Prediction Technology for Undifferentiated Marker Expression Levels via Label-Free Noninvasive Cell Imaging

Ryutaro Akiyoshi *1

The production of regenerative medicine products requires continuous monitoring of the culture process and its quality, while avoiding microbial contamination and destruction of cells. In this study, a new method was developed to monitor the undifferentiated state of stem cells by non-destructive cell image analysis, eliminating the need for conventional destructive methods. Machine learning algorithms were used to predict the expression levels of undifferentiated markers such as Oct3/4 and Nanog without physical contact. The method is based on training data from iPSC images acquired by bright-field microscopy and assessments by destructive methods such as quantitative PCR and immunostaining. The application of unsupervised and semi-supervised models enables accurate prediction of the undifferentiated state using image analysis alone. This approach contributes not only to stem cell research, but also to real-time monitoring and new quality control methods for regenerative medicine products.

INTRODUCTION

Pluripotent stem cells (PSCs), including embryonic stem cells and induced pluripotent stem cells (iPSCs), can give rise to all types of cells derived from the three embryonic germ layers. Because of their pluripotency and ability to proliferate indefinitely, these stem cells are attracting attention in various fields such as regenerative medicine, drug testing, disease modeling, and embryonic development. As a source of regenerative medicine product, PSCs must remain undifferentiated, and several marker genes and proteins have been used to detect PSCs that have deviated from the undifferentiated state[(1)]. Examples of such markers included intracellular regulators (e.g., Oct3/4 and Nanog), which are expressed in undifferentiated cells, and cell surface markers (e.g., SSEA-4 and Tra-1-60), the expression of which is downregulated in differentiated cells[(2)-(4)].

However, destructive methods are required in order to evaluate the undifferentiated state of cells using the above molecular markers, such as flow cytometry (FCM), quantitative polymerase chain reaction (qPCR), RNA sequencing (RNA-seq), and immunostaining. If destructive methods are used repeatedly during the manufacturing process of regenerative medicine products, the cumulative loss of cells, which should be used for products, becomes

*1 Life Research & Development Department, Innovation Center, Marketing Headquarters

nonnegligible. This is a particularly critical problem because regular long-term testing is generally required in cell product manufacturing[5]-[7].

In this study, to observe cell morphology and detect changes in cell status, we employed bright-field microscopy, which is routinely used in cell product manufacturing. Bright-field microscopy is a nondestructive method that does not require analytical reagents; thus, it is suited for continuous monitoring of cell status and has cost advantages. However, evaluation results are influenced by observers' experience and perception, posing a problem in terms of consistency. To overcome this, bright-field microscopy image analysis that does not rely on human judgement has recently been investigated. Christiansen et al. proposed a model that predicts CD31 protein expression from phase contrast images based on the results of fluorescence staining of CD31 protein expressed on iPSCs and endothelial cells derived from iPSCs[8]. However, this approach requires fluorescence labeling and relevant expert annotations. Schmauch et al. directly correlated pathology images of tumor specimens to RNA-seq data in order to develop a method that predicts cancer-related gene expression from tissue images[9]. Although this method achieves high accuracy without the need to identify cell morphologies or annotate individual cells, it involves staining and fixing cells, and is therefore still destructive.

Here, we developed models that predict the expression of undifferentiated cell markers based on the results of bright-field microscopy of live cells, which does not require cell staining and expert annotation. We built models using unsupervised and semi-supervised frameworks and identified models that accurately predict the results of several molecular assays.

MATERIALS AND METHODS

iPSC Preparation

The following four conditions, which might cause deviation from the undifferentiated state, were used for culturing of iPSCs (201B7).

(1) Control (Condition 1): The general maintenance culture method for iPSC was used.
(2) Low-nutrient condition (Condition 2): An inactivated medium (StemFit medium treated at 56°C for 30 min) was used.
(3) Differentiation condition (Condition 3): Dulbecco's modified Eagle's medium with 10% fetal bovine serum, 1% minimum essential medium non-essential amino acids, and 1% GlutaMax was used.
(4) Physical stimulation condition (Condition 4): Cells were suspended by pipetting 20 times at passage and then cultured using the general maintenance culture method for iPSCs.

Microscopic imaging of cells cultured under each condition was performed from day 1 to day 2, and from day 4 to day 5, and cell quality was evaluated using destructive methods (qPCR, FCM, immunostaining, RNA-seq; Figure 1).

