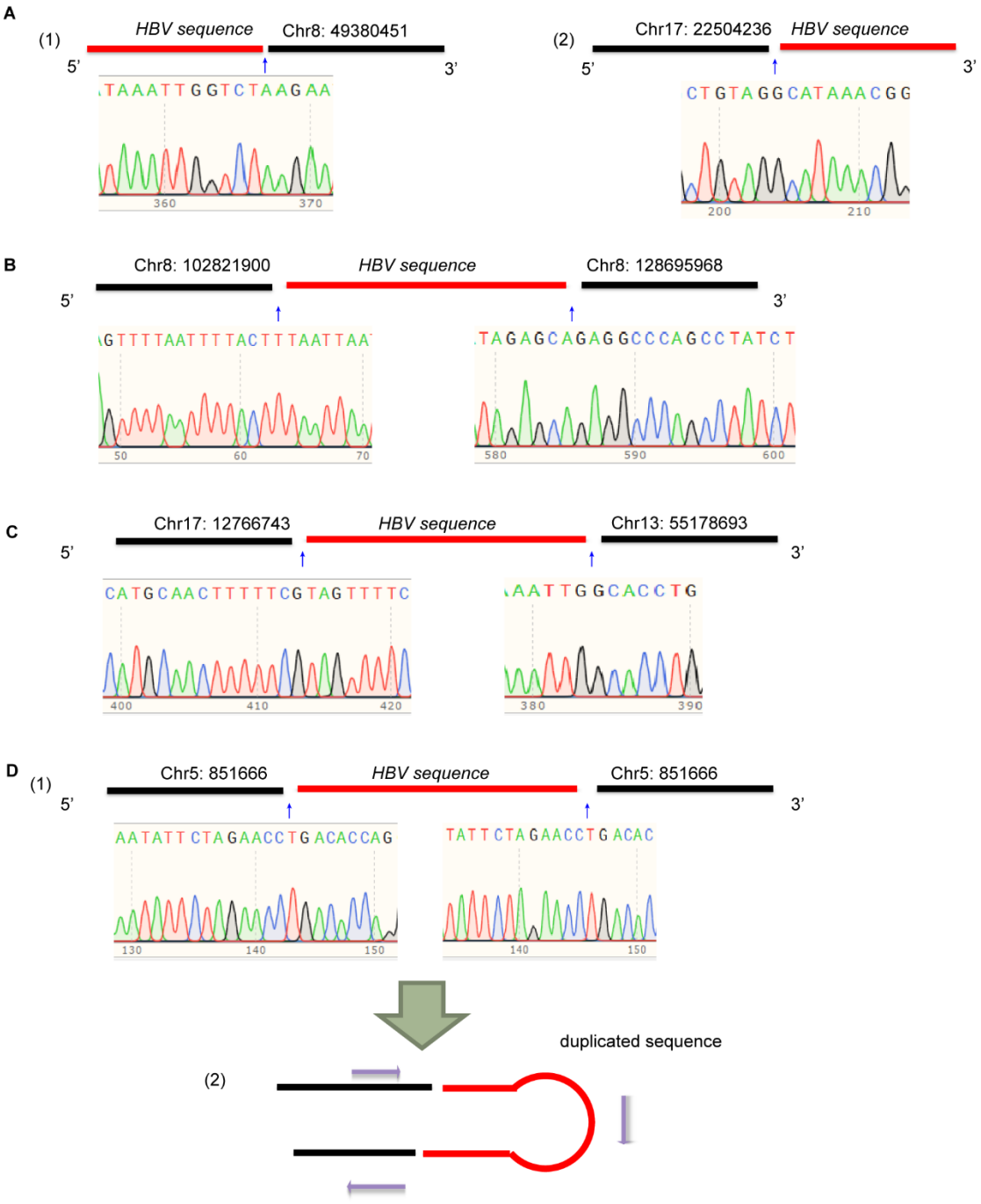
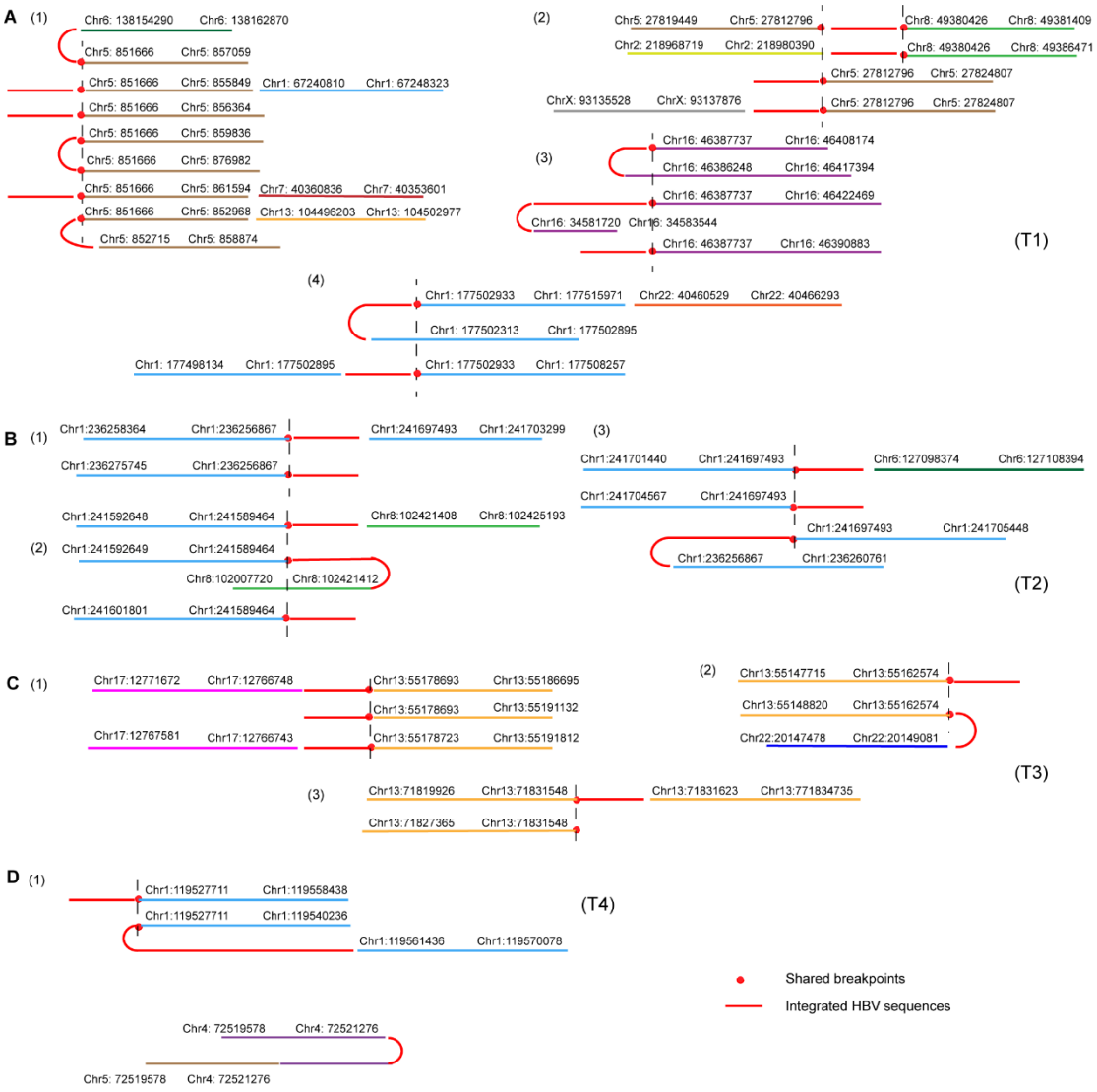
Supplemental figures

