

# Automated Ranking of Stem Cell Colonies By Translating Biological Rules to Computational Models<sup>1</sup>

Adele Peskin  
NIST  
Boulder, CO 80305  
adele.peskin@nist.gov  
Michael Halter  
NIST  
Gaithersburg, MD 20899  
Michael.halter@nist.gov

Steve Lund  
NIST  
Gaithersburg, MD 20899  
steve.lund@nist.gov  
Anne Plant  
NIST  
Gaithersburg, MD 20899  
anne.plant@nist.gov

Ya-Shian Li-Baboud  
NIST  
Gaithersburg, MD 20899  
ya-shian.li-baboud@nist.gov  
Peter Bajcsy  
NIST  
Gaithersburg, MD 20899  
peter.bajcsy@nist.gov

## ABSTRACT

This paper addresses the problem of automating an image ranking process for stem cell colonies. We automate the manual process in a novel way: instead of fitting off-the-shelf image features and colony ranks to prediction models, we define a new feature set that uniquely characterizes the visual clues from images of the colonies and biological rules experts use to rank colonies from image data. Our automation considers several factors: the inconsistency of manually assigned stem cell colony ranks, the type of image segmentation to detect stem cell colonies (manual and automated), the type of image feature set (off-the-shelf vs. custom designed), and an underlying relationship between input colony features and output stem cell colony ranks (linear and non-linear). The novelty of our work lies in automating stem cell colony ranking while preserving the connection between visually perceived quality characteristics of stem cell colonies, and image colony features combined with a computational prediction model. The main contribution of our work is in demonstrating the benefits of direct interpretation of biological rules to automation of stem cell colony ranking. We also outline a method for establishing relationships between the commonly used Haralick features and our custom-designed features.

## Categories and Subject Descriptors

I.4.7 [Image Processing and Computer Vision]: Feature Measurement

## General Terms

ACM BCB 2014 Topic: Biomedical Image Analysis

## Keywords

Image classification, Colony pluripotency prediction, Automation

## 1 Introduction

Maintaining stem cell lines currently requires manual selection of colonies for passage based on inspection under the microscope.

Even when expert biologists have defined and agreed upon a set of biological rules to rank stem cell quality, the selection process is inconsistent. One approach to increasing consistency is via computer-based automation. Automation is typically achieved by (1) adapting off-the-shelf image feature software, (2) building and validating a model to predict colony ranks from image features, and (3) predicting ranks from image feature measurements during the actual ranking process.

There are several drawbacks to such an approach. The typical approach does not specifically include features that experts look for in high quality colonies. Furthermore, there is no visual connection between off-the-shelf image features and stem cell quality. Without a biological connection between specific features and specific image characteristics, it is not possible to understand if the feature set has been fully defined. Additionally, off-the-shelf features may contain extraneous features that are included in complex prediction models (e.g., random forests and decision tree models), which may affect the outcome of such models and the accuracy of automated colony rank prediction.

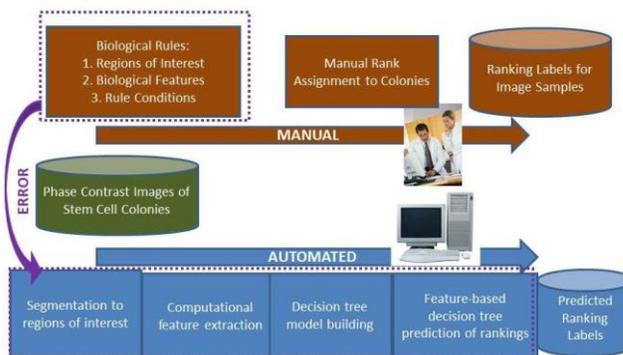


Figure 1: Overview of manual versus automated assignment of ranking labels.

Our motivation is to build a feature set that completely describes all characteristics of the pluripotency captured by experts in a set of biological rules. We assume that the biological rules are quantifiable in an image, and that they lead to a set of only relevant features. Our approach to defining a feature set to automate colony ranking addresses several of the drawbacks

<sup>1</sup> This contribution of NIST, an agency of the U.S. Government, is not subject to copyright.

mentioned above, and is illustrated in Figure 1. It shows the flow of computational steps in our automated process (bottom of Figure 1) that replace the manual steps (top of Figure 1).

The input for developing our automated process includes five biological rules (see Table 2) and 481 phase contrast images of stem cell colonies imaged at 10X magnification and ranked by two experts. The five bio-rules were broken down and mapped into 16 image features extracted over image segments obtained via automated segmentation. For comparison to more traditional methods of automation, a second set of 45 off-the-shelf features were used: 12 Haralick features, 30 wavelet features [1], area, perimeter, and circularity. Both feature sets served as input to a linear (logistic LASSO) and two non-linear (decision tree and random forest) models. All models were validated by a leave-one-out resampling technique. Table 1 summarizes all variables considered for evaluating a computational model for predicting stem cell colony ranks.

**Table 1: A summary of variables for evaluating a computational model for predicting stem cell colony ranks.**

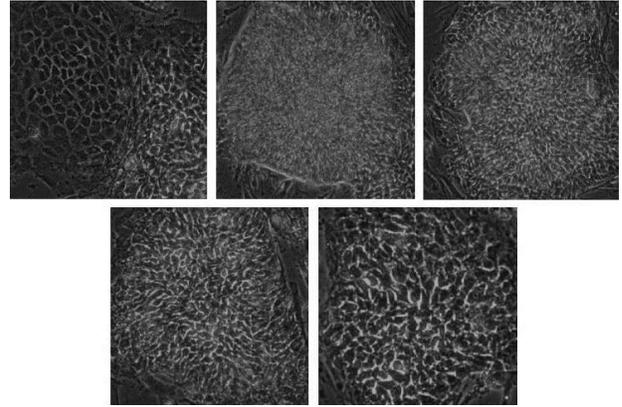
| Experts' Ranks                       | Image Segmentation | Image Features   | Prediction Model  |
|--------------------------------------|--------------------|--|---|
| <b>Consistent</b><br>383 out of 481  | <b>Manual</b>      | <b>Off-the-shelf</b><br>12 Haralick features,<br>30 wavelet features | <b>Linear</b><br>Logistic LASSO (least absolute shrinkage and selection operator) |
| <b>Inconsistent</b><br>98 out of 481 | <b>Automated</b>   | <b>Derived from biological rules</b><br>16 custom features           | <b>Non-linear</b><br>Decision tree<br>Random forests                              |

