Identification of DNA lesions using a third base pair for amplification and nanopore sequencing

Jan Riedl¹, Yun Ding¹, Aaron M. Fleming¹ & Cynthia J. Burrows¹

Damage to the genome is implicated in the progression of cancer and stress-induced diseases. DNA lesions exist in low levels, and cannot be amplified by standard PCR because they are frequently strong blocks to polymerases. Here, we describe a method for PCR amplification of lesion-containing DNA in which the site and identity could be marked, copied and sequenced. Critical for this method is installation of either the dNaM or d5SICS nucleotides at the lesion site after processing via the base excision repair process. These marker nucleotides constitute an unnatural base pair, allowing large quantities of marked DNA to be made by PCR amplification. Sanger sequencing confirms the potential for this method to locate lesions by marking, amplifying and sequencing a lesion in the KRAS gene. Detection using the α-hemolysin nanopore is also developed to analyse the markers in individual DNA strands with the potential to identify multiple lesions per strand.

¹Department of Chemistry, University of Utah, 315S 1400 East, Salt Lake City, Utah 84112-0850, USA. Correspondence and requests for materials should be addressed to C.J.B. (email: burrows@chem.utah.edu).
DNA undergoes damage caused by oxidation, deamination or alkylation leading to the formation of various base lesions, and depurination leading to abasic sites\(^1\)-\(^3\). The location in the genome in which the modification occurs is of critical interest, because it enables understanding the origin of genetic mutations resulting from these lesions. Mutations in the genome are one hallmark of melanoma, hepatic and lung carcinomas; mutations increase with age, and are observed in a number of stress-induced disorders, such as amyotrophic lateral sclerosis\(^4\),\(^5\). A method capable of identifying the chemical identity and location in which lesions appear is crucial for determining the molecular aetiology of these diseases. Moreover, recent research has highlighted lesions in close proximity to one another to be a challenge to the repair machinery and potentially more mutagenic\(^6\). However, due to the low abundance of DNA modifications in the genome, it is challenging to address these questions because the damage sites display both altered base pairing and frequently are pause or stop sites for polymerases, making them unamplifiable by PCR.

A number of methods have been developed to sequence epigenetic modifications, all of which rely on conversion of the modified base to a different, but readable base, such as the bisulfite conversion of cytosine to uracil, in contrast to the chemical stability of 5-methylcytosine\(^7\),\(^8\). Variations of this method have been developed to sequence 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxycytosine using a combination of enzymatic and chemical approaches in tandem with bisulfite sequencing\(^7\),\(^8\). Direct sequencing of epigenetic modifications has been demonstrated by single-molecule real-time sequencing (SMRT) and protein nanopores\(^9\)-\(^12\), though, both direct methods are challenged when working with actual tissue samples that contain modifications in low abundance.

In contrast to epigenetic modifications, DNA lesions such as those resulting from oxidative stress are diverse, and selective chemistry for them has not been developed. SMRT sequencing\(^13\) and ligation-mediated PCR provide limited advances in identifying base modifications\(^14\). An approach that retains the lesion location while providing a detectable signal for multiple lesions in proximity would enhance our understanding of lesions in the genome and how they contribute to mutagenesis.

An approach for mapping the precise location and identity of the lesion in DNA strands would be to label the damaged site with a marker nucleotide triphosphate during polymerase extension. The first generation of this approach attempted modified nucleotide insertion opposite an abasic site (AP)\(^13\)-\(^17\). These approaches suffered from the inability to extend past the site at which the marker was placed and were only applicable to abasic sites (AP sites). In addition, outcompeting insertion of dATP opposite an abasic site (that is, the ‘A rule’) with a modified nucleotide is challenging\(^18\). In a second generation for labelling lesion sites with a marker nucleotide, the Sturla laboratory demonstrated insertion and linear extension of a marker nucleotide opposite ω-benzylguanines\(^19\)-\(^21\). In this report, we developed a third generation approach for labelling lesion sites with an amplifiable marker nucleotide. This method utilizes the base excision repair (BER) pathway to target lesions to yield gaps in the DNA to only insert the marker nucleotide bypassing the ‘A rule’; these markers are exponentially PCR amplifiable; and lesion identity is determined by the BER enzymes avoiding lengthy synthesis of lesion-specific markers.

Critical for the development of our method is utilization of a marker nucleotide that has a selective, complementary partner allowing high-fidelity PCR amplification of the marked DNA. For this purpose, we chose the dNaM or dMMO2 nucleotides base paired with d5SICS that form an established set of unnatural base pairs (UBPs; Fig. 1)\(^22\)-\(^24\). These UBPs combined with the selectivity of the BER process allow insertion of the marker at the lesion site with excellent retention during PCR amplification, as described below.

Sequencing the UBPs is determined by a sharp stop in a Sanger sequencing chromatogram causing loss of all downstream information\(^22\). Therefore, to map more than one modification site per DNA strand, the dNaM nucleotide is also utilized to install an orthogonal functional group for attaching a detectable functionality\(^25\). The readout for this adduct is achieved using the ω-hemolsin (ω-HL) nanopore that analyses individual DNA molecules while they are electrophoretically driven through the small aperture pore\(^26\),\(^27\). Our concept is developed around common lesions found in the genome that include an abasic site (depurination lesion), 2′-deoxyuridine (dU, deamination lesion), as well as 8-oxo-7,8-dihydro-2′-deoxyguanosine (dOG) and spiroiminodihydantoin-2′-deoxynucleoside (dSp, oxidative lesions).

