# Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study

*Xin-Ran Tang*, Ying-Qin Li*, Shao-Bo Liang*, Wei Jiang*, Fang Liu*, Wen-Xiu Ge, Ling-Long Tang, Yan-Ping Mao, Qing-Mei He, Xiao-Jing Yang, Yuan Zhang, Xin Wen, Jian Zhang, Ya-Qin Wang, Pan-Pan Zhang, Ying Sun, Jing-Ping Yun, Jing Zeng, Li Li, Li-Zhi Liu, Na Liu†, Jun Ma†*

## Summary

**Background** Gene expression patterns can be used as prognostic biomarkers in various types of cancers. We aimed to identify a gene expression pattern for individual distant metastatic risk assessment in patients with locoregionally advanced nasopharyngeal carcinoma.

**Methods** In this multicentre, retrospective, cohort analysis, we included 937 patients with locoregionally advanced nasopharyngeal carcinoma from three Chinese hospitals: the Sun Yat-sen University Cancer Center (Guangzhou, China), the Affiliated Hospital of Guilin Medical University (Guilin, China), and the First People's Hospital of Foshan (Foshan, China). Using microarray analysis, we profiled mRNA gene expression between 24 paired locoregionally advanced nasopharyngeal carcinoma tumours from patients at Sun Yat-sen University Cancer Center with or without distant metastasis after radical treatment. Differentially expressed genes were examined using digital expression profiling in a training cohort (Guangzhou training cohort; n=410) to build a gene classifier using a penalised regression model. We validated the prognostic accuracy of this gene classifier in an internal validation cohort (Guangzhou internal validation cohort, n=204) and two external independent cohorts (Guilin cohort, n=165; Foshan cohort, n=158). The primary endpoint was distant metastasis-free survival. Secondary endpoints were disease-free survival and overall survival.

**Findings** We identified 137 differentially expressed genes between metastatic and non-metastatic locoregionally advanced nasopharyngeal carcinoma tissues. A distant metastasis gene signature for locoregionally advanced nasopharyngeal carcinoma (DMGN) that consisted of 13 genes was generated to classify patients into high-risk and low-risk groups in the training cohort. Patients with high-risk scores in the training cohort had shorter distant metastasis-free survival (hazard ratio [HR] 4·93, 95% CI 2·99–8·16; p<0·0001), disease-free survival (HR 3·51, 2·43–5·07; p<0·0001), and overall survival (HR 3·22, 2·18–4·76; p<0·0001) than patients with low-risk scores. The prognostic accuracy of DMGN was validated in the internal and external cohorts. Furthermore, among patients with low-risk scores in the combined training and internal cohorts, concurrent chemotherapy improved distant metastasis-free survival compared with those patients who did not receive concurrent chemotherapy (HR 0·40, 95% CI 0·19–0·83; p=0·011), whereas patients with high-risk scores did not benefit from concurrent chemotherapy (HR 1·03, 0·71–1·50; p=0·876). This was also validated in the two external cohorts combined. We developed a nomogram based on the DMGN and other variables that predicted an individual's risk of distant metastasis, which was strengthened by adding Epstein–Barr virus DNA status.

**Interpretation** The DMGN is a reliable prognostic tool for distant metastasis in patients with locoregionally advanced nasopharyngeal carcinoma and might be able to predict which patients benefit from concurrent chemotherapy. It has the potential to guide treatment decisions for patients at different risk of distant metastasis.

**Funding** The National Natural Science Foundation of China, the National Science & Technology Pillar Program during the Twelfth Five-year Plan Period, the Natural Science Foundation of Guang Dong Province, the National Key Research and Development Program of China, the Innovation Team Development Plan of the Ministry of Education, the Health & Medical Collaborative Innovation Project of Guangzhou City, China, and the Program of Introducing Talents of Discipline to Universities.

## Introduction

Nasopharyngeal carcinoma has a unique geographical distribution, with the highest incidence in southeast Asia.[1] Of the 87 000 newly diagnosed cases annually, more than 70% of patients with nasopharyngeal carcinoma are classified as having locoregionally advanced disease.[2] The standard treatment for patients with locoregionally advanced nasopharyngeal carcinoma is radiotherapy combined with chemotherapy.[3] With the use of intensity-modulated radiotherapy and combined chemoradiotherapy, locoregional control has improved substantially and distant metastasis is now the main reason for treatment failure.[4]

*Lancet Oncol* 2018

Published **Online** February 7, 2018 http://dx.doi.org/10.1016/S1470-2045(18)30080-9

See **Online/Comment** http://dx.doi.org/10.1016/S1470-2045(18)30085-8

*Contributed equally to this Article

†Senior authors contributed equally to this work

**Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine, Guangzhou, China** (X-R Tang MD, Y-Q Li PhD, S-B Liang MD, L-L Tang PhD, Y-P Mao PhD, Q-M He MD, X-J Yang BS, Y Zhang MD, X Wen MD, J Zhang MD, Y-Q Wang MD, P-P Zhang MD, Prof Y Sun PhD, Prof J-P Yun PhD, J Zeng PhD, Prof L Li PhD, L-Z Liu PhD, N Liu PhD, Prof J Ma MD); **Department of Radiation Oncology, The First People's Hospital of Foshan, Foshan, China** (F Liu PhD, S-B Liang); **Department of Radiation Oncology, Affiliated Hospital of Guilin Medical University, Guilin, China** (W Jiang PhD); **and Department of Probability and Statistics, School of Mathematical Sciences, South China Normal University, Guangzhou, China** (W-X Ge PhD)

Correspondence to: Prof Jun Ma, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine, 651 Dongfeng Road East, Guangzhou 510060, China **majun2@mail.sysu.edu.cn**

## Research in context

### Evidence before this study

We searched PubMed up to Aug 23, 2017, for research articles containing the terms "gene expression profiling AND locoregionally advanced nasopharyngeal carcinoma AND prognosis", without date or language restrictions. Our search did not identify any previous high-throughput studies that had investigated the potential prognostic role of gene-expression profiles in locoregionally advanced nasopharyngeal carcinoma.

### Added value of this study

We did a multicentre, retrospective study to test the ability of mRNA expression profiles of locoregionally advanced nasopharyngeal carcinoma tumours to predict the risk of distant metastasis at the time of diagnosis. Our study included 937 patients and is therefore, to our knowledge, the largest biomarker discovery project to be done in locoregionally advanced nasopharyngeal carcinoma. We developed and validated a distant metastasis gene signature for locoregionally advanced nasopharyngeal carcinoma (DMGN) that is based on the expression of 13 genes and can be used as a molecular classifier for distant metastasis. We show that this signature predicts risk of distant metastasis and which patients are most likely to benefit from concurrent chemotherapy in locoregionally advanced nasopharyngeal carcinoma and that the DMGN is more accurate than clinicopathological risk factors alone. We further developed a nomogram comprising DMGN, N stage, sex, and serum lactate dehydrogenase and C-reactive protein concentrations that predicted distant metastasis-free survival in the training cohort and was validated both internally and externally.

### Implications of all the available evidence

The DMGN is a reliable prognostic tool for distant metastasis in patients with locoregionally advanced nasopharyngeal carcinoma and directly quantifies mRNA expression from paraffin-embedded nasopharyngeal carcinoma tissues based on a commercially available mRNA NanoString platform, making it easy to implement in clinical practice. We believe that future prospective validation combining the DMGN with Epstein–Barr virus DNA status will improve the basis for clinical decision making for patients with locoregionally advanced nasopharyngeal carcinoma.

The tumour–node–metastasis (TNM) staging system is the key determinant for prognostic prediction and risk stratification for treatment decisions. Even though patients with locoregionally advanced nasopharyngeal carcinoma with the same TNM stage receive similar treatments, their clinical outcomes vary greatly. Approximately 30–40% of patients with locoregionally advanced nasopharyngeal carcinoma eventually develop distant metastasis after undergoing radical treatment.[5] Hence, the current anatomical-based staging system is not sufficient to predict which patients will develop distant metastasis. The differences in prognosis might be attributed to biological heterogeneity, and molecular investigation could provide biomarkers for predicting distant metastasis and guiding treatment decisions for patients in different risk groups. Great efforts have been made to search for molecular biomarkers for locoregionally advanced nasopharyngeal carcinoma, such as Epstein–Barr virus DNA (EBV DNA) concentration, lactate dehydrogenase (LDH) concentration, and miRNAs.[6–8] A randomised controlled trial using EBV DNA status to direct chemoradiotherapy in nasopharyngeal carcinoma is ongoing.[9] However, new biomarkers that reflect tumour heterogeneity still need to be identified to guide individual treatment for patients.

