Supporting Information

Nanoscale Quality Control Framework for Assessing FFPE DNA Integrity in Cancer Research

Zixuan Huang[1ab], Yunpei Si[1 a], Yi Zhang[1 a], Zicheng Huang[b], Xuehao Xiu[a], Yunshan Wang[c], YuDong Wang[b]*, Chunhai Fan[d], Ping Song[ab]*

[a] Department of Gynecologic Oncology, The International Peace Maternity and Child Health Hospital, Shanghai Jiao Tong University, Shanghai 200240, China.

[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 Clinical Laboratory, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, 250021, Shandong, China

[d] 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

*Corresponding authors.

E-mail addresses: songpingsjtu@sjtu.edu.cn (P. Song), wangyudong@shsmu.edu.cn (Y. Wang)

[1] These authors contributed equally to this work.

List of Tables

List of Figures

1. DNA sequences for qPCR amplification

1) rs206781

FP：CACTTCCTCCAGAAGGTCCAAAG

Table S1. Sequences of different RP for rs206781 amplification.

RP NO.	RP sequence	Genomic position	Amplicon length (bp)	Amplicon sequence
RP1	TGGCAGTGCTGCTTTAGGC	Chr6:119277839-119277829	50	CACTTCCTCCAGAAGGTCCAAAGCCAGAAGGGCCTAAAGCAGCACTGCCA
RP2	GAAGTCCAAGAAAGCAAGTGGC	Chr6:32978272-32978381	80	CACTTCCTCCAGAAGGTCCAAAGCCAGAAGGGCCTAAAGCAGCACTGCCACCCCCACTGCCACTTGCTTTCTTGGACTTC
RP3	GCCTAGGGCACCCCGC	Chr6:32978272-32978381	110	CACTTCCTCCAGAAGGTCCAAAGCCAGAAGGGCCTAAAGCAGCACTGCCACCCCCACTGCCACTTGCTTTCTTGGACTTCTTAGGTTGAGGTGGGCGGGGTGCCCTAGGC
RP4	CCGAGCTGGGGCCGAT	Chr6:32978242-32978381	140	CACTTCCTCCAGAAGGTCCAAAGCCAGAAGGGCCTAAAGCAGCACTGCCACCCCCACTGCCACTTGCTTTCTTGGACTTCTTAGGTTGAGGTGGGCGGGGTGCCCTAGGCCCCTTGTCATCTTCATCGGCCCCAGCTCGG
RP5	AGAAACGGAAGGCAGAGAAGC	Chr6:32978214-32978381	168	CACTTCCTCCAGAAGGTCCAAAGCCAGAAGGGCCTAAAGCAGCACTGCCACCCCCACTGCCACTTGCTTTCTTGGACTTCTTAGGTTGAGGTGGGCGGGGTGCCCTAGGCCCCTTGTCATCTTCATCGGCCCCAGCTCGGCCTCGATGCTTCTCTGCCTTCCGTTTCT
RP6	CCCAAGAGGAAAAGAGAGAAAAAAGAG	Chr6:32978180-32978381	202	CACTTCCTCCAGAAGGTCCAAAGCCAGAAGGGCCTAAAGCAGCACTGCCACCCCCACTGCCACTTGCTTTCTTGGACTTCTTAGGTTGAGGTGGGCGGGGTGCCCTAGGCCCCTTGTCATCTTCATCGGCCCCAG

				CTCGGCCTCGATGCTTCTCTGCCTTCCGTTTCTTCTTTTTCTCTTTTTTCTCTCTTTTCCTCTTGGG
RP7	TGCTCTGTCCCAGGGTCC	Chr6:32978152-32978381	230	CACTTCCTCCAGAAGGTCCAAAGCCAGAAGGGCCTAAAGCAGCACTGCCACCCCCACTGCCACTTGCTTTCTTGGACTTCTTAGGTTGAGGTGGGCGGGGTGCCCTAGGCCCCTTGTCATCTTCATCGGCCCCAGCTCGGCCTCGATGCTTCTCTGCCTTCCGTTTCTTCTTTTTCTCTTTTTTCTCTCTTTTCCTCTTGGGCTTGGATATTGGACCCTGGGACAGAGCA
RP8	TTAGCTTCGGGCAGTACATGAAC	Chr6:32978122-32978381	260	CACTTCCTCCAGAAGGTCCAAAGCCAGAAGGGCCTAAAGCAGCACTGCCACCCCCACTGCCACTTGCTTTCTTGGACTTCTTAGGTTGAGGTGGGCGGGGTGCCCTAGGCCCCTTGTCATCTTCATCGGCCCCAGCTCGGCCTCGATGCTTCTCTGCCTTCCGTTTCTTCTTTTTCTCTTTTTTCTCTCTTTTCCTCTTGGGCTTGGATATTGGACCCTGGGACAGAGCAGCCAGTTGTTCATGTACTGCCCGAAGCTAA

