HBV DNA Integration into Telomerase or MLL4 Genes and TERT Promoter Point Mutation as Three Independent Signatures in Subgrouping HBV-related HCC with distinct features

Chiao-Ling Li, Chia-Lang Hsu, You-Yu Lin, Ming-Chih Ho, Ray-Heng Hu, Chi-Ling Chen, Tung-Ching Ho, Yung-Feng Lin, Shih-Feng Tsai, Sheng-Tai Tzeng, Chin-Fang Huang, Ya-Chun Wang, Shiou-Hwei Yeh, Pei-Jer Chen

Supplementary Information

Supplementary Figures Legends

Supplementary Figure 1. Comparison of the identified HBV-human DNA junctions in HBV-HCC tumor tissue by WGS and capture-NGS (N=12). WGS generated an average of 1,766,049,526 reads and 236,640,318,450 mapped bases per 60G data set for each sample. Major junctions with high read counts meet the threshold were defined as clonal junctions. The capture-NGS platform resulted in an average of 2,639,118 126 bp high-quality reads for each sample. The X-axis indicates the position of HBV-human DNA junctions in human chromosome; the upper Y-axis indicates junction clonality calculated from WGS results, and the lower X-axis indicates the number of junction reads detected by capture-NGS; red bars indicate junctions detected and defined as clonal by both WGS and capture-NGS, purple bars indicate clonal junctions detected by either WGS or capture-NGS, and gray bars indicate non-clonal junctions detected by WGS or capture-NGS. For non-clonal junctions, only those with matching clonal junction in its counterpart data set are shown.

Supplementary Figure 2. Analysis of the Viral Integration Site DataBase (VISDB) revealed that *TERT* and *MLL4* were the most frequently recurring HBV integration genes in HCC. **a** The number and annotation status of HBV integration sites identified in HCC from 35 studies included in the VISDB [1], which can help identify the recurrent integrations found in different studies. **b** List of the five viral integrated genes (hits with gene annotation) supported by >5 studies, with frequency information. **c** Schematic representation of HBV integrated genes in the VISDB. X-axis indicates the number of integrations (hit with gene annotation) for specific genes. Y-axis indicates the number of studies that document the integration of specific genes. The HBV integrated genes supported by >5 studies were marked with red color.

Supplementary Figure 3. Mutation rate of less common somatic mutations in HCC detected by RNA-seq. A total of 80 genes extracted from 6 HCC WGS studies were examined for detection of COSMIC-reported somatic mutations. Only genes with mutations detected are shown in the figure. The 80 genes including *ACVR2A*, *ADCY2*, *ADH1B*, *ADRA1A*, *ALB*, *APC*, *APOB*, *ARID1A*, *ARID1B*, *ARID2*, *ASH1L*, *ASPM*, *ATM, ATP10B, ATRX, AXIN1, AXIN2, BAP1, BRD7, CACNA2D4, CCND1, CCNE1, CDKN1A, CDKN2A, CDKN2B, COL11A1, CPS1, CYP2E1, DOCK2, DST, EPS15, ESRRG, FAM5C, FAT4, FCRL1, FGF19, FRAS1, G6PC, GALN11, HNF1A, HNF4A, HNRNPA2B1, JAK1, KEAP1, KMT2B, KMT2D, LRP1B, MACF1, MALAT1, MAP1B, MAP2K3, MEN1, MLL, MLL3, NAV3, NCOR1, NEAT1, NF32L2, NFE2L2, NSMCE2, PEG3, PIK3CA, PRL, PTEN, PTPRB,*

RB1, RP6SKA3, RPL22, SETDB1, SLC10A1, SRCAP, SYNE2, TBL1XR1, TMEM99, TSC1, TSC2, USP34, VEGFA, WDTC1, ZNRF3.

Supplementary Figure 4. Recurrence-free survival (RFS) and overall survival (OS) stratified by HBV genotype B/C and different HBV-HCC subgroups. **a-e** RFS analysis stratified by HBV genotype in **a** all HBV-HCCs, **b** G1 tumor, **c** G2 tumor, **d** G3 tumor, and **e** G4 tumor. **f**-**j** OS analysis stratified by HBV genotype in **f** all HBV-HCCs, **g** G1 tumor, **h** G2 tumor, **i** G3 tumor, and **j** G4 tumor.

Supplementary Figure 5. Global expression profile of 48 HBV-HCCs. **a** Heatmap representation of the whole transcriptome of all samples using unsupervised hierarchical clustering. **b** Projection of the whole transcriptome of all samples using PCA. **c** Significantly enriched gene sets in G2- and G3-specific genes.

Supplementary Figure 6. Telomerase activity is higher in HBV-HCC with HBV-*TERT* (G1) and *TERT* promoter somatic mutations (G3) than G2 and G4 HBV-HCC. **a** Telomerase activity scores for G1-G4 HBV-HCC, estimated by the EXTEND algorithm. **b** Correlation between telomerase activity scores and TERT expression levels as determined by qRT-PCR.

Supplementary Figure 7. Telomerase downstream regulation is highly correlated with *TERT* expression in G3 but not G1. **a** GSEA of gene sets for *TERT*-regulated immortalization-related genes in G3 (left) or G1 (right) cells relative to G2+G4 cells. **b-c** Gene set variation analysis (GSVA) enrichment score in **b** telomerase pathway. **c** telomere maintenance. **d** Telomere length correlates with TERT expression level in G3. Correlation with TERT expression was evaluated by Spearman rank correlation.

Supplementary Figure 8. Significantly enriched hallmark gene-sets (FDR < 0.05) in G1 (blue) and G3 (yellow) HCC subgroups.

Supplementary Figure 9. Transcriptome analysis revealed activation of distinct signaling pathways in HBV-HCC subgroups.

