

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Wong, SQ;Li, J;Salemi, R;Sheppard, KE;Hongdo, D;Tothill, RW;McArthur, GA;Dobrovic, A

Title:

Targeted-capture massively-parallel sequencing enables robust detection of clinically informative mutations from formalin-fixed tumours

Date:

2013-12-13

Citation:

Wong, S. Q., Li, J., Salemi, R., Sheppard, K. E., Hongdo, D., Tothill, R. W., McArthur, G. A. & Dobrovic, A. (2013). Targeted-capture massively-parallel sequencing enables robust detection of clinically informative mutations from formalin-fixed tumours. Scientific Reports, 3 (1), <https://doi.org/10.1038/srep03494>.

Persistent Link:

<https://hdl.handle.net/11343/263102>

License:

[CC BY-NC-SA](#)



OPEN

SUBJECT AREAS:

NEXT-GENERATION
SEQUENCING

DIAGNOSTIC MARKERS

CANCER GENOMICS

MELANOMA

Received

28 June 2013

Accepted

27 November 2013

Published

13 December 2013

Correspondence and
requests for materials
should be addressed to
A.D. (alex.dobrovic@
ludwig.edu.au)

Targeted-capture massively-parallel sequencing enables robust detection of clinically informative mutations from formalin-fixed tumours

Stephen Q. Wong¹, Jason Li³, Renato Salemi¹, Karen E. Sheppard^{4,5,6}, Hongdo Do^{1,7},
Richard W. Tothill^{3,4,5}, Grant A. McArthur^{2,4,5} & Alexander Dobrovic^{1,2,7}

¹Molecular Pathology Research and Development Laboratory, Department of Pathology, Peter MacCallum Cancer Centre, East Melbourne, Victoria, 3002, Australia, ²Department of Pathology, The University of Melbourne, Parkville, Victoria, 3010, Australia, ³Bioinformatics Core, Research Division, Peter MacCallum Cancer Centre, East Melbourne, Victoria 3002, Australia, ⁴Sir Peter MacCallum Department of Oncology, The University of Melbourne, Parkville, Victoria, 3010, Australia, ⁵Division of Cancer Research, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria, 3002, Australia, ⁶Department of Biochemistry and Molecular Biology, The University of Melbourne, Parkville, Victoria, 3010, Australia, ⁷Translational Genomics and Epigenomics Laboratory, Ludwig Institute for Cancer Research, The Olivia Newton-John Cancer and Wellness Centre, Heidelberg, Victoria, 3084, Australia.

Massively parallel sequencing offers the ability to interrogate a tumour biopsy for multiple mutational changes. For clinical samples, methodologies must enable maximal extraction of available sequence information from formalin-fixed and paraffin-embedded (FFPE) material. We assessed the use of targeted capture for mutation detection in FFPE DNA. The capture probes targeted the coding region of all known kinase genes and selected oncogenes and tumour suppressor genes. Seven melanoma cell lines and matching FFPE xenograft DNAs were sequenced. An informatics pipeline was developed to identify variants and contaminating mouse reads. Concordance of 100% was observed between unfixed and formalin-fixed for reported COSMIC variants including *BRAF* V600E. mutations in genes not conventionally screened including *ERBB4*, *ATM*, *STK11* and *CDKN2A* were readily detected. All regions were adequately covered with independent reads regardless of GC content. This study indicates that hybridisation capture is a robust approach for massively parallel sequencing of FFPE samples.

The recent expansion of knowledge about genetic changes in cancer has revealed a set of actionable mutations that can be therapeutically targeted. Notable among these changes are *KIT* mutations in gastrointestinal stromal tumours, *EGFR* mutations in lung cancer, and *BRAF* mutations in melanoma^{1–3}. Many of the druggable targets involve kinase genes for which a large number of effective small molecule inhibitors have been developed⁴. Although some mutations are very frequent in certain tumour types, future studies will invariably target more boutique mutations necessitating technologies that can screen large panels of genes for mutational alterations.

In most diagnostic laboratories, DNA from formalin-fixed and paraffin-embedded (FFPE) tissues is routinely used for the detection of mutational biomarkers that can predict clinical response⁵. For example, in the context of melanoma, a malignancy that frequently displays mutations in the *BRAF* gene, there are currently both commercial and laboratory-developed companion diagnostic tests^{6–8}.

Although current PCR-based methodologies, especially those that take the relatively fragmented nature of formalin-derived DNA into account, have been successful in detecting mutations in patients, a number of challenges exist with these methodologies: (1) the amount of tumour available for analysis is often limited, (2) the quality of DNA extracted can vary greatly, with non-reproducible sequence artifacts frequently observed in FFPE DNA after PCR amplification⁹, (3) much current testing is based on non-multiplexed or low-level multiplexed amplicons and is unsuitable for the increasing number of loci from actionable gene targets that are being discovered, (4) the use of amplicons means that much of the available sequence information in a sample is not



captured, and (5) the necessity for rapid turnaround time, and the need to conserve tumour, means that it is essential that comprehensive testing be done at one site.

The advent and subsequent implementation of massively parallel sequencing (MPS) has attracted wide attention in the clinical arena^{10,11}. The appeal of MPS in a diagnostic setting is due to its ability to rapidly screen mutations in numerous gene targets with a common workflow for multiple tumour streams. Furthermore, the capability to detect low level mutations via deep sequencing will likely form part of many future treatment strategies to monitor disease response to therapy and to detect emerging resistance^{12,13}. However, the challenges mentioned above for FFPE tissues also apply in the MPS context^{10,14}.

There are currently two basic approaches for detecting mutations in a panel of genes by MPS: the targeted amplicon-based approach and the hybridisation-capture approach. Targeted amplicon-based approaches using benchtop sequencers have been adopted by diagnostic laboratories to profile more limited sets of clinically important genetic regions based on existing diagnostic workflows, shorter turnaround times, and the ability to automate¹⁵.

Hybridisation enrichment capture approaches that have been used for exome sequencing studies and are able to screen multi-exon genes with no defined hotspot^{16,17}, are also applicable to a targeted sequencing approach. However, relatively few published studies have assessed the ability to use DNA extracted from formalin-fixed material for reliable mutation detection via a capture-based MPS approach. It also remains unclear if a capture-based approach is suited for diagnostic purposes and whether sequence artifacts present in amplicon-based approaches⁸ are also apparent in capture-based methods.

This study aimed to assess targeted capture of a defined set of genes to identify mutations in FFPE samples. To address this issue, we compared formalin-fixed biopsies established from a mouse xenograft tumour model with the cell lines from which the xenografts were derived. The ability to accurately detect multiple mutations across a large array of actionable targets in the formalin-fixed samples was assessed. In particular, we investigated if the depth of sequencing was at a sufficient level to accurately identify mutations in formalin-fixed melanoma tissue with particular attention being paid to the detection of sequence artifacts.

Results

Experimental design. We chose to use formalin-fixed xenograft tumours derived from cell lines instead of clinical formalin-fixed samples as it not only provided adequate amounts of DNA for targeted capture but also comparing a xenograft to a parental cell line minimises heterogeneity issues because of the well known caveat that fresh-frozen and FFPE biopsies from the same tumour do not necessarily have the same mutational status¹⁸. In addition, xenografts are also intrinsically of interest because of the increasing use of xenografts to study tumour properties in personalised medicine^{19,20}.

Seven melanoma cell lines, representing the unfixed components, were used for this study. Five of the cell lines had mutations described on the COSMIC cancer mutation database (MeWo, SK-MEL28, LOXIMV1, A2058 and HT144) and two lines were of unknown mutational status (WM266-4, CO55-M1). Xenografts from these cell lines had previously been established in mice. Tumours were harvested, fixed in formalin, and paraffin embedded, using a protocol that was identical to that used for clinical tumour samples.

Assessing DNA fragmentation. Since DNA derived from formalin-fixed tissue tends to be extensively fragmented, we used a quality control assay to identify any samples that might have been unsuitable for sequencing (Figure 1). This PCR-based assay assesses the size range of amplifiable fragments in a DNA sample to ensure that there are sufficient DNA templates of an acceptable length for pre- and post-hybridisation PCR steps²¹.

