

Table S1. HCC samples with etiology and map counts.

Tumor ID	etiology	number of nodules	Tumor grade*	largest tumor diameter (mm)	Original name**	Total read count	Mapped on human genome	Mapped on HBV genome	Non-tumor ID	fibrosis score (Metavir)	Original sample name**	Total read count	Mapped on human genome	Mapped on HBV genome
1T	HBV	3	G1	70	17T	13,704,114	11,131,494	6,707	1NT	F4	17NT	4,952,149	4,089,802	305
2T	HBV	2	G2	50	23T	34,526,410	32,495,689	6,208	2NT	F4	23NT	31,859,986	29,345,240	58
3T	HBV	1	G2	85	52T	9,393,758	8,529,410	-	3NT	F2	52NT	15,459,585	14,346,903	-
4T	HBV	2	G3	60	71T	9,448,052	8,794,498	336	4NT	F4	71NT	24,525,111	22,857,897	388
5T	HBV	1	G1	25	81T	29,757,451	27,673,138	17,680	5NT	F4	81NT	36,083,947	33,644,047	105,051
6T	HBV	5	G3	90	86T	13,982,080	12,941,944	175	6NT	F1	86NT	13,267,499	11,990,398	893
7T	HBV	2	G2	10	133T	13,554,448	12,241,611	3,062	7NT	F2	133NT	15,809,573	13,918,837	3,816
8T	HBV	2	G2	40	137T	18,361,250	16,760,712	14,348	8NT	F4	137NT	14,928,380	13,576,095	41
9T	HBV	1	G2	160	140T	25,775,615	23,950,279	520	9NT	F3	140NT	18,848,537	16,907,315	16,395
10T	HBV	2	G3	60	141T	19,346,705	17,575,608	18,374	10NT	F4	141NT	20,099,768	18,015,093	12,960
11T	HBV	1	G2	180	154T	18,581,260	16,944,580	8,569	11NT	F4	154NT	22,278,862	20,214,152	4,545
12T	HBV	2	G1	25	155T	19,044,662	17,100,978	45,755	12NT	F4	155NT	10,352,268	9,422,626	20,565
13T	HBV	1	G2	45	156T	11,163,135	10,290,388	12,161	13NT	F4	156NT	8,100,277	7,449,314	3,745
14T	HBV	1	G2	25	173T	33,628,289	31,429,432	523	14NT	F2	173NT	31,890,433	30,102,764	834
15T	HBV	7	G2	25	174T	13,286,218	12,318,911	4,939	15NT	F4	174NT	19,465,347	18,041,909	58,555
16T	HBV	2	G1	28	175T	32,000,160	29,904,230	7,763	16NT	F4	175NT	26,287,447	25,091,661	1,483
17T	HCV	1	G2	240	11T	31,347,758	29,386,132	-	17NT	F1	11NT	16,738,091	15,788,083	-
18T	HCV	1	G1	180	12T	12,274,044	11,399,462	-	18NT	F3	12NT	22,067,957	20,430,729	-
19T	HCV	4	G1	15	29T	17,434,011	16,246,397	-	19NT	F4	29NT	8,921,719	8,238,391	-

* Tumor grade (G1 to G4) was defined according to a modified version of the Edmonson-Steiner classification . ** Sample ID as shown in reference: Hashimoto et al. 2015.

Table S2. Representative HBV genomes

Number	Accession	Genotype	Length
1	EU054331.1	A	3221
2	X02763.1	A	3221
3	AB219428.1	B	3215
4	D00331.1	B	3215
5	GQ924620.1	C	3215
6	GQ358158.1	C	3215
7	AF121240.1	D	3182
8	FJ904433.1	D	3182
9	AB106564.1	E	3212
10	AY738145.1	E	3212
11	AY090458.1	F	3215
12	X75658.1	F	3215
13	AB064313.1	G	3248
14	AF160501.1	G	3248
15	AY090454.1	H	3215
16	FJ356716.1	H	3215

