Reducing Artifactual EGFR T790M Mutations in DNA from Formalin-Fixed Paraffin-Embedded Tissue by Use of Thymine-DNA Glycosylase

Hongdo Do,1,2,3* Ramyar Molania,1,4 Paul L. Mitchell,5,6 Rita Vaiskunaite,7 John D. Murdoch,7 and Alexander Dobrovic1,2,3*

BACKGROUND: False-positive EGFR T790M mutations have been reported in formalin-fixed lung tumors, but the cause of the false positives has not been identified. The T790M mutation results from a C>T change at the cytosine of a CpG dinucleotide. The presence or absence of methylation at this cytosine has different consequences following deamination, resulting in a thymine or uracil, respectively, both of which however result in an artifactual change. Uracil-DNA glycosylase (UDG) can be used to eliminate DNA templates with uracil residues but is not active against artifactual thymines. We therefore investigated the use of thymine-DNA glycosylase (TDG) to reduce artifactual T790M mutations.

METHODS: Formalin-fixed normal lung tissues and lung squamous cell carcinomas were tested to measure the frequency of false-positive EGFR mutations by use of droplet digital PCR before and after treatment with either UDG or TDG. Methylation at the cytosine at EGFR T790 was assessed by pyrosequencing and by analysis of public databases.

RESULTS: Artifactual EGFR T790M mutations were detected in all of the archival formalin-fixed normal lung and lung squamous cell carcinomas at mutant allele frequencies of 1% or lower. The cytosine at EGFR T790 showed high levels of methylation in all lung cancer samples and normal tissues. Pretreatment of the formalin-fixed DNA with either UDG or TDG reduced the false EGFR T790M mutations, but a greater reduction was seen with the TDG treatment.

CONCLUSIONS: Both U:G and T:G lesions in formalin-fixed tissue are sources of false-positive EGFR T790M mutations. This is the first report of the use of TDG to reduce sequence artifacts in formalin-fixed DNA and is applicable to the accurate detection of mutations arising at methylated cytosines.

Lung cancer patients with activating EGFR8 mutations are selectively treated with EGFR inhibitors, often resulting in a dramatic clinical response (1, 2). All responding tumors, however, eventually develop resistance to the EGFR inhibitors through several resistance mechanisms (3). The EGFR T790M mutation is the most common resistance mechanism to first- and second-generation EGFR inhibitors and is responsible for >50% of resistant tumors (4, 5). Third-generation EGFR inhibitors have been developed to treat the T790M mutant lung tumors. Recent clinical trials have demonstrated the anti-tumor activity of third-generation EGFR inhibitors such as osimertinib against T790M mutant lung tumors (6, 7). Accurate EGFR T790M mutation testing is thus important not only to identify the mechanism of tumor resistance but also to stratify lung cancer patients for treatment with third-generation EGFR inhibitors.

Several reports have appeared claiming that the EGFR T790M mutation is present in a substantial proportion (40%–79%) of EGFR inhibitor-naïve lung tumors (8–10). In these reports, the majority of the EGFR T790M mutations were detected at levels of <1% of mutant allele frequency (11) and thus were more readily detectable by highly sensitive detection methods. It was thus considered that EGFR T790M mutant tumor cells existed before the exposure to EGFR inhibitors.

However, the results of a recent study raise the possibility of false-positive EGFR T790M mutations when
formalin-fixed lung tumors are tested (12). In the study, the \textit{EGFR} T790M mutation status was compared between fresh frozen and matched formalin-fixed lung tumors. In formalin-fixed tissues, the \textit{EGFR} T790M mutation was found at a high frequency not only in the tumors (41.7\%) but also in the adjacent normal tissues (48.5\%). These results contrasted sharply with a very low frequency observed in matched fresh frozen lung tumors (2.8\%). These results are strongly suggestive that most \textit{EGFR} T790M mutations seen in treatment-naive tumors are artifactual.

