CHERISH Supplementary Information

This file provides concise legends for supplementary figures and reviewer-facing tables that support the manuscript results and methods. The embedded supplementary tables focus on cohort denominators, validation metrics, biological-context analyses and statistical summaries.

Contents

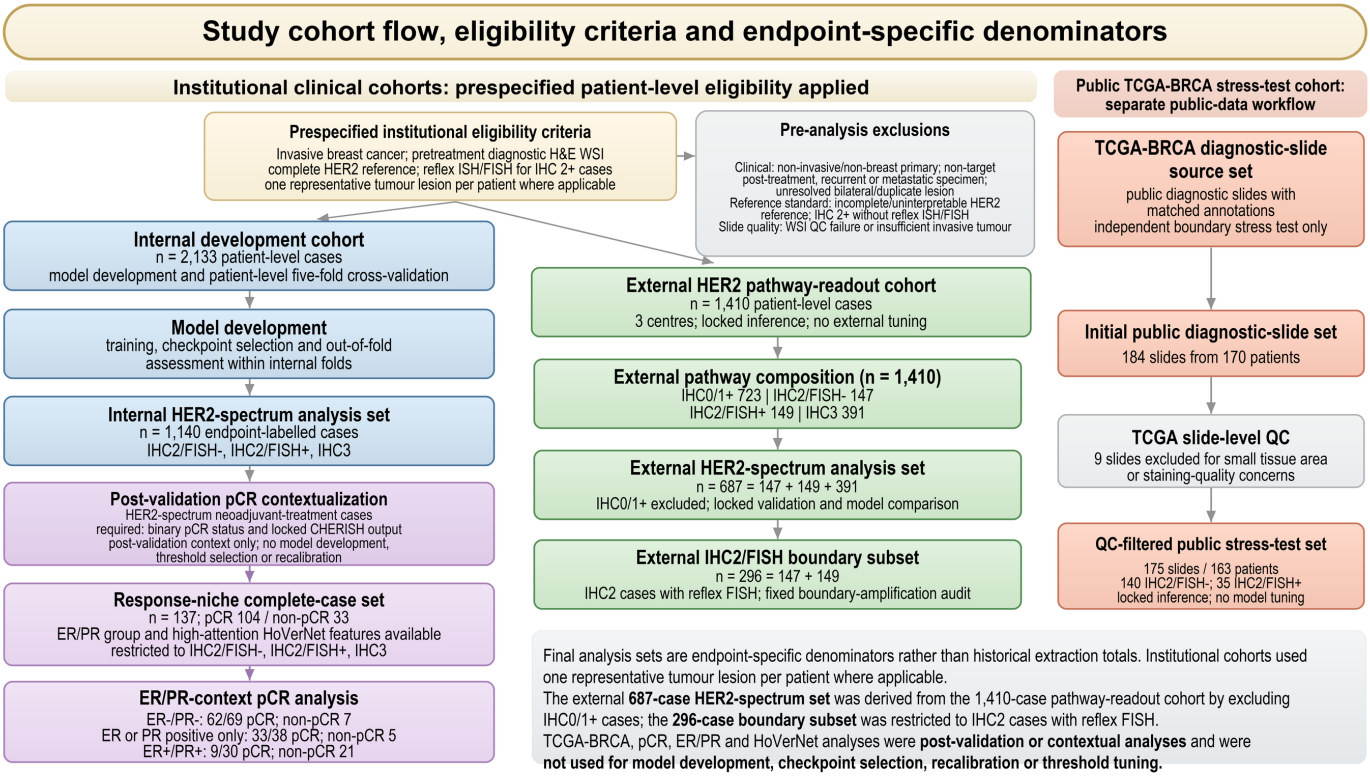
Supplementary Figures

Supplementary Tables

Supplementary Notes

Supplementary Figures

Supplementary Fig. 1

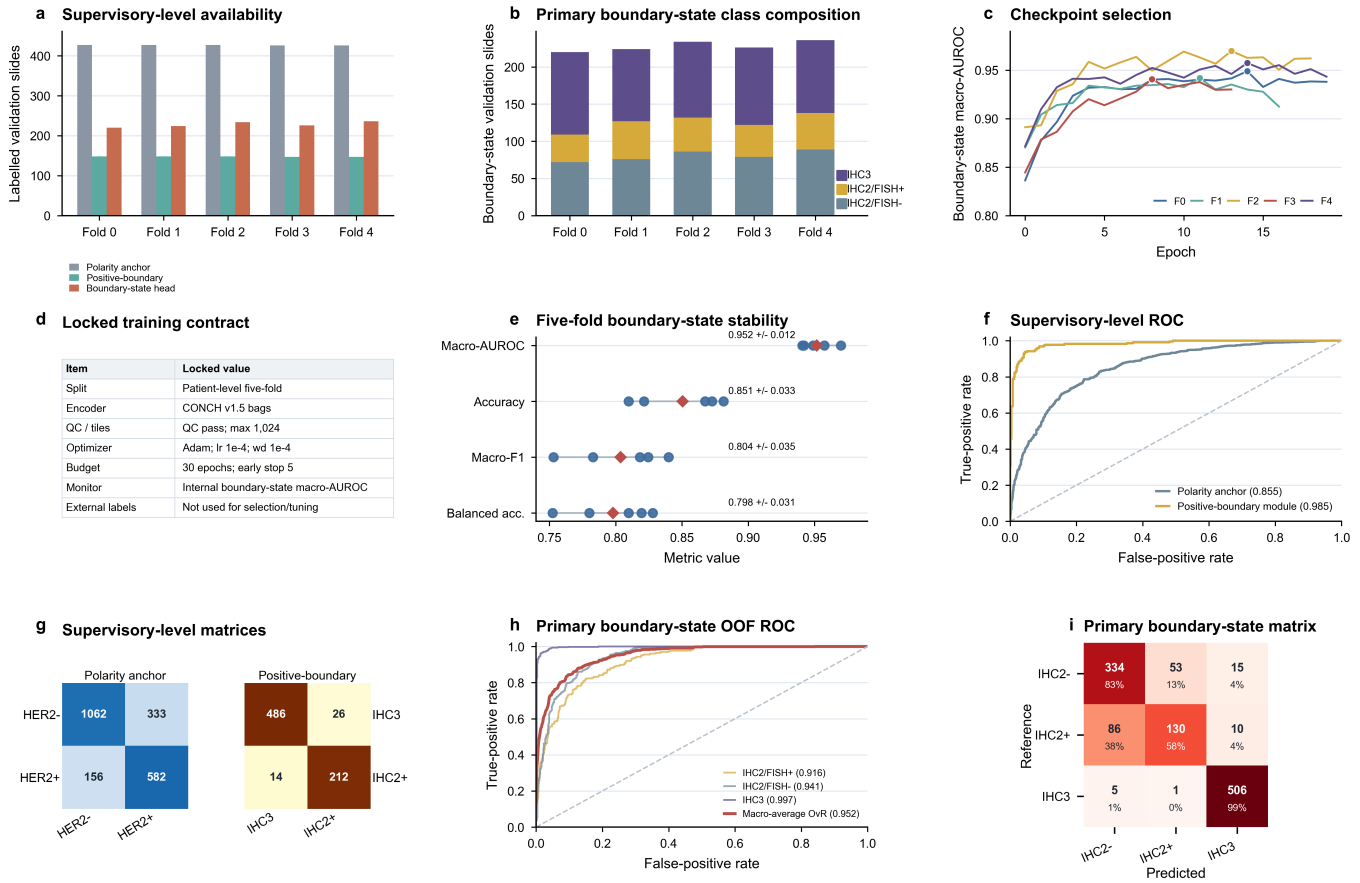


Supplementary Fig. 1

Supplementary Fig. 1 | Study cohort flow and endpoint denominators. Prespecified eligibility, exclusion categories and final analysis sets for the internal development cohort, external pathway-readout cohort, external positive-spectrum set, IHC 2+/FISH boundary subset, TCGA-BRCA stress-test set and pCR/HoVerNet context set.

Supplementary Fig. 2

Internal cross-validation audit of CHERISH supervisory levels



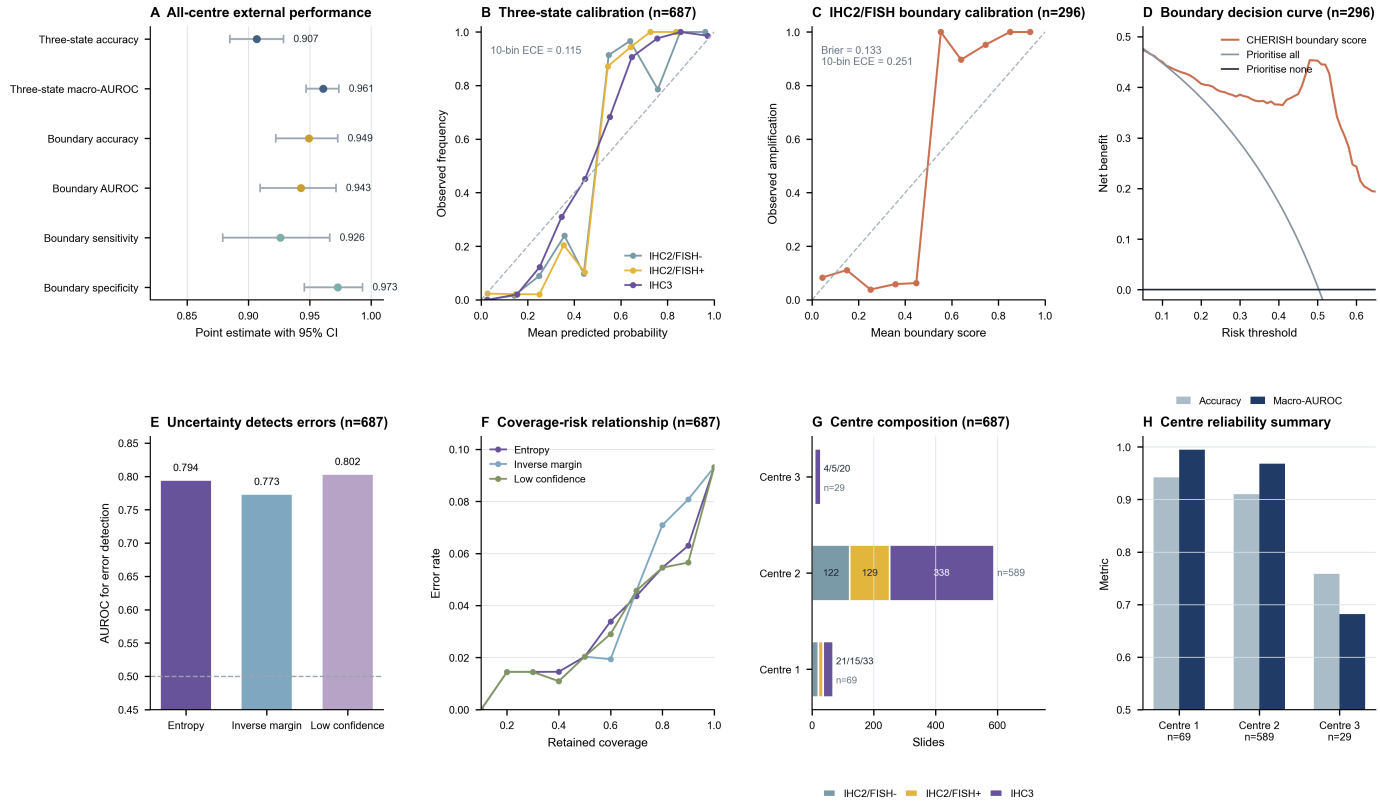
Panels use deidentified five-fold validation summaries and OOF predictions. The polarity anchor, positive-boundary module and primary boundary-state head are reported as ordered supervisory levels rather than independent clinical endpoints.

Supplementary Fig. 2

Supplementary Fig. 2 | Internal cross-validation of CHERISH supervisory levels. Five-fold internal validation of the polarity anchor, positive-boundary module and primary boundary-state head, including fold composition, checkpoint selection and out-of-fold performance.

