Supplementary Online Content

Chen D, Chen G, Jiang W, et al. Association of the collagen signature in the tumor microenvironment with lymph node metastasis in early gastric cancer. *JAMA Surg*. Published online January 30, 2019. doi:10.1001/jamasurg.2018.5249

This supplementary material has been provided by the authors to give readers additional information about their work.

I. eMethods

1. Multiphoton imaging system

The system contained a high-throughput scanning inverted Axiovert 200 microscope (LSM 510 META; Zeiss, Germany) and was equipped with a mode-locked femtosecond titanium (Ti):sapphire laser (110 fs, 76 MHz), tunable from 700 to 980 nm (Mira 900-F; Coherent, America). An acousto-optic modulator was used to control the attenuation of the laser intensity. A Plan-Apochromat 20× objective (Zeiss) was employed for focusing the excitation beam and for collecting the backward signals. The META detector collected the backward multiphoton signals from the tissue sample. The two-channel mode could achieve two-photon excitation fluorescence (TPEF) and second harmonic generation (SHG), which was separated by the dichroic mirror in the detection path. One channel corresponds to a wavelength range of 430 to 708 nm to show the morphologies of tissue components from the TPEF signals, whereas another channel covered the wavelength range from 387 to 409 nm to present the microstructures of tissue components from the SHG signals. The excitation wavelength (λ_{ex}) used in this study was 800 nm.

2. Collagen feature extraction

The extraction procedure for collagen features has previously been reported by our group.[1]

Morphological features

Six morphological features were extracted, namely, the collagen area, length, width, straightness, cross-link density and orientation. The SHG image was first segmented into collagen pixels and background pixels using the Gaussian mixture model method.[2] The binary collagen mask image was then processed using a fiber network extraction algorithm, as described in reference 3, to trace each collagen fiber in the image and to identify cross-link points, which are defined as connecting points between two or more fibers. Moreover, we quantified an orientation index to reflect the collagen alignment based on Fourier-transform spectra.[4]

Texture features

For intensity features, a histogram-based approach was used. The mean, variation, skewness, kurtosis, energy and entropy were calculated from the histogram of the SHG pixel intensity distribution.

We also included 80 gray-level co-occurrence matrix (GLCM)-based texture features from reference 5 and 48 Gabor wavelet transform features from reference 6 in the analysis. The contrast, correlation, energy and homogeneity were calculated from the GLCM with five different displacements of pixels at 1, 2, 3, 4 and 5 and four different directions at 0, 45, 90 and 135 degrees. To calculate the Gabor wavelet transform features, we convolved the SHG image with Gabor filters at five different scales and six different orientations, and the mean and variation of the magnitude of the convolution over the image at each setting were calculated.

2. Least-absolute shrinkage and selection operator (LASSO) binary regression

LASSO binary regression is an effective method for high-dimensional predictors, especially in problems wherein the number of predictors far exceeds the number of observations.[7] The method uses an L1 penalty to shrink the coefficients to zero. The penalty parameter λ, also called the tuning constant, controls the strength of the penalty. If we reduce λ and relax the penalty, then more predictors can enter the model. In this study, LASSO binary regression was performed using R (version 3.4.2)

with the "*glmnet*" package. Five-time cross validations were used to determine the optimal value of λ. Finally, the λ was selected via 1-standard error (SE) criteria.

3. Construction of the decision curve of the nomogram.

Decision curve analysis (DCA) is a novel tool for assessing the potential population impact of adopting a risk prediction instrument into clinical practice, initially introduced by Vickers and Elkin in 2006.[8, 9] The context for DCA is a situation in which individuals' risks for an undesirable outcome will be assessed, and individuals with sufficiently high risk will be recommended for some intervention or treatment. The DCA provides a net benefit, which is calculated by the following formula:

Net benefit = true-positive rate -false-positive rate * [Pt/(1-Pt)]

Pt is the threshold probability at which the expected benefit of treatment is equal to the expected benefit of avoiding treatment. In this study, Pt indicates the threshold probability of LNM. DCA was plotted using R (version 3.4.2) with the "*rmda*" package.

