HERO: Hypothesis-Driven Evidence Retrieval from Omics for Multi-Task Breast Cancer Analysis

Xiangyu Li and Ran Su*

College of Intelligence and Computing, Tianjin University, Tianjin 300072, China
xiangyuli@tju.edu.cn, ran.su@tju.edu.cn

Abstract. Matched multi-omics can improve WSI-based biomarker and prognosis prediction, but most existing pipelines use omics as a parallel feature stream or textual context rather than as an explicit retrieval constraint. HERO asks whether observed omics can be a testable morphology hypothesis: a sparse pathway-to-morphology prior maps DNA methylation and miRNA into a K-dimensional intent vector $\mathbf{m}$ (K=16), TF-IDF retrieval over structured 10× captions selects endpoint-relevant regions, and a cosine gate $c=\cos(\mathbf{m},\mathbf{v})$ triggers deterministic deficit-driven repair when $c<\tau_c$. This closed-loop design bounds VLM calls, reduces reliance on embedding-based semantic matching, and makes every retrieval and verification step lexically auditable. On TCGA-BRCA (930 WSIs, patient-level 5-fold CV), HERO sets new state-of-the-art across ER, PR, HER2, subtype, and risk prediction, outperforming both multimodal fusion and VLM-based baselines.

Keywords: Multi-omics Integration · Whole Slide Image · Evidence Retrieval · Consistency Verification · Breast Cancer

1 Introduction

Breast cancer diagnosis requires joint evidence from H&E whole-slide images (WSIs) and molecular biomarkers (ER/PR/HER2, molecular subtype, prognostic signatures) [25, 7]. A clinically useful computational system should not only predict endpoints but also produce an auditable evidence trail: which tissue regions were examined, what morphology was observed, and whether the observations support the molecular conclusion.

The central technical challenge is *endpoint-aware evidence retrieval* on gigapixel WSIs under intratumoral heterogeneity. Three families of methods address this problem, each with a specific limitation. (i) Weakly supervised MIL [11, 21] aggregates patches under slide-level labels; attention is post-hoc and does not verify evidence sufficiency. (ii) VLM-based WSI readers and multi-agent systems [16, 5, 13, 22] make reading explicit, but often rely on prompt-level guidance rather than a fixed, molecularly conditioned retrieval prior, which can cause

* Corresponding author

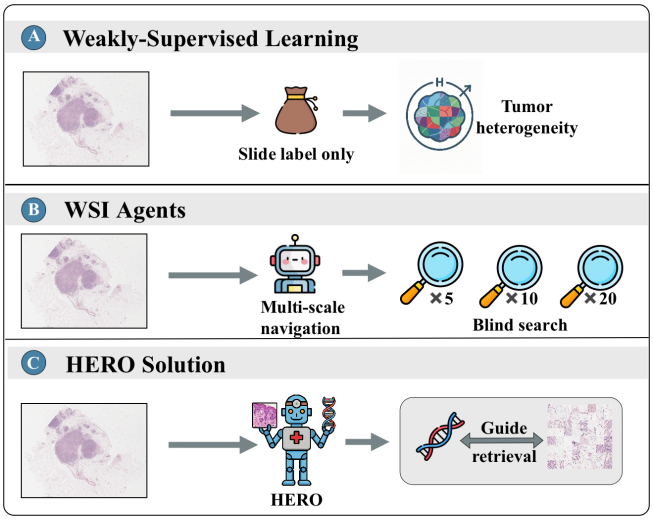

Fig. 1. (A) MIL relies on slide-level labels; attention may highlight non-diagnostic regions under intratumoral heterogeneity. (B) VLM-based WSI readers can suffer retrieval bias toward visually salient regions. (C) HERO uses omics-derived intent to control retrieval and a consistency gate to verify molecular–visual alignment.

retrieval bias toward visually salient but endpoint-irrelevant regions. (iii) Multimodal WSI–omics fusion methods [4, 12] and omics-as-context reasoning systems [9, 15] treat omics as a parallel feature stream or appended text; omics does not control *which* regions are retrieved, nor is it used to *verify* retrieval completeness.