See Online for appendix

Comparisons of the global gene-expression profiles of different cell types, developmental stages, and disease states could promote the discovery of characteristic gene signatures that might be correlated with functionally important states. Emerging evidence demonstrates that gene-expression profiling of formalin-fixed paraffin-embedded (FFPE) tissues is useful for molecular classification and prediction for prognoses and treatment responses in various types of cancers.[10–13] In nasopharyngeal carcinoma, recent evidence also indicates that certain gene-expression patterns might be used as molecular markers that allow early diagnosis and subgroup classification.[14–16] However, how expression patterns might differ between patients with different clinical outcomes, especially distant metastasis, has not been established.

We did a study of global gene expression with the aim of identifying and validating a gene-expression signature that predicts distant metastasis in patients with locoregionally advanced nasopharyngeal carcinoma. Moreover, we combined both genomic and clinical variables to generate a nomogram model with more accurate prediction than clinical risk factors for distant metastasis.

## Methods

### Clinical specimens and study design

We retrospectively collected 955 paraffin-embedded biopsy tissues from patients with stage III–IVa locoregionally advanced nasopharyngeal carcinoma, of which 937 passed quality control for the final analysis (appendix p 22). 614 samples were obtained between Jan 14, 2006, and Dec 27, 2009, from patients treated at the Sun Yat-sen University Cancer Center (Guangzhou, China). Computer-generated random numbers were used to assign these samples into a training cohort (Guangzhou training cohort) consisting of 410 samples and an internal validation cohort (Guangzhou internal validation cohort) consisting of 204 samples. 165 samples collected from the Affiliated Hospital of Guilin Medical College (Guilin, China; Guilin external validation cohort) between Feb 26, 2007, and July 23, 2011, and 158 samples collected from the First People's Hospital of Foshan (Foshan, China; Foshan

external validation cohort) between April 27, 2010, and March 3, 2014, were separately used as two independent external validation cohorts. All samples were reassessed by two pathologists (J-PY and JZ), and found to contain more than 70% tumour cells. None of the patients had received any anti-tumour therapy before biopsy sampling.

All patients underwent staging MRI, and patients with contraindications for MRI received contrast-enhanced CT scans. Two radiologists (L-ZL and LL) separately reassessed all MRI and CT scans, and any disagreements were resolved by consensus. We restaged all patients according to the 8th edition of the American Joint Committee on Cancer Staging Manual.[17] Radiotherapy dose and chemotherapy regimens are detailed in the appendix (p 2). The institutional ethical review boards of all included hospitals approved this retrospective analysis of anonymous data, and the requirement for informed consent was waived by the ethics review boards.

### Procedures

Total RNA was extracted from FFPE samples using the QIAGEN FFPE RNeasy kit (QIAGEN GmbH, Hilden, Germany). RNA was analysed using a Nanodrop 2000 spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA), and RNA integrity numbers were determined to evaluate RNA integration using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Total RNA was amplified using an Ovation FFPE WTA System (NuGEN, San Carlos, CA, USA), and an Encore Biotin Module (NuGEN) was used for fragmentation and labelling.

Total RNA was hybridised to Affymetrix Human Gene 2.0 ST GeneChips (Affymetrix, Santa Clara, CA, USA) using a GeneChip Hybridization, Wash and Stain Kit (Affymetrix) in a Hybridization Oven 645 (Affymetrix) and Fluidics Station 450 (Affymetrix) according to the manufacturer's instructions. All slides were scanned using a GeneChip Scanner 3000 (Affymetrix) and Command Console Software 3.1 (Affymetrix) with default settings.

For data processing, raw data were normalised using the oligo package in R with Robust Multi-array Analysis (RMA). We implemented Bayes statistics using empirical Bayes (eBayes) in the limma package to compute which probes were significantly differentially expressed between metastatic and non-metastatic samples. Because RNA obtained from FFPE samples is highly and randomly fragmented, we used Fisher's exact test to verify whether a transcript was indeed significantly differentially expressed. Differentially expressed transcripts were characterised according to the following criteria: an empirical fold-change greater than 1·5 and an eBayes test p value less than 0·05. The methods for data analysis are described in detail in the appendix (p 2).

### Statistical analysis

Expression profiling studies often use high-throughput assays to screen differentially expressed profiles in small subgroups of patients, then validate the findings using low-throughput methods in larger populations.[18,19] Based on this approach, we first compared gene expression profiles across 24 paired locoregionally advanced nasopharyngeal carcinoma tumour tissues from 48 patients at Sun Yat-sen University Cancer Center with or without distant metastasis after curative treatment using the Affymetrix Human Gene 2.0 ST microarray (appendix pp 5, 6). Tissues were strictly matched by sex, age, T stage, N stage, radiotherapy, and chemotherapy to rule out the influence of these parameters on metastasis. We then analysed the differentially expressed genes using digital expression profiling (NanoString nCounter system, NanoString Technologies, Seattle, WA, USA), and generated a distant metastasis gene signature for locoregionally advanced nasopharyngeal carcinoma (DMGN) in the training cohort. Raw NanoString counts for each gene within each experiment were subjected to technical normalisation before biological normalisation (appendix p 3). We used the counts obtained for the appropriate positive control probe set, which consisted of five housekeeping genes included in the NanoString CodeSet (appendix pp 7–10). The normalised data were then log2-transformed and used as the input for class prediction analysis. The methods used to select reference genes and apply quality control criteria for the NanoString data are described in detail in the appendix (pp 2, 3).

We used a penalised logistic model to select variables for constructing a prediction model. The R package glmnet was used to perform the least absolute shrinkage and selection operator (LASSO) on the logistic regression model. Bootstrap methods were used to test the robustness of the significant candidate genes that were chosen by the penalised logistic model. Then, we calculated the area under the receiver-operator characteristic (ROC) curve (AUC) to determine how many candidate genes should be chosen (appendix p 23), and we constructed the prediction model with the coefficients weighted by the Cox model in the training cohort.[20] We used X-tile software (version 3.6.1; Yale University, New Haven, CT, USA) in the training cohort to find the optimal cutoff value. The thresholds for the scores that were outputted from the predictive model that was used to separate patients into low-risk and high-risk groups were defined as the score that produced the largest $\chi^2$ value in the Mantel–Cox test.[21]

Our primary endpoint was distant metastasis-free survival. Secondary endpoints were disease-free survival and overall survival. We calculated distant metastasis-free survival as the time from day 1 of concurrent chemoradiotherapy or induction chemotherapy (if given) to the date of first distant relapse; disease-free survival to the first relapse at any site, death from any cause, or the date of the last follow-up visit, whichever occurred first; and overall survival to death from any cause.

Distant metastasis-free survival, overall survival, and disease-free survival were calculated using the Kaplan-Meier method and the log-rank test, and hazard

ratios (HRs) were calculated using a univariate Cox regression analysis. The DMGN as a predictor for concurrent chemotherapy efficacy was analysed in the combined training and internal cohorts and validated in a combination cohort of the two external cohorts.

We also did a multivariate Cox regression analysis using backward selection to test the independent significance of different factors; the p value threshold was 0·1 (p>0·1) for removing non-significant variables from the analysis, and marginally significant variables (0·05<p<0·1) remained in the final Cox model. Covariates included DMGN (high risk *vs* low risk), sex, age (≥45 years *vs* <45 years), N stage (N2–3 *vs* N0–1), T stage (T3–4 *vs* T1–2), PET–CT (no *vs* yes), viral capsid antigen immunoglobulin (VCA-IgA; titre ≥1:80 *vs* <1:80), early antigen immunoglobulin (EA-IgA; titre ≥1:10 *vs* <1:10), serum LDH (≥245 *vs* <245 U/L), C-reactive protein (≥8·2 *vs* <8·2 mg/L), haemoglobin (<120 *vs* ≥120 g/L), body-mass index (BMI; ≥18·5 *vs* <18·5 kg/m²), and concurrent chemotherapy (no *vs* yes).

For **microarray data** see https://www.ncbi.nlm.nih.gov/geo/

We formulated nomograms using the rms package in R, and included serum LDH and C-reactive protein in the nomogram as they are usually included in most prognostic models of nasopharyngeal carcinoma.[22,23] We used the coefficients of the multivariable Cox regression model to generate nomograms. Calibration curves were assessed graphically by plotting the observed rates against the nomogram predicted probabilities and a concordance index (C-index) was calculated via a bootstrap method with 1000 resamples. We investigated the prognostic or predictive accuracy of clinical features and the multi-gene-based classifier using ROC analysis.[24] We calculated the AUC to assess the sensitivity and specificity of the model for predicting distant metastasis. Statistical analyses were done in R (version 3.2.1) and SPSS (version 22.0). All statistical tests were two-sided, and p values of less than 0·05 were deemed significant.