**Results**

**Outline of labelling methodology.** The methodology for labelling lesion sites in a DNA duplex was established in a section of the KRAS gene surrounding codon 12, and then applied for lesion detection in a plasmid. A G→T transversion in the coding strand of this gene in lung cancer is proposed to result from G oxidation\(^28\). In addition, a G→A transition found in codon 12 in this gene found in colon cancer is proposed to result from dC deamination to dU\(^4\). The 65-mer model system housed 30 nucleotides centred on codon 12 of the KRAS sequence flanked by two 17-mer PCR primer regions (Fig. 2). The non-lesion-containing strand was capped with 10 dT nucleotides terminated with triethylene glycol groups to prevent unwanted ligation reactions during one of the steps below due to the presence of a 5′-phosphate required for radioactive labelling; these tails also allowed gel purification of one strand from the other (Fig. 2). The labelling protocol is a one-pot, four-step reaction sequence that harnesses the enzymes found in the BER pathway for recognizing lesion sites and replacing them with a marker nucleotide, dNaM, d5SICS or dMMO2 (Fig. 1). The marked strand was exponentially PCR amplified with the appropriate

---

**Figure 1 | Unnatural base pairs used for labelling of DNA lesions.** The dNaM–d5SICS and dMMO2–d5SICS unnatural base pairs utilized for labelling DNA lesions with a third, amplifiable marker base pair.
Glycosylase a into the KRAS the KRAS codon 12 in the melanoma cancers. All duplexes studied included site-specific deamination product, dOG, an oxidation product, and dSp, dU (Fig. 3b) either bacterial formamidopyrimidine-DNA glycosylase (Fpg) or human OG glycosylase 1 (hOGG1) for processing by APE1 to furnish the desired single-nucleotide gap (AP site), a product of spontaneous depurination; dU, a lesion (Figs 3a and 4a). The complete details of each step follow.

Site selective removal of the lesion. The selectivity of lesion labelling was defined by this step (step I; Figs 3a and 4a). As examples, four types of lesions were selected for labelling with either the dNaM or d5SICS marker nucleotides: an abasic site (AP site), a product of spontaneous depurination; dU, a deamination product; dOG, an oxidation product, and dSp, a hyperoxidized guanine lesion. These lesions have been implicated in the induction of DNA mutations found in liver and melanoma cancers. All duplexes studied included site-specific lesions incorporated at known locations by solid-phase synthesis or by literature protocols (Supplementary Methods). The dU and AP site lesions were placed in the template strand of the KRAS sequence, while dOG and dSp were placed in the coding strand at the centre of codon 12. All duplexes studied included 25-mer poly-dC tails to help facilitate entry of this strand into the s-HL nanopore.

Formation of a gap site at the lesion for marker insertion. Monofunctional glycosylases, such as UDG, yield an AP site in an intact DNA strand as the product, requiring further processing by APE1 to furnish the desired single-nucleotide gap with a 3'-OH on the 5'-oligomer and a 5'-phosphate on the 2-deoxynucleotide triphosphates (dNTPs) to yield UBP amplicons that were sequenced to identify the location of the lesion (Figs 3a and 4a). The selectivity of lesion labelling was defined by this step (step I; Figs 3a and 4a). As examples, four types of lesions were selected for labelling with either the dNaM or d5SICS marker nucleotides: an abasic site (AP site), a product of spontaneous depurination; dU, a deamination product; dOG, an oxidation product, and dSp, a hyperoxidized guanine lesion. These lesions have been implicated in the induction of DNA mutations found in liver and melanoma cancers. All duplexes studied included site-specific lesions incorporated at known locations by solid-phase synthesis or by literature protocols (Supplementary Methods). The dU and AP site lesions were placed in the template strand of the KRAS sequence, while dOG and dSp were placed in the coding strand at the centre of codon 12. All duplexes studied included 25-mer poly-dC tails to help facilitate entry of this strand into the s-HL nanopore.

Figure 2 | DNA sequences studied based on codon 12 of the KRAS gene. The KRAS-U duplex contains a dU in the template strand at the centre of codon 12 in the KRAS gene. The KRAS-OG and KRAS-Sp duplexes contain dOG or dSp, respectively, in the coding strand at the centre of codon 12. The KRAS-U, KRAS-OG and KRAS-Sp duplexes are flanked by primer regions that allow their PCR amplification. The KRAS-S- strand centrally locates the KRAS sequence with a phosphorothioate at the centre of codon 12; this sequence also has 25-mer poly-dC tails to help facilitate entry of this strand into the s-HL nanopore.

Figure 3 | Labelling of lesions that are substrates for monofunctional BER enzymes. (a) The six-step scheme for labelling, amplifying and sequencing a lesion by Sanger sequencing with a template dNaM. (b) PAGE analysis to monitor and quantify enzyme-catalysed labelling reactions on the KRAS-U duplex with a dU:dG base pair (Fig. 2). Reactions were monitored by 32P-labelling of the lesion-containing strand, and the gels were quantified by phosphorimager autoradiography. Step 0 represents the starting strand before reaction. Steps III and IV were conducted with either dNaM or d5SICS and are marked with either a superscript N or 5, respectively. (c) The percentage of yield for each enzymatic step of the labelling scheme with a template dG.
dNaMTP or d5SICSTP allowed insertion of the marker in place of the lesion (step III; Figs 3a and 4a). Because only dNaMTP or d5SICSTP exist in the reaction without any other canonical nucleotides, they will be the only nucleotides inserted. Literature precedence has demonstrated that Klenow fragment of DNA polymerase I deficient in exonuclease activity (Kf exo−) has the potential to react at a gap site in a duplex, as well as being capable of inserting non-natural nucleotides41, hence, Kf exo− was selected for study of inserting either dNaMTP or d5SICSTP opposite a template dG. On the basis of the PAGE analysis (Figs 3b and 4b), Kf exo− with either of the marker triphosphates added to the reaction from step II could insert opposite a template dG to furnish the nicked duplex in a 1-h incubation with a nearly quantitative yield (Figs 3c and 4c). The efficiency of incorporating dNaMTP or d5SICSTP opposite dA, dC or dT were also evaluated and found to be >93% (Supplementary Table 1). This step was the first forcing reaction in the protocol, and the high yields observed were encouraging for the ultimate success of the labelling reaction.

