SVCell software, powered by machine learning, enables the automated and standardized generation of patient-specific, clonal cell lines with reduced variability

Samuel V. Alworth[1], Nicholas J. Collins[2,3], Zakary A. Kenyon[1], Louis Piloco[1], Laurence Daheron[2,3], Lee L. Rubin[2,3], Chad A. Cowan[2,3], and James S.J. Lee[1]

[1] DRVision Technologies, LLC. Bellevue, WA 98008; [2] Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138; [3] Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138;

Introduction

Induced pluripotent stem cells (iPSC) are increasingly being adopted for disease modeling, and as sources of tissue for regenerative medicine. However, the generation of patient-specific cell lines can be laborious, costly and time consuming. Current processes result in unwanted variations between iPSC lines. These differences can affect functional properties in disease modeling or after transplantation. Robotic patient cell generation systems are being developed to automate and standardize the generation of patient-specific cell lines in order to reduce the variability and costs.

SVCell is a machine learning enabled image recognition software for live cell time-lapse microscopy applications. It consists of a suite of application modules, called recipes, which provide critical functions for automated patient-specific cell generation systems. The SVCell recipes can assess the growth and doubling times of input patient cells, characterize and count the colonies that emerge during reprogramming, and assess the reprogramming status (i.e. full or partially reprogrammed). The recipes incorporate innovative machine learning technologies that provide for accuracy and robustness across systems. The SVCell iPSC Score provides the label-free identification of high quality iPSC clones without fluorescent markers. This uniquely enables the generation of true clonal, rather than polyclonal, cell lines having consistent reprogramming quality and differentiation potential.

