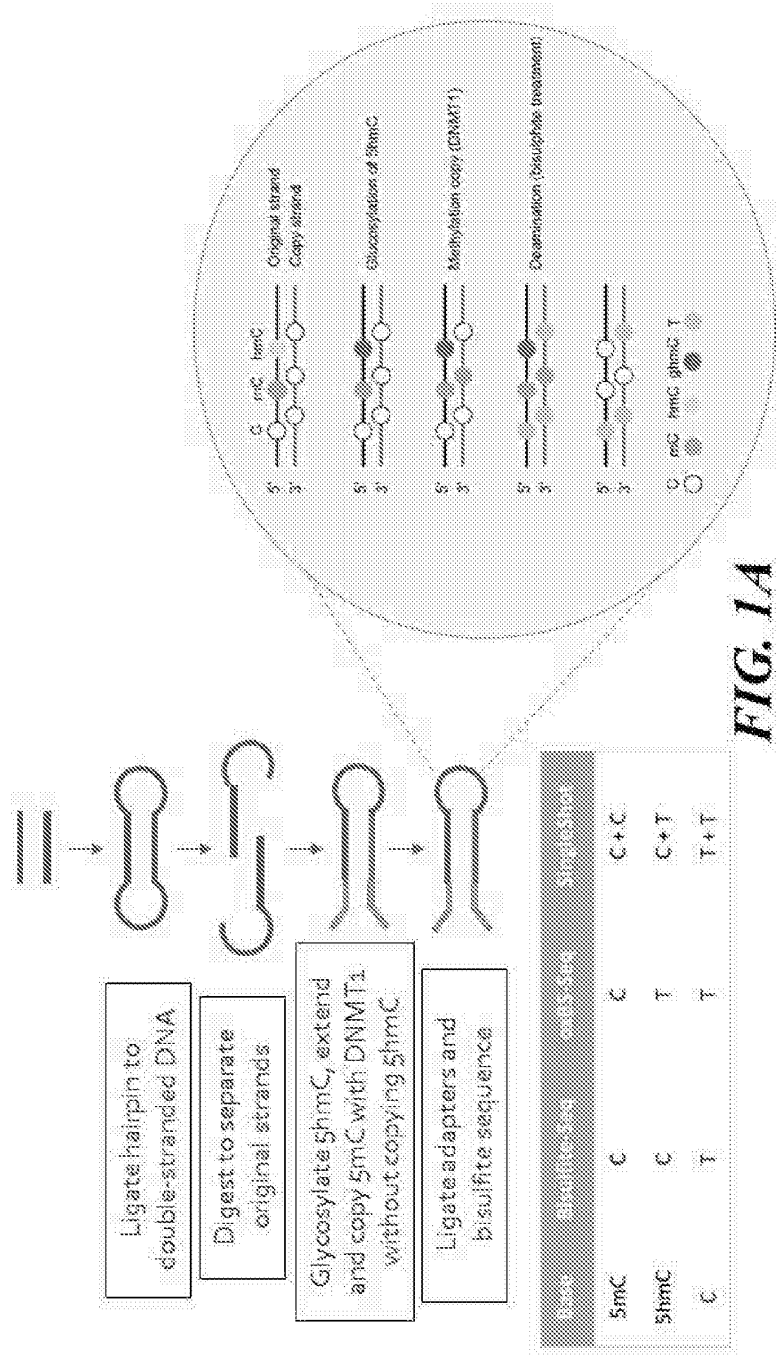


FIG. 1A

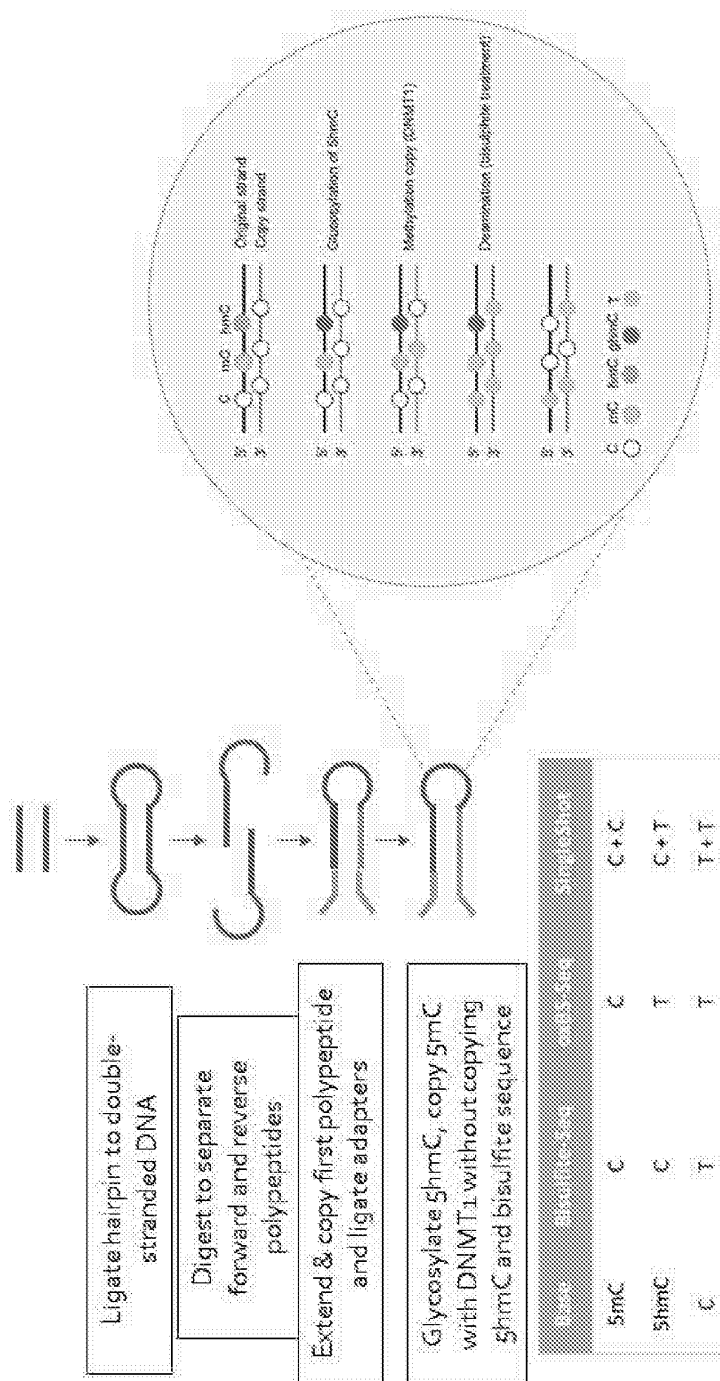
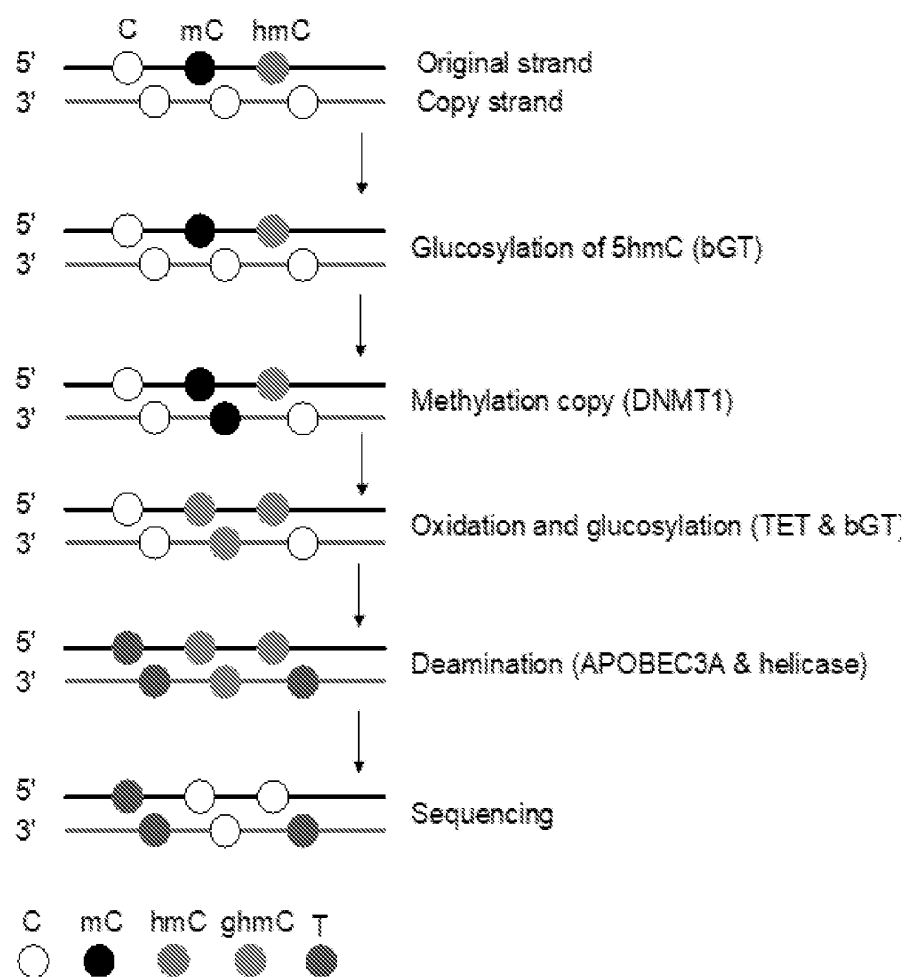
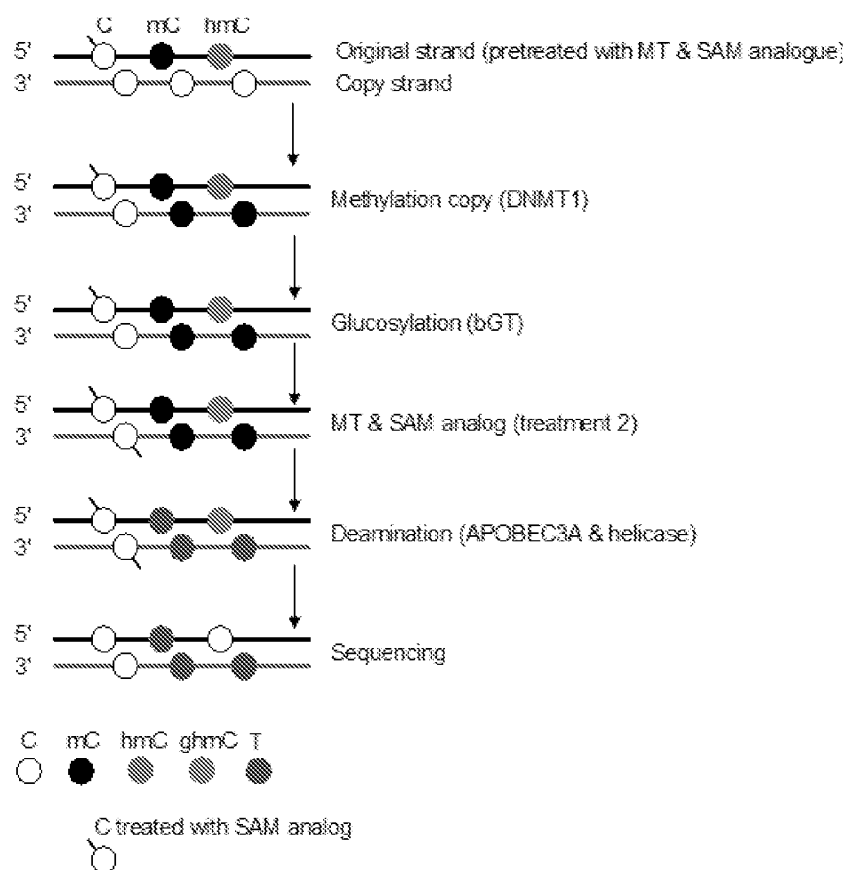


FIG. 1B

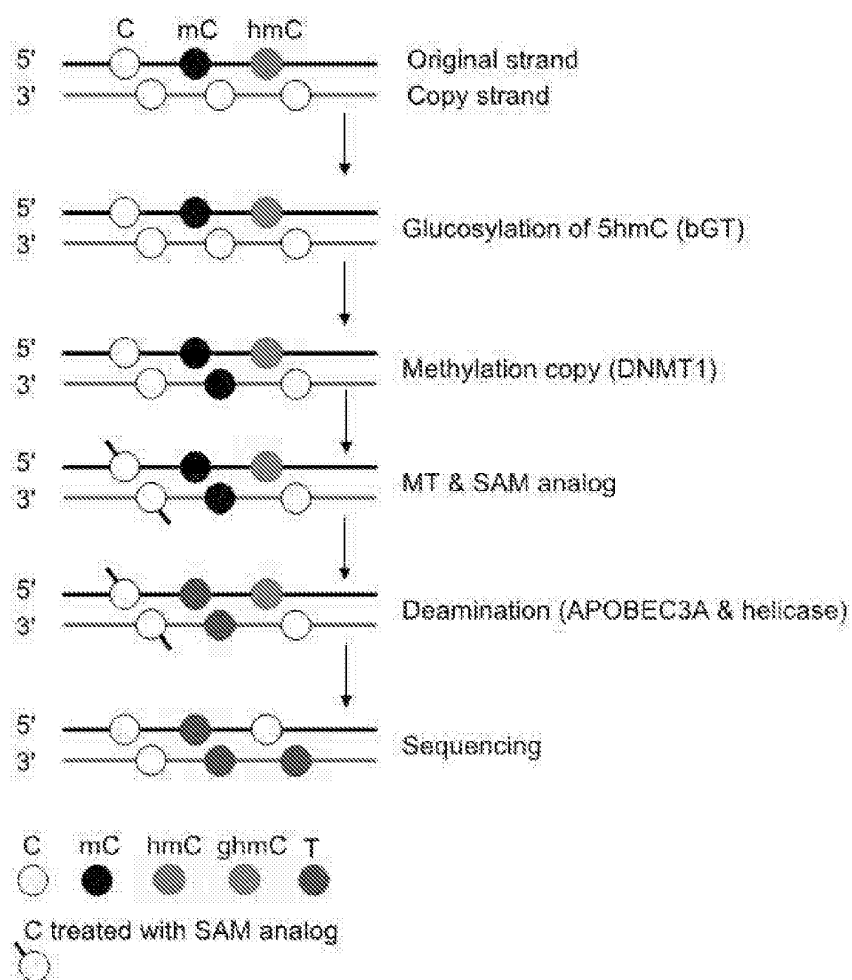
3/62

**FIG. 2A**

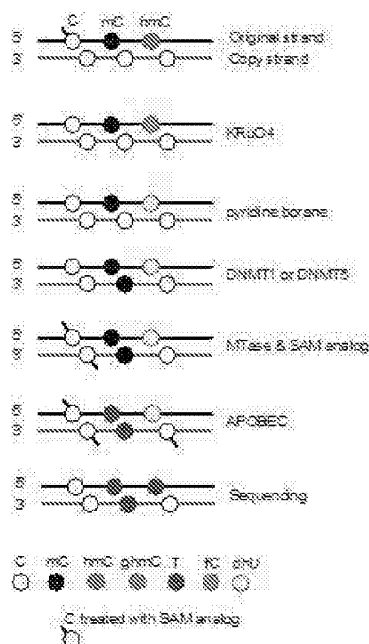
4/62

**FIG. 2B**

5/62

**FIG. 2C**

6/62



- Could also replace the FC → T with this published method: Xia, B. et al. Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. Nat. Methods 12, 1047–1050 (2015).

