Image-based pooled whole genome CRISPRi screening for subcellular phenotypes

Gil Kanfer, Shireen Sarraf, Yaakov Maman, Heather Baldwin, Eunice Dominguez Martin, Kory Johnson, Michael Ward, Martin Kampmann, Jennifer Lippincott-Schwartz, and Richard Youle

Corresponding Author(s): Richard Youle, NIH and Gil Kanfer, National Institute of Neurological Disorders and Stroke (NIH)

Review Timeline:		
	Submission Date:	2020-06-28
	Editorial Decision:	2020-08-13
	Revision Received:	2020-10-17
	Editorial Decision:	2020-11-13
	Revision Received:	2020-11-20

Monitoring Editor: Jodi Nunnari

Scientific Editor: Tim Spencer

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

August 13, 2020

Re: JCB manuscript #202006180

Dr. Richard J Youle
National Institute of Neurological Disorders and Stroke
10213 Montgomery Ave
Kensington, MD 20895

Dear Richard,

Thank you for submitting your Tools manuscript entitled "Image-based pooled whole genome CRISPR screening for Parkin and TFEB subcellular localization". We apologize for the delay in providing you with a decision. The manuscript has been evaluated by three expert reviewers, one of which also assessed the companion paper from Ron Vale's group, and their reports are appended below.

As you will see, although two of the reviewers expressed enthusiasm for the utility of this methodology for the cell biology community, they also raised a number of significant and valid concerns. One primary issue is the lack of adequate information and sufficient description of key methodology and method validation. More expansive descriptions of the methodology and pipeline/workflow are necessary. There is insufficient description and analysis of the Parkin recruitment and TFEB nuclear localization screens, which validate the approach. In addition, the reviewers point out that data presentation and writing are not accurate, and the manuscript does not place and compare this approach in the context of previously published photoactivatable screening approaches.

The extent of revisions necessary to address the reviewers' concerns are significant. Thus, we are returning it to you.
However, given interest in the topic, we would be open to an appeal and potential resubmission to JCB of a significantly revised and extended manuscript that completely addresses the reviewers' concerns. We should note that, while we agree with reviewer #1 that a mechanistic extension would be interesting and increase the impact of the work, we don't feel that it would be necessary for a Tools manuscript and so we would not require it for appeal/resubmission.

As you know, you may contact the journal office to discuss an appeal of this decision or you may submit an appeal directly through our manuscript submission system once you have completed your revisions. Please note that the paper would, of course, be subject to re-review by the same reviewers (if possible).

Regardless of how you choose to proceed, we hope that the comments below will prove constructive as your work progresses.

Thank you again for allowing us to consider this work.

Sincerely,

Jodi Nunnari, Ph.D.
Editor-in-Chief
The Journal of Cell Biology

Tim Spencer, PhD
Executive Editor
Journal of Cell Biology

Reviewer #1 (Comments to the Authors (Required)):

The manuscript by Kanfer, Youle and coworkers is built on the idea to implement a CRISPRi-based screen by using imaging of adherent cells that express a fluorescent reporter - which allows them to identify specific changes in reporter localization - and then select individual cells with a particular reporter signal by photoactivation for subsequent FACS sorting. While the idea of using imaging for feature selection for CRISPR screening is not totally new (for example Goda's group, 2018, Cell 175:266-276 used this for cells in solution), I think this is the first report of using imaging of adherent cells for CRISPR screening and also to make use of a photoactivation step for cell selection in this format. Given the large number of available fluorescent reporters based on localization changes in cells, there are many potential applications for this type of screening pipeline. In addition to the novelty of using adherent cells and selection by photoactivation in a CRISPR screen, the main novelty of their approach is the use of machine learning and convolutional neural network models for image analysis that allow them select which cells to mark by photoactivation. To demonstrate the validity of their approach, they show the use of the pipeline for two types of screens. The second is a genome-wide screen on TFEB localization that identifies potentially new regulators of the nuclear localization of TFEB which is a critical transcriptional feedback mechanisms in the mTORC1 signal transduction pathway.

Given that the use of machine learning and image analysis in a CRISPR screen is not totally new, I expected more mechanistic data demonstrating that the pipeline and approach they developed can lead to finding a mechanistically relevant new regulator of the mTORC1/TFEB signaling pathway. This is my only major point that requires that they provide some additional mechanistic data on at least one of their hits. The other minor points I am adding below can likely be addressed by changes to the representation of data and analysis.

Major point:
Some more functional characterization is needed for at least one of their hits from the CRISPRi screen. For example, they could focus on the phosphatase inhibitor PP1R1B (DARP32) for which there is a plausible hypothesis of how it may work in the mTORC1 pathway. This could strengthen their argument that this type of screen is useful to find new regulators of nuclear translocation and ultimately understand molecular mechanisms. They could for example ask whether PP1R1B acts directly in the mTORC1 pathway by directly or indirectly regulating TFEB phosphorylation (which is, as I understand, the main mechanism of TFEB nuclear localization). An effect on the rate of TFEB dephosphorylation could for example be measured by using acute inhibition of protein phosphatases (using ocadeic acid or similar compounds; or alternatively a mTORC1 inhibitor such as rapamycin) and then monitoring the nuclear localization of GFP-TFEB. A related test is to measure whether mTORC1 activity for the sgRNA targeting PP1R1B is different by measuring one or more of the known substrates such as S6K or 4E-BP.

Minor points:

1. Why did SVM classification fail to predict TFEB nuclear localization accurately - TFEB localization seems to be a clear difference in distribution that should be readily captured by an SVM algorithm. More detailed is needed why a SVM classification dies not work in this case.
2. A more detailed analysis should be provided how the three subpoolscreen in the TFEB screen differ from each other. This would allow one to better understand how reproducible the screen is. In addition, they should add a discussion how many cells and subpool screens are needed to reach saturation using this method. For example, would a doubling of the number of screens or cells still significantly increase accuracy?
3. The imaging step with one second and an analysis of 5 seconds are both quite long. Is there a way to accelerate this? A discussion should be added how a faster microscope and faster analysis can accelerate these steps.
4. They also need to discuss the limitation of photoactivation in more detail. Does it take for each cell 3 sec of photoactivation - would it be 30 seconds if they would select 10 cells in the image? Could they instead use a stronger laser or different fluorescent protein to make the photoactivation step much shorter? A discussion is needed whether/how the time to image, analyze and photoactivate could be at least 10 times faster to make the method more practically useful considering the long-time it took them to complete the subpool screens.
5. The entire Fig 2 is more of a methodological detail and should be moved into the suppl. materials section.
6. More precise experimental details are needed in all Fig legends (or in the main text describing the figures). This will allow the reader to better understand what the actual experiments were and how the data was processed without having to consult the methods section.
7. Fig 6 lacks a control sgRNA. Also, fewer time points would make the same point in the figure (a focus on the time points where the difference in localization is greatest relative to control would be best).

Reviewer #2 (Comments to the Authors (Required)):

The authors present a novel pooled screening approach based on CRISPRi and machine learning. After reading the abstract of this paper I was very excited to read the whole paper as the premise, using machine learning to identify 'hits' in relocalization screens coupled with fluorescence activation and cell sorting to retrieve them, I think is very clever and thorough validation would be a tremendous advance for the community.

This "Tool" article is divided into two main parts. Firstly, a proof-of-concept experiment by using the known Parkin recruitment to mitochondria by PINK1 and secondly, a genome-wide screen identifying new effectors of TFEB shuttling.

In both studies machine learning is used to rapidly identify candidate hits for fluorescence activation. A relatively straightforward SVM classifier was used for the proof-of-concept study and a more sophisticated deep learning approach was used for the TFEB screen.

As a "Tool" article it must be relatively straight forward for the readership of JCB to follow the implementation of the method being described. However, I found the manuscript extremely difficult

to follow and I work in the area. Furthermore, this kind of tool article should clearly demonstrate the capabilities and limitations of the "tool". Again, I found this information very difficult to find in the text and even more difficult to fully appreciate.

For the 'proof-of-principle' screen "The Draq5 channel is used for nuclear detection and the GFP-Parkin image is used to identify cell borders" Yet the first mention of Draq5 is in the methods section that follows the main manuscript and the quote above is from the caption for a supplementary figure. Even after reading the whole manuscript and looking for the details - it is not clear to me how individual cells were identified and segmented.

Looking at the figures, the segmentation is not very good which is surprising because it is usually quite facile to segment cells using only the Draq5 channel and here the authors used two channels. Figure 2a supposedly outlines the steps in segmentation but starts with an image of mitochondria which was clearly not used to create the mask in the next image. This second image is the individual cells and is used for the coordinates for photoactivation. I am not clear what the third image is or what it is for. Below they show the resulting segmentation in which there are clearly multiple instances of single segmentation boundaries surrounding multiple cells. I don't think a Draq5 image is shown anywhere.

When using the Halo-tag in the second screen the authors again do not show any halo-tag images. And the image that they do show includes "a cell" complete with nucleus and cytoplasm (olive/lime green nucleus mask near the center of figure 4f) where there is no evidence of a cell at all.

The authors show convincingly the proof of concept by detecting PINK1 as most significant hit in the Parkin translocation screen. However, assuming that I followed the manuscript and understand Figure 3c this conclusion is overstated. It was actually because I could not understand the location of the PINK1 dot in the figure that I started looking carefully in the manuscript. It appears that PINK1 guides were doped into the library at 10%. It is not at all surprising that they were able to retrieve positive control guides that constituted 10% of the total and makes the claims of sensitivity for the screen highly overstated. To claim that three replicates are needed based on this data is very suspect. Looking at the rest of the data in 3b, I see no clear delineation of potential hits - which dot is PINK1?

Feature selection is a crucial part for a successful classification. It would be of interest what features are selected in the SVM classifier. How much do the features vary? Please explain how the features were reduced. I think that the idea is that highly correlated features were dropped - which is a standard procedure but not clear here.

Actually, I fail to see the complexity in either screen and the need for machine learning. In my experience mitochondrial localization as opposed to cytoplasmic localization can be readily scored by simply measuring the variance across the cytoplasmic area of the cell. Similarly the training images shown for the nuclear cytoplasmic localization determination - probably the single most common high content screen performed by the community - are sufficiently clear that the authors should not have needed to use deep learning to score the cells.

Furthermore, they present in the TFEB screen 64 genes that cause a retention of TFEB within the nucleus. In a second validation screen 21 out of 64 were validated. But beyond that the phenomenon could be reproduced, the validation is not comprehensive. Maybe I am missing something but I don't follow the arguments at the top of page 7 suggesting the results validate the screen. In addition to standard validation approaches why don't the authors measure changes in

TFEB regulated gene expression?

While likely the problem is in the presentation of the data, I found the precision recall curves difficult to understand for the TFEB screen. How is this being calculated? What is the gold standard? If you don't know the answer how is it possible to calculate a precision-recall curve?

Overall - as attractive as the concept presented in the manuscript is, the manuscript needs a full renovation if it is to have impact in the community.

