npj Precision Oncology

https://doi.org/10.1038/s41698-026-01393-2

Article in Press

Multimodal AI model for early detection of hepatocellular carcinoma

Received: 21 June 2025

Accepted: 16 March 2026

Cite this article as: Jing, S.-y., Li, X.-l., Wen, H. *et al.* Multimodal AI model for early detection of hepatocellular carcinoma. *npj Precis. Onc.* (2026). https://doi.org/10.1038/s41698-026-01393-2

Si-yu Jing, Xue-liang Li, Hong Wen, Jing-xuan Cai, Yong-qi Bei, Xue-lian Zhao, Zheng-tao Zhang, Xiao-han Fan, Yi-ning Zou, Ling-li Chen, Yu-lin Wang, Cai-ying Wang, Xin Li, Li-li Meng, Ping Lin, Xi Yan, Yuan Ji & Hong Li

Multimodal AI model for early detection of hepatocellular carcinoma

Si-yu Jing[1]*, Xue-liang Li[1]*, Hong Wen[2,5]*, Jing-xuan Cai[1], Yong-qi Bei[1], Xue-lian Zhao[3], Zheng-tao Zhang[4], Xiao-han Fan[2], Yi-ning Zou[2], Ling-li Chen[2], Yu-lin Wang[2], Cai-ying Wang[2], Xin Li[2], Li-li Meng[2], Ping Lin[1], Xi Yan[1], Yuan Ji[2,3]†, Hong Li [1]†

Affiliations

[1] Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, 200031, People's Republic of China.

[2] Department of Pathology, Zhongshan Hospital Fudan University, Shanghai, 200032, People's Republic of China.

[3] Molecular Pathology Center, Zhongshan Hospital, Fudan University, Shanghai, 200032, People's Republic of China.

[4] State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai 200031, P. R. China.

[5] Department of Pathology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, P. R. China.

*These authors contributed equally to this work.

†Correspondence to: ji.yuan@zs-hospital.sh.cn (Y. J.); lihong01@sinh.ac.cn (H. L.)

Abstract

Detection of early hepatocellular carcinoma (eHCC) is important for timely treatment and improved prognosis. However, it is challenging to distinguish eHCC from pre-malignant high-grade dysplastic nodules (HGDN). Here we developed an artificial intelligence (AI) derived computational framework to identify potential biomarkers and built classification models for eHCC and HGDN. A two-stage multiscale deep learning model (TMC-net) captured the subtle features based on H&E images, and outperformed the pathology foundation model and traditional histopathological features. The learned features were consistent with clinical diagnostic criteria and could be highlighted on the virtual images, assisting junior pathologists in improving the diagnostic accuracy. Four marker genes were screened through comparative transcriptome analysis. The multimodal model based on marker genes and histopathological features achieved AUROC of 0.8875 and 0.9500 on the internal and external test sets, respectively. We confirmed the morpho-phenotype correlations of these genes and found that the multimodal features were associated with patient prognosis in a broader HCC cohort. This study reveals histopathological and transcriptomic features of eHCC, and provides an optimized AI solution for assistant diagnosis.

Keywords Hepatocellular carcinoma; Early diagnosis; Intelligent pathology; Multimodal integration

Introduction

Hepatocellular carcinoma (HCC) is the most common form of liver cancer and accounts for ~80% of cases, with increasing incidence and mortality rates [1,2]. Since most HCC patients are diagnosed at an advanced stage with a median survival of about 8 months, early diagnosis is crucial [1]. Early-stage HCC (eHCC) is a subset of well-differentiated HCC (WD-HCC), which is smaller in size, has no obvious capsule and contains portal elements without significantly affecting the original structure of the liver [3]. Indeed, eHCC developed in a multi-stage manner, with around 90% of eHCC cases occur in the setting of chronic liver disease and usually with small hepatic nodules [1]. Some of these nodules are benign, such as regenerative nodules, and some with malignant potential, such as dysplastic nodules (DNs). Based on the degree of atypia, DN can be further divided into low-grade DN (LGDN) and high-grade DN (HGDN), the latter of which is considered to be the precursor of HCC [4]. A major challenge in the clinical practice of liver cancer is distinguishing eHCC from HGDN.

Although HGDN is a premalignant phase, HGDN exhibits moderate cellular atypia [5,6], and share similar histological features to eHCC, such as a vague nodular shape, significantly higher cell density, and a higher degree of dysplasia [4,7]. Histopathological diagnosis is the gold standard for the diagnosis of eHCC. Although several immunohistochemical markers such as GPC3, HSP70 and GS have been reported to play a role in the diagnosis of HCC, some limitations still exist [8,9]. Currently, even for experienced liver clinicians, radiologists, and liver pathologists, the differentiation between HGDN and eHCC remains problematic [4,10-12].

The advent of deep learning and foundational models provides a new opportunity to revisit classical approaches to diagnosis, which can capture subtle features [13-16]. Several studies have used deep learning to establish prognostic models on whole slide images (WSIs) for HCC patients [17], and revealed some underlying prognostic indicators including immune gene signatures [18], different histological subtypes [19], genetic alterations [20]. But these studies have focused on advanced tumors and have largely ignored early liver cancer. Recently, classification models for five hepatocellular nodular lesions and two background tissues were built, but the distinction between precancerous HGDN and eHCC remains unclear [21]. Some studies have shown that combining histopathological imaging with other modalities can reveal multimodal relationships and generally improve the performance of these challenging cases [22,23]. Therefore, we propose that the combination of histological images and molecules can more effectively distinguish eHCC from HGDN.

This study comprehensively characterized the differences between eHCC and HGDN, by utilizing multimodal integration of pathological images and RNA data. We proposed a two-stage multi-scale deep learning model, TMC-net, which can extract more accurate diagnostic features from hematoxylin-eosin (H&E) images and performed well on both internal validation, internal test and external test datasets. In addition, we screened out key genes as potential biomarkers. By integrating the histopathological features extracted by TMC-net and the expression levels of 4 key genes, we achieved an area under the receiver operating characteristic (AUROC) of 0.92 on an independent external dataset. Further analysis found that these multimodal features could not only effectively distinguish eHCC from HGDN, but also were associated with the prognosis of 365 liver cancer patients. This study provided a novel optimization perspective for solving complex

pathology diagnosis challenges. The proposed multimodal modeling framework can provide a viable solution for the diagnosis of other visually similar early diseases in various medical fields.

Results

Overview of the data and proposed multimodal modeling framework

To identify biomarkers associated with HGDN and eHCC and build artificial intelligence (AI) models to support diagnosis, we first extracted histopathological and expression features from H&E images and transcriptome, and then built unimodal and multimodal models to classify eHCC, HGDN and other regions (**Fig. 1A**). Given the lack of distinct differences between the lesions of HGDN and eHCC on H&E images, we employed three methods to capture the subtle histopathological features: 1) latent features retrieved from supervised learning (TMC-net); 2) latent features obtained from the pathology foundation model; 3) traditional pathological features, which refer to quantifiable cellular features. Then the unimodal model was built from H&E images, which are most likely to be applied in clinical practice. On the other hand, transcriptome was analyzed to explore the differentially expressed genes (DEGs) and functional changes between eHCC and HGDN. To improve prediction accuracy, the best histopathological features and the expression levels of several key genes were combined to construct a multimodal model (eLiver). All models were evaluated via internal and external test sets.

Because eHCC and HGDN lesions are extremely occult, the number of surgical cases is low even in hospitals with a large patient volume. We screened patients recorded as having HCC or DN from Zhongshan Hospital Fudan University (ZS) and Huashan Hospital Fudan University (HS). Their H&E images were re-examined by at least two senior pathologists specializing in liver disease, and immunohistochemical analysis of multiple protein markers was performed when necessary (**Fig. 1B**; **see Methods**). Any slides containing confirmed eHCC or HGDN were retained, and the lesions of eHCC or HGDN were outlined by a pathologist. Before modeling, WSIs were segmented into tiles of 512 μm in size. The training, validation, and test sets were divided by patients to avoid information leakage. This process obtained 33,573 tiles from ZS for training, 40,318 tiles from ZS for internal testing, and an additional 3,430 tiles from HS for external testing (**Fig. 1C**; **Supplementary Table S1**; **see Methods**).

We subsequently measured the transcriptome of micro-dissected tissues of eHCC lesions (n=36), HGDN (n=34) lesions, and peripheral liver tissues (n=16). Among them, 65 ZS-derived samples were divided into training and internal test sets at a ratio of 8:2, and an additional 21 samples from HS were used as an external test set (**Fig. 1C**; **see Methods**). 50 samples with H&E images (eHCC=15, HGDN=24, peripheral liver=11) constituted a bimodal cohort (**Fig. 1C**).

Superior eHCC diagnosis from H&E using novel TMC-net

In clinical practice, a diagnosis needs to be made for each patient or slide. However, the eHCC and HGDN lesions usually account for only a small part of the WSI, and the majority of the remaining area consists of abnormal liver tissues such as cirrhosis. Therefore, we first proposed a two-stage multi-scale convolutional network (TMC-net) to predict class labels for each individual tile, then aggregated tile-level features to slide-level features to enable classify for each slide (**Fig. 2A**). The design of TMC-net was inspired by the clinical diagnostic workflow, where pathologists first examine the

whole slide at low magnification to identify regions of interest, and then closely examine these candidate regions at high magnification to make the final diagnosis. The first stage of TMC-net is a candidate region extraction network. This network takes the tile-level H&E image tiles as input and classifies them into two categories: tiles with a high risk of malignancy (candidate regions) and tiles belonging to other non-critical areas. The tiles predicted as candidate regions are then passed to the second stage, the eHCC detection network. This network performs a more fine-grained classification, distinguishing between eHCC and HGDN (**see Methods**).

To evaluate the tile-level accuracy of the TMC-net model, we performed leave-one-out cross-validation (LOOCV) evaluation on annotated tiles (**Fig. 1C**). Visualization of the tile embeddings of the last layer of the TMC-net showed a clear separation between the HGDN and eHCC classes in the latent space (**Fig. 2B**). This indicated that the multi-scale feature extraction effectively learns distinct features that distinguish eHCC from HGDN. Evaluation using LOOCV showed that the TMC-net achieved 87.1% tile-level classification accuracy with an AUROC of 0.941 (**Fig. 2C**).

