CDK1 Inhibition Targets the p53-NOXA-MCL1 Axis, Selectively Kills Embryonic Stem Cells, and Prevents Teratoma Formation

Noelle E. Huskey, Tingxia Guo, Kimberley J. Evason, Olga Momcilovic, David Pardo, Katelyn J. Creasman, Robert L. Judson, Robert Blelloch, Scott A. Oakes, Matthias Hebrok, and Andrei Goga

Supplementary Information for: Structural Phenotyping of Stem Cell Derived Cardiomyocytes

Authors: Francesco Silvio Pasqualini, Sean Paul Sheehy, Ashutosh Agarwal, Yvonne Aratyn-Schaus and Kevin Kit Parker.

Supplemental Figures

Figure S1 related to Table 1: Schematic representation of the metrics adopted in this paper.
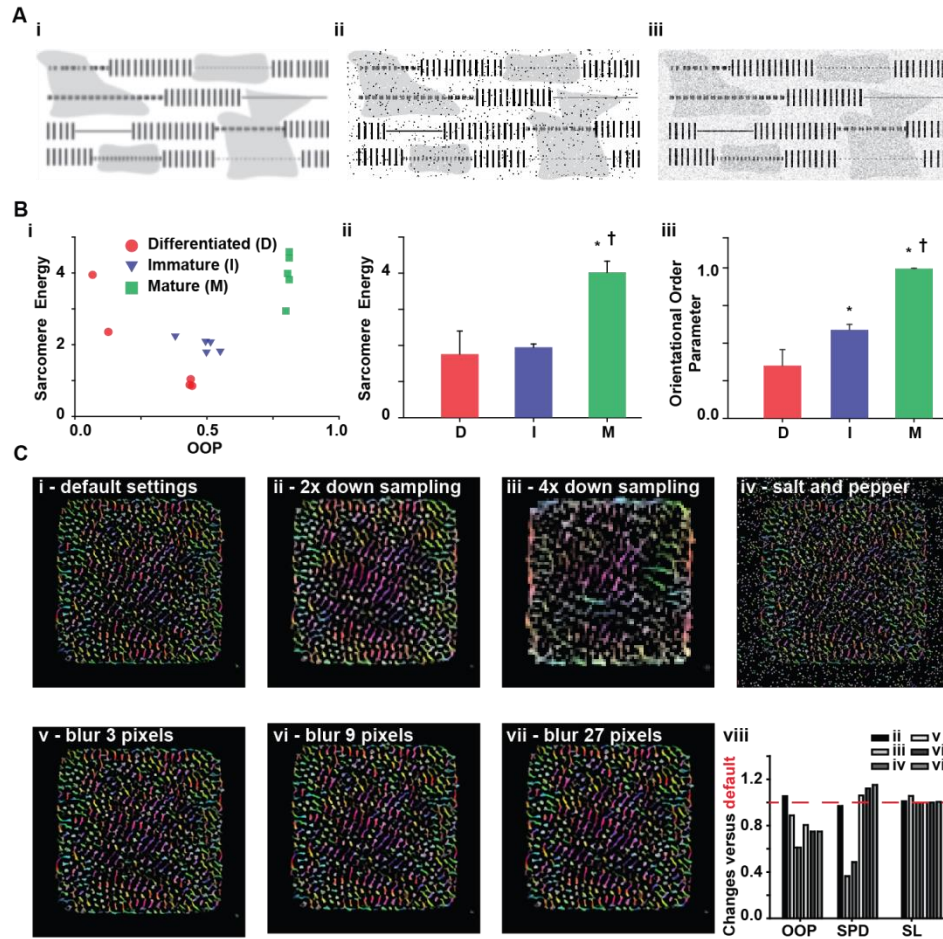
Figure S2 related to Figure 1: Effect of common imaging artifacts on feature extraction. Digital images, specifically those acquired through automatic means (Bakal et al., 2007; Collinet et al., 2010; Jones et al., 2009; Whitehurst et al., 2007), can be affected by several different types of noise and artifacts. Here we demonstrate, on the synthetic images from Fig 1, the effect of out-of-focus blurriness (**A-i**), salt-and-pepper noise (**A-ii**) or poor contrast (**A-iii**). Our metrics were robust to the effect of such noise sources, as demonstrated by the orientational order parameter (OOP) and sarcomeric energy scores. While the exact numerical values were different (**B-i**), the sarcomeric energy could statistically distinguish mature (M) myocytes from the others (**B-ii**). Furthermore, the OOP values were significantly different between mature (M), differentiated (D) and immature (I) cardiomyocytes (**B-iii**). Results are presented as mean ± SEM and analyzed with the ANOVA test (*p*-value<0.05, n=5 independent experiments). Moreover, we analyzed the pCM image in Fig 1 using default