Artificial intelligence-assisted stimulated Raman molecular cytology enables accurate and rapid intraoperative assessment of the surgical margin in pancreatic ductal adenocarcinoma

Tinghe Fang[1,#], Daoning Liu[2,#], Xun Chen[1,3,#], Keji Zhou[1], Chunyi Hao[2,*], Shuhua Yue[1,*]

[1]Key Laboratory of Biomechanics and Mechanobiology (Beihang University), Ministry of Education, Institute of Medical Photonics, Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China.
[2]Key laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Hepato-Pancreato-Biliary Surgery/Sarcoma Center, Peking University Cancer Hospital & Institute, Beijing, 100142, China.
[3]School of Engineering Medicine, Beihang University, Beijing, 100191, China.
[#]These authors contributed equally.
*Corresponding authors. Email: yue_shuhua@buaa.edu.cn (S.Y.); haochunyi@bjmu.edu.cn (C.H.)

ABSTRACT

Pancreatic ductal adenocarcinoma (PDAC) is a highly mortal cancer with surgical resection the only potentially curative treatment. The accurate intraoperative assessment of PDAC surgical margins is significant in guaranteeing resection adequacy and improving the patient's survival. However, the commonly used frozen-section pancreatic biopsy is limited by its strict requirement of practitioners and lack of objectivity. Here, we developed the multi-instance cytology with learned Raman embedding (MICLEAR), a label-free cytology based on chemical information reflected by Raman spectra. First, 4085 cells collected from 41 patients were imaged with stimulated Raman scattering (SRS) microscopy. Then, a contrastive learning (CL)-based cell embedding model was obtained to represent each cell with a concise vector that contained its morphological and componential information. Finally, a multi-instance learning (MIL)-based diagnosis model using cell vectors predicted the probability that the margin was positive. MICLEAR reached 80% sensitivity, 94.1% specificity, and 0.86 AUC on the pancreatic neck margin (PNM) samples from 27 patients. It holds promise for rapid and accurate assessment of PDAC surgical margins.

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) stands as the third leading cause of cancer-related deaths, with an abysmally low 5-year relative survival rate of approximately 10%[1,2]. The only potentially curative treatment currently available is surgical resection. Pancreatic tumors located at the head typically undergo pancreaticoduodenectomy, while those found in the body or tail of the pancreas can be managed with distal pancreatomy[3]. The assessment of surgical margin status is crucial in identifying any residual cancer cells and evaluating the adequacy of resection, which significantly impacts the patient outcome[4-5].

The surgical margin status in PDAC is classified into R0, R1, and R2 based on margin clearance quality. R2 and R1 denote positive margins with macroscopically and microscopically visible residual cancer cells, respectively. R0 denotes negative margins without residual cancer cells. Due to the dispersed growth pattern of PDAC cells, a stricter definition of R0, which requires no cancer cells within 1 mm around the surgical margin, is widely adopted[44]. The gold standard of margin assessment is permanent section (PS) pathology, which typically takes several days to provide results. For the intraoperative evaluation of resection adequacy, the frozen section (FS)

pathology, which typically takes 30 min[7], is routinely performed for assessment of significant transection margins, such as the pancreatic neck margin (PNM) and the bile duct margin (BDM)[6]. Extended resection is required when the FS pathology gives a positive diagnosis. Following the cessation of resection, the margin tissue is collected for PS pathology, providing the definitive margin status report.

However, the accurate assessment of pancreatic margins with FS pathology is challenging. The interpretation of pancreatic FSs is complicated due to thermal artifacts, confusion between PDAC and chronic pancreatitis, and pitfalls related to hyperplasia[12-14], thus requiring experienced pathologists. This reality contributes to the disparities seen in the reported false-negative rates of FS, which have been reported to range from 4.3% to 75% in studies conducted around the same period[9-11]. Besides, the surgeons' clinical impression is crucial to determine whether to re-resect when the FS diagnosis is negative[8,59]. Consequently, current intraoperative pancreatic margin assessment demands well-trained personnel and involves highly subjective decision-making. Therefore, an alternative method with improved operability and objectivity is desirable.

To meet these needs, several histological substitutes for FS have been proposed, such as fluorescent-guided surgery (FGS), photoacoustic tomography (PAT), optical coherence tomography (OCT), spontaneous Raman spectroscopy, and stimulated Raman scattering (SRS) microscopy. FGS requires costly translational studies before clinical trials. The spatiotemporal heterogeneity of tumors makes it difficult to select the optimal biomarkers and validate their efficacy in the clinical setting[45]. Its specificity is still low on the PDAC (~70%)[15,46]. PAT detects high-photoabsorbance biomolecules such as hemoglobin, lipid, and melanoma. The feasibility of identifying margin tumors with hemoglobin is still unproved[47], and assessment based on other biomolecules will substantially limit the generalizability. For example, the non-adipose breast tissues may be confused with tumors, reducing the specificity of margin assessment to 75%[48]. OCT exploits structural features of tissue but lacks chemical contrast. Its sensitivity and specificity in identifying PDAC are both less than 75%[17]. Spontaneous Raman spectroscopy has achieved sensitivity and specificity of over 90% in various margin assessments by the Raman spectra incorporating abundant componential information[49,50]. However, its data acquisition is time-consuming due to the narrow detection coverage. SRS microscopy maps the spatial distribution of Raman signals, thus combining morphological and componential information without exogenous labelling. It has been used for precise assessment of brain glioma margins[18,19]. Nevertheless, its current configuration requires large-scale tissue imaging to capture the tumor due to the substantial heterogeneity of tissues.

In addition to histological analysis, cytology has been developed to screen a broad spectrum of cancers, such as Pap test[20] for cervical cancer, urine cytology[21] for bladder cancer, and bronchoalveolar lavage fluid cytology[22] for lung cancer. The operability of cytology is much higher than that of histology. Besides, cytology collects the sample from the entire region to be detected, making benign and cancer cells evenly mixed and theoretically reducing the necessary field of view (FOV) and imaging time. However, the sensitivity of conventional cytology, which only exploits the morphological information of cells, is typically lower than 60%[23-25]. This problem can be significantly improved by detecting the altered metabolism of cancer cells, as demonstrated by cytology based on spontaneous Raman spectroscopy[51] or SRS microscopy[26]. Note that the former required several hours for spectra acquisition and thus cannot be used intraoperatively, while the latter can be finished within 30 min. Brush cytology can be used to diagnose pancreatic lesions owing to easy and rapid sampling in the pancreatic duct but only achieves ~50% sensitivity[52]. Based on the PDAC's aberrant cholesteryl esterification and accumulation of

cholesteryl ester revealed by SRS microscopy[27], we devised a refined brush cytology incorporating molecular information for PDAC margin assessment.

In this study, we developed SRS-based molecular cytology for intraoperative PDAC margin assessment and demonstrated it with PNMs. We imaged 4085 cells at the Raman bands of lipid and protein. Normal pancreatic cell samples (negative) were extracted from seven patients without pancreatic diseases, while tumor (positive) and margin cell samples (test samples) were from 34 PDAC patients. Then, a single-cell embedding model was trained with a contrastive learning (CL) framework and used to represent each cell with low-dimensional cell vectors. This method replaced manual feature extraction, which requires costly feature engineering with human bias in feature selection and extraction. Finally, with a multi-instance learning (MIL) algorithm, samples were subtyped by identifying significant cells that indicate positive samples. MIL is commonly used for artificial-intelligence-assisted pathology, in which only labels on the multi-instance (e.g., a slide) instead of single-instance (e.g., a pixel) level are accessible. The algorithms above constitute our proposed multi-instance cytology with learned Raman embedding (MICLEAR), achieving 80% sensitivity and 94.1% specificity on 27 patients, including 10 with positive PNMs and 17 with negative PNMs. The average test time was ~25 min for each patient. It holds great potential for accurate, easy-to-use, and rapid PNM assessment.

