Development and Validation of a Risk Prediction Score for Patients with Nasopharyngeal Carcinoma

Ning Xue
Affiliated Tumor Hospital of Zhengzhou University: Henan Cancer Hospital
Guoping Ou
Sun Yat-sen University Cancer Center
Weiguo Ma
Affiliated Tumor Hospital of Zhengzhou University: Henan Cancer Hospital
Lina Jia
Affiliated Tumor Hospital of Zhengzhou University: Henan Cancer Hospital
Jiahe Sheng
Affiliated Tumor Hospital of Zhengzhou University: Henan Cancer Hospital
Qingxia Xu
Affiliated Tumor Hospital of Zhengzhou University: Henan Cancer Hospital
Yubo Liu
Sun Yat-sen University Cancer Center
Miaomiao Jia (zlyyjiamm4274@zzu.edu.cn)
Zhengzhou University https://orcid.org/0000-0001-7282-1535

Primary research

Keywords: nasopharyngeal carcinoma, risk score, nomogram, overall survival

Posted Date: April 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-409633/v1

Abstract

Background: The present study aimed to develop and validate an effective predictive model for overall survival (OS) in patients with nasopharyngeal carcinoma (NPC).

Methods: 519 patients were retrospectively reviewed in this study. Random forest was proposed to identify significant prognostic factors for OS in NPC patients. Then calibration plot and concordance index(C-index) were used to evaluate the predictive accuracy of the nomogram model.

Results: We used random forest to select three most important features, dNLR, HGB, EBV DNA, which were significantly relevant to OS of NPC patients. The C-index of our model for OS were 0.733 (95% CI: 0.673~0.793) and 0.772 (95% CI: 0.691~0.853) in the two cohorts, which were higher than that of TNM stage, treatment, and EBV DNA. Based on the model risk score, patients were divided into low-risk and high-risk two groups. Kaplan-Meier curves showed that two subgroups were significantly associated with OS in the primary cohort and in the validation cohort. The nomogram for OS was established including risk score, TNM stage, EBV DNA in the two cohort. It achieved higher C-index of 0.783(95% CI: 0.730~0.836) than that of the risk score model 0.733 (95% CI: 0.673~0.793) in the primary (P = 0.005).

Conclusions: The established risk score model and nomogram resulted in more accurate prognostic prediction for individual patient with NPC.

Background

Nasopharyngeal carcinoma (NPC) is an endemic malignancy characterized by its unique geographic distribution[1]. NPC contributes to a large part of the overall cancer burden in prevalent area, such as southern China, southeast Asia and northern Africa[2, 3]. Radiotherapy is the mainstay treatment for non-metastatic NPC. Chemotherapy combined with radiotherapy is recommended for treating advanced NPC[4, 5]. However, the current TNM staging system used for guidance of the treatment regimens is insufficient as many varied clinical outcomes of patients at same stages were reported[6]. Therefore, more accurate indicators for predicting prognosis need to be developed for an adequate clinical treatment.

Recently, there have been many attempts to improve NPC prognostication using blood molecular biomarkers, such as circulating Epstein-Barr virus (EBV) DNA, serum lactate dehydrogenase (LDH)[7], globulin(GLOB), hs-CRP[8], neutrophil/lymphocyte ratio (NLR)[9]. The infection of EBV is virtually 100% associated with NPC in endemic areas. The plasma EBV DNA has gradually being used in clinical applications and is considered to be the most attractive potential biomarker to complement the TNM staging system[10]. The derived neutrophils to leukocytes ratio (dNLR) has been linked to inflammatory status and clinical outcomes in several cancers, including NPC[11]. This indicates that dNLR has prognostic value and it also has the advantages of being inexpensive and easy to calculate. Hemoglobin (HGB) levels are regarded as an important determinant of outcome in a number of cancers treated with

radiotherapy, particularly gynecological tumors and NPC[12].However, it is still a challenge to screen and incorporate biomarker into new staging systems for NPC patients now.

Random forest is an effective classifier because it can predict the class of input, and select the most important features[13]. Nomograms are currently proved to be an effective tool to predict the prognosis of patients with cancers, including lung cancer[14], rectal cancer [15], and gastric cancer[16]. In this study, we used random forest to screen the factors related to prognosis of NPC and incorporate them into new staging system by establishing a nomogram model.

Materials And Methods

Patients and clinical characteristic

We retrospectively reviewed 519 patients diagnosed nasopharyngeal carcinoma at Sun Yat-Sen University Cancer Center (SYSUCC) between 2010 and 2019. Patients included criteria: (I)All patients were diagnosed NPC by pathological, (II) without double primary cancer, (III) no clinical evidence of infection or other inflammation. All nasopharyngeal cancer patients received confirmed diagnoses by pathological examination. All the patients were classified based on the 8th edition of the AJCC TNM staging guidelines.

Gender, age, family history, smoking, body mass index(BMI), TNM stage, Treatment, white blood cells(WBC), Neutrophils, Lymphocyte, Monocyte, Platelet, HGB, NLR, neutrophils/(WBC-neutrophils) ratio (dNLR), lymphocyte/monocyte ratio (LMR), platelet/lymphocyte ratio (PLR), systemic immune-inflammation index (SII), total protein(TP), albumin(ALB), GLOB, ALB/GLOB ratio (AGR), C-reactive protein (CRP), CRP/ALB ratio (CAR), apolipoprotein B (APOA), apolipoprotein B (APOB), APOA/APOB ratio (ABR), LDH, high density lipoprotein(HDL), EBV DNA, virus capsid antigen specific immunoglobulin A (VCA-IgA), early antigen immunoglobulin A (EA-IgA), prognostic nutritional index (PNI) and prognostic index (PI) were collected in our study. PNI was calculated by the formula: Alb (g/L) + 5 × lymphocyte count × 10^9/L[17].SII was calculated by the formula: PLR × Neutrophil × 10^9/L[18]. The PI score 0 was for patients with a CRP of 10 mg/L or less and a WBC count of 11 × 10^9/L or less, patients with only one of these abnormalities were allocated a score of 1, and patients with an elevation of both levels were allocated a score of 2[19]. Patients' data were collected before any treatment.

Statistical analysis

Data analyses were performed using SPSS standard version 20.0 (SPSS, Chicago, USA) and R software version 3.6.1 (http://www.R-project. org). Cut-off values were determined by the R package "survival" and "survminer". The Kaplan-Meier method was used to estimate OS of the patients in high-risk and low-risk groups. Concordance index (C-index) was used assessing the discriminative ability and predictive accuracy of the established random forest model and nomogram. The C-index was calculated and compared by the "survcomp" package. The area under the curve (AUC) was computed by the "survivalROC" package. Calibration of the nomogram for 1-, 3-, and 5-year OS was executed by comparing

the predicted survival and observed survival. All statistical tests were two-tailed, and P value less than 0.05 was considered statistically significant.

Results

Patients and clinical characteristics

A total of 519 NPC patients were enrolled in this study. All the patients were randomly divided into primary cohort (363 NPC patients) and validation cohort (156 NPC patients). Table 1 described the patients' demographic data and clinical characteristics between the primary cohort and validation cohort. In the primary cohort, 209 (57.57%) patients with NPC were male and 154(42.43%) patients were female. The mean age (SD) of patients was 46.05 (10.87) years and the median OS was 51.0 months (interquartile range (IQR), 42.3-66.7 months). There were 92 (58.97%) male and 64 (41.03%) female patients in the validation cohort. The mean age (SD) was 46.87(11.58) years and the median OS was 50.4 months (IQR, 41.7-66.0 months). In the primary cohort and validation cohort, the 1-, 3-, 5-year OS rates were 95.0%, 84.0%, 46.8% and 98.7%, 84.0%, 45.5%, respectively.