settings (**C-i**), as well as after the application of moderate (**C-ii**) and severe (**C-iii**) down-sampling; salt-and-pepper noise (**C-iv**); and moderate (**C-v**), acute (**C-vi**) and severe (**C-vii**) blurring. We then compared the performances of OOP, sarcomeric packing density (SPD) and sarcomere length (SL) (**C-viii**) and showed that only severe down-sampling and salt-and-pepper noise induced a greater than 20% change in the numerical score assigned to the image.
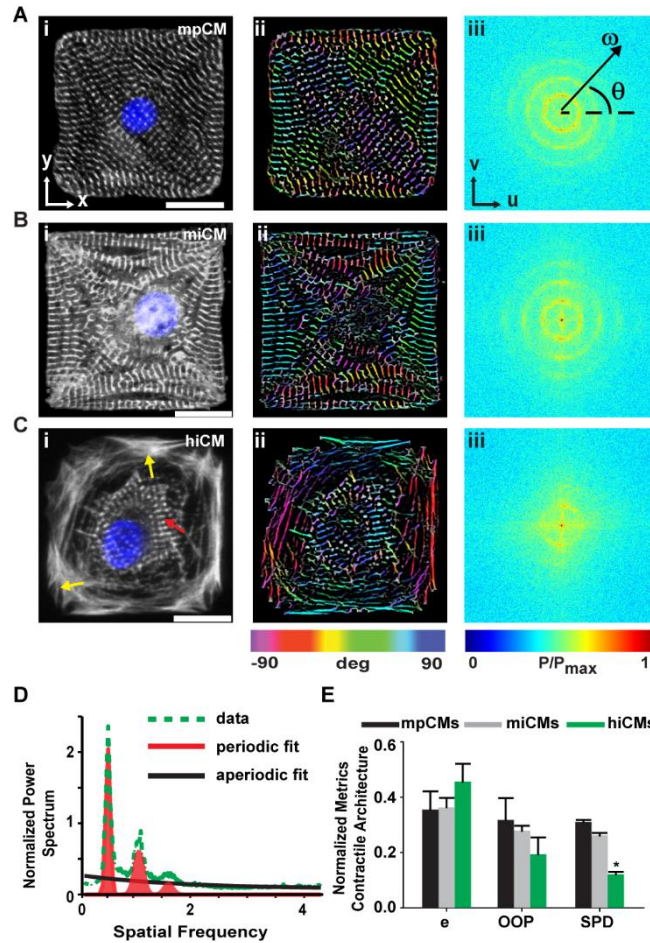
Figure S3 related to Figure 1: Metrics of myofibrillar organization were used to characterize the maturation of primary and stem cell derived cardiomyocytes. Digital images showing chromatin (blue) and sarcomeric α-actinin (white) in murine primary cardiomyocytes (mpCM, **A-i**) and murine (miCM, **B-i**) or human (hiCM, **C-i**) induced pluripotent stem cell derived cardiomyocytes were processed to detect and color-code the principal orientations of aSA positive structures (panels **A-ii, B-ii** and **C-ii** respectively for mpCMs, miCMs and hiCMs). Note that the HSV digital image representation was employed here, where Hue and Saturation channels encode orientation and coherency (Rezakhaniha et al., 2011) respectively, while the Value channel encodes the preprocessed (Sato et al., 1998) sarcomeric α-actinin image. (x,y) indicates a Cartesian system of coordinates for the spatial domain. Scalebar: 20 μm. The normalized Fourier spectra of the Value channel were reported for mpCMs (**A-iii**), miCM (**B-iii**) and hiCM (**C-iii**). (u,v) and (ω,ϑ) respectively indicate a Cartesian and polar system of coordinates for the Fourier domain. Radial integration of **A-iii** lead to the 1D representation in **D** (green dashed line) that we