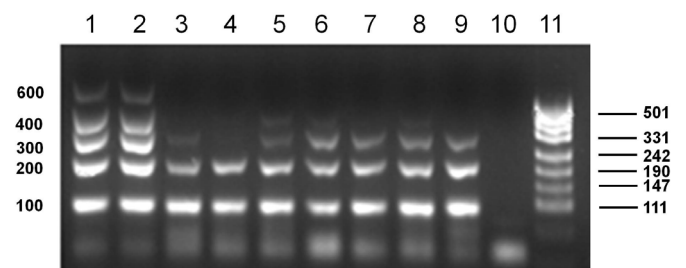


Figure 1 | Assessment of FFPE DNA integrity by a multiplex PCR assay. Sizes of the individual amplicons of the integrity assay are indicated on the left, sizes of the molecular weight marker bands are given on the right. Sizes are given in base pairs. Lanes 1 and 2 were loaded with PCR reactions obtained from genomic DNA extracted from two non-FFPE controls and served as positive controls, showing all five PCR amplification products at the expected size. Lanes 3 to 9 were loaded with PCR reactions obtained from DNA from macro-dissected FFPE xenograft biopsies derived from melanoma cell lines (Lane 3. CO55-M1, Lane 4. SK-MEL28, Lane 5. MeWo, Lane 6. A2058, Lane 7. WM266-4, Lane 8. HT144 and Lane 9. LOXIMV1). Lane 10 was loaded with the no template control (NTC). Lane 11 was loaded with the pUC19/HpaII DNA molecular weight marker. The DNAs used for lanes 3–9 all showed amplification of fragments up to 400 bp except lane 4 which showed no amplification of fragments above 200 bp.

Using the DNA isolated from the formalin-fixed xenografts, the 100 and 200 bp bands were clearly visible and most samples showed some signal at 300 bp indicating relatively good quality FFPE DNA (Figure 1). Since DNA is required to be sheared to a range of 150–200 bp to be successfully used for subsequent PCR and sequencing, this level of fragmentation meets the minimal length that is acceptable for shearing and subsequent capture enrichment.

Targeted capture results. Targeted capture was performed using the Agilent SureSelectTM Kinome Kit which comprises 26,231 120 bp biotinylated RNA oligos as probes covering a total of 10,282 exons. These probes target 566 known kinases and 46 other cancer related genes (Supplementary Table 1). The kinome kit was chosen because mutations that affect kinases have been implicated in cancer and have often been targeted therapeutically²², and because of the inclusion of many of the important non-kinase cancer related genes such as *KRAS*, *NRAS*, *TP53* and *CDKN2A*.

Sequencing was carried out using paired-end 100 bp reads with performance metrics calculated after alignment and after removal of duplicate reads. Performance metrics indicated that there was a very similar number of mean reads, mapped reads and “on target” reads for the unfixed and formalin-fixed samples (Supplementary Table 2). Overall targeting efficiency was assessed by measuring base coverage over all targeted bases for both unfixed and formalin-fixed samples (Figure 2). On average, unfixed samples had 99.86% of the targeted bases covered at least once and 99.2% covered more than 20-fold (20×). With formalin-fixed samples, 99.86% of the targeted bases were covered at least once and 98.7% of bases covered more than 20×. With a higher base coverage, targeting efficiencies for both unfixed and formalin-fixed samples decreased at a similar rate with a marginally lower targeting efficiency for formalin-fixed samples compared to unfixed samples.

In terms of actual coverage, there was a mean 800× (mean range 454–1811, $n = 7$) and 550× (mean range 218–1154, $n = 7$) target coverage based on independent reads for unfixed DNA and formalin-fixed DNA respectively (Supplementary Table 2). Both the overall lower targeting efficiency and mean coverage in formalin-fixed samples may reflect the shorter fragment sizes available for capture but is more than acceptable to call mutations at a good sensitivity.

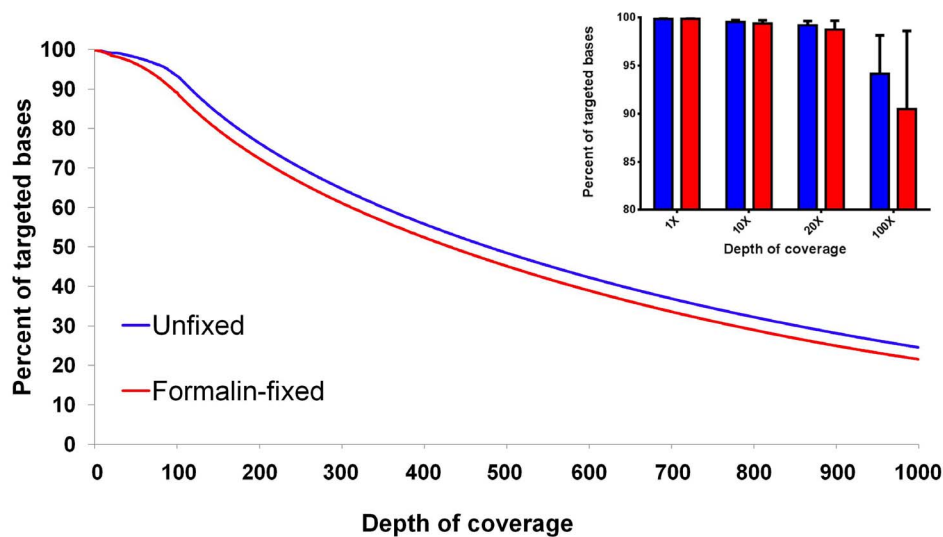


Figure 2 | Efficiency trends for unfixed vs. formalin fixed samples. Efficiency is visualised as the mean percent of total targeted bases covered at particular depths. The graph represents the average of all seven unfixed (blue line) and formalin-fixed (red line) samples. Inset: Detailed mean percentage of total targeted bases covered at lower read depths (1×, 10×, 20× and 100×). All values are the mean ± SD (n = 7).

Additionally, three times more duplicate reads (reads of identical length and sequence which have been amplified from the same template) were observed in formalin-fixed samples compared to unfixed samples indicating, not unexpectedly, that the amount of amplifiable template from formalin-fixed DNA was more limited. Nonetheless, on average more than 90% of targets exhibited more than 100× coverage in both unfixed and formalin-fixed DNA whereas less than 1% had no coverage.

Assessment of contaminating mouse sequences. A cross comparison of the percentage of mappable reads across formalin-fixed samples showed a marked decrease of mappable reads in the formalin-fixed sample from the LOXIMV1 cell line. We hypothesised that this low percentage of mappable reads was due to mouse contamination within the xenograft-derived tumour. Through BLAT alignment of each variant, there was a disproportionate number (44%) of LOXIMV1 formalin-fixed variant reads which mapped to the mouse genome when compared to its unfixed counterpart of 8% (Supplementary Figure 1). The formalin-fixed SK-MEL28 and CO55-M1 samples also showed a relatively high level of variant reads homologous to the mouse (24% and 37% respectively). All other samples showed no significant difference between unfixed (cell-line) and formalin-fixed (xenograft) counterparts. For the LOXIMV1, SK-MEL28 and CO55-M1 samples, variant reads which showed a significant alignment to mouse with a mouse BLAT score of >100 were flagged as potential mouse variants. Our interpretation of a substantial degree of contaminating mouse tissue in the LOXIMV1, SK-MEL28 and CO55-M1 fixed samples was supported by a review of the corresponding H&E slides by an experienced pathologist and confirmed that the other remaining samples had less than 1% mouse tissue.

SNV and INDEL comparison between unfixed and formalin-fixed samples. To ensure optimal analysis of tumour specific variants, only the 4 cell lines which showed no significant mouse contamination were used. The ratio of single nucleotide variants (SNVs) and insertion/deletion events (indels) for the 4 matched unfixed and formalin-fixed sample pairs was compared using different cutoffs in the variant coverage (30×, 50×, 100×, 200×) and allele frequency (>1% and >20%) as shown in Figures 3A and 3B respectively. At 30× coverage and at >1% allele frequency,

samples had an approximately 83% concordance for the called SNVs in both unfixed and formalin-fixed DNA samples. In contrast, indel variants in these four samples had a much lower concordance with an average of 62% of total indels identical between unfixed and formalin-fixed DNA. This difference in concordance rates reflects the algorithms used in the GATK variant caller which requires more coverage to accurately call indels compared to SNVs. Indels can be easily misaligned to other sequences in the genome with multiple mismatches to the reference rather than with a gap.

