

# SCIENTIFIC REPORTS



OPEN

## Minimizing carry-over PCR contamination in expanded CAG/CTG repeat instability applications

Lorène Aeschbach & Vincent Dion 

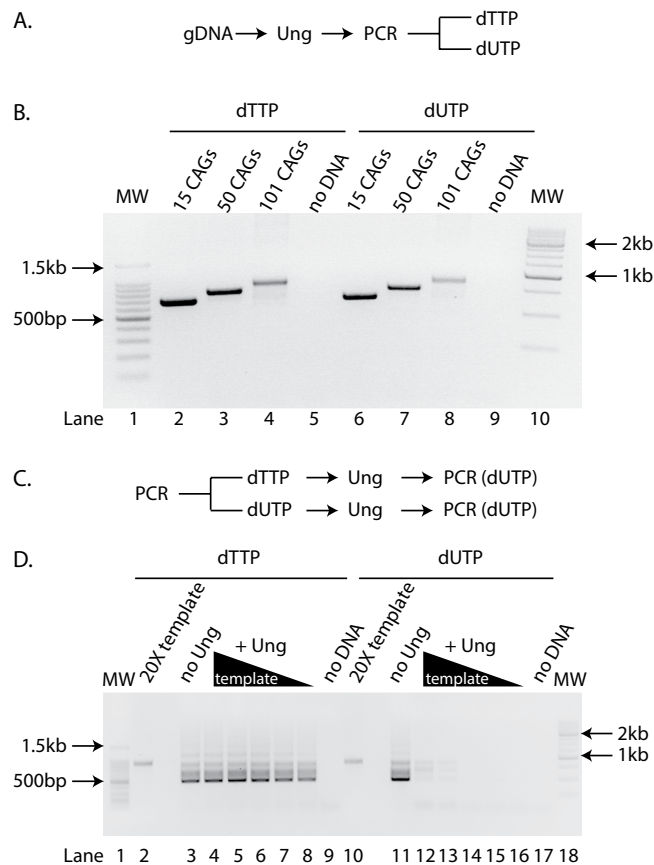
Expanded CAG/CTG repeats underlie the aetiology of 14 neurological and neuromuscular disorders. The size of the repeat tract determines in large part the severity of these disorders with longer tracts causing more severe phenotypes. Expanded CAG/CTG repeats are also unstable in somatic tissues, which is thought to modify disease progression. Routine molecular biology applications involving these repeats, including quantifying their instability, are plagued by low PCR yields. This leads to the need for setting up more PCRs of the same locus, thereby increasing the risk of carry-over contamination. Here we aimed to reduce this risk by pre-treating the samples with a Uracil *N*-Glycosylase (Ung) and using dUTP instead of dTTP in PCRs. We successfully applied this method to the PCR amplification of expanded CAG/CTG repeats, their sequencing, and their molecular cloning. In addition, we optimized the gold-standard method for measuring repeat instability, small-pool PCR (SP-PCR), such that it can be used together with Ung and dUTP-containing PCRs, without compromising data quality. We performed SP-PCR on myotonic-dystrophy-derived samples containing an expansion as large as 1000 repeats, demonstrating the applicability to clinically-relevant material. Thus, we expect the protocols herein to be applicable for molecular diagnostics of expanded repeat disorders.

Tandem trinucleotide repeats (TNRs) are common in the human genome and are polymorphic in the population<sup>1</sup>. Their size at any one locus is usually below about 35 triplets. When these sequences expand beyond this threshold, however, they become highly unstable with a bias towards expansions in both somatic and germ cells<sup>2–5</sup>. In addition, TNR expansions cause a wide variety of human neurological, neuromuscular, and neurodegenerative disorders<sup>4</sup>. Importantly, longer repeat tracts tend to provoke more severe symptoms<sup>6</sup>. Therefore, expansions from one generation to the next leads to an earlier age of onset in the progeny, providing an elegant molecular explanation for the phenomenon of anticipation. Meanwhile, ongoing expansions in somatic tissues may accelerate pathogenesis<sup>4,7–9</sup>.

The mechanisms of repeat instability remain tedious to assay in part because of the low yields during PCR amplification of expanded TNRs<sup>5</sup>. This is especially true for CAG/CTG repeats (henceforth referred to as the triplet on the non-transcribed strand of the sense transcript). Thus, several PCRs are often setup in parallel to obtain sufficient amounts of DNA for downstream applications. This opens the door to carry-over contamination in which the amplicon produced in a previous reaction contaminates the subsequent ones, skewing the results or rendering them useless. Typical methods to palliate this risk include having dedicated 'pre-PCR' room, pipettes, and solutions. A hood with a UV light source can also be used. These precautions reduce, but do not eliminate, the risk of contamination. This is particularly important in the context of small-pool PCR (SP-PCR), the gold standard assay to measure repeat instability<sup>10,11</sup>. This technique consists of performing many PCRs in parallel, each using only a few genomes as template to bypass the propensity of amplifying the shorter alleles in the population. The products are then separated on agarose gels and transferred onto a membrane for Southern blotting. This assay is exquisitely sensitive and thus many no-DNA controls must be included to ensure that there is no carry-over contamination. If the controls show amplification products, the experiment must be repeated with fresh solutions, wasting time and resources.

