Supplementary

Table S1 The clinical data of the 5 independent cohorts

Characteristic	TCGA-LIHC	GSE14520	GSE87630	GSE89377	GSE121248
Platform	Illumina HiSeq2000	GPL571 GPL3921	GPL6947	GPL6947	GPL570
Samples	427	488	94	53	107
Normal	50	241	30	13	37
Tumor	377	247	64	40	70
Survival status					
Death	128	96	NA	NA	NA
Survival	249	146	NA	NA	NA
Age, years					107
≤65	235	216	NA	NA	65
>65	141	26	NA	NA	42
Gender					107
Female	122	31	NA	NA	15
Male	255	211	NA	NA	92
TNM stage					
I	180	96	NA	NA	NA
II	89	78	NA	NA	NA
III	76	51	NA	NA	NA
IV	8	0	NA	NA	NA
T classification					
T1	182	NA	NA	NA	NA
T2	91	NA	NA	NA	NA
T3	63	NA	NA	NA	NA
T4	17	NA	NA	NA	NA
N classification					
N0	349	NA	NA	NA	NA
N1	4	NA	NA	NA	NA
M classification					
M0	349	NA	NA	NA	NA
M1	4	NA	NA	NA	NA

© Annals of Translational Medicine. All rights reserved.

Table S2 The primer sequences of eight prognostic genes and β-actin

Primer	Forward sequence	Reverse sequence
CDC20	CGGAAGACCTGCCGTTACATTC	CAGAGCTTGCACTCCACAGGTA
PTTG1	GCTTTGGGAACTGTCAACAGAGC	CTGGATAGGCATCATCTGAGGC
TOP2A	GTGGCAAGGATTCTGCTAGTCC	ACCATTCAGGCTCAACACGCTG
CXCL2	GGCAGAAAGCTTGTCTCAACCC	CTCCTTCAGGAACAGCCACCAA
CXCL14	AGATCCGCTACAGCGACGTGAA	GCAGTGCTCCTGACCTCGGTA
CYP2C9	CAGAGACGACAAGCACAACCCT	ATGTGGCTCCTGTCTTGCATGC
GHR	GCAGCTATCCTTAGCAGAGCAC	AAGTCTCTCGCTCAGGTGAACG
MT1F	GACTGATGCCAGGACAACCT	AGGAATGTAGCAAATGGGTCA
β-actin	CACCAACTGGGACGACAT	ACAGCCTGGATAGCAACG

Table S3 Top 20 genes ranked by the degree method

Rank	Gene	Score	Rank	Gene	Score
1	*PTTG1*	6	10	*IGFALS*	2
1	*CDC20*	6	10	*STAB2*	2
1	*CCNB2*	6	10	*CXCL2*	2
1	*NUSAP1*	6	10	*CXCL12*	2
1	*TOP2A*	6	10	*CYP1A2*	2
1	*PRC1*	6	10	*CYP2C9*	2
1	*ASPM*	6	10	*MT1G*	2
8	*FOS*	3	10	*MT1H*	2
8	*CYP2E1*	3	19	*CXCL14*	1
10	*MT1F*	2	19	*GHR*	1

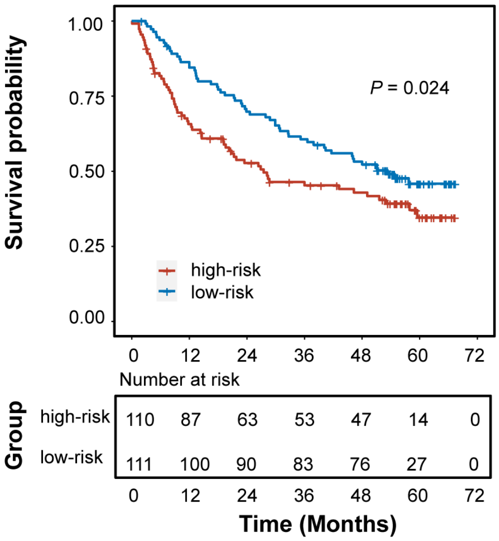

Figure S1 Kaplan-Meier analysis shows that HCC patients with high risk scores have poorer DFS in the GSE14520 validation set.

