

Grenfell et al., <http://www.jcb.org/cgi/content/full/jcb.201509079/DC1>

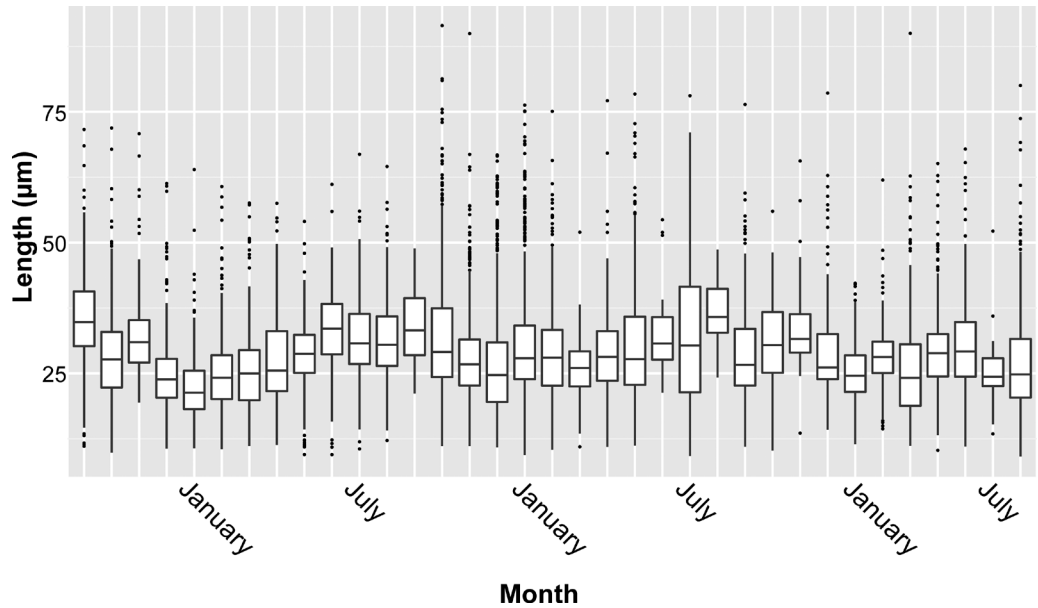


Figure S1. **Spindle length follows a cyclical pattern throughout the year.** Spindle length averages each month are plotted over 3 yr (2013–2015).

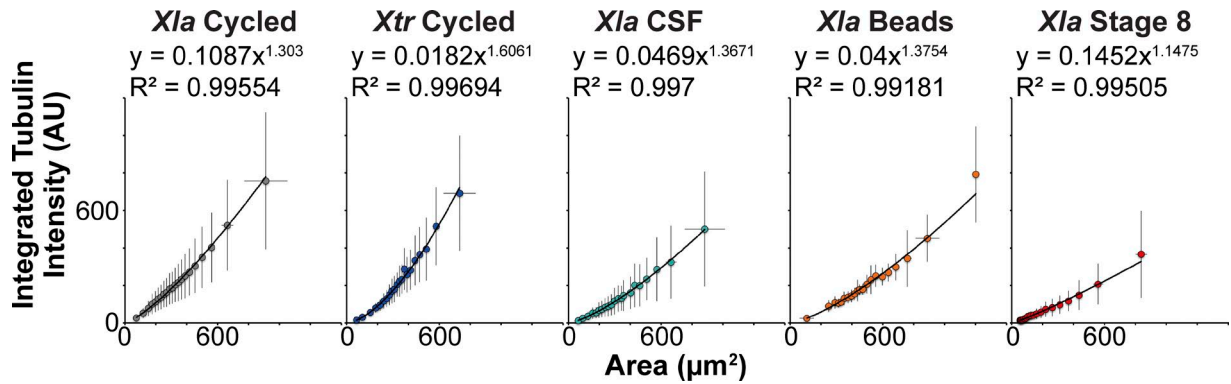


Figure S2. ***Xenopus* extract spindle types differ in composition.** Plot from Fig. 4 A separated by spindle types; equations are best-fit power functions to assess scaling relationships. Error bars represent standard deviation.