Supplemental_Fig_S1



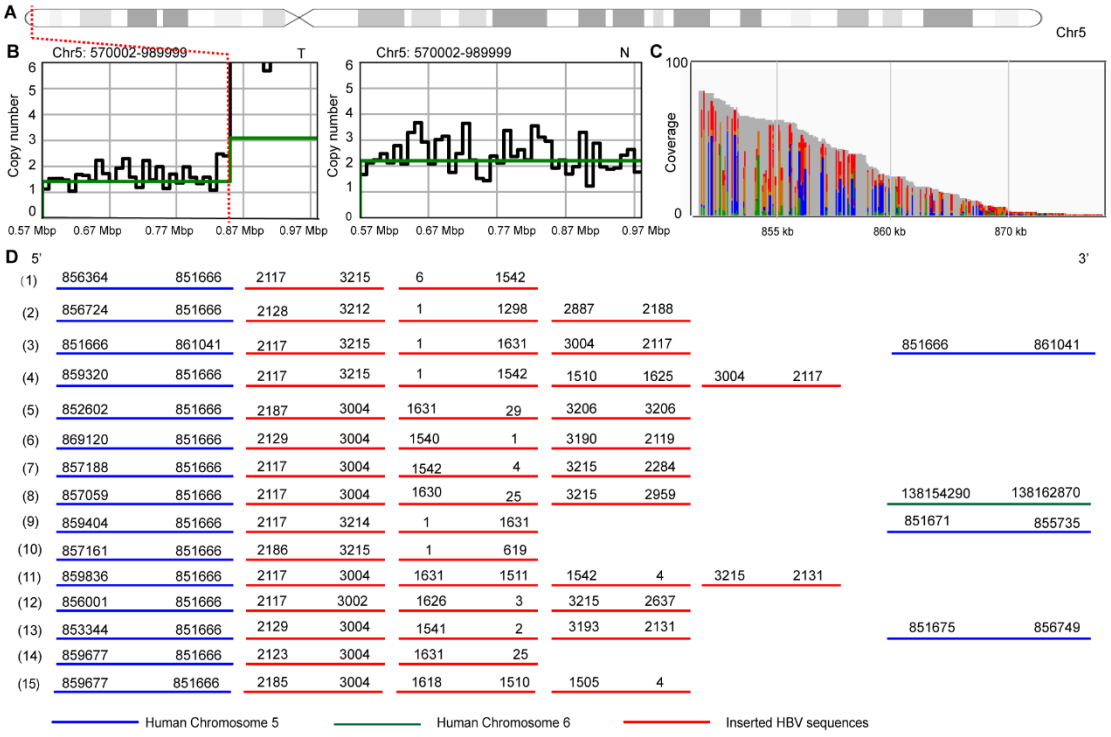
Supplemental_Fig_S1. Representative PCR validation results for the HBV–human chimeric reads. Line segments in red represent the integrated HBV sequences, and the lines segments in black represent the human genomes.

