Journal of Histochemistry & Cytochemistry 2020, Vol. 68(8) 543–552

DOI: 10.1369/0022155420945050
journals.sagepub.com/home/jhc

Why Formalin-fixed, Paraffin-embedded Biospecimens Must Be Used in Genomic Medicine: An Evidence-based Review and Conclusion

William Mathieson and Geraldine A. Thomas
Integrated Biobank of Luxembourg, Dudelange, Luxembourg (WM) and Department of Surgery and Cancer, Imperial College London, London, UK (GAT)

Summary
Fresh-frozen tissue is the "gold standard" biospecimen type for next-generation sequencing (NGS). However, collecting frozen tissue is usually not feasible because clinical workflows deliver formalin-fixed, paraffin-embedded (FFPE) tissue blocks. Some clinicians and researchers are reticent to embrace the use of FFPE tissue for NGS because FFPE tissue can yield low quantities of degraded DNA, containing formalin-induced mutations. We describe the process by which formalin-induced deamination can lead to artifactual cytosine (C) to thymine (T) and guanine (G) to adenine (A) (C:G > T:A) mutation calls and perform a literature review of 17 publications that compare NGS data from patient-matched fresh-frozen and FFPE tissue blocks. We conclude that although it is indeed true that sequencing data from FFPE tissue can be poorer than those from frozen tissue, any differences occur at an inconsequential magnitude, and FFPE biospecimens can be used in genomic medicine with confidence. (J Histochem Cytochem 68: 543–552, 2020)

Keywords
cytosine, deamination, formaldehyde, UDG, uracil-DNA glycosylase

Fixation in formalin-based solutions, followed by embedding in paraffin wax to make formalin-fixed, paraffin-embedded (FFPE) tissue blocks, is the gold standard method of preservation of human tissues for diagnosis. Processing material in this manner has a number of advantages, ranging from mitigating risks of infectious agents that may be present in the fresh material to ensuring preservation of the architectural components of the tissue.[1] Embedding in paraffin wax enables thin sections to be cut and the architecture of the tissue to be examined using simple dyes, such as hematoxylin and eosin, to delineate different components of the cell. The majority of cancer diagnosis still depends on the ability to link changes in the components of normal tissue architecture with different stages of disease and is further enhanced by the use of immunohistochemistry to examine changes in the abundance of key proteins associated with cellular function. The consistency of diagnosis across institutions is ensured by the use of individual testing protocols that have been validated on FFPE tissues and audited by appropriate bodies such as CLIA (Clinical Laboratory Improvement Amendments) in the United States and NEQAS (National External Quality Assessment Service) in the United Kingdom.

However, more recently, with the advent of tests based on genetic sequence rather than protein antigens, a return to the use of fresh-frozen (FF) tissue as the diagnostic biospecimen of choice has been advocated by some. A number of difficulties in the derivation of gene sequence data from FFPE material have been identified. These are primarily issues with chemical crosslinking that ensues from formalin fixation, DNA fragmentation, and deamination of cytosine (C) bases and generation of abasic sites, as discussed

Received for publication April 30, 2019; accepted July 6, 2020.

Corresponding Author:
William Mathieson, Integrated Biobank of Luxembourg, 1 rue Louis Rech, L-3555 Dudelange, Luxembourg.
E-mail: william.mathieson@ibbl.lu

later in this article. These can cause misinterpretation of the DNA sequence, particularly with respect to increased identification of C to thymine (T) and guanine (G) to adenine (A) (C:G > T:A) mutations which result from deamination.

Although it is commonly expected that knowledge of the DNA sequence is likely to provide more information than assessment of tissue architecture, protein abundance, and position within the cell, this has yet to be proven. Diagnosis will therefore continue to be carried out on FFPE material using histology and immunohistochemistry until a more appropriate method has been validated. This in itself presents a chicken-and-egg challenge on deciding the appropriate biospecimen format going forward. Improvements in cancer diagnosis (such as screening using sensitive techniques like ultrasound) and the use of neoadjuvant chemotherapy to shrink tumors to increase the chances of complete surgical removal and reduce damage to normal tissue structures have led to less tissue becoming available through pathology departments. The pathologist requires sufficient material to make his or her diagnosis and to complete the use of validated diagnostic tests on that material, some of which may guide future treatment of the patient, for example, PLD1 staining to guide immunotherapy.[2]

Therefore, obtaining a representative sample of tumor for freezing may not be practical in many instances, and even when it is practical, there is additional expense involved in its collection and ongoing storage. Comparison of a genetic test carried out on FF tissue with normal diagnosis and any residual FFPE material from the same tumor can be very challenging. In the pilot study for the UK's 100,000 Genomes Project, which was designed to identify whether FF or FFPE material should be used for clinical whole genome sequencing, the largest dropout (87 of 184; 48%) was due to lack of provision of a suitable FF sample.[3] In some cases, this was due to low cellularity (below 40%). Fewer patients (30 of the remaining 97; 31%) were excluded due to poor quality of DNA extracted from FFPE and a further 15 (15%) as a result of further quality issues such as low DNA yield and poor library preparation.

