Robustness of a Cancer Profiling Test Using Formalin-fixed Paraffin Embedded Tumor Specimens

SHIN ITO[1], IKURO SATO[2], MAI MOCHIZUKI[3], KAZUNORI YAMAGUCHI[1], KEIICHI TAMAI[3],
TAKAMICHI MINATO[4], NOBUHIRO TANUMA[5], HIROSHI SHIMA[5] and JUN YASUDA[1]

[1]Division of Molecular and Cellular Oncology, Miyagi Cancer Center Research Institute, Miyagi, Japan;
[2]Division of Pathology, Miyagi Cancer Center, Miyagi, Japan;
[3]Division of Cancer Stem Cell, Miyagi Cancer Center Research Institute, Miyagi, Japan
[4]Division of Gynecology, Miyagi Cancer Center, Miyagi, Japan;
[5]Division of Cancer Chemotherapy, Miyagi Cancer Center Research Institute, Miyagi, Japan

Abstract. *Background/Aim: Cancer profiling tests using formalin-fixed paraffin-embedded (FFPE) specimens with various conditions have become an essential tool for cancer treatment. The robustness of these tests needs to be addressed. Materials and Methods: A cancer profiling test, NCC oncopanel, was tested with FFPE specimens from various tissues with different storage conditions and fixation lengths. Next generation sequencing was performed with Miseq and the data were assembled using the human reference genome hg19. Results: Duration of storage and fixation affected the mapping statistics. Prolonged storage increased outward read paring and longer fixation rates caused increased singletons and unmapped reads. Conclusion: Our results indicate that a cancer profiling test with target capturing method, NCC oncopanel, shows robustness for FFPE cancer specimens with various storage conditions.*

Next generation sequencing (NGS) cancer profiling tests for surgical specimens is an essential tool for cancer treatment. In Japan, the national health insurance system recently approved the use of two cancer profiling tests, FoundationOne CDx (1) and NCC oncopanel (2), at government-certified hospitals. The costs of these cancer profiling tests are high and can only be covered once in the lifetime of a patient by the national health insurance in Japan. Therefore, the robustness of the cancer profiling test for a given specimen quality is critical.

The NCC oncopanel is a hybridization capture based NGS assay designed to examine mutations, amplifications and homozygous deletions of the entire coding regions of 114 genes of clinical or preclinical relevance along with rearrangements in 12 oncogenes. The present version of NCC oncopanel (version 4) is widely used as a cancer profiling test in the Japanese hospitals certified by the Japanese government. Version 2 of the panel is provided for research use and we can test the effect of storage of formalin-fixed paraffin-embedded (FFPE) specimen on the target capturing cancer profiling tests in a research laboratory setting.

Various clinical factors can affect the quality of FFPE specimens, from surgical procedures to the sectioning of paraffin blocks (3). For example, samples fixed in unbuffered formalin for a long period (>7 days) may not be suitable for molecular analysis (4). It is also known that formalin can induce artificial mutations *via* cytosine to thymine (C to T) mutations and that these artificial mutations can be reduced by uracil-DNA glycosylase treatment (5, 6).

It is important to understand the degree to which the output of cancer profiling tests is affected by the variations in FFPE sample quality in a clinical setting. Therefore, we investigated the mapping quality of the output of a cancer profiling test, NCC oncopanel (experimental version). We found that the NCC oncopanel showed robustness regarding the length of storage and fixation time for representative variants of the tumor tissue.

Materials and Methods

FFPE samples. We selected tissues of interest from our archive at the Miyagi Cancer Center Hospital (Table I). Surgically removed cancer tissue was treated with a 10-fold diluted 37% formaldehyde solution (069-0047; Fujifilm Wako Pure Chemical Co., Osaka, Japan). Only sample 15 was fixed with 10% neutral buffered formalin solution (060-01667; Fujifilm Wako Pure Chemical Co.). Paraffin embedding and preparation of slides were done according to standard protocols.

Correspondence to: Jun Yasuda, MD, Ph.D., Division of Molecular Carcinogenesis, Miyagi Cancer Center Research Institute, 47-1, Nodayama, Medeshima Shiode, Natori, Miyagi 981-1293, Japan. Tel: +81 223843151, e-mail: jun-ysauda@miyagi-pho.jp

Key Words: Cancer profiling test, formalin-fixed paraffin-embedded samples, next generation sequencing.

Table I. *Summary of samples*.

Sample ID	Tissue of origin	Storage (years)	Fixation (days)	$\Delta\Delta Cq$*
Sample 1	Prostate	3	1	1.387
Sample 2	Pancreas	3	1	0.928
Sample 3	Soft tissue	3	1	2.327
Sample 4	Prostate	1	1	0.748
Sample 5	Pancreas	1	1	1.089
Sample 6	Soft tissue	1	1	0.792
Sample 7	Uterus	3	1	0.884
Sample 8	Lung	3	1	1.166
Sample 9	Uterus	1	1	0.436
Sample 10	Lung	1	1	0.201
Sample 11	Prostate	5	7	1.665
Sample 12	Prostate	1	14	1.35
Sample 13	Ovary (serous adenoca.)	1	3	0.789
Sample 14	Ovary (Clear cell ca.)	1	4	0.441
Sample 15	Tongue	1	6	0.622
Panc-1	Cell Line	0	0	–0.615

*Based on the estimates by qubit data.