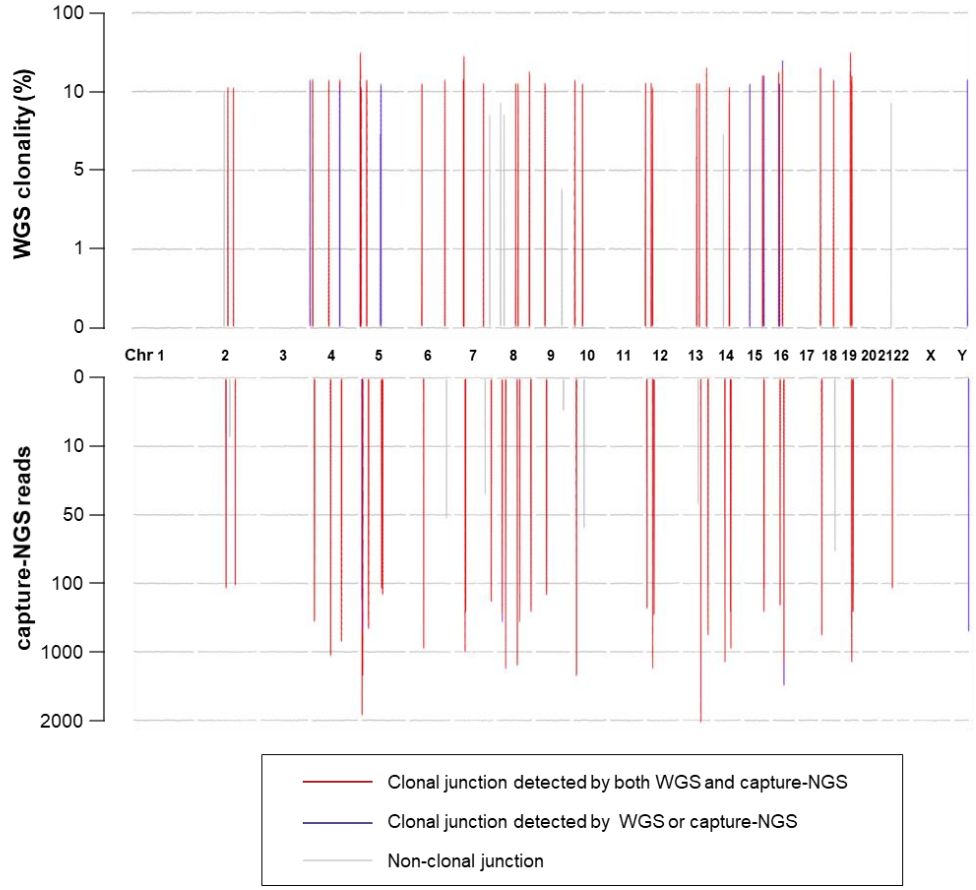
Supplementary Figure 10. The tumor immune microenvironment of different HBV-HCC subgroups. **a** Tumor-infiltrating immune cell composition characterized by CIBERSORT. **b** Significance of specific types of immune cell gene expression defined by CIBERSORT among HBV-HCC subgroups. **c** Significance of specific types of immune cell gene expression defined by Danaher et al. among HBV-HCC subgroups. **d** Dendritic cell-related gene expression score in G1-G4.

Supplementary Figure 11. Immune cell infiltration and expression of ICI response relevant markers are more abundant in G2 HCC. **a** Representative images of CD45 (leukocyte common antigen) IHC staining for immune cell infiltration status in HCC, grading with negative, weak, moderate and strong expression. The patients were divided into low expression group (negative and weak) and high expression group (moderate and strong). **b** Comparison of the CD45 expression levels in each group of HCC. **c** Representative images of the IHC staining for ICI response-relevant markers, CD8 and PD-1, in HCC, grading with negative, weak, moderate and strong expression. The patients were divided into low expression group (negative and weak) and high expression group (moderate and strong). **d-e** Expression of **d** CD8 and **e** PD-1 in each group of HCC. **f** Expression pattern of CD8 and PD-1 in individual HCC undergoing IHC staining. **a,c** Images are shown at 20x magnification. Scale bars = 100 μM. NA, not available. **b,d,e** Differences in specific markers between G2 and non-G2 HCC were estimated by Fisher's Exact test.

Supplementary Figure 12. Features of HBV-HCCs in validation cohort (N=112) divided by hotspot HBV integrations and *TERT* promoter mutation. Features were categorized into panels including genetic changes in host genome, clinical features, and integrated HBV. The P-value is the results of comparison for each parameter between the four groups.

Supplementary Figures

Supplementary Figure 1.



Supplementary Figure 2.

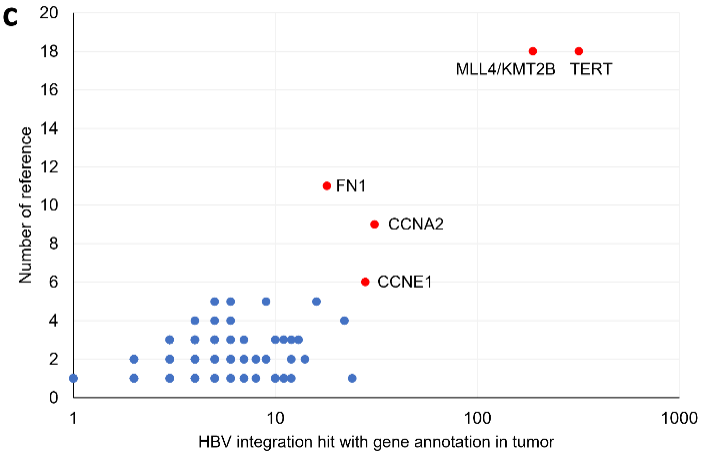
a

HBV Integration collected in VISDB	
Studies report HBV integration in tumor tissue	35
Hit without gene annotation	6413
Hit with gene annotation	3830
Annotated gene	2026

