Image-based pooled whole genome CRISPR screening for Parkin and TFEB subcellular localization

Gil Kanfer[1,3], Shireen A. Sarraf[1], Yaakov Maman[6], Heather Baldwin[1], Kory R. Johnson[5], Michael E. Ward[2], Martin Kampmann[4], Jennifer Lippincott-Schwartz[3] & Richard J. Youle[1*]

[1] Biochemistry Section, Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, 20892, USA

[2] Inherited Neurodegenerative Diseases Unit, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, 20892, USA

[3] Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, 20147, USA

[4] Institute for Neurodegenerative Diseases, Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA

[5] Bioinformatics Section, Information Technology Program, National Institute of Neurological Disorders and Stroke, Bethesda, MD, 20892, USA

[6] The Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel

*To whom correspondence should be addressed at youler@ninds.nih.gov

Abstract

Genome-wide CRISPR screens have transformed our ability to systematically interrogate human gene function, but are currently limited to a subset of cellular phenotypes. We report a novel pooled screening approach for a wider range of cellular and subtle subcellular phenotypes. Machine learning and convolutional neural network models are trained on the subcellular phenotype to be queried. Genome-wide screening then utilizes cells stably expressing dCas9 (CRISPRi), photoactivatable fluorescent protein (PA-mCherry), and a lentiviral guide RNA (gRNA) pool. Cells are screened by microscopy and classified by artificial intelligence (AI) algorithms, which precisely identify the genetically altered phenotype. Cells with the phenotype of interest are photoactivated, isolated via flow cytometry, and the gRNAs are identified by sequencing. A proof-of-concept screen accurately identified PINK1 as essential for Parkin recruitment to mitochondria. A genome-wide screen identified factors mediating TFEB relocation from the nucleus to the cytosol upon prolonged starvation. Twenty of the sixty-four hits called by the neural network model were independently validated, revealing new effectors of TFEB subcellular localization. This approach, AI-Photoswitchable Screening (AI-PS) offers a novel screening platform capable of classifying a broad range of mammalian subcellular morphologies, an approach largely unattainable with current methodologies at genome-wide scale.

Introduction

Recent advances have expanded traditional genetic screens from bacteria and yeast to mammalian cells. RNAi, CRISPRi and CRISPR screens rely on two main strategies: arrayed and pooled screens. Arrayed screens are highly specific but require the production and, by definition, individual assortment of each RNAi or CRISPR guide separately, requiring high-throughput equipment not readily available to academic labs. Pooled screens are more facile but restricted to phenotypes that affect cell growth rates or viability or result in a fluorescence increase that allows isolation of hits from the population using FACS. Single-cell RNA based pooled screens are also useful to link genetic profiles to perturbations (Horlbeck et al., 2018; Datlinger et al., 2017; Dixit et al., 2016; Adamson et al., 2016). Recent reports describe new strategies that distinguish microscopic differences in images from pooled screens. Microfluidic separation of cells from the population (Ota et al., 2018; Nitta et al., 2018), *in situ* sequencing of guides in individual cells (Wang et al., 2019; Feldman et al., 2018; Emanuel et al., 2017), microscale cell carrier technology (Wheeler et al., 2020), link genotypes to images of cell phenotypes. Though elegant, these approaches have technical limitations and are not applicable at the whole genome scale.

Recent advances in machine learning, and particularly in deep learning (convolutional neural networks) (Caicedo et al., 2018; Bzdok et al., 2018) offer novel strategies for identifying individual cells with altered organelle morphology or subcellular protein localization. We developed a unique screening method to identify genetic perturbations of subcellular morphologies that is widely applicable and high-throughput. The method is divided into four steps: first, a morphology classification model is trained on single-cell images. Second, pools of CRISPRi-perturbed target cells are imaged sequentially, and the phenotypically selected cells are labeled by laser photoactivation of a fluorescent protein. Third, the photoactivated cells are sorted and fourth, the guides within phenotypically identified cells are amplified and sequenced. The decision to select cells is made on-the-fly by pre-trained classification models allowing screening of $1x10^6$ cells within 12 hours and the whole human genome in a week.

