1

2 **Deep oncopanel sequencing reveals fixation time- and within block**
 3 **position-dependent quality degradation in FFPE processed samples**

4

5 SEQC2 Oncopanel Sequencing Working Group

6

7 **Abstract**

8 Clinical laboratories routinely use formalin-fixed paraffin-embedded (FFPE) tissue or cell block
 9 cytology samples in oncology panel sequencing to identify mutations that can predict patient
 10 response to targeted therapy. To understand the technical error due to FFPE processing, a robustly
 11 characterized normal cell line was used to create FFPE samples with four different pre-tissue processing
 12 formalin fixation times. A total of 96 FFPE sections were then distributed to different laboratories for
 13 targeted sequencing analysis by four oncopanels, and variants resulting from technical error were
 14 identified. Tissue sections that failed more frequently showed low cellularity, lower than recommended
 15 library preparation DNA input, or target sequencing depth. Importantly, sections from block surfaces were
 16 more likely to show FFPE-specific errors, akin to “edge effects” seen in histology, and the depth of
 17 formalin damage was related to fixation time. To assure reliable results, we recommend avoiding the
 18 block surface portion and restricting mutation detection to genomic regions of high confidence.

19

20 **Keywords**

21 cancer genomics, next generation sequencing, FFPE, preanalytics, precision medicine, oncopanel
 22 sequencing

1 Yifan Zhang^{1*}, Thomas M. Blomquist^{2,3*}, Rebecca Kusko^{4*}, Daniel Stetson⁵, Zhihong Zhang⁶, Lihui Yin⁷,
2 Robert Sebra⁸, Binsheng Gong¹, Jennifer S. LoCoco⁹, Vinay K. Mittal¹⁰, Natalia Novoradovskaya¹¹, Ji-Youn
3 Yeo¹², Nicole Dominiak¹², Jennifer Hipp¹³, Amelia Raymond⁵, Fujun Qiu⁶, Hanane Arib⁸, Melissa L. Smith⁸,
4 Jay E. Brock¹⁴, Daniel H. Farkas¹⁴, Daniel J. Craig¹⁵, Erin L. Crawford¹⁵, Dan Li¹, Tom Morrison¹⁶, Nikola
5 Tom^{17,18}, Wenzhong Xiao^{19,20}, Mary Yang²¹, Christopher E. Mason²², Todd A. Richmond²³, Wendell Jones²⁴,
6 Donald J. Johann Jr²⁵, Leming Shi^{26,27,28}, Weida Tong¹, James C. Willey^{29†} & Joshua Xu^{1†}

7 * These authors contributed equally to this work.

8 † Corresponding authors:

9 Dr. James C. Willey: james.willey2@utoledo.edu

10 Dr. Joshua Xu: joshua.xu@fda.hhs.gov

11

12 ¹ Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and
13 Drug Administration, Jefferson, AR 72079, USA

14 ² (formerly) Department of Pathology, College of Medicine and Life Sciences, The University of Toledo,
15 Toledo, OH 43614, USA

16 ³ Lucas County Coroner's Office, 2595 Arlington Ave., Toledo, OH 43614, USA

17 ⁴ Immuneering Corporation, One Broadway, 14th Floor, Cambridge, MA 02142, USA

18 ⁵ Astrazeneca Pharmaceuticals, 35 Gatehouse Dr, Waltham, MA 02451, USA

19 ⁶ Research and Development, Burning Rock Biotech, Shanghai 201114, China

20 ⁷ (formerly) Pathology and Laboratory Medicine Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland,
21 OH 44195, USA

22 ⁸ Icahn Institute and Dept. of Genetics and Genomic Sciences Icahn School of Medicine at Mount Sinai,

- 1 1425 Madison Ave., New York, NY 10029, USA
- 2 ⁹ Illumina Inc., 5200 Illumina Way, San Diego, CA 92122, USA
- 3 ¹⁰ Thermo Fisher Scientific, 110 Miller Ave., Ann Arbor, MI 48104, USA
- 4 ¹¹ Agilent Technologies, 11011 N Torrey Pines Rd., La Jolla, CA 92037, USA
- 5 ¹² Department of Pathology, University of Toledo, 3000 Arlington Ave., Toledo, OH 43614, USA
- 6 ¹³ Department of Pathology, Strata Oncology, Inc., Ann Arbor, MI 48103, USA
- 7 ¹⁴ Pathology and Laboratory Medicine Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, OH
- 8 44195, USA
- 9 ¹⁵ Department of Medicine, College of Medicine and Life Sciences, The University of Toledo, Toledo, OH
- 10 43614, USA
- 11 ¹⁶ Accugenomics, Inc., 1410 Commonwealth Drive, Suite 105, Wilmington, NC 20403, USA
- 12 ¹⁷ Center of Molecular Medicine, Central European Institute of Technology, Masaryk University, Kamenice
- 13 5, 625 00 Brno, Czech Republic
- 14 ¹⁸ EATRIS ERIC- European Infrastructure for Translational Medicine, De Boelelaan 1118, 1081 HZ
- 15 Amsterdam, The Netherlands
- 16 ¹⁹ Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA
- 17 ²⁰ Stanford Genome Technology Center, Stanford University, Palo Alto, CA 94304, USA
- 18 ²¹ Department of Information Science, University of Arkansas at Little Rock, 2801 S. Univ. Ave., Little Rock,
- 19 AR 72204, USA
- 20 ²² Department of Physiology and Biophysics, Weill Cornell Medicine, Cornell University, New York, NY
- 21 10065, USA
- 22 ²³ Market & Application Development Bioinformatics, Roche Sequencing Solutions Inc., 4300 Hacienda Dr.,

- 1 Pleasanton, CA 94588, USA
- 2 ²⁴ Q2 Solutions - EA Genomics, 5927 S Miami Blvd., Morrisville, NC 27560, USA
- 3 ²⁵ Winthrop P Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, 4301 W Markham
- 4 St., Little Rock, AR 72205, USA
- 5 ²⁶ State Key Laboratory of Genetic Engineering, School of Life Sciences and Shanghai Cancer
- 6 Hospital/Cancer Institute, Fudan University, Shanghai 200438, China
- 7 ²⁷ Human Phenome Institute, Fudan University, Shanghai 201203, China
- 8 ²⁸ Fudan-Gospel Joint Research Center for Precision Medicine, Fudan University, Shanghai 200438, China
- 9 ²⁹ Departments of Medicine, Pathology, and Cancer Biology, College of Medicine and Life Sciences,
- 10 University of Toledo Health Sciences Campus, 3000 Arlington Ave., Toledo, OH 43614, USA

1 Introduction

2 Next generation sequencing (NGS) is now an integral tool in the “precision” cancer care arsenal
 3 (NCI-match studies; ECOG and other substudy trials). Despite excellent performance for somatic mutation
 4 calling¹, preanalytical variation continues to limit the quality and quantity of cancer specimens, and this
 5 ultimately impacts NGS accuracy and reproducibility^{2–4}. One source of preanalytical error stems from
 6 formalin-fixation and paraffin-embedding (FFPE). FFPE processing of tumor specimens is central to
 7 histologic diagnosis of cancer and its pursuant sub-classification, grading, staging, and
 8 adequacy-assessment for ancillary studies in routine clinical workup⁵. For this reason, many ancillary
 9 prognostic and treatment markers are optimized for FFPE tissue^{6,7}. When combined with a growing trend
 10 of limited specimen size and quality⁸, in many circumstances, tissue is processed for FFPE to meet the
 11 bulk of current standard care needs⁵. However, FFPE processing harbors substantial and highly variable
 12 effects on nucleic acid quality and quantity⁹. Thus, platforms for targeted NGS analysis of FFPE clinical
 13 specimens must be subjected to rigorous analytical validation^{10–12}. It is of utmost importance that we
 14 understand factors that influence the reproducibility and accuracy of NGS testing in FFPE tissue to ensure
 15 that cancer care is fully optimized.

