

Figure S1. Calculation of common integration sites (CIS) genomic distance. The presence of CIS within a specific genomic window was calculated in our dataset containing 5036 proviral insertions, where the expected fraction (Efr) of three insertions provided an 88kb window size ($p = 0.01$) based on methods described by Mikkers et al. [58].

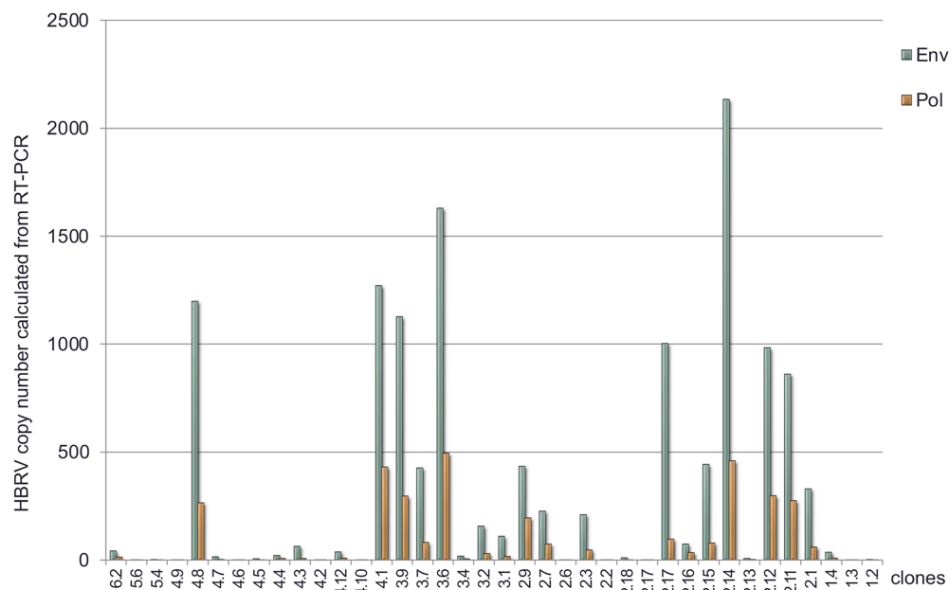


Figure S2. RT-PCR of HBRV RNA in Hs578T cells after the third round of subcloning. Cells with HBRV Env CT values < 28.1 equivalent to a copy number $> 10^3$ per sample were used for virus isolation [calibrated by dilution of RSV:MMTV plasmid].

Table S1. Representative cancer-associated genes at HBRV in vitro integration clusters (visually represented in Figure 4A).

Chromosome location ^a	Unique integration sites (+/-) ^b	First integration on site	Last integration on site	Cluster Window size (bp)	Genes in region	Gene ^c	Sanger ^d	COSMIC ^e	CCG ^f	Cancer Driver (mice) ^{g,[91]}	Relevant cancer type
3q28-29	12 (7/5)	191,671,53	3192,688,053	1,016,700	4	PYDC2 (+)	N	N	N	N	N
						FGF12 (-)	N	N	N	N	N
						MB21D2 (-)	N	Y; Tier 2	N	Y	Liver[82]
						HRASLS (+)	N	N	N	Y	N
5q32	8 (4/4)	147,054,77	7148,096,029	1,041,252	6	JAKMIP2 (-)	N	N	N	N	N
						SPINK1 (-)	N	N	N	N	Liver[74]
						SH3TC2 (-)	N	N	Y	Y	N
						HTR4 (-)	N	N	Y	Y	N
						ADRB2 (+)	N	N	N	N	Breast[92]
8q22.3	7 (3/4)	106,051,24	8106,855,522	803,698	4	LRP12 (-)	N	N	Y	Y	N
						ZFPM2 (-)	N	N	Y	Y	N
						ZFPM2-AS1 (+)	N	NL ^h	N	N	Liver[75], Breast[76]
						OXR1 (+)	N	N	N	N	
8q24.22	11 (5/6)	132,584,51	5133,541,074	956,523	6	ADCY8 (-)	N	N	N	N	Breast[93]
						EFRA3A (+)	N	NL	N	N	N
						OC90 (-)	N	N	N	N	Breast[94]
						HHLA1 (-)	N	N	N	N	N
						KCNQ3 (-)	N	N	N	N	N
						HPYR (-)	N	NL	N	N	N
8q24.22-q24.23	7 (4/3)	135,292,61	5135,846,103	553,542	6	ST3GAL1 (-)	N	N	N	N	N
						ZFAT (-)	N	N	N	N	N
						ZFAT-AS1 (+)	N	NL	N	N	N
						MIR30B (-)	N	N	N	N	Liver[77], Breast[78]
						MIR30D (-)	N	NL	N	N	N
						KHDRBS3 (+)	N	N	N	N	N
9q31.1-q31.2	5 (0/5)	107,351,49	5108,340,002	988,453	9	OR13C8 (+)	N	N	N	N	N
						OR13C5 (-)	N	N	N	N	N
						OR13C2 (-)	N	N	N	N	N
						OR13C9 (-)	N	N	N	N	N
						OR13D1 (+)	N	N	N	N	N
						FSD1L (+)	N	N	N	N	N
						FKTN (+)	N	N	N	N	N

						TAL2 (+)	Y	Y; Tier 1	Y	Y	Blood[95]
9q31.2	7 (1/6)	108,526,910	9,370,774	843,807	3	TMEM38B (+)	N	N	Y	Y	N
11q14.3-q21	6 (1/5)	92,741,519	3,768,781	1,027,264	10	MTNR1B (+)	N	N	N	N	N
						SLC36A4 (-)	N	N	N	N	N
						CCDC67 (+)	N	N	N	N	Liver[96]
						SMCO4 (-)	N	N	N	N	N
						MED17 (+)	N	N	N	N	Breast[97]
						VSTM5 (-)	N	N	N	N	N
						HEPHL1 (+)	N	N	Y	Y	N
						FOLR4 (+)	N	N	N	N	N
Xq11.1	5 (0/5)	585690285	579313	10,285	1	ZXDA (-)	N	N	N	N	N
	9 (0/9)	616822726	1723406	41,134	1	SPIN4 (-)	N	N	N	N	N

a. Genome build: GRCh37.p13; b. Integration orientation: (+) forward, (-) reverse; c. Gene orientation: as in b.; d. Sanger.; e. COSMIC (Cancer Gene Census: Tier 1 – well-documented activity relevant to cancer; Tier 2 – strong indication of a role in cancer); f. CCGD; g. Insertional mutagenesis screen in mice [91]; h. NL – not listed.