II. eTables

eTable 1. List of collagen features

No.	Feature descriptions
Morphological features	
1-2	Mean and variation of collagen area
3-4	Mean and variation of collagen length
5-6	Mean and variation of collagen width
7-8	Mean and variation of collagen straightness
9-10	Mean and variation of collagen cross-link density
11-12	Mean and variation of collagen orientation
Texture features	
13-18	Mean, variance, skewness, kurtosis, energy, entropy
19-98	Contrast, correlation, energy and homogeneity from the gray-level co-occurrence matrix (GLCM) given five different pixel distances with four different directions
99-146	Mean and variance of the convolution over the image with the Gabor filter at four scales with six orientations

eTable 2. Comparison of clinical-pathological factors in the primary and validation cohorts

Characteristics	Primary cohort (n=232)	Validation cohort (n=143)	OR (95% CI)	*P* Value
Age, Mean (SD), years	58.66 (10.77)	57.69 (11.66)	0.992 (0.974, 1.011)	0.41
Gender, No. (%)				
Male	164 (70.7)	92 (64.3)	Reference	0.20
Female	68 (29.3)	51 (35.7)	1.377 (0.858, 2.084)	
Primary site, No. (%)				
Upper	31 (13.3)	27 (18.9)	Reference	0.31
Middle	53 (22.8)	34 (23.8)	0.737 (0.376, 1.442)	
Low	148 (63.9)	82 (57.3)	0.636 (0.355, 1.139)	
Size, No. (%)				
≤2 cm	141 (60.8)	72 (50.3)	Reference	0.05
>2 cm	91 (39.2)	71 (49.7)	1.528 (1.003, 2.327)	
Macroscopic, No. (%)				
Elevated	9 (3.9)	2 (1.4)	Reference	0.05
Flat	140 (60.3)	103 (72.0)	3.311 (0.700, 15.648)	
Depressed	83 (35.8)	38 (26.6)	2.060 (0.425, 9.998)	
Differentiation, No. (%)				
Differentiated	159 (68.5)	82 (57.3)	Reference	0.05
Undifferentiated	73 (31.5)	59 (42.7)	1.530 (0.992, 2.359)	
Lymphovascular infiltration, No. (%)				
Yes	19 (8.2)	13 (9.1)	Reference	0.76
No	213 (91.8)	130 (90.9)	1.121 (0.536, 2.346)	
Depth, No. (%)				
Mucosa	105 (45.3)	61 (42.7)	Reference	0.14
Submucosa	127 (54.7)	82 (57.3)	1.368 (0.899, 2.081)	

Abbreviations: SD, standard deviation.

eTable 3. Multicollinearity assessment in the prediction model based on the collagen signature

Predictors	Collinearity statistics	
	Tolerance	Variance inflation factor
Tumor differentiation	0.973	1.027
Depth of invasion	0.924	1.083
Collagen signature	0.904	1.106

eTable 4. Effect modification assessment between the collagen signature and the depth of tumor invasion

Variables	β	OR (95% CI)	P Value
Collagen signature	1.027	2.793 (1.185-6.584)	0.02
Differentiation (Undifferentiated vs. Differentiated)	1.647	5.193 (1.412-19.095)	0.01
Depth (Submucosa vs. Mucosa)	3.048	21.082 (2.777-160.039)	0.003
Collagen signature*Depth	0.948	2.581 (0.820-8.123)	0.11

eTable 5. Effect modification assessment between the collagen signature and tumor differentiation

Variables	β	OR (95% CI)	*P* value
Collagen signature	1.637	5.137 (2.484-10.627)	<0.001
Differentiation (Undifferentiated vs. Differentiated)	1.628	5.094 (0.807-32.161)	0.08
Depth (Submucosa vs. Mucosa)	1.920	6.822 (1.635-28.471)	0.008
Collagen signature*Differentiation	0.091	1.095 (0.347-3.452)	0.88

eTable 6. The model performance in estimating the risk of LNM

Variable	Value (95% CI)		
	Primary cohort	Validation cohort	All patients
Cutoff risk probability	0.301	0.301	0.301
Sensitivity, %	86.8 (73.6-97.4)	90.0 (76.7-96.7)	87.3 (82.4-97.1)
Specificity, %	93.3 (89.7-96.4)	90.3 (84.1-94.7)	92.1 (88.9-95.3)
Accuracy, %	92.2 (89.9-96.1)	90.2 (84.3-95.5)	91.2 (88.8-95.1)
Positive predictive value, %	71.7 (63.9-85.2)	71.1 (59.1-84.4)	72.1 (64.1-81.4)
Negative predictive value, %	97.3 (95.7-99.7)	97.1 (93.9-99.5)	96.9 (95.9-99.3)

eTable 7. Univariate and multivariate logistic regression of LNM without the collagen signature