SVCell Software

SVCell Recipe Station (RS)
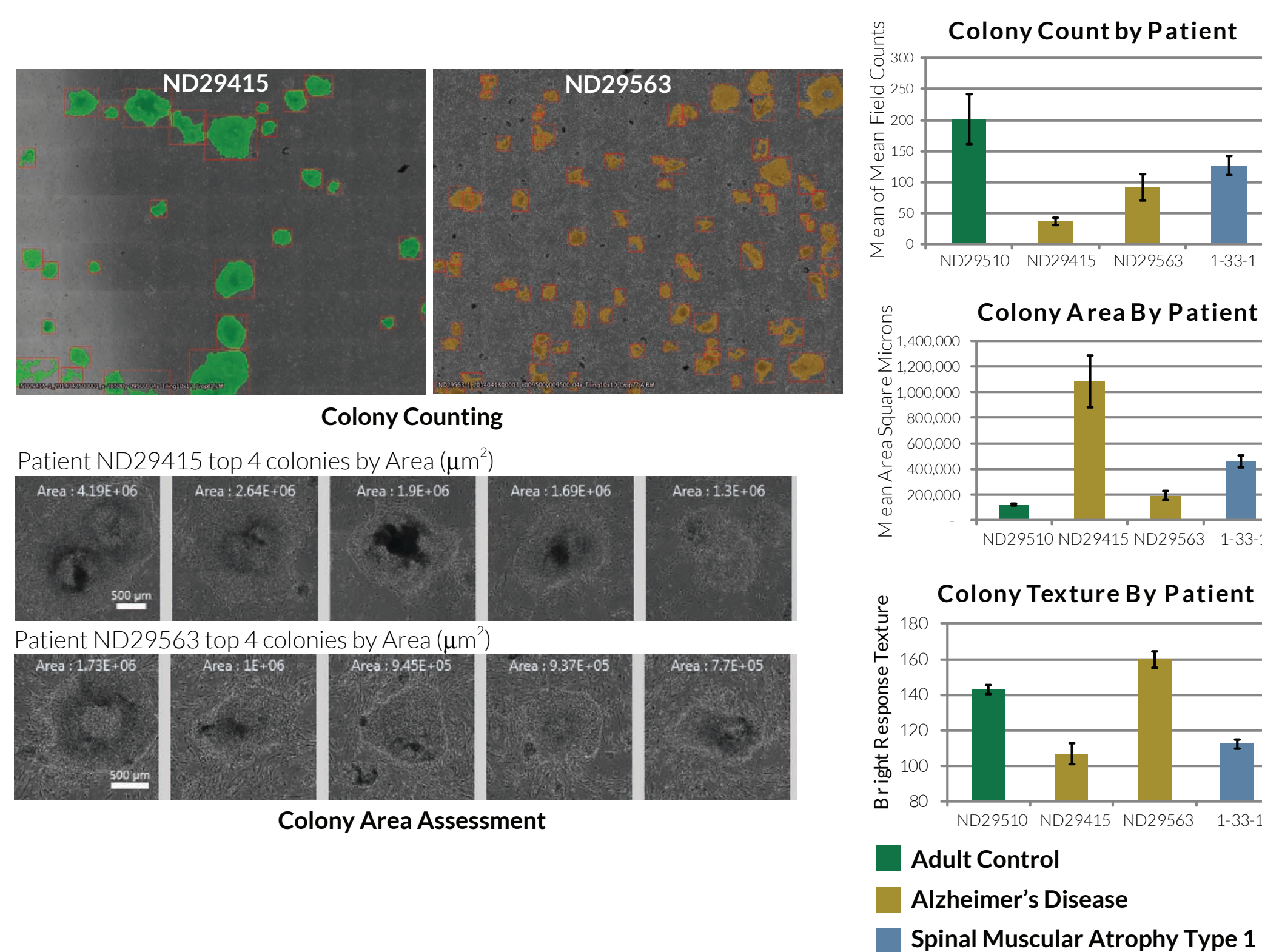
SVCell Recipe Engine
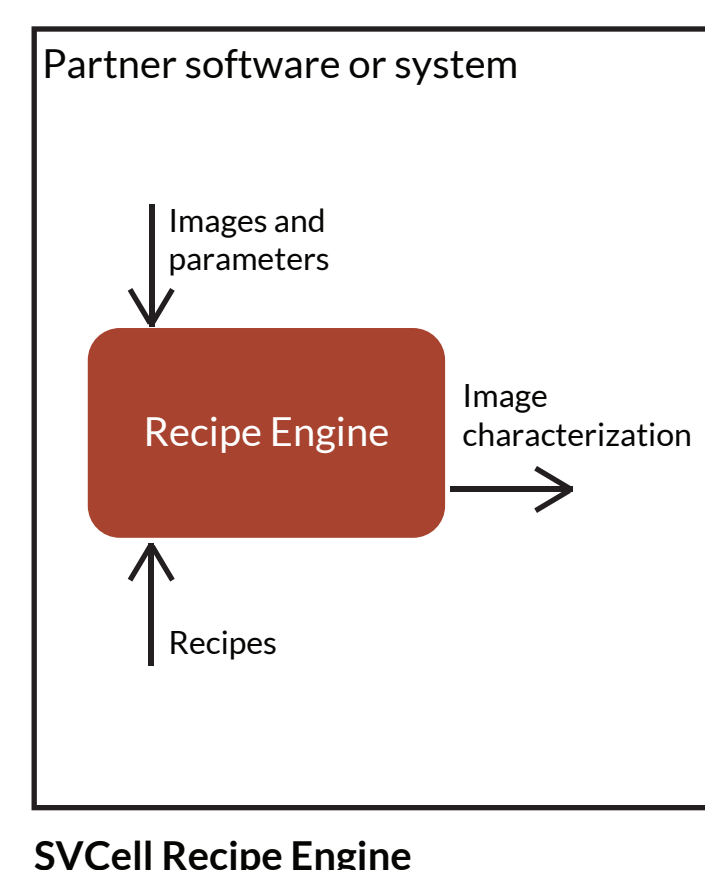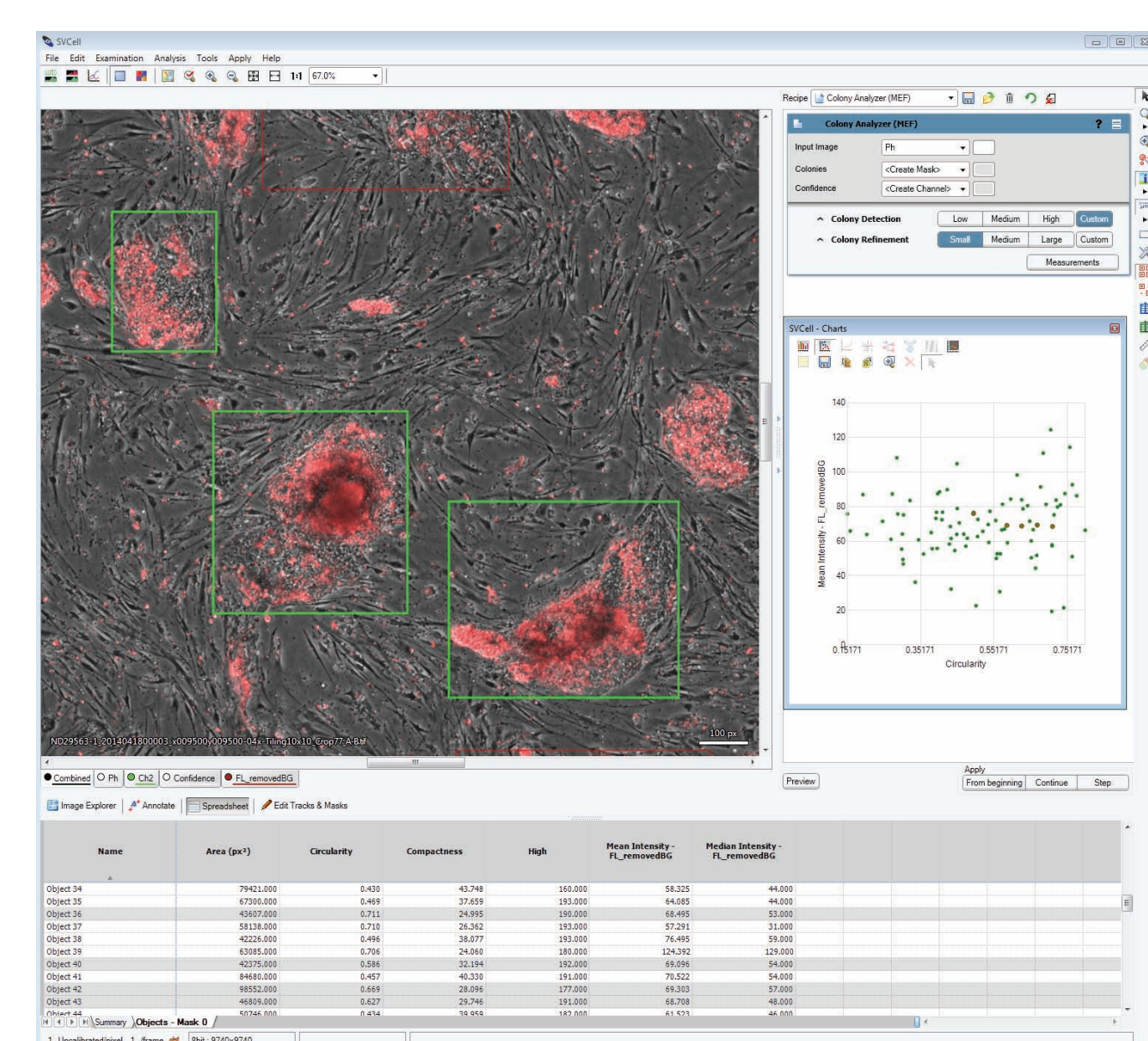
Recipes can be deployed in the easy-to-use Recipe Station or in the behind-the-scenes Recipe Engine that is deployed inside partner systems or software. The