Supplemental_Fig_S2



Supplemental_Fig_S2. Representative breakpoints shared in the tumor tissues. A–D, Sample T1–T4, respectively. The red circles represent the shared breakpoints of different HBV–human chimeric reads in one sample, the "start" and "end" positions of the human parts of the chimeric reads are labeled from 5' to 3', the red lines represent integrated HBV sequences, the colored lines represent the human sequences of different chromosomes, and the numbers in parentheses represent different groups of the shared breakpoints.

Supplemental_Fig_S3



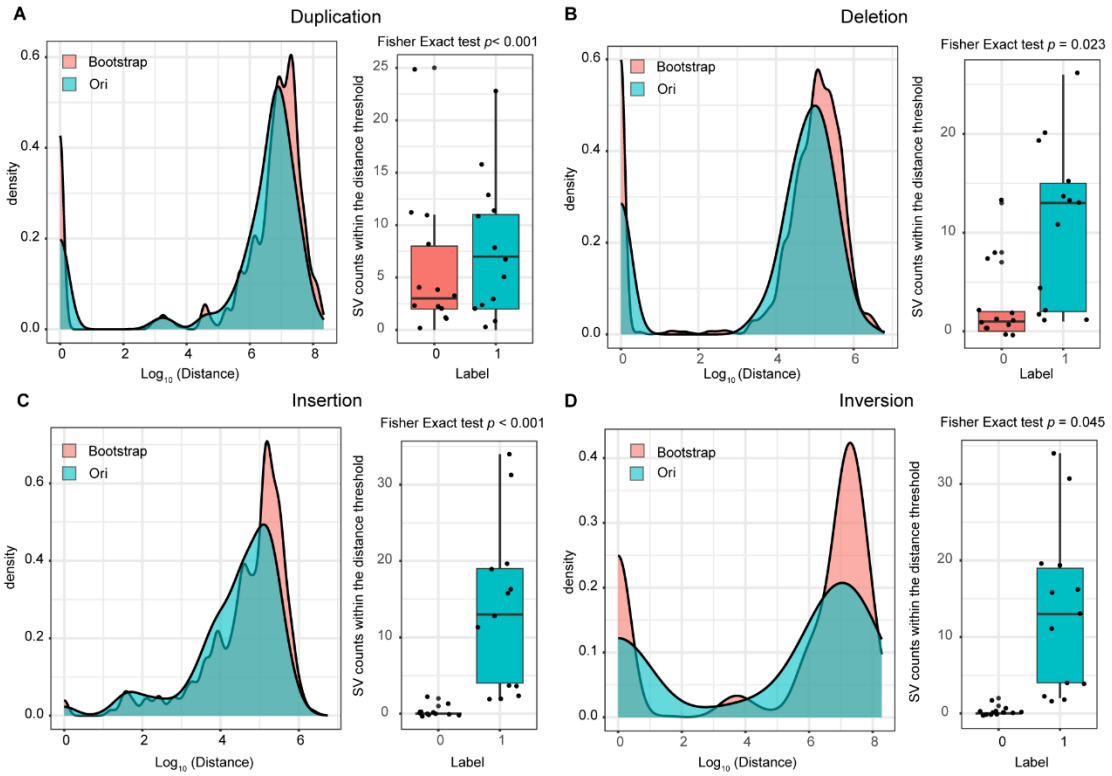
Supplemental_Fig_S3. Schematic representing various HBV inserted sequences connected to one breakpoint. (A) Schematic representation of the HBV integration site on Chr 5. (B) Copy number variations in the HBV integration region. (C) Read coverage obtained through screen capture via IGV software. (D) Representative HBV–human chimeric reads sharing the breakpoint Chr5: 851666.

