A095 · MESH-HR: Multimodal Histopathology and Somatic Genomics for Continuous Breast Cancer Receptor Phenotyping

Presenter: Shaye Carver — Harvard Medical School

Authors: Shaye Carver, Kodi Taraszka, Intae Moon, Zeyun Lu, Alexander Gusev

Session: Day 2 · Fri Oct 2, 2026 · 2:15–4:15 PM

Breast cancer treatment is guided by estrogen receptor (ER), progesterone receptor (PR), and HER2 status. These biomarkers are measured using immunohistochemical assays, but expression measurements are reduced to discrete categories using predefined thresholds. Tissue sampling, interpretive variability, and intratumoral heterogeneity can affect receptor classification, particularly in borderline tumors. Histopathology and somatic genomic profiling are routinely acquired in oncology and capture complementary dimensions of receptor-associated tumor biology. We developed MESH-HR, a multimodal model integrating H&E whole-slide images with somatic genomic profiles from 1,368 breast cancers in the Dana-Farber PROFILE cohort. MESH-HR achieved held-out AUCs of 0.94 for ER, 0.83 for PR, and 0.96 for HER2, exceeding either unimodal model; ER was driven primarily by morphology, HER2 by somatic genomics, and PR benefited most from fusion. External TCGA-BRCA AUCs were 0.90, 0.83, and 0.82. Because HR-positive breast cancers have a well-established survival advantage over HR-negative disease, we tested whether MESH-HR predictions recapitulated this expected prognostic relationship. Continuous MESH-HR probabilities improved survival discrimination over binary clinical labels (C-index 0.661 versus 0.646). Among discordant tumors, survival aligned more closely with MESH-HR predictions than with clinical labels. Among clinically HR-positive patients receiving endocrine therapy, discordant MESH-HR-negative predictions were associated with worse survival (hazard ratio=2.30, P=0.025). Applied zero-shot to cancer of unknown primary, MESH-HR recovered lineage-consistent, survival-associated receptor phenotypes. Together, these findings show that multimodal histology and somatic genomics can recover receptor-associated phenotypes that generalize across cohorts, capture clinically meaningful heterogeneity beyond binary labels, and extend biomarker inference to settings where receptor testing is unavailable.

MESH-HR: Multimodal Histopathology and Somatic Genomics for Continuous Breast Cancer Receptor Phenotyping

Shaye Carver[1,2,*], Kodi Taraszka[1], Intae Moon[1], Zeyun Lu[1], Alexander Gusev[1]
[1]Dana-Farber Cancer Institute, Boston, MA, USA; [2]Harvard Medical School, Boston, MA, USA
*Presenting author

Background. Breast cancer treatment is guided by estrogen receptor (ER), progesterone receptor (PR), and HER2 status, which determine eligibility for endocrine, CDK4/6, and HER2-directed therapies. These biomarkers are measured using established immunohistochemical assays, with continuous or semi-quantitative expression readouts subsequently categorized using discrete clinical thresholds. This categorization can be further affected by tissue sampling, interpretive variability, and intratumoral heterogeneity, particularly in borderline or spatially heterogeneous tumors. Consequently, a binary clinical label may not fully represent the molecular and morphological state associated with receptor expression. Histopathology and somatic genomics capture complementary dimensions of receptor-associated tumor biology: H&E reflects tissue architecture and cellular context, whereas tumor sequencing captures mutations, copy-number alterations, and mutational processes. We asked whether integrating these routinely acquired modalities could reconstruct receptor status while capturing clinically meaningful variation beyond binary labels, including in breast cancer of known primary and in cancer of unknown primary (CUP), where receptor-specific testing is often unavailable.

Methods. We developed MESH-HR, a multimodal framework using 1,368 breast cancers from the Dana-Farber PROFILE cohort with paired H&E whole-slide images and structured somatic genomic profiles. Clinical ER, PR, and HER2 labels were extracted from pathology reports using a GPT-4-based pipeline and used as reference labels for model training and evaluation. MESH-HR consists of an imaging model that aggregates patch-level H&E representations using attention-based multiple instance learning and a genomic model that applies XGBoost to gene-level mutations, copy-number alterations, and mutational signatures. Receptor-specific fusion combines predictions from the two modalities to generate a continuous probability for each receptor. We evaluated held-out classification performance, external generalization in TCGA-BRCA, prognostic information captured by continuous predictions, outcomes among tumors discordant between clinical and model-derived receptor status, and zero-shot transfer to CUP.

Receptor classification (AUC/F1)

Receptor	H&E only	Genomics only	MESH-HR	TCGA-BRCA
ER	0.93/0.92	0.88/0.89	0.94/0.91	0.90/0.92
PR	0.78/0.71	0.80/0.78	0.83/0.79	0.83/0.80
HER2	0.80/0.50	0.95/0.91	0.96/0.92	0.82/0.69

Internal values are from the held-out test set; TCGA-BRCA values are from external evaluation.

Results. MESH-HR achieved held-out AUCs of 0.94 for ER, 0.83 for PR, and 0.96 for HER2, exceeding either unimodal model across all three receptors. The contribution of each modality was receptor-specific: ER was predominantly encoded in morphology, HER2 in somatic genomics, and PR benefited most from multimodal fusion, indicating that histologic and genomic features capture distinct components of receptor-associated biology. External TCGA-BRCA AUCs were 0.90, 0.83, and 0.82, respectively, supporting generalization across cohorts. Because HR-positive breast cancers (ER- and/or PR-positive) generally have more favorable survival than HR-negative disease, we used overall survival as an orthogonal readout of receptor-associated biology. Continuous MESH-HR probabilities improved survival discrimination relative to binary clinical labels (C-index 0.661 versus 0.646), suggesting that probabilistic predictions preserve clinically relevant variation lost through categorical assignment. Among tumors with discordant clinical and model-predicted HR status, survival aligned more closely with MESH-HR predictions than with the recorded clinical label, suggesting that MESH-HR identifies receptor-associated tumor biology not fully captured by routine classification. Among clinically HR-positive patients receiving endocrine therapy, those with discordant MESH-HR-negative predictions had worse survival than patients with concordant HR-positive predictions (hazard ratio = 2.30, P = 0.025). Applied without retraining to CUP, receptor probabilities were enriched in biologically appropriate independently inferred tissue lineages, and predicted HR positivity was associated with improved survival in CUP (hazard ratio = 0.092, P = 0.003), demonstrating recovery of receptor-associated biology when receptor labels were unavailable.

Conclusions. Histopathology and somatic genomics provide complementary information for receptor prediction, with their relative contributions reflecting receptor-specific biology. MESH-HR extends beyond reconstruction of binary clinical labels by capturing outcome-associated heterogeneity, identifying biologically informative discordance, and generalizing to a label-sparse CUP setting. These results show how existing histology and somatic sequencing can be computationally repurposed to derive additional molecular phenotypes and prioritize confirmatory biomarker testing when conventional receptor measurements are uncertain or absent.