ChinaRxiv 中X

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-201711.02434

# Liang Can_Final Print After System Testing

**Authors:** Liang Can

**Date:** 2020-11-24T00:00:00+00:00

## Abstract

wewrew

## Full Text

## Preamble

**A Machine Learning Assisted, Label-free, Non-invasive Approach for Somatic Reprogramming in Induced Pluripotent Stem Cell Colony Formation Detection and Prediction**

Ke Fan[1],[2], , Sheng Zhang[1],[2], , Ying Zhang[1],[2], Jun Lu[1],[2], Mike Holcombe[3], , and Xiao Zhang[1],[2],*

[1]CAS Key Laboratory of Regenerative Biology, Joint School of Life Sciences, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China; Guangzhou Medical University, Guangzhou 511436, China
[2]Guangdong Provincial Key Laboratory of Biocomputing, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China
[3]Department of Computer Science, University of Sheffield, Sheffield, United Kingdom
epiGenesys, Sheffield, United Kingdom

*zhang xiao@gibh.ac.cn
These authors contributed equally to this work

## Abstract

During cellular reprogramming, the mesenchymal-to-epithelial transition is accompanied by morphological changes that occur prior to iPSC colony formation. Current approaches for detecting these morphological changes rely heavily on human expertise, which requires extensive training, is prone to human error,

offers limited quality control, and suffers from batch-to-batch variations. Here, we report a time-lapse-based bright-field imaging analysis system that enables a label-free, non-invasive method for measuring morphological dynamics. To automatically analyze and determine iPSC colony formation, we developed a machine learning-based classification, segmentation, and statistical modeling system to guide colony selection. The system can detect and monitor the earliest cellular texture changes after reprogramming induction in human somatic cells as early as day 7 in a 20–24 day process. Moreover, after quantitatively determining the reprogramming process and iPSC colony formation, we developed a mathematical model to statistically predict the optimal iPSC selection phase without requiring additional resources. All computational detection and prediction experiments were evaluated using a validation dataset, and biological verification was performed. These algorithm-detected colonies showed no significant differences (Pearson coefficient $r > 0.9$) in their biological features compared to manually processed colonies using standard molecular approaches.

## Introduction

Stem cells constitute a self-replenishing cell population whose primary function is to generate progeny that develop into terminally differentiated cell types. Tissue-specific adult stem cells or progenitors are committed to producing tissue- or lineage-specific cells, whereas pluripotent or totipotent stem cells have the potential to give rise to any of the 200+ cell types across all three germ layers. Two types of pluripotent stem cells are defined by their tissue origin: (i) embryonic stem (ES) cells obtained from early embryos, typically at the blastula stage, and (ii) induced pluripotent stem (iPS) cells derived through a reprogramming process that converts terminally differentiated somatic cells back to a pluripotent state. iPS technology offers a personalized approach that does not require embryos, thereby avoiding the ethical concerns associated with ES cells. Recently, iPSCs have been used to treat chronic and degenerative diseases, such as retinal degenerative blindness caused by retinitis pigmentosa[1]. Thus, clinical applications of iPS cells can be generated in personalized formats for cell replacement therapy. Moreover, there are strong demands for quality-assured iPSC lines from both the pharmaceutical industry and basic medical research organizations[2]. A non-invasive approach has been established to convert urinal cells into iPSCs, providing the most convenient cell source for reprogramming studies, and we implemented this method in the present study for data training and statistical modeling[3].

One major factor limiting iPSC quality and biological consistency in downstream applications is colony determination, which guides the isolation and purification of iPSC colonies during the cellular reprogramming process. The current solution relies on judgment calls from well-trained cell culture experts, though this training is time-consuming and costly. Inevitably, the natural instability of human recognition can lead to misjudgment, causing batch-to-batch variation. Quality control of iPSC colony determination is extremely important for down-

stream expansion and for maintaining a homogeneous culture of undifferentiated cells. Inconsistency in colony determination and selection leads to insufficient downstream differentiation into functional cells. Therefore, an automated quantitative method with consistent colony maturation determination would greatly assist biologists during iPSC formation.

