Experimental and Molecular Pathology 103 (2017) 294–299

Contents lists available at ScienceDirect

Experimental and Molecular Pathology

journal homepage: www.elsevier.com/locate/yexmp

A straightforward assay to evaluate DNA integrity and optimize next-generation sequencing for clinical diagnosis in oncology

Fabiana Bettoni[a,1], Fernanda Christtanini Koyama[a,1], Paola de Avelar Carpinetti[b], Pedro Alexandre Favoretto Galante[a], Anamaria Aranha Camargo[a], Paula Fontes Asprino[a,*]

[a] Hospital Sírio-Libanês, São Paulo, SP, Brazil
[b] Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Viçosa, Viçosa, MG, Brazil.

ABSTRACT

Next generation sequencing (NGS) has become an informative tool to guide cancer treatment and conduce a personalized approach in oncology. The biopsy collected for pathologic analysis is usually stored as formalin-fixed paraffin-embedded (FFPE) blocks and then availed for molecular diagnostic, resulting in DNA molecules that are invariably fragmented and chemically modified. In an attempt to improve NGS based diagnostics in oncology we developed a straightforward DNA integrity assessment assay based on qPCR, defining clear parameters to whether NGS sequencing results is accurate or when it should be analyzed with caution.

We performed DNA extraction from 12 tumor samples from diverse tissues and accessed DNA integrity by straightforward qPCR assays. In order to perform a cancer panel NGS sequencing, DNA library preparation was performed using RNA capture baits. Reads were aligned to the reference human genome and mutation calls were further validated by Sanger sequencing. Results obtained by the DNA integrity assays correlated to the efficiency of the pre-capture library preparation in up to 0.94 (Pearson's test). Moreover, sequencing results showed that poor integrity DNA leads to high rates of false positive mutation calls, specially C:G > T:A and C:G > A:T.

Poor quality FFPE DNA samples are prone to generating false positive mutation calls. These are especially perilous in cases in which subclonal populations are expected, such as in advance disease, since it could lead clinicians to erroneous conclusions and equivocated conduct.

1. Introduction

Next generation sequencing (NGS) has become an informative tool to guide cancer treatment and conduce a personalized approach in oncology. By accessing tumor's somatic variations and gene expression alterations, it is possible to refine the diagnosis and predict the tumor response to specific drugs.

The biopsy collected for pathologic analysis is usually stored as formalin-fixed paraffin-embedded (FFPE) blocks. Routinely, this material is also availed for molecular diagnostic, however, it is necessary to ensure that the material has satisfactory quality for the analysis, otherwise, the laborious and expensive NGS test may result in unreliable information.

The use of nucleic acids purified from FFPE samples imply in using biomolecules that are invariably fragmented and chemically modified, varying in its intensity according to factors such as fixation delay, thickness of tissue, formalin quality, duration of the fixation process, period of time and temperature conditions in which the sample has been stored (Do and Dobrovic, 2015; Quach et al., 2004; Srinivasan and Sedmak, 2002; von Ahlfen et al., 2007; Williams et al., 1999). Although there are DNA and RNA extraction methods optimized to circumvent the fixation effects, some characteristics may be irreversible and result in reduced efficiency of enzymatic assay or even introduction of artifactual results(Williams et al., 1999).

The best alternative to FFPE samples is fresh frozen (FF) tissues. This kind of material is ideal for molecular biology studies yielding higher quality material; nevertheless, it requires special infrastructure adaptations such as − 80 °C freezers or nitrogen tanks. According to many reports in the literature comparing NGS sequencing of paired FF and FFPE samples (Hedegaard et al., 2014; Oh et al., 2015; Schweiger et al., 2009; Spencer et al., 2013; Van Allen et al., 2014) a good correlation has been observed. However, these tests are usually performed with good quality material since all authors agree FFPE contains DNA of lower quality, susceptible to false mutation calls.

[*] Corresponding author at: Centro de Oncologia Molecular, Instituto Sírio-Libanês de Ensino e Pesquisa, Hospital Sírio-Libanês, Rua Prof. Daher Cutait, n° 69, Pesquisa-Bela Vista, São Paulo, SP CEP 01308-060, Brazil.

E-mail address: pasprino@mochsl.org.br (P.F. Asprino).

