


Artificial Intelligence-Driven Quantitative HER2 Scoring and Spatial Bystander Effect Modeling for ADC Response Stratification in HER2-Low Breast Cancer

Jiayi Liu, Xinyue Zhang 

Department of Breast Surgery, The First Hospital of China Medical University, Shenyang, Liaoning, 110001, People's Republic of China

Correspondence: Xinyue Zhang, Email mszhangxinyue@126.com

Abstract: The rapid development of antibody-drug conjugates (ADCs), particularly trastuzumab deruxtecan (T-DXd), has renewed interest in more refined assessment of low-end HER2 expression in breast cancer. However, substantial inter-observer variability in manual immunohistochemistry, especially around the IHC 0 versus 1+ boundary, remains a major challenge for consistent patient stratification. In this narrative review, we summarize emerging advances in computational pathology, including weakly supervised whole-slide image modeling and quantitative HER2 scoring approaches, and discuss their potential to improve interpretive consistency in this threshold-adjacent setting. We further examine the spatial rationale of the ADC bystander effect as a conceptual basis for integrating quantitative HER2 burden, heterogeneity, and spatial organization into response assessment. Rather than presenting these approaches as clinically established predictive tools, we argue that they should currently be viewed as exploratory and hypothesis-generating frameworks that warrant rigorous validation in outcome-linked, cross-platform, and multi-center studies. Overall, AI-assisted quantitative and spatial pathology may help refine biomarker development for ADC therapy in HER2-low breast cancer, but its clinical utility remains to be established through prospective and analytically robust validation.

Keywords: HER2-low breast cancer, artificial intelligence, computational pathology, antibody-drug conjugates, bystander effect, quantitative continuous scoring, spatial pathology omics

Introduction: Clinical Relevance of Low-End HER2 Expression and the Diagnostic Gap

The emergence of antibody-drug conjugates (ADCs), particularly trastuzumab deruxtecan (T-DXd), has expanded the clinical relevance of low-end HER2 expression in breast cancer. Historically, HER2-targeted treatment primarily benefited patients with HER2-overexpressing or gene-amplified disease, whereas tumors classified as HER2-negative were not considered candidates for HER2-directed therapy.^{1,2} This therapeutic landscape changed with DESTINY-Breast04, which showed that T-DXd significantly improved outcomes versus physician's choice chemotherapy in previously treated HER2-low metastatic breast cancer, including a median progression-free survival (PFS) of 10.1 months versus 5.4 months in the hormone receptor-positive cohort.³ DESTINY-Breast06 further extended the clinical relevance of low-end HER2 expression to an earlier-line hormone receptor-positive metastatic setting that included HER2-low and HER2-ultralow disease.⁴ Importantly, these advances expanded treatment eligibility, but they did not establish HER2-low as a distinct biological subtype.⁵ This distinction reframes HER2-low/ultralow more appropriately as a potential drug-delivery phenotype than as a stable tumor lineage. The clinically relevant question is therefore whether low, heterogeneous, and spatially distributed antigen expression generates sufficient target burden and payload-accessible neighborhoods for ADC activity. In this context, AI-assisted quantitative continuous scoring (QCS) is proposed not to create a new molecular class, but to convert a threshold-based pathology label into auditable variables describing HER2 burden, spatial accessibility, and drug-delivery biology.

At the same time, the current diagnostic framework remains imperfect for this purpose. The ASCO/CAP HER2 immunohistochemistry (IHC) scoring system was originally developed to identify HER2 overexpression or amplification for conventional anti-HER2 therapies, rather than to provide robust quantification at the low-expression end.⁵ Multi-institutional and multi-reader studies have shown substantially poorer agreement for the clinically relevant 0/non-0 and 0/1 + distinctions than for unequivocally positive categories, with persistent discordance also observed in the HER2-ultralow range.^{6–9} These limitations reflect the combined effects of pre-analytic variation, assay and platform differences, and inherent reader subjectivity when interpretation depends on faint and focal membrane staining.^{7–10}

Accordingly, the key knowledge gap is not whether HER2-low/ultralow disease is clinically relevant, but how low, heterogeneous, and spatially distributed HER2 signals can be measured reproducibly and linked to ADC-relevant drug-delivery biology. Recent pathology recommendations have emphasized standardized workflows, consensus review, and quality-control measures to improve consistency in HER2-low and HER2-ultralow interpretation.¹⁰ Within this context, AI-assisted computational pathology is best viewed as an enabling measurement approach: quantitative continuous scoring may help reduce subjectivity and improve reproducibility near threshold-adjacent cases,¹¹ whereas spatial modeling should currently be regarded as a hypothesis-generating strategy for ADC response stratification.¹² However, although early retrospective studies are encouraging, the use of AI-derived continuous or spatial features as candidate biomarkers for ADC response stratification remains exploratory and requires rigorous outcome-linked validation, independent external validation across platforms and centers, and, ideally, prospective confirmation.^{12–14} This review therefore discusses AI-based HER2 quantification and spatial modeling as emerging, hypothesis-generating tools rather than clinically established predictors.

Review Scope and Evidence Framing

This article is a narrative review rather than a systematic review. The literature was identified through focused searches of major biomedical databases, review of reference lists from relevant articles, and targeted selection of studies in four domains: clinical evidence for HER2-low/ultralow ADC therapy, HER2 diagnostic reproducibility and pathology practice, AI-based HER2 scoring and computational pathology, and mechanistic or spatial studies relevant to bystander-effect biology. Priority was given to peer-reviewed original studies, prospective clinical trials, multi-reader or multi-center investigations, and studies directly linked to treatment response or clinically relevant endpoints. Technical proof-of-concept reports and mechanistic studies were included mainly to frame translational hypotheses, but were interpreted separately from outcome-linked clinical evidence. The aim of this review is therefore to distinguish relatively established findings from exploratory concepts and future validation priorities, rather than to present QCS or spatial phenotyping as already validated clinical biomarkers. Distinct from reviews focused mainly on AI-assisted HER2 scoring or digital-pathology classification, this review specifically integrates QCS with spatial bystander-effect modeling to frame low-end HER2 expression as an auditable drug-delivery and spatial-accessibility phenotype for future ADC biomarker validation.

AI-Assisted HER2 Assessment: From Ordinal Classification to Continuous Quantification

Evolution of Deep Learning Techniques in Pathological Image Analysis

Whole-slide images (WSIs) transform the traditional “field-of-view” reading of immunohistochemistry into a globally computable entity, providing the foundational data infrastructure for enhancing the reproducibility of HER2 scoring. Given the enormous data volume of WSIs, contemporary mainstream approaches eschew end-to-end processing of entire slides in favor of tiling combined with hierarchical modeling: phenotypic features are extracted at the patch level and subsequently aggregated to yield slide-level predictions.¹⁵ Weakly supervised multiple instance learning (MIL) is particularly well-suited to real-world clinical data structures, requiring only slide-level labels for training, thereby substantially reducing the cost of per-cell or per-region annotation while attention pooling mechanisms offer interpretable insights into regions of model focus.¹⁶ Building upon these foundations, pathology-oriented weakly supervised and data-efficient training strategies have matured, demonstrating robust cross-domain generalization in independent test cohorts and under varying imaging conditions, thereby propelling computational pathology models toward greater clinical transferability.¹⁷

Importantly, the same weakly supervised WSI paradigm can be repurposed from reproducing ordinal IHC grades to learning outcome-linked representations. When trial or real-world endpoints are available, MIL features can be trained to predict objective response or time-to-event outcomes and, more critically, to test heterogeneity of treatment effect by incorporating treatment-by-feature interaction terms. This reframes WSI modeling as a biomarker discovery engine: the output is no longer a discrete HER2 class but a continuous slide-level embedding that can be evaluated for incremental predictive value beyond conventional IHC strata and clinicopathologic covariates.