RESULTS

Workflow of Multi-instance Cytology with Learned Raman Embedding

The workflow of MICLEAR is shown in **Fig. 1**. The exfoliative cell samples were collected from three locations with medical cytology brush, including pancreatic tissues in healthy pancreases (N samples), cores of PDAC tumors (T samples), and PNMs (M samples) (**Fig. 1a**). N and T samples contained cell populations without and with cancer cells, respectively. The sample images were acquired by two-color SRS microscopy at the Raman bands for CH_2 stretching in lipids around 2850 cm^{-1} and CH_3 stretching in proteins around 2930 cm^{-1} (**Fig. 1b**). Example raw SRS images of N and T samples are shown in **Fig. S1**. Single cells were segmented from the SRS images at 2930 cm^{-1} with a pre-trained deep learning model (**Fig. 1c**). The segmentation mask was applied to the corresponding image at 2850 cm^{-1} to produce single-cell two-color SRS images, which were then converted to single-cell molecular images with matrix operation (**Fig. 1d**). Then, using the molecular images of N and T samples, a cell embedding model was trained to transform each cell into a 512-dimensional vector, in which a CL framework was applied to increase the distance between vectors. The obtained cell vectors were compact representations of morphological and componential information (**Fig. 1e**). Finally, a MIL model was trained to predict the positive probability of a sample with its cell vectors. The model was evaluated with leave-one-out cross-validation (LOOCV), in which one M sample was left out for testing and the other samples were used to train the MIL model. After the results of testing on 27 M samples were assembled, the evaluation metrics, including sensitivity, specificity, and area under the receiver operating characteristic curve (AUC), were calculated (**Fig. 1f**).

Single-cell Segmentation and Extraction of Componential Information

The setup of our homebuilt SRS microscope is shown in **Fig. 2a**. The individual cells were segmented with the Cellpose[30], a well-developed cell segmentation algorithm based on intensity gradients. Compared with the

conventional watershed and the Stardist[31], the Cellpose performed significantly better (0.86 global Dice coefficient, 1.03 cell number ratio, and 0.86 cellwise Dice coefficient as defined in the *Methods* section), especially in segmenting cells with nonconvex contours (**Fig. S2a**,**b**). Then, the Raman spectra of pure chemical samples (**Fig. 2b**) were utilized to decompose the intensity of two-color SRS images into the contributions of lipid and protein, which were proportional to their concentrations according to the linear property of SRS signals[32]. This processing produced single-cell molecular images composed of two chemical channels. T samples were found to contain cells with more lipid droplets (LDs), indicating the potential contribution of componential information (**Fig. 2c**). The distributions of four representative cell features (area, shape factor, LD number, and LD dispersity as defined in the *Methods* section) were compared among different sample types to analyze how these features may help to identify positive samples (**Fig. 2d**). The feature distributions differed a lot between cells in N and T samples but obvious overlaps still existed. Besides, the difference weakened between M(Negative) and M(Positive) samples, which may be attributed to a smaller proportion of cancer cells or their less significant malignant features. To address these issues, our MICLEAR algorithm incorporated two models: a CL-based representation learning model to automatically explore high-dimensional features to distinguish cells and a MIL-based model to recognize marker cells that indicate positive samples. The manual extraction of the abovementioned four cell features and the other eight ones is shown in **Fig. S3** and described in detail in the *Methods* section.

Contrastive Learning for Single-cell Dimensionality Reduction into Cell Vectors

Inspired by SimCLR[28], a classical CL framework typically used for self-supervised model training, we applied a CL-based cell embedding model to acquire compact cell representations with minimal prior knowledge about feature selection and extraction. Specifically, a single-cell molecular image with 400×400 pixels was compressed into a 512-dimensional vector, whose components denoted values of high-dimensional cell features and were associated with both cell morphology and composition. The training of the CL-based model attempts to increase the difference between the outputted vectors of different cells and decrease the difference between the outputted vectors of the same cell with random rigid transformations (**Fig. 3a**). In comparison with SimCLR, in which the outputted vectors are fed into a linear classifier for supervised classification, we extended it to the multi-instance tasks by using the vectors for weak-supervised classification with a MIL classifier. Besides, conventional CL focuses on refining the model parameters for optimal feature space and transferability. In contrast, ours focuses on the outputted cell vectors irrespective of what parameters the model learned.

After training the cell embedding model, we analyzed the cell vectors' distribution and correlation with cell features. The ISOMAP[39], a manifold learning method that maintains the geodesic distances between points, was used to further reduce the dimensionality of cell vectors into two to visualize the vector distribution (**Fig. 3b**). The vectors in the N samples roughly constituted a subset of those in the T samples. Based on this finding, we hypothesized that the difference area between them comprised cells that indicated positive samples. These cells generally had larger areas, more irregular shapes, or more LDs. The Spearman's rank correlation tests indicated positive Spearman's rank correlation coefficients (SCCs) larger than 0.5 between the first ISOMAP components and three cell features, area, shape factor, and LD dispersity (**Fig. 3d**). These results suggest that the cell vectors obtained by our model were meaningful representations that correlated well with explicit cell features and may facilitate the following MIL for sample subtyping.

Multi-instance Learning Using Cell Vectors for Accurate PNM Assessment

We first analyzed the relationship between the degree of malignancy and the cell vector value with a representative vector component, COMP#475. On the cell level, cells with larger COMP#475 values appeared more obvious abovementioned features that may correlate with malignancies (**Fig. 4a**). On the patient level, except POS#7 and POS#8, the maximal COMP#475 values among all cells were larger in the M(Positive) samples than in the M(Negative) ones (**Fig. 4b**), which was also reflected by the values of other vector components similarly (**Fig. S5**). In comparison, such a difference could hardly be reflected by the mean values of cell features (**Fig. S6**). These results reflect the basis for MIL with cell vectors.

Then, we developed a logistic regression (LR) model named MI-LR for MIL. As conventional LR does, it assigned a coefficient for each vector component to quantify its weight in producing a positive prediction. Inspired by MI-SVM[29], our MI-LR model can predict on both the cell and patient levels. During training, the cell-level pseudo labels assigned by the model were iteratively updated until the model converged. Finally, according to whether a cell indicated a positive sample or not, the obtained MI-LR model classified the cells into two groups named significant and insignificant cells, respectively, with a threshold of 0.5. Examples of significant cells in M(Positive) samples are shown in **Fig. S7**. Examples of cells and their positive probabilities are shown in **Fig. S8**, suggesting that the positive probability was not predominantly determined by any one of cell features such as area or LD number, but by a combination of them. The details of model training are described in the *Methods* section. The patient-level positive probability was defined as the maximal cell-level positive probability in the sample, achieving 80% sensitivity, 94.1% specificity, 88.9% accuracy and an AUC of 0.86 on 27 patients with a threshold of 0.5 (**Fig. 4c**).

Moreover, we extracted the feature distributions of significant and insignificant cells (**Fig. 4d**) and compared them with the distributions shown in **Fig. 2d**. The overlaps between the feature distributions of significant and insignificant cells were substantially smaller than the overlaps between sample cells without division by the MIL model. This finding reflects the significance of our method when an unknown, relatively small proportion of cells were associated with malignancy. A comparison of other extracted cell features before and after division by the MIL model is shown in **Fig. S9**.

Interpretation of Feature Contributions in PNM Assessment by MICLEAR

To demonstrate the significance of introducing the componential information revealed by SRS microscopy and invisible to conventional cytology, we analyzed its contributions to the margin assessment by comparing the performance of MIL without morphological features and MIL without componential features. We also compared the performance of MIL with all extracted features and MIL with cell vectors outputted by the embedding model.

The algorithm achieved a higher AUC when combining two categories of features than using either of them (0.86 with all features vs 0.74 with morphological ones and 0.85 with componential ones) (**Fig. 4e**). The sensitivity increased substantially from 30% to 80% when introducing the componential features. The specificity and the accuracy increased subtly (only one sample had different prediction) from 94.1% to 100% and from 85.2% to 88.9%, respectively, when introducing the morphological features (**Fig. 4f**). In addition, the visualization through ISOMAP reflected that the distributions of cell vectors and all extracted features were more dispersed than the

distribution of either of two feature categories (**Fig. 4g**), which may help to identify significant cells and thus positive margin samples. Despite the similar performance of MIL with all extracted features and with cell vectors, the latter holds a significant advantage without manual feature engineering.

Moreover, we analyzed the contributions of each vector component, which was reflected by the coefficients assigned by the MIL model. The MIL coefficients changed subtly among different folds of the LOOCV and thus can be represented with their mean values in all folds (**Fig. 4h**). We sorted the vector components according to their MIL coefficients in ascending order. Components after COMP#245 had MIL coefficients larger than 0, indicating their positive correlation with cell-level significance. Generally, these components also correlated better with explicit cell features and had higher coefficients of variation (CVs), a nondimensionalized version of standard deviation (**Fig. 4i**). For example, COMP#462 had SCCs with area and LD dispersity larger than 0.7, and SCCs with shape factor and LD number larger than 0.3; COMP#478 had a CV larger than 3, SCCs with area and LD dispersity larger than 0.7, and a SCC with LD number larger than 0.3. These results suggest that the MIL model assigned the most weight to the vector components that correlated well with cell features and had high variation among cells. The former indicates the single-cell interpretability of the MIL model, and the latter reflects that the outcome of the CL-based model, i.e. to produce cell-level representations that distinguish cells well, is well exploited by the MIL model. All cell vector components' MIL coefficients, CVs, and SCCs with cell features are listed in **Table. S1**.

