Tissue Requirements and DNA Quality Control for Clinical Targeted Next-Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Samples: A Mini-Review of Practical Issues

Chung MJ[1,2], Lin W[1], Dong L[1] and Li X[1]*

[1]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, California, USA
[2]Department of Pathology, Chonbuk National University Medical School, San 2-20 Keumam-Dong, Jeonju 561-180, Republic of Korea

Abstract

Most molecular pathology laboratories perform mutational analyses to diagnose somatic cancer mutations and evaluate therapeutic options on formalin fixed paraffin embedded (FFPE) samples as daily practice using various methods. Recent studies show that targeted next-generation sequencing (NGS) is a promising diagnostic method with many benefits including simultaneous detection of multiple mutations in various genes in a single test. High quality DNA is essential for an efficient and successful NGS performance. However, low tumor tissue percentage and low DNA quality are the main limitations to use FFPE DNA for NGS assay. We reviewed and discussed the required tissue amount for the NGS assay, the effect of formalin fixation on DNA integrity, and the method for reduction of formalin-induced sequencing artifacts. We also review DNA extraction methods, DNA quality control methods and quality control workflow for nucleic acids and libraries. This review provides an overview for molecular pathology laboratories or researcher considering NGS to detect somatic cancer mutations using small FFPE samples.

Introduction

Next generation sequencing (NGS) is a powerful technique and rapidly spreading in clinical and research arena. Although it depends on platforms, long turnaround time, huge amount of data, and the need for bioinformatics for data analysis are considered as a major limitation for diagnostic use in clinical laboratories. Targeted NGS investigates specific areas of interest rather than an entire gene or exon. Therefore, it produces smaller, more manageable datasets, reduces turnaround time, and decreases sequencing costs. Also, as it focuses on specific regions of interest, it leads to greater depth of coverage and increases the confidence of detecting a low-level variant in cancer samples [1].

Until now, direct sequencing (DS) or real-time polymerase chain reaction (PCR) analyses are the most commonly used methods to detect somatic cancer mutations in clinical laboratories. However, demand for improved sensitivity of direct sequencing and resolving the inconvenience of multiple sequential testing practices for real-time PCR has increased [2,3]. Targeted NGS can overcome these limitations of DS or real-time PCR, and showed that NGS tests are accurate and feasible to detect clinically relevant cancer mutations as daily routine diagnostics [4-6]. Thus, targeted NGS has been suggested for routine diagnostics and is now widely used in the molecular pathology laboratories [4-6].

Archived formalin fixed paraffin embedded (FFPE) tissues are most widely available materials for targeted NGS for detection of somatic cancer mutations. However, the degradation and limited amount of FFPE DNA have posed some technical challenges in its use for molecular studies in cancer research. It is therefore important to recognize the effects of formalin on DNA quality and its effects on the sequencing outcome and to take measures to overcome it. This paper reviews formalin effect on DNA quality, required tumor content for targeted NGS, and several key steps that are important for the library preparation when conducting NGS tests using small biopsied FFPE samples.

Tumor Content of FFPE Samples for NGS

Tissues supplied for somatic cancer mutation analysis in clinical laboratory are normally not specifically for this purpose, but are excess tissue from routine diagnostic and staging purposes. Archived FFPE tissues are the main resource of these tests and most of them are small biopsies. Therefore, the amount of tumor cells in the sample may not be sufficient, which may lead to inaccurate results, such as false negative results. Also, FFPE DNA is heavily degraded due to adverse effect of formalin on nucleic acid which may cause artificial mutations. The relative frequency of artificial mutations may increase when starting material is limited. The tumor cell content of the specimen is also important to determine the significance of the detected variants. Variant allele frequency (VAF) is how often the variant was detected in a given sample. Recent report suggested more than 5% of allele frequency as a guideline for allele frequency filter for somatic variants [7]. If the amount of tumor is small, even if a variant is detected, it is less than the guideline value (<5% of allele frequency) and is likely to be reported as negative. Our laboratory uses more than 3% of the allele frequency for filtering, but recommends that the reference value should be less than 3% when the sample has low tumor content. Therefore, the percentage of tumor cells in the sample is important and should be evaluated and reported.