The traditional approach for identifying and verifying iPSC colonies uses immunofluorescence staining or a reporter system to detect pluripotent markers, which can couple well with automated fluorescence microscopy to provide a dynamic and effective method , . However, this xeno probe-based labeling can only be applied during late-stage reprogramming and cannot be delayed further due to potential safety issues for downstream clinical applications , .

Furthermore, during cellular reprogramming studies, it has been observed that alongside positive iPSC colonies, negative colonies exhibit various morphological subtypes, particularly with respect to cellular polarity. Cellular polarity is strongly linked to the specificity of gene expression, cell cycle regulation, and other cellular processes, and may explain different reprogramming mechanisms . For example, cell polarity changes during mesenchymal-to-epithelial transition (MET) or EMT are linked to iPSC colony formation or tumorigenesis [11]. Hence, the reprogramming process has a strong connection to morphological changes, and classification of these morphologies can serve as a quantitative readout to identify iPSC colonies and monitor the reprogramming process. This approach provides richer information than binary fluorescence images and opens the door for a label-free, non-invasive method.

Here, we report a label-free, non-invasive, and consistently quantitative method for accurate human iPSC detection. This automated system employs a computer vision recognition model to assist in classifying iPSC colonies from bright-field microscopic images. The model utilizes a convolutional neural network as a classifier within a sliding windows framework for colony recognition. Subsequently, a semi-supervised segmentation method was applied to locate colonies and detect their boundaries. Moreover, a Hidden Markov Model (HMM) was trained to estimate the growth phase and maturation time window of colony formation during the reprogramming process. Finally, using time-lapse data from traced colonies, this system can predict the optimal selection time window for iPSC colonies to prevent random differentiation caused by overgrowth. Our results show that algorithm-detected and predicted colonies demonstrated no significant differences (Pearson coefficient $r > 0.9$) in biological features compared to manually processed colonies, as evaluated and characterized using standard immunofluorescence staining, quantitative polymerase chain reaction (qPCR), and RNA-Seq for pluripotency verification.

## Results

### Automated iPSC Colony Detection Using Computer Vision Algorithms Verified via OCT4-GFP Reporter Mice

During the reprogramming process, cellular polarization reshaping leads to morphological changes that indicate iPSC colony formation. We developed a novel computer vision-assisted method to process microscopic images. The sliding window approach was used to scan potential colony areas, generating binary detection images. Post-processing, including removal of fragmented areas, was performed using these binary images as templates, allowing determination of iPSC colony areas based on size and location (Figure 1B [Figure 1: see original paper]). Figure 1A shows examples of cropped positive and negative colony texture mosaic windows from algorithm training. The algorithm detects positive areas throughout entire wells from 6-well cell culture plates.

We assessed algorithm performance by verifying the biological signature of iPSCs using OG mice with an OCT4-GFP knock-in reporter mouse embryonic fibroblast (MEF) system. Activation of the OCT4 pluripotent gene indicates successful MEF reprogramming to iPSCs[12]. This biological fluorescence signal served as a positive feedback mechanism for algorithm training. We then used the trained classifier to generate isolated binary images, with fluorescence signal areas masked as overlays [Figure 1B]. Observations showed that nearly 10% of cells from the total population did not express OCT4-GFP and were therefore not selected by the algorithm. Quantitative analysis revealed a strong correlation between OCT4-GFP expression and algorithm overlay, with an average Pearson' s coefficient of $r = 0.877$. This r-value was determined from three independent reprogramming experiment datasets (additional results are shown in Supplemental Figures S2 and S3).

