Methanol-based fixation is superior to buffered formalin for next-generation sequencing of DNA from clinical cancer samples

A. M. Piskorz1,*, D. Ennis2,*, G. Macintyre1,*, T. E. Goranova1, M. Eldridge1, N. Segui-Gracia5, M. Valganon5, A. Hoyle3, C. Orange2, L. Moore4,6, M. Jimenez-Linan4,6, D. Millan3, I. A. McNeish2,#, J. D. Brenton1,4

1Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK
2Institute of Cancer Sciences, University of Glasgow, Glasgow, UK
3Department of Pathology, Queen Elizabeth University Hospital, Glasgow, UK
4Addenbrooke’s Hospital, Cambridge University Hospital NHS Foundation Trust and National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge UK
5Cancer Molecular Diagnostics Laboratory, Department of Oncology, University of Cambridge, Cambridge, UK
6Department of Histopathology, Addenbrooke’s Hospital, Cambridge, UK

#Corresponding author: Professor Iain A. McNeish, Institute of Cancer Sciences, University of Glasgow, Wolfson Wohl Cancer Research Centre, Garscube Estate, Glasgow G61 1QH, UK, iain.mcneish@glasgow.ac.uk, Tel: +44 (0)141 330 3968

*Authors with equal contribution.

Key Message: “High quality tumour DNA is essential for any personalised treatment strategy based on next generation sequencing (NGS). Here we show that methanol fixation is superior to formalin, greater DNA yield, longer fragment size and more accurate copy number calling using shallow whole genome sequencing. We also show provide a new approach to understand fixation artefacts using non-negative matrix factorization.”

© The Author 2015. Published by Oxford University Press on behalf of the European Society for Medical Oncology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Conflict of Interest statement: None of the authors has any conflict of interest to declare.

Abstract

Background. Next-generation sequencing (NGS) of tumour samples is a critical component of personalised cancer treatment, but requires high quality DNA samples. Routine neutral-buffered formalin (NBF) fixation has detrimental effects on nucleic acids, causing low yields, as well as fragmentation and DNA base changes, leading to significant artefacts.

Patients and Methods. We have performed detailed comparison of DNA quality from matched samples isolated from high grade serous ovarian cancers from 16 patients fixed in methanol and NBF. These experiments use tumour fragments as well as mock biopsies to simulate routine practice, ensuring that results are applicable to standard clinical biopsies.

Results. Using matched snap frozen tissue as gold standard comparator, we show that methanol-based fixation has significant benefits over NBF, with greater DNA yield, longer fragment size and more accurate copy number calling using shallow whole genome sequencing. These data also provide a new approach to understand and quantify artefactual effects of fixation using non-negative matrix factorisation to analyse mutational spectra from targeted and whole genome sequencing data.

Conclusion. We strongly recommend the adoption of methanol fixation for sample collection strategies in new clinical trials. This approach is immediately available, is logistically simple and can offer cheaper and more reliable mutation calling than traditional NBF fixation.

Keywords: fixation, NBF, UMFIX, HGSOC, next generation sequencing, copy-number abnormalities, SNVs
Introduction

Although microscopic examination of formalin-fixed paraffin embedded (FFPE) material remains crucial in cancer diagnosis, next-generation sequencing (NGS) of tumour DNA has emerged as a powerful diagnostic tool [1] and is a central component of personalised medicine initiatives. NGS relies heavily on high-quality DNA, and snap-frozen (SF) samples are preferred because formalin fixation induces chemical modifications and degradation of DNA [2, 3].

Comprehensive diagnostic strategies and translational research protocols therefore currently demand two samples, one snap frozen for molecular analysis, and the other FFPE for routine haematoxylin and eosin staining (H&E) and immunohistochemistry (IHC). Processing of SF samples for NGS has several disadvantages, including reduced ability to microdissect tumour material and significantly increased costs [4, 5]. In particular, there are significant barriers to obtaining SF material in large-scale clinical trials, where samples are typically collected from multiple hospitals in different countries. Therefore, alternatives to formalin-based fixation are required to circumvent the need for fresh-frozen sampling.

