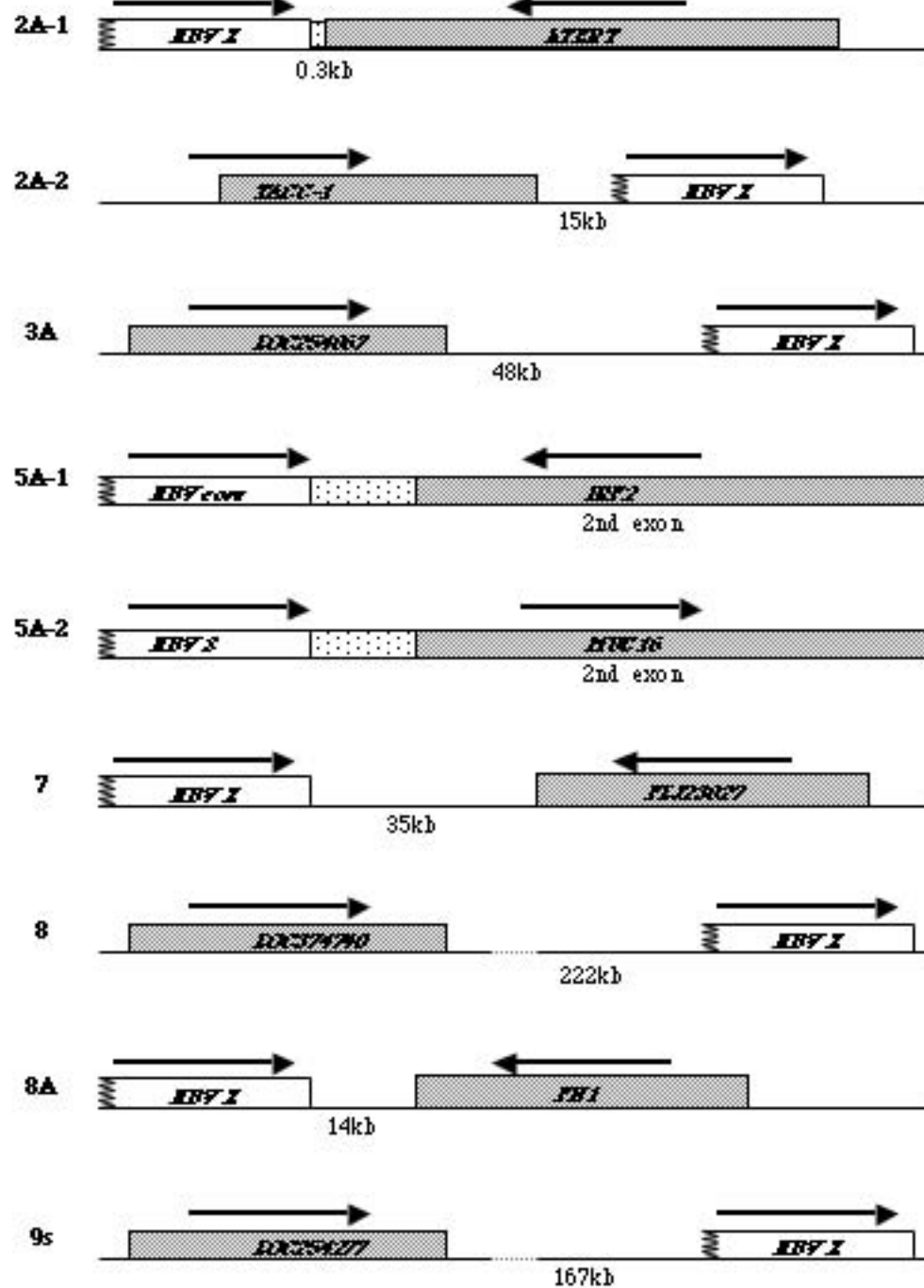


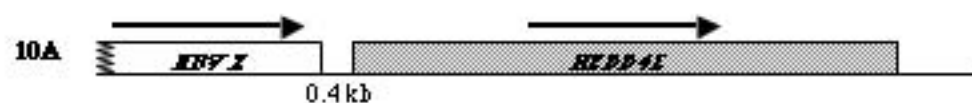
Supplementary Figure 1

Left part: Schematic drawing of HBV DNA integration in cellular genes. Open box: HBV sequence. Dotted box: cellular regions identical to GenBank sequences (cf Table 3 and 4). Shaded box: coding sequences. Bold arrowheads: orientation of the ORF. The distance between the gene ATG or stop codon and the HBV integration site is indicated (kb). The letter on the left side of the figure indicates the HCC code. The orientation of the HBV integrants as compared to the coding sequence is also indicated in Table 3 where the full name of the gene or protein is reported.

