

## Chapter 3

# Assessment of Formalin-Induced DNA Damages in FFPE Specimens

The main component of formalin is formaldehyde, which is known to induce DNA damages such as fragmentation and sequence artifacts. Our study design is comprised of 213 patients with FFPE tumour and matched blood specimens that were subjected to evaluation with the OncoPanel. Sequencing data of tumour-normal paired samples were processed and analyzed with a custom variant calling pipeline. With blood specimens serving as non-formalin-fixed controls, we assessed formalin-induced DNA damages by comparing efficiency in amplicon enrichment and sequencing results of FFPE specimens to blood. As DNA derived from blood is the gold standard for germline testing, our assessment would determine whether FFPE DNA is a reliable resource for detection of germline variants.

### 3.1 Comparison of efficiency in amplicon enrichment and sequencing results between blood and FFPE specimens

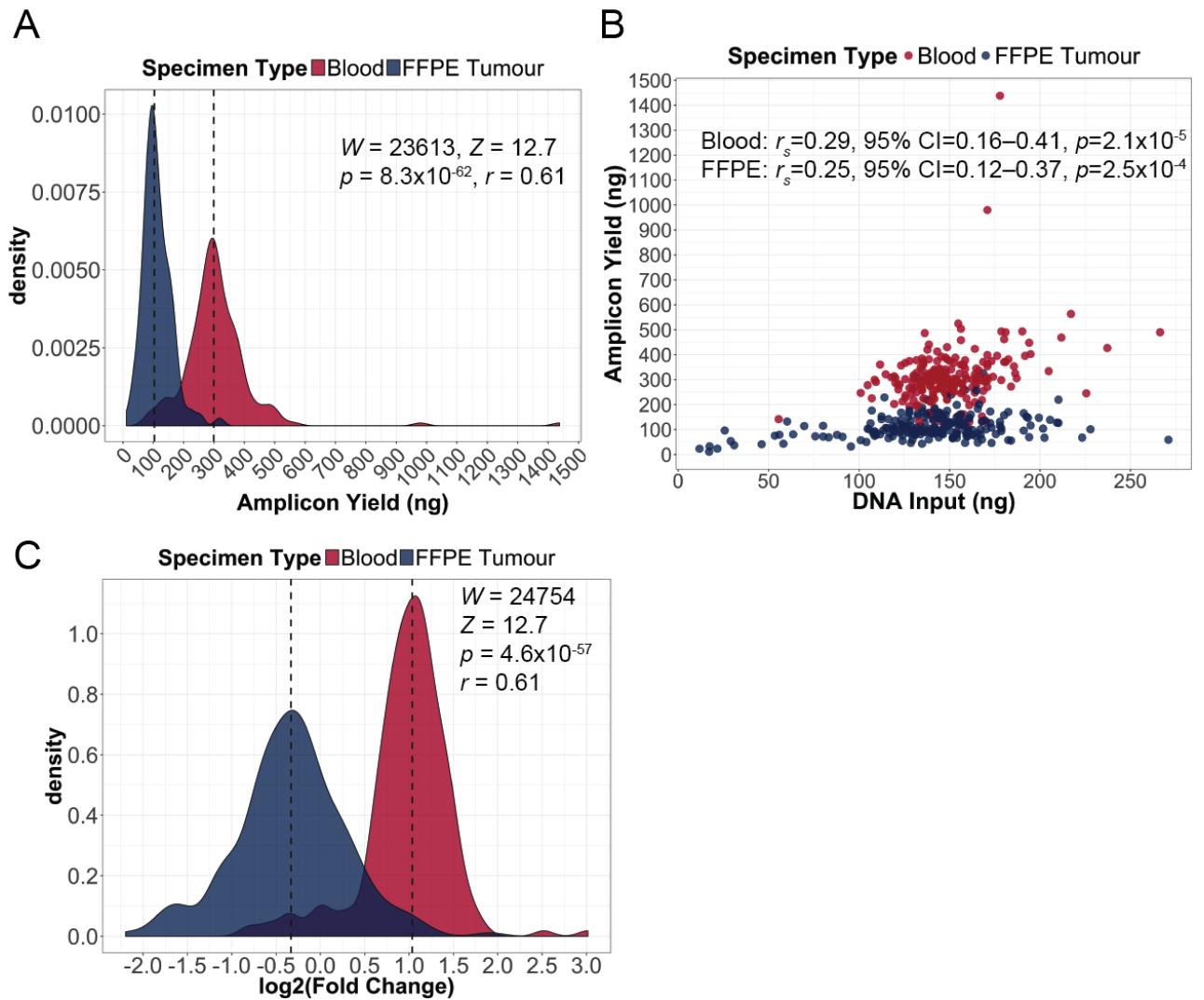
Formalin fixation causes DNA fragmentation that would reduce template DNA for PCR amplification, leading to decreased efficiency in amplicon enrichment methods for FFPE DNA. To investigate this effect, we first compared the amplicon yield between blood and FFPE specimens, and a Wilcoxon signed-rank test indicated that amplicon yield in FFPE specimens was significantly lower than blood specimens ( $W = 23613, Z = 12.7, p = 8.3 \times 10^{-62}, r = 0.61$ ; Figure 3.1A). However, the amount of DNA input for amplicon enrichment varies across specimens in our study design, and we demonstrated that amplicon yield was weakly correlated with DNA input for both blood and FFPE specimens (Spearman's rank correlation: blood,  $r_s = 0.29, 95\% \text{ CI} = 0.16\text{--}0.41, p = 2.1 \times 10^{-5}$ ; FFPE,  $r_s = 0.25, 95\% \text{ CI} = 0.12\text{--}0.37, p = 2.5 \times 10^{-4}$ ; Figure 3.1B). To account for the difference in DNA input across specimens, we derived the log<sub>2</sub> fold change between DNA input and amplicon yield ( $\log_2 (\text{Amplicon Yield}/\text{DNA Input})$ ) to measure the efficiency in amplicon enrichment. We

compared the log<sub>2</sub> fold change in FFPE specimens to blood, and we found a significant decrease in enrichment efficiency in FFPE specimens compared to blood (Wilcoxon signed-rank test,  $W = 24754$ ,  $Z = 12.7$ ,  $p = 4.6 \times 10^{-57}$ ,  $r = 0.61$ ; Figure 3.1C). This result implies that production of amplicons is less efficient in FFPE specimens compared to blood, demonstrating the drawback of using FFPE DNA in amplicon-based next-generation sequencing.

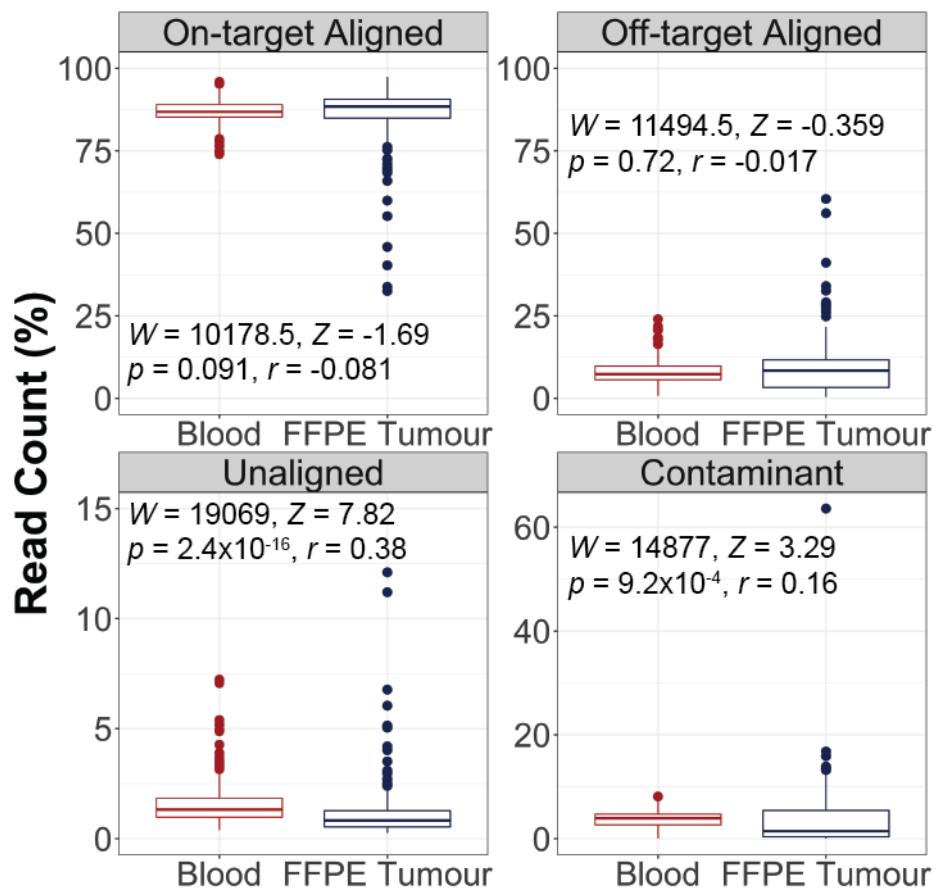
To examine whether blood and FFPE specimens produce comparable sequencing results, we compared read alignments between blood and FFPE specimens. Inspection of on-target aligned reads, which are reads that align to target regions used for variant calling, revealed no significant difference in the percentage of on-target aligned reads between blood and FFPE specimens (Wilcoxon signed-rank test,  $W = 10178.5$ ,  $Z = -1.69$ ,  $p = 0.091$ ,  $r = -0.081$ ; Figure 3.2). However, there were more outliers with slightly lower percentage of on-target aligned reads (< 75%) in FFPE specimens compared to blood, and the distribution of percentage of on-target aligned reads was also wider in FFPE specimens (range: FFPE = 32.5–97.4%, blood = 74.0–95.9%), suggesting more variability in the rate of on-target alignment in FFPE specimens than blood. Similarly, no significant difference in the percentage of off-target aligned reads, which are reads that map to the human reference genome but not to target regions, was observed between specimen types (Wilcoxon signed-rank test,  $W = 11494.5$ ,  $Z = -0.359$ ,  $p = 0.72$ ,  $r = -0.017$ ; Figure 3.2). Although a Wilcoxon signed-rank test indicated that the percentage of unaligned reads was significantly different between blood and FFPE specimens ( $W = 19069$ ,  $Z = 7.82$ ,  $p = 2.4 \times 10^{-16}$ ,  $r = 0.38$ ; Figure 3.2), there was only a small decrease in the median percentage of unaligned reads in FFPE specimens compared to blood (median: FFPE = 0.8%, blood = 1.3%). Moreover, our data showed no significant difference in percentage of contaminant reads between specimen types ( $W = 14877$ ,  $Z = 3.29$ ,  $p = 9.2 \times 10^{-4}$ ,  $r = 0.16$ ; Figure 3.2), although there was one extreme outlier in FFPE specimens (range: FFPE = 0.028–64%, blood = 0.082–8.1%). While there were minor differences in percentage of unaligned reads between sequencing libraries generated from blood and FFPE DNA, blood and FFPE libraries resulted in comparable percentage of on-target aligned reads, thereby providing equivalent amount of aligned reads for variant calling.