Supplemental_Fig_S4

Supplemental_Fig_S4. HBV SVs. The green background represents the open reading frames (ORFs) diagram of each HBV subtype, including the HBV genes encoding the S (green), P (blue), X (red), and C (orange) proteins. The yellow and light purple backgrounds respectively represent the structural variations of HBV detected in tumor and nontumor samples, including deletion (green lines), inversion (blue lines) and

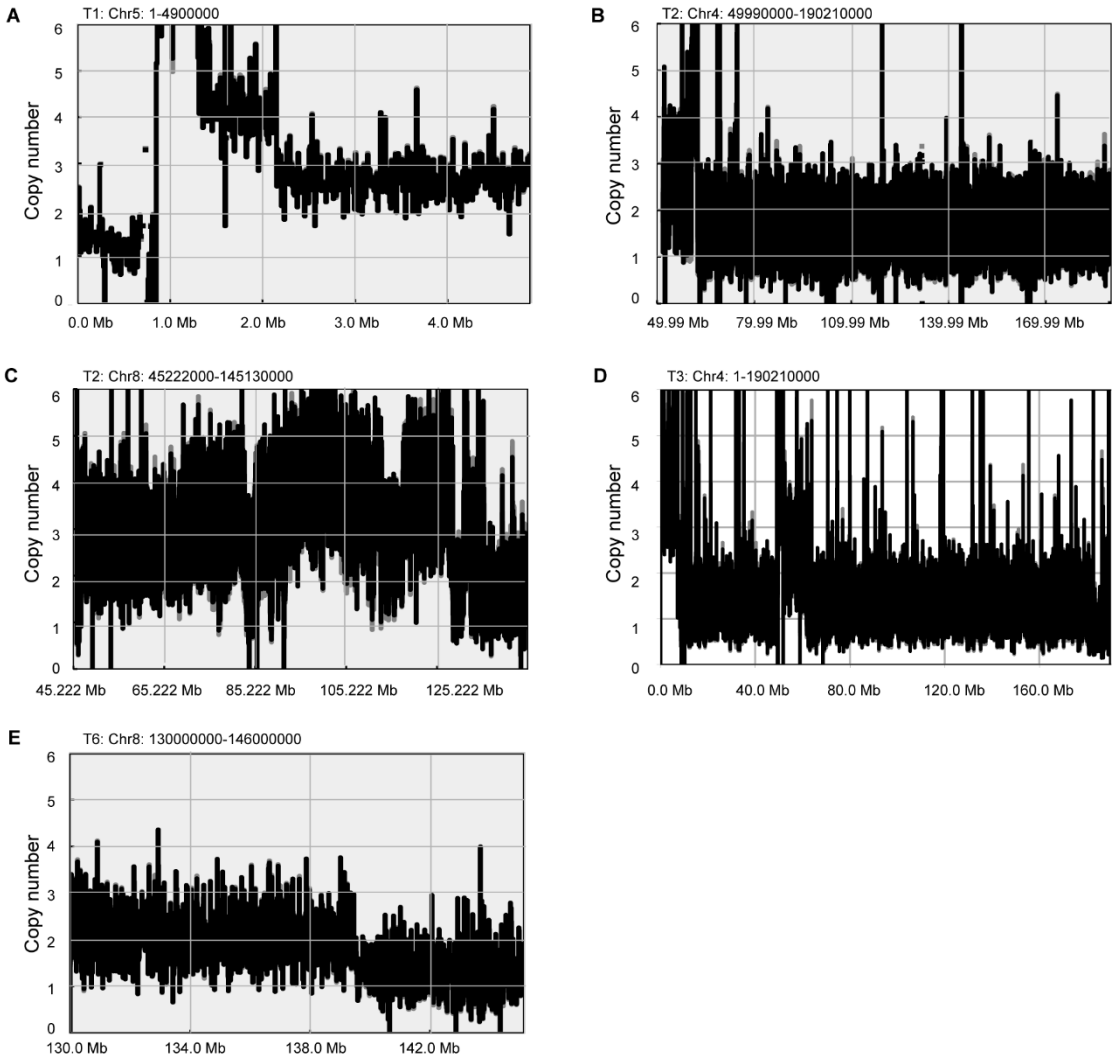
duplication (orange lines), as well as insertion (red triangle). A–D, represent subtype: AY123041.1 (A); AB602818.1 (B); AB014381.1 (C); D00330.1 (D).

