Am J Cancer Res 2022;12(4):1635-1647
www.ajcr.us /ISSN:2156-6976/ajcr0141293

Original Article

Construction and validation of a biochemical signature to predict the prognosis and the benefit of induction chemotherapy in patients with nasopharyngeal carcinoma

Xue-Song Sun[1,2*], Sai-Lan Liu[1,2*], Si-Yi Xie[1,2*], Rui Sun[1,2], Dong-Hua Luo[1,2], Qiu-Yan Chen[1,2], Hai-Qiang Mai[1,2]

[1]Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou, Guangdong 510060, P. R. China; [2]Department of Nasopharyngeal Carcinoma, Sun Yat-sen University Cancer Center, 651 Dongfeng Road East, Guangzhou, Guangdong 510060, P. R. China. [*]Equal contributors.

Received December 23, 2021; Accepted March 16, 2022; Epub April 15, 2022; Published April 30, 2022

Abstract: This study aimed to develop and validate a biochemical signature for predicting the prognosis of patients with nasopharyngeal carcinoma (NPC) and explore roles of the constructed signature for screening optimal candidates for induction chemotherapy (IC). The biochemical signature was constructed based on a retrospective cohort of 3742 patients from January 2008 to December 2010; 2078 patients from prospective studies from January 2011 to December 2012 and 2153 patients from January 2013 to December 2016 served as validation cohort A and validation cohort B. Overall survival (OS) was the primary endpoint. The least absolute shrinkage and selection operator coefficients on the Cox regression model were calculated to construct the prediction model with the data of 33 biochemical indicators. A total of six prognostic indicators, including sodium, alkaline phosphatase, lactate dehydrogenase, albumin, indirect bilirubin, and cystatin-C, were screened for constructing the biochemical signature. The patients were divided into low-risk and high-risk groups using an optimal cut-off value of 0.823. The patients in high-risk group had significantly lower OS and distant metastasis-free survival (DMFS) compared with patients in low-risk group in three cohorts ($P < 0.05$). Furthermore, among patients with high-risk scores in the combined cohort, the addition of IC to CCRT further improved their OS and DMFS, whereas patients with low-risk scores did not benefit from IC. Our study developed and validated a clinically useful biochemical signature that could predict the survival outcomes in NPC patients. This signature can help clinicians design personalized treatment strategies.

Keywords: Nasopharyngeal carcinoma, biochemical signature, prognosis, induction chemotherapy

Introduction

Nasopharyngeal carcinoma (NPC) is prevalent in southern China and has distinct epidemiology, pathological types, and therapeutic management compared with other malignant tumors of the head and neck mucosal sites [1]. Radiotherapy, particularly intensity-modulated radiotherapy (IMRT), is sought as the preferred irradiation technique for the treatment of NPC, and has been observed to greatly increase the local control rate of patients [2, 3]. However, distant metastasis in advanced NPC may result in treatment failure [4]. Several prospective randomized trials have demonstrated that induction chemotherapy (IC), followed by concurrent chemoradiotherapy (CCRT), is associated with a reduced risk of distant metastases, therefore providing survival benefits for NPC patients [5-7]. However, a considerable proportion of the patients do not benefit from IC [8, 9].

Currently, the method for predicting the prognosis and guiding the treatment for NPC is mainly according to the anatomical-based tumor node metastasis (TNM) staging system. Due to tumor heterogeneity, patients with the same stage and identical therapeutic regimen may show differences in clinical prognosis, suggesting that the TNM stage is insufficient to predict the prognosis of NPC, as well as the clinical benefits of IC. Therefore, it is necessary

to construct a novel prognostic tool to guide a personalized antitumor therapy, specifically by screening suitable candidates for IC.

A high degree of metabolic disorder is one of the main hallmarks of cancer, and is known to be a major player in tumor progression and metastasis [10]. Biochemical indicators for metabolic disorders include electrolytes, liver function, kidney function, blood lipids, blood glucose, and creatine kinase (CK). Moreover, emerging evidence demonstrates that biochemical markers such as lactate dehydrogenase (LDH), C-reactive protein (CRP), alkaline phosphatase (ALP), and hyponatremia may promote tumor progression and may therefore be used for the prediction of treatment responses and prognosis in a variety of cancers [11-15]. Although these studies have confirmed the prognostic value of some biochemical indicators, efforts are focused mainly on a few markers. Moreover, the sample size of these studies is relatively small, therefore affecting the reliability of the results. Moreover, the significance of these indicators in guiding a personalized treatment is not yet extensively studied.

A biochemical test is a routine examination for cancer patients before clinical treatment. With this, the data of 33 biochemical indicators are obtained simultaneously, all of which provide indications of human metabolism and internal environmental conditions. Thus, based on routine biochemical test data on nearly 8000 patients with non-metastatic NPC, we aimed to develop and validate a biochemical signature for predicting the prognosis of patients with NPC. We screened the above variables based on the least absolute shrinkage and selection operator (LASSO) on the COX regression model. The above method has been widely used in the establishment of prognostic models [16, 17]. Furthermore, we explored the possible roles of the constructed biochemical signature for screening candidates for IC.

Patients and methods

Patients

Patients with non-metastatic NPC were considered eligible for this study (N = 7973; January 2008-December 2016). A total of 3742 patients from 2008 to 2010 were retrospectively recruited as the training cohort; 2078 patients from previous prospective studies from 2011 to 2012 served as validation cohort A [18, 19]; and 2153 patients from 2013 to 2016 served as validation cohort B. The inclusion criteria are as follows: (1) age ≥ 18 years; (2) ECOG of 0 to 2; (3) newly biopsy-proven diagnosed NPC; (4) stage I-IVa; (5) no history of any antitumor therapy; (6) underwent radical radiotherapy with or without chemotherapy; (7) satisfactory liver and renal functions; (8) complete treatment and laboratory information; and (9) no pregnancy, lactation, and secondary malignant disease. The flowchart of this study is presented in **Figure 1**. The staging system of the 8th American Joint Committee on Cancer/Union for International Cancer Control (AJCC/UICC) was used to restage all patients. All procedures were approved by the Institutional Review Board of SYSUCC, and all participants provided written informed consent prior to treatment.

Clinical assessment and treatment

All patients underwent fasting before pretreatment evaluation using baseline biochemical tests. Moreover, plasma Epstein-barr virus (EBV) DNA detection was also performed for all patients prior to treatment. An automated immunoturbidimetric analyzer 7600-020 (Hitachi High-Technologies, Tokyo, Japan) was used to perform routine biochemical tests. Table S1 lists the detailed information of the 33 biochemical indicators. All patients received IMRT with or without chemotherapy. The detailed information on treatment is supplied in Supplementary Materials.

Follow-up

Nasopharyngeal endoscopy, MRI of the nasopharynx and neck area, chest radiography, and abdominal ultrasonography were used to assess the patients at follow-up. Patients were evaluated at the end of the treatment, at least every three months during the first three years and at least every six months after that. The primary endpoint of this study is the assessment of the patients' overall survival (OS), which was calculated from the date of diagnosis to death from any cause. Secondary endpoints include the locoregional relapse-free survival (LRFS) and distant metastasis-free survival (DMFS), which were defined as the

