Appendix 1

Genomic DNA isolation, WGBS library generation and sequencing

Genomic DNA isolation, WGBS library generation and sequencing were completed by Novogene Co., Ltd, China. Genomic DNA was isolated using Tiangen DP304 kit, and agarose gels were used to monitor degradation and contamination of DNA. DNA purity was measured by the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). DNA concentration was confirmed using Qubit® DNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies, CA, USA). All operation complied with the manufacturer's instructions. A total amount of 5.2 microgram genomic DNA spiked with 26 ng lambda DNA were fragmented by sonication to 200–300 bp with Covaris S220, followed by end repair and adenylation. Cytosine-methylated barcodes were ligated to sonicated DNA as per manufacturer's instructions. Ligated DNA fragments were treated twice with bisulfite using EZ DNA Methylation-GoldTM Kit (Zymo Research). The bisulfite-treated single-strand DNA fragments were PCR amplificated using KAPA HiFi HotStart Uracil + ReadyMix (2X). Library concentration was quantified by Qubit 2.0 Flurometer (Life Technologies) and quantitative PCR, and the insert size was assayed on Agilent Bioanalyzer 2100 system. The library preparations were sequenced on an Illumina NovaSeq platform, and 150 bp paired-end reads were generated. Image analysis and base calling were performed with Illumina CASAVA pipeline, and finally 150 bp paired-end reads were generated. At least 30 × sequencing depth was reached for each sample.

WGBS data analysis

FastQC was used to generate a quality report of raw WGBS sequencing reads (55). Fastp was used for quality control and low-quality reads filtering, which was developed in C++, and it performed fast quality control and data filtering on high-throughput sequencing reads (56). Adapters were trimmed from raw reads (AGATCGGAAGAGCACACGTCTGAACTC CAGTCA, AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT). Two ends with quality less than 3 or with N bases were also trimmed (cut_front, cut_front_window_size 1, cut_front_mean_quality 3, cut_tail, cut_tail_window_size 1, cut_tail_mean_quality 3). A sliding window of 4 bases was used to examine the average quality, and bases were trimmed for average quality lower than 15 (cut_right, cut_right_window_size 4, cut_right_mean_quality 15). Due to biased cytosine bisulfite treatment in front 10 bps, two ends reads were trimmed 10 bps from front (trim_front1 10, trim_front2 10). The minimum length to retain a trimming read was 36 bps (length_required 36). Only correctly paired reads were retained for downstream analysis, and they were clean reads for WGBS libraries.

MOABS software was used for DNA methylation analysis (18). The BSMAP module of MOABS was used to align clean reads to the reference genome (hg38) (57,58). The MCALL module of MOABS performed the single sample analysis to report methylation levels of CpG dinucleotides across the whole genome. Differentially methylated analysis was conducted using MCOMP module of MOABS.

Publicly available datasets

The PMRs were annotated using National Center for Biotechnology Information (NCBI) refseq annotations downloaded from the University of California Santa Cruz (UCSC) genome browser. A promoter region was defined 1,000 bp upstream and 500 bp downstream of transcription start site. The H3K27ac/H3K4me1 ChIP-seq data for ICC-derived cancer cell lines (CCLs) were obtained from GSE68388 (59). Transcription factors (TFs, TEAD1, TEAD4, YAP, TAZ and RPH) ChIP-seq data for CCLs were acquired from GSE68296 and GSE124430 (59,60). The ATAC-seq data for ICC samples were derived from TCGA database (61). The Chip-seq and ATAC-seq peaks from original manuscripts were employed, and genomic coordinates in hg19 were transformed into GRCh38 using LiftOver tool of UCSC. A PMR is considered to be annotated by a feature only when the overlapped length accounts for >50% of the PMR or the feature.

