Clinical Chemistry 61:1
64–71 (2015)

Mini-Reviews

Sequence Artifacts in DNA from Formalin-Fixed Tissues: Causes and Strategies for Minimization

Hongdo Do[1,2,3*] and Alexander Dobrovic[1,2,3*]

BACKGROUND: Precision medicine is dependent on identifying actionable mutations in tumors. Accurate detection of mutations is often problematic in formalin-fixed paraffin-embedded (FFPE) tissues. DNA extracted from formalin-fixed tissues is fragmented and also contains DNA lesions that are the sources of sequence artifacts. Sequence artifacts can be difficult to distinguish from true mutations, especially in the context of tumor heterogeneity, and are an increasing interpretive problem in this era of massively parallel sequencing. Understanding of the sources of sequence artifacts in FFPE tissues and implementation of preventative strategies are critical to improve the accurate detection of actionable mutations.

CONTENT: This mini-review focuses on DNA template lesions in FFPE tissues as the source of sequence artifacts in molecular analysis. In particular, fragmentation, base modification (including uracil and thymine deriving from cytosine deamination), and abasic sites are discussed as indirect or direct sources of sequence artifacts. We discuss strategies that can be implemented to minimize sequence artifacts and to distinguish true mutations from sequence artifacts. These strategies are applicable for the detection of actionable mutations in both single amplicon and massively parallel amplicon sequencing approaches.

SUMMARY: Because FFPE tissues are usually the only available material for DNA analysis, it is important to maximize the accurate informational content from FFPE DNA. Careful consideration of each step in the work flow is needed to minimize sequence artifacts. In addition, validation of actionable mutations either by appropriate experimental design or by orthogonal methods should be considered.

Recent advances in molecularly targeted therapies have greatly increased the clinical demand for the detection of actionable mutations in cancer patients. Mutational analysis is key for the stratification of cancer patients for appropriate molecularly targeted therapies. Currently, solid tumors that are selectively treated with small molecule inhibitors on the basis of mutational analysis include epidermal growth factor receptor[4] (*EGFR*)-mutant lung cancer, v-raf murine sarcoma viral oncogene homolog B (*BRAF*)-mutant melanoma, and v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (*KIT*)-mutant gastrointestinal stromal cancer *(1)*. In other cancers, such as colorectal cancer, Kirsten rat sarcoma viral oncogene homolog (*KRAS*) mutations indicate intrinsic or emerging nonresponsiveness to EGFR-directed antibodies *(2)*. Thus, accurate detection of these and other clinically actionable mutations is a crucial part of precision medicine.

Formalin-fixed paraffin-embedded (FFPE)[5] tissues are usually the primary material for detection of actionable mutations in solid tumors. Fixation of cancer tissues in buffered formalin (4% formaldehyde) is a standard procedure because formalin fixation preserves tissue and cellular morphology for assessment by anatomical pathologists. It also enables the fixed tissues to be stored at ambient conditions.

However, molecular testing with FFPE DNA is often problematic. In particular, extensive fragmentation significantly reduces the amount of amplifiable templates available for PCR amplification. A second major problem related to FFPE DNA is the occurrence of sequence artifacts, i.e., apparent sequence changes that are not present

[1] Translational Genomics and Epigenomics Laboratory, Olivia Newton-John Cancer Research Institute, Heidelberg, Victoria, Australia; [2] Department of Pathology, University of Melbourne, Parkville, Victoria, Australia; [3] School of Cancer Medicine, La Trobe University, Bundoora, Victoria, Australia.

* Address correspondence to these authors at: Translational Genomics and Epigenomics Laboratory, Olivia Newton-John Cancer Research Institute, Austin Hospital, 145 Studley Rd., Heidelberg, Victoria 3084, Australia. Fax +61-3-9496-5334; e-mail hongdo.do@onjcri.org.au, alex.dobrovic@onjcri.org.au.

Received June 19, 2014; accepted October 14, 2014.
Previously published online at DOI: 10.1373/clinchem.2014.223040

[4] Human genes: *EGFR,* epidermal growth factor receptor; *BRAF,* v-raf murine sarcoma viral oncogene homolog B; *KIT,* v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; *KRAS,* Kirsten rat sarcoma viral oncogene homolog; *UNG,* uracil-DNA glycosylase; *BRCA1,* breast cancer 1, early onset.

[5] Nonstandard abbreviations: FFPE, formalin-fixed paraffin-embedded; SNV, single nucleotide variant; UDG, uracil-DNA glycosylase; 5-mC, 5-methylcytosine; MPS, massively parallel sequencing; qPCR, quantitative real-time PCR; Safe-SeqS, Safe-Sequencing System; UID, unique identifier.

Table 1. The types of sequence artifacts detected in FFPE DNA.

