Electronic Supplementary Material (ESI) for ChemComm.
This journal is © The Royal Society of Chemistry 2021

Supporting Information for

Small molecule-induced trinucleotide repeat contractions during in vitro DNA synthesis

Chikara Dohno,* Masaki Hagihara,* Nursakinah Binti Mohd Zaifuddin, Mizuki Nihei, Kaoru Saito, and Kazuhiko Nakatani*

DNA Polymerase Stop Assay

A reaction mixture of template DNA (0.3 μM; 5'-GAACCACAACTCAAATAATGTCTAGA GATTAGAAGACATAG(CNG)nCATGAGACCGGATCCATTCATGTCGTGACTGGGAAAA CCCTGGCG-3': N = A or T, n = 10 or 20, Table S1) and 5' FAM-labeled 3'-DNA primer (0.1 μM; 5'-FAM-CGCCAGGGTTTTCCCAGTCACGAC-3') was heated to 90 ºC for 3 min in reaction buffer, and cooled to ambient temperature for 30 min. The requisite amount of ligand was added to the reaction mixture and incubated for 30 min at room temperature prior to the addition of polymerase. DNA polymerase and dNTPs were then added to the mixture, and the reaction was performed at 37 ºC (*Bsu* DNA polymerase, Klenow DNA polymerase), or 50 ºC (*96-7* DNA polymerase) for 30 min. The reaction products were diluted with HiDi Formamide (Thermo Fischer Scientific) and GeneScan 500 LIZ dye Size Standard (Thermo Fischer Scientific), and analyzed on an ABI 3500 Genetic Analyzer. The capillary electrophoresis chromatograms presented in the manuscript are the typical results among at least three experiments.

Sequencing of Shortened Products

The reaction products of *Bsu* DNA polymerase in the presence of 30 μM of NA were ethanol precipitated and then treated with T4 polynucleotide kinase (Thermo Fisher Scientific) at 20 °C for 15 min, followed by lambda exonuclease (New England BioLabs) at 37 °C for 30 min to digest 5'-phosphorylated template DNA containing $d(CAG)_{20}$. The products were PCR amplified using a primer set (Primer1: CGCCAGGGTTTTCCCAGTCACGAC and Primer2: GAACCACAACTCAAATAATGTCTAGA, Table S1), and cloned with the NEBuilder HiFi DNA Assembly Master Mix (New England BioLabs). Plasmids were extracted from each of the colonies using the FavorPrep™ Plasmid DNA Extraction Mini Kit (Favorgen) and applied to Sanger sequencing (Eurofins Genomics).

IonPGM sequencing.

The polymerase extended products were ethanol precipitated and then treated with T4 polynucleotide kinase at 20 °C for 15 min, followed by incubation with lambda exonuclease at 37 °C for 30 min. The resulting mixture of nascent strands were subjected to PCR amplification using primer set of Primer3 and Primer4 (Table S1) to add adaptor sequences necessary for IonPGM sequencing. The PCR products were purified by NucleoSpin® Gel and PCR Clean-up (MACHEREY-NAGEL) and the concentrations were determined by Qubit dsDNA HS assay kit and Qubit fluorometer (Invitrogen). Ion PGM sequencing was carried out according to manufacturer's protocol (318 v2 chip, 500 flows). The sequence reads were filtered and trimmed by 5' upstream sequence of CTG repeat region (CAGGGTTTTCCCAG), and resulting reads were

clustered by using aptamer clustering plugin. Obtained sequences containing a long repeat tract often show base alterations, deletions, additions and termination within and downstream of the repeat region (see supporting information, Table S3). These sequences were not excluded for counting sequencing reads of certain repeat numbers (Fig. 5). Since sequence reads for long repeats at around n = 20 is likely biased to lower values, we here focused mainly on distribution of the contraction products having repeat tracts shorter than 10.

