

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

22 September 2022 (22.09.2022)



(10) International Publication Number

WO 2022/197852 A2

(51) International Patent Classification:

Not classified

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(21) International Application Number:

PCT/US2022/020624

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(22) International Filing Date:

16 March 2022 (16.03.2022)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/161,736

16 March 2021 (16.03.2021)

US

(71) Applicant: UNIVERSITY OF FLORIDA RESEARCH FOUNDATION, INCORPORATED [US/US]; 223 Grinter Hall, Gainesville, FL 32611 (US).

(72) Inventors: Kladde, Michael P.; 223 Grinter Hall, Gainesville, FL 32611 (US). ZOU, Mingqi; 223 Grinter Hall, Gainesville, FL 32611 (US). NABILSI, Nancy Hisham; 223 Grinter Hall, Gainesville, FL 32611 (US).

(74) Agent: VAN DYKE, Timothy H. et al.; Wolter Van Dyke Davis, PLLC, 1900 Summit Tower Boulevard, Suite 140, Orlando, FL 32810 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: METHODS AND KITS FOR TARGETED CLEAVAGE AND ENRICHMENT OF NUCLEIC ACIDS FOR HIGH-THROUGHPUT ANALYSES OF USER-DEFINED GENOMIC REGIONS

(57) Abstract: Disclosed herein are methods, oligonucleotides, and kits for targeted cleavage and enrichment of nucleic acids for high-throughput analyses of user-defined genomic regions. Targeted sequence enrichment is an increasingly sought technology. Currently available methods exhibit biases and substantial proportions of off-target reads. Here, disclosed is FENG C, a versatile, multiplexed method, wherein oligonucleotide adapters and flap endonuclease direct 5' DNA flap formation and cutting that releases target sequences with nucleotide-level precision. The target-specific oligonucleotides are designed by a novel program called FENG C oligonucleotide designer (FOLD). Further disclosed are the oligonucleotides and kits required to perform FENG C.



WO 2022/197852 A2

Methods and kits for targeted cleavage and enrichment of Nucleic Acids for high-throughput analyses of user-defined genomic regions

5

GOVERNMENT SUPPORT

This invention was made with government support under Grant Numbers R01 CA155390 awarded by The National Institutes of Health and HDTRA1-16-1-0048 awarded by the Defense Threat Reduction Agency. The government has certain rights in the invention.

10 FIELD OF THE INVENTION

The invention is related to methods for capturing and amplifying a plurality of specific sequences in a multiplexed reaction.

BACKGROUND

Targeted enrichment of DNA sequences for next-generation sequencing (NGS) is a
15 widely used procedure in basic investigative science and medicine. Its diverse applications include medical discovery and diagnostics, genotyping, biomarker assessment, marker-assisted plant breeding, epigenetic analysis of DNA methylation, etc. Multiplexed enrichment of target sequences from a complex mixture of sequences reduces sample complexity, which enables higher sequencing coverages, improves data quality, and reduces overall cost. Enrichment of
20 many genomic loci of interest in the same reaction in combination with sample barcoding yields further increases in NGS throughput, efficiency, and cost savings. Several distinct strategies have been developed for multiplexed sequence enrichment for NGS in genetic applications and epigenetic detection of DNA methylation, for example, 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) following bisulfite treatment. Enrichment strategies that are
25 compatible with both genetic and epigenetic inquiries provide additional versatility.

Multiplexed PCR amplification (Chamberlain et al., 1988) is commonly used to detect microsatellite and single-nucleotide polymorphisms or variants (Hayden et al., 2008). However, PCR does not scale well beyond analysis of 10-20 regions of interest in the same reaction due to non-specific amplification, adverse primer-primer interactions, and amplification biases that lead

to uneven sequencing coverage. In addition, for analysis of 5mC and 5hmC by bisulfite sequencing (Darst et al., 2010), conversion of the majority of C bases to U reduces sequence complexity (without reducing genome size) and therefore available priming specificity, further limiting the ability to perform multiplex PCR.

5 To avoid these problems, several approaches for sequence enrichment have been developed that do not rely on prior PCR amplification. One such method uses a restriction endonuclease to digest DNA followed by selection of fragments of a specific size range (Meissner et al., 2005; Gu et al., 2011). The enzyme MspI (CCGG) is typically used because it cuts irrespective of the methylation status of the central CG, the most commonly methylated
10 dinucleotide in vertebrate genomes (McClelland, 1981). Isolation of MspI fragments from 40-220 bp in length represents the majority of CpG islands in the human genome (Gu et al., 2011). In principle, the resulting library of fragments could be sequenced directly to detect genetic variants; however, the approach has instead been used to examine DNA methylation in a method called reduced representation bisulfite sequencing (RRBS) (Meissner et al., 2005). The total
15 fraction and number of represented regions (e.g., promoters, CpG islands, and their adjacent shores, etc.) is increased by isolating 70- to 320-bp fragments for enhanced RRBS (Akalin et al., 2012; Garrett-Bakelman et al., 2015) or by utilizing additional methylation-insensitive restriction endonucleases (Martinez-Arguelles et al., 2014). A drawback of RRBS, however, shared with
20 many NGS technologies, is the purification of different-size fragments leads to biased PCR amplification of smaller fragments during NGS library preparation. Moreover, the need for restriction enzyme digestion and size selection severely limits the flexibility of target selection, and precludes analysis of custom panels of loci or “walking” along contiguous segments of a genome.

DESCRIPTION OF THE DRAWINGS

25 The present embodiments are illustrated by way of example, and not by way of limitation, in the figures of the accompanying drawings.

The following figures are illustrative only, and are not intended to be limiting.

FIG. 1. Preferred flap-enabled next-generation capture (FENGEC) and enrichment protocol utilizing 5' and 1-nt 3' flaps. Purified source DNA, with or without fragmentation by

sonication, restriction enzyme digestion, or other suitable method, is used directly as input.

Step 1. A plurality of double flap structures with variable-length, single-stranded DNA (ssDNA)

5' flaps and a constant 1-nt 3' flap (pink arrowheads) is formed by addition of a set of flap adapters 1 and 2 that anneal to 5' and 3' ends of each region of interest, respectively (step 1a;

only one target strand depicted in light gray for simplicity). As shown, each flap adapter comprises an oligo complex consisting of a target-specific oligo 1-N or oligo 2-N base paired at their common 3' end with a universal oligo 1 in which a 3' A, C, G, or T remains unpaired (oligo U1-N), constituting a 1-nt 3' flap. Flap adapter pairs comprised of oligo U1-T paired with both an oligo 1-T and an oligo 2-T are preferred because they demonstrate superior cleavage and capture efficiency (see FIG. 3 and FIG. 11). The 5' flap endonuclease (FEN) activity of FEN1 or *Thermus aquaticus* DNA polymerase (hereafter, Taq) is used to remove the 5' flaps (step 1b).

Phosphodiester bond cleavage occurs one nucleotide 3' of both 5' flaps (red arrowheads), within duplex DNA formed by the flap adapters annealed to their respective, complementary target sites. This releases target sequences with ligatable termini, i.e., a 5' phosphate (downstream of a

1-nt gap) and a 3' hydroxyl. Step 2. The base constituting the 1-nt 3' flap at the 3' end of oligo U1-N fills each 1-nt gap and is ligated to the 5'-phosphate of each sequence of interest, as well as to the downstream sequence (not shown). The 3' end of each target sequence is ligated to a

corresponding adapter 3 complex comprised of a target-specific oligo 3-N (nested relative to oligo 2-N) with its constant 5' end base paired with oligo U2. The 3' hydroxyl of the oligo U2 is covalently linked to a three carbon spacer (indicated by ball and stick). Step 3. This covalent

modification protects oligo U2 and 3' ends to which it has ligated against 3' to 5' degradation by addition of both exonuclease I (Exo I) and Exo III. Non-protected source DNA, unligated oligo U1-N, as well as oligos 1-N, oligos 2-N, and oligos 3 are fully degraded. Step 4. The captured

and dramatically enriched single-stranded target sequences are purified for the first time. At the end of the protocol, the enriched sequences are amplified with oligo U1 (serves as primer as well) and U2 primer using standard PCR or methyl-PCR (PCR following deamination of C to U) to construct NGS libraries. In this and all subsequent figures, lines connecting different DNA strands indicate complementary base pairing; the number of lines is not intended to reflect any specific number of base pairs, and the various molecules are not drawn to scale.

FIG. 2. Alternative FENGCC protocol utilizing a plurality of 5' flaps with no 3' flap. Taq is the preferred enzyme in this protocol because it cuts 5' flaps without a 1-nt 3' flap with more

specificity than FEN1, ultimately yielding more amplification product (see FIG. 13). Step 1. Oligo U1 (instead of U1-N) used in the alternative protocol lacks the unpaired, 3'-terminal nucleotide and therefore cannot fill the 1-nt gap created by FEN cleavage of 5' flap-only structures. Step 2. Therefore, after cleavage, at first, only oligo U2 is ligated to the 3' ends of the plurality of target sequences annealed to adapters 3. Step 3. A first round of digestion with Exo I and Exo III dramatically enriches the plurality of oligo U2-protected sequences of interest (for details see FIG. 1 legend). Step 4. Following a first purification, a set of flap adapters 1 comprising oligo U1-N and oligos 1-N is added and ligated to the 5' end of each target. Step 5. A second treatment with Exo I and Exo III removes oligos 1-N and unligated oligo U1-N. Step 6. The enriched DNA sequences are subjected to a second purification and then amplification with oligo U1 and U2 primer using standard PCR or methyl-PCR to construct NGS libraries.

FIG. 3. Preferred double flap structure to direct efficient and precise 5' flap scission. Shown is an actual oligo 1-T that hybridizes to and forms a downstream duplex at the 5' end of its respective target sequence of interest (gray). Similarly, an oligo 2-T hybridizes to the 3' end of the target sequence of interest (not shown). The constant 3' tail of all oligos 1-T (and oligos 2-T) is bound to U1-T oligo, forming an upstream duplex and reconstituting a plurality of double flaps. Note that the unpaired 3' T of the U1-T oligo 'overlaps' a T in all target sequences of interest (residues highlighted in pink, corresponding to pink arrowheads in FIG. 1), allowing it to be displaced upon formation of the downstream duplex, creating unpaired 1-nt 3'-T flaps. To ensure that the 3' flaps are only 1 nt in length, in this example, a V (A, C, or G) is selected to reside at the base of the 5' flap as shown. A V base does not pair with the first nucleotide of the 3' tail of the oligos 1-T and oligos 2-T, i.e., an A (indicated by blue arrow).

FIG. 4. Flap adapter-directed cleavage of target 200-mer ssDNA oligos. (A) Schematic of substrate to assess cleavage of a 200-mer in structures containing a 129-nt 5' flap but no 3' flap. A 200-mer with A, T, G, or C (200-N oligos) at position 130 (hydrogen bonds indicated by widest vertical line) was contacted by a flap adapter comprised of a 200 oligo 1-N annealed to a constant oligo U1 (TABLE 1, Sheet 1). In the naming convention, for example, the 200-A oligo is a proxy sequence of interest with A at nt 130, which anneals to a complementary T in the 200 oligo 1-A. Taq or FEN1 cleaves the target sequence between nt 130 and nt 131 (red arrowhead), yielding a complex with a 1-nt gap (see FIG. 2, step 1b). The oligos are not drawn to scale.

(B) Representative Agilent 2100 Bioanalyzer traces for cleavage of a 5' flap with no 3' flap. Substrates were formed by mixing 500 nM each of 200-T oligo, 200 oligo 1-T, and oligo U1 with increasing amounts of Taq as indicated (U, unitage based on polymerization activity). The cycling parameters for cleavage were 3 min at 95°C and 20 min at 65°C, followed by 14 cycles of 30 sec at 95°C and 10 min at 65°C. Digital peaks corresponding to the migration positions of the uncut, input 200-T oligo (orange arrow, 69 s (sec) aligned migration time) and cut 200-T oligo (green arrow, 53 s aligned migration time) are indicated, as are size markers added post cleavage (black arrows, 43 s and 113 s aligned migration times). With 2 U and 5 U Taq, prominent peaks of cut 200-mer were produced at the expense of uncut, input 200-mer.

FIG. 5. Optimization of Taq digestion of 5' flaps, with and without a 1-nt 3' flap. Various 5' flap substrates were formed on 200-N oligos, cleaved with Taq, and the digestion products separated on an Agilent 2100 Bioanalyzer as in FIG. 4. The areas under the peaks of cut and uncut 200-N oligo were integrated and the percentages of digestion were calculated by (mass of cut oligo)/(mass of cut oligo + mass of uncut oligo) x 100. (A) Taq dose response (unitages based on polymerization activity) in Taq buffer for 15 reaction cycles as described in the legend of FIG. 4. Reactions contained 500 nM each of 200-T oligo, 200 oligo 1-T, and oligo U1 (TABLE 1, Sheet 1). (B) Percentages of cut 200-T oligo in reactions with 500 nM each of 200-T oligo, 200 oligo 1-T, and oligo U1 incubated with 1 U Taq for the indicated numbers of reaction cycles. The first digestion cycle was 3 min at 95°C and 20 min at 65°C, and subsequent cycles were 30 sec at 95°C and 10 min at 65°C. (C) Comparison of digestion efficiencies of various combinations of indicated oligos incubated with 1 U Taq for 10 reaction cycles. Each Taq reaction contained 500 nM each of 200-N oligo and either oligo U1 (A+U1, C+U1, G+U1, and T+U1) or the indicated oligo U1-N (A+U1-C, -G, or -T; C+U1-A, -G, or -T; G+U1-A, -C, or -T; and T+U1-A, -C, or -G) mixed with the corresponding 200 oligo 1-N (e.g., bar 2, 200-A oligo, 200 oligo 1-A, and U1-C oligo) (TABLE 1, Sheet 1). Note the similar efficiencies of cutting all 200-mers irrespective of the absence/presence or identity of the base constituting the 1-nt 3' flap under these conditions employing 10 cutting cycles. All experiments were done in duplicate (n = 2) and are reported as the arithmetic mean ± range.

FIG. 6. Dose response of FEN1 cleavage of a 129-nt 5' flap, without a 1-nt 3' flap. Five hundred nanomolar each of 200-T oligo, corresponding 200 oligo 1-T, and oligo U1 (TABLE 1,

Sheet 1) were incubated with increasing amounts of FEN1 in FEN1 buffer for 15 reaction cycles as described in the legend of FIG. 5. The products were analyzed and percentages of cut 200-T oligo were calculated as also described in the legend of FIG. 5. Overall, a higher percentage of 200-T oligo cleavage was achieved with FEN1 compared with Taq as tested in FIG. 5. Data are reported as mean $\pm$ range (n = 2).

FIG. 7. Digestion efficiencies of a 129-nt 5' flap, without a 1-nt 3' flap, with Taq and FEN1 in different reaction buffers. Five hundred nanomolar each of 200-T oligo, corresponding 200 oligo 1-T, and oligo U1 (TABLE 1, Sheet 1) were incubated with Taq or FEN1 in the indicated buffers for 10 reaction cycles as described in the legend of FIG. 5. The enzymes used were: 1 U APEX Taq (Genesee Scientific); 1 U HotStar Taq (Qiagen); and 32 U FEN1 (New England Biolabs). The units of Taq were based on polymerization activity. The buffers used were: 1x PCR buffer (Qiagen; referred to as Taq buffer); 1x ThermoPol reaction buffer (New England Biolabs; referred to as FEN1 buffer); and the mixed buffer reactions contained final concentrations of 1x CutSmart buffer (New England Biolabs) and 1x Taq buffer or FEN1 buffer. The products were analyzed and the percentages of cut 200-T oligo were calculated as described in the legend of FIG. 5. Data are reported as mean $\pm$ range (n = 2; **P* < 0.05, ***P* < 0.01, Student's *t* test). As observed in FIG. 6, higher efficiencies of 200-T oligo cutting were obtained with FEN1 compared with Taq.

FIG. 8. 5-Methylcytosine (5mC) does not inhibit 5' flap cleavage. (A) Dose response of FEN1 cleavage of four different control unmethylated and four different C-5-methylated double flaps. These eight structures were formed by mixing 500 nM each of: 1) 80-nt oligos that had either zero or five internal 5mC residues (at nt 43, 47, 49, 53, and 54, all located within 1-10 bp from each cleaved phosphodiester bond); 2) either 80 oligo 1-A, -C, -G, or -T; and 3) either U1-A, -C, -G, or -T oligo, respectively (TABLE 1, Sheet 1). The -N suffixes in the key indicate the 3'-terminal base of each U1-N oligo, which overlaps the respective base in its 80 oligo 1-N to create four different double flaps. Reactions were incubated with 0, 0.01, 0.04, 0.2, or 1 U of FEN1 in a 20 μ l reaction for 3 min at 95°C and 20 min at 65°C, followed by 14 cycles of 30 sec at 95°C and 10 min at 65°C. The percentages of cut 80-mer were assessed by quantitative PCR (qPCR) and calculated using the formula: % cut = 100 $\times$ [1-2^{-(Ct of FEN1 treated - Ct of FEN1 untreated)}]. Data are mean $\pm$ SD, n = 3. (B) Effective concentration achieving 50% cleavage (EC₅₀) by FEN1

obtained from the data plotted in (A). No significant differences were observed between all pairwise comparisons of cleavage of the unmethylated and methylated substrates as determined by serial one-way ANOVA tests. Therefore, multiple 5mC in the targeted strand both adjacent to the FEN1 cleavage site and within sequences contacted by FEN1 do not significantly inhibit double flap cleavage.

FIG. 9. Effect of 5' flap length on target strand cleavage and enrichment. (A) Schematic of plasmid pGEM-3Z/601b. The selected 571-nt target region (wide dark green line) and positions of HindIII, NdeI, and DrdI restriction sites as well as their map coordinates relative to the cut HindIII end are shown. (B) Target sequence enrichment after digestion of substrates with a 1-nt 3' flap and different 5' flap lengths. pGEM-3Z/601b (20 ng) was linearized with HindIII and divided equally into four separate reactions, two of which were digested further with either NdeI or DrdI. Plasmid digested with both HindIII and NdeI was contacted at the 5' end of the 572-nt target (extra nucleotide due to location of NdeI cut) with an NdeI adapter (pGEM-3Z/601b NdeI oligo 1-T and oligo U1). This serves as a no flap positive control (no added Taq) for target strand capture by ligation and amplification of the expected 618-bp product, including 46 bp added by ligation of the oligos U1 and U2 (lane 1). Plasmid digested with HindIII-DrdI and only HindIII tested FENGC with addition of Taq and the flap adapter 1 (pGEM-3Z/601b NdeI oligo 1-T and U1-T oligo) for removal of 5' flaps of 87 nt (lane 2) and 2,453 nt (lane 3), respectively. The fourth reaction with HindIII-linearized plasmid (lane 4) but omitting Taq provides a negative control; without scission of the 2,453-nt flap, the U1-T oligo should not ligate to the 5' end of the target DNA strand and PCR amplification should also not occur. To effect cleavage of 5' flaps, all four reactions were incubated for 3 min at 95°C and 20 min at 65°C, followed by 9 cycles of 30 sec at 95°C and 10 min at 65°C. Following cleavage, a HindIII adapter (pGEM-3Z/601b HindIII oligo and oligo U2) was added to all reactions in order to ligate the oligo U2 to the common HindIII-cut 3' end of the ssDNA 571-nt target sequence. Ligation was performed using Ampligase (Lucigen), employing conditions of 3 min at 95°C followed by 100 cycles of 0.5 min at 94°C and 8 min at 65°C. After subsequent bisulfite PCR (BS-PCR), the amplification products were directly electrophoresed, i.e., not purified beforehand, on a 1% agarose gel containing 0.2 µg/ml ethidium bromide and imaged with ultraviolet light on a transilluminator. The specific 618-bp product was obtained in reactions 2 and 3, indicating precise cleavage of the 87 nt and 2,453 nt 5' flaps, respectively, as well as ligation of the liberated 571-nt target sequence to the

oligos U1-T and U2, adding 47 bp post amplification. By contrast, no amplification product was observed in the negative control reaction in which Taq was omitted. Oligos are indicated in TABLE 1, Sheet 1.

FIG. 10. FENG C enrichment with Taq after cleavage adjacent to and nearby a 5mC residue within double flap structures. (A) pGEM-3Z/601b, treated with or without M.SssI and methylation cofactor *S*-adenosyl-*L*-methionine (SAM), was digested with or without the methylation-sensitive restriction endonuclease HhaI. A 1-kb molecular size marker was electrophoresed in the leftmost lane. Complete digestion of unmethylated pGEM-3Z/601b (lane 2) and complete protection against digestion of methylated pGEM-3Z/601b (lane 4), respectively, was observed. (B) Unmethylated (U) and methylated (M) pGEM-3Z/601b was linearized with HindIII and then processed by FENG C. Three different flap adapters comprised of oligos U1-T, -G, and -C that annealed to methyl test pGEM-3Z/601b oligo 1-T, -G, or -C, respectively, were used to contact and form double flap structures with the pGEM-3Z/601b target sequence. The respective oligo U1-N was ligated to the 5' end of each target sequence after cutting by the FEN activity of Taq at the indicated phosphodiester bonds (red arrowheads; 5mC in red type). The common HindIII-cut 3' end of each target sequence was ligated to the pGEM-3Z/601b HindIII adapter, i.e., oligo U2 bound to the pGEM-3Z/601b HindIII oligo. After ligation, unprotected fragments were degraded by incubation with Exo I and Exo III. BS-PCR was performed and the amplified products were directly visualized by 1% agarose gel electrophoresis. A 100-bp molecular size marker was electrophoresed in the leftmost lane. Given that the expected amplification product was obtained in all reactions as observed in FIG. 8, it can be concluded that 5mC does not inhibit 5' flap cleavage. In addition, the presence of 5mC at the 5' end of the cut target sequence, did not block ligation to the oligo U1-N nor PCR amplification. Oligos are indicated in TABLE 1, Sheet 1.

FIG. 11. Identity of the 3'-terminal nucleotide of oligo U1-N affects the performance of preferred FENG C with double flaps. (A) Site-specific mutagenesis of pGEM-3Z/601b. The first A in the NdeI site in pGEM-3Z/601b at nt 2,453 was mutated to C, G, and T. The set of four resulting plasmids, pGEM-3Z/601b-N, introduces each of the four nucleotides immediately 5' of the site of FEN cleavage and oligo U1-N ligation, changing the base constituting the 1-nt 3' flap. After incubation with HindIII, 0.2 ng of each of the four linearized plasmids was denatured and

contacted with the pGEM-3Z/601b HindIII adapter and each of the four respective flap adapters 1 (oligo U1-N annealed to its corresponding pGEM-3Z/601b NdeI-2 oligo 1-N; TABLE 1, Sheet 1). The flap adapters directed cleavage by Taq or FEN1 between nt 2,453 and nt 2,454, using the reaction cycles described in the legend of FIG. 9. After severing the 2,453-nt 5' flap, the 572-nt target sequence released from each of the four linearized plasmids was ligated to its respective oligo U1-N and oligo U2 at the 5' and 3' ends, respectively. Reactions with no Taq and no FEN1 constituted negative controls. The ligation products were treated with Exo I and Exo III, purified with AMPure XP beads (Beckman Coulter), and subjected to BS-PCR. (B) BS-PCR products from FENG C with no enzyme (lanes 1-4), 2 U Taq (based on polymerization activity; lanes 5-8), and 32 U FEN1 (lanes 9-12) were directly analyzed by 1% agarose gel electrophoresis. Note the highest yield of the expected 619-bp product (572-bp target sequence plus 47 bp universal priming sequences) was obtained when oligo 1-T and pGEM-3Z/601b NdeI-2 oligo 1-T were incubated with pGEM-3Z/601b-T (lanes 6 and 10). Also, the use of oligo 1-G not only reduced the yield of specific product but led to an amplified smear of non-specific, high-molecular-weight products (compare lane 7 with 6 and lane 11 with 10).

FIG. 12. Identity of the 3'-terminal nucleotide of the oligo U1-N also affects the performance of alternative FENG C without a 1-nt 3' flap. (A) Schematic of 5' flap formation. The four pGEM-3Z/601b-N plasmids described in FIG. 11A were linearized with HindIII, and 0.2 ng of each was contacted by one of the four respective flap adapters (oligo U1 and corresponding pGEM-3Z/601b NdeI-2 oligo 1-N; TABLE 1, Sheet 1) in the presence of 0 U or 2 U Taq. (B) FENG C enrichment yields. The 3' end of each cut target sequence was first ligated to the oligo U2 of the HindIII adapter. After degradation of unprotected DNA fragments by Exo I and Exo III, the 5' end of the target sequence was ligated to the respective adapter (oligo U1-N and corresponding oligo 1-N). Reactions without Taq (or FEN1) are used as negative controls (lanes 1-4) and with Taq (lanes 5-6). The ligation products were treated a second time with Exo I and Exo III, amplified by BS-PCR, and directly analyzed by 1% agarose gel electrophoresis. Similar results as FIG. 11 were seen, with oligo 1-T providing the highest, specific amplification yield (lane 6). Again, use of U1-G produced a non-specific smear of high-molecular-weight products.

FIG. 13. Sensitivity of plasmid sequence enrichment using the alternative FENG C procedure without a 3' flap. Two micrograms genomic DNA (gDNA) purified from human colon cancer cell line HCT116 was spiked with the indicated masses of HindIII-linearized pGEM-3Z/601b-T and subjected to the alternative FENG C protocol as described in FIG. 2 and FIG.

12B. Sensitivity of FENG C using (A) FEN1 (32 U) in its supplied buffer and (B) FEN1 *versus* Taq (2 U) in Taq buffer. The amplified products were directly analyzed by 1% agarose gel electrophoresis. Note that in alternative FENG C procedure, Taq (lanes 1-3) yields superior sensitivity compared with FEN1 (lanes 4-6) because FEN1 produces multiple cleavage products.

FIG. 14. Specificity of plasmid sequence enrichment using the alternative FENG C procedure. pGEM-3Z/601b-T (0.2 ng) was subjected to alternative FENG C as described in FIG. 2 and FIG. 12B, using each of the four indicated flap adapters 1 in the presence of 32 U FEN1. The amplified products were directly analyzed by 1% agarose gel electrophoresis. The specific, 619-bp PCR product was only obtained when FEN activity and cognate flap adapter 1 (U1-T oligo and pGEM-3Z/601b NdeI-2 oligo 1-T were included (lane 6).

FIG. 15. Comparison of preferred and alternative FENG C enrichment of target sequences of interest from gDNA using the FEN activity of Taq. Two micrograms of human HCT116 gDNA were subjected to the indicated FENG C procedure using 2 U Taq with and without a 3' flap as described in FIG. 1 and FIG. 2, respectively. Cleavage was performed for 3 min at 95°C and 20 min at 65°C, followed by 14 cycles of 30 sec at 95°C and 10 min at 65°C. Matched flap adapters 1-T and 2-T for 11 single-copy genes were employed to capture and enrich sequences of mean length of ~300 nt (TABLE 1 and TABLE 2, Sheets Human ~300 nt Targets) using BS-PCR. Reactions without Taq were used as negative controls. The amplification products were directly analyzed by 1% agarose gel electrophoresis. Both the preferred and alternative FENG C procedures with Taq yielded the specific amplification product of ~350 bp (282-315 bp target sequences plus 47 bp of ligated oligos U1-T and U2).

FIG. 16. Comparison of Taq and FEN1 in preferred FENG C enrichment of gDNA sequences. Two micrograms of human HCT116 gDNA were processed by FENG C with a 1-nt 3' flap as described in FIG. 1. Sequences of mean length of ~300 nt were captured from the same 11 genes in FIG. 15 using matched flap adapters 1-T and 2-T. Reactions contained (A) no enzyme, 2 U Taq in 1x Taq buffer, or 32 U FEN1 in 1x FEN1 buffer and (B) No or 32 U FEN1

in FEN1 or Taq buffer as indicated. Reactions with no enzyme constituted negative controls (lanes 1 and 3). The ~350 bp amplification products of BS-PCR (282-315 bp target sequences plus 47 bp of universal primers) were directly analyzed by 1% agarose gel electrophoresis. Both panels show that FEN1 yields superior sequence enrichment compared with Taq (compare lanes 2 and 4).

FIG. 17. Effect of bisulfite conversion and amplicon length on the preferred FENG C protocol. FENG C was performed as described in FIG. 1 on 2 μ g or 4 μ g HCT116 gDNA as indicated using 32 U FEN1. In one set of reactions, the same flap adapters 1-T and 2-T were employed to enrich the same 11 target amplicons of ~350 bp (282-315 bp of target sequence plus 47 bp of universal primers; lanes 1-3; TABLE 1 and TABLE 2, Sheets Human ~300 nt Targets) as in FIG. 16. In a second set of reactions, the same flap adapters 1-T were used and new flap adapters 2-T were utilized that anneal farther downstream to capture and amplify sequences of ~500 bp (430-452 bp of target sequence plus 47 bp of universal primers; lanes 4-7; TABLE 1 and TABLE 2, Sheets Human ~450 nt Targets). The captured sequences were amplified without (lanes 1, 2, 4, and 5) and with bisulfite conversion (lanes 3, 6, and 7), i.e., subjected to standard PCR and BS-PCR, respectively. Reactions without FEN1 were included as negative controls (lanes 1 and 4). The amplification products were directly analyzed by 1% agarose gel electrophoresis. As expected, reduced PCR yields were observed with longer amplicons (compare lanes 2 with 5 and 3 with 6) and due to the known degradation of DNA by bisulfite (compare lanes 2 with 3 and 5 with 6). An ~500 bp BS-PCR product was not detected with 2 μ g gDNA input but a sufficient yield was obtained with 4 μ g gDNA input (compare lane 6 and 7).

FIG. 18. MAPit-FENG C of 10 promoter sequences of human DNA mismatch repair genes plus 1 human control gene amplified by standard PCR and BS-PCR. Two independent biological replicates (lanes 1-6 and lanes 7-12) of cultured human glioblastoma (GBM) L0 cells were assayed by Methyltransferase Accessibility Protocol for individual templates (MAPit) methylation footprinting followed by FENG C and BS-PCR (MAPit-FENG C). Cells were permeabilized, treated with M.CviPI to mark accessible GpC (hereafter, GC) sites in nuclear chromatin, and purified gDNA (2 μ g or 4 μ g as indicated) was processed by FENG C. Following target sequence scission with Taq or FEN1 and ligation capture using ~300-nt and ~450-nt oligos 1-T, 2-T, and 3-T (TABLE 1 and TABLE 2, Sheets Human ~300-nt Targets and Human ~450-nt

Targets, respectively), the samples were subjected to either standard PCR or BS-PCR. The amplicons, directly analyzed by 1% agarose gel electrophoresis, are: ~350-bp products (282-315 bp captured sequences plus ligated 47 bp universal priming sequences; lanes 1-3 and 7-9) obtained with Taq and standard PCR (lanes 1 and 7), FEN1 and standard PCR (lanes 2 and 8), and FEN1 and BS-PCR (lanes 3 and 9); ~500-bp products (430-452 bp capture regions plus ligated 47 bp universal primers; lanes 4-6 and 10-12) obtained with Taq and standard PCR (lanes 4 and 10); FEN1 and standard PCR (lanes 5 and 11); and FEN1 and BS-PCR (lanes 6 and 12). All reactions yielded PCR products of the correct size, which were purified for NGS.

FIG. 19. High correlation of HCG and GCH methylation levels between biological replicates of MAPit-FENGc. Product libraries generated by MAPit-FENGc in FIG. 18 (lanes 3, 6, 9, and 12) were purified by AMPure XP beads and subjected to single-molecule real-time (SMRT) sequencing on a Pacific Biosciences (PacBio) Sequel instrument. High-fidelity circular consensus sequencing (CCS) reads (minimum of five sequencing passes) were aligned to the reference sequences for each of the 11 enriched promoters. Plotted are the methylation level of each HCG and GCH site from the 6 promoters that had 10 or more aligned CCS reads in the combined biological replicates for both the ~300-nt and ~450-nt target sequences (TABLE 3).

FIG. 20. High correlation of HCG and GCH methylation levels between two biological replicates of MAPit-FENGc and single-gene MAPit of the *MLH1* promoter. For the MAPit-FENGc libraries constructed in FIG. 18 (lanes 6 and 12) and sequenced as described in FIG. 19, the methylation level of each HCG and GCH site was plotted for the 438-bp *MLH1* promoter sequence (upper two panels; TABLE 2, Sheet Human ~450-nt Targets). H5mCG and G5mCH levels were also plotted for the same 438 bp of the *MLH1* promoter that overlapped a single 732-bp BS-PCR product (lower two panels) that was amplified from deaminated GBM L0 gDNA using two primers specific for *MLH1* promoter sequence (TABLE 1, Sheet 1).

FIG. 21. Strong correlation of H5mCG and G5mCH levels in MAPit-FENGc for ~300-nt and ~450-nt target sequences processed by BS-PCR. High-fidelity PacBio Sequel CCS reads with five or more sequencing passes from the MAPit-FENGc product libraries constructed in FIG. 18, lanes 3, 6, 9, and 12 were aligned to the 11 promoter reference sequences (TABLE 2, Sheets Human ~300-nt Targets and Human ~450-nt Targets). The plotted values are the methylation level of each HCG and GCH site from 6 promoters with 20 or more aligned CCS

reads in the combined biological replicates (TABLE 3) and in the overlapping region between the ~300-nt and ~450-nt target sequences.

FIG. 22. Validation of MAPit-FENGc by single amplicon BS-PCR. (A) Strong correlation between the fraction of methylation of each HCG and GCH site located within 438 bp of overlapping *MLH1* sequence. PacBio CCS sequences of ~450-nt from the 11 promoter panel (TABLE 2, Sheet Human ~450-nt Targets) were enriched by MAPit-FENGc from GBM L0 gDNA followed by BS-PCR as presented in FIG. 18 and analyzed in FIG. 19 and FIG. 20. The BS-PCR CCS reads were obtained from a single 732-bp product amplified from the same deaminated GBM L0 DNA with primers MGO3791 and MGO3792, specific for *MLH1* (TABLE 1, Sheet 1). (B) Nearly identical methylation levels at HCG and GCH sites within *MLH1* regions that overlap between CCS reads from the 732-bp BS-PCR amplicon and the 300 bp and 438 bp MAPit-FENGc amplicons (TABLE 2, Sheets Human ~300-nt and Human ~450-nt Targets).

FIG. 23. FENGc sequence capture followed by deamination using enzymatic methyl conversion and amplification (EM-PCR). Genomic DNA (500 ng) from GBM L0 probed by M.CviPI was used for FENGc of 119 target sequences of ~450 nt from promoters of human genes encoding products with functions in DNA repair and cancer (TABLE 1 and TABLE 2, Sheets Human ~450 nt Targets). The captured sequences were purified by MinElute Kit (Qiagen), AMPure XP beads (Beckman Coulter), or NEBNext beads (NEB) as indicated, then subjected to EM-PCR (Sun et al., 2021). The amplification products were directly analyzed by 1% agarose gel electrophoresis. Based on the abundant ~500-bp product (~450-nt target sequences plus 47 bp universal primers), the latter two purification methods yielded favorable results.

FIG. 24. Optimization of primer concentrations for preferred FENGc followed by EM-PCR. Genomic DNA (1 µg) from HCT116 cells (not probed with M.CviPI) was input to FENGc of 119 target sequences of ~450 nt from promoters of human genes encoding products with functions in DNA repair and cancer (TABLE 1 and TABLE 2, Sheets Human ~450 nt Targets). Oligo U1-T (1 µl of indicated concentration) and a separate stock mixture of flap oligos 1-T and flap oligos 2-T (2 µl of indicated concentration) were first added to each reaction, with and without FEN1 as indicated. Next, without purification, separate stock mixtures of oligo U2 (1 µl of indicated concentration) and oligos 3-T (2 µl of indicated concentration) were added for

ligation followed by EM-PCR. The amplification products were directly analyzed by 1% agarose gel electrophoresis. (A) Titration of primer concentrations demonstrated that including 10 μ M each of oligos U1-T and U2 provides the optimal FEN1-dependent amplification of the EM-PCR product (compare lane 4 with lanes 2 and 6). (B) Specific EM-PCR product yield increased with more added flap oligos 1-T, flap oligos 2-T, and oligos 3-T. Therefore, their summed concentration should be calculated to be close to but less than the concentration of each of oligos U1-T and U2 (TABLE 12).

FIG. 25. Comparison of preferred FENG C followed by EM-PCR with DNA fragmented by SpeI digestion *versus* sonication. The indicated amounts of gDNA isolated from GBM L0 cells probed with M.CviPI were input to preferred FENG C directly after fragmentation by (A) SpeI digestion or sonication *versus* (B) sonication. The 5' flaps of the 119 sequences of interest were cleaved by 32 U FEN1 within double flap structures with 3' flaps comprised of one T. The EM-PCR amplification products were directly analyzed by 1% agarose gel electrophoresis. Note that sonication is superior to SpeI digestion in (A), and specific product was detected with as little as 50 ng input gDNA in (B).

FIG. 26. Uniform product lengths within purified libraries constructed by MAPit-FENG C enrichment. MAPit-FENG C with EM-PCR of 119 promoter sequences from DNA repair and cancer-associated genes of ~450 nt in length (TABLE 1 and TABLE 2, Sheets Human ~450 nt Targets) was performed on 500 ng of gDNA of each of the indicated cell materials. The preferred FENG C protocol (FIG. 1) was used to construct these and all libraries in subsequent figures. Two independent cultures of each cell type (NSC, human neural stem cells) were assayed as indicated by suffixes -1 and -2. Five hundred nanograms total gDNA of a 0.1%:99.9% mixture of L0:NSC gDNA were also subjected to MAPit-FENG C with EM-PCR for the 119 targets of ~450 nt. Amplified, purified products (demarcated with arrowhead) are ~500 bp (430-452 bp captured sequences plus 47 bp universal primers). The product libraries were purified for EM-seq (Sun et al., 2021) with 0.65x AMPure XP beads and analyzed with an Agilent TapeStation D5000 system.

FIG. 27. PacBio CCS read length distribution for sequenced MAPit-FENG C libraries. The purified EM-seq libraries prepared by MAPit-FENG C in FIG. 26 for the two biological replicates of NSC and GBM L0 were sequenced on a PacBio Sequel instrument. Percentages of

different-length CCS reads are plotted as a histogram with 20-nt bins. The highest frequency of obtained reads was of the expected 500 nt, including both universal oligo sequences. Small peaks of ~1,000 nt in the L0 histograms likely correspond to amplicon dimers that arose during ligation of sample-specific PacBio barcodes. Suffixes -1 and -2 denote each of the two independent

5 biological replicates.

FIG. 28. PacBio CCS read length distribution for additional sequenced MAPit-FENG C libraries. The purified EM-seq libraries prepared by MAPit-FENG C in FIG. 26 for the two biological replicates of the 0.1% L0:99.9% NSC mixtures and GBM Nx18-25 were sequenced with the PacBio Sequel platform. Percentages of different-length CCS reads plotted as a
10 histogram with 20-nt bins demonstrates the highest frequency of obtained reads was of the expected ~500 nt, including both universal oligo sequences. Small peaks of ~1,000 nt likely correspond to amplicon dimers that arose during ligation of sample-specific PacBio barcodes. Suffixes -1 and -2 denote each of the two independent biological replicates.

FIG. 29. MAPit-FENG C using EM-PCR results in low percentages of off-target reads. Percentages of CCS reads unmapped to the human genome, mapped to human genome but off
15 target for the 119 regions, and on target (TABLE 4) are shown for the eight different sequenced MAPit-FENG C libraries as described in the legends of FIG. 26, FIG. 27, and FIG. 28. Suffixes -1 and -2 denote each of the two independent biological replicates.

FIG. 30. GC content has a moderate, negative correlation with the number of obtained
20 MAPit-FENG C CCS reads. For the NSC and GBM L0 EM-seq libraries described in FIG. 26 and FIG. 27, high-fidelity CCS reads were filtered further for $\geq 95\%$ conversion of HCH to HTH, i.e., not HCG or GCH, that also covered $\geq 95\%$ of each reference sequence length in both biological replicates for each of the 119 MAPit-FENG C targets (TABLE 5). Plotted is the LOESS fit line of the natural logarithm transformation of the read number + 1 *versus* the GC
25 content for each of the 119 targets.

FIG. 31. GC content has a moderate, negative correlation with the number of obtained MAPit-FENG C CCS reads. PacBio Sequel CCS reads from the EM-seq libraries generated by MAPit-FENG C of gDNA from the duplicate Nx18-25 samples and 0.1% L0:99.9% NSC gDNA mixtures. CCS read filtering and the format of plotted data are described in the FIG. 30 legend.

FIG. 32. High reproducibility of MAPit-FENGc utilizing EM-PCR. Plotted are the correlations between the fraction of methylation of each HCG and GCH site in the two independent biological duplicates of the eight MAPit-FENGc libraries containing the enriched 119 promoter target sequences for the indicated samples described in FIG. 26, FIG. 27, and FIG. 28. Only targets with 10 or more aligned CCS reads with $\geq 95\%$ conversion of HCH and $\geq 95\%$ coverage of its reference sequence length in both replicates were included (TABLE 5). This equated to 64 targets for NSC, 52 targets for L0, 64 targets for Nx18-25, and 63 targets for the 0.1% L0:99.9% NSC gDNA mixtures.

FIG. 33. Strong correlation between MAPit-FENGc using EM and bisulfite conversion. Plotted is the fraction of methylation of each HCG and GCH site within 6 targets of ~450 nt having 20 or more reads aligned in both EM-converted (TABLE 5) and bisulfite-converted samples (TABLE 3) in the combined biological replicates for the sequenced GBM L0 libraries.

FIG. 34. The *POLD4* promoter, an example of a partially methylated promoter with a prominent, accessible nucleosome-free region (NFR) as detected by MAPit-FENGc with EM-seq. The GBM L0 libraries and sequenced CCS reads are as described in FIG. 26 (lanes 5 and 6) and FIG. 27, respectively. Shown is a methylscaper plot of 1,000 molecules randomly chosen from 2,139 obtained high-fidelity PacBio Sequel CCS reads with ≥ 5 SMRT sequencing passes, additionally filtered for $\geq 95\%$ conversion, and $\geq 95\%$ coverage of each reference sequence length (TABLE 5). To aid pattern recognition, the molecules were subjected to unsupervised hierarchical clustering and plotted by methylscaper {PMID: 34125875}. Each row of pixels depicts the epigenetic features of one molecule or epiallele from one cell recorded in one 447-nt CCS read from the *POLD4* gene promoter, -297 to +150 relative to the transcription start site (TSS; bent arrow indicating direction of transcription). Plotted CCS reads yield a methylation pattern for HCG and GCH sites (vertical hashes at top and vertical lines crossing each panel with the distances separating them drawn to scale) as plotted in the left and right panels, respectively. Overlapping GC and CG sites, i.e., GCG are omitted to avoid ambiguity of methylation by exogenously added M.CviPI (GC specificity) and endogenous DNA methyltransferases (CG specificity), respectively. As indicated in the key at bottom, two or more consecutively methylated HCG are connected by red (left panel). Two or more consecutively accessible GCH, i.e., not bound by protein and hence accessible to and methylated by M.CviPI,

are connected by yellow (right panel). Two or more consecutively unmethylated sites, either HCG or GCH, are connected by black. Gray designates border transitions between methylated and unmethylated HCG sites as well as between accessible and inaccessible GCH sites.

Unaligned sequence varying from the hg38 genome reference is plotted in white. The molecules

in both panels are displayed in the same top-to-bottom order, thereby linking the patterns of DNA methylation (left panel) and chromatin accessibility (right panel) in each molecule.

Nucleosomes impair access of M.CviPI to 147 bp DNA, the length of DNA tightly wrapped around the histone protein core. Therefore, note in the right panel that most promoter copies exhibit two nucleosome-length protections against methylation by M.CviPI (full and partial blue ellipses labeled -1 and +1, respectively, drawn to scale), flanking a prominently accessible NFR. A fraction of NFR-bearing promoter copies harbors a short span of endogenous 5mCG (left panel), which does not correlate with the presence or absence of an NFR (right panel).

FIG. 35. The *ALKBH2* promoter, a second example of MAPit-FENGc utilizing EM-seq. The GBM L0 library, sequenced CCS reads, CCS read filtering are as described in FIG. 26, FIG. 27, and FIG. 34, respectively. The majority of the 979 aligned PacBio Sequel 436-nt filtered CCS reads (TABLE 5) from -282 to +154 from L0 cells revealed 5mCG at the limit of detection, except for a low level at the TSS (left panel). In addition, a highly accessible NFR is bordered upstream by a large footprint, consistent with a variably positioned -1 nucleosome (halved blue ellipse; right panel). The NFR is prominently occupied by a relatively short, uniformly located footprint, consistent with occupancy by a sequence-specific DNA-binding factor (right panel; black rectangle at top). The CCS read filtering, content of the panels, symbols, and key at bottom are as described in the legend of FIG. 34.

FIG. 36. MAPit-FENGc with EM-PCR detects differentially methylated and differentially accessible chromatin in NSC *versus* GBM L0 cells. Both cell lines were treated separately with M.CviPI plus SAM methylation cofactor. Purified gDNA (500 ng) was used as input to construct the MAPit-FENGc libraries in FIG. 26 (NSC, lanes 1 and 2; GBM L0, lanes 5 and 6), using the panel of 119 promoter targets (TABLE 1 and TABLE 2, Sheets Human ~450-nt Targets). The shown methylscaper plot has 3,460 and 3,889 450-nt filtered CCS reads from the *EPM2AIP1* promoter (-143 to +307) from (A) NSC and (B) L0 cells, respectively (TABLE 5). The CCS read filtering, content of the panels, symbols, and key at bottom are as described in the

legend of FIG. 34. The white rectangle represents the first portion of *EPM2AIP1* protein coding sequence. In NSC (A), the *EPM2AIP1* promoter has no to low 5mCG (left panel) and harbors mostly accessible, open chromatin (right panel) occupied by a sequence-specific DNA-binding factor (black rectangle at top) and a variably positioned +1 nucleosome (blue partial and full ellipses). By contrast, in GBM L0 (B), the promoter is hypermethylated, closed, and exhibits only limited accessibility within relatively short nucleosomal linker sequences, visible in a subpopulation of cells.

FIG. 37. MAPit-FENGCC efficiently detects a 0.1% subpopulation of hypermethylated epialleles within a heterogeneous sample. Permeabilized GBM L0 cells and NSC were treated separately with M.CviPI and SAM, and purified gDNA was mixed in a ratio of 0.1% L0:99.9% NSC. Next, 500 ng of the gDNA mixture was processed by the preferred FENGCC protocol with EM-PCR to construct the EM-seq libraries in FIG. 26 (lanes 7 and 8), yielding the high-fidelity PacBio CCS read distribution shown in FIG. 28. Shown are aligned, filtered CCS reads (1,781 total; TABLE 5) plotted by methylscaper. The CCS read filtering, content of the panels and key at bottom are as described in the legend of FIG. 34. The white rectangle represents the first portion of *EPM2AIP1* protein coding sequence. Ten reads in the gDNA mixture showed dense H5mCG, consistent with originating from L0 cells (red arrowhead). A proportions test in R concluded that the observed proportion of molecules (10 hypermethylated:1,771 unmethylated epialleles) is indeed at least or greater than 0.1% (****, $P < 0.0001$).

FIG. 38. *MSH5* exemplifies a gene with a remarkably accessible promoter. MAPit-FENGCC was conducted on two independent replicate cultures of human non-cancerous NSC and second GBM Nx18-25 that were treated with M.CviPI. A 119-target panel of human promoters (TABLE 1 and TABLE 2, Sheets Human ~450-nt Targets) was enriched by preferred FENGCC to construct the EM-seq libraries in FIG. 26 (lanes 3 and 4), yielding the high-fidelity PacBio CCS read distribution shown in FIG. 28. CCS read filtering as well as the content of the panels and key at bottom are as described in the FIG. 34 legend. Methylscaper plots of 447-nt filtered CCS reads aligned to the *MSH5* promoter (−302 to +145) from (A) two independent NSC cultures (top panels; 132 and 241 reads) and (B) two independent GBM Nx18-25 cultures (bottom panels; 170 and 192 reads) (TABLE 5). Suffixes -1 and -2 denote each of the two independent biological replicates. Each HCG and GCH position (in base pairs) along the *MSH5* amplicon is indicated at

bottom of its respective panel. The black bar indicates 147 bp, the size of a nucleosome core lacking the linker; this and other features and the distances between them are drawn to scale. A heterogeneous-sized footprint is marked by a black rectangle. Almost all *MSH5* promoter molecules across all 4 samples (733 of 735) had ≥ 10 methylated GCH sites, demonstrating highly accessible chromatin. This high degree of openness rules out incomplete cell permeabilization and chromatin probing with M.CviPI as trivial reasons for differential accessibilities observed between other loci in the same or different samples.

FIG. 39. Identification of chromatin architectures differentially methylated (H5mCG), accessible (G5mCH), or both using MAPit-FENGc with EM-seq. Generalized estimating equations were used to model the effect of cell line (human NSC *versus* GBM Nx18-25) on the per molecule proportions of endogenous H5mCG (number H5mCG/(number H5mCG + number HCG); left panels) and G5mCH (number G5mCH/(number G5mCH + number GCH); right panels). Only targets with ≥ 50 total CCS reads in the combined replicates (≥ 22 in any single replicate) were considered (TABLE 5), with filtering as described in the FIG. 34 legend. Errors were modeled as normally distributed, and the correlation structure was assumed exchangeable within each replicate. The `geeglm` function from the `geepack` v1.3-2 R package was used in R version 4.1.0. *P* values were corrected for multiple testing using the Bonferroni method with an alpha of 0.05 to control the false discovery rate (TABLE 6). Using criteria of $P < 0.05$ and ≥ 0.05 differential in the proportion of either H5mCG or G5mCH between NSC and GBM Nx18-25, 57% (31/54) of the evaluated promoters showed epigenetic alterations (TABLE 6, Sheets 1 and 2). With more stringent cutoffs, the percentages of promoters with ≥ 0.1 , ≥ 0.2 , and ≥ 0.4 differentials in H5mCG or G5mCH were 26% (14/54), 13% (7/54), and 7.4% (4/54), respectively (TABLE 6, Sheets 3-5). These promoters include human (A) *CD44*, (B) *CCN4*, and (C) *HIST1HB1* for which the data were rendered as violin plots for the two independent replicates (Rep 1 and 2) of NSC and GBM Nx18-25. Plotted is the proportion of H5mCG or G5mCH for each molecule (black dots), the median (horizontal line), interquartile range of methylation levels (box), and the smoothed probability density at different methylation levels (gray area).

FIG. 40. *CD44* promoter exemplifies detection by targeted MAPit-FENGc with EM-seq of differential epigenetic alterations associated with gene silencing in cancer. (A) Methylscaper

plots of 441-nt filtered CCS reads aligned to the *CD44* promoter (−270 to +171) from NSC (left two panels; 141 reads) and GBM Nx18-25 (right two panels; 96 reads) (TABLE 5). CCS read filtering as well as the content of the panels and key at bottom are as described in the FIG. 34 legend. (B) Methylation level of each HCG (upper panel) and GCH (lower panel) motif tabulated from the filtered CCS reads plotted in (A). (C) Cumulative distribution function of NFR length, considering all aligned CCS reads from NSC and Nx18-25. The mean NFR length for NSC is 118 bp and for Nx18-25 is 72.0 bp, $P < 0.0001$. (D) Relative expression level of *CD44* in NSC and Nx18-25. Reverse transcription (RT)-qPCR was done using Taqman assay (Life Technologies, 44-449-63) with *CD44*-specific probe (ThermoFisher Scientific, Hs01075861_m1) for three independent biological replicates (mean $\pm$ SD; $**P < 0.01$). Taken together, compared with NSC where the *CD44* promoter is relatively unmethylated and open with longer NFRs, the promoter becomes hypermethylated and closed in GBM Nx18-25, consistent with strong silencing of expression.

FIG. 41. *CCN4* promoter exemplifies detection by targeted MAPit-FENGc with EM-seq of differential epigenetic alterations associated with upregulated expression in cancer. (A) Relative expression level of *CCN4* transcript in NSC and GBM Nx18-25. RT-qPCR was done using Taqman assay (Life Technologies, 44-449-63) with *CCN4*-specific probe (ThermoFisher Scientific, Hs00180245_m1) for three biological replicates (mean $\pm$ SD; $*P < 0.05$). (B) Methyldscaper plots for filtered 447-nt CCS reads of *CCN4* (−279 to +168 bp) from NSC (left two panels; 1,116 molecules) and Nx18-25 (right two panels; 1,143 molecules) (TABLE 5). The CCS read filtering, content of the panels, and key at bottom are as described in the legend of FIG. 34. The white rectangle indicates the first portion of *CCN4* protein coding exon 1. The bracketed subset of reads contains footprint evidence of a sequence-specific factor bound to a NFR (black rectangle). The difference in H5mCG between NSC and GBM Nx18-25 was statistically significant (TABLE 6); however, the difference in accessibility was not, probably because of an uncharacteristically high level of accessibility in molecules displaying H5mCG. The straight arrow marks a known GA to A variant (dbSNP, rs548251181) present in approximately half of the alleles from NSC and Nx18-25 (TABLE 7, Sheet Indels), which does not align to the hg38 reference genome and is plotted in white. This variant was verified by MAPit-FENGc genotyping using standard PCR amplification, i.e., without EM-seq conversion (TABLE 7). (C) Methylation level of each HCG (upper) and GCH (lower) motif from −279 to

+168 bp relative to the *CCN4* TSS tabulated from the filtered CCS reads plotted in (C). (D) Cumulative distribution function of NFR length, considering all aligned filtered CCS reads from NSC and Nx18-25. The mean NFR length for NSC is 89.4 bp and for Nx18-25 is 132 bp, $P < 0.0001$. Taken together, relative to NSC, the *CCN4* promoter in GBM Nx18-25 is less

5 methylated and contains more open and longer NFRs, consistent with the strong transcriptional induction in the tumor-derived cells.

FIG. 42. *HIST1H1B* promoter exemplifies detection by MAPit-FENGc with EM-seq of differential epigenetic alterations in a minority subpopulation of cells, despite no significant difference in bulk transcript abundance. (A) Methyloscaper plots of filtered 445-nt CCS reads of

10 the *HIST1H1B* promoter (−246 to +199) from NSC (left two panels; 926 reads) and GBM Nx18-25 (right two panels; 953 reads) (TABLE 5). CCS read filtering as well as the content of the panels and key at bottom are as described in the FIG. 34 legend. The white rectangle partially covered by nucleosome +1 indicates the first portion of *HIST1H1B* protein coding sequence. The black rectangle marks likely binding of a sequence-specific transcription factor, whereas the gray

15 rectangle corresponds to a relatively large footprint, but too small to be a nucleosome, that may correspond to paused RNA polymerase II. Note the range of promoter configurations (clusters marked 1-7), with less molecules populating summed clusters 1 and 2 in GBM Nx18-25 compared with NSC. In addition, the percentage of hypermethylated, inaccessible epialleles (cluster 6) was elevated in GBM *versus* NSC (4% *versus* 0.6%). (B) Methylation level of each

20 HCG (upper panel) and GCH (lower panel) motif tabulated from the CCS reads plotted in (A). The bracket indicates the area around the TSS (−80 to +45) that shows a significant decrease in accessibility in Nx18-25 (TABLE 6, $P < 0.0001$). (C) Relative expression of *HIST1H1B* in NSC and Nx18-25. RT-qPCR was done using Taqman assay (Life Technologies, 44-449-63) with *HIST1H1B*-specific probe (ThermoFisher Scientific, Hs01075861_m1) for three biological

25 replicates (mean $\pm$ SD). The expression levels are not statistically different. Taken together, the data exemplify the power of single-molecule MAPit-FENGc to detect epigenetic alterations in minority populations of cells. The hypermethylated and closed *HIST1H1B* promoter copies in GBM are highly unlikely to support active transcription but are too few in number to decrease a bulk, population-averaged measurement such as RT-qPCR.

FIG. 43. Length distribution of EM-seq high-fidelity CCS reads for ~940-nt MAPit-FENGc libraries. Two independent biological replicates of GBM Nx18-25 cells were treated with M.CviPI. Subsequently, 800 ng or 400 ng of purified gDNA from each biological replicate was processed by preferred FENGc using oligos 1-T, 2-T, and 3-T for 45 ~940-nt targets (TABLE 1 and TABLE 2, Sheets Human ~940 nt Targets). The four capture product libraries were sequenced using the PacBio Sequel platform. Shown is the length distribution of CCS reads plotted in 20-nt intervals. Suffixes -1 and -2 denote each of the two independent biological replicates.

FIG. 44. Low percentages off-target reads for MAPit-FENGc of ~940-nt sequences of interest with EM-PCR. Shown are the percentages of high-fidelity CCS reads (≥ 5 SMRT sequencing passes) unmapped to the human genome, mapped to human genome but off target for the 45 regions, and on target (TABLE 8) for the 4 different GBM Nx18-25 MAPit-FENGc libraries described in the FIG. 43 legend. Suffixes -1 and -2 denote each of the two independent biological replicates.

FIG. 45. High correlation of HCG and GCH methylation levels between two independent biological replicates of MAPit-FENGc for the ~940-nt target sequences. EM-seq high-fidelity CCS reads (≥ 5 SMRT sequencing passes) were filtered further for $\geq 95\%$ conversion, and $\geq 95\%$ coverage of the reference sequence length (TABLE 9). Plotted values are the fraction of methylation of each HCG and GCH site in 13 targets for 800 ng input and 12 targets for 400 ng input of Nx18-25 gDNA that have 14 or more CCS reads aligned in each biological replicate of each cell type (TABLE 9).

FIG. 46. High correlation of HCG and GCH methylation levels with different input amounts of gDNA for MAPit-FENGc of the ~940-nt targets. Plotted values are the methylation level of each HCG and GCH site in 13 targets from GBM Nx18-25 cells using CCS reads filtered for $\geq 15\times$ coverage in the combined biological replicates of both input gDNA amounts (TABLE 9) but otherwise as described in the FIG. 45 legend.

FIG. 47. High correlation of HCG and GCH methylation levels between two independent biological replicates of MAPit-FENGc for the ~940-nt and ~450-nt target sequences. Plotted are the methylation level of each HCG and GCH site in 13 targets from GBM Nx18-25 cells having $\geq 15\times$ coverage in the combined biological replicates of both the ~450-nt and ~940-nt MAPit-

FENGCC samples (TABLE 5 and TABLE 9). Other CCS read filtering parameters were as described in the FIG. 45 legend.

FIG. 48. Long-read MAPit-FENGCC using EM-seq phases multiple epigenetic features and reveals novel regulatory insights. The indicated lengths of promoter sequences from the locus containing the divergently transcribed *EPM2AIP1* and *MLH1* genes were captured and enriched by preferred FENGCC from gDNA isolated from two biological replicates of M.CviPI-treated GBM Nx18-25 cells (FIG. 26, lanes 3 and 4). Sequences in (A) and (B) were among those enriched by FENGCC for the 119 ~450-nt targets (TABLE 2, Sheet 2 and FIG. 28), whereas those in (C) were enriched for 45 ~940-nt targets from 800 ng gDNA (TABLE 2, Sheet 3 and FIG. 43 top). Shown and drawn to scale are methylscaper plots of molecules from the enriched regions: (A) 450 bp of *EPM2AIP1*, (B) 438 bp of *MLH1*, and (C) 937 bp of *EPM2AIP1-MLH1*, with all promoter coordinates indicated relative to *MLH1* TSSa. White rectangles depict protein coding sequence. The black bar indicates 147 bp, the size of a nucleosome core lacking the linker; this and other features and the distances between them are drawn to scale. Patterns of methylation at CHG are plotted by methylscaper in (A) and the left panel of (C); patterns of methylation at GCH are plotted in (B) and the right panel of (C). Each HCG and GCH position (in bp) on each amplicon is indicated at bottom of each panel. CCS read filtering as well as the content of the panels and key at bottom are as described in the FIG. 34 legend. The short molecules in (A) and (B) are aligned to the long molecules in (C). Hierarchical clustering of the 937-bp molecules in (C) was weighted on the *EPM2AIP1* half of the amplicon, whereas no sub-region was specified for clustering the two shorter amplicons in (A) and (B). Collectively, the three amplicons reveal three robust footprints (black rectangles numbered 1-3), corresponding to DNA-bound, sequence-specific transcription factors. The 937-bp molecules in (C), however, revealed a continuous NFR encompassing the *EPM2AIP1* TSS, *MLH1* TSSa, footprint 1 and footprint 3, the discovery of which would have otherwise required a third short amplicon. Moreover, strong co-occupancy of all three transcription factors and the two +1 nucleosomes was evident on the 937-bp molecules. In addition, based on the clustering of molecules from both short amplicons in (A) and (B), it is readily discerned that the +1 nucleosomes (blue ellipses) downstream of the *EPM2AIP1* TSS and *MLH1* TSSa occupy a continuum of positions, the *EPM2AIP1* +1 nucleosome more so. However, *MLH1* TSSa +1 nucleosome positions were randomized when hierarchical clustering of the 937-bp molecules in (C) was weighted on the

EPM2AIP1 +1 nucleosome and *vice versa* (not shown). This strongly suggests that the positioning of the two +1 nucleosomes is regulated independently, a regulatory insight that cannot be derived from short reads/molecules.

FIG. 49. Additional examples of phasing multiple epigenetic features provided by long-read MAPit-FENGc using EM-seq. Promoter sequences of (A) 938 bp of divergently transcribed *NPAT* and *ATM*, (B) 930 bp of *MSH2*, and (C) 944 bp of *CCN4* were captured and enriched by preferred FENGc from gDNA isolated from M.CviPI-treated GBM Nx18-25 cells (FIG. 26, lanes 3 and 4). All sequences were among those enriched for the 45 ~940-nt targets (TABLE 2, Sheet 3 and FIG. 43). Shown are methylscaper plots of methylation at HCG (left panels) and GCH (accessibility; right panels) of molecules from the enriched regions. The base pair position of each HCG and GCH in each amplicon is indicated at bottom of each respective panel. Promoter coordinates are indicated relative to the *ATM* TSS in (A) and relative to each single TSS in (B) and (C). White rectangles depict protein coding sequence. CCS read filtering as well as the content of the panels and key at bottom are as described in the FIG. 34 and FIG. 48 legends. The straight arrow in (C) marks a known GA to A variant (dbSNP, rs548251181; TABLE 7, Sheet Indels) plotted in white as described in the FIG. 41 legend. The pink rectangle denotes a 66-bp A-rich sequence (88% A) that exhibits variable-length non-alignment to the hg28 reference as well. Note that the *NPAT-ATM* (A) and *MSH2* (B) promoters show NFRs with robust transcription factor footprints. In addition, more heterogeneously sized footprints (cyan rectangles) co-localize with the *NPAT* and *MSH2* TSSs, possibly corresponding to paused RNA polymerase II. Furthermore, the 930-944 bp molecules from all three promoters enabled assessment of long-range nucleosome organization. The +1 nucleosome proximal to the *NPAT* TSS in (A) was particularly well positioned, whereas the more distal +2 and +3 nucleosomes were less so. By contrast, the long-range nucleosome organization in the *MSH2* (B) and *CCN4* (C) promoters was more disorganized and not apparent, respectively. Moreover, MAPit-FENGc of ~800 bp of 5' flanking sequence from *MSH2* and *CCN4* identified clear transitions between 5mCG depletion and hypermethylation.

FIG. 50. Target sequence length and GC content negatively correlate with filtered CCS read number obtained from primary mouse monocytes using MAPit-FENGc with EM-seq. Bone marrow was collected from the spines of four 4-months-old female C57BL/6J mice. Female mice

were selected to examine expected epiallelic differences on the X chromosome between active and inactive gene copies. Extraneous tissues were removed from the spine, which was crushed and the homogenate filtered through a nylon mesh. Monocytes in the filtrate were enriched by a negative isolation protocol that depletes non-monocytes (Militenyi, 130-100-629). Purified monocytes were allowed to recover from the isolation procedure for 3 hr in growth medium before processing by the MAPit-FENGc protocol. For this experiment, the primers were designed by newly developed program, FENGc oligonucleotide designer (FOLD; github.com/albertoriva/FOLD). The programs searches an input file of gene names or genome coordinates for primers that avoid repeats and satisfy criteria of FIG. 1, such as locating the 'overlapping' residues that create 1-nt 3' flaps. Other command-line options include, but are not limited to, increasing the length of default 500-nt sequences and percentage tolerance of departure from this specified length, specification of annotated TSS (e.g., RefSeq), and minimum and maximum primer melting temperature (T_m). To provide a computational solution for all desired 78 targets that satisfied all FOLD settings, target sizes were permitted to range from 474-987 nt (mean 620 nt; TABLE 1, Sheet 5 and TABLE 2, Sheet 4, Mouse ~620-nt Targets). The resulting 78-flap adapter panel, including 77 genes with known or suspected roles in the cellular inflammatory response and 1 control promoter (*Cox8a*), was used for preferred FENGc enrichment. The four resulting mouse monocyte FENGc libraries were barcoded, pooled, and sequenced on a PacBio Sequel II instrument. Demultiplexed, high-fidelity CCS reads were aligned to the complete mm9 build of the mouse genome *versus* specific target reference sequences: 20-29% did not align to either reference, yielding 71-80% on-target reads. Overall, FENGc detected 71-75 targets (91-96%) with ≥ 1 read in each sample (TABLE 10 and TABLE 11). Data are plotted as the natural logarithm transformation of filtered CCS read number + 1 *versus* (A) target length and (B) percentage GC content for each of the 78 targets.

FIG. 51. Reproducible epigenetic profiling of representative regions from primary mouse monocytes using MAPit-FENGc with EM-seq. High-fidelity CCS reads (≥ 5 SMRT sequencing passes; TABLE 10) were filtered further for $\geq 95\%$ conversion of HCH to HTH and $\geq 95\%$ coverage of each reference sequence length (TABLE 11) to avoid multiple alignments to homologous gene orthologs. Shown are methylscaper plots of methylation at HCG (left panel of each pair) and GCH (accessibility; right panel of each pair) of molecules from the enriched regions. Promoter sequences of (A) 474 bp of *Hsf1*, divergently transcribed with *Bop1*, (B)

514 bp of *Btk* from the X chromosome, (C) 605 bp of *Pik3r3*, and 576 bp of *Tlr4* were captured and enriched by FENG C from gDNA isolated from M.CviPI-treated mouse monocytes. All sequences were among those enriched for the 78 ~620-nt targets (TABLE 1, Sheet 5 and TABLE 2, Sheet 4) by the preferred FENG C protocol. Each HCG and GCH position in base pairs along each amplicon is indicated at bottom of its respective Mouse 4 panel. The black bar indicates 147 bp, the size of a nucleosome core lacking the linker; this and other features and the distances between them are drawn to scale. Promoter coordinates at top are relative to the *Hsf1* TSS in (A) and relative to each single TSS in (B), (C), and (D). The content of the panels and key at bottom are as described in the FIG. 34 legend. Note the highly consistent chromatin architectures detected at each locus (A-D) between the four different monocyte samples (Mouse 1-4). To determine rigorously if at least one mouse monocyte sample exhibited statistically significant epigenetic differences, smoothed moving averages (20-bp window) of DNA methylation and accessibility across each gene region were modeled using a mixed effects ANOVA. Testing was limited to 43 amplicons with ≥ 100 filtered CCS reads per sample and good diversity, i.e., absence of apparent duplicates (TABLE 11). Mean *P* values for H5mCG and G5mCH before correction for multiple testing were $> .90$ (range 0.17-1.0) and all *P* values after correction were 1.0 (TABLE 11). This indicates that for each target, none of the four mouse monocyte samples had significantly different levels of either DNA methylation or chromatin accessibility. On examination of the patterns of DNA methylation and chromatin accessibility on single molecules, *Hsf1* in (A) exemplifies an exceptionally open promoter in the four bone marrow-derived monocyte samples (2,333 of 2,334 molecules bearing large NFRs). As observed in FIG. 38, a locus with such a high degree of openness rules out incomplete cell permeabilization and chromatin probing as trivial bases for observed inter-locus differences in accessibility. Consistent with X inactivation, the *Btk* promoter from each mouse in (B) showed a cluster of highly accessible epialleles that were devoid of H5mCG and strongly anti-correlated with molecules displaying as few as two H5mCG. The *Pik3r3* promoter in (C) harbored a remarkably well-positioned -1 nucleosome and incremental, preferential sliding of the $+1$ nucleosome, expanding/contracting the NFR mainly on one side. This is intriguing, given that most promoters, such as *Tlr4* in (D) displayed movement of both the -1 and $+1$ nucleosomes, expanding/contracting the NFR on both sides.

FIG. 52. Reproducible epigenetic profiling of representative regions from primary mouse monocytes using MAPit-FENGc with EM-seq. Shown and drawn to scale are methylscaper plots of methylation at HCG (left panel of each pair) and GCH (accessibility; right panel of each pair) of molecules from the enriched regions. Promoter sequences of (A) 606 bp of *Hsp90ab1*, (B) 492 bp of *Cxcr4*, (C) 556 bp of *Mapk15*, and (D) 616 bp of *Src* were captured and enriched by preferred FENGc from gDNA isolated from M.CviPI-treated mouse monocytes. Promoter coordinates at top are relative to each single TSS in all panels. The FENGc target panel used to enrich sequences, CCS read filtering, and statistical test for epigenetic differences between different monocyte samples are as described in the FIG. 51 legend. As in FIG. 51, in this figure, none of the four mouse monocyte samples from each of the four targets exhibit significant epigenetic differences from the other three samples (TABLE 11). The content of the panels and key at bottom are as described in the FIG. 34 and FIG. 51 legends. As for the *Tlr4* promoter in FIG. 51D, movement of both the -1 and +1 nucleosomes led to NFR contraction/expansion on both sides of the NRF in the promoters of *Hsp90ab1* (A) and *Cxcr4* (B). Surprisingly, the *Cxcr4* promoter displayed a 36-bp zone of H5mCG (orange rectangle) in the accessible NFR, only 12 bp from a strongly footprinted TF. Compared to the above promoters, the *Mapk15* gene body (C) (+1,660 to +2,215) was hypermethylated with arrays of randomly positioned nucleosomes and short, linker-length NFRs. Chromatin -571 to +45 of the TSS of *Src* (D) was similarly organized, consistent with low-level expression of *Src* in mouse monocytes [Schaum, 2018 #190]. Interestingly, a sequence in *Src* with strong CTCF binding site homology [Hashimoto, 2017 #188] (black rectangle) conferred clear protection against endogenous 5mCG in many molecules. Taken together with FIG. 51, the data demonstrate that MAPit-FENGc is effective at discerning epigenetic landscapes of purified primary cells, with striking inter-sample reproducibility.

TABLE 1. FENGc oligonucleotides used in this study.

TABLE 2. Characteristics of designed FENGc target sequences used in this study.

TABLE 3. CCS reads aligned to 11 human targets of ~300 nt vs. ~450 nt used in FENGc assay development with standard PCR vs. BS-PCR.

TABLE 4. CCS reads on- and off-target for 119 ~450-nt human targets in MAPit-FENGc.

TABLE 5. Filtered CCS reads aligned to 119 ~450-nt human targets in MAPit-FENG C using EM-seq.

TABLE 6. Differential H5mCG and G5mCH determined by MAPit-FENG C using EM-seq of ~450-nt human targets with ≥ 50 filtered reads.

5 TABLE 7. SNPs and indels detected by FENG C using standard PCR of 119 ~450-nt human targets.

TABLE 8. CCS reads on- and off-target for 45 ~940-nt human targets in MAPit-FENG C using EM-seq.

10 TABLE 9. Filtered CCS reads aligned to 45 ~940-nt human targets in MAPit-FENG C using EM-seq.

TABLE 10. CCS reads on- and off-target for 78 ~620-nt mouse targets in MAPit-FENG C using EM-seq.

TABLE 11. Filtered CCS reads aligned to 78 ~620-nt mouse targets in MAPit-FENG C using EM-seq and *P* values.

15 TABLE 12. Recommended FENG C oligonucleotide concentrations for different numbers of targets.

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention
20 belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference.

Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics, protein, and nucleic acid
25 chemistry, and hybridization described herein are those well known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more

specific references that are cited and discussed through the present specification unless otherwise indicated.

The terms “complement,” “complementary,” or “complementarity” as used herein with reference to polynucleotides (i.e., a sequence of nucleotides such as an oligonucleotide or a genomic nucleic acid) related by the base-pairing rules. The complement of a nucleic acid sequence as used herein refers to an oligonucleotide which, when aligned with the nucleic acid sequence such that the 5' end of one sequence is paired with the 3' end of the other, is in “antiparallel association.” For example, the sequence 5'-A-G-T-3' is complementary to the sequence 3'-T-C-A-5'. Certain bases not commonly found in natural nucleic acids may be included in the nucleic acids of the present invention and include, for example, inosine, 7-deazaguanine, and 5-methylcytosine. Complementarity need not be perfect; stable duplexes may contain mismatched base pairs or unmatched bases. Those skilled in the art of nucleic acid technology can determine duplex stability empirically considering a number of variables including, for example, the length of the oligonucleotide, base composition and sequence of the oligonucleotide, ionic strength, and incidence of mismatched base pairs. Complementarity may be “partial” in which only some of the nucleic acids' bases are matched according to the base-pairing rules. Or, there may be “complete,” “total,” or “full” complementarity between the nucleic acids.

The terms “Flap endonuclease 1” and “FEN1” as used herein refer to a nucleolytic enzyme that acts as both 5'-3' exonucleases and structure-specific endonucleases on specialized DNA or RNA structures that occur during the biological processes of DNA replication, DNA repair, and DNA recombination. FENs can also cleave RNA, i.e., when the oligo complex hybridizes to an RNA target (Lyamichev et al., 1993). This contributes to the removal of RNA primers in Okazaki fragments during lagging strand DNA synthesis. The endonuclease activity of FEN1 was initially identified as acting on a DNA duplex which has a single-stranded 5' overhang on one of the strands (termed a “5' flap,” hence the name flap endonuclease). FEN1 catalyzes hydrolytic cleavage of the phosphodiester bond at the junction of single- and double-stranded DNA.

The terms “enrichment” and “enriching” as used herein refer to capturing selective genomic regions of interest for targeted sequencing of just the coding regions, specific genes, or segments of chromosomes that are relevant to a particular experiment or disease.

5 The term “oligonucleotide” as used herein refers to a short polymer composed of deoxyribonucleotides, ribonucleotides, or any combination thereof. Oligonucleotides are generally between about 10, 11, 12, 13, 14, 15, 20, 25, or 30 to about 150 nucleotides (nt) in length, more preferably about 10, 11, 12, 13, 14, 15, 20, 25, or 30 to about 70 nt.