Model construction based on clinical characteristics

The sliding windows sequential forward feature selection method (SWSFS) was used to identify the important variables by minimizing the 'out of bag (OOB)' error rate (Figure 1A). In the primary cohort, three variables that dNLR (HR = 1.14, 95% CI: 1.05-1.23, $P = 9.14\times10^{-4}$), HGB (HR = 0.98, 95% CI: 0.97-0.99, $P = 5.24\times10^{-3}$) and EBV DNA (HR = 1.59, 95% CI: 1.32-1.93, $P = 1.22\times10^{-6}$) were significantly associated with OS in NPC patients (Figure 1B). The computational formula of risk score was 0.466×DNA+0.129×dNLR-0.02×HGB. The heatmap of NPC samples in two cohort were shown in figure 2, in which red represents upregulated imaging features and blue represents downregulated imaging features. Three feature clusters (C1–C3) were identified in the heatmap by the unsupervised hierarchical clustering of 59 imaging features.

Model evaluation

ROCs were used to assess the accuracy of the established risk score model, TNM stage, treatment, and EBV DNA. In the primary cohort, for 1-year OS (Figure 3A), the AUC of TNM stage, Treatment, EBV DNA, and our established model were 0.748, 0.591, 0.751 and 0.797, respectively. Moreover, our model achieved higher AUC than TNM stage, Treatment, EBV DNA for 3-year OS (Figure 3B) and 5-year OS (Figure 3C). In the validation cohort, for 1-year OS (Figure 3D), the AUC of TNM stage, Treatment, EBV DNA, and our established model were 0.399, 0.588, 0.932 and 0.854, respectively. For 3-year and 5-year OS, the AUC of TNM stage, Treatment, EBV DNA, and our established model were 0.728, 0.573, 0.794, 0.821 and 0.725, 0.555, 0.747, 0.791(Figure 3E, 3F). The results of time-dependent ROC curve for OS in the primary cohort (Figure 4A) and validation cohort (Figure 4B) showed that the AUCs of TNM stage, EBV DNA, Treatment and our established model more detail.

Moreover, we evaluated the C-Index of the established model, TNM stage, Treatment and EBV DNA for prediction of OS in the primary cohort and validation cohort. In the primary cohort, the established model achieved higher C-index of 0.733(95%CI:0.673-0.793) than TNM stage 0.712 (95%CI:0.657~0.768), Treatment 0.542 (95%CI:0.505~0.580) and EBV DNA 0.691 (95%CI:0.626~0.756). In the validation cohort, The C-index of our model, TNM stage, Treatment and EBV DNA were 0.772 (95%CI:0.691~0.853), 0.699 (95%CI:0.628~0.770), 0.551 (95%CI:0.503~0.600), 0.739 (95%CI:0.652~0.826), respectively (Table 2).

Performance of the established model in stratifying risk

Based on the computational formula of Risk score (0.466×DNA+0.129×dNLR-0.02×HGB), high risk (risk score ≤ -0.16) and low risk (risk score > -0.16) subgroups were divided in patients of NPC. We used the R package “survival” and “survminer” to determine the Cut-off value. The optimum cut-off of our model was -1.46. The results showed that patients with high-risk score had a significantly shorter OS than low-risk score patients in the primary cohort ($P < 0.01$) (Figure 5A) and in the validation cohort ($P < 0.01$) (Figure 5E). In the primary cohort, the 1-, 3-, and 5-year survival probabilities of high-risk and low-risk patients were 90.0%, 71.3%, 37.3% and 99.0%, 93.0%, 53.5%. Meanwhile, in the validation cohort, the 1-, 3-, and 5-year survival probabilities of high-risk and low-risk patients were 97.2%, 70.4%, 35.2% and 98.8%, 95.3%, 54.1%. Moreover, in patients of stage III and stage IV, Kaplan-Meier curves showed that high-risk and low-risk subgroups were significantly associated with OS outcomes in the primary cohort ($P < 0.001$, $P = 0.011$) and in the validation cohort ($P = 0.015$, $P = 0.021$).

The nomogram for the prediction of OS

We established a nomogram for OS including risk score, TNM stage, EBV DNA in the two cohort. In the primary cohort, the nomogram model achieved a C-index of 0.783(95% CI: 0.730~0.836), which was significantly higher than that of the prognostic model 0.733(95%CI:0.673-0.793, P < 0.005) (Figure 6A, 7A). In the validation cohort, the nomogram model achieved a C-index of 0.776(95% CI: 0.709~0.844), which was significantly higher than that of the prognostic model 0.772 (95%CI:0.691~0.853, $P = 0.455$) (Figure 6E, 7B). Calibration curves for the probability of survival at 1-, 3-, 5-years showed optimal agreement between the prediction established in the nomogram and the actual observation in the two cohorts (Figure6C-D, 6F-H). The RMS curves showed a larger slope in the primary cohort for nomogram, indicating superior estimation of survival with nomogram (Figure 7).

The correlations among the variables in the nomogram model

The correlations among variables of the nomogram model were showed in Figure 8. In this figure, blue displayed positive correlations and red displayed negative corrections. Moreover, the correlation coefficients were proportional to the color intensity and the circle size. In the primary cohort, there was high significant between EBV DNA and risk score. Meanwhile, treatment was moderately associated with TNM stage. Coincidentally, we were able to get consistent results in the validation group.

Discussion

The TNM stage is commonly used for predicting prognosis and guiding clinical therapeutic regimen in many cancers. This system for NPC was updated and refined to the 8th edition in 2016[2]. However, this system has several controversies because it is completely based on the anatomical extent of the cancer, and neglect the biological heterogeneity of NPC patients. Many other important risk factors should be taken into account in current staging systems.

In the present study, we used random forest model to investigate the prognostic value of many clinical factors and selected the most significant ones. We revealed that EBV DNA, dNLR and HGB levels could be used for prediction of NPC prognosis. The established risk score model including EBV DNA, dNLR, and HGB had higher AUC and C-index than TNM stage, treatment, and EBV DNA model in the primary and validation cohort. Based on the risk score, we stratified NPC patients into two subgroups, high-risk and low-risk, which had significances in OS outcomes. Moreover, according to the results of random forest model analysis, we established nomograms predicting OS in NPC patients, which integrated risk score, TNM stage, and treatment. The nomogram model showed better predictive accuracy [C-index:0.783(95% CI: 0.730 ~ 0.836)] than the risk score model [C-index:0.733 (95% CI: 0.673 ~ 0.793) (P = 0.0x05)] in the primary cohort. However, there is no difference of C-index between nomogram [C-index:0.776(95% CI: 0.709 ~ 0.844)]and the risk score model [C-index:0.772 (95%CI:0.691 ~ 0.853) (P = 0.455)] in the validation cohort, possibly due to our small size of NPC patients in the cohort.

The infection of EBV is common in NPC in endemic areas. The level of plasma EBV DNA has been shown to be the most attractive potential biomarker for predicting the prognosis and providing accurate risk stratification in NPC[20]. Intriguingly, a recent prospective screening study involving 20174 participants showed that plasma EBV DNA detection was useful for nasopharyngeal carcinoma screening, with 97.1% sensitivity and 98.6% specificity. Nasopharyngeal carcinoma was detected significantly earlier by EBV DNA, with a significantly higher proportion of stage I or II disease than in a historical cohort (71% vs 20%), and had superior 3-year progression-free survival (97% vs. 70%; hazard ratio, 0.10)[21]. However, the EBV DNA alone for prognosis has limitations as the methodology of EBV DNA measurement is not globally standardized and the measurement is not routinely available in many medical institutions. Moreover, there is accumulating evidence indicating that inflammation plays an important role in carcinogenesis and tumor proliferation[22]. There has been reported that inflammation-based markers could be used as a potential prognostic factor for many cancers, such as CRP, neutrophils, lymphocytes, dNLR, ALB[23–25]. Neutrophils secreted proangiogenic cytokines, including IL-8, MMP-9, MMP-8, and VEGF, and these cytokines contribute to tumor angiogenesis and progression[26, 27]. Lymphocyte, especially the CD8 + T cell, which mediated immune response increasing OS of patients with gallbladder cancer[28]. Low HGB is a risk factor of cancer patient survival, and HGB level is an important predictor anemia evaluation and treatment[29, 30]. In our study, we used random forest to identify that EBV DNA, dNLR and HGB levels could be used for prediction of NPC prognosis.

