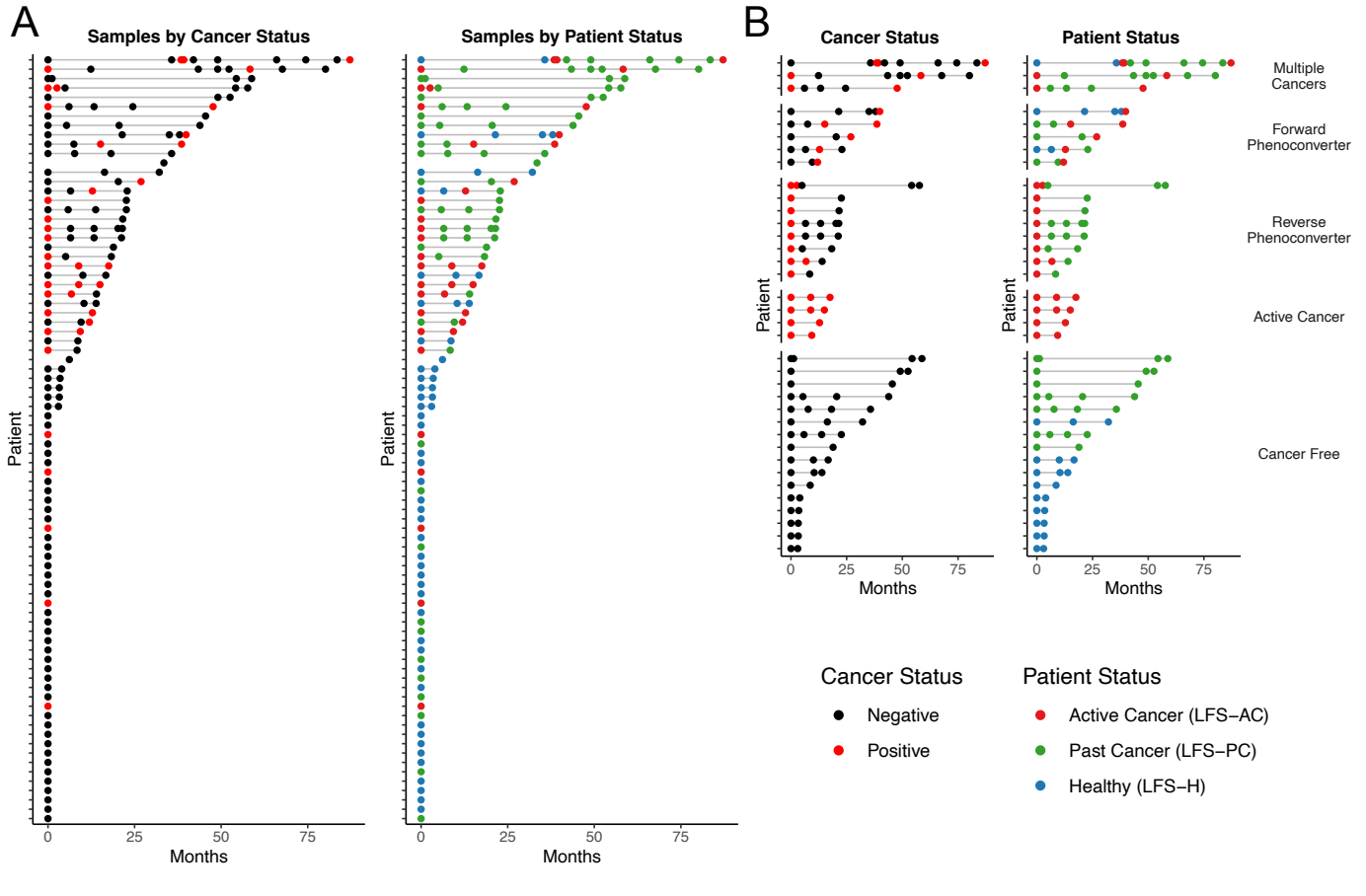
Supplementary Materials

Table of Contents:

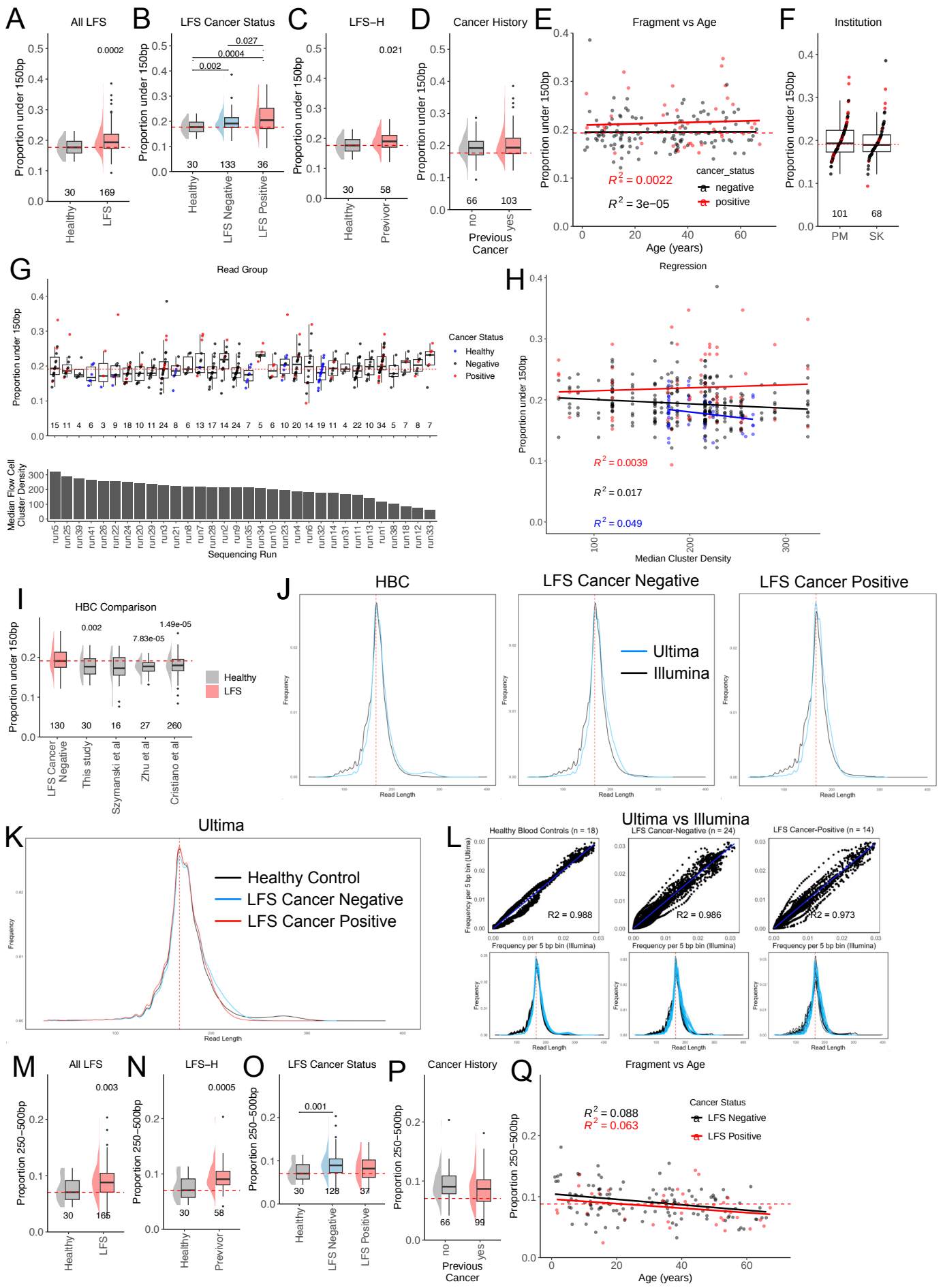

A
Samples by Cancer Status
Samples by Patient Status
Patient
Months
0
25
50
75
B
Cancer Status
Patient Status
Multiple Cancers
Forward Phenoconverter
Reverse Phenoconverter
Active Cancer
Cancer Free
Cancer Status
Negative
Positive
Patient Status
Active Cancer (LFS−AC)
Past Cancer (LFS−PC)
Healthy (LFS−H)

Supplementary Figure 1: Clinical timelines of patient cohort

A) Dot and line plots showing plasma sample schedules for each *TP53m*-carrier. Plasma samples are colored based on cancer status (left) or patient status (right).

B) Dot and line plots showing plasma sample schedules for *TP53m*-carriers with multiple timepoints. Plasma samples are colored based on cancer status (left) or patient status (right). Patients are split based on cancer status and cancer conversion status.

Source data are provided as a Source Data file.


A
All LFS
Proportion under 150bp
0.0002
30
169
Healthy
LFS
B
LFS Cancer Status
0.027
0.0004
0.002
30
133
36
Healthy
LFS Negative
LFS Positive
C
LFS−H
0.021
30
58
Healthy
Previvor
D
Cancer History
66
103
no
yes
Previous Cancer
E
Fragment vs Age
R^2 = 0.0022
R^2 = 3e−05
cancer_status
negative
positive
Age (years)
F
Institution
101
68
PM
SK
G
Read Group
Cancer Status
Healthy
Negative
Positive
15 11 4 6 3 9 18 10 11 24 8 6 13 17 14 24 7 5 6 10 20 14 19 11 4 22 10 34 5 7 8 7
Median Flow Cell Cluster Density
run5 run25 run39 run41 run26 run22 run24 run20 run29 run3 run21 run8 run7 run28 run2 run9 run35 run34 run10 run23 run4 run6 run32 run14 run31 run11 run13 run1 run38 run18 run12 run33
Sequencing Run
H
Regression
R^2 = 0.0039
R^2 = 0.017
R^2 = 0.049
Median Cluster Density
I
HBC Comparison
0.002
7.83e-05
1.49e-05
130
30
16
27
260
LFS Cancer Negative
This study
Szymanski et al
Zhu et al
Cristiano et al
Healthy
LFS
J
HBC
LFS Cancer Negative
LFS Cancer Positive
Ultima
Illumina
Frequency
Read Length
K
Ultima
Healthy Control
LFS Cancer Negative
LFS Cancer Positive
L
Ultima vs Illumina
Healthy Blood Controls (n = 18)
LFS Cancer-Negative (n = 24)
LFS Cancer-Positive (n = 14)
R2 = 0.988
R2 = 0.986
R2 = 0.973
Frequency per 5 bp bin (Ultima)
Frequency per 5 bp bin (Illumina)
M
All LFS
0.003
30
165
N
LFS−H
0.0005
30
58
O
LFS Cancer Status
0.001
30
128
37
P
Cancer History
66
99
Q
Fragment vs Age
R^2 = 0.088
R^2 = 0.063
Cancer Status
LFS Negative
LFS Positive
Proportion 250−500bp
Age (years)

Supplementary Figure 2: Comparison of cfDNA fragment lengths across clinical and technical correlates

Tukey boxplots comparing the proportion of short cfDNA fragments (10-151bp) between *TP53*-wildtype and *TP53m*-carriers (A), across LFS cancer statuses (B), LFS-H samples (C), between history of cancer in *TP53m*-carriers (D), and between institutions (F). Samples were collected in EDTA tubes at SickKids (SK) and Streck tubes at Princess Margaret (PM).

E) Scatter plot comparing the age at plasma collection and the proportion of short cfDNA fragments in *TP53m*-carriers.

G) Tukey boxplots comparing the proportion of short cfDNA fragments across sequencing flow cells (run#).

H) Scatter plot comparing the proportion of short cfDNA fragments with the flow cell cluster density.

I) Tukey boxplots comparing the proportion of short cfDNA fragments across our cohort and external healthy control cohorts.

J) Fragment length distributions comparing fragments from Illumina and Ultima sequencing.

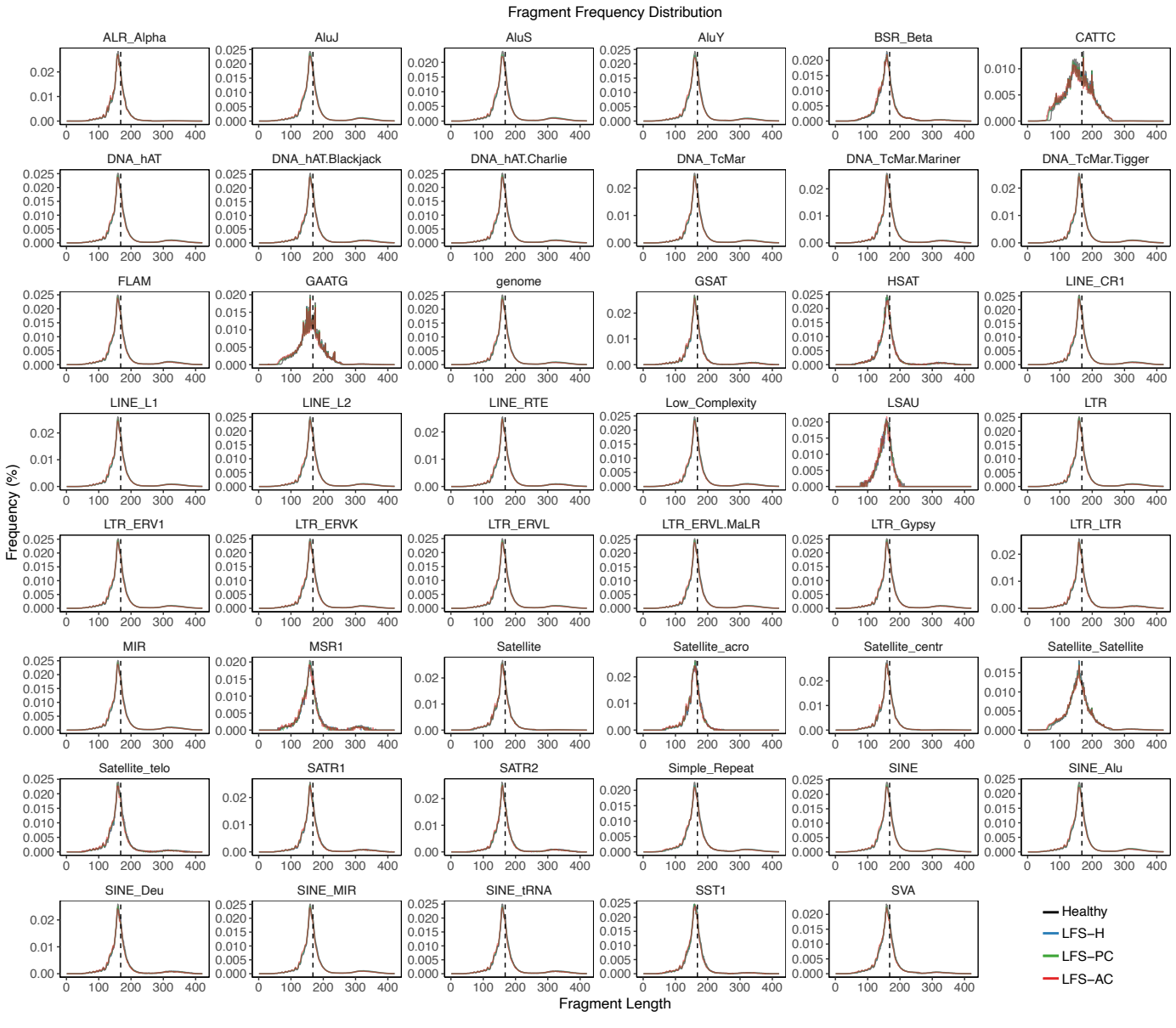
K) Fragment length distrubtions comparing fragment lengths from Ultima sequencing between *TP53*-wildtype and *TP53m*-carriers

L) Scatter plot showing a comparison of the proportion of fragments in 5bp bins between Ultima and Illumina sequencing.