The term “source sequence” as used herein refers to a sequence in which a sequence of interest is contained. The sequence of interest is cleaved from the source sequence and enriched.
10 The source sequence can be comprised of DNA, such as in a genome, and RNA.

The terms “Taq polymerase” and “Taq” as used herein refer to a thermostable DNA polymerase I named after the thermophilic eubacterial microorganism *Thermus aquaticus*, from which it was originally isolated by Chien et al. in 1976. It is frequently used in the polymerase chain reaction (PCR), a method for greatly amplifying the quantity of short segments of DNA.

15 Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit (unless the context clearly dictates otherwise), between the upper and lower limit of that range, and any other stated or intervening value in that stated range, is encompassed within the disclosure. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges and are also encompassed within the disclosure,
20 subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the disclosure.

Unless specifically stated or obvious from context, as used herein, the term “about,” or “approximately,” or symbol “~” is understood as within a range of normal tolerance in the art,
25 for example within 2 standard deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from context, all numerical values provided herein are modified by the term about.

Unless specifically stated or obvious from context, as used herein, the singular form “a,” “an,” and “the” include plural references. For example, the term “an oligonucleotide” or “an oligo” includes a plurality of oligos, including mixtures thereof.

All publications and patents cited in this specification are herein incorporated by
5 reference as if each individual publication or patent were specifically and individually indicated to be incorporated by reference and are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present disclosure is not entitled to antedate such publication by prior
10 disclosure. Further, the dates of publication provided could be different from the actual publication dates that may need to be independently confirmed.

Unless otherwise indicated, the present disclosure is not limited to particular materials, reagents, reaction materials, manufacturing processes, or the like, as such can vary. It is also to be understood that the terminology used herein is for purposes of describing particular
15 embodiments only and is not intended to be limiting. It is also possible in the present disclosure that steps can be executed in different sequence where this is logically possible.

DETAILED DESCRIPTION

The methods disclosed herein for sequence enrichment exploit the many advantages of DNA
20 strand capture and amplification afforded by patch ligation PCR (Varley and Mitra, 2008; Varley and Mitra, 2010), while eliminating the restrictive requirement to create capturable fragments using a select few restriction endonucleases. In this embodiment, two flap adapters, each consisting of an ~50-mer unmodified oligo bound to a first complementary universal PCR priming oligo, have been designed to form unpaired 5' flaps, with a preferred unpaired 1-nt 3'
25 flap, at both the 5' and 3' ends of a linear target DNA strand (FIG. 1 and FIG. 3). In some embodiments the oligos comprising the flap adapters may be shorter or longer, for example, 11 to 150 nt, contain base modifications, for example, 5mC, or some combination thereof. Cleavage of a plurality of the flap structures by the 5' flap cleavage activity of a single thermostable enzyme, such as *Thermococcus* 9°N™ FEN1 (hereafter, FEN1) or Taq DNA polymerase I

(hereafter, Taq) (Lyamichev et al., 1993; Finger et al., 2012), releases a plurality of single-stranded DNA (ssDNA) fragments with two defined ends. The 5' end of each target is ligated to the first universal oligo, whereas each 3' end is contacted by a third target-complementary oligo that positions a second universal oligo for ligation. The three oligos that hybridize to the target strands are inexpensive as they are unmodified, i.e., do not contain methylated or biotinylated bases. Only the second universal oligo is modified by covalent modification that blocks its 3' hydroxyl group, which inconsequential to overall cost as that same universal oligo is always used. All oligos, therefore, are readily synthesized at high quality and yield as well as low cost, without needing purification. Enrichment of many target sequences can be conducted in the same reaction at a fraction of the cost of strategies based on recombinant CRISPR/Cas/gRNA or oligo hybridization and pull-down. Furthermore, high specificity is achieved because neither flap incision nor ligation tolerate DNA mismatches (Wu and Wallace, 1989; Kaiser et al., 1999; Lyamichev et al., 1999; Lyamichev et al., 1999; Hall et al., 2000; Lyamichev et al., 2000; Tsutakawa et al., 2017), and target DNA strands ligated to both universal priming sequences are preferentially PCR amplified subsequent to extensive exonuclease digestion.

In addition to genotyping, because it was shown that endonuclease activity of FEN1 is unaffected by 5mC (FIG. 8), an aliquot of the same FENGCC-enriched DNA can be processed in parallel for concurrent detection of DNA methylation and chromatin accessibility, termed MAPit-FENGCC. In certain embodiments, as in bisulfite patch PCR and MAPit-patch (Varley and Mitra, 2010; Nabils et al., 2014), unmethylated C is converted to U post-DNA enrichment, allowing targeting oligos to be designed without regard to DNA methylation. As an improvement, though, enzymatic conversion of unmethylated C to U (Tahiliani et al., 2009; Schutsky et al., 2018) has been substituted for bisulfite treatment, largely eliminating chemical degradation of DNA (Tanaka and Okamoto, 2007). This facilitates enrichment of 940-bp-long and potentially longer DNA fragments for phasing multiple epigenetic features, e.g., nucleosomes and transcription factor footprints, on contiguous DNA molecules.

In certain embodiments the DNA cleavage and ligation reactions prior to PCR require only <1 hour of hands-on time. Furthermore, the entire preferred FENGCC sequence enrichment protocol is done with serial addition of reagents and only one purification step prior to PCR amplification, minimizing loss of input genetic material. The streamlined sequence enrichment

procedure therefore requires only 50 ng human gDNA and is poised for automation in clinical applications. Library preparation for FENGCC genotyping and MAPit-FENGCC requires two and three days, respectively. Both protocols generate a mean specificity of long, mapped sequencing reads of ~80%.

5 Applications of FENGCC for Target Sequence Enrichment

Sequential reagent addition protocols for FENGCC enrichment of targeted DNA sequences with nucleotide-level precision

This embodiment describes protocols in which nucleic acids are sequentially modified by reagent addition in a single tube, without multiple purifications, which minimizes losses of input biological material and maintains compatibility with robotic automation. The embodiments described herein require only one or two purifications. Toward this goal, two FENGCC protocols (FIG. 1 and FIG. 2) were devised. In certain embodiments, it is preferred that genomic DNA (gDNA) is first fragmented with sonication or, alternatively, with restriction enzyme digestion or another suitable method. Next, the gDNA is denatured and a plurality of single-stranded DNA (ssDNA) sequences of interest is contacted in solution at both ends with a matched pair of flap adapters (FIG. 1 and FIG. 2, step 1a; one sequence of interest shown for clarity). In the preferred FENGCC protocol (FIG. 1), each flap adapter consists of either a target sequence-specific oligo 1-N or oligo 2-N, both of which have the same 3' tail that anneals to and positions for ligation a complementary universal oligo 1-N (oligo U1-N; N is a 3'-terminal A, C, G, or T). In this naming scheme, flap adapters consist of: the U1-T oligo hybridized to an oligo 1-T (FIG. 3); the U1-A oligo hybridized to an oligo 1-A; and so on. The identities of the oligo U1-N 3'-terminal nucleotide and the first base-paired nucleotide of the target sequence within the downstream DNA duplex are the same (FIG. 3, highlighted in pink). This design creates a single-nucleotide "overlap," enabling the nucleotide in the target sequence to compete with and displace the oligo U1-N 3'-terminal nucleotide and base pair with the corresponding, complementary nucleotide in oligo 1-N (FIG. 3, red type). This creates an unpaired 1-nt 3' flap in addition to the unpaired 5' flap, referred to as a double flap (Lyamichev et al., 1999). The double flap with a 1-nt 3' flap increases the binding affinity for and cleavage rate of 5' flaps by FENs, including thermostable Flap endonuclease 1 (FEN1) and Taq (Lyamichev et al., 1999; Lyamichev et al., 1999; Friedrich-Heineken et al., 2003). Human FEN1 can also cleave double flaps with 2-, 10-, and 20-nt 3'

flaps, although with reduced efficiency and precision (Friedrich-Heineken et al., 2003). In a preferred embodiment, the 3' flap is restricted to 1 nt in length by ensuring that the nucleotide at the base of the 5' flap and adjacent to the overlap nucleotide cannot base pair with the nucleotide on the opposite DNA strand in oligo 1-N. For instance, V (A, C, or G) in the 5' flap will not base pair with the indicated A in oligo 1-T (FIG. 3).

FEN1, Taq, and related FENs incise the phosphodiester bond of the sequence of interest in double flaps with a 1-nt 3' flap efficiently and uniformly after both the 5' flap and the ribose of the first base pair within the downstream duplex (FIG. 1 and FIG. 2, step 1a and FIG. 3, red arrowheads). In a multiplexed reaction, this liberates a plurality of target sequences with a 5'-terminal phosphate located immediately downstream of a 1-nt gap (FIG. 1, step 1b). The 1-nt gap is filled by the oligo U1-N 3'-terminal nucleotide, which is subsequently ligated by Ampligase (Lucigen) to the 5' ends of the plurality of target sequences (FIG. 1, step 2). By contrast, because 5' flap incision severs each flap adapter 2 from the sequence of interest, additional adapter 3 is used to position the oligo U2 for ligation to the liberated 3' end. The oligo U2 is synthesized with a ligatable 5' phosphate and, at its 3' end, five phosphorothioate bonds and the 3'-terminal hydroxyl blocked by a three-carbon spacer. These 3' oligo modifications protect target strands to which the oligo U2 ligated against 3' to 5' degradation by both Exo I and Exo III (FIG. 1; step 3). By contrast, all other DNA molecules are degraded (except unligated oligo U2), dramatically enriching the targeted sequences of interest.

Note that the oligo U1-N of flap adapter 2 also ligates to the 5' end of the downstream sequence. Therefore, a series of flap adapters, each annealing successively farther downstream, can be designed to facilitate 'walking' of contiguous regions.

In certain embodiments, the ligated products can be processed for genotyping, epigenetic analysis, or both. For sequence genotyping, the ligated products are purified for the first time in the reagent addition protocol and then amplified using standard PCR with the oligo U1 and U2 primer (FIG. 1, step 4). Alternatively, the purified, enriched target sequences can be subjected to bisulfite or enzymatic conversion of C to U (deamination) prior to PCR amplification, termed methyl-PCR, to detect 5mC and 5hmC at nucleotide-level resolution. In the invention herein, the PCR or methyl-PCR amplicons are subsequently ligated to barcoded hairpin adapters to create

SMRTbell™ templates for multiplexed, long-read, and high-fidelity sequencing on a Pacific Biosciences (PacBio) instrument.

For genotyping and methyl sequencing on Illumina instruments, the P5 and P7 Illumina sequences are added to the 5' end of the U1 and U2 primers by PCR amplification without and after deamination, respectively, as previously described (Varley and Mitra, 2010; Nabils et al., 2014). It has also recently been reported that the accuracy of Oxford Nanopore Technology sequencing can be increased to 94% by the rolling circle to concatemeric consensus (R2C2) method (Volden et al., 2018). In principle, this method could be applied to captured sequences of ~1 kb, a limitation imposed by circularization of double-stranded DNA (Volden et al., 2018; Shore et al., 1981; Shore and Baldwin, 1983). Alternatively, high-accuracy, long-read Nanopore or PacBio sequencing can be obtained using unique molecular identifiers (UMIs) (Karst et al., 2020). The current embodiment can also be used to capture large megabase fragments followed by single-molecule nanopore sequencing (Bennett-Baker and Mueller, 2017; Gabrieli et al., 2018; Gilpatrick et al., 2020) in combination with UMIs to improve sequencing accuracy (Karst et al., 2020).

The FENGCC protocol can be stopped after any step and the samples stored at -20°C before proceeding. The hands-on time for FENGCC processing of one multiplexed sample is <1 hr for standard PCR and <2 hr for methyl-PCR. The entire standard PCR protocol can be completed in two days or three days for methyl-PCR.

EXAMPLES:

Cell Culture

Human fetal telencephalic NSC and human GBM cell lines (L0 and Nx18-25) were cultured in complete NSC medium (basal medium + proliferation supplement at a 9:1 ratio; NeuroCult™ NS-A Proliferation Kit, STEMCELL TECHNOLOGIES, 05751) supplemented with penicillin-streptomycin (1% final concentration; ThermoFisher Scientific Gibco, 15140122), 20 ng/ml human recombinant epidermal growth factor (STEMCELL TECHNOLOGIES, 78006.1), 10 ng/ml human recombinant basic fibroblast growth factor (STEMCELL TECHNOLOGIES, 78003.1), and 0.679 U/ml heparin (Sigma, H3149). For NSC culture, 10 ng/ml leukemia inhibitory factor (Millipore, LIF1050) was also added. The cells were

maintained in a humidified incubator at 37°C and 5% CO₂. A standard protocol was used for passaging the NSC{PMID: 27030542} and GBM cells {PMID: 22064695}, whereby the neurospheres were collected by centrifugation at 110g for 5 min) every 7-10 days. The pellet was re-suspended in 0.05% (w/v) trypsin (0.53 mM EDTA, ThermoFisher Scientific, 25300062) prewarmed to 37 °C. Soybean trypsin inhibitor (ThermoFisher Scientific, 17075029) was then added and gentle pipetting used to dissociate the neurospheres into single cells for re-plating.

Ethical Approval

Animal protocols were approved by the University of Florida Institutional Animal Care and Use Committee (IACUC Protocol Nos: 201807422 & 201910745). 'Guide for the Care and Use of Laboratory Animals' was adhered to as prepared by the Committee for the Update of the Guide for the Care and Use of Laboratory Animals of the Institute for Laboratory Animal Research, National Research Council. Mice were closely monitored for signs of dehydration, weight loss, impaired mobility or physiological signs of underlying disorders such as labored breathing or respiratory distress.

Animal Description and Treatment

C57BL/6J female mice were used to examine detection by MAPit-FENGc of the epigenetic mixture of transcriptionally inactivated and active copies of the X chromosome. All mice were 8-10 weeks of age upon arrival and individually housed on a 12 h dark/12 h light cycle at 19-22°C and 30-60% humidity, with standard chow diet and water provided *ad libitum*. Prior to cell collection, mouse anesthetization was induced and maintained with 5.0% and 1.5% isoflurane USP (NDC 14 043-704-06, Patterson Veterinary Supply, Inc.), respectively, using an Eagle Eye Model 150 anesthesia machine (Jacksonville, FL, USA). The depth of anesthesia was monitored by the absence of pedal withdrawal reflex.

Single-Molecule MAPit Methylation Footprinting

MAPit was done on permeabilized cells to mark accessible chromatin. In detail, two million cells were first washed with cold PBS with 0.015% (w/v) sodium azide. Cells are pelleted and washed with 500 µL ice-cold cell resuspension buffer (20 mM HEPES, pH 7.5, 70 mM MgCl₂, 0.25 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.5% (v/v) glycerol, freshly supplemented 10 mM DTT and 0.25 mM PMSF). Cells were pelleted and resuspend in 180 µL

cell resuspension buffer with 0.05% (w/v) digitonin and incubated on ice for 10 min. Cell permeabilization was checked by trypan blue staining, in which 100% of cells should be stained blue before preceding to the next step. Cells were then treated as indicated with or without M.CviPI (100 U/million cells) supplemented with fresh 160 μ M SAM followed by a 15 min incubation at 37 °C. The reactions were stopped by addition of equal volume stop buffer (1% (w/v) SDS, 100 mM NaCl, 10 mM EDTA) and vortexed briefly at medium speed. Nuclei were treated with RNase A for 30 min at 37 °C followed by 100 μ g/mL Proteinase K treatment at 50 °C overnight. Genomic DNA was extracted using phenol-chloroform-isoamyl alcohol (25:24:1, v/v) phase separation, followed by ethanol precipitation and resuspension in water.

10 Isolation of Monocytes from Mouse Bone Marrow

Bone marrow was collected from spines (below skull to above tail) cleaned of extraneous tissues. All subsequent steps were conducted under sterile conditions. First, the tissue was crushed at room temperature using a ceramic mortar and pestle in a sterile solution of 10 ml of phosphate-buffered saline (PBS), pH 7.2, 2 mM EDTA, 0.5% (w/v) bovine serum albumin (BSA), by mixing MACS BSA Stock Solution (Miltenyi, 130-091-376) and Biotec, autoMACS® Rinsing Solution (Miltenyi, 130-091-222) in a ratio of 1:20. Subsequently, the crushed spine was removed, and the homogenate was filtered through pre-separation filters with 30- μ m nylon mesh (Miltenyi Biotec, 130-041-407,) into a 15-mL Falcon® tube. The cellular filtrate was then washed by centrifugation at 300g for 10 min, removal of the supernatant, and resuspended in the same PBS/BSA solution for cell counting (Heska, Element HT5 Hematology Analyzer). Monocytes were then enriched using a negative isolation protocol specific to mouse bone marrow (Miltenyi Biotec, 130-100-629) according to the manufacturer's protocol. Briefly, ~40-80 million washed cells were incubated with FcR blocking reagent and cocktail of mouse biotinylated-antibodies, washed with PBS/BSA, resuspended in degassed RPMI 1640 medium (ThermoFisher, 11875101) containing 1% (w/v) penicillin-streptomycin (ThermoFisher Gibco, 15140163), and efficiently depleted of non-target cells by addition of magnetic microbeads conjugated to mouse anti-biotin monoclonal IgG1 antibodies and passage through LS ferromagnetic columns (Miltenyi Biotec, 130-042-401). The flow through, containing highly enriched bone marrow-derived monocytes, was collected into a 1-mL centrifuge tube. To allow recovery from the collection process, a mean of 1.2 million monocytes were plated in a well on a

96-well plate and incubated in a humidified 37°C incubator at 5% CO₂ in an RPMI 1640 containing 1% penicillin-streptomycin solution for 3 h before harvesting for MAPit-FENGCC.

FENGCC Oligonucleotide Designer (FOLD)

The ~620-nt panel of FENGCC primers was designed by newly developed program, FOLD. The programs searches an input file of gene names or genome coordinates for primers that avoid repeats and satisfy criteria of FIG. 1, such as locating the ‘overlapping’ residues that create 1-nt 3’ flaps. Other user-defined, command-line options include, but are not limited to, increasing the length of default 500-nt sequences and percentage tolerance of departure from this specified length, specification of annotated TSS (e.g., RefSeq), and minimum and maximum primer melting temperature (T_m). The program is available online at github.com/albertoriva/FOLD.

Statistics

For the 0.1% L0:99.9% NSC gDNA mixture (FIG. 37), a proportions test was conducted in R using the command `prop.test()`, with a null hypothesis that the observed proportion (10 hypermethylated:1781 unmethylated epialleles) is less than 0.1%. The P value of < 0.0001 indicates that the proportion of hypermethylated epialleles detected was at least or greater than 0.1%.

Generalized estimating equations were used to model the effect of cell line (human NSC *versus* GBM Nx18-25) on the per molecule proportions of endogenous H5mCG (number H5mCG/(number H5mCG + number HCG); left panels) and G5mCH (number G5mCH/(number G5mCH + number GCH); right panels). Only targets with ≥ 50 total CCS reads in the combined replicates (≥ 22 in any single replicate) were considered (TABLE 5), with filtering as described in the FIG. 34 legend. Errors were modeled as normally distributed, and the correlation structure was assumed exchangeable within each replicate. The `geeglm` function from the `geepack` v1.3-2 R package was used in R version 4.1.0. P values were corrected for multiple testing using the Bonferroni method with an alpha of 0.05 to control the false discovery rate (TABLE 6).

To determine if MAPit-FENGCC detected significant epigenetic differences in at least one of the four analyzed mouse monocyte samples, smoothed moving averages (20-bp window) of DNA methylation and accessibility across each gene region were modeled using a mixed effects ANOVA. Testing was limited to 43 amplicons with ≥ 100 CCS reads per sample and good

diversity, i.e., absence of many duplicates (TABLE 11). The model was fit using the gls function in the nlme v3.1-152 package in R version 4.1.0. Each sample was treated as a random effect and correlation along the gene region was modeled as that from an autoregression moving average model (ARMA). Autocorrelation parameters were estimated using the auto.arima function from the forecast v8.15 R package with a maximum possible value of 1 to avoid overfitting. If the differencing order was estimated as zero, then first order differences were taken of DNA methylation and accessibility. If the autoregressive and moving average parameters were both estimated as zero, then an autoregressive model of order one was used. The *P* values are of the interaction term of base pair and replicate which tested for differences between replicates across the gene region. *P* values were corrected for multiple testing using the Bonferroni method with an alpha of .05 to control the false discovery rate (TABLE 11).

EXAMPLE 1: FENG C procedure for sequence enrichment from gDNA

In certain embodiments, flap adapters 1 and 2 as well as corresponding adapters 3 were designed to capture ~300 nt, ~450 nt, and ~940 nt spanning the transcription start sites of 11, 119, and 45 human genes encoding proteins with DNA repair or cancer-associated functions, respectively (Table 1 and Table 2). In certain embodiments, flap adapters 1 and 2 as well as corresponding adapters 3 were designed to capture ~620 nt of 78 mouse genes expressing products with functions in the cellular inflammatory response.

In certain embodiments, the preferred FENG C procedure with a 1-nt 3' flap, the gDNA was fragmented by either of two methods: 1) Digestion with SpeI-HF (New England Biolabs) in the CutSmart Buffer (New England Biolabs) in 20 µl volume and incubation for 1 h at 37°C, followed by 20 min at 80°C; or 2) sonication with a UCD-200 Bioruptor (Diagenode) on the high setting for 25 sec in 100 µl of sterile distilled and deionized H₂O (ddH₂O; MilliQ), followed by SpeedVac™ reduction of the volume to 20 µl. Cleavage of 5' flaps was performed by combining 1 µl of 10 µM U1-T oligo, 2 µl of a mixture of oligos 1-T and oligos 2-T, the concentration of each depending on the number of target regions of interest (TABLE 12; calculated with formula), 1x PCR Buffer (Qiagen), 3 U APEX Taq (Genesee Scientific), and sterile ddH₂O to bring the total volume to 35 µl. For cleavage with FEN1, 32 U FEN1 and 1x FEN1 Buffer (New England Biolabs) were substituted for Taq and PCR buffer. The reactions

were incubated for 3 min at 95°C and for 20 min at 65°C, followed by 14 cycles of 30 sec at 95°C and 65°C for 10 min. Next, 2 µl of a mixture 10 µM of oligos 3-T, the concentration of each depending on the number of target regions of interest (TABLE 12; calculated with formula), 1 µl of 10 µM oligo U2, 1x Ampligase® reaction buffer (Lucigen), 10 U of Ampligase (Lucigen), and ddH₂O were added to bring the total volume to 45 µl. The ligation reaction conditions were identical to those employed above for FENG C of linearized plasmid. Unprotected DNA was removed by addition of 20 U of Exo I (New England Biolabs) and 100 U of Exo III (New England Biolabs), followed by incubation for 1 hr at 37°C and for 20 min at 80°C.

Surviving DNA sequences were subjected to standard PCR, BS-PCR, or EM-PCR. For standard PCR, the captured DNA sequences (~300-nt and ~450-nt targets) were purified with MinElute PCR Purification Kit (Qiagen) or by addition of 1.8x volume of AMPure XP beads, and PCR was performed with HotStar Taq (Qiagen). The PCR reaction was 95°C for 5 min, followed by 30 cycles of 94°C for 30 sec, 57°C for 30 sec, 72°C for 1 min. For DNA methylation analysis of ~300-nt and ~450-nt targets, the eluted, enriched sequences were bisulfite converted and PCR amplified using the same conditions, except that 35 cycles were employed. For EM-PCR of ~450-nt, ~620-nt, and ~940-nt targets, the captured sequences were purified by addition of 1.8x volumes of either AMPure XP beads or NEBNext beads, enzymatically converted according to the EM-seq manual (New England Biolabs), and eluted in 14-25 µl of sterile ddH₂O. PCR amplification with 500 nM each of U1 and U2 primers and HotStar Taq (Qiagen; ~450-nt targets). 2x KAPA HiFi HotStart Uracil+ ReadyMix (Roche; ~620-nt, and ~940-nt targets) in 50 µl for 5 min at 95°C, followed by 30 cycles of 20 sec at 98°C, 15 sec at 62°C, 30 sec at 72°C, and one final 1-min extension at 72°C. All oligos used are listed in Table 1.

For the alternate FENG C protocol without a 3' flap, digestion of purified gDNA with SpeI-HF, 5' flap cleavage, and first-round ligation to the 3' end of the cut sequences of interest were as described for the preferred protocol. After the second-round ligation using the same conditions, the reactions were purified with MinElute PCR Purification Kit (Qiagen) and eluted with 14 µl of H₂O, then incubated with 1 x QIAGEN PCR Buffer (Qiagen), 0.2 mM dNTP, 500 nM of U1 and U2 primers, 2 units of HotStar Taq (Qiagen) in 20 µl at 95°C for 5 min,

followed by 35 cycles of 30 sec at 94°C, 30 sec at 57°C, 1 min at 72°C. All oligos used are listed in Table 1.

EXAMPLE 2: Optimization of precision cleavage of ssDNA sequences of interest with Taq or FEN1

It was found that Taq and FEN1 cleave the phosphodiester bond of DNA structures with ssDNA 5' flaps, producing a 3'-terminal hydroxyl and 5'-terminal phosphate (Lyamichev et al., 1993; Finger et al., 2012; Kaiser et al., 1999; Lyamichev et al., 1999; Lyamichev et al., 1999; Hall et al., 2000; Tsutakawa et al., 2017). To develop FENGc, the configuration of oligos used for 5' flap formation and cleavage, the amount of needed enzyme activity, and the number of cleavage cycles were optimized. The efficiency of Taq incision of 5' flaps containing different combinations of base-paired nucleotides at the location immediately 5' of the scissile phosphate was tested, using annealed oligos as substrates (FIG. 4A). In this experiment, four 200-mer oligos with a sequence derived from plasmid pGEM-3Z/601 (Lowary and Widom, 1998), with either A, C, G, or T at nt 130 (200-N oligos), were used as proxy DNA sequences of interest (Table 1, Sheet 1). These oligo targets were contacted by their respective flap adapter 1, containing the oligo U1 bound to the corresponding target-complementary 200 oligo 1-N, i.e., with the nucleotide complementary to nt 130. This forms four different structures with different 5' flaps (FIG. 4A), i.e., without a 1-nt, displaced 3' flap.

Initially, a 5' flap substrate consisting of the 200-T oligo, its complementary flap-T oligo (with A base paired with T130 of the 200-T oligo), and the oligo U1 was used to determine the amount of Taq needed to achieve maximal 5' flap cleavage. The cleavage efficiency was determined using the High Sensitivity DNA Chip on the Agilent 2100 Bioanalyzer (Agilent Genomics) (FIG. 4B). The areas under the peaks of cut and uncut 200-mer oligo were integrated and used to calculate the percentages of cut oligo (FIG. 5). A digestion plateau was reached with 1 unit (U; based on polymerization activity) of Taq (FIG. 5A). Using the same 5' flap substrate and 1 U Taq, it was determined that the majority of substrate cutting was achieved with one cycle of hybridization and 5' flap formation; additional cycles yielded a trend toward a plateau with slightly increased cutting (FIG. 5B).

Next, the extent of digestion of 16 different 5' flap substrates was determined (see Table 1, Sheet 1 for oligo sequences). The identity of the nucleotide immediately 5' of the

cleavage site (in the absence of a 3' flap) on the 200-mer did not measurably affect cleavage efficiency (FIG. 5C; compare A+U1, C+U1, G+U1, and T+U1). Extending the 3' end of each oligo U1 by 1 nt (i.e., using oligo U1-Ns instead of the oligo U1), creating an unpaired 1-nt 3' flap, also did not measurably affect digestion by Taq under the employed reaction conditions (FIG. 5C; compare A+U1 with A+U1-C, A+U1-G, and A+U1-T, etc.). FEN1 achieved an overall higher digestion efficiency than Taq in a reaction containing the same 5' flap structure used in FIG. 5B (200-T oligo, 200 oligo 1-T, and oligo U1) (FIG. 6). In reactions where DNA was fragmented with a restriction enzyme, the presence of less than 1x CutSmart buffer (New England Biolabs) did not affect the percentage of subsequent flap cutting by either Taq or FEN1, the latter of which again yielded the highest level of cleavage (FIG. 7).

EXAMPLE 3: Determination of digestion efficiency of Taq DNA polymerase and FEN1

Flap adapters were designed to consist of oligos 1-N with 3' tails that anneal to the oligo U1 or oligo U1-Ns. The non-annealed 5' ends of flap adapters were designed to contact specific sequences in ssDNA target sequences of interest to form 5' flap structures. The 200-N oligos provided the 5' flap and cleavage site in the flap structure. For cleavage, 500 nM each of oligo U1, one of four 200 oligos 1-N, and respective one of four 200-N oligos were incubated with APEX Taq (Genesee Scientific) or HotStar Taq (Qiagen) in 1 x PCR Buffer (Qiagen; referred to as Taq buffer in this study) or with thermostable FEN1 (New England Biolabs) in 1 x ThermoPol Reaction Buffer (referred as FEN1 buffer in this study) in 20 µl final volume (FIG. 7). The reactions were initiated by incubation for 3 min at 95°C, then 20 min at 65°C, followed by the indicated number of cycles of 30 sec at 95°C and 65°C for 10 min. All oligos are listed in Table 1. For High Sensitivity DNA Chip assay, the DNA was purified with 5x AMPure XP Beads (Beckman Coulter) and loaded in Agilent 2100 Bioanalyzer system (Agilent Genomics). The amount of oligo with or without cleavage was indicated by digital peaks. The percentages of digestion was calculated by (mass of cut oligo)/(mass of cut oligo + mass of uncut oligo) x 100.

EXAMPLE 4: FEN1 cleavage of double flaps is insensitive to incorporated 5mC

DNA methylation interferes with cleavage by many restriction endonucleases (McClelland, 1981). Therefore, the effect of 5mC on the ability of FENs to bind and incise 5' flaps was determined. According to solved X-ray co-crystallographic structures, human FEN1 contacts the phosphodiester backbone and ribose sugars of several residues in the dsDNA

duplexes located upstream and downstream of a double flap (Tsutakawa et al., 2017; Tsutakawa et al., 2011). In FENG C of sequences with dense 5mC, only the upper DNA strand of the downstream duplex would be methylated because all of the utilized flap adapter oligos are unmethylated. Therefore, to examine the extent to which 5mC affects FEN1 activity, two 80-
 5 mers with the same sequence were synthesized. In one of these oligos, five 5mC residues were distributed near the predicted FEN1 cleavage and binding sites in downstream duplex within the double flap substrates formed by annealing to either a flap-A, -C, -G, or -T adapter. Each of the four unmethylated and four methylated double flap structures were cleaved with increasing amounts of FEN1 and the fraction of cut template was determined by quantitative real-time PCR.
 10 It is evident that 5mC did not exert a statistically significant effect on the cleavage efficiency of any of the double flap substrates (FIG. 8).

EXAMPLE 5: FENG C enrichment of a specific plasmid sequence after cutting a 2,453 nt 5' flap

In certain FENG C reactions, the lengths of 5' flaps will vary between different DNA
 15 targets of interest. To test the extent to which 5' flap length impacts FENG C efficiency, Taq was used to cut substrates containing 5' flaps of 87 nt or 2,453 nt (FIG. 9). To prepare these substrates, pGEM-3Z/601b (Dechassa et al., 2010), a modified pGEM-3Z/601, was first linearized with HindIII, localizing the 571-nt target DNA strand at one end and downstream of the 2,453 nt of the 5' flap sequence (FIG. 9A). The first and second A residues in the HindIII site
 20 (AAGCTT) were designated as positions 3025 and 1, respectively (FIG. 9A). The linearized plasmid DNA was divided equally into four reactions. Further digestion with DrrI and NdeI shortened the 5' flap sequence to 0 nt and 87 nt, respectively (FIG. 9B, bottom). A HindIII adapter, consisting of a pGEM-3Z/601b HindIII oligo and the oligo U2, was added to all four reactions to facilitate ligation of the oligo U2 to the common HindIII-cut 3' end. The NdeI-
 25 HindIII cut fragment serves as a positive control for ligation and amplification, i.e., with no 5' flap, by including an NdeI adapter (pGEM-3Z/601b NdeI oligo 1-T and oligo U1). To the remaining three reactions, a flap adapter 1 (U1-T oligo and pGEM-3Z/601b NdeI oligo 1-T) was added to test cutting of the 5' flaps of 87 nt and 2,453 nt when Taq was also included. Ligation of the cut 571-nt target sequence to the 47 nt total of U1-T and oligo U2s followed by bisulfite
 30 conversion and PCR yielded the expected 618-bp amplification product (FIG. 9B, lane 1).

Reactions with the 87-nt and 2,453-nt 5' flaps yielded a strong product indicative of site-specific 5' flap cleavage, ligation to the U1-T oligo (and oligo U2 to the HindIII-cut 3' end), and subsequent PCR amplification (FIG. 9B, lanes 2 and 3). Cutting of the 2,453-nt 5' flap was Taq dependent, as no product was detected when Taq was omitted (FIG. 9B; lane 4). It can be concluded that FENG C with Taq is able to enrich and amplify a target sequence with a 5' flap at least as long as 2,453 nt and possibly longer.

EXAMPLE 6: FENG C enrichment with Taq is insensitive to 5mC at the site of FEN cleavage

The effect of 5mC on the FENG C protocol was also tested. For this, pGEM-3Z/601b was C-5-methylated with the CpG DNA methyltransferase M.SssI in the presence of the methyl donor SAM. High-level, M.SssI-dependent methylation of the plasmid was demonstrated by inhibition of cutting by restriction enzyme HhaI (GCGC), which is sensitive to 5mC at its central CpG (FIG. 10A). FENG C was conducted on M.SssI-methylated and -untreated plasmid DNA that was linearized with HindIII (FIG. 10B, top panel). Three different flap oligo adapters, consisting of Methyl test pGEM-3Z/601b oligo 1-T, -G, or -C and the respective corresponding U1-T, -G, or -C oligo, directed cleavage of the three indicated phosphodiester bonds, releasing the 2,415-2,419 nt 5' flaps from the 605-609 nt target strands. After ligation and subsequent bisulfite conversion, PCR amplification with primers U1 and U2 showed no appreciable difference in the product of FENG C enrichment using all three unmethylated and methylated substrates (FIG. 10B, gel). This demonstrates that FENs can cleave phosphodiester bonds immediately adjacent to 5mC, and therefore the efficiency of the FENG C reaction is not overtly affected by 5mC.

EXAMPLE 7: FENG C performance of different flap complexes

With validation of the suitability of Taq and FEN1 for cleaving 5' flaps formed by flap adapters, the efficiency of FENG C enrichment was tested on different double flap structures, with overlapping A, C, G, or T. The first A in the NdeI site at position 2,453 of pGEM-3Z/601b was mutagenized to C, G, or T, creating a set of four plasmids (pGEM-3Z/601b-N) (FIG. 11A). As above, 0.2 ng (~0.1 fmole) of each HindIII-linearized plasmid was denatured and annealed to the HindIII adapter and each of the four respective flap adapters, e.g., pGEM-3Z/601b NdeI-2 oligo 1-A and U1-A oligo, pGEM-3Z/601b NdeI-2 oligo 1-C and U1-C oligo, etc. In each flap structure, the identity of the variant base at position 2,453 and the 1-nt 3' flap are the same.

Cleavage with either Taq or FEN1 of all four plasmids, followed by ligation, bisulfite conversion, and PCR with the U1 and U2 primers yielded the correct 619-bp products (FIG. 11B). Use of the flap adapter-G, however, also generated a high-molecular-weight smear consisting of repeats of primers U1-G and oligo U2s as determined by DNA sequencing.

- 5 FENG C reactions with the U1-T and flap 1-T oligos showed the highest, specific yield, and therefore this configuration was adopted as the preferred protocol.

EXAMPLE 8: Alternative FENG C procedure without an overlapping 1-nt 3' flap

It was found that FEN1 and Taq also cut 5' flap structures that lack a 1-nt 3' flap, leaving a 1-nt gap (Lyamichev et al., 1999; Lyamichev et al., 1999). In certain embodiments the 5' flap formation utilizes a flap adapter that contains the oligo U1, without the extra 3' nucleotide of the oligo U1-N. By this no 3' flap protocol (FIG. 2), target sequence cutting was accomplished as in the preferred procedure (FIG. 1 and FIG. 2, steps 1a-1b). However, after strand cutting, the 3' end of each of the plurality of target DNA strands is first ligated to the oligo U2 (FIG. 2, step 2), followed by digestion with both Exo I and Exo III (FIG. 2, step 3). Next, the 5' end of the DNA target strand is ligated to the corresponding oligo U1-N as dictated by the complementary nucleotide in the gap (FIG. 2, step 4). After a second round of incubation with Exo I and Exo III (FIG. 2, step 5), the plurality of target sequences is purified and amplified by standard PCR or methyl-PCR after deamination (FIG. 2, step 6).

This 5' flap only procedure was applied to the four pGEM-3Z/601b-N plasmids described in FIG. 11. In this experiment, four different flap structures were formed on the HindIII-linearized and denatured plasmids using flap adapters containing oligo U1 (not U1-N). Therefore, after cleavage of the target strand 5' end between nt 2,453 and nt 2,454 (FIG. 12A), the 1-nt gap precludes ligation of the oligo U1 as shown in FIG. 2, step 1b. Therefore, the HindIII adapter, Ampligase, and ATP were added in order to ligate the 3' end-protected oligo U2 to the target strand 3' end (FIG. 2, step 2). After completion of the alternative FENG C protocol, agarose gel electrophoresis verified production of the expected 619-bp amplification product only in reactions containing Taq (FIG. 12B), as was observed when substrates with a 1-nt 3' flap were employed in FIG. 11B. In particular, in this no 3' flap FENG C protocol, the reactions with no Taq also displayed a range of high-molecular-weight products (FIG. 12B) similar to those observed in the Taq and FEN1 reactions in FIG. 11B.

The sensitivity of FENG C was tested using substrates with only a 5' flap (FIG. 13). In these experiments, 2 µg of human gDNA was mixed with a 10-fold dilution series (0.002 ng to 200 ng) of HindIII-linearized plasmid pGEM-3Z/601b-T, representing human genome equivalent copy numbers of 1 to 100,000. Oligo U1 and pGEM-3Z/601b NdeI-2 oligo 1-T comprising the flap adapter were added to form a 5' flap with no 3' flap with the target sequence. Thermostable FEN1 cleaves the target DNA strand of such structures at multiple sites, according to the manufacturer (New England Biolabs). Consistent with this, after executing the alternative FENG C protocol, an amplification product was not observed until 20 ng of spike-in plasmid, equivalent to 10,000 copies (FIG. 13A, lanes 1 and 2 compared with lanes 3 and 4).

Interestingly, FENG C with FEN1 and only a 5' flap improved in Taq buffer in that an abundant PCR product was visible with 100x less spike-in (0.2 ng pGEM-3Z/601-T; 100 copy equivalents) (FIG. 13B, lane 4). By contrast, a trace PCR product of the correct size was obtained in the reaction containing Taq in its supplied buffer and only 0.002 ng plasmid, a molar equivalent of 1 copy (FIG. 13B, lane 3). Furthermore, FENG C in Taq buffer in reactions containing 2 µg human gDNA plus 0.2 ng pGEM-3Z/601-N spike-in produced the 619-bp amplification product only when FEN1 and the pGEM-3Z/601-T flap oligo were supplied (FIG. 14; lane 6 compared with lanes 1-5, 7, and 8). This demonstrates high specificity for 5' flap cleavage, a T at the 3' end of the oligo U1 in order to fill the gap and base pair with the complementary A, and ligation of the U1-T oligo to the 5' end of the FEN1-cut target sequence. Taken together, FENG C of 5' flaps without a 1-nt 3' flap achieved the highest sensitivity in reactions using oligo 1-T and Taq in its supplied buffer.

Next, enrichment of human sequences using the two FENG C procedures, one with and one without a 1-nt 3' flap, were compared in parallel. To do so, gDNA from the human colon carcinoma cell line HCT116, which has a mostly euploid genome (Mouradov et al., 2014), was used as input. Sequences were enriched from approximately -200 to +100 of the transcription start sites (TSSs) of 10 human DNA mismatch repair genes and 1 human control gene with open chromatin (Table 2). Captured sequences of the expected lengths of ~300 nt, ligated to the 47 nt of ligated universal oligos, were amplified by standard PCR dependent on Taq in both FENG C reactions, demonstrating the efficacy of both protocols (FIG. 15; lanes 2 and 4 *versus* lanes 1 and 3). Moreover, a side-by-side comparison of preferred FENG C with Taq *versus* FEN1 on double flap structures showed superior enrichment yield for the same 11 human sequences with FEN1

(FIG. 16A), and FEN1 performed well in the manufacturer-supplied buffers for both enzymes (FIG. 16B). Therefore, the FENGCC strategy employing the double flap and FEN1 in its manufacturer supplied buffer was selected for all subsequent experiments.

EXAMPLE 9: Compatibility of FENGCC with bisulfite sequencing to detect DNA methylation

To determine if 5mC could be detected by FENGCC, sequences from the same 10 human DNA mismatch repair genes plus 1 human control gene were captured from 2 µg HCT116 DNA, treated with and without sodium bisulfite, and PCR amplified. Two sets of PCR amplicons were examined, with lengths ~350 bp and ~500 bp, anchored at the same 5' end (Table 2) (FIG. 17). As expected, the FENGCC enrichment yield with and without bisulfite treatment was higher for the set of shorter amplicons than the longer set (FIG. 17, compare lane 2 with 5 and lane 3 with 6), reflecting the well-known PCR bias toward shorter amplicons. This result also underscores the advantage of matching sequence lengths as afforded by FENGCC. In addition, bisulfite deamination of both sets of captured sequences dramatically decreased amplification yield (FIG. 17, compare lane 2 with 3 and lane 5 with 6); an amplified ~500-bp product was only observed when the input gDNA was doubled to 4 µg (FIG. 17, compare lane 6 with 7). In addition, variable amplification yields were observed with different bisulfite conversion kits (data not shown). The lower PCR product yield is consistent with the degradation of as much as 99.9% of input DNA during the deamination reaction (Tanaka and Okamoto, 2007). Therefore, the quantities of captured DNA available for PCR are extremely low after bisulfite conversion, e.g., only ~1 pg within 1 µg human gDNA for 10 targets of ~300 nt.

EXAMPLE 10: MAPit-FENGCC, a versatile assay for targeted epigenetic analysis

In certain embodiments, FENGCC is combined with single-molecule MAPit methylation footprinting. In the resulting MAPit-FENGCC assay, cells are permeabilized to allow the GpC DNA methyltransferase M.CviPI (New England Biolabs) or other suitable DNA methyltransferases to enter and diffuse into nuclei to methylate accessible GpC sites in the case of M.CviPI (Xu et al., 1998) or C in other contexts in chromatin (Nabils et al., 2014; Jessen et al., 2006; Gal-Yam et al., 2006; Lin et al., 2007; Pardo et al., 2011; Kelly et al., 2012). In some aspects, nuclei may first be isolated and treated with a DNA methyltransferase. After stopping the methylation reaction, gDNA was purified and subjected to FENGCC followed by bisulfite

conversion as in FIG. 17. Because Gp5mC (hereafter, G5mC) can be discerned from endogenous 5mCpG (hereafter, 5mCG), MAPit-FENGc simultaneously detects chromatin accessibility and DNA methylation. Moreover, the assay is freed from the constraints of target selection imposed by restriction endonucleases, a limitation of previously described MAPit-patch (Nabils et al., 2014).

To test MAPit-FENGc, the same 11 FENGc amplification products described above were obtained from gDNA isolated from M.CviPI-treated human glioblastoma (GBM) cell line L0{PMID 24105770}. The *MLH1* promoter was included to serve as a positive control with a known chromatin structure (Lin et al., 2007). FENGc with Taq or FEN1 was used to capture targets with lengths of ~300 nt and ~450 nt (TABLE 1, Sheets 2 and 3 and TABLE 2, Sheets 1 and 2), which were subjected to amplification by standard PCR or BS-PCR. Specific amplification products of the expected sizes inclusive of the 47 bp of the universal oligo sequences were observed in two independent biological replicates (FIG. 18, ~350 bp, lanes 1-3 and 7-9; ~500 bp, lanes 4-6 and 10-12).

Long-read, high-fidelity sequencing is the most informative in epigenetic assays in that it provides single-molecule data, i.e., avoids population averaging, and preserves phasing, the relationship between multiple features along each sequencing read. Therefore, the amplification products were subjected to long-read, circular consensus sequencing (CCS) on a Pacific Biosciences (PacBio) Sequel instrument (Eid et al., 2009). PacBio currently provides the most accurate sequencing platform because single molecules with lengths of up to 10 kb are sequenced at least five times or passes. The PacBio Sequel II instrument has a capacity of 8 M single molecules.

After aligning the PacBio Sequel reads to their reference sequences as previously described (Nabils et al., 2014), all FENGc amplification reactions detected at least 9 of 11 targets (82%) with at least 1 read (Table 3). Few reads were detected for *MSH3* and *MSH6*, most likely due to their high GC content, a well-established NGS phenomenon. For the remaining 9 targets, both Taq and FEN1 captured more than 20 reads with standard PCR when combining two biological replicates, demonstrating that both enzymes can be used in FENGc. For the BS-PCR samples, there were 6 targets with more than 20 reads when combining two biological replicates for both the ~350-nt and ~500-nt captured sequences. These 6 targets were used to

calculate the fraction of G5mCH and H5mCG (GCG were excluded due to overlap of GpC and CpG) in reads with ≥ 5 sequencing passes and $\geq 95\%$ conversion at non-GCH and non-HCG sites, i.e., HCH. The correlation coefficients for methylation of all HCG and GCH sites between the two biological replicates were >0.92 for all conditions, indicating high reproducibility of MAPit-FENGc results (FIG. 19). Interestingly, the sequenced FENGc products of *MLH1* showed higher correlations between biological replicates than bisulfite sequencing of a single PCR amplicon, i.e., not obtained by FENGc (FIG. 20). This is perhaps related to improved amplification with universal primers as opposed to gene-specific primers (Varley and Mitra, 2008). Importantly, the correlations of HCG and GCH methylation levels between the FENGc products of different lengths were also very high (FIG. 21). FENGc and direct BS-PCR of the *MLH1* promoter also showed excellent agreement (FIG. 22). In summary, FENGc efficiently enriches specific target sequences from gDNA obtained from chromatin samples probed with exogenous DNA methyltransferases, such as M.CviPI, that are subsequently treated with bisulfite and PCR amplified.

EXAMPLE 11: MAPit (Methyltransferase Accessibility Protocol for individual templates) methylation footprinting

MAPit was performed on cells permeabilized with digitonin to mark accessible GpC sites in chromatin. In detail, two million cells were first washed with cold phosphate buffered saline containing 0.015% (w/v) sodium azide (Nabils et al., 2014). Cells were pelleted and washed with 500 μ L ice-cold cell resuspension buffer (20 mM HEPES, pH 7.5, 70 mM MgCl₂, 0.25 mM EDTA, pH 8.0, 0.5 mM EGTA, pH 8.0, 0.5% (v/v) glycerol, freshly supplemented with 10 mM DTT and 0.25 mM PMSF. Cells were next pelleted and resuspend in 180 μ L cell resuspension buffer with 0.05% (w/v) digitonin and incubated on ice for 10 min. A 1 μ L aliquot of cells was stained with trypan blue to verify 100% permeabilization before proceeding. The cell suspension was then divided in half, treated with and without M.CviPI (100 U/million cells), and supplemented with fresh 160 μ M SAM, followed by incubation for 15 min at 37°C. Methylation reactions were stopped by addition of an equal volume of stop buffer (1% (w/v) sodium dodecyl sulfate, 100 mM NaCl, 10 mM ethylenediaminetetraacetic acid (EDTA)) and vortexed briefly at medium speed. RNase A was added to 10 μ g/mL for 1 hr at 37°C followed by 100 μ g/mL proteinase K treatment overnight at 50°C. GDNA was extracted using phenol-chloroform-

isoamyl alcohol (25:24:1 (v/v)) phase separation, followed by ethanol precipitation, and resuspension in ddH₂O.

EXAMPLE 12: Enhanced FENGCC efficacy with enzymatic-based detection of 5mC

Given the extensive DNA degradation and hence decreased sensitivity of all bisulfite-based methods for detection of DNA methylation, the extent to which a nondestructive method of C to U conversion would improve the sensitivity of MAPit-FENGCC was tested. This method, Enzymatic Methyl-seq™ (EM-seq; New England Biolabs), uses an α -ketoglutarate-dependent ten-eleven translocation 2 (TET2) enzyme to oxidize 5mC to 5-hydroxymethyl-C (neb.com/products/e7120-nebnext-enzymatic-methyl-seq-kit#Citations%20&%20Technical%20Literature)(Sun et al., 2021; Zhang et al., 2013; Yu et al., 2012), which is coupled to glucosylation by T4 phage β -glucosyltransferase (Josse and Kornberg, 1962; Tomaschewski et al., 1985; Schutsky et al., 2017). The resulting glucosyl-5-hydroxymethylcytosine modification protects against subsequent C to U enzymatic deamination by APOBEC (Schutsky et al., 2018; Schutsky et al., 2017). Also, the number of distinct target sequences captured and amplified from the same L0 gDNA used to generate FIG. 18 was increased to 119 targets spanning the transcription start sites (TSSs) of 74 genes with the Gene Ontology term “metabolic process” and filtered for “DNA repair,” 42 genes associated with cancer, and 3 control genes (Table 1, Sheet 3 and TABLE 2, Sheet 2). These targets ranged from 430 nt to 452 nt in size. EM-seq-converted sequences were robustly enriched by FENGCC, using 500 ng gDNA (digested with SpeI to decrease the size of 5' flaps) and post-exonuclease purification with either AMPure XP beads or NEBNext beads (FIG. 23). Purification with the MinElute PCR Purification Kit (Qiagen) was not successful for methyl-PCR. The optimal amount of oligos comprising the flap adapters to include for FENGCC followed by EM-seq is 1 μ l each of a 10 μ M stock solution of U1-T and U2 (FIG. 24A). In addition, the summed total of flap oligos 1-T, flap oligos 2-T, and oligos 3-T to include should approach but be less than the amount of oligo U1-T or oligo U2 (FIG. 24B and Table 12, calculated with formula). Sonication of gDNA to a mean length of 1 kb to decrease 5' flap size also enriched target sequences as well as SpeI digestion, and is preferred as it avoids cutting SpeI-containing target sequences (FIG. 25A). In addition, in contrast to bisulfite conversion, EM-seq conversion produced a detectable

amplification product of the expected size with as low as 50 ng of sonicated DNA input (FIG. 25B).

EXAMPLE 13: MAPit-FENGc with enzymatic-based detection of 5mC yields a high proportion of on-target sequence reads

MAPit-FENGc analysis was conducted in duplicate on two independent cultures of NSC, GBM Nx18-25, and GBM L0. The 119 gene targets were captured from sonicated gDNA purified from each cell line itself as well as a mixture of 0.1% L0:99.9% NSC, subjected to EM-seq conversion, PCR amplified, and purified with AMPure XP beads. The amplified, purified products were of high quality and uniformity in length as gauged by the Agilent TapeStation D5000 system (FIG. 26). In the distribution of high-fidelity PacBio CCS reads, the highest proportion were 480-500 nt in length, consistent with the amplification target size (FIG. 27 and FIG. 28). The small peak at approximately 960-1,000 bp is consistent with self-ligation of amplified products during the PacBio barcoding process. CCS reads were mapped to both the whole human genome (hg38) and target reference sequences (FIG. 29; Table 4).

For the two independent biological replicates of L0, for example, percentages of unmapped, off-target, and on-target reads obtained from each of the sequenced libraries were 8-11%, 7-8%, and 82-86%, respectively. There were 105 targets (88%) detected, with at least one read in the combined data from the two biological replicates (Table 4). Reads with ≥ 5 sequencing passes, $>95\%$ HCH conversion to HTH, and aligning to $>95\%$ of the length of each reference sequence were analyzed to avoid duplicated alignment to gene homologs with high sequence similarity. After this filtering, all L0 reads were uniquely aligned and 92 targets (77%) were represented (Table 5). As reported previously for bisulfite patch PCR and MAPit-patch (Varley and Mitra, 2010; Nabils et al., 2014), filtered read number was negatively correlated with GC content for all of the sequenced libraries (FIG. 30 and FIG. 31).

EXAMPLE 14: MAPit-FENGc efficiently detects and localizes 5mCG, nucleosomes, and DNA-bound transcription factors

For the fractions of H5mCG (DNA methylation) and G5mCH (chromatin accessibility), the correlation coefficients between the two independent biological replicates of each sequenced sample were very high, above 0.91 (FIG. 32). The results from FENGc targets processed by

BS-PCR *versus* EM-seq conversion PCR (EM-PCR) were compared for six targets with a minimum of 36 sequencing reads in both data sets (FIG. 33). The fractions of methylation of all HCG and GCH sites on the six amplicons from these two data sets were highly consistent, indicating that EM-seq can be substituted for BS-seq in MAPit-FENGCC. Three representative
 5 examples of MAPit-FENGCC sequence reads using EM-seq as plotted with methylscaper {PMID 34125875} are shown in FIG. 34, FIG. 35, and FIG. 36. In these images, each row of pixels represents the pattern of HCG methylation (left; red) and GCH accessibility (right; yellow) on one chromatin copy or molecule (read) in the original GBM L0 cells. All molecules are presented in both panels in the same top-to-bottom order.

10 The *POLD4* promoter from L0 harbors two positioned nucleosomes (designated -1 and +1) flanking a prominent NFR of variable length at the TSS, i.e., present on a large proportion of epialleles (FIG. 34, right panel). Variable-length spans of H5mCG also reside in the linker DNA between the -1 and +1 nucleosomes (FIG. 34, left panel). Open chromatin and the absence of DNA methylation at the TSS are features that correlate well with active transcription.

15 In the second example, also from L0, all sequenced copies of the region encompassing the *ALKBH2* TSS are essentially unmethylated, that is, the level of 5mCG is at the detection limit, except for modest 5mCG at the TSS (FIG. 35, left panel). In addition, most promoter copies house a large region of accessibility to M.CviPI of >147 bp (147 bp DNA wraps around and contacts the histone octamer in nucleosomes), indicative of a prominent NFR (FIG. 35, right
 20 panel). Large, variable-length regions of inaccessibility or footprints at the most upstream end of the *ALKBH2* promoter likely correspond to portions of differentially positioned nucleosomes (halved blue ellipse representing the population average). By contrast, a relatively small and robust footprint occupies the NFR just upstream of the TSS. Due to its short length of ~22 bp and uniform position, this footprint most likely corresponds to occupancy by a sequence-specific,
 25 non-histone regulatory factor, perhaps a transcriptional activator that orchestrates nucleosome eviction from the TSS.

The third promoter, *EPM2AIP1*, exemplifies a locus that is differentially methylated in two different cell types (FIG. 36). In cultured human neural stem cells (NSC), the region around the TSS of *EPM2AIP1* is highly accessible, except for a variably positioned +1 nucleosome and
 30 a likely DNA-bound, sequence-specific transcriptional activator just upstream of the TSS

(FIG. 36A, right panel) and essentially unmethylated (FIG. 36A, left panel). By contrast, in cultured GBM L0 cells, the region around the TSS of *EPM2AIP1* exhibited high, aberrant levels of H5mCG (FIG. 36B, left panel) and largely inaccessible chromatin, except for accessible, relatively short nucleosomal linkers (FIG. 36B, right panel), consistent with transcriptional silencing (data not shown). The above three examples illustrate that MAPit-FENGCC employing methylation detection by EM-seq is a powerful strategy for detecting multiple epigenetic features at high resolution on single molecules.

EXAMPLE 15: MAPit-FENGCC has high sensitivity, identifying 1 in 1,000 hypermethylated epialleles

To examine the sensitivity of MAPit-FENGCC in the detection of gDNA derived from abnormal cells, 0.1% L0 gDNA was spiked into that of human NSC, i.e., 0.1% L0:99.9% NSC. The 119 products were enriched as described above in FIG. 26, lanes 7 and 8. The read lengths distribution (FIG. 28), percentages of mapped CCS reads (FIG. 29), and correlations of the fractions H5mCG and G5mCH between the two independent biological replicates (FIG. 32) were also high as observed with MAPit-FENGCC of L0 described above in FIG. 27, FIG. 29, and FIG. 32. With more than three times more high-fidelity CCS reads compared with L0, this 0.1% L0:99.9% NSC sample detected 109 targets (92%) with ≥ 1 read (Table 4). After more stringent filtering of CCS reads for ≥ 5 sequencing passes, $\geq 95\%$ HCH conversion, and $\geq 95\%$ coverage of the length of each target reference sequence, 103 targets (87%) were detected (Tables 4 and 5). MAPit-FENGCC successfully detected the dense H5mCG in the 450 bp area encompassing the *EPM2AIP1* TSS derived from the 0.1% L0 gDNA spike-in (10 of 1,781 molecules), with high statistical significance ($P < 0.0001$) (FIG. 37). By contrast, the remaining molecules in the mixture showed a pattern of GCH accessibility similar to that of NSC and therefore were likely derived from that sample. These data demonstrate that FENGCC is highly sensitive, quantifiable, and reproducible for the detection of epigenetic features.

EXAMPLE 16: Efficient detection of differential epigenetic alterations by MAPit-FENGCC

The ability of MAPit-FENGCC to detect differential epigenetic signatures was tested further by assaying independent duplicate cultures of non-cancerous NSC and a different GBM primary cell line, Nx18-25, using the same 119 target-gene panel. Libraries for the 119 captured products were generated as described above (FIG. 26, lanes 1-4) and yielded the expected length

distributions of high-fidelity PacBio CCS read lengths (FIG. 27 and FIG. 28). The percentages of mapped CCS reads were high (Tables 4 and 5). Among these samples, Nx18-25 had similar read numbers compared with the 0.1% L0 spike-in sample, and also showed the same number of detected targets (109 or 92%). In addition, an identical number of targets (103 or 87%) was represented among high-quality reads having ≥ 5 sequencing passes, additionally filtered for $\geq 95\%$ HCH conversion, and covering $\geq 95\%$ of each reference sequence (Tables 4 and 5). While 2.3x as many reads were obtained from NSC as compared to Nx18-25, the numbers of targets detected (mean 110 or 93%) and targets covered by filtered reads (mean 105 or 87%) were similar (Tables 4 and 5).

Epigenetic differences between NSC and GBM Nx18-25 were assessed by calculating P values for 54 targets with ≥ 50 total CCS reads in both replicates, but no less than 22 reads in either of the two replicates (TABLE 6). Using criteria of $P < 0.05$ and $\geq 5\%$ differential in the level of either H5mCG or G5mCH between NSC and GBM Nx18-25, 57% (31/54) of the evaluated promoters showed epigenetic alterations (TABLE 6, Sheets 1 and 2). Applying more stringent cut-offs, the percentages of promoters with ≥ 0.1 , ≥ 0.2 , and ≥ 0.4 differentials in H5mCG or G5mCH were 26% (14/54), 13% (7/54), and 7.4% (4/54), respectively (TABLE 6, Sheets 3-5). The quantitative nature of these results is supported by the highly correlated ($r > 0.91$) H5mCG and G5mCH levels between individual replicates (FIG. 32). In addition, 733 of 735 of *MSH5* promoter molecules across all 4 samples had ≥ 10 methylated GCH sites (FIG. 38). This rules out incomplete cell permeabilization and chromatin probing as trivial reasons for the differential accessibilities observed between loci in the same or different samples in TABLE 6 and as further discussed below.

Three representative genes from the NSC and Nx18-25 data sets were chosen to illustrate the reproducibility and power of MAPit-FENGCC to reveal differential epigenetic landscapes. These promoters include human *CD44*, *CCN4*, and *HIST1HB1* for which the data were rendered as violin plots for the two independent replicates (Rep 1 and 2) of NSC and GBM Nx18-25 (FIG. 39). Plotted is the proportion of H5mCG or G5mCH for each molecule (black dots), the median (horizontal line), interquartile range of methylation levels (box), and the smoothed probability density at different methylation levels (gray area). These plots, similar to the H5mCG

and G5mCH correlation plots (FIG. 32), further illustrate the quantitative nature and reproducibility of MAPit-FENGc results.

CD44 encodes a transmembrane glycoprotein with functions in cell adhesion, proliferation, and apoptosis (Naor et al., 1997; Naor, 2016). It has also been reported to be a marker for astrocyte-restricted precursor cells (Liu et al., 2004). High expression of *CD44* in GBM tissue has been particularly linked to the mesenchymal subtype (Phillips et al., 2006; Verhaak et al., 2010) and GBM cancer stem cells (Anido et al., 2010; Fu et al., 2013). MAPit-FENGc of non-cancerous NSC showed undetectable H5mCG in the vicinity of the *CD44* TSS, with limited H5mCG accumulating farther upstream (FIG. 40A, first panel and FIG. 40B, upper panel). The majority of promoter copies from NSC harbored an NFR flanked by nucleosomes, which appeared to move in register to occupy multiple positions (FIG. 40A, second panel and FIG. 40B, lower panel). In striking contrast, in GBM Nx18-25, H5mCG had apparently spread to different extents across most promoter epialleles, leading to relatively promoter hypermethylation that correlated with a dramatic reduction in accessibility (FIG. 40A, third and fourth panels and FIG. 40B, lower panel), which manifests as shortened NFRs compared with NSC (FIG. 40C). The revealed epigenetic signatures of *CD44* are consistent with the observed strong transcriptional silencing in Nx18-25 compared with NSC (FIG. 40D).

CCN4, also known as WNT1-Inducible Signaling Pathway Protein 1 (WISP1), has been shown to contribute to the tumorigenesis and progression of a wide array of human cancers (Gaudreau et al., 2019; Liu et al., 2019; Deng et al., 2019). More importantly, *CCN4* gene expression has been reported to be upregulated in GBM compared with normal tissues (Jing et al., 2017). Indeed, the transcript level of *CCN4* was markedly enhanced in GBM Nx18-25 as opposed to the undetectable level in NSC (FIG. 41A). MAPit-FENGc of the *CCN4* promoter from NSC showed about half of the cells contained a broken span of H5mCG immediately downstream of the TSS (FIG. 41B, first panel and FIG. 41C, upper panel). Reads derived from these and many other cells contained random, small spans of accessibility, demonstrating occupancy by randomly positioned nucleosomes (FIG. 41A, second panel). In addition, a cluster of reads at the top had a visible -1 nucleosome; the bracketed subset of these reads contained footprint evidence of a DNA-bound, sequence-specific factor (black rectangle). Consistent with the dramatic increase in *CCN4* transcription in Nx18-25 compared with NSC seen in FIG. 41A,

the number of *CCN4* promoter molecules with H5mCG was depleted and there were increases in accessibility and NFR length upstream of the TSS (FIG. 41B, compare the third with first panel as well as fourth with second panel; FIG. 41C, and FIG. 41D).

The gene *HIST1H1B* encodes the linker histone protein H1.5, which is involved in maintaining higher-order chromatin structure as well as regulating DNA repair and cell proliferation (Albig et al., 1997; Sancho et al., 2008; Happel and Doenecke, 2009). By MAPit-FENGc, the *HIST1H1B* promoter from NSC was largely devoid of H5mCG, except for 0.6% that were hypermethylated and relatively inaccessible across the whole analyzed region (FIG. 42A, first panel, cluster 6 and FIG. 42B, upper panel). In Nx18-25, the fraction of epialleles populating cluster 6 increased to ~4% (FIG. 42A, compare panel 3 with 1). Interestingly, in both cell types, the -1 nucleosome was uniformly positioned, whereas the +1 nucleosome was slid downstream to various extents (FIG. 42A, second and fourth panels). On close examination, overall, accessibility decreased at the *HIST1H1B* TSS (FIG. 42A, compare second and fourth panels, sum of clusters 1 and 2; FIG. 42B, bottom panel, bracket). Despite these observed alterations in promoter accessibility and apparent epigenetic silencing of 4% of epialleles in Nx18-25, there was no significant difference in transcript abundance between NSC and Nx18-25 (FIG. 42D). Therefore, MAPit-FENGc is able to identify frequent as well as rare epigenetic differences between biological samples, and the profiled epigenetic heterogeneity provides key insights into gene regulatory mechanisms.

EXAMPLE 17: Detection of genetic alterations by FENGc

For genotyping, target DNA sequences captured and enriched by FENGc are sequenced directly, without deamination. To demonstrate feasibility, the same 119 target promoter sequences were captured from NSC and Nx18-25 gDNA by FENGc, amplified with standard PCR, barcoded, and then directly sequenced on a PacBio Sequel instrument. Ninety-seven (82%) targets were detected with at least 1 read in at least one condition (Table 4, Sheet 2). Among these 97 regions, 54 single-nucleotide polymorphisms (SNPs) and 18 indels were identified, in which 9 SNPs and 2 indels were GBM-specific (Table 7). Three GBM-specific variants affected CG or GC sites. There were 43 SNPs and 16 indels already recorded in dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>). The C-to-A substitution in the 5' upstream flanking region of the *CDH1* gene at chr16:68737131 was identified as GBM cell-specific, and indeed has been

labeled as a risk-factor with clinical significance (rs16260). The SNP A allele was identified within 0% of reads in NSC and 37% of reads in Nx18-25. Among the 11 SNPs and 2 indels not yet included in dbSNP, a T-to-A substitution in the 5' upstream flanking region of the *DDB2* gene at chr11:47214657 was also GBM-specific. The SNP A allele was observed in 2.7% of reads in NSC and 21% of reads in Nx18-25. The ability to identify both alleles of known polymorphisms and novel variants demonstrates the high genotyping sensitivity of FENG C that is reproducible across samples as well as its potential for application in clinical diagnosis of genetic disorders. In addition, no SNPs or indels were identified in the three promoters chosen above to exemplify the identification of differential DNA methylation and chromatin accessibility between NSC and Nx18-25 GBM by MAPit-FENG C (FIG. 40, FIG. 41, and FIG. 42). This demonstrates *bona fide* epigenetic alterations between these cell lines, which, by definition, require the absence of mutations.

EXAMPLE 18: MAPit-FENG C captures and detects epigenetic features on ~940-nt products and delineates long-range regulatory relationships.

Long-read sequencing allows examination of epigenetic landscapes at a distance, i.e., relationships between individual regulatory modules such as multiple positioned nucleosomes and *cis*-acting sequences bound by transcription factors. MAPit-FENG C was therefore applied to the primary GBM cell line, Nx18-25, for 45 targets with lengths of ~940 nt (Table 1, Sheet 4; Table 2, Sheet 3). Two gDNA input amounts, 800 ng and 400 ng, were tested from two biological replicates. The distribution of obtained CCS read lengths showed that most of the captured and amplified products were mostly of the expected ~990 nt (~940-nt targets plus 47-nt PCR primers; FIG. 43). Total percentages of on-target ~940 bp reads for the 800 ng and 400 ng input samples were similar to the above ~450 bp products in FIG. 29, approximately 81% and 77%, respectively (FIG. 44; Table 8). Less than 8% of reads did not map to the human genome and, among the reads aligned to human genome, the off-target percentage was ~12% for 800 ng gDNA input and 18% for 400 ng gDNA input. There were 38 (84%) targets detected by ≥ 1 read for both of the replicates for each input DNA mass (Table 8). The coverage, however, of each target was variable (Table 9). Nevertheless, H5mCG and G5mCH levels were highly correlated between the two biological replicates for both the same (FIG. 45) and different input gDNA

amounts (FIG. 46), as well as between the overlapped regions between the ~940-nt and ~450-nt targets (FIG. 47).

MAPit-FENGc reads of 937 bp containing the divergent TSSs from both the *EPM2AIP1* and *MLH1* promoters were compared with the reads from two overlapping, shorter products of 438 bp and 450 bp obtained from Nx18-25 gDNA (FIG. 48). A very low level of H5mCG was detected along the entire promoter in these cells (FIG. 48A and FIG. 48C, left panel). The 450 bp of overlap harboring the *EPM2AIP1* TSS showed a variably positioned +1 nucleosome and therefore range of NFR lengths, and a short footprint (labeled 1) likely corresponding to occupancy by a sequence-specific transcriptional activator (FIG. 48B, left panel and FIG. 48C, right panel). Both of these features were seen in NSC as well (FIG. 36A). By contrast, in the overlapping *MLH1* sequences, the +1 nucleosome occupied a much more constrained range of positions (FIG. 48B, right panel), and occupancy of a second sequence-specific transcription factor was identified upstream of *MLH1* TSSb (labeled 2) (FIG. 48B and FIG. 48C, right panels). A third, robust transcription factor footprint (labeled 3) was detected upstream of *MLH1* TSSa on the 937-bp amplicon (FIG. 48C, right panel). Robust co-occupancy of all three TFs and the two +1 nucleosomes was evident on these long *EPM2AIP1-MLH1* molecules (FIG. 48C, right panel).

In addition to detecting additional molecular information, these long reads provided an opportunity to examine the extent to which multiple epigenetic features are coordinated or co-regulated. For example, the hierarchical organization of reads from both shorter amplicons clearly shows that the positions of both +1 nucleosomes range from farther to closer to each TSS (FIG. 48B). However, on the 937-bp amplicon, this ordered organization is no longer apparent for the *MLH1* TSSa +1 nucleosome when the reads are hierarchically clustered on the *EPM2AIP1* TSS +1 nucleosome (FIG. 48C, right panel) and *vice versa* (data not shown). This indicates that these two nucleosomes shift independently of each other. In sum, the longer amplicon netted an additional TF footprint and deduction of independent, dynamic mobilization of two nucleosomes, novel regulatory insights precluded by short reads.

MAPit-FENGc of other ~940-nt-targets from Nx18-25 cells availed additional organizational features compared to their shorter counterparts. The NFR of the divergent *NPAT-ATM* promoter showed a robust transcription factor footprint, a heterogeneously sized footprint (cyan rectangle) at the *NPAT* TSS, and a well-positioned +1 nucleosome followed by

progressively less well-positioned nucleosomes +2 and +3 (FIG. 49A). By contrast, the upstream *MSH2* promoter nucleosomes were much more disorganized (FIG. 49B), and the NFR of a sizeable number of molecules was punctuated with ~55-bp footprints at the TSS (cyan rectangle), possibly corresponding to paused RNA polymerase II. MAPit-FENGc of a longer *CCN4* promoter fragment from Nx18-25 (FIG. 49C) than assayed in FIG. 41B revealed NFR expansion to ~400 bp on a subset of molecules, but no upstream positioned nucleosomes were discernable. Furthermore, MAPit-FENGc of ~800 bp of 5' flanking sequence from *MSH2* (FIG. 49B) and *CCN4* (FIG. 49C) identified clear transitions between 5mCG depletion and hypermethylation.

EXAMPLE 19: MAPit-FENGc reproducibly informs epigenetic architectures within primary cells.

Having successfully applied FENGc to detect epigenetic and genetic variation in cultured NSC and GBM, the epigenetic arm of the protocol was tested on primary monocytes isolated from the bone marrows of four female mice. These cells were treated with M.CviPI and 600 ng gDNA from each sample was processed with primers targeting 78 promoters of cellular inflammatory response genes (Table 1, Sheet 5; Table 2, Sheet 4). For this experiment, the primers were designed by newly developed program, FENGc oligo designer (FOLD). To provide a computational solution for all 78 targets that avoided repetitive sequences and satisfied other rigorous command-line settings, target sizes were permitted to range from 474-987 nt (Table 2, Sheet 4).

High-quality CCS reads from the four libraries were mapped to the complete mm9 build of the mouse genome and the specific target reference sequences: 20-29% did not align to either reference or were removed by filtering ($\geq 95\%$ HCH conversion and $\geq 95\%$ alignment), yielding 80-71% on-target reads. FENGc detected 71-75 targets (91-96%) with ≥ 1 read in each sample (Tables 10 and 11).

A mixed effects ANOVA was used to determine the extent to which the levels of DNA methylation and chromatin accessibility determined by MAPit-FENGc across each target were statistically different in at least one bone marrow-derived monocyte sample (Table 11). This testing was based on the total percentage of H5mCG or G5mCH per molecule and limited to 43 amplicons with ≥ 100 CCS reads per sample and good diversity, i.e., absence of duplicates

apparent on visual inspection. CCS read numbers obtained from these targets showed negative correlations with target sequence length and GC content that ranged from 37-70% GC content and 474-760 nt, respectively (FIG. 50). Mean *P* values for H5mCG and G5mCH before correction for multiple testing were >0.90 (range 0.17-1.0) and all *P* values after correction with the Bonferroni method were 1.0. This indicates that for each target, none of the four mouse monocyte samples had significantly different levels of DNA methylation or chromatin accessibility (Table 11).

The reproducibility of chromatin architecture between independent mice is evident in single-molecule plots of H5mCG and G5mCH from eight representative loci that also illustrate interesting chromatin biology (FIG. 51 and FIG. 52). *Hsf1* exemplifies a promoter in bone marrow-derived monocytes that is exceptionally open, with 2,333 of 2,334 molecules showing relatively large NFRs (range 223-445 bp; FIG. 51A). Again, such a locus indicates that observed changes in accessibility between different loci within a sample or at a specific locus between different samples are not attributable to variable cell permeabilization or M.CviPI activity. Therefore, over the population of *Btk* promoter molecules from the X chromosome of the four female mice samples, a mean maximal accessibility of 46% (range 43-50%) occurred at the TSS within NFRs up to 195 bp long, localized almost exclusively to epialleles bearing 0% H5mCG (FIG. 51B). By contrast, consistent with X inactivation, long NFRs were highly depleted from epialleles with ≥ 2 H5mCG.

Among other intriguing loci, the *Pik3r3* promoter harbored a remarkably well-positioned -1 nucleosome and incremental, preferential sliding of the +1 nucleosome, expanding or contracting NFR length in individual cells mainly on the side of the NFR downstream of the TSS (FIG. 51C). By comparison, NFR contraction/expansion occurred on both sides of the NRF in the promoters of *Tlr4* (FIG. 51D), *Hsp90ab1* (FIG. 52A), *Irf7* (data not shown), and *Cxcr4* (FIG. 52B) due to movement of both the -1 and +1 nucleosomes. Surprisingly, the *Cxcr4* promoter displayed a 36-bp zone of H5mCG (orange rectangle) in the accessible NFR, only 12 bp from a strongly footprinted TF.

Compared to the above promoters, the *Mapk15* gene body (+1,660 to +2,215) was heavily methylated with arrays of randomly positioned nucleosomes with short, linker-length NFRs (FIG. 52C). Chromatin -571 to +45 of the *Src* TSS was similarly organized (FIG. 52D),

consistent with low-level expression of *Src* in mouse monocytes {Schaum, 2018 #190}.

Interestingly, a sequence in *Src* with strong CTCF binding site homology {Hashimoto, 2017 #188} conferred clear protection of 50 bp against endogenous 5mCG in many cells.