We reviewed the research articles describing the amount of tumor cells required for accurate testing of targeted NGS for cancer-related genes using FFPE specimens. A near-linear decline in mutation frequency is observed as tumor-cell content in the titration decreases.

*Corresponding author: Dr. Xinmin Li, Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, California, USA, Tel: 3108253664; Fax: 3108253570; E-mail: xinminli@mednet.ucla.edu

Most studies showed successful sequencing results in samples with >20% tumor content [3,5,8]. An *EGFR* mutation was called for in all samples with tumor-cell content >22.8% [3]. Macro-or micro-dissection has been used to enrich specimens with <20% tumor content. One study showed that FFPE tissue with satisfied both >20% tumor content and >1000 tumor cells have sufficient undamaged templates to dominate the amplification process and resulted in highly reproducible and sensitive results for targeted sequencing of 409 genes [9].

Also, optimal amount of DNA is necessary for successful sequencing. The required quantity of DNA varies (10 ng–10 μg) depending on the genotyping aim and the sequencing platform. For targeted sequencing, the required amount of DNA to detect mutation in about 50 genes is about 10-100 ng and 10 ng for Illumina MiSeq and Ion Personal Genome Machine, respectively. Needle biopsies usually yield sufficient DNA for targeted sequencing but biopsy samples are usually less than resected specimen (% of samples proceeded to sequencing: 63% vs. 96%) [10]. The DNA yield in biopsy tissue varies from tissue to tissue, but DNA yield in FFPE tissue is known to be 0.1-3.5 μg DNA per mg. Average DNA yield from FFPE tissue was 1.2 μg DNA/mm^3 and 3.1 μg DNA/mm^3 in needle biopsy of lung tissue and colonic tissue, respectively [11]. A 40-μm prostate needle biopsy with 10 mm^2 surface area produces 55 ng of DNA [10].

Formalin Effect on Nucleic Acids and Uracil-DNA Glycosylase

Formaldehyde as 10% neutral buffered formalin is the most widely used fixative because it preserves a wide range of tissues and maintains the morphology of the tissue. Recently archived FFPE tissues are becoming a more and more important DNA resource in cancer research. However, in general, DNA extracted from FFPE is damaged, which has some limitations for downstream molecular analysis. Therefore, it is important for both clinicians and researchers to be aware of the effects of formalin on DNA quality.

The most important steps to affect high quality DNA from FFPE samples are the pre-extraction steps, such as fixation duration, fixative type, fixative composition (formalin concentration, pH, and salt concentration), tissue type, and temperature. Summarized recommendations for using formaldehyde as a tissue nucleic acid fixative are: 1) minimal prefixation time lag <2 hours; 2) use cold 10% neutral formalin; 3) low salt concentration; 4) cold fixation (at 4°C); 5) duration of fixation 3–6 hours; 6) 20–50 mM ethylene-diamine-tetra-acetic acid (EDTA) as an additive; and 7) avoid a low pH environment [12].

Common types of DNA damage induced by formalin fixation are DNA fragmentation and deamination of cytosine. In a study of the relationship between formalin fixation time and DNA degradation, the amount of fragmented DNA was relatively small in short term fixation (4 hours) and DNA degradation was increased with increasing fixation time. The samples fixed for 72 hours had only 15% of the compared to DNA obtained from fresh frozen tissues [13]. There are three main chemical interactions between formaldehyde and nucleic acids, which are known to cause DNA degradation and affect subsequent DNA analysis: 1) Formaldehyde can induce DNA adducts and cause the formation of methylene bridges between amino groups. 2) It can generate apurinic and apyrimidinic (AP) sites. 3) It may cause hydrolysis of the phosphodiester bonds leading to DNA fragmentation [12,14]. The methylene bridges between affected DNA bases can inhibit denaturation of double-stranded DNA and lead to failed PCR amplification. AP sites can induce template breakage because the AP site cannot form a base pair during PCR [12,15].