Next, we explored the effect of integrating multi-scale features at different field-of-view sizes on model accuracy and AUROC. The results showed that the best configuration was to utilize a 512 μm-sized field-of-view to extract complementary features at 20x and 40x magnifications as model input, which achieved the best differentiation between eHCC and HGDN (**Supplementary Fig. S1A**). With the optimal tile size configuration established, we then conducted ablation studies using different network settings to understand the importance of the two-stage design and multi-scale architecture of TMC-net (**see Methods**). The results showed that removing any individual module from the TMC-net led to in a decrease in accuracy and AUROC (**Supplementary Fig. S1B**). This indicated that both the "two-stage framework " and "multi-scale CNN" elements made crucial contributions to superior performance of the model.

Subsequently, we calculated 34 aggregated features across the entire slide. These features were derived from the distribution patterns, central tendency, and variability characteristics of the tile-level predictions (**Supplementary Fig. S2** and **Supplementary Table S2**; **see Methods**). These features were then utilized as input for a support vector machine to enable whole-slide-level prediction. The LOOCV evaluation achieved a classification accuracy of 84.2% at the slide level, with an AUROC of 0.918 (**Fig. 2D** and **Fig. 3A**). This integrated approach effectively combined tile-wise deep learning outputs with higher-order statistical characteristics of the full tissue sample (**Fig. 2A**), thereby enhancing the robustness and discriminative power of the classification model.

In addition to the novel TMC-net approach, we evaluated two alternative pathology feature extraction methods: based on the pathology foundation model PLIP [24], and based on quantifiable cellular features, including quantitative cell density ratio and cellular atypia index (**Fig. 1A**). The slide-level predictions were obtained from these features using the same modeling framework (**see Methods**). The results confirmed that TMC-net outperformed the other two methods on all three datasets (**Fig. 3A** and **3B**). TMC-net achieved an accuracy of 0.8415 ± 0.0182 and an AUROC of 0.9178 ± 0.0175 on the validation set (**Fig. 3B**, left). On the internal test set, the accuracy was 0.8136 ± 0.0338 and the AUROC was 0.8354 ± 0.0214 (**Fig. 3B**, middle), demonstrating that TMC-net exhibits consistent performance when extended to new relevant data. Importantly, TMC-net achieved excellent accuracy of 0.9714 ± 0.0286 and AUROC of 0.9916 ± 0.0084 on an independent external test set (**Fig. 3B**, right), suggesting its generalizability even across different institutions and imaging protocols, and promising potential for validation and optimization in more external cohorts.

Taken together, TMC-net, as a two-stage multi-scale convolutional network, demonstrated the ability to extract superior

local image features. Leveraging the TMC-net architecture, we have developed a whole-slide level diagnostic framework for eHCC from H&E images. This framework combines deep learning-based tile-level feature extraction with slide-level statistical modeling, and has maintained consistent performance across an external dataset. This consistent performance underscored the potential generalizability and clinical applicability of the TMC-net.

Interpretability and assisted diagnosis of TMC-net

The whole-slide diagnosis framework not only provides prediction on H&E slides, but also supports human pathologists. By highlighting regions with high HCC probability (**Fig. 4A**), TMC-net can help pathologists focus on potentially malignant areas and reduce the time required for diagnosis. We invited pathologists with different experience levels to evaluate the validation dataset, including two early-career junior pathologists with less than 5 years of experience, a mid-level pathologist, and a senior pathologist with more than 15 years of experience. Each pathologist independently diagnosed the WSI without any other assistance. The HCC probabilities predicted by TMC-net for the tiles were then visualized as risk heatmaps, which highlighted areas that were more likely to represent malignant lesions. A few months later, the junior pathologist re-evaluated the WSI with the help of risk heatmaps generated by TMC-net. The diagnosis results showed that the inter-pathologist agreement kappa score between junior pathologists was improved by 9% with the help of the risk heatmaps. Their diagnostic accuracy improved from 0.60 without assistance to 0.77 with heatmap support, and the F1 score improved from 0.61 to 0.80 (**Supplementary Fig. S3**).

To further elucidate the interpretability of TMC-net, we collaborated with experienced pathologists to analyze the key histopathological features highlighted by the model on the tiles (**see Methods**). This interpretability analysis revealed that the high-attention areas of TMC-net parsed by Grad-CAM [25] were generally consistent with the diagnostic criteria recognized by pathologists and widely used in clinical practice. For example, although eHCC is less overtly malignant, the model accurately focused on areas showing portal invasion, a critical criterion for clinical HCC diagnosis (**Fig. 4B**) [26]. For HGDN, TMC-net directed attention towards areas with increased cellularity and steatosis (**Fig. 4C**), the former being the important feature and the latter indicating a high risk of HGDN. In addition, we also extended the analysis to several WD-HCC slides, on which TMC-net highlighted regions showing marked cellular atypia, prominent nucleoli and hyperchromasia, which are all hallmarks of malignant transformation (**Fig. 4D**).

Taken together, the key histopathological features highlighted by TMC-net predictions reflect the diagnostic criteria recognized by experienced pathologists. This AI-assisted approach represents a valuable decision support system to improve early detection rates of HCC in clinical settings.

Transcriptome alterations and key genes

Next, we sought to further elucidate the molecular underpinnings and identify potential biomarkers that distinguishes eHCC from HGDN. We measured the transcriptome from micro-dissected tissues of eHCC, HGDN, and the control tissues of surrounding livers. By comparing DEGs between HGDN and eHCC samples, we observed distinct molecular signatures consistent with pathological progression from a pre-malignant to an early malignant state. In HGDN, upregulated genes were significantly enriched in metabolic pathways related to xenobiotic metabolism, amino acid and organic acid processing, oxidative phosphorylation, and adipogenesis (**Supplementary Fig. S4A** and **S4B**). In contrast, eHCC samples exhibited significant upregulation of genes associated with epithelial-mesenchymal transition, cell

adhesion, mitosis, and cell cycle regulation (**Supplementary Fig. S4C** and **S4D**). These transcriptional signatures are consistent with the more aggressive, proliferative phenotype characteristic of eHCC.

Next, a feature importance analysis algorithm (**see Methods**) was employed to select potential biomarker genes that enable good classification among different lesions. The top 10 genes are *AARS2*, *ARHGEF11*, *RABEPK*, *ATP6V0A2*, *CKAP5*, *ADK*, *CEP250*, *DCXR*, *CYP4F2* and *AADAT* (**Fig. 5A**). Notably, these key genes exhibited significant expression differences not only in the internal cohort, but also in the independent external dataset (**Fig. 5B**; **Supplementary Table S3**). Among them, *ARHGEF11*, *RABEPK*, *ATP6V0A2*, *ADK*, *CEP250*, *DCXR*, *CYP4F2* and *AADAT* showed significantly different expression patterns in HGDN and eHCC samples (**Fig. 5C**), consistent with the previously observed enrichment of pathways related to vascular smooth muscle contraction and various metabolic processes (**Supplementary Fig. S4A**, **S4B**, **S4C** and **S4D**).

Multimodal classifier integrating histopathology and molecular data

To explore whether the combination of histopathology and RNA expression can more comprehensively characterize eHCC features and improve diagnostic performance, we first integrated 34 TMC-net-based histopathological features and the expression of 10 key genes to build a multimodal model (called eLiver).

Considering that the input features were potential biomarkers identified through univariate analysis, carrying clear biological significance. Additionally, the bimodal cohort was limited in size (eHCC=15, HGDN=24) due to the preciousness of the samples. Therefore, we opted for a relatively simple logistic regression model as the classifier, to reduce the risk of overfitting and facilitate interpretability of the results (**see Methods**). The internal cohort was randomly divided into 80% training set and 20% internal test set, and this process was repeated 10 times to evaluate the robustness of the model (**Fig. 6A**; **see Methods**). The results showed that the multimodal eLiver model outperformed any single modality method in distinguishing eHCC from HGDN, with an accuracy of 0.8231 ± 0.0308 (**Fig. 6B**). In comparison, the accuracy of using only RNA features was 0.7962 ± 0.0731, and the accuracy of using only H&E features was 0.7308 ± 0.0455 (**Fig. 6B**). The multimodal model showed better generalization and noise immunity than the unimodal model. The performance of eLiver on the external test set did not degrade, achieving an accuracy of 0.8462 (**Fig. 6B**). In addition, the AUROC of the eLiver model was 0.8938 ± 0.0230 on the internal test set and 0.9250 on the external test set, both of which were higher than the classification performance of the unimodal model (**Fig. 6C**).

To explore the minimum number of molecular features required for accurate diagnosis, we performed a feature selection process within the same modeling framework (**Fig. 6A**; **see Methods**). We added RNA expression level to TMC-net-derived H&E features in order according to the importance scores of the genes and constructed multimodal models with different numbers of features. The results showed that as the number of genes increased, the accuracy and AUROC of the model on both the internal and external test sets improved, indicating that molecular data can effectively supplement H&E-based lesion diagnosis (**Fig. 6D** and **6E**). Notably, by adding the H&E features and the top 4 key genes, the AUROC increased from 0.7308 ± 0.0455 to 0.8875 ± 0.0459 on the internal test set and from 0.6154 to 0.9500 on the external test set (**Fig. 6E**). However, the accuracy and AUROC of the model on the test set did not show obvious improvement with further increase in the number of genes (**Fig. 6E**). Then, we explored the expression of four common biomarker proteins - GPC3, HSP70, GS and AFP [8]. The results showed that the corresponding genes had limited classification performance compared to our potential biomarkers, especially on the external test set (**Supplementary Fig. S5A** and

S5B). In contrast to the diagnostic outcomes based on IHC markers, our multimodal approach may serve as a complementary or alternative tool to enhance early liver cancer detection (**Supplementary Fig. S5C**). This suggests that combining H&E images and the RNA expression of the top 4 key genes (*AARS2*, *ARHGEF11*, *RABEPK*, and *ATP6V0A2*) may represent the best trade-off between diagnostic performance and actual clinical feasibility, particularly in settings where specialized pathological expertise is not readily available.