Single-amplicon MAPit was used as an independent method to evaluate the chromatin structures of the *Btk*, *Cxcr4*, *Hsp90ab1*, and *Tlr4* targets. Identical patterns of chromatin accessibility and DNA methylation were seen (data not shown), validating the MAPit-FENG C results. Taken together, the data demonstrate that MAPit-FENG C is effective at discerning epigenetic landscapes of purified primary cells, with striking inter-sample reproducibility.

In sum, FENG C permits facile, multiplexed, and cost-effective capture and enrichment of cohorts of user-defined sequences for either genotyping or detection of DNA methylation and chromatin accessibility in a single experiment. The high on-target coverage of long sequencing reads provides an unprecedented and exquisite level of molecular detail for applications in basic science and medicine.

REFERENCES

1. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. Target-enrichment strategies for next-generation sequencing. *Nat Methods*. 2010;7(2):111-8. doi: 10.1038/nmeth.1419. PubMed PMID: 20111037.
2. Myllykangas S, Ji HP. Targeted deep resequencing of the human cancer genome using next-generation technologies. *Biotechnol Genet Eng Rev*. 2010;27:135-58. PubMed PMID: 21415896; PMCID: PMC4340661.
3. Kozarewa I, Armisen J, Gardner AF, Slatko BE, Hendrickson CL. Overview of target enrichment strategies. *Curr Protoc Mol Biol*. 2015;112:7 21 1-3. doi: 10.1002/0471142727.mb0721s112. PubMed PMID: 26423591.
4. Chamberlain JS, Gibbs RA, Ranier JE, Nguyen PN, Caskey CT. Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Res*. 1988;16(23):11141-56. doi: 10.1093/nar/16.23.11141. PubMed PMID: 3205741; PMCID: PMC339001.
5. Hayden MJ, Nguyen TM, Waterman A, Chalmers KJ. Multiplex-ready PCR: a new method for multiplexed SSR and SNP genotyping. *BMC Genomics*. 2008;9:80. doi: 10.1186/1471-2164-9-80. PubMed PMID: 18282271; PMCID: PMC2275739.
6. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci USA*. 1992;89(5):1827-31. doi: 10.1073/pnas.89.5.1827. PubMed PMID: 1542678; PMCID: PMC48546.

7. Darst RP, Pardo CE, Ai L, Brown KD, Kladde MP. Bisulfite sequencing of DNA. *Curr Protoc Mol Biol.* 2010;Chapter 7:Unit 7 9 1-17. PubMed PMID: 20583099.
8. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 2005;33(18):5868-77. doi: 10.1093/nar/gki901. PubMed PMID: 16224102; PMCID: PMC1258174.
9. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc.* 2011;6(4):468-81. doi: 10.1038/nprot.2010.190. PubMed PMID: 21412275.
10. McClelland M. The effect of sequence specific DNA methylation on restriction endonuclease cleavage. *Nucleic Acids Res.* 1981;9(22):5859-66. doi: 10.1093/nar/9.22.5859. PubMed PMID: 6273810; PMCID: PMC327569.
11. Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttill J, Zhang L, Khrebtukova I, Milne TA, Huang Y, Biswas D, Hess JL, Allis CD, Roeder RG, Valk PJ, Lowenberg B, Delwel R, Fernandez HF, Paietta E, Tallman MS, Schroth GP, Mason CE, Melnick A, Figueroa ME. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet.* 2012;8(6):e1002781. doi: 10.1371/journal.pgen.1002781. PubMed PMID: 22737091; PMCID: PMC3380828 Illumina.
12. Garrett-Bakelman FE, Sheridan CK, Kacmarczyk TJ, Ishii J, Betel D, Alonso A, Mason CE, Figueroa ME, Melnick AM. Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. *J Vis Exp.* 2015(96):e52246. doi: 10.3791/52246. PubMed PMID: 25742437; PMCID: PMC4354670.
13. Martinez-Arguelles DB, Lee S, Papadopoulos V. *In silico* analysis identifies novel restriction enzyme combinations that expand reduced representation bisulfite sequencing CpG coverage. *BMC Res Notes.* 2014;7:534. doi: 10.1186/1756-0500-7-534. PubMed PMID: 25127888; PMCID: PMC4141122.
14. Sun Z, Vaisvila R, Hussong LM, Yan B, Baum C, Saleh L, Samaranayake M, Guan S, Dai N, Correa IR, Jr., Pradhan S, Davis TB, Evans TC, Jr., Eitwiller LM. Nondestructive enzymatic deamination enables single-molecule long-read amplicon sequencing for the determination of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Genome Res.* 2021. doi: 10.1101/gr.265306.120. PubMed PMID: 33468551; PMCID: PMC7849414.
15. Lyamichev V, Brow MA, Dahlberg JE. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science.* 1993;260(5109):778-83. doi: 10.1126/science.7683443. PubMed PMID: 7683443.
16. Chien A, Edgar DB, Trela JM. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J Bacteriol.* 1976;127(3):1550-7. doi: 10.1128/JB.127.3.1550-1557.1976. PubMed PMID: 8432; PMCID: PMC232952.
17. Varley KE, Mitra RD. Nested Patch PCR enables highly multiplexed mutation discovery in candidate genes. *Genome Res.* 2008;18(11):1844-50. doi: 10.1101/gr.078204.108. PubMed PMID: 18849522; PMCID: PMC2577855.

18. Varley KE, Mitra RD. Bisulfite Patch PCR enables multiplexed sequencing of promoter methylation across cancer samples. *Genome Res.* 2010;20(9):1279-87. doi: 10.1101/gr.101212.109. PubMed PMID: 20627893; PMCID: PMC2928506.
19. Finger LD, Attack JM, Tsutakawa S, Classen S, Tainer J, Grasby J, Shen B. The wonders of flap endonucleases: structure, function, mechanism and regulation. *Subcell Biochem.* 2012;62:301-26. doi: 10.1007/978-94-007-4572-8_16. PubMed PMID: 22918592; PMCID: PMC3728657.
20. Wu DY, Wallace RB. Specificity of the nick-closing activity of bacteriophage T4 DNA ligase. *Gene.* 1989;76(2):245-54. doi: 10.1016/0378-1119(89)90165-0. PubMed PMID: 2753355.
21. Kaiser MW, Lyamicheva N, Ma W, Miller C, Neri B, Fors L, Lyamichev VI. A comparison of eubacterial and archaeal structure-specific 5'-exonucleases. *J Biol Chem.* 1999;274(30):21387-94. doi: 10.1074/jbc.274.30.21387. PubMed PMID: 10409700.
22. Lyamichev V, Brow MA, Varvel VE, Dahlberg JE. Comparison of the 5' nuclease activities of Taq DNA polymerase and its isolated nuclease domain. *Proc Natl Acad Sci USA.* 1999;96(11):6143-8. doi: 10.1073/pnas.96.11.6143. PubMed PMID: 10339555; PMCID: PMC26849.
23. Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T, Kwiatkowski RW, Sander TJ, de Arruda M, Arco DA, Neri BP, Brow MA. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat Biotechnol.* 1999;17(3):292-6. doi: 10.1038/7044. PubMed PMID: 10096299.
24. Hall JG, Eis PS, Law SM, Reynaldo LP, Prudent JR, Marshall DJ, Allawi HT, Mast AL, Dahlberg JE, Kwiatkowski RW, de Arruda M, Neri BP, Lyamichev VI. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proc Natl Acad Sci USA.* 2000;97(15):8272-7. doi: 10.1073/pnas.140225597. PubMed PMID: 10890904; PMCID: PMC26937.
25. Lyamichev VI, Kaiser MW, Lyamicheva NE, Vologodskii AV, Hall JG, Ma WP, Allawi HT, Neri BP. Experimental and theoretical analysis of the invasive signal amplification reaction. *Biochemistry.* 2000;39(31):9523-32. doi: 10.1021/bi0007829. PubMed PMID: 10924149.
26. Tsutakawa SE, Thompson MJ, Arvai AS, Neil AJ, Shaw SJ, Algasaier SI, Kim JC, Finger LD, Jardine E, Gotham VJB, Sarker AH, Her MZ, Rashid F, Hamdan SM, Mirkin SM, Grasby JA, Tainer JA. Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. *Nat Commun.* 2017;8:15855. doi: 10.1038/ncomms15855. PubMed PMID: 28653660; PMCID: PMC5490271.
27. Nabilsi NH, Deleyrolle LP, Darst RP, Riva A, Reynolds BA, Kladde MP. Multiplex mapping of chromatin accessibility and DNA methylation within targeted single molecules identifies epigenetic heterogeneity in neural stem cells and glioblastoma. *Genome Res.* 2014;24(2):329-39. doi: 10.1101/gr.161737.113. PubMed PMID: 24105770; PMCID: PMC3912423.
28. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in

- mammalian DNA by MLL partner TET1. *Science*. 2009;324(5929):930-5. doi: 10.1126/science.1170116. PubMed PMID: 19372391; PMCID: PMC2715015.
29. Schutsky EK, DeNizio JE, Hu P, Liu MY, Nabel CS, Fabyanic EB, Hwang Y, Bushman FD, Wu H, Kohli RM. Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. *Nat Biotechnol*. 2018. doi: 10.1038/nbt.4204. PubMed PMID: 30295673; PMCID: PMC6453757.
30. Tanaka K, Okamoto A. Degradation of DNA by bisulfite treatment. *Bioorg Med Chem Lett*. 2007;17(7):1912-5. doi: 10.1016/j.bmcl.2007.01.040. PubMed PMID: 17276678.
31. Friedrich-Heineken E, Henneke G, Ferrari E, Hubscher U. The acetyltable lysines of human Fen1 are important for endo- and exonuclease activities. *J Mol Biol*. 2003;328(1):73-84. doi: 10.1016/s0022-2836(03)00270-5. PubMed PMID: 12683998.
32. Volden R, Palmer T, Byrne A, Cole C, Schmitz RJ, Green RE, Vollmers C. Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *Proc Natl Acad Sci USA*. 2018;115(39):9726-31. doi: 10.1073/pnas.1806447115. PubMed PMID: 30201725; PMCID: PMC6166824.
33. Shore D, Langowski J, Baldwin RL. DNA flexibility studied by covalent closure of short fragments into circles. *Proc Natl Acad Sci USA*. 1981;78(8):4833-7. doi: 10.1073/pnas.78.8.4833. PubMed PMID: 6272277; PMCID: PMC320266.
34. Shore D, Baldwin RL. Energetics of DNA twisting. I. Relation between twist and cyclization probability. *J Mol Biol*. 1983;170(4):957-81. doi: 10.1016/s0022-2836(83)80198-3. PubMed PMID: 6315955.
35. Shore D, Baldwin RL. Energetics of DNA twisting. II. Topoisomer analysis. *J Mol Biol*. 1983;170(4):983-1007. doi: 10.1016/s0022-2836(83)80199-5. PubMed PMID: 6644817.
36. Stenberg J, Dahl F, Landegren U, Nilsson M. PieceMaker: selection of DNA fragments for selector-guided multiplex amplification. *Nucleic Acids Res*. 2005;33(8):e72. doi: 10.1093/nar/gni071. PubMed PMID: 15860769; PMCID: PMC1087790.
37. Karst SM, Ziels RM, Kirkegaard RH, Sørensen EA, McDonald D, Zhu Q, Knight R, Albertsen M. Enabling high-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *bioRxiv*. 2020:645903. doi: 10.1101/645903.
38. Bennett-Baker PE, Mueller JL. CRISPR-mediated isolation of specific megabase segments of genomic DNA. *Nucleic Acids Res*. 2017;45(19):e165. doi: 10.1093/nar/gkx749. PubMed PMID: 28977642; PMCID: PMC5737698.
39. Gabrieli T, Sharim H, Fridman D, Arbib N, Michaeli Y, Ebenstein Y. Selective nanopore sequencing of human *BRCA1* by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Res*. 2018;46(14):e87. doi: 10.1093/nar/gky411. PubMed PMID: 29788371; PMCID: PMC6101500.
40. Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, Downs B, Sukumar S, Sedlazeck FJ, Timp W. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat Biotechnol*. 2020. doi: 10.1038/s41587-020-0407-5. PubMed PMID: 32042167.

41. Lowary PT, Widom J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol.* 1998;276(1):19-42. doi: 10.1006/jmbi.1997.1494. PubMed PMID: 9514715.
- 5 42. Tsutakawa SE, Classen S, Chapados BR, Arvai AS, Finger LD, Guenther G, Tomlinson CG, Thompson P, Sarker AH, Shen B, Cooper PK, Grasby JA, Tainer JA. Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell.* 2011;145(2):198-211. doi: 10.1016/j.cell.2011.03.004. PubMed PMID: 21496641; PMCID: PMC3086263.
- 10 43. Dechassa ML, Sabri A, Pondugula S, Kassabov SR, Chatterjee N, Kladde MP, Bartholomew B. SWI/SNF has intrinsic nucleosome disassembly activity that is dependent on adjacent nucleosomes. *Mol Cell.* 2010;38(4):590-602. doi: 10.1016/j.molcel.2010.02.040. PubMed PMID: 20513433; PMCID: PMC3161732.
- 15 44. Mouradov D, Sloggett C, Jorissen RN, Love CG, Li S, Burgess AW, Arango D, Strausberg RL, Buchanan D, Wormald S, O'Connor L, Wilding JL, Bicknell D, Tomlinson IP, Bodmer WF, Mariadason JM, Sieber OM. Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res.* 2014;74(12):3238-47. doi: 10.1158/0008-5472.CAN-14-0013. PubMed PMID: 24755471.
- 20 45. Xu M, Kladde MP, Van Etten JL, Simpson RT. Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC. *Nucleic Acids Res.* 1998;26(17):3961-6. Epub 1998/08/15. PubMed PMID: 9705505; PMCID: 147793.
46. Jessen WJ, Hoose SA, Kilgore JA, Kladde MP. Active *PHO5* chromatin encompasses variable numbers of nucleosomes at individual promoters. *Nat Struct Mol Biol.* 2006;13(3):256-63. doi: 10.1038/nsmb1062. PubMed PMID: 16491089.
- 25 47. Gal-Yam EN, Jeong S, Tanay A, Egger G, Lee AS, Jones PA. Constitutive nucleosome depletion and ordered factor assembly at the *GRP78* promoter revealed by single molecule footprinting. *PLoS Genet.* 2006;2(9):e160. doi: 10.1371/journal.pgen.0020160. PubMed PMID: 17002502; PMCID: PMC1574359.
- 30 48. Lin JC, Jeong S, Liang G, Takai D, Fatemi M, Tsai YC, Egger G, Gal-Yam EN, Jones PA. Role of nucleosomal occupancy in the epigenetic silencing of the *MLH1* CpG island. *Cancer Cell.* 2007;12(5):432-44. doi: 10.1016/j.ccr.2007.10.014. PubMed PMID: 17996647; PMCID: PMC4657456.
- 35 49. Pardo CE, Darst RP, Nabils NH, Delmas AL, Kladde MP. Simultaneous single-molecule mapping of protein-DNA interactions and DNA methylation by MAPit. *Curr Protoc Mol Biol.* 2011;Chapter 21:Unit 21 2. doi: 10.1002/0471142727.mb2122s95. PubMed PMID: 21732317; PMCID: PMC3214598.
50. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* 2012;22(12):2497-506. doi: 10.1101/gr.143008.112. PubMed PMID: 22960375; PMCID: PMC3514679.

51. Lay FD, Kelly TK, Jones PA. Nucleosome Occupancy and Methylome Sequencing (NOMe-seq). *Methods Mol Biol.* 2018;1708:267-84. doi: 10.1007/978-1-4939-7481-8_14. PubMed PMID: 29224149.
52. Deleyrolle LP, Reynolds BA. Isolation, expansion, and differentiation of adult Mammalian neural stem and progenitor cells using the neurosphere assay. *Methods Mol Biol.* 2009;549:91-101. doi: 10.1007/978-1-60327-931-4_7. PubMed PMID: 19378198.
53. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. Real-time DNA sequencing from single polymerase molecules. *Science.* 2009;323(5910):133-8. doi: 10.1126/science.1162986. PubMed PMID: 19023044.
54. Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* 2012;40(10):e72. doi: 10.1093/nar/gks001. PubMed PMID: 22323520; PMCID: PMC3378858.
55. Zhang L, Szulwach KE, Hon GC, Song CX, Park B, Yu M, Lu X, Dai Q, Wang X, Street CR, Tan H, Min JH, Ren B, Jin P, He C. Tet-mediated covalent labelling of 5-methylcytosine for its genome-wide detection and sequencing. *Nat Commun.* 2013;4:1517. doi: 10.1038/ncomms2527. PubMed PMID: 23443545; PMCID: PMC3679896.
56. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, Min JH, Jin P, Ren B, He C. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell.* 2012;149(6):1368-80. doi: 10.1016/j.cell.2012.04.027. PubMed PMID: 22608086; PMCID: PMC3589129.
57. Josse J, Kornberg A. Glucosylation of deoxyribonucleic acid. III. α - and β -Glucosyl transferases from T4-infected *Escherichia coli*. *J Biol Chem.* 1962;237:1968-76. PubMed PMID: 14452558.
58. Tomaschewski J, Gram H, Crabb JW, Ruger W. T4-induced α - and β -glucosyltransferase: cloning of the genes and a comparison of their products based on sequencing data. *Nucleic Acids Res.* 1985;13(21):7551-68. doi: 10.1093/nar/13.21.7551. PubMed PMID: 2999696; PMCID: PMC322070.
59. Schutsky EK, Nabel CS, Davis AKF, DeNizio JE, Kohli RM. APOBEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA. *Nucleic Acids Res.* 2017;45(13):7655-65. doi: 10.1093/nar/gkx345. PubMed PMID: 28472485; PMCID: PMC5570014.
60. Naor D, Sionov RV, Ish-Shalom D. CD44: structure, function, and association with the malignant process. *Adv Cancer Res.* 1997;71:241-319. doi: 10.1016/s0065-230x(08)60101-3. PubMed PMID: 9111868.

61. Naor D. Editorial: interaction between hyaluronic acid and its receptors (CD44, RHAMM) regulates the activity of inflammation and cancer. *Front Immunol.* 2016;7:39. doi: 10.3389/fimmu.2016.00039. PubMed PMID: 26904028; PMCID: PMC4745048.
- 5 62. Liu Y, Han SS, Wu Y, Tuohy TM, Xue H, Cai J, Back SA, Sherman LS, Fischer I, Rao MS. CD44 expression identifies astrocyte-restricted precursor cells. *Dev Biol.* 2004;276(1):31-46. doi: 10.1016/j.ydbio.2004.08.018. PubMed PMID: 15531362.
- 10 63. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, Misra A, Nigro JM, Colman H, Soroceanu L, Williams PM, Modrusan Z, Feuerstein BG, Aldape K. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell.* 2006;9(3):157-73. doi: 10.1016/j.ccr.2006.02.019. PubMed PMID: 16530701.
- 15 64. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, Spellman PT, Wilson RK, Speed TP, Gray JW, Meyerson M, Getz G, Perou CM, Hayes DN, Cancer Genome Atlas Research N. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell.* 2010;17(1):98-110. doi: 10.1016/j.ccr.2009.12.020. PubMed PMID: 20129251; PMCID: PMC2818769.
- 20 65. Anido J, Saez-Borderias A, Gonzalez-Junca A, Rodon L, Folch G, Carmona MA, Prieto-Sanchez RM, Barba I, Martinez-Saez E, Prudkin L, Cuartas I, Raventos C, Martinez-Ricarte F, Poca MA, Garcia-Dorado D, Lahn MM, Yingling JM, Rodon J, Sahuquillo J, Baselga J, Seoane J. TGF- β receptor inhibitors target the CD44^{high}/Id1^{high} glioma-initiating cell population in human glioblastoma. *Cancer Cell.* 2010;18(6):655-68. doi: 10.1016/j.ccr.2010.10.023. PubMed PMID: 21156287.
- 25 66. Fu J, Yang QY, Sai K, Chen FR, Pang JC, Ng HK, Kwan AL, Chen ZP. TGM2 inhibition attenuates ID1 expression in CD44-high glioma-initiating cells. *Neuro Oncol.* 2013;15(10):1353-65. doi: 10.1093/neuonc/not079. PubMed PMID: 23877317; PMCID: PMC3779037.
- 30 67. Gaudreau PO, Clairefond S, Class CA, Boulay PL, Chrobak P, Allard B, Azzi F, Pommey S, Do KA, Saad F, Trudel D, Young M, Stagg J. WISP1 is associated to advanced disease, EMT and an inflamed tumor microenvironment in multiple solid tumors. *Oncoimmunology.* 2019;8(5):e1581545. doi: 10.1080/2162402X.2019.1581545. PubMed PMID: 31069142; PMCID: PMC6492985.
- 35 68. Liu Y, Song Y, Ye M, Hu X, Wang ZP, Zhu X. The emerging role of WISP proteins in tumorigenesis and cancer therapy. *J Transl Med.* 2019;17(1):28. doi: 10.1186/s12967-019-1769-7. PubMed PMID: 30651114; PMCID: PMC6335850.
- 40 69. Deng W, Fernandez A, McLaughlin SL, Klink DJ, 2nd. WNT1-inducible signaling pathway protein 1 (WISP1/CCN4) stimulates melanoma invasion and metastasis by promoting the epithelial-mesenchymal transition. *J Biol Chem.* 2019;294(14):5261-80. doi: 10.1074/jbc.RA118.006122. PubMed PMID: 30723155; PMCID: PMC6462510.

70. Jing D, Zhang Q, Yu H, Zhao Y, Shen L. Identification of *WISP1* as a novel oncogene in glioblastoma. *Int J Oncol*. 2017;51(4):1261-70. doi: 10.3892/ijo.2017.4119. PubMed PMID: 28902353.
- 5 71. Albig W, Meergans T, Doenecke D. Characterization of the H1.5 gene completes the set of human H1 subtype genes. *Gene*. 1997;184(2):141-8. doi: 10.1016/s0378-1119(96)00582-3. PubMed PMID: 9031620.
72. Sancho M, Diani E, Beato M, Jordan A. Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. *PLoS Genet*. 2008;4(10):e1000227. doi: 10.1371/journal.pgen.1000227. PubMed PMID: 18927631; PMCID: PMC2563032.
- 10 73. Happel N, Doenecke D. Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene*. 2009;431(1-2):1-12. doi: 10.1016/j.gene.2008.11.003. PubMed PMID: 19059319.
- 15 74. Knight P, Gauthier ML, Pardo CE, Darst RP, Kapadia K, Browder H, Morton E, Riva A, Kladde MP, Bacher R. Methyloscaper: an R/shiny app for joint visualization of DNA methylation and nucleosome occupancy in single-molecule and single-cell data. *Bioinformatics*. 2021 Jun 14;37(24):4857–9. doi: 10.1093/bioinformatics/btab438. Epub ahead of print. PMID: 34125875; PMCID: PMC8665741.
- 20 Fortin JM, Azari H, Zheng T, Darioosh RP, Schmoll ME, Vedam-Mai V, Deleyrolle LP, Reynolds BA. Transplantation of Defined Populations of Differentiated Human Neural Stem Cell Progeny. *Sci Rep*. 2016 Mar 31;6:23579. doi: 10.1038/srep23579. PMID: 27030542; PMCID: PMC4814839.
- Azari H, Millette S, Ansari S, Rahman M, Deleyrolle LP, Reynolds BA. Isolation and expansion of human glioblastoma multiforme tumor cells using the neurosphere assay. *J Vis Exp*. 2011 Oct 30;(56):e3633. doi: 10.3791/3633. PMID: 22064695; PMCID: PMC3227195.

Table 1: FENG C oligonucleotides used in this study.

Oligo name	SEQ ID NO	Oligo sequence	Notes
Oligo U1 (U1 primer)	1	GGTGTGTGTGGGTAGTTGGTGAT	Oligo does not form a 1-nt 3' flap when base paired with an oligo 1-N or an oligo 2-N annealed to a sequence of interest
Oligo U1-A	2	GGTGTGTGTGGGTAGTTGGTGATA	Nucleotide in red forms the 1-nt 3' flap when base paired with an oligo 1-A or an oligo 2-A annealed to a sequence of interest
Oligo U1-C	3	GGTGTGTGTGGGTAGTTGGTGATC	Nucleotide in red forms the 1-nt 3' flap when base paired with an oligo 1-C or an oligo 2-C annealed to a sequence of interest
Oligo U1-G	4	GGTGTGTGTGGGTAGTTGGTGATG	Nucleotide in red forms the 1-nt 3' flap when base paired with an oligo 1-G or an oligo 2-G annealed to a sequence of interest
Oligo U1-T	5	GGTGTGTGTGGGTAGTTGGTGATT	Nucleotide in red forms the 1-nt 3' flap when base paired with an oligo 1-T or an oligo 2-T annealed to a sequence of interest
Oligo U2	6	/5Phos/AGAGAATGAGGAAGGTGG*G*G*T*A*G/3SpC3/	/5Phos/, 5' Phosphate; *, Phosphorothioate Bonds; /3SpC3/, 3' C3 Spacer (C, carbon)
U2 primer	7	CTACCCACCTTCCTATTCTCT	
200-A oligo	8	CGGTACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGC GTCAGCGGGTGTGGCGGGTGTGGGGCTGGCTTAACTATGCGGCATCAGAGCAG ATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGA GAAAATACCGCATCAGGCGCCATTGCCATTTCAGG	200-mer substrate A; nucleotide in red is immediately 5' of FEN cleavage site

200-C oligo	9	CGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGC GTCAGCGGGTGTGGCGGGTGTGGGGCTGGCTTAACTATGCGGCATCAGAGCAG ATTGTAAGTGAAGTGCACCTTATGCGGTGTGAAATACCGCACAGATGCGTAAGGA GAAAATACCGCATCAGGCGCCATTGCCATTTCAGG	200-mer substrate C; nucleotide in red is immediately 5' of FEN cleavage site
200-G oligo	10	CGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGC GTCAGCGGGTGTGGCGGGTGTGGGGCTGGCTTAACTATGCGGCATCAGAGCAG ATTGTAAGTGAAGTGCACCTTATGCGGTGTGAAATACCGCACAGATGCGTAAGGA GAAAATACCGCATCAGGCGCCATTGCCATTTCAGG	200-mer substrate G; nucleotide in red is immediately 5' of FEN cleavage site
200-T oligo	11	CGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGC GTCAGCGGGTGTGGCGGGTGTGGGGCTGGCTTAACTATGCGGCATCAGAGCAG ATTGTAAGTGAAGTGCACCTTATGCGGTGTGAAATACCGCACAGATGCGTAAGGA GAAAATACCGCATCAGGCGCCATTGCCATTTCAGG	200-mer substrate T; nucleotide in red is immediately 5' of FEN cleavage site
200 oligo 1-A	12	CGCATCTGTGCGGTATTTACACCGCATATATCACCAACTACCCACACACACC	Nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1
200 oligo 1-C	13	CGCATCTGTGCGGTATTTACACCGCATAGATCACCAACTACCCACACACACC	Nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1
200 oligo 1-G	14	CGCATCTGTGCGGTATTTACACCGCATACATCACCAACTACCCACACACACC	Nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1
200 oligo 1-T	15	CGCATCTGTGCGGTATTTACACCGCATAAATCACCAACTACCCACACACACC	Nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1
Unmethyla ted 80-nt oligo	16	GTCTGAACTATGCAGCATCAGAGCAGATTGTAAGATACAGTGCATGCGCGTGCCG ATGCGAATACACAGACGTGTCAGGA	Control unmethylated C in red
Methylated 80-nt oligo	17	GTCTGAACTATGCAGCATCAGAGCAGATTGTAAGATACAGTGS _{5mC} ATGS _{5mC} CG _{5mC} CG TG _{5mC} CG _{5mC} GATGCGAATACACAGACGTGTCAGGA	5mC in red: 5- Methylcytosine
80 oligo 1- A	18	TATTCGCATCGGCACGCGCATATCACCAACTACCCACACACACC	Non-bolded nucleotides are complementary to the 5' end of target sequence;

			nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1
80 oligo 1-C	19	GTGTATTTCGCATCGGCACGCGATC CACCAACTACCCACACACACC	Non-bolded nucleotides are complementary to the 5' end of target sequence; nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1
80 oligo 1-G	20	TGTATTTCGCATCGGCACGCGATC CACCAACTACCCACACACACC	Non-bolded nucleotides are complementary to the 5' end of target sequence; nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1
80 oligo 1-T	21	GTATTTCGCATCGGCACGCGCAATC CACCAACTACCCACACACACC	Non-bolded nucleotides are complementary to the 5' end of target sequence; nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1
80 qPCR-F primer	22	GTCTGAACTATGCAGCATCAGAGC	Forward qPCR primer to assess cutting of unmethylated and methylated 80-nt oligo substrate.
80 qPCR-R primer	23	TCCTGACACGTCTGTGTATTTCG	Reverse qPCR primer to assess cutting of unmethylated and methylated 80-nt oligo substrate.
pGEM-3Z/601b NdeI oligo 1-T	24	CTGTGCGGTATTT CACACCGCAT ATC CACCAACTACCCACACACACC	Forms 5' flap with U1-T oligo on wild-type pGEM-3Z/601b

pGEM-3Z/601b HindIII oligo	25	CTACCCACCTTCCTCATTCTCTTGCATGCCTGCAGGTCGGG	Ligates to 3' end of HindIII-cut target
Methyl test pGEM-3Z/601b oligo 1-C	26	TCAGTACAATCTGCTCTGATGCCGATCACCAACTACCCACACACACC	Directs flap cleavage 5' of nt 2419 in pGEM-3Z/601b
Methyl test pGEM-3Z/601b oligo 1-G	27	AGTACAATCTGCTCTGATGCCGATCACCAACTACCCACACACACC	Directs flap cleavage 5' of nt 2418 in pGEM-3Z/601b
Methyl test pGEM-3Z/601b oligo 1-T	28	GTACAATCTGCTCTGATGCCGATATCACCAACTACCCACACACACC	Directs flap cleavage 5' of nt 2415 in pGEM-3Z/601b
pGEM-3Z/601b mutagenesis to C oligo 1	29	TGCACCTATGCGGTGT	Nucleotide in NdeI site in red was mutated from A to C to create pGEM-3Z/601b-C
pGEM-3Z/601b mutagenesis to G oligo 1	30	TGCACCTATGCGGTGT	Nucleotide in NdeI site in red was mutated from A to G to create pGEM-3Z/601b-G
pGEM-3Z/601b mutagenesis to T oligo 1	31	TGCACCTATGCGGTGT	Nucleotide in NdeI site in red was mutated from A to T to create pGEM-3Z/601b-T
pGEM-3Z/601b mutagenesis oligo 2	32	CTCTCAGTACAATCTGCTCTGA	To create three of four pGEM-3Z/601b-N substrates
pGEM-3Z/601b NdeI-2 oligo 1-A	33	CGCATCTGTGCGGTATTTACACCGCATATATCACCAACTACCCACACACACC	Suffix -2 indicates a 1-nt shift in hybridization to the target sequence compared with above pGEM-3Z/601b NdeI oligo 1-T. Non-bolded nucleotides are complementary to the 5' end of target sequence; nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1.
pGEM-3Z/601b NdeI-2 oligo 1-C	34	CGCATCTGTGCGGTATTTACACCGCATATATCACCAACTACCCACACACACC	Suffix -2 indicates a 1-nt shift in hybridization to the target sequence compared with

			above pGEM-3Z/601b NdeI oligo 1-T. Non-bolded nucleotides are complementary to the 5' end of target sequence; nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1.
pGEM-3Z/601b NdeI-2 oligo 1-G	35	CGCATCTGTGCGGTATTTACACCCGCATACAT CACCAACTACCCACACACACC	Suffix -2 indicates a 1-nt shift in hybridization to the target sequence compared with above pGEM-3Z/601b NdeI oligo 1-T. Non-bolded nucleotides are complementary to the 5' end of target sequence; nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1.
pGEM-3Z/601b NdeI-2 oligo 1-T	36	CGCATCTGTGCGGTATTTACACCCGCATA AA T CACCAACTACCCACACACACC	Suffix -2 indicates a 1-nt shift in hybridization to the target sequence compared with above pGEM-3Z/601b NdeI oligo 1-T. Non-bolded nucleotides are complementary to the 5' end of target sequence; nucleotide in red is immediately 5' of FEN cleavage site; bolded nucleotides are complementary to oligo U1.
MGO3791	37	tttAtAGAGTTGAGAAATTTGAtT	Primer used with MGO3792 for BS-PCR of a 732-bp amplicon from

			the <i>MLH1</i> promoter, with lower case t denoting C to t conversion.
MGO3792	38	TaaATaTaAACAAaTaATTCTaaAATaAAaTa	Primer used with MGO3791 for BS-PCR of a 732-bp amplicon from the <i>MLH1</i> promoter, with lower case a denoting G to a conversion.

Oligo name	SEQ ID NO	Oligo sequence	Notes	
300-nt-EPM2AIP1-oligo 1-T	39	GCGCCACTACGAGGCTGAGCACGAATATCACCAACTACCCACACACACC	1. All oligos were designed to capture target sequences of ~300 nt in length.	Same as ~450-nt oligo 1-T
300-nt-EXO1-oligo 1-T	40	AGCCCCAGTTCTGGAGGCTGAATCACCAACTACCCACACACACC		Same as ~450-nt oligo 1-T
300-nt-HSPA5-oligo 1-T	41	GTCGGAGTAGGTCCAGCAGGAATCACCAACTACCCACACACACC		Same as ~450-nt oligo 1-T
300-nt-MLH1-oligo 1-T	42	GGAAACGTCTAGATGCTCAACGGAAATCACCAACTACCCACACACACC	2. Non-bolded nucleotides are complementary to the 5' end of target sequence.	Same as ~450-nt oligo 1-T
300-nt-MLH3-oligo 1-T	43	CTGGCCTATGGTAGGCACCGAAATCACCAACTACCCACACACACC		Same as ~450-nt oligo 1-T
300-nt-MSH2-oligo 1-T	44	AAACTAGCCTTCTCCTCCTCCAGTAATCACCAACTACCCACACACACC	3. Nucleotide in red is complementary to nucleotide immediately 5' of FEN cleavage site.	Same as ~450-nt oligo 1-T
300-nt-MSH3-oligo 1-T	45	CCGACTCCCGGCGAGAGGAATCACCAACTACCCACACACACC		Same as ~450-nt oligo 1-T
300-nt-MSH6-oligo 1-T	46	GATAACCGGGGGGACCCGAAATCACCAACTACCCACACACACC		Same as ~450-nt oligo 1-T
300-nt-PMS1-oligo 1-T	47	GCTTTGCCTTCCAGCTGAGATGTAAATCACCAACTACCCACACACACC	4. Bolded nucleotides are complementary to oligo U1.	Same as ~450-nt oligo 1-T
300-nt-PMS2-oligo 1-T	48	CTCTCCGTGTGGAGCTTCCCAATCACCAACTACCCACACACACC		Same as ~450-nt oligo 1-T
300-nt-TDG-oligo 1-T	49	CTGTGTTCATTATCAGGTAATGAGATGTGAGGGAAATCACCAACTACCCACACACACC		Same as ~450-nt oligo 1-T
300-nt-EPM2AIP1-oligo 2-T	50	AACCGGGCTTGTGTGCCTCTGCTGAATCACCAACTACCCACACACACC		
300-nt-EXO1-oligo 2-T	51	GCCCCGCCAGAGGATGTGAATCACCAACTACCCACACACACACC		
300-nt-HSPA5-oligo 2-T	52	CGTCGTGGAGGCGCGCGAATCACCAACTACCCACACACACACC		

300-nt-MLH1-oligo 2-T	53	AGGAGAAGGCCTGACTGGCAGCTCAATCACC AACTACCCA CACACACC		
300-nt-MLH3-oligo 2-T	54	CAAGGATCCGCATCCCAGGGTTCATCACC AACTACCCAC ACACACC		
300-nt-MSH2-oligo 2-T	55	GTCTGCCGCGGCAACCAATCAATCACC AACTACCCACACA CACC		
300-nt-MSH3-oligo 2-T	56	CGTCTGCGGGGCGGAGCAATCACC AACTACCCACACACA CC		
300-nt-MSH6-oligo 2-T	57	GGGCGGTGCGCGCC ATCACC AACTACCCACACACACC		
300-nt-PMS1-oligo 2-T	58	GTGACGTGCGGCCGCGAGAAATCACC AACTACCCACACA CACC		
300-nt-PMS2-oligo 2-T	59	GGCTGCTTTGACGTTGTGCTCCAATCACC AACTACCCAC ACACACC		
300-nt-TDG-oligo 2-T	60	ACGCGTACTGGCCGTCACGTGAATCACC AACTACCCACAC ACACC		
300-nt-EPM2AIP1-oligo 3-T	61	CTACCCACCTTCCTCATTCTCTAGGTGATCTGGCGCAGA GCGGAGG	1. All oligos were designed to capture target sequences of ~300 nt in length.	
300-nt-EXO1-oligo 3-T	62	CTACCCACCTTCCTCATTCTCTACGTCAACGCGTATCCC GCAACCC		
300-nt-HSPA5-oligo 3-T	63	CTACCCACCTTCCTCATTCTCTATTGGTGAAGGCCCTGC TCGTTGG		
300-nt-MLH1-oligo 3-T	64	CTACCCACCTTCCTCATTCTCTAGGGAACCCGGCGGCCT C	2. Non-bolded nucleotides are complementary to the 3' end of target sequence.	
300-nt-MLH3-oligo 3-T	65	CTACCCACCTTCCTCATTCTCTAGTGAACCGGTACCTAC AGTCTCAATCCTTATCC		
300-nt-MSH2-oligo 3-T	66	CTACCCACCTTCCTCATTCTCTATAAGCAGACGCGGCC AGC		
300-nt-MSH3-oligo 3-T	67	CTACCCACCTTCCTCATTCTCTACGCCGCGACCTGCGT G	3. Nucleotide in red is complementary to 3' end of target sequence, i.e., immediately 5' of FEN cleavage site.	
300-nt-MSH6-oligo 3-T	68	CTACCCACCTTCCTCATTCTCTAGGCCCGCCCCGG		
300-nt-PMS1-oligo 3-T	69	CTACCCACCTTCCTCATTCTCTAGGAAGGGCGGGTTG GGG		
300-nt-PMS2-oligo 3-T	70	CTACCCACCTTCCTCATTCTCTACCCTTCCGGGCGGGG C	4. Bolded nucleotides are complementary to oligo U2.	
300-nt-TDG-oligo 3-T	71	CTACCCACCTTCCTCATTCTCTATGTAGGACAGCCCAAT CACGCTCACC		

Oligo name	SEQ ID NO	Oligo sequence	Notes
------------	-----------	----------------	-------

450-nt-ALKBH1-oligo 1-T	72	CAGCGTGGTGGCTGAATTAGGAG AATCACC AACTACCCACACACACC	1. All oligos were designed to capture target sequences of ~450 nt in length.	
450-nt-ALKBH2-oligo 1-T	73	AGGAGTTCTGCAGTCGGATATCGTGG AATCACC AACTACCCACACACACC		
450-nt-ALKBH3-oligo 1-T	74	GCATCGGCATCTTTGTGGCAGAAAG AATCACC AACTACCCACACACACC		
450-nt-APEX1-oligo 1-T	75	GGTCTCCGCTCGCGGG AATCACC AACTACCCACACACACACC	2. Non-bolded nucleotides are complementary to the 5' end of target sequence.	
450-nt-APEX2-oligo 1-T	76	GCCACTTTTAAAGCGACGGAAACCCTGA AATCACC AACTACCCACACACACC		
450-nt-ATM-oligo 1-T	77	GTGGTTCCTGCTGTGGTTTGATCA AATCACC AACTACCCACACACACC		
450-nt-ATRX-oligo 1-T	78	CGGCACTACACTAATCCTTTATCCGGGTTATCTCATTTAA ATCA CCA ACTACCC ACACACACACC	3. Nucleotide in red is complementary to nucleotide immediately 5' of FEN cleavage site.	
450-nt-ATXN3-oligo 1-T	79	CCGCTTCCCA CCCCATTTCTATGGATTCTTAATCACC AACTACCCACACACACC		
450-nt-AXIN2-oligo 1-T	80	GTTTCCCTGGGAATTTGGGCTTGG AATCACC AACTACCCACACACACC		
450-nt-BRCA1-oligo 1-T	81	CCCTAATGGAGGTCTCCAGTTTCGGTAAATATA AAATCACC AACTACCCACACACACACC	4. Bolded nucleotides are complementary to oligo U1.	
450-nt-BRCA2-oligo 1-T	82	GTGGGCGCCCTCGGG AATCACC AACTACCCACACACACACC		
450-nt-CCN4-oligo 1-T	83	CCTGCCCTAGGTCTTTCACTACTGTAGGG AATCACC AACTACCCACACACACACC		
450-nt-CD44-oligo 1-T	84	GGGGAGTTGGTGAATCTTCCAAGTGGAAAG AATCACC AACTACCCACACACACACC		
450-nt-CDH13-oligo 1-T	85	TGGTGCCTCCCTGCCATCTCACAA AATCACC AACTACCCACACACACACC		
450-nt-CDH1-oligo 1-T	86	AGTGCAGTGGTGCATCACAGCTCA AATCACC AACTACCCACACACACACC		
450-nt-CDH2-oligo 1-T	87	CCGCAAGGGGGTGAGGCG AATCACC AACTACCCACACACACACC		
450-nt-CDKN2A-oligo 1-T	88	GGCAGCTCGCCCCCTAGCT AATCACC AACTACCCACACACACACC		
450-nt-CDKN2D-oligo 1-T	89	GCAGGCGGGAAAGAAGGCTGGTTTA AATCACC AACTACCCACACACACC		
450-nt-CETN1-oligo 1-T	90	GACCCAGTTTATAGGTTCTTTAGGG AATCACC AACTACCCACACACACC		
450-nt-CHEK1-oligo 1-T	91	TGGGGAACATAGGGAGACCCCATCTCTA AATCACC AACTACCCACACACACC		
450-nt-CHEK2-oligo 1-T	92	GCCAGATGTGGCGGGGGAG AATCACC AACTACCCACACACACACC		
450-nt-CSDE1-oligo 1-T	93	GGCCAACAATTCTGATGGCATATTCAGGAGTCA AATCACC AACTACCCACACACACC		
450-nt-DBP2-oligo 1-T	94	CAGCCTCTAAGTAGATGGGACTACAA AATCACC AACTACCCACACACACC		
450-nt-DRD3-oligo 1-T	95	TGATGTCAATATCCTTACCTGCCTATTTA AATCACC AACTACCCACACACACC		

450-nt-DRD4-oligo 1-T	96	CGCGACCCGAGAAACCGACAATCACCAACTACCCACACACACC		
450-nt-EGFR-oligo 1-T	97	CCCCCTTCCCTTCTTTTGTTTACAAAGCAAACCTGTATCACCAACTACCCACACACACC		
450-nt-EPM2AIP1-oligo 1-T	98	GCGCCACTACGAGGCTGAGCACGAATATCACCAACTACCCAACACACACC		Same as ~300-nt oligo 1-T
450-nt-EPS11-oligo 1-T	99	CAGGCAACTCTAATTTGAGGTTACCGCTGGATCACCAACTACCCACACACACC		
450-nt-ERCC1-oligo 1-T	100	CTCAGACATTTGACTTGGCTTCAGTTTCCTCAATCACCAACTACCCACACACACC		
450-nt-ERCC2-oligo 1-T	101	CCCAGGAGTTGAGATCAGCTTGAGCAAGAATCACCAACTACCCACACACACC		
450-nt-ERCC8-oligo 1-T	102	ATCCCGCAGTACAAGAAGCTGGAGATCACCAACTACCCACACACACC		
450-nt-EXO1-oligo 1-T	103	AGCCCCAGTTCTGGAGGCTGAATCACCAACTACCCACACACACACC		Same as ~300-nt oligo 1-T
450-nt-FANCC-oligo 1-T	104	GTAGGGGAGGTAAAAATGGAAATTAGTTCCTTTTGAATCACCAACTACCCACACACACC		
450-nt-FAT4-oligo 1-T	105	GAGTTCCAGAAAAAGTTTCTCAAGCCTCTTATATAATCACCAACTACCCACACACACC		
450-nt-FEN1-oligo 1-T	106	CGAGAGGCGTGGGGCGAATCACCAACTACCCACACACACACC		
450-nt-FHIT-oligo 1-T	107	ACGTGTAACGAAAGGCTGTACCGCAATCACCAACTACCCACACACACC		
450-nt-FOXO2-oligo 1-T	108	TGGGCCCCGCGGAGGAATCACCAACTACCCACACACACACC		
450-nt-FZD7-oligo 1-T	109	CGCACCGCAGCGCGCAATCACCAACTACCCACACACACACC		
450-nt-FZD8-oligo 1-T	110	GGCAGGGCCCTTCCGCAATCACCAACTACCCACACACACACC		
450-nt-GPR158-oligo 1-T	111	CGCTCTCATTTGAGGCGGTTTCAAGAAATCACCAACTACCCACACACACC		
450-nt-GSTA1-oligo 1-T	112	CCATGTAAATGTGAATTTTCAAGCAAAACAATCAATCACCAACTACCCACACACACACC		
450-nt-HIST1H1B-oligo 1-T	113	CATACTTGTGTGTGTTTCAAGAATCTTGTGCAATTGAATCACCAACTACCCACACACACACC		
450-nt-HMGA2-oligo 1-T	114	CGGCGGGGGGTTAAAGGGATCACCAACTACCCACACACACACC		
450-nt-HOTAIR-oligo 1-T	115	CCGAGCCTTAATGTAGGAGCTAAAGCCAGATCACCAACTACCCACACACACACC		
450-nt-HOXB2-oligo 1-T	116	GCCTCCGTTTCTCCAGAAAAGCTAGAATCACCAACTACCCACACACACACC		
450-nt-HSPA5-oligo 1-T	117	GTCGGAGTAGGTCCAGCAGGAATCACCAACTACCCACACACACACC		Same as ~300-nt oligo 1-T
450-nt-INO80-oligo 1-T	118	CGCAGGAGAAAAATGTCGCGAGAGCTTTCCAATCACCAACTACCCACACACACACC		
450-nt-KLF8-oligo 1-T	119	GGGCTGTCTAAGAGGGAAAGGAGAAAAACAAAAATCACCAACTACCCACACACACACC		
450-nt-LIG1-oligo 1-T	120	GCAGCTGGACGTGGGGGGATCACCAACTACCCACACACACACACC		
450-nt-MBD4-oligo 1-T	121	GCTTGCTGCTGGGCGGATCACCAACTACCCACACACACACACC		
450-nt-MCM9-oligo 1-T	122	CCCCAAATTTCTTCATATTAACCCCTGCAATCACCAACTACCCACACACACACC		
450-nt-MECP2-oligo 1-T	123	GCAAGTGAAGGCCTAGGCTTGGGAATCACCAACTACCCACACACACACC		
450-nt-MEG3-oligo 1-T	124	ATTGAGGCACAATGTGTGGGGGGAATCACCAACTACCCACACACACACC		
450-nt-MGMT-oligo 1-T	125	GATGGCCCTTCGGCCGGTACAATCACCAACTACCCACACACACACC		
450-nt-MLH1-oligo 1-T	126	GGAAACGTCTAGATGCTCAACGGAAATCACCAACTACCCACACACACACC		Same as ~300-nt oligo 1-T
450-nt-MLH3-oligo 1-T	127	CTGGCCTATGGTAGGCACCGAATCACCAACTACCCACACACACACC		Same as ~300-nt oligo 1-T
450-nt-MNAT1-oligo 1-T	128	GATTTATGAGCGGTGTACTTACTTTCTTTGGACCAATCACCAACTACCCACACACACACC		

450-nt-MPG-oligo 1-T	129	ACGTGTTTTCCGGTGGGGTGCCAAATCACCAACTACCCACACACACC		
450-nt-MRE11-oligo 1-T	130	CCGCTTCGTGGTCCCCAGGTAATCACCAACTACCCACACACACC		
450-nt-MSH2-oligo 1-T	131	AAACTAGCCTTCTCTCTCTCCAGTAATCACCAACTACCCACACACACC		Same as ~300-nt oligo 1-T
450-nt-MSH3-oligo 1-T	132	CCGACTCCCGGCGAGAGGAATCACCAACTACCCACACACACC		Same as ~300-nt oligo 1-T
450-nt-MSH4-oligo 1-T	133	GCACACTAGCTTATCTTTAAATGTAGAAACCACCAACAATCACTCAACTACCCACACACACC		
450-nt-MSH5-oligo 1-T	134	GGGTTTTGAGGGGGAGGGCCAAACTTAAAGAAATCACCAACTACCCACACACACC		
450-nt-MSH6-oligo 1-T	135	GATAACCGGGGGGACCCGAAATCACCAACTACCCACACACACC		Same as ~300-nt oligo 1-T
450-nt-MUTYH-oligo 1-T	136	CAAGCCCTACAAAATGTTTCTGTCAGGAATCACCAACTACCCACACACACC		
450-nt-NEIL2-oligo 1-T	137	CTGGCGCCCGGCTAATTTTGTATTTTGTAGAGAAATCACCAACTACCCACACACACC		
450-nt-NEIL3-oligo 1-T	138	CCTCTTAATCTGTCCTGTCTGTGACAATCACCAACTACCCACACACACC		
450-nt-NKX2-5-oligo 1-T	139	GGGTGGGAAGCGGTGGGGAATCACCAACTACCCACACACACC		
450-nt-OGG1-oligo 1-T	140	CGTCTCCCCCGAGCCTTTGCAATCACCAACTACCCACACACACC		
450-nt-OLIG2-oligo 1-T	141	GGGGATGGAGGGAGGGCTGAATCACCAACTACCCACACACACC		
450-nt-PARP1-oligo 1-T	142	AGGCGCTCACCACAACCTGGCTAATCACCAACTACCCACACACACC		
450-nt-PCNA-oligo 1-T	143	AGCCCTGCTCCCGGGAGAAATCACCAACTACCCACACACACC		
450-nt-PELO-oligo 1-T	144	ACCACCCCCGACAGGCTCCAATCACCAACTACCCACACACACC		
450-nt-PMS1-oligo 1-T	145	GCTTTGCCCTTCCAGCTGAGATGTAAATCACCAACTACCCACACACACC		Same as ~300-nt oligo 1-T
450-nt-PMS2-oligo 1-T	146	CTCTCCGTGTGGAGCTTCCCAATCACCAACTACCCACACACACC		Same as ~300-nt oligo 1-T
450-nt-PMS2P1-oligo 1-T	147	CCCATCACCATGCCTGGCTAATTTTGTAAATCACCAACTACCCACACACACC		
450-nt-PMS2P1-oligo 1-T	148	GAGTTCGAGACCAGTCTGGCCAATCACCAACTACCCACACACACC		
450-nt-PMS2P2-oligo 1-T	149	CCTTGGTGTCTGGCGGTTTCCAATCACCAACTACCCACACACACC		
450-nt-PMS2P3-oligo 1-T	150	CTGGAGCCCAAGTCTCTGGGTTAATCACCAACTACCCACACACACC		
450-nt-PMS2P5-oligo 1-T	151	GCGTTGGTGTCTGGCGGTTTCTGAATCACCAACTACCCACACACACC		
450-nt-POLD1-oligo 1-T	152	TGGGGTGTTCGCGCGCGCAATCACCAACTACCCACACACACACC		
450-nt-POLD4-oligo 1-T	153	GCTTCGCTAATCCCCGTGGCCAATCACCAACTACCCACACACACC		
450-nt-POLG-oligo 1-T	154	GCCGCGGTGGAAGGGAGGTTATGATTAACCTTTAATCACCAACTACCCACACACACC		
450-nt-POLK-oligo 1-T	155	AACGCAGGGGTCACGGCGAATCACCAACTACCCACACACACACC		
450-nt-POLQ-oligo 1-T	156	AGGTAGATCAGGGAAGTAGCTCTTCTTAATCACCAACTACCCACACACACC		
450-nt-POLR2B-oligo 1-T	157	AGGGTCACCACGCCCGGGTAATCACCAACTACCCACACACACC		
450-nt-POLR2G-oligo 1-T	158	TGTGCCACCACACCTGACTAATTTTGTAAATCACCAACTACCCAACACACACC		
450-nt-PRDM2-oligo 1-T	159	CAGACAAGCTGTGCCAAGAAGTAATCAATCACCAACTACCCACACACACC		
450-nt-PRKCG-oligo 1-T	160	GGGCGCTAAGGTTCTCACTTCTCCGAATCACCAACTACCCAACACACACC		
450-nt-PRUNE2-oligo 1-T	161	GGAGGTGCAGGGGTATGGCGAATCACCAACTACCCACACACACC		
450-nt-PXMP4-oligo 1-T	162	GGAGGAGCAGGGCCCCAATCACCAACTACCCACACACACACC		

450-nt-PYCARD-oligo 1-T	163	CATGTCGTAGAATCTTGGGAATCATGACAGCTAGAAATCACCAACTACCCACACACACC		
450-nt-RNASEH2A-oligo 1-T	164	CGTAGCTTCGAAGACCCAGCCTAATCACCAACTACCCACACACACC		
450-nt-RNF111-oligo 1-T	165	CAACATTTAAGCCAAATCTTGACTACATCTTGGTTCAATCACCAACTACCCACACACACC		
450-nt-RPA2-oligo 1-T	166	GGTCACATTTCTCTTCGTGTAATGCCACCAATCACCAACTACCCACACACACC		
450-nt-RPA3-oligo 1-T	167	GGGGAAATGTATGTGTGTTTAAAGCTTCCCAATCACCAACTACCCACACACACC		
450-nt-RPA4-oligo 1-T	168	CTTTGGGAGGCTGAGGCAGGTTGATCACCAACTACCCACACACACC		
450-nt-SH3TC1-oligo 1-T	169	CTTCCATTTACTGCACACACACACACAATCACCAACTACCCACACACACC		
450-nt-SIM2-oligo 1-T	170	GCGGCAGGGAGGGCGGAATCACCAACTACCCACACACACACC		
450-nt-SMAD4-oligo 1-T	171	GCGTAGAGTGGGCGTCCAGTAATCACCAACTACCCACACACACC		
450-nt-SMUG1-oligo 1-T	172	CATTCACTATTTTCCCAAAGGGATTGCAATCAATCACCAACTACCCACACACACC		
450-nt-SNAI1-oligo 1-T	173	GGGCGGGGACTCCTCCGAATCACCAACTACCCACACACACACC		
450-nt-SNAI2-oligo 1-T	174	GCCTGGCTTCCAGATGTGGTGCAATCACCAACTACCCACACACACC		
450-nt-TAF1-oligo 1-T	175	GGAGCTGCTCGAGTCACGTGGCTTAATCACCAACTACCCACACACACC		
450-nt-TDG-oligo 1-T	176	CTGTGTTGCATTATCAGGTAATGAGATGTGAGGGATCACCAACTACCCACACACACC		Same as ~300-nt oligo 1-T
450-nt-TP53-oligo 1-T	177	CTCTCAGCTGGATCCTTTCTTCTCATCAGTTAAATCACCAACTACCCACACACACC		
450-nt-TREX1-oligo 1-T	178	CATGCCACCTGTGGGTCCCAATCACCAACTACCCACACACACACC		
450-nt-TRIM28-oligo 1-T	179	TTCGCCGCCGTAGAGTCGTCTCCAATCACCAACTACCCACACACACACC		
450-nt-TWIST1-oligo 1-T	180	GCAGCGCCCCCGAACCCTAATCACCAACTACCCACACACACACC		
450-nt-TWIST2-oligo 1-T	181	AGTCCCAGATGCGCTCCAGAAATCACCAACTACCCACACACACACC		
450-nt-UBB-oligo 1-T	182	ACCTCCGCCCTCCGGGTCAAAATCACCAACTACCCACACACACACC		
450-nt-USP47-oligo 1-T	183	CTTCTCTAGTAGTAAGCATGGTGCTTTGTGTTATAATCACCAACTACCCACACACACACC		
450-nt-WDR48-oligo 1-T	184	GGCCCAAGCAAAACGAAAGCACTGAAAGAAATCACCAACTACCCACACACACACC		
450-nt-WNT1-oligo 1-T	185	TCCGGAGCCCGTCGAGGACTAATCACCAACTACCCACACACACACACC		
450-nt-XPC-oligo 1-T	186	AGGCTGCTTTTCCAGACTCAGGAAAAAATCACCAACTACCCACACACACACC		
450-nt-XRCC2-oligo 1-T	187	GGCCCCCTTTGTCCGTTCTCGGGTTAAATCACCAACTACCCACACACACACC		
450-nt-XRCC3-oligo 1-T	188	TCCGCGCGCGGCTGACAATCACCAACTACCCACACACACACACC		
450-nt-ZEB1-oligo 1-T	189	GGGCAGGGAGGGATCTGGCTGAATCACCAACTACCCACACACACACC		
450-nt-ZNF830-oligo 1-T	190	GACAGCGGTTGGAGCGGCCAATCACCAACTACCCACACACACACACC		
450-nt-ALKBH1-oligo 2-T	191	CTGCGTGGGCGCGGAATCACCAACTACCCACACACACACC		
450-nt-ALKBH2-oligo 2-T	192	CGATTAGAAAGCGAGACGCAAGAGGTTAAACAATCACCAACTACCCACACACACACC		Same as ~940-nt oligo 2-T
450-nt-ALKBH3-oligo 2-T	193	GGGTGGTACTCGCACCCCCAATCACCAACTACCCACACACACACACC		
450-nt-APEX1-oligo 2-T	194	CCTCTTATGAGCAAAAGAGCAACCCCGTAATCACCAACTACCCACACACACACC		Same as ~940-nt oligo 2-T
450-nt-APEX2-oligo 2-T	195	CGGCACAGTTGCTGGGTTCTGAATCACCAACTACCCACACACACACC		Same as ~940-nt oligo 2-T

450-nt-ATM-oligo 2-T	196	GCCAAGCAGCCGCAGAGCAATCACCAACTACCCACACACAC C		Same as ~940-nt oligo 2-T
450-nt-ATRX-oligo 2-T	197	CGTCACTGTAGCTGCTGCTGGAAATCACCAACTACCCACACA CACC		Same as ~940-nt oligo 2-T
450-nt-ATXN3-oligo 2-T	198	CCCCTCGCCGGGCGAGAAATCACCAACTACCCACACACACC		
450-nt-AXIN2-oligo 2-T	199	GAGTTGCCAGGACCTTATCAAAGCGCAATCACCAACTACCCA CACACACC		
450-nt-BRCA1-oligo 2-T	200	GGGCCCCAGTTATCTGAGAAACCCCAATCACCAACTACCCACA CACACC		Same as ~940-nt oligo 2-T
450-nt-BRCA2-oligo 2-T	201	GGTGCAGTTCTTTTGGCCGGAGTAAATCACCAACTACCCA CACACACC		Same as ~940-nt oligo 2-T
450-nt-CCN4-oligo 2-T	202	AGGCAGGACTCACCGTGGCCAATCACCAACTACCCACACAC ACC		Same as ~940-nt oligo 2-T
450-nt-CD44-oligo 2-T	203	GGCTGTGACTAAGCAGCCCGCAATCACCAACTACCCACACAC ACC		Same as ~940-nt oligo 2-T
450-nt-CDH13-oligo 2-T	204	TGGTCATCAGCCTCTACCCAATGCTGCAATCACCAACTACCC ACACACACC		
450-nt-CDH1-oligo 2-T	205	GGCTGGAGCGGGCTGGAATCACCAACTACCCACACACACACC		
450-nt-CDH2-oligo 2-T	206	CCCAGAGATCAAGGAGCTGGGGAAATCACCAACTACCCACAC ACACC		
450-nt-CDKN2A-oligo 2-T	207	ACCATGTTCTCGCCGCCTCCAATCACCAACTACCCACACACA CC		
450-nt-CDKN2D-oligo 2-T	208	ACGGCGAGGGGCTGGGAATCACCAACTACCCACACACACACC		
450-nt-CETN1-oligo 2-T	209	AAATGCTTCCCGAACTTCTTGCTTCTGAATCACCAACTACCCA CACACACC		Same as ~940-nt oligo 2-T
450-nt-CHEK1-oligo 2-T	210	GAAAGCGAGCAGTTTATGTTGGAGCTTGAAATCACCAACTAC CCACACACACC		Same as ~940-nt oligo 2-T
450-nt-CHEK2-oligo 2-T	211	CACACACTCCCCCTTCAGCTCAAAAATCAATCACCAACTAC CCACACACACC		Same as ~940-nt oligo 2-T
450-nt-CSDE1-oligo 2-T	212	GGCTTCTCGTTAGCCTTGGTCTTATTTCTCAATCACCAACTACC CACACACACC		Same as ~940-nt oligo 2-T
450-nt-DBP2-oligo 2-T	213	GGGTACTGTGCTCTATGCGAAGATCAATCACCAACTACCCAC ACACACC		Same as ~940-nt oligo 2-T
450-nt-DRD3-oligo 2-T	214	CCTGAGACTGGATAATTTATAATGAACAGAAATCACCAACT ACCCACACACACC		
450-nt-DRD4-oligo 2-T	215	CACGAGCGAGTTCCCCGCGAATCACCAACTACCCACACACAC C		
450-nt-EGFR-oligo 2-T	216	CCGCGTCGGGGCGCTCAATCACCAACTACCCACACACACACC		
450-nt-EPM2AIP1-oligo 2-T	217	GCTGGGTCTCTTCGTCCTCCCTGAAATCACCAACTACCCACA CACACC		
450-nt-EPST11-oligo 2-T	218	CGAGCCGGAGTTCACCACTCTAATCACCAACTACCCACACA CACC		Same as ~940-nt oligo 2-T
450-nt-ERCC1-oligo 2-T	219	GGCGCCAATCTCCACCCGCAATCACCAACTACCCACACACAC C		
450-nt-ERCC2-oligo 2-T	220	ACCCAGGCGCGCCCAATCACCAACTACCCACACACACACC		
450-nt-ERCC8-oligo 2-T	221	GGCCTGTTAGCCAAAAAGTAAGGTTTCTTTAATCACCAACTA CCCACACACACC		
450-nt-EXO1-oligo 2-T	222	CTCCTCTCCCCCTTGAGCGGAATCACCAACTACCCACACACA CC		Same as ~940-nt oligo 2-T
450-nt-FANCC-oligo 2-T	223	TCCTCGCGGAGCTGCTTCAATCACCAACTACCCACACACAC C		
450-nt-FAT4-oligo 2-T	224	GCTGCTGGAAGAGCAACCTCCAATCACCAACTACCCACACAC ACC		
450-nt-FEN1-oligo 2-T	225	GCAACCAGTCCCTCCAGGGAGAAATCACCAACTACCCACACA CACC		Same as ~940-nt oligo 2-T
450-nt-FHIT-oligo 2-T	226	CTCCCCGCCCTACCTTCCAATCACCAACTACCCACACACAC C		
450-nt-FOXO2-oligo 2-T	227	TAGCTGCCCCGAGCCCGGTAATCACCAACTACCCACACACAC C		
450-nt-FZD7-oligo 2-T	228	GCCAGCACCAGGGCACAGAATCACCAACTACCCACACACAC C		

450-nt-FZD8-oligo 2-T	229	GGCGAGCACTGGATCTCCACCAATCACCAACTACCCACACAC ACC		
450-nt-GPR158-oligo 2-T	230	TCGCGGCTGGCGCCAACAATCACCAACTACCCACACACACC		
450-nt-GSTA1-oligo 2-T	231	GTCACCGTCCTGGCTCGACAATCACCAACTACCCACACACA CC		Same as ~940-nt oligo 2-T
450-nt-HIST1H1B-oligo 2-T	232	TTAGAAGCAGCCACAGCCTTGGTGAATCACCAACTACCCACA CACACC		Same as ~940-nt oligo 2-T
450-nt-HMGA2-oligo 2-T	233	CAGAGGACAGAGTAGTGGGTGGCAATCACCAACTACCCACA CACACC		
450-nt-HOTAIR-oligo 2-T	234	GGGGCAAGGAGGCATAGGCCAATCACCAACTACCCACACAC ACC		Same as ~940-nt oligo 2-T
450-nt-HOXB2-oligo 2-T	235	GAGCGACGGCTGGCTGTTTATAAATCACCAACTACCCACAC ACACC		Same as ~940-nt oligo 2-T
450-nt-HSPA5-oligo 2-T	236	ACCAGGCGAAGGGCAGGTCTAATCACCAACTACCCACACAC ACC		Same as ~940-nt oligo 2-T
450-nt-INO80-oligo 2-T	237	GCGGCTCCTGGGACCCCAATCACCAACTACCCACACACACC		
450-nt-KLF8-oligo 2-T	238	CCCCACGCGAGTCAGAACCTAAATCACCAACTACCCACACA CACC		
450-nt-LIG1-oligo 2-T	239	CACTGGGAGTCGTAGTCTCCCGAAATCACCAACTACCCACAC ACACC		Same as ~940-nt oligo 2-T
450-nt-MBD4-oligo 2-T	240	GCCATTTCCACGCTCCACCAATCAAAATCACCAACTACCCAC ACACACC		Same as ~940-nt oligo 2-T
450-nt-MCM9-oligo 2-T	241	GATTTCCATGTTGGTCTCAAACAGAGTCAATCACCAACTACC CACACACACC		Same as ~940-nt oligo 2-T
450-nt-MECP2-oligo 2-T	242	GCGGCGGCCATTTTCCGGAATCACCAACTACCCACACACACC		Same as ~940-nt oligo 2-T
450-nt-MEG3-oligo 2-T	243	CAGAGCCCATCCAGAAAGCGGAATCACCAACTACCCACACA CACC		
450-nt-MGMT-oligo 2-T	244	GAGGCTGCCACCGTCCCGAATCACCAACTACCCACACACACC		
450-nt-MLH1-oligo 2-T	245	GCATACATAATTTTAAATTAAGTGGCTTCCTTAATCACCAACT ACCCACACACACC		
450-nt-MLH3-oligo 2-T	246	GAGTCCCGCTTAACAAATTCCTCGAATCACCAACTACCCACA CACACC		
450-nt-MNAT1-oligo 2-T	247	GGGGTTCCGATATTTGGTGGTCTTAATCACCAACTACCCACA CACACC		
450-nt-MPG-oligo 2-T	248	CCAGCAGCCGCCGACCAATCACCAACTACCCACACACACC		
450-nt-MRE11-oligo 2-T	249	GAAGCCCCACAATCGCGACCCAATCACCAACTACCCACACA CACC		Same as ~940-nt oligo 2-T
450-nt-MSH2-oligo 2-T	250	GCACCGCCATGTGAAACCTCCTCAATCACCAACTACCCACA CACACC		Same as ~940-nt oligo 2-T
450-nt-MSH3-oligo 2-T	251	GCTTGCTCGCAGGGGCTGAATCACCAACTACCCACACACAC C		
450-nt-MSH4-oligo 2-T	252	GGGAAACCGCCGGGGCAATCACCAACTACCCACACACACC		Same as ~940-nt oligo 2-T
450-nt-MSH5-oligo 2-T	253	CTCCCCAGCTGCCGCCAATCACCAACTACCCACACACACC		Same as ~940-nt oligo 2-T
450-nt-MSH6-oligo 2-T	254	AGGCCCCAACGTTCTGTCCGAATCACCAACTACCCACACACA CC		
450-nt-MUTYH-oligo 2-T	255	GACGGTGAGCGGCTTCCCAATCACCAACTACCCACACACACC		
450-nt-NEIL2-oligo 2-T	256	ATGCGGCCCTCCGTCCCTTAATCACCAACTACCCACACACAC C		
450-nt-NEIL3-oligo 2-T	257	CGCGCGCGAATCTTCTCTCCAATCACCAACTACCCACACACA CC		
450-nt-NKX2-5-oligo 2-T	258	CCCCTGGCAGCTTCCTTGCATCACCAACTACCCACACACAC C		
450-nt-OGG1-oligo 2-T	259	TGCTTCGCCCTTTGCGGGTCAATCACCAACTACCCACACACA CC		Same as ~940-nt oligo 2-T
450-nt-OLIG2-oligo 2-T	260	GAGGAGAGTCTGTCCCTGGGATCTAATCACCAACTACCCACA CACACC		Same as ~940-nt oligo 2-T
450-nt-PARP1-oligo 2-T	261	CCAGAGCCGCCACCGAACAATCACCAACTACCCACACACAC C		
450-nt-PCNA-oligo 2-T	262	ACCGCTTGGCAAAGCAAAGCAGAATCACCAACTACCCACAC ACACC		

450-nt-PELO-oligo 2-T	263	CGCGCTCTTGGTCTCTGAAATCTCAATCACCAACTACCCACACACACC		Same as ~940-nt oligo 2-T
450-nt-PMS1-oligo 2-T	264	TAGGGAACCTGCTGGCGGCCCTAATCACCAACTACCCACACACACACC		
450-nt-PMS2-oligo 2-T	265	GGGGACACCGGAAGACTGCGAATCACCAACTACCCACACACACACC		Same as ~940-nt oligo 2-T
450-nt-PMS2P11-oligo 2-T	266	CCACATGGAGAAAAACATACATTGTATCTCTCAAATTACCAAAAATCACCAACTACCCACACACACC		
450-nt-PMS2P1-oligo 2-T	267	GTGGCCTGGCGTGGATTCCGAATCACCAACTACCCACACACACACC		
450-nt-PMS2P2-oligo 2-T	268	CACGAGGCGGCCCTCCCAATCACCAACTACCCACACACACACC		
450-nt-PMS2P3-oligo 2-T	269	CGCGCGAACCCTGGTCTGCAATCACCAACTACCCACACACACACC		
450-nt-PMS2P5-oligo 2-T	270	CCTCGGCCATGTTCCCCCAATCACCAACTACCCACACACACACC		Same as ~940-nt oligo 2-T
450-nt-POLD1-oligo 2-T	271	GGCCAAAGGGTGGCCTCGCAATCACCAACTACCCACACACACACC		
450-nt-POLD4-oligo 2-T	272	GAGGCAACTGCGGGGAAGAGGAATCACCAACTACCCACACACACACC		Same as ~940-nt oligo 2-T
450-nt-POLG-oligo 2-T	273	ACCCAGGCCCCACGGCAATCACCAACTACCCACACACACACC		
450-nt-POLK-oligo 2-T	274	GGGATACTACCCGTTACCTCAGGAATCACCAACTACCCACACACACACC		
450-nt-POLQ-oligo 2-T	275	CGAGAACGAATCTGAGCCTGATTCTGAAATCACCAACTACCCACACACACC		
450-nt-POLR2B-oligo 2-T	276	AGGATCTGCGCCTGCGCAAATTGTAATCACCAACTACCCACACACACACC		
450-nt-POLR2G-oligo 2-T	277	GGCCTTGCTCTTCTTTCACCCCAATCACCAACTACCCACACACACACC		
450-nt-PRDM2-oligo 2-T	278	GCCTTCTAACAAGTGGTAACAGAGCTTCAATCACCAACTACCCACACACACC		
450-nt-PRKCG-oligo 2-T	279	AGCGGGCGAGTGGGGAGGAATCACCAACTACCCACACACACACACC		
450-nt-PRUNE2-oligo 2-T	280	CCTCCTCTGCTGCAACACCTTTTCCCTCTCCAATCACCAACTACCCACACACACC		Same as ~940-nt oligo 2-T
450-nt-PXMP4-oligo 2-T	281	CAGTGCCTGACGACTACGAGCAATCACCAACTACCCACACACACACC		
450-nt-PYCARD-oligo 2-T	282	CCGCTCACCCCGCTGCAATCACCAACTACCCACACACACACC		
450-nt-RNASEH2A-oligo 2-T	283	CCCGCCTCATCGACGCCCAATCACCAACTACCCACACACACACACC		
450-nt-RNF111-oligo 2-T	284	GCCTGGACTCGGCCAGCTAATCACCAACTACCCACACACACACACC		
450-nt-RPA2-oligo 2-T	285	GCCAATCGGTGCTCAOGGATGCTAATCACCAACTACCCACACACACACC		Same as ~940-nt oligo 2-T
450-nt-RPA3-oligo 2-T	286	CTTGACTATGTGTGTCTTCAAAAATAATACCTTTAAGTTAATCACTCAACTACCCACACACACACC		
450-nt-RPA4-oligo 2-T	287	GGCCACCTTCTCACCTGATCGCAATCACCAACTACCCACACACACACC		Same as ~940-nt oligo 2-T
450-nt-SH3TC1-oligo 2-T	288	CATACCTGCGGCGTGTCTCTCAATCACCAACTACCCACACACACACC		
450-nt-SIM2-oligo 2-T	289	GGAGCCCAGCGGAGGGGAATCACCAACTACCCACACACACACC		
450-nt-SMAD4-oligo 2-T	290	CTGGGCGGCGCGGGAATCACCAACTACCCACACACACACC		
450-nt-SMUG1-oligo 2-T	291	CTTACCCCATACGGAGGGAGGGGAATCACCAACTACCCACACACACACC		
450-nt-SNAI1-oligo 2-T	292	GCAGGCGCGCTCCCTCAATCACCAACTACCCACACACACACC		
450-nt-SNAI2-oligo 2-T	293	CGGATAACGGTCCGCGGGGAATCACCAACTACCCACACACACACACC		
450-nt-TAF1-oligo 2-T	294	AATGGGCGCGCTCCAGCGGGAATCACCAACTACCCACACACACACC		Same as ~940-nt oligo 2-T
450-nt-TDG-oligo 2-T	295	CTCACTGCGGTAAGACAACCCCAATCACCAACTACCCACACACACACC		