Table II. *Correlation and coefficient of variables among duration of storage or fixation of samples with mapping statistics*.

Mapping statistics	CV	R^2 Storage	R^2 Fixation	R^2 $\Delta\Delta Cq$
Singletons	0.388	0.001	0.406	0.01696
Unmapped reads	0.346	0.000000273	0.412	0.01352
Non-primary alignments	0.318	0.520	0.027	0.676
Outward oriented pairs	0.227	0.457	0.00969	0.592
Pairs on different chromosomes	0.210	0.229	0.196	0.260
Inward oriented pairs	0.014	0.512	0.00141	0.637
Reads properly paired	0.003	0.144	0.262	0.137
Bases mapped (cigar)	0.002	0.500	0.147	0.509
Reads mapped and paired	0.002	0.000301	0.410	0.015

This study was approved by the in-house ethical committee of the Miyagi Cancer Center (registration number 2018-069).

DNA extraction from FFPE specimens. Tissues were cut from paraffin blocks (5 μM thick, 10 pieces) and DNA was extracted with the Maxwell RSC DNA FFPE Kit - PKK, Custom (AX2500; Promega Corporation, Madison, WI, USA) and Maxwell RSC Instrument (Promega Corporation), according to the manufacturer's instructions.

Cell culture and cellular DNA extraction. PANC-1 human pancreatic carcinoma cells were cultured in RPMI-1640 medium (189-02025; Fujifilm Wako Pure Chemical Co.) containing 9% fetal bovine serum (SFBM30; Equitech-Bio, Kerrville, TX, USA). Cells were incubated at 37°C with 5% CO_2 and collected for DNA extraction at 60% confluence in a 10-cm dish. DNA was extracted from cells using a DNeasy Blood & Tissue Kit (69504; Qiagen Gmbh, Hilden, Germany) according to the manufacturer's instructions.

Measurement of extracted DNA concentration and quality control. A Qubit dsDNA HS Assay Kit (Q32851; Thermo Fisher Scientific, Waltham, MA, USA) and a Qubit Fluorometer (Thermo Fisher Scientific) was used to measure the concentration of DNA extracted from surgical specimens and cultured cells. We also used the Agilent NGS FFPE QC Kit (G9700A; Agilent Technologies, Santa Clara, CA, USA) and LightCycler 480 System (Roche Diagnostics, Indianapolis, IN, USA) for quantitative polymerase chain reaction (qPCR) assessment of DNA concentration and sample DNA fragmentation. qPCR was performed according the manufacturer's instructions. DNA concentrations of the samples with a $\Delta\Delta Cq$ value less than 1 were based on the DNA concentration measured by the Qubit.

Library preparation for NGS. 200 ng of each DNA sample was used for library preparation. DNA was fragmented with SureSelect XT HS

and XT Low Input Enzymatic Fragmentation (5191-4079; Agilent Technologies). SureSelect XT HS Reagents (G9702A, Agilent Technologies) and SureSelect NCC oncopanel (931195; Agilent Technologies) was used for the preparation of the NCC oncopanel capture library for the Illumina Next Generation sequencers. The TapeStation 2200 system (Agilent Technologies) and D1000 ScreenTape and its reagent (5067-5582 and 5067-5583, respectively) were used for pre-capture and post-capture library size checks and quantification. All processes from DNA fragmentation to library quality control were performed according to the manufacturer's instructions.

NGS sequencing. The prepared NGS libraries were diluted to 8 nM each based on the quantitative results from TapeStation. A library pool was created by mixing equal amounts of these libraries. Paired end sequencing was performed with the MiSeq Reagents Kit v2 300 Cycle (MS102-2002; Illumina, Inc., San Diego, CA, USA) and MiSeq (Illumina) in 151-cycles run.

Bioinformatics. The raw reads were trimmed of adapter sequences and low-quality bases by Trimmomatic 0.39 with default options (7). Only paired reads were mapped onto the hg19 reference genome with BWA-MEM (8) and the generated sam files were sorted and compressed with samtools (9). The mapping statistics data were obtained with samtools stat and flagstat options. Variant calling was then performed with Mutect2 in the Genome Analysis Toolkit (GATK), version 4.1.4.1. The filtering conditions were as following: false-discovery-rate=0.01, unique-alt-read-count=20, min-allele-fraction=0.1, minimum depth=100, and within the regions defined by 0471501_Padded.bed in the NCC oncopanel target region [(911 kb)+100 bp flanking sequences] (1.4 Mb total).