Carry-over contamination is a problem for molecular diagnostics<sup>12</sup>, ancient DNA amplification<sup>13</sup>, and for any application that requires amplifying the same few genomic loci<sup>14</sup>. A common solution is to pre-treat the samples with a bacterial uracil *N*-glycosylase, Ung, that degrades any contaminating DNA containing uracil and to replace dTTP with dUTP in all PCRs<sup>15</sup>. It was unclear whether this method is suitable for molecular applications

Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, 1015, Lausanne, Switzerland. Correspondence and requests for materials should be addressed to V.D. (email: [vincent.dion@unil.ch](mailto:vincent.dion@unil.ch))



**Figure 1.** dTTP substituted with dUTP allows PCR amplification of expanded CAG repeats. **(A)** Experimental scheme used to generate the PCR amplicons found in B. gDNA = genomic DNA **(B)** Agarose gel of PCR amplicons after pre-treatment with Ung and containing either dTTP (lanes 2–5) or dUTP (lanes 6–9) from genomic DNA isolated from GFP(CAG)<sub>15</sub> (lanes 2, 6), GFP(CAG)<sub>50</sub> (lanes 3, 7), GFP(CAG)<sub>101</sub> (lanes 4, 8), or without DNA added (lanes 5, 9). MW = molecular weight marker (lanes 1, 10). **(C)** Experimental scheme used for panel D. **(D)** Agarose gel of PCR amplicons obtained after pre-treatment with Ung of a template containing either dTTP (lanes 2–9) or dUTP (lanes 10–17). The template was generated from amplifying GFP(CAG)<sub>101</sub> genomic DNA. The PCRs with no Ung treatment (lanes 3, 11) contained 1X concentration of the template amplicon. Ten-fold dilution series (lanes 4–8 et 12–16) of the template (20X loaded directly without PCR in (lanes 2, 10)) and no DNA controls are also shown (lanes 9 and 17). MW = molecular weight marker (lanes 1, 18).

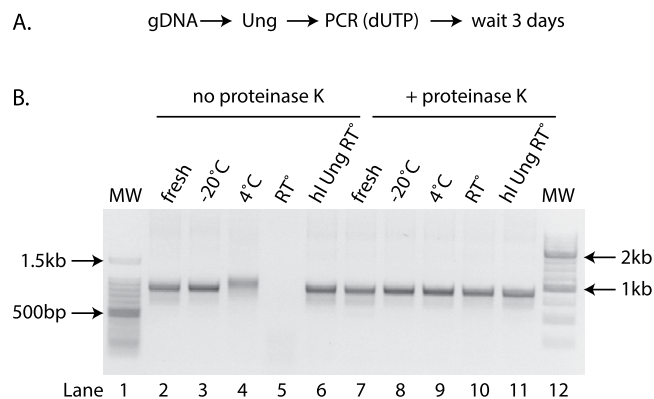
dNTP used	no. of repeats*
dTTP	98, 99, 100, 101, 102 (2)
dUTP	92 (2), 101 (2), 102, 103

**Table 1.** dUTP and dTTP yield the similar repeat sizes upon Sanger sequencing. \*The number in parenthesis is the number of times we obtained this repeat size (for  $n > 1$ ).

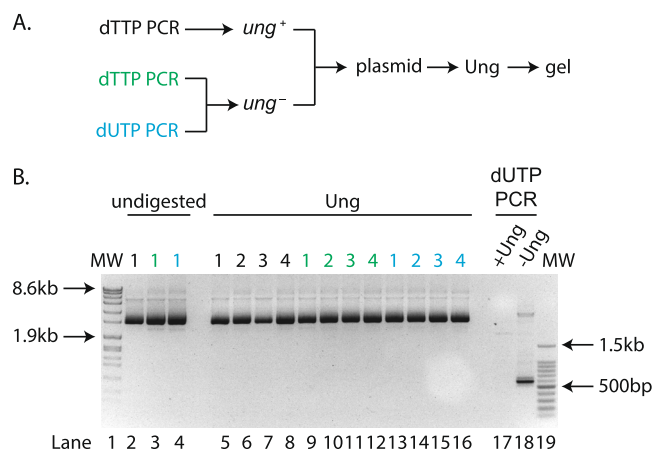
involving expanded CAG repeats given the difficulties associated with their PCR amplification. Indeed, adding an enzyme that degrades DNA to a difficult-to-amplify template was not trivial. Here we adapted protocols and show the applicability of this method in four molecular biology techniques commonly used when working with expanded CAG repeats: PCR amplification, sequencing, molecular cloning, and SP-PCR.