Associations between gene markers and histopathological characteristics

To further investigate the potential associations between these four key genes and histopathological features, we correlated gene expression data with histopathological features in the larger TCGA-LIHC cohort (**see Methods**). Based on the nuclei of malignant epithelial cells, lymphocytes, and fibroblasts segmented on H&E images, we calculated a series of histopathological features, including nuclear morphological heterogeneity, nuclear spatial distribution architecture, etc. Interestingly, expression level of *AARS2* and *ARHGEF11* were found to be significantly positively correlated with cellular atypia in malignant cells (**Supplementary Fig. S6A**; Pearson correlation $p < 0.001$, r = 0.29 for *AARS2* and r = 0.24 for *ARHGEF11*). *AARS2* is involved in protein synthesis, energy metabolism, and cellular respiration, while *ARHGEF11* is involved in cell signaling, migration, and adhesion - processes that are frequently dysregulated during carcinogenesis. In addition, we clustered the spatial distribution of fibroblasts by affinity propagation clustering[27,28]. The expansion of the clusters was found to be correlated with the expression level of *RABEPK*, a gene related to membrane trafficking and cell signaling (**Supplementary Fig. S6A**; Pearson correlation r = 0.16, $p < 0.001$), suggesting a possible role in tumor microenvironment remodeling. *ATP6V0A2* regulates intracellular pH homeostasis and cell signaling, and its expression level was found to be correlated with polygonal areas of the Delaunay plot constructed by malignant cells (**Supplementary Fig. S6A**; Pearson correlation r = 0.29, $p < 0.001$), suggesting that it may be involved in the spatial organization of tumor cells. These findings provide insights into the potential mechanistic links between key molecular features and tumor histopathological characteristics.

Biomarkers from eHCC indicate prognosis for a broader HCC cohort

In addition to the diagnostic performance of the multimodal model, we also sought to explore the clinical relevance of the multimodal features of H&E combined with gene expression. Specifically, we examined the association between the multimodal features and overall survival (OS) in the TCGA-LIHC cohort (**see Methods**). We found that the risk score defined by combining the expression levels of 4 genes (*AARS2*, *ARHGEF11*, *RABEPK*, and *ATP6V0A2*) was able to distinguish survival differences between patients ($p = 0.017$; **Supplementary Fig. S6B**). Moreover, the proportion of high-risk areas calculated based on the TMC-net model made the survival difference more obvious ($p = 0.0066$; **Supplementary Fig. S6C**). When we combined the expression of these 4 genes with the proportion of high-risk areas based on H&E, a significantly correlated survival difference was also obtained ($p = 0.014$; **Supplementary Fig. S6D**). These results indicate that the molecular features captured by the top 4 biomarkers as well as the risk areas derived from H&E images have strong prognostic relevance, highlighting that they could serve as potential biomarkers in the clinical management of HCC.

Discussion

Through comprehensive analysis of histological and transcriptomic characteristics, this study significantly improved the clinical diagnostic accuracy of eHCC and its precursor lesion HGDN. The innovation is that the proposed two-stage multi-scale deep learning model, TMC-net, can effectively capture the details of high-level coarse tissue structure and cell morphology in H&E images. It also possesses a certain degree of interpretability, being able to identify clinically recognized histopathological features, including portal vein invasion, which are closely associated with tumor staging and prognosis [29]. This modeling approach offers a new method for optimizing feature extraction in complex pathological diagnostic scenarios. Furthermore, this study demonstrates that using AI-generated risk heatmaps can improve the diagnostic accuracy and consistency of junior pathologists, providing an avenue for AI-assisted collaboration in clinical practice.

In addition, through comparative transcriptome analysis, the study identified molecular markers that can distinguish eHCC from HGDN. These biomarkers may facilitate future clinical detection through reverse transcription polymerase chain reaction (RT-PCR) [30]. Furthermore, we linked and integrated gene expression from 10 genes with H&E features extracted from deep learning models, resulting in multimodal biomarkers with clinical value. These biomarkers not only demonstrated utility for the eHCC screening but were also found to be associated with the prognosis of HCC patients, highlighting their potential for enabling personalized management of HCC. Overall, this multimodal modeling framework could also serve as a feasible approach to address other complex clinical diagnostic challenges.

This study also has the limitations. First, the limited sample size may restrict the further optimization and generalizability of the diagnostic model and molecular biomarkers. The diagnoses of these two diseases are very challenging in clinical practice, resulting in a very limited number of cases. Future studies should encourage collaborative multicenter research to increase sample sizes and better validate the robustness and clinical applicability. In addition, we conducted a preliminary exploration of the associations between different modalities, but future research should incorporate clinical information and imaging features, such as computed tomography. These cross-modal relationships can help identify new biomarkers to support large-scale screening, improve the feasibility of clinical trial enrollment, and facilitate the development of less invasive diagnostic alternatives.

Methods

Patients and H&E images

The internal cohort of this study was from Zhongshan Hospital Fudan University (ZS; Shanghai, China) and contained H&E slides of patients diagnosed with eHCC or DN between 2017 and 2023. These slides were reviewed by at least two senior pathologists specializing in liver diseases and supplemented with clinical information and immunohistochemistry (IHC) examination of proteins such as Ki67, HSP70, GS, and GPC3. eHCC is clearly defined as HCC in the early stage of malignancy (generally ≤ 2 cm), which is usually sparsely vascularized, irregularly bordered, and contains portal elements on diagnostic imaging, but has no significant impact on the original architecture of the liver [3]. HGDN is characterized by hepatocellular hyperplasia with atypical cytological and/or architectural features that are insufficient for diagnosis of HCC. HGDN shows higher cellularity and often exhibits small cell changes [31].

These slides were reviewed by at least two senior pathologists specializing in liver diseases, and IHC examination

of multiple proteins was performed when necessary. All H&E slides confirmed as eHCC or HGDN were retained and scanned using Aperiod AT2 (500 nm/px) or Nano Zoomer S360 (220 nm/px) digital slide scanners. Because most HCC patients were in advanced stages at the time of diagnosis, only 112 high-quality slides from 75 patients in the ZS were included in the study, resulting in the generation of 73,891 tiles (**Supplementary table S1**). In addition, 7 eHCC or HGDN slides from 6 patients from Huashan Hospital Fudan University (HS; Shanghai, China) were included as independent external datasets, resulting in the generation of 3,430 tiles (**Supplementary table S1**).

All patients provided informed consent. The use of clinical samples and all of the procedures were approved by the Zhongshan Hospital Fudan University Research Ethics Committee (B2024-340) and Huashan Hospital Fudan University Research Ethics Committee (2017M-003). The study conformed to the principles of the Helsinki Declaration.

Each H&E image was loaded and tiled into non-overlapping tiles of 512×512 μm (corresponding to 1024 px for Aperiod AT2 and 2327 px for Nano Zoomer S360) using openslide (v1.1.2). The background or blank images were deleted according to the intensity of the RGB channels, and only the tiles containing more than 80% tissue pixels were retained for analysis. Pathologists annotated the images using Qupath [32]. We extracted the coordinates from the annotation files using xml in Python (v3.9.16), shapely (v2.0.1) and geojson (v3.0.1).

Design of TMC-net

TMC-net is a novel two-stage, multi-scale convolutional network designed to emulate the diagnostic workflow of pathologists. In clinical practice, pathologists first assess overall morphological patterns at low magnification before conducting a detailed examination of localized regions at higher magnification. TMC-net replicates this process by incorporating two different magnification levels as input data for each tile[33].

Specifically, the tile at 20x magnification is denoted as X_0, and used to input into the Low-resolution convolutional neural network (CNN) to extract global features. While the corresponding tile at 40x magnification is subdivided into four sub-regions: upper-left, upper-right, lower-left, and lower-right, denoted as X_1, X_2, X_3 and X_4, and input into the high-resolution CNN to extract detail features.

Specifically, a tile at 20× magnification, denoted as X_0, serves as input to the Low-resolution CNN to extract global features. In parallel, the corresponding tile at 40× magnification is subdivided into four sub-regions: upper-left, upper-right, lower-left, and lower-right, denoted as X_1, X_2, X_3 and X_4. These subregions are processed by the high-resolution CNN to extract fine-grained features.

The multi-scale architecture of TMC-net is built upon ResNet18, where both low-resolution features (W_0) and high-resolution features (W_1, W_2, W_3, W_4) are learned through multi-layer CNNs. These high-resolution representations are subsequently concatenated with the global representation W_0, and the final prediction $\hat{y}$ is generated using a multilayer perceptron (MLP) applied to the concatenated feature vector. The entire network is trained in an end-to-end manner, ensuring the joint optimization of all layers. The mathematical formulation of the model is as follows:

$$W_0 = CNN_{low-resolution}(X_0), \tag{1}$$

$$W_l = CNN_{high-resolution}(X_l), l = 1,2,3,4 \tag{2}$$

$$\hat{y} = MLP([W_0; W_1; W_2; W_3; W_4]) \tag{3}$$

In order to explore the most suitable input data feature type for early liver cancer diagnosis, this study evaluated the impact of multi-scale features under different field of view sizes on model performance, including 256 μm, 512 μm and 1024 μm. The original tile was first scaled to 448×448 px and cut into four parts to obtain a 224×224 px high-power field

of view image. All field of view images are eventually scaled to 224×224 px input model

To assess the contributions of the multi-scale CNN and the two-stage framework, we conducted ablation studies under various network configurations. The baseline model TMC-net, which integrates both the multi-scale CNN and two-stage modeling, is denoted as "TwoStage_MultiScale". Variations of the model were constructed as follows:

1) 'TwoStage_SingleScale' - This variant retains the two-stage framework but replaces the multi-scale CNN with a single-scale CNN that processes only 20× magnification images.
2) 'OneStage_MultiScale' - This variant preserves the multi-scale CNN but adopts a one-stage modeling framework, directly classifying regions as non-critical areas, HGDN, or eHCC, without a candidate region extraction step.
3) 'OneStage_SingleScale' - This model, representing the most conventional approach, employs a single-scale CNN within a one-stage framework.

Other histopathological feature extraction methods

The proposed TMC-net was compared with 1) lightweight CNN, 2) the pathology foundation model and 3) the machine learning model based on traditional cell-level features.

1) We replace the basic ResNet18 neural network in TMC-net with a lightweight CNN SqueezeNet to evaluate the performance of the model with a smaller number of parameters. Because SqueezeNet performs poorly, 'TMC-net' reported in the Results are all ResNet18 neural networks (**Supplementary Table S4**).

2) The PLIP model is a multimodal pathology-based basic model based on contrastive learning, which has been trained on OpenPath. The PLIP image encoder has been shown to be the preferred pre-trained backbone for various pathology image classification tasks. Therefore, this study used the PLIP image encoder to extract the feature representation of small tiles in 512 dimensions, and we trained two commonly used machine learning classifiers (logistic regression and random forest) for the classification of other non-critical regions, HGDN and eHCC.