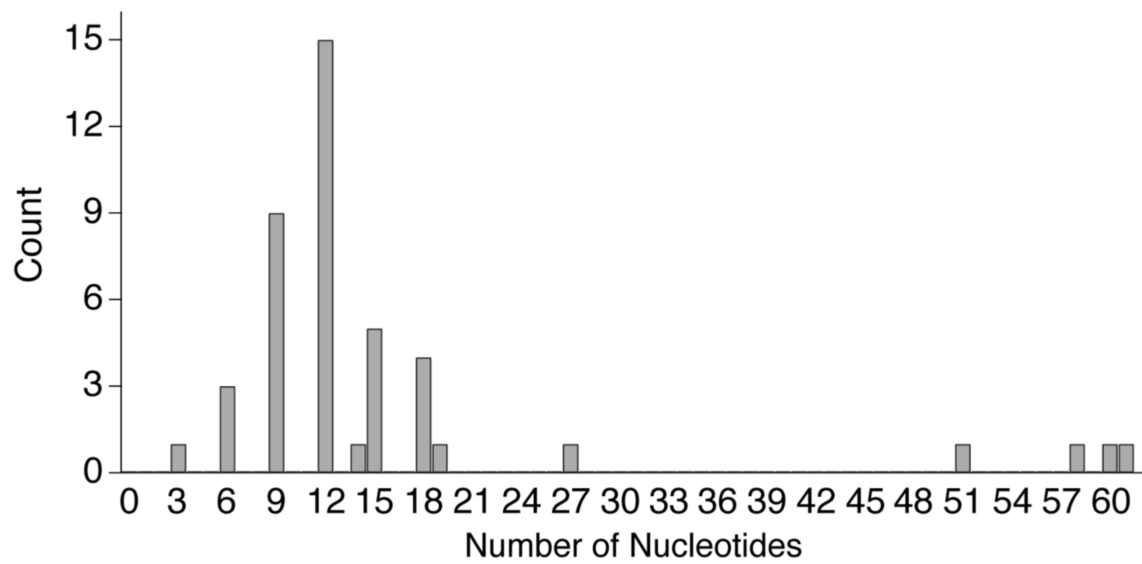

Fig. S1. Distributions of contraction products obtained by Sanger method. Counts of reads are plotted against number of nucleotides in the repeat region.

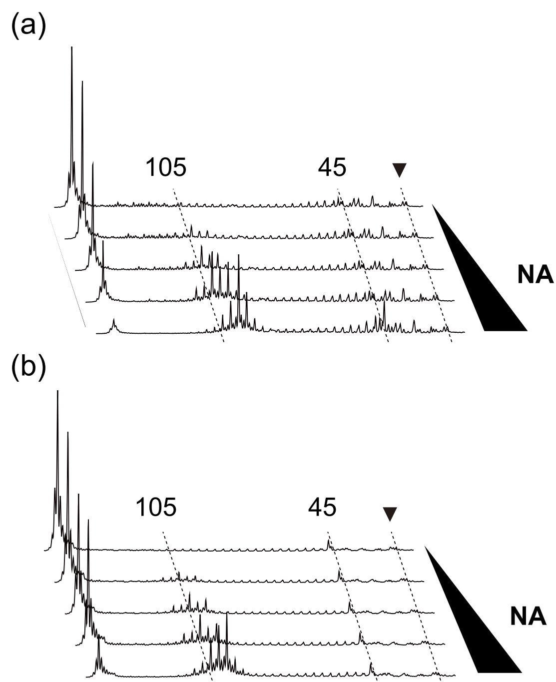

Fig. S2. Fragment analysis of the elongated primers by (a) Klenow fragment and (b) *96-7* DNA polymerase. From top to bottom, NA concentrations are 0, 1, 3, 10, and 30 μM, respectively.

