Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mapping microstructural features of pathological tissues by pixel clustering of Mueller matrix images

Hui Ma (mahui@tsinghua.edu.cn)
Tsinghua University

Jiachen Wan
Tsinghua University

Yang Dong
Tsinghua University

Yue Yao
Tsinghua University

Weijin Xiao
Fujian Cancer Hospital

Ruqi Huang
Tsinghua-Berkeley Shenzhen Institute https://orcid.org/0000-0001-5942-3671

Jing-Hao Xue
University College London

Ran Peng
Fujian Medical University

Haojie Pei
Tsinghua University

Xuewu Tian
University of Chinese Academy of Sciences Shenzhen Hospital

Ran Liao
Tsinghua University

Honghui He
Tsinghua University https://orcid.org/0000-0001-7369-7433

Nan Zeng
Tsinghua University

Chao Li
Fujian Cancer Hospital

Article

Keywords:

Posted Date: March 10th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2483307/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: (Not answered)

Version of Record: A version of this preprint was published at Communications Engineering on December 8th, 2023. See the published version at https://doi.org/10.1038/s44172-023-00136-6.

Abstract

In histopathology, doctors identify diseases by characterizing abnormal cells and their spatial organization within tissues. Polarization microscopy and supervised learning have been proved as an effective tool for extracting polarization parameters to highlight pathological features. Here we present an alternative approach based on unsupervised learning to group polarization-pixels into clusters, which correspond to distinct pathological structures. For pathological samples from different patients, it is confirmed that such unsupervised learning technique can decompose the histological structures into a stable basis of characteristic microstructural clusters, some of which correspond to distinctive pathological features for clinical diagnosis. Using hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC) samples, we demonstrate how the proposed framework can be utilized for segmentation of histological image, visualization of microstructure composition associated with lesion, and identification of polarization-based microstructure markers that correlates with specific pathology variation. This technique is capable of unraveling invisible microstructures in non-polarization images, and turn them into visible polarization features to pathologists and researchers.

Introduction

In histopathology doctors study the abnormality of tissue microstructure caused by diseases [1]. For instance, cancers are closely related to various microstructure alterations, such as changes in cellular nucleus size and anisotropy [1], collagen organization [2], or even composition of organelles in cytoplasm [3,4]. As the gold standard in pathology, hematoxylin and eosin (H&E) staining reveals tissue morphology and microstructure by dyeing the nucleus and extracellular matrix with varying colors of blue and pink [5]. Since malignant lesions induce changes in tissue microstructures, cancer-type-specific microstructural patterns observed on H&E slides provide crucial diagnostic information.

It has been known that interactions between polarized photons and complex media encode comprehensive microstructural information. A Mueller matrix describes the sample's polarization properties which are closely related to the microstructural features [6,7], particularly super resolution features at subwavelength scale [8,9]. Mueller matrix microscopy is capable of mapping tissue architecture down to subcellular level [9,10]. Being treated as an inverse problem, the pixel-level correlation between microstructure and polarization features are usually established by taking supervised learning approaches [11–14], as illustrated in Fig. 1. Pathologist segments regions on the H&E-stained pathological images to provide spatial labeling of the interested pathological structures, which in turn induces labeled pixels on the co-registered polarization images as the ground truth. Then polarization features corresponding to the labeled pixels are identified using supervised machine learning techniques. Such analysis can be performed on a low-magnification optical system, based on the finding that polarimetric imaging's contrast mechanism is insensitive to image resolution [15]. This approach has been used to study a variety of biological tissue samples through the extraction of microstructure-specific polarization features [11,14,16,17]. In particular, microstructures in liver tissue during lesion have been analyzed to extract

the simplest form of polarization feature parameter for cancerous cell nucleus identification in hepatocellular carcinoma samples[16].

A Mueller matrix image encodes abundant microstructural patterns that potentially correlate with pathological variation. In this paper, we advocate an unsupervised learning approach to clustering pixels based on distinctive polarization features, which allows for identifying specific spatial organization via projection from clustered pixels to H&E images. Such unsupervised learning approaches provide advantages for experienced pathologists to exploit the rich information in Mueller matrix images and their medical expertise to identify pathological features which may not be clearly seen under normal optical microscopes in clinics.

In this study, we take full advantage of the microstructural information encoded in Mueller matrix images by adopting the unsupervised approach, as illustrated in Fig. 1. We cluster the polarization pixels and discover that these polarization clusters correspond to distinct pathological structures. Such correspondence between clusters and pathological structures appears resilient to individual differences of patients. It indicates that by combining Mueller matrix imaging with unsupervised pixel-level clustering, we can effectively decompose the histological structure into its basic constituent microstructure subtypes. We can further extract the microstructure subtypes that are sensitive to pathological variations, identifying them as polarization markers for assisted diagnostic purposes. We demonstrate the viability of the proposed method by applying it to analyze the H&E-stained pathological slides of liver cancer, which has been examined previously using Mueller polarimetry and supervised learning [16], and show that the identified clusters in polarization feature space segment a variety of pathological structures. We propose a set of promising polarization markers that can distinguish intrahepatic cholangiocarcinoma (ICC) from hepatocellular carcinoma (HCC), recognize HCC lesion, and identify HCC with different cellular differentiation degrees. This framework of dividing tissues into microstructural subtypes will be referred to as the tissue microstructural composition analysis in this article. The framework allows the exploration of the super-resolution microstructures, which are invisible in the non-polarization optical images for routine clinical diagnosis. Optimistically, it may assist pathological structure identification and quantitative diagnostic evaluation, and become a new tool for pathological research and practice.

Method

Samples. Hepatocellular carcinoma (HCC) accounts for 70% of primary liver tumor, one of the cancers with the worst survival rate (20%) [18]. Undoubtedly, accurate staging of HCC subtypes has significant clinical implications. As the second most common type of primary liver tumor, intrahepatic cholangiocarcinoma (ICC) is much more aggressive than HCC, and the differentiation between them remains challenging [19–22]. Here, we use HCC and ICC samples to demonstrate the validity of our proposed method. Pathological slides of all patients are provided by Fujian Medical University Cancer Hospital. Sampled from 41 patients, a total of 222 ROIs are used for the unsupervised model construction, in which 181 ROIs are labeled. Among them, 71 ROIs are from ICC samples, 85 ROIs are

from HCC samples, and 25 ROIs are from normal tissues around lesions. The study was approved by the Ethics Committee of Fujian Medical University Cancer Hospital.

Mueller matrix images. Histological slides obtained from biopsy are measured using Mueller matrix microscope under 20x 0.4NA objective lens. The choice of this moderate NA allows us to achieve spatial resolution of 1 micron for resolving subcellular structures but maintain optimal accuracy in the Mueller matrix measurements. The homemade Mueller matrix microscope was upgraded from a low-cost commercial upright transmission microscope by installing polarization state generator and analyzer in the existing optical path, based on fixed polarizer and rotating waveplate setup [10,23]. It is capable of taking a Mueller matrix image in about 3 minutes with less than 0.02 average root mean square error (RMSE). Details for optimization and calibration of the microscope can be found in previous studies [10,23–25]. Recently, the development of DoFP (division of focal plane) Mueller microscope [26] pushes both imaging speed and accuracy to the next level. Compared to the dual waveplate method, DoFP method's acquisition time decreased by more than a factor of 10 while improving the RMSE by a factor of 2 [26]. RMSE can be reduced further by rotating both the polarizer and the wave plate in the polarization state generator [27]. Taking advantage of such instrumental advancement, DoFP Mueller microscope is used in complement to the dual rotating waveplate Mueller microscope during polarimetric data acquisition.