A complementary literature predicts molecular signals directly from H&E morphology [1, 20, 8, 18, 23, 24], showing that morphology encodes molecular information. HERO operates in the reverse direction: it uses *observed* omics to control retrieval and verify evidence completeness, which is the step most existing pipelines leave implicit. We also note interpretable fusion with spatial transcriptomics [19].

We introduce HERO, a four-stage pipeline (Fig. 1) with three contributions:

1. Omics-guided retrieval: a sparse pathway-to-morphology prior converts DNA methylation and miRNA into a K-dimensional intent vector $\mathbf{m}$ and keyword set $\mathcal{Q}$, enabling endpoint-aware patch selection via TF-IDF over structured captions.
2. Consistency gate with deficit-driven repair: a cosine gate $c = \cos(\mathbf{m}, \mathbf{v})$ detects molecular–visual mismatch; when $c < \tau_c$, deterministic re-ranking retrieves patches that cover under-supported morphology axes, bounding compute to one repair round.
3. Robustness analysis: we evaluate sensitivity to K, τ_c, and perturbations of the prior $\mathbf{W}$, providing evidence that the mapping encodes informative biological structure rather than arbitrary assignments.

2 Method

Problem formulation. Given a WSI I with matched DNA methylation $\mathbf{x}_{\text{dna}}$ and miRNA $\mathbf{x}_{\text{mirna}}$, we predict ER/PR/HER2, subtype, and risk while producing an auditable evidence chain. A shared K-dimensional evidence space (K=16) contains a molecular intent vector $\mathbf{m} \in \mathbb{R}^K$, a visual evidence vector $\mathbf{v} \in \mathbb{R}^K$, and a consistency score $c = \cos(\mathbf{m}, \mathbf{v})$. We use TF-IDF and word-boundary checklist matching throughout: pathology descriptors that are semantically similar can be diagnostically distinct (e.g., "high-grade nuclear atypia" vs. "reactive atypia"), and embedding-based retrieval can conflate such terms. Structured prompts elicit standardized terminology for reliable lexical scoring.

2.1 Stage 1: Omics to intent and keywords

Given $\mathbf{x}_{\text{dna}} \in \mathbb{R}^{F_d}$ and $\mathbf{x}_{\text{mirna}} \in \mathbb{R}^{F_m}$, this stage produces an intent vector $\mathbf{m} \in \mathbb{R}^K$, a keyword set $\mathcal{Q}$, and a molecular narrative.

Pathway scoring. We z-normalize each omics vector and compute pathway scores via pre-computed binary masks $\mathbf{M}_{\text{dna}} \in \mathbb{R}^{P \times F_d}$, $\mathbf{M}_{\text{mirna}} \in \mathbb{R}^{P \times F_m}$ derived from MSigDB Hallmark gene sets [17] (P=50) and validated miRNA-target interactions (e.g., miRTarBase) [6]. For pathway p:

$$
\begin{aligned}
s_p^{\text{dna}} &= -\frac{\sum_{j=1}^{F_d} M_{p,j}^{\text{dna}} z(x_j^{\text{dna}})}{\sum_{j=1}^{F_d} M_{p,j}^{\text{dna}} + \epsilon}, \\
s_p^{\text{mirna}} &= -\frac{\sum_{j=1}^{F_m} M_{p,j}^{\text{mirna}} z(x_j^{\text{mirna}})}{\sum_{j=1}^{F_m} M_{p,j}^{\text{mirna}} + \epsilon}, \qquad s_p = \frac{s_p^{\text{dna}} + s_p^{\text{mirna}}}{2}.
\end{aligned}
\tag{1}
$$

The denominator is a pathway-wise average over assigned features, not a global matrix norm; the negation maps hypermethylation and miRNA-mediated repression to pathway down-regulation scores. This linear scoring provides interpretability, resists overfitting on a 930-WSI cohort, and is regularized by the sparse prior $\mathbf{W}$ below.

