



## Primary liver cancer classification from routine tumour biopsy using weakly supervised deep learning

Aurélie Beaufrère, Nora Ouzir, Paul Emile Zafar, Astrid Laurent-Bellue, Miguel Albuquerque, Gwladys Lubuela, Jules Grégory, Catherine Guettier, Kévin Mondet, Jean-Christophe Pesquet, et al.

### ► To cite this version:

Aurélie Beaufrère, Nora Ouzir, Paul Emile Zafar, Astrid Laurent-Bellue, Miguel Albuquerque, et al.. Primary liver cancer classification from routine tumour biopsy using weakly supervised deep learning. JHEP Reports Innovation in Hepatology, 2024, 6 (3), pp.101008. 10.1016/j.jhepr.2024.101008 . hal-04534805

**HAL Id: hal-04534805**

**<https://hal.science/hal-04534805>**

Submitted on 5 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Primary liver cancer classification from routine tumour biopsy using weakly supervised deep learning

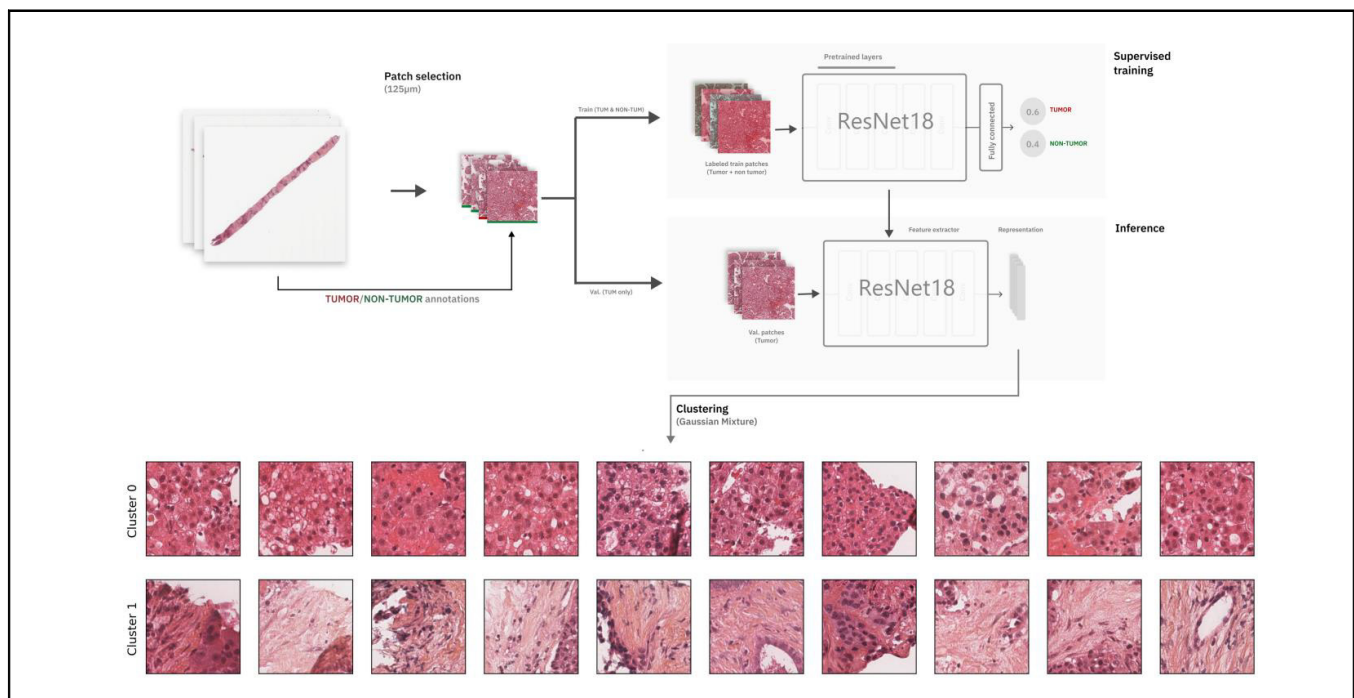
## Authors

**Aurélie Beaufrère, Nora Ouzir**, Paul Emile Zafar, Astrid Laurent-Bellue, Miguel Albuquerque, Gwladys Lubuela, Jules Grégory, Catherine Guettier, Kévin Mondet, Jean-Christophe Pesquet, Valérie Paradis

## Correspondence

[aurelie.beaufriere@aphp.fr](mailto:aurelie.beaufriere@aphp.fr) (A. Beaufrère).

## Graphical abstract



## Highlights

- Weakly supervised learning method identified specific features of HCC and iCCA.
- No specific features of cHCC-CCA were recognized.
- Our model allowed for assessment of the proportion of HCC and iCCA tiles within a cHCC-CCA biopsy.
- Our model correctly predicted the diagnosis in 97% of HCC cases and 83% of iCCA cases.
- For cHCC-CCA, we observed a highly variable proportion of HCC and iCCA tiles.

## Impact and implications

The diagnosis of primary liver cancers can be challenging, especially on biopsies and for combined hepatocellular-cholangiocarcinoma (cHCC-CCA). We automatically classified primary liver cancers on routine-stained biopsies using a weakly supervised learning method. Our model identified specific features of hepatocellular carcinoma and intrahepatic cholangiocarcinoma. Despite no specific features of cHCC-CCA being recognized, the identification of hepatocellular carcinoma and intrahepatic cholangiocarcinoma tiles within a slide could facilitate the diagnosis of primary liver cancers, and particularly cHCC-CCA.



# Primary liver cancer classification from routine tumour biopsy using weakly supervised deep learning



Aurélien Beaufrère,<sup>1,2,3,\*†</sup> Nora Ouzir,<sup>4,†</sup> Paul Emile Zafar,<sup>1,4</sup> Astrid Laurent-Bellue,<sup>5</sup> Miguel Albuquerque,<sup>1</sup> Gwladys Lubuela,<sup>2</sup> Jules Grégory,<sup>1,2,6</sup> Catherine Guettier,<sup>5</sup> Kevin Mondet,<sup>1</sup> Jean-Christophe Pesquet,<sup>4</sup> Valérie Paradis<sup>1,2,3</sup>

<sup>1</sup>AP-HP, Nord, Department of Pathology, FHU MOSAIC, Beaujon Hospital, Clichy, France; <sup>2</sup>Université Paris Cité, Paris, France; <sup>3</sup>Centre de Recherche sur l'Inflammation, INSERM UMR 1149, Paris, France; <sup>4</sup>University of Paris-Saclay, CentraleSupélec, CVN, OPIS Inria, Gif-sur-Yvette 91190, France; <sup>5</sup>AP-HP, Department of Pathology, Hôpital Bicêtre, Le Kremlin-Bicêtre, France; UMR-S 1193, Université Paris-Saclay, Kremlin-Bicêtre, France; <sup>6</sup>AP-HP-Nord, Department of Imaging, Beaujon Hospital, Clichy, France

JHEP Reports 2024. <https://doi.org/10.1016/j.jhepr.2024.101008>

**Background & Aims:** The diagnosis of primary liver cancers (PLCs) can be challenging, especially on biopsies and for combined hepatocellular-cholangiocarcinoma (cHCC-CCA). We automatically classified PLCs on routine-stained biopsies using a weakly supervised learning method.

**Method:** We selected 166 PLC biopsies divided into training, internal and external validation sets: 90, 29 and 47 samples, respectively. Two liver pathologists reviewed each whole-slide hematein eosin saffron (HES)-stained image (WSI). After annotating the tumour/non-tumour areas, tiles of 256x256 pixels were extracted from the WSIs and used to train a ResNet18 neural network. The tumour/non-tumour annotations served as labels during training, and the network's last convolutional layer was used to extract new tumour tile features. Without knowledge of the precise labels of the malignancies, we then applied an unsupervised clustering algorithm.

**Results:** Pathological review classified the training and validation sets into hepatocellular carcinoma (HCC, 33/90, 11/29 and 26/47), intrahepatic cholangiocarcinoma (iCCA, 28/90, 9/29 and 15/47), and cHCC-CCA (29/90, 9/29 and 6/47). In the two-cluster model, Clusters 0 and 1 contained mainly HCC and iCCA histological features. The diagnostic agreement between the pathological diagnosis and the two-cluster model predictions (major contingent) in the internal and external validation sets was 100% (11/11) and 96% (25/26) for HCC and 78% (7/9) and 87% (13/15) for iCCA, respectively. For cHCC-CCA, we observed a highly variable proportion of tiles from each cluster (cluster 0: 5-97%; cluster 1: 2-94%).