Our previous studies have demonstrated that artifactual C>T and G>A mutations arise from damaged DNA in formalin-fixed tissues (13, 14). In formalin-fixed DNA, uracil lesions resulting from cytosine deamination are the major source of the C>T and G>A sequence artifacts, which can be reduced by treating formalin-fixed DNA with uracil-DNA glycosylase (UDG)\(^9\) (13). These findings have been confirmed by other groups (15, 16).

UDG is a DNA repair enzyme that removes uracil from single- and double-stranded DNA by hydrolyzing the N-glycosidic bond between the uracil base and the sugar phosphate backbone (17). We observed that artifactual C>T variants that were resistant to the UDG treatment mainly occurred at CpG dinucleotides, consistent with the presence of methylation and the resulting deamination of 5-methylcytosine (5-mC) to thymine (13, 14).

\textit{EGFR} T790M mutations occur at the CpG dinucleotide in codon 790 of \textit{EGFR} exon 20. Cytosines at a CpG dinucleotide are either methylated (5-mC) or unmethylated, but nonpromoter CpGs, including those at intragenic CpG sites of transcribed genes, are generally methylated. Deamination of cytosine and 5-mC results in uracil and thymine bases, creating U:G and T:G mismatches, respectively.

As cytosine is deaminated in formalin-fixed tissues (13, 14), it is likely that 5-mC undergoes corresponding deamination to thymine. This study is the first report to examine 5-mC deamination as the source of sequence artifacts in DNA from formalin-fixed tissues by use of \textit{EGFR} T790M as an example of a clinically important mutation that would be affected by such artifacts. In addition, we assessed the use of thymine-DNA glycosylase (TDG) to minimize these artifacts.

---

\textbf{Materials and Methods}

\textbf{SAMPLES AND DNA EXTRACTION}

Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen) from 72 formalin-fixed lung squamous cell carcinomas (SCCs) that were previously reported to have high levels of C>T and G>A sequence artifacts (18) and 10 formalin-fixed normal lung tissues. Genomic DNA was also extracted from peripheral blood mononuclear cells from 2 healthy individuals, 11 lung cancer cell lines, and 23 formalin-fixed non-small cell lung tumors for DNA methylation studies. This study was approved by the Human Research Ethics Committee at the Austin Hospital, Melbourne, Australia.

\textbf{DNA METHYLATION STATUS OF \textit{EGFR} EXON 20 IN THE CANCER GENOME ATLAS AND NORMAL TISSUES}

There are 9 CpG dinucleotides, including the CpG at T790, within the human \textit{EGFR} exon 20 sequence. The methylation status of those 9 CpG dinucleotides was assessed in lung cancer tissue, normal lung tissue, and other normal tissues by use of The Cancer Genome Atlas and The Human Epigenome Atlas. At the time of data download, complete whole-genome bisulfite sequencing data were available for 5 adenocarcinomas (ADCs) and 5 SCCs of lung, 1 normal lung, and 18 different types of normal tissues. The details of the normal tissues are shown in Table 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol63/issue9.

\textbf{METHYLATION-SENSITIVE HIGH-RESOLUTION MELTING}

Genomic DNA was bisulfite modified by use of the EZ DNA Methylation-Lightning Kit (Zymo Research) according to the manufacturer’s instructions. A methylation-sensitive high-resolution melting (MS-HRM) assay was developed to screen the methylation status at the \textit{EGFR} c.2369 cytosine and 3 upstream CpG dinucleotides by following the guidelines described previously (19). MS-HRM primers were designed to amplify the sense strand of the bisulfite-modified DNA sequence to avoid the influence of a common single-nucleotide polymorphism (rs1050171, C>T) on melting analysis; forward 5'-gaaggTagTYgaaggTattag-3’ and reverse 5’-biotin-gtAatAAccaAcRtAacacccccca-3’. The capital “T” in the forward and the capital “A” in the reverse primer sequences indicate the position of cytosine before the bisulfite conversion. The capital “Y” in the forward and the capital “R” in the reverse primer sequences indicate the position of cytosine at a CpG site before the bisulfite conversion. The reverse primer was biotinylated to enable subsequent pyrosequencing. Methylation dilution standards were prepared as previously described (20).

