High-throughput gene screen reveals modulators of nuclear shape

Andrew C. Tamashunas\textsuperscript{a}, Vincent J. Tocco\textsuperscript{a}, James Matthews\textsuperscript{b}, Qiao Zhang\textsuperscript{a}, Kalina R. Atanasova\textsuperscript{b}, Lauren Paschall\textsuperscript{c}, Shreya Pathak\textsuperscript{c}, Ranjala Ratnayake\textsuperscript{b}, Andrew D. Stephens\textsuperscript{c}, Hendrik Luesch\textsuperscript{b}, Jonathan D. Licht\textsuperscript{d}, and Tanmay P. Lele\textsuperscript{a,\ast}

\textsuperscript{a}Department of Chemical Engineering and \textsuperscript{b}Department of Medicinal Chemistry and Center for Natural Products, Drug Discovery and Development (CNPD3), University of Florida, Gainesville, FL 32610; \textsuperscript{c}Biology Department, University of Massachusetts Amherst, Amherst, MA 01003; \textsuperscript{d}Division of Hematology/Oncology, University of Florida Health Cancer Center, Gainesville, FL 32610

**Abstract** Irregular nuclear shapes characterized by blebs, lobules, micronuclei, or invaginations are hallmarks of many cancers and human pathologies. Despite the correlation between abnormal nuclear shape and human pathologies, the mechanism by which the cancer nucleus becomes misshapen is not fully understood. Motivated by recent evidence that modifying chromatin condensation can change nuclear morphology, we conducted a high-throughput RNAi screen to identify epigenetic regulators that are required to maintain normal nuclear shape in human breast epithelial MCF-10A cells. We silenced 608 genes in parallel using an epigenetics siRNA library and used an unbiased Fourier analysis approach to quantify nuclear contour irregularity from fluorescent images captured on a high-content microscope. Using this quantitative approach, which we validated with confocal microscopy, we significantly expand the list of epigenetic regulators that impact nuclear morphology.

**Introduction**

The nucleus of the mammalian cell organizes and protects the genome, and ensures normal gene expression and cell function. While the nuclei of most normal epithelial cell types have smooth circular or ellipsoidal shape, the nucleus of a cancer cell has irregularities in its contour including blebs, lobules, micronuclei, or invaginations. These nuclear irregularities are characteristic of many types of malignancies, including breast cancer (Abdalla et al., 2009; Williams et al., 2014), melanoma (Mijovic et al., 2013), squamous cell carcinoma (Speight, 2007), lung cancer (Petersen et al., 2009), and prostate cancer (Diamond et al., 1982; Epstein et al., 1984). Pathologists commonly evaluate nuclear shape irregularities to diagnose and prognosticate the severity of cancer in biopsies of human tissue (Gil et al., 2002; Kashyap et al., 2018). Additionally, irregular nuclear shapes are characteristic of several major human diseases, including laminopathies and progerias (Goldman et al., 2004; De Vos et al., 2010; Ho et al., 2011; Zwerger et al., 2011; Siu et al., 2012; Taniq et al., 2017; Steele-Stallard et al., 2018). Despite the well-established correlation between abnormal nuclear shape and human pathologies, the mechanisms by which nuclei of cells become misshapen are not fully understood.

Numerous cancers are characterized by a decrease in the expression of nuclear lamins (Capo-chichi et al., 2011a,b; Denais and Lammerding, 2014; Irianto et al., 2016). Nuclear lamins impart rigidity to the nucleus at nuclear extensions greater than 3 μm (Stephens et al., 2017) and down-regulation of lamins results in an abnormally shaped nucleus (Lammerding et al., 2004, 2006). Cytoskeletal forces that act on the nucleus can cause or amplify existing nuclear shape abnormalities (Hatch et al., 2013; Cain et al., 2014; Takaki et al., 2017; Tocco et al., 2018). There is also evidence that nuclear shape abnormalities can be caused by alterations to chromatin regulatory proteins. Knockdown of the chromatin remodeling enzyme BRG-1 (encoded by SMARCA4) increased the percentage of irregular nuclei in MCF-10A (human breast fibrocystic disease) cells, characterized by curves, bulges, folds, and invaginations in the nuclear contour (Imbalzano et al., 2013). Treatment with drugs that increase euchromatin or decrease heterochromatin resulted in nuclear blebbing in HT1080 cells and mouse embryonic fibroblasts, without...
altered lamin (Stephens et al., 2018a). Altering chromatin condensation through changes in histone modification states also mechanically softened nuclei, which may make them susceptible to developing abnormal morphologies (Stephens et al., 2018a). Conversely, condensing chromatin by increased levels of heterochromatin reduced the incidence of blebbled nuclei and partially rescued nuclear shape in Hutchinson-Gilford progeria syndrome (HGPS) cells, a disease caused by mutant lamin A (Stephens et al., 2018a, 2019b).

Motivated by these previous results, we conducted a high-throughput RNAi screen to identify individual epigenetic regulators that are required to maintain normal nuclear shape in noncancerous human breast epithelial MCF-10A cells. For our screen, we developed an automated elliptical Fourier analysis technique to quantify the elliptic Fourier coefficient (EFC) ratio which is a quantitative metric of nuclear shape abnormalities. Knockdown of a library of 608 epigenetic-related genes revealed novel gene products which caused statistically significant nuclear shape aberrations. Gene depletions with a mean EFC ratio below approximately 5.3 were visually abnormal. We used confocal microscopy to study nuclear abnormalities in a small set of 21 genes and found reductions in EFC ratio similar to results with the high-throughput screen. Our results offer new directions in understanding the mechanisms of aberrant nuclear shaping in malignancy.