Sealing the marker at the lesion site via ligation. The duplex product from step III has a nick with a non-native base pair on the 3′ terminus of the 5′ oligomer (Figs 3a and 4a). Previous studies from our laboratory demonstrated that this side of the nick is the most challenging to seal when a non-native base pair is present42; therefore, initial test reactions were set up to ligate either dNaM or d5SICS base paired with a dG on the 5′-side of the nick (step IV). This system allowed optimization of the ligation reaction to yield the desired duplex product. The best conditions found to seal the nick required the addition of dimethylsulphoxide (DMSO, 20% (v/v))44, decreased ATP (0.1 mM) and a large excess of T4-DNA ligase (800 units (U)) that were allowed to react for 16 h at 25 °C. These conditions gave a yield up to ~70%, and further refinement did not achieve higher yields. To complete these studies, tests were conducted with all combinations of dNaM or d5SICS paired with the canonical nucleotides to elucidate if there are any pairing combinations that might be better or worse for ligation (Fig. 2). These studies determined that dNaM or d5SICS paired opposite a dG at the nick site led to the highest ligation yields (~70%; Supplementary Table 2), and all other cases gave lower yields (~56–65%; Supplementary Table 2). Finally, a test of the ability to label multiple lesions with this approach was conducted with two dU lesions placed nine base pairs apart. PAGE analysis found the overall ligation yield to be ~45% (Supplementary Fig. 9). Because high yields (>90%) could not be achieved for the ligation step, an alternative approach was pursued to install the marker nucleotide (dNaM or d5SICS) at the gap resulting from BER activity on multiple lesions.

Polymerase extension to seal the marker nucleotides. Ligation at the site of the marker nucleotide was the most challenging step in our scheme with the lowest overall yield (56–72%; Supplementary Fig. 3). Because of the low abundance of lesions in the genome, it is preferred to maximize the yield in all steps. To bypass the ligation reaction, the nicked duplex bearing a marker nucleotide at the lesion site can be subjected to polymerase extension with a strand displacing polymerase, such as Kf exo− (Fig. 5a). The extension reaction will initiate from the marker nucleotide on the 3′-end of the priming strand (Fig. 5a). To conduct this extension, we simply added the four natural dNTPs after completion of step III and allowed Kf exo− to perform the extension reaction. One limitation to this approach is that any product from step III that does not have the marker nucleotide present will also lead to full length duplexes, and a method for bypassing this issue is

**Figure 4 | Labelling of lesions that are substrates for bifunctional BER enzymes.** (a) The six-step scheme for labelling, amplifying and sequencing a lesion. (b) PAGE analysis to monitor and quantify enzyme-catalysed labelling reactions on the KRAS-OG duplex with Fpg (Fig. 2). Reactions were monitored by 5′ 32P-labelling of the lesion-containing strand, and the gels were quantified by phosphorimagery autoradiography. Step 0 represents the starting strand before reaction. Steps III and IV were conducted with either dNaM or d5SICS, in which these differences are marked with either a superscript N or 5, respectively. (c) The percentage of yield for each enzymatic step of the labelling scheme with a template dG.
NATURE COMMUNICATIONS | DOI: 10.1038/ncomms9807

Figure 5 | Labelling of a lesion with a marker nucleotide utilizing polymerase extension. (a) The alternative scheme for labelling, amplifying and sequencing a lesion. (b) PAGE analysis to monitor and quantify enzyme-catalysed labelling reactions on the KRAS-U duplex with a U→G base pair (Fig. 2). Reactions were monitored by 5′-32P-labelling of the lesion-containing strand, and the gels were quantified by phosphorimager autoradiography. Step 0 represents the starting strand before reaction. Steps III and IV were conducted with either dNaM or d5SICS, in which the differences are marked with either a superscript N or 5, respectively. (c) The percentage of yield for each enzymatic step of the labelling scheme with a template dG.

Figure 6 | The dMMO2ssbioTP unnatural nucleotide utilized for affinity purification.  

Proposed below. The other limitation to this approach is that only one marker nucleotide per strand, and therefore one lesion per strand, can be detected because during the polymerase extension step any other marker present will be displaced by the polymerase. In spite of this limitation, polymerase extension using the strand with the marker nucleotide on the 3′-end was conducted. On the basis of PAGE analysis (Fig. 5b), the intact duplex was observed in high yield (90%), and the only impurity in this reaction was the presence of strands that failed to extend (10%; Fig. 5c).

PCR amplification of duplex DNA with a marker nucleotide. The sealed duplex with a dNaM or d5SICS marker nucleotide at the lesion site from step IV was PCR amplified. At this stage, the duplex product can directly be subjected to PCR amplification (Figs 3a and 4a); however, the strand that does not contain the marker nucleotide will be more favourably amplified at this step than the strand with the dNaM or d5SICS marker. For initial analysis of marker nucleotide incorporation, the two strands of the duplex were of different lengths allowing PAGE purification of the marker-containing strand away from the template strand (Fig. 2). Next, the marker-containing strand was subjected to 20 cycles of PCR following literature protocols22 to yield a population of amplicons (step V). Locating the position of the UBP within the amplified DNA was determined in the sequencing step (step VI; Figs 3a and 4a).