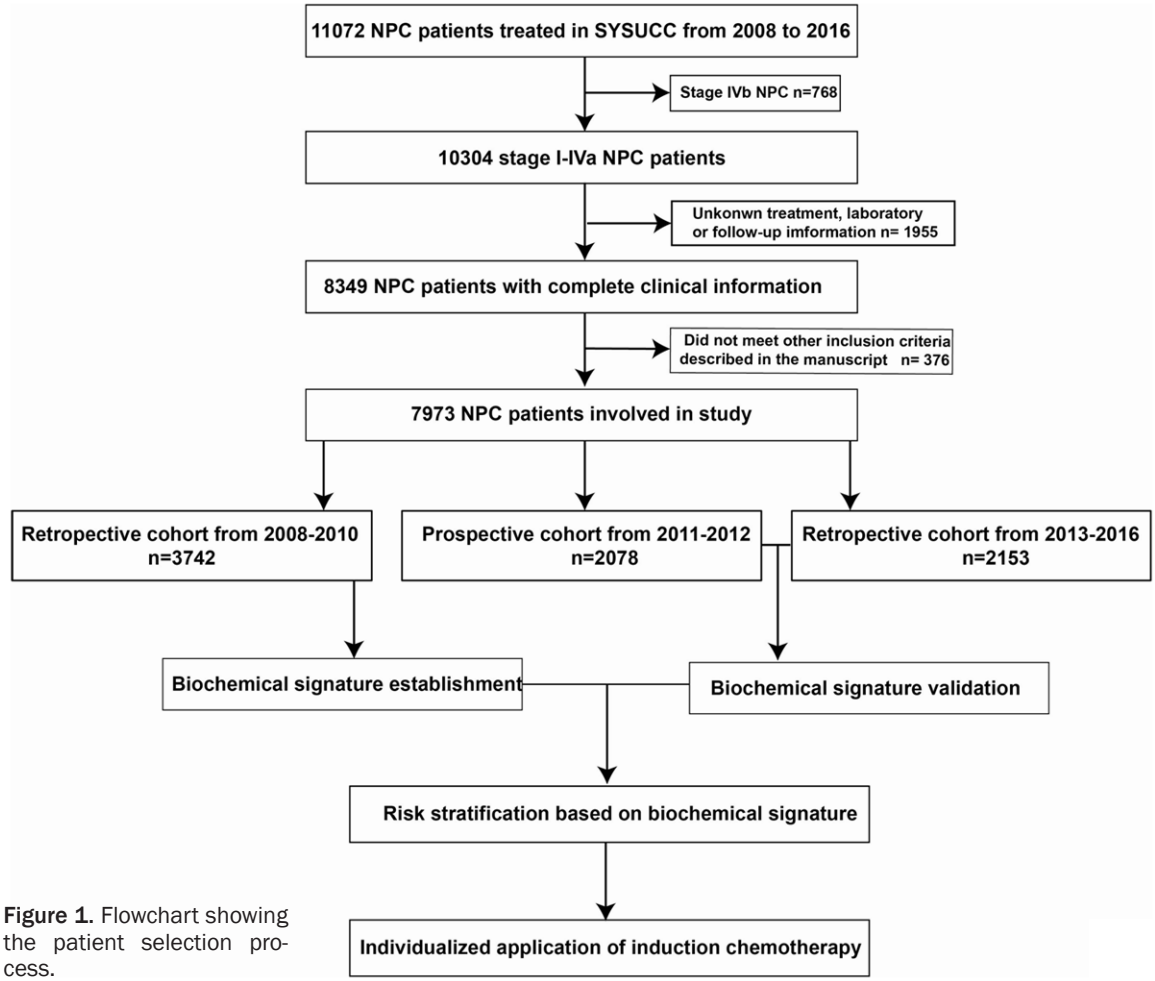

Figure 1. Flowchart showing the patient selection process.

interval between the date of diagnosis and the first event.

Statistical analysis

The obtained data from all biochemical indicators were included in the analysis. We used a penalized cox model to select indicators for constructing a biochemical signature in the training group. Using the R package *glmnet*, the LASSO coefficients on the Cox regression model were calculated to construct the prediction model. In selecting the optimum data threshold of biochemical signature, we used the X-tile software (V.3.6.1; Yale University, New Haven, Connecticut, USA) to obtain the highest χ^2 value (minimum *p*-value) defined by the Kaplan-Meier survival analysis and the log-rank test. The propensity score matching (PSM) was applied to balance potential confounders between different groups with a ratio of 1:1. The chi-square test was then used for assessing categorical variables. The time-to-event endpoints were analyzed through the Kaplan-Meier survival curves, and the differences between the groups were tested using the log-rank test. The Cox proportional hazards model was used to perform multivariate analyses and to calculate the hazard ratios (HRs). The concordance index (C-index) was applied to compare the prognostic value of the biochemical signature, EBV DNA, and TNM stage (8^{th} AJCC/UICC) [20]. Analyses were performed using R (V.3.6.0) and SPSS (V.22.0, IBM). All statistical tests were two-tailed, and the statistical significance was indicated as $P < 0.05$.

Results

The summary of patient characteristics is listed in **Table 1**. In the combined cohort, the median age of patients was 46 years old, and the ratio of males to females was 3.2:1. Stage

Table 1. Patients' characteristics

Characteristic	Combined cohort n (%)	Training cohort n (%)	Validation cohort A n (%)	Validation cohort B n (%)
Total	7973	3742	2078	2153
Age				
≤ 46	4055 (50.9)	1846 (49.3)	1044 (50.2)	1165 (54.1)
> 46	3918 (49.1)	1896 (50.7)	1034 (49.8)	988 (45.9)
Sex				
Female	2110 (26.5)	1006 (26.9)	550 (26.5)	554 (25.7)
Male	5863 (73.5)	2736 (73.1)	1528 (73.5)	1599 (74.3)
Smoking history				
No	5124 (64.3)	2272 (60.7)	1301 (62.6)	1551 (72.0)
Yes	2849 (35.7)	1470 (39.3)	777 (37.4)	602 (28.0)
NPC family history				
No	7123 (89.3)	3320 (88.7)	1856 (89.3)	1947 (90.4)
Yes	850 (10.7)	422 (11.3)	222 (10.7)	206 (9.6)
T stage*				
T1	614 (7.7)	321 (8.6)	192 (9.2)	101 (4.7)
T2	1528 (19.2)	840 (22.4)	384 (18.5)	304 (14.1)
T3	3988 (50.0)	1735 (46.6)	1043 (50.2)	1210 (56.2)
T4	1843 (23.1)	846 (22.6)	459 (22.1)	538 (25.0)
N stage*				
N0	1303 (16.3)	712 (19.0)	365 (17.6)	226 (10.5)
N1	2850 (35.7)	1295 (34.6)	836 (40.2)	719 (33.4)
N2	2943 (36.9)	1425 (38.1)	706 (34.0)	812 (37.7)
N3	877 (11.0)	310 (8.3)	171 (8.2)	396 (18.4)
Overall stage*				
I	233 (2.9)	126 (3.4)	69 (3.3)	38 (1.8)
II	982 (12.3)	525 (14.0)	283 (13.6)	174 (8.1)
III	4264 (53.5)	2007 (53.6)	1149 (55.3)	1108 (51.5)
IV	2494 (31.3)	1084 (29.0)	577 (27.8)	833 (38.7)
EBV DNA				
≤ 1500 copies/ml	3593 (45.1)	1534 (41.0)	1012 (48.7)	1047 (48.6)
1500 copies/ml	4380 (54.9)	2208 (59.0)	1066 (51.3)	1106 (51.4)
Treatment method				
RT alone	1211 (15.2)	725 (19.4)	286 (13.8)	200 (9.3)
CCRT	2901 (36.4)	1213 (32.4)	854 (41.1)	834 (38.7)
IC + RT	1322 (16.6)	844 (22.6)	331 (15.9)	147 (6.8)
IC + CCRT	2362 (29.6)	857 (22.9)	571 (27.5)	934 (43.4)
Others	177 (2.2)	103 (2.8)	36 (1.7)	38 (1.8)

P value was calculated with the Pearson χ^2 test. *According to the 8th edition of the UICC/AJCC staging system. Abbreviations: NPC, Nasopharyngeal Carcinoma; EBV DNA, Epstein-Barr Virus DNA; RT, Radiotherapy; CCRT, Concurrent Chemoradiotherapy; IC, Induction Chemotherapy.

III-IV patients accounted for more than 80% in each cohort. A total of 6762 (84.8%) patients received radiotherapy combined with chemotherapy. CCRT (36.4%) and IC + CCRT (29.6%) were the two most common treatment methods, as observed in the combined cohort. The cut-off value of pretreatment EBV DNA (1500 copies/ml) was based on a previous study [21]. There were 54.9% of patients with an EBV DNA load of more than 1500 copies/ml.

The LASSO Cox regression model was used to build a biochemical signature for predicting death in the training cohort (Figure S1). The six most useful prognostic indicators, namely sodium (Na), alkaline phosphatase (ALP), lactate dehydrogenase (LDH), albumin (ALB), indirect bilirubin (IBIL), and cystatin-C (CYSC), were selected for further analysis. Each patient had a risk score, calculated based on a formula derived from the levels of these biochemical indicators weighted by their corresponding regression coefficient. The formula is as follows:

Risk score = [-0.00095 × level of Na (mmol/L)] + [0.0036 × level of ALP (U/L)] + [0.00058 × level of LDH (U/L)] - [0.045 × level of ALB (g/L)] - [0.010 × level of IBIL (umol/L)] + [0.278 × level of CYSC (mg/L)].

The patients were divided into low-risk and high-risk groups using an optimal cut-off value of 0.823, which was generated from the X-tile plots.

In the training cohort, 3260 patients were included in the low-risk group, while the remaining 482 patients comprised the high-risk group. Through the Kaplan-Meier curves, we found that patients in the high-risk group had significantly lower three-year OS and DMFS compared with the patients in the low-risk group. The three-year OS of the low- and high-risk groups were 93.4% and 81.9%, respectively ($P_{log\text{-}rank}$ < 0.001); and the three-year DMFS were 90.2% and 77.1%, respectively ($P_{log\text{-}rank}$ < 0.001). However, there was no significant difference in the LRFS between the two groups ($P_{log\text{-}rank}$ = 0.270) (**Figure 2A-C**). Next, we confirmed the prognostic value of the established biochemical signature in the two validation cohorts. Patients were also divided into low- and high-risk using the same cut-off value. In the validation cohort A, the three-year OS was 94.8% in the low-risk group compared with 85.9% in the high-risk group ($P_{log\text{-}rank}$ < 0.001), and the three-year DMFS was 91.9% in the low-risk group compared with 80.2% in the high-risk group ($P_{log\text{-}rank}$ < 0.001). The same trend is observed in the validation cohort B. The three-year OS of the low- and high-risk groups were 96.4% and 92.8%, respectively ($P_{log\text{-}rank}$ = 0.004); and the three-year DMFS were 90.8% and 86.4%, respectively ($P_{log\text{-}rank}$ = 0.014). Similar to the training cohort, there was no significant difference in the LRFS between the two groups in both validation cohorts (**Figure 2D-I**). We further established the matched cohort using the PSM method to eliminate potential confounders. As shown in Table S2, all the clinical characteristics were balanced between low- and high-risk groups. In the matched cohort, the three-year OS of the low- and high-risk groups were 92.3% and 86.5%, respectively ($P_{log\text{-}rank}$ < 0.001); the three-year DMFS were 87.8% and 81.5%, respectively ($P_{log\text{-}rank}$ < 0.001). No significant difference was observed in LRFS between the two groups ($P_{log\text{-}rank}$ = 0.410) (Figure S2).