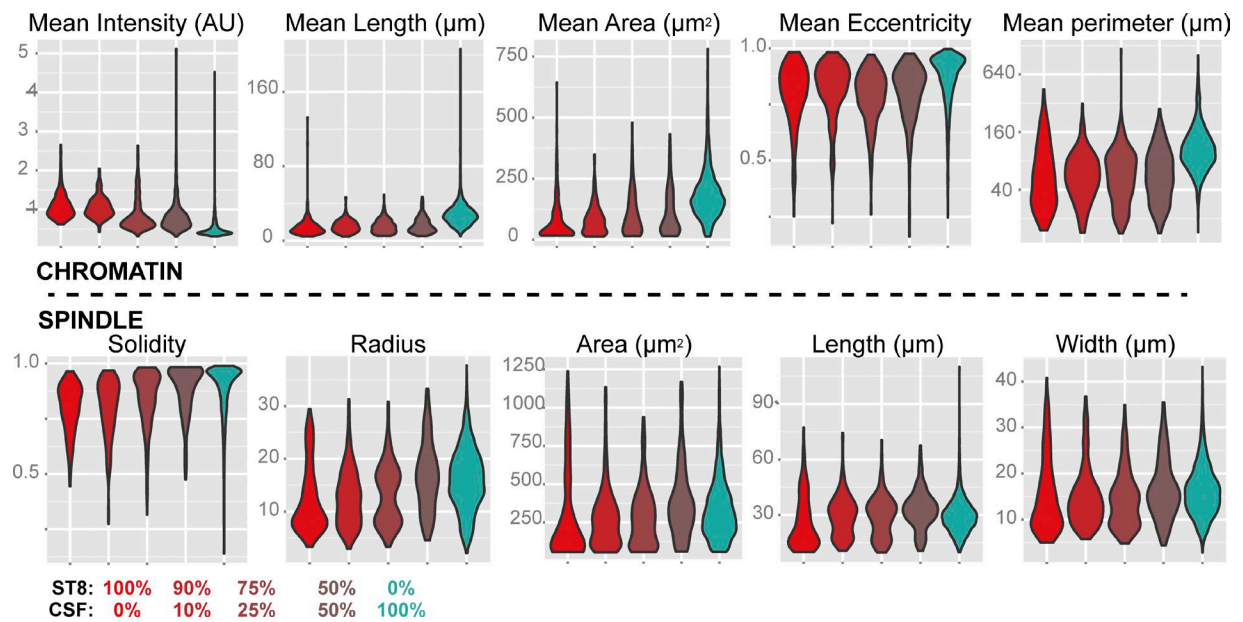


Figure S3. **Some spindle and chromatin features titrate in mixing experiments while others display a bimodal distribution.** Distributions of selected chromatin and spindle that differ most between CSF and embryonic stage 8 spindles.