PMS formula

(–0.353)*chr1:171,854,763-171,854,992 + (–0.202)*chr2:21,710,188-21,710,438 + (–0.232)*chr2:62,464,507-62,465,170 + (–0.593)*chr3:179,767,886-179,768,188 + (–0.742)*chr6:14,396,913-14,397,254 + (–0.187)*chr7:149,987,544-149,987,810 + (–0.464)*chr9:22,005,150-22,006,798 + (–0.899)*chr10:239,222-239,510 + (–0.387)*chr11:57,165,910-57,166,366 + (–0.626)*chr12:64,671,264-64,671,938 + (–0.477)*chr12:103,319,240-103,319,760 + (–0.363)*chr15:85,545,689-85,545,931 + (–0.358)*chr15:97,383,669-97,383,930 + (–0.261)*chr18:26,881,728-26,881,977

https://dx.doi.org/10.21037/hbsn-21-424

References

55. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc
56. Chen S, Zhou Y, Chen Y, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884-90.
57. Xi Y, Bock C, Müller F, et al. RRBSMAP: a fast, accurate and user-friendly alignment tool for reduced representation bisulfite sequencing. Bioinformatics 2012;28:430-2.
58. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics 2009;10:232.
59. Galli GG, Carrara M, Yuan WC, et al. YAP Drives Growth by Controlling Transcriptional Pause Release from Dynamic Enhancers. Mol Cell 2015;60:328-37.
60. Kitchen P, Lee KY, Clark D, et al. A Runaway PRH/HHEX-Notch3-Positive Feedback Loop Drives Cholangiocarcinoma and Determines Response to CDK4/6 Inhibition. Cancer Res 2020;80:757-70.
61. Corces MR, Granja JM, Shams S, et al. The chromatin accessibility landscape of primary human cancers. Science 2018;362:eaav1898.