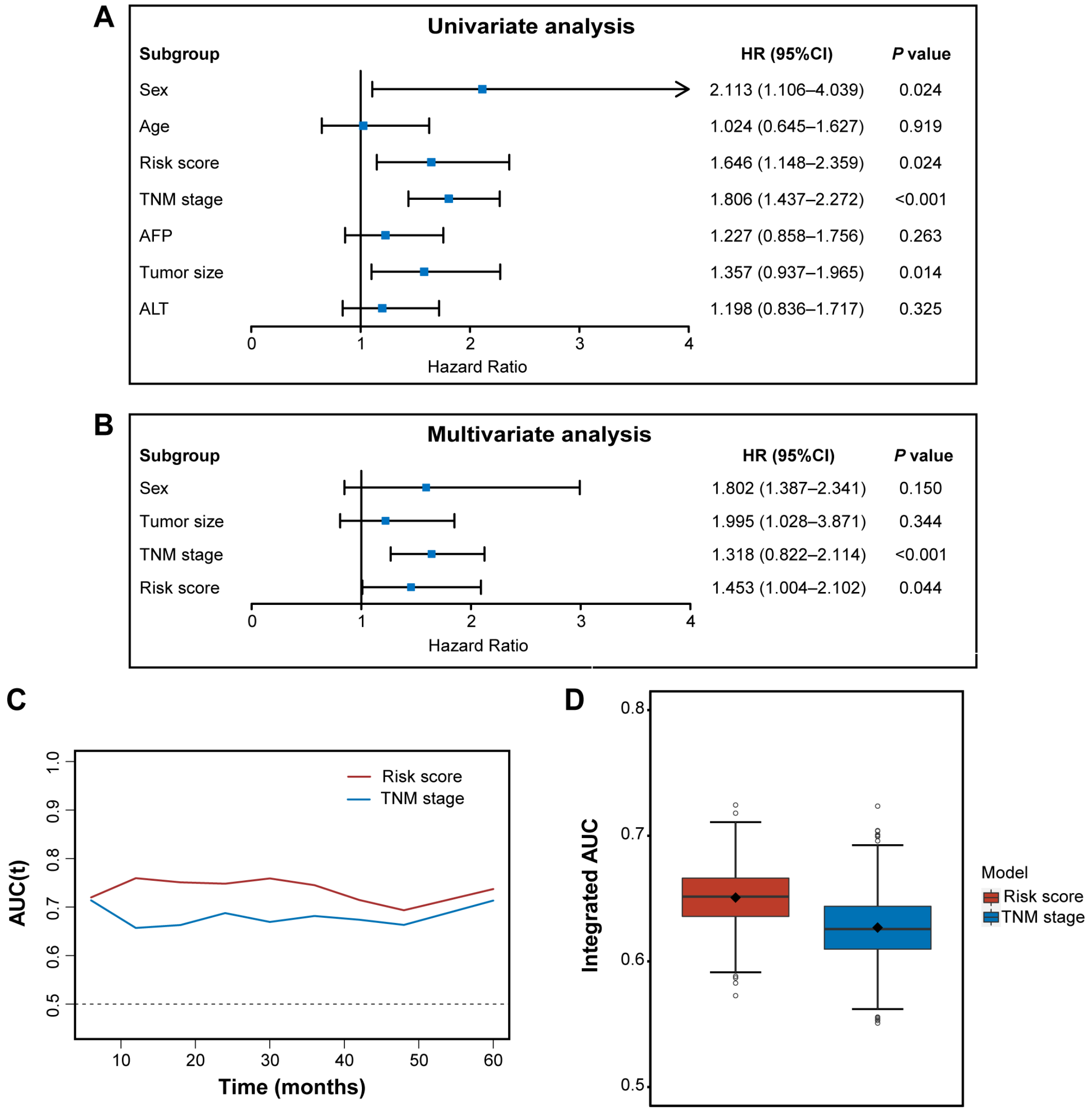

Figure S2 (A,B) Univariate and multivariate Cox regression analysis for DFS in the GSE14520 validation set. (C) Time-dependent AUC for DFS shows that risk score has better predictive accuracy than TNM stage in the GSE14520 validation set. (D) The iAUC indicates integrated area under the ROC curve, which shows that risk score has better predictive performance than TNM stage. DFS, disease-free survival; AUC, area under the curve; iAUC, integrated value of time-dependent AUC.