Morphology axes, checklist, and prior $\mathbf{W}$. We define K=16 axes: mitotic activity, nuclear pleomorphism, tubule formation, necrosis, lymphocytic infiltrate, stromal desmoplasia, solid growth, cribriform pattern, mucin production, vascular invasion, nuclear grade, apoptotic bodies, inflammatory response, cell density, architectural disorder, and calcification. Each axis k has a checklist $\mathcal{C}_k$ of 10–30 descriptors compiled from breast pathology terminology; synonyms are merged and ambiguous terms removed. The checklist is fixed before cross-validation and used unchanged in all folds. Matching uses case-insensitive word boundaries. Instead of unconstrained black-box mapping, we construct a signed, weighted sparse prior $\mathbf{W} \in \mathbb{R}^{K \times P}$ guided by WHO terminology and biological associations (E2F/MYC $\rightarrow$ proliferation; EMT/TGF-β $\rightarrow$ stroma; hypoxia $\rightarrow$ necrosis); rows are ℓ_1-normalized and fixed before cross-validation. This fixed prior does not

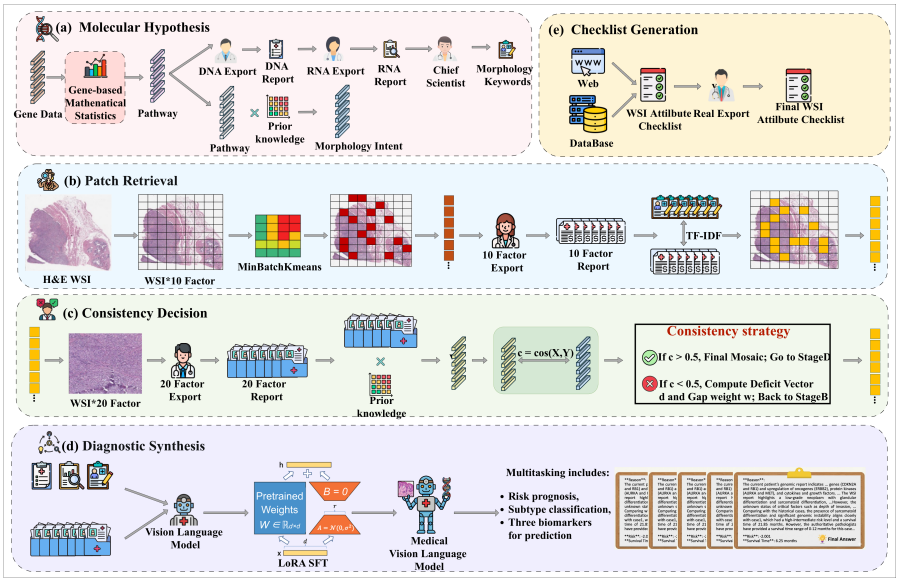

Fig. 2. Overview of HERO. (a) Stage 1: omics→intent via pathway scoring and committee; (b) Stage 2: 10× representative mining and TF-IDF retrieval; (c) Stage 3: 20× consistency gate and deficit-driven repair; (d) Stage 4: LoRA-tuned VLM diagnosis; (e) morphology axis checklist shared across stages.

by itself rule out shortcut learning. Instead, it makes the omics-to-morphology mapping explicit, fold-invariant, and auditable; $\mathbf{W}$ is never optimized with endpoint labels. $\mathbf{W}$ serves as an auditable prior and explicit regularizer: $\mathbf{W}$-shuffle degrades HER2 by 0.033 and Risk by 0.032, while noise/drop cause $\leq$0.009 degradation (Table 3).

Intent vector.

$$\mathbf{m} = \ell_2\text{-normalize}(\text{ReLU}(\mathbf{W}\mathbf{s})) \in \mathbb{R}^K. \tag{2}$$

Committee module. A role-based chain of three LLM calls returns a structured record with up/down-regulated pathways, morphology axes, a molecular narrative, and a 15–40 term keyword set $\mathcal{Q}$ after synonym merging and removal of ambiguous descriptors. The keyword set is used only for Stage 2 lexical retrieval.

2.2 Stage 2: 10× retrieval (300→Top-64)

Representative mining (N_r=300). Tissue patches (256×256 px at 20×) are encoded with ℓ_2-normalized UNI [3] embeddings (1024-dim). MiniBatchKMeans (k=N_r=300, batch size 1024) clusters the embeddings; the patch closest to each centroid is selected as a representative, ensuring coverage of the slide's global morphological diversity.