16
 17 The process of formalin fixation and paraffin embedding includes several steps with varying degrees of: 1)
 18 fixation, 2) progressive dehydration, 3) clearing, and 4) molten paraffin infiltration¹³. For each step, clinical
 19 labs independently establish the ideal amount of time, temperature, pressure, agitation, and reagent
 20 composition, typically with the primary goal of optimizing histologic quality^{13,14}. Some guidelines do exist
 21 for specific preanalytical FFPE metrics (e.g., fixation time). For example, the College of American
 22 Pathologists recognizes that fixation time can influence accuracy and reproducibility of ancillary tests

1 performed on FFPE such as ER/PgR immunohistochemistry and HER2 in-situ hybridization; and for these
 2 specimens, formalin fixation should be limited to >6 or <72 hours of total formalin exposure^{6,7}. Still, false
 3 somatic mutation calling rates by NGS vary greatly for FFPE specimens from different, and even within the
 4 same, anatomic laboratories¹⁵. Of even greater concern, FFPE-derived sequencing errors may arise at
 5 clinically relevant loci and at actionable allelic frequencies¹⁶. Working groups are beginning to formulate
 6 and extend more detailed recommendations on pre-analytical controls⁴. While it is important that these
 7 practice recommendations are based on human biospecimen research data, some important questions
 8 may benefit from additional studies using carefully contrived sample sets as conducted here.

9
 10 In this study, the Oncopanel Sequencing Working Group of the FDA-led Sequencing Quality Control Phase
 11 II (SEQC2) consortium extended the study¹ to investigate the impact of FFPE processing. This line of
 12 inquiry has served to address the insufficiency of well-controlled data regarding FFPE tissue, its
 13 preanalytical metrics, and impact on NGS accuracy and reproducibility¹⁷. Here, we adopted an easy to
 14 follow FFPE preparation protocol (see Methods for details) that is highly analogous to clinically obtained
 15 cell block cytology sample processing and can be readily applied to existing reference cell line
 16 materials¹⁸⁻²⁰. The benefits of the approach used here are that: 1) reference variant data and the ability to
 17 compare with numerous orthologous molecular methods provide robust information for accuracy and
 18 reproducibility studies, 2) genetic heterogeneity across replicate measures is minimized (a common
 19 challenge in clinical FFPE NGS studies), and perhaps most importantly, 3) preanalytical FFPE effects on
 20 technical artifacts can be well-documented and controlled, enabling future meta-analysis and subsequent
 21 guideline recommendations. As an initial investigation, cultured cell samples from a normal cell line¹⁸
 22 (Agilent Male Lymphoblast Cell Line) were subjected to varying formalin fixation times between 1 and 24

1 hours prior to tissue processing^{4,21,22}. FFPE sectioning samples at multiple locations within each FFPE
2 block were selected and distributed to four independent laboratories for targeted NGS following amplicon
3 or hybrid capture enrichment. Based on reference variant data, we identified false positive (FP) calls and
4 estimated FP rate (FPR) within each panel's targeted region—a key quality control metric demonstrated
5 first in our cross oncopanel investigation¹. Comprehensive analysis was then conducted on FP calls and
6 FPR to decipher the effects of FFPE factors including formalin fixation time and tissue block section
7 position.

8

9 **Results**

10 *Overview of study design and analysis*

11 To investigate the effect of formalin fixation time and tissue block section position on targeted NGS
12 analysis of FFPE specimens, we designed a comprehensive study querying several key components. Fig. 1
13 displays the flow of three major components: FFPE sample preparation (Fig. 1A), sequencing experiments
14 with four diverse oncopanels (Fig. 1B), and data quality control and analysis (Fig. 1C). High quality
15 genomic DNA from a single normal cell line (Agilent Male Lymphoblast Cell Line) was sequenced with
16 multiple oncopanels and technical replicates in our companion study¹. These datasets enabled the
17 establishment of a known variant set for each panel and the subsequent detection of artifacts induced by
18 the FFPE process (Fig. 1C). Cultured cell populations were used to make FFPE samples with four different
19 formalin fixation times. Equal amounts of cultured cells were mixed in each gel matrix mold, followed by
20 formalin fixation, routine tissue processing, and paraffin embedding (Fig. 1A). Samples were created by
21 sectioning FFPE blocks as described in the Methods section. Based on their estimated cell counts and
22 positions in the FFPE blocks, samples were assigned to two categories: surface (either top or bottom) or

1 inner FFPE samples (see Methods for details). Each laboratory extracted and quantified genomic DNA
2 from 24 samples evenly distributed across eight distinct FFPE blocks. NGS sequencing experiments and
3 subsequent bioinformatics processing were conducted following vendor recommended protocol (Fig. 1B,
4 see Methods). These FFPE samples were sequenced by four oncopanels (Table S1): AstraZeneca 650
5 genes Oncology Research Panel (AZ650), Burning Rock DX OncoScreen Plus (BRP), Illumina TruSight Tumor
6 170 (ILM), and Thermo Fisher Oncomine Comprehensive Assay v3 (TFS). Variant calling results and QC
7 data were collected and submitted to the Working Group for integrated analysis (Fig. 1C, see Methods for
8 details).

9
10 Five QC check steps were included in our workflow (Fig. 1): 1) cell count estimation, 2) DNA extraction
11 yield, 3) library construction yield, 4) median depth and library complexity computed from mapped reads,
12 and 5) variant histograms by variant type and VAF for the detection of oxidative damage or contamination.
13 All QC data are provided in **Supplementary File 1**. Six samples failed in the ILM experiments based on very
14 low target sequencing depth and low library yield, which may be related to the lower than recommended
15 library input. Four samples failed both attempts of library preparation for TFS and were not sequenced. In
16 each case, those failed samples appeared to have been processed together in one batch, potentially
17 failing at the same step between DNA extraction and library construction. Five samples were possibly
18 contaminated before or during the NGS experiments (two in AZ650, two in BRP, and one in ILM,
19 respectively). These samples were excluded from further analysis on FFPE effects.

20
21 *The main constituents of the known variant set were homozygous or heterozygous germline variants*

22 To differentiate FPs from true variants, it was imperative to first build the set of reference (known)

1 variants for each panel. By aggregating the variants called by over 75% of the fresh gDNA samples that
2 were free of FFPE damage, we generated a set of known variants for each panel. The lone exception was
3 AZ650 whose known variant set was generated from FFPE samples that passed stringent QC filters
4 because AZ650 was not included in our companion study of multiple oncopanels using fresh gDNA
5 samples¹. To determine FPs introduced by FFPE processing, variant calls from each FFPE sample were
6 compared with the known variant set for each respective panel.

7

8 We expected to detect only homozygous and heterozygous variants in the known variant sets since a
9 normal cell line was used. In general, the variant allele frequencies (VAFs) were close to either 0.5 or 1,
10 with heterozygous variants (near 0.5) showing more dispersion in allele frequency than homozygous
11 variants (**Fig. 2A**). Variants in the known variant sets located outside of the consensus high confidence
12 targeted region (CTR, see Methods for details) showed greater VAF dispersion than those within the CTR
13 (**Fig. 2A**), particularly for the BRP, ILM, and TFS panels. As expected, more reference variants of low VAF
14 were reported for AZ650 than for the other three panels.

15

16 The distribution plot revealed few variants falling between VAF 0.1-0.3 and 0.7-0.9 (**Fig. 2A**). Thus, we
17 choose 0.2 and 0.8 as the boundaries for heterozygous and homozygous variants. We classified the
18 known variant set into four groups by the VAF value: homozygous variants (VAF > 0.8), heterozygous
19 variants (0.2 < VAF ≤ 0.8), and two additional low VAF ranges separated by a VAF at 0.1. For variants called
20 within the CTR, only two variants were in the two low VAF ranges for AZ650 while none were called for
21 the other three panels. Among the variants called outside of the CTR, slightly more fell into the low VAF
22 ranges and their proportion relative to all variants remained low (**Fig. 2B**). Furthermore, we confirmed the

1 high concordance (>98%) among known variants across panels, thus the known variants were reliable.