Data preprocessing: polarization super-pixel. Once the Mueller matrix images are obtained, a Gaussian filter is applied to increase the signal-to-noise ratio. Through the sum decomposition of Mueller matrix introduced by Cloude, the physically unrealizable part of Mueller matrices is filtered [28].

Considering the large number of polarization pixels to be processed, a multi-scale strategy is taken in our scheme. Instead of preserving the entirety of local information by considering all the pixels, we characterize the local subset of neighboring points in polarization feature space with their centroids, and perform unsupervised algorithms on the set of centroids. Effectively, computing at the scale of super-pixel level instead of pixel level allows us to ignore the detailed local structure of polarization data and focus on the global structure in polarization space, improving robustness of the result. Such super-pixel approach is a proven method to reduce data redundancy while decreasing computational complexity for downstream tasks [29–32]. On demand, the super-pixel can be unpacked and analyzed for more detailed features once the supe-pixels of interests are selected.

Using the described super-pixel approach, pixels with similar polarization characterization are grouped, creating N polarization super-pixels for each ROI. Note that the spatial constraint of generic super-pixels is waived since it increases the compression rate for polarization data. N can be chosen based on the complexity of the images in polarization space. A larger N preserves more polarization information, while a smaller N leads to more compact data representation. Since Mueller matrix is a comprehensive representation of polarimetric data, the super-pixels are created based on the Mueller matrix images combined with the minibatch KMeans algorithm [33]. The super-pixel method is utilized to compress the volume of polarimetric data while preserving its main polarization features [13]. It is otherwise impossible

to process the vast amount of pixel-level data points with unsupervised techniques. Through the aggressive use of super-pixels, the data volume is decreased by 3 orders of magnitude.

Polarimetric basis parameters. Mueller matrix describes comprehensively the sample's polarization properties, which provides extensive microstructural information. However, the explicit relationships between Mueller matrix elements and the characteristic microstructures are usually unclear, and many of the elements are sensitive to sample orientation. To accommodate this shortcoming, various decomposition [34–36] and transformation [37–40] methods have been proposed to provide polarization parameters which are functions of Mueller matrix elements but tend to have more explicit interpretability from the perspective of physics. These polarization parameters are used as the polarization basis parameters (PBP) for further statistical analysis [11,14].

UMAP and clustering. Polarimetric data lie on a nonlinear high-dimensional manifold, so we used uniform manifold approximation and projection (UMAP) for analysis, which offers nonlinear dimensionality reduction and projection of unseen data [41]. UMAP reveals the underlining structure in the polarization data, and the clusters in the UMAP space may correspond to specific pathological structures, which are annotated and studied. UMAP is used to visualize the global structure of polarization features in this paper. After normalizing the super-pixels into zero mean and unit variance, they are fed into the UMAP model using Canberra distance and 30 neighbors. We chose the Canberra distance since it is robust to outliers and sensitive to data around the origin. With polarimetric data, Canberra distance produces superior clustering results than conventional metrics such as Euclidean and Manhattan distance, a phenomenon observed in other data types as well [29,42,43]. In UMAP model, the number of neighbors parameter controls the size of local neighborhood during the manifold approximation process. Decreasing the parameter means preserving more local information and shortening the computation time. Empirically, number of neighbors is set as 30 so it is near the sweet spot for both global structure preservation and optimal computation time.

Once the data is projected onto the two-dimensional UMAP axes, hierarchal clustering is applied [44]. Single linkage is used, where the distance between two clusters is determined by the pair of points that are the closest to each other, one from each cluster. To focus on the main structure of the data, the points with low surrounding density are removed before clustering, by using a 200x200 regular grid, and using 11 as the cut-off threshold to remove the low-density areas. The cluster labels of the removed points are recovered using a semi-supervised approach, determined by label spreading after clustering is conducted using the trimmed data points [45]. In particular, a 41-fold cross-validation is conducted, where ROIs from 1 patient is reserved as unseen test data while the rest is used to construct the UMAP model. Especially, in each fold, we compute the Euclidean distance between the projection onto the UMAP axis of the test ROIs and its projection onto the UMAP axis established with the whole data from 41 patients. The mean distance is 0.012 with variance 0.00068, normalized with respect to the range of x- and y-axis, suggesting a stable UMAP configuration with respect to the training set.

To verify the plausibility of decomposing tissue structures into a stable basis of microstructural clusters using polarization imaging, we first analyzed 37 ROIs containing characteristic lesion structures, and closely monitored the changes in UMAP projection as we added more ROIs from new patients covering a variety of pathological structures. Using as little as 5% of the total dataset, we observe clusters forming in the UMAP projection. As we gradually increase the number of patients used during UMAP training process, the overall configuration converges, as shown in Fig. 2. Different structures require different amount of data to converge. For instance, the configuration of the two clusters on the left (cluster 1 and 2) starts to emerge with as few as 2 patients, and remains stable. Other clusters, especially the two clusters on the right (cluster 3 and 4), struggle during convergence due to their complex microstructure characteristic. Overall, we confirm that the macroscopic structure of the UMAP projection converges after roughly 20 patients (or equivalently 100 ROIs), and the layout remains consistent as the data size increases even further.

Results

Polarimetric feature based microstructural mapping. Since Mueller matrix describes how the optical and structural properties affect polarization states as the photons propagate and scatter in the media, the differences in polarization feature imply differences in the microstructures. Polarization super-pixels obtained from Mueller images are clustered in polarization space, creating a map of tissue microstructural clusters. Our study reveals that the pixels from the sample clusters are spatially correlated, and each microstructural cluster corresponds with a specific pathological feature.

Several distinct clusters are observed, as shown in Fig. 3a. The clusters are clearly separated into two main groups, and detailed clusters can be seen in each group. Separated by a clear gap, two subgroups (cluster 1 and 2) are identified on the left, along with a few scattered clusters in the bot left corner. The gap separating the clusters indicates that there is a distinct difference in polarization characteristic between cluster 1 and 2. On the other hand, the clusters on the right side are more convoluted compared to the left side. Cluster 6 is noticeably separated from the other clusters, but cluster 3, 4 and 5 are close to each other, with many visible subclusters. To interpret these clusters pathologically, we project the points from the polarization feature space onto the co-registered histological images. After the observation and evaluation by pathologists, the sample segmented structures from each polarimetric cluster are identified as displayed in Fig. 3a. Pathologists infer that the two clusters 1 and 2 on the left correspond to normal and cancerous cell nuclei respectively. Furthermore, cluster 2 potentially correlates with differentiation degree, which will be discussed in the next section. Recall that the clear gap dividing cluster 1 and 2 indicates distinctive polarization characteristics, which in turn infer that the cancerous nuclei is microstructurally distinctive from benign nuclei. On the top-right corner, cluster 3 and 4 are involved with cell cytoplasm in liver H&E-stained pathological tissues. The two clusters are connected, indicating similarities in polarization feature, and the corresponding structural features, between the clusters. Unlike the other clusters, cluster 3 and 4 has many distinguishable detailed subsets. It suggests cell cytoplasm has various complicated subtypes of different microstructure or optical properties, which is plausible considering the rich variety in sizes and shapes for organelles within cells. Cluster 5 and 6 mainly consist

of collagen fiber and fibrocytes respectively. In cluster 6, there are two identifiable subgroups, implying there are subtypes of fiber with different microstructure property. The unlabeled clusters correspond to noises or imaging artifacts. In short, the polarization pixels form clusters on the UMAP axes, and each cluster has a distinctive microstructure characterization. We discover that pixels clustered in polarization feature space are clustered spatially in histopathology images as well, segmenting pathologically meaningful structures such as normal and cancerous cell nucleus, cytoplasm, fibrocytes, and collagen fiber.