**Conclusion:** Our method applied to PLC HES biopsy could identify specific morphological features of HCC and iCCA. Although no specific features of cHCC-CCA were recognized, assessing the proportion of HCC and iCCA tiles within a slide could facilitate the identification of cHCC-CCA.

**Impact and implications:** The diagnosis of primary liver cancers can be challenging, especially on biopsies and for combined hepatocellular-cholangiocarcinoma (cHCC-CCA). We automatically classified primary liver cancers on routine-stained biopsies using a weakly supervised learning method. Our model identified specific features of hepatocellular carcinoma and intrahepatic cholangiocarcinoma. Despite no specific features of cHCC-CCA being recognized, the identification of hepatocellular carcinoma and intrahepatic cholangiocarcinoma tiles within a slide could facilitate the diagnosis of primary liver cancers, and particularly cHCC-CCA.

© 2024 The Author(s). Published by Elsevier B.V. on behalf of European Association for the Study of the Liver (EASL). This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

Primary liver cancers (PLCs) are the third leading cause of cancer-related death worldwide, with an increasing incidence in Western countries.<sup>1,2</sup> PLCs define a heterogeneous group of tumours associated with distinct risk factors, clinical findings,

imaging, and histologic and molecular characteristics. Among these tumours, hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (iCCA) are the most common and represent the two ends of the PLC spectrum. Halfway between HCC and iCCA, combined hepatocellular-cholangiocarcinoma (cHCC-CCA), a much rarer tumour, shares features with both types.<sup>3-5</sup>

The diagnosis of HCC is mainly based on imaging in the context of cirrhosis, whereas the gold standard for diagnosing iCCA and cHCC-CCA is histological analysis.<sup>6-8</sup> The pathological definition of cHCC-CCA has significantly evolved. The last 2019 WHO classification endorsed a cHCC-CCA diagnosis definition based on the unequivocal presence of both hepatocytic and cholangiocytic differentiation within the same tumour on

**Keywords:** Primary liver cancer; biopsy; histological slides; artificial intelligence; weakly supervised learning.

Received 31 August 2023; received in revised form 9 December 2023; accepted 17 December 2023; available online 13 January 2024

† A.B. and N.O. contributed equally to this work.

\* Corresponding author. Address: Department of Pathology, Beaujon Hospital 100 boulevard du Général Leclerc, Clichy, 92110, France; Tel.: +33 1 40 87 54 60, fax: +33 1 40 87 00 77.

E-mail address: [aurelie.beaufriere@aphp.fr](mailto:aurelie.beaufriere@aphp.fr) (A. Beaufrère).





routine histopathology (hematein eosin or hematein eosin saffron [HES] staining). Additional immunostainings can help confirm the diagnosis.<sup>9</sup> Nonetheless, diagnosing PLCs, particularly cHCC-CCA, is challenging on biopsy samples because histological analysis may miss one tiny tumour component. Our previous study concluded that although diagnostic performance remains relatively low for cHCC-CCA on routine biopsy staining, additional immunohistochemistry (IHC) and, mainly, a two-step strategy combining imaging and histology, significantly improved the diagnostic performance of biopsy for cHCC-CCA.<sup>10</sup>

Histology slides are readily available in pathology departments, and the recent advance of high-throughput scanning devices has enabled digital captures of an entire conventional glass slide, called a whole-slide image (WSI).<sup>11,12</sup> WSIs contain valuable information that can be invisible to the human eye. Artificial intelligence (AI), particularly deep-learning methods, can exploit this information at different levels by extracting deep morphological features from WSIs after a tiling step.<sup>13,14</sup> The literature on AI applications has flourished in recent years. Most studies in computational pathology rely on supervised methods, where ground truth is known only for the slide (globally) and hardly ever for every single tile.<sup>15</sup> The effort of an exhaustive annotation has been a significant obstacle to many supervised AI methods in medical applications. In histology, in which the image size is huge, no reasonable human effort will likely succeed in labelling a massive dataset (at the tile scale) for fully supervised AI methodologies.<sup>16</sup>

Weakly supervised methods circumvent this difficulty by using a feasible amount of annotations, and several studies have already shown the effectiveness of these methods in extracting meaningful features for medical applications, including histology.<sup>16,17</sup> The essence of the solved task remains unsupervised, although not entirely blind to the application specificities anymore. In imaging applications, this method could help provide a few categorical labels for WSI segmentation<sup>16</sup> or a limited number of classes to solve a more complex and broad classification problem.<sup>18</sup> In liver pathology, weak annotations of HCC histological slides have been used to extract features for prognosis prediction.<sup>19–21</sup> Different AI models trained on histology slides of HCC treated by surgical resection showed that weakly supervised models outperformed others in predicting the activation of immune and inflammatory gene signatures.<sup>14</sup> Although some of these studies have investigated weakly supervised methods on WSIs from surgical specimens, biopsy studies remain scarce.

This proof-of-concept study aimed to automatically classify PLCs from biopsy samples using a weakly supervised deep-learning approach.

## Materials and methods

### Pathological analysis

#### Sample selection

We selected 119 formalin-fixed paraffin-embedded biopsies of PLC archived between 2012 and 2021 in the Pathology Department of Beaujon Hospital (Clichy, France), divided into 90 training and 29 internal validation samples. Both sets comprised a balanced proportion of HCC, iCCA and cHCC-CCA to train and validate the model correctly. We constructed an external validation set of 47 biopsies of PLC archived in the Pathology Department of Bicêtre Hospital between 2012 and 2023 (Le Kremlin-Bicêtre, France) (Fig. 1). All patients gave written

consent for using the biopsies as required by French legislation. This study was registered at the Commission Nationale de l'Informatique et des Libertés and was approved by the ethics committee (EDS APHP no. CSE-20-85 MOSAIC-EDS). The recorded data included age, sex, presence of surgery, main risk factors for chronic liver diseases (e.g., viral hepatitis, chronic alcohol consumption, metabolic syndrome, hemochromatosis), Child-Pugh score, METAVIR fibrosis score in the non-tumour liver, and the number and size of tumours at diagnosis.

#### HES and IHC

For each case, an HES staining was selected for digitization. An additional IHC analysis was performed for cases included in the internal validation set when sufficient material was available (20/29 cases). Whole-biopsy sections were cut on charged slides (Superfrost slides; Thermo Scientific, Waltham, MA, USA) and were immunostained for routine markers for HCC and iCCA (glypican 3 [1G12, 1/100, Zytomed, Germany], anti-hepatocyte antibody [OCH1E5, 1/250, Agilent, USA], and CK7 [50V-TL 12/30, 1/500, Agilent, USA]). HES and IHC slides were scanned at x20 magnification (Aperio Slide Scanner).

#### Pathology review

Two liver pathology experts (AB and VP) examined all HES WSIs, classifying each case as HCC, iCCA, or cHCC-CCA (Fig. 2). cHCC-CCA diagnosis was retained when biphenotypic differentiation (hepatocytic and cholangiocytic) was evident on HES WSIs according to the definition of the last WHO classification.<sup>9</sup>