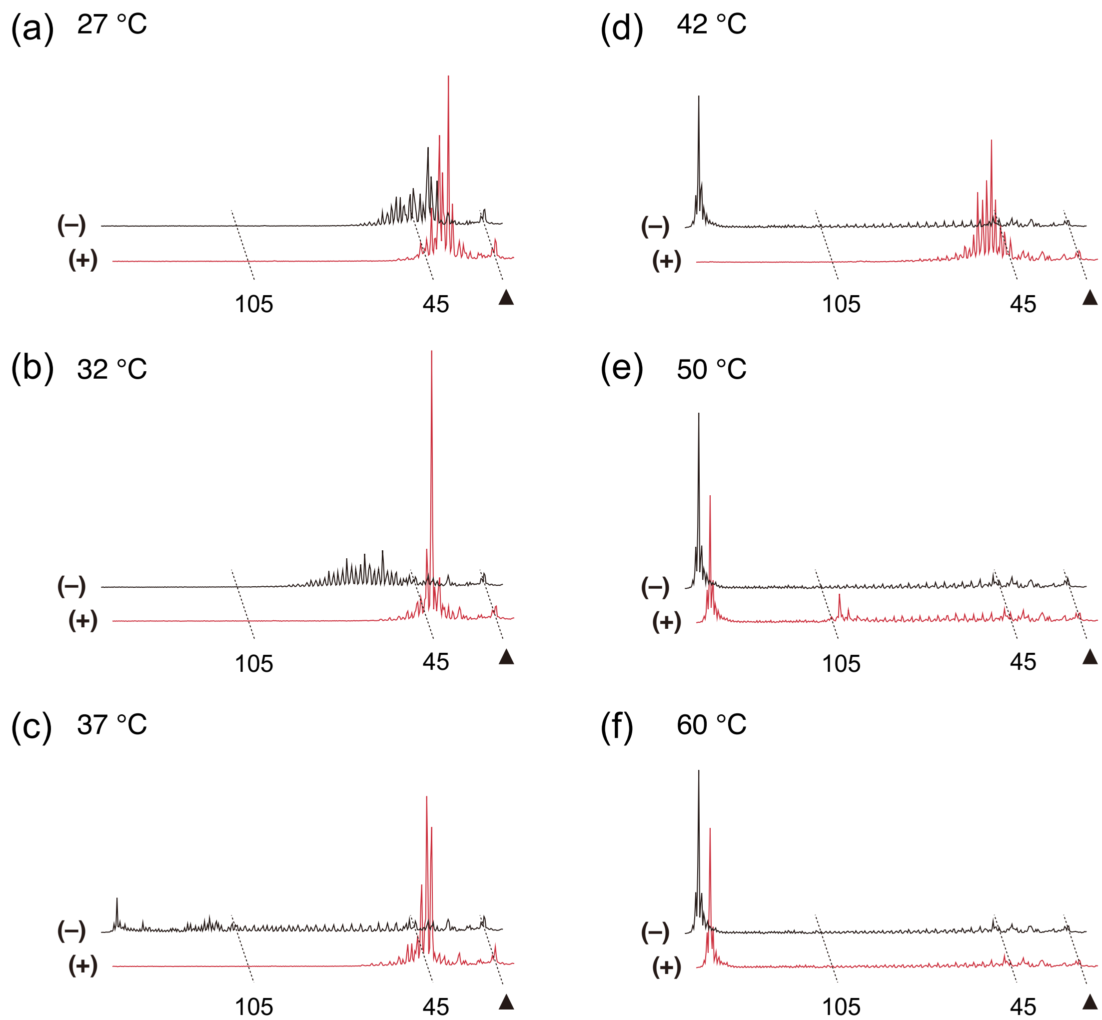

Fig. S3. Fragment analysis of primer extension reactions through CAG 20 repeat template by *ExTaq* polymerase in the presence of NA (30 μM). The primer extension reactions were carried out at designated temperatures. (a) Primer extension at 27 ºC; (b) 32 ºC; (c) 37 ºC; (d) 42 ºC; (e) 50 ºC; (f) 60 ºC. Fragment analysis data in the absence (top, black) and presence of 30 μM NA (bottom, red) are shown.