Although blood and FFPE specimens demonstrated no significant difference in the percentage of on-target aligned reads, this result does not reflect the coverage depth of target regions in blood and FFPE specimens. To examine whether discrepancy in coverage depth exists between specimen types, we obtained coverage depth of target bases for all sequencing libraries and normalized per base coverage depth to account for difference in library size. We derived the average per base coverage depth for each library and compared this sequencing metric between blood and FFPE specimens. The average per base coverage depth was significantly different between FFPE and blood specimens (Wilcoxon signed-rank test,  $W = 20864$ ,  $Z = 9.76$ ,  $p = 2.5 \times 10^{-26}$ ,  $r = 0.47$ ), but there was only a slight decrease in the average per base coverage depth in FFPE specimens compared to blood (median: FFPE = 1194, blood = 1271). We also calculated the percentages of target bases

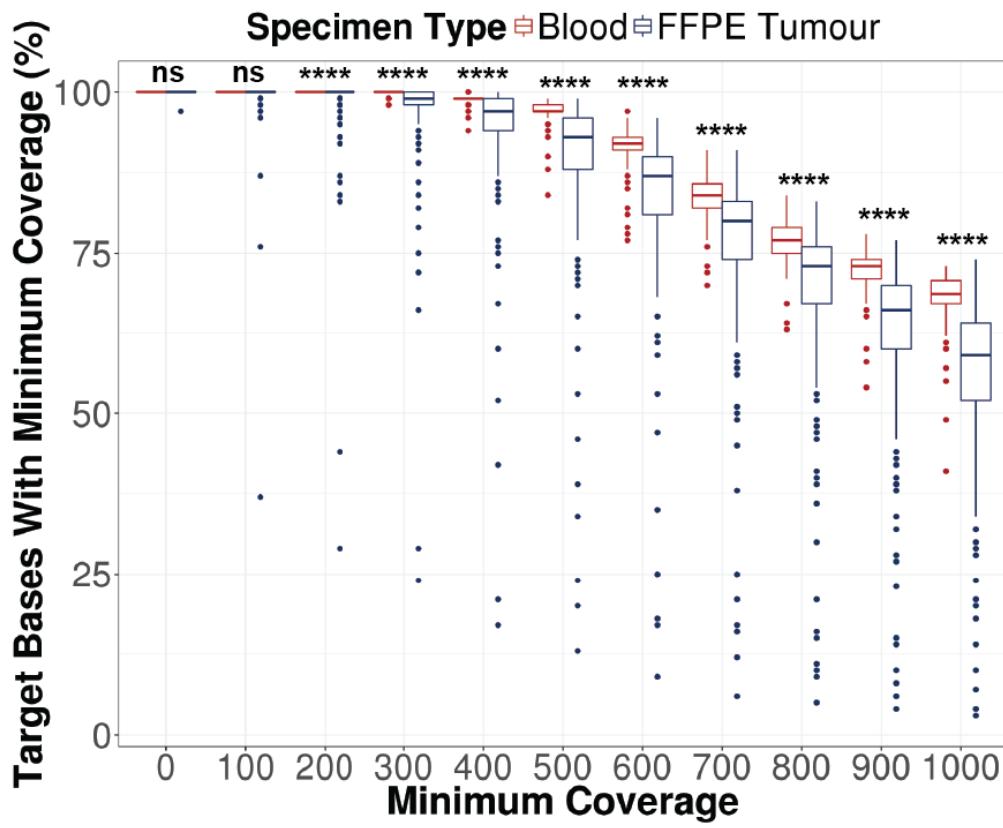
that met coverage thresholds ranging from zero to 1000x to evaluate coverage uniformity of target bases between blood and FFPE specimens. While coverage uniformity was significantly different between blood and FFPE specimens at coverage levels except at the zero and 100x coverage depth cut-off (Wilcoxon signed-rank test,  $p < 0.0001$ ; Figure 3.3), we considered these discrepancies to be subtle because the absolute difference in median percentage of target bases only exceeded 5% at 500x, 900x, and 1000x coverage thresholds (Table 3.1). Nevertheless, there were more outliers with lower percentage of target bases than median values in FFPE specimens at coverage thresholds between 100x to 1000x, implying that poor coverage uniformity is more profound for a subset of FFPE specimens. Together, our findings reveal that FFPE specimens demonstrated lower efficiency in amplicon enrichment and minor discrepancies in coverage depth and uniformity compared to blood specimens, whereas comparable proportion of on-target read alignments could be attained between specimen types.



**Figure 3.1:** Comparison of efficiency in amplicon enrichment between blood and FFPE specimens. (A) The distributions of amplicon yield in blood and FFPE specimens (Wilcoxon signed-rank test). Dashed lines indicate median amplicon yield in blood and FFPE specimens, which are 299.3 ng and 103.6 ng, respectively. (B) The correlations between amplicon yield and the amount of DNA input for amplicon enrichment in blood and FFPE specimens (Spearman's rank correlation). (C) The distributions of fold change between DNA input and amplicon yield ( $\log_2$ ), which is used to measure efficiency in amplicon enrichment in blood and FFPE specimens (Wilcoxon signed-rank test). Dashed lines indicate median  $\log_2$  fold change in blood and FFPE specimens, which are 1.04 and -0.332, respectively.



**Figure 3.2:** Assessment of read alignments between blood and FFPE specimens (Wilcoxon signed-rank test). Box plots show the median (horizontal bar within) and interquartile range (IQR) of percentage of reads, with whiskers representing the range of data not exceeding 1.5x the IQR and circles indicating outliers.



**Figure 3.3:** Evaluation of coverage uniformity in blood and FFPE specimens (Wilcoxon signed-rank test, \*\*\* $p < 0.0001$ , ns = not significant). Per base coverage was normalized to account for difference in library size. Percentage of target bases that met various coverage thresholds was calculated. Box plots show the median (horizontal bar within) and IQR of percentage of target bases that met the respective coverage thresholds, with whiskers representing the range of data not exceeding 1.5x the IQR and circles indicating outliers.