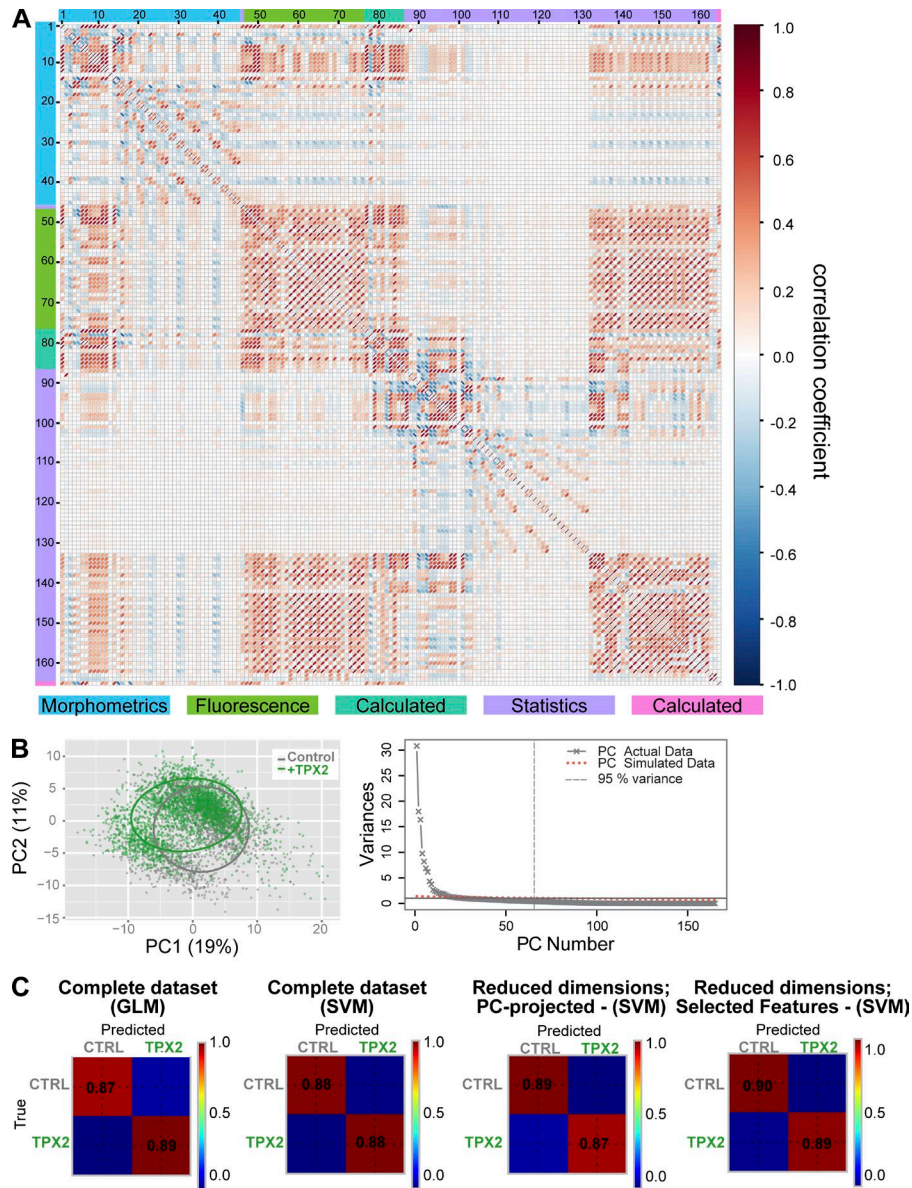


Figure S4. **Approaches taken to define feature space to build a model for unambiguous classification of TPX2-addition phenotype spindles.** (A) Correlation matrix for the complete TPX2 dataset. Pairwise correlations are color-coded by correlation coefficient. Variables are numbered according to their order of appearance in the output comma-separated values file (CSV), and color-coded according to CellProfiler module. The matrix shows that several features strongly positively (red) or negatively (blue) correlated with each other, thus introducing redundant information to the classification model. (B) Projection of the complete dataset (166 features) on first two principle components (left), which capture 30% of the variance. As this is only a small fraction of the variation in the dataset, further components were considered. Scree plot for principle component analysis (right) to visualize fraction of total variation captured as additional principle components are included. 64 components account for 95% of total variance (gray dashed vertical line). (C) Performance of classification models tested to automatically assign spindles to TPX2-supplemented and control classes. Simple, generalized linear model (GLM; ridge-penalized logistic regression) and complex, support vector machines (SVM) with radial kernel. Models were trained on complete dataset (166 features), as well as on datasets with reduced dimensions (64 principle components; PC, principle component, projected) and features selected based on pair-wise univariate significance tests between the two classes (e.g., Kolmogorov-Smirnov test; Fig. 6). ST8, stage 8.