After adjustment by other risk factors, multivariate analyses showed that the biochemical signature remained an independent risk factor for OS in the three cohorts (training: hazard ratio [HR] = 2.159, 95% confidence interval [CI]: 1.757 to 2.653; P < 0.001; validation A: HR = 2.147, 95% CI: 1.473 to 3.129; P < 0.001; validation B: HR = 1.624, 95%CI: 1.114 to 2.504; P = 0.012). Additionally, the high-risk patients of the three cohorts also had a significantly higher risk of distant metastasis than the low-risk patients (training: HR = 1.987, 95% CI: 1.604 to 2.463; P < 0.001; validation A: HR = 1.974, 95% CI: 1.458 to 2.673; P < 0.001; validation B: HR = 1.328, 95% CI: 1.013 to 1.765; P = 0.036). Similar to the univariate analysis results, no associations were observed between the biochemical signature and the LRFS (**Table 2**).

Next, we compared the performance of our biochemical signature against the TNM stage (8^{th} AJCC/UICC) and EBV DNA for prognostication in the combined cohort. The C-index of biochemical signature in predicting OS was 0.683, which was significantly higher than that of the current staging system, with values of 0.631 (P < 0.001). Similarly, the C-index for DMFS prediction was 0.691 based on the biochemical signature, which was significantly higher than the C-index by staging system, with a value of 0.626 (P < 0.001). There was no significant difference between biochemical signature and EBV DNA in the C index for OS and DMFS prediction (P > 0.05). However, the predictive value of biochemical signature on LRFS is lower than that of TNM stage and EBV DNA. The C-indexes of biochemical signature, TNM stage and EBV DNA are shown in Table S3.

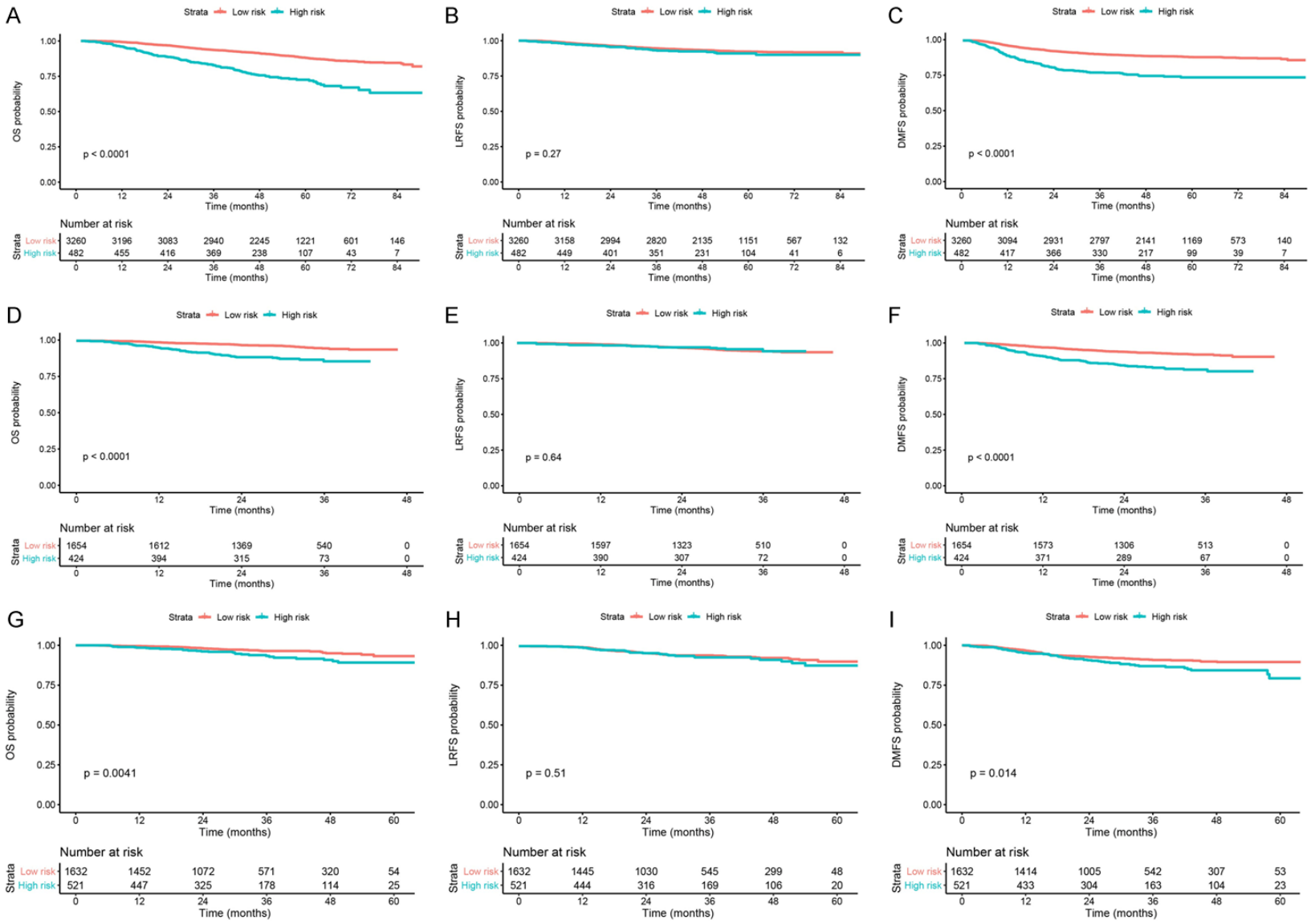

Figure 2. Kaplan-Meier curves of NPC patients in the training cohort: (A) Overall survival, (B) Locoregional relapse-free survival, (C) Distant metastasis-free survival; In the validation cohort A: (D) Overall survival, (E) Locoregional relapse-free survival, (F) Distant metastasis-free survival; In the validation cohort B: (G) Overall survival, (H) Locoregional relapse-free survival, (I) Distant metastasis-free survival.

Table 2. Multivariate analysis of OS, LRFS, and DMFS in three cohorts

Variable	Training cohort		Validation cohort A		Validation cohort B	
	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
Overall survival						
Age	1.603 (1.328-1.935)	< 0.001	1.535 (1.049-2.247)	0.028	2.069 (1.281-3.342)	0.003
Sex	1.728 (1.336-2.236)	< 0.001	1.782 (1.060-2.995)	0.029	0.829 (0.471-1.459)	0.515
Smoking history	1.111 (0.913-1.351)	0.293	0.891 (0.603-1.317)	0.563	1.217 (0.716-2.067)	0.468
NPC family history	0.865 (0.646-1.159)	0.330	1.089 (0.623-1.904)	0.765	1.033 (0.473-2.253)	0.936
T stage	1.570 (1.252-1.967)	< 0.001	1.673 (1.006-2.782)	0.047	1.536 (0.787-2.999)	0.208
N stage	1.417 (1.174-1.710)	< 0.001	2.584 (1.694-3.942)	< 0.001	2.026 (1.194-3.436)	0.009
EBV DNA level	2.358 (1.860-2.989)	< 0.001	2.277 (1.407-3.685)	0.001	1.681 (1.004-2.837)	0.046
Biochemical signature	2.159 (1.757-2.653)	< 0.001	2.147 (1.473-3.129)	< 0.001	1.624 (1.114-2.504)	0.012
Loco-regional relapse-free survival						
Age	1.031 (0.806-1.319)	0.810	0.892 (0.59-1.349)	0.589	0.857 (0.581-1.263)	0.435
Sex	1.252 (0.906-1.730)	0.173	0.710 (0.419-1.203)	0.203	1.365 (0.840-2.216)	0.209
Smoking history	1.105 (0.840-1.453)	0.477	1.632 (1.007-2.647)	0.047	1.012 (0.659-1.556)	0.955
NPC family history	0.957 (0.651-1.408)	0.823	1.355 (0.754-2.437)	0.310	1.112 (0.610-2.029)	0.729
T stage	1.234 (0.927-1.642)	0.150	1.155 (0.709-1.882)	0.564	1.030 (0.638-1.661)	0.905
N stage	1.000 (0.775-1.290)	1.000	0.854 (0.555-1.312)	0.470	1.363 (0.914-2.031)	0.129
EBV DNA level	2.258 (1.669-3.053)	< 0.001	2.119 (1.332-3.372)	0.002	1.139 (0.785-1.628)	0.553
Biochemical signature	1.053 (0.736-1.506)	0.778	0.762 (0.438-1.327)	0.337	1.149 (0.740-1.783)	0.537
Distant metastasis-free survival						
Age	1.169 (0.971-1.406)	0.099	0.956 (0.715-1.280)	0.763	1.111 (0.820-1.505)	0.496
Sex	1.628 (1.268-2.090)	< 0.001	1.726 (1.164-2.560)	0.007	1.217 (0.828-1.788)	0.318
Smoking history	0.971 (0.796-1.184)	0.771	0.840 (0.616-1.144)	0.268	1.166 (0.832-1.632)	0.372
NPC family history	0.766 (0.562-1.044)	0.091	1.287 (0.851-1.947)	0.232	0.978 (0.585-1.637)	0.934
T stage	1.529 (1.218-1.919)	< 0.001	1.622 (1.099-2.395)	0.015	1.717 (1.076-2.738)	0.023
N stage	1.502 (1.237-1.822)	< 0.001	1.847 (1.351-2.525)	< 0.001	2.148 (1.505-3.066)	< 0.001
EBV DNA level	2.554 (1.997-3.266)	< 0.001	3.359 (2.266-4.979)	< 0.001	2.173 (1.534-3.078)	< 0.001
Biochemical signature	1.987 (1.604-2.463)	< 0.001	1.974 (1.458-2.673)	< 0.001	1.328 (1.013-1.765)	0.036

A Cox proportional hazards model was used to conduct multivariate analyses. All variables were transformed into categorical variables. HRs were calculated for age (> 46 vs. ≤ 46); sex (male vs. female); smoking history (yes vs. no); family history of NPC (yes vs. no); T stage (T3-4 vs. T1-2); N stage (N2-3 vs. N0-1); EBV DNA level (> 1500 copies/ml vs. ≤ 1500 copies/ml); Biochemical signature (high risk vs. low risk). Abbreviations: HR, Hazard Ratio; CI, Confidence Interval; NPC, Nasopharyngeal Carcinoma; EBV, Epstein-Barr Virus.