10× captioning and TF-IDF ranking. Each representative p_i is revisited at 10× and captioned by Qwen2.5-VL-7B [2] with a structured prompt covering tumor architecture, stromal composition, inflammatory infiltrate, and necrosis, producing cap_i^{10}. A TF-IDF vectorizer $\phi(\cdot)$ is fit on the 300-caption corpus; each representative is scored against the keyword set $\mathcal{Q}$:

$$\text{score}_{10}(p_i) = \cos\big(\phi(\mathcal{Q}),\ \phi(\text{cap}_i^{10})\big). \tag{3}$$

The vectorizer is fit separately per slide, so IDF is a within-slide rarity weight rather than a cohort-level learned parameter; no endpoint label is used. The Top-64 by score_{10} form the evidence set for Stage 3, corresponding to an 8×8 mosaic (single VLM forward pass).

2.3 Stage 3: 20× consistency gate and repair

Mosaic construction and captioning. The Top-64 regions are cropped at 20× (256×256 px), arranged into a coordinate-labeled 8×8 mosaic, and captioned with the same structured prompt, now requesting cytologic features such as nuclear grade, mitoses, chromatin pattern, and cell-level morphology.

Visual evidence vector $\mathbf{v}$. For caption i and axis k, $u_{ik} = 1$ if any checklist term in $\mathcal{C}_k$ appears in cap_i^{20} under word-boundary matching. We aggregate $v_k = \sum_{i=1}^{64} u_{ik}$ and ℓ_2-normalize $\mathbf{v}$; if all v_k=0, $\mathbf{v}$=$\mathbf{0}$.

Consistency gate and repair. The initial Top-64 mosaic is captioned once at 20×. We compute:

$$c = \cos(\mathbf{m}, \mathbf{v}) = \frac{\mathbf{m}^\top \mathbf{v}}{\|\mathbf{m}\|_2 \|\mathbf{v}\|_2}, \tag{4}$$

$$\mathbf{w}' = \text{softmax}(\text{ReLU}(\mathbf{m} - \mathbf{v})/\tau), \quad \tau{=}0.3, \tag{5}$$

$$\text{score}_{\text{repair}}(p_i) = \text{score}_{10}(p_i) + \mathbf{w}'^{\top} \mathbf{a}_i. \tag{6}$$

We set c=0 when $\mathbf{v}$=$\mathbf{0}$ and accept the mosaic if $c \geq \tau_c$ (τ_c=0.5). If $c < \tau_c$, the deficit $\mathbf{m} - \mathbf{v}$ identifies under-supported axes. Each unselected representative receives an axis-hit vector $\mathbf{a}_i \in \{0,1\}^K$ from its 10× caption; the top-32 repair candidates are merged with the original 64, the final Top-64 is rebuilt from this 96-patch pool, and one additional 20× captioning pass is performed. No further repair round is performed.

2.4 Stage 4: Diagnosis module

The final 20× mosaic, molecular narrative, and consistency summary (including c and top deficit axes when repair was triggered) are fed into Qwen2.5-VL-7B [2] with LoRA [10] (rank 64, α=128, `q/k/v/o_proj`; backbone frozen). The molecular narrative is textual context, not an additional learned omics feature; the

consistency summary reports c and whether visual evidence already supports the molecular hypothesis or required repair. The model outputs JSON predictions for ER, PR, HER2, subtype, and risk. Training uses teacher forcing with cross-entropy loss on the structured JSON tokens; at inference, greedy decoding is used. Only Stage 4 is tuned because Stages 1–3 are intended to remain fixed retrieval and verification modules; tuning the retrieval prior or repair rule with endpoint labels would make it harder to isolate evidence selection from final diagnosis. All adaptivity is deterministic (gate + re-ranking); the VLM budget per slide is 300 (10×) + 64 (20×) + at most 32 repair (20×, triggered on 37% of slides). $\{\mathcal{C}_k\}$ and $\mathbf{W}$ are fixed across folds.