Minor points:
For the authors consideration.
Why is there a probability value discrepancy in Fig. 6 a (top panel, 1h, sgTGFBR1). Two nuclei side by side with similar intensity but a notable difference in the probability values. Given the problems with the segmentation alluded to above and no segmentation data shown here it is hard for me to interpret the numbers provided.

Both dCas9 (pC13N-dCas9-BFP-KRAB) and sgRNAs (mU6-BstXI-BlpI-BFP vector) are tagged with BFP. It would be helpful if the authors briefly describe (in addition to the citation "Tian et al., 2019") in the method section how only the sgRNA expression is assessed by the BFP signal.

Fig 1 g and Fig. 2 f: typo "field of view"
Figure caption 3: typo "fluorescence"
Page 5 wrong citation formatting
Throughout the text the authors write 12,500 sgRNAs, however on page 4 they mention 12,775 guides.
For better visualization it is recommended to rethink the font color, e.g. in Fig S2b. Some yellow digits are hard to read, especially on a white background. Also in Fig S4, the red digits are not easy to read. The red circle is a bit too thin and might be overlooked.
Fig. S 2 b, it would be helpful if the authors could label the Pa mCh channel like they did in Fig. 4 f with "pre-activated" and "activated".
Fig. 4 b, TC medium was used in the figure, however the authors used complete medium (cm) in the figure caption.
Could you specify (page 17, "initially trained 2,234 images") if images means individual cells.
Please be consistent in writing PINK1.
On page 23, please could the authors write out FC as fold change once and then abbreviate it.
In Fig. S1 c, the authors show an immunofluorescence experiment on CDH2 and TRANS to assess potent dCas9 clones. However, in the method section this experiment is not described.
The authors should provide sequences of sgRNAs (e.g. sgTRANS, sgCDH2, sgPPP1R1B, sg-mTOR, sgCREB53, sgTGFBR1).

Reviewer #3 (Comments to the Authors (Required)):

The authors present AI-PS, a method for performing pooled optical screens in cultured cell lines. Pooled optical screening is an important application, and advances in this area would be of broad impact. Like other similar approaches, AI-PS combines automated microscopy with a photoactivatable fluorescent protein to allow the marking of cells with a desired phenotype. The manuscript is extremely vague, making evaluation of the work difficult. However, based on what I was able to glean, AI-PS appears to be significantly worse than these previous approaches.

Important controls are missing and critical analyses are incorrect. Moreover, no fair attempt is made to compare AI-PS to previous photoactivatable protein-based approaches or to other methods for conducting arrayed or pooled optical screens Thus, my enthusiasm for this manuscript is low and I strongly suggest that it not be published in its current form.

General comments

-The authors completely ignore over a decade of prior work by many groups using photactivatable fluorescent proteins to enable the recovery of cells of interest. Either they were unaware of this previous work or, cynically, chose to ignore it. In particular, this work is similar to a manuscript published recently (PMID 32500953). The authors should cite and fairly discuss previous related work, both in the introduction when mentioning approaches for pooled screens, and later when discussing the strengths and weaknesses of their approach.

-Unfortunately, I do not believe that the data presented substantiate either the performance claims made about the method or the results of the screens performed. In particular, evaluation of the performance both of the cell classification algorithms and the fidelity of cell photoactivation and sorting is lacking.

-The lack of separation of the photoactivated populations from the unactivated populations in Figure 3b and S3b suggests that the method does not work very well. Other photoactivation-based approaches have done much better in this regard. Obviously, the authors are not going to repeat everything they've done, but their approach is clearly worse than other similar approaches.

-The description of the experiments and analyses is oftentimes confusing and occasionally impenetrable. See below for many examples.

-Two key statistical analyses, of the power of AI-PS and of the GFP-TFEB screen, appear to be incorrect.

-The figures are confusing, with panels occurring out of order and some missing citations in the text.

-Given that the main point is to present a method for pooled optical screening, the manuscript should contain a careful comparison of AI-PS to other methods for pooled optical screening. The lack of this comparison must be remedied.

Specific comments

Page 3/Figure 1 - The organization of Figure 1 is extremely confusing, reflected by the fact that the first panel called out in the text is Figure 1e. The authors should reorganize this figure to match the flow of the text.

Page 3/Figure 1 - CCCP should be defined.

Pages 3, 4 - General readability would be improved if the Parkin experimental system was described at least a little in the text (e.g. what do the drugs do, etc).

Page 4 - The description of the SVM is inadequate both here and in the methods. More information

regarding the details should be given. Importantly, the performance of the model should be more thoroughly described, especially in terms of biological replicate performance (image classification algorithms are notoriously challenged by batch effects).

Page 4 - Related to the above comment, it is not at all clear how cross-validation and test sets were handled. In particular, it is not clear whether final performance was evaluated using a test set never used in model training.

Page 4 - Clearly, there was feature selection but no details are given here or in the methods about how feature selection was performed. These must be added.

Page 4 - The authors do not present sufficient detail regarding the fidelity of AI-PS. For example, given the photoactivation conditions chosen, how many photoactivated (positive) cells are recovered? And, among recovered cells, how many are false positives? How do these quantities change as positive cells become rarer in the screened population?

Page 4 - Related to the previous comment, there appears to be some sort of validation experiment shown in SF 2c. But, this experiment is not referenced in the text and it is not described in sufficient detail in the figure legend to ascertain what was done. Presumably, it is some sort of demixing experiment where positive (PINK1+) cells are marked in some way and then sorted out? If so, that's great and should be explained, discussed. However, details like the number of replicates, etc, are missing.

Page 4 - It is great that the authors built a Shiny GUI for their SVM. But, Figure 2 seems like a waste for most readers. Critical methodological and performance details (see previous q's) could be answered using main figure space instead.

Page 4 - "Extended data figure" I think this should be Figure S2?

Page 5 - "The most abundant sgRNAs identified in the photoactivated samples were targeted against PINK1…" This should be "The most enriched…" because a fold change plot is shown and, if I understand correctly, the authors are looking at sorted/initial sgRNA frequencies.

Page 5 - I don't fully understand the power calculation that was performed and shown in SF2d. The methods description needs clarifying. I puzzled over it for a while and honestly can't figure it out. But, I'm pretty sure its the wrong thing to do, because the authors are using a t-test based metric to compare two samples (I am guessing it's control and PINK1/positive read counts in the sorted and naive populations across their biological replicates). Of course, in a real screen many sgRNAs are compared in each replicate. Much more complicated math is typically used to analyze such data (e.g. the negative binomial-based models used in EdgeR, which the authors use/cite) which incorporate per-sgRNA variance, per-gene variance, replicate level variance and false discovery rate control. It seems to me that any meaningful power analysis would have to do the same. Nonetheless, it is an important question and the authors should either more fully describe and defend what they did or fix it.

Page 5 - Related to the previous comment, "…sample size estimation indicated that three biological repeats are sufficient for detecting the desired genetic link in our experimental setup…" This is extremely confusing. The authors claim they detected the PINK1 "genetic link" and indeed that is what is shown in Figure 3c. But here they say they would need 3 replicates to detect it? What is meant by this statement?

Page 5/Figure 3c - Where did the log2 fold change threshold come from. If arbitrary, the authors should say that.

Page 5 - The authors show that their SVM approach performs poorly for TFEB nuclear translocation. But, more information is needed. Presumably they went through the same feature selection process as for the Parkin screen? All the same questions as raised above apply regarding what was done, how the SVM was trained, etc.

Pages 5, 6 - How did the CNN perform once trained. A model testing set of ~5,000 images is mentioned in the methods (which is great!) but the performance is not shown on the test set.

Page 6 and SF3c - Related to the previous comment "TFEB-GFP phenotype classification performance by SVM. Precision-Recall Curve from ~5,000 single cell images obtained from starved cells. Image collection began 8 hours after starvation initiated and continued for another 10 hours. The accuracy was computed from the integral area under the Precision-Recall Curve (AUC, area under the curve). The AUC was calculated per subpooled library (designated by color), from a pool of 3 biological repeats." This makes no sense. A PR AUC from a test set (I guess I learned that their test set was derived from starved cells in this SF legend quote - it should be in the main text and methods, see previous comment) makes sense. But then "The AUC was calculated per subpooled library…" makes no sense. The test images were collected and the model evaluated on them. I don't see how (additional) images from the library could be used to evaluate CNN performance.

Page 6 - " The entire photoactivated and sorted gene abundance ranking list…" The authors should say a little bit in the main text about how they combined the replicates and scored each gene.

Page 6 - The authors should give a sense of how many cells were sorted in each sublibrary/replicate. In fact, a supplementary table is needed summarizing each replicate of both screens in terms of number of cells sorted, reads acquired, etc.

Pages 6, 23 - From page 6 "A second validation screen was conducted of the 64 enriched genes using two new sgRNAs." From page 23 "For the secondary validation, the best two sgRNA with FC higher than two standard deviations from the non-targeting-sgRNA controls and roast test FDR < 15%." Even though the second sentence is a fragment, I think it contradicts the first and means that the two best sgRNAs for each significant hit were chosen, and not "two new sgRNAs" targeting hits, as the first sentence implies. This is a good example of the major readability problems plaguing this paper, and must be clarified.

Page 6 - The authors claim to have validated 21 of 64 hits, but appear to not have corrected for the 64 ANOVA tests they performed.

Page 7 - "The speed of AI-PS screening relies on the simultaneous execution of four steps: image capture, segmentation, generation of classification region of interest, and photoactivation of the region of interest." In fact, these steps occur sequentially, not simultaneously.

Page 8 - The authors state " We validated this by identifying PINK1 as the only significant hit required for Parkin translocation to damaged mitochondria within the genome guide sub-library of kinases, phosphatases and the druggable genome, demonstrating an exceptional signal-to-noise ratio when using the method." Is there strong evidence that PINK1 is the only hit that should be found? The authors should cite and discuss such evidence. Also, I strongly disagree that this one

example somehow validates the "exceptional signal-to-noise ratio" of the method, since the TFEB screen (generously) had a false positive rate of something like 39/64 or about 60%.

Pages 8, 9 - The discussion should contain a clear, fair comparison of the performance of AI-PS to arrayed screening methods in terms of cost, time, accuracy and the like. The most important drawback of AI-PS ignoring all the technical issues I raised, is that the phenotype must be pre-selected to allow for model training. Arrayed methods (and also in-situ sequencing) do not suffer from this limitation, which must be mentioned.

Page 25 - In the validation screen, the actual metric used to quantify nuclear GFP-TFEB is not clearly stated.

Page 17 - "For optimization of the model, we performed iterations and calculated performance by area under the receiver operating characteristic (ROC) curve or precision-recall curve (in the case of asymmetric phenotype representation). The performance values were plotted against iteration to prevent data overfitting." This is inadequate. Where are these ROC/PR curves? How many iterations (presumably of X-fold cross-validation)?