2 The known variants for each panel are listed in **Supplementary File 2**.

3

4 *Multiple factors can lead to a high false positive rate in FFPE samples*

5 Oncopanel experiments can be marked as failed or flagged for further examination by collecting many
6 quality measurements using pre-analytical and analytical instruments, bioinformatic QC tools, and
7 reference thresholds from large cohorts. In our analysis, several QC measurements (Fig. 3) were found
8 practically useful to determine whether the experiments failed or to discover outliers that required
9 further investigation. These QC measurements, including cell count, DNA input, deduplicated sequencing
10 depth, library complexity, G:C to T:A transversion count, and VAF distribution, are described below in
11 detail. Cell count was an essential measurement for the quality of sequential FFPE sample sections and
12 was the very first measurement for our pre-analytical QC check. Samples with a lower cell count usually
13 showed a much higher FPR (Fig. 3A, circled in a red dashed line). However, several samples were found to
14 have sufficient cell counts but high FPR (Fig. 3A, circled in an orange dashed line). In subsequent QC
15 evaluation, these experiments were identified as failed or were flagged as outliers. Next, DNA input was
16 another important measurement, particularly useful prior to library preparation. For example, four TFS
17 libraries with DNA input lower than 2.5 ng showed an FPR (per million bases) greater than 50 (Fig. 3B). As
18 seen in Fig. 3B, slightly lower DNA input substantially increased FPR. Mapping of FASTQ files to the
19 reference genome enabled calculation of additional QC measurements. Lower sequencing depth usually
20 correlated with a higher FPR, especially in the regions outside of the CTR (Fig. 3C, showcased by ILM data
21 but observed across all panels). In addition, library complexity was a similar QC measurement for
22 assessing the extent of PCR-associated read duplication. Taking BRP results as an example, the FPR

1 significantly increased when the library complexity was lower than 0.25 (Fig. 3D). After variant calling,
 2 several measurements can be calculated and visualized for discovering outliers. For example, G:C to T:A
 3 transversion count is a known indicator of oxidative nucleotide damage²³. Higher G:C to T:A transversion
 4 count may be the natural state of a sample, or it may be due to exposure to oxidative radicals. Since all
 5 samples were created from a normal cell line, the samples with a high G:C to T:A transversion count were
 6 flagged as failed (Fig. 3E). As we were plotting the VAF histogram for each sample, we surprisingly found a
 7 handful experiments with an unexpected tail on the left side of 100%. Homozygous germline mutations
 8 are expected to be observed with a VAF of 100%, with very little or no leftward shift. Taking sample
 9 6_G_7 as an example, the wider shift may have been the result of a small amount of contamination (Fig.
 10 3F). In summary, we demonstrated how multiple QC checks can be utilized to understand the factors
 11 contributing to high FPR for some FFPE samples.

12

13 *False positive rates were impacted by genomic regions and can be effectively controlled by a VAF cutoff*

14 The majority of surface FFPE samples failed one or more of the QC checks described above.
 15 Understandably, the surface FFPE samples were exposed to longer and more intensive chemical damage
 16 than the inner FFPE samples. Using the inner FFPE samples that passed all QC checks, we evaluated their
 17 FPRs compared against fresh samples to investigate the impacts of genomic regions and VAF cutoffs on
 18 the FPR. Given that variant detection performance may be influenced by the chosen alternate allele
 19 depth (ADP) threshold for variant calling, we also evaluated the combination of VAF cutoffs with
 20 additional ADP cutoffs. All panels showed the lowest FP rates within the CTR compared against outside
 21 the CTR (Fig. 4A).

22

1 We then employed a one-tailed t-test to check whether the FPR of the QC-passing inner FFPE sample
 2 group was higher than that of the fresh sample group (Fig. 4A). For the whole panel region, ILM was the
 3 only panel with higher FPRs in FFPE samples. That may be explained by the lower DNA input amount (≤ 30
 4 ng) for the FFPE samples in comparison to 50 ng for fresh DNA samples. Within the CTR, only the TFS
 5 panel showed a significant difference in FPR between the two sample groups (Fig. 4A). Again, the variable
 6 DNA input amounts may have been a contributing factor. Taken together, these observations suggested
 7 no consistent impact on the FP rate by the FFPE process. The standardized FFPE procedure together with
 8 the rigorous quality control checks described above could thus achieve an FP rate close to that of fresh
 9 gDNA samples.

10

11 We also applied a range of additional VAF and ADP cutoffs beyond each panel's internal thresholds to
 12 discern the influence of VAF and ADP thresholds on FP rates (Fig. 4B-C). Across all of the panels, applying
 13 both criteria reduced FP rates as expected. VAF cutoffs ranging from 1% to 10% led to a higher degree of
 14 FP rate reduction than applying additional ADP cutoffs. Of note, the benefits were not consistent across
 15 sample types (FFPE samples vs fresh samples) or oncopanels. When considering FFPE samples only (Fig.
 16 4C), an increase in VAF cutoff from 2.5% to 10% resulted in only modest changes in BRP's FPR but notable
 17 changes in TFS's FPR. The VAF and ADP cutoffs were interdependent criteria; how much each one
 18 influenced reduction of the FPR depended on analytical filters such as total sequencing depth. Based on
 19 the results of this study, we suggest that cutoff selection should be based on the desired or required FPR.

20

21 These results suggested that the CTR is a more reliable region for variant reporting, as evidenced by the
 22 much lower FPRs achieved in the CTR. Additionally, we confirmed the effectiveness of stricter VAF cutoffs

1 to control the FFPE sample FP rate.

2

3 *Surface FFPE samples showed significantly more FFPE damage and artifacts due to hydrolytic deamination*

4 We extended the analysis of FP variant calls to include “surface” FFPE samples and QC-failed “inner”
5 samples. Since the FPR was much higher outside the CTR, the analysis was confined to the CTR. The FPs
6 were further classified into four variant categories: 1) indels, 2) hydrolytic deamination introduced
7 artifacts (G:C>A:T SNVs), 3) oxidative damage artifacts (G:C>T:A SNVs), and 4) other FP calls. The FP count
8 per sample was plotted for each sample category (Fig. 5A).

9

10 Overall, compared with fresh and QC-passed inner FFPE samples, both surface and QC-failed inner FFPE
11 samples produced more FP calls. Except for ILM, the surface FFPE samples consistently made more FP
12 calls per sample than the QC-failed inner FFPE samples (Fig. 5A). Furthermore, across all panels, the
13 surface FFPE samples yielded more hydrolytic deamination artifacts per sample than the QC-failed inner
14 FFPE samples. This clear pattern indicated that the main driver of surface FFPE samples FP calls was
15 hydrolytic deamination introduced artifacts (G:C>A:T). Within the known variant set, the G:C>A:T variants
16 constituted approximately 40% of all variants for each panel (Fig. 5B). In comparison, the proportion of
17 G:C>A:T FP calls for surface FFPE samples was much higher than 40%, which was the baseline value for
18 the known variant sets. The lone exception was BRP whose surface FFPE samples also produced a high
19 number of FP G:C>T:A calls per sample, presumably due to some oxidative damage (Fig. 5C).
20 Understandably, the surface FFPE samples were physically exposed to more intensive formalin treatment,
21 leading to more FFPE process-induced uracil lesions²⁴. Our results clearly demonstrated that more DNA
22 damage occurred on the surface portions of the sample during the FFPE process.

1

2 It was understandable that all five surface samples with low cell counts (1600 - 6200) failed the
3 sequencing experiments. Even after excluding them, the QC-passing rate among the surface sample group
4 (1 out of 5) was drastically lower than that of the inner FFPE sample group (60 out of 71). The samples
5 taken from the inner portion of a FFPE block performed significantly better than samples from the surface
6 portion ($p < 0.004$, Pearson's Chi-squared test with Yates' continuity correction).

7

8 *Longer formalin fixation reduced the portion of high-quality samples*

9 A previous study reported that formalin fixation times longer than 24 hours could lead to more
10 deamination artifacts²⁵. In this study, we limited the fixation time to 24 hours maximum. Within the
11 QC-passed inner FFPE samples, we examined whether the fixation time would lead to any consistent
12 differences in FPR. This analysis was first performed for each panel (Fig. 6A-D). Except for TFS, there was
13 no observable effect of formalin fixation time on the FPR of inner FFPE samples. For TFS, the longer
14 fixation group (i.e., 6 hours and 24 hours) appeared to show more elevated FPRs than the shorter fixation
15 group (i.e., 1 hour and 2 hours). However, when pooling all samples together for analysis, there were no
16 significant differences in FPR between the long and short fixation groups (Fig. S1). Using the FPR as a
17 quality measure for the sample, this result indicated that longer formalin fixation time had no detrimental
18 effect on the quality of inner FFPE samples. We then calculated the average cell count of inner FFPE
19 samples for each block and tested whether there was any difference across the fixation time groups. The
20 FFPE blocks were made in the same dimension with similar cell densities. As expected, the average cell
21 count showed no effect by formalin fixation time (Fig. 6E). Taken together, regardless of the fixation time,
22 inner FFPE samples showed no FFPE damage and could achieve the same low false positive rates as fresh

1 DNA samples (Fig. 4A, Fig. 5A).

2

3 Finally, we counted all inner samples within each FFPE block and calculated their aggregated thickness
4 (i.e., depth of the inner sample portion) within each block and compared it across the fixation time
5 groups. Factoring the alternative sectioning slide stained for cell counting, each sample comprised nine 5
6 μm sections and thus was equivalent to a portion of 45 μm thickness within the block. The shorter
7 fixation groups (i.e., 1 hour and 2 hours) clearly contained more inner samples and the inner portion was
8 thus thicker than that in the longer fixation group (i.e., 6 hours and 24 hours) (Fig. 6F and Supplementary
9 File 1 – Sample Cell Counts Table). This formalin fixation-related shrinkage artifact is well documented⁵,
10 and this condition combined with longer fixation times could have led to fewer high quality inner FFPE
11 samples available for oncopanel testing.