450-nt-TP53-oligo 2-T	296	GGAAGCGTGTACCGTCTGGGAAATCACCAACTACCCACACACACC		Same as ~940-nt oligo 2-T
450-nt-TREX1-oligo 2-T	297	CCCTCCCCTAGTGGCCCCCTTAATCACCAACTACCCACACACACC		
450-nt-TRIM28-oligo 2-T	298	CCGCCGAAACCCGACCAATCACCAACTACCCACACACACACC		
450-nt-TWIST1-oligo 2-T	299	GCGGGAGGGGGAGGGGAAATCACCAACTACCCACACACACACC		
450-nt-TWIST2-oligo 2-T	300	ACACGCTCTCGGAGTTGCTGGAAATCACCAACTACCCACACACACC		
450-nt-UBB-oligo 2-T	301	AGTTTAGGCTTTCAACTGAGCCCCAAATCACCAACTACCCACACACACC		Same as ~940-nt oligo 2-T
450-nt-USP47-oligo 2-T	302	CCTGGACGTCATCCCCTCTTCAATCACCAACTACCCACACACACC		
450-nt-WDR48-oligo 2-T	303	TGGCGCCGGGACCTTCTAATCACCAACTACCCACACACACACC		
450-nt-WNT1-oligo 2-T	304	CTCTGTGCTGCAGTTGCGGCGAATCACCAACTACCCACACACACC		
450-nt-XPC-oligo 2-T	305	GGGCCTTGCTCTTGGCCTTGGAAATCACCAACTACCCACACACACC		
450-nt-XRCC2-oligo 2-T	306	TCCCAACCCGGCGGCTCTCAATCACCAACTACCCACACACACACC		Same as ~940-nt oligo 2-T
450-nt-XRCC3-oligo 2-T	307	CGCCCCCGCCCCCTTGAATCACCAACTACCCACACACACACC		
450-nt-ZEB1-oligo 2-T	308	CGGGTTCGCTGCTTTCTGCGCTTAATCACCAACTACCCACACACACC		
450-nt-ZNF830-oligo 2-T	309	CGGTTGTACTTCGCGAATGGAGATTCTAATCACCAACTACCCACACACACC		Same as ~940-nt oligo 2-T
450-nt-ALKBH1-oligo 3-T	310	CTACCCACCTTCCTCATTTCTCTAGAAGTCGATGACCCCTTCAGGT	1. All oligos were designed to capture target sequences of ~450 nt in length.	
450-nt-ALKBH2-oligo 3-T	311	CTACCCACCTTCCTCATTTCTCTACGGGTAAACCGCACGCAA AATTCTG		Same as ~940-nt oligo 3-T
450-nt-ALKBH3-oligo 3-T	312	CTACCCACCTTCCTCATTTCTCTAGCCCCGCGCACCCA		
450-nt-APEX1-oligo 3-T	313	CTACCCACCTTCCTCATTTCTCTATCTGCACCGGCCGATG	2. Non-bolded nucleotides are complementary to the 5' end of target sequence.	Same as ~940-nt oligo 3-T
450-nt-APEX2-oligo 3-T	314	CTACCCACCTTCCTCATTTCTCTATTTGCCACCCCTTGCAAGGGTCT		Same as ~940-nt oligo 3-T
450-nt-ATM-oligo 3-T	315	CTACCCACCTTCCTCATTTCTCTACGGTATGCCCATGCGCATCCAGT		Same as ~940-nt oligo 3-T
450-nt-ATRX-oligo 3-T	316	CTACCCACCTTCCTCATTTCTCTACCTCCCCACAGCTCAAAGGC	3. Nucleotide in red is complementary to nucleotide immediately 5' of FEN cleavage site.	Same as ~940-nt oligo 3-T
450-nt-ATXN3-oligo 3-T	317	CTACCCACCTTCCTCATTTCTCTATCGGCATGGGGGCGACTCG		
450-nt-AXIN2-oligo 3-T	318	CTACCCACCTTCCTCATTTCTCTAGCCGGCTCCAGCCGACT		
450-nt-BRCA1-oligo 3-T	319	CTACCCACCTTCCTCATTTCTCTACAGCCTGTCCCCGTCCA	4. Bolded nucleotides are complementary	Same as ~940-nt oligo 3-T
450-nt-BRCA2-oligo 3-T	320	CTACCCACCTTCCTCATTTCTCTAGCTGACAAAAACCGCGCCGG		Same as ~940-nt oligo 3-T

			entary to oligo U2.	
450-nt-CCN4-oligo 3-T	321	CTACCCACCTTCCTCATTTCTCTAGGACGGTGGCGGGCT		Same as ~940-nt oligo 3-T
450-nt-CD44-oligo 3-T	322	CTACCCACCTTCCTCATTTCTCTAGAGGTTTFAAGAAGTAGC AGCCCTCCC		Same as ~940-nt oligo 3-T
450-nt-CDH13- oligo 3-T	323	CTACCCACCTTCCTCATTTCTCTATTGGGTAGAGGCTGATGA CCAGGACCAATG		
450-nt-CDH1- oligo 3-T	324	CTACCCACCTTCCTCATTTCTCTAGTCTGAACTGACTTCCGC AAGCTCAC		
450-nt-CDH2- oligo 3-T	325	CTACCCACCTTCCTCATTTCTCTAGGGGGCGGGGGAACAGG		
450-nt-CDKN2A- oligo 3-T	326	CTACCCACCTTCCTCATTTCTCTAGGGCCGAGCTCGGCAGC		
450-nt-CDKN2D- oligo 3-T	327	CTACCCACCTTCCTCATTTCTCTAGCCGGGCCCAACCCCT		
450-nt-CETN1- oligo 3-T	328	CTACCCACCTTCCTCATTTCTCTATCCTCAGTGAGCTCGGGC TTA		Same as ~940-nt oligo 3-T
450-nt-CHEK1- oligo 3-T	329	CTACCCACCTTCCTCATTTCTCTAGTCAGGATAAACAGTTTT GCCTTTTTCTGTG		Same as ~940-nt oligo 3-T
450-nt-CHEK2- oligo 3-T	330	CTACCCACCTTCCTCATTTCTCTAGACAAAGCGAAGCTCAGG AGACTCC		Same as ~940-nt oligo 3-T
450-nt-CSDE1- oligo 3-T	331	CTACCCACCTTCCTCATTTCTCTAGCCCTTCCCTGGGTTCT GCTA		Same as ~940-nt oligo 3-T
450-nt-DDB2- oligo 3-T	332	CTACCCACCTTCCTCATTTCTCTATGGAGGGAGGATTGAGAG GCCTC		Same as ~940-nt oligo 3-T
450-nt-DRD3- oligo 3-T	333	CTACCCACCTTCCTCATTTCTCTATTTATTTGGCTCACAGTTC TGGAGTC		
450-nt-DRD4- oligo 3-T	334	CTACCCACCTTCCTCATTTCTCTAGCACCGCGCCGATGAGCA		
450-nt-EGFR-oligo 3-T	335	CTACCCACCTTCCTCATTTCTCTACACCGTGCGGGGGGC		
450-nt- EPM2AIP1-oligo 3-T	336	CTACCCACCTTCCTCATTTCTCTAGCGGCTACTGCCCGCTAC CTAGAA		
450-nt-EPSTI1- oligo 3-T	337	CTACCCACCTTCCTCATTTCTCTATTGCGGGTGTTCATGGTTC ACAGC		Same as ~940-nt oligo 3-T
450-nt-ERCC1- oligo 3-T	338	CTACCCACCTTCCTCATTTCTCTAGTCTCCGCTCCCGCACCT		
450-nt-ERCC2- oligo 3-T	339	CTACCCACCTTCCTCATTTCTCTACCGATGACCCCATCTCA AGACCC		
450-nt-ERCC8- oligo 3-T	340	CTACCCACCTTCCTCATTTCTCTACCTCCGTGTTGACTCTGCT CT		
450-nt-EXO1-oligo 3-T	341	CTACCCACCTTCCTCATTTCTCTAGCCGATTGCGGTTGGAGA AACC		Same as ~940-nt oligo 3-T
450-nt-FANCC- oligo 3-T	342	CTACCCACCTTCCTCATTTCTCTAGCAGTTTGGGCAGTGGCG GAAA		
450-nt-FAT4-oligo 3-T	343	CTACCCACCTTCCTCATTTCTCTAGCTTGGAAACCTGGCGC G		
450-nt-FEN1-oligo 3-T	344	CTACCCACCTTCCTCATTTCTCTACAGAGTCTCTCAGCGACC CTG		Same as ~940-nt oligo 3-T
450-nt-FHIT-oligo 3-T	345	CTACCCACCTTCCTCATTTCTCTAGGATGTTGACAGCTGGGA ATGAAAGGCAG		
450-nt-FOX2- oligo 3-T	346	CTACCCACCTTCCTCATTTCTCTAGTAATTCTGCTCGCTCAG GTAGGG		
450-nt-FZD7-oligo 3-T	347	CTACCCACCTTCCTCATTTCTCTAGGCCAGGGACGAAAGC GGA		
450-nt-FZD8-oligo 3-T	348	CTACCCACCTTCCTCATTTCTCTAGCGGCCAGAACTGGTGCA CCT		
450-nt-GPR158- oligo 3-T	349	CTACCCACCTTCCTCATTTCTCTAGCTCCCAATCCCAGCTGA GCAA		
450-nt-GSTA1- oligo 3-T	350	CTACCCACCTTCCTCATTTCTCTACTGAATCCAGGTCCTAA TGTATTATAAGC		Same as ~940-nt oligo 3-T
450-nt-HIST1H1B- oligo 3-T	351	CTACCCACCTTCCTCATTTCTCTATCAGCTCTGAGACTGGGG GC		Same as ~940-nt oligo 3-T
450-nt-HMGA2- oligo 3-T	352	CTACCCACCTTCCTCATTTCTCTACCGCGCCTCTAGGGTGG		

450-nt-HOTAIR-oligo 3-T	353	CTACCCACCTTCCTCATTCTCTACAACACTCACCCACGGA GCAG		Same as ~940-nt oligo 3-T
450-nt-HOXB2-oligo 3-T	354	CTACCCACCTTCCTCATTCTCTACCCAATCTCCCTCTCAAA TCAAAAATTCATGGC		Same as ~940-nt oligo 3-T
450-nt-HSPA5-oligo 3-T	355	CTACCCACCTTCCTCATTCTCTAGAAATACAGGCCGCGCG CT		Same as ~940-nt oligo 3-T
450-nt-INO80-oligo 3-T	356	CTACCCACCTTCCTCATTCTCTAGATGGCGGCGGCGCTGA		
450-nt-KLF8-oligo 3-T	357	CTACCCACCTTCCTCATTCTCTAGTGCTCTCTGGGCAACCT GAG		
450-nt-LIG1-oligo 3-T	358	CTACCCACCTTCCTCATTCTCTATGGGAGGAGGGCGGGAA AAG		Same as ~940-nt oligo 3-T
450-nt-MBD4-oligo 3-T	359	CTACCCACCTTCCTCATTCTCTATCCATTCTCGAGGAAGAC GCACCG		Same as ~940-nt oligo 3-T
450-nt-MCM9-oligo 3-T	360	CTACCCACCTTCCTCATTCTCTATGGCAATTAACCACAACCTG GGTAATGA		Same as ~940-nt oligo 3-T
450-nt-MECP2-oligo 3-T	361	CTACCCACCTTCCTCATTCTCTACGGCTTTTACCACAGCCC TCTCT		Same as ~940-nt oligo 3-T
450-nt-MEG3-oligo 3-T	362	CTACCCACCTTCCTCATTCTCTATGCCGCTTCTCTTTGCAC ATCCTT		
450-nt-MGMT-oligo 3-T	363	CTACCCACCTTCCTCATTCTCTAGGGAGAGCTCCGCACTCT TC		
450-nt-MLH1-oligo 3-T	364	CTACCCACCTTCCTCATTCTCTACTTAGTTAACGGTACCAT TAAGACAAC		
450-nt-MLH3-oligo 3-T	365	CTACCCACCTTCCTCATTCTCTACACCAACCGCTCGGACG C		
450-nt-MNAT1-oligo 3-T	366	CTACCCACCTTCCTCATTCTCTACACCGAGGGCAACCCCTGA TC		
450-nt-MPG-oligo 3-T	367	CTACCCACCTTCCTCATTCTCTACGGAAGCACCCCTAGGAC C		
450-nt-MRE11-oligo 3-T	368	CTACCCACCTTCCTCATTCTCTAGGCTGTCTCTTTTCGGGA AGAAAGG		Same as ~940-nt oligo 3-T
450-nt-MSH2-oligo 3-T	369	CTACCCACCTTCCTCATTCTCTACCTCCTGGTTGAAGAAAA TGCGCGACC		Same as ~940-nt oligo 3-T
450-nt-MSH3-oligo 3-T	370	CTACCCACCTTCCTCATTCTCTAGCTGGAGGCAGCGAGGCC		
450-nt-MSH4-oligo 3-T	371	CTACCCACCTTCCTCATTCTCTAGAAGGCGAGGTTGATGAG ATCTCA		Same as ~940-nt oligo 3-T
450-nt-MSH5-oligo 3-T	372	CTACCCACCTTCCTCATTCTCTATCGCAGCGTTTCCCGCCT TTTC		Same as ~940-nt oligo 3-T
450-nt-MSH6-oligo 3-T	373	CTACCCACCTTCCTCATTCTCTACGGAGCTCCTAAAAGCAC CGCAT		
450-nt-MUTYH-oligo 3-T	374	CTACCCACCTTCCTCATTCTCTAGAGGTAGCCTTCAAAGCC TCTGCG		
450-nt-NEIL2-oligo 3-T	375	CTACCCACCTTCCTCATTCTCTAGGGGACGCCCTGTAGGT		
450-nt-NEIL3-oligo 3-T	376	CTACCCACCTTCCTCATTCTCTATTACAGGTACAGCCTGGT CCTT		
450-nt-NKX2-5-oligo 3-T	377	CTACCCACCTTCCTCATTCTCTATGGTGCCGCCGCCCG		
450-nt-OGG1-oligo 3-T	378	CTACCCACCTTCCTCATTCTCTATCAAAGCCCCGCTTCGTG CCCCAA		Same as ~940-nt oligo 3-T
450-nt-OLIG2-oligo 3-T	379	CTACCCACCTTCCTCATTCTCTAGGCTGCGGCTGCTGTGG		Same as ~940-nt oligo 3-T
450-nt-PARP1-oligo 3-T	380	CTACCCACCTTCCTCATTCTCTACGCCGCACCCGCCA		
450-nt-PCNA-oligo 3-T	381	CTACCCACCTTCCTCATTCTCTACTGAAGCGTGGGTTTCAT TTCTCATCATGTG		
450-nt-PELO-oligo 3-T	382	CTACCCACCTTCCTCATTCTCTACTGGCAGCTTCCAGGCT CTG		Same as ~940-nt oligo 3-T
450-nt-PMS1-oligo 3-T	383	CTACCCACCTTCCTCATTCTCTAGGGGTCGTCTGAGAGGG AAA		
450-nt-PMS2-oligo 3-T	384	CTACCCACCTTCCTCATTCTCTAGCCCCGCTCACCTCGAGC		Same as ~940-nt oligo 3-T
450-nt-PMS2P11-oligo 3-T	385	CTACCCACCTTCCTCATTCTCTACCTTTGGCAACAATGGTG TCTTCTTC		
450-nt-PMS2P1-oligo 3-T	386	CTACCCACCTTCCTCATTCTCTAGCGTTGGTGTCTGGCGGT TTC		

450-nt-PMS2P2-oligo 3-T	387	CTACCCACCTTCCTCATTCTCTACGCGCCTCGGCCATGTTC		
450-nt-PMS2P3-oligo 3-T	388	CTACCCACCTTCCTCATTCTCTACAGCTCCGCGTCCGCTAG		
450-nt-PMS2P5-oligo 3-T	389	CTACCCACCTTCCTCATTCTCTATTTCCAGGGAGGTTGGAA GCCGTG		Same as ~940-nt oligo 3-T
450-nt-POLD1-oligo 3-T	390	CTACCCACCTTCCTCATTCTCTACCGTCTCTCAGATCTATGT GCCTGGCT		
450-nt-POLD4-oligo 3-T	391	CTACCCACCTTCCTCATTCTCTAGAGACCGGCCAGTGGGC		Same as ~940-nt oligo 3-T
450-nt-POLG-oligo 3-T	392	CTACCCACCTTCCTCATTCTCTACGGCGTCCCCCTTCGC		
450-nt-POLK-oligo 3-T	393	CTACCCACCTTCCTCATTCTCTATCGCGACTACAACCTCCCA GGA		
450-nt-POLQ-oligo 3-T	394	CTACCCACCTTCCTCATTCTCTACGCCGCCGTTCCCACT		
450-nt-POLR2B-oligo 3-T	395	CTACCCACCTTCCTCATTCTCTAGGTCAAGCTTGGGCCGGG GTT		
450-nt-POLR2G-oligo 3-T	396	CTACCCACCTTCCTCATTCTCTAGCCCTTGCCGCCACC		
450-nt-PRDM2-oligo 3-T	397	CTACCCACCTTCCTCATTCTCTACAATTAAGAATCTAAATT TATCCCAAGAGCCA		
450-nt-PRKCG-oligo 3-T	398	CTACCCACCTTCCTCATTCTCTAGAGCCAAAGGCAGGGGC G		
450-nt-PRUNE2-oligo 3-T	399	CTACCCACCTTCCTCATTCTCTACCTTCCATTAGCATACGC GCGCG		Same as ~940-nt oligo 3-T
450-nt-PXMP4-oligo 3-T	400	CTACCCACCTTCCTCATTCTCTAGAGCCCTTAGCTGCGGC		
450-nt-PYCARD-oligo 3-T	401	CTACCCACCTTCCTCATTCTCTAGCCGCCGACCAGGAGGAA GTC		
450-nt-RNASEH2A-oligo 3-T	402	CTACCCACCTTCCTCATTCTCTAGGACGCAAGGCTCCTTGC G		
450-nt-RNF111-oligo 3-T	403	CTACCCACCTTCCTCATTCTCTAGCCGGCGTGGGTTGAGAA TGCCTAATTT		
450-nt-RPA2-oligo 3-T	404	CTACCCACCTTCCTCATTCTCTACGCTTGTTCCTGTCTCCTC TGC		Same as ~940-nt oligo 3-T
450-nt-RPA3-oligo 3-T	405	CTACCCACCTTCCTCATTCTCTATTTTGAAGGAAAAAACTA TGTATCCAACAAATTCAAAC		
450-nt-RPA4-oligo 3-T	406	CTACCCACCTTCCTCATTCTCTATAATCCTTCCTGTGTGTC AGTGGTTGC		Same as ~940-nt oligo 3-T
450-nt-SH3TC1-oligo 3-T	407	CTACCCACCTTCCTCATTCTCTAGAGGCAGCTCCAGGCC		
450-nt-SIM2-oligo 3-T	408	CTACCCACCTTCCTCATTCTCTAGCGAGTCGCAGGCGGCTG		
450-nt-SMAD4-oligo 3-T	409	CTACCCACCTTCCTCATTCTCTAGCCTAGGGGAGAGCAGG AAG		
450-nt-SMUG1-oligo 3-T	410	CTACCCACCTTCCTCATTCTCTACCCCAAAATATAGAAATT CTAACCCCTCTCGGC		
450-nt-SNAI1-oligo 3-T	411	CTACCCACCTTCCTCATTCTCTAGGTGCCCCCACGCAG		
450-nt-SNAI2-oligo 3-T	412	CTACCCACCTTCCTCATTCTCTAGGACACGGCGGTCCCTAC AGCAT		
450-nt-TAF1-oligo 3-T	413	CTACCCACCTTCCTCATTCTCTATCTTCGTGCTGTCCGTGT CTGACA		Same as ~940-nt oligo 3-T
450-nt-TDG-oligo 3-T	414	CTACCCACCTTCCTCATTCTCTAGAGACCTGGGCCCGATGG CT		
450-nt-TP53-oligo 3-T	415	CTACCCACCTTCCTCATTCTCTAGCACGCTCCAGCCCGAA C		Same as 950-nt- oligo 3-T
450-nt-TREX1-oligo 3-T	416	CTACCCACCTTCCTCATTCTCTAGTCTGGGGTGGGGGAACC CA		
450-nt-TRIM28-oligo 3-T	417	CTACCCACCTTCCTCATTCTCTACGCGCCACAAGCACAAGC ACAA		
450-nt-TWIST1-oligo 3-T	418	CTACCCACCTTCCTCATTCTCTACGGTGTGGATGGCCCCGA		
450-nt-TWIST2-oligo 3-T	419	CTACCCACCTTCCTCATTCTCTAGGCTCTGATTTCAAGGCC GGCTTGT		

450-nt-UBB-oligo 3-T	420	CTACCCACCTTCCTCATTCTCTATTCTCATAGCCGTAAGA AAGGCTCCT		Same as ~940-nt oligo 3-T
450-nt-USP47- oligo 3-T	421	CTACCCACCTTCCTCATTCTCTAGATACCCAGGCCTCGTC CT		
450-nt-WDR48- oligo 3-T	422	CTACCCACCTTCCTCATTCTCTAGCTGGAGTGGGCCCGGA		
450-nt-WNT1- oligo 3-T	423	CTACCCACCTTCCTCATTCTCTACTTTGGTTGTTGCCCGCGA CGG		
450-nt-XPC-oligo 3-T	424	CTACCCACCTTCCTCATTCTCTATTTCTGGCTGCGCAGITCG CGT		
450-nt-XRCC2- oligo 3-T	425	CTACCCACCTTCCTCATTCTCTACCTCGGTCCAGACTCAG C		Same as ~940-nt oligo 3-T
450-nt-XRCC3- oligo 3-T	426	CTACCCACCTTCCTCATTCTCTACCTGCCTCACCCCTCAC G		
450-nt-ZEB1-oligo 3-T	427	CTACCCACCTTCCTCATTCTCTACACCTGGGGCCATCCGCC AT		
450-nt-ZNF830- oligo 3-T	428	CTACCCACCTTCCTCATTCTCTATCCGTTCCGACTGGTGCT CAGAC		Same as ~940-nt oligo 3-T

Oligo name	SEQ ID NO	Oligo sequence	Notes	
940-nt-ALKBH2- oligo 1-T	429	CTCCCAAATGTTGGGATTACAGGCGTGAATCACCAACTA CCCACACACACC	1. All oligos were designed to capture target sequence s of ~940 nt in length.	
940-nt-APEX1-oligo 1-T	430	ACTGAGCGCATTTTATGTGCTAAGTATTCATCACCAACT ACCCACACACACC		
940-nt-APEX2-oligo 1-T	431	GAGACGTCGAATGAATTCATTTCATTCTCACAGTTCAATC ACCAACTACCCACACACACC		
940-nt-ATM-oligo 1- T	432	GAGAGACCTTCCTGGCGCCCTTCTAATAAATCACCAACTA CCCACACACACC	2. Non- bolded nucleotid es are complem entary to the 5' end of target sequence	
940-nt-ATRX-oligo1	433	CAGAGGAAGCTTGGCTGGATGATCTGTAAATCACCAAC TACCCACACACACC		
940-nt-BRCA1-oligo 1-T	434	TCCGTCGACGCAATCGCCACCAATCACCAACTACCCACA CACACC		
940-nt-BRCA2-oligo 1-T	435	ACCACGCACGGCTAACTTTGCAATCACCAACTACCCACA CACACC	3. Nucleoti de in red is complem entary to nucleotid e immedia tely 5' of FEN cleavage site.	
940-nt-CCN4-oligo 1-T	436	GCATTGCGCCCAACCCTACTTCTGGTTTTCATCACCAAC TACCCACACACACC		
940-nt-CD44-oligo 1- T	437	CCCATCCACCCACATATCTGTTTATTGATATATCCACTCAA TCACCAACTACCCACACACACC		
940-nt-CETN1-oligo 1-T	438	CCATGTGTTCAACTTGACATAAATGGGAAACAGAAATCACC AACTACCCACACACACC	4. Bolded nucleotid es are complem	
940-nt-CHEK1-oligo 1-T	439	AGAGGCTGCAAGTGAGCAGAGAAATCACCAACTACCCACAC ACACC		
940-nt-CHEK2-oligo 1-T	440	GAGGCTAGAGGGGAGGGAGATTGAAATCACCAACTACCC ACACACACC		
940-nt-CSDE1-oligo 1-T	441	CCCAGATTCAATGCTTCTGTTTTCACAGATTCAATCACCAA CTACCCACACACACC	4. Bolded nucleotid es are complem	
940-nt-DDB2-oligo 1-T	442	GCAGTGTCTCTCTTCCGTTACCCAATCACCAACTACCCAC ACACACC		

			entary to oligo U1.	
940-nt-EPM2AIP1- MLH1-oligo 1-T	443	ACGAATACTACGAGCGGTATGTGGCGGAATCACCAACTA CCCACACACACC		
940-nt-EPST11-oligo 1-T	444	GCTACCTGCATTTCTAGGAGTATCCTAGTAGACAATCACCA ACTACCCACACACACC		
940-nt-EXO1-oligo	445	GATAATGCAGACTCAATTCAAGGAATTGAAGGGGAATC ACCAACTACCCACACACACC		
940-nt-FEN1-oligo 1- T	446	GGCTGACCGGTGACAATGAAGTTAACAATTTGTAAATCACCA ACTACCCACACACACC		
940-nt-GSTA1-oligo 1-T	447	TGGAACACAGAGGCGAGAAATTAGAGTGAATCACCAACTA CCCACACACACC		
940-nt-HIST1H1B- oligo 1-T	448	GAGTGTTTTCTCTTTTGTCTCTGGGCAAATCCAATCACCC AACTACCCACACACACC		
940-nt-HOTAIR- oligo 1-T	449	TGGGGTGTCCCGGGTGCAAAATCACCAACTACCCACACAC ACC		
940-nt-HOXB2-oligo 1-T	450	CCTTATGTACCCCTGCTTTCTCTCACTTTCAATCACCAACTA CCCACACACACC		
940-nt-HSPA5-oligo 1-T	451	TTTACTTTCCCTCCGCTGCCCAATCACCAACTACCCACA CACACC		
940-nt-LIG1-oligo 1- T	452	GGAAGGAGGAGTTCAGATGAATACGAGAGTTAATCACCC AACTACCCACACACACC		
940-nt-MBD4-oligo 1-T	453	GAGATCTCGTTTCTCTAAAAGTGGAGATGGGCAATCACCA ACTACCCACACACACC		
940-nt-MCM9-oligo 1-T	454	GGTTCTAGCTTTTGTATTATGAACACAGTGTCAATCACCA ACTACCCACACACACC		
940-nt-MECP2-oligo 1-T	455	CTGTTAGCTAACCAGGTGTTCAGGACAGGAATCACCAACT ACCCACACACACC		
940-nt-MRE11-oligo 1-T	456	GTTTCTTAACTGTAAGTCGGGTAGATGCGCCTAATCACCA ACTACCCACACACACC		
940-nt-MSH2-oligo 1-T	457	AGGGCTGAGGCGAAAGGATCACCTGAATCACCAACTACC CACACACACC		
940-nt-MSH4-oligo 1-T	458	CGGTTTACCCACATGCAGTGACAACAATCACCAACTACC CACACACACC		
940-nt-MSH5-oligo 1-T	459	GCGGTGAACATAATGAAGTCACTGCTCCATAAAGCTTAAT CACCAACTACCCACACACACC		
940-nt-OGG1-oligo 1-T	460	CTGGGCCATTTCTACCTTTTAGCATAGGGAAATCACCAACT ACCCACACACACC		
940-nt-OLIG2-oligo 1-T	461	CTCTCCTTTAGGTTGAGGATGGTGTGGGTAAATCACCAACT ACCCACACACACC		
940-nt-PELO-oligo 1-T	462	GCGGTTTTGAGAATTTCAAAGACAGAAAAGAATAAGTGGAA TCACCAACTACCCACACACACC		
940-nt-PMS2-oligo 1-T	463	CCATTGCACTCGAGCCTGGGCAATCACCAACTACCCAC ACACACC		
940-nt-PMS2P5- oligo 1-T	464	GGCCAACATGGTGAAACCCCGTGTCTAATCACCAACTAC CCACACACACC		
940-nt-POLD4-oligo 1-T	465	CCATGTGCCAAGTACTGTTCTAAGCACTTTCCCAACAATCA CCAATACCCACACACACC		
940-nt-PRUNE2- oligo 1-T	466	TAGGTGCCCTTATGTACGCCCTGGGACTAAATCACCAACT ACCCACACACACC		
940-nt-RPA2-oligo 1- T	467	TGTCCAGGCTGCAGTGAGCTGTGAATCACCAACTACCCA CACACACC		
940-nt-RPA4-oligo 1- T	468	CAAGGCTCTGTGGACAGGACAGAAAAGTACAAGAATCAC CAACTACCCACACACACC		
940-nt-TAF1-oligo 1- T	469	CGGAAAATGTGAAAGTTTAGTAAAATGAACAGCTGACCCC TCAATCACCAACTACCCACACACACC		
940-nt-TP53-oligo 1- T	470	GCTTGATGGGAACGGGAAACCTTCTAACCTTTCAATCACCC AACTACCCACACACACC		
940-nt-UBB-oligo 1- T	471	CTTGAACCTAGGAGGACAGAGATTGCATCACCAACTACC CACACACACC		
940-nt-XRCC2-oligo 1-T	472	CCCACCTAGTTTGTGTTTCAAAAACGCTAGGAAAAGAGCAA TCACCAACTACCCACACACACC		
940-nt-ZNF830-oligo 1-T	473	GCACCTTTTCTATAGACAAGAGTGAGGAAAAGAGAATCA CCAATACCCACACACACC		
940-nt-ALKBH2- oligo 2-T	474	CGATTGAGAAGCGAGACGCAAGAGGTTAAACAATCACCC AACTACCCACACACACC		Same as ~450-nt oligo 2-T

940-nt-APEX1-oligo 2-T	475	CCTCTTATGAGCAAAAGAGCAACCCCGTAATCACCAACTA CCCACACACACC	Same as ~450-nt oligo 2-T
940-nt-APEX2-oligo 2-T	476	CGGCACAGTTGCTGGGTTCCTGAATCACCAACTACCCAC ACACACC	Same as ~450-nt oligo 2-T
940-nt-ATM-oligo 2-T	477	GCCAAGCAGCCGCAGAGCAATCACCAACTACCCACACAC ACC	Same as ~450-nt oligo 2-T
940-nt-ATRX-oligo 2-T	478	CGTCACTGTAGCTGCTGCTGGAAATCACCAACTACCCAC ACACACC	Same as ~450-nt oligo 2-T
940-nt-BRCA1-oligo 2-T	479	GGGCCCAGTTATCTGAGAAACCCCAATCACCAACTACCC ACACACACC	Same as ~450-nt oligo 2-T
940-nt-BRCA2-oligo 2-T	480	GGTGCAGTTCCTTTTGGCCGGAGTAAATCACCAACTACC CACACACACC	Same as ~450-nt oligo 2-T
940-nt-CCN4-oligo 2-T	481	AGGCAGGACTCACCGTGGCCAAATCACCAACTACCCACAC ACACC	Same as ~450-nt oligo 2-T
940-nt-CD44-oligo 2-T	482	GGCTGTGACTAAGCAGCCCGCAATCACCAACTACCCACA CACACC	Same as ~450-nt oligo 2-T
940-nt-CETN1-oligo 2-T	483	AAATGCTTCCCAGAACTTCTTGCTTCTGAATCACCAACTAC CCACACACACC	Same as ~450-nt oligo 2-T
940-nt-CHEK1-oligo 2-T	484	GAAAGCGAGCAGTTTATGTTGGAGCTTGAATCACCAACT ACCCACACACACC	Same as ~450-nt oligo 2-T
940-nt-CHEK2-oligo 2-T	485	CACACACTCCCCCTCAGCTCAAAACTACAATCACCAACT ACCCACACACACC	Same as ~450-nt oligo 2-T
940-nt-CSDE1-oligo 2-T	486	GGCTTCTCGTTAGCCTTGGTCTTATTCTCAATCACCAACTA CCCACACACACC	Same as ~450-nt oligo 2-T
940-nt-DDB2-oligo 2-T	487	GGTACTGTGCTCTATGCGAAGATCAATCACCAACTACCC ACACACACC	Same as ~450-nt oligo 2-T
940-nt-EPM2AIP1- MLH1-oligo 2-T	488	AACCGGGCTTGTGTGCTCTGCTGAATCACCAACTACCCA CACACACC	Same as ~300-nt MLH1 oligo 2-T
940-nt-EPSTI1-oligo 2-T	489	CGAGCCCGGAGTTCACCACTCTAATCACCAACTACCCAC ACACACC	Same as ~450-nt oligo 2-T
940-nt-EXO1-oligo 2-T	490	CTCTCTCTCCCTTGAGCGGAATCACCAACTACCCACACA CACC	Same as ~450-nt oligo 2-T
940-nt-FEN1-oligo 2-T	491	GCAACCAGTCCCTCCAGGGAGAAATCACCAACTACCCAC ACACACC	Same as ~450-nt oligo 2-T
940-nt-GSTA1-oligo 2-T	492	GTCACCGTCTGCTCGACAATCACCAACTACCCACAC ACACC	Same as ~450-nt oligo 2-T
940-nt-HIST1H1B- oligo 2-T	493	TTAGAAGCAGCCACAGCCTTGGTGAATCACCAACTACCC ACACACACC	Same as ~450-nt oligo 2-T
940-nt-HOTAIR- oligo 2-T	494	GGGGCAAGGAGGCATAGGCCAATCACCAACTACCCACAC ACACC	Same as ~450-nt oligo 2-T
940-nt-HOXB2-oligo 2-T	495	GAGCGACGGCTGGCTGTTTATAAATCACCAACTACCCA CACACACC	Same as ~450-nt oligo 2-T
940-nt-HSPA5-oligo 2-T	496	ACCAGGCGAAGGGCAGGTCTAATCACCAACTACCCACAC ACACC	Same as ~450-nt oligo 2-T
940-nt-LIG1-oligo 2-T	497	CACTGGGAGTCGTAGTCTCCCGAAATCACCAACTACCCA CACACACC	Same as ~450-nt oligo 2-T
940-nt-MBD4-oligo 2-T	498	GCCATTTCCACGCTCCACCAATCAAAATCACCAACTACCC ACACACACC	Same as ~450-nt oligo 2-T
940-nt-MCM9-oligo 2-T	499	GATTTCCATGTTGGTCTCAAACAGAGTCAATCACCAACTA CCCACACACACC	Same as ~450-nt oligo 2-T
940-nt-MECP2-oligo 2-T	500	GCGGCGGCCATTTCCGGAAATCACCAACTACCCACACAC ACC	Same as ~450-nt oligo 2-T
940-nt-MRE11-oligo 2-T	501	GAAGCCCCACAATCGCGACCCAAATCACCAACTACCCAC ACACACC	Same as ~450-nt oligo 2-T
940-nt-MSH2-oligo 2-T	502	GCACCGCCATGTGAAACCTCTCAATCACCAACTACCCA CACACACC	Same as ~450-nt oligo 2-T
940-nt-MSH4-oligo 2-T	503	GGGAAACCGCCGGGGCAATCACCAACTACCCACACACA CC	Same as ~450-nt oligo 2-T
940-nt-MSH5-oligo 2-T	504	CTCCCCAGCTGCGGCCAATCACCAACTACCCACACACA CC	Same as ~450-nt oligo 2-T
940-nt-OGG1-oligo 2-T	505	TGCCTCGCCCTTTCGGGTCAATCACCAACTACCCACACA CACC	Same as ~450-nt oligo 2-T
940-nt-OLIG2-oligo 2-T	506	GAGGAGAGTCTGTCCCTGGGATCTAATCACCAACTACCC ACACACACC	Same as ~450-nt oligo 2-T
940-nt-PELO-oligo 2-T	507	CGCGTCTTGGTCTCTGAAATCTCAATCACCAACTACCCA CACACACC	Same as ~450-nt oligo 2-T
940-nt-PMS2-oligo 2-T	508	GGGACACCGGAAGACTGCGAATCACCAACTACCCACAC ACACC	Same as ~450-nt oligo 2-T

940-nt-PMS2P5-oligo 2-T	509	CCTCGGCCATGTTCCCCCAATCACC AACT ACCCACACACACC		Same as ~450-nt oligo 2-T
940-nt-POLD4-oligo 2-T	510	GAGGCAACTGCGGGGAAGAGGAATCACC AACT ACCCACACACACC		Same as ~450-nt oligo 2-T
940-nt-PRUNE2-oligo 2-T	511	CCTCCTCTGCTGCAACACCTTTTCCTCTCC AACT ACCCAACTACCCACACACACC		Same as ~450-nt oligo 2-T
940-nt-RPA2-oligo 2-T	512	GCCAATCGGTGCTCACGGATGCTAATCACC AACT ACCCAACACACACC		Same as ~450-nt oligo 2-T
940-nt-RPA4-oligo 2-T	513	GGCCACCTTCTCACCTGATCGCAATCACC AACT ACCCACACACACC		Same as ~450-nt oligo 2-T
940-nt-TAF1-oligo 2-T	514	AATGGGCGCCTCCAGCGGA AACT ACCCAACTACCCACACACACC		Same as ~450-nt oligo 2-T
940-nt-TP53-oligo 2-T	515	GGAAGCGTGTCACCGTCGTGGAA AACT ACCCAACTACCCAACACACACC		Same as ~450-nt oligo 2-T
940-nt-UBB-oligo 2-T	516	AGTTTAGGCTTTTCAACTGAGCCCCAA AACT ACCCAACTACCCACACACACC		Same as ~450-nt oligo 2-T
940-nt-XRCC2-oligo 2-T	517	TCCCAACCCGGCGGCTCTCAATCACC AACT ACCCACACACACC		Same as ~450-nt oligo 2-T
940-nt-ZNF830-oligo 2-T	518	CGGTTGTACTTCGGAATGGAGATTCTAATCACC AACT ACCCACACACACC		Same as ~450-nt oligo 2-T
940-nt-ALKBH2-oligo 3-T	519	CTACCC CACTT CTCTCATTTCTCTACGGGTAAACCGCACGCAAAATTCTG	1. All oligos were designed to capture target sequences of ~940 nt in length	Same as ~450-nt oligo 3-T
940-nt-APEX1-oligo 3-T	520	CTACCC CACTT CTCTCATTTCTCTATCTGCACCGGCCCGATG		Same as ~450-nt oligo 3-T
940-nt-APEX2-oligo 3-T	521	CTACCC CACTT CTCTCATTTCTCTATTTGCCACCCCTTGCAGGGTCT		Same as ~450-nt oligo 3-T
940-nt-ATM-oligo 3-T	522	CTACCC CACTT CTCTCATTTCTCTACGGGTATGCCCATGCGCATCCAGT		Same as ~450-nt oligo 3-T
940-nt-ATRX-oligo 3-T	523	CTACCC CACTT CTCTCATTTCTCTACCTCC CACT ACAGCTCAAGGC	2. Non-bolded nucleotides are complementary to the 5' end of target sequence	Same as ~450-nt oligo 3-T
940-nt-BRCA1-oligo 3-T	524	CTACCC CACTT CTCTCATTTCTCTACAGCCTGTCCCGCTCA		Same as ~450-nt oligo 3-T
940-nt-BRCA2-oligo 3-T	525	CTACCC CACTT CTCTCATTTCTCTAGCTGACAAAAACCGCGCCGG		Same as ~450-nt oligo 3-T
940-nt-CCN4-oligo 3-T	526	CTACCC CACTT CTCTCATTTCTCTAGGACGGTGCTGGCGGCT		Same as ~450-nt oligo 3-T
940-nt-CD44-oligo 3-T	527	CTACCC CACTT CTCTCATTTCTCTAGAGGTTTAAAGAAGTAGCAGCCCTCCC	3. Nucleotide in red is complementary to nucleotide immediately 5' of FEN cleavage site.	Same as ~450-nt oligo 3-T
940-nt-CETN1-oligo 3-T	528	CTACCC CACTT CTCTCATTTCTCTATCCTCAGTGAGCTCGGGCTTA		Same as ~450-nt oligo 3-T
940-nt-CHEK1-oligo 3-T	529	CTACCC CACTT CTCTCATTTCTCTAGTCAGGATAACCAGTTTGGCTTTTTCTGTG		Same as ~450-nt oligo 3-T
940-nt-CHEK2-oligo 3-T	530	CTACCC CACTT CTCTCATTTCTCTAGACAAAGCGAAGCTCAGGAGACTCC		Same as ~450-nt oligo 3-T
940-nt-CSDE1-oligo 3-T	531	CTACCC CACTT CTCTCATTTCTCTAGCCCTTCCCCTGGGTTCTGCTA	4. Bolded nucleotides are complementary to	Same as ~450-nt oligo 3-T
940-nt-DDB2-oligo 3-T	532	CTACCC CACTT CTCTCATTTCTCTATGGAGGGAGGATTGAGAGGCCTC		Same as ~450-nt oligo 3-T

			oligo U2.	
940-nt-EPM2AIP1- MLH1-oligo 3-T	533	CTACCCACCTTCCTCATTCCTCTAGGTGATCTGGCGCAG AGCGGAGG		Same as ~300-nt MLH1 oligo 3-T
940-nt-EPSTI1-oligo 3-T	534	CTACCCACCTTCCTCATTCCTCTATTGCGGGTGTTCATGG TTCACAGC		Same as ~450-nt oligo 3-T
940-nt-EXO1-oligo 3-T	535	CTACCCACCTTCCTCATTCCTCTAGCCGATTGCGGTTGG AGAAACC		Same as ~450-nt oligo 3-T
940-nt-FEN1-oligo 2- T	536	CTACCCACCTTCCTCATTCCTCTACAGAGTCTCTCAGCG ACCCTG		Same as ~450-nt oligo 3-T
940-nt-GSTA1-oligo 3-T	537	CTACCCACCTTCCTCATTCCTCTACTGAATCCAGGTCTT AATGTATTATAAGC		Same as ~450-nt oligo 3-T
940-nt-HIST1H1B- oligo 3-T	538	CTACCCACCTTCCTCATTCCTCTATCAGCTCTGAGACTG GGGGC		Same as ~450-nt oligo 3-T
940-nt-HOTAIR- oligo 3-T	539	CTACCCACCTTCCTCATTCCTCTACAACACTCACCCAC GGAGCAG		Same as ~450-nt oligo 3-T
940-nt-HOXB2-oligo 3-T	540	CTACCCACCTTCCTCATTCCTCTACCCAATCTCCCTCTCA AATTCAAAATTCATGGC		Same as ~450-nt oligo 3-T
940-nt-HSPA5-oligo 3-T	541	CTACCCACCTTCCTCATTCCTCTAGAAATACAGGCCGCG GCGCT		Same as ~450-nt oligo 3-T
940-nt-LIG1-oligo 3- T	542	CTACCCACCTTCCTCATTCCTCTATGGGAGGAGGGCGGG AAAAG		Same as ~450-nt oligo 3-T
940-nt-MBD4-oligo 3-T	543	CTACCCACCTTCCTCATTCCTCTATCCATTCTCGAGGAA GACGCACCG		Same as ~450-nt oligo 3-T
940-nt-MCM9-oligo 3-T	544	CTACCCACCTTCCTCATTCCTCTATGGCATTAAACCACAA CTGGGTAAATGA		Same as ~450-nt oligo 3-T
940-nt-MECP2-oligo 3-T	545	CTACCCACCTTCCTCATTCCTCTACGGCTTTTACCACAGC CCTCTCT		Same as ~450-nt oligo 3-T
940-nt-MRE11-oligo 3-T	546	CTACCCACCTTCCTCATTCCTCTAGGCTGTCTCTTTTCG GGAAGAAAAGG		Same as ~450-nt oligo 3-T
940-nt-MSH2-oligo 3-T	547	CTACCCACCTTCCTCATTCCTCTACCTCCTGGTGAAGA AAATGCGCGACC		Same as ~450-nt oligo 3-T
940-nt-MSH4-oligo 3-T	548	CTACCCACCTTCCTCATTCCTCTAGAAGGCGAGGTTGAT GAGATCTCA		Same as ~450-nt oligo 3-T
940-nt-MSH5-oligo 3-T	549	CTACCCACCTTCCTCATTCCTCTATCGCAGCGTTTCCCG CCTTTTC		Same as ~450-nt oligo 3-T
940-nt-OGG1-oligo 3-T	550	CTACCCACCTTCCTCATTCCTCTATCAAAGCCCCGCTTCG TGCCCCAA		Same as ~450-nt oligo 3-T
940-nt-OLIG2-oligo 3-T	551	CTACCCACCTTCCTCATTCCTCTAGGCTGCGGCTGCTGT GG		Same as ~450-nt oligo 3-T
940-nt-PELO-oligo 3-T	552	CTACCCACCTTCCTCATTCCTCTACTGGCAGCTTCCCAG GCTCTG		Same as ~450-nt oligo 3-T
940-nt-PM52-oligo 3-T	553	CTACCCACCTTCCTCATTCCTCTAGCCCCGCTCACCTCGA GC		Same as ~450-nt oligo 3-T
940-nt-PMS2P5- oligo 3-T	554	CTACCCACCTTCCTCATTCCTCTATTTCCAGGGAGGTTG GAAGCCGTG		Same as ~450-nt oligo 3-T
940-nt-POLD4-oligo 3-T	555	CTACCCACCTTCCTCATTCCTCTAGAGACCGGCCAGTGG GC		Same as ~450-nt oligo 3-T
940-nt-PRUNE2- oligo 3-T	556	CTACCCACCTTCCTCATTCCTCTACCTTTCCATTAGCATA CGCGCGCG		Same as ~450-nt oligo 3-T
940-nt-RPA2-oligo 3- T	557	CTACCCACCTTCCTCATTCCTCTACGCTTGTCTCTGTCTC CTCTGC		Same as ~450-nt oligo 3-T
940-nt-RPA4-oligo 3- T	558	CTACCCACCTTCCTCATTCCTCTATAATCCTTCCCTGTGT GTCAGTGGTTGC		Same as ~450-nt oligo 3-T
940-nt-TAF1-oligo 3- T	559	CTACCCACCTTCCTCATTCCTCTATCTTCGTGCTGTCCG TGTCTGACA		Same as ~450-nt oligo 3-T
940-nt-TP53-oligo 3- T	560	CTACCCACCTTCCTCATTCCTCTAGCACGCTCCAGCCC GAAC		Same as ~450-nt oligo 3-T
940-nt-UBB-oligo 3- T	561	CTACCCACCTTCCTCATTCCTCTATTCTCATAGCCGTAA GAAAGGCTCCT		Same as ~450-nt oligo 3-T
940-nt-XRCC2-oligo 3-T	562	CTACCCACCTTCCTCATTCCTCTACCTCGGTCCAGACTC AGC		Same as ~450-nt oligo 3-T
940-nt-ZNF830-oligo 3-T	563	CTACCCACCTTCCTCATTCCTCTATCCGTTCCGACTGGT GCTCAGAC		Same as ~450-nt oligo 3-T

Oligo name	SEQ ID NO	Oligo sequence	Notes
620-nt-Ash11-oligo 1-T	564	GCTGGCCACGGTGACCGGAATGTTTTGAAATC CACCAACTACCCACAC ACACC	1. All oligos were designed to capture target sequences of ~620 nt in length.
620-nt-Btk-oligo 1-T	565	AGGGGAGAATCGGGATCACACAAGTCA ATCACCAACTACCCACAC ACACC	
620-nt-Ccl5-oligo 1-T	566	GGTCAAGCCATTGGTCTTCACTTGGCAA ATCACCAACTACCCACAC ACACC	
620-nt-Cd180-oligo 1-T	567	TGGTTAGGAGAAGGGTAGAGCAGGGCAG ATCACCAACTACCCACA CACACC	2. Non-bolded nucleotides are complementary to the 5' end of target sequence.
620-nt-Cox8a-oligo 1-T	568	ACTGTTATCTCTCAGAAACCCGTTTCC ATCACCAACTACCCACACA CACC	
620-nt-Cxcl15-oligo 1-T	569	CAGAAGAAATGGAGTCCAGACTGACACA ATCACCAACTACCCACA CACACC	
620-nt-Cxcr4-oligo 1-T	570	CCGCCAGGGTCTATACTCTGAGCTGTCA ATCACCAACTACCCACAC ACACC	3. Nucleotide in red is complementary to nucleotide immediately 5' of FEN cleavage site.
620-nt-Dnm3b-oligo 1-T	571	CCTGCTTTCACCTAGCTTGGCCTTTGTC ATCACCAACTACCCACACA CACC	
620-nt-Foxo1-oligo 1-T	572	CAGCCTCCACTCATCCTGAACCCAGAG ATCACCAACTACCCACAC ACACC	
620-nt-Gdf5-oligo 1-T	573	GGGGCTCCCCATCCAGAGTAGACTCA ATCACCAACTACCCACACA CACC	4. Bolded nucleotides are complementary to oligo U1.
620-nt-Hdac3-oligo 1-T	574	GCTAGACTGGTGGTCAAAGCCTCTCTGA ATCACCAACTACCCACAC ACACC	
620-nt-Hsf1-oligo 1-T	575	GGAACCATGCAGGCACCTGGCAAGACCA ATCACCAACTACCCACACA CACC	
620-nt-Hsf2bp-oligo 1-T	576	GAGCTCCTGGCCCTGAGTCCAAGCTCA ATCACCAACTACCCACACAC ACC	
620-nt-Hsf4-oligo 1-T	577	GATCCCCGATCCCCCACAGTCACAGTA ATCACCAACTACCCACACACA CC	
620-nt-Hsf5-oligo 1-T	578	AGAGGAACCCACTCCAGTCTTCCCTAA ATCACCAACTACCCACAC ACACC	
620-nt-Hsfy2-oligo 1-T	579	ATCGCCCAAAGTCAGGAAAACAAA ATCACCAACTACCCACACACA CC	
620-nt-Hsp90aa1-oligo 1-T	580	AGCTGTCTGGAGGCCTTCCAATAGGAACA ATCACCAACTACCCACAC ACACC	
620-nt-Hsp90ab1-oligo 1-T	581	AAGAGTCCAGAACTCAGGGGACTGAAAG ATCACCAACTACCCACA CACACC	
620-nt-Hspal-oligo 1-T	582	ACTTAAGGGCTAAAACCTCTCCCCCCA ATCACCAACTACCCACAC ACACC	
620-nt-Hspalb-oligo 1-T	583	TGGTGAGAGTGTGGGACTCAGACCAGGG ATCACCAACTACCCACAC ACACC	
620-nt-Hspa2-oligo 1-T	584	CGGTTGCATCGAATACAGAACGGCTGGG ATCACCAACTACCCACAC ACACC	
620-nt-Ifna1-oligo 1-T	585	CAAGGATGGTTGCAATGTTTCTCTCA ATCACCAACTACCCACACAC ACC	
620-nt-Ifna2-oligo 1-T	586	CTAAGGTAGAAGTATTGTTGTGGCCTGG ATCACCAACTACCCACAC ACACC	
620-nt-Ifng-oligo 1-T	587	GCCTTGTGTTGTGGTTGGGGACGGCTGA ATCACCAACTACCCACAC ACACC	
620-nt-Ikbb-oligo 1-T	588	GAGGCTGGGTATTGTCTTCTCCACCAGG ATCACCAACTACCCACAC ACACC	
620-nt-Ikbbg-oligo 1-T	589	AAGTCATAACACCGATCGCTCCAGCACA ATCACCAACTACCCACAC ACACC	
620-nt-Il10-oligo 1-T	590	AACGTCCGATATTTCTTGCTGACTCGGA ATCACCAACTACCCACAC ACACC	
620-nt-Il12a-oligo 1-T	591	ACCATCTCCATAACCACAAGCCTTTGGG ATCACCAACTACCCACAC ACACC	
620-nt-Il12b-oligo 1-T	592	CCACCTTTCCCCATGTTTCTTGTTAGCA ATCACCAACTACCCACACA CACC	
620-nt-Il3ra-oligo 1-T	593	TCCTGTATCCGTTGCAACTGTACTCGCA ATCACCAACTACCCACACA CACC	

620-nt-Ilf6-oligo 1-T	594	AGACACACACTCCCTCTCCCACCCCCAATCACCAACTACCCACACAC ACC	
620-nt-Irak1-oligo 1-T	595	TGCTTGTGTGCTAGATCTCCATGCCCCCAATCACCAACTACCCACACA CACC	
620-nt-Irak4-oligo 1-T	596	CCTGTCTGCCCCGCTCTTGAAATCACCAACTACCCACACACACC	
620-nt-Irf3-oligo 1-T	597	CGTTTGCCGCAAAGCATACTAGGATTAAATCACCAACTACCCACACAC ACC	
620-nt-Irf5-oligo 1-T	598	CTTGGGCCACTCTGTCTCCTCAGCCTTAATCACCAACTACCCACACA CACC	
620-nt-Irf7-oligo 1-T	599	CAGGAAAGCACATTGGGACAAGGCAGGGAATCACCAACTACCCACA CACACC	
620-nt-Jun-oligo 1-T	600	ACTTCCGGGGACAGAGCGAATCACCAACTACCCACACACACC	
620-nt-Kdm6b-oligo 1-T	601	CTGCTACCTCCAACCCCTAACAAGACACCAATCACCAACTACCCACAC ACACC	
620-nt-Lbp-oligo 1-T	602	AACCTTGCAATGGGATAGCTGTCTGAACAATCACCAACTACCCACAC ACACC	
620-nt-Ly86-oligo 1-T	603	AACAGGAGCTTGTCTTAGGAGCCCACACAATCACCAACTACCCACAC ACACC	
620-nt-Map3k1-oligo 1-T	604	AGAGGTTACTCCCCCGTCTTGGAAGGCAATCACCAACTACCCACAC ACACC	
620-nt-Mapk1-oligo 1-T	605	GTGGGACAGAAACCAAAAAGCATAACCCAATCACCAACTACCCACA CACACC	
620-nt-Mapk13-oligo 1-T	606	AAGGGAGAGGAGAGACTAGAGAGGAGAGAATCACCAACTACCCACA CACACC	
620-nt-Mapk14-oligo 1-T	607	CAGGCGCTCCTACATGTTTTAGGCAGAAATCACCAACTACCCACAC ACACC	
620-nt-Mapk15-oligo 1-T	608	GCTTTGTGCTCAGGTTTACGGGGACAAATCACCAACTACCCACACACAC C	
620-nt-Mapk3-oligo 1-T	609	GCCTGAGCCATCATACAGATTTCAGGAATCACCAACTACCCACACACA CC	
620-nt-Mapk4-oligo 1-T	610	AAGTAAAACCTCACAGATGGGAACCGAGAATCACCAACTACCCACAC ACACC	
620-nt-Mapk6-oligo 1-T	611	GAATGAGTGAGGGGTCTTGGGAAAACCTGAATCACCAACTACCCACAC ACACC	
620-nt-Mapk8-oligo 1-T	612	TTCAAGGGCCTACCACCCACCAAGGAGTAATCACCAACTACCCACAC ACACC	
620-nt-Mme-oligo 1-T	613	AGTGCTGATTGCAGGCTCTCATTCTCTAAATCACCAACTACCCACACA CACC	
620-nt-Myd88-oligo 1-T	614	CGGGGTCTGCAAAAAACACGCGGAACCTAATCACCAACTACCCACAC ACACC	
620-nt-Nfkb1-oligo 1-T	615	AGCCTGGGGCTAGGTAGCTCCCAAGTGAATCACCAACTACCCACACA CACC	
620-nt-Nfkb2-oligo 1-T	616	CAAACCTCCTCTCTGTGGAAAATTGCCAATCACCAACTACCCACACA CACC	
620-nt-Nfkb1a-oligo 1-T	617	CGCTTCTCCTAAATTTCTGCGACGACGGAATCACCAACTACCCACAC ACACC	
620-nt-Nfkb1b-oligo 1-T	618	CGCATTTCTCGTACTTTGGCTGCTGAGGAATCACCAACTACCCACACA CACC	
620-nt-Nr2c2-oligo 1-T	619	GTGCCCCCTCTCCTTCGCGATTGAGAAAAATCACCAACTACCCACAC ACACC	
620-nt-Pik3ap1-oligo 1-T	620	ATAGGGGAAGGCCTGCAGGAAGGGGAAGGAATCACCAACTACCCACA CACACC	
620-nt-Pik3c2a-oligo 1-T	621	TGAGCGCGGACTGCAATTGGGCTTGTAATCACCAACTACCCACAC ACACC	
620-nt-Pik3c2g-oligo 1-T	622	TGTGGGAAATCAAAGAGGACTCTTGGAAATCACCAACTACCCACAC ACACC	
620-nt-Pik3c3-oligo 1-T	623	CGGGTTTTGAATCTTCTGAGTGCCTGTAAATCACCAACTACCCACACA CACC	
620-nt-Pik3ca-oligo 1-T	624	CTTTCGGCTCACATCGTGTAAACAAACAATCACCAACTACCCACAC ACACC	
620-nt-Pik3r1-oligo 1-T	625	CTGCCTCTTTCGAATCCACTAGAGCTTCAATCACCAACTACCCACACA CACC	
620-nt-Pik3r2-oligo 1-T	626	CCGCCGATAGCTTCCGGAAATTGCCGAATCACCAACTACCCACACA CACC	
620-nt-Pik3r3-oligo 1-T	627	ACCCCTAGCCTCGCTCAGAATGGTACAATCACCAACTACCCACACA CACC	

620-nt-Pik3r5-oligo 1-T	628	ACTTCCAGCACTGCTTCTGCAACCTCTGAATCACCAACTACCCACACA CACC	
620-nt-Pik3r6-oligo 1-T	629	TGTCGCCACATTTGCATCTTAGACCACAATCACCAACTACCCACACA CACC	
620-nt-Ptpn6-oligo 1-T	630	AAAGGGTCTAAAGGAGGGGAAAGGGTAGAATCACCAACTACCCACA CACACC	
620-nt-Rela-oligo 1-T	631	GAGGACTAGGCAGACACCTGTACCTCTAATCACCAACTACCCACAC ACACC	
620-nt-Src-oligo 1-T	632	ACGTCATCTCATCGTGTACCTCTTGCCATCACCAACTACCCACACAC ACC	
620-nt-Stat3-oligo 1-T	633	TTTGTGAGCTCCTTACCACGTGGGAATCAATCACCAACTACCCACAC ACACC	
620-nt-Tab1-oligo 1-T	634	ACCCCAATGCCAAACTAGCTGAAGTTAGATCACCAACTACCCACAC ACACC	
620-nt-Tab2-oligo 1-T	635	AGTTTGGCTCGGCCGCTGGATCACCAACTACCCACACACACC	
620-nt-Tet2-oligo 1-T	636	GCCAAGCCGCTCCCCATTTCCCCAAATCACCAACTACCCACACACA CC	
620-nt-Tlr2-oligo 1-T	637	CCTCAGAGGTTGCAACTTGGGCTCAGGAAATCACCAACTACCCACAC ACACC	
620-nt-Tlr4-oligo 1-T	638	TGTTCAATTCGTTACGGGGGGGATGTATATCACCAACTACCCACAC ACACC	
620-nt-Tnf-oligo 1-T	639	ACTGAAACCCATTCTCTCTGTCTCTCCATCACCAACTACCCACACA CACC	
620-nt-Traf6-oligo 1-T	640	AGATGCCCCAAAAGATTTTACCCTCAATCACCAACTACCCACACACA CC	
620-nt-Vav1-oligo 1-T	641	GTTCACCTTGCTTTGGGGCCAAAATCAATCACCAACTACCCACAC ACACC	
620-nt-Ash1-oligo 2-T	642	AGAGCGCCGAGTGCAGGCAATCACCAACTACCCACACACACC	
620-nt-Btk-oligo 2-T	643	AGTTCTGACTTAATGCAGGTAGCTCCCCAATCACCAACTACCCACAC ACACC	
620-nt-Ccl5-oligo 2-T	644	GCATCGAGGCAGAGTGGGGCTGTGAATCACCAACTACCCACACACA CC	
620-nt-Cd180-oligo 2-T	645	GAAGAGGGCCACTAAAAAGAAGCAGCTGAATCACCAACTACCCACA CACACC	
620-nt-Cox8a-oligo 2-T	646	TCAGCAGCAGTGGCGTCAGGACAGACAATCACCAACTACCCACACAC ACC	
620-nt-Cxcl15-oligo 2-T	647	AGAAAAATTCCCCGTTAGCTTACCGAATCACCAACTACCCACACAC ACC	
620-nt-Cxcr4-oligo 2-T	648	GGATGCAAGTGGACTTACACTCAGCTGAATCACCAACTACCCACAC ACACC	
620-nt-Dnmt3b-oligo 2-T	649	AAAAGTTGCTGCAGGTCTCCTTCCCTCAATCACCAACTACCCACAC ACACC	
620-nt-Foxo1-oligo 2-T	650	CGCTGGGTACCGGTGTCTAATCACCAACTACCCACACACACC	
620-nt-Gdf5-oligo 2-T	651	GTGCACCGTCTCCAGTCAGCAGCTGAAAAATCACCAACTACCCACAC ACACC	
620-nt-Hdac3-oligo 2-T	652	GCCACATCGGGTCTGTAGAAATACGCCATCACCAACTACCCACAC ACACC	
620-nt-Hsf1-oligo 2-T	653	GGCCTTTTGAAGAGTGGCTGAAGGCTGAATCACCAACTACCCACAC ACACC	
620-nt-Hsf2bp-oligo 2-T	654	ATCAAACAGCCGTAGGCGTTCTGCTGGAATCACCAACTACCCACAC ACACC	
620-nt-Hsf4-oligo 2-T	655	GCTGGGTGGACAGGGGTGTTACGAATCACCAACTACCCACACACAC C	
620-nt-Hsf5-oligo 2-T	656	CTGAGCAGCTCGGCCTCGAAGATCACCAACTACCCACACACACC	
620-nt-Hsfy2-oligo 2-T	657	ACTGTCTTCATCTGGCTCTGAGGCACAGAATCACCAACTACCCACAC ACACC	
620-nt-Hsp90aa1-oligo 2-T	658	ATCTGAGGTGAACCTCACTCTTACCCGCTAATCACCAACTACCCACAC ACACC	
620-nt-Hsp90ab1-oligo 2-T	659	GTACAGCCCGAACCTCTAACCTCTAGACAATCACCAACTACCCACAC ACACC	
620-nt-Hspala-oligo 2-T	660	GCTGGAACACGCCACGCAAGAGTATCACCAACTACCCACACACAC C	
620-nt-Hspa1b-oligo 2-T	661	GCTGGAACACGCCACGCAAGAGTATCACCAACTACCCACACACAC C	
620-nt-Hspa2-oligo 2-T	662	AAGCGCAAGCCGGGACCTGAATCACCAACTACCCACACACACC	

620-nt-Ilna1-oligo 2-T	663	TCAGCTCACTCAGGACAGGGATGGCTTGAATCACCAACTACCCACAC ACACC	
620-nt-Ilna2-oligo 2-T	664	TGACCACTAGCTCATCACTATCAGCATCAATCACCAACTACCCACAC ACACC	
620-nt-Ilna3-oligo 2-T	665	AGCGTTCATTGTCTCAGAGCTAGGCCGCAATCACCAACTACCCACAC ACACC	
620-nt-Ilna4-oligo 2-T	666	AGGGGAGGCTAGGACACCGGTGTAGTTAAATCACCAACTACCCACAC ACACC	
620-nt-Ilna5-oligo 2-T	667	TTCATCTTGAAGCAGTGACACTGTAGGAAATCACCAACTACCCACAC ACACC	
620-nt-Ilna6-oligo 2-T	668	CCCCTGCTGATCCTCATGCCAGTCAGTAAATCACCAACTACCCACAC ACACC	
620-nt-Ilna7-oligo 2-T	669	CGGAGCGCGTTGCAAGTTGTGATAAATCACCAACTACCCACACACACC	
620-nt-Ilna8-oligo 2-T	670	AGATGCAGGGAGTTAGCGACAGGGAAGAATCACCAACTACCCACAC ACACC	
620-nt-Ilna9-oligo 2-T	671	CGCGAGTAAACAGTGTGAGGCAGGGGGAATCACCAACTACCCACAC ACACC	
620-nt-Ilna10-oligo 2-T	672	GTGGTTGTACCAGCATCAGTCCCAAGAAATCACCAACTACCCACAC ACACC	
620-nt-Ilna11-oligo 2-T	673	GGCTCCAGGGCGTCCATCACTTTGTAGAAATCACCAACTACCCACAC ACACC	
620-nt-Ilna12-oligo 2-T	674	GGGACAGAGGACAGCGGAGAGAATCACCAACTACCCACACACACC	
620-nt-Ilna13-oligo 2-T	675	GCTCCCAAAACAGTAGTTCTGAGACTCGCAATCACCAACTACCCACAC ACACC	
620-nt-Ilna14-oligo 2-T	676	AGGCGCAGTGGGGATGTCCAATCACCAACTACCCACACACACC	
620-nt-Ilna15-oligo 2-T	677	GATCACTCACTACTCTGTAGAGGGGACAAATCACCAACTACCCACAC ACACC	
620-nt-Ilna16-oligo 2-T	678	TGTCTCTGCCACACTCAGTGCAACTCTGAATCACCAACTACCCACACA CACC	
620-nt-Ilna17-oligo 2-T	679	CAAGCTTACCAGATGGAGTTGGAAGTGGAAATCACCAACTACCCACAC ACACC	
620-nt-Ilna18-oligo 2-T	680	CTGTATGGACAGGAACAGCAGCCCCAATCACCAACTACCCACACACA CC	
620-nt-Ilna19-oligo 2-T	681	AAGCCAGCCCCAATAAAAAGACCTCCGAATCACCAACTACCCACACA CACC	
620-nt-Ilna20-oligo 2-T	682	CGCCATTTTCTCTCGCGGGCTAATCACCAACTACCCACACACACC	
620-nt-Ilna21-oligo 2-T	683	GGCGATCGGGAACGAGGAAGGAGGACAAATCACCAACTACCCACAC ACACC	
620-nt-Ilna22-oligo 2-T	684	CCCACGAACTCCGAGATCCCTTTTAAAAAATCACCAACTACCCACAC ACACC	
620-nt-Ilna23-oligo 2-T	685	CAGATGGTCTTGTTCAGCTCCTGCCGGTAATCACCAACTACCCACACA CACC	
620-nt-Ilna24-oligo 2-T	686	GGGCCAGATTCTCAGGGAAGCCAGGGAATCACCAACTACCCACACA CACC	
620-nt-Ilna25-oligo 2-T	687	GAAGTCACTGCACGATACTTCCCCTGCGAATCACCAACTACCCACAC ACACC	
620-nt-Ilna26-oligo 2-T	688	GGCAGCGAGAGAAAGAAAGCGGGCGAATCACCAACTACCCACACAC ACC	
620-nt-Ilna27-oligo 2-T	689	ACACACTCAGGGGACTTCTCCGACTCAATCACCAACTACCCACACAC ACC	
620-nt-Ilna28-oligo 2-T	690	TCACGTCCGCACAAAGAAACCGCCCCAATCACCAACTACCCACACACA CC	
620-nt-Ilna29-oligo 2-T	691	GGTGACTCACCAGTGGGATTGGGGAATCACCAACTACCCACACACAC C	
620-nt-Ilna30-oligo 2-T	692	GTAAAGCGCGACCAAGGGTATGGAGAACAATCACCAACTACCCACAC ACACC	
620-nt-Ilna31-oligo 2-T	693	CTGGCTGGAAATCCCACTGACGTGAGAAATCACCAACTACCCACAC ACACC	
620-nt-Ilna32-oligo 2-T	694	GTGCCCAGAGGTTGGCTATCTCTGAGACAATCACCAACTACCCACAC ACACC	
620-nt-Ilna33-oligo 2-T	695	CACCATTTGCTCGTACTCCTCGTCCCTTCAATCACCAACTACCCACACA CACC	
620-nt-Ilna34-oligo 2-T	696	CGCCCCGTACTGCACCACAGCCAAATCACCAACTACCCACACACAC C	

620-nt-Nr2c2-oligo 2-T	697	GGGACCGCGCTTTGTTATTGTTTTGTCAATCACCAACTACCCACACA CACC	
620-nt-Pik3ap1-oligo 2-T	698	CTAAAGAAGTAGGGAGACAGGTCTCAAGAATCACCAACTACCCACA CACACC	
620-nt-Pik3c2a-oligo 2-T	699	GCCCCGAGGAAGGAGTCAGAGAATCACCAACTACCCACACACACC	
620-nt-Pik3c2g-oligo 2-T	700	TGGAAGGCAGGTCCAGGCCAGCAGAATCACCAACTACCCACACACA CC	
620-nt-Pik3c3-oligo 2-T	701	CAGGAAGAAACCGCCGCTGCATCAAAACAATCACCAACTACCCACAC ACACC	
620-nt-Pik3ca-oligo 2-T	702	CCTCCGCTGGTCTTCCATCTACCCACGTAATCACCAACTACCCACACA CACC	
620-nt-Pik3r1-oligo 2-T	703	TCGGCTCGCTTCCCTCTCTGCCCATCACCAACTACCCACACACACC	
620-nt-Pik3r2-oligo 2-T	704	GGTATCGCGGAACCCCCAATGCAAAAGGAATCACCAACTACCCACAC ACACC	
620-nt-Pik3r3-oligo 2-T	705	AAAAGCTCCCACCGCCTCCTAATCTTTCAATCACCAACTACCCACACA CACC	
620-nt-Pik3r5-oligo 2-T	706	CCTACCGGCTGCAGTTGGAGCACACTAATCACCAACTACCCACACAC ACC	
620-nt-Pik3r6-oligo 2-T	707	CAGCCAATGAGCTCCTGAGGTCTGTTCATCACCAACTACCCACAC ACACC	
620-nt-Ptpn6-oligo 2-T	708	ATTTCTGTTTCTTGGCCACTTCCTCCTCTAATCACCAACTACCCACACA CACC	
620-nt-Rela-oligo 2-T	709	CGGCGCGCTAAAGTAAAGCCATTGCGCAATCACCAACTACCCACACA CACC	
620-nt-Src-oligo 2-T	710	ACCTGAAGGAGAAACTCACCAGGACAATCACCAACTACCCACACA CACC	
620-nt-Stat3-oligo 2-T	711	CGACACCGAGCGGGGTGTACCTCAATCACCAACTACCCACACACAC C	
620-nt-Tab1-oligo 2-T	712	CGAGGCGGTCCACACAATGCCAGAGAATCACCAACTACCCACACAC ACC	
620-nt-Tab2-oligo 2-T	713	TCCGCATCAAAATACGGCCACAATCACCAACTACCCACACACACC	
620-nt-Tet2-oligo 2-T	714	CCTTTCTAACAAATCCTACAGGGCAGCCAATCACCAACTACCCACAC ACACC	
620-nt-Tlr2-oligo 2-T	715	CACAGGGGTCTGCAAGGGGAACGGAATCACCAACTACCCACACACA CC	
620-nt-Tlr4-oligo 2-T	716	CAGGGAGGCATCATCTGGCAATCACCAACTACCCACACACACC	
620-nt-Tnf-oligo 2-T	717	CTGTGCTCATGGTGTCTTTTCTGGAGGGAATCACCAACTACCCACACA CACC	
620-nt-Traf6-oligo 2-T	718	ACTCTAGTTTGAAGCCGCTCACTCACCGAATCACCAACTACCCACAC ACACC	
620-nt-Vav1-oligo 2-T	719	TGGCCTGGGGAAGCAGGTGTGTAAGCAATCACCAACTACCCACACA CACC	
620-nt-Ash11-oligo 3-T	720	CTACCCCACTTCTCTATTCTCTAGTGGAGAGAGAGAGGGTGGCAG CCTG	1. All oligos were designed to capture target sequences of ~620 nt in length.
620-nt-Btk-oligo 3-T	721	CTACCCCACTTCTCTATTCTCTAGACGTGTGAGATGCCAGGACTT GGAAG	
620-nt-Ccl5-oligo 3-T	722	CTACCCCACTTCTCTATTCTCTATGGTCAGAACCTCGACTTACAT GGTGA	
620-nt-Cd180-oligo 3-T	723	CTACCCCACTTCTCTATTCTCTATGTCCGGAGCCATCCCAGGTCCA GAG	
620-nt-Cox8a-oligo 3-T	724	CTACCCCACTTCTCTATTCTCTATGACGGCGCGGAACGACCA	2. Non-bolded nucleotides are complementary to the 5' end of target sequence.
620-nt-Cxcl15-oligo 3-T	725	CTACCCCACTTCTCTATTCTCTAGATGCCTAGGTTAAGGACAGCCA GCAGG	
620-nt-Cxcr4-oligo 3-T	726	CTACCCCACTTCTCTATTCTCTATCGGTTCCATGGCAACACTCGCTC TACA	
620-nt-Dnmt3b-oligo 3-T	727	CTACCCCACTTCTCTATTCTCTACTCTTCAGAGTGGCGGAGTGCAG CG	
620-nt-Foxo1-oligo 3-T	728	CTACCCCACTTCTCTATTCTCTAGGAGCCTCCGAGTTCGTGCCTAG GAC	3. Nucleotide in red is complementary to
620-nt-Gdf5-oligo 3-T	729	CTACCCCACTTCTCTATTCTCTATAACTCGTTCTTGAAAGGAGAAA GCCGA	

620-nt-Hdac3-oligo 3-T	730	CTACCCACCTTCCTCATTCTCTACGGTCTTGGCCATGGTGCTGG	nucleotide immediately 5' of FEN cleavage site.
620-nt-Hsf1-oligo 3-T	731	CTACCCACCTTCCTCATTCTCTAGCTAGTGTGCTGCTGCCCT	
620-nt-Hsf2bp-oligo 3-T	732	CTACCCACCTTCCTCATTCTCTAGGAGACCTGACCTCCAACCTAACC GCCCT	4. Bolded nucleotides are complementary to oligo U2.
620-nt-Hsf4-oligo 3-T	733	CTACCCACCTTCCTCATTCTCTACATTGGAGGGATCCCGGGGACCA CTG	
620-nt-Hsf5-oligo 3-T	734	CTACCCACCTTCCTCATTCTCTAGCGGCTGGTCGATGAGCAG	
620-nt-Hsfy2-oligo 3-T	735	CTACCCACCTTCCTCATTCTCTAGGGTATGCGGTGGCCTCTTTATCA AGAA	
620-nt-Hsp90aa1-oligo 3-T	736	CTACCCACCTTCCTCATTCTCTACACCAGGGAGGGCACACAAGCAC C	
620-nt-Hsp90ab1-oligo 3-T	737	CTACCCACCTTCCTCATTCTCTACGTTATCTCCACAGGCTGCCCCGC	
620-nt-Hspa1a-oligo 3-T	738	CTACCCACCTTCCTCATTCTCTAGGTGGTGGCCAGGTCGATGCCGA T	
620-nt-Hspa1b-oligo 3-T	739	CTACCCACCTTCCTCATTCTCTAGGTGGTGGCCAGGTCGATGCCGA T	
620-nt-Hspa2-oligo 3-T	740	CTACCCACCTTCCTCATTCTCTAGCCAACGGTGGCTTCTCACAACCTC TCCA	
620-nt-Ifna1-oligo 3-T	741	CTACCCACCTTCCTCATTCTCTAGCCTTCTTGATCTGCTGGGCATCC AC	
620-nt-Ifna2-oligo 3-T	742	CTACCCACCTTCCTCATTCTCTACGAGGAAAGCACAGAGTCTAGCC ATTGT	
620-nt-Ifng-oligo 3-T	743	CTACCCACCTTCCTCATTCTCTAGGAGGAGAAGCCCAGAACTTCTG TCTCA	
620-nt-Ikbkb-oligo 3-T	744	CTACCCACCTTCCTCATTCTCTACAAATAAAAGTTGCGGCGGCCGG GC	
620-nt-Ikbkg-oligo 3-T	745	CTACCCACCTTCCTCATTCTCTAGCTAGGCGGCTCAGCTTTAGCAC AAGAG	
620-nt-Il10-oligo 3-T	746	CTACCCACCTTCCTCATTCTCTAGAGCAGGCAGCATAGCAGTGCT	
620-nt-Il12a-oligo 3-T	747	CTACCCACCTTCCTCATTCTCTACTGGTGGGGATATATCTGGTCATT GCCG	
620-nt-Il12b-oligo 3-T	748	CTACCCACCTTCCTCATTCTCTAGGAGAGAACTTACCTTGCTTTGCT GCGA	
620-nt-Il3ra-oligo 3-T	749	CTACCCACCTTCCTCATTCTCTATGACCACGAGTTAGTTTCTGAGG CAGGG	
620-nt-Il6-oligo 3-T	750	CTACCCACCTTCCTCATTCTCTAGGCAACTGGATGGAAGTCTCCTG CGTGG	
620-nt-Irak1-oligo 3-T	751	CTACCCACCTTCCTCATTCTCTAGCGGCACATAACCCAGGGTGGCA C	
620-nt-Irak4-oligo 3-T	752	CTACCCACCTTCCTCATTCTCTATCGTGGCTTGAGATCAGCCTTCCT CAT	
620-nt-Irf3-oligo 3-T	753	CTACCCACCTTCCTCATTCTCTAGCCACGCAACCCCTTCTTTGTCA GC	
620-nt-Irf5-oligo 3-T	754	CTACCCACCTTCCTCATTCTCTACTGGGCCCTTGGGACTTTCC	
620-nt-Irf7-oligo 3-T	755	CTACCCACCTTCCTCATTCTCTAGAATTTCCCTGTGGCACTAGCTC ACCT	
620-nt-Jun-oligo 3-T	756	CTACCCACCTTCCTCATTCTCTAGTCCTTATCCAGCCTGAGCTCAAC ACTT	
620-nt-Kdm6b-oligo 3-T	757	CTACCCACCTTCCTCATTCTCTACTGTCTCATGGAACAAGCACTCC AGCAG	
620-nt-Lbp-oligo 3-T	758	CTACCCACCTTCCTCATTCTCTAGCAATGTAGAGAGCAGAGGTCCA GTCCC	
620-nt-Ly86-oligo 3-T	759	CTACCCACCTTCCTCATTCTCTAGCTCGTCTCCCTGGAAGGAGT AGT	
620-nt-Map3k1-oligo 3-T	760	CTACCCACCTTCCTCATTCTCTACATTGCTCGTGCCTGC	
620-nt-Mapk1-oligo 3-T	761	CTACCCACCTTCCTCATTCTCTACACAGAGGAGAGAATTAACCGCC GGTAG	
620-nt-Mapk13-oligo 3-T	762	CTACCCACCTTCCTCATTCTCTAGTTACAGCCTTGGAGAACCAGAT TCGCC	

620-nt-Mapk14-oligo 3-T	763	CTACCCACCTTCCTCATTCTCTAGAACGTGGGCTCTCTGCGACA TCTTC
620-nt-Mapk15-oligo 3-T	764	CTACCCACCTTCCTCATTCTCTACCTGTGGGCACAAGGTGGAGCCA AC
620-nt-Mapk3-oligo 3-T	765	CTACCCACCTTCCTCATTCTCTACACTCCCTGAGCCCCCATGT
620-nt-Mapk4-oligo 3-T	766	CTACCCACCTTCCTCATTCTCTAGCAGGGAGCCACTCACCGA
620-nt-Mapk6-oligo 3-T	767	CTACCCACCTTCCTCATTCTCTATCGACAGAGTGCAGGGCCC
620-nt-Mapk8-oligo 3-T	768	CTACCCACCTTCCTCATTCTCTACAGTCAAAACACGCTCCCCGTGG GTG
620-nt-Mme-oligo 3-T	769	CTACCCACCTTCCTCATTCTCTAGAAGTTGGAGGGCGACTGAGCAGA CACA
620-nt-Myd88-oligo 3-T	770	CTACCCACCTTCCTCATTCTCTATGAAGGAGTCCAGGGACCCGGAT CCC
620-nt-Nfkb1-oligo 3-T	771	CTACCCACCTTCCTCATTCTCTAGAGAGCAGACAGACGGACGGAC CCAC
620-nt-Nfkb2-oligo 3-T	772	CTACCCACCTTCCTCATTCTCTACAGGAAACCGGGTCAGTCCC
620-nt-Nfkb1a-oligo 3-T	773	CTACCCACCTTCCTCATTCTCTATGGAGTCCAGGCCGCTGTCTGT
620-nt-Nfkb1b-oligo 3-T	774	CTACCCACCTTCCTCATTCTCTATCAGAAAAATGAGGTGGACATTC GCTCG
620-nt-Nr2c2-oligo 3-T	775	CTACCCACCTTCCTCATTCTCTAGTGAAAGGTCAGAGTTCGTGACC CCGGT
620-nt-Pik3ap1-oligo 3-T	776	CTACCCACCTTCCTCATTCTCTACTGGGAGGGTGGTGTGAGAGAGC CTCA
620-nt-Pik3c2a-oligo 3-T	777	CTACCCACCTTCCTCATTCTCTAGGATGGAGTGACCAAGGGGCACG G
620-nt-Pik3c2g-oligo 3-T	778	CTACCCACCTTCCTCATTCTCTATTTGTGCTGTGGGTGGCAGAGTC AATCC
620-nt-Pik3c3-oligo 3-T	779	CTACCCACCTTCCTCATTCTCTATGAAAACCTTCCTGTTCGGCCATC TGCC
620-nt-Pik3ca-oligo 3-T	780	CTACCCACCTTCCTCATTCTCTAGAGCGGAGGCTGACACCGT
620-nt-Pik3r1-oligo 3-T	781	CTACCCACCTTCCTCATTCTCTAGCCTGGGCTTCAGCGGCCTCTGA A
620-nt-Pik3r2-oligo 3-T	782	CTACCCACCTTCCTCATTCTCTAGGGGCTACGATGCACGCATAGT CTGTG
620-nt-Pik3r3-oligo 3-T	783	CTACCCACCTTCCTCATTCTCTAGCGGAAGCCACTTTGTGTCTCTC G
620-nt-Pik3r5-oligo 3-T	784	CTACCCACCTTCCTCATTCTCTACTCCAGAATGGTCCCAACCTGCTC TGGC
620-nt-Pik3r6-oligo 3-T	785	CTACCCACCTTCCTCATTCTCTAGGAGGCCAGCAACCCCAAGGACC A
620-nt-Ptpn6-oligo 3-T	786	CTACCCACCTTCCTCATTCTCTATCCAGCCTGTCAAACCAATGGGA CTCTG
620-nt-Rela-oligo 3-T	787	CTACCCACCTTCCTCATTCTCTAGAGGGGAAATGCGCG
620-nt-Src-oligo 3-T	788	CTACCCACCTTCCTCATTCTCTAGCCCTTCCTCCAGCCACGTCTGCA
620-nt-Stat3-oligo 3-T	789	CTACCCACCTTCCTCATTCTCTAGCGATCCGGTTAGGGCTCG
620-nt-Tab1-oligo 3-T	790	CTACCCACCTTCCTCATTCTCTATACAGCACGCCTTACACAAGCGG GGG
620-nt-Tab2-oligo 3-T	791	CTACCCACCTTCCTCATTCTCTACAATATGGCGGGCGCGG
620-nt-Tet2-oligo 3-T	792	CTACCCACCTTCCTCATTCTCTATCTTGCTTCCTCTGCTCCTCTAC GAAA
620-nt-Tlr2-oligo 3-T	793	CTACCCACCTTCCTCATTCTCTAGCCTGGCATCAGAGCTGGCAGGA G
620-nt-Tlr4-oligo 3-T	794	CTACCCACCTTCCTCATTCTCTATTTTCTGCAACCAGTGGAGGCCAT GAGA
620-nt-Tnf-oligo 3-T	795	CTACCCACCTTCCTCATTCTCTAGATGTGGCGCCTTGGGCCAGTGA G
620-nt-Traf6-oligo 3-T	796	CTACCCACCTTCCTCATTCTCTATCGTACACGCTGAGACTTCACAA CCAAC
620-nt-Vav1-oligo 3-T	797	CTACCCACCTTCCTCATTCTCTATTGGCACAAGAGGACACCGTCCC

Table 2: Characteristics of designed FENGCC target sequences used in this study.