There are several limitations in this study. First, this study is a retrospective research, and there may be a selection bias during data collection. Second, this is a single-center study with a limited number of NPC patients. Third, this study established models for predicting OS of patients with NPC, but the models of

disease-free survival were unknown. Therefore, our next aiming is to validate our models on a large-scale with multi-center study.

Conclusions

this study established a risk score model based on EBV DNA, dNLR and HGB levels. Compared with TNM stage, treatment and EBV DNA models, the risk score model achieved higher AUC. This easy-to-use scoring prognostic model may provide more precise estimation for clinicians and patients.

Abbreviations

BMI: body mass index; TNM: Tumor Node Metastasis stage; WBC: white blood cell; HGB: hemoglobin; NLR: neutrophil/lymphocyte ratio; dNLR: neutrophil/WBC-neutrophil ratio; LMR: lymphocyte/monocyte ratio; PLR: platelet/lymphocyte ratio; SII: systemic immune-inflammation index; TP: total protein; ALB: albumin; GLOB: globulin; AGR: ALB/GLOB ratio; CRP: C-reactive protein; CAR: C-reactive protein/albumin ratio; APOA: apolipoprotein AI; APOB: apolipoprotein B; ABR: APOA/APOB ratio; LDH: lactic dehydrogenase; HDL: high density lipoprotein; EBV: Epstein-Barr virus; VCA-IgA: viral capsid antigen specific immunoglobulin A; EA-IgA: early antigen immunoglobulin A; PNI: prognostic nutritional index; PI: prognostic index.

Declarations

Authors' contributions

QXX, YBL, and MMJ are design this study. NX, and GPO collected clinical data and wrote the manuscript. WGM, LNJ, and JHS performed data analysis. All authors read and approved the final manuscript.

Funding:

This study was supported by Science and Technology Department of Henan Province (Grant No. LHGJ20190637).

Availability of data and materials

The datasets analyzed during the current study are not publicly available due to patient privacy concerns, but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was approved by the Clinical Research Ethics Committee of Sun Yat-sen University Cancer Center.

Consent for publication

Not applicable.

Competing Interests:

The authors have declared no conflicts of interest.

References

1. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. Lancet. 2019;394(10192):64–80.
2. Tang LL, Chen YP, Mao YP, Wang ZX, Guo R, Chen L, Tian L, Lin AH, Li L, Sun Y, et al: **Validation of the 8th Edition of the UICC/AJCC Staging System for Nasopharyngeal Carcinoma From Endemic Areas in the Intensity-Modulated Radiotherapy Era**. *J Natl Compr Canc Netw* 2017, **15**(7):913–919.
3. Wei KR, Zheng RS, Zhang SW, Liang ZH, Li ZM, Chen WQ. Nasopharyngeal carcinoma incidence and mortality in China, 2013. Chin J Cancer. 2017;36(1):90.
4. Colevas AD, Yom SS, Pfister DG, Spencer S, Adelstein D, Adkins D, Brizel DM, Burtness B, Busse PM, Caudell JJ, et al: **NCCN Guidelines Insights**: **Head and Neck Cancers**, **Version 1**.2018. *J Natl Compr Canc Netw* 2018, **16**(5):479–490.
5. Wierzbicka M, Napierala J. Updated National Comprehensive Cancer Network guidelines for treatment of head and neck cancers 2010–2017. Otolaryngol Pol. 2017;71(6):1–6.
6. Liu MZ, Tang LL, Zong JF, Huang Y, Sun Y, Mao YP, Liu LZ, Lin AH, Ma J. Evaluation of sixth edition of AJCC staging system for nasopharyngeal carcinoma and proposed improvement. Int J Radiat Oncol Biol Phys. 2008;70(4):1115–23.
7. Huang L, Sim AYL, Wu Y, Liang Z, Li K, Du Y, Ong EHW, Tan HQ, Wee JTS, Xie Y, et al. Lactate dehydrogenase kinetics predict chemotherapy response in recurrent metastatic nasopharyngeal carcinoma. Ther Adv Med Oncol. 2020;12:1758835920970050.
8. Li QJ, Mao YP, Guo R, Huang CL, Fang XL, Ma J, Tang LL, Chen L. A Nomogram Based on Serum Biomarkers and Clinical Characteristics to Predict Survival in Patients With Non-Metastatic Nasopharyngeal Carcinoma. Front Oncol. 2020;10:594363.
9. Feng Y, Zhang N, Wang S, Zou W, He Y, Ma JA, Liu P, Liu X, Hu C, Hou T. Systemic Inflammation Response Index Is a Predictor of Poor Survival in Locally Advanced Nasopharyngeal Carcinoma: A Propensity Score Matching Study. Front Oncol. 2020;10:575417.
10. Tang LQ, Li CF, Li J, Chen WH, Chen QY, Yuan LX, Lai XP, He Y, Xu YX, Hu DP, et al: **Establishment and Validation of Prognostic Nomograms for Endemic Nasopharyngeal Carcinoma**. *J Natl Cancer Inst* 2016, 108(1).
11. Zhao L, Chen H, Hu B, Zhang H, Lin Q. Prognostic significance of Ki67 expression and the derived neutrophil-lymphocyte ratio in nasopharyngeal carcinoma. Cancer Manag Res. 2018;10:1919–26.
12. Guo SS, Tang LQ, Chen QY, Zhang L, Liu LT, Huang PY, Cao KJ, Guo L, Mo HY, Guo X, et al: **Is Hemoglobin Level in Patients with Nasopharyngeal Carcinoma Still a Significant Prognostic Factor**

in the Era of Intensity-Modulated Radiotherapy Technology? *PLoS One* 2015, **10**(8):e0136033.

13. He B, Ji T, Zhang H, Zhu Y, Shu R, Zhao W, Wang K. MRI-based radiomics signature for tumor grading of rectal carcinoma using random forest model. Journal of cellular physiology. 2019;234(11):20501–9.
14. Huang Z, Xing S, Zhu Y, Qu Y, Jiang L, Sheng J, Wang Q, Xu S, Xue N. Establishment and Validation of Nomogram Model Integrated With Inflammation-Based Factors for the Prognosis of Advanced Non-Small Cell Lung Cancer. Technology in cancer research treatment. 2020;19:1533033820971605.
15. Li C, Qin X, Yang Z, Guo W, Huang R, Wang H, Wang H. A nomogram to predict the incidence of permanent stoma in elderly patients with rectal cancer. Annals of translational medicine. 2021;9(4):342.
16. Ding Y, Chen Y, Wu M, Li L, Huang Y, Wang H, Wang H, Yu X, Xu N, Teng L. Identification of genes associated with gastric cancer survival and construction of a nomogram to improve risk stratification for patients with gastric cancer. Oncology letters. 2020;20(1):215–25.
17. He X, Xue N, Liu X, Tang X, Peng S, Qu Y, Jiang L, Xu Q, Liu W, Chen S. A novel clinical model for predicting malignancy of solitary pulmonary nodules: a multicenter study in chinese population. Cancer cell international. 2021;21(1):115.
18. Chen JH, Zhai ET, Yuan YJ, Wu KM, Xu JB, Peng JJ, Chen CQ, He YL, Cai SR. Systemic immune-inflammation index for predicting prognosis of colorectal cancer. World journal of gastroenterology. 2017;23(34):6261–72.
19. Kasymjanova G, MacDonald N, Agulnik JS, Cohen V, Pepe C, Kreisman H, Sharma R, Small D. The predictive value of pre-treatment inflammatory markers in advanced non-small-cell lung cancer. Current oncology. 2010;17(4):52–8.
20. Li W, Chen J, Liang B, Li Z, Li J, Yuan X, Wu S, Zeng F, Peng X, Li Y, et al. Long-term monitoring of dynamic changes in plasma EBV DNA for improved prognosis prediction of nasopharyngeal carcinoma. Cancer medicine. 2021;10(3):883–94.
21. Chan KCA, Woo JKS, King A, Zee BCY, Lam WKJ, Chan SL, Chu SWI, Mak C, Tse IOL, Leung SYM, et al. Analysis of Plasma Epstein-Barr Virus DNA to Screen for Nasopharyngeal Cancer. N Engl J Med. 2017;377(6):513–22.
22. McMillan DC. Cancer and systemic inflammation: stage the tumour and stage the host. Br J Cancer. 2013;109(3):529.
23. Proctor MJ, McMillan DC, Morrison DS, Fletcher CD, Horgan PG, Clarke SJ. A derived neutrophil to lymphocyte ratio predicts survival in patients with cancer. Br J Cancer. 2012;107(4):695–9.
24. Partl R, Lukasiak K, Thurner EM, Renner W, Stranzl-Lawatsch H, Langsenlehner T. **The Elevated Pre-Treatment C-Reactive Protein Predicts Poor Prognosis in Patients with Locally Advanced Rectal Cancer Treated with Neo-Adjuvant Radiochemotherapy**. *Diagnostics* 2020, 10(10).
25. Guo S, He X, Chen Q, Yang G, Yao K, Dong P, Ye Y, Chen D, Zhang Z, Qin Z, et al. The C-reactive protein/albumin ratio, a validated prognostic score, predicts outcome of surgical renal cell carcinoma patients. BMC Cancer. 2017;17(1):171.