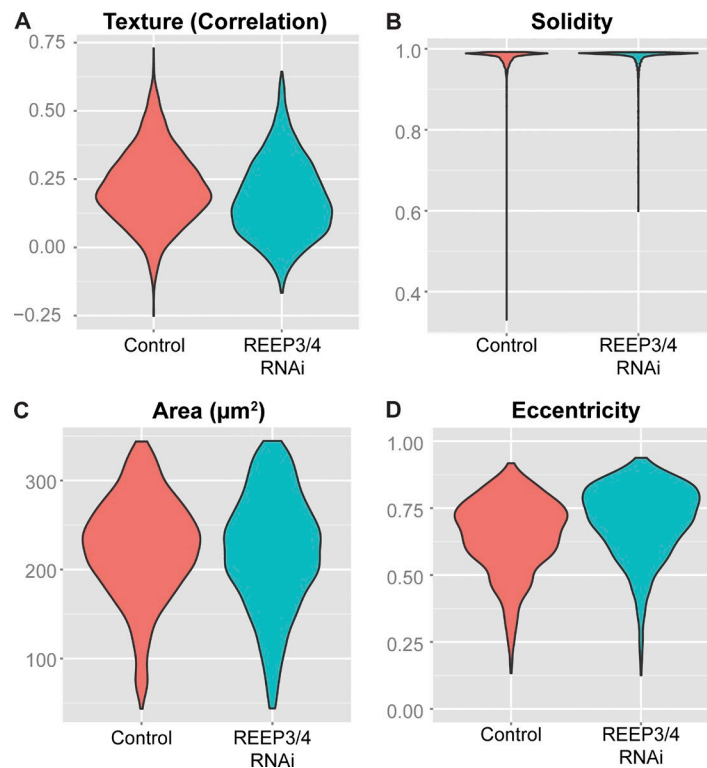


Figure S5. **An independently developed analysis pipeline reveals similar nuclear characteristics upon REEP3/4 depletion.** (A) Lamin B1 texture of control and REEP3/4-depleted cells. (B) Nuclear solidity of control and REEP3/4-depleted cells. (C) Nuclear area of control and REEP3/4-depleted cells. (D) Eccentricity of control and REEP3/4-depleted cells.

Table S1. **Locations of feature distribution peaks**

ID	CSF	Stage 8	Solidity		Aspect ratio		Length	
1	0	100	0.88	<i>0.79</i>	0.62	<i>0.76</i>	15	<i>44</i>
2	10	90	0.9	—	0.42	<i>0.66</i>	32	<i>17</i>
3	25	75	0.94	—	0.46	—	33	<i>16</i>
4	50	50	0.96	—	0.44	<i>0.56</i>	33	<i>17</i>
5	100	0	0.96	—	0.58	—	28	—

Locations of peaks marked with asterisks in Fig. 3 B. Values were determined as locations of maxima of Gaussian kernel smoothed distributions as described in Materials and methods. Minor peaks are marked in *italics*. Minor peaks, which have values 10 times lower than major peaks, are filtered out.

Provided on the following pages is a PDF of instructions for the CellProfiler pipeline.

Pipeline instructions

The instructions below are for the CellProfiler pipeline published in **A versatile multivariate image analysis pipeline reveals features of *Xenopus* extract spindles**. Feel free to contact Andrew Grenfell (awgrenfell.lab@gmail.com) or Magdalena Strzelecka (strzelecka.em@gmail.com) with any questions. Images analyzed in our publication are available through Xenbase (<ftp://xenbaseturbofrog.org/>, [choose guest login when prompted](#)). Raw data and pipelines (for analyzing spindles, microtubule asters, and Lamin B1 stained nuclei) are available through Xenbase and github (<ftp://xenbaseturbofrog.org/> and <https://github.com/MultivariateSpindleResource>, respectively). All data available through Xenbase is in the sequence_information folder.

Hardware recommendations

Essentially all modern laptops have the computational power to run any of the CellProfiler pipelines that we developed. However, we recommend using a desktop computer for analysis simply because they are much more efficiently cooled, and laptops tend to overheat (especially when CellProfiler runs are long). If a desktop computer is not available, we recommend using some kind of cooling system to help prevent heat-related damage to your laptop. Also, it is not at all necessary to build a computer for analysis as we did; however, in case the runs become a routine activity, it may be much more cost effective to do so.