3) Cell density ratio and cellular atypia index are two common quantifiable cellular features. In each annotated region on each WSI image, we randomly selected 100 tiles, and if there were less than 100 tiles, we selected all of them. The previously published cell segmentation classification network[34] was used to segment epithelial cells on these tiles. For nuclear density, it was defined as the number of cells divided by the area. For the morphology atypia [35], the score of epithelial cells in tile i was defined as s_i. Then the cell morphology heterogeneity score at the WSI level is defined as $\max(s_i)$.

$$s_i = SD\left(\frac{p_{cj}}{2\sqrt{\pi a_{cj}}}\right), j \in \{1, \ldots, n\}, \quad (4)$$

where SD is the standard deviation function, p_{cj} is the perimeter of the nucleus of the jth epithelial cell, a_{cj} is the nuclear area of the jth epithelial cell, and n is the total number of epithelial cells on tile i. These features were input into logistic regression to obtain tile-level prediction results.

Aggregation of classifications at the tile level to the slide level

To obtain the slide-level prediction, the output from the models (such as TMC-net and the above-mentioned multiple logistic regression classifiers) were further integrated. Specifically, the predicted labels (HGDN, eHCC or others) of all tiles and the corresponding probabilities were used to calculate the statistical features of each WSI. Each WSI is

represented by a 34-dimensional feature vector Z. These features include: the frequency of each label category (e.g., the number of tiles with HGDN), the proportion of each label category (e.g., the proportion of tiles with HGDN), the predicted probability distribution of each label category (e.g., mean, standard deviation, and quantile). The full list of features is shown in **Supplementary Table S2**.

To diagnose at the WSI level, we built a support vector machine (SVM) classifier based on the statistical features of tiles to predict whether the WSI is HGDN or eHCC. Specifically, SVM outputs the prediction of the WSI based on the 34-dimensional feature vector **Z**.

$$Y = SVM(Z) \tag{5}$$

Model training and evaluation

In TMC-net, we implemented two neural networks: one for candidate lesion region extraction and another for eHCC detection, both based on multi-scale CNN architectures. The Resnet18 network was initialized with the best set of parameters obtained from training on the ImageNet dataset. To enable the networks to learn pathology-specific features, all layers were unfrozen during training. Each network was trained for 50 epochs on an Nvidia 4090 Ti GPU, with a batch size of 128. The cross-entropy loss function was used, and the Adam optimizer was employed with a learning rate of 5×10^{-6}, a first-order moment estimates decay rate of 0.9, and a second-order moment estimate decay rate of 0.999. To enhance the diversity of the training data, random vertical, horizontal flips and random rotations were applied. The deep learning models were trained using the PyTorch (v1.13.1), while scikit-learn (v1.3.2) was utilized for training machine learning models and evaluating.

All tile images were processed through the candidate lesion region extraction network, with a threshold of 0.5. Tiles with a probability greater than 0.5 were classified as candidate regions, while others were deemed irrelevant. The candidate region tiles were subsequently input into the eHCC detection network, which also employed a 0.5 threshold. Tiles with a probability greater than 0.5 were classified as eHCC, and the others were HGDN.

HE models were evaluated using the following methods: Tile-level performance was evaluated on the internal validation dataset using LOOCV. Slide-level performance was evaluated not only on the internal validation set, but also on the internal test set and an independent external test set. Evaluation metrics used included AUROC, accuracy, and F1 score. The LOOCV process was repeated 10 times, and we report the average performance of each method along with 95% confidence intervals.

Model interpretability by saliency maps

The Grad-CAM++ class activation maps (CAMs) are generated by visualizing the weighted gradients flowing into the final convolutional layer of the network (referred to as the feature layer A). Since the feature maps of the final convolutional layer contain high-level semantic information, and spatial information is lost in the subsequent fully connected layers, this layer is chosen as the optimal feature layer for visualization. Compared to the traditional CAM, Grad-CAM[25] does not require modifications to the network or retraining, making it more convenient to use. Grad-CAM++ improves upon Grad-CAM by not only considering positive gradients but also incorporating additional weight factors to more precisely reflect the contribution of each pixel to the class prediction.

Based on the fundamental assumption that the prediction score Y^c for class c is a linear combination of the globally averaged feature maps A^k of the feature layer:

$$Y^c = \sum_k w_k^c . \sum_i \sum_j A_{ij}^k . \tag{6}$$

In Grad-CAM++, the weight w_k^c is a weighted average designed to capture the importance of the feature map A^k:

$$w_k^c = \sum_i \sum_j \alpha_{ij}^{kc}.relu(\frac{\partial Y^c}{\partial A_{ij}^k}), \quad (7)$$

where $\frac{\partial Y^c}{\partial A^k}$ represents the gradient of class c with respect to the k-th channel of the feature layer A. The weight α^{kc} is the importance of each pixel in the k-th channel of feature layer A, derived from the second-order partial derivatives:

$$\alpha_{ij}^{kc} = \frac{\frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2}}{2\frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2} + \sum_a \sum_b A_{ab}^k \{\frac{\partial^3 Y^c}{(\partial A_{ij}^k)^3}\}} \quad (8)$$

Thus, the saliency map for an image L can be computed as a linear combination of the feature maps:

$$L_{ij}^c = relu(\sum_k w_k^c.A_{ij}^k). \quad (9)$$

Differential expression and functional analysis of transcriptome

The RNA-seq data was obtained from a collaborative study (Z.Z., H.L., L.C., Y.J. and L.H., unpublished data). The count data included 60 samples from the ZS cohort and 26 samples from the HS cohort, comprising a total of 36 eHCC tissues, 34 HGDN tissues, and 16 surrounding liver tissues. To address potential batch effects associated with differences in sequencing time, we applied the combat_seq() function from the sva package (v3.38) for batch correction, with the sample groups (HGDN, eHCC, and control tissue of surrounding liver) set as the holdout parameter. We confirmed the effectiveness of batch removal using PCA and proceeded with the corrected expression profiles for downstream modeling.

Differential expression analysis was conducted using DESeq2 (v1.30) on the count data. The DEGs between all groups were compared as one-to-other. For the DEGs and functional pathways of HGDN and eHCC, only the two were compared. Genes were considered significantly differentially expressed if they had an adjusted p-value < 0.05 (Wald test, false discovery rate method). The gene sets were sourced from the public MSigDB database, including Hallmark gene sets and KEGG pathways, and analyzed using the ClusterProfiler (v4.0.5) package.

Screening of potential gene markers

Given the effectiveness of XGBoost in handling high-dimensional expression data, its ability to capture non-linear relationships between gene expression values, and its robustness to noise and outliers, we employed an XGBoost machine learning model to analyze the internal dataset. Within the ensemble learning framework, each base estimator (i.e., the XGBoost model) independently computed feature importance using the 'estimator.feature_importances_' attribute. During the modeling process, the internal cohort was split into 80% training and 20% testing sets. The model was trained on the training data, and feature importance scores were calculated. This data partitioning and model training were performed 10 times randomly. The importance scores obtained each time were summed to obtain a combined feature importance ranking. The top 10 genes with the highest cumulative importance scores were selected as potential transcriptome biomarkers.

In the model building and evaluation stages, the same data split method was used. Evaluation metrics used included AUROC, accuracy, and F1 score.

Combining images with different numbers of genes

The model construction pipeline is illustrated in Figure 6A and includes four parts: feature combination,

hyperparameter tuning, model refitting, and model evaluation.

The cohort was split into train cohort (80%) and test cohort (20%) for the classification task. The normalized features were concatenated together. We extracted a 34-dimensional feature vector for each WSI from the TMC-net. In the process of integrating RNA expression profiles, we added one gene at a time according to the order of importance score from high to low. This process generated 11 logistic regression models based on different feature combinations: 34-dimensional pathology features, pathology and top 1 gene, pathology and top 2 genes, …, pathology and all top 10 genes. All models followed the same machine learning training and evaluation framework. To balance the calculation burden and the model accuracy, we used the randomized search cross validation strategy to search the optimal hyperparameter, setting the roc_auc as the scoring function and 5-fold cross validation. In the model refitting and model evaluation phases, to evaluate the model performance, we randomly re-split cohort and refit the model for 10 times, and evaluate the AUROC, balanced accuracy, F1, precision, and recall on the test cohort. The machine learning pipeline was built in Python (v3.9.15) using the following libraries: scikit-learn (v1.2.0), numpy (v1.24.1), scipy (v1.9.3), and pandas (v1.5.2).

Baseline model based on IHC markers

The genes corresponding to GS, AFP, GPC3, and HSP70 proteins are *GLUL*, *AFP*, *GPC3*, *HSPA2*/*HSPA1A*/*HSPE1*. Among them, *GLUL, AFP*, *GPC3*, and *HSPA2* showed significant differential expression between the eHCC and HGDN groups, so they were combined with the image features extracted by TMC-net to establish a multimodal baseline model.

In addition, considering the impact of post-transcriptional regulation, we also collected the positive or negative information of these proteins in the samples of bimodal cohort. Under the guidance of pathology experts, we considered the possibility of HCC to be high when two or more of these protein markers were positive. Based on this criterion, a simple logistic classifier was established as another baseline model.

Association between histopathological features and gene expression

To identify histopathological features associated with gene expression, we computed histopathological features for samples from the TCGA-LIHC cohort. First, H&E tissue images were tiled into 512 μm. These tiles were then fed into the previously published TSFD-Net model[34] to perform cell segmentation and classification. The centroids of cells labeled as malignant epithelial cells were extracted. For each tile, if the proportion of malignant epithelial cells exceeded 80%, the tile was considered to represent a tumor region. histopathological feature computation was then performed only within these tumor regions. Using the centroids of malignant epithelial cells, we performed affinity propagation clustering (using the sklearn.cluster module, v0.24.1) and constructed Delaunay plots (using the scipy.spatial module, v1.9.3). In addition, we computed the morphological heterogeneity of nuclei based on the segmented nuclear contours (using numpy, v1.24.1).

For patients with multiple H&E slides, the computed H&E features were averaged to obtain a representative set of features for each patient. In total, features of 358 patient slides were successfully calculated. These normalized H&E features were then subjected to Pearson correlation analysis with gene expression data using the scipy.stats module (v1.6.1) to determine the association between histopathological patterns and expression of key genes.