	Artifactual base changes						
Study	C:G>T:A	C:G>A:T	C:G>G:C	A:T>G:C	Others	Gene	Method
Do and Dobrovic *(9)*	60%	35%	0%	5%	0%	*EGFR*	Sanger sequencing
Lamy et al. *(14)*	52%	36%	11%	0%	1%	*KRAS*	SNaPshot multiplex PCR assay
Akabari et al. *(53)*	100%	0%	0%	0%	0%	*UNG*[a]	Sanger sequencing
Wong et al. *(62)*	42%	13%	0%	35%	10%	*BRCA1*	Sanger sequencing
Do et al. *(41)*	80%	3%	2%	6%	9%	48 Genes	Targeted amplicon sequencing

[a] *UNG*, uracil-DNA glycosylase; *BRCA1*, breast cancer 1, early onset.

in the original sample (Table 1). Several studies have demonstrated that the number of sequence variants seen in formalin-fixed tissues is higher than that in matched frozen tissues *(3, 4)*.

It is often difficult to distinguish true sequence changes from artifactual sequence changes, thus increasing the risk of false-positive mutation calls *(5, 6)*. In some cases, sequence artifacts can be falsely interpreted as clinically important mutations. Tsao and colleagues reported multiple novel *EGFR* mutations in FFPE DNA *(7)* that have never been found in over 2000 fresh-frozen samples of non–small cell lung cancer *(8)*. Other studies have confirmed that multiple artifactual sequence alterations in the *EGFR* gene can arise in FFPE lung tissues *(9, 10)*. A systemic review on the 3381 somatic *EGFR* mutations detected in 12 244 patients with non–small cell lung cancer found that 71% of the *EGFR* mutations were seen in only a single case *(11)*, suggesting that many of the reported *EGFR* mutations may be sequence artifacts.

Importantly, sequence artifacts can display the same base changes as recurrent canonical mutations. For example, *KRAS* mutations and *EGFR* T790M mutations are predictive markers for resistance to anti-EGFR monoclonal antibodies and EGFR tyrosine kinase inhibitors, respectively *(12, 13)*. Sequence artifacts corresponding to *KRAS* and *EGFR* T790M mutations have been reported in DNA from formalin-fixed colorectal and non–small cell lung cancers *(14, 15)*. Lamy and colleagues reported that artifactual codon 12 and 13 *KRAS* variants were present in 53 of 993 (5%) formalin-fixed colorectal cancers *(14)*. Similarly, Ye and colleagues found a high rate of *EGFR* T790M mutations in formalin-fixed lung tumors (41.7%, 15 of 36 cases) and adjacent normal tissues (48.5%, 16 of 33 cases), but in only 1 matching fresh-frozen lung tumor *(15)*, indicating that there were false positives in the tested formalin-fixed tissues.

Sequence artifacts can arise from various sources, including damaged templates preexisting in FFPE DNA *(3, 4)*, oxidative DNA damage during sample preparation *(16)*, DNA polymerase error *(17)*, pseudogene amplification *(18)*, adaptor sequences and adaptor chimeras *(19)*, sequencing chemistry *(20)*, sequence alignment *(21)*, and spontaneous deamination of nucleotides during thermocycling *(22, 23)*. Understanding these issues is important for accurate detection of actionable mutations and thus for implementation of precision medicine into the clinic. In this mini-review we specifically focus on preexisting damage to template DNA as a major source of sequence artifacts in FFPE DNA and discuss the strategies for minimization of sequence artifacts generated from damaged FFPE DNA.

Types of DNA Damage in Formalin-Fixed Tissues

Several types of DNA damage have been identified in formalin-fixed tissues as sources of sequence artifacts (Fig. 1) and this section addresses these DNA damage types in more detail.

FORMALDEHYDE-INDUCED CROSSLINKS

Formaldehyde, the main component of formalin, is a reactive electrophilic chemical that creates various crosslinks between intracellular macromolecules such as protein and DNA *(24)*. The formaldehyde-induced crosslinks include protein–protein, protein–DNA, and DNA–formaldehyde adducts and interstrand DNA crosslinks. The interaction of formaldehyde with the functional groups of amino acids (e.g., primary amines and thiols) forms methylol adducts that can further crosslink with other amino acids through methylene bridge formation *(25)*.

Formaldehyde also crosslinks DNA by reacting with the imino groups of DNA bases *(26)*. Because the atoms in the imino groups are involved in hydrogen bonds mediating base pairing, formaldehyde-induced DNA adducts weaken the bonding strength of double-stranded DNA by reducing the number of hydrogen bonds in the

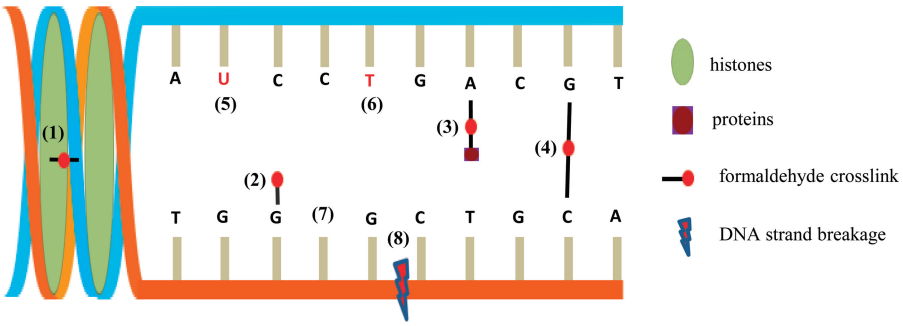

Fig. 1. DNA damage present in formalin-fixed tissues.