RESULTS

Building the single-cell imaging screening approach

We developed a new platform which assesses images of cells and uses machine learning to distinguish their subcellular phenotypes. Using laser activation of a fluorescent probe to denote the selected cell phenotypes and FACS to separate the cells for guide sequencing, one essentially converts the individual cells exposed to pooled CRISPRi libraries into an arrayed screen (Fig. 1). Using this approach, every imaged cell is referred to as an independent entity, and a predicted phenotype score is produced based on a classification machine-learning model (Fig. 1 e). Making the AI platform entails three steps: training and creation of the phenotype classification model; model deployment on pooled imaged cells; and validation of the model's screening performance. We utilized Pink1-dependent Parkin translocation to mitochondria as a proof-of-concept (Fig. 1 b). In cells with unimpaired polarized mitochondria, Parkin is in the cytoplasm; however, upon mitochondrial depolarization, it translocates to mitochondria(Narendra et al., 2008) (Fig. 1 b,c). This binary switch in Parkin location is suitable for detection by a support vector machine (SVM) classification model. An SVM classification model was trained on images of cells with either cytosolic or mitochondrial GFP-Parkin. For each single-cell image, the program calculates a broad set of features based on measurements of the GFP signal pixels (Fig. 1 d,e). Feature variations were assessed, and redundant contributions were excluded from the model (Fig. S1a,b). Manually selected cell images annotated by phenotype and image features were computationally applied on a radial basis kernel SVM to create the classification model (Fig. S1 a and b). For this step, we generated an easy-to-use graphical user interface program to facilitate image segmentation, measurement and model building (Fig 2). The R-based script for image segmentation and analysis, and the SVM-classification model (Fig. 2b) were deployed on-the-fly to identify cells exhibiting the desired phenotype (GFP-Parkin mitochondrial localization). During live cell image acquisition, single cell images were captured following segmentation and stored on a local computer. The SVM-based model classified the individual cells and generated a mask corresponding to the live image field identifying the location of cells with the phenotype of interest (Fig. 1g, Extended Data Fig. 2a,b, and Video 1). In cells identified with this mask, PA-mCh was then laser photoactivated. This 10 second process was iterated across serial images of the entire chamber slide (an average of 500,000 cells for one subgenomic CRISPRi guide pool (Gilbert et al., 2014; Horlbeck et al., 2016)). Finally, the photoactivated samples were sorted using flow cytometry and deep sequenced to determine sgRNA abundance in the activated sample compared to untreated cells.

Parkin translocation screen validation

For platform validation, U2OS cells stably expressing GFP-Parkin, PA-mCh, and dCAS9 were infected with subpool of the version 2 CRISPRi library comprising 12,775 guides targeting kinases, phosphatases and the druggable genome (Horlbeck et al., 2016). Cells were treated with carbonyl cyanide m-chlorophenyl hydrazone (CCCP) to depolarize mitochondria and GFP-Parkin localization was assessed using the SVM classification model (Fig. 3 a). From approximately 200,000 cells, 1,132 were called, photoactivated, sorted and sequenced (Fig. 3 b). The most abundant sgRNAs identified in the photoactivated samples were targeted against PINK1 (Fig. 3 c), known to be required for Parkin translocation[16], exhibiting a nearly 30-fold increase compared to the unsorted control sample (FDR adjusted p-value < 0.0001, Table S1). Thus, the single known Parkin modifier targeted by the subpool library, PINK1, was identified, validating the method. In addition, sample size estimation indicated that three biological repeats are sufficient for detecting the desired genetic link in our experimental setup (Fig. S2 d).

TFEB nuclear localization screen: convolutional neural network (CNN) based screen

To explore a subcellular phenotype with more complex regulation, we screened for genes affecting the nuclear localization of the transcription factor, TFEB. Upon nutrient starvation, TFEB moves from the cytosol to the nucleus, where it activates the transcription of lysosome- and autophagy-related genes (Settembre et al., 2011). Upon prolonged starvation, mTOR is reactivated, presumably due to replenishment of nutrients through autophagy, lysosomes repopulate the cells (Yu et al., 2010) and TFEB returns to the cytosol (Fig. 4 b, Fig S4 and Video 2). As the import of TFEB to the nucleus is well elucidated (Puertollano et al., 2018), we assessed TFEB reappearance in the cytosol following prolonged starvation-induced nuclear import. U2OS cells stably expressing GFP tagged TFEB, PA-mCH, and dCas9 (designated as TFEB-GFP) were infected with a lentiviral library expressing sgRNAs against the entire genome divided in seven separate subpools (Horlbeck et al., 2016). The screen was split into seven subscreens, one per day for seven days. To increase reproducibility, each subpool screen was repeated at least 3 times.