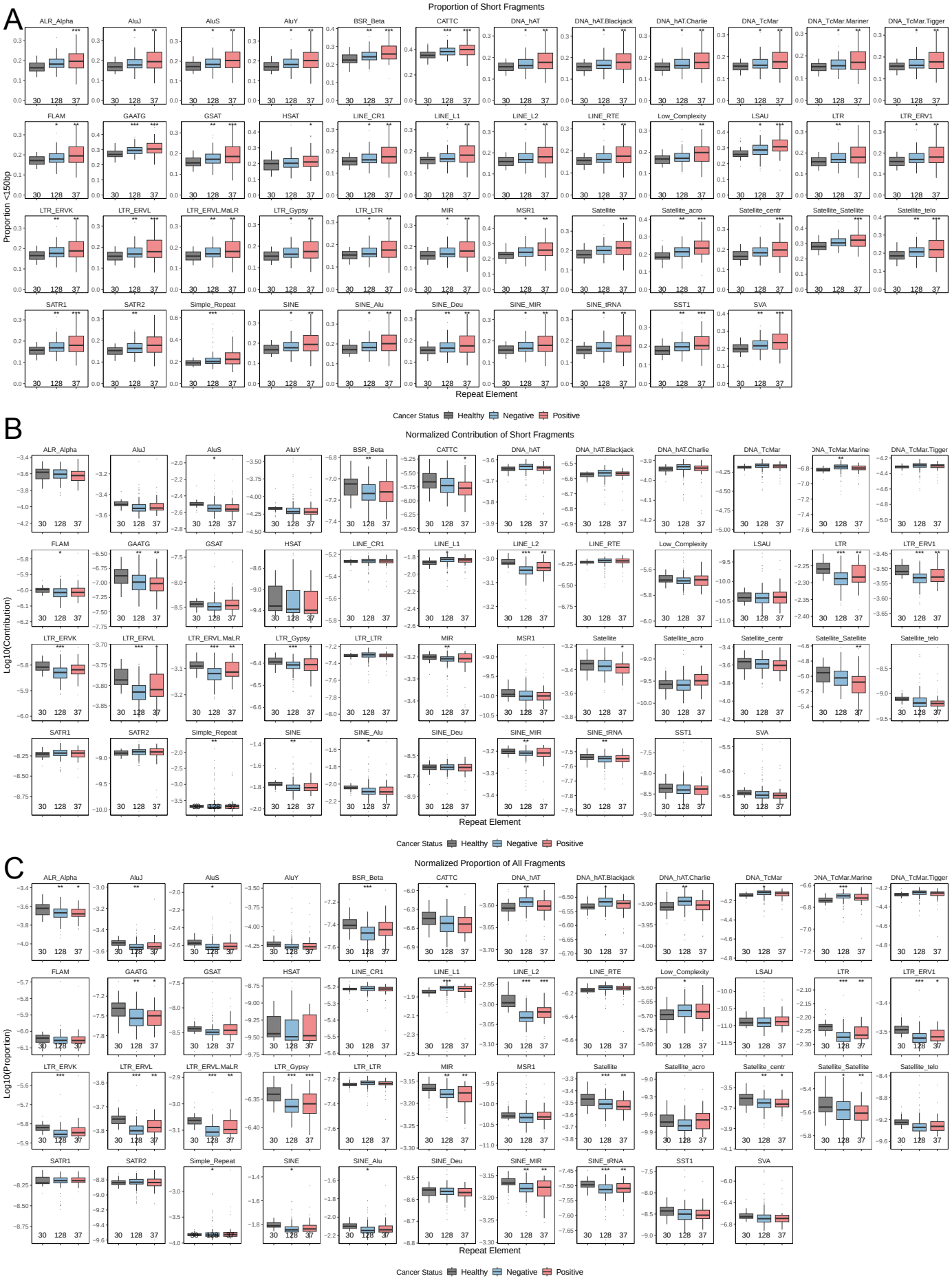
Tukey boxplots comparing the proportion of dinucleosome fragments (250-500bp) between *TP53*-wildtype and *TP53m*-carriers (M), across LFS cancer statuses (N), LFS-H samples (O), and between history of cancer in *TP53m*-carriers (P)

Q) Scatter plot comparing the age at plasma collection and the proportion of dinucleosome fragments in *TP53m*-carriers.

R^2 values calculated using Pearson's correlation and p-values calculated using two-sided Student's t-test. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Source data are provided as a Source Data file.


Fragment Frequency Distribution
ALR_Alpha
AluJ
AluS
AluY
BSR_Beta
CATTC
DNA_hAT
DNA_hAT.Blackjack
DNA_hAT.Charlie
DNA_TcMar
DNA_TcMar.Mariner
DNA_TcMar.Tigger
FLAM
GAATG
genome
GSAT
HSAT
LINE_CR1
LINE_L1
LINE_L2
LINE_RTE
Low_Complexity
LSAU
LTR
LTR_ERV1
LTR_ERVK
LTR_ERVL
LTR_ERVL.MaLR
LTR_Gypsy
LTR_LTR
MIR
MSR1
Satellite
Satellite_acro
Satellite_centr
Satellite_Satellite
Satellite_telo
SATR1
SATR2
Simple_Repeat
SINE
SINE_Alu
SINE_Deu
SINE_MIR
SINE_tRNA
SST1
SVA
Frequency (%)
Fragment Length
Healthy
LFS−H
LFS−PC
LFS−AC

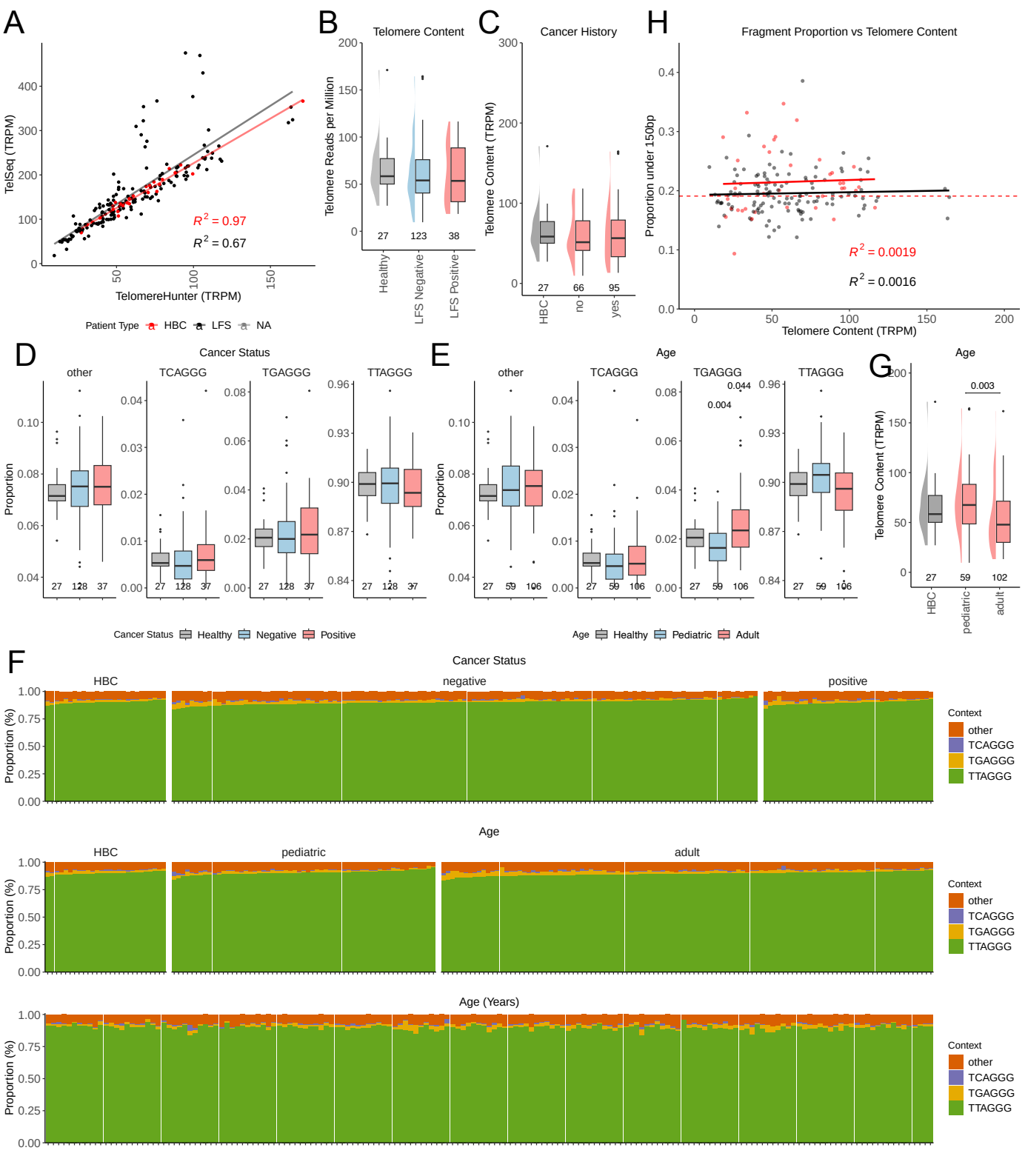
Supplementary Figure 3: Fragment length frequency distribution across genomic repeat elements
Fragment size distributions across 46 repeat element families. Each plot compares the distribution between *TP53*-wildtype and *TP53m*-carriers. A genome-wide comparison is also included. Source data are provided as a Source Data file.


A
Proportion of Short Fragments
ALR_Alpha
AluJ
AluS
AluY
BSR_Beta
CATTC
DNA_hAT
DNA_hAT.Blackjack
DNA_hAT.Charlie
DNA_TcMar
DNA_TcMar.Mariner
DNA_TcMar.Tigger
FLAM
GAATG
GSAT
HSAT
LINE_CR1
LINE_L1
LINE_L2
LINE_RTE
Low_Complexity
LSAU
LTR
LTR_ERV1
LTR_ERVK
LTR_ERVL
LTR_ERVL.MaLR
LTR_Gypsy
LTR_LTR
MIR
MSR1
Satellite
Satellite_acro
Satellite_centr
Satellite_Satellite
Satellite_telo
SATR1
SATR2
Simple_Repeat
SINE
SINE_Alu
SINE_Deu
SINE_MIR
SINE_tRNA
SST1
SVA
Proportion <150bp
Repeat Element
Cancer Status Healthy Negative Positive
B
Normalized Contribution of Short Fragments
Log10(Contribution)
Repeat Element
Cancer Status Healthy Negative Positive
C
Normalized Proportion of All Fragments
Log10(Proportion)
Repeat Element
Cancer Status Healthy Negative Positive

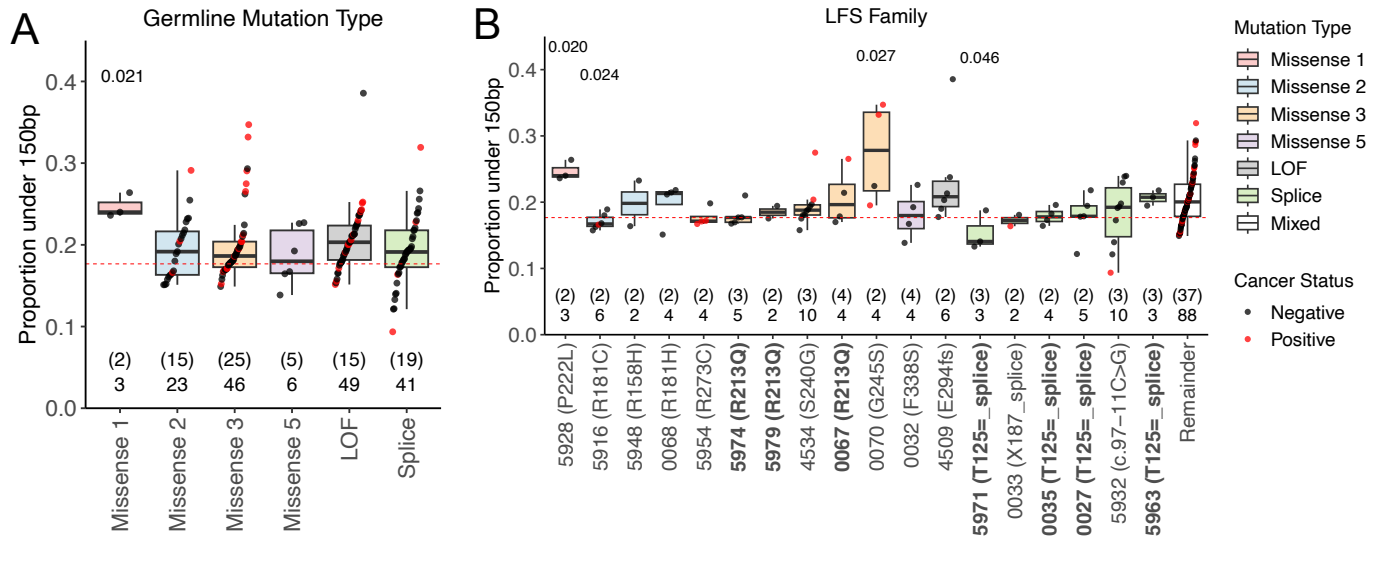
Supplementary Figure 4: Comparison of the proportion and contribution of genomic repeat elements to cfDNA fragment lengths

Tukey boxplots comparing the proportion of short cfDNA fragments (A), the normalized contribution of short cfDNA fragments (B), and the total proportion of all fragments (C) across all repeat families. The normalized contribution was calculated by normalizing the proportion of short cfDNA fragments by the global proportion of short cfDNA fragments and the total proportion of all fragments mapped to that specific repeat element.

p-values calculated using two-sided Student's t-test. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Exact p-values are provided in Supplementary Data 3. Source data are provided as a Source Data file.


