Suitability of melanoma FFPE samples for NGS libraries: time and quality thresholds for downstream molecular tests

David Millán-Esteban[‡,1], Diana Reyes-García[‡,2], Zaida García-Casado[1], José Bañuls[3], José Antonio López-Guerrero[1], Celia Requena, Aranzazu Rodríguez-Hernández[4], Víctor Traves[5] & Eduardo Nagore[*,4]

[1]*Laboratorio de Biología Molecular, Instituto Valenciano de Oncología, Valencia, Spain;* [2]*Universidad Católica de Valencia San Vicente Mártir, Valencia, Spain;* [3]*Servicio de Dermatología, Hospital General Universitario de Alicante, Alicante, Spain;* [4]*Servicio de Dermatología, Instituto Valenciano de Oncología, Valencia, Spain;* [5]*Servicio de Anatomía Patológica, Instituto Valenciano de Oncología, Valencia, Spain*

BioTechniques 65: 79–85 (August 2018) 10.2144/btn-2018-0016
Keywords: DIN • FFPE • library • melanoma • NGS • QC • quality • threshold • time

The use of NGS in clinical practice for precision diagnosis requires a quality starting material. Despite the broadly established use of formalin-fixed paraffin-embedded (FFPE) samples in molecular testing, these usually have low-quality DNA. We established a method to determine the suitability of melanoma FFPE samples for an amplicon-based NGS custom panel analysis. DNA was extracted from unstained melanoma samples and wide local excision samples. Amplicon-based libraries were constructed and tested using time and quality parameters as variables. Time elapsed from sample retrieval >7 years, a quality control value > 5.63 and a DNA integrity value < 2.05 indicated samples were not suitable. A decision tree is provided with rate of samples suitable for analysis according to the combination of these parameters.

Formalin-fixed paraffin-embedded (FFPE) samples represent a valuable under-exploited resource for large, retrospective and prospective studies with long-term clinical follow-up [1]. The existence of FFPE samples collections worldwide, and their advantages such as easy handling, long-term inexpensive storage, suitability for immunohistochemical analyses, and low-cost of large-scale application, make this type of storage a relevant option [2].

The primary goal of formalin fixation is preserving cellular structure for histological examination and diagnosis, so some challenges arise regarding the suitability of genetic material extracted from FFPE samples for molecular testing [3].

Many factors during the fixation process influence nucleic acids conditions. Such as, at the pre-fixation step, the type and amount of tissue or the time elapsed from surgery to fixation. At the fixation step, this could be the type of fixative, temperature, pH or duration. At the post-fixation step, it could be the storage time and conditions [2].

The effects on the genetic material include cross-linking, which decreases the efficacy of posterior PCR [3]. Hence, steps must be taken to improve the use of FFPE in molecular testing, especially those factors affecting downstream techniques, because this will permit the reevaluation of stored FFPE samples independently of their origin [4].

Recently, one of the most commonly used downstream techniques is NGS, which allows sequencing multiple cancer-related genes in a high-throughput manner [5]. Despite the tested use of FFPE tissue in NGS [6], further studies are required to check whether archival FFPE samples are suitable for NGS [7,8]. A recent study found that FFPE samples stored for more than 7 years were not worth being included in NGS studies [9].

It would be helpful to have predictive quality indicators to know whether samples are suitable for NGS. This would avoid wasting precious samples that could undergo alternative molecular studies if not suitable for NGS, an as-yet expensive molecular approach.

In this context, there are parameters to elucidate the quality of the genetic material from FFPE samples. Among others, DNA integrity number (DIN) and RNA integrity number (RIN) from Agilent Technologies (CA, USA), in a range of 1 to 10, give information regarding the fragmentation of the DNA or RNA based on electropho-

METHOD SUMMARY

To maximize sample usability and minimize costs, we proposed a procedure to identify samples with enough quality for massive sequencing. Our proposal included the use of quality control and DNA integrity number quality parameters together with the time elapsed from surgery to develop an affordable protocol that can be easily implemented in research and clinical laboratories.

resis [10,11]. Also, the quality control (QC) value from Illumina (CA, USA) enables the researcher to test the quality of the DNA based on a qPCR, giving values closer to zero when the quality is the highest [12,13].

Our aim was to evaluate the suitability of melanoma FFPE samples for an amplicon-based NGS custom panel analysis according to the storage time, type of sample, QC and DIN values.