Because SVM classification failed to predict TFEB nuclear localization accurately (performance compression between area under the precision-recall curve of 72% for the TFEB SVM classification model versus 99% for the Parkin model (Fig. S2 c and Fig. S3 a), we used deep learning via a convolutional neural network (CNN) (Fig. 4 c,d). The training set was composed of 100,000, 150 pixel X 150 pixel single-cell images. The CNN architecture was based on ImageNet (Deng et al., 2010) architecture and composed of three deconvolutions and four Max pooling processes, which were followed by a fully connected dense network (Fig. 4 c,d).

TFEB-GFP cells expressing guide libraries were grown under complete nutrient deprivation conditions for eight hours prior to the commencement of screening after which those cells retaining TFEB in the nucleus were photoactivated (Fig. 5 a), isolated by FACS and deep sequenced (Fig. 4 e,f and Fig. S3b, Video 3). Among the seven subpooled libraries, a mean accuracy of 90% was calculated from the approximation of the area under the precision-recall curve (Fig. S3 c). The entire photoactivated and sorted gene abundance ranking list analysed for ontology clusters revealed enrichment in mitochondrial and kinase complex gene sets (Puertollano et al., 2018; Nezich et al., 2015), that may relate to energetic consequences to mitochondrial states and TFEB post-translational regulation, respectively (Fig. 5b and Fig. S5). Plasma membrane proteins were also enriched, perhaps related to cell division rates or nutrient import.

Differential sgRNA abundance analysis between unsorted and photoactivated/sorted samples showed a significant fold-change enrichment in 64 genes (Fig. 5c, Table S2). A second validation screen was conducted of the 64 enriched genes using two new sgRNAs. As with the whole-genome screen, TFEB-GFP nuclear localization following validation guide transduction during prolonged starvation was recorded 8 hours after starvation for 10 hours. The perturbation effect on TFEB positioning was compared to a non-targeting control sgRNA. We found that 21 of the 64 sgRNAs from the whole genome analysis significantly extended nuclear TFEB retention ($p < 0.05$, repeated-measures ANOVA, Fig. 5 d). Interestingly, these 21 validated hits were amongst the genes with the highest ranked p-value significance in the whole genome screen (Table S2). Amongst the validated genes, the signaling receptor TGFBR1 was enriched in the secondary TFEB screen (Fig. 5 d,e and Fig. 6a, Video 4). This may be related to a previous report of the induction of another MITF family transcription factor, TFE3, by the loss of TGFBR1 (Sun et al., 2016). Validating the screen, one of the strongest hits is the transcription factor, CREB (Fig. 5d,e and Fig. 6b, Video 5), which has been shown previously to mediate autophagy and induce the expression of several autophagy genes including Ulk1, Atg5 and Atg7 upstream of TFEB following starvation and TFEB itself (Seok et al., 2014). Certain autophagy genes are more predominantly activated by CREB and others more by TFEB. In addition, loss of another hit, Pitx2, *in vivo* causes an increase in mitophagy that has been linked to TFEB activation (Nezich et al., 2015; Chang et al,. 2019). Additionally, the membrane protein, TMEM184b, has been reported previously to play a role in autophagy (Fig. 5 d,e) (Bhattacharya et al., 2016; Agod et al., 2018). Loss of the phosphatase, PPP1R1B, which also scored amongst the top validated hits, resulted in significant retention in TFEB in the nucleus upon starvation (Fig. 5 d,e and Fig. 6c, Video 6). As phosphorylation of TFEB is intimately linked to its activation and subcellular localization (Puertollano et al., 2018), this hit deserves further mechanistic study. The extensively studied TFEB regulator mTOR was not significantly enriched in our photoactivated samples. To explore this explicitly, live-cell imaging of starved GFP-TFEB infected with two distinct sgRNAs targeting mTOR showed an accumulation of TFEB on lysosomes, which resulted in punctate cytosolic foci, similar to previous reports (Martina and Puertollano, 2013; Settembre et al., 2012) (Fig. 6d, Video 7). Therefore, mTOR was not identified in the enrichment analysis owing to the lack of classification of this specific mTOR phenotype, which is distinct from the deep learning model trained for nuclear localization.

Discussion

Here, we present a platform that applies machine learning and deep learning algorithms to allow pooled genetic screening for subcellular image phenotypes. This method, which we call Artificial Intelligence-Photoswitchable Screening (AI-PS), reduces the time, cost and complexity compared to standard screening methods that have required arrayed RNAi or CRISPR libraries.

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. For a field of approximately 200 cells, these four steps together take an average of 10 seconds, which is then iterated across an entire plate. Therefore, a screen of 600,000 cells infected with 1/7th of the genome guide library, composed of 12,500 sgRNAs, takes ~12 hours. Hence, this accelerated platform coupled with a user-friendly interface should accelerate the utility of pooled genomic screens.