The assessment of epithelial content when fresh material is sampled is something of a lottery—cellularity can only be assessed properly once the material has been rendered solid enough (i.e., frozen) to cut and stain a section of the material. Even then, assessing tumor stage and grade using an FF tissue section is much less precise than with an FFPE section. Although frozen sections are unquestionably valuable when immediate intraoperative diagnosis is required, they have repeatedly been associated with higher diagnostic error rates (principally specificity) than FFPE.[4–7] Repeated sampling of a fresh operative specimen is often not practical because there is limited material available due to small tumor size and so on. Complete excision of tumors and excisional biopsies are becoming more infrequent, and needle biopsies require the highest quality of histomorphology, especially for immuno-oncology.

Studies that use two different biospecimen formats are also biased as a result of the sampling of different areas of the tumor and by the fact that FF samples often contain a much smaller area of the tumor than FFPE blocks that are usually much larger in size. The "noisier" bioinformatic readouts that are associated with the FFPE material may actually be caused by the inclusion of multiple different clones of tumor within one specimen compared with the smaller number that may be present in the smaller amount of tumor area taken for FF biospecimen preparation.

Common sense therefore suggests that the best way forward would be to optimize protocols for sequencing on DNA derived from the FFPE material, rather than seeking to use two different biospecimen formats. The result of this approach should be better for the patient's diagnosis and treatment than the sum of the parts.

The Problem of Formalin-induced Sequence Artifacts

A problem that users of FFPE tissue face is formalin-induced sequence artifacts, which appear as changes in the DNA sequences following next-generation sequencing (NGS) that were not present in the sample before it was fixed. It is imperative to be able to distinguish genuine mutations from the artifactual ones caused by fixation. Accurate quantification of DNA enables sufficient DNA to be used for library preparation, which is crucial because random sequence artifacts become more noticeable as the number of input DNA templates decreases.[8] Ensuring that the primers used for generating libraries produce amplicons that are short and therefore well suited to the fragmented DNA extracted from FFPE biospecimens maximizes the number of templates that are amplified and then sequenced. This in turn reduces the likelihood that random artifactual mutations will be detected above the background noise and data-filtering thresholds.

Sequence artifacts can also be reduced by selecting DNA polymerases that have low efficiencies at bypassing DNA lesions that are artifactual. As described below, the predominant artifactual mutation is caused by formaldehyde-induced deamination of C bases to uracil (U).[9] If a DNA polymerase such as Pfu

polymerase is used, it can recognize U bases in the template DNA strand and terminate elongation, rather than read across the U, as happens with a conventional DNA polymerase, which results in an incorrect A being incorporated into the newly formed DNA strand.[10] Independently sequencing both sense and antisense strands of the DNA is another approach that can identify artifactual lesions with extremely high accuracy (e.g., an error rate of one artifactual mutation per billion bases sequenced) on account of the mutation only being present in one of the two strands.[11]

Formaldehyde-induced Cytosine Deamination and Uracil-DNA Glycosylase Treatment

When C becomes deaminated, the result is a base substitution to U. The process occurs naturally at an estimated rate of 70 to 200 events per day in a living cell.[12] When it occurs in vivo, the incorrect U is removed by a uracil-DNA glycosylase (UDG), leaving an abasic site. There are at least four different UDGs in mammalian cells (UNG, SMUG1, MBD4, and TDG).[13] The cell's base excision repair mechanism is then able to restore C to its correct position, on account of the still-correct G that is opposite the abasic site in the complementary DNA (cDNA) strand. This intracellular repair mechanism is non-functional in a formalin-fixed biospecimen because the tissue is biologically inert, so when C deamination occurs as a consequence of formalin fixation, the U remains in situ following DNA extraction. Then, when a conventional DNA polymerase encounters the U in its template strand during PCR (for library preparation), it responds by incorporating an A into the amplicon it is making (which is complementary in sequence to the template strand). This results in the G > A artifactual mutation in the complementary strand. The consequence of the A is that in the following round of PCR, the DNA polymerase incorporates a T into the new amplicon it is making, at the site where the C was before fixation, thus resulting in a C > T artifactual mutation. From then on, the artifactual C.G > T.A mutation becomes logarithmically amplified as the PCR reaction proceeds.

Only a small proportion of C bases are deaminated as a consequence of formalin fixation, and the process occurs randomly, resulting in low-frequency and unpredictable, artifactual single-nucleotide variant (SNV) calls. However, low-frequency C > T mutations also occur in cancer and can be clinically important, so addressing the issue of artifactual C deamination is important in the context of sequencing DNA from FFPE biospecimens, even when occurring at low allele frequency. Artifactual SNVs resulting from formalin fixation have been cited as being as high as 1% of the total number of SNV calls when the coverage is low (e.g., 20-fold).[14–16]

An approach to addressing this issue is to incorporate UDG treatment in the DNA extraction protocol using commercially available kits such as GeneRead DNA FFPE Kit (Qiagen, Hilden, Germany), AmpliSeq Direct FFPE DNA Kit (Illumina, San Diego, CA), or NEBNext DNA Repair Mix (New England Biolabs, Ipswich, MA). The result of such UDG treatment is that the U is enzymatically removed, leaving an abasic site. However, the cDNA strand still contains the correct G opposite the abasic site, so when the G-containing strand becomes the template in PCR, the DNA polymerase restores the C to the correct position in the amplicon it produces, and the correct sequence is amplified in subsequent rounds of PCR. It should be noted, however, that the C present in CpG sites is often naturally methylated, and methylcytosine is converted directly into thymidine following deamination; this results in C:G > T:A substitutions after PCR amplification that UDG treatment cannot correct.[9,17]