At this coverage cutoff of 30× and at >1% allele frequency, more unique variants in the unfixed samples were identified than unique variants in the formalin-fixed samples for A2058, HT144 and WM266-4. This observation could possibly be due to low allele frequency variants being called more robustly in unfixed samples than those in formalin-fixed samples. In line with this, an allele frequency cutoff of >20% demonstrated a much higher level of concordance of common variants between unfixed and formalin-fixed counterparts in these three sample pairs (mean 88% concordance for SNVs and 82% for indels at >30× coverage). While the number of common variants decreased at this allele frequency cutoff, this analysis indicates that at least for high frequency variants there is high degree of concordance between formalin-fixed and unfixed datasets.

The pattern of higher unique unfixed variants was not apparent when examining the MeWo pair. This can be explained by a higher mean coverage in the formalin-fixed sample compared to matching unfixed counterpart. Supporting this observation, only 26 variants were identified in unfixed MeWo sample compared to 54 in the formalin-fixed MeWo sample below the 5% allele frequency. Concordance of common variants >1% and >20% allele frequency were also very similar. The MeWo formalin-fixed sample was one of the less fragmented samples based on the FFPE integrity assay results (Figure 1) and appears to have been sequenced in comparable if not marginally better level than its unfixed counterpart.

As expected, at increasing coverage cutoffs of 50×, 100× and 200×, the number of common SNVs and indels between formalin-fixed and unfixed samples decreased progressively for all 4 sample pairs. For A2058, HT144 and MeWo, both unique unfixed and formalin-fixed variants decreased with increasing coverage cutoffs. However, this trend was surprisingly absent in the WM266-4 sample where there was a noticeable increase in variants called uniquely for the unfixed sample with increasing coverage cutoffs. However, close

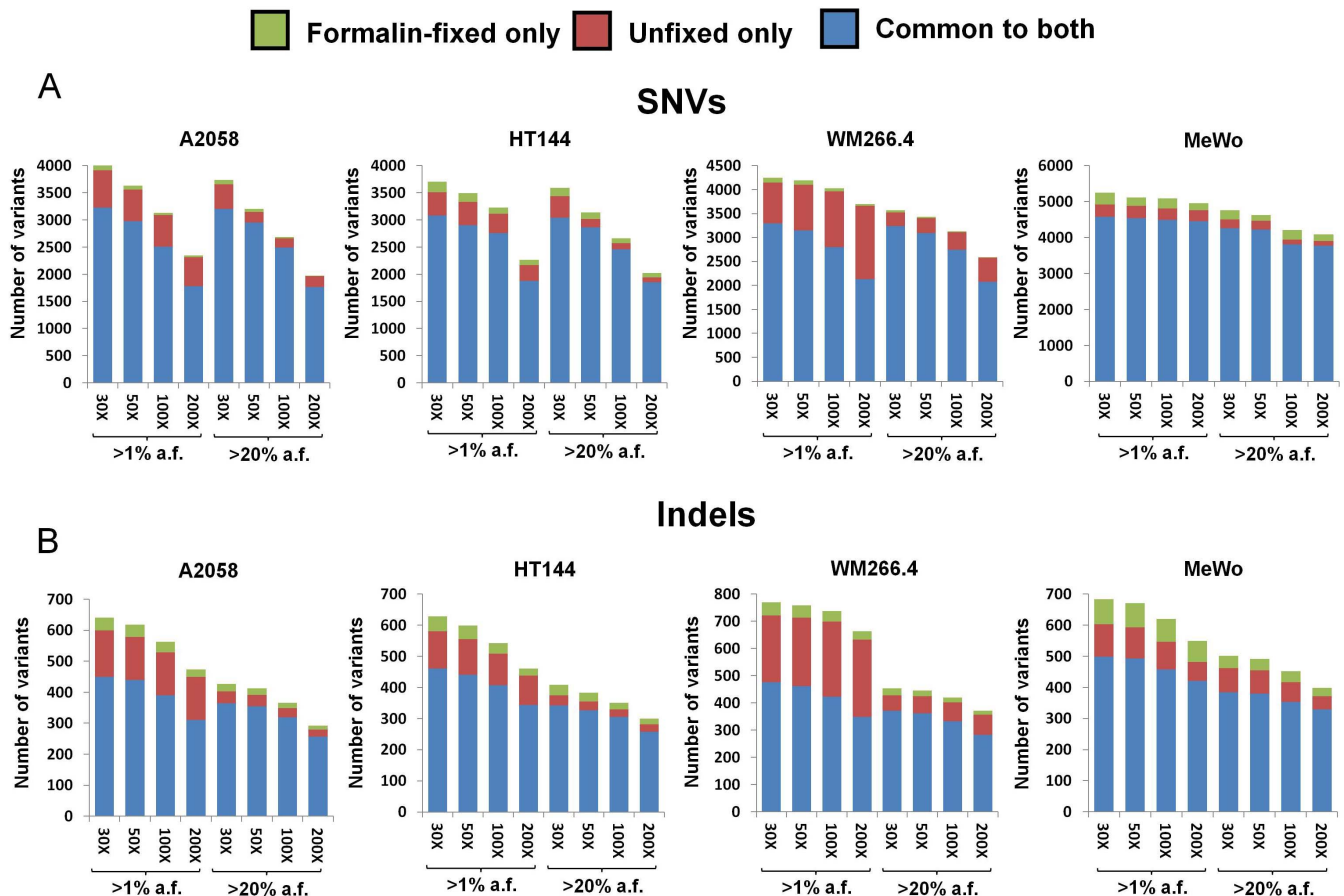


Figure 3 | SNV and INDEL comparison between unfixed and formalin-fixed pairs. For all graphs, variants unique to formalin-fixed samples are shown in green, variants unique to the unfixed formalin cell line samples are shown in red, and variants common to both unfixed formalin and formalin-fixed samples are shown in blue. Variant coverage cutoffs of 30 $\times$, 50 $\times$, 100 $\times$ and 200 $\times$ were applied at >1% and at >20% allele frequency (a.f.). (A) Filtered SNVs for four melanoma unfixed formalin/formalin-fixed pairs. (B) Filtered indels for four melanoma unfixed formalin/formalin-fixed pairs.

examination of the mean coverage showed a large difference between the unfixed sample (1811 $\times$) compared with the formalin-fixed sample (470 $\times$) and was considerably higher than other unfixed samples (mean coverage for other unfixed samples = 633 $\times$). This indicated that many low frequency variants remained in this unfixed sample despite high coverage cutoffs and may reflect increasing mutational heterogeneity in this cell line in comparison to the other cell lines.

Detection of COSMIC mutations in unfixed and formalin-fixed DNA. Five out of the seven cell lines appear on the COSMIC database. All have mutations in at least one of the genes listed in the COSMIC database (Table 1)²³. Importantly, concordance between the unfixed cell line and formalin-fixed samples in the somatic variants listed on COSMIC database was 100% (Variant reads for both fixed and unfixed samples shown in Supplementary Table 3). This is strong confirmation that formalin-fixed samples can be used reliably for the detection of mutations using the capture approach.

To assess how accurate the allele frequencies between formalin-fixed samples compared to unfixed counterparts, the coverage for both the normal and mutant alleles for the *BRAF* c.1799T > A mutation were calculated for three representative unfixed/formalin-fixed pairs (Figure 4A). A slight decrease in overall number of independent reads for formalin-fixed samples for the mutation was observed with an average of 14.4% decrease in the number reads in formalin-fixed vs. unfixed samples ($p = 0.17$, t-test). Examination of the quality of reads across these three formalin-fixed/unfixed pairs showed an average 17.4% decrease in quality scores in formalin-fixed

over unfixed pairs ($p = 0.36$, t-test). Importantly, the proportion of variant reads was very similar between unfixed and formalin-fixed samples. These results clearly indicate no marked difference in the ability to detect the *BRAF* V600E variant using formalin-fixed samples.