Page 17 - "A training set composed of 107,226 single-cell example images of GFP-TFEB in the nucleus or cytosol was produced." Based on the SF3 legend, I think they collected starved and non-starved cell images and used these as labels for training. If so, this is probably not a great idea. In particular, I am virtually certain there is variance over GFP-TFEB translocation into the nucleus in the starved condition (and also probably variance over GFP-TFEB cytoplasmic localization in non-starved conditions). At a minimum, the authors should go back to their imaging data and score, for a non-trivial number of cells, the degree of TFEB nuclear localization in each condition using any number of parametric methods (e.g. average pixel intensity in nucleus vs cytoplasm, etc). Showing this would at least give the reader a sense of how good we could expect the CNN, as trained, to be. A better way to do things would have been to manually curate GFP-TFEB nuclear and GFP-TFEB cytoplasmic cells and use transfer learning to re-train an existing CNN.

Pages 17, 18 - "Model performance" The description here is inadequate. Was this done for both types of models or just the CNN? How were images collapsed into single cells?

Page 19 - " To handle outlier cells, several features were computed and the outlier features were removed." This is insufficiently detailed. Which, and what was the decision boundary for feature removal?

Figure 1c - The panel as shown is inadequate. Details like the number of cells tracked, replicates performed, etc are needed. Also, presumably, error estimates could be generated per timepoint (since at least many cells must have been tracked).

Figure 1d - It is hard to appreciate that the upper left quadrant mitochondria are green (or at least they should be green because that is where GFP is in those cells).

Figure 3a - Same comment as above - hard to see mito+GFP.

Figure 3b - I can't quite figure out what is being shown here, but I think it is library cells following photoactivation. If that is so, an unactivated control population should be shown. Troublingly, the activated population (presumably the ones in red) are not distinctly separated from the unactivated population (presumably blue). If this is so, it is a major problem and significantly worse than other

similar methods. Also, all flow plots should include a description of the number of cells sorted in total.

Figure 4a - Not cited in text.

Figure S1 - A table explaining the full names/meaning of the features in S1a should be included.

Figures S3/4 - These are cited out of order.

DEPARTMENT OF HEALTH & HUMAN SERVICES   **National Institutes of Health
National Institute of Neurological
Disorders and Stroke**

Richard J. Youle
35 Convent Drive, 2C917
Bethesda, Maryland 20892
Phone: (301) 496-6628
E-mail: youle@helix.nih.gov

October 14th, 2020

Dear Jodi and Tim,

On behalf of all authors, we would like to thank you and reviewers for sharing the time and expertise in reading and providing valuable comments on our manuscript "Image-based pooled whole genome CRISPR screening for intracellular phenotypes – Parkin and TFEB subcellular localization".

The comments are very helpful and addressing them greatly improves our manuscript. Reviewers one and two agree that the method is novel and interesting. One major complaint by reviewer three is the lack of comparison of our method to existing similar methods. We add to the discussion a comparison with several other papers that somewhat overlap with ours including (Hasle et al., 2020; Feldman et al., 2018; Wheeler et al., 2020).

A consensus of all the reviewers is that the manuscript needs to be substantially revised and rewritten with more information added to the main text from the methods section. The reviewers indicate that important technical information is missing. More specifically, there are three subject areas we are requested to elaborate on: 1. Tool performance, 2. Image analysis and processing 3. Statistical methods. The revised manuscript answers all the reviewer comments, including reanalyzing of the data, changing data presentation and further validated one hit derived from our whole genome screen. In the following rebuttal, comments to the reviewers and editors are labeled in Orange, direct citations from the revised manuscript are in red and the location of the changes with page and line numbers are in blue.

With appreciation,

Gil Kanfer and Richard Youle

Reviewer #1 (Comments to the Authors (Required)):

The manuscript by Kanfer, Youle and coworkers is built on the idea to implement a CRISPRi-based screen by using imaging of adherent cells that express a fluorescent reporter - which allows them to identify specific changes in reporter localization - and then select individual cells with a particular reporter signal by photoactivation for subsequent FACS sorting. While the idea of using imaging for feature selection for CRISPR screening is not totally new (for example Goda's group, 2018, Cell 175:266-276 used this for cells in solution), I think this is the first

report of using imaging of adherent cells for CRISPR screening and also to make use of a photoactivation step for cell selection in this format. Given the large number of available fluorescent reporters based on localization changes in cells, there are many potential applications for this type of screening pipeline. In addition to the novelty of using adherent cells and selection by photoactivation in a CRISPR screen, the main novelty of their approach is the use of machine learning and convolutional neural network models for image analysis that allow them select which cells to mark by photoactivation. To demonstrate the validity of their approach, they show the use of the pipeline for two types of screens. The second is a genome-wide screen on TFEB localization that identifies potentially new regulators of the nuclear localization of TFEB which is a critical transcriptional feedback mechanisms in the mTORC1 signal transduction pathway.

Given that the use of machine learning and image analysis in a CRISPR screen is not totally new, I expected more mechanistic data demonstrating that the pipeline and approach they developed can lead to finding a mechanistically relevant new regulator of the mTORC1/TFEB signaling pathway. This is my only major point that requires that they provide some additional mechanistic data on at least one of their hits. The other minor points I am adding below can likely be addressed by changes to the representation of data and analysis.

Major point:

1) Some more functional characterization is needed for at least one of their hits from the CRISPRi screen. For example, they could focus on the phosphatase inhibitor PP1R1B (DARP32) for which there is a plausible hypothesis of how it may work in the mTORC1 pathway. This could strengthen their argument that this type of screen is useful to find new regulators of nuclear translocation and ultimately understand molecular mechanisms. They could for example ask whether PP1R1B acts directly in the mTORC1 pathway by directly or indirectly regulating TFEB phosphorylation (which is, as I understand, the main mechanism of TFEB nuclear localization). An effect on the rate of TFEB dephosphorylation could for example be measured by using acute inhibition of protein phosphatases (using ocadeic acid or similar compounds; or alternatively a mTORC1 inhibitor such as rapamycin) and then monitoring the nuclear localization of GFP-TFEB. A related test is to measure whether mTORC1 activity for the sgRNA targeting PP1R1B is different by measuring one or more of the known substrates such as S6K or 4E-BP.

We further investigated the TFEB redistribution regulator CREB5 because it was our top hit with the strongest effect. We found that the knock down (KD) of CREB5 reduced the expression of LC3B and two lysosomal genes. Further validating CREB5 activity, rescuing CREB5 in CREB5 KD cells rescued TFEB cytosolic redistribution and the expression of the TFEB target genes. A paragraph has been added to the result section, titled: "**TFEB nuclear translocation is regulated by CREB5"** page 11, line 312. Fig 9, b and c

Minor points:

1. Why did SVM classification fail to predict TFEB nuclear localization accurately - TFEB localization seems to be a clear difference in distribution that should be readily captured by an SVM algorithm. More detailed is needed why a SVM classification dies not work in this case.

To address this issue, we added the following text to the results and discussion sections.

Added to results: page 8, line 235

Overall, it is not clear why the CNN model shows higher performance than the SVM. Our speculation is that the difference is most likely that uneven illumination of image examples introducing in the training set influenced the segmentation step of our CNN model,

Added to discussion: page 14, line 418

"Prediction of TFEB nuclear translocation by the deep learning approach was more accurate than the SVM classification model, possibly because of discrepancies in classification accuracy owing to uneven fluorescence levels of the TFEB signal. Although the cell line was carefully generated from a single clone, over several passages the TFEB expression level diverged across the population. The use of low magnification objective (20X) with a low numerical aperture value of 0.75 further amplified these variations. To address this, uneven illuminated images were introduced in our CNN classifier builder by adding an augmentation step to our image batch generator prior to training. In future screening designs there are several steps that can be used to overcome this issue. First, knocking in GFP into the TFEB or gene of interest locus may decrease expression variability. In addition, higher magnification objectives equipped with better NA lenses would decrease the illumination heterogenicity."

2a. A more detailed analysis should be provided how the three subpool screen in the TFEB screen differ from each other. This would allow one to better understand how reproducible the screen is.

To assess the variation between the triplicate read counts of each subpool, we plotted the coefficient of variation between the triplicate screens against the log2-CPM normalized mean count per sgRNA. From this analysis we conclude that the overall in group variation between the triplicate screens is minimal. Every subpooled library contains 500 non-targeting gRNAs. The distribution of these gRNAs and the number of detectable gRNA per subpooled library also support a minimal variation. However, there is considerable variation between the different guide subpool samples comparing photoactivated and unactivated cells. The between subpool sgRNA's variation is reflected in our abundance analysis since in one pool, the membrane protein pool (new Fig. 6b) related genes were highly enriched in our gene-set analysis indicating a higher false positive rate and higher false negative rate than, for example, subpool H3 or H4. Hence, we cannot exclude that some hits were missed in our analysis. This also explains the relatively high false positive hit rate, which falls away in our secondary validation. The coefficient of variation has been added to results: page 9, line 243 and Fig. 6 b i-vii

2b) In addition, they should add a discussion how many cells and subpool screens are needed to reach saturation using this method. For example, would a doubling of the number of screens or cells still significantly increase accuracy?

This point is similar to that of reviewer #3, point 4 and we thank the reviewers for pointing out this issue. Indeed, our choice of power analysis strategy was not well thought out and we should have considered the parameters the reviewers mention. To fix the power analysis issue we used a simulation strategy. We used the R based package PROPER to estimate sample size based on a negative binomial model which includes dispersion distribution (Wu et al., 2015). We added the statistical power of analysis to the results section on page 7, line 188 and Fig 3, e

2c) Page 5/Figure 3c - Where did the log2 fold change threshold come from. If arbitrary, the authors should say that.

3

We have added to the results the following "The $\log_2$-fold change was modeled based on the non-targeting negative control distribution." We also added a histogram plot showing non-targeting negative control distribution. Please see:
page 7, line 180 and Fig 3, c

3. The imaging step with one second and an analysis of 5 seconds are both quite long. Is there a way to accelerate this? A discussion should be added how a faster microscope and faster analysis can accelerate these steps.

In the discussion we have added the following text to explain a deep learning based object segmentation approach which could lead to five to ten fold computational time reduction for the segmentation step.
See page 15, line 432

We also elaborate on how a faster CMOS camera and larger FOV capabilities also will improve our tool in the discussion.
page 15, line 440
"in the future, simultaneous imaging with two CEMOS cameras will reduce capture time. And finally, large-format camera sensors with larger FOV capturing will greatly improve the overall screen since more cells can be screened and analyzed. "

4. They also need to discuss the limitation of photoactivation in more detail. Does it take for each cell 3 sec of photoactivation - would it be 30 seconds if they would select 10 cells in the image? Could they instead use a stronger laser or different fluorescent protein to make the photoactivation step much shorter? A discussion is needed whether/how the time to image, analyze and photoactivate could be at least 10 times faster to make the method more practically useful considering the long-time it took them to complete the subpool screens.

We apologize for the lack of clarity and now understand how Fig. 1g was misleading. We improved the wording in the results as follows.
page 5, line 136
"The selected cells were photoswitched by illumination of 50 msec/pixel dwell time with 80% UV laser intensity. This parameter was chosen so as to reduce the photoactivation time, eliminate unwanted activation of adjacent cells and maximize signal intensity."
In the discussion we add: page 15, line 452
"The tool we present here is best suited for low phenotype alteration hit rates, i.e., when only 0.5% to 1% of the cells are called per field of view captured to minimize photoactivation time. For example, in the current TFEB screen, a mean of three cells were detected and activated per field of view. Therefore, for the current screen a galvo-miniscanner photoactivation unit was sufficient, however, in a scenario where the phenotype altering hit rate is much higher, a faster photoactivation unit such as DMD illumination module would be more suitable."