Methanol-based fixation has emerged as a promising such alternative [5-7] (Supplementary Table 5). Universal molecular fixative (UMFIX) has been shown to be superior for IHC to neutral-buffered formalin (NBF), and gives higher yield and molecular weight of extracted DNA and RNA [5, 6, 8]. In addition, prolonged exposure to methanol fixatives may have fewer deleterious effects on DNA/RNA quantity and quality than NBF [3, 5]. However, potential NGS sequencing artefacts from methanol fixation have not been studied.

Here, we have tested the suitability of DNA extracted after methanol-based fixation for NGS assays compared to DNA from matched NBF and fresh-frozen tissues. We studied high-grade serous ovarian cancer (HGSOC) samples because they have ubiquitous TP53 mutation and TP53 sequences have been extensively studied for fixation artefacts [9, 10]. HGSOC also has marked
genomic rearrangement and copy number abnormalities (CNA), which allow stringent inspection of the effects of DNA fragment length size on CNA profiling.

Patients and methods

Sample acquisition and processing

Three equal fragments were macrodissected from tumour specimens removed from 16 patients, median age 62, with HGSOC undergoing debulking surgery. In addition, mock biopsies of the tumour were taken from 12 cases with a 16G core biopsy gun. All samples were reviewed by at least two pathologists and fixed in 10% NBF (Genta Medical, York, UK), UMFIX (Sakura Finetek, Thatcham, UK) or snap-frozen (liquid nitrogen). Matched normal tissue controls were processed in parallel. Full clinical details are given in Supplementary table 1.

Immunohistochemistry

5µm sections of NBF and UMFIX fixed material were stained for CK7, p53, PAX8, WT1 and CK20 using established clinical protocols in the Department of Pathology, Queen Elizabeth University Hospital, Glasgow, with additional optimization for WT1 staining of UMFIX tissues. Staining and image analysis protocols, as well as all histoscore data, are described in the Supplementary Material.

DNA extraction and quantification

DNA was extracted using QIAmp DNA Micro and AllPrep DNA/RNA Micro Kit for UMFIX/NBF-fixed and SF tumours, respectively. DNA size distribution and quality were assessed by qPCR with Illumina FFPE QC Kit and KAPA hgDNA Quantification and QC Kit, respectively.

Tagged-Amplicon Sequencing (TAm-Seq)

The coding regions of TP53, PTEN, EGFR, PIK3CA, KRAS and BRAF were sequenced by TAm-Seq as described previously [11] on an Illumina MiSeq using PE-125bp protocols. Data analysis is described in Supplementary Material.
Shallow Whole Genome sequencing (sWGS)

WGS libraries were prepared from 100ng DNA using modified TruSeq Nano DNA LT Sample Prep Kit protocol. Library quality and quantity were assessed with DNA-7500 kit on 2100 Bioanalyzer and with Kapa Library Quantification kit according to the original protocols, respectively. Eighteen barcoded libraries were pooled together in equimolar amounts and each pool was sequenced on HiSeq2500 in SE-50bp mode. Analysis methods are described in Supplementary Material.

Mutation signature analysis

Non-negative matrix factorisation was performed to identify mutation signatures [12] in relation to different fixation (Supplementary Material). All non-reference base changes observed across the sequencing data were interrogated from both TAm-Seq and sWGS data.

Results

Figure 1 summarises the study design and the flow of samples through the study. Additional REMARK data are provided in Supplementary Material.

Methanol fixation yields higher yield and size of DNA fragments than buffered formalin

There was no significant difference in tumour cellularity and TP53 allele fraction between UMFIX and NBF samples, thus allowing direct comparison of DNA metrics (Supplementary Figure 1). Quantification of extracted DNA showed similar yields of small (90bp) fragments from UMFIX and SF samples, both of which were significantly higher than from NBF (Figure 2a). As expected, SF samples showed the highest yields of large fragments (129bp, 305bp), but yields from UMFIX samples were still significantly higher than NBF (Figure 2b).