Importantly, other clinically relevant mutations described on COSMIC for these cell lines in the *TP53*, *PTEN*, *EGFR*, *NF1* and *CDKN2A* genes were also readily identified in the formalin-fixed samples (Table 1). Allele frequencies for these variants were very similar between unfixed and formalin-fixed pairs even at low allele frequencies such as the *PTEN* p.174–186del mutation in A2058 which was detected at 15% in unfixed and 9% in its formalin-fixed counterpart (Table 1 and Supplementary Table 3).

Detection of other clinically relevant mutations in unfixed and formalin-fixed DNA. Current diagnostic tests that screen single or small number of exons such as Sanger sequencing, HRM analysis or pyrosequencing exist for the screening of mutations in melanoma samples. These include *BRAF*, *NRAS*, *KIT* and *TP53*. However, many other large sized genes such as *PTEN*, *ERBB4* and *ATM* which have prognostic value to melanoma patients can only effectively be screened using these capture-based approaches. Hence, to test the clinical utility of capture-based approaches, we examined if other clinically relevant mutations were retained and could be reliably detected after the formalin-fixation process.

Based on the list of cancer genes which form part of the kinome panel, we examined those genes which are commonly mutated in melanoma. This has identified mutations in both unfixed and form-



Table 1 | Concordance of reported COSMIC mutations for five melanomas cell lines (Unfixed vs. Formalin-fixed)

Name	Mutation	Unfixed sample		Formalin-fixed sample	
		Mutation identified	Allele frequency	Mutation identified	Allele frequency
LOXIMV1 A2058	<i>BRAF</i> c.1799T > A, V600E	Yes	62%	Yes	58%
	<i>BRAF</i> c.1799T > A, V600E	Yes	33%	Yes	29%
	<i>PTEN</i> c.335T > C, L112Q	Yes	49%	Yes	45%
	<i>PTEN</i> c.1879-1913del, p.174-186del	Yes	15%	Yes	8%
	<i>TP53</i> c.1010 C > A, V274F	Yes	100%	Yes	100%
HT144 SK-MEL28	<i>BRAF</i> c.1799T > A, V600E	Yes	100%	Yes	99%
	<i>BRAF</i> c.1799T > A, V600E	Yes	100%	Yes	94%
	<i>EGFR</i> c.2257C > T, p.P753S	Yes	98%	Yes	99%
	<i>TP53</i> c.624A > C, L145R	Yes	97%	Yes	98%
	<i>CDKN2A</i> 509G > A, R80X	Yes	100%	Yes	100%
MeWo	<i>TP53</i> c.1139 G > A, Q317X	Yes	50%	Yes	53%
	<i>TP53</i> c.962 C > T, E258K	Yes	49%	Yes	52%
	<i>NF1</i> c.4006 G > A, Q1136X	Yes	100%	Yes	99%

alin-fixed samples for genes such as *PTEN*, *STK11*, *ERBB4*, *NF1* and *ATM* (Table 2.). The *BRAF* V600E mutation was also identified in both unfixed and formalin-fixed components for the two samples (WM266-4, CO55-M1) which were not in the COSMIC database.

Kinases have been a frequent target for targeted therapies and therefore identification of mutation in these genes will be critical for future treatment regimens²⁴. A number of exome studies have identified recurrent mutations in the mitogen-activated protein kinase genes *MAP3K9*, *MAP3K5* and *MAP2K1*^{25,26}. Mutations in

these genes were identified in three of our unfixed/formalin-fixed pairs (A2058, HT144 and MeWo) (Table 2). Once again, these mutations could be readily identified by examination of the formalin-fixed samples alone. Therefore, these results show that it is possible to robustly detect clinically relevant mutations from formalin-fixed material, even from samples with high mutation loads.

Relative coverage in large multi-exon genes and GC-rich genes in unfixed and formalin-fixed DNA. Whereas *BRAF* represents a gene

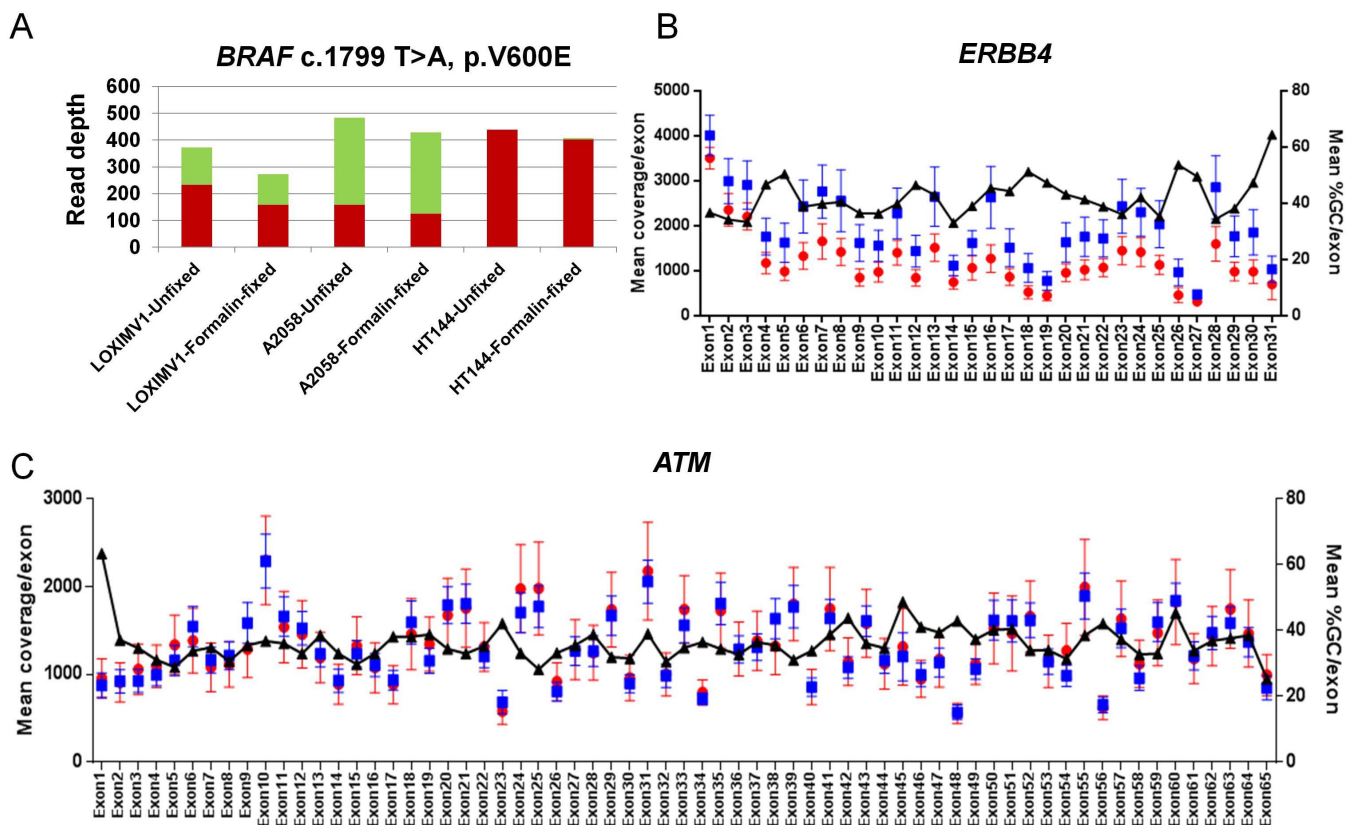


Figure 4 | Coverage of the *BRAF* V600E mutation and the *ATM* and *ERBB4* genes. (A) The coverage of the c.1799T > A, p. V600E mutation in exon 15 of the *BRAF* gene across three representative unfixed/formalin-fixed pairs. The counts for the wildtype T base (green) and the variant A base (red) are shown for each sample. (B) The average coverage for each exon in the *ERBB4* gene in unfixed (blue squares) versus formalin-fixed (red circles) samples across all seven melanoma cell lines. All values are the mean $\pm$ SD ($n = 7$). The mean percentage of GC content in each exon for these genes are also shown (black triangles). (C) The average coverage for each exon in the *ATM* gene in unfixed (blue squares) versus formalin-fixed (red circles) samples across all seven melanoma cell lines. All values are the mean $\pm$ SD ($n = 7$). The mean percentage of GC content in each exon for these genes are also shown (black triangles).