The method enables the detection and labelling of cells according to subcellular protein localization. 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.

We also applied AI-PS to explore a completely different protein translocation process, one which again would be undetectable via FACS separation of whole cells based on a change in overall fluorescence intensity. The transcription factor, TFEB, is retained in the cytosol in growing cells and upon starvation relocalizes to the nucleus, where it induces transcription of lysosomal- and autophagy-related genes (Settembre et al., 2011; Sardiello et al., 2009). Upon prolonged starvation, TFEB returns to the cytosol via an undefined process. Either nuclear TFEB migrates back to the cytosol or nuclear TFEB is degraded while newly synthesized TFEB repopulates the cytosol. As we found minimal evidence for a role of cytoskeletal or nuclear transporter proteins, whether the appearance of TFEB in the cytosol is due to physically shuttling of preexisting TFEB or to an increase in translation of new TFEB remains an open question. Beyond protein localization screens, our method will be useful to identify genes involved in regulation of organelle abundance, size and shape.

Similar to the concepts presented in our study, machine learning-based image analysis has been used for calling and sorting of cells (Ota et al., 2018; Nitta et al., 2018). However, AI-PS conveys distinct advantages. First, the microscopic resolution of AI-PS is much higher than that utilized during dissociated cell sorting (Ota et al., 2018; Nitta et al., 2018), allowing the identification of more difficult to detect subcellular structures.

Specifically, the detection of minor subcellular events such as alteration in protein distribution, positioning and motion requires high spatial-temporal resolution image acquisition. Previously published methods utilized low magnification objectives (4x and 10x) and very short exposure times (below 50 msec), which resulted in low signal-to-noise ratios, and are not suitable for the resolution of subcellular events.

Another advantage of the AI-PS platform is its wide accessibility – there is no need for specialized flow instrumentation, and the algorithms and code presented here can be adapted easily for a variety of microscope systems. AI-PS is compatible with adherent tissue culture cells, unlike sorting-based approaches for which cells must be in suspension, further allowing a more accurate examination of subcellular events in regular culture conditions. One current limitation of AI-PS is that the cells must be screened live to allow trypsinization to produce single cell suspension for FACS. Because some phenotypes would be better screened in fixed cells, we are developing methods enabling single cell release of fixed cells to allow screening of additional cell biology processes.

Another step to improve AI-PS would be to reduce the segmentation time per image to speed up the screen. Fortunately, a huge improvement in cell segmentation, specifically the development of deep learning-based techniques, such as U-Net segmentation (Caicedo et al., 2019; Toth, 2018; Ronneberger et al., 2015) can be utilized in AI-PS. This new deep learning-based segmentation has the potential for at least a five-fold reduction in analysis time. Increasing the speed will make it possible to increase the sample size and thereby increase the sgRNA coverage in the sorted samples, decreasing the false discovery rate. Another strategy for increasing specificity would be to utilize single-cell DNA sequence analysis.

In conclusion, our platform demonstrates novel implementation of machine learning to improve cell biology research and discovery and enables phenotypic-based screening at the subcellular level – an approach previously largely unavailable. Additionally, AI-PS can be implemented for drug target exploration and may prove valuable in methods targeting single cells within complex human samples.

References

Adamson, B., T.M. Norman, M. Jost, M.Y. Cho, J.K. Nuñez, Y. Chen, J.E. Villalta, L.A. Gilbert, M.A. Horlbeck, M.Y. Hein, R.A. Pak, A.N. Gray, C.A. Gross, A. Dixit, O. Parnas, A. Regev, and J.S. Weissman. 2016. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell*. 167:1867-1882.e21. doi:10.1016/j.cell.2016.11.048.

Agod, Z., K. Pazmandi, D. Bencze, G. Vereb, T. Biro, A. Szabo, E. Rajnavolgyi, A. Bacsi, P. Engel, and A. Lanyi. 2018. Signaling lymphocyte activation molecule family 5 enhances autophagy and fine-tunes cytokine response in monocyte-derived dendritic cells via stabilization of interferon regulatory factor 8. *Front. Immunol.* 9. doi:10.3389/fimmu.2018.00062.

Bhattacharya, M.R.C., S. Geisler, S.K. Pittman, R.A. Doan, C.C. Weihl, J. Milbrandt, and A. DiAntonio. 2016. TMEM184b promotes axon degeneration and neuromuscular junction maintenance. *J. Neurosci.* 36:4681–4689. doi:10.1523/JNEUROSCI.2893-15.2016.