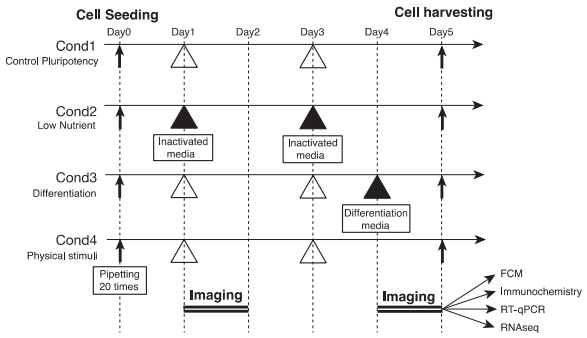

Figure 1 Differences in four iPSC culture conditions and timing of cell assessment
White triangles: changing of the common culture media
Black triangles: changing of the condition-specific media

Bright-field Imaging of iPSCs

Images of cells cultured on 6-well plates were acquired using CellVoyager CQ1 (Yokogawa Electric Corporation) with a 20 × high numerical aperture objective. In 4 of the 6 wells on each plate, there are 5 regions consisting of 5 × 8 squares, for a total of 200 fields of view per well. In each field, five planes were acquired at 5-µm intervals along the z-axis (Figure 2). Further, time-lapse imaging at 1-h intervals was performed for a total acquisition time of 20 h. Contrast-enhanced (CE) bright-field images were created using High Content Analysis Software System CellVoyager CellPathfinder (Yokogawa Electric Corporation) from images captured at different z-axis positions.

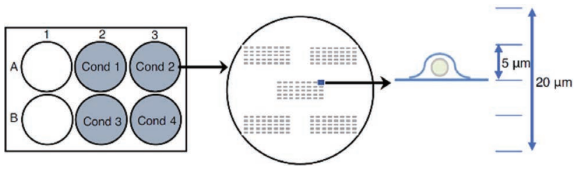

Figure 2 Sample arrangement in a 6-well plate for cell imaging and the data acquisition method in each field

Creation of Models for Predicting Results of Destructive Methods

CE bright-field images were resized, and the brightness of each pixel was tuned. These preprocessed images were used to build unsupervised and semi-supervised models. The one-class support vector machine (One-Class SVM) was used to build unsupervised models, and two models with different characteristics were obtained based on the correlation coefficient and root mean square error of correlations with results of destructive techniques (qPCR, FCM, immunostaining, and RNA-seq). A classifier model by Waisman et al. was used as a guide model for labeling images of iPSCs cultured under the four conditions as either undifferentiated or differentiated[10]. Highly reliable images for prediction of differentiated and undifferentiated iPSCs were extracted and used for retraining and optimization of models, and one model was built.

RESULTS

iPSC Imaging

It was difficult to detect differences among Conditions 1, 2, and 4 by visual inspection of bright-field images of iPSCs taken on day 5 (Figure 3). In contrast, in Condition 3, the morphology of iPSCs was clearly unusual (flattened) and the confluency was markedly higher than in the other conditions.

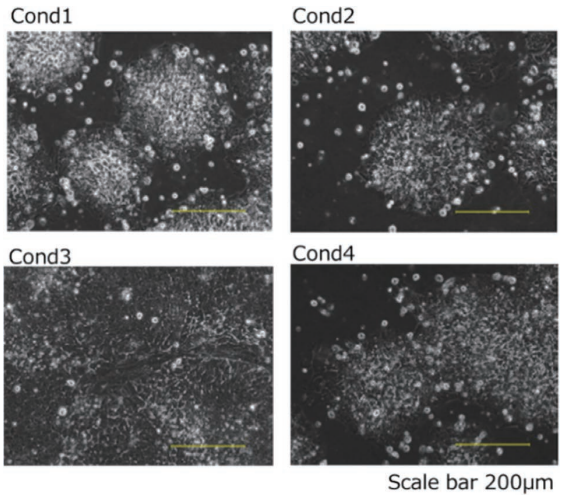

Figure 3 Bright-field images of iPSC in each condition (day 5)
Condition 1: control; Condition 2: low nutrient condition, Condition 3: differentiation condition, Condition 4: physical stimulation condition

Analysis of Various Biological Data

Protein Expression Analysis (FCM)

Expression of SSEA-4 and Tra-1-602, which are markers for undifferentiated cells, was examined by FCM. The ratio of undifferentiated cells was defined as the percentage of double-positive cells determined by manual gating or automated gating (mindensity2) out of the total live cells. The ratios of undifferentiated cells were relatively high regardless of the gating method, and the mean ratios of undifferentiated cells were > 90% in all conditions (Figure 4). Figure 4 shows the ratio of undifferentiated cells in each condition. To visualize differences in the ratio of undifferentiated cells by condition, the vertical axis shows the ratio of undifferentiated cells and the horizontal axis shows the condition. The possibility of unequal variance between groups was taken into account, and Welch's t test was used to determine statistical significance ($p < 0.1$).