3 Experiments

3.1 Setup

We evaluate on TCGA-BRCA (930 WSIs, 878 patients) with matched DNA methylation (14,115 CpGs) and miRNA-seq (315 miRNAs) for ER/PR/HER2, subtype (5-class), and survival risk, using patient-level 5-fold stratified cross-validation (seed 42). Labels for ER, PR, HER2, clinical subtype, and risk are from TCGA clinical annotations; missing or equivocal receptor status is excluded per endpoint. These endpoint labels are not produced by HERO from its input methylation and miRNA vectors. Only the Stage 4 model is LoRA-tuned; UNI features (1024-dim, 20×) are pre-extracted. Baselines: omics-only (SNN [14], SNNTrans [27]), WSI-only MIL (ABMIL [11], TransMIL [21]), multimodal fusion (MCAT [4], MOTCat [26], SurvPath [12]), VLM/agent (GPT-4o, MDAgent [13]). For fair comparison, all multimodal baselines were implemented to consume the exact same concatenated DNA methylation and miRNA feature vectors as our method. This unified protocol controls input modality across methods, but may differ from the originally optimized configuration of some baselines; results are interpreted under a common methylation+miRNA setting rather than as a claim that each baseline is globally optimized. Inference: 57 s/WSI on 2×A100, 22 GB peak; the gate triggers on 37% of test slides.

3.2 Comparative results

Table 1 reports results across all endpoints. HERO obtains the highest scores on every metric. The largest improvements over the best fusion baseline (SurvPath) are on HER2 (+0.050 AUC) and Risk (+0.030 C-index), the two endpoints where diagnostically relevant morphology is sparse and heterogeneous, making retrieval quality a bottleneck.

To isolate the contribution of omics-guided retrieval from VLM capacity, we evaluate two controlled baselines that use the *same* LoRA-tuned Qwen2.5-VL-7B diagnosis head as HERO but replace the retrieval mechanism: VLM + TransMIL Top-64 (attention-selected patches) and VLM + Random Top-64 (uniformly sampled patches). HERO improves over VLM + TransMIL Top-64 by +0.090

Table 1. Results on TCGA-BRCA (patient-level 5-fold CV, mean±std). AUC for ER/PR/HER2, macro AUC for Subtype, C-index for Risk. "*": our reimplementation. VLM baselines use the same LoRA-tuned Qwen2.5-VL-7B as HERO with Random or TransMIL-selected Top-64 evidence.

Method	ER (AUC)	PR (AUC)	HER2 (AUC)	Subtype (AUC)	Risk (C-idx)
SNN* [14]	0.952±0.021	0.885±0.035	0.725±0.045	0.958±0.015	0.620±0.045
SNNTrans* [27]	0.968±0.018	0.912±0.028	0.765±0.038	0.970±0.012	0.645±0.038
ABMIL [11]	0.885±0.032	0.755±0.045	0.655±0.052	0.955±0.022	0.590±0.052
TransMIL [21]	0.901±0.028	0.763±0.041	0.687±0.048	0.972±0.015	0.615±0.048
MCAT* [4]	0.975±0.022	0.935±0.031	0.768±0.042	0.975±0.018	0.665±0.042
MOTCat* [26]	0.982±0.019	0.948±0.025	0.795±0.036	0.980±0.014	0.685±0.035
SurvPath* [12]	0.991±0.015	0.957±0.022	0.815±0.032	0.985±0.012	0.705±0.030
GPT-4o*	0.945±0.035	0.880±0.045	0.710±0.055	0.950±0.025	0.610±0.065
MDAgent* [13]	0.965±0.025	0.915±0.038	0.755±0.048	0.972±0.018	0.640±0.050
VLM + Rand. Top-64	0.845±0.045	0.720±0.055	0.625±0.060	0.865±0.035	0.565±0.050
VLM + TransMIL Top-64	0.962±0.020	0.905±0.030	0.775±0.040	0.978±0.015	0.665±0.035
HERO (Ours)	**0.994±0.010**	**0.978±0.015**	**0.865±0.022**	**0.995±0.008**	**0.735±0.025**

Table 2. Ablation on HER2 and Risk (mean±std). Gate-only computes c without repair; Random repair adds 32 random patches.

Setting	HER2 (AUC)	Risk (C-idx)
Visual-only Top-64	0.758±0.048	0.645±0.050
Text-only retrieval	0.835±0.030	0.690±0.035
w/o consistency gate	0.822±0.036	0.695±0.038
Gate-only	0.828±0.032	0.702±0.032
Random repair	0.830±0.040	0.700±0.045
HERO (full)	**0.865±0.022**	**0.735±0.025**

Table 3. Sensitivity to K, τ_c, and $\mathbf{W}$ perturbations (mean±std; HER2 AUC / Risk C-index).