RESULTS
A Fourier analysis approach for quantification of nuclear contour irregularities
Commonly used methods to detect irregular nuclear shapes, such as manual classification, are subject to observer bias and poor reproducibility (Imbalzano et al., 2013). Here we used elliptical Fourier analysis to quantitatively measure nuclear morphology, which has previously been used to analyze irregular nuclear shapes (Diaz et al., 1989; Lammerding et al., 2006). This method approximates the nuclear contour with a series of diminishing harmonic ellipses, and an EFC ratio is calculated. The EFC ratio provides a relative estimate of the regularity of each nucleus, with larger EFC ratios corresponding to more circular or ellipsoidal (regular) nuclei, and smaller EFC ratios to nuclei with more complex, irregular shapes which require larger contributions of subsequent ellipses to accurately fit the nuclear contour. In our analysis of MCF-10A nuclei, a series of 15 harmonic ellipses were found to be sufficient for approximating nuclear contours (compare contour and reconstructed contour; Figure 1, A–E). The EFC ratio correlated well with visually obvious irregularities in the nuclear contour (Figure 1F), and there was a significant separation in EFC ratios between normal and visually irregular shapes (ranging from ~10 for normal shapes to ~2 for irregular shapes). In particular, nuclei with EFC ratios below approximately 5.3 appeared visually abnormal. Another objective shape descriptor commonly used to compare nuclear morphology in the literature is solidity, which is defined as the ratio of the measured area of the nucleus to that of a bounding convex shape. Therefore, nuclei with more irregular nuclear shapes would, in theory, have a significantly larger bounding convex area than regular nuclei with circular or ellipsoidal shapes. However, in contrast with the EFC ratio, the widely utilized metric of solidity (Choi et al., 2011; Scholgin et al., 2014; Hulsman et al., 2015; Jagielska et al., 2017; Verschuuren et al., 2017; Konstandinou et al., 2018; Stephens et al., 2018a, b) yielded relatively weak numerical separation for the same set of regular and visually irregular nuclei (Figure 1F).

We quantitatively assessed the accuracy of the Fourier approximation of nuclear shapes by calculating the root mean-square (RMS) error; that is, the RMS difference between the Fourier reconstruction and the nuclear boundary. The RMS error was 500 nm or less (Figure 1G), which is comparable to the spatial resolution of the 40x objective (0.95 numerical aperture) used for imaging. The small RMS error in fitting the 2-D nuclear contour highlights the accuracy of the Fourier method and validates the visual agreement between estimated and measured contours in Figure 1, A–F.

If the error in estimation is dependent on the degree of nuclear irregularity, that would bias the Fourier approach and reduce its reliability. We therefore examined the correlation between the RMS error and the EFC ratio and found no correlation between these quantities (Figure 1H). Together with Figure 1G, these results validate the Fourier approximation method as an effective tool for detecting nuclear irregularities. We developed custom programs for automated quantification of the EFC ratio from fluorescent images of nuclei. The algorithm was able to detect the majority of isolated nuclei from fluorescent images (Figure 1I) and approximate nuclear contours through Fourier decomposition while excluding clumped nuclei.

Next, we performed imaging of MDA-MB-231 cancer nuclei, which are known to have abnormal shapes, using a high-throughput microscope and calculated EFC ratios. The mean EFC ratio of MDA-MB-231 cancer nuclei was 3.05 ± 0.03 and significantly lower than the mean EFC ratio of MCF-10A cells (EFC ratio of 6.5 ± 0.01), validating the use of the EFC ratio for quantifying nuclear abnormalities in cancer cells (Figure 1, J and K). Solidity, plotted using an adjusted solidity metric (defined as (1-solidity) multiplied by 100) which is inversely related to solidity in terms of nuclear regularity, was also statistically lower in MDA-MB-231 cells (mean adjusted solidity of 4.03 ± 0.03) than MCF-10A control nuclei (mean adjusted solidity of 3.74 ± 0.01), but the numerical separation between control and cancer nuclei was again much weaker (Figure 1K).

A high-throughput screen for nuclear shape abnormalities
We next developed an experimental and computational analysis pipeline for a high-throughput gene screen of the effects of depletion of chromatin proteins on nuclear shape (Figure 2A). MCF-10A cells were seeded into 384-well plates at an initial density of 500 cells/well. After seeding, cells were incubated with individual siRNAs at a concentration of 25 nM, which was found to maximize depletion of mRNA levels while minimizing loss of cell viability (Supplemental Figure S4), in 384-well plates for 72 h, which was determined to be sufficient time for knockdown via RT-qPCR, consistent with previous RNAi screens (Bolt et al., 2015) (Figure 2, B and C). Cells were then fixed and immunolabeled for target proteins (including lamin A), and fluorescent images of nuclei stained with Hoechst 33342 were acquired through high-throughput imaging at 40x resolution. Nuclear morphology was then quantified for in-focus nuclei using elliptical Fourier analysis. Gene depletions that resulted in abnormal nuclear shapes (decreased EFC ratios) were identified using nonparametric statistical analysis and a subsequent, refined criterion for mean EFC ratios below a cut-off value (Figure 2D).

Using the analysis pipeline in Figure 2A, we screened 608 siRNAs targeting genes associated with epigenetic and chromatin regulatory roles. The full list of genes whose depletion caused statistically significant effects on EFC ratio is listed in Supplemental Table S2. The computational analysis was highly sensitive to nuclear shape abnormalities and therefore was able to detect subtle changes in nuclear shape. However, such subtle changes may not be physiologically impactful. Because nuclei below a mean value of approximately 5.3 looked visually abnormal (see Figure 1, A–F), we reasoned that imposing such a cut-off might be one way to narrow down the list of
genes that significantly impact nuclear morphology. Genes that had visually detectable effects are listed in Supplemental Table S3. Using this criterion, our list of depletion conditions, which were found to statistically significantly induce irregular nuclear morphologies, was narrowed down to 33 genes, encoding a diverse array of proteins (Supplemental Table S3 and Supplemental Figure S5).

