Supplementary Note

The potential association and implications of HBV integration at known and putative cancer genes of TERT, MLL4, CCNE1, SENP5, and ROCK1 on tumor development were discussed.
Human telomerase reverse transcriptase (***TERT***), which plays important roles to maintain telomere length and promote viral oncogenesis[30], is the most prevalent gene integrated by the HBV genome in HCC. Of the18 tumor samples with HBV-TERT integration, 15 HBV breakpoints were found in the promoter region, driving the upregulation of *TERT* in these tumor samples (but not in normal liver tissues) as supported by gene expression data (**Fig. 4**) and RNA-Seq data (Supplementary **Fig. 4a**). Overall, *TERT* gene expression is up-regulated by an average of 12.4 fold in the 18 tumor samples in comparison to their respective normal tissues (P=4.8x10^{-11}, paired t-test). The present findings support the previous preliminary observation that HBV inserted in the *TERT* promoter [13,15,16]. In addition, we identified three novel HBV insertion breakpoints in the introns 2 and 6, which also cause increased *TERT* expression level in the affected tumors. Although the detailed regulatory mechanism how HBV intronic insertion causing gene upregulation remains to be defined, it has been previously reported that HBV integration occurred in intron 3 of *TERT* in a HCC cell line SNU449 and upregulated *TERT* expression[16].

Another recurrent event on ***MLL4*** (mixed-lineage leukemia 4) was observed in 9 HCC samples. In addition to the previously reported integration breakpoints in the intron 3[15], we also found HBV integration breakpoints 3 in exon 3 and 1 in exon 6 of *MLL4* (**Fig. 3b**). RNA-Seq also revealed HBV integration on exon 3 and subsequent over-expression of *MLL4* transcripts on samples 70T (but not on 70N) as shown in **Supplementary Fig. 4b.** A total of 48 RNA-Seq PE reads spanning the HBV and the 3rd exon of *MLL4* was detected in the 70T but not in 70N, and the transcript expression level was >5-fold higher in the tumor. Gene expression array data has also shown that *MLL4* is overexpressed by an average of 5.4 fold in the 9 tumors (P=5.6x10^{-5}, paired t-test). The *MLL* gene family contains a unique SET domain that possesses histone H3 lysine 4 (H3K4)-

specific methyltransferase activity and has critical roles in gene activation and epigenetics and potentially regulates mRNA processing[31]. Despite as a potential cancer driver in multiple tumor types through epigenetic gene regulation, its precise role in HCC progression and its nature of viral oncogenesis remain elusive.

HBV integration in *CCNE1* (cyclin E1) gene is a hitherto unidentified event but was found in 4 tumor-specific samples of HCC in this cohort **(Fig. 3c**), which was further confirmed by Sanger sequencing. RNA-Sequencing analysis further validated this novel HBV integration site and identified the transcript where HBV gene fused with the last exon of *CCNE1* in sample #200T (**Supplementary Fig. 4c**). Significantly increased expression of *CCNE1* transcript is also observed in the 4 tumors when compared with their respective non-tumor tissues (P=6.8x10^{-4}, paired t-test) by an average of 29.7 fold. Overexpression of CCNE1 was also observed in breast cancer and knockdown of CCNE1 shown to suppress tumor development in a breast cancer mouse model[32]. Cyclin E1 protein plays an important role in oncogenesis and its over-expression is associated with shorter disease-free survival[33]. The human cyclins are mainly involved in regulating cell cycle events in all eukaryotic cells and are major targets for oncogenic signals[34]. Recently, HBV integration within an intron of cyclin A gene has been reported in early stage of a liver tumor potentially disrupting cell cycle division and causing tumorigenesis[35]. Therefore, HBV integration at the *CCNE1* gene locus has provided an alternative molecular mechanism driving aberrant cell cycle control in HCC development and progression.

For SENP5, three HBV integration events are discovered occurring in three tumor samples, and strikingly all occur in intron2 validated by Sanger sequencing. SENP5, which belongs to SUMO-specific protease family (SENP1-8) that mediates protein degradation, is overexpressed in oral squamous cell carcinomas and associated with cell differentiation[36].

For ROCK1, 2 HBV integrations are discovered occurring in two tumor samples, one in the promoter region while another one in the first intron. ROCK1 is a serine/threonine protein kinase that regulates focal adhesion pathway and cell mobility, and activation of the Rho/ROCK pathway has been associated with more aggressive tumor properties such as metastasis and invasion[37,38].

Supplementary references:

30. Horikawa, I. & Barrett, J.C. *cis*-Activation of the human telomerase gene (hTERT) by the hepatitis B virus genome. *J. Natl. Cancer Inst.* **93**, 1171–1173 (2001).

31. Ansari, K.I. & Mandal, S.S. Mixed lineage leukemia: roles in gene expression, hormone signaling and mRNA processing. *FEBS J.* **277**, 1790–1804 (2010).

32. Liang, Y. *et al.* siRNA-based targeting of cyclin E overexpression inhibits breast cancer cell growth and suppresses tumor development in breast cancer mouse model. *PLoS ONE* **5**, e12860 (2010).