### Human iPSC Colony Detection Using Computer Vision Algorithms Verified via Pluripotency Biomarkers

Human iPSC regeneration opens the door to personalized cell sources for cell replacement therapy in regenerative medicine. This algorithm extends the non-invasive, label-free computer-assisted method to human urinal cell reprogramming[3]. Unlike the OG mouse system, OCT4-GFP is not applicable for urinal cells, so we built the training dataset using manual cropping performed by well-trained cell biologists (Figure 2A). As described previously, our iPSC detection framework employs a mosaic sliding window-based classification process. If a small patch was located within a positive iPSC colony texture, the area was masked out by the overlay; otherwise, it was classified as a negative patch. The training session utilized numerous manually cropped samples, followed by annotation of positive colonies, which indicated region locations after post-processing for entire wells in 6-well plates (Figure 2B).

We implemented this computer-guided selection by combining the generated binary image with the bright-field image, making colony-picking decisions based

solely on computer vision results. To verify the computer-guided iPSC colony detection, we performed manual picking followed by standard iPSC characterization (Figures 2C-D). Expanded colonies that passed characterization, including karyotyping by G-band analysis (data not shown), were further assessed for pluripotency. Here, we show iPSCs generated from sample c4p2. Fluorescent immunostaining demonstrated expression of pluripotent surface markers, including TRA-1-60, TRA-1-81, and SSEA4 (Figure 2C). Endogenous pluripotent genes such as OCT4, SOX2, and NANOG were fully activated at levels comparable to the human embryonic stem cell line H1 (Figure 2D). Finally, genomic PCR confirmed that UC-iPSCs expanded at passage 21 no longer harbored exogenous reprogramming factors from the original episomal plasmid (Figure 2E). Thus, this computer vision approach works for both human iPSC detection and mouse models.

### Quantitative Detection of Synchronized Time-Lapse Data for iPSC Selection Prediction in Human Somatic Cell Reprogramming

Somatic cell reprogramming is indicated by cell polarization changes in morphology. Based on experience, urinal cell reprogramming systems require more than 17 days to reach full maturation. The computer vision machine learning method allowed us to determine the stage of cellular reprogramming without being limited by experience-based criteria. To quantitatively monitor complete reprogramming, we used new sets of time-lapse training data to model iPSC formation and classify phases in this process.

First, individual colonies were back-traced as described earlier. Combined with the detection method, the earliest features of cellular polarity changes—iPSC texture formation—could be recognized at day 7 after reprogramming induction. An example of detected time-lapse data is shown in Figure 3A [Figure 3: see original paper]. Subsequently, individual iPSC texture features were detected using segmented boundaries, with each feature digitally registered and tagged (Figure 3B). Each iPSC texture feature was monitored individually throughout the reprogramming process. We could easily exclude over-grown and under-grown colonies using the average growth rate, providing quantitative measurement of colony formation across growth phases, where each curve represents a qualified iPSC clone (Figure 3C). Moreover, we applied a Hidden Markov Model (HMM) with four quantitative features to analyze this process. Under comparable reprogramming conditions with the same cell types, the model classified four different phases in the growth curve to mathematically characterize iPSC clone formation. Finally, this model provided posterior probabilities for each phase; for example, between days 12 and 17, the probability score for the maturation phase increased from 0 to 1. The model provides correlation between iPSC harvest time and probability score. Image data for optimal picking decisions were selected manually and fed into the model to calculate the selection threshold. A probability score of 0.3 was established as the picking threshold for the urinal cell reprogramming system. In this dataset, the algorithm yielded the closest

score to 0.3 on day 14, indicating that colony picking could be triggered. Therefore, the optimal picking time window can be predicted. If the probability factor reaches 1, the cells are overgrown, implying a risk of random differentiation.