Copy-number calling in methanol-fixed material is superior to formalin

Copy number profiles from shallow whole genome sequencing (sWGS) were compared for correlation and variance of Copy Number Abnormality (CNA) estimation, using SF as gold standard. UMFIX showed superior copy number profiles compared to NBF, with 9/11 biopsies and
10/12 surgical samples showing higher correlation with the matched SF (Figure 2c). UMFIX also had lower noise for segmental copy-number estimation than NBF (Figure 2d).

Single nucleotide sequencing noise from methanol-fixed material is comparable to SF and NBF

We analysed low-level sequence noise using 255,376 observed non-reference bases in the sWGS and TAm-Seq data. All analysed mutations were filtered using dbSNP specifically to exclude germline SNPs. Analysis of the flanking bases around each base change revealed three mutation signatures (Figure 3a): Signature 1 was dominated by non-CpG C>A transversions and C>T transitions; Signature 2 had high rates of T>A, C>A, T>C and C>T transitions, with the latter enriched in the trinucleotide context NCA (where N indicates any base); Signature 3 showed T>C and CpG-related C>T transitions. A breakdown of the contribution of each signature across four categories (base changes common to all samples, changes unique to SF, UMFIX and NBF) showed that the common changes (containing a collection of both true SNVs and typical errors) were dominated by Signature 3, whereas the other categories were a mix of Signatures 1 and 2 (Figure 3b).

Single nucleotide variant calling from methanol-fixed material is comparable to fresh-frozen

SNVs were called using TAm-Seq of 66 samples yielding 546 variants. Manual curation of these variants revealed lower average sensitivity and specificity for NBF compared to SF and UMFIX, albeit not significantly (Figure 3c).

Methanol fixation permits high quality H&E and IHC analyses

Tissue morphology (H&E staining) of UMFIX samples was comparable to NBF fixation. Overall, differences between UMFIX and NBF were not diagnostically significant (Figure 4a). Statistically significant correlation was found between quantitative IHC histoscores in UMFIX and NBF-fixed samples for key HGSOC markers (p53, CK7, PAX8, WT1). CK20 was uniformly negative in all tumour samples, regardless of fixative (data not shown). There was no significant difference in median histoscore between the two sample sets for p53, CK7 and PAX8 (Figure 4b).
Discussion

The most important variables for NGS assays are DNA quality and yield. Formalin fixation can induce severe effects on the structure and integrity of DNA causing C>T, A>G, G>T, G>C and A>T base changes, methylene bridge formation, DNA denaturation and DNA fragmentation [6, 13-15]. After NGS, these chemical modifications result in greater SNV artefacts, higher sequence duplication rates, smaller insert sizes and lower fractions of mappable reads [16, 17].

We evaluated whether methanol-based fixation can reduce these detrimental artefacts when attempting to identify true somatic SNVs and accurate copy-number from clinical material. We show that UMFIX fixation yields longer amplifiable DNA fragments, in agreement with previous reports [3, 5, 8], which improves our ability to call DNA copy-number accurately. We show that SNV calling from UMFIX DNA has similar performance to DNA from SF tissues and that traditional H&E staining and IHC scoring can be performed on UMFIX-embedded samples with minimal optimisation.

These findings are clinically highly important: although attempts have been made to reduce noise induced by formalin fixation (e.g. increasing targeted sequencing coverage or reducing C>T transitions with UDG treatment), these methods only mitigate some sources of noise when calling SNVs, and do not improve the ability to call CNA [18, 19]. CNA detection is more challenging than SNV detection and remains the major clinical need for personalised treatment approaches in HGSOC.