DNA extracted from formalin-fixed tissues contains various types of damage. Formaldehyde, the main component of formalin, is highly reactive with DNA bases and proteins, generating histone–DNA crosslinks (1), formaldehyde–DNA adducts (2), DNA–protein crosslinks (3), and DNA–DNA crosslinks (4). Uracil (5) and thymine (6), which result from deamination of cytosine and 5-mC, respectively, are also present in FFPE DNA. DNA bases are also lost, resulting in abasic sites (7), and DNA strands are broken, leading to fragmentation of DNA (8).

DNA double helix *(27)*. Furthermore, crosslinking of DNA bases with nearby histones, a dominant form of DNA damage in formaldehyde-exposed cells, results in a conformational change of DNA *(28)*. Thus, formaldehyde-induced crosslinks of DNA reduce the stability of double-stranded DNA, resulting in a partial denaturation of DNA *(27)*.

DNA FRAGMENTATION

Fragmentation, often extensive, is the common form of DNA damage found in formalin-fixed tissues *(5)*. Fragmentation of DNA in formalin-fixed tissues was shown to be increased with longer storage time and lower pH of formalin used in tissue fixation *(29)*. Compared to DNA from fresh formalin-fixed tissues, the PCR success rate of DNA from older formalin-fixed tissues was shown to be decreased *(29)*, indicating that DNA fragmentation may continuously occur during storage. Fragmentation damage in FFPE DNA directly influences the amount of templates available for PCR amplification *(30)*. Thus, the same quantity of FFPE DNA from different samples may contain significantly different amounts of amplifiable templates, depending on the degree of fragmentation damage *(6)*.

ABASIC SITES

Formaldehyde is readily oxidized to formic acid in the reaction with atmospheric oxygen. The formation of formic acid reduces the pH of formalin. Formalin is thus usually buffered to maintain a neutral pH level. The *N*-glycosidic bonds of the purine bases to the sugar backbone are susceptible to hydrolysis at low pH *(31)*, generating abasic sites in the DNA. Thus, fixation of tissues in unbuffered formalin will significantly lower the amount of amplifiable DNA templates *(32)*.

The depurination rate of single-stranded DNA is 4 times higher than that of double-stranded DNA *(33)*. Purine bases at the terminals of DNA strands are more readily depurinated than those located at internal positions *(31)*. The aldehyde residue of abasic sites can generate an interstrand crosslink by reacting with the exocyclic amino group of a guanine base *(34)*. Furthermore, abasic sites in DNA strongly destabilize the double helix *(35)*, leading to local denaturation of the DNA. Because the rate of DNA damage in single-stranded DNA is higher than in double-stranded DNA, the DNA denaturation, induced by formaldehyde, may promote further DNA damage.

Abasic sites cause problems in sequence analysis. DNA polymerases have generally low bypass efficiencies at abasic sites *(36)*, preventing amplification of DNA templates with abasic sites *(37)*. However, when DNA polymerases read through abasic sites, sequence artifacts can be generated. Adenines are preferentially incorporated opposite to abasic sites by many DNA polymerases, but guanines or short deletions (1 to 3 bases) are also incorporated to a lesser extent *(36)*. As a result, various types of artifactual single nucleotide variants (SNVs) and deletions can arise from abasic sites. In addition, abasic sites can undergo spontaneous cleavage through the β-elimination reaction leading to breakage of DNA strands *(38)*.

DEAMINATION OF CYTOSINE BASES

Hydrolytic deamination of cytosine bases to uracil takes place at an estimated rate of 70–200 events/day in a living cell *(39)*. In living cells, uracil lesions in DNA are removed by uracil-DNA glycosylase (UDG). In the resulting abasic site, the cytosine is then correctly restored by base excision repair due to the guanine in the com-

plementary strand. However, when cytosine is deaminated outside the context of a living cell, the uracil lesions remain unrepaired. When DNA templates with uracil lesions are amplified by PCR, artifactual C:G>T:A SNVs are generated because DNA polymerase incorporates an adenine opposite to the uracil lesions.

Recently, uracil lesions have been identified as major sources of sequence artifacts in FFPE DNA *(40–42)*. Among the sequence artifacts detected in FFPE DNA, transitional C:G>T:A variants are the most frequent type of SNVs. Sequence artifacts are more readily detectable when low copy numbers of FFPE DNA are tested *(9, 18)*, as is often the case in amplicon-based protocols. Such artifactual C:G>T:A variants can be markedly reduced after treatment of FFPE DNA with UDG before PCR amplification, indicating that uracil lesions are a major source of artifactual C:G>T:A variants in FFPE DNA *(40–42)*.