Characteristics	Univariate logistic regression		Multivariate logistic regression	
	OR (95% CI)	*P* Value	OR (95% CI)	*P* Value
Age	0.991 (0.960-1.024)	0.60	NA	NA
Gender				
Male	Reference	1		
Female	1.314 (0.627-2.753)	0.47	NA	NA
Primary site				
Upper	Reference	1		
Middle	1.191 (0.276-5.145)	0.81	NA	NA
Low	2.275 (0.647-8.002)	0.20	NA	NA
Size				
≤2 cm	Reference	1		
>2 cm	3.249 (1.578-6.693)	0.001 [a]	NA	NA
Macroscopic				
Elevated	Reference	1		
Flat	0.314 (0.072-1.363)	0.12	NA	NA
Depressed	0.478 (0.108-2.118)	0.33	NA	NA
Differentiation				
Differentiated	Reference	1	Reference	1
Undifferentiated	2.956 (1.452-6.017)	0.002	2.576 (1.167-5.685)	0.02
Lymphovascular invasion				
No	Reference	1	Reference	1
Yes	7.341 (2.743-19.646)	<0.001	3.333 (1.145-9.703)	0.03
Depth				
Mucosa	Reference	1	Reference	1
Submucosa	9.231 (3.155-27.010)	<0.001	9.323 (3.305-29.793)	0.008

eTable 8. Multicollinearity assessment in the traditional model

Predictors	Collinearity statistics	
	Tolerance	Variance Inflation Factor
Tumor differentiation	0.970	1.031
Depth of invasion	0.942	1.061
Lymphovascular infiltration	0.920	1.087

eAppendix 1. Collagen signature calculation formula

Collagen signature=0.4065624

+82.7081949* Variation of straightness

+69.7400305* Variation of cross-link density

+0.2508164* Skewness

-3.7531736* GLCM_contrast_0°_2 pixel

-3.7486890*Texture.GLCM_45°_5 pixel

-0.7642914*Texture.GLCM_90°_4 pixel

eAppendix 2. Sample size calculation for logistic regression

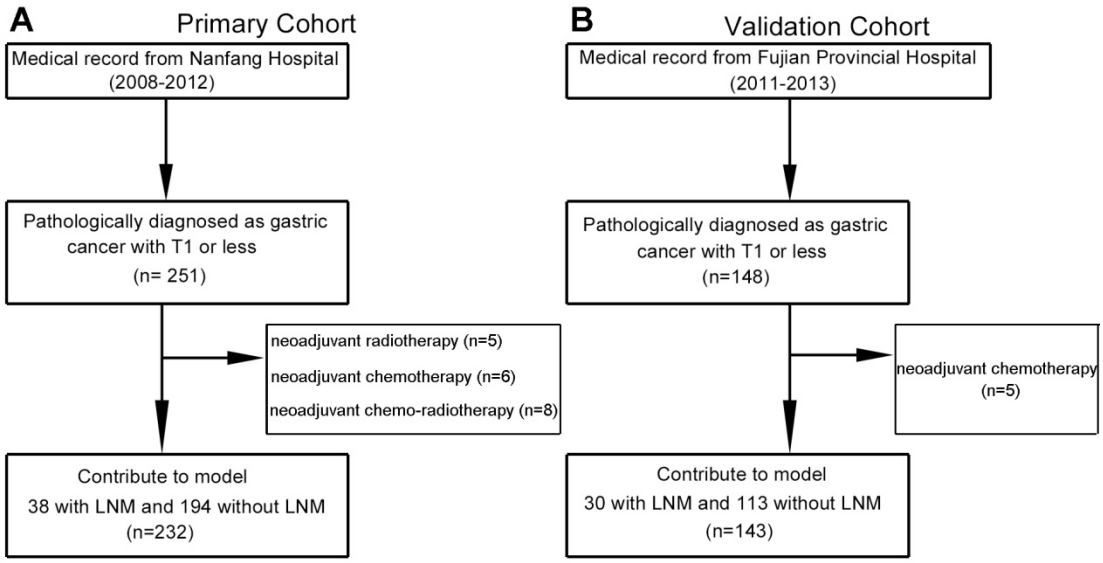
There were two methods used to calculate the sample size. One was as the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement, for which at least 10 events per variable (EPV) were needed.[10] However, as the statement noted, the requirement of 10 EPV was based on two empirical investigations.[11, 12] Some researchers suspected that 10 EPV was too lenient[13] or too strict.[14] Only for a planned prospective prediction model development study would the sample size be predetermined on statistical grounds.[10] The other was based on the variance inflation factor (VIF), which does not explicitly require knowledge of the number of variables in the regression model.[15] The formula was as follows:

$$N= \frac{\{Z_{1-\alpha/2}[p(1-p)/B]^{1/2}+Z_{1-\beta}[p_0(1-p_0)+p_1(1-p_1)(1-B)/B]^{1/2}\}^2}{(p_0-p_1)^2(1-B)} \text{ *VIF}$$

We could take our research as an example to illustrate the formula. The *p* represents the incidence of LNM in EGC, and B represents when the tumor invaded the submucosa in the proportion of patients in all EGC. The p_0 and p_1 represent the incidence of LNM when the tumor invaded the mucosa and submucosa, respectively. From our data, the incidence of LNM was 3.8% (4/105) and 26.8% (34/127) with mucosa and submucosa invasion, respectively. The incidence of LNM was 16.4% (38/232) and the proportion of submucosa invasion in EGC was 54.7% (127/232). In our multicollinearity assessment, the VIF of the depth of tumor invasion was 1.083. We hypothesized that the type I error α was 0.05 and the type II error β was 0.1, and the minimum sample size was 110 according to our data. Our sample size might be not adequate for the TRIPOD method but was enough for the VIF method. Thus, we hope that this limitation will be solved in our upcoming clinical trial. Fewer studies have been reported for sample size calculation in the validation cohort. Lei et al.[16] revealed that the ratio between primary and validation cohort was 7:3. In our study, the validation cohort contained 143 patients, which

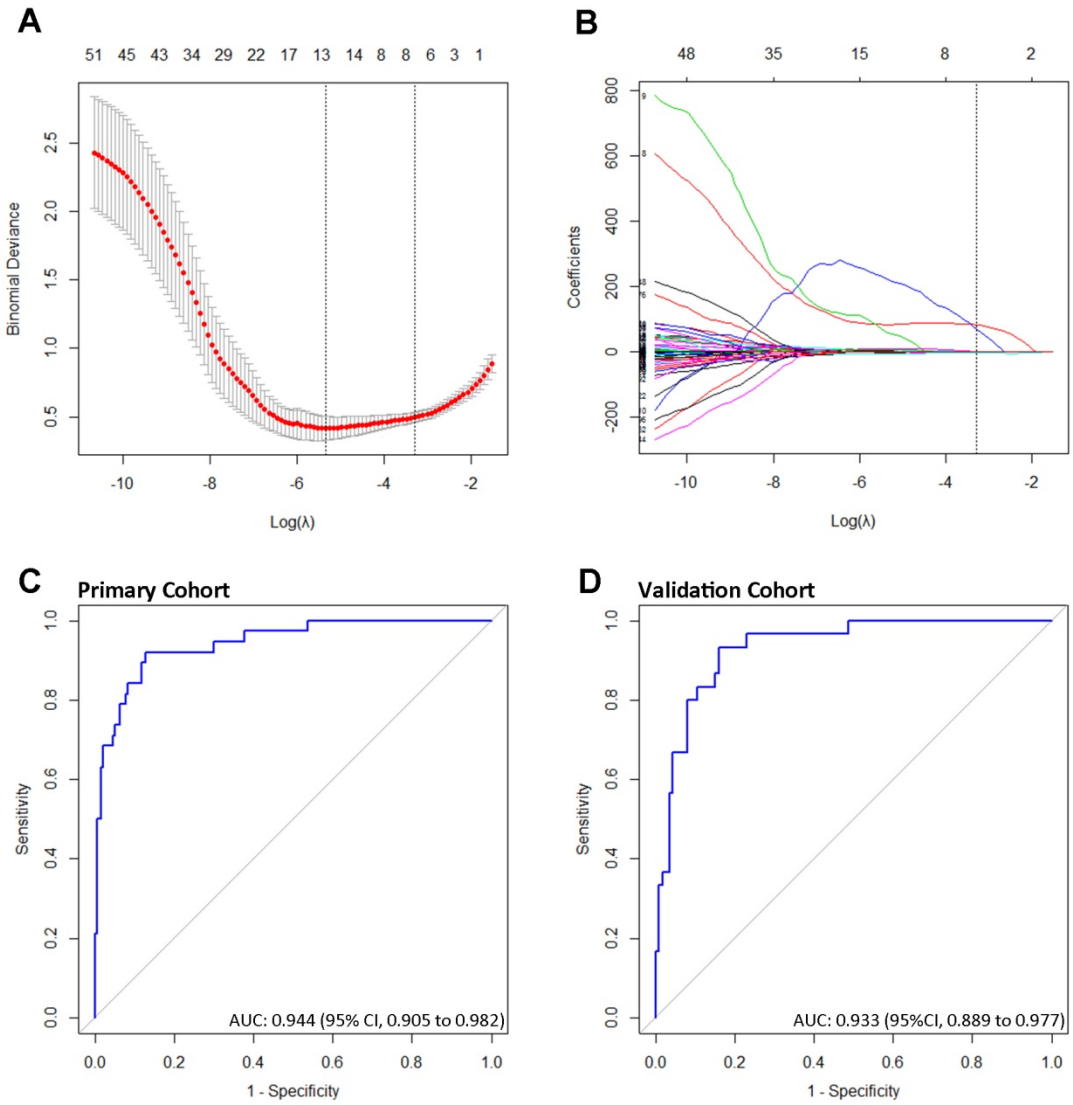
was enough to validate the prediction model.