Lesion induced microstructural alteration. Pathological alteration of tissues may induce changes in tissue microstructure, and here we aim to visualize the microstructure transition at different stages of cellular differentiation. Specifically, cellular differentiation affects tissue microstructures, and such changes in microstructure composition can be revealed in the polarization feature space.

To explore the concept of visualizing microstructure transition during pathological changes, we collected HCC samples at different stages of differentiation degrees, and visualized the microstructure composition at each stage using a density heatmap on the UMAP atlas. Figure 3b demonstrates how ROIs from normal, well differentiation, moderate differentiation, and poor differentiation HCC samples differ in microstructural composition, represented by the UMAP density heatmap from the respective tissue samples. There seems to be a characteristic density distribution of polarization signatures at each stage of cellular differentiation degree, where the overall layout of the map remains stable, but the proportion and distribution of pixels that belongs to each cluster varies. The density ratios of several clusters appear to alter distinctly with different cellular differentiation degree: the density of cluster 2 (identified as cell nuclei by pathologists in pathological images) and cluster 6 (recognized as fibrocyte) both increase monotonically as the differentiation degree decreases. On the other hand, the lower right part of cluster 3 (realized as cytoplasm) vanishes gradually as differentiation degree decreases, while the upper left part of cluster 3 changes abruptly between well and moderately differentiated cellular states. Cluster 2 is clearly absent in benign tissues but present in malignant tissues, which implies that the presence of this specific microstructure characterization correlates strongly with tumor malignancy. Such polarization signature can be potentially used as the marker for HCC malignancy detection, providing a quantitative way to identify HCC tumor. The idea is thoroughly explored in the polarization marker section. Another point we noticed is that the configuration of the primary clusters is fixed, while the composition of clusters and the local density of data points varies with respect to cellular differentiation. To shortly summarize: (1) polarization based microstructural clusters contain pathologically significant information such as cellular differentiation degree, and (2) the principal configuration of the microstructural map in UMAP representation remains stable; only the inter-cluster proportion changes with pathological variation.

Cluster dendrogram. Each cluster has its unique polarization characteristic and corresponding pathological feature, and the dendrogram plot in Fig. 4 illustrates how the pixels are separated into clusters using polarization features. Dendrogram illustrates how the clusters are formed, where the node indicates two clusters are merged into one, and the abscissa-axis indicates the distance between clusters.

Each branch corresponds to a certain variation in polarization features, and variation in polarization features infer variation in the microstructure. As shown in Fig. 4, the entire set first splits into two main groups, which contain nucleus structures (cluster 1, 2) and the rest (cluster 3, 4, 5, 6), including collagen fiber and cytoplasm. This indicates that cell nucleus and collagen fiber have very different microstructural properties, unsurprisingly. In the top strand, the two main nucleus clusters are extracted by stripping away the artifact clusters, mostly utilizing the equalities of Mueller matrices. In the bottom strand cluster 6 is the easiest to identify, as indicated by the between-cluster distance. The polarization microstructural map and dendrogram plot demonstrate that each subdivided cluster corresponds to a meaningful pathological microstructure in the H&E images, identifiable with distinctive polarization features.

Polarization markers for diagnosis. Tissue microstructure variation is pathologically meaningful. Here, we aim to extract polarization-based microstructure markers and explore their diagnostic value. Subwavelength microstructures, such as that of the cytoplasm, are potential polarization markers as well. Once the tissue microstructural subtypes are mapped using polarization features, it enables the visualization of pathological transition during primary liver cancer progression and the identification of potential polarization markers for tumor classification. To demonstrate, we attempt to tackle two different tasks: identification of HCC with different differentiation degrees, and classification of ICC from HCC.

We first study the microstructural variation during HCC progression and recognize cancerous tissues with different differentiation degrees. We observe the tissue microstructural composition alteration correlates with pathological variation in HCC, from normal liver tissues to well, moderate, and poor differentiation degree of HCC, which is clearly visualized in Fig. 5a, and the animated heatmap provided in *Supplementary 1*. As indicated in Fig. 4a and Fig. 5a, we identified cluster 2 and 6 as the polarization marker to distinguish HCC from normal tissue. The cluster area proportion (area of pixels belonging to that specific cluster divided by the total area in the ROI) of cluster 2 and cluster 6 are calculated for normal tissue, highly differentiated, moderately differentiated, and poorly differentiated cancer samples respectively. The resulting box-whisker plots are shown in Fig. 5b, and their corresponding p-values for the t-test are calculated to test for statistical significance. When using cluster 2 as the marker, the distinctions between neighboring sets are statistically significant, but not between the well differentiated and moderately differentiated samples ($p = 2.22x10^{-5}$ for normal vs well, $p = 1.00\ x10^{-4}$ for moderate vs poor, $p = 1.16x10^{-1}$ for well vs moderate). The monotonically increasing trend of cluster 2 proportion is observed as well. To discriminate malignant HCC samples from normal tissues, we have separated all the samples into two sets, namely the malignant set (containing HCC samples of all three differentiation degrees) and the normal set (containing all the normal ROIs). Under this binary classification scenario, cluster 2 area proportion achieves an AUC of 94.84%. Using cluster 6 as the polarization marker shows a similar trend, as seen in Fig. 5c. The area proportions of cluster 6 in normal tissues are relatively low, most of which are under 5%. The separation between normal and highly differentiated samples ($p = 2.69x10^{-2}$), and between moderately and highly differentiated samples ($p = 1.48x10^{-4}$) are statistically significant, but not between highly and moderately differentiated samples ($p = 5.87x10^{-1}$).

As labeled in Fig. 5d, we identified a cytoplasm structure in cluster 3 that is present in moderately differentiated samples, but not in highly differentiated samples. Using the labeled region as polarization marker, an AUC of 88.59% is achieved for distinguishing well differentiated from moderately differentiated samples, as shown in Fig. 5e. This observation is quite unconventional and requires validation in future works.

We now study the differences in tissue microstructural composition between HCC and ICC pathological samples, in an attempt to distinguish them. Likewise, we can first visualize the variation of tissue microstructural composition by observing the animated heatmap (provided in *Supplementary 2*) that samples gradually from HCC to ICC tissues. The microstructures characterized by cluster 5, i.e., collagen fiber, is abundant in ICC samples, but not in HCC samples, as seen in Fig. 6a. In contrast, cluster 3, a subtype of cytoplasm, is abundant in HCC sample while not in ICC samples. Cluster 5 is the main focus as the polarization marker for HCC and ICC distinction. For comparison, Fig. 6b shows the box-whisker plot of the cluster 5 area proportion for both HCC and ICC samples. It is observed that most of HCC samples' cluster 5 proportion ratio is less than 10%, while that of the majority of ICC samples are above 10%. For discriminating ICC from HCC samples, the area proportion of cluster 5 yields an AUC of 84.94%. We also experimented with cluster 3, using it as the polarization marker yields an AUC of 71.69%, as seen in Fig. 6c. This implies that to a certain degree, the cytoplasm composition in HCC and ICC are different.