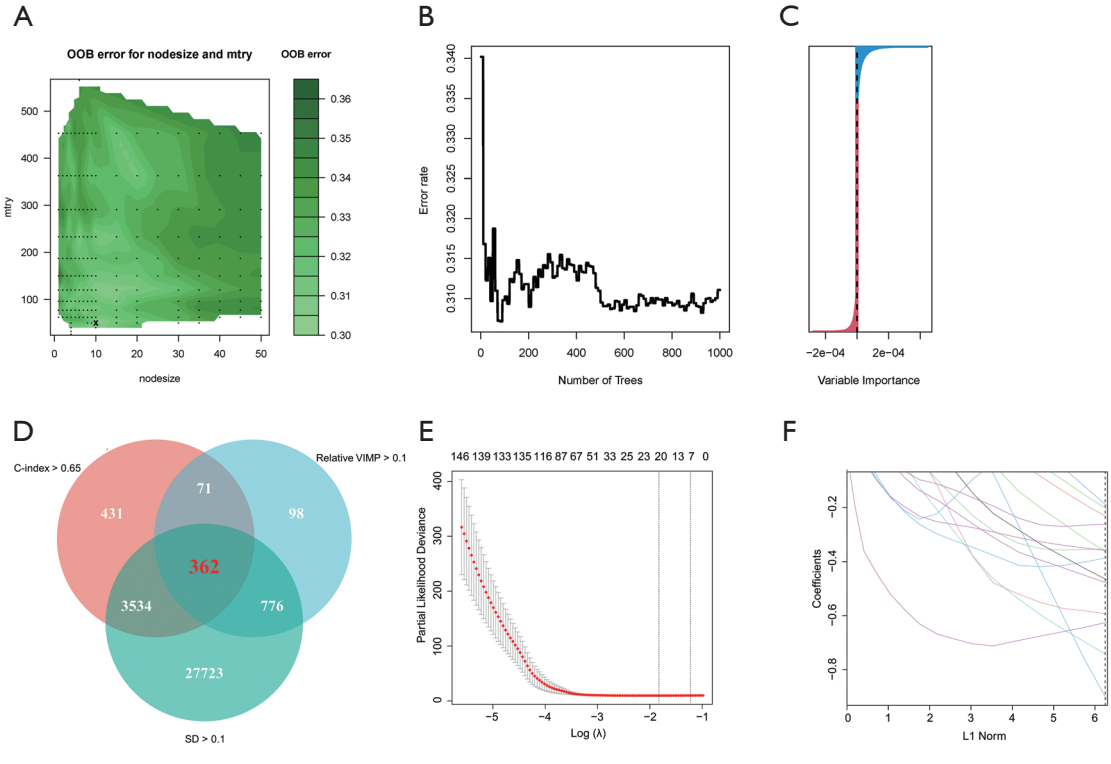

Figure S1 Random forest and LASSO Cox results for the PMS construction. (A) Grid search for optimal mtry and nodesize combination (marked by "x"). (B) The association between error rate distribution and number of trees. (C) VIMP for 35,023 PMRs, y-axis represents PMRs, x-axis represents VIMP for each PMR, "blue" indicates a significant association between PMRs and OS, and "red" implies an insignificant association between PMRs and OS. (D) Overlap of PMRs according to following criteria: relative VIMP >0.1, C-index >0.65 and SD >0.1. (E) Penalized LASSO Cox regression analysis to select survival-associated PMRs in the discovery cohort. (F) Fourteen selected PMRs includded in the PMS, L1 Norm represents the summation of absolute nonzero coefficients at each λ, y-axis represents the values of nonzero coefficients at each λ. OOB, out-of-bag; VIMP, variable importance; SD, standard deviation; LASSO, least absolute shrinkage and selector operation; PMS, prognostic methylation score; PMRs, prognostically methylated regions; OS, overall survival.

https://dx.doi.org/10.21037/hbsn-21-424

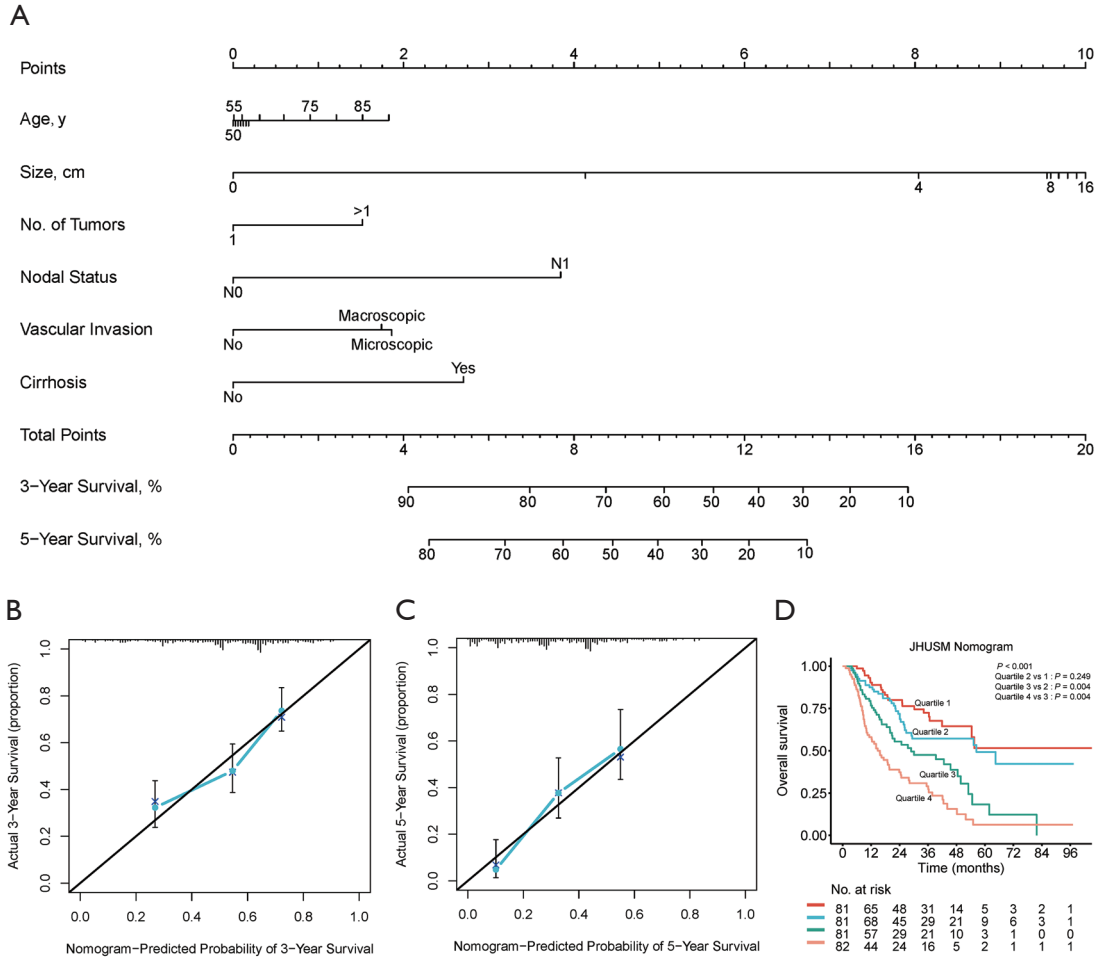

Figure S2 JHUSM nomogram. (A) JHUSM nomogram. (B) Calibration curve for predicting survival at 3 years. (C) Calibration curve for predicting survival at 5 years. (D) Kaplan-Meier curves of OS by JHUSM nomogram quartile 1–4. JHUSM, Johns Hopkins University School of Medicine.

