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

Running title: Automatic identification for Cell-in-cell structures

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

[1] *Laboratory of Cell Engineering, Institute of Biotechnology; 20 Dongda Street, Beijing 100071, China.*

[2] *Department of GI Surgery, Beijing Shijitan Hospital of Capital Medical University, 10 TIEYI Road, Beijing 100038, China.*

[3] *Department of Oncology, Beijing Shijitan Hospital of Capital Medical University, 10 TIEYI Road, Beijing 10038, China.*

[4] *School of Mathematical Sciences, Peking University, Beijing 100871, China.*

[5] *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.*

These authors contributed equally: Meng Tang, Yan Su, Wei Zhao, Zubiao Niu

Correspondence:

Hongyan Huang
Email: hhongy1999@126.com

Hanping Shi
Email: shihp@ccmu.edu.cn

Qiang Sun
Email: sunq@bmi.ac.cn

Abstract

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

Key words: cell-in-cell structure; artificial intelligence; AIM-CICs; cell death; entosis; convolutional neural network

48 ## Introduction

49 Cell-in-cell structures (CICs) typically referred to the unusual eukaryotic cells involving the whole
50 objects internalized partially or completely inside of others, which had been observed in diverse
51 physiological and pathological samples [1, 2]. The presence of CICs was reported to be correlated
52 with patient prognosis in a group of human tumors, such as breast cancer [3], head and neck
53 squamous carcinoma [4, 5], and pancreatic ductal adenocarcinoma [6]. Functional studies
54 implicated CICs in a number of biomedical processes, including embryonic development [7],
55 mitotic surveillance [8], tumor evolution [9], and immune homeostasis [10] and the forth. As an
56 evolutionarily conserved process, CICs formation was underlain by multiple mechanisms, such as
57 entosis [11], cannibalism [12] and emperitosis [13]. Among which, entosis was one of the best
58 studied processes that generally ended up with the death of the internalized cells in an acidified
59 lysosome-dependent way [11, 14]. The formation of entotic CICs turned out to be a genetically
60 controlled process, where cell internalization was driven cell-autonomously by polarized
61 actomyosin resulted from the E-cadherin-mediated adherens junctions [15, 16], and coordinated
62 by a mechanical ring interfacing in between them [17]. Additionally, an ever-expanding set of
63 factors, acting through either actomyosin, or adherent junctions or mechanical ring, were
64 identified as important regulators [18, 19, 20, 21].

65 Despite great progress made over the past decade, the studies on CICs formation were,
66 however, based on the over-simplified readout of CICs counts that was performed manually,
67 which is not only labor-intensive and time-consuming, but also sharply incompatible with the
68 complex CICs formation per se. First, since CICs formation is a dynamic process preceding

through sequential steps including cell-cell contact, penetration and closing [22], therefore, it generally gives rise to CICs at different stages displaying morphologies of partial or complete. Second, the CICs morphologies were further complicated by the involvement of multiple cells, which frequently resulted in structures of "cell-in-cell-in-cell" or even more. Third, due to personal experiences and preferences, the CICs judgment and inclusion-exclusion criteria for analysis varied from investigators to investigators, 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 CICs quantification reported results of less informative, hardly comparable and low-throughput, which calls for more efficient and informative ways for the quantification of CICs.

Recent years had witnessed the rapid development of image-based artificial intelligence (AI) technology in assisting biomedical practices. For example, by using a single convolutional neural networks (CNN) algorithm, Esteva *et al* demonstrated the classification of skin lesions in performance on par with all tested experts [23]. Lin *et al* 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 [24]. 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 [25], tumor cell malignancy [26], mitosis staging [27], and the like. The remarkable potentials in accuracy and efficiency make AI-based image analysis an ideal method for comprehensive and reliable CICs quantification.

In this study, based on RGB fluorescent microscopic images, we employed the deep CNN

algorithms (Faster-RCNN and ResNet) to evaluate a large amount of cell candidates with defined subtypes and trained a multiclassfier for the recognition of subdivided CICs, which was named as AIM-CICs abbreviated from Automatic Identification Method of Cell-In-Cell structures. The AIM-CICs exhibited a high level of sensitivity and specificity, as evidenced by AUC values of > 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 hold the promise of speeding up CICs-related studies, such as deciphering the molecular controls of CICs 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 set, and then performed inspections in the test set (Fig. 1). For an RGB-format image, the proposed system performs two consecutive steps. First, a Faster-RCNN [28] network with ResNet-50 [29] backbone was formulated to find the cell regions and extract the candidate patches. Second, each candidate, representing one cell or CICs, was classified by an ResNet-101 network based on the cellular morphology. Subsequently, those 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 red channel for membrane

113 and blue channel for nucleus. Along with the bright field, the merged images could be further
114 composited into RGB format with variant cell quantities and brightness values (Fig. S1a). The
115 extraction of cell candidate aims to propose regions of interest (ROI) that potentially involved CIC
116 structures. This step served to reduce the searching space and improve efficiency of subsequent
117 steps in a high-content study. Initially, four pieces of MCF7 images and four pieces of MCF10A
118 images, which included 2164 cells in total, were used as the training set for cell region detection.
119 Through manually annotating these images using VGG Image Annotator (VIA,
120 https://www.robots.ox.ac.uk/~vgg/software/via/) (Fig. 2a), cell region detection was further treated
121 as a classic 1-class object detection task through the Faster-RCNN [28] network with ResNet-50
122 [29] backbone. Specifically, during training, we have performed random flip, random rotation and
123 random scale for data augmentation, which greatly expanded the data's diversity. Following the
124 training process (Fig. 2b), we ensured the applicability of this step with average of 0.88 precision
125 and 0.96 recall (IoU 0.1) by randomly testing on 10 pieces of MCF7 and MCF10A images, which
126 covered 2398 cells (Fig. 2c).
127 It is believed that factors, such as cell morphology, sample density, as well as image brightness, do
128 impact the accuracy of target detection and recognition. In the data collected this study, MCF10A
129 samples generally displayed a larger cell size and much more complicated pattern of CICs as
130 compared with MCF7 samples (Fig. S1a). Based on the precisely manual labeling, we could
131 minimize the effect of target varieties among MCF7 and MCF10A samples (Fig. S1b), except for
132 the over-exposed fluorescent images that should be excluded in the processing of the primary
133 images. Eventually, we exported the patches of detected cell regions of the entire RGB-format
134 images for the following analysis.

135 **Definition of the structural subtypes of CICs**

136 To classify the cell-in-cell structures, we first divided the traditional CICs into five structural
137 subtypes, including (a) partial, with more than 30% of the internalizing cells were enclosed, but
138 not fully, by the outer cells; (b) one-in-one, with only one cell fully internalized, (c) two-in-one,
139 with two cells were fully internalized; (d) in turn, a nested CICs with multiple cells sequentially
140 internalized into neighboring cells; (e) complicated, a complex CICs generated by four or more
141 cells (Fig. 3a). Considering the potential complexity, two kinds of breast cell lines including
142 MCF7 and MCF10A were investigated, in which the total rate of CICs and its subtypes showed
143 great discrepancy according to the manually labeling (Fig. 3b). In total, 17 pieces of MCF7 images
144 and 85 pieces of MCF10A images were enrolled in this study, the cell number of each image
145 ranged from 100 to 600, and from 30 to 200, respectively (Fig. 3c). The overall CICs rate of each
146 image counted from 1% to 85% (Fig. 3d).