2) rs2638145

FP：GGATGGGACTCCAATGCAAAACT

Table S2. Sequences of different RP for rs2638145 amplification.

RP NO.	RP sequence	Genomic position	Amplicon length (bp)	Amplicon sequence
RP1	CCCTAGACATCCTCACACTGATAC	Chr3:22276759-22276809	51	GGATGGGACTCCAATGCAAAACTCAATGTATCAGTGTGAGGATGTCTAGGG
RP2	TTTATTATTTTGTTCCATTCTATATTTCTTACCCTAG	Chr3:22276759-22276840	82	GGATGGGACTCCAATGCAAAACTCAATGTATCAGTGTGAGGATGTCTAGGGTAAGAAATATAGAATGGAACAAAATAATAAA
RP3	ACAGGAAAAGAAACTAAAATTGTACCCTT	Chr3:22276759-22276867	109	GGATGGGACTCCAATGCAAAACTCAATGTATCAGTGTGAGGATGTCTAGGGTAAGAAATATAGAATGGAACAAAATAATAAAGGGTACAATTTTAGTTTCTTTTCCTGT
RP4	ATATTATCTATTTTCAGATATAAAAATTTTAGAACAGGAAAAG	Chr3:22276759-22276900	142	GGATGGGACTCCAATGCAAAACTCAATGTATCAGTGTGAGGATGTCTAGGGTAAGAAATATAGAATGGAACAAAATAATAAAGGGTACAATTTTAGTTTCTTTTCCTGTTCTAAAATTTTTATATCTGAAAATAGATAATAT
RP5	TCTCTTGTTATTTCTGACAATTTTATTCATATTATCTATT	Chr3:22276759-22276928	170	GGATGGGACTCCAATGCAAAACTCAATGTATCAGTGTGAGGATGTCTAGGGTAAGAAATATAGAATGGAACAAAATAATAAAGGGTACAATTTTAGTTTCTTTTCCTGTTCTAAAATTTTTATATCTGAAAATAGATAATATGAATAAAATTGTCAGAAATAACAAGAGA
RP6	TCTTAAGTGAATGCATGTGTTGAAATAAC	Chr3:22276759-22276960	202	GGATGGGACTCCAATGCAAAACTCAATGTATCAGTGTGAGGATGTCTAGGGTAAGAAATATAGAATGGAACAAAATAATAAAGGGTACAATTTTAGTTTCTTTTCCTGTTCTAAAATTTTTATATCTGAAAATAGATAATATGAATAAAATTGTCAG

				AAATAACAAGAGATTTGTTATTTCAACACATGCATTCACTTAAGA
RP7	TCTTCATTACCAACCATGTAAAATAAAGTCT	Chr3:22276759-22276988	230	GGATGGGACTCCAATGCAAAACTCAATGTATCAGTGTGAGGATGTCTAGGGTAAGAAATATAGAATGGAACAAAATAATAAAGGGTACAATTTTAGTTTCTTTTCCTGTTCTAAAATTTTTATATCTGAAAATAGATAATATGAATAAAATTGTCAGAAATAACAAGAGATTTGTTATTTCAACACATGCATTCACTTAAGACTTTATTTTACATGGTTGGTAATGAAGA
RP8	CATTATATAACTACTGCTTGTGTTTTCTGTTC	Chr3:22276759-22277018	260	GGATGGGACTCCAATGCAAAACTCAATGTATCAGTGTGAGGATGTCTAGGGTAAGAAATATAGAATGGAACAAAATAATAAAGGGTACAATTTTAGTTTCTTTTCCTGTTCTAAAATTTTTATATCTGAAAATAGATAATATGAATAAAATTGTCAGAAATAACAAGAGATTTGTTATTTCAACACATGCATTCACTTAAGACTTTATTTTACATGGTTGGTAATGAAGAACAGAAAACACAAGCAGTAGTTATATAATG

3) rs2510152

FP：GGATGGGACTCCAATGCAAAACT

Table S3. Sequences of different RP for rs2510152 amplification.