Organizing the input images

Image metadata is assigned using a Regular Expression that reads the file path of the images. This can be easily changed in the Metadata Input Module (found at the top left of CellProfiler window). However, our organization scheme is laid out below for those that do not want to make changes to the metadata handling of the pipeline.

Folder structure is case sensitive and follows the format:

Raw/AWG/Exp00/1_000_CYC/40xMM

Raw: folder containing all of the raw images, must be an exact match of text (i.e. Raw)

AWG: one or more capital letters corresponding to the experimenter

Exp00: the letters Exp (must match case) followed by one or more numbers corresponding to the experiment number

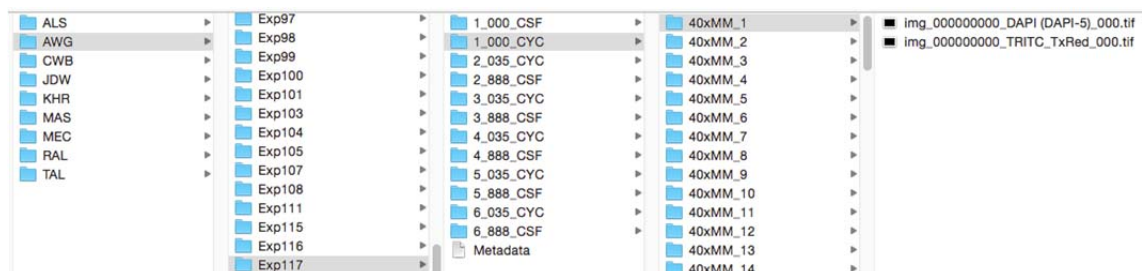
1_000_CYC:

- A. **1:** one or more numbers corresponding to the coverslip number
- B. **000:** one or more numbers corresponding to the treatment
- C. **CYC:** one or more capital letters corresponding to the reaction type (e.g., CYC for cycled)

40xMM: any text, we use this to describe the microscope and objective used. This is the folder containing the raw image set (i.e. corresponding tubulin and DNA images)! Each of these folders contains a single tubulin and a single DNA image.

Images are identified as either red (Rhodamine tubulin) or blue (DNA) channels by matching unique text found in the image file name. For our images, TRITC is found in the red image file name and DNA is found in the blue image file name. If your images do not follow this convention, simply change the matching criteria in the NamesAndTypes Input modules (found at the top left side of the CellProfiler window). Make sure that the text you choose is unique to a single channel!

For an example of this layout please see the screenshot below:



Preparing the pipeline and checking the segmentation scheme

The instructions below outline steps to prepare the *Xenopus* spindle image analysis pipeline for use, and to verify that the segmentation scheme is performing well.

1. Download and install CellProfiler (<http://cellprofiler.org/download.html>).
 - a. **Note:** we used CellProfiler v. 2.1.0 and 2.1.1 for our analyses. Pipelines are generally forward compatible, but previous releases of CellProfiler are available for download (<http://cellprofiler.org/previousReleases.html>).
2. Open CellProfiler and load the pipeline.
 - a. **Note:** if loading from a .cppipe file go to File>Import>Pipeline from file, CellProfiler will convert the .cppipe file to a .cpproj file automatically. If loading from a .cppipe file the pipeline can be saved as .cpproj.
3. Change the object diameter (these values are expressed in pixels) in the IdentifyPrimaryObject modules (modules 8, 19 and 20) to account for the scale of your imaging setup. The current values are for an imaging setup where 1 μm = 6.18 pixels.
 - a. The values in the CalculateMath modules that convert the pixel measurements to μm must also be changed (Modules 28, 30, 31, 34–36).
4. Drag and drop image folders into the open box that says “Drop files and folders here” in the Images Input Module.
 - a. If there are images in this box remove them before adding your images by right clicking in the box and selecting “Clear File List.”
5. Select the output folder (where the data will be stored) by clicking the “View output settings” at the bottom left side of the window and navigating to the appropriate folder/entering the correct path in the text box.
6. Check the segmentation by entering test mode (hitting F5 on Mac or the “Start Test Mode” button at the bottom left of the screen), making the IdentifyPrimaryObjects modules visible (modules 8, 19, 20) by clicking the eye next to each module, making the FilterObjects modules visible (modules 23 and 24), and then pushing the “Run” button at the bottom left of the screen.