Discussion

Here, we have introduced a tissue microstructural composition analysis framework using Mueller microscopy and unsupervised learning methods. It is demonstrated that the resulting pixel clusters of Mueller images correspond to compositions of characteristic microstructure which contain pathologically significant information. We analyzed the cancerous tissues and the normal tissues around lesions in liver pathological samples, and split the tissue structures into 6 components, each with its own distinctive polarimetric and microstructural feature. The clustered pixels in polarization feature space are correlated spatially, and by projecting the cluster labels onto histological images we find that they segment various pathological structures. In particular, the six clusters can be respectively associated with the normal nucleus, cancerous nucleus, HCC and ICC cytoplasm, collagen fiber, and fibrocytes. The difference in polarization characteristics implies a difference in microstructural characteristics. It is inferred that there are certain underlining microstructural differences between the identified clusters, causing their differences in polarization signature.

We demonstrate how polarization markers can be extracted and potentially utilized for pathological applications. Tissue microstructural composition analysis enables visualization of tissue structure transition in polarization space during pathological variation. It is shown that cluster 2 correlates strongly with tumor. Evidence shows that cluster 2 can serve as a polarization marker to quantitatively characterize HCC samples with different differentiation degrees and separate HCC from normal liver tissues with an AUC of 95%. Differentiation of ICC from HCC is a known challenging pathological task. In this study, we identify a polarization marker to distinguish ICC from HCC on the H&E-stained slides,

achieving an AUC of 84.94%. Through the transformation from polarization feature space to pathological feature space, it informs pathologists that cytoplasm and fibrocytes contribute the most toward classification, which is consistent with the findings of medicine [46]. Therefore, this method may supplement the immunostaining analysis, providing interpretable and quantitative polarization markers for pathological diagnosis and prognosis prediction, and even reveal “invisible” super-resolution features to pathologists in an intuitive and visualized manner.

By adjusting the size of super-pixels and selecting interested microstructural clusters, it is possible to split an identified structure into even finer clusters, revealing detailed subcellular microstructures. To demonstrate, pixels from cluster 1 and 2 are collected and analyzed using the proposed method, only at a much smaller super-pixel size. As shown in Fig. 7a, the two clusters in the original UMAP are now transformed into four or more clusters. Figure 7b compares the heatmap between benign and malignant nuclei on the newly generated UMAP projection, it is clear that two distinct clusters correlate strongly with malignancy. Such a detailed pattern is only revealed after focusing on the nucleus structure and increasing the spatial resolution of polarimetric images.

In conclusion, by clustering the pixels of Mueller matrix images, we develop a tissue microstructural composition analysis framework to separate the intricate tissue structures into detailed microstructural components with distinctive polarization characteristics. Evidences show that the microstructure clusters are pathologically meaningful and their overall layout remain stable in polarization space with respect to samples from different individuals. The framework is applied to analyze the two most common primary liver cancers, HCC and ICC. The detailed map of biological microstructure in polarization space allowed the visualization of microstructure composition variation during HCC occurrence and progression, and the difference in microstructure between HCC and ICC. We also derived sets of polarization markers to tackle them quantitatively, yielding comparable performance with other state-of-the-art methods. Most importantly, even the “invisible” microstructure in pathology, the detailed structures of cytoplasm, could be presented in a “visible” manner and used for pathological diagnosis. Our results highlight the potential of polarization-based tissue compositional analysis framework as a tool to assist routine pathological tasks and explore the unknown pathological mechanism for pathologists and researchers.

Declarations

Author contributions

Conceptualization: JW, YD, HM, CL

Methodology: JW, YD, YY, HM, HH, RL, NZ, HP

Investigation: JW, YD, HM, CL, RH

Visualization: JW, HP

Supervision: HM, CL

Writing—original draft: JW, YD, HM, CL

Writing—review & editing: HM, CL, RH, YY, WX, RP, XT

Acknowledgement We thank Professor Jie Chen from Peking University, Chao He from Oxford University, Dr. Anli Hou and Dr. Shan Du from University of Chinese Academy of Sciences Shenzhen Hospital, for their invaluable feedback on the manuscript.

Funding National Natural Science Foundation of China (11974206, 61527826); Shenzhen Bureau of Science and Innovation (JCYJ20170412170814624); Joint Funds for the innovation of science and Technology, Fujian province (2021Y9212)

Disclosures The authors declare no conflicts of interest.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Kumar, V., Abbas, A. K., Aster, J. C. & Elsevier, C. *Robbins & Cotran pathologic basis of disease*. Tenth edition, (Elsevier, 2020).
2. Xu, S. *et al.* The role of collagen in cancer: from bench to bedside. *J. Transl. Med.* **17**, 309 (2019).
3. Osborn, M. & Weber, K. Tumor diagnosis by intermediate filament typing: a novel tool for surgical pathology. *Lab. Invest.* **48**, 372-394 (1983).
4. Nelson, R. S. & Epstein, J. I. Prostatic Carcinoma with Abundant Xanthomatous Cytoplasm: Foamy Gland Carcinoma. *Am. J. Surg. Pathol.* **20**, 419-426 (1996).
5. Fischer, A. H., Jacobson, K. A., Rose, J. & Zeller, R. Hematoxylin and eosin staining of tissue and cell sections. *Cold Spring Harb. Protoc.* **2008**, pdb.prot4986 (2008).
6. Ghosh, N. & Vitkin, A. Tissue polarimetry: concepts, challenges, applications, and outlook. *J. Biomed. Opt.* **16**, 110801 (2011).
7. He, C. *et al.* Polarisation optics for biomedical and clinical applications: a review. *Light Sci. Appl.* **10**, 194 (2021).
8. Adhikari, A., Dev, K. & Asundi, A. Subwavelength metrological chracterization by Mueller matrix polarimeter and finite difference time domain method. *Opt. Lasers Eng.* **86**, 242-247 (2016).
9. Du, E. *et al.* Mueller matrix polarimetry for differentiating characteristic features of cancerous tissues. *J. Biomed. Opt.* **19**, 076013 (2014).