**Table 3.1:** Comparison of coverage uniformity between blood and FFPE specimens using the Wilcoxon signed-rank test.

Threshold	Blood		FFPE Tumour		$D^{\dagger}$ (%)	$p (< 0.0001^*)$
	Median (%)	Range (%)	Median (%)	Range (%)		
$\geq 0x$	100	100–100	100	97.0–100	0.0	1.0
$\geq 100x$	100	100–100	100	37.0–100	0.0	$2.3 \times 10^{-4}$
$\geq 200x$	100	100–100	100	29.0–100	0.0	$2.9 \times 10^{-11}^*$
$\geq 300x$	100	98.0–100	99.0	24.0–100	1.0	$4.1 \times 10^{-18}^*$
$\geq 400x$	99.0	94.0–100	97.0	17.0–100	2.0	$5.0 \times 10^{-28}^*$
$\geq 500x$	97.0	84.0–99.0	89.5	13.0–99.0	7.5	$2.1 \times 10^{-38}^*$
$\geq 600x$	92.0	77.0–97.0	87.0	9.0–96.0	5.0	$1.5 \times 10^{-32}^*$
$\geq 700x$	84.0	70.0–91.0	80.0	6.0–91.0	4.0	$5.7 \times 10^{-25}^*$
$\geq 800x$	77.0	63.0–84.0	73.0	5.0–83.0	4.0	$4.7 \times 10^{-27}^*$
$\geq 900x$	73.0	54.0–78.0	66.0	4.0–77.0	7.0	$4.6 \times 10^{-40}^*$
$\geq 1000x$	68.5	41.0–73.0	59.0	3.0–74.0	9.5	$3.6 \times 10^{-42}^*$

<sup>†</sup>Absolute difference between median of blood and FFPE specimens.

### 3.2 Reduced coverage depth in FFPE specimens is more pronounced for longer amplicons

The OncoPanel consists of 416 amplicons that interrogate coding exons and mutational hotspots of 21 genes, and these amplicons vary in length and GC content. Since we observed discrepancy in sequencing coverage between blood and FFPE specimens, we sought to determine whether this discrepancy is influenced by amplicon length and GC content. We attained the coverage depth for each amplicon and normalized the coverage depth to account for difference in library size. We considered a fold change of 1.5x in either direction to be a substantial change in amplicon coverage depth between blood and FFPE specimens. Using this threshold, we found significant differences in coverage depth between blood and FFPE specimens for 44 out of 416 amplicons (Wilcoxon signed-rank test with Benjamini-Hochberg correction, adjusted  $p < 0.0001$ ; Figure 3.4). To quantify the amplicon-specific differences in coverage depth, we derived the log2 fold change in the median coverage depth between blood and FFPE specimens ( $\log_2 (\text{Median Coverage}_{\text{FFPE}}/\text{Median Coverage}_{\text{Blood}})$ ) for each amplicon. Hence, a negative fold change indicates lower coverage depth of the amplicon in FFPE specimens relative to blood specimens, whereas a positive fold change indicates higher coverage depth of the amplicon in FFPE specimens relative to blood specimens. Our assessment showed that 23 out of the 44 amplicons have negative log2 fold changes, whereas 21 out of the 44 amplicons have positive log2 fold changes (Figure 3.4). These results indicate that there are differ-

ences in coverage depth between FFPE and blood specimens for certain amplicons in the panel, with slightly more amplicons exhibiting lower coverage depth in FFPE specimens than blood specimens.

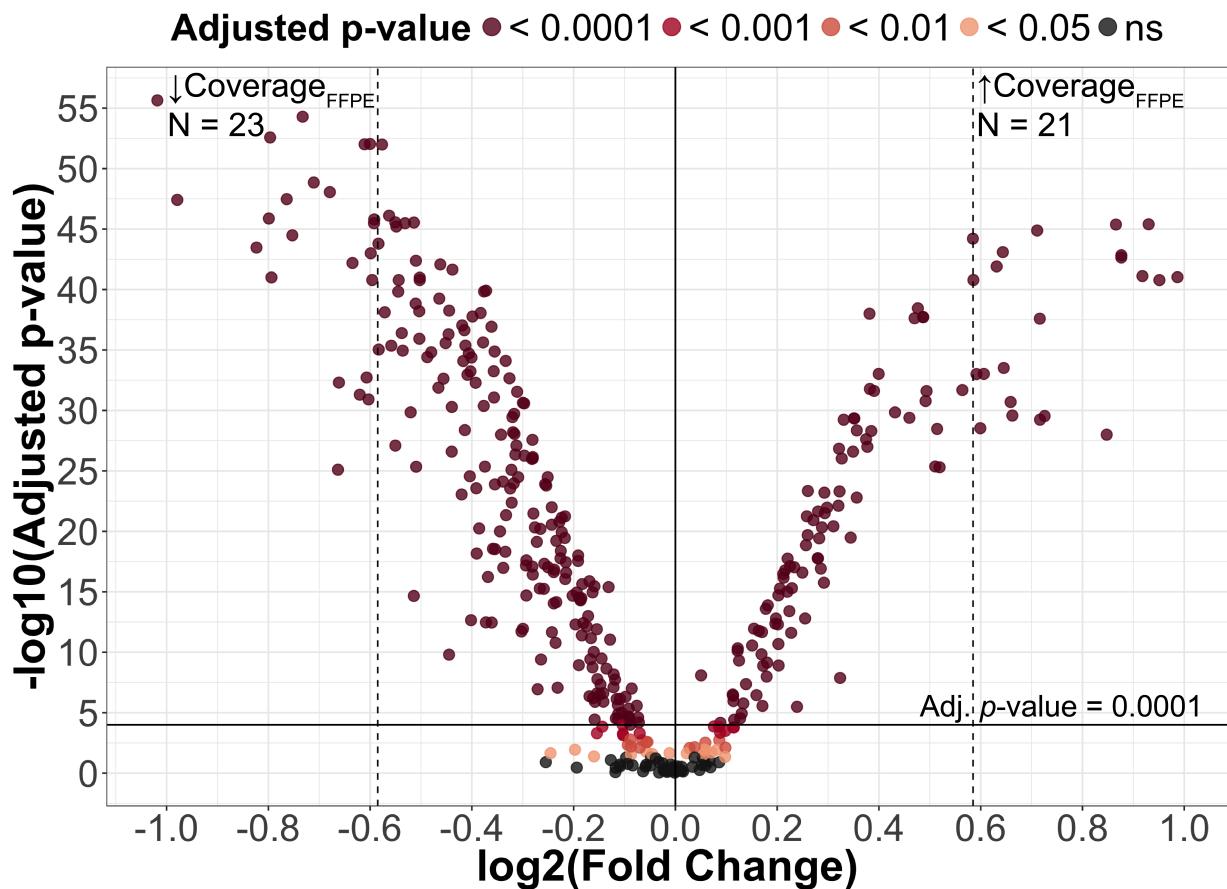
We subsequently examined the impact of amplicon length and GC content on the amplicon-specific differences in coverage depth between specimen types, which we measured as the log<sub>2</sub> fold change in median coverage depth between blood and FFPE specimens. We first confirmed that no significant correlation exists between amplicon GC content and length (Pearson's correlation,  $r = 0.045$ , 95% CI = -0.051–0.14,  $p = 0.36$ ; Figure 3.5). We then evaluated the correlation between log<sub>2</sub> fold change in amplicon coverage depth and amplicon length, and Pearson's correlation demonstrated a strong, negative correlation between the two variables ( $r = -0.79$ , 95% CI = -0.82–-0.75,  $p = 1.4 \times 10^{-88}$ ; Figure 3.6A). This result indicates that coverage depth in FFPE specimens tend to be lower relative to blood specimens as amplicon length increases. On the other hand, coverage depth tend to be enriched in FFPE specimens relative to blood for shorter amplicons. We also assessed the correlation between log<sub>2</sub> fold change in amplicon coverage depth and amplicon GC content, and Pearson's correlation demonstrated a weak, negative correlation between the two variables ( $r = -0.31$ , 95% CI = -0.40–-0.22,  $p = 1.1 \times 10^{-10}$ ; Figure 3.6B). Although the correlation is weak, this finding still implies that coverage depth in FFPE specimens tend to be lower relative to blood specimens as amplicon GC content increases, whereas enriched coverage depth in FFPE specimens with respect to blood was observed for amplicons with lower GC content.