However, the value of routinely applying UDG treatment to correct artifactual C deamination is open to question: Studies consistently show a limited effect of UDG treatment. Bonnet et al.[14] compared C:G > T:A substitution rates in 25 matched FFPE and FF tumor samples using three different FFPE extraction kits, one including and the other two not including UDG treatment. They found that, overall, FFPE samples had 1% higher C:G > T:A substitution rates than FF (presumably the effect of formalin-induced C deamination), but for the FFPE samples, the extraction kit that included the UDG treatment only reduced the substitutions by 0.01%. In another study in which the UDG-containing GeneRead DNA extraction method was compared with a non-UDG-containing extraction method, C:G > T:A substitution differences only became apparent when the allele frequency cutoff was reduced from 5% to 2.5%, at which point the call was correct when extractions had been performed using the GeneRead kit.[18] A further study found that up to 94% of the observed C:G > T:A substitutions present in FFPE samples were artifactual and could be reversed using UDG the treatment, but the frequency was so low that there was "little impact on sensitivity."[19]

The extent of formalin-induced C deamination has been shown in some studies to depend on fixation time. Fixation time is unfortunately not stated in the Bonnet et al.[14] study, but Prentice et al.[20] found that fixing samples for 48 hr caused more C:G > T:A mutations than fixing them for ≤24 hr, and although UDG treatment could reverse this enhanced substitution, the allelic frequency of the formalin-induced substitutions always fell below the allelic acceptance cutoff threshold

of 5%. A consequence of longer fixation times can be that DNA extracted from the blocks is more fragmented, so it is not surprising that there is a tendency for samples that are more fragmented (with higher cycle threshold numbers in a quantitative PCR [qPCR] assay) to benefit more from UDG treatment.[1,21] For example, in the study by Serizawa et al.,[22] UDG treatment reduced C:G > T:A mutations by 60% when the DNA was "highly fragmented," but in less fragmented DNA, UDG treatment had no effect. The reason UDG treatment appears more effective in poorer quality FFPE blocks could also be stochastic—there will be fewer template molecules of DNA of amplifiable length, meaning that if one of these amplifiable fragments contains a random deamination artifact, there will be a higher probability that it will be amplified to a magnitude that is above the allelic frequency cutoff threshold applied in the NGS data analysis (usually 5%).

Comparing NGS Data From Patient-matched FFPE and Frozen Biospecimens

We are aware of 17 published studies in which FFPE and FF biospecimens from the same patient have been sequenced in the context of NGS (summarized in Table 1).[3,14,15,21,23–35] A common finding is that DNA sequenced from FFPE biospecimens has a lower percentage of mapped reads (i.e., reads that were aligned to the reference genome) than that from FF. Seven of the studies report that DNA from FFPE has lower coverage than that from FF samples, but still above the usually applied quality thresholds for NGS.[3,14,24,27,30,34,35] Another seven studies found or show no statistically significant difference in coverage between FFPE and FF.[15,21,23,25,29,31,32] One paper reports greater coverage in FFPE than FF, and two papers do not report any.[26,28,33]

Of the 17 comparative studies, 16 did not apply UDG treatment to the FFPE extractions, and the 1 study that did applied it to one of the three DNA kits tested (discussed in the previous section).[14] An enrichment in C:G > T:A mutations in the FFPE DNA compared with the FF DNA was reported in four of the reports (average coverage per study, 77–130×),[3,14,30,32] and there were no statistically significant enrichments in five studies.[15,21,24,25,35] One study only found formaldehyde-induced artifactual mutations when the FFPE DNA was particularly degraded,[23] one study only found it in CpG sites,[29] and six studies did not report any.[26–28,31,33,34]

Formaldehyde-induced artifacts are random in their nature, and so they become less likely to materialize when sequencing coverage increases. Kerick et al.[15] noted that a 0.98% false positive mutation rate is detected in DNA from FFPE at 20× coverage, which was eliminated when coverage was increased to 80×. A reduction in the false positive rate from 30% to 10% was noted when coverage was increased from 4× to 8× in another study.[31]

The amount of sequencing coverage is dependent on the quality of the library, which in turn is dependent on the quality and quantity of the DNA. Astolfi et al.[23] sequenced DNA (from FFPE blocks) that had been defined as "good quality" and "poor quality" depending on whether they passed or failed a qPCR quantification and QC Assay: "Good-quality" DNA had the same coverage as FF and "poor-quality" DNA did not. Vanni et al.[34] demonstrated that both 10 and 20 ng DNA could be used for successful library preparation, but 20 ng DNA returned higher sequencing coverage (2/90 amplicons <500×) as opposed to 10 ng DNA (6/90 amplicons <500×) in the Ion Torrent platform. It has also been suggested that the random, fragmented nature of DNA from FFPE enables higher coverage, thus reducing the false positive rate compared with FF.[15]

The extent to which there is homology between FFPE and FF sequencing results depends on the study, but the authors cite values between 55% and 100% (Table 1). However, the different studies define "correlation" differently, with some referring to mapped reads, some to mutational status, and others not providing a definition. The different sequencing platforms and applications used in the studies have radically different depths and breadths of coverage, with different false-discovery rate thresholds applied. These factors will determine the degree to which FFPE and FF sequences correlate. The choice of bioinformatics pipeline is also crucial, as demonstrated by Betge et al.,[24] who used three different bioinformatic pipelines to analyze their data and found that the choice of bioinformatics pipeline had more influence on the results than selecting an FF or FFPE biospecimen. In addition, tissue heterogeneity means that the mutational status of the FFPE and FF will be inherently different. Finally, the different preanalytical factors that FFPE biospecimens undergo during preparation, storage and DNA extraction compared with FF biospecimens will also be an influential factor.