A
TelSeq (TRPM)
TelomereHunter (TRPM)
R^2 = 0.97
R^2 = 0.67
Patient Type HBC LFS NA
B
Telomere Content
Telomere Reads per Million
Healthy
LFS Negative
LFS Positive
27 123 38
C
Cancer History
Telomere Content (TRPM)
HBC
no
yes
27 66 95
H
Fragment Proportion vs Telomere Content
Proportion under 150bp
Telomere Content (TRPM)
R^2 = 0.0019
R^2 = 0.0016
D
Cancer Status
other
TCAGGG
TGAGGG
TTAGGG
Proportion
27 128 37
Cancer Status Healthy Negative Positive
E
Age
other
TCAGGG
TGAGGG
TTAGGG
0.044
0.004
Proportion
27 59 106
Age Healthy Pediatric Adult
G
Age
0.003
Telomere Content (TRPM)
27 59 102
HBC
pediatric
adult
F
Cancer Status
HBC
negative
positive
Proportion (%)
Context
other
TCAGGG
TGAGGG
TTAGGG
Age
HBC
pediatric
adult
Age (Years)

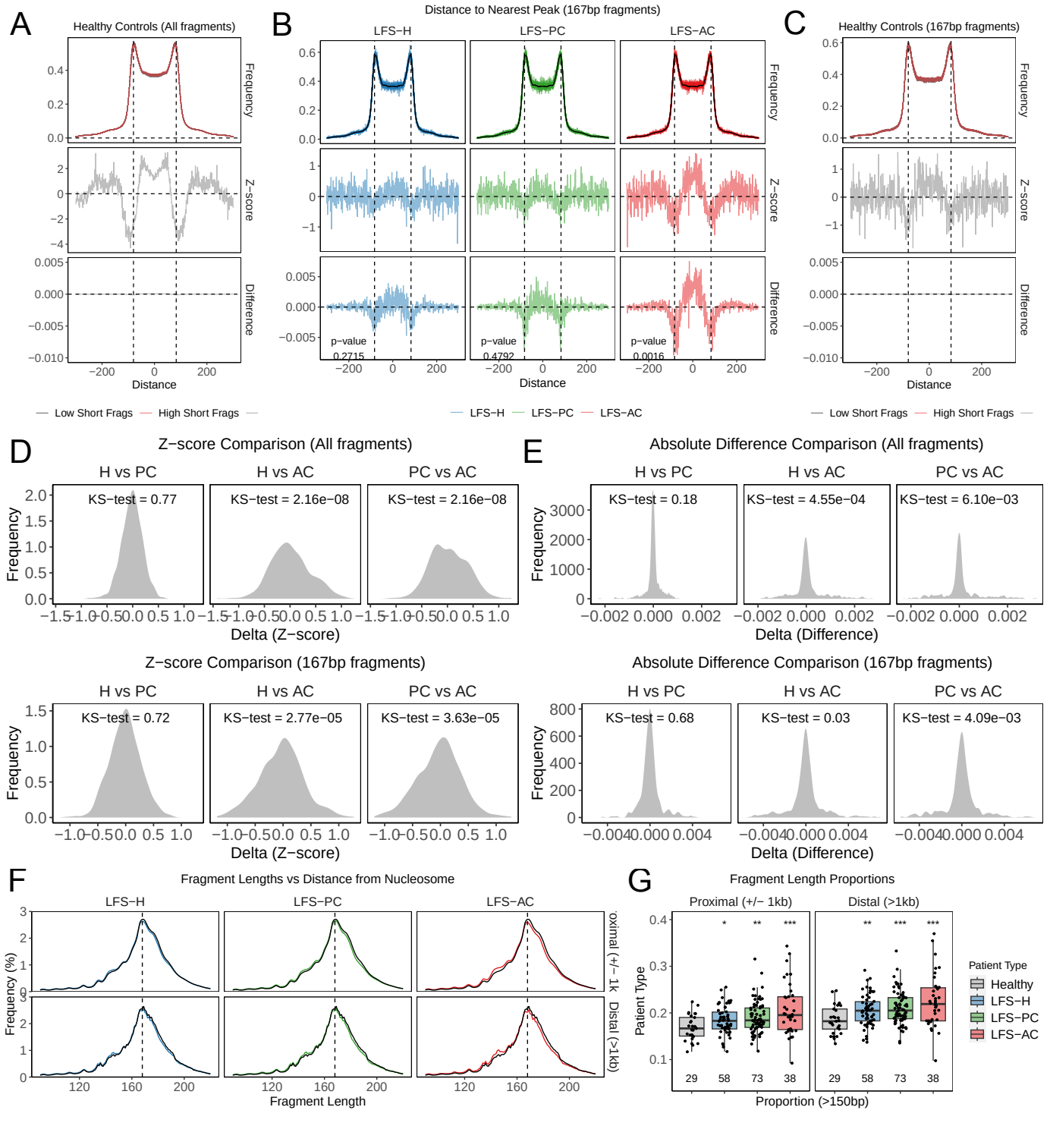
Supplementary Figure 5: Correlation of telomere length and content to cfDNA fragment length
A) Scatter plot comparing telomere content estimates from TelSeq and TelomereHunter.
Tukey boxplots comparing telomere content estimates between *TP53*-wildtype and *TP53m*-carriers across cancer status (B), history of cancer (C), and age (D).
E) Scatter plot comparing the proportion of short (<150bp) fragments and telomere content in *TP53*-carriers.
Tukey boxplots comparing the contribution from each telomere repeat sequencing across cancer status (F) and age (G)
H) Stacked barplot showing the frequency of telomere contexts between *TP53*-carriers and healthy *TP53*-wildtype controls stratified by cancer status (top), patient type (middle), and age (bottom).
R^2 values calculated using Pearson's correlation and p-values calculated using two-sided Student's t-test. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Source data are provided as a Source Data file.


A
Germline Mutation Type
Proportion under 150bp
0.021
(2) (15) (25) (5) (15) (19)
3 23 46 6 49 41
Missense 1
Missense 2
Missense 3
Missense 5
LOF
Splice
B
LFS Family
0.020
0.024
0.027
0.046
(2) (2) (2) (2) (2) (3) (2) (3) (4) (2) (4) (2) (3) (2) (2) (2) (3) (3) (37)
3 6 2 4 4 5 2 10 4 4 4 6 3 2 4 5 10 3 88
5928 (P222L)
5916 (R181C)
5948 (R158H)
0068 (R181H)
5954 (R273C)
5974 (R213Q)
5979 (R213Q)
4534 (S240G)
0067 (R213Q)
0070 (G245S)
0032 (F338S)
4509 (E294fs)
5971 (T125=_splice)
0033 (X187_splice)
0035 (T125=_splice)
0027 (T125=_splice)
5932 (c.97−11C>G)
5963 (T125=_splice)
Remainder
Mutation Type
Missense 1
Missense 2
Missense 3
Missense 5
LOF
Splice
Mixed
Cancer Status
Negative
Positive

Supplementary Figure 6: Comparison of the proportion of short cfDNA fragments to *TP53* germline mutation classes and patient families

Tukey boxplots comparing the proportion of short cfDNA fragments across *TP53* mutation classes (A) and LFS families (B). Number of samples is shown at the bottom, with the number of patients in brackets. Germline mutations with multiple families are bolded.

p-values calculated using two-sided Student's t-test. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Source data are provided as a Source Data file.


A
Healthy Controls (All fragments)
Frequency
Z-score
Difference
Distance
Low Short Frags
High Short Frags
B
Distance to Nearest Peak (167bp fragments)
LFS−H
LFS−PC
LFS−AC
p−value
0.2715
p−value
0.4792
p−value
0.0016
Distance
LFS−H
LFS−PC
LFS−AC
C
Healthy Controls (167bp fragments)
Low Short Frags
High Short Frags
D
Z−score Comparison (All fragments)
H vs PC
H vs AC
PC vs AC
KS−test = 0.77
KS−test = 2.16e−08
KS−test = 2.16e−08
Frequency
Delta (Z−score)
Z−score Comparison (167bp fragments)
KS−test = 0.72
KS−test = 2.77e−05
KS−test = 3.63e−05
E
Absolute Difference Comparison (All fragments)
KS−test = 0.18
KS−test = 4.55e−04
KS−test = 6.10e−03
Delta (Difference)
Absolute Difference Comparison (167bp fragments)
KS−test = 0.68
KS−test = 0.03
KS−test = 4.09e−03
F
Fragment Lengths vs Distance from Nucleosome
LFS−H
LFS−PC
LFS−AC
Proximal (+/− 1k)
Distal (>1kb)
Frequency (%)
Fragment Length
G
Fragment Length Proportions
Proximal (+/− 1kb)
Distal (>1kb)
Patient Type
Proportion (>150bp)
Patient Type
Healthy
LFS−H
LFS−PC
LFS−AC
29
58
73
38

Supplementary Figure 7: Nucleosome positioning is altered in *TP53m*-carriers

A) Frequency distribution (top), z-scores (middle), and absolution difference (bottom) of the distance of fragment starts and ends to the closest nucleosome peak in healthy *TP53*-wildtype controls separated by those with high (red) and low (black) proportion of short (<150bp) fragments (cohort split by median). Y-axis of the absolute difference panel is at the same scale as in Figure 4A and Supplementary Figures 9B/C for comparison.

B) Frequency distribution (top), z-scores (middle), and absolution difference (bottom) of the distance of 167bp fragment starts and ends to the closest nucleosome peak in *TP53*-carriers separated by cancer status compared to healthy *TP53*-wildtype controls (black).

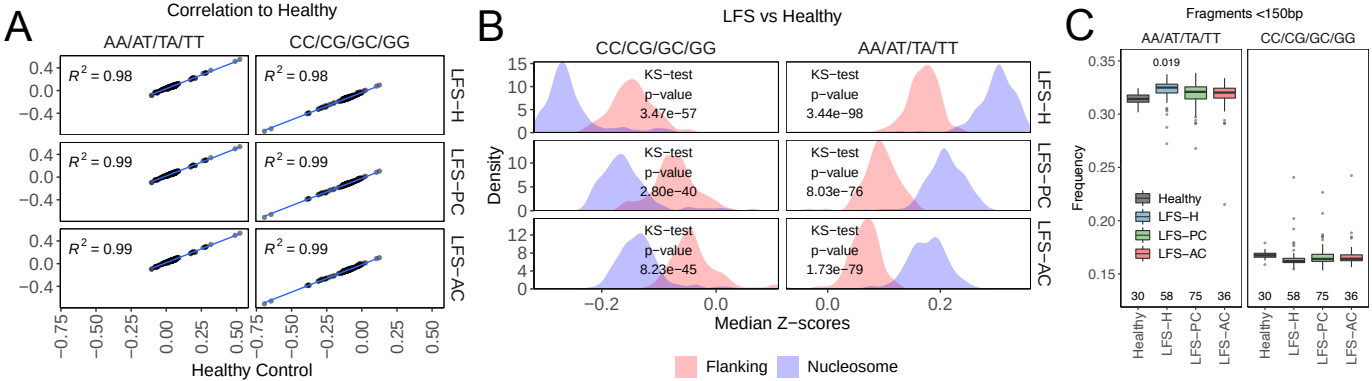
C) Frequency distribution (top), z-scores (middle), and absolution difference (bottom) of the distance of 167bp fragment starts and ends to the closest nucleosome peak in healthy *TP53*-wildtype controls separated by those with high (red) and low (black) proportion of short (<150bp) fragments (cohort split by median). Y-axis of the absolute difference chart is the same as in Supplementary Figure 9B for scale.

Histogram showing the difference in Z-score values (compared to *TP53*-wildtype) (D) and absolute difference (E) within *TP53m*-carriers using all fragments (top) and only nucleosome spanning fragments (167bp, bottom).

