# Assessment of an imaging protocol for real-time selection of human iPSC colonies using live cell microscopy and image recognition software

William T. Hendriks[3,4], Samuel V. Alworth[1], Zakary A. Kenyon[1], Hoyin Lai[1], Chieko Nakada[2], Laurence Daheron[3], Yasujiro Kiyota[2], James S.J. Lee[1], Lee L. Rubin[3,4], and Chad A. Cowan[3,4]

[1] *DRVision Technologies, LLC. Bellevue, WA 98008;* [2] *Nikon Corporation, Yokohama, Kanagawa Prefecture 244-8533, Japan;* [3] *Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138;* [4] *Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138*

## Abstract

Previously, we have created a custom image recognition analytic using CL-Quant (Nikon Corporation) for the classification of induced pluripotent stem cell (iPSC) colonies vs. non-iPSC colonies in human fibroblasts undergoing reprogramming using Klf4, Oct 3/4, Sox 2 and c-Myc, and achieved a classification accuracy of 99.83% (correctly classifies 4070 out of 4077 colonies) using only the phase contrast images without fluorescence information (1). The study used static images and measurements at three weeks post infection. We are interested in answering whether iPSC formation can be predicted at much earlier time points, and whether kinetic pattern recognition could detect features for early prediction. Early selection and selective monitoring has the promise to greatly reduce the cost and labor of iPSC production. Image recognition for colony selection is also critical to reduce the need for highly skilled labor and / or enable selection automation. In this study we've used a novel, full course image set acquired on the BioStation CT (Nikon Corporation). We have discovered an early kinetic pattern of colony formation that can reliably predict which colonies will progress to form iPSC colonies (as judged by viral GFP silencing). This decision can be made within the first 72 hours of colony emergence, well before the GFP is silenced. We have implemented a decision rule to automatically select the colonies at this early time point.

## Ideal imaging and selection

Figure 1 shows the ideal protocol for imaging and colony selection. Colonies are monitored as they emerge and the vast majority of them are rejected. A small number of promising colonies could then be isolated for further monitoring as they continue to undergo reprogramming, after which a final decision could be made.

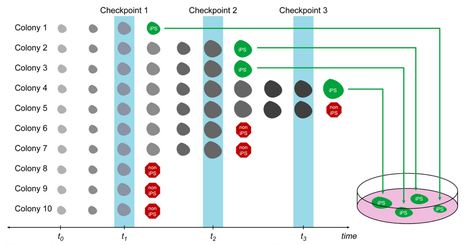



***Figure 1.*** *Ideal imaging and selection protocol. Colonies are monitored automatically from the time of initial formation $t_0$. The colonies are subjected to a series of checkpoints at increments offset from $t_0$, where they are rejected or selected.*

## iPSC reprogramming

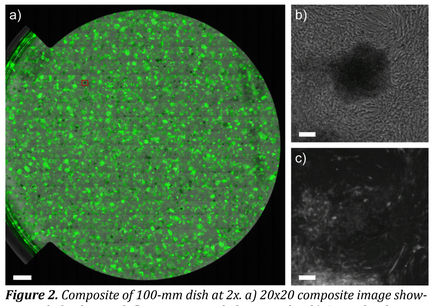



***Figure 2.*** *Composite of 100-mm dish at 2x. a) 20x20 composite image showing a whole plate with fluorescence and phase overlay. b) Example of an iPSC colony in phase contrast. c) The same iPSC colony in fluorescence, note the lack of GFP signal in the iPSC colony. Scale bars are 5 mm for (a) and 200 µm for (b) and (c).*

Adult fibroblasts were plated (100,000 cells into one well of a 6-well plate) and transduced for two consecutive days with four retroviruses (1ml MIG-Oct4, 1ml MIG-Klf4, 1ml MIG-Sox2 and 200ul MIG c-Myc). Five days later, the cells were trypsinized and replated on irradiated embryonic mouse fibroblasts in one 10 cm dish. The cell culture dish was then placed in the BioStation CT and incubated over a period of four weeks. Scans of the entire dish were obtained by 2x magnification in phase contrast and fluorescence (GFP) every 6 hours (Figure 2a). After 23.5 days of imaging, 28.5 days post infection, 26 iPSC colonies (Figure 2b) were identified by viral GFP-silencing to establish ground truth data (iPSC vs. non iPSC). Loss of GFP signal (Figure 2c) indicates successful reprogramming into iPSC colonies.