For biological samples, the strands will not have different lengths, and PAGE separation of the strands will not be possible. When applying this method, we envision three approaches for purification of the marker-containing strand. (1) Immuno precipitation of DNA strands containing the lesion of interest, followed by labelling and sequencing. This method is limited by the availability of immunoprecipitation grade antibodies for the lesion of interest. (2) Placement of an affinity purification tag on one of the marker nucleotides to allow purification such as the biotinylated dMMO2ssbioTP that can be substituted for dNaMTP as the UBP in this approach.25 Insertion of d5SICSTP in step III followed by amplification using d5SICSTP and biotinylated dMMO2ssbioTP (Fig. 6) provides the affinity purification tag25. As demonstrated below, utilization of the d5SICS-dMMO2ssbio UBP was an excellent approach for purification of the labelled strands. (3) Utilization of the phosphorothionucleotide triphosphates of dNaM or d5SICS, dNaMTP25 or d5SICS25TP, respectively, during step III (Figs 3a and 4a) would incorporate a functional group into duplex DNA that has orthogonal reactivity for purification purposes. Literature precedence has demonstrated reacting phosphorothioate-bearing DNA with iodoacetamide functionalized biotin for labelling to conduct affinity purification. Hence, the use of phosphorothionucleotide triphosphates of dNaM or d5SICS during step III would provide a third approach for applying our method to in vivo samples, as has been proposed by the Ramesberg laboratory.24 Therefore, we tested this proposal by synthesizing the phosphorothionucleotide triphosphates of dNaM or d5SICS and incorporated it into step III when Kf exonase inserted the marker at the gap generated by the lesion. PAGE analysis determined that dNaM25TP could be substituted for dNaMTP during this protocol (Supplementary Fig. 4).

Sequencing of the marker base pair in place of a DNA lesion. Once amplicons have been generated with the dNaM or d5SICS UBP located at the lesion site in the original duplex (Figs 3a and 4a), sequencing provides the readout of the base pair’s location. In the present iteration, the method for sequencing the UBP was conveniently achieved by the observation of an abrupt stop during Sanger sequencing45. Application of this method with either the dNaM or d5SICS nucleotide in the template strand correctly identified the location of the modification (Fig. 3a; Supplementary Fig. 5). As a test of the robustness of the method, the template containing either the dNaM or d5SICS nucleotide was subjected to 20 cycles of PCR before Sanger sequencing. The sequencing chromatograms demonstrated the ability to retain the marker nucleotides in the template strand; however, experiments did show a small level of impurity that was amplified during the PCR workup (Supplementary Fig. 5). These results verify the ability to mark and amplify at lesion sites in a homogenous solution of lesion-bearing oligomers, leading us to question how our method will perform with samples that model those found from biological sources.

Lesion sequencing in simulated biological contexts. Demonstration of this sequencing approach was then performed in a plasmid bearing a dOG nucleotide. Site-specific incorporation of an dOG nucleotide in a plasmid was achieved via an approach developed in the Wang laboratory46. The dOG was
Lesion sequencing with a large excess of background DNA. DNA samples from a biological source would have lesions in very low amounts. Next, a test of the feasibility of our approach to find and sequence a lesion with a large background of undamaged DNA was then pursued. Mixtures that contained a dU lesion strand with non-damaged strands in a ratio of 1:10, 1:100 and 1:1,000, respectively, were processed and sequenced to verify that DNA was then pursued. Mixtures that contained a dU lesion strand with non-damaged strands in a ratio of 1:10, 1:100 and 1:1,000, respectively, were processed and sequenced to verify that a large background of DNA does not interfere with the processing steps. Again, the dMMO2ssbioTP was used to affinity purify the labelled amplicons. The sequencing results illustrated that the method outlined was capable of correctly identifying the location of dOG (Fig. 7b). This approach verifies the ability to sequence a lesion within a much larger DNA context.

Detection of marker nucleotides by the z-HL nanopore. Although Sanger sequencing is a widely available method, the abrupt stop points by the polymerase encountering the third base pair will not be favourable for identification of more than one marker UBP that might exist in real samples. Therefore, we sought another approach for detecting the marker nucleotides that would have the potential to read more than one marker per strand. Our laboratory has an interest in studying DNA modifications by electrophoretically passing DNA strands through the z-HL protein nanopore embedded in a lipid bilayer. The z-HL nanopore analyses one DNA molecule at a time to give ion current versus time traces. Differences in the patterns of these traces allow the analysis of variations between populations of molecules. This feature will be advantageous when determining if more than one modification site exists in a strand of DNA, which cannot be readily achieved by current Sanger sequencing technology. Synthesis of a fifth and sixth dye pair and their incorporation into Sanger sequencing would advance this method; in contrast, the z-HL nanopore is a label-free technique actively being explored for DNA sequencing.

Free translocation of DNA through the nanopore under an electrical bias is too fast to read the individual nucleotide sequence without the use of molecular motors. To overcome this limitation, our laboratory has focused on the use of current-modulating adducts for detection of lesions in DNA. Previous studies identified 18-crown-6 to be an excellent current-modulating marker in NaCl electrolyte solutions that allowed detection of one to three of these adducts in a single strand of DNA, and we therefore adopted our previous method to the detection of the one and two marker nucleotide installed by the previous steps described herein.

Critical to our adaptation for detecting the dNaM or dSICS marker nucleotides is the ability to use their z-phosphorothio-nucleotide triphosphates for introduction of a site-specific sulfur atom that does not interfere with any of the other steps (Fig. 8a). As previously stated, the sulfur can be adducted specifically in high yield with an iodoacetamide. Therefore, we used the

Figure 7 | Detection of a dOG lesion in plasmid DNA. (a) The scheme comprising dOG labelling by unnatural base followed by amplification. (b) Sanger sequencing showing stop of the sequencing indicating the position of the dOG lesion.
dNaM previously reported to incorporate a group that can be site specifically reacted on. Alkylation of the phosphorothioate was achieved with the N-iodoacetamide of 2-aminomethyl-18-crown-6 (I-18-c-6), which we also synthesized via a method adapted from literature resources (Supplementary Methods)\textsuperscript{57}. On alkylation of the phosphorothioate with I-18-c-6, the strand could be analysed by the \( \alpha \)-HL nanopore to detect the ion current modulations diagnostic of the crown ether that also identifies the lesion that was in the original strand from step I (Figs 3a and 4a).