RP NO.	RP sequence	Genomic position	Amplicon length (bp)	Amplicon sequence
RP1	AGCTTGTGTCCAGTGATATGGTTAT	Chr11:119277839-119277829	50	ACCCAGGTGAGTTTTGTTTCACATGATAACCATATCACTGGACACAAGCT
RP2	TGCTGTCAGTTCTATAAAGAATATACTAAAGC	Chr11:119277839-119277917	79	ACCCAGGTGAGTTTTGTTTCACATGATAACCATATCACTGGACACAAGCTTTAGTATATTCTTTATAGAACTGACAGCA
RP3	TAGTCAATGGGTTCCAATGAATTCAATTAT	Chr11:119277839-119277946	108	ACCCAGGTGAGTTTTGTTTCACATGATAACCATATCACTGGACACAAGCTTTAGTATATTCTTTATAGAACTGACAGCATAATTGAATTCATTGGAACCCATTGACTA
RP4	TGAAACCACATACACACAAATTCACT	Chr11:119277839-119277976	138	ACCCAGGTGAGTTTTGTTTCACATGATAACCATATCACTGGACACAAGCTTTAGTATATTCTTTATAGAACTGACAGCATAATTGAATTCATTGGAACCCATTGACTAGATTAGTGAATTTGTGTGTATGTGGTTTCA
RP5	GGGTCCTATTTTAAGCTCCAGGG	Chr11:119277839-119277859	168	ACCCAGGTGAGTTTTGTTTCACATGATAACCATATCACTGGACACAAGCTTTAGTATATTCTTTATAGAACTGACAGCATAATTGAATTCATTGGAACCCATTGACTAGATTAGTGAATTTGTGTGTATGTGGTTTCACTTTAAACCCTGGAGCTTAAAATAGGACCC
RP6	TTTATTTTTTATTAAACCAGAAAGCATCTAGTCTG	Chr11:119277839-119278040	202	ACCCAGGTGAGTTTTGTTTCACATGATAACCATATCACTGGACACAAGCTTTAGTATATTCTTTATAGAACTGACAGCATAATTGAATTCATTGGAACCCATTGACTAGATTAGTGAATTTGTGTGTATGTGGTTTCACTTTAAACCCTGGAGCTTAAAATAGGACCCAGACTAGATGCTTTCTGGTTTAATAAAAA

				ATAAA
RP7	TGCTTATATAAAAATGTCACAACAGTGGT	Chr11:119277839-119278068	230	ACCCAGGTGAGTTTTGTTTCACATGATAACCATATCACTGGACACAAGCTTTAGTATATTCTTTATAGAACTGACAGCATAATTGAATTCATTGGAACCCATTGACTAGATTAGTGAATTTGTGTGTATGTGGTTTCACTTTAAACCCTGGAGCTTAAAATAGGACCCAGACTAGATGCTTTCTGGTTTAATAAAAAATAAACCACTGTTGTGACATTTTTATATAAGCA
RP8	AAAGTGAAGACTTGTTTCCTATACAAATTTTG	Chr11:119277839-119278098	260	ACCCAGGTGAGTTTTGTTTCACATGATAACCATATCACTGGACACAAGCTTTAGTATATTCTTTATAGAACTGACAGCATAATTGAATTCATTGGAACCCATTGACTAGATTAGTGAATTTGTGTGTATGTGGTTTCACTTTAAACCCTGGAGCTTAAAATAGGACCCAGACTAGATGCTTTCTGGTTTAATAAAAAATAAACCACTGTTGTGACATTTTTATATAAGCAAAATTTGTATAGGAAACAAGTCTTCACTTT

Table S4. List of SNPs used for qPCR multi-amplicon length testing.