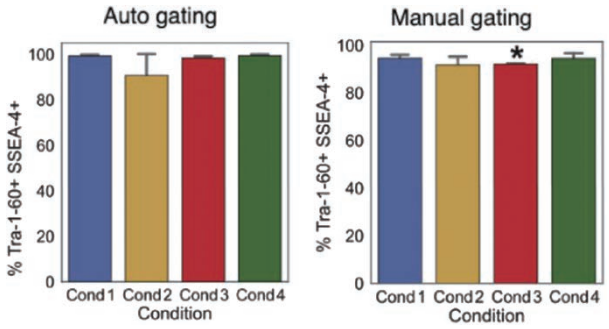

Figure 4 Ratios of undifferentiated cells determined by FCM (with automated or manual gating) in each condition
*: $p < 0.1$ (Welch's t-test)

Protein Expression Analysis (Immunostaining)

Cells were stained using antibodies to Oct3/4 and Nanog, and positive and negative staining was judged by automated thresholding using the BASC algorithm and manual thresholding. The number of positive cells with an intensity above each threshold was used to calculate the ratio of undifferentiated cells (Figure 5). In Condition 1, the ratio of Nanog-positive cells was lower with automated thresholding than with manual thresholding. With automated thresholding, the Nanog-positive ratio was significantly lower in Conditions 2 and 3 than in Condition 1 ($p < 0.05$), with an especially large decrease in Condition 3. On the other hand, with manual thresholding, the Nanog-positive ratio was significantly lower in only Condition 2 compared with Condition 1 ($p < 0.1$), and the profile was similar to that of the ratios of undifferentiated cells determined by FCM with manual gating (Pearson correlation coefficient, 0.93). In Condition 1, the ratios of Oct3/4-positive cells were lower with automated thresholding than with manual thresholding. With automated thresholding, the ratios of Oct3/4-positive cells were generally higher than the ratios of Nanog-positive cells. With automated thresholding, the ratios of Oct3/4-positive cells were slightly lower in Condition 2 and 3 than in Condition 1, and the difference between Conditions 1 and 2 was significant ($p < 0.1$). On the other hand, with manual thresholding, the ratios of Oct3/4-positive cells were significantly lower in Condition 2 and significantly higher in Condition 3 than in Condition 1.

Gene Expression Analysis (qPCR, RNA-seq)

Expression of Nanog and Pou5f1 (Oct3/4 gene) was analyzed by qPCR (Figure 6A) and RNA-Seq (Figure 6B). The expression levels of Nanog and Pou5f1 were significantly lower in Condition 2 than in Condition 1 by both methods (Figure 6A and 6B). The difference in Nanog expression level was greater between Conditions 1 and 3 than between Conditions 1 and 2, but there was no significant difference in Pou5f1 expression between Conditions 1 and 3.

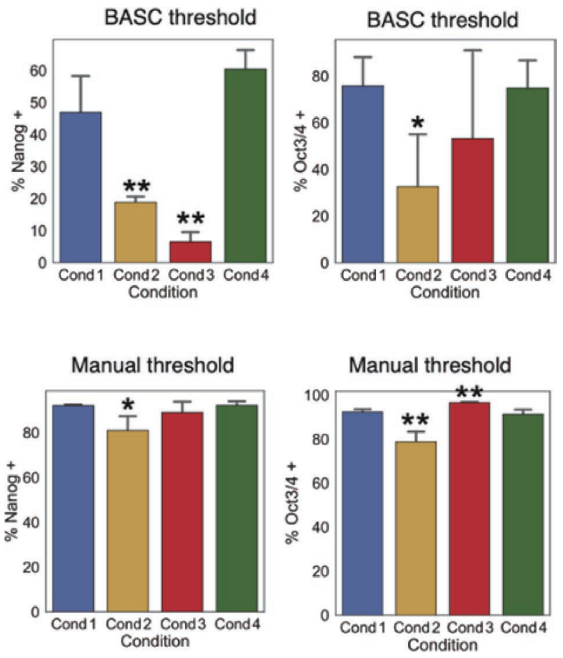

Figure 5 Ratios of undifferentiated cells determined by immunostaining in each condition
Top: Results with BASC algorithm-based automated thresholding
Bottom: Results with manual thresholding
*: p < 0.1, **: p < 0.05 (Welch's t-test)

