Use of uracil-DNA glycosylase enzyme to reduce DNA-related artifacts from formalin-fixed and paraffin-embedded tissues in diagnostic routine

C. M. Berra†, G. T. Torrezan†, C. A. de Paula†, R. Hsieh‡, S. V. Lourenço‡ and D. M. Carraro†*

Abstract

Background: Detection of somatic mutations is a mandatory practice for therapeutic definition in precision oncology. However, somatic mutation detection protocols use DNA from formalin-fixed and paraffin-embedded (FFPE) tumor tissues, which can result in detection of nonreproducible sequence artifacts, especially C:G > T:A transitions, in DNA. In recent studies, DNA pretreatment with uracil DNA glycosylase (UDG), an enzyme involved in base excision repair, significantly reduced the number of DNA artifacts after mutation detection by next-generation sequencing (NGS) and other methods, without affecting the capacity to detect real mutations. This study aimed to evaluate the effects of UDG enzymatic pretreatment in reducing the number of DNA sequencing artifacts from FFPE tumor samples, to improve the accuracy of genetic testing in the molecular diagnostic routine.

Methods: We selected 12 FFPE tumor samples (10 melanoma, 1 lung, and 1 colorectal tumor sample) with different storage times. We compared sequencing results of a 16-hotspot gene panel of NGS libraries prepared with UDG-treated and untreated samples.

Results: All UDG-treated samples showed large reductions in the total number of transitions (medium reduction of 80%) and the transition/transversion ratio (medium reduction of 75%). In addition, most sequence artifacts presented a low variant allele frequency (VAF < 10%) which are eliminated with UDG treatment.

Conclusion: Including UDG enzymatic treatment before multiplex amplification in the NGS workflow significantly decreased the number of artifactual variants detected in FFPE samples. Thus, including this additional step in the current methodology should improve the rate of true mutation detection in the molecular diagnostic routine.

Keywords: Next-generation sequencing, Diagnostic routine, FFPE, Uracil DNA glycosylase

Background

Next-generation sequencing (NGS) is the term used to describe numerous modern sequencing technologies that have been advancing genomic progress and revolutionizing biomedical research and clinical practice [1]. In oncology, a genetic variant can be used as a biomarker when it influences disease diagnosis or prognosis and predicts sensitivity to specific treatments [2]. Correct and precise mutation detection is an essential step in precision medicine [3]. Although the benefits of NGS are unquestionable, the complexities of these methodologies and the required DNA integrity demand an evolving set of standards to ensure testing quality [4].

Use of formalin fixation and paraffin embedding of solid tumor tissues is an obligatory approach in clinical pathology because this method preserves the cellular morphology (a necessary condition for further pathological analyses) and allows long-term room-temperature sample storage. DNA from formalin-fixed and paraffin-embedded (FFPE) tumor tissues is frequently used when detecting somatic mutations in defining the tumor profile and...
selecting the proper target therapy in cancer treatment [5]. However, such DNA presents extensive DNA degradation, which reduces the efficiency of polymerase chain reaction (PCR) amplification of DNA templates for further sequencing. Moreover, DNA from FFPE samples presents sequence artifacts that may interfere with detection of true gene mutations and increase the possibility of false-positive mutation calls [6, 7].

Sequence artifacts can arise from several types of DNA damage generated in FFPE tissues, including deamination of cytosine to uracil or deamination of 5-methylcytosine (5-mC) to thymine [7]. Hydrolytic deamination of cytosine or 5-mC in DNA, which contributes significantly to the appearance of spontaneous mutations, occurs spontaneously in an aqueous environment and in response to temperature changes [8]. In living cells, uracil lesions in DNA are removed by uracil-DNA glycosylase (UDG/UNG). G-mispaired base lesions, including thymine, are repaired by either methyl-binding domain protein 4 (MBD4) or thymine-DNA glycosylase (TDG), enzymes involved in the base excision repair (BER) pathway [9]. However, in an artificial context, where there are no repair enzymes and lesions are not repaired, deamination of cytosine or 5-mC elicits guanine mispairing, which causes DNA polymerase to incorporate an adenine opposite to both uracil and thymine, causing C:G > T:A transitions [6].