- a. **Note:** the pause buttons next to each module allow you to run only part of the pipeline in test mode. If any of these are yellow it means that the pipeline will pause at that module and will not continue until you hit the Run button again.
7. One window per module will pop up as the pipeline analyzes the first image set. First check the IdentifyPrimaryObjects windows (focusing on the “Chromatin” and “Spindles” modules (Modules 8 and 20, respectively); windows corresponding to these modules can be brought to the top by double clicking on the relevant module in the CellProfiler window).
8. If the identifiable microtubule and chromatin masses are correctly segmented (see note below) then continue to Step 8. If identifiable microtubule and chromatin masses are not correctly segmented then see the “**Tuning the Segmentation Scheme**” section below for tips on modifying the segmentation scheme.
 - a. **Note:** Objects that are outlined in green are putative chromatin or spindle objects, those outlined in magenta did not pass segmentation, and those outlined in yellow were rejected because they are touching the image border. Identifying too many objects at this stage is not a problem. Also, structures segmented in the PreSpindles module (#19) will likely be much larger than the actual spindles; this is okay as long as the structures in the Spindles module (#20) are segmented correctly.
9. Check the FilterObjects modules to insure that microtubule and chromatin objects are retained after filtering.
 - a. **Note:** Similar to the segmentation modules, objects outlined in green passed filtering and will be measured, and those outlined in red did not pass filtering and will be discarded. If a significant number of the segmented objects retained after filtering are not spindles, or if a significant number of spindles fail to be retained then the segmentation scheme should be altered; see “**Tuning the Segmentation Scheme**” below
10. Check the next image set by clicking the “Next Image Set” button at the bottom left of the screen.
 - a. **Note:** It is very important that you check images from both control and treated reactions to ensure that the segmentation performs well across the entire range of spindle morphologies in your experiment. We recommend checking the segmentation of at least 10 images from each treatment.
11. Hide the IdentifyPrimaryObjects and FilterObjects modules (by clicking the eye again) before running the pipeline. This is not strictly necessary, but leaving them visible increases the execution time.
 - a. **Note:** you must exit test mode before you can run the pipeline for analysis.

Tuning the segmentation scheme

If the microtubule and chromatin structures retained after the FilterObjects modules do not accurately represent the structures present in the image it may be necessary to alter the segmentation scheme. The instructions below are the easy fixes. If you are still having problems, feel free to email us or consult the CellProfiler manual (<http://d1zypm9ayga15t.cloudfront.net/CellProfilerManual.pdf>). There are two general scenarios: identifying too many objects/identifying objects larger than the true objects and identifying too few/identifying objects that are smaller than the true objects.

Too many objects identified/identified objects are larger than true objects

1. The easiest fix for this is to increase the Threshold Correction Factor in the IdentifyPrimaryObjects module.
 - a. For chromatin this is Module 8
 - b. For microtubules this is Module 20. The Threshold Correction Factor can also be changed in Module 19, but we recommend changing Module 20 first.
2. If the first approach does not work, it is possible that there is a problem with the composite image. Before making any of the changes below, make sure you reset the Threshold Correction Factor to 1.
 - a. Reveal the ImageMath module corresponding to the structure that is not being segmented properly (click on the eye next to the module name)
 - i. For chromatin this is Module #6
 - ii. For Microtubules this is Module #17.
 - b. Run the pipeline in test mode and examine the window from the ImageMath module.
 - c. If the structures of interest (e.g., microtubule structures) are difficult to distinguish from background in any of the images (e.g., Rhodamine, RhodamineEdgedRescaled, etc.) then set the image weight to zero by changing the “Multiple the nth image by” box to 0.
 - d. The weighting of the component images can also be changed to intermediate levels or further increased by changing the “Multiply the nth image by” box.
 - e. **Note:** CellProfiler has an excellent feature that lets you systematically test different image weightings, which we highly recommend if you are going to change the segmentation scheme in this way. This feature is under the Sampling menu; please refer to the CellProfiler manual for more information.