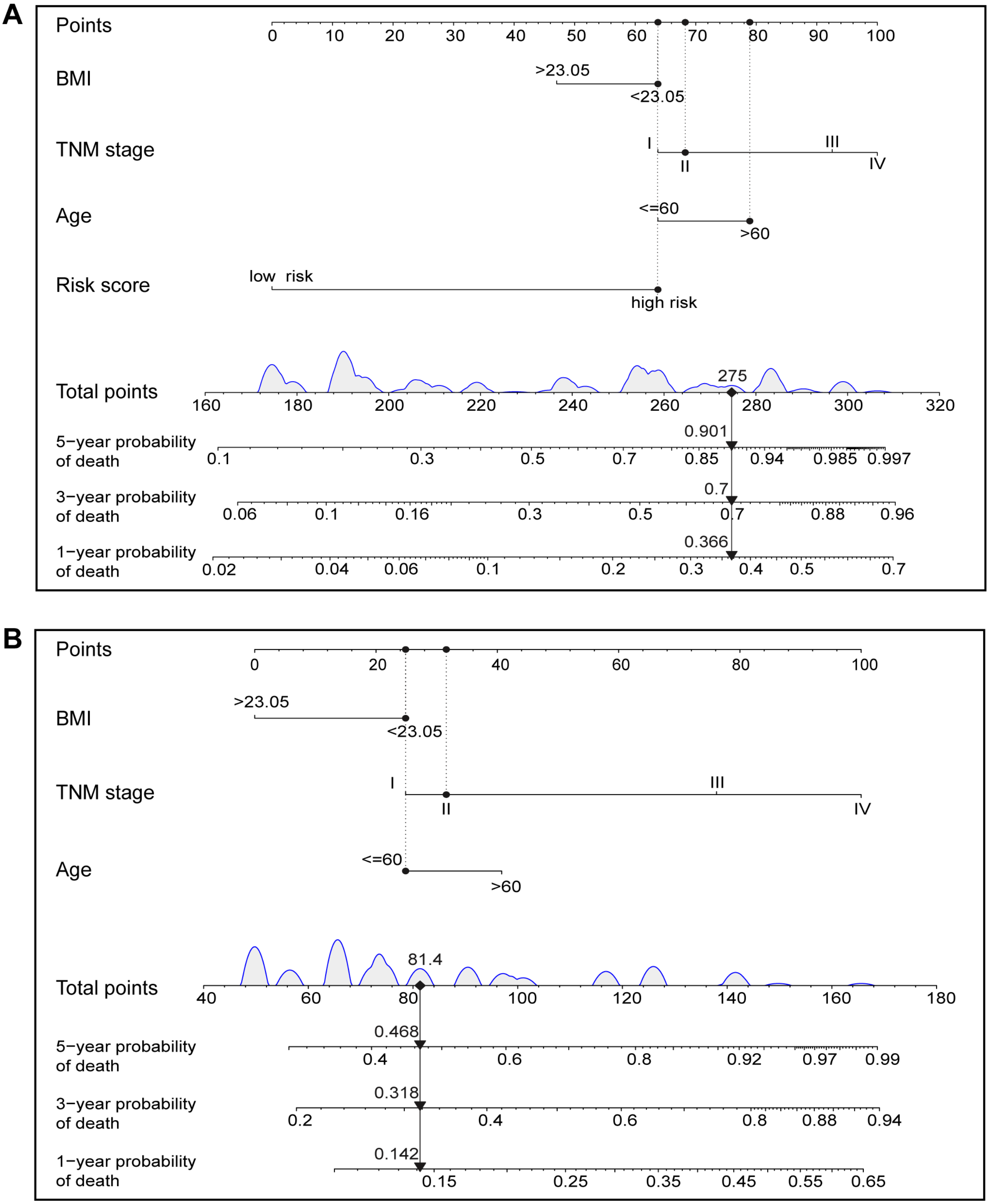

Figure S3 (A) The nomogram to forecast the 1-, 3-, and 5-year death likelihood of HCC patients in the TCGA-LIHC training set. The nomogram model is constructed based on BMI, TNM stage, age, and risk score of the 8 risk genes. (B) The nomogram without risk score in the TCGA-LIHC training set.