#### Quantitative IHC analysis

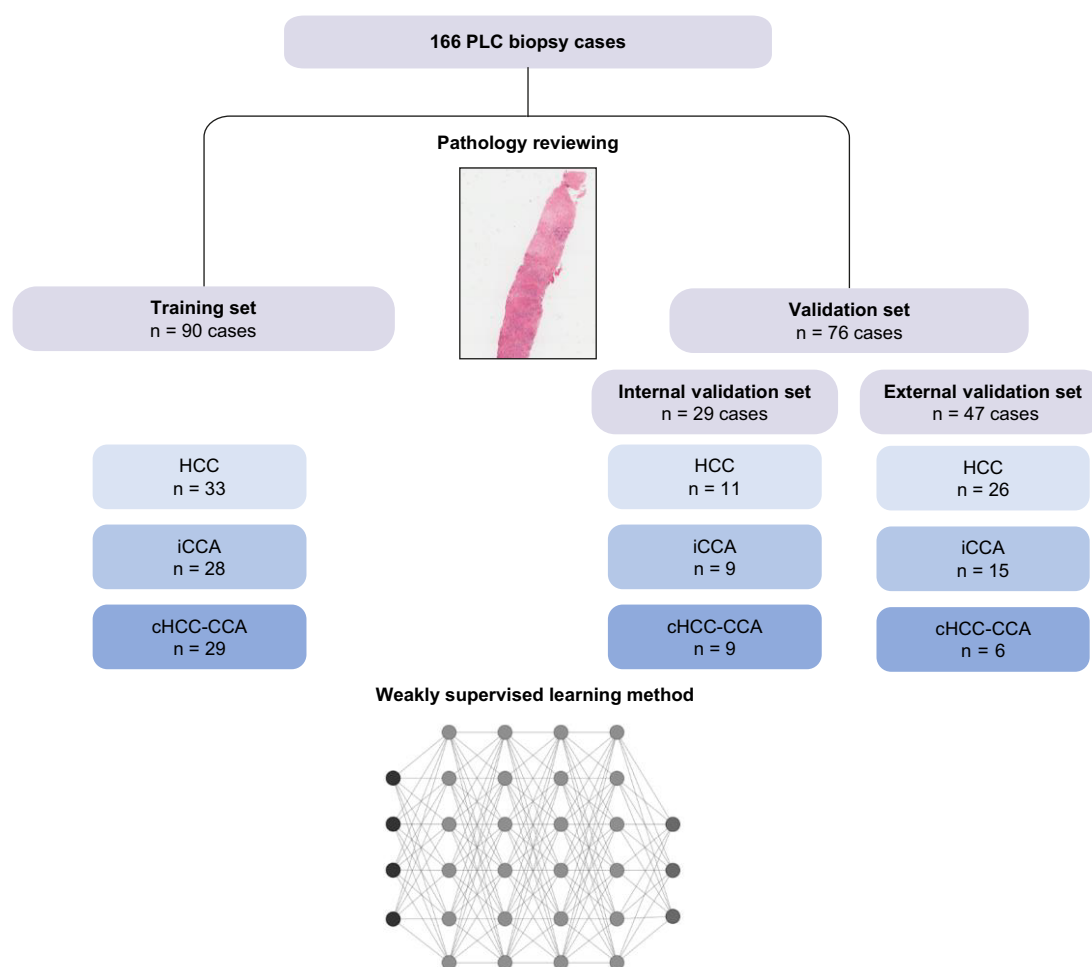
Quantitative IHC analysis was performed in the internal validation set to improve the assessment of the proportion of iCCA and HCC contingents, particularly for cHCC-CCA. Single-cell-based analyses were carried out on validation WSIs using the haematoxylin channel to segment cell nuclei in the open-source WSI analysis software QuPath (v0.2.3).<sup>22</sup> A stepwise procedure followed segmentation for further cell subclassification. First, cells were classified as tumorous or stromal cells using QuPath's built-in machine-learning features as described.<sup>23</sup> Tumour and stromal cells were further subclassified using intensity thresholds in the relevant chromogen channels (DAB). This subclassification comprised the identification of glypican 3-, anti-hepatocyte antibody, and CK7-positive cells within tumour areas (Fig. 3). A ratio of positive tumour cells to tumour cells was calculated for each immunostaining. The percentage of HCC IHC contingent was determined with the glypican 3 or anti-hepatocyte antibody ratio (the highest positive ratio between the two antibodies was retained). The percentage of iCCA IHC contingent was determined with the CK7 ratio. We used the Pearson correlation coefficient to assess the diagnostic agreement between the AI model and pathology.

### Weakly supervised tumour classification

#### Patch processing

Processing WSIs is computationally challenging because of their high dimension ( $\sim 10^4$  pixels wide, which results in images of 100 to 1,000 megapixels). Because a straightforward resizing of the WSIs can lead to loss of crucial information at the microscopic scale, we adopted a patch-wise strategy, with patches (or tiles) of a fixed, much smaller size extracted from the images and processed individually. After testing different scales (i.e., patch sizes), we chose a scale of 125  $\mu\text{m}$ , corresponding to a





**Fig. 1. Flowchart of the study.** We selected 166 formalin-fixed paraffin-embedded biopsies of PLC, divided into 90 training, 29 internal validation and 47 external validation samples. cHCC-CCA, combined hepatocellular-cholangiocarcinoma; HCC, hepatocellular carcinoma; iCCA, intrahepatic cholangiocarcinoma; PLC, primary liver cancer.

patch size of 256x256. This choice allowed us to study the samples at an adequate level of tissue detail while guaranteeing a reasonable computational time. Because tissue areas are scarce within biopsies, this procedure also maximized the number of extracted patches compared to choosing a larger patch size. Large blank background areas cover the rest of the images; they were processed using a masking approach, ignoring the background and extracting patches lying on the tissue areas only (along the shape of the biopsy). We further processed border patches (*i.e.*, patches containing background and tissue) by selecting the ones containing mainly tissue pixels.

#### Data augmentation

After patch extraction, we further augmented the data to bring more variability to the dataset. The data augmentation was restricted to plausible transformations to preserve meaningful patches from the pathological point of view: colour variations ( $\pm 2\%$  hue,  $[-80\%$  to  $+60\%]$  saturation, and  $\pm 20\%$  brightness), rigid transformations such as horizontal and vertical flips, random rotations of  $(-90^\circ$  to  $+90^\circ)$ , and Gaussian blurring with a kernel of size 3x3. We varied the 125  $\mu\text{m}$  scaling by a small factor between 1 and 1.2 to add variability while preserving similar structures in

the patches. The final number of patches was 86,936, including 63,132 ( $\sim 70\%$ ) training patches and 23,804 ( $\sim 30\%$ ) validation patches. To avoid over-fitting, training and validation patches were extracted from biopsies of distinct patients. Both sets contained a balanced proportion ( $\sim 1/3$ ) of patients with each type of cancer (Fig. 1).

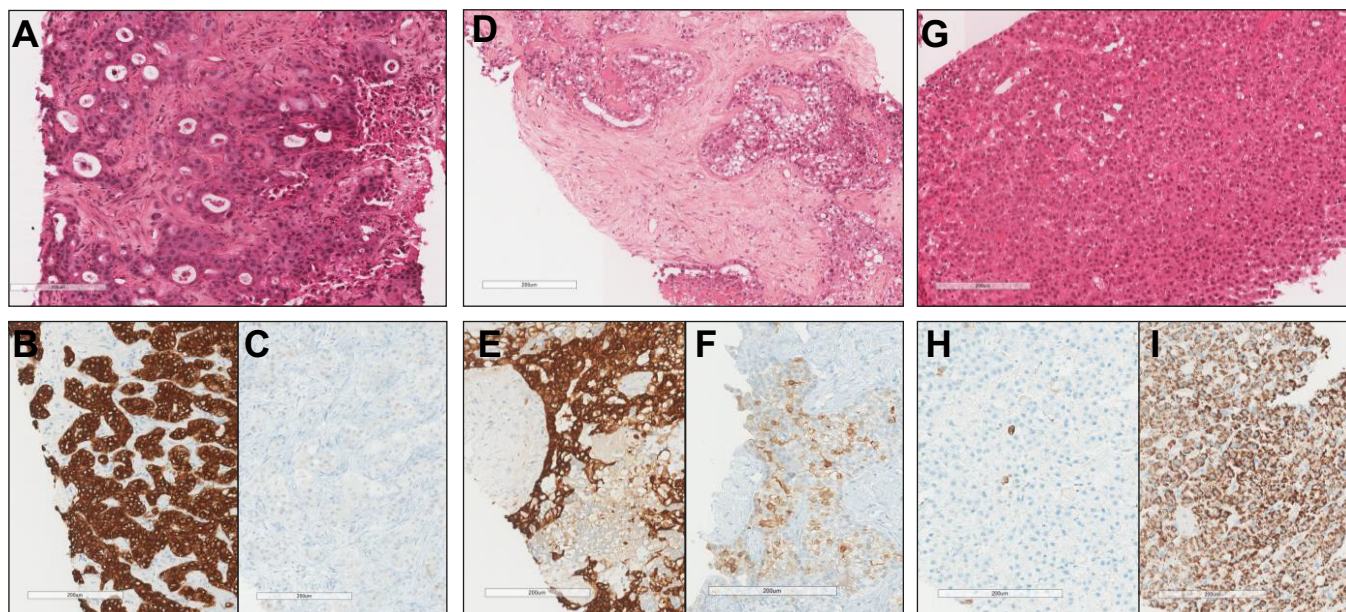
#### Tumour/non-tumour annotations

The weakly supervised approach involved using tumour/non-tumour annotations, which are more easily accessible than detailed tumour-type annotations. The experts manually outlined the tumoural regions, and their annotations were transformed into a binary mask of tumour/non-tumour regions inside/outside the outlined area. Border patches were considered tumoural if most pixels were labelled as such. All biopsies of training and validation cohorts contained tumoural and non-tumoural tiles. We used both tumoural and non-tumoural tiles for training but only tumoural tiles for validation.