## Results and Discussion

**dTTP substituted with dUTP allows PCR amplification of expanded CAG repeats.** We first wanted to determine whether expanded CAG repeats of varying lengths could be amplified using dUTP instead of dTTP in the PCR and that Ung did not interfere with the amplification. We therefore pre-treated samples with Ung and set up PCRs containing either dUTP and dTTP (Fig. 1A). We used genomic DNA isolated from HEK293-derived cells harbouring a single copy transgene with CAG sizes ranging from 15 to 101 (GFP(CAG)<sub>x</sub>)<sup>16</sup>. We found that the Ung-treated and dUTP-containing reactions consistently had a slightly lower yield for repeat sizes ranging from 15 to 101 CAG repeats compared to dTTP-containing reactions (Fig. 1B). This did not affect the apparent size of the repeat tract after agarose gel electrophoresis. Importantly, uracil-containing



**Figure 2.** Using heat-labile Ung and/or Proteinase K prevents the degradation of uracil-containing amplicons. **(A)** Experimental scheme used. **(B)** Agarose gel of amplicons from genomic DNA of GFP(CAG)<sub>101</sub> containing uracil and treated post-PCR with (lanes 7–11) or without (lanes 2–6) proteinase K three days after the PCR was performed (lanes 3–6 and 8–11) or immediately after the PCR (fresh, lanes 2, 7). The samples were stored at -20°C (lanes 3, 8), 4°C (lanes 4, 9), room temperature (RT°, lanes 5, 10), or were amplified using a heat labile (hl) Ung instead of the conventional Ung and stored at room temperature (lanes 6, 11). MW = molecular weight marker (lanes 1 and 12).



**Figure 3.** Plasmids cloned into *ung*<sup>-</sup> cells contain undetectable levels of uracil. **(A)** Experimental scheme used. **(B)** Gel of plasmids isolated from *ung*<sup>+</sup> cells transformed with a thymine containing insert (black – lanes 2, 5–8) or *ung*<sup>-</sup> cells transformed with dTTP (green – 3, 9–12) or dUTP (blue – 4, 13–16). The samples remained untreated (lanes 2–4) or were incubated with 0.5U of Ung (lanes 5–16). Controls for Ung digestion consist of SUMO1 cDNA amplicons containing uracil digested with Ung (lane 17) or left untreated (lane 18). MW = molecular weight markers (lanes 1, 19). The numbers above the lanes refer to the name of the clones tested.

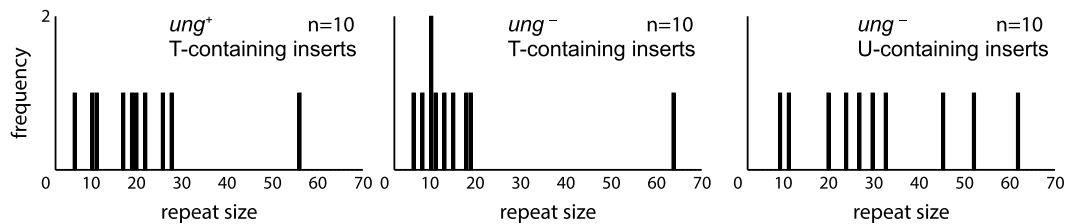
Nucleotide used in PCR	DH5α*	DH5α Δ <i>ung</i> *
dTTP	60 ± 49	200 ± 34
dUTP	0 ± 0	142 ± 26
No insert†	0 ± 0	0 ± 0

**Table 2.** Number of colonies obtained when cloning thymine or uracil-containing amplicons of the SUMO1 cDNA. \*Number of colonies derived from 3 independent transformations of 3 independent PCR amplicons ± standard deviation. The competence was similar with 10<sup>8</sup> colony/μg of pUC19 for both strains. The same amount of DNA was used for the cloning and transformation reactions (see methods). †The cloning reaction was performed in the absence of any insert to test for self-circularization of the plasmid backbone.

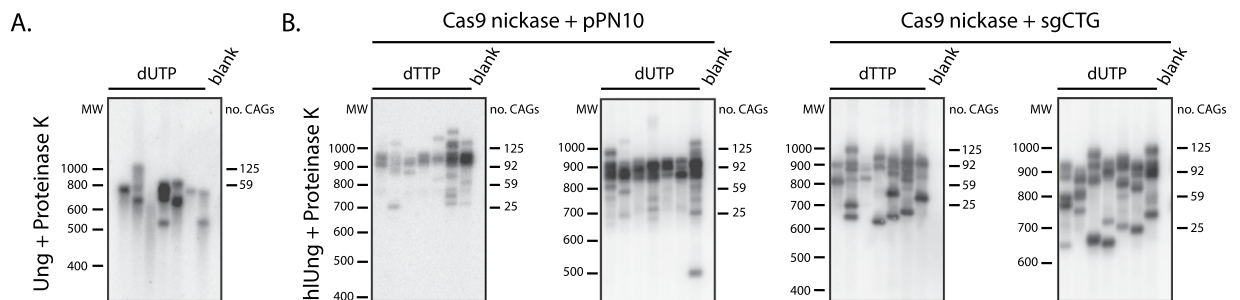
amplicons were efficiently degraded by Ung prior to PCR (Fig. 1C,D). We conclude that pre-treatment with Ung and the use of dUTP in the PCR allows for an efficient amplification of expanded CAG repeats and prevents carry-over contamination.