To deal with this mechanism of sequence artifact formation, which could represent an obstacle for mutational analysis and precision medicine, the use of UDG enzymatic treatment before PCR amplification in the NGS routine has been proposed. Studies of UDG treatment have shown significant reductions in the frequency of C:G > T:A artifactual mutations, with no effects on the ability of NGS to detect real mutations [3, 6, 10, 11]. Therefore, the aim of this work was to validate whether pretreatment with commercial UDG enzyme could reduce the number of sequencing artifacts of DNA extracted from FFPE tumors, to improve the genetic testing used in the molecular diagnostic routine.

**Methods**

**Sample preparation**

Twelve FFPE tissue samples were used in this study, derived as follows: 10 samples from patients with melanoma, 1 sample from a patient with lung adenocarcinoma, and 1 sample from a patient with colorectal cancer. FFPE tissues were subjected to histological analysis by a pathologist to assess the percentage of tumor cells and mark tumor areas. Manual dissection of tumor regions was performed in unstained slides. Genomic DNA was extracted by using the QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany), according to the manufacturer’s instructions. DNA quantity and quality were assessed by using Nanodrop 1000 and Qubit dsDNA HS kit (Thermo Fisher Scientific, Wilmington, DE, USA).

**Gene mutation analysis**

Tumor somatic mutations were investigated by target-sequencing by using a custom Ion Ampliseq™ Panel (Thermo Fisher Scientific), containing hotspot regions of 16 genes frequently mutated in solid tumors (BRAF, CDH1, EGFR, ERBB2, HRAS, IDH1, IDH2, JAK2, KIT, KRAS, MET, NRAS, PDGFRα, PIK3CA, RET, and ROS1). Multiplex amplification was performed with 10 ng of DNA using Ion AmpliSeq Library Kit 2.0, with NGS performed with the Ion Proton platform (Thermo Fisher Scientific). To compare the effects of UDG treatment before multiplex amplification, 15 ng of DNA was treated with 0.5 μL (1 unit/μL) of a commercial enzyme (Thermo Fisher Scientific) for 30 min at 37 °C and inactivated for 5 min at 95 °C.

Mapping of sequencing reads and variant calling were performed by using Torrent Suite Browser and TVC (Thermo Fisher Scientific). Somatic mutations were considered if the variant allele was present in more than 2% of the reads, considering a minimum coverage depth of 100×. Called variants were imported, annotated, and filtered in VarSeq software (Golden Helix). The transition/transversion ratio (Ts/Tv) was calculated by considering all identified variants.

**Statistics**

Descriptive statistics was used to describe the absolute and relative numbers of each mutation type (transitions and transversions) and the percentage of reduction after UDG treatment. The comparison of the number of transitions and transversions and Ts/Tv ratio between treated and untreated samples was done by using the paired t-test (and nonparametric tests) with 95% confidence intervals in GraphPad V5 software. Variant allele frequency (VAF) values of detected variants were compared between treated and untreated samples using the unpaired t-test. Results were considered statistically significant if the p-value was < 0.05.

**Results**

We performed NGS sequencing of a 16-gene panel using UDG-treated and untreated DNA obtained from 12 FFPE tumors samples previously known to harbor high levels of transitions in routine mutation analysis, to evaluate the capacity of reducing sequencing artifacts with UDG treatment. The medium number of variants was 82 for untreated samples (range: 2–338) and 12 for treated samples (range: 3–37) (Table 1).

