Table of Contents

Figure S1. Summary of long-read sequencing data and somatic SV detection workflow.

A. The N50 distribution of nanopore sequencing reads for each sample. Orange bars show the average ± standard deviation (SD) of each population.

B. The distribution of nanopore sequencing depth for each sample. Orange bars show the average ± standard deviation (SD) of each population.

C. Flowchart depicting the workflow for somatic SV identification using paired long-read sequencing samples.

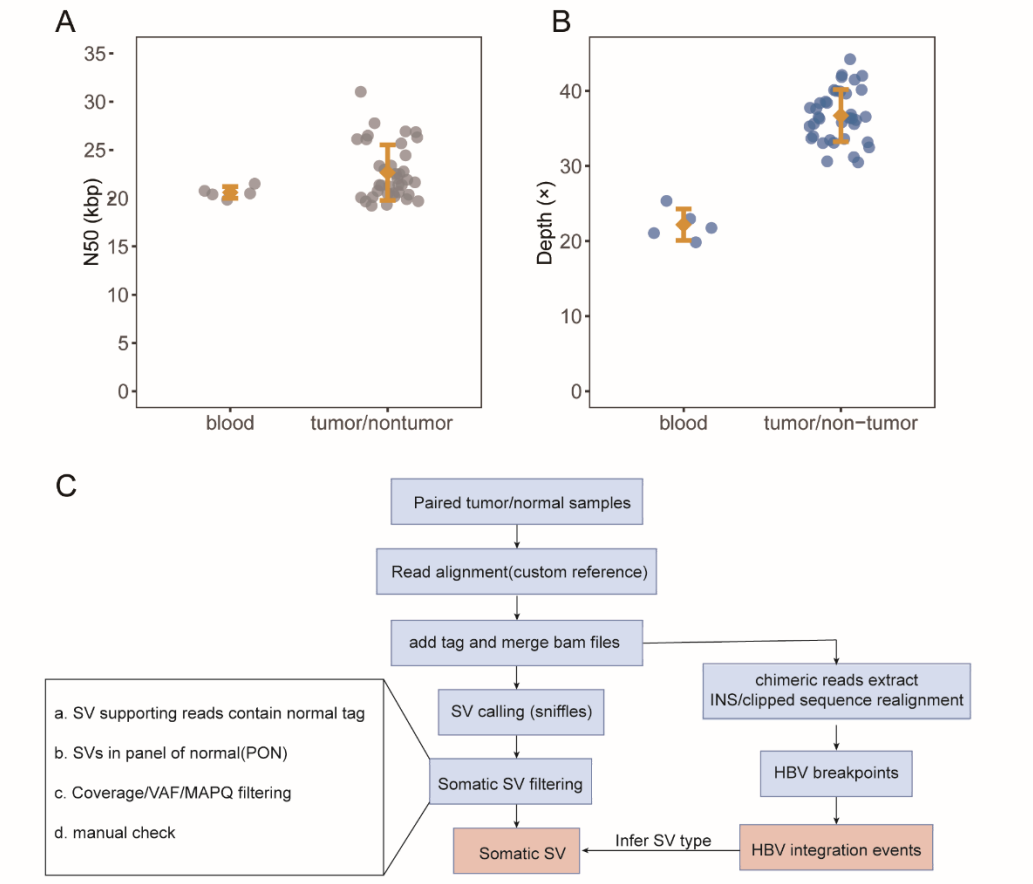


Figure S2. Annotation of genomic elements for somatic SVs.

A. The proportion of different repeat types in INS and DEL sequences with lengths ranging from 160 to 170 bp.

B. The proportion of different repeat types in INS and DEL sequences with lengths ranging from 310 to 320 bp.

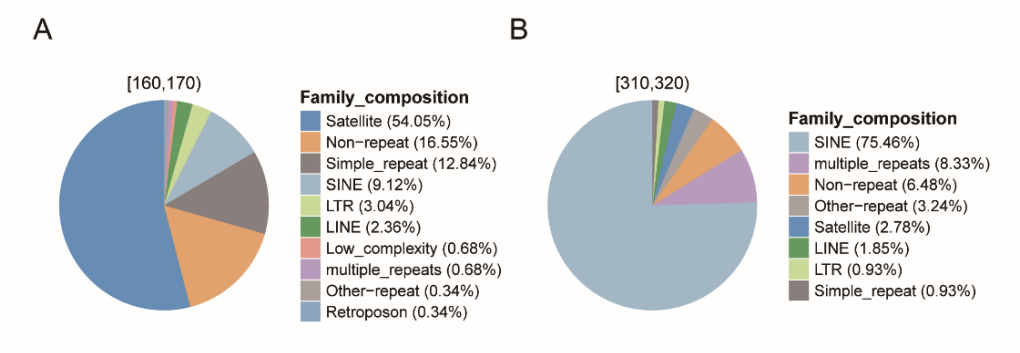


Figure S3. VAF of somatic SVs and SNVs.

A. The distribution of VAF for somatic SNVs in both tumor and adjacent nontumor tissues across patients.

B. The distribution of VAF for somatic SVs in both tumor and adjacent nontumor tissues across patients.

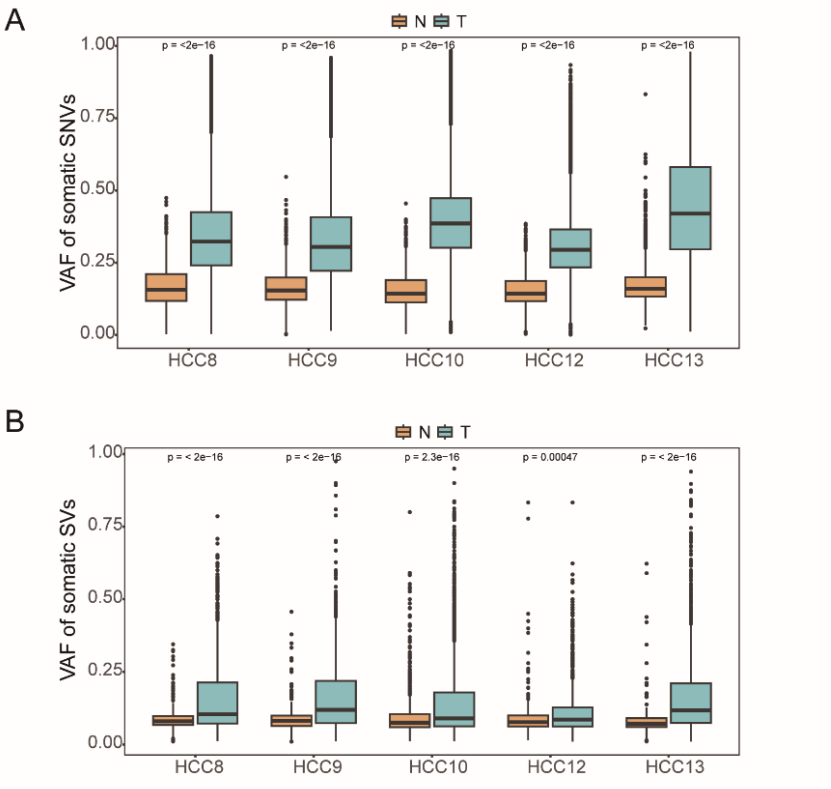
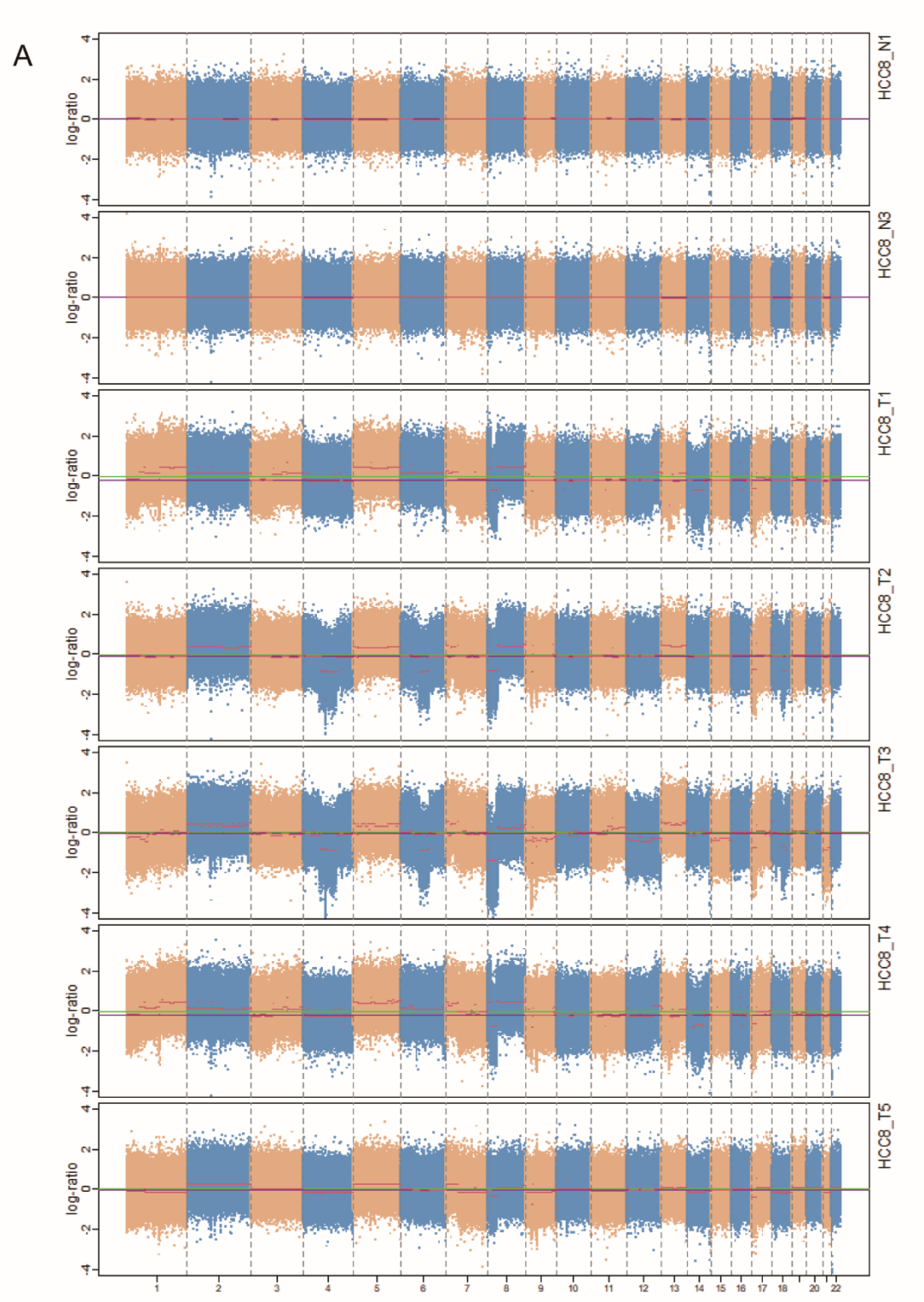
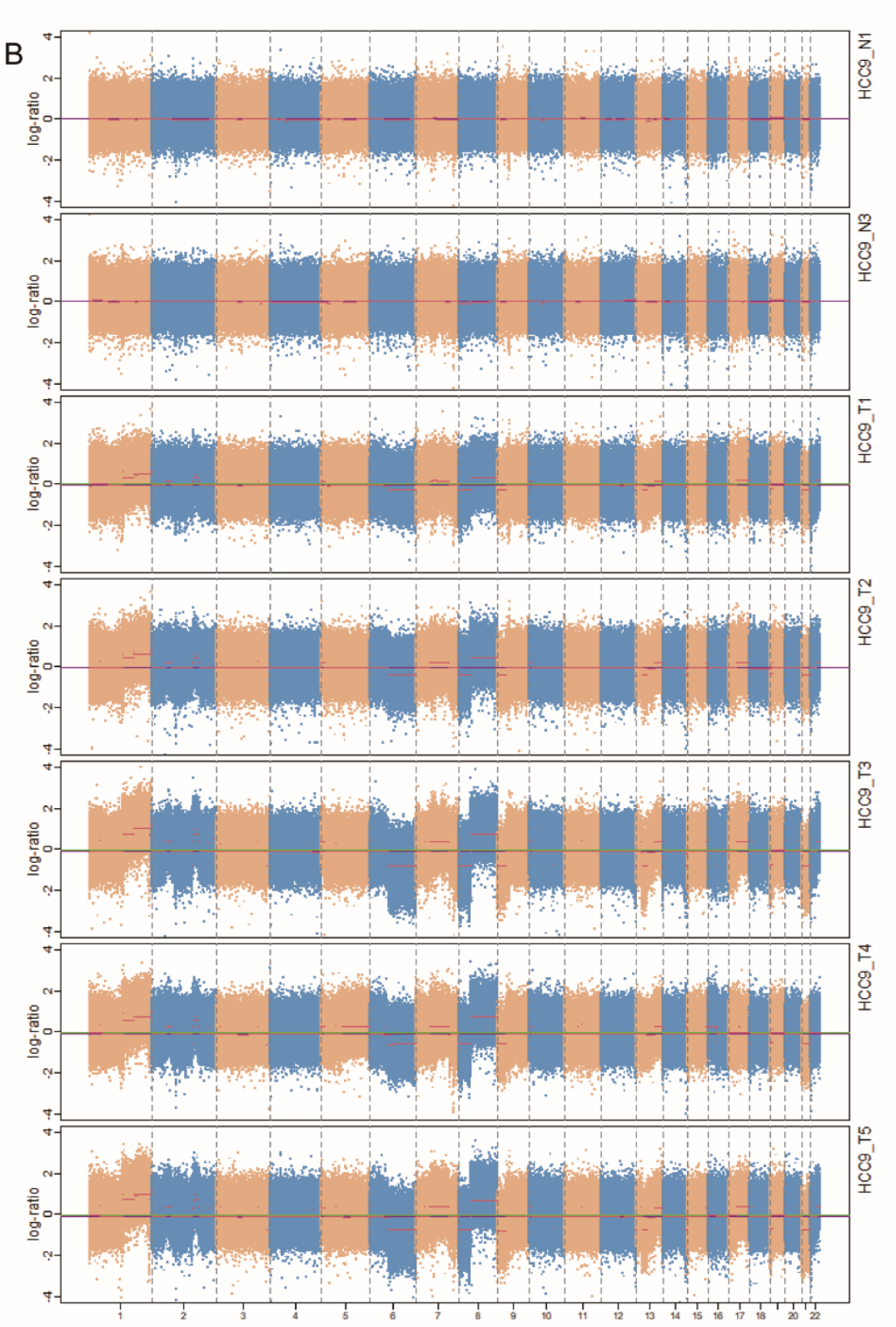