further fitted to identify the periodic (red) and aperiodic (black) components needed to derive the sarcomere packing density (SPD). (**E**) Note that SPD could discriminate the maturation of the cell contractile cytoskeleton in this dataset where the nuclear eccentricity (Bray et al., 2010) (e) and the traditional orientational order parameter (Sheehy et al., 2012) (OOP) fell short due to the central symmetry in the cell geometry (Grosberg et al., 2011; Sheehy et al., 2012). Results are mean ± SEM, and were analyzed with ANOVA ($p<0.05$, n=3 independent experiments)
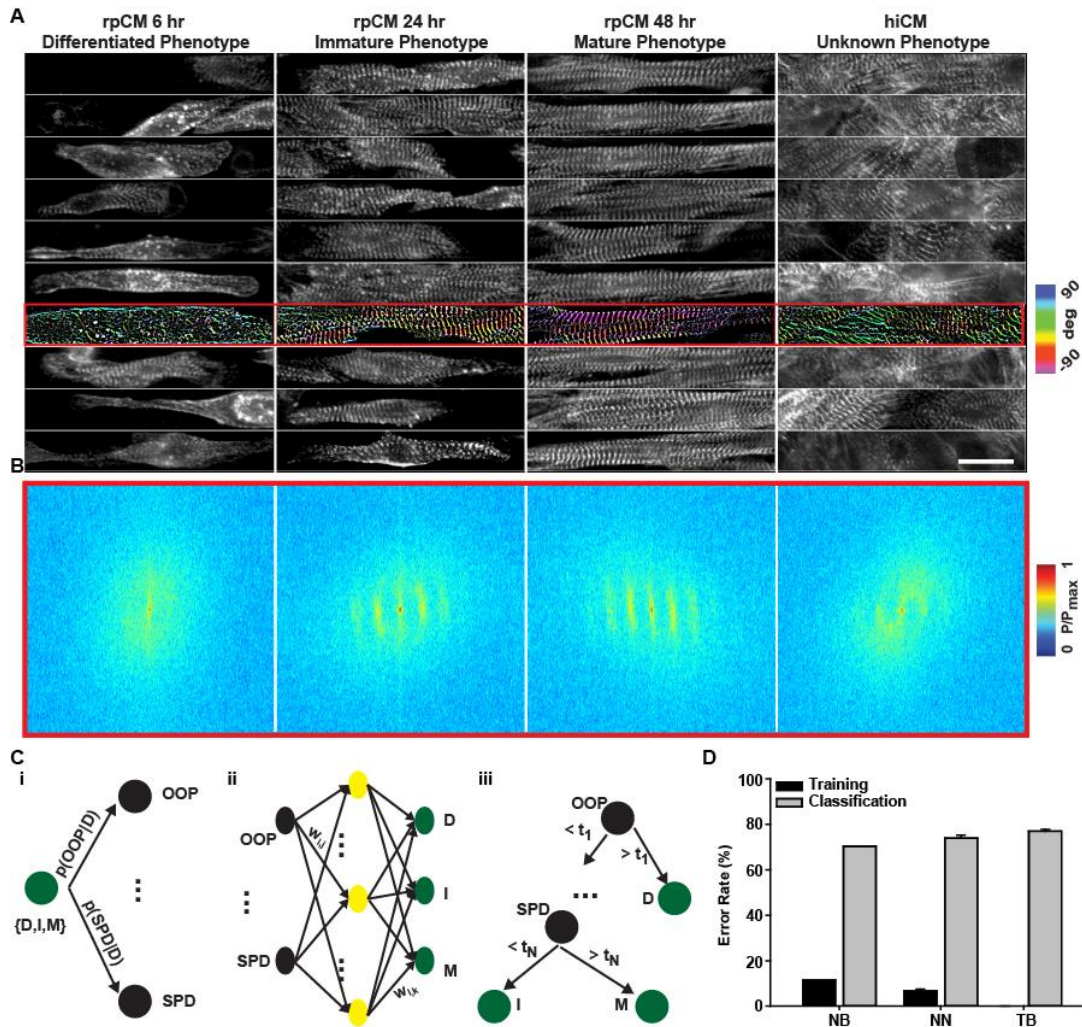
Figure S4 related to Figure 2 and Table 2: Myofibrillogenesis dataset. **A**) Representative images from the dataset used in this study: columns from left to right represent sarcomeric α-actinin images collected at 6 hr, 24 hr and 48 hrs after seeding of rat primary cardiomyocytes (rpCMs) as well as commercially available human induced pluripotent stem cell derived myocytes (hiCMs). Scale bar is 20 um. The set of image in the red inset were color-coded to highlight the orientation of each α-sarcomeric positive pixel in the image, demonstrating how more mature myocytes have aligned sarcomeres sharing the same color. **B**) 2D Fourier power spectra for the same set of image in the red inset in panel **A**. Note how increasingly more mature myocytes display more intense periodic signal in the Fourier domain. **C**) Schematic representations of the classifiers used to analyze this dataset: naïve Bayes (NB, **i**), neural network (NN, **ii**) and tree bagging (TB, **iii**). **D**) Error rates for training (black) and classification (gray) as measured in ten randomly seeded and optimized iterations for each classifier. Results are presented as mean ± SEM.