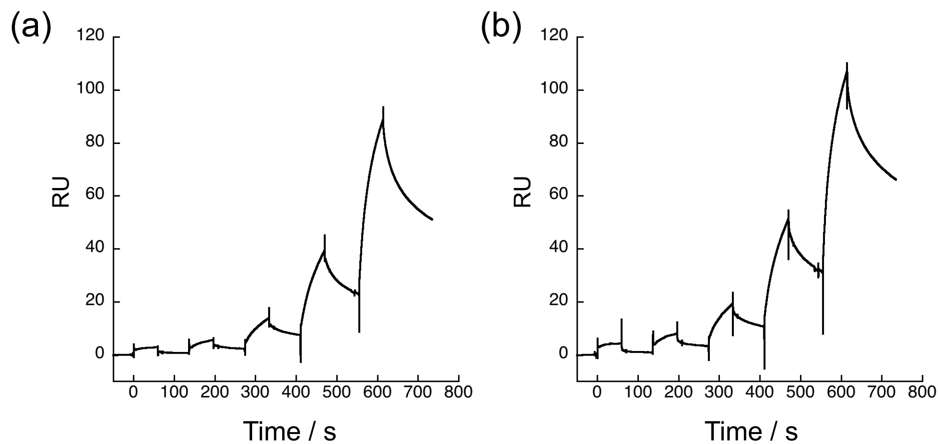

Fig. S4 NA binding to (CAG) repeat was analyzed by SPR single cycle kinetics using Biacore T200 system. d(CAG)10 and d(CAG)20 were immobilized onto SPR sensor chip (Series S sensor chip SA, cytiva). NA was sequentially injected for 60 sec each at concentrations of 0.063, 0.13, 0.25, 0.5, 1 μM. (a) d(CAG)10; (b) d(CAG)20. Immobilized amounts onto the sensor ship surface were 937.8 and 833.3, for d(CAG)10 and d(CAG)20, respectively. Note that amounts of immobilization are similar in this measurement, *i.e.*, the number of CAG units are similar but molar amounts of immobilized strands are smaller for d(CAG)20.

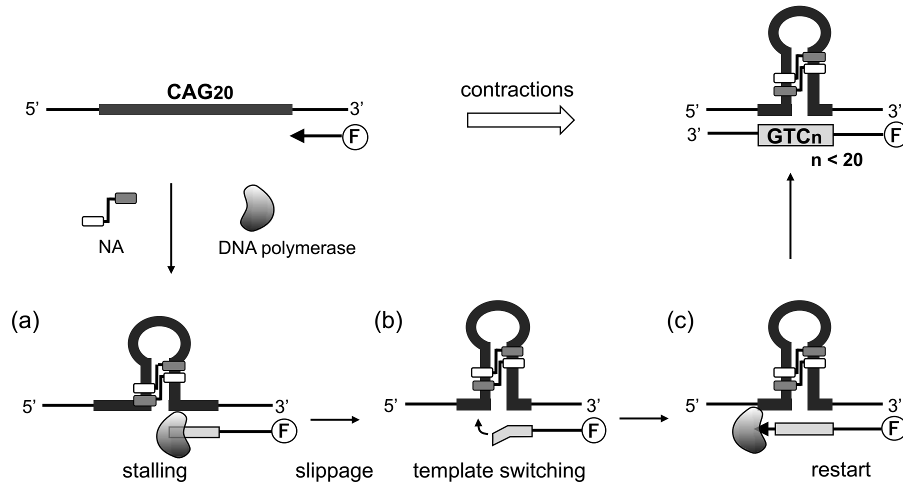

Fig. S5 Proposed mechanism of NA-induced repeat contractions during primer extension reaction. Strand slippage mechanism includes (a) stalling of DNA polymerase by NA-bound hairpin, then (b) slippage and switching template, followed by (c) restarting polymerase extension.