Right part: Sequence of the last 20 HBV nucleotides found at the integration site in the isolated HBV/cellular DNA junctions. The nucleotide position in the viral genome at the junction is indicated, (numbering from the hypothetical EcoRI site of subtype adw⁴⁹). The sequence of the 20 first cellular nucleotides at the integration site is on the right.



viral sequence	cellular sequence
TG TTC ACC AGC ACC ATG CCA 1817	GGG TG CAG GGG AC G CC CAG CGA
CTT CAA AGA CTG TTT GTT TA 1725	CAC ACA CTT AA ATTT CTT TT
ATA AAT TGG TCT GTT CAC CA 1808	AAG AA ATGT CG AA ATTT AAG
AAT TTGG AGA AACT GTGG AG 1942	AACTT AAC CCTTTT GGATTA
GAT TGG AAAGT ATGT CAA AG 997	ATGGGG TTTGG CTTTGTGCC
ACC AC GGG GCGT ACC TCT CT 1535	GTAG AAG TTGT GCG CCT CTC
GTT CACC AGC ACC ATG CAA C 1820	TGT CCC GAAT GGT ATTT CCT
GTACT AGG AGG CTGTAG GCA 1789	ATGTT CCAC CGG TTTGA ACA
GTCC CGT CGG CGCTGA ATCC 1596	AGCCT GGG CGAC AGAG CAAG



