



A nanoscale quality control framework for assessing FFPE DNA integrity in cancer research†

Cite this: *Nanoscale Horiz.*, 2025, 10, 1692

Received 24th March 2025,
Accepted 29th May 2025

DOI: 10.1039/d5nh00176e

rsc.li/nanoscale-horizons

Zixuan Huang,^{‡,ab} Yunpei Si,^{‡,ab} Yi Zhang,^{‡,ab} Zicheng Huang,^c Xuehao Xiu,^{ab}
Yunshan Wang,^d YuDong Wang,^{*c} Chunhai Fan^{id}^e and Ping Song^{id}^{*ab}

Formalin-fixed paraffin-embedded (FFPE) samples are widely used in cancer research and clinical diagnostics for preserving tissue morphology and enabling long-term storage. However, FFPE-induced DNA degradation, crosslinking, and inconsistent quality control significantly hinder their utility in molecular analyses. In this study, we established a robust nanoscale quality control (QC) framework incorporating gel electrophoresis and quantitative polymerase chain reaction (qPCR) to evaluate DNA integrity in clinical tissue FFPE samples. Our findings demonstrate a quantifiable inverse correlation between the degree of DNA fragmentation and amplification efficiency in FFPE samples. Further analysis of 26 single nucleotide polymorphism loci using targeted next-generation sequencing demonstrated substantial improvements in DNA integrity after enzymatic repair. A comparative whole-exome sequencing analysis of endometrial carcinoma samples with different archival durations demonstrated significantly increased damage levels across multiple genomic features in long-term stored specimens, highlighting the cumulative impact of archival duration. These findings emphasize the detrimental effects of prolonged storage on FFPE DNA quality. Our QC framework enables effective sample stratification, facilitating the selection of high-integrity specimens for sequencing and guiding heavily degraded samples toward targeted short-amplicon assays. This strategy provides a standardized approach to assess the integrity of FFPE-derived DNA, supporting accurate and reproducible use of archival biospecimens in clinical genomics.

New concepts

We present a novel nanoscale-resolved DNA integrity assessment platform that integrates qPCR, gel electrophoresis and next generation sequencing (NGS) to unravel the molecular effects of formalin-fixed paraffin-embedded (FFPE) processing and archival storage. Unlike existing studies that rely on isolated methods, our work uniquely combines macro- and nanoscale techniques to reveal formaldehyde-induced DNA fragmentation. A key innovation is the development of a rapid quality control framework using qPCR and gel electrophoresis for efficient pre-screening of archival FFPE samples, addressing a critical bottleneck in clinical genomics. Additionally, through NGS, we reveal storage-dependent biases in sequencing uniformity, variant allele frequencies, and GC-rich sequence retention, offering direct molecular evidence of DNA lesions affecting clinical sequencing. Moreover, our findings demonstrate that enzymatic repair strategies reduce base substitution artifacts while notably improving amplification efficiency at genomic sites that were previously underrepresented or undetectable prior to repair. Our work establishes a pipeline that bridges nanoscale DNA damage characterization with actionable genomic workflows, improving the reliability of FFPE-derived data in precision oncology. This approach not only enhances the use of archival biospecimens but also provides new insights into optimizing FFPE sample utility for retrospective cancer research and beyond.

1. Introduction

Clinical genetic research heavily relies on diverse nucleic acid sources, including cell-free DNA (cfDNA),¹ fresh-frozen (FF)

^a The International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University, National Center for Translational Medicine, Shanghai, 200030, China. E-mail: songpingsjtu@sjtu.edu.cn

^b School of Biomedical Engineering, Zhangjiang Institute for Advanced Study and National Center for Translational Medicine, Shanghai Jiao Tong University, Shanghai 200240, China

^c Department of Gynecologic Oncology, The International Peace Maternity and Child Health Hospital, Shanghai Jiao Tong University, Shanghai, 200240, China. E-mail: wangyudong@shsmu.edu.cn

^d Department of Clinical Laboratory, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, 250021, Shandong, China

^e State Key Laboratory of Synergistic Chem-Bio Synthesis, School of Chemistry and Chemical Engineering, New Cornerstone Science Laboratory, Frontiers Science Center for Transformative Molecules, Zhangjiang Institute for Advanced Study and National Center for Translational Medicine, Shanghai Jiao Tong University, Shanghai, 200240, China

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d5nh00176e>

‡ These authors contributed equally to this work.

tissues, and formalin-fixed paraffin-embedded (FFPE) specimens.² Each of these sources poses unique challenges, such as degradation, chemical damage, and contamination interference. Among them, FFPE samples are particularly valuable due to their long-term storage stability and widespread availability in clinical archives.³ Globally preserved FFPE specimens constitute an extensive repository, offering unique research potential for retrospective analyses in precision oncology and molecular epidemiology.^{4–8} However, the use of FFPE-derived DNA is significantly hindered by formalin-induced damage.^{9–11} During fixation, formaldehyde causes chemical modifications, including DNA–protein crosslinks, cytosine deamination (which leads to artifactual C > T mutations), and oxidative base lesions.^{12–15} These modifications not only reduce nucleic acid extraction efficiency but also compromise the accuracy of downstream analyses, such as quantitative PCR (qPCR) and next-generation sequencing (NGS).^{16–20} The paraffin embedding process further exacerbates DNA degradation due to heat and dehydration, resulting in fragmented and damaged DNA with non-uniform ends, which complicates library preparation and sequencing.^{21–23}

FFPE DNA degradation poses significant challenges over time, particularly in archival storage settings.^{24–26} Even under controlled conditions, progressive fragmentation and depurination lead to shorter amplifiable fragments,²⁷ which impede PCR amplification efficiency and sequencing uniformity.^{28,29} Research indicates that DNA integrity declines substantially after years of storage, with FFPE samples stored for over 7 years frequently failing to meet quality thresholds for reliable genomic analysis.³⁰ This degradation manifests as reduced library yields, increased shifts in variant allele frequencies (VAFs), and biases in GC-rich sequence retention.³¹ Despite these challenges, recent advances in sample preparation, repair, and analytical techniques have improved the utility of FFPE samples for genomic studies, enabling researchers to extract valuable information from these historically challenging specimens.^{32,33} Enzymatic repair kits, such as PreCR, aim to restore DNA integrity by addressing base damage, including the excision of deaminated cytosines and the repair of oxidized guanine.³⁴ This repair capacity enables the recovery of PCR-amplifiable templates from degraded DNA, thereby improving the fidelity of downstream genomic analyses by mitigating FFPE-induced sequencing artifacts.³⁵

FFPE processing and storage introduce variable degrees of DNA damage, necessitating thorough evaluation prior to genomic applications. Therefore, we employed gel electrophoresis and qPCR to systematically compare DNA extracted from FF and FFPE specimens, establishing a standardized QC framework for the rapid assessment and screening of FFPE samples. To further elucidate the impact of DNA degradation and repair on sequencing accuracy, we performed targeted NGS on both untreated FFPE samples and those treated with a commercial DNA repair kit, focusing on 26 single nucleotide polymorphism (SNP) loci. Additionally, we conducted whole-exome sequencing (WES) on FFPE samples archived for different durations to evaluate the effects of DNA fragmentation and mutation profiles on sequencing performance. This multi-tiered analytical

approach enables the efficient screening of FFPE samples, facilitating optimal resource allocation for downstream applications. Specifically, high-integrity samples are prioritized for applications requiring long DNA fragments, such as whole-exome sequencing (WES) and gene fusion detection,³⁶ while severely degraded samples are directed toward targeted short-amplicon assays. By bridging the gap between the characterization of FFPE-induced damage and the implementation of practical QC strategies, our study supports the reliable utilization of archived biospecimens in clinical genomics and contributes to improving the accuracy and reproducibility of FFPE-based genomic research.