The annotations for the called variants were performed with Annovar (10) with a custom-made 4.7 KJPN variant dataset and COSMIC 90 (11). The 4.7KJPN variant data, consisted of whole genome sequencing variant data of more than 4,700 Japanese individuals, were downloaded from the jMorp database (https://jmorp.megabank.tohoku.ac.jp/) (12). Variant extraction was performed with bcftools (9, 13). The variants were also annotated with the COSMIC database version 90 (11).

Table III. *Cosmic mutations identified in this study.*

Chrom.	Position	Ref	Alt	Ref. Gene	AA change	ClinVar	Detected tumors
1	11188183	C	T	MTOR	A1971T	Likely_pathogenic	Prostate
2	212288934	G	A	ERBB4	R938C	UNK	Sarcoma
3	178952072	A	G	PIK3CA	M1043V	Likely_pathogenic	Sacoma
3	178952085	A	G	PIK3CA	H1047R	Pathogenic	Pancreas
4	1807896	C	T	FGFR3	T652I	UNK	Ovary
7	55259515	T	G	EGFR	L858R	drug_response	Panc-1, Prostate
10	43607221	C	G	RET	(Intronic)	UNK	Panc-1
10	123247515	T	A	FGFR2	K659M	UNK	Tongue
12	25398284	C	T	KRAS	G12D	Pathogenic	Panc-1, Lung, Tongue
12	25398284	C	A	KRAS	G12V	Pathogenic	Panc-1, Lung, Tongue
16	3779343	G	A	CREBBP	T1902M	UNK	Lung
17	7577120	C	T	TP53	R273H	Pathogenic/Likely_pathogenic	Panc-1

Verification of variants. Prrimers were designed to amplify the regions containing the 12 mutations and 4 indels selected for validation (primer sequences are available upon request). The targeted regions were amplified by PCR with KOD FX (KFX-101; TOYOBO CO., Osaka, Japan), PCRx Enhancer System (11495017; Thermo Fisher Scientific), and the corresponding primer sets. After denaturation at 94°C for 2 min, 35 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 30 s, and extension at 68°C for 30 s were performed, followed by a 7-min extension reaction at 68°C. After PCR amplification, PCR products were subjected to ExoSAP-IT Express PCR Cleanup Reagents (75001; Thermo Fisher Scientific) and processed at 37°C for 4 min and at 80°C for 1 min to remove single-strand DNA and dNTPs. Sanger sequence analysis of each PCR product was performed by Eurofins Genomics K.K. (Tokyo, Japan). Electrophoresis was performed with NuSieve GTG Agarose (50081; Lonza, Rockland, ME) and Agarose S (316-01191; Nippon Gene, Tokyo, Japan) 3:1 mixture gel for verification of large indels. The presence or absence of indels in the PCR products was verified by the differentiation of electrophoretic mobility. The molecular marker was Quick-Load 100 bp DNA Ladder (N0467; NEB, Ipswich, MA, USA).

Results

The effects of the storage and fixation lengths of FFPE samples on genomic DNA quality may vary. The initial aim of the present study was to identify "thresholds" for the cancer profiling test regarding the storage and fixation periods for FFPE samples. Table I summarizes the conditions of the samples tested in this study. The storage and fixation durations ranged from 1-5 years and 1-14 days, respectively. Quick estimates of the genomic DNA quality of FFPE samples can be undertaken by quantitative PCR. A higher $\Delta\Delta Cq$ value indicates more damaged DNA templates. The same storage and fixation times resulted in a variety of $\Delta\Delta Cq$ values among the samples of different origins, indicating that a number of clinical factors may affect the quality of genomic DNA extracted from FFPE specimens (Table I). The storage length (years) are more correlated to $\Delta\Delta Cq$ than the fixation length (days) for all samples analyzed in this study (Pearson's correlation coefficients are 0.70 and 0.26, respectively).

Prolonged fixation and storage of FFPE samples may affect the mapping quality of the NGS output. The obtained sequence reads from damaged genomic DNA showed high-quality base calls, while the mapping data of the sequence reads to the reference genome may have been affected by formalin fixation-derived artifacts. Table II summarizes the relationship between the mapping statistics and the condition of the FFPE samples. The coefficient variables (CVs) among samples normalized with the number of total sequencing reads (Table II) were used for comparison with the length of storage (years) or fixation (days) or the $\Delta\Delta Cq$ value of the samples, regardless of their origins (Table II). Some of the mapping statistics were very invariable; the CVs of "Reads properly paired", "Bases mapped (cigar)", and "Reads mapped and paired" were less than 0.005 (Table II). The CVs of "Singletons", "Unmapped reads", "Non-primary alignments", "Outward oriented pairs", and "Pairs on different chromosomes" appeared as >0.2, indicating these mapping statistics may have been affected by the prolonged storage and/or formalin fixation. These events were relatively rare and seem to be caused by either biological mutations occurring in small subclones in cancer tissue, or through storage-fixation artifacts, or both.