Survival analysis

Considering that integrating H&E features with *AARS2*, *ARHGEF11*, *RABEPK* and *ATP6V0A2* can detect eHCC

well, we further explored the association between these four genes and prognosis, including 1) individual gene, 2) the combination of the four key genes, and 3) the multimodal feature values. These analyses were performed by the TCGA-LIHC dataset. The H&E images, RNA expression clinical data were downloaded from cBioPortal. Samples with missing values are deleted.

1) The effect of individual genes on survival was analyzed using the Cox Proportional Hazards Model.

2) For the gene combination, considering that integrating H&E signature with *AARS2*, *ARHGEF11*, *RABEPK* and *ATP6V0A2* can detect eHCC well, we defined a risk score for these four genes [36]:

$$Risk\ Score = \sum_{i=1}^{3} \beta_i * Exp_i, \quad (10)$$

where *Exp* is the expression level of each gene, and β is its regression coefficient obtained from the single-variate Cox model. The TCGA-LIHC patients were divided into high-risk and low-risk groups based on the median risk score. The overall survival rates in the two groups were compared using Kaplan–Meier curves and log-rank tests in the R package survminer and survival.

3) For the multimodal feature, we extracted 'predicted HCC area ratio' from TMC-net based on H&E images. If a patient has multiple slides, choose the one with the larger value. Similarly, we defined a risk score for 'predicted HCC area ratio' (β = 1), *AARS2*, *ARHGEF11*, *RABEPK* and *ATP6V0A2* expression using the same approach and performed survival analysis.

Data availability

The pathology data of this study are available upon reasonable request from the corresponding author (ji.yuan@zs-hospital.sh.cn). Data of the TCGA-LIHC cohort can be downloaded from cBioPortal (http://www.cbioportal.org/). The raw RNA sequence data reported in this paper have been deposited in the GSA database (HRA008950, https://ngdc.cncb.ac.cn/gsa-human/).

Code availability

Custom code related to the image preprocessing, extraction, model builder and model evaluation was written in Python. Expression profile analysis was performed by R. These codes are available in https://github.com/LiHongCSBLab/AI-model-for-eHCC

Acknowledgements

This work was supported by National Natural Science Foundation of China (T2122018, 32470707 to H.L. and 32300555 to P.L.), Shanghai Post-doctoral Excellence Program (2025190), Shanghai Municipal Science and Technology Major Project, Shanghai Sailing Program (22YF1458000 to P.L.), and Shanghai Municipal Science and Technology Commission (Project Number: 22JC1403002 to Y.J.)

Author Contributions

H.L. and Y.J. designed and supervised the research work. X.Z. and Xin.L. conducts data application and agreement signing. S.J., H.W., Y.W. and Z.Z. collected the original slides and scan all the hepatocellular carcinoma slides. H.W., C.L. manually labelled whole-slide images, and Y.J. checked the annotation. J.C. and Y.B. wrote the code for digital pathology workflow. Xue.L. and S.J. built models and led the deep learning algorithm development and evaluation. L.M., X.F., Y.Z., and C.W. participated in the evaluation of AI-assisted diagnosis as pathologists. P.L. and X.Y. provided molecular analysis workflow, J.S., Xue.L. and H.L. wrote the manuscript.

Competing interests

The authors declare no competing financial or non-financial interests.

Reference

1. Llovet, J.M., Kelley, R.K., Villanueva, A., Singal, A.G., Pikarsky, E., Roayaie, S., Lencioni, R., Koike, K., Zucman-Rossi, J., and Finn, R.S. (2021). Hepatocellular carcinoma. Nat Rev Dis Primers 7, 6. 10.1038/s41572-020-00240-3.
2. Wang, W., and Wei, C. (2020). Advances in the early diagnosis of hepatocellular carcinoma. Genes Dis 7, 308-319. 10.1016/j.gendis.2020.01.014.
3. Kudo, M. (2013). Early hepatocellular carcinoma: definition and diagnosis. Liver Cancer 2, 69-72. 10.1159/000343842.
4. Pathologic diagnosis of early hepatocellular carcinoma: a report of the international consensus group for hepatocellular neoplasia. (2009). Hepatology 49, 658-664. 10.1002/hep.22709.
5. Beaufrère, A., Paisley, S., Ba, I., Laouirem, S., Priori, V., Cazier, H., Favre, L., Cauchy, F., Lesurtel, M., Calderaro, J., et al. (2024). Differential diagnosis of small hepatocellular nodules in cirrhosis: surrogate histological criteria of TERT promoter mutations. Histopathology 84, 473-481. 10.1111/his.15086.
6. Sato, T., Kondo, F., Ebara, M., Sugiura, N., Okabe, S., Sunaga, M., Yoshikawa, M., Suzuki, E., Ogasawara, S., Shinozaki, Y., et al. (2015). Natural history of large regenerative nodules and dysplastic nodules in liver cirrhosis: 28-year follow-up study. Hepatol Int 9, 330-336. 10.1007/s12072-015-9620-6.
7. Ojima, H., Masugi, Y., Tsujikawa, H., Emoto, K., Fujii-Nishimura, Y., Hatano, M., Kawaida, M., Itano, O., Kitagawa, Y., and Sakamoto, M. (2016). Early hepatocellular carcinoma with high-grade atypia in small vaguely nodular lesions. Cancer Sci 107, 543-550. 10.1111/cas.12893.
8. Di Tommaso, L., Destro, A., Seok, J.Y., Balladore, E., Terracciano, L., Sangiovanni, A., Iavarone, M., Colombo, M., Jang, J.J., Yu, E., et al. (2009). The application of markers (HSP70 GPC3 and GS) in liver biopsies is useful for detection of hepatocellular carcinoma. J Hepatol 50, 746-754. 10.1016/j.jhep.2008.11.014.
9. Di Tommaso, L., Franchi, G., Park, Y.N., Fiamengo, B., Destro, A., Morenghi, E., Montorsi, M., Torzilli, G., Tommasini, M., Terracciano, L., et al. (2007). Diagnostic value of HSP70, glypican 3, and glutamine synthetase in hepatocellular nodules in cirrhosis. Hepatology 45, 725-734. 10.1002/hep.21531.
10. Bolondi, L., Gaiani, S., Celli, N., Golfieri, R., Grigioni, W.F., Leoni, S., Venturi, A.M., and Piscaglia, F. (2005).

Characterization of small nodules in cirrhosis by assessment of vascularity: the problem of hypovascular hepatocellular carcinoma. Hepatology 42, 27-34. 10.1002/hep.20728.

11. Kojiro, M. (2004). Focus on dysplastic nodules and early hepatocellular carcinoma: an Eastern point of view. Liver Transpl 10, S3-8. 10.1002/lt.20042.
12. Nagtegaal, I.D., Odze, R.D., Klimstra, D., Paradis, V., Rugge, M., Schirmacher, P., Washington, K.M., Carneiro, F., and Cree, I.A. (2020). The 2019 WHO classification of tumours of the digestive system. Histopathology 76, 182-188. 10.1111/his.13975.
13. Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyo, D., Moreira, A.L., Razavian, N., and Tsirigos, A. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med 24, 1559-1567. 10.1038/s41591-018-0177-5.
14. Gehrung, M., Crispin-Ortuzar, M., Berman, A.G., O'Donovan, M., Fitzgerald, R.C., and Markowetz, F. (2021). Triage-driven diagnosis of Barrett's esophagus for early detection of esophageal adenocarcinoma using deep learning. Nat Med. 10.1038/s41591-021-01287-9.
15. Lu, M.Y., Chen, T.Y., Williamson, D.F.K., Zhao, M., Shady, M., Lipkova, J., and Mahmood, F. (2021). AI-based pathology predicts origins for cancers of unknown primary. Nature 594, 106-110. 10.1038/s41586-021-03512-4.
16. Calderaro, J., Seraphin, T.P., Luedde, T., and Simon, T.G. (2022). Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. J Hepatol 76, 1348-1361. 10.1016/j.jhep.2022.01.014.
17. Saillard, C., Schmauch, B., Laifa, O., Moarii, M., Toldo, S., Zaslavskiy, M., Pronier, E., Laurent, A., Amaddeo, G., Regnault, H., et al. (2020). Predicting Survival After Hepatocellular Carcinoma Resection Using Deep Learning on Histological Slides. Hepatology 72, 2000-2013. 10.1002/hep.31207.
18. Zeng, Q., Klein, C., Caruso, S., Maille, P., Laleh, N.G., Sommacale, D., Laurent, A., Amaddeo, G., Gentien, D., Rapinat, A., et al. (2022). Artificial intelligence predicts immune and inflammatory gene signatures directly from hepatocellular carcinoma histology. J Hepatol 77, 116-127. 10.1016/j.jhep.2022.01.018.
19. Wang, H., Jiang, Y., Li, B., Cui, Y., Li, D., and Li, R. (2020). Single-Cell Spatial Analysis of Tumor and Immune Microenvironment on Whole-Slide Image Reveals Hepatocellular Carcinoma Subtypes. Cancers (Basel) 12. 10.3390/cancers12123562.
20. Gao, Q., Shi, J.Y., Wang, X., Ding, G.Y., Dong, Z., Han, J., Guan, Z., Ma, L.J., Zheng, Y., Zhang, L., et al. (2020). Exploring prognostic indicators in the pathological images of hepatocellular carcinoma based on deep learning. Gut. 10.1136/gutjnl-2020-320930.
21. Cheng, N., Ren, Y., Zhou, J., Zhang, Y., Wang, D., Zhang, X., Chen, B., Liu, F., Lv, J., Cao, Q., et al. (2022). Deep Learning-Based Classification of Hepatocellular Nodular Lesions on Whole-Slide Histopathologic Images. Gastroenterology 162, 1948-1961 e1947. 10.1053/j.gastro.2022.02.025.
22. Boehm, K.M., Khosravi, P., Vanguri, R., Gao, J., and Shah, S.P. (2022). Harnessing multimodal data integration to advance precision oncology. Nat Rev Cancer 22, 114-126. 10.1038/s41568-021-00408-3.
23. Chen, R.J., Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Lipkova, J., Noor, Z., Shaban, M., Shady, M., Williams, M., Joo, B., and Mahmood, F. (2022). Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer Cell 40, 865-878.e866. 10.1016/j.ccell.2022.07.004.
24. Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T.J., and Zou, J. (2023). A visual-language foundation model for pathology image analysis using medical Twitter. Nat Med 29, 2307-2316. 10.1038/s41591-023-02504-3.
25. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 22-29 Oct. 2017. pp. 618-626.