Our *objectives* are to (a) incorporate bio-rules and experts' ranks into a computational prediction model, (b) compare the accuracy of a prediction model for off-the-shelf image features in linear and non-linear prediction models against the accuracy of features based on bio-rules in a prediction model, and (c) explore relations between custom-designed and off-the-shelf features. The *novelty* of our work lies in automating stem cell colony ranking while preserving the connection between visually perceived characteristics of stem cell colonies with a computational prediction model. This is achieved by directly translating biological rules to computational image features and a prediction model.

## 2 Input Data

Phase contrast images of stem cell colonies were acquired at 10X magnification 24 hours after seeding. The images were provided to us from the Lieber Institute for Brain Development together with ranks of 481 stem cell colonies assigned by two experts and five biological rules that were established before ranking. The biological rules were a summary of the experts' ranking decision process. Rankings were divided by us into a Low pluripotency group (Low refers to 1,2 and 3 assigned by experts) and a High pluripotency group (High refers to 4 and 5 assigned by experts). Figure 2 shows examples of stem cell colonies in the Low and

High groups, as well as the 1 to 5 ranks assigned by experts. Table 2 presents the five biological rules and pluripotency group labels assigned according to each rule.



**Figure 2: Examples of stem cell colony images: top: ranks 1,2,3 (Low); bottom: ranks 4,5 (High).**

## 3 Previous Work

The closest previous work on predicting stem cell colony growth by image processing and pattern analyses have been reported in [2], [3], [4]. Bradhurst in his Master's thesis [2] reports 44 previously unseen images of bone marrow stem cell colonies that are classified with 88 % accuracy in comparison to 72 % accuracy of manual classification. The key emphasis is on normalization and colony segmentation. Jeffreys in his Master's thesis [3] focused on image texture features to discriminate between differentiated and undifferentiated stem cells while using a support vector machine prediction model and Kullback-Leibler distance [4]. The texture features are selected to satisfy three criteria: textural homogeneity, textural tightness, and border sharpness and circularity.

**Table 2: Biological rules**

| IF <condition is met>   | THEN <rank> | THEN <pluripotency group > |
|---|-------------|----------------------------|
| (a) distinct margins, homogeneous phase-bright internal cells, small inter-nuclear distance, epithelial characteristics (elongated cells throughout) (b) distinct margins, homogeneous phase-bright internal cells, small inter-nuclear distance, epithelial characteristics at edge only | 5           | High                       |
| Intermediate  | 4           | High                       |
| distinct margins, heterogeneous phase bright/phase dark cells, variable inter-nuclear distance  | 3           | Low                        |
| Intermediate  | 2           | Low                        |
| fully differentiated, indistinct margins, heterogeneous internal cell types, large inter-nuclear distance   | 1           | Low                        |

In comparison to the previous work, we focus more on the selection of the individual features of our feature set for characterization of the colonies including automated colony segmentation. The overall challenges in automating the process of stem cell colony ranking (classification) lie in modeling inconsistent ranks assigned by experts, segmenting colonies from images, and selecting features and prediction models optimally. We briefly describe related work for each challenging problem.

### 3.1 Modeling inconsistent ranks of stem cell colonies

Inconsistent ranks, diagnoses and expert’s interpretations occur often during examinations of medical records and medical images [5]. One approach to modeling the inconsistencies is to view the ranks as a random variable. In our case, the five ranks were converted into two pluripotency groups that enable us to describe the inconsistency with a binomial distribution. Let  $Y_{ij}$  be 1 (pluripotent) if expert  $j$  assigned a score of 4 or 5 to image  $i$  and let  $Y_{ij}$  be 0 (not pluripotent) if expert  $j$  assigned a score of 1, 2 or 3. We assume that for colony image  $i$ , the proportion of experts that would declare the imaged colony pluripotent is  $p_i$ . Expert scores are assumed to be independent from one another, conditional on the colony specific pluripotency proportion. That is, we assume  $Y_{ij} \sim \text{Binomial}(n=I, p_i)$  where  $n=I$  refers to the number of agreeing pluripotency labels among  $N$  trials. We seek to model the proportions  $p_i$  as a function of features computed from colony image  $i$ .

### 3.2 Segmentation

The input images of stem cell colonies are difficult to segment due to surrounding epithelial feeder cells in the growth medium. Many different approaches were tried using ImageJ open source software utilizing routines for gradient edge detection, Gaussian blurring kernels to remove the feeder cells, and thresholding [6]. The most successful of all ImageJ routines found the colony boundaries but could not separate touching colonies. Areas of cells inside of a colony that were visually very distinct from the rest of the colony were also hard to capture (see for example, the rank 1 colony in Figure 2). We evaluated multiple methods against manual segmentation. Due to unsatisfactory accuracy of existing methods we decided to develop a custom segmentation technique.

### 3.3 Feature selection

Several surveys have been written on feature selection either focusing on the algorithmic aspects [7], [8], [9] or bio-domain specific aspects [10]. Previous feature selection methods can be categorized as supervised or unsupervised. The types of feature selection algorithms can be classified as those using an initial feature filter, or those using machine learning to select a subset of relevant features [7]. The key challenges in a feature selection problem are (a) in searching a large space of features ( $2^n$  where  $n$  is the number of features [11]), (b) in coping with feature dependencies and degrees of relevancy (relevant, irrelevant, redundant, and noisy features [12]), and (c) in having sufficient and realistic reference data to evaluate algorithmic performance (measured or synthetic data [13]). In our work, we approached the feature selection problem by translating biological rules into computational image features. Although this translation can be

viewed as a one-to-many mapping, it reduces the feature search space significantly while preserving the connection between visually perceived quality characteristics of stem cell colonies in biological rules and image colony features combined with a computational prediction model.