Table S3. Normalized expression values for 17 TSSs

Genbank ID: GQ358158.1

						Tumor (Expression: TPM)																
Peak	Gene	Start	End	Total count (T)	Strand	1	2	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	Med
1	PreS/S	2802	2807	2446	+	17.2	0.0	3.2	17.8	0.8	0.9	3.3	0.8	0.2	0.0	15.8	52.7	0.1	1.1	19.9	8.9	1.1
2	PreS/S	3183	19	98266	+	14.6	153.6	23.3	497.3	5.9	210.8	387.2	17.1	773.1	390.7	1397.9	972.2	8.5	260.0	119.4	348.8	210.8
3	S	25	121	13817	+	407.0	13.5	0.4	16.5	0.3	0.6	23.7	0.3	88.7	23.7	143.7	14.9	0.4	14.8	48.7	53.1	14.9
4	S	132	133	184	+	5.5	0.2	0.1	0.2	0.0	0.0	0.4	0.0	1.6	0.3	1.7	0.0	0.0	0.6	0.3	0.7	0.2
5	S	203	207	219	+	1.2	0.6	0.4	2.5	0.0	0.5	0.7	0.0	0.4	0.3	2.1	0.5	0.0	0.5	0.5	0.7	0.5
6	S	258	262	126	+	0.6	0.5	0.2	0.7	0.0	0.0	0.4	0.0	0.5	0.2	1.3	0.1	0.0	0.5	0.8	0.4	0.4
7	S	284	285	216	+	0.4	1.1	0.1	3.1	0.1	0.0	1.0	0.0	0.7	0.0	0.0	0.4	0.0	0.4	1.2	0.6	0.4
8	S	302	303	60	+	0.0	0.1	0.0	0.4	0.0	0.4	1.1	0.0	0.0	0.1	0.6	0.1	0.0	0.1	0.1	0.2	0.1
9	S	759	760	368	+	0.0	0.6	0.0	1.2	0.0	0.1	2.5	0.0	0.9	1.8	3.6	1.3	0.0	0.5	3.9	1.1	0.6
10	X	1210	1211	48	+	0.0	0.1	0.0	0.7	0.0	0.0	0.3	0.0	0.3	0.0	0.2	0.3	0.0	0.2	0.1	0.1	0.1
11	X	1523	1524	395	+	0.1	0.1	0.2	1.2	0.1	0.1	0.0	0.0	2.8	0.3	2.2	3.9	0.1	1.5	5.6	1.2	0.2
12	AS	1557	1610	107	-	0.0	0.1	0.0	0.1	0.0	0.0	0.3	0.0	0.8	3.7	0.3	0.0	0.0	0.0	0.1	0.4	0.0
13	X	1572	1573	49	+	0.1	0.0	0.0	0.1	0.0	0.0	0.3	0.0	0.4	0.0	0.3	0.4	0.0	0.6	0.5	0.2	0.1
14	X	1655	1656	159	+	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.5	1.5	1.3	0.2	4.0	1.3	0.6	0.1
15	PreC/C	1745	1746	402	+	0.0	0.0	0.0	0.3	0.0	0.1	0.0	0.2	8.0	0.6	7.8	0.0	0.2	5.1	0.0	1.5	0.1
16	pgRNA	1816	1819	20869	+	0.1	0.0	5.3	13.6	4.6	0.1	282.7	0.3	0.0	22.0	734.9	0.4	4.2	44.0	0.5	74.2	4.2
17	AS	1822	1834	21	-	0.0	0.0	0.0	0.1	0.0	0.0	0.8	0.0	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.1	0.0