5. The entire Fig 2 is more of a methodological detail and should be moved into the suppl. materials section.

We moved Fig. 2 to the Suppl. Materials as suggested.

6. More precise experimental details are needed in all Fig legends (or in the main text describing

the figures). This will allow the reader to better understand what the actual experiments were and how the data was processed without having to consult the methods section.

We have extended and elaborated all figure legends and the main text.

7. Fig 6 lacks a control sgRNA. Also, fewer time points would make the same point in the figure (a focus on the time points where the difference in localization is greatest relative to control would be best).

Figure 9a was modified according to the reviewer suggestions.

Reviewer #2 (Comments to the Authors (Required)):

The authors present a novel pooled screening approach based on CRISPRi and machine learning. After reading the abstract of this paper I was very excited to read the whole paper as the premise, using machine learning to identify 'hits' in relocalization screens coupled with fluorescence activation and cell sorting to retrieve them, I think is very clever and thorough validation would be a tremendous advance for the community.

This "Tool" article is divided into two main parts. Firstly, a proof-of-concept experiment by using the known Parkin recruitment to mitochondria by PINK1 and secondly, a genome-wide screen identifying new effectors of TFEB shuttling.

In both studies machine learning is used to rapidly identify candidate hits for fluorescence activation. A relatively straightforward SVM classifier was used for the proof-of-concept study and a more sophisticated deep learning approach was used for the TFEB screen.

As a "Tool" article it must be relatively straight forward for the readership of JCB to follow the implementation of the method being described. However, I found the manuscript extremely difficult to follow and I work in the area. Furthermore, this kind of tool article should clearly demonstrate the capabilities and limitations of the "tool". Again, I found this information very difficult to find in the text and even more difficult to fully appreciate.

1) For the 'proof-of-principle' screen "The Draq5 channel is used for nuclear detection and the GFP-Parkin image is used to identify cell borders" Yet the first mention of Draq5 is in the methods section that follows the main manuscript and the quote above is from the caption for a supplementary figure. Even after reading the whole manuscript and looking for the details - it is not clear to me how individual cells were identified and segmented.

We add the Draq5 images to fig 1f. We also add the following to the discussion.
page 12, line 341.
"The effective segmentation of live cells is critical in order to ensure efficiency in training and avoid erroneous predictions. We found that the best way to segment mammalian cells utilizing the R package, EBImage, was to use two cellular markers in two different channels. Draq5 was used to mark the nuclei, which provided the "seeds" for segmentation. The other marker provides the cellular "borders, or cytosolic volume of the cell. The latter is important for the effective segmentation at higher confluency of cells, which maximizes the number of cells screened. Similar two-channel-approaches are commonly used in cellular segmentation (Wählby et al.,

5

2002; Quelhas et al., 2010; Al-Kofahi et al., 2018). Three of the four most commonly differentiated fluorescence channels are used for nuclei detection (far red), photo-activation (red), and CRISPRi guide RNA expression (blue). Thus, AI-PS utilizes the remaining green / GFP channel to visualize both the phenotype queried and the cell borders. Deep learning models are becoming a more popular tool, but any gain in accuracy they provide is countered by the computational power and time required to deploy such models during the AI-PS segmentation step."

2) Looking at the figures, the segmentation is not very good which is surprising because it is usually quite facile to segment cells using only the Draq5 channel and here the authors used two channels.

We apologize for our poor selection of images showing the segmentation step we used. We present more representative examples of our Image segmentation procedure and we extend the description by adding the following text and data to the results section.
page 5, line 129 and Fig. S1c
"The accuracy of the segmentation procedure was compared to the gold standard manual segmentation using the NIS-elements imaging software. The segmentation was evaluated by calculating the intersection over union (IOU). Comparing the IOU of the current segmentation procedure to CellProfiler showed very similar segmentation scores (Fig S1c). "

3) Figure 2a supposedly outlines the steps in segmentation but starts with an image of mitochondria which was clearly not used to create the mask in the next image. This second image is the individual cells and is used for the coordinates for photoactivation. I am not clear what the third image is or what it is for. Below they show the resulting segmentation in which there are clearly multiple instances of single segmentation boundaries surrounding multiple cells. I don't think a Draq5 image is shown anywhere.

The reviewer is correct, the screen snapshot from our Shiny APP was misleading. We are very sorry for that and added better examples of the Shiny-APP segmentation procedure. See Fig. S2 for a better version of the APP.
As stating in Point #1 Draq5 images were added to Fig. 1 f.

4) When using the Halo-tag in the second screen the authors again do not show any halo-tag images. And the image that they do show includes "a cell" complete with nucleus and cytoplasm (olive/lime green nucleus mask near the center of figure 4f) where there is no evidence of a cell at all.

Please see the corrected image in Fig 5b. It was an image contrast problem.

5) The authors show convincingly the proof of concept by detecting PINK1 as most significant hit in the Parkin translocation screen. However, assuming that I followed the manuscript and understand Figure 3c this conclusion is overstated. It was actually because I could not understand the location of the PINK1 dot in the figure that I started looking carefully in the manuscript. It appears that PINK1 guides were doped into the library at 10%. It is not at all surprising that they were able to retrieve positive control guides that constituted 10% of the total and makes the claims of sensitivity for the screen highly overstated. To claim that three replicates are needed based on this data is very suspect. Looking at the rest of the data in 3b, I see no clear delineation of potential hits - which dot is PINK1?

Please refer also to a similar issue with Reviewer #3 Point 3. We failed to explain this well in the text so we entirely rewrote this section. To be clear, we retrieved PINK1 guides from a guide library that was not doped with PINK1 guides. A paragraph was added to the result section, titled: "**Photoactivation accuracy and performance**" page 11, line 148. See fig. 2 a, for the method illustration. In Fig 3 d, we improved the volcano plot visibility.

6) Feature selection is a crucial part for a successful classification. It would be of interest what features are selected in the SVM classifier. How much do the features vary? Please explain how the features were reduced. I think that the idea is that highly correlated features were dropped - which is a standard procedure but not clear here.

We now explain the feature selection in building the SVM classification model. We moved the PCA plot from the supplement (Fig. S1a) to the main figures. We added a new Table to summarize the five-features selected. The following text has been added to the result section: page 4, line 109. and Fig. 1 d,e.

"To build the SVM classification model, 19 features were computed from 2500 single cell images of cytosolic or mitochondrial GFP-Parkin (Fig. 1d). The features were computed using the R image processing and analysis package, EBImage (Pau, 2009). To prevent classifier over-fitting and reduce the computational cost, five cellular features measuring the 5% intensity quantile, standard deviation of intensity, minimum radius, eccentricity and area, that showed distinct variation were selected (Fig 1e). The selected features and labelled cell images were computationally applied on a nonlinear SVM algorithm for creating the classification model (Fig. S1 a)."

7) Actually, I fail to see the complexity in either screen and the need for machine learning. In my experience mitochondrial localization as opposed to cytoplasmic localization can be readily scored by simply measuring the variance across the cytoplasmic area of the cell. Similarly, the training images shown for the nuclear cytoplasmic localization determination - probably the single most common high content screen performed by the community - are sufficiently clear that the authors should not have needed to use deep learning to score the cells.

We agree that both of these phenotypes likely could be properly analysed and detected with standard image analysis techniques. However, we present here an alternative way of preforming "on the fly" analysis. We provide the codes and user friendly open source program for the community so future phenotypic genetic screen for more subtle phenotypes will be feasible.

8) Furthermore, they present in the TFEB screen 64 genes that cause a retention of TFEB within the nucleus. In a second validation screen 21 out of 64 were validated. But beyond that the phenomenon could be reproduced, the validation is not comprehensive. Maybe I am missing something but I don't follow the arguments at the top of page 7 suggesting the results validate the screen. In addition to standard validation approaches why don't the authors measure changes in TFEB regulated gene expression? Some more functional characterization is needed for at least one of their hits from the CRISPRi screen.

This point is similar to that of reviewer #1 major point 1. Please refer there for how we further assessed CREB5 activity.

9) While likely the problem is in the presentation of the data, I found the precision recall curves difficult to understand for the TFEB screen. How is this being calculated? What is the gold standard? If you don't know the answer how is it possible to calculate a precision-recall curve?

We moved the explanation of the test from the Methods section to the Results section and also provide the code we built to address it. We understand that the Machine learning performance evaluation was not clear and caused confusion. We add the following to the main text:
page 8, line 217.
"The models were trained using two datasets, one for each phenotypic classification, that was made up of image files of individual cells. The single-cell images were generated using the R-based segmentation script deployed by AI-PS and manually classified."
In addition, we also added the following details to the Methods section.
page 25, line 720.
"To train the CNN model, the files of each dataset were split into three groups: training (80%), validation (15%), and testing (5%). The validation set was used during model development to evaluate the model's performance during training and tuning classification hyperparameters. Validation accuracy was important for detecting model overfitting. After training, the model was then evaluated using the testing image set. The validation and testing designated images were never used during training, allowing for the assessment of a model's generalizability. Both SVM and CNN models were evaluated for their performance on the testing dataset; ability to produce prediction values matching the cell image's "true" class label."

10) Overall - as attractive as the concept presented in the manuscript is, the manuscript needs a full renovation if it is to have impact in the community.

We apologize for the writing and extensively revamped the text. All the answers to reviewer comments have been now integrated in the text and more tool performance and validation data are added as mentioned to individual reviewer comments.

Minor points:

For the authors consideration.

11) Why is there a probability value discrepancy in Fig. 6 a (top panel, 1h, sgTGFBR1). Two nuclei side by side with similar intensity but a notable difference in the probability values. Given the problems with the segmentation alluded to above and no segmentation data shown here it is hard for me to interpret the numbers provided.
The reviewer is correct, the probability value difference is most likely because of the segmentation. We add the following explanation to the discussion.
page 15, line 443.
"Another limitation of AI-PS is that to complete a whole genome screen we image 600,000 cell batches in three repeats. To minimize the overall screening time we reduce the number of fields of view to be screened by seeding cells at 90% confluency. The high confluency allows more cells to be screened, but results in slight reduction in the accuracy of segmentation. Therefore, in the future, to improve segmentation by allowing lower cell seeding density one could compensate by increasing screen image acquisition times. Screening of fixed cells with a reversible fixation method to allow cell sorting following photoactivation would help extend screen image acquisition time."

8

12) Both dCas9 (pC13N-dCas9-BFP-KRAB) and sgRNAs (mU6-BstXI-BlpI-BFP vector) are tagged with BFP. It would be helpful if the authors briefly describe (in addition to the citation "Tian et al., 2019") in the method section how only the sgRNA expression is assessed by the BFP signal.

