

Supporting Materials

Supporting Result 1. Comparison with previous models

Supporting Result 2. Interpretability of transformer model

Supporting Result 3. Tumor region delineation

Supporting Result 4. Mechanistic coherence of modules

Table S1. Comparison of the AUCs among Different Models in the Training and Validation Cohort

Table S2. MR scanning parameters across different institutions

Table S3. Comparison of model performance across different models in three cohorts

Table S4. Top 20 enriched pathways across transcriptomic and proteomic GSVA scores

Table S5. Comparison of model performance across this study and other studies in the external validation cohort.

Table S6. Hub gene validation for plum1 and saddlebrown WGCNA modules.

Table S7. Decision curve analysis net benefit of the stacking model and comparators across different threshold probabilities.

Table S8. Comprehensive functional enrichment analysis of pulm1 and saddlebrown WGCNA modules.

Supplementary Fig 1. Additional model-comparison and clinical-utility analyses.

Supplementary Fig 2. Additional multi-omics concordance and WGCNA hub-gene analyses.

Supporting Result 1: Comparison with previous models

In addition to conventional machine learning approaches, we reimplemented three representative published deep learning models for ALN prediction on our dataset under identical preprocessing and evaluation conditions: the CNN-based multiparametric MRI model from Wang et al. (2022), the CNN-RNN kinetic heterogeneity model from Gao et al. (2024), and the CNN-based model from Chen et al. (2023). In the external validation cohort, these models achieved AUCs of 0.779 (95% CI, 0.750 to 0.808), 0.795 (95% CI, 0.767 to 0.823), and 0.768 (95% CI, 0.738 to 0.798), respectively. Our stacking model (AUC=0.830) significantly outperformed all three published approaches (DeLong test $p < 0.001$ for all comparisons), as well as all conventional machine learning models, confirming that the transformer-based multi-modal integration provides consistent performance advantages over prior CNN-based architectures in this prediction task (Supplementary Table S3).

Supporting Result 2. Interpretability of transformer model

To quantitatively evaluate the correspondence between model attention and clinically relevant imaging regions, two radiologists with over ten years of breast imaging experience independently annotated tumor margins, spiculation extent, and peritumoral stromal reaction zones on MRI and MMG images in a randomly selected subset of 120 cases from the external validation cohort (60 ALN-positive, 60 ALN-negative), without knowledge of the model predictions or ALN outcomes. Grad-CAM activation maps were thresholded at the top 30% of activation intensity to generate binary high-activation masks, and the Dice similarity coefficient between these masks and the

radiologist-annotated regions was computed. The mean Dice coefficient was 0.61 (SD 0.12) for MRI and 0.57 (SD 0.14) for MMG. Inter-reader agreement for the annotations was high (ICC=0.83; 95% CI, 0.77 to 0.88). Spatial agreement was significantly greater in ALN-positive cases (mean Dice 0.65) than in ALN-negative cases (mean Dice 0.53; Mann-Whitney U test $p=0.007$), indicating that the model's attention was most concentrated in clinically actionable regions in cases with confirmed nodal involvement. These quantitative results substantiate the correspondence between model attention patterns and expert radiological assessment.

Supporting Result 3. Tumor region delineation

Tumor regions were manually delineated on both MRI and MMG images by two board-certified radiologists each with over 10 years of breast imaging experience, using 3D Slicer software (version 4.10.2). Both readers were fully blinded to all clinical, histopathological, and outcome data, including pathological ALN status, throughout the entire annotation process, and were not informed of any patient outcomes until after all delineations were finalized. To quantify the impact of manual annotation on model generalizability, we additionally evaluated the stacking model using automatically generated tumor ROIs derived from a pretrained nnU-Net breast tumor segmentation pipeline without manual correction. In the external validation cohort, the model achieved an AUC of 0.811 (95% CI, 0.785 to 0.837) with automated ROIs versus 0.830 with manually annotated ROIs (DeLong test $p=0.04$), a difference of 0.019 AUC points. This modest reduction indicates that the model retains strong predictive performance with automated segmentation inputs and is not

critically dependent on expert manual delineation, supporting the feasibility of an automated deployment pipeline.