Samples showed significantly fewer transitions after UDG treatment (mean 18.8 ± 10.3) when compared with no enzyme treatment (mean 157.7 ± 99.1, p < 0.0007).
with an average reduction of 80% (range: 21–95%) (Table 1). UDG treatment had no consistent effect on the number of transversion variants \( (p = 0.3774) \), with an average reduction of 5.5%, as some samples presented extra variants and others showed fewer variants after UDG treatment (Table 1). This effect probably occurred due to other sequencing-associated factors that are unrelated to cytosine deamination. Tvs/Ts ratio were higher in untreated compared to UDG-treated samples \( (p < 0.0007) \) and the decrease in the Tvs/Ts ratio varied from 2 to 95% (mean 75%) (Table 1). One sample (A43) presented a very small, very divergent reduction of the Tvs/Ts ratio compared to other samples (2%). When this value was removed from the analysis, the average Tvs/Ts ratio was 83% (data not shown). For one sample (A20), multiplex amplification after UDG treatment failed completely, and sequencing did not result in any mapped reads.

To visualize differences between true and artifactual variants, we aligned and visually inspected sequenced reads from untreated and UDG-treated samples. Figure 2 shows results of sequencing alignments of the KRAS gene from two representative samples, A11 and K2568. A11 is a melanoma sample in which a KRAS c.38G > A \( (p\text{-Gly13Asp}) \) variant was detected with 8.8% VAF before UDG treatment. K2568 is a colorectal tumor sample in which a KRAS c.35G > A \( (p\text{-Gly12Asp}) \) variant was detected with 13.0% VAF before treatment. After UDG treatment, only the mutation detected in sample K2568 remained as a result of a true variant; UDG removed all mis-incorporated thymine in sample A11, confirming it as an artifactual variant.

### Discussion
Use of NGS techniques has revolutionized the practice of personalized oncology. Identification of real somatic variants, especially in driver genes such as EGFR for lung tumors, KRAS for colorectal cancer, and BRAF for melanoma, is a crucial step for defining the correct molecular target therapy [3]. Here, we evaluated the effects of UDG enzymatic pretreatment in reducing sequencing artifacts of DNA from FFPE tumor samples, to improve genetic testing used in the molecular diagnostic routine. Our results clearly showed that UDG enzymatic pretreatment eliminated most of the sequence artifacts that appeared at a frequency of lower than 10% in a 16-hotspot gene panel.

### Table 1
| Samples | Transition (Ts) | Transversion (Tv) | Ts/Tv ratio |
|---------|-----------------|------------------|-------------|
|         | non UDG | with UDG | reduction (%) | non UDG | with UDG | reduction (%) | non UDG | with UDG | reduction (%) |
| PP209   | 139      | 18      | 87         | 9      | 6      | 33         | 15.4    | 3      | 81         |
| K2568   | 194      | 9       | 95         | 7      | 5      | 29         | 27.7    | 1.8    | 94         |
| A11     | 220      | 37      | 83         | 9      | 12     | -33        | 24.4    | 3.1    | 87         |
| A14     | 83       | 21      | 75         | 7      | 4      | 43         | 11.9    | 5.3    | 56         |
| A19     | 24       | 7       | 71         | 3      | 3      | 0          | 8       | 2.3    | 71         |
| A20     | 310      | NA      | NA         | 7      | NA     | NA         | 44.3    | NA     | NA         |
| A25     | 338      | 28      | 92         | 9      | 3      | 67         | 37.6    | 9.3    | 75         |
| A43     | 28       | 22      | 21         | 5      | 4      | 20         | 5.6     | 5.5    | 2          |
| O10     | 179      | 30      | 83         | 5      | 6      | -20        | 35.8    | 5      | 86         |
| O15     | 152      | 9       | 94         | 3      | 3      | 0          | 50.7    | 3      | 94         |
| O16     | 144      | 20      | 86         | 7      | 9      | -29        | 20.6    | 2.2    | 89         |
| O6      | 81       | 6       | 93         | 2      | 3      | -50        | 40.5    | 2      | 95         |