Supplementary Fig. 3

Supplementary Fig. 3 | Locked external validation and reliability audit

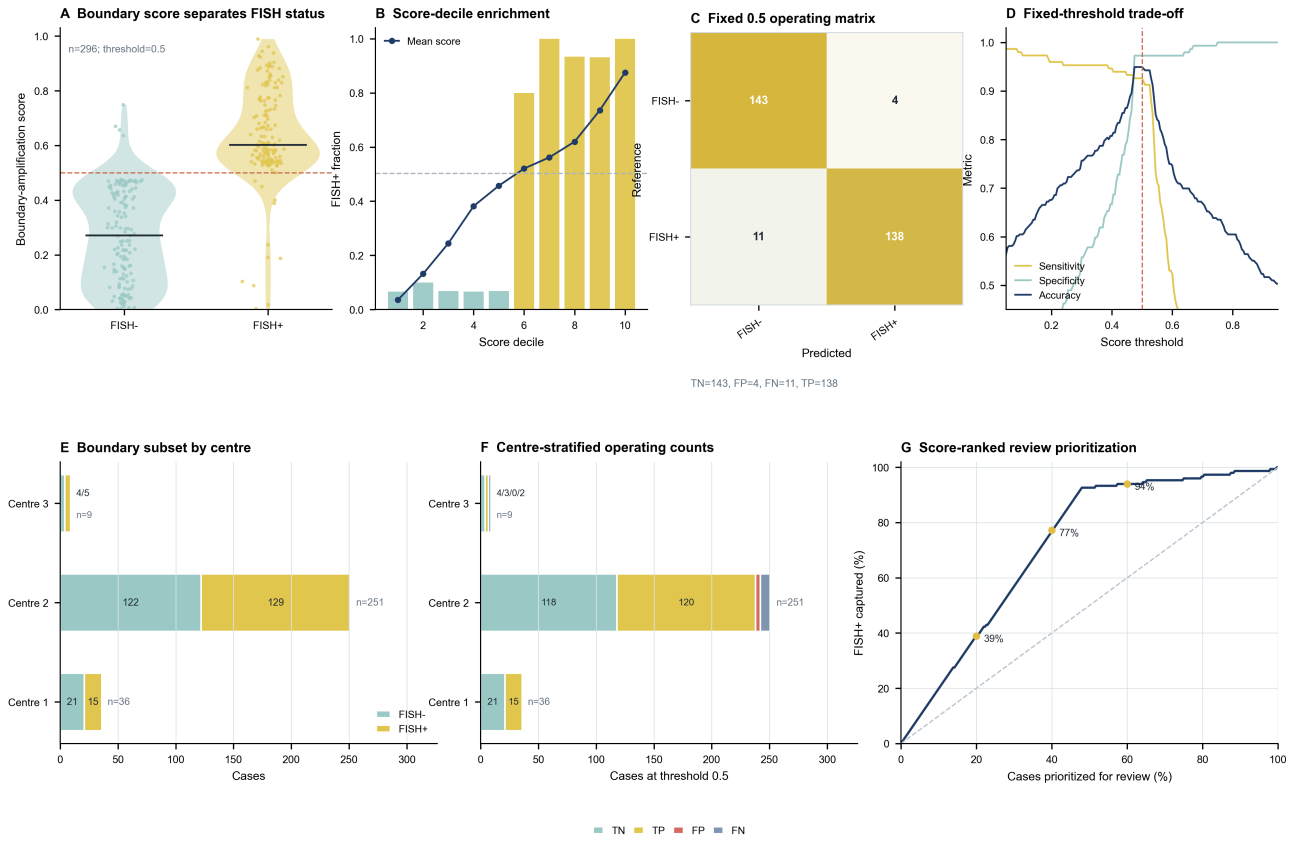


Supplementary Fig. 3

Supplementary Fig. 3 | Locked external validation and reliability assessment. External validation of the finalized model in the positive-spectrum and IHC 2+/FISH boundary analysis sets, with calibration, uncertainty and centre-level summaries.

Supplementary Fig. 4

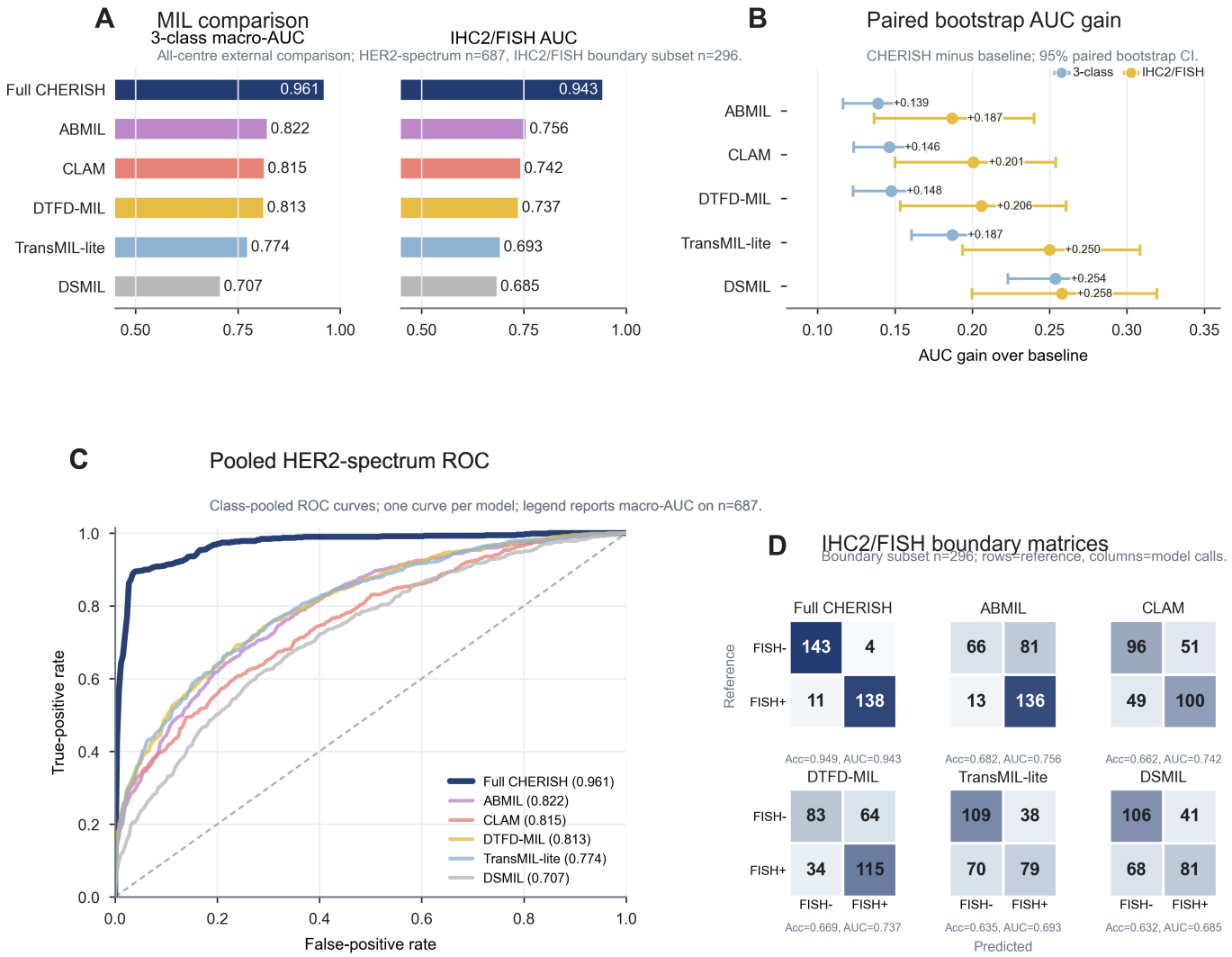
Supplementary Fig. 4 | IHC2/FISH boundary-use and review-prioritization audit



Supplementary Fig. 4

Supplementary Fig. 4 | IHC 2+/FISH boundary use and review prioritization. Fixed-threshold boundary-amplification assessment in the external IHC 2+ subset, centre-stratified operating counts and score-ranked review-prioritization analysis. Four-stratum external readout counts are reported in Fig. 1e, Table 1 and Supplementary Tables 7 and 8.

Supplementary Fig. 5



Supplementary Fig. 5

Supplementary Fig. 5 | MIL model comparison and component ablation. Matched comparison of CHERISH with representative MIL baselines and component-removal ablations using locked external denominators.

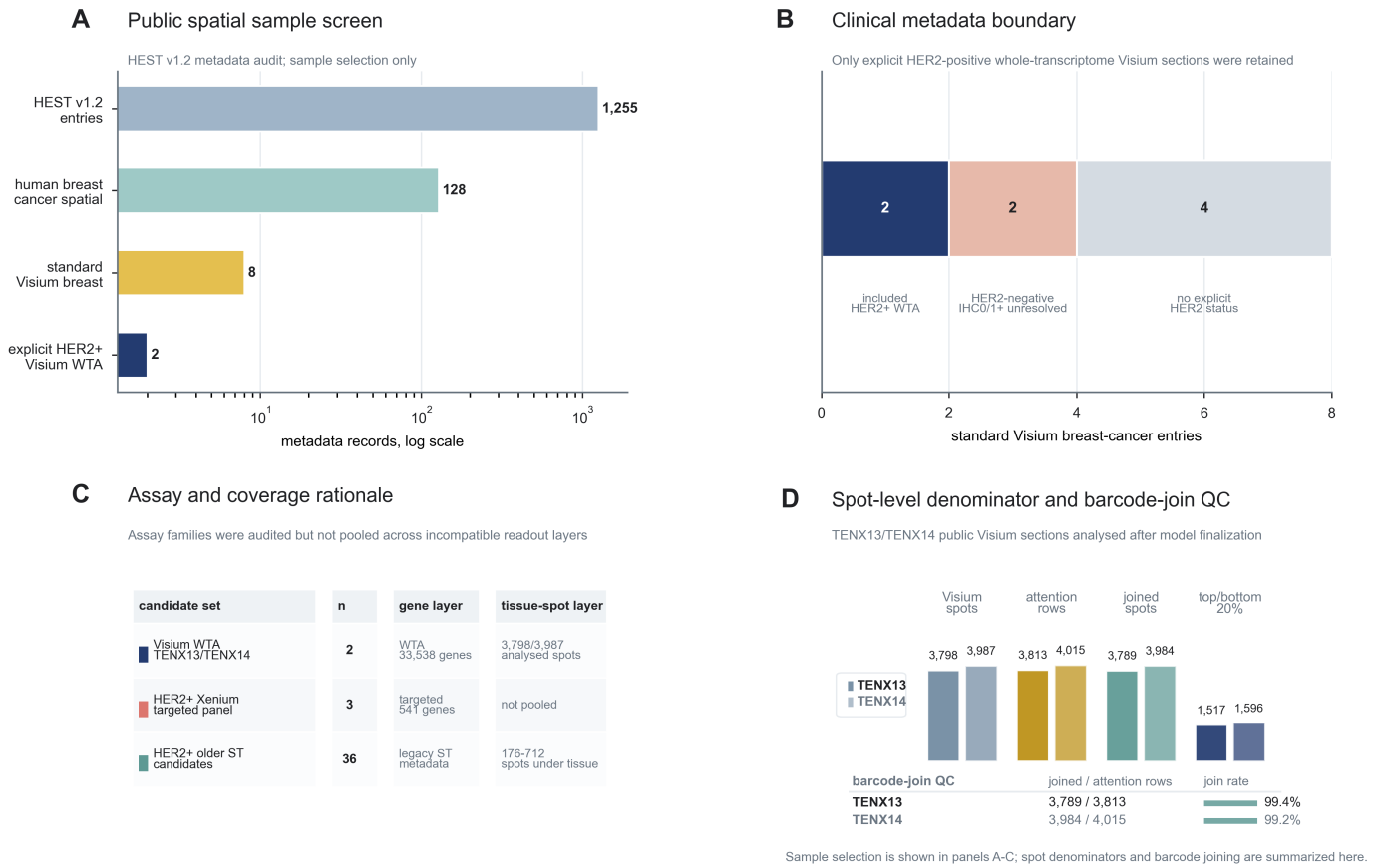
Supplementary Fig. 6

Supplementary tables provide the full numeric values for the HoVerNet differences summarized in the main figure.

Supplementary Fig. 6

Supplementary Fig. 6 | HoVerNet metric definitions and cellular evidence. HoVerNet metric families and quantitative cellular evidence supporting the epithelial-immune interface interpretation of CHERISH-attended morphology.

Supplementary Fig. 7

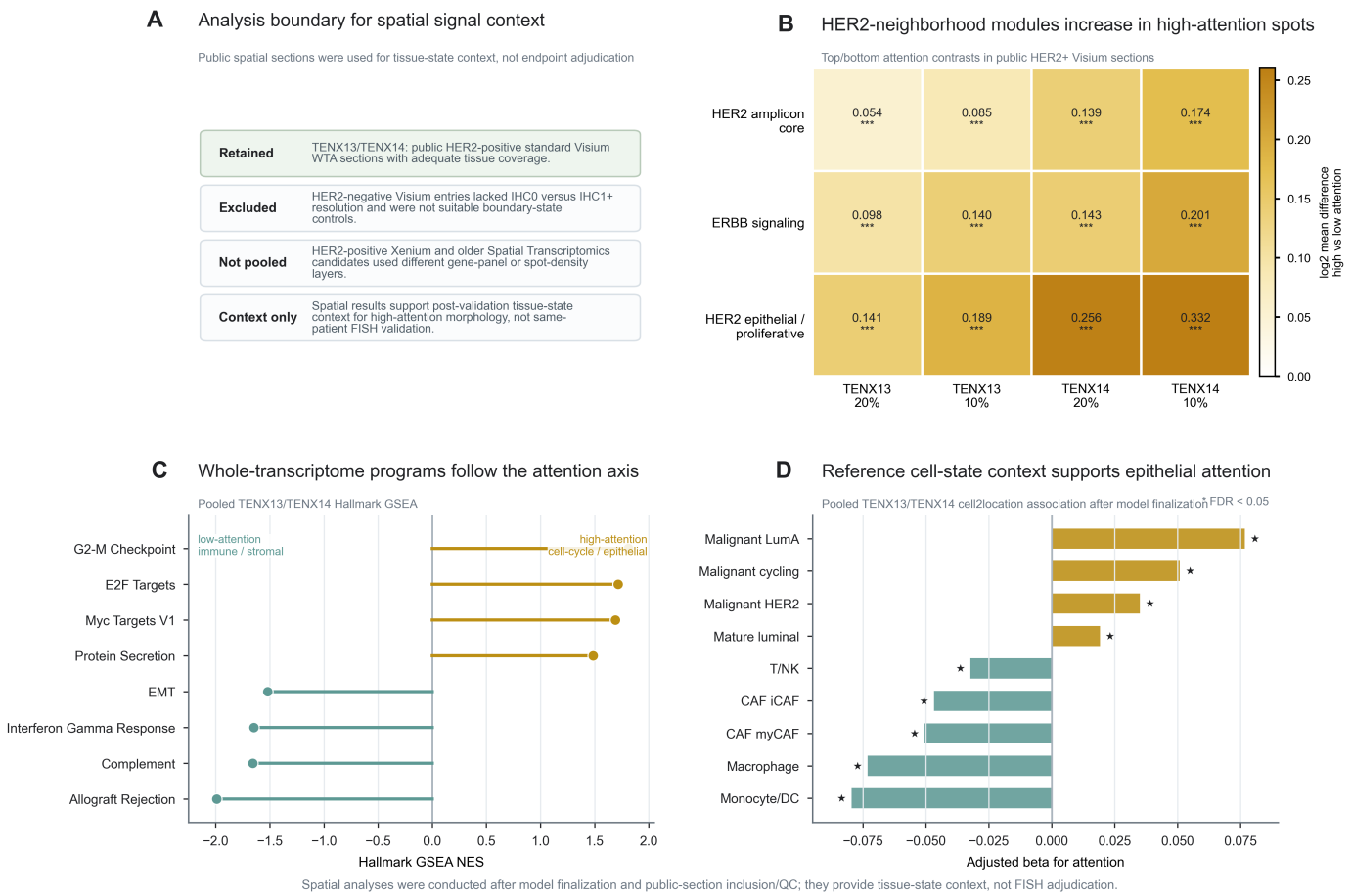


Public spatial sections were selected and audited before tissue-state signal analysis; raw HEST data are not redistributed.