F) Frequency distributions of fragment lengths for fragments proximal (<1000bp) and distal (>1000bp) from nucleosome peaks from *TP53*-carriers separated by cancer status. Healthy median is in black.

G) Tukey boxplots showing the proportion of short (<150bp) fragments proximal (<1000bp) and distal (>1000bp) from nucleosome peaks in *TP53*-carriers separated by cancer status.

R^2 values calculated using Pearson's correlation and p-values calculated using two-sided Student's t-test or Kolmogorov–Smirnov. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Exact p-values are provided in Supplementary Data 3. Source data are provided as a Source Data file.


A
Correlation to Healthy
AA/AT/TA/TT
CC/CG/GC/GG
LFS−H
LFS−PC
LFS−AC
R^2 = 0.98
R^2 = 0.99
Healthy Control
B
LFS vs Healthy
CC/CG/GC/GG
AA/AT/TA/TT
KS−test
p−value
3.47e−57
3.44e−98
2.80e−40
8.03e−76
8.23e−45
1.73e−79
Density
Median Z−scores
Flanking
Nucleosome
C
Fragments <150bp
AA/AT/TA/TT
CC/CG/GC/GG
0.019
Frequency
Healthy
LFS−H
LFS−PC
LFS−AC
30
58
75
36

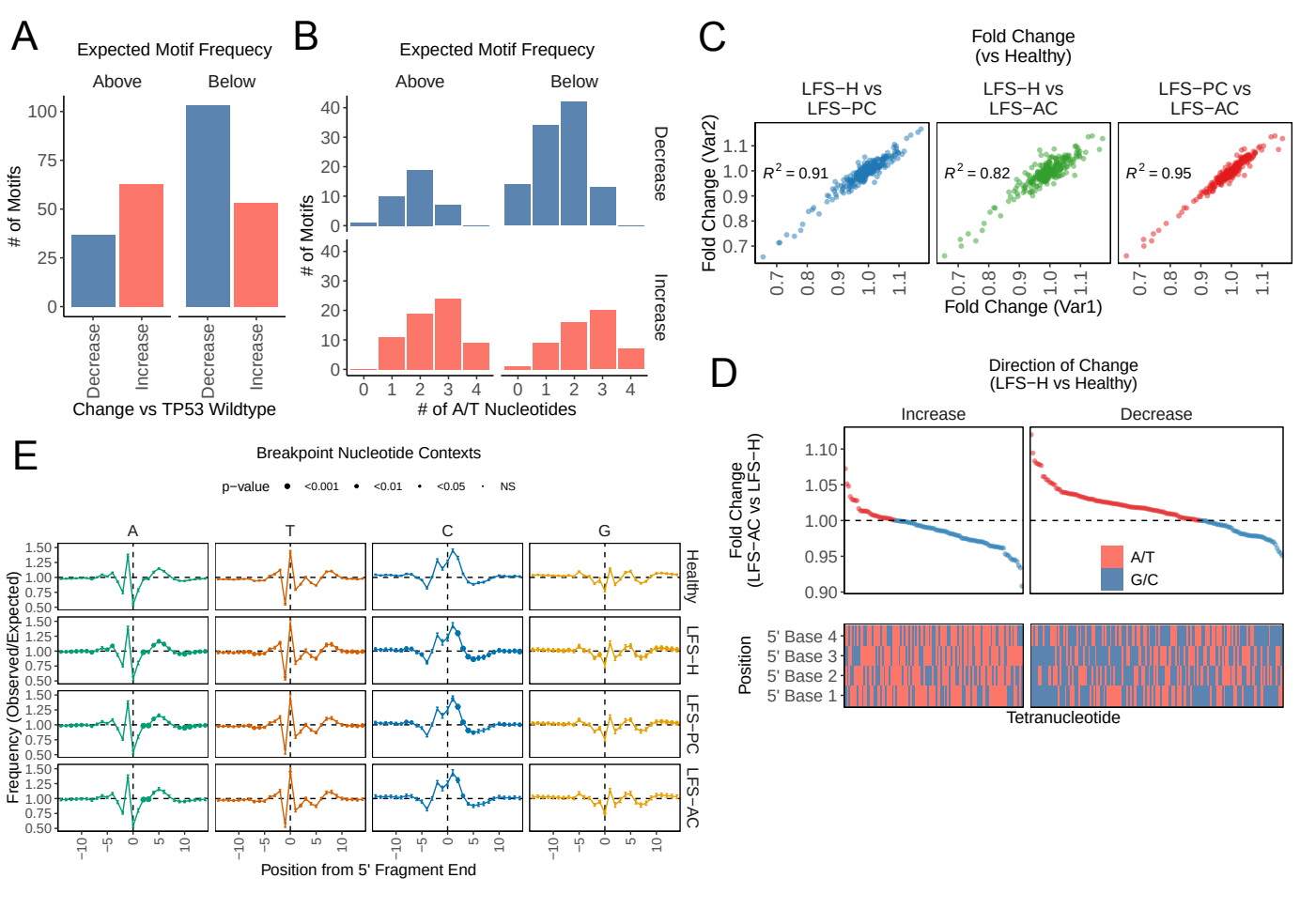
Supplementary Figure 8: Fragmentation is biased towards AA/AT/TA/TT dinucleotide contexts in *TP53m*-carriers

A) Scatter plots comparing the frequencies of AA/AT/TA/TT and CC/CG/GC/GG dinucleotide contexts at each position between *TP53*-carriers split by cancer status and healthy *TP53*-wildtype controls.

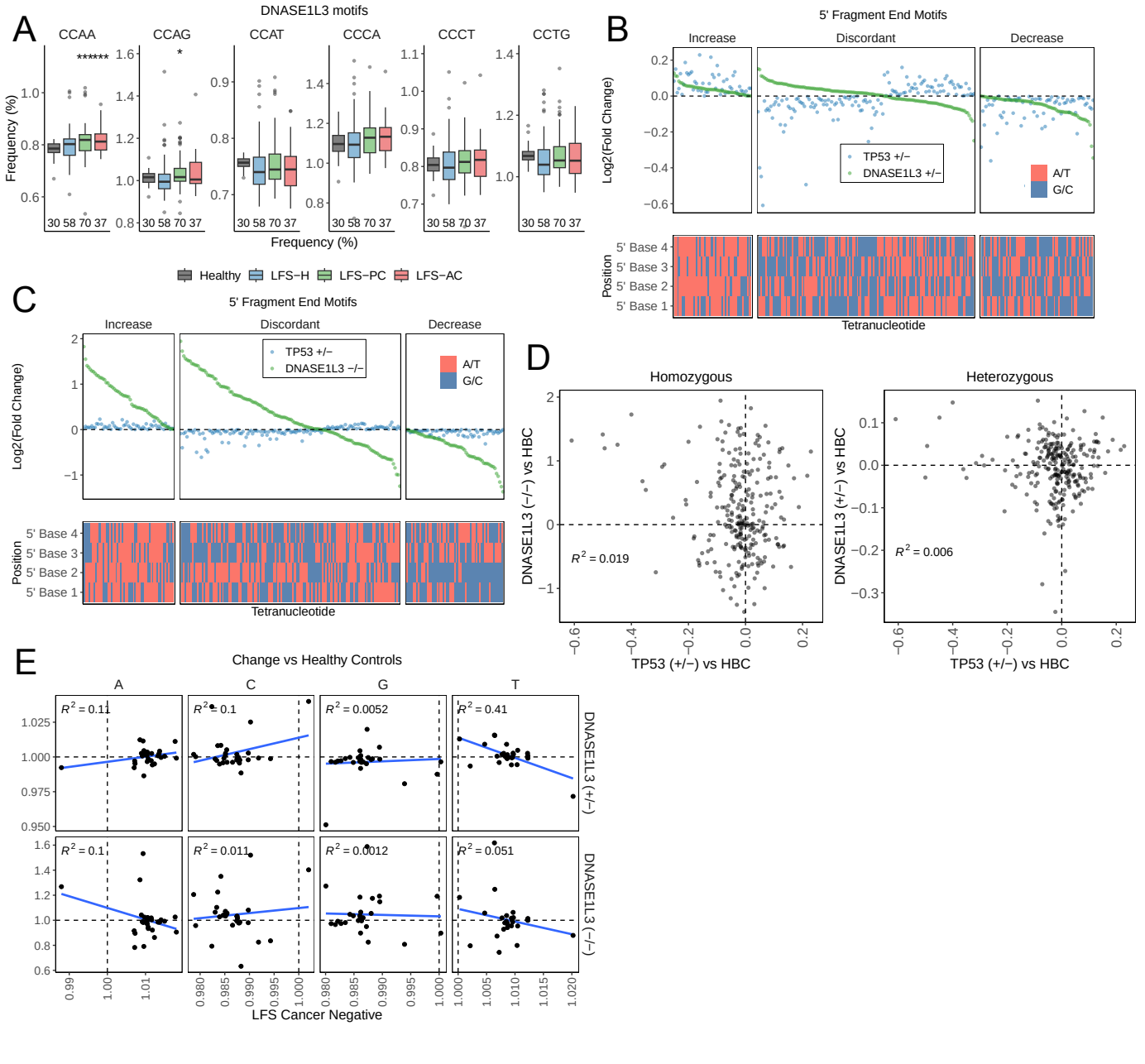
B) Density plots of z-scores calculated using the frequencies of CC/CG/GC/GG and AA/AT/TA/TT dinucleotide contexts in *TP53*-carriers, split by cancer status, and healthy *TP53*-wildtype controls. Z-scores are further split by positions flanking (+/-50bp) and spanning (167bp fragment) the nucleosome. P-values were calculated using the Kolmogorov–Smirnov test.

C) Tukey boxplots showing the proportion of A/T and G/C dinucleotide contexts using only short (<150bp) fragments.

R^2 values calculated using Pearson's correlation and p-values calculated using two-sided Student's t-test or Kolmogorov–Smirnov. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Source data are provided as a Source Data file.


A
Expected Motif Frequecy
Above
Below
# of Motifs
Decrease
Increase
Change vs TP53 Wildtype
B
Expected Motif Frequecy
Above
Below
Decrease
Increase
# of Motifs
# of A/T Nucleotides
C
Fold Change
(vs Healthy)
LFS−H vs
LFS−PC
LFS−H vs
LFS−AC
LFS−PC vs
LFS−AC
R^2 = 0.91
R^2 = 0.82
R^2 = 0.95
Fold Change (Var2)
Fold Change (Var1)
D
Direction of Change
(LFS−H vs Healthy)
Increase
Decrease
Fold Change
(LFS−AC vs LFS−H)
A/T
G/C
Position
5' Base 4
5' Base 3
5' Base 2
5' Base 1
Tetranucleotide
E
Breakpoint Nucleotide Contexts
p−value
<0.001
<0.01
<0.05
NS
A
T
C
G
Healthy
LFS−H
LFS−PC
LFS−AC
Frequency (Observed/Expected)
Position from 5' Fragment End

Supplementary Figure 9: Increase representation of A/T rich end-motifs in *TP53m*-carriers.
A) Barplot showing the number of end motifs that increased or decreased in frequency, in *TP53m*-carriers compared to *TP53*-wildtype, split by end motifs that occur above and below the expected random frequency.
B) Barplots showing data from panel A split by the number of A/T nucleotides within the motif.
C) Scatter plot comparing the fold changes of *TP53*-carriers between different cancer statuses. Fold changes were calculated compared to the healthy *TP53*-wildtype control cohort.
D) Dotplot showing the fold change of each tetranucleotide motif between LFS-H and LFS-AC separated by motifs that were increased or decreased in LFS-H versus healthy *TP53*-wildtype controls (top). Tile plot showing the GC/AT content of each tetranucleotide motif (bottom).
E) Line plot showing the nucleotide frequency at each position up and downstream from the fragment cut site. Error bars represent the standard deviation and points are sized by p-value compared to the *TP53*-wildtype controls.
R^2 values calculated using Pearson's correlation and p-values calculated using two-sided Student's t-test. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Source data are provided as a Source Data file.