## 4 Modeling Manual Stem Cell Colony Ranking Process

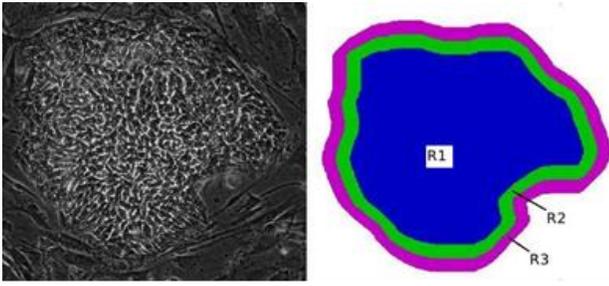
### 4.1 Translating Biological Rules

The five bio-rules in Table 2 were broken down and mapped into 16 computational image features. Our work in translating these five bio-rules into image features involves three steps:

- (1) image segmentation of stem cell colonies into regions of interests,
- (2) word-for-word translation of the attributes found in biological rules (in an IF x THEN y format) into image characteristics, and
- (3) identification of computational operations that quantify image characteristics.

**Table 3: Translation of attributes into regions of interests (R1,R2,R3), image characteristics, and computational features.**

| Attributes of the IF-THEN biological rules | Region of Interest       | Image Characteristics to Measure  | Quantitative Computational Feature  |
|--|--------------------------|---|---|
| Distinct margins at colony border          | border between R2 and R3 | Sharp differences between R2 and R3; Homogeneity of border  | Edge quality, Edge contrast, Entropy ratio (Fgbg)   |
| Phase-bright internal cells                | R1                       | Pixel intensity contrasts in R1; Homogeneity of R1; Textures in R1                                    | Local std., Entropy, Histogram of local std. (Hg1, Hg2), Fraction of dark/light transitions (areaRatio) |
| Inter-nuclear distance                     | R1                       | Size of individual cells in R1  | Holes/Area  |
| Epithelial characteristics                 | R2, R3                   | Appearance of elongated cells in R2; Homogeneity of R2; Epithelial cells transverse at colony borders | Stats of shapes in R2 and R3 regions (Edge circles, Edge ratio, E1, E2, ep-slope)                       |



**Figure 3: Three regions of interest: exterior margins (R3), interior margins (R2), and center region (R1).**

Table 3 summarizes the process of mapping the five biological rules (“IF<condition is met>THEN<colony rank>”) and parsing the “<condition>” parts into attributes that correspond to image characteristics and their corresponding computational features evaluated over image sub-regions, which represent cell colonies. Figure 3 shows the three regions of interests R1, R2 and R3 that can be derived from stem cell colony segmentation. We describe the stem cell colony segmentation and the computational features presented in Table 3 (right most column).

#### 4.1.1 Segmentation

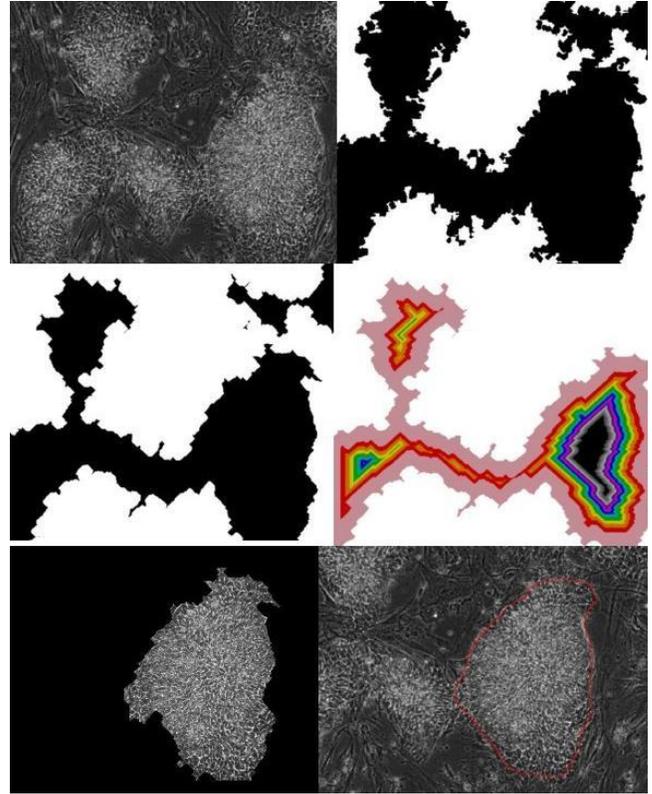
We have developed an automated segmentation technique consisting of the following steps (as illustrated in Figure 4):

- Step 1: In a 20 x 20 neighborhood around each pixel, we compute both an average intensity value and a local intensity standard deviation value. Average values and standard deviations over a whole image provide levels of intensity and contrast to classify each pixel.
- Step 2: We collect pixels with either high intensity (> 0.5 standard deviation above image mean) or high contrast (> 1.0 standard deviation above image mean), and fill in holes.
- Step 3: We apply morphological erosion (kernel size = 2) to separate colonies from one another and from epithelial cells surrounding them.
- Step 4: Finally, we split a colony based on the geodesic distance of each interior pixel from a colony boundary [14].

Figure 4 shows the resulting segmentation mask and corresponding manually segmented mask. Once a stem cell colony is automatically segmented, the three image regions of interest shown in Figure 3 are found by morphological erosion (to find R1) or dilation (to find R3) of the segmentation mask (kernel size = 50).