In separate experiments, we synchronized iPSC colony detection using a clone-by-clone approach and triggered colony selection as described above. These colonies were then verified using high-throughput RNA-Seq gene expression analysis. Pluripotency- and germ layer-specific marker genes were plotted for comparison. We analyzed global gene expression patterns from algorithm-detected urinal cell-derived iPSCs (Alg iPSCs), UCs, and the embryonic stem cell line H1. Hierarchical clustering heatmap analysis showed that gene expression in Alg iPSCs is highly distinct from UCs (Figure 3d). Comparison of expression profiles between UCs and Alg iPSCs revealed that 1,759 genes were up-regulated in Alg iPSCs, while 1,538 genes were up-regulated in UCs (Figure 3e). The Gene Ontology project provides controlled vocabulary to describe differentially expressed genes. Up-regulated genes in UCs were enriched in GO terms related to epithelial, angiogenic, or fibroblast function, whereas up-regulated genes in Alg iPSCs were associated with DNA replication and mitotic division. These GO terms also strongly demonstrated differentially expressed genes between H1 cells and UCs (Supplemental data Figure S4). Further analysis revealed that pluripotency genes were more highly expressed in both UC-iPSCs and H1 cells compared to their low expression in UCs. The data showed no distinctive expression differences between Alg iPSCs and manually picked iPSCs (Man iPSCs) or between Alg iPSCs and H1 cells. In contrast, the UC gene expression profile clustered in a completely different manner.

## Discussion

Recent developments in computer vision techniques provide advanced tools for cellular morphology description and classification ,[13]. Application of a computer vision system enables quantitative analysis in cell biology, reducing human error and improving consistency[1] [2] . Historically, digital pathology has focused primarily on low-level image feature extraction[21] [23] (e.g., color and intensity, nuclear segmentation, typical morphology features), followed by construction of classification models using classical machine learning methods such as Support Vector Machines[2] ,[2] and Random Forests[2] . Manual design and selection of features based on human observations and experience make it difficult to achieve optimal performance. Moreover, conventional classifiers are designed for relatively small training datasets.

Deep learning has recently attracted attention in the machine learning field, motivated by the use of big data to directly train multi-layer neural networks with different deep structures that combine feature extraction and classification. Convolutional Neural Network (CNN), introduced by LeCun[2] , is widely used in computer vision. Its most successful applications involve image data classification, segmentation, detection, and retrieval tasks. For example, CNN serves as the current baseline approach in breast cancer classification and diagnosis[2] .

Chen et al.[2] presented a multi-feature, label-free cell classification system using deep learning techniques. The advantage of deep learning is its end-to-end learning procedure that combines feature extraction with classification, with performance benefiting from expanded training datasets.

Here, we present a machine learning-based system for automated detection and prediction of iPSC formation during cellular reprogramming. This model works in both mouse and human reprogramming systems, and the HMM-based model estimates phase probabilities to trigger iPSC colony selection. Our approach uses a Deep Convolutional Neural Network (DCNN) end-to-end learning framework that avoids suboptimal manual design of extractors and classifiers when faced with complex cell textures and morphological changes, providing optimized performance and convenience.

From an application standpoint, this computer-aided machine learning approach for iPSCs can detect cells early and predict optimal selection times. This tool can quantitatively evaluate different somatic reprogramming approaches, such as those using engineered transcription factors or small molecules, for iPSC generation or further differentiation. If time-lapse data have high frame rates, this automated high-throughput analysis could potentially group different stages as individual conditions for cellular reprogramming in cell fate mapping models.

Finally, since the entire reprogramming process is serum-free, feeder-free, and uses episomal-based induction, this computer vision-guided, label-free, and non-invasive approach was fully verified by standard biological methods and RNA sequencing. We expect this combined approach will become a routine technique for quantitative cell biology studies. It should not be limited to cellular reprogramming work; the system can be further developed to study downstream cell differentiation and cell line development, enabling identification of appropriate cells in a fully traceable and quantitative manner, or even guiding automated robotics in regenerative medicine applications.