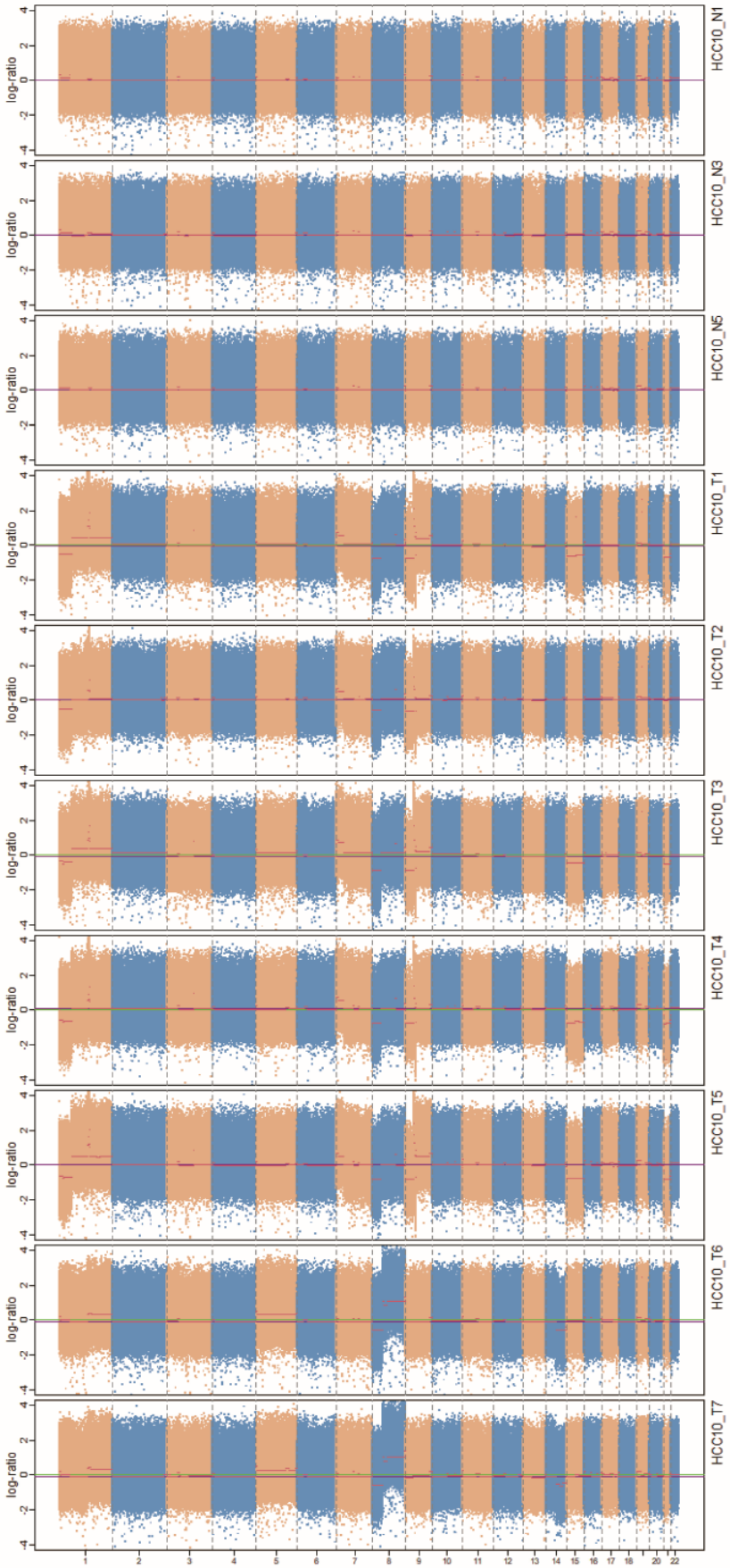
Figure S4. Distribution of somatic CNVs across different sampling sites in each patient.