Supporting Result 4. Mechanistic coherence of modules

Clustering identified two modules, plum1 and saddlebrown, with statistically significant correlations with the stacking model output score ($r=0.22$ and 0.24 , respectively; $P<0.01$). Although these module-level correlations are modest in magnitude, which is consistent with typical WGCNA module-trait relationships in heterogeneous clinical cohorts, they correspond to medium effect sizes when comparing module eigengene scores between the top and bottom tertiles of the stacking model output distribution (Cohen's $d=0.53$ for plum1, 95% CI 0.27 to 0.79 ; Cohen's $d=-0.48$ for saddlebrown, 95% CI -0.74 to -0.22 ; $p<0.001$ for both). To further validate the mechanistic coherence of these modules, we identified the top five hub genes within each module by intramodular connectivity (kME score) and assessed their individual associations with the stacking model score. Hub genes of the plum1 module (KIF14, MCM6, PCNA, RFC4, and RAD51AP1) showed individual Spearman correlations with the model score of 0.33 to 0.41 (all $FDR<0.05$), and their functional annotations as components of the DNA replication and repair machinery are mechanistically consistent with tumor proliferative activity and ALN metastatic potential. Hub genes of the saddlebrown module (ITGB4, LAMC2, CD44, VCL, and FN1) showed Spearman correlations of -0.30 to -0.38 with the model score (all $FDR<0.05$), consistent with their established roles in integrin-mediated adhesion and cytoskeletal regulation during metastatic dissemination (Supplementary Table S6).

Table S1. Comparison of the AUCs among Different Models in the Training and Entire Validation Cohort

Model	Training Cohort (n=658)		Validation Cohorts (n=1447)	
	AUC (95% CI)	P-value	AUC (95% CI)	P-value
Stacking model	0.939 (0.921, 0.958)	reference	0.840 (0.818, 0.861)	reference
MRI model	0.901 (0.877, 0.925)	.014*	0.792 (0.766, 0.817)	.005*
MMG model	0.833 (0.802, 0.864)	<.001*	0.748 (0.722, 0.773)	<.001*
AUS	0.709 (0.675, 0.743)	<.001*	0.701 (0.677, 0.725)	<.001*

Note. - the p-value represented the DeLong test on AUCs between models.

AUC = the area under the curve; CI = confident interval. * Indicates statistical significance; *p* value < .05.

Table S2. MR Scanning Parameters across Different Institutions

Cohort	Scanner	Sequence	TR/TE (ms)	FOV (mm)	Matrix	Slice Thickness (mm)	Slice Gap (mm)	Flip Angle
I	GE 3.0T (Signa HDxt)	T1+C	5.6/2.2	380×380	320×320	3.0	0	10°
II	Siemens 1.5T (Avanto)	T1+C	4.43/1.5	340×340	448×336	1.7	0.34	10°
III	GE 1.5T (Signa HDx)	T1+C	6.5/2.1	380×342	256×256	3.0	0	10°
IV	Siemens 3.0T (Skyra)	T1+C	5.65/2.46	360×360	384×384	2.5	0.5	15°
V	GE 1.5T (Optima)	T1+C	5.2/2.15	340×340	256×244	1.5	0.7	12°
VI	Siemens 3.0T (Skyra)	T1+C	3.6/1.25	380×380	320×320	2.0	0.5	10°
VII	Philips 3.0T (Ingenia)	T1+C	4.8/2.1	280×340	280×339	1.0	0	12°
VIII	GE 3.0T (Signa HDx)	T1+C	5.6/2.2	380×380	320×320	3.0	0	10°
IX	Philips 1.5T (Achieva)	T1+C	5.0/2.2	360×360	320×320	1.5	0.3	12°
X	Philip 3.0T (Ingenia)	T1+C	4.3/2.1	280×339	280×337	2.0	0.6	12°

Note. - The institutions I-III was used as the training cohort, and institutions IV-X were used as the validation cohort. T1, T1-weighted imaging, FOV, Field of View, TR, Repetition Time, TE, Echo Time.

Table S3. Comparison of model performance across different models in three cohorts

Model	Training AUC (95% CI)	Internal Validation AUC (95% CI)	External Validation AUC (95% CI)	<i>P</i> -value
Stacking model	0.939 (0.921, 0.958)	0.867 (0.836, 0.898)	0.830 (0.806, 0.854)	Reference
MRI model	0.901 (0.877, 0.925)	0.823 (0.779, 0.867)	0.782 (0.755, 0.809)	0.003
MMG model	0.833 (0.802, 0.864)	0.764 (0.712, 0.816)	0.739 (0.710, 0.768)	<0.001
Random forest	0.835 (0.801, 0.869)	0.771 (0.720, 0.822)	0.758 (0.728, 0.788)	<0.001
SVM	0.821 (0.785, 0.857)	0.763 (0.711, 0.815)	0.741 (0.710, 0.772)	<0.001
Logistic regression	0.812 (0.774, 0.850)	0.748 (0.695, 0.801)	0.726 (0.694, 0.758)	<0.001
AUS	0.709 (0.675, 0.743)	0.696 (0.641, 0.751)	0.701 (0.677, 0.725)	<0.001