Standardization: Technical Strategies for Eliminating Staining Variability

The low-expression range of HER2 IHC is highly sensitive to color and contrast drift. A critical limitation is that currently used chromogenic HER2 IHC assays were optimized and clinically validated mainly for categorical assessment of over-expression or amplification, rather than for strictly linear quantification across the ultralow range.^{5–10} This distinction is essential: AI-based quantification cannot create molecular sensitivity beyond the underlying assay or convert background chromogenic variation into biological signal. AI-based quantification is therefore most useful when it makes visible tumor-cell membrane staining more reproducibly measurable, identifies cases in which signal and noise cannot be confidently separated, and provides auditable outputs for review, calibration, and uncertainty handling. In this setting, AI is best understood as a measurement and quality-control layer for a diagnostic corridor in which manual ordinal interpretation is particularly fragile. Color normalization remains the most widely employed foundational operation, ranging from classical Macenko normalization to structure-preserving stain separation/reconstruction methods, all aimed at maximally decoupling staining style variations from the feature space to direct model attention toward morphological evidence of membrane staining.¹⁸ In the absence of paired data, unsupervised image translation techniques such as CycleGAN may be leveraged for style alignment or data augmentation; however, in tasks like HER2 assessment where boundaries are exceedingly subtle, rigorous quality control and threshold calibration are imperative, as style transfer itself risks altering the visibility of faint membrane signals and introducing systematic bias. Ideally, color normalization and domain adaptation should be framed as auditable preprocessing and quality control steps, with explicit documentation of their impact on interpretive stability near low-expression thresholds.

For clinical translation, within-platform performance is insufficient; a quantitative HER2 algorithm also requires locked-model transportability. A quantitative HER2 algorithm should preserve calibration and boundary-case performance when applied across different assay and scanner ecosystems, including settings involving different staining platforms and Leica/Aperio or comparable WSI scanning workflows. Cross-platform validation should therefore examine site-stratified score distributions, low-expression-band performance, scanner qualification, preprocessing sensitivity, prespecified recalibration rules, version control, and drift monitoring. Without such evidence, a continuous HER2 score may remain center-specific and unsuitable as a transportable candidate predictor of ADC benefit.

Subcellular-Level Quantification: Converting “Faint Membrane Signals” from Subjective Interpretation to Measurable Entities

Controversy in HER2 scoring centers on the distinction between 0 and 1+, which relies on faint, often barely perceptible incomplete membrane staining, inherently limiting inter-observer concordance. Dedicated HER2 algorithms typically employ stepwise, interpretable pipelines: tissue detection and slide quality control, tumor region localization, tumor cell segmentation, followed by cellular-level quantification of membrane staining intensity and completeness, culminating in output of expression distributions across cells or regions. A multi-center, multi-reader validation study demonstrated that automated AI systems can generate whole-slide HER2 scores concordant with ASCO/CAP criteria while significantly improving inter-pathologist consistency and accuracy in the most contentious 0/1+ differentiation, thereby providing more reliable decision support for cases in the threshold-adjacent gray zone.¹¹

Notably, HER2 IHC interpretation is further complicated by interference from non-invasive components. A recent system incorporating ductal carcinoma in situ (DCIS) recognition and exclusion into the workflow—leveraging adjacent sections and annotation data to filter non-target elements—achieved an increase in overall interpretive accuracy from 0.710 to 0.902 across two-stage validation, with a marked reduction in misclassification of 1+ as 0 (from 65/279 to 32/279 cases), a finding of particular relevance to clinical risk mitigation in the HER2-low era.¹⁹

Cell-level quantification enables a direct translation from pathology observations to pharmacologically interpretable variables. Beyond a slide-level ordinal label, the pipeline yields distributions of membrane-positive tumor cells (fraction, intensity percentiles) and, when coupled with cell coordinates, spatial descriptors such as clustering versus admixture of positive cells. These outputs naturally support “bystander-relevant” hypotheses: for a given positive-cell burden, diffuse admixture increases the positive–negative interface and may enlarge the population of antigen-negative cells within effective payload reach, whereas tight clustering may limit interface density. Such variables are explicitly testable against response and PFS in T-DXd-treated cohorts and can be contrasted with conventional IHC strata to quantify incremental predictive signal.

Transition from Classification to Continuous Spectrum: Leveraging Continuous Quantification to Accommodate Evolving Therapeutic Thresholds

The conventional HER2 IHC ordinal system (0/1+/2+/3+) was originally developed to identify patients likely to benefit from prior anti-HER2 therapies—those with unequivocal HER2 positivity—thus ordinal grading near low-expression thresholds inadequately preserves continuous expression information.^{5,8} In contrast, continuous outputs (eg., proportion of membrane-positive tumor cells, H-score, or calibrated model-derived continuous values) are better positioned to interrogate two critical questions: first, whether reproducible expression gradients exist within IHC 0; and second, whether random interpretive error can be mitigated and reproducibility enhanced in threshold-proximal cases. Recent Transformer-based automated HER2 scoring studies have introduced continuous scoring frameworks and reported high inter-model consistency in continuous values, laying technical groundwork for cross-model and cross-center calibration; nonetheless, their clinical relevance awaits validation in cohorts correlated with treatment outcomes or drug exposure-response relationships.²⁰

A key advantage of continuous scoring is that it supports prespecified modeling strategies rather than post-hoc threshold hunting. Practically, QCS outputs can be summarized as a small set of robust features (eg., positive-cell fraction, median/upper-tail intensity, within-slide heterogeneity indices) and evaluated as continuous predictors, with thresholds—if needed for clinical reporting—derived via outcome-linked calibration and decision-analytic criteria rather than purely statistical cutoffs. This shifts the discussion from “what is the correct IHC bin” to “what continuous phenotype best discriminates benefit”, enabling transparent comparisons between categorical ASCO/CAP strata and continuous spatial phenotypes in predicting T-DXd efficacy endpoints. The conceptual transition from ordinal HER2 scoring to quantitative continuous and spatial phenotyping is summarized in [Figure 1](#).

Evidence-Based Assessment of Clinical Validation Studies: Consistency Gains Are Relatively Well-Established, While Outcome Correlations Remain a Gap

From an evidence-based perspective, the principal value of AI integration into HER2 IHC interpretation lies in rendering the gray zone most impactful on treatment access (the HER2 0/1+ boundary) more reproducible and reducing misclassification. A two-arm multi-reader study by Krishnamurthy et al reported improvements in overall concordance from 75.0% to 83.7% and accuracy from 85.3% to 88.0% with AI assistance, with gains predominantly in 0/1+ discrimination (concordance 69.8%→87.4%, accuracy 81.9%→88.8%) and reaching 92.1% concordance on high-confidence reference slides.¹¹ However, analytic performance does not equate to clinical net benefit: systematic reviews and meta-analyses of diagnostic studies indicate pooled sensitivity of 0.97 (95% CI 0.96–0.98) and specificity of 0.82 (95% CI 0.73–0.88) for AI in T-DXd eligibility-oriented classification tasks, yet performance often declines with substantial heterogeneity in external validation and commercial algorithm contexts, underscoring the need for greater data transparency and cross-platform external validation to delineate thresholds, calibration, and transferability limits.¹³

It must be emphasized that enhanced consistency does not imply complete resolution of low-expression interpretive challenges. The Friends of Cancer Research Digital PATH initiative’s cross-comparison of 10 independent AI models across 1124 WSIs revealed a median overall concordance of approximately 65.1% ($\kappa \approx 0.51$), with discrepancies concentrated in low-expression categories such as 1+ and 2+, suggesting that AI does not inherently transcend inherent IHC noise but rather computationalizes and explicates it to some degree.¹⁴ Accordingly, the methodological inflection point is no longer whether AI can improve reproducibility, but whether AI-derived continuous and spatial phenotypes demonstrate predictive utility for