147 **Multi-Subtype classification achieved by the AIM-CICs**

148 The obtained cell candidates were used to train ResNet101 model for the purpose of CICs
149 recognition (Fig. S2a). Practically, we used 13 pieces of MCF7 images and 32 pieces of MCF10A
150 images as the training set, which had 4026 MCF7 cells with a CICs rate of 11% and 3912
151 MCF10A cells with a CICs rate of 32% (Fig. S2b). Based on the morphological features of cell
152 candidates, five subtypes of CICs were manually labeled for each cell candidate in the training and
153 validation set. The distribution of each subtype of CICs showed remarkable discrepancy, as well
154 as in the test set (Fig. 4a-b). To improve the practicality of the model, we defined a F-category
155 from the non-CIC candidates. The F-category contains ambiguous structures that were hard to tell
156 their identities by both experienced experts and AI algorithm, therefore, were generally removed

157 from the sample counting (Fig S2c-d).

158 As shown in Fig. 4c, data training progressively increased the prediction accuracy to a
159 considerable level for each subtype. In both training and validation sets, the comprehensive
160 accuracy of integrated CICs (involving a, b, c, d, e types) and non-CIC type (including F category)
161 revealed approving performance (Fig. 4d). Moreover, the AIM-CICs also exhibited impressive
162 performance as indicated by the AUC of more than 0.97 for each CICs subtype (partial 0.9761,
163 one-in-one 0.9807, two-in-one 0.9872, in turn 0.9709, complicated 0.9984) (Fig. 4e-f) in the test
164 set. Additionally, for the low-quality images in the test set that displayed unclear cell regions and
165 were eventually removed for further analysis, their recognition also reached an ideal AUC of 0.99
166 (Fig. S2e). Together, the AIM-CICs performed accurate recognition of CICs on independent
167 datasets of MCF10A and MCF7 cells, suggesting the generalizability of this model.

168 **Visualization of morphological features and output**

169 To better understand what the model learnt from the annotated data, we extract features from the
170 output of network's global average pooling layer and applied t-SNE to reduce dimension to 2D for
171 visualization. For the training set, each group of cell samples represented independent clusters,
172 except for cell candidates in the circled region (Fig. 5a). Backtracking the training data identified
173 that these were candidates categorized into two subtypes due to erroneous manual annotation.
174 Thus, the t-SNE-based clustering would be a visualized way for error-correction in recognizing
175 CICs. For the test set (Fig. 5b), subtypes of CICs were clustered into close, but clearly distinct,
176 regions, whereas F-category was neighboring to the area of non-CIC as expected. Moreover,
177 following the comprehensive recognition under a specified confidence threshold, we were able to
178 accurately locate each structure with a predicted value on the original images (Fig. 5c).

179 **Application of AIM-CICs in an experimental setup**

180 To explore the potential implication of AI-based recognition of CICs in a biological context, we
181 included a functional experiment as an example of subtype profiling. In this analysis, the
182 confidence threshold was set to 0.2 for more informative identification (Fig. 6a). As the results
183 showed, though all of the three truncations of ARHGAP36, a molecule identified to be a regulator
184 of CICs formation in a screening study [20], resulted in impaired formation of CICs, the
185 alterations of CICs subtypes were rather different (Fig. 6b-c). While the truncated GAP36 (1-194)
186 had little impact on the formation of partial CICs (Fig. 6b-d), the majority of CICs were in
187 completed form (including all CICs subtypes except for the partial) in cells expressing the
188 truncated GAP (118-194) or GAP (195-395) (Fig. 6b-c), suggesting that the N-terminal region
189 (1-117) of ARHGAP36 might function to slow down the process of cell internalization.
190 Meanwhile, the C-terminal region of ARHGAP36 was likely to be responsible for the closing step
191 of CICs formation as evidenced by comparable formation of completed CICs between control and
192 GAP (195-395)-expressing cells (Fig. 6b and 6e-h). Moreover, the GAP (118-194) seemed to be
193 the major region that drives cell internalization as it promoted the formation of completed CICs at
194 a rate comparable to the GAP (1-194) region. Furthermore, though the N-terminal region might
195 negatively regulate the speed of CICs formation, it did function positively to promote cell
196 internalization as its truncation significantly reduced the formation of both partial and total CICs
197 (Fig. 6b-d). Thus, the AIM-CICs algorithm allows us, for the first time, to accurately dissect the
198 impacts of different domains or molecules on CICs formation in a heretofore underappreciated
199 resolution.

200 **Discussion**

Fluorescent microscope images recorded the cellular structures such as CICs, but inevitably provided a great number of morphological variations. To provide recognition with sufficient accuracy and potentially featured insights, we, for the first time, explored the application of Convolution Neural Network (CNN) in the profiling of subtyped CICs formed during entosis, a non-apoptotic cell death process occurred via cell-in-cell invasion [11]. Based on the tons of images accumulated from previous studies, we developed the AI-based identification algorithm AIM-CICs, which was trained with distinct illumination, textures, and density, in order to deliver an optimal performance in cell region detection and multiple subtypes classification, despite of the unseen perturbations.

In the proposed system, we set up two tasks, of which, a classic 1-class object detection model was formulated to find cell regions as the first task, followed by multi-class object recognition as the second task. Comparing 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 will achieve the flexibility for the raw samples to be recategorized and repurposed. In AIM-CICs developed in this study, the second task included a well-trained 7-category classifier (5 CICs subtypes plus one non-CICs and one F-category) to define the multiple subtypes of CIC structures, which is compatible with the cell candidates from the first step. 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 which could utilize a shared feature extraction backbone.

Among all the defined cell death programs, CICs-mediated death is unique in that it can only be

223 accomplished with the involvement of at least two cells, but not one cell in other programs like
224 apoptosis, necrosis and the forth [30]. Therefore, mechanistic study is a challenging task for the
225 field of CICs-mediated death, which was further complicated by the fact of lacking a reliable
226 biochemical marker. Current studies on CICs relayed on the morphology-based binary
227 quantification, that is, CICs or non-CICs. Here, CICs were usually defined as structures with more
228 than 1/2, or 2/3 in some studies, of the inner cell body being internalized/enclosed by the outer cell.
229 This oversimplified quantification of CICs, did move the field forward over the past decade,
230 however, provided rather coarse information over a more complicated process [22, 31]. CICs
231 formation is a stepwise process that could be empirically subdivided into three major stages: 1) the
232 early initiation stage from cell-cell contact to about 1/3 of the inner cell body being internalized,
233 this stage was primarily driven by cell-cell adhesion and assisted by cytoskeleton remodeling; 2)
234 the middle internalization stage covering the whole process of cell internalization that was
235 primarily driven by active actomyosin contraction within the inner cells, and coordinately assisted
236 by the outer cells; 3) the final closing stage that may involve in tail cutting and membrane fusion,
237 this a process rarely being investigated largely because it is technically challenging. Furthermore,
238 CICs formation is a dynamic process that may have multiple cells, either sequentially or
239 simultaneously, form a complicated structure that may contain more than one cell inside (Fig. 3a).
240 The regulation of this feature is completely unknown for the field yet, but might be conceptually
241 feasible as it was reported in phagocytosis that the number of corpses engulfed by a phagocyte
242 was genetically controlled [32]. Taking these two factors (stage and cell number) into account will
243 produce an even higher dimensional complexity, which, however, was missed from the traditional
244 analysis by the binary quantification. The implantation of AIM-CICs enables us to make a more

245 sophisticated description of CICs phenotypes, which would help identify finer molecular control.
246 For example, though expression of the GAP (195-395) domain did not influence the frequency of
247 simple CICs, where only one cell was enclosed (one-in-one in Fig. 6e), it did result in significantly
248 reduced formation of complex CICs, where more cells were enclosed by one outer cell
249 (two-in-one and in-turn in Fig. 6f-g). These results suggested that the truncated N-terminal domain
250 may facilitate the internalization of multiple cells to form complex CICs, which warrants further
251 functional validation.