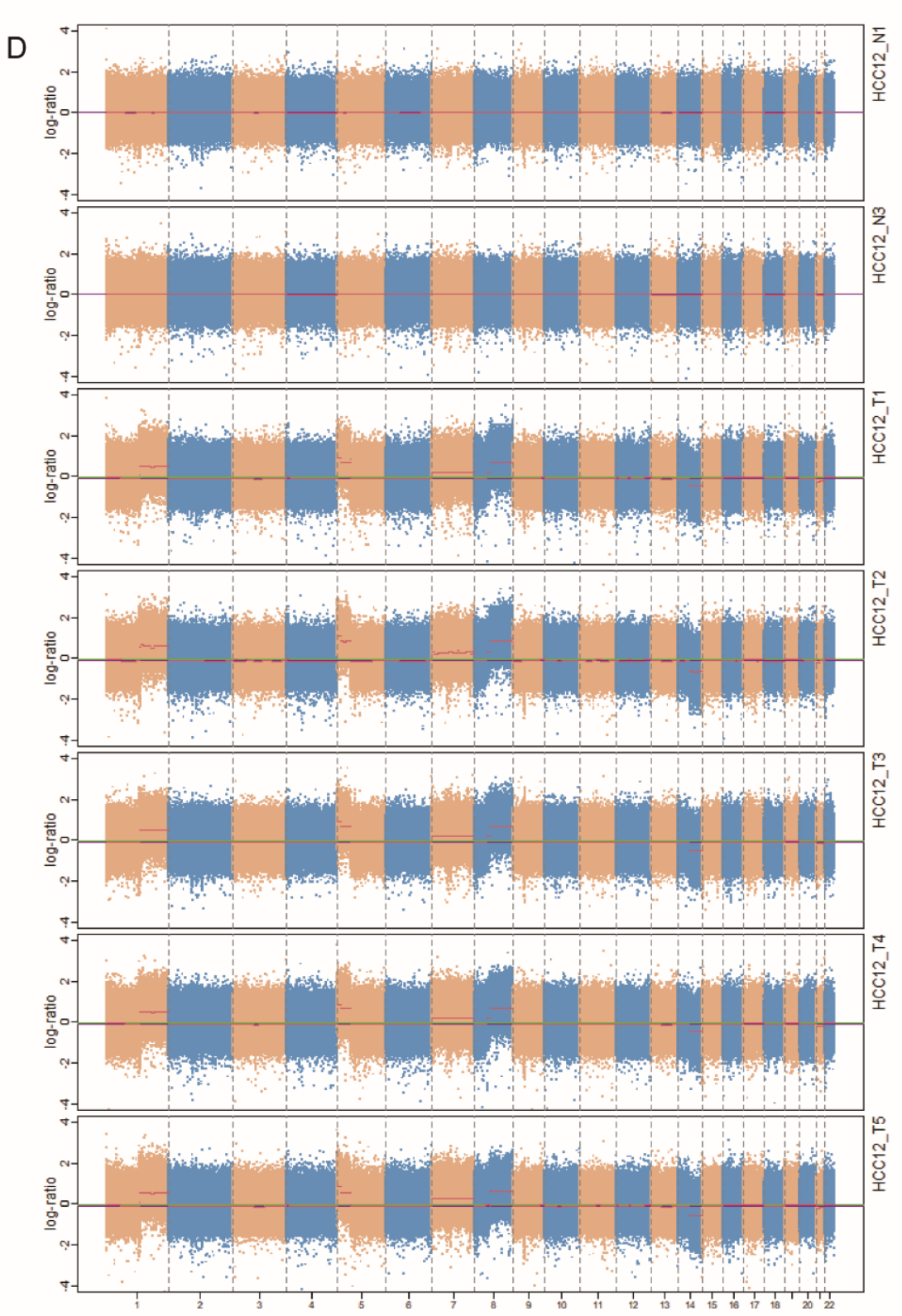
The total copy number log-ratio information for each segment is displayed. The red line represents the median of the segment.




B
log-ratio
HCC9_N1
HCC9_N3
HCC9_T1
HCC9_T2
HCC9_T3
HCC9_T4
HCC9_T5
4
2
0
-2
-4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
20
22

C


log-ratio
HCC10_N1
HCC10_N3
HCC10_N5
HCC10_T1
HCC10_T2
HCC10_T3
HCC10_T4
HCC10_T5
HCC10_T6
HCC10_T7
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
20
22


D
log-ratio
HCC12_N1
HCC12_N3
HCC12_T1
HCC12_T2
HCC12_T3
HCC12_T4
HCC12_T5
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
20
22

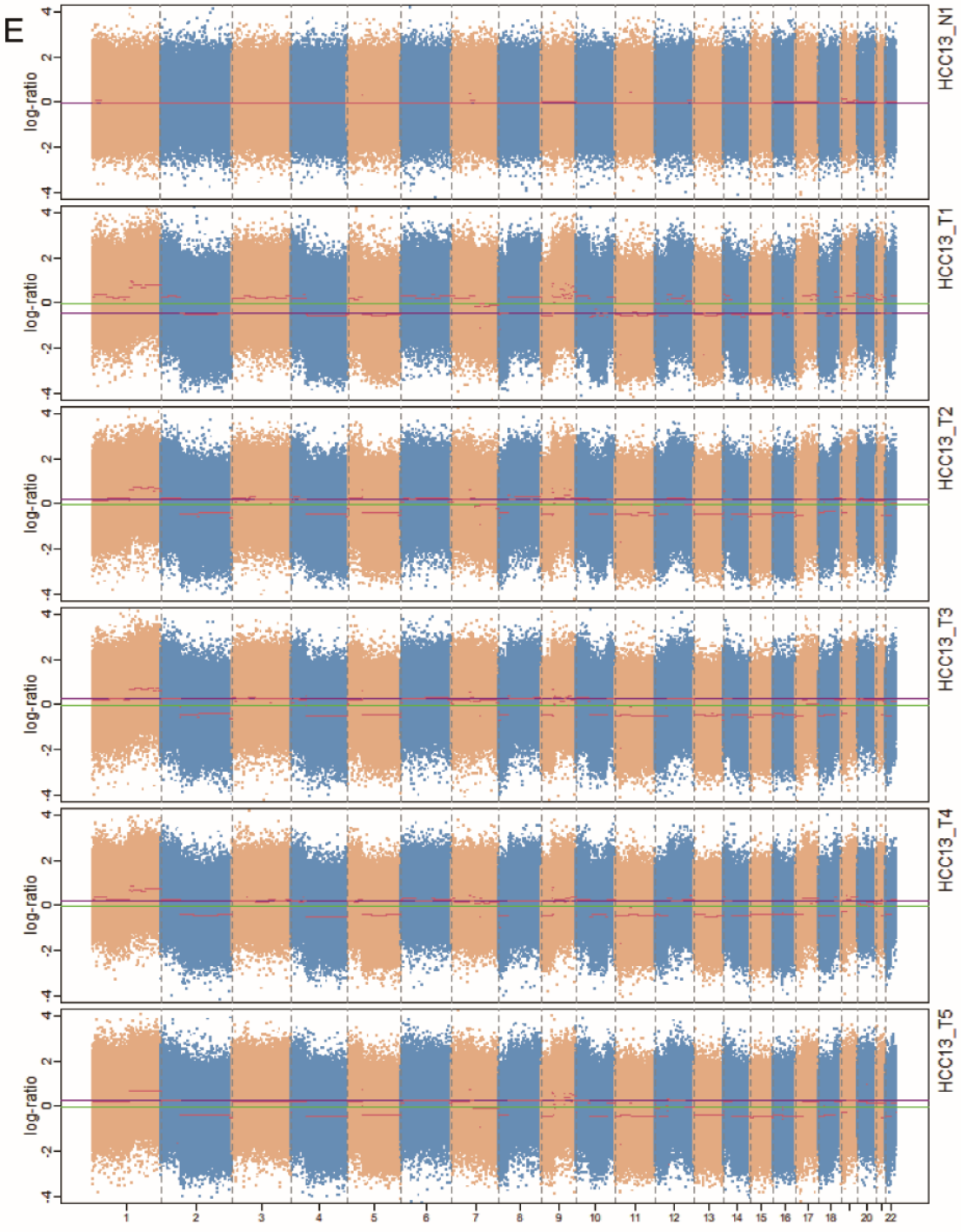
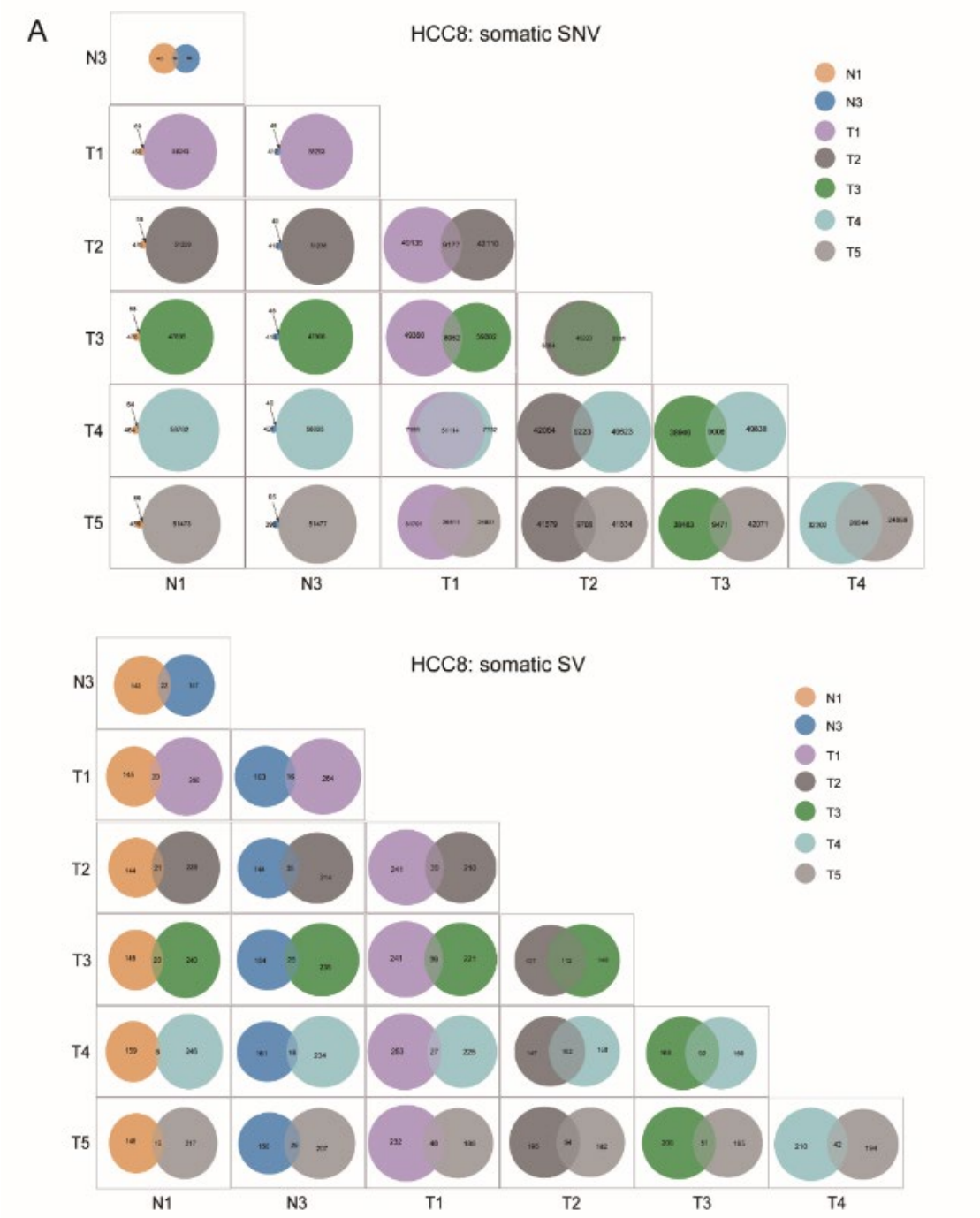
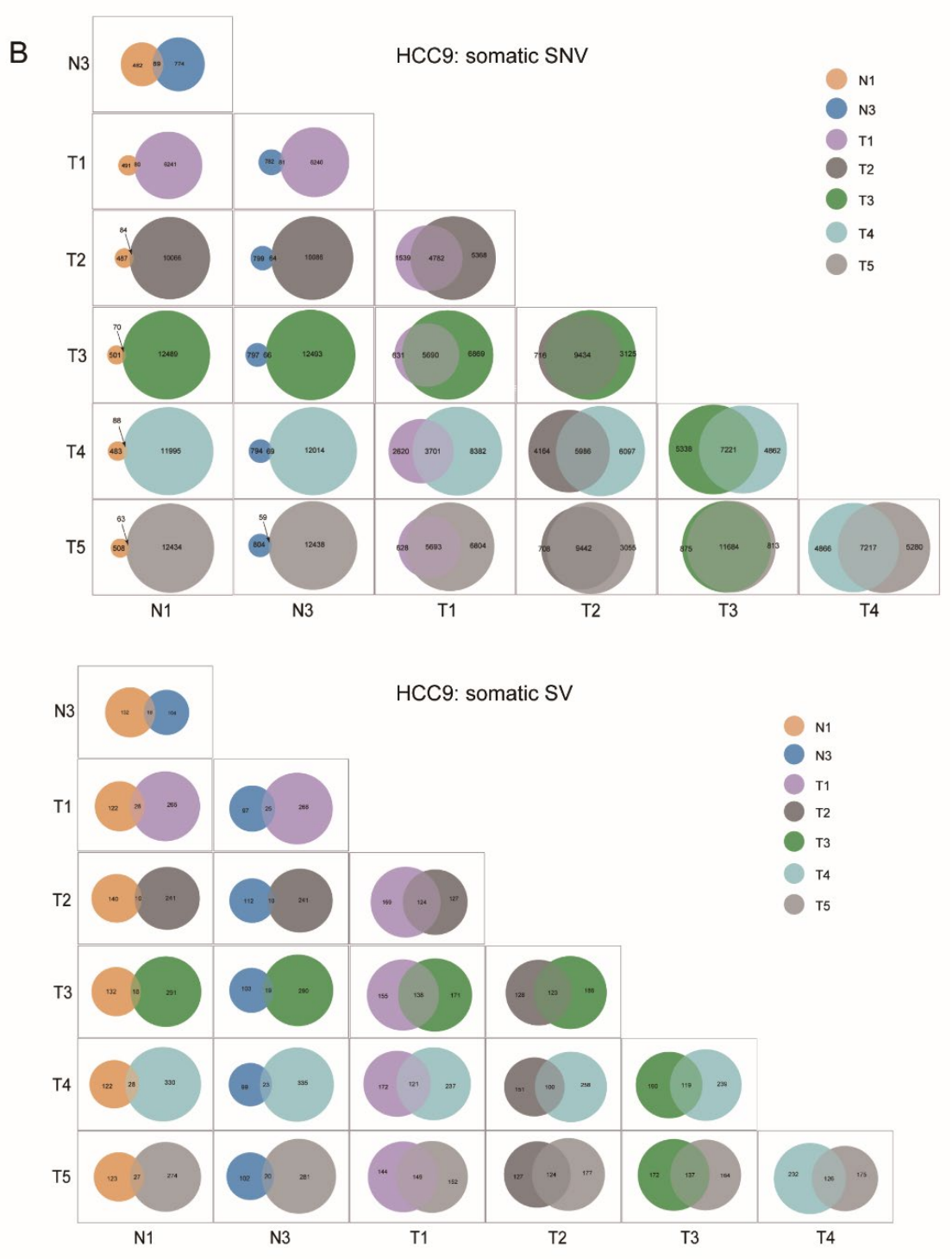