In terms of describing the amenability of FFPE biospecimens to NGS, the consensus of the 17 studies is that raw NGS data from FFPE biospecimens are slightly poorer than those from FF biospecimens, but this is not surprising, given that FF is the "gold standard" for NGS. The value of these higher quality NGS data must be offset against the higher relevance that the FFPE biospecimens have to the patient's

Table 1. A Summary (in Alphabetical Order of Primary Author) of Studies That Have Applied NGS to Patient-matched FFPE and FF Biospecimens.

Reference	Sample Type (Paired FFPE and FF)	Type of NGS	Quantity of DNA Used for Library Prep	UDG Treatment	QC Method	Sequence Concordance Between Frozen and FFPE	% FFPE DNA Extractions Passing QC	% Patients Where FF Samples Available[a]	% Mapped Reads	Authors' Conclusions
Astolfi et al.[23]	Gastrointestinal tumor and normal (n=4)	WES	100 ng	No	Multiplex and qPCR[b] (crude DNA), nano electrophoresis ABio (library prep)	95% (good-quality FFPE), 55% (poor-quality FFPE)	100%, but 40-50% are defined as "high quality"	N/A	98.6% (FFPE) 98.8% (FF)	"WES is feasible on FFPE specimens and it is possible to easily select FFPE samples of high quality that yield sequencing results comparable to the FF counterpart."
Betge et al.[24]	Colorectal cancer (n=10)	TES	250 ng (FFPE) 150 ng (FF)	No	qPCR[c] (crude DNA) and nano electrophoresis ABio (library prep)	91%	53%	58%	96% (FFPE) 78% (FF)	"We illustrate the general feasibility of amplicon sequencing in FFPE tissue."
Bonfiglio et al.[25]	Lung adenocarcinoma (n=5)	WES	200 ng	No	Multiplex PCR, nano electrophoresis DIN, qPCR[d] (crude extract), and nano electrophoresis ABio (library prep)	Not stated	100% (no acceptance criteria specified)	N/A	> 99% (FFPE and FF)	"High-quality and reliable data can be successfully obtained from WES of FFPE samples."
Bonnet et al.[14]	Liver and colon tumors and normal (n=25)	WES	200 ng	No (2 extraction kits), Yes (1 kit)	Nano electrophoresis DIN (crude extract)	> 98.4%	100% (no acceptance criteria specified)	N/A	97% (FFPE) 98.4% (FF) *$p<0.01$	"FFPE samples have less good indicators than FF samples, yet the coverage quality remains above accepted thresholds."
De Paoli-Iseppi et al.[26]	Melanoma lymph node metastasis (n=10).	WES	1 µg	No	Nano electrophoresis ABio (library prep)	55%	40% or 70% depending on exome capture kit	N/A	55% and 69% (FFPE) 99% (FF), *$p<0.04$	"Further refinements will be necessary before FFPE could be used for routine clinical decision making."
Einaga et al.[21]	Normal liver (n=4)	TES	80 ng	No	Multiplex PCR and qPCR (crude DNA)	Not stated	100% (no acceptance criteria specified)	N/A	Not stated	"FFPE tissues used for routine clinical diagnosis can be utilized to obtain reliable NGS data if appropriate conditions of fixation and validation are applied."
Hedegaard et al.[27]	Colon, bladder, and prostate carcinoma (n=38)	TES	1.2 µg	No	Nano electrophoresis ABio (crude DNA and library prep)	Not stated	100% (no acceptance criteria specified)	N/A	95% (FFPE) 98% (FF)	"Our results suggest that NGS can be used to study FFPE specimens in both prospective and retrospective archive-based studies in which FF specimens are not available."
Kerick et al.[15]	Prostate tumors (n=5)	TES, WES	500 ng	No	qPCR and nano electrophoresis ABio	FF and FFPE had "similar coverage profiles"	30% success rate for library preparation in FFPE	N/A	75% both FF and FFPE	"FF tissues remain the preferred source of DNA, but FFPE tissue can be used for SNV and InDel detection instead if the coverage is 80× or higher."
Mafficini et al.[28]	Pancreatic tumors (n=5)	TES	20 ng	No	Multiplex PCR (crude DNA) and nano electrophoresis ABio (library prep)	"Similar for FFPE and FF DNA"	100%	N/A	Not stated	"NGS can be applied on DNA from routinely prepared paraffin tissues; the data produced are quantitative and thus permit the description of the molecular subclonal composition of a tumor."
Munchel et al.[29]	Ovarian, breast, and colon tumors and normal (n=13)	WES, TES, WGS	50 ng	No	nano electrophoresis ABio and qPCR (crude DNA), nano electrophoresis ABio (library prep)	> 96% (WES) > 95% (TES) 100% (WGS)	100% (no acceptance criteria specified)	N/A	Not stated	"Results from our studies suggest the potential use of diagnostic FFPE samples for cancer genomic studies."
Oh et al.[30]	Tumor, tissue type not stated (n=4)	WES	Not stated	No	Agarose gel electrophoresis (crude DNA)	Not stated	100% (no acceptance criteria specified)	N/A	90% (FFPE) 99% (FF)	"DNA derived from routinely processed FFPE specimens produced NGS data that were similar in quality to DNA from frozen samples."