PCR cycling and HRM were performed on the RotorGene 6000 (Corbett). The PCR mixture was

---

\(^9\) Nonstandard abbreviations: UDG, uracil-DNA glycosylase; 5-mC, 5-methylcytosine; ADC, adenocarcinoma; ddPCR, droplet digital PCR; FFPE, formalin-fixed paraffin-embedded; MS-HRM, methylation-sensitive high-resolution melting; SCC, squamous cell carcinoma; TDG, thymine-DNA glycosylase; XNA, xenonucleic acid.
prepared in a final volume of 20 μL and contained 10 ng of bisulfite modified templates, 1× PCR buffer, 2.5 MgCl₂, 200 nmol/L each primer, 200 μmol/L deoxy-nucleotide triphosphate, 5 μmol/L SYTO 9, and 0.5 U HotStar Taq (Qiagen). The PCR cycling and HRM conditions were as follows: 1 cycle of 95 °C for 15 min; 55 cycles of 95 °C for 10 s, 60 °C for 20 s, 72 °C for 30 s; 1 cycle of 97 °C for 1 min, and a melt from 70 °C to 95 °C with an increment of 0.2 °C/s. The melting profiles were analyzed by use of the RotorGene 6000 software. All samples were tested in duplicate.

**ENRICHMENT OF EGFR T790M MUTANT TEMPLATES BY USE OF AN XNA BLOCKER**

A xenonucleic acid (XNA) blocker was used to enrich EGFR T790M mutant templates by selectively suppressing the amplification of wild-type templates during PCR cycling. Three formalin-fixed DNAs extracted from the archival SCCs of lung were amplified both with and without the XNA blocker (DiaCarta) by use of the following primers: forward 5′-gtcggcactgctcacc-3′ and reverse 5′-biotin-tctttgtgttcccggacatagtc-3′. The reverse primer was biotinylated to enable subsequent pyrosequencing. PCR amplification was performed on the RotorGene 6000. The PCR mixture was prepared in a final volume of 23 μL and contained 50 ng of formalin-fixed paraffin-embedded (FFPE) DNA, 1× Qclamp master mix (DiaCarta), 200 nmol/L each primer, 200 μmol/L deoxynucleotide triphosphate, and 10 μL of XNA. The PCR cycling conditions were as follows: 1 cycle of 95 °C for 5 min; 55 cycles of 95 °C for 20 s, 70 °C for 40 s, 66 °C for 30 s, and 72 °C for 40 s. After PCR amplification, the PCR products were used for subsequent pyrosequencing.

**PYROSEQUENCING**

Pyrosequencing was performed on the Qseq instrument (Bio Molecular Systems) by use of the PyroMark Gold Q24 reagents (Qiagen) and Streptavidin Mag Sepharose beads (GE Healthcare) according to the manufacturer’s protocol. The forward primers were used as pyrosequencing primers. Pyrosequencing data were analyzed by use of the Qseq analysis software version 2.0.11 (Bio Molecular Systems).

**DROPLET DIGITAL PCR (ddPCR) FOR DETECTION OF EGFR T790M MUTATION**

The PrimePCR ddPCR mutation assay (Bio-Rad) was used for detection of the EGFR T790M mutation by use of the QX200 ddPCR system (Bio-Rad). The digital PCR reaction mixture was prepared in a final volume of 23 μL, and the manufacturer’s instructions were followed for droplet generation, PCR cycling, and droplet reading. ddPCR data were analyzed by use of QuantaSoft Software (Bio-Rad).

**EGFR T790M DDPCR WITH AND WITHOUT UDG AND TDG TREATMENT**

Two units of UDG (New England Biolabs) were directly added to the ddPCR sample mixture. After droplet generation, the droplets underwent the standard PCR cycling after an incubation step at 37 °C for 2 h. Noncommercial TDG was obtained from New England Biolabs as a cloned, expressed, and purified protein at a concentration of 5.2 μg/μL. For TDG treatment, 300 ng of FFPE DNA was mixed with 10.4 μg of the New England Biolabs TDG and 1× TDG buffer, followed by incubating at 37 °C for 2 h and 95 °C for 10 min.