ACTAGGAGGCTGTAGGCAAA
1789

CCCATGAATACTTGGGAAGGC



GGATCCTGCGCGGGACGTCC
1419

TGACCCAGGCCTTCGCCGCC



GACTCTTGACTCTCAGCAA
1680

CACCACTGCCATCACTGCCA



CTAGCCGCTTGTTTTGCTCG
1449

CCAGCTATGGGATTCAGGAA



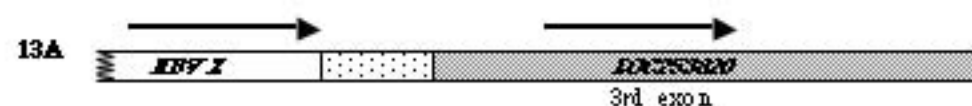
ACCTCTGCACGTGCATGGA
1612

AAGGGAAAACCTGGCAGTGCC



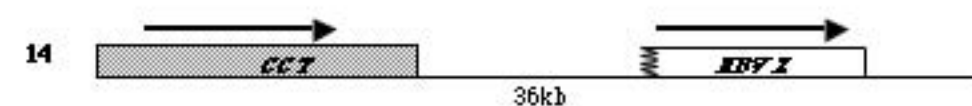
TGTGCCTTCTCATCTGCCGG
1721

AAGGGAAAACCTGGCAGTGCC



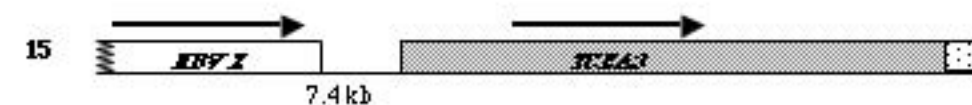
GGCTGTAGGCATAAATTGGT
1573

CCATTTACATTTAAAGTTAA



AGTTGGGGGAGGAGATTAGG
1757

CAGAGTTGGGGGAGACTAG



AGATTAGGTTAATGATCTTT
1771

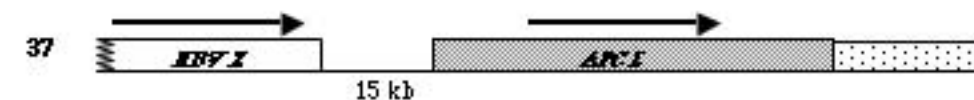
ACAGGGCTTGAGGGGAGGAG



AAATCTCATCGGAAGTGAACA

GGATGTGGCCTTGCATGGCT

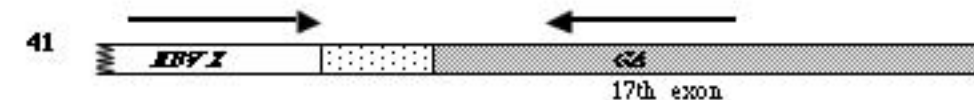
1336



AGGTCTTTGTACTAGGAGGC

CTTCCGACTGCAGACAGAGC

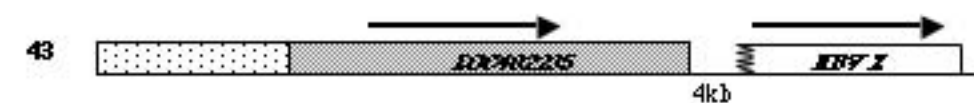
1781



AGGCTGCAGGTTTAAAAGGG

AAAAATATATATGTGTTTGC

1783



TTGCCCAAGGTCTTACATAA

TGCAAGTGTGTTGCTGACGCC

1657



TCTGTGCACCAGCACCTGGC

