

Nomogram model combining macro and micro tumor-associated collagen signatures obtained from multiphoton images to predict the histologic grade in breast cancer: supplement

GANGQIN XI,^{1,6} JIAJIA HE,^{1,6} DEYONG KANG,^{2,6} SHUOYU XU,³ WENHUI GUO,⁴ FANGMENG FU,⁴ YULAN LIU,¹ LIQIN ZHENG,¹ LIDA QIU,^{1,5} LIANHUANG LI,^{1,7} CHUAN WANG,^{4,8} AND JIANXIN CHEN^{1,9} 

¹Key Laboratory of OptoElectronic Science and Technology for Medicine of Ministry of Education, Fujian Provincial Key Laboratory of Photonics Technology, Fujian Normal University, Fuzhou 350007, China

²Department of Pathology, Fujian Medical University Union Hospital, Fuzhou 350001, China

³Department of General Surgery, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China

⁴Department of Breast Surgery, Department of General Surgery, Fujian Medical University Union Hospital, Fuzhou 350001, China

⁵College of Physics and Electronic Information Engineering, Minjiang University, Fuzhou 350108, China

⁶These authors contributed equally to this work.

⁷lhli@fjnu.edu.cn

⁸dr_chuanwang@fjmu.edu.cn

⁹chenjianxin@fjnu.edu.cn

This supplement published with Optica Publishing Group on 27 September 2021 by The Authors under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) in the format provided by the authors and unedited. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Supplement DOI: <https://doi.org/10.6084/m9.figshare.16628965>

Parent Article DOI: <https://doi.org/10.1364/BOE.433281>

Supplementary materials for

A nomogram model combining macro and micro tumor-associated collagen signatures obtained from multiphoton images to predict histologic grade in breast cancer

TACS-score and TCMF-score calculation formulas

Using the ridge regression analysis, we obtain a TACS-score for each patient based on the combined TACS1-8:

$$\text{TACS-score} = 0.6021201 - (0.3301427 * \text{TACS1}) + (1.1208062 * \text{TACS2}) - (2.1103586 * \text{TACS3}) + (1.3892094 * \text{ACS4}) - (1.6522228 * \text{TACS5}) + (2.4391089 * \text{TACS6}) + (0.1563863 * \text{TACS7}) - (1.2751252 * \text{TACS8})$$

Using the LASSO logistic regression analysis (Fig. S2), we obtain a TCMF-score for each patient based on the 7 selected microscopic features:

$$\text{TCMF-score} = 1.7237065 - 0.4824760 * \text{Area} - 0.1013971 * \text{Width} + 0.4286551 * \text{Histogram energy} + 0.2521051 * \text{GLCM_energy_0^\circ_1 pixel} + 0.2433832 * \text{Gabor_variance_30^\circ_1 scale} + 0.5310991 * \text{Gabor_variance_120^\circ_1 scale} + 0.2019598 * \text{Gabor_variance_90^\circ_3 scale}$$

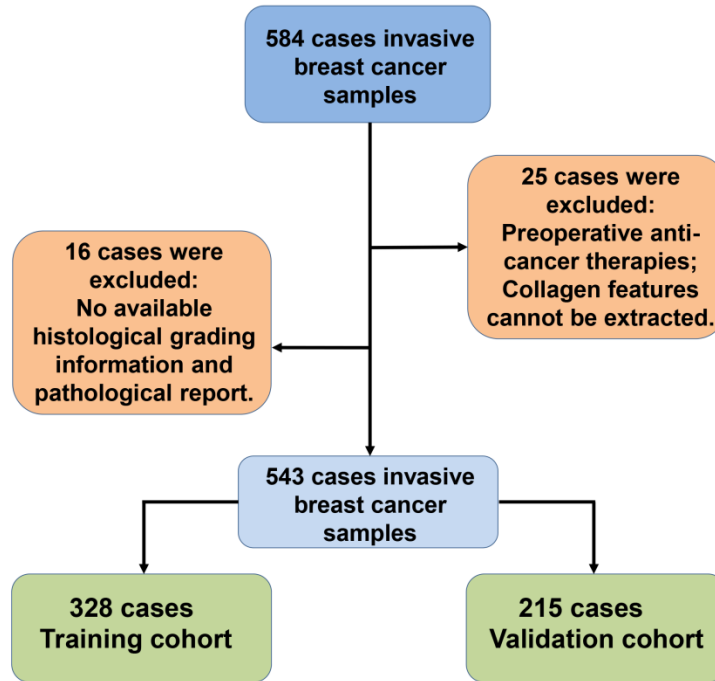


Fig. S1. A flowchart of patient selection.