The degree of cellular spreading drives nuclear flattening and the assembly of actomyosin networks further generate mechanical force on the nuclear surface which can drive dysmorphia (Takaki et al., 2017). In addition, local motion of the cellular membrane can induce cancer nuclear deformations in 2-D cultures (Lele et al., 2018; Tocco et al., 2018; Kent et al., 2019). Consequently, potential confounding effects of gene depletions on cellular spreading may indirectly impact nuclear shape (Lele et al., 2018). We therefore quantified cell spreading area in a random subset of gene depletions. We found no significant correlation between cell spreading area and EFC ratio (Supplemental Figure S1A). Additionally, EFC ratio did not correlate significantly with nuclear aspect ratio (Figure 2E) or with nuclear cross-sectional area (Figure 2F). Nuclear contour irregularities, quantified by the EFC ratio, therefore appear largely independent of effects of siRNA transfection on cell spreading (Neelam et al., 2016).

FIGURE 1: Quantification of nuclear contour irregularities with elliptical Fourier analysis. (A) Example images of single nuclei from nontransformed human MCF-10A mammary epithelial cells. Images of nuclei are arranged in descending order of regularity, from left to right. Scale bar is 5 µm. (B) Representative images of the detected nuclear contour in A overlaid onto the original raw image; nuclear contours were detected using MATLAB image processing algorithms. (C) Nuclear contours shown in B without the overlay of the original image (blue). (D) Elliptic Fourier approximations of the nuclear contours from C, calculated using 15 elliptic harmonics to fit each nuclear shape (Diaz et al., 1989; Lammerding et al., 2006). (E) Overlays of the detected nuclear contour and the elliptic Fourier approximation. (F) Representative EFC ratios (computed in MATLAB) and solidity measurements (where solidity = area/convex area; computed in ImageJ) for each nuclear shape detected in A. (G) Quantification of the RMS error between the elliptic Fourier approximation and the nuclear boundary; each data point corresponds to a single nucleus (N = 3, n = 20 where N represents the number of biological replicates and n represents the total number of cells analyzed). (H) EFC ratio plotted against RMS error for each nucleus in (G; N = 3, n = 20). Error bars denote SEM. (I) Representative image after segmentation and quantification of EFC ratio. White outlines represent the detected nuclear boundary; green numbers overlaid onto nuclear centroids represent the EFC ratio for each respective nucleus. Scale bar is 20 µm. (J) Representative images of MCF-10A human breast epithelial cells and MDA-MB-231 human breast adenocarcinoma cells taken at 40× on a high-content microscope. Scale bar is 25 µm. (K) Mean EFC ratio and adjusted solidity (defined as (1-solidity)*100) for MCF-10A and MDA-MB-231 cells (N ≥ 15 and 3, n = 71,191 and 6952 for MCF-10A and MDA-MB-231, respectively). Error bars represent SEM. *p < 0.05 by the Bonferroni-corrected nonparametric Dunn’s test.
Validation with confocal fluorescence microscopy

We next performed a limited confocal imaging analysis at 60x magnification (1.49 NA) on a subset of cells depleted for a random subset of 20 genes (Supplemental Table S4 and Supplemental Figure S6), 15 of which caused statistically significant effects on the EFC ratio in the high-throughput screen and 5 genes that had no effect on the EFC ratio (Table 1). As these experiments were done in 96-well plates (as opposed to the 384-well format in the high-throughput approach), we reconfirmed knockdown of gene expression using RT-qPCR in a subset of siRNA transfections (Supplemental Figure S1, B and C, for RT-qPCR results of genes which did or did not produce a statistically significant effect on EFC ratio, respectively). A series of confocal z-stack images were taken for each siRNA condition and used to create maximum intensity projections of nuclei, which were then used to calculate EFC ratios. Of siRNA conditions that were classified in the high-throughput screen to decrease EFC ratios (Supplemental Table S5), 60% were confirmed by confocal analysis to have the same effect (compare Table 1 with Table 2; Supplemental Table S6). On the other hand, 100% of siRNA conditions identified to have no effect on nuclear shape in the high-throughput screen (Supplemental Table S7) were confirmed to have no effect by confocal analysis (compare Table 1 with Table 2; Supplemental Table S8). These results validate the efficacy of the high-throughput screen in detecting effects on nuclear shape. Also, because siRNA knockdown efficiencies can be variable, we used three sets of siRNA smart pools that targeted unique sequences different from our original experiments (Supplemental Table S9) toward three top hits, TP53, DIDO1, and MTA2, and
reperformed the knockdowns. We confirmed that the siRNA treatment caused a depletion of the mRNA levels for these genes (Supplemental Figure S2A) and also found as before that this depletion caused a statistically significant decrease in the EFC ratio (Supplemental Figure S2B).

As nuclear volume is an important morphological parameter that can change during cancer progression, we quantified nuclear volume from confocal fluorescence images using our previously published methods (Tocco et al., 2018). We found no correlation between the EFC ratio and the nuclear volume (Figure 3A). Furthermore, in agreement with the results of the high-throughput screen, we found no correlation between the EFC ratio and the nuclear aspect ratio nor between the EFC ratio and the 2-D nuclear cross-sectional area (Figure 3, B and C). These results confirm the conclusions from the high-throughput screen and highlight the fact that the EFC ratio is a specific readout of nuclear contour irregularities.