Bzdok, D., M. Krzywinski, and N. Altman. 2018. Points of significance: Machine learning: Supervised methods. *Nat. Methods*. 15:5–6. doi:10.1038/nmeth.4551.

Caicedo, J.C., A. Goodman, K.W. Karhohs, B.A. Cimini, J. Ackerman, M. Haghighi, C.K. Heng, T. Becker, M. Doan, C. McQuin, M. Rohban, S. Singh, and A.E. Carpenter. 2019. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nat. Methods*. 16:1247–1253. doi:10.1038/s41592-019-0612-7.

Carpenter, A.E., T.R. Jones, M.R. Lamprecht, C. Clarke, I.H. Kang, O. Friman, D.A. Guertin, J.H. Chang, R.A. Lindquist, J. Moffat, P. Golland, and D.M. Sabatini. 2006. CellProfiler : image analysis software for identifying and quantifying cell phenotypes. Genome Biology. 7. doi:10.1186/gb-2006-7-10-r100.

Chih-Ning Chang, Arun J. Singh, Michael K. Gross, and C.K. 2019. Requirement of Pitx2 for Skeletal Muscle Homeostasis. *Dev Biol*. 176:90–102. doi:10.1016/j.gde.2016.03.011.

Dai, Z., J.M. Sheridan, L.J. Gearing, D.L. Moore, S. Su, S. Wormald, S. Wilcox, L.O. Connor, R.A. Dickins, M.E. Blewitt, and M.E. Ritchie. 2019. edgeR : a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens [version 2 ; peer review : 3 approved]. F1000Research 2014, 3:95

Datlinger, P., A.F. Rendeiro, C. Schmidl, T. Krausgruber, P. Traxler, J. Klughammer, L.C. Schuster, A. Kuchler, D. Alpar, and C. Bock. 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods*. 14:297–301. doi:10.1038/nmeth.4177.

Deng, J., W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei. 2010. ImageNet: A large-scale hierarchical image database. Institute of Electrical and Electronics Engineers (IEEE). 248–255.

Dixit, A., O. Parnas, B. Li, J. Chen, C.P. Fulco, L. Jerby-Arnon, N.D. Marjanovic, D. Dionne, T. Burks, R. Raychowdhury, B. Adamson, T.M. Norman, E.S. Lander, J.S. Weissman, N. Friedman, and A. Regev. 2016. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*. 167:1853-1866.e17. doi:10.1016/j.cell.2016.11.038.

Emanuel, G., J.R. Moffitt, and X. Zhuang. 2017. High-throughput, image-based screening of pooled genetic-variant libraries. *Nat. Methods*. 14:1159–1162. doi:10.1038/nmeth.4495.

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.

Fu, G., and L. Yi. 2019. Tuning model parameters in class-imbalanced learning with precision-recall curve. BIMJ.61:3:652–664. doi:10.1002/bimj.201800148.

Gilbert, L.A., M.A. Horlbeck, B. Adamson, J.E. Villalta, Y. Chen, E.H. Whitehead, C. Guimaraes, B. Panning, H.L. Ploegh, M.C. Bassik, L.S. Qi, M. Kampmann, and J.S. Weissman. 2014. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. 159:647–661. doi:10.1016/j.cell.2014.09.029.

Grau, J., I. Grosse, and J. Keilwagen. PRROC : computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinfo.* 31:15:2595-2597.

Horlbeck, M.A., L.A. Gilbert, J.E. Villalta, B. Adamson, R.A. Pak, Y. Chen, A.P. Fields, C.Y. Park, J.E. Corn, M. Kampmann, and J.S. Weissman. 2016. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife*. 5:1–20. doi:10.7554/eLife.19760.

Horlbeck, M.A., A. Xu, M. Wang, N.K. Bennett, C.Y. Park, D. Bogdanoff, B. Adamson, E.D. Chow, M. Kampmann, T.R. Peterson, K. Nakamura, M.A. Fischbach, J.S. Weissman, and L.A. Gilbert. 2018. Mapping the Genetic Landscape of Human Cells. *Cell*. 174:953-967.e22. doi:10.1016/j.cell.2018.06.010.

Laufer, C., B. Fischer, M. Billmann, W. Huber, and M. Boutros. 2013. Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat. Methods*. 10:427–31. doi:10.1038/nmeth.2436.