From ordinal HER2 scoring to quantitative continuous and spatial phenotyping

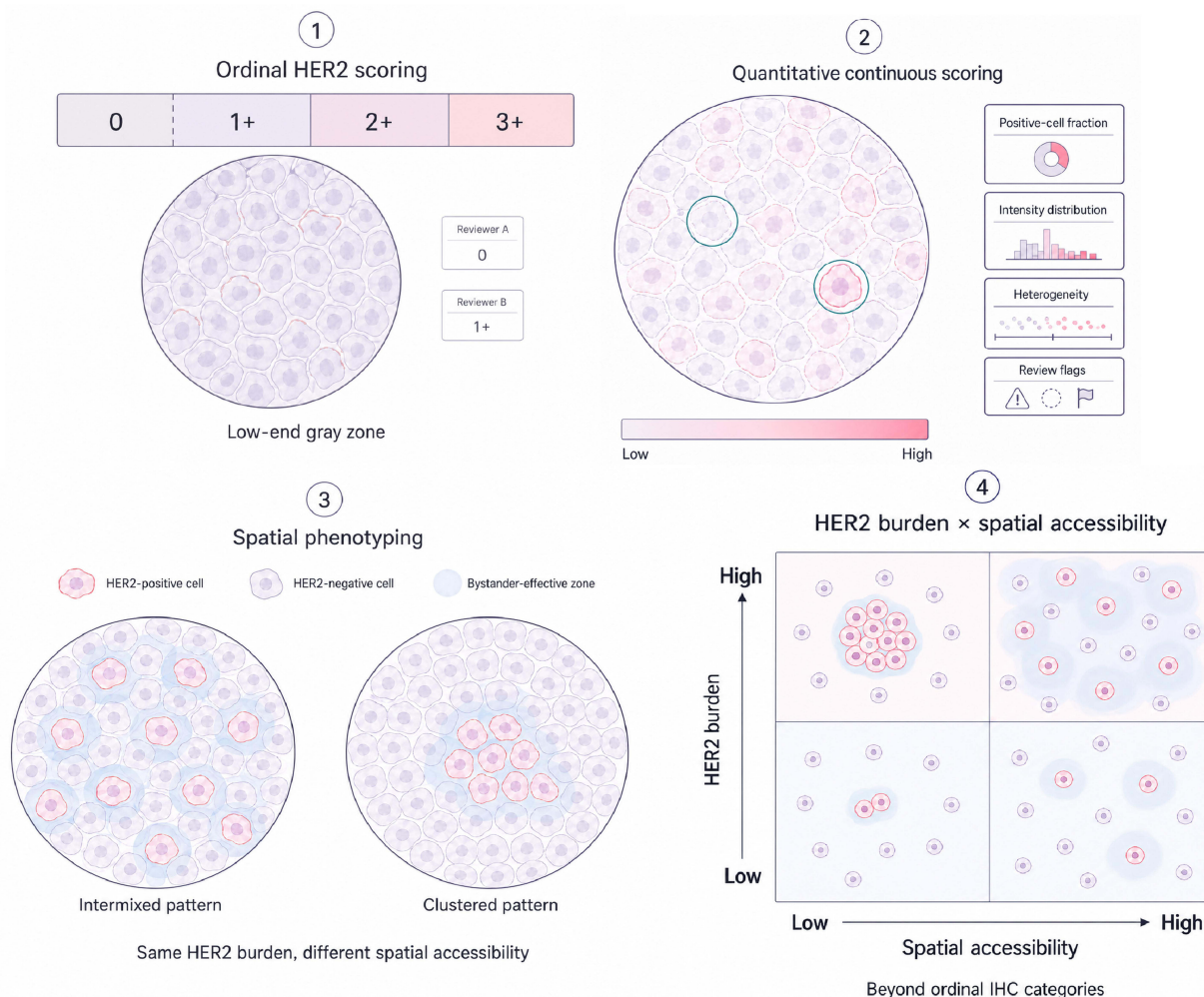


Figure 1 From ordinal HER2 scoring to quantitative continuous and spatial phenotyping. This schematic illustrates the conceptual transition from conventional ordinal HER2 immunohistochemistry scoring to AI-assisted quantitative continuous scoring and spatial phenotyping. Ordinal scoring compresses low-end HER2 expression into discrete categories and is particularly vulnerable to interpretive variability around the 0/1+ boundary. Quantitative continuous scoring provides more granular descriptors, including positive-cell fraction, intensity distribution, heterogeneity, and review flags. Spatial phenotyping further evaluates how HER2-positive and HER2-negative tumor cells are arranged in tissue. Even when overall HER2 burden is similar, intermingled versus clustered patterns may generate different bystander-effective zones and spatial accessibility, supporting the rationale for integrating HER2 burden with spatial organization rather than relying on ordinal IHC categories alone.

T-DXd benefit. This requires study designs and analyses that explicitly distinguish prognostic association from treatment-effect modification: first, by testing treatment-by-biomarker interactions within randomized cohorts; second, by quantifying incremental value over conventional IHC strata; and finally, by demonstrating transportability through external validation.

Bridging Diagnostic Consistency to Therapeutic Prediction: A Treatment-Effect–Aligned Validation Roadmap

Improved reproducibility at the IHC 0/1+ boundary is clinically necessary, but it does not by itself establish that AI-derived quantitative continuous scoring (QCS) or spatial phenotypes are predictive of T-DXd benefit. The central translational question is whether a biomarker modifies the treatment effect of T-DXd versus a relevant comparator on endpoints such as objective response rate (ORR) or PFS, rather than whether it better reproduces ordinal IHC strata. Therefore, single-arm correlations between QCS and response should be framed as hypothesis-generating, whereas decisive evidence should come from randomized trials or prospective–retrospective analyses of archived trial specimens using prespecified classifiers with independent replication.²¹ In such datasets, biomarker-by-treatment interaction should

be tested, and incremental utility beyond conventional IHC categories and clinicopathologic covariates should be quantified using calibration, discrimination, and decision-analytic measures such as net benefit (decision-curve analysis).²² Operationally, QCS pipelines already yield interpretable primitives—positive-cell fraction/intensity distributions, within-slide heterogeneity, and interface/admixture metrics—aligned with the spatial logic suggested by DAISY and by retrospective QCS+proximity enrichment studies.^{12,23} For transportability, continuous outputs should be harmonized across centers through prespecified calibration and reported under TRIPOD+AI to support transparent external validation and drift monitoring.²⁴ Taken together, current evidence more strongly supports improved analytic consistency than validated prediction of differential ADC benefit. At present, QCS should therefore be regarded as an analytically promising but still clinically unproven biomarker framework, pending rigorous outcome-linked validation.

Mechanisms of ADC Drugs and the Spatial Logic of the Bystander Effect

Structural and Pharmacological Characteristics of Next-Generation ADCs

ADCs are not merely a simplistic amalgamation of antibody and chemotherapeutic agent. Their therapeutic window is collectively determined by three key factors: the antibody's binding and internalization behavior toward the target antigen, the *in vivo* stability and cleavage site of the linker, and the physicochemical properties of the payload (particularly membrane permeability and affinity for intracellular targets). These design parameters ultimately translate into two clinically relevant dimensions: the extent to which effective payload exposure within the tumor can encompass heterogeneous clones, and the risk of systemic toxicity arising from off-target release.²⁵

T-DXd exemplifies a design paradigm centered on a high-potency payload, cleavable linker, and elevated drug-to-antibody ratio (DAR). It features an average DAR of approximately 8, with the topoisomerase I inhibitor DXd conjugated via a cleavable peptide linker. In contrast, trastuzumab emtansine (T-DM1) employs a non-cleavable linker to conjugate the microtubule inhibitor DM1 to trastuzumab, with a typical DAR of about 3.5; payload release occurs primarily following internalization into target cells and lysosomal degradation, rendering it more dependent on antigen density and endocytic efficiency.^{26,27} These structural distinctions not only modulate intracellular cytotoxic potency but also critically determine whether the payload can egress from the target cell, thereby providing the mechanistic foundation for the bystander effect.