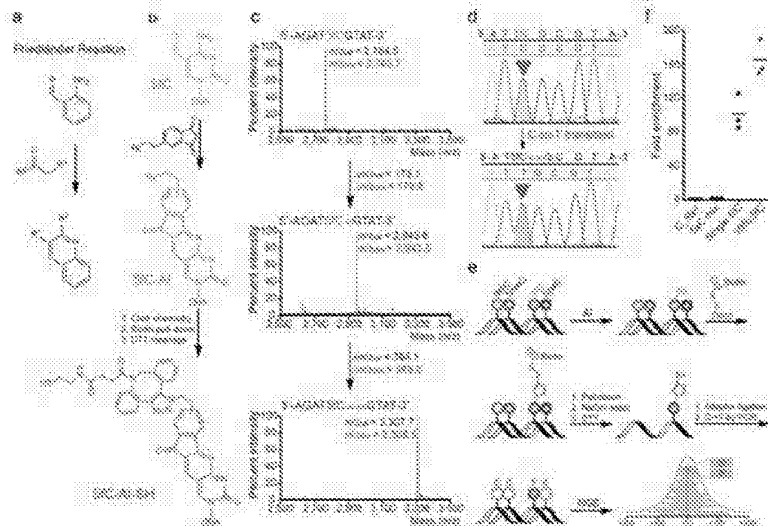
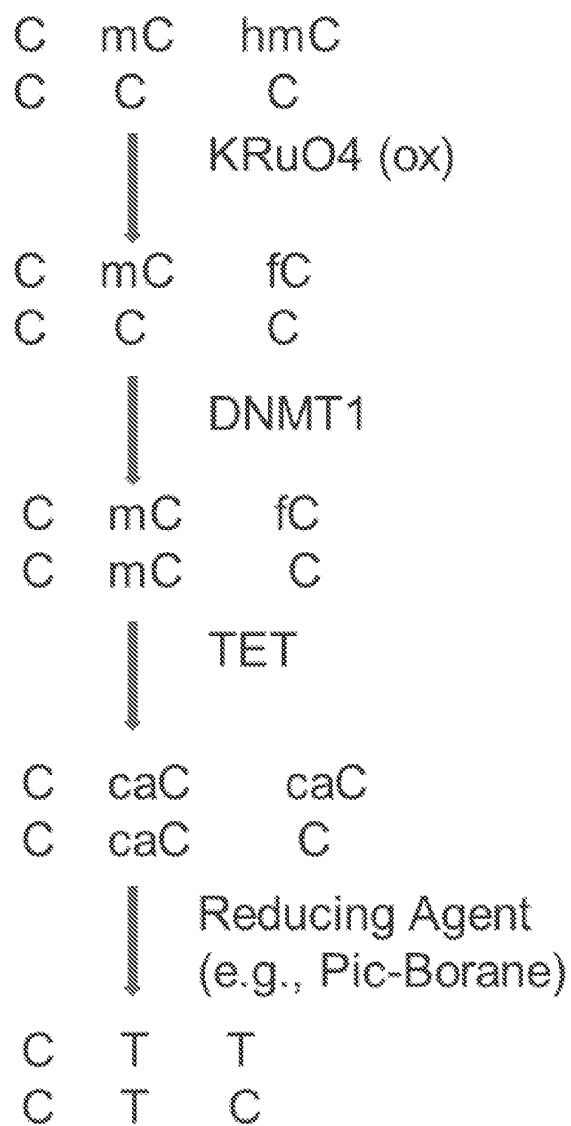
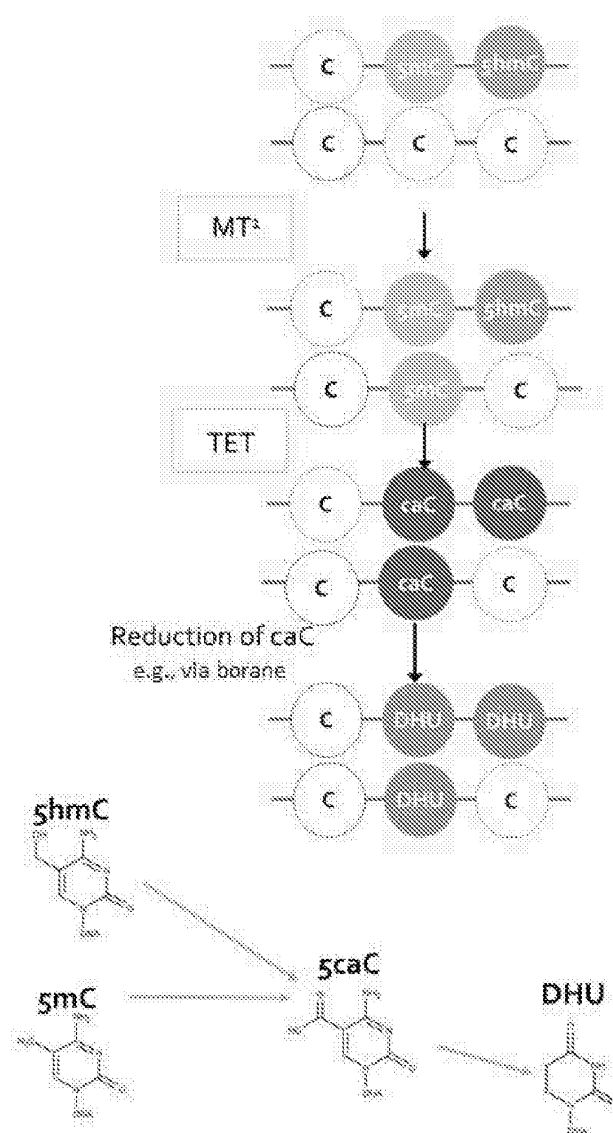


