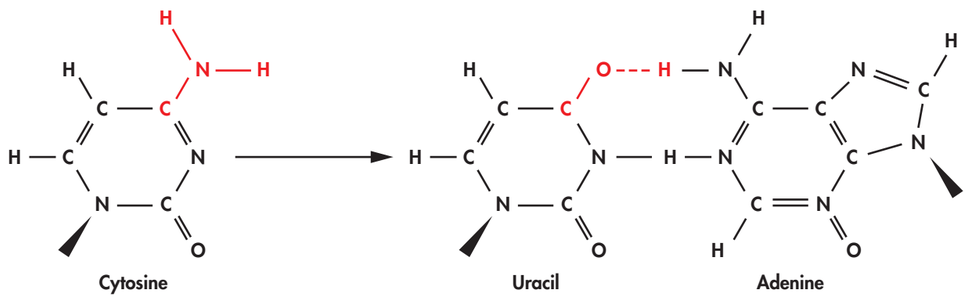
# A new method for DNA extraction and artifact removal from FFPE samples for next-generation sequencing experiments

Dominic O'Neil, Stefanie Schroeer, Yanquan Luo, and Tanya Sperling

QIAGEN GmbH, QIAGEN Strasse 1, 40724 Hilden, Germany

## Introduction

Formalin fixation and long storage periods in non-ideal conditions contribute to the severe damage and sequence artifacts often seen in DNA isolated from formalin-fixed, paraffin-embedded (FFPE) tissue. Cytosine-to-thymine transitions constitute the majority of these artifacts (1, 2). While the exact mechanism for this is not known, one explanation is that deamination of cytosine leads to a uracil in that position. This then pairs with adenine (3), and upon sequencing, this modified base reads as a C>T transition. If strand information is not preserved during library construction, either strand may be sequenced, so the artifact may appear as a C>T or a G>A transition. Removing these artifacts is particularly crucial when analyzing cancer samples, as they may otherwise appear as false positive mutations.




Adapted from Klug and Cummings, 1997

**Deamination of cytosine leads to false adenine pairing.** In FFPE samples, cytosine may become deaminated, leading to the cytosine pairing with adenine as uracil. In sequencing reactions, this will manifest as a C>T|G>A transition.

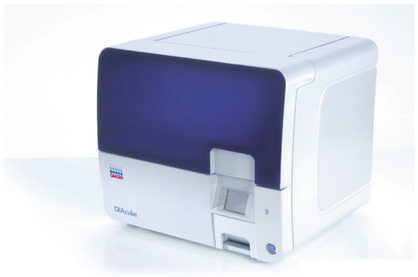
## Novel technology for FFPE DNA purification

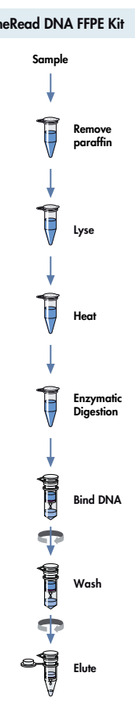
We have modified the existing, efficient QIAGEN DNA purification technology for FFPE tissue samples to significantly reduce the sequence artifacts resulting from formalin fixation and storage, and to optimize yields from small amounts of starting material.

**The new GeneRead® DNA FFPE Kit* provides**

- Enzymatic removal of sequence artifacts resulting from fixation and storage
- High yields from low sample inputs
- Efficient, non-toxic deparaffinization

Furthermore, the protocol is designed to be automatable on the QIAcube® instrument.

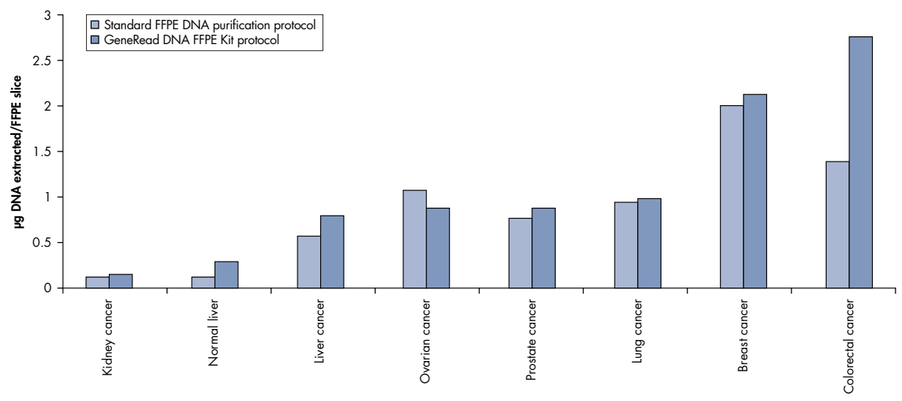





* The GeneRead DNA FFPE Kit is currently in development and will be available soon.