SNP	Position (GRCh38)	FP sequence
rs206781	chr6:32978356	GTCTCTGCTTCCTACTCTAACTCAC
rs10230708	chr7:26625472	ACCAATGGGAGTCACTGCTG
rs2301720	chr7:27156450	GTAGCCGCTTCTCTGTGAGTT
rs4712476	chr6:20291800	CCCCGGATGTCAGGGAATG

Table S5. Sequences of different RP corresponding to various amplicon lengths for SNPs used in qPCR multi-amplicon length testing.

SNP	RP Sequence(50bp)	RP Sequence (150bp)	RP Sequence (250bp)
rs206781	GACCTTCTGGAGGAAGTGGC	GGGTCCAATATCCAAGCCCA	GGGTCCAATATCCAAGCCCA
rs10230708	GTCCTAACACCTCTTATCACAGGG	GTGTCAAGTCAGGTACCCTGAC	GCCATTTTACTTAGAAACTAGCAGAATTTAA
rs2301720	GCCGAGCCGACTTCTTGC	CTGCCGGACAACAAATCACAG	GCTGCGGCGAGCTCC
rs4712476	TTGTGAGACGGGAACAGTGTT	ACCCTTCCGGTGTGGGC	GTACTGAAGAAAAGCAGATCTTAATTTGG

Table S6. Full amplicon sequences (250 bp) for SNPs used in qPCR multi-amplicon length testing.

SNP	Amplicon sequence (250bp)
rs206781	CTAGTCTCTGCTTCCTACTCTAACTCACTTGGTGCCACTTCCTCCAGAAGGTCCAAAGCCaGAAGGGCCTAAAGCAGCACTGCCACCCCCACTGCCACTTGCTTTCTTGGACTTCTTAGGTTGAGGTGGGCGGGGTGCCCTAGGCCCCTTGTCATCTTCATCGGCCCCAGCTCGGCCTCGATGCTTCTCTGCCTTCCGTTTCTTCTTTTTCTCTTTTTTCTCTCTTTTCCTCTTGGGCTTGGATATTGGACCCTGGGA
rs10230708	ACCAATGGGAGTCACTGCTGTAGGCGCCCTGTGATAAGAGGTGTTAGGACTCATACTGCCCAGATGAGGACACTGATGTCCAGGGAGGTAGAGCAGCGTGCCCAAACCCACTCCCTACCATGGGAGGTGTCAGGGTACCTGACTTGACACCCCAGTTCTTTCCACTTAACCCCACAACTTAAGCCAGCCTGCTCTGCAGTTAGTTAAATTTGGCCATTCTTAAATTCTGCTAGTTTCTAAGTAAAATGGCAAGGTATCCATGCCTTTC
rs2301720	GTAGCCGCTTCTCTGTGAGTTGGGAGCAAAGGAGCAAGAAGTCGGCTCGGCATTTTGGAACAGAGAAGCCCCCGCCGTATATTTGCTAAAAAGCGCGTTCACATAATACGAAGAACTCATAATTTTGACCTGTGATTTGTTGTCCGGCAGCTTTCAGTGTCGGTTTTACGAGGTAGAGTGATATATGATAACATTACACCCCCAGATTTACACCAAACCCCATTTTCTTTTGGACGGAGCTCGCCGCAGCACGTGACCGCCCACATGACCGCCTC
rs4712476	CCCCGGATGTCAGGGAATGCTCTGTACAAAACACTGTTCCCGTCTCACAAAGCATCTTGAAGATGGAGACGTGCATACTCACAGTTCACCACAAGAGGGCGGTCATCAAAAGGGAAAGGTCACCCTGCCCTAAGCCCACACCGGAAGGGTGAATCCATGTTTAAGGAGGTTGAAGGTATTTTTATAATCATGGGCTGCTCCTGGAATTGATAAAACACGATCCAAATTAAGATCTGCTTTTCTTCAGTACTATTTTTATGTTGTTAATCCCATTT

Table S7. List of SNPs used in NGS.