33. Nakayama, N. *et al.* Gene amplification *CCNE1* is related to poor survival and potential therapeutic target in ovarian cancer. *Cancer* **116**, 2621–2634 (2010).

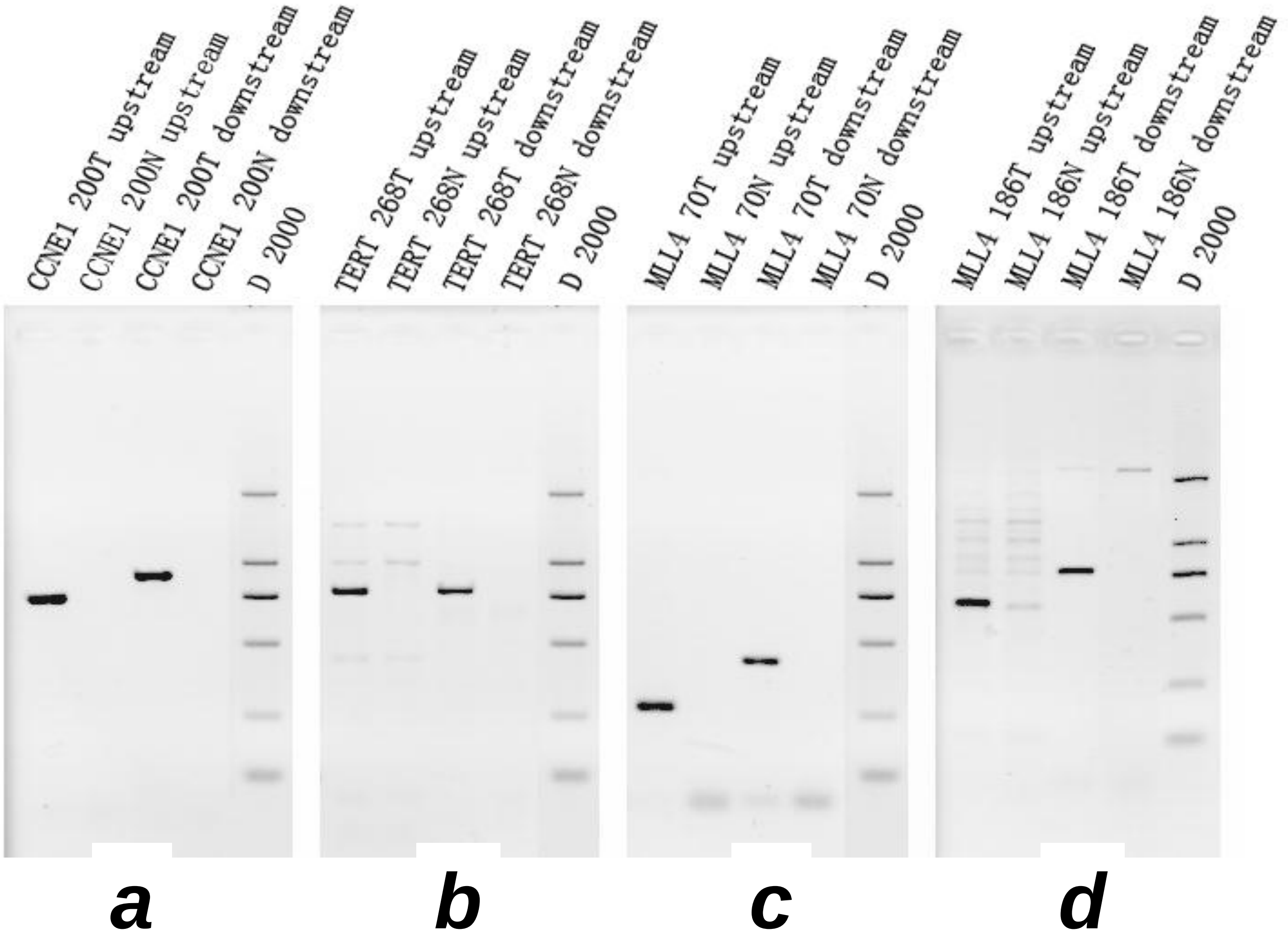
34. Pagano, M., Pepperkok, R., Verde, F., Ansorge, W. & Draetta, G. Cyclin A is required at two points in the human cell cycle. *EMBO J.* **11**, 961–971 (1992).

35. Wang, J., Chenivesse, X., Henglein, B. & Brechot, C. Hepatitis B virus integration in a cyclin A gene in a hepatocellular carcinoma. *Nature* **343**, 555–557 (1990).