Materials & methods

Sample collection

FFPE blocks came from the Biobank at the *Instituto Valenciano de Oncología* (Valencia, Spain). A cohort of 59 samples were analyzed including 37 primary melanoma tumors and 22 wide local excision tissues retrieved and stored at the Pathological Anatomy Department from January 2000 to April 2017. Therefore, the time elapsed from the surgery (time of storage) was up to 17 years and the distribution was made based on a CART analysis regarding the library functionality (see below): 42.4% of the samples were referred to as ‘old’ (>7 years) and 57.6% of the samples were referred to as ‘recent’ (≤7 years). The time of sample fixation in formalin solution was estimated based on the date of the surgery when the sample was taken. Two categories were defined: <1 day (usually overnight fixation), when the day after the surgery was a work day and >1 day when it was a holiday (usually >1–2 days).

This study took place as part of a bigger project that had the approval of the Ethics Committee at the Instituto Valenciano de Oncología, and patients signed a voluntary cession of the samples to the Biobank.

The main outcome variable was the functionality of the sample, which was defined as the ability of a sample to construct an amplicon-based library with a length of 300–350 bp, visible as a single band on an electrophoresis test.

FFPE processing

The Pathological Anatomy Department of our center had a standardized protocol for the processing of FFPE samples. A first formalin fixation of the sample was followed by the block preparation procedure. This was performed in the Excelsior ES automatic processor (ThermoScientific, CA, USA) and included a formalin fixation-step (30 min), an increasing multiple-step dehydration (9 h 45 min), a triple clearing step with xylene (2 h 15 min), and a final three-step embedding in paraffin wax (4 h). Then, blocks were kept at room temperature at the Biobank.

DNA isolation & quantification

From each FFPE block, a 3-μm section was used for hematoxylin and eosin (H&E) stain. Then, a pathologist evaluated and selected the area with tumor-enriched cells for macrodissection. Using the H&E slide as a reference, three 0.6-mm needle biopsies were taken from every primary tumor. For the wide local excision tissue, three 10-μm sections were cut and collected into 1.5-ml tube (Eppendorf).

DNA was extracted using the QIAamp® DNA Investigator kit (QIAGEN, Hilden, Germany), with the following modifications: given the toughness of the skin, we established an overnight incubation at 56°C for the proteinase K to assure a complete digestion. Also, we introduced the optional carrier addition to maximize the extraction yield.

Quantification was obtained using Quant-iT™ PicoGreen™ dsDNA Assay kit (Invitrogen, MA, USA). All samples had a concentration above 2.5 ng/μl and were accepted for the study.

DNA reparation

The NEBNext® FFPE Repair Mix (New England Biolabs, Hertfordshire, UK) kit was used to repair the C:G > T:A changes induced by nucleotide deamination, usually present in FFPE samples.

Quality assessment tests

Real-time PCR was performed using 1X Sybr Green PCR Master Mix (Applied Biosystems) and FFPE QC Kit (Illumina) following the manufacturer’s instructions. Briefly, 2 μl of diluted DNA (1:100) was added to 8 μl of the mix containing SybrGreen and Illumina primers. All runs were processed in an ABI7500 Fast PCR system (Applied Biosystems) using the default run protocol: 50°C/2 min–95°C/10 min–40 cycles of 95°C/30 s, 57°C/30 s, 72°C/30 s). All reactions were performed in triplicates. The resulting QC value was an indicator of the sample quality, with a lower value being the better the quality indicator.

Gel electrophoresis was performed using Genomic DNA ScreenTape in a 4200 TapeStation (Agilent Technologies). Briefly, 1 μl of DNA and 3 μl of sample buffer were added to each well.

DIN value obtained was an indicator of the integrity of the DNA, thus a higher DIN value meant a better quality.

Library construction

Low input DNA libraries of the gene panel containing 21 melanoma-related genes were constructed according to the manufacturer’s instructions using a custom GeneRead DNAseq Panel (QIAGEN). Shortly, DNA fragments were amplified in a multiplex PCR to obtain a total of 633 amplicons of 200 bp in length (GeneRead DNAseq Panel PCR kit V2 Qiagen). At this point, a normalization step was included and 100 ng of each sample continued the process. The ends of the molecules were enzymatically repaired and universal adaptors were ligated, then unique combinations of MID adaptors were ligated (NEBNext Ultra DNA Library Prep Kit, New England Biolabs).