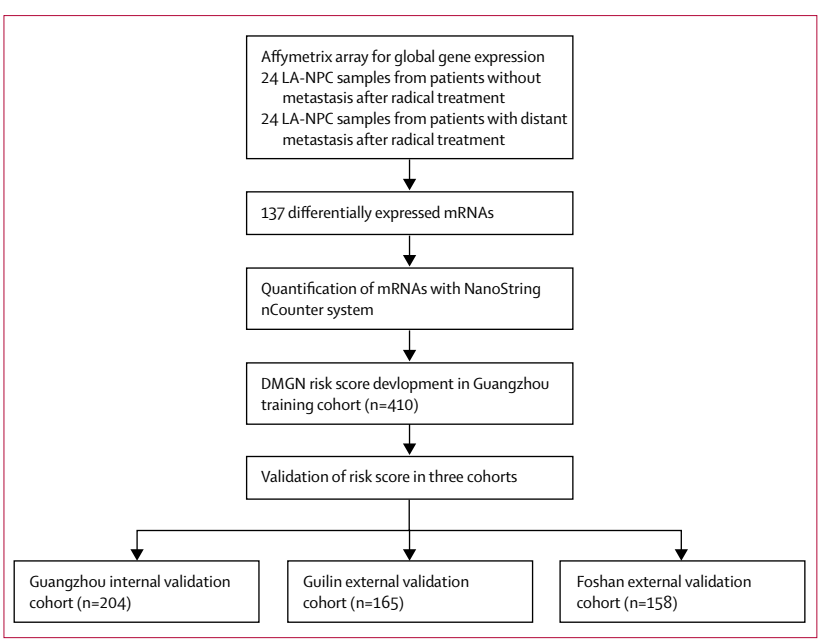


***Figure 1:*** **Study flow**
LA-NPC=locoregionally advanced nasopharyngeal carcinoma. DMGN=distant metastasis gene signature for locoregionally advanced nasopharyngeal carcinoma.

### Data sharing

The key raw data have been uploaded onto the Research Data Deposit public platform (RDD), with the approval RDD number of RDDB2017000204. The microarray data have been deposited online under accession number GSE103611.

### Role of the funding source

The sponsors had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to the data and had final responsibility for the decision to submit the manuscript for publication.

## Results

We included in our analysis 937 pretreatment, non-metastatic nasopharyngeal carcinoma specimens that were obtained from three academic institutions in China (figure 1). The table shows the clinical characteristics of the patients in the Guangzhou training cohort (n=410), Guangzhou internal validation cohort (n=204), and two external validation cohorts (Guilin, n=165; Foshan, n=158). All patients underwent radical radiotherapy, and platinum-based chemotherapy was administered to 516 (84%) of 614 patients in the Guangzhou cohort, 154 (93%) of 165 in the Guilin cohort, and 147 (93%) of 158 in the Foshan cohort (appendix p 11). Median follow-up was 78·2 months (IQR 49·0–89·4) for patients in the Guangzhou cohort, 68·2 months (49·9–84·4) for those in the Guilin cohort, and 49·9 months (38·9–64·4) for those in the Foshan cohort.

In the microarray analysis, 137 mRNAs were found to be differentially expressed between 24 patients with non-metastatic nasopharyngeal carcinoma and 24 patients with nasopharyngeal carcinoma who developed distant metastasis after treatment. We found 13 of these differentially expressed genes were associated with distant metastasis-free survival in the training cohort (appendix p 23). A risk score was calculated for each patient using a formula derived from the expression levels of these 13 genes weighted by their regression coefficient (appendix p 12):

Risk score = (0·1846 × expression of *YBX3*) – (0·3007 × expression of *CBR3*) – (0·1383 × expression of *CXCL10*) – (0·3661 × expression of *CLASP1*) + (0·2381 × expression of *DCTN1*) – (0·4004 × expression of *FNDC3B*) + (0·60 × expression of *WSB2*) + (0·1093 × expression of *LRIG1*) – (0·1162 × expression of *GRM4*) + (0·1327 × expression of *ANXA1*) + (0·1485 × expression of *WNK1*) + (0·0714 × expression of *HDLBP*) + (0·1774 × expression of *POLR2M*)

| | Guangzhou training cohort (n=410) | | | Guangzhou internal validation cohort (n=204) | | | Guilin external validation cohort (n=165) | | | Foshan external validation cohort (n=158) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Patients (n) | Low risk | High risk | Patients (n) | Low risk | High risk | Patients (n) | Low risk | High risk | Patients (n) | Low risk | High risk |
| **Age (years)** | | | | | | | | | | | | |
| <45 | 181 | 104 (57%) | 77 (43%) | 89 | 44 (49%) | 45 (51%) | 64 | 39 (60%) | 25 (40%) | 63 | 34 (54%) | 29 (46%) |
| ≥45 | 229 | 103 (45%) | 126 (55%) | 115 | 61 (50%) | 54 (50%) | 101 | 57 (56%) | 44 (44%) | 95 | 53 (56%) | 42 (44%) |
| **Sex** | | | | | | | | | | | | |
| Male | 314 | 147 (47%) | 167 (53%) | 153 | 75 (49%) | 78 (51%) | 121 | 62 (51%) | 59 (49%) | 121 | 63 (52%) | 58 (48%) |
| Female | 96 | 60 (63%) | 36 (38%) | 51 | 30 (59%) | 21 (41%) | 44 | 34 (77%) | 10 (23%) | 37 | 24 (65%) | 13 (35%) |
| **WHO pathological type** | | | | | | | | | | | | |
| Undifferentiated non-keratinising | 398 | 201 (51%) | 197 (49%) | 198 | 101 (51%) | 97 (49%) | 164 | 96 (59%) | 68 (41%) | 158 | 87 (55%) | 71 (45%) |
| Differentiated non-keratinising | 12 | 6 (50%) | 6 (50%) | 6 | 4 (67%) | 2 (33%) | 1 | 0 (0%) | 1 (100%) | 0 | 0 (0%) | 0 (0%) |
| **T stage** | | | | | | | | | | | | |
| T1 | 15 | 6 (40%) | 9 (60%) | 9 | 4 (44%) | 5 (56%) | 4 | 3 (75%) | 1 (25%) | 12 | 5 (42%) | 7 (58%) |
| T2 | 60 | 28 (47%) | 32 (53%) | 28 | 16 (57%) | 12 (43%) | 31 | 18 (58%) | 13 (42%) | 11 | 6 (55%) | 5 (45%) |
| T3 | 253 | 138 (55%) | 115 (45%) | 118 | 65 (55%) | 53 (45%) | 84 | 46 (55%) | 38 (45%) | 75 | 45 (60%) | 30 (40%) |
| T4 | 82 | 35 (43%) | 47 (57%) | 49 | 20 (41%) | 29 (59%) | 46 | 29 (63%) | 17 (37%) | 60 | 31 (52%) | 29 (48%) |
| **N stage** | | | | | | | | | | | | |
| N0 | 44 | 23 (52%) | 21 (48%) | 17 | 7 (41%) | 10 (59%) | 12 | 7 (58%) | 5 (42%) | 16 | 7 (44%) | 9 (56%) |
| N1 | 192 | 105 (55%) | 87 (45%) | 80 | 40 (50%) | 40 (50%) | 54 | 35 (65%) | 19 (35%) | 71 | 41 (48%) | 30 (42%) |
| N2 | 103 | 48 (47%) | 55 (53%) | 65 | 38 (58%) | 27 (42%) | 82 | 46 (56%) | 36 (44%) | 62 | 34 (55%) | 28 (45%) |
| N3 | 71 | 31 (44%) | 40 (56%) | 42 | 20 (48%) | 22 (52%) | 17 | 8 (47%) | 9 (53%) | 9 | 5 (56%) | 4 (44%) |
| **TNM stage** | | | | | | | | | | | | |
| III | 264 | 145 (55%) | 119 (45%) | 120 | 68 (57%) | 52 (43%) | 103 | 61 (59%) | 42 (41%) | 91 | 52 (57%) | 39 (43%) |
| IV | 146 | 62 (42%) | 84 (58%) | 84 | 37 (44%) | 47 (56%) | 62 | 35 (56%) | 27 (44%) | 67 | 35 (52%) | 32 (48%) |
| **VCA-IgA** | | | | | | | | | | | | |
| <1:80 | 45 | 24 (53%) | 21 (47%) | 22 | 12 (55%) | 10 (45%) | 119 | 66 (55%) | 53 (45%) | NA | NA | NA |
| ≥1:80 | 365 | 183 (50%) | 182 (50%) | 182 | 93 (51%) | 89 (49%) | 46 | 30 (65%) | 16 (35%) | NA | NA | NA |
| **EA-IgA** | | | | | | | | | | | | |
| <1:10 | 76 | 41 (54%) | 35 (46%) | 43 | 27 (63%) | 16 (37%) | 119 | 66 (55%) | 53 (45%) | NA | NA | NA |
| ≥1:10 | 334 | 166 (50%) | 168 (50%) | 161 | 78 (48%) | 83 (52%) | 46 | 30 (65%) | 16 (35%) | NA | NA | NA |
| **LDH concentration (U/L)** | | | | | | | | | | | | |
| <245 | 385 | 198 (51%) | 187 (49%) | 183 | 92 (50%) | 91 (50%) | NA | NA | NA | 145 | 78 (54%) | 67 (46%) |
| ≥245 | 25 | 9 (36%) | 16 (64%) | 21 | 13 (62%) | 8 (38%) | NA | NA | NA | 13 | 9 (69%) | 4 (31%) |
| **C-reactive protein concentration (mg/L)** | | | | | | | | | | | | |
| <8·20 | 339 | 171 (50%) | 168 (50%) | 167 | 90 (54%) | 77 (46%) | NA | NA | NA | 145 | 79 (54%) | 66 (46%) |
| ≥8·20 | 71 | 36 (51%) | 35 (49%) | 37 | 15 (41%) | 22 (59%) | NA | NA | NA | 13 | 8 (62%) | 5 (39%) |
| **Haemoglobin concentration (g/L)** | | | | | | | | | | | | |
| <120 | 111 | 58 (52%) | 53 (48%) | 47 | 19 (40%) | 28 (60%) | 70 | 49 (70%) | 21 (30%) | 20 | 9 (45%) | 11 (55%) |
| ≥120 | 299 | 149 (50%) | 150 (50%) | 157 | 86 (55%) | 71 (45%) | 95 | 47 (49%) | 48 (51%) | 138 | 78 (57%) | 60 (43%) |
| **Body-mass index (kg/m²)** | | | | | | | | | | | | |
| <18·5 | 21 | 13 (62%) | 8 (38%) | 8 | 4 (50%) | 4 (50%) | 17 | 14 (82%) | 3 (18%) | 21 | 11 (52%) | 10 (48%) |
| ≥18·5 | 389 | 194 (50%) | 195 (50%) | 196 | 101 (52%) | 95 (48%) | 148 | 82 (55%) | 66 (45%) | 137 | 76 (55%) | 61 (45%) |
| **Concurrent chemotherapy** | | | | | | | | | | | | |
| Yes | 204 | 100 (49%) | 104 (51%) | 122 | 59 (48%) | 63 (52%) | 77 | 45 (58%) | 32 (42%) | 127 | 65 (51%) | 62 (49%) |
| No | 206 | 107 (52%) | 99 (48%) | 82 | 46 (56%) | 36 (44%) | 88 | 51 (58%) | 37 (42%) | 31 | 22 (71%) | 9 (29%) |