12

13 Discussion

14 The goal of this study was to identify and characterize poorly understood sources of technical variation
15 associated with targeted NGS analysis of variants in FFPE tissue and cell block samples. Through the effort
16 of the Oncopanel Sequencing Working Group of the SEQC2 consortium, we used multiple oncopanels to
17 systematically survey the effect of formalin fixation time, genomic location, and sectioning position within
18 an FFPE block. Owing to this investigation's focus on technical variation, a single robustly characterized
19 normal cell line (Agilent Male Lymphoblast Cell Line) was used to create FFPE cell blocks with four
20 different formalin fixation times. We adopted an FFPE procedure that is highly analogous to cell block
21 cytology sample processing so that the findings from our study would be relevant to oncopanel
22 sequencing of clinical FFPE samples. We identified multiple previously unrecognized and avoidable

1 sources of variation that, if addressed by appropriate QC measures, should enable more reliable use of
2 FFPE samples.

3

4 We leveraged results from the SEQC2 Oncopanel Sequencing working group that extensively sequenced
5 this cell line in a companion study¹ to create variant reference sets. However, the variant reference set for
6 AZ650 was generated from FFPE samples that passed stringent QC filters because AZ650 was not included
7 in the companion study of multiple oncopanels with fresh gDNA samples. This might have led to an
8 underestimation of FP rates for samples sequenced by AZ650 because few variant calls due to FFPE
9 damage may have been included as reference variants. This did not alter our overall conclusion based on
10 the analysis results of FPR because FPR was drastically different for any panel in the pairwise comparisons
11 of genomic regions (Fig. 4A), QC status, and sample groups (Fig. 5A).

12

13 Given that a QC related issue in any single step of FFPE processing, sample preparation, library
14 preparation, sequencing, and/or bioinformatics can increase false positive/negative rates, we sought to
15 survey QC checks throughout the entire process. For the first step (pre-analytical QC), cell count and DNA
16 input were key drivers of the FPR. For the second step (Oncopanel NGS QC), there was a strong and
17 significant correlation between minimum deduplicated sequencing depth (DSD) and FPR. In support of
18 this finding, experiments were far more likely to fail if library complexity was below 0.25. After removing
19 contaminated samples, a minimum median DSD of 600X effectively controlled FPR and separated high
20 quality FFPE samples from failed samples for the three capture-based panels. For each hybrid-capture
21 panel, we observed some moderate correlation between median DSD and DNA input amount for library
22 preparation. However, there was no correlation between cell count and DNA extraction yield for any

1 panel after excluding those surface samples with very low cell count. The loss of correlation may have
 2 been due to variations in cell count estimation and DNA extraction efficiency. For the third step
 3 (post-variant call QC), G:C to T:A transversions may be employed to pinpoint outliers associated with
 4 oxidative damage. A per sample VAF histogram also provided an important QC metric. A long tail from the
 5 left side of 100% VAF was likely an indicator of sample contamination. This sample contamination check is
 6 applicable to impure tumor samples that may contain DNA from stromal cells. While the exact QC cutoffs
 7 employed in this study may not extrapolate perfectly to different panels, tissue samples, and FFPE
 8 processes, we expect these QC metrics to remain relevant across oncopanels.

9
 10 Not all variants fall into equally easy-to-call genomic regions, thus we queried the impact of variant
 11 location on FFPE sample calling. All panels generally showed the lowest FPRs in the CTR. Thus, restriction
 12 of variant calls to within the CTR boosts reliability in the context of FFPE samples. Genomic region also
 13 showed interactions with other QC metrics. For variants falling outside of the CTR, the impact of alternate
 14 read depth cutoff was more pronounced (lower alternate read depth cutoff led to higher FPR). In the case
 15 of precious clinical samples from an individual patient, it may not always be possible to employ such
 16 rigorous QC metrics because choices are limited. Here, we found that it was possible to reduce the FPR
 17 through higher VAF cutoff, variant allele depth cutoff, and restriction of variant calling to the consensus
 18 high confidence region. The trade-off was reduced sensitivity.

19
 20 Interestingly, physical position within the FFPE block showed a significant effect on the amount of
 21 measured FFPE damage and technical artifacts. Analogous to this, “edge effect” in histologic tissue
 22 section staining is observed regularly by pathologists as a source of error in prognostic and treatment

1 marker interpretation²⁶. Its cause is not completely understood but is thought to stem from the surfaces
2 of tissue sections experiencing a combination of drying artifact, oxidative damage, as well as formalin
3 fixation damage. In our study, sections taken from the inner portion of the FFPE block were less likely to
4 fail our QC metrics than samples from the surface portions, which almost always failed QC; perhaps due
5 to the same underlying factors causing “edge effect” seen in traditional histologic markers. This leads to
6 an actionable recommendation that samples from the surface portions of an FFPE block should be
7 avoided if possible when selecting samples for oncopanel sequencing. Follow-up studies should be
8 conducted to confirm and further characterize this newly identified intra-sample source of FFPE quality
9 variation. In addition, samples that underwent longer formalin fixation times were more likely to produce
10 less high quality FFPE sections suitable for genomic sequencing. To the extent possible, shortening
11 formalin fixation time is recommended to enhance the quality of genomic sequencing. Importantly, our
12 study employed cell culture samples rather than surgically cut tissue samples. Intact tissue samples
13 harbor widely varying formalin penetration rates, and this can greatly impact formalin fixation time (e.g.
14 brain [fatty] vs muscle [high water content]). For clinical samples, it is important to make samples as thin
15 as possible to improve uniformity of fixation in the shortest amount of time possible. Our findings may
16 not extend quantitatively to FFPE samples from intact tissue samples, but we anticipate our findings will
17 apply qualitatively. Currently there are no clear guidelines for a specimen formalin fixation window to
18 optimize FFPE samples for NGS testing. A recent publication resulted from the CAP Preanalytics for
19 Precision Medicine Project which recommended that specimen sample thickness be less than 5 mm and
20 total fixation time be between 6 and 24 hours for nonfatty tissues to ensure molecular integrity of
21 cancer specimens⁴. To extend this recommendation, we suggest that the total fixation time be limited to 6
22 hours and the surface portions be avoided when choosing FFPE sections for oncopanel sequencing.

1

2 Taken together, our work advocates for a robust set of QC metrics querying various steps in the process
3 from sample to sequencing to bioinformatics. For the first time, through comprehensive multi-laboratory
4 oncopanel sequencing of 96 samples created under well-controlled FFPE processing, we quantitatively
5 evaluated the effects of formalin fixation duration and within-block position on data quality. We specified
6 the portion of FFPE block that would be suitable for oncopanel sequencing experiments and discovered
7 that the size of such portion was dependent on formalin fixation time. Thus, shorter fixation time is
8 recommended to the extent possible. To ensure reliable results, our results support the application of
9 strict threshold criteria for cell count, DNA input, allele frequency, and restriction of analysis to genomic
10 regions of high confidence. FFPE samples are archived on a routine basis in pathology departments
11 around the world. By identifying specific quality control factors that affect targeted NGS analysis of FFPE
12 samples we hope to increase their value in research and clinical diagnostics.