https://dx.doi.org/10.21037/hbsn-21-424

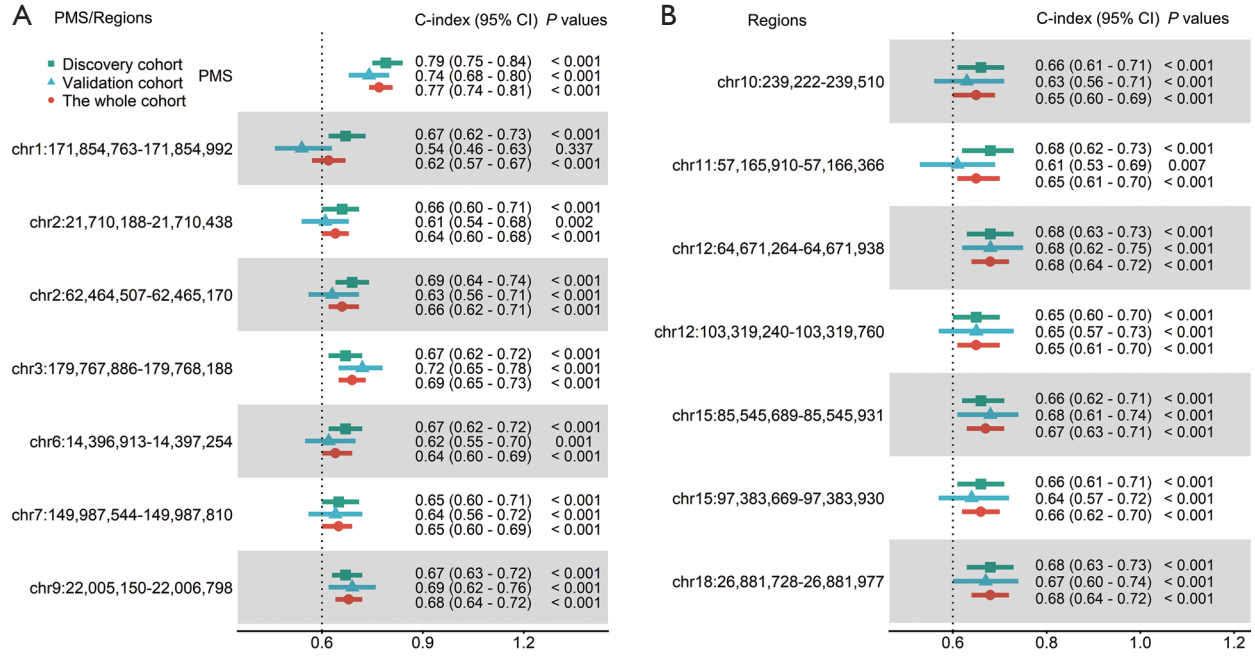

Figure S3 C-indices of the PMS and 14 PMRs for constructing the PMS in the training, validation and whole cohorts. PMS, pronostic methylation score; PMRs, prognostically methylated regions.