use of SVCell by customers and partners is protected by 48 issued U.S. patents covering the SVCell interface [1], technologies [2,3], and usage [4,5].
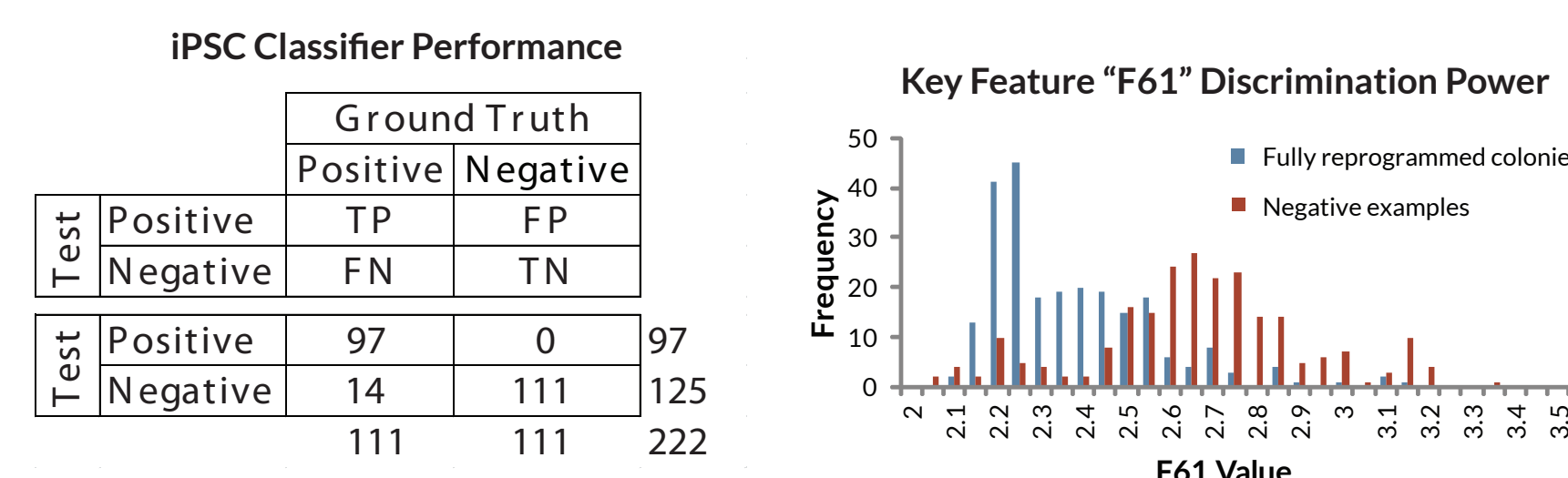
Machine Learning Inside
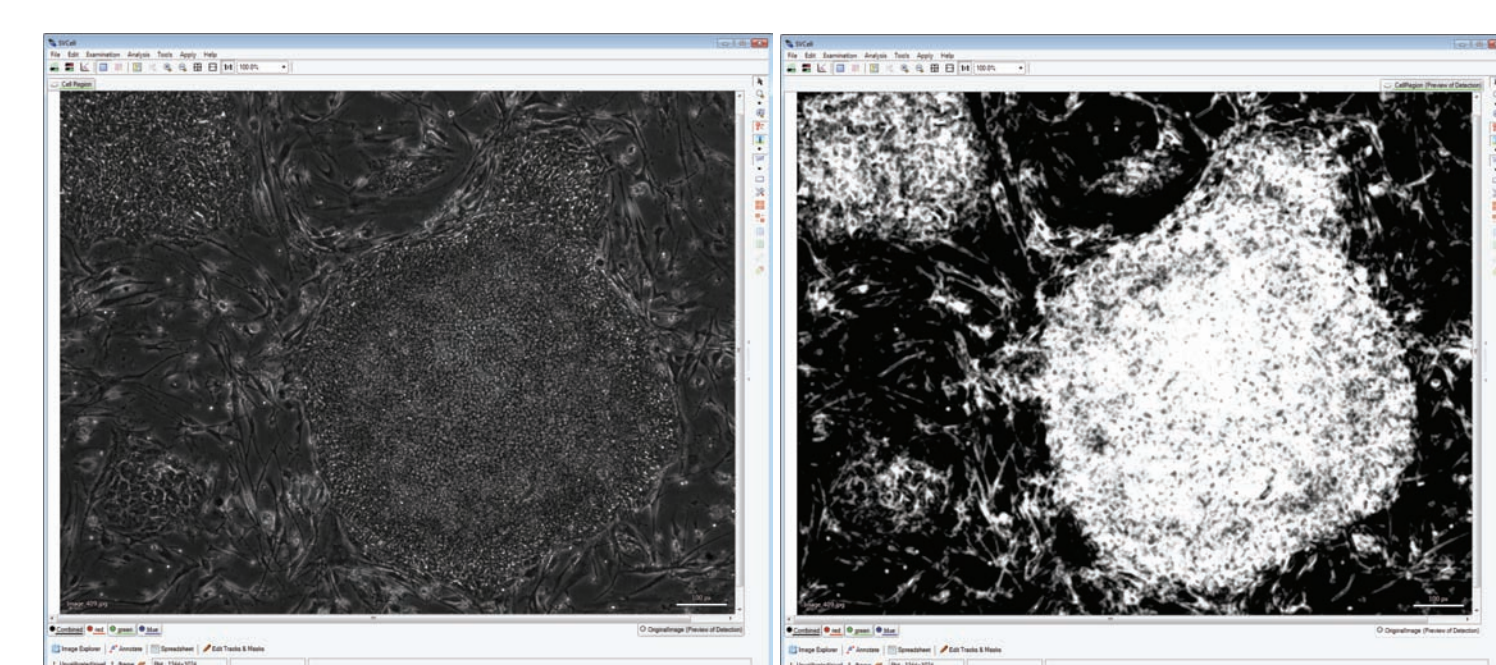
SVCell recipes incorporate highly innovative machine learning technologies for both image transformations and object (e.g. cells, colonies) classification and scoring. Soft Matching™ technology [2,6,7], shown left, uniquely enables the processing of phase contrast and other challenging image modalities by transforming the input image into a confidence image that is amenable to traditional image processing methods. SVCell recipes contain many types of scoring and classification technologies including linear and non linear regression and DRVision's proprietary regulated decision tree (shown here) [3]. These technologies are deployed inside the easy-to-use Recipe Station where end users get results with the click of a button; no machine training required.