To develop this concept, the alkylation reactions had to be optimized. An 8-mer 5\textsuperscript{′}-ATGSCATGC-3\textsuperscript{′} was synthesized with one phosphorothioate (SC) in the backbone that allowed optimization to achieve high yield and characterization of the product resulting from alkylation with I-18-c-6 (Fig. 8b,c). Controls with strands that do not have a phosphorothioate failed to give any detectable product. These observations convinced us all signals observed in the \( \alpha \)-HL experiment will result from adduct formation at the phosphorothioate.

The \( \alpha \)-HL nanopore platform was then used to analyse the original KRAS sequence with one and two phosphorothioates installed at the lesion sites (\( \text{KRAS-S}^\text{−} \) and \( \text{KRAS-2S}^\text{−} \) (Fig. 2); Supplementary Fig. 6). This strand flanked the KRAS sequence with 25-mer tails of poly-dC to enhance entry of this strand into the \( \alpha \)-HL nanopore. After alkylation of this strand with I-18-c-6, it was subjected to analysis by the \( \alpha \)-HL nanopore using 3 M NaCl as the electrolyte solution. Entry of the strand into the \( \alpha \)-HL channel causes the current to decrease from the open channel value (\( I_0 \)) to a deep blockage level (11\% \( I_0 \)) when the DNA spans the channel (Fig. 8d). When the 18-c-6 is driven through the \( \beta \)-barrel, the current recorded becomes more deeply blocked (5\% \( I_0 \)), diagnostic of the 18-c-6 traversing the narrowest constriction of the pore and signalling the presence of the marker.

Figure 8 | New method for detection of marker nucleotides using the \( \alpha \)-HL nanopore. (a) PCR amplification with dNaM\textsuperscript{25}TP and labelling of DNA containing phosphorothioate by an iodoacetamide. (b) Reaction for phosphorothioate-containing DNA labelling by I-18-c-6. (c) Gel-shift analysis of labelled 8-mer by I-18-c-6 to confirm the reaction yield. (d) Translocation of DNA (\( \text{KRAS-S}^\text{−} \)) labelled by I-18-c-6 through the \( \alpha \)-HL nanopore providing a modulation in the deep-blockage current level observed as the 11\% \( I_0 \) signal decreasing to 5\% \( I_0 \) that signals the presence of the crown ether and the marker nucleotide (that is, lesion). Nanopore measurements for one 18-c-6-labelled site were conducted at 140 mV bias and for two 18-c-6 labels were conducted at 160 mV (trans versus cis) bias.
nucleotide (Fig. 8d; Supplementary Fig. 6). From this study, the modulation of the deep-blockage ion current level was similar to the blockages we previously recorded when 18-c-6 methylamine was added to an AP site in a DNA strand. In the last study, two 18-c-6 added phosphorothioate sites were installed in a strand followed by α-HL nanopore analysis. In this analysis, two current modulations were observed (Fig. 8d; Supplementary Fig. 6), reproducing the previously published results on a closely related adduct. The present data confirmed that the α-HL nanopore can be an alternative platform for detecting the presence of one or two marker nucleotides installed by this method at lesion sites in DNA strands. The emerging power of other nanopore systems to rapidly sequence DNA provide additional avenues to explore direct sequencing of these UBP5,14. Finally, detection of more than one dNaM or d5SICS by Sanger sequencing is not currently achievable; thus, it is anticipated that detection of the dNaM through alkylation of its phosphorothioate with 18-c-6 by α-HL will provide a platform for counting more than one of these marker nucleotides per DNA strand.

Discussion

Herein, a method was developed to label a DNA lesion with the dNaM-d5SICS UBP that marks the lesion’s location in the primary sequence of DNA. To achieve this goal, a four-step reaction sequence was developed followed by PCR amplification and sequencing (Figs 3a and 4a). The overall reaction yield varied between 50 and 65% (Figs 3c and 4c). A low reaction yield was observed when ligation was used to seal a nicked site in the DNA duplex with dNaM or d5SICS. The ligation yield varied with the canonical base paired with dNaM or d5SICS (Supplementary Fig. 3). This approach will label more than one lesion per strand, as long as they are not too close to prevent the reactions in steps I, II, III and IV. As an alternative, polymerase extension past the incorporated dNaM or d5SICS could label a single lesion site in DNA in a nearly 90% overall yield (Fig. 5c). This alternative approach provides a much needed increase in yield that is required when dealing with the low levels of lesions in the genome. The excellent retention and amplification of the dNaM-d5SICS UBP provides the ability to conduct PCR that effectively amplifies amplification of lesion-containing DNA. This step is critical in obtaining enough amplicons for performing DNA sequencing, particularly Sanger sequencing. While Sanger sequencing is only capable of identifying one dNaM or d5SICS UBP per duplex, we demonstrate the use of chemical labels in tandem with the single-molecule profiling capability of the α-HL nanopore to provide the ability to observe multiple markers per strand. This straightforward method can be applied to DNA samples leading to lesion labelling in <24 h when ligation of the marker nucleotide is the approach taken; alternatively, if polymerase extension is used for sealing the marker nucleotide into the DNA, the process can be achieved in about 8 h. This method was developed with commercially available enzymes and instruments; while advantageous, it also limits the ability to mark lesion sites in one strand in quantitative yield, as well as to sequence the markers by single-molecule profiling at single-nucleotide resolution. Nevertheless, future engineering of ligases that efficiently accept UBP at the nick site would provide a much needed boost in the yields for labelling of multiple modifications. Furthermore, nanopore sequencing of genomic data has made enormous strides, and adaptation of the present labelling and PCR amplification method with these sequences will likely be the approach applied for routinely sequencing lesions from biological sources.