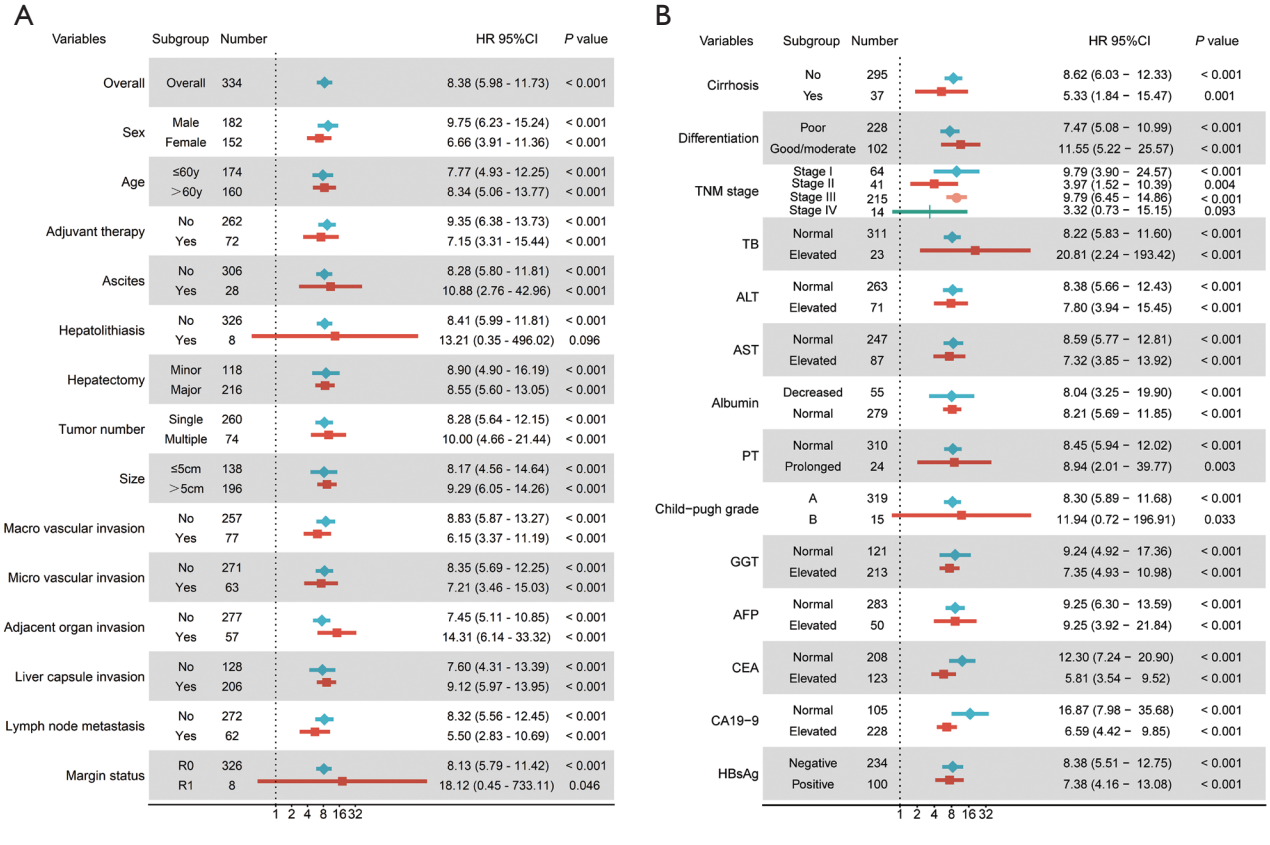

Figure S4 Subgroup analyses according to clinical variables. HR, hazard ratio; TNM stage, tumour-node-metastasis stage; TB, total bilirubin; ALT, alanine transaminase; AST, aspartate aminotransferase; PT, prothrombin time; GGT, gamma-glutamyl transpeptidase; AFP, alpha-fetoprotein; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; HBsAg, hepatitis B surface antigen.