26. Kusumanto YH, Dam WA, Hospers GA, Meijer C, Mulder NH. Platelets and granulocytes, in particular the neutrophils, form important compartments for circulating vascular endothelial growth factor. Angiogenesis. 2003;6(4):283–7.
27. Weitzman SA, Gordon LI. Inflammation and cancer: role of phagocyte-generated oxidants in carcinogenesis. Blood. 1990;76(4):655–63.
28. Fluxa P, Rojas-Sepulveda D, Gleisner MA, Tittarelli A, Villegas P, Tapia L, Rivera MT, Lopez MN, Catan F, Uribe M, et al. High CD8(+) and absence of Foxp3(+) T lymphocytes infiltration in gallbladder tumors correlate with prolonged patients survival. BMC Cancer. 2018;18(1):243.
29. Rodgers GM 3rd, Becker PS, Blinder M, Cella D, Chanan-Khan A, Cleeland C, Coccia PF, Djulbegovic B, Gilreath JA, Kraut EH, et al. Cancer- and chemotherapy-induced anemia. J Natl Compr Canc Netw. 2012;10(5):628–53.
30. Wan S, Lai Y, Myers RE, Li B, Palazzo JP, Burkart AL, Chen G, Xing J, Yang H. Post-diagnosis hemoglobin change associates with overall survival of multiple malignancies - results from a 14-year hospital-based cohort of lung, breast, colorectal, and liver cancers. BMC Cancer. 2013;13:340.

Tables

Table 1. Demographics and clinical characteristics of patients in the primary and

validation cohort

Characteristic	Primary cohort	Validation cohort
	n= (363) No. (%) or Mean ± sd	n= (156) No. (%) or Mean ± sd
Gender		
Male	209 (57.57%)	92 (58.97%)
Female	154 (42.43%)	64 (41.03%)
Age (years)	46.05±10.87	46.87±11.58
Smoking		
No	277 (76.31%)	123 (78.85%)
Yes	86 (23.69%)	33 (21.15%)
Family history		
Yes	87 (23.97%)	50 (32.05%)
No	276 (76.03%)	106 (77.95%)
BMI (kg/m^2)	23.17±6.74	22.96±3.37
TNM stage[a]		
I	11 (3.03%)	6 (3.85%)
II	47 (12.95%)	22 (14.10%)
III	179 (49.31%)	69 (44.23%)
IV	126 (34.71%)	59 (37.82%)
Treatment		
Radiotherapy	300 (82.64%)	129 (82.69%)
Chemotherapy	63 (17.36%)	27 (17.31%)
WBC ($10^9/L$)	7.03±3.33	7.04±3.32
Neutrophils ($10^9/L$)	4.55±2.87	4.35±2.25
Lymphocyte ($10^9/L$)	1.71±0.73	1.65±0.81
Monocyte ($10^9/L$)	0.47±0.24	0.43±1.20
Platelet ($10^9/L$)	225.02±69.18	214±67.46

HGB (g/L)	136.63±15.84	137.38±15.62
NLR	3.33±3.82	3.34±2.87
dNLR	2.30±2.28	2.48±4.13
LMR	4.67±4.59	4.58±3.49
PLR	157.39±89.41	156.89±86.88
SII	757.39±822.18	719.40±692.30
TP (g/L)	73.28±5.89	75.56±5.24
ALB (g/L)	43.40±3.29	43.17±3.23
GLOB (g/L)	29.88±4.67	29.39±4.50
AGR	1.49±0.24	1.50±0.26
CRP (mg/L)	4.69±9.93	5.42±9.81
CAR	0.11±0.26	0.13±0.27
APOA (g/L)	1.31±0.25	1.32±0.27
APOB (g/L)	0.99±0.24	0.98±0.25
ABR	1.40±0.42	1.42±0.47
LDH (U/L)	174.47±54.26	176.76±119.25
HDL (U/L)	1.23±0.31	1.221±0.31
EBV DNA (copy/mL)		
$<10^3$	167 (48.8%)	72 (40.5%)
10^3-9,999	77 (20.8%)	31 (20.8%)
10^4-99,999	66 (16.8%)	31 (22.5%)
10^5-999,999	36 (8.4%)	10 (9.8%)
$\geq 10^6$	17 (5.2%)	12 (6.4%)
VCA-IgA		
<1:80	59 (17.1%)	28 (16.2%)
1:80–1:320	221 (60.1%)	93 (61.3%)
≥ 1:640	83 (22.8%)	35 (22.5%)
EA-IgA		
<1:10	111 (32.7%)	51 (28.3%)

1:10–1:20	121 (31.8%)	49 (34.7%)
≥1:40	83 (35.5%)	56 (37.0%)
PNI	51.94±5.09	51.4±5.17
PI		
0	297 (81.82%)	119 (76.28%)
1	59 (16.25%)	35 (22.44%)
2	7 (1.93%)	2 (1.28%)

a: TNM stage was classified according to the AJCC 8th TNM staging system;

Abbreviations: BMI: body mass index; TNM: Tumor Node Metastasis stage; WBC: white blood cell; HGB: hemoglobin; NLR: neutrophil/lymphocyte ratio; dNLR: neutrophil/WBC-neutrophil ratio; LMR: lymphocyte/monocyte ratio; PLR: platelet/lymphocyte ratio; SII: systemic immune-inflammation index; TP: total protein; ALB: albumin; GLOB: globulin; AGR: ALB/GLOB ratio; CRP: C-reactive protein; CAR: C-reactive protein/albumin ratio; APOA: apolipoprotein AI; APOB: apolipoprotein B; ABR: APOA/APOB ratio; LDH: lactic dehydrogenase; HDL: high density lipoprotein; EBV: Epstein-Barr virus; VCA-IgA: viral capsid antigen specific immunoglobulin A; EA-IgA: early antigen immunoglobulin A; PNI: prognostic nutritional index; PI: prognostic index.