252 In addition to mechanistic investigation, AIM-CICs is also promising in enabling high-content
253 based screening for therapeutic compounds that target CICs formation considering their pivotal
254 roles in multiple biomedical processes such as cancer [1]. While high throughput screening
255 generally relies critically on a reliable biochemical marker that is currently unavailable to CICs
256 formation, the related systemic screening, which would be labor-intensive and time-consuming if
257 worked out by manual annotation, had yet been reported. Empowered with AIM-CICs and
258 high-content microscopy, the systemic screening would be feasible in the near future.

259 **Acknowledgements**

260 We thank Dr Lulin Zhou, Xiaoyi Jiang, He Ren, Yichao Zhu, Yuqi Wang, Lihua Gao, Zhaolie
261 Chen, and the members of the Sun lab for the constructive discussions and assisting manual
262 labeling.

263 **Funding statement**

264 This work was supported by Beijing Municipal Natural Science Foundation (KZ202110025029 to
265 HYH), the National Natural Science Foundation of China (31970685 to QS), and Beijing

266 Municipal Administration of Hospitals Incubating Program (PX2021033 to HYH), and Youth
267 Science Foundation of Beijing Shijitan Hospital of CMU (No. 2019-q02 to MT).

268 Author contributions

269 Concept and design: QS; CNN Model Training and Analysis: WZ; Data collection: MT, ZBN,
270 CXW, and BZR; Data interpretation: QS, WZ and MT; Figures: MT, YS, WZ and QS; Manuscript:
271 MT, QS and WZ, with input from ZBN, CXW, BZR, YZ, BZ, QQL, YZ, HYH, FXZ, and HPS.
272 Funding acquisition: QS, HYH, HPS, and MT. All authors have read and approved the final
273 manuscript.

274 Declaration of interests

275 The authors declare no competing interests.

276 Ethics statement

277 This study did not involve ethical approval.

278 Materials and methods

279 Image processing and softwares

280 An entire dataset involving 17 pieces of MCF7 images and 85 pieces of MCF10A images were
281 obtained from Sun's lab. As detailed protocol described [33], the fluorescently labeled cells were
282 necessary to be stained with discrepant colors for each cell components, such as, red for
283 cytomembrane (E-Cadherin, 1:200, BD Biosciences, 610181) with secondary antibodies Alexa
284 Fluor 568 anti-rabbit (1:500; Invitrogen; A11036), and blue for cytoblast (DAPI, Sigma D8417).
285 Random fields were taken under corresponding channels of laser lights through fluorescent

286 microscopy (Nikon Ti-E microscope, Nikon NIS-Elements AR 4.5 software), along with bright
287 color for the background. For algorithm performing, each sample with three single-channel images
288 was transformed into an RGB format with value rescaled to 0 - 255. Softwares used and
289 algorithms developed in this study include: Python (http://www.python.org./); PyTorch
290 (https://pytorch.org/); VIA Annotation Tools (https://www.robots.ox.ac.uk/~vgg/software/via/);
291 Detectron2 (https://github.com/facebookresearch/detectron2).
292

293 **Cell region labeling and candidates extraction**

294 After acquiring the processed images, we manually annotated the cell regions through VGG Image
295 Annotator (https://www.robots.ox.ac.uk/~vgg/software/via/). Based on the annotated images, a
296 classic 1-class object detection task was carried out for cellular morphological learning. The
297 model we used is a Faster-RCNN [28] network with ResNet-50 [29] backbone. Since the original
298 resolution of microscopic image is 2160 x 2560 which is too large for Faster-RCNN training, we
299 first split each image into 4x4 grids, then follow the common practice to train the model. For data
300 augmentation, we use random flip, random rotation, and random scale to expand diversity of data.
301 As for other hyper-parameters, we set batch-size to 24 and iterate 50000 steps using SGD
302 optimizer with momentum 0.9. As the output of the Faster-RCNN network, the patches of detected
303 cell regions were exported as candidate sequences for further steps.
304

305 **Manual classifications of cell-in-cell structures**

306 The manual definition of cell-in-cell structural classification primarily included bipartite-class,
307 CICs and non-CICs. CICs category were further subdivided into 5 subtypes, including (a) partial,

308 with more than 30% of the internalizing cells were enclosed, but not fully, by the outer cells; (b)
309 one-in-one, with only one cell fully internalized, (c) two-in-one, with two cells were fully
310 internalized; (d) in turn, a nested CICs with multiple cells sequentially internalized into
311 neighboring cells; (e) complicated, a complex CICs generated by four or more cells. To refine the
312 output results, we added a F-category among non-CICs, which was defined as unclear or not sure
313 for the cell recognition and needs to be removed for the quantitative analysis. The cell candidates
314 involved in the training set were verified together by an expert group consisted of 6 members in
315 the lab.
316
317 **Multiple classification model**
318 We used the ResNet101 model as our classifier and the input size was set to 224. Since this model
319 could take the detection model's output as input, we cropped cell samples using detection model
320 and manually labeled them with corresponding cell types. During training, each sample was first
321 padded to square and then resize to 224 x 224. Both horizontal and vertical random flip were
322 performed. We trained our model for 250 epochs with batch size of 32, using SGD optimizer with
323 learning rate of 0.001 and momentum of 0.9. To alleviate overfitting, a dropout layer with $p = 0.25$
324 was set right before feature went into the final fully connected layers. To choose hyper-params, we
325 kept 20% samples as validation set. Eventually, the prediction results could be visualized on the
326 original image with detected cell region and a predictive score of CICs, as well as in the output
327 folders of each cell type.
328 Importantly, when applying our model for inference, the test samples should be pad and resize in
329 the same way as training. Our model is a 7 classes classifier, and it outputs a 7 elements vector

representing the probability for the test sample to belong to each type. Traditionally, the predicted type should be the type with maximum probabilities. In practice, to increase precision, we predict cells that have predicted probability lower than 0.2 as non-CIC, even if the non-CIC probability is not the maximum for it. For example, if the predicted output is [0.1,0.18,0.12,0.15,0.15,0.13,0.17] (for a, b, c, d, e, f, non-CIC), we will use non-CIC as model's prediction. Ultimately, we could output the classifications into specific folders of each cell type, and obtain the visualized results that marked with individual colors on the original image.

Performance analysis of detection model

In deep learning community, the most common metric used for quantitatively comparing detection models' performance is mean average precision (mAP), as proposed in [34] and [35]. However, since our work mainly focused on multi-type CICs classification instead of general object detection technique, we reported our detection result in a more practical recall/precision manner. In detail, we kept detection model's output instances with confidence > 0.1 as model's prediction and calculate metrics at two different Intersection over Union (IoU) thresholds 0.5 / 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.

Features visualization

To better understand what the classification model learns from labeled samples, we extracted features from each cell sample and visualize them in a 2D space. The feature we used is the output of network's global average pooling layer, which is right before the final classification layer. This

2048D feature is the deepest and the most semantic so it can represent the information extracted by the network from a corresponding input image. To visualize these 2048D features, we uses the t-SNE method for dimensionality reduction to transform each feature to 2D [36]. 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 models

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

(1) Se (recall) = TP / (TP+FN)

(2) Sp = TN / (TN+FP)

(3) Precision = TP / (TP+FP)

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 correctly recognized as negative CICs.

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 $p < 0.05$.