To highlight the efficacy of the nuclear EFC ratio as a readout of nuclear contour, representative maximum projection images of nuclei for the scrambled control and the top seven gene depletion conditions from the confocal screen are shown in Figure 3D. Notably, depletion of each of these six genes with the lowest mean EFC ratios in the confocal screen produced visually pronounced nuclear aberrations compared with the scrambled siRNA negative control (compare MTA2, TP53, and Scrambled in Figure 3D and Table 2). Overall, these results significantly expand the list of epigenetic regulators that impact nuclear shape.

**DISCUSSION**

Nuclear pleomorphisms are characteristic of many types of cancers and have been used both diagnostically and prognostically in clinical settings (Papanicolaou and Traut, 1997; de Las Heras and Schirmer, 2014). Recent studies suggest that alterations to chromatin and epigenetic regulators may cause abnormal nuclear shapes (Imbalzano et al., 2013; Furusawa et al., 2015; Schreiner et al., 2015; Stephens et al., 2017, 2018a; Senigagliesi et al., 2019) associated with nuclear dysfunction (Stephens et al., 2019a). In this study, we used noncancerous MCF-10A breast epithelial cells in conjunction with high-throughput nuclear imaging and knockdown using an RNAi library to systematically screen epigenetic regulators of nuclear shape. The screen successfully revealed a number of genes encoding epigenetic regulatory proteins which caused nuclear shape abnormalities on depletion. Our findings that chromatin regulatory genes are implicated in abnormal nuclear shaping are in agreement with previous reports that knockdown of the SWI/SNF chromatin remodeling complex ATPase BRG1 (encoded by SMARCA4) increases the frequency of irregular nuclei (Imbalzano et al., 2013). A commonly used metric for quantifying nuclear shape abnormalities is solidity. We found that solidity does detect differences between nuclear shapes as expected. However, in comparison with solidity, the EFC ratio has a larger range of values that should allow more facile discrimination between normal and abnormal nuclear shapes. Another metric commonly used for nuclear morphology is roundness or circularity.

### Table 1: Summary of results from the high-throughput screen for a subset of 21 randomly selected genes to be screen using confocal microscopy.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Number of nuclei</th>
<th>Mean EFC ratio</th>
<th>Standard error</th>
<th>Statistically significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>CXXC1</td>
<td>805</td>
<td>4.85</td>
<td>0.07</td>
<td>Yes</td>
</tr>
<tr>
<td>HIST1H3G</td>
<td>1823</td>
<td>4.97</td>
<td>0.04</td>
<td>Yes</td>
</tr>
<tr>
<td>HIST1H1T</td>
<td>533</td>
<td>5.00</td>
<td>0.07</td>
<td>Yes</td>
</tr>
<tr>
<td>SUZ12</td>
<td>807</td>
<td>5.03</td>
<td>0.06</td>
<td>Yes</td>
</tr>
<tr>
<td>WSB2</td>
<td>1524</td>
<td>5.08</td>
<td>0.05</td>
<td>Yes</td>
</tr>
<tr>
<td>MTA2</td>
<td>909</td>
<td>5.11</td>
<td>0.07</td>
<td>Yes</td>
</tr>
<tr>
<td>ZRANB3</td>
<td>692</td>
<td>5.25</td>
<td>0.08</td>
<td>Yes</td>
</tr>
<tr>
<td>INO80B</td>
<td>1795</td>
<td>5.29</td>
<td>0.05</td>
<td>Yes</td>
</tr>
<tr>
<td>FOXP1</td>
<td>505</td>
<td>5.34</td>
<td>0.10</td>
<td>Yes</td>
</tr>
<tr>
<td>DIDO1</td>
<td>687</td>
<td>5.35</td>
<td>0.09</td>
<td>Yes</td>
</tr>
<tr>
<td>ARID4A</td>
<td>1015</td>
<td>5.37</td>
<td>0.06</td>
<td>Yes</td>
</tr>
<tr>
<td>MAP3K7</td>
<td>2048</td>
<td>5.42</td>
<td>0.05</td>
<td>Yes</td>
</tr>
<tr>
<td>TAF7</td>
<td>850</td>
<td>5.43</td>
<td>0.08</td>
<td>Yes</td>
</tr>
<tr>
<td>RAD54L2</td>
<td>1946</td>
<td>5.48</td>
<td>0.04</td>
<td>Yes</td>
</tr>
<tr>
<td>TP53</td>
<td>139</td>
<td>6.17</td>
<td>0.23</td>
<td>Yes</td>
</tr>
<tr>
<td>RAD51</td>
<td>305</td>
<td>6.20</td>
<td>0.14</td>
<td>No</td>
</tr>
<tr>
<td>LEO1</td>
<td>1283</td>
<td>6.30</td>
<td>0.07</td>
<td>No</td>
</tr>
<tr>
<td>GADD45A</td>
<td>902</td>
<td>6.83</td>
<td>0.08</td>
<td>No</td>
</tr>
<tr>
<td>SCRAM</td>
<td>15677</td>
<td>6.90</td>
<td>0.04</td>
<td>N/A</td>
</tr>
<tr>
<td>FBL</td>
<td>1036</td>
<td>6.92</td>
<td>0.08</td>
<td>No</td>
</tr>
<tr>
<td>SMARCB1</td>
<td>819</td>
<td>7.04</td>
<td>0.09</td>
<td>No</td>
</tr>
</tbody>
</table>