Setting	HER2 (AUC)	Risk (C-idx)
K=8	0.852±0.026	0.724±0.028
K=16	0.865±0.022	0.735±0.025
K=32	**0.867±0.023**	**0.736±0.026**
τ_c=0.3	0.858±0.024	0.728±0.027
τ_c=0.5	**0.865±0.022**	**0.735±0.025**
τ_c=0.7	0.861±0.025	0.731±0.026
$\mathbf{W}$ (ours)	**0.865±0.022**	**0.735±0.025**
$\mathbf{W}$-noise	0.860±0.025	0.732±0.028
$\mathbf{W}$-drop	0.856±0.027	0.729±0.029
$\mathbf{W}$-shuffle	0.832±0.035	0.703±0.038

on HER2 and +0.070 on Risk, indicating that gains originate from molecular-guided retrieval and verification rather than model scale. VLM + Random Top-64 is much lower (HER2 0.625, Risk 0.565), confirming that a 7B VLM without informed retrieval is insufficient.

3.3 Ablation study

Component ablations (Table 2). Text-based retrieval outperforms visual-only selection (+0.077 HER2, +0.045 Risk), indicating that caption semantics capture endpoint-relevant cues beyond embedding saliency. Removing the consistency gate reduces HER2 by 0.043 and Risk by 0.040. Gate-only (computing c without repair) and Random repair (adding 32 random patches) recover only a fraction of the gap, confirming that deficit-driven re-ranking is the operative component. All ablations use the same LoRA-tuned diagnosis head; only the retrieval/gating pipeline differs.

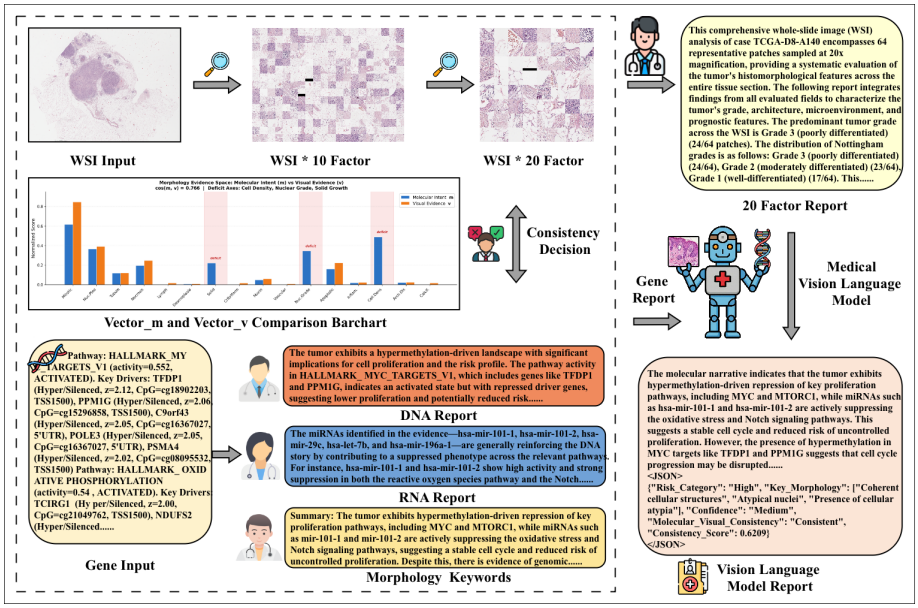

Fig. 3. Case-level evidence chain. Omics→intent $\mathbf{m}$ guides initial retrieval; captions yield $\mathbf{v}$ and $c = \cos(\mathbf{m}, \mathbf{v})$. When $c < \tau_c$, repair candidates rebuild the final mosaic; dense molecular narratives are summarized into intent axes for readability.

Sensitivity and robustness (Table 3). K=16 balances interpretability and coverage: K=8 merges distinct cytologic and stromal patterns, while K=32 adds little gain but requires a larger checklist. τ_c=0.5 is fixed and endpoint-independent; performance is stable from 0.3 to 0.7. Among $\mathbf{W}$ perturbations, additive noise and entry dropout cause $\leq$0.009 degradation; shuffling pathway-to-axis assignments degrades HER2 by 0.033 and Risk by 0.032, supporting that $\mathbf{W}$ encodes informative biological structure rather than arbitrary connectivity.