[1] Both authors contributed equally to this work and are listed alphabetically.

https://doi.org/10.1016/j.yexmp.2017.11.011
Received 4 May 2017; Received in revised form 16 October 2017; Accepted 14 November 2017
Available online 21 November 2017

In a study with diverse tumor tissues using FFPE samples, preparation of library for sequencing was unsuccessful for 70% (43/61) owning to bad DNA integrity (Hedegaard et al., 2014). Using low quality material may result in the obvious unfeasibility of performing the NGS assay, but another concern is the high rate of false positive mutation calls, especially in cases which extra modified DNA mass is necessary to perform the test. As described by previous studies, the false call rate can be as high as 40%, especially when the alteration is expected to be subclonal and the nucleotide substitution inquired is especially prone to artifacts, as the insidious C > T alteration. The EGFR c.2369C > T (T790 M) substitution, inquired during the treatment with EGFR inhibitors is a good example of substitution that may be mistaken for a real mutation (Ye et al., 2014), leading to equivocated therapeutic approach.

There are commercial kits available to indicate the integrity of FFPE extracted DNA based on its fragmentation and modification level. However, some kits do not inform the molecule size evaluated, other kits evaluate the integrity of DNA fragments much shorter (~40–120 bp) than what is used by most sequencing library preparation methods (~150–250 bp), informing a biased and unreal quality.

Some studies have also described alternative methods for evaluating FFPE extracted DNA and the form of circumventing modifications is by starting library prep with higher DNA mass (Dang et al., 2016; Sah et al., 2013). Still, all suggest enhancing DNA mass and increasing PCR cycling to compensate for poor quality DNA but does not shown the implications of conducing such approaches. Importantly, neither establish a cutoff in which the FFPE material should not be used for sequencing tests at the cost of calling artifacts for mutations.

Taken together, testing FFPE samples for integrity and establishing integrity parameters is essential to perform accurate clinical diagnostic NGS assays. The lack of integrity standards may generate inconsistent and non-reproducible results. In an attempt to improve NGS based diagnostics in oncology, we developed a straightforward DNA integrity assessment assay which can be used to estimate fragmentation and modification levels for DNA extracted from FFPE samples. We also established integrity parameters to optimize NGS library preparation and demonstrate the implications of sequencing poor quality samples.

2. Methods

2.1. Sample preparation

Twelve FFPE samples, including tumor tissues from breast, colon, lung, pancreas, uterus and thyroid were used. Glass slides were prepared, one of each submitted to hematoxylin and eosin staining, used to guide macrodissection. Four to ten unstained slides, depending on the area to be macrodissecated, of 5 μm thick tissue were manually scraped. The minimum accepted area was 25 mm^2. Tumor cellularity varied from 50 to 80% of the area. Genomic DNA was extracted using GeneRead DNA FFPE kit (Qiagen), containing Uracyl-D Glycosylase, according to the manufacturer's instructions. DNA concentration was measured in a Qubit fluorometer using dsDNA BR Assay Kit (Thermo Fisher Scientific).

2.2. Primers design

Primers were designed in the intronic region of the MLH1 gene in regions of low SNP content, checking for specificity in the genome using primer Blast (Jian et al., 2012) and in-silico PCR (Kent et al., 2002). This region is not prone to copy number variation and does not present retrocopies or pseudogenes, as based on our methodology of retrocopies and pseudogenes identification described elsewhere (Galante, 2015; Navarro and Galante, 2013).

PCR efficiency was estimated using the formula described by Pffafl (Pfaffl, 2001), checking melting curves for a single amplification product (supplementary).

Primer 78bp_F: AAAGGCCCAAAGTGTGAAATG; Primer 78bp_R: CAACTGGGACAGAGCAAATGG; Primer 254bp_F: AATTCCCAAGCAGA TGAATGC; Primer 254bp_R: ATAACGGAGGAACAAGGGTTG.

2.3. Quality assessment test

Real time PCR was performed using 1X Sybr Green PCR master mix (Thermo Fisher Scientific), 360 nM of primers defining 78 bp and 254 bp amplicons, separately, 1 ng template DNA in a 20 μL reaction. All runs were processed in a 7300 PCR system (Thermo Fisher Scientific) using the default run protocol of the equipment: 95 °C/10 min – 40 cycles of 95 °C/15 s, 60 °C/60 s – melting curve program: 95 °C/15 s, 60 °C/30 s, 95 °C/15 s). DNA obtained from leucocytes of a healthy person was quantified using a fluorometer (Qubit) and used as reference to define a concentration curve comprised of 5 points ranging from 10 ng to 16 pg. The percentage of amplifiable DNA was calculated as the quantity obtained by qPCR, divided by the initial quantity of DNA assumed (quantity obtained/1 ng). All reactions were performed in duplicates.