Manually and automatically segmented colony masks are compared using the Dice similarity metric [15]. The Dice metric (D) reports an index that varies between 0.0 (no match), and 1.0 (a perfect match). Equation (2) presents the Dice metric computation:

$$D = 2A / (2A + B + C), \quad (1)$$

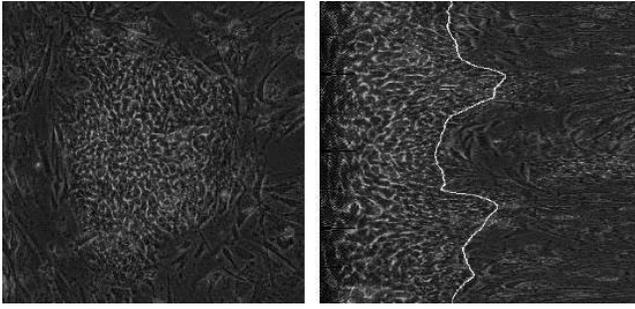


**Figure 4: Top: original image; clusters of high intensity pixels. Middle: clusters after erosion; clusters colored by distance to colony border; Bottom: final automated vs. manual segmentation masks.**

where  $A$  is the number of image pixels shared between the two samples and  $B$  and  $C$  are pixels found only in the first and second samples respectively. An index of 0.7 and higher qualifies as substantial similarity [15].

#### 4.1.2 Derived Quantitative Computational Features

Many computational features in Table 1 are computed by incorporating well known techniques such as thresholding, local standard deviation measurements, and entropy calculations. Colony edge features are derived by linearizing the colony boundary using a polar coordinate transformation (see Figure 5). Each image is transformed to polar  $(r, \theta)$  coordinates with respect to the center of the colony, where the  $x$  axis represents the distance of a pixel to the colony center and the  $y$  axis represents an angle  $\theta$  about the center,  $(\theta: [0, 360])$  degrees). Additional details about computational feature computations are provided in Appendix A.



**Figure 5: top: Sample colony image, bottom: image transformed to polar coordinates, colony boundary in white.**

## 4.2 Prediction Modeling

We included two non-linear prediction models (decision tree and random forest) and one linear prediction model (logistic LASSO - least absolute shrinkage and selection operator). The decision tree prediction model intuitively fits our effort to translate biological rules in a form “IF<condition>THEN<rank assignment>” statements to a computational model. All models were validated by a leave-one-out cross validation.

## 5 Experimental Results

### 5.1 Segmentation

Manual segmentation was performed using an ImageJ plug-in, Segmentation\_manual\_514 [6]. The accuracy of the developed segmentation technique is evaluated per stem cell colony in terms of the Dice index presented in Equation (2), comparing automated with manually generated colony masks. According to Dice indices computed over all segmented colonies, 96.25 % of 481 segmented colonies are above 0.7 (substantial similarity), and 88.96 % are above 0.8. Figure 6 shows the Dice indices for all 481 colonies.

### 5.2 Performance of Prediction Models

We compare the accuracy of our automated bio-rule based ranking system with the accuracy of a system using off-the-shelf image features. The evaluations are performed for both linear and non-linear prediction models (see Table 1).

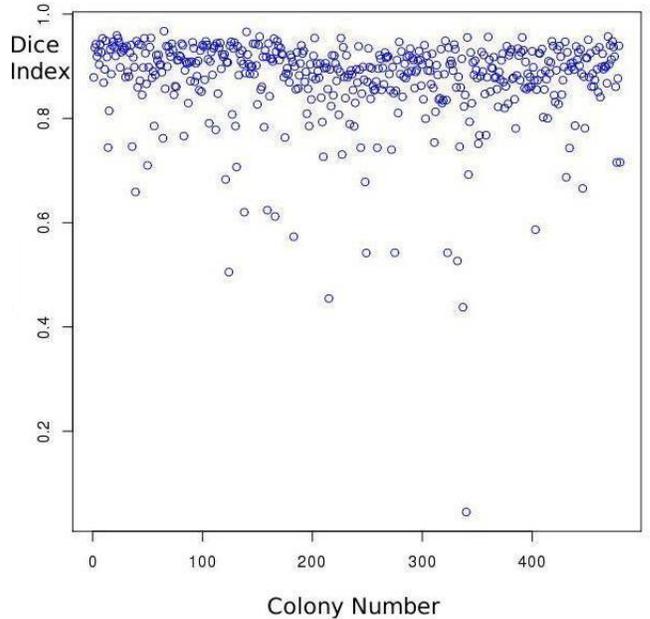
The comparison of prediction models is slightly complicated by the coarse reference information the experts provide. That is, as evidenced by the frequency with which the experts disagree with each other (66 of 449 colonies scored by both experts), their scores are not absolute truth and should be considered random. As such, even when both experts classify a colony as pluripotent (or not), the true pluripotency state of that colony remains unknown. To make use of the at most two (binary) scores per colony as “reference” when assessing a given model’s performance, we characterize how likely the reported scores from colony  $i$  were to occur under that model’s fit to the log-likelihood from the binomial distribution:

$$\ell_i^{(\text{mod}, \text{set})} = \sum_j Y_{ij} \ln p_i^{(\text{mod}, \text{set})} + (1 - Y_{ij}) \ln(1 - p_i^{(\text{mod}, \text{set})}), \quad (2)$$

where  $p_i^{(\text{mod}, \text{set})}$  denotes the pluripotency probability or proportion of a large number of experts that would score colony  $i$  as pluripotent, as estimated by applying modeling technique  $\text{mod} = \{\text{Logistic LASSO, Random Forest, Decision Tree}\}$  to feature set  $\text{set} = \{\text{Bio Rules, Off Shelf, All}\}$ , and  $Y_{ij}$  is as defined in Section 3.1.

The results reported below correspond to an analysis in which the  $p_i^{(\text{mod}, \text{set})}$  were estimated via leave-one-out cross validation. Similar results were obtained when instead  $p_i^{(\text{mod}, \text{set})}$  was estimated using 10-fold and 20-fold cross-validation, respectively, illustrating a robustness of our findings to the size of the testing sets used during cross validation.