3.4 Visualization

Fig. 3 shows a held-out case where Stage 1 intent $\mathbf{m}$ highlights proliferation and stromal axes. Initial retrieval yields c=0.38 $< \tau_c$ with necrosis and lymphocytic infiltrate under-represented; after one repair round the rebuilt mosaic reaches c=0.74. Globally, the gate triggers on 37% of test slides, raising c from 0.41 to 0.78 on average and improving HER2 AUC by +0.045 and Risk C-index by +0.038 on the triggered subset. For the remaining 63% non-triggered cases, the initial zero-shot retrieval was highly aligned (mean c=0.81), ensuring safety without extra compute overhead.

4 Conclusion

Limitations. Evaluation is limited to internal TCGA-BRCA cross-validation and does not establish robustness across institutions, scanners, stains, or popula-

tions. The breast-cancer-specific 16-axis checklist would need auditing for other tumor types. VLM captions remain fallible: structured prompts and checklist matching make evidence inspectable, but do not guarantee pathologic correctness or remove same-model bias between captioning and diagnosis. Inference is also costly at 57 s/WSI on 2×A100, so faster multi-stage VLM inference is needed before clinical deployment.

HERO turns matched multi-omics into a testable morphology hypothesis by mapping omics to an intent vector, retrieving endpoint-relevant regions with TF-IDF captions, and verifying molecular–visual alignment with a cosine consistency gate and deterministic repair. On TCGA-BRCA (930 WSIs, 5-fold CV), HERO achieves state-of-the-art across ER/PR/HER2, subtype, and risk.

Acknowledgements

This work was supported by the Open Project Program of the State Key Laboratory of Medical Proteomics (SKLP-O202406), the Emerging Frontiers Cultivation Program of Tianjin University Interdisciplinary Center, the Seed Foundation of Tianjin University, the Internal Research Grants of Macao Polytechnic University (No. RP/CAI02/2023), and the Science and Technology Development Fund (No. 0177/2023/RIA3).

Disclosure of Interests

The authors have no competing interests to declare.

References

1. Arslan, S., Schmidt, J., Bass, C., et al.: A systematic pan-cancer study on deep learning-based prediction of multi-omic biomarkers from routine pathology images. Communications Medicine **4**(48) (2024). https://doi.org/10.1038/s43856-024-00471-5
2. Bai, S., Chen, K., Liu, X., et al.: Qwen2.5-VL technical report. arXiv preprint arXiv:2502.13923 (2025)
3. Chen, R.J., Ding, T., Lu, M.Y., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**(3), 850–862 (Mar 2024). https://doi.org/10.1038/s41591-024-02857-3
4. Chen, R.J., Lu, M.Y., Weng, W.H., Chen, T.Y., Williamson, D.F., Manz, T., Shady, M., Mahmood, F.: Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3995–4005. IEEE (Oct 2021). https://doi.org/10.1109/iccv48922.2021.00398
5. Chen, Y., Wang, G., Ji, Y., et al.: SlideChat: A large vision-language assistant for whole-slide pathology image understanding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5134–5143 (Jun 2025)