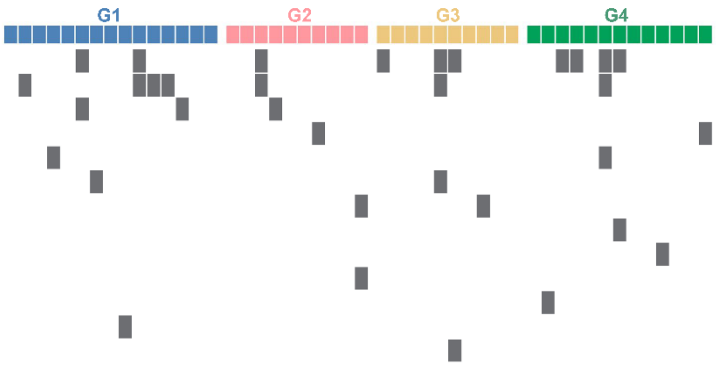
b

Target gene	Hit with gene annotation	Number of studies	Frequency (in hit with gene annotation)
TERT	319	18	8.3%
KMT2B/MLL4	189	18	4.9%
FN1	18	11	0.5%
CCNA2	31	9	0.8%
CCNE1	28	6	0.7%

c

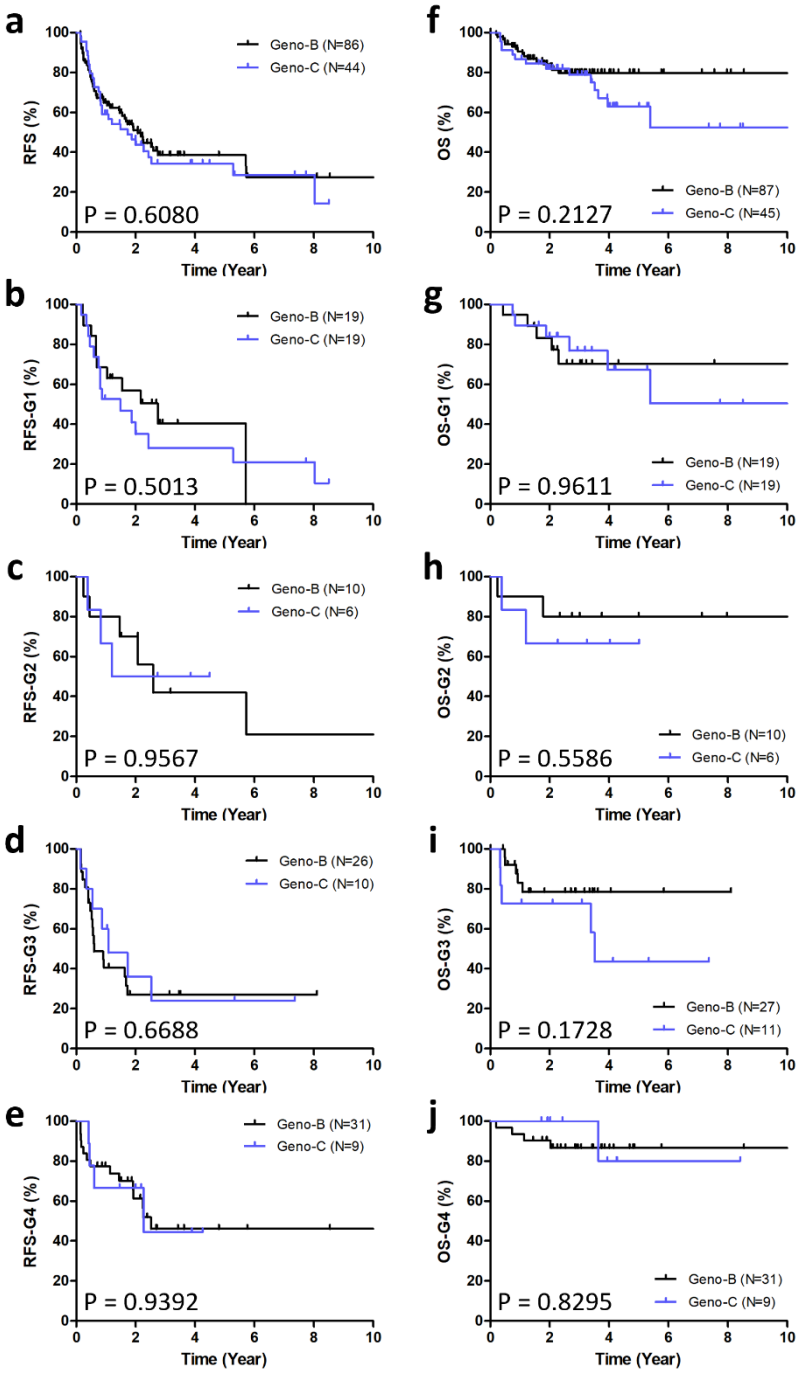


Supplementary Figure 3.