Supplementary Fig. 7

Supplementary Fig. 7 | Public spatial sample inclusion and spot QC. Selection of public HEST/10x Visium HER2-positive breast-cancer samples and spot/barcode denominator checks for TENX13 and TENX14.

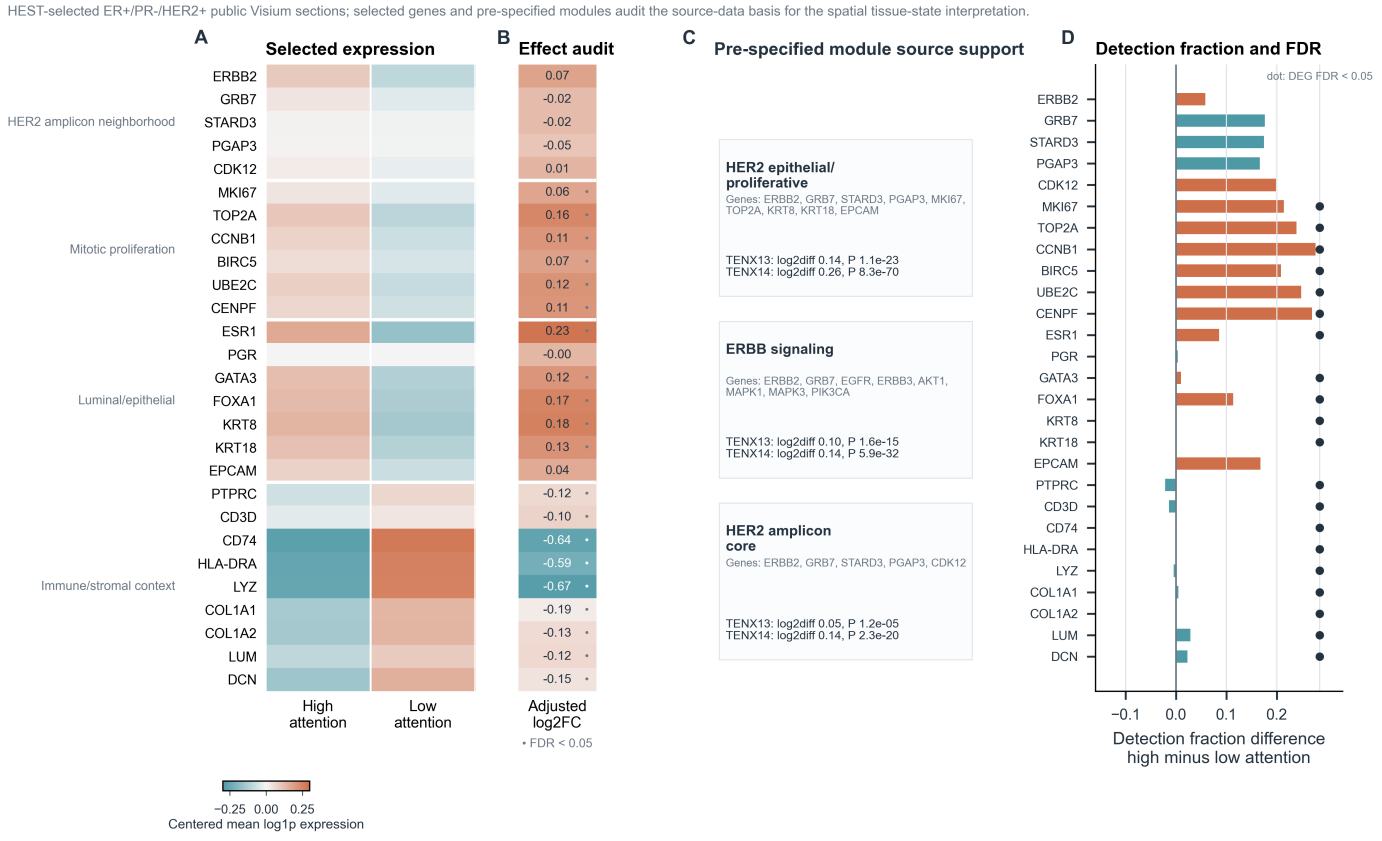
Supplementary Fig. 8



Supplementary Fig. 8

Supplementary Fig. 8 | Public Visium spatial signal context. Spatial transcriptomic contrasts linking CHERISH high-attention regions to HER2-neighbourhood, proliferation and cell-state readouts in TENX13/TENX14.

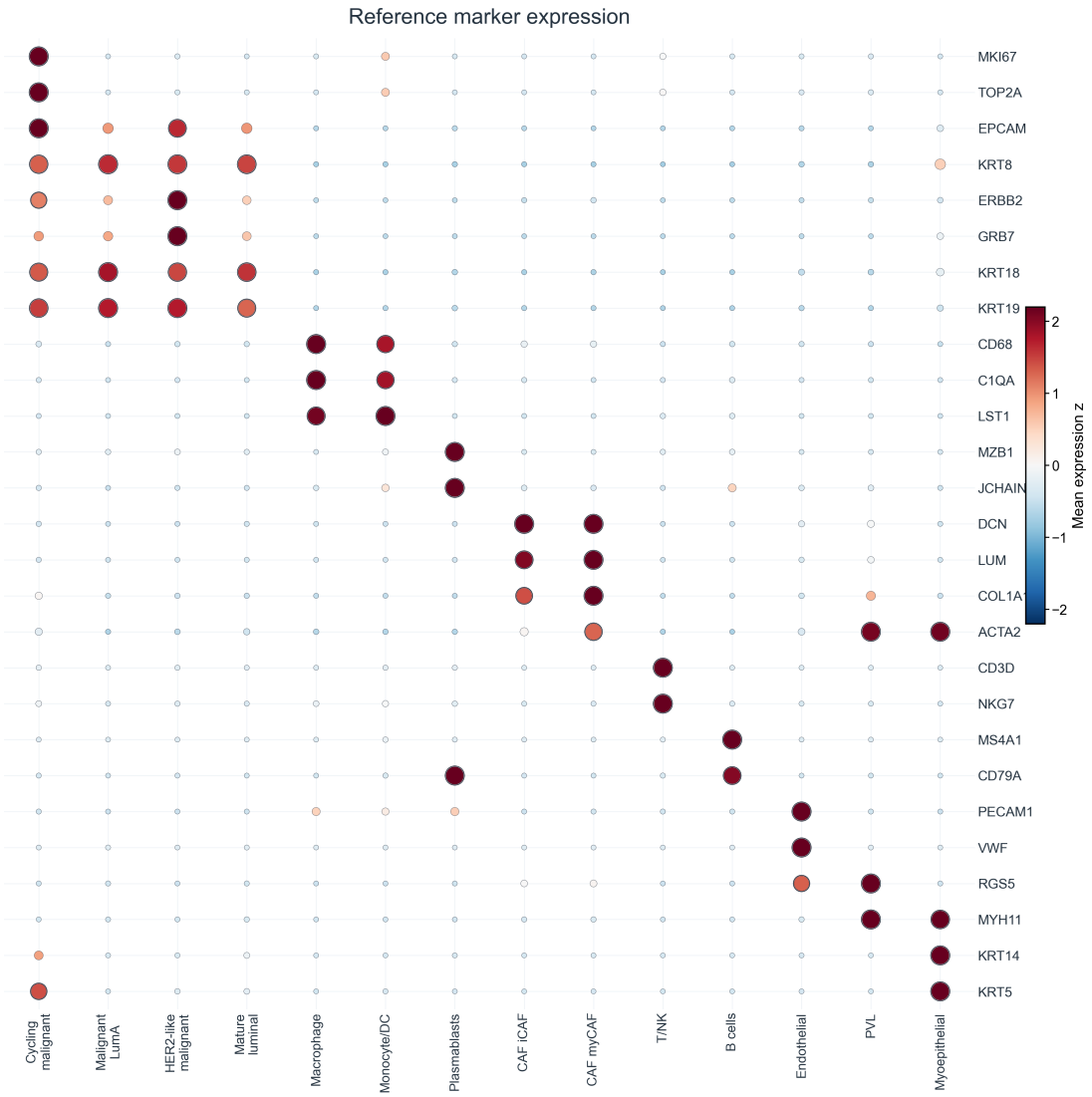
Supplementary Fig. 9



Supplementary Fig. 9

Supplementary Fig. 9 | Gene-level and module evidence for spatial context. Gene-level and module-level checks supporting the spatial transcriptomic interpretation of high-attention regions.

Supplementary Fig. 10



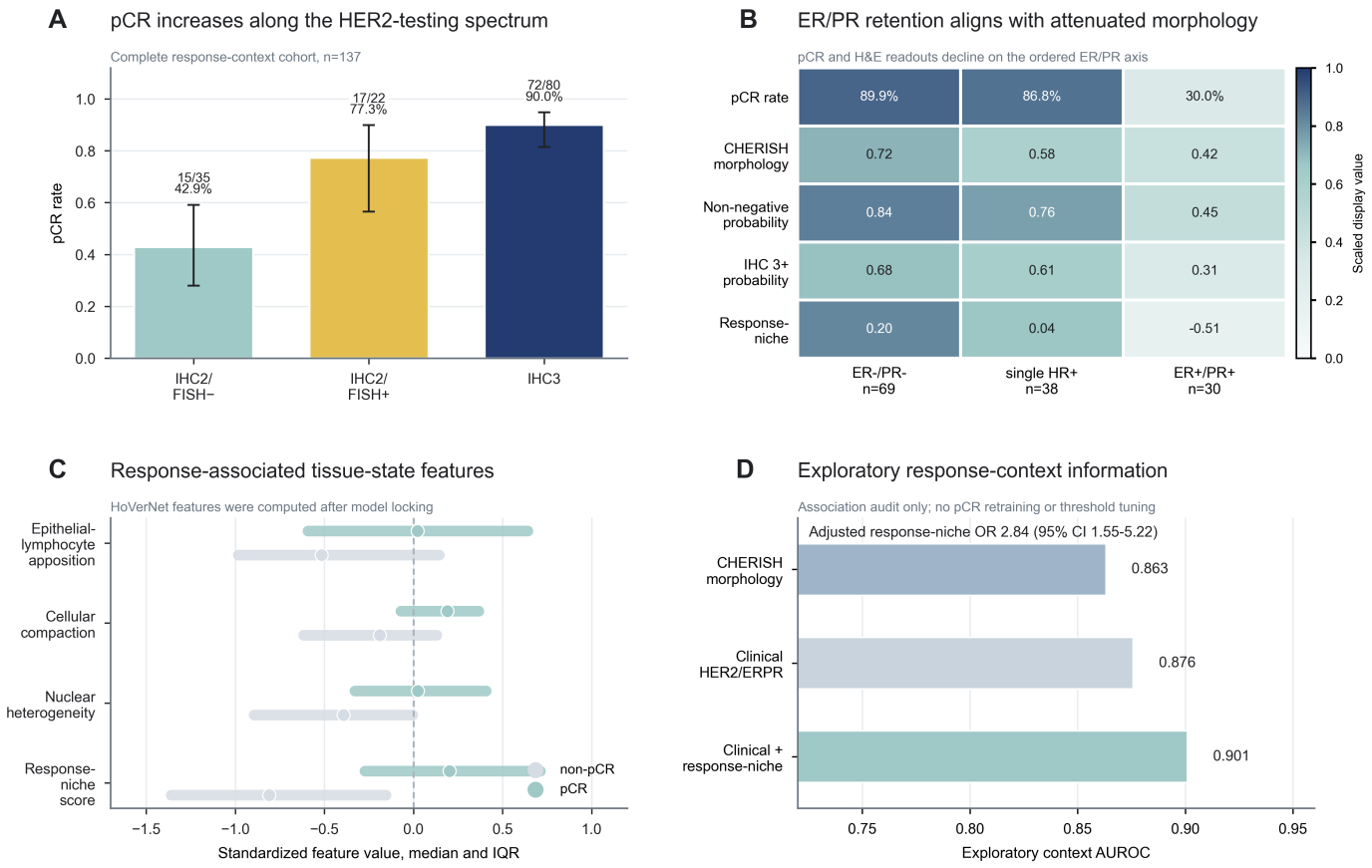
Color shows within-marker log1p-expression z-score across reference cell states; point area shows relative log1p expression.