E
log-ratio
HCC13_N1
HCC13_T1
HCC13_T2
HCC13_T3
HCC13_T4
HCC13_T5
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
20
22

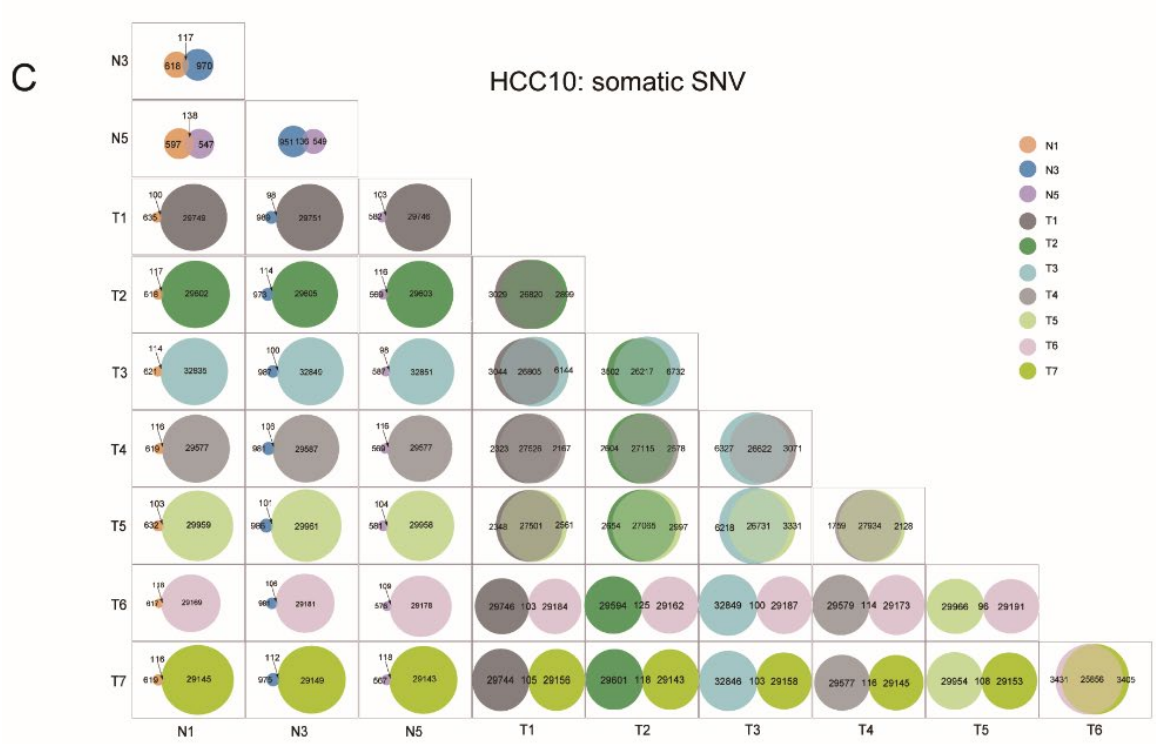
Figure S5. The pairwise comparisons of somatic SNVs and SVs across different sampling sites within each patient.

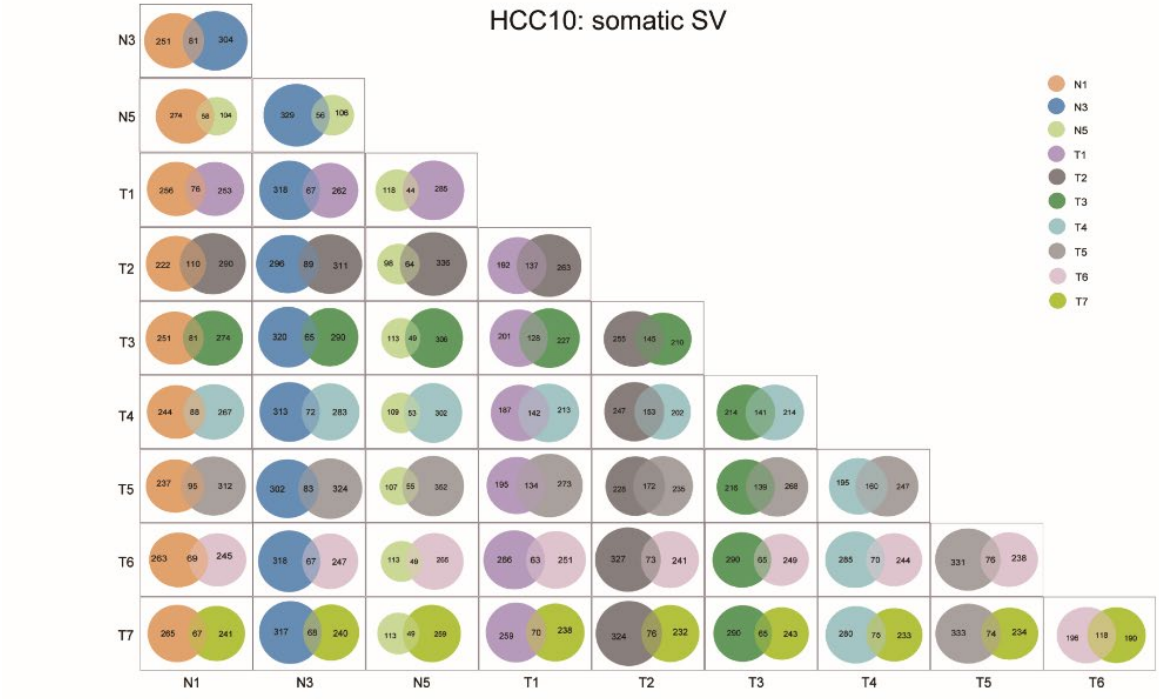
Venn diagram illustrates the pairwise comparison of somatic SNVs/SVs between different sampling sites.