Comparison with Other PNM assessment Algorithms Using Cell Vectors

We also used two categories of algorithms without MIL to compare their performance with the MI-LR. One category utilized a conventional LR model and patient-level statistics derived from the cell vectors (**Fig. S10a**). Therefore, they cannot identify significant cells as our MIL model did. When using the mean values of cell vectors, the sensitivity declined sharply from 80% to 50% because the characteristics related to malignancy were seriously diluted. When using the maximum of all embedding components, the specificity declined from 94.1% to 82.4% because this was equivalent to the situation that there was a cell incorporating the malignancy-related characteristics of all cells in the sample, which may cause false positives. The additional introduction of the minimal values was not helpful. In summary, these methods have inherent limitations in the PNM assessment due to the inability to focus on the subset of sample cells crucial for detecting malignancy.

The other category calculated patient-level statistics only within a cluster of cell vectors, which was produced by dimensionality reduction (principle component analysis (PCA) or ISOMAP) and then K-Means clustering. The selected cluster was similar to the difference area shown in **Fig. 3b**. This was similar to the K-PCA in our previous study[26], but different in that the dimensionality reduction and clustering were performed on all N and T samples and then applied to each M sample instead of directly on each M sample because the exfoliated cells collected from a PNM sample were much less. The methods used here were named K-N-PCA or K-N-ISOMAP for distinguishment (**Fig. S10b**). The performance of K-N-ISOMAP was better than K-N-PCA (sensitivity of 80% vs 60%, specificity of 88.2% vs 82.4%, and accuracy of 81.5% vs 77.8%) and similar to MI-LR (specificity of 88.2% vs 94.1% and accuracy of 81.5% vs 88.9%). Note that all these methods incorporated the prior knowledge that the significant cells were centered in a specific cluster. In comparison, MI-LR did not require prior knowledge to identify significant cells and thus is more generalizable.

Moreover, we modified the classification model and the model training scheme of MI-LR and compared their performance (**Fig. S11**). MI-models and mi-models[29] are similar methods requiring iterative training to make the model converge. The difference is that mi-models took all cells in positive samples for the next training iteration. In contrast, MI-models only took the cell with the highest positive probability in each positive sample. Therefore, MI-models conformed more to the knowledge that cells whose positive probabilities are not the highest almost make no difference to the patient-level prediction, while mi-SVM may regard too many cells as significant and thus cause more false positives. Replacing mi-models with MI-models increased the specificity from 64.7% to 100% for the support vector machine (SVM) and 70.6% to 94.1% for the LR. For the MI-models, replacing the SVM with the LR slightly reduced the specificity from 100% to 94.1% but significantly increased the sensitivity from 50% to 80%.

DISCUSSION

In this study, we have developed the MICLEAR, a methodology to apply label-free SRS microscopy-based cytology for PNM assessment. Two-color SRS microscopy was utilized to obtain single-cell morphological and componential information. Our cell embedding method based on CL condensed the spatially-resolved Raman profile of cells into low-dimensional vectors that distinguished cells as much as possible. Then, a MIL model, MI-LR, was used to predict the patient-level positive probability with the cell vectors belonging to the sample. For the samples of exfoliated margin cells collected from 27 patients, MICLEAR achieved 80% sensitivity, 94.1% specificity, 88.9% accuracy and AUC of 0.86 as evaluated by LOOCV. The procedure can be finished within approximately 25 min for each patient.

The cell embedding model minimizes the prior knowledge necessary for diagnosis. Conventional machine learning involves manual feature engineering that designs practical features. The task-specific prior knowledge required for this process is costly to accumulate through many statistical analyses. Besides, designing robust algorithms that work well in an extensive dataset resistant to data variability is challenging. In comparison, by incorporating customized architectures and loss functions, neural networks can automatically compress high-dimensional image data into proper low-dimensional feature vectors. CL framework is an efficient method to embed data while adjusting the similarities between the outputted vectors. It has been widely applied in medical data to facilitate downstream diagnosis tasks and multimodal fusion. For example, weakly-supervised CL can adjust the similarities between representations of image patches according to the patient-level labels[53], and self-supervised CL does not require exogenous labels[54,55]. MICLEAR utilized self-supervised SimCLR for cell embedding in consideration of high intercellular heterogeneity occurring in the positive samples of exfoliated cells, which may introduce huge label noise when using the patient-level labels directly.

Our scheme of MIL using cell vectors maintains the single-cell interpretability. Current artificial intelligence-assisted digital histology[41] and cytology[60] using MIL typically comprises image patch embedding, data aggregation and prediction, and the diagnosis result is then interpreted with the attention mechanism[42] or the saliency map[43]. These methods guide the attribution of model predictions to particular regions on the image patches. However, such patch-based MIL cannot provide a clear interpretation on the single-cell level. In this case, the analysis of how the prediction is associated with the cell features has to be done within the so-called focal regions of the model, which is indirect and may be inaccurate because the regions with high attention are not necessarily of diagnostic significance. In comparison, MICLEAR directly recognizes cells that indicate positive samples and allows comparison between their features and others'. As shown in **Fig. 4d** and **Fig. S10**, in addition to larger size and nucleocytoplasmic ratio in morphology, the altered lipid metabolism of significant cells was revealed, including increasing and more dispersive LDs and higher lipid intensity within the cytoplasm.

We investigated two false negatives (POS#7, POS#8) and one false positive (NEG#2) in the PNM assessment. Two false negatives had vector components similar to those of M(Negative) samples, as shown in **Fig. S5**. Cells belonging to them appeared dissimilar from the significant cells in other samples and had a maximal cell-level positive probability of only 0.1% and 14.5%, respectively (**Fig. S12**). These false negatives may be caused by cancer cell loss in sample processing or failure to find them under the microscope. For POS#7, the patient was diagnosed as negative by intraoperative FS pathology but positive by postoperative PS pathology, so our method had the same finding as FS pathology. The false positive was produced by a cell with a positive

probability of 53.4%, which was slightly higher than the threshold of 50% (**Fig. S13**). Incorporating more samples whose features are near the MIL model's decision boundary may help alleviate this problem. Notably, for POS#10, which was diagnosed as negative by intraoperative FS pathology but positive by postoperative PS pathology like POS#7, the diagnosis provided by our method matched the PS pathology with a positive probability near 100%. Because the cases where FS and PS pathology mismatched were rare in this study, more samples are required to evaluate how often our method will match the PS pathology when the FS pathology gives false negative findings.

Our method may have the potential to overcome the limitations of FS-based intraoperative PNM assessment in sampling and detection. Limited sampling region is a significant cause of false negatives of FS pathology[8,10]. In comparison, the wider coverage of brush cytology may alleviate this problem. The efficiency of sampling and sample processing can be further improved by refined sampling instruments and devices for high-throughput SRS microscopy, such as Raman imaging flow cytometry and microfluidic chips. Besides, we significantly improved the diagnosis accuracy by incorporating componential information, which conventional cytology cannot exploit. Similar improvements were also shown in fine-needle aspiration (FNA) biopsy combined with mass spectrometry[57,58], which can also reflect the componential information but is time-consuming, requires frequent calibration and complicated data processing. The comparison between PNM assessment with different feature sets reflects the predominant significance of componential information. In comparison, its relative contribution was relatively smaller in the SRS microscopy-based peritoneal metastasis detection as we studied before[26]. Such difference may be attributed to a more distinct abnormality in the lipid metabolism of PDAC cells[27].

The scheme of embedding the spatially-resolved Raman profile facilitates incorporating multimodal information to improve the accuracy further. The imaging channels can be extended from two-color to hyperspectral SRS microscopy to explore more significant markers for disease diagnosis. The SRS microscopy may be replaced with similar modalities that reflect chemical components, such as all-fiber-based coherent anti-Stokes Raman scattering (CARS) microscopy or spontaneous Raman spectroscopy. Besides, techniques derived from Raman spectroscopy can serve as an omics approach to combine with other single-cell omics[56].