https://dx.doi.org/10.21037/hbsn-21-424

Table S1 Baseline characteristics of training and validation cohort

Variable	Training cohort	Validation cohort	P value
Age (years)	59 (51–65)	62 (55–69)	0.004
Sex			
Male	87 (53.0)	95 (55.9)	0.603
Female	77 (47.0)	75 (44.1)	
Adjuvant treatment			
No	151 (92.1)	111 (65.3)	<0.001
Yes	13 (7.9)	59 (34.7)	
Ascites			
No	143 (87.2)	163 (95.9)	0.004
Yes	21 (12.8)	7 (4.1)	
Hepatectomy			
Minor	44 (26.8)	74 (43.5)	0.001
Major	120 (73.2)	96 (56.5)	
Hepatolithiasis			
No	163 (99.4)	163 (95.9)	0.067[b]
Yes	1 (0.6)	7 (4.1)	
Size (cm)	5.8 (4.5–7.2)	5.5 (4.0–7.6)	0.155
Tumor number			
Single	125 (76.2)	135 (79.4)	0.483
Multiple	39 (23.8)	35 (20.6)	
Macro vascular invasion			
No	140 (85.4)	117 (68.8)	<0.001
Yes	24 (14.6)	53 (31.2)	
Micro vascular invasion			
No	144 (87.8)	127 (74.7)	0.002
Yes	20 (12.2)	43 (25.3)	
Adjacent organ invasion			
No	138 (84.1)	139 (81.8)	0.563
Yes	26 (15.9)	31 (18.2)	
Liver capsule invasion			
No	51 (31.1)	77 (45.3)	0.008
Yes	113 (68.9)	93 (54.7)	
Lymph node metastasis			
No	125 (76.2)	147 (86.5)	0.016
Yes	39 (23.8)	23 (13.5)	
Distal metastasis			
M0	158 (96.3)	162 (95.3)	0.633
M1	6 (3.7)	8 (4.7)	
Cirrhosis			
No	145 (88.4)	150 (88.2)	0.367
Yes	19 (11.6)	18 (10.6)	
Missing	0 (0)	2 (1.2%)	
Differentiation			
Poor	118 (72.0)	110 (64.7)	0.076
High/moderate	46 (28.0)	56 (32.9)	
Missing	0 (0)	4 (2.4)	
Margin status			
R0	157 (95.7)	169 (99.4)	0.034
R1	7 (4.3)	1 (0.6)	
AJCC TNM stage			
IA	12 (7.3)	32 (18.8)	0.002
IB	11 (6.7)	9 (5.3)	
II	13 (7.9)	28 (16.5)	
IIIA	72 (43.9)	55 (32.4)	
IIIB	50 (30.5)	38 (22.4)	
IV	6 (3.7)	8 (4.7)	
TB (μmol/L)	13.2 (10.3, 17.0)	11.2 (8.7, 16.2)	0.002
ALT (IU/L)	23 (16.0, 37.0)	22 (15.0, 34.0)	0.260
AST (IU/L)	29 (23.0, 37.0)	24 (19.0, 30.0)	<0.001
Albumin (g/L)	44.1 (40.9, 46.7)	43.8 (40.3, 46.0)	0.342
PT (second)	11.5 (10.9, 12.1)	11.2 (10.7, 11.7)	0.002
Child-Pugh grade			
A	159 (97.0)	160 (94.1)	0.211
B	5 (3.0)	10 (5.9)	
GGT (IU/L)	64.5 (33.0, 119.0)	62.5 (31.0, 137.5)	0.544
AFP (ng/mL)	3.5 (2.4, 5.5)	2.9 (2.2, 5.4)	0.200
CEA (ng/mL)	2.5 (1.4, 4.9)	2.6 (1.6, 4.6)	0.634
CA19-9 (ng/mL)	63.2 (16.6, 431.2)	51.9 (17.5, 284.4)	0.774
HBsAg			
Negative	118 (72.0)	116 (68.2)	0.459
Positive	46 (28.0)	54 (31.8)	

Categorical variables are presented as number of patients; continuous variables are presented as median (IQR). [b], Fisher's exact test. AJCC, American Joint Committee on Cancer; TNM, tumor-node-metastasis; TB, total bilirubin; ALT, alanine transaminase; AST, aspartate aminotransferase; PT, prothrombin time; GGT, gamma-glutamyl transpeptidase; AFP, alpha-fetoprotein; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; HBsAg, hepatitis B surface antigen; IQR, interquartile range.