Within each modeling method, following fitting, we conduct pairwise comparisons of the feature sets by tallying across the colonies, to find which set produced the largest likelihood of the reported scores. (Model fits for which the observed scores are more likely are better.)



**Figure 6: Dice similarity index measurements comparing manual and automatic segmentation masks.**

Figures 7-9 show a summary of pairwise comparisons of the feature sets within each model. Figure 10 reports the median cross validation (CV) log likelihoods for each combination of model and feature set. The median likelihood of our bio-rule feature set is statistically significantly higher than the off-the-shelf feature set using non-linear prediction models (decision tree and random forest models). Sets are statistically significantly higher where p-values from a sign test for equality of medians (shown in Figures 7-9) are less than  $p=0.05$ . For example, looking at the results of the random forest model, Figure 8 shows a p-value of  $3.71e-05$

comparing our bio-rules feature set with the off-the-shelf feature set and  $9.62e-05$  comparing our bio-rules set with all 3 features taken together. This is also an illustration of the effect of extraneous features on the output of the model. Within each pairwise comparison shown in Figures 7-9, the number of colonies for which each feature set produced the larger likelihood is reported underneath the respective label. A p-value from a sign test for equality of medians is reported at the top of each pairwise comparison. The green section characterizes the number of colonies for which the two methods produced equal likelihoods.

### 5.3 Relating Off-The-Shelf and Biologically Motivated Features

We have shown the benefits of direct interpretation of biological rules into image features in comparison to the more traditional approach of using off-the-shelf feature sets. However, there is a trade-off between the labor needed to custom-design image features according to biological rules and the loss of biological interpretation of features in off-the-shelf sets. For this reason we were motivated to explore whether one can relate off-the-shelf Haralick features to our custom-designed features, so that the off-the-shelf features could be used in the future for automation of colony ranking based on similar biological rules.

#### Logistic LASSO

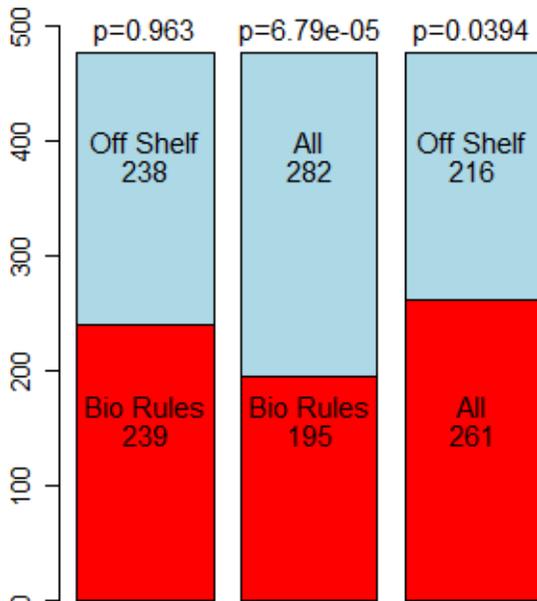


Figure 7: Barplots displaying pairwise comparisons of the feature sets using the Logistic LASSO model.

#### Random Forest

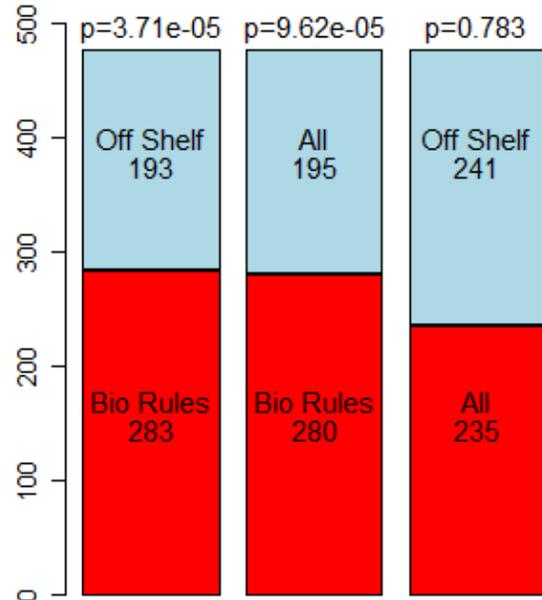


Figure 8: Barplots displaying pairwise comparisons of the feature sets using the Random Forest model.

#### Decision Tree

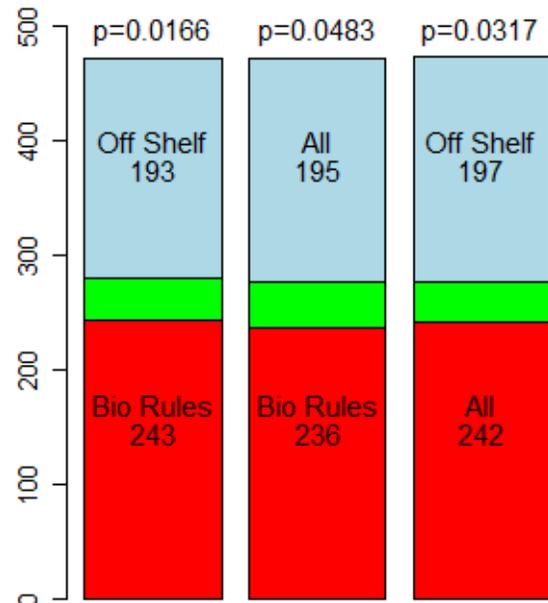


Figure 9: Barplots displaying pairwise comparisons of the feature sets using the Decision Tree model.

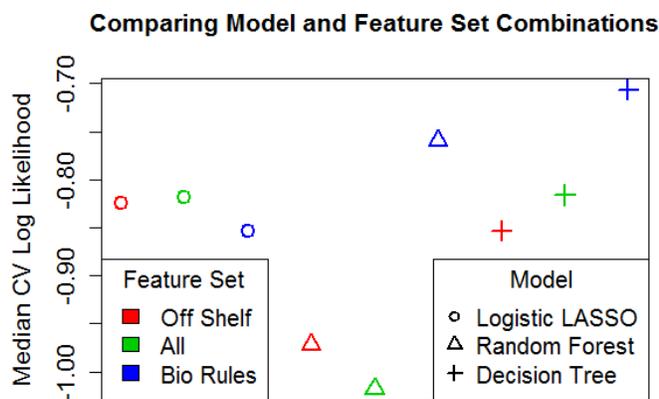


Figure 10: Median cross validation (CV) log likelihoods for each combination of model and feature set.