References

1. Fais S, Overholtzer M. Cell-in-cell phenomena in cancer. *Nat Rev Cancer* 2018, **18**(12)**:** 758-766.
2. Wang X. Cell-in-cell phenomenon: A New Paradigm in Life Sciences. *Current molecular medicine* 2015, **15**(9)**:** 810-818.
3. Zhang X, Niu Z, Qin H, Fan J, Wang M, Zhang B*, et al.* Subtype-Based Prognostic Analysis of Cell-in-Cell Structures in Early Breast Cancer. *Front Oncol* 2019, **9:** 1-12.
4. Schwegler M, Wirsing AM, Schenker HM, Ott L, Ries JM, Buttner-Herold M*, et al.* Prognostic Value of Homotypic Cell Internalization by Nonprofessional Phagocytic Cancer Cells. *BioMed research international* 2015, **2015:** 359392.
5. Fan J, Fang Q, Yang Y, Cui M, Zhao M, Qi J*, et al.* Role of Heterotypic Neutrophil-in-Tumor Structure in the Prognosis of Patients With Buccal Mucosa Squamous Cell Carcinoma. *Front Oncol* 2020, **10**(2284)**:** 541878.
6. Huang H, He M, Zhang Y, Zhang B, Niu Z, Zheng Y*, et al.* 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* 2020, **5**(1)**:** 246-248.
7. Lee Y, Hamann JC, Pellegrino M, Durgan J, Domart MC, Collinson LM*, et al.* Entosis Controls a Developmental Cell Clearance in C. elegans. *Cell Rep* 2019, **26**(12)**:** 3212-3220 e3214.
8. Liang J, Niu Z, Zhang B, Yu X, Zheng Y, Wang C*, et al.* p53-dependent elimination of aneuploid mitotic offspring by entosis. *Cell Death Differ* 2020, 10.1038/s41418-020-00645-3.
9. Sun Q, Luo T, Ren Y, Florey O, Shirasawa S, Sasazuki T*, et al.* Competition between human cells by entosis. *Cell Res* 2014, **24:** 1299-1310.
10. Davies SP, Reynolds GM, Wilkinson AL, Li X, Rose R, Leekha M*, et al.* Hepatocytes Delete Regulatory T Cells by Enclysis, a CD4(+) T Cell Engulfment Process. *Cell Rep* 2019, **29**(6)**:** 1610-1620.e1614.
11. Overholtzer M, Mailleux AA, Mouneimne G, Normand G, Schnitt SJ, King RW*, et al.* A nonapoptotic cell death process, entosis, that occurs by cell-in-cell invasion. *Cell* 2007, **131**(5)**:** 966-979.
12. Sharma N, Dey P. Cell cannibalism and cancer. *Diagn Cytopathol* 2011, **39**(3)**:** 229-233.
13. Wang S, He M, Chen YH, Wang MY, Yu XM, Bai J*, et al.* Rapid reuptake of granzyme B leads to emperitosis: an apoptotic cell-in-cell death of immune killer cells inside tumor cells. *Cell death & disease* 2013, **4**(10)**:** e856.
14. Su Y, Ren H, Tang M, Zheng Y, Zhang B, Wang C*, et al.* Role and dynamics of vacuolar pH during cell-in-cell mediated death. *Cell death & disease* 2021, **12**(1)**:** 119.
15. Sun Q, Cibas ES, Huang H, Hodgson L, Overholtzer M. Induction of entosis by epithelial cadherin expression. *Cell Res* 2014, **24**(11)**:** 1288-1298.
16. Wang M, Ning X, Chen A, Huang H, Ni C, Zhou C*, et al.* Impaired formation of homotypic cell-in-cell structures in human tumor cells lacking alpha-catenin expression. *Scientific reports* 2015, **5:** 12223.
17. Wang M, Niu Z, Qin H, Ruan B, Zheng Y, Ning X*, et al.* Mechanical Ring Interfaces between Adherens Junction and Contractile Actomyosin to Coordinate Entotic Cell-in-Cell Formation. *Cell Rep* 2020, **32**(8)**:** 108071.

18. Wang C, Chen A, Ruan B, Niu Z, Su Y, Qin H*, et al.* PCDH7 Inhibits the Formation of Homotypic Cell-in-Cell Structure. *Front Cell Dev Biol* 2020, **8**(329)**:** 1-12.
19. Liang J, Fan J, Wang M, Niu Z, Zhang Z, Yuan L*, et al.* CDKN2A inhibits formation of homotypic cell-in-cell structures. *Oncogenesis* 2018, **7**(6)**:** 1-8.
20. Ruan B, Wang C, Chen A, Liang J, Niu Z, Zheng Y*, et al.* Expression profiling identified IL-8 as a regulator of homotypic cell-in-cell formation. *BMB Rep* 2018, **51**(8)**:** 412-417.
21. Hinojosa LS, Holst M, Baarlink C, Grosse R. MRTF transcription and Ezrin-dependent plasma membrane blebbing are required for entotic invasion. *J Cell Biol* 2017, **216**(10)**:** 3087-3095.
22. Niu Z, He M, Sun Q. Molecular mechanisms underlying cell-in-cell formation: Core machineries and beyond. *J Mol Cell Biol* 2021.
23. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM*, et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017, **542**(7639)**:** 115-118.
24. Lin C, Zhao G, Yin A, Yang Z, Guo L, Chen H*, et al.* A novel chromosome cluster types identification method using ResNeXt WSL model. *Med Image Anal* 2020, **69:** 101943.
25. Waisman A, La Greca A, Mobbs AM, Scarafia MA, Santin Velazque NL, Neiman G*, et al.* Deep Learning Neural Networks Highly Predict Very Early Onset of Pluripotent Stem Cell Differentiation. *Stem Cell Reports* 2019, **12**(4)**:** 845-859.
26. Oei RW, Hou G, Liu F, Zhong J, Zhang J, An Z*, et al.* Convolutional neural network for cell classification using microscope images of intracellular actin networks. *PLoS One* 2019, **14**(3)**:** e0213626.
27. Mao Y, Han L, Yin Z. Cell mitosis event analysis in phase contrast microscopy images using deep learning. *Med Image Anal* 2019, **57:** 32-43.
28. Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2015, **39**(6)**:** In: NIPS. (2015).
29. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision & Pattern Recognition* 2016**:** In: CVPR. (2016).
30. Galluzzi L, Vitale I, Aaronson SA, Abrams JM, Adam D, Agostinis P*, et al.* Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ* 2018, **25**(3)**:** 486-541.
31. Rizzotto D, Villunger A. P53 clears aneuploid cells by entosis. *Cell Death Differ* 2020, 10.1038/s41418-020-00659-x.
32. Park D, Han CZ, Elliott MR, Kinchen JM, Trampont PC, Das S*, et al.* Continued clearance of apoptotic cells critically depends on the phagocyte Ucp2 protein. *Nature* 2011, **477**(7363)**:** 220-224.
33. Sun Q, Overholtzer M. Methods for the study of entosis. *Methods Mol Biol* 2013, **1004:** 59-66.
34. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision* 2009, **88**(2)**:** 303-338.
35. Lin TY, Maire M, Belongie S, Hays J, Zitnick CL. Microsoft COCO: Common Objects in Context. *European Conference on Computer Vision* 2014**:** In: ECCV. (2014).
36. Maaten. Lvd, G. H. Visualizing data using t-SNE. *Journal of Machine Learning Research* 2008, **9:** 2579-2605.

466 ## Figure legends

467 **Fig. 1 The workflow for the development of AIM-CICs**.

468

469 **Fig. 2 Cell regions detection and extraction**. **a** Using VGG Image Annotator to annotate cell regions
470 in the training set. Each box represents an individual cell region with numeric order on it. **b** Detection
471 Model-based prediction of the cell region in the validation set. Each cell region was indicated by a box
472 with a predicted confidence score. **c** Performance of cell region detection model. The MCF7 and
473 MCF10A images in the validation set belong to the same batch of training set images, and test set is
474 composed of MCF7 and MCF10A images from independent experiments. N_GT: number of ground
475 truth cells. N_prediction: cell number of model's prediction (confidence threshold set to 0.1). IoU:
476 Intersection over Union.