https://dx.doi.org/10.21037/hbsn-21-424

Table S2 Univariate and multivariate analysis of the whole cohort

Variables	Number or median (IQR)	Univariate analysis	P value	Multivariate analysis HR (95% CI) (without PMS)	P value	Multivariate analysis HR (95% CI) (with PMS)	P value
Age (years)	60 (53, 67)	0.99 (0.98–1.01)	0.399	–	–	–	–
Sex (male/female)	182/152	1.13 (0.83–1.53)	0.434	–	–	–	–
Preoperative treatment (yes/no)	5/329	1.08 (0.27–4.36)	0.916	–	–	–	–
Adjuvant (yes/no)	72/262	0.83 (0.56–1.22)	0.337	–	–	–	–
Ascites (yes/no)	28/306	2.53 (1.56–4.10)	<0.001	2.72 (1.60–4.63)	<0.001	2.68 (1.59–4.51)	<0.001
Hepatectomy (major/minor)	216/118	1.22 (0.88–1.69)	0.241	–	–	–	–
Hepatolithiasis (yes/no)	8/326	1.56 (0.50–4.91)	0.444	–	–	–	–
Tumor number (multiple/single)	74/260	1.45 (1.02–2.05)	0.038	1.33 (0.93–1.92)	0.123	1.54 (1.06–2.24)	0.022
Size (cm)	5.5 (4.2, 7.4)	1.10 (1.04–1.16)	0.001	1.07 (1.01–1.14)	0.029	1.12 (1.05–1.19)	<0.001
Macro vascular invasion (yes/no)	77/257	1.67 (1.19–2.35)	0.003	1.61 (1.10–2.35)	0.015	1.27 (0.84–1.92)	0.266
Micro vascular invasion (yes/no)	63/271	1.97 (1.38–2.82)	<0.001	1.35 (0.90–2.02)	0.150	1.34 (0.88–2.06)	0.173
Adjacent organ invasion (yes/no)	57/277	1.44 (0.99–2.10)	0.054	–	–	–	–
Liver capsule invasion (yes/no)	206/128	1.08 (0.79–1.49)	0.631	–	–	–	–
Lymph node metastasis (yes/no)	62/272	2.94 (2.10–4.11)	<0.001	2.57 (1.78–3.71)	<0.001	1.90 (1.33–2.72)	<0.001
Distal metastasis (yes/no)	14/320	2.66 (1.44–4.91)	0.002	1.32 (0.66–2.64)	0.427	1.11 (0.57–2.19)	0.754
Cirrhosis (yes/no)	37/295	1.92 (1.23–2.99)	0.004	1.41 (0.83–2.39)	0.204	0.98 (0.58–1.64)	0.926
Differentiation (poor/high or moderate)	228/102	1.65 (1.16–2.35)	0.005	1.70 (1.17–2.47)	0.005	1.26 (0.86–1.85)	0.236
Margin status (yes/no)	8/326	1.14 (0.51–2.59)	0.746	–	–	–	–
AJCC TNM stage			0.001	–	–	–	–
Stage I	64	–	Reference				
Stage II	41	2.06 (1.11–3.84)	0.023				
Stage III	215	1.79 (1.10–2.91)	0.019				
Stage IV	14	4.43 (2.10–9.36)	<0.001				
TB (μmol/L)	12.0 (9.3, 16.4)	1.01 (1.00–1.01)	0.013	1.01 (1.00–1.01)	0.156	1.01 (1.00–1.01)	0.099
ALT (IU/L)	23 (16, 36)	–	0.778	–	–	–	–
AST (IU/L)	27 (21, 35)	–	0.304	–	–	–	–
Albumin (g/L)	44.0 (40.9, 46.4)	0.98 (0.95–1.02)	0.340	–	–	–	–
PT (second)	11.3 (10.8, 11.9)	1.23 (1.09–1.38)	0.001	1.15 (0.99–1.33)	0.071	1.08 (0.94–1.25)	0.290
Child-Pugh grade (B/A)	15/319	1.82 (0.93–3.57)	0.083	–	–	–	–
GGT (IU/L)	64 (32, 125)	1.00 (1.00–1.00)	0.006	1.00 (1.00–1.00)	0.527	1.00 (1.00–1.00)	0.474
AFP (ng/mL)	3.2 (2.3, 5.4)	1.00 (1.00–1.00)	0.514	–	–	–	–
CEA (ng/mL)	2.6 (1.5, 4.7)	1.00 (1.00–1.00)	0.203	–	–	–	–
CA19-9 (elevated/normal)	228/105	2.29 (1.59–3.32)	<0.001	1.87 (1.28–2.75)	0.001	1.22 (0.83–1.81)	0.312
HBsAg (positive/negative)	100/234	1.40 (1.01–1.94)	0.042	1.15 (0.80–1.64)	0.455	1.18 (0.83–1.68)	0.352
PMS	–4.44 (–4.79, –4.02)	8.38 (5.98–11.74)	<0.001	–	–	8.12 (5.48–12.04)	<0.001

Categorical variables are presented as number of patients; continuous variables are presented as median (IQR). IQR, interquartile range; HR, hazard ratio; CI, confidence interval; PMS, promoter methylation score; AJCC, American Joint Committee on Cancer; TNM, tumor-node-metastasis; TB, total bilirubin; ALT, alanine transaminase; AST, aspartate aminotransferase; PT, prothrombin time; GGT, gamma-glutamyl transpeptidase; AFP, alpha-fetoprotein; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; HBsAg, hepatitis B surface antigen.

https://dx.doi.org/10.21037/hbsn-21-424