(continued)

Table 1. (continued)

Reference	Sample Type (Paired FFPE and FF)	Type of NGS	Quantity of DNA Used for Library Prep	UDG Treatment	QC Method	Sequence Concordance Between Frozen and FFPE	% FFPE DNA Extractions Passing QC	% Patients Where FF Samples Available[a]	% Mapped Reads	Authors' Conclusions
Robbe et al.[3]	Breast, colorectal, endometrial, prostatic, renal, and thoracic tumors (n=52)	WGS	250 or 400 ng (FFPE) depending on results of QC.	No	qPCR[c] (crude DNA)	63-71%	69%	52%	Not stated	"Our results support the use of optimized FFPE cancer samples as an alternative source of DNA for WGS cancer diagnostics if FF specimens are not available."
Schweiger et al.[31]	Breast tumor (n=1), macrodissection for tumor in FFPE only	WGS	1.5 µg	No	Nano electrophoresis [ABio] (library prep)	χ^2= 0.82	100% (no acceptance criteria specified)	N/A	Not stated	"The data obtained from both preparation methods is of comparable quality."
Spencer et al.[32]	Lung adenocarcinoma (n=16)	TES	Not stated	No	Multiplex PCR (crude DNA) and nano electrophoresis [ABio] (library prep)	99.99%	100%	N/A	97.7% (FFPE) 98.6% (FF) *p=0.02	"Differences between fixed and frozen samples are minor and do not affect clinical diagnostic calls."
Van Allen et al.[33]	Lung adenocarcinoma (n=11)	TES	100 ng	No	Not stated	92.6%	100%	N/A	Not stated	"The ability to detect base mutations that were sufficiently powered was equivalent regardless of whether frozen or FFPE-derived genomic DNA was used."
Vanni et al.[34]	Non–small cell lung cancer (n=12)	TES	20 ng	No	Multiplex PCR (crude DNA) and nano electrophoresis [ABio] (library prep)	Not stated	100%	N/A	73–100%, depending on degradation state (FFPE), 99–100% (FF)	"Our NGS-based workflow was successfully able to screen FFPE tissues, demonstrating the possibility of transferring this novel technology into routine clinical context."
Wong et al.[35]	Mouse melanoma xenograft (FFPE) and cell line used for xenograft (FF) (n=7)	TES	1 µg	No	Multiplex PCR (crude extract)	83% (SNP), 62% (InDel)	100%	N/A	91.9% (FFPE) 98.5% (FF)	"Formalin fixed xenografts can be accurately assessed for their mutational profile with very similar results to the unfixed corresponding cell lines."

Abbreviations: NGS, next-generation sequencing; FFPE, formalin-fixed, paraffin-embedded; FF, fresh-frozen; qPCR, quantitative PCR; WES, whole exome sequencing; TES, targeted exome sequencing; WGS, whole genome sequencing; ABio, Bioanalyzer (Agilent Technologies, Santa Clara, CA); LGX, LabChip GX (Perkin Elmer, Waltham, MA); DIN, DNA Integrity Numbers (obtained using Agilent TapeStation); UDG, uracil-DNA glycosylase.
[a]N/A: The percentage of patients without a matched FF biospecimen is not stated, or patient-matched FF and FFPE biospecimens obtained from a Biobank.
[b]hgDNA Quantification and QC Kit (KAPA Biosystems, Wilmington, MA).
[c]FFPE QC Assay (Illumina).
[d]ProNex DNA QC Assay (Promega, Madison, WI).
*Statistical significance.

diagnosis, given the same FFPE block will also have been used for IHC and histomorphological analyses. However, the salient question is, "Are the FFPE samples good enough?" which in the clinical setting means "is there a consequential difference in clinical decision making when NGS data from FFPE, as opposed to data from FF biospecimens, are used?" At one end of the scale, De Paoli-Iseppi et al.[26] found the differences between FFPE and FF biospecimens to be so large that FFPE "shows little promise for use in clinical whole exome sequencing" and the FFPE data were "too inaccurate to confidently inform clinical decisions." The authors of the 16 other studies take the opposite view, however, concluding that accurate NGS data can be obtained from FFPE biospecimens and clinically actionable variants identified.

One explanation as to why the conclusions of De Paoli-Iseppi et al.[26] are so discordant compared with the other studies could be because melanoma is a cancer with particularly high mutation rates and heterogeneity.[36] Also, the methods are not entirely clear as to how spatially distant FF and FFPE blocks were: The FF samples were "surgically resected lymph node melanoma metastases" and the FFPE biospecimens (from the same patients) were "routinely collected tumor tissue blocks from the same specimen" that were "generally adjacent to the tumor sample."