A
DNASE1L3 motifs
CCAA
CCAG
CCAT
CCCA
CCCT
CCTG
Frequency (%)
30 58 70 37
Frequency (%)
Healthy
LFS−H
LFS−PC
LFS−AC
B
5' Fragment End Motifs
Increase
Discordant
Decrease
Log2(Fold Change)
TP53 +/−
DNASE1L3 +/−
A/T
G/C
Position
5' Base 4
5' Base 3
5' Base 2
5' Base 1
Tetranucleotide
C
5' Fragment End Motifs
Increase
Discordant
Decrease
Log2(Fold Change)
TP53 +/−
DNASE1L3 −/−
A/T
G/C
Position
5' Base 4
5' Base 3
5' Base 2
5' Base 1
Tetranucleotide
D
Homozygous
Heterozygous
DNASE1L3 (−/−) vs HBC
DNASE1L3 (+/−) vs HBC
R^2 = 0.019
R^2 = 0.006
TP53 (+/−) vs HBC
TP53 (+/−) vs HBC
E
Change vs Healthy Controls
A
C
G
T
R^2 = 0.11
R^2 = 0.1
R^2 = 0.0052
R^2 = 0.41
R^2 = 0.1
R^2 = 0.011
R^2 = 0.0012
R^2 = 0.051
DNASE1L3 (+/−)
DNASE1L3 (−/−)
LFS Cancer Negative

Supplementary Figure 10: Differences in cfDNA end-motifs observed in *TP53m*-carriers is not driven by dysregulation of *DNASE1L3*

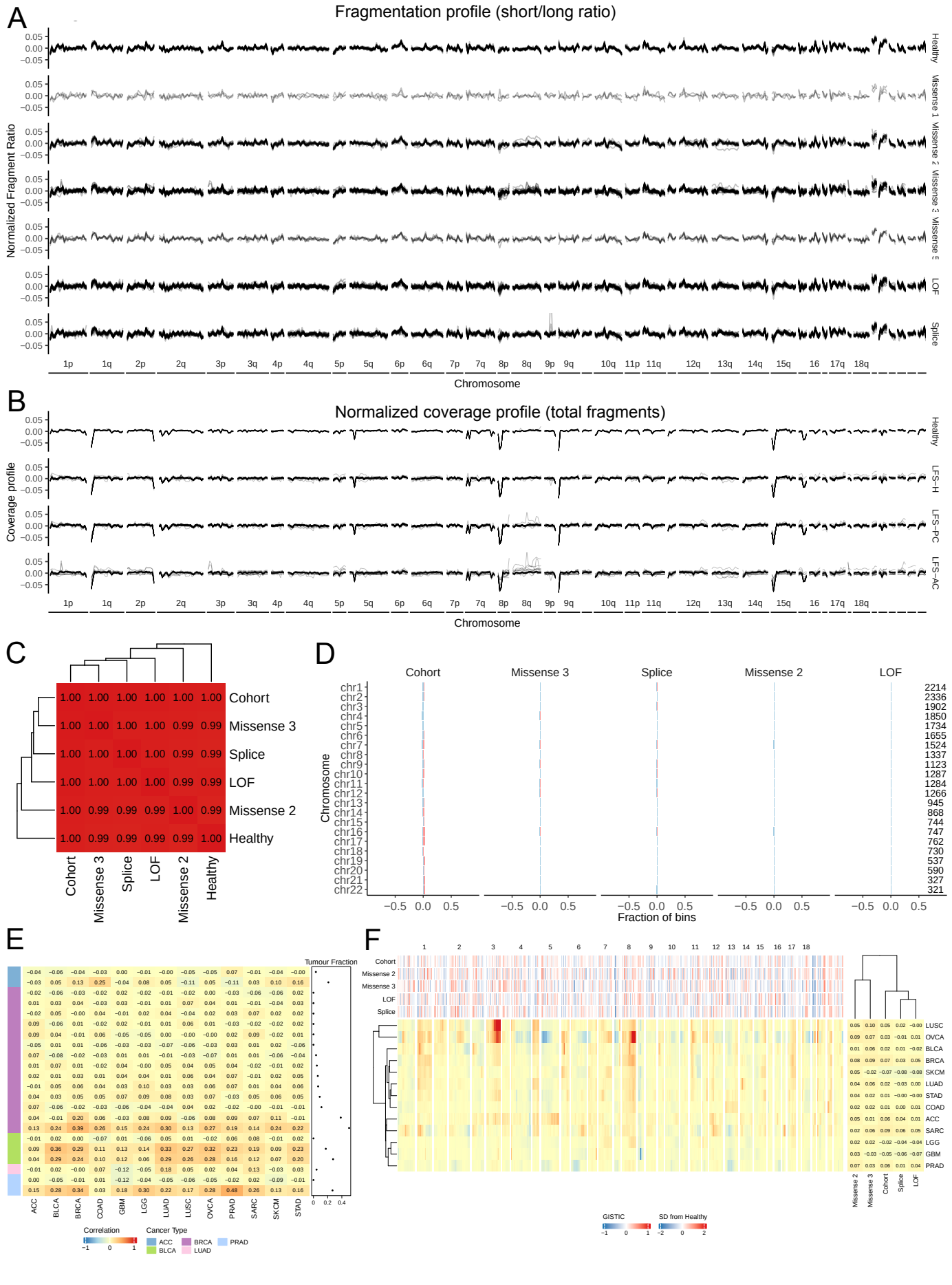
A) Tukey boxplots comparing the frequency of the top 6 end-motifs recognized by DNASE1L3 across healthy *TP53*-wildtype controls and *TP53*-carriers.

Dotplots showing the fold change of *TP53*-carriers (blue) versus heterozygous (B) and homozygous (C) *DNASE1L3* patients (green) versus healthy *TP53*-wildtype controls. Tile plots showing the GC/AT content of each tetranucleotide motif are shown below. Motifs are sorted by those that are increased or decreased in both and those that are discordant.

D) Scatter plots comparing the fold change in end motif frequency in *TP53m*-carriers and *DNASE1L3* mutant carriers.

E) Scatter plots comparing the nucleotide frequencies 10 basepairs up and downstream from the fragment cut site in *TP53m*-carriers and *DNASE1L3* mutant carriers.

R^2 values calculated using Pearson's correlation and p-values calculated using two-sided Student's t-test. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Exact p-values are provided in Supplementary Data 3. Source data are provided as a Source Data file.


A
Fragmentation profile (short/long ratio)
Normalized Fragment Ratio
Healthy
Missense 1
Missense 2
Missense 3
LOF
Splice
Chromosome
B
Normalized coverage profile (total fragments)
Coverage profile
Healthy
LFS−H
LFS−PC
LFS−AC
Chromosome
C
Cohort
Missense 3
Splice
LOF
Missense 2
Healthy
D
Cohort
Missense 3
Splice
Missense 2
LOF
Chromosome
Fraction of bins
E
Tumour Fraction
Correlation
Cancer Type
F
GISTIC
SD from Healthy

Supplementary Figure 11: cfDNA fragmentation profiles are not driven by differences in cfDNA coverage or by somatic copy number alterations in *TP53m*-carriers

A) Genome-wide fragment ratio profiles in healthy *TP53*-wildtype controls and *TP53*-carriers split by functional *TP53* mutation class.

B) Normalized genome-wide coverage profiles in healthy *TP53*-wildtype controls and *TP53*-carriers split by cancer status. Coverage profiles are affected by the discarding of fragments in regions of low mappability and those within the ENCODE blacklist.

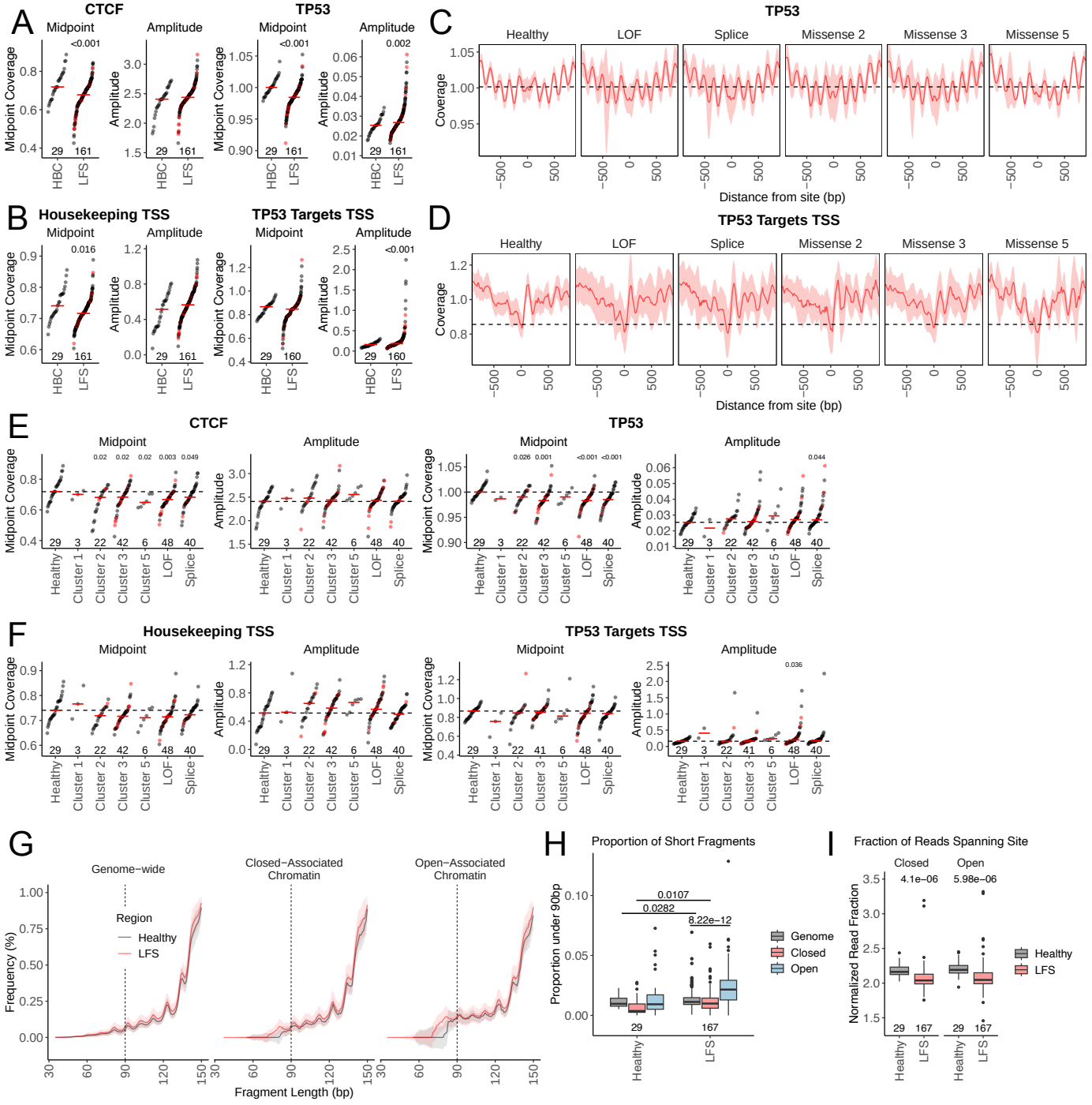
C) A heatmap showing the Pearson's correlation scores between each *TP53* mutation class and healthy *TP53*-wildtype controls using coverage profiles.

D) Barcharts showing the proportion of bins with significantly increased or decreased coverage in each *TP53* mutation class compared to the healthy *TP53*-wildtype control median.

E) Heatmap of Pearson's correlation scores comparing fragmentation ratio profiles from samples from *TP53*-carriers with active cancer and GISTIC copy number scores. IchorCNA tumor fractions are displayed on the right and cancer type displayed on the left.

F) Heatmap of TCGA GISTIC scores for cancer types associated with LFS. Heatmap of Pearson's correlation scores comparing the median fragmentation ratio profile of each germline mutation class to each GISTIC copy number profile on the right. Heatmap of the median number of standard deviations away from the healthy median for each germline mutation class on top.

Source data are provided as a Source Data file.


A
CTCF
Midpoint
Amplitude
TP53
Midpoint
Amplitude
Midpoint Coverage
Amplitude
<0.001
<0.001
0.002
29
161
HBC
LFS
B
Housekeeping TSS
TP53 Targets TSS
0.016
<0.001
160
C
TP53
Healthy
LOF
Splice
Missense 2
Missense 3
Missense 5
Coverage
Distance from site (bp)
D
TP53 Targets TSS
E
CTCF
TP53
0.02
0.02
0.02
0.003
0.049
0.026
0.001
<0.001
<0.001
0.044
3
22
42
6
48
40
Healthy
Cluster 1
Cluster 2
Cluster 3
Cluster 5
LOF
Splice
F
Housekeeping TSS
TP53 Targets TSS
41
0.036
G
Genome−wide
Closed−Associated Chromatin
Open−Associated Chromatin
Region
Healthy
LFS
Frequency (%)
Fragment Length (bp)
H
Proportion of Short Fragments
Proportion under 90bp
0.0107
0.0282
8.22e−12
Genome
Closed
Open
I
Fraction of Reads Spanning Site
Closed
Open
4.1e−06
5.98e−06
Normalized Read Fraction
Healthy
LFS

Supplementary Figure 12: Chromatin accessibility is not dependent on germline mutation context in *TP53m*-carriers

A) Dotplot showing the midpoint and amplitude of nucleosome positioning curves at CTCF (left) and TP53 (right) binding sites in healthy *TP53*-wildtype controls and *TP53*-carriers.

B) Dotplot showing the midpoint and amplitudes of nucleosome positioning curves at transcription start sites of housekeeping (left) and p53 target (right) genes in healthy *TP53*-wildtype controls and *TP53*-carriers.

C) Nucleosome positioning curves at p53 binding sites across *TP53* mutation classes.

D) Nucleosome positioning curves at the transcription start sites of p53 target genes across *TP53* mutation classes.

E) Dotplot showing the midpoint and amplitudes of nucleosome positioning curves at the transcription start sites of p53 binding sites in healthy across *TP53* mutation classes.