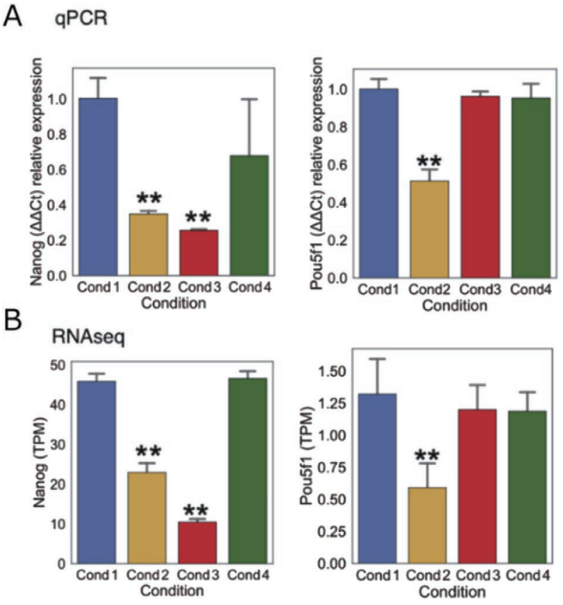

Figure 6 Comparison of gene expression levels of Nanog and Pou5f1
A: qPCR, B: RNA-seq
*: p < 0.1, **: p < 0.05 (Welch's t-test)

Predictions of Destructive Test Results by Models

CE bright-field images of cells in the four culture conditions were input into the three models for predicting destructive test results, and the expression levels of undifferentiated cell markers were predicted (Figure 7). Model I with the One-Class SVM framework showed the strongest correlation to the immunostaining results for Oct3/4 with manual thresholding (bottom right panel in Figure 5, r = 0.998), and also showed strong correlations to the qPCR and RNA-seq results for the Oct3/4 gene. Model I also showed strong correlations to the FCM results for undifferentiated cell marker proteins with auto-gating (left panel in Figure 4, r = 0.912). Model II, which also employs the One-Class SVM framework, showed the strongest correlation to the immunostaining results for Nanog with automated thresholding (top left panel in Figure 5, r = 0.961), and correlations to Nanog gene expression levels (left panels in Figure 6). On the other hand, Model III with a semi-supervised framework showed a very strong correlation to the immunostaining results for Nanog with automated thresholding (top left panel in Figure 5, r = 0.978), and correlations to Nanog gene expression levels (left panels in Figure 6). Furthermore, Models II and III showed strong correlations (r = 0.908 and r = 0.938, respectively) to FCM results with manual gating (left panel in Figure 4). Taken together, each model predicted expression levels of undifferentiated cell markers with high accuracy.

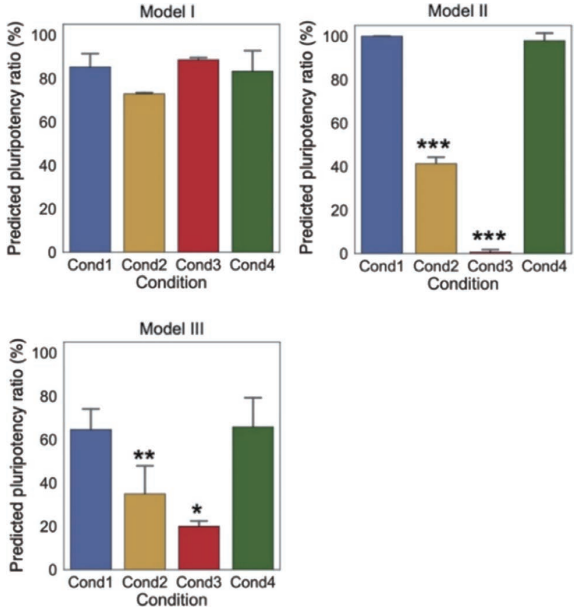

Figure 7 Predictions of expression levels of undifferentiated cell markers in each condition by models for predicting the results of destructive methods
*: p < 0.1, **: p < 0.05, ***: p < 0.01 (Welch's t-test)

DISCUSSION

Various cellular states, such as epigenetic, transcriptional, and metabolic states, change over a period of hours, weeks, or longer; thus, abnormalities can be missed if their detection relies on a single marker or a single method[11]-[13]. This study examined in detail the expression of undifferentiated cell markers in iPSC using multiple destructive methods and found discrepancies between the expression of Oct3/4 gene and

Oct3/4 protein (Figures 5 and 6). This suggests a limitation of methods that rely solely on gene expression or protein expression for accurate capturing changes in cell status. In this study, we optimized independent models predicting either protein or gene expression and showed that changes in cell status are detectable even with the above discrepancies. By using changes in the expression of various molecular markers determined by different techniques, those prediction models can be optimized as signal detection models that comprehensively capture the cellular states. Further, this nondestructive label-free approach can be used for continuous nondestructive monitoring of cell status in the closed culturing environment, which may facilitate process control or process stoppage when abnormalities occur.