2.4. Library preparation, sequencing and mapping

Low input DNA libraries of the institutional cancer gene panel containing the coding region of 493 cancer genes plus intronic regions of 19 genes frequently rearranged in cancer (2.5 Mb) were constructed according to the manufacturer's instructions using Custom SureSelect capture kit (Agilent). Briefly, 200 ng of genomic DNA were sheared into 100–300 bp fragments using a Covaris S2 sonicator (Covaris Inc). The ends of the molecules were enzymatically repaired and adenylated at the 3´end before submitting to paired-end adaptors ligation. The adaptor ligated molecules were PCR amplified (10 cycles), purified and concentration was accessed by a Qubit fluorometer; while the profile was checked using the DNA 1000 Bioanalyzer assay (Agilent). Samples were then hybridized to capture libraries using 120 nt RNA baits and selected using streptavidin beads. The final library was PCR amplified (12 cycles), purified, checked for concentration using a Qubit fluorometer (Thermo Fisher Scientific) and quality, using the Bioanalyzer High Sensitivity DNA analysis kit (Agilent).

Libraries were sequenced on MiSeq platform (Illumina) and reads were aligned to the human reference genome (version GRCh377/hg19) using bwa (Li and Durbin, 2009). Variations were called using GATK (DePristo et al., 2011; McKenna et al., 2010; Van der Auwera et al., 2013) and SNV annotation was performed using wANNOVAR (Chang and Wang, 2012; Wang et al., 2010). Only variation on our target region and covered by a total of at least 100 reads in total (reference and variant allele), 15% of variant reads and 0.01 of minor allele frequency (MAF) reported by the Exome Aggregation Consortium (ExAC) (Consortium et al., 2015), were considered as mutation.

2.5. Sanger sequencing

Primers flanking the reported mutations were designed. Sequencing reactions were performed using Big Dye Terminator (Thermo Fisher Scientific) and conditions followed the manufacturer's indications. Chromatograms were manually inspected.

2.6. Statistical analysis

Statistical analyses were performed using GraphPad Prism 4. The correlation was performed using Pearson distribution in a 95% of confidence interval, the same parameter was used for linear regression.

3. Results

Twelve FFPE tumor samples from six tissues (breast, colon, lung, thyroid, pancreas and uterus) were analyzed in this study. We evaluated

Table 1
Parameter of sample quality was assessed by pre capture library yield (≥ 750 ng).

	Library yield (ng)	gDNA (ng)	% amplifiable (78 bp)	% amplifiable (254 bp)	Ratio 78/254 bp	Log10 (%254 bp)
Thyroid 1	1176	3400	123.517	45.845	2.7	1.7
Uterus 1	899	1056	64.259	24.739	2.6	1.4
Lung 3	860	860	116.239	34.878	3.3	1.5
Colon 2	843	4284	64.074	7.295	8.8	0.9
Lung 1	570	425	28.79	0.824	34.9	− 0.1
Breast 1	492	3672	36.894	0.468	78.9	− 0.3
Colon 3	489	3744	27.817	0.351	79.2	− 0.5
Colon 4	412	1264	74.965	0.628	119.4	− 0.2
Colon 1	351	1296	19.103	0.254	75.2	− 0.6
Lung 2	324	677	8.437	0.041	207.3	− 1.4
Breast 2	105	3348	7.478	0.001	6989	− 3
Pancreas 1	*NA*	82	60.744	0.486	125	− 0.3

Samples were also evaluated for initial gDNA mass, percentage of molecules amplified at 78 bp and 254 bp and fragmentation status, based on the ratio of 78 bp/254 bp. NA: not available.

the amount of purified DNA, as well as the percentage of amplifiable material at 78 bp and 254 bp. As an additional parameter to estimate DNA fragmentation, we established a ratio of 78 bp/254 bp amplifiable fragments. The higher this value, the more fragmented the sample. In a high integrity DNA sample both values would be very close to 100% and the ratio would be approximately 1.

Although all libraries were prepared with the same quantity of DNA input (200 ng), there was a great variability in library yield, with no correlation to the tissue of origin or initial DNA extraction yield (Table 1). Amplification depends on the availability of proper size DNA fragments to serve as template for the reaction as well as DNA molecules deprived of modifications that impede polymerase from extending. In the presence of such modifications, DNA polymerase may stall at the damaged base, restricting PCR, or continue through the lesion, allowing amplification at the price of a potential false mutation call. Therefore, we considered the mass obtained for pre capture library as a parameter of DNA integrity (Table 1) since this step depends on adaptor ligation and PCR amplification, influenced by the degree of DNA modification and fragmentation.