26. Choi, K.K., Kim, S.H., Choi, S.B., Lim, J.H., Choi, G.H., Choi, J.S., and Kim, K.S. (2011). Portal venous invasion: the single most independent risk factor for immediate postoperative recurrence of hepatocellular carcinoma. J Gastroenterol Hepatol 26, 1646-1651. 10.1111/j.1440-1746.2011.06780.x.
27. Frey, B.J., and Dueck, D. (2007). Clustering by passing messages between data points. Science 315, 972-976. 10.1126/science.1136800.
28. Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K.R., Zhao, T., Batiste, R., et al. (2018). Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. Cell Reports 23, 181-193.e187. 10.1016/j.celrep.2018.03.086.
29. Sun, Y.F., Wu, L., Liu, S.P., Jiang, M.M., Hu, B., Zhou, K.Q., Guo, W., Xu, Y., Zhong, Y., Zhou, X.R., et al. (2021). Dissecting spatial heterogeneity and the immune-evasion mechanism of CTCs by single-cell RNA-seq in hepatocellular carcinoma. Nat Commun 12, 4091. 10.1038/s41467-021-24386-0.
30. Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., et al. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351, 2817-2826. 10.1056/NEJMoa041588.
31. Vij, M., and Calderaro, J. (2021). Pathologic and molecular features of hepatocellular carcinoma: An update. World J Hepatol 13, 393-410. 10.4254/wjh.v13.i4.393.
32. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open source software for digital pathology image analysis. Sci Rep 7, 16878. 10.1038/s41598-017-17204-5.
33. Xiang, H., Shen, J., Yan, Q., Xu, M., Shi, X., and Zhu, X. (2023). Multi-scale representation attention based deep multiple instance learning for gigapixel whole slide image analysis. Med Image Anal 89, 102890. 10.1016/j.media.2023.102890.
34. Ilyas, T., Mannan, Z.I., Khan, A., Azam, S., Kim, H., and De Boer, F. (2022). TSFD-Net: Tissue specific feature distillation network for nuclei segmentation and classification. Neural Netw 151, 1-15. 10.1016/j.neunet.2022.02.020.
35. Heindl, A., Khan, A.M., Rodrigues, D.N., Eason, K., Sadanandam, A., Orbegoso, C., Punta, M., Sottoriva, A., Lise, S., Banerjee, S., and Yuan, Y. (2018). Microenvironmental niche divergence shapes BRCA1-dysregulated ovarian cancer morphological plasticity. Nat Commun 9, 3917. 10.1038/s41467-018-06130-3.
36. Jing, S.Y., Liu, D., Feng, N., Dong, H., Wang, H.Q., Yan, X., Chen, X.F., Qu, M.C., Lin, P., Yi, B., et al. (2024). Spatial multiomics reveals a subpopulation of fibroblasts associated with cancer stemness in human hepatocellular carcinoma. Genome Med 16, 98. 10.1186/s13073-024-01367-8.

Figure Legends

Figure 1| Overview of the data and multimodal modeling framework for eHCC detection.

A. Schematic diagram of feature extraction and classification modeling from multimodal data for early liver cancer detection.

B. An example of diagnosis based on hematoxylin-eosin (H&E) images and immunohistochemistry (IHC) of marker proteins.

C. Cohorts and datasets for modeling.

Figure 2| Predicting early HCC by the TMC-net model from pathological features.

A. The schematic diagram of TMC-net for classifying eHCC and HGDN. Left, the aggregation method from tile prediction to WSI-level result. Right, two-stage and multi-scale convolutional network for tile-level prediction. WSI, whole slide image; SVM, support vector machine.

B. PCA plot of deep learning-based pathological features at the tile-level.

C. Tile-level confusion matrix (left) and AUROC (right) of the TMC-net through leave-one-out cross-validation.

D. Slide-level confusion matrix of the TMC-net on the validation and test datasets.

Figure 3| Comparison TMC-net with other methods.

A. ROC curves of slide-level prediction for different models. These models were trained by pathological features extracted from different strategies in Figure 1A.

B. Accuracy and AUROC of slide-level prediction for different models.

Figure 4 | Interpretability of TMC-net prediction.

A. Visualization of the typic prediction results. The red circle in the first row draws the lesion area annotated by the pathologist. The second row shows the probability heatmaps that illustrate the predicted probability of the corresponding category by TMC-net.

B. Magnified tile area in the eHCC lesion and the corresponding fine-grained visualization heatmap. The redder the color, the more important the area is to TMC-net. Pathological feature descriptions are annotated below.

C. Same for (B), but for HGDN lesion.

D. Same for (B), but for WD-HCC lesion.

Figure 5| Comparative analyzes in transcriptome between eHCC and HGDN.

A. Feature importance scores and corresponding functions of the 10 key genes obtained based on the feature importance analysis algorithm.

B. Expression patterns of 10 key genes on internal (left) and external cohorts (right).

C. Expression levels of the 10 key genes in different lesions.

Figure 6| Multimodal data integration facilitates early HCC diagnosis.

A. Strategies for building multimodal models by integrating pathological and gene features.

B. Compare the accuracy of unimodal models (e.g., H&E model and RNA model) and multimodal models (combining HE and RNA) on the test set.

C. Same for (B), but for AUROC.

D. The accuracy of the model established by combining images with different numbers of genes.

E. Same for (D), but for AUROC.

Figure 1

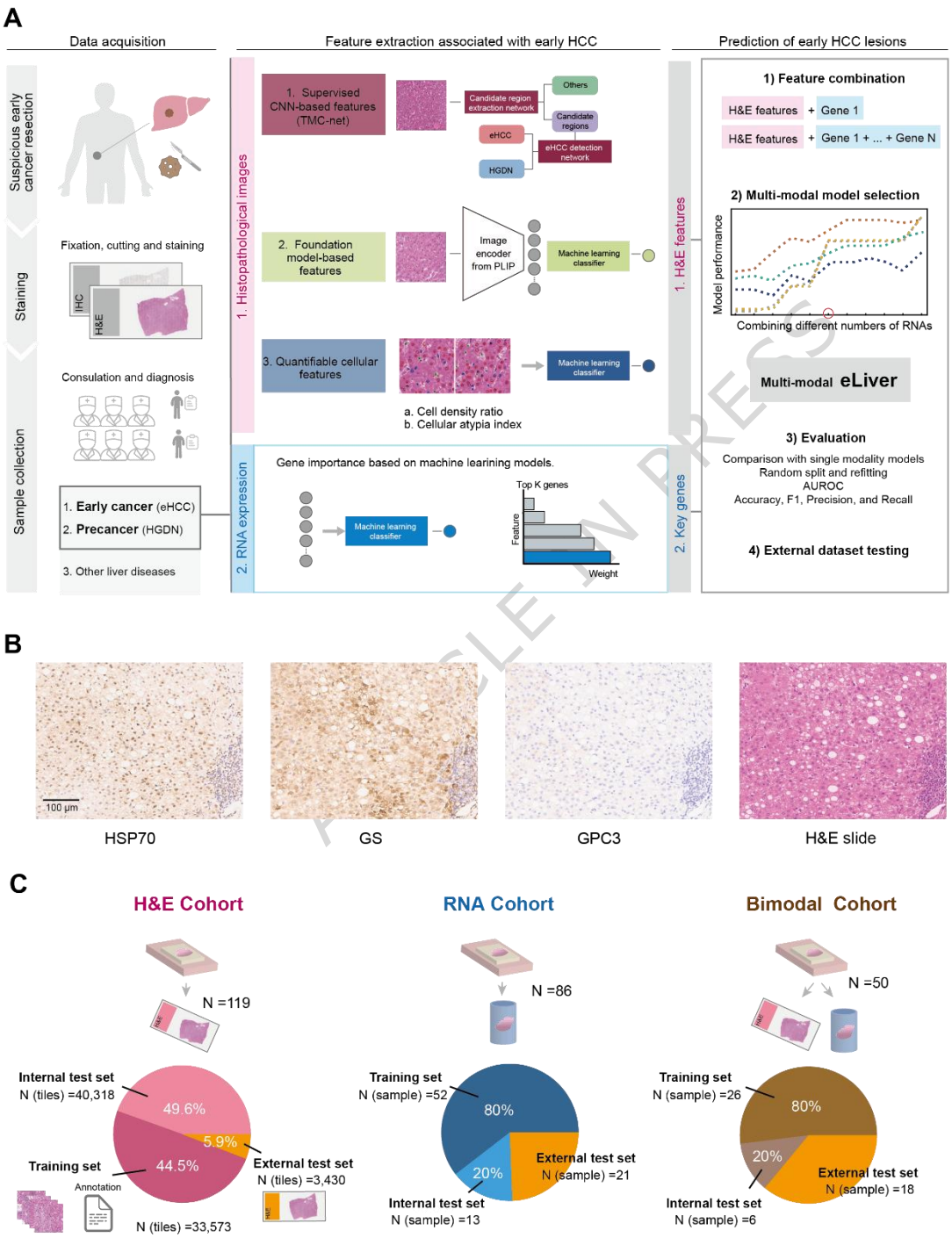

A
Data acquisition
Feature extraction associated with early HCC
Prediction of early HCC lesions
Suspicious early cancer resection
Staining
Sample collection
Fixation, cutting and staining
IHC
H&E
Consulation and diagnosis
1. Early cancer (eHCC)
2. Precancer (HGDN)
3. Other liver diseases
1. Histopathological images
1. Supervised CNN-based features (TMC-net)
Candidate region extraction network
Others
Candidate regions
eHCC
eHCC detection network
HGDN
2. Foundation model-based features
Image encoder from PLIP
Machine learning classifier
3. Quantifiable cellular features
Machine learning classifier
a. Cell density ratio
b. Cellular atypia index
2. RNA expression
Gene importance based on machine learining models.
Machine learning classifier
Top K genes
Feature
Weight
1. H&E features
2. Key genes
1) Feature combination
H&E features + Gene 1
H&E features + Gene 1 + ... + Gene N
2) Multi-modal model selection
Model performance
Combining different numbers of RNAs
Multi-modal eLiver
3) Evaluation
Comparison with single modality models
Random split and refitting
AUROC
Accuracy, F1, Precision, and Recall
4) External dataset testing
B
100 μm
HSP70
GS
GPC3
H&E slide
C
H&E Cohort
N =119
Internal test set
N (tiles) =40,318
49.6%
5.9%
44.5%
Training set
Annotation
External test set
N (tiles) =3,430
N (tiles) =33,573
RNA Cohort
N =86
Training set
N (sample) =52
80%
20%
External test set
N (sample) =21
Internal test set
N (sample) =13
Bimodal Cohort
N =50
Training set
N (sample) =26
80%
20%
External test set
N (sample) =18
Internal test set
N (sample) =6