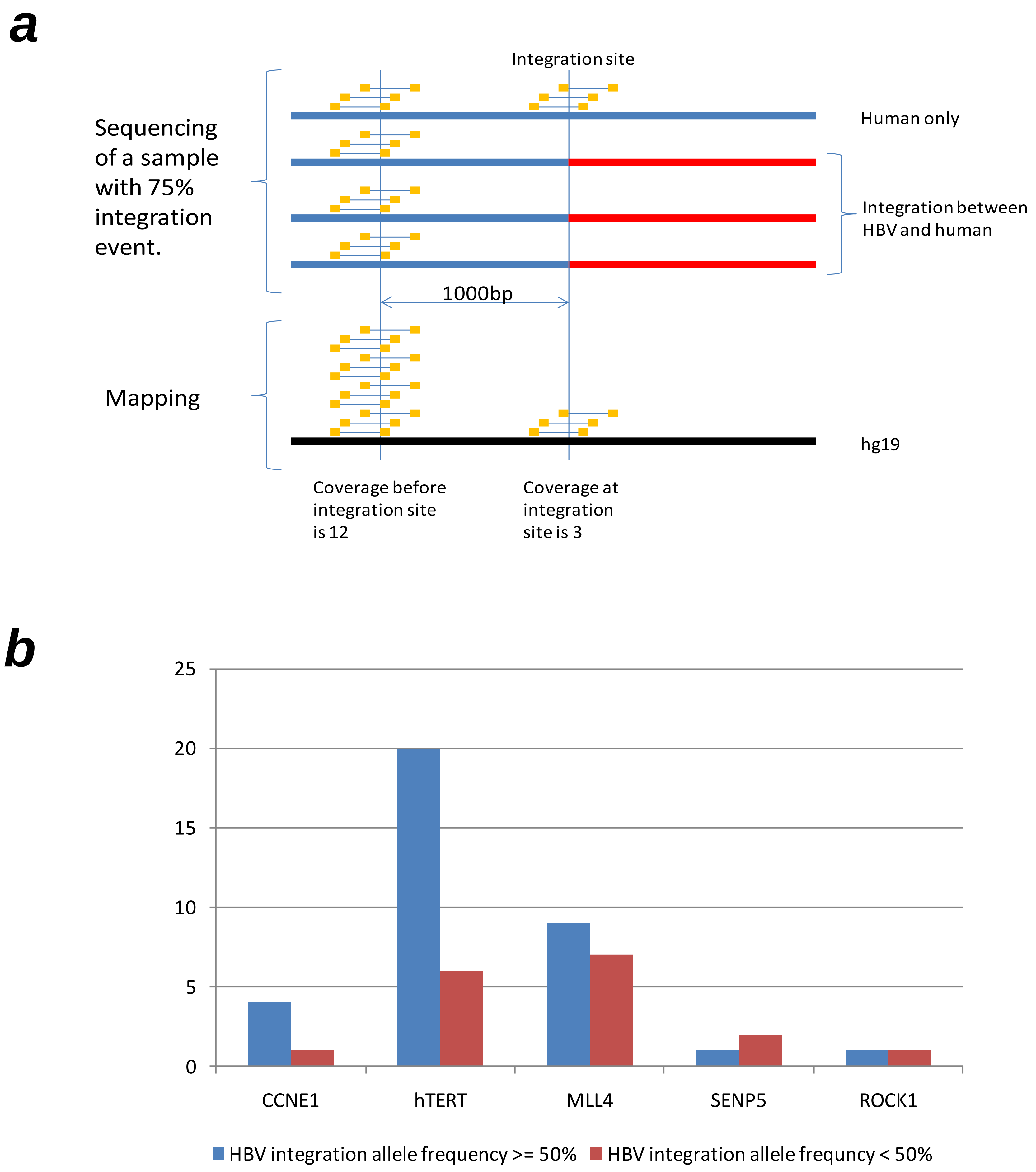
36. Ding, X. *et al.* Overexpression of SENP5 in oral squamous cell carcinoma and its association with differentiation. *Oncol. Rep.* **20**, 1041–1045 (2008).

37. Wilkinson, S., Paterson, H.F. & Marshall, C.J. Cdc42-MRCK and Rho-ROCK signalling cooperate in myosin phosphorylation and cell invasion. *Nat. Cell Biol.* **7**, 255–261 (2005).