Marker profiles audit reference cell-state labels used for post-lock spatial contextualization, not FISH adjudication.

Supplementary Fig. 10

Supplementary Fig. 10 | cell2location marker-expression support. Reference marker-expression profiles used to interpret malignant epithelial, cycling, immune and stromal cell-state estimates.

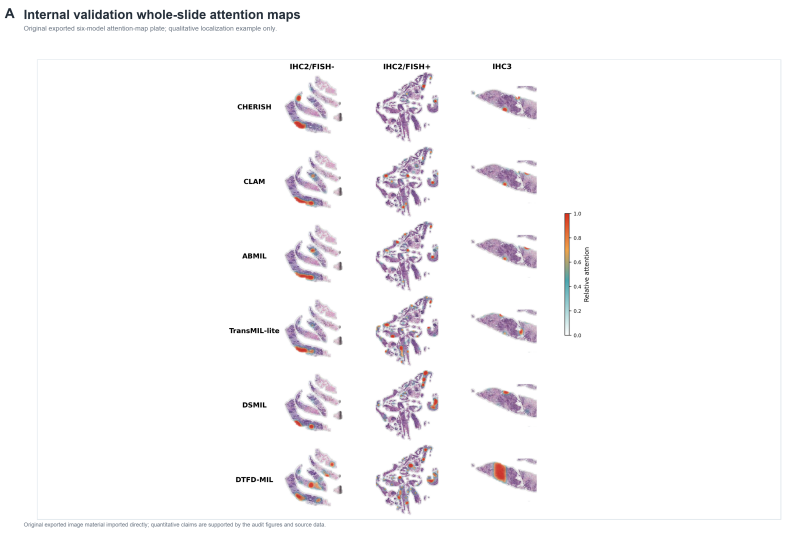
Supplementary Fig. 11

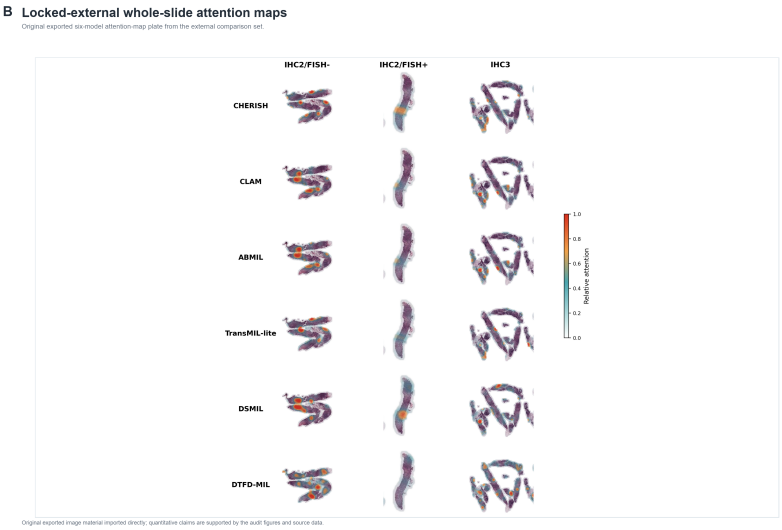


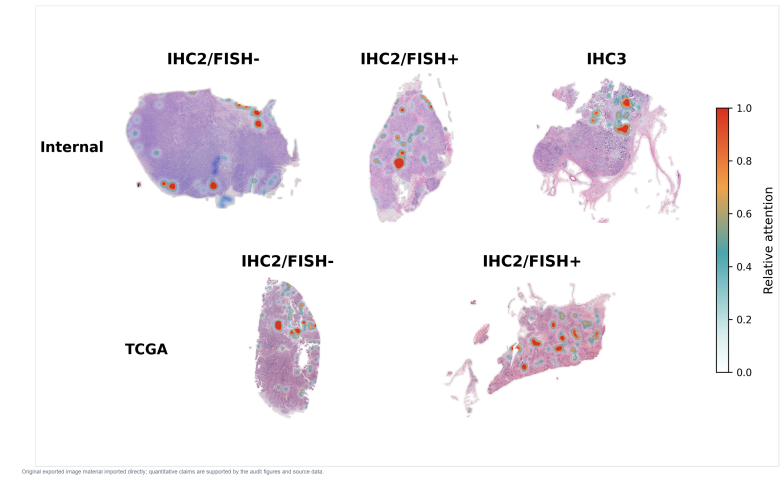
Supplementary Fig. 11

Supplementary Fig. 11 | Treatment-response and receptor-context assessment. Post-lock pCR, response-niche and ER/PR receptor-context analyses in the response-follow-up cohort.

Supplementary Fig. 12







Supplementary Fig. 12 | Whole-slide attention and heatmap examples. Representative qualitative CHERISH attention maps from internal, external, surgery and public TCGA examples. Quantitative interpretation is provided in the corresponding HoVerNet, spatial, pCR and review-alert analyses.

Supplementary Tables

Supplementary Tables 1-23 are embedded in full in the accompanying reviewer-facing SI PDF and retained as the source workbook CHERISH_Supplementary_Tables.xlsx. The concise index below defines the role of each result-supporting table.

Supplementary Table 1. Study cohort flow, eligibility criteria and endpoint-specific denominators.

Defines the analysis denominators used throughout the manuscript, including the internal development cohort, external 1,410-case readout cohort, external 687-case positive-spectrum set, IHC 2+/FISH boundary subset, TCGA-BRCA stress-test set and pCR/HoVerNet context set. The machine-readable table contains 7 data rows and 4 columns.

Supplementary Table 2. Validation lock and no-retuning record.

Documents the separation between model development, model/classification locking, external validation and post-validation biological-context analyses. The machine-readable table contains 4 data rows and 7 columns.

Supplementary Table 3. Internal five-fold training cohort composition.

Summarizes patient-level fold composition for the internal five-fold development split. The machine-readable table contains 5 data rows and 5 columns.

Supplementary Table 4. Internal primary boundary-state fold-level performance and checkpoint selection.

Reports fold-level primary boundary-state performance and checkpoint-selection information. The machine-readable table contains 5 data rows and 11 columns.

Supplementary Table 5. Pooled internal primary boundary-state out-of-fold summary.

Reports pooled out-of-fold internal primary boundary-state metrics, confusion rows and bootstrap confidence intervals for internal AUROC summaries. The machine-readable table contains 13 data rows and 5 columns.

Supplementary Table 6. All-centre finalized external performance summary.

Reports the all-centre external positive-spectrum and IHC 2+/FISH boundary performance estimates with confidence intervals. The machine-readable table contains 10 data rows and 8 columns.

Supplementary Table 7. External 1,410-case clinically matched H&E classifications.

Reports external four-stratum clinically matched H&E readouts in the 1,410-case cohort. The machine-readable table contains 6 data rows and 7 columns.

Supplementary Table 8. Centre-stratified cohort composition and final classification matrices.

Provides centre-stratified reference-state composition and final classification matrices. The machine-readable table contains 12 data rows and 8 columns.

Supplementary Table 9. External IHC 2+/FISH boundary metric summary.

Reports the fixed-threshold IHC 2+/FISH boundary metrics in the external 296-case subset. The machine-readable table contains 1 data rows and 22 columns.

Supplementary Table 10. Public TCGA image-level FISH-boundary test.

Reports the public TCGA-BRCA image-level FISH-boundary stress-test metrics after slide-level QC. The machine-readable table contains 1 data rows and 20 columns.

Supplementary Table 11. External acquisition metadata availability and transportability boundary.

Summarizes external acquisition-metadata visibility and transportability boundaries. The machine-readable table contains 4 data rows and 6 columns.

Supplementary Table 12. Internal and external comparison with representative MIL baselines.

Reports matched internal and external performance for CHERISH and representative MIL baselines. The machine-readable table contains 6 data rows and 8 columns.

Supplementary Table 13. Paired bootstrap AUC differences against MIL baselines.

Reports paired bootstrap AUC differences between CHERISH and MIL baselines. The machine-readable table contains 5 data rows and 12 columns.

Supplementary Table 14. Component-removal ablation analysis.

Reports component-removal ablations used to localize the contribution of the hierarchy and residual-refinement modules. The machine-readable table contains 7 data rows and 14 columns.

Supplementary Table 15. HoVerNet metric-definition and calculation dictionary.

Defines HoVerNet, attention-region, neighbourhood and response-niche metrics used for biological interpretation. The machine-readable table contains 22 data rows and 6 columns.

Supplementary Table 16. HoVerNet ordered HER2-spectrum metric summary.

Reports selected HoVerNet metrics following the ordered HER2-spectrum gradient. The machine-readable table contains 11 data rows and 14 columns.

Supplementary Table 17. HoVerNet FISH-positive versus FISH-negative feature-evidence summary.

Reports selected FISH-positive versus FISH-negative HoVerNet feature evidence in IHC 2+ cases. The machine-readable table contains 24 data rows and 15 columns.

Supplementary Table 18. Public spatial sample-selection summary.

Summarizes public spatial sample selection from HEST v1.2. The machine-readable table contains 6 data rows and 6 columns.

Supplementary Table 19. Public Visium inclusion and spot-denominator QC.

Reports TENX13/TENX14 Visium inclusion, spot and join denominators. The machine-readable table contains 2 data rows and 11 columns.

Supplementary Table 20. Treatment-response and receptor-context assessment.

Reports pCR, ER/PR receptor-context, CHERISH morphology-score and Wilson/Fisher statistical summaries in the response-follow-up cohort. The machine-readable table contains 21 data rows and 6 columns.

Supplementary Table 21. pCR response-niche covariate-adjusted models.

Reports pCR response-niche logistic models and covariate-adjusted sensitivity analyses. The machine-readable table contains 5 data rows and 10 columns.

Supplementary Table 22. Positive-boundary review-alert summary across cohorts.

Reports positive-boundary support among initially undercalled FISH-amplified cases in internal, external and TCGA cohorts. The machine-readable table contains 3 data rows and 7 columns.

Supplementary Table 23. Spatial transcriptomic module and cell-state source statistics.

Reports reviewer-facing statistics for public Visium spatial module contrasts, cell2location high-versus-low attention cell-state contrasts, continuous attention-cell-state associations and marker/module concordance supporting Fig. 4. The machine-readable table contains 34 data rows and 12 columns.

Supplementary Notes

Supplementary Note 1. Residual refinement and clinical-pathway base distribution. For slide i, the intermediate pathway modules define a three-element clinical-pathway base distribution b_i over IHC 2+/FISH-negative, IHC 2+/FISH-amplified and IHC 3+ states. The morphology residual branch outputs a three-dimensional residual logit vector r_i, and final state logits are computed as log{clip(b_i, epsilon, 1)} + r_i. When the residual branch contributes no additional morphology information, the prediction degenerates to the structured clinical-pathway base distribution. The residual branch therefore refines, rather than replaces, the clinical-pathway prior.

Supplementary Note 2. Training provenance and effective sample accounting. The full CHERISH training configuration, fold-level effective sample counts, checkpoint-selection summaries, ROC/calibration source tables and deidentified epoch summaries are retained in the controlled source-data archive and summarized in Supplementary Fig. 2 and Supplementary Tables 3-5. These files document the patient-level five-fold split, tile budget, loss weights, hierarchy-loss schedule, optimizer settings and validation endpoint used for checkpoint selection. Raw slide identifiers, patient identifiers, institution-specific feature-bag references and checkpoint binaries are not redistributed.

Supplementary Note 3. HoVerNet response-niche score derivation. The response-niche score was defined as the arithmetic mean of three slide-level z-scored high-attention HoVerNet features: epithelial-lymphocyte apposition, local cellular compaction and epithelial nuclear heterogeneity. The z-score transformation was fit within the complete pCR biological-context analysis set before logistic regression. Descriptive summaries, bootstrap intervals and multivariable sensitivity models are summarized in the Supplementary Tables; extended source tables are retained in the controlled source-data archive for the journal checklist or final source-data stage.

Supplementary Note 4. TCGA-BRCA HER2/IHC/FISH label curation. HER2/IHC/FISH labels for the public TCGA-BRCA stress-test cohort were abstracted from GDC-accessible pathology reports and phenotype annotations. Case-level and slide-level records were harmonized, available IHC scores, FISH amplification calls, HER2/CEP17 ratio and copy-number information were extracted where reported, and locked slide-level boundary labels were generated through scripted filtering followed by manual review. This public cohort was used as an independent image-level stress test under heterogeneous public-data conditions.

Controlled Source-Data Archive 4. Biological-context source data. Contains HoVerNet source tables and metric definitions, pCR/ERPR context tables, receptor-context summaries, TCGA image-level FISH-boundary summaries, HEST sample-selection summaries and spatial transcriptomic source tables.

Supplementary Table 1. Cohort composition and main diagnostic endpoints.