Bystander Effect: From Intracellular Release to Tissue-Level Coverage

Rather than a purely pharmacologic “feature,” the bystander effect is best viewed as a spatially constrained transport process. For patient stratification, this shifts attention from HER2 burden alone to the spatial accessibility of payload release: how many neighboring cells, including HER2-negative cells, fall within a plausible payload-reach neighborhood. In other words, the translational hypothesis is mechanistically explicit: (QCS-defined target burden and heterogeneity) × (spatial proximity/coverage variables) should better align with the multi-mechanistic activity of T-DXd than ordinal IHC strata alone.

The canonical bystander killing effect underscores propagation at the tissue scale: following intracellular release within antigen-positive cells, a payload with sufficient membrane permeability can diffuse into the extracellular space and be taken up by adjacent antigen-negative cells, ultimately inducing DNA damage or cell death. This provides a plausible geometric “bridge” across patchy antigen expression and clonal admixture typical of solid tumors; conversely, when payloads (or active metabolites) are poorly membrane-permeant, bystander killing is markedly attenuated, leaving antigen-negative “blind spots” more susceptible to therapeutic escape.^{25,27} Importantly, experimental ADC models further support that the physicochemical properties and the amount of released payload are central determinants of bystander killing, with membrane-permeable payloads showing substantially stronger bystander activity than less permeable counterparts.²⁸

Crucially, the bystander effect is not solely governed by payload membrane permeability. Factors such as interstitial fluid pressure within the tumor, stromal barriers, and antigen-mediated binding site barriers coupled with clearance mechanisms collectively shape the “effective diffusion radius” and payload accessibility in tissue.^{29,30} Accordingly, when interpreting the observed overall efficacy superiority of T-DXd over T-DM1 in clinical settings, a more cautious formulation is warranted: this advantage likely stems from the superposition of multiple mechanisms, with the bystander effect representing one biologically plausible contributor rather than the sole attribution.³¹ Indeed, in computational

transport analyses, heterogeneous tumoral distribution was shown to materially impact efficacy; bystander-capable payload diffusion can partially compensate for poor antibody penetration, but direct antigen-positive cell killing remains more efficient than bystander killing—implying that spatial predictors should not be framed as a substitute for target burden, but as a multiplicative modifier that is most informative in heterogeneous, marginal-expression settings.³²

Recent mechanistic investigations have further extended the boundaries of the bystander concept: in certain contexts, T-DXd activity may not fully depend on the canonical sequence of binding → internalization → lysosomal cleavage. Tsao et al suggested that cathepsin L (CTSL) present in the extracellular tumor microenvironment can facilitate linker cleavage and extracellular payload release, with immunomodulatory interactions shaping the overall therapeutic effect.²⁶ Such extracellular release routes are consistent with broader discussions that bystander killing may occur via mechanisms not strictly requiring antigen-dependent internalization, depending on linker–payload chemistry.³³ Nonetheless, at the current stage, CTSL should be positioned as a candidate mechanism and companion indicator rather than an established standalone predictive biomarker; its clinical utility requires prospective validation and careful control for confounding by microenvironmental state.

Spatial Pathology Omics: Why the Bystander Effect Inherently Requires Spatial Metrics

Because bystander killing is fundamentally spatial, the most relevant question is not simply “what fraction of cells are HER2-positive,” but “how positive cells are arranged relative to negative neighbors.” Even with identical HER2-positive fractions, two tumors can have profoundly different bystander-reachable negative-cell volumes: one with diffuse admixture, in which positive cells are embedded among negative counterparts, versus one with clustered segregation, in which positive cells are contiguous with one another. To make this concept operational, we define a “bystander-effective zone” as a candidate neighborhood around a HER2-expressing tumor cell or HER2-positive cluster within which adjacent HER2-negative cells may be exposed to released payload. Its radius should be modeled rather than assumed: a value such as 50 μm can be treated as one prespecified candidate distance, but the biologically relevant range is likely to vary with linker–payload chemistry, tissue architecture, stromal barriers, and local transport conditions. AI can therefore estimate spatial features across multiple prespecified radii, including bystander coverage, defined as the proportion of HER2-negative tumor cells within radius r of HER2-expressing cells; uncovered-negative fraction; positive–negative interface density; neighborhood HER2 burden; and admixture or clustering indices. Their value will depend on reproducibility and on whether they add outcome-linked information beyond conventional IHC category and non-spatial HER2 burden.

Spatial transcriptomics provides a general framework for localizing gene expression on tissue sections, while high-multiplex in situ hybridization methods such as multiplexed error-robust fluorescence in situ hybridization (MERFISH) extend spatial resolution toward single-cell or subcellular multi-gene readouts.³⁴ Multiplexed protein imaging platforms, including CODEX and related technologies, further enable simultaneous quantification of multiple proteins on the same section while preserving proximity relations, thereby more directly delineating cell–cell interaction networks.³⁵ In the present framework, these modalities should be viewed as orthogonal biological anchors rather than immediate replacements for routine HER2 IHC. Spatial transcriptomics can relate AI-derived spatial features to ERBB2 expression gradients and local microenvironmental programs, whereas multiplexed protein imaging can test whether slide-derived variables such as positive-cell fraction, membrane intensity distribution, positive–negative boundary density, neighborhood HER2 burden, and bystander-effective-zone coverage correspond to molecularly defined tissue states. Together, these modalities outline a scalable validation pathway in which routine WSI/IHC provides the clinical substrate, AI extracts quantitative and spatial phenotypes, spatial transcriptomics or multiplex imaging provides biological anchoring, and ADC-treated outcome cohorts determine whether these features are clinically predictive.

Graph Neural Networks and Spatial Proximity Analysis: Transforming Tissue Topology Into Learnable Predictive Variables

Despite their mechanistic appeal, spatial omics assays currently face constraints in throughput, cost, and standardization, limiting near-term scalability for routine clinical decision-making. A pragmatic translational route is therefore to extract spatial structure from conventional H&E/IHC digital slides at scale. This begins with cell detection and phenotypic classification, proceeds to constructing cellular graphs (cells as nodes, proximity relations as edges), and then uses graph

neural networks (GNNs) to learn mappings between local topology and slide- or patient-level outcomes. Feasibility for capturing clinically and molecularly relevant topology has been demonstrated in colorectal cancer pathology: Ding et al converted whole-slide scans into graph-structured data and achieved molecular profile prediction with multi-cohort external validation, supporting the premise that graph models can encode tissue architectures that patch-based classifiers may miss.³⁶ In the HER2-low/ADC setting, the priority lies not in algorithmic novelty, but in explicitly defining biologically aligned spatial primitives—boundary density, admixture, neighborhood burden, and coverage surrogates—that correspond to pharmacologic reach and can be learned, audited, and transported across centers.

From Expression Level to Spatial Accessibility: An Evidence-Structured Rationale for Spatial Predictors of T-DXd Benefit

A central observation across contemporary trials is that “HER2-low” functions primarily as an eligibility category rather than a complete predictor of benefit magnitude. Exploratory biomarker analyses from ongoing programs have reported no statistically significant interaction for treatment benefit by baseline HER2 IHC status or HER2 gene-expression level, motivating predictors that encode tissue accessibility and heterogeneity beyond ordinal intensity strata.³⁷

Mechanistically, the bystander effect implies that spatial arrangement—where HER2-positive cells sit relative to HER2-negative neighbors—should influence effective payload coverage. The DAISY trial, spanning HER2-expressing through HER2 IHC 0 disease, documented meaningful activity even in the non-expressing cohort and highlighted HER2 expression and its spatial distribution as key unresolved determinants of efficacy and resistance.²³ Although DAISY was not a computational pathology study, it strengthens the biological plausibility that “how HER2 is distributed” can be outcome-relevant even when categorical IHC labels appear similar.