Data are n (%) unless otherwise stated. TNM=tumour–node–metastasis.VCA-IgA=viral capsid antigen immunoglobulin A. EA-IgA=early antigen immunoglobulin A. LDH=serum lactate dehydrogenase. NA=not available.

***Table:*** **Clinical characteristics of patients in the training, internal, and external validation cohorts**

We used X-tile plots to generate an optimal cutoff value (5·01) to separate patients into low-risk and high-risk groups in the Guangzhou training cohort (appendix p 24). This assay assigned 207 (50·5%) of 410 patients to the low-risk group and 203 (49·5%) to the high-risk group. 5-year distant

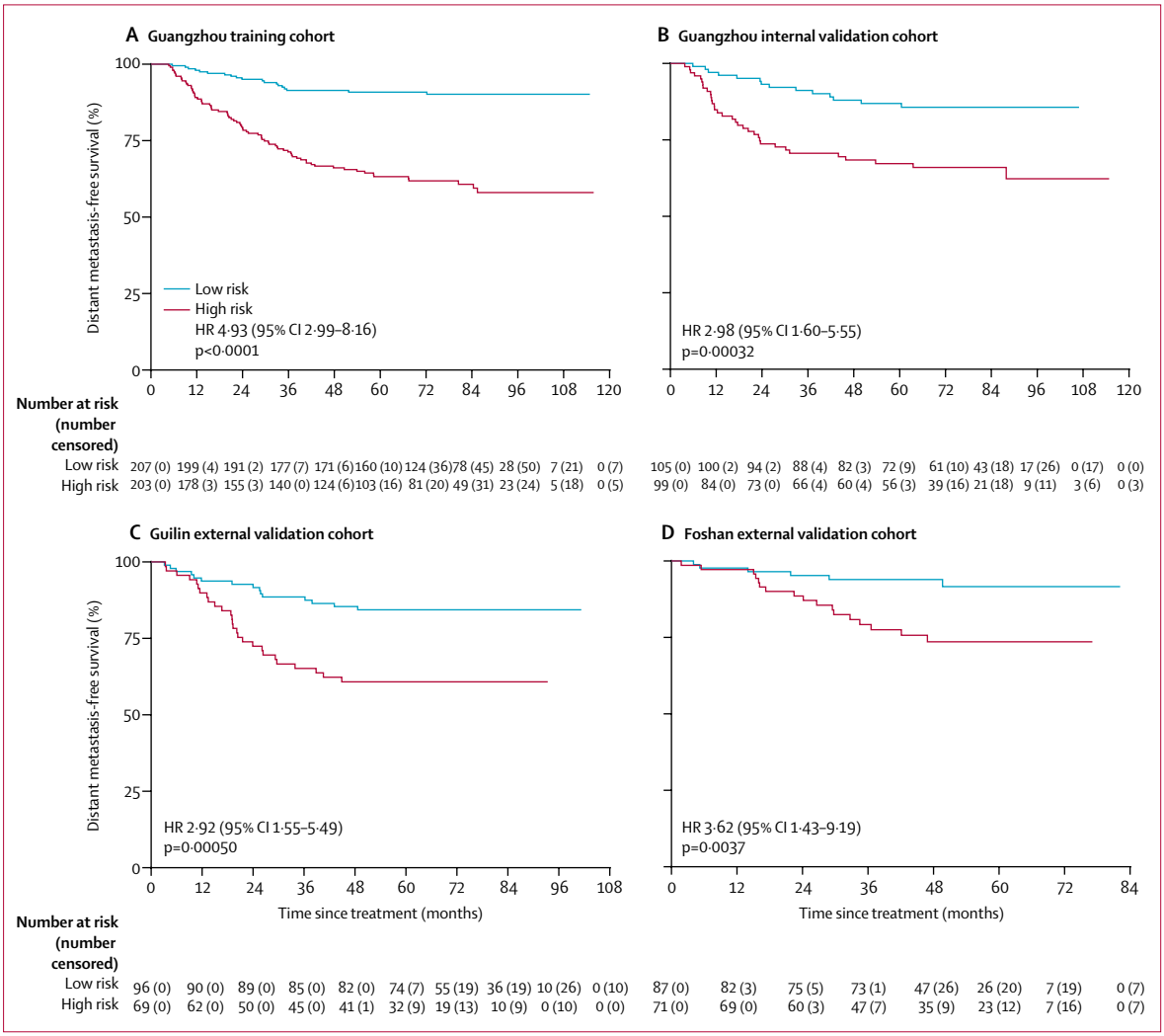


***Figure 2:*** **Kaplan-Meier curves of distant metastasis-free survival according to the DMGN**
(A) Guangzhou training cohort (n=410), (B) Guangzhou internal validation cohort (n=204), (C) Guilin external validation cohort (n=165), and (D) Foshan external validation cohort (n=158). We calculated p values using the unadjusted log-rank test and hazard ratios using a univariate Cox regression analysis. DMGN=distant metastasis gene signature for locoregionally advanced nasopharyngeal carcinoma. HR=hazard ratio.

metastasis-free survival was 63·3% (95% CI 56·1–69·6) for the high-risk group and 90·9% (85·9–94·2) in the low-risk group (HR 4·93, 95% CI 2·99–8·16; p<0·0001; figure 2A; appendix pp 13, 14). Patients with high-risk scores also had shorter disease-free survival (HR 3·51, 95% CI 2·43–5·07; p<0·0001; appendix pp 13, 25) and overall survival (HR 3·22, 2·18–4·76; p<0·0001; appendix pp 13, 26) than patients with low-risk scores. The 5-year distant metastasis-free survival, disease-free survival, and overall survival in each risk group and the number of patients who had an event for each risk group are listed in the appendix (pp 13, 14).