Because amplicon length and GC content demonstrated significant correlations with amplicon-specific differences in coverage depth, we determined which contributing factor has a greater effect. We used a multiple linear regression to predict log<sub>2</sub> fold change in amplicon coverage depth based on amplicon length and GC content (Table 3.2). A significant equation was found ( $F(2, 411) = 471$ ,  $p = 4.65 \times 10^{-107}$ ), with an adjusted  $R^2$  of 0.695. Predicted log<sub>2</sub> fold change in amplicon coverage depth between blood and FFPE specimens is equal to

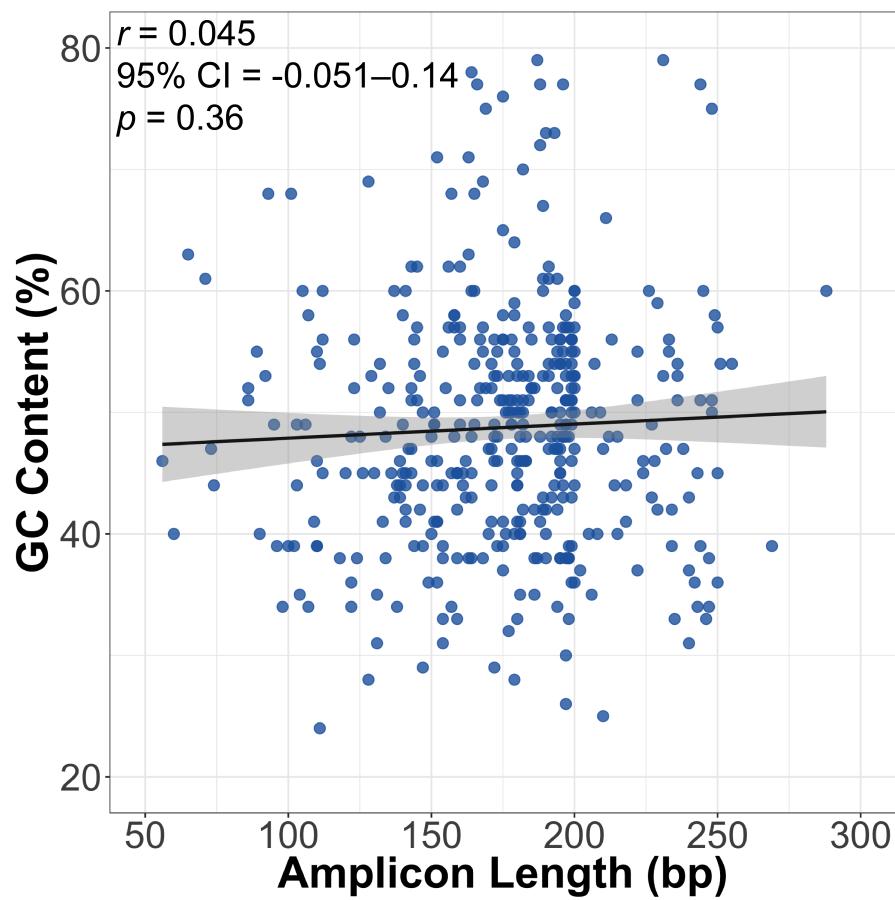
$$1.66 - 7.24 \times 10^{-3}(\text{Length}) - 9.92 \times 10^{-3}(\text{GC Content}),$$

in which amplicon length is expressed in base pairs (bp) and GC content is expressed as percentage (%). Both amplicon length and GC content were significant predictors of log<sub>2</sub> fold change in amplicon coverage depth. Based on the standardized coefficients, we compared the strength of predictors within the model to identify the predictor with a greater effect on the response variable. Our assessment showed that one standard deviation increase in amplicon length would lead to a 0.775 standard deviation decrease in log<sub>2</sub> fold change in amplicon coverage depth, whereas one standard deviation increase in amplicon GC content would lead to a 0.277 standard deviation decrease in log<sub>2</sub> fold change in amplicon coverage depth. This result indicates that amplicon length has a stronger association with amplicon-specific differences in coverage depth between specimen types, which we measured as the log<sub>2</sub> fold change in amplicon coverage depth between blood and FFPE speci-

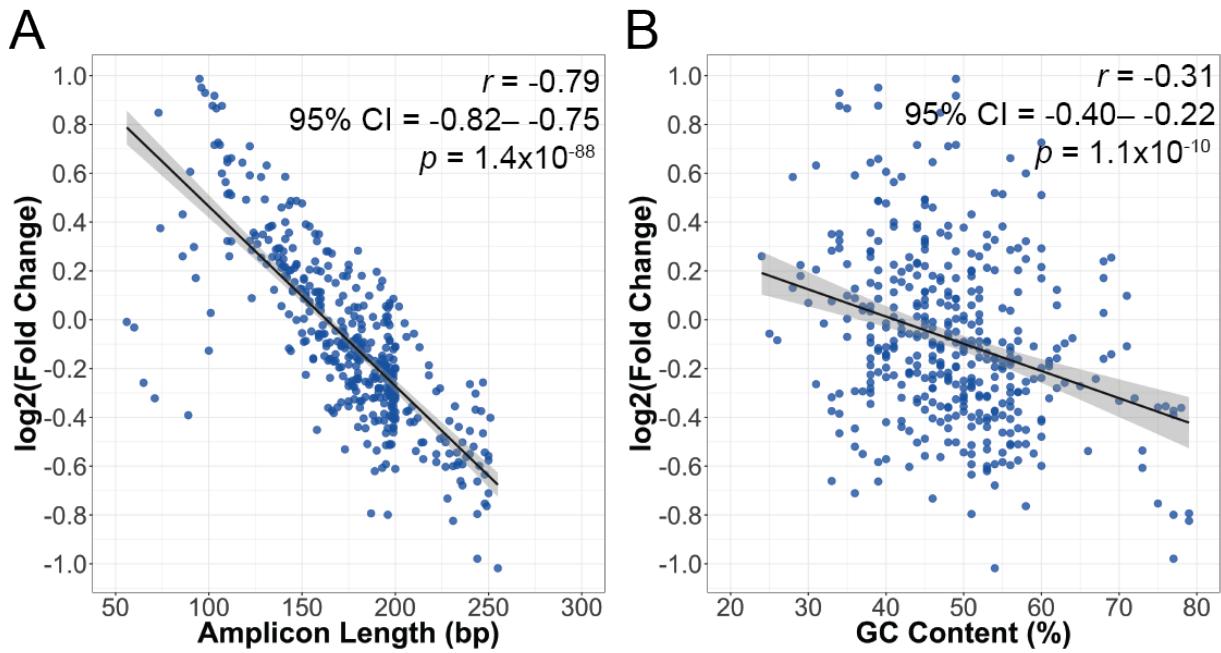
mens, than GC content. Collectively, these findings reveal the challenge imposed by fragmentation damages in FFPE DNA, which results in shorter template DNA that would be more difficult for PCR amplification to produce longer amplicons.



**Figure 3.4:** Amplicon-specific differences in coverage depth between blood and FFPE specimens. Difference in amplicon coverage depth between specimen types was determined using the Wilcoxon signed-rank test with Benjamini-Hochberg correction (adjusted  $p < 0.0001$ ). Volcano plot illustrates the  $-\log_{10}$  adjusted  $p$ -value in relation to  $\log_2$  fold change between median coverage depth in blood and FFPE specimens ( $\log_2(\text{Median Coverage}_{\text{FFPE}}/\text{Median Coverage}_{\text{Blood}})$ ) for amplicons in the panel. Negative  $\log_2$  fold change indicates lower coverage depth of the amplicon in FFPE specimens relative to blood ( $\downarrow \text{Coverage}_{\text{FFPE}}$ ), whereas positive  $\log_2$  fold change indicates higher coverage depth of the amplicon in FFPE specimens relative to blood ( $\uparrow \text{Coverage}_{\text{FFPE}}$ ). A fold change of 1.5x in either direction, indicated by the dashed lines, was considered to be a substantial change in amplicon coverage depth between blood and FFPE specimens. N = number of amplicons; ns = not significant



**Figure 3.5:** The relationship between amplicon GC content and amplicon length (Pearson's correlation). Solid line represents the fitted linear relationship between the two variables, and the shaded band indicates pointwise 95% confidence interval of the fitted linear regression line.



**Figure 3.6:** Scatter plots showing log<sub>2</sub> fold change between amplicon coverage depth in blood and FFPE specimens ( $\log_2(\text{Median Coverage}_{\text{FFPE}}/\text{Median Coverage}_{\text{Blood}})$ ) in relation to (A) amplicon length and (B) GC content (Pearson's correlation). Solid line represents the fitted linear relationship between the two variables, and the shaded band indicates pointwise 95% confidence interval of the fitted linear regression line.

**Table 3.2:** Multiple linear regression to predict log<sub>2</sub> fold change between amplicon coverage depth in blood and FFPE specimens ( $\log_2(\text{Median Coverage}_{\text{FFPE}}/\text{Median Coverage}_{\text{Blood}})$ ) based on amplicon length and GC content.