Statistical significance of gene depletion EFC ratios, relative to the scrambled siRNA control, was determined by the two-tailed Bonferroni-corrected nonparametric Dunn’s test for \( p < 0.05 \).
TABLE 2: Summary of results from the confocal microscopy screen for a subset of 21 randomly selected genes.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Number of nuclei</th>
<th>Mean EFC ratio</th>
<th>Standard error</th>
<th>Statistically significant?</th>
<th>Hit in the high-throughput screen?</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTA2</td>
<td>176</td>
<td>4.62</td>
<td>0.15</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>TAF7</td>
<td>255</td>
<td>5.07</td>
<td>0.10</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FOXP1</td>
<td>153</td>
<td>5.19</td>
<td>0.16</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>DIDO1</td>
<td>146</td>
<td>5.14</td>
<td>0.15</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SUZ12</td>
<td>164</td>
<td>5.19</td>
<td>0.15</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>TP53</td>
<td>235</td>
<td>5.35</td>
<td>0.13</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ARID4A</td>
<td>176</td>
<td>6.01</td>
<td>0.20</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>RAD54L2</td>
<td>165</td>
<td>5.49</td>
<td>0.13</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CXCL1</td>
<td>199</td>
<td>5.93</td>
<td>0.17</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>MAP3K7</td>
<td>137</td>
<td>5.97</td>
<td>0.15</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>FBL</td>
<td>246</td>
<td>6.26</td>
<td>0.13</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>RAD51</td>
<td>188</td>
<td>6.33</td>
<td>0.16</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>WS2B</td>
<td>492</td>
<td>6.38</td>
<td>0.11</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>HIST1H1T</td>
<td>209</td>
<td>6.31</td>
<td>0.15</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>HIST1H3G</td>
<td>210</td>
<td>6.80</td>
<td>0.18</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>ZRANB3</td>
<td>209</td>
<td>6.63</td>
<td>0.14</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>LEO1</td>
<td>202</td>
<td>6.85</td>
<td>0.14</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SCRAM</td>
<td>467</td>
<td>6.93</td>
<td>0.11</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>GADD45A</td>
<td>148</td>
<td>7.09</td>
<td>0.16</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SMARCB1</td>
<td>158</td>
<td>7.48</td>
<td>0.21</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>INO80B</td>
<td>124</td>
<td>7.79</td>
<td>0.23</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Statistical significance of gene depletion EFC ratios, relative to the scrambled siRNA control, was determined by the two-tailed Bonferroni-corrected nonparametric Dunn’s test for \( p < 0.05 \).

(defined as \( \frac{4 \pi \text{area}}{\text{perimeter}^2} \)). The circularity metric for the genes in our confocal images similarly has a relatively narrow range of values (Supplemental Figure S3) and a weak correlation with the EFC ratio. These results suggest that the EFC ratio is a sensitive measure of nuclear shape abnormalities.

Our gene screen reveals one of the most well-known cancer-related genes, TP53, as a regulator of nuclear shape. TP53 is mutated in 92.98% of uterine carcinomas, 91.51% of ovarian serous cystadenocarcinomas, and 36.51% of all breast cancers with most mutations being loss-of-function (Grossman et al., 2016). A role for p53 in nuclear shape determination is consistent with an earlier study which reported increased nuclear shape variability in high-grade tumors in which p53 was frequently mutated (associated with loss-of-function) compared with low-grade tumors with infrequent p53 mutations (Friedrich et al., 1997). More recently, it was shown that p53 knockdown can cause nuclear rupture (Yang et al., 2017). Interestingly, another study has recently demonstrated that overexpression or activation of p53 induces nuclear deformation reminiscent of the irregular nuclear shapes found in HGPS cells and induces cellular senescence (Yoon et al., 2019). Our findings thus corroborate several studies which have linked dysregulation of p53 with aberrant nuclear shapes in a novel cell type.

Top hits that satisfied the EFC ratio less than \( \sim 5.3 \) criterion in the high-throughput screen included TCF19, KMT5C, YAF2, RBBP5, POGZ, SIN3A, and CXCL1 (Supplemental Table S3). Our choice of a mean EFC ratio of below approximately 5.3 was based on visually noticeable contour irregularities, quantified by four independent observers, binned into binary “regular” and “irregular” groups based on the presence of concavities in the nuclear contour. If we were to not impose this cutoff, then the number of genes that induce statistically significant nuclear abnormalities is much larger. In the absence of a cutoff, of the 608 siRNAs screened (Supplemental Table S1), 71% (429/608 genes) significantly affected nuclear EFC ratios (mean EFC range of 4.85 ± 0.07 to 9.46 ± 0.09) compared with the scrambled control (mean EFC of 6.90 ± 0.04; Supplemental Table S2). A majority (58% of gene depletions or 354/608; Supplemental Table S5) decreased EFC ratio (indicating an increase in irregular nuclear shapes), whereas about 12% of gene depletions increased EFC ratio (Figure 2D and Supplemental Table S10). The reason this high-throughput screen is sensitive enough to detect subtle changes in nuclear abnormalities is that per gene depletion, hundreds of nuclei were analyzed on average, which vastly increases the statistical power. Genes which significantly decreased EFC ratio included histone variants, histone acetyltransferases, histone deacetylases, DNA binding proteins, transcription factors, and members of prominent chromatin remodeling complexes. However, it is likely that very subtle effects on nuclear shape, which are not immediately obvious to the eye, may not be physiologically relevant.

Interestingly, our high-throughput screen for epigenetic regulators of nuclear shape revealed several families of genes and protein complexes whose individual components have effects on nuclear...
shape. For instance, in the high mobility group nucleosomal binding domain (HMGN) family of genes, we found that depletion of HMGN2 significantly decreases EFC ratio (Supplemental Table S5), while HMGN3 significantly increased the EFC ratio (Supplemental Table S10). Members of this family of genes have previously been identified to regulate nuclear mechanics. For example, overexpression of HMGN5 was shown to decrease nuclear stiffness in mouse cardiomyocytes (Rochman et al., 2010; Furusawa et al., 2015). The high-throughput screen also revealed that depletion of specific subunits of the ATP-dependent INO80 chromatin remodeling complex has differential effects on EFC ratio. Depletion of subunits encoded by INO80E and INO80D was found to significantly decrease EFC ratio (Supplemental Table S5), while depletion of INO80B was found to significantly increase EFC ratio (Supplemental Table S10).