Supplemental_Fig_S5



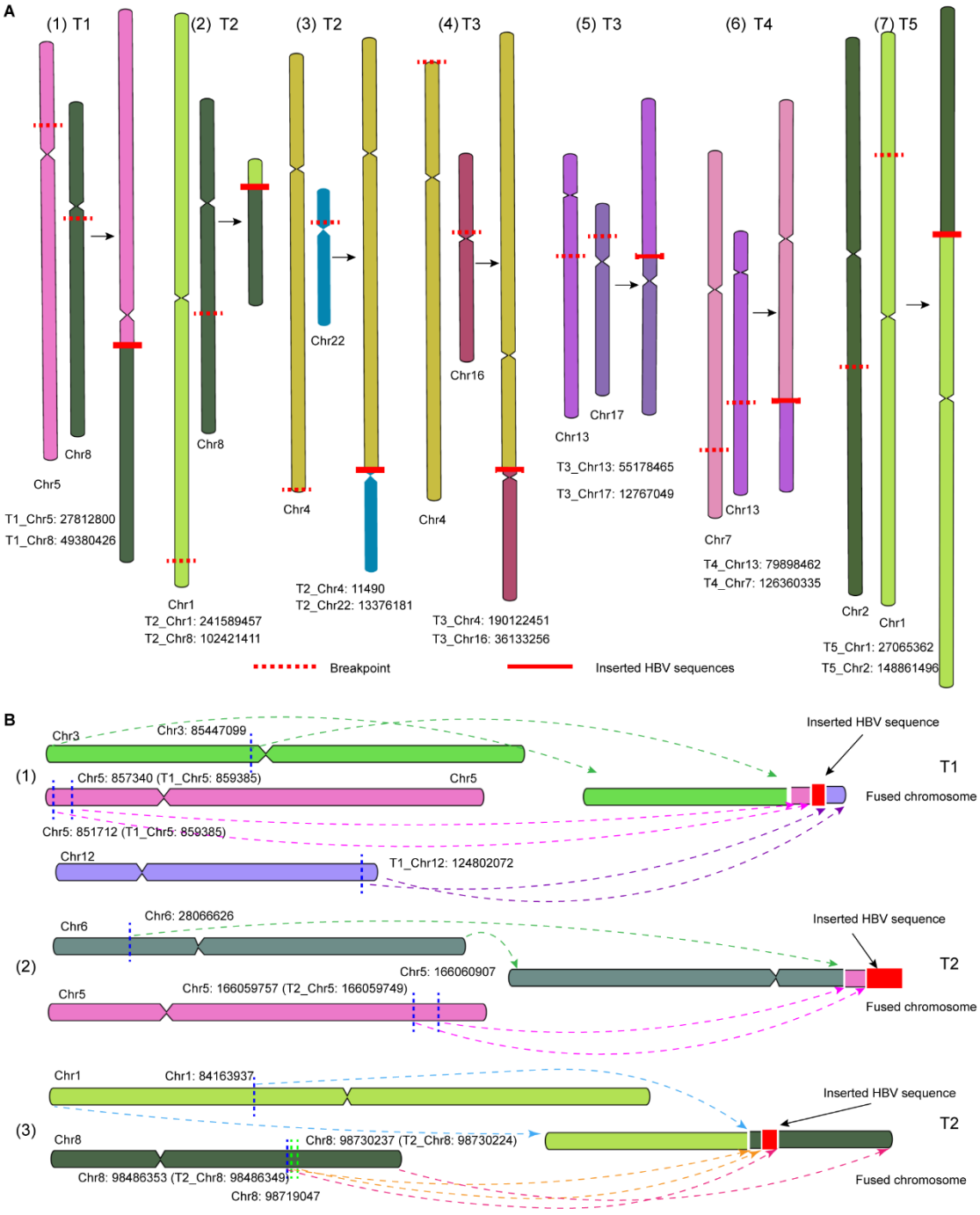
Supplemental_Fig_S5. The distribution of nearest distances between different types of SVs and the integrated loci determined with bootstrap analysis, as well as the SV counts comparison of in and out the region within 10 Mb. (A) Duplication; (B) Deletion; (C) Insertion; (D) Inversion.

Supplemental_Fig_S6



Supplemental_Fig_S6. Copy number plots of the samples with telomeric-like deletions associated with HBV DNA insertions in tumor samples.