Nucleotide used in PCR	DH5 $\alpha$ *	DH5 $\alpha$ $\Delta$ ung*
dTTP	193 $\pm$ 117	88 $\pm$ 59
dUTP	0 $\pm$ 0	54 $\pm$ 45
No insert <sup>†</sup>	0 $\pm$ 0	0 $\pm$ 0

**Table 3.** Number of colonies obtained when cloning thymine or uracil-containing amplicons with 89 CAGs. \*Number of colonies derived from 3 independent transformations of 3 independent PCR amplicons. The competence was similar with  $10^8$  colony/ $\mu$ g of pUC19 for both strains. The same amount of DNA was used for the cloning and transformation reactions (see methods). <sup>†</sup>The cloning reaction was performed in the absence of any insert to test for self-circularization of the plasmid backbone. This experiment was done twice in the  $\Delta$ ung strain.



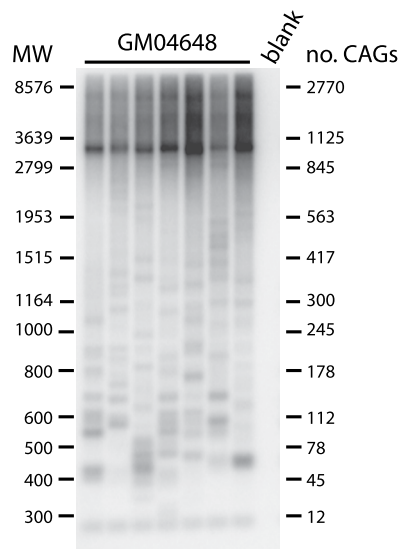
**Figure 4.** Repeat sizes recovered after molecular cloning of amplicons containing uracil or thymidine using DH5 $\alpha$  and DH5 $\alpha$   $\Delta$ ung strains.



**Figure 5.** Small-pool PCR of expanded repeats using Ung and dUTP. (A) SP-PCR blot from reactions pretreated with the standard Ung, amplified with dUTP, and stopped using Proteinase K. The visible bands are shorter than in the samples seen in B., and the smearing below the bands is evident. Each lane contains the product of a PCR with 50 pg of template DNA derived from GFP(CAG)<sub>101</sub> treated with sgCTG and the Cas9 nickase. (B) SP-PCR blots of amplicons generated using dTTP or dUTP, pretreated with hl Ung, and stopped after the PCR with Proteinase K. Each lane contains the product of a PCR that used 250 pg of genomic DNA from GFP(CAG)<sub>101</sub> treated with the Cas9 nickase and either the sgCTG or an empty guide RNA vector (pPN10). In all cases, the probe used was generated with dUTP.

**dUTP-containing amplicons can be sequenced with ease.** We next determined whether the amplicons containing uracil could be sequenced as efficiently as those containing thymine. To do so, we set up 6 different sequencing reactions from two PCR products amplified with either dUTP or dTTP from genomic DNA obtained from GFP(CAG)<sub>101</sub> cells. We found no difference in the apparent size determination between the two nucleotides used (Table 1,  $P = 1.0$  between uracil and thymidine-containing amplicons using a Wilcoxon U-test). These observations show that dUTP can be substituted for dTTP when sequencing of expanded CAG repeats.

**Keeping residual activity of Ung in check.** We noticed that the Ung-treated and uracil-containing amplicons degraded completely at room temperature over a period of 3 days (Fig. 2). We saw some degradation even at 4 °C, but none when the samples were stored at  $-20$  °C. This is consistent with the observation that Ung retains enzymatic activity even after many PCR cycles<sup>17</sup>. Indeed, the Ung that we used is active at temperatures below 55 °C even after repeated denaturation-renaturation cycles, according to the manufacturer. Given that some applications, such as SP-PCR and molecular cloning, require small amounts of DNA, degradation may have a large impact. We circumvented this problem using two different strategies. The first was to deactivate the Ung by digesting with Proteinase K immediately after the reaction. The second approach was to use a heat-labile (hl) Ung<sup>18</sup> (Fig. 2). This enzyme works well at 20 °C, but an incubation at 95 °C for 7 minutes inactivates it.



**Figure 6.** Small-pool PCR of 1000 CTGs at the *DMPK* locus from sample GM04648. 3 ng/reaction was used for these SP-PCRs.

**Cloning dUTP-containing expanded repeats.** We next sought to determine whether it is possible to clone uracil-containing amplicons reliably. This was previously done using *ung<sup>-</sup> dut<sup>-</sup> E. coli* such that the plasmids isolated from these strains contained up to 20% of uracil<sup>13</sup>. Wild-type strains, by contrast, have less than one uracil residue per 10<sup>6</sup> nucleotides<sup>19</sup>. Here we wanted to know whether we could clone uracil-containing amplicons using a *ung<sup>-</sup>* strain and convert uracil to thymine residues in the process. Thus, we knocked out *ung* in DH5 $\alpha$  cells. This strain background is used for a variety of molecular cloning purposes. We used the well-established lambda Red recombination-based approach<sup>20,21</sup> to replace the endogenous *ung* with the coding sequence of a gene that provides resistance to spectinomycin, generating a  $\Delta$ *ung* allele, *ung-9*. Thus, the more commonly used antibiotics, including ampicillin, chloramphenicol, and kanamycin, remain available for cloning.