Figure 2

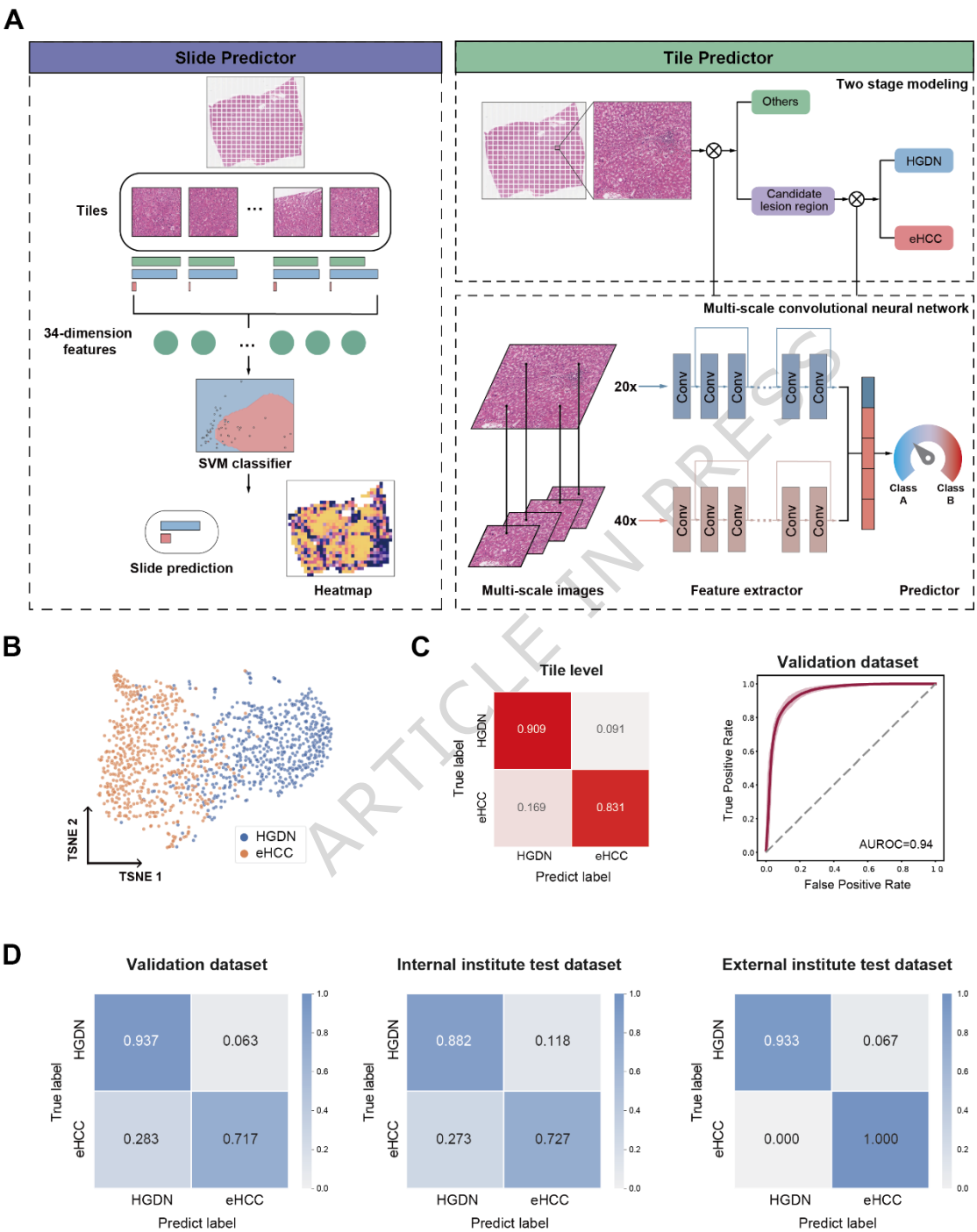


Figure 3

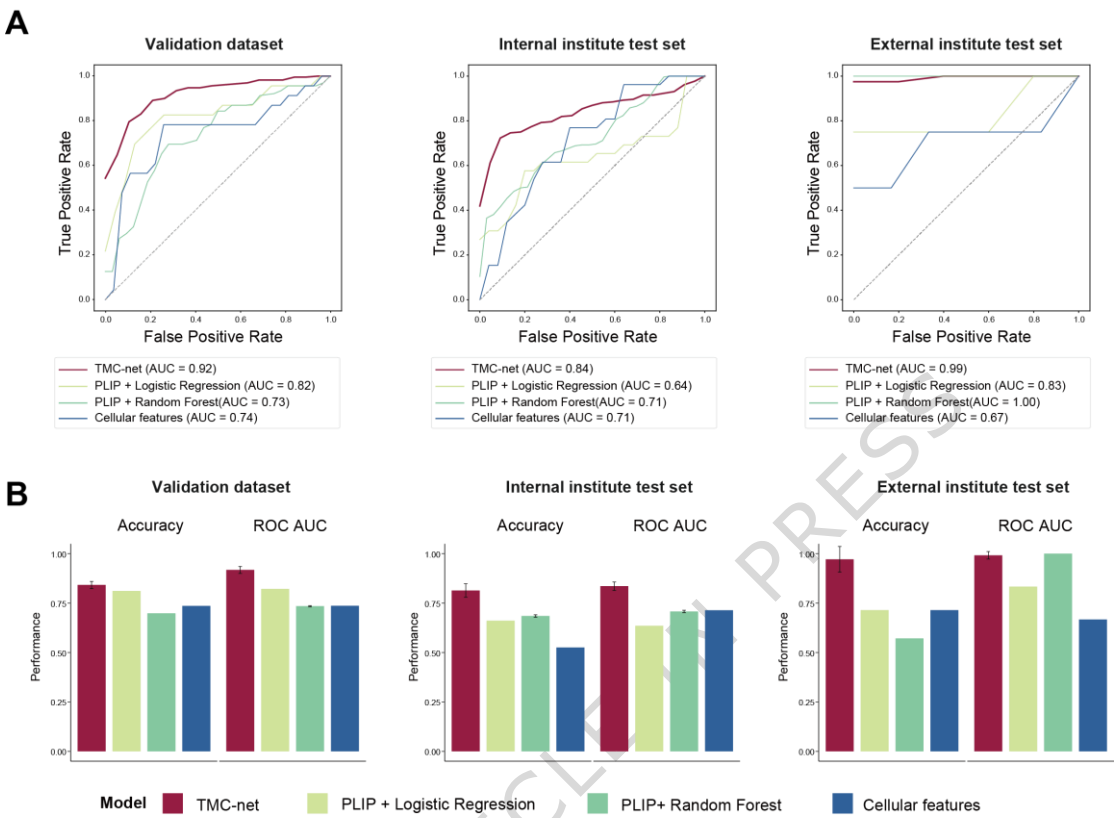


Figure 4

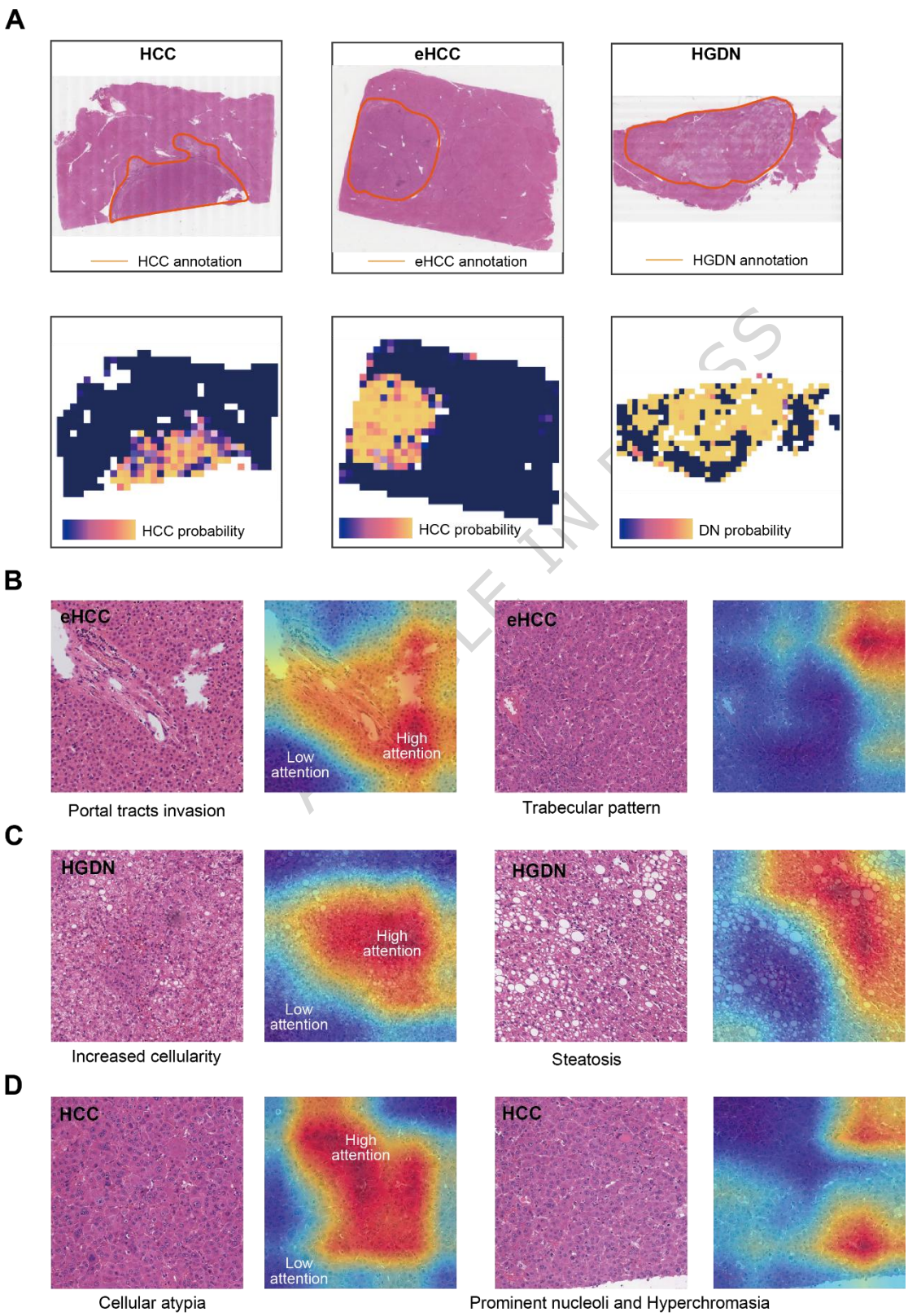

A
HCC
eHCC
HGDN
HCC annotation
eHCC annotation
HGDN annotation
HCC probability
HCC probability
DN probability
B
eHCC
Low attention
High attention
eHCC
Portal tracts invasion
Trabecular pattern
C
HGDN
High attention
Low attention
HGDN
Increased cellularity
Steatosis
D
HCC
High attention
Low attention
HCC
Cellular atypia
Prominent nucleoli and Hyperchromasia