Variable	Unstandardized Coefficient	Standard Error	Standardized Coefficient	p-value
Length (bp)	$-7.24 \times 10^{-3}$	$2.54 \times 10^{-4}$	$-7.75 \times 10^{-1}$	$2.47 \times 10^{-99}$
GC Content (%)	$-9.92 \times 10^{-3}$	$9.77 \times 10^{-4}$	$-2.77 \times 10^{-1}$	$8.70 \times 10^{-22}$
				Intercept = 1.66, Adjusted R <sup>2</sup> = 0.695 $F(2, 411) = 471$ , p-value = $4.65 \times 10^{-107}$

### 3.3 Deamination effects lead to increased C>T/G>A transitions in FFPE specimens

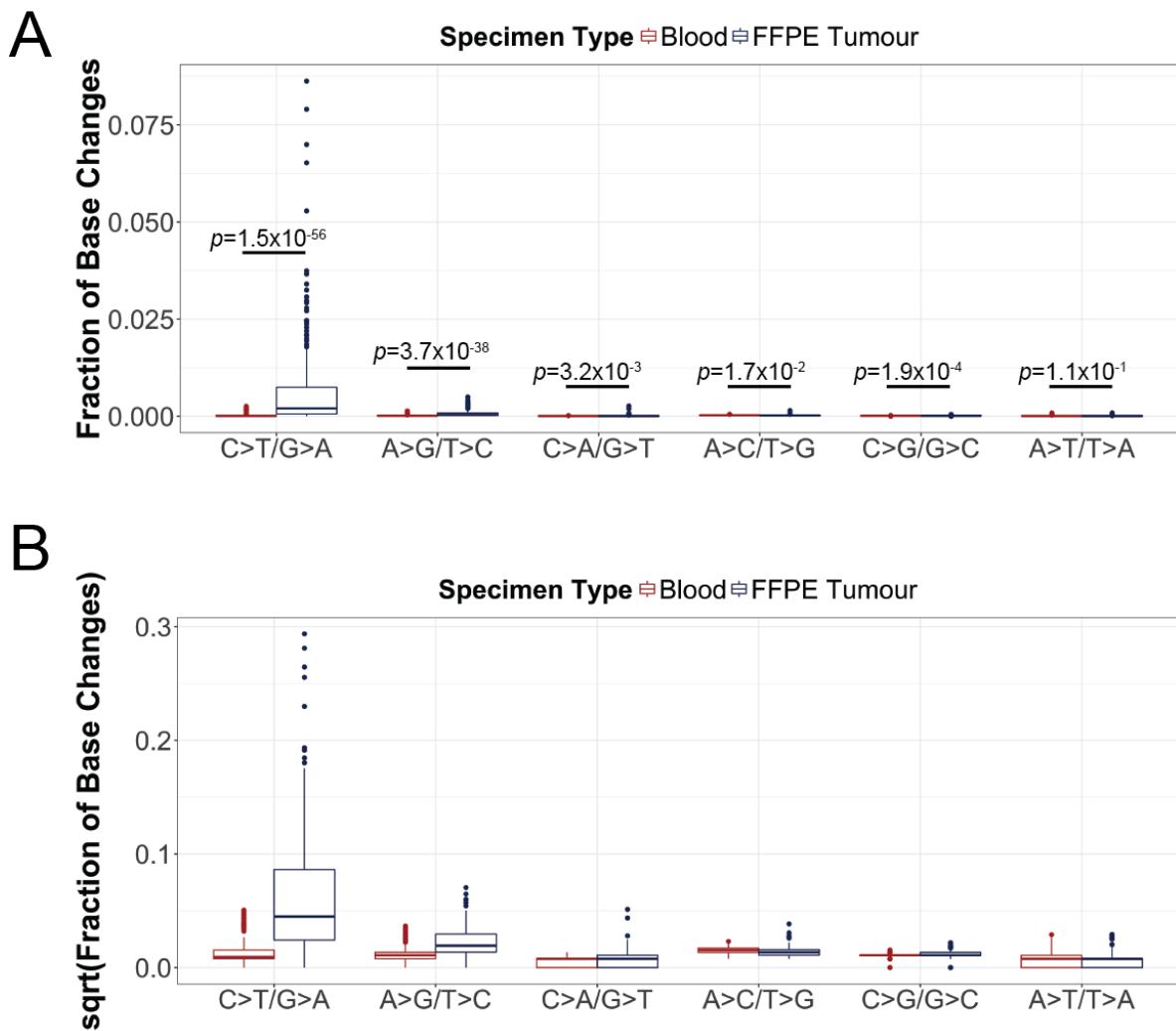
Formalin fixation not only induces DNA fragmentation, but also base modifications that give rise to sequence artifacts. A prominent type of formalin-induced sequence artifact is C>T/G>A transitions as a result of deamination of cytosine bases. To measure the level of formalin-induced artifacts

in FFPE specimens, we quantified the fraction of base changes that were not identified as true SNVs by our variant calling pipeline. We only considered high quality bases (Phred-scaled base quality score  $\geq 20$ ) and base changes that were  $\geq 1\%$  allele frequency to exclude sequencing errors from our analysis. Base changes were categorized into C>T/G>A and A>G/T>C, which are nucleotide transitions, as well as C>A/G>T, A>C/T>G, C>G/G>C, and A>T/T>A, which are nucleotide transversions. We compared the fraction of base changes between specimen types and found significant differences in fraction of C>T/G>A and A>G/T>C between blood and FFPE specimens (Wilcoxon signed rank test,  $p < 0.0001$ ; Figure 3.7A). As blood DNA is not affected by formalin fixation, we evaluated the prevalence of artifactual base changes in FFPE specimens with respect to blood by calculating the fold change between the median fraction of base changes in blood and FFPE specimens (Table 3.3). We noted a substantially higher fold change for C>T/G>A compared to A>G/T>C: fraction of C>T/G>A was 23 times higher in FFPE specimens relative to blood, whereas fraction of A>G/T>C was 3.1 times higher in FFPE specimens relative to blood. This result is consistent with deamination effects that are reportedly predominant in FFPE DNA and considering that base transitions occur at a higher frequency compared to transversions, it is possible that the modest increase in A>G/T>C transitions was caused by the presence of true SNVs that were not called by our variant calling pipeline.

To assess the relative difference in fraction of base changes in FFPE specimens compared to blood specimens, we calculated the log<sub>2</sub> fold change in fraction of base changes between paired blood and FFPE specimens ( $\log_2(\text{Fraction of Base Changes}_{\text{FFPE}}/\text{Fraction of Base Changes}_{\text{Blood}})$ ). We compared the relative difference in fraction of base changes across different types of base changes, and a Kruskal-Wallis test indicated that type of base changes has a significant effect on the relative difference in fraction of base changes ( $H = 428.5$ ,  $p = 2.1 \times 10^{-90}$ ; Figure 3.8). Multiple pairwise comparison of the relative difference in fraction of base changes was performed using a post-hoc Dunn's test with Benjamini-Hochberg correction. Relative difference in fraction of C>T/G>A was significantly different compared to the five other types of base changes, and this was similar for A>G/T>C (adjusted  $p < 0.0001$ ; Table 3.4). Although both C>T/G>A and A>G/T>C were elevated in FFPE specimens compared to the other base transversions, the magnitude of difference was larger for C>T/G>A than A>G/T>C (median log<sub>2</sub> fold change: C>T/G>A = 4.2, A>G/T>C = 1.6), which further confirms that deamination of cytosine bases is the most frequent form of sequence artifacts in FFPE DNA.

Formalin-induced sequence artifacts often occur at low allele frequency; hence, we examined the prevalence of sequence artifacts at different ranges of allele frequency, including 1–10%, 10–20%, and 20–30%. Because variants were not called within the 1–10% allele frequency range, we did not remove true SNVs detected by our variant calling pipeline to ensure consistency when comparing fraction of base changes across different ranges of allele frequency. Nevertheless, we adhered to the previous criterion of only including base changes with Phred-scaled base quality

score  $\geq$  20 in this analysis. For all types of base changes, we noted that the range of allele frequency has a significant effect on fraction of base changes in blood and FFPE specimens (Kruskal-Wallis test,  $p < 0.0001$ ; Figure ??), with increased levels of base changes at the 1-10% allele frequency range compared to 10-20% and 20-30%. Because blood DNA represents good quality DNA that is unaffected by formalin fixation, we also compared the fraction of base changes at the 1-10% allele frequency range in FFPE specimens to blood. Similar to previous analyses, there was a marked increase in C>T/G>A and a modest increase in A>G/T>C in FFPE specimens relative to blood within the 1-10% allele frequency (fold change: C>T/G>A = 33, A>G/T>C = 3.1; Table 3.5). Collectively, our assessment demonstrates that high frequency of C>T/G>A transitions is present and detectable in FFPE specimens, which indicates that deamination of cytosine is the primary form of formalin-induced sequence artifacts, and these artifactual transitions are more prevalent at low allele frequency.