We tested whether cloning with uracil-containing amplicons was possible by amplifying a non-repetitive sequence, the SUMO1 cDNA, and then cloning it using the Zero Blunt TOPO PCR kit. Significantly, we were unsuccessful in cloning uracil-containing amplicons when we used the standard Ung during the pre-PCR treatment (data not shown). This was presumably because of degradation of the amplicons prior to transformation. We overcame this problem by using the hUng before and proteinase K treatment after the PCR. Under these conditions, we found that cloning was efficient with thymine-containing amplicons in both *ung<sup>+</sup>* and *ung<sup>-</sup>* strains (Table 2). By contrast, cloning uracil-containing amplicons with the *ung<sup>+</sup>* strain yielded no colony, presumably because the uracil residues are digested and the plasmid degraded once inside the cell. In *ung<sup>-</sup>* cells, however, the cloning efficiency was similar whether uracil or thymine was present in the amplicon (Two-tailed Student's t-test  $p = 0.12$ ). We next performed colony PCR and/or Sanger sequencing of the plasmids from 8 independent clones from each of the conditions that yielded colonies. In every case, we could confirm the presence of the SUMO1 cDNA insert (not shown). Taken together, our results demonstrate that it is possible to clone uracil-containing amplicons in *ung<sup>-</sup>* bacteria.

To determine whether our *ung<sup>-</sup>* strain could convert uracil to thymine residues, we isolated plasmids from both *ung<sup>+</sup>* and *ung<sup>-</sup>* strains and digested them with Ung *in vitro*. We found no evidence that plasmids isolated from our *ung<sup>-</sup>* strain contained any uracil (Fig. 3), suggesting that these residues are indeed converted to thymine *in vivo*.

We considered two important parameters when cloning expanded repeats. First, the efficiency of the cloning reaction itself and, second, the instability incurred during the cloning procedure. To assess the former, we used the same approach as for cloning the SUMO1 cDNA and calculated the number of colonies obtained when cloning a PCR product amplified from a plasmid sporting an unstable repeat tract of up to 89 CAGs and containing uracil or thymidine into *ung<sup>+</sup>* and *ung<sup>-</sup>* strains (Table 3). We found similar numbers as for the SUMO1 cloning: uracil-containing amplicons only survived in the strain lacking *ung*. These results suggest that cloning uracil- and expanded CAG-containing amplicons can be achieved at high efficiencies in a *ung<sup>-</sup>* strain.

To assess whether uracil-containing amplicons retained the expanded repeats throughout the cloning procedure, we isolated plasmid DNA from 10 different colonies from all three conditions yielding colonies and counted the number of CAGs blindly after Sanger sequencing. As expected, the plasmids isolated contained between 5 and 64 CAGs from the plasmids isolated, all contained the repeat tract. Interestingly, we found that the size distribution was slightly longer when we transformed in a *ung<sup>-</sup>* strain uracil-containing inserts compared to thymine-containing ones (Fig. 4 -  $P = 0.03$  Wilcoxon U-test), implying that *ung<sup>-</sup>* strains and uracil-containing amplicons are conducive to cloning expanded repeat tracts.

Primer	Sequence	References
M13For	TGTAACACGACGCCAG	—
M13Rev	CAGGAAACAGCTATGAC	—
oVIN-460	TCTGCAAATTCAGTGATGC	16
oVIN-1251	GAGCGTGGGTCTCCGCCAG	This study
oVIN-1252	CACTTTGCGAACCAACGATA	This study
oVIN-1425	GACCTCATAAGATAGGCTT	This study
oVIN-1448	AAAGAAGCAGTTAAGCTAGGCCGATTGAAGATTCGCAGGAGAGCGAGATGTGCTGCCCAAGGTTGCCGGG	This study
oVIN-1449	ACTCTGCCATCCGGCATTTCGCCGAAATTTACTCACTCTCTGCCGTAATGTTTGCAAGCAGCAGATTA	This study
oVIN-1451	GTACCCCAAAAAACAGTCA	This study
oVIN-1452	ATTTGTTTCAGCCTGTCGC	This study
oVIN-1453	GAAAATGGTACTGGTGTGG	This study
oVIN-1450	GCAGATCAGTTGGAAGAATT	This study
SP6	TAATACGACTCACTATAGG	—
T7	ATTTAGGTGACACTATAG	—

**Table 4.** Primers used in this study.

Plasmid	Content	References
pKM208	Ampicillin resistance, temperature sensitive origin of replication, Lambda RED proteins for recombineering.	23
pVIN-109	GFP transgene containing 40 CAG repeats. Ampicillin resistance.	This study
pVIN-117	GFP transgene containing between 40 and 89 CAG repeats and a modified intron to be described elsewhere.	This study
pPN10	Kanamycin resistance, U6 promoter, guide RNA scaffold without a target sequence	16
pPN10-gCTG	Kanamycin resistance, single-guide RNA with 6 CTG repeats as target sequence	16
pDONR-SUMO1	SUMO1 cDNA, Kanamycin resistance.	This study