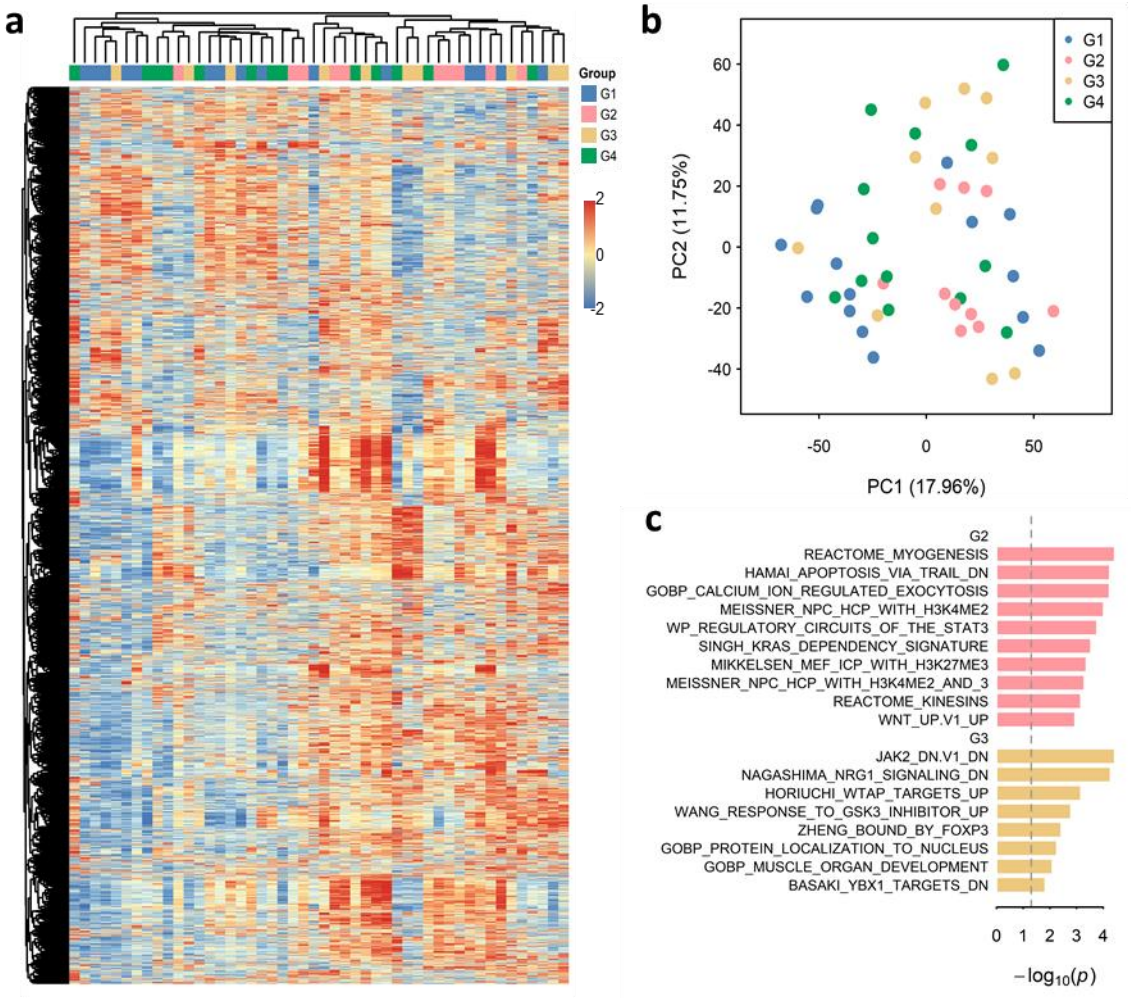


	Mutation rate				
	All	G1	G2	G3	G4
SETDB1	21%	13%	10%	30%	31%
APOB	15%	27%	10%	10%	8%
TSC2	6%	13%	10%	0%	0%
HNF4A	4%	0%	10%	0%	8%
DST	4%	7%	0%	0%	8%
ARID1B	4%	7%	0%	10%	0%
CDKN2A	4%	0%	10%	10%	0%
HNF1A	2%	0%	0%	0%	8%
AXIN1	2%	0%	0%	0%	8%
SRCAP	2%	0%	10%	0%	0%
MEN1	2%	7%	0%	0%	0%
MACF1	2%	7%	0%	0%	0%
ATM	2%	0%	0%	10%	0%

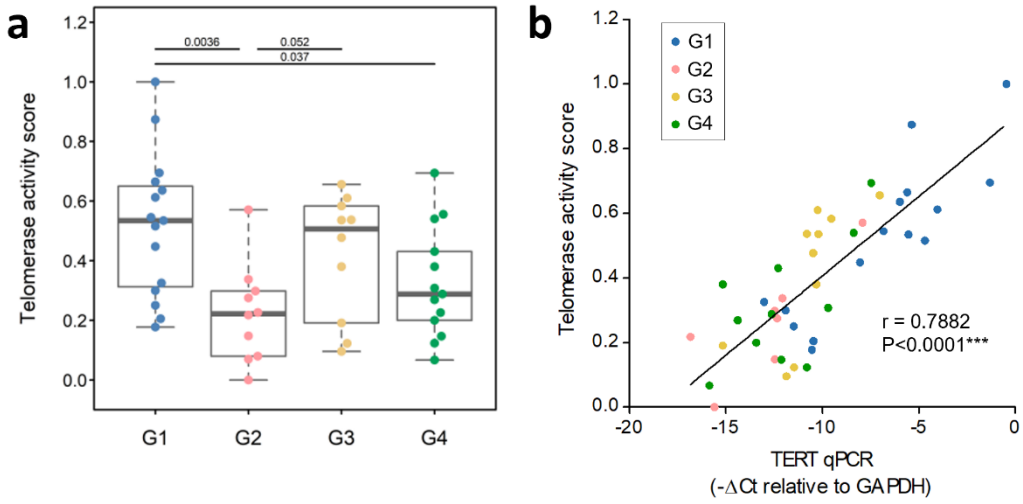
Supplementary Figure 4.