Table 2. The C-index of the prognostic model, TNM staging, Treatment, and EBV DNA for prediction of OS in the training cohort and validation cohort

Factors	C-index (95% CI)	*P*
For training cohort		
Our model	0.733 (0.673~0.793)	
TNM staging	0.712 (0.657~0.768)	
Treatment	0.542 (0.505~0.580)	
EBV DNA	0.691 (0.626~0.756)	
Prognostic model vs TNM staging		0.531
Prognostic model vs Treatment		<0.001
Prognostic model vs EBV DNA		0.035
For validation cohort		
Our model	0.772 (0.691~0.853)	
TNM staging	0.699 (0.628~0.770)	
Treatment	0.551 (0.503~0.600)	
EBV DNA	0.739 (0.652~0.826)	
Prognostic model vs TNM staging		0.099
Prognostic model vs Treatment		<0.001
Prognostic model vs EBV DNA		0.259

C-index = concordance index; CI = confidence interval; P values are calculated based on normal approximation using function rcorrp.cens in Hmisc package.

Table 3. OS and OS rate in high-risk and low-risk groups according to the established model risk score in the primary and validation cohort

Parameter	Primary cohort			Validation cohort		
	High-Risk Group	Low-Risk Group	Total	High-Risk Group	Low -Risk Group	Total
No. of patients	150	213	363	71	85	156
Median (IQR)	46.5 (27.0-65.9)	61.0 (45.1-68)	51.0 (42.3-66.7)	45.8 (30.5-66.0)	61.2 (45.3-66.0)	50.4 (41.7-66.0)
No. of OS						
1-Year	135 (90.0%)	211 (99.0%)	345 (95.0%)	69 (97.2%)	84 (98.8%)	154 (98.7%)
3-Year	107 (71.3%)	198 (93.0%)	305 (84.0%)	50 (70.4%)	81 (95.3%)	131 (84.0%)
5-Year	56 (37.3%)	114 (53.5%)	170 (46.8%)	25 (35.2%)	46 (54.1%)	71 (45.5%)

Abbreviations: OS: overall survival; IQR: interquartile range.

Figures

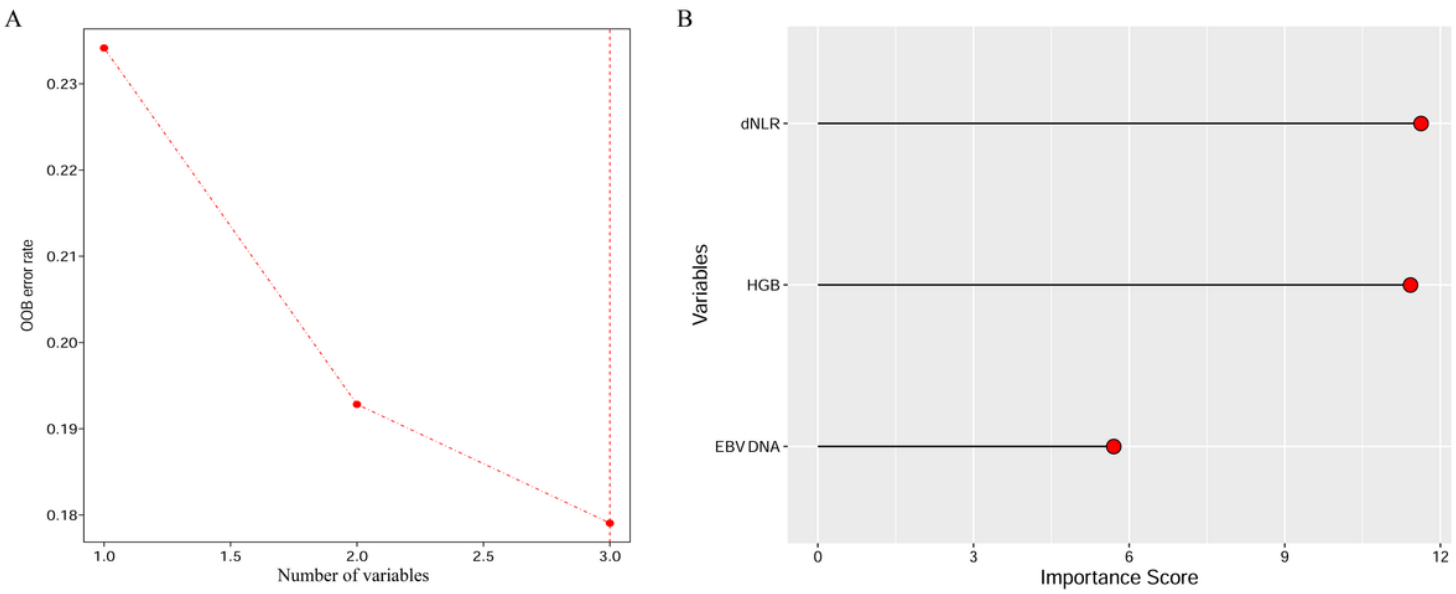

Figure 1

Ranger provides variable importance score for variables for NPC patients in the primary cohort. 'Out of bag (OOB)' error rate of top 3 variables in the model, when probes were included one by one based on their variable importance score ranks.

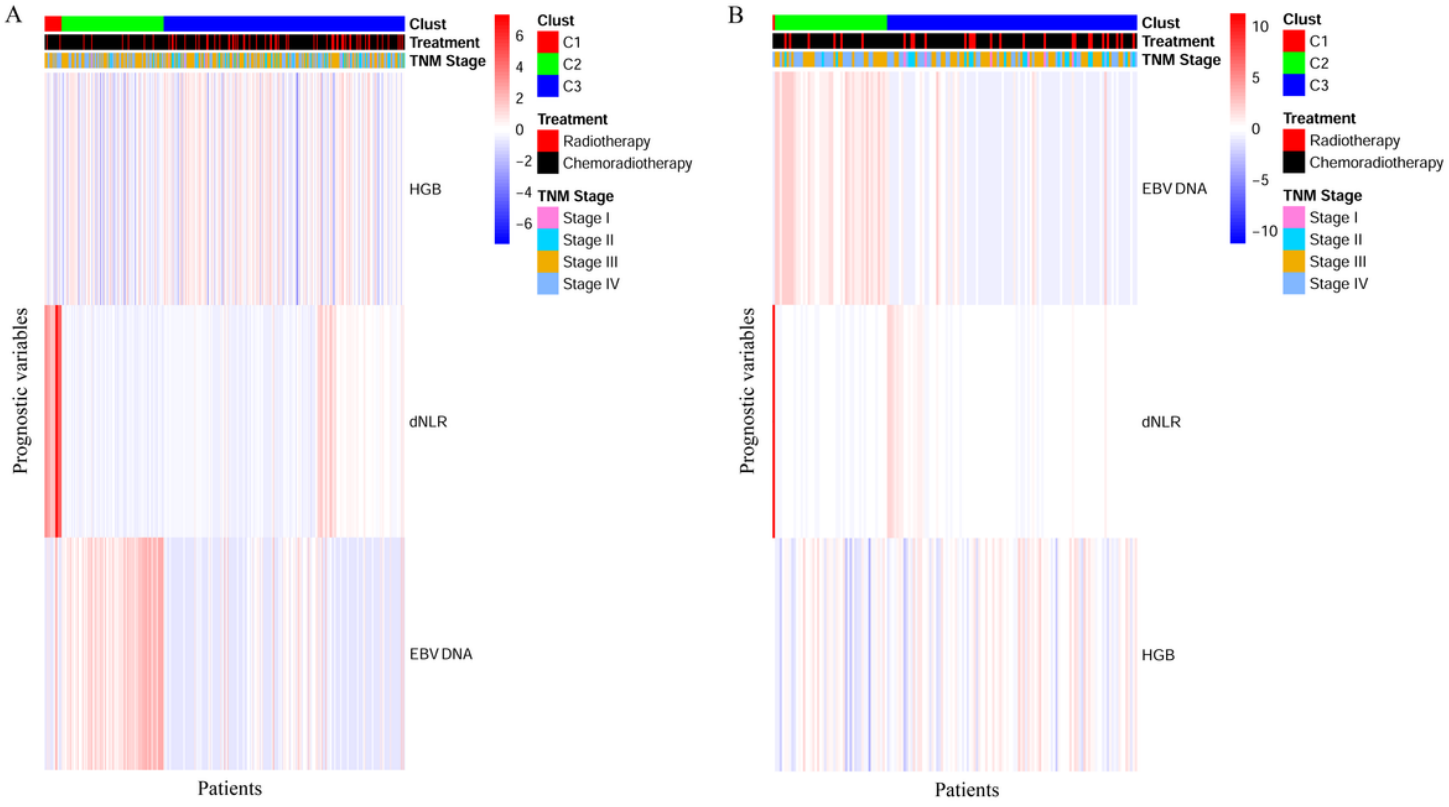

Figure 2

Heatmap were generated by unsupervised hierarchical clustering of 3 features (Y axis) across identified NPC patients on X-axis revealed 3 major image feature patterns in the primary cohort (A) and in the validation cohort (B). The corresponding treatment and TNM stage that the tumor was derived from are shown above the color bars.