Library functionality

Final library size was checked with a bioanalyzer using D1000 DNA ScreenTape (Agilent Technologies). The final amplicon size including the MID adapters made an average of 350 bp. Thus, the presence of a single band in the range of 300–350 bp classified the library as functional.

Statistical analysis

The statistics were performed using IBM SPSS Statistics 20.0. Normal distribution of continuous variables was checked using the Kolmogorov–Smirnov test. Pearson test was used to study the correlation among parametric variables and Spearman test was used for non-parametric variables. A 1-factor ANOVA test was used to compare means of continuous variables and qualitative ones. Also, continuous variables were categorized with a CART analysis using library functionality as a filter. Diagnostic parameters, including sensitivity, specificity, predicted positive and negative value, accuracy, and area under the curve from a ROC test were calculated to evaluate the capacity of each parameter or algorithm to predict library functionality.

Results & discussion

The study included a total of 59 unpaired FFPE samples corresponding to 37 melanoma primary tumors and 22 unmatched wide local excision tissues stored for a median of 5 years (range: 1–17 years). The characteristics of the samples are displayed in Table 1.

Influence of date of surgery on library functionality

There was a great variability in library yield independent of the tissue origin or the time of storage, despite the existence of a normalization step to include 100 ng in the end reparation step. A 1-factor ANOVA test showed that the mean storage time was significantly lower in samples with functional library than in samples with non-functional library (4.93 vs 9.49 years, respectively; $p < 0.001$). The best cut-off that differentiated functional from non-functional samples was established at 7 years by CART analysis, and recent samples (≤7 years) showed a significantly higher percentage of library functionality than old samples (>7 years; 70.6 vs 20%; $p < 0.001$) (Figure 1A). The fixation time did not influence the functionality (Table 1).

QC & DIN values as quality predictor parameters

QC values were inversely correlated with time of storage ($r = -0.616$; $p < 0.001$) (Figure 1B) and library functionality ($r_s = -0.334$; $p = 0.009$). A 1-factor ANOVA test showed that QC means were statistically different between functional and non-functional samples (4.03 vs 5.7; $p = 0.002$).

According to the CART analysis, cut-off value for QC was established at 5.63 (93.1% of samples with a QC ≤ 5.63 produced functional libraries compared with 6.9% of samples with a QC > 5.63; $p = 0.004$) (Figure 1C).

DIN values showed a direct correlation with the time of storage ($r = 0.523$; $p < 0.001$) (Figure 1D), as well as with library functionality ($r_s = 0.319$; $p = 0.016$). We established a cut-off for DIN value at 2.05 based on CART analysis, and results showed that DIN values greater than 2.05 gave functional libraries in a higher proportion than those less than or equal to 2.05 (82.1 vs 17.9%; $p = 0.023$) (Figure 1E). As expected, QC and DIN were inversely correlated ($r = -0.535$; $p < 0.001$) (Figure 1F).

When looking at the possible differences between tumor tissue and wide local excision samples within old and recent groups, it was found that for the old cohort, lower QC values were more frequent in tumor samples (13/14; 92.9%) than in wide local excision samples (1/14; 7.1%; $p = 0.009$). No difference was found for DIN or functionality in this group. For the recent cohort, functional libraries corresponded in a higher proportion to the wide local excision (15/24; 62.5%) rather than tumor samples (9/24; 37.5%). No difference was found for DIN or QC in this group.

Convergence of parameters in a decision tree

Time of storage, QC and DIN were simultaneously assessed by CART test to analyze their impact on the library functionality, and a decision tree was developed (Figure 2). The storage time was the parameter that better discriminated the library functionality. Samples stored for 7 years or less gave functional libraries in 70.6% of the cases. For samples stored for more than 7 years the value of QC in the first place, and of DIN in the second place, discriminated samples by their functionality. Thus, samples with QC less than or equal to 5.63 and DIN greater than 2.05, allowed identification of a group with 44% of functional libraries.

The diagnostic parameters were evaluated for each variable individually and for the algorithm obtained by CART analysis. The latter was also evaluated only for the oldest samples (Table 2).

The decision tree provided better values in the diagnostic parameters (sensitivity = 91.3%; specificity = 72.9; predicted positive values = 65.1%; predicted negative values = 93.8; accuracy = 79.5%; ROC area under the curve = 0.733) than each parameter individually. In addition, the decision tree restricted to the old samples also showed acceptable figures (ROC = 0.78).