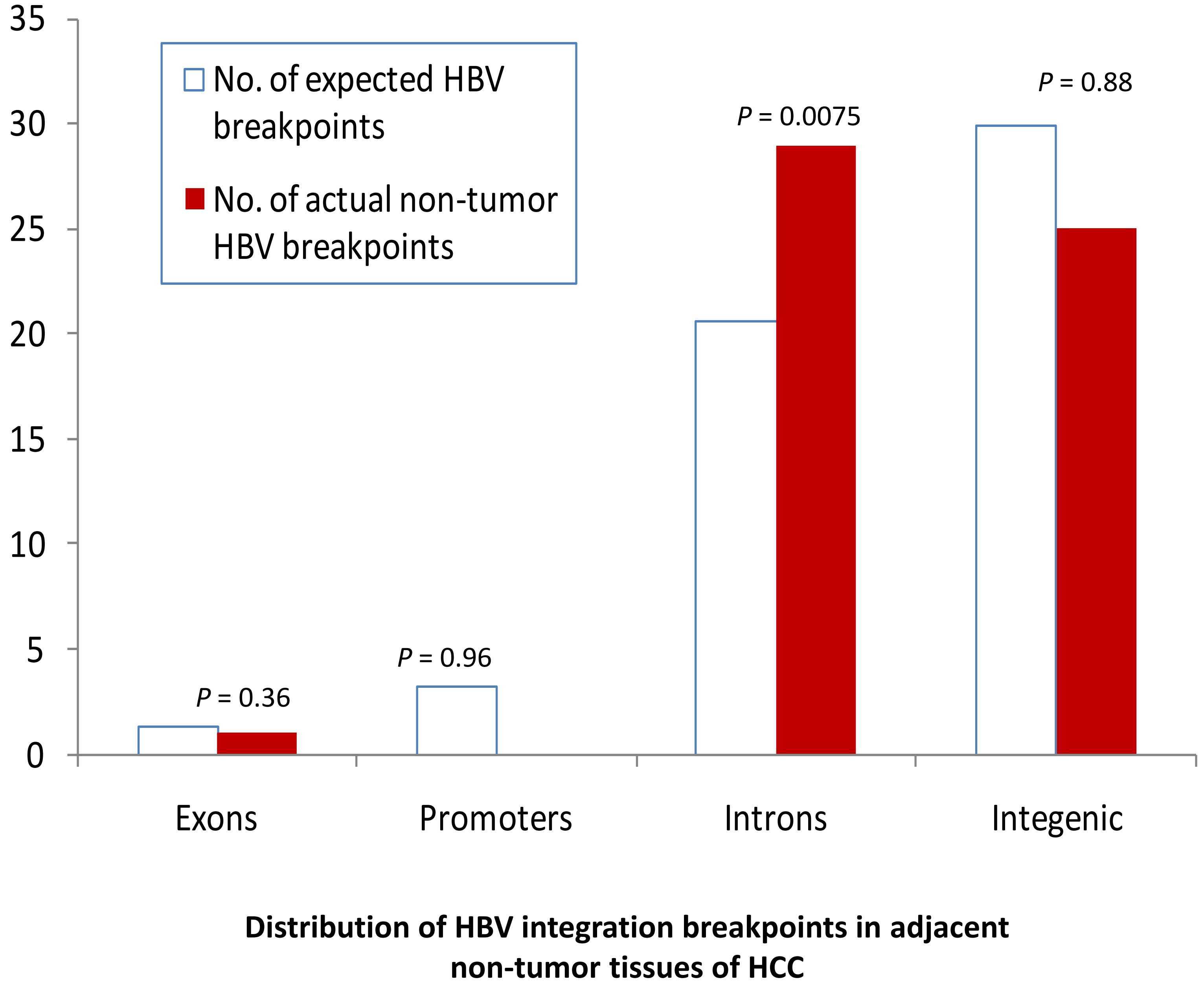
38. Wong, C.C. *et al.* Deleted in liver cancer 1 (DLC1) negatively regulates Rho/ROCK/MLC pathway in hepatocellular carcinoma. *PLoS ONE* **3**, e2779 (2008).



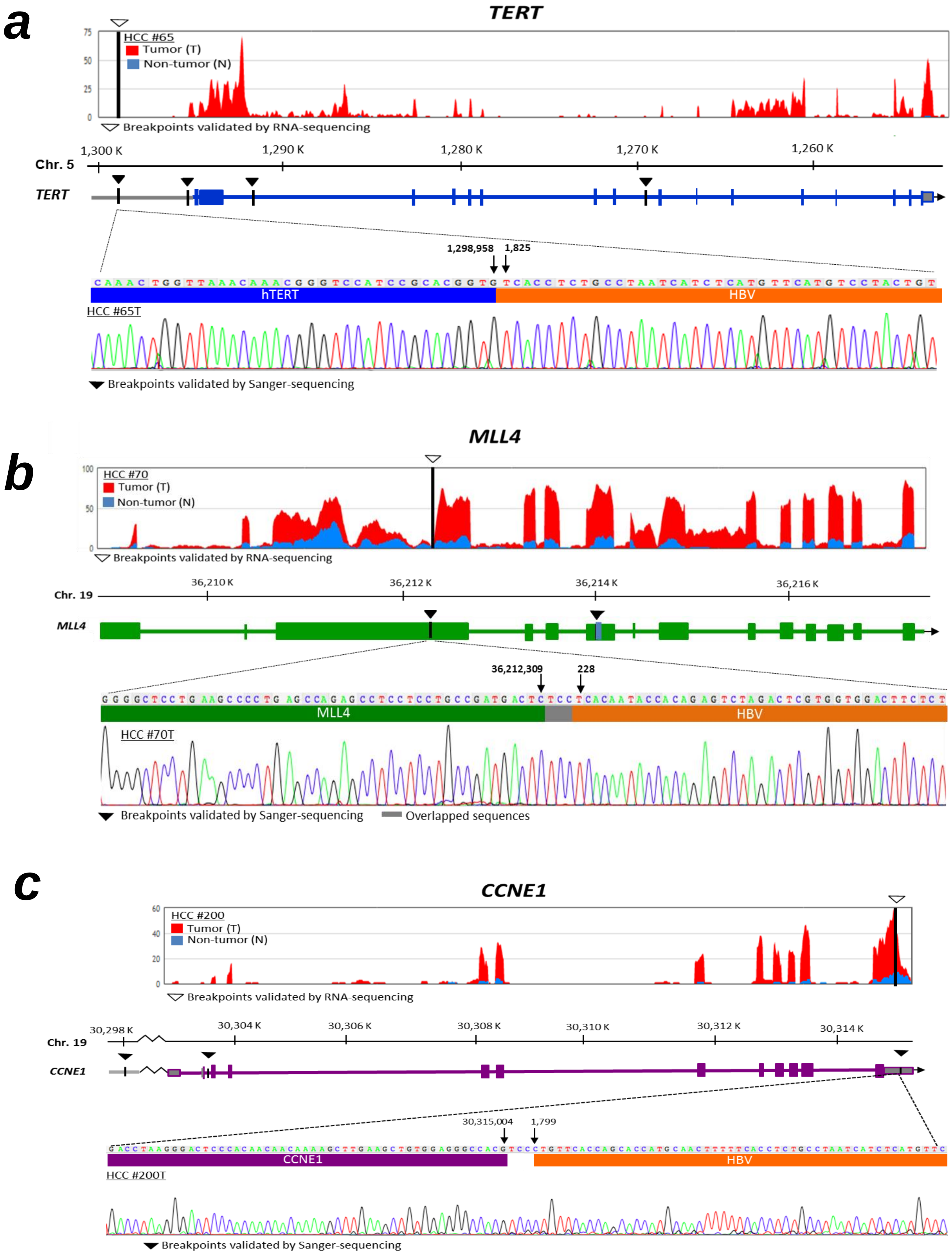
Supplementary Fig. 1. Validating that the HBV breakpoints are somatic. Eight HBV breakpoints appearing on 3 recurrent genes (*CCNE1*, *TERT* and *MLL4*) are selected for validation in both tumors and the adjacent non-tumors. The gel shows that all breakpoints appear in the tumor and are not detectable in the matched adjacent non-tumor. samples Hence, they are considered as somatic events.