2. Materials and methods

2.1. DNA extraction and quantification

A total of 33 FFPE samples of endometrial carcinoma (EC) tissue and 4 paired FF EC tissue samples were collected from the International Peace Maternity & Child Health Hospital of the China Welfare Institute; and 11 FFPE samples of hepatocellular carcinoma (HCC) tissue were obtained from Shandong Provincial Hospital affiliated with Shandong First Medical University. All samples were prepared between 2018 and 2024 and subsequently analyzed. Genomic DNA (gDNA) was extracted from these samples using the QIAamp DNA FFPE tissue kit (Qiagen) according to the manufacturer's protocol. FFPE repair was performed using the PreCR repair mix (NEB, M0309). The quality of the extracted DNA was assessed using a fluorometric assay (Qubit 4.0 Fluorometer), and the concentrations were subsequently adjusted to 20 ng μL^{-1} .

2.2. PCR amplification of varying lengths

To assess the quality of the extracted DNA, we employed singleplex qPCR to amplify six SNPs and multiplex qPCR for targeted library preparation followed by NGS. This approach enabled us to evaluate DNA integrity, amplifiability, and suitability for downstream NGS applications, ensuring high-quality and reliable sequencing results.

The qPCR analysis was conducted on a CFX96 Real-Time PCR Thermal System (Bio-Rad) with a reaction volume of 10 μL , comprising 5 μL of 2 \times SYBR Green master mix, 1 μL of 4 μM forward primer, 1 μL of 4 μM reverse primer, 2 μL of nuclease-free water, and 1 μL of extracted gDNA. The qPCR was initiated at 95 $^{\circ}\text{C}$ for 2 min, followed by a thermal cycle consisting of denaturation for 10 s at 95 $^{\circ}\text{C}$, annealing, and extension at 60 $^{\circ}\text{C}$ for 30 s.

2.3. Gel electrophoresis of extracted DNA

To verify DNA integrity, we employed agarose gel electrophoresis using a standardized protocol. A 1% agarose gel was prepared by dissolving 1 g of agarose powder in 100 mL of 1 \times TAE buffer, which was heated in a microwave until fully dissolved. After cooling to approximately 50 $^{\circ}\text{C}$, GelRed dye was added at a final concentration of 1 \times , and the mixture was poured into a gel mold equipped with a comb for well formation. The gel was allowed to

solidify at room temperature for 30 min before being transferred to the electrophoresis tank. Subsequently, 10 μL of extracted DNA samples were mixed with 2 μL of 6 \times loading buffer (6 \times Ficol Gel Loading Buffer III) and loaded alongside a 50–10 000 bp molecular weight ladder. Electrophoresis was performed at 100 V for 60 min in TAE buffer until the dye front migrated sufficiently. The gel was then visualized under UV light using a documentation system to assess the band size and intensity.

Additionally, denaturing polyacrylamide gel electrophoresis (PAGE) was conducted using a 10% denaturing gel. The gel was prepared by dissolving 20 g of urea in 10 mL of 5 \times TBE buffer using sonication. Following this, 100 μL of 10% ammonium persulfate and 5 μL of TEMED were added to initiate polymerization, ensuring a final TBE concentration of 1 \times . The gel was cast and allowed to polymerize at room temperature. For sample preparation, 5 μL of extracted DNA was mixed with 5 μL of 2 \times urea-based denaturing sample buffer and heated at 95 $^{\circ}\text{C}$ for 5 min. After denaturation, the samples were stored on ice until loading onto the gel. Electrophoresis was carried out at 120 V in 1 \times TBE buffer at room temperature, with progress monitored as the samples migrated through the denaturing gel matrix.

2.4. Library preparation and data analysis

For WES, gDNA was fragmented into 180–280 bp segments through random shearing. These fragments underwent end repair, A-tailing, and Illumina adapter ligation, followed by PCR amplification and size selection. Hybridization capture was performed using biotin-labeled probes and streptavidin-coated magnetic beads to isolate exonic regions, with non-hybridized fragments removed through washing. The captured

libraries were then enriched *via* PCR and subjected to quality control, which included Qubit quantification, real-time PCR, and bioanalyzer size distribution analysis.

Libraries were sequenced on Illumina platforms using a paired-end 150 bp (PE150) strategy after pooling based on concentration and data requirements. Raw sequencing data underwent quality trimming with Fastp, which removed read pairs containing adapter contamination (>10 nucleotide alignment with $\leq 10\%$ mismatches), reads with $>10\%$ uncertain bases, or those with $>50\%$ low-quality bases (Phred score < 5). Cleaned reads were aligned to the GRCh38 reference genome using BWA-MEM v0.7.17. Base quality recalibration was implemented through GATK 4.0.2.1 in a two-step workflow: first, generating covariate tables with BaseRecalibrator using known variants, followed by the application of recalibration parameters genome-wide using ApplyBQSR. Variant calling was conducted with GATK HaplotypeCaller, and low-confidence variants were excluded based on filtering criteria (GQ < 20 , DP < 10 or > 500 , QUAL < 30) prior to downstream analysis.

3. Results and discussion

3.1. Assessment of DNA integrity in FFPE samples

DNA samples extracted from FF cell lines and FFPE samples of HCC and EC, prepared between 2018 and 2024, were studied (Fig. 1a). FFPE tissue samples represent a critical resource for retrospective studies in cancer research, bridging clinical data with long-term patient outcomes. However, formalin fixation, paraffin embedding, prolonged storage and harsh extraction can cause DNA degradation, leading to single-strand breaks (SSBs), double-strand breaks (DSBs), DNA-protein

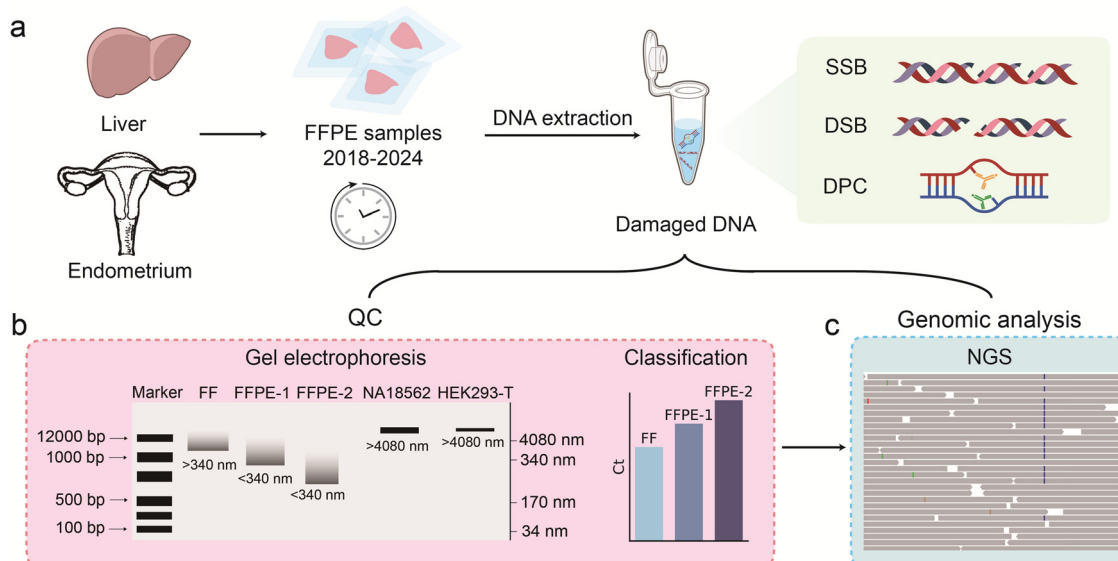


Fig. 1 Schematic illustration of DNA quality assessment from FF and FFPE samples. (a) DNA extraction from FF and FFPE samples prepared in different years from cancer tissue samples using the QIAamp DNA FFPE tissue kit. (b) Evaluation of DNA fragmentation in FFPE samples *via* gel electrophoresis and assessment of amplicon amplifiability of varying lengths by qPCR. FFPE-1 and FFPE-2 denote samples with different storage durations; FFPE-2 represents an older sample with more extensive DNA degradation. (c) Bioinformatic analysis of FFPE samples by NGS, highlighting the impact of FFPE-induced DNA damage on sequencing outcomes.

crosslinks (DPCs), and other forms of damage, such as oxidative damage.^{31,32,37} Such DNA damage can significantly compromise downstream molecular analyses (*e.g.*, NGS), leading to inaccurate or unreliable results.