In the Guangzhou internal validation cohort, the DMGN categorised 105 (51%) of 204 patients into the low-risk group and 99 patients (49%) into the high-risk group, which were significantly different in terms of distant metastasis-free survival (HR 2·98, 95% CI 1·60–5·55; p=0·00032; figure 2B), disease-free survival (HR 2·14, 1·30–3·54; p=0·0023; appendix p 25), and overall survival (2·00, 1·22–3·27; p=0·0049; appendix p 26). The 5-year distant metastasis-free survival, disease-free survival, and overall survival in each risk group and the number of patients who had an event for each risk group are listed in the appendix (pp 13, 14).

In the Guilin external validation cohort, the DMGN classified 96 (58%) of 165 patients into the low-risk group and 69 (42%) patients into the high-risk group. Patients with high-risk scores had shorter distant metastasis-free survival (HR 2·92, 95% CI 1·55–5·49; p=0·00050; figure 2C), disease-free survival (HR 2·12, 1·29–3·48; p=0·0024; appendix p 25), and overall survival (HR 2·33, 1·35–4·04; p=0·0018; appendix p 26)

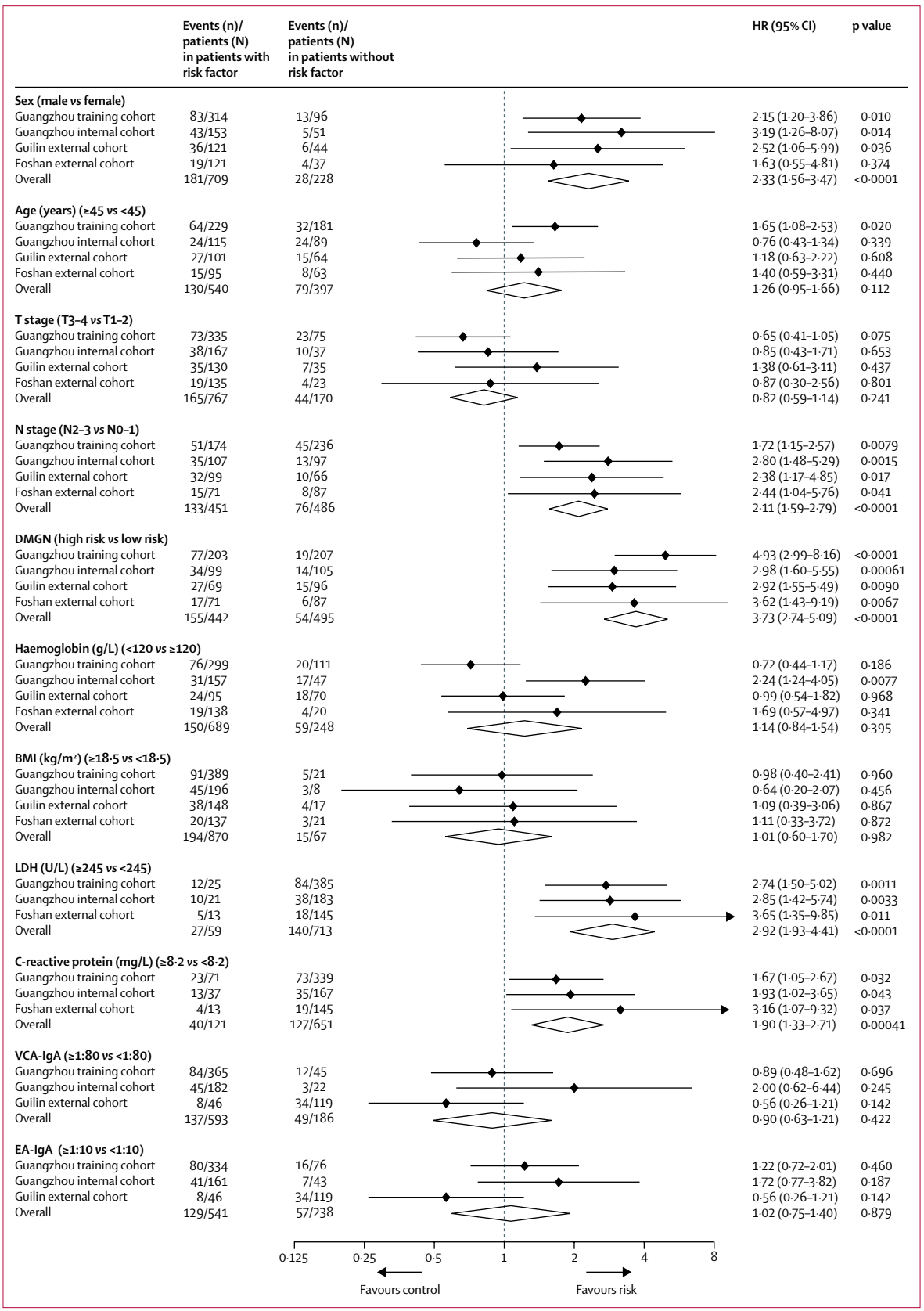


***Figure 3*: Univariate association of DMGN and clinicopathological characteristics with distant metastasis-free survival**
DMGN=distant metastasis gene signature for locoregionally advanced nasopharyngeal carcinoma. BMI=body-mass index. LDH=serum lactate dehydrogenase. VCA-IgA=viral capsid antigen immunoglobulin A. EA-IgA=early antigen immunoglobulin A. HR=hazard ratio.

than patients with low-risk scores. Similarly, in the Foshan external validation cohort, 87 (55%) of 158 patients were stratified into the low-risk group and 71 (45%) patients were stratified into the high-risk group; patients with high-risk scores had shorter distant metastasis-free survival (HR 3·62, 95% CI 1·43–9·19; p=0·0037; figure 2D), disease-free survival (HR 3·10, 1·48–6·47; p=0·0016; appendix p 25), and overall survival (HR 2·45, 1·19–5·02; p=0·012; appendix p 26). 5-year distant metastasis-free survival, disease-free survival, and overall survival in each risk group and the number of patients who had an event for each risk group are listed in the appendix (pp 13, 14).

Figure 3 shows univariate analysis of distant metastasis-free survival by clinical and biological subgroups in the four cohorts. The DMGN was significantly associated with distant metastasis-free survival in all four cohorts (figure 3). Analysis using different categorisations for N stage (N3 *vs* N0–2) and T stage (T4 *vs* T1–3) showed similar results to the original dichotomisations of N stage (N2–3 *vs* N0–1) and T stage (T3–4 *vs* T1–2; appendix p 27). After multivariable adjustment by clinicopathological variables and several serum markers, the DMGN remained a strong independent prognostic factor for distant metastasis-free survival in the Guangzhou training cohort (HR 4·51, 95% CI 2·72–7·47; p<0·0001;