Supplementary Fig. 2. Illustration of HBV integration allele frequency in tumor sample. *(a) An example illustrates how to compute the HBV integration allele frequency for a HBV breakpoint in the tumor sample. For the above example, the HBV integration allele frequency is (12-3)/12=75%.; (b) The number of HBV integrations in CCNE1, TERT, MLL4, SENP5, and ROCK1 whose HBV integration allele frequency ≥ 0.5.*



Supplementary Fig. 3. Histogram of the HBV integration breakpoints in non-tumor tissues according to the locations. The open bar shows the expected number of HBV integration breakpoints, assuming the breakpoints are uniformly randomly distributed. he red bar shows the actual number of non-tumor HBV breakpoints. The P values are computed assuming the binomial distribution.

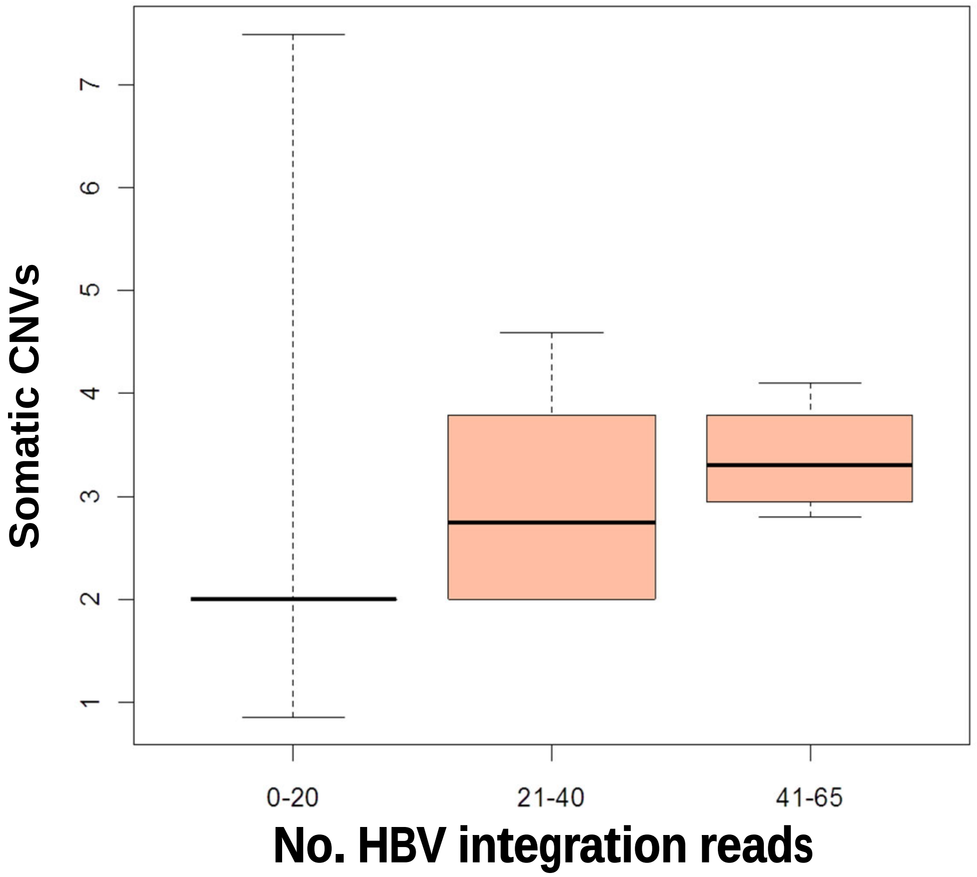


Supplementary Fig. 4. RNA-Seq data analysis (upper panel) and Sanger sequencing (lower panel) validation of selected recurrent HBV integrated genes: *(a) TERT*; *(b) MLL4*; and *(c) CCNE1*.