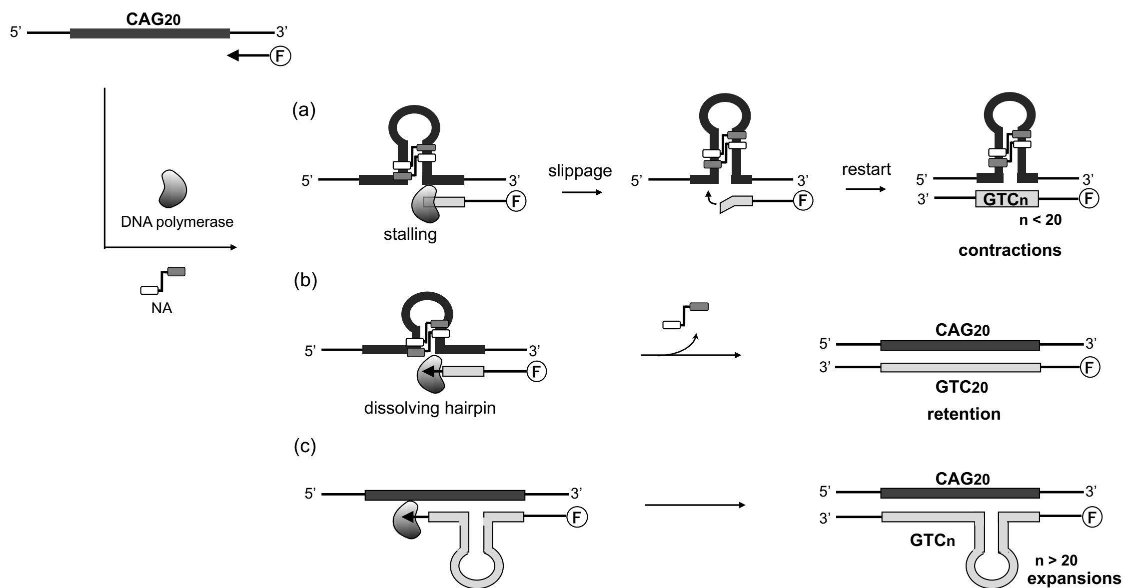

Fig. S6. Proposed mechanism of the NA-induced repeat contractions and possible scenarios during primer extension reaction. (a) NA-induced contractions of nascent strand by strand slippage. (b) Retention of the repeat length. (c) Repeat expansion involving formation of hairpin structure in the nascent strand. Note repeat expansions were not observed under the conditions in this study.

Plausible mechanism for the NA-induced contractions:

Strand slippage mechanism involves the formation of a hairpin structure that stalls polymerase (Fig. S5a), followed by switching template (Fig. S5b) and restarting polymerase extension (Fig.

S5c). NA-stabilized hairpin is a key to stalling the polymerase and the strand slippage. According to the proposed mechanism, contraction products keep CTG triads in the repeat region, *i.e.* contractions always occurs with the base triad as a unit. Although repeat expansions may occur if stable hairpins form in the nascent strand during polymerase elongation (Fig. S6c), no expansion were detected under the conditions examined here (for example, Fig. 3).

There were apparently preferred repeat lengths produced by NA-induced contractions (Fig. 2 and 5), which may be accounted for the proposed mechanism (Fig. S5 & S6a).[1] First, the abundance of contracted products having intermediate repeats (n > 10) was much smaller than shorter repeats (Fig. 2 and 5). Considering size of NA-bound hairpin defines length of skipped repeat units (Fig. S5b), the hairpin consisting of more than 10 CAG units is required for the NA-induced strand slippage. No contraction was detected using a DNA template having CAG 10 repeats (Fig. 3C), suggesting that threshold repeat number for the NA-induced contractions is more than 10. We speculate that contractions occur when the size of the NA-stabilized hairpin exceeds a threshold (producing contraction products with ~6 repeats) and then shorter products appear as increasing NA concentration. Short hairpins (n < 10) will be either unfolded by the polymerase progression or converted into more stable longer hairpins by further NA bindings (Figure S6b, retention of the repeat length). Second, little or no contraction products having very short (n = 0 and 1) repeats was observed (Fig. 2 and 5). In the proposed slippage mechanism, 3' terminal of elongating strand needs to re-hybridize downstream of the hairpin structure in the template switching step (Fig. S5b, c). Contractions to produce 0 or 1 repeat is unfavorable, because a short segment (≤ 3 base pairs) is only available for hybridization step to restart elongation (Fig. S5c).[2,3]