FIG. 2D

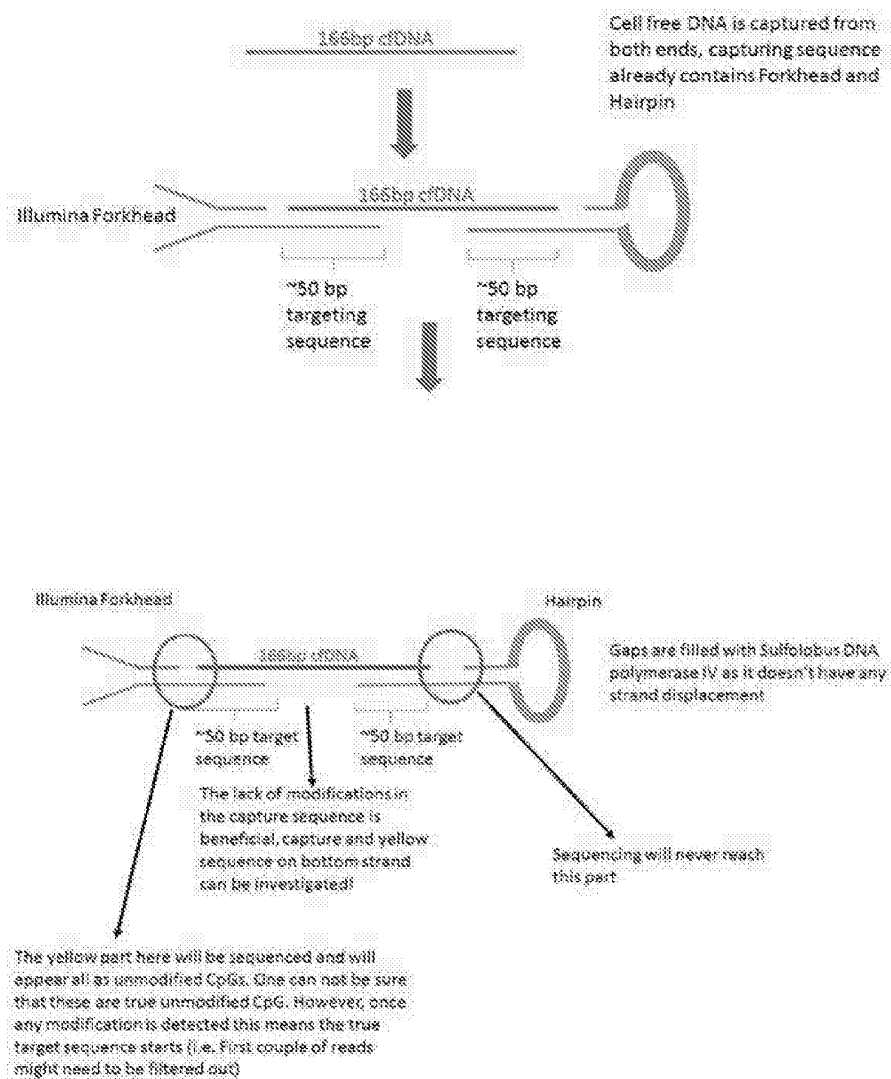
7/62

**FIG. 3**

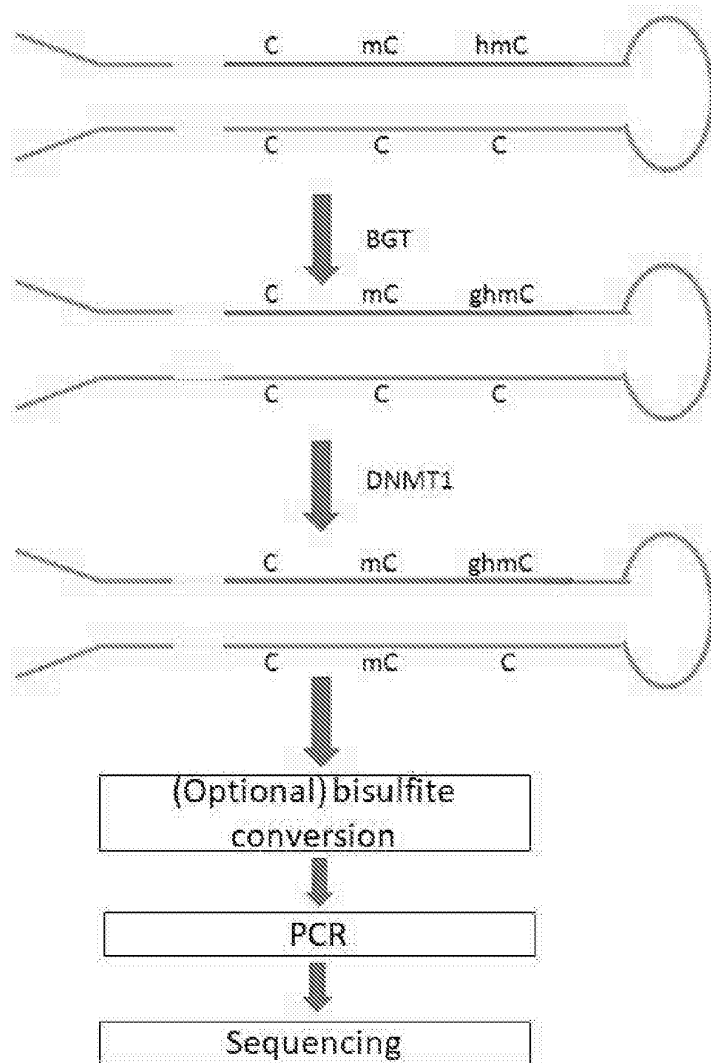
8/62

**FIG. 4**

9/62

**FIG. 5**

10/62

**FIG. 6**

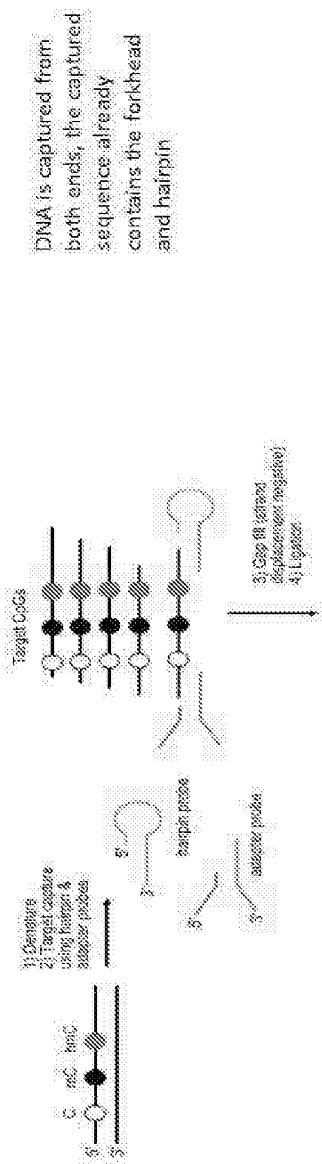


FIG. 7A

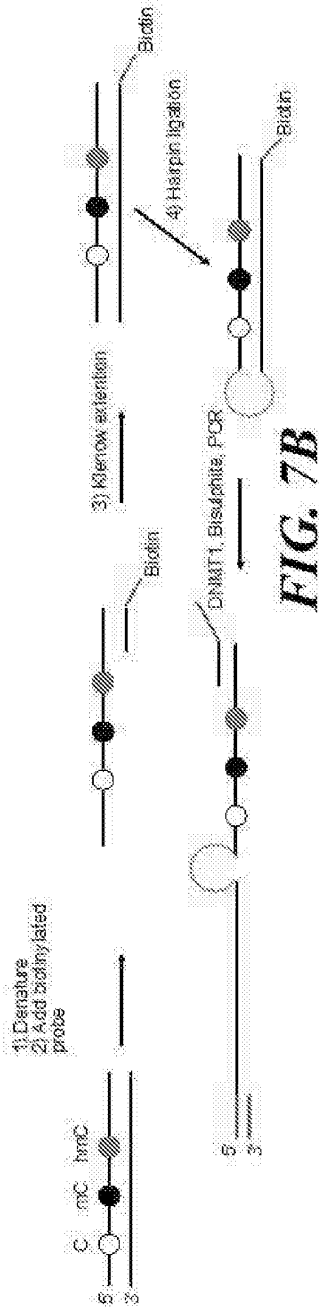


FIG. 7B

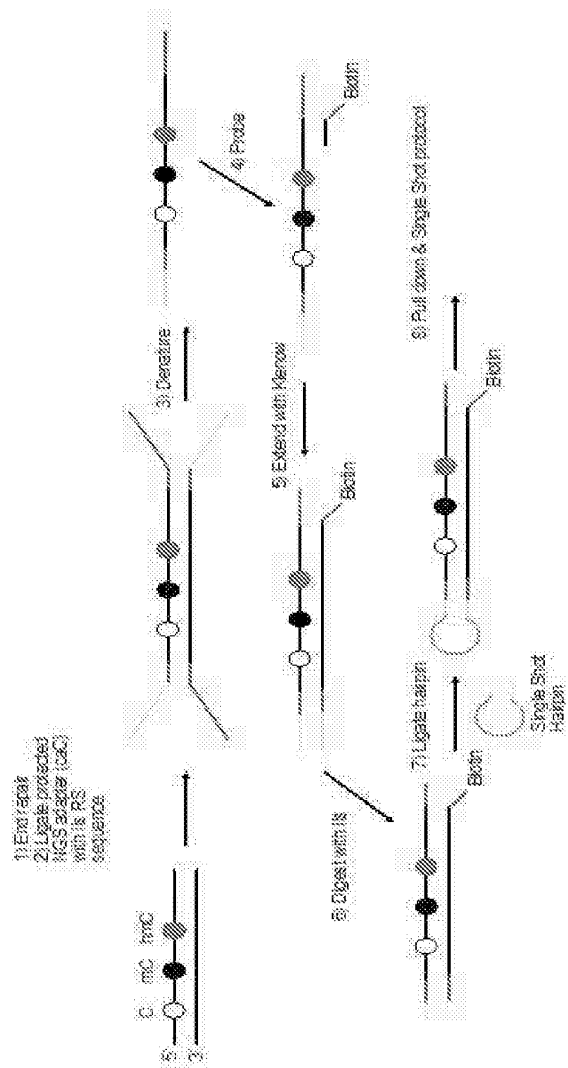
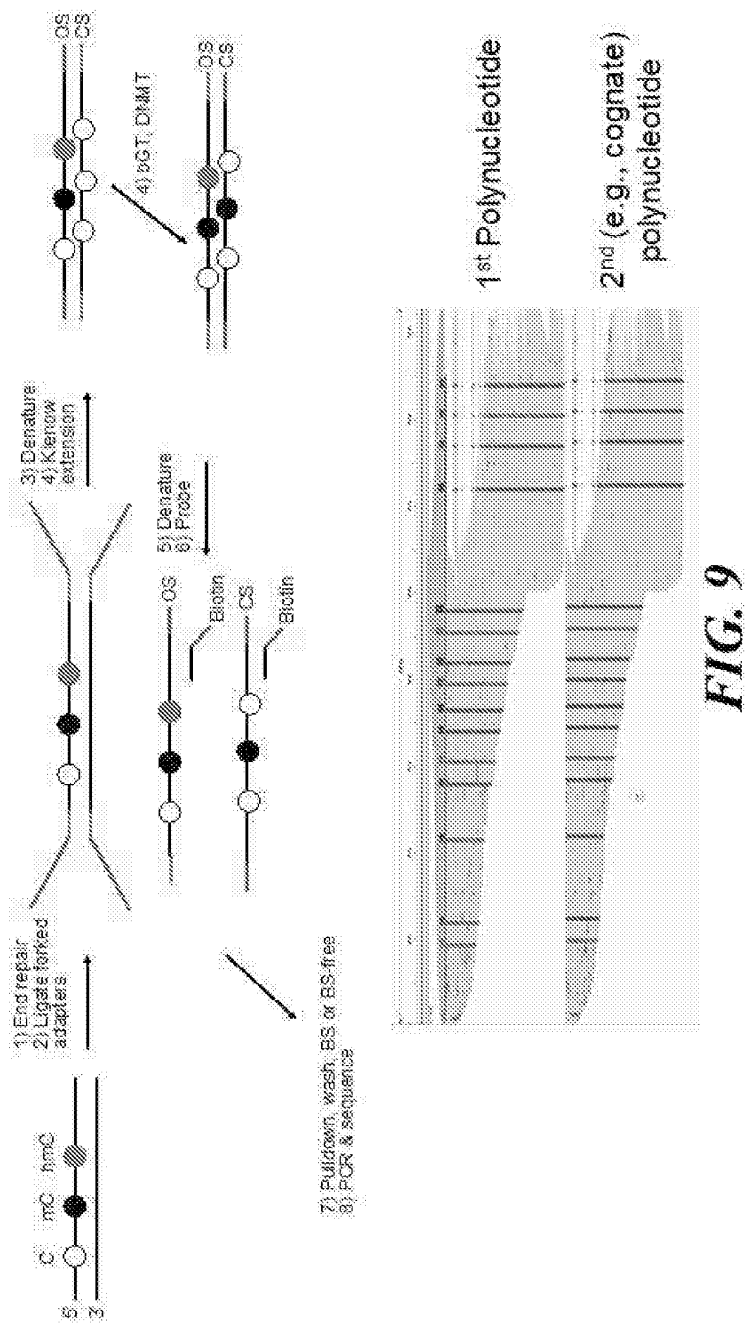
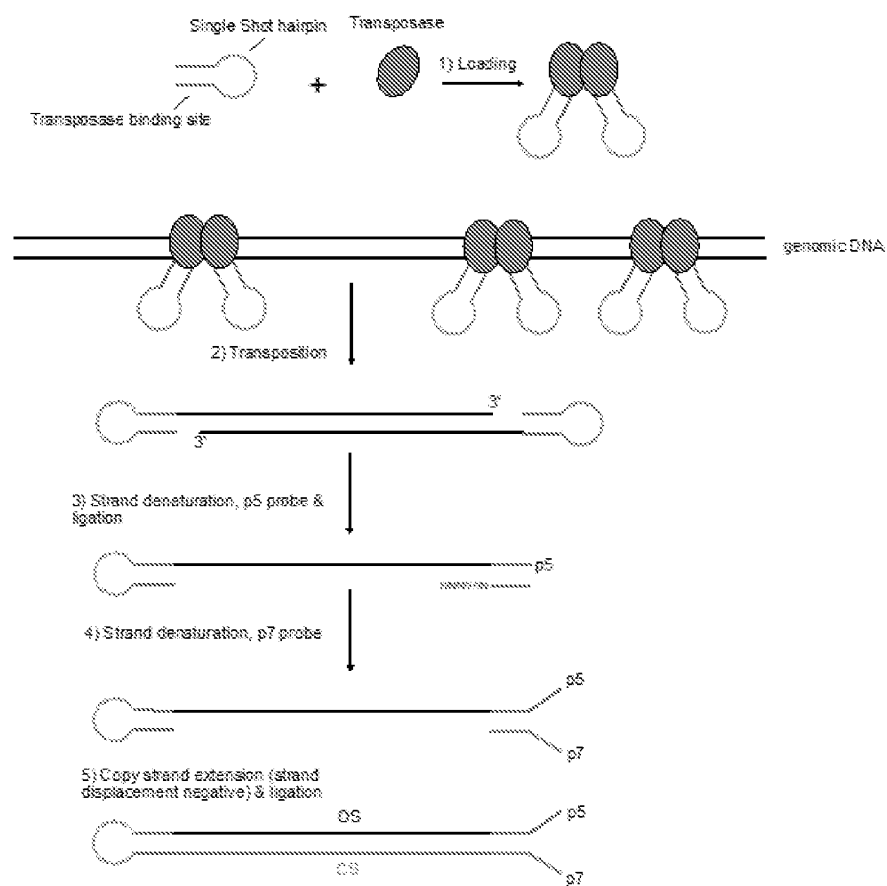


FIG. 8



14/62

**FIG. 10**

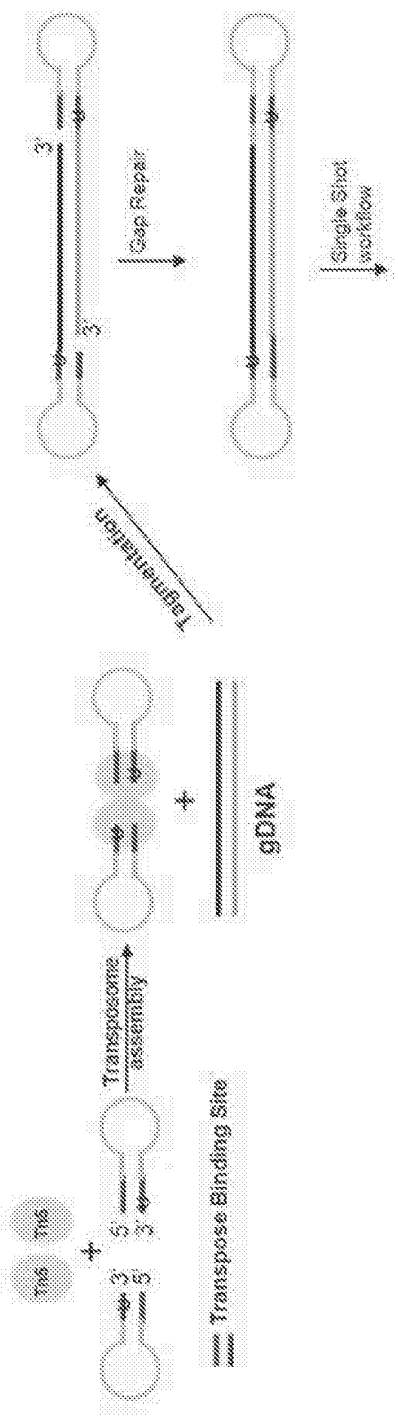
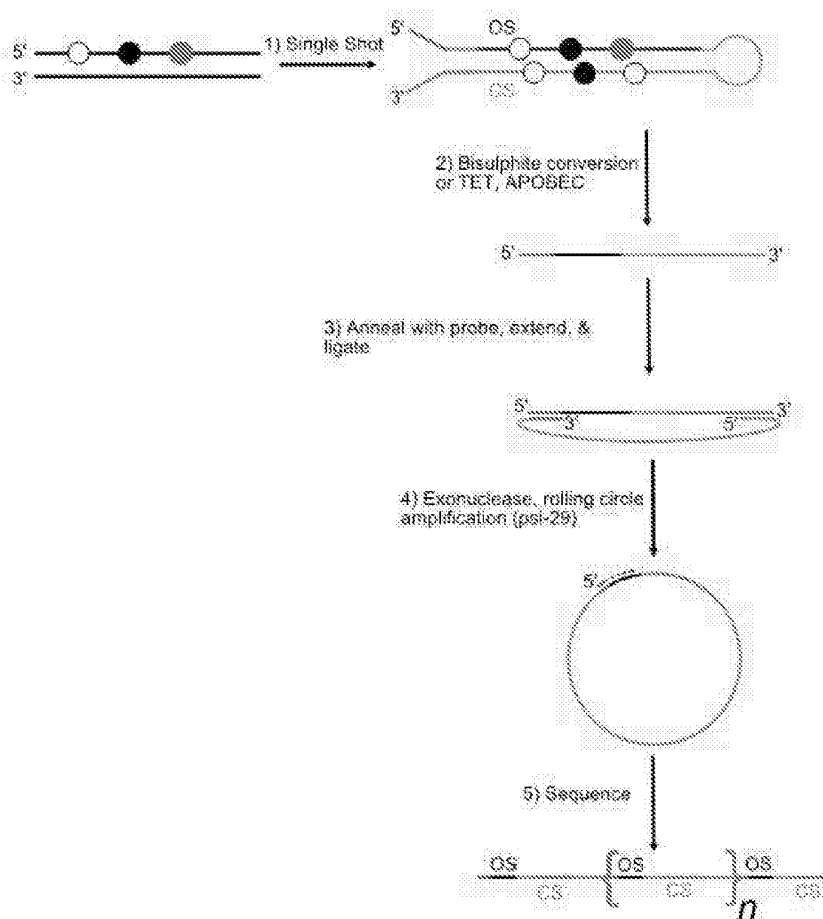