Another common type of DNA damage that occurs under formalin fixation is the cytosine deamination. FFPE DNA exhibits a higher sequence variation compared to DNA isolated from freshly frozen tissue, and one mutant artifact per 500-2,050 bases is reported [14]. An artificial C-T or G-A mutation often accounts for 50–90% of all artifactual single nucleotide changes in FFPE samples [14,16]. Formalin fixation results in a deamination of cytosine bases, making the Taq DNA polymerase incapable of recognizing cytosine and eventually incorporating other bases. Cytosine deamination means the amine group (NH_2) of the cytosine base is removed spontaneously, which causes conversion of cytosine into uracil. This occurs at a rate of 70–200 events per day in the human genome and is repaired by uracil-DNA glycosylase (UDG) in living cells. However, cytosine deamination can occur in FFPE samples during tissue collection, formalin fixation, or storage and the replaced uracil cannot be repaired because of the absence of UDG. The resulting uracil lesions cause a C-T sequence artifact because uracil pairs with adenine. Deamination of methylcytosine forms thymine and then it causes incorporation an adenosine instead of a guanine. Do et al. demonstrated that deamination of cytosine bases in FFPE samples is the primary cause of C-T or G-A sequence artifacts [16].

Studies have reported that UDG pretreatment markedly reduces the incidence of such artificial mutations and greatly facilitates the accurate discrimination of mutations in FFPE samples by use of amplicon-based approaches [16,17]. Sequencing PCR products without UDG pre-treatment resulted in 53 C-T or G-A sequence artifacts from 85 sequencing replicates (60%) in *AKT1, BRAF*, and *EGFR* assays. A marked reduction of those sequence artifacts was observed when the samples were tested after UDG treatment [16,17].

DNA Extraction Methods: Manual vs. Automated Methods

DNA extraction methods can be classified into three categories: manual phenol-chloroform (PC) method, manual extraction method using a commercial kit, and automatic extraction method. Here, we briefly review DNA extraction methods from FFPE tissues. The manual PC method showed different efficiency for each report, it was reported to have poor DNA yield when compared to commercial kit, and other studies reported that the highest yields of amplifiable DNA were extracted [18,19]. G. Turashvili et al. compared the efficiency of the four nucleic acid extraction methods [20]. The Wax Free DNA kit (Trimgen, MD, USA) yielded the highest amounts of DNA, followed by in-house PC method, QIA amp DNA FFPE tissue kit (Qiagen Inc., ON, Canada) and Recover All kit (Applied Biosystems/Ambion, ON, Canada). The DNA quality of FFPE was the best in PC method and the Wax Free DNA kit showed the lowest quality. However, the Wax Free method appeared to yield the highest PCR success rates and the Recover All method yielded the lowest success rates with the FFPE. The different results associated with the PC method may be due to the different skills of the workers and the different in-house protocols. However, a major disadvantage of the PC method is that it requires more experimental steps than the kit methods and uses phenol and chloroform which are harmful to human body.

Studies have compared DNA extraction yields from a variety of commercial kits, with results varying depending on the type of kit being compared. However, most of the currently widely used kits were not significantly different [21,22]. DNA fragmentation is one of the main problems of FFPE DNA. The DNA extracted from some kits showed good results in the amplification of longer DNA fragments [21].

