STROBE statement

	Item No.	Recommendation
Title and abstract	1	iCEMIGE: Integration of CEll-morphometrics, MIcrobiome, and GEne biomarker signatures for risk stratification in breast cancers
		We found that iCEMIGE score had an independent prognostic value for OS and PFS over clinical factors and PAM50-based molecular subtype. Importantly, iCEMIGE score significantly increased the power for predicting OS and PFS compared to CMPS, GEPS, or MAPS alone. Our study demonstrates a novel and generic AI framework for multimodal data integration towards improving prognosis risk stratification of BC patients, which can be extended to other types of cancers.
Introduction		
Background/rationale	2	The biomarkers to stratify individual risk are critical to precision therapies.
Objectives	3	We aimed to investigate whether iCEMIGE (**i**ntegration of **CE**ll-morphometrics, **MI**crobiome, and **GE**ne biomarker signatures) improves risk stratification of breast cancer (BC) patients
Methods		
Study Design	4	TCGA data was used for this study. The patient diagnostic tissue histology slides were downloaded from The Cancer Genome Atlas (TCGA) breast cancer (TCGA-BRCA) cohort. TCGA-BRCA microbiome, transcriptome, and clinical data, including PAM50-based molecular subtypes, were downloaded from the cBioPortal (https://www.cbioportal.org/).
Setting	5	N/A
Participants	6	(a) All patients from the TCGA-BRCA public cohort with diagnostic slides, microbiome, gene expression, and clinical data available.
		(b) N/A
Variables	7	Overall survival, progression free survival, age, molecular subtype, diagnostic slides, gene expression, tumor microbiome
Data sources/measurement	8	All data are downloaded from TCGA-BRCA cohort and cBioPortal
Bias	9	All patients with diagnostic slides, microbiome, gene expression, and clinical data available were included in this study.

Study size	10	All patients with diagnostic slides, microbiome, gene expression, and clinical data available are included in this study.
Quantitative variables	11	Data was downloaded from TCGA. None of any additional modifications were made to the downloaded data during our analyses.
Statistical methods	12	(a) All patients were then divided into three groups (Good: bottom third, Intermediate: middle third, and Poor: top third) based on CMPS or iCEMIGE. The multivariate Cox regression was used to assess the independent prognostic impact of CMPS and iCEMIGE by adjusting for the clinical factors (age, stage, ER, and PR status) and PAM50 molecular subtype. All statistical analyses were performed through either R (version 4.0.2, https://www.r-project.org/) or SPSS 24.0 (IBM, NY, USA). Graphic visualizations were generated using R (ggplot2 package, Version 3.3.3; ggpubr package, Version 0.4.0) or SPSS. The statistical significance was defined as p<0.05 (two-tails).
		(b) The Kaplan-Meier log-rank test was used in each subgroup, and multivariate Cox regression method was used to assess independent effects.
		(c) Patients with a missing value of any variables were excluded in multivariate analysis.
		(d) The area under the ROC curve and C-Index were used to assess the predictive values of different models.
Results		
Participants	13*	(a) All patients with diagnostic slides, microbiome, gene expression, and clinical data available were included in this study.
		(b) Patients with missing value of any variables were excluded in multivariate analysis.
		(c) N/A
Descriptive data	14*	(a) diagnostic slides, microbiome, gene expression, and clinical data
		(b) Number of patients was indicated in each figure
		(d) Clinical data was downloaded from cBioPortal. None of any additional modifications were made to the downloaded data during our analyses.
Outcome data	15*	(a) We analyzed clinical data that was downloaded from cBioPortal.
		(b) We analyzed clinical data that was downloaded from cBioPortal.
		(c) We analyzed clinical data that was downloaded from cBioPortal.
Main results	16	(a) 95% CI was reported.

		(b) The cut-points were provided in a supplementary table.
		(c) N/A
Other analyses	17	The area under the ROC curve and C-Index were used to assess the predictive values of different models.
Discussion		
Key results	18	Key results were summarized
Strengths and Limitations	19	Strengths and limitations were discussed.
Interpretation	20	A cautious overall interpretation of results was stated.
Generalizability	21	The generalizability (external validity) of the study results was described.
Other information		
Funding	22	The funding information for supporting this study was provided