Figure 5

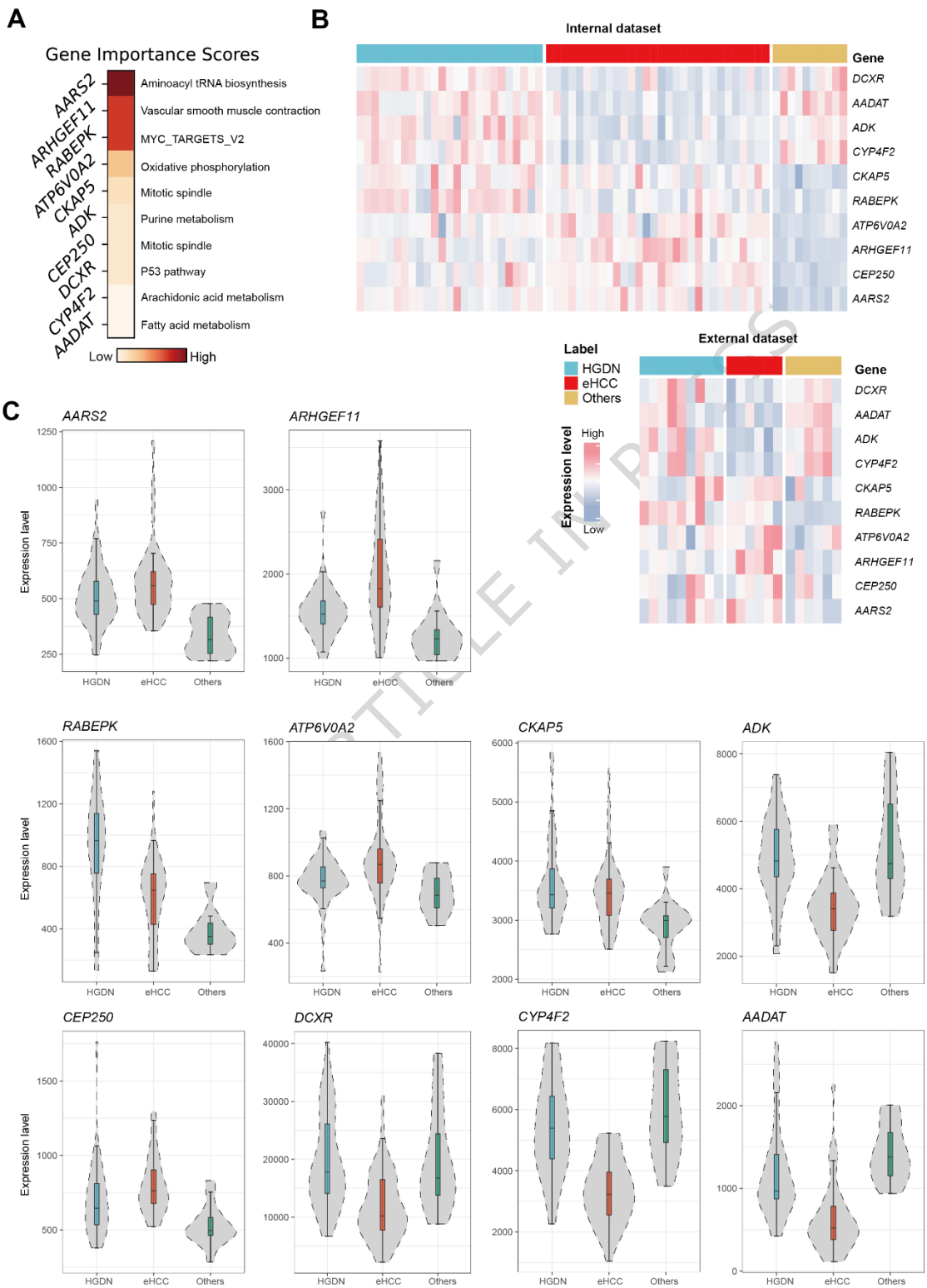
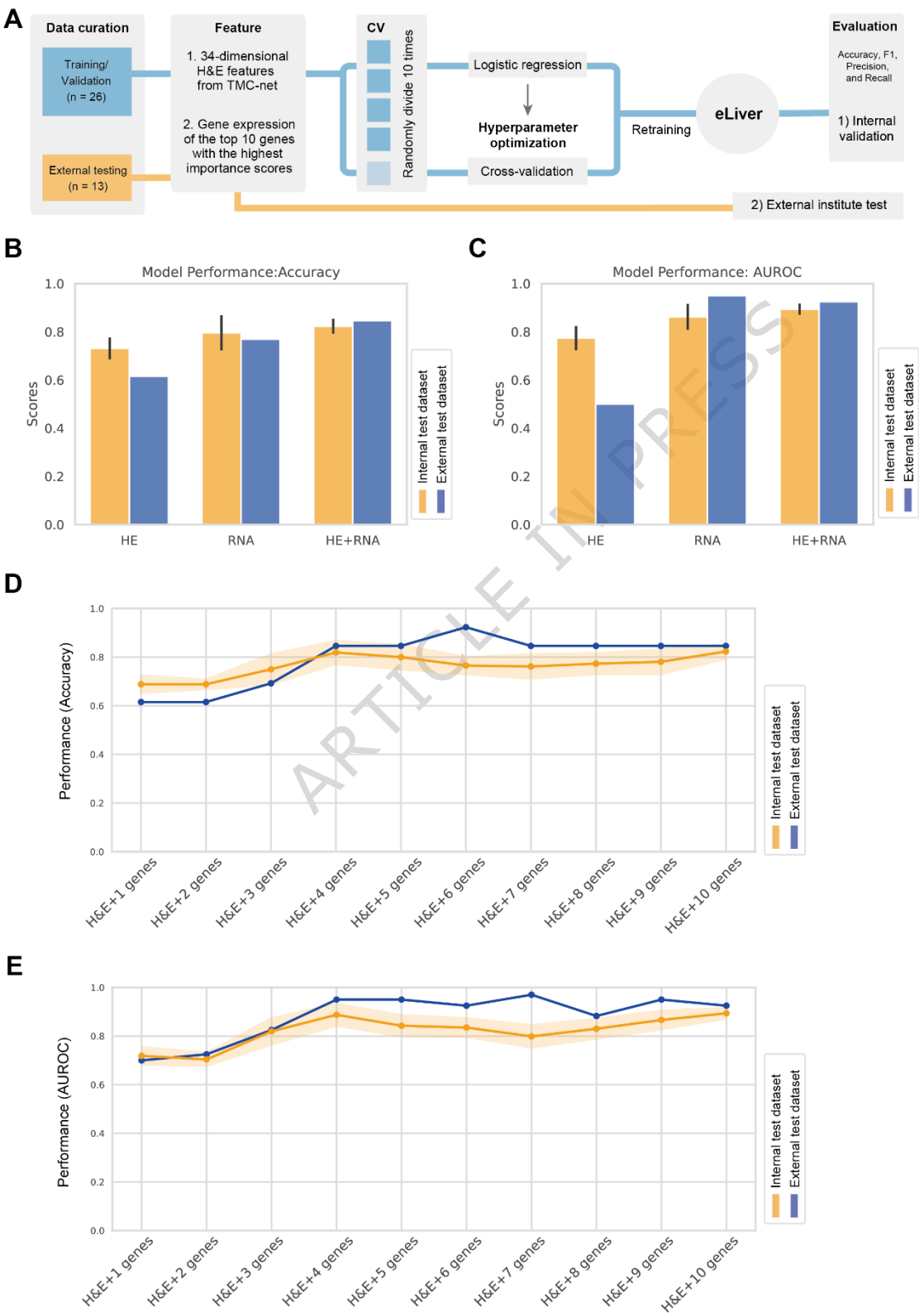

A
Gene Importance Scores
AARS2
ARHGEF11
RABEPK
ATP6V0A2
CKAP5
ADK
CEP250
DCXR
CYP4F2
AADAT
Aminoacyl tRNA biosynthesis
Vascular smooth muscle contraction
MYC_TARGETS_V2
Oxidative phosphorylation
Mitotic spindle
Purine metabolism
Mitotic spindle
P53 pathway
Arachidonic acid metabolism
Fatty acid metabolism
Low
High
B
Internal dataset
Gene
DCXR
AADAT
ADK
CYP4F2
CKAP5
RABEPK
ATP6V0A2
ARHGEF11
CEP250
AARS2
External dataset
Label
HGDN
eHCC
Others
Expression level
High
Low
C
AARS2
ARHGEF11
RABEPK
ATP6V0A2
CKAP5
ADK
CEP250
DCXR
CYP4F2
AADAT
Expression level
HGDN
eHCC
Others

Figure 6


A
Data curation
Training/ Validation (n = 26)
External testing (n = 13)
Feature
1. 34-dimensional H&E features from TMC-net
2. Gene expression of the top 10 genes with the highest importance scores
CV
Randomly divide 10 times
Logistic regression
Hyperparameter optimization
Cross-validation
Retraining
eLiver
Evaluation
Accuracy, F1, Precision, and Recall
1) Internal validation
2) External institute test
B
Model Performance:Accuracy
Scores
HE
RNA
HE+RNA
Internal test dataset
External test dataset
C
Model Performance: AUROC
D
Performance (Accuracy)
H&E+1 genes
H&E+2 genes
H&E+3 genes
H&E+4 genes
H&E+5 genes
H&E+6 genes
H&E+7 genes
H&E+8 genes
H&E+9 genes
H&E+10 genes
E
Performance (AUROC)