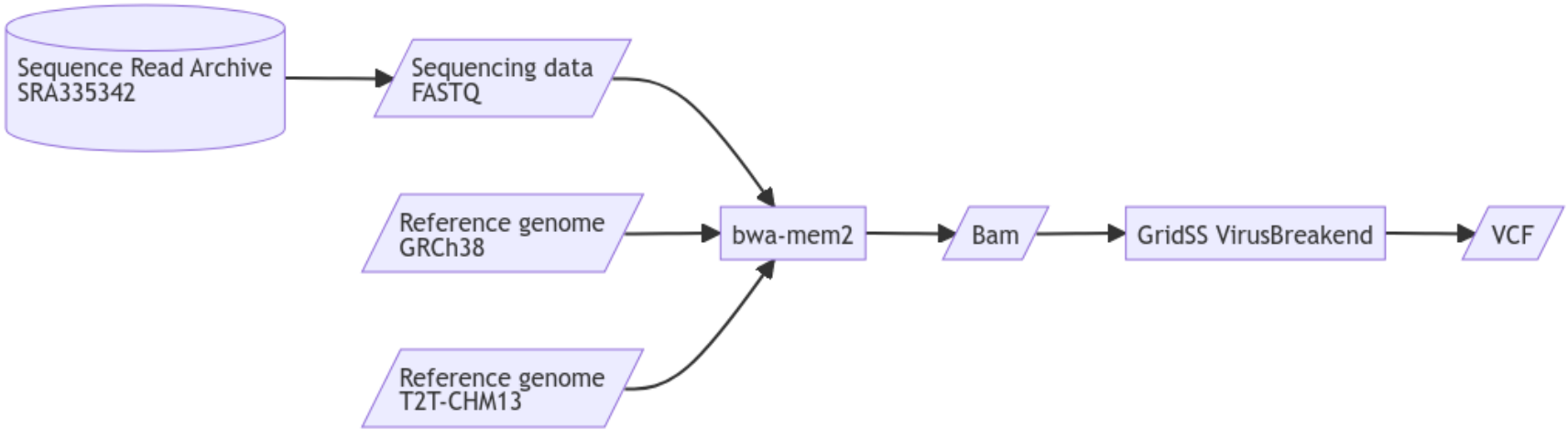
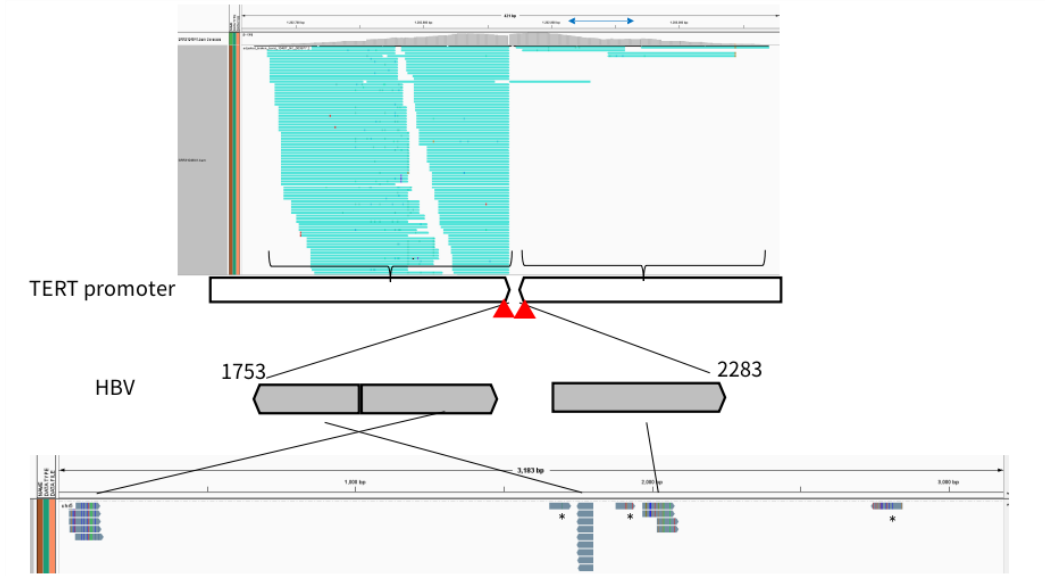