Against this clinical-to-translational backdrop, a concrete computational pathology proof-of-concept has already operationalized this logic retrospectively. In a Phase 1 DS-8201a-treated cohort, Kapil et al reported that a HER2 QCS system—particularly when combined with spatial proximity metrics—enriched for responders among conventionally HER2-negative/IHC 0 patients, demonstrating a median PFS of 14.8 months in the QCS+spatially selected subset versus 8.6 months in the overall HER2-negative cohort, and capturing 76.4% of responders compared with 56.9% by conventional IHC scoring.¹² Notably, the conceptual core is transferable: QCS provides a continuous measure of target burden and heterogeneity, while spatial proximity features approximate the neighborhood coverage needed for bystander-mediated reach. While this evidence remains retrospective and hypothesis-generating—requiring prospective, cross-platform validation—it offers an operational template for converting “bystander effect as a spatial process” into measurable, falsifiable, outcome-linked spatial phenotypes suitable for pre-specified testing in trials. However, biological plausibility should not be conflated with clinical proof. Current support for spatial phenotyping remains largely mechanistic, retrospective, or proof-of-concept, rather than derived from prespecified biomarker analyses within randomized therapeutic comparisons. Accordingly, spatial phenotyping should currently be viewed as a hypothesis-generating extension of HER2 assessment, with its clinical utility contingent on reproducible feature definition and prospective outcome-linked validation.

Clinical Translation: Toward Trial-Grade and Auditable QCS×Spatial Phenotyping

Workflow Integration: Prioritizing Stable Quantitation Over Binary Diagnosis

A practical way to frame workflow integration is to begin with the clinical decision the system is intended to support. In the HER2-low/ultralow era, treatment access to antibody–drug conjugates can hinge on a narrow interpretive corridor around the IHC 0/1+ boundary, where faint, focal membrane staining, specimen limitations, and inter-laboratory variability collectively amplify misclassification risk. This is not a niche edge case: the clinical stakes have increased as evidence continues to expand for trastuzumab deruxtecan across low and ultralow expression strata. Accordingly, the central translational promise of QCS and spatial phenotyping is not marginal gains in retrospective discrimination, but the reliable production of cross-center comparable continuous and spatial variables that can be endpoint-linked, audited, and—if validated—used for therapy stratification.

Translation therefore requires treating the algorithm as one component of an end-to-end measurement system. The operational objective should be an auditable report that quantifies HER2 signal on a continuous scale, characterizes heterogeneity and positive–negative admixture, and explicitly communicates uncertainty when the specimen or signal is intrinsically fragile. A pragmatic deployment pattern is “triage plus evidence overlay” within routine digital pathology: following pre-analytic control and whole-slide scanning, the system performs slide-level QC and returns a prespecified set of variables rather than a single categorical label. At minimum, these variables should include a calibrated continuous summary of HER2 burden (eg., positive-cell fraction and intensity distribution summaries), quantitative descriptors of heterogeneity (eg., regional variability or clustering/admixture indices), and spatial accessibility surrogates aligned with bystander logic (eg., positive–negative boundary density, neighborhood positive burden within radius r , mixing indices, and coverage surrogates estimating the fraction of negative cells within bystander-reach neighborhoods). Each output should be coupled to an audit trail—traceable ROIs, cell/patch overlays supporting computation, and uncertainty flags enriched for boundary-adjacent cases, low tumor cellularity, and staining instability—so that review effort is concentrated where consequences are largest.

Clear responsibility boundaries are essential. Pathologists retain interpretive authority; the system should prioritize efficient audit rather than replace expert judgment. The laboratory owns pre-analytics and staining standard operating procedures (SOPs); the digital pathology program owns scanner qualification and image QC; the AI system is accountable for versioned, reproducible outputs; and the clinical team uses the report as decision support, including escalation to repeat testing or multidisciplinary adjudication when uncertainty is high.³⁸ The proposed AI-assisted HER2 QCS workflow and translational validation pathway are summarized in [Figure 2](#).

Prerequisites for Clinical Utility: Verification, Calibration, and Lifecycle QA

If QCS×spatial phenotyping is advanced as a predictor (or effect modifier) of trastuzumab deruxtecan benefit, “one-time validation” is not sufficient. The translational claim is defensible only if three conditions are met, and the evidence package is reported with contemporary AI-specific transparency standards.

First, features, thresholds, and calibration strategies must be prespecified, and continuous outputs must be demonstrably comparable across centers. Without calibration, a numeric score can become “center-specific,” reflecting local staining, scanning, and preprocessing idiosyncrasies rather than biology. A trial-grade calibration plan must inherently examine site-stratified distributions and quantify signal sensitivity to analytic choices, subsequently applying a prespecified mapping to preserve interpretability across platforms. The same discipline is required for spatial variables: if spatial metrics depend on cell-detection thresholds or membrane-signal segmentation behavior, these parameters must be fixed, documented, stress-tested, and shown stable in the low-expression regime. Importantly, recent multi-model evaluations indicate that algorithmic agreement itself can vary materially across AI systems, reinforcing the need to treat “model choice and versioning” as part of measurement uncertainty rather than as an implementation detail.

Second, external validation must match the intended clinical claim. For diagnostic assistance, concordance and repeatability are central; for therapeutic prediction, validation must be endpoint-linked and aligned to treatment effect rather than label agreement alone. Evidence should be structured as analytic validity (measurement accuracy and repeatability, especially at the 0/1+ and 1+/2+ boundaries), clinical validity (association with response or time-to-event endpoints, ideally via prospective–retrospective analyses on trial specimens with prespecified plans), and—when feasible—clinical utility (incremental net benefit beyond conventional IHC strata and clinicopathologic covariates). Critically, predictive claims require designs capable of separating prognostic association from treatment-effect modification, such as biomarker-by-treatment interaction testing in randomized or otherwise well-annotated comparative cohorts. Absent such evidence, QCS×spatial phenotyping should be framed as hypothesis-generating rather than as a decision rule.

Third, drift monitoring must be operationalized as an auditable SOP with explicit triggers for re-validation. Drift may arise from scanner upgrades, reagent or protocol changes, case-mix shifts, or software updates. A defensible lifecycle plan includes periodic sampling re-reviews (oversampling boundary-adjacent and low-cellularity cases), boundary-focused audits, and predefined criteria that mandate re-validation when material pipeline components change.³⁹ For AI/ML-enabled device software functions, contemporary regulatory discussions explicitly emphasize predetermined change control and lifecycle governance; in practice, this means writing down what is allowed to change, what must