FIG. 11

16/62

**FIG. 12**

Base in reference sequence (position)	R1 Sum of depth	R1 Sum of correct calls	R1 Average base call accuracy (per base call)	R2 Sum of depth	R2 Sum of correct calls	R2 Average base call accuracy (per base call)	R3 Sum of correct calls	R2 Average base call accuracy (per base call)	R1 Average base call accuracy (per base call)
A	2,895,287	2,635,657	98.1%	2,893,383	2,628,299	98.0%		98.0%	97.9%
C (non-CpG)	2,806,775	2,192,040	87.9%	2,198,847	2,161,947	98.3%		98.3%	98.2%
C (CpG)	827,332	478,532	90.4%	524,313	518,329	99.0%		99.0%	98.9%
G (non-CpG)	2,042,811	2,055,925	91.1%	278,589	274,510	98.9%		98.9%	98.9%
G (non-CpG)	34,719	28,469	82.1%	33,921	33,454	98.7%		98.7%	98.8%
C (non-CpG)	283,842	259,458	88.2%	282,841	281,321	99.1%		99.1%	99.4%
C (non-CpG)	87,225	79,323	91.0%	88,083	86,190	98.0%		98.0%	99.4%
G (non-CpG)	1,481,993	1,432,278	97.8%	1,481,891	1,436,237	98.2%		98.2%	97.8%
G (CpG)	828,898	521,998	68.7%	528,139	473,926	90.1%		90.1%	90.5%
G (non-CpG)	293,386	279,917	95.4%	290,590	287,731	99.0%		99.0%	99.0%
G (non-CpG)	287,284	284,202	99.0%	287,289	285,159	99.1%		99.1%	99.1%
T	2,188,891	2,134,588	98.0%	2,139,905	2,112,181	98.7%		98.7%	98.7%

FIG. 13

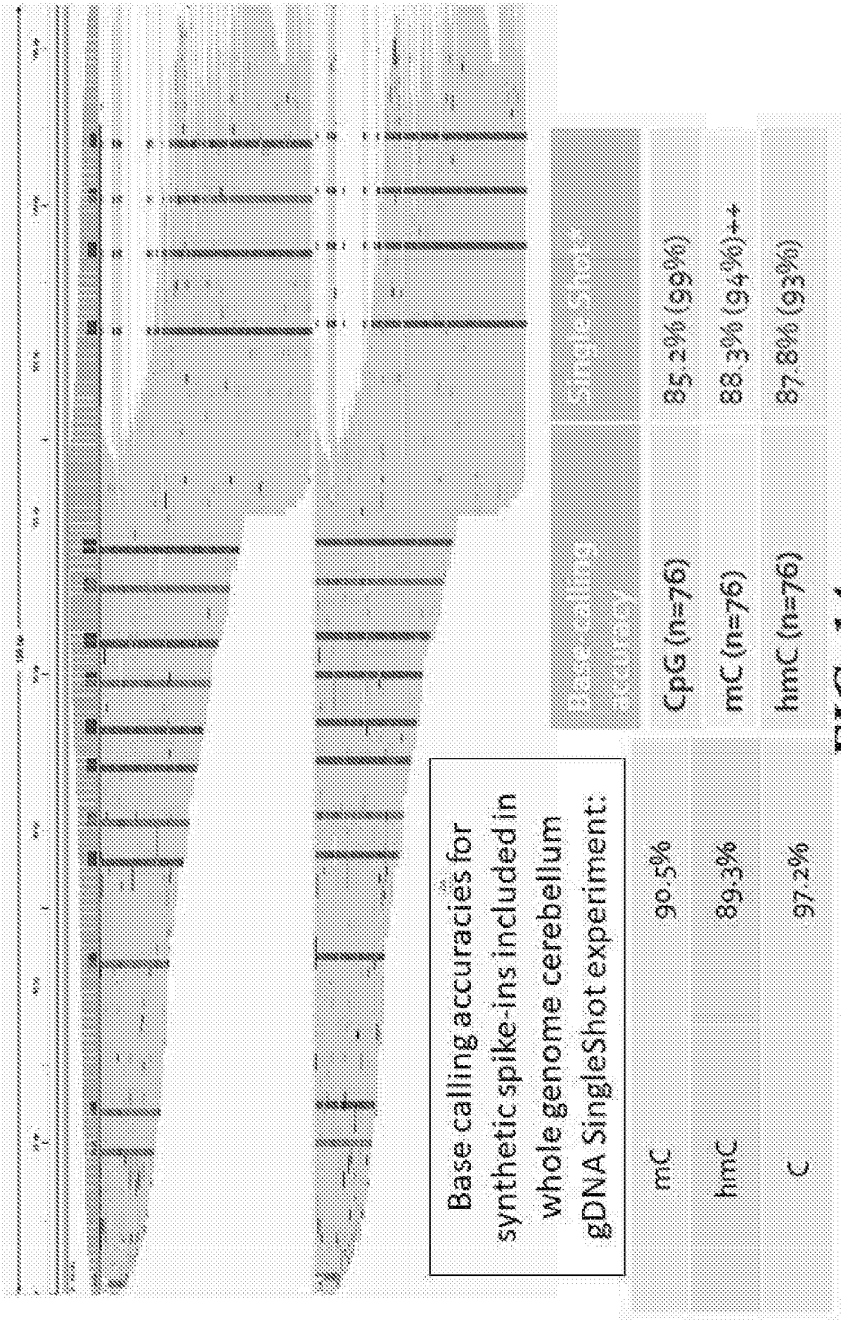
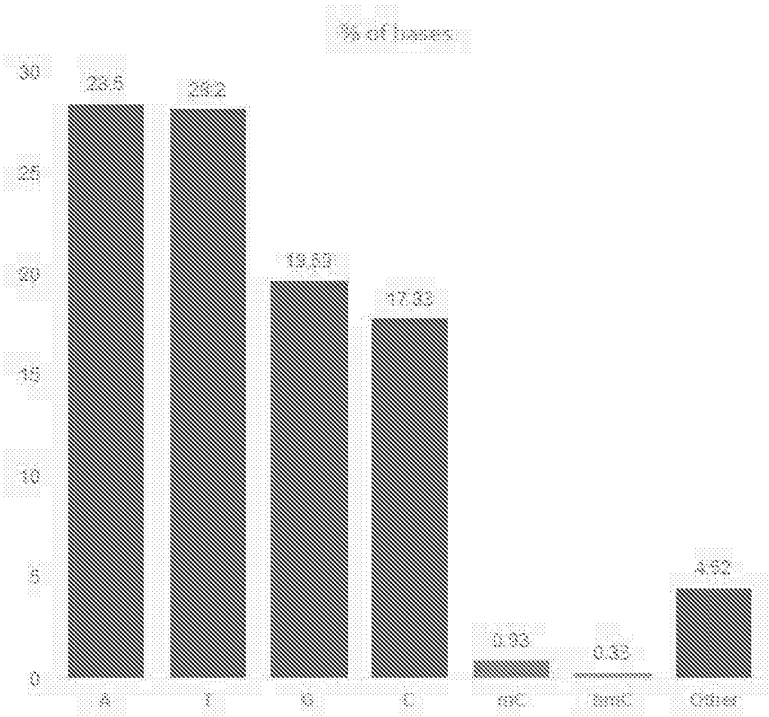


FIG. 14



	Global measures in sample source ¹	SingleShot
mC	4.9% of Cs	5.2% of Cs
hmC	1.14% of Cs	1.8% of Cs