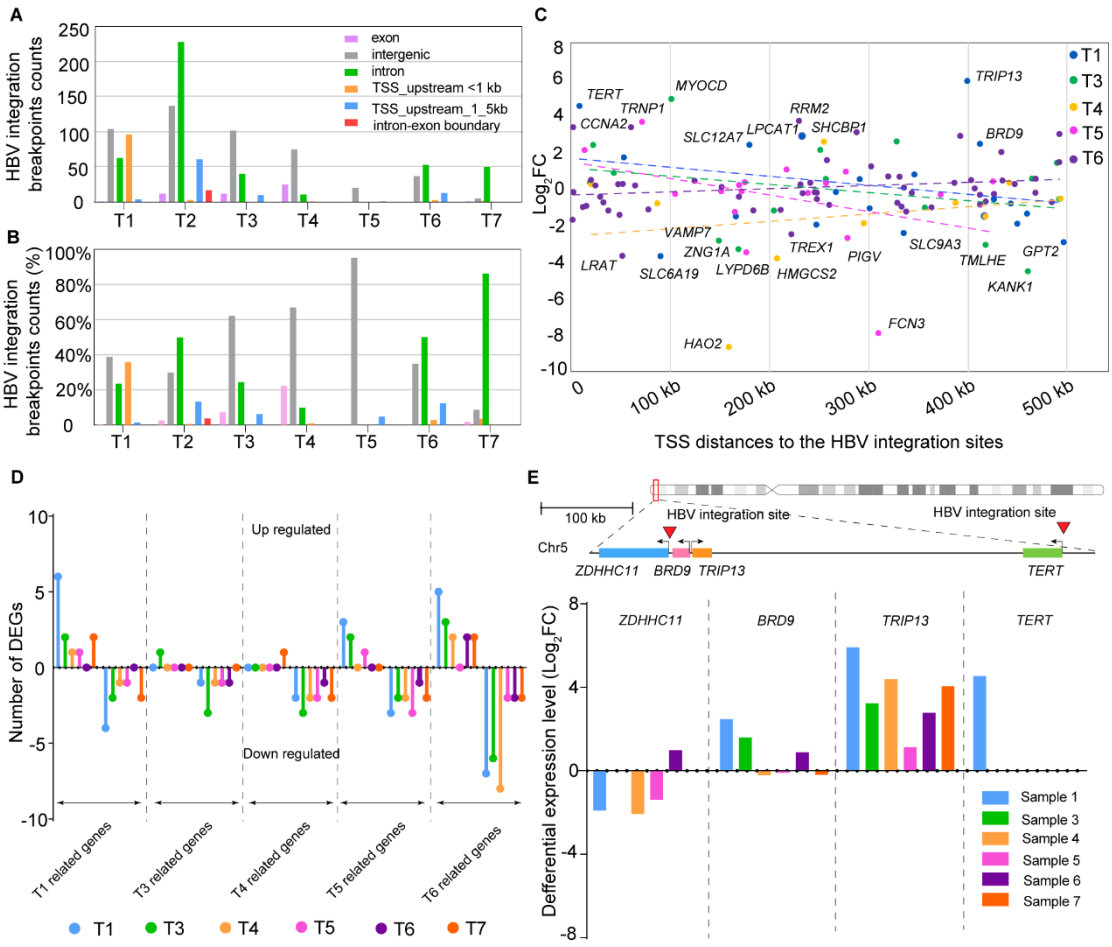
Supplemental_Fig_S7



Supplemental_Fig_S7. Interchromosomal translocations in tumor samples. (A) Major chromosomal translocation in each tumor tissue (each chromosomal translocation supported by more than 3 HBV–human chimeric reads). (1) – (7) represent different cluster of translocations in each sample; the integration locus IDs

are indicated. (B) Three patterns of interchromosomal fusion associated with HBV integration. The breakpoints are indicated with the related integration locus IDs in the parentheses.