Table 4 shows the statistical correlation between each Haralick feature and the features of our biologically motivated feature set. Mathematical descriptions of the Haralick features can be found in [16] with the indices following the definitions of Angular Second Moment (h1), Correlation (h2), Contrast (h3), Sum of the Squares of Variance (h4), Inverse Difference Moment (h5), Sum Average (h6), Sum Variance (h7), Sum Entropy (h8), Entropy (h9), Difference Variance (h10), Difference Entropy (h11), and Information Measure of Correlation 1 (h12). Information Measure of Correlation 2 is not included because these values were all negligible when applied to the stem cell colony images. The co-occurrence matrix needed for Haralick features was computed along the horizontal orientation with pairs of pixels apart by one pixel.

Table 4: A correlation matrix between the Haralick features and the biological rule derived features. The high/low values of columns containing high correlations are highlighted.

| Feature     | h1    | h2    | h3    | h4    | h5    | h6    | h7    | h8    | h9    | h10   | h11   | h12   |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Size        | -0.58 | -0.08 | -0.39 | -0.29 | -0.35 | 0.39  | 0.34  | -0.01 | 0.74  | -0.16 | 0.03  | 0.86  |
| Local.std   | 0.03  | 0.76  | 0.08  | 0.92  | -0.07 | -0.22 | -0.07 | 0.64  | -0.07 | 0.82  | 0.68  | -0.41 |
| Perim       | -0.53 | -0.16 | -0.23 | -0.31 | -0.21 | 0.3   | 0.24  | -0.04 | 0.68  | -0.22 | -0.08 | 0.82  |
| Circularity | -0.13 | 0.29  | -0.3  | 0.17  | -0.32 | 0.26  | 0.29  | 0.24  | 0.07  | 0.24  | 0.32  | -0.06 |
| Entropy     | -0.89 | 0.54  | -0.33 | 0.44  | -0.69 | 0.6   | 0.68  | 0.85  | 0.86  | 0.42  | 0.67  | 0.53  |
| Fgbg        | -0.45 | 0.4   | -0.21 | 0.35  | -0.38 | 0.3   | 0.36  | 0.56  | 0.34  | 0.35  | 0.46  | 0.09  |
| Edge qual   | 0.28  | 0.24  | 0.08  | 0.31  | 0.15  | -0.3  | -0.25 | 0.07  | -0.26 | 0.29  | 0.17  | -0.34 |
| Hg1         | 0.07  | 0.7   | -0.01 | 0.79  | -0.04 | -0.15 | -0.01 | 0.47  | -0.09 | 0.75  | 0.58  | -0.35 |
| Hg2         | 0.07  | 0.66  | -0.02 | 0.75  | -0.04 | -0.13 | 0     | 0.43  | -0.08 | 0.71  | 0.54  | -0.31 |
| E1          | 0.1   | 0.33  | -0.01 | 0.37  | -0.11 | -0.05 | 0.02  | 0.29  | -0.19 | 0.33  | 0.32  | -0.38 |
| E2          | -0.06 | -0.4  | 0.02  | -0.46 | 0.12  | 0.06  | -0.01 | -0.37 | 0.11  | -0.41 | -0.41 | 0.33  |
| area ratio  | -0.46 | -0.2  | -0.27 | -0.38 | -0.48 | 0.72  | 0.66  | 0.1   | 0.39  | -0.34 | -0.06 | 0.38  |
| Edge cont   | -0.61 | -0.12 | -0.37 | -0.34 | -0.61 | 0.8   | 0.74  | 0.19  | 0.56  | -0.28 | 0.05  | 0.52  |
| Edge ratio  | -0.52 | 0.28  | -0.27 | 0.18  | -0.49 | 0.44  | 0.45  | 0.43  | 0.53  | 0.19  | 0.37  | 0.38  |
| Ep-slope    | 0.03  | 0.01  | -0.03 | -0.01 | -0.07 | 0.03  | 0.02  | 0.01  | -0.03 | -0.01 | 0.03  | -0.06 |
| Edge circ1  | 0.59  | 0.04  | 0.3   | 0.21  | 0.3   | -0.35 | -0.32 | -0.06 | -0.73 | 0.11  | -0.05 | -0.81 |
| MAX         | 0.59  | 0.76  | 0.3   | 0.92  | 0.3   | 0.8   | 0.74  | 0.85  | 0.86  | 0.82  | 0.68  | 0.86  |
| MIN         | -0.89 | -0.4  | -0.39 | -0.46 | -0.69 | -0.35 | -0.32 | -0.37 | -0.73 | -0.41 | -0.41 | -0.81 |

Based on the high absolute correlation results in Table 4 (highlighted in yellow), one could approximate the custom-designed features with a subset of Haralick features based on input biological rules. The highest absolute correlation of 0.92 was found between h4 (Sum of the Squares of Variance) and local standard deviation. The smallest absolute correlation of 0.39 was