FIG. 15

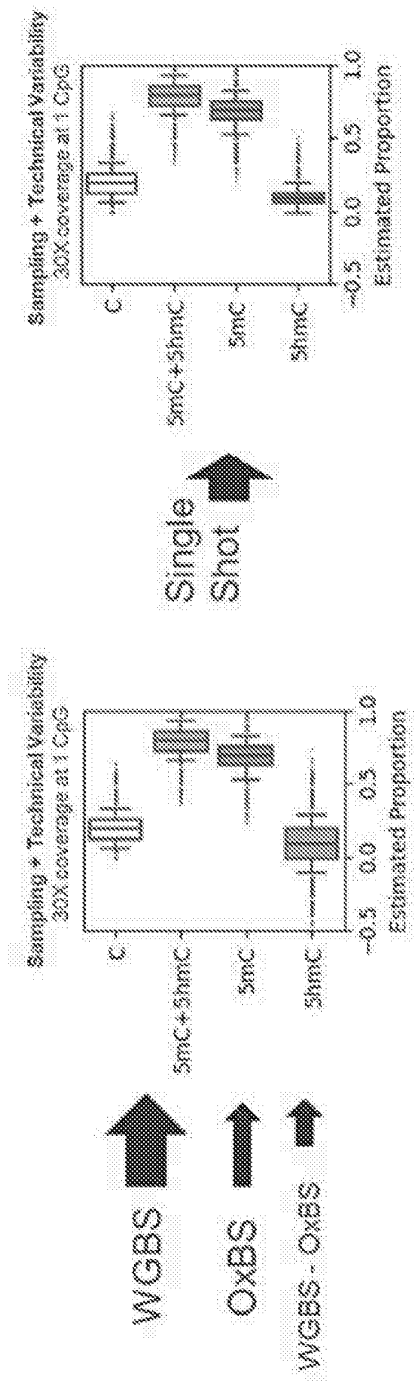


FIG. 16

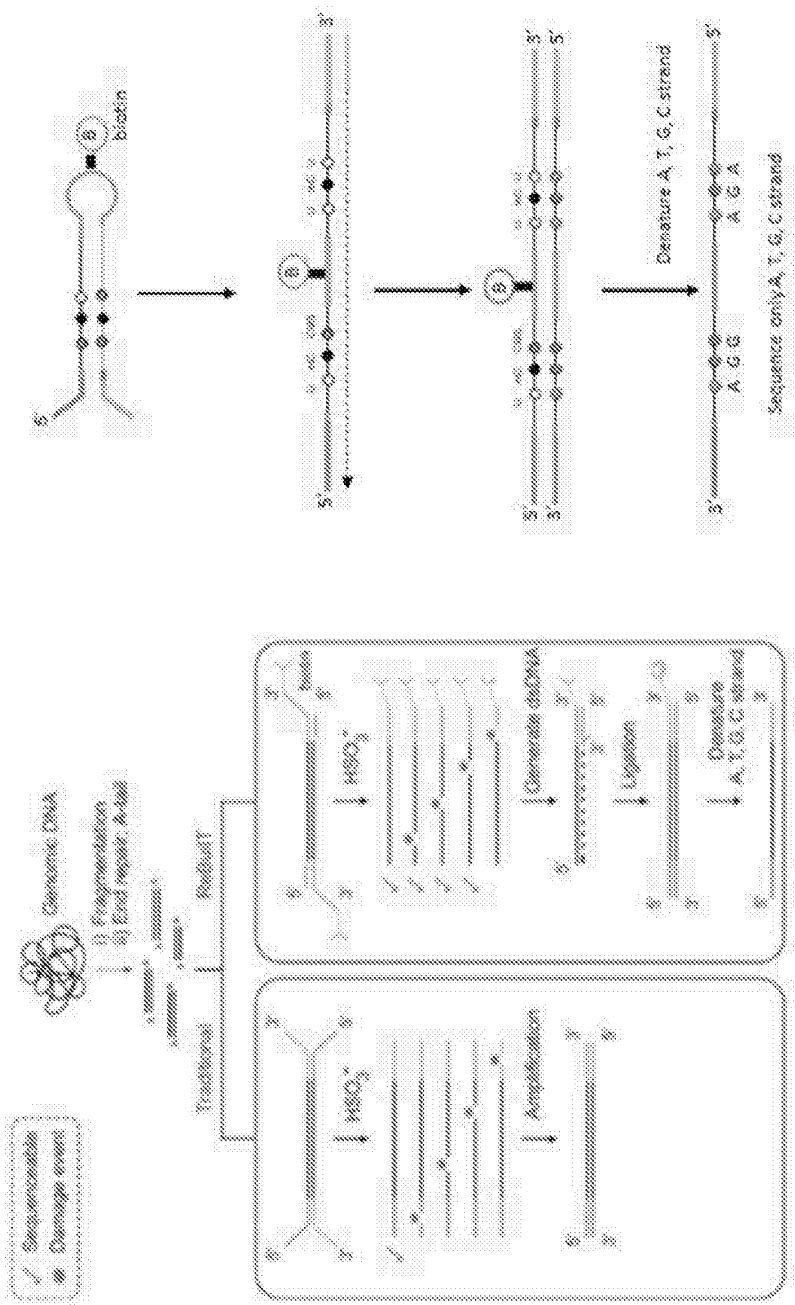


FIG. 17

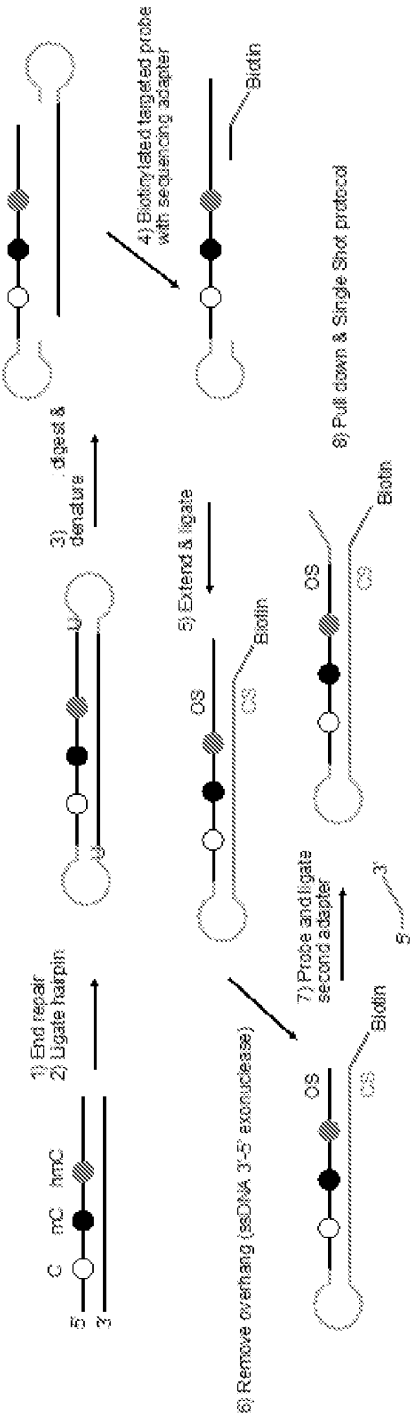
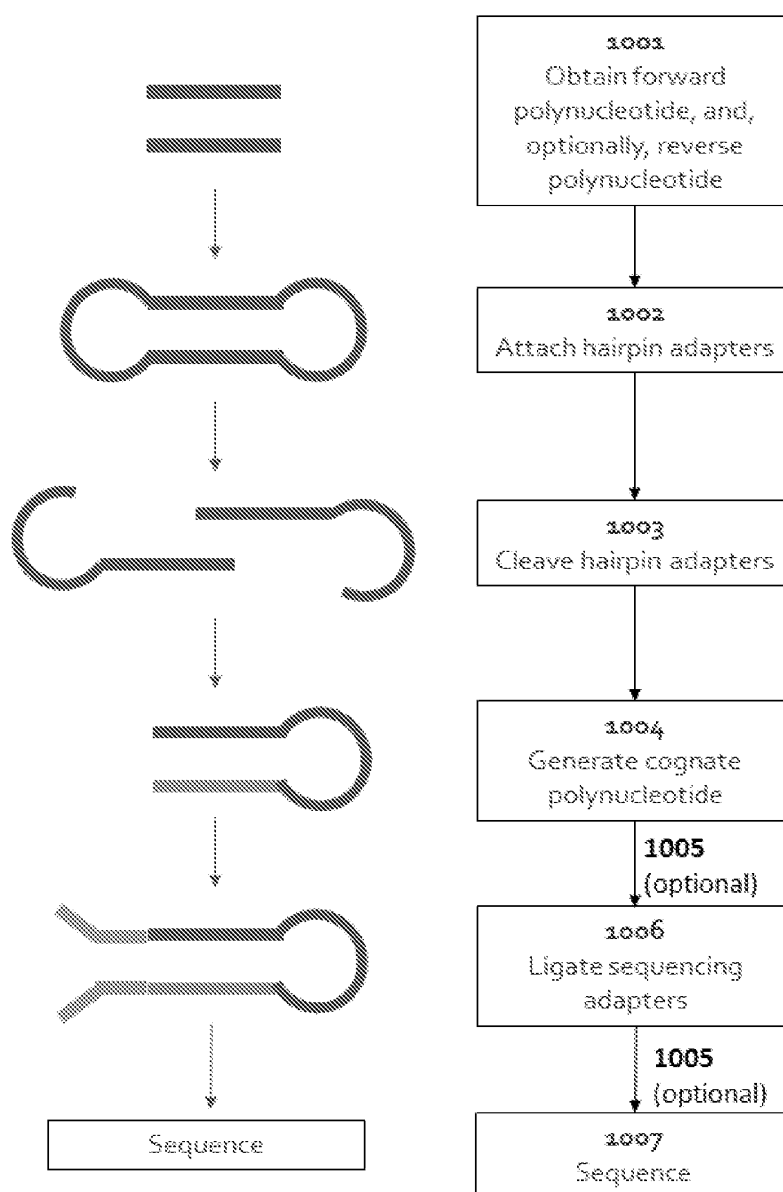
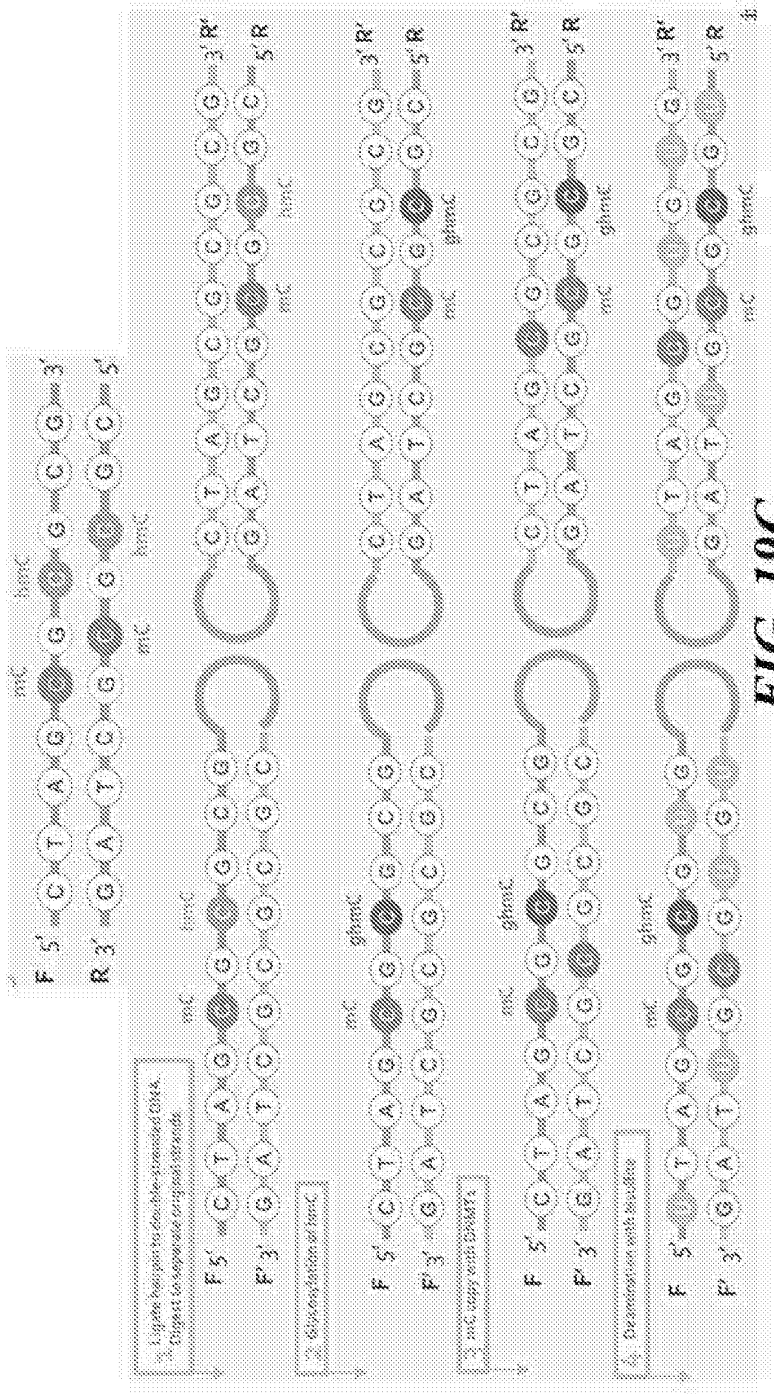


FIG. 18

23/62

**FIG. 19A****FIG. 19B**



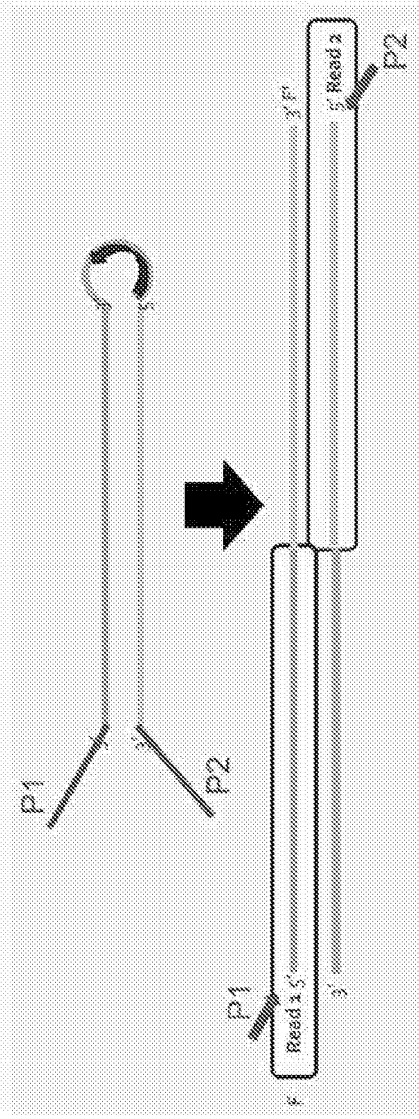


FIG. 19D

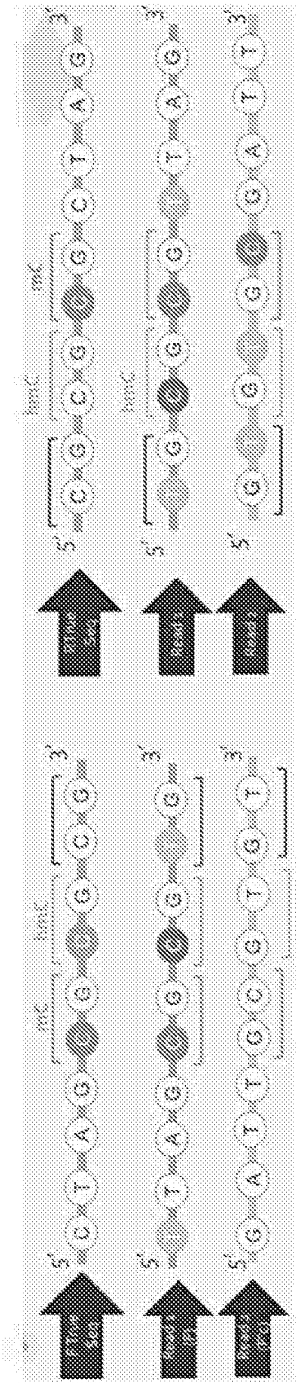


FIG. 19E

True base	F strand	F' strand
A	A	T
C	C	G
G	G	C
T	T	A
Error (F/F')	A/A, A/C, A/G, C/A, C/C, C/T, G/A, G/G, G/T, T/C, T/G, T/T	

FIG. 20A

True base	F strand	F' strand
A	A	T
C	T	G
G	G	T
T	T	A
Methylated C	C	G
Error (F/F')	A/A, A/C, A/G, C/A, C/C, C/T, G/A, G/C, G/G, T/C, T/T	

FIG. 20B

True base	F strand	F' strand
A	A	T
C or 5hmC	T	G
G	G	T
T	T	A
5mC	C	G
Error (F/F')	A/A, A/C, A/G, C/A, C/C, C/T, G/A, G/C, G/G, T/C, T/T	

FIG. 20C

True base	F strand	F' strand
A	A	T
C	C	G
G	G	C
T	T	A
5mCpG	T	G
	G	T
5hmCpG	T	G
	G	C
Error (F/F')	A/A, A/C, A/G, C/A, C/C, C/T, G/A, G/G, T/C, T/T + G/T or T/G in non-CpG context	

FIG. 20D

True base	F strand	F' strand
A	A	T
C	T	G
G	G	T
T	T	A
5mCpG	C	G
	G	C
5hmCpG	C	G
	G	T
Error (F/F')	A/A, A/C, A/G, C/A, C/C, C/T, G/A, G/G, T/C, T/T + C/G or G/C in non-CpG context	