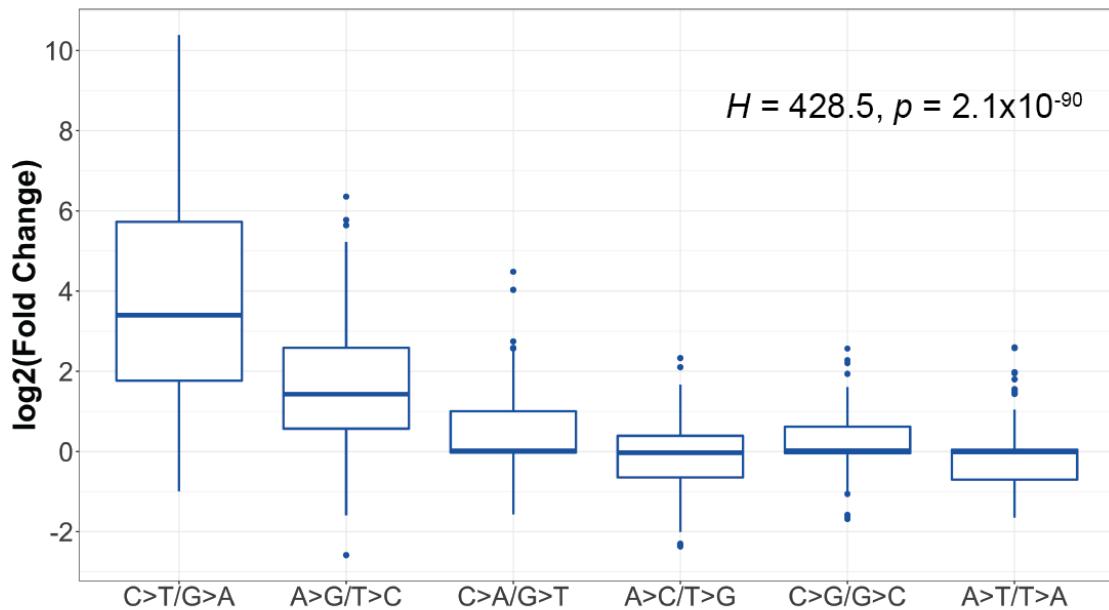


**Figure 3.7:** Assessment of formalin-induced sequence artifacts in FFPE specimens. (A) Comparison of fraction of base changes in blood and FFPE specimens (Wilcoxon signed-rank test). Box plots show the median (horizontal bar within) and IQR of fraction of base changes for different types of base changes, with whiskers representing the range of data not exceeding 1.5x the IQR and circles indicating outliers. (B) Box plots showing square root-transformed fraction of base changes on the Y-axis.

**Table 3.3:** Summary statistics of fraction of base changes in blood and FFPE specimens.

Type of Base Changes	Blood		FFPE Tumour		FC <sup>†</sup>
	Median	Range	Median	Range	
C>T/G>A	$8.9 \times 10^{-5}$	$0\text{--}2.6 \times 10^{-3}$	$2.0 \times 10^{-3}$	$0\text{--}8.6 \times 10^{-2}$	23
A>G/T>C	$1.2 \times 10^{-4}$	$0\text{--}1.3 \times 10^{-3}$	$3.7 \times 10^{-4}$	$0\text{--}5.0 \times 10^{-3}$	3.1
C>A/G>T	$6.0 \times 10^{-5}$	$0\text{--}1.8 \times 10^{-4}$	$6.0 \times 10^{-5}$	$0\text{--}2.6 \times 10^{-3}$	1.0
A>C/T>G	$2.4 \times 10^{-4}$	$5.9 \times 10^{-5}\text{--}5.3 \times 10^{-4}$	$1.8 \times 10^{-4}$	$5.8 \times 10^{-5}\text{--}1.4 \times 10^{-3}$	0.77
C>G/G>C	$1.2 \times 10^{-4}$	$0\text{--}2.4 \times 10^{-4}$	$1.2 \times 10^{-4}$	$0\text{--}4.8 \times 10^{-4}$	1.0
A>T/T>A	$6.0 \times 10^{-5}$	$0\text{--}8.4 \times 10^{-4}$	$5.9 \times 10^{-5}$	$0\text{--}8.6 \times 10^{-4}$	0.99

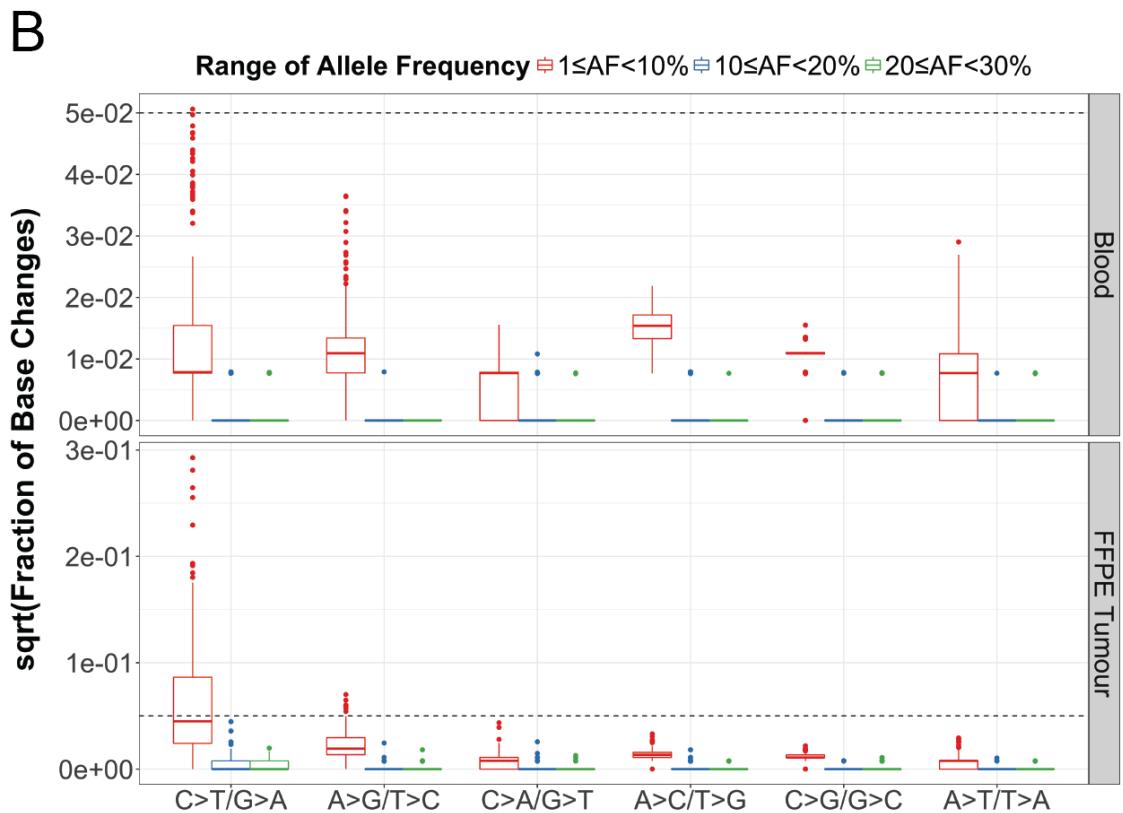
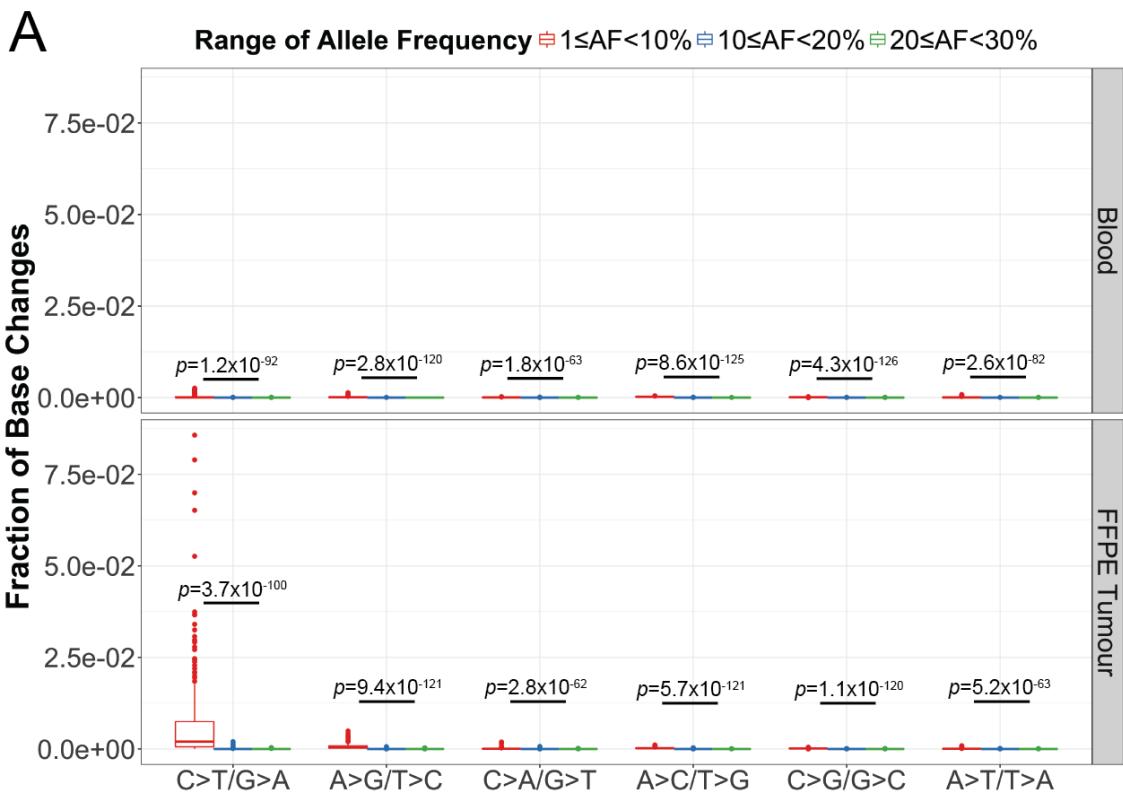
<sup>†</sup>Fold change (FC) between the median of blood and FFPE specimens.