In this study, we employed a combined approach of qPCR and gel electrophoresis for the preliminary assessment of FFPE samples (Fig. 1b). The gDNA extracted from human cell lines including NA18562, NA18537, HEK293T, and HeLa was used as the reference for comparisons. Given that DNA measures 0.34 nm per base pair, gel electrophoresis reveals that DNA from FF references typically exceeds 1000 base pairs (approximately 340 nm), whereas DNA from FFPE samples is shorter than 340 nm.³⁸ This fragmentation was further corroborated by qPCR analysis, the reference DNA typically shows decreasing Ct values with increasing amplicon length. In contrast, damaged DNA extracted from FFPE samples does not follow this trend and often yields higher Ct values as the amplicon length increases.

To further evaluate the impact of FFPE-induced DNA damage on sequencing outcomes, we performed NGS analysis for FFPE samples (Fig. 1c). The results revealed distinct patterns of DNA degradation, with older samples exhibiting higher levels of fragmentation and artifactual mutations, which were consistent with prolonged archival storage. These findings underscore the importance of implementing a standardized QC workflow to ensure the reliability of FFPE-derived DNA for genomic analyses.

3.2. Evaluation of DNA fragmentation and amplification efficiency in FFPE samples

To investigate the extent of DNA damage in FFPE samples, we compared commercially available gDNA references (NA18562 and NA18572), FF and FFPE samples stored for 4 years. We evaluated the amplification efficiency and fragmentation patterns of DNA across 3 SNPs (rs206781, rs2638145, and rs2510152, see Tables S1–S3, ESI[†]) with amplicon lengths ranging from 50 to 260 bp using qPCR and gel electrophoresis. This approach allowed us to assess the impact of FFPE processing on DNA quality and its implications for downstream analyses.

For intact reference DNA samples, the qPCR amplification Ct value decreases as the amplicon length increases due to enhanced fluorescence signal accumulation per cycle in longer fragments. Consequently, the Δ Ct between the Ct value of the long amplicon and the short amplicon is negative. In FF samples, Δ Ct values ($C_{t_{260\text{ bp}}} - C_{t_{50\text{ bp}}}$) ranged from -2.0 to -0.7 for the 3 SNPs, closely matching those of the reference NA18562 (-2.0 to -1.0), both exhibiting a consistent trend of decreasing Ct with increasing amplicon length. In contrast, FFPE samples consistently exhibited an increase in Ct values as amplicon length increased from 50 bp to 260 bp, with final Ct increases ranging from 3.9 to 5.5 for these loci (Fig. 2a), indicating substantial DNA fragmentation.

This divergence reflects the limited availability of FFPE DNA templates for long template amplification, highlighting the fragmented nature of FFPE-derived DNA. For example, at the rs206781 locus, the Ct difference between FFPE and FF samples was 2.1 cycles for the amplification of a 50 bp sequence,

suggesting that FFPE samples retained only 23.3% of the effective DNA concentration compared to FF samples. However, for a 260 bp amplicon, the Ct difference increased to 8.3 cycles, with FFPE DNA showing just 0.3% of the effective DNA concentration relative to FF DNA (Fig. 2b). This dramatic reduction in amplifiable DNA as amplicon size increases highlights the severe fragmentation of FFPE-derived DNA, which significantly impedes efficient PCR amplification.

Agarose gel electrophoresis further revealed distinct fragmentation patterns among sample types (Fig. 2c). The reference NA18572 exhibited superior DNA integrity, characterized by longer fragment lengths and well-defined band distributions. In contrast, FF samples predominantly retained high-molecular-weight DNA (>1 kb), while FFPE samples exhibited significant fragmentation, with most fragments being less than 1 kb. This progressive degradation in FFPE samples is consistent with the formalin-induced DNA damage observed in qPCR analysis, further highlighting the impact of FFPE processing on DNA quality.

Additionally, denaturing PAGE further distinguished the quality differences between FF and FFPE samples (Fig. 2d). The FF DNA exhibited discrete high-molecular-weight bands, whereas FFPE samples displayed a heterogeneous fragment distribution with significant migration of sub-100 nt fragments. This smearing pattern confirms single-stranded DNA (ssDNA) fragmentation in FFPE samples, consistent with formalin-induced crosslinking and hydrolytic damage mechanisms.

The observed fragmentation patterns directly account for the differential PCR amplification efficiencies between FF and FFPE samples. Formaldehyde fixation induces extensive DNA fragmentation through crosslinking and hydrolytic cleavage, severely limiting the availability of intact templates for long amplicons. As a result, FFPE-derived DNA exhibits increased Ct values with amplicon length, reflecting a progressive loss of amplifiable template regions. In contrast, FF samples, which retain higher molecular weight DNA, maintain efficient amplification across a range of target sizes, with decreasing Ct values for longer amplicons. The commercially available reference DNA further confirmed this trend, demonstrating amplification kinetics characteristic of intact gDNA. Collectively, these findings highlight the substantial impact of FFPE processing on DNA integrity, emphasizing the need for rigorous quality control measures to optimize the use of FFPE-derived DNA in molecular analyses.

To further investigate the impact of sample preparation time on DNA integrity, we analyzed EC FFPE tissue specimens spanning multiple archival years (2018–2024). Similarly, we employed the same FP and different reverse primers to generate amplicons of varying lengths (50 bp to 250 bp) targeting 4 SNPs (Fig. 3a and b and Tables S4–S6, ESI[†]). This approach enabled a quantitative assessment of DNA fragmentation through differential amplification efficiency. Taking rs206781 as an example, we compared Δ Ct values between 250 bp and 50 bp amplicons for FF and FFPE samples, revealing a significant correlation between Δ Ct values and archival age (Fig. 3b and Fig. S1, ESI[†]). One-way ANOVA demonstrated significant