**Table 5.** Plasmids used in this study.

**Small-pool PCR using Ung pre-treatment and dUTP.** We next asked whether the gold standard assay to measure repeat instability frequencies, SP-PCR, was amenable to Ung treatment and dUTP-containing PCR. We expected this assay to be a challenge because of the small amounts of DNA used as templates and of the residual activity of Ung may lead to degradation of the products before we can visualize them. Indeed, we found that using the standard Ung led to blots so smeary that the results were unusable, which implied a considerable level of degradation. The treatment with proteinase K decreased the smearing somewhat (Fig. 5A). We observed sharp bands on the membranes only once we combined proteinase K and hUng (Fig. 5B). These SP-PCR blots were done using a probe amplified after a Ung pre-treatment and containing dUTP, showing that the presence of uracil does not impair nick translation. We conclude that with these changes to the standard protocols, SP-PCR can be done efficiently with hUng and dUTP.

We confirmed that neither the hUng treatment nor the use of dUTP affect the accuracy of SP-PCR by performing the assay on GFP(CAG)<sub>101</sub> cells that were treated with the Cas9 nickase and a single-guide RNA targeting the CAG tract<sup>16</sup>. We found previously that this treatment led to an increase in the number of contractions found in the sample. Indeed, we observed here the increase in the number of contractions in both the SP-PCR made with dTTP and dUTP (Fig. 5B). These results demonstrate that the measures used to prevent carry-over contamination do not affect the result interpretation.

To determine whether we could combine Ung and dUTPs on samples with longer repeat tracts, we performed SP-PCR on DNA isolated from lymphoblastoids from a myotonic dystrophy type 1 individual (GM04648). This individual is reported to harbour one allele with 5 CTGs in the 3' untranslated region of the *DMPK* gene, the other contains ~1000 CTGs. SP-PCR with Ung-treated and dUTP-containing samples amplified bands at the expected size as well as a number of contractions (Fig. 6). Therefore, our method is applicable to a wide variety of samples, including clinically relevant material.

Our experiments show that a pre-treatment with Ung and the use of dUTP in PCRs are compatible with the most widely used molecular biology techniques involving expanded CAG repeats. In particular, we can PCR-amplify, sequence, and clone expanded repeats with this method as well as perform SP-PCR with minor modifications to the standard protocols. Since the introduction of this technology in our laboratory, however, we have found that for some loci, we had to redesign primers. We have demonstrated that this method is useful for repeat sizes as long as 1000 CTG, does not cloud data interpretation, and can be used for clinically relevant material. Thus, we predict that Ung and dUTP can be incorporated to current protocols used for the molecular

diagnostics of expanded repeat disorders. Although thorough testing is still required, the inclusion of Ung and dUTP to clinical settings is already commonplace and thus we expect implementation to be straightforward. Taken together, our results demonstrate the broad application of this method to both routine molecular biology and molecular diagnostics of expanded repeat disorders.

## Materials and Methods

**Cell culture.** GFP(CAG)<sub>x</sub> cells used here are described elsewhere<sup>16,22</sup>. They were propagated in DMEM supplemented with 10% fetal bovine serum, 100 U ml<sup>-1</sup> penicillin, 100 µg ml<sup>-1</sup> streptomycin, 15 µg ml<sup>-1</sup> blasticidin and 150 µg ml<sup>-1</sup> hygromycin and were maintained at 37 °C in a 5% CO<sub>2</sub> incubator. They tested negative for mycoplasma at the beginning of our experiments using the MycoAlert Detection kit (Lonza LT07-218) and at the end using the service from GATC Biotech AG (Constance, Germany). The lymphoblastoids cells GM04648 were obtained from the Coriell repository (coriell.org). They were grown in RPMI with 15% fetal bovine serum, 2 mM L-Glutamine, and penicillin/streptomycin. They were also maintained at 37 °C in a 5% CO<sub>2</sub> incubator. They tested negative for mycoplasma using the service provided by GATC Biotech AG.

The nickase treatment was done as before<sup>16</sup>, with GFP(CAG)<sub>101</sub> cells transfected with either pPN10, an empty plasmid, or pPN10-gCTG together with a vector expressing the Cas9 D10A nickase variant using Lipofectamine 2000 (Life Technologies) on day 1 and again on day 5 and 8. On day 12, we harvested the cells and found that those transfected with pPN10-gCTG used for the SP-PCR in Fig. 5 had an increase of 2.6 fold in the brightest 1% of the cells compared to pPN10-transfected cells. GFP intensities were determined using an Accuri C6 flow cytometer from BD.