Table S4. Top 20 enriched pathways across transcriptomic and proteomic GSEA scores

Rank	Pathway	Transcriptomic GSEA Mean (SD)	Proteomic GSEA Mean (SD)	Spearman r	FDR
1	Integrin-mediated cell adhesion	0.31 (0.18)	0.28 (0.16)	0.54	<0.001
2	Actin cytoskeleton organization	0.28 (0.21)	0.25 (0.19)	0.51	<0.001
3	MHC class II antigen presentation	0.22 (0.17)	0.20 (0.15)	0.49	0.001
4	DNA damage response	0.35 (0.22)	0.31 (0.20)	0.48	0.001
5	VEGF signaling pathway	0.19 (0.14)	0.17 (0.13)	0.46	0.002
6	TGF-beta receptor signaling	0.24 (0.19)	0.21 (0.17)	0.45	0.002
7	Epithelial-to-mesenchymal transition	0.33 (0.24)	0.29 (0.21)	0.44	0.003
8	Cell adhesion molecule binding	0.26 (0.16)	0.23 (0.15)	0.44	0.003
9	MHC class I antigen presentation	0.20 (0.15)	0.18 (0.14)	0.43	0.004
10	Regulation of T cell activation	0.18 (0.13)	0.16 (0.12)	0.43	0.004
11	Cytokine-cytokine receptor interaction	0.29 (0.20)	0.25 (0.18)	0.42	0.005
12	Extracellular matrix organization	0.34 (0.23)	0.30 (0.21)	0.42	0.005
13	PI3K-Akt signaling pathway	0.27 (0.18)	0.24 (0.17)	0.41	0.006
14	mTORC1 signaling	0.23 (0.16)	0.20 (0.15)	0.41	0.006
15	Focal adhesion	0.30 (0.21)	0.26 (0.19)	0.41	0.007
16	Regulation of apoptotic process	0.21 (0.15)	0.18 (0.14)	0.41	0.007
17	NK cell-mediated cytotoxicity	0.17 (0.13)	0.15 (0.12)	0.40	0.009
18	Protein ubiquitination pathway	0.25 (0.17)	0.22 (0.16)	0.40	0.009
19	Complement and coagulation cascades	0.22 (0.16)	0.19 (0.15)	0.40	0.010
20	Wnt signaling pathway	0.20 (0.15)	0.17 (0.14)	0.40	0.010

Table S5. Comparison of model performance across this study and other studies in the external validation cohort

Model	Architecture	Input	External Validation AUC (95% CI)	<i>P</i> vs. Stacking
Stacking model (this study)	Swin Transformer	MRI + MMG + clinical	0.830 (0.806, 0.854)	Reference
MRI model (this study)	Swin Transformer	MRI	0.782 (0.755, 0.809)	0.003
MMG model (this study)	Swin Transformer	MMG	0.739 (0.710, 0.768)	<0.001
Wang et al. 2022 (reimplemented)	CNN	Multi-parametric MRI	0.779 (0.750, 0.808)	<0.001
Gao et al. 2024 (reimplemented)	CNN-RNN	MRI kinetics	0.795 (0.767, 0.823)	<0.001
Chen et al. 2023 (reimplemented)	CNN	Multi-parametric MRI	0.768 (0.738, 0.798)	<0.001
Random forest	Ensemble	Radiomics + clinical	0.758 (0.728, 0.788)	<0.001
SVM	Kernel	Radiomics + clinical	0.741 (0.710, 0.772)	<0.001
Logistic regression (radiomics)	Linear	Radiomics + clinical	0.726 (0.694, 0.758)	<0.001
AUS	Clinical assessment	Sonographic features	0.701 (0.677, 0.725)	<0.001