Sardiello, M., M. Palmieri, A. di Ronza, D.L. Medina, M. Valenza, V.A. Gennarino, C. Di Malta, F. Donaudy, V. Embrione, R.S. Polishchuk, S. Banfi, G. Parenti, E. Cattaneo, and A. Ballabio. 2009. A Gene Network Regulating Lysosomal Biogenesis and Function. Science (80-.). 325:473 LP – 477. doi:10.1126/science.1174447.

Martina, J.A., and R. Puertollano. 2013. Rag GTPases mediate amino acid-dependent recruitment of TFEB and MITF to lysosomes. *J. Cell Biol.* 200:475–491. doi:10.1083/jcb.201209135.

Narendra, D., A. Tanaka, D.F. Suen, and R.J. Youle. 2008. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *J. Cell Biol.* 183:795–803. doi:10.1083/jcb.200809125.

Nezich, C.L., C. Wang, A.I. Fogel, and R.J. Youle. 2015. MiT/TFE transcription factors are activated during mitophagy downstream of Parkin and Atg5. *J. Cell Biol.* 210:435–450. doi:10.1083/jcb.201501002.

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.

Ota, S., R. Horisaki, Y. Kawamura, M. Ugawa, I. Sato, K. Hashimoto, R. Kamesawa, K.

Setoyama, S. Yamaguchi, K. Fujiu, K. Waki, and H. Noji. 2018. Ghost cytometry. *Science.* 360:1246–1251. doi:10.1126/science.aan0096.

Pau, G. 2009. Analysis of microscopy images and cellular phenotypes using EBImage. *Mol. Biol.* 1–6.

Pickrell, A.M., and R.J. Youle. 2015. The roles of PINK1, Parkin, and mitochondrial fidelity in parkinson's disease. *Neuron*. 85:257–273. doi:10.1016/j.neuron.2014.12.007.

Puertollano, R., S.M. Ferguson, J. Brugarolas, and A. Ballabio. 2018. The complex relationship between TFEB transcription factor phosphorylation and subcellular localization. *EMBO J.* 37:e98804. doi:10.15252/embj.201798804.

Reimand, J., R. Isserlin, V. Voisin, M. Kucera, C. Tannus-lopes, A. Rostamianfar, L. Wadi, M. Meyer, J. Wong, C. Xu, D. Merico, and G.D. Bader. 2019. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA , Cytoscape and EnrichmentMap. *Nat Protoc*. 14, 482–517 (2019). https://doi.org/10.1038/s41596-018-0103-9

Robinson, M., D. Mccarthy, and G.K. Smyth. 2010. edgeR : differential expression analysis of digital gene expression data. *Bioinformatics*, Volume 26, Issue 1, 1 January 2010, Pages 139–140, https://doi.org/10.1093/bioinformatics/btp616

Ronneberger, O., P. Fischer, and T. Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*. 9351:234–241. doi:10.1007/978-3-319-24574-4_28.

Settembre, C., C. Di Malta, V.A. Polito, M.G. Arencibia, F. Vetrini, S. Erdin, S.U. Erdin, T. Huynh, D. Medina, P. Colella, M. Sardiello, D.C. Rubinsztein, and A. Ballabio. 2011. TFEB Links Autophagy to Lysosomal Biogenesis. Science. 332:1429 LP – 1433. doi:10.1126/science.1204592.

Settembre, C., R. Zoncu, D.L. Medina, F. Vetrini, S. Erdin, S. Erdin, T. Huynh, M. Ferron, G. Karsenty, M.C. Vellard, V. Facchinetti, D.M. Sabatini, and A. Ballabio. 2012. A lysosome-to-nucleus signalling mechanism senses and regulates the lysosome via mTOR and TFEB. *EMBO J.* 31:1095–1108. doi:10.1038/emboj.2012.32.

Sun, Z.-J., G.-T. Yu, C.-F. Huang, L.-L. Bu, J.-F. Liu, S.-R. Ma, W.-F. Zhang, B. Liu, and L. Zhang. 2016. Hypoxia induces TFE3 expression in head and neck squamous cell carcinoma. *Oncotarget*. 7:11651–11663. doi:10.18632/oncotarget.7309.

Tian, R., M.A. Gachechiladze, C.H. Ludwig, M.T. Laurie, J.Y. Hong, D. Nathaniel, A. V. Prabhu, M.S. Fernandopulle, R. Patel, M. Abshari, M.E. Ward, and M. Kampmann. 2019. CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron*. 104:239-255.e12. doi:10.1016/j.neuron.2019.07.014.

Hollandi, R. and Horvath, P. A deep learning framework for nucleus segmentation using image style transfer. Preprint available at https://www.biorxiv.org/content/10.1101/580605v1 (2019).