(a) RNA-Seq expression of *TERT* in 65T (tumor) and 65N (normal). The triangle (∇) and vertical line indicate the HBV breakpoint location (promoter) of the in 65T. It is shown that *TERT* is over-expressed in 65T (red) but not in 65N (blue). Lower panel indicates the Sanger sequencing result of the *TERT*-HBV chimera in HCC sample 65T.

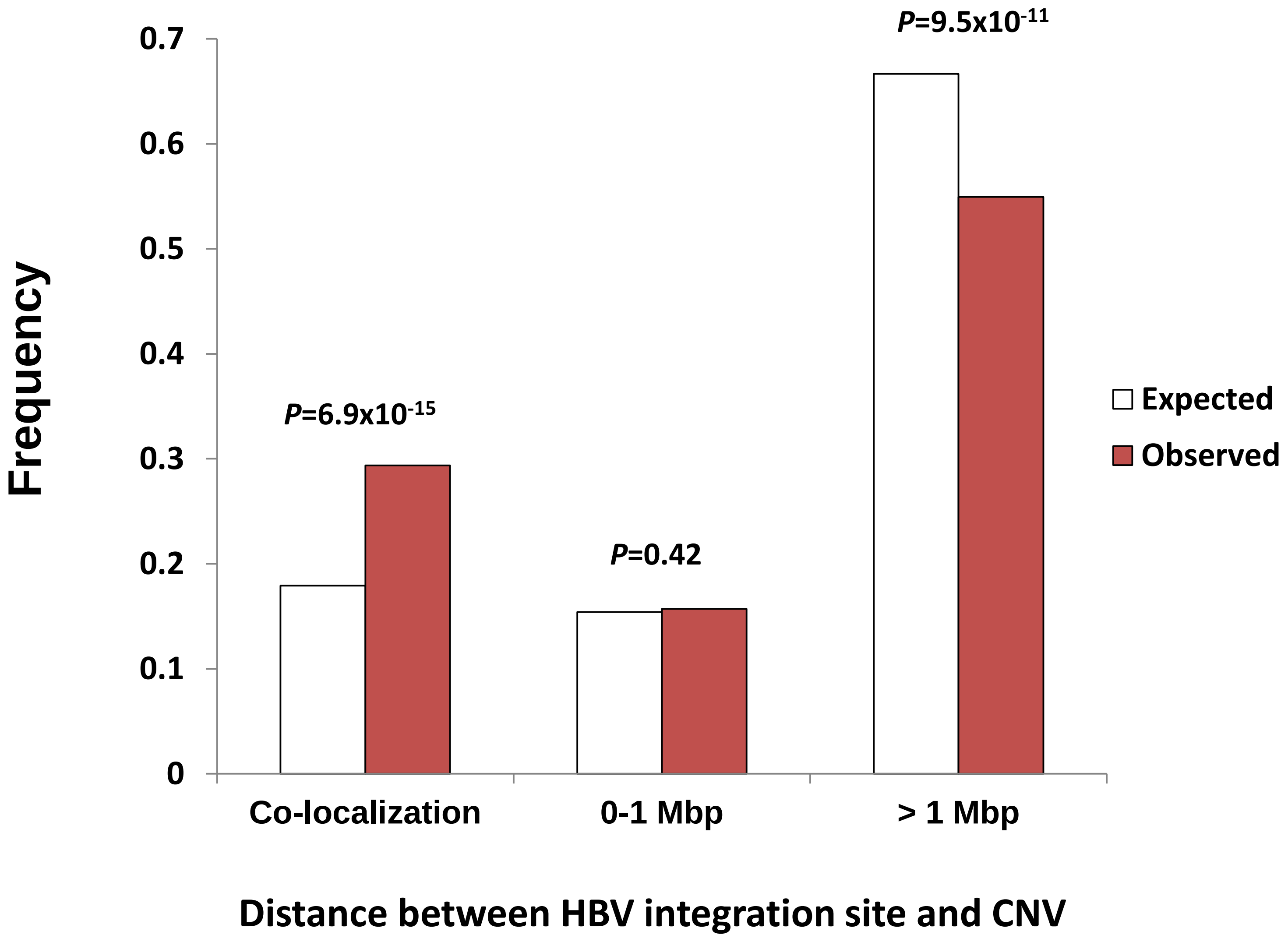
(b) RNA-Seq expression of *MLL4* in 70T and 70N. The vertical line indicates the HBV integration site in 70T (which is on the 3rd exon). RNA read signals were higher in the tumor compared with the normal. Sanger sequencing result shows the HBV fuses in frame with the *MLL4* gene in HCC sample 70T.