In summary, we propose a less operationally challenging and more objective approach compared to existing intraoperative pancreatic margin assessment. It holds great potential for development to generalizable cytology with single-cell interpretability and may find application in clinical scenarios such as FNA biopsy, ascitic fluid cytology and evaluation of superior mesenteric artery (SMA) margin status of PDAC.

METHODS

Exfoliative Cell Sample Collection

Exfoliative cell samples were collected using the medical cytology brush from 41 patients (7 N samples, 28 T samples, 10 M(Pos.) samples, and 17 M(Neg.) samples) in the Peking University Cancer Hospital. Normal pancreatic cell samples were collected from seven patients without pancreatic diseases. PDAC tumor and PNM cell samples were collected from the other 34 patients diagnosed with PDAC. The sample collection locations of each patient are described in **Table. S2**. Each sample contained ~1 mL cell suspension. The inclusion criteria were: diagnosis as PDAC by the pathology both before and after operation; no distant metastasis discovered before or during operation; not suffering from multiple primary tumors; no anti-tumor therapy before operation. This study was approved by the Institutional Review Boards of Peking University Cancer Hospital and Beihang University. Written informed consent was obtained from all patients before the study was conducted.

PNM assessment with FS and PS

FSs were used for PNM assessment during the resection. The tissue around the PNM could be re-resected repeatedly until the resected tissue was assessed as negative by FS. Exfoliative cells were collected from the first PNM without any re-resection. The tissue was collected from the last PNM when no re-resection was required and then postoperatively assessed with PS, which serves as the gold standard above the intraoperative FS. The PNM status was determined using the 1 mm rule that R0 was defined as absence of cancer cells within 1 mm of the PNM. The resection can be classified into several cases according to the first FS finding and the PS finding. Their corresponding ground truth and labels for model training were discussed separately, as shown in **Table. S3**. Note that when re-resection was performed, the PNM cell samples for our study corresponded with the first FS finding instead of the PS finding due to their different sampling locations. Moreover, for the assessment of the final PNM right before all resection was over, most FS findings were consistent with the PS findings. Therefore, we chose the FS findings as the labels for model training, but used the PS finding for model evaluation, i.e. the ground truth, to analyze how our model predicted in the LOOCV when the FS and the PS mismatched. The information of each margin samples is shown in **Table. S4**.

Processing of PNM Exfoliative Cell Samples Before SRS Microscopy

The cell samples were centrifuged at 1200 rpm for 4 min. Then, the cell sediments were treated by red blood cell lysate for 5 min at the temperature of 4 °C and centrifuged again at 1200 rpm for 4 min. Finally, the cell sediments (~3 μL) were dropped on the glass slide and evenly smeared for the following SRS microscopy.

Two-color SRS Microscopy

The setup of our homebuilt SRS microscope is shown in **Fig. 2a**. It employed a dual-output picosecond pulse laser (picoEmeraldTM S, Applied Physics and Electronics) with a repetition rate at 80 MHz and a pulse duration of 2 ps. The outputted beams included the pump beam with tunable wavelength from 700 nm to 960 nm and the Stokes beam with fixed wavelength at 1031 nm. The Stokes beam was modulated at ~20 MHz by an electronic optic modulator. Two beams were overlapped in space and time, coupled to a 2-D scanning galvanometer

(GVS012-2D, Thorlabs), and then imported into an inverted microscope (IX73, Olympus). The lasers were focused by a 60× water immersion objective lens (LUMPlanFL N, 1.0 numerical aperture, Olympus) onto the sample and then collected by a 60× water immersion condenser (LUMPlanFL N, 1.0 numerical aperture, Olympus). The pump beam with stimulated Raman loss (SRL) was selected by a short-pass filter (ET980SP, Chroma) and then detected by a photodiode (S3994-01, Hamamatsu, Japan) that was equipped with a resonant circuit selectively amplifying the signal at the optical modulation frequency. The stimulated Raman loss signal was then extracted by a digital lock-in amplifier (HF2LI, Zurich Instrument, Zurich, Switzerland). The output voltage, i.e. the SRS signal, was sampled by a data acquisition card (PCI- 6363, National Instruments, Austin, TX). A LabVIEW program synchronized the 2-D scanning by galvanometer and acquisition of SRS signals. The two-color SRS microscopy was performed by tuning the pump wavelength to 796.8 nm and 791.8 nm to match the Raman bands of 2850 cm^{-1} and 2930 cm^{-1}, which correspond to the C-H stretching regions of CH_2 (rich in lipid) and CH_3 (rich in protein), respectively. The output power was ~50 mW and ~300 mW for the pump beam and the Stokes beam, respectively. Each SRS image contained 400 × 400 pixels and its FOV was 100 × 100 μm. The pixel dwell time was 10 μs.

Single-cell Segmentation and Quantitative Evaluation

The single-cell segmentation used the python library cellpose 2.2.2. The Cellpose model requires the information of cell diameter to produce segmentation masks. Cells smaller than the preset value may be neglected, while cells larger than it may be over-segmented. Because the heterogeneity of cell size can be very high in positive samples, single preset cell diameter was not proper for accurate segmentation. We used a series of preset values to recognize cells with different sizes. The resulting prediction results were integrated to fuse over-segmented cell parts together. After automated single-cell segmentation, inaccurate results were manually rectified. In comparison, single-cell segmentation with watershed and Stardist model was performed by Fiji (after Otsu threshold to create binary masks) and python library stardist 0.8.3, respectively.

The segmentation results were quantified with three metrics, global Dice coefficient, cell number ratio, and cellwise Dice coefficient. The ground truth was produced by visual judgement. Global Dice coefficient quantifies the overall overlap degree between the prediction and the ground truth, as defined by

$$Global\ Dice\ coefficient = \frac{2|Pred \cap GT|}{|Pred| + |GT|} \tag{1}$$

where |·| denotes the overall area of cell regions.

Cell number ratio quantifies the degree of under- or over-segmentation, as defined by

$$Cell\ number\ ratio = \frac{N_{pred}}{N_{GT}} \tag{2}$$

where N_{pred} and N_{GT} denote the cell number in the prediction and the ground truth, respectively.

Cellwise Dice coefficient quantifies the mean overlap degree between each segmented cell and the corresponding cell in the ground truth, as defined by

$$Cellwise\ Dice\ coefficient = \sum_i \frac{2|C_{Pred,i} \cap C_{GT,i}|}{|C_{Pred,i}| + |C_{GT,i}|} / N_{pred} \tag{3}$$

where $C_{pred,i}$ and $C_{GT,i}$ denote the i th cell in the segmentation map and its corresponding cell in the ground truth, respectively. The corresponding cell in the ground truth was the cell with the largest intersection over union (IoU) with the segmented cell. This metric reflects the single-cell segmentation accuracy better than global Dice coefficient, which is insensitive to under- or over-segmentation.

Single-cell Feature Extraction

We extracted 12 features for each cell, including five morphological features (area, perimeter, shape factor, cytoplasm area, and cytoplasm area fraction) and seven componential features (lipid intensity, lipid intensity within cytoplasm, protein intensity, protein intensity within cytoplasm, lipid-protein intensity ratio, LD number, and LD dispersity). The morphological features involved cell shapes and structures. The componential features involved distribution of lipid and protein as calculated before.

For the two-color SRS images of each FOV, the molecular image, i.e. the concentration of lipid and protein at each pixel, was calculated by

$$\begin{bmatrix} c_{lipid} \\ c_{protein} \end{bmatrix} = \begin{bmatrix} k_{lipid,2850} & k_{protein,2850} \\ k_{lipid,2930} & k_{protein,2930} \end{bmatrix}^{-1} \begin{bmatrix} S_{2850} \\ S_{2930} \end{bmatrix} \tag{4}$$

where k denotes a component's contribution to the SRS signal at a certain Raman shift and is obtained from the Raman spectra of pure chemical samples; S denotes the pixel intensity of SRS images at a certain Raman shift; c denotes the calculated component concentration. This is based on the SRS property that the SRL is linearly correlated with the concentration of Raman-active vibrational modes. The single-cell composite images including lipid channel (green) and protein channel (blue) were used for extraction of componential features and later training of the cell embedding model. Each image had 400×400 pixels and the segmented cell was located in the center.