In addition, we have used a state-of-the-art computational approach to perform in-depth exploration of the low-level sequence noise introduced by fixation and sample processing. In an advance over previous approaches, we modelled the trinucleotide context of each base change and de-convolved distinct trinucleotide noise signatures. This computational approach has previously been used to identify signatures in collections of SNVs observed across thousands of tumours, and these signatures used to infer underlying mutational processes [12]. In our data, we
identified three distinct trinucleotide signatures. Signature 3 has high similarity to a previously-identified CpG-age-related cytosine deamination (C>T) signature (Signature 1B [12]), and a recently uncovered sequence error signature [20]. However, signatures 1 and 2 are novel and have no similarity to previously described signatures. In particular, they show high rates of C>T transition but not in CpG dinucleotide contexts. As expected, signature 3 contributed only to the set of base changes common to all samples across a patient. In contrast, signatures 1 and 2 only contributed to the base changes exclusive to SF, UMFIX or NBF samples. This suggests that the sequencing noise represented by these two signatures (C>T not at CpG) is induced through sample processing. The two fixative conditions showed a slightly increased contribution to signature 1 compared to SF, suggesting that fixation may have a specific effect. However, larger studies are required to achieve the power to discern this. This approach to modelling sequence noise provides powerful tools to explore sequencing artefacts and an analytical framework to understand the mechanisms behind their creation. Further studies with high coverage WGS are now underway to refine these data.

There are no data on the effects of long-term methanol fixation on DNA quality or quantity, and this study utilised samples collected no more than six months prior to analysis. With FFPE material, it is possible to isolate DNA from long-term archived samples [21], although factors such as duration of fixation, age of the sample, exposure to heat and light, as well as the concentration, buffering and age of the formalin, can all influence DNA quality and extent of sequence artefact [22]. Careful longitudinal analyses will be required to ascertain whether similar problems emerge in UMFIX samples.

We specifically did not examine RNA in this study. There are several previous publications on the utility of RNA extracted from methanol-fixed specimens in PCR and microarray assays, including from samples stored at room temperature for up to eight weeks [5, 23]. However, we are not aware of any study assessing RNA sequencing or RNA profiling of samples extracted from methanol-fixed tissue – again, future studies will be required to confirm whether RNA extracted from methanol can be reliably used in such assays.
In summary, whilst snap frozen samples remain the gold standard for nucleic acid extraction from tumour material at present, there are significant costs associated with such samples in clinical trials and NGS-based personalised medicine studies. A key advantage of methanol fixation is that it allows easy collection and embedding of tumour material with associated economies for pathological verification and microdissection. Based on our findings of superior DNA quality, we recommend that UMFIX be routinely adopted for collection and storage of clinical cancer specimens for large-scale genomic analysis.

Acknowledgments

This work was supported by Ovarian Cancer Action [BriTROC project grant: IMcN, JDB], Cancer Research UK [grant numbers A15973, A15601, A18072]; NHS Greater Glasgow and Clyde Biorepository; the Universities of Cambridge and Glasgow; National Institute for Health Research Cambridge Biomedical Research Centre; National Cancer Research Network; Cambridge and Glasgow Experimental Cancer Medicine Centres and Hutchison Whampoa Limited. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

We acknowledge the NIHR Cambridge Biomedical Research Centre for their support of this work. The authors would also like to thank Colin Nixon (CRUK Beatson Institute, Glasgow) for assistance with IHC optimisation.

Authors’ contribution statement

Experimental design: IMcN, JDB, AP, DE, GM, TG
Pathological assessment: DM, AH, LM, MJL
Data acquisition: AP, DE, GM, TG, NSG, MV
Quantitative IHC: CO
Data analysis: GM, AP, ME, DE, TG, IMcN, JDB
Manuscript preparation: IMcN, JDB, AP, GM, TG, DE

All authors reviewed the manuscript before submission.

References

5. Vincek V, Nassiri M, Nadji M, Morales AR. A tissue fixative that protects macromolecules (DNA, RNA, and protein) and histomorphology in clinical samples. Lab Invest 2003; 83: 1427-1435.