FIG. 20E

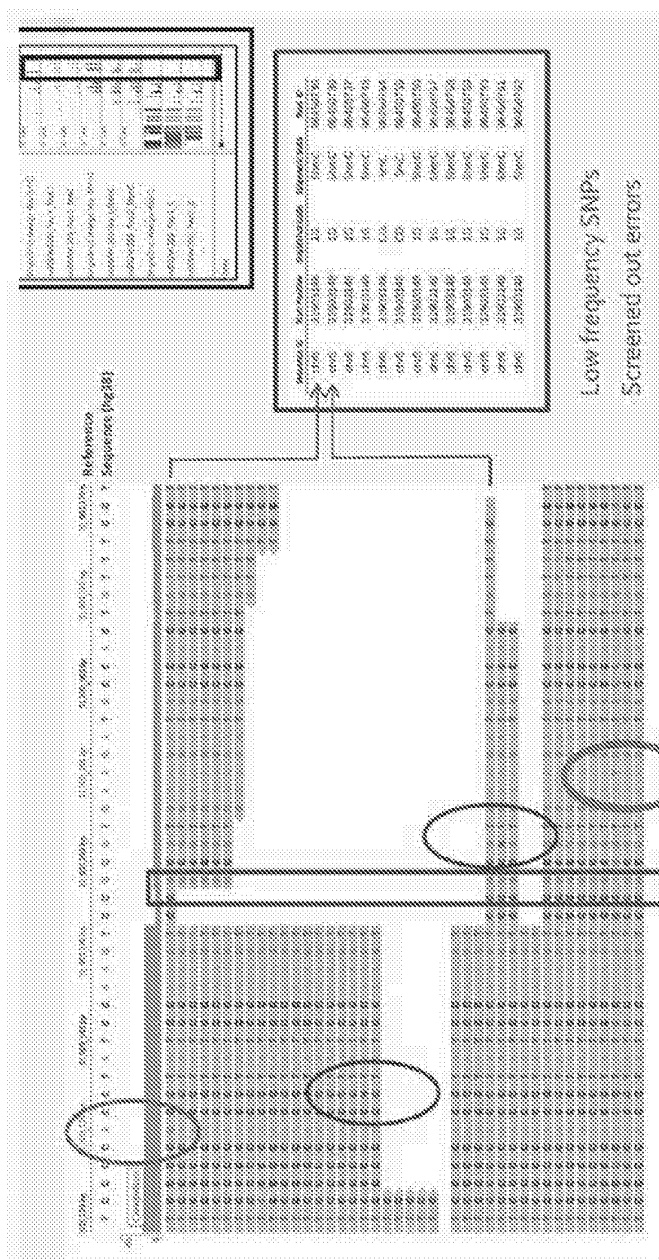


FIG. 21

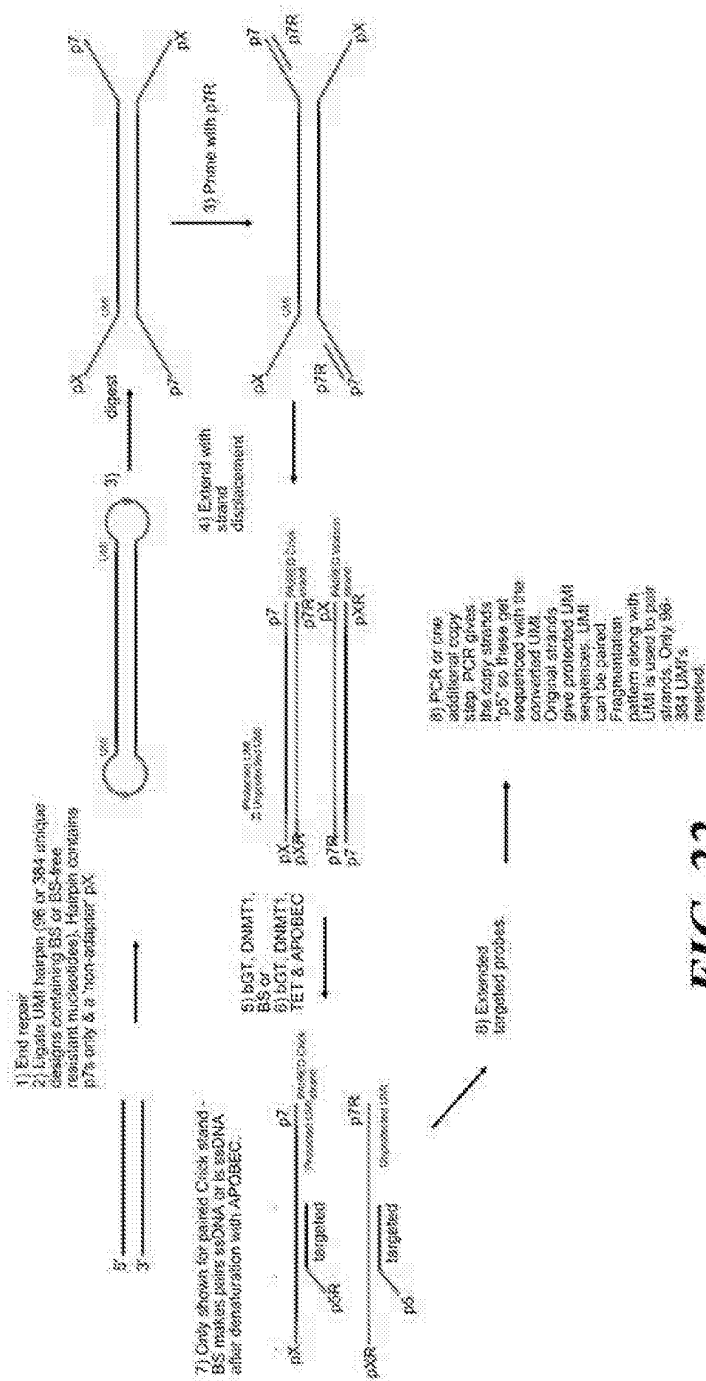


FIG. 22

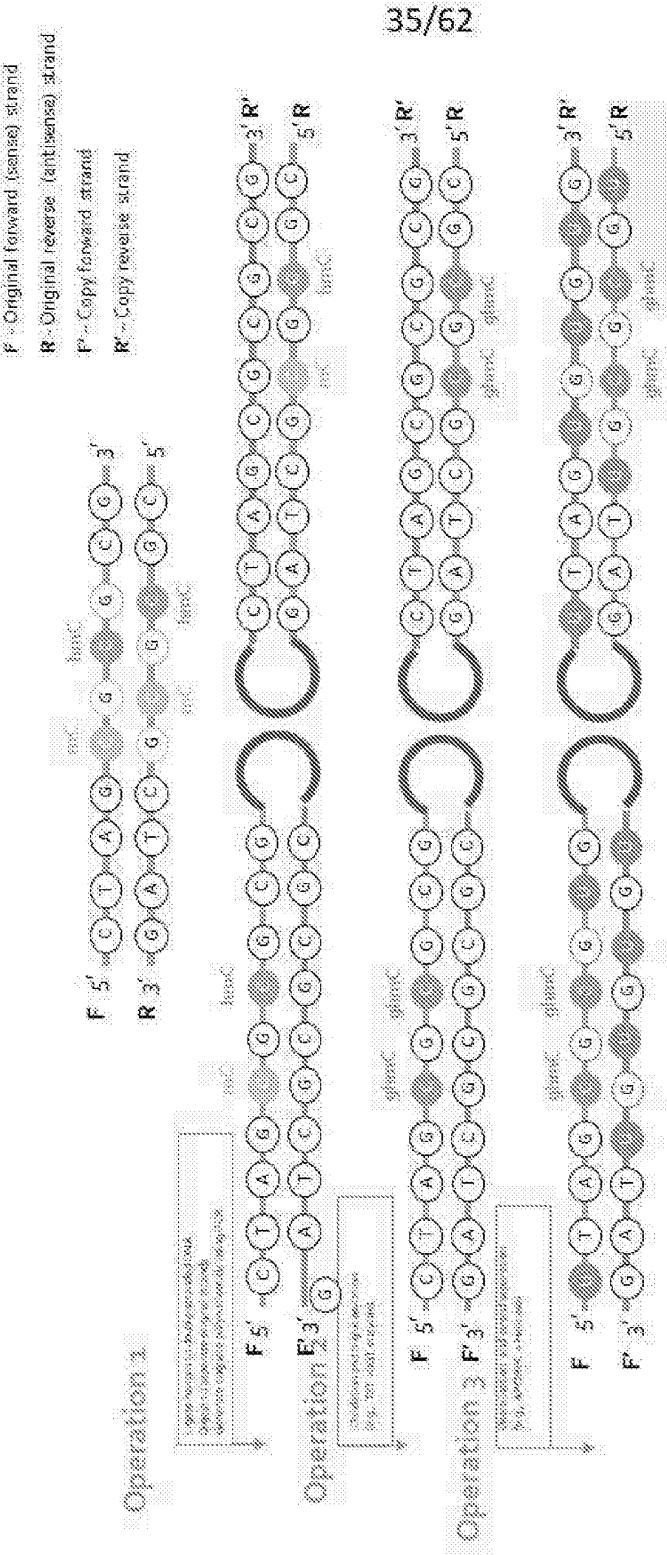


FIG. 23

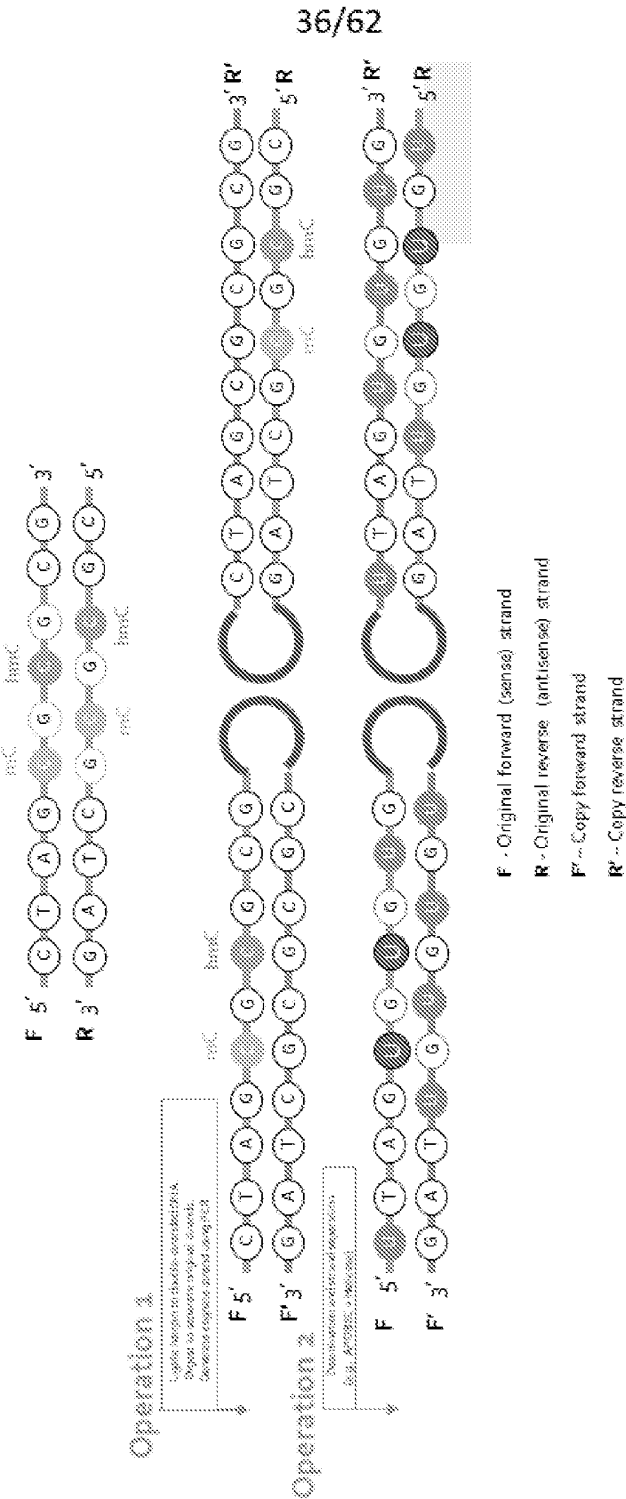


FIG. 24

37/62

True base	F strand	F' strand
A	A	T
C	T	G
G	G	T
T	T	A
5mCpG/ 5hmCpG	C	G
Error (F/F')	A/A, A/C, A/G, C/A, C/C, C/T, G/A, G/G, T/C, T/T, G/C, C/G in non-CpG context	

FIG. 25

38/62

True base	F strand	F' strand
A	A	T
C	U	G
G	G	T
T	T	A
Error (F/F')	A/A, A/C, A/G, C/A, C/C, C/T, C/G, G/A, G/G, G/C, T/C T/T, T/G	

FIG. 26

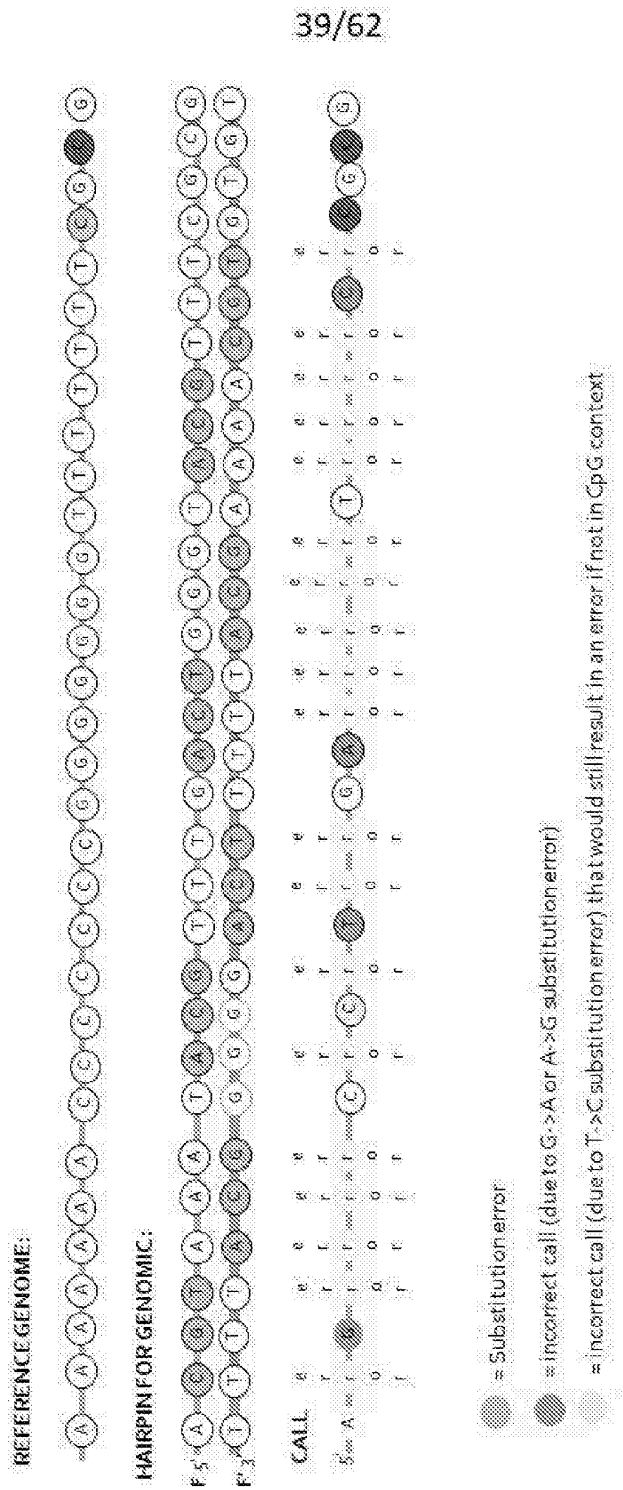


FIG. 27

40/62

	NA24385	NA24631
# Reads Aligned	2.19 billion	2.42 billion
# Bases Aligned	241 billion	266 billion
Strand Balance	0.50	0.50
Unfiltered Coverage	82.5x	91.2x
Duplication Rate	63.5%	67.6%
De-duplicated Coverage	30.1x	29.6x

FIG. 28

41/62

	NA24385	NA24631
Total SNPs	3.35 million	3.32 million
% in dbSNP (SNPs)	97.1%	97.2%
dbSNP Ti/Tv	2.074	2.069
Novel Ti/Tv	1.93	2.01
Total Indels	642K	639K
% in dbSNP (Indels)	93.9%	94.2%
Ins/Del (known, novel)	0.90, 0.93	0.90, 0.96