Table S6. Hub gene validation for plum1 and saddlebrown WGCNA modules

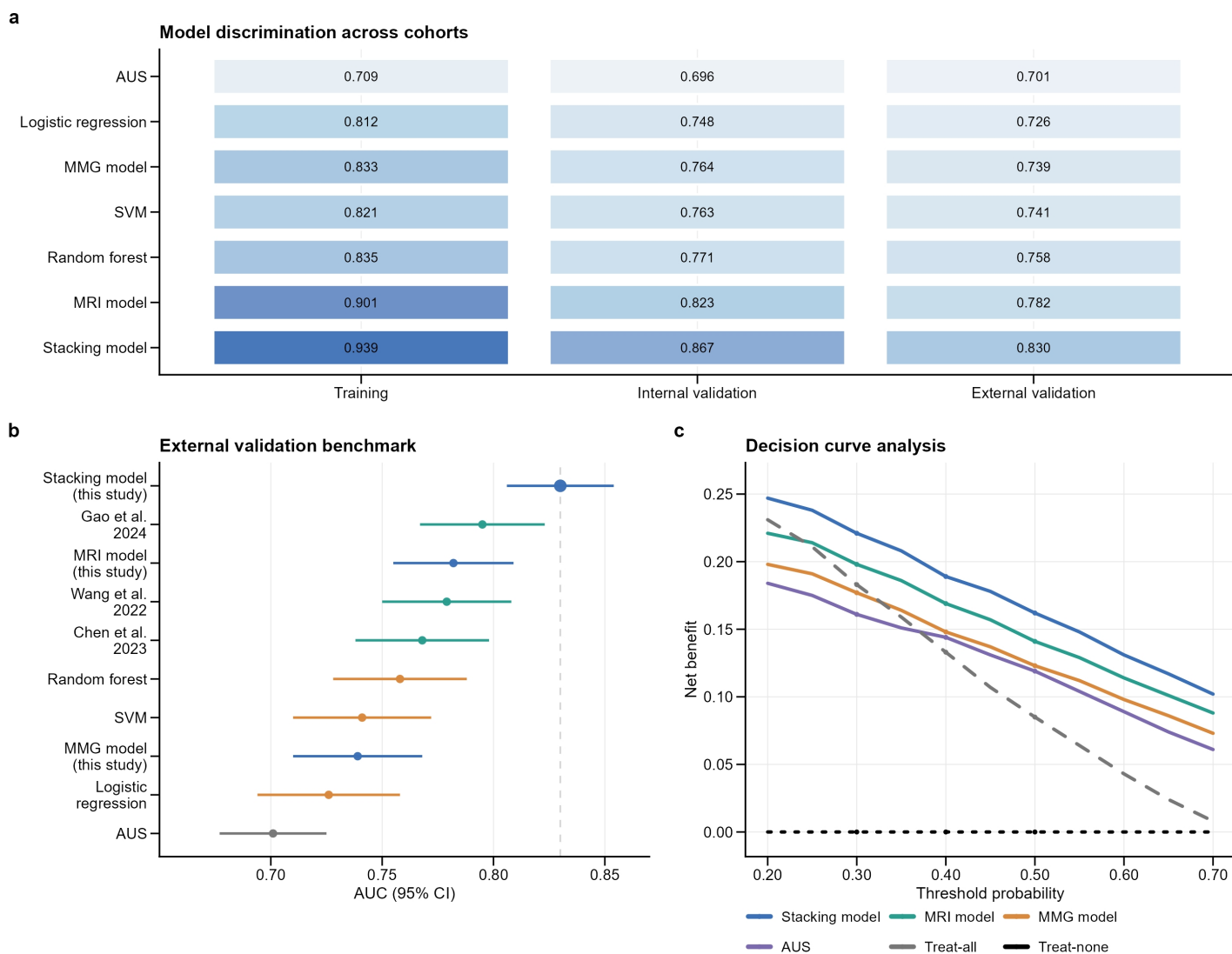
Module	Gene	kME Score	Spearman r	P-value	FDR	Known Function
Plum1	KIF14	0.89	0.41	<0.001	<0.001	Mitotic kinesin; DNA replication fidelity
Plum1	MCM6	0.87	0.39	<0.001	<0.001	DNA replication licensing factor
Plum1	PCNA	0.85	0.37	<0.001	<0.001	DNA sliding clamp; replication and repair
Plum1	RFC4	0.83	0.35	<0.001	0.001	Replication factor C; replication fork
Plum1	RAD51AP1	0.81	0.33	<0.001	0.002	Homologous recombination; DNA repair
Saddlebrown	ITGB4	0.88	-0.38	<0.001	<0.001	Integrin beta-4; hemidesmosome formation
Saddlebrown	LAMC2	0.85	-0.36	<0.001	<0.001	Laminin gamma-2; basement membrane
Saddlebrown	CD44	0.83	-0.35	<0.001	0.001	Adhesion receptor; cancer stem cell marker
Saddlebrown	VCL	0.81	-0.32	<0.001	0.002	Vinculin; focal adhesion scaffolding
Saddlebrown	FN1	0.79	-0.30	<0.001	0.003	Fibronectin-1; ECM remodeling and migration