Gene	Genome coordinates (hg38)	Captured length (nt)	Captured sequence GC content (%)	T _m			Captured sequence relative to TSS (+/-)	Gene/product name ^{a,b}	Notes
				Oligo 1-1	Oligo 2-2	Oligo 3-3			
<i>EPAM2</i>	chr3:36,992,888-36,993,182	295	62	69	70	71	+13 to +307	EPAM2-1-interacting protein 1	Divergently transcribed with <i>MLH1</i> (target sequence is -13 to +287 relative to <i>MLH1a</i> TSS)
<i>AIP1</i>	chr1:241,847,874-241,848,155	282	66	66	69	69	-330 to -49	Exonuclease 1	
<i>EXO1</i>	chr9:125,241,405-125,241,709	305	61	64	70	67	-366 to +84	Endoplasmic reticulum chaperone BiP	Control promoter with open chromatin
<i>HSP45</i>	chr3:36,993,504-36,993,803	300	60	61	69	70	-13 to +287 (MLH1 TSSa)	DNA mismatch repair protein Mlh1	Divergently transcribed with <i>EPAM2</i> (target sequence is +13 to +307 relative to <i>EPAM2</i> TSS)
<i>MLH1</i>	chr14:75,051,561-75,051,853	293	36	64	66	65	-396 to +53	DNA mismatch repair protein Mlh3	
<i>MLH3</i>	chr2:47,402,729-47,403,020	292	56	65	67	67	-338 to -47	DNA mismatch repair protein MSH2	
<i>MSH2</i>	chr5:80,654,325-80,654,639	315	70	68	71	71	-323 to -9	DNA mismatch repair protein MSH3	
<i>MSH3</i>	chr2:47,782,759-47,783,061	303	78	65	72	74	-323 to -21	DNA mismatch repair protein MSH6	
<i>MSH6</i>	chr2:189,783,763-189,784,054	292	62	62	69	70	-322 to -31	PMS1 protein homolog 1	
<i>PMS1</i>	chr7:6,009,126-6,009,426	301	61	64	67	71	-320 to -20	Mismatch repair endonuclease PMS2	Divergently transcribed with <i>AIMP2</i> (target sequence is -2? to +?) relative to <i>AIMP2</i> TSS)
<i>PMS2</i>	chr12:103,965,49-103,965,791	293	56	64	69	67	-316 to -24	GTT mismatch-specific ligase DNA glycosylase	
<i>TDG</i>									

Gene Ontology term
"metabolic process," filtered
for "DNA repair"
Control loci with open
chromatin

Gene	Genome coordinates (hg38)	Captured length (nt)	Capture sequence GC content (%)	T _m			Captured sequence relative to TSS (+/-)	Gene/product name ^{abz}	Notes
<i>ALKB</i>	chr14:77,707,866-77,708,314	449	57	64	69	64	-291 to +158	Nucleic acid dinucleoside	Divergently transcribed with <i>SLIRP1</i> (target sequence is -205 to +244 relative to <i>SLIRP1</i> TSS)
<i>HL</i>								ALKBH1	
<i>ALKB</i>	chr12:109,093,335-109,093,770	436	62	65	65	65	-282 to +154	DNA oxidative demethylase	
<i>H2</i>								ALKBH2	
<i>ALKB</i>	chr11:43,880,520-43,880,969	450	62	64	68	68	-311 to +193	Alpha-ketoglutarate-dependent dioxygenase (RRB homolog 3)	
<i>H3</i>								DNA-(cytosine or pyrimidine site) endonuclease	
<i>APEX</i>	chr14:20,454,833-20,455,278	446	58	67	64	64	-393 to +53	DNA-(cytosine or pyrimidine site) endonuclease	
<i>I</i>								DNA-(cytosine or pyrimidine site) endonuclease 2	
<i>APEX</i>	chrX:55,000,041-55,000,485	445	59	66	67	68	-322 to +123	Serine-protein kinase ATM	
<i>2</i>								Transcriptional regulator ATRX	
<i>ATM</i>	chr11:108,222,537-108,222,984	448	61	62	68	68	-530 to -83	Ataxin-2	Divergently transcribed with <i>NPAT</i> (target sequence is -346 to +102 relative to <i>NPAT</i> TSS)
<i>ATR</i>	chrX:77,786,121-77,786,569	449	49	64	64	65	-353 to +96	Ataxin-2	
<i>ATXN</i>	chr14:92,106,418-92,106,858	441	67	64	70	69	-276 to +165	Ataxin-2	
<i>3</i>								Ataxin-2	
<i>AXIN</i>	chr17:65,561,578-65,562,018	441	66	64	66	68	-370 to +71	Breast cancer type 1 susceptibility protein	
<i>BRC</i>	chr17:43,125,334-43,125,779	446	52	62	65	67	-415 to +31	Breast cancer type 2 susceptibility protein	
<i>AI</i>								WISP1 isoform 1	
<i>BRC</i>	chr13:32,315,183-32,315,631	449	63	68	64	67	-325 to +124	CD44 antigen	
<i>A2</i>								Cadherin-1	
<i>CCN</i>	chr8:133,190,760-133,191,206	447	56	65	69	69	-279 to +168	Cadherin-13	
<i>4</i>								Cadherin-2	Divergently transcribed with <i>NPAT</i> (target sequence is -346 to +102 relative to <i>NPAT</i> TSS)
<i>CD44</i>	chr11:35,138,612-35,139,052	441	62	66	68	64	-270 to +171	Cyclin-dependent kinase inhibitor 2A	
<i>CDH</i>	chr16:68,736,898-68,737,344	447	60	67	68	65	-394 to +53	Cyclin-dependent kinase 4 inhibitor D	
<i>I</i>								Cadherin-13	
<i>CDH</i>	chr16:82,626,498-82,626,943	446	50	65	68	67	-471 to -26	Cadherin-2	
<i>13</i>								Cyclin-dependent kinase 4 inhibitor D	
<i>CDH</i>	chr18:28,177,299-28,177,747	449	73	69	66	70	-300 to +149	Cadherin-2	
<i>2</i>								Cyclin-dependent kinase inhibitor 2A	
<i>CDK</i>	chr9:21,994,358-21,994,794	448	75	68	67	70	-402 to +46	Cyclin-dependent kinase 4 inhibitor D	
<i>N2A</i>								Cadherin-2	
<i>CDK</i>	chr19:10,568,868-10,569,301	434	69	66	69	67	-345 to +89	Cadherin-2	Divergently transcribed with <i>NPAT</i> (target sequence is -346 to +102 relative to <i>NPAT</i> TSS)
<i>N2D</i>								Cadherin-2	
<i>CEP</i>	chr18:580,059-580,492	434	63	60	62	62	-284 to +150	Cadherin-2	
<i>NI</i>								Cadherin-2	
<i>CHE</i>	chr11:125,624,871-125,625,316	446	45	66	63	62	-265 to +181	Cadherin-2	
<i>KI</i>								Cadherin-2	

<i>CHE</i>	chr22:28,741,673-	437	59	68	66	65	-289 to +148	Serine/threonine-protein kinase Chk2
<i>K2</i>	28,742,109	451	56	65	64	67	-271 to +180	Cold shock domain-containing protein E1
<i>CSD</i>	chr1:114,757,871-	447	63	61	62	65	-327 to +120	DNA damage-binding protein 2
<i>DDDB</i>	chr11:47,214,647-	447	39	57	56	59	-297 to +150	D(3) dopamine receptor
<i>2</i>	47,215,093	448	76	65	68	69	-302 to +146	D(4) dopamine receptor
<i>DRD</i>	chr3:114,199,258-	444	70	65	69	69	-383 to +61	Epidermal growth factor receptor EPHA-interacting protein 1
<i>3</i>	114,199,704	450	60	69	69	69	-143 to +307	Epithelial Stromal Interaction 1
<i>DRD</i>	chr11_K1270832v1_	430	55	66	66	65	-267 to +163	DNA excision repair protein ERCC-1
<i>4</i>	alt.166,876-167,323	450	59	64	69	69	-295 to +155	General transcription and DNA repair factor IIIH helicase subunit XPD
<i>EGF</i>	chr7:55,018,634-	445	60	67	68	65	-325 to +120	DNA excision repair protein ERCC-8
<i>R</i>	55,019,077	444	64	66	66	65	-330 to +114	Exonuclease 1
<i>EPM</i>	chr3:36,992,888-	442	63	60	67	67	-283 to +159	Fatoma anemina group C protein
<i>2AIP</i>	36,993,337	446	54	61	65	67	-293 to +153	FAT-Associated Cadherin 4
<i>1</i>	chr13:42,992,151-	447	63	67	66	63	-570 to -124	Flap endonuclease 1; transcribed divergently from FENM258
<i>EPST</i>	chr13:42,992,151-	433	58	65	67	66	-276 to +157	Bis(5'-adenosyl)-triphosphatase
<i>11</i>	42,992,580	448	78	70	70	63	-290 to +158	Forkhead box protein C2
<i>ERC</i>	chr19:45,423,763-	430	76	70	67	69	-325 to +105	Fuzled-7
<i>C1</i>	45,424,212	435	69	68	67	69	-205 to +230	Fuzled-8
<i>ERC</i>	chr19:45,370,469-	436	61	66	70	66	-313 to +123	G-protein coupled receptor 158
<i>C2</i>	45,370,913	447	41	60	65	58	-449 to -2	Glutathione S-transferase A1
<i>ERC</i>	chr5:60,944,930-	445	53	61	65	64	-246 to +199	Histon H1.5
<i>C8</i>	60,945,380	448	67	69	65	69	-307 to +141	High mobility group protein HMGI-C
<i>EYO1</i>	chr1:241,847,874-	449	48	65	68	67	-297 to +152	Homo sapiens HOX transcript antisense RNA, transcript variant 1
<i>FAN</i>	chr9:95,317,552-	446	64	66	66	65	-330 to +114	Exonuclease 1
<i>CC</i>	95,317,993	442	63	60	67	67	-283 to +159	Fatoma anemina group C protein
<i>FAT4</i>	chr4:125,316,119-	446	54	61	65	67	-293 to +153	FAT-Associated Cadherin 4
<i>FEN1</i>	chr11:61,792,341-	447	63	67	66	63	-570 to -124	Flap endonuclease 1; transcribed divergently from FENM258
<i>FHIT</i>	chr3:61,251,306-	433	58	65	67	66	-276 to +157	Bis(5'-adenosyl)-triphosphatase
<i>FOX</i>	chr16:86,566,961-	448	78	70	70	63	-290 to +158	Forkhead box protein C2
<i>C2</i>	86,567,408	430	76	70	67	69	-325 to +105	Fuzled-7
<i>FZD7</i>	chr2:202,034,262-	435	69	68	67	69	-205 to +230	Fuzled-8
<i>FZD8</i>	chr10:35,641,206-	436	61	66	70	66	-313 to +123	G-protein coupled receptor 158
<i>GPR1</i>	chr10:25,175,048-	447	41	60	65	58	-449 to -2	Glutathione S-transferase A1
<i>58</i>	25,175,483	445	53	61	65	64	-246 to +199	Histon H1.5
<i>GSTA1</i>	chr6:52,803,818-	448	67	69	65	69	-307 to +141	High mobility group protein HMGI-C
<i>1</i>	52,804,264	449	48	65	68	67	-297 to +152	Homo sapiens HOX transcript antisense RNA, transcript variant 1
<i>HIST</i>	chr6:27,867,390-	445	53	61	65	64	-246 to +199	Histon H1.5
<i>HIB</i>	chr12:65,824,153-	448	67	69	65	69	-307 to +141	High mobility group protein HMGI-C
<i>HMG</i>	chr12:65,824,153-	449	48	65	68	67	-297 to +152	Homo sapiens HOX transcript antisense RNA, transcript variant 1
<i>A2</i>	65,824,600	449	48	65	68	67	-297 to +152	Homo sapiens HOX transcript antisense RNA, transcript variant 1
<i>HOT</i>	chr12:53,968,606-	449	48	65	68	67	-297 to +152	Homo sapiens HOX transcript antisense RNA, transcript variant 1
<i>AIR</i>	53,969,054	449	48	65	68	67	-297 to +152	Homo sapiens HOX transcript antisense RNA, transcript variant 1

Divergently transcribed with *MLH1* (target sequence is -13 to +425 relative to *MLH1a* TSS)

[illegible]

OGG	chr3:9,749,663-9,750,103	442	64	68	70	70	-282 to +160	N-glycosylase/DNA lyase
OLIG	chr21:33,025,611-33,026,059	449	67	67	63	68	-324 to +125	Oligodeoxyribose transcription factor 2
PARP	chr1:226,408,056-226,408,497	442	66	66	66	69	-396 to +46	Poly (ADP-ribose) polymerase 1
PCN	chr20:5,126,395-5,126,840	446	61	68	65	64	-217 to +229	Proliferating cell nuclear antigen
PEL	chr5:82,787,645-52,788,092	448	60	67	63	67	-271 to +177	Protein pelota homolog
PMS1	chr2:189,783,763-189,784,201	439	64	62	66	65	-322 to +117	PMS1 protein homolog 1
PMS2	chr7:6,008,985-6,009,426	442	62	64	67	69	-320 to +122	Mismatch repair endonuclease PMS2
PMS2	chr7:100,336,159-100,336,604	446	67	64	68	66	-296 to +150	Puative postmeiotic segregation increased 3-like protein 1
PMS2	chr7:77,011,149-77,011,595	447	49	63	62	63	-298 to +149	Puative postmeiotic segregation increased 3-like protein 11
PMS2	chr7:75,358,848-75,359,299	452	61	65	67	67	-302 to +150	Puative postmeiotic segregation increased 2-like protein 2
PMS2	chr7:75,527,982-75,528,429	448	61	65	69	68	-305 to +143	Puative postmeiotic segregation increased 2-like protein 3
PMS2	chr7:74,890,474-74,890,918	445	62	67	69	66	-294 to +151	Postmeiotic segregation increased 2-like protein 5
POL	chr19:50,384,048-50,384,479	432	68	70	69	67	-275 to +157	DNA polymerase delta catalytic subunit
POL	chr11:67,353,447-67,353,893	447	65	68	67	67	-297 to +150	DNA polymerase delta subunit 4
POL	chr15:89,334,728-89,335,168	441	64	66	70	68	-372 to +69	DNA polymerase subunit gamma-1
POL	chr5:75,511,457-75,511,901	445	64	68	64	63	-293 to +152	DNA polymerase kappa
POL	chr3:121,545,842-121,546,288	447	52	60	62	66	-281 to +166	DNA polymerase theta
POL	chr4:56,978,353-56,978,799	447	48	69	66	69	-287 to +160	DNA-directed RNA polymerase II subunit RPB2
R2B	chr11:62,761,244-62,761,691	448	62	62	67	66	-336 to +112	DNA-directed RNA polymerase II subunit RPB7
R2G	chr1:13,704,558-13,705,004	447	34	61	62	57	-297 to +150	PR domain zinc finger protein 2
PRD	chr19:53,881,833-53,882,277	445	68	67	70	67	-351 to +94	Protein kinase C gamma type
PRU	chr9:76,905,871-76,906,311	441	68	67	69	67	-223 to +218	Protein prunz homolog 2
NE2	chr20:3,372,0179-33,720,629	451	58	67	64	64	-298 to +153	Proximal membrane protein 4
PYM	chr16:31,202,733-31,203,170	449	65	62	68	69	-393 to +45	Apoptosis-associated speck-like protein containing a CARD
PYCA	chr19:12,806,317-12,806,765	450	66	63	69	66	-298 to +152	Ribonuclease H2 subunit A

Transcribed from same promoter as *ITGAL*Divergently transcribed with *AIMP2* (target sequence is -287 to +155 relative to *AIMP2* TSS)Divergently transcribed with *STAG3L1* (target sequence is -366 to +86 relative to *STAG3L1* TSS)Divergently transcribed with *STAG3L2* (target sequence is -306 to +139 relative to *STAG3L2* TSS)

<i>RNF11</i>	chr15:58,987,414-58,987,859	446	55	61	68	68	-252 to +194	E3 ubiquitin-protein ligase Arkadia Replication protein A 32 kDa subunit
<i>RP42</i>	chr1:27,914,634-27,915,080	447	56	64	67	64	-282 to +165	Replication protein A 13 kDa subunit
<i>RP43</i>	chr7:7,718,466-7,718,915	450	34	62	57	58	-307 to +143	Replication protein A 30 kDa subunit
<i>RP44</i>	chrX:96,883,621-96,884,066	446	44	66	67	64	-287 to +159	SH3 domain and tetraoxypolide repeat-containing protein 1
<i>SH3T</i>	chr4:8,198,945-8,199,389	445	66	62	65	69	-299 to +146	Single-strand homolog 2
<i>C1</i>	chr21:36,698,824-36,699,273	450	77	68	69	70	-297 to +153	Mothers against decapentaplegic homolog 4
<i>SMA</i>	chr18:51,029,915-51,030,359	445	67	64	70	64	-298 to +147	Single-strand selective monofunctional uracil DNA glycosylase
<i>SMU1</i>	chr12:54,188,843-54,189,292	450	53	61	65	66	-297 to +153	Snail family transcriptional repressor 1
<i>SNAI1</i>	chr20:49,982,773-49,983,210	438	68	69	68	68	-203 to +235	Zinc finger protein SNAI2
<i>SNAI2</i>	chr8:48,921,315-48,921,748	434	57	67	68	70	-307 to +127	Transcription initiation factor TFIIID subunit 1
<i>TAF1</i>	chrX:71,365,956-71,366,401	446	55	68	70	66	-283 to +163	G/T mismatch-specific thymine DNA glycosylase
<i>TDG</i>	chr12:103,965,499-103,965,940	442	59	64	64	69	-316 to +126	Cellular tumor antigen p53
<i>TP53</i>	chr17:7,687,409-7,687,855	447	53	62	67	69	-304 to +143	Three-prime repair exonuclease 1
<i>TRRX1</i>	chr3:48,465,929-48,466,361	448	67	68	66	69	-302 to +146	Transcription intermediary factor 1-beta
<i>TRIM28</i>	chr19:58,544,079-58,544,508	430	69	69	68	68	-390 to +40	Twist-related protein 1
<i>TWIS1</i>	chr7:19,117,521-19,117,969	449	65	69	70	68	-296 to +153	Twist-related protein 2
<i>TWIS2</i>	chr2:238,847,667-238,848,112	446	67	66	65	67	-365 to +81	Polyubiquitin-B
<i>UBB</i>	chr17:16,380,497-16,380,942	446	54	69	63	63	-282 to +164	Ubiquitin carboxyl-terminal hydrolase 47
<i>USP47</i>	chr11:11,841,122-11,841,567	446	58	60	64	65	-301 to +145	WD repeat-containing protein 48
<i>WDR48</i>	chr3:39,051,689-39,052,135	447	58	66	68	68	-297 to +150	Proto-oncogene Wnt-1
<i>WNT1</i>	chr12:48,978,163-48,978,609	447	74	68	68	68	-290 to +157	DNA repair protein complementing XP-C cells
<i>XPC</i>	chr3:14,178,508-14,178,952	445	59	63	68	67	-279 to +166	DNA repair protein XRCC2
<i>XRCC2</i>	chr7:152,676,039-152,676,482	444	59	67	70	65	-316 to +128	DNA repair protein XRCC3
<i>XRCC3</i>	chr14:103,715,325-10,3715,761	437	69	70	69	68	-274 to +163	Zinc finger E-box-binding
<i>ZEB1</i>	chr10:31,318,813-	443	65	69	69	70	-359 to +84	

31,319,255		homedbox 1		Bidirectional with CCT6B (two TSSs only 91 bp apart, target sequence is -219 to +227 relative to CCT6B TSS)	
ZNF830	chr17:34,961,234-34,961,679	446	57	62	66
				-306 to +140	

Gene Ontology term "metabolic process," filtered further for "DNA repair"
Known cancer-associated functions
Control loci with open chromatin

Gene	Genome coordinates (hg38)	Captured length (nt)	Capture d sequence GC content (%)	T _m			Captured sequence relative to TSS (+1)	Notes
				Oligo 1-T _m	Oligo 2-T _m	Oligo 3-T _m		
ALKBH2	chr12:109,093,335-109,094,280	946	54	65	65	65	-791 to +155	
APEX1	chr14:20,454,342-20,455,278	937	56	60	64	64	-789 to +148	
APEX2	chrX:54,999,546-55,000,485	940	55	63	67	68	-777 to +163	
ATM	chr11:108,222,047-108,222,984	938	58	66	68	68	-785 to +153	Divergently transcribed with NPAT (-346 to +592)
ATRX	chrX:77,786,121-77,787,050	930	43	64	64	65	-780 to +150	
BRCA1	chr17:43,125,334-43,126,270	937	51	69	65	67	-786 to +151	
BRCA2	chr13:32,314,697-32,315,631	935	57	65	64	67	-783 to +152	
CCN4	chr8:133,190,263-133,191,206	944	53	69	69	69	-776 to +168	
CD44	chr11:35,138,111-35,139,052	942	53	64	68	64	-759 to +183	
CEIN1	chr18:579,563-580,492	930	54	62	62	62	-780 to +150	
CHEK1	chr11:125,624,364-125,625,316	953	44	63	63	62	-772 to +181	
CHEK2	chr22:28,741,673-28,742,622	950	61	64	66	65	-787 to +163	
CSDPE1	chr11:114,757,871-114,758,815	945	47	62	64	67	-764 to +181	

<i>DDB2</i>	chr11:47,214,152-47,215,093	942	58	63	62	65	-790 to +152 -609 to +328 for EPMDAIP1; -650 to +287 for MLH1
<i>EPMDAIP1- MLH1</i>		937	59	67	69	70	TSSa MLH1
<i>EPSTH1</i>	chr3:36,992,868-36,993,803	934	47	63	66	65	-771 to +163
<i>EXO1</i>	chr13:42,992,151-42,993,084	951	53	63	66	65	-824 to +127
<i>FEN1</i>	chr1:241,847,367-241,848,317	942	54	64	66	63	-791 to +151
<i>GSTA1</i>	chr11:61,791,846-61,792,787	939	39	61	65	58	-788 to +151
<i>HIST1H1B</i>	chr6:52,803,818-52,804,756	952	45	63	65	64	-759 to +193
<i>HOTAIR</i>	chr6:27,867,390-27,868,341	959	53	68	68	67	-807 to +152
<i>HOXB2</i>	chr12:53968605-53969564	935	49	64	62	64	-783 to +152
<i>HSPA45</i>	chr17:48,544,881-48,545,815	951	59	66	66	67	-823 to +128
<i>LIG1</i>	chr9:125,241,260-125,242,210	941	65	64	65	65	-776 to +165
<i>MBD4</i>	chr19:48,170,440-48,171,380	951	56	64	65	65	-802 to +149
<i>MCM9</i>	chr3:129,440,032-129,440,982	939	33	60	61	60	-789 to +150
<i>MECP2</i>	chr6:118,931,590-118,932,528 chrX:154,097,673-154,098,629	957	59	65	68	65	-897 to +60
<i>MRE11</i>	chr11:94,493,724-94,494,679	956	54	64	68	63	-804 to +152
<i>MSH2</i>	chr2:47,402,248-47,403,177	930	52	69	68	67	-819 to +111
<i>MSH4</i>	chr1:75,796,091-75,797,021	931	46	64	68	62	-780 to +151
<i>MSH5</i>	chr6:31,739,162-31,740,092	931	53	65	70	65	-786 to +145
<i>OGG1</i>	chr3:9,749,158-9,750,103	946	61	66	70	70	-786 to +160
<i>OLLIG2</i>	chr21:33,025,108-33,026,059	952	63	65	63	68	-800 to +152
<i>PELO</i>	chr5:52,787,147-52,788,092	946	50	63	63	67	-793 to +153
<i>PMS2</i>	chr7:6,008,985-6,009,928	944	63	68	67	69	-822 to +122
<i>PMS2P5</i>	chr7:74,889,984-74,890,918	935	64	68	69	66	-784 to

Control promoter with open chromatin

Gene	Genome coordinates (mm10)	Captured length (nt)	Capture sequence GC content (%)	T _m			Captured sequence relative to TSS (+1)	Notes
				Oligo 1-T	Oligo 2-T	Oligo 3-T		
<i>Ash1l</i>	chr3:88,950,129-88,950,957	829	73	65.1	61.4	64.3	-493 to +336	
<i>Btk</i>	chrX:134,583,042-134,583,555	514	50	61	60.8	63.4	-415 to +99	
<i>Ccl5</i>	chr11:83,530,364-83,530,887	524	54	61.8	64.4	60.8	-369 to +155	
<i>Cd180</i>	chr13:102,693,205-102,693,713	509	46	63.2	60.8	64.6	-353 to +156	
<i>Cox8a</i>	chr19:7,217,511-7,218,156	646	52	59.5	65.3	61.5	-540 to +106	Control promoter with open chromatin
<i>Cxcl15</i>	chr5:90,794,068-90,794,625	558	44	58.6	57	63	-466 to +92	
<i>Cxcr4</i>	chr1:128,592,199-128,592,690	492	57	62.6	60.4	63.4	-400 to +92	
<i>Dnm2b</i>	chr2:153,648,958-153,649,971	987	70	62.7	62	64.4	-492 to +495	
<i>Foxo1</i>	chr3:52,267,705-52,268,609	905	65	64.1	58.5	64.6	-631 to +274	Divergently transcribed with <i>Hdac3</i> (-420 to +117)
<i>Gdf5</i>	chr2:155,945,300-155,945,790	491	46	65.2	64.5	58.5	-423 to +68	
<i>Rela</i>	chr18:37,954,900-37,955,436	537	61	62.7	65.3	61.6	-250 to +287 (Rela)	
<i>Hsf1</i>	chr15:76,477,070-76,477,543	474	67	64.9	62.1	61.9	-375 to +99	Divergently transcribed with <i>Bop1</i> (-274 to +200)

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

<i>Mapk13</i>	chr17:28,768,866-28,769,660	795	65	59.6	59.1	62.1	-431 to +364
<i>Mapk14</i>	chr17:28,690,874-28,691,764	891	68	62.7	63.4	65.1	-455 to +436
<i>Mapk15</i>	chr15:75,995,428-75,995,983	556	57	61.1	64.8	64.8	+1660 to +2215
<i>Mapk3</i>	chr7:126,759,191-126,759,897	707	67	58.2	63.6	61.2	-435 to +272
<i>Mapk4</i>	chr18:74,064,823-74,065,577	753	69	57.9	65.1	59.5	-628 to +127
<i>Mapk6</i>	chr9:75,409,879-75,410,441	563	68	61.4	62.6	58.5	-436 to +127
<i>Mapk8</i>	chr14:33,446,723-33,447,594	872	72	63.8	64.9	64.8	-436 to +436
<i>Mme</i>	chr5:63,295,269-63,295,864	596	65	59.8	61.6	63.2	-166 to +430
<i>Myc88</i>	chr9:119,339,910-119,340,537	628	51	64.9	62.4	64.2	-496 to +132
<i>Nfkb1</i>	chr3:135,691,529-135,692,148	620	61	65.3	63.6	64.7	-602 to +18
<i>Nfkb2</i>	chr19:46,303,888-46,304,398	511	50	58.4	62.8	58.1	-432 to +79
<i>Nfkbia</i>	chr12:55,492,417-55,493,106	690	60	62.6	61.9	62.2	-459 to +231
<i>Nfkbib</i>	chr7:28,766,590-28,767,119	530	57	62.8	65	59.7	-475 to +55
<i>Nr2c2</i>	chr6:92,091,014-92,091,520	507	67	62.7	61	64.3	-376 to +131
<i>Pik3ap1</i>	chr19:41,384,869-41,385,481	613	67	65.2	57	65.2	-385 to +228
<i>Pik3c2a</i>	chr7:116,443,198-116,444,081	883	65	64.8	59.1	63.9	-632 to +251
<i>Pik3c2g</i>	chr6:139,736,328-139,736,987	660	37	58.3	65.3	62.9	-567 to +93
<i>Pik3c3</i>	chr18:30,272,232-30,272,871	640	39	59.3	63.9	63	-515 to +125
<i>Pik3ca</i>	chr3:32,397,129-32,397,798	670	75	58.9	64.1	59.9	-542 to +128
<i>Pik3r1</i>	chr13:101,768,008-101,768,630	623	71	60.1	65.3	65.8	-413 to +210
<i>Pik3r2</i>	chr8:70776347-70777162	815	66	64.9	64.5	65.2	-449 to +366
<i>Pik3r3</i>	chr4:116,221,346-116,221,950	605	57	63.8	61.7	61.8	-315 to +290
<i>Pik3r5</i>	chr11:68,431,620-68,432,347	728	60	63.5	63.8	65.1	-501 to +227
<i>Pik3r6</i>	chr11:68,502,642-68,503,181	540	51	61.6	63.4	65.3	-377 to +163
<i>Pip6</i>	chr6:124,738,642-124,739,124	483	59	60.1	59.7	62.1	-415 to +68
<i>Rela</i>	chr19:5,637,111-5,637,764	654	71	60	65.6	55.4	-372 to +282
<i>Src</i>	chr2:157,417,873-157,418,488	616	51	61.3	60.7	65.5	-571 to +45
<i>Stat3</i>	chr11:100,939,384-100,939,932	549	60	61.6	64.1	58.4	-392 to +157
<i>Tab1</i>	chr15:80,132,766-80,133,525	760	63	59.9	64.4	64.5	-361 to +399
<i>Tab2</i>	chr10:7,956,254-7,956,828	575	78	61.1	57.2	55.4	-705 to +131

Exon 5 of 14

<i>Tet2</i>	chr3:133,544,278-133,544,951	674	75	65.4	60.4	59.6	-561 to +113
<i>Tlr2</i>	chr3:83,841,701-83,842,365	665	47	63.7	64.1	64.2	-598 to +67
<i>Tlr4</i>	chr4:66,827,250-66,827,825	576	48	61.1	58.3	62.1	-334 to +242
<i>Tnfr</i>	chr17:35,201,859-35,202,357	499	55	59.4	61.8	64.9	-350 to +149
<i>Trqf6</i>	chr2:101,678,076-101,678,618	543	58	57	64.2	61	-364 to +179
<i>Vav1</i>	chr17:57,278,729-57,279,333	605	61	62.1	64	60.9	-371 to +234

Table 3: CCS reads aligned to 11 human targets of ~300 nt vs. ~450 nt used in FENGCC assay development with standard PCR vs. BS-PCR.

~300-nt targets aligned CCS read number ^a						
Sample/Promoter	~300 nt, Taq, PCR-1	~300 nt, Taq, PCR-2	~300 nt, FEN1, PCR-1	~300 nt, FEN1, PCR-2	~300 nt, FEN1, BS-PCR-1	~300 nt, FEN1, BS-PCR-2
<i>EPM2AIP1</i>	3896	1994	27	34	165	54
<i>EXO1</i>	37	17	3215	2913	2500	3105
<i>HSPA5</i>	21	10	1534	1528	1311	1149
<i>MLH1</i>	5081	2545	456	603	1477	2054
<i>MLH3</i>	141	171	1355	1743	48	34
<i>MSH2</i>	1395	643	810	729	287	159
<i>MSH3</i>	0	0	0	1	0	1
<i>MSH6</i>	0	0	0	0	0	0
<i>PMS1</i>	3140	1759	1610	2500	104	45
<i>PMS2</i>	210	149	716	1059	304	278
<i>TDG</i>	2018	1068	2537	2396	59	33
~450-nt targets aligned CCS read number ^a						
Sample/Promoter	~450 nt, Taq, PCR-1	~450 nt, Taq, PCR-2	~450 nt, FEN1, PCR-1	~450 nt, FEN1, PCR-2	~450 nt, FEN1, BS-PCR-1	~450 nt, FEN1, BS-PCR-2
<i>EPM2AIP1</i>	4321	3865	236	156	109	41
<i>EXO1</i>	14	21	2678	3371	318	234
<i>HSPA5</i>	103	55	2859	3607	933	501
<i>MLH1</i>	540	277	1682	1362	2325	1527
<i>MLH3</i>	31	33	111	208	5	2
<i>MSH2</i>	454	398	2196	2132	121	50
<i>MSH3</i>	0	0	0	0	6	3
<i>MSH6</i>	0	0	0	0	7	10
<i>PMS1</i>	454	347	817	971	4	10
<i>PMS2</i>	11	14	1150	1431	194	113
<i>TDG</i>	22	28	1148	1259	8	1
Aligned CCS read number ^a						
Sample/Promoter	BS-PCR control-1	BS-PCR control-2				

<i>MLH1</i>	5021	2780				
^a High-fidelity CCS reads with ≥ 5 SMRT sequencing passes.						
^b Suffixes -1 and -2 refer to the two independent biological replicates.						
^c PCR control was a PCR product amplified from bisulfite-converted gDNA.						

Table 4: CCS reads on- and off-target for 119 ~450-nt human targets in MAPit-FENGCC.^{a,b}

Measurement/Sample	NSC-1	NSC-2	L0-1	L0-2	0.1%:99.9% L0:NSC-1	0.1%:99.9% L0:NSC-2	Nx18-25-1	Nx18-25-2
Input genomic DNA (ng)	500	500	500	500	500	500	500	500
Total reads	91,751	189,941	15,937	20,601	67,757	53,670	66,882	54,385
Reads aligned on genome	85,408	168,764	14,223	18,884	59,928	48,017	59,286	47,266
% Reads aligned on genome	93.1%	88.9%	89.2%	91.7%	88.4%	89.5%	88.6%	86.9%
Reads aligned on target	77,014	156,295	13,056	17,642	55,750	41,368	53,392	43,269
% Reads off target but on genome	9.8%	7.4%	8.2%	6.6%	7.0%	13.8%	9.9%	8.5%
% Total reads on target	83.9%	82.3%	81.9%	85.6%	82.3%	77.1%	79.8%	79.6%
Number of targets detected ^c	111		105		109		109	
Number of targets detected using reads with ≥95% coverage of each reference sequence and ≥95% conversion	105		92		103		104	

^a Genomic DNA from M.CviPI-treated cells subject to EM-PCR for detection of DNA methylation.

^b High-fidelity CCS reads with ≥ 3 SMRT sequencing passes.

^c Suffixes -1 and -2 refer to the two independent biological replicates.

^d At least one read in at least one condition.

Measurement/Sample	NSC-1	NSC-2	Nx18-25-1	Nx18-25-2
Input genomic DNA (ng)	500	500	500	500
Total reads	32,701	27,442	25,062	22,905
Reads aligned on genome	32,512	27,191	22,819	22,652
% Reads aligned on genome	99.4%	99.1%	91.1%	98.9%
Reads aligned on target	28,414	22,963	18,170	20,377
% Reads off target but on genome	12.6%	15.5%	20.4%	10.0%
% Total reads on target	86.9%	83.7%	72.5%	89.0%
Number of targets detected ^d	97			

^a Genomic DNA from M.CviPI-treated cells subject to standard PCR for genotyping.

^b High-fidelity CCS reads with ≥ 3 SMRT sequencing passes.

^c Suffixes -1 and -2 refer to the two independent biological replicates.

^d At least one read in at least one condition.

Upon received as blank

Table 5: Filtered CCS reads aligned to 119 ~450-nt human targets in MAPit-FENGCC using EM-seq.^a

	Filtered CCS read number ^b							
Promoter/Sample	NSC-1	NSC-2	L0-1	L0-2	0.1%:99.9% L0:NSC-1	0.1%:99.9% L0:NSC-2	Nx18-25-1	Nx18-25-2
<i>ALKBH1</i>	0	6	0	0	2	0	1	0
<i>ALKBH2</i>	3152	7899	476	503	5595	2842	3381	3072
<i>ALKBH3</i>	5	6	0	0	3	3	2	2
<i>APEX1</i>	159	421	33	32	198	70	124	81
<i>APEX2</i>	701	2995	318	365	1928	813	5865	4908
<i>ATM</i>	807	1866	70	93	418	195	1292	964
<i>ATRX</i>	202	730	40	62	194	79	936	623
<i>ATXN3</i>	0	0	0	0	0	0	0	0
<i>AXIN2</i>	0	0	0	0	0	0	0	0
<i>BRCA1</i>	398	712	217	179	709	236	560	479
<i>BRCA2</i>	129	778	226	218	445	280	785	539
<i>CCN4</i>	333	783	31	66	456	147	566	577
<i>CD44</i>	48	93	7	15	46	33	53	43
<i>CDH1</i>	6	15	0	1	9	4	8	8
<i>CDH13</i>	0	0	0	0	0	0	0	0
<i>CDH2</i>	15	9	0	1	9	4	0	3
<i>CDKN2A</i>	0	0	0	0	0	0	0	0
<i>CDKN2D</i>	2	0	0	0	1	0	0	0
<i>CETN1</i>	280	1555	35	53	133	103	472	434
<i>CHEK1</i>	184	299	22	40	177	101	233	131
<i>CHEK2</i>	3197	12014	1887	1982	4333	2314	5409	5098
<i>CSDE1</i>	63	112	57	62	84	43	145	128
<i>DDB2</i>	69	412	17	20	73	38	182	150
<i>DRD3</i>	48	62	0	2	7	1	26	26
<i>DRD4</i>	0	0	4	6	3	0	0	1
<i>EGFR</i>	0	0	0	0	0	0	1	0
<i>EPM2AIP1</i>	1138	2322	1961	1928	1183	598	687	518
<i>EPSTU1</i>	70	135	13	17	51	33	142	80
<i>ERCC1</i>	249	281	93	137	915	362	422	383
<i>ERCC2</i>	7	6	7	3	10	6	13	23
<i>ERCC8</i>	613	2453	67	82	1295	458	1351	1386
<i>EXO1</i>	4	16	12	24	17	12	30	25
<i>FANCC</i>	1	11	0	1	2	0	1	2

<i>FAT4</i>	0	0	0	0	0	0	0	0
<i>FEN1</i>	705	1828	153	181	986	502	466	480
<i>FHIT</i>	46	95	9	7	71	36	22	26
<i>FOXC2</i>	9	6	2	0	14	10	1	3
<i>FZD7</i>	0	0	0	0	0	0	0	0
<i>FZD8</i>	7	3	0	0	2	4	2	3
<i>GPR158</i>	10	31	6	8	10	11	6	8
<i>GSTA1</i>	1580	2952	54	52	780	320	2468	1718
<i>HIST1H1B</i>	293	633	19	34	371	121	430	523
<i>HMG A2</i>	1	2	1	1	0	0	1	1
<i>HOTAIR</i>	1063	1877	448	528	1175	597	1546	968
<i>HOXB2</i>	31	43	104	140	59	26	57	24
<i>HSPA5</i>	0	28	84	87	3	1	28	11
<i>INO80</i>	3	1	0	0	0	1	0	6
<i>KLF8</i>	0	5	0	0	2	0	0	0
<i>LIG1</i>	14	34	2	3	2	2	21	17
<i>MBD4</i>	50	144	19	27	96	30	112	93
<i>MCM9</i>	655	1721	29	43	267	157	1008	977
<i>MECP2</i>	7	31	0	0	4	3	6	7
<i>MEG3</i>	97	84	8	8	19	19	48	47
<i>MGMT</i>	15	29	0	2	42	31	5	3
<i>MLH1</i>	2872	8950	1339	1532	1741	2054	3558	3287
<i>MLH3</i>	7	15	0	2	2	1	3	6
<i>MNAT1</i>	35	142	5	8	9	6	194	116
<i>MPG</i>	0	1	0	1	4	0	1	0
<i>MRE11</i>	435	1215	43	83	595	299	500	485
<i>MSH2</i>	1117	1939	185	242	2485	801	1531	1036
<i>MSH3</i>	2	23	1	9	9	12	8	8
<i>MSH4</i>	13	99	12	33	21	32	33	25
<i>MSH5</i>	132	241	24	57	198	107	170	192
<i>MSH6</i>	5	13	1	10	0	7	9	10
<i>MUTYH</i>	7	2	0	3	0	2	7	14
<i>NEIL2</i>	0	0	0	0	0	0	0	0
<i>NEIL3</i>	1	1	0	0	0	0	3	0
<i>NKX2-5</i>	0	5	0	0	0	1	0	0
<i>OGG1</i>	131	275	617	549	700	397	189	183
<i>OLIG2</i>	116	291	6	19	46	51	86	86
<i>PARP1</i>	0	11	1	1	0	1	2	0
<i>PCNA</i>	408	552	20	18	66	109	358	216
<i>PELO</i>	103	235	17	23	121	89	89	95
<i>PMS1</i>	27	63	0	1	8	12	33	19

<i>PMS2</i>	56	214	30	30	68	93	82	72
<i>PMS2P1</i>	26	124	16	28	56	33	32	30
<i>PMS2P11</i>	501	991	25	52	340	147	680	489
<i>PMS2P2</i>	5	17	0	1	5	3	12	9
<i>PMS2P3</i>	33	58	336	324	67	36	64	41
<i>PMS2P5</i>	1078	5423	423	547	2232	1425	3784	3841
<i>POLD1</i>	0	0	2	2	2	0	0	1
<i>POLD4</i>	1491	1719	1043	1096	3924	1853	1491	1441
<i>POLG</i>	0	0	0	0	0	0	0	0
<i>POLK</i>	13	82	1	2	5	2	41	31
<i>POLQ</i>	215	456	56	89	195	66	595	469
<i>POLR2B</i>	47	76	9	12	40	32	15	19
<i>POLR2G</i>	3	24	4	0	2	5	5	4
<i>PRDM2</i>	407	641	47	50	310	135	662	413
<i>PRKCG</i>	24	7	226	209	89	33	47	33
<i>PRUNE2</i>	238	673	24	28	247	145	280	303
<i>PXMP4</i>	1	3	0	1	0	2	7	6
<i>PYCARD</i>	0	3	3	4	3	2	3	2
<i>RNASEH2A</i>	11	4	0	6	8	4	10	9
<i>RNF111</i>	130	294	12	5	56	31	159	126
<i>RPA2</i>	1026	5542	890	878	1755	1012	4577	3062
<i>RPA3</i>	187	206	1	8	68	36	234	154
<i>RPA4</i>	3285	10938	423	471	5404	2499	3858	3811
<i>SH3TC1</i>	11	10	59	74	44	14	11	5
<i>SIM2</i>	0	3	0	0	4	2	2	0
<i>SMAD4</i>	12	3	0	0	5	5	1	4
<i>SMUG1</i>	8	24	1	1	10	2	16	10
<i>SNAI1</i>	0	0	0	0	0	0	0	0
<i>SNAI2</i>	1	4	0	3	2	1	1	0
<i>TAF1</i>	168	370	88	134	216	102	711	546
<i>TDG</i>	6	3	0	2	5	1	3	4
<i>TP53</i>	59	144	14	21	72	38	160	107
<i>TREX1</i>	0	3	0	0	0	0	2	0
<i>TRIM28</i>	32	95	7	6	24	22	70	42
<i>TWIST1</i>	4	2	0	0	3	4	5	1
<i>TWIST2</i>	9	10	1	1	16	4	4	3
<i>UBB</i>	66	118	3	5	17	22	43	32
<i>USP47</i>	7	6	0	2	1	0	2	0
<i>WDR48</i>	3	9	0	0	2	0	6	1
<i>WNT1</i>	71	55	82	72	106	57	18	12
<i>XPC</i>	4	0	0	0	0	0	0	0

<i>XRCC2</i>	9	20	28	29	21	11	23	41
<i>XRCC3</i>	0	0	0	0	0	0	0	0
<i>ZEB1</i>	0	0	0	0	0	0	0	0
<i>ZNF830</i>	6095	24059	1156	1234	11769	7706	10358	8753

^a High-fidelity CCS reads with ≥ 5 SMRT sequencing passes.

^b Filtered for $\geq 95\%$ coverage of each reference sequence and $\geq 95\%$ HCH conversion.

^c Suffixes -1 and -2 refer to the two independent biological replicates.

Table 6: Differential H5mCG and G5mCH determined by MAPit-FENG C using EM-seq of ~450-nt human targets with ≥ 50 filtered reads.^a

Differential DNA methylation (H5mCG)						
Promoter/Parameter	Read number ^b	Methylation change ^c	SE ^d	Wald	P value ^e	Adjusted P value ^f
<i>CCN4</i>	333	-0.085	0.0098	73.8911	< 0.0001	< 0.0001
<i>OLIG2</i>	86	-0.043	0.0027	256.219	< 0.0001	< 0.0001
<i>EPSTI1</i>	70	-0.355	0.0412	74.0241	< 0.0001	< 0.0001
<i>FHIT</i>	22	-0.017	0.0108	2.3591	0.1246	1
<i>RPA3</i>	154	-0.231	0.0336	47.4977	< 0.0001	< 0.0001
<i>PMS2P1</i>	26	-0.003	0.0024	1.5259	0.2167	1
<i>HDXB2</i>	24	0.506	0.0021	57278.7	< 0.0001	< 0.0001
<i>HIST1H1B</i>	293	0.041	0.0008	2569.17	< 0.0001	< 0.0001
<i>PMS2P3</i>	33	-0.010	0.0088	1.1892	0.2755	1
<i>ATM</i>	807	-0.024	0.0055	18.6618	< 0.0001	0.0011
<i>MCM9</i>	655	0.109	0.0038	799.869	< 0.0001	< 0.0001
<i>MBD4</i>	50	-0.023	0.0161	2.0394	0.1533	1
<i>MEG3</i>	47	0.009	0.0244	0.122	0.7269	1
<i>RPA2</i>	1026	0.025	0.0015	263.429	< 0.0001	< 0.0001
<i>UBB</i>	32	-0.050	0.0069	53.7674	< 0.0001	< 0.0001
<i>FEN1</i>	466	-0.011	0.0013	80.5095	< 0.0001	< 0.0001
<i>CHEK2</i>	3197	-0.005	0.0006	69.6601	< 0.0001	< 0.0001
<i>RPA4</i>	3285	-0.111	0.0018	3966.6	< 0.0001	< 0.0001
<i>ATRX</i>	202	-0.035	0.0102	11.704	0.0006	0.0424
<i>MNAT1</i>	35	0.006	0.0006	114.129	< 0.0001	< 0.0001
<i>BRCA2</i>	129	0.044	0.0049	79.2924	< 0.0001	< 0.0001
<i>DRD3</i>	26	-0.015	0.0282	0.3002	0.5838	1
<i>MSH2</i>	1036	-0.011	0.0036	9.5303	0.002	0.1374
<i>CETN1</i>	280	0.036	0.011	10.9478	0.0009	0.0637
<i>PCNA</i>	216	-0.016	0.001	289.171	< 0.0001	< 0.0001
<i>MLH1</i>	2872	-0.006	0.0017	11.7929	0.0006	0.0404
<i>BRCA1</i>	398	-0.004	0.0012	12.9407	0.0003	0.0219
<i>DDX2</i>	69	-0.068	0.0012	3029.65	< 0.0001	< 0.0001
<i>POLD4</i>	1491	-0.028	0.0059	22.3205	< 0.0001	0.0002
<i>POLQ</i>	215	-0.004	0.0007	23.7005	< 0.0001	0.0001
<i>MSH5</i>	148	-0.017	0.0013	159.303	< 0.0001	< 0.0001
<i>TP53</i>	59	0.000	0.0019	0.0127	0.9103	< 0.0001
<i>ERCC1</i>	249	0.000	0.0004	1.1144	0.2911	< 0.0001

<i>PMS2P5</i>	1078	-0.007	0.0005	254.038	< 0.0001	< 0.0001
<i>PRDM2</i>	407	0.031	0.0025	417.962	< 0.0001	< 0.0001
<i>MRE11</i>	435	-0.006	0.0014	19.5723	< 0.0001	0.0007
<i>CSDE1</i>	63	-0.005	0.0037	2.1073	0.1466	1
<i>ALKBH2</i>	3072	-0.047	0.0013	1234.02	< 0.0001	< 0.0001
<i>RNF111</i>	126	-0.006	0.0008	61.5392	< 0.0001	< 0.0001
<i>APEX2</i>	701	0.145	0.0125	134.271	< 0.0001	< 0.0001
<i>HOTAIR</i>	968	0.491	0.0032	23831.2	< 0.0001	< 0.0001
<i>ZNF830</i>	6095	-0.004	0.0007	31.3671	< 0.0001	< 0.0001
<i>APEX1</i>	81	-0.008	0.0027	9.2998	0.0023	0.1558
<i>CHEK1</i>	131	-0.051	0.0019	711.188	< 0.0001	< 0.0001
<i>HIST1H1B (TSS) ^g</i>	293	0.033	0.0009	1295.97	< 0.0001	< 0.0001
<i>ERCC8</i>	613	-0.011	0.0003	1340.62	< 0.0001	< 0.0001
<i>GSTA1</i>	1580	-0.148	0.0028	73279.5	< 0.0001	< 0.0001
<i>OGG1</i>	131	-0.004	0.0019	4.2566	0.0391	1
<i>PMS2</i>	56	-0.003	0.0018	3.6271	0.0568	1
<i>TRIM28</i>	32	-0.031	0.0124	6.0277	0.0141	0.9576
<i>EPM2AIP1</i>	518	0.007	0.0014	21.6546	< 0.0001	0.0002
<i>EAF1</i>	168	0.233	0.0168	193.082	< 0.0001	< 0.0001
<i>PELO</i>	89	0.067	0.0021	965.227	< 0.0001	< 0.0001
<i>PRUNE2</i>	238	-0.065	0.0084	60.2971	< 0.0001	< 0.0001
<i>CD44</i>	43	0.575	0.0327	309.346	< 0.0001	< 0.0001

Differential Accessibility (G5mCH)					
Promoter/Parameter	Methylation change ^c	SE ^d	Wald	P value ^e	Adjusted P value ^f
<i>CCN4</i>	0.079	0.0333	5.564	0.0183	1
<i>OLIG2</i>	0.067	0.019	12.4442	0.0004	0.0285
<i>EPSTH1</i>	0.051	0.0278	3.3488	0.0673	1
<i>FHIT</i>	0.038	0.0172	4.9677	0.0258	1
<i>RPA3</i>	0.027	0.0099	7.4999	0.0062	0.4196
<i>PMS2P1</i>	0.015	0.0265	0.3395	0.5601	1
<i>HOXB2</i>	0.015	0.0049	9.255	0.0023	0.1597
<i>HIST1H1B</i>	-0.046	0.0067	46.4778	< 0.0001	< 0.0001
<i>PMS2P3</i>	0.006	0.071	0.006	0.9382	1
<i>ATM</i>	0.003	0.0087	0.1242	0.7246	1
<i>MCM9</i>	-0.001	0.0109	0.0106	0.9182	1
<i>MBD4</i>	-0.006	0.019	0.0972	0.7552	1
<i>MEG3</i>	-0.008	0.0157	0.2823	0.5952	1
<i>RPA2</i>	-0.010	0.0078	1.5891	0.2075	1

<i>UBB</i>	-0.017	0.0232	0.5242	0.469	1
<i>FEN1</i>	-0.019	0.0038	25.1641	< 0.0001	< 0.0001
<i>CHEK2</i>	-0.020	0.0047	18.05	< 0.0001	0.0015
<i>RPA4</i>	-0.027	0.0061	20.3553	< 0.0001	0.0004
<i>ATRX</i>	-0.030	0.0287	1.0965	0.295	1
<i>MNAT1</i>	-0.032	0.0055	33.6365	< 0.0001	< 0.0001
<i>BRCA2</i>	-0.032	0.0215	2.2664	0.1322	1
<i>DRD3</i>	-0.034	0.0118	8.3239	0.0039	0.2661
<i>MSH2</i>	-0.035	0.0099	12.0732	0.0005	0.0348
<i>CETN1</i>	-0.036	0.007	26.6114	< 0.0001	0.0001
<i>PCNA</i>	-0.037	0.0116	10.3639	0.0013	0.0874
<i>MLH1</i>	-0.038	0.0118	10.2262	0.0014	0.0942
<i>BRCA1</i>	-0.040	0.0132	8.9771	0.0027	0.1859
<i>DDB2</i>	-0.043	0.0231	3.5194	0.0607	1
<i>POLD4</i>	-0.044	0.0252	2.991	0.0837	1
<i>POLQ</i>	-0.045	0.005	81.5233	< 0.0001	< 0.0001
<i>MSH5</i>	-0.046	0.0109	17.5415	< 0.0001	0.0019
<i>TP53</i>	-0.046	0.0074	39.6643	< 0.0001	0.0001
<i>ERCC1</i>	-0.050	0.0142	12.587	0.0004	0.0264
<i>PMS2P5</i>	-0.050	0.0098	26.5186	< 0.0001	< 0.0001
<i>PRDM2</i>	-0.051	0.005	107.727	< 0.0001	< 0.0001
<i>MRE11</i>	-0.053	0.0076	48.6682	< 0.0001	< 0.0001
<i>CSDE1</i>	-0.057	0.0122	21.6445	< 0.0001	0.0002
<i>ALKBH2</i>	-0.059	0.0041	203.229	< 0.0001	< 0.0001
<i>RNF111</i>	-0.061	0.0179	11.556	0.0007	0.0459
<i>APEX2</i>	-0.066	0.0192	11.6487	0.0006	0.0437
<i>HOTAIR</i>	-0.067	0.0049	189.452	< 0.0001	< 0.0001
<i>ZNF830</i>	-0.067	0.0058	133.521	< 0.0001	< 0.0001
<i>APEX1</i>	-0.068	0.0158	18.2191	< 0.0001	0.0013
<i>CHEK1</i>	-0.072	0.0183	15.6263	0.0001	0.0052
<i>HIST1H1B (TSS) *</i>	-0.074	0.0017	1886.56	< 0.0001	< 0.0001
<i>ERCC8</i>	-0.074	0.0077	91.6536	< 0.0001	< 0.0001
<i>GSTA1</i>	-0.080	0.0057	192.088	< 0.0001	< 0.0001
<i>OGG1</i>	-0.084	0.0209	16.0631	< 0.0001	0.0042
<i>PMS2</i>	-0.086	0.0132	42.1014	< 0.0001	< 0.0001
<i>TRIM28</i>	-0.087	0.072	1.4488	0.2287	1
<i>EPM2AIP1</i>	-0.102	0.0191	28.259	< 0.0001	< 0.0001
<i>TAF1</i>	-0.102	0.0051	401.629	< 0.0001	< 0.0001
<i>PELO</i>	-0.113	0.0153	53.8195	< 0.0001	< 0.0001
<i>PRUNE2</i>	-0.113	0.0108	108.798	< 0.0001	< 0.0001
<i>CD44</i>	-0.201	0.0054	1398.66	< 0.0001	< 0.0001

-
-

^a For loci in which ≥ 22 reads were obtained in each of the four samples, i.e., both NSC and both Nx18-15 replicates.

^b Read number of sample with the least amount of reads.

^c Proportion of difference in H5CG or accessibility (GBM Nx18-25 minus NSC non-cancerous cells).

^d SE, standard error.

^e All *P* values $< .0001$ (non-corrected and corrected) were rounded off to $< .0001$.

^f Bonferroni method corrected ($\alpha = .05$).

^g *P* value calculated for HCG or GCH sites for -80 to $+45$ relative to TSS within full-length target from -246 to $+199$.

Differential DNA methylation (H5mCG)						
Promoter/Parameter	Read number ^b	Methylation change ^c	SE ^d	Wald	<i>P</i> value ^e	Adjusted <i>P</i> value ^f
<i>CD44</i>	43	0.575	0.0327	309.346	< 0.0001	< 0.0001
<i>HOXB2</i>	24	0.506	0.0021	57278.7	< 0.0001	< 0.0001
<i>HOTAIR</i>	968	0.491	0.0032	23831.2	< 0.0001	< 0.0001
<i>TAF1</i>	168	0.233	0.0168	193.082	< 0.0001	< 0.0001
<i>APEX2</i>	701	0.145	0.0125	134.271	< 0.0001	< 0.0001
<i>MCM9</i>	655	0.109	0.0038	799.869	< 0.0001	< 0.0001
<i>PELO</i>	89	0.067	0.0021	965.227	< 0.0001	< 0.0001
<i>PRDM2</i>	407	0.051	0.0025	417.962	< 0.0001	< 0.0001
<i>HIST1H1B</i> (TSS) ^g	293	0.033	0.0009	1295.97	< 0.0001	< 0.0001
<i>EPM2AIP1</i>	518	0.007	0.0014	21.6546	< 0.0001	0.0002
<i>ERCC1</i>	249	0.000	0.0004	1.1144	0.2911	< 0.0001
<i>PMS2</i>	56	-0.003	0.0018	3.6271	0.0568	1
<i>ZNF830</i>	6095	-0.004	0.0007	31.3671	< 0.0001	< 0.0001
<i>OGG1</i>	131	-0.004	0.0019	4.2566	0.0391	1
<i>CSDE1</i>	63	-0.005	0.0037	2.1073	0.1466	1
<i>MRE11</i>	435	-0.006	0.0014	19.5723	< 0.0001	0.0007
<i>RNF111</i>	126	-0.006	0.0008	61.5392	< 0.0001	< 0.0001
<i>PMS2P5</i>	1078	-0.007	0.0005	254.038	< 0.0001	< 0.0001
<i>APEX1</i>	81	-0.008	0.0027	9.2998	0.0023	0.1558
<i>ERCC8</i>	613	-0.011	0.0003	1340.62	< 0.0001	< 0.0001
<i>OLIG2</i>	86	-0.043	0.0027	256.219	< 0.0001	< 0.0001
<i>ALKBH2</i>	3072	-0.047	0.0013	1234.02	< 0.0001	< 0.0001
<i>UBB</i>	32	-0.050	0.0069	53.7674	< 0.0001	< 0.0001
<i>CHEK1</i>	131	-0.051	0.0019	711.188	< 0.0001	< 0.0001

<i>PRUNE2</i>	238	-0.065	0.0084	60.2971	<0.0001	<0.0001
<i>DDB2</i>	69	-0.068	0.0012	3029.65	<0.0001	<0.0001
<i>CCN4</i>	333	-0.085	0.0098	73.8911	<0.0001	<0.0001
<i>RPA4</i>	3285	-0.111	0.0018	3966.6	<0.0001	<0.0001
<i>RPA3</i>	154	-0.231	0.0336	47.4977	<0.0001	<0.0001
<i>EPSTH1</i>	70	-0.355	0.0412	74.0241	<0.0001	<0.0001
<i>GSTA1</i>	1580	-0.748	0.0028	73279.5	<0.0001	<0.0001

Differential Accessibility (G5mCH)					
Promoter/Parameter	Methylation change ^c	SE ^d	Wald	P value ^e	Adjusted P value ^f
<i>CD44</i>	-0.201	0.0054	1398.66	<0.0001	<0.0001
<i>HOXB2</i>	0.015	0.0049	9.255	0.0023	0.1597
<i>HOTAIR</i>	-0.067	0.0049	189.452	<0.0001	<0.0001
<i>TAF1</i>	-0.102	0.0051	401.629	<0.0001	<0.0001
<i>APEX2</i>	-0.066	0.0192	11.6487	0.0006	0.0437
<i>MCM9</i>	-0.001	0.0109	0.0106	0.9182	1
<i>PELO</i>	-0.113	0.0153	53.8195	<0.0001	<0.0001
<i>PRDM2</i>	-0.051	0.005	107.727	<0.0001	<0.0001
<i>HIST1H1B (TSS) ^g</i>	-0.074	0.0017	1886.56	<0.0001	<0.0001
<i>EPM2AIP1</i>	-0.102	0.0191	28.259	<0.0001	<0.0001
<i>ERCC1</i>	-0.050	0.0142	12.587	0.0004	0.0264
<i>PMS2</i>	-0.086	0.0132	42.1014	<0.0001	<0.0001
<i>ZNF830</i>	-0.067	0.0058	133.521	<0.0001	<0.0001
<i>OGG1</i>	-0.084	0.0209	16.0631	<0.0001	0.0042
<i>CSDE1</i>	-0.057	0.0122	21.6445	<0.0001	0.0002
<i>MRE11</i>	-0.053	0.0076	48.6682	<0.0001	<0.0001
<i>RNF111</i>	-0.061	0.0179	11.556	0.0007	0.0459
<i>PMS2P5</i>	-0.050	0.0098	26.5186	<0.0001	<0.0001
<i>APEX1</i>	-0.068	0.0158	18.2191	<0.0001	0.0013
<i>ERCC8</i>	-0.074	0.0077	91.6536	<0.0001	<0.0001
<i>OLIG2</i>	0.067	0.019	12.4442	0.0004	0.0285
<i>ALKBH2</i>	-0.059	0.0041	203.229	<0.0001	<0.0001
<i>UBB</i>	-0.017	0.0232	0.5242	0.469	1
<i>CHEK1</i>	-0.072	0.0183	15.6263	<0.0001	0.0052
<i>PRUNE2</i>	-0.113	0.0108	108.798	<0.0001	<0.0001
<i>DDB2</i>	-0.043	0.0231	3.5194	0.0607	1
<i>CCN4</i>	0.079	0.0333	5.564	0.0183	1
<i>RPA4</i>	-0.027	0.0061	20.3553	<0.0001	0.0004
<i>RPA3</i>	0.027	0.0099	7.4999	0.0062	0.4196
<i>EPSTH1</i>	0.051	0.0278	3.3488	0.0673	1
<i>GSTA1</i>	-0.080	0.0057	192.088	<0.0001	<0.0001

-
-

^a For loci in which ≥ 22 reads were obtained in each of the four samples, i.e., both NSC and both Nx18-15 replicates.

^b Read number of sample with the least amount of reads.

^c Proportion of difference in H5CG or accessibility (GBM Nx18-25 minus NSC non-cancerous cells).

^d SE, standard error.

^e All *P* values $< .0001$ (non-corrected and corrected) were rounded off to $< .0001$.

^f Bonferroni method corrected ($\alpha = .05$).

^g *P* value calculated for HCG or GCH sites for -80 to $+45$ relative to TSS within full-length target from -246 to $+199$.

Differential DNA methylation (H5mCG)						
Promoter/Parameter	Read number ^b	Methylation change ^c	SE ^d	Wald	<i>P</i> value ^e	Adjusted <i>P</i> value ^f
<i>CD44</i>	43	0.575	0.0327	309.346	< 0.0001	< 0.0001
<i>HOXB2</i>	24	0.506	0.0021	57278.7	< 0.0001	< 0.0001
<i>HOTAIR</i>	968	0.491	0.0032	23831.2	< 0.0001	< 0.0001
<i>TAF1</i>	168	0.233	0.0168	193.082	< 0.0001	< 0.0001
<i>APEX2</i>	701	0.145	0.0125	134.271	< 0.0001	< 0.0001
<i>MCM9</i>	655	0.109	0.0038	799.869	< 0.0001	< 0.0001
<i>PELO</i>	89	0.067	0.0021	965.227	< 0.0001	< 0.0001
<i>EPM2AIP1</i>	518	0.007	0.0014	21.6546	< 0.0001	0.0002
<i>PRUNE2</i>	238	-0.065	0.0084	60.2971	< 0.0001	< 0.0001
<i>RPA4</i>	3285	-0.111	0.0018	3966.6	< 0.0001	< 0.0001
<i>RPA3</i>	154	-0.231	0.0336	47.4977	< 0.0001	< 0.0001
<i>EPSTH1</i>	70	-0.355	0.0412	74.0241	< 0.0001	< 0.0001
<i>GSTA1</i>	1580	-0.748	0.0028	73279.5	< 0.0001	< 0.0001

Differential Accessibility (G5mCH)					
Promoter/Parameter	Methylation change ^c	SE ^d	Wald	<i>P</i> value ^e	Adjusted <i>P</i> value ^f
<i>CD44</i>	-0.201	0.0054	1398.66	< 0.0001	< 0.0001
<i>HOXB2</i>	0.015	0.0049	9.255	0.0023	0.1597
<i>HOTAIR</i>	-0.067	0.0049	189.452	< 0.0001	< 0.0001
<i>TAF1</i>	-0.102	0.0051	401.629	< 0.0001	< 0.0001
<i>APEX2</i>	-0.066	0.0192	11.6487	0.0006	0.0437
<i>MCM9</i>	-0.001	0.0109	0.0106	0.9182	1

<i>PELO</i>	-0.113	0.0153	53.8195	<0.0001	<0.0001
<i>EPM2AIP1</i>	-0.102	0.0191	28.259	<0.0001	<0.0001
<i>PRUNE2</i>	-0.113	0.0108	108.798	<0.0001	<0.0001
<i>RPA4</i>	-0.027	0.0061	20.3553	<0.0001	0.0004
<i>RPA3</i>	0.027	0.0099	7.4999	0.0062	0.4196
<i>EPSTI1</i>	0.051	0.0278	3.3488	0.0673	1
<i>GSTA1</i>	-0.080	0.0057	192.088	<0.0001	<0.0001

-
-

^a For loci in which ≥ 22 reads were obtained in each of the four samples, i.e., both NSC and both Nx18-15 replicates.

^b Read number of sample with the least amount of reads.

^c Proportion of difference in H5CG or accessibility (GBM Nx18-25 minus NSC non-cancerous cells).

^d SE, standard error.

^e All *P* values $<< .0001$ (non-corrected and corrected) were rounded off to $< .0001$.

^f Bonferroni method corrected ($\alpha = .05$).

Differential DNA methylation (H5mCG)						
Promoter/Parameter	Read number ^b	Methylation change ^c	SE ^d	Wald	<i>P</i> value ^e	Adjusted <i>P</i> value ^f
<i>CD44</i>	43	0.575	0.0327	309.346	0.0001	<0.0001
<i>HOXB2</i>	24	0.506	0.0021	57278.7	0.0001	<0.0001
<i>HOTAIR</i>	968	0.491	0.0032	23831.2	0.0001	<0.0001
<i>TAF1</i>	168	0.233	0.0168	193.082	0.0001	<0.0001
<i>RPA3</i>	154	-0.231	0.0336	47.4977	0.0001	<0.0001
<i>EPSTI1</i>	70	-0.355	0.0412	74.0241	0.0001	<0.0001
<i>GSTA1</i>	1580	-0.748	0.0028	73279.5	0.0001	<0.0001

Differential Accessibility (G5mCH)					
Promoter/Parameter	Methylation change ^c	SE ^d	Wald	<i>P</i> value ^e	Adjusted <i>P</i> value ^f
<i>CD44</i>	-0.201	0.0054	1398.66	0.0001	<0.0001

<i>HOXB2</i>	0.015	0.0049	9.255	0.0023	0.1597
				<	
<i>HOTAIR</i>	-0.067	0.0049	189.452	0.0001	<0.0001
				<	
<i>TAF1</i>	-0.102	0.0051	401.629	0.0001	<0.0001
<i>RPA3</i>	0.027	0.0099	7.4999	0.0062	0.4196
<i>EPSTI1</i>	0.051	0.0278	3.3488	0.0673	1
				<	
<i>GSTA1</i>	-0.080	0.0057	192.088	0.0001	<0.0001

-
-

^a For loci in which ≥ 22 reads were obtained in each of the four samples, i.e., both NSC and both Nx18-15 replicates.