Source workbook: Table 1; 7 data rows and 4 columns.

analysis set	role	n	reference composition
Internal development cohort	Model development and patient-level five-fold cross-validation	2,133	Patient-level invasive breast cancer cases with available diagnostic H&E WSI and eligible clinical HER2 reference information
Internal HER2-positive-spectrum analysis set	Internal out-of-fold validation of the primary HER2 testing-pathway boundary-state endpoint	1,140	IHC2/FISH-; IHC2/FISH+; IHC3 endpoint-labelled cases
External HER2 pathway-readout cohort	Clinical-facing H&E readout assessment under locked inference	1,410	IHC0/1+ 723; IHC2/FISH- 147; IHC2/FISH+ 149; IHC3 391
All-centre external HER2-positive-spectrum validation/model-comparison cohort	Locked three-state validation, reliability assessment and model comparison	687	IHC2/FISH- 147; IHC2/FISH+ 149; IHC3 391
External IHC2/FISH boundary subset	Conditional fixed-threshold boundary-amplification assessment	296	FISH non-amplified 147; FISH amplified 149
Public TCGA-BRCA stress-test set	Independent public boundary stress test after slide-level QC	175 slides / 163 patients	IHC2/FISH- 140 slides; IHC2/FISH+ 35 slides
pCR/HoVerNet response-niche complete-case set	Post-validation biological-context analysis	137	pCR 104; non-pCR 33; restricted to IHC2/FISH-, IHC2/FISH+ and IHC3 cases with ER/PR group, CHERISH output and high-attention HoVerNet features

Supplementary Table 2. Detailed internal development cohort composition.

Source workbook: Table 2; 4 data rows and 7 columns.

validation item	cohort / step	n	permitted use	finalized before next step	external labels used for tuning	manuscript role
Model and classification finalization	Final CHERISH estimator and prespecified classification formulas		Freeze neural-network outputs and deterministic classification layer	Yes	No	Prevents post-hoc threshold or classifier retuning
Clinical-facing external classifications	External HER2 pathway-readout cohort	1,410	Evaluate clinically matched H&E readouts across HER2 clinical questions	Yes	No	External clinical readout alignment
All-centre locked validation	External HER2-positive-spectrum validation cohort	687	Evaluate the finalized three-state endpoint and reliability behaviour	Yes	No	Primary all-centre external validation
IHC2/FISH boundary subset	External IHC2/FISH boundary subset	296	Evaluate the conditional boundary-amplification score and fixed 0.5 operating rule	Yes	No	Boundary-amplification assessment

Supplementary Table 3. Detailed external validation cohort composition.

Source workbook: Table 3; 5 data rows and 5 columns.

validation fold	validation slides	Polarity-anchor labels	Positive-boundary labels	Primary boundary-state labels
Fold 0	427	427 (HER2-negative 279; HER2-positive 148)	148 (IHC3 111; IHC2/FISH+ 37)	220 (IHC2/FISH- 72; IHC2/FISH+ 37; IHC3 111)
Fold 1	427	427 (HER2-negative 279; HER2-positive 148)	148 (IHC3 97; IHC2/FISH+ 51)	224 (IHC2/FISH- 76; IHC2/FISH+ 51; IHC3 97)
Fold 2	427	427 (HER2-negative 279; HER2-positive 148)	148 (IHC3 102; IHC2/FISH+ 46)	234 (IHC2/FISH- 86; IHC2/FISH+ 46; IHC3 102)
Fold 3	426	426 (HER2-negative 279; HER2-positive 147)	147 (IHC3 104; IHC2/FISH+ 43)	226 (IHC2/FISH- 79; IHC2/FISH+ 43; IHC3 104)
Fold 4	426	426 (HER2-negative 279; HER2-positive 147)	147 (IHC3 98; IHC2/FISH+ 49)	236 (IHC2/FISH- 89; IHC2/FISH+ 49; IHC3 98)

Supplementary Table 4. TCGA-BRCA cohort composition.

Source workbook: Table 4; 5 data rows and 11 columns.

validation fold	validation slides	primary boundary-state slides	selected epoch	epochs reviewed	accuracy	macro-F1	balanced accuracy	macro-AUROC	ECE (10 bins)	Brier score
Fold 0	427	220	14	20	0.8727	0.8245	0.8194	0.9490	0.0520	0.2047
Fold 1	427	224	11	17	0.8214	0.7829	0.7801	0.9419	0.0234	0.2257
Fold 2	427	234	13	19	0.8675	0.8183	0.8097	0.9698	0.0406	0.1736
Fold 3	426	226	8	14	0.8097	0.7531	0.7525	0.9407	0.0612	0.2379
Fold 4	426	236	14	20	0.8814	0.8399	0.8279	0.9574	0.0353	0.1796

Supplementary Table 5. Full-spectrum HER2 reference-state distribution.

Source workbook: Table 5; 13 data rows and 5 columns.

summary block	item	value	denominator or counts	note
Pooled primary boundary-state OOF	Denominator	n=1140; IHC2/FISH- 402; IHC2/FISH+ 226; IHC3 512	Five-fold deidentified out-of-fold predictions	Patient-level internal validation folds
Pooled primary boundary-state OOF	Accuracy	0.8509	n=1140; IHC2/FISH- 402; IHC2/FISH+ 226; IHC3 512	Pooled across validation folds
Pooled primary boundary-state OOF	Macro-F1	0.8041	n=1140; IHC2/FISH- 402; IHC2/FISH+ 226; IHC3 512	Unweighted mean across the three boundary states
Pooled primary boundary-state OOF	Balanced accuracy	0.7981	n=1140; IHC2/FISH- 402; IHC2/FISH+ 226; IHC3 512	Mean class recall
Pooled primary boundary-state OOF	Macro-AUROC	0.9512	n=1140; IHC2/FISH- 402; IHC2/FISH+ 226; IHC3 512	One-vs-rest macro average
Pooled primary boundary-state OOF	ECE (10 bins)	0.0270	n=1140; IHC2/FISH- 402; IHC2/FISH+ 226; IHC3 512	Top-label calibration error
Pooled primary boundary-state OOF	Brier score	0.2038	n=1140; IHC2/FISH- 402; IHC2/FISH+ 226; IHC3 512	Multiclass probability score
Pooled confusion matrix	Reference IHC2/FISH-	83.1%	IHC2/FISH- 334; IHC2/FISH+ 53; IHC3 15	reference n=402
Pooled confusion matrix	Reference IHC2/FISH+	57.5%	IHC2/FISH- 86; IHC2/FISH+ 130; IHC3 10	reference n=226
Pooled confusion matrix	Reference IHC3	98.8%	IHC2/FISH- 5; IHC2/FISH+ 1; IHC3 506	reference n=512
Internal OOF bootstrap confidence interval	Polarity anchor AUROC	0.8544 (95% CI 0.8378-0.8706)	n=2133; 5,000 label-stratified bootstrap replicates	Derived summary value reported in this workbook.
Internal OOF bootstrap confidence interval	Positive-boundary module AUROC	0.9824 (95% CI 0.9729-0.9906)	n=738; 5,000 label-stratified bootstrap replicates	Derived summary value reported in this workbook.
Internal OOF bootstrap confidence interval	Primary boundary-state macro-AUROC	0.9512 (95% CI 0.9417-0.9598)	n=1140; 5,000 class-stratified bootstrap replicates	Derived summary value reported in this workbook.

Supplementary Table 6. External three-centre HER2-category performance.

Source workbook: Table 6; 10 data rows and 8 columns.

endpoint	metric	n	point estimate	95% CI	estimate (95% CI)	confusion counts	supporting source
HER2-positive-spectrum boundary-state endpoint	Accuracy	687	0.9068	0.8850-0.9287	0.9068 (0.8850-0.9287)		Derived metric table and bootstrap confidence-interval table
HER2-positive-spectrum boundary-state endpoint	Balanced accuracy	687	0.8909	0.8627-0.9172	0.8909 (0.8627-0.9172)		Derived metric table and bootstrap confidence-interval table
HER2-positive-spectrum boundary-state endpoint	Macro-F1	687	0.8915	0.8649-0.9171	0.8915 (0.8649-0.9171)		Derived metric table and bootstrap confidence-interval table
HER2-positive-spectrum boundary-state endpoint	Macro-AUROC	687	0.9610	0.9471-0.9737	0.9610 (0.9471-0.9737)		Derived metric table and bootstrap confidence-interval table
IHC2/FISH amplification-boundary endpoint	Accuracy	296	0.9493	0.9223-0.9730	0.9493 (0.9223-0.9730)	TN=143, FP=4, FN=11, TP=138	Derived metric table and bootstrap confidence-interval table
IHC2/FISH amplification-boundary endpoint	Balanced accuracy	296	0.9495	0.9225-0.9731	0.9495 (0.9225-0.9731)	TN=143, FP=4, FN=11, TP=138	Derived metric table and bootstrap confidence-interval table
IHC2/FISH amplification-boundary endpoint	F1	296	0.9485	0.9204-0.9728	0.9485 (0.9204-0.9728)	TN=143, FP=4, FN=11, TP=138	Derived metric table and bootstrap confidence-interval table
IHC2/FISH amplification-boundary endpoint	AUROC	296	0.9429	0.9097-0.9715	0.9429 (0.9097-0.9715)	TN=143, FP=4, FN=11, TP=138	Derived metric table and bootstrap confidence-interval table
IHC2/FISH amplification-boundary endpoint	Sensitivity	296	0.9262	0.8792-0.9664	0.9262 (0.8792-0.9664)	TN=143, FP=4, FN=11, TP=138	Derived metric table and bootstrap confidence-interval table
IHC2/FISH amplification-boundary endpoint	Specificity	296	0.9728	0.9456-0.9932	0.9728 (0.9456-0.9932)	TN=143, FP=4, FN=11, TP=138	Derived metric table and bootstrap confidence-interval table

Supplementary Table 7. Reference-state-specific external accuracy.

Source workbook: Table 7; 6 data rows and 7 columns.

clinical question	analysis set	classification used	matched / total	percent	95% CI	CI method/source
IHC 0/1+	Three-centre external HER2 cohort	model-derived negative-side classification	698/723	96.5%	94.9-97.6	Wilson score interval
IHC 2+/FISH–	Three-centre external HER2 cohort	boundary-amplification classification below the prespecified threshold	143/147	97.3%	93.2-98.9	Wilson score interval
IHC 2+/FISH+	Three-centre external HER2 cohort	boundary-amplification classification above the prespecified threshold	138/149	92.6%	87.3-95.8	Wilson score interval
IHC 3+	Three-centre external HER2 cohort	model-derived strong-positive classification	366/391	93.6%	90.7-95.6	Wilson score interval
IHC2/FISH boundary accuracy	IHC2/FISH subset in 1,410-case cohort	boundary-amplification classification at the prespecified 0.5 threshold	281/296	94.9%	92.2-97.3	5,000 label-stratified bootstrap
IHC2/FISH boundary AUC	IHC2/FISH subset in 1,410-case cohort	continuous conditional boundary-amplification score		0.9429	91.0-97.2	5,000 label-stratified bootstrap

Supplementary Table 8. External IHC 2+/FISH-boundary confusion and performance.

Source workbook: Table 8; 12 data rows and 8 columns.

centre	reference state	0/1+ routed	FISH- within IHC2	FISH+ within IHC2	IHC3 recalled	matched / total	row percent
Centre 1	IHC 0/1+	63	1	0	3	63/67	94.0%
Centre 1	IHC 2+/FISH-	0	21	0	0	21/21	100.0%
Centre 1	IHC 2+/FISH+	0	0	15	0	15/15	100.0%
Centre 1	IHC 3+	0	0	0	33	33/33	100.0%
Centre 2	IHC 0/1+	588	0	1	16	588/605	97.2%
Centre 2	IHC 2+/FISH-	0	118	4	0	118/122	96.7%
Centre 2	IHC 2+/FISH+	0	9	120	0	120/129	93.0%
Centre 2	IHC 3+	0	6	17	315	315/338	93.2%
Centre 3	IHC 0/1+	47	1	1	2	47/51	92.2%
Centre 3	IHC 2+/FISH-	0	4	0	0	4/4	100.0%
Centre 3	IHC 2+/FISH+	0	2	3	0	3/5	60.0%
Centre 3	IHC 3+	0	1	1	18	18/20	90.0%

Supplementary Table 9. Internal development performance summary.

Source workbook: Table 9; 1 data rows and 22 columns.

cohort	analysis subset	n	FISH non-amplified	FISH amplified	threshold	TN	FP	FN	TP	accuracy	accuracy 95% CI	AUC	AUC 95% CI	sensitivity	sensitivity 95% CI	specificity	specificity 95% CI	PPV	NPV	F1	F1 95% CI
External three-centre HER2 cohort	IHC2/FISH boundary	296	147	149	0.500	143	4	11	138	0.9493	0.9223-0.9730	0.9429	0.9097-0.9715	0.9262	0.8792-0.9664	0.9728	0.9456-0.9932	0.9718	0.9286	0.9485	0.9204-0.9728

Supplementary Table 10. TCGA-BRCA locked validation summary.

Source workbook: Table 10; 1 data rows and 20 columns.

cohort	slides / patients	FISH non-amplified slides	FISH amplified slides	TN	FP	FN	TP	accuracy	accuracy 95% CI	AUC	AUC 95% CI	sensitivity	sensitivity 95% CI	specificity	specificity 95% CI	PPV	NPV	F1	F1 95% CI
Public TCGA image-level test	175 / 163	140	35	129	11	12	23	0.8686	0.8229-0.9143	0.8571	0.7853-0.9186	0.6571	0.4857-0.8000	0.9214	0.8714-0.9643	0.6765	0.9149	0.6667	0.5397-0.7826

Supplementary Table 11. Calibration and threshold audit summary.