**Figure 3.8:** Comparison of relative difference in fraction of base changes in FFPE specimens compared to blood (Kruskal-Wallis test). Relative difference was measured as  $\log_2$  fold change between fraction of base changes in blood and FFPE specimens ( $\log_2(\text{Fraction of Base Changes}_{\text{FFPE}}/\text{Fraction of Base Changes}_{\text{Blood}})$ ). Box plots show the median (horizontal bar within) and IQR of  $\log_2$  fold change for different types of base changes, with whiskers representing the range of data not exceeding 1.5x the IQR and circles indicating outliers.

**Table 3.4:** Multiple pairwise comparison of log2 fold change in fraction of base changes between blood and FFPE specimens using Dunn's test with Benjamini-Hochberg multiple hypothesis testing correction. Top values represent Dunn's pairwise  $z$  statistics, whereas bottom values represent adjusted  $p$ -value. Asterisk(\*) indicates significance level of adjusted  $p$ -value  $< 0.0001$ .

Type of Base Changes	A>C/T>G	A>G/T>C	A>T/T>A	C>A/G>T	C>G/G>C
A>G/T>C	-11.7 $4.15 \times 10^{-31}*$				
A>T/T>A	-0.399 $3.45 \times 10^{-1}$	9.57 $1.31 \times 10^{-21}*$			
C>A/G>T	-3.46 $4.00 \times 10^{-4}$	6.39 $1.52 \times 10^{-10}*$	-2.73 $3.99 \times 10^{-3}$		
C>G/G>C	-3.02 $1.73 \times 10^{-3}$	8.63 $6.76 \times 10^{-18}*$	-2.17 $1.71 \times 10^{-2}$	0.918 $1.92 \times 10^{-1}$	
C>T/G>A	-17.1 $7.78 \times 10^{-65}*$	-5.60 $1.76 \times 10^{-8}*$	-14.3 $5.10 \times 10^{-46}*$	-11.1 $1.32 \times 10^{-28}*$	-14.1 $6.46 \times 10^{-45}*$



**Figure 3.9:** Assessment of formalin-induced sequence artifacts in FFPE specimens at different ranges of allele frequency. (A) Comparison of fraction of base changes across different ranges of allele frequency (Kruskal-Wallis test). Box plots show the median (horizontal bar within) and IQR of fraction of base changes for different types of base changes, with whiskers representing the range of data not exceeding 1.5x the IQR and circles indicating outliers. (B) Box plots demonstrating square root-transformed fraction of base changes across different ranges of allele frequency. Dashed lines equal to 0.05 to indicate that the Y-axis scales are different for blood and FFPE tumour plots.

**Table 3.5:** Summary statistics of fraction of base changes in blood and FFPE specimens within 1-10% allele frequency.

Type of Base Changes	Blood		FFPE Tumour		†FC
	Median	Range	Median	Range	
C>T/G>A	$6.2 \times 10^{-5}$	$0-2.6 \times 10^{-3}$	$2.0 \times 10^{-3}$	$0-8.6 \times 10^{-2}$	33
A>G/T>C	$1.2 \times 10^{-4}$	$0-1.3 \times 10^{-3}$	$3.7 \times 10^{-4}$	$0-4.9 \times 10^{-3}$	3.1
C>A/G>T	$6.0 \times 10^{-5}$	$0-2.4 \times 10^{-4}$	$6.0 \times 10^{-5}$	$0-1.9 \times 10^{-3}$	1.0
A>C/T>G	$2.4 \times 10^{-4}$	$5.9 \times 10^{-5}-4.8 \times 10^{-4}$	$1.8 \times 10^{-4}$	$0-1.1 \times 10^{-3}$	0.77
C>G/G>C	$1.2 \times 10^{-4}$	$0-2.4 \times 10^{-4}$	$1.2 \times 10^{-4}$	$0-4.8 \times 10^{-4}$	1.0
A>T/T>A	$6.0 \times 10^{-5}$	$0-8.4 \times 10^{-4}$	$5.9 \times 10^{-5}$	$0-8.6 \times 10^{-4}$	0.99

†Fold change (FC) between the median of blood and FFPE specimens.

### 3.4 Increased age of paraffin block results in reduced amplicon yield and elevated level of C>T/G>A sequence artifacts

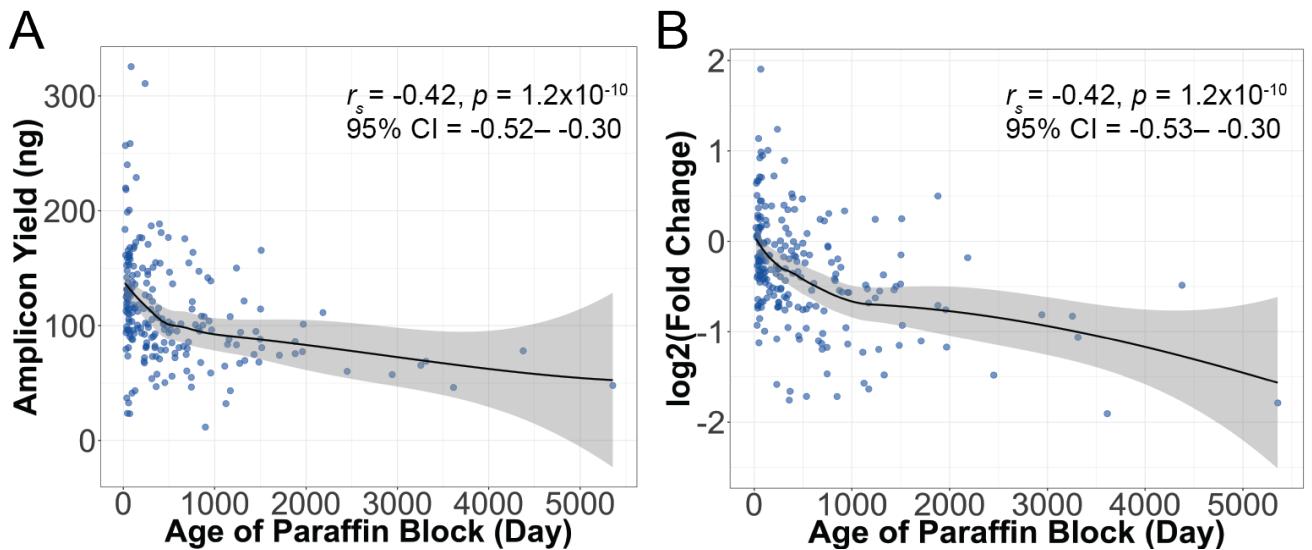
The amount of amplifiable DNA derived from FFPE specimens is dependent on the degree of fragmentation damages. Given two FFPE DNA samples of similar quantity, the sample with more extensive DNA fragmentation would yield reduced amount of PCR amplicons compared to the less fragmented sample. Several studies reported increased fragmentation damages in DNA isolated from older paraffin blocks due to longer exposure to environmental conditions. As the age of paraffin blocks in our study ranges from 18 to 5356 days, we hypothesized that older paraffin blocks would result in more extensively fragmented DNA, leading to reduced efficiency in amplicon enrichment. Inspection of the relationship between amplicon yield and age of paraffin block demonstrated a moderate, negative correlation (Spearman's rank correlation,  $r_s = -0.42$ , 95% CI =  $-0.52- -0.30$ ,  $p = 1.2 \times 10^{-10}$ ; Figure 3.10A), suggesting that DNA extraction from older paraffin blocks tend to yield lower amount of amplicons. Because the amount of DNA input for production of amplicons varies across specimens in our study design, a representation of efficiency in amplicon enrichment would be the log2 fold change between DNA input and amplicon yield. Thus, we assessed the correlation between log2 fold change and the storage time of FFPE blocks. Similarly, there was a moderate, negative correlation between log2 fold change and age of paraffin block (Spearman's rank correlation,  $r_s = -0.42$ , 95% CI =  $-0.53- -0.30$ ,  $p = 1.2 \times 10^{-10}$ ; Figure 3.10B), implying that production of amplicons is less efficient in FFPE DNA extracted from older paraffin blocks, which is likely caused by more substantial DNA fragmentation.