We thank the reviewer for this comment. As both dCAS9 and the sgRNA vectors are BFP labeled, we noticed that the best KD effects are observed in single cells in which the dCas9 BFP signal is the dimmer. We add the following explanation in the methods section:
"Cells were single-cloned and selected for dim dCas9 BFP signal that yielded the largest knockdown effects"

13) Fig 1 g and Fig. 2 f: typo "field of view"
We appreciate this and the following nine corrections. We have corrected all the following as requested up to point 23 below.

14) Figure caption 3: typo "fluorescence"

15) Page 5 wrong citation formatting

16) Throughout the text the authors write 12,500 sgRNAs, however on page 4 they mention 12,775 guides.

17) For better visualization it is recommended to rethink the font color, e.g. in Fig S2b. Some yellow digits are hard to read, especially on a white background. Also in Fig S4, the red digits are not easy to read. The red circle is a bit too thin and might be overlooked.

18) Fig. S 2 b, it would be helpful if the authors could label the Pa mCh channel like they did in Fig. 4 f with "pre-activated" and "activated".

19) Fig. 4 b, TC medium was used in the figure, however the authors used complete medium (cm) in the figure caption.

20) Could you specify (page 17, "initially trained 2,234 images") if images means individual cells.

21) Please be consistent in writing PINK1.

22) On page 23, please could the authors write out FC as fold change once and then abbreviate it.

23) In Fig. S1 c, the authors show an immunofluorescence experiment on CDH2 and TRANS to assess potent dCas9 clones. However, in the method section this experiment is not described.

We have further described the dCas9 clone test. The description below was added to the method section page 23 line 658.
"The U2OS-dCas9-KRAB cell line was then subcloned and the dCas9-KRAB activity assessed to select the most potent clones for further use by live plasma membrane immunostaining. In brief, dCas9-KRAB U2OS clones were induced with lentivirus expressing gRNA targeting Transferrin receptor or N-Cadherin. Follow four days of induction, cells were seeded on an

24) The authors should provide sequences of sgRNAs (e.g. sgTRANS, sgCDH2, sgPPP1R1B, sg-mTOR, sgCREB53, sgTGFBR1).

In Table S2, second tab, we added asterisks to the guide sequences used for the secondary screen.

Reviewer #3 (Comments to the Authors (Required)):

The authors present AI-PS, a method for performing pooled optical screens in cultured cell lines. Pooled optical screening is an important application, and advances in this area would be of broad impact. Like other similar approaches, AI-PS combines automated microscopy with a photoactivatable fluorescent protein to allow the marking of cells with a desired phenotype. The manuscript is extremely vague, making evaluation of the work difficult. However, based on what I was able to glean, AI-PS appears to be significantly worse than these previous approaches. Important controls are missing and critical analyses are incorrect. Moreover, no fair attempt is made to compare AI-PS to previous photoactivatable protein-based approaches or to other methods for conducting arrayed or pooled optical screens Thus, my enthusiasm for this manuscript is low and I strongly suggest that it not be published in its current form.

General comments

1) -The authors completely ignore over a decade of prior work by many groups using photoactivatable fluorescent proteins to enable the recovery of cells of interest. Either they were unaware of this previous work or, cynically, chose to ignore it. In particular, this work is similar to a manuscript published recently (PMID 32500953). The authors should cite and fairly discuss previous related work, both in the introduction when mentioning approaches for pooled screens, and later when discussing the strengths and weaknesses of their approach.

We apologize for the incomplete referencing – it was unintentional. We had focused our literature searches on image based genetic screening methods not on the uses of photoactivatable FP selection methods related to ours. We revised the Introduction accordingly and add mention of the Fowler paper (Hasle et al., 2020) as suggested. We also cite an early paper using photoactivation to select denoted cells from tissues (Victora et al., 2010). We appreciate the importance of this criticism and if there are other papers we are unaware of, please tell us.

2) -Unfortunately, I do not believe that the data presented substantiate either the performance claims made about the method or the results of the screens performed. In particular, evaluation of the performance both of the cell classification algorithms and the fidelity of cell photoactivation and sorting is lacking.

Please see our answers to specific issues below.

3) -The lack of separation of the photoactivated populations from the unactivated populations in Figure 3b and S3b suggests that the method does not work very well. Other photoactivation-based approaches have done much better in this regard. Obviously, the authors are not going to

repeat everything they've done, but their approach is clearly worse than other similar approaches.

This relates to reviewer #2, point 5. The performance of the phenotype classification and sorting of photoactivated cells were reevaluated separately to address this issue. For testing the cell separation performance, we evaluated the sorting accuracy of the detected and photoactivated cells from the entire population by experimentally mixing cells blocked for Parkin recruitment by PINK1 KD with wildtype cells. A full paragraph is added to the result section, titled: "**Photoactivation accuracy and performance**" page 11, line 148. See Fig. 2.

4) -Two key statistical analyses, of the power of AI-PS and of the GFP-TFEB screen, appear to be incorrect.

Thank you, this is the same issue raised by reviewer #1. Please see that explanation and the new data included to address this under comments to Reviewer #1, point 2b.

5) -The figures are confusing, with panels occurring out of order and some missing citations in the text.
We apologize - we rearranged the text and Figure panels to follow in order.

6) -Given that the main point is to present a method for pooled optical screening, the manuscript should contain a careful comparison of AI-PS to other methods for pooled optical screening. The lack of this comparison must be remedied.
We added to the introduction:
Page 2 line 55
        The usage of image based pooled genetic screens linking phenotypes to genotypes was previously reported in three independent studies in which in-situ barcoded sequencing was coupled to phenotypes. First this approach was used to identify photostable and brighter variants of a fluorescent protein, by testing 60,000 mutation variants (Emanuel et al., 2017). Then an in-situ platform was integrated with Crispr genetic screens for identify genes involved in RNA nuclear localization, while another Crispr screen used in-situ sequencing imaging to identify factors associated with NFkB translocation regulation. These later two methods screened 162 CRISPR guides in Wang et al. (Wang et al., 2019) and 3063 guides in Feldman et al. (Feldman et al., 2018). More recently a semi arrayed 12,500 gRNA Crispr screen was used to identify regulators of stress granule formation (Wheeler et al., 2020). These methods enable the investigation of protein pathways regulating subcellular organization and positioning in an unbiased manner. In addition to unbiased CRISPR screens linking microscopic phenotypes to genotypes, single-cell images linking microscopic phenotypes to genotypes was established by a new method called single cell magneto-optical capture (Binan et al., 2019). Although these processes are elegant and will improve genetic studies, they are not well suited for high throughput large scale screens. Hence, we propose that a simple photoactivation of cells with desired phenotypes coupled to cell sorting will reduce image based screen complexity. Previously, B lymphocytes isolation and characterization were conducted from photoactivatable transgenic mice, by coupling photoactivation and flow cytometry (Victora et al., 2010). In addition, in a more recent study, photoactivation coupled to flow cytometry enabled the investigation of the link between the morphology response to a drug and the genetic profile at single cell resolution (Hasle et al., 2020).
To the discussion we add:
Page 14 line 403.

"The use of machine learning, photoconversion and deep sequencing in separate applications are not new. In the current AI-PS we show an improvement of scalability of pooled optical screens in comparison to similar approaches already reported. However, we show that in comparison to a previous pooled visual genetic screen (Feldman et al., 2018), only 32% of the primary screen hits were validated to directedly affect TFEB translocation in a secondary assay. We cannot rule out that this lower validation rate is a result of the large scale of the current screen which increases the complexity and might increase variation. In the future, in order to increase the discovery rate of large AI-PS screens, a few considerations are recommended. First, in the current study we observed that increasing the biological replicates from three to five, resulted in significant power increase. Second, as discussed previously, faster and larger imaging fields will allow greater screening sample size. In addition, from our Flow-cytometry data we learned that greater than 0.5% frequency of the desired cell phenotype decreases false positives."

7) Page 3/Figure 1 - The organization of Figure 1 is extremely confusing, reflected by the fact that the first panel called out in the text is Figure 1e. The authors should reorganize this figure to match the flow of the text.

The figures were reorganized according to the reviewers comments.

8) Page 3/Figure 1 - CCCP should be defined.

Thank you, we did this.

9) Pages 3, 4 - General readability would be improved if the Parkin experimental system was described at least a little in the text (e.g. what do the drugs do, etc).

We revised accordingly.

10) Page 4 - The description of the SVM is inadequate both here and in the methods. More information regarding the details should be given. Importantly, the performance of the model should be more thoroughly described, especially in terms of biological replicate performance (image classification algorithms are notoriously challenged by batch effects).

Similar to reviewer #1, point 2 – we add the requested information as described in detail above to reviewer #1. A more detailed description of SVM is included in the main text: Page 4 line 106 Fig. 1 d and e. In addition, we also included replicate variation analysis, Page 9 line 244 and Fig. 6a and b.

11) Page 4 - Related to the above comment, it is not at all clear how cross-validation and test sets were handled. In particular, it is not clear whether final performance was evaluated using a test set never used in model training.

We indeed assessed performance using a test set not used for training. Please see our detailed response above to the similar point made by reviewer #2, point 9.

12) Page 4 - Clearly, there was feature selection but no details are given here or in the methods about how feature selection was performed. These must be added.

13) Page 4 - The authors do not present sufficient detail regarding the fidelity of AI-PS. For example, given the photoactivation conditions chosen, how many photoactivated (positive) cells are recovered? And, among recovered cells, how many are false positives? How do these quantities change as positive cells become rarer in the screened population?

14) Page 4 - Related to the previous comment, there appears to be some sort of validation experiment shown in SF 2c. But, this experiment is not referenced in the text and it is not described in sufficient detail in the figure legend to ascertain what was done. Presumably, it is some sort of demixing experiment where positive (PINK1+) cells are marked in some way and then sorted out? If so, that's great and should be explained, discussed. However, details like the number of replicates, etc, are missing.

15) Page 4 - It is great that the authors built a Shiny GUI for their SVM. But, Figure 2 seems like a waste for most readers. Critical methodological and performance details (see previous q's) could be answered using main figure space instead.

16) Page 4 - "Extended data figure" I think this should be Figure S2?

17) Page 5 - "The most abundant sgRNAs identified in the photoactivated samples were targeted against PINK1..." This should be "The most enriched..." because a fold change plot is shown and, if I understand correctly, the authors are looking at sorted/initial sgRNA frequencies.

18) Page 5 - I don't fully understand the power calculation that was performed and shown in SF2d. The methods description needs clarifying. I puzzled over it for a while and honestly can't figure it out. But, I'm pretty sure its the wrong thing to do, because the authors are using a t-test based metric to compare two samples (I am guessing it's control and PINK1/positive read counts in the sorted and naive populations across their biological replicates). Of course, in a real screen many sgRNAs are compared in each replicate. Much more complicated math is typically used to analyze such data (e.g. the negative binomial-based models used in EdgeR, which the authors

use/cite) which incorporate per-sgRNA variance, per-gene variance, replicate level variance and false discovery rate control. It seems to me that any meaningful power analysis would have to do the same. Nonetheless, it is an important question and the authors should either more fully describe and defend what they did or fix it.