To verify the effect of TDG on the reduction of artifactual EGFR T790M changes, commercially available TDG was obtained from Trevigen and 3 FFPE DNA samples were tested by use of the EGFR T790M ddPCR assay before and after Trevigen TDG treatment. For the TDG treatment, each FFPE DNA sample was mixed with 2.5 U of TDG and 1× TDG buffer, followed by incubation at 65 °C for 1 h. The TDG-treated FFPE DNA was purified by use of the DNA Clean & Concentrator kit (Zymo Research) before the ddPCR assay.

**Results**

**ARTIFICIAL EGFR T790M MUTATION IN FORMALIN-FIXED LUNG TISSUES**

We first assessed whether artifactual EGFR T790M mutations are detectable in formalin-fixed normal lung tissues. Ten formalin-fixed normal lung tissues were tested by use of a ddPCR assay for both sensitive detection and accurate quantification of artifactual EGFR T790M mutations. DNA was freshly extracted from the 10 archival normal lung tissues dating from 2014 and 2015. The EGFR T790M ddPCR assay contained the hexachloro-6-carboxyfluorescein- and 6-carboxyfluorescein-labeled fluorescent probes for separate detection of the wild-type and T790M mutant templates, respectively, enabling accurate quantitative analysis. Artifactual T790M mutations were detected by use of this sensitive assay in all of the normal lung tissues. However, they were detected at low allelic frequencies of around 0.2% (Fig. 1A).

We then tested 72 formalin-fixed, EGFR inhibitor-naïve archival lung SCCs. The lung SCCs were chosen, as the EGFR T790M mutation is not normally found in this tumor type (21). DNA was extracted from the archival lung SCCs dating from 1992 to 2008. Artifactual EGFR T790M mutations were detected in the lung SCCs at low allelic frequencies of around 0.4%, ranging from 0.1% to 1%, with a tendency to higher frequencies in older samples (Fig. 1B). These results indicate that artifactual EGFR T790M mutations are frequently detectable in formalin-fixed tissues when highly sensitive detection methods are used.
Fig. 1. False-positive EGFR T790M mutations in archival lung tissues.

False-positive EGFR T790M mutations were assessed in formalin-fixed lung tissues of 10 normal lungs (A) and 72 lung SCCs (B) by use of a droplet digital PCR assay. Artifactual T790M mutations were detected in all the normal lungs and lung SCCs at allelic frequencies of up to 1%. Sample numbers were given according to the date of collection. The black trend line in Fig. 1B shows the tendency to higher false-positive EGFR T790M mutations in the older samples. (C), The presence of templates causing the false-positive EGFR T790M mutations was confirmed in selected cases by XNA-based enrichment, followed by pyrosequencing. The H1975 lung cancer cell line DNA, which carries the T790M mutation, and 3 lung SCC DNAs were amplified with and without a XNA blocker that specifically inhibits the amplification of the EGFR wild-type allele. The position of the mutant “T” allele of the T790M mutation is indicated by red arrows. In H1975, the level of the mutant allele is increased from 67% to 91%, confirming the wild-type suppressive function of the XNA blocker. In the 3 lung SCC samples, the EGFR T790M mutation was not detectable by pyrosequencing without the XNA enrichment, concordant with their low allelic frequencies of 0.7 (SCC23), 0.6 (SCC30), and 0.7 (SCC35). The use of the XNA blocker enabled the T790M mutant allele to be enriched to 59% (SCC23), 66% (SCC30), and 65% (SCC35), helping visualization of the artifactual T790M mutation.
Reducing Artifactual \textit{EGFR} T790M in FFPE DNA

We used pyrosequencing to confirm the artifactual \textit{EGFR} T790M mutations in formalin-fixed tissues. Because the artifactual \textit{EGFR} T790M mutations were present below the analytic sensitivity of pyrosequencing, we used an XNA blocker for selective enrichment of the artifactual T790M allele by blocking the amplification of the wild-type templates during PCR cycling.