Table S7. Decision curve analysis net benefit of the stacking model and comparators across different threshold probabilities

Threshold Probability	Stacking Model	MRI Model	MMG Model	AUS	Treat-All	Treat-None
0.20	0.247	0.221	0.198	0.184	0.231	0
0.25	0.238	0.214	0.191	0.175	0.211	0
0.30	0.221	0.198	0.177	0.161	0.183	0
0.35	0.208	0.186	0.164	0.151	0.159	0
0.40	0.189	0.169	0.148	0.144	0.133	0
0.45	0.178	0.157	0.137	0.131	0.107	0
0.50	0.162	0.141	0.123	0.119	0.085	0
0.55	0.148	0.129	0.112	0.104	0.064	0
0.60	0.131	0.114	0.098	0.089	0.043	0
0.65	0.117	0.101	0.086	0.074	0.024	0
0.70	0.102	0.088	0.073	0.061	0.008	0

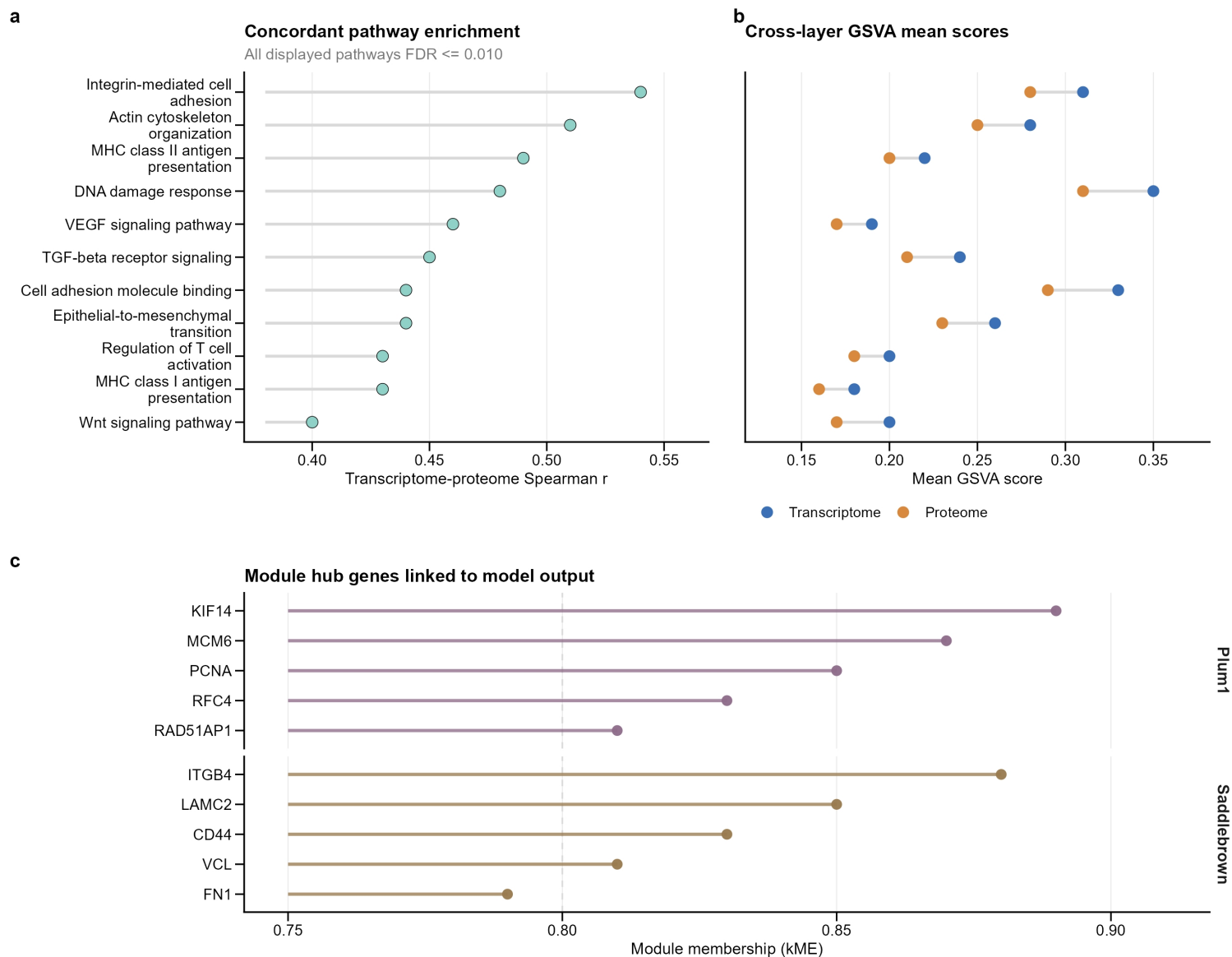
Table S8. Comprehensive functional enrichment analysis of plum1 and saddlebrown WGCNA modules