Reviewer #1 made the same point. Indeed, our choice of power analysis strategy was not correct. Please see out detailed comments on this issue to Reviewer #1, point #3. The revised statistical power estimation was added to the main text on page 7 line 188, and Fig. 3 e. Also see page 9 line 259, and Fig. 7 b.

19) Page 5 - Related to the previous comment, "...sample size estimation indicated that three biological repeats are sufficient for detecting the desired genetic link in our experimental setup..." This is extremely confusing. The authors claim they detected the PINK1 "genetic link" and indeed that is what is shown in Figure 3c. But here they say they would need 3 replicates to detect it? What is meant by this statement?

Please see our detailed comment on this issue to reviewer #1 point 3.
The statistical power simulation was conducted on the data acquired in the current study and shows that 3 repeats are sufficient. However, for future usage of the methods 5 repeats would result in greater detection performance. We also add to the discussion the following for clarification:
Page 14 line 411:
"First, in the current study we observed that increasing the biological replicates from three to five, resulted in significant power increase. Second, as discussed previously, faster and larger imaging fields will allow greater screening sample size. In addition, from our Flow-cytometry data we learned that greater than 0.5% frequency of the desired cell phenotype decreases false positives."

20) Page 5/Figure 3c - Where did the log2 fold change threshold come from. If arbitrary, the authors should say that.

Please see our comments on this issue to Reviewer #1 point 2c and new Figure panel.
The is now in the text on page 6, line 177 and in Fig 3c.

21) Page 5 - The authors show that their SVM approach performs poorly for TFEB nuclear translocation. But, more information is needed. Presumably they went through the same feature selection process as for the Parkin screen? All the same questions as raised above apply regarding what was done, how the SVM was trained, etc.

Thank you, we explained this issue above for reviewer #1, minor point 1.
see results: page 8, line 235.
discussion: page 14, line 418.

22) Pages 5, 6 – How did the CNN perform once trained. A model testing set of ~5,000 images is mentioned in the methods (which is great!) but the performance is not shown on the test set.

This is the same issue as Reviewer #2, point 9. We apologize and understand that our text was not clear and important information was missing. The performance analysis is always on the test set and not on the training set. We add the following information to the main text:

14

"The models were trained using two datasets, one for each phenotypic classification that was made up of image files of individual cells. The single-cell images were generated using the R-based segmentation script deployed by AI-PS and manually classified."
In addition, we will also added the following additional details to the method section.
"To train the CNN model, the files of each dataset were split into three groups: training (80%), validation (15%), and testing (5%). The validation set was used during model development to evaluate the model's performance during training and tuning classification hyperparameters. Validation accuracy was important for detecting model overfitting. After training, the model was then evaluated using the testing set. The validation and testing designated images were never used during training, allowing for the assessment of a model's generalizability. Both SVM and CNN models were evaluated for their performance on the testing dataset; ability to produce prediction values matching the cell image's "true" class label."

23) Page 6 and SF3c - Related to the previous comment "TFEB-GFP phenotype classification performance by SVM. Precision-Recall Curve from ~5,000 single cell images obtained from starved cells. Image collection began 8 hours after starvation initiated and continued for another 10 hours. The accuracy was computed from the integral area under the Precision-Recall Curve (AUC, area under the curve). The AUC was calculated per subpooled library (designated by color), from a pool of 3 biological repeats." This makes no sense. A PR AUC from a test set (I guess I learned that their test set was derived from starved cells in this SF legend quote - it should be in the main text and methods, see previous comment) makes sense. But then "The AUC was calculated per subpooled library..." makes no sense. The test images were collected and the model evaluated on them. I don't see how (additional) images from the library could be used to evaluate CNN performance.

The reviewer is absolutely correct, we now have rewritten the section and added it to the main text for clarity. It is similar to reviewer #2, point 9. In addition, we are sorry for this incorrect description and re-wrote the figure legend:
"GFP-TFEB phenotype classification performance by SVM, Precision-Recall curve from 7,848 single cell images obtained from HBSS starved and fed cells. For the starved cell population, image collection began 8 hours after starvation was initiated and continued for another 10 hours. Accuracy is computed from the integral area under the Precision-Recall Curve (AUC, Area Under the Curve)."

24) Page 6 - " The entire photoactivated and sorted gene abundance ranking list..." The authors should say a little bit in the main text about how they combined the replicates and scored each gene.

We add the following explanation to the main text:
"For calculating gene enrichment we subjected the sgRNA list to the rotation gene set test, provided by the R package EdgeR."

25) Page 6 - The authors should give a sense of how many cells were sorted in each sublibrary/replicate. In fact, a supplementary table is needed summarizing each replicate of both screens in terms of number of cells sorted, reads acquired, etc.

15

This is similar to Reviewer #1, point 2a. We added a Table as requested by this reviewer. See new Table S3 and comments to Rev. 1.
"Table S3. TFEB cell numbers and NGS read numbers. In the first tab, total number of cells sorted from the photoactivated samples per library per biological repeat. In the second tab, Total number of reads detected from the NGS analysis. "

26) Pages 6, 23 - From page 6 "A second validation screen was conducted of the 64 enriched genes using two new sgRNAs." From page 23 "For the secondary validation, the best two sgRNA with FC higher than two standard deviations from the non-targeting-sgRNA controls and roast test FDR < 15%." Even though the second sentence is a fragment, I think it contradicts the first and means that the two best sgRNAs for each significant hit were chosen, and not "two new sgRNAs" targeting hits, as the first sentence implies. This is a good example of the major readability problems plaguing this paper, and must be clarified.

We apologize for this mistake – we clarified it as follows.
Page 10, line 276
"A second validation screen was conducted of the 64 enriched genes using the top ranked primary screen identified sgRNAs."

27) Page 6 - The authors claim to have validated 21 of 64 hits, but appear to not have corrected for the 64 ANOVA tests they performed.

We added the sentence, "p-values were Benjamini-Hochberg corrected" on
Page 10 line 285

28) Page 7 - "The speed of AI-PS screening relies on the simultaneous execution of four steps: image capture, segmentation, generation of classification region of interest, and photoactivation of the region of interest." In fact, these steps occur sequentially, not simultaneously.

The reviewer is correct – we changed the wording to sequentially.

29) Page 8 - The authors state " We validated this by identifying PINK1 as the only significant hit required for Parkin translocation to damaged mitochondria within the genome guide sub-library of kinases, phosphatases and the druggable genome, demonstrating an exceptional signal-to-noise ratio when using the method." Is there strong evidence that PINK1 is the only hit that should be found? The authors should cite and discuss such evidence. Also, I strongly disagree that this one example somehow validates the "exceptional signal-to-noise ratio" of the method, since the TFEB screen (generously) had a false positive rate of something like 39/64 or about 60%.

We softened the claim with the following wording.
page 12 line 358.
"We validated this by identifying PINK1 as the only known reported hit required for Parkin translocation to damaged mitochondria within the genome guide sub-library of kinases, phosphatases and the druggable genome, demonstrating the validity of the method".

30) Pages 8, 9 - The discussion should contain a clear, fair comparison of the performance of AI-PS to arrayed screening methods in terms of cost, time, accuracy and the like. The most

important drawback of AI-PS ignoring all the technical issues I raised, is that the phenotype must be pre-selected to allow for model training. Arrayed methods (and also in-situ sequencing) do not suffer from this limitation, which must be mentioned.

The author is correct - we now stress this limitation in the discussion with the following text. Page 14 line 397.
"While machine learning methods require larger training datasets, they have a clear advantage over standard image analysis algorithms in the classification and prediction of subtle subcellular phenotypes. Classification models built with deep learning are the least influenced by human bias, since they independently decide on the image features important for distinguishing the two (or more) phenotypes. However, one limitation for such screens is that only pre-defined phenotypes can be assessed."'

31) Page 25 - In the validation screen, the actual metric used to quantify nuclear GFP-TFEB is not clearly stated.
We added the following text to the results section.
Page 10 line 280.
"For validating the screen, the TFEB-GFP positioning score was computed using a CNN based classification algorithm. The mean prediction score over time was calculated and subtracted from the non-targeting control sgRNA. To determine if there is a significant prediction score difference between the non-targeting control sgRNA and the target sgRNA we used repeated measure ANOVA."

32) Page 17 - "For optimization of the model, we performed iterations and calculated performance by area under the receiver operating characteristic (ROC) curve or precision-recall curve (in the case of asymmetric phenotype representation). The performance values were plotted against iteration to prevent data overfitting." This is inadequate. Where are these ROC/PR curves? How many iterations (presumably of X-fold cross-validation)?
We added the following text to the results section.
Page 5 line 117
"For optimization of the model, we performed iterations and calculated performance by area under the precision-recall curve. To prevent overfitting, we shuffled the feature data and split it into two unique groups, a test set and a training image set. Then, we fit an SVM model on the training set and evaluated it on the test set. Then an accuracy score was calculated. This procedure was iterated 100 times where every observation was allowed to be used in the training set or test set only once."
In addition, we added a Table indicating the features used for creating the SVM model in Fig. 1 d and e.

33) Page 17 - "A training set composed of 107,226 single-cell example images of GFP-TFEB in the nucleus or cytosol was produced." Based on the SF3 legend, I think they collected starved and non-starved cell images and used these as labels for training. If so, this is probably not a great idea. In particular, I am virtually certain there is variance over GFP-TFEB translocation into the nucleus in the starved condition (and also probably variance over GFP-TFEB cytoplasmic localization in non-starved conditions). At a minimum, the authors should go back to their imaging data and score, for a non-trivial number of cells, the degree of TFEB nuclear localization in each condition using any number of parametric methods (e.g. average pixel intensity in nucleus vs cytoplasm, etc). Showing this would at least give the reader a sense of how good we could expect the CNN, as trained, to be. A better way to do things would have

been to manually curate GFP-TFEB nuclear and GFP-TFEB cytoplasmic cells and use transfer learning to re-train an existing CNN.

As requested we compared the accuracy of the CNN model to the average pixel intensity in the nucleus vs the cytoplasm. The accuracy of pixel intensity computing is slightly greater, 90% vs 88%, however, the CNN model classification prediction is better in specificity, 97% vs 83%. In the nature of the current screen, as the frequency of the desired cell phenotype is low, specificity is more important than sensitivity. These results indicate that in the case of the TFEB translocation classification problem, both of the methods preform almost equally and sufficiently for this task. The explanation is included in the result section:
Page 8 line 223 and see also fig. S4d and e.

34) Pages 17, 18 - "Model performance" The description here is inadequate. Was this done for both types of models or just the CNN? How were images collapsed into single cells?

Yes, this was done for both the Machine Learning and CNN models. We add the following sentence to the method section under "cell segmentation……"
Page 25 line 729
"After the mask was generated, the images were collapsed into single cell images using the EBimage function stackObjects according to the mask. The function generates 150px by 150px boxes and assigned zero for all the pixels outside of the ROI mask".

35) Page 19 - " To handle outlier cells, several features were computed and the outlier features were removed." This is insufficiently detailed. Which, and what was the decision boundary for feature removal?

We address this issue by adding the following explanation to the Methods section:
Page 27 line 785
 "To handle outlier cells, the mean intensity and area of the segmented cell outline were calculated. Using the R package SCORE significant outliers values are calculated and removed".