We first assessed the suppressive effect of the XNA blocker on the wild-type templates by use of the H1975 lung cancer cell line as an \textit{EGFR}T790M mutant control. Without the XNA blocker, the T790M mutant allele was detected in 67% of total templates, consistent with a 2:1 mutant to wild-type allelic ratio. After the XNA-based enrichment, the allelic proportion of the T790M mutation was increased to 91% (Fig. 1C), confirming the suppressive role of the XNA blocker on the wild-type templates.

We then tested 3 of the archival lung SCCs by pyrosequencing with and without the XNA-based enrichment. The \textit{EGFR} T790M mutation was not detectable (i.e., less than the approximate 10% analytical sensitivity of the pyrosequencing assay) in any of the lung SCCs without the enrichment, consistent with the ddPCR results showing lower than 1% mutant allele frequencies. However, the artifactual \textit{EGFR} T790M mutation was clearly detected after enrichment with an allelic frequency of around 60% (Fig. 1C). This confirmed that the ddPCR results were correctly detecting c.2369 C>T changes, which however were artifacts arising from DNA modified because of DNA damage.

\textbf{METHYLATION STATUS OF \textit{EGFR} c.2369 IN LUNG CANCER AND NORMAL TISSUES}

We have shown that C>T artifacts remaining after UDG treatment are often found at CpG sites (13, 14). As UDG cannot remove the thymine because of the deamination of a 5-mC, and the T790M mutation arises at a CpG site, we therefore examined whether deamination of 5-mC to thymine is associated with artifactual T790M mutations.

We first examined the methylation status of the cytosine at the CpG site within T790 and the surrounding 8 CpG sites in \textit{EGFR} exon 20 in The Cancer Genome Atlas lung cancer data set. Whole-genome bisulfite sequencing data were available for 5 ADCs and 5 SCCs of lung. All of the 9 CpG dinucleotides were shown to be methylated at high levels in the 5 ADCs and 5 SCCs of lung (Fig. 2, A and B).

We also examined the methylation status of the 9 CpG sites in \textit{EGFR} exon 20 in normal lung and 18 different normal tissues in The Human Epigenome Atlas database. Similar to the results seen in lung cancer samples, the 9 CpG dinucleotides were highly methylated in the normal lung and various normal tissues (Fig. 2C and 2D). When the levels of methylation were examined, the \textit{EGFR} c.2369 position (CpG 6) was methylated in a range of 40%–100% in the lung ADCs and 75%–100% in the lung SCCs, 98% in normal lung, and 8%–100% in other normal tissues. The levels were an approximation owing to low sequencing coverage of the genomic region examined. The levels of cytosine methylation were also detected in other normal tissues.

To further confirm the cytosine methylation status at the \textit{EGFR} c.2369 position, an independent sample cohort was tested by use of a different detection method, MS-HRM followed by pyrosequencing. The sample cohort included 11 lung cancer cell lines, 23 lung tumors, and 2 mononuclear cell samples from healthy individuals. The MS-HRM assay was used to screen the methylation status at the \textit{EGFR} c.2369 and 3 adjacent upstream CpG dinucleotides. Consistent with The Cancer Genome Atlas lung cancer and normal tissue results, cytosine methylation was detected in all of the samples by MS-HRM (Fig. 2E and 2F).

The MS-HRM PCR products were then pyrosequenced to quantify the levels of cytosine methylation (Fig. 2G). The \textit{EGFR} c.2369 position (CpG 6) was highly methylated in all samples: 11 lung cancer cell lines (mean, range: 86.2%, 19%–98%), 23 lung tumors (mean, range: 70.6%, 48%–90%), and 2 mononuclear cell samples (mean 89%). Similar high levels of cytosine methylation were also detected in other CpG dinucleotides. These results indicate that the cytosine at the \textit{EGFR} c.2369 position is part of a CpG island that is methylated at high levels in both normal lung tissue and lung cancer tissue.

\textbf{THYMINE AND URACIL LESIONS AS THE SOURCES OF ARTIFICIAL T790M MUTATIONS}

As the cytosines at the \textit{EGFR} c.2369 position were highly but not fully methylated, thymine and uracil lesions resulting from 5-mC and cytosine deamination were examined as the sources of artifactual T790M mutations. The level of artifactual \textit{EGFR} T790M mutations was compared in 10 normal lung tissues and 10 lung SCCs before and after the treatment with UDG or TDG (Fig. 3).