#### Feature extraction and clustering

The proposed deep learning-based approach involved a three-step process (Fig. 4). First, supervised learning trained a feature extractor guided by the tumour/non-tumour annotations. The

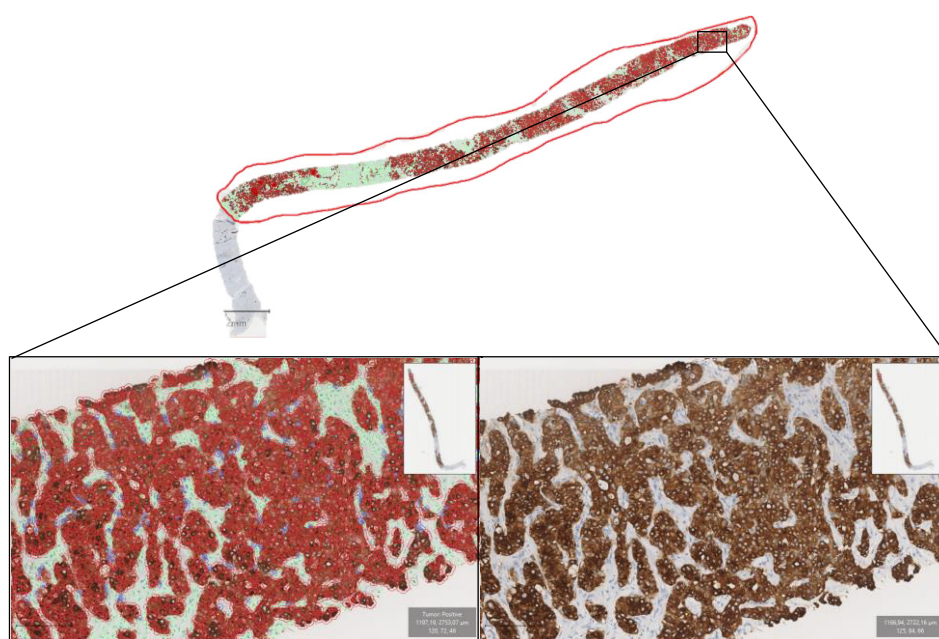




**Fig. 2. Examples of morphological and immunohistochemistry features of iCCA, cHCC-CCA and HCC.** (A) iCCA, HES staining: glandular architecture with fibrous stroma. (B) iCCA, CK7 staining: intense labelling of tumour cells. (C) iCCA, glypican 3 staining: no labelling of tumour cells. (D) cHCC-CCA, HES staining: glands surrounded by cuboidal tumour cells mixed with trabeculae and nests of clear hepatoid tumour cells. (E) cHCC-CCA, CK7 staining: expression of the majority of tumour cells. (F) cHCC-CCA, glypican 3 staining: expression of tumour cells located in the trabeculae and the nests. (G) HCC, HES staining: hepatoid tumour cells organized in trabeculae and pseudoglands without fibrous stroma. (H) HCC, CK7 staining: rare tumour cells marked. (I) HCC, anti-hepatocyte antibody staining: intense expression of all of the tumour cells. cHCC-CCA, combined hepatocellular-cholangiocarcinoma; HCC, hepatocellular carcinoma; HES, hematein eosin saffron; iCCA, intrahepatic cholangiocarcinoma.

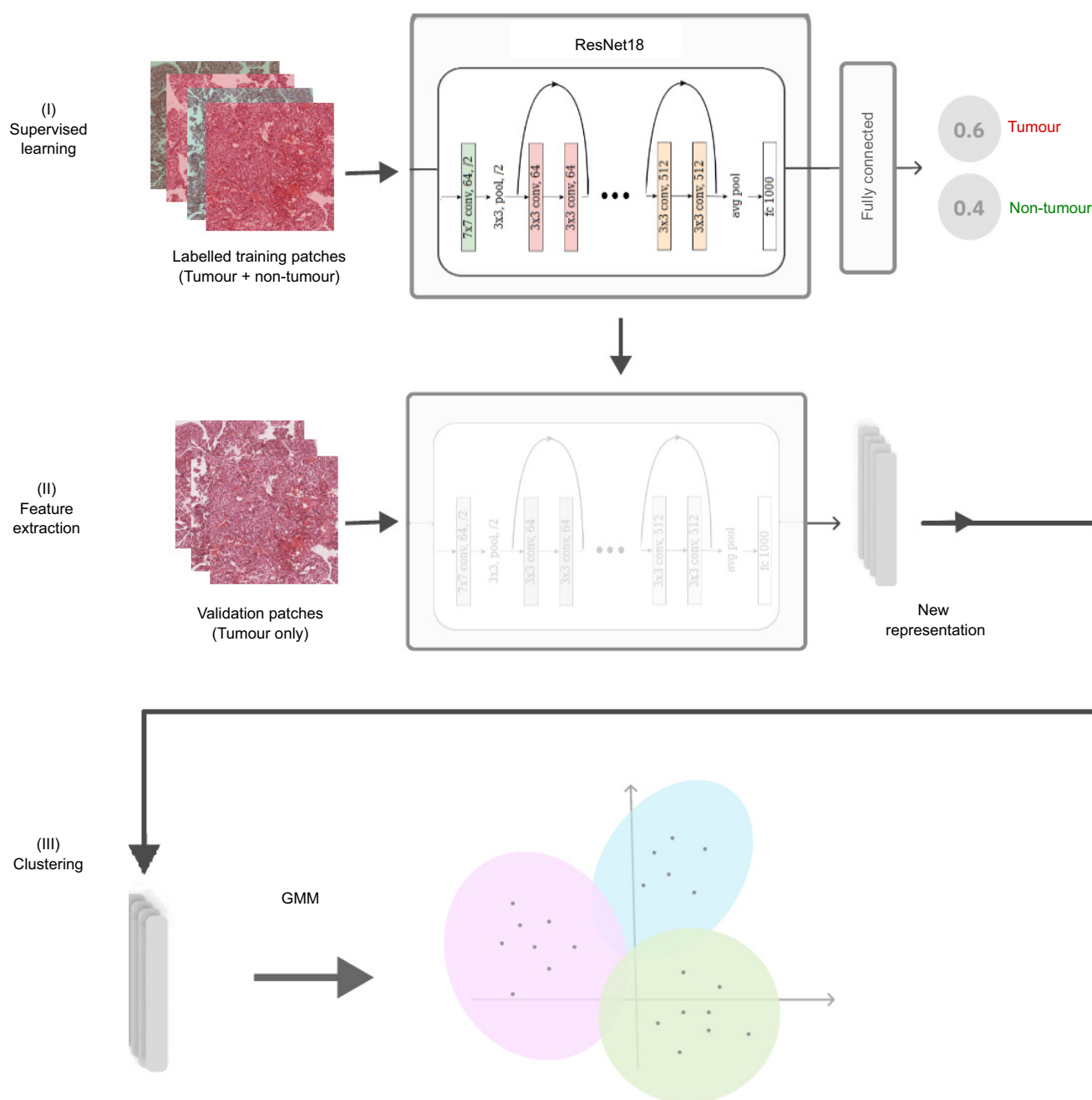
feature extractor was applied to tumoural validation patches in the second phase, transforming them according to the learned representation. Finally, an unsupervised learning method

clustered the transformed validation patches into  $K = 2$  or  $K = 3$  groups. We used principal component analysis to reduce the number of features by half before clustering.



**Fig. 3. Quantitative immunohistochemistry analysis with QuPath software.** Example analysis of CK7 staining in an iCCA case. The tumour area was manually annotated. Cells were classified as tumour cells (red or blue) or stromal cells (green) by using the QuPath machine-learning features. Then CK7-positive tumour cells (red) were further subclassified by using intensity thresholds in the tumour area relevant chromogen channels (DAB). iCCA, intrahepatic cholangiocarcinoma.





**Fig. 4. Main steps of the proposed weakly supervised method.** After annotating the tumour/non-tumour areas, tiles of 256x256 pixels were extracted from the whole-slide images and used to train a ResNet18 neural network (step I). The tumour/non-tumour annotations served as labels during training, and the network's last convolutional layer was used to extract new tumour tile features (step II). Without knowledge of the precise labels of the malignancies, we then applied an unsupervised clustering algorithm (GMM) (step III). GMM, Gaussian mixture model.