Patient Colony Characterization

Colony Counting

Patient ND29415 top 4 colonies by Area (µm²)

Patient ND29563 top 4 colonies by Area (µm²)

Colony Area Assessment

Colony Count by Patient

Colony Area By Patient

Colony Texture By Patient

- Adult Control
- Alzheimer's Disease
- Spinal Muscular Atrophy Type 1

It has been shown that colony characteristics, such as number of colonies and colony morphology [8] vary between patients. These characteristics should be used to optimize the timing of colony picking in automated systems. We use the SVCell Colony Analyzer recipe to characterize patient cells undergoing reprogramming. Fibroblasts from three disease patients and one healthy adult control were subject to Sendai virus-mediated reprogramming with Klf4, Oct3/4, Sox2 and c-Myc. At approximately 25 days after infection, whole-well image composites were acquired on the Nikon BioStation CT under phase contrast optics at 4x magnification. We applied the SVCell Colony Analyzer recipe to quantitatively characterize the colonies that formed for each patient. We find significant differences in colony counts, colony area sizes and texture measurements (error bars show 95% confidence intervals) between patients. In a real time system, this type of information should be used to assess the quality of the reprogramming run, and guide the timing of colony picking in a patient-specific fashion.
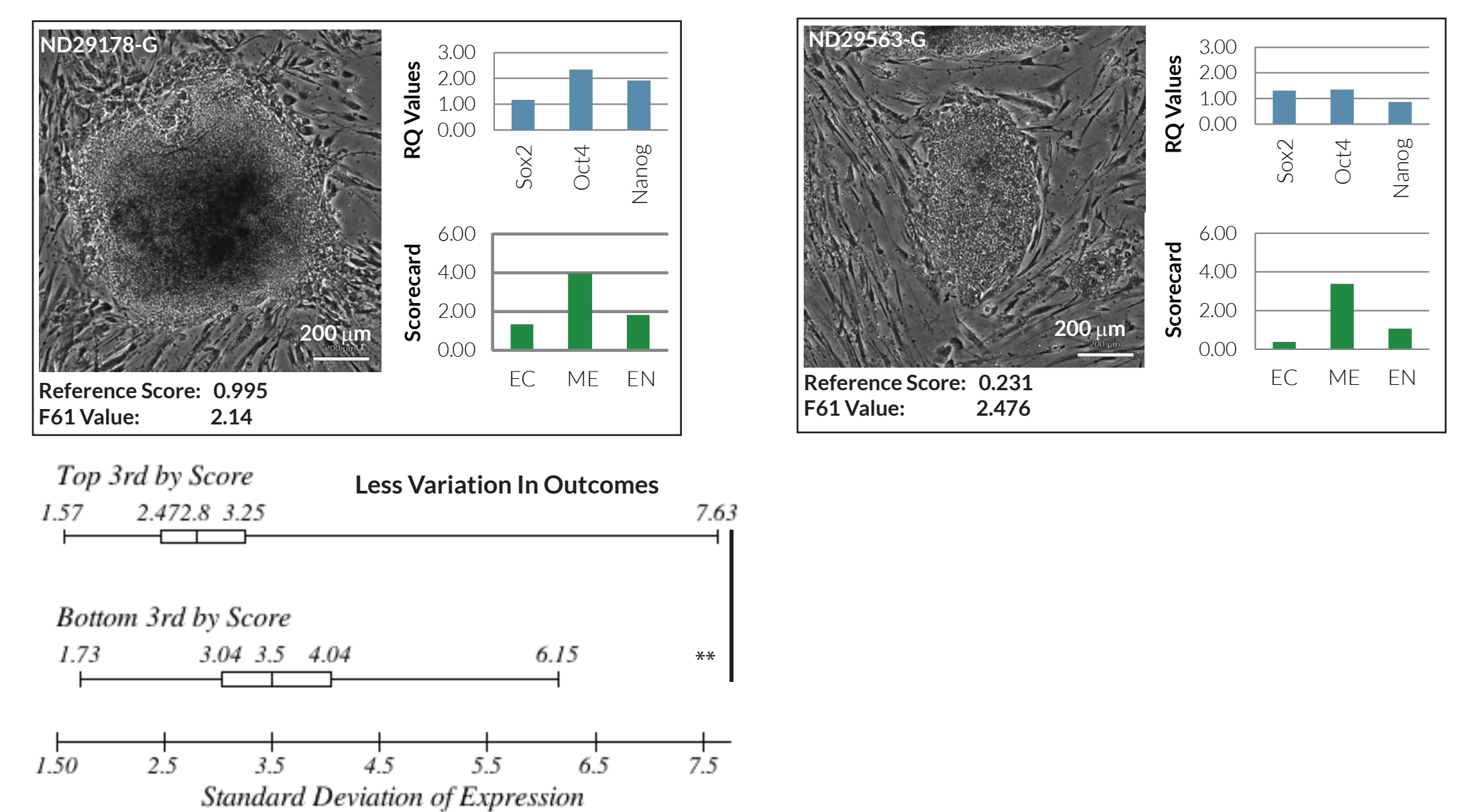
Real-Time iPSC Classification

iPSC Classifier Performance

		Ground Truth	
		Positive	Negative
Test	Positive	TP	FP
	Negative	FN	TN
Test	Positive	97	0	97
	Negative	14	111	125
		111	111	222