Source workbook: Table 11; 4 data rows and 6 columns.

analysis set	n	centre / source stratum	image source / input type	acquisition metadata visible in SI	transportability interpretation
External HER2 pathway-readout cohort	1,410 (Centre 1 136; Centre 2 1194; Centre 3 80)	Three-centre institutional pathway-readout cohort	Diagnostic H&E WSI feature-bag inputs	Centre-level reference-state composition is visible in Supplementary Table 8; restricted slide-level identifiers and scanner serial numbers are not reported	Clinical pathway-readout denominator reserved from checkpoint selection, recalibration and threshold tuning
All-centre external HER2-positive-spectrum validation/model-comparison cohort	687 (Centre 1 69; Centre 2 589; Centre 3 29)	Three-centre HER2-positive-spectrum endpoint set	Diagnostic H&E WSI feature-bag inputs	Centre-level composition and reliability are visible in Supplementary Fig. 3 and Supplementary Table 8	External validation spans three clinical sites; hardware-domain effects such as scanner vendor or model require prospective device-stratified evaluation
External IHC2/FISH boundary subset	296 (Centre 1 36; Centre 2 251; Centre 3 9)	IHC2 cases with reflex FISH	Diagnostic H&E WSI feature-bag inputs; FISH reference labels are provided only as deidentified derived categories	Centre-level composition and operating counts are visible in Supplementary Fig. 4 and Supplementary Tables 8-9	Fixed-threshold boundary assessment using the prespecified 0.5 conditional score threshold; centre-stratified counts are descriptive because Centre 3 is small
Public TCGA-BRCA stress-test set	175 slides / 163 patients	Public TCGA diagnostic-slide workflow after slide-level QC	Public diagnostic DX slides with matched public clinical annotation	Public-data source and QC denominator are visible in Supplementary Fig. 1 and Supplementary Table 10	Independent public image-level stress test only; not pooled with the institutional validation denominators

Supplementary Table 12. Multiple-instance learning baseline comparison.

Source workbook: Table 12; 6 data rows and 8 columns.

model	internal validation accuracy	internal validation macro-AUROC	external state accuracy	external state macro-AUROC	external IHC2/FISH accuracy	external IHC2/FISH AUROC	external IHC2/FISH F1
Full CHERISH	0.8506	0.9517	0.9068	0.9610	0.9493	0.9429	0.9485
ABMIL	0.8040	0.9354	0.6157	0.8220	0.6824	0.7560	0.7432
CLAM	0.8121	0.9372	0.5604	0.8146	0.6622	0.7423	0.6667
DSMIL	0.7651	0.9066	0.5415	0.7073	0.6318	0.6851	0.5978
DTFD-MIL	0.8289	0.9458	0.6230	0.8134	0.6689	0.7371	0.7012
TransMIL-lite	0.8429	0.9493	0.6317	0.7738	0.6351	0.6929	0.5940

Supplementary Table 13. Paired bootstrap comparison with ABMIL.

Source workbook: Table 13; 5 data rows and 12 columns.

baseline	3-class AUC CHERISH	3-class AUC baseline	3-class AUC difference	3-class difference 95% CI	3-class empirical P	IHC2/FISH AUC CHERISH	IHC2/FISH AUC baseline	IHC2/FISH AUC difference	IHC2/FISH difference 95% CI	IHC2/FISH empirical P	bootstrap resamples
ABMIL	0.9610	0.8220	0.1390	0.1164-0.1624	<0.0004	0.9429	0.7560	0.1869	0.1365-0.2399	<0.0004	5000
CLAM	0.9610	0.8146	0.1464	0.1233-0.1707	<0.0004	0.9429	0.7423	0.2006	0.1500-0.2540	<0.0004	5000
DSMIL	0.9610	0.7073	0.2537	0.2230-0.2860	<0.0004	0.9429	0.6851	0.2578	0.1998-0.3193	<0.0004	5000
DTFD-MIL	0.9610	0.8134	0.1476	0.1230-0.1741	<0.0004	0.9429	0.7371	0.2058	0.1534-0.2607	<0.0004	5000
TransMIL-lite	0.9610	0.7738	0.1872	0.1608-0.2145	<0.0004	0.9429	0.6929	0.2500	0.1937-0.3084	<0.0004	5000

Supplementary Table 14. Hierarchy-consistency ablation and external stability under domain shift.

Source workbook: Table 14; 7 data rows and 14 columns.

model / ablation	ablation definition	polarity anchor	positive-boundary module	clinical base distribution	hierarchy consistency	residual refinement	internal 3-class macro-AUC (5-fold mean)	internal 3-class macro-AUC SD	external 3-class accuracy	external 3-class macro-AUC	external IHC2/FISH accuracy	external IHC2/FISH AUC	external IHC2/FISH F1
Full CHERISH	Full CHERISH ensemble	included	included	included	included	included	0.9517	0.0121	0.9068	0.9610	0.9493	0.9429	0.9485
Boundary-state head only	Primary boundary-state head without auxiliary clinical-pathway components	removed	removed	removed	removed	removed	0.9511	0.0142	0.7132 (Delta -0.194; 95% CI -0.223 to -0.165)	0.8870	0.6588 (Delta -0.291; 95% CI -0.338 to -0.240)	0.7946	0.5738
No hierarchy consistency	Clinical-pathway modules retained but hierarchy-consistency loss removed	included	included	included	removed	included	0.9458	0.0125	0.7613	0.9378	0.6081 (Delta -0.341; 95% CI -0.378 to -0.301)	0.8818	0.3696 (Delta -0.579; 95% CI -0.674 to -0.495)
No residual refinement	Clinical base distribution retained but morphology residual refinement removed	included	included	included	included	removed	0.9288	0.0165	0.6652 (Delta -0.242; 95% CI -0.280 to -0.204)	0.8798	0.7297	0.8144	0.7561
No clinical base distribution	Clinical-pathway modules retained but clinical base distribution removed	included	included	removed	included	included	0.9483	0.0154	0.7453 (Delta -0.162; 95% CI -0.192 to -0.131)	0.8910	0.7230	0.7999 (Delta -0.143; 95% CI -0.189 to -0.099)	0.7172
No positive-boundary module + no hierarchy	Positive-boundary module and hierarchy-consistency loss removed together	included	removed	included	removed	included	0.9459	0.0152	0.8064	0.9279	0.7635	0.8714	0.7445
No positive-boundary module	Positive-boundary module removed while hierarchy-consistency loss retained where defined	included	removed	included	included	included	0.9471	0.0159	0.8108	0.9400	0.7804	0.9273	0.7347

Supplementary Table 15. HoVerNet cellular-feature dictionary.

Source workbook: Table 15; 22 data rows and 6 columns.

metric category	metric or concept	calculation	denominator / unit	interpretation	interpretation boundary
Image-analysis scale	Processed tile coordinate system	HoVerNet and attention-region metrics are computed from processed tile and nuclear coordinates in the source tables. The analysis uses high-attention region tiles generated at the prespecified processed-tile setting.	Analysis-resolution pixels and processed tile coordinates; not a direct micrometre-calibrated pathology measurement.	Gives readers the coordinate frame for graph distances, neighbourhood radii and tile-region geometry.	Pixel-based distances, areas and radii should be interpreted as computational image-analysis units unless an explicit physical calibration is provided in the relevant source table.
Attention-region definition	High-attention tiles	Tiles with within-slide CHERISH attention values at or above the 80th percentile.	All attention tiles from the same slide at the prespecified processed-tile resolution.	Defines the morphologic regions most emphasized by CHERISH within each slide.	Quantiles are computed within slide, so the definition is relative to each WSI and is not an external threshold.
Attention-region definition	Low-attention tiles	Tiles with within-slide CHERISH attention values at or below the 20th percentile.	All attention tiles from the same slide at the prespecified processed-tile resolution.	Provides an internal comparator for within-slide high-versus-low attention tissue-context contrasts.	Low-attention regions are contrast regions, not negative-control tissue by themselves.
HoVerNet cell classes	Epithelial fraction	Epithelial-class nuclei divided by epithelial + lymphocyte + macrophage + neutrophil nuclei.	Tile, high-attention region, low-attention region or whole-slide region, depending on the source table.	H&E-derived epithelial/tumour-compartment proxy.	HoVerNet classes are morphology-derived labels, not multiplex immunophenotyping.
HoVerNet cell classes	Immune-like fraction	Lymphocyte + macrophage + neutrophil nuclei divided by epithelial + lymphocyte + macrophage + neutrophil nuclei.	Same region denominator as the corresponding epithelial fraction.	Compact summary of non-epithelial immune-like morphology inside the analysed region.	Immune-like is a pragmatic H&E-derived compartment label; it should not be interpreted as lineage-resolved immune profiling.
HoVerNet cell classes	Immune:epithelial ratio	Immune-like fraction divided by epithelial fraction. Tile mixed-model count ratios use log2((immune-like count + 0.5)/(epithelial count + 0.5)).	Region-level fractions or tile-level counts, as specified by the source table.	Measures the balance between immune-like and epithelial compartments.	Large values can occur when epithelial counts are low; ratio interpretation should be checked with the paired fraction columns.
Attention-coupled composition	Attention-weighted epithelial or immune-like fraction	Weighted average of per-tile fractions using CHERISH attention values as weights; negative weights, if present, are shifted before weighting.	Valid HoVerNet-mapped tiles with nonzero analysed nuclei.	Asks which cell compartment is enriched in regions receiving higher model attention.	Used for biological interpretation after model finalization.
High-minus-low contrast	Within-slide high-versus-low attention contrast	Metric in high-attention tiles or nuclei minus the matched low-attention value from the same slide.	Within-slide paired high and low attention regions.	Controls for slide-level tissue composition by contrasting model-emphasized and less-emphasized areas in the same WSI.	A high-versus-low value is a relative contrast, not an absolute whole-slide abundance estimate.
Attention spatial patterning	Moran's I / Geary's C	Spatial autocorrelation of tile-level attention or cell-fraction values over 4-neighbour or 8-neighbour tile grids.	Neighbouring attention tiles with finite values.	Higher Moran's I indicates stronger local clustering; Geary's C summarizes local dissimilarity.	Spatial autocorrelation is descriptive; it does not establish a causal spatial mechanism.
Attention spatial patterning	Getis-Ord local hotspot summaries	Local neighbourhood z scores comparing each tile's attention neighbourhood with the slide-wide attention distribution.	Tile plus adjacent neighbourhood in the attention grid.	Summarizes local high-attention and low-attention clustering.	Hotspot z-score fractions are descriptive tissue-map summaries and depend on analysed tile geometry.
Attention-region geometry	Fragmentation index	Number of connected components divided by the number of tiles in a high- or low-attention region using grid connectivity.	Tiles assigned to the specified attention region.	Higher values indicate more spatially fragmented attention fields.	Fragmentation is sensitive to tissue mask geometry and tile coverage; source tables include valid-tile denominators.
Attention-region geometry	Largest component fraction / compactness / elongation	Connected-component, perimeter, convex-hull and covariance summaries of high- or low-attention tile coordinates.	Tiles assigned to the specified attention region.	Describes whether attention concentrates into compact fields or elongated/dispersed fields.	These are morphology-map summaries, not direct cellular phenotypes.
Cell-neighbour graph	Nearest-neighbour distance and heterotypic nearest-neighbour fraction	Nearest-neighbour distances between nuclei; heterotypic fraction is the fraction whose nearest neighbour has a different HoVerNet class.	Nuclei in all, high-attention or low-attention regions.	Measures local cell mixing and spatial proximity across HoVerNet classes.	Pixel-distance units are computational image-analysis units from the processed coordinate system and are not reported as calibrated micrometres.
Cell-neighbour graph	Delaunay edge fractions	Delaunay graph is built from nuclear centroids; edge fractions summarize homotypic, heterotypic and specified epithelial-immune edge classes.	Nuclei with valid centroids; very large graphs were skipped or flagged during QC.	Captures local tissue adjacency beyond single nearest neighbours.	Graph metrics are derived from centroids and do not imply physical membrane contact.
Cell-neighbour graph	Minimum-spanning-tree edge length	Mean, median or percentile of edge lengths from the centroid-based minimum spanning tree.	Nuclear centroids connected by the centroid graph.	Summarizes cellular compaction or dispersion in the analysed region.	Interpret together with cell density and total cell counts.
Spatial cross-type interaction	G-cross / Ripley's L-cross epithelial-lymphocyte metrics	Counts or transformed K-function summaries of neighbouring nuclei of one class around nuclei of another class within fixed radii.	Nuclei of the specified source and target HoVerNet classes.	Measures local epithelial-lymphocyte apposition or mixing.	These statistics are spatial-context descriptors and are not lineage-resolved immune validation.
Spatial colocalization	Morisita-Horn epithelial-lymphocyte tile colocalization	Similarity of epithelial and lymphocyte count distributions across attention tiles.	Tile-level epithelial and lymphocyte count vectors.	Higher values indicate stronger co-occurrence of the two compartments across the analysed tile grid.	Co-occurrence is affected by tile resolution and tissue coverage.
Neighbourhood motifs	Epithelial-only, lymphocyte-only and mixed neighbourhood fractions	For each nucleus, neighbours within a 50-pixel radius are classified by whether they contain epithelial, lymphocyte and/or myeloid-like classes.	Nuclei with valid centroid coordinates.	Provides an interpretable local-niche summary of pure versus mixed cellular neighbourhoods.	The 50-pixel radius is a computational neighbourhood definition, not a histologic structure boundary.