Our results also implicate members of the histone family in regulation of nuclear shape. H1F0, which encodes the linker histone H1.0, has been shown to be down-regulated in cancer self-renewing cells in multiple cancer types. Loss of this particular histone was identified to destabilize AT-rich genomic regions thereby activating oncogenic pathways and has been correlated with worse patient outcomes across multiple cancers (Torres et al., 2016). While Torres
et al. (2016) did not investigate nuclear shape, depletion of H1.0 significantly decreased the EFC ratio (mean EFC ratio of 5.40 ± 0.05) compared with the scrambled negative control (mean EFC ratio of 6.90 ± 0.04) in our high-throughput screen (Supplemental Table S7). It is possible that destabilization of AT-rich genomic regions by loss of H1.0 may induce these nuclear shape irregularities.

Validation of a subset of genes from the high-throughput screen using confocal microscopy revealed a number of genes which cause significant nuclear abnormalities on depletion. Top hits include MT2A, TAF7, FOXP1, SUZ12, DIDD1, TP53, RADS4L2, and CXXC1. Of these, MT2A had the lowest mean EFC ratio of 4.62 ± 0.15 (mean EFC ratio of scrambled siRNA negative control was 6.93 ± 0.12) and resulted in visually apparent nuclear aberrations (Figure 3D; Supplemental Table S6). Similarly, SUZ12, DIDD1, and TP53 were also among top hits which resulted in visually aberrant nuclear morphologies (Figure 3D). SUZ12 tends to be overexpressed in cancer and is a subunit of the PRC2 complex that creates the repressive histone H3 lysine residue 27 trimethyl mark with the central enzymatic component EZH2; changes in nuclear shape might result from dysfunction of the PRC2 complex. DIDD1 located on chromosome 20q is a region that is frequently deleted in myeloproliferative neoplasms (Futterer et al., 2005). DIDD1 was recently shown to recognize the histone 3 lysine residue 4 trimethylation mark in a pH-dependent manner through a PHD domain (Tencer et al., 2017) and was identified as a possible germline variant in familial colorectal cancer (Thut-kawkorapin et al., 2019). In a pan-cancer analysis in cBioPortal (Cerami et al., 2012; Gao et al., 2013), DIDD1 is mutated in ~4% of cases, with frequent truncating mutations. Surprisingly, knockdown of lamin A/C, which is encoded by the LMNA gene, did not have comparably large effects on the mean EFC ratio in MCF-10A cells, even though the effects were statistically significant. The mean EFC ratio was 6.49 ± 0.04 for LMNA compared with 6.90 ± 0.04 for the scrambled siRNA in the high-throughput screen and 6.10 ± 0.13 for LMNA compared with 6.93 ± 0.12 for the scrambled siRNA in the confocal screen.

Since nuclear shape abnormalities can be induced purely by altering chromatin histone modification state (Stephens et al., 2018a), it is possible that a number of epigenetic regulators discovered in this paper impact the nucleus through modifications of chromatin structure. Given the significant number of genes revealed by our screen that can regulate nuclear shape, the mechanisms by which nuclear shape becomes abnormal in cancer are likely to be highly complex. Understanding the underlying mechanisms remains a central challenge.

**MATERIALS AND METHODS**

**Cell culture**

All cell types were maintained at 37°C in a humidified 5% (vol/vol) CO2 environment during culture. Human breast fibrocytic disease MCF-10A cells (ATCC CRL-10317) were cultured in DMEM/F-12 (Thermo Fisher Scientific) media supplemented with 5% donor horse serum (Fisher Scientific), 100 ng/ml cholera toxin (Sigma-Aldrich), 20 ng/ml epidermal growth factor (PeproTech Online), 0.5 mg/ml hydrocortisone (Sigma-Aldrich), 100 ng/ml insulin (Sigma-Aldrich), and 1% penicillin and streptomycin (Fisher Scientific). Human breast adenocarcinoma MDA-MB-231 cells were cultured in DMEM (Thermo Fisher Scientific) supplemented with 10% donor bovine serum (Invitrogen), 1% MEM nonessential amino acids (Corning), 1% 200 mM L-glutamine (Fisher Scientific), and 1% penicillin and streptomycin (Fisher Scientific). Cell cultures were maintained in T-75 flasks prior to experiments. To begin the high-throughput experiment, cells were trypsinized, counted, and diluted to 1 × 10^6 cells/ml in fresh cell media; 50 µl of this cell suspension were dispensed into each well of 384-well plates using a Matrix WellMate automatic cell dispenser (Thermo Fisher Scientific). Cells were allowed to spread overnight prior to transfection. For the confocal microscopy screen, cells were seeded manually at a density of 2000 cells per well into 96-well plates and allowed to adhere and spread overnight prior to transfection. MCF-10A cells were periodically validated with short tandem repeat analysis at the Interdisciplinary Center for Biotechnology at the University of Florida and checked for mycoplasma contamination.