Table 2 | List of other clinically relevant mutations detected in both unfixed and formalin-fixed samples

Gene	Variants identified (samples)
<i>BRAF</i>	c.1799T > A, V600E (WM266-4, CO55-M1)
<i>PTEN</i>	c.1856A > G, T167A (SK-MEL28)
<i>STK11</i>	c.1260T > G, Y49D (A2058)
<i>ERBB4</i>	c.882A > G, P294P (MeWo) c.1657G > A, S449F (MeWo) c.2609C > T, M766I (MeWo)
<i>ATM</i>	c.8919 G > A, W2845X (HT144) c.4643 C > T, L1420F (MeWo)
<i>NF1</i>	c.3520 C > T, Q1174X (LOXIMV1)
<i>MAP2K1</i>	c.901 C > T, P124S (A2058)
<i>MAP3K5</i>	c.3951T > G, p.D3590A (HT144) c.4169 G > A, p.L1270F (MeWo)
<i>MAP3K9</i>	c.1127T > C, p.E376G (MeWo) c.1162 G > A, p.P388S (MeWo) c.1678G > A, p.R560X (MeWo)

with a defined hotspot region and relatively few variant changes, other genes that are associated with melanoma have fewer or no defined hotspots^{27–29}. It is thus important to ensure that there is adequate coverage over the full length of the genes. We examined the average number of reads for the multi-exon *ERBB4* and *ATM* genes across all 7 melanoma cell lines. For formalin-fixed samples, there was an average of 1460 reads (312–3029) for *ERBB4* (Figure 4B) and an average of 1276 reads (315–2234) for *ATM* (Figure 4C) respectively. This compares well to the reads from unfixed samples which displayed an average of 1821 reads (338–5417) for *ERBB4* and an average of 1331 reads (331–3029) for *ATM* respectively. Assuming a minimum of 20 independent reads are required to call a variant confidently, this level of coverage in the formalin-fixed sample is more than sufficient to detect mutations at a level of greater than 5%, and in most cases greater than 2% for these genes.

In general, the GC content for both *ERBB4* and *ATM* genes is relatively low with a mean percentage of GC/exon of 36.2 and 42.2 respectively. In fact, only one exon from each of these genes had a mean percentage of GC > 60%. The *CDKN2A* and *STK11* genes on the other hand represent genes with relatively high GC content that have proven to be problematic in amplicon-based approaches. Despite exon 1B, 2 and 3 of *CDKN2A* all having >60% GC, the kinome capture kit was able to achieve independent reads in the hundreds across the three exons as shown in Figure 5A (mean coverage of 347 for unfixed samples versus 333 for formalin fixed samples). Another GC-rich gene, *STK11*, also had adequate coverage even when 9 out of its 11 exons were >60% GC as shown in Figure 5B (mean coverage of 617 for unfixed samples versus 706 for formalin fixed samples). Whilst exon 1 of *STK11* displayed a marked drop in the coverage compared to other exons because of a 80% GC content, the average coverage of exon 1 was still adequate for mutation detection with an average 170× coverage for unfixed vs. 200× coverage for formalin-fixed samples. Mutations were detected in both these genes for the MeWo and A2058 formalin fixed samples (Supplementary Table 3). This demonstrates that hybridisation capture is a suitable platform to achieve sufficient coverage in GC-rich regions.

FFPE specific sequence artifacts. FFPE DNA restricted changes could be explained because of sequence artifacts derived from formalin fixation. Our lab has previously shown that the predominant sequence artifacts in formalin-fixed DNA are C:G > T:A changes which are caused by deamination of cytosine bases to uracil^{9,14}. Thus, if the DNA polymerase that amplifies templates derived from FFPE DNA is capable of reading across uracil,

artificial C:G > T:A changes will be observed. This is particularly important in melanoma where C:G > T:A changes are often true changes.

Previously, we developed an informatics pipeline which is able to specifically identify the degree of sequence artifacts¹⁴. Using an amplicon-based MPS approach, we found that formalin fixed samples often displayed sequence artifacts in the range of 1–10% allele frequency¹⁴. Part of the analysis includes the removal of known cell line variants including low frequency variants. This was to ensure that FFPE specific sequence artifacts could only be due to the formalin fixation process. To examine if this was similar in our samples, this pipeline was applied to the four samples which showed few mouse contaminating sequences. To fully examine if sequence artifacts had also been amplified, the aligned kinome capture data included duplicate reads to account for sequence artifacts that can be informatically removed.

Figure 6 illustrates that across these samples we did not observe any marked difference in the profile of the nucleotide substitutions when comparing the unfixed versus formalin-fixed samples. Since C:G > T:A substitutions represented the major artifact from the formalin fixation process when amplicons were analysed¹⁴, it illustrates that artifactual C:G > T:A substitutions are less frequent using a capture-based process compared to those observed in conventional amplicon-based procedures¹⁴. This reflects the increased amounts of templates available for capture-based approaches and the ability to assess independent reads.

Discussion

The ability to use DNA from formalin-fixed material is mandatory for any methodology that seeks to be diagnostically applicable. In this communication, we provide evidence that targeted capture is a robust MPS approach when using FFPE DNA. We consider that targeted capture is preferable to the amplicon-based approach that is commonly used in diagnostics when using FFPE DNA as it can utilise many more of the templates present in a FFPE DNA sample.

FFPE tissues will remain the standard material for clinical analysis in the foreseeable future. Not only are they preferred by pathologists for the examination of tumour morphology, the preservation of morphology enables the isolation and analysis of relatively enriched tumour material space as well as the ability to evaluate intra-tumoural heterogeneity. Ease and cost of storage is also a huge advantage compared to frozen tissues.

Hybridisation capture has been used for whole exome studies in a large number of cancer types^{16,17,25,30,31}. Hybridisation capture first involves the preparation of a ‘shotgun’ library of fragmented DNA followed by end repair, A-tailing and adapter ligation. To then enrich specific DNA sequences, a capture using RNA or DNA ‘baits’ is used on the library followed by limited PCR amplification to yield enough product for sequencing³².

We chose the commercially available SureSelect™ kinome kit as proof-of-principle as it not only captured exons from a large set of potentially actionable kinase genes but also from a set of commonly mutated cancer genes. Working with a targeted panel of genes allows greater cost efficiency compared to exome sequencing as the use of fewer gene targets reduces the cost of sequencing which then also simplify the bioinformatics analysis. The higher mean coverage per sample that is economically possible compared to exome sequencing also increases the ability to detect low frequency mutations that are harder to detect in cancer specimens because of variations of ploidy and normal tissue contamination.