V. eFigures



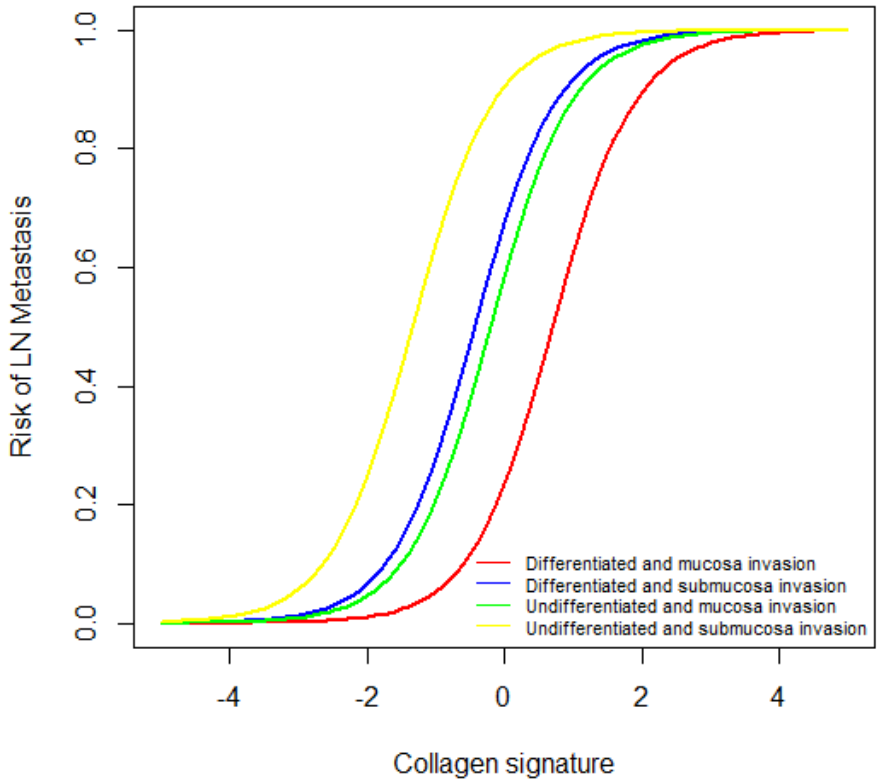
eFigure 1. Recruitment pathways for patients in the primary and validation cohorts

(A) In the primary cohort, in total, 251 consecutive patients were diagnosed with T1 stage gastric cancer January 2008 to December 2012. Nineteen patients were excluded because 5 had neoadjuvant radiotherapy, 6 had neoadjuvant chemotherapy and 8 had neoadjuvant chemo-radiotherapy. Thus, 232 consecutive EGC patients were enrolled. (B) For external validation, 148 consecutive patients were diagnosed with T1 stage gastric cancer from January 2011 to December 2013 at Fujian Provincial Hospital and meeting the same inclusion criteria of the primary cohort, and 5 were excluded because of neoadjuvant chemotherapy. Therefore, an additional 143 consecutive EGC patients were enrolled in the validation cohort.