(c) RNA-Seq expression of *CCNE1* in 200T and 200N. The vertical line indicates the HBV integration site in 200T (which is on the last exon). Expression of signal in 200T is higher than the 200N. Sanger sequencing in sample 200T also reveals the chimeric HBV-*CCNE1* fusion transcript.



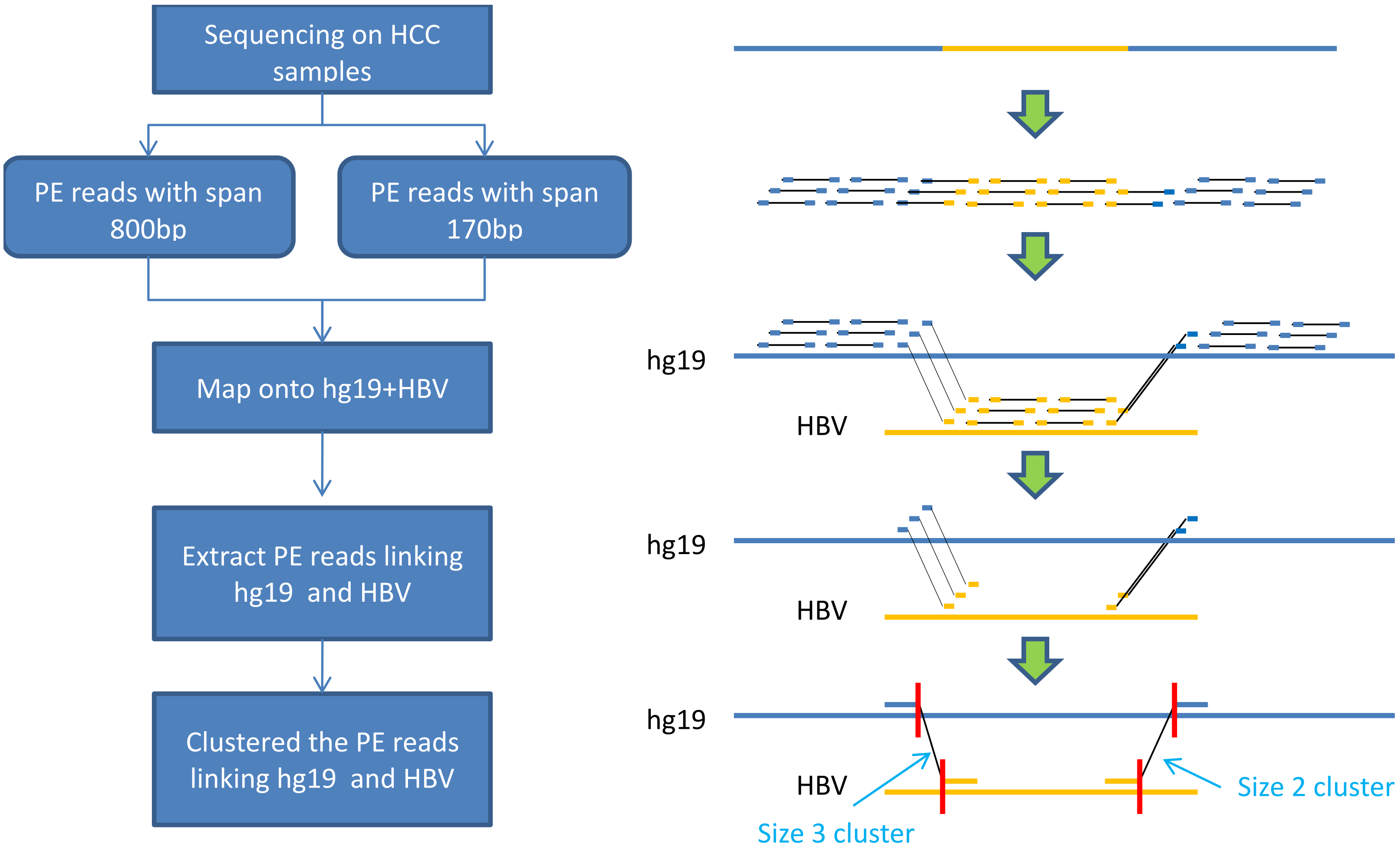
Supplementary Fig. 5. Relationship between the somatic copy number (sCNV) and the number of HBV integration reads. The boxplot compares the copy number around HBV breakpoints with (a) 0-20, (b) 21-40, and (c) 41-60 read support.