© Annals of Translational Medicine. All rights reserved.
https://dx.doi.org/10.21037/atm-22-1934

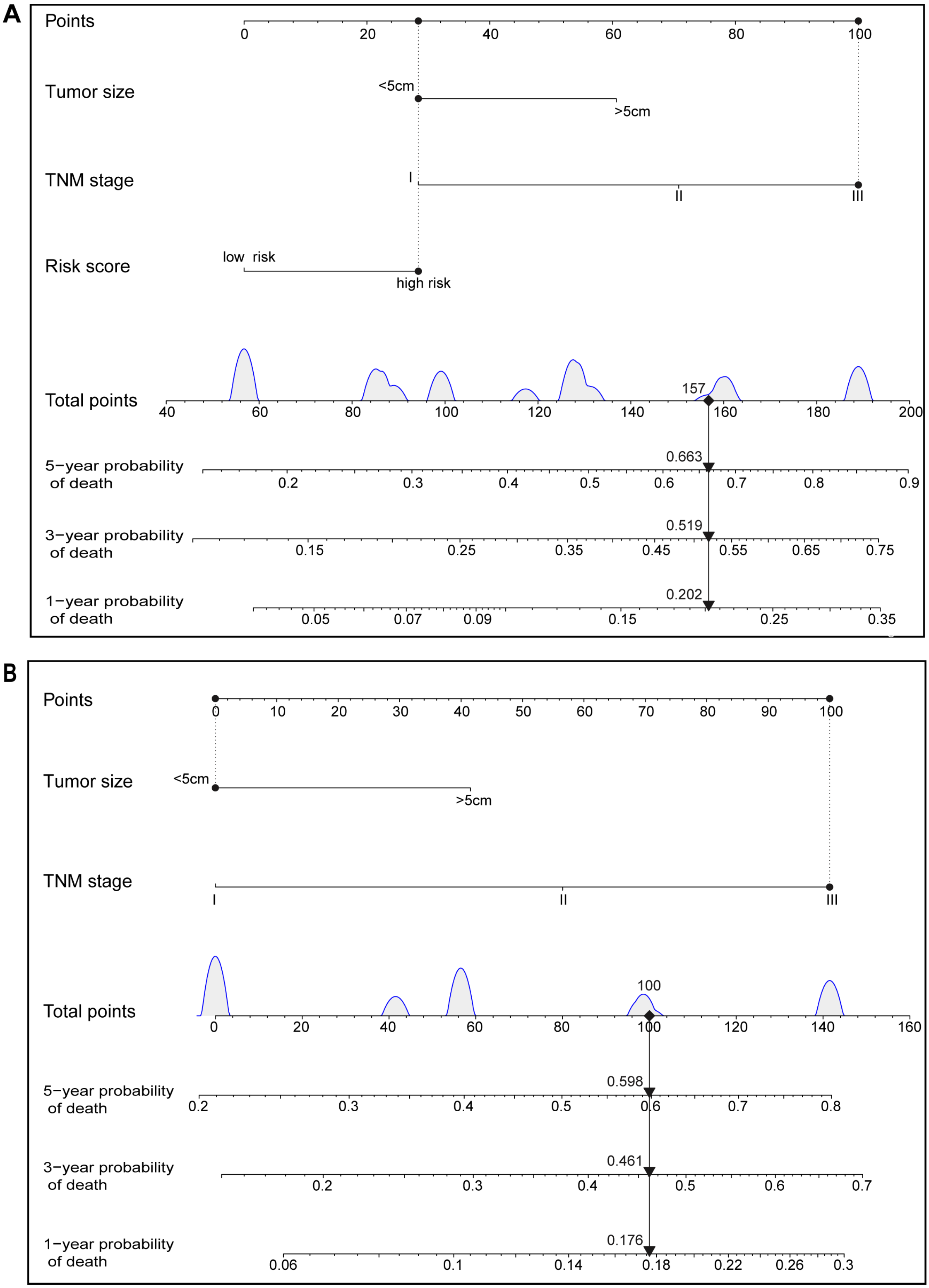

Figure S4 (A) The nomogram to forecast the death likelihood of HCC patients at 1, 3, and 5 years in the GSE14520 validation set. The nomogram model is constructed based on TNM stage, tumor size, and risk score of the 8 risk genes. (B) The nomogram without risk score constructed in the GSE14520 validation set.