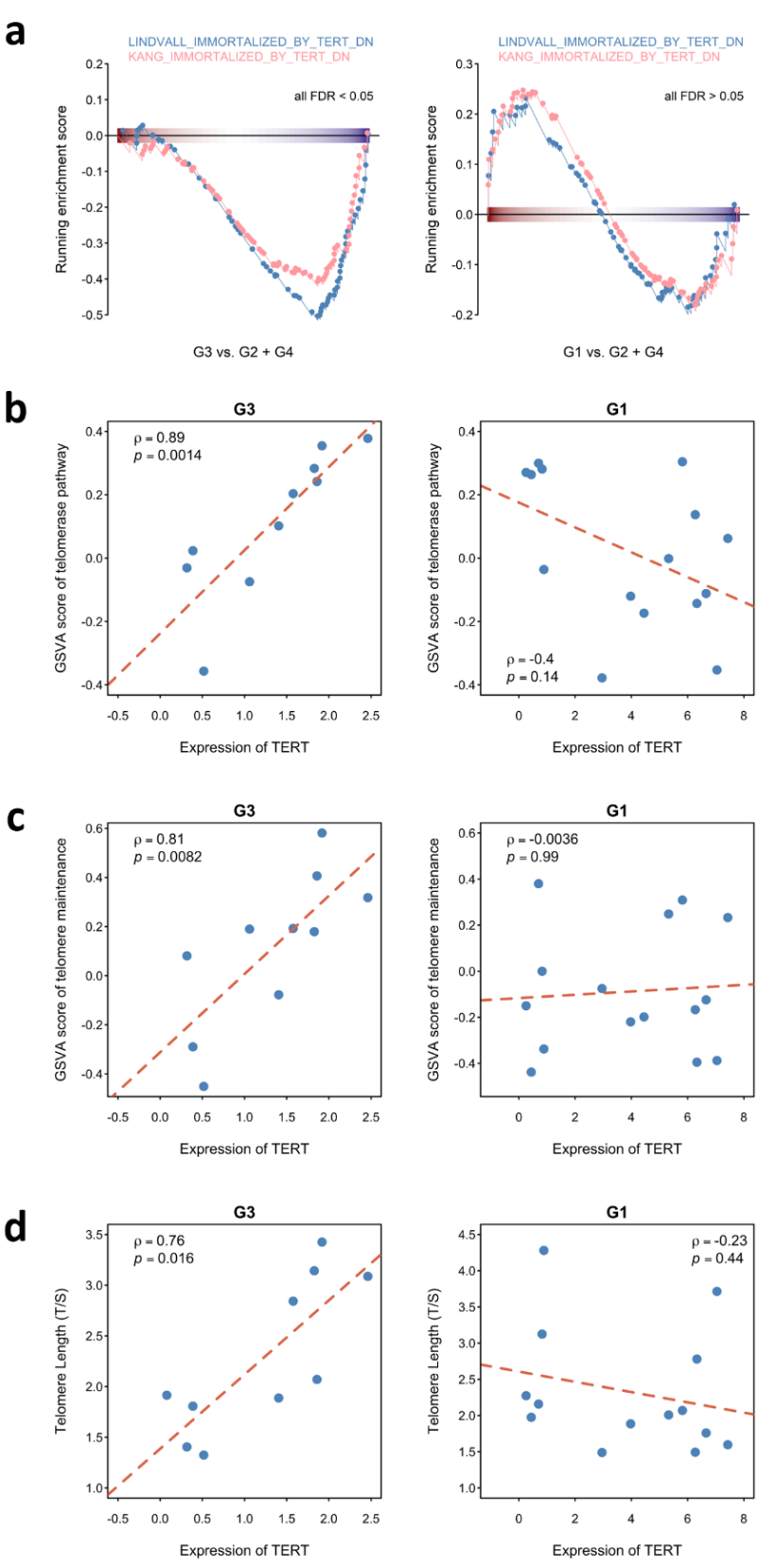
Supplementary Figure 5.



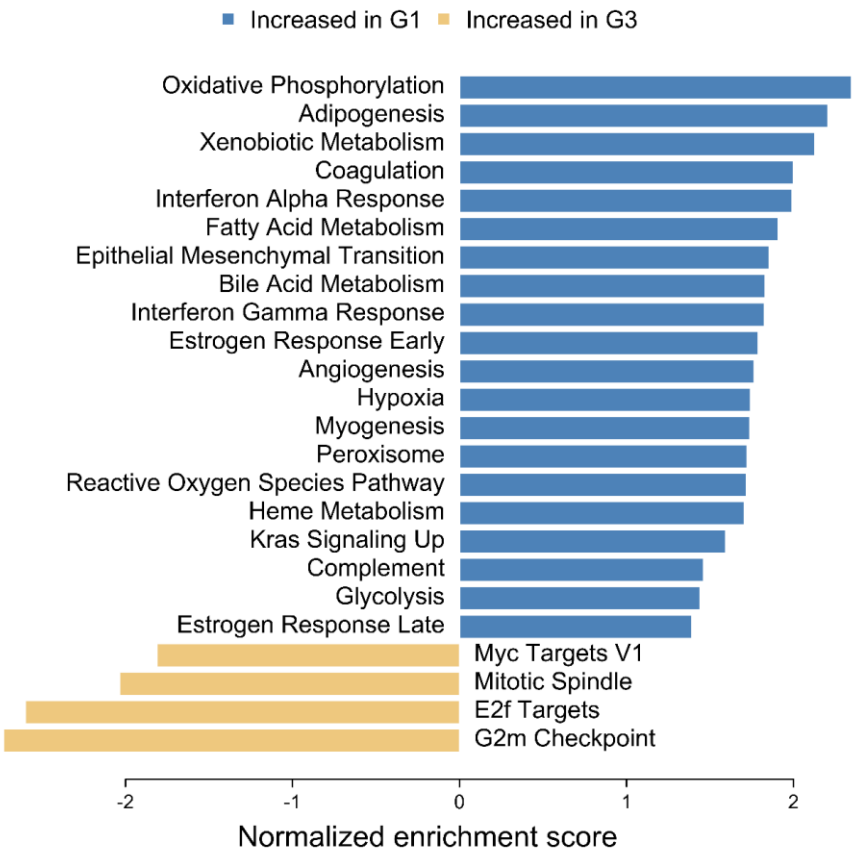
Supplementary Figure 6.