SNP	Position (GRCh38)	FP sequence	RP sequence
rs10230708	chr7:26625472	ACCAATGGGAGTCACTGCTG	TAAGTGGAAAGAACTGGGGTGTC
rs10104396	chr8:23839899	GAGGGGTATTAGAAGAATGACTATGTGA	ACATGGTTAGATATTAGCCTGACCTATG
rs199032	chr6:23438217	GCTCTTCCTCTCACATCTTTATTTAACC	GCAGCCAAGTGTGAAAGTATTGA
rs926850	chr6:24898851	CAGAGTAAAATTTACTGCTCCGTCATAA	TGATGTTGAGTTGAGACAGGTTACA
rs17149369	chr7:24372574	GGATTCCCTAAGCTCTTCAATATTGC	AAATGTAGTTCTATTATGGTCAGCACAC
rs869720	chr1:38326229	CCTCATCTGTAAAGCAGGGAGAGA	AGTATCCCCAAAAGGTTGCAGAT
rs12478327	chr2:20174488	ACTTCTGCCAACATTCAAATTCAGG	GTGCAAGCTGGAGGCACT
rs2638145	chr3:22276782	GGATGGGACTCCAATGCAAAACT	ACAGGAAAAGAAACTAAAATTGTACCCTT
rs2170091	chr3:22088347	CATCTTGCTCTTCATAGATAGCTTCAGA	GAAGCCAGATCTCAAAGTGTCCT
rs2043583	chr3:24598078	CCTGAATGTCAGTTTTGTTAGAGCAAC	GTTATTGGGAATGCTATGAAAGAGACA
rs955456	chr4:23651593	CAGACTTAATCAAAGCCCTTGAAAAGA	AGAACTCATTTCCTTATAGCTGAAGAACT
rs966516	chr4:24775000	CCTCCCATAGTGATTCTTATGAAGTCA	GCAGACACTTAGGATGTTTCCAGT
rs354169	chr1:49189758	AATGCTTTGCTTGCTGAGAACTT	GAGCCTTAGTTCCTCCATCAGTAAA
rs1898170	chr4:29349716	AATGGGAAAACACATTTTAAGGGCA	AAATTTACGTTGGTAATTGGGTCTTGT
rs11247921	chr1:26339111	CCACACTCTGCCTCTCATGGTAT	CACAGAGGTGACAGAACACAGT
rs1635718	chr1:35508868	ACTTAAGAGGTCAACACAGATGAAAATTATC	TAGTTATTCATGGTGGGAAGGCAA
rs10510620	chr3:29288795	TCCGCAAAACCTACAATCTCTGAA	AAAAGATAATGTTCTTGTTTATATGCCCTTG
rs7104025	chr11:21126827	TCAGATGCTTTAGGCTCATGAGTTA	TACAGCAACTCACAAACTAATGACTCT
rs2246745	chr2:29717663	CTCCTTGGAATCACCAACAAACAT	GGCTGCGATGAGACAGGAA
rs3789806	chr7:140749271	CTTGTATATAGACGGTAAAATAAACACCAAG	AGGCACCAGAAGTCATCAGAATG

A			
rs706714	chr5:68227023	TGAAGCAGATGTTGAACAACAAGG	GACCAAGCTTTTATGCACCACA
rs1884444	chr1:67168129	TTCCTGCTTCCAGACATGAATCA	TGAAAGATAGCAATAGATACATAAAACACCA
rs2510152	chr11:119277863	ACCCAGGTGAGTTTTGTTTCACAT	TGAAACCACATACACACAAATTCACT
rs16754	chr11:32396399	CTCTCTGCCTGCAGGATGTG	CTTCCTGCTGTGCATCTGTAAGT
rs206781	chr6:32978356	CACTTCCTCCAGAAGGTCCAAAG	AAAAAGAAGAAACGGAAGGCAGAG
rs28932178	chr5:177210575	ACTAAGAGTGCAGAGCCTGGAA	TGCTGCCCCACCCTTTATTAAC