Melanomas have a much higher somatic mutation load compared to most solid tumours^{33,34}, making our samples a powerful set to examine variant concordance. We first focused on the detection of the most common mutation in melanoma, the *BRAF* V600E mutation⁸. Accurate determination of the *BRAF* mutation status is crucial in deciding the use of Braf inhibitors in individual patients. Any MPS

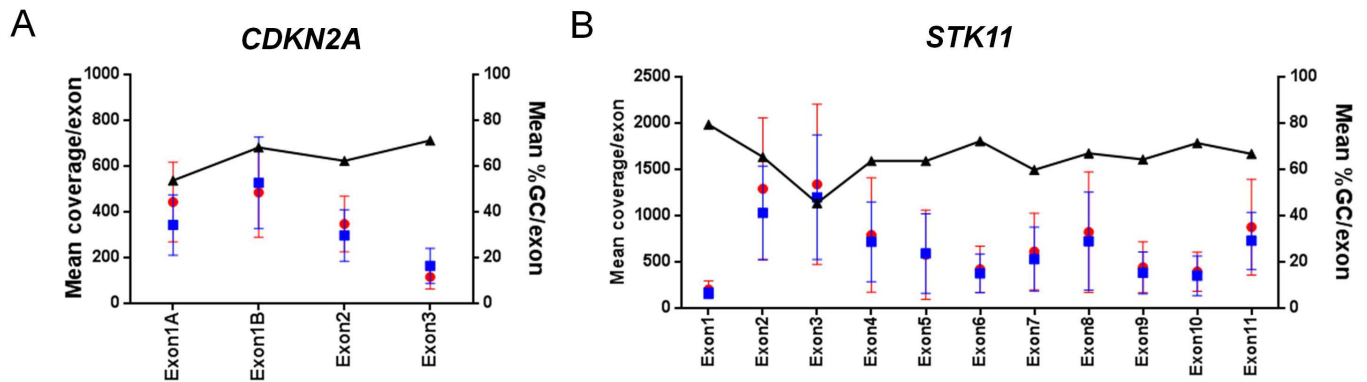


Figure 5 | Coverage of the GC rich *CDKN2A* and *STK11* genes. (A) The average coverage for each exon in the *CDKN2A* gene in unfixed (blue squares) versus formalin-fixed (red circles) samples across all seven melanoma cell lines. All values are the mean $\pm$ SD ($n = 7$). The mean percentage of GC content in each exon for these genes are also shown (black triangles). (B) The average coverage for each exon in the *STK11* gene in unfixed (blue squares) versus formalin-fixed (red circles) samples across all seven melanoma cell lines. All values are the mean $\pm$ SD ($n = 7$). The mean percentage of GC content in each exon for these genes are also shown (black triangles).

protocol that can be diagnostically applied to formalin-fixed tissues must detect such clinically relevant mutations particularly in terms of coverage and sensitivity.

Importantly, our findings did show reliable detection of the *BRAF* V600E mutation in both our unfixed and our formalin-fixed samples, indicating the robustness of this approach for the detection of specified mutations. We demonstrated 100% concordance between both the unfixed and formalin-fixed samples for these COSMIC mutations (Table 1). Significantly, our study also demonstrated robust identification of other mutations, especially those that occur throughout multi-exon genes including *ATM*, *ERBB4* and *PTEN*. Prior to massively parallel sequencing, these genes were rarely tested as the absence of major hotspots required considerable expense in reagents and time, and consumed a large amount of sample DNA.

Recently, exome sequencing studies on melanoma cell lines^{25,26,35–38} have identified multiple mutations in kinase genes that have not previously been linked to melanoma including *MAP3K5*, *MAP3K9*, *MAP2K1*, *MAP2K2*. In this study, we detected the activating *MAP2K1* P124S mutation in A2058 which is known to result in erk1/2 phosphorylation in the absence of serum²⁵ and the inactivating *MAP3K9* R560X mutation detected in MeWo.

Of note, some mutations could be missed depending on the coverage stringencies used. For example, the *PTEN* 174–186del and *STK11* Y49D mutations that we observed would have been filtered in the formalin-fixed samples if a high variant cutoff of 100 $\times$ was used. Whereas, a higher cutoff might be useful to determine mutations in FFPE samples, it could also lead to true low frequency variants being lost in low coverage samples.

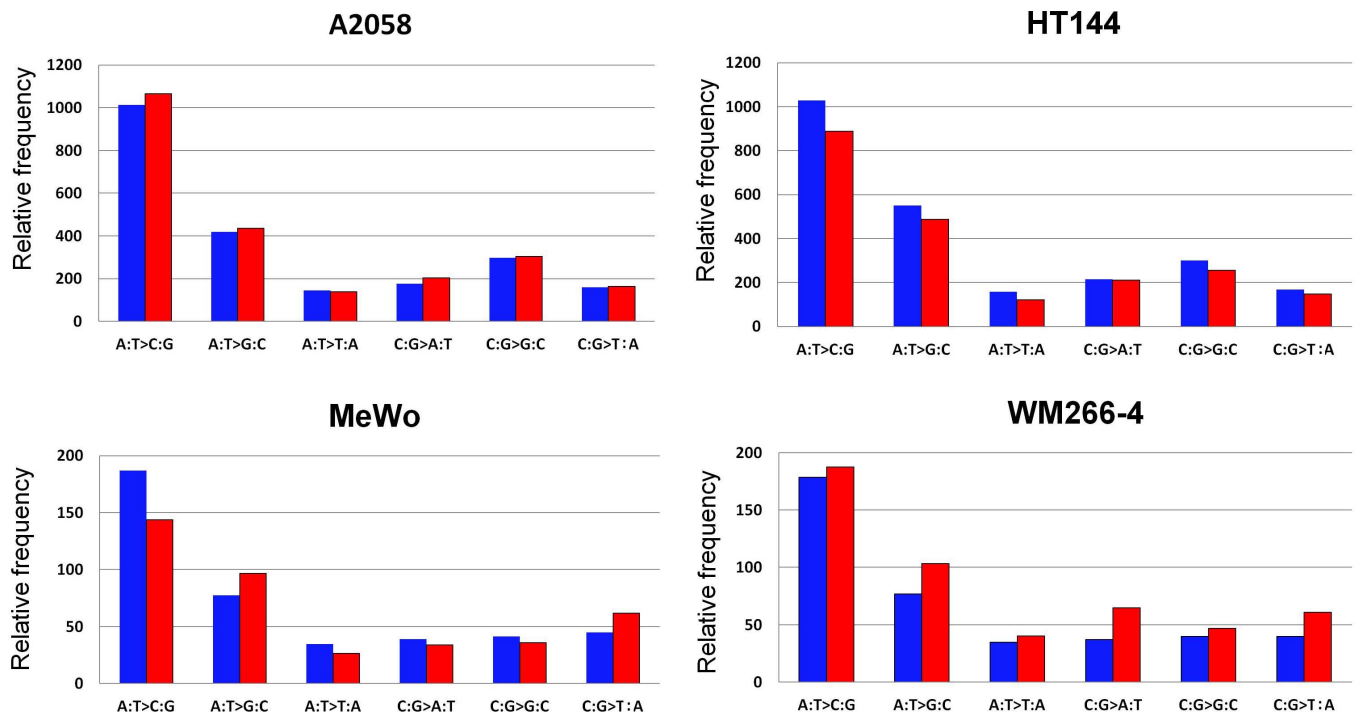


Figure 6 | Nucleotide substitution rates for four unfixed and formalin fixed pairs. For all graphs, the relative frequency of each nucleotide change is shown in blue for unfixed samples and shown in red for formalin-fixed samples. Single nucleotide changes were filtered using lower and upper thresholds of 1% and 10% to minimise contamination by sequencing errors and true changes. Known cell line variants including low frequency variants are excluded. To calculate the relative prevalence of each type of single nucleotide change, the sum of correct base calls (AA, CC, GG, and TT) was used as the denominator for data normalisation, and the normalised value was then multiplied by 10^6 .



Many mutations occur in GC-rich genes such as *CDKN2A* or *STK11*. However, the amplification of GC-rich regions is often problematic when using an amplicon-based sequencing method³⁹. This is much less of an issue in hybridisation capture as our results demonstrate an adequate and consistent coverage of both the *CDKN2A* and *STK11* genes from both unfixed and formalin-fixed material. While we did not examine regions with extremely high GC-content, i.e. >80%, reliable detection of mutations in both *CDKN2A* and *STK11* for unfixed and formalin-fixed samples indicate that this technology is suitable for targeting areas which have relatively high GC content.

Other studies have tested the feasibility of obtaining reliable variant information from formalin-fixed material using limited panels of genes. Wagle and colleagues previously reported a targeted capture approach on FFPE DNA⁴⁰. Their study showed that it was possible to screen for actionable mutations that could be validated. Similarly, Adams and colleagues sequenced 5 human colon cancer samples using a capture approach for 140 driver genes and found high concordance with results obtained from Sanger sequencing⁴¹. However, neither study made a direct comparison of variant concordance between formalin fixed samples compared to a matching unfixed sample. Kerick and colleagues reported reliable mutation detection after formalin processing in five prostate samples based on minimal coverage (5×–80×)⁴². However, the detection of clinically relevant mutations in formalin-fixed samples compared to fresh frozen counterparts was not examined as their report was focused on the detection of SNPs⁴².