Interestingly, different mapping statistics were affected by the variations in storage and fixation duration. The storage length of FFPE samples affected the "Inward oriented pairs", "Non-primary alignment", and "Bases mapped (cigar)" ($R^2 \geq 0.5$) (Figure 1a and c). Meanwhile, the fixation length affected the "Unmapped reads" and "Singletons" ($R^2 \geq 0.4$) (Figure 1b and d). $\Delta\Delta Cq$ shows similar but stronger pattern of correlations with mapping parameters to the storage length (Table II).
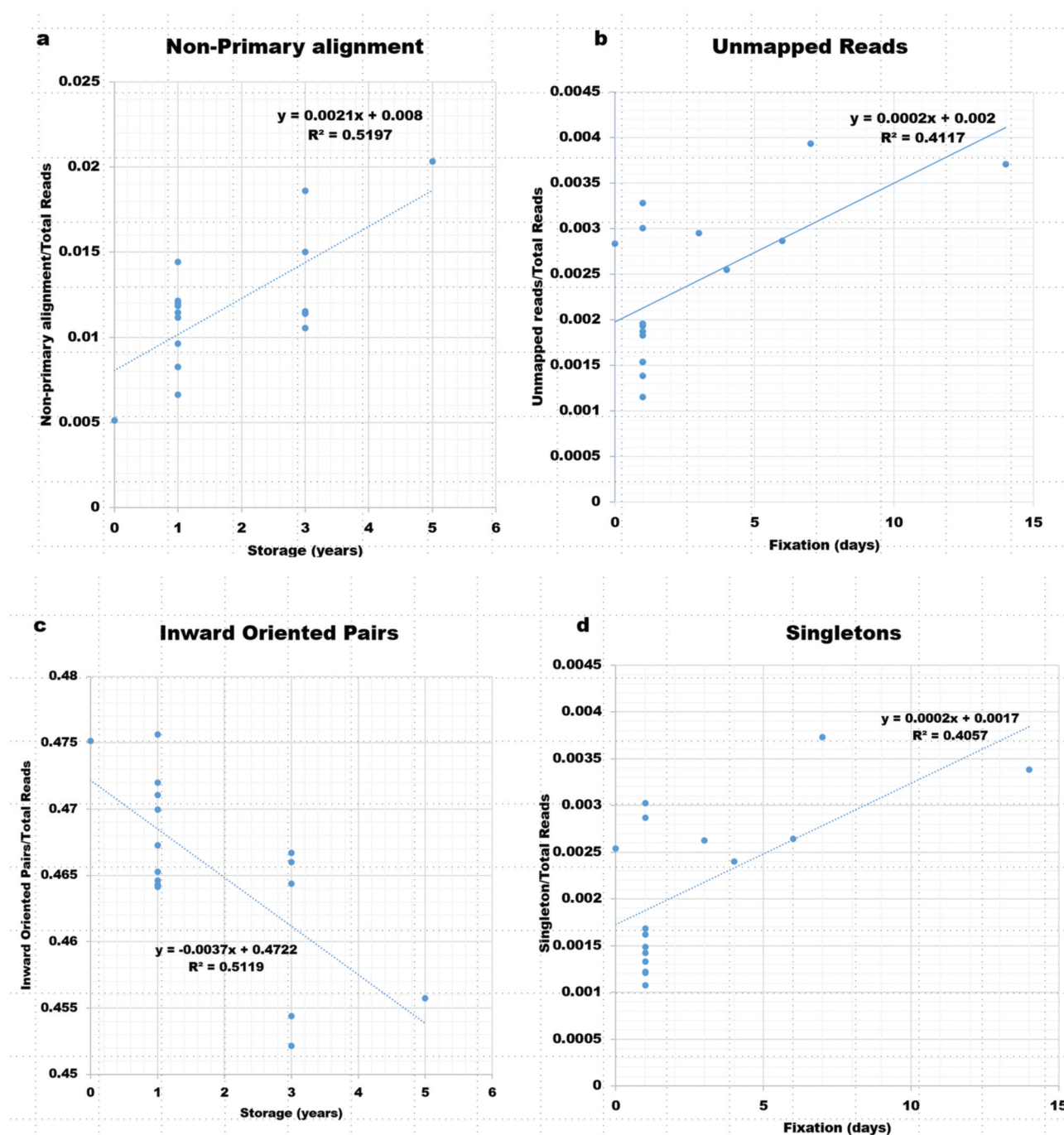
Figure 1. *Correlation between storage and fixation length and mapping statistics. (a) Correlation between storage length and ratio of non-primary mapping (i.e. duplication). Vertical axis indicates the ratio of non-primary mapping reads against the total reads. Horizontal axis indicates the storage years of the FFPE samples. Dotted line indicates the linear regression. (b) Correlation between fixation length and ratio of unmapped reads. (c) Correlation between storage length and ratio of inward-oriented pairs. (d) Correlation between fixation length and ratio of singletons.*

Storage and fixation length may not affect the detection of variants of major cancer mutations. Cancer profiling tests are mainly used to detect drug-targetable mutations that should be found in the majority of the sample analyzed.