477

478 **Fig. 3 Image processing and cell candidate regions. a** Representative images of five CICs subtypes.
479 Cell membrane: E-cadherin in red, nucleus: DAPI in blue, and background is green. Original images
480 are listed in supplementary Fig. S1a. **b** Percentage of different CICs subtypes for MCF7 and MCF10A
481 cells. **c** The number of cell candidates extracted from each MCF7 and MCF10A image. Samples
482 contained 17 pieces of MCF7 images and 85 pieces of MCF10A images. The columns in orange
483 and purple represent for samples used in the training set. **d** The count and frequency of CICs for
484 individual MCF7 or MCF10A image.

485

486 **Fig. 4 Training and testing process of the multiple classification. a,b** The percentages of CICs
487 subtypes in the training set (**a**) and test set (**b**). The CICs counts of MCF7 were 437 (**a**) and 340 (**b**),

488 respectively. In the MCF10A samples, the CIC counts were 1269 (**a**) and 1948 (**b**), respectively.
489 Supplementary figures associated: Supplementary Fig. S2. **c** The prediction accuracy of AIM-CICS for
490 each subtype in a 250 epochs of learning process. **d** The integrated accuracy of AIM-CICS for CICs
491 and non-CIC type in the training and validation set in a 250 epochs of learning process. CICs included
492 the partial, one-in-one, two-in-one, in turn, and complicated type. Non-CIC referred to F-category and
493 non-CIC. **e** Representative images of each CICs subtype predicted in the training set. **f** The ROC curves
494 for each CICs subtype in the test set.
495
496 **Fig. 5 Visualization of sample features and output. a,b** The two-dimension visualization of CICs
497 subtypes in the training set (**a**) and the test set (**b**). The samples circled out were those predicted by the
498 AIM-CICs to be miscategorized manually. **c** A representative image showing the recognition result of
499 AIM-CICs. Colored Frames indicate structures in different categories, the predicated confidence scores
500 were marked on the up-left corner of each structure. The structures in the right were cropped from the
501 image in the left.
502
503 **Fig. 6 Analysis of CICs subtypes in an experimental setup by AIM-CICs. a** The visualized
504 recognition results of AIM-CICs in an experiment, where MCF10A cells expressed the empty vector
505 (control) and three truncated mutants of ARHGAP36 (1-194, 118-194, 195-395), respectively. **b,c**
506 Graphs show the absolute frequency (**b**) or relative frequency (**c**) of subtyped CICs in MCF10A cells
507 expressing different ARHGAP36 mutants. (n = 934 cells for control, 1060 cells for 1-194, 1392 cells
508 for 118-194, 852 cells for 195-395). **d-h** The frequencies of subtyped CICs in MCF10A cells
509 expressing different ARHGAP36 mutants. Data were shown as box-plots with means and individual

data points. * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$ (two-tailed Student's t test).

Fig. S1 Format transformation and cell region detection. a Representative images of different brightness in RGB format for MCF7 and MCF10A cells, respectively. Cell membrane: E-cadherin in red, nucleus: DAPI in blue, and background is green. The intensity value was rescaled to 0 - 255. **b,c** Representative images of MCF7 (**b**) and MCF10A (**c**) samples predicted by the cell region detection model. Each cell region was indicated by a box with a predicted confidence score.

Fig. S2 Binary subtyping by the AIM-CICs. a Representative images for CICs and non-CICs. **b** The quantification of CICs for MCF7 and MCF10A cells in the training set. **c** Percentage analysis of F-category relative to the non-CIC category in training set and test set. **d** Representative images of F-category, images belonging to this category were unclear and hard to be classified into other categories. **e** The ROC curves for the F-category in the entire dataset, MCF7 group, and MCF10A group of test set, respectively.

Figure 1

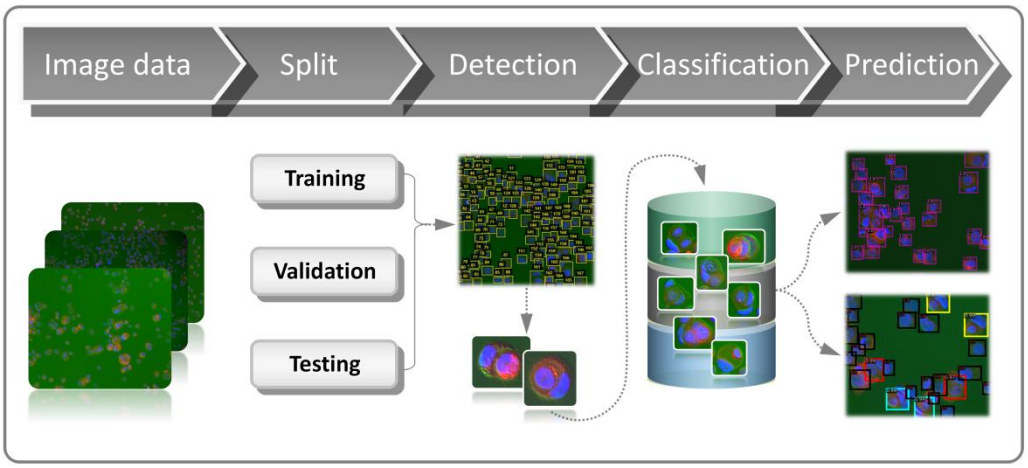
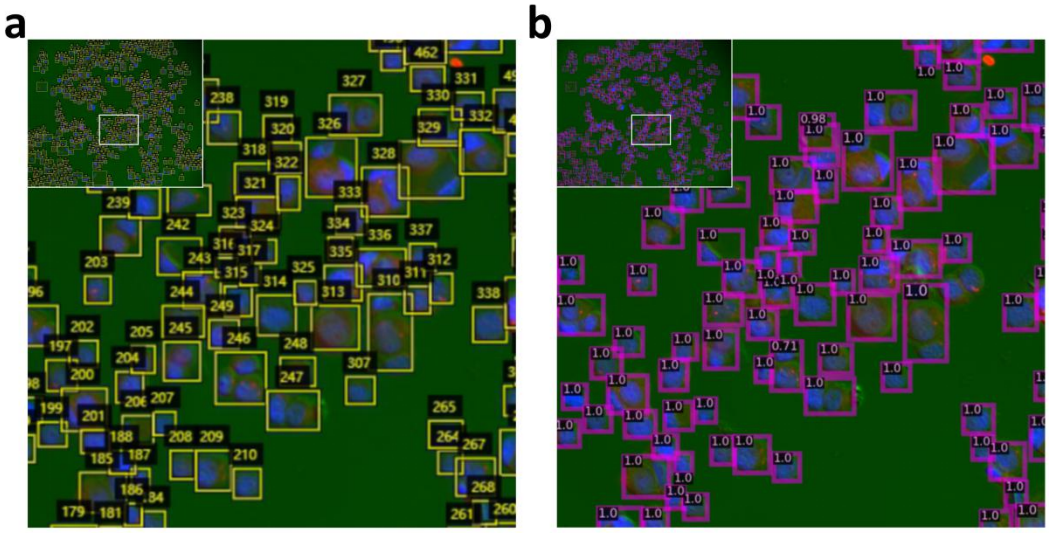