The cytosine in CpG dinucleotides is often present as 5-methylcytosine (5-mC). When 5-mC undergoes hydrolytic deamination, it is converted to thymine, generating a T:G mismatch in the DNA *(43)*. The deamination rate of 5-mC is approximately 2-fold higher than for unmethylated cytosine in double-stranded DNA *(44)*. The 5-mC base is the most susceptible DNA base to deamination damage *(45)*. Deamination of 5-mC to thymine causes artifactual C:G>T:A SNVs because DNA polymerase incorporates an adenine opposite to the thymine lesions. It is not surprising that high levels of artifactual C:G>T:A SNVs are found at CpG dinucleotide sites in FFPE DNA, strongly indicative of deamination of 5-mC bases *(41)*.

Strategies for Minimization of Sequence Artifacts from FFPE DNA

Minimizing sequence artifacts is crucial for the accurate detection of actionable mutations in formalin-fixed clinical tissues. Accurate detection of actionable mutations enables the identification of patients who will respond to targeted treatments but will also avoid unnecessary adverse effects arising from inappropriate treatment of nonresponsive patients. Suggested strategies for the minimization of sequence artifacts are discussed in this section and summarized in Table 2.

PREANALYTIC ASSESSMENT OF FFPE DNA

Preanalytic assessment of FFPE DNA is crucial to both optimizing the experimental conditions for mutation detection and ensuring a considered interpretation of results. The key components of preanalytic assessment are review of tumor tissues by an experienced pathologist and estimation of amplifiable templates.

Pathological review is necessary to identify tumor-rich areas for macrodissection or coring of tumor tissues and to estimate the tumor purity within the sampled area. Tumor purity information is important for interpretation of the results because mutations will be present at lower frequency if there is a predominant amount of normal tissue. The analytic sensitivity of detection methods used for molecular testing varies substantially, and the minimum required levels of tumor purity differ depending on the detection method used. For Sanger sequencing, a minimum purity of 20% is desirable but deep sequencing by massively parallel sequencing (MPS) allows mutations to be detectable with lower tumor purities. However, at these lower tumor purities, it becomes more difficult to distinguish sequence artifacts from true mutations because both will be present at similar frequencies.

The quantity of DNA can be measured by spectrophotometry or fluorometry. Importantly, the same measured quantities of DNA from different FFPE samples can contain widely different amounts of amplifiable templates, depending on the degree of fragmentation *(6)*. Both spectrophotometry and fluorometry tend to overestimate, often seriously, the actual amount of amplifiable templates in FFPE DNA *(18)*. For this reason, PCR-based methods such as quantitative real-time PCR (qPCR) and digital PCR are preferable to quantify the amount of amplifiable templates in FFPE DNA. The amplicon size used in estimating targets should reflect the mean amplicon size used in the sequencing protocol *(18, 46)*.

Information on the number of amplifiable templates will enable determination as to how reliable the accurate detection of variants might be. In low amplifiable template situations, the allele frequency of true variants cannot be reliably measured because of stochastic variation in allelic representation. Thus, the fewer the templates used in mutation analysis, the higher the risk of false negatives, especially in the case of low tumor purity, in which even true mutations are present at reduced frequencies. In addition, artifactual sequence variants arising from DNA damage will be more frequently detected because of stochastic enrichment in the low copy number context, increasing the risk of false positives *(9, 18)*.

REMOVAL OF CROSSLINKS BY HEAT TREATMENT

Formaldehyde-induced DNA–DNA and DNA–protein crosslinks adversely affect the isolation of DNA from formalin-fixed tissues and the amount of amplifiable DNA templates by PCR. Formaldehyde-induced crosslinks are reversible by heat treatment *(47)*. The reversal rate of formaldehyde crosslinks is closely dependent on the temperature and pH of the buffer solution *(48, 49)*. The half-life of formaldehyde crosslinks is inversely correlated with temperature *(48)*. High-

Table 2. Strategies for minimization of sequence artifacts from FFPE DNA.

Step	Strategy
DNA extraction	Assessment of tumor purity and identification of tumor-enriched areas by a pathologist
	Macrodissection or coring of the tumor-enriched areas
	Use of sufficient tissue, whenever possible, to ensure that a sufficient quantity of DNA is isolated for subsequent molecular testing
	Heat treatment to remove formaldehyde-induced crosslinks and to facilitate subsequent tissue digestion with proteinase
	Extended proteinase K treatment to digest tissue and to remove proteins cross-linked to DNA
DNA assessment	Assessment of double-stranded DNA quantity using fluorometry
	Quantification of amplifiable templates using qPCR or digital PCR, especially for massively parallel sequencing. Use amplicons sizes that correspond to the mean amplicon size of the sequencing assay
Sample library preparation	In vitro removal of uracil prior to PCR amplification of FFPE DNA
	Using assays generating short amplicons to increase the number of templates for PCR
	Capture-based target enrichment allowing the recognition of the initial templates in sequence reads using their unique start and end sites
	Using primers specific for each strand of the DNA template in amplicon-based target enrichment approach
	Molecularly tagging DNA templates for identification of sequence artifacts
PCR amplification	Use of specific DNA polymerases (e.g. Pfu and KAPA) that have low bypass efficiency over DNA lesions such as uracil and abasic sites
	Use a high-fidelity DNA polymerase to reduce polymerase errors
Validation of sequence variants from amplicon-based MPS	Running each test in duplicate so that separate pools of templates are used
	Using orthogonal methods for clinically actionable mutations

temperature heating methods, usually at >90 °C, have been shown to be effective not only for the yield of DNA *(50)*, but also for the yield of amplifiable templates from FFPE tissues *(49, 51)*.