## Image analysis

Early colony emergence is detected using a CL-Quant segmentation procedure incorporating Soft Matching technology (2,3). Early events are qualified to remove noise, and the colonies are tracked. Seventy-eight colonies were used in this study, and any errors in segmentation were corrected manually. Kinetic measures are quantified using a measurement procedure. Summary kinetic statistics were made on these measurements for each colony, and these statistics were used to create a decision rule to classify colonies into fully reprogrammed (iPSC) and partially reprogrammed (non-iPSC) groups.

## Results

### A Pattern of Early Colony Formation

Our major finding is the discovery of an early pattern predictive of iPSC colony formation that occurs within the first 72 hours after a colony is first discernible (Figures 4 and 5). iPSC colonies form as the steady emergence of a small dense core. Non iPSC colonies seem to emerge all at once, and from multiple regions. Results are summarized in Figure 3.

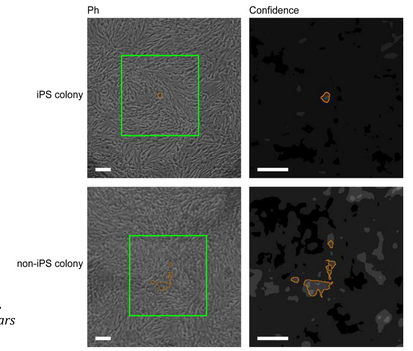



***Figure 3.*** *Early formation differences between iPSC and non-iPSC colonies. Initial detection mask is outlined in orange. Confidence channel images are generated using CL-Quant's segmentation procedure and used for detection, and are insets of the Ph images (green square). The scale bars are 200 µm for all images.*

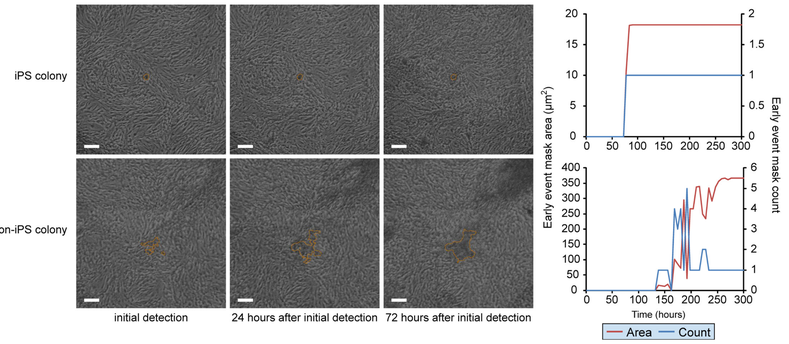



***Figure 4.*** *Change in colony formation mask over time. Measurements of mask component count and area over time define the formation pattern. Early event mask is smaller, has fewer components and is more stable in iPSC colonies than non-iPSC colonies over time. Scale bars for all the images are 200 µm.*

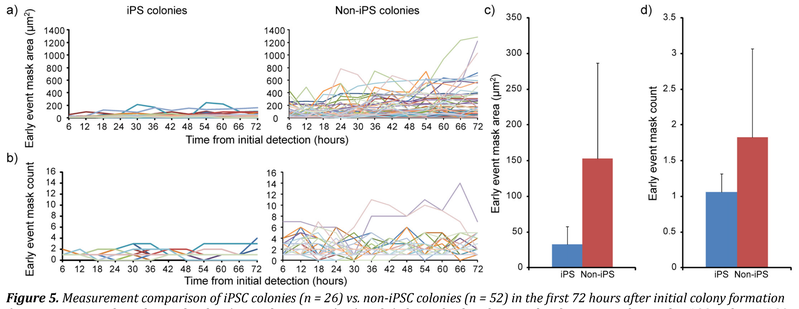
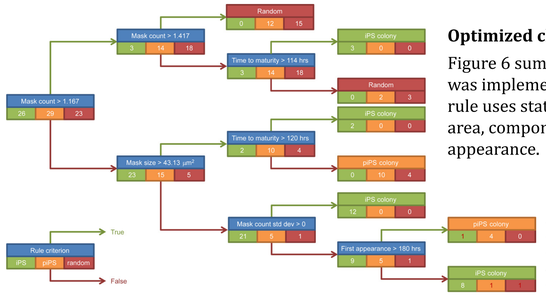



***Figure 5.*** *Measurement comparison of iPSC colonies (n = 26) vs. non-iPSC colonies (n = 52) in the first 72 hours after initial colony formation (time courses are aligned to each colony's initial time point). a) and c) shows the distribution of early event mask area for iPSC and non-iPSC colonies (error bars show + 1 standard deviation). iPSC colonies are smaller compared to non-iPSC colonies. b) and d) shows the distribution of early event mask detected with iPSC having fewer component counts.*




### Optimized classification rules

Figure 6 summarizes the decision rule that was implemented using kinetic statistics. The rule uses statistics on three measurements: area, component count, and time of initial appearance.