## High yields from limited starting material

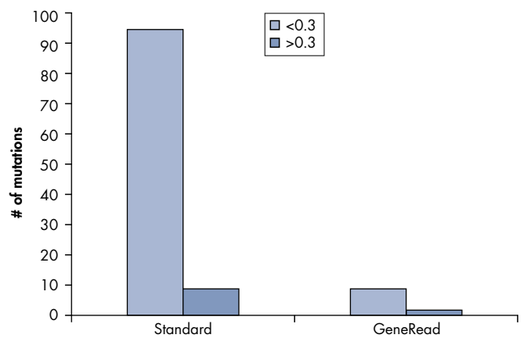
Advances in library preparation and sequencing technologies have made it attractive to perform high-throughput sequencing on the large amounts of biobanked FFPE tissues that are available. These technologies have also lowered the frequency threshold at which sequence mutations can be reliably detected. However, sequencing FFPE samples involves special challenges. Yields from such samples may be limited due to the compromised status of the DNA. Additionally, FFPE samples are often irreplaceable. There is a need to get the maximum amount of nucleic acid from the smallest amount of starting material. Artifact suppression also becomes critical when sequencing FFPE samples, as the relative frequency of false mutations will be increased when starting with limited material.



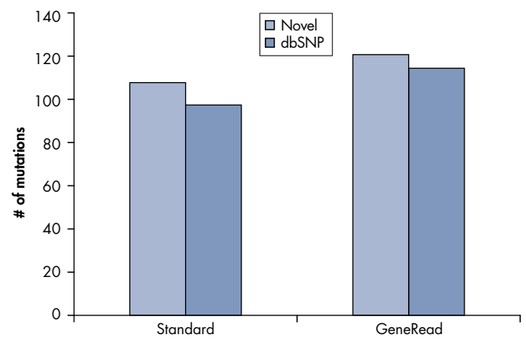

**The GeneRead DNA FFPE kit provides high yields of double-stranded DNA.** Single 10 µm slices of FFPE tissue were used to compare the results of DNA purification using standard and GeneRead protocols for FFPE samples. The wide range of DNA yields seen is due to the variation in the starting material. This is common for DNA purification from FFPE samples.

## Efficient removal of artifactual C>T|G>A transitions

DNA damage caused by formalin fixation and storage is essentially random, with the resulting altered sites distributed across the sequence. Only a few copies of the genome are damaged at any site, leading to a generally low frequency of these artifacts. As low-frequency novel mutations can be very important, for example in cancer analysis, it is important to distinguish between true and false low-frequency mutations. By examining low-frequency novel mutations, we can estimate the number of artifacts in a sample and the effectiveness of the artifact removal process of our new protocol.




**Dramatic reduction in artifactual C>T|G>A mutations.** The GeneRead DNA FFPE Kit removes over 90% of the low-frequency novel mutations that are most likely to be artifactual in nature.




**High frequency C>T|G>A transitions are retained.** High frequency C>T|G>A transitions, whether novel or present in dbSNP, are retained when the new GeneRead DNA FFPE Kit protocol is applied.