^b Read number of sample with the least amount of reads.

^c Proportion of difference in H5CG or accessibility (GBM Nx18-25 minus NSC non-cancerous cells).

^d SE, standard error.

^e All *P* values $< < .0001$ (non-corrected and corrected) were rounded off to $< .0001$.

^f Bonferroni method corrected ($\alpha = .05$).

Differential DNA methylation (H5mCG)						
Promoter/Parameter	Read number ^b	Methylation change ^c	SE ^d	Wald	<i>P</i> value ^e	Adjusted <i>P</i> value ^f
<i>CD44</i>	43	0.575	0.0327	309.346	0.0001	<0.0001
					<	
<i>HOXB2</i>	24	0.506	0.0021	57278.7	0.0001	<0.0001
					<	
<i>HOTAIR</i>	968	0.491	0.0032	23831.2	0.0001	<0.0001
					<	
<i>GSTA1</i>	1580	-0.748	0.0028	73279.5	0.0001	<0.0001

Differential Accessibility (G5mCH)					
Promoter/Parameter	Methylation change ^c	SE ^d	Wald	<i>P</i> value ^e	Adjusted <i>P</i> value ^f
				<	
<i>CD44</i>	-0.201	0.0054	1398.66	0.0001	<0.0001
<i>HOXB2</i>	0.015	0.0049	9.255	0.0023	0.1597
				<	
<i>HOTAIR</i>	-0.067	0.0049	189.452	0.0001	<0.0001

<i>GSTA1</i>	-0.080	0.0057	192.088	0.0001	<0.0001
--------------	--------	--------	---------	--------	---------

-
-

^a For loci in which ≥ 22 reads were obtained in each of the four samples, i.e., both NSC and both Nx18-15 replicates.

^b Read number of sample with the least amount of reads.

^c Proportion of difference in H5CG or accessibility (GBM Nx18-25 minus NSC non-cancerous cells).

^d SE, standard error.

^e All *P* values $< .0001$ (non-corrected and corrected) were rounded off to $< .0001$.

^f Bonferroni method corrected ($\alpha = .05$).

Table 7: SNPs and indels detected by FENGCC using standard PCR of 118 ~450-nt human targets.^a

Promoter/Target Sample	Reads NS C-1	Reads NS C-2	Reads Nx 18-25-1	Reads Nx 18-25-2	Captured strand (+/-)	Num SNPs	Location using hg38	dbSNP	Reference base	Variant	Amino acid change	Fraction of reads with variant				Notes	
												NSC-1	NSC-2	Nx18-25-1	Nx18-25-2		
ALK	6	5	24	17	-	0											
BHL																	
ALK	233	201	1634	1879	-	2	chr12: 109,093,535	rs76	G	A	N/A	10/22	45/0	10/20	51/1	11/63	Eliminates a GCG and a GCH
BH2	4	7										1/22	7/20	0/20	9/8	6/17	
							chr12: 109,093,559	rs16	C	T	N/A	1/22	0/10	7/20	0/10	6/17	
								45				22/1	0/20	0/00	20/17	4/4	
ALK	4	6	15	9	+	0											
BH3																	
APE	130	159	210	240	+	1	chr14: 20,454,990	rs17	T	G	N/A	53/15	47/5	3/8	0/00	0/92	NSC-specific SNP
X1								609				3/5	2/8	0/2	0/92	0/00	
APE	390	321	1251	1293	+	0											
X2																	
ATM	290	349	702	720	+	0											
ATR	207	340	551	492	-	0											
X																	
ATX	0	0	0	0	-	0											
N3																	
AXIN	0	0	0	0	-	0											
2																	
BRC	222	182	254	278	-	0											
AJ																	
BRC	73	128	240	274	+	0											
A2																	
CCN	249	292	499	497	+	0											
4																	
CD4	6	12	11	19	-	0											
4																	
CDH	28	23	68	49	+	1	chr16: 68,737,131	rs16	C	A	N/A	0/0	0/22	0/00	0/1	0/8	GBM-specific SNP
1								260				25/0	0/00	0/00	0/00	0/00	
CDH	0	2	2	6	+	0											
13																	

[illegible]

[illegible]

<i>FAN</i> <i>CC</i>	2	3	0	0	-	0	chr11: 61,79 2,350	G	A	N/A	2/ 37	5/ 41	3/ 61	4/ 92	0/ 68	0/ 97	9/ 77	11/ 6	Eliminates an HCG and a GCH
<i>FEN</i> <i>i</i>	447	367	268	339	+	2	chr11: 61,79 2,609	G	A	N/A	29 75	50 8	32 6	0 0	69 7	10 3	99 2	11 0	Higher prevalence of SNP in GBM that eliminates a GCH site
<i>FHIT</i>	97	89	103	100	-	0													
<i>FOX</i> <i>C2</i>	0	0	0	0	+	0													
<i>FZD</i> <i>7</i>	0	0	0	0	+	0													
<i>FZD</i> <i>8</i>	3	1	7	1	-	0													
<i>GPR</i> <i>158</i>	9	16	21	35	+	0													
<i>GST</i> <i>A1</i>	277 0	254 9	2850	3147	-	4	chr6:5 2,803, 872	T	C	N/A	0/ 50	0/ %	0/ 50	0/ %	37 5	40 4	97 6		GBM-specific SNP at -59 bp from TSS
<i>GST</i> <i>A1</i>	277 0	254 9	2850	3147	-	4	chr6:5 2,803, 889	A	G	N/A	0/ 50	0/ %	0/ 50	0/ %	43 4	40 3	95 3		GBM-specific SNP -74 bp from TSS
<i>HIST</i> <i>H1</i> <i>B</i>	560	461	545	531	-	0	chr6:5 2,803, 986	C	A	N/A	16 7	10 %	17 7	0/ %	70 0	62 6	98 4		
<i>HMG</i> <i>A2</i>	2	10	15	23	+	0	chr6:5 2,804, 145	T	C	N/A	4/ 15	0/ %	5/ 17	0/ %	75 7	63 6	10 00		
<i>HOT</i> <i>AIR</i>	246 7	213 1	1896	2089	-	2	chr12: 53,96 8,724	C	T	N/A	24 96	12 %	27 20	2/ 7	0/ 50	0/ 21	0/ 3		Eliminates a GCH
<i>HOT</i> <i>AIR</i>	246 7	213 1	1896	2089	-	2	chr12: 53,96 8,802	T	C	N/A	62 64	30 %	65 33	9/ 0	0/ 4	0/ 6	0/ %		NSC-specific SNP

<i>HOXB2</i>	428	421	350	337	-	1	chr17: 48,544,929	G	T	N/A	0/74	0/00	0/73	0/00	0/11	0/00	15/7	19/8
<i>HSPA5</i>	1	9	10	32	-	0											7	%
<i>INOS0</i>	0	0	0	0	-	0												
<i>KLF8</i>	9	0	18	11	+	0												
<i>LIG1</i>	128	85	109	149	-	0												
<i>MBD4</i>	13	24	71	61	-	0												
<i>MC9</i>	383	310	2793	2917	-	1	chr6.1 18,931,610	G	T	Tyr to Stop codon	36/23	16/4	42/26	18/5	0/25	0/00	39/1	19/0
<i>MECP2</i>	0	0	0	0	-	0											97	%
<i>MEG3</i>	91	81	108	172	+	1	chr14: 100,882,25,826	C	A	N/A	5/54	9/26	7/64	10/4	8/55	14/5	16/6	
<i>MGM1</i>	0	0	0	0	+	0												
<i>MLH1</i>	187	136	1122	1300	+	1	chr3.3 6,993,455	G	A	N/A	0/91	0/00	4/68	5/88	33/0	43/7	64/9	
<i>MLH3</i>	21	29	63	56	-	1	chr14: 75,051,627	G	A	N/A	15/5	10/00	21/2	10/00	45/4	10/00	38/3	10/00
<i>MNA1</i>	26	30	130	84	+	0												
<i>MPG</i>	0	0	0	0	+	0												
<i>MRE11</i>	185	173	312	279	-	1	chr11: 94,490,39,59	C	A	N/A	31/7	40/7	27/7	37/5	44/9	46/3	53/2	
<i>MSH2</i>	881	529	737	762	-	0												
<i>MSH3</i>	0	0	0	0	+	0												
<i>MSH4</i>	30	39	48	50	+	3	chr1.7 5,796,838	G	A	N/A	0/30	0/00	2/31	6/45	18/9	16/3	43/4	

GBM-specific SNP -62 bp from TSSa

Eliminates a GCH

Eliminates an HCG and GCH, i.e., a CGC, at -115 bp from MRE11 TSS

GBM-specific SNP that eliminates a GCH

[illegible]

PMS	13	15	66	77	-	0
2P3						
PMS	737	968	1291	1533	+	0
2P5						
POL	0	0	0	0	+	0
D1						
POL	731	379	512	739	-	0
D4						
POL	0	0	0	2	-	0
G						
POL	50	135	166	209	+	0
K						
POL	658	679	1029	987	-	1
Q						
POL	30	8	14	24	+	1
R2B						
POL	2	2	11	26	+	0
R2G						
PRD	822	875	1145	1250	+	1
M2						
PRK	9	3	9	22	+	0
CG						
PRU	7	8	0	14	-	0
NE2						
FAM	4	14	66	76	-	0
P4						
PTC	0	0	0	2	-	0
ARD						
RNA						
SEH	1	2	16	7	+	0
2A						
RVF	132	135	218	204	+	1
111*						
RPA	309	509	754	739	-	0
2						
RPA	135	103	1134	1265	-	0
3	3	4				
RPA	121	104	1075	1161	+	0
4	1	7				
SH3T	3	2	4	9	+	0
CI						
SIM2	0	0	0	0	+	0

[illegible]

^a High-fidelity CCS reads with ≥ 5 SMRT sequencing passes.

^b Suffixes -1 and -2 refer to the two independent biological replicates.

^c Comparing NSC and GBM, locus with differential variant frequency and statistically significant epigenetic differences (TABLE S6).

Novel SNP detected specifically in GBM cells with GQ > 30 and depth > 20 in both two biological replicates.

GBM-specific SNP

NSC-specific SNP

Promoter/Sample	Reads_NSC-1	Reads_NSC-2	Reads_Nx18-25-1	Reads_Nx18-25-2	Capture strand (+/-)	Num indels (GQ > 30 , Depth > 15)	Location using hg38	dbSNP	Reference base	Variant	Annotation	Fraction of reads with variant				Notes
												NSC-1	NSC-2	Nx18-25-1	Nx18-25-2	
ALKB H1	6	5	24	17	-	0										
ALKB H2	2334	2017	1634	1879	-	0										
ALKB H3	4	6	15	9	+	0										
APEX 1	130	159	210	240	+	0										
APEX 2	390	321	1251	1293	+	0										
ATM	290	349	702	720	+	0										
ATRX	207	340	551	492	-	0										
ATXN 3	0	0	0	0	-	0										
AXIN 2	0	0	0	0	-	0										
BRCA 1	222	182	254	278	-	0										
BRCA 2	73	128	240	274	+	0										
CCN4	249	292	499	497	+	1	chr8.133,190,920-133,190,921	rs548251181	GA	G	N/A	29/59	49/59	14/30	25/30	Unmapped sequence plotted in white in FIG. 41B and FIG. 49C.
CD44	6	12	11	19	-	0										

CDH1	28	23	68	49	+	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
------	----	----	----	----	---	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Gene	3	1	7	1	-	0	chr	rs	AT	A	N/	0/	0.0	0/	0.0	47	27	46	1	GBM-specific deletion
<i>FZD8</i>	3	1	7	1	-	0		rs19			N/	0/	0.0	0/	0.0	47	27	46	1	
<i>GPR158</i>	9	16	21	35	+	0		9734	AT	A	A	19	0%	22	0%	71	101	1	2	
<i>GSTA1</i>	2770	2549	2850	3147	-	0		232			A	6		2		34	%	43	9	
<i>HIST1H1B</i>	560	461	545	531	-	0		rs71			N/	14	69	15	68	0/	0/	0/	0	
<i>HMG A2</i>	2	10	15	23	+	0		2272	T	T	N/	2/	61	8/	10	17	0%	14	00	
	2467	2131	1896	2089	-	2		79		C	A	20	%	23	%	4		7	%	
<i>HOTAIR</i>												4		2						
								rs11			N/	79	96	78	91	76	98	76	90	
								1979	A	A	N/	8	34	8	76	77	70	78	4	
								968		C	A	2	%	5	%	7	%	4	%	
												7/	7.8	13	13	3/	3.6	1/	1	
<i>HROXB2</i>	428	421	350	337	-	3		rs13	AC	A	A	89	7%	7	%	82	6%	87	15	
								8169			N/	3/	3.3	2/	2.0	23	28	32	7	
								880		C	A	89	7%	97	6%	8	105	8	8	
										C						2	%	7	9	
								rs90			N/	31	28	41	34	40	37	35	35	
								1811			N/	1/	44	1/	75	71	38	79	3	
								939	CT	C	A	09	%	18	%	07	%	9	5	
<i>HSPA5</i>	1	9	10	32	-	1		rs32	GC	G	N/	0/	0.0	2/	40	4/	57	7/	31	
								1673			A	1	0%	5	%	14	14	22	2	
								3								7	%		%	
<i>INO80</i>	0	0	0	0	-	0														
<i>KLF8</i>	9	0	18	11	+	0														
<i>LIG1</i>	128	85	109	149	-	0														
<i>MBD4</i>	13	24	71	61	-	0														
<i>MCM9</i>	3836	3105	2793	2917	-	1		rs76	GA	G	N/	14	41	15	41	10	35	12	46	
								8549			N/	8/	57	8/	91	9/	28	8/	0	
								255			A	35	%	37	%	30	%	27	4	
<i>MECP2</i>	0	0	0	0	-	0						6		7		9		8	%	

[illegible]

[illegible]

[illegible]

$\frac{3}{8}$

Table 8: CCS reads on- and off-target for 45 ~940-nt human targets in MAPit-FENG C using EM-seq.^a

Sample ^b	800 ng input-1	800 ng input-2	400 ng input-1	400 ng input-2
Total reads	179,458	130,943	54,023	57,231
Reads aligned on genome	165,646	121,108	50,873	53,592
% Reads aligned on genome	92.3%	92.5%	94.2%	93.6%
Reads aligned on target	145,070	105,794	41,467	44,214
% Reads off target but on genome	12.4%	12.6%	18.5%	17.5%
% Total reads on target	80.8%	80.8%	76.8%	77.3%
Number of targets detected	38		38	
Number of targets detected using reads with $\geq 95\%$ coverage of each reference sequence and $\geq 95\%$ conversion	25		20	

^a High-fidelity CCS reads with ≥ 5 SMRT sequencing passes.

^b Suffixes -1 and -2 refer to the two independent biological replicates.

Table 9: Filtered CCS reads aligned to 45 ~940-nt human targets in MAPit-FENGCC using EM-seq.^a

Promoter/Parameter	Capture strand length (nt)	%GC	Filtered CCS read number ^b			
			800 ng input-1 ^c	800 ng input-2	400 ng input-1	400 ng input-2
<i>ALKBH2</i>	946	54	0	0	0	0
<i>APEX1</i>	937	56	418	302	109	99
<i>APEX2</i>	940	55	60	50	21	16
<i>ATM</i>	938	58	730	547	155	136
<i>ATRX</i>	930	43	0	0	0	0
<i>BRCA1</i>	937	51	0	0	0	0
<i>BRCA2</i>	935	57	0	0	0	0
<i>CCN4</i>	944	53	255	191	62	49
<i>CD44</i>	942	53	0	0	0	0
<i>CETN1</i>	930	54	0	2	0	0
<i>CHEK1</i>	953	44	0	0	0	0
<i>CHEK2</i>	950	61	0	0	1	0
<i>CSDE1</i>	945	47	1	0	0	0
<i>DDB2</i>	942	58	3	3	1	0
<i>EPM2AIP1-MLH1</i>	937	59	47174	33315	8989	8286
<i>EPSTI1</i>	934	47	1	0	0	0
<i>EXO1</i>	951	53	4	2	0	0
<i>FEN1</i>	942	54	1	1	0	0
<i>GSTA1</i>	939	39	0	0	0	0
<i>HIST1H1B</i>	952	45	0	0	0	0
<i>HOTAIR</i>	959	53	0	0	0	0
<i>HOXB2</i>	935	49	0	0	0	0
<i>HSPA5</i>	951	59	25971	18679	6675	6472
<i>LIG1</i>	941	65	0	0	0	0
<i>MBD4</i>	951	56	9	5	2	1
<i>MCM9</i>	939	33	0	0	0	0
<i>MECP2</i>	957	59	0	0	0	0
<i>MRE11</i>	956	54	3	1	1	2
<i>MSH2</i>	930	52	462	331	108	86
<i>MSH4</i>	931	46	80	55	20	14
<i>MSH5</i>	931	53	4347	3234	1029	965
<i>OGG1</i>	946	61	10	16	2	1
<i>OLIG2</i>	952	63	0	0	0	0
<i>PELO</i>	946	50	0	0	0	0
<i>PMS2</i>	944	63	22554	16530	5198	4918

<i>PMS2P5</i>	935	64	28081	19985	8261	7756
<i>POLD4</i>	939	59	4512	3222	976	923
<i>PRUNE2</i>	945	66	29	24	7	9
<i>RPA2</i>	951	53	7	5	4	2
<i>RPA4</i>	951	42	0	0	0	0
<i>TAF1</i>	958	53	0	0	0	0
<i>TP53</i>	938	49	0	1	0	0
<i>UBB</i>	943	56	1	0	0	1
<i>XRCC2</i>	939	50	0	0	0	0
<i>ZNF830</i>	950	53	0	0	0	0

^a High-fidelity CCS reads with ≥ 5 SMRT sequencing passes.

^b CCS reads with $\geq 95\%$ coverage of each reference sequence and $\geq 95\%$ HCH conversion.

^c Suffixes -1 and -2 refer to each of the two independent biological replicates.

Table 10: CCS reads on- and off-target for 78 ~620-nt mouse targets in MAPit-FENGCC using EM-seq.^a

Sample	Mouse 1	Mouse 2	Mouse 3	Mouse 4
Input gDNA (ng)	600	600	600	600
Total reads	172,876	147,858	168,272	108,306
Reads aligned to other amplicons in library after filtering ^b	19,537	16,271	18,598	4,588
Reads aligned to FENGCC targets after filtering ^b	122,568	101,787	106,040	76,389
Unaligned reads and reads removed by filtering ^{b,c}	30,771	29,800	43,634	27,329
% On-target reads, ignoring off-target reads from other amplicons	80%	77%	71%	74%
Number of targets detected ^c	73	71	73	75
Number of targets detected in each library using filtered reads ^{b,d}	70	68	67	71
Number of targets detected in four combined libraries using filtered reads ^{b,d}	74			

^a High-fidelity CCS reads with ≥ 5 SMRT sequencing passes.

^b Filtered for $\geq 95\%$ conversion and $\geq 95\%$ coverage of each mouse reference sequence (mm9 genome build).

^c May include off-target reads from other amplicons in library.

^d At least one read.

Table 11: Filtered CCS reads aligned to 78 ~620-nt mouse targets in MAPit-FENGCC using EM-seq and *P* values.^a

Promoter/Sample	CCS read number ^b				Average CCS read number	Amplicon length	Averaged % H5mCG (endogenous methylation) <i>P</i> value ^{c,d}	Averaged % G5mCH (accessibility) <i>P</i> value ^{c,d}
	Mouse 1	Mouse 2	Mouse 3	Mouse 4				
<i>Ash1l</i>	9	14	0	10	8	829	Excluded; insufficient reads	Excluded; insufficient reads
<i>Btk</i>	8412	6930	2687	1761	4948	514	0.878	0.989
<i>Ccl5</i>	14885	12096	7268	5135	9846	524	1.000	0.956
<i>Cd180</i>	216	201	290	515	306	509	0.300	0.885
<i>Cox8a</i>	567	376	739	570	563	646	0.877	0.905
<i>Cxcl15</i>	178	253	633	701	441	558	1.000	0.652
<i>Cxcr4</i>	1586	1738	2805	2253	2096	492	1.000	0.998
<i>Dnmt3b</i>	78	125	92	68	91	987	Excluded; duplicates	Excluded; duplicates
<i>Foxo1</i>	0	0	0	0	0	905	Excluded; insufficient reads	Excluded; insufficient reads
<i>Gdf5</i>	15135	15154	13733	7529	12888	491	1.000	1.000
<i>Hdac3</i>	2642	2129	1568	816	1789	537	1.000	0.998
<i>Hsf1</i>	812	566	649	307	584	474	1.000	0.841
<i>Hsf2bp</i>	792	475	645	260	543	484	1.000	0.582
<i>Hsf4</i>	301	120	87	50	140	623	Excluded; duplicates	Excluded; duplicates
<i>Hsf5</i>	0	0	0	0	0	636	Excluded; insufficient reads	Excluded; insufficient reads
<i>Hsfy2</i>	516	710	784	727	684	662	0.903	0.963
<i>Hsp90aa1</i>	16707	13095	19155	11218	15044	522	0.482	1.000
<i>Hsp90ab1</i>	3523	2401	2625	1523	2518	606	1.000	0.879
<i>Hspa1a</i>	0	4	0	4	2	709	Excluded; insufficient reads	Excluded; insufficient reads
<i>Hspa1b</i>	16	1	2	10	7	640	Excluded; insufficient reads	Excluded; insufficient reads
<i>Hspa2</i>	118	103	166	81	117	605	Excluded; duplicates	Excluded; duplicates
<i>Ifna1</i>	0	0	0	0	0	696	Excluded; insufficient reads	Excluded; insufficient reads
<i>Ifna2</i>	3174	3855	3031	4350	3603	533	1.000	0.958
<i>Ifng</i>	3892	2805	3667	3159	3381	476	0.594	0.974
<i>Ikbkb</i>	8	0	3	19	8	720	Excluded; insufficient reads	Excluded; insufficient reads
<i>Ikbkg</i>	211	279	647	666	451	554	Excluded; duplicates	Excluded; duplicates
<i>Il10</i>	4907	3722	3439	3110	3795	551	0.173	0.981
<i>Il12a</i>	14249	14043	8757	7683	11183	537	0.995	0.995
<i>Il12b</i>	3217	3150	5053	4153	3893	532	1.000	0.997
<i>Il3ra</i>	1377	1142	1412	1081	1253	615	0.999	1.000
<i>Il6</i>	18	8	19	16	15	807	Excluded; insufficient reads	Excluded; insufficient reads
<i>Irak1</i>	65	25	78	50	55	678	Excluded; insufficient reads	Excluded; insufficient reads

<i>Irak4</i>	1035	834	451	385	676	606	1,000	0.951
<i>Irf3</i>	10	0	11	12	8	557	Excluded; insufficient reads	Excluded; insufficient reads
<i>Irf5</i>	219	61	241	83	151	747	Excluded; duplicates	Excluded; duplicates
<i>Irf7</i>	952	1043	2468	2601	1766	485	0.990	0.928
<i>Jun</i>	1110	798	966	718	898	630	1,000	0.966
<i>Kdm6b</i>	14370	13361	15486	9225	13111	489	0.687	0.981
<i>Lbp</i>	659	850	749	758	754	564	1,000	0.968
<i>Ly86</i>	7615	5536	8098	5683	6733	534	0.996	0.618
<i>Map3k1</i>	0	88	0	0	22	677	Excluded; insufficient reads	Excluded; insufficient reads
<i>Mapk1</i>	1815	1389	1589	1004	1449	496	1,000	0.924
<i>Mapk13</i>	0	0	0	0	0	795	Excluded; insufficient reads	Excluded; insufficient reads
<i>Mapk14</i>	18	202	57	31	77	891	Excluded; insufficient reads	Excluded; insufficient reads
<i>Mapk15</i>	9818	8824	6756	5199	7649	556	0.998	0.994
<i>Mapk3</i>	85	102	23	21	58	707	Excluded; insufficient reads	Excluded; insufficient reads
<i>Mapk4</i>	507	402	398	232	385	755	1,000	0.891
<i>Mapk6</i>	6	11	0	3	5	563	Excluded; insufficient reads	Excluded; insufficient reads
<i>Mapk8</i>	1	0	2	0	1	872	Excluded; insufficient reads	Excluded; insufficient reads
<i>Mme</i>	1993	1088	1758	1255	1524	596	1,000	0.992
<i>Myd88</i>	118	70	28	46	66	628	Excluded; insufficient reads	Excluded; insufficient reads
<i>Nfkb1</i>	19	36	16	4	19	620	Excluded; insufficient reads	Excluded; insufficient reads
<i>Nfkb2</i>	10038	8147	6037	4969	7298	511	1,000	0.992
<i>Nfkbia</i>	55	118	158	181	128	690	Excluded; duplicates	Excluded; duplicates
<i>Nfkbib</i>	141	116	105	162	131	530	0.997	0.492
<i>Nr2c2</i>	54	67	17	7	36	507	Excluded; insufficient reads	Excluded; insufficient reads
<i>Pik3ap1</i>	172	84	73	51	95	613	Excluded; duplicates	Excluded; duplicates
<i>Pik3c2a</i>	0	0	1	0	0	883	Excluded; insufficient reads	Excluded; insufficient reads
<i>Pik3c2g</i>	88	102	23	408	155	660	Excluded; insufficient reads	Excluded; insufficient reads
<i>Pik3c3</i>	306	232	251	309	275	640	1,000	1,000
<i>Pik3ca</i>	48	80	90	74	73	670	Excluded; duplicates	Excluded; duplicates
<i>Pik3r1</i>	24	0	13	6	11	623	Excluded; insufficient reads	Excluded; insufficient reads
<i>Pik3r2</i>	51	179	0	29	65	815	Excluded; insufficient reads	Excluded; insufficient reads
<i>Pik3r3</i>	3304	2329	2313	1533	2370	605	0.831	0.998
<i>Pik3r5</i>	231	233	289	270	256	728	1,000	0.863
<i>Pik3r6</i>	12852	10250	13647	8810	11390	540	1,000	0.955
<i>Ptpn6</i>	2311	1920	5209	3882	3331	483	0.906	0.923
<i>Rela</i>	50	0	16	66	33	654	Excluded; insufficient reads	Excluded; insufficient reads
<i>Src</i>	4417	3135	3658	1978	3297	616	0.986	0.937
<i>Stat3</i>	717	320	490	225	438	549	1,000	0.985

<i>Tab1</i>	721	473	404	224	456	760	1.000	0.971
<i>Tab2</i>	0	18	0	11	7	575	Excluded; insufficient reads	Excluded; insufficient reads
<i>Tet2</i>	131	82	267	83	141	674	Excluded; duplicates	Excluded; duplicates
<i>Tlr2</i>	580	644	630	346	550	665	0.890	0.971
<i>Tlr4</i>	2542	1936	3654	2704	2709	576	0.836	0.231
<i>Tnf</i> ^a	1	5	0	6	3	499	Excluded; insufficient reads	Excluded; insufficient reads
<i>Traf6</i>	1531	1150	2002	1400	1521	543	1.000	0.998
<i>Vav1</i>	811	729	280	257	519	605	0.979	0.907

^a High-fidelity CCS reads with ≥ 5 SMRT sequencing passes.

^b Filtered for $\geq 95\%$ conversion of HCH to HTH and $\geq 95\%$ coverage of the length of each reference sequence.

^c Limited to targets with ≥ 100 reads with good pattern diversity, i.e., lacking many duplicates, for each of the four mice.

^d Not corrected for multiple testing; all Bonferroni method corrected (alpha .05) *P* values were 1.000.

Table 12: Recommended FENG C oligonucleotide concentrations for different numbers of targets.

Number of targets	Concentration of each of oligos 1-N, 2-N, and 3 (nM) in stock oligo mixtures ^{a-e}
10	200
20	100
30	67
40	50
50	40
60	33
70	29
80	25
90	22
100	20
110	18
120	17
130	15
140	14
150	13
160	13
170	12
180	11
190	11
200	10
210	10
220	9
230	9
240	8
250	8
260	8
270	7
280	7
290	7
300	7
310	6
320	6
330	6
340	6
350	6
360	6
370	5
380	5

390	5
400	5
410	5
420	5
430	5
440	5
450	4
460	4
470	4
480	4
490	4
500	4

^a Concentration of separate stocks of U1-N (or U1) and U2 oligos is fixed at empirically determined $10\ \mu\text{M} = 10,000\ \text{nM}$ ($1\ \mu\text{L}$ of each is added; FIG. 24).

^b The total amount of all oligo 1-N and 2-N should be close to but less than that of U1-N (or U1).

^c Flap oligo 1-N or 2-N concentration (nM) $\times$ number of targets $\times$ 4 (i.e., total added volume of oligo 1-N and 2-N is 4x of U1-N (or U1)), empirically determined to be $8,000\ \text{nM}$ (close to but $< 10,000\ \text{nM}$; FIG. 24).

^d Concentration of each oligo 3-N should be the same as that of oligo 1-N or oligo 2-N in this separate stock mixture.

^e Therefore, the concentration of each of oligo 1-N, 2-N, and 3-N is calculated by $8,000\ \text{nM}/(\text{Number of targets} \times 4)$.

CLAIMS

What is claimed is:

1. An oligonucleotide complex for enriching at least one sequence of interest from a source sequence comprising flap oligonucleotides 1-N and 2-N each hybridized to a universal oligonucleotide 1-N (U1-N) or universal oligonucleotide 1 (U1) to form a first and second flap adapter, wherein the 5' ends of the flap oligonucleotides 1-N and 2-N are sufficiently complementary to anneal to separate ends of the sequence of interest to facilitate cleavage of the sequence of interest from the source sequence.
2. The oligonucleotide complex of claim 1, wherein oligonucleotides 1-N and 2-N are comprised of a common 3' tail portion that anneals to a universal oligonucleotide, either U1 or a respective U1-N.
3. The oligonucleotide complex of claim 1, wherein oligonucleotide U1-N or U1 is sufficiently complementary to anneal to a 3' tail portion of flap oligonucleotide 1-N and flap oligonucleotide 2-N.
4. The oligonucleotide complex of claim 1, wherein an unpaired 3' flap on oligonucleotide U1-N is at least one nucleotide in length and wherein no 3' flap is present on oligonucleotide U1.
5. The oligonucleotide complex of claim 3, wherein the unpaired 3' flap on oligonucleotide U1-N is one deoxynucleotide or dideoxynucleotide in length and can be A, C, G, or T.
6. The oligonucleotide complex of claim 3, wherein the 3' flap end comprises at least one nucleotide and the most 5' nucleotide of the 3' flap matches the nucleotide at point of cleavage of the source sequence where the flap oligonucleotides 1-N and 2-N are annealed at and downstream of the target sequence, respectively.
7. The oligonucleotide complex of claims 4 or 6, wherein the 3' flap cannot pair with the 3' tail of the flap oligonucleotides 1-N and 2-N.
8. The oligonucleotide complex of claim 1, wherein the first and second flap adapters create cleavage sites for a structure-specific endonuclease.
9. The oligonucleotide complex of claim 8, wherein the endonuclease is a flap endonuclease or Taq.
10. The oligonucleotide complex of any of claims 1-9, wherein the 5' cleaved end is ligated to a first universal oligonucleotide (U1-N) and the 3' cleaved end is ligated to a second universal

oligonucleotide (U2) hybridized to an oligonucleotide 3-N thereby forming a DNA strand comprising an arrangement of U1-N-Sequence of interest-U2.

11. The oligonucleotide complex of claim 10, wherein the ligated universal oligonucleotides U1-N, U2, or both are modified to protect against degradation by one or more exonucleases.

5 12. The oligonucleotide complex 10 or 11, wherein upon subjecting the DNA strand to one or more exonucleases an enriched single strand of DNA comprising an arrangement of U1-N-Sequence of interest-U2 is produced.

13. A kit for enriching at least one sequence of interest from a source sequence comprising; flap oligonucleotide 1-N stock, flap oligonucleotide 2-N stock, oligonucleotide 3-N stock, universal
10 oligonucleotide 1-N (oligo U1-N), universal oligonucleotide 2 (U2), U1 primer (oligonucleotide U1), and U2 primer.

14. The kit of claim 13, wherein the flap oligonucleotide 1-N and 2-N stocks comprise of flap oligonucleotides 1-N and 2-N diluted in a solution with the concentration dependent on the number of targets sequences being analyzed.

15 15. The kit of either of claims 13 and 14, wherein for each target sequence of interest there are flap oligonucleotides 1-N and 2-N that are comprised of a 5' end that is target sequence specific and a 3' end that is sufficiently complementary to the oligonucleotide U1-N or U2.

16. The kit of any of claims 13-15, wherein for each target sequence there is an oligonucleotide 3-N that is comprised of a 3' end that is target sequence specific and a 5' end that is sufficiently
20 complementary to the oligonucleotide U2.

17. The kit of any of claims 13-16, wherein oligonucleotide U2 is synthesized with a ligatable 5' phosphate and, at its 3' end, one or more chemical moieties that protect against exonuclease degradation.

18. The kit of any of claims 13-17, wherein oligonucleotide U2 is synthesized with five
25 phosphorothioate bonds at the 3' end and wherein a three-carbon spacer is covalently attached to the 3'-terminal hydroxyl.

19. The kit of any of claims 13-18, wherein the U1-N is synthesized with a ligatable 3'-terminal hydroxyl and, at its 5' end, one or more chemical moieties that protect against exonuclease degradation.

20. The kit of any of claims 13-19, wherein the 5' end of oligonucleotide U1-N is synthesized with five phosphorothioate bonds and either covalently attached to a three-carbon spacer or without a 5' phosphate.

21. A method for enriching a sequence of interest from a source sequence for sequencing

5 comprising:

I. obtaining a source DNA,

II. denaturing the source DNA,

III. annealing a pair of flap oligonucleotides 1-N and 2-N to a first and second region, respectively, of the denatured source sequence, each flap oligonucleotide 1-N and 2-N

10 hybridized to a universal oligonucleotide 1-N (U1-N), wherein the first region comprises a portion of the sequence of interest and wherein the second region is downstream of the sequence of interest,

IV. cleaving the denatured source DNA using a structure-specific endonuclease to produce a cleaved sequence of interest comprising 5' and 3' cut ends,

15 V. ligating the 5' cut end to the oligonucleotide U1-N such that the sequence of interest comprises a ligated U1-N hybridized with the flap oligonucleotide 1-N,

VI. ligating the 3' cut end to an oligonucleotide U2 hybridized to an oligonucleotide 3-N, wherein a 5' portion of the oligonucleotide 3-N is sufficiently complementary to U2 and a 3' tail portion is sufficiently complementary to a portion of the sequence of interest such that the
20 sequence of interest comprises a ligated oligonucleotide U2 hybridized to oligonucleotide 3-N, and

VII. subjecting the sequence of interest from step (V), step (VI), or both, to one or more exonucleases to produce an enriched single strand of DNA comprising an arrangement of U1-N-Sequence of interest-U2.

25 22. The method of claim 21, wherein fragmenting the source DNA comprises subjecting the source DNA to sonication, a restriction enzyme, a DNase, micrococcal nuclease, a chemical, or any other means to fragment DNA.

23. The method of claims 21 or 22, wherein ligating comprises ligating the 5' end of the sequence of interest to the 3' end of oligonucleotide oligo U1-N and the 3' end of the sequence of
30 interest to the 5' end of oligonucleotide oligo U2 followed by exonuclease digestion.

24. The method of claim 23, wherein the exonuclease digestion does not degrade the sequence of interest.
25. The method of any of claims 21-24, wherein the method further comprises subjecting a source sequence to DNA methyltransferase prior to step I.
- 5 26. The method of claim 22, further comprising purifying the U1-N-Sequence of interest-U2 or amplifying the enriched single strand of DNA comprising an arrangement of U1-N-Sequence of interest-U2 by standard PCR, BS-PCR, EM-PCR, both standard PCR and BS-PCR, or both standard PCR and EM-PCR, using U1 and U2 primers.
27. The method of any of claims 21-26, wherein the structure-specific endonuclease comprises a
10 flap endonuclease or Taq polymerase.
28. The method of any of claims 1-12, wherein the sequence of interest is enriched for genotyping or epigenetic analysis.
29. The kit of any of claims 13-20, wherein the sequence of interest is enriched for genotyping or epigenetic analysis.
- 15 30. A method for enriching a sequence of interest from a source sequence comprising:
- I. obtaining a source DNA,
 - II. denaturing the source DNA,
 - III. annealing a pair of flap oligonucleotides 1-N and oligonucleotide 2-N to a first and second region, respectively, of the denatured source sequence, each flap oligonucleotides 1-N
20 and 2-N hybridized to a universal oligonucleotide-1 (U1), wherein the first region comprises a portion of the sequence of interest and wherein the second region is downstream of the sequence of interest,
 - IV. cleaving the denatured source DNA using a structure-specific endonuclease to produce a cleaved sequence of interest comprising 5' and 3' cut ends,
 - 25 V. ligating the 3' cut end to an oligonucleotide oligo U2 hybridized to an oligonucleotide 3-N, wherein a 5' portion of the oligonucleotide 3-N is complementary to oligo U2 and a 3' tail portion is complementary to a portion of the sequence of interest such that the sequence of interest comprises a ligated oligonucleotide U2 hybridized to oligonucleotide 3,
 - VI. subjecting the sequence of interest from step (V), to one or more exonucleases to produce a
30 single strand of DNA comprising an arrangement of Sequence of interest-U2,

VII. annealing to the enriched sequence of interest-U2 to a flap oligonucleotide 1-N hybridized to a universal oligonucleotide 1-N (U1-N),

VIII. ligating oligonucleotide U1-N to the 5' end of the sequence of interest-U2,

IX. subjecting the sequence of interest-U2 of step VIII to one or more exonucleases to produce a single strand of DNA comprising an arrangement of U1-N-Sequence of interest-U2.

31. The method of claim 30, wherein fragmenting the source DNA comprises subjecting the source DNA to sonication, a restriction enzyme, a DNase, micrococcal nuclease, a chemical, or any other means to fragment DNA.

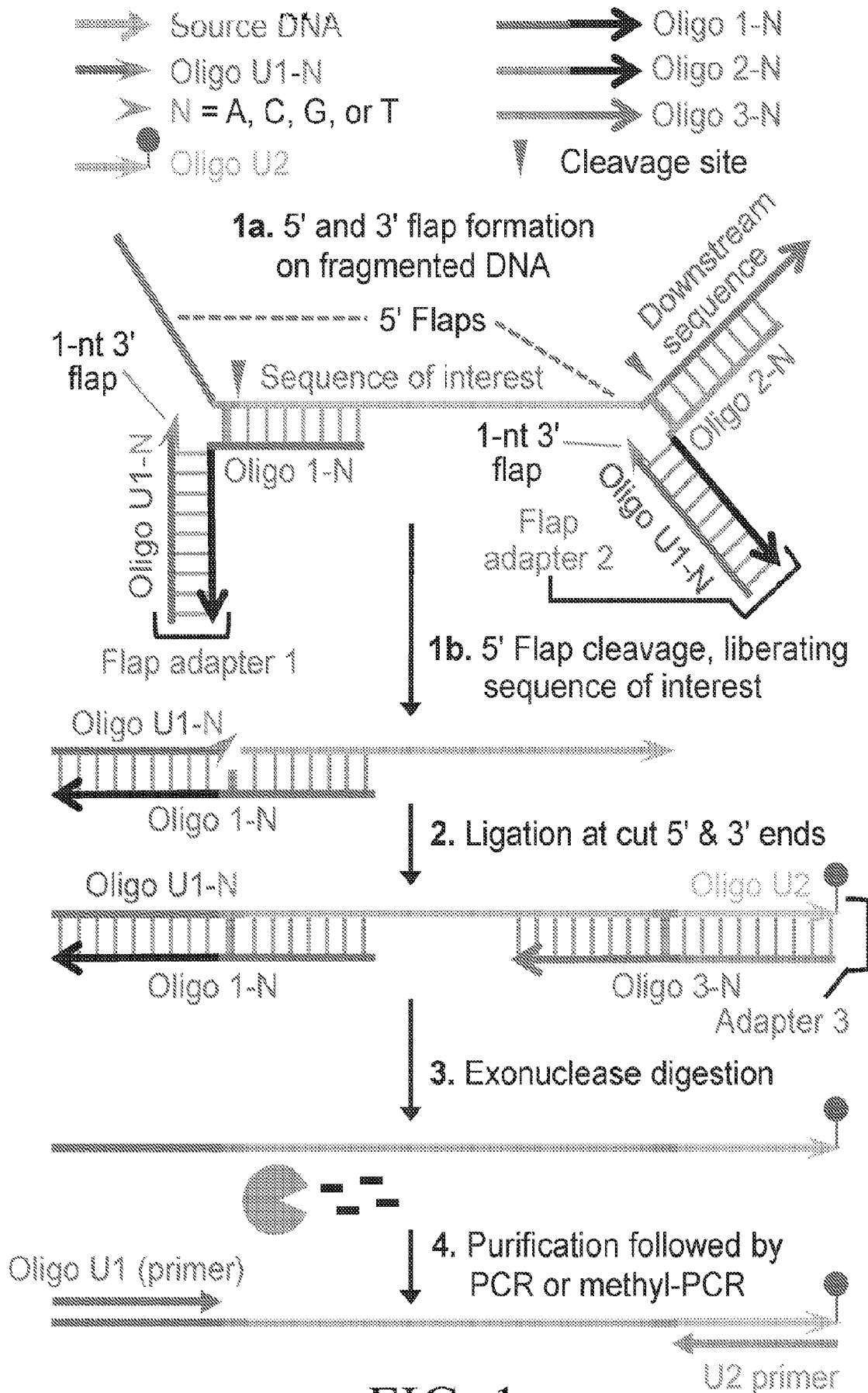
32. The method of claims 30 or 31, wherein the exonuclease digestion of Step VI and IX does not degrade the sequence of interest.

33. The method of any of claims 30-32, wherein the method further comprises subjecting a source sequence to treatment with a DNA methyltransferase prior to step I.

34. The method of any of claims 30-33, further comprising purifying the U1-N-Sequence of interest-U2 or amplifying U1-N-Sequence of Interest-U2 by standard PCR, BS-PCR, EM-PCR, both standard PCR and BS-PCR, or both standard PCR and EM-PCR, using U1 and U2 primers.

35. The method of any of claims 30-34, wherein the structure-specific endonuclease comprises flap endonuclease or Taq polymerase.

36. The method of any of claims 30-35, wherein the sequence of interest is enriched for genotyping or epigenetic analysis.



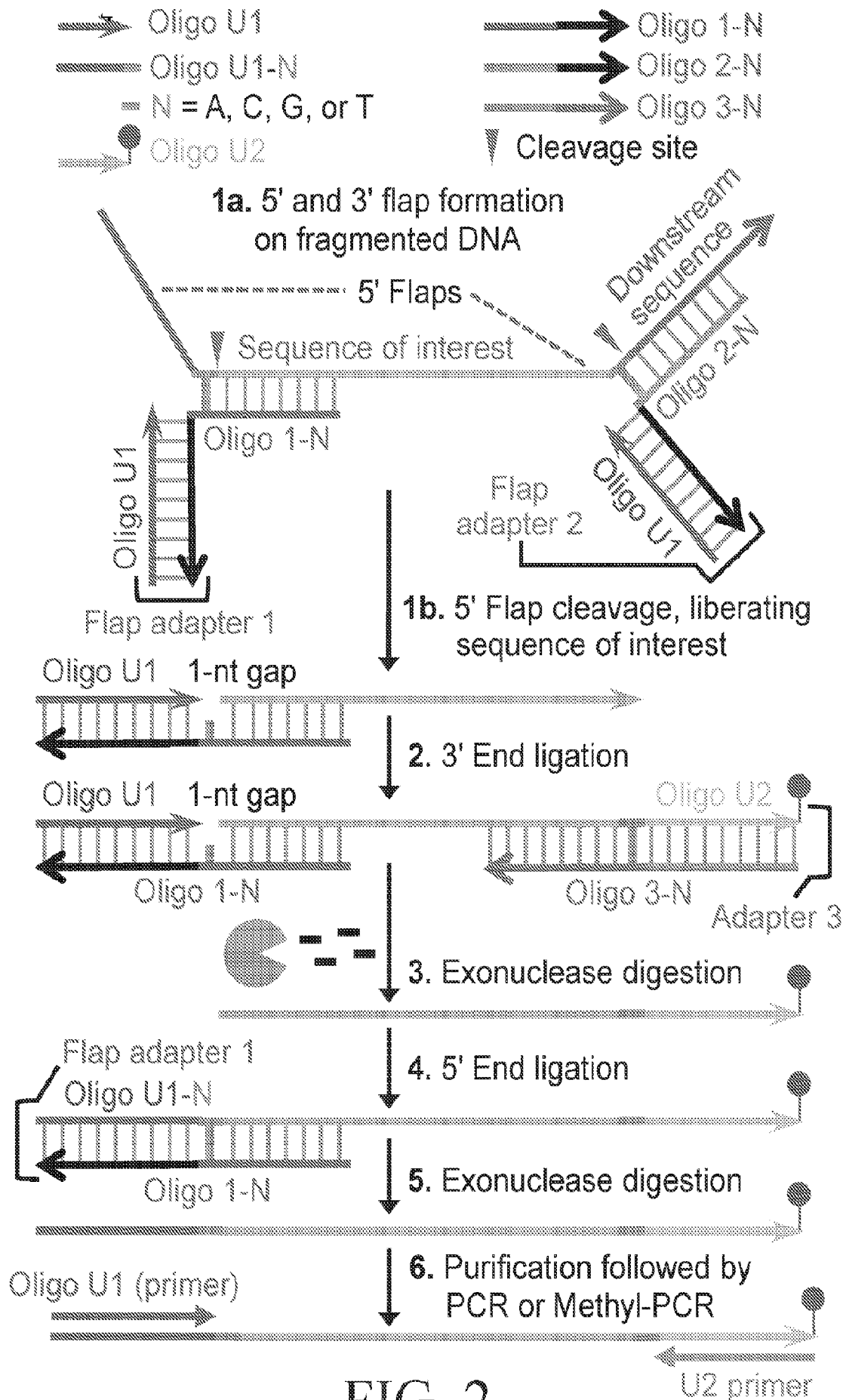


FIG. 2

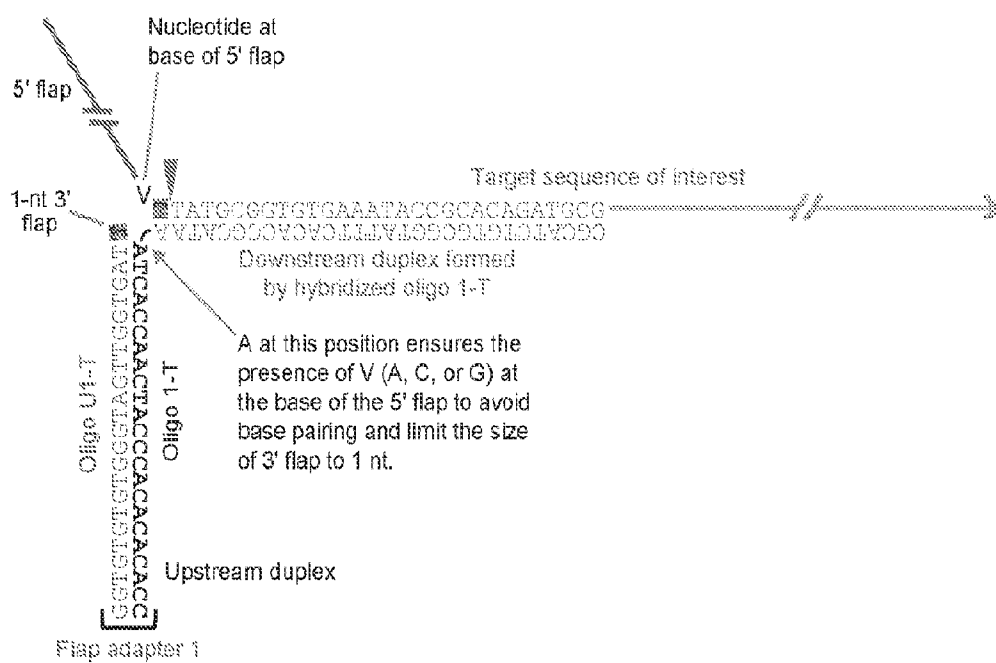


FIG. 3

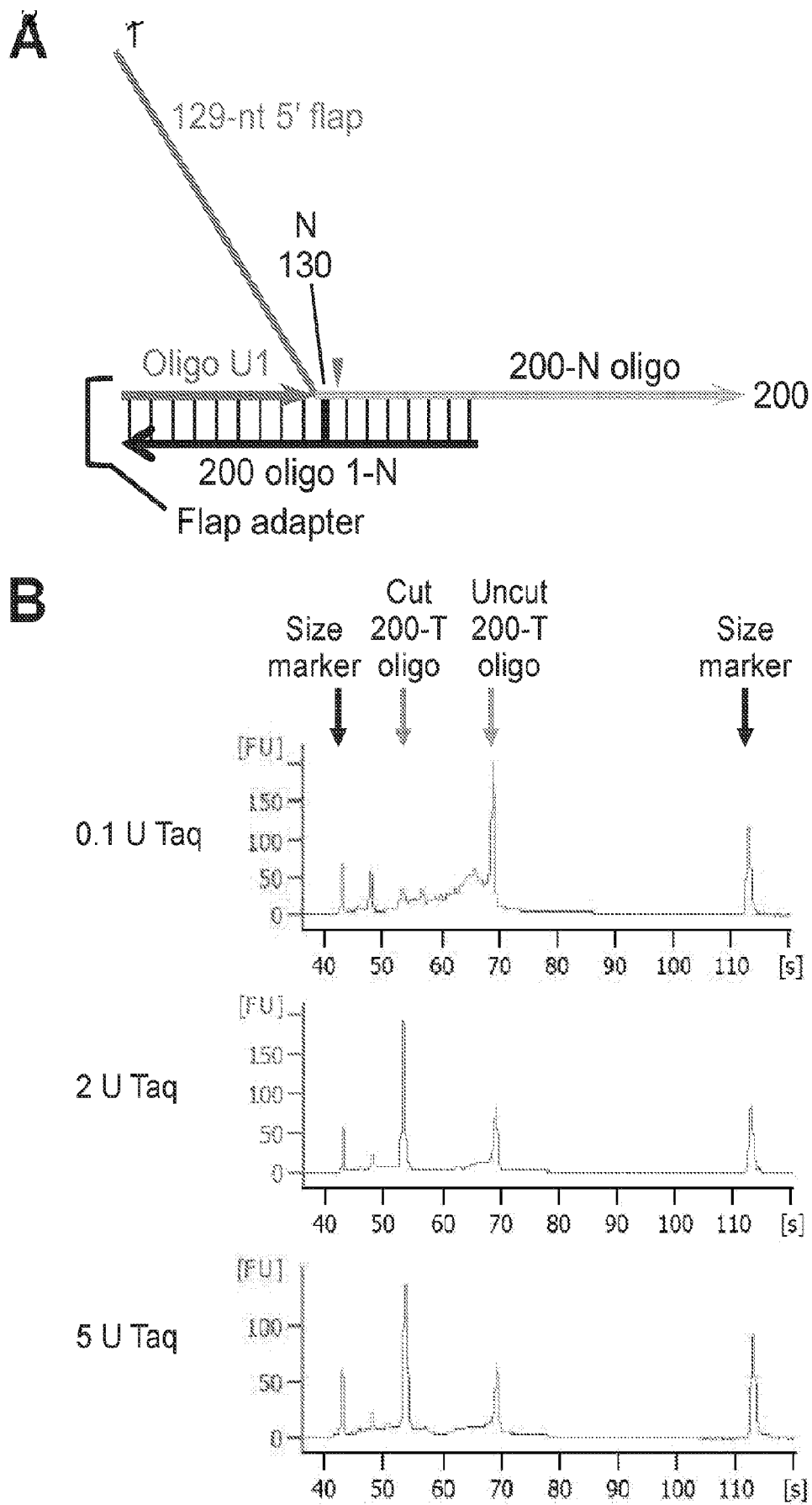


FIG. 4

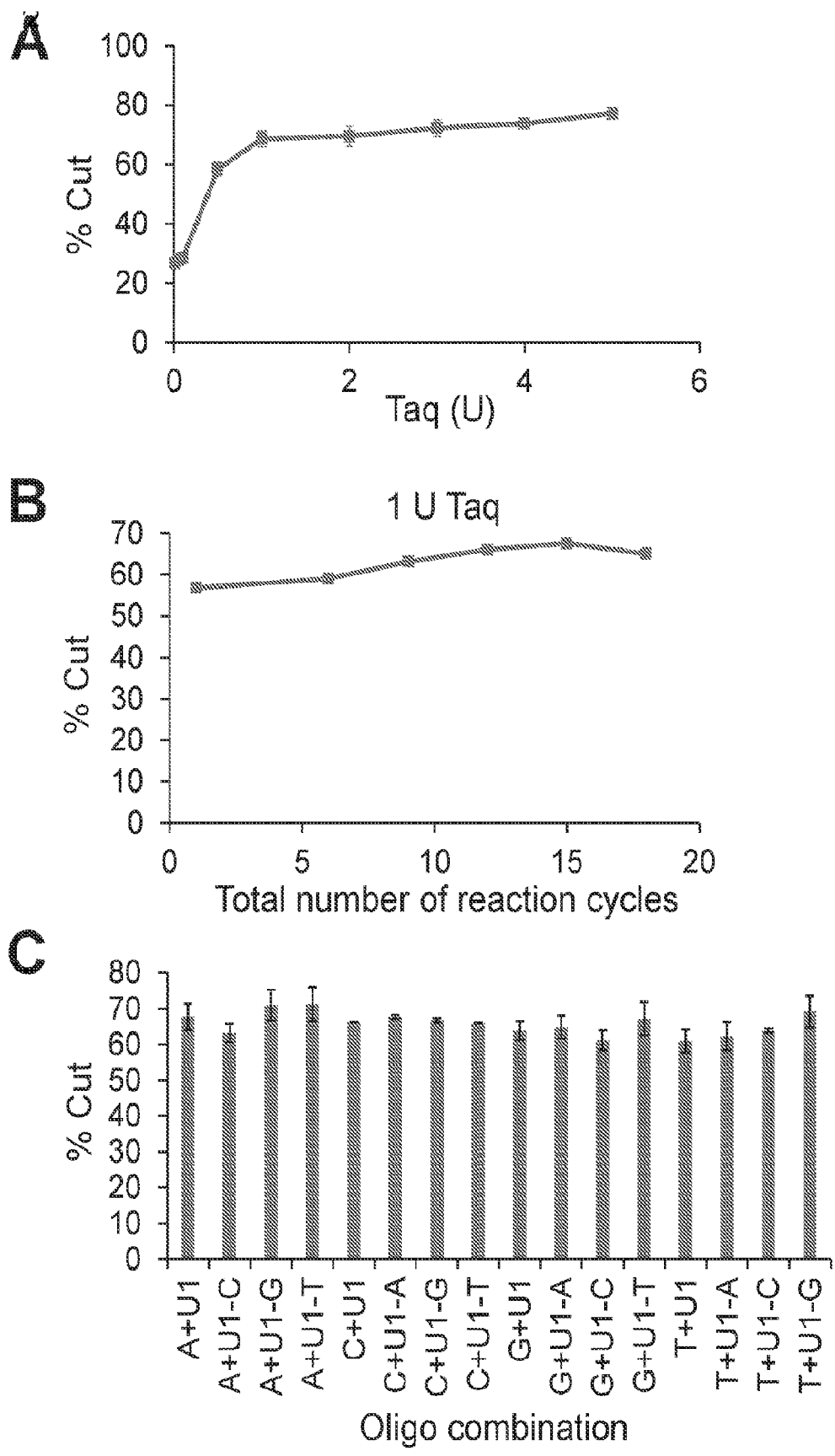


FIG. 5

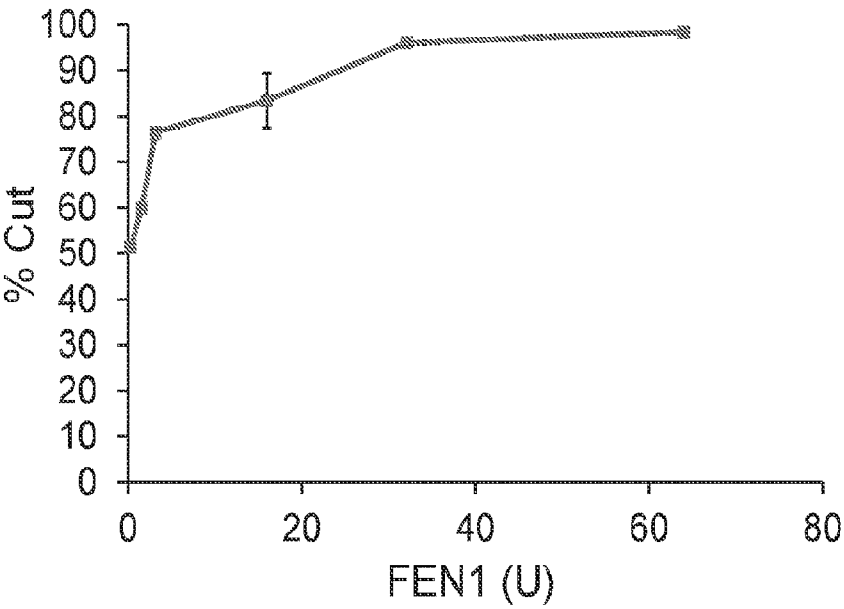


FIG. 6

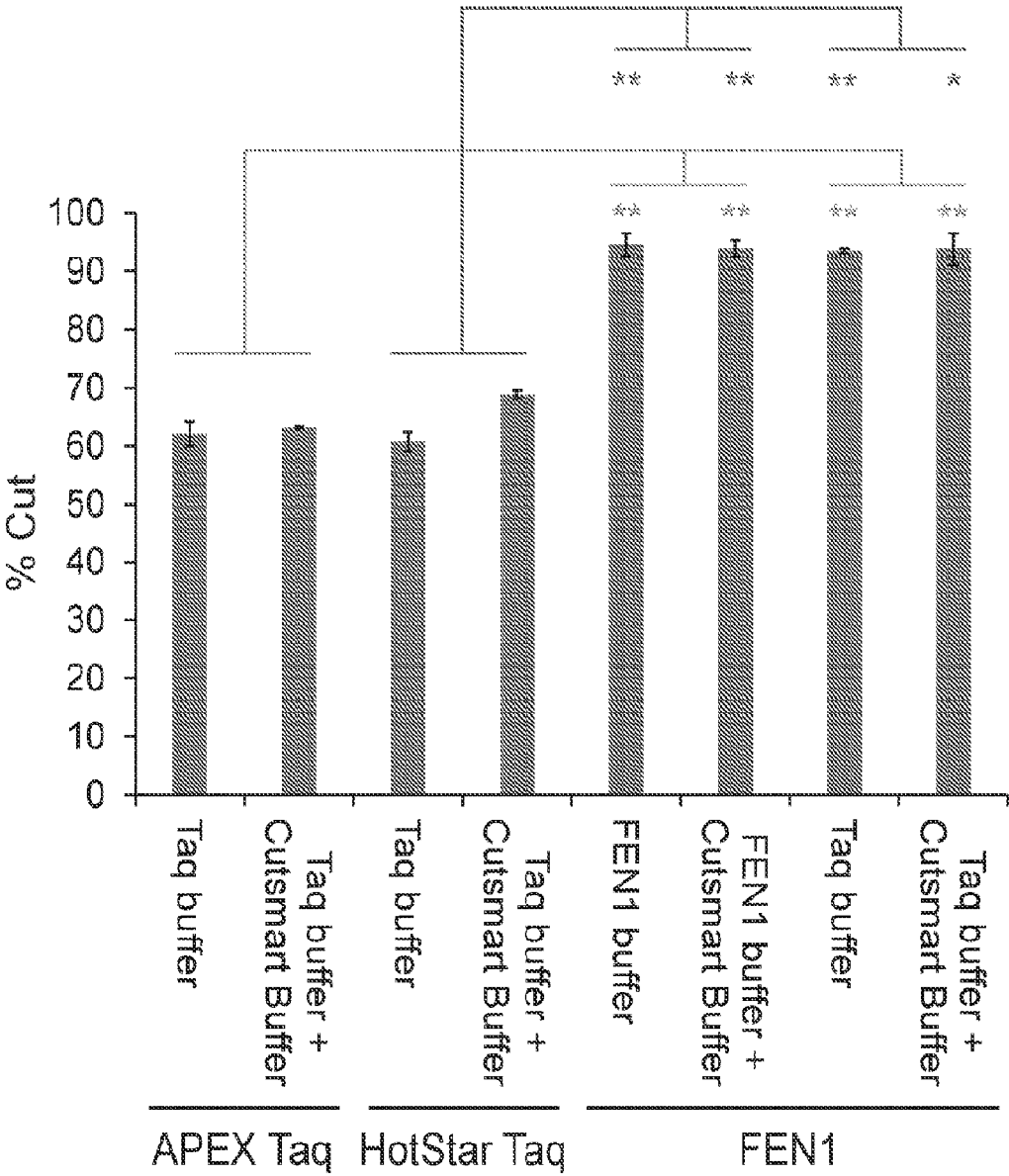


FIG. 7

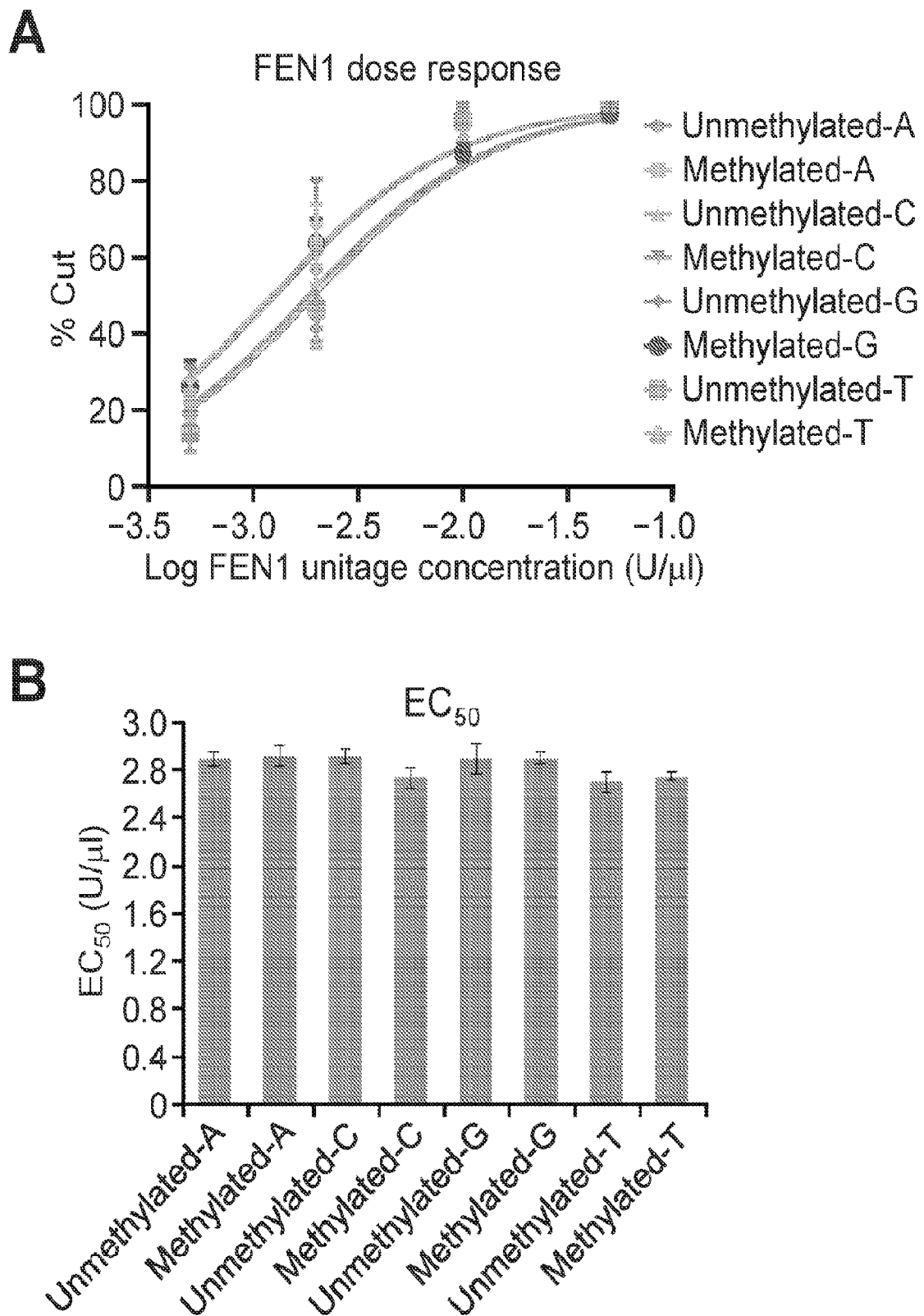


FIG. 8

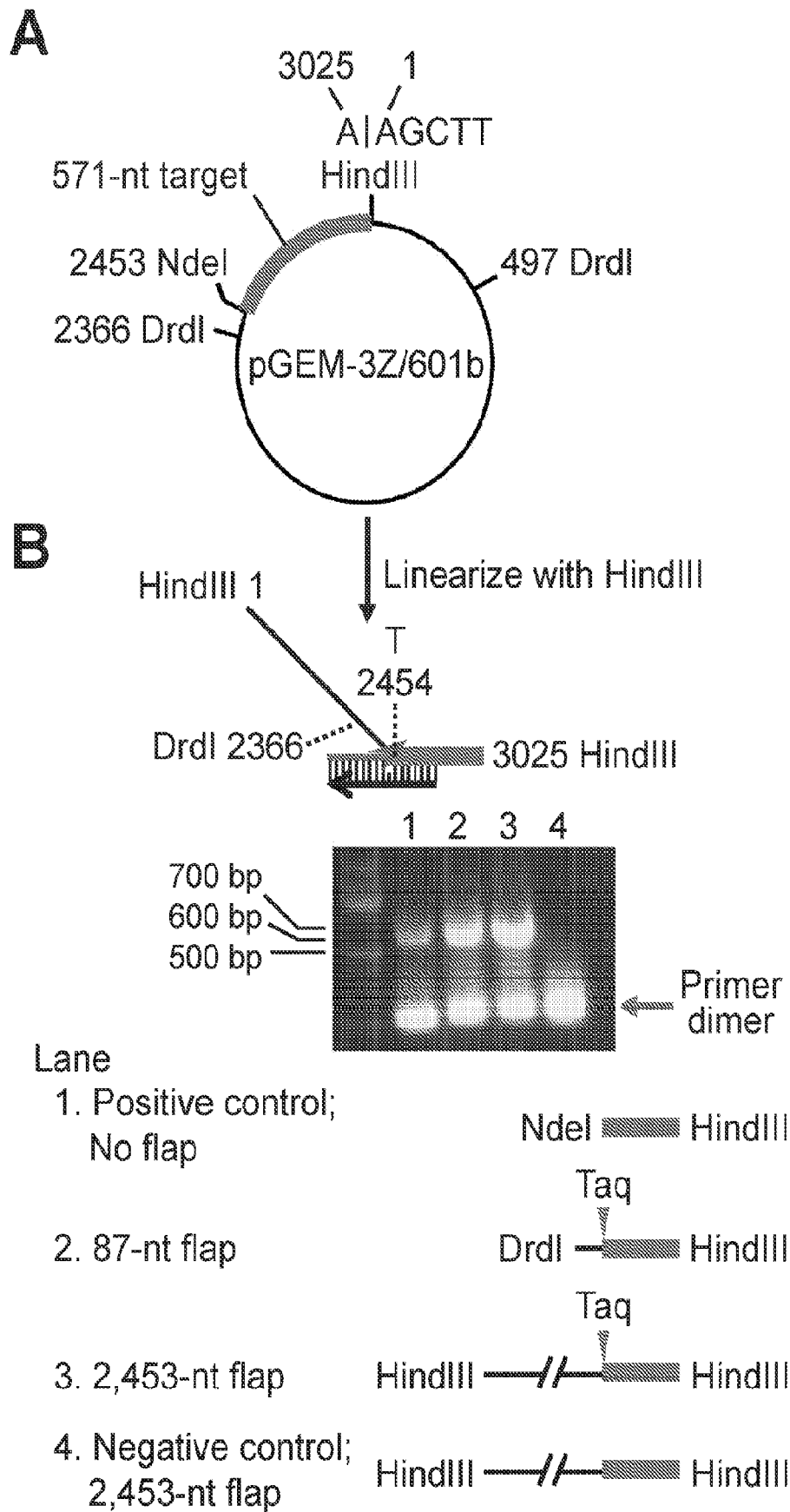


FIG. 9

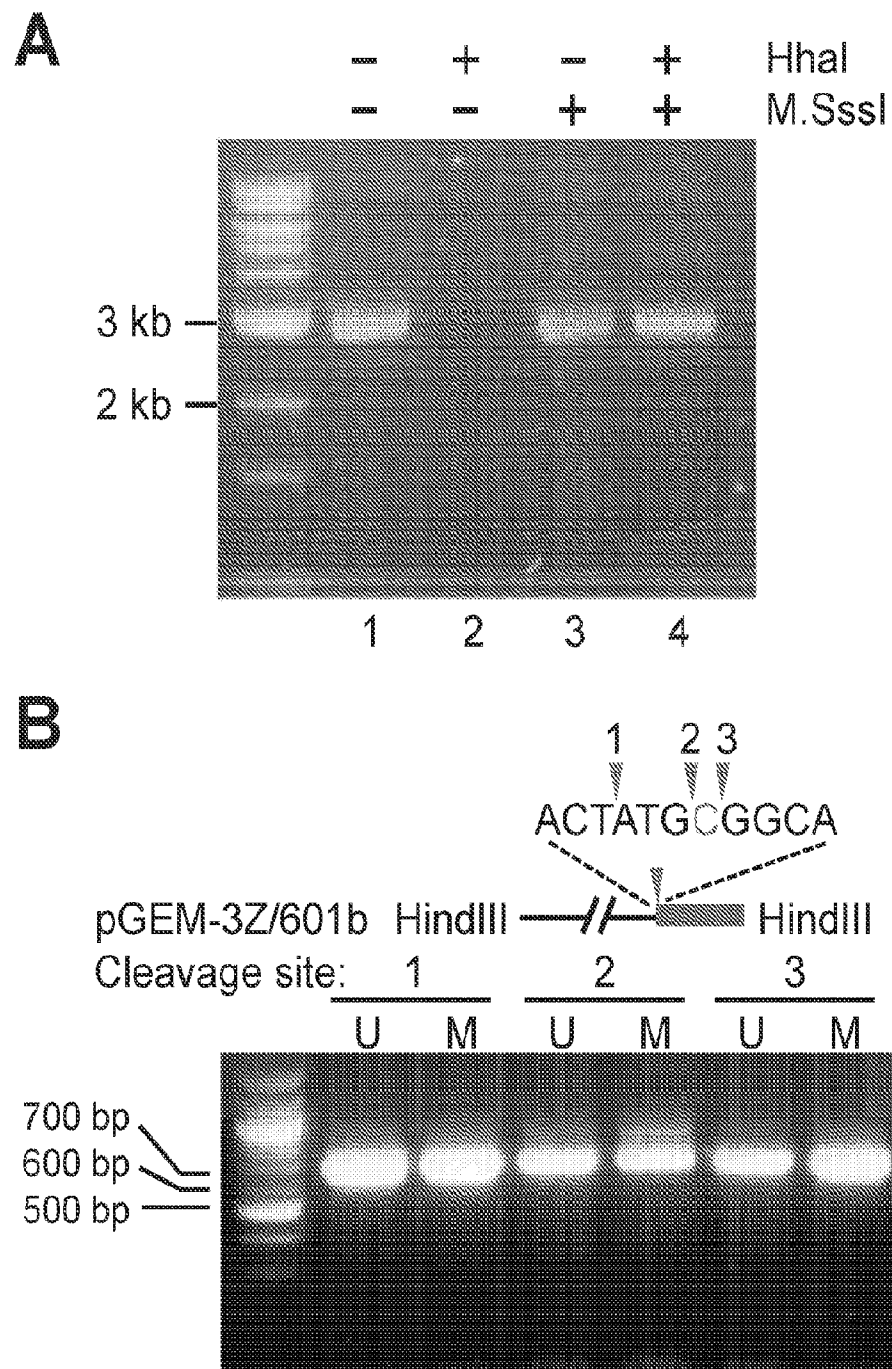
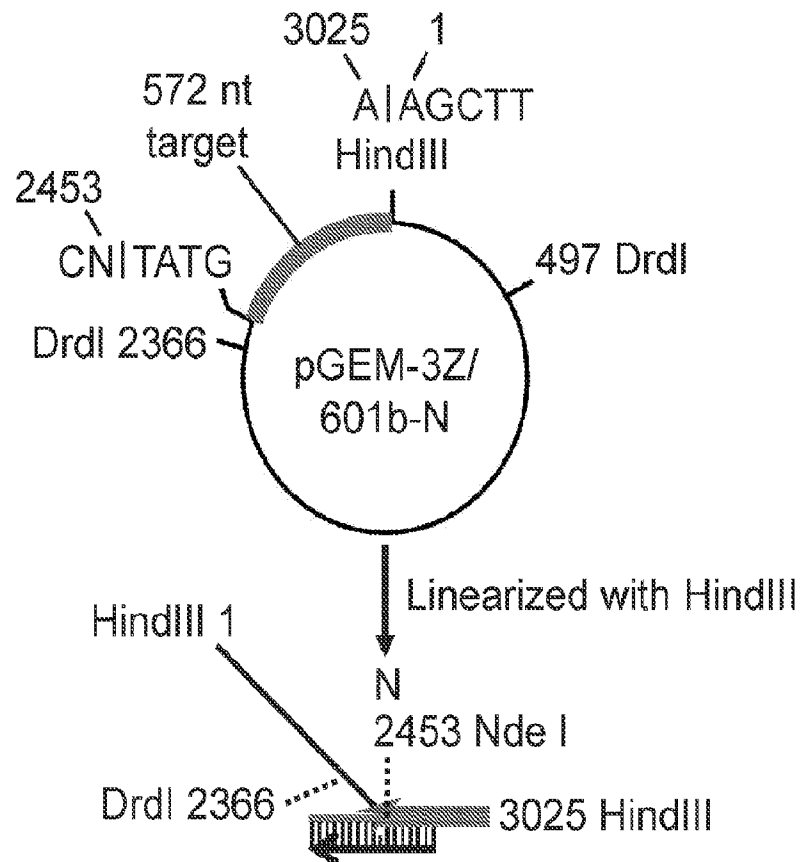
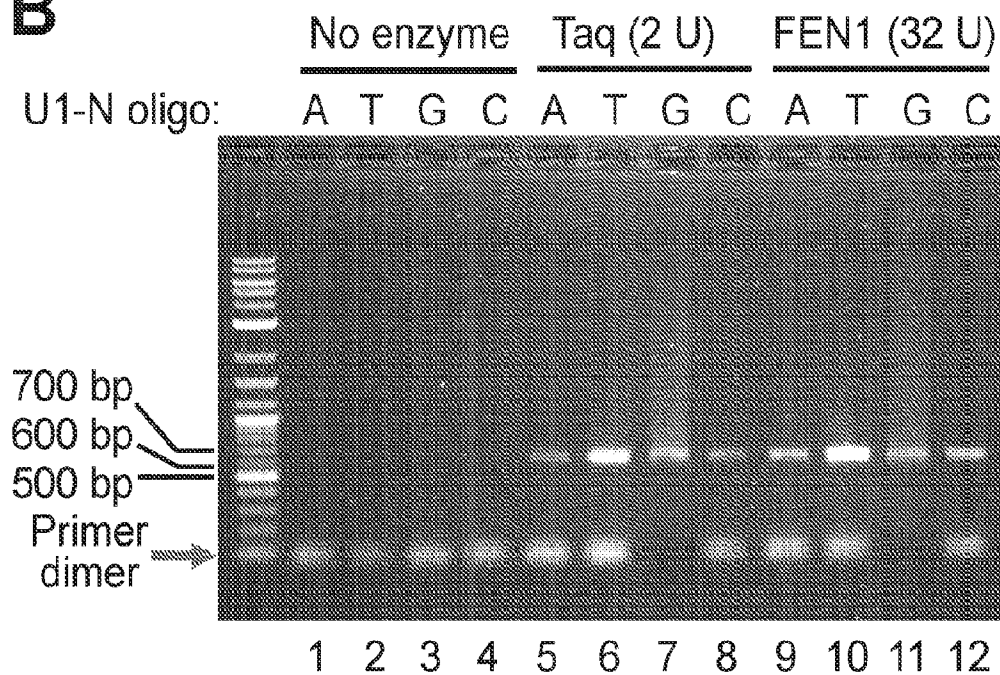