13

14 **Materials and Methods**

15 **Sample Preparation, Quality Check, and Distribution**

16 The Agilent male lymphoblast cell line was cultured in T75 flasks (Corning Catalog No. 10-126-28) and
17 harvested according to vendor product specifications. The harvested cellular material was combined into
18 a single 15 mL conical tube (Falcon Catalog No. 14-959-53A) and resuspended to a total volume of 1 mL
19 with 10% neutral buffered formalin (NBF) (StatLab Catalog No. 28600). In separate vials, HistoGel
20 specimen processing gel matrix (ThermoFisher Catalog No. HG-4000-012) was heated to 60 °C for 2 hours
21 to liquify and then allowed to cool and equilibrate to 45 °C in a vendor supplied thermal block
22 (ThermoFisher Catalog No. HGSK-2050-1). Replicate square shape cell-block molds were set up

1 (Fisherbrand Catalog No. EDU00553), and into each mold, 500 μ L of 45 °C HistoGel was added; Then, 100
2 μ L of NBF- suspended cell line mixture was added. These were immediately and gently stirred to ensure
3 homogeneity of cells within the cooling HistoGel matrix, and then they were allowed to sit and solidify on
4 the bench top for at least 5 minutes. Next, for each mold, a micro-spatula was used to carefully dislodge
5 the formed HistoGel embedded cell mixtures, and these were carefully placed into nylon mesh bags
6 (ThermoFisher Scientific Catalog No. 6774010) to prevent disaggregation during subsequent tissue
7 processing. Each cell mixture block was 2.67 mm thick with a square cross section of 225 mm². These
8 formed HistoGel cell mixtures in nylon bags were placed into individual tissue processing cassettes
9 (ThermoFisher Scientific Catalog No. 1000957), and then were submerged in a plastic pail filled with 10%
10 NBF to simulate pre-tissue-processing time-in-formalin delay before batch tissue processing steps.

11

12 The sequence described above was performed at 1-, 2-, 6-, and 24-hour time points prior to batch tissue
13 processing. All cassettes were then placed into a tissue processor for a “routine” tissue processing run at
14 the University of Toledo Medical Center Department of Pathology (Sakura Tissue Tek VIP 5 Tissue
15 Processor). The processor with 14 stations was programmed as follows: 1) 10% NBF for 1 hour, 2) 10%
16 NBF for 1 hour, 3) 70% ethanol for 1 hour, 4) 80% ethanol for 1 hour, 5) 95% ethanol for 45 minutes, 6) 95%
17 ethanol for 45 minutes, 7) 100% ethanol for 45 minutes, 8) 100% ethanol for 45 minutes, 9) xylene for 45
18 minutes, 10) xylene for 45 minutes, 11) paraffin at 60 °C for 30 minutes, 12) paraffin at 60 °C for 30
19 minutes, 13) paraffin at 60 °C for 30 minutes, and 14) paraffin at 60 °C for 0 minutes. The processed
20 formalin fixed paraffin infiltrated cell blocks were then embedded in paraffin (Sakura Tissue Tek TEC 5
21 Tissue Embedding Station) to create formalin-fixed paraffin embedded (FFPE) cell blocks.

22

1 Immunohistochemistry of FFPE materials was prepared using Pan Keratin (Ventana Catalog No. 760-2135)
2 and CONFIRM-anti-CD45 (Ventana Catalog No. 760-2505) reagents using a Benchmark Ultra Ventana
3 Automated IHC slide staining system. These two IHC stains were used to ensure purity of cell cultures
4 after tissue processing.

5

6 Each FFPE cell block was serially sectioned at 5 μ m thickness with a microtome. Smaller groups of 8
7 sections ("curl samples") were placed into individual low-binding Eppendorf tubes for intra-block
8 comparison of FFPE sampling variation. For cell count and quality control purposes, one alternating
9 section was taken for routine hematoxylin and eosin (H&E) staining. Alternating sets of one H&E slide and
10 8 sections of FFPE material ("curl sets") were sectioned until the block was exhausted. The relative
11 position "number" was recorded for each H&E slide and used to mark the corresponding "curl-set" tube.
12 To estimate the cellularity for each tube containing formalin fixed and paraffin embedded materials, the
13 average cell count was first computed for the two flanking H&E slides and multiplied by four as a human
14 lymphoblast cell (10-20 μ m in size) would likely appear in two adjacent sections of 5 μ m thickness. Most
15 cell count estimates ranged from approximately 10,000 to 40,000.

16

17 Samples with obviously low counts at the top end of each block were excluded prior to the first recording
18 of relative position. Based on the relative positions and cell count estimates, samples were grouped into
19 three categories: surface (top or bottom) samples with cell count estimates below 50% of the average
20 count per sample, the adjacent surface samples with similar cell counts, and the inner samples. Most
21 inner samples showed cell counts above 20,000. Three adjacent samples were assigned to a second
22 surface category at each end of an FFPE cell block. Occasionally one or two more adjacent samples were

1 assigned to the second surface category if their cell counts were much closer to the surface samples than
2 the inner samples. Samples of the first surface category usually showed cell counts below 10,000. Cell
3 counts for all available samples and their categories are provided in **Supplementary File 1**. All samples
4 were coded by a concatenated string of three fields separated by an underscore “_”: up to two digits for
5 formalin fixation time in hours, a letter in uppercase for FFPE block (“G” or “H”), and up to two digits for
6 FFPE block position. For each formalin fixation time, two FFPE blocks were used, and three samples were
7 taken evenly from each block. A set of 24 FFPE curl samples were then distributed to each testing
8 laboratory.

10 **Cross-panel targeted NGS testing of FFPE samples**

11 Four laboratories participated in this study and each tested one distinct oncopanel with support from the
12 panel provider. These four panels were AstraZeneca 650 genes Oncology Research Panel (AZ650), Burning
13 Rock DX OncoScreen Plus (BRP), Illumina TruSight Tumor 170 (ILM), and Thermo Fisher Oncomine
14 Comprehensive Assay v3 (TFS). Each laboratory extracted and quantified genomic DNA from the FFPE
15 sections. NGS sequencing experiments were conducted following vendor recommended protocols, with
16 QC data collected as well. Sequencing data was then processed by the respective oncopanel vendor
17 recommended bioinformatics pipeline. Detailed information regarding targeted NGS experiment and
18 variant calling method is provided below for each panel. Variant calling results and QC data were
19 collected and submitted to the Working Group for integrated analysis.

21 **AZ650 panel NGS testing and variant calling**

1 The AZ650 assay is a hybrid capture panel designed by AstraZeneca to perform next-generation
2 sequencing on solid tumors for exploratory evaluation of pan-cancer biomarkers. AZ650 was designed
3 with reference sequences from human genome HG38. DNA Extraction was performed on FFPE tissue
4 using the Omega M6958 Kit performed on the KingFisher Flex instrument. Extracted DNA was quantitated
5 using the Qubit dsDNA High Sensitivity Kit (cat # Q32854). Each sample was quantitated in duplicate 2 μ L
6 reactions, and the average was calculated as the final DNA concentration (ng/ μ L).

7

8 DNA whole genome libraries were constructed using the Kapa Biosystems HyperPlus kit (cat #
9 07962428001, Roche) onboard the Beckman Coulter Biomek FxP liquid handling platform with an
10 integrated on-deck Biometra TRobot thermal cycler. A DNA aliquot was normalized for each sample in 10
11 mM TRIS-HCl buffer. Enzymatic fragmentation was performed to shear DNA prior to adapter ligation.
12 Unique dual-indexed adapters containing a 6 base-pair UMI sequence were ligated to the fragmented
13 DNA. The DNA whole genome libraries were quantitated using the Agilent TapeStation D1000 (cat #
14 5067-5582, Agilent). Quantitation values and fragment lengths sourced from the TapeStation D1000 were
15 used for quality control prior to hybridization capture reaction.

16 Hybridization capture was performed to enrich for the regions of the genome that comprise the targeted
17 panel. Prior to hybrid capture, whole genome libraries were multiplexed together in equimolar ratios, and
18 concentrated using a SPRI bead method. The hybridization capture protocol was performed manually
19 using probes produced by IDT and the Roche NimbleGen SeqCap Hybridization and Wash Kit. Hybrid
20 capture libraries were quantitated using both the Agilent TapeStation D1000 ScreenTape and the Kapa
21 Biosystems Library Quantification kit (qPCR).