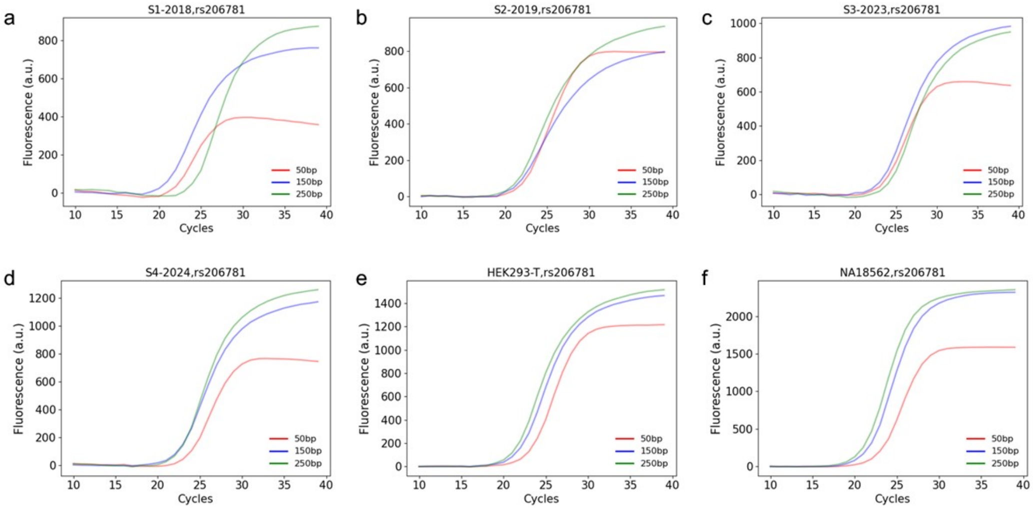

Fig. S1. qPCR amplification curves for FFPE samples of different years and cell lines at rs206781 with multiple amplicon lengths.

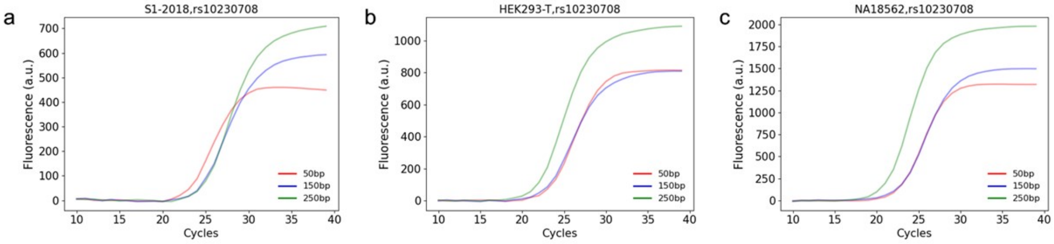

Fig. S2. qPCR amplification curves for FFPE sample S1-2018, and cell lines HEK293-T and NA18562 at locus rs10230708 with multiple amplicon lengths.

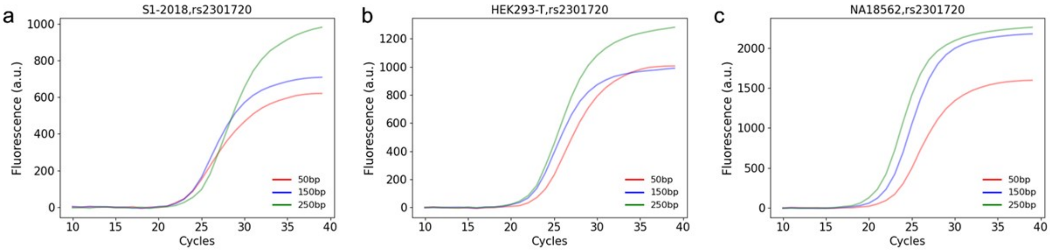

Fig. S3. qPCR amplification curves for FFPE sample S1-2018, and cell lines HEK293-T and NA18562 at locus rs2301720 with multiple amplicon lengths.

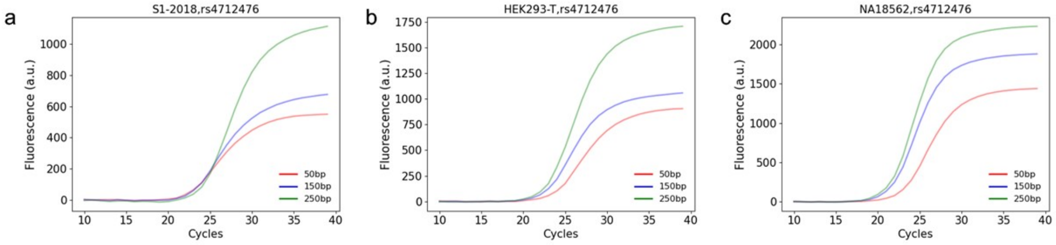

Fig. S4. qPCR amplification curves for FFPE sample S1-2018, and cell lines HEK293-T and NA18562 at locus rs4712476 with multiple amplicon lengths.