Subgroup analyses showed that, after adjusting for other factors, high-risk patients in the age, sex, T stage, N stage, total stage, and EBV DNA subgroups had significantly lower OS than those with low-risk scores (P < 0.05). In terms of distant metastasis, high-risk patients had significantly lower DMFS than low-risk patients in all subgroups (P < 0.05), except the female subgroup (P > 0.05). However, there were no associations between the biochemical signature and the locoregional recurrence in all subgroups (P > 0.05, **Figure 3**).

Finally, we analyzed the accuracy of the constructed biochemical signature in predicting the efficacy of IC in patients with Stage II-IVA NPC. To reduce the impact of different treatment regimens on the results, all patients included in this analysis had received CCRT ± IC. Of the 5246 patients in the combined cohort, 2360 (45.0%) patients received IC + CCRT, and 2886 (55.0%) received CCRT alone. The clinical characteristics of the two treatment groups are shown in Table S4. In the low-risk subgroup (risk score < 0.8), non-significant differences were observed in the OS (P = 0.700), LRFS (P = 0.120), and DMFS (P = 0.100) between the IC + CCRT and CCRT groups (**Figure 4A-C**). However, in the high-risk subgroup, patients who received IC + CCRT achieved higher OS, and DMFS compared with patients who received CCRT alone (three-year OS rate: 92.8 vs. 86.4%, P = 0.001; three-year DMFS rate: 86.9 vs. 81.2%, P = 0.012). There was no significant difference between the two treatment groups regarding their LRFS (**Figure 4D-F**). In the multivariate analysis, no significant differences in survival endpoints were observed between the two treatment groups of the low-risk subgroup (P > 0.05 for all survival

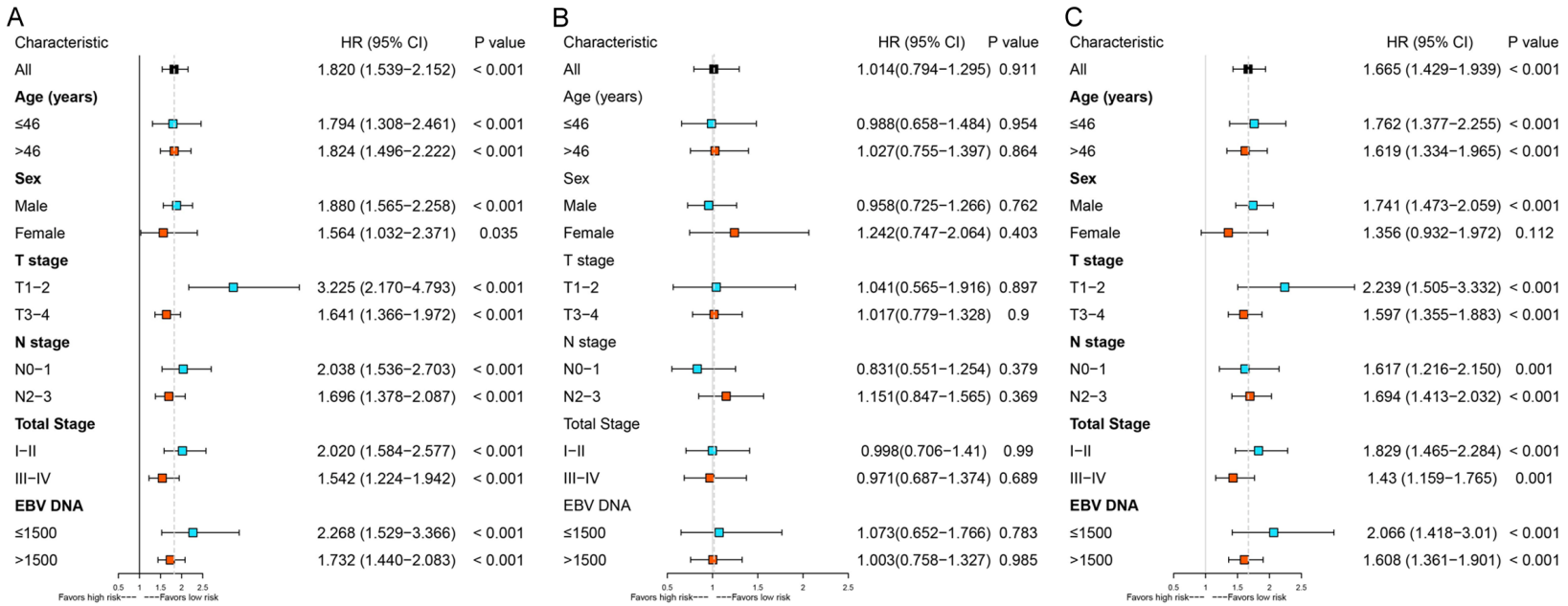

Figure 3. Forest plots of the associations between biochemical signature and survival endpoints. Multivariate hazard ratios were adjusted for the following factors: age, sex, T stage, N stage, and EBV DNA. Low and high refer to the lower and upper limits of the 95% confidence intervals, respectively. (A) Overall survival, (B) Locoregional relapse-free survival, (C) Distant metastasis-free survival.

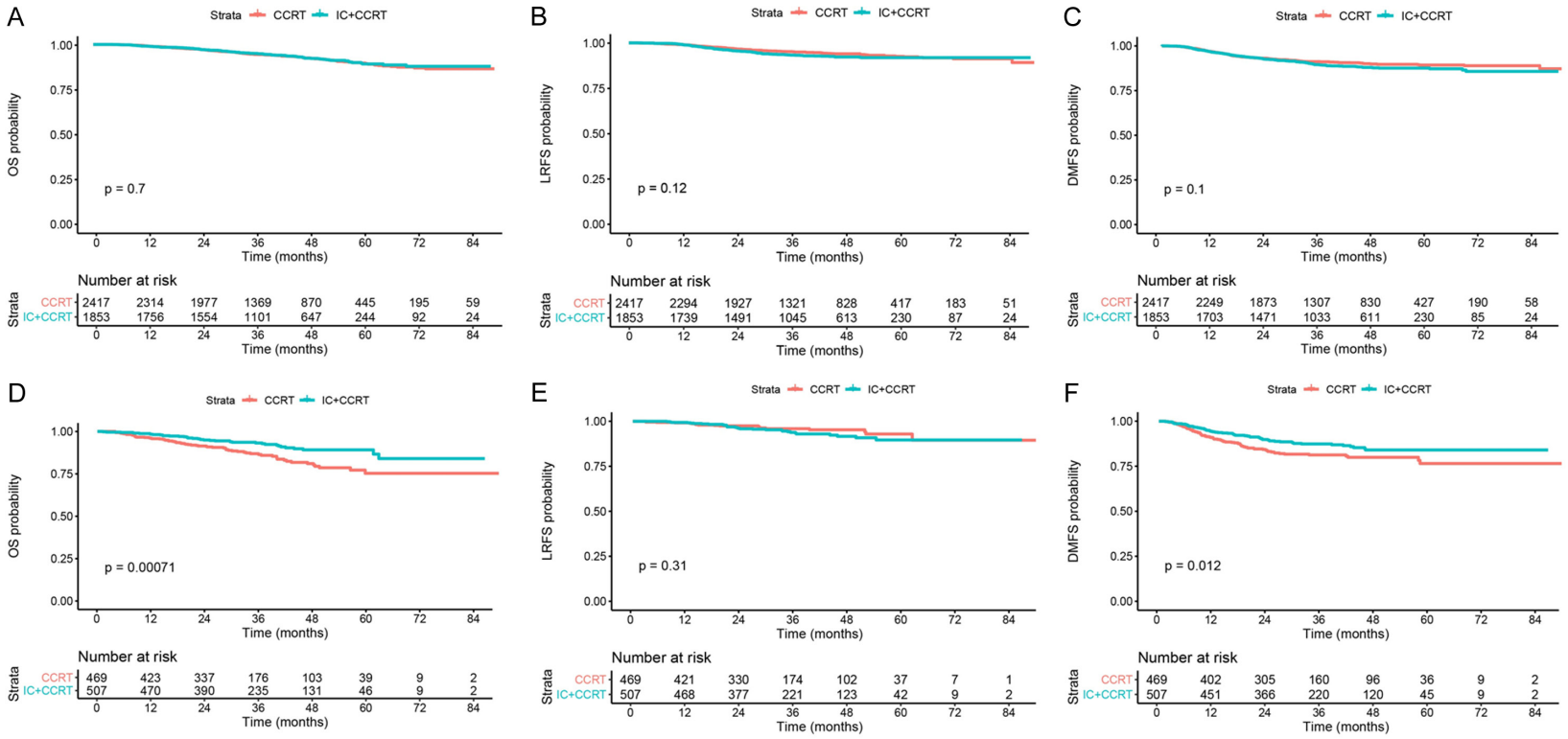

Figure 4. Kaplan-Meier curves of patients with stage II-IVA NPC receiving concurrent chemoradiotherapy (CCRT), and induction chemotherapy combined with concurrent chemoradiotherapy (IC + CCRT) in the low-risk group: (A) Overall survival, (B) Locoregional relapse-free survival, (C) Distant metastasis-free survival; In the high-risk group: (D) Overall survival, (E) Locoregional relapse-free survival, (F) Distant metastasis-free survival.