Mean (Std Dev) | 157.7 (99.1) | 18.8 (10.3) | 80 (21) | 6.1 (2.5) | 5.3 (2.9) | 5.5 (36.3) | 26.9 (14.9) | 3.9 (2.3) | 75.5 (27.1) |
CI (95%) | 82.7-204.9 | 11.9-25.7 | 65.9-94.1 | 4.3-7.7 | 3.3-7.2 | -18.8-29.8 | 15.5-35.1 | 2.4-5.4 | 57.28-93.6 |
p-value | 0.0007 | 0.3774 | 0.0007 | 0.0007 | 0.0007 | 0.0007 | 0.0007 | 0.0007 | 0.0007 |

NA: not analyzed due to failed in multiplex amplification after UDG treatment.
which includes the most important actionable mutations used for clinical decision-making in oncology.

Formalin is a formaldehyde-based fixative solution that is frequently used for the long-term storage of tumor biopsy samples. After fixation, tissue samples are usually embedded in paraffin, which promotes tissue preservation and provides a platform for tissue sectioning. Evaluation of DNA from FFPE solid tumor tissues for detection of somatic mutations is regularly performed for selecting patients to specific molecular target therapies in cancer treatment [5]. The fixation process preserves the tissue ultrastructure and cellular morphology by causing several different types of chemical interactions between adjacent macromolecules, including DNA molecules, within the tissue sample. However, this process can also lead to DNA damage through different mechanisms, including [7]: (i) extensive DNA fragmentation, the presence of which increases with longer storage time and lower pH due to formaldehyde oxidation in unbuffered solutions; (ii) protein-DNA, DNA-DNA, and DNA-formaldehyde crosslinking, which creates adducts; (iii) the formation of abasic sites (e.g., AP-sites) in the presence of water and/or reduced pH, which releases a free base and leaves a gap; and (iv) the deamination of cytosine to uracil or 5-mC to thymine, especially in CpG dinucleotides (where cytosine is commonly methylated). Indeed, some recent works have identified uracil lesions as a major source of sequence artifacts in FFPE DNA [3, 6, 10, 11].

In living cells, BER is the main repair pathway involved in correction of nonbulky lesions produced by oxidation, alkylation, deamination, abasic sites (AP-sites), and single-strand DNA breaks to prevent the mutagenic effect of these lesions [12]. DNA glycosylases initiate BER by catalyzing cleavage of the N-glycosidic bond between the damaged base and its deoxyribose, resulting in an AP-site that is further processed by other BER enzymes, ending with insertion of the correct nucleotide [9].

UDGs are monofunctional glycosylases that belong to a conserved family of DNA repair enzymes that initiate the BER pathway and remove uracil from both single- and double-stranded DNA, with greater affinity for single-stranded DNA [13], leaving an AP-site. Upon cleavage, UDG appears to remain bound to its AP-sites.
Because AP-sites are highly mutagenic and cytotoxic, this process may indicate a protective role of UDG in vivo until further action of the subsequent enzymes in the BER pathway [13]. Excision of 5-mC intermediates is, in turn, preferentially initiated by both TDG and MBD4, which excise thymine from T:G mispairs. Thymine is then replaced by cytosine in the BER pathway [14].

In an in vitro context, where there are no repair enzymes and the lesions artificially formed during experimental process cannot be repaired, deamination of cytosine gives rise to guanine mispaired with uracil, and deamination of 5-mC gives rise to guanine mispaired with thymine. These two situations can lead to the DNA polymerase incorporation of an adenine opposite to both uracil and thymine, causing C:G > T:A transition after PCR amplification. These C:G > T:A mutations can be either intrinsic to the sample before isolation (i.e., biological) or an artifact of the methodology steps, including DNA isolation, PCR amplification, and/or sequencing [6]. A recent work showed that most of the publicly available datasets have signatures of damage, leading to erroneous calls in at least one third of the G-to-T variant reads. This situation corresponds to almost one incorrect call per cancer gene, thus confounding the identification of real somatic mutations [10].