[1] (a) D. Canceill, S. D. Ehrlich, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 6647–6652. (b) D. Canceill, E. Viguera, S. D. Ehrlich, *J. Biol. Chem.* **1999**, *274*, 27481–27490. (c) E. Viguera, D. Canceill, S. D. Ehrlich, *EMBO J.* **2001**, *20*, 2587–2595.

[2] R. Patel, C. Lin, M. Laney, N. Kurn, S. Rose, E. F. Ullman, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 2969–2974.

[3] N. B. Leontis, W. Kwok, J. S. Newman, *Nucleic Acids Res.* **1991**, *19*, 759–766.

Table S1. Oligonucleotide sequences used in this study [a,b)]

CAG20: GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG**(CAG)**$_{20}$CATGAGACCGGATC CATTCATGTCGTGACTGGGAAAACCCTGGCG

CAG10: GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG**(CAG)**$_{10}$CATGAGACCGGATC CATTCATGTCGTGACTGGGAAAACCCTGGCG

CAG5: GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG**(CAG)**$_{5}$CATGAGACCGGATCC ATTCATGTCGTGACTGGGAAAACCCTGGCG

CTG20: GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG**(CTG)**$_{20}$CATGAGACCGGATC CATTCATGTCGTGACTGGGAAAACCCTGGCG

FAM-Primer1: FAM-CGCCAGGGTTTTCCCAGTCACGAC

Primer1: CGCCAGGGTTTTCCCAGTCACGAC

Primer2: GAACCACAACTCAAATAATGTCTAGA

Primer3: CCTCTCTATGGGCAGTCGGTGATGAACCACAACTCAAATAATGTCTAGA

Primer4: CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNNNNNNNCGCCAGGGTTTTCCCAGTC ACGAC

a) Sequences are shown from 5’ to 3’.

b) NNNNNNNNNN is barcode sequence for IonPGM sequencing. Hybridized region between templates and primers are underlined.

Table S2. Sequences of elongated primers through CAG20 template in the presence of 30 μM NA obtained by Sanger sequencing[a)]

	Count	Count %	repeat number	length in repeat	5' sequence	Sequence in repeat region	3' sequence
			20	60	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
1	13	29.5	4	12	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
2	7	15.9	3	9	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
3	5	11.4	5	15	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
4	3	6.8	6	18	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCAGCAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
5	2	4.5	2	6	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
6	1	2.3	19	61	GAACCACAACTCAAATAATGTCTAGAGATTAGAACACATAG	CAGCAGCAGCAGGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
7	1	2.3	19	60	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAACAGCAGCAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
8	1	2.3	18	58	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCACAGCAGCAGCAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
9	1	2.3	17	51	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
10	1	2.3	8	27	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCAGCAGCAGCAGCATCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
11	1	2.3	5	19	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAAGCAGCAGCAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
12	1	2.3	5	18	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCAGCAGCAGCAA	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
13	1	2.3	3	14	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAACACAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
14	1	2.3	4	12	GAACCACAACTCAAATAATGTCTAGAGGATTAGAAGACATAG	CAGCAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
15	1	2.3	4	12	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAA-CCCTGGCG
16	1	2.3	3	9	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACATAG	CAGCAGCAG	CATAAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
17	1	2.3	3	9	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACA**CAG**	CAGCAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
18	1	2.3	2	6	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGAC-TAG	CAGCAG	CATGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG
19	1	2.3	1	3	GAACCACAACTCAAATAATGTCTAGAGATTAGAAGACA**CAG**	CAG	C-TGAGACCGGATCCATTCATGTCGTGACTGGGAAAACCCTGGCG

a) 44 clones. Mutations except for repeat expansion and contraction are shown in blue.