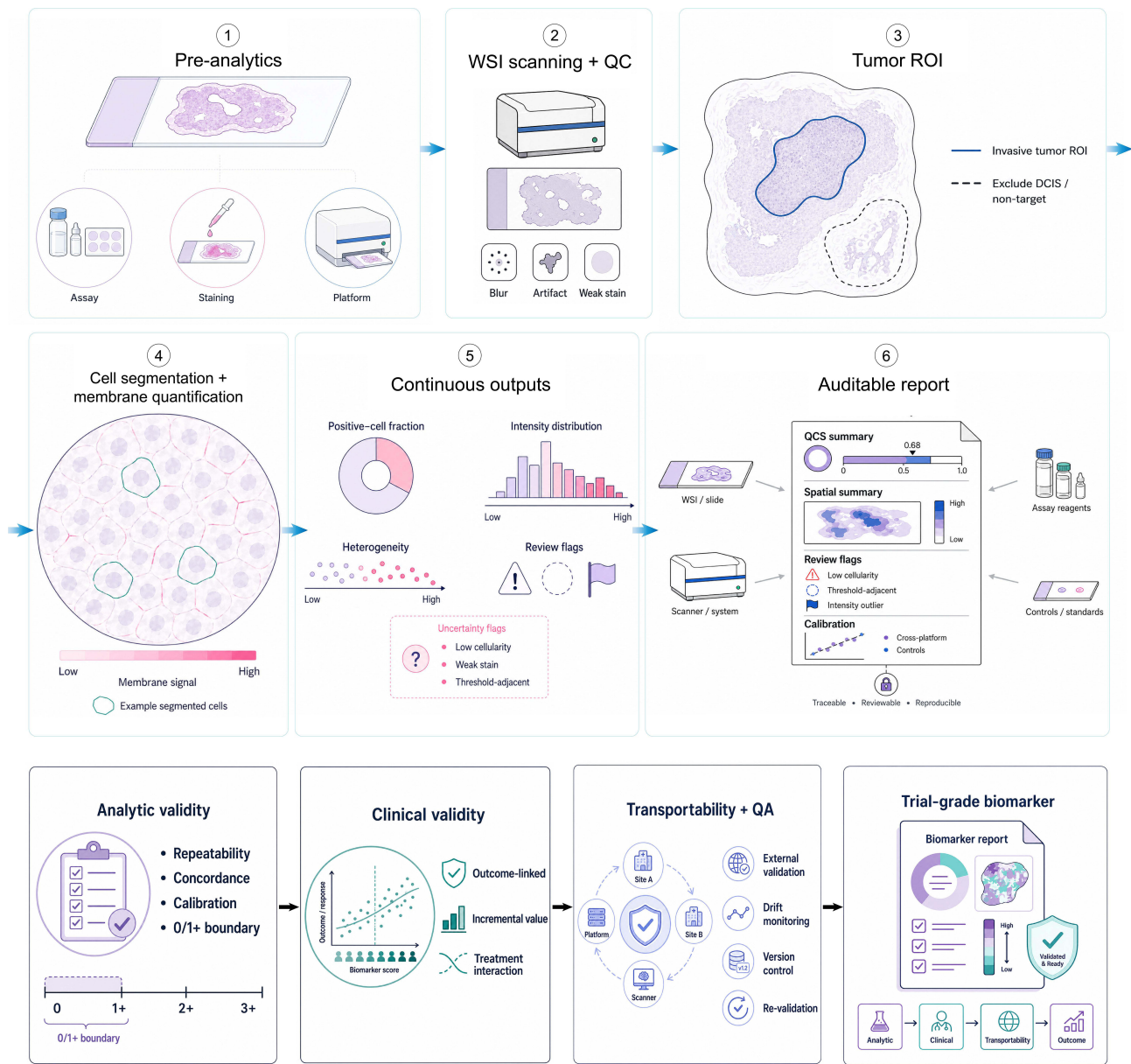


Figure 2 AI-assisted HER2 quantitative continuous scoring workflow and translational validation pathway. This schematic summarizes an end-to-end AI-assisted HER2 QCS workflow and its validation requirements for potential ADC biomarker development. The workflow begins with pre-analytic control, whole-slide image scanning and quality control, tumor-region selection, cell segmentation, and membrane-signal quantification. The system then generates continuous outputs, including HER2-positive cell fraction, intensity distribution, heterogeneity metrics, uncertainty flags, and an auditable report linking quantified results to traceable image evidence and calibration information. For translational use, analytic validity, clinical validity, transportability, lifecycle quality assurance, and trial-grade biomarker development are required before AI-derived QCS or spatial phenotypes can be considered clinically actionable predictors of ADC benefit.

remain locked, how updates are validated, and what performance bounds must be preserved, so that iterative updates do not silently erode cross-center comparability of continuous and spatial outputs.⁴⁰

Overall, QCS×spatial phenotypes can function as transportable biomarkers only when outputs are prespecified, calibrated, externally validated for the intended clinical claim, and continuously monitored with enforceable re-validation triggers.⁴¹

Sampling, Re-Testing, and Heterogeneity: Protecting the Biomarker from Tissue-Level Fragility

Spatial phenotyping is only as credible as the specimen. HER2-low and ultralow disease is frequently heterogeneous and characterized by faint, focal membrane staining; localized biopsies may under-sample the interfaces where bystander-

relevant geometry is expressed. This creates a practical vulnerability: even a well-calibrated algorithm cannot compensate for non-representative tissue. Two implications follow.

First, QCS×spatial outputs should be treated as specimen-conditional measurements, and reports should explicitly include adequacy descriptors (tumor cellularity, tissue area, staining quality, and heterogeneity-related uncertainty flags) to inform whether repeat testing is warranted. Second, repeat testing should be tied to decision pressure and biological evolution. When eligibility or therapeutic choice hinges on a narrow boundary, escalation to repeat testing or targeted re-biopsy can reduce decision error more effectively than further refinement of a single archived specimen—particularly when receptor expression and spatial organization can shift over time and across sites. Operationally, repeat testing can be prioritized when the initial specimen is low-adequacy, when treatment decisions hinge on the 0/1+ corridor, or where clinical evolution suggests discordance between archived and current disease. Finally, heterogeneity should be elevated from a descriptive add-on to a prespecified component of the phenotype: rather than reporting only an average continuous score, the system should provide heterogeneity summaries (eg., distribution tails, regional variability, clustering/admixture indices) that can be prospectively tested against outcome heterogeneity.

Health Economics: Aligning Reimbursement Models with Clinical Utility

Economic arguments are most compelling when tied to measurable levers that QCS×spatial phenotyping is expected to move. Relevant inputs include scanning and storage, software licensing and maintenance, personnel training, calibration and re-validation exercises, and ongoing quality monitoring.⁴² Relevant outputs should be limited to consequences plausibly concentrated in the 0/1+ corridor: changes in boundary-case adjudication and second-read burden, reductions in repeat testing and controversy consultations, shifts in access attributable to more reproducible low-end stratification, and net changes in turnaround time. A minimal reimbursement narrative can then be constructed from site-specific volumes and the cost trade-off between missed eligibility and unnecessary escalation, avoiding generic claims of cost-effectiveness.

Challenges and Future Directions

Technical Bottlenecks and Resolution Pathways

Data availability and annotation remain the primary bottlenecks constraining the upper performance limits of AI pathology models. This is particularly acute for HER2: conventional IHC scoring exhibits systematic inconsistency in the low-expression range, such that the ostensibly established “ground truth” itself drifts with reader, center, and threshold interpretation; in multi-institutional, multi-reader real-world evaluations, concordance for critical strata such as 0/1+ and 1+/2+ is markedly lower than for high-expression intervals like 3+, implying that models trained on single-reader labels risk entrenching subjective biases as algorithmic regularities.^{6,8} Consequently, a more rational approach for training and validation entails constructing reference standards via multi-expert consensus or multi-reader statistical frameworks, designating “threshold-adjacent cases” as discrete quality control and incremental learning sets; where in situ hybridization (ISH)/fluorescence in situ hybridization (FISH) or other molecular assay results are available, these may be leveraged for weakly or semi-supervised constraints to mitigate noise propagation from reliance solely on ordinal IHC labels.

Multi-center heterogeneity constitutes the second decisive barrier. Variations in antibody lots, staining platforms, scanners, and resolution engender substantial domain shifts, causing models performing well at center A to degrade at center B. Stain standardization and color normalization remain foundational for cross-center deployment (eg., classical stain normalization methods serving as preprocessing baselines), upon which federated learning offers a pragmatically viable engineering pathway for inter-institutional collaboration: model updates are accomplished without central aggregation of raw WSIs, thereby lowering data-sharing barriers while better aligning with privacy and regulatory constraints.⁴³

Core algorithmic challenges center on interpretability and robustness. Pathologists require visibility into “what the model bases its decision on and why,” rather than merely a score; decomposing model outputs into cellular- or patch-level evidence (eg., overlaid positive-cell annotations, counterfactual explanations, or critical region highlighting) facilitates transformation of AI from opaque conclusions into auditable decision support.⁴⁴ Concurrently, distribution drift and performance decay represent enduring risks: shifts in case mix, laboratory workflows, or equipment upgrades may precipitate inconsistencies for the same model across time or centers. Deployment, therefore, should not be

construed as a one-time rollout but integrated into ongoing quality systems: establishing local periodic sampling and re-validation protocols, with model modifications and updates governed by auditable change management frameworks.⁴⁵

Priority Agenda for Clinical Validation

Building on the mechanistic framework in Standardization: Technical Strategies for Eliminating Staining Variability–3.5 (continuous HER2 burden by QCS coupled with spatial accessibility/coverage variables), the evidentiary pathway must progress from concordance studies to outcome-associated investigations. Contemporary multi-reader evaluations suggest AI can improve reproducibility in the low-expression interval, but concordance alone is insufficient to justify therapeutic re-stratification.^{11,46} Next-phase designs with higher informational yield should prioritize prospective–retrospective correlative studies embedded in ADC trials to test incremental associations with outcomes. These efforts must be complemented by prospective cohorts enriched for threshold-adjacent cases to validate responder identification, alongside multi-center assessments ensuring that spatial predictors remain robust against workflow variations.