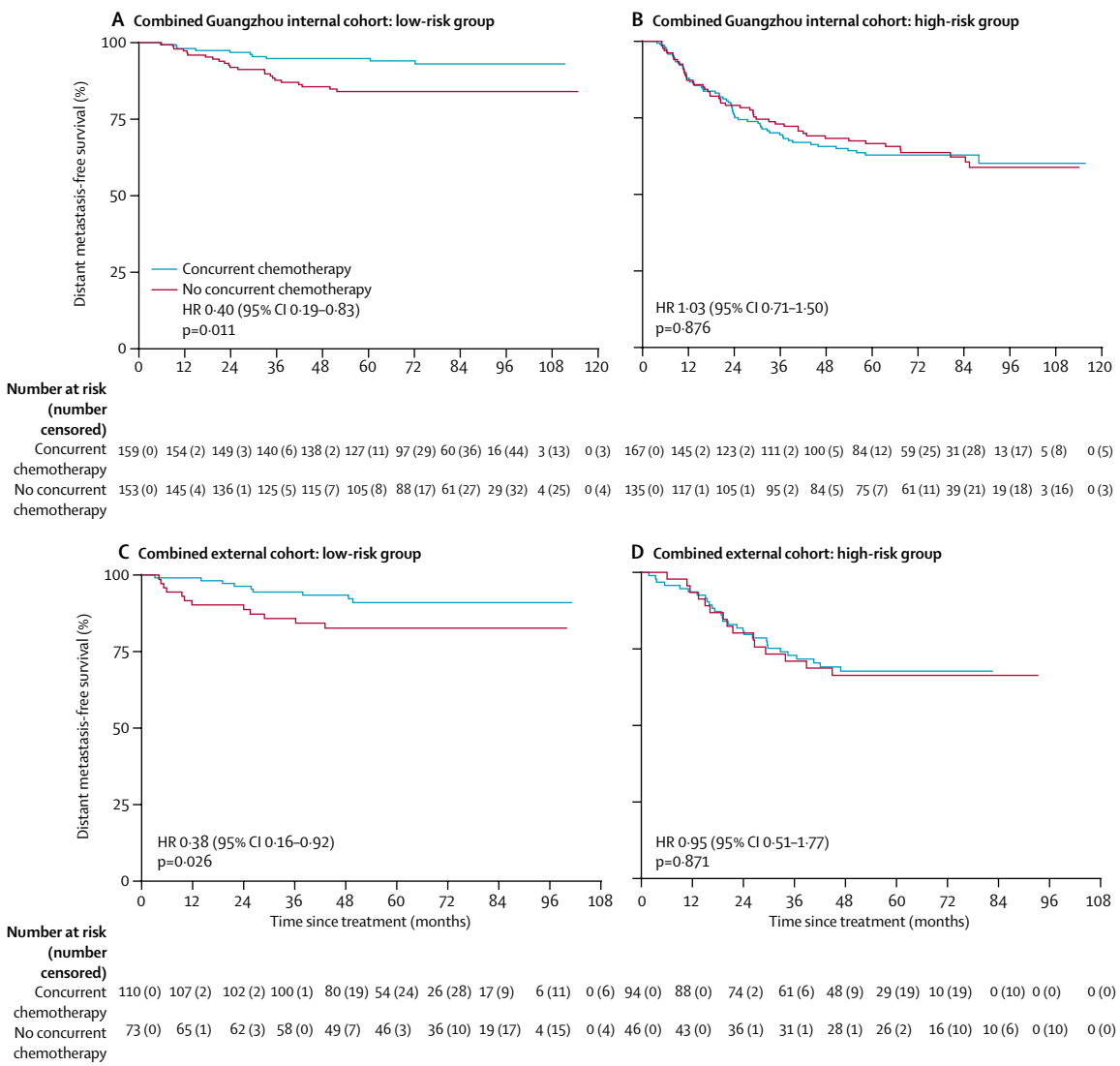


***Figure 4:*** **Kaplan-Meier curves of distant metastasis-free survival according to low-risk or high-risk scores stratified by administration of concurrent chemotherapy**
(A) DMGN-defined low-risk patients in the combined Guangzhou internal cohort (n=312), (B) DMGN-defined high-risk patients in the combined Guangzhou internal cohort (n=302), (C) DMGN-defined low-risk patients in the combined external cohort (n=183), and (D) DMGN-defined high-risk patients in the combined external cohort (n=140). We calculated p values using the unadjusted log-rank test and hazard ratios using a univariate Cox regression analysis. DMGN=distant metastasis gene signature for locoregionally advanced nasopharyngeal carcinoma. HR=hazard ratio.

appendix p 15), as well as in the Guangzhou internal validation cohort (HR 3·32, 1·74–6·33; p=0·00028; appendix p 16) and the two external cohorts (Guilin: HR 2·48, 1·31–4·71; p=0·0055; and Foshan: HR 4·88, 1·85–12·8; p=0·0014; appendix pp 17, 18).

We analysed whether the DMGN could be used to predict the efficacy of concurrent chemotherapy in locoregionally advanced nasopharyngeal carcinoma. 326 (53%) of 614 patients in the combined Guangzhou cohort received concurrent chemoradiation. When the Kaplan-Meier survival analysis was stratified with the DMGN, treatment with concurrent chemotherapy compared with no concurrent chemotherapy was associated with improved distant metastasis-free survival (HR 0·40, 95% CI 0·19–0·83; p=0·011), disease-free survival (HR 0·56, 0·34–0·94; p=0·025), and overall survival (HR 0·58, 0·35–0·97; p=0·036) in patients with low DMGN scores but not in patients with high DMGN scores (distant metastasis-free survival: HR 1·03, 0·71–1·50; p=0·876; figure 4; appendix pp 28, 29 shows the disease-free and overall survival results). This finding was validated in the combined external validation cohort, in which 204 (63%) of 323 patients received concurrent chemotherapy (figure 4; appendix pp 28, 29). The 5-year distant metastasis-free survival, disease-free survival, and overall survival in each risk group and the number of patients who had an event for each risk group are listed in the appendix (pp 13, 14).

We did a multivariate analysis of distant metastasis-free survival (appendix p 15) to generate nomogram A to predict distant metastasis-free survival in the training cohort (figure 5A). The predictors included DMGN, sex, N stage, and serum LDH and C-reactive protein concentrations. Among these, DMGN had the highest C-index (appendix p 19). The calibration plots for the 5-year distant metastasis-free survival were predicted well in the Guangzhou training cohort (C-index 0·727, 95% CI 0·679–0·775), the Guangzhou internal validation cohort (0·725, 0·654–0·796), the Guilin external validation cohort (0·693, 0·617–0·769), and the Foshan validation cohort (0·739, 0·639–0·839; figure 5B–E). ROC analysis to compare the sensitivity and specificity of the nomogram comprising the DMGN signature with clinicopathological risk factors showed that distant metastasis-free survival was more accurately predicted by the nomogram than by the risk factors in all cohorts (appendix p 30).

Since plasma EBV DNA concentration has been shown to be a useful prognostic biomarker in nasopharyngeal carcinoma, we did a subgroup analysis to explore whether or not it enhances the value of our present model. Pre-treatment EBV DNA data were available for 247 patients in the Guangzhou internal cohort and 122 patients in the Foshan external cohort (appendix p 21). First, we developed nomogram B using plasma EBV DNA status, sex, N stage, serum LDH, and C-reactive protein, and then built a final nomogram C using these five risk factors plus the DMGN using the 247 patients in the Guangzhou internal cohort (appendix p 31). Nomogram C (C-index 0·748, 95% CI 0·680–0·816) had better accuracy for predicting distant metastasis-free survival than nomogram B (0·686, 0·619–0·753; appendix p 20), which was confirmed using ROC analysis (p=0·037; appendix p 32). These findings were validated in the 122 patients from the Foshan external cohort (appendix pp 20, 32).