						Non-Tumor (Expression: TPM)																
Peak	Gene	Start	End	Total count (NT)	Strand	1	2	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	Med
1	PreS/S	2802	2807	10302	+	6.9	0.0	1.5	183.7	6.3	1.9	0.1	77.1	11.6	0.0	16.2	37.3	0.8	64.5	1.9	27.3	6.3
2	PreS/S	3183	19	103617	+	28.1	0.7	5.1	1104.0	43.8	175.5	1.1	668.1	538.9	162.4	1696.5	370.1	4.1	625.6	8.1	362.1	162.4
3	S	25	121	5187	+	19.2	1.0	0.2	39.2	2.4	13.8	0.0	11.1	40.5	13.6	87.4	8.4	0.2	52.7	2.2	19.5	11.1
4	S	132	133	64	+	0.0	0.0	0.0	0.4	0.0	0.3	0.0	0.1	0.3	0.1	1.3	0.1	0.0	0.9	0.1	0.2	0.1
5	S	203	207	308	+	0.0	0.0	0.0	6.4	0.2	0.3	0.1	0.5	0.2	0.4	1.3	0.6	0.0	1.2	0.0	0.8	0.2
6	S	258	262	156	+	0.0	0.0	0.0	3.1	0.2	0.1	0.0	0.4	0.1	0.1	0.1	0.0	0.0	1.3	0.1	0.4	0.1
7	S	284	285	312	+	0.0	0.0	0.0	7.3	0.1	0.3	0.0	0.9	0.3	0.0	0.0	0.0	0.0	1.0	0.0	0.7	0.0
8	S	302	303	58	+	0.0	0.0	0.0	1.2	0.0	0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.3	0.0	0.1	0.0
9	S	759	760	361	+	0.2	0.0	0.0	3.4	0.2	0.4	0.0	6.7	1.3	0.1	1.9	0.9	0.0	2.3	0.1	1.2	0.2
10	X	1210	1211	76	+	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.3	0.1	0.0	0.1	0.0	0.0	0.4	0.0	0.2	0.0
11	X	1523	1524	618	+	0.0	0.0	0.1	4.8	0.1	2.0	0.0	11.8	2.9	0.3	1.1	2.2	0.0	3.9	0.7	2.0	0.7
12	AS	1557	1610	50	-	0.0	0.0	0.0	0.9	0.0	0.2	0.0	0.0	0.1	0.0	0.3	0.0	0.0	0.1	0.4	0.1	0.0
13	X	1572	1573	68	+	0.0	0.0	0.0	0.1	0.2	0.1	0.0	0.8	0.8	0.0	0.1	0.0	0.0	1.1	0.3	0.2	0.1
14	X	1655	1656	157	+	0.0	0.0	0.0	0.1	0.0	0.4	0.0	0.8	0.5	0.1	2.0	0.1	0.1	4.7	0.1	0.6	0.1
15	PreC/C	1745	1746	704	+	0.0	0.0	0.0	0.1	0.0	0.0	0.0	26.5	7.8	0.2	0.0	0.6	0.0	1.7	0.0	2.5	0.0
16	pgRNA	1816	1819	91583	+	0.0	0.0	7.5	1391.1	10.7	23.0	1.2	0.2	1.1	21.3	68.1	5.3	18.1	1956.1	29.5	235.6	10.7
17	AS	1822	1834	186	-	0.0	0.0	0.0	4.2	0.1	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.6	0.6	0.4	0.0

Table S4. HBV model systems for CAGE analysis

Num	species	Name	Description	Read count	Mapped on host genomes	Mapped on HBV genome (AF121240.1)
1	human	Hep2.2.15	HBV genome integrated HepG2	23,093,242	19,406,656	11,870
2	human	Hep2.2.15	HBV genome integrated HepG2	17,222,419	15,071,334	9,280
3	human	Hep2.2.15	HBV genome integrated HepG2	22,395,701	18,589,376	11,473
4	human	PHH (primary human hepatocyte)	HBV infected primary human hepatocytes	9,876,339	7,967,062	13,812
5	human	PHH (primary human hepatocyte)	HBV infected primary human hepatocytes	6,747,370	5,611,204	6,152
6	human	PHH (primary human hepatocyte)	HBV infected primary human hepatocytes	5,952,961	4,790,630	10,289
7	human	HepAD38	hepatoma cell line derived from HepG2	8,109,433	6,270,257	54,258
8	human	HepAD38	hepatoma cell line derived from HepG2	11,442,448	9,402,212	51,788
9	mouse	AAV-HBV	mouse liver transduced with AAV-HBV	10,276,986	8,641,287	27,196
10	mouse	AAV-HBV	mouse liver transduced with AAV-HBV	13,764,553	11,488,641	46,682
11	mouse	AAV-HBV	mouse liver transduced with AAV-HBV	7,032,187	6,032,999	12,012