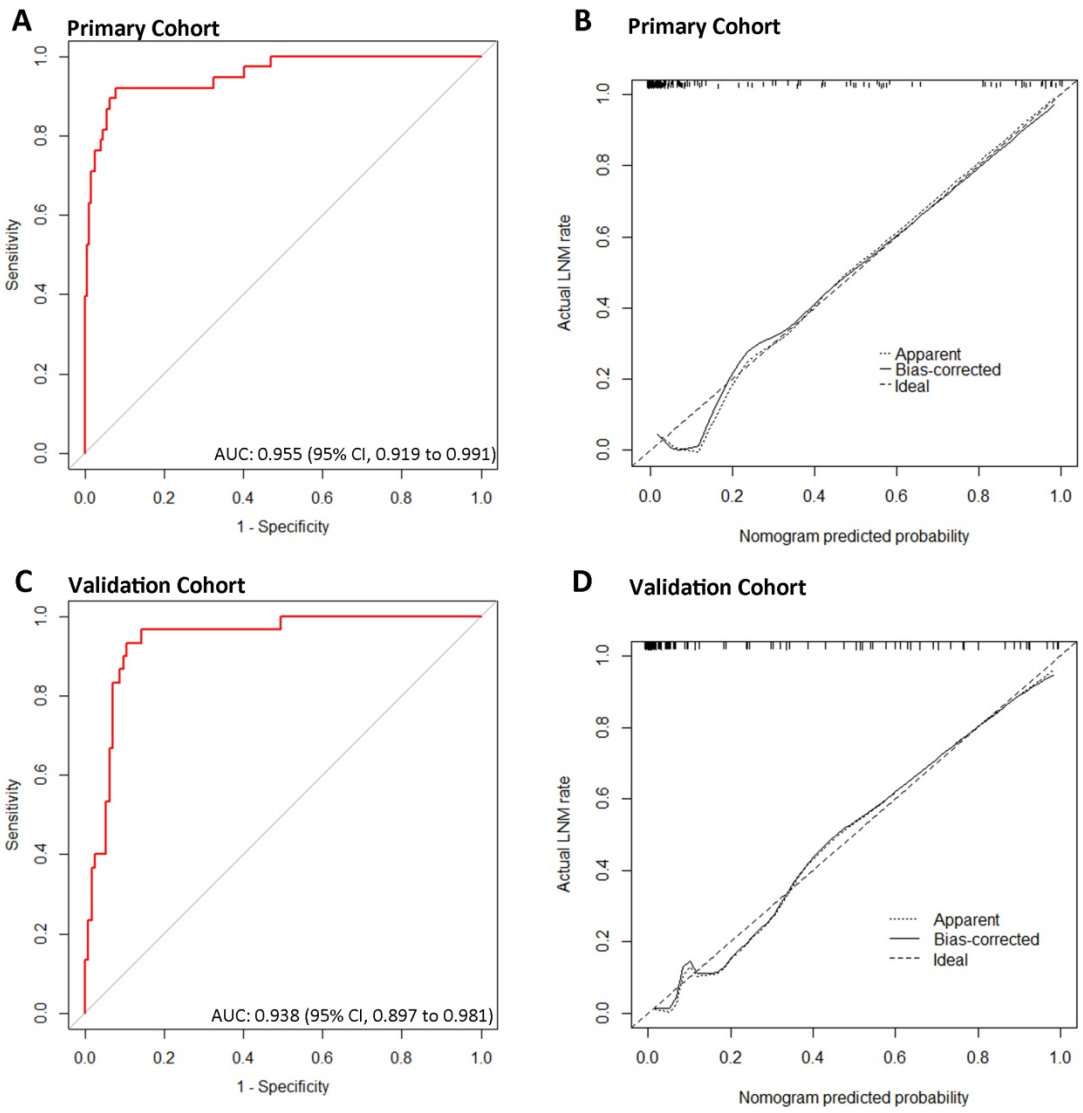


eFigure 2. Collagen feature selection using LASSO binary logistic regression and discrimination of the collagen signature