CTCACGGATCTGGTAGGGAA

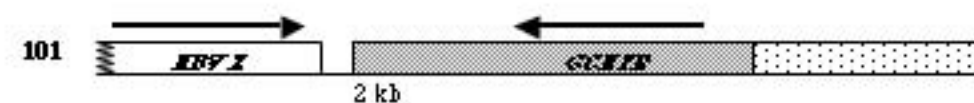
1657



GAGGCTGTAGGCATAAATTG

CCTGGCCCATAATGTGTTTC

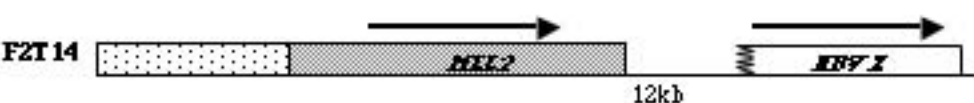
1796



CCAGCACCATGCAACTTTTT

TCCTCTTGACAACTTCAATA

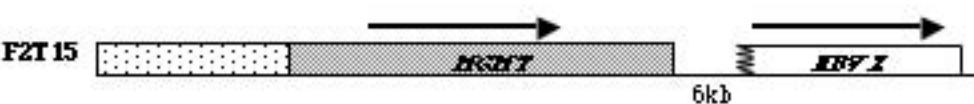
1824



AGACTGGGAGAATTGGGGGA

GTACAACCCCTCACGGTCAG

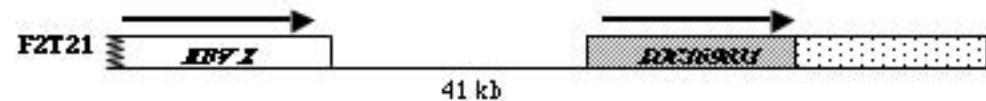
1747



GAAACTAAGTTAATGATCTT

GATTCGCCCCGTTTACACGGT

1768



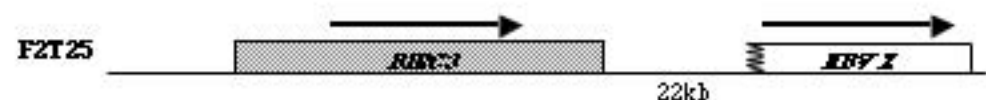
TTGCCCAAGGTCTTACATAA
1657

CAATGTTTAADDGTGCATAC



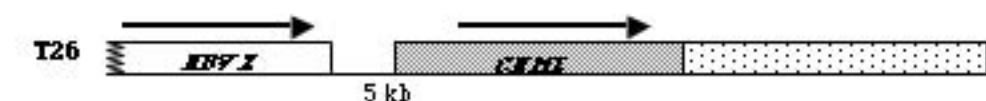
TTGGTCTGTTACCAGCACC
1768

AAGTGGTTTCTCTTTCTCCA



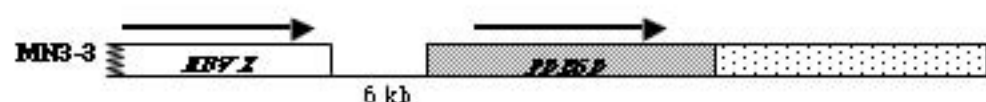
ACCAGCACCATGCAACTTTT
1824

GTTTCTCTTTCTCCAAGAGG



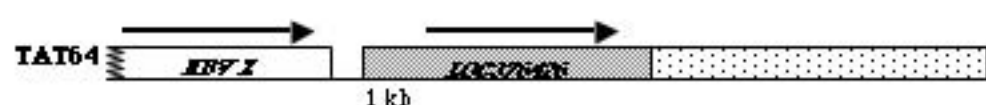
TAAATTGTCTGTTACCAGC
1811

GCGGTGCGCCGCTCTAGCCG



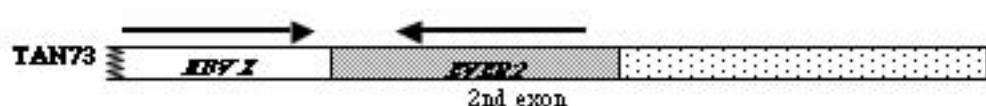
ACTAGAGCTGTAGCATAAAT
1792

GCGTGCGCCGCTCTAGCCCG



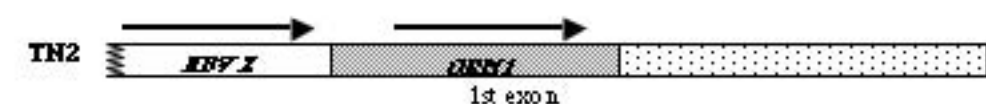
TGTTCCCAGCACCATGCAAC
1818