10. Wang, Y. *et al.* Differentiating characteristic microstructural features of cancerous tissues using Mueller matrix microscope. *Micron.* **79**, 8-15 (2015).
11. Dong, Y. *et al.* Deriving Polarimetry Feature Parameters to Characterize Microstructural Features in Histological Sections of Breast Tissues. *IEEE Trans. Biomed. Eng.* **68**, 881-892 (2021).
12. Ivanov, D. *et al.* Polarization and depolarization metrics as optical markers in support to histopathology of ex vivo colon tissue. *Biomed. Opt. Express* **12**, 4560-4572 (2021).
13. Heinrich, C. *et al.* Mueller polarimetric imaging of biological tissues: classification in a decision-theoretic framework. *J. Opt. Soc. Am. A* **35**, 2046 (2018).
14. Dong, Y. *et al.* A Polarization-Imaging-Based Machine Learning Framework for Quantitative Pathological Diagnosis of Cervical Precancerous Lesions. *IEEE Trans. Med. Imaging* **40**, 3728-3738 (2021).
15. Liu, Y. *et al.* Comparison between image texture and polarization features in histopathology. *Biomed. Opt. Express* **12**, 1593-1608 (2021).
16. Wan, J. *et al.* Polarization-based probabilistic discriminative model for quantitative characterization of cancer cells. *Biomed. Opt. Express.* **13**, 3339-3354 (2022).
17. Li, P. *et al.* Polaromics: deriving polarization parameters from a Mueller matrix for quantitative characterization of biomedical specimen. *J. Phys. D: Appl. Phys.* **55**, 034002 (2021).
18. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2021. *CA Cancer J. Clin.* **71**, 7-33 (2021).
19. Ryu, H. S. *et al.* Comparative analysis of immunohistochemical markers for differential diagnosis of hepatocelluar carcinoma and cholangiocarcinoma. *Tumori.* **98**, 478-484 (2012).
20. Vijgen, S., Terris, B. & Rubbia-Brandt, L. Pathology of intrahepatic cholangiocarcinoma. *Hepatobiliary Surg. Nutr.* **6**, 22-34 (2017).
21. Buettner, S., van Vugt, J. L., JN, I. J. & Groot Koerkamp, B. Intrahepatic cholangiocarcinoma: current perspectives. *Onco. Targets Ther.* **10**, 1131-1142 (2017).
22. Banales, J. M. *et al.* Expert consensus document: Cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). *Nat. Rev. Gastroenterol Hepatol.* **13**, 261-280 (2016).
23. Azzam, R. M. Photopolarimetric measurement of the Mueller matrix by Fourier analysis of a single detected signal. *Opt. Lett.* **2**, 148-150 (1978).
24. Smith, M. H. Optimization of a dual-rotating-retarder Mueller matrix polarimeter. *Appl. Opt.* **41**, 2488-2493 (2002).
25. Meng, R. *et al.* Comparison of different calibration methods for Mueller matrix microscopy of cells. *Appl. Opt.* **60**, 1380-1386 (2021).
26. Huang, T. *et al.* Fast Mueller matrix microscope based on dual DoFP polarimeters. *Opt. Lett.* **46**, 1676-1679 (2021).

27. Zhao, Q. *et al.* Geometric Optimization Method for Polarization State Generator of Mueller Matrix Microscope. *Opt. Lett.* **46**, 5631-5634 (2021).
28. Cloude, S. Conditions For The Physical Realisability Of Matrix Operators In Polarimetry. Proc. SPIE **1166**, Polarization Considerations for Optical Systems II (1990).
29. Bouguettaya, A., Yu, Q., Liu, X., Zhou, X. & Song, A. Efficient agglomerative hierarchical clustering. *Expert Syst. Appl.* **42**, 2785-2797 (2015).
30. Achanta, R. *et al.* SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. on Pattern Anal. and Mach. Intell.* **34**, 2274-2282 (2012).
31. Subudhi, S., Patro, R. N., Biswal, P. K. & Dell'Acqua, F. A Survey on Superpixel Segmentation as a Preprocessing Step in Hyperspectral Image Analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **14**, 5015-5035 (2021).
32. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226-1232 (2019).
33. Sculley, D. Web-scale k-means clustering. *WWW '10: Proceedings of the 19th international conference on World wide web*, 1177-1178 (2010).
34. Cloude, S. R. Group theory and polarisation algebra. *Optik* **75**, 26-36 (1986).
35. Lu, S.-Y. & Chipman, R. A. Interpretation of Mueller matrices based on polar decomposition. *J. Opt. Soc. Am. A* **13**, 1106-1113 (1996).
36. Ossikovski, R. Differential matrix formalism for depolarizing anisotropic media. *Opt. Lett.* **36**, 2330-2332 (2011).
37. Fry, E. S. & Kattawar, G.W. Relationships between elements of the Stokes matrix. *Appl. Opt.* **20**, 2811-2814 (1981).
38. Gil, J. Invariant quantities of a Mueller matrix under rotation and retarder transformations. *J. Opt. Soc. Am. A* **33**, 52-58(2015).
39. He, H. *et al.* Mueller Matrix Polarimetry—An Emerging New Tool for Characterizing the Microstructural Feature of Complex Biological Specimen. *J. Light. Technol.* **37**, 2534-2548 (2019).
40. Li, P., Lv, D., He, H. & Ma, H. Separating azimuthal orientation dependence in polarization measurements of anisotropic media. *Opt. Express* **26**, 3791-3800 (2018).
41. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1802.03426 (2018).
42. Rivera, A. J., Pérez-Godoy, M. D., Elizondo, D., Deka, L. & del Jesus, M. J. Analysis of clustering methods for crop type mapping using satellite imagery. *Neurocomputing* **492**, 91-106 (2022).
43. Yang, Y. M. *et al.* Determining the Number of Instars in Simulium quinquestriatum (Diptera: Simuliidae) Using k-Means Clustering via the Canberra Distance. *J. Med. Entomol.* **55**, 808-816 (2018).
44. Zepeda-Mendoza, M.L. & Resendis-Antonio, O. Hierarchical Agglomerative Clustering. in *Encyclopedia of Systems Biology* (eds. Dubitzky, W., Wolkenhauer, O., Cho, KH., & Yokota, H.), 886-887

(Springer New York, 2013).

45. Zhou, D., Bousquet, O., Lal, T., Weston, J. & Olkopf, B. Learning with Local and Global Consistency. *Adv. Neural Inf. Process. Syst.* **16** (2004).
46. Nagtegaal, I. D. *et al.* The 2019 WHO classification of tumours of the digestive system. *Histopathology* **76**, 254-259 (2020).

Figures

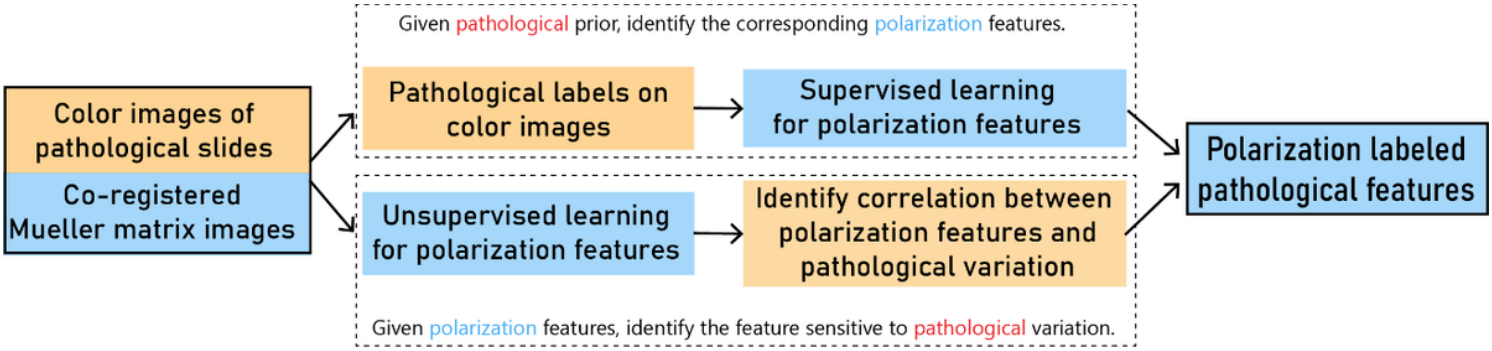


Figure 1

Two different frameworks to identify correlation between polarization and pathological features.