Area, centroid, and perimeter were directly extracted from the segmentation map with the regionprops function in the python library scikit-image. Cells on the edge of each FOV were excluded from feature extraction. Then, the shape factor was calculated by

$$Shape\ factor = 1 - 4\pi \frac{Area}{Perimeter^2} \tag{5}$$

The shape factor of a circle is 0. The value is larger for cells with slender shape or concave contour, whose perimeter is larger when the area is fixed.

For other features, the feature extraction was performed in each cell region provided by the segmentation mask. Mean lipid intensity, mean protein intensity and their ratio were calculated with the molecular image. LDs were detected with the blob_log function in the python library scikit-image, and then LD number and LD dispersity (defined as the mean distance from all LDs of the cell to its centroid) were obtained. Nucleus could be detected by applying an Otsu threshold in the lipid channel of each cell because the lipid concentration in nucleus was much lower than that in cytoplasm. Then, cytoplasm area, cytoplasm area fraction, and mean lipid and protein intensity within cytoplasm were extracted.

Contrastively learned Cell Embedding

The cell embedding model was trained with the contrastive framework SimCLR. CL typically learns the low-dimensional representation of high-dimensional data (e.g., images) by making similar data attract each other and different kinds of data repel each other. SimCLR produces similar images by applying certain transformations that do not change their essence. For example, proper color distortion does not change the animal type in an image. The transformed results of the same image are regarded as similar ones, while the transformed results of different images are regarded as different ones. Therefore, this contrastive framework does not need any label. The contrastive loss is calculated by

$$l(i,j) = -log\frac{\exp(s_{i,j}/\tau)}{\sum_{k=1}^{2N} 1_{[k\neq i]}\exp(s_{i,k}/\tau)} \tag{6}$$

$$L = \frac{1}{2N}\sum_{k=1}^{N}[l(2k-1,2k) + l(2k,2k-1)] \tag{7}$$

where $s_{i,j}$ denotes the cosine similarity between the projections z_i and z_j; the projection is acquired by passing a transformed result through two networks, f and g, sequentially; the projections z_{2k-1} and z_{2k} come from the transformed results of the kth image in the minibatch containing N images (N=8 for our model); τ denotes temperature hyperparameter and is 0.07 for our model.

The image dataset containing all single-cell molecular images belonging to the N and T samples was divided into training and validation set randomly with a proportion of 4:1. Images belonging to the M samples were not used to make the model embed cells without being seen. The sequence of image transformations included random rotation, random horizontal flipping, and random vertical flipping. Each transformation was performed with a probability of 50%. The model was optimized with AdamW and trained for 200 epochs with a NVIDIA GeForce RTX 3070 Laptop GPU, which required ~2 hrs. The checkpoint with the smallest contrastive loss on the validation set was chosen. After the model containing f and g was trained, the network g was removed for better information preservation[28]. Finally, the cell vector was outputted by the network f and used for the following MIL.

Multi-instance Learning Model

The MIL was performed with a LR model for single-cell classification. Its C parameter is 1.0 and the solver is L-BFGS. Due to the finding that the negative cells in the positive samples had similar cells in the negative samples, we first assigned pseudo-labels for all cells according to their neighbors in the ISOMAP distribution. A cell in a negative sample was assigned a negative pseudo-label. For a cell in a positive sample, in its ten nearest neighbors, if there were less than nine belonging to a positive sample, the cell was assigned a negative pseudo-label because its characteristics were considered to be not unique to positive cells. Otherwise, it was assigned a positive pseudo-label. The LR model was initialized by training with these pseudo-labels. Then, the cells used for next training iteration included cells in the negative samples (negative pseudo-labels were reassigned to them) and cells with the largest positive probability in each positive sample (positive pseudo-labels were reassigned to them). The LR model was updated after being trained with them. This iteration was repeated until cells that were in the positive samples and used for training did not change, and then the model training was stopped. The MI-SVM was trained similarly but the model was replaced with a SVM classifier with C parameter of 1.0 and a linear kernel. The mi-LR and mi-SVM were different in that they did not exclude cells without the largest positive probability in the positive samples so all cells were used in the training iterations.

Statistical Analysis

The violin plots represent medians and two quartiles, and the box plots represent medians (center lines), two quartiles (box edges), 2.5 and 97.5 percentiles (top and bottom lines outside the box), and minimum and maximum (outliers). They were both drawn with GraphPad Prism 9. The unpaired comparisons between cell features in different groups were performed with Mann-Whitney U test, and the paired comparisons between cell segmentation metrics with different methods were performed with Wilcoxon matched-pairs signed rank test. Two-tailed p-values of all comparisons were calculated in GraphPad Prism 9. p-value < 0.05 was considered statistically significant. In the comparisons shown in violin plots, outliers were removed to focus on the major bodies of distributions. Outliers were identified using the ROUT method with Q=0.5%. The removed outliers of **Fig. 2d** were shown in **Fig. S4**. Especially in T and M(Pos.) samples, they may be crucial in producing a positive prediction but diluted by a huge proportion of cells similar to ones in the N and M(Neg.) samples. The ROC curve and other metrics were calculated by assembling 27 PNM assessment results in the LOOCV.

REFERENCES

[1] Siegel R L, Miller K D, Wagle N S, et al. Cancer statistics, 2023[J]. CA: a cancer journal for clinicians, 2023, 73(1): 17-48.

[2] Adamska A, Domenichini A, Falasca M. Pancreatic ductal adenocarcinoma: current and evolving therapies[J]. International journal of molecular sciences, 2017, 18(7): 1338.

[3] Mizrahi J D, Surana R, Valle J W, et al. Pancreatic cancer[J]. The Lancet, 2020, 395(10242): 2008-2020.

[4] Holm M B, Verbeke C S. Prognostic impact of resection margin status on distal pancreatectomy for ductal adenocarcinoma[J]. Current Oncology, 2022, 29(9): 6551-6563.

[5] Crippa S, Ricci C, Guarneri G, et al. Improved survival after pancreatic re-resection of positive neck margin in pancreatic cancer patients. A systematic review and network meta-analysis[J]. European Journal of Surgical Oncology, 2021, 47(6): 1258-1266.

[6] Ethun C G, Kooby D A. The importance of surgical margins in pancreatic cancer[J]. Journal of surgical oncology, 2016, 113(3): 283-288.

[7] Novis D A, Zarbo R J. Interinstitutional comparison of frozen section turnaround time[J]. Archives of pathology & laboratory medicine, 1997, 121(6): 559.

[8] Cioc A M, Ellison E C, Proca D M, et al. Frozen section diagnosis of pancreatic lesions[J]. Archives of pathology & laboratory medicine, 2002, 126(10): 1169-1173.

[9] Witz M, Shkolnik Z, Dinbar A. Intraoperative pancreatic biopsy—a diagnostic dilemma[J]. Journal of surgical oncology, 1989, 42(2): 117-119.

[10] Campanale R P, Frey C F, Farias L R, et al. Reliability and sensitivity of frozen-section pancreatic biopsy[J]. Archives of Surgery, 1985, 120(3): 283-288.

[11] Harris P L, Rumley T O, Lineaweaver W C, et al. Pancreatic cancer: unreliability of frozen section in diagnosis[J]. Southern Medical Journal, 1985, 78(9): 1053-1056.

[12] Mohamed A, Hassan M M, Zhong W, et al. A Quantitative and qualitative assessment of frozen section diagnosis accuracy and deferral rate across organ systems[J]. American Journal of Clinical Pathology, 2022, 158(6): 692-701.

[13] Theisen B K, DiCianno R, Singhi A D. Pancreatic frozen section nightmares[J]. Diagnostic Histopathology, 2016, 22(6): 236-242.

[14] Bellizzi A M, Frankel W L. Pancreatic pathology: a practical review[J]. Laboratory medicine, 2009, 40(7): 417-426.

[15] Tummers W S, Miller S E, Teraphongphom N T, et al. Intraoperative pancreatic cancer detection using tumor-specific multimodality molecular imaging[J]. Annals of surgical oncology, 2018, 25: 1880-1888.

[16] Li R, Wang P, Lan L, et al. Assessing breast tumor margin by multispectral photoacoustic tomography[J]. Biomedical optics express, 2015, 6(4): 1273-1281.

[17] van Manen L, Stegehuis P L, Farina-Sarasqueta A, et al. Validation of full-field optical coherence tomography in distinguishing malignant and benign tissue in resected pancreatic cancer specimens[J]. PLoS One, 2017, 12(4): e0175862.