The genome is under constant threat by the inherent reactivity of DNA bases and by chemical insults to which they are exposed. These chemical changes lead to mutations, in which the original chemical modification is rarely known, but rather implied based on the mutagenic profile. To address this knowledge gap, we developed a method that enables selective replacement of four types of lesions—uracil, 8-oxo-7,8-dihydroguanine, spiroiminodihydantoin and abasic sites by the dNaM-d5SICS UBP in a one-pot, four-step reaction sequence (Figs 3a and 4a). The elegance of this system lies in the ability to PCR amplify the third base pair, which effectively allows amplification of DNA lesions chosen selectively by BER enzymes for replacement with the marker nucleotides. Moreover, this simple method can easily be extended to other types of lesions or modifications to the genome, as long as a BER enzyme exists to conduct step I of the method as outlined (Figs 3a or 4a); for example, lesions/modifications with known glycosylases include thymine glycol (NTHL1), 3-methyladenine (ALK A), the T in a T-G mismatch (TDG) and 5-methylcytosine (ROS 1), all of which would also be exciting avenues for study by this method. Furthermore, this method avoids competition with native nucleotides to maximize labelling yields without interfering side reactions involving insertion of native nucleotides (for example, due to the ‘A rule’). The sequencing protocol for the dNaM-d5SICS UBP is demonstrated by Sanger sequencing, leading to an abrupt stop in the sequence chromatogram for identification of the base pair’s location (Fig. 3a). Accordingly, we further developed chemistry around the dNaM-d5SICS UBP that will permit detection of more than one lesion per strand and allow population analysis of lesion location by the α-HL nanopore method (Fig. 8d). The labelling methodology developed can be applied to questions surrounding the underlying chemical reactions that initiate mutations in the genome leading to disease. For example, mutations in codon 12 of the KRAS gene can be probed to determine if dOG or dU are responsible for the mutations observed in lung or colon cancer, respectively. This information will be paramount in preventative medicine for cancers and other deleterious stress-induced diseases resulting from these mutations. The new α-HL nanopore sequencing methodology developed has the potential to become a robust approach for detecting and sequencing non-native nucleotides that is currently only achievable by SMRT sequencing.

Methods

Materials. UDG, Fpg, hOGG1, endonuclease IV, T4-polymerase kinase, Klenow(exo- ) DNA polymerase, T4-DNA ligase and OneTaq DNA polymerase were purchased from New England Biolabs. APE 1 (ref. 60) and hNEILL-DNA glycosylase were a kind gift from Professor Sheila S. David. dNaM and d5SICS UBP were a kind gift from Professor Rosalind A. Fisch Factor (NTHL1). Oligonucleotides were synthesized by previously published procedures. Other chemicals were of molecular biology grade.

DNA preparation and purification procedures. DNA was prepared from commercially available phosphoramidites (Glen Research, VA) by the DNA/Peptide Core Facility at the University of Utah. The DNA was cleaved and deprotected following the manufacturer’s protocol, followed by purification using an ion-exchange high-performance liquid chromatography (HPLC) column running a linear gradient of B from 20 to 100% over 30 min while monitoring ultraviolet absorbance at 260 nm (A = 20 mM NaPi, 1 M NaCl, pH 7 in 10% CH3CN/90% ddH2O; B = 10% CH3CN/90% ddH2O, flow rate = 3 ml min-1). Oligonucleotides were dialysed and water was evaporated. The oligonucleotide concentrations were determined by ultraviolet-visible spectroscopy using the primary sequence to determine the extinction coefficients.

Site-specific replacement of an uracil lesion. Reaction mixture (30 µl) containing KRAS-U (10 pmol), reaction buffer (25 mM HEPES, 10 mM MgCl2, 5 mM KCl, 1 mM DTT, 1 mM EDTA) and UDG (1 U) was incubated at 37 °C for 30 min. Then AP endonuclease (1 µl, 150 mU) was added to the reaction mixture, incubated at 37 °C for another 1 h and heated to 95 °C for 10 min. Next, dNaMTP or d5SICSTP (3 µl, 500 µM) and Klenow(exo- ) DNA polymerase (7 U) were added to the reaction mixture and heated at 37 °C for 1 h. The reactions were quenched by heating to 95 °C for 10 min. Finally, 5 µl of DMSO, 1 µl of 3 mM ATP and 2 µl (800 U) of DNA ligase were added to the reaction and kept at 25 °C for 16 h.

© 2015 Macmillan Publishers Limited. All rights reserved.
For the extension assay, a mixture of natural dNTPs was added in step IV, and the template strand without poly T tails and triethylene glycol capping was used.

Site-specific replacement of uracil lesions, nine bases apart. A reaction mixture (50 μl) containing KRAS-U9 (10 pmol), reaction buffer (25 mM HEPES, 10 mM MgCl2, 5 mM KCl, 1 mM DTT, 1 mM EDTA) and UDG (1 U) was added to the reaction mixture, incubated at 37 °C for 30 min. Then AP endonuclease (1 μl, 150 mM solution) was added to the reaction mixture, incubated at 37 °C for another 1 h and heated to 95 °C for 10 min. Next, dNAMTP or dSSCSTP (3 μl, 500 μM) and Klenow(exo−) DNA polymerase (7 μl) were added to the reaction mixture and heated at 37 °C for 1 h. The reactions were quenched by heating to 95 °C for 10 min. Finally, 5 μl of DMSO, 1 μl of 3 mM ATP and 2 μl (800 U) of DNA ligase were added to the reaction and kept at 25 °C for 16 h.

Dilution study of labelling uracil lesion at different ratios. Reaction mixture (50 μl) containing duplex KRAS sequence without lesion (20 μg), KRAS-U (2 μg, 200 ng and 20 ng) reaction buffer (25 mM HEPES, 10 mM MgCl2, 5 mM KCl, 1 mM DTT and 1 mM EDTA) and UDG (1 U) was added to the reaction mixture, incubated at 37 °C for 1 h and heated to 95 °C for 10 min. Next, dNAMTP or dSSCSTP (3 μl, 500 μM) and Klenow(exo−) DNA polymerase (7 μl) were added to the reaction mixture and heated at 37 °C for 1 h. The reactions were quenched by heating to 95 °C for 10 min. Finally, 5 μl of DMSO, 1 μl of 3 mM ATP and 2 μl (800 U) of DNA ligase were added to the reaction and kept at 25 °C for 16 h.