**Transfection of siRNAs**

For high-throughput siRNA interference, MCF-10A cells were seeded into 384-well plates at a density of 500 cells/well in appropriate culture medium and incubated for 24 h. This seeding density was chosen to account for the mild cytotoxicity associated with transfection while ensuring that the cells were not confluent at the time of fixation. Prior to transfection, cells were washed once with 1× phosphate-buffered saline (PBS; Corning) and replenished with appropriate antibiotic-free culture medium. siRNA transfections were carried out using Dharmafect 1 (Dharmacon) in OptiMEM serum-free media (Thermo Fisher Scientific) per the manufacturer’s protocols, diluting to a final siRNA concentration of 25 nM. Posttransfection, cells were incubated for 72 h to allow siRNA-mediated interference and turnover of endogenous proteins to occur. The siGENOME SMARTpool siRNA library targeting 608 genes encoding epigenetic-associated proteins was obtained from Dharmacon. One additional condition was screened in quadruplicate wells per 384-well plate which served as a negative control: scrambled siRNA control (negative control; Dharmacon). For the confocal screen, cells were transfected for 72 h using the protocol described above, scaling initial cell seeding density and volumes proportionally to the increased culture dish surface area. For the siRNA optimization experiments, cells were seeded at an initial seeding density of 400 cells per well in 384-well plates and transfected using the protocol described above, varying final siRNA transfection concentration. Posttransfection, cells were incubated for 120 h to allow for siRNA-mediated interference and turnover of endogenous proteins to occur. A final siRNA transfection concentration of 25 nM was chosen to maximize development of phenotype while minimizing cell death associated with transfection (data not shown).

**Fixing, immunofluorescence, and confocal imaging**

MCF-10A cells in 384-well plates were fixed following the 72 h siRNA incubation period using an equal volume amount of 8% paraformaldehyde (Alfa Aesar) in PBS and added directly to the culture media supplemented with 1% bovine serum albumin (Fisher Scientific) for 45 min. Prior to fixation, cells were washed once with 1× PBS (Corning) and then fixed using 4% paraformaldehyde (Alfa Aesar) diluted in 1× PBS. Cells in 384-well plates were then washed with 1× PBS using a Matrix WellMate high-throughput pipetting system (Thermo Fisher Scientific). Cells in 96-well plates were manually washed with 1× PBS. Cells were then simultaneously permeabilized and blocked using a permeabilization buffer consisting of 0.1% Triton X-100 (Sigma-Aldrich) supplemented with 1% bovine serum albumin (Fisher Scientific) for 45 min. Primary antibody toward E-cadherin (ab40772, Abcam) was diluted in the permeabilization buffer described above and used at 1:500 concentration. Following incubation at 4°C overnight in the dark, cells were washed once with 1× PBS and incubated with goat anti-rabbit Alexa Fluor 594 (Life Technologies) diluted in permeabilization buffer to a final concentration of 1:500 for 1 h at room temperature in the dark. Cells were then stained with Hoechst 33342.
(Thermo Fisher Scientific) diluted to 1:200 and/or Alexa Fluor 594 Phalloidin (Thermo Fisher Scientific) diluted to 1:40 in 1x PBS for 1 h at room temperature in the dark. For confocal imaging, fixed and stained nuclei were imaged using a Nikon A1 laser scanning confocal microscope (Nikon), equipped with a 60x (1.49 NA) oil immersion objective and Nikon DU-4 CCD camera using Nikon NIS-Elements software. Z-stacks were acquired using an axial step size of 0.3 µm, and both maximum intensity projections and sum-of-slice projections were generated in ImageJ (National Institutes of Health [NIH]).

High-throughput imaging and image preprocessing

The 384-well plates were imaged using an Operetta CLS high-content imaging system (PerkinElmer) equipped with a 40x (NA of 0.95) air objective (PerkinElmer, #HH12000507). A total of 20 images, spanning the majority of the available well surface area excluding edges, were taken for each well; a minimum of three biological replicates were imaged for each siRNA condition. Images were corrected for illumination gradients using a custom CellProfiler (Carpenter Lab; Carpenter et al., 2006; Kamentsky et al., 2011) script, and both out-of-focus and mitotic nuclei were subsequently masked using a set of exclusion criteria generated from a machine learning algorithm in CellProfiler Analyst (Jones et al., 2009). For each masking process, a 15-parameter exclusion criterion was generated using a random forest classifier and training sets of approximately 800 nuclei, manually classified as either “in-focus” or “out-of-focus/mitotic.” The final images, corrected for uneven illumination and out-of-focus and mitotic nuclei, were used for all image analyses.

Nuclear morphometric analysis

To compute the EFC ratio in the high-throughput screen, raw images of Hoechst 33342-labeled nuclei imaged at the brightest equatorial plane were directly loaded into MATLAB (MATLAB 2018a and Image Processing Toolbox; MathWorks Inc) for processing. To compute the EFC ratio from confocal images, z-stack images were first loaded into ImageJ. Maximum-projection images of the Hoechst 33342 channel were then generated and saved. These maximum-projection images of the Hoechst channel were then loaded into MATLAB for processing.

Images were loaded into MATLAB (MATLAB2018a and Image Processing Toolbox; MathWorks Inc), and a built-in function was used to saturate the top and bottom 1% of all pixel values to improve image contrast. Nuclei touching the border of the image were then removed, and the image was segmented using an Otsu segmentation algorithm which automatically thresholds each image. A pixel area-size exclusion filter with empirically determined upper and lower bounds was then applied to segmented objects to exclude small imaging artifacts and large masses of touching nuclei which confound segmentation. Masks of the resulting nuclei were obtained and applied back onto the original image which was then fit with a Fourier decomposition using a custom MATLAB script. Elliptical Fourier analysis approximates nuclear shapes through decomposition of the shape into a series of harmonic ellipses (Diaz et al., 1989; Lammerding et al., 2006). The first four coefficients of each harmonic ellipse were used to calculate the major and minor axes of the harmonic ellipse fit. Nuclear contours were fit using a total of 15 harmonic ellipses, which were found to accurately approximate even the most complex, irregular nuclei. An EFC ratio was then calculated by the sum of the major and minor axes of the first harmonic ellipse, divided by the sum of the major and minor axes for the subsequent 14 harmonic ellipses. Physically, this measurement represents a relative ratio of the fit of the first harmonic ellipse to the following 14 harmonic ellipses; therefore, regular nuclear shapes (circular) will have larger EFC ratios than irregular nuclear shapes (nuclei containing blebs, invaginations, etc.).