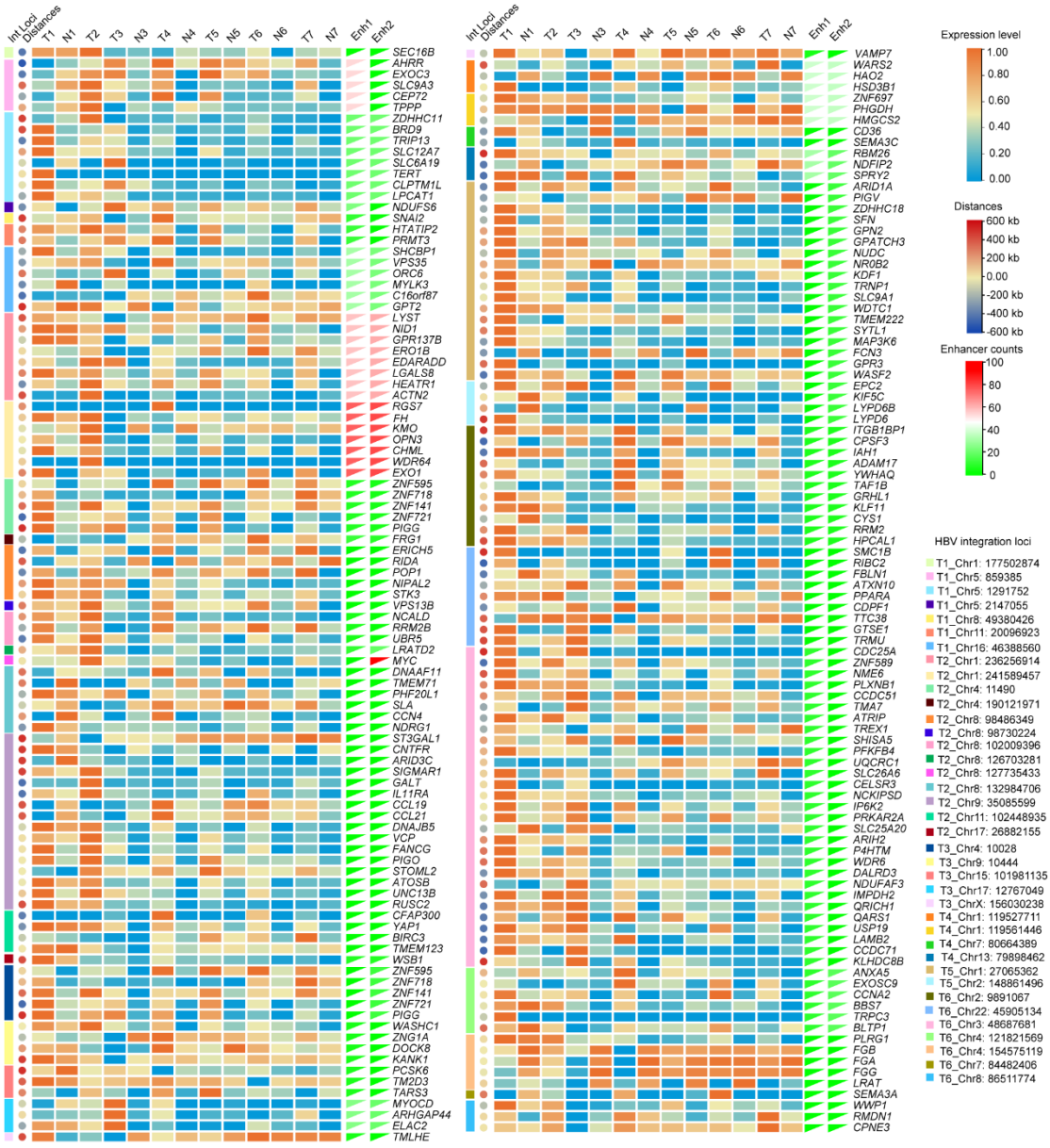
Supplemental_Fig_S8



Supplemental_Fig_S8. HBV integration mediated aberrant gene expression. (A) and (B) Distribution of the HBV integration breakpoints. The x-axis represents the sample number, different colors represent different annotations of the HBV integration loci (exon, intergenic, etc.); the y-axis represents the HBV integration breakpoints' counts within the regions (A); the y-axis represents the proportion of the HBV integration HBV breakpoints located within the regions) (B). (C) The impact of the distances between the TSSs and HBV integration sites on gene expression. The x-axis represents distances between the human genome breakpoints with HBV integration and

the adjacent gene transcription start site; the y-axis represents the relative expression level of the genes located within 500 kb upstream and downstream of the main HBV integration sites. The differentially expressed genes are annotated. (D) The x-axis represents the sample number related to the genes associated with HBV integration; the y-axis represents the number of differentially expressed genes. (E) HBV integration in the proximal telomere region of Chromosome 5 and its effect on gene expression. Schematic representation of the HBV integration locus and the up- and downstream genes (top) and differential gene expression levels of the related genes (bottom).