6. Chou, C.H., Shrestha, S., Yang, C.D., et al.: miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Research **46**(D1), D296–D302 (Jan 2018). https://doi.org/10.1093/nar/gkx1067
7. Ding, T., Wagner, S.J., Song, A.H., et al.: A multimodal whole-slide foundation model for pathology. Nature Medicine **31**, 3749–3761 (Nov 2025). https://doi.org/10.1038/s41591-025-03982-3
8. Ekholm, A., Wang, Y., Vallon-Christersson, J., et al.: Prediction of gene expression-based breast cancer proliferation scores from histopathology whole slide images using deep learning. BMC Cancer **24**, 1510 (2024). https://doi.org/10.1186/s12885-024-13248-9
9. Fallahpour, A., Ma, J., Munim, A., et al.: MedRAX: Medical reasoning agent for chest x-ray. arXiv preprint arXiv:2502.02673 (2025)
10. Hu, E.J., Shen, Y., Wallis, P., et al.: LoRA: Low-rank adaptation of large language models. In: International Conference on Learning Representations (ICLR 2022) (2022)
11. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: Proceedings of the 35th International Conference on Machine Learning (ICML). Proceedings of Machine Learning Research, vol. 80, pp. 2127–2136. PMLR (2018)
12. Jaume, G., Vaidya, A., Chen, R.J., Williamson, D.F., Liang, P.P., Mahmood, F.: Modeling dense multimodal interactions between biological pathways and histology for survival prediction. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11579–11590. IEEE (Jun 2024). https://doi.org/10.1109/cvpr52733.2024.01100
13. Kim, Y., Park, C., Jeong, H., et al.: MDAgents: An adaptive collaboration of LLMs for medical decision-making. arXiv preprint arXiv:2404.15155 (2024)
14. Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-normalizing neural networks. In: Advances in Neural Information Processing Systems (NeurIPS 2017). pp. 971–980 (2017)
15. Li, B., Yan, T., Pan, Y., et al.: MMedAgent: Learning to use medical tools with multi-modal agent. In: Findings of the Association for Computational Linguistics: EMNLP 2024. pp. 8745–8760 (Nov 2024). https://doi.org/10.18653/v1/2024.findings-emnlp.510
16. Liang, Y., Lyu, X., Chen, W., et al.: WSI-LLaVA: A multimodal large language model for whole slide image. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 22718–22727 (Oct 2025)
17. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., Tamayo, P.: The molecular signatures database hallmark gene set collection. Cell Systems **1**(6), 417–425 (Dec 2015). https://doi.org/10.1016/j.cels.2015.12.004
18. Liu, H., Xie, X., Wang, B.: Deep learning infers clinically relevant protein levels and drug response in breast cancer from unannotated pathology images. npj Breast Cancer **10**(18) (2024). https://doi.org/10.1038/s41523-024-00620-y
19. Liu, L., Pan, X., Yuan, Y., et al.: Prototype-driven fusion of pathology and spatial transcriptomics for interpretable survival prediction. arXiv preprint arXiv:2602.12441 (2026)
20. Mondol, R.K., Millar, E.K.A., Graham, P.H., Browne, L., Sowmya, A., Meijering, E.: hist2RNA: An efficient deep learning architecture to predict gene expression from breast cancer histopathology images. Cancers **15**(9), 2569 (2023). https://doi.org/10.3390/cancers15092569
21. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., Zhang, Y.: TransMIL: Transformer based correlated multiple instance learning for whole slide image clas-

sification. In: Advances in Neural Information Processing Systems (NeurIPS 2021). pp. 2136–2147 (2021)
22. Tang, X., Zou, A., Zhang, Z., Li, Z., Zhao, Y., Zhang, X., Cohan, A., Gerstein, M.: MedAgents: Large language models as collaborators for zero-shot medical reasoning. In: Findings of the Association for Computational Linguistics: ACL 2024. pp. 599–621 (Aug 2024). https://doi.org/10.18653/v1/2024.findings-acl.33
23. Wang, Y.K., Tydlitatova, L., Kunz, J.D., et al.: Screen them all: High-throughput pan-cancer genetic and phenotypic biomarker screening from H&E whole slide images. arXiv preprint arXiv:2408.09554 (2024)
24. Wu, S., Xu, S.: Virtual immunohistochemistry for breast cancer biomarker prediction from H&E-stained images using generative network. Image Analysis and Stereology **44**(3), 159–170 (2025). https://doi.org/10.5566/ias.3613
25. Xu, H., Usuyama, N., Bagga, J., et al.: A whole-slide foundation model for digital pathology from real-world data. Nature **630**(8015), 181–188 (May 2024). https://doi.org/10.1038/s41586-024-07441-w
26. Xu, Y., Chen, H.: Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 21184–21194. IEEE (Oct 2023). https://doi.org/10.1109/iccv51070.2023.01942
27. Zhou, H., Zhou, F., Chen, H.: Cohort-individual cooperative learning for multimodal cancer survival analysis. IEEE Transactions on Medical Imaging (2024). https://doi.org/10.1109/TMI.2024.3455931, early access