Key Feature "F61" Discrimination Power

- Fully reprogrammed colonies
- Negative examples

We developed a real-time iPSC colony classifier using a patient reference panel of 46 clones from 10 patients reprogrammed and imaged as described above, and having associated gene expression, immunocytochemistry (TRA-1-60) and pluripotency outcomes confirming their *bona fide* iPSC status. In addition, negative examples of partially reprogrammed colonies that are similar in size and shape but judged not to be fully reprogrammed on the basis of TRA-1-60 live staining and appearance were used. The rule is trained using a probabilistic, regulated decision tree, where the score is the normalized probability that the sample is fully reprogrammed[4] based on the reference panel.

A key morphometric feature, "F61", is revealed through machine learning. A high value indicates a relative degree of dark dynamic intensity along the colony border compared with the interior; high values reduce the probability that the colony is fully reprogrammed. A histogram of the "F61" feature shows good separation of the reprogrammed and negative classes. We then applied the scoring rule to the test data and results are shown in the contingency table. Overall tree performance on the testing data is 98.58% +/- 1.56% accuracy, with specificity of 100% +/- 0.06%.

Real-Time iPSC Quality Score

Reference Score: 0.995
F61 Value: 2.14

Reference Score: 0.231
F61 Value: 2.476

Less Variation In Outcomes

Top 3rd by Score

Bottom 3rd by Score

Standard Deviation of Expression

The Real-Time iPSC score can be used for ranking and selection from within a set of fully reprogrammed iPSC clones. We use the score to rank the confirmed iPSC clones, and compared the variation in pluripotency scorecard gene expression (EB assay) between the top and bottom third of test and training clones ranked by score. We see a significant reduction in variation between the top and bottom groups (p < 0.01 by Wilcoxon signed ranks test) indicating that use of the reference metric may improve clonal line selection and reduce process variations.

Conclusions

We have shown that the SVCell Colony Analyzer Recipe and Real Time iPSC Score can be used for patient-specific colony characterization including counts, area and texture data. This can be used to assess the quality of the reprogramming run and determine picking time. In addition, the Real Time iPSC score is successfully used to identify and rank fully reprogrammed clones, resulting in reduced variation between clones. These results show that SVCell could enable automated and clonal patient-specific cell generation in next generation robotic systems to standardize patient-specific cell line generation with reduced variability.

Acknowledgements

References

1. Lee JSJ (2015). Recipe based method for time-lapse image analysis. US patent number number 9,196,038; November 24.
2. Lee JSJ, Oh S (2007). Learnable object segmentation. US patent number number 7,203,360; April 10.
3. Lee JSJ, Alworth SV (2015). Progressive decision for cellular process selection. US patent number 9,123,120; September 1.
4. Lee JSJ (2006). Regulation of hierarchic decisions in intelligent systems. US patent number 7,031,948; April 18.
5. Lee JSJ (2015). Image guided protocol for cell generation. US patent number 9,173,909; November 3.
6. Alworth SV, Watanabe H, Lee JSJ (2010). Teachable, high-content analytics for live-cell, phase contrast movies. J Biomol Screen. Sep;15(8):968-77. PMID: 20639505.
7. Lin S, Foteno S, Satish S, Bhanu B, Talbot P (2010). Video bioinformatics analysis of human embryonic stem cell colony growth. J Vis Exp May 20; (39). PMID 20495527.
8. Paull D et al (2015). Automated, high-throughput derivation, characterization and differentiation of induced pluripotent stem cells. Nat Methods. Sep; 12(9):885-92.

SVCell RS

DRVision Technologies LLC