The manufacturer recommends the use of 750 ng of pre capture library DNA (minimum recommended of 500 ng) to proceed to hybridization. Using this value as a cutoff, we had 4 good quality samples (Thyroid 1, Uterus 1, Lung 3, and Colon 2), all of which amplified at least 5% of the 254 bp assay and yielded good NGS mutation calling results (Table 1).

As the availability of shorter fragments is higher, the efficiency of the 78 bp reaction is superior. Although this reaction may serve as a positive input control, performing the 254 bp qPCR reaction is a better approach since it contemplates both fragmentation and modification status of DNA molecules. The 254 bp test alone presented the best correlation to pre capture library yield ($\rho = 0.86$; p-value = 0.0007; Person's correlation test). Since there is a wide variation in amplification efficiency in this test among the samples, ranging from 0.001 to 45%, we used logarithmic scale to plot these data. In this case, Pearson's correlation test reached even higher scores: $\rho = 0.94$, p-value < 0.0001 (Fig. 1).

Samples amplifying at least 0.25% of the 254 bp assay yielded a minimum of 350 ng of pre-capture library and sequencing results were informative, being considered “acceptable samples”. However, those amplifying < 0.25% of the 254 bp assays yielded as low as 105 ng of pre-capture library and, proceeding through NGS sequencing, presented considerable number of false positive mutation calls (Fig. 2).

Sequencing runs were performed with the available material at that point and somatic mutations were called. Only mutations presenting coverage equal or higher than 100 × and percentage of variant reads higher than 15% and MAF above 1%, reported by the ExAC, were maintained.

Although some variation in the number of mutations is expected, the number of mutations called for the poor quality samples (those amplifying < 0.25% of 254 bp assay) was much higher. For example, the poorest sample in the cohort, “breast 2” (amplified 0.001% of 254 bp assay) presented 296 mutations, 10.5 times more mutations than what was observed in “breast 1” (amplified 0.47% of 254 bp assay) (Table 2).

As depicted in Fig. 2, it was evident the presence of two types of substitutions in this poor quality sample: C:G > T:A, the most frequent artifact described, caused by deamination of cytosine. We also found C:G > A:T putative artifacts, caused by the acoustic shearing of DNA samples in the presence of reactive contaminants from the extraction process, during the library preparation process (Costello et al., 2013).

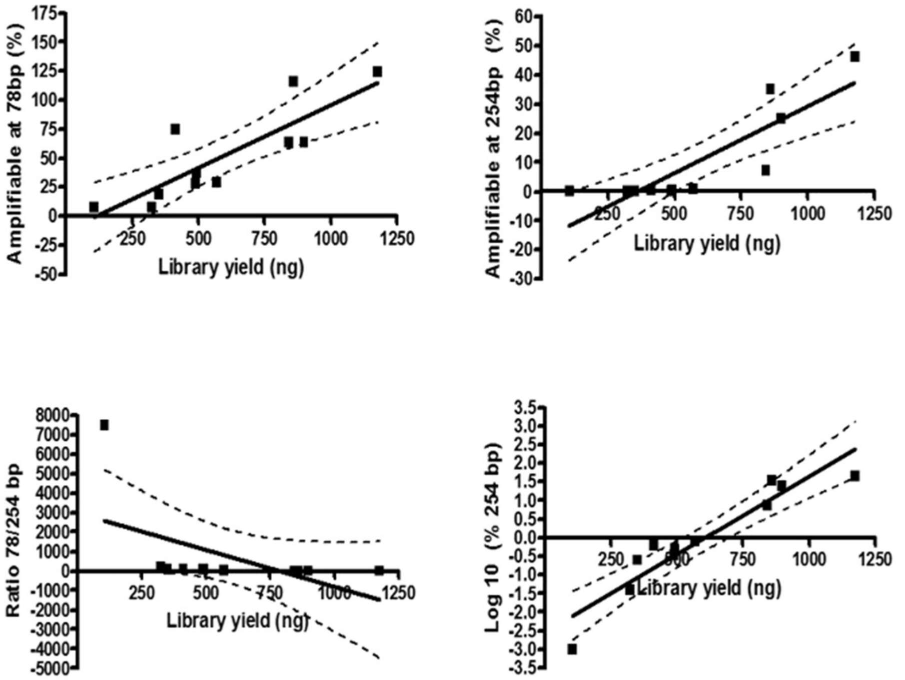
Some mutations were randomly assigned for validation by Sanger sequencing as there was insufficient material to validate them all. Besides the quantity availability, validation of poor quality material is technically challenging. As it happens for NGS library constructions, PCR reactions fail to amplify either because of fragmented and modified DNA or due to the presence of reaction inhibitors. While good quality samples confirmed 100% of tested variations, poor quality samples were not as veracious. For the poorest sample, breast 2, only 16% of called mutation that could be effectively tested were validated (Table 2).