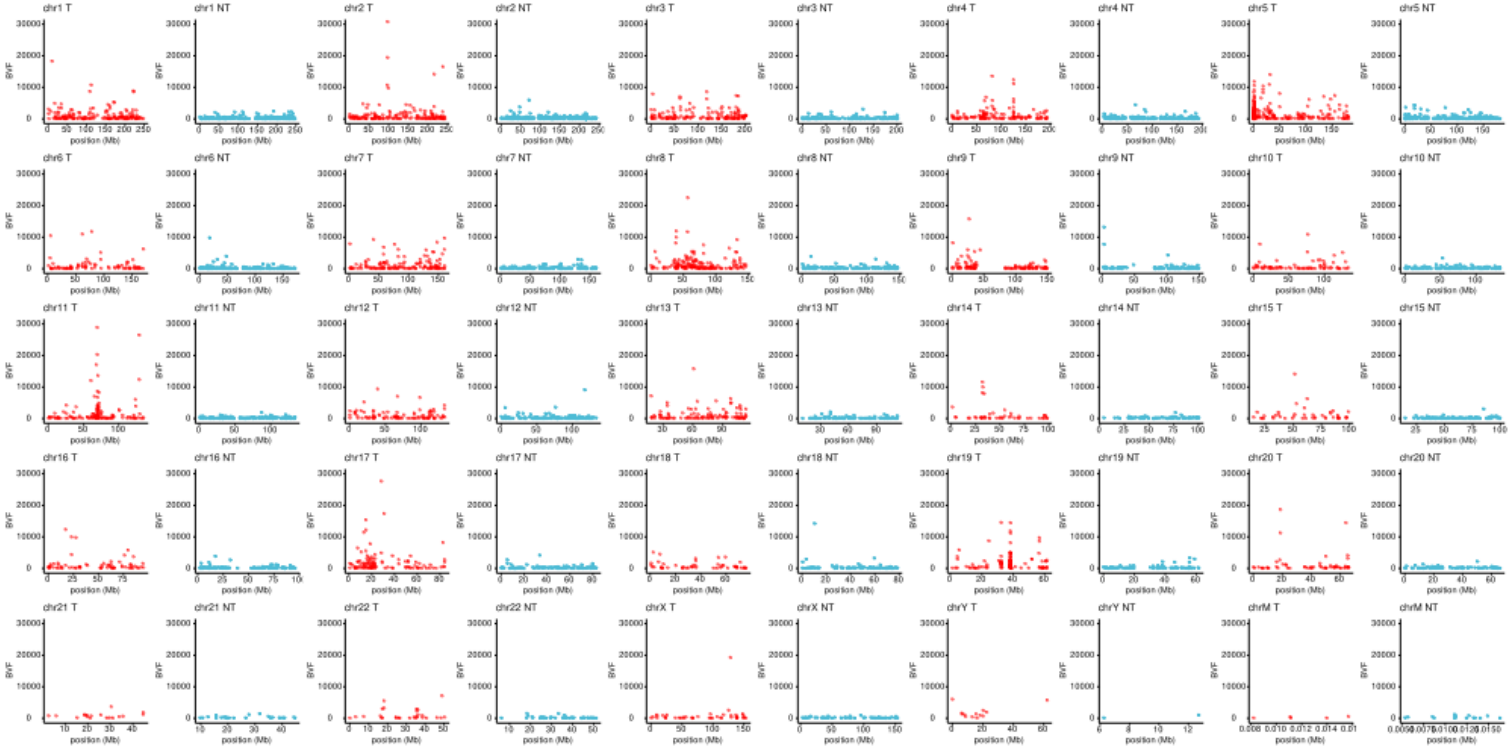