IN VITRO REMOVAL OF MODIFIED BASES USING DNA-GLYCOSYLASES

Transitional C:G>T:A SNVs are the most frequent sequence artifacts arising from deamination of cytosine in FFPE DNA *(8, 40)*. In vitro removal of uracil bases from FFPE DNA using UDG before PCR amplification markedly reduces the artifactual C:G>T:A SNVs *(40)*. This can be as high as 60%–80% in some FFPE DNAs *(41)*.

UDG removes uracil bases from U:G mismatches in double-stranded DNA, generating abasic sites. The resulting abasic sites significantly hinder the amplification of templates by diminishing the DNA polymerase extension rate and causing thermal cleavage of templates under PCR cycling conditions *(36)*. Thus, UDG pretreatment of FFPE DNA before PCR amplification can selectively prevent the enrichment of artifactual sequence reads from uracil lesions when a polymerase is used that does not read through abasic sites.

After UDG treatment, a certain number of artifactual C>T changes can still be observed. Many of these are at CpG sites which are presumably methylated *(40, 41)*. Thymine lesions generated by deamination of 5-mC are theoretically removable from double-stranded DNA using either of the base excision repair enzymes MBD4 and thymine-DNA glycosylase *(52)*. Up to now, this strategy has not, to our knowledge, been used in the context of FFPE DNA.

USING SHORT AMPLICONS

There is a significant relationship between the amount of amplifiable DNA used in mutational analysis and the frequency of sequence artifacts *(9, 18)*. Sequence artifacts in FFPE DNA are observed more frequently as the number of input DNA templates decreases *(53)*. The fewer amplifiable templates that are used in molecular analysis, the more chance that DNA templates with lesions leading to subsequent sequencing errors will be de-

tected above the background noise level as a consequence of stochastic variation *(9)*.

Fragmentation of FFPE DNA directly influences the amount of amplifiable templates available for PCR *(30)*. The PCR success rate of FFPE DNA is strongly correlated with the size of the amplicon *(49, 54)*, confirming the benefit of designing shorter amplicons. Thus, the use of short amplicon (e.g., 120 bp or less) should be considered to maximize the number of templates to be used for PCR.

REDUCED AMPLIFICATION OF DAMAGED TEMPLATES BY HIGH FIDELITY POLYMERASES

DNA polymerases with low bypass efficiencies over various DNA lesions can be used for reduction of sequence artifacts. Many DNA polymerases incorporate adenine opposite to uracil during extension. However, other DNA polymerases, especially from the family B DNA polymerases (e.g., Pfu and KAPA), have a read-ahead function to recognize uracil lesions and terminate extension before misincorporation of adenine *(55)*. Pfu polymerase has been shown to terminate the extension over uracil lesions in 70%–99% of templates *(36)*. The bypass efficiency of DNA polymerases also differs markedly at abasic sites *(36)*. Thus, the use of specific DNA polymerases with low bypass efficiencies over various DNA lesions would be a simple but effective way to minimize sequence artifacts generated from the DNA lesions.

SEQUENCING BOTH STRANDS OF DNA

Accuracy in variant calling can be improved by protocols that have the capacity to sequence the sense and antisense strands of target sequences independently, e.g., molecular inversion probes and other extension-ligation techniques. Because each template lesion will be present in only 1 of the DNA strands, this approach can distinguish these and other sequence artifacts arising from DNA lesions from true mutations.

MOLECULAR TAGGING OF DNA TEMPLATES

Tagging of DNA templates with unique sequences is a powerful approach which effectively reduces sequence artifacts. Recently, 2 methods, Safe-Sequencing System (Safe-SeqS) and duplex sequencing, were reported to enable more sensitive and accurate rare variant detection by stringently eliminating sequence errors *(17, 56)*.

In the Safe-SeqS method, individual single-stranded DNAs are tagged with a unique identifier (UID) of 14-bp degenerate sequences to allow the tracking of the initiating templates in the sequence reads *(17)*. The traceability of templates in their sequence reads enables the allelic frequency of each sequence variant to be readily counted. True mutations are present in all daughter molecules, whereas any errors introduced during the various experimental steps are present only in a lower proportion of a UID family. By eliminating variants present at less than <95% in the sequence reads with the same UID, this approach has been shown to reduce the error rate by approximately 20-fold *(17)*.

In duplex sequencing, both the sense and the antisense strands of each DNA template are tagged with a unique double-stranded sequence at each end *(56)*. This strategy enables any sequence variant detected in 1 strand of DNA to be crosschecked using the corresponding sequence reads from the other strand of the same template. Thus, sequence artifacts arising from DNA lesions are readily distinguished from true mutations because artifactual sequence variants are detectable in only 1 of the strands but true mutations are present in both strands. Duplex sequencing thus enables sequence artifacts to be readily recognized by their strand specificity, resulting in exquisite sensitivities of 1 mutant molecule in 10 000 wild-type molecules *(56)*.