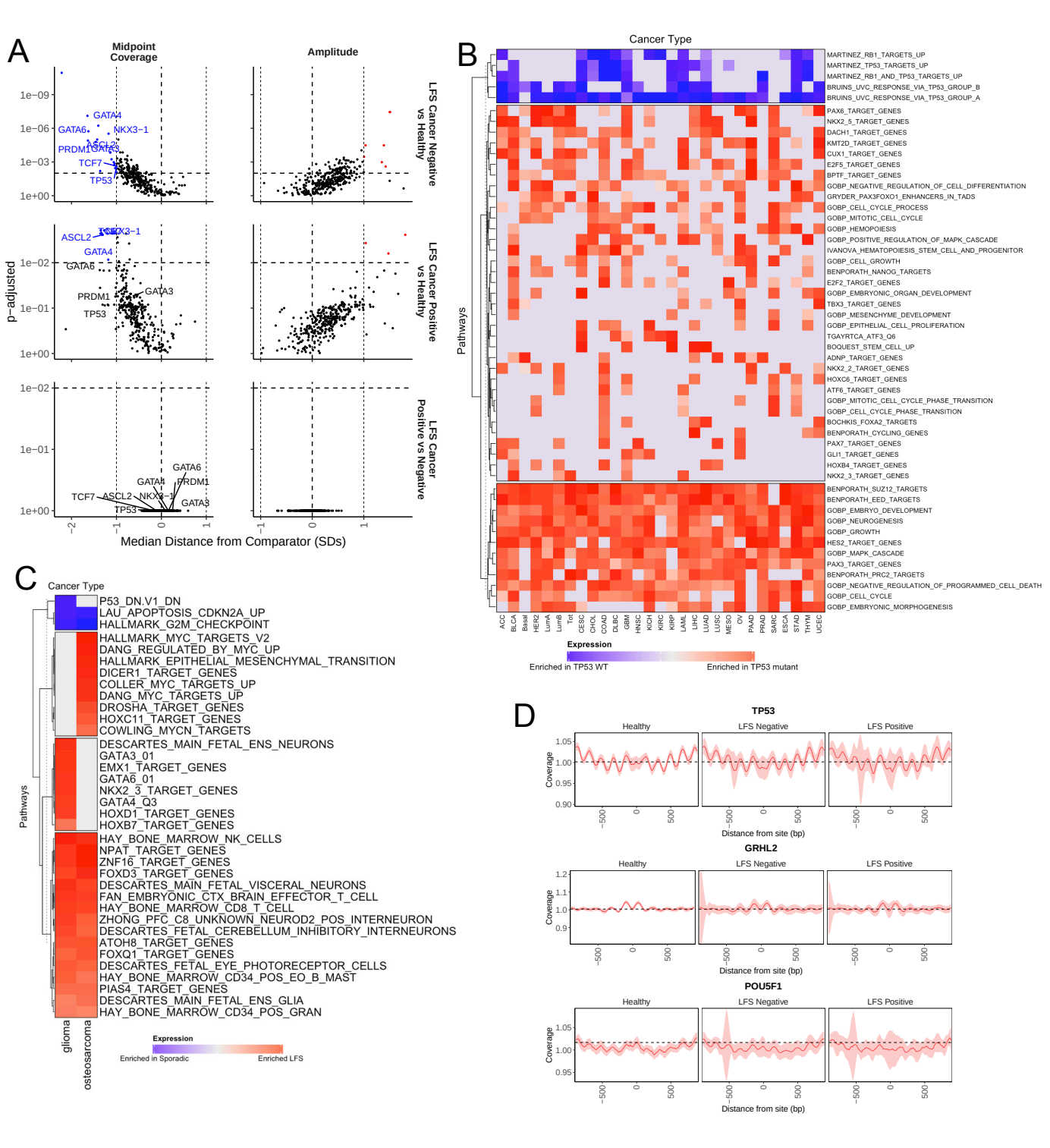
F) Dotplot showing the midpoint and amplitudes of nucleosome positioning curves at the transcription start sites of p53 target genes across *TP53* mutation classes.

G) Frequency distributions of fragment lengths for fragments genome-wide and fragments spanning hematopoietic open chromatin and closed chromatin sites in healthy *TP53*-wildtype controls and cancer negative *TP53m*-carriers.

H) Tukey boxplots showing the proportion of fragments under 90bp in healthy *TP53*-wildtype controls and *TP53m*-carriers at open and closed chromatin sites.

I) Tukey boxplots showing the fraction of total reads overlapping sites used in Supplementary Figures 5G and 5H.

p-values calculated using two-sided Student's t-test. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Source data are provided as a Source Data file.


A
Midpoint Coverage
Amplitude
LFS Cancer Negative vs Healthy
LFS Cancer Positive vs Healthy
LFS Cancer Positive vs Negative
p-adjusted
Median Distance from Comparator (SDs)
GATA4
GATA6
NKX3-1
ASCL2
PRDM1
GATA3
TCF7
TP53
B
Cancer Type
Pathways
MARTINEZ_RB1_TARGETS_UP
MARTINEZ_TP53_TARGETS_UP
MARTINEZ_RB1_AND_TP53_TARGETS_UP
BRUINS_UVC_RESPONSE_VIA_TP53_GROUP_B
BRUINS_UVC_RESPONSE_VIA_TP53_GROUP_A
PAX6_TARGET_GENES
NKX2_5_TARGET_GENES
DACH1_TARGET_GENES
KMT2D_TARGET_GENES
CUX1_TARGET_GENES
E2F5_TARGET_GENES
BPTF_TARGET_GENES
GOBP_NEGATIVE_REGULATION_OF_CELL_DIFFERENTIATION
GRYDER_PAX3FOXO1_ENHANCERS_IN_TADS
GOBP_CELL_CYCLE_PROCESS
GOBP_MITOTIC_CELL_CYCLE
GOBP_HEMOPOIESIS
GOBP_POSITIVE_REGULATION_OF_MAPK_CASCADE
IVANOVA_HEMATOPOIESIS_STEM_CELL_AND_PROGENITOR
GOBP_CELL_GROWTH
BENPORATH_NANOG_TARGETS
E2F2_TARGET_GENES
GOBP_EMBRYONIC_ORGAN_DEVELOPMENT
TBX3_TARGET_GENES
GOBP_MESENCHYME_DEVELOPMENT
GOBP_EPITHELIAL_CELL_PROLIFERATION
TGAYRTCA_ATF3_Q6
BOQUEST_STEM_CELL_UP
ADNP_TARGET_GENES
NKX2_2_TARGET_GENES
HOXC6_TARGET_GENES
ATF6_TARGET_GENES
GOBP_MITOTIC_CELL_CYCLE_PHASE_TRANSITION
GOBP_CELL_CYCLE_PHASE_TRANSITION
BOCHKIS_FOXA2_TARGETS
BENPORATH_CYCLING_GENES
PAX7_TARGET_GENES
GLI1_TARGET_GENES
HOXB4_TARGET_GENES
NKX2_3_TARGET_GENES
BENPORATH_SUZ12_TARGETS
BENPORATH_EED_TARGETS
GOBP_EMBRYO_DEVELOPMENT
GOBP_NEUROGENESIS
GOBP_GROWTH
HES2_TARGET_GENES
GOBP_MAPK_CASCADE
PAX3_TARGET_GENES
BENPORATH_PRC2_TARGETS
GOBP_NEGATIVE_REGULATION_OF_PROGRAMMED_CELL_DEATH
GOBP_CELL_CYCLE
GOBP_EMBRYONIC_MORPHOGENESIS
ACC BLCA Basal HER2 LumA LumB Tot CESC CHOL COAD DLBC GBM HNSC KICH KIRC KIRP LAML LIHC LUAD LUSC MESO OV PAAD PRAD SARC ESCA STAD THYM UCEC
Expression
Enriched in TP53 WT
Enriched in TP53 mutant
C
Cancer Type
Pathways
P53_DN.V1_DN
LAU_APOPTOSIS_CDKN2A_UP
HALLMARK_G2M_CHECKPOINT
HALLMARK_MYC_TARGETS_V2
DANG_REGULATED_BY_MYC_UP
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
DICER1_TARGET_GENES
COLLER_MYC_TARGETS_UP
DANG_MYC_TARGETS_UP
DROSHA_TARGET_GENES
HOXC11_TARGET_GENES
COWLING_MYCN_TARGETS
DESCARTES_MAIN_FETAL_ENS_NEURONS
GATA3_01
EMX1_TARGET_GENES
GATA6_01
NKX2_3_TARGET_GENES
GATA4_Q3
HOXD1_TARGET_GENES
HOXB7_TARGET_GENES
HAY_BONE_MARROW_NK_CELLS
NPAT_TARGET_GENES
ZNF16_TARGET_GENES
FOXD3_TARGET_GENES
DESCARTES_MAIN_FETAL_VISCERAL_NEURONS
FAN_EMBRYONIC_CTX_BRAIN_EFFECTOR_T_CELL
HAY_BONE_MARROW_CD8_T_CELL
ZHONG_PFC_C8_UNKNOWN_NEUROD2_POS_INTERNEURON
DESCARTES_FETAL_CEREBELLUM_INHIBITORY_INTERNEURONS
ATOH8_TARGET_GENES
FOXQ1_TARGET_GENES
DESCARTES_FETAL_EYE_PHOTORECEPTOR_CELLS
HAY_BONE_MARROW_CD34_POS_EO_B_MAST
PIAS4_TARGET_GENES
DESCARTES_MAIN_FETAL_ENS_GLIA
HAY_BONE_MARROW_CD34_POS_GRAN
glioma
osteosarcoma
Expression
Enriched in Sporadic
Enriched LFS
D
TP53
GRHL2
POU5F1
Healthy
LFS Negative
LFS Positive
Coverage
Distance from site (bp)

Supplementary Figure 13: *TP53m*-carriers display differential chromatin accessibility across transcription factor binding sites which are recapitulated in expression profiles of *TP53* mutant cancers

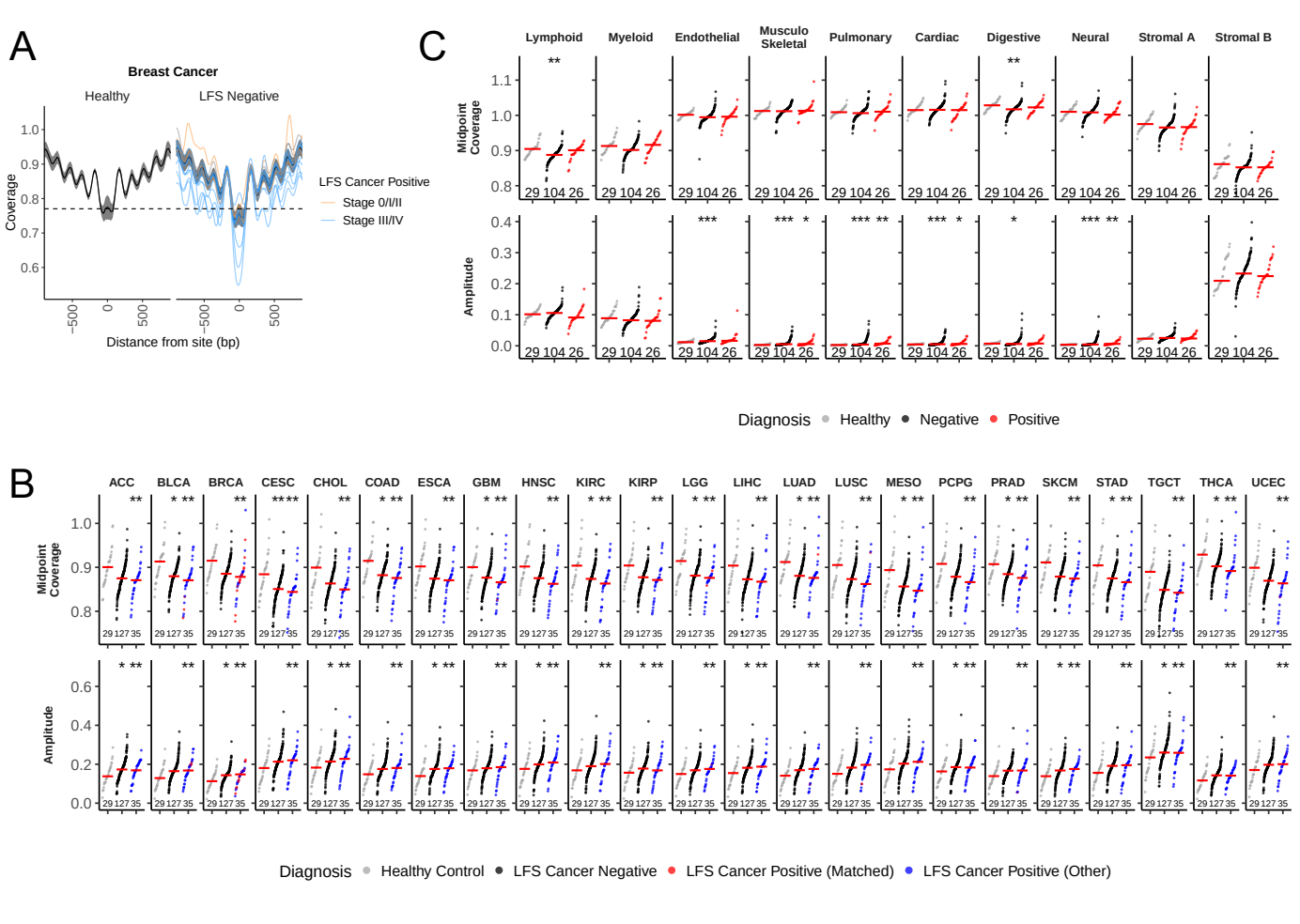
A) Volcano plots of transcription factors with differential midpoint coverage and amplitude across several comparisons.

B) Heatmap showing a selection of differentially expressed gene pathways in *TP53* inactive cancers vs *TP53* intact cancers across 26 cohorts in the TCGA. Scores are shown as a ratio of enriched in *TP53* inactive over encriched in *TP53* intact.

C) Heatmap showing a selection of differentially expressed gene pathways in LFS-derived vs sporadic osteosarcomas and glioblastomas. Scores are shown as a ratio of enriched in *TP53* inactive over encriched in *TP53* intact.

D) Nucleosome positioning curves for p53 (top), GRHL2 (middle), and POU5F1 (bottom) in healthy *TP53*-wildtype controls, cancer negative and cancer positive *TP53m*-carriers. Dotted line is the midpoint coverage of healthy *TP53*-wildtype controls.

p-values calculated using two-sided Student's t-test. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Source data are provided as a Source Data file.