metric category	metric or concept	calculation	denominator / unit	interpretation	interpretation boundary
Nuclear morphology	Nuclear area, eccentricity, solidity and coefficient of variation	Contour- or box-derived nuclear geometry summaries. Area is measured in analysis-resolution pixel area; perimeter and axis lengths are measured in analysis-resolution pixels; eccentricity and solidity are dimensionless. Coefficient of variation is standard deviation divided by mean for the relevant nucleus set.	HoVerNet-segmented nuclei with valid contour or box geometry.	Quantifies nuclear size/shape heterogeneity in epithelial or all-cell compartments.	Chromatin-intensity features were not used because the redistributed HoVerNet outputs do not provide validated intensity channels.
Statistical summary	FISH-positive versus FISH-negative tests	Two-sided Mann-Whitney U tests on slide-level metrics with Benjamini-Hochberg FDR correction.	Slide-level metric values in IHC2/FISH-negative and IHC2/FISH-positive groups.	Supports FISH-boundary feature-evidence summaries.	These tests support biological interpretation after the model and thresholds were finalized.
Statistical summary	Ordered HER2-spectrum trends	Spearman correlation with ordered group score IHC2/FISH-negative = 0, IHC2/FISH-positive = 1 and IHC3 = 2; BH-FDR is reported across tested metrics.	Slide-level metric values across the three HER2-spectrum groups.	Tests whether tissue-context metrics follow an ordered HER2-spectrum gradient.	Ordered trends support tissue-state interpretation, not causal HER2 amplification mechanisms.
Statistical summary	Tile mixed model	Mixed linear model: metric ~ within-slide attention z score x FISH-positive status + random intercept for slide.	Valid HoVerNet-mapped top/bottom attention tiles.	Checks whether tile-level attention-metric coupling persists with slide-level clustering accounted for.	Included as supporting quantitative evidence for readers who want tile-level modelling details.

Supplementary Table 16. HoVerNet whole-slide and attention-ranked cellular architecture summary.

Source workbook: Table 16; 11 data rows and 14 columns.

measurement layer	metric family	metric	region	slides with valid metric	IHC2/FISH- n	IHC2/FISH- median	IHC2/FISH+ n	IHC2/FISH+ median	IHC3 n	IHC3 median	ordered HER2-spectrum Spearman rho	BH-FDR	pattern across HER2 spectrum
nuclear spatial metrics	epithelial composition / epithelial morphology	High-low epithelial fraction		1139	402	0.0165	226	-0.0225	511	-0.0866	-0.3439	5.22e-31	stepwise decrease
nuclear spatial metrics	immune infiltration / immune-epithelial mixing	High-low lymphocyte fraction		1139	402	-0.0168	226	0.0180	511	0.0733	0.3117	2.01e-25	stepwise increase
cell-composition metrics	immune infiltration / immune-epithelial mixing	High-attention immune:epithelial ratio	high	1140	402	0.5477	226	0.8134	512	0.8384	0.2843	3.09e-21	FISH-positive peak
cell-composition metrics	epithelial composition / epithelial morphology	High-attention epithelial fraction	high	1139	402	0.6461	226	0.5515	511	0.5441	-0.2839	3.09e-21	stepwise decrease
cell-composition metrics	immune infiltration / immune-epithelial mixing	High-attention immune-like fraction	high	1139	402	0.3539	226	0.4485	511	0.4559	0.2839	3.09e-21	stepwise increase
tile spatial metrics	epithelial composition / epithelial morphology	Attention-weighted epithelial fraction		1140	402	0.5547	226	0.4754	512	0.4610	-0.2828	3.94e-21	stepwise decrease
tile spatial metrics	immune infiltration / immune-epithelial mixing	Attention-weighted immune-like fraction		1140	402	0.4453	226	0.5246	512	0.5390	0.2828	3.94e-21	stepwise increase
tile spatial metrics	spatial organization / heterogeneity	High-attention fragmentation		1140	402	0.3732	226	0.3853	512	0.4247	0.2670	6.28e-19	stepwise increase
cell-composition metrics	immune infiltration / immune-epithelial mixing	High-attention lymphocyte fraction	high	1139	402	0.3198	226	0.4153	511	0.4150	0.2428	9.74e-16	FISH-positive peak
nuclear spatial metrics	immune infiltration / immune-epithelial mixing	High-low heterotypic edge fraction		1032	394	0.0043	222	0.0092	416	0.0302	0.2191	8.05e-12	stepwise increase
nuclear spatial metrics	immune infiltration / immune-epithelial mixing	All-region heterotypic edge fraction		854	349	0.2743	215	0.2900	290	0.2971	0.1798	4.85e-07	stepwise increase

Supplementary Table 17. HoVerNet high-attention cellular-neighbourhood statistics.

Source workbook: Table 17; 24 data rows and 15 columns.

metric family	metric	interpretive feature family	evidence rank among tested metrics	IHC2/FISH- n	IHC2/FISH+ n	IHC2/FISH- mean	IHC2/FISH+ mean	mean delta FISH+ minus FISH-	IHC2/FISH- median	IHC2/FISH+ median	median delta FISH+ minus FISH-	direction	P value	BH-FDR
epithelial compartment	Whole-slide epithelial fraction	whole-slide cellular composition	whole-slide summary	402	226			-0.0804	0.6516	0.5743	-0.0772	lower in FISH+		3.74e-07
immune infiltration / epithelial-immune interface	Whole-slide immune-like fraction	whole-slide cellular composition	whole-slide summary	402	226			0.0804	0.3484	0.4257	0.0772	higher in FISH+		3.74e-07
immune infiltration / epithelial-immune interface	Whole-slide lymphocyte fraction	whole-slide cellular composition	whole-slide summary	402	226			0.0787	0.3128	0.3879	0.0751	higher in FISH+		1.03e-06
epithelial compartment	Whole-slide epithelial nuclear area CV	whole-slide epithelial nuclear heterogeneity	whole-slide summary	402	226			0.0419	0.5575	0.5963	0.0388	higher in FISH+		1.61e-13
attention composition	High-attention epithelial fraction	composition inside high-attention regions	34	402	226	0.6358	0.5372	-0.0986	0.6461	0.5515	-0.0946	lower in FISH+	1.19e-11	1.78e-10
attention composition	High-attention lymphocyte fraction	composition inside high-attention regions	41	402	226	0.3287	0.4251	0.0964	0.3198	0.4153	0.0955	higher in FISH+	5.65e-11	7.47e-10
attention composition	High-attention immune:epithelial ratio	composition inside high-attention regions	33	402	226	0.7799	1.1886	0.4088	0.5477	0.8134	0.2656	higher in FISH+	1.18e-11	1.78e-10
attention composition	Immune-interface axis score	composition inside high-attention regions	19	402	226	-0.3737	0.0926	0.4663	-0.3802	0.1675	0.5477	higher in FISH+	9.61e-13	2.54e-11
attention-weighted ecology	Attention-weighted immune:epithelial ratio	cellular balance weighted by CHERISH attention	40	402	226	0.7489	1.2175	0.4687	0.5631	0.8818	0.3187	higher in FISH+	1.68e-11	2.38e-10
attention-weighted ecology	Attention-weighted immune-like fraction	cellular balance weighted by CHERISH attention	46	402	226	0.4494	0.5373	0.0879	0.4453	0.5246	0.0793	higher in FISH+	8.16e-11	1.01e-09
attention spatial patterning	Local attention clustering	spatial clustering and autocorrelation of attention	108	402	226	-1.3212	-1.0863	0.2349	-1.3268	-0.8572	0.4696	higher in FISH+	7.88e-08	4.14e-07
attention spatial patterning	Attention spatial autocorrelation	spatial clustering and autocorrelation of attention	170	402	226	0.0996	0.0808	-0.0188	0.0890	0.0572	-0.0318	lower in FISH+	5.17e-06	1.73e-05
attention-region geometry	High-attention elongation	shape and compactness of attention fields	50	402	226	3.4553	2.6497	-0.8056	3.1024	2.3377	-0.7647	lower in FISH+	1.08e-10	1.22e-09
attention-region geometry	Low-attention compactness	shape and compactness of attention fields	94	402	226	0.0078	0.0113	0.0035	0.0064	0.0083	0.0019	higher in FISH+	2.52e-08	1.52e-07
attention-region geometry	High-attention convex-hull fill	shape and compactness of attention fields	143	402	226	0.0605	0.0726	0.0121	0.0559	0.0688	0.0129	higher in FISH+	1.19e-06	4.75e-06
attention-region geometry	High-low centroid separation	shape and compactness of attention fields	119	402	226	14.3072	10.7867	-3.5204	12.6999	8.8524	-3.8476	lower in FISH+	1.81e-07	8.64e-07
neighborhood ecology	Epithelial-only neighborhood fraction	local epithelial or lymphocyte neighborhoods	16	402	226	0.4606	0.3625	-0.0981	0.4550	0.3527	-0.1023	lower in FISH+	6.58e-13	2.34e-11
neighborhood ecology	Lymphocyte-only neighborhood fraction	local epithelial or lymphocyte neighborhoods	59	402	226	0.1685	0.2407	0.0722	0.1526	0.2194	0.0669	higher in FISH+	4.80e-10	4.62e-09
cell-neighbor graph	Heterotypic nearest-neighbor fraction	heterotypic cell-contact structure	138	402	226	0.1867	0.2080	0.0213	0.1884	0.2076	0.0192	higher in FISH+	7.04e-07	2.88e-06
cell-neighbor graph	Epithelial-neutrophil edge fraction	heterotypic cell-contact structure	149	398	224	0.0029	0.0048	0.0019	0.0020	0.0031	0.0011	higher in FISH+	1.72e-06	6.56e-06
nuclear morphology	Epithelial nuclear area CV	epithelial nuclear size and heterogeneity	1	402	226	0.5730	0.6215	0.0485	0.5666	0.6166	0.0500	higher in FISH+	5.19e-18	1.66e-15
nuclear morphology	High-attention nuclear area median	epithelial nuclear size and heterogeneity	5	402	226	280.0908	252.1040	-27.9868	276.2500	245.5000	-30.7500	lower in FISH+	2.93e-16	3.33e-14
nuclear morphology	High-attention nuclear major-axis length median	epithelial nuclear size and heterogeneity	8	402	226	24.4467	23.2050	-1.2417	24.2895	22.9331	-1.3564	lower in FISH+	1.83e-14	1.30e-12
nuclear morphology	High-attention nuclear perimeter median	epithelial nuclear size and heterogeneity	13	402	226	68.8301	65.4251	-3.4049	68.5269	64.9056	-3.6213	lower in FISH+	1.72e-13	7.52e-12

Supplementary Table 18. Spatial transcriptomics dataset-selection and sample metadata.

Source workbook: Table 18; 6 data rows and 6 columns.

screening step	n	screened / included IDs	key count	screening decision	interpretation boundary
HEST v1.2 total entries	1255			Screening universe only.	Metadata screen; not clinical validation.
Human breast-cancer spatial entries	128		Spatial Transcriptomics 108; Xenium 9; Visium 8; Visium HD 3	Candidate pool for public breast-cancer spatial support.	No cross-assay pooling.
Standard Visium human breast-cancer entries	8	TENX68; TENX53; TENX39; TENX24; TENX23; TENX14; TENX13; NCBI776	HER2-positive 2; HER2-negative 2; no explicit HER2 status 4	Screened for explicit HER2-positive metadata and WTA compatibility.	No HER2-negative control use without IHC0/1+ resolution.
Standard Visium with explicit HER2-positive metadata	2	TENX14; TENX13	Public 10x Block A Sections 1 and 2	Included as primary public Visium tissue-context support.	Serial sections; not independent validation.
HER2-positive Xenium candidates	3	NCBI785; NCBI784; NCBI783	Targeted 541-gene in situ panel	Reviewed but not pooled into main spatial analysis.	Not WTA-comparable.
HER2-positive Spatial Transcriptomics candidates	36	not listed in compact table	Median tissue spots 320; range 176-712	Reviewed but not pooled into main spatial analysis.	Older geometry and lower spot coverage.

Supplementary Table 19. Visium spot joining and attention-projection QC.

Source workbook: Table 19; 2 data rows and 11 columns.

public section	sample ID	assay	receptor metadata	HEST tissue spots	analysed Visium spots	attention rows	joined spots	joined / attention rows	top/bottom 20% spots	SI role
(Block A Section 1)	TENX13	Visium	ER positive; PR negative; HER2 positive	3813	3798	3813	3789	3789/3813 (99.4%)	759 high; 758 low	Included public HER2-positive Visium WTA context section
(Block A Section 2)	TENX14	Visium	ER positive; PR negative; HER2 positive	4015	3987	4015	3984	3984/4015 (99.2%)	797 high; 799 low	Included public HER2-positive Visium WTA context section

Supplementary Table 20. Neoadjuvant response-context and receptor-background analyses.