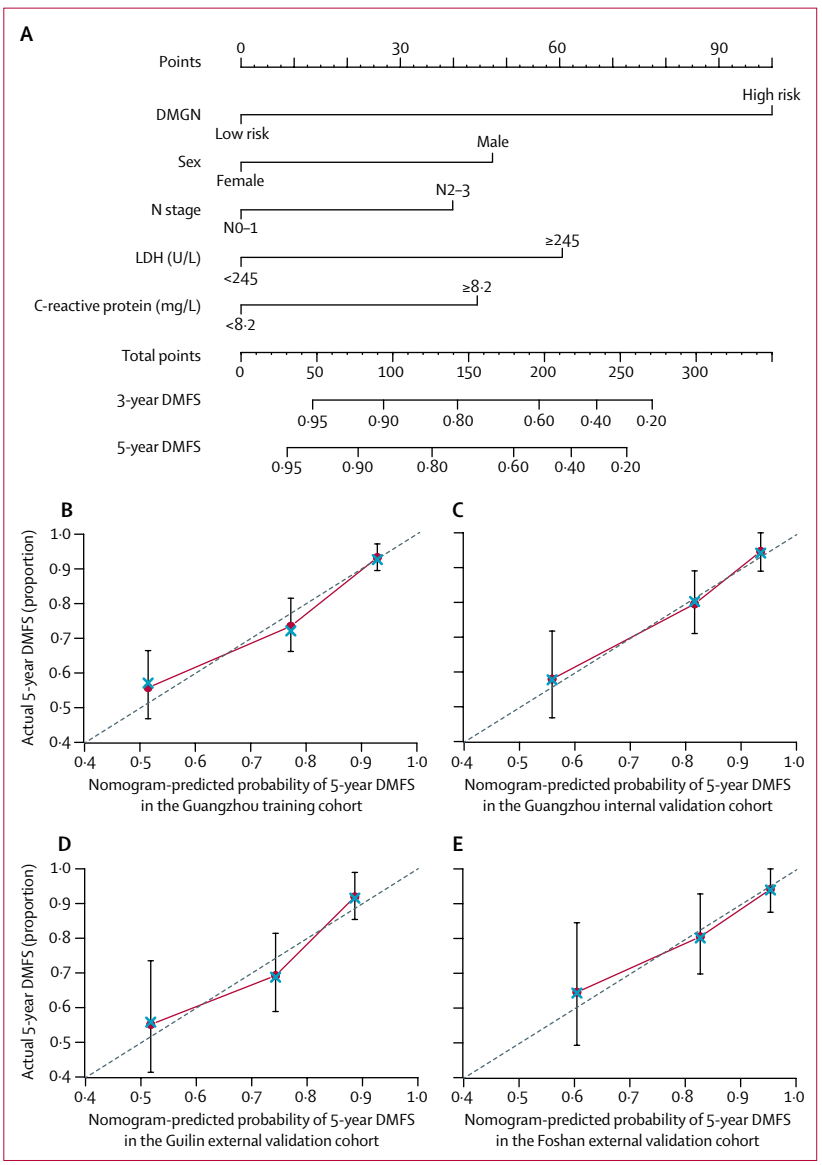


***Figure 5:*** **Nomogram A to predict the risk of distant metastasis in locoregionally advanced nasopharyngeal carcinoma**

(A) Nomogram A to predict distant metastasis-free survival. Calibration curves of the nomogram to predict distant metastasis-free survival at 5 years in (B) the Guangzhou training cohort, (C) the Guangzhou internal validation cohort, (D) the Guilin external validation cohort , and (E) the Foshan external validation cohort. The actual distant metastasis-free survival is plotted on the y-axis; nomogram predicted probability is plotted on the x-axis. DMGN=distant metastasis gene signature for locoregionally advanced nasopharyngeal carcinoma. LDH=serum lactate dehydrogenase. DMFS=distant metastasis-free survival.

## Discussion

In this multicentre, retrospective cohort study, we developed and validated a novel prognostic tool based on the expression of 13 genes that compared with clinical risk factors improves the ability to predict distant metastasis in patients with locoregionally advanced nasopharyngeal carcinoma. Our results showed that the DMGN developed in this study categorised patients into high-risk and low-risk groups of patients who had significantly different distant metastasis-free survival. Furthermore, our DMGN classifier was significantly better than other clinicopathological risk factors at predicting distant metastasis-free survival in patients with locoregionally advanced nasopharyngeal carcinoma. We also built a nomogram based on the DMGN and other variables that predicts individual risk of distant metastasis, which was strengthened by adding EBV DNA status.

Advances in molecular biology have resulted in the generation of large amounts of data that have been used to construct multigene profiles, which can be used for risk stratification and to guide strategies for chemotherapy treatment in various types of cancers.[10–13] Although a few gene-expression profiling studies have been done to identify genes that are differentially expressed between nasopharyngeal carcinoma and non-cancerous tissues, differences in expression profiles between patients with different prognoses are largely unexplored.[14–16] Moreover, while previous gene-expression profiling studies have yielded large numbers of markers, they did not subject them to large-scale, independent validation. In this study, we used microarray analysis to compare global gene-expression profiles between tissue samples obtained from paired locoregionally advanced nasopharyngeal carcinoma tumours with and without distant metastasis after radical treatment. Digital expression profiling was then used to generate a DMGN classifier that predicted the risk of distant metastasis in locoregionally advanced nasopharyngeal carcinoma, and the results were validated in an internal validation cohort and two external independent validation cohorts. Importantly, this classifier was applicable to patients receiving either 2D radiotherapy (Guilin cohort) or intensity-modulated radiation therapy (Foshan cohort). Considering that PET–CT is able to discover 6–12% more cases of distant metastasis than staging work-up with a conventional abdominal B ultrasound, chest radiography, and bone scan,[25,26] DMGN remained an independent prognostic factor after inclusion of PET–CT as a covariate in multivariable analysis. Additionally, the fact that our approach uses the US Food and Drug Administration-approved NanoString platform makes it easy to implement in clinical practice. Moreover, the mRNA platform is Clinical Laboratory Improvement Amendments certified, and was used to implement PAM50 for breast cancer subtyping.[27] This technology has been shown to be robust and reliable for quantifying RNA species that were extracted from FFPE tissues, and it is therefore a suitable platform for a gene expression-based clinical test.[13,28]

In this study, we showed that concurrent chemotherapy versus no concurrent chemotherapy provides a survival benefit to patients classified as low risk by our DMGN tool, but not in those classified as high risk. These results suggest that low-risk patients could be cured by concurrent chemotherapy, whereas high-risk patients are candidates for more aggressive therapeutic strategies to prevent tumour metastasis. Concurrent chemotherapy, due to its additive or synergistic effect with radiotherapy, provides a survival benefit mainly by improving locoregional control but not metastasis.[29] It is possible that concurrent chemotherapy does not effectively eradicate micrometastases in the high-risk group.

Cancer is a heterogeneous disease, and exploring the dysregulated genes involved in carcinogenesis and development might help to improve prognostic and therapeutic strategies. In the current study, we identified a group of 13 genes that effectively predict nasopharyngeal carcinoma distant metastasis and concurrent chemosensitivity. Among these genes, *FNDC3B* and *CLASP1* were previously shown to regulate cell motility and to thereby influence metastatic ability; and *ANXA1* and *CXCL10* were found to be involved in immune responses and the NF-κB signalling pathway, which influences tumour apoptosis, angiogenesis, and invasiveness.[30–35] Clinically promising results have been achieved by inhibiting VEGF and using immunotherapies, such as adoptive T-cell transfer, in patients with recurrent nasopharyngeal carcinoma.[36,37] Therefore, our DMGN could potentially be used as a predictive tool in personalised therapy, and might offer potential targets for therapeutic intervention in the clinical management of nasopharyngeal carcinoma.

We built a nomogram to predict an individual's risk of distant metastasis. Despite the fact that such methods generally use traditional prognostic factors, such as N stage and sex, we propose that including our DMGN reflects the biological heterogeneity of these tumours, whereas blood parameters, such as serum C-reactive protein and LDH concentrations, provide insights into a patient's systemic inflammatory status and elevated concentrations of these are associated with inferior survival in nasopharyngeal carcinoma.[7,22] The performance of the nomogram was verified in all validation cohorts. Moreover, the addition of EBV DNA status to this nomogram improved predictive accuracy, possibly because EBV DNA concentration reflects tumour burden. Thus, our nomogram might provide a simple and accurate method for predicting prognoses in locoregionally advanced nasopharyngeal carcinoma.

Our study had several limitations. First, our data were obtained in an endemic area in China where the nasopharyngeal cancer pathological subtype is mainly undifferentiated non-keratinising carcinoma; the pathological subtype and distribution of clinical

characteristics might be different in other areas. Second, the biological mechanisms by which many candidate markers that have been included in the DMGN, such as *POLR2M* and *YBX3*, contribute to nasopharyngeal carcinoma metastasis remain unknown, and further investigations into their functions might provide novel targets and treatment strategies. Moreover, our DMGN and nomogram require further validation in prospective studies and multicentre clinical trials; we are proceeding with a prospective, multicentre study (NCT03025854).

To our knowledge, this is the first study to include such a large sample size that has evaluated the global gene-expression profiles in locoregionally advanced nasopharyngeal carcinoma tumours with different prognoses. Our findings show that the DMGN prognostic tool can effectively classify patients with locoregionally advanced nasopharyngeal carcinoma into groups with different risks of distant metastasis, thereby adding prognostic value to the traditional clinico-pathological risk factors used to assess patient prognoses. Moreover, in our study, we show that the DMGN classifier might be a useful predictive tool for identifying patients who would benefit from concurrent chemotherapy. A nomogram comprising DMGN might help clinicians in directing personalised therapeutic regimen selection for patients with locoregionally advanced nasopharyngeal carcinoma.

#### Contributors

JM and NL designed the study. X-RT, NL, Y-QL, S-BL, WJ, FL, Q-MH, X-JY, and YZ collected the data. JM, NL, X-RT, W-XG, Y-QL, Q-MH, L-LT, and Y-PM analysed and interpreted the data. X-RT, JM, and NL wrote the manuscript. Y-QL, XW, JZ, Y-QW, P-PZ, YS, J-PY, JZ, L-ZL, and LL revised the manuscript. X-RT, W-XG, and NL did the statistical analysis. All authors reviewed the manuscript and approved the final version.