Many companies have launched automated nucleic acid extractors, and one study compared five automated DNA extraction systems,
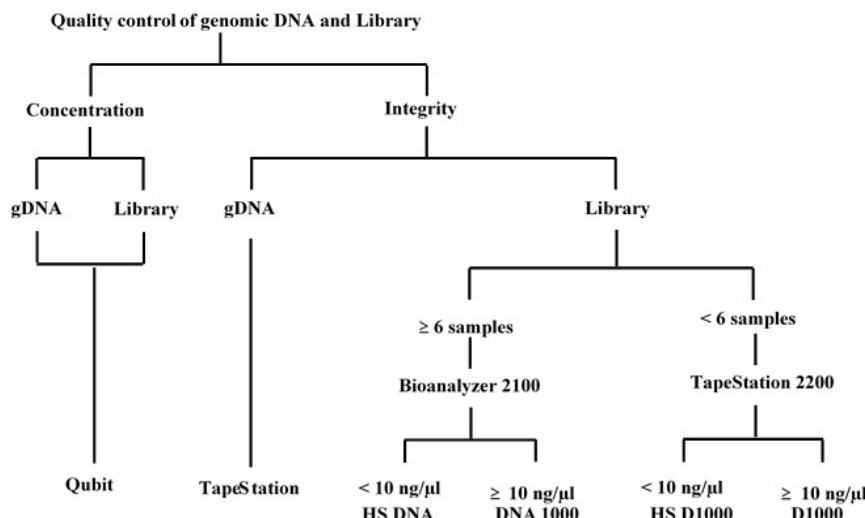
Figure 1: Quality control workflow for genomic DNA (gDNA) and library with regard to performance in downstream sequencing. High sensitivity (HS) DNA, DNA 1000, HS D1000, and D1000 are the names of assay kits.

such as three from Qiagen (Hilden, Germany), the Maxwell 16 from Promega (Mannheim, Germany), and the InnuPure C16 from Analytik Jena (Jena, Germany) [23]. The extract from the Maxwell 16 system had 1.3–24.6-fold higher DNA concentrations compared to those of other extraction systems, and DNA quality from the Maxwell 16 extract was most suitable for downstream applications because the Maxwell 16 extract had higher molecular weight DNA present compared to that from the other automated methods [23]. However, when the Maxwell 16 system was compared with manual DNA extraction kit, the mean DNA concentration obtained by the manual kit was greater than that from the automated system [24]. The known benefits of automated DNA extractors are lack of cross-contamination and less hands-on time. Processing times (hands-on time/total operation time) for the two extraction methods were 25 min/2.5 hr and 40 min/6 hr for the Maxwell 16 and the Qiagen FFPE kit, respectively [24]. Cost of the instrument is a limiting factor the automated method for some laboratories. Because PC methods, commercial kit methods, and automated systems have their advantages and disadvantages, researchers will need to make decisions based on available resources, personal experience, and existing infrastructure along with these points when choosing DNA extraction methods.

Comparison of DNA Quantification Methods

Accurate quantitative analysis of input genomic DNA and libraries is critical for successful targeted sequencing. The use of a defined amount of genomic DNA is important for obtaining consistent and reproducible results in library preparation, and the use of equivalent libraries between samples is critical for even coverage with minimal bias during sequencing. Spectrophotometric methods (e.g., NanoDrop instrument, Thermo Scientific) and fluorometric methods (e.g., Qubit Fluorometer, Life Technologies or Quantus Fluorometer, Promega Corp) are the two most commonly used methods for quantifying nucleic acids. These methods have advantages and disadvantages [23,25,26]. The advantage of a spectrophotometric assay is that it is fast and simple and provide data regarding possible impurities with absorbance ratios at 260/280 and 260/230. However, the biggest drawback is that it cannot differentiate between RNA, single-stranded DNA, and double-stranded DNA (dsDNA), which can result in overestimates of DNA quantity. A fluorescence assay offers improved tolerance to contamination because it

uses an intercalating dye that only interacts with dsDNA. This increased measurement precision of the fluorescence assay is required for FFPE samples because NanoDrop overestimates DNA concentrations in FFPE samples. FFPE tissue quantified by the NanoDrop alone gave poor performances in a NGS library set-up [25,26]. The vulnerability of Qubit is that it does not provide DNA quality data. Therefore, library yield can be different despite standardization of the input DNA amount [27]. However, since the main function of the instrument for quantification of genomic DNA is to accurately quantify dsDNA, and most of the sequencing system manufacturers recommend dsDNA-specific fluorometric methods.