36) Figure 1c - The panel as shown is inadequate. Details like the number of cells tracked, replicates performed, etc are needed. Also, presumably, error estimates could be generated per timepoint (since at least many cells must have been tracked).

Thank you. We regenerated the curves that are now shown in Fig. 1c and we plotted all the data as requested. We switched this new analysis for the old Fig. 1c.

37) Figure 1d - It is hard to appreciate that the upper left quadrant mitochondria are green (or at least they should be green because that is where GFP is in those cells).

Thank you, we made the mitochondrial borders thinner to fix this issue.

38) Figure 3a - Same comment as above - hard to see mito+GFP.

Thank you, we addressed as above.

39) Figure 3b - I can't quite figure out what is being shown here, but I think it is library cells following photoactivation. If that is so, an unactivated control population should be shown.

18

Troublingly, the activated population (presumably the ones in red) are not distinctly separated from the unactivated population (presumably blue). If this is so, it is a major problem and significantly worse than other similar methods. Also, all flow plots should include a description of the number of cells sorted in total.

We realize that the FACS plots in the current form are misleading. The reason is that we sorted an average of approximately between 500-2000 cells per sort. To get better assessment of the scatter plot we aggregated all the sorts together and normalized the signal intensity. Since this approach is misleading, we now show a single example plot from our screen instead the aggregate plot.
Fig. 3 b and Fig. 5 c
In addition, the flow cytometry (fcs) files from the TFEB screen and the R script analysis and gating information is provided and can be found in our GitHub repository:
https://github.com/gkanfer/AI-PS/tree/master/facs

40) Figure 4a - Not cited in text.

Thank you, we changed the figures, so it is not relevant anymore.

41) Figure S1 - A table explaining the full names/meaning of the features in S1a should be included.

This is now in Fig 1e

42) Figures S3/4 - These are cited out of order.

Thank you, we corrected this.

References

1. Al-Kofahi, Y., A. Zaltsman, R. Graves, W. Marshall, and M. Rusu. 2018. A deep learning-based algorithm for 2-D cell segmentation in microscopy images. *BMC Bioinformatics*. 19:1–11. doi:10.1186/s12859-018-2375-z.
2. Binan, L., F. Bélange, M. Uriarte, J.F. Lemay, J.C.P. de Koninck, J. Roy, E.B. Affar, E. Drobetsky, H. Wurtele, and S. Costantino. 2019. Opto-magnetic capture of individual cells based on visual phenotypes. *eLife*. 8:1–21. doi:10.7554/eLife.45239.
3. Feldman, D., A. Singh, J.L. Schmid-Burgk, A. Mezger, A.J. Garrity, R.J. Carlson, F. Zhang, and P. Blainey. 2018. Pooled optical screens in human cells. *Cell*. 179:383943. doi:10.1101/383943.
4. Hasle, N., A. Cooke, S. Srivatsan, H. Huang, J.J. Stephany, Z. Krieger, D. Jackson, W. Tang, S. Pendyala, R.J. Monnat, C. Trapnell, E.M. Hatch, and D.M. Fowler. 2020. High-throughput, microscope-based sorting to dissect cellular heterogeneity. *Molecular Systems Biology*. 16. doi:10.15252/msb.20209442.
5. Kuo, C.T., A.M. Thompson, M.E. Gallina, F. Ye, E.S. Johnson, W. Sun, M. Zhao, J. Yu, I.C. Wu, B. Fujimoto, C.C. Dufort, M.A. Carlson, S.R. Hingorani, A.L. Paguirigan, J.P. Radich, and D.T. Chiu. 2016. Optical painting and fluorescence activated sorting of single adherent cells labelled with photoswitchable Pdots. *Nature Communications*. 7:1–11. doi:10.1038/ncomms11468.

6. Nitta, N., T. Sugimura, A. Isozaki, H. Mikami, K. Hiraki, S. Sakuma, T. Iino, F. Arai, T. Endo, Y. Fujiwaki, H. Fukuzawa, M. Hase, T. Hayakawa, K. Hiramatsu, Y. Hoshino, M. Inaba, T. Ito, H. Karakawa, Y. Kasai, K. Koizumi, S.W. Lee, C. Lei, M. Li, T. Maeno, S. Matsusaka, D. Murakami, A. Nakagawa, Y. Oguchi, M. Oikawa, T. Ota, K. Shiba, H. Shintaku, Y. Shirasaki, K. Suga, Y. Suzuki, N. Suzuki, Y. Tanaka, H. Tezuka, C. Toyokawa, Y. Yalikun, M. Yamada, M. Yamagishi, T. Yamano, A. Yasumoto, Y. Yatomi, M. Yazawa, D. di Carlo, Y. Hosokawa, S. Uemura, Y. Ozeki, and K. Goda. 2018. Intelligent Image-Activated Cell Sorting. *Cell*. 175:266-276.e13. doi:10.1016/j.cell.2018.08.028.

7. Patterson, G.H., and J. Lippincott-Schwartz. 2002. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science*. 297:1873–1877. doi:10.1126/science.1074952.

8. Pau, G. 2009. Analysis of microscopy images and cellular phenotypes using EBImage. *Molecular Biology*. 1–6.

9. Quelhas, P., M. Marcuzzo, A.M. Mendonça, and A. Campilho. 2010. Cell nuclei and cytoplasm joint segmentation using the sliding band filter. *IEEE Transactions on Medical Imaging*. 29:1463–1473. doi:10.1109/TMI.2010.2048253.

10. Robinson, M., D. Mccarthy, and G.K. Smyth. 2010. edgeR : differential expression analysis of digital gene expression data How to get help Reading data. *Most*. 23:58.

11. Victora, G.D., T.A. Schwickert, D.R. Fooksman, A.O. Kamphorst, M. Meyer-Hermann, M.L. Dustin, and M.C. Nussenzweig. 2010. Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. *Cell*. 143:592–605. doi:10.1016/j.cell.2010.10.032.

12. Wählby, C., J. Lindblad, M. Vondrus, E. Bengtsson, and L. Björkesten. 2002. Algorithms for cytoplasm segmentation of fluorescence labelled cells. *Analytical Cellular Pathology*. 24:101–111. doi:10.1155/2002/821782.

13. Wang, C., T. Lu, G. Emanuel, H.P. Babcock, and X. Zhuang. 2019. Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proceedings of the National Academy of Sciences*. 116:10842–10851. doi:10.1073/pnas.1903808116.

14. Wheeler, E.C., A.Q. Vu, J.M. Einstein, M. DiSalvo, N. Ahmed, E.L. van Nostrand, A.A. Shishkin, W. Jin, N.L. Allbritton, and G.W. Yeo. 2020. Pooled CRISPR screens with imaging on microraft arrays reveals stress granule-regulatory factors. *Nature Methods*. 17:636–642. doi:10.1038/s41592-020-0826-8.

15. Wu, H., C. Wang, and Z. Wu. 2015. PROPER: Comprehensive power evaluation for differential expression using RNA-seq. *Bioinformatics*. 31:233–241. doi:10.1093/bioinformatics/btu640.

November 13, 2020

RE: JCB Manuscript #202006180R-A

Dr. Richard J Youle
NIH
NINDS
35 Convent Drive MSC 3704
Bethesda, MD 20892-3704

Dear Richard:

Thank you for submitting your revised manuscript entitled "Image-based pooled whole genome CRISPR screening for Parkin and TFEB subcellular localization". The paper has been seen again by the original reviewers and they all now recommend acceptance. Therefore, we would be happy to publish your paper in JCB pending final revisions necessary to meet our formatting guidelines (see details below).

**As you will see, reviewer #2 has one remaining concern that we'd like for you to address in the final revision. This should be quite straightforward and will only require changes to the text. Please be sure to include a brief note indicating how you addressed this concern when submitting your final revision **

To avoid unnecessary delays in the acceptance and publication of your paper, please read the following information carefully.

A. MANUSCRIPT ORGANIZATION AND FORMATTING:

Full guidelines are available on our Instructions for Authors page, https://jcb.rupress.org/submission-guidelines#revised. **Submission of a paper that does not conform to JCB guidelines will delay the acceptance of your manuscript.**

B. FINAL FILES:

Please upload the following materials to our online submission system. These items are required prior to acceptance. If you have any questions, contact JCB's Managing Editor, Lindsey Hollander (lhollander@rockefeller.edu).

1) Text limits: Character count for Tools is < 40,000, not including spaces. Count includes title page, abstract, introduction, results, discussion, and acknowledgments. Count does not include materials and methods, figure legends, references, tables, or supplemental legends. You are currently well below this limit but please bear it in mind when revising.

2) Figure formatting: Scale bars must be present on all microscopy images, including inset magnifications. Molecular weight or nucleic acid size markers must be included on all gel electrophoresis, including cropped gels. Therefore, please add molecular weight markers to the gels in Figure 9b.

3) Statistical analysis: Error bars on graphic representations of numerical data must be clearly described in the figure legend. The number of independent data points (n) represented in a graph must be indicated in the legend. Statistical methods should be explained in full in the materials and methods. For figures presenting pooled data the statistical measure should be defined in the figure legends. Please also be sure to indicate the statistical tests used in each of your experiments (both in the figure legend itself and in a separate methods section) as well as the parameters of the test (for example, if you ran a t-test, please indicate if it was one- or two-sided, etc.). Also, since you used parametric tests in your study (e.g. t-tests, ANOVA, etc.), you should have first determined whether the data was normally distributed before selecting that test. In the stats section of the methods, please indicate how you tested for normality. If you did not test for normality, you must state something to the effect that "Data distribution was assumed to be normal but this was not formally tested."

4) Title: The title should be less than 100 characters including spaces. While your current title is certainly accurate, we think that the appeal of this work will extend beyond those researchers interested in Parkin and/or TFEB function. Thus, in order to enhance the accessibility of the paper for a broad cell biology, we recommend broadening the title somewhat. One possible title is: "Image-based pooled whole genome CRISPR screening for subcellular phenotypes". Feel free to edit that as you see fit.

5) Materials and methods: Should be comprehensive and not simply reference a previous publication for details on how an experiment was performed. Please provide full descriptions (at least in brief) in the text for readers who may not have access to referenced manuscripts. The text should not refer to methods "...as previously described."

6) Please be sure to provide the sequences for all of your primers/oligos and RNAi constructs in the materials and methods. You must also indicate in the methods the source, species, and catalog numbers (where appropriate) for all of your antibodies.

7) Microscope image acquisition: The following information must be provided about the acquisition and processing of images:
a. Make and model of microscope
b. Type, magnification, and numerical aperture of the objective lenses
c. Temperature
d. imaging medium
e. Fluorochromes
f. Camera make and model
g. Acquisition software
h. Any software used for image processing subsequent to data acquisition. Please include details and types of operations involved (e.g., type of deconvolution, 3D reconstitutions, surface or volume rendering, gamma adjustments, etc.).

8) References: There is no limit to the number of references cited in a manuscript. References should be cited parenthetically in the text by author and year of publication. Abbreviate the names of journals according to PubMed.