Site-specific replacement of 8-oxoguanosine lesion. The synthesis of KRAS-Sp was achieved by two protocols and purified via the HPLC method outlined above45. Reaction mixture (50 μl) containing KRAS-Sp (1 pmol), reaction buffer (20 mM Tris, 10 mM MgCl2, 5 mM KCl, 1 mM DTT and 1 mM EDTA) and Fpg-DNA glycosylase (40 U) or hOGG1-DNA glycosylase (20 U) were incubated at 37 °C for 2 or 3 h, respectively. Then endonuclease IV (20 U) was combined with T4-polymerase kinase (1 U) and added to the reaction mixture followed by incubation at 37 °C for 2 h followed by heating to 95 °C for 10 min. Next, dNAMTP or dSSCSTP (3 μl, 500 μM) and Klenow(exo−) DNA polymerase (7 μl) were added to the reaction mixture and heated at 37 °C for 1 h. The reactions were quenched by heating to 95 °C for 10 min. DMSO (5 μl), ATP (1 μl, 3 mM) and DNA ligase (2 μl, 800 U) were added to the reaction and kept at 25 °C for 16 h.

Site-specific conversion of spiroiminodihydrantinolone synthesis. Synthesis of KRAS-sp was achieved by two protocols and purified via the HPLC method outlined above45. Reaction mixture (50 μl) containing KRAS-Sp (1 pmol), reaction buffer (200 mM Tris, 10 mM MgCl2, 5 mM KCl, 1 mM DTT and 1 mM EDTA) and hNEIL1-DNA glycosylase (100 μM) were incubated at 25 °C for 2 h. Then endonuclease IV (20 U) was combined with T4-polymerase kinase (1 U) and added to the reaction mixture, incubated at 37 °C for another 2 h and heated to 95 °C for 10 min.

Ion channel measurements. The glass nanopore membrane (GNM; with radius 800 nm) was fabricated as previously reported85. 1,2-Diphtyrylso-sn-glycero-3- phosphocholine (DPhPC) bilayers (DPHPC) across the orifice of the GNM were prepared by immersing a 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) bilayer into a solution of DPhPC in decane to a concentration of 100 μg ml−1 under denaturing conditions (27 mA, 3 h) using TRIS-borate-EDTA (TBE) electrolyte buffer.

32P-labelling. To the oligonucleotide (6 pmol) sample, 10 μl of polymerase kinase buffer (10 μl), γ-32P ATP (15 μCi) and T4-polymerase kinase (0.5 μl) was added, and deionized water to a total volume of 100 μl. The mixture was heated at 73 °C for 1 h, and inactivated at 95 °C for 10 min.

Insertion of VEGF promoter sequence to plasmid DNA. The plasmid pBR322 (1 μg) was subjected to restriction-free cloning using 5′-GGCCAGTTGTTCACCTCACAAGATGGCGGGGCGGGGGGGTGATCCACTACACT-3′, 5′-CATTGCATTGCGAGTTACAG-3′ and 5′-AACATGCCCGTACTG-3′. The plasmid DNA was purified with an UltraClean PCR clean-up kit from MO BIO laboratories, Inc. The purified plasmid DNA was transformed to competent Esherichia coli and isolated with a Qiagen plasmid max kit.
Detection of an dOG in plasmid DNA. The reaction mixture (10 μl) containing the pBR322 plasmid with an OG lesion (30 ng), reaction buffer (25 mM HEPES, 10 mM MgCl₂, 5 mM KCl, and 1 mM EDTA) and Fpg (5 U), endonuclease IV (1 U) was incubated at 37 °C for 30 min, followed by addition of dSSCSTP (3 μl, 500 μM). Finally, ATP (0.5 μl of 2 mM) and 1.5 μl (600 U) of DNA ligase were added to the reaction and kept at 25 °C for 2 h. The processed pBR322 plasmid (2 μl from the previous reaction) was PCR amplified using forward primer: 5'-CCCTGAGATTGTCTACAGTGCGGGC-3' (1 μM), and reverse primer: 5'-ACGGAGAGGAGATCTCGAGATTGGTACGTTGAC-3' (1 μM), natural dNTPs (100 μM), dSSCSTP (100 μM), dMMOSS8BTP (100 μM) and OneTAq DNA polymerase (2 U) in a 20-μl reaction. The PCR procedure consisted of initial denaturation at 95 °C for 2 min, followed by 20 cycles of PCR. Each cycle consisted of 95 °C for 1 min, 55 °C for 30 s and 72 °C for 2 min, and a final extension for 5 min 30 s to complete the PCR. The PCR products were purified by 2% agarose gel electrophoresis using TAE buffer for 45 min.

References

1. Lonkar, P. & Dedon, P. C. Reactive species and DNA damage in chronic inflammation: reconciling chemical mechanisms and biological fates. Int. J. Cancer 128, 1999–2009 (2011).

2. Gates, K. An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alklylation, and reactions with radicals. Chem. Res. Textil. 22, 1747–1770 (2009).

3. Cadet, J., Wagner, J. R., Shatirovich, V. & Geacintov, N. E. One-electron oxidation reactions of purine and pyrimidine bases in cellular DNA. Int. J. Radiat. Biol. 90, 423–432 (2014).

4. Burmer, G. C. & Loeb, L. A. Mutations in the KRAS2 oncogene during progressive stages of human colon carcinoma. Proc. Natl Acad. Sci. USA 86, 2407–2410 (1989).

5. Irani, K. et al. Mitogenic signaling mediated by oxidants in Ras-transformed fibroblasts. Science 275, 1649–1652 (1997).

6. Cunniffe, S., O’Neill, P., Greenberg, M. M. & Lomax, M. E. Reduced repair capacity of a DNA clustered damage site comprised of 8-oxo-7,8-dihydro-2'-deoxyguanosine and a thymine lesion. Proc. Natl Acad. Sci. USA 106, 19953–19958 (2009).