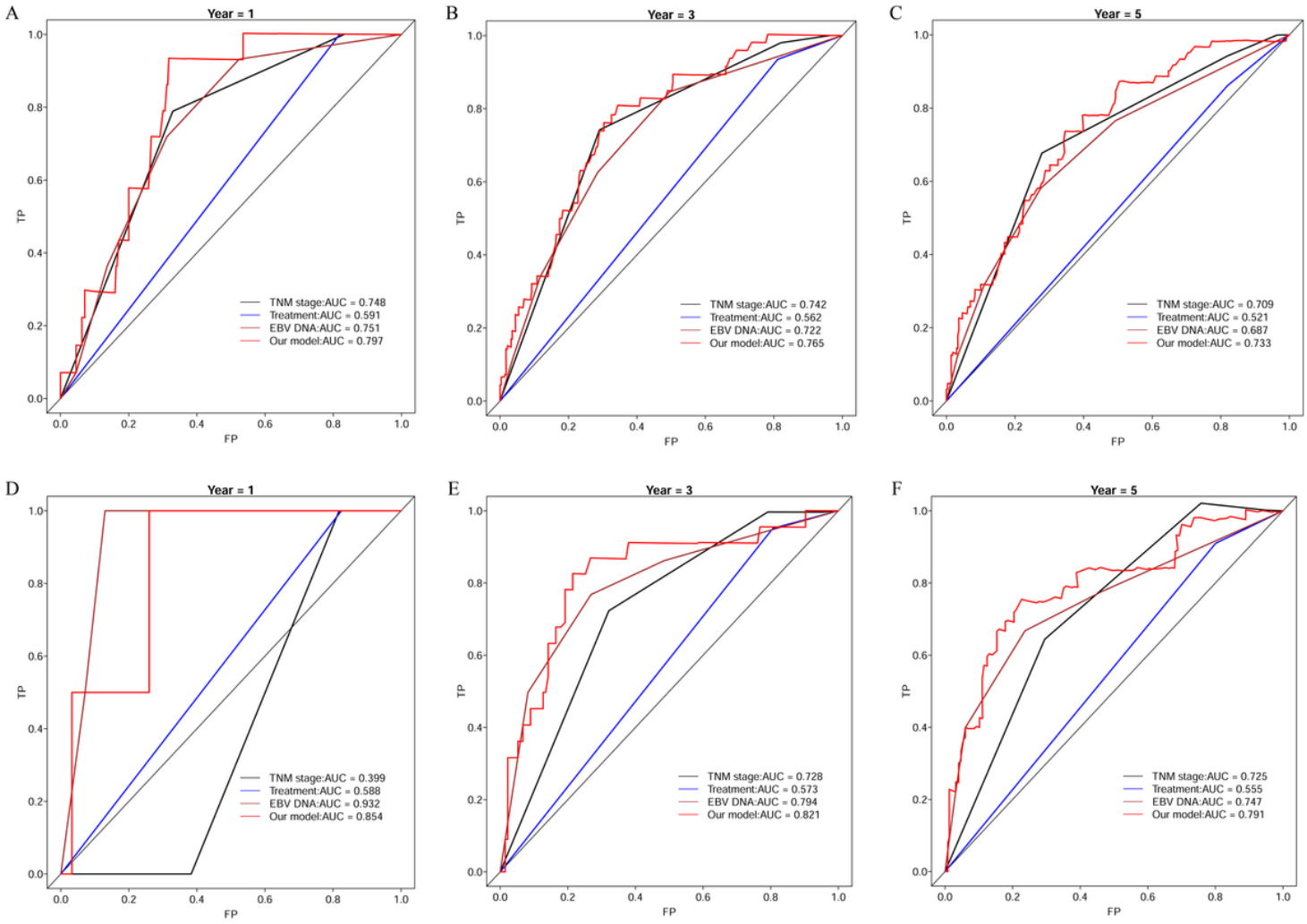

Figure 3

Comparison of AUC among TNM stage, treatment, EBV DNA, and our model in 1-year overall survival (OS), 3-year OS, and 5-year OS in the primary (A, B, C) cohort and validation cohort (D, E, F).

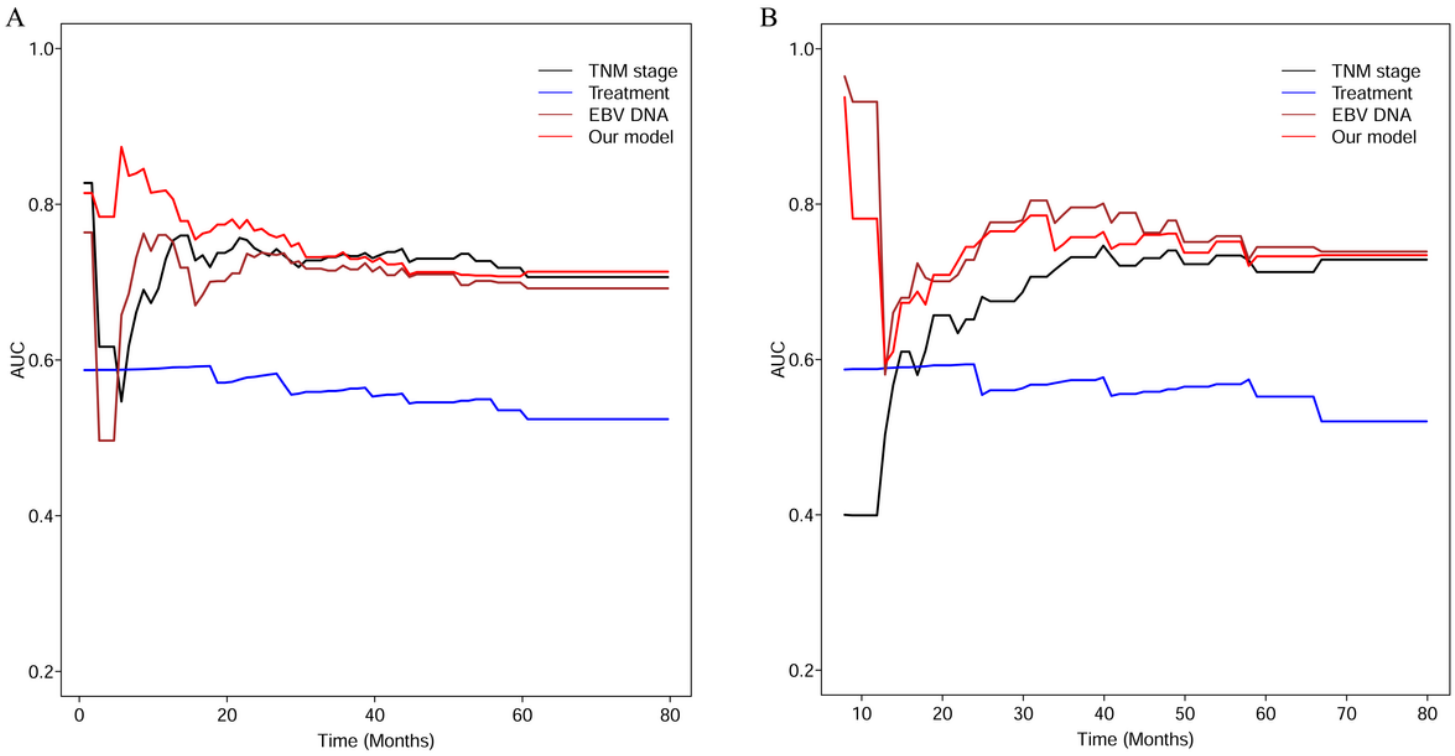

Figure 4

Time-dependent ROC curve for OS in the primary cohort (A)and validation cohort(B). ROC, receiving operative characteristics; OS, overall survival.