Module	Database	Term ID	Term Description	Overlap (n)	BH-FDR	Representative Genes
plum1	GO-BP	GO:0045321	Leukocyte activation	28	0.0003	CD8A, CD4, CD3E, PDCD1, CD274
plum1	GO-BP	GO:0050776	Regulation of immune response	31	0.0004	CTLA4, LAG3, TIGIT, FOXP3, HAVCR2
plum1	GO-CC	GO:0005886	Plasma membrane	78	0.0004	CD274, PDCD1, CD3E, CD8A, CTLA4
plum1	GO-CC	GO:0042613	MHC class II protein complex	9	0.0008	HLA-DRA, HLA-DRB1, CD74, HLA-DPA1
plum1	GO-MF	GO:0005125	Cytokine activity	22	0.0004	IL2, IFNG, CXCL9, CXCL10, IL10
plum1	GO-MF	GO:0042379	Chemokine receptor binding	12	0.0006	CXCL9, CXCL10, CCL5, CXCL11
plum1	KEGG	hsa04514	Cell adhesion molecules	18	0.0002	CD274, PDCD1, CD8A, CD4, HLA-DRA
plum1	KEGG	hsa05235	PD-L1 and PD-1 checkpoint	14	0.0004	CD274, PDCD1, CD8A, CTLA4, LAG3
saddlebrown	GO-BP	GO:0030036	Actin cytoskeleton organization	31	0.0002	ACTN1, FN1, ITGB1, VCL, RHOA
saddlebrown	GO-BP	GO:0016477	Cell migration	36	0.0003	FN1, ITGB1, VIM, CDH2, RHOA
saddlebrown	GO-CC	GO:0005925	Focal adhesion	28	0.0002	VCL, ITGB1, ACTN1, PXN, TLN1
saddlebrown	GO-CC	GO:0005578	Proteinaceous extracellular matrix	29	0.0005	FN1, COL1A1, COL1A2, VIM, LOX
saddlebrown	GO-MF	GO:0003779	Actin binding	24	0.0003	ACTN1, VCL, DIAPH1, MYH9, FLNB
saddlebrown	GO-MF	GO:0005178	Integrin binding	12	0.0005	FN1, ITGB1, VCAM1, ICAM1, VCL
saddlebrown	KEGG	hsa04510	Focal adhesion	33	0.0001	ITGB1, VCL, FN1, ACTN1, PTK2
saddlebrown	KEGG	hsa04151	PI3K-Akt signaling pathway	29	0.0008	ITGB1, FN1, PTK2, PIK3CA, COL1A1



Supplementary Fig. 1. Additional model-comparison and clinical-utility analyses. a,

Heatmap summarizing AUC values across the training, internal validation, and external validation cohorts for the stacking model, single-modality models, conventional machine-learning baselines, and axillary ultrasound (AUS). b, External validation benchmark comparing the stacking model with study-specific single-modality models, reimplemented prior deep-learning architectures, conventional radiomics baselines, and AUS. Points indicate AUC values and horizontal lines indicate 95% confidence intervals. The dashed vertical line marks the external validation AUC of the stacking model. c, Decision curve analysis in the external validation cohort. Net benefit is shown across threshold probabilities from 0.20 to 0.70 for the stacking model, MRI model, MMG model, AUS, treat-all strategy, and treat-none strategy.



Supplementary Fig. 2. Additional multi-omics concordance and WGCNA hub-gene analyses. a, Concordant pathway enrichment across transcriptomic and proteomic GSVA scores. Points show pathway-level Spearman correlations between transcriptomic and proteomic enrichment scores. b, Mean transcriptomic and proteomic GSVA scores for the same concordantly enriched pathways. c, WGCNA hub genes associated with model output in the plum1 and saddlebrown modules. Points indicate module membership (kME), with colors denoting module identity. The dashed vertical line marks kME = 0.80.