The Confounding Problem of Intratumor Heterogeneity

To assign differences in NGS data between patient-matched FF and FFPE tissue to the formalin fixation, the extent of inherent tissue heterogeneity must be controlled for. Tumor heterogeneity has been the subject of various studies. For example, using exome sequencing, the average number of mutations per tumor have been cited as 84 for breast (n=11), 76 for colorectal (n=11), 48 for pancreatic (n=48), and 47 for glioblastoma (n=21).[37] In renal cancer, whole exome sequencing has found that 63% to 69% of mutations were heterogeneous and thus not detectable in every sequenced region of the tumor.[38] In another paper, when 19 renal cancer biopsies were evaluated in respect of tumor heterogeneity, 25% to 50% of the variants were not detected elsewhere in the tumor, despite sequencing coverage being >250×.[39] An analysis of intratumoral heterogeneity in 12 cancer types found that 86% of tumors contained at least two clones.[40] Navin et al.[41] present a study using breast cancer biopsies, in which they separate subpopulations of tumor cells from a single biopsy using sector ploidy profiling, apply single-nucleus sequencing to tumor cells, and then demonstrate that 100 cells from a single breast cancer biopsy contained three clones likely involved in clonal expansion.

Intratumoral heterogeneity is clearly commonplace, both within individual tumor biopsies and in spatially separated tissue blocks taken from the same tumor, as in the comparative FF vs FFPE studies. Indeed, one paper describes the magnitude of tumor heterogeneity as "bewildering."[39] In our opinion, the magnitude of this heterogeneity is large enough to account for the differences seen between FFPE and FF samples summarized in Table 1. Indeed, some authors publishing papers that compare sequences derived from paired samples readily make the point that an unknown proportion of the differences they find are a consequence of intrasample heterogeneity.[3,42]

Tissue heterogeneity does not account for the poorer metrics seen in DNA extracted from FFPE biospecimens in some of the studies, however, and the fact remains that when extracted from FFPE, DNA is more degraded and more likely to fail QC assays than if it had been extracted from FF samples. In our experience, the typical QC failure rate in a cohort of clinical FFPE samples is 25% to 40%, with the DNA either being too degraded or being insufficient in quantity. This failure rate is comparable to that reported by others.[3,27] It would therefore be understandable for clinicians to hesitate before embracing FFPE as a biospecimen type when QC failure rates are of this magnitude, but this downside must be offset by the universal availability of FFPE tissue and the problems in collecting FF tissue.

In the majority of Biobanks, FF biospecimens are either much scarcer than FFPE tissue blocks or not available at all. So, researchers open to using FFPE biospecimens have a much greater selection of biospecimens to select their samples from, thus enabling them to use larger cohorts of samples and more precisely match potentially confounding parameters such as patient age, gender, and percentage of tumor within and between their study groups. For clinical diagnosis, minimal sampling methods such as needle biopsies are too small to yield sufficient tissue to have both an FF and an FFPE sample. For these patients, selecting the FF option is not feasible because the error rate of diagnosing cancer, defining tumor grade and stage, and performing immunohistochemistry is too high. Consequently, when NGS is performed using FF tissue, an additional FFPE block is required for diagnosis, but conversely when an FFPE block has been obtained for diagnosis, an additional FF block for NGS is not a necessity, but surplus to requirements in most instances.

We therefore argue that the focus should be on improving the QC failure rates in FFPE biospecimens rather than attempting to change clinical practice and collect FF and FFPE tissues. It would be desirable to better understand how events in tissue processing impact DNA integrity. In addition, DNA yields and integrities can be improved at the point of DNA extraction, by individual labs optimizing their extraction protocols. For example, we recently demonstrated that optimizing the Proteinase K digest and deparaffinization steps of a DNA extraction kit's protocol reduced the percentage of FFPE DNA extractions that failed to meet the QC acceptance criteria used in the 100,000 Genomes Project's pilot study from 33% to 7% in a cohort of 54 FFPE clinical tissue blocks.[43] This equates to considerably fewer patients than those who were excluded because an FF tissue block could not be obtained (48%).[3]

Improving the percentage of FFPE biospecimens that can be sequenced would have benefits for patients (fewer patients needing to return to clinic for another biopsy) and for research (larger cohort sizes mean more robust results). Individual labs can improve their data by taking simple steps to optimize their DNA extraction protocols. Biospecimen Science will enable us to better understand what precise factors are responsible for the poorer quality found when DNA is extracted from FFPE biospecimens and drive improvements in DNA extraction protocols, QC assays, NGS methodology, and bioinformatic pipelines. We are optimistic that FFPE biospecimens will be more commonly used for NGS, and the additional data therefore obtained will also be valuable in matching data quality to preanalytical variables. In the meantime, however, there is sufficient information available to make it clear that NGS can be performed on FFPE biospecimens with confidence.

Competing Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Author Contributions

WM wrote the manuscript from and including the section "The Problem of Formalin-induced Sequence Artifacts" and GAT wrote the preceding text.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