9) Supplemental materials: There are strict limits on the allowable amount of supplemental data. Tools may have up to 5 supplemental figures. At the moment, you are below this limit but please bear it in mind when revising.

Please also note that tables, like figures, should be provided as individual, editable files. A summary of all supplemental material should appear at the end of the Materials and methods section.

10) eTOC summary: A ~40-50 word summary that describes the context and significance of the findings for a general readership should be included on the title page. The statement should be written in the present tense and refer to the work in the third person. It should begin with "First author name(s) et al..." to match our preferred style.

11) Conflict of interest statement: JCB requires inclusion of a statement in the acknowledgements regarding competing financial interests. If no competing financial interests exist, please include the following statement: "The authors declare no competing financial interests." If competing interests are declared, please follow your statement of these competing interests with the following statement: "The authors declare no further competing financial interests."

12) A separate author contribution section is required following the Acknowledgments in all research manuscripts. All authors should be mentioned and designated by their first and middle initials and full surnames. We encourage use of the CRediT nomenclature (https://casrai.org/credit/).

13) ORCID IDs: ORCID IDs are unique identifiers allowing researchers to create a record of their various scholarly contributions in a single place. At resubmission of your final files, please consider providing an ORCID ID for as many contributing authors as possible.

-- An editable version of the final text (.DOC or .DOCX) is needed for copyediting (no PDFs).

-- High-resolution figure and video files: See our detailed guidelines for preparing your production-ready images, https://jcb.rupress.org/fig-vid-guidelines.

-- Cover images: If you have any striking images related to this story, we would be happy to consider them for inclusion on the journal cover. Submitted images may also be chosen for highlighting on the journal table of contents or JCB homepage carousel. Images should be uploaded as TIFF or EPS files and must be at least 300 dpi resolution.

It is JCB policy that if requested, original data images must be made available to the editors. Failure to provide original images upon request will result in unavoidable delays in publication. Please ensure that you have access to all original data images prior to final submission.

The license to publish form must be signed before your manuscript can be sent to production. A link to the electronic license to publish form will be sent to the corresponding author only. Please take a moment to check your funder requirements before choosing the appropriate license.

Thank you for your attention to these final processing requirements. Please revise and format the manuscript and upload materials within 7 days. If complications arising from measures taken to prevent the spread of COVID-19 will prevent you from meeting this deadline (e.g. if you cannot retrieve necessary files from your laboratory, etc.), please let us know and we can work with you to determine a suitable revision period.

Please contact the journal office with any questions, cellbio@rockefeller.edu or call (212) 327-8588.

Thank you for this interesting contribution, we look forward to publishing your paper in Journal of

Cell Biology.

Sincerely,

Jodi Nunnari, PhD
Editor-in-Chief
The Journal of Cell Biology

Tim Spencer, PhD
Executive Editor
Journal of Cell Biology

Reviewer #1 (Comments to the Authors (Required)):

The authors have addressed all my minor comments. They have also added data on the role of CREB5 which suggests that different hits from their screen are functionally relevant which increases the usefulness of the screen for other researchers.

Reviewer #2 (Comments to the Authors (Required)):

In this revision the authors have responded satisfactorily to the requests of the reviewers. However, the issue of the general utility of the CNN compared to a parametric classifier is not fully addressed. At the bottom of page 10 and top of page 11 the authors indicate that mTOR was not scored as a hit because the mTOR phenotype is not sufficiently close to the deep learning model. The authors need to point out that in this case the parametric intensity measure "pixel intensity computation" would have identified mTOR as a hit. Therefore in any specific screen there can be both computational and accuracy advantages to using parametric classifiers designed based on knowledge of the system being analyzed.

Reviewer #3 (Comments to the Authors (Required)):

The authors have adequately addressed my concerns.

1) Text limits: Character count for Tools is < 40,000, not including spaces. Count includes title page, abstract, introduction, results, discussion, and acknowledgments. Count does not include materials and methods, figure legends, references, tables, or supplemental legends. You are currently well below this limit but please bear it in mind when revising.

The word count is: ~29,000

2) Figure formatting: Scale bars must be present on all microscopy images, including inset magnifications. Molecular weight or nucleic acid size markers must be included on all gel electrophoresis, including cropped gels. Therefore, please add molecular weight markers to the gels in Figure 9b.

Done

3) Statistical analysis: Error bars on graphic representations of numerical data must be clearly described in the figure legend. The number of independent data points (n) represented in a graph must be indicated in the legend. Statistical methods should be explained in full in the materials and methods. For figures presenting pooled data the statistical measure should be defined in the figure legends. Please also be sure to indicate the statistical tests used in each of your experiments (both in the figure legend itself and in a separate methods section) as well as the parameters of the test (for example, if you ran a t-test, please indicate if it was one- or two-sided, etc.). Also, since you used parametric tests in your study (e.g. t-tests, ANOVA, etc.), you should have first determined whether the data was normally distributed before selecting that test. In the stats section of the methods, please indicate how you tested for normality. If you did not test for normality, you must state something to the effect that "Data distribution was assumed to be normal but this was not formally tested."

For Fig3 c and Fig8 c we added the sentence affirming distributions were assumed to be normal.

4) Title: The title should be less than 100 characters including spaces. While your current title is certainly accurate, we think that the appeal of this work will extend beyond those researchers interested in Parkin and/or TFEB function. Thus, in order to enhance the accessibility of the paper for a broad cell biology, we recommend broadening the title somewhat. One possible title is: "Image-based pooled whole genome CRISPR screening for subcellular phenotypes". Feel free to edit that as you see fit.

5) Materials and methods: Should be comprehensive and not simply reference a previous publication for details on how an experiment was performed. Please provide full descriptions (at least in brief) in the text for readers who may not have access to referenced manuscripts. The text should not refer to methods "...as previously described."

6) Please be sure to provide the sequences for all of your primers/oligos and RNAi constructs in the materials and methods. You must also indicate in the methods the source, species, and catalog numbers (where appropriate) for all of your antibodies.

7) Microscope image acquisition: The following information must be provided about the acquisition and processing of images:
a. Make and model of microscope
b. Type, magnification, and numerical aperture of the objective lenses
c. Temperature
d. imaging medium
e. Fluorochromes
f. Camera make and model
g. Acquisition software
h. Any software used for image processing subsequent to data acquisition. Please include details and types of operations involved (e.g., type of deconvolution, 3D reconstitutions, surface or volume rendering, gamma adjustments, etc.).

8) References: There is no limit to the number of references cited in a manuscript. References should be cited parenthetically in the text by author and year of publication. Abbreviate the names of journals according to PubMed.

9) Supplemental materials: There are strict limits on the allowable amount of supplemental data. Tools may have up to 5 supplemental figures. At the moment, you are below this limit but please bear it in mind when revising.

Please also note that tables, like figures, should be provided as individual, editable files. A summary of all supplemental material should appear at the end of the Materials and methods section.

Summary follows:

Fig. S1 shows the SVM classification plot and the SVM classification and segmentation performance. Fig. S2 presents a summary of the AI-PS shiny APP platform. Fig. S3 shows the CNN classification architecture and performance while Fig. S4 addresses TFEB translocation prediction by the SVM classification model. Fig. S5 summarizes the network interaction and clustering of the hits retrieved from the whole genome CRISPR screen.

10) eTOC summary: A ~40-50 word summary that describes the context and significance of the findings for a general readership should be included on the title page. The statement should be written in the present tense and refer to the work in the third person. It should begin with "First author name(s) et al..." to match our preferred style.

Kanfer et al. developed a novel forward genetic CRISPR screening approach termed Artificial-Intelligence Photoswitchable Screening (AI-PS). Using this platform to "convert" single cells into "separate" wells through application of deep learning algorithms that detect subcellular phenotypes allowed identification of novel regulators involved in nuclear translocation of TFEB.

11) Conflict of interest statement: JCB requires inclusion of a statement in the acknowledgements regarding competing financial interests. If no competing financial interests exist, please include the following statement: "The authors declare no competing financial interests." If competing interests are declared, please follow your statement of these competing interests with the following statement: "The authors declare no further competing financial interests."

Done

12) A separate author contribution section is required following the Acknowledgments in all research manuscripts. All authors should be mentioned and designated by their first and middle initials and full surnames. We encourage use of the CRediT nomenclature (https://casrai.org/credit/).

Done

13) ORCID IDs: **ORCID IDs** are unique identifiers allowing researchers to create a

record of their various scholarly contributions in a single place. At resubmission of your final files, please consider providing an ORCID ID for as many contributing authors as possible.

Done

-- An editable version of the final text (.DOC or .DOCX) is needed for copyediting (no PDFs).

-- High-resolution figure and video files: See our detailed guidelines for preparing your production-ready images, https://jcb.rupress.org/fig-vid-guidelines.

-- Cover images: If you have any striking images related to this story, we would be happy to consider them for inclusion on the journal cover. Submitted images may also be chosen for highlighting on the journal table of contents or JCB homepage carousel. Images should be uploaded as TIFF or EPS files and must be at least 300 dpi resolution.

It is JCB policy that if requested, original data images must be made available to the editors. Failure to provide original images upon request will result in unavoidable delays in publication. Please ensure that you have access to all original data images prior to final submission.

The license to publish form must be signed before your manuscript can be sent to production. A link to the electronic license to publish form will be sent to the corresponding author only. Please take a moment to check your funder requirements before choosing the appropriate license.

Thank you for your attention to these final processing requirements. Please revise and format the manuscript and upload materials within 7 days. If complications arising from measures taken to prevent the spread of COVID-19 will prevent you from meeting this deadline (e.g. if you cannot retrieve necessary files from your laboratory, etc.), please let us know and we can work with you to determine a suitable revision period.

Please contact the journal office with any questions, cellbio@rockefeller.edu or call (212) 327-8588.

Thank you for this interesting contribution, we look forward to publishing your paper in Journal of Cell Biology.

Sincerely,

Jodi Nunnari, PhD
Editor-in-Chief
The Journal of Cell Biology

Tim Spencer, PhD
Executive Editor
Journal of Cell Biology

Reviewer #1 (Comments to the Authors (Required)):

The authors have addressed all my minor comments. They have also added data
on the role of CREB5 which suggests that different hits from their screen are
functionally relevant which increases the usefulness of the screen for other
researchers.

Reviewer #2 (Comments to the Authors (Required)):

In this revision the authors have responded satisfactorily to the requests of the
reviewers. However, the issue of the general utility of the CNN compared to a
parametric classifier is not fully addressed. At the bottom of page 10 and top of
page 11 the authors indicate that mTOR was not scored as a hit because the
mTOR phenotype is not sufficiently close to the deep learning model. The authors
need to point out that in this case the parametric intensity measure "pixel
intensity computation" would have identified mTOR as a hit. Therefore in any
specific screen there can be both computational and accuracy advantages to
using parametric classifiers designed based on knowledge of the system being
analyzed.

We have added the following sentence on lines 308-310 as suggested – "A parametric
pixel intensity computation model may have detected such an unanticipated
phenotype."

Reviewer #3 (Comments to the Authors (Required)):

The authors have adequately addressed my concerns.