Supplemental Experimental Procedures

Machine Learning

The machine learning problem we wanted to solve could be divided in 2 parts. In the training phase we were interested in solving the following direct problem: given a set of classes $C = \{D, I, M\}$ (respectively differentiated, immature and mature myocytes) and a set of features F= $\{F_1, F_2, ..., F_M\}$ (related to the myofibrillar architecture of cells pertaining to those classes, such that e.g. $F_1 = SPD, F_2 = OOP, ...$) we wanted to find an algorithm that mapped a given combination of features $f = \{f_1, f_2, ..., f_M\}$ to a single class $C = c$. In the classification phase, we wanted to solve the inverse problem: given a new set of features, that was not used during training, we asked the algorithm to assign it to one of the available classes.

We collected a dataset comprising sarcomeric α-actinin digital images from rpCMs obtained at 6, 24 and 48 hrs after seeding: based on a-priori knowledge (Dabiri et al., 1997; Parker et al., 2008; Sheehy et al., 2014) we took these time points to represent the classes of differentiated, immature and mature myocytes, respectively. We then implemented and trained several classifiers (see Table 2) on this dataset and utilized them to classify features measured on independently acquired sarcomeric α-actinin images from hiCMs samples. In particular we counted the number of times the myofibrillar architecture of hiCMs was not classified as mature.

Since different classifiers are based on different assumptions and rely on different stochastic algorithms (Sun et al., 2012), we tested three different frameworks and ten different randomly seeded iterations, to ensure the classification was robust to the choice of a specific machine learning strategy and particular initialization.

Classifier 1: Naïve Bayes

In the framework of Bayesian classification, a classifier is based on a conditional model for the probability that a certain set of features belongs to a given class (Fig S4C-i). Under the naïve hypothesis of conditional independence between features and once a suitable prior is selected, all model parameters can be derived applying the maximum likelihood estimation algorithm on the training dataset. We chose kernel distributions as priors, since they only require the random variables to be continued and not normally distributed, and randomly seeded ten different iterations of the Naïve Bayes classifier. A 10-fold cross-validation test was adopted to determine the performance of the classifier.

Classifier 2: Neural Network

In the framework of neural network a classifier is a network that possesses: i) an input layer, with as many neurons as there are features; ii) at least one hidden layer, with a number of neurons that can be optimized; and iii) an output layer, with as many neurons as there are classes. The neurons n_i and n_j are connected through a weight $w_{i,j}$ and training the network is equivalent to assign the weights $w_{i,j}$ such that when the input layer receives the set of features $\{F_1, F_2, ..., F_M\}$ pertaining to the class C, the output node associated with C exhibits the highest value (Fig S4C-ii). We utilized Matlab Neural Network toolbox to design a perceptron network and we optimized the number of hidden neurons (from 1 to 20) for 10 random iterations. We also trained the network using the back-propagation algorithm adopting 70% of the dataset for Training, 15% for Validation and 15% for final Testing.

Classifier 3: Tree Bagging

Tree bagging stands for bootstrap aggregation of decision trees. Bootstrap aggregation is an ensemble meta-algorithm that optimally subdivides the entire dataset and uses each part to train a simpler classifier (in this case a binary decision tree). The final classification is obtained by voting: that is, if the majority of trees has assigned the set of features $\{F_1, F_2, ..., F_M\}$ to the class $C = c$ than $C = c$ will be the result of the

global classification. The number of decision trees to be used is thus the parameter optimized in our implementation.