Due to the sensitivity of elliptical Fourier analysis to subtle variations in nuclear contours, we opted to forego watershed algorithms which are traditionally employed to discern boundaries between closely bordering cells or nuclei in our method as inaccuracies in segmentation could propagate throughout the analysis and bias results. This analysis algorithm is ideal for high-throughput screening applications where image and cell numbers screened will be large to compensate for excluded nuclei.

To compute solidity (where solidity = area/convex area), raw images of Hoechst 33342-labeled nuclei imaged at the brightest equatorial plane were loaded into ImageJ. Nuclei were thresholded using the default ImageJ threshold algorithm, and holes in the resulting binary image were filled. ImageJ’s “Analyze Particles” feature was then used to quantify the shape descriptor solidity for each nucleus in Figure 1A. Adjusted solidity was calculated according to the following formula [(1-solidity)*100] to allow a measure of solidity to be plotted on the same axes as EFC ratio.

To compute circularity, defined as (4π*area)/(perimeter^2), nuclear areas and nuclear perimeters were first measured from the final nuclear contour detected after the image processing algorithm described above for use in calculation of the EFC ratio in MATLAB. Circularity was then computed in MATLAB per the described definition and used for subsequent statistical analysis.

Cell spreading area and nuclear volume measurements

Cell spreading area was quantified for several gene depletions by manually tracing the boundary of individual cells in ImageJ and using ImageJ’s built-in measurement function. E-cadherin immunostained images were used to discern individual cell boundaries and quantify spreading area. The coordinates of the final traced cell boundary region of interest were saved and subsequently used to crop the Hoechst 33342-stained nucleus corresponding to the measured cell. These cropped images were then analyzed using the method described above to obtain corresponding EFC ratios. Automated algorithms were not used to detect cell boundaries due to difficulties in segmenting cells in monolayer cultures.

To measure nuclear volume, images of individual nuclei stained with Hoechst 33342 were cropped in ImageJ and nuclear volumes were measured using the “3D Objects Counter” tool after application of an intensity threshold calculated at the equatorial x-y focal plane of the nucleus, applied to all subsequent x-y planes in the z-stack image.

Validation of control siRNA knockdown with RT-qPCR

RT-qPCR experiments were run in parallel to the bulk RNAi screen to ensure that knockdown occurred in every plate screened with the following 11 randomly chosen genes: LMNA, SMARCA4, HIST1H1T, ZRNAB3, INO80B, RAD51, GADD45A, CXXC1, HIST1H3G, TP53, and SMARCB1. These genes were further tested for knockdown on the confocal screen as well, with one additional gene target, DDO1. All SMARTpool siRNAs were obtained from Dharmacon (see Supplemental Table S1 for information on the siRNAs, including sequences targeted).

Whole cells were lysed using the SingleShot Probes One-Step Kit (Bio-Rad), designed for low-cell number inputs, 72 h posttransfection. Lysates were then combined with complete RT-qPCR reaction mix per the manufacturer’s protocol and presdesigned fluorogenic probes targeting LMNA, SMARCA4, or GAPDH (Bio-Rad). Reactions were processed on a CFX-96 Real-time PCR detection system.
(Bio-Rad) using the following thermocycler conditions: 10 min at 50°C, 3 min at 95°C, followed by 40 cycles of 15 s at 95°C and 30 s at 60°C. All results were analyzed using the 2-ΔΔCT method and normalized against GAPDH expression (Livak and Schmittgen, 2001).

Statistics/replicates
Confidence in the EFC ratios was determined using a Kruskal–Wallis one-way analysis of variance nonparametric statistical test (Kruskal and Wallis, 1952) to determine if any of the measured gene conditions originated from a different distribution than other gene depletion conditions; specifically, comparisons to the scrambled siRNA negative control condition were used for interpretation of statistical significance. Data derived from biological samples and cell culture experiments rarely follow normal distributions and, when maintaining consistent microscopy image acquisition conditions, rarely allow for equal sample numbers between conditions. Therefore, the Kruskal–Wallis one-way analysis of variance test was used for most analyses as this test does not assume equal sample sizes between conditions or that samples follow a specific distribution. Subsequently, a two-tailed Dunn's test which allows for multiple comparisons was used to determine which, if any, of the gene depletion conditions originated from a different distribution than any other gene depletion condition; again, specifically, comparison to the scrambled siRNA negative control condition was used to determine statistical significance for our experiments. As the high-throughput screen has many unique gene deletions (genes screened), familywise error rates were corrected for using a two-tailed Bonferroni correction to a significance level of 5%. For our analysis, we were specifically interested if experimental conditions originated from a different distribution than our scrambled siRNA negative control condition. However, as all possible comparisons between conditions were still made and type I error was corrected for using Bonferroni corrections, our analysis represents a conservative listing of conditions identified to influence nuclear morphology. We note that there is high statistical power due to the large sample numbers in the high-throughput screen.

ACKNOWLEDGMENTS
This work was supported by the NIH (R01 EB014869 to T.P.L.; K99GM123195 to A.D.S.), the National Cancer Institute (R01 CA172310 to H.L.; US4CA193419 to J.D.L.), the National Cancer Institute Research Specialist Award (R50CA211487 to R.R.), and a supplement from the Northwestern PSONC to T.P.L. and J.D.L.

REFERENCES

CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 7, R100.

Imbalzano KM, Cohet N, Wu Q, Underwood JM, Imbalzano AN, Nickerson JA (2013). Nuclear shape changes are induced by knockdown of the SWI/SNF ATPase BRG1 and are independent of cytoskeletal connections. PLoS One 8, e55628.