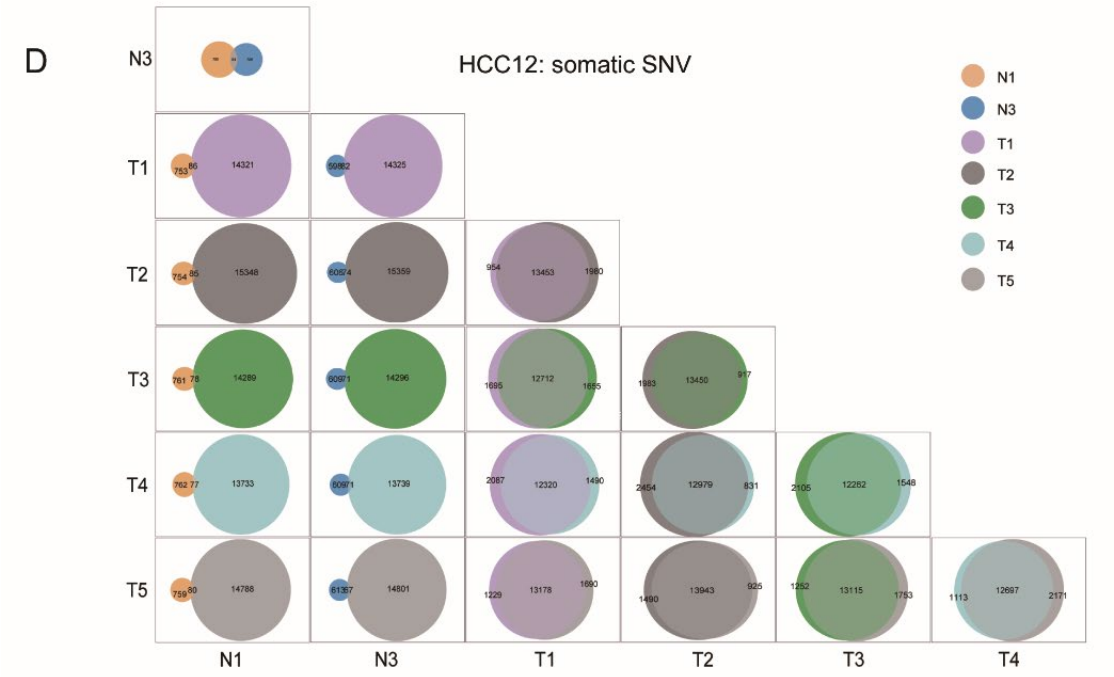

B
HCC9: somatic SNV
N1
N3
T1
T2
T3
T4
T5
HCC9: somatic SV

C


HCC10: somatic SNV
N1
N3
N5
T1
T2
T3
T4
T5
T6
T7


HCC10: somatic SV
N1
N3
N5
T1
T2
T3
T4
T5
T6
T7

D


HCC12: somatic SNV
N1
N3
T1
T2
T3
T4
T5
N3
T1
T2
T3
T4
T5
753 86 14321
59982 14325
754 85 15348
60974 15359
954 13453 1980
761 78 14289
60971 14296
1695 12712 1655
1983 13450 917
762 77 13733
60971 13739
2087 12320 1490
2454 12979 831
2105 12262 1548
759 80 14788
61357 14801
1229 13178 1690
1490 13943 925
1252 13115 1753
1113 12697 2171

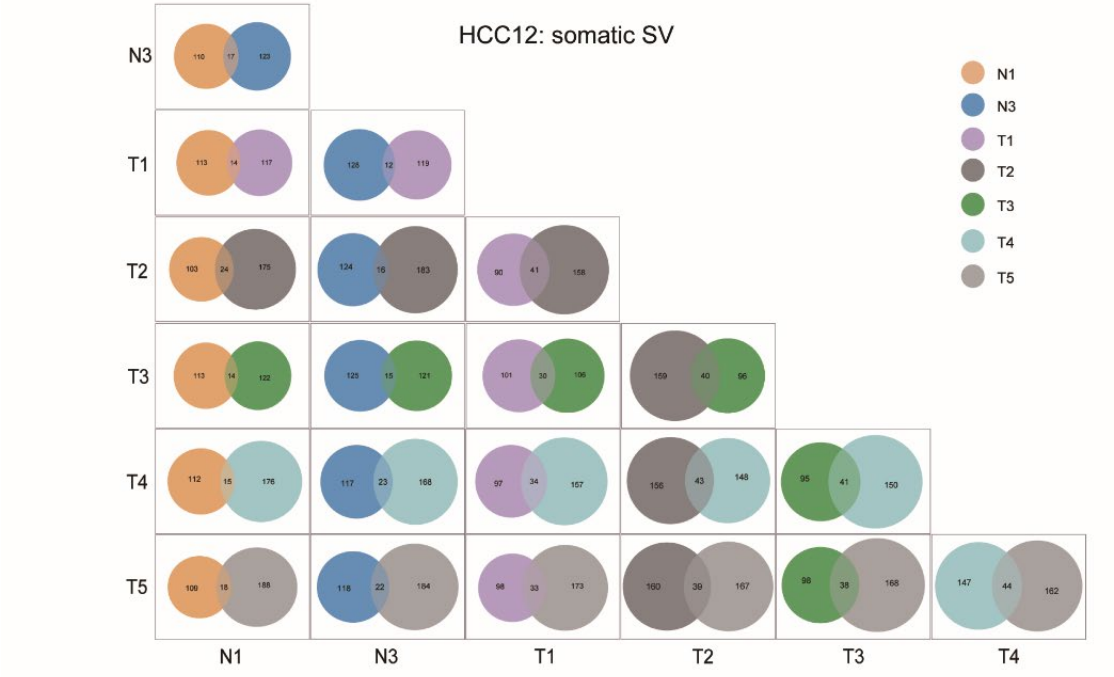

HCC12: somatic SV
N1
N3
T1
T2
T3
T4
T5
N3
T1
T2
T3
T4
T5
110 17 123
113 14 117
126 12 119
103 24 175
124 16 183
90 41 158
113 14 122
125 15 121
101 30 106
159 40 96
112 15 176
117 23 168
97 34 157
156 43 148
95 41 150
109 18 188
118 22 164
98 33 173
160 39 167
98 38 168
147 44 162

Figure S6. The proportion of shared somatic SNVs/SVs between tumor and adjacent nontumor tissues.

Boxplots depicting the proportion of shared mutations between adjacent nontumor and tumor tissues for somatic SVs and SNVs. Proportions were calculated through pairwise comparisons between adjacent nontumor and tumor samples for each patient. Statistical significance was determined by a paired *t*-test (***P < 0.001).

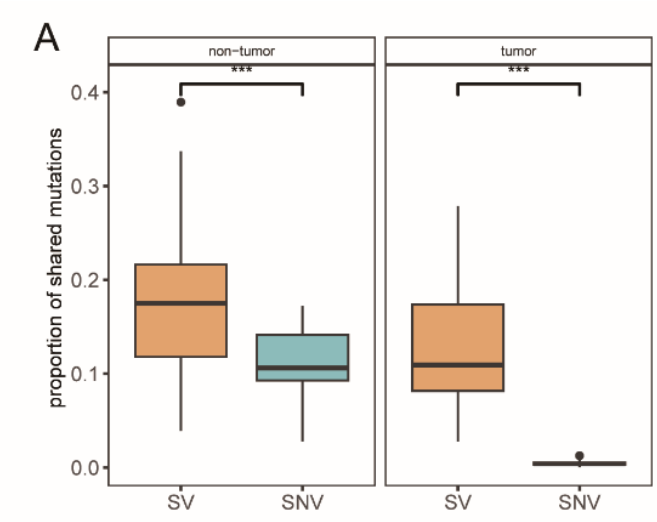


Figure S7. The phylogenetic tree constructed based on somatic SNVs/SVs.

A. Tumor phylogenetic trees for each patient were constructed using the somatic SV mutation matrix. The gray lines indicate branches shared between tumor and adjacent nontumor tissues.

B. Tumor phylogenetic trees for each patient were constructed using the somatic SNV mutation matrix. The gray lines indicate branches shared between tumor and adjacent nontumor tissues.