**Bacterial strains.** The *ung-9* allele was made by amplifying the spectinomycin resistance gene from pENTR223.1 with oVIN-1448 and oVIN-1449 (see Table 4 for primer sequences) and transforming the resulting products into DH5α cells harbouring pKM208. This strain and plasmid were a gift of Kenan Murphy (Addgene plasmid # 13077). pKM208 contains the Red λ phage proteins under a lactose-inducible promoter, a temperature sensitive origin of replication, and a gene expressing β-lactamase<sup>23</sup>. We followed a previously published protocol<sup>20</sup>, except that we generated electrocompetent cells using our standard laboratory protocol from cells grown to exponential phase at 30 °C in LB supplemented with 100 µg ml<sup>-1</sup> ampicillin and induced with 0.1 mM IPTG. After electroporation, the cells were selected at 37 °C on LB agar plates containing 100 µg ml<sup>-1</sup> of spectinomycin. The bacteria were then tested for the loss of ampicillin resistance (and thus loss of pKM208) at 30 °C. Colony PCR was used to confirm the loss of *ung* and the gain of the spectinomycin resistance gene both 5' and 3' of the insertion site using primers oVIN-1451 and oVIN-1452 for the 5' and oVIN-1450 and oVIN-1453 for the 3' end. DH5α *ung*<sup>+</sup> control strains were made by curing the strain containing pKM208 obtained from Addgene by streaking the cells twice sequentially on LB agar plates incubated at 37 °C and then selecting colonies that could not grow on LB agar plates containing 100 µg ml<sup>-1</sup> ampicillin at 30 °C. Competence was calculated by transforming a known amount of pUC19 and counting the resulting colonies on LB agar plates containing 100 µg ml<sup>-1</sup> ampicillin.

**Ung treatment, PCR amplification, and sequencing.** The standard Ung was obtained from Thermo Fisher Scientific, whereas hUng was from Roche. The polymerase used was the Phusion U Green Hot Start DNA Polymerase (Thermo scientific). We used 0.5 U of the Ung and 2 U of the Taq Polymerase in our reactions. The primers used to amplify the repeat tract from the GFP locus were oVIN-460 and oVIN-1425. The PCR program was as follow: 50 °C for 2 min and 5 min at 95 °C for the standard Ung, or 20 °C for 10 min followed by 7 minutes at 95 °C for hUng. Then there were 35 cycles at 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 1 min 30 s. Ung was neutralized by the addition of Proteinase K (Promega) immediately after the PCR run to a final concentration of 100 µg ml<sup>-1</sup> and incubated at 37 °C for 1 h. Sanger sequencing was done by Microsynth AG (Balgach, Switzerland).

**Molecular cloning.** All plasmids used are in Table 5. For cloning the SUMO1 cDNA, we used pDONR-SUMO1 as PCR template, primers M13for and M13rev, hUng, and Proteinase K to stop the reaction. The amplicons were run on an agarose gel, the band was cut out and purified using the Nucleospin Gel and PCR clean up kit from Macherey-Nagel. The isolated DNA was quantified using a Nanodrop spectrophotometer. 220 ng of this insert DNA was used with the Zero Blunt Topo kit (Thermo Fisher Scientific). We diluted the ligation products 4-fold and used 2 µl of the reaction to electroporate 40 µl of bacteria, which were then plated onto LB agar plates containing 50 µg ml<sup>-1</sup> of kanamycin. Colony PCR was done with the SP6 and T7 primers, sequencing of the resulting plasmids was done with SP6.

The degradation assays performed in Fig. 3 were done using 1 µg of plasmid DNA and 0.5U of Ung. They were incubated at 50 °C for 30 min in PCR buffer. The amplicons used as controls for the degradation were the SUMO1 cDNA amplified as to make inserts for cloning. The PCRs contained dUTPs. We used the full content of a 50 µl PCR per reaction.

For cloning of the expanded repeats, we used the same protocol as for SUMO1 cDNA cloning, but the primers were oVIN-1425 and oVIN-460 and the template was pVIN-117, which harbours the GFP reporter assay with an unstable repeat tract containing up to 89 repeats. The resulting plasmids were sequenced with oVIN-460. The sequences were analyzed blindly of the genotype of the cells and of the nucleotide present in the insert. Of note, we found that sequencing in the other direction led to poorer results presumably due to the higher stability of CTG-containing secondary structures versus CAG-containing ones<sup>24</sup>.

**SP-PCR.** The SP-PCR protocol was derived from<sup>25</sup>, using the primers and conditions used in<sup>16</sup> for dTTP-containing blots, with the exception that we used dUTP to generate the probe. A full step-by-step protocol is found in the supplementary material. The SP-PCR of the *DMPK* locus was done in a similar fashion with the

following changes. We used oVIN-1251 and oVIN-1252 for the amplification. The probe used was an oligo of 10 CAGs labelled with [ $\gamma$ - $^{32}$ P] ATP using the polynucleotide kinase from New England Biolabs. The temperature of hybridization was lowered to 52 °C.

**Statistics.** We used a two-tailed Student's t-test that assumes equal variance to compare cloning efficiencies. When comparing the distribution of repeat sizes in Table 1 and Fig. 4, we used a two-tailed Wilcoxon U-test since the repeat size distributions were not normal. Nevertheless, the same conclusions were obtained with a two-tailed Student's t-test. Statistical tests were performed using R version 3.1.3.

**Data availability statement.** The datasets and protocols generated during the current study are available from the corresponding author.