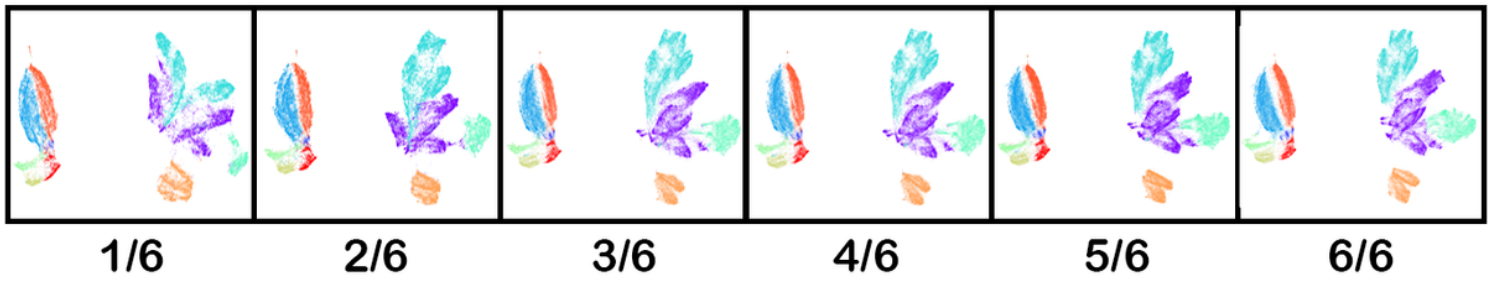


Figure 2

Convergence of the overall configuration in UMAP space as the number of ROIs increases. The numbers indicate the fraction of data used, increasing at a step of 1/6, equivalent to 37 ROIs from samples of 6-7 patients. Above the numbers are the corresponding UMAP projection using the specified fraction of total data.

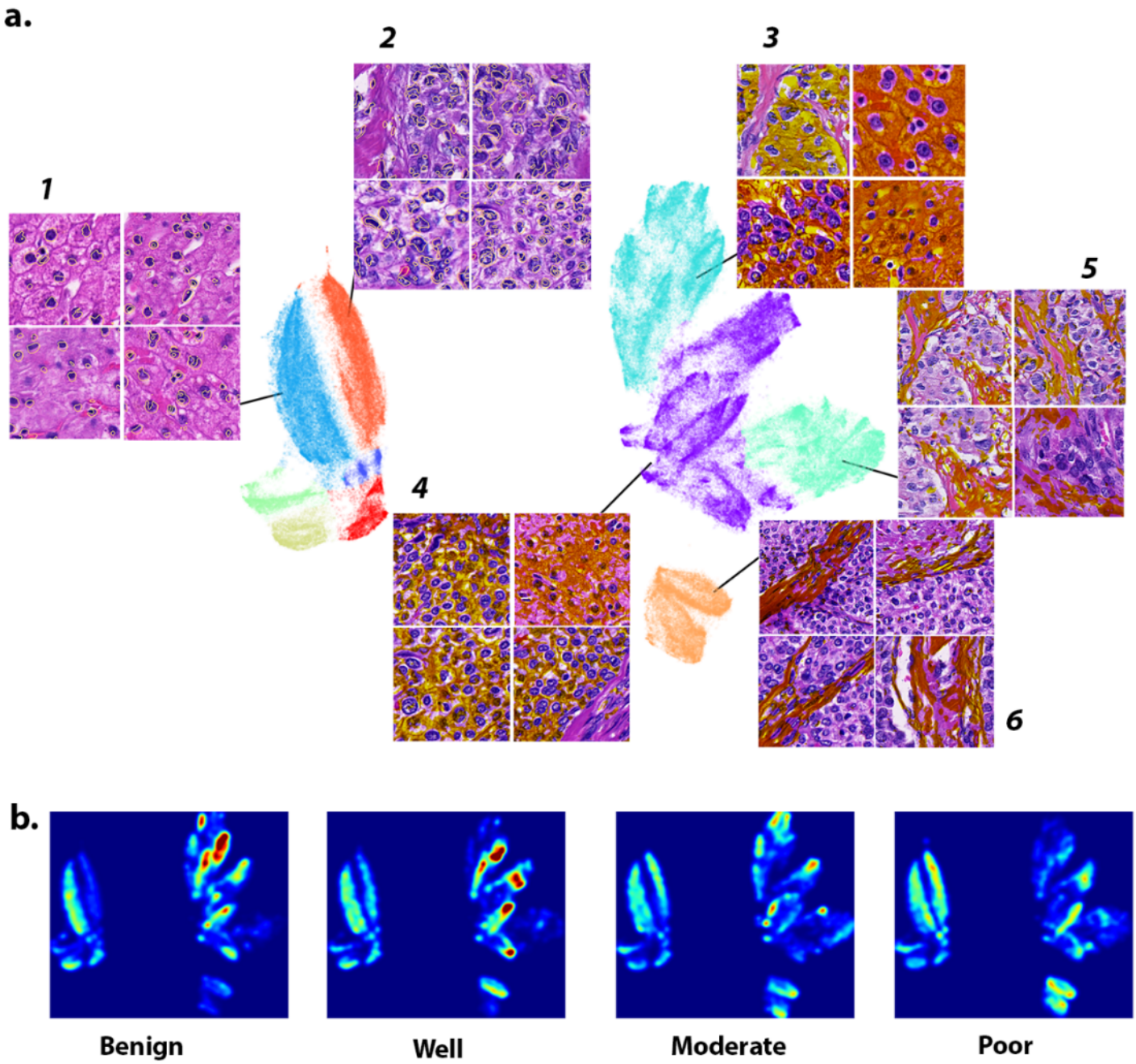


Figure 3

The UMAP clustering result of Mueller matrix pixel images and their corresponding pathological features. The 6 polarization feature clusters segment pathological features on the histological images (a), and pathologists identify them as cell nucleus (cluster 1 and 2), cytoplasm (cluster 3 and 4), and collagen fiber and fibrocyte (cluster 5 and 6). The heatmap of the UMAP projection is drawn at various HCC differentiation degrees (b), to help visualize variations in polarization features corresponding to pathological variation. An animated version of (b) is supplied in supplementary.

