Journal of Molecular Cell Biology (2022), 14(6), mjac044 https://doi.org/10.1093/jmcb/mjac044
Published online July 23, 2022

Article

AIM-CICs: an automatic identification method for cell-in-cell structures based on convolutional neural network

Meng Tang[1,2,6,†], Yan Su[2,†], Wei Zhao[3,†], Zubiao Niu[2,†], Banzhan Ruan[2], Qinqin Li[1], You Zheng[2], Chenxi Wang[2], Bo Zhang[1,2], Fuxiang Zhou[4], Xiaoning Wang[5], Hongyan Huang[1,*], Hanping Shi[1,*], and Qiang Sun[2,*]

[1] Beijing Shijitan Hospital of Capital Medical University, Beijing 100038, China
[2] Laboratory of Cell Engineering, Institute of Biotechnology, Research Unit of Cell Death Mechanism, Chinese Academy of Medical Science, 2021RU008, Beijing 100071, China
[3] School of Mathematical Sciences, Peking University, Beijing 100871, China
[4] Department of Radiation and Medical Oncology, Hubei Key Laboratory of Tumor Biological Behaviors, Hubei Clinical Cancer Study Center, Zhongnan Hospital, Wuhan University, Wuhan 430071, China
[5] National Clinic Center of Geriatric & State Key Laboratory of Kidney, Chinese PLA General Hospital, Beijing 100853, China
[6] Comprehensive Oncology Department, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China
[†] These authors contributed equally to this work.
* Correspondence to: Qiang Sun, E-mail: sunq@bmi.ac.cn; Hongyan Huang, E-mail: hhongy1999@126.com; Hanping Shi, E-mail: shihp@ccmu.edu.cn

Edited by Luonan Chen

Whereas biochemical markers are available for most types of cell death, current studies on non-autonomous cell death by entosis rely strictly on the identification of cell-in-cell structures (CICs), a unique morphological readout that can only be quantified manually at present. Moreover, the manual CIC quantification is generally over-simplified as CIC counts, which represents a major hurdle against profound mechanistic investigations. In this study, we take advantage of artificial intelligence technology to develop an automatic identification method for CICs (AIM-CICs), which performs comprehensive CIC analysis in an automated and efficient way. The AIM-CICs, developed on the algorithm of convolutional neural network, can not only differentiate between CICs and non-CICs (the area under the receiver operating characteristic curve (AUC) > 0.99), but also accurately categorize CICs into five subclasses based on CIC stages and cell number involved (AUC > 0.97 for all subclasses). The application of AIM-CICs would systemically fuel research on CIC-mediated cell death, such as high-throughput screening.

Keywords: cell-in-cell structure, artificial intelligence, AIM-CICs, cell death, entosis, convolutional neural network

Introduction

Cell-in-cell structures (CICs) typically refer to the unusual eukaryotic cells involving the whole objects internalized partially or completely inside of others, which have been observed in diverse physiological and pathological samples for centuries (Steinhaus, 1891; Wang, 2015; Fais and Overholtzer, 2018a; Zheng et al., 2021). The presence of CICs was reported to be correlated with patient prognosis in a group of human tumors, such as breast cancer (Zhang et al., 2019), head and neck squamous carcinoma (Schwegler et al., 2015; Fan et al., 2020), and pancreatic ductal adenocarcinoma (Huang et al., 2020). Functional studies implicated CICs in a number of biomedical processes, including embryonic development (Lee et al., 2019), mitotic surveillance (Liang et al., 2020), tumor evolution (Lugini et al., 2003; Sun et al., 2014b), immune homeostasis (Davies et al., 2019; Sun and Chen, 2022), cancer immunotherapy (Su et al., 2022), and the forth (Fais and Fauvarque, 2012). As an evolutionarily conserved process, CIC formation is underlain by multiple mechanisms, such as entosis (Overholtzer et al., 2007), cannibalism (Lugini et al., 2006; Sharma and Dey, 2011), and emperitosis (Wang et al., 2013). Among these, entosis is one

Received December 16, 2021. Revised April 1, 2022. Accepted July 20, 2022.

of the best-studied processes, which generally ends up with the death of the internalized cells in an acidified lysosome-dependent way (Overholtzer et al., 2007; Magtanong et al., 2016; Su et al., 2021), but different from autophagy (Marino et al., 2010, 2012; Fais and Overholtzer, 2018b). The formation of entotic CICs turned out to be a genetically controlled process, where cell internalization is driven cell-autonomously by polarized actomyosin resulting from the E-cadherin-mediated adherens junctions (Sun et al., 2014a; Wang et al., 2015) and coordinated by a mechanical ring interfacing in between (Wang et al., 2020b). Additionally, an ever-expanding set of factors, acting through either actomyosin or adherent junctions or mechanical rings, were identified as important regulators (Hinojosa et al., 2017; Liang et al., 2018; Ruan et al., 2018; Wang et al., 2020a).

Despite great progress made over the past decade, the studies on CIC formation were, however, based on the over-simplified readout of CIC counts that were performed manually, which is not only labor-intensive and time-consuming but also sharply incompatible with the complex CIC formation *per se*. First, since CIC formation is a dynamic process preceding through sequential steps including cell–cell contact, penetration, and closing (Galluzzi et al., 2018; Niu et al., 2021), CICs formed at different stages display partial or complete morphologies. Second, the CIC morphologies are further complicated by the involvement of multiple cells, which frequently results in structures of 'cell-in-cell-in-cell' or even more. Third, due to personal experience and preference, the CIC judgment and inclusion–exclusion criteria for analysis vary from investigator to investigator, making it hard to compare across studies from different labs, or even studies from different investigators in one lab. In addition, manual quantification is rather inefficient in dealing with a large number of samples that may serve the screening purpose. Thus, the traditional CIC quantification reported less informative, hardly comparable, and low-throughput results, which calls for more efficient and informative ways for the quantification of CICs.

Recent years have witnessed the rapid development of image-based artificial intelligence (AI) technology in assisting biomedical practices. For example, by using a single convolutional neural network (CNN) algorithm, Esteva et al. (2017) demonstrated the classification of skin lesions in performance on par with all tested experts. Lin et al. (2020) developed a ResNeXt WSL model that achieved impressive performance (94.09% accuracy, 92.79% sensitivity, and 98.03% specificity) in making chromosome cluster type identification. Actually, simply based on microscopic images, AI algorithms were quite competent in analyzing most, if not all, biological events, such as the early onset of pluripotent stem cell differentiation (Waisman et al., 2019), tumor cell malignancy (Oei et al., 2019), mitosis staging (Mao et al., 2019), and the like. The remarkable potentials in accuracy and efficiency make AI-based image analysis an ideal method for comprehensive and reliable CIC quantification.

In this study, based on RGB fluorescent microscopic images, we employed the deep CNN algorithms (Faster-RCNN and ResNet) to evaluate a large number of cell candidates with defined subtypes and trained a multiclassifier for the recognition of subdivided CICs, which was named AIM-CICs abbreviated from automatic identification method of CICs. The AIM-CICs exhibited a high level of sensitivity and specificity, as evidenced by the area under the receiver operating characteristic (ROC) curve (AUC) values >0.97 for all tasks, in differentiating CICs from non-CICs and identifying subtyped CICs from multiple cells. The development and application of AIM-CICs holds the promise of speeding up CIC-related studies, such as deciphering the molecular controls of CIC formation in a finer resolution and enabling image-based systemic screening by high-content microscopy.