[18] Ji M, Orringer D A, Freudiger C W, et al. Rapid, label-free detection of brain tumors with stimulated Raman scattering microscopy[J]. Science translational medicine, 2013, 5(201): 201ra119-201ra119.

[19] Ji M, Lewis S, Camelo-Piragua S, et al. Detection of human brain tumor infiltration with quantitative stimulated Raman scattering microscopy[J]. Science translational medicine, 2015, 7(309): 309ra163-309ra163.

[20] Nanda K, McCrory D C, Myers E R, et al. Accuracy of the Papanicolaou test in screening for and follow-up

of cervical cytologic abnormalities: a systematic review[J]. Annals of internal medicine, 2000, 132(10): 810-819.
[21] Brimo F, Vollmer R T, Case B, et al. Accuracy of urine cytology and the significance of an atypical category[J]. American journal of clinical pathology, 2009, 132(5): 785-793.
[22] Domagala-Kulawik J. The relevance of bronchoalveolar lavage fluid analysis for lung cancer patients[J]. Expert Review of Respiratory Medicine, 2020, 14(3): 329-337.
[23] Kassirian S, Hinton S N, Cuninghame S, et al. Diagnostic sensitivity of pleural fluid cytology in malignant pleural effusions: Systematic review and meta-analysis[J]. Thorax, 2023, 78(1): 32-40.
[24] Zgajnar J, Frkovic-Grazio S, Besic N, et al. Low sensitivity of the touch imprint cytology of the sentinel lymph node in breast cancer patients—results of a large series[J]. Journal of surgical oncology, 2004, 85(2): 82-86.
[25] Arnold D T, De Fonseka D, Perry S, et al. Investigating unilateral pleural effusions: the role of cytology[J]. European Respiratory Journal, 2018, 52(5).
[26] Chen X, Wu Z, He Y, et al. Accurate and Rapid Detection of Peritoneal Metastasis from Gastric Cancer by AI-Assisted Stimulated Raman Molecular Cytology[J]. Advanced Science, 2023: 2300961.
[27] Li J, Gu D, Lee S S Y, et al. Abrogating cholesterol esterification suppresses growth and metastasis of pancreatic cancer[J]. Oncogene, 2016, 35(50): 6378-6388.
[28] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//International conference on machine learning. PMLR, 2020: 1597-1607.
[29] Andrews S, Tsochantaridis I, Hofmann T. Support vector machines for multiple-instance learning[J]. Advances in neural information processing systems, 2002, 15.
[30] Stringer C, Wang T, Michaelos M, et al. Cellpose: a generalist algorithm for cellular segmentation[J]. Nature methods, 2021, 18(1): 100-106.
[31] Schmidt U, Weigert M, Broaddus C, et al. Cell detection with star-convex polygons[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11. Springer International Publishing, 2018: 265-273.
[32] Ji M, Orringer D A, Freudiger C W, et al. Rapid, label-free detection of brain tumors with stimulated Raman scattering microscopy[J]. Science translational medicine, 2013, 5(201): 201ra119-201ra119.
[39] Tenenbaum J B, Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. science, 2000, 290(5500): 2319-2323.
[40] Wang Y, Yao H, Zhao S. Auto-encoder based dimensionality reduction[J]. Neurocomputing, 2016, 184: 232-242.
[41] Lipkova J, Chen R J, Chen B, et al. Artificial intelligence for multimodal data integration in oncology[J]. Cancer cell, 2022, 40(10): 1095-1110.
[42] Li G, Li C, Wu G, et al. Multi-view attention-guided multiple instance detection network for interpretable breast cancer histopathological image diagnosis[J]. IEEE Access, 2021, 9: 79671-79684.
[43] Mahapatra D, Poellinger A, Shao L, et al. Interpretability-driven sample selection using self supervised learning for disease classification and segmentation[J]. IEEE transactions on medical imaging, 2021, 40(10): 2548-2562.
[44] Verbeke C S. Resection margins in pancreatic cancer[J]. Surgical Clinics, 2013, 93(3): 647-662.
[45] Hernot S, van Manen L, Debie P, et al. Latest developments in molecular tracers for fluorescence image-guided cancer surgery[J]. The lancet oncology, 2019, 20(7): e354-e367.
[46] Tummers W S, Miller S E, Teraphongphom N T, et al. Detection of visually occult metastatic lymph nodes

using molecularly targeted fluorescent imaging during surgical resection of pancreatic cancer[J]. HPB, 2019, 21(7): 883-890.

[47] Piras D, Xia W, Steenbergen W, et al. Photoacoustic imaging of the breast using the twente photoacoustic mammoscope: present status and future perspectives[J]. IEEE Journal of Selected Topics in Quantum Electronics, 2009, 16(4): 730-739.

[48] Li R, Wang P, Lan L, et al. Assessing breast tumor margin by multispectral photoacoustic tomography[J]. Biomedical optics express, 2015, 6(4): 1273-1281.

[49] Haka A S, Volynskaya Z, Gardecki J A, et al. In vivo margin assessment during partial mastectomy breast surgery using Raman spectroscopy[J]. Cancer research, 2006, 66(6): 3317-3322.

[50] Feng X, Fox M C, Reichenberg J S, et al. Biophysical basis of skin cancer margin assessment using Raman spectroscopy[J]. Biomedical optics express, 2019, 10(1): 104-118.

[51] Traynor D, Behl I, O'dea D, et al. Raman spectral cytopathology for cancer diagnostic applications[J]. Nature Protocols, 2021, 16(7): 3716-3735.

[52] Volmar K E, Vollmer R T, Routbort M J, et al. Pancreatic and bile duct brushing cytology in 1000 cases: review of findings and comparison of preparation methods[J]. Cancer Cytopathology: Interdisciplinary International Journal of the American Cancer Society, 2006, 108(4): 231-238.

[53] Hollon T, Jiang C, Chowdury A, et al. Artificial-intelligence-based molecular classification of diffuse gliomas using rapid, label-free optical imaging[J]. Nature Medicine, 2023, 29(4): 828-832.

[54] Tavolara T E, Gurcan M N, Niazi M K K. Contrastive multiple instance learning: An unsupervised framework for learning slide-level representations of whole slide histopathology images without labels[J]. Cancers, 2022, 14(23): 5778.

[55] Yang M, Yang Y, Xie C, et al. Contrastive learning enables rapid mapping to multimodal single-cell atlas of multimillion scale[J]. Nature Machine Intelligence, 2022, 4(8): 696-709.

[56] Cutshaw G, Uthaman S, Hassan N, et al. The Emerging Role of Raman Spectroscopy as an Omics Approach for Metabolic Profiling and Biomarker Detection toward Precision Medicine[J]. Chemical Reviews, 2023.

[57] DeHoog R J, Zhang J, Alore E, et al. Preoperative metabolic classification of thyroid nodules using mass spectrometry imaging of fine-needle aspiration biopsies[J]. Proceedings of the National Academy of Sciences, 2019, 116(43): 21401-21408.

[58] Capitoli G, Piga I, L'Imperio V, et al. Cytomolecular classification of thyroid nodules using fine-needle washes aspiration biopsies[J]. International Journal of Molecular Sciences, 2022, 23(8): 4156.

[59] Campanale R P, Frey C F, Farias L R, et al. Reliability and sensitivity of frozen-section pancreatic biopsy[J]. Archives of Surgery, 1985, 120(3): 283-288.

[60] Teramoto A, Kiriyama Y, Tsukamoto T, et al. Weakly supervised learning for classification of lung cytological images using attention-based multiple instance learning[J]. Scientific reports, 2021, 11(1): 20317.