Too few objects identified/identified objects are smaller than true objects

1. The easiest fix for this is to decrease the Threshold Correction Factor in the IdentifyPrimaryObjects module.
 - a. For chromatin this is Module 8
 - b. For microtubules this is Module 19. The Threshold Correction Factor can also be changed in Module 20, but we recommend changing Module 19 first.
 - i. **Note:** this is opposite to what we recommend above.
2. If the first approach does not work, it is possible that there is a problem with the composite image. Before making any of the changes below, make sure you reset the Threshold Correction Factor to 1.
 - a. Reveal the ImageMath module corresponding to the structure that is not being segmented properly (click on the eye next to the module name)
 - i. For chromatin this is Module #6
 - ii. For Microtubules this is Module #17.
 - b. Run the pipeline in test mode and examine the window from the ImageMath module.
 - c. If the structures of interest (e.g., microtubule structures) are difficult to distinguish from background in any of the images (e.g., Rhodamine, RhodamineEdgedRescaled, etc.) then set the image weight to zero by changing the “Multiple the nth image by” box to 0.
 - d. Use the Sampling feature (see note in previous section) to systematically test the combinations of image weights.

Running the pipeline

1. Start CellProfiler and open the pipeline (with the modified segmentation scheme if necessary)
2. Drag and drop image sets into the Images Input Module. Make sure to clear the file list before adding your images
3. Click the analyze images button at the bottom left of the CellProfiler window
 - a. **Note:** if you manually entered a file path in the Output Settings a warning will appear asking if you want to create the output folder, click yes
4. The pipeline saves an overlay of the spindle outlines and the tubulin image. We recommend checking these images to make sure that your segmentation scheme is effective; however, if this is no longer necessary, delete the OverlayOutlines and SaveImages modules at the end of the pipeline.

Analyzing the output data

For quick data exploration it is possible to use CellProfiler Analyst (Jones et al., 2008) since it is integrated with CellProfiler output data and is, in general, quite user friendly (i.e. it provides a graphical user interface). For more in-depth analyses we prefer to use any of the scripting languages (e.g., R, Python), since they provide better and more versatile functionality and larger community support. However, it is also possible to analyze the output data using Excel. The measurements for chromatin component of the spindles are contained in FilteredChromatin.csv and the measurements for microtubule component of the spindles are contained in FilteredSpindles.csv. For those interested in using a scripting language for data exploration and analysis, (including data modeling approaches) data, code, additional example plots and instructions are available at: <https://github.com/MultivariateSpindleResource>.

Note: each row of the CSV outputs corresponds to a single chromatin structure or a single microtubule structure (i.e. object). The metadata associated with each image set (outlined in **Organizing the input images**) are also contained in the CSV allowing you to sort by experiment number, treatment, etc. Also, since the metaphase plate is occasionally discontinuous there may be more chromatin objects than spindle objects. Individual output files can, however, be merged.

Applying the pipeline to Lamin B1–stained nuclei

Below are the steps taken to adapt the pipeline to segmenting HeLa cell nuclei stained with Lamin B1. When making changes such as those outlined below, CellProfiler will mark modules that are configured incorrectly, so finding downstream modules affected by a change is easily accomplished.

1. Modifications to the input modules (top left side of the screen) were as follows:
 - a. In the Metadata module, the full image name was used as the image “Set.”
 - b. In the NamesandTypes module the Rhodamine image was removed (these datasets only contain images from a single channel), and the DNA image was renamed Lamin with matching text LaminB1 contained in the image file name.
2. The chromatin segmentation scheme was used to segment nuclei, modifications were as follows:
 - a. Object diameter range was changed in IdentifyPrimaryObjects (Module #8) to 50-200 pixels.