Considering the target sizes (~1.4 MB), around 660 germline variants per individual sample would be expected. Most of the variants found in this study were annotated as positive in 4.7 KJPN (93.7% to 99.8%), suggesting that

Table IV. *Comparison between NCC oncopanel and Sanger sequencing.*

Chr	Start	End	IGV	Sanger	Cocordance	Databases*	gnomAD MNV DB**
chr1	65311116	65311117	GC/CT	GC/CT	Concordant	0.2782 and 0.2782	Positive
chr2	29447523	29447524	AC/CA	AC/CA	Concordant	0.0184 and 0.0184	Negative
chr3	178952072	178952072	A/G	A/G	Concordant	COSMIC	NA
chr3	178952085	178952085	A/G	A/G	Concordant	COSMIC	NA
chr7	55222610	55222610	-/A	-/A	Concordant	0.0002	NA
chr7	140485334	140485335	GT/AC	GT/AC	Concordant	0.0693 and 0.1333	Positive
chr7	55259515	55259515	T/G	T/G	Concordant	COSMIC	NA
chr9	139377498	139377499	GG/CA	GG/CA	Concordant	0.2483 and 0.1264	Positive
chr10	104850568	104850569	TC/GT	TC/GT	Concordant	0.0809 and 0.0809	Positive
chr12	25398284	25398284	C/T	C/T	Concordant	COSMIC	NA
chr16	33373381	33373382	TG/CA	TG/CA	Concordant	0.4298 and 0.4298	Negative

*Numbers are minor allele frequencies in 4.7KJPN (DNP is splitted as two SNPs) and COSMIC indicates the variation is found in the COSMIC database. **The gnomAD MNV DB indicates the variation is in the gnomAD multinucleotide variants database.
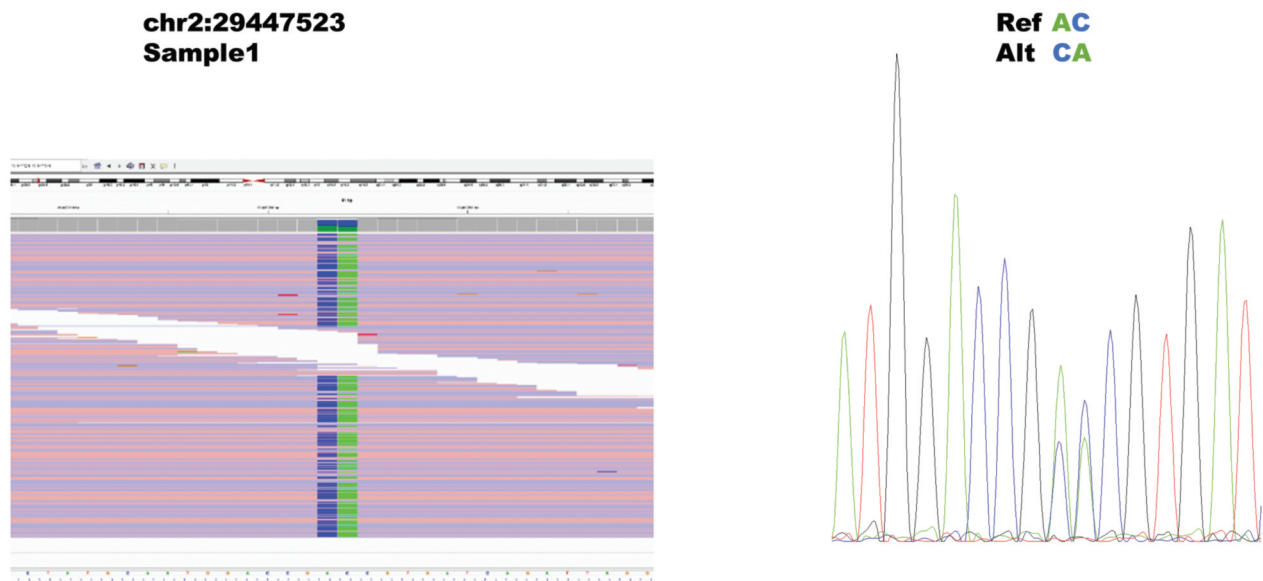

Figure 2. *Verification of DNVs found in the NCC oncopanel data by Sanger sequencing. Verification of dinucleotide variants (DNVs). Left image is generated with the Integrative Genomics Viewer (16). Right image is Sanger fluorogram corresponding to the DNV.*

most came from germline mutations. In addition, PANC-1 pancreatic cancer cells, which originated from a Caucasian patient, showed relatively high numbers (34 variants) of 4.7KJPN-negative variants. More than half of the samples showed variants in the COSMIC database but not in the 4.7KJPN. All of them (13 variants) are single nucleotide variants (SNVs) and detected in most of the tissues analyzed (Table III).