Treatment of the normal lung and lung SCC DNAs with UDG showed a reduction in the artifactual T790M mutations in a few samples (NL1, SCC7, and SCC26), but the level of the artifactual T790M mutations was largely unchanged in most of the samples. However, the treatment with the TDG resulted in a marked reduction in all of the normal lung samples (mean, 65%; range, 38%–81%) and also in all of the lung SCCs (mean, 49%; range, 23%–75%). To confirm the effect of TDG from New England Biolabs on the reduction of artifactual \textit{EGFR} T790M changes in FFPE DNA, commercial TDG from a different supplier was used to treat 3 of the SCC DNA samples. Consistent with our previous results,
DNA methylation status of the 9 CpG sites in EGFR exon 20 was analyzed by use of whole bisulfite sequencing data from publicly available databases, The Cancer Genome Atlas and The Human Epigenome Atlas (A–D). High levels of DNA methylation were seen at each CpG site; 5 lung ADCs (A), 5 lung SCCs (B), 1 normal lung tissue sample (C), and 23 other normal tissues (D). To confirm the DNA methylation status, an independent sample cohort (11 lung cancer cell lines, 23 lung tumors, and 2 mononuclear cell samples from healthy individuals) was tested by use of MS-HRM and pyrosequencing (E–G). The MS-HRM amplicon contains 4 CpG sites including the T790M locus and 3 adjacent upstream CpG sites. High levels of DNA methylation were detected by MS-HRM in all of the samples as shown by an increased melting temperature relative to the unmethylated control. The melting profiles of fully methylated (red) and unmethylated (blue) DNA controls, and the representative samples (green) of mononuclear cells (MNC25; E) and a lung tumor (LC3; F) are shown. Pyrosequencing of the MS-HRM products confirmed the DNA methylation status (G).
the level of artifactual EGFR T790M changes was markedly reduced by 45%, 72%, and 74% after the TDG treatment (data not shown). These results indicate that T:G lesions resulting from 5-mC deamination are the primary source of the false EGFR T790M mutations in formalin-fixed tissues.

Discussion

Our previous studies identified that uracil lesions in the DNA of formalin-fixed tissues are the major source of C>T (and G>A) sequence artifacts (13, 14). Cytosine is known to undergo deamination to uracil. Because many DNA polymerases incorporate an adenine opposite to the uracil lesions, uracil lesions in formalin-fixed DNA are hotspots for C>T sequence artifacts. In this study, we identified that methylated cytosines at CpG sites are also hotspots for C>T sequence artifacts as a result of the deamination of 5-mC to thymine.

The EGFR T790M mutation is a clinically important example, particularly because testing for this mutation is a requisite for treatment with third-generation EGFR inhibitors. We demonstrated that the cytosine at the EGFR T790M position is methylated at varying levels in lung tumors and normal cells. We further showed that deamination of cytosine and 5-mC to uracil and thymine is responsible for the artifactual EGFR T790M mutations by pretreating formalin-fixed DNA with either UDG or TDG. The greater reduction was seen with the TDG pretreatment, whereas UDG had little effect on most of the samples.

The reason for the greater reduction in the artifactual EGFR T790M mutations by TDG treatment is the high proportion of methylated cytosines leading to T:G mismatches that can be removed by only TDG, but not by UDG. Furthermore, the uracils in the U:G mismatches are also removed by TDG (21). Thus, TDG pretreatment alone can remove both thymine and uracil lesions in formalin-fixed DNA (Fig. 4).

Although TDG pretreatment helped reduce the artifactual EGFR T790M mutations, residual artifacts remained. It is unclear what causes the residual artifacts, but incomplete enzymatic removal of thymine and uracil lesions by the TDG pretreatment may be the reason. We are currently investigating the cause of these residual artifacts after TDG treatment.