- i. **Note:** Both FIJI and CellProfiler have tools that allow you to determine object diameter.
 - b. In the same module (IdentifyPrimaryObjects, Module #8), both “Method to distinguish clumped objects” and “Method to draw dividing lines between clumped objects” were changed to “Shape.”
 - c. In the ImageMath module corresponding to the Lamin image (Module #6), the weighting of DNAEdgedRescaled and DNARescaled were set to 0 by entering 0 into the boxes labeled “Multiple second/third image by.”
3. Modifications to the measurement modules were as follows:
 - a. MeasureObjectIntensity:
 - i. Removed Rhodamine Image
 - ii. Changed DNA image to Lamin image in “Select image to measure”
 - iii. Removed FilteredSpindles from “Select objects to measure”
 - iv. Changed FilteredChromatin to Chromatin in “Select objects to measure”
 - b. MeasureObjectSizeShape:
 - i. Removed FilteredSpindles from “Select objects to measure”
 - ii. Changed FilteredChromatin to Chromatin in “Select objects to measure”
 - c. Added MeasureObjectRadialDistribution, selected Lamin as the image to measure, and selected Chromatin as the objects to measure.
 - d. Added MeasureTexture, selected Lamin as the image to measure, and selected Chromatin as the objects to measure.
 - i. Also added 5, 10, 15, 20, and 25 pixel scales to use for measuring texture (see CellProfiler documentation and citations therein for more information on texture measurements).
 - e. Removed all CalculateMath modules.

Building an alternative pipeline for segmentation and quantification of human nuclei

As segmentation of human nuclei is a fairly established procedure, we used the following approach to build an alternative pipeline to segment Lamin B1 stained HeLa cell nuclei (use Fiji [ImageJ]) to estimate certain starting parameters before you attempt adapting/building the pipeline).

1. We extracted individual optical slices from the original LSM z-stacks using custom ImageJ macro (it should be noted that the CellProfiler can deal with stacks in lsm format); the macro (batch_split_Z_stack.ijm) can be found at <https://github.com/MultivariateSpindleResource/Super-Spindle-DataSet/tree/master/Code>.
2. We renamed images to allow for extraction of metadata from image names (in the Terminal window, once in correct directory, type: rename 's/R3_4/R3-4/' *).
3. We specified in the Image module of CellProfiler that only files in tiff format should be analyzed.
4. We specified the regular expression for identification of images and extracting metadata as: `^(?P<ExperimentID>.*)(?P<Treatment>[a-z,A-Z,0-9,-]+)(?P<Stain>.*)(?P<ImageID>[0-9]+).tiff`.
5. Given the quality of staining and imaging, the preprocessing steps were limited to smoothing with a Gaussian filter to limit the impact of noise on segmentation steps.
6. Segmentation was done following maximum correlation thresholding method (suitable for images with sparse foreground signal densities, e.g., the outlines of nuclei from lamin

staining; see CellProfiler documentation and citations therein for more information) and thresholding was done on tile-by-tile basis to account for local illumination differences.

7. Additional steps were added to capture potential phenotypical differences identified upon initial image evaluation in FIJI (disclaimer: all of these approaches are approximate and designed to capture differences, not to derive absolute quantification results):
 - a. To calculate approximate signal ratio between the lamina and nuclear interior, detected nuclei were shrunk and the shrunk version of objects was subtracted from the original objects to obtain lamina rings.

Quantification steps were partially adapted from the original pipeline (see: above and <https://github.com/MultivariateSpindleResource/Super-Spindle-DataSet/tree/master/Pipelines>) and adapted to also capture variables of interest described above; texture quantification was also added.

Reference

Jones, T.R., I. Kang, D.B. Wheeler, R.A. Lindquist, A. Papallo, D.M. Sabatini, P. Golland, and A.E. Carpenter. 2008. CellProfiler Analyst: data exploration and analysis software for complex image-based screens. *BMC Bioinformatics*. 9:482.