Results

The deep-learning framework of AIM-CICs

In this work, we conducted a framework of object detection and classification based on manual annotation in the training and validation sets and then performed inspections in the test set (Figure 1). The pipeline of the AIM-CICs task includes image processing, cell detection, and CIC classification. For an RGB-format image, the proposed system performs two consecutive steps. First, a Faster-RCNN (Ren et al., 2017) network with a ResNet-50 (He et al., 2016) backbone was formulated to find the cell regions and extract the candidate patches. Second, each candidate, representing one cell or CIC, was classified by a ResNet-101 network based on the cellular morphology. Subsequently, these subdivided candidates of the predicted results were grouped into different folders and marked out on the original locations of the corresponding images.

Cell region detection and extraction

Cell region detection is the initial task to investigate microscopic images. According to the basic cell components, we acquired the fluorescent microscopic images with a red channel for the membrane and a blue channel for the nucleus. Along with the bright field, the merged images could be further composited into RGB format with variant cell quantities and brightness values (Supplementary Figure S1A). The extraction of cell candidates aimed to propose regions of interest that potentially involved CICs. This step served to reduce the searching space and improve the efficiency of subsequent steps in a high-content study. Initially, four pieces of MCF7 images and four pieces of MCF10A images, which included 2164 cells in total, were used as the training set for cell region detection. Through manually annotating these images using VGG Image Annotator (VIA, https://www.robots.ox.ac.uk/~vgg/software/via/) (Figure 2A), cell region detection was further treated as a classic 1-class object detection task through the Faster-RCNN (Ren et al., 2017) network with the ResNet-50 (He et al., 2016) backbone. Specifically, during training, we performed random flip, random rotation, and random scale for data augmentation, which greatly expanded the data diversity. Following the training process (Figure 2B), we ensured the applicability of this step with an average of 88% precision and 96% recall, at an intersection over union (IoU) 0.1, by randomly testing 10 pieces of MCF7 and MCF10A images, which covered 2398 cells (Figure 2C).

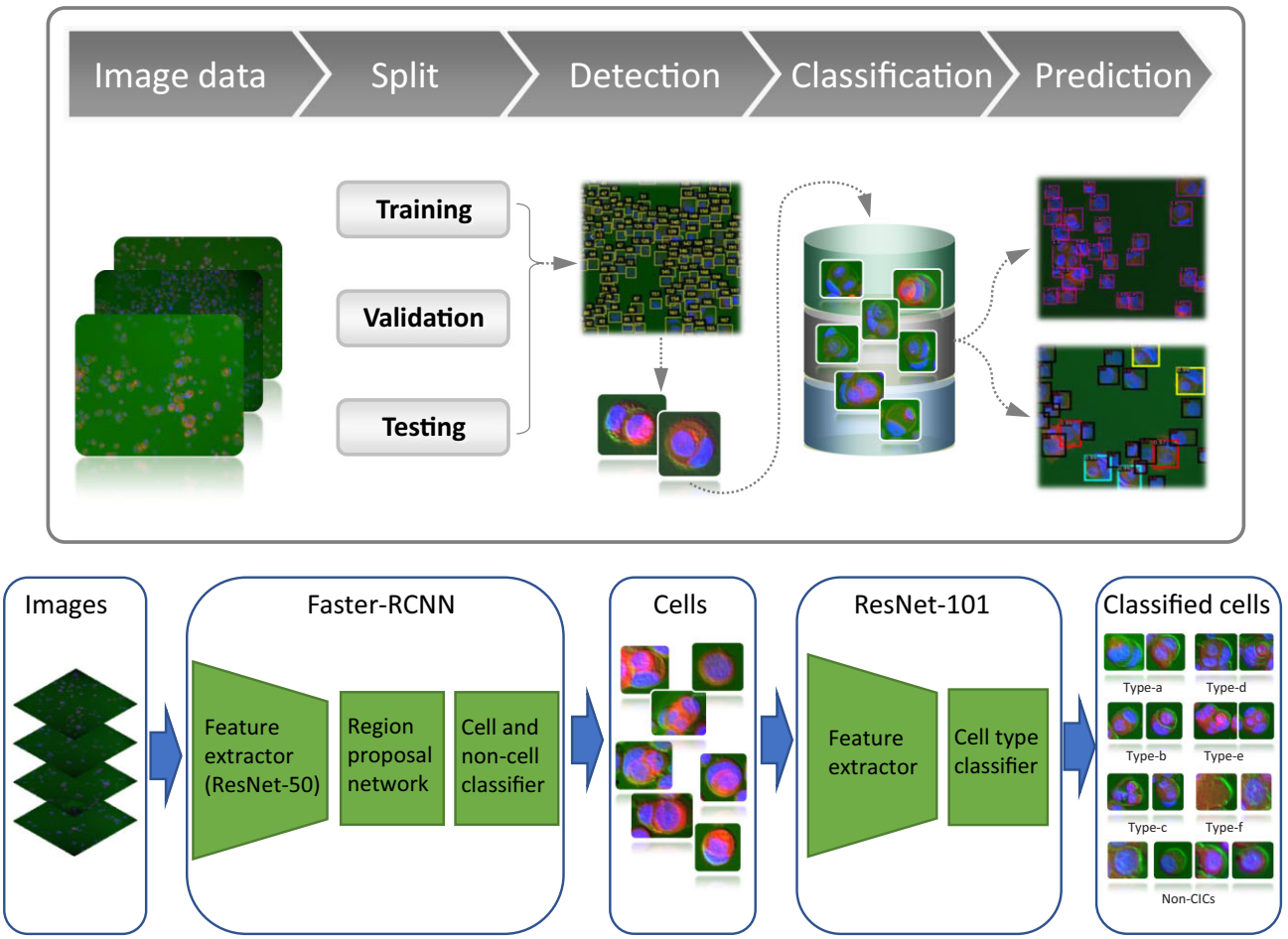


Figure 1 The workflow for the development of AIM-CICs. Image data: the image processing step. Detection: a classic 1-class object detection task based on a Faster-RCNN network with a ResNet-50 backbone was carried out for cellular morphological learning. Classification: ResNet-101 model was applied as our classifier with a 7-class classifier, which outputs a 7-element vector representing the probability for the test sample to belong to each type. Type-a: partial. Type-b: one-in-one. Type-c: two-in-one. Type-d: in turn. Type-e: complicated. Type-f: F category among non-CICs, which was defined as unclear or not sure for the cell recognition.

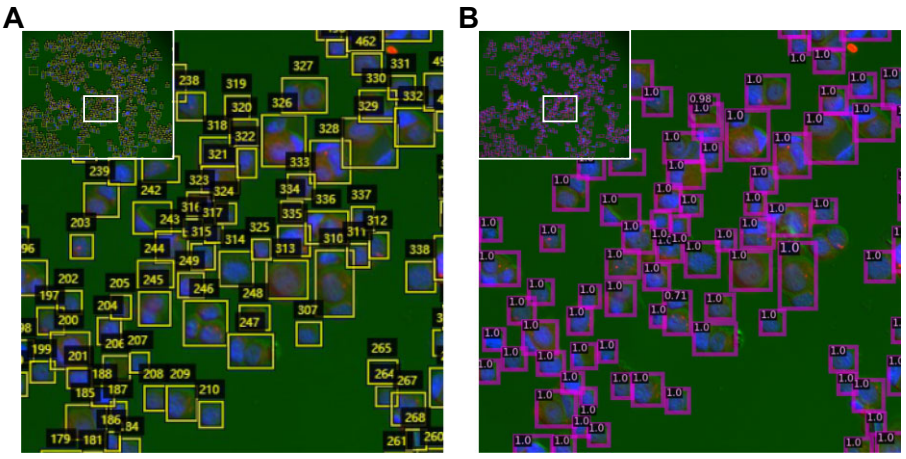
It is believed that factors, such as cell morphology, sample density, as well as image brightness, do impact the accuracy of target detection and recognition. In the data collected in this study, MCF10A samples generally displayed a larger cell size and a much more complicated pattern of CICs as compared with MCF7 samples (Supplementary Figure S1A). Based on the precise manual labelling, we could minimize the effect of target varieties among MCF7 and MCF10A samples (Supplementary Figure S1B), except for the over-exposed fluorescent images that should be excluded in the processing of the primary images. Eventually, we exported the patches of detected cell regions of the entire RGB-format images for the following analysis.