It is now well recognised that genetic intratumoural heterogeneity can exist between biopsies from even the same site⁴³. Therefore, while actual clinical samples were not used in this current study, the design of our experiments lessened issues of tumour heterogeneity that could confound the accurate comparison of variants in unfixed and formalin-fixed samples. However, it is possible that clonal selection may make xenografts less genetically complex than parental cell lines. Hence, while low allele frequency variants were generally observed at higher rates in unfixed samples because of increased coverage, some of these variants may also represent low level heterogeneity from the cell line. Nevertheless, we consider the formalin-fixed samples to be representative of actual clinical tumour biopsies with varying degrees of fragmentation observed despite similar times of fixation.

In many tumour samples, the amount of DNA that can be extracted is limited. The ability to take live cells from a tumour and transplant them into a mouse model gives the opportunity to expand tumour tissue of the same genetic makeup. The use of xenografts in this paper also reflects the likelihood that patient-derived xenografts as a mode of identifying optimal treatment for individual tumours are becoming more common^{19,44}. Our results indicate inherent issues when using xenografted DNA samples for global genomic studies. Xenografted samples used for MPS need to be examined for contaminating mouse tissue to provide accurate results as any residue mouse variant can potential be misinterpreted as a somatic mutation. The need to assess morphology, determine tumour purity and ability to evaluate intra-tumoral heterogeneity also means that formalin fixation will be of use in the assessment of xenografts.

Amplicon-based MPS methods have been useful in detecting mutational hotspots in a limited number of genes. While no amplicon-based MPS comparison was performed in this study, we have had previous experience with this methodology. There are many inherent advantages to using a targeted-capture method compared to an amplicon-based MPS approach for formalin-fixed material. Importantly, hybridisation capture uses baits covering considerably shorter regions than those used in amplicon-based approaches. Therefore as more templates are effectively available for sequencing, capture in this respect is more suitable for analysis of formalin-fixed derived DNA. In addition, capture enables bioinformatic alignment

based on non-duplicate reads providing confidence that independent templates are analysed in the process. By contrast, amplicon-based sequencing alignment is performed using many reads from the same PCR product that can not be differentiated according to the origin of the template. The more limited number of templates available in amplicon-based sequencing increases the likelihood of a template error being interpreted as a low frequency mutation¹⁴. This is being solved by the use of overlapping amplicons but other problems with the amplicon based approach remain.

In the past, hybridisation capture was considered unsuitable for clinical diagnostics because of high input DNA requirements, long turnaround times and low depths of coverage required to accurately measure low levels of mutations. This is increasingly less the case as improvements in automation and instrumentation now offer increased amounts of coverage in a more efficient and uniform manner. More efficient library preparation methods are also becoming increasing available with improved repair/adaptor tagging chemistry and have no intermediate cleanup steps. It is hoped that these refinements will bridge the gap between the input DNA amounts used in this study and those routinely obtained from clinical specimens. In one study by Kozarewa et al., a minimal input amount of 50 nanograms of DNA was demonstrated to be sufficient to obtain good quality sequencing data⁴⁵. This is comparable to the amount required for amplicon-based assays¹⁴.

No study to date has examined if sequence artifacts are present in formalin fixed samples through a capture approach. Artifacts due to cytosine deamination were not readily apparent in this study⁴⁶. We consider that this derives from a combination of the following reasons: (i) the FFPE samples were freshly prepared whereas cytosine deamination probably gradually accumulates with time in a storage-condition related manner, (ii) the proofreading enzymes used may have a restricted ability to read through uracil containing templates (iii) there are more available templates for capture-based approaches than for amplicon-based approaches and (iv) the use of the Picard algorithm in our bioinformatic pipeline means that only independent amplicons are assessed minimising the effect of individual damaged templates.

In this proof-of-principle study we have shown that formalin-fixed xenografts can be accurately assessed for their mutational profile with very similar results to the unfixed corresponding cell lines. Further assessment of actual clinical fresh frozen versus FFPE samples with varying degrees of heterogeneity would be desirable to confirm the clinical utility of this approach.

In conclusion, we describe a capture-based method that has the capability to analyse a large number of clinically relevant genes from formalin-fixed derived DNA that is comparable to DNA from unfixed sources. Ultimately, the implementation of these methodologies in a clinical setting will allow the sequencing of retrospective and prospective formalin-fixed material to form a comprehensive genetic profile of patients and aid in advancing a personalised medicine approach by directing targetable therapeutics and providing prognostic outcomes.

Methods

Tumour tissue and cell line DNA. DNA was extracted from 7 melanoma cell lines (Table 1) using the QIAamp DNA Blood Mini kit (Qiagen, Hilden, Germany). Five of the seven cell lines have somatic mutations which have been described on the Catalogue of Somatic Mutations in Cancer (COSMIC) database²³ (<http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/>). For xenograft studies, female Balb/c nude mice were injected subcutaneously with 5×10^6 cells in 0.05 mL of 50% Matrigel. When the tumours reached ~100 mm³, mice were euthanised and tumours were immediately excised. Animal experiments followed the National Health and Medical Research Council (NHMRC) Australian Code of Practice for the Care and Use of Animals for Scientific Purposes and were approved by the Peter MacCallum Cancer Centre Animal Experimentation Ethics Committee. Tumours were then formalin-fixed for a 24 hours period in a manner that is identical to those of routine pathology department protocols and paraffin embedded. For each FFPE block, ten 5 µm sections were prepared and mounted on glass slides. After baking the sections for five minutes on a hotplate at 70°C, the sections were deparaffinised in three



changes of xylene for two minutes each and were taken to water by three changes of 100% ethanol for two minutes each and DEPC water for two minutes. Subsequently, the sections were stained with 0.5% methyl green to assist with scraping of cells from slides. The scraped cells then underwent proteinase K digestion for 3 days at 56°C. DNA from digested FFPE DNA and the cell line DNA were extracted using the Genra Puregene kit (Qiagen) according to the manufacturer's protocol. Extracted DNA was quantified using a Qubit fluorometer (Invitrogen, Carlsbad, CA) and a Quant-iT dsDNA HS kit (Invitrogen).

FFPE DNA integrity assay. We used a PCR-based quality control assay to determine the degree of fragmentation in a DNA sample. This allowed the quality control of samples and selection of the samples with least compromised DNA integrity for use in MPS based discovery. We adopted a modified protocol based on the previously published protocol²¹. Modifications to the original protocol were as follows. The reaction mixture included 1× PCR buffer, 2.5 mM MgCl₂, the 100, 200, 300, and 400 bp primers used at 2.5 pmol each and the 600 bp primers at 5 pmol, 200 μM of dNTPs, 0.5U of HotStarTaq polymerase (Qiagen), 10 ng DNA and PCR grade water in a total volume of 20 μL. PCR conditions included an activation step of 15 minutes at 95°C followed by 45 cycles of 94°C for 1 minute, annealing for 1 minute and extension at 72°C for 1 minute and a final extension at 72°C for 7 minutes. The PCR products were run on a 2% (w/v) agarose gel in 1× TBE and stained with ethidium bromide. The wells were loaded with 20 μL of the PCR product mixture with 5 μL 5× loading dye. One μL pUC19/HpaII DNA. Molecular Weight Marker (GeneWorks, Hindmarsh, Australia) was run alongside the PCR products to determine their size.

Hybridisation capture. Capture was performed using the Agilent® SureSelect™ Human Kinome kit (Agilent Technologies, Santa Clara, CA). In the kit, capture probes target the coding region and associated UTRs of 612 genes including more than 500 known kinases and selected cancer-specific genes covering 3.2 Mb of the human genome. A list of genes targeted is shown in Supplementary Table 1. A total of 1 μg of DNA from both unfixed and formalin-fixed DNA based on Qubit readings was used for shearing. Samples were sheared using a Covaris and processed using XT SureSelect reagents following prescribed protocols.

Illumina sequencing. Samples were sequenced on the Illumina HiSeq 2000 with version 4 flow cells according to the manufacturer's instructions. Short insert, paired-end 100 bp reads were generated. Samples were loaded in an indexed pool of 6 samples across a single lane of a flow cell.