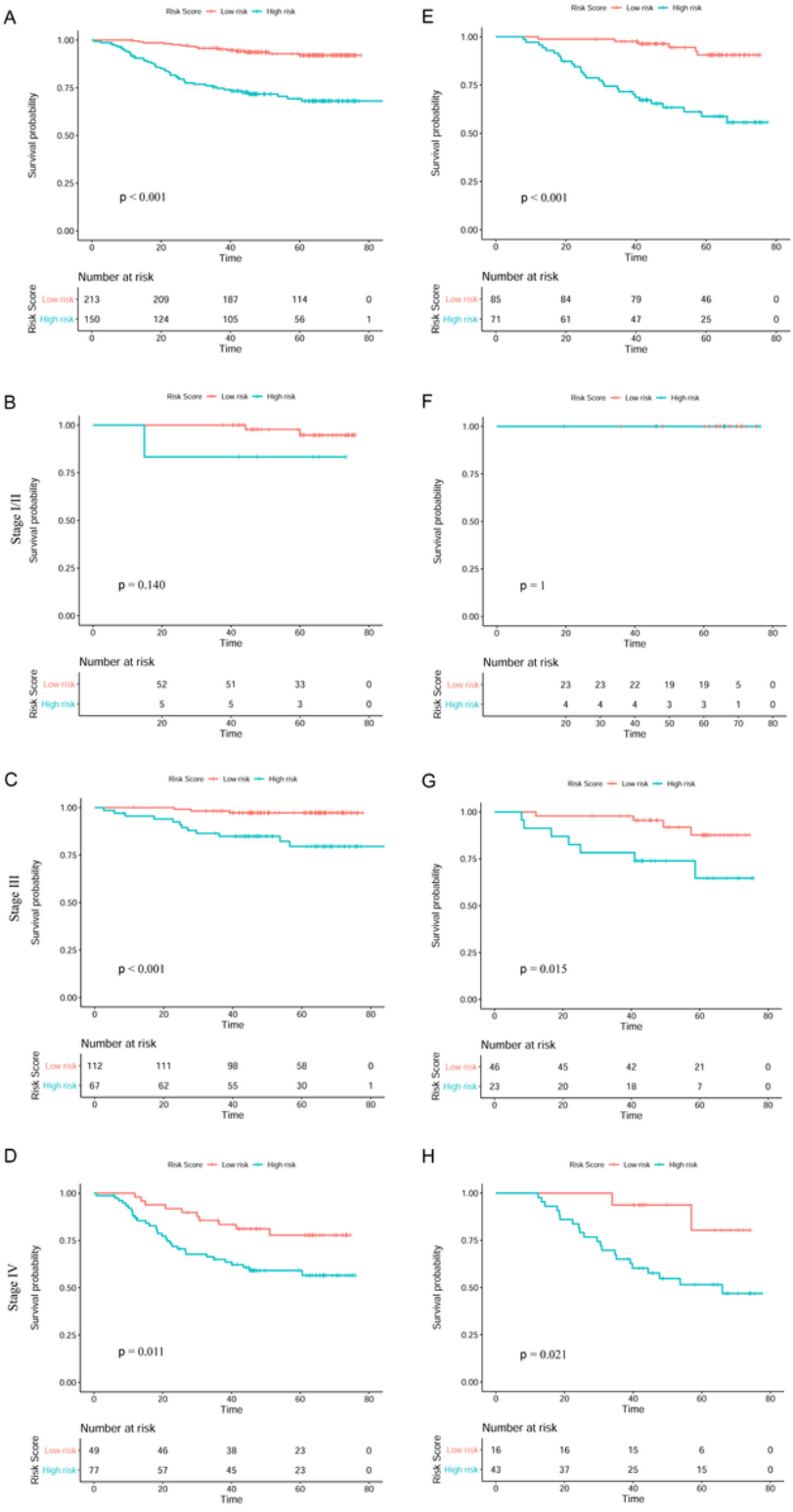

Figure 5

Risk group stratification within I/II, III and IV stage in the patients of NPC. Kaplan-Meier curves of OS according to the risk score in the primary cohort and validation cohort.