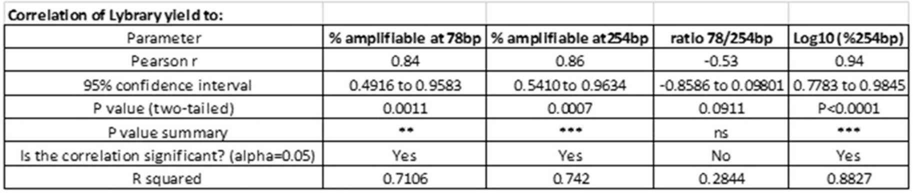
Combining the efficiency obtained in the library prep step with the results from final sequencing we assumed that good quality samples are those amplifying at least 5% of the 254 bp assay (Colon 2, Uterus 1, Lung 3, and Thyroid 1). Samples amplifying between 0.25% and 5% are acceptable and should be carried on with caution. Samples below 0.25% of efficiency are of very poor quality and NGS sequencing is not recommended at the cost of high level of false positive mutation calls.

4. Discussion

Nowadays we are living in a new paradigm in which, more than defining the original site of the primary tumor, it is necessary to describe the molecular profile in order to describe the pathology as well as define the treatment. In this new scenario, immunohistochemistry is being complemented by NGS approaches. The preservation of tissue samples in FFPE has been the method of choice for decades, mostly because it preserves morphological features of the original tissue, is practical for storage and useful for immunohistochemistry analysis. However, this is not ideal for molecular biology analysis. Consequently, evaluation of quality of the biomolecules is crucial for analyzing critically NGS data, since it might convey false results that would impair the right decision for patient treatment.

Here we performed simple and cost effective qPCR assays capable of evaluating DNA modification and fragmentation status. We also showed that the percentage of amplifiable material at 78 bp fragments indicates predominantly the level of DNA modification, which is responsible for





Correlation of Lybrary yield to:				
Parameter	% amplifiable at 78bp	% amplifiable at 254bp	ratio 78/254bp	Log10 (%254bp)
Pearson r	0.84	0.86	-0.53	0.94
95% confidence interval	0.4916 to 0.9583	0.5410 to 0.9634	-0.8586 to 0.09801	0.7783 to 0.9845
P value (two-tailed)	0.0011	0.0007	0.0911	P<0.0001
P value summary	**	***	ns	***
Is the correlation significant? (alpha=0.05)	Yes	Yes	No	Yes
R squared	0.7106	0.742	0.2844	0.8827

Fig. 1. Sample quality assessment considering pre capture library efficiency. Pearson's analysis indicates the correlations between library yield (ng) and percentage of amplifiable molecules at 78 bp PCR assay, percentage of amplifiable molecules at 254 bp PCR assay, ratio of 78 bp/254 bp assays and log of the amplifiable molecules at 254 bp assay.

decreasing molecular reactions efficiency and generating false positive mutation calls. This reaction therefore serves as a template control and is a good indicator of the presence of PCR inhibitors, which frequently contaminate FFPE extracted DNA.

We also showed that the percentage of amplifiable material evaluated with the 254 bp amplicon reflects both DNA modification and fragmentation status, being an excellent indicator of DNA integrity. Although still a relatively small DNA fragment, it represents the size necessary to prepare most of the libraries used to conduce NGS sequencing. Frequently the percentage of material available to perform such test is limited and, consequently, NGS library yield is extremely low. Samples that do not amplify at the 254 bp test are either too modified and/or too fragmented and are not indicated for further NGS analysis.

The concentration curve is used not only to quantify the samples but also to infer the efficiency of PCR reaction in every run. We make sure the dissociation curve shows no unspecific amplification and that the standard curve represents an efficient reaction (supplementary).