William Mathieson https://orcid.org/0000-0002-1328-5629

Literature Cited

1. Fox CH, Johnson FB, Whiting J, Roller PP. Formaldehyde fixation. J Histochem Cytochem. 1985;33(8):845–53.
2. Tsai KK, Daud AI. The role of anti-PD-1/PD-L1 agents in melanoma: progress to date. Drugs. 2015;75(6):563–75.
3. Robbe P, Popitsch N, Knight SJL, Antoniou P, Becq J, He M, Kanapin A, Samsonova A, Vavoulis DV, Ross MT, Kingsbury Z, Cabes M, Ramos SDC, Page S, Dreau H, Ridout K, Jones LJ, Tuff-Lacey A, Henderson S, Mason J, Buffa FM, Verrill C, Maldonado-Perez D, Roxanis I, Collantes E, Browning L, Dhar S, Damato S, Davies S, Caulfield M, Bentley DR, Taylor JC, Turnbull C, Schuh A. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. Genet Med. 2018;20:1196–205.
4. Catalona WJ, Stein AJ. Accuracy of frozen section detection of lymph node metastases in prostatic carcinoma. J Urol. 1982;127(3):460–1.
5. Grabenstetter A, Moo TA, Hajiyeva S, Schuffler PJ, Khattar P, Friedlander MA, McCormack MA, Raiss M, Zabor EC, Barrio A, Morrow M, Edelweiss M. Accuracy of intraoperative frozen section of sentinel lymph nodes after neoadjuvant chemotherapy for breast carcinoma. Am J Surg Pathol. 2019;43(10):1377–83.
6. Harris PL, Rumley TO, Lineaweaver WC, Copeland EM 3rd. Pancreatic cancer: unreliability of frozen section in diagnosis. South Med J. 1985;78(9):1053–6.
7. Mat Zin AA, Zulkarnain S. Diagnostic accuracy of cytology smear and frozen section in glioma. Asian Pac J Cancer Prev. 2019;20(2):321–5.
8. 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(1):36–9.
9. Do H, Dobrovic A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. Clin Chem. 2015;61(1):64–71.
10. 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(16):e161.
11. 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 USA. 2012;109(36):14508–13.
12. Kavli B, Otterlei M, Slupphaug G, Krokan HE. Uracil in DNA—general mutagen, but normal intermediate in acquired immunity. DNA Repair. 2007;6(4):505–16.
13. Krokan HE, Nilsen H, Skorpen F, Otterlei M, Slupphaug G. Base excision repair of DNA in mammalian cells. FEBS Lett. 2000;476(1–2):73–7.
14. Bonnet E, Moutet ML, Baulard C, Bacq-Daian D, Sandron F, Mesrob L, Fin B, Delepine M, Palomares MA, Jubin C, Blanche H, Meyer V, Boland A, Olaso R,

Deleuze JF. Performance comparison of three DNA extraction kits on human whole-exome data from formalin-fixed paraffin-embedded normal and tumor samples. PLoS ONE. 2018;13(4):e0195471.
15. Kerick M, Isau M, Timmermann B, Sultmann H, Herwig R, Krobitsch S, Schaefer G, Verdorfer I, Bartsch G, Klocker H, Lehrach H, Schweiger MR. Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. BMC Med Genomics. 2011;4:68.
16. Loudig O, Brandwein-Gensler M, Kim RS, Lin J, Isayeva T, Liu C, Segall JE, Kenny PA, Prystowsky MB. Illumina whole-genome complementary DNA-mediated annealing, selection, extension and ligation platform: assessing its performance in formalin-fixed, paraffin-embedded samples and identifying invasion pattern-related genes in oral squamous cell carcinoma. Hum Pathol. 2011;42(12):1911–22.
17. Duncan BK, Miller JH. Mutagenic deamination of cytosine residues in DNA. Nature. 1980;287(5782):560–1.
18. Darwanto A, Hein AM, Strauss S, Kong Y, Sheridan A, Richards D, Lader E, Ngowe M, Pelletier T, Adams D, Ricker A, Patel N, Kuhne A, Hughes S, Shiffman D, Zimmermann D, Te Kaat K, Rothmann T. Use of the Qiagen GeneReader NGS system for detection of KRAS mutations, validated by the Qiagen Therascreen PCR kit and alternative NGS platform. BMC Cancer. 2017;17(1):358.
19. Bourgon R, Lu S, Yan Y, Lackner MR, Wang W, Weigman V, Wang D, Guan Y, Ryner L, Koeppen H, Patel R, Hampton GM, Amler LC, Wang Y. High-throughput detection of clinically relevant mutations in archived tumor samples by multiplexed PCR and next-generation sequencing. Clin Cancer Res. 2014;20(8):2080–91.
20. Prentice LM, Miller RR, Knaggs J, Mazloomian A, Hernandez RA, Franchini P, Parsa K, Tessier-Cloutier B, Lapuk A, Huntsman D, Schaeffer DF, Sheffield BS. Formalin fixation increases deamination mutation signature but should not lead to false positive mutations in clinical practice. PLoS ONE. 2018;13(4):e0196434.
21. Einaga N, Yoshida A, Noda H, Suemitsu M, Nakayama Y, Sakurada A, Kawaji Y, Yamaguchi H, Sasaki Y, Tokino T, Esumi M. Assessment of the quality of DNA from various formalin-fixed paraffin-embedded (FFPE) tissues and the use of this DNA for next-generation sequencing (NGS) with no artifactual mutation. PLoS ONE. 2017;12(5):e0176280.
22. Serizawa M, Yokota T, Hosokawa A, Kusafuka K, Sugiyama T, Tsubosa 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 Genet. 2015;208(9):415–27.
23. Astolfi A, Urbini M, Indio V, Nannini M, Genovese CG, Santini D, Saponara M, Mandrioli A, Ercolani G, Brandi G, Biasco G, Pantaleo MA. Whole exome sequencing (WES) on formalin-fixed, paraffin-embedded (FFPE) tumor tissue in gastrointestinal stromal tumors (GIST). BMC Genomics. 2015;16:892.
24. Betge J, Kerr G, Miersch T, Leible S, Erdmann G, Galata CL, Zhan TZ, Gaiser T, Post S, Ebert MP, Horisberger K, Boutros M. Amplicon sequencing of colorectal cancer: variant calling in frozen and formalin-fixed samples. PLoS ONE. 2015;10(5):e0127146.
25. Bonfiglio S, Vanni I, Rossella V, Truini A, Lazarevic D, Dal Bello MG, Alama A, Mora M, Rijavec E, Genova C, Cittaro D, Grossi F, Coco S. Performance comparison of two commercial human whole-exome capture systems on formalin-fixed paraffin-embedded lung adenocarcinoma samples. BMC Cancer. 2016;16:692.
26. De Paoli-Iseppi R, Johansson PA, Menzies AM, Dias KR, Pupo GM, Kakavand H, Wilmott JS, Mann GJ, Hayward NK, Dinger ME, Long GV, Scolyer RA. Comparison of whole-exome sequencing of matched fresh and formalin fixed paraffin embedded melanoma tumours: implications for clinical decision making. Pathology. 2016;48(3):261–6.
27. Hedegaard J, Thorsen K, Lund MK, Hein A-MK, Hamilton-Dutoit SJ, Vang S, Nordentoft I, Birkenkamp-Demtroder K, Kruhoffer M, Hager H, Knudsen B, Andersen CL, Sorensen KD, Pedersen JS, Orntoft TF, Dyrskjot L. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. PLoS ONE. 2014;9(5):e98187.
28. Mafficini A, Amato E, Fassan M, Simbolo M, Antonello D, Vicentini C, Scardoni M, Bersani S, Gottardi M, Rusev B, Malpeli G, Corbo V, Barbi S, Sikora KO, Lawlor RT, Tortora G, Scarpa A. Reporting tumor molecular heterogeneity in histopathological diagnosis. PLoS ONE. 2014;9(8):e104979.
29. Munchel S, Hoang Y, Zhao Y, Cottrell J, Klotzle B, Godwin AK, Koestler D, Beyerlein P, Fan JB, Bibikova M, Chien J. Targeted or whole genome sequencing of formalin fixed tissue samples: potential applications in cancer genomics. Oncotarget. 2015;6(28):25943–61.
30. Oh E, Choi YL, Kwon MJ, Kim RN, Kim YJ, Song JY, Jung KS, Shin YK. Comparison of accuracy of whole-exome sequencing with formalin-fixed paraffin-embedded and fresh frozen tissue samples. PLoS ONE. 2015;10(12):e0144162.
31. Schweiger MR, Kerick M, Timmermann B, Albrecht MW, Borodina T, Parkhomchuk D, Zatloukal K, Lehrach H. Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. PLoS ONE. 2009;4(5):e5548.
32. Spencer DH, Sehn JK, Abel HJ, Watson MA, Pfeifer JD, Duncavage EJ. Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. J Mol Diagn. 2013;15(5):623–33.
33. Van Allen EM, Wagle N, Stojanov P, Perrin DL, Cibulskis K, Marlow S, Jane-Valbuena J, Friedrich DC, Kryukov G, Carter SL, McKenna A, Sivachenko A, Rosenberg M, Kiezun A, Voet D, Lawrence M, Lichtenstein LT, Gentry