Definition of the structural subtypes of CICs

To classify the CICs, we first divided the traditional CICs into five structural subtypes, including (a) partial, with >30% of the internalizing cells enclosed, but not fully, by the outer cells; (b) one-in-one, with only one cell fully internalized; (c) two-in-one, with two cells fully internalized; (d) in turn, a nested CIC with multiple cells sequentially internalized into neighboring cells; and (e) complicated, a complex CIC generated by four or more cells (Figure 3A). Considering the potential complexity, two kinds of breast cell lines, MCF7 and MCF10A, were investigated, in which the total rate of CICs and their subtypes showed great discrepancy according to the manual labelling (Figure 3B). In total, 17 pieces of MCF7 images and 85 pieces of MCF10A images were enrolled in this study, the cell number of each image ranged from 100 to 600 and from 30 to 200, respectively (Figure 3C). The overall CIC rate of each image counted from 1% to 85% (Figure 3D).

Multi-subtype classification achieved by the AIM-CICs

The obtained cell candidates were used to train the ResNet101 model for CIC recognition (Supplementary Figure S3A). Practically, we used 13 pieces of MCF7 images and 32 pieces of MCF10A images as the training set, which had



C

Groups	N_GT	N_ prediction	IoU 0.1				IoU 0.5			
			N_recall	Recall (%)	Precision (%)	F-score (%)	N_recall	Recall (%)	Precision(%)	F-score (%)
Val_MCF7	1783	1782	1691	94.7 ± 2.1	94.7 ± 1.7	94.7 ± 1.8	1650	92.3 ± 4.5	92.3 ± 4.0	92.3 ± 4.3
Val_MCF10A	347	367	336	96.8 ± 1.3	90.4 ± 4.9	93.5 ± 3.0	333	96.0 ± 2.5	89.7 ± 5.1	92.7 ± 3.5
Testing	268	331	258	96.2 ± 2.5	78.1 ± 1.8	86.2 ± 0.2	238	89.3 ± 5.8	72.5 ± 2.5	80.0 ± 3.6

Figure 2 Cell region detection and extraction. (**A**) Using a VGG Image Annotator to annotate cell regions in the training set. Each box represents an individual cell region with numeric order on it. (**B**) Model-based prediction of the cell region in the validation set. Each cell region was indicated by a box with a predicted confidence score. (**C**) Performance of cell region detection model. The MCF7 and MCF10A images in the validation set belong to the same batch of training set images, and the test set is composed of MCF7 and MCF10A images from independent experiments. N_GT, number of ground truth cells. N_prediction, cell number of the model's prediction (confidence threshold set to 0.1). Val_MCF7, validation set of MCF7 samples. Val_MCF10A, validation set of MCF10A samples. Testing, test set. Data are presented as mean ± SD.

4026 MCF7 cells with a CIC rate of 11% and 3912 MCF10A cells with a CIC rate of 32%, respectively (Supplementary Figure S3B). Based on the morphological features of cell candidates, five subtypes of CICs were manually labelled for each cell candidate in the training and validation sets. The distribution of each subtype of CICs showed a remarkable discrepancy, as in the test set (Figure 4A and B). To improve the practicality of the model, we defined an F category for the non-CIC candidates. The F category contains ambiguous structures that were hard to tell their identities by both experienced experts and AI algorithm and thus were generally removed from the sample counting (Supplementary Figure S3C and D).

As shown in Figure 4C, data training progressively increased the prediction accuracy to a considerable level for each subtype. In both training and validation sets, the comprehensive accuracy of integrated CICs (involving a, b, c, d, and e types) and non-CIC types (including the F category) revealed approving performance (Figure 4D). Moreover, the AIM-CICs also exhibited impressive performance in the test set, as indicated by the AUC > 0.97 for each CIC subtype (partial 0.9761, one-in-one 0.9807, two-in-one 0.9872, in turn 0.9709, and complicated 0.9984) (Figure 4E and F). Additionally, for the low-quality images in the test set that displayed unclear cell regions and were eventually removed for further analysis, their recognition also reached an ideal AUC of 0.99 (Supplementary Figure S3E). Moreover, the AIM-CICs performed accurate recognition of CICs on independent datasets of different cell types, such as colon cancer cells (SW480-E and SW620) and cervical cancer cells (Hela), suggesting the generalizability of this model (Supplementary Figure S5).

Visualization of morphological features and output

To better understand what the model learned from the annotated data, we extract features from the output of the network's global average pooling layer and applied t-SNE to reduce the dimension to 2D for visualization. For the training set, each group of cell samples represented independent clusters, except for cell candidates in the circled region (Figure 5A). Backtracking the training data identified that these were candidates categorized into two subtypes due to erroneous manual annotation. Comparing the manual efficacy with the AIM-CICs model among well-trained biologists and newly trained biological students, the AIM-CICs model showed superiority over different investigations (Supplementary Figure S6). Thus, the t-SNE-based clustering would be a visualized way for error correction in recognizing CICs. For the test set (Figure 5B), subtypes of CICs were clustered into the close, but clearly distinct, regions, whereas the F category was neighboring to the area of non-CICs as expected. Furthermore, the multi-subtype classification model

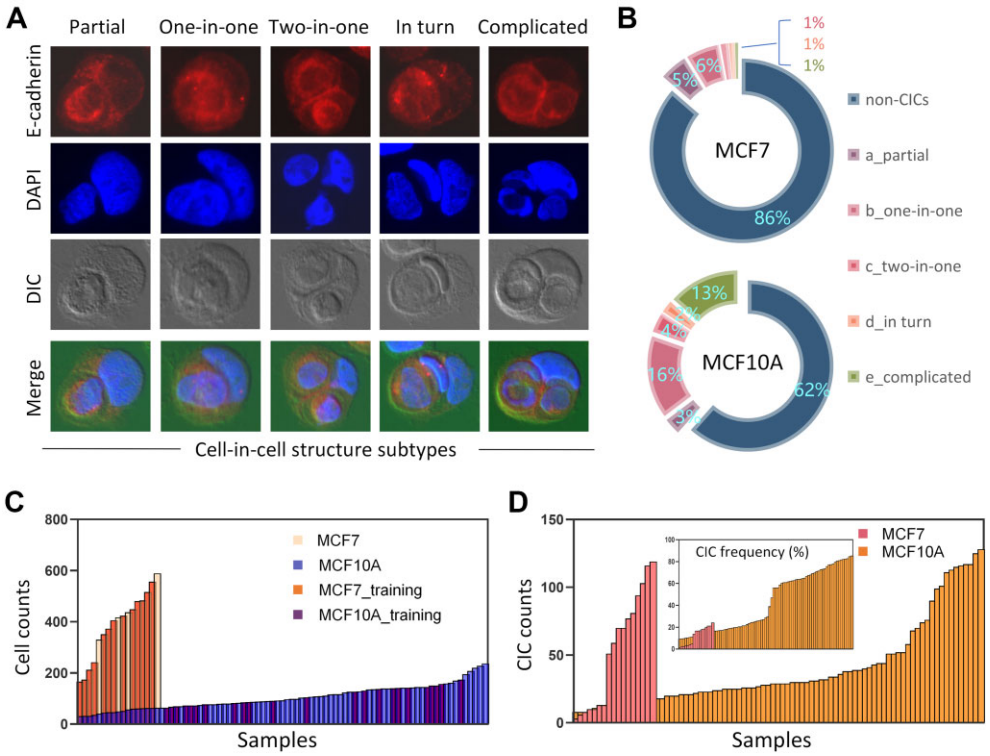


Figure 3 Image processing and cell candidate regions. (**A**) The representative images of five CIC subtypes. Cell membrane (E-cadherin) in red, nucleus (DAPI) in blue, and background in green. Original images are listed in Supplementary Figure S1A. (**B**) The percentages of different CIC subtypes for MCF7 and MCF10A cells. (**C**) The number of cell candidates extracted from each MCF7 or MCF10A image. Samples contained 17 pieces of MCF7 images and 85 pieces of MCF10A images. The columns in orange and purple represent samples used in the training set. (**D**) The count and frequency of CICs for individual MCF7 or MCF10A image.