## Methods

### Ethics Statement

Mus musculus (mouse model) were maintained and cared for in our Experimental Animal Centre' s facility in accordance with Guangzhou Institutes of Biomedicine and Health Institutional Animal Care and Use Committee protocols (approval number N2017039). All experiments were performed in accordance with guidelines set by the Human Subject Research Ethics Committee at Guangzhou Institutes of Biomedicine and Health (GIBH) and the Chinese Academy of Sciences (CAS), and the Committee approved the experiments (approval number GIBH-IRB07-2017039). Formal informed consent was obtained from all subjects.

### Overall Computer-Assisted System

iPSC colonies in bright-field microscopy images typically exhibit distinct morphological features compared to other cells. However, detecting single cells is difficult because cell edges and boundaries overlap and appear fuzzy. We therefore did not adopt traditional segmentation and recognition procedures. Instead, our strategy identified typical colony textures rather than single-cell features. We used the sliding window method to detect potential colony areas (Figure 4 [Figure 4: see original paper]).

A $96 \times 96$ patch was defined as a fixed detection window that "slides" across the entire test image from the top left corner. For each window, we extracted the region and applied CNN to determine whether the patch represented an iPSC area. This yielded a binary map indicating iPSC colony locations and sizes in the test images. All input images were normalized to reduce illumination condition effects. Since background is dominated by overall image brightness, we assumed that brightness normalization would minimize inter-image differences. Our recognition system uses a modified Autolevels (AL) algorithm for pre-processing (Supplemental data 1.1) that normalizes brightness and enhances acquired images. In this study, CNN achieved the best recognition rate compared to other methods (Supplemental data 1.2).

After binary image creation, morphological transformation was applied to fill holes. We ran a Gaussian filter and performed re-thresholding to remove isolated noise points. Using GFP reporter signals as positive labels in mouse data allowed us to train an effective classifier that could directly identify contours and boundaries from binary images. Finally, the image was processed to filter out sparse residuals counted as noise.

To locate human urinal cell-derived iPSC colonies and their boundaries, initial data collection selected day 12 as a seed frame for segmentation. For data before this time point, positive binary window points were counted as areas within colony contours. In more mature cases, a semi-supervised Random Walker algorithm[3] was implemented to expand areas and complete boundaries using labels to mark detection. A segmentation example on day 20 is presented in Figure 3B.

### Growth Curve Modeling

After colony boundaries were identified, each colony could be measured in terms of area and time after reprogramming induction. Growth curves were plotted based on detected signal area versus time after reprogramming induction (Figure 3C). Abnormal growth conditions (overgrowth/undergrowth) were manually defined by a stem cell specialist after reviewing all acquired image files from the reprogramming process. The abnormal phase was then discriminated based on the mean of first-order differences in the growth curve between days 10 and 20. The remaining 97 normal samples were used to train the HMM model for prediction using four features: (1) first-order difference in growth rate, (2)

second-order difference in growth rate, (3) area, and (4) number of growth days. Two stages were labeled to find the best-fit HMM model. Human experts labeled the optimal picking day for each colony. The period from one day before the picking day to the end was labeled the mature stage, while the period from day 1 to two days after positive sample appearance was labeled the first stage. We used the Baum–Welch algorithm to train the HMM model with 2,000 different initializations. The best model fit the labeled data with 4 hidden stages, with the middle portion divided into two periods of similar length. Finally, the Viterbi algorithm[31] was used to predict hidden stages of growth curves, with an example shown in Figure 3C.

### Image and Dataset Acquisition

MEF cells were cultured and imaged using Zeiss Z1 microscopy. Images of each well were captured every 2 hours. Human cells in each well were imaged daily with the Solentim Cell Metric. Images were acquired at 16-bit intensity depth, with MEF resolution of $3322 \times 2496$ per region and human cell resolution of $17702 \times 17684$ per well. More than 50 images for both cell types were acquired from 3 different plates for training.