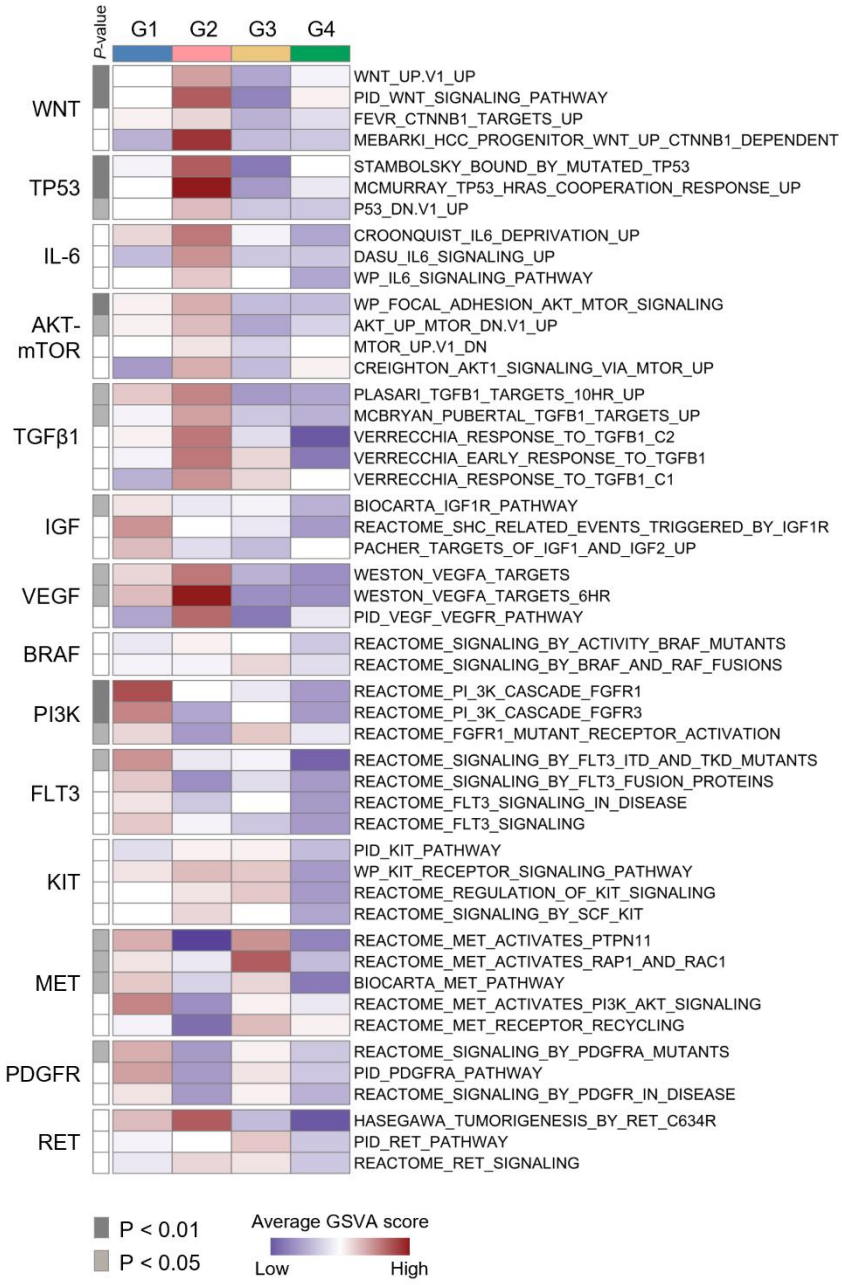
Supplementary Figure 7.



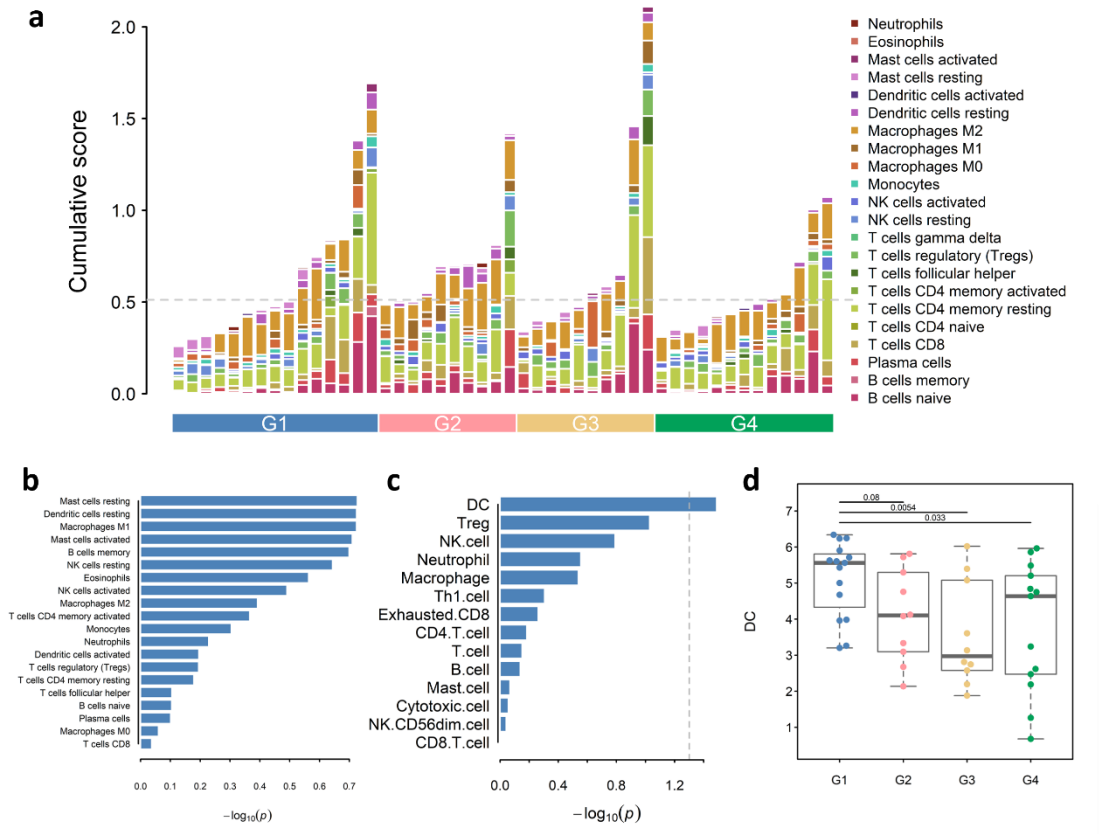
Supplementary Figure 8.