imitating the dynamic process exhibited comparable performance with the conventional bisection in CIC analysis (Supplementary Figure S6). Therefore, following the comprehensive recognition under a specified confidence threshold, we were able to accurately locate each structure with a predicted value on the original images (Figure 5C).

Application of the AIM-CICs in an experimental setup

To explore the potential implications of AI-based recognition of CICs in a biological context, we included a functional experiment as an example of subtype profiling. In this analysis, the confidence threshold was set to 0.2 for more informative identification (Figure 6A). As the results showed, though all of the three truncations of ARHGAP36, a molecule identified to be a regulator of CIC formation in a screening study (Ruan et al., 2018), impaired the formation of CICs, the alterations of CIC subtypes were rather different (Figure 6B and C). While the truncated GAP (1–194) had little impact on the formation of partial CICs (Figure 6B–D), the majority of CICs were in completed form (including all CIC subtypes except the partial type) in cells expressing the truncated GAP (118–194) or GAP (195–395) (Figure 6B and C), suggesting that the N-terminal region (1–117) of ARHGAP36 might function to slow down the process of cell internalization. Meanwhile, the C-terminal region of ARHGAP36 was likely to be responsible for the closing step of CIC formation, as evidenced by the comparable formation of completed CICs between control and GAP (195–395)-expressing cells (Figure 6B and E–H). Moreover, GAP (118–194) seemed to be the major region that drives cell internalization, as it promoted the formation of completed CICs at a rate comparable to the GAP (1–194) region. Furthermore, though the N-terminal region might negatively regulate the speed of CIC formation, it did function positively to promote cell internalization, as its truncation significantly reduced the formation of both partial and completed CICs (Figure 6B–D). Thus, the AIM-CICs algorithm allows us, for the first time, to accurately dissect the impacts of different domains or molecules on CIC formation in a heretofore underappreciated resolution.

Discussion

Fluorescent microscopic images record the cellular structures, such as CICs, but inevitably provide a great number of morphological variations (Cieri et al., 2017; Zhang et al., 2021). To provide recognition with sufficient accuracy and potentially featured insights, we, for the first time, explored the application of CNN in the profiling of subtyped CICs formed during entosis, a non-apoptotic cell death process that occurs via cell-in-cell invasion (Overholtzer et al., 2007). Based on the tons of images accumulated from previous studies, we developed the AI-based

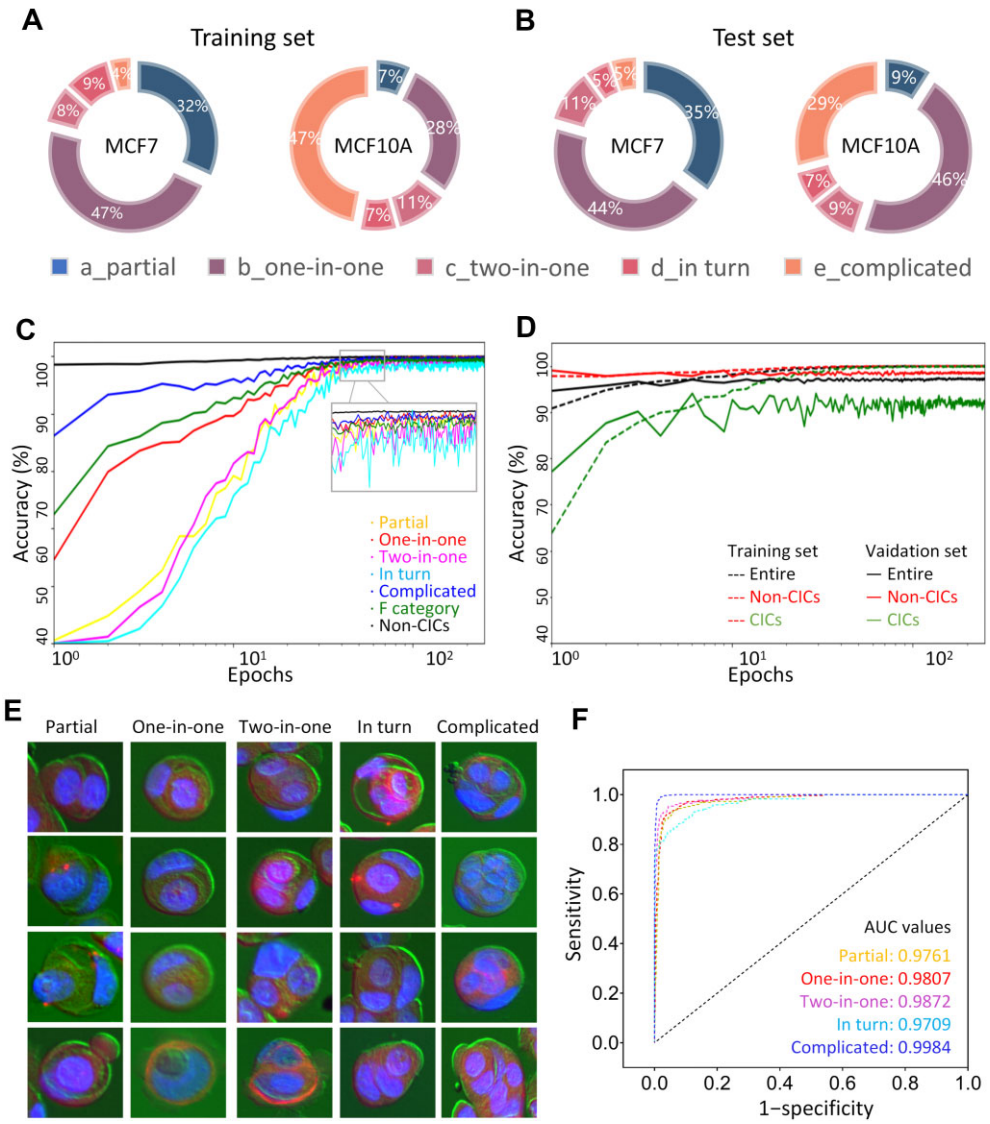


Figure 4 Training and testing processes of the multiple classifications. (**A** and **B**) The percentages of CIC subtypes in the training set (**A**) and the test set (**B**). The CIC counts of MCF7 were 437 (**A**) and 340 (**B**), respectively. In the MCF10A samples, the CIC counts were 1269 (**A**) and 1948 (**B**), respectively. Associated data are listed in Supplementary Figure S2. (**C**) The prediction accuracy of AIM-CICs for each subtype in a 250-epoch learning process. (**D**) The integrated accuracy of AIM-CICs for CIC and non-CIC types in the training and validation sets in a 250-epoch learning process. CICs included the partial, one-in-one, two-in-one, in turn, and complicated types. Non-CICs referred to F category and non-CICs. (**E**) Representative images of each CIC subtype predicted in the training set. (**F**) The ROC curves for each CIC subtype in the test set.

identification algorithm AIM-CICs, which was trained with distinct illumination, textures, and density, in order to deliver optimal performance in cell region detection and multi-subtype classification, despite the unseen perturbations.