Table 3. Multivariate analysis of OS, LRFS, and DMFS in low- and high-risk subgroups

Variable	Low-risk subgroup		High-risk subgroup	
	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
Overall survival				
Age	1.508 (1.189-1.912)	0.001	1.892 (1.200-2.982)	0.006
Sex	1.268 (0.908-1.772)	0.164	2.035 (1.113-3.721)	0.021
Smoking history	1.504 (1.156-1.958)	0.002	0.790 (0.521-1.199)	0.268
NPC family history	0.856 (0.567-1.291)	0.457	0.952 (0.493-1.837)	0.883
T stage	1.611 (1.174-2.210)	0.003	0.886 (0.534-1.469)	0.638
N stage	1.877 (1.448-2.433)	< 0.001	2.108 (1.313-3.386)	0.002
EBV DNA level	2.592 (1.915-3.510)	< 0.001	2.473 (1.401-4.367)	0.002
Treatment method	0.827 (0.671-1.027)	0.090	0.411 (0.274-0.617)	< 0.001
Loco-regional relapse-free survival				
Age	0.903 (0.699-1.167)	0.435	1.195 (0.657-2.175)	0.559
Sex	0.917 (0.658-1.278)	0.610	1.756 (0.777-3.969)	0.176
Smoking history	1.586 (1.190-2.116)	0.002	0.759 (0.411-1.400)	0.377
NPC family history	1.282 (0.884-1.859)	0.191	0.66 (0.205-2.129)	0.487
T stage	0.969 (0.716-1.310)	0.836	1.234(0.520-2.927)	0.633
N stage	1.016 (0.781-1.323)	0.904	1.509 (0.782-2.910)	0.220
EBV DNA level	1.539 (1.166-2.031)	0.002	1.630 (0.764-3.477)	0.206
Treatment method	1.125 (0.869-1.458)	0.371	1.141 (0.630-2.067)	0.662
Distant metastasis-free survival				
Age	0.992 (0.813-1.209)	0.933	0.963 (0.686-1.353)	0.829
Sex	1.414 (1.083-1.847)	0.011	2.006 (1.197-3.361)	0.008
Smoking history	1.144 (0.922-1.420)	0.221	0.900 (0.631-1.283)	0.559
NPC family history	0.834 (0.592-1.175)	0.299	1.406 (0.864-2.287)	0.170
T stage	1.603 (1.229-2.091)	0.001	1.235 (0.760-2.009)	0.394
N stage	1.731 (1.395-2.146)	< 0.001	2.845 (1.832-4.418)	< 0.001
EBV DNA level	2.618 (2.048-3.347)	< 0.001	2.410 (1.492-3.892)	< 0.001
Treatment method	0.865 (0.709-1.056)	0.155	0.486 (0.347-0.680)	< 0.001

A Cox proportional hazards model was used to conduct multivariate analyses. All variables were transformed into categorical variables. HRs were calculated for age (> 46 vs. ≤ 46); sex (male vs. female); smoking history (yes vs. no); family history of NPC (yes vs. no); T stage (T3-4 vs. T1-2); N stage (N2-3 vs. N0-1); EBV DNA level (> 1500 copies/ml vs. ≤ 1500 copies/ml); treatment method (IC + CCRT vs. CCRT). Abbreviations: HR, Hazard Ratio; CI, Confidence Interval; NPC, Nasopharyngeal Carcinoma; EBV, Epstein-Barr Virus; IC, Induction Chemotherapy; CCRT, Concurrent Chemoradiotherapy.

endpoints). However, in the high-risk subgroup, the addition of IC was observed to be a protective factor for OS (HR = 0.411; 95% CI: 0.274-0.617; P < 0.001), and DMFS (HR = 0.486; 95% CI: 0.347-0.680; P < 0.001). The treatment method was not associated with differences in locoregional recurrence in the two risk groups (**Table 3**).

Discussion

In this paper, we reported the results of a retrospective study that aimed to utilize biochemical indicators in categorizing locoregionally advanced NPC patients into high-risk and low-risk groups. In the three cohorts, low-risk patients yielded better OS and DMFS. In addition, patients identified as high risk by the biochemical signature were predicted to have greater therapeutic benefits from IC + CCRT.

With the development of radiotherapy technology, IMRT is widely applied in treating NPC, effectively improving the survival rate and local control rate of patients. In recent years, IC has brought some survival benefits for NPC patients [6, 7]. However, distant metastasis remains a hindrance in treatment efficacy. Moreover, due to tumor heterogeneity, the prognosis and response to IC vary even in indi-

viduals with the same clinical TNM stage. Therefore, there is an urgent need for an effective prognostic model that will guide clinicians in choosing the suitable treatment option for NPC patients.

In our study, 33 biochemical indicators were screened and assessed, six of which were selected to construct the prediction model through the LASSO cox regression model. LASSO regression is characterized by variable selection and complexity adjustment while fitting the generalized linear model. Therefore, it is feasible to screen the set of variables that have the strongest explanatory power on the outcome. Using the biochemical signature, patients were divided into high-risk and low-risk groups. Prior to treatment, Na, ALB, and IBIL were considered protective factors, while high concentrations of ALP, LDH, and CYSC indicated rapid disease progression and poor prognosis.

Our results showed consistency with other studies. Doshi *et al.* reported that hyponatremia in cancer patients is associated with longer hospital stays and higher mortality [14]. Additionally, in a study by Zhang, a combination of Na and globulin levels in the serum was used to reflect electrolyte homeostasis and inflammatory state in cancer patients. It was demonstrated that the Na to globulin ratio (SGR) obtained a high accuracy in predicting cancer patient survival, where the low-SGR group exhibited significantly worse OS compared with the high-SGR group [22]. On the other hand, a study that focused on the albumin-to-alkaline phosphatase ratio (AAPR) indicated that an AAPR of over 0.4876 is associated with better OS and LRFS in locoregionally advanced NPC patients [13]. Another protective indicator in our study, IBIL (> 7.15 μmol/L), was also used as an independent protective prognostic factor for determining the PFS, OS, and DMFS of patients with advanced NPC [23]. Furthermore, an elevation in LDH levels serves as an effective negative biomarker for cancer prognosis. LDH is a key enzyme in cancer metabolism and contributes to immune evasion by allowing cancer cells to suppress and evade the immune system by altering the tumor microenvironment [24]. Lastly, studies have also shown that a high concentration of pretreatment CYSC is implicated in rapid disease progression [25, 26].

Our study is the first to involve all 33 indicators and six representative indicators based on a large sample to establish a biochemical signature. We reported that the constructed biochemical signature was an independent risk factor for estimating the OS in all three cohorts. The high-risk patients of all three cohorts also had a significantly higher risk of distant metastasis than the low-risk patients. Subgroup analyses showed that the biochemical signature applied to all ages, sex, T stage, N stage, total stage, and EBV DNA subgroups. We further confirmed that the biochemical signature might also be used to predict the efficacy of the IC in patients with Stage II-IVA NPC. In patients with high-risk biochemical signatures, it was observed that the addition of IC was a protective factor for both OS and DMFS.

Previous studies have adopted the use of plasma EBV DNA detection to monitor NPC patients before and during treatment and to predict the outcome of treatment [21, 27]. Patients with a higher EBV DNA load prior to treatment often have a worse prognosis than those with a lower EBV DNA load, and the presence of detectable EBV DNA in the plasma by the end of treatment often suggests a poor prognosis. Nevertheless, there are several obstacles to the widespread use of this biomarker: (1) some medical centers lack the facilities or technicians to assess the plasma EBV DNA load; (2) there is heterogeneity in the results of each center; and (3) no objective and unified detection method is generally accepted among medical centers. In contrast, the relatively low-cost biochemical indicators have been developed and optimized, and test results are stably and widely recognized. Therefore, as a predictive index, biochemical signatures offer great advantages in terms of clinical applications.