Biologists manually annotated iPSC colonies and cropped hundreds of iPSC areas and numerous negative samples from acquired images. Cropped images had various widths and heights ranging from 200 to 400 pixels. In each selected area, a $96 \times 96$ pixel bounding box was randomly placed and cropped to create a standard patch. Class label 0 was assigned to each iPSC feature vector, while class label 1 was assigned to each non-iPSC patch. We finally obtained more than $2 \times 10$  standard patches for both cell types, including iPSCs and non-iPSCs. Repeated random sub-sampling evaluation (Supplemental data 1.2) on this dataset showed that CNN achieved the best accuracy for our study.

### Neural Network Design and Iterative Training

We used a modified AlexNet[32] structure as our neural network base. Since our colony detection application is a binary classification problem, network size could be reduced to a smaller scale without performance loss. A batch-normalized layer was added before each convolutional layer to accelerate training and slightly improve performance. To fine-tune network performance, biologists reviewed automatic recognition results for newly captured images and marked incorrect regions after initial training. Incorrect data were added to the training set with corrected labels. After more than 3 fine-tuning iterations, the detector achieved near-human recognition rates for iPSC colonies.

### Mouse Embryonic Fibroblast Cell Reprogramming

Following the previous OKS protocol[33], Oct4-GFP transgenic MEFs were plated at 4,000 cells/cm$^2$ and transfected with retrovirus (retroviral vectors carrying murine cDNAs for Oct4, Sox2, Klf4, and c-Myc were purchased from Addgene)

using Plat-E packaging cells. Forty-eight hours post-transfection, viral supernatants were removed, culture medium was added, and this time point was defined as day 0. iPSC-inducing medium contained DMEM with vitamin C, bFGF, CHIR99021, and other chemical components. Cell images in wells were captured daily using Cell Metric or SteREO Lumar.V12 (Zeiss).

### Human Urinal Cell Culture and Reprogramming

Urine cell (UC) collection, culture, and reprogramming were performed as described in Xue et al.[33] with modifications. Briefly, primary urine cells were cultured in urine cell medium consisting of equal parts DMEM/F12 and MEF medium containing 10% fetal bovine serum (FBS, PAA), 0.1 mM NEAA (Gibco), 1 mM L-glutamax (Gibco), and a SingleQuot Kit CC-4127 REGM (Lonza). Approximately 1–2 × 10 urine cells were dissociated with 0.25% trypsin treatment. After centrifugation, 6 µg of the T2k vector and 4 µg of the microRNA302 vector were co-transfected into UCs using an Amaxa™ Basic Nucleofector™ Kit (Program T-020, LONZA) according to manufacturer instructions. Transfected UCs were grown on Matrigel-coated P6 wells in UC medium. After 24 hours, UC medium was replaced with TESR supplement containing 4i (4i: CHIR99021 (3 µM), A83-01 (0.5 µM), thiazovivin (0.5 µM), and PD0325901 (1 µM)). After 12 days, the four small molecule cocktail was removed. Bright-field images were captured daily by Cell Metric and analyzed by our detection system. The system detected and identified mature iPSC colonies around day 25. Colonies were manually marked at their corresponding locations, picked, and plated on Matrigel-coated 96-well plates containing mTesR medium.

### Immunocytochemical and Quantitative RT-PCR (qRT-PCR) Analysis

Immunofluorescence was used to characterize iPS cells as previously described (Anti-TRA-1-60, Millipore; Anti-SSEA4, Life Technologies). DAPI was used to stain cell nuclei. Fluorescence-labeled cells were imaged on the ImageXpress® Micro Confocal High-Content Imaging System. Total RNA was extracted using the RNeasy Mini kit (Qiagen Cat No. 74104). First-strand cDNA was synthesized from 1 µg of total RNA using the GoScript™ Reverse Transcription System (Promega Cat No. A5000). qRT-PCR was performed on a CFX96 machine (BIO-RAD) with three biological replicates using the SYBR PCR Kit (SsoAdvanced™ Universal SYBR® Green, Cat No. 1725272). The CT method was used to calculate relative gene expression levels. Primers used for genomic PCR and qPCR are listed in the supplemental table (Supplemental Table S2).