In the proposed system, we set up two tasks: a classic 1-class object detection model formulated to find the cell regions as the first task, followed by multi-class object recognition as the second task. Compared with the traditional end-to-end manner, i.e. to train a multi-class detection model with different kinds of cells marked simultaneously, our model of separated detection achieved the flexibility for the raw samples to be recategorized and repurposed. We chose Faster-RCNN for our detection model, since it is a classic, widely used two-stage detection model with good flexibility and robustness. Compared to Faster-RCNN, one-stage detection models (e.g. YOLO series) and anchor-free models (e.g. FCOS) require many tricks to work (Redmon et al., 2016; Redmon and Farhadi, 2018; Tian et al., 2022), but exhibit only comparable performance. In the AIM-CICs developed in this study, the second task included a well-trained 7-category classifier (five CIC subtypes plus non-CICs and F category) based on the ResNet-101 model to define the multiple subtypes of CICs, which was compatible with the cell candidates from the first

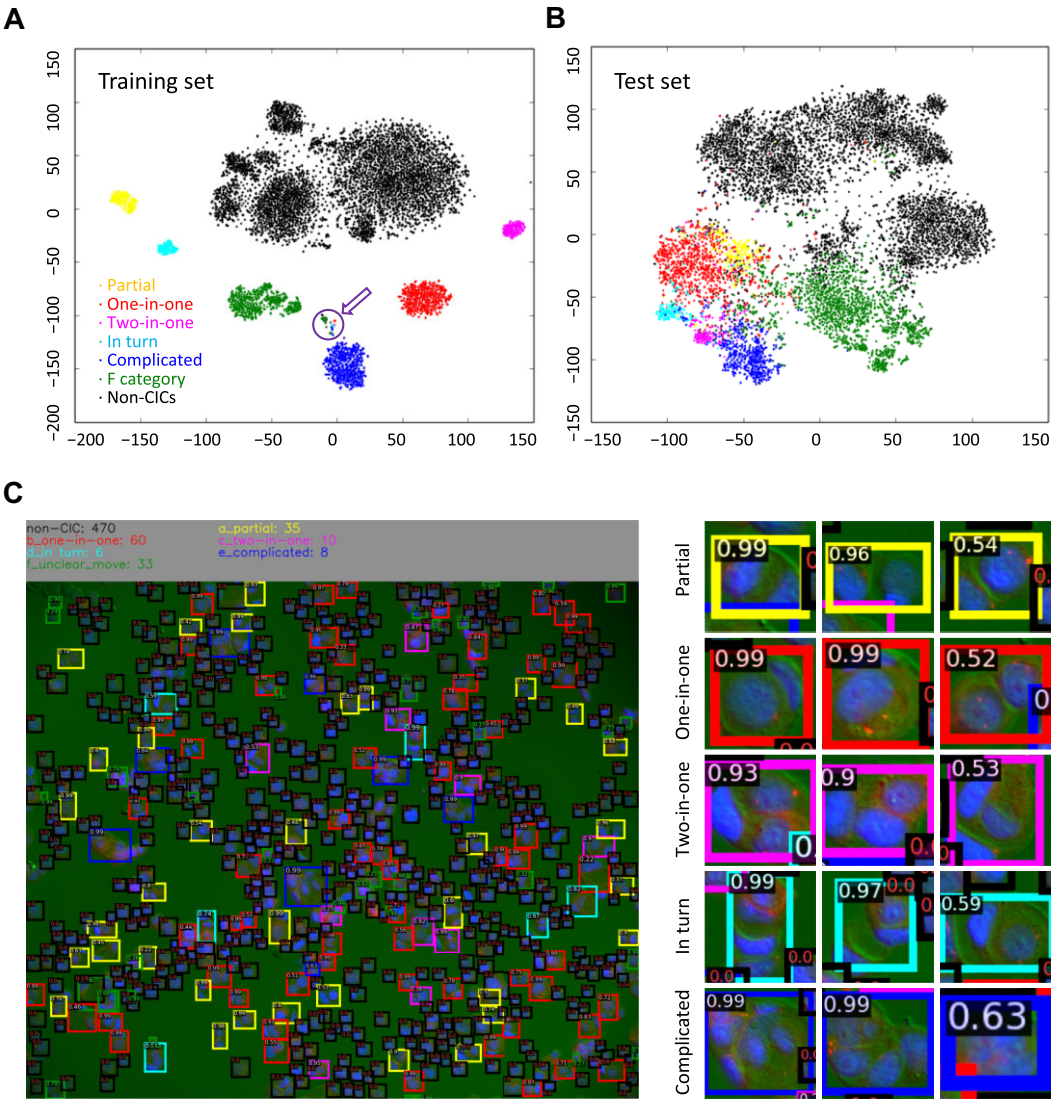


Figure 5 Visualization of sample features and output. (**A** and **B**) The two-dimension visualization of CIC subtypes in the training set (**A**) and test set (**B**). The samples circled out were those predicted by the AIM-CICs to be miscategorized manually. (**C**) A representative image showing the recognition result of AIM-CICs. Colored frames indicate structures in different categories. The predicated confidence scores are marked on the up-left corner of each structure. The structures in the right were cropped from the image in the left.

step. Unlike most modern classification models (i.e. NAS-based and ViT-based) requiring large datasets and many tricks to train, the traditional ResNet model is more suitable for our dataset scale (He et al., 2016; Tan et al., 2019). This two-step algorithm is also advantageous in debugging the possible mechanisms leading to inferior final prediction outcomes, as each step could be optimized separately. Meanwhile, this two-step algorithm may fall short of efficiency (speed) as compared to the end-to-end multi-class detection model that could utilize a shared feature extraction backbone.

Among all the defined cell death programs, CIC-mediated death is unique in that it can only be accomplished with the involvement of at least two cells, but not one cell in other programs like apoptosis, necrosis, and the forth (Galluzzi et al., 2018). Therefore, the mechanistic study is a challenging task for the field of CIC-mediated death, which is further complicated by the fact of lacking a reliable biochemical marker. Current studies on CICs rely on morphology-based binary quantification, i.e. CICs or non-CICs, where CICs are usually defined as structures with $>1/2$, or $>2/3$ in some studies, of the inner cell body being internalized/enclosed by the outer cell. This over-simplified quantification of CICs did move the field forward over the past decade but provided rather coarse information over a more complicated process (Rizzotto and Villunger, 2020; Niu et al., 2021). CIC formation is a stepwise process that could be empirically subdivided into three major stages: (i) the early initiation stage from cell–cell contact to $\sim 1/3$ of the inner cell body being internalized, which is primarily driven by cell–cell adhesion and assisted