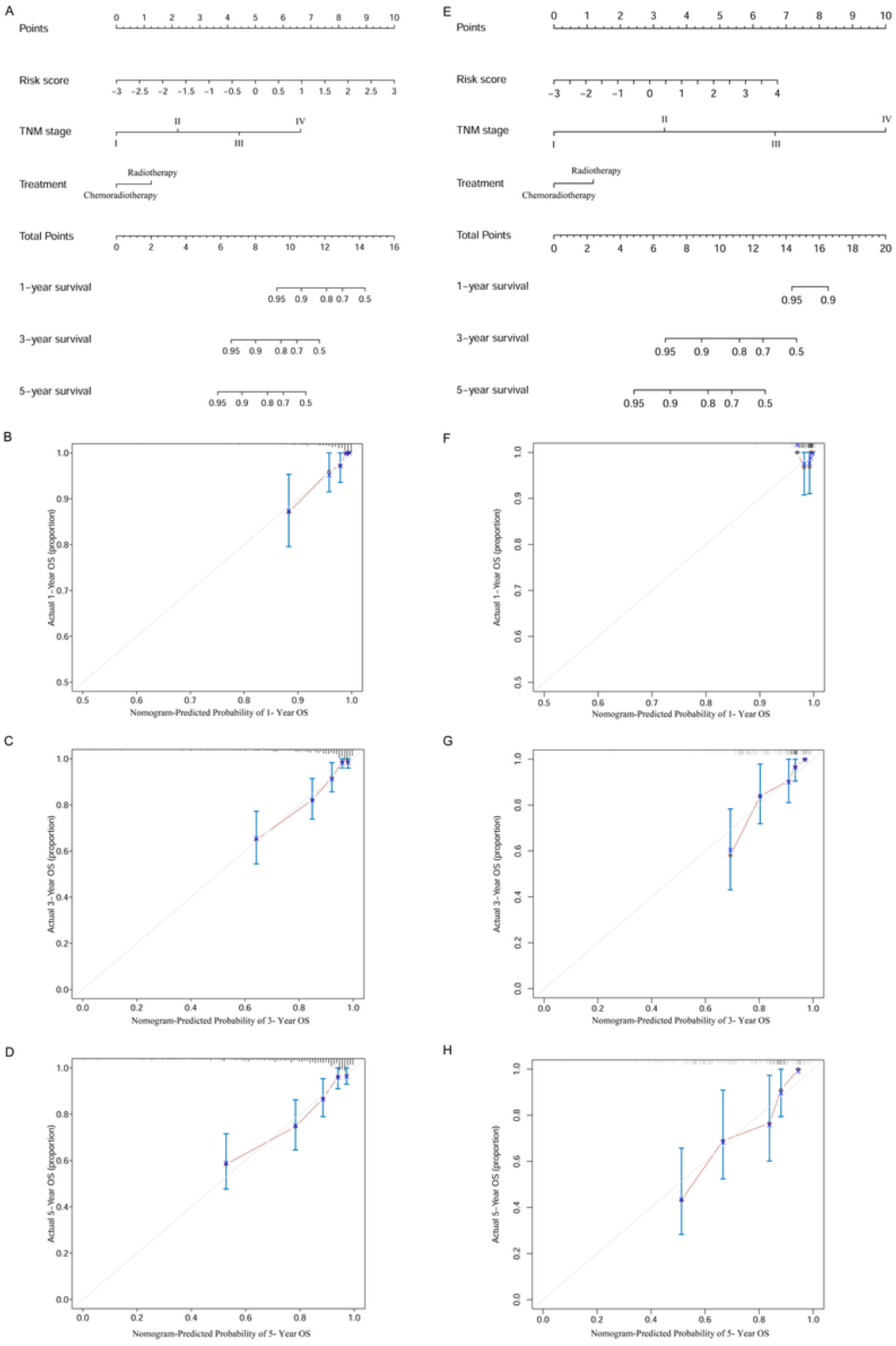

Figure 6

The prognostic model based on risk score, TNM stage and treatment predicting OS in primary cohort (A left) and validation cohort (A right). The calibration curves for predicting patient OS at 1-year, 3-year, 5-year in the primary cohort and in the validation cohort. Total points projected on the bottom scales indicate the probability of 1- , 3-year and 5-year survival.

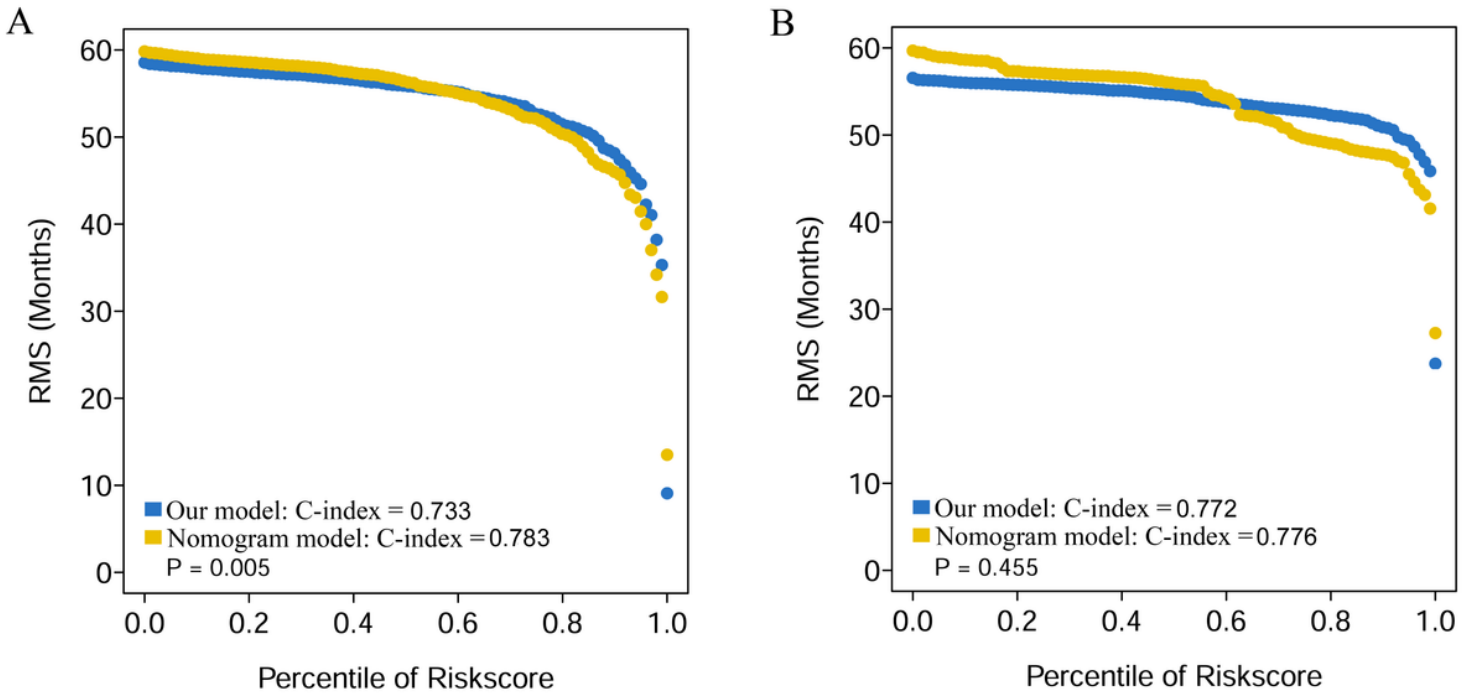

Figure 7

Restricted mean survival (RMS) curves for our model and the nomogram model in primary cohort(A) and validation cohort(B). Each point represents the RMS time of corresponding our model and nomogram model scores.

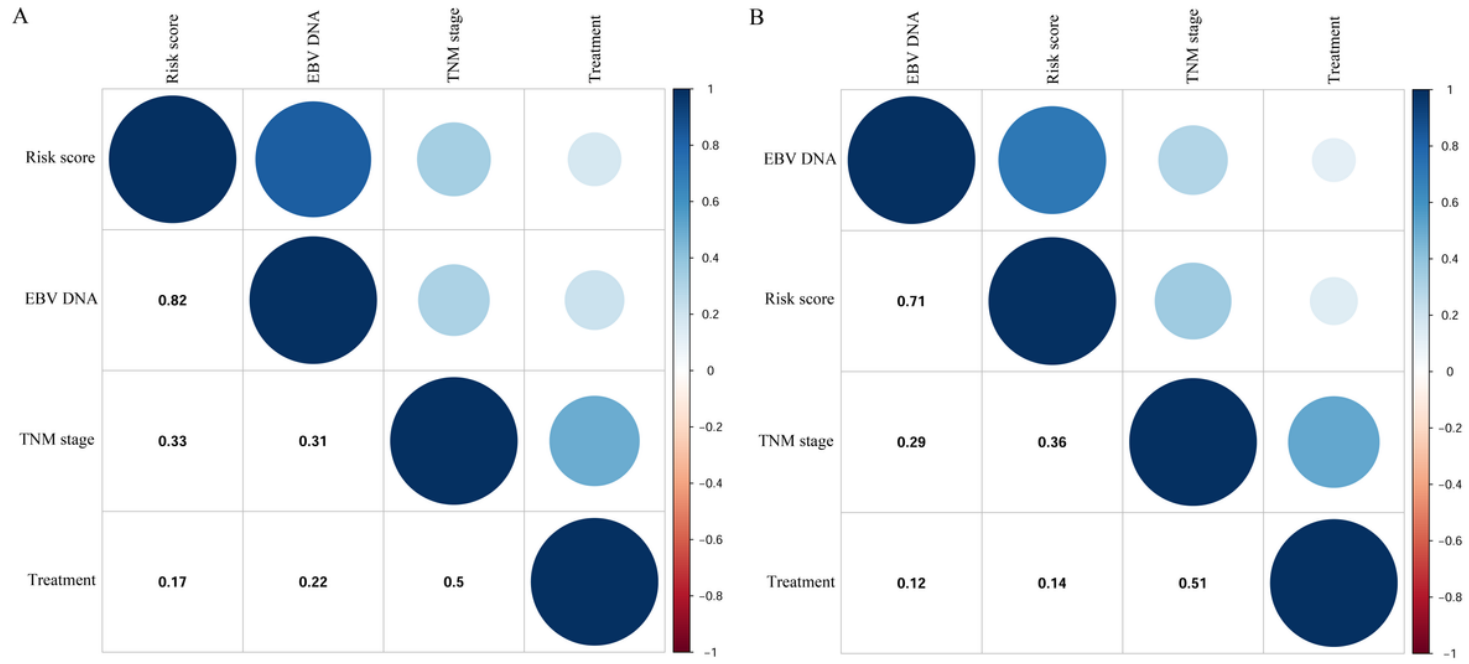

Figure 8

The correlations among the various variables of the nomogram model.