## Removal of false positive COSMIC variants

| Chrom | Pos | COSMIC ID | dbSNP ID | Gene Name | Ref | Var | Standard FFPE | GeneRead FFPE |
|---|---|---|---|---|---|---|---|---|
| chr1 | 226595647 | COSN392383 | rs907187 | PARP1 | C | G | 0.95 | 1.00 |
| chr2 | 29416481 | COSM1130802 | rs1881420 | ALK | T | C | 1.00 | 1.00 |
| chr2 | 48032105 | COSM13342 | – | MSH6 | C | T | 0.13 | 0.00 |
| chr3 | 30713126 | COSM149346 | rs11466512 | TGFBR2 | T | A | 0.51 | 0.45 |
| chr3 | 47125385 | COSM149376 | rs4082155 | SETD2 | G | A | 0.59 | 0.63 |
| chr4 | 1807130 | COSM327089 | – | FGFR3 | C | T | 0.16 | 0.00 |
| chr4 | 55152040 | COSM22413 | rs2228230 | PDGFRA | C | T | 0.63 | 0.67 |
| chr4 | 55595519 | COSM12708 | rs121913516 | KIT | C | T | 0.31 | 0.56 |
| chr5 | 35861068 | COSM149813 | rs1494558 | IL7R | T | C | 0.53 | 0.50 |
| chr5 | 35871190 | COSM149814 | rs1494555 | IL7R | G | A | 0.49 | 0.47 |
| chr5 | 35875593 | COSN167436 | rs987106 | IL7R | A | T | 0.66 | 0.49 |
| chr5 | 180036871 | COSN167671 | rs2242219 | FLT4 | C | G | 0.60 | 0.59 |
| chr7 | 55214348 | COSM42978 | rs2072454 | EGFR | C | T | 1.00 | 1.00 |
| chr9 | 21968199 | COSM14251 | rs11515 | CDKN2A | C | G | 0.99 | 0.99 |
| chr9 | 139397707 | COSM33747 | rs10521 | NOTCH1 | G | A | 1.00 | 1.00 |
| chr11 | 64572018 | COSM255213 | rs2959656 | MEN1 | T | C | 0.99 | 1.00 |
| chr12 | 121426785 | COSM46438 | – | HNF1A | G | A | 0.20 | 0.00 |
| chr16 | 3828705 | COSM970602 | – | CREBBP | C | T | 0.10 | 0.00 |
| chr17 | 41244000 | COSM148277 | rs16942 | BRCA1 | T | C | 0.40 | 0.52 |

**COSMIC mutations found in a 15-year old carcinoma from liver tissue.** The tissue was processed in parallel using a standard FFPE kit and the new GeneRead DNA FFPE Kit with artifact reduction. Both samples were amplified with the GeneRead DNAseq Comprehensive Cancer Panel, and sequenced with massively parallel sequencing. The frequency of the mutation is found in the last two columns. Four mutations found when using the standard kit were not found in the tissue processed using the GeneRead kit. These were the lowest frequency mutations, and all of them were C>T|G>A transitions, thus likely artifacts resulting from formalin fixation and storage. All other mutations were found in both tissues with very similar frequencies, indicating that the new kit specifically targets artifactual mutations while leaving true mutations unchanged.

## Conclusion

Sequence changes due to DNA damage from formalin fixation and FFPE sample storage are most commonly C>T|G>A transitions. The majority are low-frequency mutations with no clear biological significance, but a significant portion of the sequenced mutations with dbSNP and COSMIC identifiers may also be due to these artifactual changes. Removing these artifacts is particularly important when sequencing small amounts of starting material, where the relative frequency of these mutations will be increased. The GeneRead DNA FFPE kit maximizes yields and efficiently removes C>T|G>A artifacts resulting from formalin fixation and storage.

1. Williams, C., et al. (1999). A high frequency of sequence alterations is due to formalin fixation of archival specimens. American Journal of Pathology **155**, 1467
2. Yost, S.E., et al. (2012). Identification of high-confidence somatic mutations in whole genome sequence of formalin-fixed breast cancer specimens. Nucleic Acids Research **40:14**, 1
3. Do, H., Wong, S.Q., Li, J., and Dobrovic, A., et al. (2013). Reducing sequence artifacts in amplicon-based massively parallel sequencing of formalin-fixed paraffin-embedded DNA by enzymatic depletion of uracil-containing templates. Clinical Chemistry **59:9**, 18
4. Kerick, M., et al. (2011). Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. BMC Medical Genomics, **4**, 68.

The applications presented here are for research use only. Not for use in diagnostic procedures. For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at www.qiagen.com or can be requested from QIAGEN Technical Services or your local distributor.



1080932 02/2014