Unexpectedly, we found a few dinucleotide variants (DNVs) per sample in our dataset. Sanger sequencing of the selected variants (SNVs and DNVs) was undertaken for 12 loci (Table IV and Figures 2 and 3). One locus, g.chr17: 29663624–29663625, is in the middle of the T stretch and A stretch inverted repeat (11 T followed by 11 A: chr17: 29663615–29552636). We failed to confirm the sequence of this locus for one sample and thus we could only verify 11 loci (Table III). All the DNVs except one were confirmed by the Sanger method and we speculated whether the two nucleotide changes occurred in one chromosome or were split in two chromosomes, as suggested by the database. Figure 2 indicates that the DNVs were basically localized in single reads, suggesting that most of the DNVs that appeared in this study should be considered as one variant rather than two SNVs (14). Two DNVs showed different
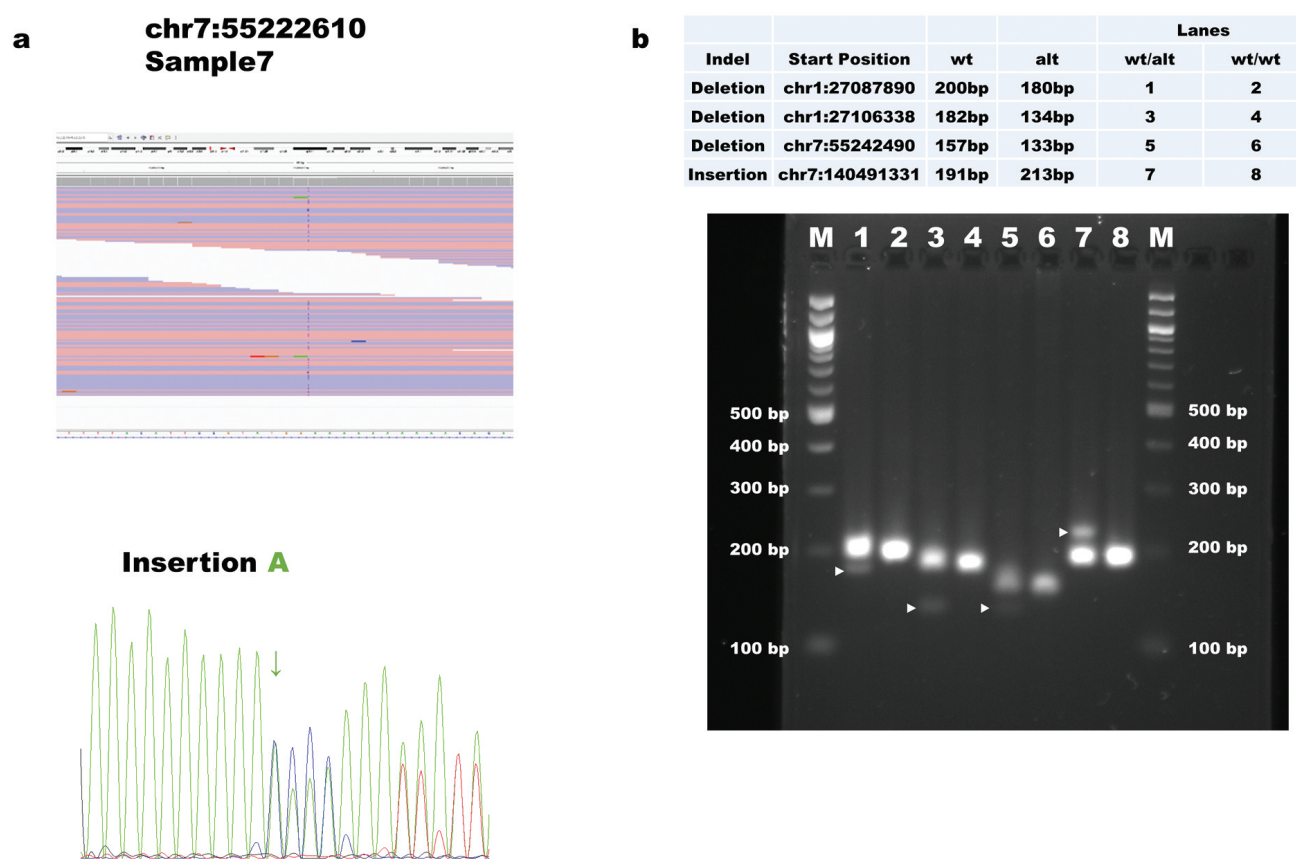
a

**chr7:55222610
Sample7**

Insertion A

b

Indel	Start Position	wt	alt	Lanes	
				wt/alt	wt/wt
Deletion	chr1:27087890	200bp	180bp	1	2
Deletion	chr1:27106338	182bp	134bp	3	4
Deletion	chr7:55242490	157bp	133bp	5	6
Insertion	chr7:140491331	191bp	213bp	7	8