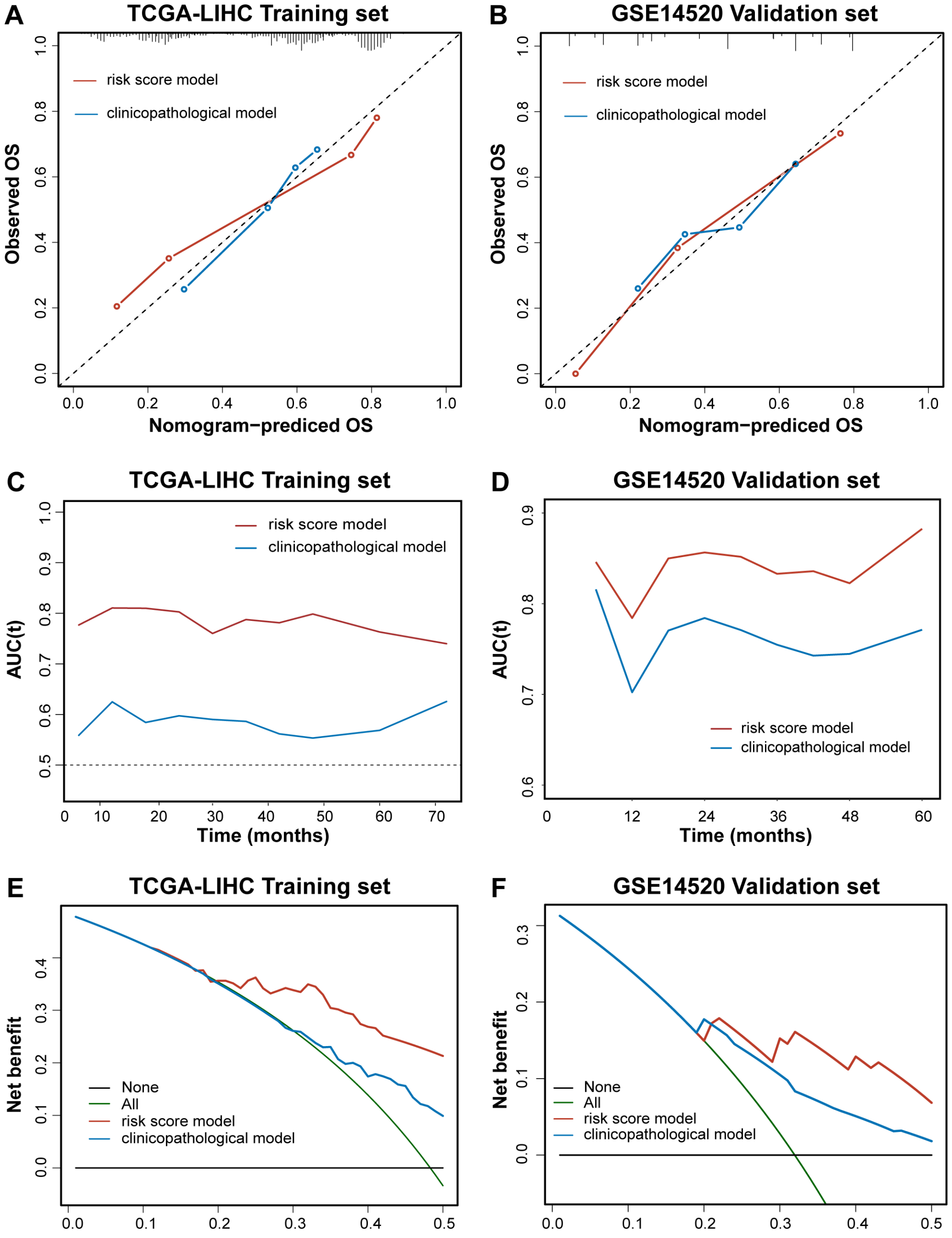

Figure S5 (A,B) The nomogram calibration curves in the TCGA-LIHC dataset and GSE14520 validation set. (C,D) Time-dependent AUC of the risk score model and clinicopathological model in the training set and validation set. (E,F) DCA curves of the risk score model and clinicopathological model in the training set and validation set.

© Annals of Translational Medicine. All rights reserved.