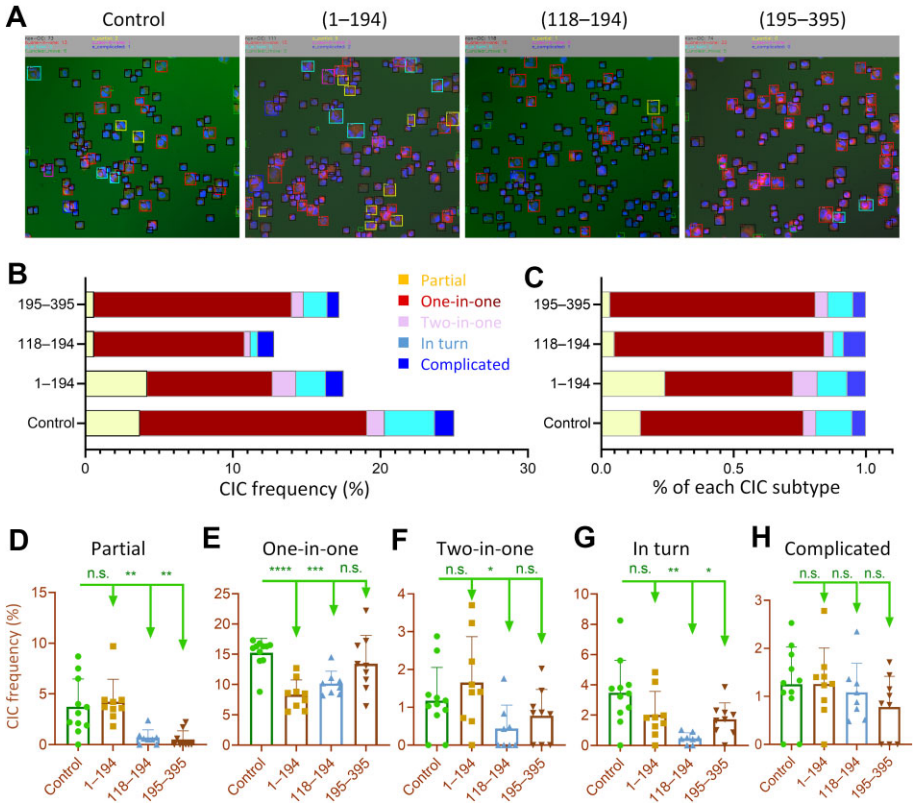


Figure 6 Analysis of CIC subtypes in an experimental setup by the AIM-CICs. (**A**) The visualized recognition results of AIM-CICs in an experiment where MCF10A cells expressed the empty vector (control) and three truncated mutants of ARHGAP36 (1–194, 118–194, and 195–395), respectively. (**B** and **C**) Graphs show the absolute frequency (**B**) or relative frequency (**C**) of subtyped CICs in MCF10A cells expressing different ARHGAP36 mutants ($n = 934$ cells for control, 1060 cells for 1–194, 1392 cells for 118–194, and 852 cells for 195–395). (**D–H**) The frequencies of subtyped CICs in MCF10A cells expressing different ARHGAP36 mutants. Data are shown as box-plots with means and individual data points. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ (two-tailed Student's *t*-test).

by cytoskeleton remodeling; (ii) the middle internalization stage covering the whole process of cell internalization, which is primarily driven by active actomyosin contraction within the inner cells and coordinately assisted by the outer cells; and (iii) the final closing stage that may involve tail cutting and membrane fusion, which is rarely investigated largely because it is technically challenging. Furthermore, CIC formation is a dynamic process that multiple cells, either sequentially or simultaneously, form a complicated structure that may contain more than one cell inside (Figure 3A). The regulation of this feature is completely unknown for the field yet, but might be conceptually feasible, as it was reported in phagocytosis that the number of corpses engulfed by a phagocyte was genetically controlled (Park et al., 2011). Taking these two factors (stage and cell number) into account produces an even higher dimensional complexity, which, however, was missed from the traditional analysis by the binary quantification. The implantation of AIM-CICs enables us to make a more sophisticated description of CIC phenotypes, which would help identify finer molecular control. For example, though the expression of the GAP (195–395) domain did not influence the frequency of simple CICs, where only one cell was enclosed (one-in-one in Figure 6E), it did result in a significantly reduced formation of complex CICs, where more cells were enclosed by one outer cell (two-in-one and in turn in Figure 6F and G). These results suggest that the truncated N-terminal domain may facilitate the internalization of multiple cells to form complex CICs, which warrants further functional validation.

In addition to mechanistic investigation, the AIM-CICs are also promising in enabling high-content-based screening for therapeutic compounds that target CIC formation considering the pivotal roles of CICs in multiple biomedical processes such as cancer (Fais and Overholtzer, 2018a). Since high-throughput screening relies critically on a reliable biochemical marker that is currently unavailable for CIC formation, the related systemic screening, which would be labor-intensive and time-consuming if worked out by manual annotation, has not been reported. Empowered with the AIM-CICs and high-content microscopy, systemic screening would be feasible in the near future.

Materials and methods

Image processing and software

An entire dataset involving 17 pieces of MCF7 images and 85 pieces of MCF10A images was obtained from the Sun lab. As

the detailed protocol described (Sun and Overholtzer, 2013), the fluorescently labelled cells were necessary to be stained with discrepant colors for each cell component, such as red for cytomembrane (E-Cadherin, 1:200, BD Biosciences, 610181), with secondary antibodies Alexa Fluor 568 anti-rabbit (1:500, Invitrogen, A11036), and blue for cytoblast (DAPI, Sigma, D8417) (Supplementary Figure S1A). Random fields were taken under corresponding channels of laser lights through a fluorescent microscope (Nikon Ti-E microscope, Nikon NIS-Elements AR 4.5 software), along with bright colors for the background. For the algorithm performing, each sample with three single-channel images was transformed into an RGB format with a value rescaled to 0–255. Softwares used and algorithms developed in this study include Python (http://www.python.org/), PyTorch (https://pytorch.org/), VIA Annotation Tools (https://www.robots.ox.ac.uk/~vgg/software/via/), and Detectron2 (https://github.com/facebookresearch/detectron2).

Cell region labelling and candidate extraction

After acquiring the processed images, we manually annotated the cell regions through the VGG Image Annotator (https://www.robots.ox.ac.uk/~vgg/software/via/). Based on the annotated images, a classic 1-class object detection task was carried out for cellular morphological learning. The model used was a Faster-RCNN (Ren et al., 2017) network with a ResNet-50 (He et al., 2016) backbone. Since the original resolution of the microscopic image was 2160 × 2560, too large for Faster-RCNN training, we first split each image into 4 × 4 grids, and then followed the common practice to train the model. For data augmentation, we used random flip, random rotation, and random scale to expand the diversity of data. For other hyper-parameters, we set batch size to 24 and iterated 50000 steps using an SGD optimizer with momentum 0.9. For comparison, we trained the data with the FCOS (Tian et al., 2019) and YOLOv3 (Redmon and Farhadi, 2018) models, respectively, in the same process. The average recall (F-score) (%) at IoU 0.1 of Faster-RCNN, FCOS, and YOLOv3 models was 96.2 (86.2), 95.9 (87.3), and 97.7 (84.6), respectively, in the test set (Supplementary Figure S2). As the output of the Faster-RCNN network, the patches of detected cell regions were exported as candidate sequences for further steps (Supplementary Figure S3B and C).

Manual classifications of CICs

The manual definition of CIC classification primarily included bipartite-class CICs and non-CICs. CICs were further subdivided into five subtypes, including (a) partial, (b) one-in-one, (c) two-in-one, (d) in turn, and (e) complicated. To refine the output results, we added an F category among non-CICs, which was defined as unclear or not sure for the cell recognition and needed to be removed for the quantitative analysis (Supplementary Figure S3). The cell candidates involved in the training set were verified together by an expert group consisting of six members in the lab.