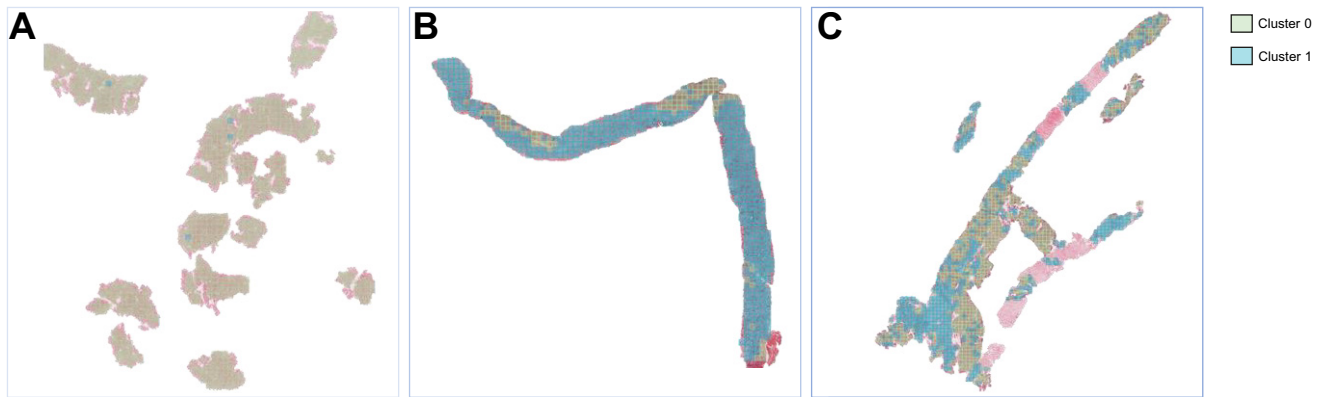
#### Implementation details

The first stage of the method used a pre-trained ResNet18 architecture that was fine-tuned on its last eight convolutional layers with the annotated 63,132 training patches. The weights of the first nine layers pre-trained on ImageNet<sup>24</sup> were unchanged. The ResNet18 architecture, although less complex than its deeper 50- and 150-layer versions,<sup>25</sup> provided a solid baseline for assessing the potential of transfer learning for our task. Recent works using models pre-trained on ImageNet have demonstrated that more complex architectures did not necessarily lead

to clear improvements and significantly increased the computational load.<sup>26</sup> Hence, ResNet18, with fewer parameters, is less prone to over-fitting and over-specialization.

We used an Adam optimizer with an initial learning rate of 0.03 and an exponential decay scheduler with a multiplicative factor of 0.5. The obtained training and validation accuracies were 0.75 and 0.71, respectively. After training, the final dense level of the network was discarded. The last convolutional layer was used to extract the features from the 23,804 validation patches, which led to a new patch representation of size





**Fig. 5. Spatial distribution of the tumour labels for the two-cluster model in examples of PLCs.** Examples of HCC (A), iCCA (B), and cHCC-CCA patients (C). cHCC-CCA, combined hepatocellular-cholangiocarcinoma; HCC, hepatocellular carcinoma; iCCA, intrahepatic cholangiocarcinoma.

256x256x512 (*i.e.*, 512 features representing each pixel). After principal component analysis, the patch size was reduced to 256x256x256. The unsupervised clustering algorithm was the Gaussian mixture model, which was solved with the expectation-maximization algorithm.<sup>27</sup> After clustering, the WSI can be reconstructed to visualize the spatial distribution of the tumour labels (Fig. 5).

## Results

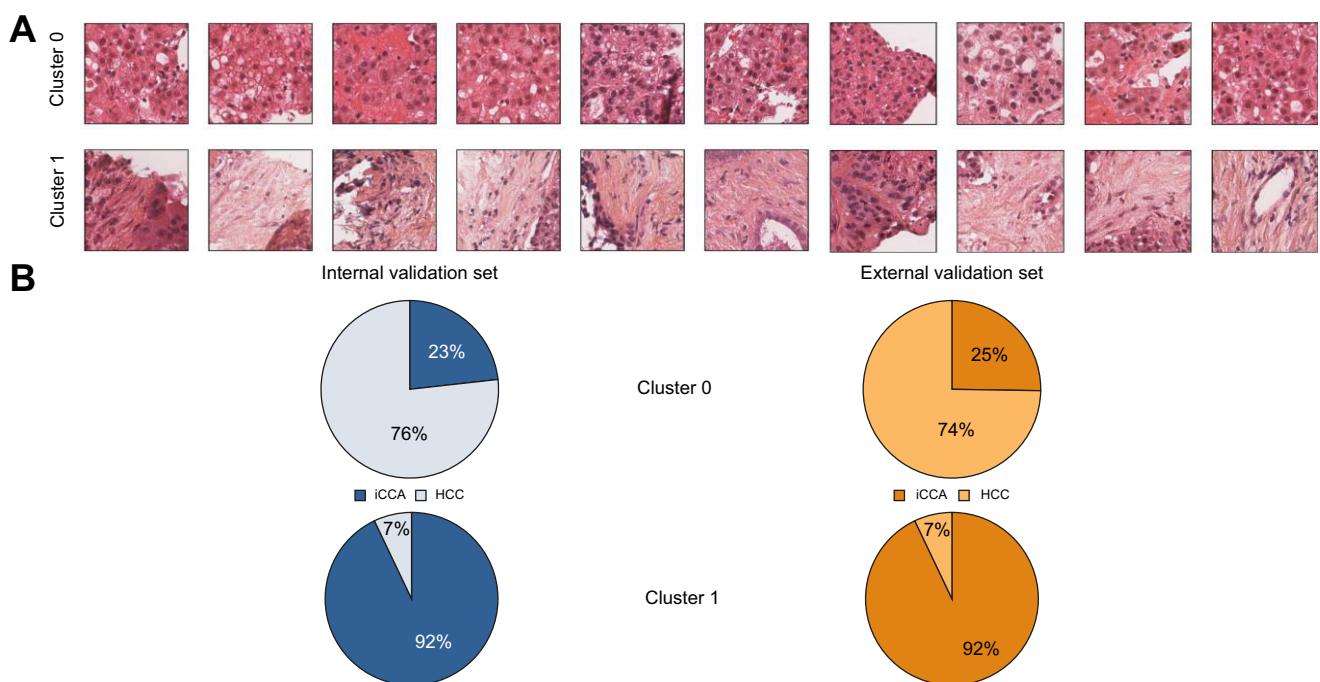
### Clinical and pathological features of cases

Clinical features of cases included in the two datasets are summarized in Table S1. Pathological review classified the training, internal and external validation sets into HCC (33/90, 11/29 and 26/47, respectively), iCCA (28/90, 9/29 and 15/47, respectively),

and cHCC-CCA (29/90, 9/29 and 6/47, respectively) (Fig. 1). In the internal validation set, all cHCC-CCA ( $n = 9$ ) cases presented a majority contingent of iCCA. The mean percentage of iCCA IHC contingent was 90% (range 79–99) in iCCA cases ( $n = 7$ ), 2% (0–3) in HCC cases ( $n = 5$ ), and 64% (1–100) in cHCC-CCA cases ( $n = 8$ ). The mean percentage of HCC IHC contingent was 4% (0–13) in iCCA cases, 76% (52–93) in HCC cases, and 17% (1–41) in cHCC-CCA cases. In the external validation set, most cHCC-CCA cases (5/6, 83%) presented a majority contingent of iCCA.