To our knowledge, this is the first study to combine all 33 biochemical indicators to predict the prognosis of locoregionally advanced NPC patients. We analyzed 7973 cases of non-metastatic NPC with one training cohort and two internal validation cohorts. The large sample size allowed for subgroup analyses with sufficient testing. However, there are several limitations to this study. First, the retrospective cohorts might be biased in case selection, and there may be information loss or inaccurate records in medical records. Second, all cases

were enrolled in an NPC-endemic area in China, and the pathological subtype was mainly undifferentiated non-keratinizing carcinoma. The application of this study should be revalidated in areas with distinct pathological subtypes and clinical features.

Conclusion

Our findings reported a prognostic index, a biochemical signature that could effectively predict the prognosis of locoregionally advanced NPC patients, particularly the risk for distant metastasis. Furthermore, we found that patients identified as high-risk benefit more from IC + CCRT. This prognostic index may therefore offer clinicians more practical and accessible information for the prediction of NPC prognosis.

Disclosure of conflict of interest

None.

Address correspondence to: Drs. Hai-Qiang Mai and Qiu-Yan Chen, Department of Nasopharyngeal Carcinoma, Sun Yat-sen University Cancer Center, 651 Dongfeng Road East, Guangzhou, Guangdong 510060, P. R. China. Tel: +86-20-87343380; Fax: +86-20-87343392; E-mail: maihq@mail.sysu.edu.cn (HQM); Tel: +86-20-87343155; Fax: +86-20-87343392; E-mail: chenqy@sysucc.org.cn (QYC)

References

[1] Wee JT, Ha TC, Loong SL and Qian CN. Is nasopharyngeal cancer really a "Cantonese cancer"? Chin J Cancer 2010; 29: 517-526.

[2] Peng G, Wang T, Yang KY, Zhang S, Zhang T, Li Q, Han J and Wu G. A prospective, randomized study comparing outcomes and toxicities of intensity-modulated radiotherapy vs. conventional two-dimensional radiotherapy for the treatment of nasopharyngeal carcinoma. Radiother Oncol 2012; 104: 286-293.

[3] Zhang MX, Li J, Shen GP, Zou X, Xu JJ, Jiang R, You R, Hua YJ, Sun Y, Ma J, Hong MH and Chen MY. Intensity-modulated radiotherapy prolongs the survival of patients with nasopharyngeal carcinoma compared with conventional two-dimensional radiotherapy: a 10-year experience with a large cohort and long follow-up. Eur J Cancer 2015; 51: 2587-2595.

[4] Lai SZ, Li WF, Chen L, Luo W, Chen YY, Liu LZ, Sun Y, Lin AH, Liu MZ and Ma J. How does intensity-modulated radiotherapy versus conventional two-dimensional radiotherapy influence the treatment results in nasopharyngeal carcinoma patients? Int J Radiat Oncol Biol Phys 2011; 80: 661-668.

[5] Yang Q, Cao SM, Guo L, Hua YJ, Huang PY, Zhang XL, Lin M, You R, Zou X, Liu YP, Xie YL, Wang ZQ, Mai HQ, Chen QY, Tang LQ, Mo HY, Cao KJ, Qian CN, Zhao C, Xiang YQ, Zhang XP, Lin ZX, Li WX, Liu Q, Li JB, Ling L, Guo X, Hong MH and Chen MY. Induction chemotherapy followed by concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: long-term results of a phase III multicentre randomised controlled trial. Eur J Cancer 2019; 119: 87-96.

[6] Zhang Y, Chen L, Hu GQ, Zhang N, Zhu XD, Yang KY, Jin F, Shi M, Chen YP, Hu WH, Cheng ZB, Wang SY, Tian Y, Wang XC, Sun Y, Li JG, Li WF, Li YH, Tang LL, Mao YP, Zhou GQ, Sun R, Liu X, Guo R, Long GX, Liang SQ, Li L, Huang J, Long JH, Zang J, Liu QD, Zou L, Su QF, Zheng BM, Xiao Y, Guo Y, Han F, Mo HY, Lv JW, Du XJ, Xu C, Liu N, Li YQ, Chua MLK, Xie FY, Sun Y and Ma J. Gemcitabine and cisplatin induction chemotherapy in nasopharyngeal carcinoma. N Engl J Med 2019; 381: 1124-1135.

[7] Sun Y, Li WF, Chen NY, Zhang N, Hu GQ, Xie FY, Sun Y, Chen XZ, Li JG, Zhu XD, Hu CS, Xu XY, Chen YY, Hu WH, Guo L, Mo HY, Chen L, Mao YP, Sun R, Ai P, Liang SB, Long GX, Zheng BM, Feng XL, Gong XC, Li L, Shen CY, Xu JY, Guo Y, Chen YM, Zhang F, Lin L, Tang LL, Liu MZ and Ma J. Induction chemotherapy plus concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase 3, multicentre, randomised controlled trial. Lancet Oncol 2016; 17: 1509-1520.

[8] Liu SL, Sun XS, Yan JJ, Chen QY, Lin HX, Wen YF, Guo SS, Liu LT, Xie HJ, Tang QN, Liang YJ, Li XY, Lin C, Du YY, Yang ZC, Xiao BB, Yang JH, Tang LQ, Guo L and Mai HQ. Optimal cumulative cisplatin dose in nasopharyngeal carcinoma patients based on induction chemotherapy response. Radiother Oncol 2019; 137: 83-94.

[9] Liu LT, Tang LQ, Chen QY, Zhang L, Guo SS, Guo L, Mo HY, Zhao C, Guo X, Cao KJ, Qian CN, Zeng MS, Bei JX, Hong MH, Shao JY, Sun Y, Ma J and Mai HQ. The prognostic value of plasma Epstein-Barr viral DNA and tumor response to neoadjuvant chemotherapy in advanced-stage nasopharyngeal carcinoma. Int J Radiat Oncol Biol Phys 2015; 93: 862-869.

[10] Vander Heiden MG and DeBerardinis RJ. Understanding the Intersections between Metabolism and Cancer Biology. Cell 2017; 168: 657-669.

[11] Jin Y, Ye X, Shao L, Lin BC, He CX, Zhang BB and Zhang YP. Serum lactic dehydrogenase strongly predicts survival in metastatic nasopharyngeal carcinoma treated with palliative

chemotherapy. Eur J Cancer 2013; 49: 1619-1626.
[12] Xia WX, Zhang HB, Shi JL, Lu X, Wang L, Ye YF, Cao KJ, Qian CN, Guo X and Xiang YQ. A prognostic model predicts the risk of distant metastasis and death for patients with nasopharyngeal carcinoma based on pre-treatment serum C-reactive protein and N-classification. Eur J Cancer 2013; 49: 2152-2160.
[13] Kim JS, Keam B, Heo DS, Han DH, Rhee CS, Kim JH, Jung KC and Wu HG. The prognostic value of albumin-to-alkaline phosphatase ratio before radical radiotherapy in patients with non-metastatic nasopharyngeal carcinoma: a propensity score matching analysis. Cancer Res Treat 2019; 51: 1313-1323.
[14] Doshi SM, Shah P, Lei X, Lahoti A and Salahudeen AK. Hyponatremia in hospitalized cancer patients and its impact on clinical outcomes. Am J Kidney Dis 2012; 59: 222-228.
[15] Zhou Q, Zhou C, Yin Y, Chen W, Liu C, Atyah M, Weng J, Shen Y, Yi Y and Ren N. Development and validation of a nomogram combining hematological and imaging features for preoperative prediction of microvascular invasion in hepatocellular carcinoma patients. Ann Transl Med 2021; 9: 402.
[16] Tang XR, Li YQ, Liang SB, Jiang W, Liu F, Ge WX, Tang LL, Mao YP, He QM, Yang XJ, Zhang Y, Wen X, Zhang J, Wang YQ, Zhang PP, Sun Y, Yun JP, Zeng J, Li L, Liu LZ, Liu N and Ma J. Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study. Lancet Oncol 2018; 19: 382-393.
[17] Liu SL, Sun XS, Chen QY, Liu ZX, Bian LJ, Yuan L, Xiao BB, Lu ZJ, Li XY, Yan JJ, Yan SM, Li JM, Bei JX, Mai HQ and Tang LQ. Development and validation of a transcriptomics-based gene signature to predict distant metastasis and guide induction chemotherapy in locoregionally advanced nasopharyngeal carcinoma. Eur J Cancer 2022; 163: 26-34.
[18] Tang LQ, Li CF, Li J, Chen WH, Chen QY, Yuan LX, Lai XP, He Y, Xu YX, Hu DP, Wen SH, Peng YT, Zhang L, Guo SS, Liu LT, Guo L, Wu YS, Luo DH, Huang PY, Mo HY, Xiang YQ, Sun R, Chen MY, Hua YJ, Lv X, Wang L, Zhao C, Cao KJ, Qian CN, Guo X, Zeng YX, Mai HQ and Zeng MS. Establishment and validation of prognostic nomograms for endemic nasopharyngeal carcinoma. J Natl Cancer Inst 2016; 108: djv291.
[19] Tang LQ, Li CF, Chen QY, Zhang L, Lai XP, He Y, Xu YX, Hu DP, Wen SH, Peng YT, Chen WH, Liu H, Guo SS, Liu LT, Li J, Zhang JP, Guo L, Zhao C, Cao KJ, Qian CN, Zeng YX, Guo X, Mai HQ and Zeng MS. High-sensitivity C-reactive protein complements plasma Epstein-Barr virus deoxyribonucleic acid prognostication in nasopharyngeal carcinoma: a large-scale retrospective and prospective cohort study. Int J Radiat Oncol Biol Phys 2015; 91: 325-336.
[20] Harrell F, Califf R, Pryor D, Lee K and Rosati R. Evaluating the yield of medical tests. JAMA 1982; 247: 2543-2546.
[21] Lin JC, Wang WY, Chen KY, Wei YH, Liang WM, Jan JS and Jiang RS. Quantification of plasma Epstein-Barr virus DNA in patients with advanced nasopharyngeal carcinoma. N Engl J Med 2004; 350: 2461-2470.
[22] Zhang L, Wang Z, Xiao J, Chen H, Zhang Z, Li H, Wang Y, Piao H, Li F, Zhang L and Zhang J. Sodium to globulin ratio as a prognostic factor for patients with advanced gastric cancer. J Cancer 2020; 11: 7320-7328.
[23] Yao JJ, Kou J, Peng QH, Dong J, Zhang WJ, Lawrence WR, Zhang F, Zhou GQ, Wang SY and Sun Y. Prognostic value of serum bilirubin in southern Chinese patients with advanced nasopharyngeal carcinoma. Clin Chim Acta 2018; 484: 314-319.
[24] Ding J, Karp JE and Emadi A. Elevated lactate dehydrogenase (LDH) can be a marker of immune suppression in cancer: interplay between hematologic and solid neoplastic clones and their microenvironments. Cancer Biomark 2017; 19: 353-363.
[25] Zhao W, He Z, Li Y, Jia H, Chen M, Gu X, Liu M, Zhang Z, Wu Z and Cheng W. Nomogram-based parameters to predict overall survival in a real-world advanced cancer population undergoing palliative care. BMC Palliat Care 2019; 18: 47.
[26] Wang H, Shan D, Dong Y, Yang X, Zhang L and Yu Z. Correlation analysis of serum cystatin C, uric acid and lactate dehydrogenase levels before chemotherapy on the prognosis of small-cell lung cancer. Oncol Lett 2021; 21: 73.
[27] Liu LT, Chen QY, Tang LQ, Guo SS, Guo L, Mo HY, Li Y, Tang QN, Sun XS, Liang YJ, Zhao C, Guo X, Qian CN, Zeng MS, Bei JX, Hong MH, Shao JY, Sun Y, Ma J and Mai HQ. Neoadjuvant or adjuvant chemotherapy plus concurrent CRT Versus concurrent CRT alone in the treatment of nasopharyngeal carcinoma: a study based on EBV DNA. J Natl Compr Canc Netw 2019; 17: 703-710.