FIG. 10

A**B****FIG. 11**

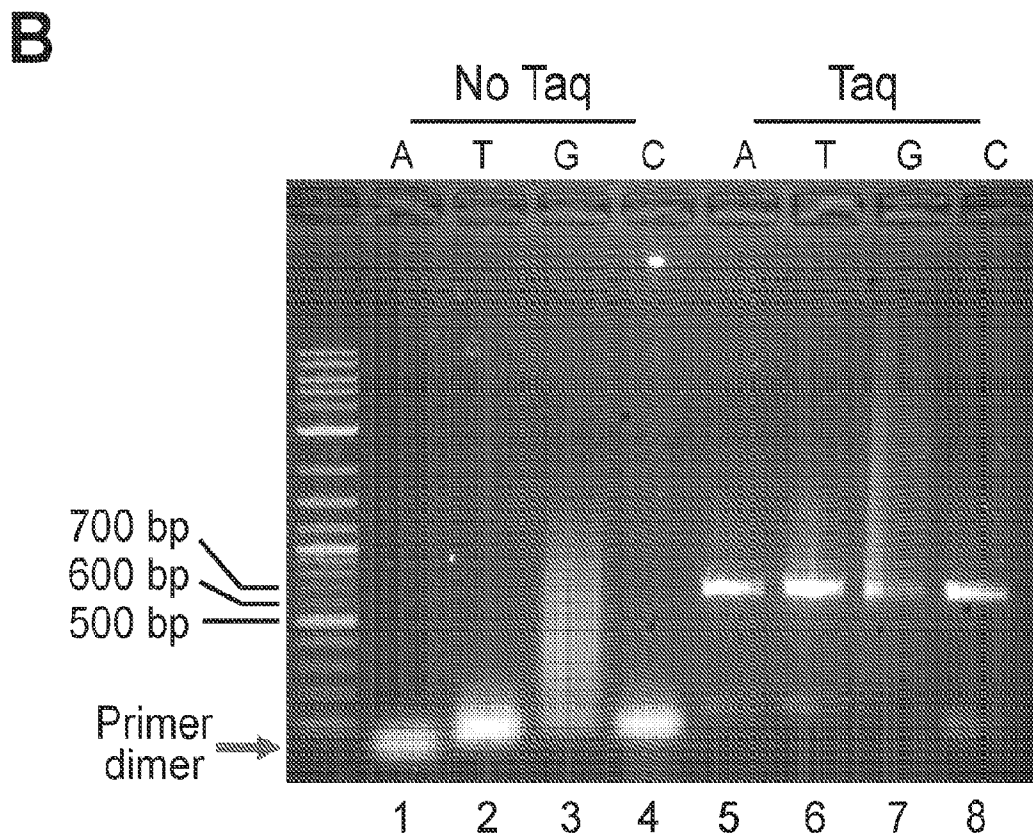
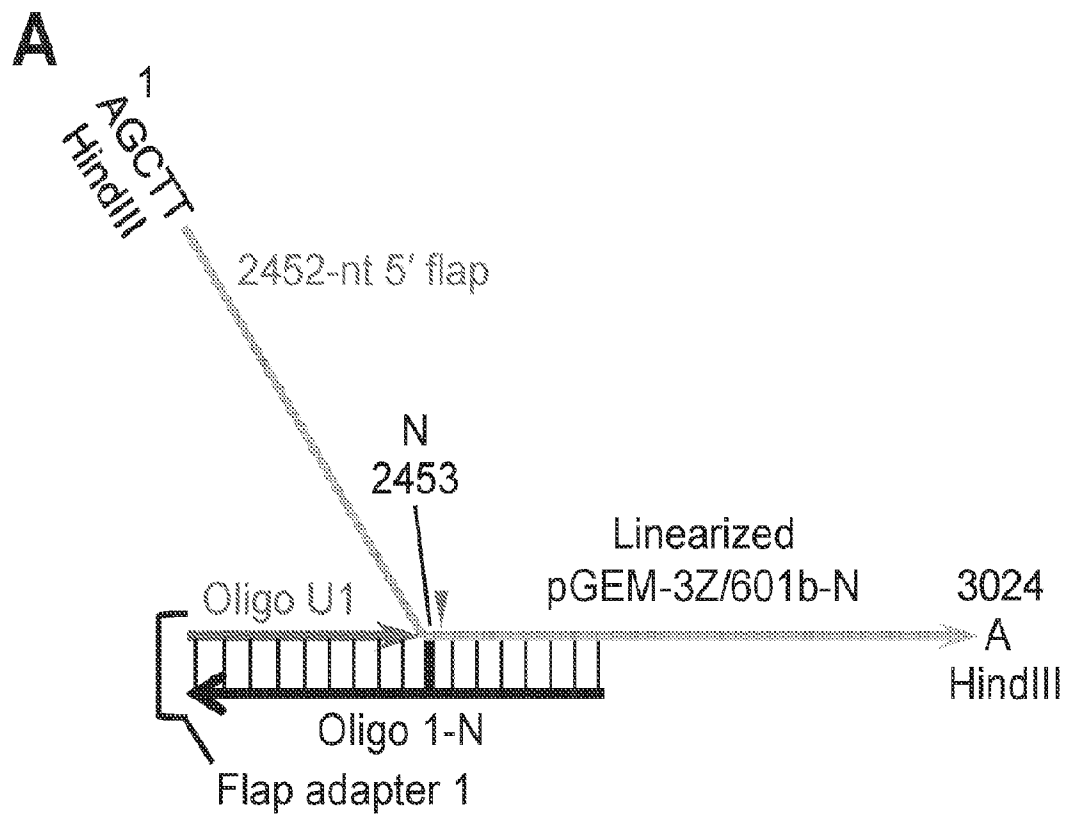


FIG. 12

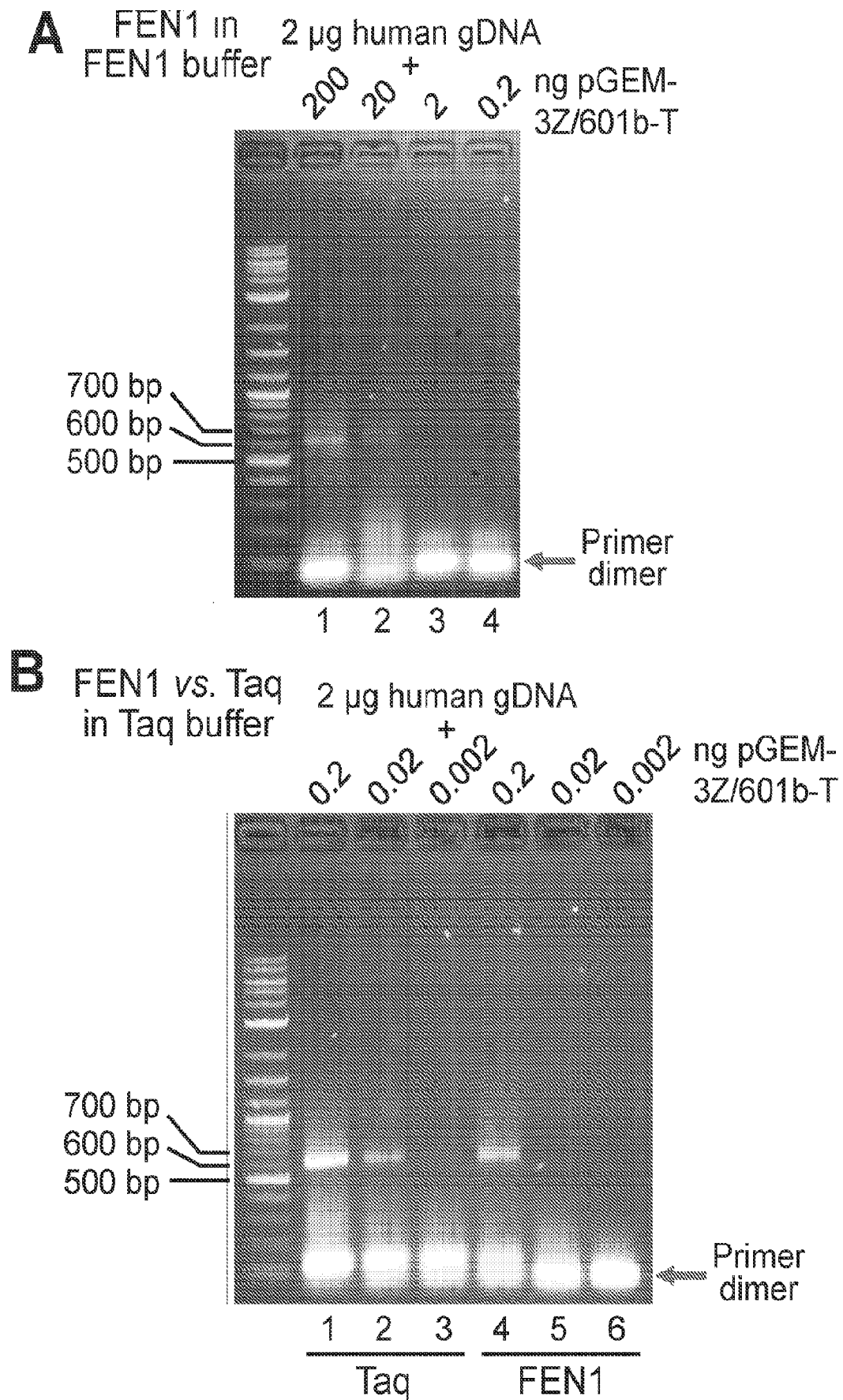


FIG. 13

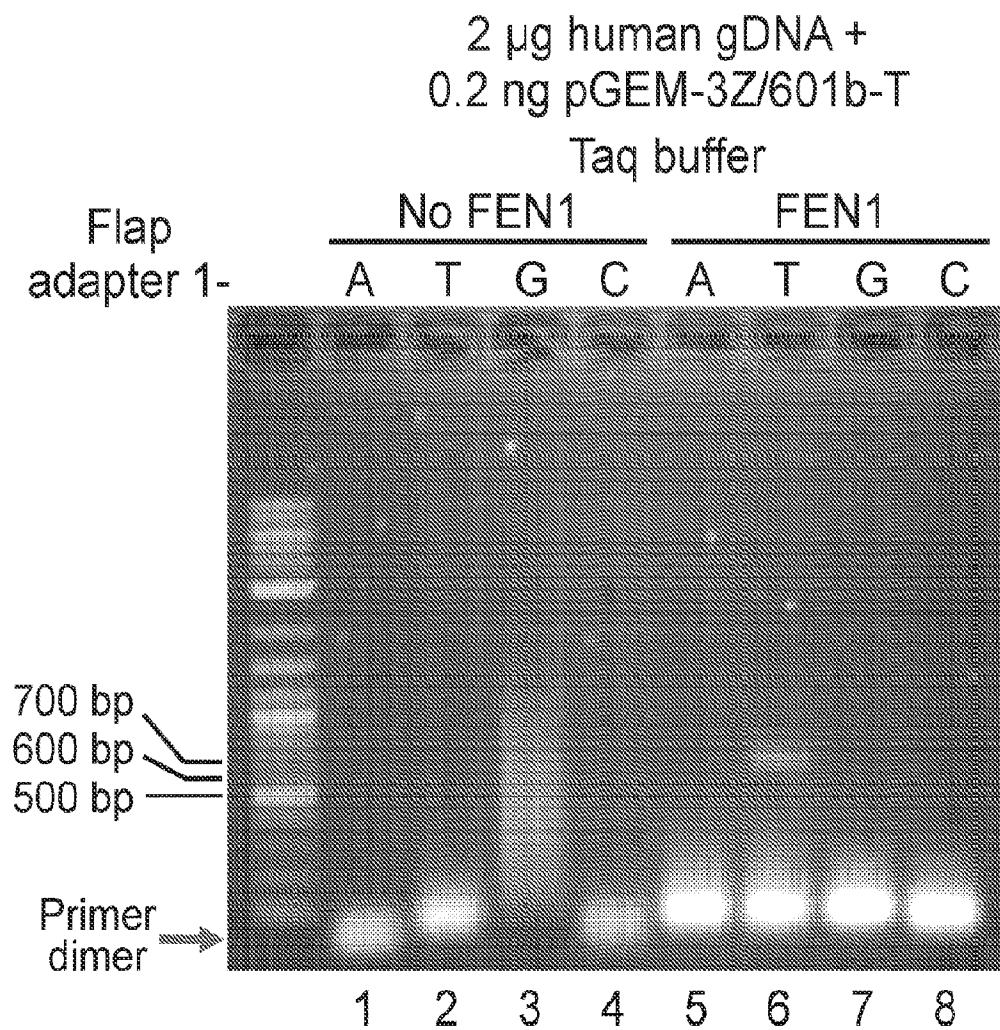


FIG. 14

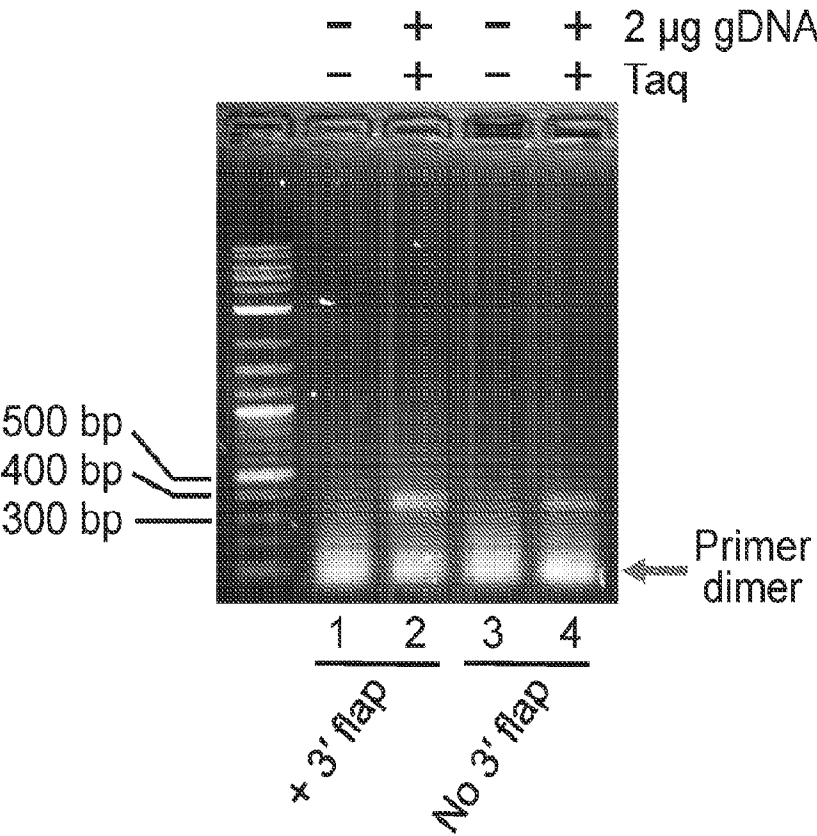


FIG. 15

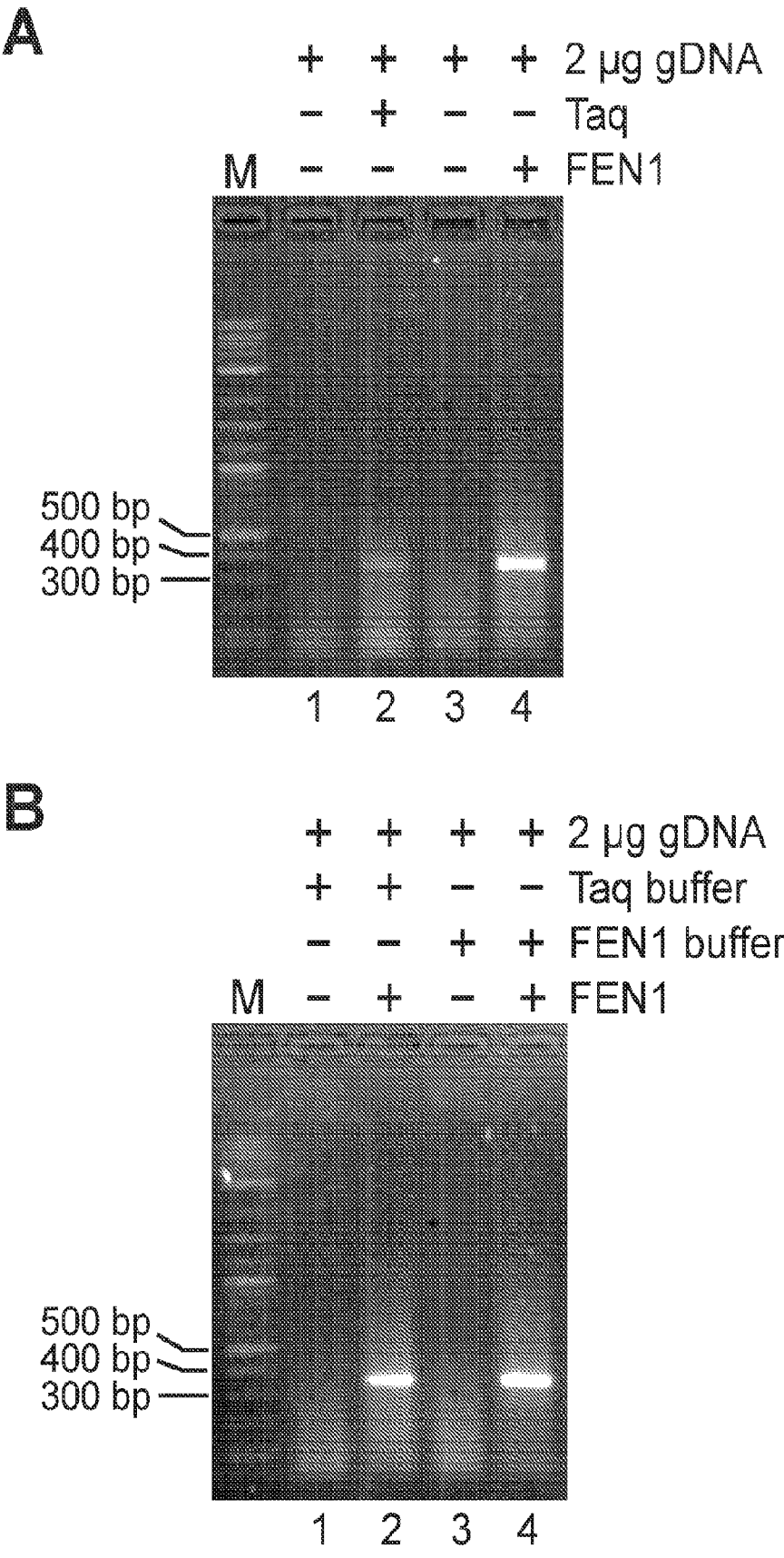


FIG. 16

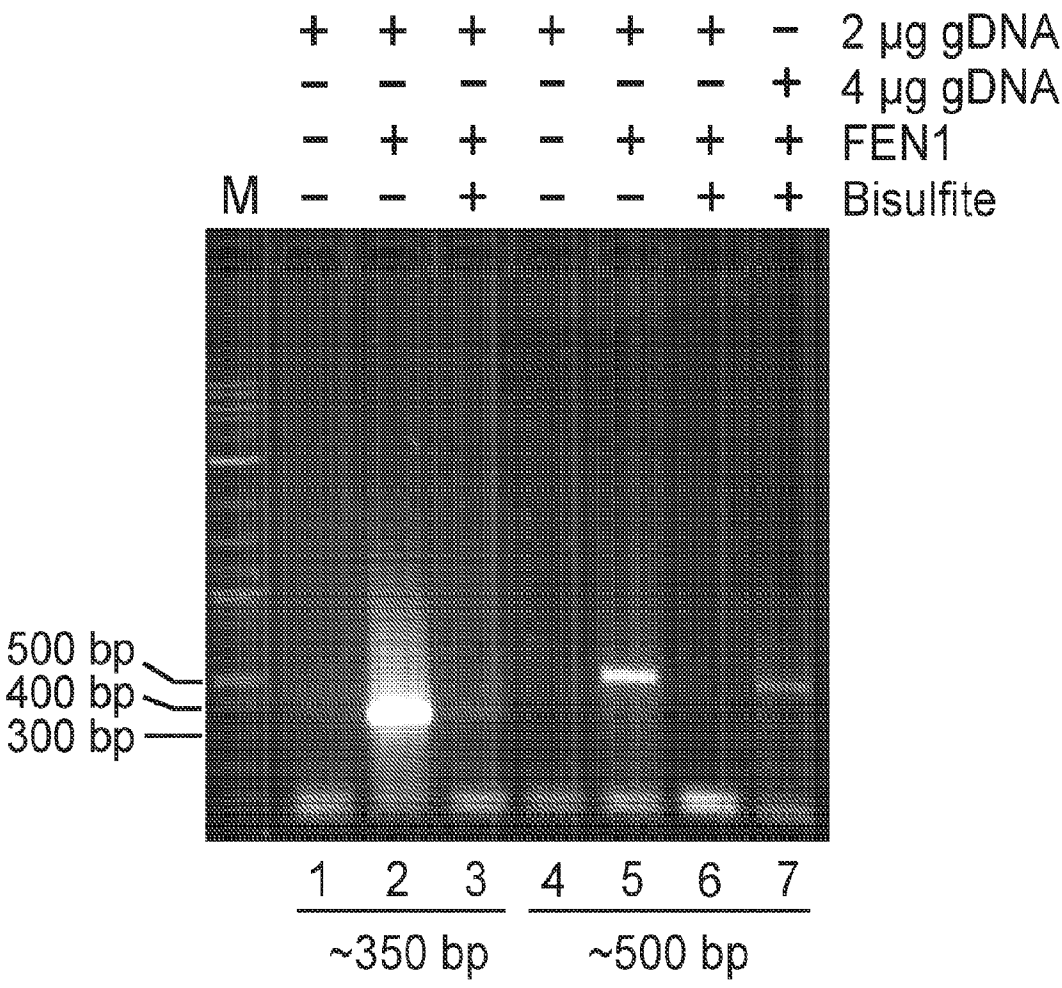


FIG. 17

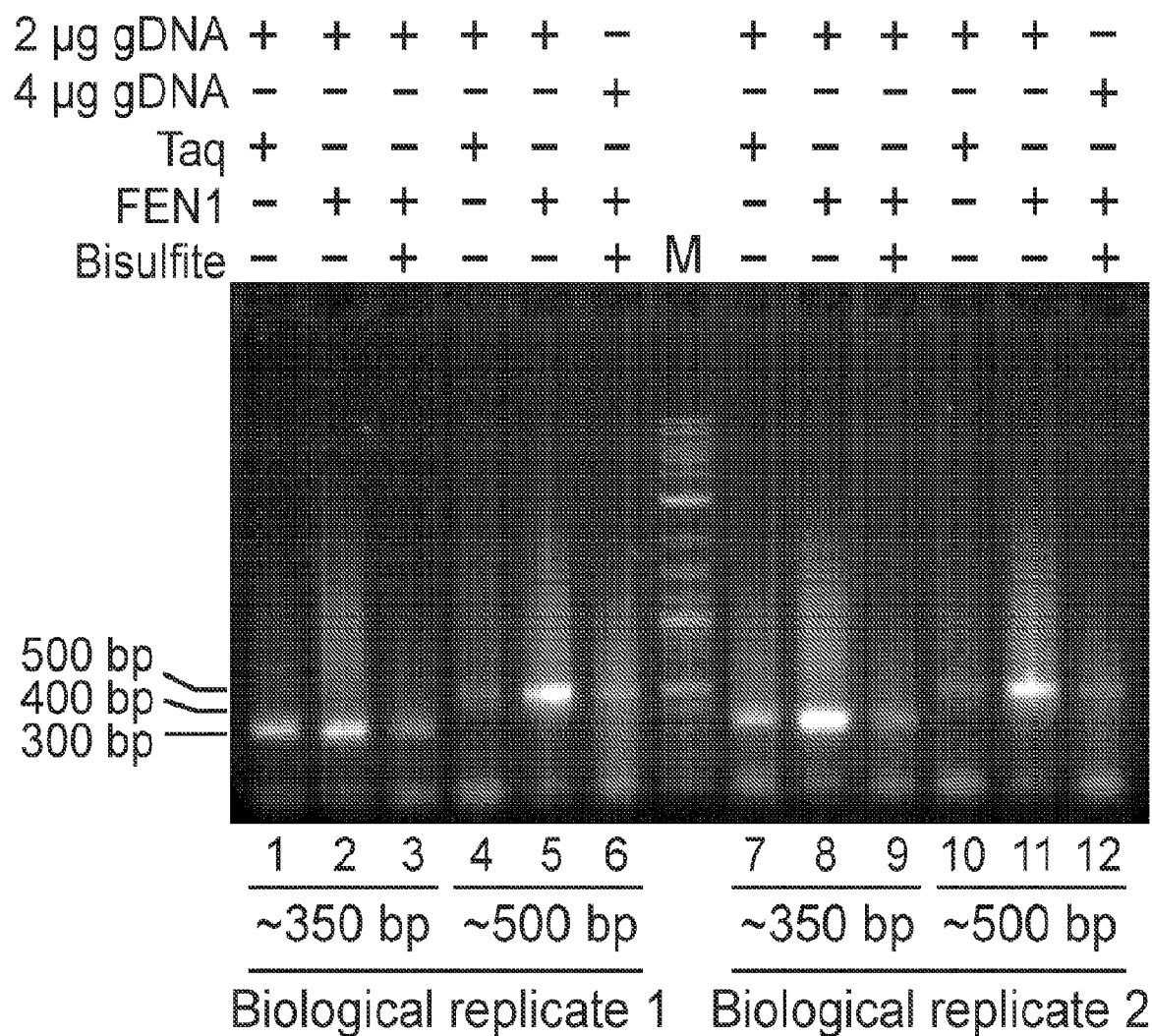


FIG. 18

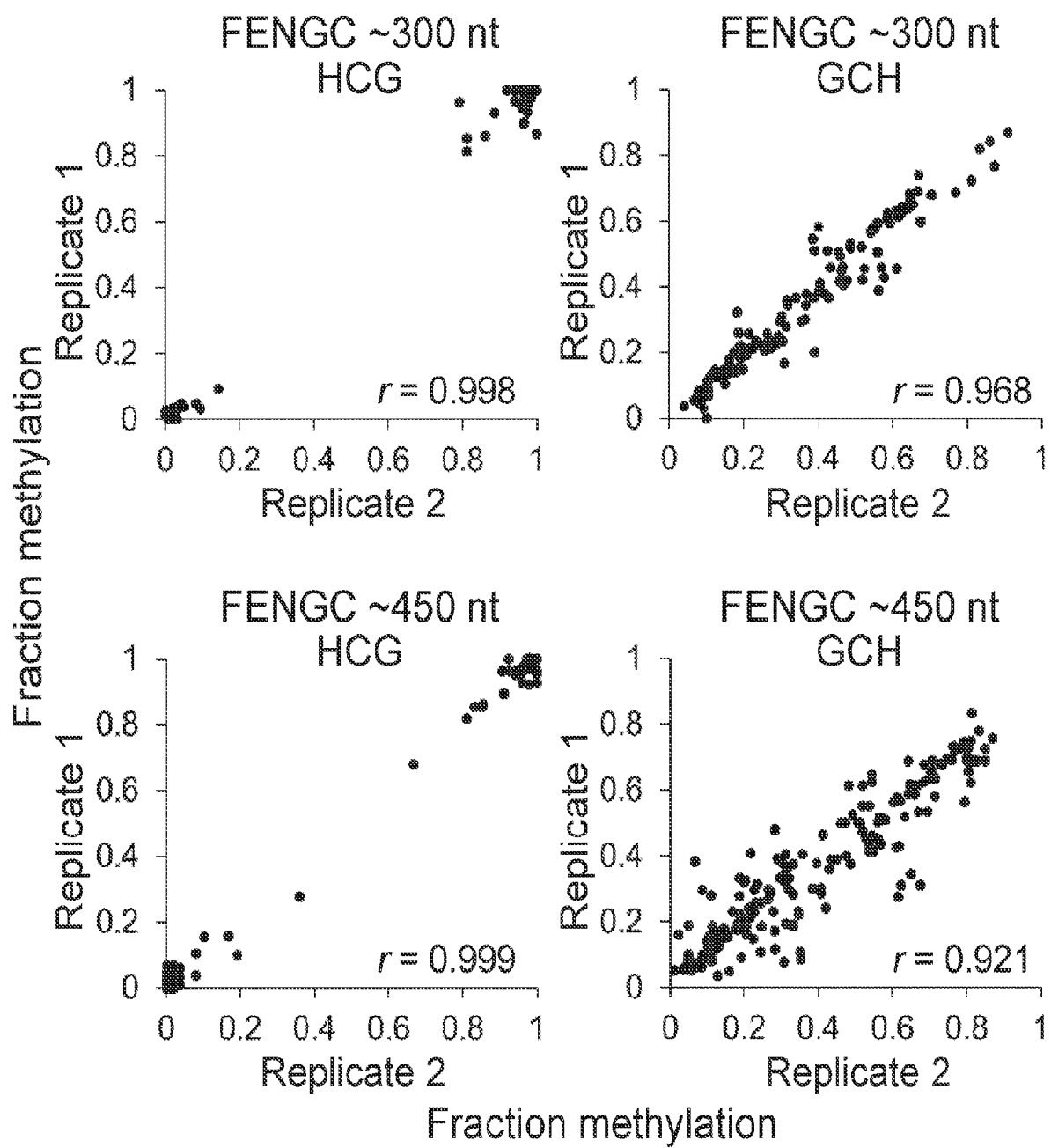


FIG. 19

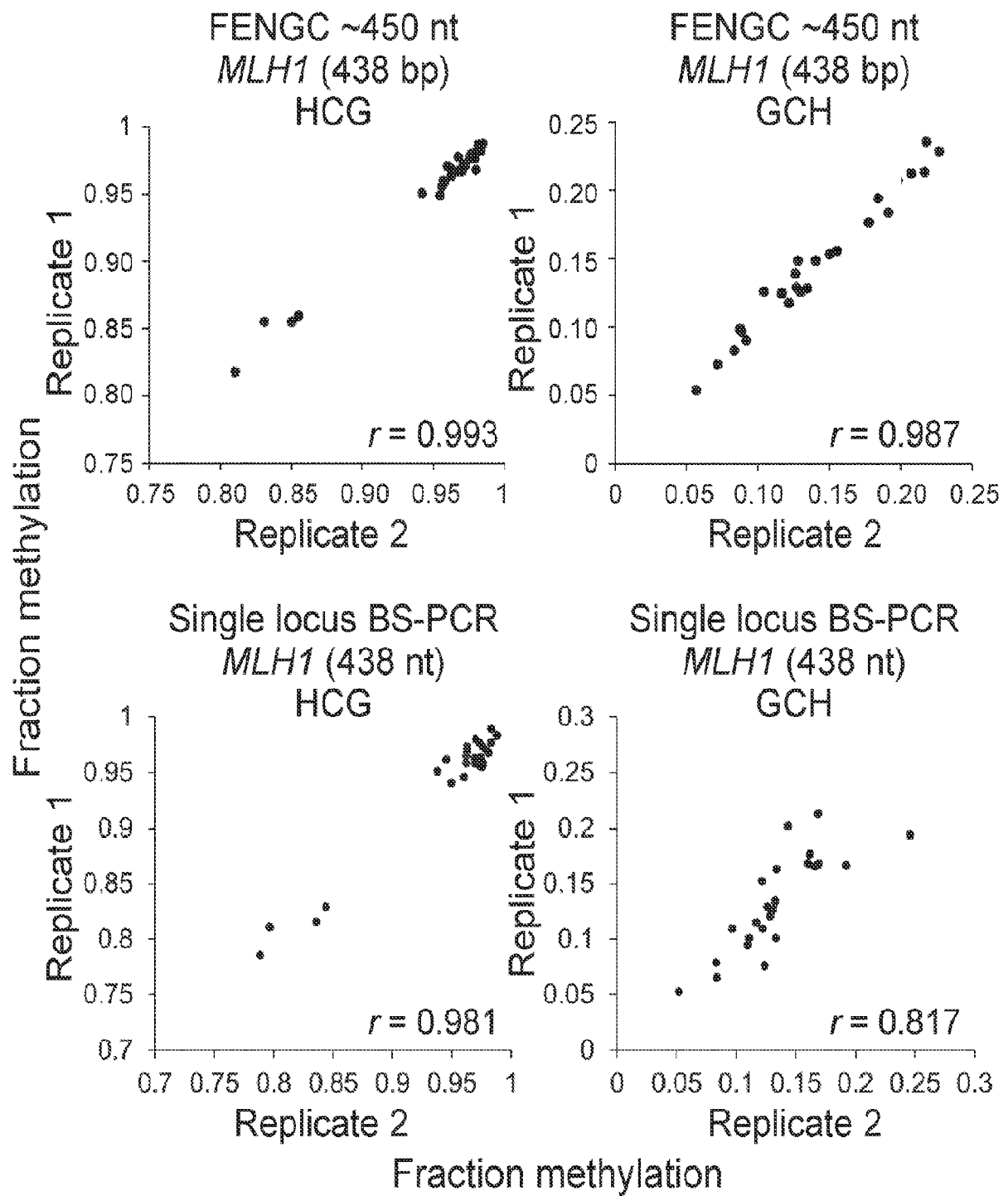


FIG. 20

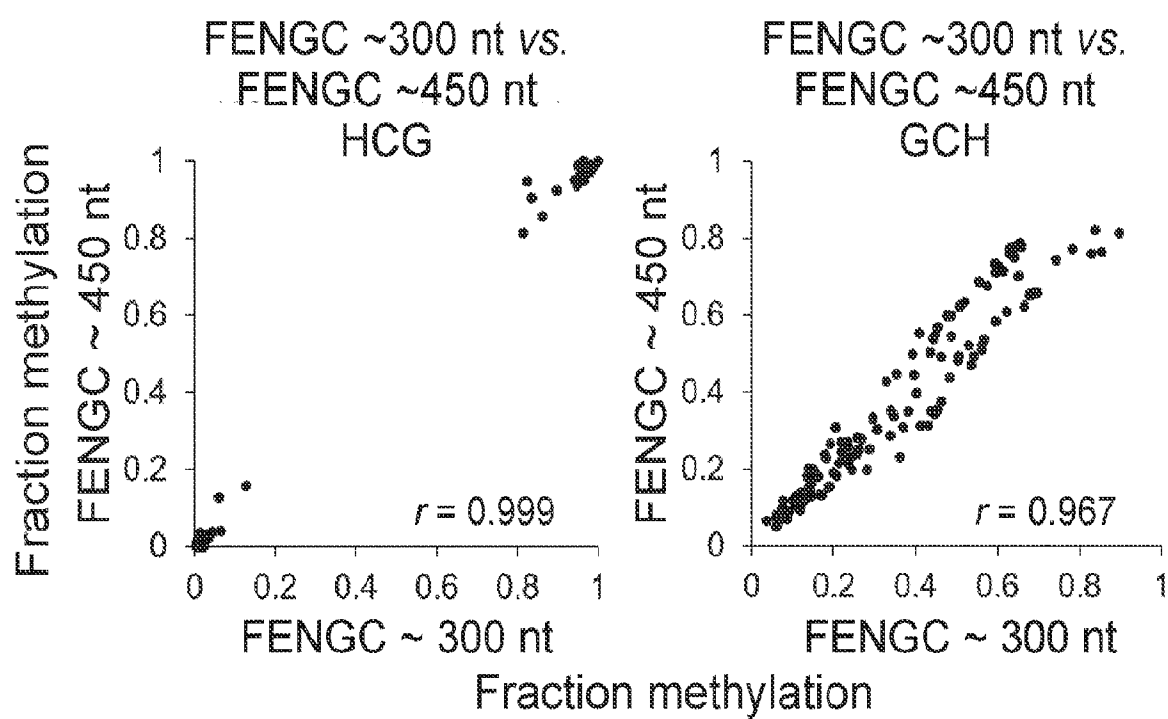


FIG. 21

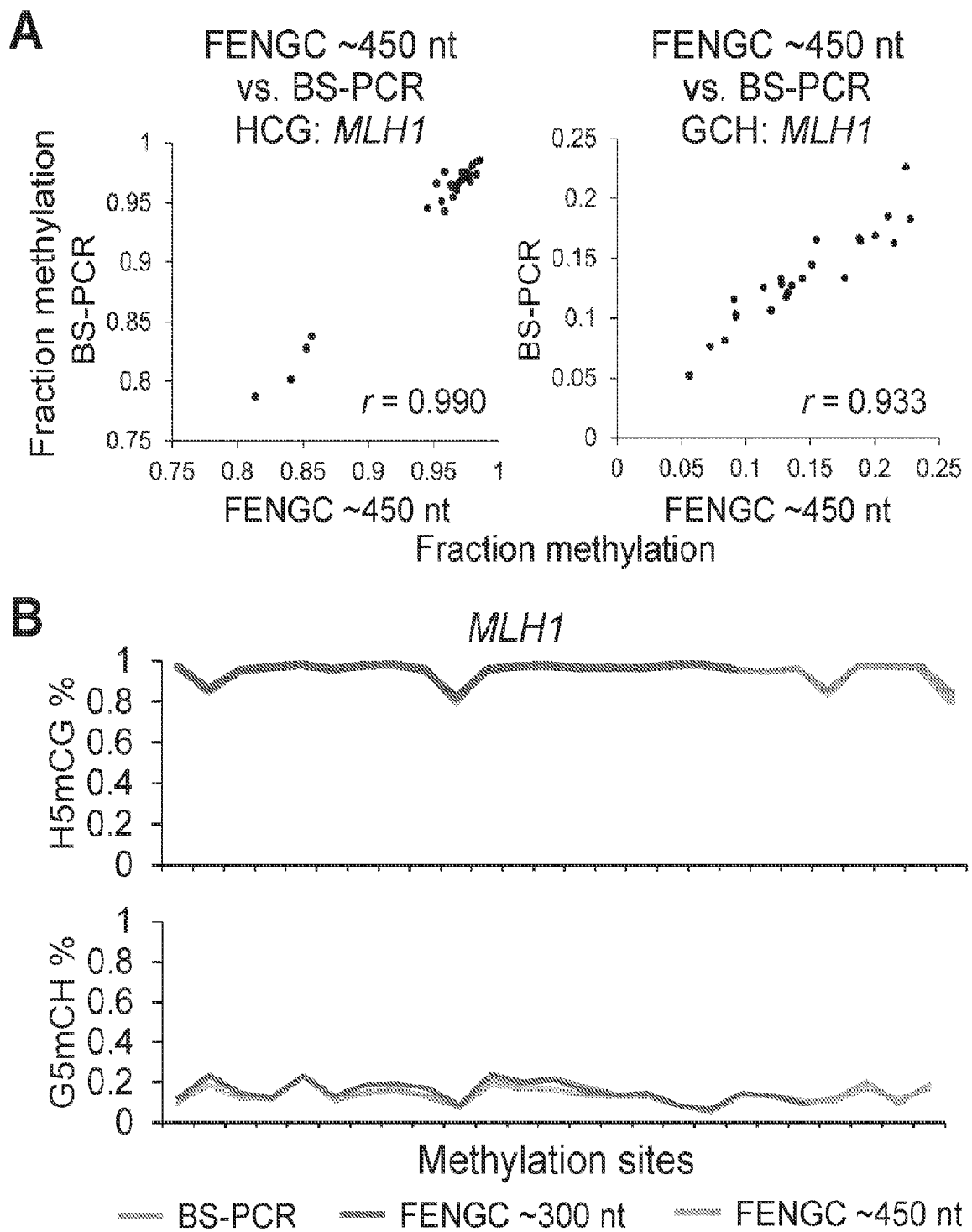


FIG. 22

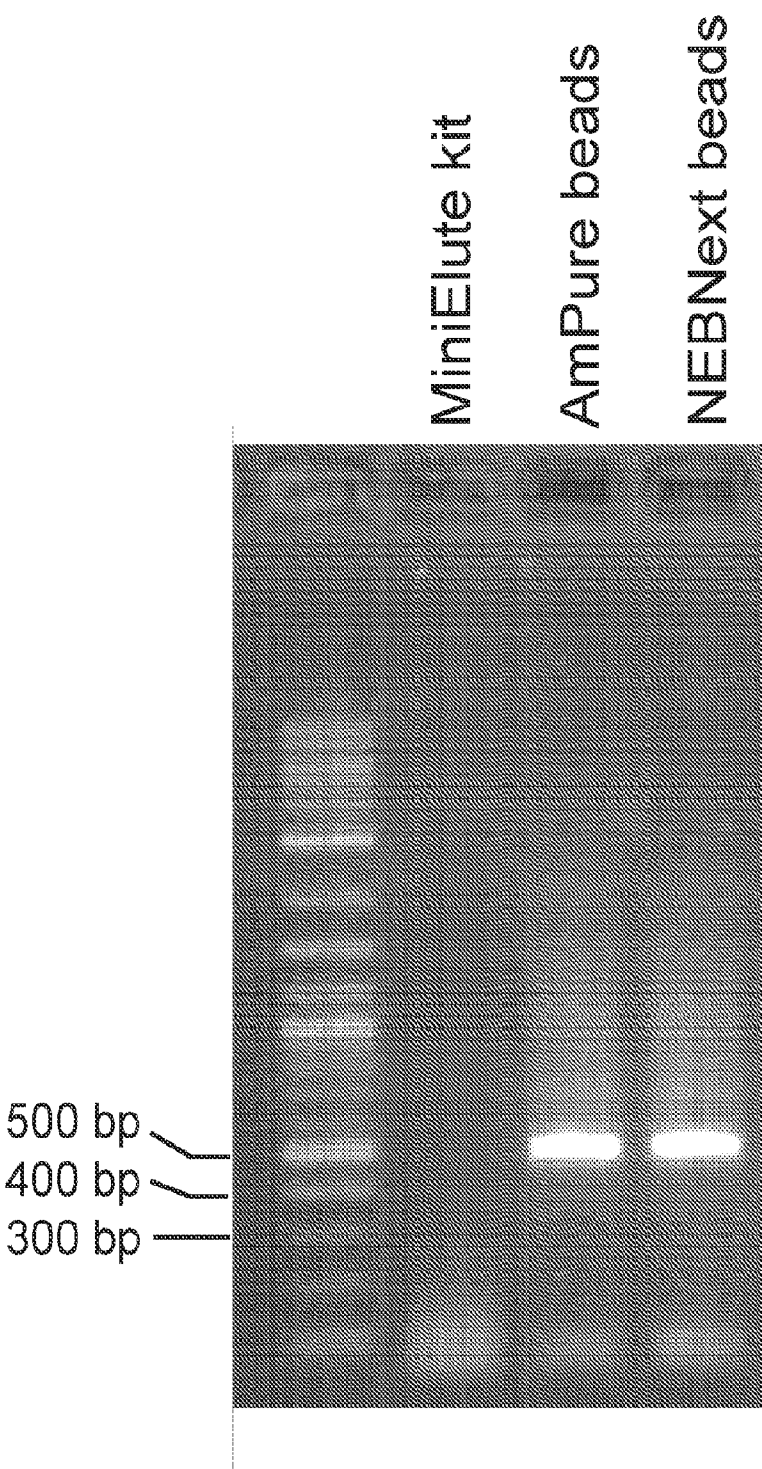
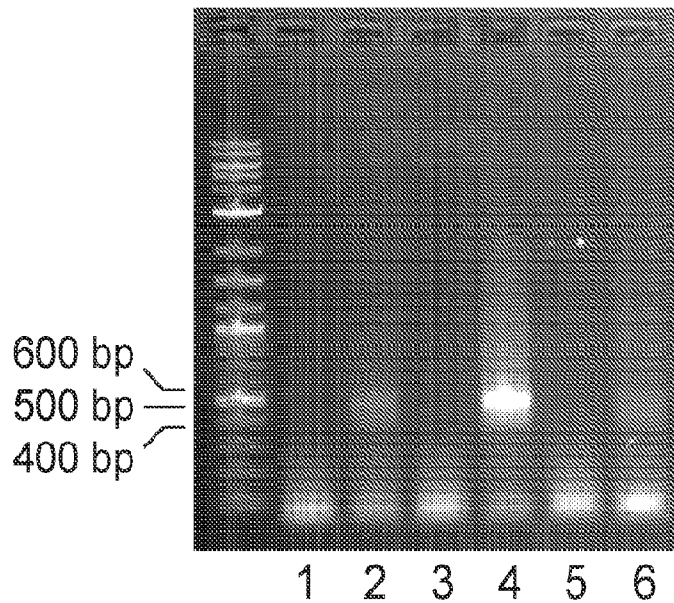


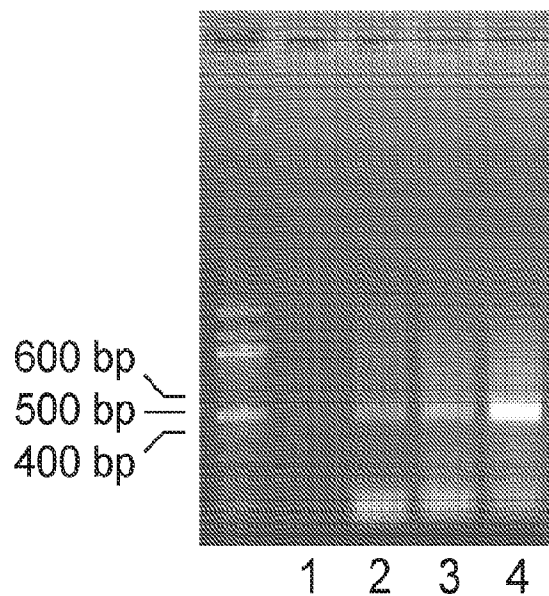
FIG. 23

A

25 25 10 10 2.5 2.5 nM each oligos 1-T, 2-T, 3-T
 25 25 10 10 2.5 2.5 μ M each oligos U1-T, U2
 - + - + - + FEN1

**B**

0 5 10 15 nM each oligos 1-T, 2-T, 3-T
 10 10 10 10 μ M oligos U1-T, U2
 + + + + FEN1