(A) Tuning parameter (λ) selection in the LASSO method using five-fold cross-validation via minimum criteria. Solid vertical lines represent the binomial deviance ± standard error (SE). Dotted vertical lines were drawn at the optimal values by the minimum criteria and the 1-SE criteria. A λ value of 0.0371, with log (λ) = -3.294 was chosen by five-fold cross-validation via 1-SE criteria. (B) LASSO coefficient profiles of the 146 collagen features. A dotted vertical line was drawn at the value [log (λ) = -3.294] using five-fold cross-validation, wherein the optimal λ resulted in 6 nonzero coefficients. (C) The ROC curve of the collagen signature for LNM in the primary cohort. (D) The ROC curve of the collagen signature for LNM in the validation cohort.

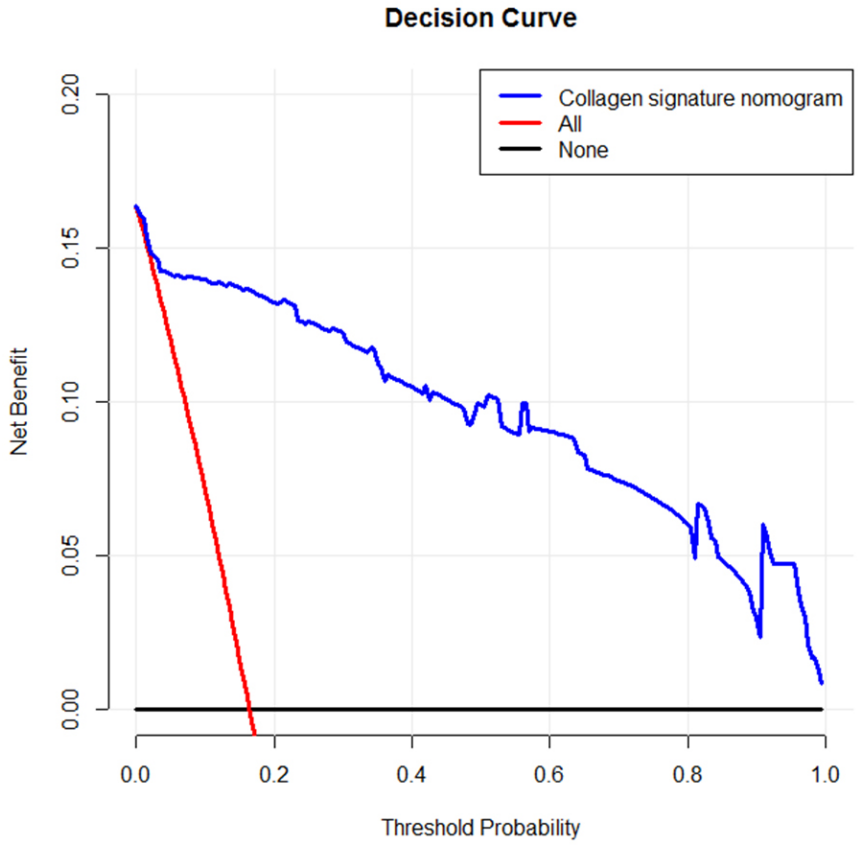


eFigure 3. Association between the collagen signature and the risk of LNM with different types of tumor differentiation and depths of tumor invasion