JG, Huang FW, Fostel J, Farlow D, Barbie D, Gandhi L, Lander ES, Gray SW, Joffe S, Janne P, Garber J, MacConaill L, Lindeman N, Rollins B, Kantoff P, Fisher SA, Gabriel S, Getz G, Garraway LA. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. Nat Med. 2014;20(6):682–8.
34. Vanni I, Coco S, Truini A, Rusmini M, Dal Bello MG, Alama A, Banelli B, Mora M, Rijavec E, Barletta G, Genova C, Biello F, Maggioni C, Grossi F. Next-generation sequencing workflow for NSCLC critical samples using a targeted sequencing approach by ion torrent PGM platform. Int J Mol Sci. 2015;16(12):28765–82.
35. Wong SQ, Li J, Salemi R, Sheppard KE, Do H, Tothill RW, McArthur GA, Dobrovic A. Targeted-capture massively-parallel sequencing enables robust detection of clinically informative mutations from formalin-fixed tumours. Sci Rep. 2013;3:3494.
36. Grzywa TM, Paskal W, Wlodarski PK. Intratumor and intertumor heterogeneity in melanoma. Transl Oncol. 2017;10(6):956–75.
37. Salk JJ, Fox EJ, Loeb LA. Mutational heterogeneity in human cancers: origin and consequences. Annu Rev Pathol. 2010;5:51–75.
38. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883–92.
39. Swanton C. Intratumor heterogeneity: evolution through space and time. Cancer Res. 2012;72(19):4875–82.
40. Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, Ji HP, Maley CC. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. Nat Med. 2016;22(1):105–13.
41. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, Muthuswamy L, Krasnitz A, McCombie WR, Hicks J, Wigler M. Tumour evolution inferred by single-cell sequencing. Nature. 2011;472(7341):90–4.
42. Kim S, Park C, Ji Y, Kim DG, Bae H, van Vrancken M, Kim DH, Kim KM. Deamination effects in formalin-fixed, paraffin-embedded tissue samples in the era of precision medicine. J Mol Diag. 2017;19(1):137–46.
43. Frazer Z, Yoo C, Sroya M, Bellora C, DeWitt BL, Sanchez I, Thomas GA, Mathieson W. Effect of different proteinase K digest protocols and deparaffinization methods on yield and integrity of DNA extracted from formalin-fixed, paraffin-embedded tissue. J Histochem Cytochem. 2020;68(3):171–84.