### RNA Sequencing

RNA sequencing was performed using a protocol from a previous publication[3] . After recognizing and picking iPS clones, cells were lysed with 200 µl of Trizol (Invitrogen). Total RNA was prepared with the Direct-zol RNA MiniPrep

kit (Zymo Research). RNA was then quantified, purified, and used to generate cDNA sequencing libraries using the TruSeq RNA Sample Prep Kit (Illumina). The Qubit dsDNA HS Assay kit (Invitrogen) was used to detect cDNA library concentrations. Sequencing was performed on a MiSeq system with MiSeq Reagent Kits v2 (50 cycles) (Illumina).

## References


1. Wiley, L. A. et al. cGMP production of patient-specific iPSCs and photoreceptor precursor cells to treat retinal degenerative blindness. *Sci Rep* 6, 30742 (2016).
2. Zhao, T. et al. Humanized Mice Reveal Differential Immunogenicity of Cells Derived from Autologous Induced Pluripotent Stem Cells. *Cell Stem Cell* 17, 353–359 (2015).
3. Xue, Y. et al. Generating a non-integrating human induced pluripotent stem cell bank from urine-derived cells. *PLOS ONE* 8, e70573 (2013).
4. Zhou, X., Cao, X., Perlman, Z. & Wong, S. T. A computerized cellular imaging system for high content analysis in Monastrol suppressor screens. *J Biomed Inform* 39, 115–125 (2006).
5. Wang, M., Zhou, X., King, R. W. & Wong, S. T. Context based mixture model for cell phase identification in automated fluorescence microscopy. *BMC Bioinformatics* 8, 32 (2007).
6. Nattkemper, T. W., Ritter, H. J. & Schubert, W. A neural classifier enabling high-throughput topological analysis of lymphocytes in tissue sections. *IEEE Trans Inf Technol Biomed* 5, 138–149 (2001).
7. Nattkemper, T. W., Twellmann, T., Ritter, H. & Schubert, W. Human vs machine: evaluation of fluorescence micrographs. *Comput Biol Med* 33, 31–43 (2003).
8. Butler, M. T. & Wallingford, J. B. Planar cell polarity in development and disease. *Nat Rev Mol Cell Biol* (2017).
9. Li, X., Pei, D. & Zheng, H. Transitions between epithelial and mesenchymal states during cell fate conversions. *Protein & Cell* 5, 580–591 (2014).
10. Shu, X. & Pei, D. The function and regulation of mesenchymal-to-epithelial transition in somatic cell reprogramming. *Current Opinion in Genetics & Development* 28, 32–37 (2014).
11. Rejon, C. A., Almasri, M. & Mccaffrey, L. Cell polarity proteins in breast cancer progression. *Journal of Cellular Biochemistry* 117, 2215 (2016).
12. Chen, J. et al. BMPs functionally replace Klf4 and support efficient reprogramming of mouse fibroblasts by Oct4 alone. *Cell Res* 21, 205–212 (2011).
13. Wang, L. & He, D.-C. Texture classification using texture spectrum. *Pattern Recognition* 23, 905–910 (1990).
14. Tokunaga, K. et al. Computational image analysis of colony and nuclear morphology to evaluate human induced pluripotent stem cells. *Sci Rep* 4, 6996 (2014).
15. Maddah, M., Shoukat-Mumtaz, U., Nassirpour, S. & Loewke, K. A sys-