Binary decision trees are simple yet powerful machine learning algorithms. For each feature $\{F_1, , \dots, F_M\}$ the algorithm chooses thresholds $\{t_1, , \dots, t_M\}$ and an ordering scheme, such that the set of features can be traveled from "the root to the leaves" coherently. The value of a given feature F_k is considered, compared with the relative threshold and a decision is made: either to examine the next feature in the tree F_{k+1}, or to assign a specific class to that combination (Fig S4C-iii)..

We performed 10 random initialization of the algorithm, selected the optimal number of trees (in the range 5-100) and then trained each decision tree using the Gini's diversity index. A 10-fold cross-validation test was adopted to determine the performance of the classifier.

References:

Bakal, C., Aach, J., Church, G., and Perrimon, N. (2007). Quantitative morphological signatures define local signaling networks regulating cell morphology. Science (New York, NY) *316*, 1753-1756.

Bray, M.-A.P., Adams, W.J., Geisse, N.A., Feinberg, A.W., Sheehy, S.P., and Parker, K.K. (2010). Nuclear morphology and deformation in engineered cardiac myocytes and tissues. Biomaterials *31*, 5143-5150.

Collinet, C., Stöter, M., Bradshaw, C.R., Samusik, N., Rink, J.C., Kenski, D., Habermann, B., Buchholz, F., Henschel, R., Mueller, M.S.*, et al.* (2010). Systems survey of endocytosis by multiparametric image analysis. Nature *464*, 243-249.

Dabiri, G.A., Turnacioglu, K.K., Sanger, J.M., and Sanger, J.W. (1997). Myofibrillogenesis visualized in living embryonic cardiomyocytes. Proceedings of the National Academy of Sciences of the United States of America *94*, 9493-9498.

Grosberg, A., Kuo, P.-L., Guo, C.-L., Geisse, N.A., Bray, M.-A., Adams, W.J., Sheehy, S.P., and Parker, K.K. (2011). Self-organization of muscle cell structure and function. PLoS computational biology *7*, e1001088.

Jones, T.R., Carpenter, A.E., Lamprecht, M.R., Moffat, J., Silver, S.J., Grenier, J.K., Castoreno, A.B., Eggert, U.S., Root, D.E., Golland, P.*, et al.* (2009). Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. Proceedings of the National Academy of Sciences of the United States of America *106*, 1826-1831.

Parker, K.K., Tan, J., Chen, C.S., and Tung, L. (2008). Myofibrillar architecture in engineered cardiac myocytes. Circulation research *103*, 340-342.

Rezakhaniha, R., Agianniotis, A., Schrauwen, J.T.C., Griffa, A., Sage, D., Bouten, C.V.C., van de Vosse, F.N., Unser, M., and Stergiopulos, N. (2011). Experimental investigation of collagen waviness and orientation in the arterial adventitia using confocal laser scanning microscopy. Biomechanics and modeling in mechanobiology *11*, 461-473.

Sato, Y., Nakajima, S., Shiraga, N., Atsumi, H., Yoshida, S., Koller, T., Gerig, G., and Kikinis, R. (1998). Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. Medical image analysis *2*, 143-168.

Sheehy, S.P., Grosberg, A., and Parker, K.K. (2012). The contribution of cellular mechanotransduction to cardiomyocyte form and function. Biomechanics and Modeling in Mechanobiology *11*, 1227-1239.

Sheehy, Sean P., Pasqualini, F., Grosberg, A., Park, Sung J., Aratyn-Schaus, Y., and Parker, Kevin K. (2014). Quality Metrics for Stem Cell-Derived Cardiac Myocytes. Stem Cell Reports *2*, 282-294.

Sun, N., Yazawa, M., Liu, J., Han, L., Sanchez-Freire, V., Abilez, O.J., Navarrete, E.G., Hu, S., Wang, L., Lee, A.*, et al.* (2012). Patient-specific induced pluripotent stem cells as a model for familial dilated cardiomyopathy. Sci Transl Med *4*, 130ra147.

Whitehurst, A.W., Bodemann, B.O., Cardenas, J., Ferguson, D., Girard, L., Peyton, M., Minna, J.D., Michnoff, C., Hao, W., Roth, M.G.*, et al.* (2007). Synthetic lethal screen identification of chemosensitizer loci in cancer cells. Nature *446*, 815-819.