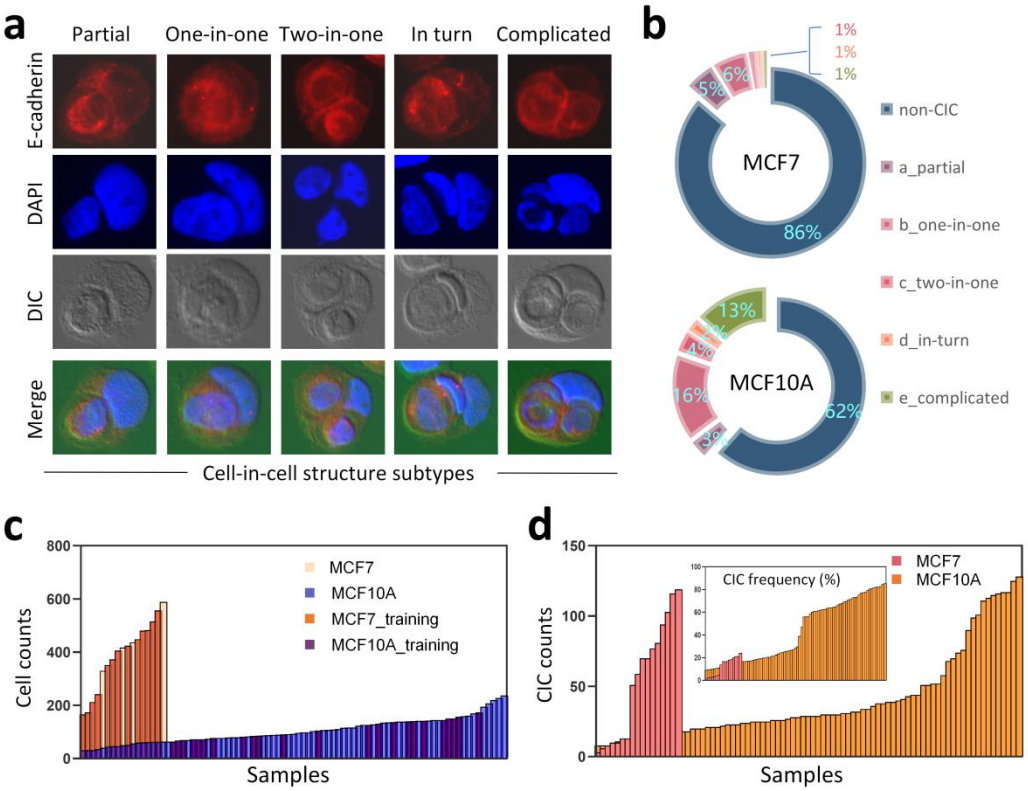
Figure 2



c

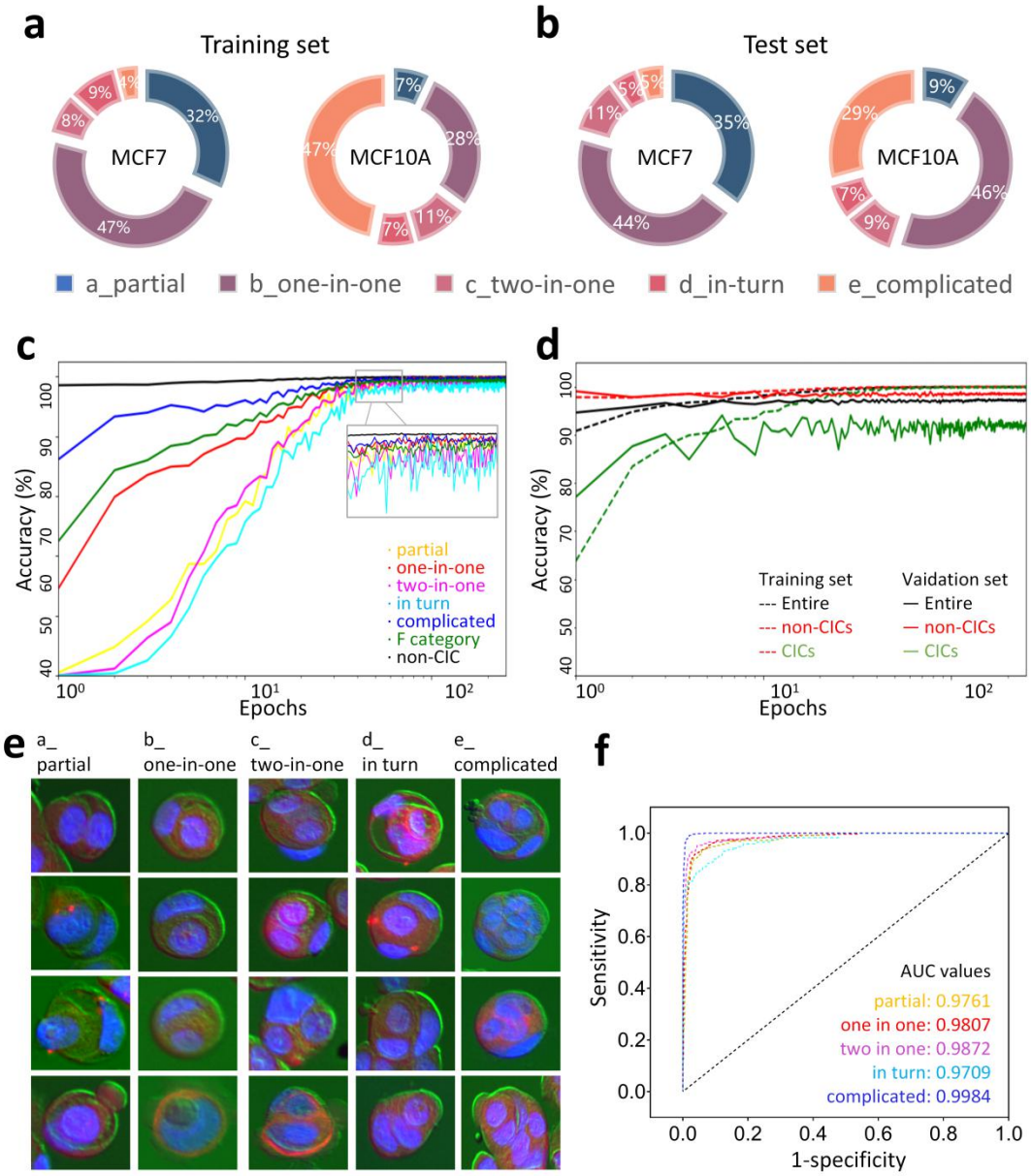
Group	N_GT	N_prediction	IoU 0.1			IoU 0.5		
			N_RECALL	RECALL	PRECISION	N_RECALL	RECALL	PRECISION
validation_MCF7	1783	1782	1691	0.947	0.947	1650	0.923	0.923
validation_MCF10A	347	367	336	0.968	0.904	333	0.960	0.897
testing	268	331	258	0.962	0.781	238	0.893	0.725

Figure 3


a
Partial
One-in-one
Two-in-one
In turn
Complicated
E-cadherin
DAPI
DIC
Merge
Cell-in-cell structure subtypes
b
1%
1%
1%
5%
6%
MCF7
86%
non-CIC
a_partial
b_one-in-one
c_two-in-one
d_in-turn
e_complicated
13%
16%
MCF10A
62%
c
Cell counts
MCF7
MCF10A
MCF7_training
MCF10A_training
Samples
d
CIC counts
CIC frequency (%)
MCF7
MCF10A
Samples