Figure Legends

Figure 1 - Study design

a. Operative specimens from women undergoing surgery for HGSOC were sampled with a scalpel to acquire three surgical tumour samples and a 16G needle was used to obtain 3 mock biopsies. Matched surgical and biopsy samples from each case, with matched control tissue, were processed in parallel with fixation in NBF or UMFIX, or snap-frozen before downstream analysis.

b. Sample workflow: numbers of patients (P) and samples (S) used for analysis. Bx, biopsy; Tu, surgical tumour fragment; Ctrl, control tissue.

Figure 2 - DNA yield and copy-number calling performance

Boxplots show results of PCR assays for DNA size after extraction from snap frozen (SF), buffered formalin (NBF) and methanol (UMFIX) fixation from matched biopsy and surgical samples from 11 HGSOC patients.

a. Observed ∆ Cq values for DNA yield of 90bp fragments (negative ∆ Cq values are shown for convenience). * p<0.05, ** p< 0.005, ***p<0.0005.

b. Observed Q ratios for 129bp/41bp (top) and 305bp/41bp (bottom) fragments. Vertical brackets indicate Wilcoxon rank sum test for difference in means: * p<0.05, ** p< 0.005, ***p<0.0005.

c. Scatter plots show correlation between median normalised copy-number profiles from shallow WGS of SF compared to NBF or UMFIX biopsy and surgical samples from 12 patients.
Spearman’s rank correlation rho is shown. Green background indicates plots with the highest correlation between UMFIX and NBF for each patient sample (biopsy or surgery).

d. Boxplots show observed variance for each copy-number segment (n=90,312) in 69 samples from 12 patients.

Figure 3: Single nucleotide noise profiles and variant calling performance

a. Bar plots of the three somatic mutation signatures (S1–S3) identified by non-negative matrix factorisation using all non-reference bases observed in sWGS and TAm-SEQ sequencing data in 69 samples from 12 patients (n=255,376). Bar plots are grouped by the observed base change with individual bars showing the proportion observed at different tri-nucleotide sequences.

b. Stacked bar plots show the proportion of the three mutation signatures observed only in SF and NBF or UMFIX fixed samples compared to signatures present in all samples from an individual patient (common).

c. Sensitivity (top) and specificity (bottom) for manually curated SNV calls (n=546) from TAm-SEQ of biopsy and surgical samples from 11 patients. Bars indicate the 95% confidence interval around the indicted mean.

Figure 4: H&E staining and IHC scoring

a. H&E staining of tumour fragments (left) and biopsies (right) from matched tissues fixed in NBF and UMFIX. Bars represent 100 µm.

b. Representative images (left) show IHC staining for p53, PAX8, CK7, WT1 on matched NBF and UMFIX tissues. Quantitative histoscores (middle) of intensity of staining for each IHC marker on tumour fragments (blue points), biopsies (red points) and control tissue (green points) samples. Spearman’s rank correlation rho is shown (p<0.001 for all analyses). Paired data plots (right) show comparison of median histoscores from paired UMF and NBF fixed tissues for each IHC marker. Median scores were only significantly different for WT staining (p = 0.011)
Figure 2

A  Illumina DNA quantification (90bp)

B  KAPA DNA quantification (129 and 305 bp)

C  Segmented copy-number correlation with SF (x-axis)

D  Segmented copy-number noise
Figure 3

A

Single nucleotide (error) signatures

<table>
<thead>
<tr>
<th>Motif</th>
<th>C&gt;A</th>
<th>C&gt;G</th>
<th>C&gt;T</th>
<th>T&gt;A</th>
<th>T&gt;C</th>
<th>T&gt;G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Contribution

B

Signature contribution

C

SNV calling

<table>
<thead>
<tr>
<th>Biopsy</th>
<th>Surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4

A

NBF - Tumour

UMFIX - Tumour

NBF - Biopsy

UMFIX - Biopsy

B

p53

PAX8

CK7

WT1