22

1 Sequencing of each hybrid capture pool was performed on either the Illumina HiSeq 4000 or NovaSeq
2 6000 sequencers. Each pool was normalized to 1 nM and quantitated via TapeStation D1000, then diluted
3 to a final concentration of 200 pM prior to flowcell loading. Sequencing Analysis Viewer (SAV) and the
4 MultiQC tools²⁷ were used to review the quality metrics generated from the sequencer. Sequencing data
5 was demultiplexed, passed through a bcl-to-fastq conversion program²⁸ (bcl2fastq v2.20.0.422). FASTQ
6 files were analyzed using pipeline software bcbio-nextgen²⁹. Reads were aligned to the hg38 reference
7 using bwa mem v0.7.17³⁰, and sequencing duplicates for each UMI were collapsed into a single consensus
8 read using fgbio³¹ v1.0.0. Variant calling was performed using VarDict v1.7.0³², down to a variant allele
9 frequency (VAF) of 1% (before filtering and curation) and variant effects annotated by snpEff v4.3.1t³³. All
10 software was run using best practice parameters established within the bcbio workflow or in-house.
11 Mapped UMI consensus reads (in BAM files) and variant calling results (in VCF files) were then provided
12 to the working group for further analysis. The following variant filters were recommended by the panel
13 provider to minimize false positive calls: 1) a total depth threshold of 100; 2) at least four forward, four
14 reverse, and ten total support reads for the alternative allele; 3) VAF threshold of 2%; and 4) mean
15 position in support reads (pMEAN)³² greater than 15.

16

17 **BRP panel NGS testing and variant calling**

18 The DNA from 24 individual formalin fixed cell pellets in paraffin scrolls were extracted using the AllPrep
19 DNA/RNA FFPE kit (Qiagen) following the manufacturer's genomic DNA purification protocol. The
20 extraction process involved deparaffinization, protease digestion, DNA-containing pellet separation,
21 second protease digestion, de-crosslinking, column binding, washing, and elution. After purification, the
22 DNA concentration was quantified using a Qubit Fluorometer with dsDNA HS assay kit (Life Technologies,

1 Carlsbad, CA). The library prep and enrichment process were performed using a Burning Rock HS library
2 preparation kit. The procedure was described previously¹. In brief, DNA shearing was performed on each
3 FFPE DNA samples using a Covaris M220 for 240s, with Peak Incident Power=50W, Duty Factor: 20%, Cycle
4 Per Burst: 200, at 2-8 °C, followed by end repair, adaptor ligation, and PCR enrichment. Approximately
5 750 ng of purified pre-enrichment library was hybridized to the OncoScreenPlus panel and further
6 enriched following manufacturer instructions. The OncoScreenPlus panel is approximately 1.7M bp in size
7 and covers 520 human cancer related genes. Final DNA libraries were quantified using a Qubit
8 Fluorometer with dsDNA HS assay kit (Life Technologies, Carlsbad, CA). Agilent 4200 TapeStation D1000
9 Screen Tape was then performed to assess the quality and size distribution of the library. The libraries
10 were sequenced on a NovaSeq 6000 instrument (Illumina, Inc., San Diego, CA) with 2 × 150 bp pair-end
11 reads with unique dual index.

12

13 After demultiplexing using bcl2fastq v2.20²⁸ (Illumina), sequence data were filtered using the
14 Trimmomatic 0.36³⁴ with parameters “TRAILING:20 SLIDINGWINDOW:30:25 MINLEN:50”.
15 Sequence data in fastq format were mapped to the human genome (hg19) using BWA aligner 0.7.10³⁰.
16 Local alignment optimization, variant calling and annotation were performed using GATK v3.2.2³⁵ with
17 parameters “--interval_padding 100 -known 1000G_phase1.indels.b37.vcf -known
18 Mills_and_1000G_gold_standard.indels.b37.vcf” and VarScan v2.4.3³⁶ with parameters “-min-coverage 50
19 --min-var-freq 0.005 --min-reads2 5 --output-vcf 1 --strand-filter 0 --variants 1 --p-value 0.2”. For SNV
20 and small indels, variants were further filtered using an in-house variant filter pipeline. For each valid
21 variant, the covered raw depth was required to be greater or equal than 50 (DP>=50), and at least 5
22 mutation supporting count (AD>=5); minor allele frequency was required to be greater than 0.01

1 (AF>=0.01). In order to further filter out false positives, only variants with at least 6 unique fragments
2 support or 2 unique paired fragment support, i.e., within overlapping region between read pairs, were
3 kept. After filtering, remaining valid variants were annotated with ANNOVAR 20160201³⁷ and SnpEff
4 v3.6³³.

5

6 **ILM panel NGS testing and variant calling**

7 FFPE curls were de-paraffinized with xylene followed by an ethanol wash. Briefly, 1 mL xylene was added
8 to each tube with the FFPE sample, vortexed vigorously for 10 seconds, and centrifuged at 20,000 x *g* for
9 2 minutes. The supernatant was carefully removed. Then 1 mL 96-100% ethanol was added to the pellet,
10 mixed by vortexing, and centrifuged at 20,000 x *g* for 2 minutes. Ethanol was removed and the pellet was
11 air dried. DNA extraction from the pellet was performed using the QIAGEN Allprep DNA/RNA FFPE kit
12 (QIAGEN) on a QIAcube. DNA concentrations were determined by fluorometric quantitation using a Qubit
13 2.0 Fluorimeter with a Qubit DNA dsDNA HS Assay Kit (Thermo Fisher Scientific).

14

15 Library preparation was carried out using the TruSight Tumor 170 Assay (Illumina) following
16 manufacturer's instructions, except with a lower amount of DNA. Briefly, 30 ng DNA from each sample,
17 except 24G15, 24H10, and 24H15 with 25, 21, and 8 ng DNA respectively, were fragmented on a Covaris
18 Ultrasonicator (Covaris) using the following setting: peak incident power 50 watts, duty factor 30%, cycles
19 per burst 1000, treatment time 270 seconds, and temperature 20 °C. The fragmented DNA was processed
20 through end repair, A-tailing, adapter ligation, and index PCR, enriched by the hybridization-capture
21 method, amplified by final PCR, normalized by bead based normalization, and pooled for sequencing.
22 Twenty-four samples were batched for each library prep, and libraries from every 8 samples were pooled

1 and sequenced on the Illumina NextSeq 550 instrument using the NextSeq High Output Kit. BaseSpace
2 Sequence Hub was used to set up the sequencing run, perform the initial quality control, and generate
3 fastq files for each sample.

4
5 Variant calling was performed in BaseSpace Sequence Hub. Briefly, high level sequencing run metrics
6 were evaluated to generate a Run QC Metrics report. Next reads were converted into the fastq format
7 using bcl2fastq²⁸, adapters were trimmed and then reads were aligned to the human genome version
8 hg19 using the iSAAC aligner³⁸. Indel realignment was performed and then candidate variants were
9 identified using the Pisces variant caller³⁹, with a fixed lower limit cutoff for variant allele fraction of at
10 least 2.6%. Variant calls were further compared against a baseline of normal samples to remove
11 systematic false positives.

12

13 **TFS panel NGS testing and variant calling**

14 Genomic DNA was extracted from 24 samples of sectioned material using the ALLPrep DNA/RNA FFPE kit
15 (cat # 80234). Samples were eluted in 15 µL from which 1 µL of material was used for quantification.

16 Extracted material was prepared for quantification with a Qubit™ dsDNA HS (High Sensitivity) Assay Kit
17 using 1 µL of sample material in 200 µL of Qubit solution. Concentration readings in (ng/µL) from the
18 Qubit Fluorometer were used to calculate 20 ng DNA input in a maximum volume of 7.5 µL into the
19 library preparation.

20

21 Libraries were generated using the Ion AmpliSeq Oncomine Comprehensive panel versions 3.0 from
22 Thermofisher Scientific as described⁴⁰ and 17 amplification cycles were performed as suggested for FFPE

1 samples. Final library quantification was performed using real-time PCR (QuantStudio) and the values
2 were given by the instrument in pM. Fifteen out of 24 libraries passed the QC threshold of 50 pM and
3 were successfully sequenced. Five out of 24 libraries did not pass the QC threshold of 50 pM (libraries <
4 50pM). These samples were cleaned using 1.8 X Ampure beads, eluted in 15 µL then Qubit quantified
5 using 1 µL from the cleaned material. The input DNA amount to be used into library preparation was
6 determined using the samples' post clean-up Qubit values multiplied by 7.5, which is the maximum
7 volume of input material for library preparation. The newly constructed libraries passed the QC threshold
8 of 50 pM and were successfully sequenced.

9
10 The remaining 4 out of 24 libraries did not pass the QC threshold of 50 pM (libraries < 50pM). These
11 samples were cleaned using 1.8 X Ampure beads, eluted in 15 µL then Qubit quantified using 1 µL from
12 the cleaned material. Qubit values were too low to be detected, so 7.5 µL was used as input for library
13 preparation. The newly constructed libraries did not pass the QC threshold of 50 pM and therefore were
14 not sequenced (libraries < 50pM).