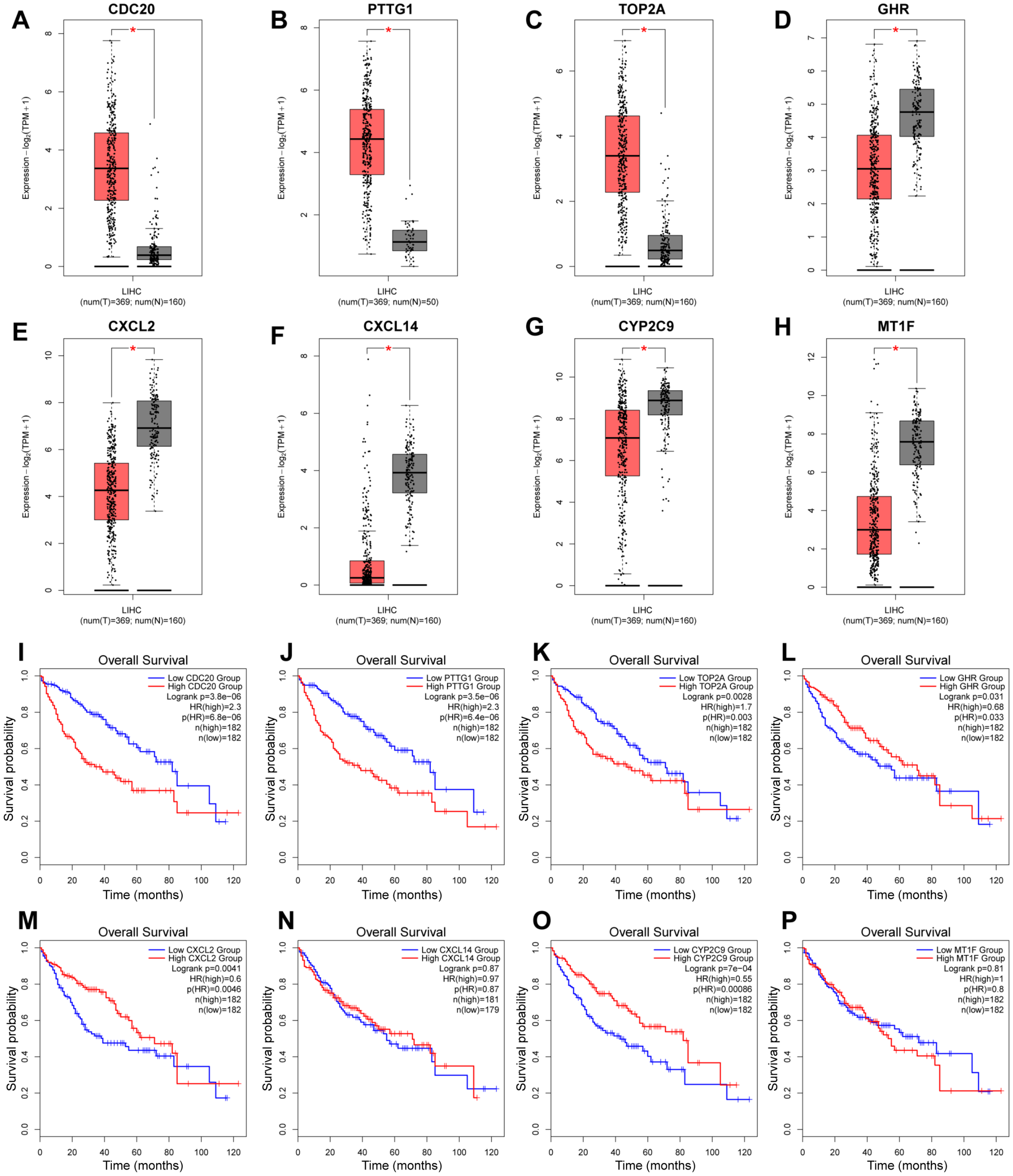

Figure S6 (A-H) Comparisons of mRNA expression of each gene in HCC tissues versus adjacent normal tissues in TCGA-LIHC via GEPIA. (I-P) Validation of the prognostic role of each gene by Kaplan-Meier survival analysis via GEPIA. *, P<0.05.