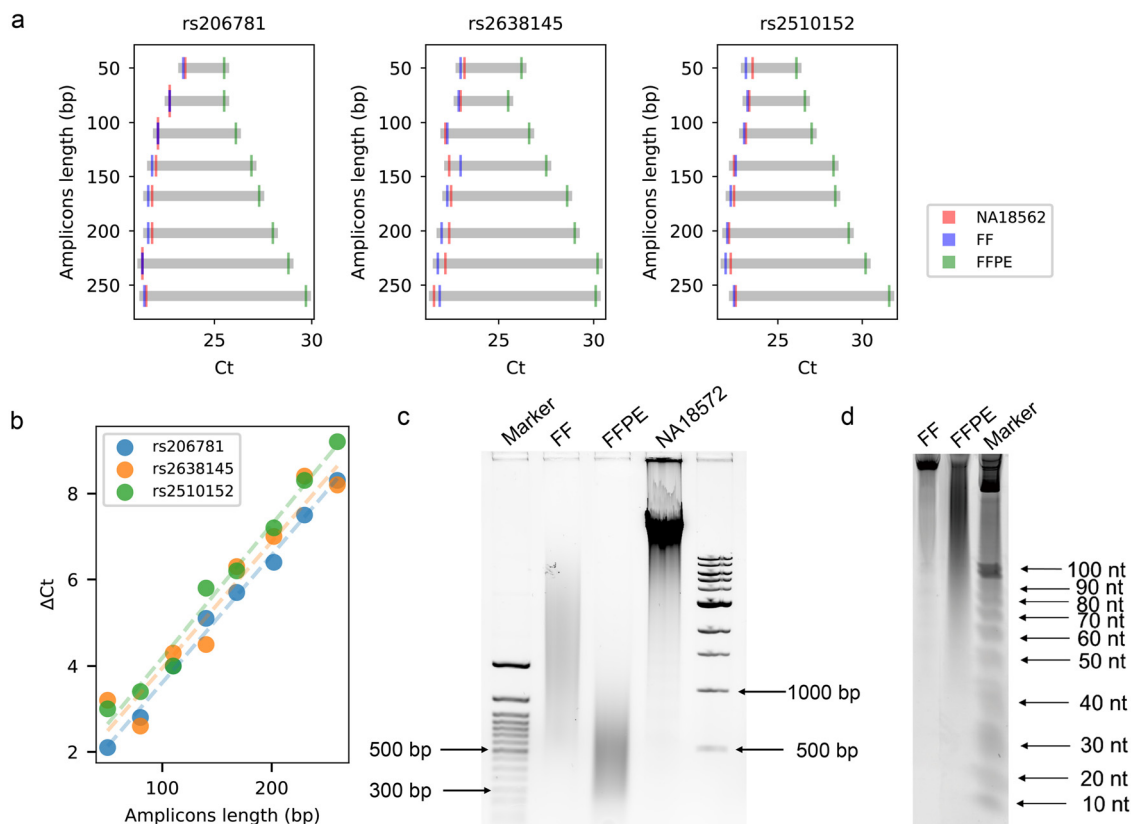


Fig. 2 Comparative analysis of DNA fragment length in FFPE and FF tissue samples preserved for 4 years. (a) qPCR analysis of NA18562, FF, and FFPE samples at multiple SNP loci using reverse primers at varying distances from the forward primer. FFPE samples exhibited increased Ct values with longer amplicons, indicating DNA fragmentation. (b) Relationship between Δ Ct values and amplicon length, showing a progressive increase in Δ Ct for FFPE samples compared to FF controls. Here, Δ Ct represents the difference in Ct values between FFPE and FF samples. (c) Agarose gel electrophoresis comparing DNA fragmentation patterns in NA18572, FF, and FFPE samples. FFPE samples showed marked fragmentation, with most fragments being less than 1 kb. (d) Denaturing PAGE analysis of FF and FFPE samples, highlighting ssDNA fragmentation in FFPE samples.

differences in Δ Ct values among groups ($F(4,32) = 20.2$, $p < 0.001$). Tukey HSD *post hoc* comparisons revealed distinct patterns of nucleic acid degradation across storage durations. References exhibited significantly lower Δ Ct values than samples from 2018 (MD = -6.48 , $p < 0.001$), 2019 (MD = -4.82 , $p < 0.001$), and 2023 (MD = -3.05 , $p = 0.025$), supporting their use as undamaged controls and highlighting the extent of damage in samples stored for more than one year. However, the difference between references and samples prepared in 2024 was not statistically significant (MD = -1.59 , $p = 0.459$), suggesting minimal damage in the most recently prepared FFPE samples. Samples prepared in 2018 demonstrated significantly higher degradation compared to 2023 (MD = 3.43 , $p < 0.001$) and 2024 samples (MD = 4.89 , $p < 0.001$). Similarly, 2019 samples showed significantly increased Δ Ct values relative to 2024 samples (MD = 3.23 , $p = 0.014$). These findings demonstrate a clear time-dependent increase in nucleic acid damage, with critical degradation occurring in samples stored beyond 6 years. Furthermore, HCC samples mirrored this degradation profile ($F(2,12) = 27.5$, $p < 0.0001$), with *post-hoc* Tukey HSD tests confirming significant Δ Ct differences between specimens prepared during 2018–2020 and 2024-prepared specimens (MD = -2.87 , $p = 0.0049$), as well as between the 2018–2020 specimens

and references (MD = -5.66 , $p < 0.0001$). These findings collectively demonstrate time-dependent nucleic acid damage in FFPE-preserved tissues across multiple cancer types.

Cross-tissue analysis of 2024-prepared FFPE specimens revealed significant Δ Ct value disparities between HCC and EC groups (permutation test: median difference = -1.218 , $p = 0.015$, Fig. 3c). This pronounced divergence suggests that HCC-associated DNA undergoes accelerated formalin-induced fragmentation during FFPE processing. Such inter-tumor variability in DNA degradation may be attributed to tumor-specific intrinsic factors combined with pre-analytical variables.³⁹

Control experiments revealed genome-wide DNA degradation in the S1-2018 specimen, with systematically elevated Δ Ct values across all tested loci (Δ Ct > 1.47) compared to reference samples (Δ Ct < -0.86) (Fig. 3d and Fig. S2–S5, ESI[†]). This pan-genomic degradation pattern suggests that the observed fragmentation results from intrinsic genome-wide degradation processes rather than sequence-specific amplification artifacts, thereby validating the reliability of qPCR for assessing global DNA integrity in archival specimens.

Agarose gel electrophoresis further validated the progressive DNA fragmentation patterns in FFPE samples (Fig. 3e). Notably, the main bands for FF samples were generally larger than

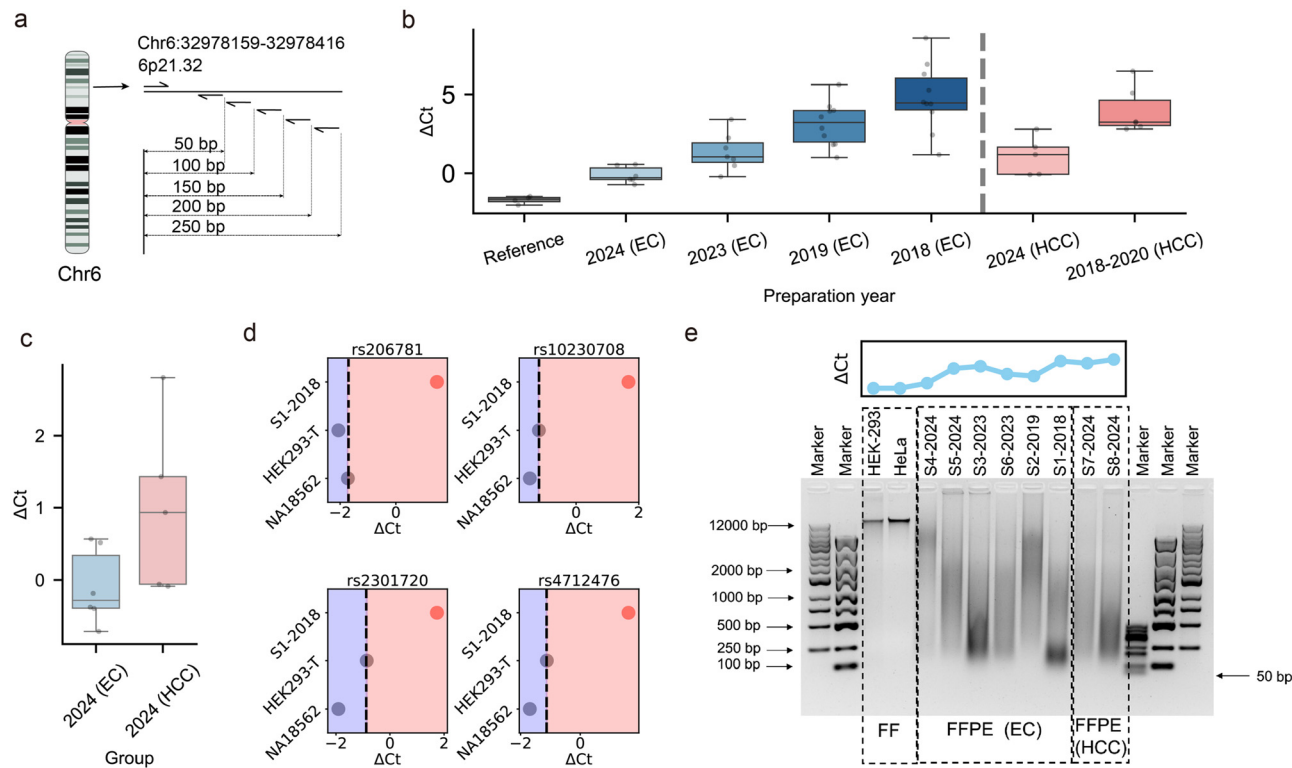


Fig. 3 Investigation of amplification efficiency and fragment lengths in different tissue samples. (a) Schematic showing amplicons with different lengths amplified using designed forward primers and different reverse primers. (b) ΔC_t values of rs206781 locus amplification in 33 EC and 11 HCC FFPE samples prepared from 2018 to 2024 years, with 4 DNA references (NA18562, NA18537, HEK293T, and HeLa). Box plots show median and interquartile range. (c) ΔC_t between 250 and 50 bp amplicons of rs206781 for EC and HCC FFPE samples prepared in 2024. (d) ΔC_t values for the same sample at different loci with amplicon lengths of 250 and 50 bp, showing uniformity across loci. (e) Agarose gel electrophoresis highlighting DNA integrity and fragmentation patterns of FF and FFPE samples.