Source workbook: Table 20; 21 data rows and 6 columns.

analysis	group	n	events / total	value	definition
HER2 spectrum pCR	IHC2/FISH-	35	15/35	42.9% (Wilson 95% CI 28.0-59.1)	IHC2 with non-amplified FISH
HER2 spectrum pCR	IHC2/FISH+	22	17/22	77.3% (Wilson 95% CI 56.6-89.9)	IHC2 with amplified FISH
HER2 spectrum pCR	IHC3	80	72/80	90.0% (Wilson 95% CI 81.5-94.8)	IHC3 HER2-positive by immunohistochemistry
ER/PR axis pCR	ER-/PR-	69	62/69	89.9% (Wilson 95% CI 80.5-95.0)	ER/PR receptor-context axis
ER/PR axis pCR	single HR+	38	33/38	86.8% (Wilson 95% CI 72.7-94.2)	ER/PR receptor-context axis
ER/PR axis pCR	ER+/PR+	30	9/30	30.0% (Wilson 95% CI 16.7-47.9)	ER/PR receptor-context axis
CHERISH pCR discrimination	CHERISH morphology score	137	104/137	AUROC 0.8631 (95% CI 0.793-0.920); AUPRC 0.9553 (95% CI 0.932-0.976)	Locked morphology score; no pCR retraining, calibration or threshold tuning; bootstrap confidence intervals.
HoVerNet response-niche pCR discrimination	Response-niche score	137	104/137	AUROC 0.8150; AUPRC 0.9232; response-niche score P=5.4e-08; component FDRs: epithelial-lymphocyte apposition 0.002, cellular compaction 7.5e-04, nuclear-size heterogeneity 5.1e-04	Post-validation composite of epithelial-lymphocyte apposition, local compaction and nuclear heterogeneity; component tests use two-sided Mann-Whitney U with Benjamini-Hochberg FDR across displayed HoVerNet axes.
HER2 spectrum pCR trend	Ordered IHC2/FISH- to IHC2/FISH+ to IHC3	137	104/137	ordered-trend P=9.54e-08	pCR rates 42.9%, 77.3% and 90.0% across IHC2/FISH-, IHC2/FISH+ and IHC3
Incremental response-niche model	CHERISH morphology score plus response-niche score	137	104/137	AUROC 0.863054 to 0.885198; delta AUROC 0.022145; LR P=5.3e-04	Nested logistic comparison against the locked CHERISH morphology-score model; no pCR retraining of CHERISH.
ER/PR axis CHERISH morphology score	ER-/PR-	69		Polarity anchor 0.718; non-negative probability 0.839; IHC3-like probability 0.682	Mean model readouts by receptor-context group; Spearman rho values across the ordered ER/PR axis: polarity anchor -0.3873, non-negative probability -0.3780, IHC3-like probability -0.3662.
ER/PR axis CHERISH morphology score	single HR+	38		Polarity anchor 0.584; non-negative probability 0.761; IHC3-like probability 0.610	Mean model readouts by receptor-context group; source-level values support Fig. 5d heatmap rounding.
ER/PR axis CHERISH morphology score	ER+/PR+	30		Polarity anchor 0.417; non-negative probability 0.452; IHC3-like probability 0.306	Mean model readouts by receptor-context group; source-level values support Fig. 5d heatmap rounding.
ER/PR axis response-niche score	ER-/PR-	69		0.201	Mean value by receptor-context group; Spearman rho -0.2818, P=0.0008.
ER/PR axis response-niche score	single HR+	38		0.035	Mean value by receptor-context group; Spearman rho -0.2818, P=0.0008.
ER/PR axis response-niche score	ER+/PR+	30		-0.506	Mean value by receptor-context group; Spearman rho -0.2818, P=0.0008.
ER/PR axis pCR trend	Ordered ER-/PR- to single HR+ to ER+/PR+	137	104/137	Spearman rho -0.4628; P=1.24e-08	pCR rates 89.9%, 86.8% and 30.0% across the ordered ER/PR receptor-context axis; receptor-context HoVerNet attenuation metrics in Fig. 5e: nuclear regularity FDR=0.004 and reduced epithelial-immune proximity FDR=0.005.
IHC2/FISH-amplified undercall by ER status	ER-positive vs ER-negative	223	62/139 vs 24/84	Fisher exact test P=0.0228	ER-positive and ER-negative undercall rates among true IHC2/FISH-amplified cases in internal development OOF predictions
IHC2/FISH-amplified undercall by ER/PR status	ER+/PR+	76	38/76	50.0%; internal undercall composition ER-/PR- 24/86, single HR+ 24/86, ER+/PR+ 38/86	ER+/PR+ undercall rate and displayed receptor-context composition among true IHC2/FISH-amplified internal OOF undercalled cases.
TCGA IHC2/FISH-amplified undercall by HR status	HR-positive among underestimated slides	12	12/12	100.0%; TCGA undercall composition ER-/PR- 0/12, single HR+ 2/12, ER+/PR+ 10/12	Hormone-receptor-positive composition and displayed receptor-context composition among false-negative undercalled true FISH-amplified TCGA-BRCA slides; the 12 undercalled slides represented 12 unique patients.
External boundary error by HR status	HR-positive vs HR-negative	285	25/221 vs 9/64	11.3% vs 14.1%; external undercalled FISH-amplified composition ER-/PR- 2/11, single HR+ 1/11, ER+/PR+ 8/11	Overall boundary errors in the locked external IHC2/FISH subset with complete ER/PR; no hormone-receptor-positive enrichment was observed. Displayed undercall composition is restricted to false-negative FISH-amplified cases.

Supplementary Table 21. Response-niche covariate-adjusted pCR association models.

Source workbook: Table 21; 5 data rows and 10 columns.

model	n	pCR / non-pCR patients	adjustment variables	response-niche OR	response-niche OR 95% CI	response-niche P value	apparent AUROC	AIC	converged
Unadjusted	137	104/33	response-niche score only	3.4457	1.9518-6.0832	1.99e-05	0.8150	128.4169	True
HER2-spectrum adjusted	137	104/33	HER2-spectrum group	3.0328	1.7096-5.3801	0.0001	0.8488	112.1014	True
HER2-spectrum and ER/PR adjusted	137	104/33	HER2-spectrum group + ER/PR group	2.8420	1.5460-5.2242	0.0008	0.9009	98.1822	True
Stage/biomarker adjusted	136	103/33	HER2-spectrum group + ER/PR group + age + cT + cN + Ki67	3.3906	1.6941-6.7860	0.0006	0.9088	100.8279	True
Treatment-available adjusted	127	98/29	HER2-spectrum group + ER/PR group + age + cT + cN + Ki67 + regimen/cycles/anti-HER2 availability	5.3387	2.0323-14.0242	0.0007	0.9490	83.6528	True

Supplementary Table 22. Receptor-context undercall composition across cohorts.

Source workbook: Table 22; 3 data rows and 7 columns.

cohort	analysis set	true FISH+ called non-amplified	true FISH+ called amplified	positive-boundary support among initially undercalled FISH+ cases	support percent	interpretation
Internal development OOF	primary boundary-state IHC2/FISH boundary assessment	86	140	81/86	94.2%	Internal out-of-fold review-alert summary
Locked external validation	All-centre external IHC2/FISH boundary subset	11	138	7/11	63.6%	Locked external boundary-subset review-alert summary
Public TCGA stress test	TCGA-BRCA diagnostic-slide boundary stress test (n=175)	12	23	12/12	100.0%	Public TCGA boundary stress-test review-alert summary

Supplementary Table 23. Spatial transcriptomic module and cell-state source statistics.

Source workbook: Table 23; 34 data rows and 12 columns.

evidence layer	readout	sample / scope	n high	n low	mean high	mean low	effect estimate	P value	FDR	direction	manuscript use
primary_spatial_readout	ERBB2-neighbourhood epithelial/proliferative module	TENX13 top/bottom 20pct					log2diff=0.141	1.07e-23		high-attention enriched	Fig. 4d-f; main text module contrast
primary_spatial_readout	ERBB2-neighbourhood epithelial/proliferative module	TENX14 top/bottom 20pct					log2diff=0.256	8.34e-70		high-attention enriched	Fig. 4d-f; main text module contrast
targeted_sensitivity	ERBB2 amplicon core	TENX13 top/bottom 20pct					log2diff=0.054	1.17e-05		high-attention enriched	Fig. 4d-f; main text module contrast
targeted_sensitivity	ERBB2 amplicon core	TENX14 top/bottom 20pct					log2diff=0.139	2.27e-20		high-attention enriched	Fig. 4d-f; main text module contrast
targeted_sensitivity	custom ERBB signaling module	TENX13 top/bottom 20pct					log2diff=0.098	1.65e-15		high-attention enriched	Fig. 4d-f; main text module contrast
targeted_sensitivity	custom ERBB signaling module	TENX14 top/bottom 20pct					log2diff=0.143	5.94e-32		high-attention enriched	Fig. 4d-f; main text module contrast
cell2location high-low contrast	Macrophage	high_20pct_vs_low_20pct	1556	1553	0.437615	0.590573	-0.152958	1.588e-47	2.699e-46	low_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	Malignant_LumA_SC	high_20pct_vs_low_20pct	1556	1553	0.78211	0.424524	0.357586	5.256e-42	4.468e-41	high_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	Monocyte_DC	high_20pct_vs_low_20pct	1556	1553	0.415355	0.456734	-0.041379	6.228e-36	3.529e-35	low_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	Malignant_Cycling	high_20pct_vs_low_20pct	1556	1553	0.644428	0.340313	0.304115	8.338e-35	3.544e-34	high_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	Plasmablasts	high_20pct_vs_low_20pct	1556	1553	0.350316	0.515348	-0.165032	2.038e-24	6.931e-24	low_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	CAF_iCAF	high_20pct_vs_low_20pct	1556	1553	0.424229	0.495863	-0.0716344	2.041e-13	5.783e-13	low_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	Myoepithelial	high_20pct_vs_low_20pct	1556	1553	0.553307	0.3781	0.175207	6.081e-11	1.477e-10	high_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	Malignant_Basal_SC	high_20pct_vs_low_20pct	1556	1553	0.575044	0.364716	0.210328	1.722e-06	3.660e-06	high_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	PVL	high_20pct_vs_low_20pct	1556	1553	0.519045	0.423137	0.0959079	1.315e-04	2.484e-04	high_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	CAF_myCAF	high_20pct_vs_low_20pct	1556	1553	0.493501	0.538855	-0.0453539	0.00232005	0.00394408	low_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	Mature_Luminal	high_20pct_vs_low_20pct	1556	1553	0.742943	0.459764	0.283179	0.00358709	0.00554369	high_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	Malignant_LumB_SC	high_20pct_vs_low_20pct	1556	1553	0.694656	0.471882	0.222774	0.0220787	0.0312782	high_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	T_NK	high_20pct_vs_low_20pct	1556	1553	0.565397	0.547016	0.0183801	0.0261818	0.0342377	high_attention_enriched	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	B_cells	high_20pct_vs_low_20pct	1556	1553	0.550615	0.517689	0.0329257	0.435642	0.528994	higher mean in high attention; not FDR-significant	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	Malignant_HER2_SC	high_20pct_vs_low_20pct	1556	1553	0.623052	0.422647	0.200405	0.587673	0.66603	higher mean in high attention; not FDR-significant	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	Luminal_Progenitors	high_20pct_vs_low_20pct	1556	1553	0.587265	0.394971	0.192293	0.794026	0.843653	higher mean in high attention; not FDR-significant	Fig. 4g-j; main text cell-state contrast
cell2location high-low contrast	Endothelial	high_20pct_vs_low_20pct	1556	1553	0.503507	0.47396	0.029547	0.92607	0.92607	higher mean in high attention; not FDR-significant	Fig. 4g-j; main text cell-state contrast
cell2location continuous association	CAF_iCAF	pooled continuous attention					-0.0467216	2.964e-05	6.572e-05	negative with attention	Fig. 4j; continuous attention-cell-state association
cell2location continuous association	CAF_myCAF	pooled continuous attention					-0.0505541	8.615e-07	2.584e-06	negative with attention	Fig. 4j; continuous attention-cell-state association
cell2location continuous association	Macrophage	pooled continuous attention					-0.0730941	1.640e-14	2.788e-13	negative with attention	Fig. 4j; continuous attention-cell-state association
cell2location continuous association	Malignant_Cycling	pooled continuous attention					0.0509251	1.137e-11	8.281e-11	positive with attention	Fig. 4j; continuous attention-cell-state association
cell2location continuous association	Malignant_HER2_SC	pooled continuous attention					0.0350746	5.834e-07	1.860e-06	positive with attention	Fig. 4j; continuous attention-cell-state association
cell2location continuous association	Malignant_LumA_SC	pooled continuous attention					0.0766869	2.454e-17	6.258e-16	positive with attention	Fig. 4j; continuous attention-cell-state association
cell2location continuous association	Monocyte_DC	pooled continuous attention					-0.0794706	1.930e-17	6.258e-16	negative with attention	Fig. 4j; continuous attention-cell-state association
cell2location continuous association	Plasmablasts	pooled continuous attention					-0.062936	2.655e-10	1.354e-09	negative with attention	Fig. 4j; continuous attention-cell-state association
marker/module concordance	Malignant_Cycling vs Proliferation	pooled marker/module correlation					0.593108	0	0	positive concordance	Supplementary Fig. 10; reference-marker support
marker/module concordance	Malignant_HER2_SC vs HER2_amplicon	pooled marker/module correlation					0.510308	0	0	positive concordance	Supplementary Fig. 10; reference-marker support
marker/module concordance	Malignant_LumA_SC vs Luminal	pooled marker/module correlation					0.646605	0	0	positive concordance	Supplementary Fig. 10; reference-marker support