© Annals of Translational Medicine. All rights reserved. https://dx.doi.org/10.21037/atm-22-1934

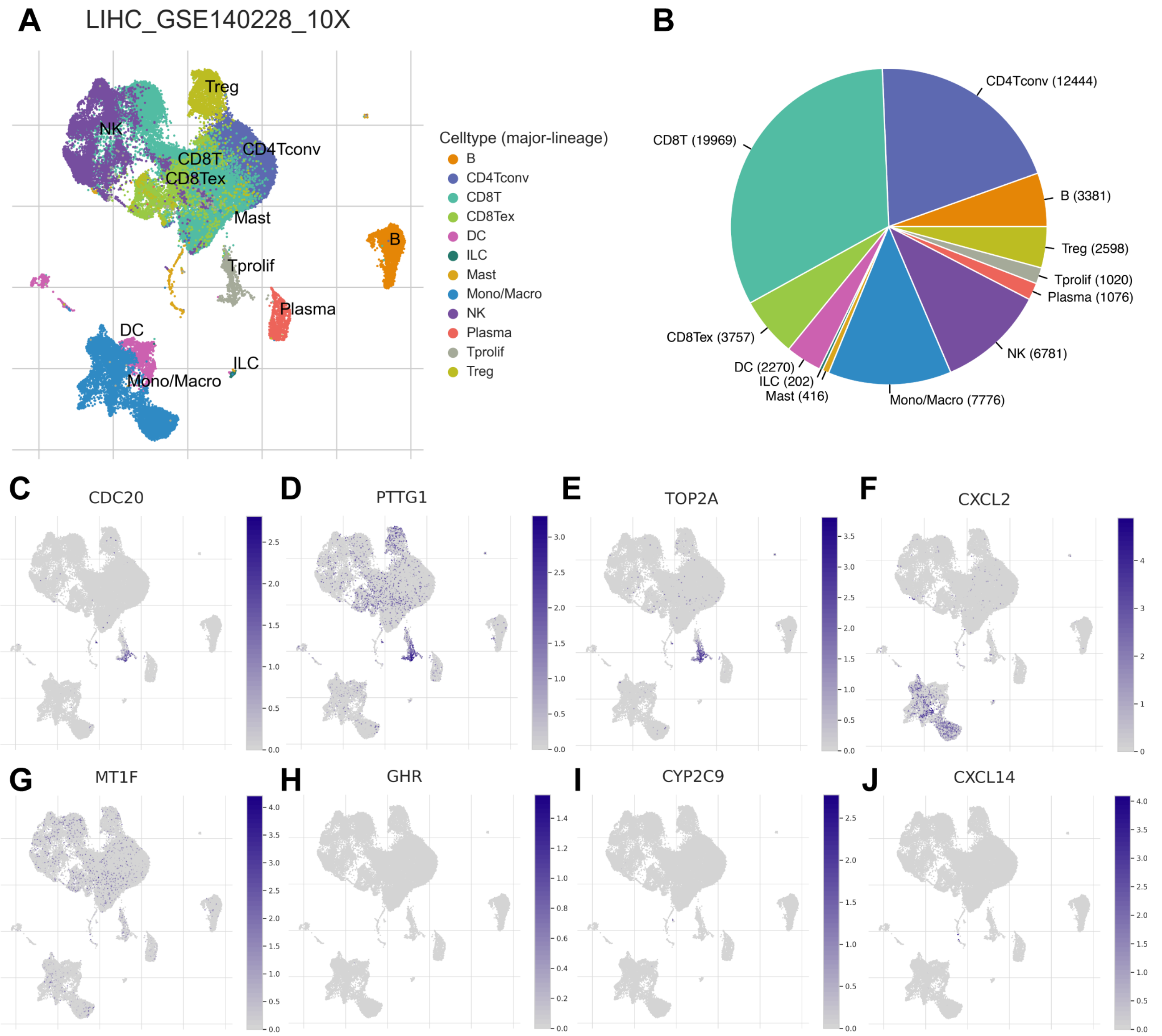

Figure S7 (A) The landscape of different cell types in the LIHC_GSE140228_10X single-cell RNA sequencing dataset. (B) The cell types as well as their distribution in the LIHC_GSE140228_10X dataset. (C-J) The expression of *CDC20*, *PTTG1*, *TOP2A*, *CXCL2*, *MT1F*, *GHR*, *CYP2C9*, and *CXCL14* in different cell types.