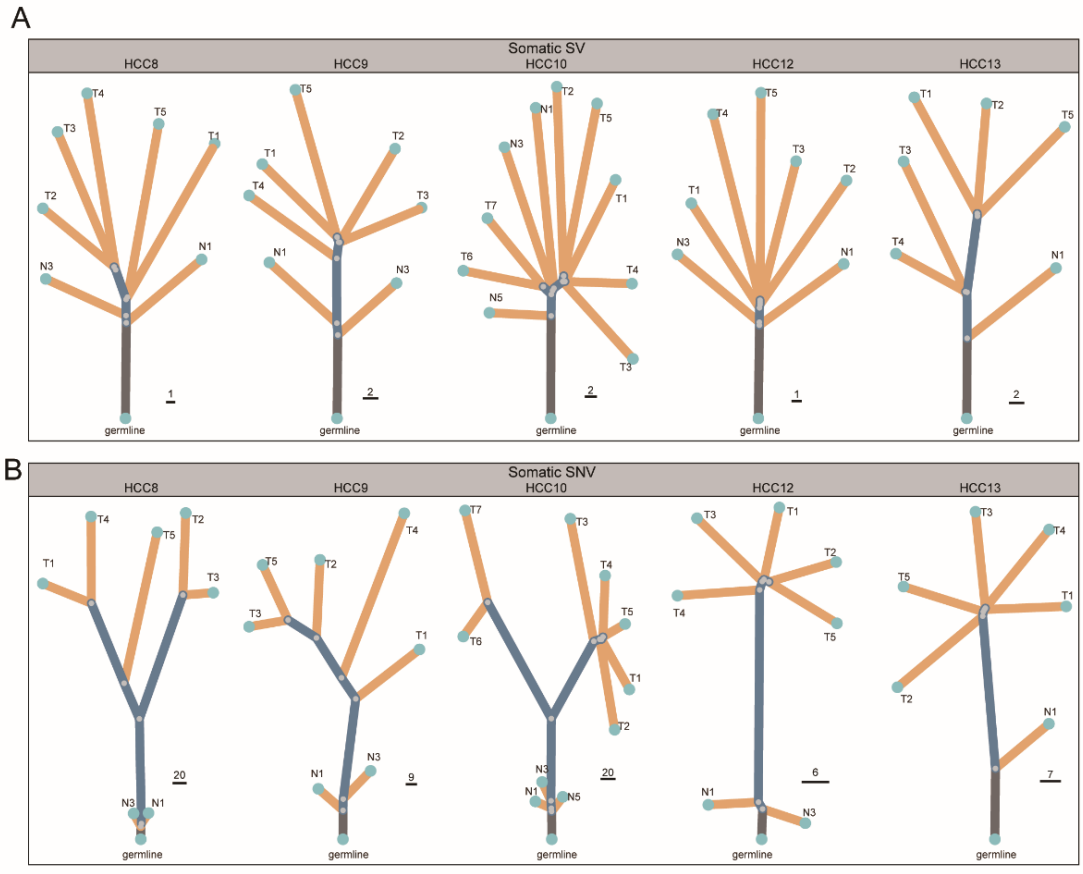


Figure S8. Somatic DUPs and INVs in tumor samples.

A. The Venn diagram illustrates the pairwise comparison of somatic DUPs across different samples of HCC9.

B. The Venn diagram illustrates the pairwise comparison of somatic INVs across different samples of HCC13.

C. The distribution of somatic INVs across the genome in HCC13.

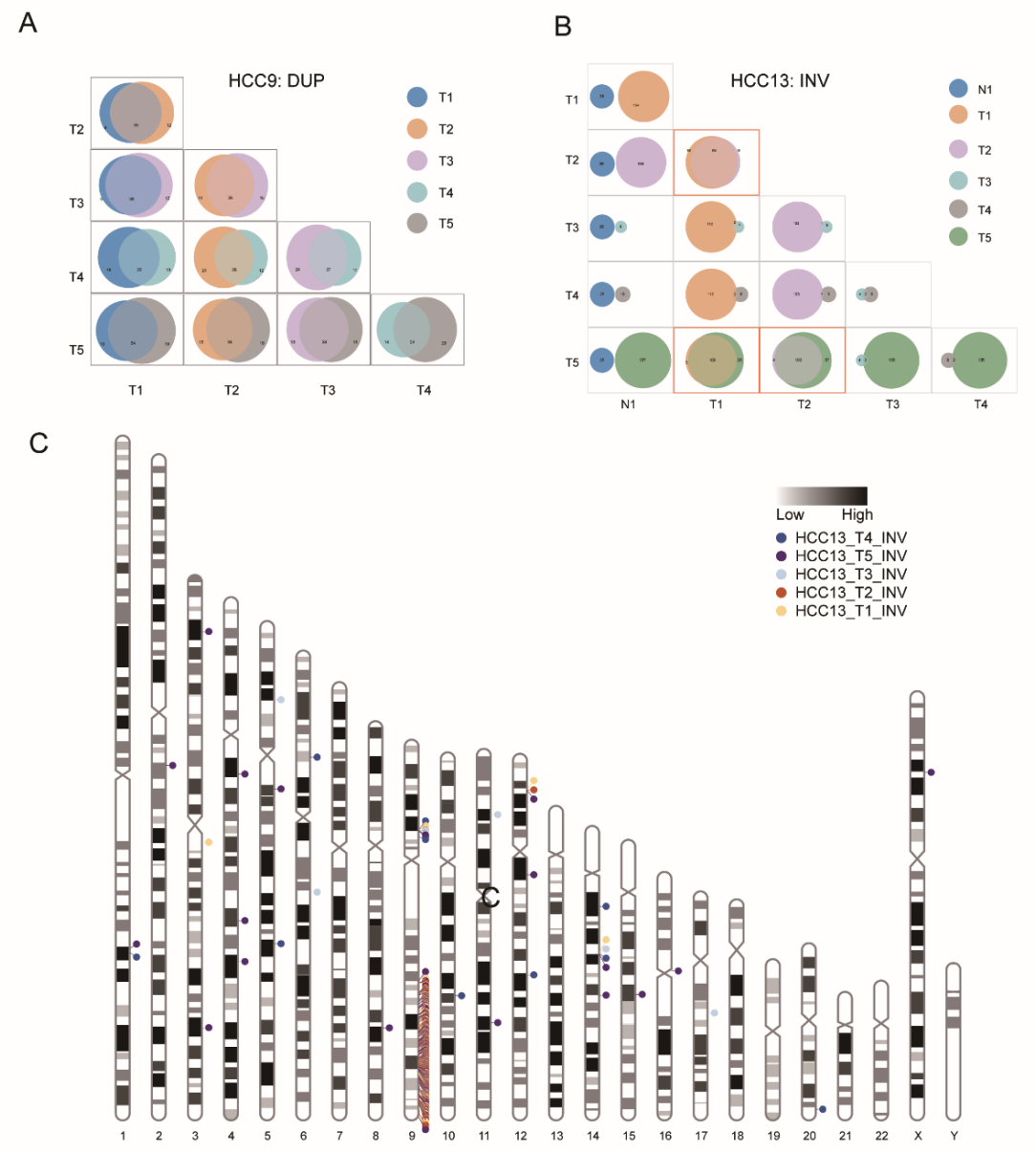


Figure S9. The repeat types of shared somatic SVs between tumor and adjacent nontumor tissues.

A. The pie chart illustrates the proportion of different repeat sequence types in shared somatic SVs between tumor and adjacent nontumor tissues, with "Non-repeat" referring to sequences that are not annotated as repetitive by RepeatMasker.

B. The pie chart illustrates the proportion of different repeat sequence types in somatic SVs specific to tumor or adjacent nontumor tissues, with "Non-repeat" referring to sequences that are not annotated as repetitive by RepeatMasker.

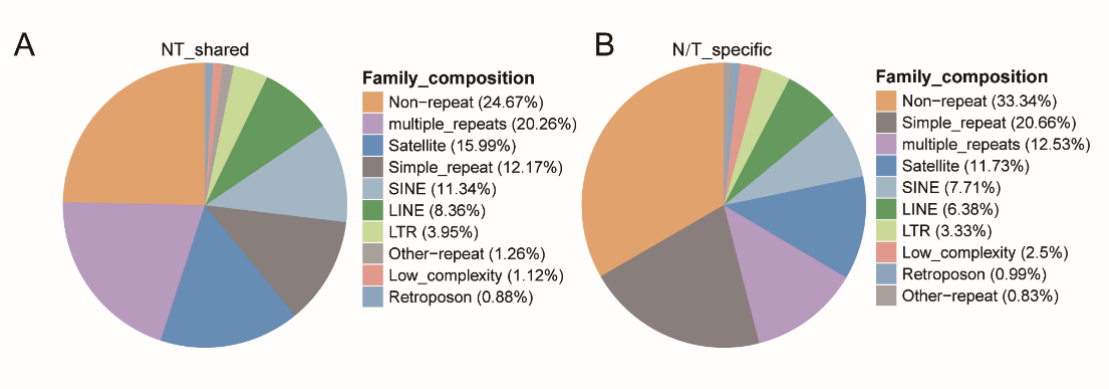


Figure S10. The somatic deletions on *EVA1C*.

A. The IGV snapshots illustrating the alignment of long-read sequencing data (top panel) and short-read sequencing data (middle panel) on the deletion region within the *EVA1C* gene. The bottom panel shows the genomic repeat annotation of RepeatMasker, with the DEL region highlighted in orange.

B. The DEL sequence within the *EVA1C* gene is annotated, with different colors representing various repeat types, and the underlined regions denote the deleted segments.