To deal with this problem, some researchers have begun to incorporate a commercial UDG enzymatic treatment in their NGS protocols, before the PCR amplification step. UDG recognizes and cleaves the uracil-containing DNA molecules, generating an AP-site without destroying the DNA sugar-phosphodiester backbone. The resulting AP-sites are susceptible to hydrolytic cleavage at the elevated temperatures used in PCR amplification cycles, resulting in DNA fragmentation. This additional step significantly reduces the frequency of C:G > T:A mutations, without affecting the capacity of NGS to detect real mutations [3, 6, 10, 11]. On the other hand, although the thymine lesions generated by deamination of 5-mC are removable from double-stranded DNA by MBD4 and TDG in a cellular environment [14], until now there has been no methodology describing the use of these enzymes as a strategy to reduce sequence artifacts of FFPE DNA in an artificial situation [7].

A recent work showed that subclonal mutations in KRAS with a very low VAF (<3%) detected in FFPE samples of metastatic colorectal carcinoma may be artifactual, reinforcing the notion that UDG pretreatment of DNA is a mandatory step to identify true mutations that can govern the choice of a therapeutic compound [15]. Similarly, our results showed that most sequence artifacts presented a low VAF (<10%), and most of them were eliminated after
UDG treatment. Moreover, we found a large reduction in the number of transitions (mean 80% reduction) by pretreatment with UDG, consistent with published data.

Some reports have demonstrated a high concordance in the reduction of mutation artifacts in fragmented or degraded DNA by UDG enzymatic pretreatment. Nevertheless, one study showed that in the case of low DNA input (30 ng of DNA from FFPE tumor tissue or 10 μL of cell-free DNA obtained from liquid biopsies), use of UDG treatment decreased PCR sensitivity sufficiently to hamper distinction between artifactual and true mutations. Thus, some caution should be exercised in the use of UDG pretreatment because of its potentially deleterious effects under some conditions [16]. Interestingly, in one of our samples, UDG treatment led to a failure of the amplification reaction, possibly due to an increase in the DNA degradation and low amount of starting material.

Conclusions

Here, we validated that pretreatment with UDG enzyme before multiplex amplification in the NGS sequencing workflow significantly decreased the number of artifactual variants, especially transitions, which were detected in FFPE samples. Our results suggest that including this additional step should improve the rate of true mutation detection in the molecular diagnostic routine.

Abbreviations

5-mC: 5-methylcytosine; AP-sites: Abasic sites; BER: Base excision repair; FFPE: Formalin-fixed and paraffin-embedded; MBD4: Methyl-binding domain protein 4; NGS: Next-generation sequencing; PCR: Polymerase chain reaction; TDG: Thymine-DNA glycosylase; Ts: Transition; Tv: Transversion; UDG: Uracil DNA glycosylase; VAF: Variant allele frequency

Acknowledgements

We acknowledge Dr. Cláudia Malheiros Coutinho Camillo and Katia Klug Oliveira from the International Research Center, A.C. Camargo Cancer Center, São Paulo, S.P. Brazil for performing DNA extraction of the tumors.

Authors’ contributions

CM analyzed and interpreted the data and wrote the manuscript. GTT designed the study, edited the manuscript, and contributed reagents and funding. All authors read and approved the final manuscript.

Funding

São Paulo Research Foundation (FAPESP).

Availability of data and materials

Datasets used and/or analyzed in the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

CAAE: 14385613.5.1001.0075.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

1 Genomics and Molecular Biology Group, International Research Center, A.C. Camargo Cancer Center, São Paulo, SP, Brazil. 2 Stomatology Department, School of Dentistry, University of São Paulo, São Paulo, SP, Brazil.