A
Breast Cancer
Healthy
LFS Negative
Coverage
1.0
0.9
0.8
0.7
0.6
−500
0
500
Distance from site (bp)
LFS Cancer Positive
Stage 0/I/II
Stage III/IV
C
Lymphoid
Myeloid
Endothelial
Musculo Skeletal
Pulmonary
Cardiac
Digestive
Neural
Stromal A
Stromal B
Midpoint Coverage
Amplitude
29 104 26
Diagnosis
Healthy
Negative
Positive
B
ACC
BLCA
BRCA
CESC
CHOL
COAD
ESCA
GBM
HNSC
KIRC
KIRP
LGG
LIHC
LUAD
LUSC
MESO
PCPG
PRAD
SKCM
STAD
TGCT
THCA
UCEC
Midpoint Coverage
Amplitude
29 127 35
Diagnosis
Healthy Control
LFS Cancer Negative
LFS Cancer Positive (Matched)
LFS Cancer Positive (Other)

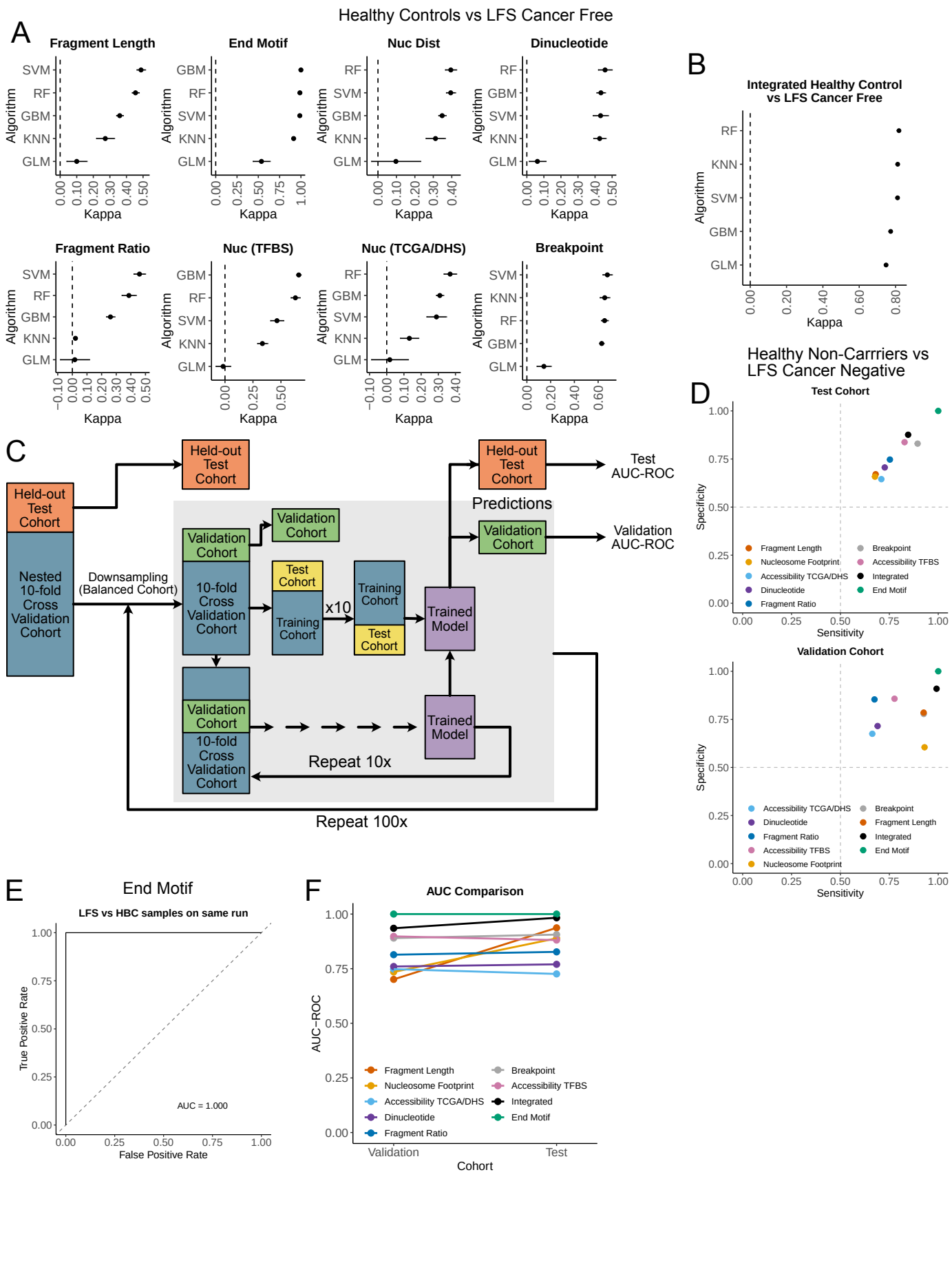
Supplementary Figure 14: *TP53m*-carriers show differential chromatin accessibility at cancer-associated sites but not normal tissue sites

A) Nucleosome positioning tracks showing chromatin accessibility at breast cancer-associated open chromatin sites. Patients with active breast cancer are also displayed.

B) Midpoint coverage and amplitude of nucleosome positioning curves of open chromatin sites associated with all TCGA cancer types.

C) Midpoint coverage and amplitude of nucleosome positioning curves of open chromatin sites associated with an array of normal tissues.

p-values calculated using two-sided Student's t-test. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Exact p-values are provided in Supplementary Data 3. Source data are provided as a Source Data file.


Healthy Controls vs LFS Cancer Free
A
Fragment Length
End Motif
Nuc Dist
Dinucleotide
Fragment Ratio
Nuc (TFBS)
Nuc (TCGA/DHS)
Breakpoint
Algorithm
Kappa
B
Integrated Healthy Control vs LFS Cancer Free
C
Held-out Test Cohort
Nested 10-fold Cross Validation Cohort
Downsampling (Balanced Cohort)
10-fold Cross Validation Cohort
Validation Cohort
Test Cohort
Training Cohort
x10
Training Cohort
Test Cohort
Trained Model
Predictions
Test AUC-ROC
Validation AUC-ROC
Repeat 10x
Repeat 100x
Healthy Non-Carrriers vs LFS Cancer Negative
D
Test Cohort
Validation Cohort
Specificity
Sensitivity
Fragment Length
Nucleosome Footprint
Accessibility TCGA/DHS
Dinucleotide
Fragment Ratio
Breakpoint
Accessibility TFBS
Integrated
End Motif
E
End Motif
LFS vs HBC samples on same run
True Positive Rate
False Positive Rate
AUC = 1.000
F
AUC Comparison
AUC−ROC
Validation
Test
Cohort

Supplementary Figure 15: Validation of cfDNA fragmentation-based classification between cancer negative *TP53m*-carrier and *TP53*-wildtype samples

A) Dotplot of kappa values and 95% confidence intervals for each machine learning algorithm tested in each fragmentomic metric using healthy *TP53*-wildtype controls and cancer negative *TP53m*-carriers.

B) Dotplot of kappa values and 95% confidence intervals for each machine learning algorithm tested in an integrated classifier using healthy *TP53*-wildtype controls and cancer negative *TP53*-carriers. Fragment end-motif was excluded.

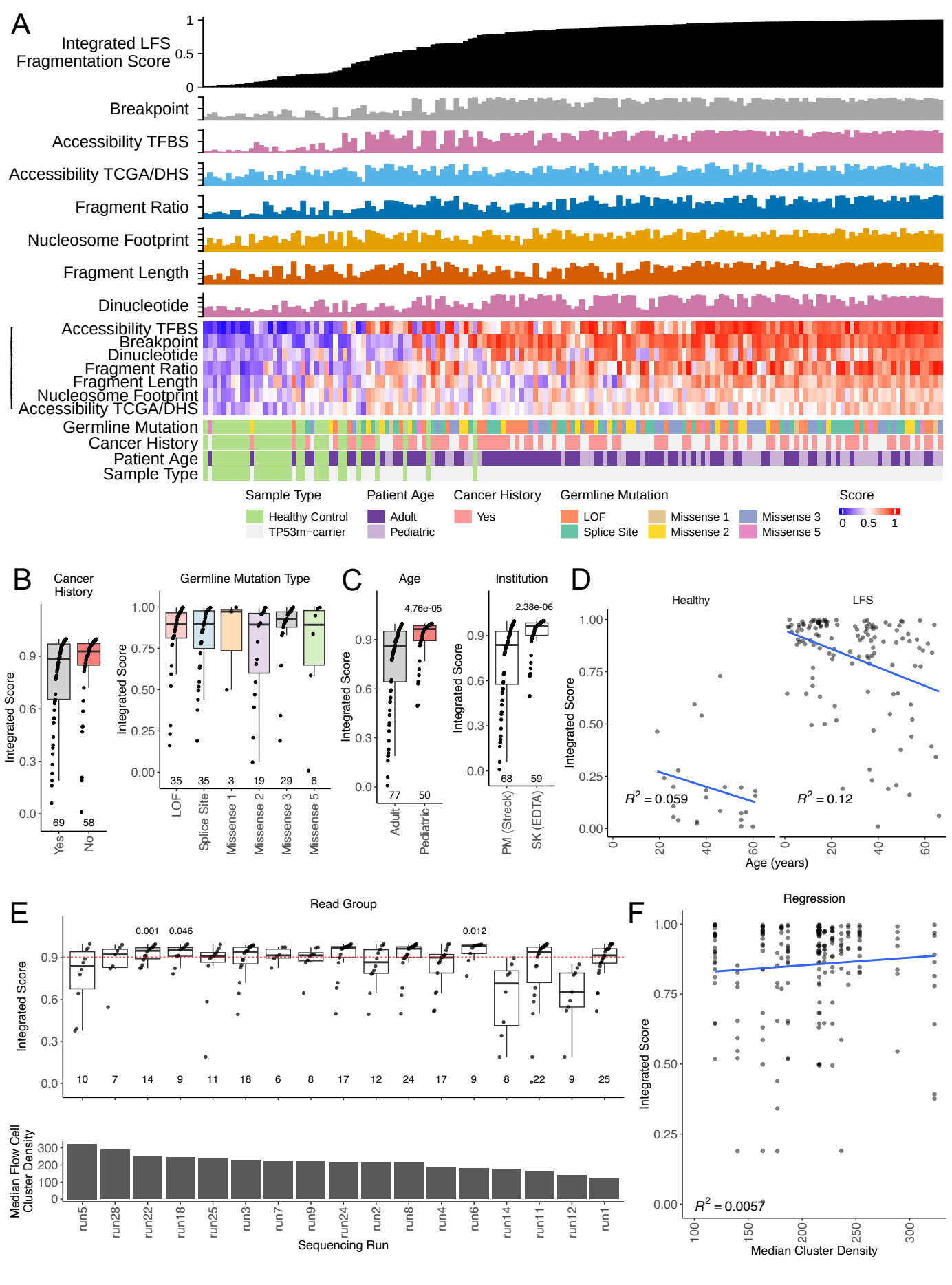
C) Schematic outlining methodology for machine learning models.

D) Dotplot of sensitivity and specificity values for each fragmentomic feature used for classifying *TP53*-wildtype controls and cancer negative *TP53*-carriers in the test set (left) and validation set (right).

E) ROC curve showing the performance of a classification between *TP53*-wildtype and cancer negative *TP53m*-carriers using fragment end-motifs. For this analysis, we identified and used only samples that were on sequencing runs where both *TP53*-wildtype and *TP53m*-carriers were pooled to control for technical variability. AUC is displayed.

F) Dot-plot comparing the AUC performance between test and validation sets for the classification for each fragmentomic feature using *TP53*-wildtype and *TP53m*-carriers. Similar performance was observed across fragmentomic features tested.

Source data are provided as a Source Data file.


A
Integrated LFS Fragmentation Score
Breakpoint
Accessibility TFBS
Accessibility TCGA/DHS
Fragment Ratio
Nucleosome Footprint
Fragment Length
Dinucleotide
Germline Mutation
Cancer History
Patient Age
Sample Type
Sample Type
Healthy Control
TP53m−carrier
Patient Age
Adult
Pediatric
Cancer History
Yes
Germline Mutation
LOF
Splice Site
Missense 1
Missense 2
Missense 3
Missense 5
Score
B
Cancer History
Integrated Score
Yes
No
Germline Mutation Type
LOF
Splice Site
Missense 1
Missense 2
Missense 3
Missense 5
C
Age
4.76e-05
Adult
Pediatric
Institution
2.38e-06
PM (Streck)
SK (EDTA)
D
Healthy
LFS
R^2 = 0.059
R^2 = 0.12
Age (years)
E
Read Group
0.001
0.046
0.012
Median Flow Cell Cluster Density
Sequencing Run
F
Regression
R^2 = 0.0057
Median Cluster Density

Supplementary Figure 16: Analysis of differential contribution of fragmentation features towards classification and technical correlates

A) Heatmap and barcharts of prediction scores for each fragmentomic metric input into the integrated *TP53*-wildtype vs *TP53m*-carriers classifier. Additional clinical information is also displayed on the bottom (germline mutation class, cancer history, patient age).

B) Tukey Tukey boxplots showing comparison of integrated fragmentation scores in *TP53m*-carriers separated by history of cancer (left) and germline mutation class (right).