Nature Genetics: doi:10.1038/ng.2295

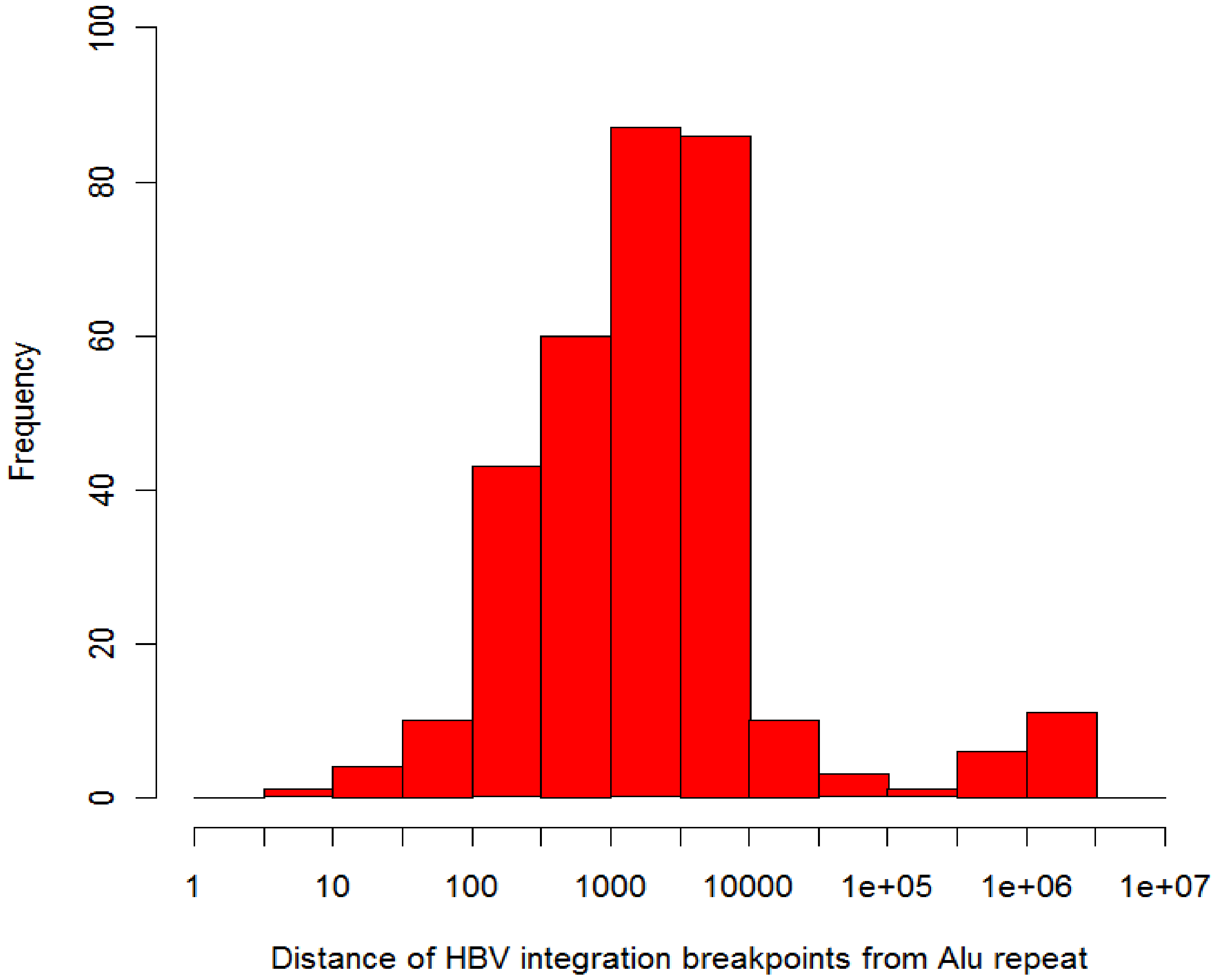


Supplementary Fig. 6. Histogram showing the distance of the HBV integration breakpoints from the CNV regions. The open bar shows the expected number of HBV integration breakpoints co-localize, within 1million and more than 1 million from CNV regions. The red bar shows the actual number of non-tumor HBV breakpoints.

Nature Genetics: doi:10.1038/ng.2295



Supplementary Fig. 7. Workflow on processing the sequencing data. The flowchart on the left shows the flow of our method to identify the HBV integration breakpoints. On the right, we show an example to illustrate the whole process.



Supplementary Fig. 8. Distribution of HBV integration breakpoints versus *Alu repeat.* The histogram shows the distribution of the distance of the HBV integration breakpoints from the Alu repeat.

Supplementary Table 1: Demographic and clinicopathologic characteristics of 88 HCC patients

Variable Name	Mean±SD/Median or %
Male	78.4%
Age (year)	56.1 ± 11.2/56.5
HBsAg status	
Positive	92.0%
Negative	8.0%
Liver function parameters	
AFP [log$_{10}$] (ng/mL)	2.2 ± 1.4/1.8
SGPT (U/L)	61.4 ± 56.0/46.5
SGOT (U/L)	62.4 ± 52.8/47
BILIRUBIN (μM)	15.6 ± 24.8/12
Tumor size (cm)	7.2 ± 3.9/6
Tumor recurrence	58.0%
Non-tumorous liver histology	
Cirrhotic	66.3%
Non-cirrhotic	7.2%
Chronic hepatitis	26.5%
Child's grade	
A	96.3%
B	3.7%
Edmondson grade	
Poorly differentiated	25.0%
Moderately differentiated	59.2%
Well differentiated	15.8%
AJCC stage	
I	41.0%
II	27.7%
IIIA/B	31.3%
Number of tumor nodule(s)	
Single	71.6%
Multiple ≥ 2	28.4%
Survival (month)	58.4 ± 34.3/58.2
Event	
Deceased	38.6%
Censored	61.4%

† All analyses were performed based on available data (last updated on Aug 8, 2011).