Mapping and variant calling. Image analysis and base calling was performed using Illumina's CASAVA v1.8. Sequence reads were aligned to the human genome (hg19 assembly) using BWA. Local realignment around indels and base quality score recalibration were performed using the Genome Analysis Tool Kit (GATK) software and duplicate reads removed using Picard (<http://picard.sourceforge.net>). Single nucleotide variants (SNVs) and indels were identified using the GATK Unified Genotyper. Variants were further filtered by requiring a minimum read depth of 30 and the presence of reads in both strands. Identification of single nucleotide changes between the 1–10% allele frequency range was performed according to a previously published analysis pipeline¹⁴.

Identification of contaminating mouse sequences. To determine if there was contamination of mouse sequences after capture from the xenograft-derived formalin-fixed samples, we assessed the proportion of variants caused by aligning mouse DNA against the human reference genome (hg19 build). 100 bp flanking sequences around the called variants were extracted from hg19, which were then aligned against the mouse reference (MM9) using BLAT⁴⁷. A BLAT score of 100 was used as a cutoff to indicate significance of homology. The proportion of variants with significant mouse homology was then compared between corresponding unfixed and formalin-fixed samples.

1. Hirota, S. *et al.* Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* **279**, 577–580 (1998).
2. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
3. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004).
4. Zhang, L. & Daly, R. J. Targeting the human kinome for cancer therapy: current perspectives. *Critical reviews in oncogenesis* **17**, 233–246 (2012).
5. Ferte, C., Andre, F. & Soria, J. C. Molecular circuits of solid tumors: prognostic and predictive tools for bedside use. *Nat Rev Clin Oncol* **7**, 367–380 (2010).
6. Liu, W. *et al.* Distinct clinical and pathological features are associated with the BRAF(T1799A(V600E)) mutation in primary melanoma. *J Invest Dermatol* **127**, 900–905 (2007).
7. Handolias, D. *et al.* Mutations in KIT occur at low frequency in melanomas arising from anatomical sites associated with chronic and intermittent sun exposure. *Pigment Cell Melanoma Res* **23**, 210–215 (2010).
8. Richter, A. *et al.* A multisite blinded study for the detection of BRAF mutations in formalin-fixed, paraffin-embedded malignant melanoma. *Sci. Rep.* **3**, 1659 (2013).
9. Do, H. & Dobrovic, A. Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil- DNA glycosylase. *Oncotarget* **3**, 546–558 (2012).

10. Meldrum, C., Doyle, M. A. & Tothill, R. W. Next-generation sequencing for cancer diagnostics: a practical perspective. *Clin Biochem Rev* **32**, 177–195 (2011).
11. Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* **11**, 685–696 (2010).
12. Harismendy, O. *et al.* Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. *Genome Biol* **12**, R124 (2011).
13. Forshew, T. *et al.* Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* **4**, 136ra168 (2012).
14. Do, H., Wong, S. Q., 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**, 1376–1383 (2013).
15. Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* **30**, 434–439 (2012).
16. Mar, V. J. *et al.* BRAF/NRAS wild-type melanomas have a high mutation load correlating with histologic and molecular signatures of UV damage. *Clinical cancer research: an official journal of the American Association for Cancer Research* **19**, 4589–4598 (2013).
17. Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395–399 (2012).
18. Swanton, C. Intratumor heterogeneity: evolution through space and time. *Cancer Res* **72**, 4875–4882 (2012).
19. Li, S. *et al.* Endocrine-Therapy-Resistant ESRI Variants Revealed by Genomic Characterization of Breast-Cancer-Derived Xenografts. *Cell reports* (2013).
20. Roychowdhury, S. *et al.* Personalized oncology through integrative high-throughput sequencing: a pilot study. *Sci Transl Med* **3**, 111ra121 (2011).
21. van Dongen, J. J. *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* **17**, 2257–2317 (2003).
22. Zhang, J., Yang, P. L. & Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nature reviews. Cancer* **9**, 28–39 (2009).
23. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic acids research* **39**, D945–950 (2011).
24. Biankin, A. V. & Grimmond, S. M. Novel cancer drivers: mining the kinome. *Genome Med* **5**, 19 (2013).
25. Nikolaev, S. I. *et al.* Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nat Genet* **44**, 133–139 (2012).
26. Stark, M. S. *et al.* Frequent somatic mutations in MAP3K5 and MAP3K9 in metastatic melanoma identified by exome sequencing. *Nat Genet* **44**, 165–169 (2012).
27. Handolias, D. *et al.* Clinical responses observed with imatinib or sorafenib in melanoma patients expressing mutations in KIT. *Br J Cancer* **102**, 1219–1223 (2010).
28. Prickett, T. D. *et al.* Analysis of the tyrosine kinome in melanoma reveals recurrent mutations in ERBB4. *Nat Genet* **41**, 1127–1132 (2009).
29. Lavin, M. F. Ataxia-telangiectasia: from a rare disorder to a paradigm for cell signalling and cancer. *Nat Rev Mol Cell Biol* **9**, 759–769 (2008).
30. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157–1160 (2011).
31. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
32. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27**, 182–189 (2009).
33. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
34. Mar, V. J. *et al.* BRAF/NRAS Wild-Type Melanomas Have a High Mutation Load Correlating with Histologic and Molecular Signatures of UV Damage. *Clin Cancer Res* **19**, 4589–4598 (2013).
35. Berger, M. F. *et al.* Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature* **485**, 502–506 (2012).
36. Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251–263 (2012).
37. Krauthammer, M. *et al.* Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat Genet* **44**, 1006–1014 (2012).
38. Wei, X. *et al.* Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nat Genet* **43**, 442–446 (2011).
39. Mamedov, T. G. *et al.* A fundamental study of the PCR amplification of GC-rich DNA templates. *Comput Biol Chem* **32**, 452–457 (2008).
40. Wagle, N. *et al.* High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer discovery* **2**, 82–93 (2012).
41. Adams, M. D. *et al.* Global mutational profiling of formalin-fixed human colon cancers from a pathology archive. *Mod Pathol* **25**, 1599–1608 (2012).
42. Kerick, M. *et al.* Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. *BMC Med Genomics* **4**, 68 (2011).
43. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883–892 (2012).
44. Schatton, T. *et al.* Identification of cells initiating human melanomas. *Nature* **451**, 345–349 (2008).
45. Kozarewa, I. *et al.* A modified method for whole exome resequencing from minimal amounts of starting DNA. *PLoS One* **7**, e32617 (2012).



46. Do, H. & Dobrovic, A. Limited copy number-high resolution melting (LCN-HRM) enables the detection and identification by sequencing of low level mutations in cancer biopsies. *Mol Cancer* **8**, 82 (2009).
47. Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656–664 (2002).

Acknowledgments

This project was enabled by the Peter MacCallum Pfizer Translational Oncology Research Collaborative Hub (TORCH) funded by Pfizer Australia, and the Melbourne Melanoma Project funded by the Victorian Cancer Agency Translational research program and established through support of the Victor Smorgon Charitable Fund. We thank Carleen Cullinane for preparing the xenograft blocks and Jia-Min Pang for review of slides for contaminating normal mouse tissue in xenografted formalin-fixed sections. We also thank Andrew Fellowes, Anthony Bell, and Stephen Fox for the TruSeq reagent and access to the MiSeq instrument.

Author contributions

G.M., S.W., R.S., H.D., R.T. and A.D. were involved in the conceptualisation and design of

the project. The major experimental work for this project was carried out by S.W. and R.S. J.L. performed the bioinformatics analysis. K.S. coordinated the xenograft work. S.W. and A.D. wrote and revised the main manuscript text and prepared the figures. All authors reviewed and approved the manuscript.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: G.M. has received commercial research funding from Pfizer, Novartis, and Millennium. K.S. and A.D. have received commercial research funding from Pfizer.

How to cite this article: Wong, S.Q. *et al.* Targeted-capture massively-parallel sequencing enables robust detection of clinically informative mutations from formalin-fixed tumours. *Sci. Rep.* **3**, 3494; DOI:10.1038/srep03494 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>