GCGGTGCGCCGCTCTAGCCC



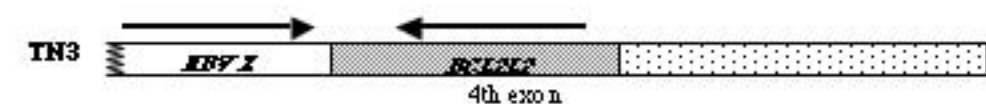
ACCACACCATGCAACTTTTT
1825

GCGGTGCGCCGCTCTACCGG



TTGCCCAAGGTCTTACATAA
1659

CAAGTGGTTTTATATCGCAT



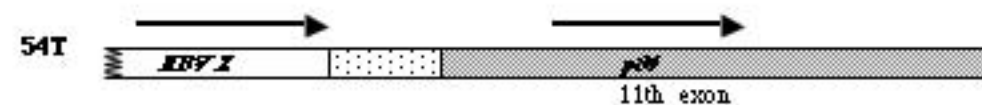
AGACTGGGAGAATTGGGGGA
1747

CAGGTCTCTGCCAAGTGGTT



TTAAAACTGGGAGATTGGGG
1745

CGCGGTGCGCCGCTCTAGCC



AATTGGTCTGTTACCAGCA
1810

TTAAAATATGACTTTGTTA



ACCACCAGGCATCTTTTTCA
1827

AGTGAATTGGACCTGAGTA



CTGTAGGCATAAATGTTTTTC
1790

TGTATTCTGATTTGCCTTCT



GATCTTTGTACTAGGAGGCT
1781

GTAGGCATAAATTGGTCTGT



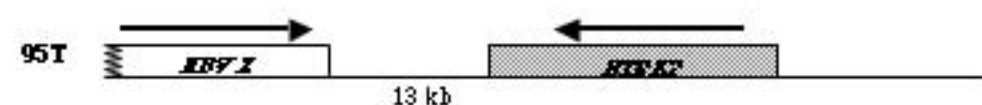
TAGGTTAAAGGTCTTTGTAC
1772

GGAAATACTCAAGAAGAAAA



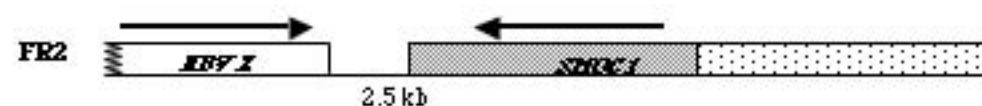
TGTTACCAGCACCATGCAA
1819

GCTGGTGATACAGCAGTTTG



AATTGGTCTGTTACCAGCA
1810

TCTTTGCAGCCATTTAAAT



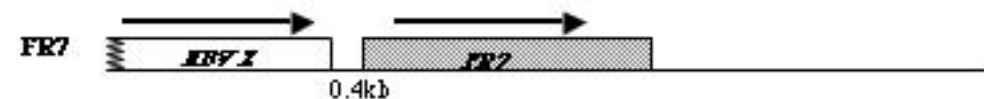
TTTGTACTGGGAGGCTGGAG
1785

TAAGTAAATATGCTTCAGAT



AGCACCATGCAACTTTTTCA
1827

TGCTATTGCATACTTTATAG



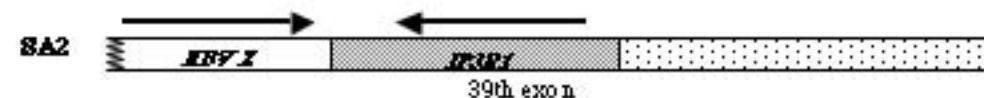
GACCGACCTTGAGGCAAAC