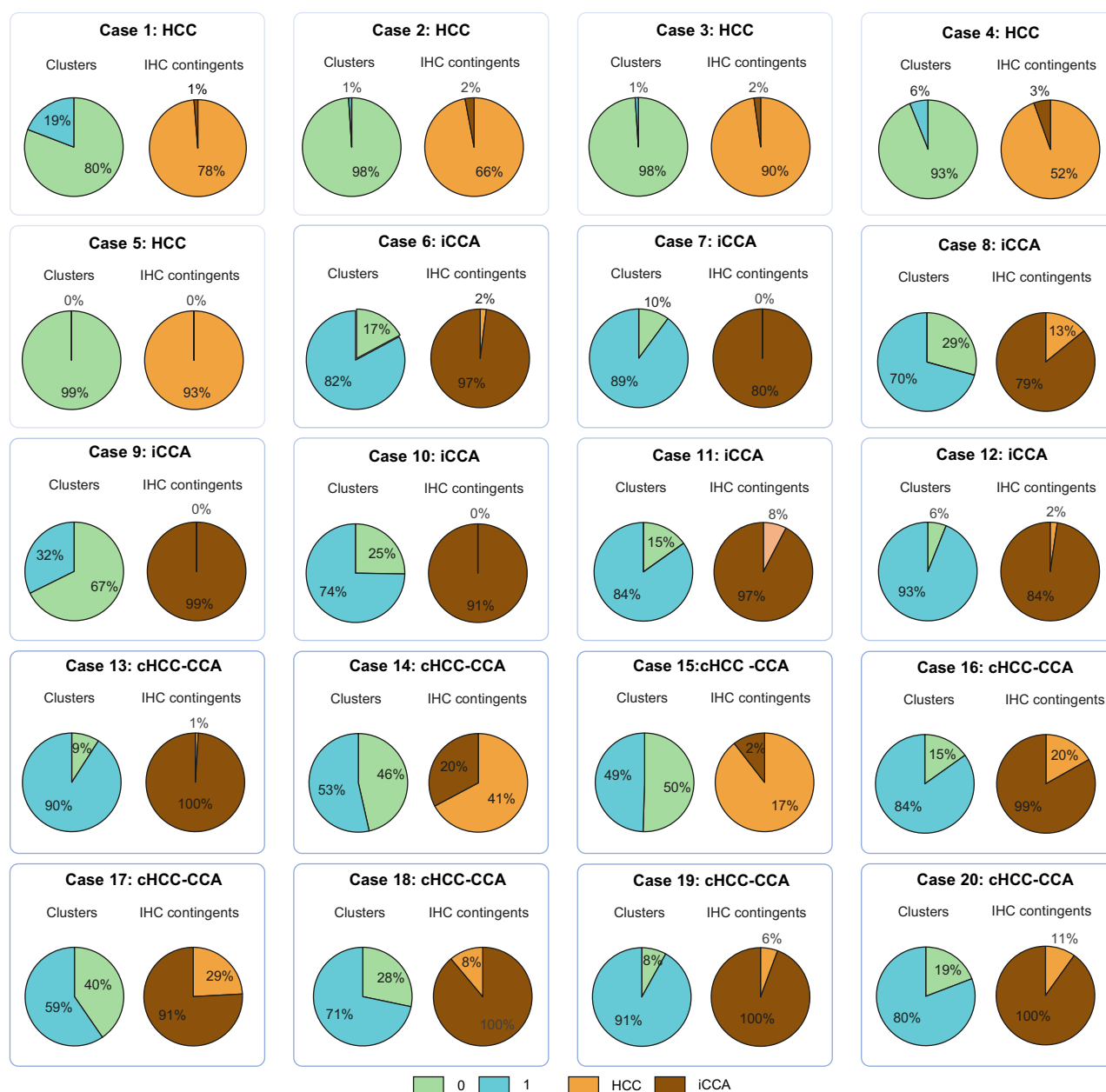
### Clustering results

After feature extraction, the validation set was clustered with a Gaussian mixture model. Different numbers of clusters were tested and analysed. The most coherent results were obtained for  $K = 2$  clusters. In this two-cluster model, Cluster 0 contained tiles



**Fig. 6. Results of the two-cluster model.** Examples of patches in Clusters 0 and 1 (A) and distribution of the HCC and iCCA tumour tiles within each cluster in the internal and external validation sets (B). HCC, hepatocellular carcinoma; iCCA, intrahepatic cholangiocarcinoma.





**Fig. 7. Comparison of the two-cluster predictions and immunohistochemistry contingents in the internal validation cohort (n = 20).** Pie charts showing (i) the proportion of Cluster 0 (green) and 1 (blue) tiles within each slide and (ii) the proportion of HCC (orange) and iCCA IHC contingents (brown) within each slide. The associated pathological diagnosis is displayed above each chart. cHCC-CCA, combined hepatocellular-cholangiocarcinoma; HCC, hepatocellular carcinoma; iCCA, intrahepatic cholangiocarcinoma.

with homogeneous tissue and large cells with abundant eosinophil cytoplasm, features characteristic of HCC (76% of the tiles). The second cluster (Cluster 1) contained tiles with glandular structures within a fibrous stroma (stained orange in HES), usually present in iCCA tumours (92% of the tiles) (Fig. 6). Tiles from cHCC-CCA cases were spread among the two clusters.

We created a three-cluster model ( $K = 3$ ) to test the potential of extracting a separate cluster for the combined cHCC-CCA tumour. The three-cluster model divided the previous Cluster 0 into two clusters (Clusters 2 and 3), with the previous Cluster 1 remaining separate (Cluster 4). The cHCC-CCA patches were spread among the three clusters. Half the cHCC-CCA cases (51%)

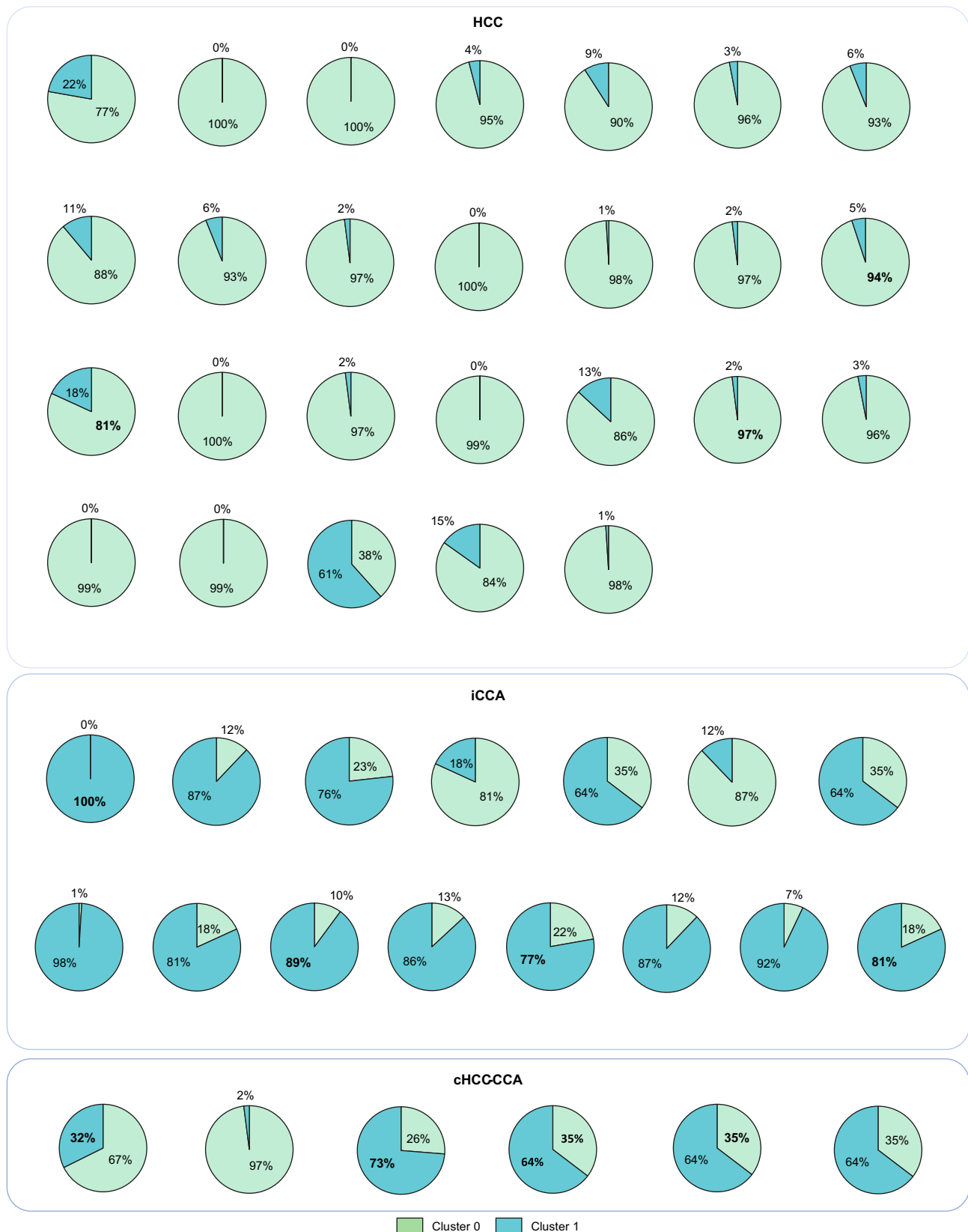
were assigned to the cluster with a majority of iCCA (Cluster 4, corresponding to Cluster 1 in the two-cluster model), whereas the two new HCC clusters, Clusters 2 and 3, contained 26% and 31% of cHCC-CCA cases, respectively (Fig. S1).