15
16 All barcoded samples were sequenced on the Ion S5 XL System. Analyses of sequencing raw data were
17 performed with the Ion Torrent Suite (5.10.0). After base calling, reads were aligned with the TMAP
18 module and variant calling was performed with Torrent Variant Caller (TVC), a variant calling module
19 optimized for Ion Torrent data. The default thresholds used for SNVs and indels was 2.5%. Finally, a series
20 of post-calling filters were applied to variant calls to eliminate potential artifacts by fitting the statistical
21 model of flow signals for the observed reference and non-reference alleles. For more details, please see

1 the description of its bioinformatics pipeline in the cross-laboratory oncopanel study¹ where this panel
2 was used to test a set of reference samples including Agilent male DNA control.

3

4 **Identification and Analysis of false positive calls**

5 *Consensus high confidence targeted region (CTR)*

6 In our collective consortium effort to establish a verified genomic reference material¹⁸ suitable for
7 assessing the performance of oncology panels in detecting small variants of low allele frequency, a
8 consensus high confidence region for targeted NGS was adopted and later tested in a cross-laboratory
9 oncopanel study¹ of eight oncology panels where the CTR was shown to reduce the rate of false positive
10 variant calls. The CTR resulted from intersecting exonic coding regions, the NIST high confidence region,
11 and the targeted regions of four whole-exome sequencing panels. Low complexity regions were then
12 removed from the CTR.

13

14 *VCF files cleaning procedures*

15 VCF files provided by each panel vendor were cleaned through a series of procedures. First, each original
16 VCF file was converted into standardized VCF format. INDELs were then normalized with left alignment
17 and trimming using GATK³⁵. Complex variants were decomposed with RTG “vcfdecompose”. After that,
18 only variants within the panel regions (excluding the low complexity regions) were kept. Specifically, for
19 VCF files from the TFS panel, we also removed the blacklist variants provided by Thermo Fisher Scientific.
20 To further remove the less confident variants, VAF thresholds recommended by the panel providers were
21 applied for each panel: 2% for AZ650, 1% for BRP, 2.6% for ILM, and 2.5% for TFS. All downstream
22 analyses were applied to the cleaned VCF files.

1

2 *Generation of known variant set for each panel*

3 The known variant set was generated from the fresh DNA samples of each panel. Any variants called in
4 over 75% of all samples were each considered as a known variant. However, due to only three fresh DNA
5 samples that were sequenced for AZ650 panel, instead of using the fresh DNA samples, we adopted high
6 quality FFPE samples with deep sequencing depth (after excluding two contaminated samples) for
7 generation of the known variant set. For the ILM panel, we excluded 6 fresh DNA samples for their low
8 median depth (less than 850). Finally, 20 FFPE samples from AZ650, 12 fresh DNA samples from BRP, 10
9 fresh DNA samples from ILM, and 12 fresh DNA samples from TFS were used for the known variant set
10 generation for each panel.

11

12 *The false positive estimation and type classification*

13 After the cleaning procedures, any variants called by a panel that were not in the known variant set for
14 that panel were determined as false positive calls. Two metrics were used in the study. The false positive
15 call count was the number of false positive variants for each library, and the false positive rate was then
16 estimated as the ratio of false positive calls out of every million positions defined by each panel. Besides
17 the false positive estimation from the originally cleaned library, false positives could be further reduced
18 by applying more stringent cutoffs. Two types of additional filters were adopted in this study, VAF cutoffs
19 from the panel default VAF threshold to 10% and alternative allele depth cutoffs from 0 to 30. For analysis
20 purposes, besides evaluating the impact of various factors on the FP rate, we also examined the effect per
21 variant type by grouping variant calls into four types: INDELs, G:C to A:T transitions, G:C to T:A
22 transversions, and any other variants.

1

2 **Availability of data and materials**

3 DNA-seq data (in FASTQ or BAM format) from four oncopanels will be deposited in the National Center for
4 Biotechnology Information (NCBI) BioProject repository. Variant call results (in VCF format) and
5 dependent files (e.g., panel BED files, genome FASTA files, and others) will be made publicly available
6 through FigShare.

7

8 **Acknowledgements**

9 All SEQC2 participants freely donated their time, reagents, and computing resources for the completion
10 and analysis of this project. Part of this work was carried out with the support to Thomas Blomquist
11 through the University of Toledo College of Medicine Academic Affiliation with ProMedica research funds.
12 Leming Shi was supported by the National Key R&D Project of China (2018YFE0201600), the National
13 Natural Science Foundation of China (31720103909), and Shanghai Municipal Science and Technology
14 Major Project (2017SHZDZX01). Donald J. Johann, Jr. acknowledges the support by FDA BAA grant
15 HHSF223201510172C. Nikola Tom was supported by research infrastructure EATRIS-CZ, ID number
16 LM2018133 funded by MEYS CR and MEYS CR project CEITEC 2020 (LQ1601).

17

18 The contents of these published materials are solely the responsibility of the administering institution, a
19 participating institution, or individual authors, and they do not reflect the views of any funding body
20 listed above.

21

22 **Disclaimer**

1 The views presented in this article do not necessarily reflect those of the U.S. Food and Drug
2 Administration. Any mention of commercial products is for clarification and is not intended as an
3 endorsement.

4

5

6 **References**

7

- 8 1. SEQC2 OncoPanel Sequencing Working Group. Cross-oncopanel study reveals high sensitivity and
9 accuracy with overall analytical performance depending on genomic regions. *Accept. Genome Biol.*
10 *GBIO--20-01201*.
- 11 2. Jennings, L. J. *et al.* Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels:
12 A Joint Consensus Recommendation of the Association for Molecular Pathology and College of
13 American Pathologists. *J. Mol. Diagn.* **19**, 341–365 (2017).
- 14 3. Agrawal, L., Engel, K. B., Greytak, S. R. & Moore, H. M. Understanding Preanalytical Variables and
15 their Effects on Clinical Biomarkers of Oncology and Immunotherapy. *Semin. Cancer Biol.* **52**, 26–38
16 (2018).
- 17 4. Compton, C. C. *et al.* Preanalytics and Precision Pathology: Pathology Practices to Ensure Molecular
18 Integrity of Cancer Patient Biospecimens for Precision Medicine. *Arch. Pathol. Lab. Med.* **143**,
19 1346–1363 (2019).
- 20 5. *AJCC Cancer Staging Manual, 8th edition.* (Springer International Publishing, 2017).
- 21 6. Wolff, A. C. *et al.* Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast
22 Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice

- 1 17. Aziz, N. *et al.* College of American Pathologists' Laboratory Standards for Next-Generation
2 Sequencing Clinical Tests. *Arch. Pathol. Lab. Med.* **139**, 481–493 (2015).
- 3 18. Jones, W. D. & SEQC2 WG2 group. A Validated Genomic Reference Material for Assessing
4 Performance of Cancer Panels Detecting Small Variants of Low Allele Frequency. *Genome Biol.*
5 *Submiss. ID GBIO--20-01193*.
- 6 19. Zook, J. M. *et al.* Integrating human sequence data sets provides a resource of benchmark SNP and
7 indel genotype calls. *Nat. Biotechnol.* **32**, 246–251 (2014).
- 8 20. Fang, L. T. & SEQC2 WG1 working group. Establishing reference samples for benchmarking detection
9 of somatic mutations germline and variants with NGS technologies. *Nat. Biotechnol.* *NBT-RS47789*
10 *Rev.*
- 11 21. Van Ooyen, S., Loeffert, D. & Korfhage, C. Overcoming constraints of genomic DNA isolated from
12 paraffin-embedded tissue. *Qiagen White Pap.* 1–6.
- 13 22. Spencer, D. H. *et al.* Comparison of Clinical Targeted Next-Generation Sequence Data from
14 Formalin-Fixed and Fresh-Frozen Tissue Specimens. *J. Mol. Diagn.* **15**, 623–633 (2013).
- 15 23. Sekiguchi, M. & Tsuzuki, T. Oxidative nucleotide damage: consequences and prevention. *Oncogene* **21**,
16 8895–8904 (2002).
- 17 24. Do, H. & Dobrovic, A. Sequence Artifacts in DNA from Formalin-Fixed Tissues: Causes and Strategies
18 for Minimization. *Clin. Chem.* **61**, 64–71 (2015).
- 19 25. Prentice, L. M. *et al.* Formalin fixation increases deamination mutation signature but should not lead
20 to false positive mutations in clinical practice. *PLOS ONE* **13**, e0196434 (2018).
- 21 26. Gown, A. M. Current issues in ER and HER2 testing by IHC in breast cancer. *Mod. Pathol.* **21**, S8–S15
22 (2008).