1707

CTTCTGGTATAAAGTGGGGA



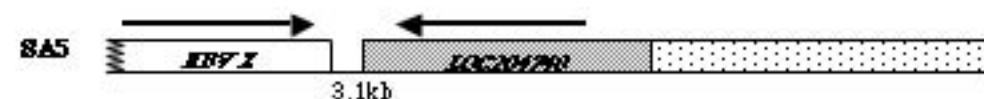
AGCAATGTCAACGACCGACC
1693

AGTTTTGTCAAAGTTTCAGG



GGGGAGGAGATTAGGTTAAT
1760

TATTAGGAGGCTGTAGGCAT



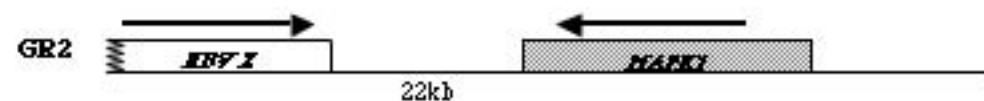
AGGCTGTAGGCATAAATTGG
1796

TTCAGTTGGAAAAGATTTGT



TGCGCACCAGCACCATGCAA
1818

AAACAAAAACAAAAACTTGG



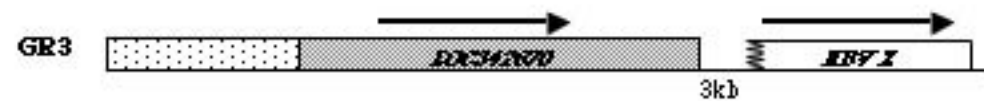
GCACCAGCACCATGTAACCT
1825

CGAGACCGGGGAACACTGTG



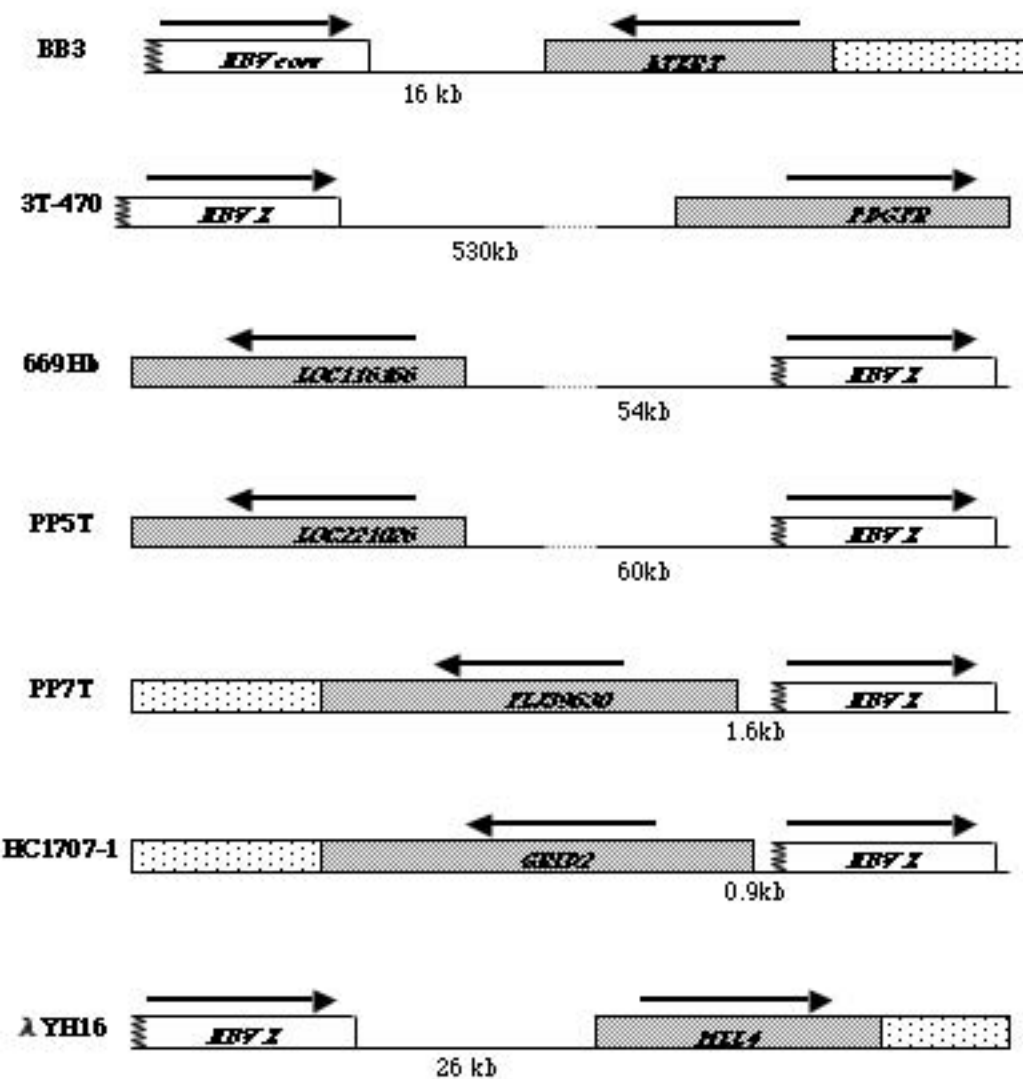
AATTGGTCTGCGCACCAGCA
1810

ATATTGCAATGTCGGACCCA



GTCTGCGCACCAGCACCATG
1815

AGCCAGGACCCGCGGCCCTC



(5')

viral sequence

cellular sequence (3')

TCCATGCCCCAAAGCCACCC
1862

TATATCACAGACCGAGGTAG

CTTTCCTTCCGTCAGAGATC
1961

CTATTCAACTTGTTTAACCC

CCCATGCCCTTTCGAGGTTT
2871

GGAAGACCCCCTGCCTAATC

CCCTGCTCGTGTTACAGGCG
205

AAACTGAATAGAACACATTA

AAGGCAGCCTACTCCCATCT
3146

TAACATGCTATCTTATCTAT

CCAGCACCATGCAACTTTTT
1798

GCTCATCACATTAAAAGCAG

ATTTGGAGCTTCTGTGGAGT
1916

AAGTTTACCCCAAGCGAAGC