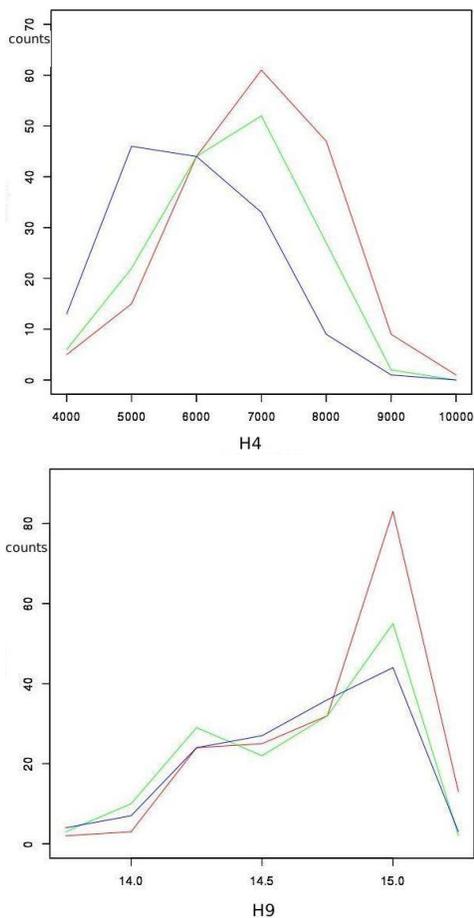
found between h3 (Contrast) and size. However, the prediction modeling results indicate that non-linear models are more accurate than linear models and hence discovering feature relations has to be conducted for each non-linear model separately.

In addition, we have observed that some of the Haralick features vary with imaging conditions more than with colony attributes. For example, one can contrast the variations of Haralick feature h4 (Sum of the Squares of Variance) and h9 (Entropy), obtained from colonies imaged over three wells as shown in Figure 11. The values for h4 vary over the different imaging conditions for the 3 different wells, whereas values for h9 remain similar over the 3 wells.

## 6 Conclusions

We demonstrated the benefits of using biological rules to select image features for the automation of stem cell colony ranking. Based on our analyses of segmentation and prediction models, the performance of automation, measured with well-defined metrics, benefited from additional information presented as a set of biological rules. We concluded this based on the comparison of using our new feature sets with off-the-shelf features including Haralick texture features and wavelet features. The analyses also suggest that the model we are looking for may not be a strict linear model. The improvement we saw in our model was seen only in the non-linear (random forest and decision tree) models. Using these 2 models we showed an improvement in the outcome of the model using a selected set of features over the outcome of either using only off-the-shelf features, or using a combination of all feature sets together. Finally, we attempted to establish relations between the custom-designed features and off-the-shelf features based on correlation in order to minimize future labor investments into custom-development of image features.

In the future, we plan to further investigate the process of establishing biological rules for stem cell colony ranking. There is a need to incorporate biological knowledge into automated systems and to decouple sources of errors contributing to the automations based on biological rules. Having a biological connection between image features and the visual characteristics that biological experts are looking for provides additional value to biologists in understanding stem cell colonies.



**Figure 11: Top: Haralick feature h4 and Bottom: Haralick feature h9, computed over colony masks from well 1 (red), well 2 (green), and well 3 (blue).**

## 7 Acknowledgment

This work has been supported by NIST. We acknowledge Dr. Daniel Hoepfner and his colleagues at the Lieber Institute for Brain Development for providing the image data and the corresponding colony rankings. We acknowledge the team members of the computational science in biological metrology project at NIST for providing invaluable inputs to our work.

## 8 Disclaimer

Commercial products are identified in this document in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products identified are necessarily the best available for the purpose.

## 9 References

- [1] K. Huang and R. F. Murphy, "Boosting accuracy of automated classification of fluorescence microscope images for location proteomics.," *BMC Bioinformatics*, vol. 5, p. 78, Jun. 2004.
- [2] C. J. Bradhurst, "Monitoring Mesenchymal Stem Cell Cultures Using Image Processing and Pattern Recognition Techniques," Queensland University of Technology, Brisbane, Queensland, 2010.
- [3] Mangoubi, R., Jeffreys, C., Copeland, A., Desai, M., Sammak, P., Non-invasive Image Based Support vector Machine Classification of Human Embryonic Stem Cells, 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007.
- [4] R. Mangoubi, C. Jeffreys, A. Copeland, M. Desai, P. Sammak, and M. A. Massachusetts, "Non-Invasive Image Based Support Vector Machine Classification of Human Embryonic Stem Cells," in *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2007, pp. 284–287.
- [5] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: the kappa statistic.," *Fam. Med.*, vol. 37, no. 5, pp. 360–3, May 2005.
- [6] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: an open-source platform for biological-image analysis.," *Nat. Methods*, vol. 9, no. 7, pp. 676–82, Jul. 2012.
- [7] L. C. Molina, L. Belanche, and A. Nebot, "Feature selection algorithms: a survey and experimental evaluation," in *2002 IEEE International Conference on Data Mining*, 2002, pp. 306–313.
- [8] A. L. Bluma and P. Langley, "Artificial Intelligence Selection of relevant features and examples in machine.," *Elsevier Artif. Intell.*, vol. 97, no. 97, pp. 245–271, 1997.
- [9] H. Liu, S. Member, L. Yu, and S. Member, "Algorithms for Classification and Clustering," vol. 17, no. 4, pp. 491–502, 2005.
- [10] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics.," *Bioinformatics*, vol. 23, no. 19, pp. 2507–17, Oct. 2007.
- [11] P. Bajcsy and P. Groves, "Methodology For Hyperspectral Band Selection," *Photogramm. Eng. Remote Sens. J.*, vol. 70, no. 7, pp. 793–802, 2004.
- [12] H. Peng, F. Long, and C. Ding, "Feature Selection Based on Mutual Information : Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [13] A. Jain and D. Zongker, "Feature selection: evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, 1997.
- [14] J. Chalfoun, M. Majurski, A. Dima, C. Stuelten, A. Peskin, M. Brady, FogBank: A Single Cell Segmentation across Multiple Cell Lines and Image Modalities, *BMC Bioinformatics*, under review.

- [15] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [16] NIST, "Image Texture Feature Definitions," *Web page*, 2014. [Online]. Available: [https://isg.nist.gov/deepzoomweb/stemcellfeatures#\\_RefHeading\\_909\\_1766305140](https://isg.nist.gov/deepzoomweb/stemcellfeatures#_RefHeading_909_1766305140). [Accessed: 16-Mar-2014].