CAPTURE-BASED SEQUENCING APPROACHES

Amplicons generated by PCR do not retain the information on the number of initiating templates of sequence reads, making it difficult to distinguish true mutations from sequence artifacts without adequate validation. For this reason, a capture-based approach is particularly helpful because the varying insert sequences of each captured template enable the differentiation of the templates *(5)*. Because all of the sequence reads from the same template have the same insert sequence, the number of independent templates harboring the same sequence variants can be readily determined using a bioinformatic tool like Picard.

An important advantage of the capture-based approach is that the capture baits are shorter than amplicons and can be overlapped, enabling more templates to be captured. Orthogonal validation is usually not required when a mutation is seen in multiple independent templates. If a sequence variant is found in only 1 template, although detected in multiple sequence reads of the same template, the variant should be interpreted with caution because it may be a sequence artifact. Although the capture-based approach requires more setup time and may require shearing of DNA before library generation, it is amenable to automation *(57)*. It is thus possible that capture-based approaches will become the preferred technology to analyze FFPE DNA for mutations.

Validation of Sequence Variants

Sequence artifacts are often present above the intrinsic background level of sequencing variation of MPS, which is operationally considered to be 1% *(23)*. It can be difficult to distinguish artifacts from true low-level mutations which are present as the result of low tumor purity or tumor heterogeneity. Bioinformatic filtering has been

suggested as a potential strategy for artifact reduction *(58, 59)*, but the bioinformatic removal of sequence variants can also increase the risk of false negatives for clinically important mutations, especially resistance mutations present at low levels.

When locus-specific singleplex assays are used in mutational analysis, all sequence variants detected can readily be verified by using independent PCR products. However, validation of every single variant detected in MPS-based approaches is not feasible because of the high number of sequence variants.

It is, however, desirable to validate clinically important (actionable) mutations. The simplest approach is to consider only variants present in 2 independent MPS runs of separate aliquots of each sample *(60)*. Another MPS methodology can also be used for validation *(60)* but is unlikely to be practical in diagnostic situations. Alternatively, validation can be built into the assay design, e.g., if the design of the assay means that independent templates can be identified, as in capture-based methods or by molecular tagging, or with strand-specific amplification.

Orthogonal methodologies not using MPS can also be used, particularly when only 1 or a few mutations need to be validated. Typically, a singleplex sequencing method such as Sanger sequencing or pyrosequencing is used, although the lower sensitivity can be an issue *(61)*.

Conclusions

Detection of actionable mutations from formalin-fixed tissues is often problematic because of sequence artifacts arising from DNA damage. DNA fragmentation not only reduces the amount of amplifiable templates but also increases the sequence artifact rate due to stochastic enrichment of artifactual changes.

A number of measures need to be implemented to reduce the danger of false-positive and false-negative calls in the diagnostic context. First, preanalytical assessment of amplifiable templates in FFPE DNA should be implemented into the work flow for reliable interpretation of mutational results. For conventional amplicon-based approaches, the removal of damaged templates is desirable directly and/or indirectly by the use of enzymes that do not read through modified or abasic sites. Marked reduction of C:G>T:A sequence artifacts by UDG pretreatment of FFPE DNA in combination with an enzyme that does not read through abasic sites demonstrates the validity of this approach.

To avoid false positives arising from sequence artifacts, sequence variants detected in FFPE DNA may need to be validated by 1 of several approaches, such as molecular barcodes to tag individual DNA templates to enable the origin of templates to be traced from the sequence reads or duplicate sequencing reads. It may be possible to simultaneously repair the multiple types of DNA damage seen in formalin-fixed tissues using a mixture of multiple DNA repair enzymes. Implementation of these approaches in mutational analysis will greatly improve accurate detection of clinically important mutations in formalin-fixed tissues.

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.
Research Funding: The work that underlies this review was supported by funding from the Cancer Council of Victoria and Cancer Australia.
Expert Testimony: None declared.
Patents: None declared.

Acknowledgments: We thank Jonathan Weiss, Giada Zapparoli, and Tom Witkowski for their critical reading of this manuscript.