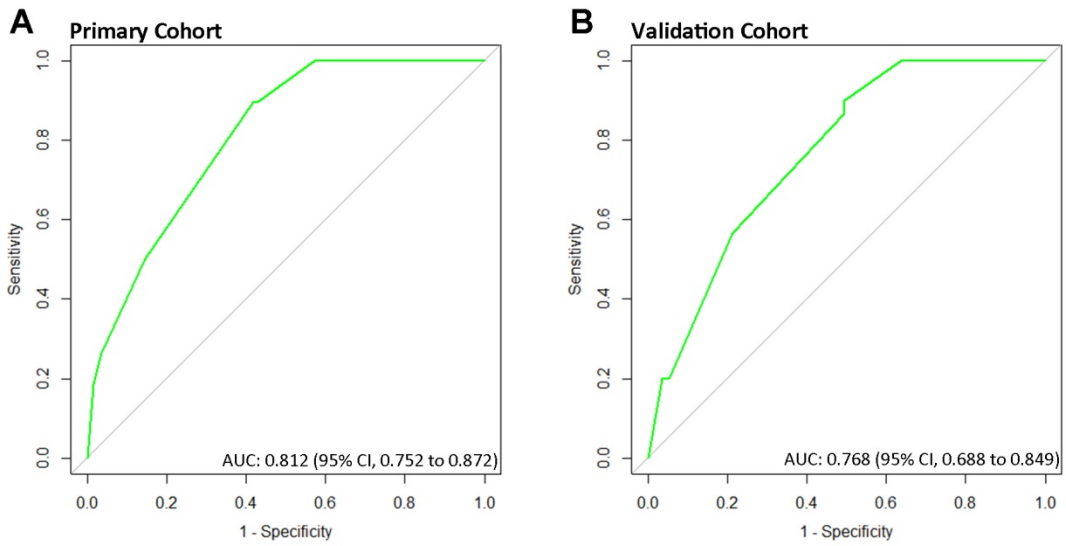
eFigure 4. The performance of the prediction model based on the collagen signature

(A) The ROC curve of the nomogram for LNM in the primary cohort. (B) Calibration curve of the nomogram in the primary cohort, with a mean absolute error of 0.025. (C) The ROC curve of the nomogram for LNM in the validation cohort. (D) Calibration curve of the nomogram in the validation cohort, with a mean absolute error of 0.024. In the calibration curve, the y-axis represents the actual LNM rate, and the x-axis represents the predicted LNM probability. The calibration curve describes the calibration of each model in terms of the agreement between the predicted probability of LNM and the observed outcome of LNM. The diagonal dotted line represents a perfect prediction by an ideal model. The other dotted line represents the performance of the nomogram. The solid line represents the bias-corrected performance of the nomogram.



eFigure 5. Decision curve analysis of the nomogram

The y-axis represents the net benefit, and the blue line represents the nomogram. The red line and black line represent the assumption of all patients with and without LNM, respectively.



eFigure 6. The ROC curve to predict LNM in the primary and validation cohorts without the collagen signature

(A) The ROC curve of the nomogram for LNM in the primary cohort without the collagen signature with an AUC of 0.812 (95% CI, 0.752-0.872). (B) The ROC curve of the nomogram for LNM in the validation cohort without the collagen signature with an AUC of 0.768 (95% CI, 0.688-0.849).

References:

1. Xu S, Kang CH, Gou X, et al. Quantification of liver fibrosis via second harmonic imaging of the Glisson's capsule from liver surface. *J Biophotonics*. 2016;9(4): 351-363.
2. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Statistical Society Series B*. 1977;39(1): 1-38.
3. Stein AM, Vader DA, Jawerth LM, Weitz DA, Sander LM. An algorithm for extracting the network geometry of three-dimensional collagen gels. *J Microsc*. 2008; 232(3): 463-475.
4. Frisch KE, Duenwald-Kuehl SE, Kobayashi H, Chamberlain CS, Lakes RS, Vanderby J R. Quantification of collagen organization using fractal dimensions and Fourier transforms. *Acta Histochemica*. 2012; 114(2): 140-44.
5. Haralick RM, Shanmugam K. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*. 1973; 6(3): 610-621.
6. Daugman, JG. Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on acoustics, speech, and signal processing*. 1988; 36(7): 1169-1179.
7. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. 2007;26(30): 5512-5528.
8. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6): 565-574.
9. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol*. 2016;34(21): 2534-2540.
10. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1): 55-63.
11. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol.* 1996;49(12):1373-1379.
12. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol.* 1995;48(12):1503-1510.
13. Wynants L, Bouwmeester W, Moons KG, et al. A simulation study of sample size demonstrated the importance of the number of events per variable to develop prediction models in clustered data. *J Clin Epidemiol.* 2015;68(12):1406-1414.
14. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol.* 2007;165(6):710-718.
15. Hsieh FY. Sample size tables for logistic regression. *Stat Med.* 1989;8(7):795-802.
16. Lei Z, Li J, Wu D, et al. Nomogram for Preoperative Estimation of Microvascular Invasion Risk in Hepatitis B Virus-Related Hepatocellular Carcinoma Within the Milan Criteria. *JAMA Surg*. 2016;151(4):356-363.