FIG. 29

42/62

	NA24385	NA24631
SNP het/hom sensitivity	91.9%, 92.0%	94.2%, 94.7%
SNP het/hom PPV	99.0%, 99.9%	98.9%, 99.9%
SNP genotype concordance	99.9%	99.9%
SNP non-ref GC	91.4%	93.8%
Indel het/hom sensitivity	84.9%, 84.3%	88.6%, 90.2%
Indel het/hom PPV	96.7%, 99.4%	97.4%, 99.3%
Indel genotype concordance	99.9%	99.9%
Indel non-ref GC	82.5%	87.5%

FIG. 30

43/62

	SS NA24585	NIIST NA24585	SS NA24631	NIIST NA24631
Total SNPs	3.35 million	3.5 million	3.32 million	3.57 million
% in dbSNP (SNPs)	97.1%	90.3%	97.2%	91.6%
dbSNP Ti/Tv	2.074	2.066	2.069	2.06
Novel Ti/Tv	1.93	0.709	2.01	0.76
Total Indels	642K	1.25 million	639K	1.12 million
% in dbSNP (indels)	93.9%	47.5%	94.2%	51.6%
Ins/Del (known, novel)	0.90, 0.93	0.91, 2.63	0.90, 0.96	0.91, 3.36

FIG. 31

44/62

	NA24385	NIST NA24385	NA24631	NIST NA24631
SNP het/hom sensitivity	91.9%, 92.0%	87.6%, 90.9%	94.2%, 94.7%	93.1%, 95.7%
SNP het/hom PPV	99.0%, 99.9%	87.5%, 98.7%	98.9%, 99.9%	89.6%, 99.1%
SNP genotype concordance	99.9%	99.9%	99.9%	99.9%
SNP non-ref GC	91.4%	82.2%	93.8%	88.5%
Indel het/hom sensitivity	84.9%, 84.3%	81.2%, 89.4%	88.6%, 90.2%	86.1%, 93.2%
Indel het/hom PPV	96.7%, 99.4%	30.7%, 97.5%	97.4%, 99.3%	32.1%, 97.3%
Indel genotype concordance	99.9%	99.9%	99.9%	99.9%
Indel non-ref GC	82.5%	38.3%	87.5%	41.1%

FIG. 32

45/62

	Sample	Chunks
Lane 01	NA24385	119
Lane 02	NA24385	121
Lane 03	NA24631	131
Lane 04	NA24631	134

FIG. 33A

	NA24385	NA24631	1% Mixin
# Reads Aligned	2.19 billion	2.42 billion	2.45 billion
# Bases Aligned	241 billion	266 billion	270 billion
Strand Balance	0.50	0.50	0.50
Unfiltered Coverage	82.5x	91.2x	92.3x
Duplication Rate	63.5%	67.6%	N/A
De-duplicated Coverage	30.1x	29.6x	N/A

FIG. 33B

46/62

	6x dups removed	6x no dup marking
# Reads Aligned	156 million	156 million
# Bases Aligned	17.3 billion	17.3 billion
Strand Balance	0.5005	0.5005
Unfiltered Coverage	5.9x	5.9x
Duplication Rate	14.5%	0%
De-duplicated Coverage	5.0x	5.9x

FIG. 34

47/62

	NA24385 original	NA24385 5x	NA24385 6x
Total SNPs	3.35 million	2.43 million	3.29 million
% in dbSNP (SNPs)	97.1%	96.3%	72.1%
dbSNP Ti/Tv	2.074	2.056	2.070
Novel Ti/Tv	1.93	3.06	26.6
Total Indels	642K	328K	340K
% in dbSNP (indels)	93.9%	95.7%	95.1%
Ins/Del (known, novel)	0.90, 0.93	0.90, 0.86	0.89, 0.85

FIG. 35

	NA24805 minicells
A → C	61
A → G	8085 (88%)
A → T	205
C → A	201
C → G	62
C → T	853 (88%)
G → A	859 (85%)
G → C	60
G → T	228
T → A	221
T → C	8841 (88%)
T → G	52

FIG. 36A

	NA24805	NA24801
A → C	128,030	211,389
A → G	26,523,863 (88%)	28,572,695
A → T	528,519	462,050
C → A	530,385	437,120
C → G	175,377	258,460
C → T	2,423,628 (88%)	2,306,728
G → A	2,425,785 (88%)	2,302,532
G → C	176,964	338,989
G → T	531,670	437,680
T → A	527,010	448,359
T → C	26,523,282 (88%)	28,598,802
T → G	126,656	332,539

FIG. 36B

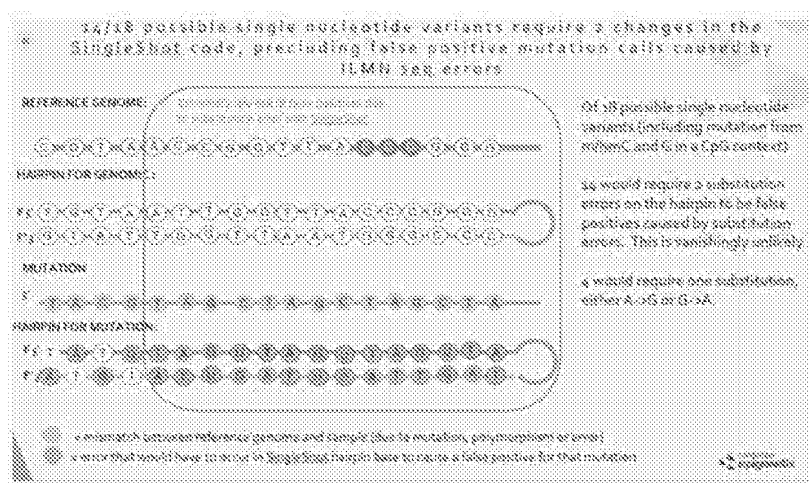
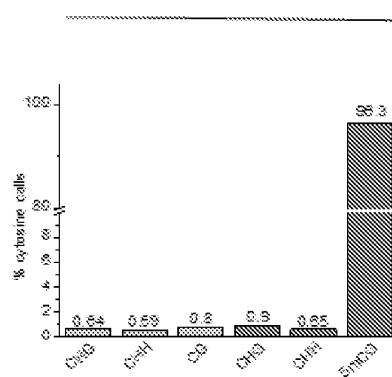
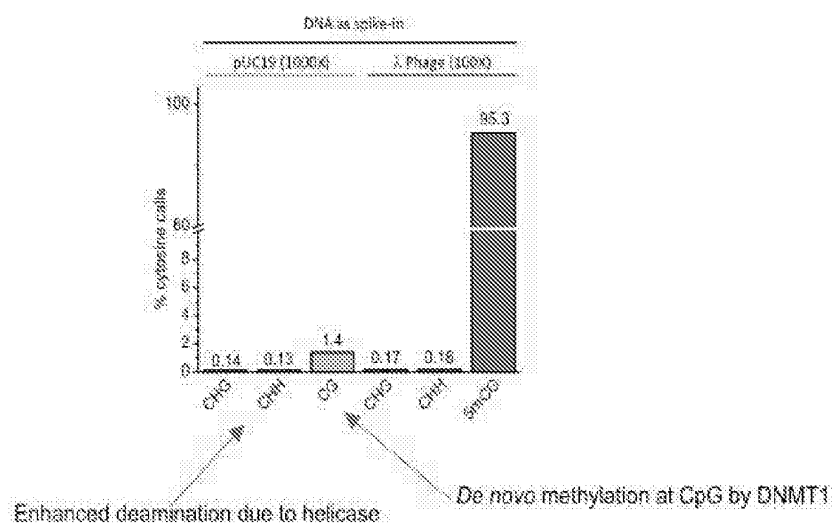


FIG. 37

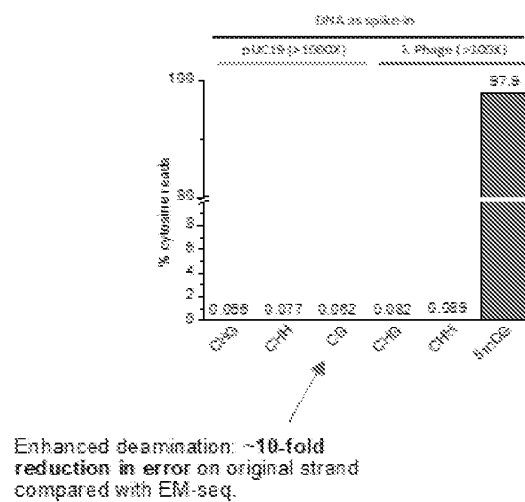
50/62

**FIG. 38**

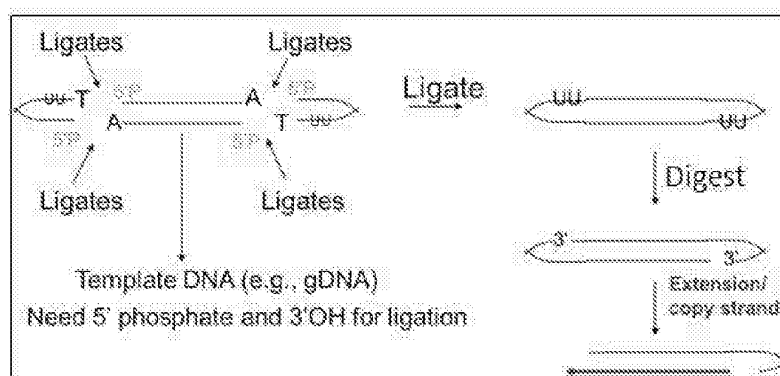
51/62

**FIG. 39**

52/62

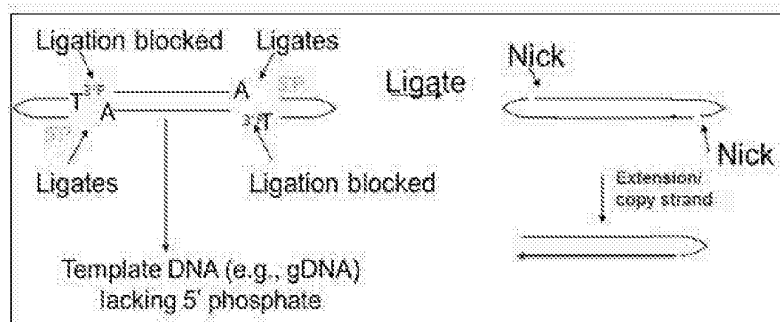
**FIG. 40**

53/62



Standard pre-deamination workflow involving the use of enzyme digestion to generate a nick on hairpin adapter (that contains uridines)

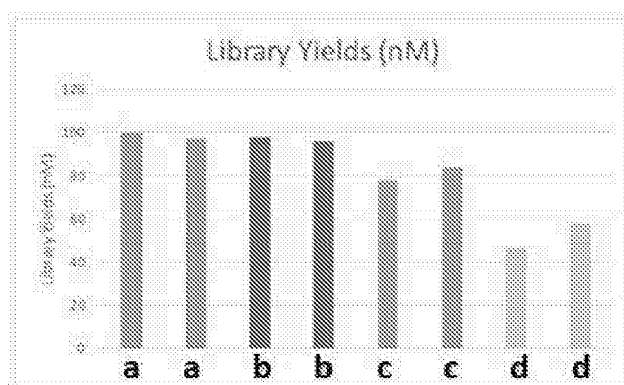
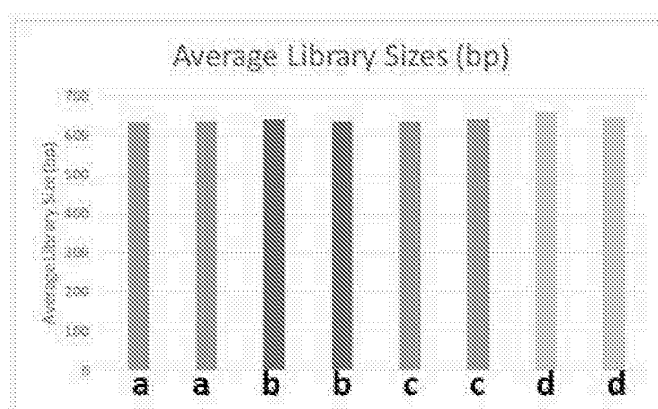
FIG. 41A



Pre-deamination workflow involving no enzyme digestion and a new hairpin lacking uridines (with 3' phosphate) and the use of template DNA lacking 5' phosphate

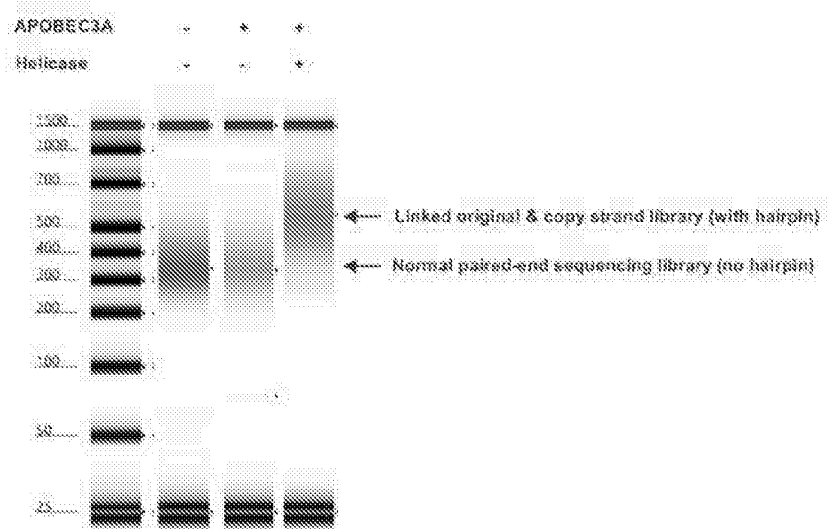
FIG. 41B

54/62

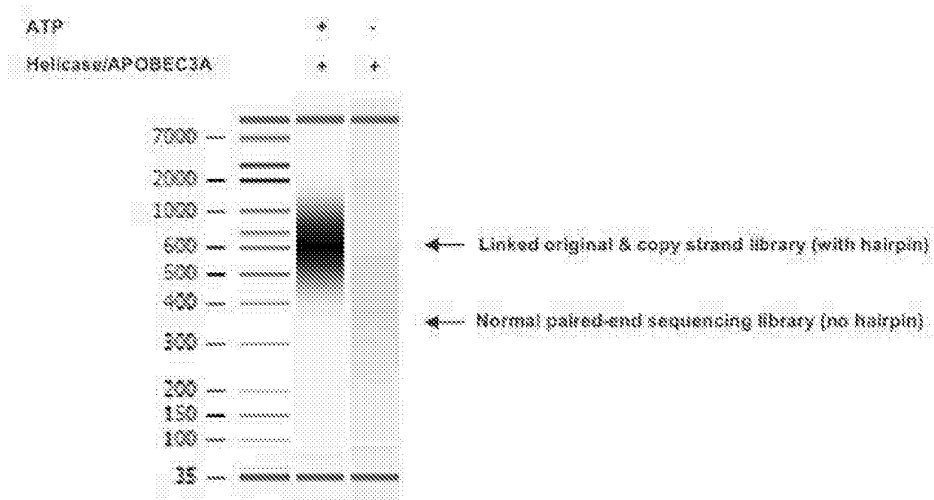
**FIG. 42A****FIG. 42B**

- a) Standard workflow with U hairpin + Digest
Ligation to 5' and 3' end of hairpin adapter. Can include digestion (e.g., with a mixture of Uracil DNA Glycosylase (UDG) and Endonuclease VIII) before extension step.
- b) 3'phosphate hairpin
Blocked ligation by 3'phosphate on hairpin.
- c) 3'phosphate hairpin + PNK exchange reaction
Blocked ligation by 3'phosphate on hairpin and by removing 5'phosphate on gDNA in presence of ADP and PNK.
- d) 3'phosphate hairpin + rSAP phosphatase
Blocked ligation by 3'phosphate on hairpin and by removing 5'phosphate on gDNA with rSAP phosphatase.