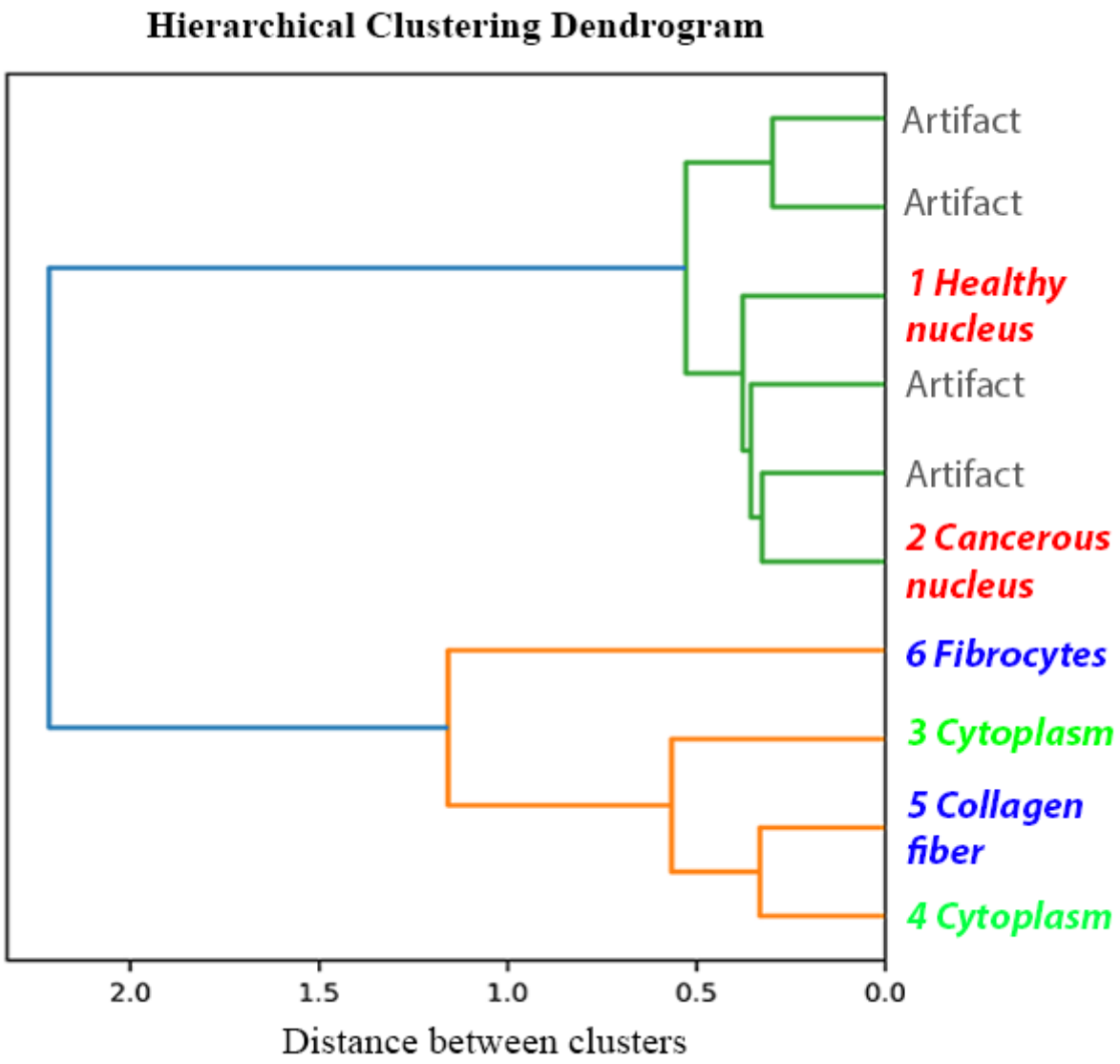


Figure 4

Dendrogram generated from hierarchical clustering process. X-axis shows the distance between clusters, and the y-axis labels the different clusters and their corresponding pathological structures.

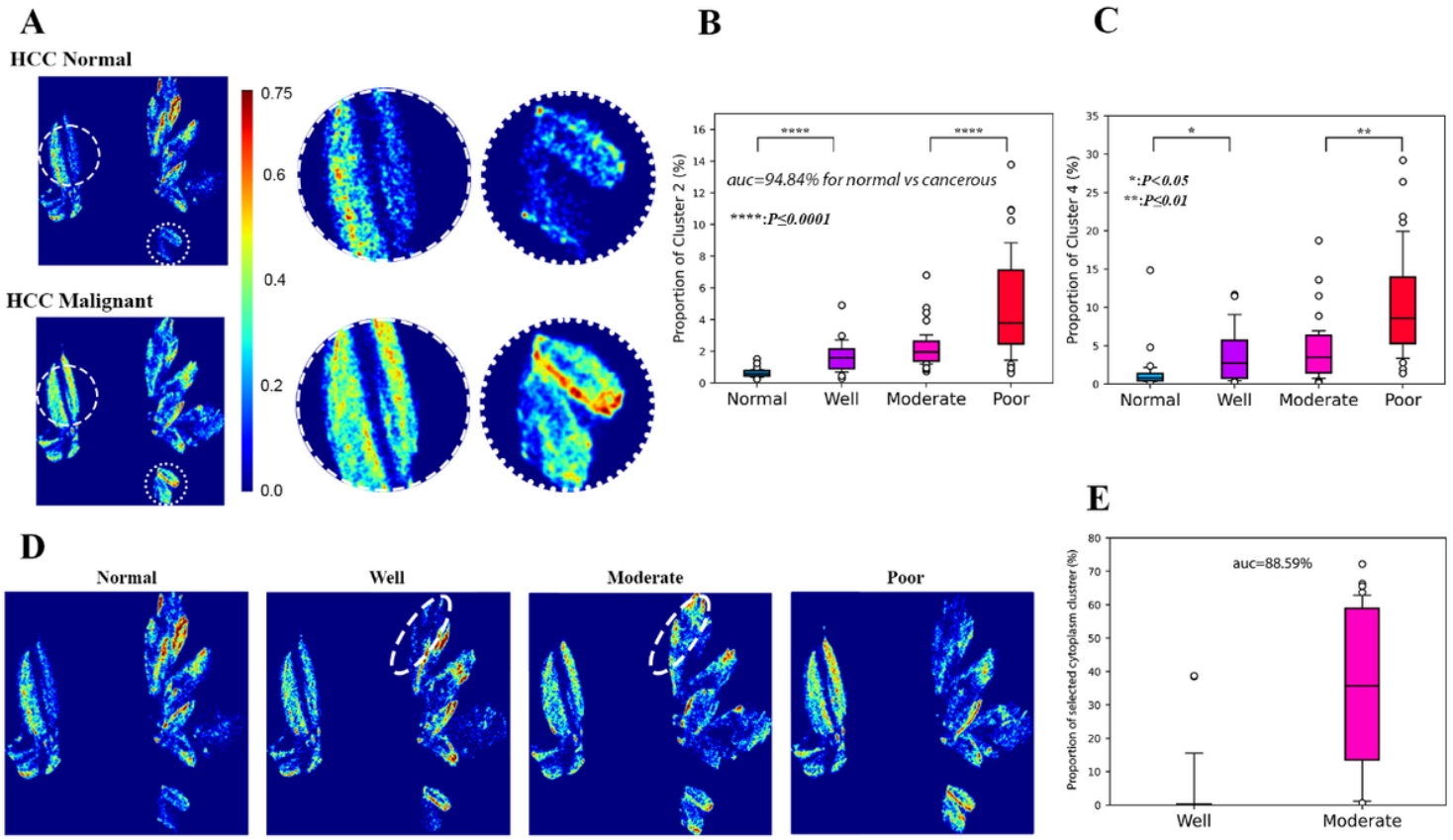


Figure 5

Identification of polarization markers that correlates with HCC differentiation degree. The logarithm scale density heat maps of normal and malignant HCC samples are compared in (a), and the two potential polarization markers are enlarged for better view. Proportion of pixels belongs to cluster 2 and 6 in each ROI are calculated, and the box whisker plots are shown in (b) and (c) for each differentiation degree. The box and the whiskers mark the 10,25,50,75, and 90 percentiles respectively. (d) displays the heat map for each differentiation degree, and the dashed line indicates the selected polarization marker in the cytoplasm cluster to differentiate well and moderately differentiated samples. (e) shows the box-whisker plot of the selected polarization marker in (d).