Wang, C., T. Lu, G. Emanuel, H.P. Babcock, and X. Zhuang. 2019. Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proc. Natl. Acad. Sci.* 116:10842–10851. doi:10.1073/pnas.1903808116.

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. arrays reveals stress granule-regulatory factors. *Nat. Methods*. 17, 636–642 (2020). https://doi.org/10.1038/s41592-020-0826-8

Seok, S., Fu, T., Choi, S. et al. Transcriptional regulation of autophagy by an FXR–CREB axis. Nature 516, 108–111 (2014). https://doi.org/10.1038/nature13949

Yu, L., C.K. McPhee, L. Zheng, G.A. Mardones, Y. Rong, J. Peng, N. Mi, Y. Zhao, Z. Liu, F. Wan, D.W. Hailey, V. Oorschot, J. Klumperman, E.H. Baehrecke, and M.J. Lenardo. 2010. Termination of autophagy and reformation of lysosomes regulated by mTOR. *Nature*. 465:942–946. doi:10.1038/nature09076.

Methods

Cell lines, constructs and reagents

U2OS and HEK293T were cultured in a humidified incubator at 37°C and 5% CO_2 and maintained in DMEM (Life Technologies) supplemented with 10% (v/v) FBS (Gemini Bio Products), 10 mM HEPES (Life Technologies), 1 mM sodium pyruvate (Life Technologies), 1 mM non-essential amino acids (Life Technologies) and 2 mM glutamine (Life Technologies). Testing for mycoplasma contamination was performed bimonthly using the PlasmoTest kit (InvivoGen).

For constituting a stably expressing dCas9-KRAB U2OS cell line we took a similar approach as was described here. In brief, pC13N-dCas9-BFP-KRAB (127968, Addgene) was integrated in the U2OS genome using F-Talen and R-Talen (pZT-C13-R1 and pZT-C13-L1, Addgene:62196, 62197) targeting the human CLYBL intragenic safe harbor locus between exons 2 and 3 (as was described previously (Tian et al., 2019)). The U2OS-dCas9 cell line was then subcloned and the dCas9 activity assessed to select the most potent clones for further use (Fig. S1 c).

To generate the parental U2OS-dCas9-PA-mCh, photoactivatable-mCherry was PCR-amplified from the plasmid N-PA-mCh and assembled into the retroviral vector pBABE-puro using HiFi DNA Assembly (E5520S, NEB). To create the stable U2OS-dCas9-PA-mCh/GFP-Parkin and U2OS-dCas9-PA-mCh/TFEB-GFP cell lines, Parkin or TFEB was inserted into the lentiviral pHAGE vector by HiFi DNA Assembly (E5520S, NEB). The cell lines were subcloned and cells expressing low levels of the GFP-tagged proteins were selected to prevent overexpression artifacts. For nucleus segmentation we used a lentiviral plasmid expressing nuclear-localized Halo-tag, hU6-bsd-NLS-Halo. Prior to the screen, HBSS was supplemented with 2 μM of the pa Janelia Dye 646, SE, (TOCRIS). For the Parkin screen, the nucleus was detected using 1000x dilution of Draq5 (62251, ThermoFisher).

For Parkin-induced mitophagy, GFP-Parkin cells were treated with 10 μM Carbonyl Cyanide Chlorophenylhydrazone (CCCP) (Sigma-Aldrich) and 0.1 μM Bafilomycin A (Sigma-Aldrich). For TFEB screening, cells were starved in Hank’s Balanced Salt solution (HBSS) without calcium and magnesium (14170112, ThermoFisher).

Parkin-GFP and TFEB-GFP positioning classification by support vector machine (SVM)

To create the classification model, we initially trained 2,234 images of each of the binary phenotypes: Parkin or TFEB translocation. GFP-Parkin signal was mitochondrial versus cytosolic while TFEB-GFP was nuclear versus cytosolic. The model was created using the R library e1071. In brief, we used a radial basis Kernel with a cost violation of 10 computed for an example set of phenotypes using the radial Kernel formula: $e^{(-\gamma|u-v|\text{^}2)}$ 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.

GFP-TFEB positioning classification by Convolutional neural network

For TFEB localization classification, an ImageNet (Deng et al., 2010) architecture CNN model was created using TensorFlow and the KERAS library. A training set composed of 107,226 single-cell example images of GFP-TFEB in the nucleus or cytosol was produced. 80% of the data was used for training and 15% for validation. The remaining 5% was used for testing the model performance. Image input size was 150 pixel X 150 pixel and three steps of convolution and max pooling were conducted at a learning rate of $1e^{-4}$.