Beyond HER2: Methodological Transferability and Precision Stratification for ADCs

The value of the HER2 paradigm extends beyond resolving a single scoring challenge to furnishing a transferable methodological toolkit: cellular/patch-level segmentation and quantification, weakly supervised WSI modeling, spatial graph-structured feature extraction, and cross-center quality control. These capabilities are not proprietary to HER2. For instance, the clinical benefit of TROP2-directed ADCs in metastatic triple-negative breast cancer illustrates that, once efficacy is established, constraints on accessibility often shift toward “how to more reliably select potential beneficiaries”⁴⁷ Accordingly, directions of broader generalizability lie in expanding from “single-target intensity” to composite phenotypes integrating “target burden + heterogeneity + spatial accessibility,” with validation via stratification or enrichment designs in clinical trials. The confluence of spatial omics and digital pathology provides conceptual and technical scaffolding: spatial transcriptomics and analogous modalities have demonstrated that *in situ* spatial information can reshape understanding of tumor microenvironment and cell-cell interactions,³⁴ while graph representation learning and cross-level molecular profile prediction on routine sections have evidenced feasibility in linking spatial structure to tissue morphology and molecular/prognostic outcomes^{36,48}—affording an engineerable entry point for testing the spatial logic of ADC bystander effects.

Ethics, Equity, and Accessibility

Risks in AI pathology derive not solely from technical misalignment but also from equity and governance lacunae. Real-world investigations have shown that medical algorithms may, under ostensibly neutral inputs, systematically amplify population disparities, thereby altering resource allocation and treatment access. Thus, model development and deployment phases should incorporate “population representativeness, subgroup performance, and bias auditing” into metric hierarchies of equivalent primacy to overall area under the curve (AUC); for resource-constrained settings, pragmatic trade-offs among computational power, storage, and workflow costs are requisite, promoting lightweight inference, tiered deployment, and remote collaboration to preclude concentration of technological dividends in high-resource centers alone.

Conclusion

The HER2-low/ultralow era has increased the importance of more reproducible assessment at the low-expression end, where conventional IHC remains vulnerable to interpretive variability. Current evidence suggests that AI-assisted computational pathology may improve analytic consistency and provide auditable quantitative descriptors of HER2 burden, heterogeneity, and tissue organization. However, the role of QCS and spatial phenotypes as predictive biomarkers of ADC benefit has not yet been established. Most supporting data remain retrospective, exploratory, or proof-of-concept, and further work is needed to define robust features, ensure cross-platform comparability, and validate these approaches in outcome-linked and independently replicated cohorts. At present, these methods are best regarded as candidate biomarker frameworks for further study rather than tools ready for routine clinical decision-making.

Declaration of Generative AI Use

Authors declare no AI use during the preparation of this work.

Abbreviations

ADCs, antibody-drug conjugates; T-DXd, trastuzumab deruxtecan; QCS, quantitative continuous scoring; AI, artificial intelligence; PFS, progression-free survival; IHC, immunohistochemistry; WSIs, Whole-slide images; MIL, multiple instance learning; DCIS, ductal carcinoma in situ; DAR, drug-to-antibody ratio; T-DM1, trastuzumab emtansine; CTSL, cathepsin L; GNNs, graph neural networks; ISH, in situ hybridization; FISH, fluorescence in situ hybridization.

Data Sharing Statement

No new datasets were generated or analyzed for this narrative review.

Ethics Statement

Ethical approval was not required because this narrative review analyzed previously published literature and did not involve human participants, animals, identifiable personal data, or biological specimens.

Author Contributions

All authors contributed to data analysis, drafting or revising the article, have agreed on the journal to which the article will be submitted, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Funding

There is no funding to report.

Disclosure

The authors declare no competing interests.

References

- Dai LJ, Li YW, Ma D, Shao ZM, Jiang YZ. Next-generation antibody-drug conjugates revolutionize the precise classification and treatment of HER2-expressing breast cancer. *Cancer Biol Med*. 2023;20(10):689–694. doi:10.20892/j.issn.2095-3941.2023.0286
- Qu F, Lu R, Liu Q, et al. Antibody-drug conjugates transform the outcome of individuals with low-HER2-expression advanced breast cancer. *Cancer*. 2024;130(S8):1392–1402. doi:10.1002/encr.35205
- Modi S, Jacot W, Yamashita T, et al. Trastuzumab deruxtecan in previously treated HER2-Low advanced breast cancer. *N Engl J Med*. 2022;387(1):9–20. doi:10.1056/NEJMoa2203690
- Bardia A, Hu X, Dent R, et al. Trastuzumab deruxtecan after endocrine therapy in metastatic breast cancer. *N Engl J Med*. 2024;391(22):2110–2122. doi:10.1056/NEJMoa2407086
- Wolff AC, Somerfield MR, Dowsett M, et al. Human epidermal growth factor receptor 2 testing in breast cancer: ASCO-College of American Pathologists Guideline Update. *J Clin Oncol*. 2023;41(22):3867–3872. doi:10.1200/jco.22.02864
- Robbins CJ, Fernandez AI, Han G, et al. Multi-institutional assessment of pathologist scoring HER2 immunohistochemistry. *Mod Pathol*. 2023;36(1):100032. doi:10.1016/j.modpat.2022.100032
- Fernandez AI, Liu M, Bellizzi A, et al. Examination of low ERBB2 protein expression in breast cancer tissue. *JAMA Oncol*. 2022;8(4):1–4. doi:10.1001/jamaoncol.2021.7239
- Baez-Navarro X, van Bockstal MR, Nawawi D, et al. Interobserver variation in the assessment of immunohistochemistry expression levels in HER2-Negative breast cancer: can we improve the identification of low levels of HER2 expression by adjusting the criteria? An International Interobserver Study. *Mod Pathol*. 2023;36(1):100009. doi:10.1016/j.modpat.2022.100009
- Wu S, Shang J, Li Z, et al. Interobserver consistency and diagnostic challenges in HER2-ultralow breast cancer: a multicenter study. *ESMO Open*. 2025;10(2):104127. doi:10.1016/j.esmoop.2024.104127
- Tozbikian G, Bui MM, Hicks DG, et al. Best practices for achieving consensus in HER2-low expression in breast cancer: current perspectives from practising pathologists. *Histopathology*. 2024;85(3):489–502. doi:10.1111/his.15275
- Krishnamurthy S, Schnitt SJ, Vincent-Salomon A, et al. Fully automated artificial intelligence solution for human epidermal growth factor receptor 2 immunohistochemistry scoring in breast cancer: a Multireader Study. *JCO Precis Oncol*. 2024;8e2400353. doi:10.1200/po.24.00353
- Kapil A, Spitzmüller A, Brieu N, et al. HER2 quantitative continuous scoring for accurate patient selection in HER2 negative trastuzumab deruxtecan treated breast cancer. *Sci Rep*. 2024;14(1):12129. doi:10.1038/s41598-024-61957-9
- Albuquerque DAN, Vianna MT, Sampaio LAF, Vasiliu A, Neves Filho EHC. Systematic review and meta-analysis of artificial intelligence in classifying HER2 status in breast cancer immunohistochemistry. *NPJ Digit Med*. 2025;8(1):144. doi:10.1038/s41746-025-01483-8
- McKelvey B, Torres-Saavedra PA, Li J, et al. Agreement across 10 artificial intelligence models in assessing human epidermal growth factor receptor 2 (HER2) expression in breast cancer whole-slide images. *Mod Pathol*. 2026;39(2). doi:10.1016/j.modpat.2025.100944
- Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. 2019;25(8):1301–1309. doi:10.1038/s41591-019-0508-1