Table S3-1. Top 20 sequences of elongated primers through CAG20 template in the presence of 30 μM NA obtained by IonPGM.[a)]

	Count	Count %	repeat number	5' trimmed sequence	Sequence
1	117580	13.86	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
2	48726	5.74	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
3	41928	4.94	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
4	16787	1.98	5	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
5	15964	1.88	2	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
6	8231	0.97	1	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
7	6691	0.79	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTTGTGGTTC
8	6354	0.75	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTG**CTATGTCTTCTAATCTCTAGTACATTATTTGAGTTGTGGTTC
9	4246	0.50	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTTGAGTTGTGGTTC
10	3404	0.40	5	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
11	3141	0.37	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTG**CTATGTCTTCTAATCTCTAGACTATTATTTGAGTTGTGGTTC
12	2982	0.35	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTG**______TCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
13	2726	0.32	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACTATTATTTGAGTTGTGGTTC
14	2719	0.32	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTG**CTATGTCTTCTAATCTCTAGACTAT_ATTTTGAGTTGTGGTTC
15	2598	0.31	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
16	2032	0.24	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
17	1955	0.23	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTG**___TGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
18	1859	0.22	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTG**CTATGTCTTCCTAATCTCTAGACATTATTTGAGTTGTGGTTC
19	1730	0.20	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTTATT_GAGTTGTGGTTC
20	1701	0.20	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTG**CTATGTCTTCTAATCTCTAGTACATTATTTGAGTTTGTGGTTC

a) Total processed read is 881803. Number of read after filtering by the trimmed sequence (shown as bold and underlined cases) is 848477. Possible mutations and sequencing errors are shown in blue.

Table S3-2. Top 20 sequences of elongated primers through CAG20 template in the presence of 10 μM NA[a)]

	Count	Count %	repeat number	5' trimmed sequence	Sequence
1	22769	5.96	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
2	17682	4.63	5	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
3	8430	2.21	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
4	7094	1.86	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
5	4501	1.18	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
6	2226	0.58	7	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGTACATTATTTGAGTTGTGGTTC
7	1667	0.44	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGATCATTATTTGTAGTTGCGGTTC
8	1625	0.43	5	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
9	1513	0.40	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
10	1485	0.39	2	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
11	1365	0.36	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACTATTATTTGAGTTGTGGTTC
12	1329	0.35	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGATCATTTATTTGTAGTTGCGGTTC
13	1186	0.31	8	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
14	963	0.25	7	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGTACATTATTTGAGTTTGTGGTTC
15	849	0.22	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTTATT_GAGTTGTGGTTC
16	738	0.19	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGATCATTTATT_____________
17	722	0.19	5	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGATCATTATTTGAGTTGTGGTTC
18	708	0.19	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGATCATTTATTT____________
19	567	0.15	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATACGGTCTCATG**CTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
20	552	0.14	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTTATTGATGTTGTGGTTC

a) Total processed read is 403115. Number of read after filtering by the trimmed sequence (shown as bold and underlined cases) is 381726. Possible mutations and sequencing errors are shown in blue.

Table S3-3. Top 20 sequences of elongated primers through CAG20 template in the presence of 3 μM NA[a)]