***Figure 6.*** *Optimized decision rule for colony classification. Three error cases remain upon optimization of the decision tree and are shown in red numbering. Analysis of the error cases are shown in Figure 8 below.*

## Discussions

### iPSC colony occurrence in decision set

Occurrence of iPS outcome in the decision rule set is significantly higher than the occurrence of iPSC in the study set (equivalent to random selection as control) with a p-value $7.48 \times 10^{-9}$ (one sided t-test). The overall accuracy of the rule is 96.2% +/- 4.3% (75 correct assignments out of 78 colonies).

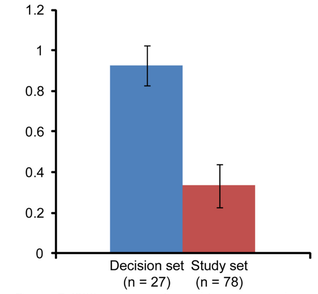



***Figure 7.*** *iPSC occurrence in teaching set vs. study set. The decision set has statistically significantly higher rate of iPSC occurrence.*

### Error analysis

Figure 8 shows the three misclassified colonies. Features that are both iPSC-like and non-iPSC-like is found in these error cases.

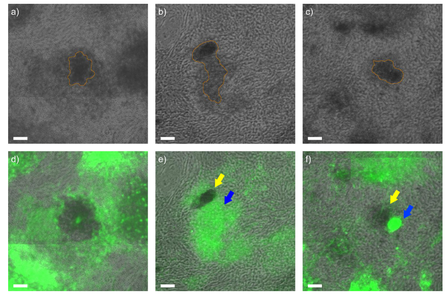



***Figure 8.*** *Error case analysis of misclassified colonies. Outline of the detected colony is shown in the top row. a) and d) shows an iPSC colony that is mis-classified as a non-iPSC. It may be acceptable to miss some iPSC colonies so long as specificity is very high. b) and e) was considered a non iPSC colony in our ground truth (determined at week 3). However, close inspection of the time sequence reveals two colonies - an iPSC colony (yellow arrow) and a non iPSC colony (blue arrow) - that are overlapping each other, resulting incorrect truth assignment. Similarly c) and f) shows a colony considered non iPSC in the ground truth. However, review of the time series shows that it is actually two colonies. The non-iPSC colony (blue arrow) overlaps a colony lacking GFP expression (yellow arrow) that is likely iPSC. Scale bars are 200 µm for all images.*

## Conclusion

In this study, we set out to discover early kinetic features for colony classification using image recognition. Our findings are summarized below:

- Colony reprogramming was recorded in real-time using phase-contrast live-cell microscopy.
- New algorithms and decision rules for image analysis of early colony formation were developed.
- The algorithm enables iPSC colonies prediction within first 72 hours of colony formation - well before viral-GFP silencing.
- These analyses would become useful tools for early iPSC colonies selection.

We are confident that quantitative descriptors of early colony formation will enable iPSC colony formation prediction at early time points as we improve the decision rule in future studies.

## Acknowledgements

The project described was supported by Grant Number R44HL106863 from the National Heart, Lung, And Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, And Blood Institute or the National Institutes of Health.

## Works cited

1. **Lai H, Hendriks WT, Wadadekar TV, Alworth SV, Daheron L, Nakada C, Kiyota Y, Cowan CA, Rubin LL, Lee JSJ (2011).** Teachable tool for standardization of human induced pluripotent stem cell colony selection from live cell microscopy image sequences. Poster presented at the American Society for Cell Biology annual conference, Denver CO.

2. **Alworth SV, Watanabe H, Lee JSJ (2010).** Teachable, high content analytics for live cell, phase contrast movies. J Biomol Screening.

3. **Lee JSJ, Oh S (2007).** Learnable object segmentation. US Patent # 7203360.