Table S5. Normalized expression values for 17 TSSs

Genbank ID: GQ358158.1

Genbank ID: GQ358158.1						Normalized expression values (tpm) for 11 samples													
						Hep2.2.15			PHH		HepAD38		AAV-HBV						
Peak	Gene	Start	End	Total count	Strand	1	2	3	4	5	6	7	8	9	10	11	Mean	Med	
1	PreS/S	2802	2807	2446	+	21.3	34.1	30.6	3.1	8.2	7.4	330.6	310.2	93.3	94.6	60.6	90.4	34.1	
2	PreS/S	3183	3234	98266	+	45.4	72.6	42.3	38.7	47.1	50.4	136.8	90.4	1122.7	1322.3	796.9	342.3	72.6	
3	S	25	121	13817	+	4.2	6.0	3.0	3.4	6.2	4.9	15.7	10.7	23.0	28.6	14.8	11.0	6.2	
4	S	132	133	184	+	0.2	0.1	0.1	0.0	0.3	0.0	0.5	0.9	2.1	1.2	1.7	0.7	0.3	
5	S	203	207	219	+	3.2	5.2	2.9	0.8	5.3	5.0	12.2	15.7	10.7	10.3	5.7	7.0	5.3	
6	S	258	262	126	+	1.0	1.1	0.8	0.0	0.9	1.2	1.8	1.9	0.6	1.7	1.1	1.1	1.1	
7	S	284	285	216	+	3.3	6.8	3.0	0.7	2.8	2.0	7.6	12.4	4.7	4.5	4.0	4.7	4.0	
8	S	302	303	60	+	0.1	0.1	0.2	0.0	0.1	0.5	0.4	0.5	0.7	0.5	0.4	0.3	0.4	
9	S	759	760	368	+	4.9	5.0	3.8	1.4	5.5	4.0	13.9	18.4	17.3	11.3	10.2	8.7	5.5	
10	X	1210	1211	48	+	0.7	0.6	0.6	0.0	0.0	0.0	0.4	0.2	3.9	2.2	1.0	0.9	0.6	
11	X	1523	1524	395	+	0.9	0.9	0.4	0.4	1.0	0.8	2.2	3.1	2.1	3.0	2.3	1.6	1.0	
12	AS	1557	1610	107	-	1.6	1.3	0.9	0.1	0.0	0.0	0.7	0.7	0.3	0.9	0.4	0.6	0.7	
13	X	1572	1573	49	+	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
14	X	1655	1656	159	+	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	
15	PreC/C	1745	1746	402	+	0.3	0.3	0.1	0.0	0.3	0.0	1.4	1.0	0.6	0.6	0.1	0.4	0.3	
16	pgRNA	1816	1819	20869	+	281.4	217.2	####	1250.8	705.8	1518.4	5771.8	#####	822.7	1281.0	469.1	1484.7	822.7	
17	AS	1822	1834	21	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.4	0.1	0.0	

Table S6. Peaks for HBV model systems

Genbank ID: GQ358158.1

Peak	Start	End	Total count	Strand	Corresponding peaks in clinical samples
1	25	52	1093	+	Peak #3
2	206	209	1185	+	Peak #5
3	283	284	841	+	Peak #7
4	388	391	1431	+	-
5	759	763	1323	+	Peak #9
6	1004	1005	164	+	-
7	1210	1211	123	+	Peak #10
8	1240	1276	1013	+	-
9	1467	1472	134	+	-
10	1523	1525	275	+	Peak #11
11	1557	1610	112	-	Peak #12
12	1780	1821	178952	+	Peak #16
13	2800	2812	12102	+	Peak #1
14	2822	2824	661	+	-
15	3183	3233	41689	+	Peak #2