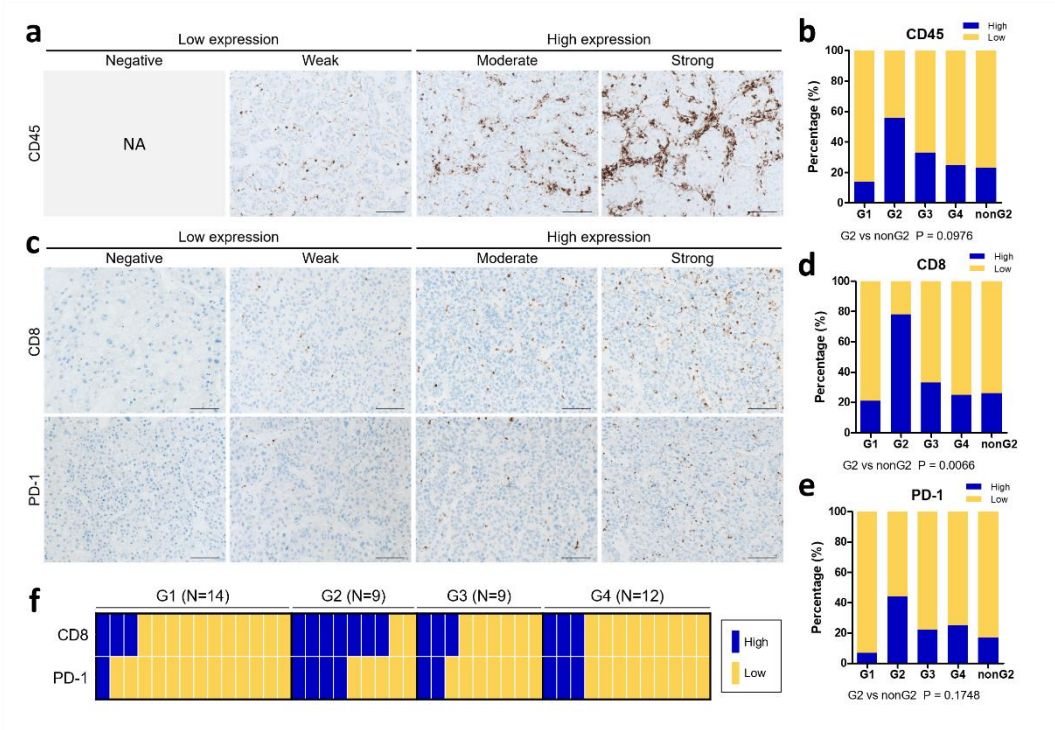
Supplementary Figure 9.



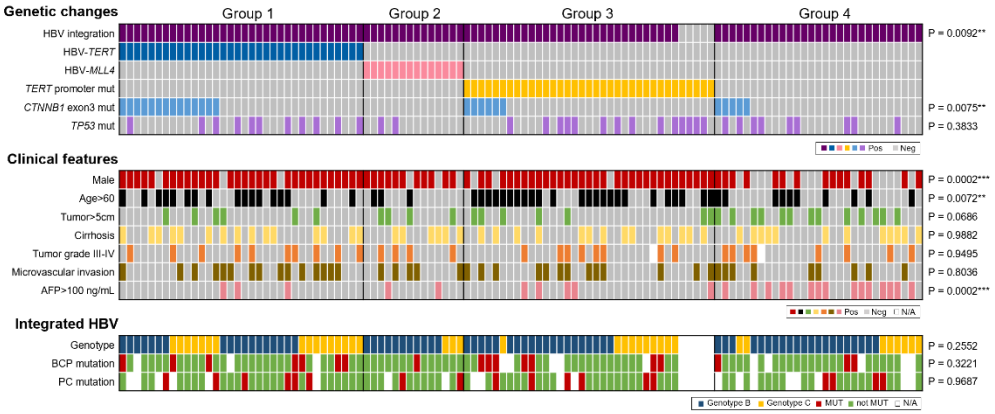
Supplementary Figure 10.



Supplementary Figure 11.



Supplementary Figure 12.



Supplementary Tables

Supplementary Table 1. Logistic regression model for variables in predicting 2-year recurrence of HBV-HCC.