**Supplementary Figure 1 Overview of HBV integration discovery pipeline.** All reads in the dataset were aligned to the GRCh38 and T2T-CHM13 reference genomes using bwa_mem2. VIRUSBreakend was used to detect integration sites, and the analysis was performed using Nextflow on Amazon Web Service. HBV integration sites were detected using GRIDSS VIRUSBreakend. Integration sites were compared with the count of fragments providing breakend for the variant allele (BVF) in the variant call format specification (VCF) file. HTSlib and its Ruby binding, ruby-htslib, were used to collect BVF values from the VCF files. Statistical analysis and visualization were performed using R software.

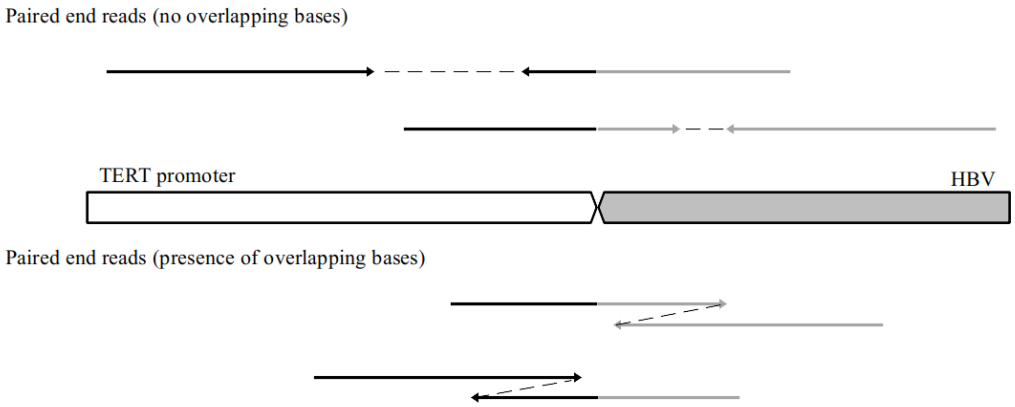



**Supplementary Figure 2 An example of HBV integration breakpoints in the *TERT* promoter region of SRR3104641.** An example of HBV integration in the *TERT* region (promoter). Upper; Turquoise reads in the figure have their paired mates aligned to the HBV genome. The red arrowheads indicate the breakpoints. VIRUSBreakend detected two breakpoints (at chr5: 1,202,865 (HBV: 1,753) and chr5: 1,202,871 (HBV: 2,283)), while the original study only detected the latter and four nearby breakpoints in the low-complexity region (blue arrow). T2T-CHM13 was used as the human reference genome and NC_003977 was used as the HBV reference genome. Lower; Three peaks of reads aligned to the HBV genome that have mates aligned to the human *TERT* region, due to structural variation in the HBV genome. The white box represents the human genome and the gray box represents the HBV genome. * These reads indicate integration breakpoints outside of the *TERT* gene on chromosome 5.




**Supplementary Figure 3 Scatterplot of BVF across chromosomes.**

**Supplementary Figure 4 An example of HBV integration breakpoints in the *TERT* promoter region of SRR3104862.** The figure shows an example of HBV integration sites in the *TERT* region (distal intergenic) that were not detected by the original study, but were detected by VIRUSBreakend. T2T-CHM13 was used as the human reference genome, and NC_003977 was used as the HBV reference genome. Left; For turquoise reads shown in the figure, their paired mate reads aligned to HBV. Two integration breakpoints were detected at chr5: 1,207,615 (HBV: 1,986) and at chr5: 1,207,620 (HBV: 1,806) (located in the center). There are several paired-end reads without overlapping bases. The integration pattern suggested 1 to 4 bp deletion of human genome between the breakpoints. Right; There are two peaks of reads aligned to HBV whose mate is aligned to human *TERT* region. The 3′ end of their reads corresponds to the breakpoints mentioned above. In the viral end, they are located near the direct repeat 1 of HBV genome.




**Supplementary Figure 5 Paired-end reads lacking overlapping bases.** Upper; some paired-end reads could not be assembled because of the absence of overlapping bases. Lower; paired-end reads can be assembled. White box, human genome; gray box, HBV; Black line, read aligned to the human genome; Gray line, read aligned to the HBV genome; Arrows indicate the direction of the genome.

The URL of the page used in Figure 2D to display the copy number of liver cancer on cBioPortal.

https://www.cbioportal.org/results/cnSegments?
cancer_study_list=lihc_tcga_pan_can_atlas_2018&Z_SCORE_THRESHOLD=2.0&
RPPA_SCORE_THRESHOLD=2.0&profileFilter=mutations%2Cstructural_variants%2Cgistic&
case_set_id=lihc_tcga_pan_can_atlas_2018_cnaseq&gene_list=CCND1&geneset_list=%20&
tab_index=tab_visualize

Supplementary Figure 2 and Supplementary Figure 4 were created using Integrative Genomics Viewer (IGV).

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol 2011; 29: 24-26 [PMID: 21221095 DOI: 10.1038/nbt.1754]