There are several studies in the literature that have evaluated QC and DIN/RIN values prior to NGS [11]. Yakovleva *et al.* established in 2017 a cut-off of RIN of > 2.00 for using FFPE samples in downstream processes [10]. Similarly, Bonfiglio *et al.* proposed in 2016 the use of a DIN cutoff at 3.0 [13]. All these previous studies agree with our findings, but all of them used FFPE samples that had been stored for a maximum of 6 years. Hence, we contribute

Table 1. Sample distribution according to their final library functionality.

	Total		Functional		Non-functional		p-value
	N	%	N	%	N	%	
QC							
QC ≤ 5.63	45	76.2	27	93.1	18	60	0.005
QC > 5.63	14	23.7	2	6.9	12	40	
DIN							
DIN ≤ 2.05	19	33.9	5	17.9	14	50	0.023
DIN > 2.05	37	66.1	23	82.1	14	50	
Time of storage							
Old	25	51	5	17.2	20	66.7	<0.001
Recent	34	57.6	24	82.8	10	33.3	
Fixation time							
<1 day	25	44.6	15	51.7	10	33.3	0.3
≥1 day	11	19.6	5	17.2	6	20	
Not available	20	35.7	8	27.6	12	40	

p-value by Chi-squared test. DIN: DNA integrity value; N: Number; QC: Quality control.

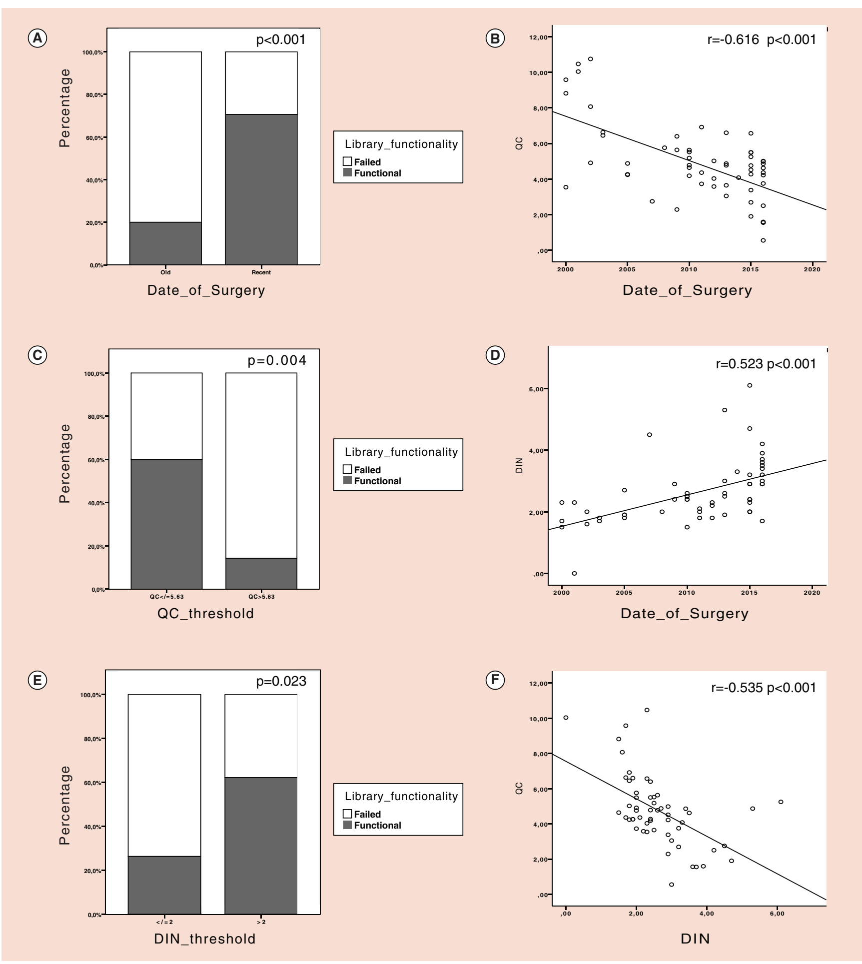


Figure 1. Statistical analysis. Here are presented the graphical distribution of failed (white) and functional (gray) libraries for the different variables (Date of surgery, QC threshold and DIN threshold) **(A, C, E)**. Also, the correlations studied between QC and date of surgery **(B)**, DIN and date of surgery **(D)**, and QC and DIN **(F)**. DIN: DNA integrity value; QC: Quality control.

valuable information regarding samples stored for longer periods, with a functional analysis on constructed amplicon-based libraries and not only on genomic DNA. These findings suggest that older FFPE samples should preferably be used for pathological analysis and molecular tests that do not require such quality starting material.