FIGURES

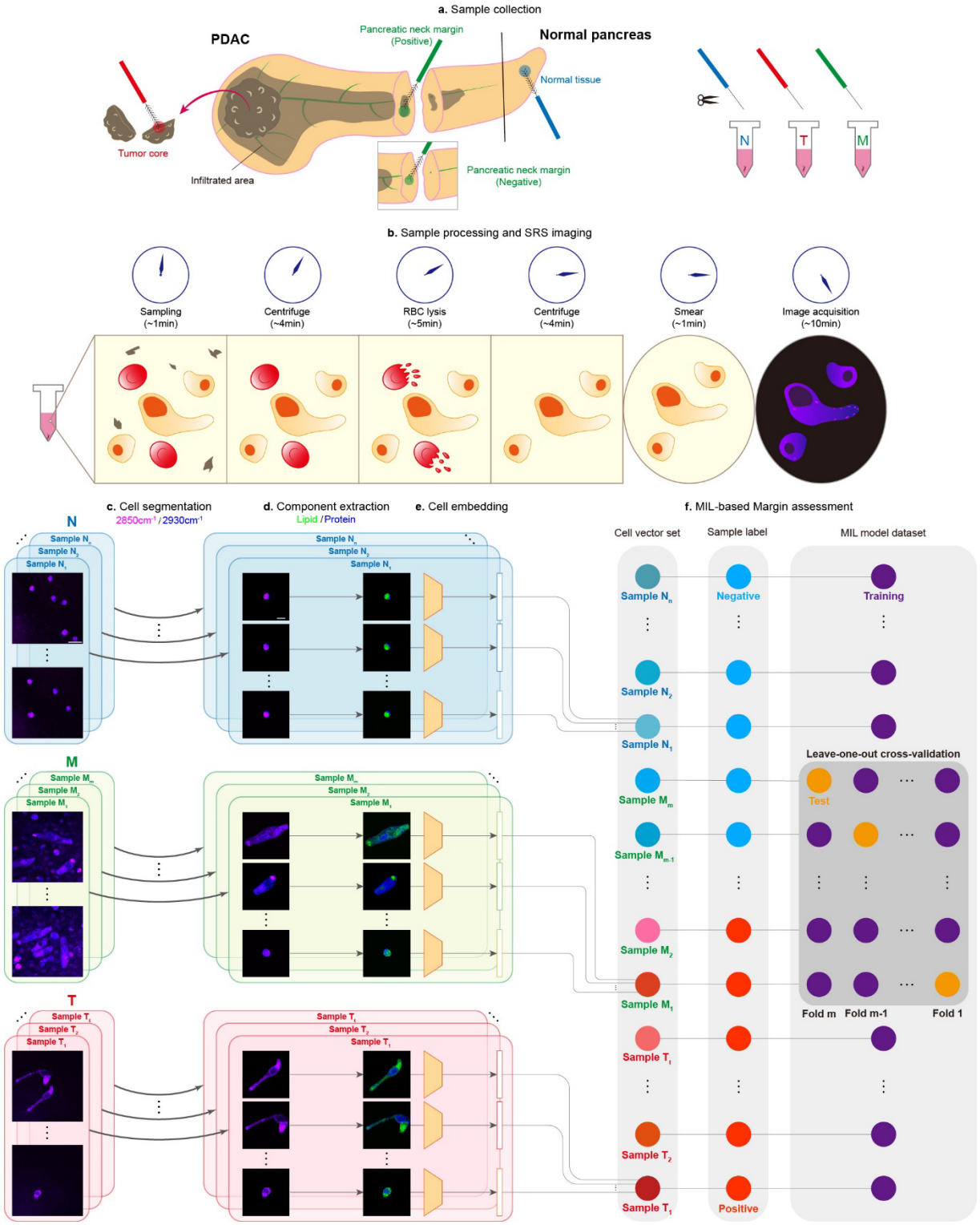

Figure 1 | Overview of the workflow. a, Sample collection at normal tissue, tumor core, and PNM. **b**, Processing and SRS imaging of cell samples. **c**, Single-cell segmentation based on the Cellpose model. **d**, Extraction of molecular images containing lipid and protein channels with matrix operation. **e**, Cell embedding with CL framework. **f**, Margin assessment with MIL model. Scale bars, 20 μm (**c**), 10 μm (**d**).

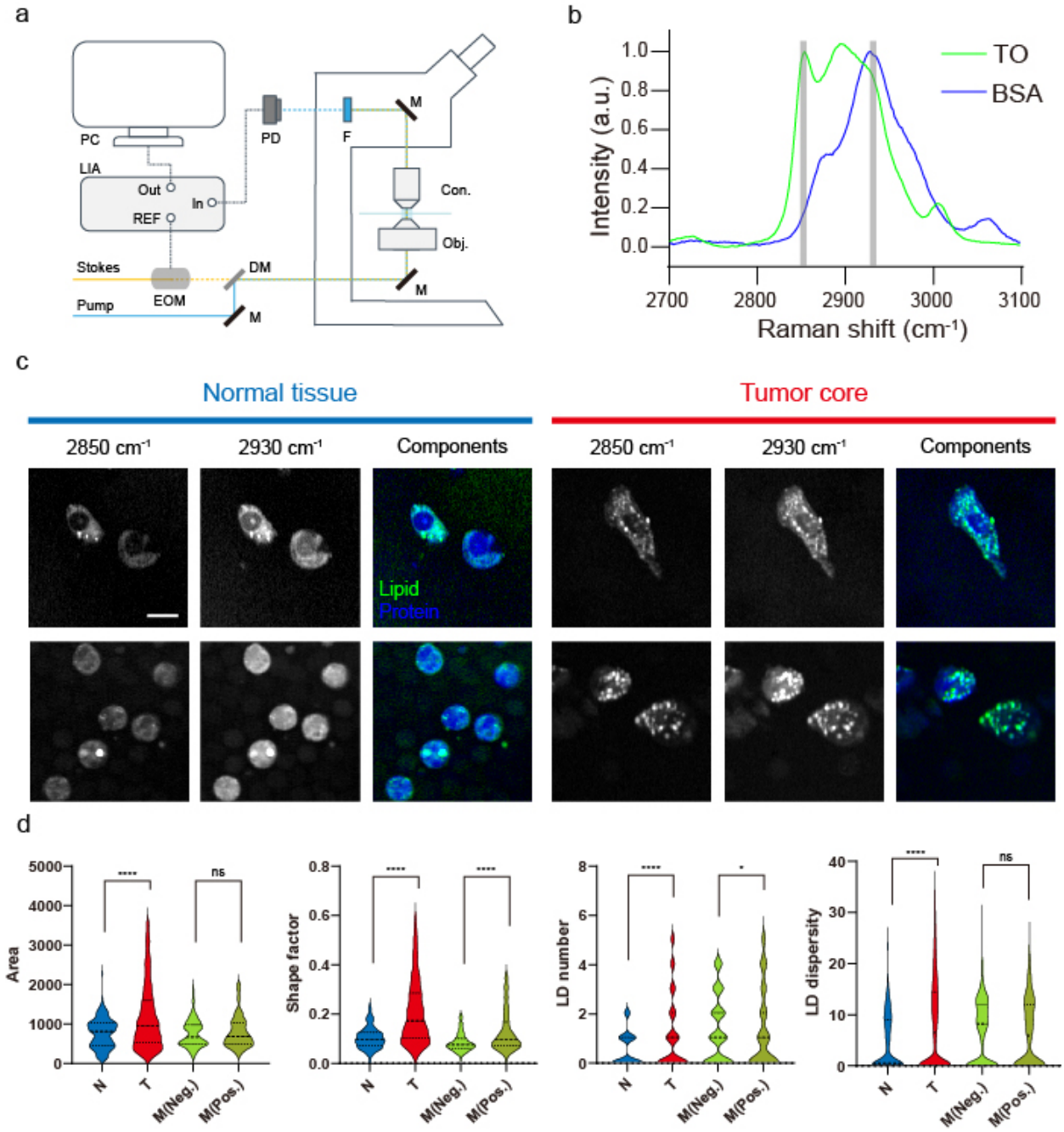

Figure 2 | SRS image acquisition and preprocessing. a, Setup of SRS imaging. **b**, Raman spectra of triolein (TO) and bovine serum albumin (BSA) with imaging channels 2850 cm^{-1} and 2930 cm^{-1} shown in gray columns. **c**, two-color SRS images and corresponding molecular images. **d**, Feature distribution in four sample groups, N, T, negative M, and positive M; outliers were removed as described in *Methods*. Scale bars, 10 μm. ns: no significant difference, *: $p<0.05$, ****: $p<0.0001$.