References

1. Ferte C, Andre F, Soria JC. Molecular circuits of solid tumors: prognostic and predictive tools for bedside use. Nat Rev Clin Oncol 2010;7:367–80.
2. Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N Engl J Med 2008;359:1757–65.
3. Williams C, Ponten F, Moberg C, Soderkvist P, Uhlen M, Ponten J, et al. A high frequency of sequence alterations is due to formalin fixation of archival specimens. Am J Pathol 1999;155:1467–71.
4. Quach N, Goodman MF, Shibata D. In vitro mutation artifacts after formalin fixation and error prone translesion synthesis during PCR. BMC Clin Pathol 2004;4:1.
5. Wong SQ, Li J, Salemi R, Sheppard KE, Do H, Tothill RW, et al. Targeted-capture massively-parallel sequencing enables robust detection of clinically informative mutations from formalin-fixed tumours. Sci Rep 2013;3:3494.
6. Wong SQ, Li J, Tan AY, Vedururu R, Pang JM, Do H, et al. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. BMC Med Genomics 2014;7:23.
7. Tsao MS, Sakurada A, Cutz JC, Zhu CQ, Kamel-Reid S, Squire J, et al. Erlotinib in lung cancer: molecular and clinical predictors of outcome. N Engl J Med 2005;353:133–44.
8. Marchetti A, Felicioni L, Buttitta F. Assessing EGFR mutations. N Engl J Med 2006;354:526–8; author reply 526–8.
9. Do H, Dobrovic A. Limited copy number-high resolution melting (LCN-HRM) enables the detection and identification by sequencing of low level mutations in cancer biopsies. Mol Cancer 2009;8:82.
10. Gallegos Ruiz MI, Floor K, Rijmen F, Grunberg K, Rodriguez JA, Giaccone G. EGFR and K-ras mutation analysis in non-small cell lung cancer: comparison of paraffin embedded versus frozen specimens. Cell Oncol 2007;29:257–64.
11. Murray S, Dahabreh IJ, Linardou H, Manoloukos M, Bafaloukos D, Kosmidis P. Somatic mutations of the ty-

rosine kinase domain of epidermal growth factor receptor and tyrosine kinase inhibitor response to TKIs in non-small cell lung cancer: an analytical database. J Thorac Oncol 2008;3:832–9.

