Reference metrics for human induced pluripotent stem cell colony selection

Samuel V. Alworth¹, Nicholas J. Collins²,³, Zakary A. Kenyon¹, Louis Plioco¹, Laurence Daheron²,³, Lee L. Rubin²,³, Chad A. Cowan²,³, and James S.J. Lee¹

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

Abstract

Induced pluripotent stem cells (iPSC) have become increasingly adopted for disease modeling, and have the potential to become a source of tissue for regenerative medicine. However, picking fully reprogrammed iPSC colonies can be unreliable, costly and time consuming.

We have created an iPSC scoring module that generates reference metrics from label-free phase contrast images taken on standard microscopes. The reference metrics indicate the likelihood that a colony morphologically resembles a reference panel of colonies known to be successfully reprogrammed. The reference metrics are created using supervised machine learning methods. These metrics could be used to automate colony picking or to ease implementation of reprogramming in research laboratories by providing a real-time, label-free indication of full reprogramming [1].

Reference Panel Creation

To develop the reference metrics, we created an image database showing the reprogramming of individual clones along with associated gene expression data, immunocytochemistry (TRA-1-60), and pluripotency outcomes. The database includes 46 lines from 10 patients, imaged every 6 hours for 4 weeks with two TRA-1-60 live stainings on the BioStation CT. Fibroblast samples from healthy donors and disease were subject to Sendai virus-mediated reprogramming with Klf4, Oct3/4, Sox2 and c-Myc.

Image & Feature Normalization

Images and measurements are normalized so that scoring may be reliable across imaging systems. Images are acquired from the BioStation CT, EVOS and Olympus IX71 systems. Images are normalized using image processing (background flattening) and histogram normalization. Robust measurements such as 5% trim mean are used [2], as well as self-normalized measurements (i.e. ratios). The dot plots show that these methods improve the correlation of measurements extracted from images of the same field taken on different systems.

Reference Metric Training

The colony scoring rule is created using images of colonies validated to be fully reprogrammed based on gene expression and pluripotency analysis, and also negative examples of 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. The rule is trained using a probabilistic, regulated decision tree, where the score is the normalized probability that the sample is fully reprogrammed [3] based on the reference panel. Overall rule accuracy on the training data is 97.07% +/- 2.31% accuracy, with specificity of 100% +/- 0.06%.

Reference Metric Validation

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%. Lastly, we compare 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.

Acknowledgements

The project described was supported in part by Grant Number R44HL106863 from the National Heart, Lung and Blood Institute. We would like to thank Yasujiro Kiyota and Nikon Corp. for their assistance with the BioStation CT on the project.

References