12 000 bp, while both EC and HCC FFPE samples exhibited significant degradation and smearing. Based on the primary bands, we assessed the degradation severity as follows: S4-2024 < S2-2019 < S5-2024 < S7-2024 < S6-2023 < S8-2024 < S3-2023 < S1-2018.

This fragmentation hierarchy directly paralleled the ΔC_t rankings observed in qPCR analysis, where specimens with more severe fragmentation exhibited systematically elevated ΔC_t values. The correlation between reduced amplification efficiency (reflected by higher ΔC_t) and lower molecular weight DNA underscores the mechanistic link between strand break accumulation and polymerase accessibility limitations. These findings validate ΔC_t measurements as a robust, PCR-based QC metric for assessing DNA integrity in archival biospecimens. The approach of standardizing ΔC_t thresholds against reference, combined with gel electrophoresis validation, provides a robust framework for assessing FFPE sample suitability in downstream genomic applications. Previous methods for assessing the integrity of FFPE-derived DNA, such as NanoDrop spectrophotometry only assess DNA concentration but do not reflect DNA integrity,⁴⁰ while gel electrophoresis offers visual assessment with inherent subjectivity and variability.⁴¹ The Q-ratio method evaluates DNA integrity by comparing the qPCR amplification of long and short fragments.³⁰ Herein, we

combined gel electrophoresis and qPCR to comprehensively assess DNA integrity in clinical FFPE samples, providing a cost-effective quality control strategy for clinical workflows and uncovering time-dependent nucleic acid damage across different carcinoma tissue types.

3.3. Fragmentation analysis of FFPE samples *via* targeted NGS

To further investigate the impact of DNA fragmentation and damage in FFPE samples, we selected 26 SNPs and designed 100–180 bp amplicons for targeted NGS (Table S7, ESI[†]). We studied 5 FF and 8 FFPE samples, including 4 matched pairs, collected during 2018–2024. In NGS analysis, FFPE samples prepared in 2024 showed significantly higher read count uniformity compared to those from 2018/2019, as indicated by a lower coefficient of variation (CV; 1.165 vs. 1.535) and Gini index (0.569 vs. 0.691), with statistical significance (Mann-Whitney $U = 16.00$, $p < 0.05$ for both metrics, Fig. 4a). These results confirm that prolonged FFPE storage aggravates DNA damage, reducing available template numbers at certain loci and causing greater amplification bias.

To further elucidate time-dependent degradation effects, we analyzed a matched pair of FF and FFPE samples preserved for 4 years, as well as the FFPE samples repaired using the preCR kit. The observed disparity in read counts across samples

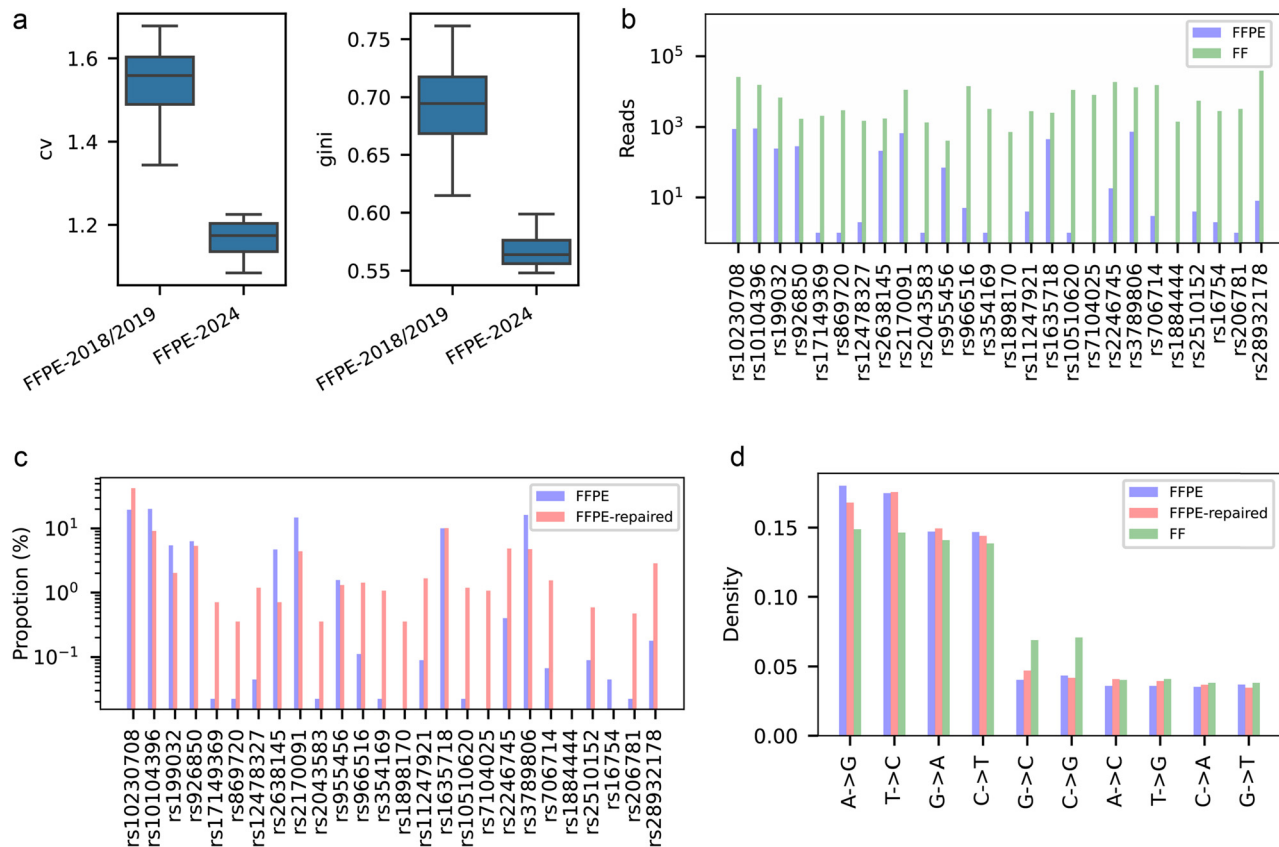


Fig. 4 Comparative analysis of FFPE, FFPE-repaired, and FF samples using targeted NGS. (a) Uniformity metrics comparison (CV and Gini index) between FFPE samples prepared in different years. (b) Reads at 26 predefined SNP loci for FFPE and FF samples, highlighting differences in sequencing read counts. (c) Read proportions of 26 predefined SNP loci in FFPE versus FFPE-repaired samples. (d) Distribution of base substitution patterns observed in the different sample types.

reflects a complex interplay of DNA fragmentation dynamics, repair-induced template redistribution, and PCR amplification biases. FF samples demonstrated significantly higher mean read counts at the 26 SNP loci, reaching 8155, compared to 172 in FFPE. The distribution of read counts across these loci was also more uniform in FF samples, with a CV of 1.46 and a Gini index of 0.65, compared to a CV of 1.53 and a Gini index of 0.69 in FFPE samples. This discrepancy reflects the reduced number of amplifiable templates caused by FFPE-associated DNA fragmentation (Fig. 4b, more details in Fig. S6, ESI†). Notably, loci with initially high read proportions, defined as the percentage of total reads across all 26 loci in FFPE samples, generally exhibited reduced proportions following PreCR treatment (Fig. 4c). For instance, the read proportion of rs10104396 decreased from 20.16% to 9.14% after PreCR treatment. In contrast, loci with negligible or undetectable initial read proportions showed substantial increases, such as rs869720, which rose from 0.02% to 3.56%, and rs1898170, which increased from 0% to 0.36%. This phenomenon may be attributed to the enhanced amplification and capture of certain loci that benefited from better repair, leading to an increase in their read proportion. Given the fixed sequencing data volume, this resulted in a relative decrease in the read proportion of other loci.