12. Misale S, Yaeger R, Hobor S, Scala E, Janakiraman M, Liska D, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature 2012;486:532–6.
13. Kobayashi S, Boggon TJ, Dayaram T, Janne PA, Kocher O, Meyerson M, et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N Engl J Med 2005;352:786–92.
14. Lamy A, Blanchard F, Le Pessot F, Sesboue R, Di Fiore F, Bossut J, et al. Metastatic colorectal cancer KRAS genotyping in routine practice: results and pitfalls. Mod Pathol 2011;24:1090–100.
15. Ye X, Zhu ZZ, Zhong L, Lu Y, Sun Y, Yin X, et al. High T790M detection rate in TKI-naive NSCLC with EGFR sensitive mutation: truth or artifact? J Thorac Oncol 2013;8:1118–20.
16. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res 2013; 41:e67.
17. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A 2011;108:9530–5.
18. Sah S, Chen L, Houghton J, Kemppainen J, Marko AC, Zeigler R, Latham GJ. Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies. Genome Med 2013;5:77.
19. Kircher M, Heyn P, Kelso J. Addressing challenges in the production and analysis of Illumina sequencing data. BMC Genomics 2011;12:382.
20. Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: promises and challenges. Cancer Lett 2013;340:284–95.
21. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 2008;18:1851–8.
22. Hogrefe HH, Hansen CJ, Scott BR, Nielson KB. Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation. Proc Natl Acad Sci U S A 2002;99:596–601.
23. Lin MT, Mosier SL, Thiess M, Beierl KF, Debeljak M, Tseng LH, et al. Clinical validation of KRAS, BRAF, and EGFR mutation detection using next-generation sequencing. Am J Clin Pathol 2014;141:856–66.
24. Feldman MY. Reactions of nucleic acids and nucleoproteins with formaldehyde. Prog Nucleic Acid Res Mol Biol 1973;13:1–49.
25. Fraenkel-Conrat H, Olcott HS. The reaction of formaldehyde with proteins; cross-linking between amino and primary amide or guanidyl groups. J Am Chem Soc 1948;70:2673–84.
26. McGhee JD, von Hippel PH. Formaldehyde as a probe of DNA structure. r. Mechanism of the initial reaction of formaldehyde with DNA. Biochemistry 1977;16: 3276–93.
27. McGhee JD, von Hippel PH. Formaldehyde as a probe of DNA structure. II. Reaction with endocyclic imino groups of DNA bases. Biochemistry 1975;14:1297–303.
28. Ohba Y, Morimitsu Y, Watarai A. Reaction of formaldehyde with calf-thymus nucleohistone. Eur J Biochem 1979;100:285–93.
29. Ludyga N, Grunwald B, Azimzadeh O, Englert S, Hofler H, Tapio S, Aubele M. Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. Virchows Arch 2012;460:131–40.
30. Didelot A, Kotsopoulos SK, Lupo A, Pekin D, Li X, Atochin I, et al. Multiplex picoliter-droplet digital PCR for quantitative assessment of DNA integrity in clinical samples. Clin Chem 2013;59:815–23.
31. Suzuki T, Ohsumi S, Makino K. Mechanistic studies on depurination and apurinic site chain breakage in oligodeoxyribonucleotides. Nucleic Acids Res 1994;22: 4997–5003.
32. Zsikla V, Baumann M, Cathomas G. Effect of buffered formalin on amplification of DNA from paraffin wax embedded small biopsies using real-time PCR. J Clin Pathol 2004;57:654–6.
33. Lindahl T, Nyberg B. Rate of depurination of native deoxyribonucleic acid. Biochemistry 1972;11:3610–8.
34. Dutta S, Chowdhury G, Gates KS. Interstrand cross-links generated by abasic sites in duplex DNA. J Am Chem Soc 2007;129:1852–3.
35. Vesnaver G, Chang CN, Eisenberg M, Grollman AP, Breslauer KJ. Influence of abasic and anucleosidic sites on the stability, conformation, and melting behavior of a DNA duplex: correlations of thermodynamic and structural data. Proc Natl Acad Sci U S A 1989;86:3614–8.
36. Heyn P, Stenzel U, Briggs AW, Kircher M, Hofreiter M, Meyer M. Road blocks on paleogenomes-polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. Nucleic Acids Res 2010;38: e161.
37. Sikorsky JA, Primerano DA, Fenger TW, Denvir J. DNA damage reduces Taq DNA polymerase fidelity and PCR amplification efficiency. Biochem Biophys Res Commun 2007;355:431–7.
38. Lindahl T, Andersson A. Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. Biochemistry 1972;11:3618–23.
39. Kavli B, Otterlei M, Slupphaug G, Krokan HE. Uracil in DNA–general mutagen, but normal intermediate in acquired immunity. DNA Repair 2007;6:505–16.
40. 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.
41. Do H, Wong SQ, Li J, Dobrovic A. Reducing sequence artifacts in amplicon-based massively parallel sequencing of formalin-fixed paraffin-embedded DNA by enzymatic depletion of uracil-containing templates. Clin Chem 2013;59:1376–83.
42. Chen G, Mosier S, Gocke CD, Lin MT, Eshleman JR. Cytosine deamination is a major cause of baseline noise in next-generation sequencing. Mol Diagn Ther 2014.
43. Duncan BK, Miller JH. Mutagenic deamination of cytosine residues in DNA. Nature 1980;287:560–1.
44. Shen JC, Rideout WM 3rd, Jones PA. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res 1994;22:972–6.
45. Rideout WM 3rd, Coetzee GA, Olumi AF, Jones PA. 5-methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. Science 1990; 249:1288–90.
46. Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, et al. Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat Methods 2013;10:1003–5.
47. Jackson V. Studies on histone organization in the nucleosome using formaldehyde as a reversible cross-linking agent. Cell 1978;15:945–54.
48. Kennedy-Darling J, Smith LM. Measuring the formaldehyde protein-DNA cross-link reversal rate. Anal Chem 2014;86:5678–81.
49. Shi SR, Cote RJ, Wu L, Liu C, Datar R, Shi Y, et al. DNA extraction from archival formalin-fixed, paraffin-embedded tissue sections based on the antigen retrieval principle: heating under the influence of pH. J Histochem Cytochem 2002;50:1005–11.
50. Sepp R, Szabo I, Uda H, Sakamoto H. Rapid techniques for DNA extraction from routinely processed archival tissue for use in PCR. J Clin Pathol 1994;47:318–23.
51. Wu L, Patten N, Yamashiro CT, Chui B. Extraction and amplification of DNA from formalin-fixed, paraffin-embedded tissues. Appl Immunohistochem Mol Morphol 2002;10:269–74.
52. Yoon JH, Iwai S, O'Connor TR, Pfeifer GP. Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair. Nucleic Acids Res 2003;31:5399–404.
53. Akbari M, Hansen MD, Halgunset J, Skorpen F, Krokan HE. Low copy number DNA template can render polymerase chain reaction error prone in a sequence-dependent manner. J Mol Diagn 2005;7:36–9.
54. Gillio-Tos A, De Marco L, Fiano V, Garcia-Bragado F, Dikshit R, Boffetta P, Merletti F. Efficient DNA extraction from 25-year-old paraffin-embedded tissues: study of 365 samples. Pathology 2007;39:345–8.
55. Greagg MA, Fogg MJ, Panayotou G, Evans SJ, Connolly BA, Pearl LH. A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. Proc Natl Acad Sci U S A 1999;96:9045–50.
56. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci U S A 2012;109:14508–13.
57. Metzker ML. Sequencing technologies: the next generation. Nat Rev Genet 2010;11:31–46.
58. Li M, Stoneking M. A new approach for detecting low-level mutations in next-generation sequence data. Genome Biol 2012;13:R34.
59. Harismendy O, Schwab RB, Bao L, Olson J, Rozenzhak S, Kotsopoulos SK, et al. Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. Genome Biol 2011;12:R124.
60. Robasky K, Lewis NE, Church GM. The role of replicates for error mitigation in next-generation sequencing. Nat Rev Genet 2014;15:56–62.
61. Querings S, Altmuller J, Ansen S, Zander T, Seidel D, Gabler F, et al. Benchmarking of mutation diagnostics in clinical lung cancer specimens. PLoS One 2011;6: e19601.
62. Wong C, DiCioccio RA, Allen HJ, Werness BA, Piver MS. Mutations in BRCA1 from fixed, paraffin-embedded tissue can be artifacts of preservation. Cancer Genet Cytogenet 1998;107:21–7.