Figure 3. *Verification of indels in NCC oncopanel data. (a) Verification of one nucleotide insertion by Sanger sequencing. Top image is generated with the Integrative Genomics Viewer and one base (Adenine) insertion is visible in the middle of the image. Bottom image is Sanger fluorogram corresponding to one-base insertion (b) Verification of indel variants. The genomic positions, amplicon sizes, and sample genotypes are indicated at the top of the gel image. The nucleotide lengths are indicated on the left and right sides of the gel image. The PCR templates are sample 14 (lanes 1, 3, and 7), sample 13 (lanes 2, 4, and 8), sample 8 (lane 5) and sample 10 (lane 6). The alt alleles are indicated with triangles in the gel image. "wt" and "alt" indicate the lengths of wild-type and of alternate alleles, respectively.*

minor allele frequencies between the first and second variant nucleotides (chr7:140185334–140185334 and chr9:13977498–13977499). In one of these, chr9:13977498–13977499, we observed three different alleles in our collection. One allele corresponded to the reference, the DNV, and g.chr9: 13977499G>C SNV among our samples, indicating that the overlapped nucleotide between the DNV and SNV showed a larger minor allele frequency in the 4.7KJPN database. We could not identify overlapping SNVs for the other, chr7:140185334–140185334, in our dataset. We also verified indels by Sanger sequencing and agarose gel electrophoresis of the amplicons including the indel-detected regions. Figure 3 shows that all the four indels tested showed the expected one-base shift in the fluorogram (Figure 3a) or in the band sizes of the amplicon products (Figure 3b), showing the accuracy of the NCC oncopanel variant call data.

Discussion

The present study was focused on the robustness of a cancer profiling test for the quality of genomic DNA extracted from FFPE surgical specimens from various cancerous tissues with different storage conditions. To address this issue, we investigated a cancer profiling test, NCC oncopanel. Initial quality estimates by real-time PCR indicated that the quality of the extracted genomic DNA was not strongly correlated with the storage conditions. Similarly, the quality of the next generation sequencing reads was not different significantly, regardless of their storage conditions. However, the mapping quality of the sequence reads was affected by the FFPE storage conditions. Interestingly, the length of storage and fixation affected distinct mapping statistics of the NCC oncopanel cancer profiling test reads. Prolonged storage increased non-primary alignments and longer fixation caused

an increase in unmapped reads. However, appropriate filtration of the variant calls produced reasonable results as an NCC oncopanel cancer profiling test, regardless of the variation in the storage or fixation length of FFPE samples. Our results indicate that the NCC oncopanel is a very robust cancer profiling test for most FFPE cancer specimens.

In the present study, we tested the quality of FFPE samples with real-time PCR and adjusted the amount of DNA sample based on the results. This may not be possible for most of the hospitals in Japan because the real-time PCR kits for the quality control of FFPE samples are not covered by Japanese health insurance. Therefore, judging the quality of FFPE samples for cancer profiling tests by storage periods or fixation length should be critical.

One of our interesting observations was that the mapping statistics can differ with different storage and fixation lengths. Increases in non-primary alignment resulted in increases in PCR duplication and may cause a decrease in SNV detection (15). Most likely, prolonged storage could cause an overall decrease in the "amplifiable DNA" in the FFPE samples and consequentially a decrease in the complexity of the NGS libraries. At the bedside, biopsy specimens are frequently used for cancer profiling tests. In the case of biopsy samples with prolonged storage, more sections may be required for cancer profiling tests than for newer biopsy samples. In the case of fixation, its effects are mainly to increase the sequence variations compared with the reference genome. C to T deamination is the mechanism of the increase in "unmapped reads" and "singletons" (5, 6). "Unmapped reads" and "singletons" would be caused by a few nucleotide changes in sequence reads and cause mismapping.

We identified a substantial number of DNVs that were not found in the genome variant databases of the general population. Wang *et al*. pointed out that there are three major mutational mechanisms for DNVs: combinations of independent SNVs, replication errors by DNA polymerase-zeta, and slippage at the repeat junctions (14). The first type of DNV mutation mechanism was detected in our dataset and showed different allele frequencies in the 4.7KJPN (Table IV). The third type of DNV was hg19/g.chr17:29663624-29663625, which we failed to verify in one sample because of the slippage during Sanger sequencing reactions. We believe some of the DNVs in Table IV are a result of the second mutational DNV mechanism, although the pattern of variation was not typical for polymerase-zeta (TC or GC>AA) (14). The variant databases did not have DNVs because the default option of the variant caller (GATK HaplotypeCaller) splits the MNVs into SNVs for subsequent joint genotyping in population studies.

The recommended conditions for FFPE samples for cancer profiling panel tests are as following: firstly, the prolonged stored samples (more than 4 years) may be detrimental for cancer profiling tests. Secondly, prolonged fixation (more

than 3 days) should also be avoided. Ideally, the resected tumor samples should be cut for the better permeability of formalin. In summary, our results show that a cancer profiling test, the NCC oncopanel, is robust for FFPE sample quality. The storage and fixation conditions of samples tested in this study are within the expected range of FFPE samples at most Japanese hospitals. Hence, we conclude that most FFPE samples can be examined using cancer profiling tests.

Conflicts of Interest

There are no conflicts of interest to disclose.

Authors' Contributions

J.Y. and I.S. planned the study. S.I and T.M. performed the experiments. M.M., K.Y., and K.T. helped conduct the experimental procedures. Bioinformatics analysis was performed by S.I., N.T., and J.Y. The manuscript was written by S.I. and J.Y. H.S. contributed to the general management of the project.