	Count	Count %	repeat number	5' trimmed sequence	Sequence
1	24890	3.23	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
2	23512	3.05	5	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
3	13528	1.75	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
4	11047	1.43	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
5	5642	0.73	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
6	3572	0.46	7	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGTACATTATTTGAGTTGTGGTTC
7	2802	0.36	8	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
8	2792	0.36	5	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
9	2574	0.33	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
10	2318	0.30	2	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
11	2019	0.26	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACTATTATTTGAGTTGTGGTTC
12	1969	0.26	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGATCATTTATTTGTAGTTGCGGTTC
13	1852	0.24	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTTATT_GAGTTGTGGTTC
14	1819	0.24	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGATCATTATTTGTAGTTGCGGTTC
15	1688	0.22	7	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGTACATTATTTGAGTTTGTGGTTC
16	1492	0.19	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGATCATTTATT____________
17	1352	0.18	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGATCATTTATTT___________
18	1331	0.17	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCTA_GTCT
19	1150	0.15	5	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGATCATTATTTGAGTTGTGGTTC
20	1016	0.13	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCTA_GTCTCT

a) Total processed read is 806271. Number of read after filtering by the trimmed sequence (shown as bold and underlined cases) is 770883. Possible mutations and sequencing errors are shown in blue.

Table S3-4. Top 20 sequences of elongated primers through CAG20 template in the presence of 1 μM NA[a)]

	Count	Count %	repeat number	5' trimmed sequence	Sequence
1	9834	1.56	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
2	9152	1.45	5	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
3	4911	0.78	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
4	3418	0.54	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
5	2747	0.44	3	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
6	1643	0.26	7	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGTACATTATTTGAGTTGTGGTTC
7	1639	0.26	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCTA_GTCT
8	1249	0.20	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCTA_GTCT_CT
9	1210	0.19	8	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
10	1206	0.19	2	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
11	1189	0.19	4	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACTATTATTTGAGTTGTGGTTC
12	1171	0.19	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCTA_GTCT_C
13	1155	0.18	5	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
14	1139	0.18	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCTA_GTCT_CTA
15	994	0.16	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGATCATTTATTTGTAGTTGCGGTTC
16	937	0.15	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGACATTTATTTGAGTTGTGGTTC
17	929	0.15	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCTA_GTC
18	885	0.14	7	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGTACATTATTTGAGTTTGTGGTTC
19	796	0.13	6	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTATGTCTTCTAATCTCTAGATCATTATTTGTAGTTGCGGTTC
20	758	0.12	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCTA_GTCT_CTAT

a) Total processed read is 657666. Number of read after filtering by the trimmed sequence (shown as bold and underlined cases) is 630706. Possible mutations and sequencing errors are shown in blue.

Table S3-5. Top 20 sequences of elongated primers through CAG20 template in the absence of NA[a)]

	Count	Count %	repeat number	5' trimmed sequence	Sequence
1	1767	0.40	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTCTA_GTCT
2	1144	0.26	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTCTA_GTCT_CT
3	1064	0.24	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTCTA_GTCT_C
4	992	0.22	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTCTA_GTCT_CTA
5	912	0.21	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTCTA_GTC
6	867	0.20	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTCTA_GTCTTC
7	814	0.18	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTCTA_GTCTT
8	791	0.18	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTCTA_GTCTTCT
9	765	0.17	5	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTG**CTATGTCTTCTAATCTCTAGACATTATTTGAGTTGTGGTTC
10	683	0.15	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTCTA_GTCT_CTAT
11	681	0.15	18	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTG**CTGCTG**CT
12	651	0.15	18	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTG**CTGCTG**CTT
13	650	0.15	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTGCTATCGT
14	631	0.14	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTGCTATCG
15	618	0.14	20	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTATCGTC
16	610	0.14	17	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTG**CTG**CT
17	571	0.13	20	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTATCGTCT
18	567	0.13	19	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTCTAGT
19	567	0.13	18	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTG**CTGCTG**
20	565	0.13	18	CGC**CAGGGTTTTCCCAG**	TCACGACATGAATGGATCCGGTCTCATG**CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG**CTTG**CTGCTG**C

a) Total processed read is 463992. Number of read after filtering by the trimmed sequence (shown as bold and underlined cases) is 441411. Sequencing of the CAG/CTG 20 repeat is difficult under the conditions we used. From near end of the repeat region (CTG17~20), the sequences showed significant variations and the sequencing were often aborted. Possible mutations and sequencing errors are shown in blue.