A flow-chart proposal was developed including all variables to determine the best approach when working with FFPE samples, taking into consideration the quality parameters QC and DIN as well as the time of storage (Figure 2). A deeper analysis of the diagnostic parameters of this proposal pointed out the utility of our approach given its accuracy of up to 79.5%. This could be translated into the practical work and suggests that, if willing to have a small percentage of compromise, all recent samples could be used. Then, only by using the QC value as a predictor parameter, those samples stored for >7 years could be included if their QC is ≤ 5.63 with an expected adequate functionality in 35.7% of samples. The use of the DIN value would increase the specificity in those studies where genetic material is precious and must be highly optimized and the expected adequate functionality could be increased up to 44.4% of the samples.

This study has demonstrated that QC (≤5.63) and DIN (>2.05) are able to discriminate between functional and non-functional samples beyond storage time, particularly for old samples. It showed that QC and DIN were appropriate quality parameters, for which values differed between old and recent samples, and between functional and non-functional libraries. On the other hand, differences were found in QC between tumor samples and wide local excision samples, although they were not relevant for functionality in old samples. In recent samples, wide local excision samples worked better than tumors in terms of functionality, which might be explained by the role of melanin in tumor samples, which can lead to inhibition of PCR due to the association of remnants of pigment melanin with genomic DNA [14,15].

The strengths of this study include the performance over amplicon-based libraries instead of genomic DNA, as seen in previous studies. Also, the results can be extrapolated to different approaches. As a limitation, the modest size of the cohort should be highlighted, so future studies must corroborate these findings.

In conclusion, the storage time was the most important variable that influenced sample viability for amplicon-based library construction. The addition of QC and DIN helped refine the rate of samples suitable for NGS and particularly to identify which ones within old samples could be used in this regard.

Table 2. Diagnostic test parameters for each variable and the algorithm obtained by CART analysis.

	S (%)	SP (%)	PPV (%)	PNV (%)	A (%)	ROC (AUC)
QC	80.8	68.2	60.0	85.7	72.9	0.33
DIN	70.3	66.1	62.2	73.7	68.0	0.66
Time of storage	77.9	73.1	70.6	80.0	75.3	0.75
CART†	91.3	72.9	65.1	93.8	79.5	0.73
CART (for old samples)‡	87.7	62.8	44.4	93.8	69.1	0.78

'CART' includes the results for the diagnostic parameters of the proposed flowchart obtained by CART analysis.
†Positive for the test if date of surgery < 7 years or if QC ≤ 5.63 and DIN > 2.05.
‡Only for old samples, with a date of surgery > 7 years. Positive if QC ≤ 5.63 and DIN > 2.05
A: Accuracy; PNV: Predicted negative values; PPV: Predicted positive values; QC: Quality control; S: Sensitivity; SP: Specificity.