In an attempt to overcome some types of DNA modifications, the use of uracil-DNA-glycosylase (UDG) in the extraction step is critical, since

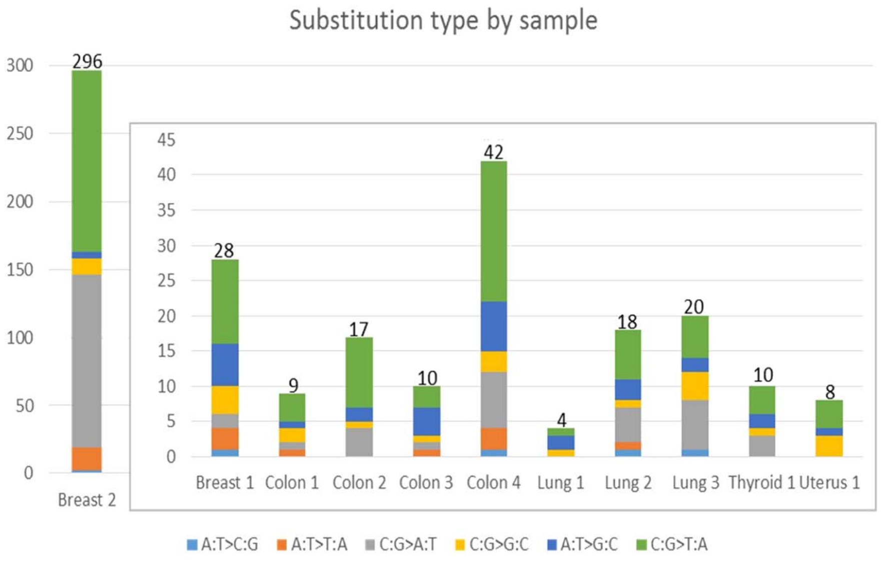


Fig. 2. Distribution of mutation type by sample. Each color represents a specific mutation type. Above each bar, numbers indicate the total number of mutations.

Table 2
Number of mutations called and the validation performance. Comparison between samples.

	% amplifiable (254 bp)	# mutations	Tested	Validated	% validated
Breast 2	0.001	296	6	1	16.6
Lung 2	0.041	18	4	3	75
Colon 1	0.254	9	1	1	100
Colon 3	0.351	10	5	5	100
Breast 1	0.468	28	5	5	100
Colon 4	0.628	42	5	2	40
Lung 1	0.824	4	NA	NA	NA
Colon 2	7295	17	9	9	100
Uterus 1	24,739	8	NA	NA	NA
Lung 3	34,878	20	3	3	100
Thyroid 1	45,845	10	4	4	100

"% amplifiable (254 bp)" represents the amount of 254 bp DNA amplified; "# mutations", represents the number of mutations called by NGS; "tested" represents the number of mutations tested by Sanger sequencing and "validated" represents the number of mutations called by NGS that was validated by Sanger sequencing. NA: not available, there was insufficient DNA for performing Sanger sequencing validation.

it guarantees reduction of the main artifact described in FFPE DNA samples, the cytosine deamination, which implicates in C:G > T:A transition. However, depending on the circumstances in which the sample has been prepared and stored, UDG may not fulfill its potential and artifactual C > T transitions may still be detected (Kim et al., 2016). Another consideration is that by releasing free uracils from the altered DNA, UDG impairs the polymerase from replicating the error by staling the polymerase and reduces PCR efficiency.

Another frequent artifact, C:G > A:T transversion, is common when the library preparation step requires shearing by sonication. Powerful acoustic sonication alone is not sufficient to cause this type of artifact; yet, it has been demonstrated that in the presence of oxidative contaminants the DNA is more susceptible to damage due to the generation of 8-oxoG (Costello et al., 2013). Although other protocols to prepare libraries can substitute this step, the alternatives depend on restriction enzyme DNA recognition and polymerase amplification, likewise influenced by the level of DNA modification. Despite FFPE extracted DNA being already fragmented, breaking points were mostly created in regions of lesion. For this reason, it is recommended to prepare the molecules so that the ligation to the library preparation primers can be effective.