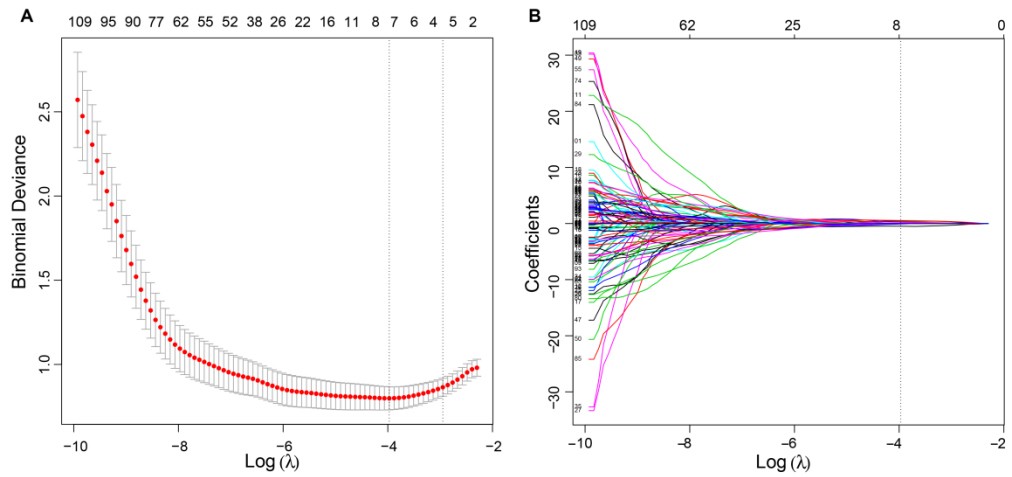


Fig. S2. LASSO logistic regression analysis. (A) A plot showing the relationship between the binomial deviance and $\log(\lambda)$. The left dotted vertical line was at the optimal lambda value point by using the minimum criteria, and the right line was at the optimal lambda value point by using one standard error of the minimum criteria (the 1-SE criteria). (B) LASSO coefficient profiles of the 142 features. A dotted vertical line was drawn at the value selected using ten-fold cross-validation, where the optimal lambda results in 7 nonzero coefficients.

Table S1. Characteristics of patients with breast cancers in the training and validation cohorts.

Characteristics	Training cohort (n=328)	Validation cohort (n=215)	P
Age			0.603
≤50	192 (58.5%)	121 (56.3%)	
>50	136 (41.5%)	94 (43.7%)	
Molecular subtype			0.325
Luminal A	73 (22.3%)	35 (16.3%)	
Luminal B	133 (40.5%)	91 (42.3%)	
HER2-enriched	70 (21.3%)	47 (21.9%)	
Triple-negative	52 (15.9%)	42 (19.5%)	
Tumor size			0.905
≤2cm	131 (39.9%)	87 (40.5%)	
2-5cm	174 (53.0%)	111 (51.6%)	
>5cm	23 (7.0%)	17 (7.9%)	
Nodal status			0.086
0	171 (52.1%)	99 (46.0%)	
1-3	80 (24.4%)	47 (21.9%)	
≥4	77 (23.5%)	69 (32.1%)	
Histological grade			0.467
Grade 1	63 (19.2%)	36 (16.7%)	
Grade 2/3	265 (80.8%)	179 (83.3%)	

Table S2 Univariate and multivariate logistic regression analyses of the association of variables with pathologic grades in the validation cohort.

Variable	Univariate analysis			Multivariate analysis				
	OR	(95%CI)	P Value	OR	(95%CI)	P Value		
Age								
≤50	Reference							
>50	1.106	0.536	2.284	0.785	0.901	0.343	2.367	0.833
Molecular subtype								
Luminal A	Reference							
Luminal B	3.794	1.482	9.718	0.005	2.921	0.833	10.244	0.094
HER2-enriched	2.543	0.906	7.141	0.076	3.722	0.918	15.081	0.066
Triple-negative	3.861	1.203	12.388	0.023	5.715	1.215	26.872	0.027
Tumor size								
≤2cm	Reference							
2-5cm	1.549	0.738	3.251	0.247	1.472	0.541	4.008	0.449
≥5cm	1.957	0.410	9.348	0.400	0.976	0.102	9.297	0.983
Nodal status								
0	Reference							
1-3	1.313	0.533	3.230	0.554	1.843	0.551	6.165	0.321
≥4	2.385	0.952	5.972	0.064	2.983	0.804	11.063	0.102
TACS-score	2.719	1.621	4.561	1.50E-04	3.555	1.824	6.926	1.94E-04
TCMF-score	4.140	2.481	6.909	5.39E-08	5.028	2.711	9.324	2.96E-07

Table S3 Univariate logistic regression analysis on the association of seven features with histologic grades in the training cohort.

Variable	Univariate analysis			
	OR	(95%CI)		P Value
Area	0.515	0.411	0.644	6.62E-09
Width	0.621	0.506	0.763	5.39E-06
Histogram energy	2.204	1.645	2.954	1.22E-07
GLCM_ energy _0°_1 pixel	1.840	1.390	2.436	2.06E-05
Gabor_ variance_ 30°_1 scale	1.657	1.328	2.068	7.87E-06
Gabor_ variance_ 120°_1 scale	1.684	1.348	2.105	4.63E-06
Gabor_ variance_ 90°_3 scale	1.370	1.100	1.706	4.98E-03