Additionally, the repair process may have altered DNA integrity, fragment length distribution, and amplification preferences, collectively contributing to the observed changes in read proportion distribution. These bidirectional shifts underscore the efficacy of the PreCR kit in repairing FFPE-induced DNA lesions. The treatment not only restored amplifiable DNA, enabling the detection of previously undetectable loci, but also enhanced the amplification of underrepresented loci, thereby improving the overall quality of genomic analysis.

A comparative analysis of mutation patterns across FFPE, FFPE-repaired, and FF samples reveals critical insights into formalin-induced artifacts and repair efficacy (Fig. 4d). The most prevalent mutation observed was A > G, with proportions of 18.02% in FFPE, 16.81% in FFPE-repaired, and 14.88% in FF samples. While A > G transitions are traditionally attributed to sequencing errors rather than formalin damage, the elevated rate in FFPE compared to FF suggests potential compounding effects of formalin fixation on error-prone sequencing contexts. This trend aligns with reports of oxidative damage exacerbating sequencing inaccuracies in archival samples.

Notably, the C > T transition, a well-documented formalin artifact,^{32,42,43} decreased progressively from 14.68% in FFPE to 14.40% in FFPE-repaired and 13.85% in FF. This gradation demonstrates the partial restoration of DNA integrity through

preCR repair, bridging the molecular disparity between FFPE and FF specimens. The repair process appears particularly effective against deamination damage, though complete normalization to FF levels remains elusive. Interestingly, G > A mutations exhibited minimal variation across sample types, ranging from 14.08% to 14.94%, potentially indicating mechanisms that are independent of fixation damage. Similarly, T > C mutations displayed a range of 14.64% to 17.56%, also suggesting mechanisms that are not influenced by fixation damage.

3.4. Impact of FFPE archival duration on WES quality and genomic analysis

To investigate the effect of FFPE preparation time on NGS library quality and mutation analysis, we selected two FFPE samples prepared in different years (S1-2018 and S5-2024) for WES analysis (Fig. S7–S9, ESI[†]). Despite shearing DNA fragments to lengths between 180 bp and 280 bp during library preparation, the 2018-prepared FFPE sample (S1-2018) exhibited a lower average fragment length of 215 bp compared to 225 bp for the sample prepared in 2024 (S5-2024) (Fig. 5a). This indicates that the 2018-prepared FFPE sample (S1-2018) underwent more severe fragmentation, consistent with previous qPCR and gel electrophoresis results.

To further investigate the impact of DNA fragmentation, we visualized read alignment at rs206781 using IGV (Fig. 5b). The results revealed that S5-2024 contained a higher number of

longer reads at this locus compared to S1-2018. This discrepancy likely explains why Ct values in multi-amplicon length qPCR increased more rapidly for S1-2018 as amplicon length increased, highlighting the challenge of amplifying longer DNA templates in older FFPE samples. These findings underscore the progressive degradation of DNA in long-stored FFPE samples, which poses significant challenges for applications requiring high-quality sequencing data.

Furthermore, we performed mutation analysis on both samples. After filtering out common SNPs and known COSMIC mutations, S1-2018 exhibited a higher mutation density at VAF ranges of 35.8–56.0%, while S5-2024 showed greater density at VAF < 35.8%. (Fig. 5c). One possible explanation for this pattern is the accumulation of formalin-induced artifacts over time, particularly deamination events that lead to C > T transitions. These artifacts tend to manifest at higher VAFs due to their widespread occurrence across DNA molecules.⁴⁴ Taken together, these findings highlight the compounded effects of prolonged storage on both DNA integrity and mutation profiles in FFPE samples.

To further elucidate the impact of storage duration on sequencing performance, we analyzed differences in read density between S1-2018 and S5-2024 under identical GC content and sequencing depth conditions (Fig. 5d). Our results demonstrated that S1-2018 exhibited higher read density than S5-2024 in low-depth regions (50–150 reads) with GC content ranging

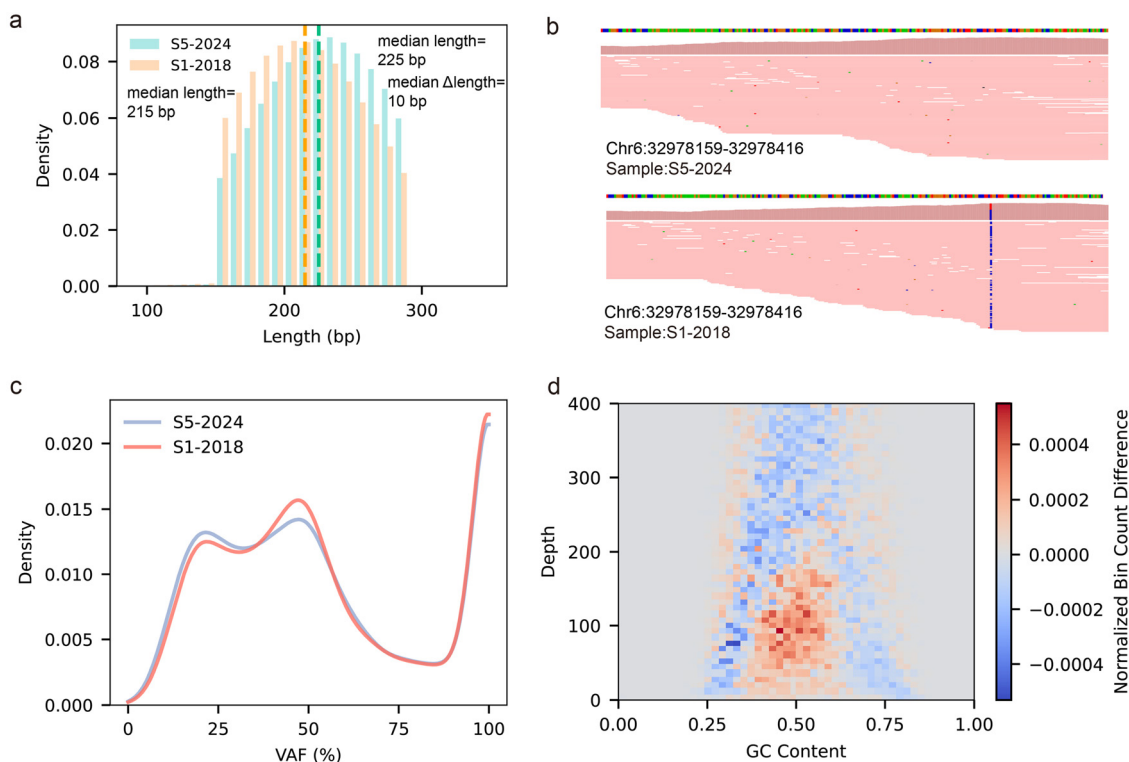


Fig. 5 WES analysis of FFPE samples stored for different durations. (a) Distribution of read lengths for FFPE samples S5-2024 and S1-2018, sequenced on the Illumina NovaSeq with paired-end 150 bp reads. (b) Visualization of read alignment near SNP rs206781 for both S5-2024 and S1-2018. (c) VAF distribution of detected mutations in S5-2024 and S1-2018. (d) Density differences between samples S1-2018 and S5-2024 under identical GC content and depth conditions.

from 0.4 to 0.6. However, across a broader depth range (150–400 reads), S1-2018 consistently displayed lower read density. This pattern likely reflects the more extensive DNA damage in S1-2018, which reduces amplification efficiency during library preparation and consequently results in lower sequencing coverage in mid-to-high depth regions. Conversely, the relatively higher read density observed for S1-2018 in low-depth regions may be attributed to the preferential sequencing of shorter, more degraded DNA fragments, which fail to contribute meaningfully to deeper coverage. Moreover, S5-2024 exhibited higher read density in regions with lower GC content, a finding consistent with previous studies indicating that DNA with low GC content is more susceptible to degradation over extended storage periods.⁴⁵ This observation supports the hypothesis that prolonged storage selectively exacerbates the degradation of AT-rich regions, likely due to their lower thermodynamic stability and increased susceptibility to hydrolytic damage.