### Internal validation of the two-cluster model

#### Quantification of clusters in the validation set

The two-cluster model was assessed with the validation set by quantifying the proportions of Cluster 0 (representative of HCC) and Cluster 1 (representative of iCCA) in each case. Overall, in validation set samples, the mean proportion of Cluster 0 was 51% (range 2-100%) and Cluster 1 was 48% (0-97%).





**Fig. 8. Comparison of the two-cluster predictions and the pathological diagnosis within each slide of the external validation set (n = 47).** Pie charts showing the proportion of Cluster 0 (green) and 1 (blue) tiles within each slide. The associated pathological diagnosis is displayed above each group. cHCC-CCA, combined hepatocellular-cholangiocarcinoma; HCC, hepatocellular carcinoma; iCCA, intrahepatic cholangiocarcinoma.



### Two-cluster model and pathology diagnosis

For HCC and iCCA, the diagnostic agreement between the pathological diagnosis and the model predictions (major contingent) was 100% for HCC (11/11 cases) and 78% for iCCA (7/9 cases). For cHCC-CCA, we observed a highly variable proportion of each cluster type (Cluster 0 [5-50%] and Cluster 1 [9-94%]). The diagnostic agreement between the major contingent in conventional pathology and the one predicted by the model was 89% (8/9 cases) (Fig. S2).

### Two-cluster model and quantitative IHC analyses

The diagnostic agreement between the IHC diagnosis and the two-cluster model predictions (major contingent) was 100% in HCC (5/5 cases), 86% in iCCA (6/7 cases) and 75% in cHCC-CCA (6/8 cases) (Fig. 7). The correlation coefficient between the proportions of Cluster 0 predicted by the two-cluster model and the HCC IHC contingent was 0.87, and between the proportions of Cluster 1 predicted by the two-cluster model and the iCCA IHC contingent was 0.77.

### External validation of the two-cluster model

The results obtained with the two-cluster model for the internal validation set generalized well to the external validation set. Cluster 0 contained a majority of HCC tiles (74% of the tiles), and Cluster 1 contained mainly iCCA tiles (92% of the tiles) (Fig. 6). For HCC and iCCA, the diagnostic agreement between the pathological diagnosis and the model predictions (major contingent) was 96% (25/26) for HCC and 87% (13/15) for iCCA, respectively.

For cHCC-CCA, we observed a highly varying proportion of each cluster type (Cluster 0 [26-97%] and Cluster 1 [2-73%]). The diagnostic agreement between the major contingent in conventional pathology and the one predicted by the model was 83% (5/6 cases) (Fig. 8).

## Discussion

This proof-of-concept study has shown that weakly supervised deep learning can extract discriminative features of different PLCs from routinely stained tumour biopsies. A two-cluster model based on Gaussian mixtures of these features has illustrated the discriminative power of biopsies and their potential for classifying different liver cancers.

At the core of the proposed approach was a combination of supervised transfer learning and unsupervised clustering. For the supervised part, we used a pre-trained convolutional neural network (ResNet18) with weak annotations of tumour/non-tumour areas in WSIs. As explained before, a complete annotation of these images would have required tremendous effort. Hence, a comparison with a fully supervised approach was impossible. Nonetheless, the adopted transfer learning approach successfully retrieved meaningful features with reasonable annotation effort and showed promising results for classifying PLC in biopsies.

Our model has identified two clusters, one specific to HCC and another specific to iCCA, corresponding to the two tumour types located at opposite ends of the malignant liver spectrum. The distinction between the two clusters aligns with the morphological variations between HCC, characterized by notable tumour eosinophilic hepatocytes without fibrous stroma, and iCCA, characterized by smaller cells with abundant fibrous stroma. Among HCC and iCCA cases in the validation set, two cases of iCCA were misclassified as HCC by the two-cluster model. A

pathological review of these two cases revealed one moderately differentiated tumour and one poorly differentiated tumour, both exhibiting extensive architectural areas and minimal stroma. These factors could potentially explain the model's error. In the two-cluster model, cHCC-CCA tiles included varying proportions of both HCC and iCCA clusters. These results are consistent with the current WHO definition of the unequivocal presence of both hepatocytic and cholangiocytic differentiation within the same tumour on routine histopathology.<sup>9</sup>

The proposed two-cluster model provides the accurate percentage of HCC and iCCA components in cHCC-CCA, which could be valuable information for choosing a treatment strategy. Currently, there is no systemic treatment for unresectable cHCC-CCA, and patients with cHCC-CCA receive standard advanced iCCA or HCC treatments<sup>28-31</sup> without any specific recommendation. Hence, determining the main component of cHCC-CCA holds potential for guiding the best treatment strategy.

Our two-cluster model involved only routine HES slides. Although diagnosis is based on routine HES staining, additional IHC is still routinely performed to confirm the diagnosis of cHCC-CCA.<sup>7,9</sup> Comparing each contingent's percentage with the pathological diagnosis (based on both HES and IHC staining) showed that our HES only model achieved equivalent tumour classification. This consistency is remarkable as biopsy samples may be limited in size, preventing the use of complementary IHC. For example, 9 out of 29 cases (31%) in our cohort could not benefit from this additional analysis because of a shortage in tumour tissue. Interestingly, the model has identified HCC tiles within all iCCA cases, confirmed by IHC analysis in half of them, suggesting that cHCC-CCA could be more frequent than reported by the pathologist in the iCCA group. Thus, our model, providing a more accurate tumour tissue characterization than conventional histology, confirms previous findings showing tiny HCC or iCCA areas (not diagnosed by histology) within cHCC-CCA samples using MALDI imaging, an *in situ* proteomic approach.<sup>32</sup>

The potential therapeutic outcomes and the ability to circumvent material difficulties highlight the importance of developing automatic AI methodologies for biopsy specimens. Currently, most AI studies of PLC have focused on surgical samples,<sup>13,14,19,33</sup> but most patients do not have such samples during their entire cancer history, which introduces a selection bias. In contrast, tumour biopsies are increasingly performed in the context of PLC, which could be helpful for AI-based approaches. With this in mind, deep learning can be challenging with biopsies, as their representativeness may be limited for heterogeneous tumours, and their small sample size and intricate shape reduce the number of tiles useable for analysis. Artefacts such as fragmentation, folds and tears (less visible in surgical samples) can also hinder the use of biopsies. Nevertheless, a few studies have laid the groundwork for using biopsies for other cancers, particularly prostate cancer.<sup>34-36</sup> The present study confirms that biopsies can be exploited despite these challenges and indicates that encouraging deep learning-based results can be obtained for liver cancer. Our study promotes the use of biopsy for PLC diagnosis and supports recent results showing that tumour biopsy led to accurate diagnosis in 11% of nodules classified as LIRADS-5 on radiology.<sup>37</sup>

The main limitations of this work are its retrospective design and the relatively low number of studied cases. To the best of our knowledge, no open-access database of PLC biopsies is available. Moreover, HCC routine biopsies are scarce, and cHCC-CCA samples are even more so because of the extreme rarity of this subtype. In



order to avoid unbalanced learning, we have selected a similar proportion of each tumour subtype in the training and internal validation sets, which has impacted the total number of studied cases. Additional prospective validation has yet to confirm the full potential of our model in assisting the pathologist.

## Abbreviations

AI, artificial intelligence; cHCC-CCA, Combined hepatocellular carcinoma; HES, hematein eosin saffron; HCC, hepatocellular carcinoma; IHC, Immunohistochemistry; iCCA, Intrahepatic cholangiocarcinoma; PLC, Primary liver cancer; WSI, Whole-slide image.

## Financial support

The authors received no financial support to produce this manuscript.

## Conflicts of interest

The authors declare no conflicts of interest that pertain to this work.

Please refer to the accompanying ICMJE disclosure forms for further details.

## Authors' contributions

Study concept and design (AB, KM, JG, JCP, VP), acquisition of data (AB, PEZ, NO, MA, GL); analysis and interpretation of data (AB, PEZ, NO, JCP, VP); drafting of the manuscript (AB, NO, KM); critical revision of the manuscript for important intellectual content (JCP, VP); statistical analysis (AB); study supervision (JCP, VP).

## Data availability statement

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## Acknowledgement

We thank the Bernoulli Lab (Inria-APHP) for supporting this study.

## Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhepr.2024.101008>.