#### Declaration of interests

We declare no competing interests.

#### Acknowledgments

We thank Hanqi Yin (Shanghai Biotechnology Corporation, Shanghai, China) for statistical consultation, and Yiducloud (Beijing) Technology Ltd for supporting part of the data extraction and processing. This study was supported by grants from the National Natural Science Foundation of China (81230056, 81402516), the National Science & Technology Pillar Program during the Twelfth Five-year Plan Period (2014BAI09B10), Natural Science Foundation of Guang Dong Province (2017A030312003), the National Key Research and Development Program of China (2016YFC0902000), the Innovation Team Development Plan of the Ministry of Education (IRT_17R110), Health & Medical Collaborative Innovation Project of Guangzhou City, China (201400000001), and the Program of Introducing Talents of Discipline to Universities (B14035).

#### References

1 Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015; **65:** 87–108.

2 Pan JJ, Ng WT, Zong JF, et al. Prognostic nomogram for refining the prognostication of the proposed 8th edition of the AJCC/UICC staging system for nasopharyngeal cancer in the era of intensity-modulated radiotherapy. *Cancer* 2016; **122:** 3307–15.

3 National Comprehensive Cancer Network guidelines. 2017. http://www.nccn.org/professionals/physician_gls/pdf (accessed Aug 23, 2017).

4 Lai SZ, Li WF, Chen L, et al. 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:** 661–68.

5 Hui EP, Leung SF, Au JS, et al. Lung metastasis alone in nasopharyngeal carcinoma: a relatively favorable prognostic group. A study by the Hong Kong Nasopharyngeal Carcinoma Study Group. *Cancer* 2004; **101:** 300–06.

6 Lin JC, Chen KY, Wang WY, et al. Detection of Epstein-Barr virus DNA in the peripheral-blood cells of patients with nasopharyngeal carcinoma: relationship to distant metastasis and survival. *J Clin Oncol* 2001; **19:** 2607–15.

7 Zhou GQ, Tang LL, Mao YP, et al. Baseline serum lactate dehydrogenase levels for patients treated with intensity-modulated radiotherapy for nasopharyngeal carcinoma: a predictor of poor prognosis and subsequent liver metastasis. *Int J Radiat Oncol Biol Phsy* 2012; **82:** e359–65.

8 Liu N, Chen NY, Cui RX, et al. Prognostic value of a microRNA signature in nasopharyngeal carcinoma: a microRNA expression analysis. *Lancet Oncol* 2012; **13:** 633–41.

9 Kim KY, Le Q, Yom SS, et al. Clinical utility of Epstein-Barr virus DNA testing in the treatment of nasopharyngeal carcinoma patients. *Int J Radiat Oncol Biol Phys* 2017; **98:** 996–1001.

10 Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med* 2016; **375:** 717–29.

11 Zhao SG, Chang SL, Spratt DE, et al. Development and validation of a 24-gene predictor of response to postoperative radiotherapy in prostate cancer: a matched, retrospective analysis. *Lancet Oncol* 2016; **17:** 1612–20.

12 Rini B, Goddard A, Knezevic D, et al. A 16-gene assay to predict recurrence after surgery in localised renal cell carcinoma: development and validation studies. *Lancet Oncol* 2015; **16:** 676–85.

13 Scott DW, Chan FC, Hong F, et al. Gene expression-based model using formalin-fixed paraffin-embedded biopsies predicts overall survival in advanced-stage classical Hodgkin lymphoma. *J Clin Oncol* 2013; **31:** 692–700.

14 Shi W, Bastianutto C, Li A, et al. Multiple dysregulated pathways in nasopharyngeal carcinoma revealed by gene expression profiling. *Int J Cancer* 2006; **119:** 2467–75.

15 Zeng Z, Zhou Y, Xiong W, et al. Analysis of gene expression identifies candidate molecular markers in nasopharyngeal carcinoma using microdissection and cDNA microarray. *J Cancer Res Clin Oncol* 2007; **133:** 71–81.

16 Wang S, Li X, Li ZG, et al. Gene expression profile changes and possible molecular subtypes in differentiated-type nonkeratinizing nasopharyngeal carcinoma. *Int J Cancer* 2011; **128:** 753–62.

17 Amin MB, Edge S, Greene F, et al. AJCC cancer staging manual, 8th edn. New York, NY: Springer, 2016.

18 Zhang JX, Song W, Chen ZH, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol* 2013; **14:** 1295–306.

19 Nault JC, De Reyniès A, Villanueva A, et al. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. *Gastroenterology* 2013; **145:** 176–87.

20 Guo P, Luo Y, Mai G, et al. Gene expression profile based classification models of psoriasis. *Genomics* 2014; **103:** 48–55.

21 Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res* 2004; **10:** 7252–59.

22 Xia WX, Zhang HB, Shi JL, et al. 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:** 2152–60.

23 Tang LQ, Li CF, Li J, et al. Establishment and validation of prognostic nomograms for endemic nasopharyngeal carcinoma. *J Natl Cancer Inst* 2016; **108:** pii: djv291.

24 Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; **56:** 337–44.

25 Chang JT, Chan SC, Yen TC, et al. Nasopharyngeal carcinoma staging by (18)F-fluorodeoxyglucose positron emission tomography. *Int J Radiat Oncol Biol Phys* 2005; **62:** 501–07.

26 Tang LQ, Chen QY, Fan W, et al. Prospective study of tailoring whole-body dual-modality [18F]fluorodeoxyglucose positron emission tomography/computed tomography with plasma Epstein-Barr virus DNA for detecting distant metastasis in endemic nasopharyngeal carcinoma at initial staging. *J Clin Oncol* 2013; **31:** 2861–69.

27 Sestak I, Cuzick J, Dowsett M, et al. Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian Breast and Colorectal Cancer Study Group 8 and Arimidex, Tamoxifen Alone or in Combination randomized trials using the PAM50 risk of recurrence score. *J Clin Oncol* 2015; **33:** 916–22.

28 Ramaswamy V, Remke M, Bouffet E, et al. Recurrence patterns across medulloblastoma subgroups: an integrated clinical and molecular analysis. *Lancet Oncol* 2013; **14:** 1200–07.

29 Lin JC, Jan JS, Hsu CY, et al. Phase III study of concurrent chemotherapy versus radiotherapy alone for advanced nasopharyngeal carcinoma: positive effect on overall and progression-free survival. *J Clin Oncol* 2003; **21:** 631–37.

30 Zhang X, Liu S, Hu T, et al. Up-regulated microRNA-143 transcribed by nuclear factor kappa B enhances hepatocarcinoma metastasis by repressing fibronectin expression. *Hepatology* 2009; **50:** 490–99.

31 Bouchet BP, Noordstra I, van Amersfoort M, et al. Mesenchymal cell invasion requires cooperative regulation of persistent microtubule growth by SLAIN2 and CLASP1. *Dev Cell* 2016; **39:** 708–23.

32 Zhang Z, Huang L, Zhao W, et al. Annexin 1 induced by anti-inflammatory drugs binds to NF-κB and inhibits its activation: anticancer effects in vitro and in vivo. *Cancer Res* 2010; **70:** 2379–88.

33 Vacchelli E, Enot DP, Pietrocola F, et al. Impact of pattern recognition receptors on the prognosis of breast cancer patients undergoing adjuvant chemotherapy. *Cancer Res* 2016; **76:** 3122–26.

34 Brownell J, Polyak SJ. Molecular pathways: hepatitis C virus, CXCL10, and the inflammatory road to liver cancer. *Clin Cancer Res* 2013; **19:** 1347–52.

35 Harris DP, Bandyopadhyay S, Maxwell TJ, et al. Tumor necrosis factor (TNF)-α induction of CXCL10 in endothelial cells requires protein arginine methyltransferase 5 (PRMT5)-mediated nuclear factor (NF)-κB p65 methylation. *J Biol Chem* 2014; **289:** 15328–39.

36 Hui EP, Ma BBY, King AD, et al. Hemorrhagic complications in a phase II study of sunitinib in patients of nasopharyngeal carcinoma who has previously received high-dose radiation. *Ann Oncol* 2011; **22:** 1280–87.

37 Chia WK, Teo M, Wang WW, et al. Adoptive T-cell transfer and chemotherapy in the first-line treatment of metastatic and/or locally recurrent nasopharyngeal carcinoma. *Mol Ther* 2014; **22:** 132–39.