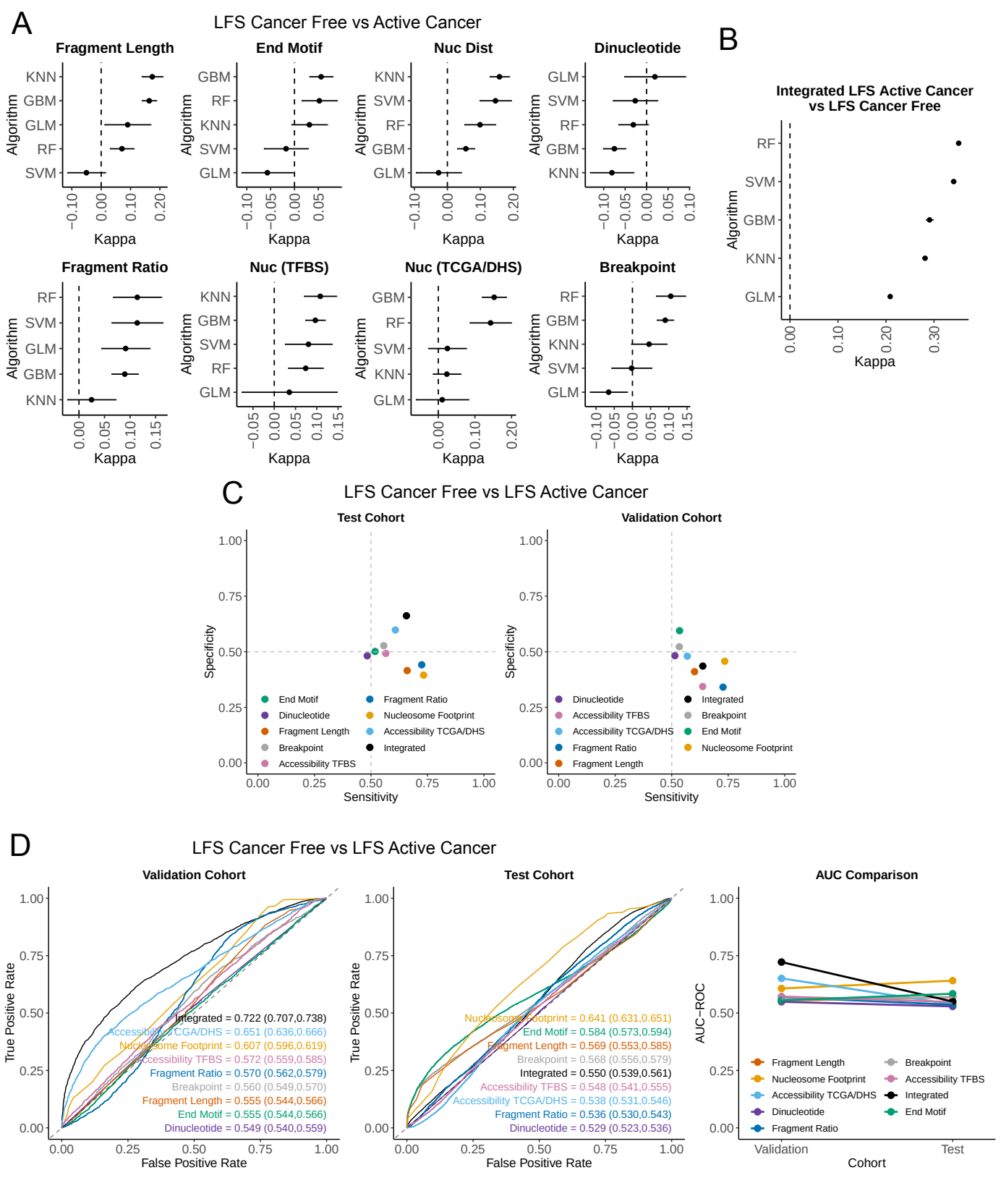
C) Tukey Tukey boxplots showing comparison of integrated fragmentation scores in *TP53m*-carriers separated by age (left) and institution/collection tube (right).

D) Scatter plot showing the age vs LFS fragmentation scores in *TP53m*-carriers and *TP53*-wildtype samples.

E) Tukey Tukey boxplots (top) and barplot (bottom) comparing LFS fragmentation scores across sequencing runs ordered by the cluster density.

F) Scatter plot comparing the LFS fragmentation score and the flow cell cluster density.

R^2 values calculated using Pearson's correlation and p-values calculated using two-sided Student's t-test. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Source data are provided as a Source Data file.


A
LFS Cancer Free vs Active Cancer
Fragment Length
End Motif
Nuc Dist
Dinucleotide
Fragment Ratio
Nuc (TFBS)
Nuc (TCGA/DHS)
Breakpoint
Algorithm
Kappa
B
Integrated LFS Active Cancer vs LFS Cancer Free
C
LFS Cancer Free vs LFS Active Cancer
Test Cohort
Validation Cohort
Specificity
Sensitivity
End Motif
Dinucleotide
Fragment Length
Breakpoint
Accessibility TFBS
Fragment Ratio
Nucleosome Footprint
Accessibility TCGA/DHS
Integrated
D
LFS Cancer Free vs LFS Active Cancer
Validation Cohort
Test Cohort
AUC Comparison
True Positive Rate
False Positive Rate
AUC−ROC
Integrated = 0.722 (0.707,0.738)
Accessibility TCGA/DHS = 0.651 (0.636,0.666)
Nucleosome Footprint = 0.607 (0.596,0.619)
Accessibility TFBS = 0.572 (0.559,0.585)
Fragment Ratio = 0.570 (0.562,0.579)
Breakpoint = 0.560 (0.549,0.570)
Fragment Length = 0.555 (0.544,0.566)
End Motif = 0.555 (0.544,0.566)
Dinucleotide = 0.549 (0.540,0.559)
Nucleosome Footprint = 0.641 (0.631,0.651)
End Motif = 0.584 (0.573,0.594)
Fragment Length = 0.569 (0.553,0.585)
Breakpoint = 0.568 (0.556,0.579)
Integrated = 0.550 (0.539,0.561)
Accessibility TFBS = 0.548 (0.541,0.555)
Accessibility TCGA/DHS = 0.538 (0.531,0.546)
Fragment Ratio = 0.536 (0.530,0.543)
Dinucleotide = 0.529 (0.523,0.536)
Validation
Test
Cohort

Supplementary Figure 17: Validation of cfDNA fragmentation-based clasification of cancer positive and cancer negative *TP53m*-carriers

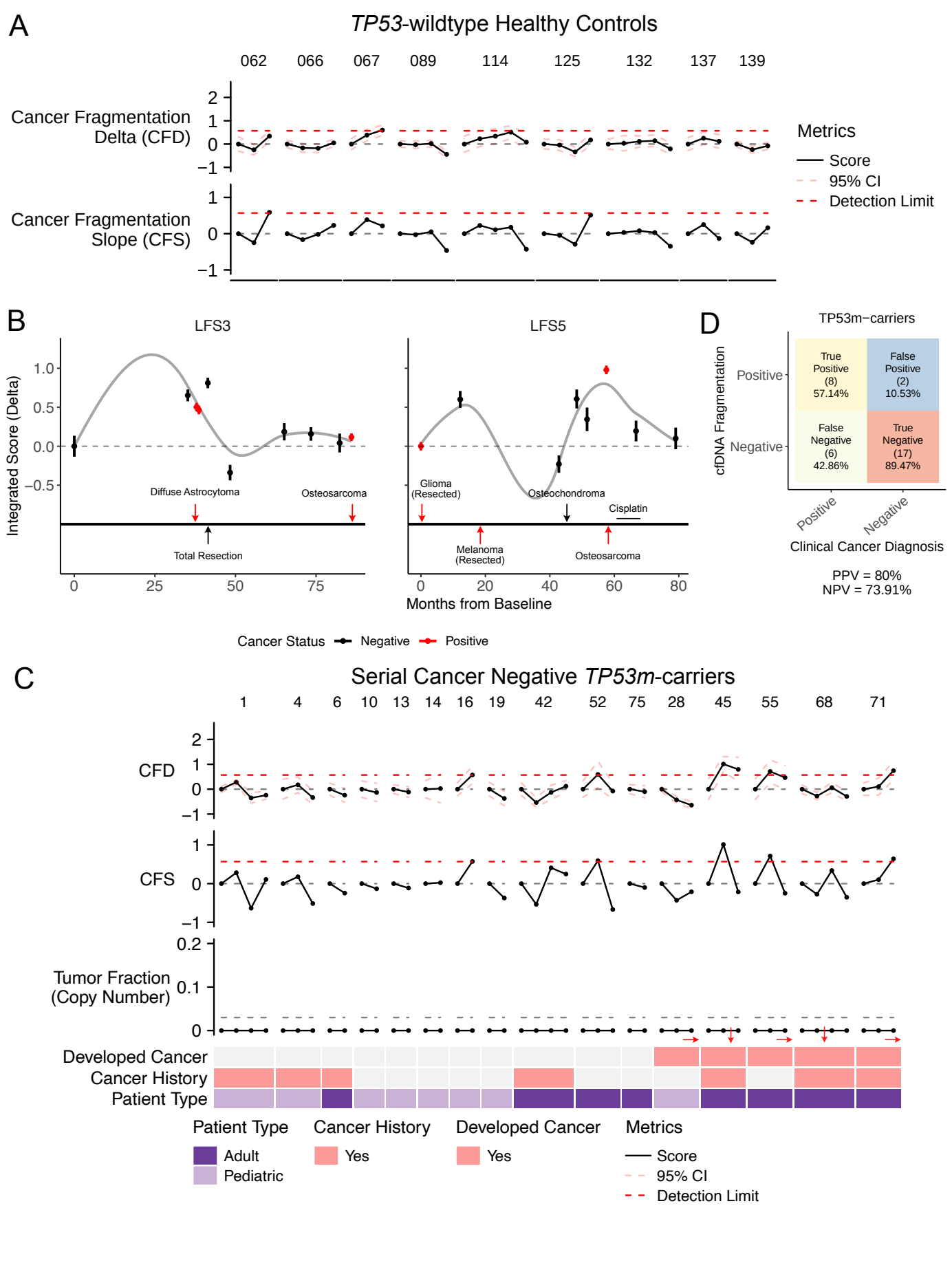
A) Dotplot of kappa values and 95% confidence intervals for each machine learning algorithm tested in each fragmentomic metric using cancer negative and cancer positive *TP53*-carriers.

B) Dotplot of kappa values and 95% confidence intervals for each machine learning algorithm tested in an integrated classifier using cancer negative and cancer positive *TP53*-carriers.

C) Dotplot of sensitivity and specificity values for each fragmentomic feature used for classifying cancer negative and cancer positive *TP53*-carriers in the test set (left) and validation set (right).

D) ROC curves for test set (left), validation set (middle) and a dot plot comparing AUC performance (right) for each fragmentomic feature. Models were trained using cancer negative and cancer positive *TP53*-carriers.

Source data are provided as a Source Data file.


A
TP53-wildtype Healthy Controls
062
066
067
089
114
125
132
137
139
Cancer Fragmentation Delta (CFD)
Cancer Fragmentation Slope (CFS)
Metrics
Score
95% CI
Detection Limit
B
LFS3
LFS5
Integrated Score (Delta)
Diffuse Astrocytoma
Osteosarcoma
Total Resection
Glioma (Resected)
Osteochondroma
Cisplatin
Melanoma (Resected)
Osteosarcoma
Months from Baseline
Cancer Status
Negative
Positive
D
TP53m−carriers
cfDNA Fragmentation
Positive
Negative
True Positive (8) 57.14%
False Positive (2) 10.53%
False Negative (6) 42.86%
True Negative (17) 89.47%
Clinical Cancer Diagnosis
PPV = 80%
NPV = 73.91%
C
Serial Cancer Negative TP53m-carriers
1
4
6
10
13
14
16
19
42
52
75
28
45
55
68
71
CFD
CFS
Tumor Fraction (Copy Number)
Developed Cancer
Cancer History
Patient Type
Patient Type
Adult
Pediatric
Cancer History
Yes
Developed Cancer
Yes
Metrics
Score
95% CI
Detection Limit

Supplementary Figure 18: Analysis of longitudinal patient-specific cancer framgnetaion scores in healthy *TP53*-wildtype and cancer negative *TP53m*-carriers.
A) Longitudinal CFS for *TP53*-wildtype controls with serial samples. Scores were converted into delta scores using the baseline timepoint as reference.
B) Longitudinal CFS for cancer negative patients with serial samples. Scores were converted into delta scores using the baseline timepoint as reference. Patient type, cancer history, and whether the patient eventually developed cancer are displayed. Arrows denote where cancer developed in relation to the plasma timepoints analyzed (between timepoints or following last timepoint).
C) To-scale timelines and integrated fragment scores for three patients that developed multiple cancers. Cancers and any clinical treatments are annotated.
D) Confusion matrix comparing the rates of cancer detection using cell-free cancer fragmentation. Clinical diagnosis was used as the ground truth.
Source data are provided as a Source Data file.

Supplementary Table 1

Fragmentomic Feature	Comparison	Optimal Algorithm
Fragment Length	Healthy Control vs LFS Cancer Free	Support Vector Machine
End Motif	Healthy Control vs LFS Cancer Free	Gradient Boosted Machine
Breakpoint	Healthy Control vs LFS Cancer Free	Support Vector Machine
Nucleosome Footprint	Healthy Control vs LFS Cancer Free	Random Forest
Dinucleotide Context	Healthy Control vs LFS Cancer Free	Random Forest
Fragmentation Profile	Healthy Control vs LFS Cancer Free	Support Vector Machine
Nucleosome Accessibility (TCGA/DHS)	Healthy Control vs LFS Cancer Free	Gradient Boosted Machine
Nucleosome Accessibility (TFBS)	Healthy Control vs LFS Cancer Free	Random Forest
Integrated	Healthy Control vs LFS Cancer Free	Random Forest
Fragment Length	LFS Cancer Free vs LFS Active Cancer	K Nearest Neighbors
End Motif	LFS Cancer Free vs LFS Active Cancer	Gradient Boosted Machine
Breakpoint	LFS Cancer Free vs LFS Active Cancer	Random Forest
Nucleosome Footprint	LFS Cancer Free vs LFS Active Cancer	K Nearest Neighbors
Dinucleotide Context	LFS Cancer Free vs LFS Active Cancer	Generalized Linear Model
Fragmentation Profile	LFS Cancer Free vs LFS Active Cancer	Random Forest
Nucleosome Accessibility (TCGA/DHS)	LFS Cancer Free vs LFS Active Cancer	K Nearest Neighbors
Nucleosome Accessibility (TFBS)	LFS Cancer Free vs LFS Active Cancer	Gradient Boosted Machine
Integrated	LFS Cancer Free vs LFS Active Cancer	Random Forest