Multiple classification models

We used the ResNet-101 model as our classifier, and the input size was set to 224. Since this model could take the detection model's output as input, we cropped cell samples using the detection model and manually labelled them with corresponding cell types. During training, each sample was first padded to square and then resized to 224 × 224. Both horizontal and vertical random flips were performed. We trained our model for 250 epochs with a batch size of 32, using an SGD optimizer with a learning rate of 0.001 and momentum of 0.9. To alleviate over-fitting, a dropout layer with $P = 0.25$ was set right before the feature went into the final fully connected layers. To choose hyper-parameters, we kept 20% of samples as a validation set. For comparison, we trained three other representative methods, DenseNet-201 (Huang et al., 2017), MnasNet (Tan et al., 2019), and MobileNetV2 (Sandler et al., 2018), in the CIC classification task via the same process. The average recall (F-score) (%) of ResNet-101, DenseNet-201, MnasNet, and MobileNetV2 models was 80.3 (78.1), 75.1 (73.2), 54.6 (54.9), and 64.6 (65.9), respectively (Supplementary Figure S4). Eventually, the prediction results could be visualized on the original images with the detected cell regions and a predictive score of CICs, as well as in the output folders of each cell type.

Importantly, when applying our model for inference, the test samples should be padded and resized in the same way as training. Our model was a 7-class classifier, and it output a 7-element vector representing the probability for the test sample belonging to each type. Traditionally, the predicted type should be the type with maximum probabilities. In practice, to increase precision, we predicted cells that had the predicted probability <0.2 as non-CICs, even if the non-CIC probability was not the maximum for it. For example, if the predicted output was [0.1, 0.18, 0.12, 0.15, 0.15, 0.13, 0.17] (for a, b, c, d, e, f, non-CIC), we would use non-CICs as the model's prediction. Ultimately, we could output the classifications into specific folders of each cell type and obtain the visualized results that individual colors were marked on the original images. The typical size of our test images was 2160 × 2560, and it took our detection model 869.6 ms to infer such an image. The time for our classification model was 2.7 ms per sample (1080TI GPU).

Performance analysis of detection model

In the deep-learning community, the most common metric used for quantitatively comparing the performances of detection models is mean average precision, as proposed in Everingham et al. (2009) and Lin et al. (2014). However, since our work mainly focused on multi-type CIC classification instead of general object detection techniques, we reported our detection results in a more practical recall/precision manner. In detail, we kept the detection model's output instances with confidence >0.1 as the model's prediction and calculated metrics at two different IoU thresholds, 0.5 and 0.1. Under IoU threshold 0.5, the model must output an accurate prediction box to get a match, while 0.1 requires only loosely overlapping.

Feature's visualization

To better understand what the classification model learnt from labelled samples, we extracted features from each cell sample and visualized them in a 2D space. The feature we used is the output of the network's global average pooling layer, which is right before the final classification layer. This 2048D feature is the deepest and most semantic, so it can represent the information extracted by the network from a corresponding input image. To visualize these 2048D features, we use the t-SNE method for dimensionality reduction to transform each feature to 2D (Van der Maaten and Hinton, 2008). t-SNE is a popular method for visualizing high-dimension data, since it can keep most of the original data structure during dimensionality reduction.

Evaluation criteria for classification model

The output of the classification model was evaluated by the universal criteria, such as sensitivity (Se or recall), specificity (Sp), precision, the ROC curve, and the AUC. The equations were listed as follows:

$$\text{Se (recall)} = \text{TP}/(\text{TP} + \text{FN}) \quad (1),$$

$$\text{Sp} = \text{TN}/(\text{TN} + \text{FP}) \quad (2),$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (3),$$

$$\text{F} - \text{score} = 2 \times \text{recall} \times \text{precision}/(\text{recall} + \text{precision}) \quad (4).$$

True positive (TP) stands for the accounts of positive CICs, which are correctly recognized as positive CICs. False-positive (FP) stands for the number of negative CICs that are incorrectly recognized as positive CICs. False-negative (FN) stands for the accounts of positive CICs, which are incorrectly recognized as negative CICs. True negative (TN) stands for the number of negative CICs that are correctly recognized as negative CICs.

Data and code availability

The data used in this study are deposited at Github.com (github.com/tammyvv/AIM-CIC). Any additional information required to reproduce this work is available from Dr Qiang Sun.

Statistical analysis

Categorical data are expressed as frequencies (%) and were tested by a two-tailed Student's *t*-test. *P*-values were calculated by Excel or GraphPad Prism software. The level of significance was set at $P < 0.05$.

Supplementary material

Supplementary material is available at *Journal of Molecular Cell Biology* online.

Acknowledgements

We thank Dr Lulin Zhou, Wenzhao Zhou, Nannan Du, Xiaoyi Jiang, He Ren, Yichao Zhu, Yuqi Wang, Lihua Gao, Zhaolie Chen, and the members of the Sun lab for the constructive discussions and assisting manual labelling.

Funding

This work was supported by Beijing Municipal Natural Science Foundation (KZ202110025029 to H.H.), the National Key R&D Program of China (2022YFC3600100 to Q.S. and H.H.), the National Natural Science Foundation of China (32100608 to C.W., 82002918 and 31970685 to Q.S.), Beijing Municipal Administration of Hospitals Incubating Program (PX2021033 to H.H.), and Beijing Postdoctoral Research Foundation (2021-ZZ-027 to M.T.).

Conflict of interest: none declared.

Author contributions: concept and design: Q.S.; CNN model training and analysis: W.Z. and M.T.; data collection: M.T., Z.N., C.W., and B.R.; data interpretation: Q.S., W.Z., and M.T.; figure preparation: M.T., Y.S., W.Z., and Q.S.; manuscript preparation: M.T., Q.S., and W.Z., with input from Z.N., C.W., B.R., Y.Z., B.Z., Q.L., H.H., X.W., F.Z., and H.S.; funding acquisition: Q.S., H.H., and M.T. All authors have read and approved the final manuscript.

References

Cieri, D., Vicario, M., Giacomello, M., et al. (2017). SPLICS: a split green fluorescent protein-based contact site sensor for narrow and wide heterotypic organelle juxtaposition. Cell Death Differ. *25*, 1131–1145.

Davies, S.P., Reynolds, G.M., Wilkinson, A.L., et al. (2019). Hepatocytes delete regulatory T cells by enclysis, a $CD4^+$ T cell engulfment process. Cell Rep. *29*, 1610–1620.e4.

Esteva, A., Kuprel, B., Novoa, R.A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature *542*, 115–118.

Everingham, M., Van Gool, L., Williams, C.K.I., et al. (2010). The Pascal Visual Object Classes (VOC) challenge. Int. J. Comput. Vision *88*, 303–338.

Fais, S., and Fauvarque, M.-O. (2012). TM9 and cannibalism: how to learn more about cancer by studying amoebae and invertebrates. Trends Mol. Med. *18*, 4–5.

Fais, S., and Overholtzer, M. (2018a). Cell-in-cell phenomena in cancer. Nat. Rev. Cancer *18*, 758–766.

Fais, S., and Overholtzer, M. (2018b). Cell-in-cell phenomena, cannibalism, and autophagy: is there a relationship? Cell Death Dis. *9*, 95.

Fan, J., Fang, Q., Yang, Y., et al. (2020). Role of heterotypic neutrophil-in-tumor structure in the prognosis of patients with buccal mucosa squamous cell carcinoma. Front. Oncol. *10*, 541878.

Galluzzi, L., Vitale, I., Aaronson, S.A., et al. (2018). Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. Cell Death Differ. *25*, 486–541.

He, K., Zhang, X., Ren, S., et al. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778, doi: 10.1109/CVPR.2016.90.

Hinojosa, L.S., Holst, M., Baarlink, C., et al. (2017). MRTF transcription and ezrin-dependent plasma membrane blebbing are required for entotic invasion. J. Cell Biol. *216*, 3087–3095.

Huang, G., Liu, Z., Van Der Maaten, L, et al. (2017). Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2261–2269, doi: 10.1109/CVPR.2017.243.

Huang, H., He, M., Zhang, Y., et al. (2020). Identification and validation of heterotypic cell-in-cell structure as an adverse prognostic predictor for young patients of resectable pancreatic ductal adenocarcinoma. Signal Transduct. Target. Ther. *5*, 246.