Most of all, we showed that library preparation of very poor quality samples is viable. Although it is not ideal, some protocols try to compensate poor quality material by increasing the input DNA mass, when available, and/or increasing the number of PCR cycles. Trying to compensate quality with quantity is deceptive since by adding more of the same damaged DNA would only enhance the variety of damaged positions to be mistakenly called for mutations, increasing even more the rate of false positive calls. Here we show that even though libraries from poor quality DNA can be successfully prepared, a very high number of false positive mutations is called, being hard to distinguish real mutations from artifacts. What could be considered in this situation is restricting the mutations called by frequency, adding bioinformatics tools to filter out some know specific artifacts (Costello et al., 2013), but also loosing low frequency clonal mutations and accepting beforehand that much of the real information may be lost. In these cases, validation by orthogonal methods is highly recommended.

Nowadays protocols use small amounts of DNA to prepare libraries. Although it represents a gain to study restricted materials such as biopsies, it also means that artifacts can be amplified, making it even harder to tell real mutations apart from errors, reason why quality of starting material should be considered beforehand.

5. Conclusion

Here we present a straightforward assays that can be used to guide the decision of using FFPE samples in NGS runs, defining objective parameters and illustrating the consequences of running poor quality specimen. Preparing and sequencing poor quality samples may implicate in description of artifacts mistaken for mutations, leading to wrong judgment calls.

Funding

This work was supported by Sociedade Beneficente de Senhoras Hospital Sírio-Libanês and Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP. The funding sources had no involvement in the conduct of the research or in the manuscript preparation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yexmp.2017.11.011.

References

von Ahlfen, S., Missel, A., Bendrat, K., Schlumpberger, M., 2007. Determinants of RNA quality from FFPE samples. PLoS One 2, 1–7. http://dx.doi.org/10.1371/journal.pone.0001261.

Chang, X., Wang, K., 2012. wANNOVAR: annotating genetic variants for personal genomes via the web. J. Med. Genet. 49, 433–436. http://dx.doi.org/10.1136/jmedgenet-2012-100918.

Consortium, E.A., Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fennell, T., O'Donnell-Luria, A., Ware, J., Hill, A., Cummings, B., Tukiainen, T., Birnbaum, D., Kosmicki, J., Duncan, L., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G., Poplin, R., Rivas, M., Ruano-Rubio, V., Rose, S., Ruderfer, D., Shakir, K., Stenson, P., Stevens, C., Thomas, B., Tiao, G., Tusie-Luna, M., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Roberto, E., Florez, J., Gabriel, S., Getz, G., Glatt, S., Hultman, C., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M., McGovern, D., McPherson, R., Neale, B., Palotie, A., Purcell, S., Saleheen, D., Scharf, J., Sklar, P., Sullivan, P., Tuomilehto, J., Tsuang, M., Watkins, H., Wilson, J., Daly, M., MacArthur, D., 2015. Analysis of Protein-Coding Genetic Variation in 60,706 Humans, bioRxiv. http://dx.doi.org/10.1101/030338.

Costello, M., Pugh, T.J., Fennell, T.J., Stewart, C., Lichtenstein, L., Meldrim, J.C., Fostel, J.L., Friedrich, D.C., Perrin, D., Dionne, D., Kim, S., Gabriel, S.B., Lander, E.S., Fisher, S., Getz, G., 2013. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. 41, 1–12. http://dx.doi.org/10.1093/nar/gks1443.

Dang, J., Mendez, P., Lee, S., Kim, J.W., Yoon, J.H., Kim, T.W., Sailey, C.J., Jablons, D.M., Kim, I.J., 2016. Development of a robust DNA quality and quantity assessment qPCR assay for targeted next-generation sequencing library preparation. Int. J. Oncol. 49, 1755–1765. http://dx.doi.org/10.3892/ijo.2016.3654.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498. http://dx.doi.org/10.1038/ng.806.

Do, H., Dobrovic, A., 2015. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. Clin. Chem. 61, 64–71. http://dx.doi.org/10.1373/clinchem.2014.223040.

Galante, P.A.F., 2015. A genome-wide landscape of retrocopies in primate. Genome Biol. Evol. 7, 2265–2275. http://dx.doi.org/10.1093/gbe/evv142.

Hedegaard, J., Thorsen, K., Lund, M.K., Hein, A.M.K., Hamilton-Dutoit, S.J., Vang, S., Nordentoft, I., Birkenkamp-Demtröder, K., Kruhøffer, M., Hager, H., Knudsen, B., Andersen, C.L., Sørensen, K.D., Pedersen, J.S., Ørntoft, T.F., Dyrskjøt, L., 2014. 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 9. http://dx.doi.org/10.1371/journal.pone.0098187.