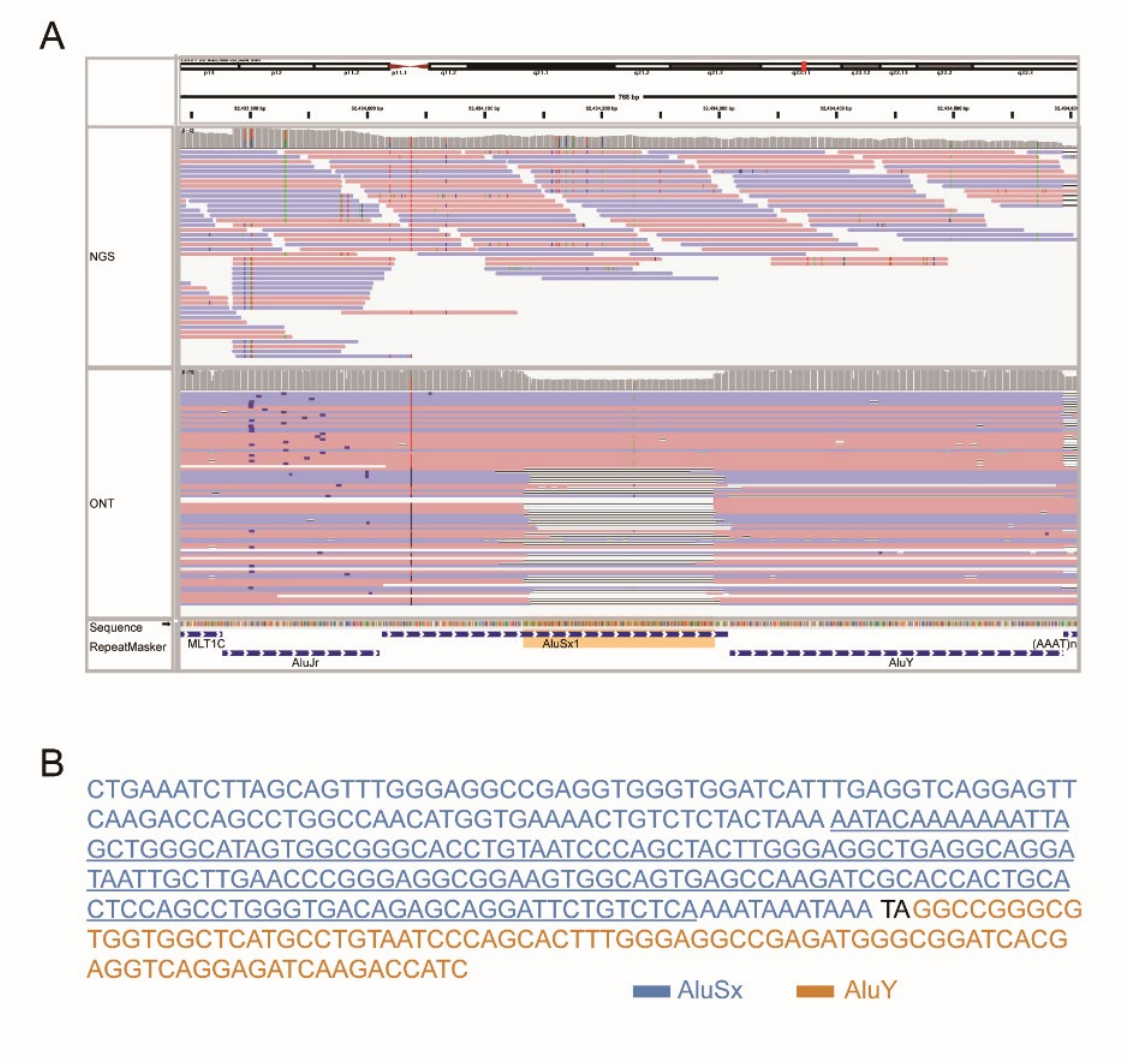


Figure S11. The impact of shared somatic SVs between tumor and adjacent nontumor tissues on genes.

A. The IGV image illustrating the alignment of long RNA reads on the *EVA1C* gene. The orange line indicates the location of deletions. Deletion sites are marked by orange lines. The upper panel corresponds to sample without deletion, while the middle panel represents sample with deletion.

B. The IGV plot illustrates the transcript loss of *GSTM1* and *GSTM2* in samples exhibiting DEL. The red regions indicate the areas of deletion.

C. The relative expression levels of *EVA1C* determined using RT–PCR in samples with and without the deletion.

D. Kaplan-Meier analysis of the overall survival rates of the high- and low-expression groups of *EVA1C* (P=0.0056, log-rank test).

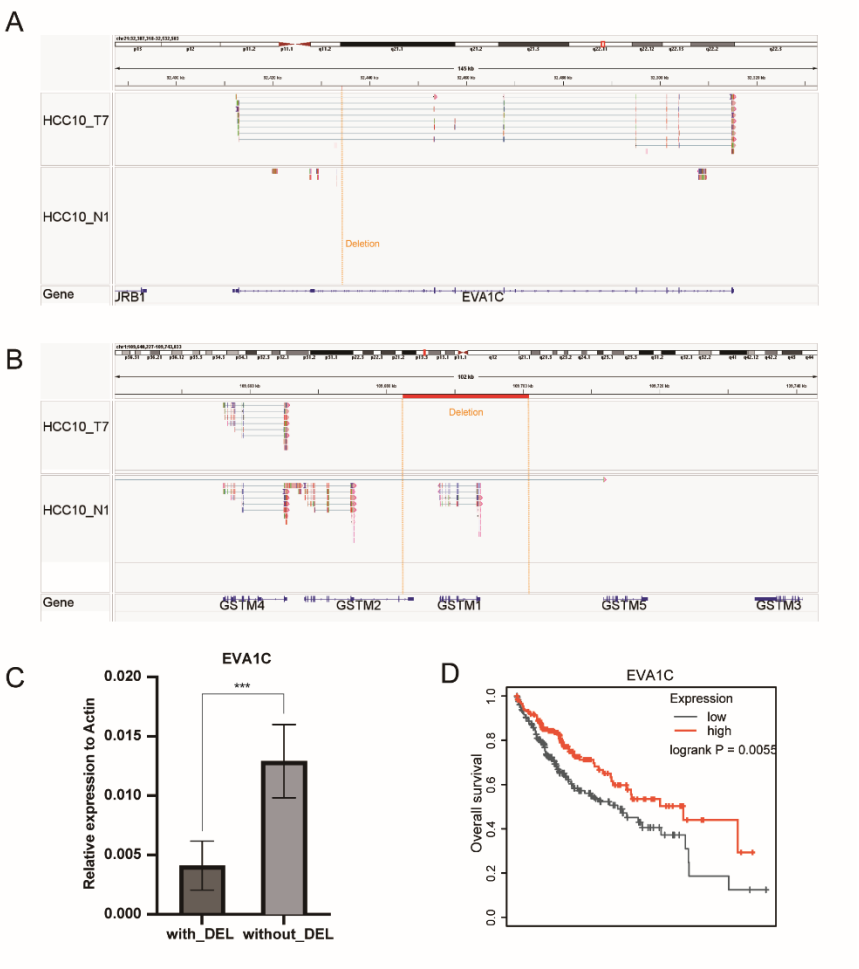


Figure S12. Inferring SV types based on HBV DNA integration patterns.

A. An illustrative IGV plot showcasing an HBV integration event. Long reads may align as chimeric reads, mapping to both human and HBV sequences, or they may solely align to the human genome, with the integrated HBV DNA sequence appearing as insertions or being clipped.

B. The length distribution of HBV DNA sequences integrated into the human genome.

C. The schematic diagram illustrating the integration of HBV DNA into the human genome in a simple insertion form, without inducing chromosomal rearrangements.

D. Illustration of HBV DNA integration inducing formation of TRAs in the human genome. The feature is the alignment of human sequences flanking the HBV sequence to different chromosomes.

E. Illustration of HBV DNA integration inducing formation of INVs in the human genome. The feature is the alignment of human sequences flanking the HBV sequence in the opposite direction with consistent order.

F. Illustration of HBV DNA integration inducing formation of DELs in the human genome. The feature is the alignment of human sequences flanking the HBV sequence in the same direction with consistent order, along with the presence of a gap.

G. Illustration of HBV DNA integration inducing formation of DUPs in the human genome. The feature is the alignment of human sequences flanking the HBV sequence in the same direction with consistent order, along with the presence of an overlap.

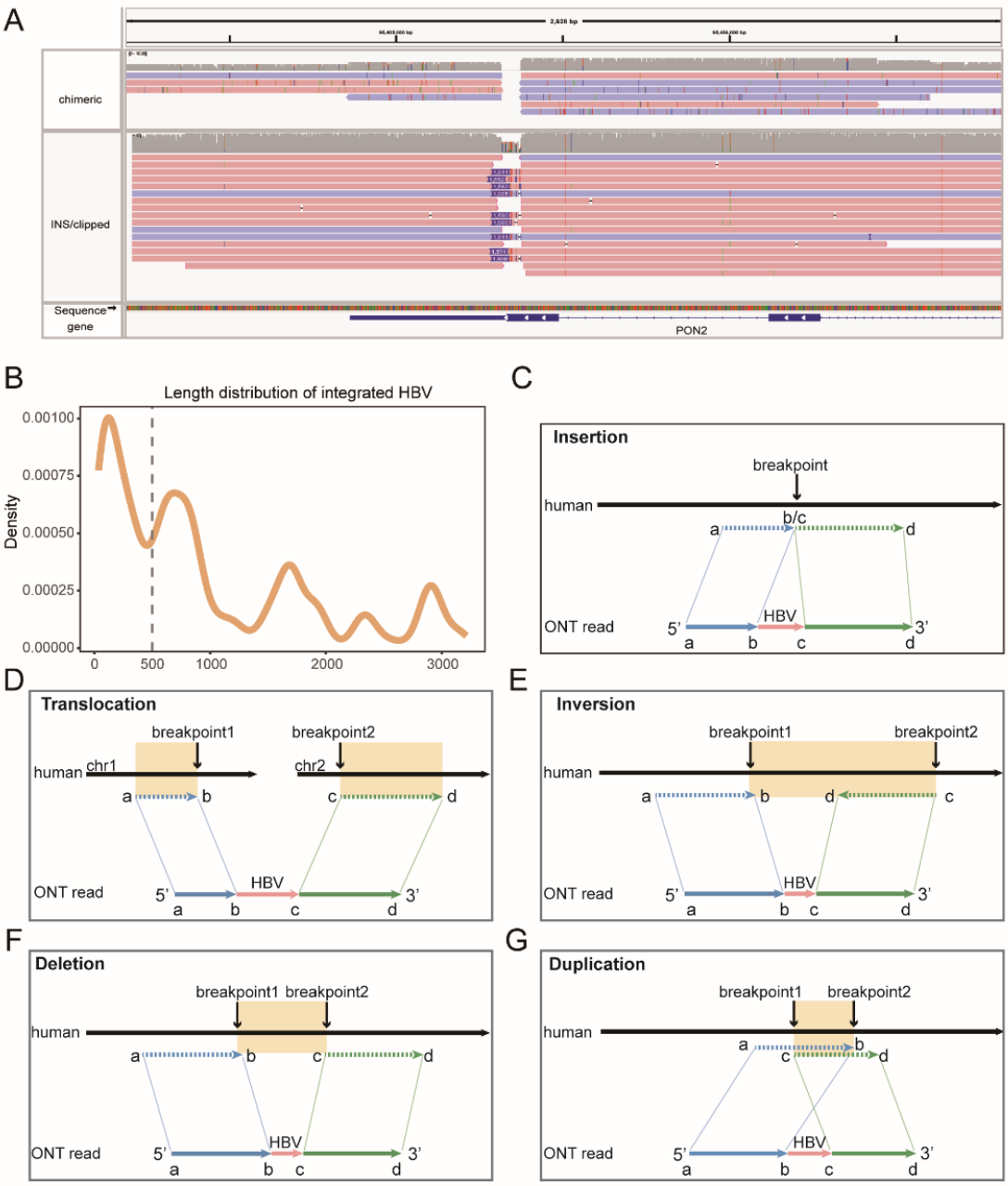

A
chimeric
INS/clipped
Sequence
gene
PON2
B
Length distribution of integrated HBV
Density
0.00100
0.00075
0.00050
0.00025
0.00000
0
500
1000
2000
3000
C
Insertion
breakpoint
human
b/c
a
d
ONT read
5'
HBV
3'
a
b
c
d
D
Translocation
breakpoint1
breakpoint2
human
chr1
chr2
a
b
c
d
ONT read
5'
HBV
3'
E
Inversion
breakpoint1
breakpoint2
human
a
b
d
c
ONT read
5'
HBV
3'
F
Deletion
breakpoint1
breakpoint2
human
a
b
c
d
ONT read
5'
HBV
3'
G
Duplication
breakpoint1
breakpoint2
human
a
b
c
d
ONT read
5'
HBV
3'

Figure S13. Somatic SVs were enriched at HBV integration sites.