					Multivariate logistic regression											
End Point and Variables	**Univariate logistic regression**				**Group1**				**Group2**				**Group3**				**Group4**			
Recurrence in 2yr	**OR**	**95% CI**	**P-value**		**OR**	**(95% CI)**	**P-value**		**OR**	**(95% CI)**	**P-value**		**OR**	**(95% CI)**	**P-value**		**OR**	**(95% CI)**	**P-value**	
G1 vs non-G1	1.16	0.55-2.42	0.699		1.09	0.46-2.61	0.838													
G2 vs non-G2	0.64	0.22-1.86	0.409						0.60	0.18-2.02	0.410									
G3 vs non-G3	2.12	1.03-4.37	0.042	*									2.35	0.97-5.68	0.058					
G4 vs non-G4	0.54	0.27-1.07	0.078														0.56	0.25-1.26	0.163	
Age>60 vs <=60	1.08	0.56-2.07	0.822		1.30	0.62-2.73	0.490		1.26	0.60-2.66	0.543		1.11	0.51-2.39	0.794		1.23	0.58-2.60	0.589	
Male vs Female	1.15	0.55-2.41	0.714		1.39	0.59-3.28	0.457		1.40	0.59-3.31	0.448		1.14	0.47-2.79	0.770		1.22	0.50-2.95	0.667	
Tumor size >5cm vs <=5cm	3.6	1.83-7.10	<0.001	***	3.77	1.61-8.84	0.002	**	3.76	1.60-8.83	0.002	**	4.17	1.75-9.93	0.001	**	4.05	1.70-9.64	0.002	**
Microinvasion Y vs N	1.79	0.91-3.54	0.092		2.04	0.93-4.49	0.075		2.16	0.97-4.82	0.059		1.99	0.90-4.39	0.090		1.90	0.86-4.20	0.114	
Cirrhosis Y vs N	2.55	1.29-5.03	0.007	**	1.99	0.89-4.42	0.093		1.90	0.84-4.27	0.121		1.76	0.78-3.97	0.177		1.91	0.85-4.27	0.116	
Tumor grade 34 vs 12	4.1	1.99-8.45	<0.001	***	2.04	0.93-4.44	0.074		2.03	0.94-4.38	0.071		2.29	1.04-5.03	0.039	*	2.05	0.95-4.43	0.069	
AFP >100 ng/mL vs <=100ng/mL	1.95	1.01-3.78	0.047	*	0.91	0.39-2.13	0.828		0.89	0.39-2.04	0.779		0.75	0.32-1.78	0.519		0.91	0.39-2.10	0.825	

n.s., not significant. *, *P*<0.05; **, *P*<0.01; ***, *P*<0.001.

Supplementary Table 2. Characteristics and HBV-related features of HBV-HCCs in validation cohort.

Group	All	G1 HBV-*TERT*	G2 HBV-*MLL4*	G3 *TERT*p mut	G4 Others	Statistics
N	112 (100%)	34 (30%)	14 (13%)	35 (31%)	29 (26%)	
Genetic changes						
HBV integration	107 (96%)	34 (100%)	14 (100%)	30 (86%)	29 (100%)	G3 vs non-G3, P=0.0024***
						G1 vs non-G1, P=0.0016**
CTNNB1 exon3 mutation	25 (22%)	14 (41%)	0 (0%)	6 (17%)	5 (17%)	G2 vs non-G2, P=0.0368*
						G1 vs G2, P=0.0041#
TP53 mutation	36 (32%)	11 (32%)	2 (14%)	14 (40%)	9 (31%)	n.s.
Clinical Features						
Gender						
Male	89 (79%)	31 (91%)	11 (79%)	32 (91%)	15 (52%)	G1 vs non-G1, P=0.0458* G3 vs non-G3, P=0.0433* G4 vs non-G4, P<0.0001***
Female	23 (21%)	3 (9%)	3 (21%)	3 (9%)	14 (48%)	G1 vs G4, P=0.0006# G3 vs G4, P=0.0005#
Age						
>60yr	54 (48%)	17 (50%)	3 (21%)	24 (69%)	10 (34%)	G2 vs non-G2, P=0.0445* G3 vs non-G3, P=0.0037**
<=60yr	58 (52%)	17 (50%)	11 (79%)	11 (31%)	19 (66%)	G2 vs G3, P=0.0041# G3 vs G4, P=0.0065#
Tumor Size						
>5cm	30 (27%)	6 (18%)	4 (29%)	7 (20%)	13 (45%)	G4 vs non-G4, P=0.0108*
<=5cm	82 (73%)	28 (82%)	10 (71%)	28 (80%)	16 (55%)	
Cirrhosis						
Positive	47 (42%)	15 (44%)	6 (43%)	14 (40%)	12 (41%)	n.s.
Negative	65 (58%)	19 (56%)	8 (57%)	21 (60%)	17 (59%)	
Tumor Grade (Edmondson)						
G1-G2	80 (73%)	24 (71%)	10 (71%)	26 (76%)	20 (71%)	n.s.
G3-G4	30 (27%)	10 (29%)	4 (29%)	8 (24%)	8 (29%)	
Microvascular invasion						
Yes	42 (38%)	15 (44%)	5 (36%)	13 (37%)	9 (32%)	n.s.
No	69 (62%)	19 (56%)	9 (64%)	22 (63%)	19 (68%)	
AFP						

>100ng/mL	25	(22%)	3	(9%)	2	(14%)	5	(14%)	15	(52%)	G1 vs non-G1, P=0.0266* G4 vs non-G4, P<0.0001***
<=100ng/mL	87	(78%)	31	(91%)	12	(86%)	30	(86%)	14	(48%)	G1 vs G4, P=0.0002# G3 vs G4, P=0.0013#
Integrated HBV-related features											
Number of clonal junctions											
Range	1-35		2-21		2-15		1-28		1-35		n.s.
Average±SD	5.5±5.0		5.9±4.3		6.6±3.8		5.3±5.4		4.9±6.1		
HBV genotype											
B	70	(65%)	18	(53%)	11	(79%)	20	(67%)	21	(72%)	n.s.
C	37	(35%)	16	(47%)	3	(21%)	10	(33%)	8	(28%)	
Basal core promoter mutation (A1762T/G1764A)											
not Mutant	76	(81%)	24	(77%)	13	(93%)	18	(72%)	21	(88%)	n.s.
Mutant	18	(19%)	7	(23%)	1	(7%)	7	(28%)	3	(13%)	
Pre-core mutation (G1896A)											
not Mutant	69	(79%)	22	(79%)	10	(83%)	20	(77%)	17	(81%)	n.s.
Mutant	18	(21%)	6	(21%)	2	(17%)	6	(23%)	4	(19%)	

n.s., not significant.

*, *P*<0.05; **, *P*<0.01; ***, *P*<0.001; #, *P*<0.0083 in pairwise comparison by Chi-square test or Fisher's Exact test.

Supplementary Reference

1 Tang D, Li B, Xu T, Hu R, Tan D, Song X, et al. VISDB: a manually curated database of viral integration sites in the human genome. Nucleic Acids Res. 2020;48(D1):D633-D41.