Overall, these findings provide valuable insights into how FFPE storage duration influences DNA integrity and sequencing outcomes. Specifically, prolonged storage appears to intensify GC-biased degradation while also altering fragment length distribution and sequencing depth profiles. These results underscore the critical importance of accounting for storage duration when interpreting sequencing data from archival FFPE samples, particularly in the context of mutation analysis and genomic profiling.

4. Conclusions

This study provides a comprehensive evaluation of FFPE-induced DNA degradation and its implications for genomic analyses, addressing critical challenges in the utilization of archival biospecimens. By integrating gel electrophoresis and qPCR, we established a robust QC framework to assess DNA integrity and optimize resource allocation for downstream applications. Our findings highlight the progressive nature of FFPE degradation, with older samples exhibiting severe fragmentation, reduced amplification efficiency, and altered mutation profiles. Enzymatic repair strategies reduce base substitution artifacts and rescue amplification at insufficiently amplified loci. WES analysis revealed that prolonged storage intensifies GC-biased degradation and increases mutation densities, particularly in AT-rich regions, emphasizing the importance of accounting for storage duration in genomic studies. The multi-tiered QC approach developed in this study enables efficient screening of FFPE samples, ensuring optimal utilization of high-integrity specimens for WES while directing degraded samples toward targeted assays. These advancements bridge the gap between FFPE damage characterization and practical QC strategies, enhancing the reliability of archival biospecimens in precision oncology and molecular epidemiology. Future research should focus on refining repair techniques and developing standardized QC metrics to further improve the accuracy and reproducibility of FFPE-based genomic studies.

Author contributions

Zixuan Huang: investigation, methodology, formal analysis, writing original draft, and writing – review & editing; Y. Si: investigation, methodology, formal analysis, and writing – review & editing; Y. Zhang: investigation, formal analysis and writing – review & editing; X. Xiu: writing – review & editing and methodology; P. Song: conceptualization, data analysis, writing and editing, and supervision. Y. Wang and Zicheng Huang provide clinical samples and editing; Y.D. Wang and C. Fan: clinical discussion, data analysis and writing manuscript.

Data availability

Supplementary data associated with this article are available in the ESI.†

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the National Key R&D Program of China (No. 2022YFF1201800), the National Natural Science Foundation of China (No. 22174094, 22404109), the Fundamental Research Funds for the Central Universities (YG2023QNA33), the major difficult diseases of Chinese and Western clinical cooperation construction project of the National Health Commission of the People's Republic of China (ZXXTQJ-2024), and the Shanghai Science and Technology Committee (24Y22800300, YDZX20223100003006, and 22XD1403500). This work was supported by the Young Leading Scientists Cultivation Plan of the Shanghai Municipal Education Commission (ZXWH1082101), the Chun-Tsung Program (2024-02-05), the Shanghai Leading Talent Program of Eastern Talent Plan (LJ2024085), the Shanghai Municipal Key Clinical Specialty (No. shslczdzk06302), the Shanghai Jiao Tong University (YG2022ZD027, YG2025ZD29), and the SJTU Trans-med Awards Research Program (20240202).

References

- 1 L. Gardner, J. Warrington, J. Rogan, D. G. Rothwell, G. Brady, C. Dive, K. Kostarelos and M. Hadjidemetriou, *Nanoscale Horiz.*, 2020, 5, 1476–1486.
- 2 P. Robbe, N. Popitsch, S. J. L. Knight, P. Antoniou, J. Becq, M. He, A. Kanapin, A. Samsonova, D. V. Vavoulis, M. T. Ross, Z. Kingsbury, M. Cabes, S. D. C. Ramos, S. Page, H. Dreau, K. Ridout, L. J. Jones, A. Tuff-Lacey, S. Henderson, J. Mason, F. M. Buffa, C. Verrill, D. Maldonado-Perez, I. Roxanis, E. Collantes, L. Browning, S. Dhar, S. Damato, S. Davies, M. Caulfield, D. R. Bentley, J. C. Taylor, C. Turnbull, A. Schuh and G. P. On be half of the, *Genet. Med.*, 2018, 20, 1196–1205.
- 3 N. Blow, *Nature*, 2007, 448, 959–960.

