Removal of sequence artifacts in FFPE samples during DNA extraction improves next-generation sequencing results

Dominic O'Neil, Stefanie Schröer, Tanya Sperling, and Dirk Löffert
QIAGEN GmbH, QIAGEN Strasse 1, 40724 Hilden, Germany

DNA from FFPE tissues shows sequencing artifacts

Objectives: Fixation in formalin and storage causes sequence alterations in formalin-fixed paraffin-embedded (FFPE) samples. The most prevalent alteration is a cytosine to thymine transition caused by deamination of cytosines. The existence of these artifactual mutations can complicate analysis of next-generation sequencing (NGS) results, particularly when analyzing low-frequency mutations. We modified standard FFPE DNA extraction to include an enzymatic digestion of these artifacts. DNA extraction, gene panel amplification, and sequencing was performed on a panel of samples to verify the effectiveness of the new method in removing artifacts from FFPE tissues.

Methods: DNA was extracted from FFPE tissues using the new method and a standard DNA FFPE extraction kit. Samples were amplified with a cancer panel for 124 genes and sequenced. The overall artifact rate and the effectiveness of the new method in removing these artifacts was estimated, by examining low-frequency novel mutations found in the same sample.

Results: Sequencing results on DNA extracted from a panel of tissue samples confirmed that C>T and G>A were the most prevalent mutations at low frequencies. An enzymatic artifact removal step during the extraction procedure removed the majority (>80%) of low-frequency C>T|G>A variants. High frequency mutations were not affected by the artifact digestion step, indicating that the new method specifically targets artificial mutations while leaving true mutations unchanged.

Conclusion: The new method efficiently removes sequence artifacts caused by fixation and storage of FFPE material. DNA extracted using this method is of high quality and well suited for NGS experiments. This method eliminates artifacts in the nucleic acid directly, requiring no special downstream library preparation or bioinformatics manipulations to excise the artifacts. Artifact reduction during extraction simplifies data analysis and improves the resolution of low-prevalence mutations.

Background

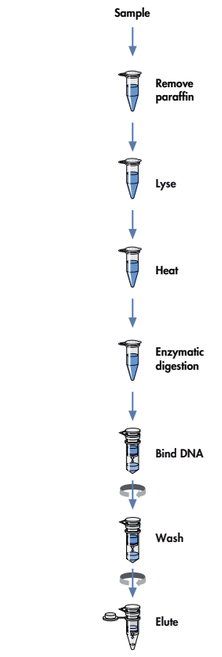
Advances in library preparation and sequencing technologies have made it attractive to perform high-throughput sequencing on the large amounts of biobanked FFPE tissues available. Such technologies have also lowered the frequency threshold at which sequence mutations can be reliably detected. However, sequencing DNA from FFPE samples presents 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.

The GeneRead™ DNA FFPE Kit

We modified the existing, efficient FFPE DNA purification technology from QIAGEN to provide a significant reduction of sequence artifacts resulting from formalin fixation and storage, and optimize yields from small amounts of starting material.

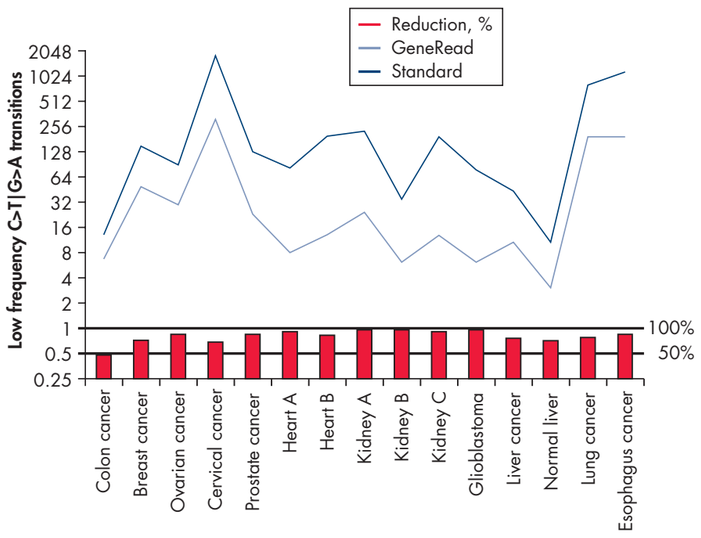
Key features of the GeneRead™ DNA FFPE Kit:

- Enzymatic removal of sequence artifacts resulting from fixation and storage
- Automatable on the QIAcube®
- High yield from low input (>500 ng double-stranded DNA from a 10 µm slice)
- Efficient, non-toxic deparaffinization with QIAGEN Deparaffinization Solution



Modified GeneRead DNA FFPE Kit workflow, automatable on the QIAcube.