**FIG. 24**

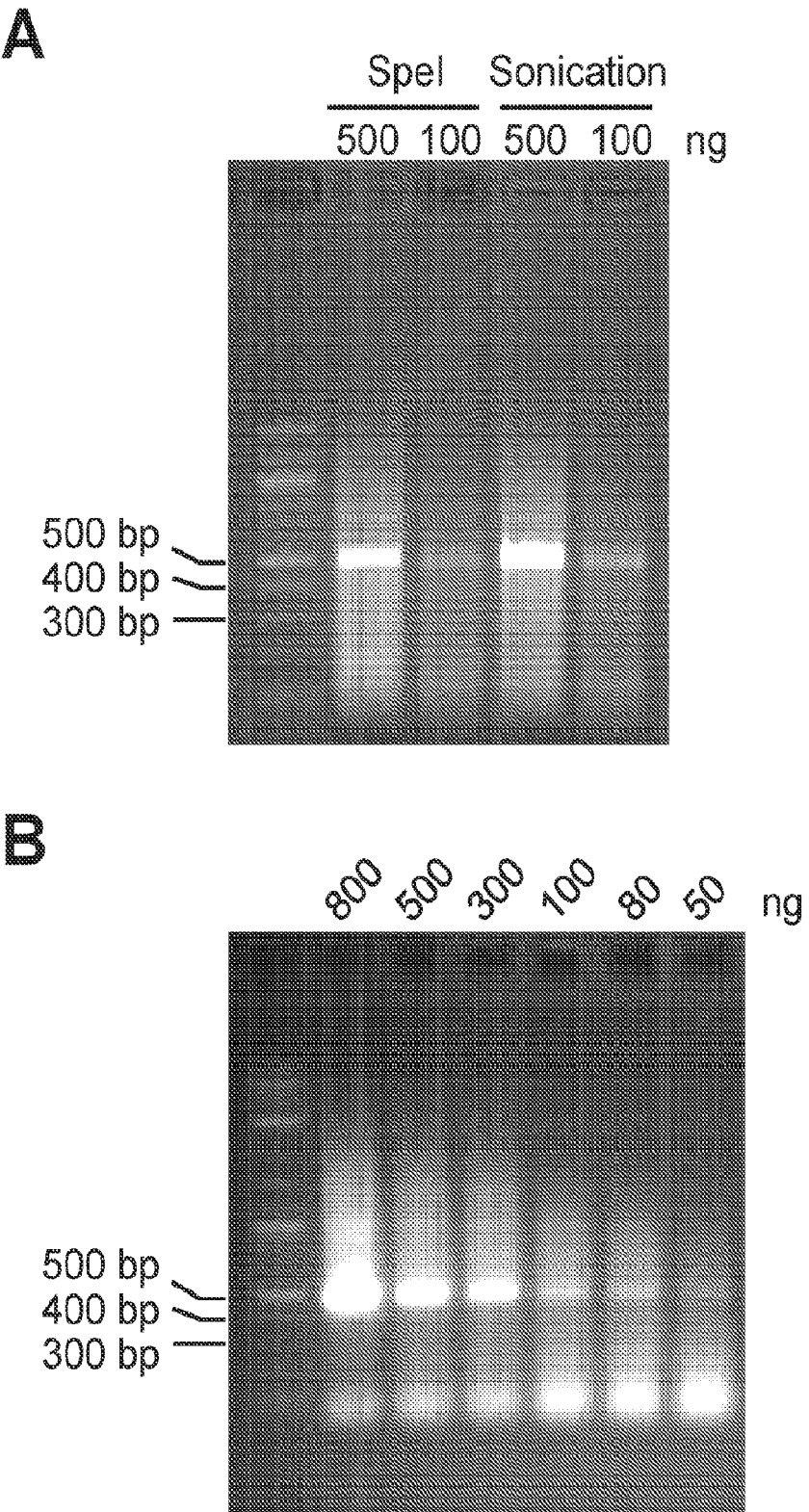


FIG. 25

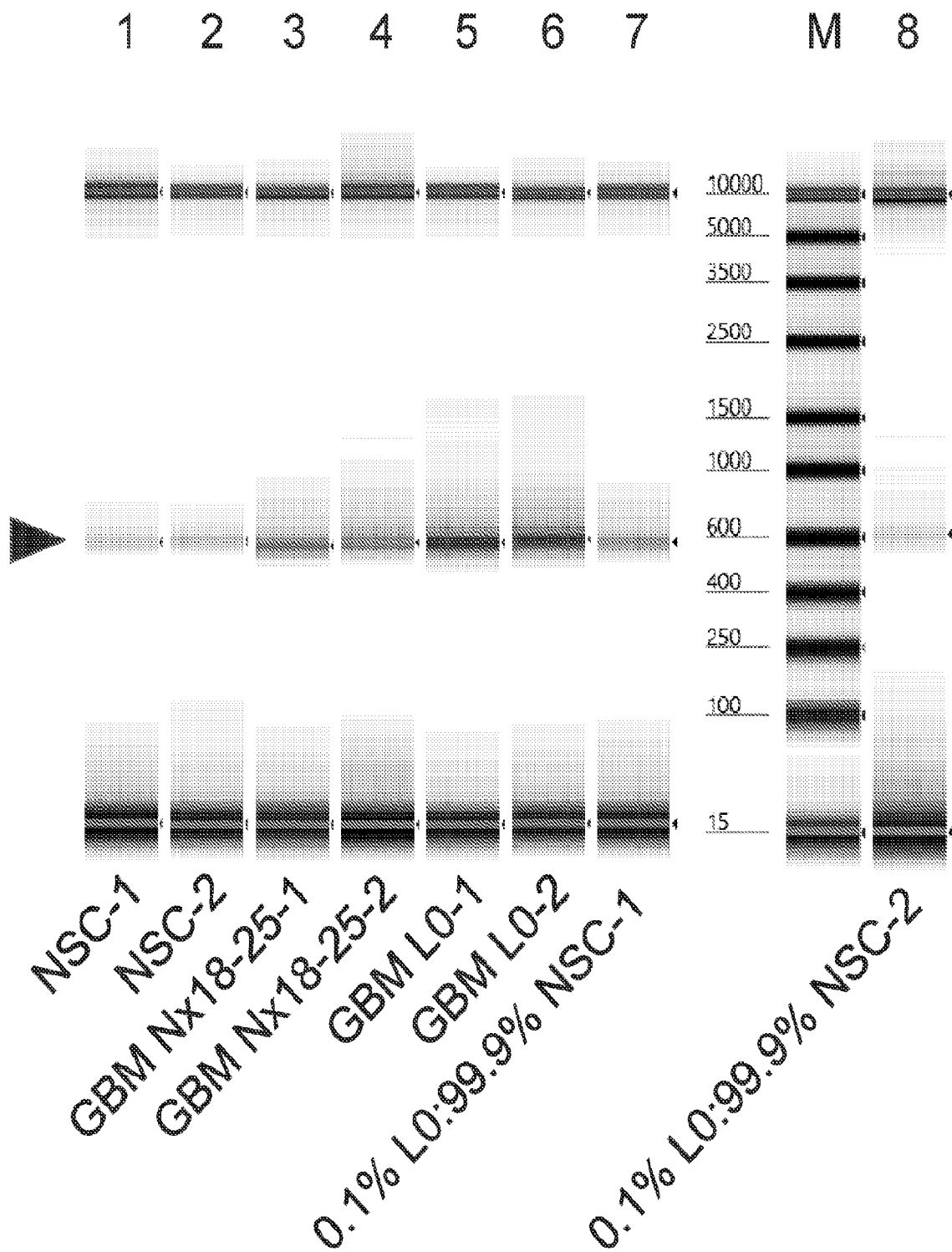


FIG. 26

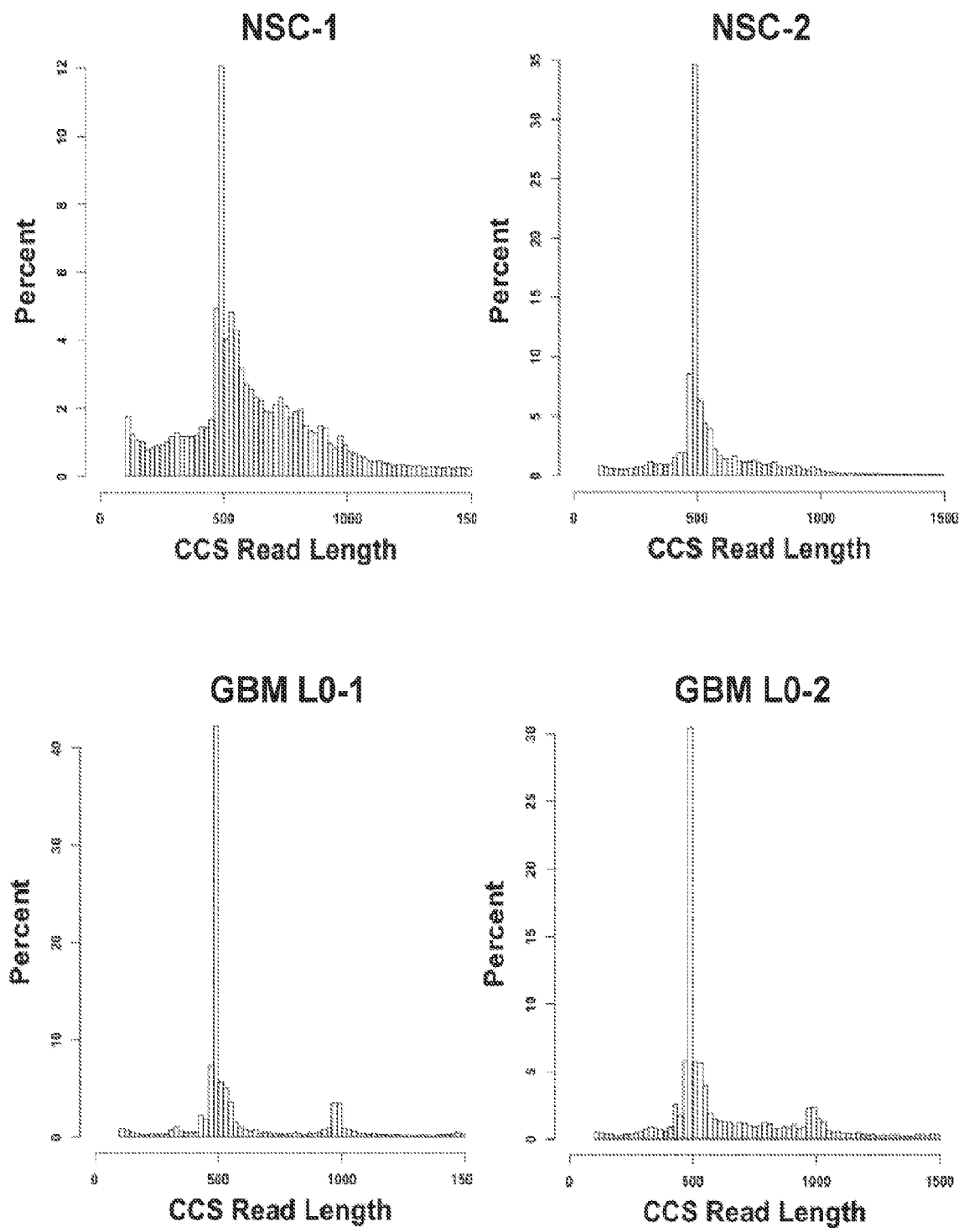


FIG. 27

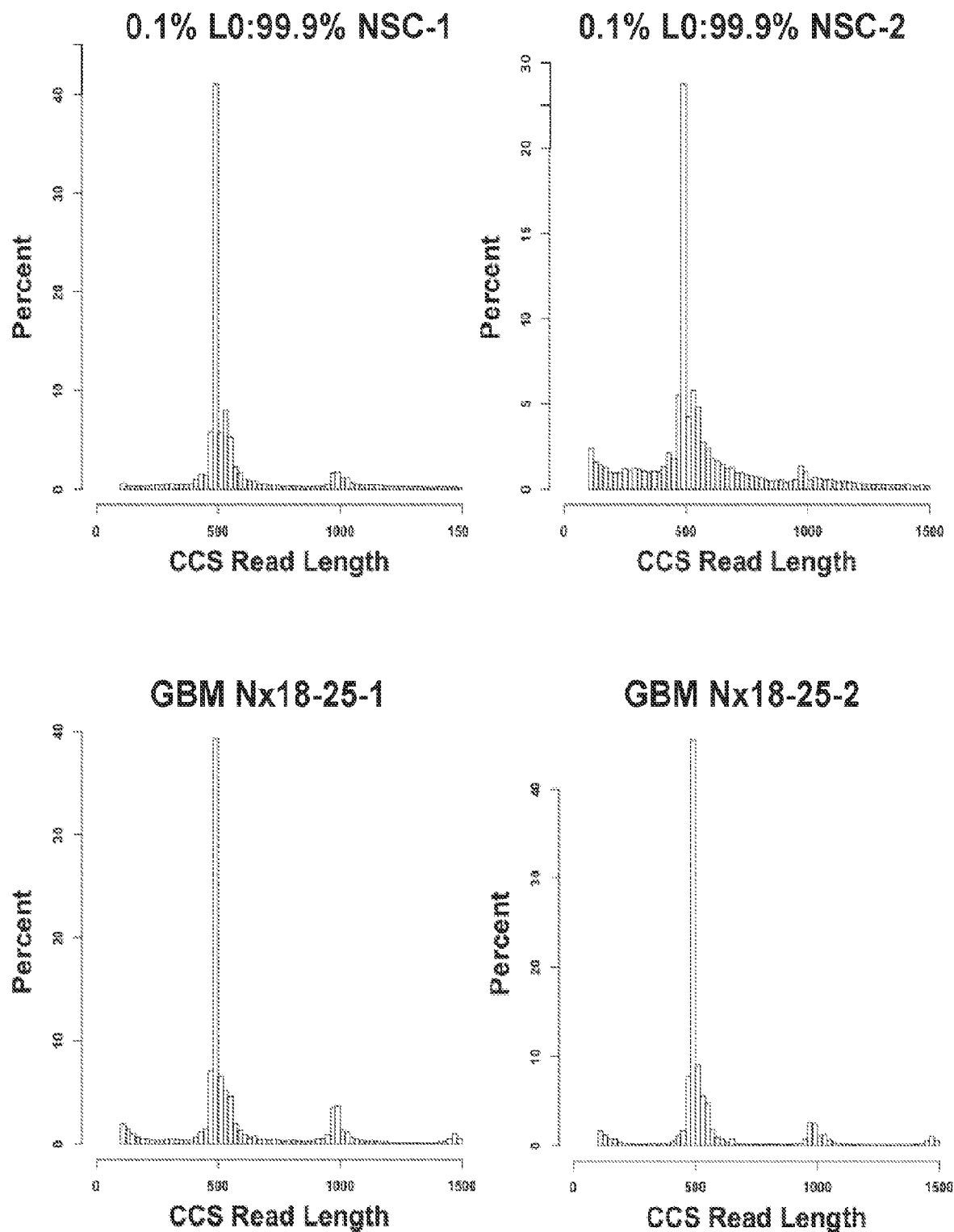


FIG. 28

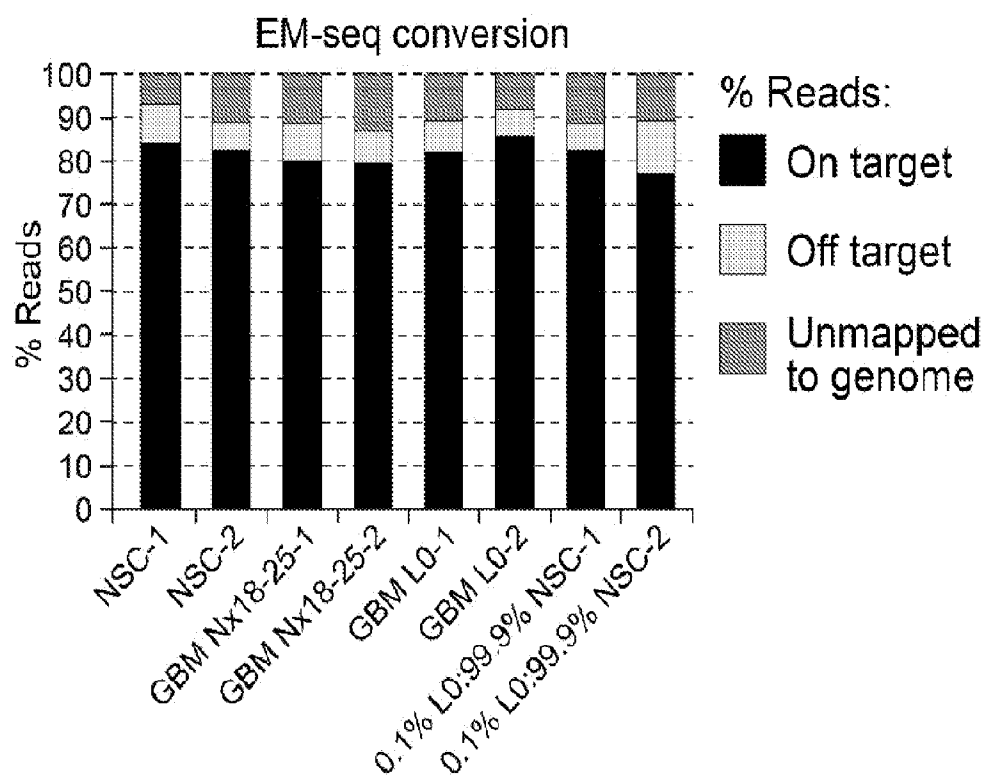


FIG. 29

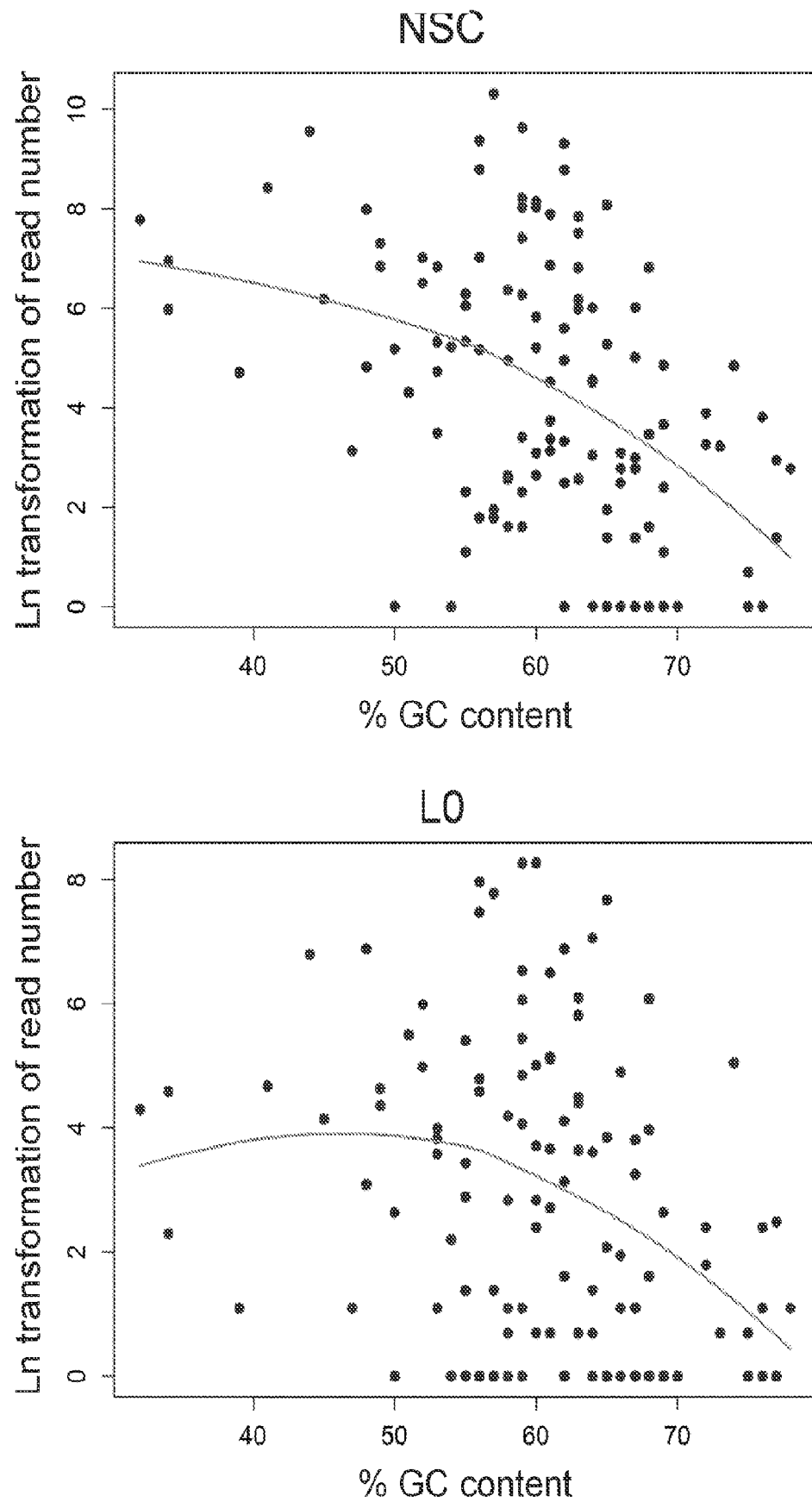


FIG. 30

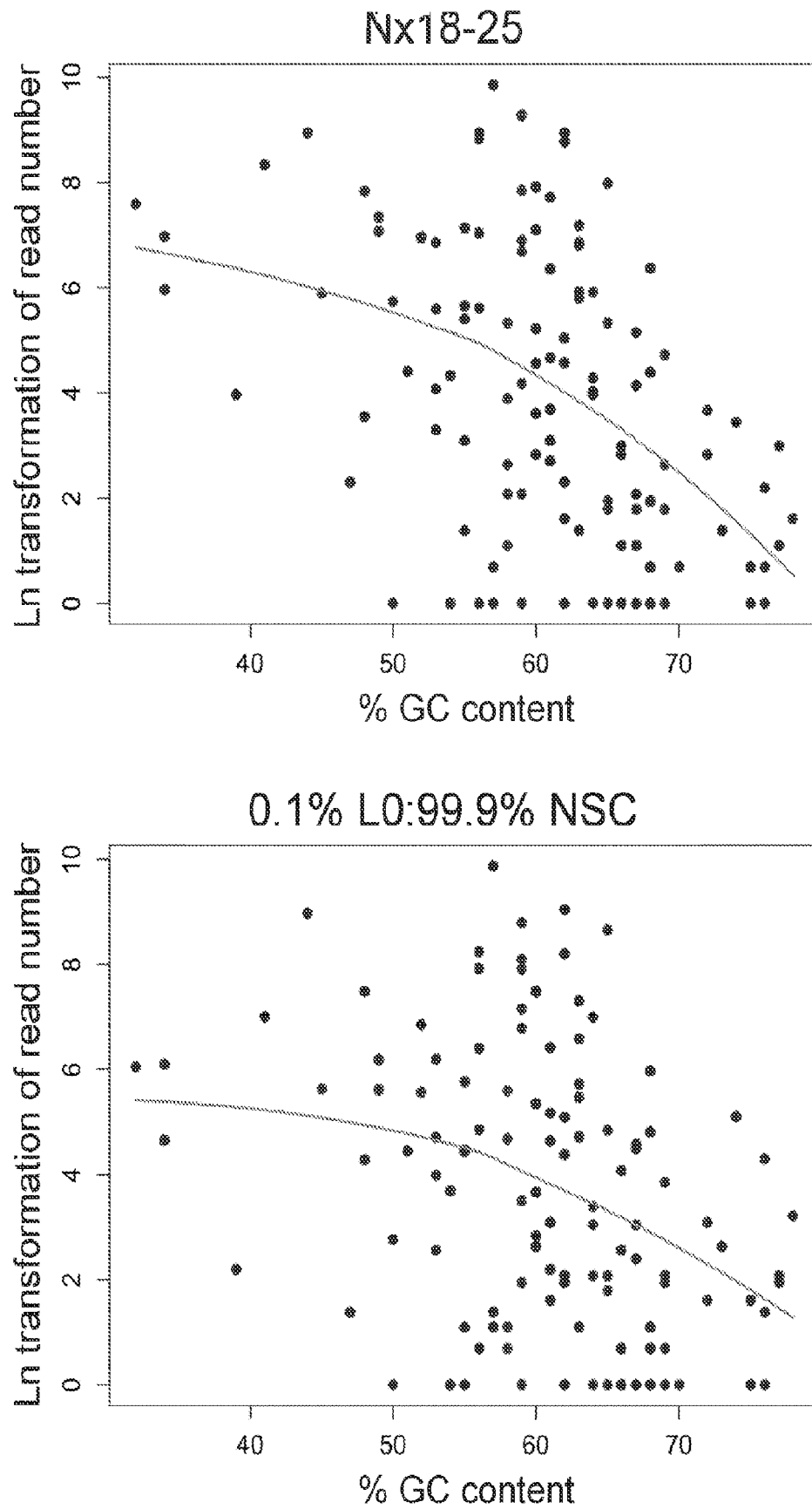


FIG. 31

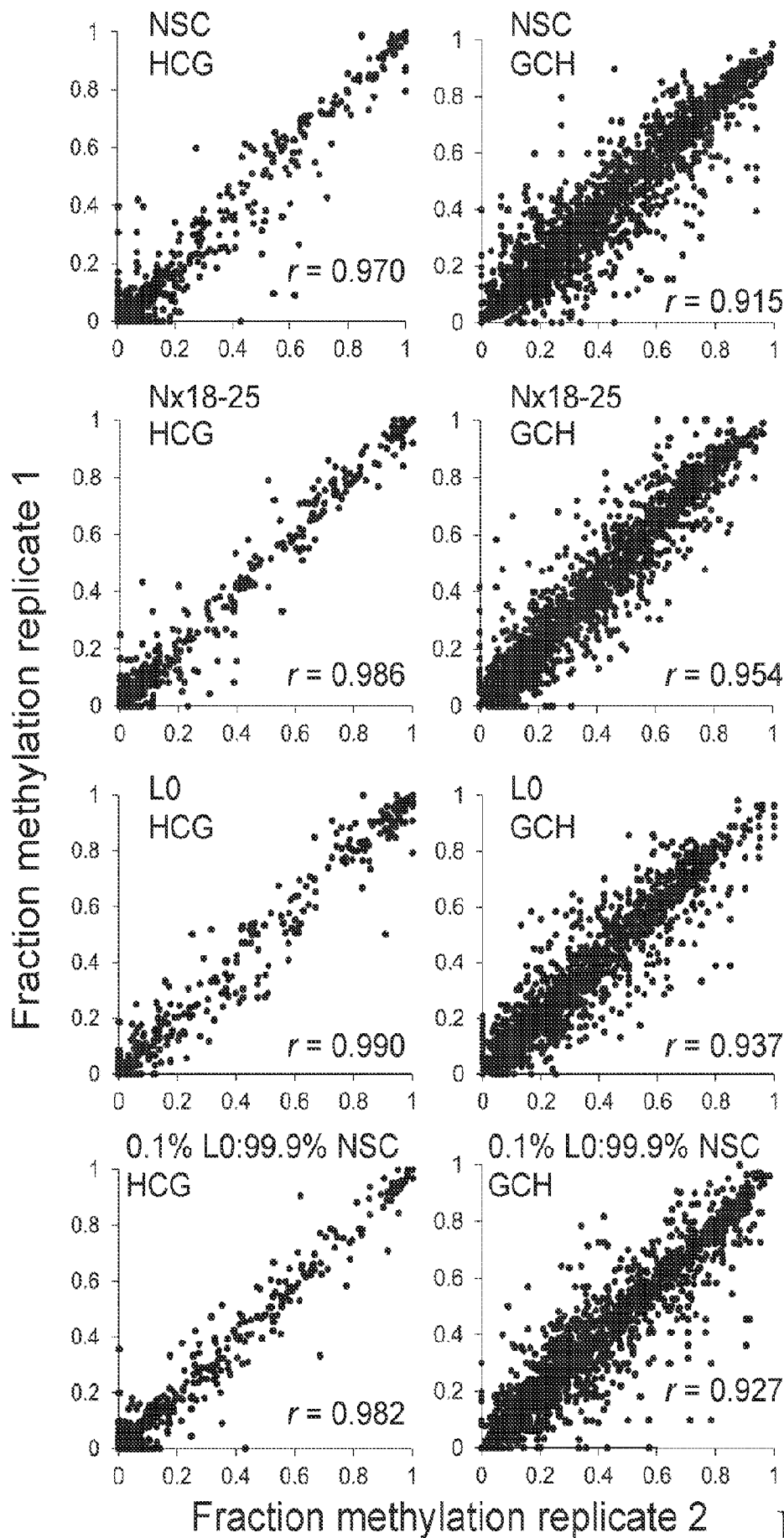


FIG. 32

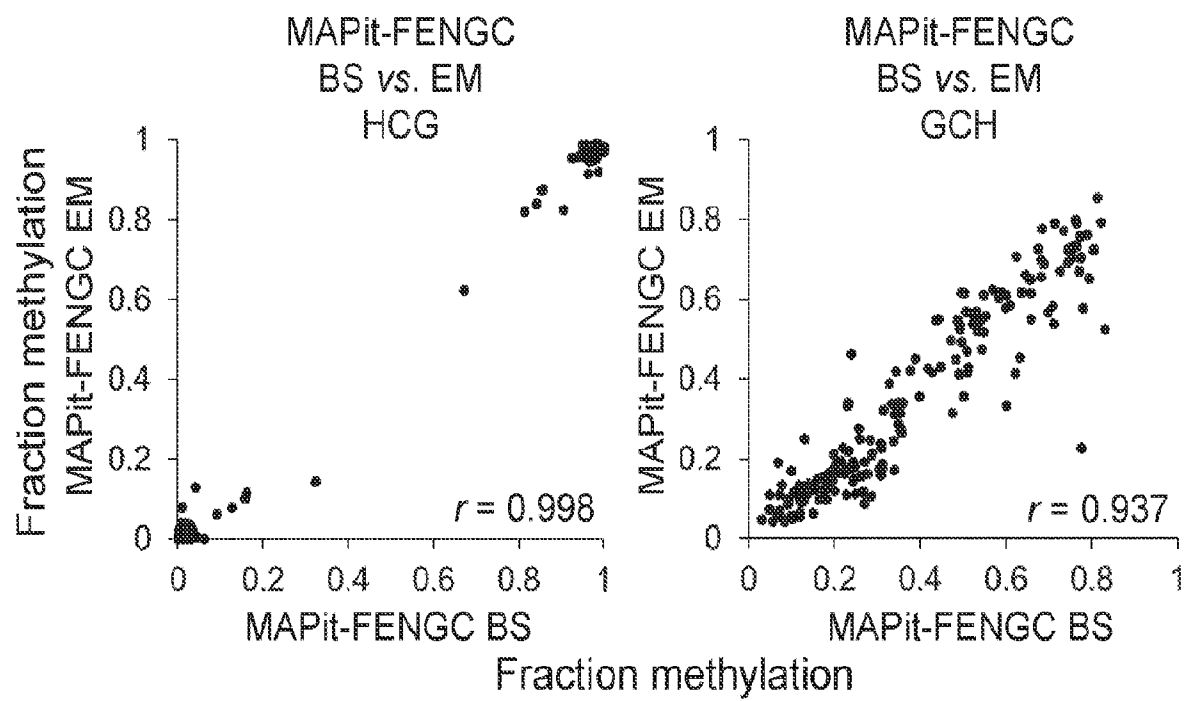


FIG. 33

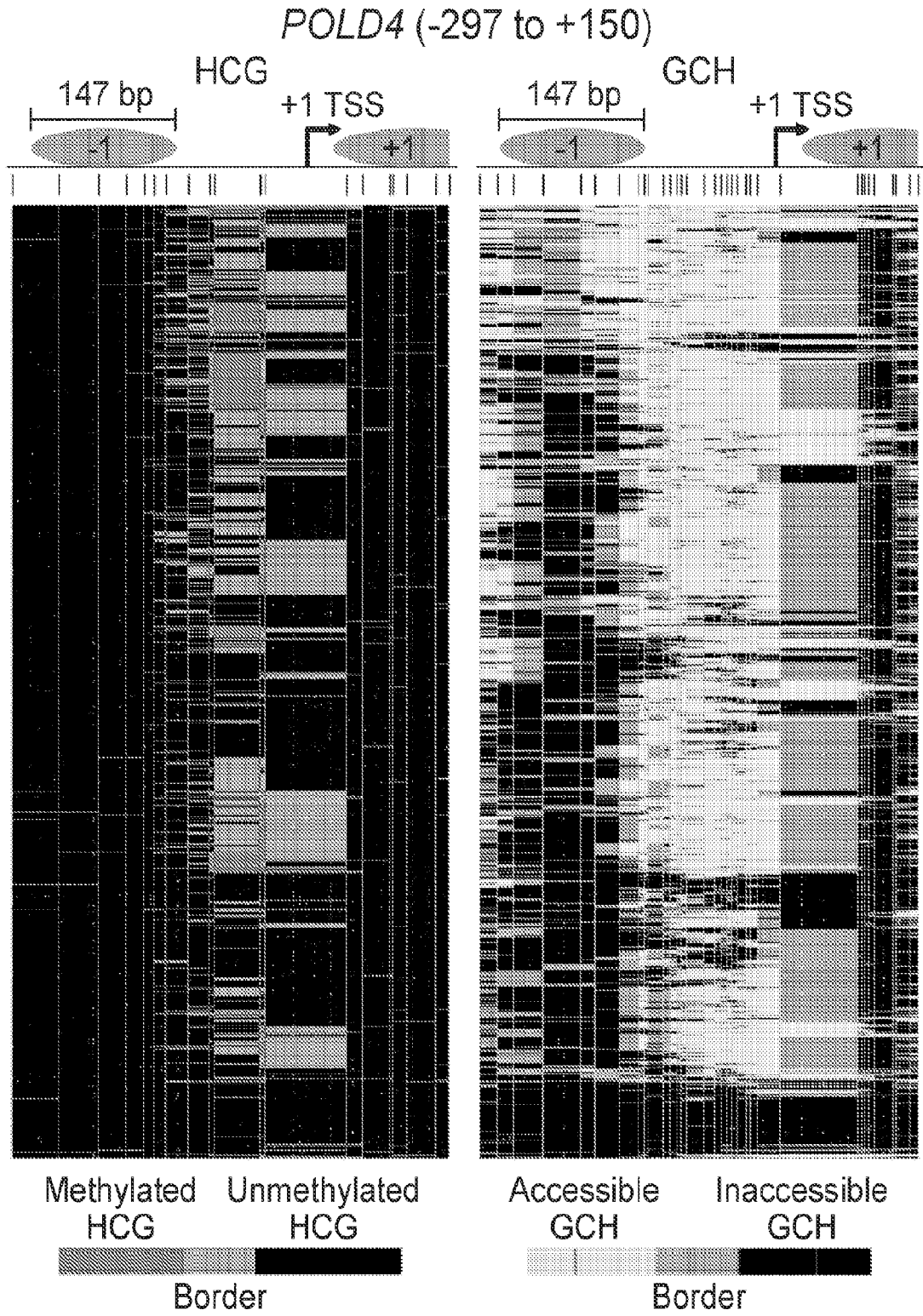


FIG. 34

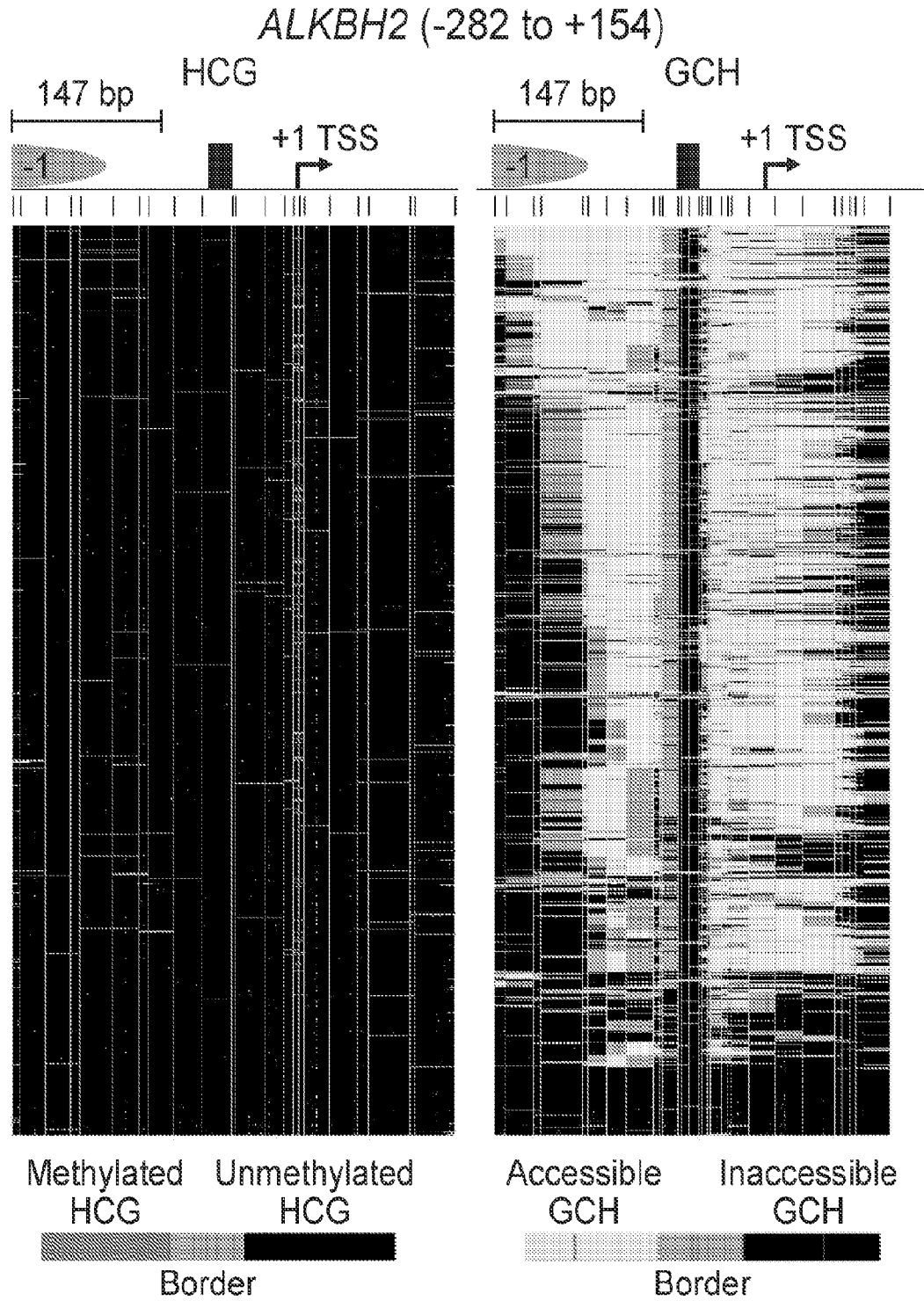


FIG. 35

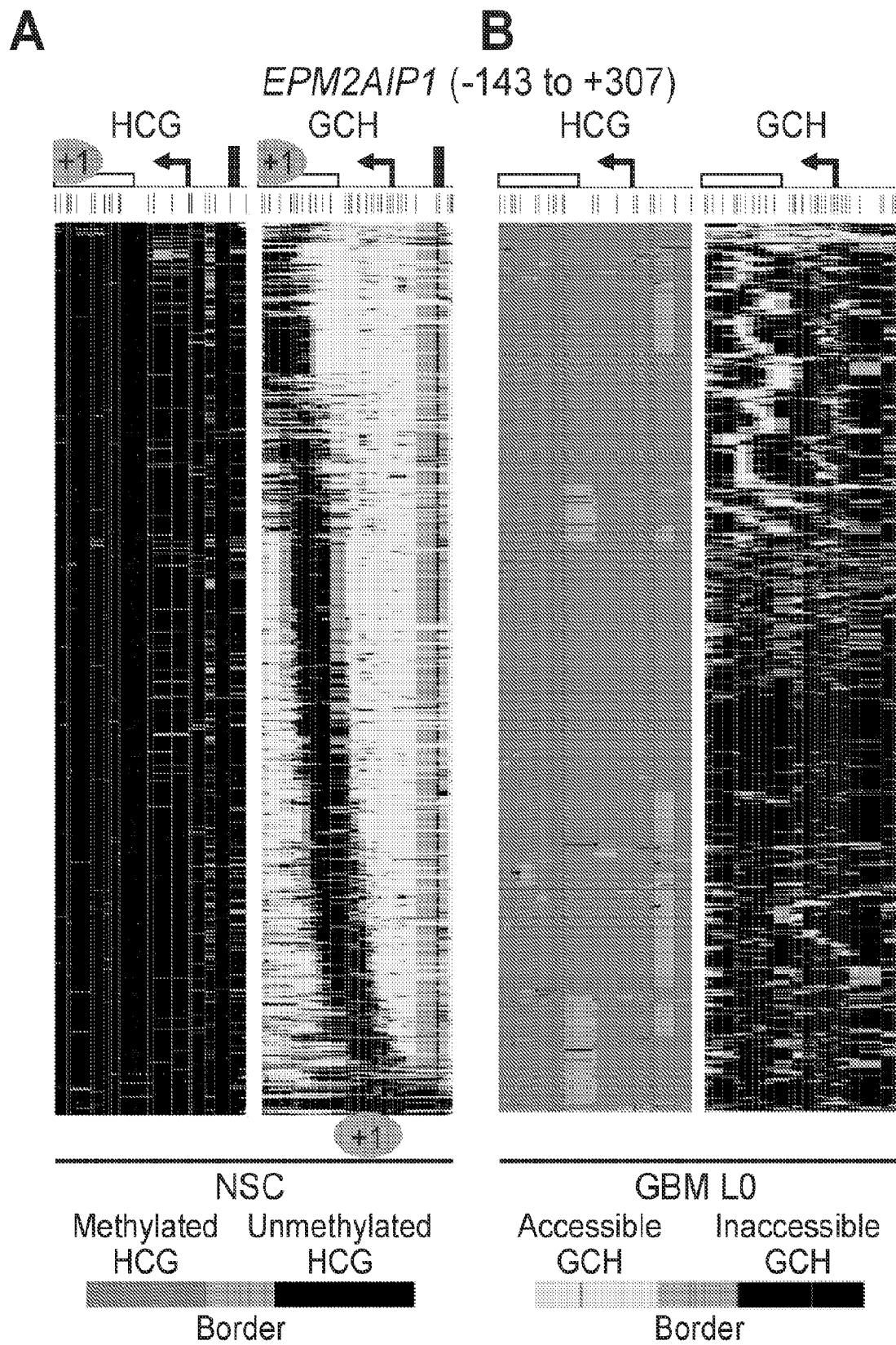


FIG. 36

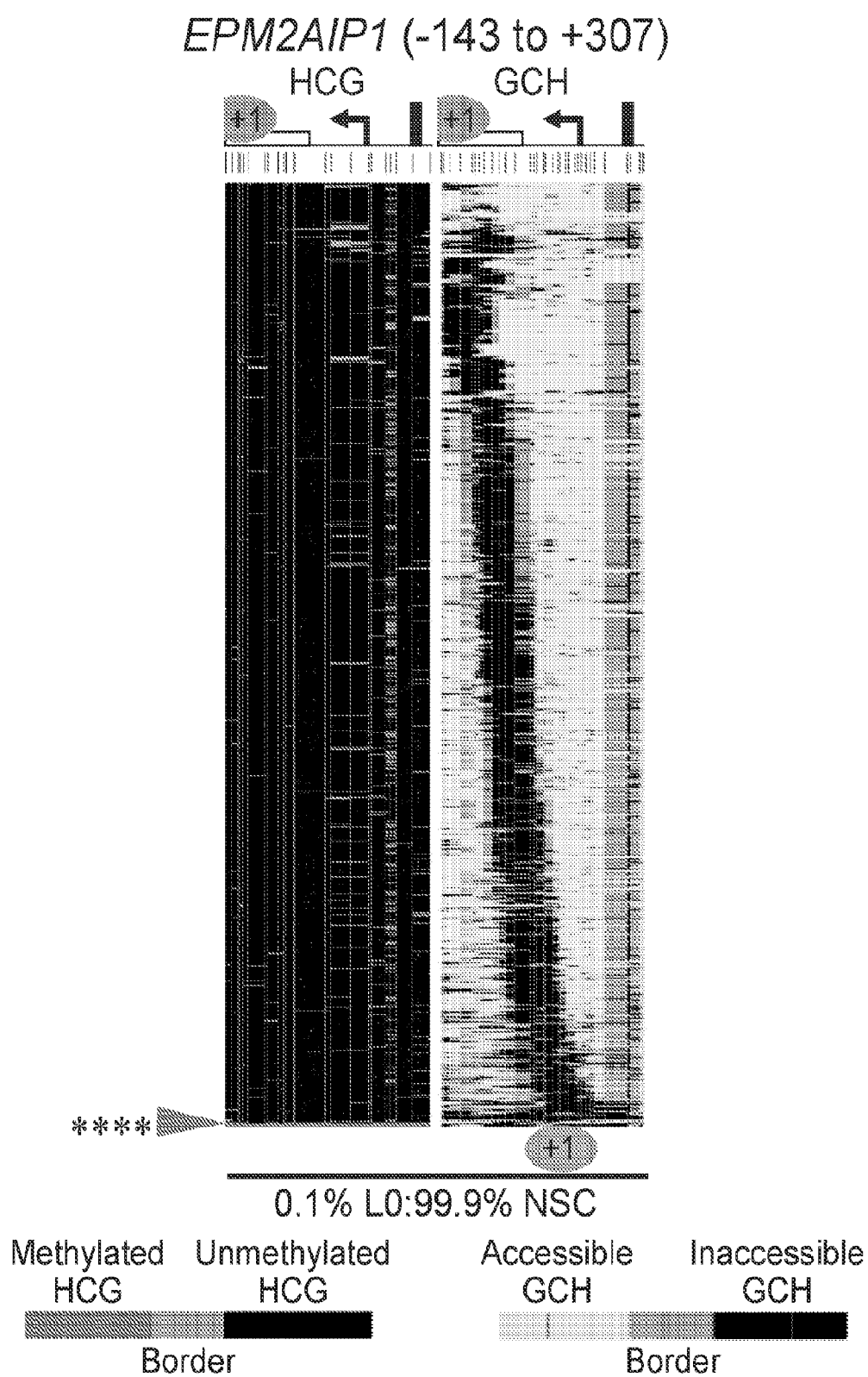


FIG. 37

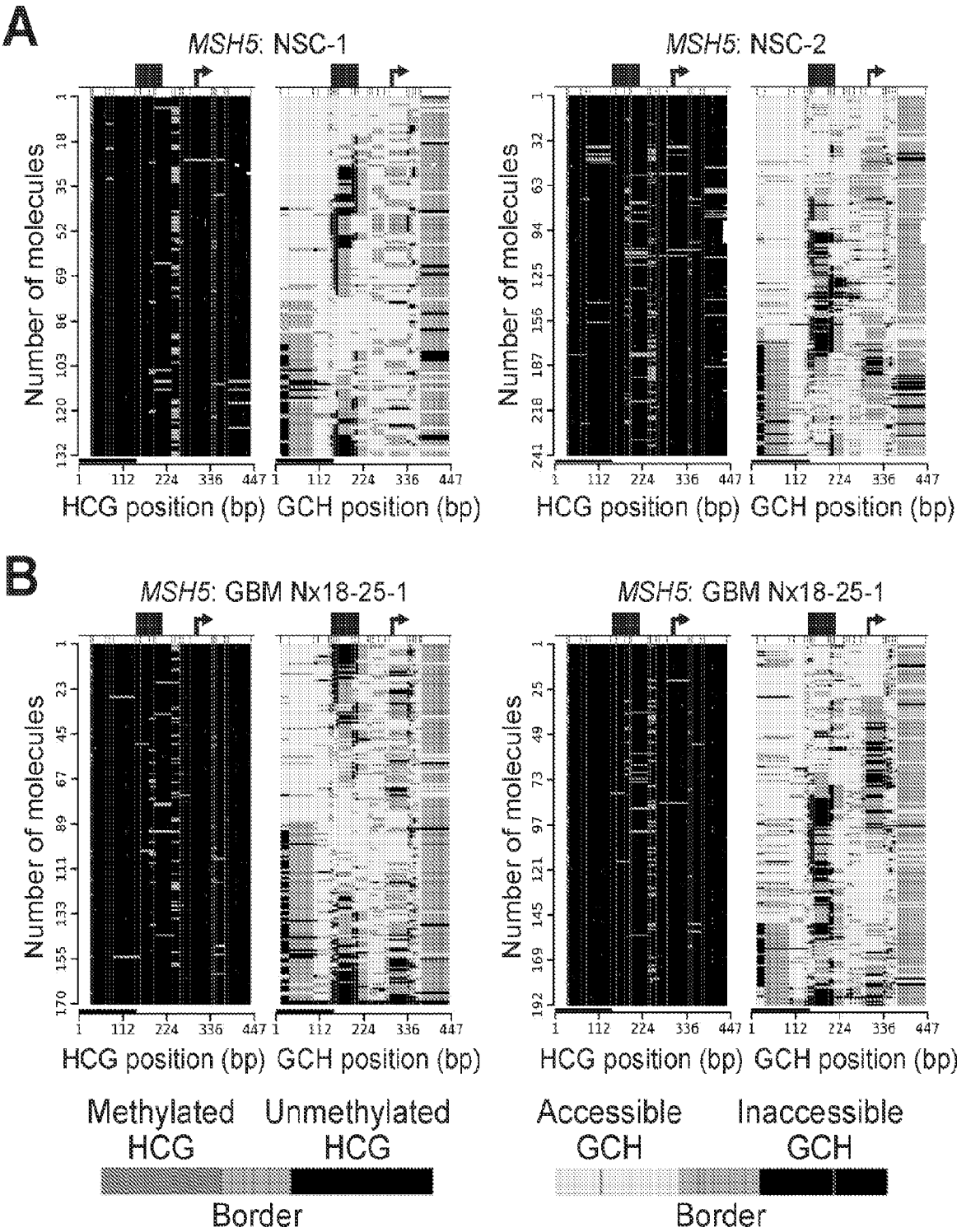


FIG. 38

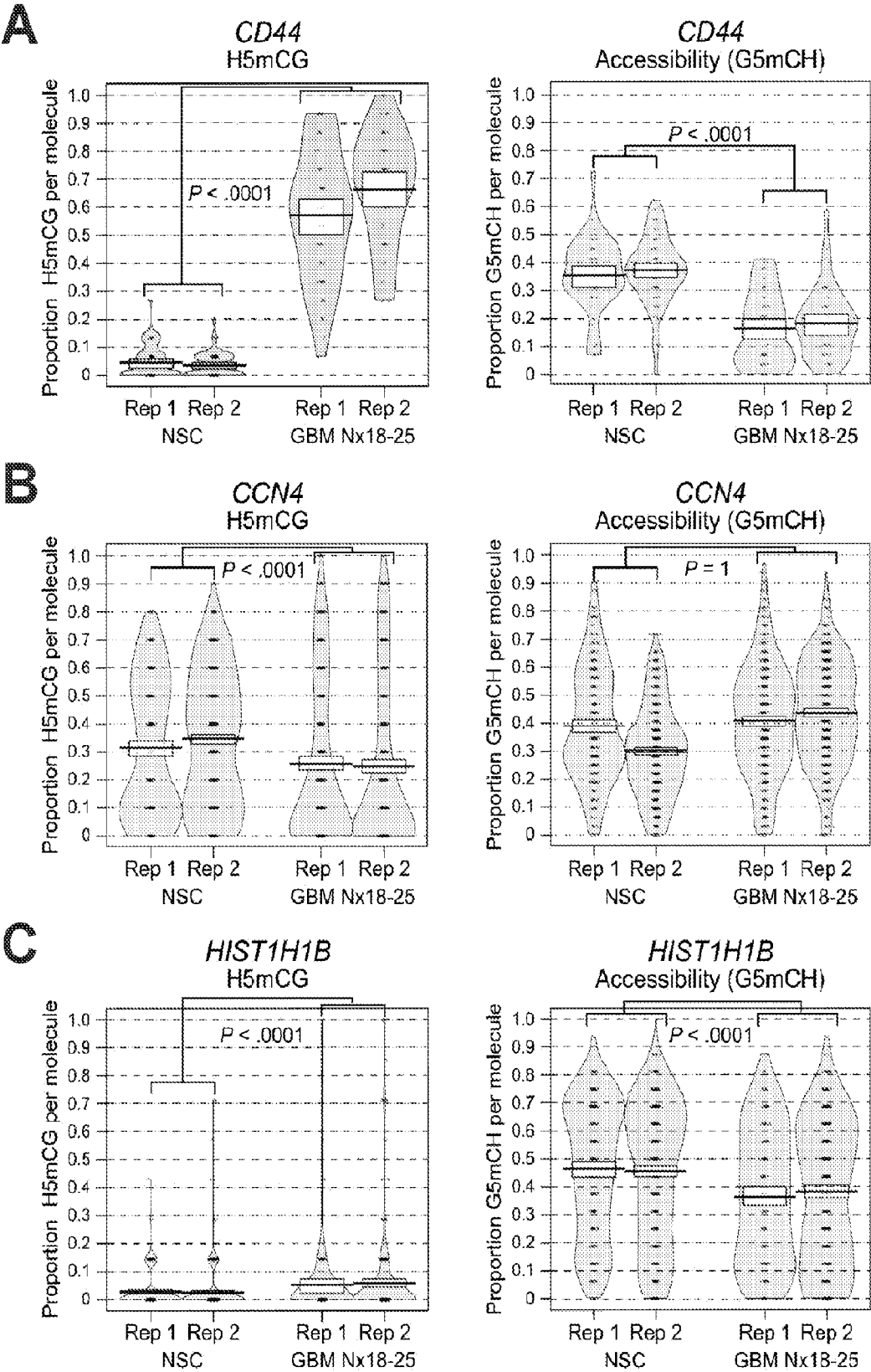


FIG. 39

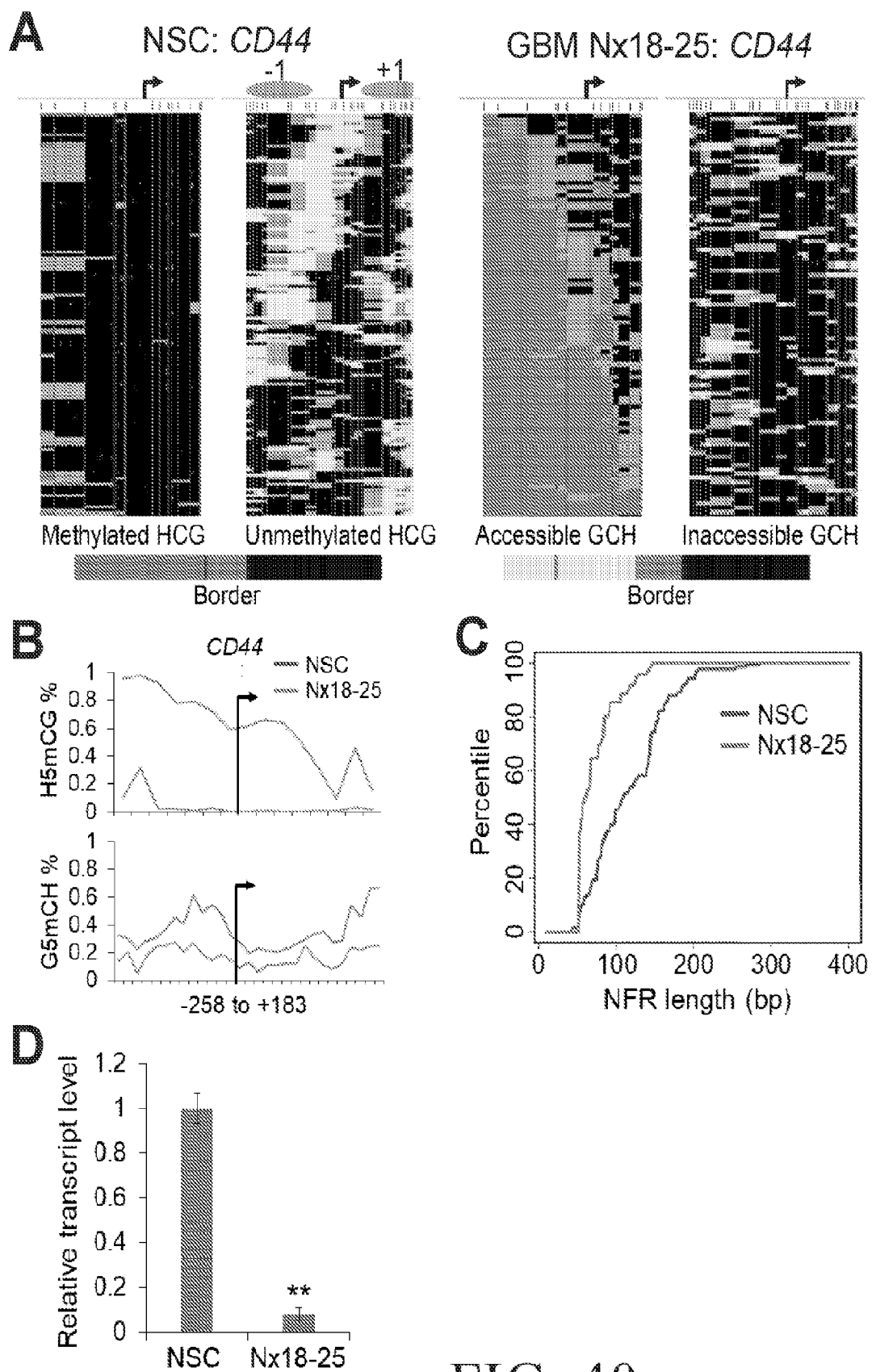


FIG. 40

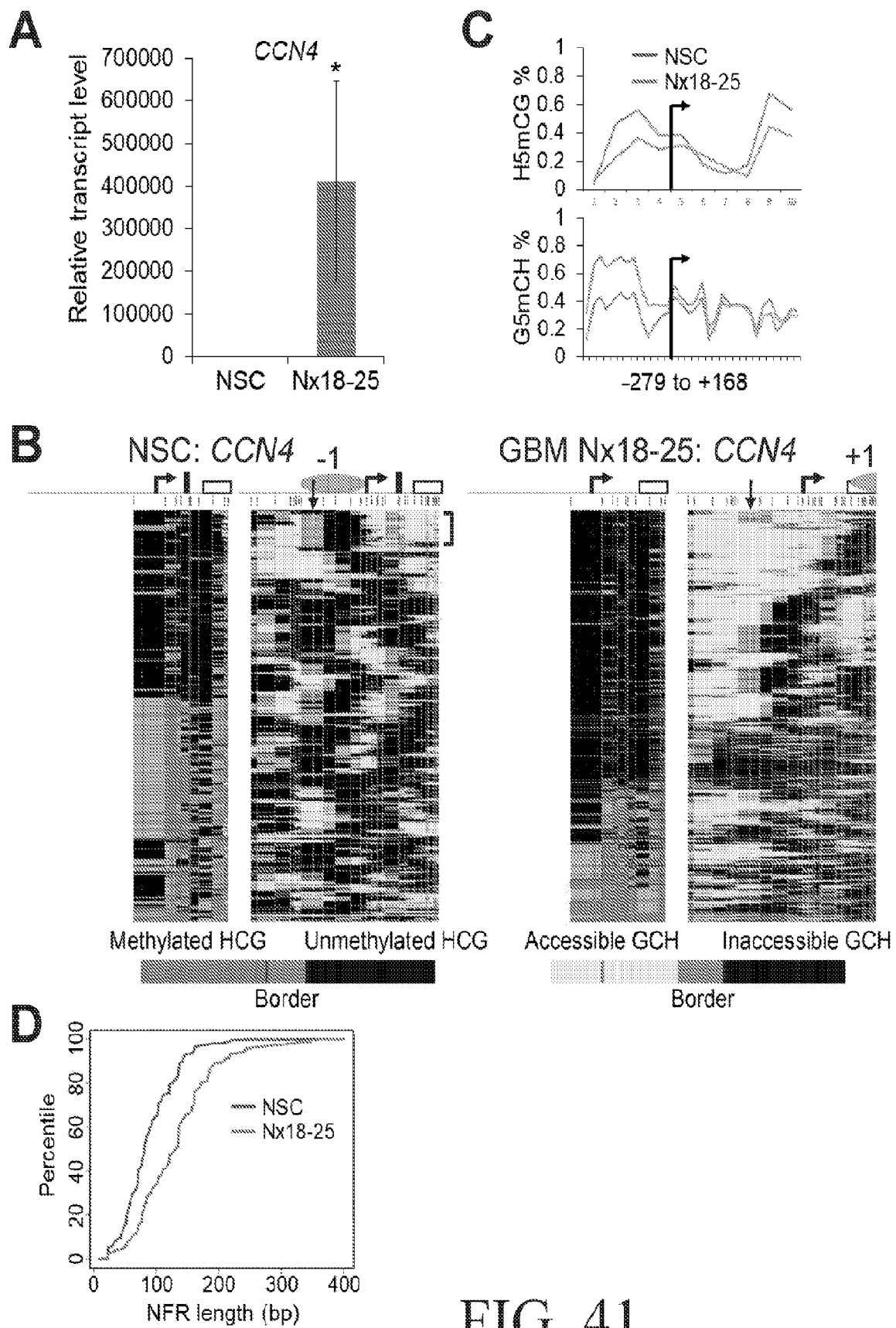


FIG. 41

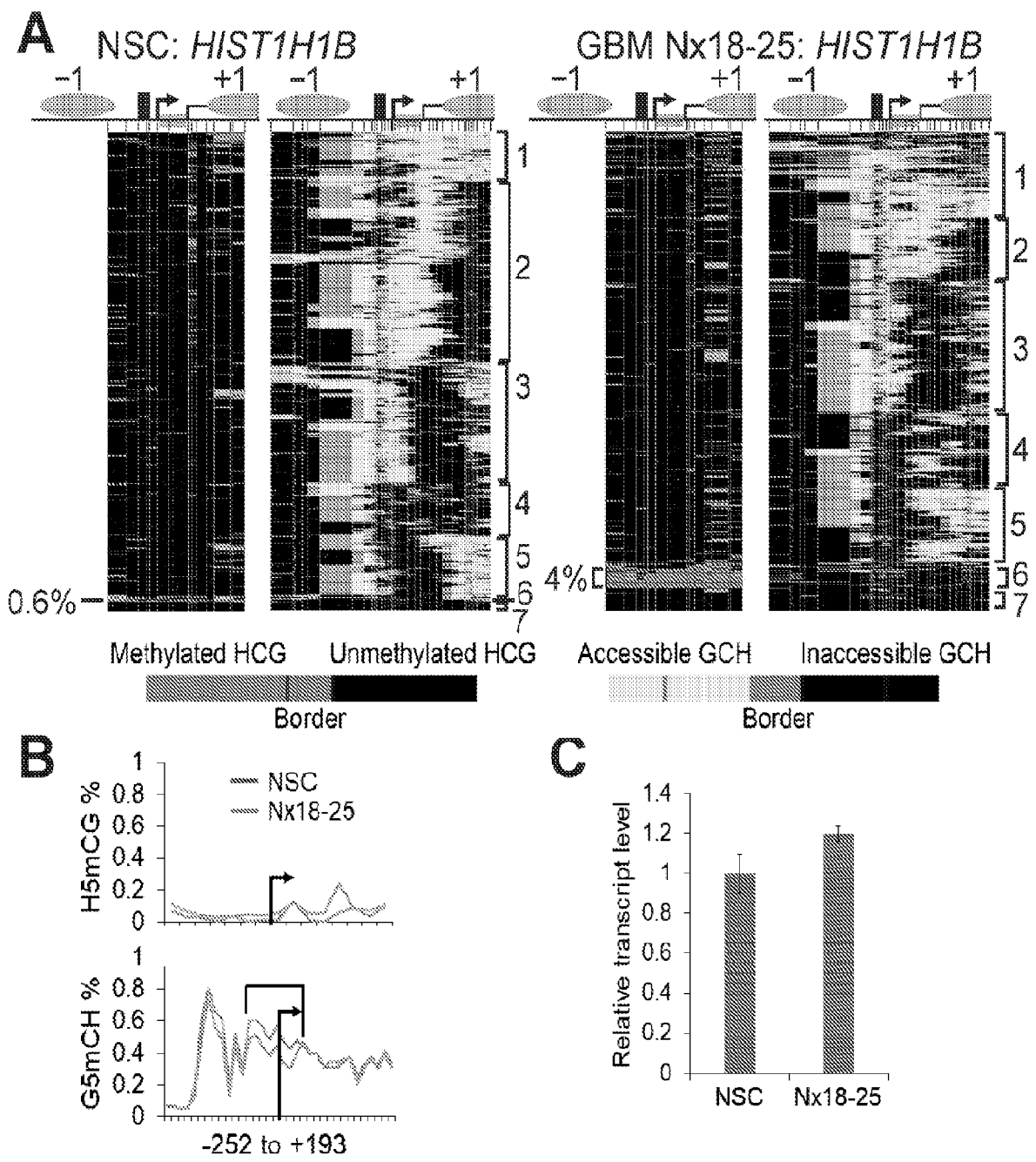


FIG. 42

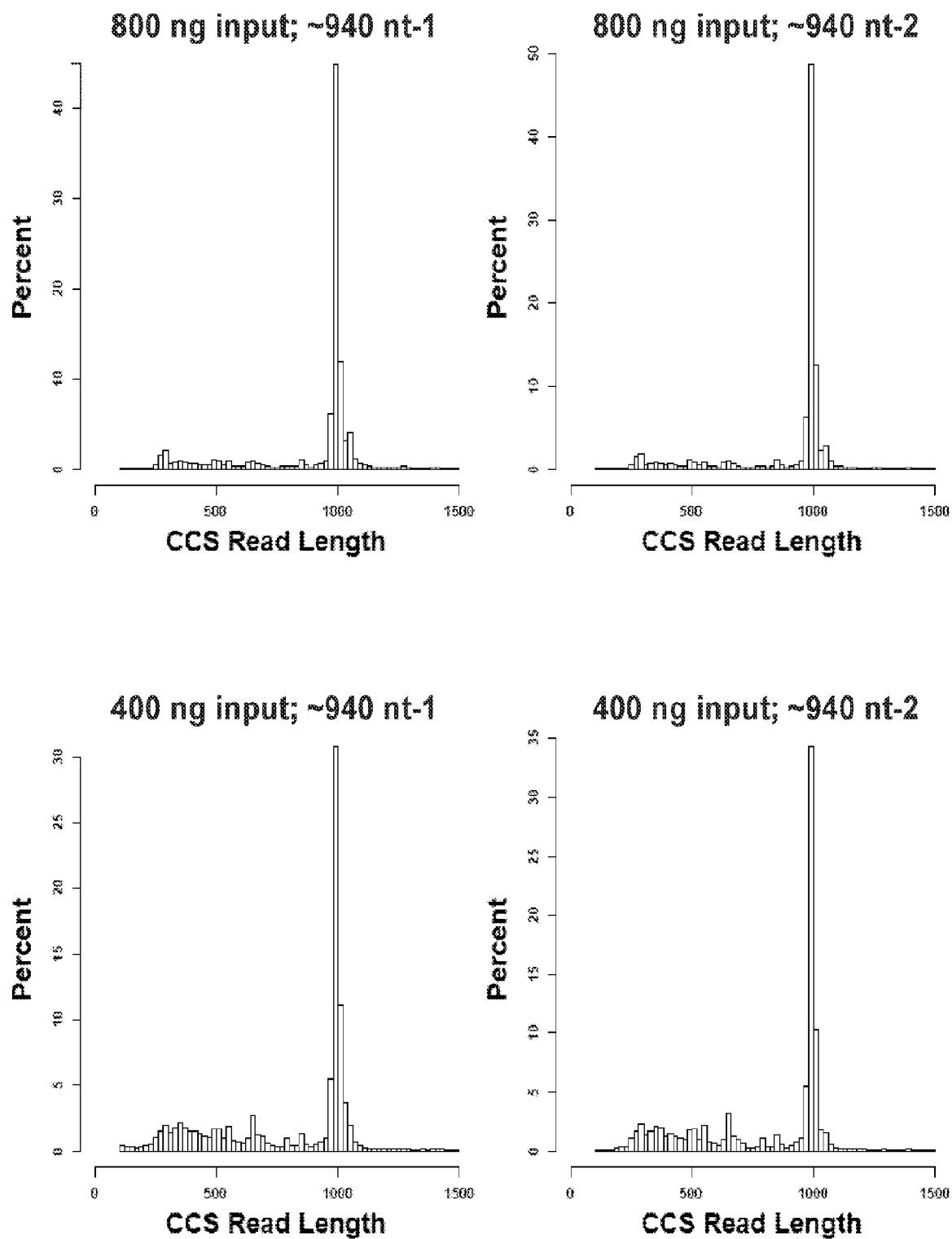


FIG. 43

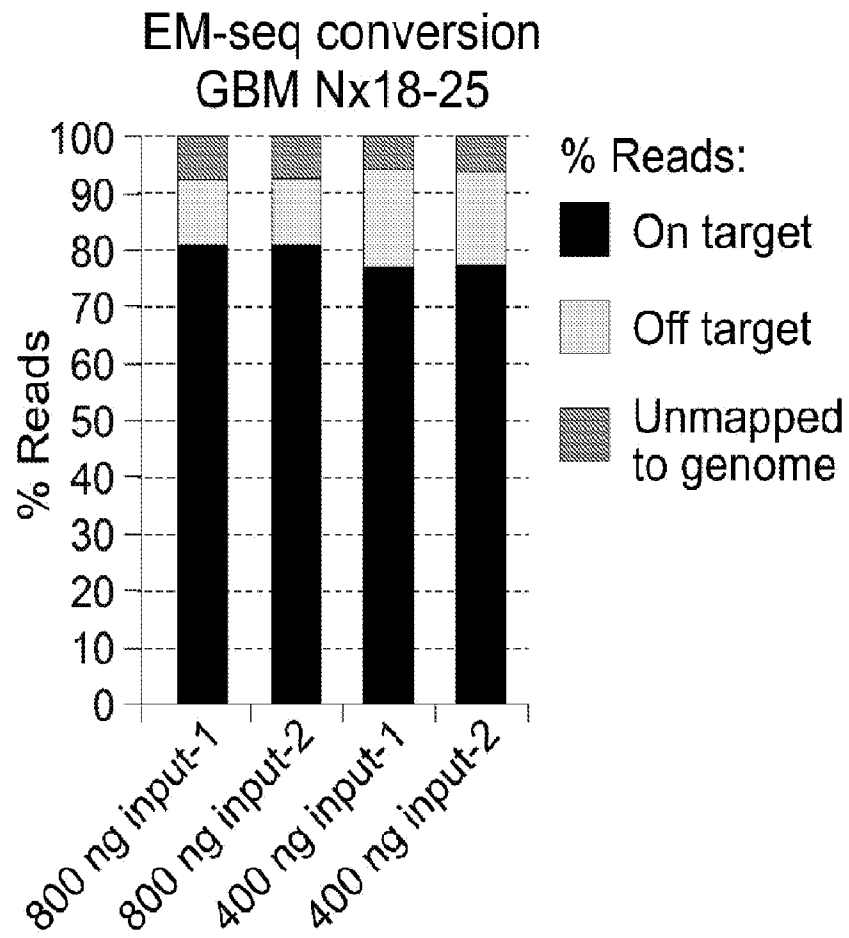


FIG. 44

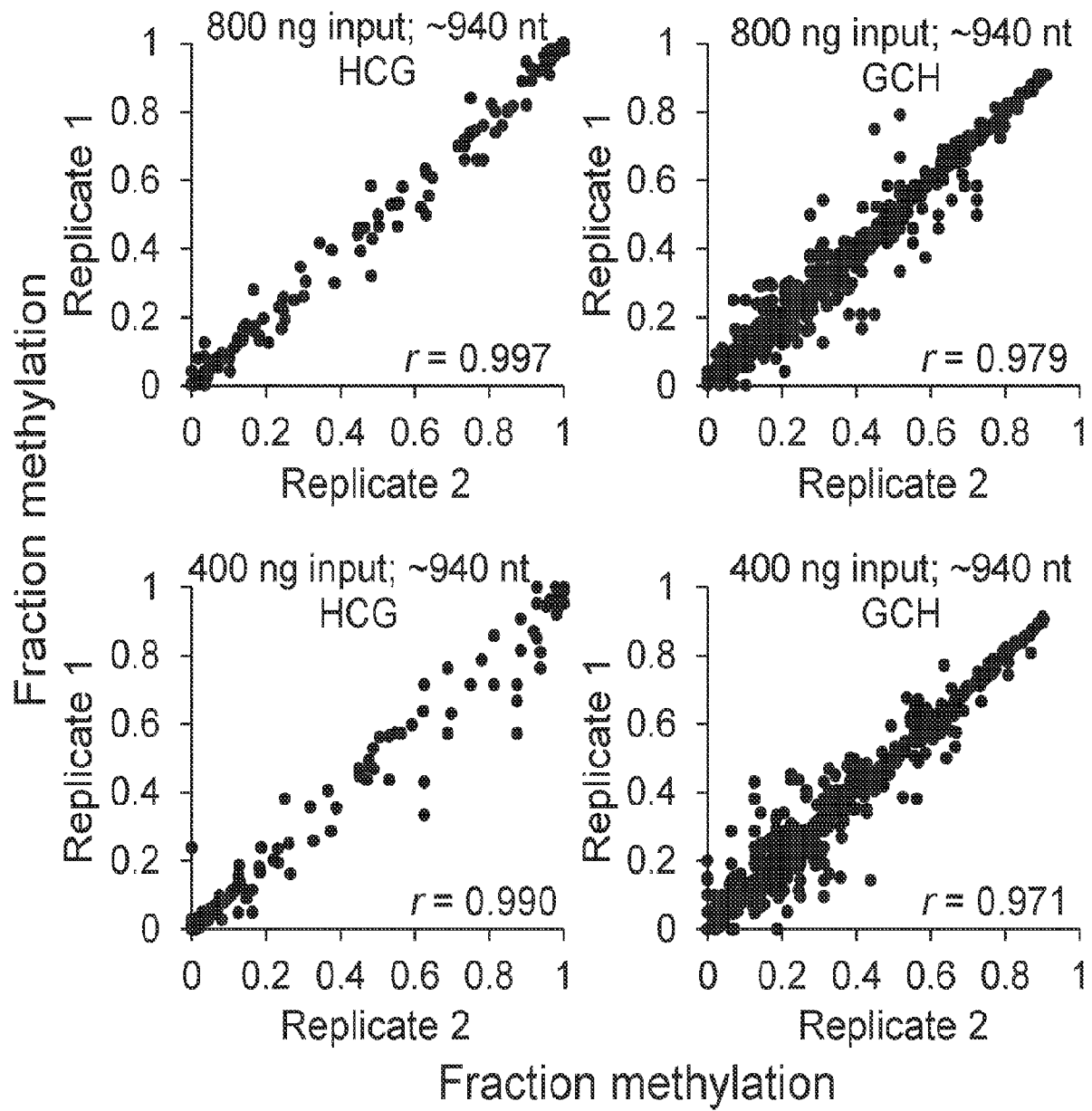


FIG. 45

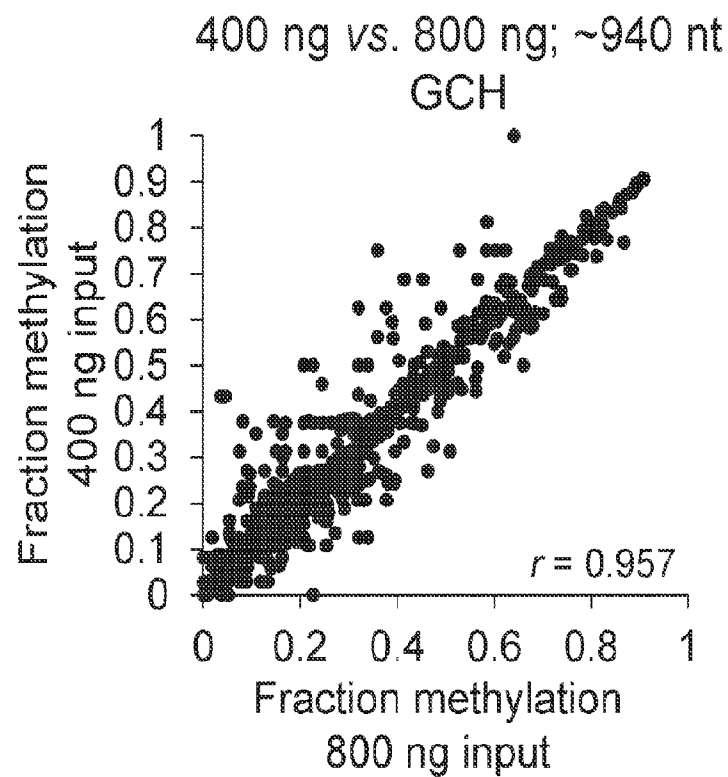
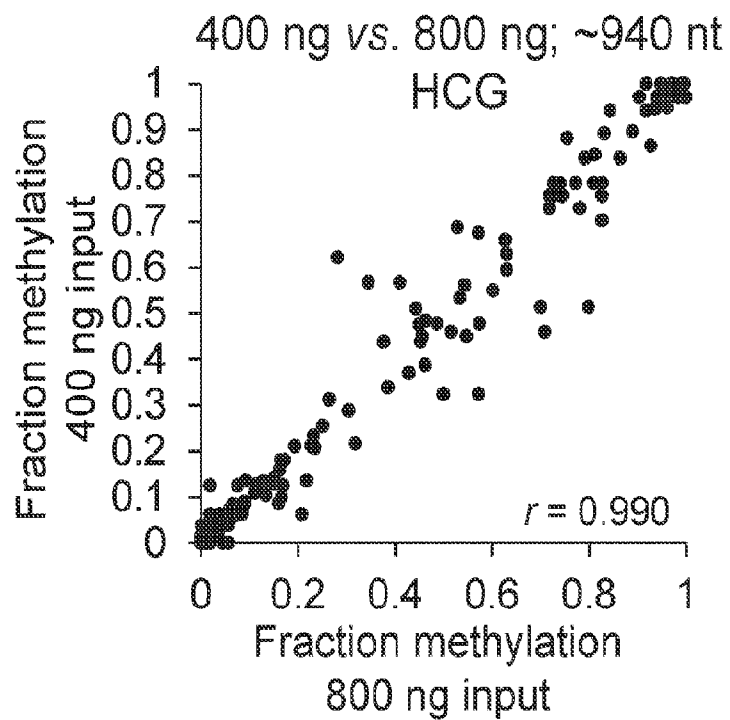


FIG. 46

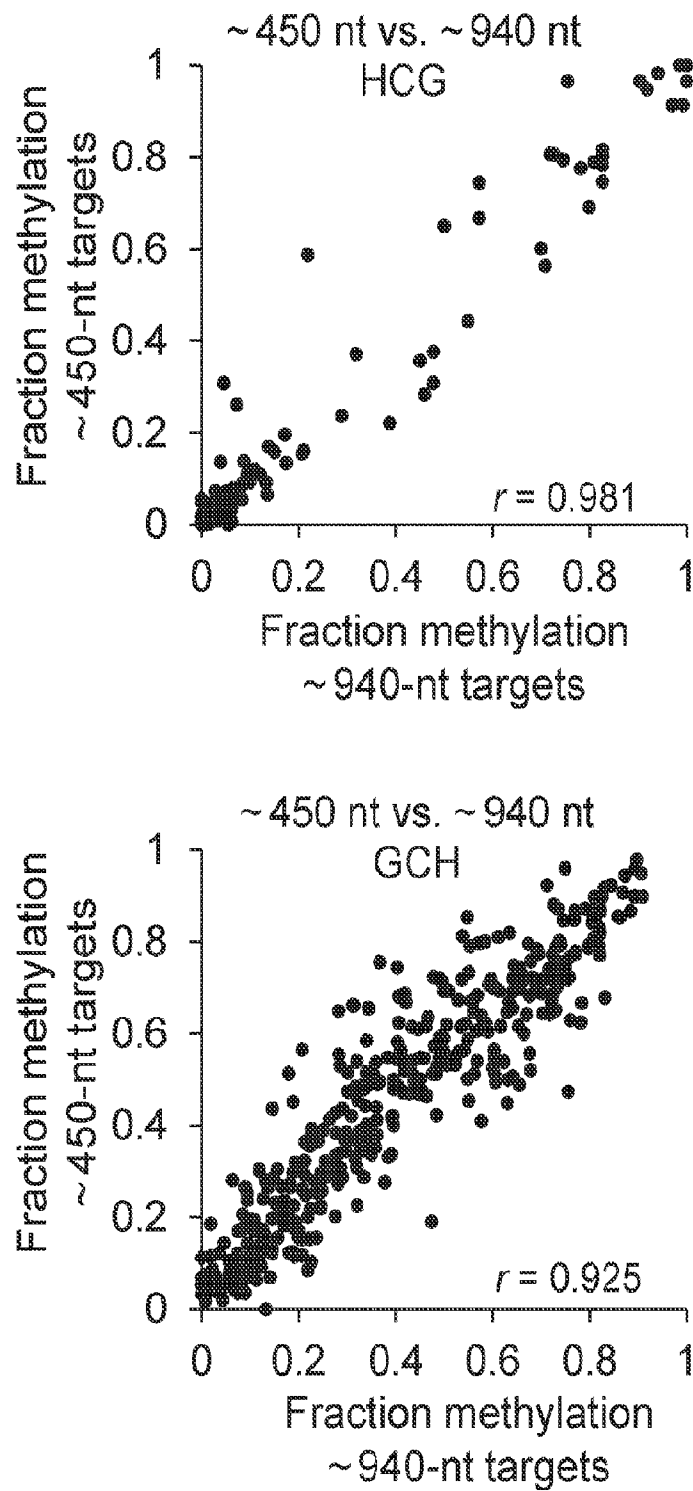


FIG. 47

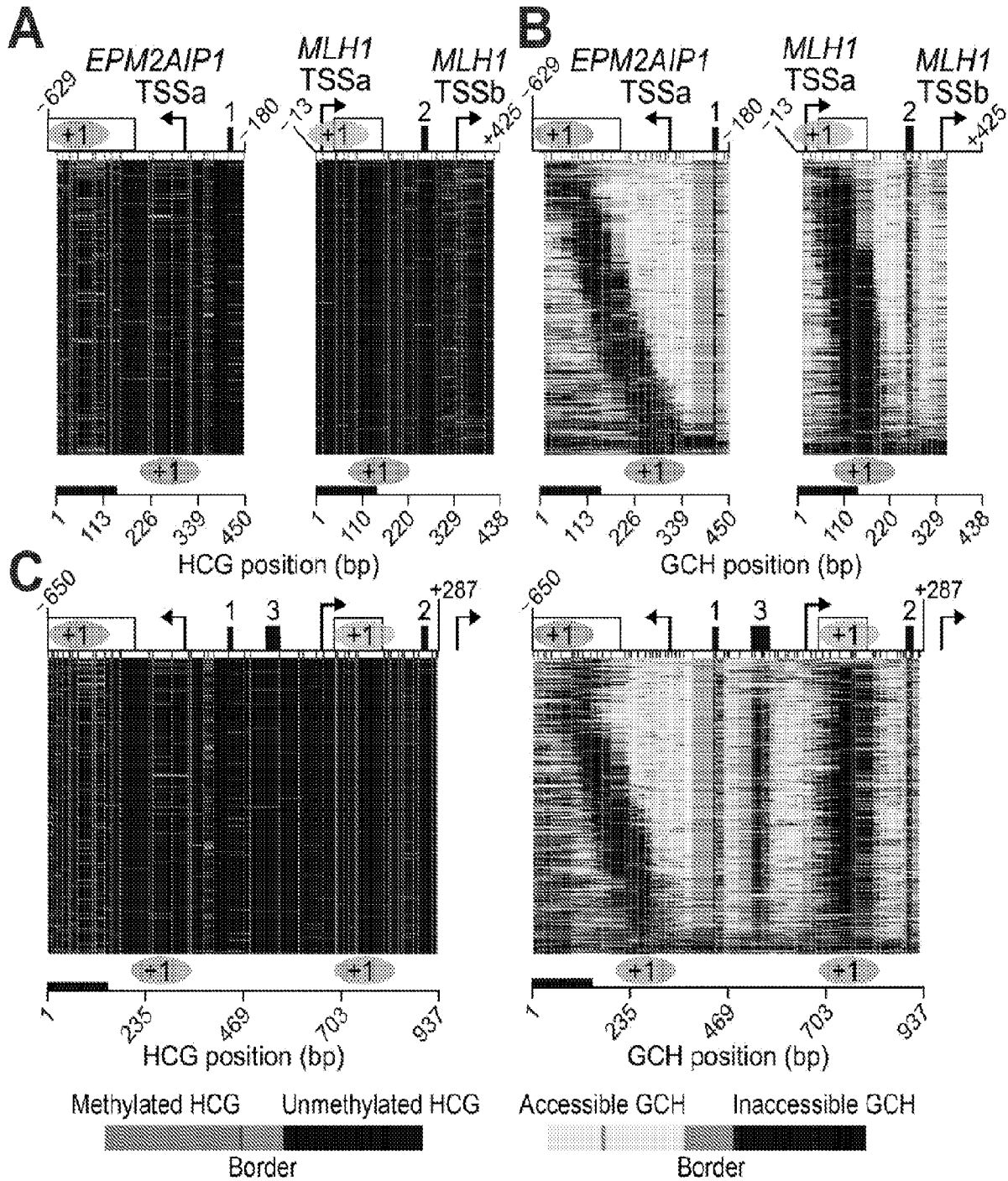


FIG. 48

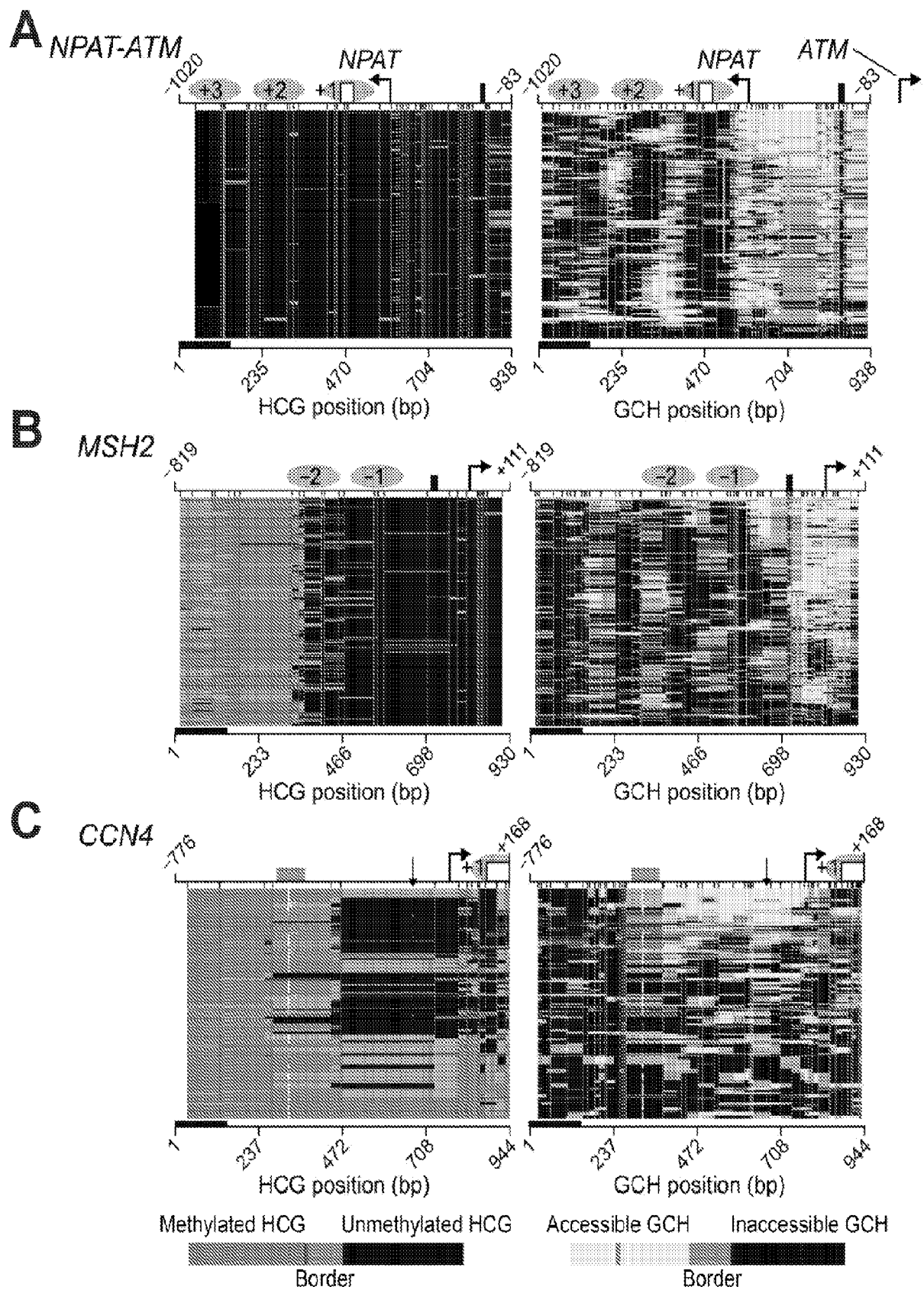


FIG. 49

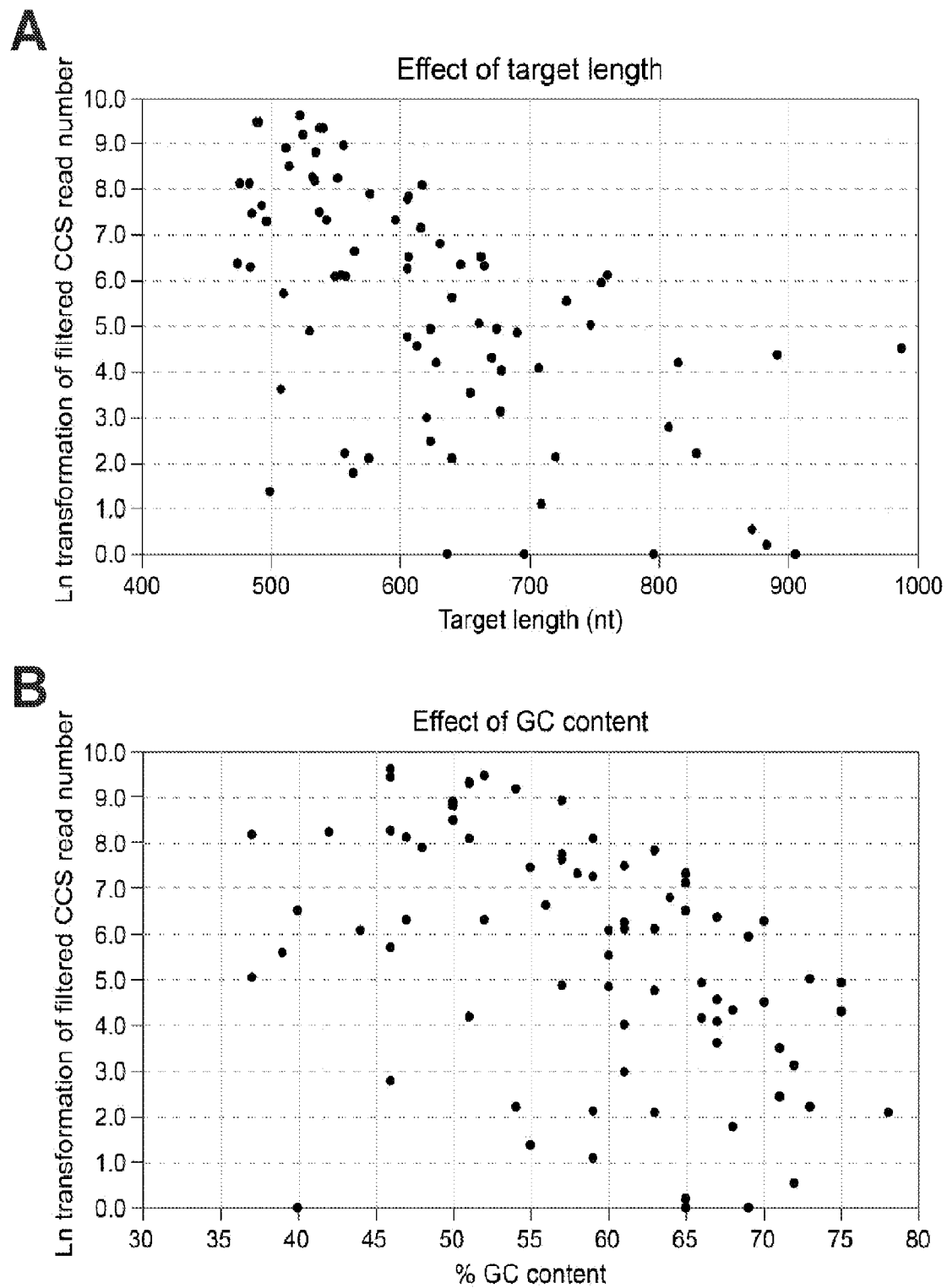


FIG. 50

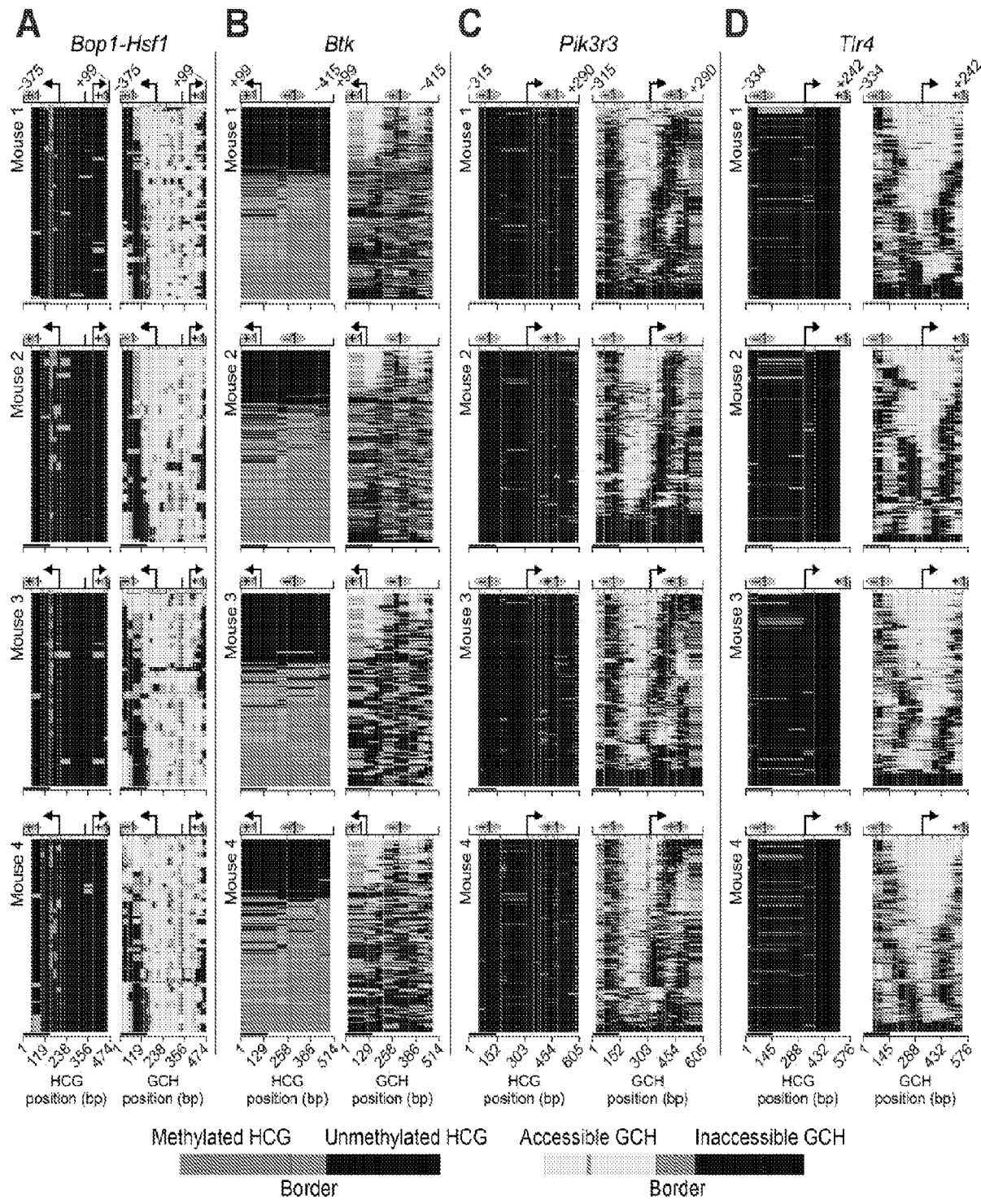


FIG. 51

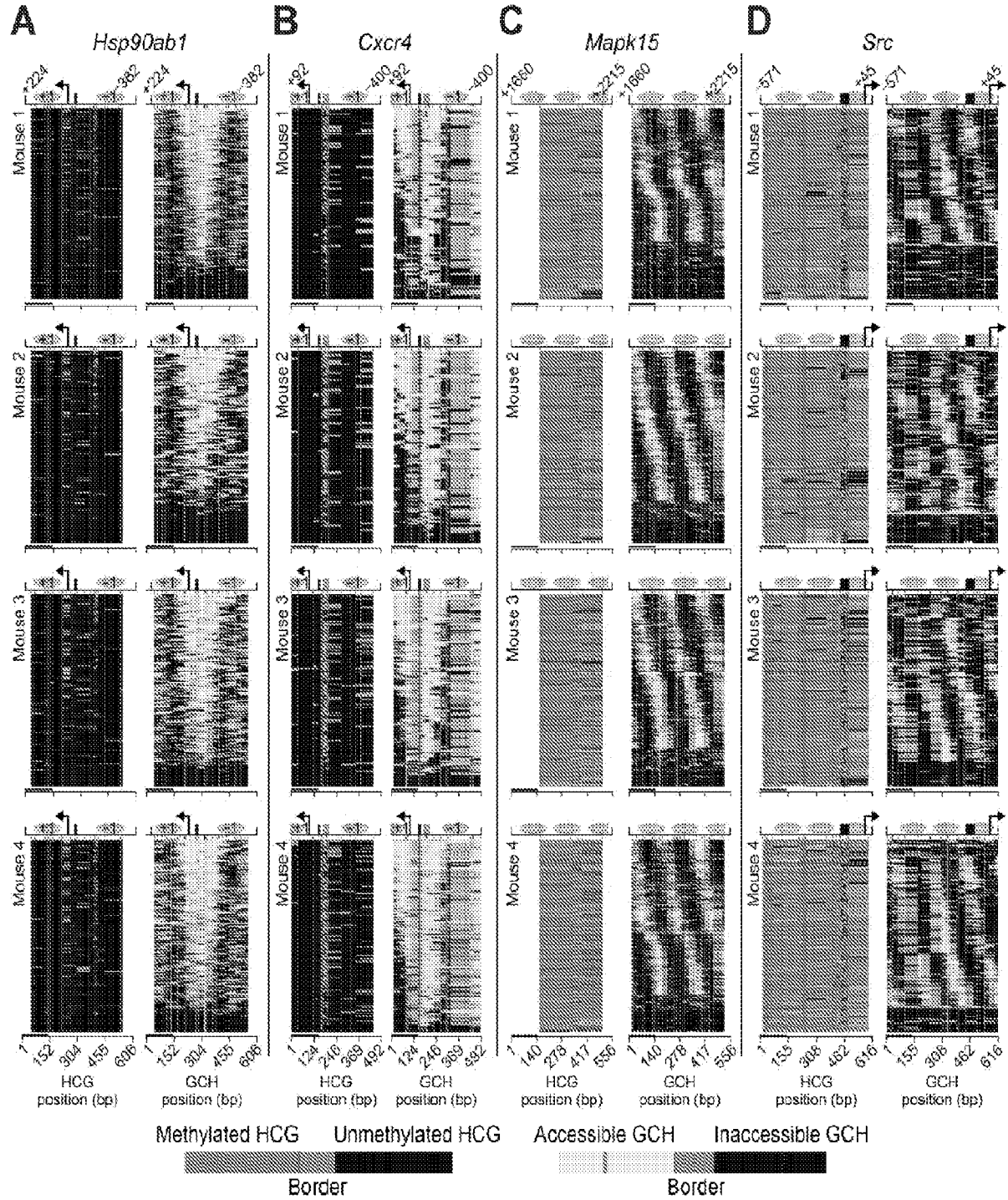


FIG. 52