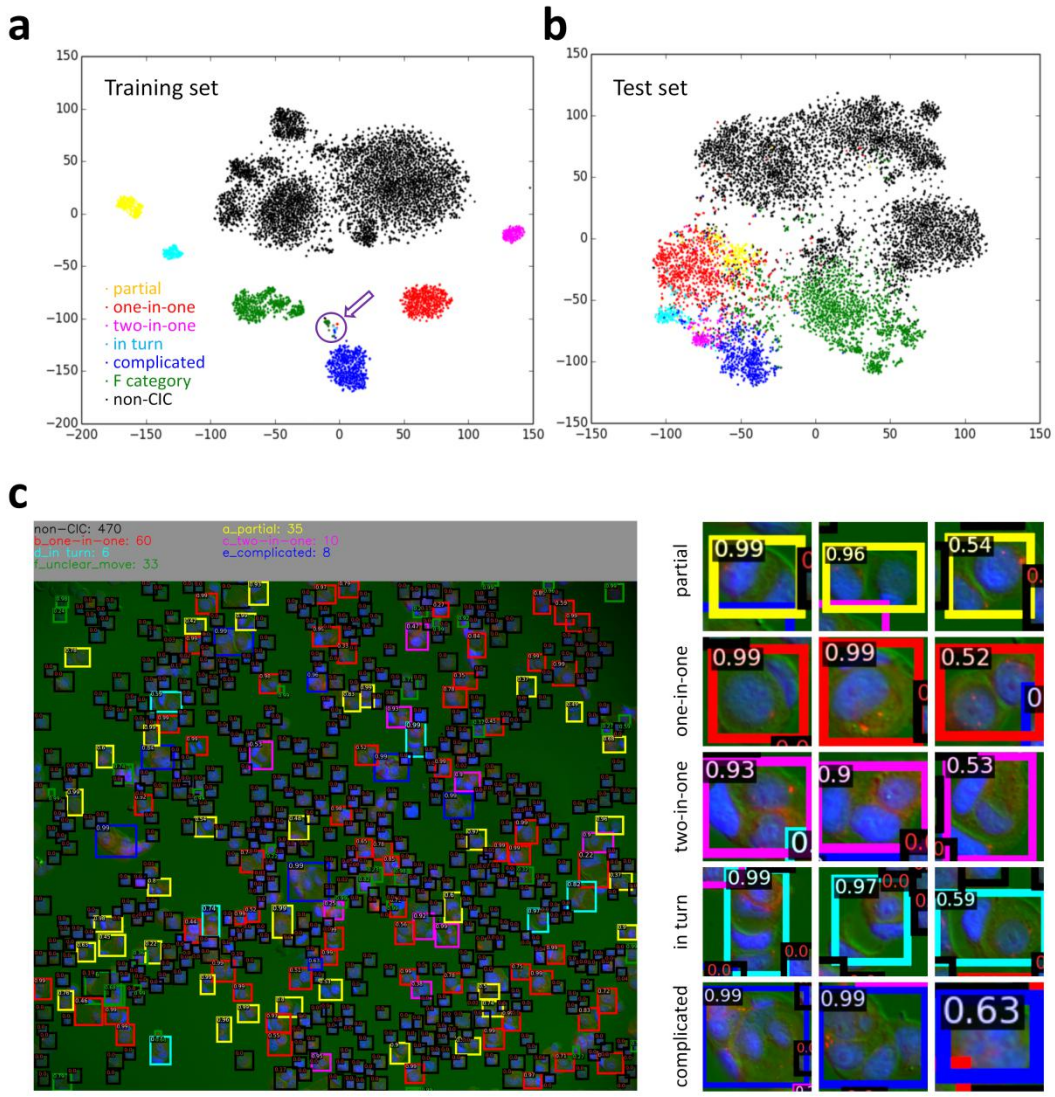
527

Figure 4


a
Training set
MCF7
32%
47%
8%
9%
4%
MCF10A
7%
28%
11%
7%
47%
b
Test set
MCF7
35%
44%
11%
5%
5%
MCF10A
9%
46%
9%
7%
29%
a_partial
b_one-in-one
c_two-in-one
d_in-turn
e_complicated
c
Accuracy (%)
Epochs
partial
one-in-one
two-in-one
in turn
complicated
F category
non-CIC
d
Accuracy (%)
Epochs
Training set
--- Entire
--- non-CICs
--- CICs
Vaidation set
— Entire
— non-CICs
— CICs
e
a_ partial
b_ one-in-one
c_ two-in-one
d_ in turn
e_ complicated
f
Sensitivity
1-specificity
AUC values
partial: 0.9761
one in one: 0.9807
two in one: 0.9872
in turn: 0.9709
complicated: 0.9984

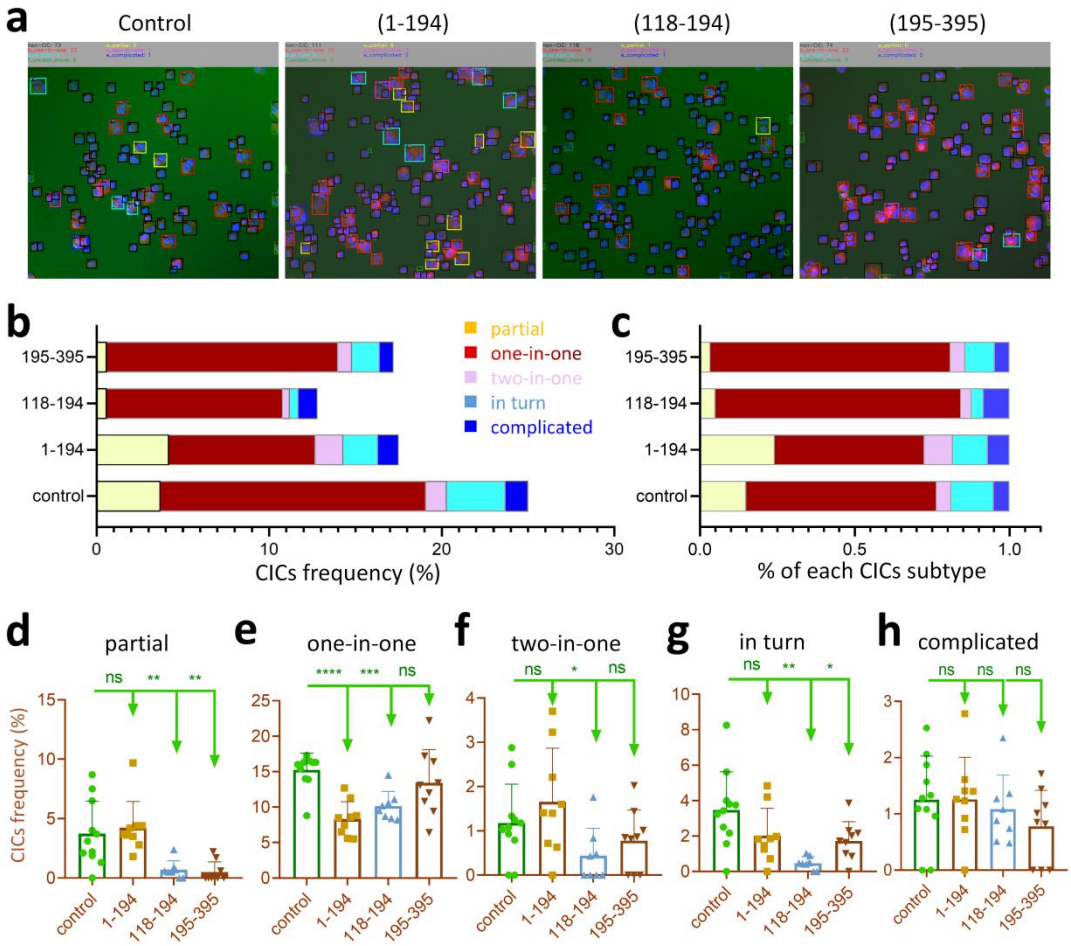
528

Figure 5


a
Training set
partial
one-in-one
two-in-one
in turn
complicated
F category
non-CIC
b
Test set
c
non-CIC: 470
b_one-in-one: 60
d_in turn: 6
f_unclear_move: 33
a_partial: 35
c_two-in-one: 10
e_complicated: 8
partial
one-in-one
two-in-one
in turn
complicated