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

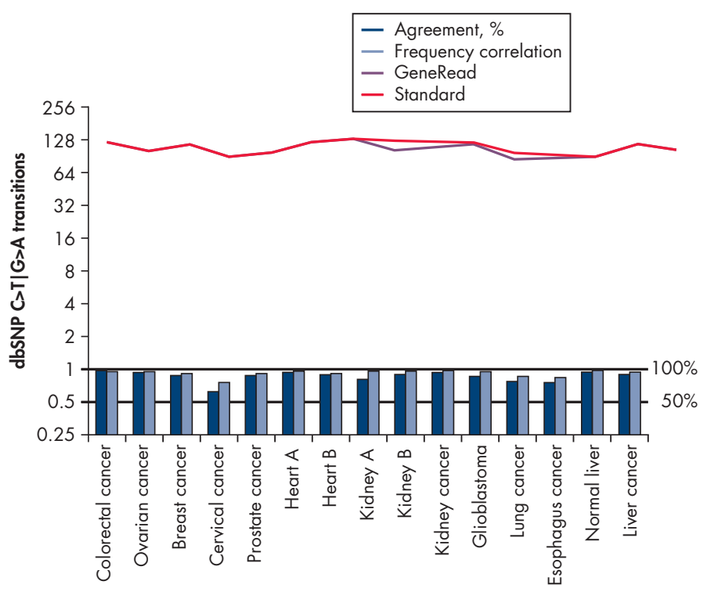


Artifact suppression is important when sequencing FFPE samples, as the relative frequency of false mutations is increased when starting with limited material.[1]

C>T|G>A transitions resulting from DNA damage in FFPE tissues will be distributed across the sequence, and will occur at low frequencies due to the random nature of these damaging events. Low-frequency, novel mutations make up the majority of these artifacts and a reduction of these markers is an effective measure of artifact removal.

Artifactual transitions constituted, on average, 73.1% (range 57.8–84.8%) of all measured novel variants in the control samples. The GeneRead DNA FFPE Kit removes the majority of these mutations.

Retention of true C>T|G>A mutations



It is essential that true C>T|G>A mutations are retained. High frequency C>T|G>A dbSNP transitions are unlikely to be artifacts.

The graph shows the number of such mutations found for each method (red lines), as well as the percentage agreement and the correlation of frequencies for those mutations found with both methods.

For this analysis, C>T|G>A transitions with a frequency above 0.2 were analyzed. There is excellent agreement between both methods for C>T|G>A transitions which are unlikely to be artifacts.

Removal of false positive COSMIC mutations

Sample	Chrom	Pos	COSMIC ID	dpSNP ID	Ref	Variant	Frequency Standard	Frequency GeneRead
Heart	chr5	149435759	COSN167607	rs216136	G	A	0.286	–
	chr10	123243197	COSN163794	rs2278202	G	A	1	–
Kidney	chr3	47061325	COSM1423513	–	C	T	0.143	–
	chr3	128202761	COSM249854	–	C	T	0.528	-
	chr4	55599268	COSM1307	rs55789615	C	T	0.364	–
	chr13	25043207	COSM1365994	rs116339023	G	A	0,14	–
Breast cancer	chr9	139391380	COSM99629	–	G	A	0.103	–
	chr10	43622127	COSM1505473	–	C	T	0.157	–
	chr10	43613912	COSM35666	–	G	A	0.088	–
	chr17	7579433	COSM45837	–	G	A	0.106	–
	chr21	36206778	COSM211651	–	G	A	–	0.167
	chr21	36206818	COSM96514	–	G	A	0.282	–
Prostate cancer	chr2	29446296	COSM28502	–	C	T	–	0.171
	chr7	55249100	COSM53289	–	G	A	0.128	–
	chr9	98248106	COSM1464214	–	C	T	0.111	–
	chr17	7578476	COSM43582	–	G	A	0.273	–
	chr17	37881349	COSM95865	rs140231769	C	T	0.158	–
	chr20	57484421	COSM94388	rs121913495	G	A	0.173	–
	chr21	36206778	COSM211651	–	G	A	0.102	–
Liver cancer	chr2	48032105	COSM13342	–	C	T	0.13	–
	chr4	1807130	COSM327089	–	C	T	0.16	–
	chr12	121426785	COSM46438	–	G	A	0.2	–
	chr16	3828705	COSM970602	–	C	T	0.1	–

The table shows COSMIC mutations found in a variety of old (10–15 years) FFPE tissues. The tissues were processed in parallel using a standard FFPE kit and the new GeneRead 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. Only the discrepant C>T|G>A mutations are shown. Most of the discrepants are found only in the FFPE samples processed using the normal kit, indicating that these COSMIC mutations are false positives.

Conclusion

- FFPE samples show frequent sequence changes due to DNA damage resulting from formalin fixation and storage, most commonly manifesting as a cytosine to thymine transition.
- The majority of these artifacts are low frequency mutations with no clear biological significance, but a significant portion of sequenced mutations with dbSNP and COSMIC identifiers may also be caused by these artifactual changes.
- Removal of 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 yield from small starting amounts and efficiently removes C>T|G>A artifacts resulting from formalin fixation and storage.

Reference
1. Do H, et al. 2013. Clin. Chem. **59**, 1376.

The GeneRead DNA FFPE Kit is intended for molecular biology applications. This product is not intended for the diagnosis, prevention, or treatment of a disease.

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.

1082041 04/2014