## References

*Author names in bold designate shared co-first authorship*

- [1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–249. <https://doi.org/10.3322/caac.21660>.
- [2] GLOBOCAN Cancer Fact Sheets. Liver Cancers n.d. <http://globocan.iarc.fr/old/FactSheets/cancers/liver-new.asp>. [Accessed 27 November 2017].
- [3] Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet Lond Engl* 2018;391:1301–1314. [https://doi.org/10.1016/S0140-6736\(18\)30010-2](https://doi.org/10.1016/S0140-6736(18)30010-2).
- [4] Llovet JM, Montal R, Sia D, Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. *Nat Rev Clin Oncol* 2018;15:599–616. <https://doi.org/10.1038/s41571-018-0073-4>.
- [5] Bridgewater J, Galle PR, Khan SA, et al. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol* 2014;60:1268–1289. <https://doi.org/10.1016/j.jhep.2014.01.021>.
- [6] Aghemo A. Update on HCC management and review of the new EASL guidelines. *Gastroenterol Hepatol* 2018;14:384–386.
- [7] Beaufrère A, Calderaro J, Paradis V. Combined hepatocellular-cholangiocarcinoma: an update. *J Hepatol* 2021;74:1212–1224. <https://doi.org/10.1016/j.jhep.2021.01.035>.
- [8] Holczbauer Á, Factor VM, Andersen JB, et al. Modeling pathogenesis of primary liver cancer in lineage-specific mouse cell types. *Gastroenterology* 2013;145:221–231. <https://doi.org/10.1053/j.gastro.2013.03.013>.
- [9] WHO classification of Tumours Editorial Board. Digestive system tumours. vol.1. 5th ed. Lyon: International Agency for Research on cancer; 2019.
- [10] Gigante E, Ronot M, Bertin C, et al. Combining imaging and tumour biopsy improves the diagnosis of combined hepatocellular-cholangiocarcinoma. *Liver Int Off J Int Assoc Study Liver* 2019;39:2386–2396. <https://doi.org/10.1111/liv.14261>.
- [11] Zarella MD, Rivera Alvarez K. High-throughput whole-slide scanning to enable large-scale data repository building. *J Pathol* 2022;257:383–390. <https://doi.org/10.1002/path.5923>.
- [12] Calderaro J, Kather JN. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. *Gut* 2021;70:1183–1193. <https://doi.org/10.1136/gutjnl-2020-322880>.
- [13] Kather JN, Calderaro J. Development of AI-based pathology biomarkers in gastrointestinal and liver cancer. *Nat Rev Gastroenterol Hepatol* 2020;17:591–592. <https://doi.org/10.1038/s41575-020-0343-3>.
- [14] Zeng Q, Klein C, Caruso S, et al. Artificial intelligence predicts immune and inflammatory gene signatures directly from hepatocellular carcinoma histology. *J Hepatol* 2022;77:116–127. <https://doi.org/10.1016/j.jhep.2022.01.018>.
- [15] Ghaffari Laleh N, Muti HS, Loeffler CML, et al. Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology. *Med Image Anal* 2022;79:102474. <https://doi.org/10.1016/j.media.2022.102474>.
- [16] Lerousseau M, Classe M, Battistella E, et al. Weakly supervised pan-cancer segmentation tool. In: de Bruijne M, Cattin PC, Cotin S, Padoy N, Speidel S, Zheng Y, et al., editors. *Med. Image Comput. Comput. Assist. Interv. – MICCAI 2021*. Cham: Springer International Publishing; 2021. p. 248–256. [https://doi.org/10.1007/978-3-030-87237-3\\_24](https://doi.org/10.1007/978-3-030-87237-3_24).
- [17] Chen RJ, Krishnan RG. Self-supervised vision transformers learn visual concepts in histopathology 2022. <https://doi.org/10.48550/arXiv.2203.00585>.
- [18] Martin S, Boudiaf M, Chouzenoux E, et al. Towards practical few-shot query sets: transductive minimum description length inference. *Adv Neural Inf Process Syst* 2022;35:34677–34688.
- [19] Saillard C, Schmauch B, Laifa O, et al. Predicting survival after hepatocellular carcinoma resection using deep-learning on histological slides. *Hepatol Baltim Md* 2020. <https://doi.org/10.1002/hep.31207>.
- [20] Shi J-Y, Wang X, Ding G-Y, et al. Exploring prognostic indicators in the pathological images of hepatocellular carcinoma based on deep learning. *Gut* 2021;70:951–961. <https://doi.org/10.1136/gutjnl-2020-320930>.
- [21] Yamashita R, Long J, Saleem A, et al. Deep learning predicts postsurgical recurrence of hepatocellular carcinoma from digital histopathologic images. *Sci Rep* 2021;11:2047. <https://doi.org/10.1038/s41598-021-81506-y>.
- [22] Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep* 2017;7:16878. <https://doi.org/10.1038/s41598-017-17204-5>.
- [23] Aulasevich N, Haist M, Förtsch S, et al. The role of the immune phenotype in tumor progression and prognosis of patients with mycosis fungoides: a quantitative immunohistology whole slide approach. *Cells* 2022;11:3570. <https://doi.org/10.3390/cells11223570>.
- [24] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database. 2009. *IEEE Conf. Comput. Vis. Pattern Recognit.*; 2009. p. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>.
- [25] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2015. <https://doi.org/10.48550/arXiv.1512.03385>.
- [26] Ke A, Ellsworth W, Banerjee O, et al. CheXtransfer: performance and parameter efficiency of ImageNet models for chest X-Ray interpretation. *Proc. Conf. Health inference learn*. New York, NY, USA: Association for Computing Machinery; 2021. p. 116–124. <https://doi.org/10.1145/3450439.3451867>.
- [27] Moon TK. The expectation-maximization algorithm. *IEEE Signal Process Mag* 1996;13:47–60. <https://doi.org/10.1109/79.543975>.
- [28] Gigante E, Hobeika C, Le Bail B, et al. Systemic treatments with tyrosine kinase inhibitor and platinum-based chemotherapy in patients with unresectable or metastatic hepatocellular carcinoma. *Liver Cancer* 2022;11:460–473. <https://doi.org/10.1159/000525488>.



- [29] Salimon M, Prieux-Klotz C, Tougeron D, et al. Gemcitabine plus platinum-based chemotherapy for first-line treatment of hepatocholangiocarcinoma: an AGEO French multicentre retrospective study. *Br J Cancer* 2018;118:325–330. <https://doi.org/10.1038/bjc.2017.413>.
- [30] Kobayashi S, Terashima T, Shiba S, et al. Multicenter retrospective analysis of systemic chemotherapy for unresectable combined hepatocellular and cholangiocarcinoma. *Cancer Sci* 2018;109:2549–2557. <https://doi.org/10.1111/cas.13656>.
- [31] Trikalinos NA, Zhou A, Doyle MBM, et al. Systemic therapy for combined hepatocellular-cholangiocarcinoma: a single-institution experience. *J Natl Compr Cancer Netw JNCCN* 2018;16:1193–1199. <https://doi.org/10.6004/jnccn.2018.7053>.
- [32] Gigante E, Cazier H, Albuquerque M, et al. MALDI imaging, a powerful multiplex approach to decipher intratumoral heterogeneity: combined hepato-cholangiocarcinomas as proof of concept. *Cancers* 2023;15:2143. <https://doi.org/10.3390/cancers15072143>.
- [33] Cheng N, Ren Y, Zhou J, et al. Deep learning-based classification of hepatocellular nodular lesions on whole-slide histopathologic images. *Gastroenterology* 2022;162:1948–1961.e7. <https://doi.org/10.1053/j.gastro.2022.02.025>.
- [34] Pantanowitz L, Quiroga-Garza GM, Bien L, et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *Lancet Digit Health* 2020;2:e407–e416. [https://doi.org/10.1016/S2589-7500\(20\)30159-X](https://doi.org/10.1016/S2589-7500(20)30159-X).
- [35] Nagpal K, Foote D, Tan F, et al. Development and validation of a deep learning algorithm for gleason grading of prostate cancer from biopsy specimens. *JAMA Oncol* 2020;6:1372–1380. <https://doi.org/10.1001/jamaoncol.2020.2485>.
- [36] **Xu F, Zhu C**, Tang W, et al. Predicting axillary lymph node metastasis in early breast cancer using deep learning on primary tumor biopsy slides. *Front Oncol* 2021;11:759007. <https://doi.org/10.3389/fonc.2021.759007>.
- [37] Brusset B, Jacquemin M, Teyssier Y, et al. Radiological diagnosis of hepatocellular carcinoma does not preclude biopsy before treatment. *JHEP Rep* 2023;0. <https://doi.org/10.1016/j.jhepr.2023.100957>.