16. Ilse M, Tomczak JM, Welling M. Attention-based deep multiple instance learning. *ArXiv*. 2018;abs/1802.04712.
17. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng*. 2021;5(6):555–570. doi:10.1038/s41551-020-00682-w
18. Vahadane A, Peng T, Sethi A, et al. Structure-Preserving color normalization and sparse stain separation for histological images. *IEEE Trans Med Imaging*. 2016;35(8):1962–1971. doi:10.1109/tmi.2016.2529665
19. Xiong Z, Liu K, Liu S, et al. Precision HER2: a comprehensive AI system for accurate and consistent evaluation of HER2 expression in invasive breast cancer. *BMC Cancer*. 2024;24(1):1204. doi:10.1186/s12885-024-12980-6
20. Wang YH, Chang MH, Tsai HH, Chien CJ, Wang JC. Transformer-Based HER2 scoring in breast cancer: comparative performance of a foundation and a lightweight model. *Diagnostics*. 2025;15(17). doi:10.3390/diagnostics15172131
21. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst*. 2009;101(21):1446–1452. doi:10.1093/jnci/djp335
22. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565–574. doi:10.1177/0272989x06295361
23. Mosele F, Deluche E, Lusque A, et al. Trastuzumab deruxtecan in metastatic breast cancer with variable HER2 expression: the Phase 2 DAISY trial. *Nat Med*. 2023;29(8):2110–2120. doi:10.1038/s41591-023-02478-2
24. Collins GS, Moons KG, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:q902. doi:10.1136/bmj.q902
25. Lambert JM, Chari RV. Ado-trastuzumab emtansine (T-DM1): an antibody-drug conjugate (ADC) for HER2-positive breast cancer. *J Med Chem*. 2014;57(16):6949–6964. doi:10.1021/jm500766w
26. Tsao LC, Wang JS, Ma X, et al. Effective extracellular payload release and immunomodulatory interactions govern the therapeutic effect of trastuzumab deruxtecan (T-DXd). *Nat Commun*. 2025;16(1):3167. doi:10.1038/s41467-025-58266-8
27. Ogitani Y, Aida T, Hagihara K, et al. DS-8201a, A Novel HER2-Targeting ADC with a novel DNA topoisomerase i inhibitor, demonstrates a promising antitumor efficacy with differentiation from T-DM1. *Clin Cancer Res*. 2016;22(20):5097–5108. doi:10.1158/1078-0432.Ccr-15-2822
28. Li F, Emmerton KK, Jonas M, et al. Intracellular released payload influences potency and bystander-killing effects of antibody-drug conjugates in preclinical models. *Cancer Res*. 2016;76(9):2710–2719. doi:10.1158/0008-5472.Can-15-1795
29. Heldin CH, Rubin K, Pietras K, Ostman A. High interstitial fluid pressure - an obstacle in cancer therapy. *Nat Rev Cancer*. 2004;4(10):806–813. doi:10.1038/nrc1456
30. Thurber GM, Schmidt MM, Wittrup KD. Antibody tumor penetration: transport opposed by systemic and antigen-mediated clearance. *Adv Drug Deliv Rev*. 2008;60(12):1421–1434. doi:10.1016/j.addr.2008.04.012
31. Cortés J, Kim SB, Chung WP, et al. Trastuzumab deruxtecan versus trastuzumab emtansine for breast cancer. *N Engl J Med*. 2022;386(12):1143–1154. doi:10.1056/NEJMoa2115022
32. Khera E, Cilliers C, Bhatnagar S, Thurber GM. Computational transport analysis of antibody-drug conjugate bystander effects and payload tumoral distribution: implications for therapy. *Mol Syst Des Eng*. 2018;3(1):73–88. doi:10.1039/C7ME00093F
33. Staudacher AH, Brown MP. Antibody drug conjugates and bystander killing: is antigen-dependent internalisation required? *Br J Cancer*. 2017;117(12):1736–1742. doi:10.1038/bjc.2017.367
34. Ståhl PL, Salmén F, Vickovic S, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*. 2016;353(6294):78–82. doi:10.1126/science.aaf2403
35. Goltsev Y, Samusik N, Kennedy-Darling J, et al. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell*. 2018;174(4):968–981.e15. doi:10.1016/j.cell.2018.07.010
36. Ding K, Zhou M, Wang H, Zhang S, Metaxas DN. Spatially aware graph neural networks and cross-level molecular profile prediction in colon cancer histopathology: a retrospective multi-cohort study. *Lancet Digit Health*. 2022;4(11):e787–e795. doi:10.1016/s2589-7500(22)00168-6
37. Dent RA, Curigliano G, Hu X, et al. Exploratory biomarker analysis of trastuzumab deruxtecan (T-DXd) vs physician’s choice of chemotherapy (TPC) in HER2-low/ultralow, hormone receptor-positive (HR+) metastatic breast cancer (mBC) in DESTINY-Breast06 (DB-06). *J Clin Oncol*. 2025;43(16_suppl):1013. doi:10.1200/JCO.2025.43.16_suppl.1013
38. Sounderajah V, Guni A, Liu X, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat Med*. 2025;31(10):3283–3289. doi:10.1038/s41591-025-03953-8
39. Ebad SA, Alhashmi A, Amara M, Miled AB, Saqib M. Artificial intelligence-based software as a medical device (AI-SaMD): a systematic review. *Healthcare*. 2025;13(7). doi:10.3390/healthcare13070817
40. Carvalho E, Mascarenhas M, Pinheiro F, et al. Predetermined change control plans: guiding principles for advancing safe, effective, and high-quality AI-ML technologies. *JMIR AI*. 2025;4:e76854. doi:10.2196/76854
41. Evans AJ, Brown RW, Bui MM, et al. Validating whole slide imaging systems for diagnostic purposes in pathology. *Arch Pathol Lab Med*. 2022;146(4):440–450. doi:10.5858/arpa.2020-0723-CP
42. Ardon O, Asa SL, Lloyd MC, et al. Understanding the financial aspects of digital pathology: a dynamic customizable return on investment calculator for informed decision-making. *J Pathol Inform*. 2024;15:100376. doi:10.1016/j.jpi.2024.100376
43. Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ Digit Med*. 2020;3:119. doi:10.1038/s41746-020-00323-1
44. Plass M, Kargl M, Kiehl TR, et al. Explainability and causability in digital pathology. *J Pathol Clin Res*. 2023;9(4):251–260. doi:10.1002/cjp.2.322
45. Benjamins S, Dhunoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med*. 2020;3:118. doi:10.1038/s41746-020-00324-0
46. Wu S, Yue M, Zhang J, et al. The role of artificial intelligence in accurate interpretation of HER2 immunohistochemical scores 0 and 1+ in breast cancer. *Mod Pathol*. 2023;36(3):100054. doi:10.1016/j.modpat.2022.100054
47. De Moura A, Loirat D, Vaillant S, et al. Sacituzumab govitecan in metastatic triple-negative breast cancer patients treated at Institut Curie Hospitals: efficacy, safety, and impact of brain metastases. *Breast Cancer*. 2024;31(4):572–580. doi:10.1007/s12282-024-01565-7
48. Pati P, Jaume G, Foncubierta-Rodríguez A, et al. Hierarchical graph representations in digital pathology. *Med Image Anal*. 2022;75:102264. doi:10.1016/j.media.2021.102264

Breast Cancer: Targets and Therapy

Publish your work in this journal

Breast Cancer - Targets and Therapy is an international, peer-reviewed open access journal focusing on breast cancer research, identification of therapeutic targets and the optimal use of preventative and integrated treatment interventions to achieve improved outcomes, enhanced survival and quality of life for the cancer patient. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/breast-cancer—targets-and-therapy-journal>

Dovepress
Taylor & Francis Group