The Quantidex DNA assay is a novel quantitative PCR (qPCR) assay that measures the absolute copy number of PCR-amplifiable DNA in a sample and reports PCR inhibition. The quantitative functional index (QFI) score is defined as the fraction of haploid DNA templates available for amplification compared to the calibrator DNA standard curve. Therefore, a QFI of 100% means all templates (n = 3,030) are available for amplification, and QFI of 5% means 5% of the input templates are amplifiable [26]. Noticeable disadvantages of qPCR over Nanodrop and Qubit are high cost and labor intensive. Simbolo et al. compared the costs and labor involved in DNA quantification by Nanodrop, Qubit, and qPCR [25]. The qPCR-based method is expensive in terms of platform and sample costs. qPCR sample cost was 20 and 3.8-fold higher than those for Nanodrop and Qubit, respectively ($3.00, $0.15, $0.8, respectively). The qPCR operation time is much longer than that for Nanodrop or Qubit (2 hr, 30sec, and 5.3 min respectively) [25].

Here we briefly introduce our quality control workflow for nucleic acids and libraries with regard to performance in downstream sequencing. We use fluorometric method to determine genomic DNA and library concentrations because of the dsDNA specificity and accuracy of the fluorometric methods described above. However, the spectrophotometric method is used for RNA quantitation because its accuracy is similar to the fluorometric methods and is faster and simpler than the fluorometric method. To check genomic DNA integrity, we use the Agilent tape station because the Bioanalyzer does not have a chip for genomic DNA. The Agilent tapestation or Bioanalyzer method is most commonly used to verify library integrity. We are choosing test methods based on the criteria presented in (Figure 1), namely, library concentration and cost efficiency.

Acknowledgment

References

1. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, et al. (2013) ACMG clinical laboratory standards for next-generation sequencing. Genet Med 15: 733-747.

2. Angulo B, Conde E, Suárez-Gauthier A, Plaza C, Martínez R, et al. (2012) A comparison of EGFR mutation testing methods in lung carcinoma: Direct sequencing, real-time PCR and immunohistochemistry. PLoS One 7: e43842.

3. Endris V, Penzel R, Warth A, Muckenhuber A, Schirmacher P, et al. (2013) Molecular diagnostic profiling of lung cancer specimens with a semiconductor-based massive parallel sequencing approach: Feasibility, costs and performance compared with conventional sequencing. J Mol Diagn 15: 765-775.

4. Spencer DH, Sehn JK, Abel HJ, Watson MA, Pfeifer JD, et al. (2013) Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. J Mol Diagn 15: 623-633.

5. Tuononen K, Mäki-Nevala S, Sarhadi VK, Wirtanen A, Rönty M, et al. (2013) Comparison of targeted next-generation sequencing (NGS) and real-time PCR in the detection of EGFR, KRAS, and BRAF mutations on formalin-fixed, paraffin-embedded tumor material of non-small cell lung carcinoma-superiority of NGS. Genes Chromosomes Cancer 52: 503-511.

6. Hadd AG, Houghton J, Choudhary A, Sah S, Chen L, et al. (2013) Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. J Mol Diagn 15: 234-247.

7. Sukhai MA, Craddock KJ, Thomas M, Hansen AR, Zhang T, et al. (2016) A classification system for clinical relevance of somatic variants identified in molecular profiling of cancer. Genet Med 18: 128-136.

8. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, et al. (2013) Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol 31: 1023-1031.

9. Singh RR, Patel KP, Routbort MJ, Aldape K, Lu X, et al. (2014) Clinical massively parallel next-generation sequencing analysis of 409 cancer-related genes for mutations and copy number variations in solid tumours. Br J Cancer 111: 2014-2023.