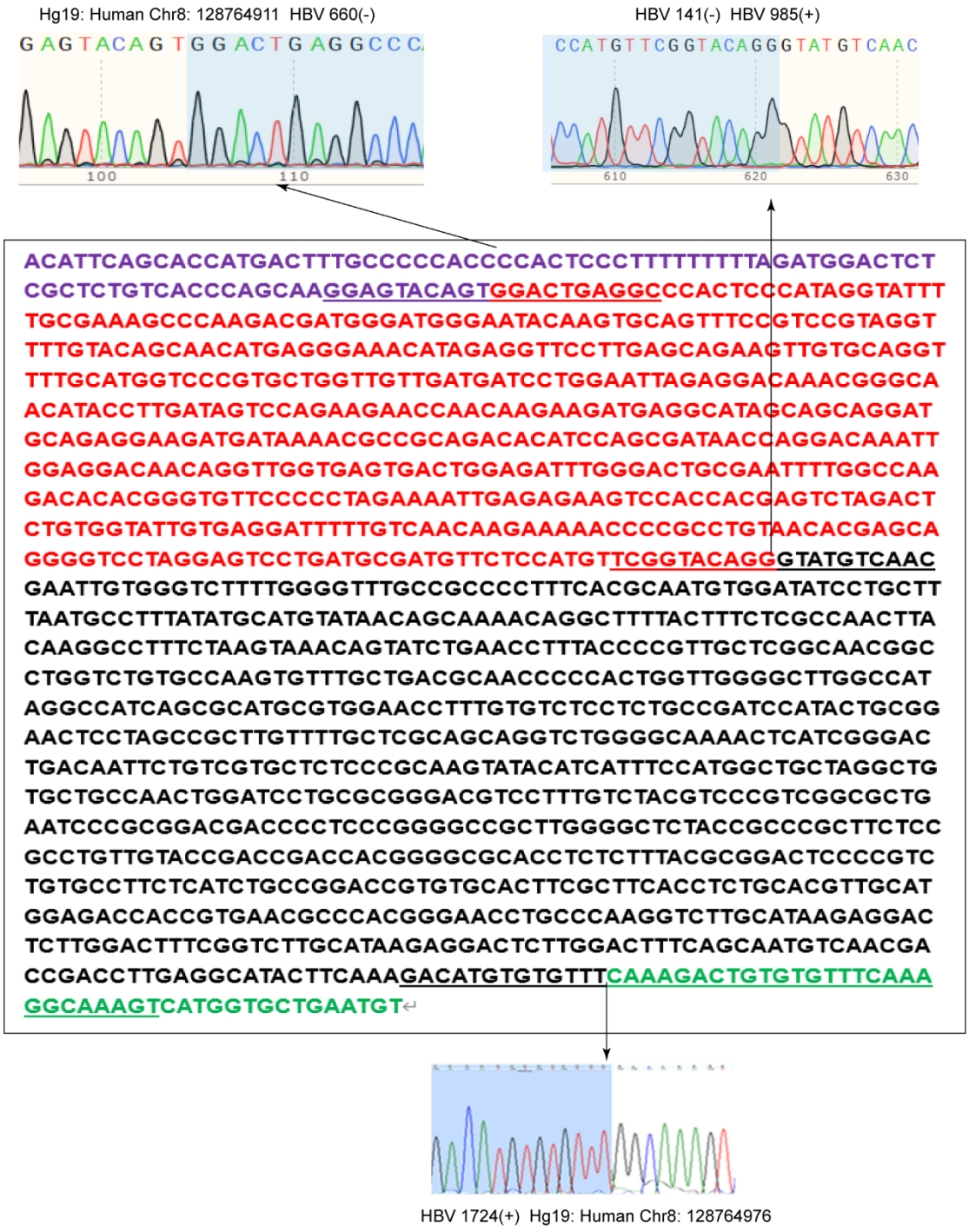
Supplemental_Fig_S9



Supplemental_Fig_S9. Heatmap of the genes' expression levels nearby the HBV integration sites (+- 500,000 bp). The first color-coded bar plot panel (leftmost column) represents HBV integration loci, while the second panel from the left displays genomic distances between integration sites and transcription start sites (TSSs) of adjacent genes. The coordinate system is defined with positive values indicating upstream positioning

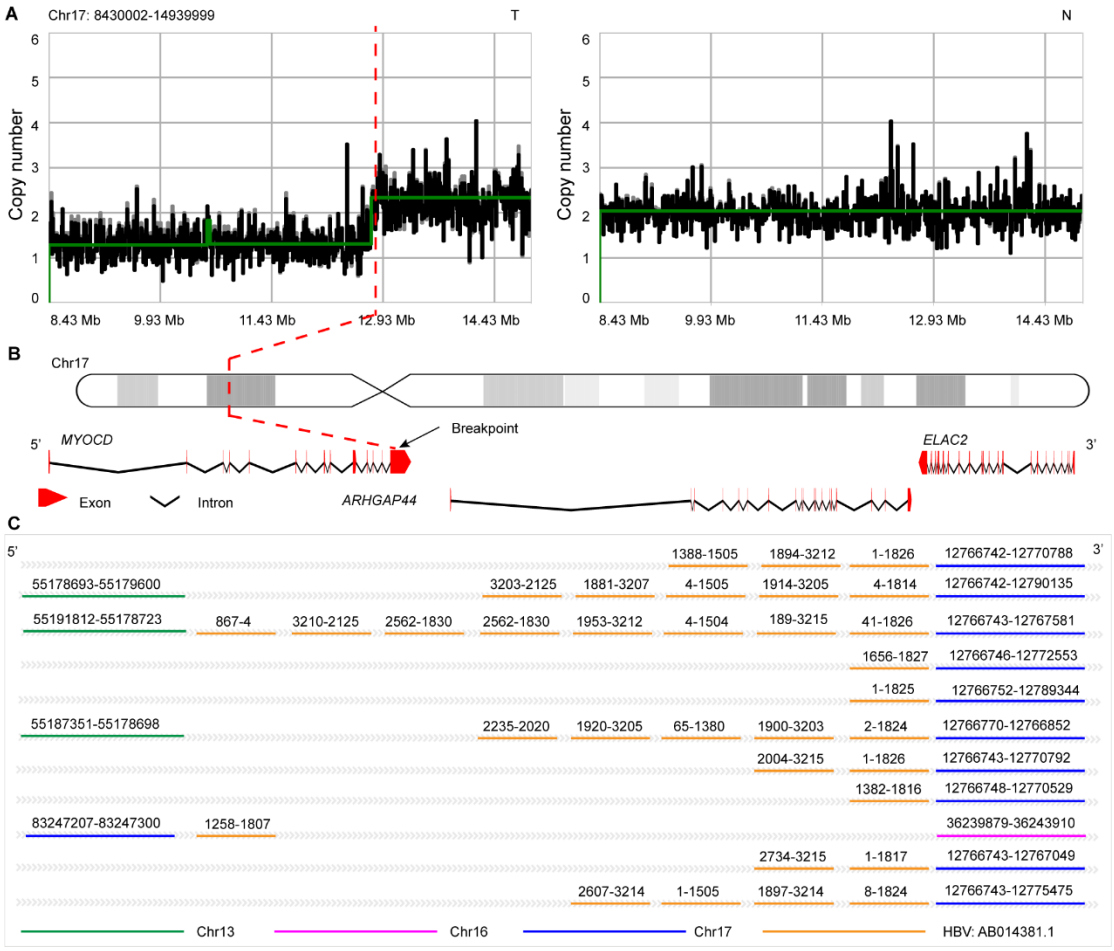
of integration loci relative to gene TSSs (towards the 5’ direction), and negative values corresponding to downstream localization (3’ orientation).

Supplemental_Fig_S10



Supplemental_Fig_S10. Full-length sequence of the inserted HBV sequence in the hotspot between *MYC* and *PVT1* and the representing Sanger sequencing results.

Supplemental_Fig_S11



Supplemental_Fig_S11. HBV integration in the *MYOCD* region. (A) Copy number variations of Chr17: 8430002–14939999 in tumor (left) and nontumor (right) tissues of sample 3; (B) Schematic representation of the HBV integration sites of the *MYOCD* region on Chr17; (C) HBV sequence insertions in the *MYOCD* region containing HBV-*Enh I*.