## References

- Riggins, G. J. *et al.* Human genes containing polymorphic trinucleotide repeats. *Nat Genet* **2**, 186–91 (1992).
- Usdin, K., House, N. C. & Freudenreich, C. H. Repeat instability during DNA repair: Insights from model systems. *Crit Rev Biochem Mol Biol* **50**, 142–67 (2015).
- Lee, D. Y. & McMurray, C. T. Trinucleotide expansion in disease: why is there a length threshold? *Curr Opin Genet Dev* **26**, 131–40 (2014).
- Lopez Castel, A., Cleary, J. D. & Pearson, C. E. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol* **11**, 165–70 (2010).
- Dion, V. Tissue specificity in DNA repair: lessons from trinucleotide repeat instability. *Trends Genet* **30**, 220–9 (2014).
- Jones, L., Houlden, H. & Tabrizi, S. J. DNA repair in the trinucleotide repeat disorders. *Lancet Neurol* **16**, 88–96 (2017).
- Budworth, H. *et al.* Suppression of Somatic Expansion Delays the Onset of Pathophysiology in a Mouse Model of Huntington's Disease. *PLoS Genet* **11**, e1005267 (2015).
- Kovalenko, M. *et al.* Msh2 acts in medium-spiny striatal neurons as an enhancer of CAG instability and mutant huntingtin phenotypes in Huntington's disease knock-in mice. *PLoS One* **7**, e44273 (2012).
- Wheeler, V. C. *et al.* Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum. *Hum Mol Genet* **12**, 273–81 (2003).
- Monckton, D. G., Wong, L. J., Ashizawa, T. & Caskey, C. T. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum Mol Genet* **4**, 1–8 (1995).
- Gomes-Pereira, M., Bidichandani, S. I. & Monckton, D. G. Analysis of unstable triplet repeats using small-pool polymerase chain reaction. *Methods Mol Biol* **277**, 61–76 (2004).
- Borst, A., Box, A. T. & Fluit, A. C. False-positive results and contamination in nucleic acid amplification assays: suggestions for a prevent and destroy strategy. *Eur J Clin Microbiol Infect Dis* **23**, 289–99 (2004).
- Pruvost, M., Grange, T. & Geigl, E. M. Minimizing DNA contamination by using UNG-coupled quantitative real-time PCR on degraded DNA samples: application to ancient DNA studies. *Biotechniques* **38**, 569–75 (2005).
- Greub, G., Sahli, R., Brouillet, R. & Jatou, K. Ten years of R&D and full automation in molecular diagnosis. *Future Microbiol* **11**, 403–25 (2016).
- Longo, M. C., Berninger, M. S. & Hartley, J. L. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene* **93**, 125–8 (1990).
- Cinesi, C., Aeschbach, L., Yang, B. & Dion, V. Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nat Commun* **7**, 13272 (2016).
- Thornton, C. G., Hartley, J. L. & Rashtchian, A. Utilizing uracil DNA glycosylase to control carryover contamination in PCR: characterization of residual UDG activity following thermal cycling. *Biotechniques* **13**, 180–4 (1992).
- Sobek, H., Schmidt, M., Frey, B. & Kaluza, K. Heat-labile uracil-DNA glycosylase: purification and characterization. *FEBS Lett* **388**, 1–4 (1996).
- Lari, S. U., Chen, C. Y., Vertessy, B. G., Morre, J. & Bennett, S. E. Quantitative determination of uracil residues in Escherichia coli DNA: Contribution of ung, dug, and dut genes to uracil avoidance. *DNA Repair (Amst)* **5**, 1407–20 (2006).
- Murphy, K. C. Targeted chromosomal gene knockout using PCR fragments. *Methods Mol Biol* **765**, 27–42 (2011).
- Yu, D. *et al.* An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci USA* **97**, 5978–83 (2000).
- Santillan, B. A., Moye, C., Mittelman, D. & Wilson, J. H. GFP-based fluorescence assay for CAG repeat instability in cultured human cells. *PLoS One* **9**, e113952 (2014).
- Murphy, K. C. & Campellone, K. G. Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic *E. coli*. *BMC Mol Biol* **4**, 11 (2003).
- Pearson, C. E., Ewel, A., Acharya, S., Fishel, R. A. & Sinden, R. R. Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum Mol Genet* **6**, 1117–23 (1997).
- Dion, V., Lin, Y., Hubert, L. Jr, Waterland, R. A. & Wilson, J. H. Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Hum Mol Genet* **17**, 1306–17 (2008).

## Acknowledgements

We would like to thank Onya Opota for the initial suggestion of using Ung and dUTP in our PCRs. We also thank him, Fisun Hamaratoglu, and members of the Dion lab for critical reading of the manuscript. We are grateful to Yashashvi Singh Bhugowon for technical help and for blinding the sequencing data. This work was supported by a professorship from the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (#144789 to V.D.) and by Gebert RUF Stiftung and UNISCIENTIA STIFTUNG within the programme «Rare Diseases – New Approaches» (GRS-060/14 to V.D.).

## Author Contributions

L.A. and V.D. designed the experiments; L.A. performed them. V.D. performed the statistical tests and wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-18168-2>.

**Competing Interests:** The authors declare that they have no competing interests.



**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017