The findings of this study can be applied to improve the quality of molecular testing in which damaged DNA is extracted from formalin-fixed tissues. We studied the artifactual EGFR T790M mutations, but the benefit of TDG pretreatment is not limited to the EGFR T790M locus. As TDG can remove uracil, thymine, and 5-hydroxymethyluracil lesions, TDG treatment can be used to reduce sequence artifacts arising from those lesions.
It should be noted that our formalin-fixed DNA samples were from archival lung SCCs in which high rates of sequence artifacts arising from cytosine deamination were previously detected \((13, 14)\). Sequence artifacts may be detected at lower rates in DNA from recently formalin-fixed samples as seen in the normal lung tissues. However, the use of the archival samples allowed us to more readily explore the effects of UDG and TDG treatment due to the higher proportion of artifactual mutations.

In addition, there are several reports of sequence artifacts detected from recently formalin-fixed samples \((22, 23)\). Thus, it cannot be assumed that recently fixed samples are entirely sequence artifact free. This is of particular importance when highly sensitive or enrichment methodologies are used. Enrichment techniques such as blocker \((24)\) or cold-PCR \((25)\) can lead to damaged templates getting selectively amplified during PCR cycling. Our results with the XNA blocker are a good example of the enrichment of damaged templates leading to false-positive results.

Interestingly, ancient DNA shows the same cytosine deamination phenomenon \((26)\) and thus can benefit from the use of TDG treatment. UDG treatment of ancient DNA has shown to improve sequencing accuracy by reducing C>T sequence artifacts \((27)\). However, C>T sequence artifacts present at CpG dinucleotides were resistant to the UDG treatment \((27)\), consistent with deamination of 5mC to thymine.

Massively parallel sequencing is a widely used platform for detection of sequence variants. Accurate detection of somatic sequence variants in formalin-fixed tissues by massively parallel sequencing is often hampered by numerous low levels of sequence artifacts \((14, 22, 28)\). Thus, TDG pretreatment could be implemented in the massively parallel sequencing workflow to minimize artifactual variants in formalin-fixed tumors.

In conclusion, the presence of low-level C>T mutations in formalin-fixed DNA should be interpreted with caution, as sequence artifacts are difficult to distinguish from true mutations \((29)\). We identified that U:G and T:G mismatches because deamination of cytosine and 5-mC are the primary sources of the artifactual EGFR T790M mutations in formalin-fixed lung tumors. This is the first report demonstrating the use of TDG to reduce sequence artifacts in damaged DNA.

**Fig. 4. Deamination of cytosine and 5-mC as the sources of false-positive EGFR T790M mutations.**
Cytosine and 5-mC are deaminated to uracil and thymine, respectively. When amplified by PCR, the resultant U:G and T:G mismatches become the sources of C>T sequence artifacts. UDG removes the uracil from U:G mismatches. TDG removes the uracil from U:G and thymine from T:G mismatches. Thus, pretreatment of formalin-fixed DNA with UDGs or TDGs enables artifactual C>T (and G>A) sequence artifacts to be minimized.

**Author Contributions:** All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

**Authors’ Disclosures or Potential Conflicts of Interest:** Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

**Employment or Leadership:** None declared.

**Consultant or Advisory Role:** None declared.

**Stock Ownership:** None declared.

**Honoraria:** None declared.
Reducing Artifactual EGFR T790M in FFPE DNA

Research Funding: The Olivia Newton-John Cancer Research Institute is supported by the Operational Infrastructure Support Program Funding of the Victorian State Government. H. Do, the Cancer Council of Victoria and the Austin Medical Research Foundation; P.L. Mitchell, the Australian National Health and Medical Research Council; A. Dobrovic, the Australian National Health and Medical Research Council, Cancer Australia. QClamp EGFR Mutation Test kit was provided by Mike Powell from DiaCarta, noncommercial TDG was provided by New England Biolabs, and normal lung tissues were supplied by Gavin Wright.

Expert Testimony: None declared.

Patents: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, and final approval of manuscript.

Acknowledgments: We would like to thank Mike Powell from DiaCarta for providing the QClamp EGFR Mutation Test kit and New England Biolabs for noncommercial TDG. We also would like to thank Gavin Wright for supplying the normal lung tissues.

References


Clinical Chemistry 63:9 (2017)