Lee, Y., Hamann, J.C., Pellegrino, M., et al. (2019). Entosis controls a developmental cell clearance in C. elegans. Cell Rep. *26*, 3212–3220.e4.

Liang, J., Fan, J., Wang, M., et al. (2018). CDKN2A inhibits formation of homotypic cell-in-cell structures. Oncogenesis *7*, 50.

Liang, J., Niu, Z., Zhang, B., et al. (2021). p53-dependent elimination of aneuploid mitotic offspring by entosis. Cell Death Differ. *28*, 799–813.

Lin, C., Zhao, G., Yin, A., et al. (2021). A novel chromosome cluster types identification method using ResNeXt WSL model. Med. Image Anal. *69*, 101943.

Lin, T.Y., Maire, M., Belongie, S., et al. (2014). Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., et al. (eds). Computer Vision—ECCV 2014. Cham: Springer, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Lugini, L., Lozupone, F., Matarrese, P., et al. (2003). Potent phagocytic activity discriminates metastatic and primary human malignant melanomas: a key role of ezrin. Lab. Invest. *83*, 1555–1567.

Lugini, L., Matarrese, P., Tinari, A., et al. (2006). Cannibalism of live lymphocytes by human metastatic but not primary melanoma cells. Cancer Res. *66*, 3629–3638.

Magtanong, L., Ko, P.J., and Dixon, S.J. (2016). Emerging roles for lipids in non-apoptotic cell death. Cell Death Differ. *23*, 1099–1109.

Mao, Y., Han, L., and Yin, Z. (2019). Cell mitosis event analysis in phase contrast microscopy images using deep learning. Med. Image Anal. *57*, 32–43.

Marino, M.L., Fais, S., Djavaheri-Mergny, M., et al. (2010). Proton pump inhibition induces autophagy as a survival mechanism following oxidative stress in human melanoma cells. Cell Death. Dis. *1*, e87.

Marino, M.L., Pellegrini, P., Di Lernia, G., et al. (2012). Autophagy is a protective mechanism for human melanoma cells under acidic stress. J. Biol. Chem. *287*, 30664–30676.

Niu, Z., He, M., and Sun, Q. (2021). Molecular mechanisms underlying cell-in-cell formation: core machineries and beyond. J. Mol. Cell Biol. *13*, 329–334.

Oei, R.W., Hou, G., Liu, F., et al. (2019). Convolutional neural network for cell classification using microscope images of intracellular actin networks. PLoS One *14*, e0213626.

Overholtzer, M., Mailleux, A.A., Mouneimne, G., et al. (2007). A nonapoptotic cell death process, entosis, that occurs by cell-in-cell invasion. Cell *131*, 966–979.

Park, D., Han, C.Z., Elliott, M.R., et al. (2011). Continued clearance of apoptotic cells critically depends on the phagocyte Ucp2 protein. Nature *477*, 220–224.

Redmon, J., and Farhadi, A. (2018). YOLOv3 an incremental improvement. arXiv, https://doi.org/10.48550/arXiv.1804.02767

Redmon, J., Divvala, S., Girshick, R., et al. (2016). You only look once: unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788, doi: 10.1109/CVPR.2016.91.

Ren, S., He, K., Girshick, R., et al. (2017). Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. *39*, 1137–1149.

Rizzotto, D., and Villunger, A. (2021). p53 clears aneuploid cells by entosis. Cell Death Differ. *28*, 818–820.

Ruan, B., Wang, C., Chen, A., et al. (2018). Expression profiling identified IL-8 as a regulator of homotypic cell-in-cell formation. BMB Rep. *51*, 412–417.

Sandler, M., Howard, A., Zhu, M., et al. (2018). MobileNetV2: inverted residuals and linear bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4510–4520, doi: 10.1109/CVPR.2018.00474.

Schwegler, M., Wirsing, A.M., Schenker, H.M., et al. (2015). Prognostic value of homotypic cell internalization by nonprofessional phagocytic cancer cells. Biomed. Res. Int. *2015*, 359392.

Sharma, N., and Dey, P. (2011). Cell cannibalism and cancer. Diagn. Cytopathol. *39*, 229–233.

Steinhaus, J. (1891). Ueber carcinom-einschlüsse. Archiv f. pathol. Anat. *126*, 533–541. https://doi.org/10.1007/BF01937613

Su, Y., Huang, H., Luo, T., et al. (2022). Cell-in-cell structure mediates in-cell killing suppressed by CD44. Cell Discov. *8*, 35.

Su, Y., Ren, H., Tang, M., et al. (2021). Role and dynamics of vacuolar pH during cell-in-cell mediated death. Cell Death Dis. *12*, 119.

Sun, Q., and Chen, W. (2022). Cell-in-cell: an emerging player in COVID-19 and immune disorders. Natl Sci. Open *1*, 20220001.

Sun, Q., Cibas, E.S., Huang, H., et al. (2014). Induction of entosis by epithelial cadherin expression. Cell Res. *24*, 1288–1298.

Sun, Q., Luo, T., Ren, Y., et al. (2014). Competition between human cells by entosis. Cell Res. *24*, 1299–1310.

Sun, Q., and Overholtzer, M. (2013). Methods for the study of entosis. Methods Mol. Biol. *1004*, 59–66.

Tan, M., Chen, B., Pang, R., et al. (2019). MnasNet: platform-aware neural architecture search for mobile. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2815–2823, doi: 10.1109/CVPR.2019.00293.

Tian, Z., Shen, C., Chen, H., et al. (2019). FCOS: fully convolutional one-stage object detection. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 9626–9635, doi: 10.1109/ICCV.2019.00972.

Tian, Z., Shen, C., Chen, H., et al. (2022). FCOS: a simple and strong anchor-free object detector. IEEE Trans. Pattern Anal. Mach. Intell. *44*, 1922–1933.

Van der Maaten., L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Waisman, A., La Greca, A., Mobbs, A.M., et al. (2019). Deep learning neural networks highly predict very early onset of pluripotent stem cell differentiation. Stem Cell Rep. *12*, 845–859.

Wang, C., Chen, A., Ruan, B., et al. (2020a). PCDH7 inhibits the formation of homotypic cell-in-cell structure. Front. Cell Dev. Biol. *8*, 329.

Wang, M., Ning, X., Chen, A., et al. (2015). Impaired formation of homotypic cell-in-cell structures in human tumor cells lacking α-catenin expression. Sci. Rep. *5*, 12223.

Wang, M., Niu, Z., Qin, H., et al. (2020b). Mechanical ring interfaces between adherens junction and contractile actomyosin to coordinate entotic cell-in-cell formation. Cell Rep. *32*, 108071.

Wang, S., He, M., Chen, Y.H., et al. (2013). Rapid reuptake of granzyme B leads to emperitosis: an apoptotic cell-in-cell death of immune killer cells inside tumor cells. Cell Death Dis. *4*, e856.

Wang, X. (2015). Cell-in-cell phenomenon: a new paradigm in life sciences. Curr. Mol. Med. *15*, 810–818.

Zhang, X., Niu, Z., Qin, H., et al. (2019). Subtype-based prognostic analysis of cell-in-cell structures in early breast cancer. Front. Oncol. *9*, 895.

Zheng, Y., Zhou, L., Su, Y., et al. (2021). Cell fusion in the pathogenesis of COVID-19. Mil. Med. Res. *8*, 68.

Zhang, Z., Zheng, Y., Niu, Z., et al. (2021). SARS-CoV-2 spike protein dictates syncytium-mediated lymphocyte elimination. Cell Death Differ. *28*, 2765–2777.