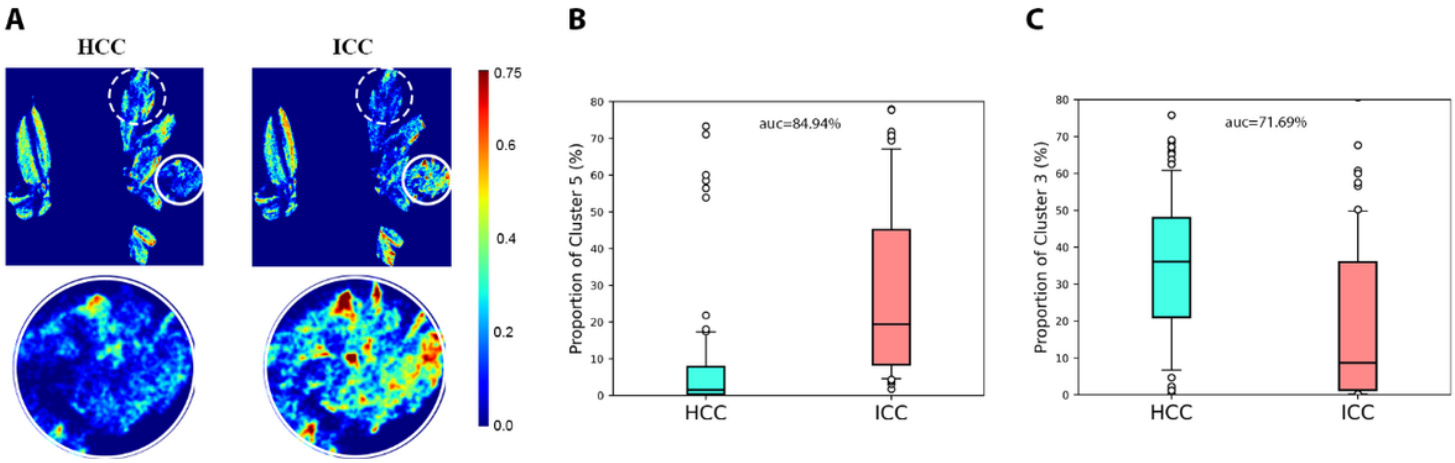


Figure 6

Identification of polarization marker for distinguishing HCC and ICC. (a) shows the logarithm scale density heat map for HCC and ICC samples, and two clusters are labeled, one with solid line (cluster 5) and one with dashed line (cluster 3). The box whisker plots for the two polarization markers are shown in (b) and (c), the box and the whiskers mark the 10,25,50,75, and 90 percentiles respectively.

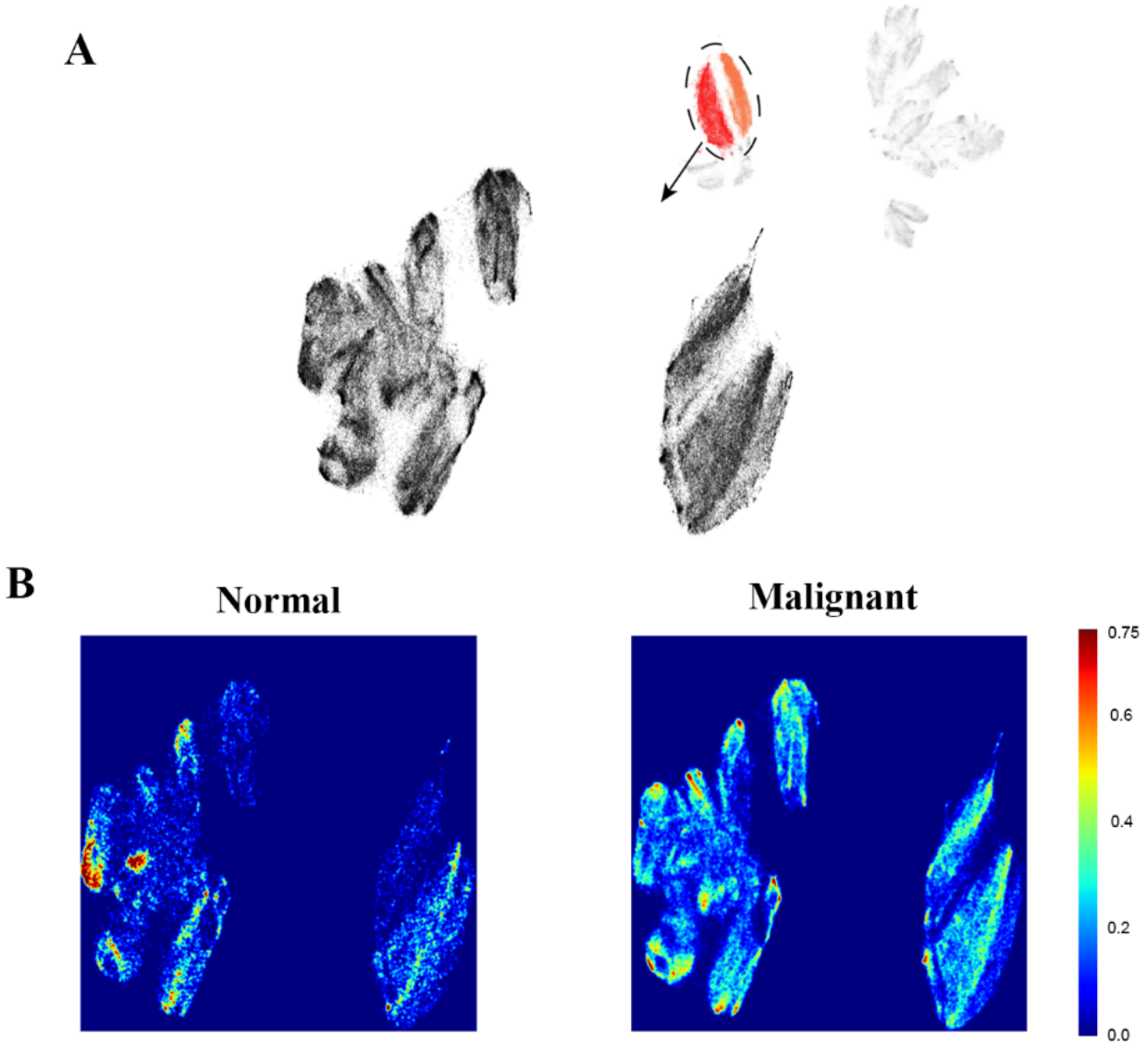


Figure 7

The pixels nucleus clusters, cluster 1 and 2, are collected and remapped using smaller sized super-pixels, as seen in (a). Detailed nucleus structures are revealed, two of which corelates with lesion malignant, as shown by the logarithm scale density plots in (b) of normal and malignant HCC tissues.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary1.avi
- Supplementary2.avi