Supplementary materials

Treatment method

All patients were treated with radical intensity-modulated radiotherapy (IMRT) as a primary treatment, while immobilized in the supine position using a thermoplastic head and shoulder mask. Contrast-enhanced planning computed tomography (CT; 3 mm-slice thickness) images from the superior border of the frontal sinus to 2 cm below the sterno-clavicular joint were obtained and transferred to the Monaco treatment planning system (version 3.02; Elekta AB, Stockholm, Sweden).

Target volumes and organs at risk (OARs) were delineated on each slice of the CT images, as previously described [1], in agreement with International Commission on Radiation Units and Measurements Reports 62 [2] and 83 [3]. The gross tumor volume (GTV) including primary nasopharyngeal tumor (GTVp) and GTVnd was delineated on the basis of clinical, endoscopic and MRI findings. Gross disease at primary site together with enlarged retropharyngeal lymph nodes was designated GTVp; clinically-involved cervical lymph nodes was designated GTVnd. Two clinical target volumes (CTVs) were delineated according to the GTV: CTV1, high-risk regions encompassing GTVp plus 5-10 mm, including entire nasopharyngeal mucosa and 5 mm submucosal region; and CTV2, low-risk regions containing CTV1 plus 5-10 mm, encompassing sites of microscopic extension and lymphatic regions. The planning target volumes (PTVs), termed PTVp, PTV1, PTV2 and PTVnd, were constructed by expanding the GTVp, CTV1, CTV2 and CTVnd, respectively, by 3 mm; a 3 mm margin was added to the brainstem and spinal cord to generate planning organ at risk volume (PRV). The prescribed doses were 66-72 Gy/28-33 fractions to the planning target volume (PTV) of the primary gross tumour volume (GTVnx), 64-70 Gy/28-33 fractions to the PTV of the GTV of the involved lymph nodes (GTVnd), 60-63 Gy/28-33 fractions to the PTV of the high-risk clinical target volume (CTV1), and 54-56 Gy/28-33 fractions to the PTV of the low-risk clinical target volume (CTV2).

Institutional guidelines recommended IMRT for stage I NPC, platinum-based concurrent chemoradiotherapy ± induction chemotherapy/adjuvant chemotherapy for stage II to IVB NPC. For concurrent chemotherapy, the accumulated dose for cisplatin/nedaplatin was 80-100 mg/m^2, administered every three weeks or 30-40 mg/m^2 administered every week. The induction chemotherapy (IC) and adjuvant chemotherapy (AC) regimens that were administered included TPF (docetaxel [60 mg/m^2, day 1], paclitaxel [110-135 mg/m^2, day 1] or paclitaxel liposome [110-135 mg/m^2, day 1], cisplatin [20-25 mg/m^2/day, days 1-3], and 5-fluorouracil [3000-3750 mg/m^2, 120 h of continuous intravenous infusion]), PF (cisplatin [20-25 mg/m^2/day, days 1-3], and 5-fluorouracil [3200-4000 mg/m^2, 96 h of continuous intravenous infusion]), or TP (docetaxel [60 mg/m^2, day 1], paclitaxel [110-135 mg/m^2, day 1] or paclitaxel liposome [110-135 mg/m^2, day 1], cisplatin [20-25 mg/m^2/day, days 1-3]). IC and AC were administered every 3 weeks.

References

[1] Lai SZ, Li WF, Chen L, Luo W, Chen YY, Liu LZ, Sun Y, Lin AH, Liu MZ and Ma J. How does intensity-modulated radiotherapy versus conventional two-dimensional radiotherapy influence the treatment results in nasopharyngeal carcinoma patients? Int J Radiat Oncol Biol Phys 2011; 80: 661-8.

[2] ICRU report. Vol. 62: prescribing, recording, and reporting photon beam therapy. Maryland: International Commission on Radiation Units and Measurements; 1999.

[3] ICRU Report. Vol. 83: prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT). Maryland: International Commission on Radiation Units and Measurements; 2010.

Table S1. The detailed information of the thirty-three biochemical indicators

Abbreviation	Full name
K	Kalium
Na	Sodium
Cl	Chlorine
CO_2	Carbon dioxide
Gap	Serum anion gap
Ca	Calcium
Mg	Magnesium
ALT	Glutamic-pyruvic transaminase
AST	Glutamic-oxalacetic transaminase
AS/AL	Glutamic-oxalacetic transaminase/Glutamic-pyruvic transaminase
TBA	Total bile acid
ALP	Alkaline phosphatase
GGT	Gamma-glutamyl transpeptidase
LDH	Lactate dehydrogenase
AFU	Alpha-L-fucosidase
TP	Total protein
ALB	Albumin
GLOB	Globulin
A/G	Albumin/Globulin
TBIL	Total bilirubin
IBIL	Indirect bilirubin
BUN	Blood urea nitrogen
CRE	Creatinine
UA	Uric acid
CYSC	Cystatin-C
CHO	Total cholesterol
TG	Triglyceride
GLU	Blood glucose
HDL	High-density lipoprotein
LDL	Low-density lipoprotein
ApoAI	Apolipoprotein A1
ApoB	Apolipoprotein B
CRP	C-reaction protein

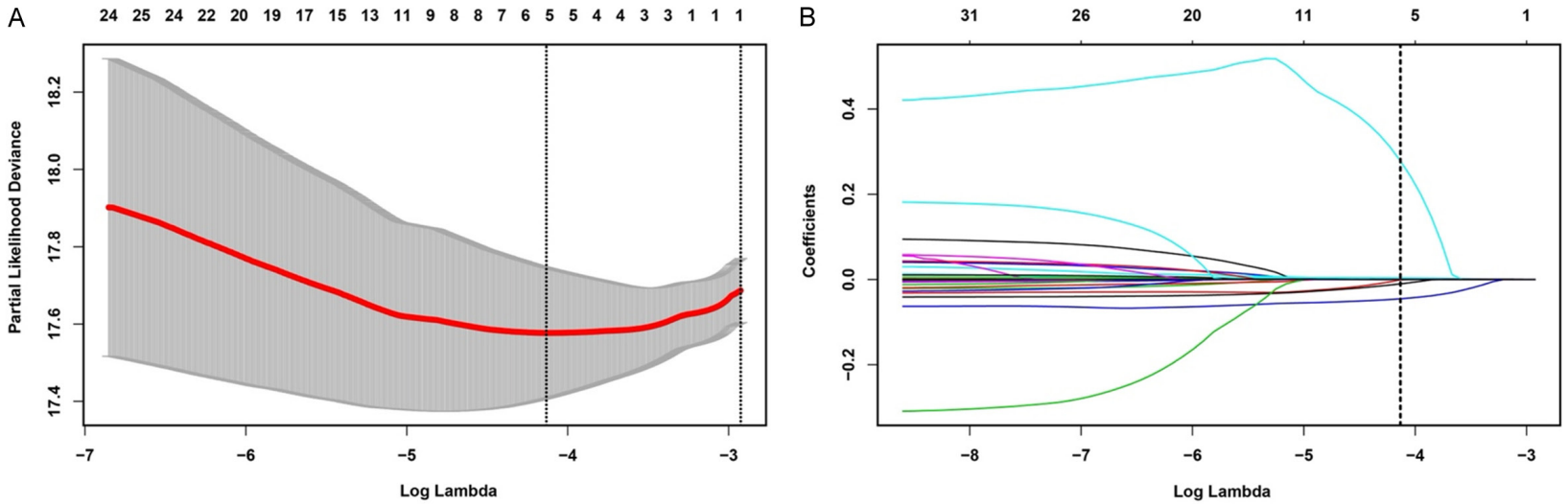

Figure S1. Selection of biochemical indicators using LASSO Cox regression. A. Tuning parameter lambda (λ) selection in the LASSO method; B. LASSO coefficient profiles of the candidate biochemical indicators.