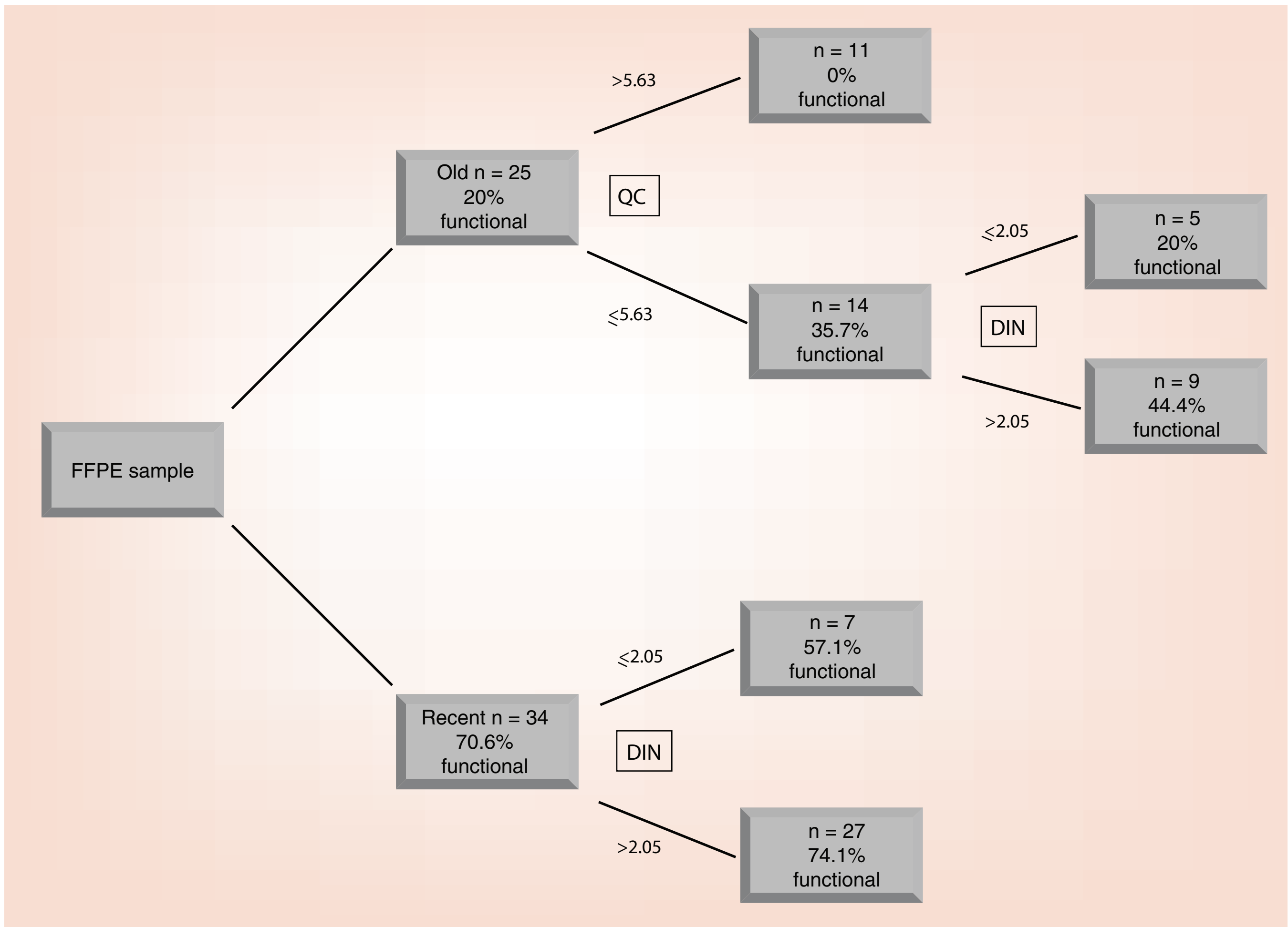


Figure 2. Decision tree to optimize the use of FFPE samples according to our results.
DIN: DNA integrity value; QC: Quality control.

Author contributions

DM, ZG and EN conceived and designed the experiments. JAL provided the technology. JB, CR and VT assessed sample identification. DR and AR managed sample collection and preparation. DM and DR performed the experiments. DM and EN analyzed the data. ZG and EN supervised the study. DM wrote the paper. EN corrected the manuscript.

Acknowledgments

We would like to acknowledge the patients and the IVO Biobank, integrated in the Spanish National Biobank Network and in the Valencian Biobank Network, for their collaboration.

Financial & competing interests disclosure

This study has been funded by the Instituto de Salud Carlos III grant number PI15/01860, the Junta Provincial de Valencia de la Asociación Española contra el Cancer through a PhD grant, and by the Universidad Católica de Valencia San Vicente Mártir. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Creative Commons Attribution 4.0 License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

References

1. Seiler C, Sharpe A, Barrett J, Harrington E, Jones E, Marshall G. Nucleic acid extraction from formalin-fixed paraffin-embedded cancer cell line samples: a tradeoff between quantity and quality? *BMC Clin. Pathol.* 16, 17 (2016).
2. Ludyga N, Grünwald B, Azimzadeh O *et al.* Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. *Virchows Arch.* 460(2), 131–140 (2012).
3. Wong S, Li J, Tan A *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* 7, 23 (2014).
4. Greytak S, Engel K, Bass B, Moore H. Accuracy of molecular data generated with FFPE biospecimens: lessons from the literature. *Cancer Res.* 75(8), 1541–1547 (2015).
5. Do H, Wong S, 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.* 59(9), 1376–1383 (2013).
6. Bettoni F, Koyama F, de Avelar Carpinetti P, Galante P, Camargo A, Asprino P. A straightforward assay to evaluate DNA integrity and optimize next-generation sequencing for clinical

diagnosis in oncology. *Exp. Mol. Pathol.* 103(3), 294–299 (2017).