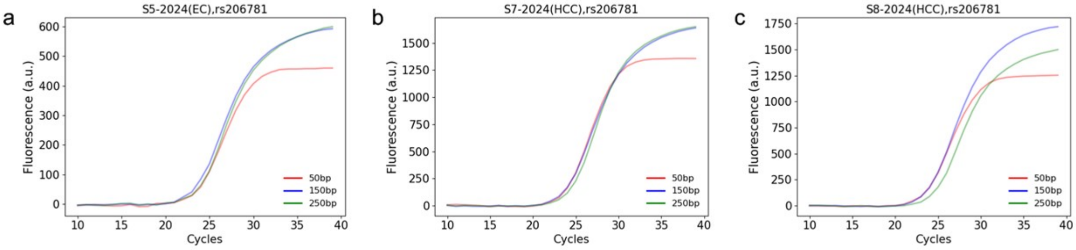

Fig. S5. qPCR amplification curves for samples of different tissue origins: S5-2024 (EC), S7-2024 (HCC), and S8-2024 (HCC) at locus rs206781 with multiple amplicon lengths.

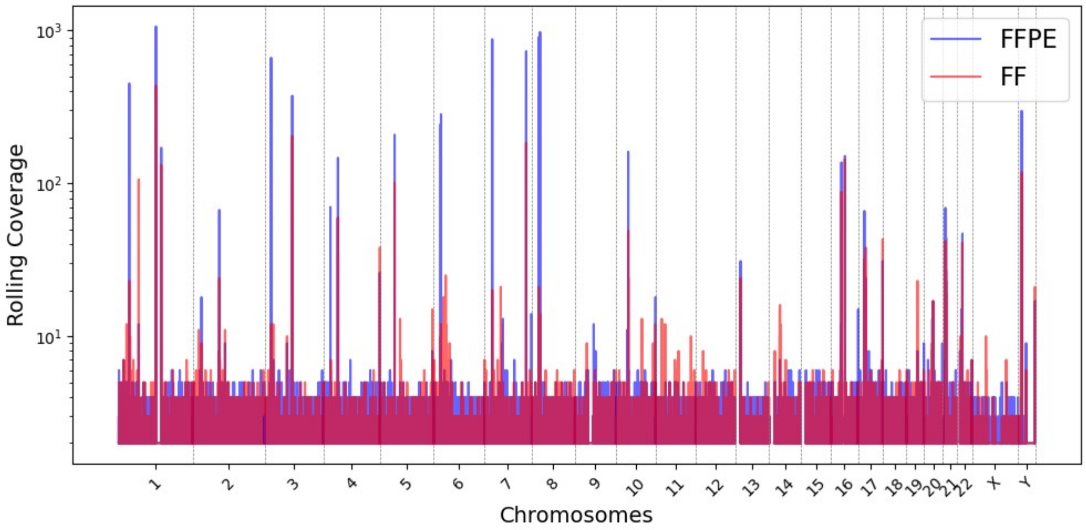

Fig. S6. Chromosomal coverage of FF and FFPE samples after targeted library preparation and NGS.

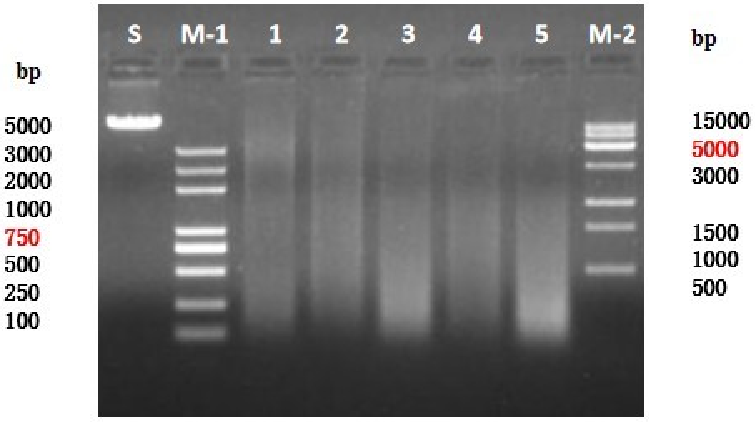

Fig. S7. Gel electrophoresis results of samples S4-2024 (lane 1), S5-2024 (lane 2), S9-2023 (lane 3), S1-2018 (lane 4), and S10-2019 (lane 5) processed by NovaSeq.