Table S2. Clinical characteristics of patients in matched cohort

Characteristic	Low-risk patients n (%)	High-risk patients n (%)	*P*-value
Total	1415	1415	
Age, y			0.904
≤ 46	451 (31.9)	447 (31.6)	
> 46	964 (68.1)	968 (68.4)	
Sex			1.000
Female	323 (22.8)	322 (22.8)	
Male	1092 (77.2)	1093 (77.2)	
Smoking history			0.252
No	816 (62.6)	797 (56.3)	
Yes	579 (37.4)	618 (43.7)	
NPC family history			0.711
No	1265 (89.4)	1272 (89.9)	
Yes	150 (10.6)	143 (10.1)	
T stage*			0.585
T1	54 (3.82)	62 (4.38)	
T2	175 (12.4)	166 (11.7)	
T3	726 (51.3)	701 (49.5)	
T4	460 (32.5)	486 (34.3)	
N stage*			0.657
N0	170 (12.0)	162 (11.4)	
N1	443 (31.3)	448 (31.7)	
N2	589 (41.6)	570 (40.3)	
N3	213 (15.1)	235 (16.6)	
Overall stage*			0.970
I	18 (1.27)	20 (1.41)	
II	87 (6.15)	84 (5.94)	
III	683 (48.3)	676 (47.8)	
IV	627 (44.3)	635 (44.9)	
EBV DNA level			0.873
≤ 1500 copies/ml	470 (33.2)	465 (32.9)	
> 1500 copies/ml	945 (66.8)	950 (67.1)	
Treatment method			0.980
RT alone	178 (12.6)	186 (13.1)	
CCRT	471 (33.3)	463 (32.7)	
IC + RT	236 (16.7)	235 (16.6)	
IC + CCRT	507 (35.8)	505 (35.7)	
Others	23 (1.63)	26 (1.84)	

P value was calculated with the Pearson χ^2 test. *According to the 8th edition of the UICC/AJCC staging system. Abbreviations: NPC, Nasopharyngeal Carcinoma; EBV DNA, Epstein-Barr Virus DNA; RT, Radiotherapy; CCRT, Concurrent Chemoradiotherapy; IC, Induction Chemotherapy.

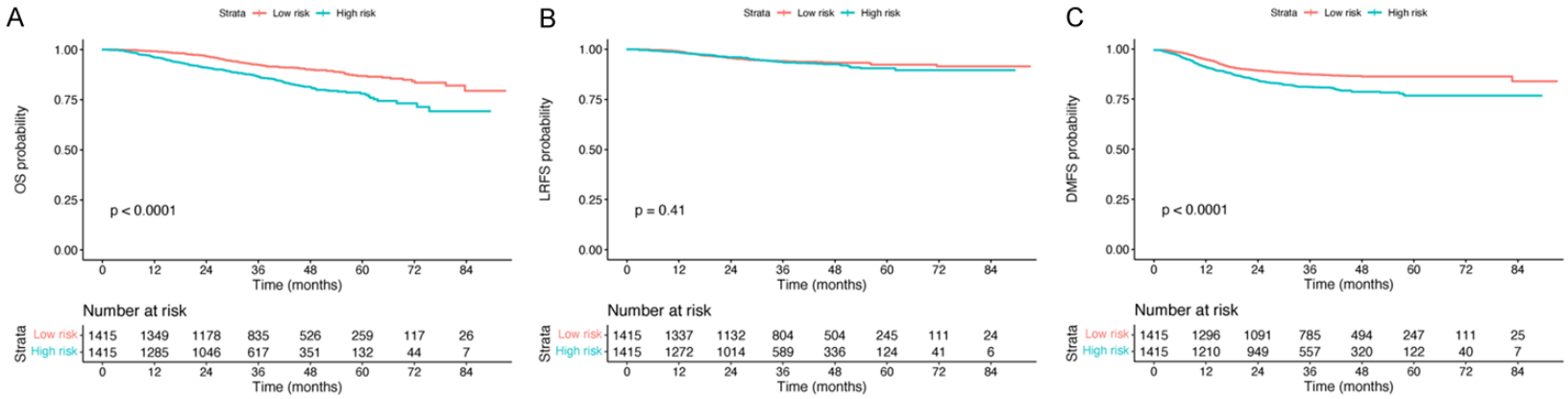

Figure S2. Kaplan-Meier curves of NPC patients in the matched cohort. A. Overall survival; B. Locoregional relapse-free survival; C. Distant metastasis-free survival.

Table S3. Concordance indexes of biochemical signature, TNM stage and EBV DNA

Characteristic	C-index (95% CI)	*P-value*
Overall survival		
Biochemical signature	0.683 (0.660-0.705)	Reference
TNM stage*	0.631 (0.611-0.650)	< 0.001
EBV DNA	0.689 (0.667-0.710)	0.281
Locoregional relapse-free survival		
Biochemical signature	0.502 (0.473-0.530)	Reference
TNM stage*	0.573 (0.548-0.597)	< 0.001
EBV DNA	0.614 (0.587-0.640)	< 0.001
Distant metastasis-free survival		
Biochemical signature	0.691 (0.671-0.710)	Reference
TNM stage*	0.626 (0.609-0.642)	< 0.001
EBV DNA	0.702 (0.684-0.719)	0.160

*According to the 8th edition of the UICC/AJCC staging system.

Table S4. Clinical characteristics of patients in low- and high-risk subgroups

Characteristic	Low-risk patients n (%)			High-risk patients n (%)		
	CCRT	CCRT + IC	*P*-value	CCRT	CCRT + IC	*P*-value
Total	2417	1853		469	507	
Age, y			0.024			0.033
≤ 46	1367 (56.6)	1112 (60.0)		153 (32.6)	199 (39.3)	
> 46	1050 (43.4)	741 (40.0)		316 (67.4)	308 (60.7)	
Sex			0.463			0.396
Female	652 (27.0)	481 (26.0)		109 (23.2)	106 (20.9)	
Male	1765 (73.0)	1372 (74.0)		360 (76.8)	401 (79.1)	
Smoking history			0.795			0.796
No	1583 (65.5)	1221 (65.9)		26 (56.7)	292 (57.6)	
Yes	834 (34.5)	632 (34.1)		203 (43.3)	215 (42.4)	
NPC family history			0.155			0.911
No	2153 (89.1)	1676 (90.4)		428 (91.3)	461 (90.9)	
Yes	264 (10.9)	177 (9.6)		41 (8.7)	46 (9.1)	
T stage*			< 0.001			< 0.001
T1	145 (6.0)	59 (3.2)		18 (3.8)	10 (2.0)	
T2	462 (19.1)	262 (14.1)		53 (11.3)	45 (8.9)	
T3	1435 (59.4)	920 (49.6)		275 (58.6)	229 (45.2)	
T4	375 (15.5)	612 (33.0)		123 (26.2)	223 (44.0)	
N stage*			< 0.001			< 0.001
N0	342 (14.1)	150 (8.1)		42 (9.0)	23 (4.5)	
N1	1031 (42.7)	533 (28.8)		185 (39.4)	129 (25.4)	
N2	880 (36.4)	819 (44.2)		191 (40.7)	227 (44.8)	
N3	164 (6.8)	351 (18.9)		51 (10.9)	128 (25.2)	
Overall stage*			< 0.001			< 0.001
II	300 (12.4)	69 (3.7)		29 (6.2)	9 (1.8)	
III	1603 (66.3)	904 (48.8)		279 (59.5)	193 (38.1)	
IV	514 (21.3)	880 (47.5)		161 (34.3)	305 (60.2)	
EBV DNA level			< 0.001			< 0.001
≤ 1500 copies/ml	1287 (53.2)	619 (33.4)		194 (41.4)	118 (23.3)	
> 1500 copies/ml	1130 (46.8)	1234 (66.6)		275 (58.6)	389 (76.7)	

P value was calculated with the Pearson χ^2 test. *According to the 8th edition of the UICC/AJCC staging system. Abbreviations: NPC, Nasopharyngeal Carcinoma; EBV DNA, Epstein-Barr Virus DNA; RT, Radiotherapy; CCRT, Concurrent Chemoradiotherapy; IC, Induction Chemotherapy.