10. Beltran H, Yelensky R, Frampton GM, Park K, Downing SR, et al. (2013) Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. Eur Urol 63: 920-926.

11. Austin MC, Smith C, Pritchard CC, Tait JF (2016) DNA yield from tissue samples in surgical pathology and minimum tissue requirements for molecular testing. Arch Pathol Lab Med 140: 130-133.

12. Srinivasan M, Sedmak D, Jewell S (2002) Effect of fixatives and tissue processing on the content and integrity of nucleic acids. Am J Pathol 161: 1961-1971.

13. Isola J, De-Vries S, Chu L, Ghazvini S, Waldman F (1994) Analysis of changes in DNA sequence copy number by comparative genomic hybridization in archival paraffin-embedded tumor samples. Am J Pathol 145: 1301-1308.

14. Williams C, Pontén F, Moberg C, Söderkvist P, Uhlén M, et al. (1999) A high frequency of sequence alterations is due to formalin fixation of archival specimens. Am J Pathol 155: 1467-1471.

15. Douglas MP, Rogers SO (1998) DNA damage caused by common cytological fixatives. Mutat Res 401: 77-88.

16. Do H, Dobrovic A (2012) Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil- DNA glycosylase. Oncotarget 3: 546-558.

17. Do H, Wong SQ, Li J, Dobrovic A (2013) Reducing sequence artifacts in amplicon-based massively parallel sequencing of formalin-fixed paraffin-embedded DNA by enzymatic depletion of uracil-containing templates. Clin Chem 59:1376-1383.

18. Sengüven B, Baris E, Oygur T, Berktas M (2014) Comparison of methods for the extraction of DNA from formalin-fixed, paraffin-embedded archival tissues. Int J Med Sci 11: 494-499.

19. Okello JB, Zurek J, Devault AM, Kuch M, Okwi AL, et al. (2010) Comparison of methods in the recovery of nucleic acids from archival formalin-fixed paraffin-embedded autopsy tissues. Anal Biochem 400: 110-117.

20. Turashvili G, Yang W, McKinney S, Kalloger S, Gale N, et al. (2012) Nucleic acid quantity and quality from paraffin blocks: defining optimal fixation, processing and DNA/RNA extraction techniques. Exp Mol Pathol 92: 33-43.

21. Huijsmans CJ, Damen J, Van der Linden JC, Savelkoul PH, Hermans MH, et al. (2010) Comparative analysis of four methods to extract DNA from paraffin-embedded tissues: Effect on downstream molecular applications. BMC Res Notes 3:239-248.

22. Schweiger MR, Kerick M, Timmermann B, Albrecht MW, Borodina T, et al. (2009) Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. PLoS One 4: e5548.

23. Heydt C, Fassunke J, Künstlinger H, Ihle MA, König K, et al. (2014) Comparison of pre-analytical FFPE sample preparation methods and their impact on massively parallel sequencing in routine diagnostics. PLoS One 9: e104566.

24. Khokhar SK, Mitui M, Leos NK, Rogers BB, Park JY (2011) Evaluation of Maxwell® 16 for automated DNA extraction from whole blood and formalin-fixed paraffin embedded (FFPE) tissue. Clin Chem Lab Med 50: 267-272.

25. Simbolo M, Gottardi M, Corbo V, Fassan M, Mafficini A, et al. (2013) DNA qualification workflow for next generation sequencing of histopathological samples. PLoS One 8: e62692.

26. Sah S, Chen L, Houghton J, Kemppainen J, Marko AC, et al. (2013) Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies. Genome Med 5: 77-89.

27. Sie D, Snijders PJ, Meijer GA, Doeleman MW, Van Moorsel MI, et al. (2014) Performance of amplicon-based next generation DNA sequencing for diagnostic gene mutation profiling in oncopathology. Cell Oncol 37: 353-361.