Acknowledgements

References

1 Hellmann MD, Ciuleanu TE, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, Minenza E, Linardou H, Burgers S, Salman P, Borghaei H, Ramalingam SS, Brahmer J, Reck M, O'Byrne KJ, Geese WJ, Green G, Chang H, Szustakowski J, Bhagavatheeswaran P, Healey D, Fu Y, Nathan F and Paz-Ares L: Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. N Engl J Med *378(22)*: 2093-2104, 2018. PMID: 29658845. DOI: 10.1056/NEJMoa1801946

2 Sunami K, Ichikawa H, Kubo T, Kato M, Fujiwara Y, Shimomura A, Koyama T, Kakishima H, Kitami M, Matsushita H, Furukawa E, Narushima D, Nagai M, Taniguchi H, Motoi N, Sekine S, Maeshima A, Mori T, Watanabe R, Yoshida M, Yoshida A, Yoshida H, Satomi K, Sukeda A, Hashimoto T, Shimizu T, Iwasa S, Yonemori K, Kato K, Morizane C, Ogawa C, Tanabe N, Sugano K, Hiraoka N, Tamura K, Yoshida T, Fujiwara Y, Ochiai A, Yamamoto N and Kohno T: Feasibility and utility of a panel testing for 114 cancer-associated genes in a clinical setting: A hospital-based study. Cancer Sci *110(4)*: 1480-1490, 2019. PMID: 30742731. DOI: 10.1111/cas.13969

3 Srinivasan M, Sedmak D and Jewell S: Effect of fixatives and tissue processing on the content and integrity of nucleic acids.

Am J Pathol *161(6)*: 1961-1971, 2002. PMID: 12466110. DOI: 10.1016/S0002-9440(10)64472-0

4 Vitošević K, Todorović M, Varljen T, Slović Ž, Matić S and Todorović D: Effect of formalin fixation on pcr amplification of DNA isolated from healthy autopsy tissues. Acta Histochem *120(8)*: 780-788, 2018. PMID: 30224244. DOI: 10.1016/j.acthis.2018.09.005

5 Do H and Dobrovic A: Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil- DNA glycosylase. Oncotarget *3(5)*: 546-558, 2012. PMID: 22643842. DOI: 10.18632/oncotarget.503

6 Prentice LM, Miller RR, Knaggs J, Mazloomian A, Aguirre Hernandez R, Franchini P, Parsa K, Tessier-Cloutier B, Lapuk A, Huntsman D, Schaeffer DF and Sheffield BS: Formalin fixation increases deamination mutation signature but should not lead to false positive mutations in clinical practice. PLoS One *13(4)*: e0196434, 2018. PMID: 29698444. DOI: 10.1371/journal.pone.0196434

7 Bolger AM, Lohse M and Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30(15)*: 2114-2120, 2014. PMID: 24695404. DOI: 10.1093/bioinformatics/btu170

8 Li H: Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. arXive: 1303.3997v1301 [q-bio.GN], 2013.

9 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and 1000 Genome Project Data Processing Subgroup: The sequence alignment/map format and SAMtools. Bioinformatics *25(16)*: 2078-2079, 2009. PMID: 19505943. DOI: 10.1093/bioinformatics/btp352

10 Wang K, Li M and Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res *38(16)*: e164, 2010. PMID: 20601685. DOI: 10.1093/nar/gkq603

11 Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, Fish P, Harsha B, Hathaway C, Jupe SC, Kok CY, Noble K, Ponting L, Ramshaw CC, Rye CE, Speedy HE, Stefancsik R, Thompson SL, Wang S, Ward S, Campbell PJ and Forbes SA: COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res *47(D1)*: D941-D947, 2019. PMID: 30371878. DOI: 10.1093/nar/gky1015

12 Tadaka S, Saigusa D, Motoike IN, Inoue J, Aoki Y, Shirota M, Koshiba S, Yamamoto M and Kinoshita K: jMorp: Japanese multi omics reference panel. Nucleic Acids Res *46(D1)*: D551-D557, 2018. PMID: 29069501. DOI: 10.1093/nar/gkx978

13 Li H: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27(21)*: 2987-2993, 2011. PMID: 21903627. DOI: 10.1093/bioinformatics/btr509

14 Wang Q, Pierce-Hoffman E, Cummings BB, Alföldi J, Francioli LC, Gauthier LD, Hill AJ, O'Donnell-Luria AH, Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, Karczewski KJ and MacArthur DG: Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. Nat Commun *11(1)*: 2539, 2020. PMID: 32461613. DOI: 10.1038/s41467-019-12438-5

15 Liu Q, Guo Y, Li J, Long J, Zhang B and Shyr Y: Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data. BMC Genomics *13 Suppl 8*: S8, 2012. PMID: 23281772. DOI: 10.1186/1471-2164-13-S8-S8

16 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G and Mesirov JP: Integrative genomics viewer. Nat Biotechnol *29(1)*: 24-26, 2011. PMID: 21221095. DOI: 10.1038/nbt.1754