- 4 W. Mathieson and G. Thomas, *Curr. Pathobiol. Rep.*, 2019, **7**, 35–40.
- 5 E. M. Van Allen, N. Wagle, P. Stojanov, D. L. Perrin, K. Cibulskis, S. Marlow, J. Jane-Valbuena, D. C. Friedrich, G. Kryukov, S. L. Carter, A. McKenna, A. Sivachenko, M. Rosenberg, A. Kiezun, D. Voet, M. Lawrence, L. T. Lichtenstein, J. G. Gentry, F. W. Huang, J. Postel, D. Farlow, D. Barbie, L. Gandhi, E. S. Lander, S. W. Gray, S. Joffe, P. Janne, J. Garber, L. MacConaill, N. Lindeman, B. Rollins, P. Kantoff, S. A. Fisher, S. Gabriel, G. Getz and L. A. Garraway, *Nat. Med.*, 2014, **20**, 682–688.
- 6 S. Bonin, F. Petrera, B. Niccolini and G. Stanta, *Mol. Pathol.*, 2003, **56**, 184–186.
- 7 M. Simbolo, M. Fassan, A. Ruzzenente, A. Mafficini, L. D. Wood, V. Corbo, D. Melisi, G. Malleo, C. Vicentini, G. Malpeli, D. Antonello, N. Sperandio, P. Capelli, A. Tomezzoli, C. Iacono, R. T. Lawlor, C. Bassi, R. H. Hruban, A. Guglielmi, G. Tortora, F. D. Braud and A. Scarpa, *Oncotarget*, 2014, **5**, 2839–2852.
- 8 O. E. Eremina, C. Vazquez, K. N. Larson, A. Mouchawar, A. Fernando and C. Zavaleta, *Nanoscale Horiz.*, 2024, **9**, 1896–1924.
- 9 S. M. Hewitt, F. A. Lewis, Y. Cao, R. C. Conrad, M. Cronin, K. D. Danenberg, T. J. Goralski, J. P. Langmore, R. G. Raja, P. M. Williams, J. F. Palma and J. A. Warrington, *Arch. Pathol. Lab. Med.*, 2008, **132**, 1929–1935.
- 10 X. Ye, Z.-Z. Zhu, L. Zhong, Y. Lu, Y. Sun, X. Yin, Z. Yang, G. Zhu and Q. Ji, *J. Thorac. Oncol.*, 2013, **8**, 1118–1120.
- 11 B. P. Bass, K. B. Engel, S. R. Greytak and H. M. Moore, *Arch. Pathol. Lab. Med.*, 2014, **138**, 1520–1530.
- 12 M. Srinivasan, D. Sedmak and S. Jewell, *Am. J. Clin. Pathol.*, 2002, **161**, 1961–1971.
- 13 M. T. Gilbert, T. Haselkorn, M. Bunce, J. J. Sanchez, S. B. Lucas, L. D. Jewell, E. Van Marck and M. Worobey, *PLoS One*, 2007, **2**, e537.
- 14 W. Mathieson and G. A. Thomas, *J. Histochem. Cytochem.*, 2020, **68**, 543–552.
- 15 K. Lu, W. Ye, L. Zhou, L. B. Collins, X. Chen, A. Gold, L. M. Ball and J. A. Swenberg, *J. Am. Chem. Soc.*, 2010, **132**, 3388–3399.
- 16 S. R. Greytak, K. B. Engel, B. P. Bass and H. M. Moore, *Cancer Res.*, 2015, **75**, 1541–1547.
- 17 T. Kuwata, M. Wakabayashi, Y. Hatanaka, E. Morii, Y. Oda, K. Taguchi, M. Noguchi, Y. Ishikawa, T. Nakajima, S. Sekine, S. Nomura, W. Okamoto, S. Fujii and T. Yoshino, *Pathol. Int.*, 2020, **70**, 932–942.
- 18 W. Xiao, L. Ren, Z. Chen, L. T. Fang, Y. Zhao, J. Lack, M. Guan, B. Zhu, E. Jaeger, L. Kerrigan, T. M. Blomquist, T. Hung, M. Sultan, K. Idler, C. Lu, A. Scherer, R. Kusko, M. Moos, C. Xiao, S. T. Sherry, O. D. Abaan, W. Chen, X. Chen, J. Nordlund, U. Liljedahl, R. Maestro, M. Polano, J. Drabek, P. Vojta, S. Köks, E. Reimann, B. S. Madala, T. Mercer, C. Miller, H. Jacob, T. Truong, A. Moshrefi, A. Natarajan, A. Granat, G. P. Schroth, R. Kalamegham, E. Peters, V. Petitjean, A. Walton, T.-W. Shen, K. Talsania, C. J. Vera, K. Langenbach, M. de Mars, J. A. Hipp, J. C. Willey, J. Wang, J. Shetty, Y. Kriga, A. Raziuddin, B. Tran, Y. Zheng, Y. Yu, M. Cam, P. Jailwala, C. Nguyen, D. Meerzaman, Q. Chen, C. Yan, B. Ernest, U. Mehra, R. V. Jensen, W. Jones, J.-L. Li, B. N. Papas, M. Pirooznia, Y.-C. Chen, F. Seifuddin, Z. Li, X. Liu, W. Resch, J. Wang, L. Wu, G. Yavas, C. Miles, B. Ning, W. Tong, C. E. Mason, E. Donaldson, S. Lababidi, L. M. Staudt, Z. Tezak, H. Hong, C. Wang and L. Shi, *Nat. Biotechnol.*, 2021, **39**, 1141–1150.
- 19 H. Do and A. Dobrovic, *Clin. Chem.*, 2015, **61**, 64–71.
- 20 A. Astolfi, M. Urbini, V. Indio, M. Nannini, C. G. Genovese, D. Santini, M. Saponara, A. Mandrioli, G. Ercolani, G. Brandi, G. Biasco and M. A. Pantaleo, *BMC Genomics*, 2015, **16**, 892.
- 21 M. Stiller, A. Sucker, K. Griewank, D. Aust, G. B. Baretton, D. Schadendorf and S. Horn, *Oncotarget*, 2016, **7**, 59115–59128.
- 22 Y. Zhao, L. T. Fang, T.-W. Shen, S. Choudhari, K. Talsania, X. Chen, J. Shetty, Y. Kriga, B. Tran, B. Zhu, Z. Chen, W. Chen, C. Wang, E. Jaeger, D. Meerzaman, C. Lu, K. Idler, L. Ren, Y. Zheng, L. Shi, V. Petitjean, M. Sultan, T. Hung, E. Peters, J. Drabek, P. Vojta, R. Maestro, D. Gasparotto, S. Köks, E. Reimann, A. Scherer, J. Nordlund, U. Liljedahl, J. Fook, C. E. Mason, C. Xiao, H. Hong and W. Xiao, *Sci. Data*, 2021, **8**, 296.
- 23 R. Menon, M. Deng, D. Boehm, M. Braun, F. Fend, D. Boehm, S. Biskup and S. Perner, *Int. J. Mol. Sci.*, 2012, **13**, 8933–8942.
- 24 C. Bolognesi, C. Forcato, G. Buson, F. Fontana, C. Mangano, A. Doffini, V. Sero, R. Lanzellotto, G. Signorini, A. Calanca, M. Sergio, R. Romano, S. Gianni, G. Medoro, G. Giorgini, H. Morreau, M. Barberis, W. E. Corver and N. Manaresi, *Sci. Rep.*, 2016, **6**, 20944.
- 25 G. P. Pfeifer, Y.-H. You and A. Besaratinia, *Mutat. Res., Fundam. Mol. Mech. Mutagen.*, 2005, **571**, 19–31.
- 26 V. I. Bruskov, L. V. Malakhova, Z. K. Masalimov and A. V. Chernikov, *Nucleic Acids Res.*, 2002, **30**, 1354–1363.
- 27 V. Ademà, E. Torres, F. Solé, S. Serrano and B. Bellosillo, *Biopreserv. Biobanking*, 2014, **12**, 281–283.
- 28 E. M. Golenberg, A. Bickel and P. Weihs, *Nucleic Acids Res.*, 1996, **24**, 5026–5033.
- 29 S. K. Nam, J. Im, Y. Kwak, N. Han, K. H. Nam, A. N. Seo and H. S. Lee, *Korean J. Pathol.*, 2014, **48**, 36–42.
- 30 M. Nagahashi, Y. Shimada, H. Ichikawa, S. Nakagawa, N. Sato, K. Kaneko, K. Homma, T. Kawasaki, K. Kodama, S. Lyle, K. Takabe and T. Wakai, *J. Surg. Res.*, 2017, **220**, 125–132.
- 31 S. Basyuni, L. Heskin, A. Degasperri, D. Black, G. C. C. Koh, L. Chmelova, G. Rinaldi, S. Bell, L. Grybowicz, G. Elgar, Y. Memari, P. Robbe, Z. Kingsbury, C. Caldas, J. Abraham, A. Schuh, L. Jones, M. Tischkowitz, M. A. Brown, H. R. Davies, S. Nik-Zainal, P. T. Group and G. Personalised Breast Cancer Program, *Nat. Commun.*, 2024, **15**, 7731.
- 32 Q. Guo, E. Lakatos, I. A. Bakir, K. Curtius, T. A. Graham and V. Mustonen, *Nat. Commun.*, 2022, **13**, 4487.
- 33 A. N. Hosein, S. Song, A. E. McCart Reed, J. Jayanthan, L. E. Reid, J. R. Kutasovic, M. C. Cummings, N. Waddell,

- S. R. Lakhani, G. Chenevix-Trench and P. T. Simpson, *Lab. Invest.*, 2013, **93**, 701–710.
- 34 S. S. David, V. L. O'Shea and S. Kundu, *Nature*, 2007, **447**, 941–950.
- 35 Y. Flores Bueso, S. P. Walker and M. Tangney, *Biol. Methods Protoc.*, 2020, **5**, bpaa015.
- 36 X. Su, Q. Zheng, X. Xiu, Q. Zhao, Y. Wang, D. Han and P. Song, *Med-X*, 2024, **2**, 14.
- 37 J. Kennedy-Darling and L. M. Smith, *Anal. Chem.*, 2014, **86**, 5678–5681.
- 38 J. Sponer and J. Kypr, *Gen. Physiol. Biophys.*, 1989, **8**, 257–272.
- 39 S. Bonin, F. Hlubek, J. Benhattar, C. Denkert, M. Dietel, P. L. Fernandez, G. Höfler, H. Kothmaier, B. Kruslin, C. M. Mazzanti, A. Perren, H. Popper, A. Scarpa, P. Soares, G. Stanta and P. J. Groenen, *Virchows Arch.*, 2010, **457**, 309–317.
- 40 A. M. García-Alegría, I. Anduro-Corona, C. J. Pérez-Martínez, M. A. Guadalupe Corella-Madueño, M. L. Rascón-Durán and H. Astiazaran-García, *Int. J. Anal. Chem.*, 2020, **2020**, 8896738.
- 41 N. Einaga, A. Yoshida, H. Noda, M. Suemitsu, Y. Nakayama, A. Sakurada, Y. Kawaji, H. Yamaguchi, Y. Sasaki, T. Tokino and M. Esumi, *PLoS One*, 2017, **12**, e0176280.
- 42 S. Graw, R. Meier, K. Minn, C. Bloomer, A. K. Godwin, B. Fridley, A. Vlad, P. Beyerlein and J. Chien, *Sci. Rep.*, 2015, **5**, 12335.
- 43 D.-h Heo, I. Kim, H. Seo, S.-G. Kim, M. Kim, J. Park, H. Park, S. Kang, J. Kim, S. Paik and S.-E. Hong, *Sci. Rep.*, 2024, **14**, 2559.
- 44 C. Williams, F. Pontén, C. Moberg, P. Söderkvist, M. Uhlén, J. Pontén, G. Sitbon and J. Lundeberg, *Am. J. Pathol.*, 1999, **155**, 1467–1471.
- 45 B. Gold, M. P. Stone and L. A. Marky, *Acc. Chem. Res.*, 2014, **47**, 1446–1454.