7. Laszlo, A. H. et al. DNA repair and transcription in a human cell line with a G54S mutation in the TP53 tumor suppressor gene. Proc. Natl Acad. Sci. USA 91, 3282–3286 (1994).

8. Cao, C., Jiang, Y. L., Stivers, J. T. & Song, F. Dynamic opening of DNA during transcription. Nat. Struct. Mol. Biol. 16, 374–378 (2009).
45. Malyshev, D. A. et al. A semi-synthetic organism with an expanded genetic alphabet. *Nature* **509**, 385–388 (2014).

46. You, C. et al. A quantitative assay for assessing the effects of DNA lesions on transcription. *Nat. Chem. Biol.* **8**, 817–822 (2012).

47. Fleming, A. M., Zhou, J., Wallace, S. S. & Burrows, C. J. A role for the fifth G-track in G-quadruplex forming oncogene promoter sequences during oxidative stress: Do these “spare tires” have an evolved function? *ACS Cent. Sci.* **1**, 226–233 (2015).

48. Schibel, A. E. P. et al. Nanopore detection of 8-oxo-7,8-dihydro-2'-deoxyguanosine in immobilized single-stranded DNA via adduct formation to the DNA damage site. *J. Am. Chem. Soc.* **132**, 17992–17995 (2010).

49. An, N., White, H. S. & Burrows, C. J. Modulation of the current signatures of DNA basic site adducts in the alpha-hemolysin ion channel. *Chem. Commun.* **48**, 11410–11412 (2012).

50. Jin, Q., Fleming, A. M., Ding, Y., Burrows, C. J. & White, H. S. Structural destabilization of DNA duplexes containing single-base lesions investigated by nanopore measurements. *Biochemistry* **52**, 7870–7877 (2013).

51. Schibel, A. et al. Sequence-specific single-molecule analysis of 8-oxo-7,8-dihydroguanine lesions in DNA based on unzipping kinetics of complementary probes in ion channel recordings. *J. Am. Chem. Soc.* **133**, 14778–14784 (2011).

52. Johnson, R. P., Fleming, A. M., Jin, Q., Burrows, C. J. & White, H. S. Temperature and electrolyte optimization of the α-hemolysin latch sensing zone for detection of base modification in double-stranded DNA. *Biophys. J.* **107**, 924–931 (2014).

53. Wanunu, M. Nanopores: a journey towards DNA sequencing. *Phys. Life Rev.* **9**, 125–138 (2012).

54. Laszlo, A. H. et al. Decoding long nanopore sequencing reads of natural DNA. *Nat. Biotechnol.* **32**, 829–833 (2014).

55. Wolna, A. H. et al. Electrical current signatures of DNA base modifications in single molecules immobilized in the α-hemolysin ion channel. *Isr. J. Chem.* **53**, 417–430 (2013).

56. An, N., Fleming, A. M., White, H. S. & Burrows, C. J. Nanopore detection of 8-oxoguanine in the human telomere repeat sequence. *ACS Nano* **9**, 4296–4307 (2015).

57. Qiao, X. et al. Imidazolium-based iodoacetamide functional tags: design, synthesis, and property study for cysteinyll-peptide analysis by mass spectrometry. *Analyst* **139**, 705–708 (2014).

58. Jain, M. et al. Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* **12**, 351–356 (2015).

59. Zhu, J.-K. Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* **43**, 143–166 (2009).

60. Krishnamurthy, N., Zhao, X., Burrows, C. J. & David, S. S. Superior removal of hydantoin lesions relative to other oxidized bases by the human DNA glycosylase hNEIL1. *Biochemistry* **47**, 7137–7146 (2008).

61. Bandaru, V., Sunkara, S., Wallace, S. S. & Bond, J. P. A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease VIII. *DNA Repair (Amst)* **1**, 517–529 (2002).

62. Leconte, A. M. et al. Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *J. Am. Chem. Soc.* **130**, 2336–2343 (2008).

63. Kornyushyna, O., Berges, A. M., Muller, J. G. & Burrows, C. J. *In vitro* nucleotide misinsertion opposite the oxidized guanosine lesions spiroiminodihydantoin and guanidinoxyhdantoin and DNA synthesis past the lesions using *Escherichia coli* DNA polymerase I (Klenow fragment). *Biochemistry* **41**, 15304–15314 (2002).

64. Zhang, B. et al. Bench-top method for fabricating glass-sealed nanodisk electrodes, glass nanopore electrodes, and glass nanopore membranes of controlled size. *Anal. Chem.* **79**, 4778–4787 (2007).

65. White, R. J. et al. Single ion-channel recordings using glass nanopore membranes. *J. Am. Chem. Soc.* **129**, 11766–11775 (2007).

**Acknowledgements**

This work was supported by a research grant from the National Institutes of Health (R01 GM093099). The oligonucleotides were provided by the DNA/Peptide Core Facility at the University of Utah that is supported in part by an NCI Cancer Center Support Grant (P30 CA042014). The dNaMTP, d5SICSTP and dMMO2ssbioTP samples were a generous gift from Professor Floyd Romesberg, Dr Denis Malyshev, Dr Thomas Langerman and Mr Aaron Feldman at The Scripps Research Institute, La Jolla, CA. The APE1 and hNEIL1 enzymes were a gift from Professor Sheila David at the University of California, Davis. Instruments and software were donated by Electronic BioSciences, San Diego, CA, and are gratefully acknowledged.

**Author contributions**

J.R., A.M.F. and C.J.B. designed the research, and wrote the manuscript with contributions from all authors. J.R. conducted the enzyme reactions, synthesis and Sanger sequencing. Y.D. conducted the α-HI nanopore experiments. J.R. and Y.D. analysed the data.

**Additional information**

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Riedl, J. et al. Identification of DNA lesions using a third base pair for amplification and nanopore sequencing. *Nat. Commun.* **6**, 8807 doi: 10.1038/ncomms9807 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/