tem for automated, noninvasive, morphology-based evaluation of induced pluripotent stem cell cultures. *J Lab Autom* 19, 454–460 (2014).
16. Maddah, M. & Loewke, K. Automated, non-invasive characterization of stem cell-derived cardiomyocytes from phase-contrast microscopy. In *Medical Image Computing and Computer Assisted Intervention*, vol. 17, 57–64 (2014).
17. Kato, R. et al. Parametric analysis of colony morphology of non-labelled live human pluripotent stem cells for cell quality control. *Sci Rep* 6, 34009 (2016).
18. Joutsijoki, H. et al. Histogram-based classification of iPSC colony images using machine learning methods. In *Systems, Man and Cybernetics*, 2611–2617 (2015).
19. Joutsijoki, H., Haponen, M., Rasku, J., Aaltosetala, K. & Juhola, M. Error-correcting output codes in classification of human induced pluripotent stem cell colony images. *BioMed Research International* 2016, 3025057 (2016).
20. Joutsijoki, H., Haponen, M., Rasku, J., Aaltosetala, K. & Juhola, M. Machine Learning Approach to Automated Quality Identification of Human Induced Pluripotent Stem Cell Colony Images. *Computational and Mathematical Methods in Medicine* 2016, 1–15 (2016).
21. Liu, S., Mundra, P. A. & Rajapakse, J. C. Features for cells and nuclei classification. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 6601–6604 (IEEE, 2011).
22. Paci, M. et al. Non-Binary Coding for Texture Descriptors in Sub-Cellular and Stem Cell Image Classification. *Current Bioinformatics* 8, 208–219 (2013).
23. Masuda, A. et al. Automatic detection of good/bad colonies of iPS cells using local features. In *International Conference on Machine Learning*, 153–160 (2015).
24. Mammone, A., Turchi, M. & Cristianini, N. Support vector machines. *Engineering* 13, 18–28 (2009).
25. Han, J. W., Breckon, T. P., Randell, D. A. & Landini, G. The application of support vector machine classification to detect cell nuclei for automated microscopy. *Machine Vision and Applications* 23, 15–24 (2010).
26. Raytchev, B. et al. Detection of differentiated vs. undifferentiated colonies of iPS cells using random forests modeled with the multivariate polya distribution. In Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G. & Wells, W. (eds.) *MICCAI 2016*, 667–675.
27. Bengio, Y. & Lecun, Y. Convolutional networks for images, speech, and time-series. *The handbook of brain theory and neural networks* (1995).
28. Wang, J. et al. Discrimination of breast cancer with microcalcifications on mammography by deep learning. *Sci Rep* 6, 27327 (2016).
29. Chen, C. L. et al. Deep learning in label-free cell classification. *Sci Rep* 6, 21471 (2016).
30. Grady, L. Random walks for image segmentation. *IEEE Trans. Pattern*

*Anal. Mach. Intell.* 28, 1768–1783 (2006).
31. Viterbi, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory* 13, 260–269 (1967).
32. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Neural Information Processing Systems*, 1097–1105 (2012).
33. Chen, J. et al. Towards an optimized culture medium for the generation of mouse induced pluripotent stem cells. *Journal of Biological Chemistry* 285, 31066–31072 (2010).
34. Wang, L. et al. Tgf signaling regulates the choice between pluripotent and neural fates during reprogramming of human urine derived cells. *Scientific Reports* 6, 22484 (2016).

## Acknowledgements

Support for this work was provided by the Ministry of Finance "Life Instrumentation Development Program" (ZDYZ2012-3), the Hundred Talents Program of the Chinese Academy of Sciences, the Bureau of International Cooperation at the Chinese Academy of Sciences (154144KY-SB20130006), and the Instrument Developing Project of the Chinese Academy of Sciences, Grant No. Y74H0111K1. We thank Dr. Xiaoye Li for her advice on the manuscript.

## Author Contributions Statement

K.F., S.Z., and X.Z. conceived the experiments. K.F. and M.H. designed and implemented the computational system. S.Z. performed the cell culture and validation experiments. K.F., Y.Z., and J.L. analyzed the results. K.F., S.Z., and X.Z. wrote the manuscript. All authors reviewed the manuscript.

## Additional Information

The authors declare no competing financial interests.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv —Machine translation. Verify with original.*