55/62

**FIG. 43**

56/62

**FIG. 44**

Step 1: Align sequencing read to genome (BAM file)

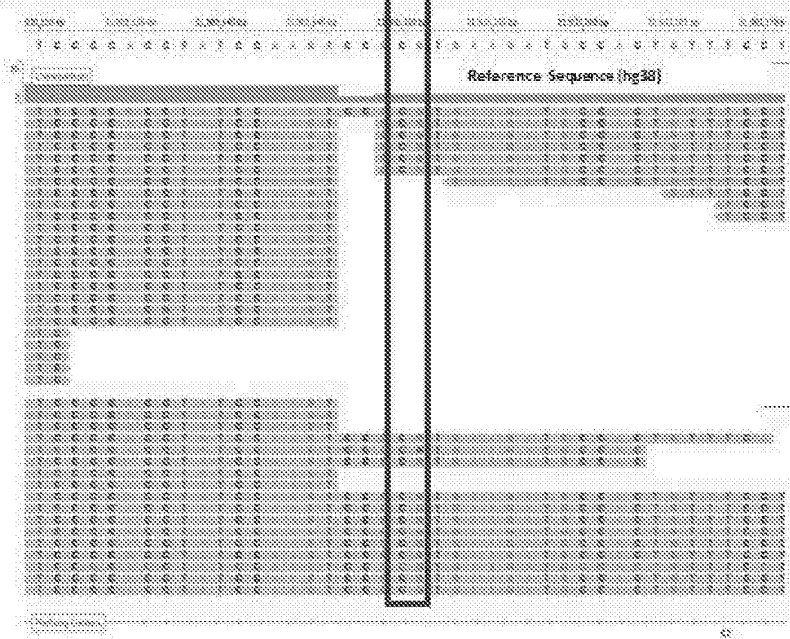


FIG. 45A

Step 2: Epigenetic Code

Sequence ID	Start Position	Epigenetic Code	Epigenetic State	Read ID
chr1	31903148	110	StateC	38456710
chr1	31903148	110	StateC	38456710
chr1	31903148	110	StateC	38456717
chr1	31903148	110	StateC	38456720
chr1	31903148	110	StateC	38456724
chr1	31903148	110	StateC	38456730
chr1	31903148	110	StateC	38456740
chr1	31903148	110	StateC	38456757
chr1	31903148	110	StateC	38456758
chr1	31903148	110	StateC	38456760
chr1	31903148	110	StateC	38456766
chr1	31903148	110	StateC	38456771
chr1	31903148	110	StateC	38456782

FIG. 45B

Step 3: Genomic Quantification

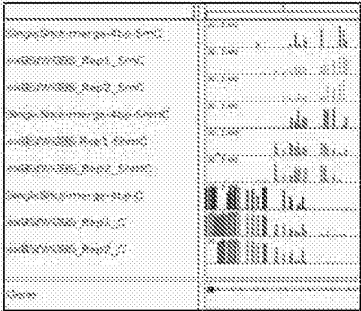


FIG. 45C

58/62

Step 1: BAM processing

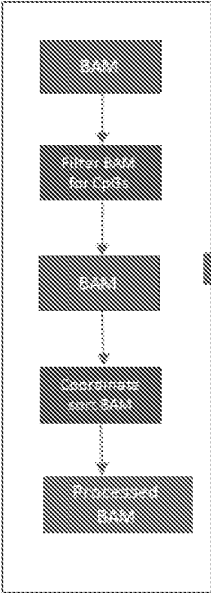
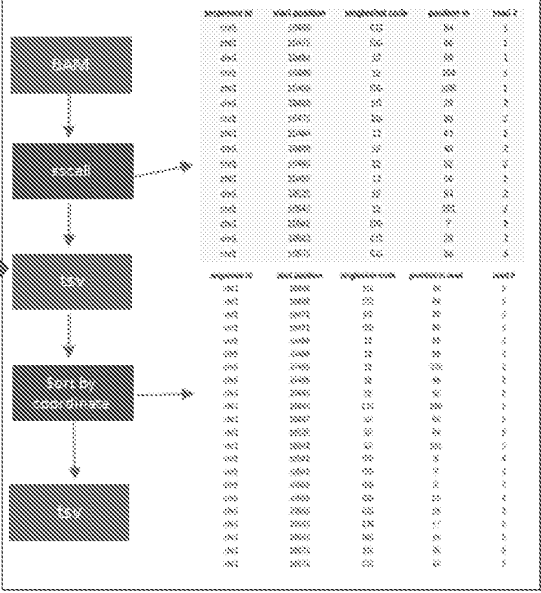
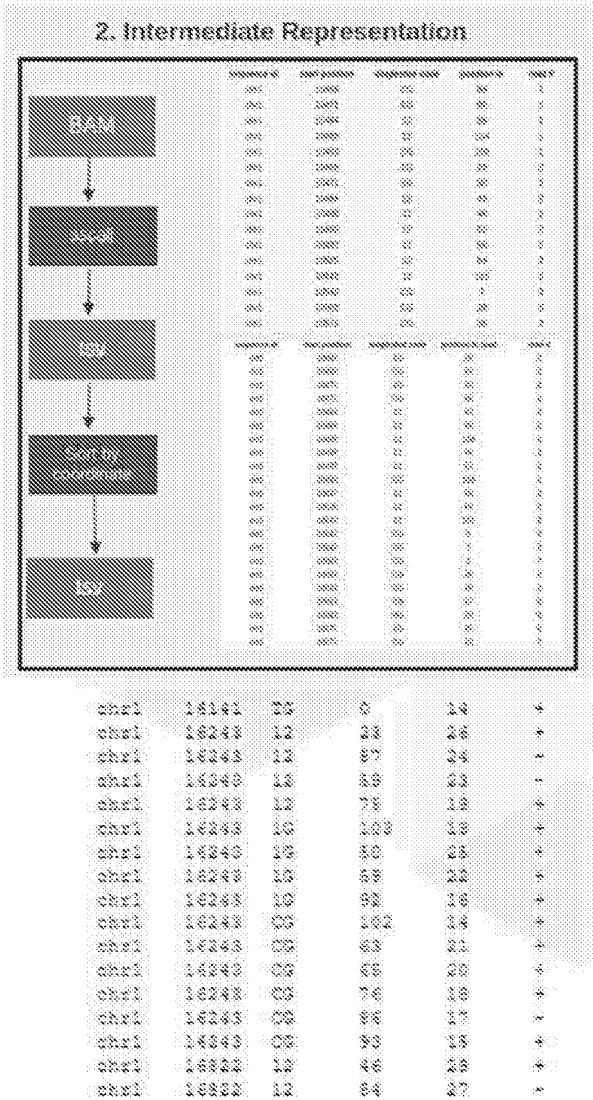


FIG. 46A

Step 2: Intermediate Representation

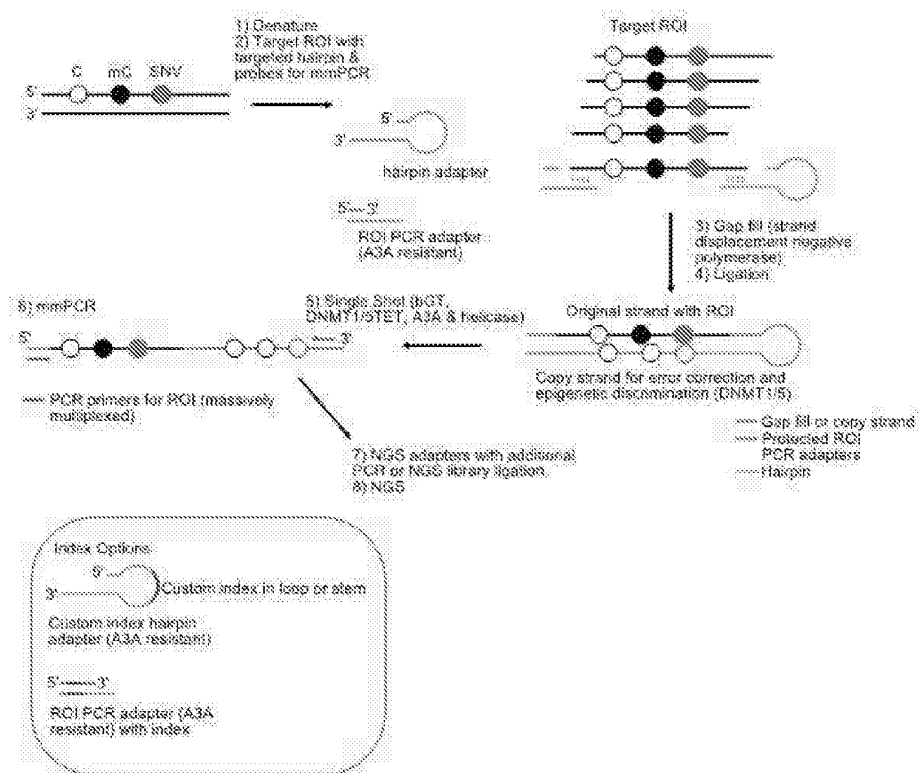




To identify stranded reads we add in a “stranded” column

FIG. 47

60/62

**FIG. 48**

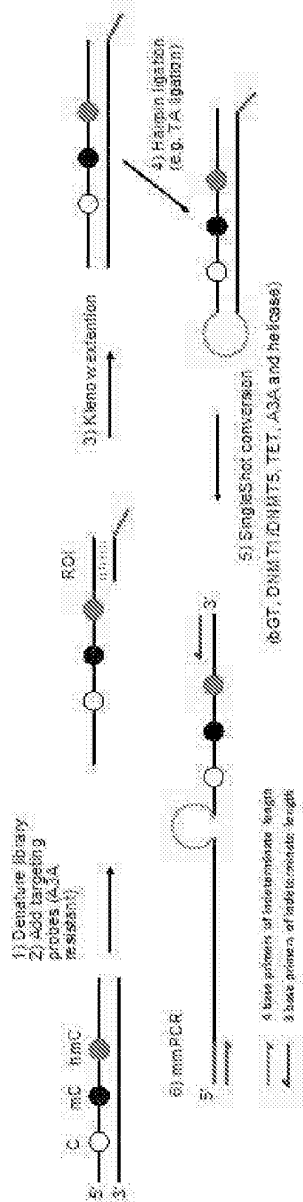


FIG. 49

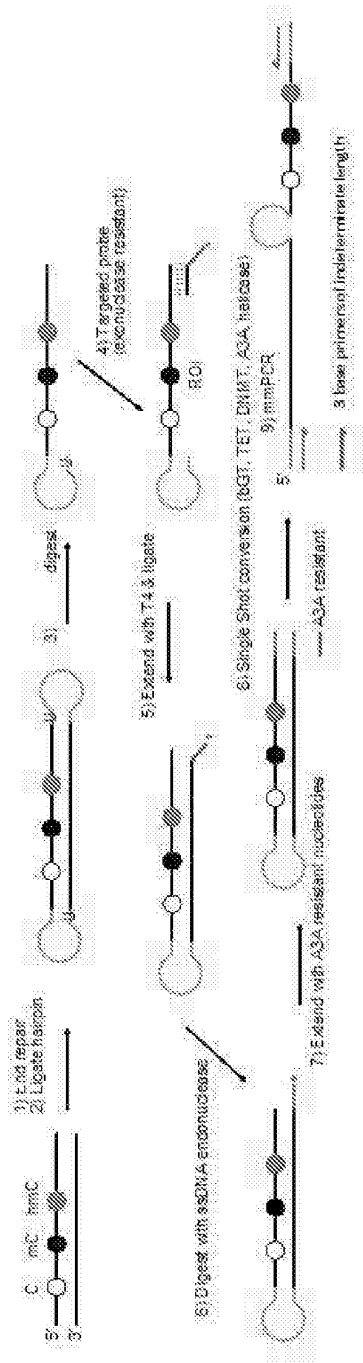


FIG. 50