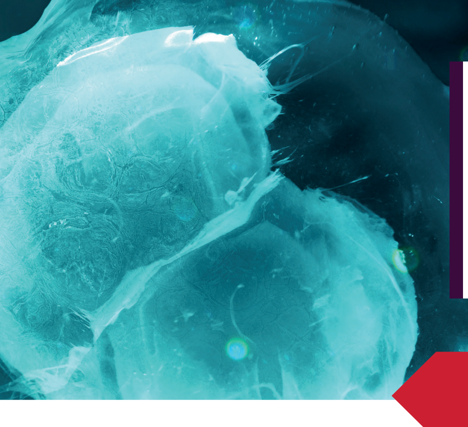
7. Arreaza G, Qiu P, Pang L *et al.* Pre-analytical considerations for successful next-generation sequencing (NGS): challenges and opportunities for formalin-fixed and paraffin-embedded tumor tissue (FFPE) samples. *Int. J. Mol. Sci.* 17(12), 1579 (2016).
8. Carrick D, Mehaffey M, Sachs M *et al.* Robustness of next generation sequencing on older formalin-fixed paraffin-embedded tissue. *PLoS One* 10(7), e0127353 (2015).
9. Nagahashi M, Shimada Y, Ichikawa H *et al.* Formalin-fixed paraffin-embedded sample conditions for deep next generation sequencing. *J. Surg. Res.* 220, 125–132 (2017).
10. Yakovleva A, Plieskatt J, Jensen S *et al.* Fit for genomic and proteomic purposes: sampling the fitness of nucleic acid and protein derivatives from formalin fixed paraffin embedded tissue. *PLoS One* 12(7), e0181756 (2017).
11. Bonnet E, Moutet M, Baulard C *et al.* Performance comparison of three DNA extraction kits on human whole-exome data from formalin-fixed paraffin-embedded normal and tumor samples. *PLoS One* 13(4), e0195471 (2018).
12. Bonfiglio S, Vanni I, Rossella V *et al.* Performance comparison of two commercial human whole-exome capture systems on formalin-fixed paraffin-embedded lung adenocarcinoma samples. *BMC Cancer* 16(1) (2016).
13. Mathieson W, Guljar N, Sanchez I, Sroya M, Thomas G. Extracting DNA from FFPE tissue biospecimens using user-friendly automated technology: is there an impact on yield or quality? *Biopreserv. Biobank* 16(13) (2018).
14. Price K, Linge C. The presence of melanin in genomic DNA isolated from pigmented cell lines interferes with successful polymerase chain reaction. *Melanoma. Res.* 9(1), 5–9 (1999).
15. Chung J, Choi J, Sears J *et al.* A melanin-bleaching methodology for molecular and histopathological analysis of formalin-fixed paraffin-embedded tissue. *Lab. Invest.* 96(10), 1116–1127 (2016).

First draft submitted: 19 March 2018; Accepted for publication: 6 June 2018

Address correspondence to: Eduardo Nagore; Servicio de Dermatología, Instituto Valenciano de Oncología, Valencia, Spain; E-mail: eduardo_nagore@ono.com

‡Authors contributed equally to this paper

To purchase reprints of this article contact: s.cavana@future-science.com



Regenerative Medicine

ISSN: 1746-0751
Frequency per year: 8

IMPACT FACTOR:
2.868 (2016)

In association with

www.regmednet.com

INDEXING

MEDLINE/Index Medicus, EMBASE/Excerpta Medica, EMCare, Chemical Abstracts, Science Citation Index Expanded™ (SciSearch®), Biological Abstracts, BIOSIS Previews, Biotechnology Citation Index®, Journal Citation Reports/Science Edition®, Scopus

www.futuremedicine.com

To claim your free trial, contact us at: trials@futuremedicine.com