This study developed models that predict the expression of undifferentiated cell markers from cell images. This approach indicates that changes in expression of any molecules that influence the cell population dynamics (e.g., appearance and confluency) can be predicted from cell images. By extending these models to detect variation in expression of intracellular molecules, advance quality assessments, beyond assessment of the differentiation status of cells, will be achievable in cell research and cell product manufacturing.

Moreover, integration of the model-building workflow developed in this study with new imaging technologies, such as holographic imaging and the RM-DIC (retardation-modulated differential interference contrast) microscopy, will enable us to visualize the structure of living cells at higher resolution[14][15], thereby expanding possibilities of a robust and reliable pipeline for assessing cell quality in the manufacturing of regenerative medicine products.

CONCLUSION

We developed models that can predict gene and protein expression levels in cells from iPSC images with high accuracy in a nondestructive manner. Future challenges include demonstrating the robustness of our models' predictions when applied to several cell lines, and building similar prediction models for differentiated cells. We further validate the application of this technology to actual manufacturing and demonstrate its value at our clients' sites in order to provide solutions to challenges in the manufacturing of regenerative medicine products.

ACKNOWLEDGEMENT

This technology was developed with SBX BioSciences, Inc.

REFERENCES

(1) P. W. Andrews, I. Barbaric, et al., "The consequences of recurrent genetic and epigenetic variants in human pluripotent stem cells," Cell Stem Cell, Vol. 29, No. 12, 2022, pp. 1624-1636
(2) I. Chambers, J. Silva, et al., "Nanog safeguards pluripotency and mediates germline development," Nature, 450, 2007, pp. 1230-1234
(3) J. A. Hackett, M. A. Surani, "Regulatory principles of pluripotency: From the ground state up," Cell Stem Cell, Vol. 15, No. 4, 2014, pp. 416-430
(4) J. Silva, J. Nichols, et al., "Nanog is the gateway to the pluripotent ground state," Cell, Vol. 138, No. 4, 2009, pp. 722-737
(5) A. Mamaeva, O. Krasnova, et al., "Quality control of human pluripotent stem cell colonies by computational image analysis using convolutional neural networks," Int. J. Mol. Sci., Vol. 24, No. 1, 2023, p. 140
(6) T. Kawase, K. Okuda, et al., "Non-invasive, quantitative assessment of the morphology of γ-irradiated human mesenchymal stem cells and periosteal cells using digital holographic microscopy," Int. J. Radiat. Biol., Vol. 92, No. 12, 2016, pp. 796-805
(7) M.-J. Kang, Y.-W. Cho, et al., "Progress in nano-biosensors for non-invasive monitoring of stem cell differentiation," Biosensors, Vol. 13, No. 5, 2023, p. 501
(8) E. M. Christiansen, S. J. Yang, et al., "In silico labeling: Predicting fluorescent labels in unlabeled images," Cell, Vol. 173, No. 3, 2018, pp. 792–803.e19
(9) B. Schmauch, A. Romagnoni, et al., "A deep learning model to predict RNA-Seq expression of tumours from whole slide images," Nat. Commun., Vol. 11, 2020, p. 3877
(10) A. Waisman, A. La Greca, et al., "Deep learning neural networks highly predict very early onset of pluripotent stem cell differentiation," Stem Cell Rep., Vol. 12, No. 4, 2019, pp. 845-859
(11) N. Rachinger, S. Fischer, et al., "Loss of gene information: Discrepancies between RNA sequencing, cDNA microarray, and qRT-PCR," Int. J. Mol. Sci., Vol. 22, No. 17, 2021, p. 9349
(12) E. Abruzzese, M. Bocchia, et al., "Minimal residual disease detection at RNA and leukemic stem cell (LSC) levels: Comparison of RT-qPCR, d-PCR and CD26+ stem cell measurements in chronic myeloid leukemia (CML) patients in deep molecular response (DMR)," Cancers, Vol. 15, No. 16, 2023, p. 4112
(13) The International Stem Cell Initiative, "Assessment of established techniques to determine developmental and malignant potential of human pluripotent stem cells," Nat. Commun., Vol. 9, 2018, p. 1925
(14) I. Moon, B. Javidi, "Three-dimensional identification of stem cells by computational holographic imaging," J. R. Soc. Interface, Vol. 4, No. 13, 2007, pp. 305-313
(15) K. Nishimura, H. Ishiwata, et al., "Live-cell imaging of subcellular structures for quantitative evaluation of pluripotent stem cells," Sci. Rep., Vol. 9, 2019, p. 1777