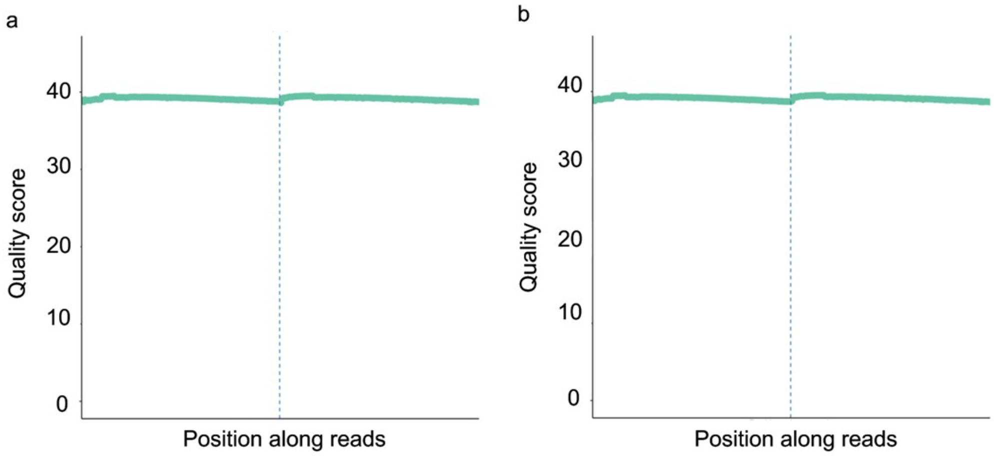

Fig. S8. Quality score distribution of sequencing reads. (a) Quality score distribution of S5-2024 reads. (b) Quality score distribution of S1-2018 reads. The x-axis represents the base position in the reads, and the y-axis represents the quality score per base. The first half shows the quality distribution of the forward reads, while the second half represents the reverse reads.

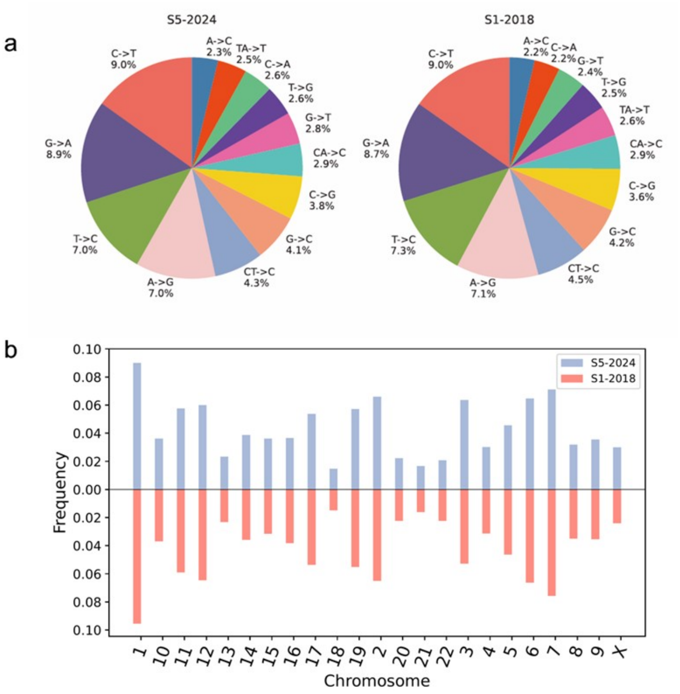

Fig. S9. WES mutation analysis of samples S5-2024 and S1-2018.(a) Proportional distribution of base substitution types in somatic mutations. (b) Chromosomal distribution pattern of detected mutations.