Jian, Y., George, C., Irena, Z., Ioana, C., Steve, R., Madden Thomas, L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13, 134.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, a.D., 2002. The human genome browser at UCSC. Genome Res. 12, 996–1006. http://dx.doi.org/10.1101/gr.229102.

Kim, S., Park, C., Ji, Y., Kim, D.G., Bae, H., van Vrancken, M., Kim, D.-H., Kim, K.-M., 2016. Deamination effects in formalin-fixed, paraffin-embedded tissue samples in the era of precision medicine. J. Mol. Diagn. 1–11. http://dx.doi.org/10.1016/j.jmoldx.2016.09.006.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25, 1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.

McKenna, A.H., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., Depristo, M., 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. http://dx.doi.org/10.1101/gr.107524.110.

Navarro, C.P., Galante, P.A.F., 2013. Databases and ontologies RCPedia: a database of retrocopied genes. Bioinformatics 29, 1235–1237. http://dx.doi.org/10.1093/bioinformatics/btt104.

Oh, E., Choi, Y.-L., Kwon, M.J., Kim, R.N., Kim, Y.J., Song, J.-Y., Jung, K.S., Shin, Y.K., 2015. Comparison of accuracy of whole-exome sequencing with formalin-fixed paraffin-embedded and fresh frozen tissue samples. PLoS One 10, e0144162. http://dx.doi.org/10.1371/journal.pone.0144162.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29, e45. http://dx.doi.org/10.1093/nar/29.9.e45.

Quach, N., Goodman, M.F., Shibata, D., 2004. In vitro mutation artifacts after formalin fixation and error prone translesion synthesis during PCR. BMC Clin. Pathol. 4, 1. http://dx.doi.org/10.1186/1472-6890-4-1.

Sah, S., Chen, L., Houghton, J., Kemppainen, J., Marko, A.C., Zeigler, R., Latham, G.J., 2013. Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies. Genome Med. 5, 77. http://dx.doi.org/10.1186/gm481.

Schweiger, M.R., Kerick, M., Timmermann, B., Albrecht, M.W., Borodina, T., Parkhomchuck, D., Zatloukal, K., Lehrach, H., 2009. Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number-and mutation-analysis. PLoS One 4. http://dx.doi.org/10.1371/journal.pone.0005548.

Spencer, D.H., Sehn, J.K., Abel, H.J., Watson, M.A., Pfeifer, J.D., Duncavage, E.J., 2013. Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. J. Mol. Diagn. 15, 623–633. http://dx.doi.org/10.1016/j.jmoldx.2013.05.004.

Srinivasan, M., Sedmak, D., 2002. Review content and integrity of nucleic acids. Am. J. Pathol. 161, 1961–1971. http://dx.doi.org/10.1016/S0002-9440(10)64472-0.

Van Allen, E.M., Wagle, N., Stojanov, P., Perrin, D.L., Cibulskis, K., Marlow, S., Jane-Valbuena, J., Friedrich, D.C., Kryukov, G., Carter, S.L., McKenna, A., Sivachenko, A., Rosenberg, M., Kiezun, A., Voet, D., Lawrence, M., Lichtenstein, L.T., Gentry, J.G., Huang, F.W., Fostel, J., Farlow, D., Barbie, D., Gandhi, L., Lander, E.S., Gray, S.W., Joffe, S., Janne, P., Garber, J., MacConaill, L., Lindeman, N., Rollins, B., Kantoff, P., Fisher, S. a, Gabriel, S., Getz, G., Garraway, L. a, 2014. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. Nat. Med. 20, 682–688. http://dx.doi.org/10.1038/nm.3559.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinformatics. http://dx.doi.org/10.1002/0471250953.bi1110s43.

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164. http://dx.doi.org/10.1093/nar/gkq603.

Williams, C., Pontén, F., Moberg, C., Söderkvist, P., Uhlén, M., Pontén, J., Sitbon, G., Lundeberg, J., 1999. A high frequency of sequence alterations is due to formalin fixation of archival specimens. Am. J. Pathol. 155, 1467–1471. http://dx.doi.org/10.1016/S0002-9440(10)65461-2.

Ye, X., Zhu, Z.-Z., Zhong, L., Lu, Y., Sun, Y., Yin, X., Yang, Z., ZHU, G., Ji, Q., 2014. High T790M detection rate in TKI-naive NSCLC with EGFR sensitive mutation: truth or artifact? Mol. Cancer Ther. 12, A36. http://dx.doi.org/10.1158/1535-7163.TARG-13-A36.