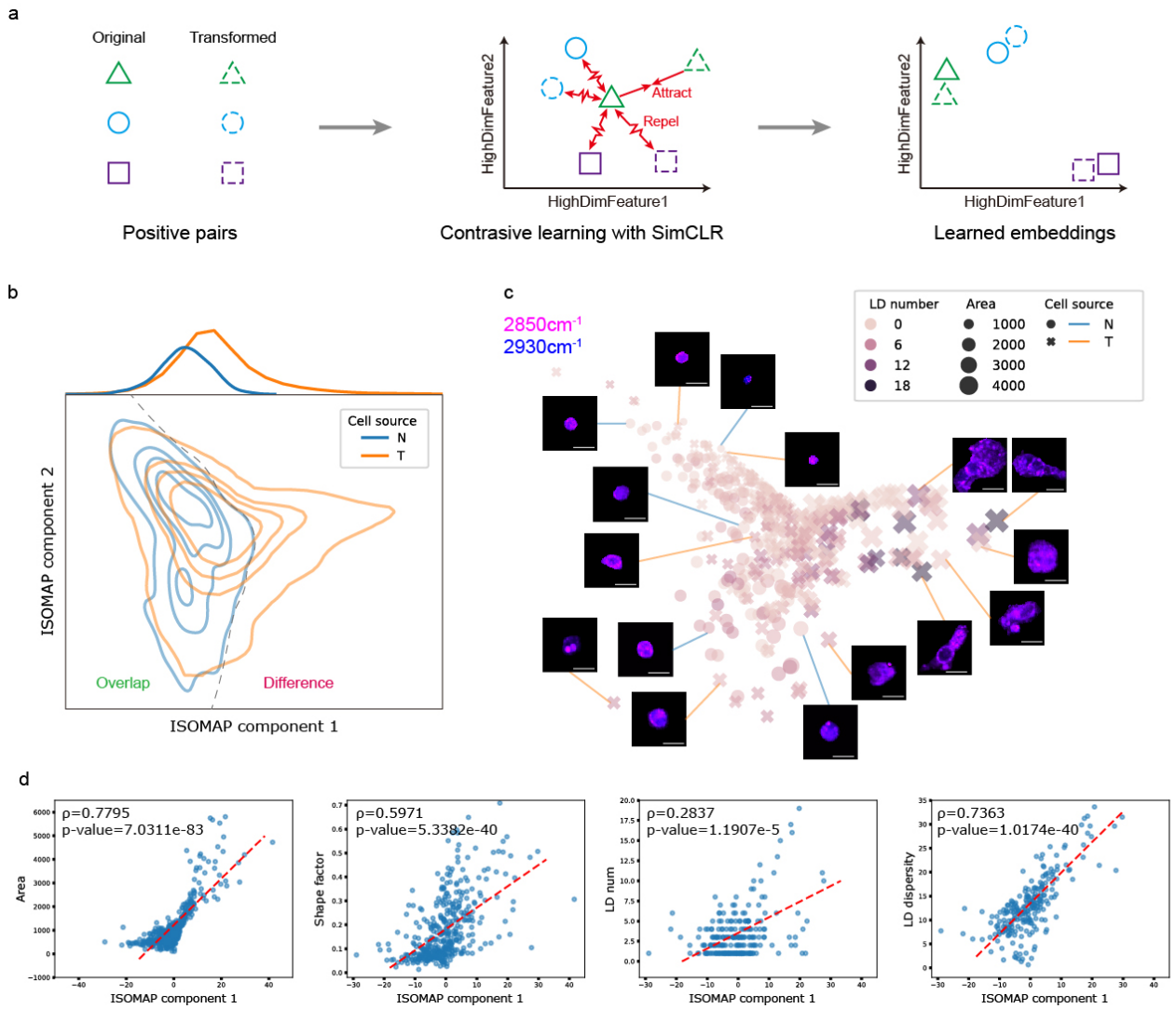

Figure 3 | CL-based model for cell embedding. a, CL based on the SimCLR. **b**, ISOMAP distribution of cell vectors in N and T samples. **c**, ISOMAP distribution in **b** associated with cell images and features. **d**, Correlation between cell features and the first component of ISOMAP, tested by Spearman's rank correlation test. Cells without LDs were excluded from the correlation analysis between ISOMAP component 1 and two LD-related features. Red dashed lines denote linear regression between ISOMAP component 1 and cell features. Cells are down-sampled by a factor of 10 for visual clarity in **c** and **d**. Scale bars, 10 μm.

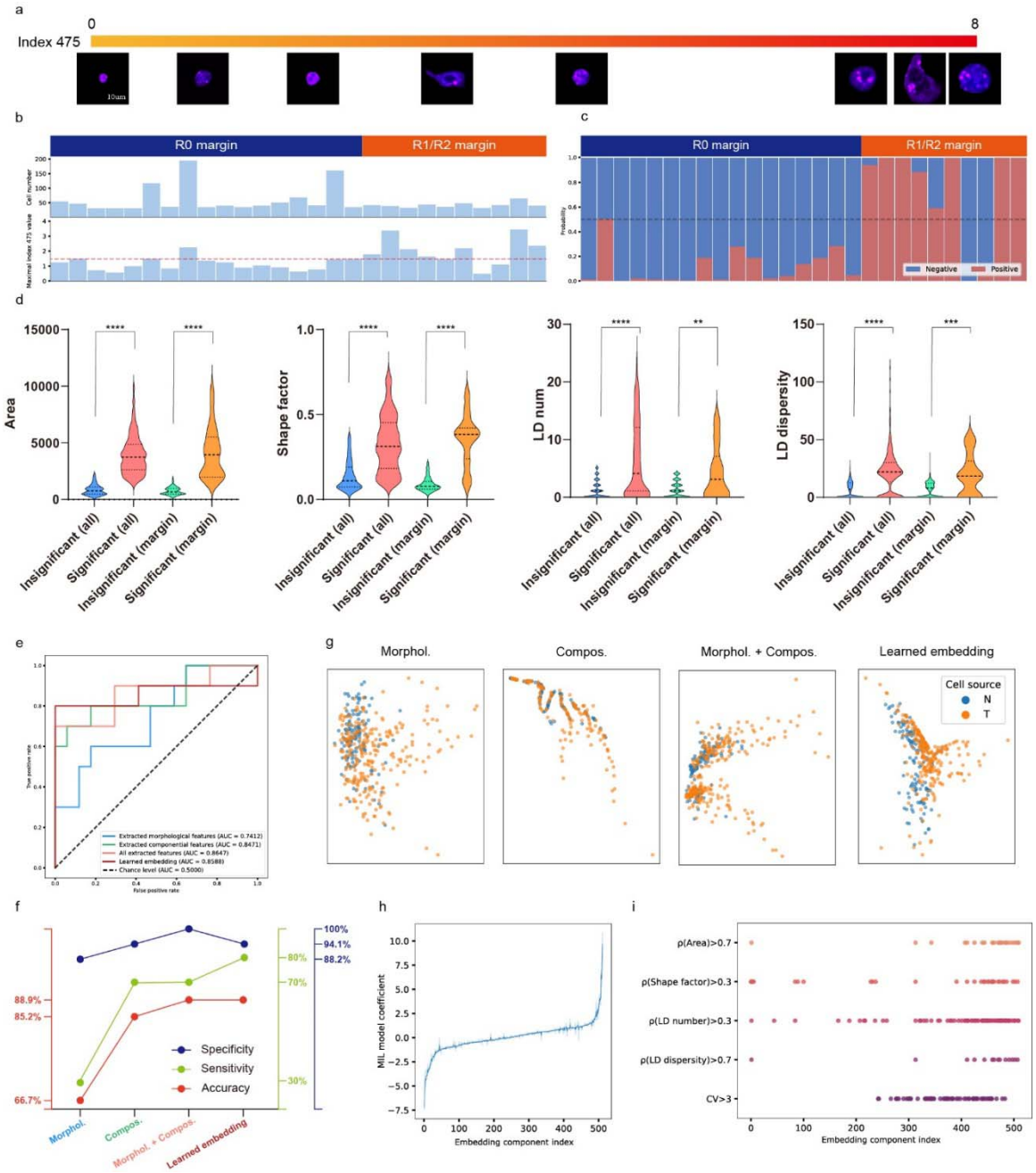

Figure 4 | Multi-instance learning for PNM assessment. **a**, Cell images with different values of a representative embedding component. **b,** Exfoliated cell numbers of M samples and their cells' maximal values of the embedding component in **a**. **c**, Positive probability of M samples. **d**, Feature distribution in significant and insignificant cells. The ranges of feature values are different from **Fig. 2d** because of removed outliers. **e**, ROC curves of PNM assessment in different feature extraction. **f**, Metrics (Specificity, sensitivity, and accuracy) of PNM assessment in different feature extraction. **g**, ISOMAP distribution in different feature extraction. Cells are down-sampled by a factor of 10 for visual clarity. **h**, MIL model coefficients sorted by their mean values in all folds of the LOOCV. Light blue shows their ranges. **i**, Embedding components with high SCCs with cell features and high CV. **: p<0.01, ***: p<0.001, ****: p<0.0001.