Received: 18 February 2019 Accepted: 20 May 2019

Published online: 10 July 2019

References

1. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niaz F, Turcotte CL, Izyk GP, Lupski JR, Chirn S, Song XZ, Liu Y, Yuan Y, Niaareth L, Qin X, Muzy N, Margules M, Weinstock GM, Gibbs RA, Rothenberg JM. The complete genome of an individual by massively parallel DNA sequencing. Nature. 2008;452:872–6.
2. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NL, Roy S, Tsimberidou AM, Vennak-Jones CL, Wolff DJ, Younes A, Nikiforova MN. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. J Mol Diagn. 2017;19:4–23.
3. Do H, Dobrovic A. Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil-DNA glycosylase. Oncotarget. 2012;3:546–58.
4. Gargis AS, Kalman L, Lubin IM. Assuring the quality of next-generation sequencing in clinical microbiology and public health laboratories. J Clin Microbiol. 2016;54(2):2857–65.
5. Hastings RI, Bown N, Tibielli MG, Debiec-Rychter M, Vanni R, Espinet B, van Roy N, Roberts P, van den Berg-de Ruiter E, Bernheim A, Schoumans J, Chatters S, Zemanova Z, Stevens-Kroef M, Simons A, Heim S, Salido M, Ylstra B, Betts DR. Guidelines for cytogenetic investigations in tumours. Eur J Hum Genet. 2016;24:6–13.
6. Chen G, Mosier S, Gocke CD, Lin MT, Eslehammer JR. Cytosine deamination is a major cause of baseline noise in next-generation sequencing. Mol Diag Ther. 2014;18:587–93.
7. Do H, Dobrovic A. Sequence artefacts in DNA from formalin-fixed tissues: causes and strategies for minimization. Clin Chem. 2015;61:694–71.
8. Lewis CA Jr, Crayle J, Zhou S, Swanstrom R, Wolfenden R. Cytosine deamination and the precipitous decline of spontaneous mutation during Earth’s history. Proc Natl Acad Sci U S A. 2016;113:8194–9.
9. Jacobs AL, Schar P. DNA glycosylases: in DNA repair and beyond. Chromosoma. 2012;121:1–20.
10. Chen L, Liu P, Evans TC Jr, Ettwiller LM. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. Science. 2017;355:752–6.
11. Serizawa M, Yokota T, Hosokawa A, Kusakuka K, Sugiyama T, Tsutoba Y, Yasui H, Nakajima T, Koh Y. The efficacy of uracil DNA glycosylase pretreatment in amplicon-based massively parallel sequencing with DNA extracted from archived formalin-fixed paraffin-embedded esophageal cancer tissues. Cancer Gene. 2015;208:415–7.
12. Dianov GL, Hubscher U. Mammalian base excision repair: the forgotten archangel. Nucleic Acids Res. 2013;41:3483–90.
13. Schormann N, Ricciardi R, Chattopadhyay D. Uracil-DNA glycosylases: structural and functional perspectives on an essential family of DNA repair enzymes. Protein science: a publication of the Protein Society. 2014;23:1667–85.
14. Moreira S, Grin I, Vigoour A, Couve S, Henniot V, Sarapaev M, Ichikane AA. Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA. Nucleic Acids Res. 2012;40:9917–26.
15. Mariani S, Bertero L, Osella-Abate S, Di Bello C, Francia Di Celle P, Coppola V, Sapino A, Cassoni P, Marchio C. Extreme assay sensitivity in molecular diagnostics further unveils intratumour heterogeneity in metastatic colorectal cancer as well as artifactual low-frequency mutations in the KRAS gene. Br J Cancer. 2017;117:538–66.
16. Mariani S, Bertero L, Coppola V, Saracco G, Arezzo A, Francia Di Celle P, Metovic J, Marchio C, Cassoni P. Awareness of mutational artefacts in suboptimal DNA samples: possible risk for therapeutic choices. Expert Rev Mol Diagn. 2018;18:467–75.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.