## Appendix A: Translation of Biological Rules

### Distinct margins at colony border; homogeneous border

Pluripotent colonies have distinct colony borders, which we translated into pixel intensity differences at the R2-R3 border, and pixel entropy differences at the border. After polar transformation, each colony border is represented by a nearly linear region on the transformed image. For each border pixel we found the ratio of the five inner pixels to the five outer pixels. The edge quality feature is an average of all of these ratios. We also computed the entropy of pixel intensities 50 pixels inside and 50 pixels outside the R2-R3 border, and the Fgbg (foreground/background) features is the ratio of these two entropies. Homogeneous border is interpreted to mean consistent pixel intensities at the R2-R3 border. An Otsu thresholding provided a threshold value for most of the bright cells of each colony. Holes in the resulting mask are filled, leaving most of the unfilled pixels near the colony border. The edge contrast feature is a ratio of the filled area to the total area; the closer to 1.0, the smoother the edge.

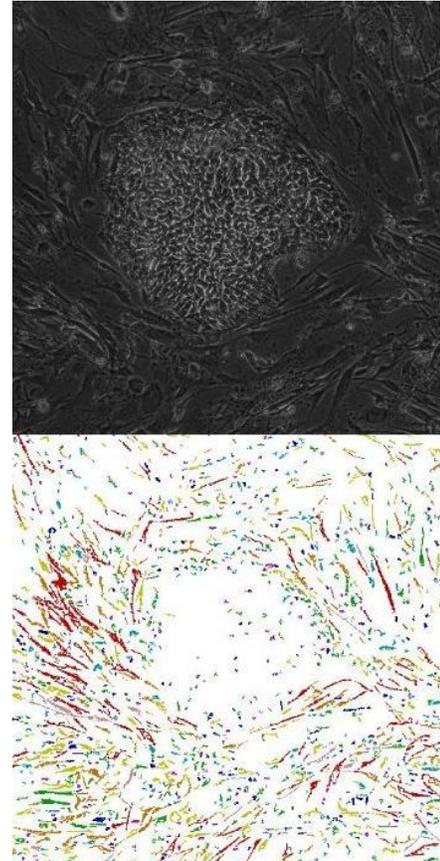
### Phase-bright internal cells; homogeneous region

Pluripotent colonies also are characterized by contrasting intensities between the center and the outsides of individual cells within the colony. We looked for very high contrast in pixel intensity at the level of a single cell, which is approximately a 20 x 20 pixel area on our 10X images. For each colony pixel, we found a local standard deviation of pixel intensity in this small neighborhood. We average the local standard deviation values over the whole image to find the fraction of pixels in the mask that are 3 standard deviations above the image average. This fraction is our local standard deviation (local std) feature. (Below we also use the local std to measure contrast at the colony edge.) Because the cells are round, the phase bright intensities will create an expected textural pattern on high-scoring colonies, so we compute the entropy of the colonies to quantify this texture. We also measure the phase-bright characteristic with the overall ratio of dark to light pixels, performing Otsu segmentation on each original image to threshold between light and dark, and then finding the ratio of the resulting thresholded mask to the entire mask (denoted as areaRatio feature).

To measure phase-bright homogeneity, we look at the similarity in local std. values in R1. We separate the pixels into different levels, determined by how many standard deviations they lie above the mean value for the image, five levels in  $0.5 * \text{stddev}$  increments. Groups of pixels are ordered by how many pixels are in each group. One feature (Hg1) measures the fraction of pixels in the groups with the largest number of pixels:  $(\text{highest} + 0.5 * \text{second highest} - 0.25) / 0.75$ . A second feature (Hg2) is the standard deviation of the number of pixels in each group.

### Inter-nuclear distance

Size of individual cells in a colony indicates the state of that colony, and we measured the ratio of internal dark cell pixels to edge pixels in the colony as a measure of the size of an average individual cell. Again we used Otsu segmentation to separate light and dark pixels, and measured the fraction of the pixels in the resulting mask without filling holes, for a feature called holes/area. The smaller the cells, the more cell edge pixels we should see, and the larger that ratio.



**Figure 12: Sample colony image with dark pixel clusters colored by roundness, low to high: red, orange, yellow, green, blue, and purple.**

### Elongated cells along colony edge; epithelial cells transverse to R2-R3 border

For the edge cell characteristics, we identify both the elongated edge cells and the outlying epithelial cells starting with the same technique. Dark clusters of pixels are isolated by selecting pixels at least  $3/4$  of a standard deviation below the image mean. Each cluster is labeled based on its roundness. Cluster shape analysis is performed before the image is transformed to polar coordinates. The elongated cells just inside the R2-R3 border appear as small very round clusters (blue and purple in Figure 12), and the epithelial cells outside the R2-R3 border appear as more linear clusters (red and yellow in Figure 12).

To quantify the elongated cells inside the R2-R3 border, we define a feature we call edge-circles. After transforming the images to polar coordinates as shown in , we see that the area just inside the border is filled with the round clusters from the original

image. Edge-circles is the measure of what fraction of the area just inside (within 50 pixels of) the R2-R3 border contains these cells. A feature called ep-slope measures the alignment of the border with epithelial cells 50 pixels outside the R2-R3 border, an average of the slope of each cluster in that region. Slopes are measured in the transformed image, looking for linear clusters.

Image features E1 and E2 measure the edge homogeneity, using local std values at the colony edge. E1 is the fraction of high contrast pixels (local std more than half a standard deviation above the mean value) for the image. E2 quantifies how low the areas of lowest contrast border areas are, a value based on the lowest value of the local std measured at the border. Another measure of the homogeneity is quantified using a feature called edge-ratio. Here the size of the largest cluster produced from Otsu thresholding is compared with the size of the colony mask. The more homogeneously the edge is filled with elongated cells, the higher the ratio of thresholded pixels to mask pixels.

### **Colony size**

The size of the colonies are measured by area (number of pixels), perimeter (number of pixels at the border), and circularity ( $4*\pi*area/perimeter^2$ ), since smaller colonies tend to be more circular.