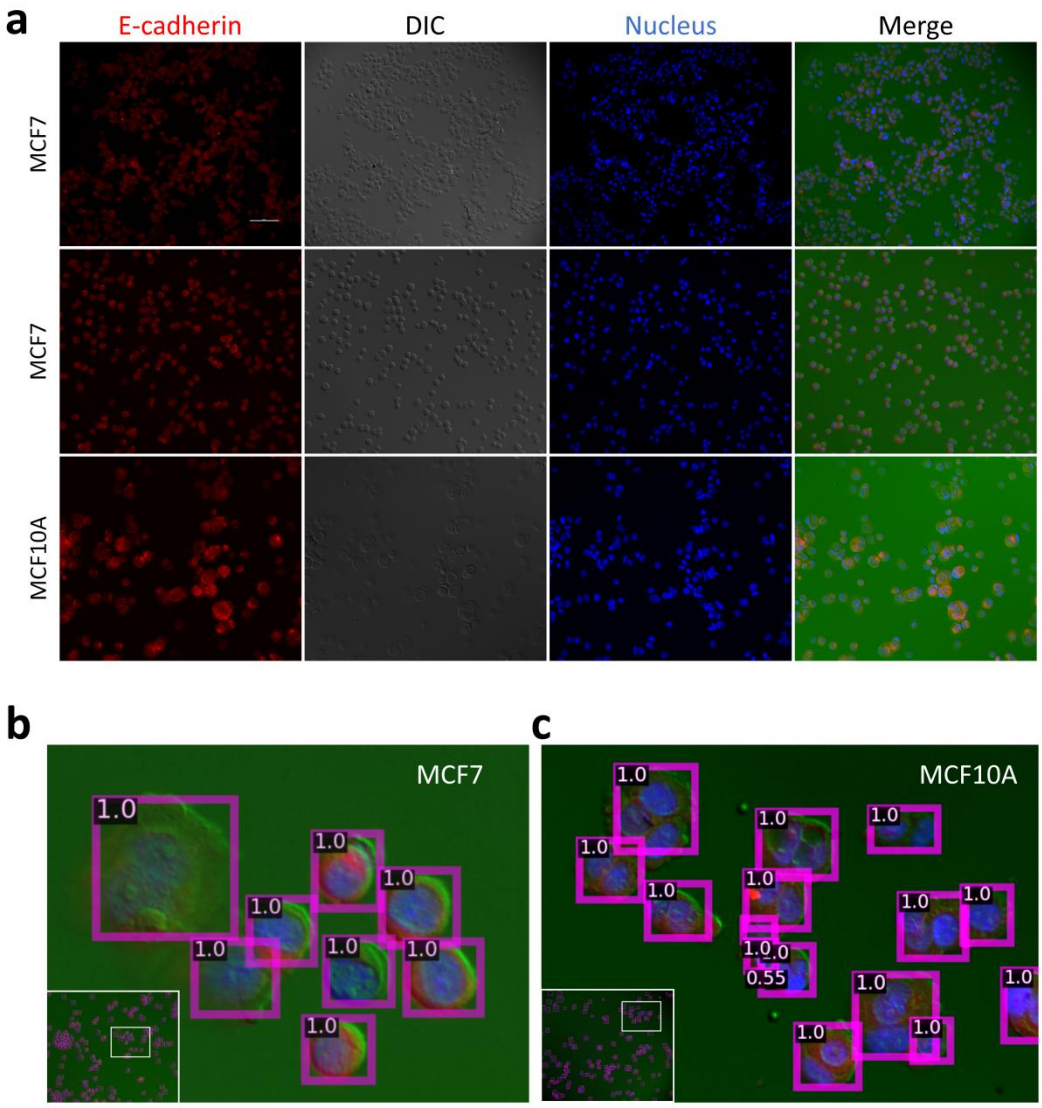
529

Figure 6


a
Control
(1-194)
(118-194)
(195-395)
b
195-395
118-194
1-194
control
partial
one-in-one
two-in-one
in turn
complicated
0
10
20
30
CICs frequency (%)
c
0.0
0.5
1.0
% of each CICs subtype
d
partial
e
one-in-one
f
two-in-one
g
in turn
h
complicated
CICs frequency (%)
ns
**
****
***
*
control
1-194
118-194
195-395

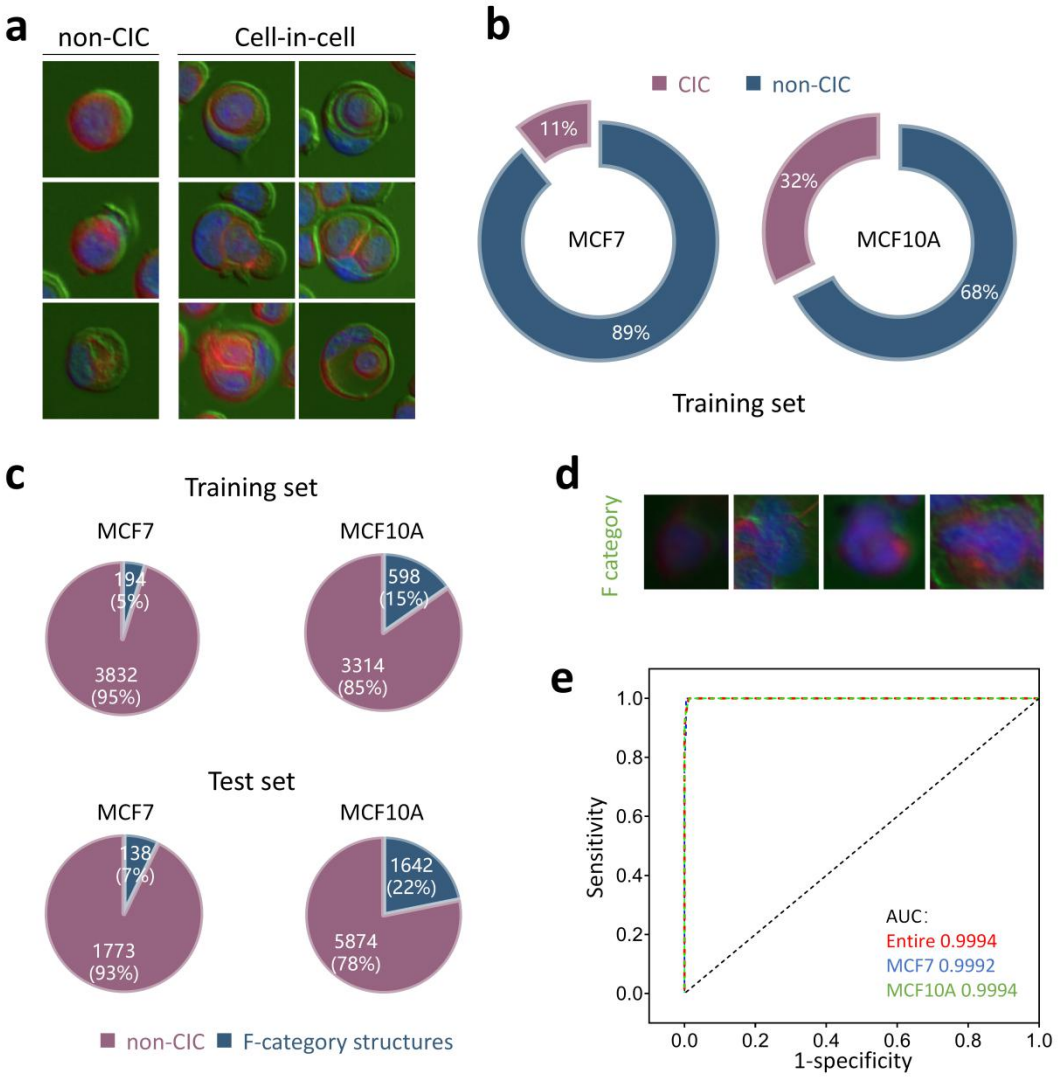
530

Figure S1



531

Figure S2



532