A. The distance distribution of different types of HBV-induced SVs relative to the CNV boundaries.

B. Enrichment analysis of somatic SVs at HBV integration sites. Two-sided binomial test was used for enrichment analysis.

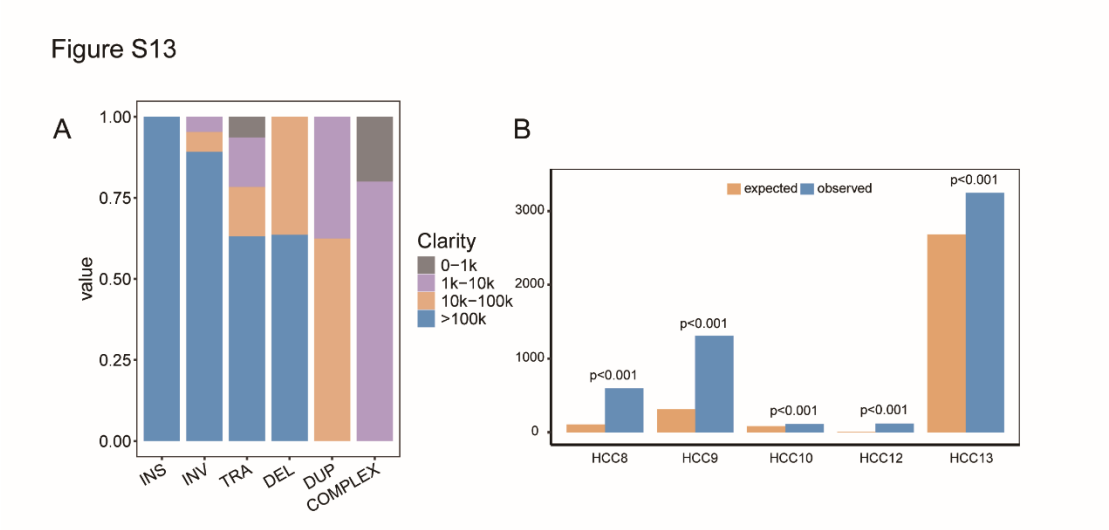


Figure S14. Breakpoint locations of HBV integration in both the human and HBV genomes.

A. Shared HBV DNA integrations were depicted, showing the correspondence of breakpoints between the human and HBV genomes. Different colors represent distinct samples.

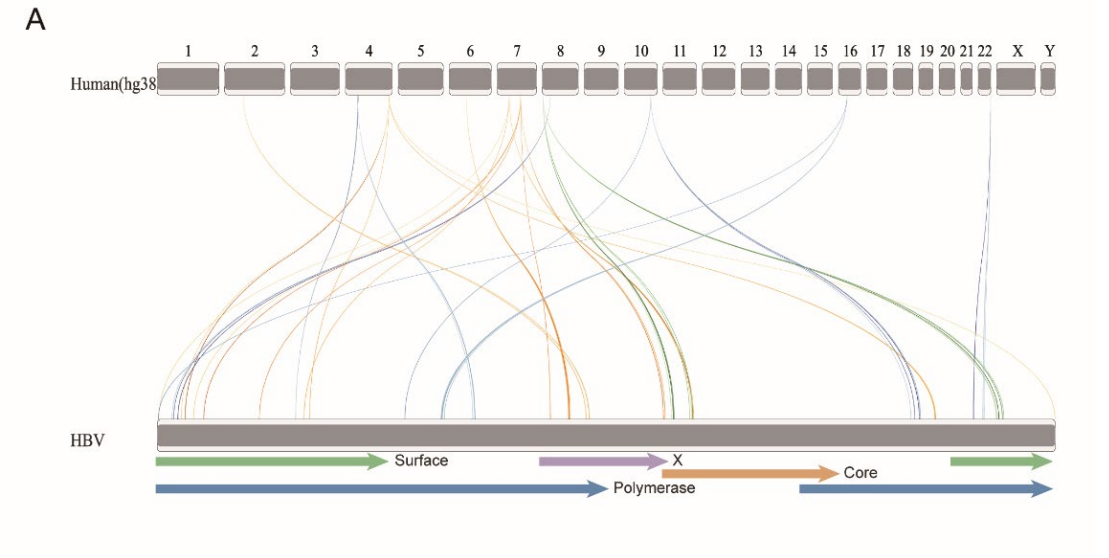


Figure S15. Shared breakpoints of HBV-induced SVs.

A. The percentage of shared breakpoints at different frequencies across all samples. "1" denotes those occurring in only one sample, while "2-5" represents the number of samples associated with shared breakpoints.

B. The number of adjacent nontumor and tumor samples at shared breakpoints of different frequencies.

C. Boxplot illustrating the distribution of Variant Allele Frequencies (VAF) for shared breakpoints and single breakpoints. Statistical significance was determined by Wilcoxon rank-sum test (***P < 0.001).

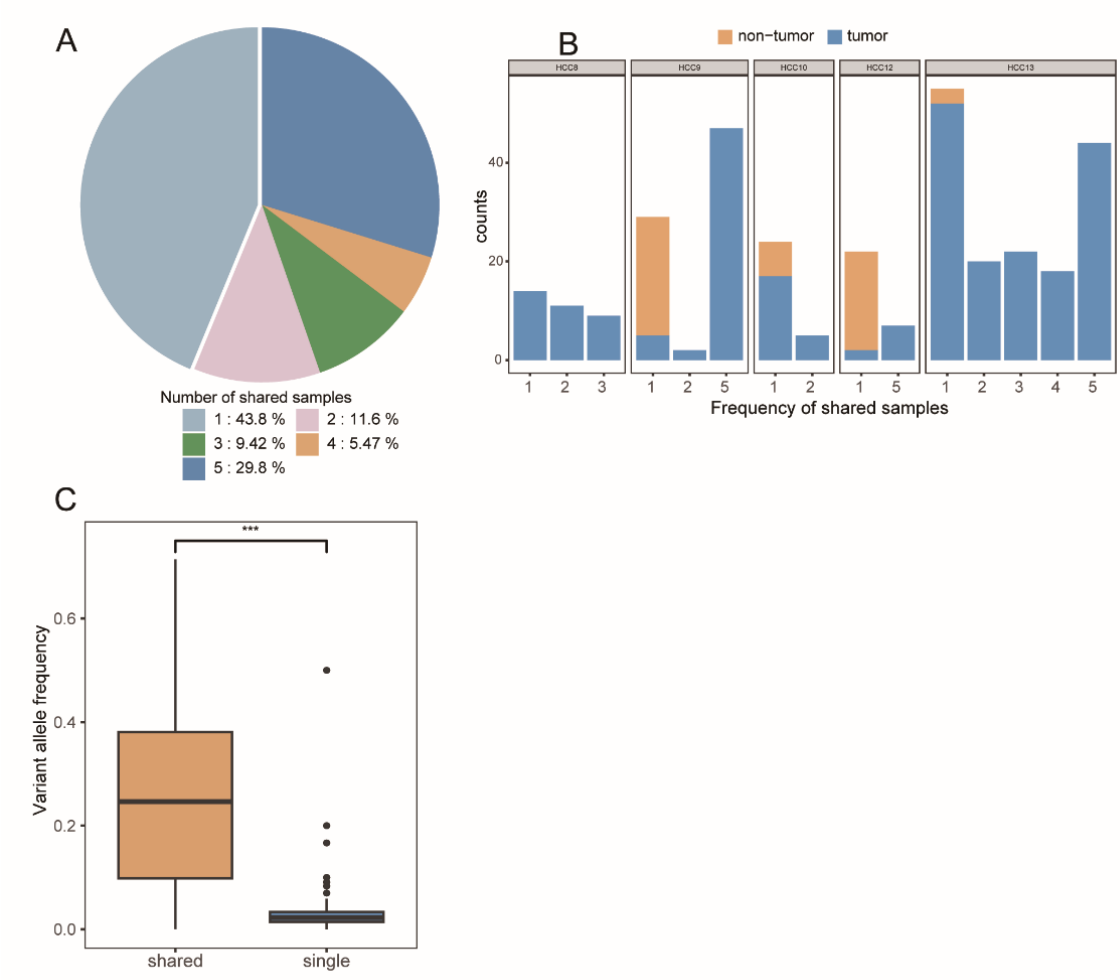


Figure S16. Genes disrupted by HBV-induced SVs.

A. KEGG pathway enrichment analysis of genes affected by HBV-induced SVs.

B. The percentage of shared and single breakpoints positioned within genes and their flanking ±5 kb regions, as well as other genomic loci.

C. The relative expression levels of *NRG1* determined using RT–PCR in paired tumor and nontumor tissues. The grouping on the x-axis represents primers spanning different exons.