There are also studies that revealed increased frequency of sequence artifacts in FFPE DNA that are exceedingly fragmented. As DNA fragmentation results in reduced template DNA for PCR amplification, this leads to a higher probability for enrichment of sequence artifacts. Our previous

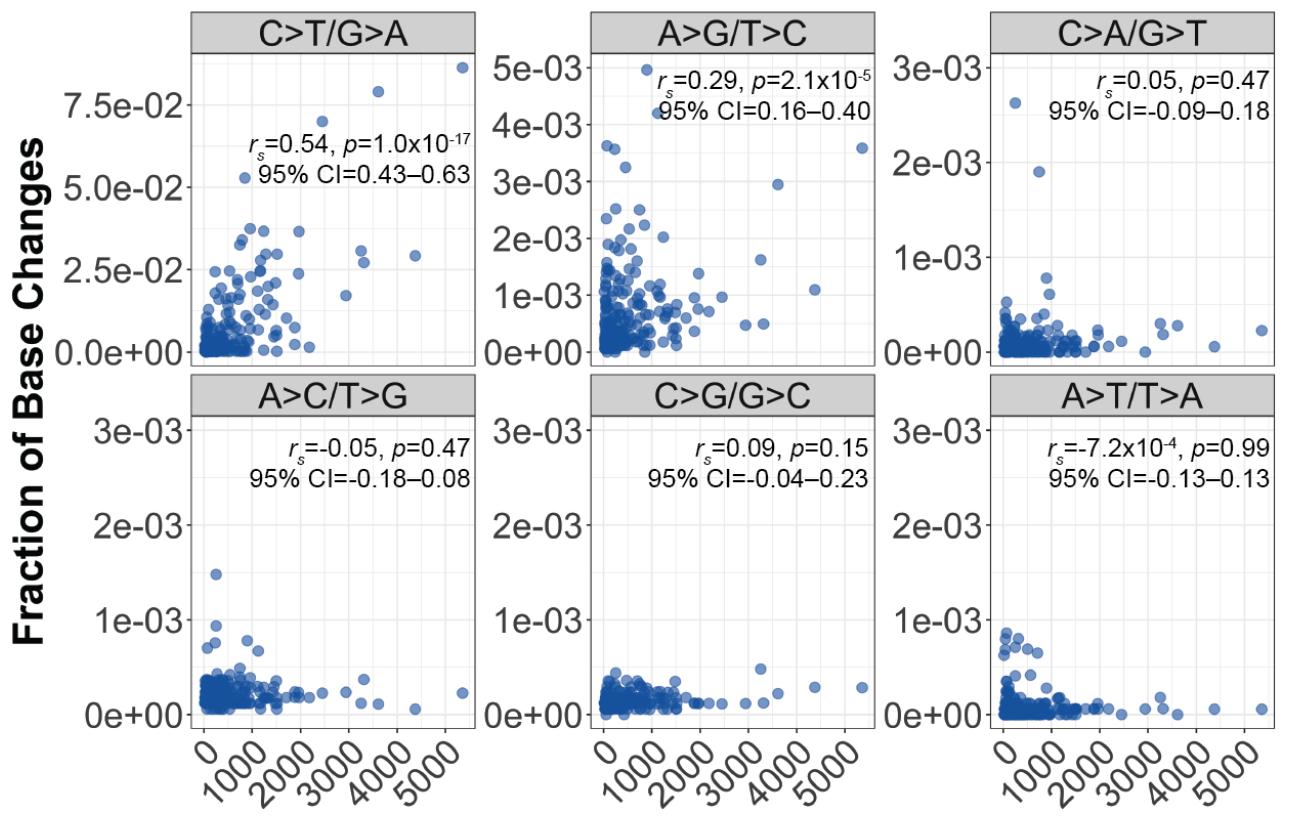
evaluation indicated that older paraffin blocks were associated with lower efficiency in amplicon enrichment, which is possibly due to increased fragmentation damages in the extracted DNA. This leads to our hypothesis that older paraffin blocks would yield elevated levels of sequence artifacts, particularly C>T/G>A transitions, which are the most prominent type of formalin-induced base modifications. To address our hypothesis, we assessed the relationship between fraction of base changes and age of paraffin blocks for different types of base changes (Figure 3.11). There was a moderate, positive correlation between fraction of C>T/G>A transitions and age of paraffin block (Spearman's rank correlation,  $r_s = 0.54$ , 95% CI = 0.43–0.63,  $p = 1.0 \times 10^{-17}$ ). We also noted a positive correlation between fraction of A>G/T>C and age of paraffin block (Spearman's rank correlation,  $r_s = 0.29$ , 95% CI = 0.16–0.40,  $p = 2.1 \times 10^{-5}$ ), albeit a weak one. As previously mentioned, it is possible that SNVs with low allele frequencies as a result of clonal heterogeneity were not removed by our variant calling pipeline, resulting in the observed weak correlation for A>G/T>C. As for transversion base changes (i.e. C>A/G>T, A>C/T>G, C>G/G>C, and A>T/T>A), no significant correlations with age of paraffin block were observed (Spearman's rank correlation,  $p < 0.05$ ). These findings reveal that increased detection of sequence artifacts, especially the common C>T/G>A changes in FFPE specimens, is associated with long term storage of FFPE blocks.

We subsequently examined how pre-sequencing variables such as age of paraffin block and efficiency in amplicon enrichment correlate with sequencing metrics, which include average per base coverage (normalized to account for library size), percentage of on-target alignments, and fraction of C>T/G>A changes (Table 3.6). This assessment would provide insight on how pre-sequencing variables can affect sequencing results, thereby facilitating sample selection before sequencing. We noted a moderate, negative correlation between average per base coverage and age of paraffin block (Spearman's rank correlation,  $r_s = -0.47$ , 95% CI = -0.57– -0.36,  $p = 4.7 \times 10^{-7}$ ), and a weak, negative correlation between percentage of on-target aligned reads and age of paraffin block (Spearman's rank correlation,  $r_s = -0.35$ , 95% CI = -0.46– -0.23,  $p = 8.2 \times 10^{-3}$ ). Conversely, we observed a moderate, positive correlation between average per base coverage and efficiency in amplicon enrichment (Spearman's rank correlation,  $r_s = 0.52$ , 95% CI = 0.42–0.61,  $p = 2.3 \times 10^{-11}$ ), and a weak, positive correlation between percentage of on-target aligned reads and efficiency in amplicon enrichment (Spearman's rank correlation,  $r_s = 0.35$ , 95% CI = 0.22–0.45,  $p = 2.9 \times 10^{-5}$ ). Since efficiency in amplicon enrichment is inversely correlated with storage time of FFPE blocks, opposing correlations with sequencing metrics were expected for both pre-sequencing variables. Furthermore, there was also a moderate, negative correlation between fraction of C>T/G>A and efficiency in amplicon enrichment (Spearman's rank correlation,  $r_s = -0.55$ , 95% CI = -0.64– -0.45,  $p = 2.0 \times 10^{-20}$ ). As reduced efficiency in amplicon enrichment is an indicator for low amount of template DNA, the consequent increase in C>T/G>A changes is the outcome of stochastic enrichment of sequence artifacts. Together, these results reveal that pre-sequencing variables such as

age of paraffin block and efficiency in amplicon enrichment could be predictors of sequencing metrics, in which older FFPE blocks are more likely to yield lower efficiency in amplicon enrichment, leading to poorer sequencing results and increased prevalence of artifactual C>T/G>A transitions.



**Figure 3.10:** Scatter plots showing (A) amplicon yield and (B) efficiency in amplicon enrichment, which is represented by the log<sub>2</sub> fold change between the amount of DNA input for producing amplicons and amplicon yield, in relation to age of paraffin blocks (Spearman's rank correlation). Solid lines represent locally weighted smoothing (LOESS) curves, with shaded bands indicating 95% confidence interval of the LOESS curves.



**Figure 3.11:** The relationship between fraction of base changes and age of paraffin block for different types of base changes (Spearman's rank correlation).

**Table 3.6:** Spearman's rank correlation between pre-sequencing variables and sequencing metrics. Top values represent Spearman's *rho* and 95% confidence interval in brackets, whereas bottom values represent *p*-value. Asterisk(\*) indicates significance level of *p*-value < 0.05.

Variable	Enrichment Efficiency <sup>†</sup>	Age of Paraffin Block (Day)	Fraction of C>T/G>A	Average Per Base Normalized Coverage
Age of Paraffin Block (Day)	-0.42 (-0.53– -0.30) $9.3 \times 10^{-10}*$			
Fraction of C>T/G>A	-0.55 (-0.64– -0.45) $2.0 \times 10^{-20}*$	0.54 (0.43–0.63) $1.0 \times 10^{-17}*$		
Average Per Base Normalized Coverage	0.52 (0.42–0.61) $2.3 \times 10^{-11}*$	-0.47 (-0.57– -0.36) $4.7 \times 10^{-7}*$	-0.80 (-0.84– -0.75) $7.5 \times 10^{-17}*$	
On-target Aligned Reads (%)	0.34 (0.22–0.45) $2.9 \times 10^{-5}*$	-0.35 (-0.46– -0.23) $8.2 \times 10^{-3}*$	-0.57 (-0.65– -0.47) $4.2 \times 10^{-8}*$	0.73 (0.66–0.79) $3.1 \times 10^{-58}*$

<sup>†</sup>log2 fold change between DNA input for amplicon enrichment and amplicon yield.

### 3.5 Discussion

Our study design consists of paired tumour-blood samples, which enabled us to extensively assess the degree of formalin-induced DNA damages in FFPE specimens through comparison with blood specimens.