- 1 27. Ewels, P. MultiQC: Summarize results from bioinformatics analysis across many samples into a single
2 report. *Bioinformatics* **32**(19), 3047–8 (2016).
- 3 28. Illumina. bcl2fastq2 Conversion Software v2.20.
4 <https://support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html>.
- 5 29. Brad Chapman *et al.* *bcbio/bcbio-nextgen: v1.2.4*. (Zenodo, 2020). doi:10.5281/zenodo.4041990.
- 6 30. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
7 *ArXiv13033997 Q-Bio* (2013).
- 8 31. fulcrumgenomics, fulcrumgenomics, & fulcrumgenomics. *fgbio*. (Fulcrum Genomics, 2020).
- 9 32. Lai, Z. *et al.* VarDict: a novel and versatile variant caller for next-generation sequencing in cancer
10 research. *Nucleic Acids Res.* **44**, e108 (2016).
- 11 33. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide
12 polymorphisms, SnpEff. *Fly (Austin)* **6**, 80–92 (2012).
- 13 34. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data.
14 *Bioinformatics* **30**, 2114–2120 (2014).
- 15 35. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing
16 next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- 17 36. Koboldt, D. C. *et al.* VarScan 2: Somatic mutation and copy number alteration discovery in cancer by
18 exome sequencing. *Genome Res.* **22**, 568–576 (2012).
- 19 37. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from
20 high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
- 21 38. Racz, C. *et al.* Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms.
22 *Bioinformatics* **29**, 2041–2043 (2013).

- 1 39. Dunn, T. *et al.* Pisces: an accurate and versatile variant caller for somatic and germline
2 next-generation sequencing data. *Bioinformatics* **35**, 1579–1581 (2019).
- 3 40. Pain, M. *et al.* Treatment-associated TP53 DNA-binding domain missense mutations in the
4 pathogenesis of secondary gliosarcoma. *Oncotarget* **9**, 2603–2621 (2017).

5

6 Figures

7 **Figure 1 | Overview of study design.** (A) FFPE sample preparation workflow with four different formalin
8 fixation times: 1-, 2-, 6-, and 24 hours. (B) OncoPanel sequencing experiments with in-laboratory DNA
9 extraction. (C) Panel-specific variant calling followed by uniform and integrated analysis to assess the
10 impacts of genomic region, VAF threshold, depth, formalin fixation time, and sample position in the FFPE
11 block.

12

13 **Figure 2 | Histogram and pie chart distributions of known variants across VAF ranges confirming that**
14 **most were homozygous or heterozygous germline variants.** (A) Distribution of the number of known
15 variants within (green) and outside (blue) of the consensus high confidence targeted region (CTR). (B)
16 Number and percentage of known variants across four VAF ranges.

17

18 **Figure 3 | Illustration of complementary QC checks for identifying bad experiments and outliers.** QC
19 checks can be done during three major processes: pre-analytical, oncopanel NGS, and post variant calling.
20 Multiple measurements can be used to evaluate the quality of the experiments and the data. Here we
21 showcase a few measurements which are important in our study. (A) Experiments (circled with a red
22 dashed line) with lower cell count (x-axis) usually showed a higher false positive rate (FPR, y-axis). Some

1 experiments (circled with an orange dashed line) with cell counts (>20K) and a higher FPR failed one or
 2 more subsequent QC checks. (B) Experiments with low DNA input (x-axis) show high FPR (y-axis). (C)
 3 Lower sequencing depth (which can be measured by median depth, total depth) tends to result in higher
 4 FPR. Regions out of CTR are more affected. (D) Experiments were considered to have failed if library
 5 complexity was too low (below 0.25 as observed in our data). (E) G:C to T:A transversion count could be
 6 used to identify outliers for further investigation. (F) In this VAF histogram, a long tail from the left side of
 7 100% VAF was observed in sample 6_G_7, which indicates potential contamination.

8

9 **Figure 4 | Similar false positive rates (per million base) were achieved by QC-passed inner FFPE samples**
 10 **in comparison to fresh DNA samples.** (A) Violin plots of the false positive rate for fresh DNA samples
 11 versus QC-passed inner FFPE samples in three panel regions (whole panel, within the CTR, or outside of
 12 the CTR), asterisk symbols represent the significance level of the true difference in means is less than 0 (*:
 13 $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). (B) Average false positive rate in different panel regions for fresh DNA
 14 samples after applying various additional VAF and alternative allele depth cutoffs. (C) Average false
 15 positive rate in different panel regions for QC-passed FFPE samples after applying various additional VAF
 16 and alternative allele depth cutoffs.

17

18 **Figure 5 | Counts and distribution of false positive calls by variant types within the consensus targeted**
 19 **region (CTR) indicated more FFPE damage in surface FFPE samples.** (A) The average number of false
 20 positive calls is plotted with standard error of the mean (SEM) for four variant types within the CTR for
 21 the fresh DNA and various FFPE sample groups. The variant types were 1) indels, 2) hydrolytic
 22 deamination introduced artifacts (G:C>A:T transitions), 3) oxidative damage artifacts (G:C>T:A

1 transversions), and 4) other FP calls. The number of samples for each sample group is inserted at the top
 2 right corner of each subplot. (B) Number and percentage of known variants within the CTR are shown in a
 3 pie chart over four variant types for each panel. (C) Average number and percentage of false positive calls
 4 per sample within the CTR are shown in a pie chart over four variant types for each panel and two sample
 5 groups (QC-passed inner FFPE samples vs surface FFPE samples).

6

7 **Figure 6 | Longer formalin fixation time had no effect on the quality of inner FFPE samples but reduced**
 8 **their quantity. (A-D)** For each panel, QC-passed inner FFPE samples are plotted by formalin fixation time
 9 (x-axis). Each circle represents a sample with its false positive rate per million bp shown on the y-axis.
 10 Samples are color coded per panel: red for AZ650, green for BRP, blue for ILM, and black for TFS. Except
 11 for TFS, there was no observable effect of formalin fixation time on the quality (measured by false
 12 positive rate) of inner FFPE samples. (E) The average cell count of inner FFPE samples remained consistent
 13 across different formalin fixation times. Two FFPE blocks were prepared under each formalin fixation time.
 14 Red circles denote the average cell counts of QC-passed inner FFPE samples taken from each block while
 15 block crosses denote the average cell counts of all inner FFPE samples within each block. These two
 16 values are shown to be very close to each other for each block. Finally, the red dots and dashed line show
 17 that the overall average cell counts are consistent across four formalin fixation times. (F) The quantity of
 18 inner FFPE samples is plotted for each FFPE block. The number of inner samples in each block is shown on
 19 the left y-axis while the aggregated thickness of all inner samples (i.e., thickness of inner sample portion)
 20 is shown on the right y-axis. Longer formalin fixation time clearly reduced the quantity of inner FFPE
 21 samples.

22

1 List of Supplementary Materials

2 **Supplementary File 1.xlsx | List of FFPE sample information and extensive quality check data**

3 **Supplementary File 2.xlsx | List of known variants for four participating oncology panels**

4 **Supplementary Table 1 | List of detailed information for four participating oncology panels**

5 **Supplementary Figure 1 | Plot of false positive rate for QC-passed inner FFPE samples showed no**

6 **observable effect of formalin fixation time.** QC-passed inner FFPE samples were aggregated from all four

7 panels after Z-score normalization of false positive rate within the sample group per panel. Within each

8 panel's sample group, the mean and standard deviation of the false positive rate were first calculated.

9 Z-scored false positive rate was calculated by dividing its difference from the mean by the standard

10 deviation. This normalization step removed the differences in false positive rate between panels. Samples

11 were then grouped by formalin fixation time (x-axis) and plotted together. Each circle represents a sample

12 with its normalized false positive rate shown on the y-axis. Samples are color coded per panel: red for

13 AZ650, green for BRP, blue for ILM, and black for TFS. There was no observable effect of formalin fixation

14 time on false positive rate.

15

Figure 1

A **B**

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6