Training was performed with 50 epochs and a batch size of 200. Overfitting was prevented by employing the built-in Keras callbacks API feature to save the model weights after each epoch. The selected model was chosen from the epoch* at which the validation and training loss curves were no longer decreasing. The variation in fluorescence signal intensity was accounted for by randomly applying brightness augmentation (10% to 90%) to the images in the training data set.

Model performance

To assess classification model performance, we performed a precision-recall curve in which the curve integral was a measurement of accuracy (Fu and Yi, 2019). In brief, 5-10% of

images in the data set from our experiment were arbitrary selected for performance testing. Images were collapsed into single cells. The parameters extracted for constructing the precision-recall curve were the corresponding CNN prediction value against the ground truth class. The curve and the AUC were plotted and calculated using the R package PRROC (Grau et al.).

Image acquisition and model deployment

SVM deployment live-image acquisition was done on a Nikon Ti-2 CSU-W1 spinning disk confocal system equipped with a high-speed electron-multiplying charge-coupled device camera (Evolve 512; Photometrics) using a 20X air objective (NA 0.75) with an environmental control chamber (temp controlled at 37^0C and CO_2 at 5%) operated by Nikon elements AR microscope imaging software.

Cells were seeded at for screening at $1x10^5$ cells per well of a 2 well Lab-Tek chamber slide (Thermofisher, 155360). The on-the-fly real-time capture was done using the 488 nm laser channel for excitation and using the 520 nm emission detector to collect the GFP signal and the 647 nm excitation laser and 667 nm emission detector for the segmentation channel. Saved images were segmented live using a bash file script (https://github.com/gkanfer/AI-PS), and the classifications were deployed by the SVM model. A mask file containing the selected cells was generated and stored on the local computer. The mask image was used to photoactivate the called regions by exciting with a 405 nm wavelength using a Bruker minscanner XY galvo photostimulation scanner. The process was iterated across more than 1000 fields of view (512 px by 512 px for the Parkin screen and 2048 px by 2044 px for the TFEB screen). The Nikon Imaging System (NIS) elements AR microscope software was used in JOB mode to allow integration of the deployment code on-the-fly (the JOB file can be found on our https://github.com/gkanfer/AI-PS/). In brief, following capture and saving of the 488 nm and 647 nm channel images on the local computer, the NIS JOB module OUTPROC is activated and directed to run the segmentation and deployment R script. Next, the region of interest mask is generated, uploaded back on the local microscope computer hard drive on the outproc folder path, after which the NIS-JOB continues by saving the mask coordinates and preforming the photoactivation of the selected regions of interest with a 405 nm laser. The microscope stage then moves to the next field of view to repeat the process.

Live-cell image acquisition and deployment of the CNN-based screen were carried out on the Eclipse Ti2-E (Nikon) with CSU-W1 spinning disk system equipped with an ORCA-FLASH 4.0 V3 sCMOS (Hamamatsu) and an Opti-Microscan XY Galvo Scanning Unit and a Nikon LUN-F laser unit with 90mW 405NM, rated 90mW output at fiber tip, using a 20X objective (NA 0.75) and environmental control chamber (temp controlled at 37^0C and CO_2 at 5%). The microscope was controlled by the NIS elements AR microscope imaging software.
The on-the-fly real-time acquisition and deployment for the CNN-based screen were performed as described above with one major modification: the TensorFlow deployment script was running the backend "while-loop" throughout the acquisition (https://github.com/gkanfer/AI-PS/tree/master/TFEB_screen).

Cell segmentation analysis and processing

For image manipulation, the R package EBimage (Pau, 2009) was used similar to a previous report (Laufer et al., 2013). In brief, the two-channel images were min/max-normalized and nuclear staining was used as a seed to identify individual cells. For nucleus segmentation, thresholding with a 5x5 filter map and Watershed transformation were applied. Then, the target channel (designated GFP), was used to identify cell borders and edges for segmentation, after which it was used for classification. High-pass filtering and local thresholding followed by global thresholding were used to create global and local masks. Together with the nucleus mask generated in the first step, this mask was used for the Cellprofiler (Carpenter et al., 2006)-based EBimage propagation function. To handle outlier cells, several features were computed and the outlier features were removed. For SVM classification, preselected features were computed and used for classification. For the CNN classification, single cells were extracted and stacked into tensor array configuration which is compatible with CNN-based prediction analysis.

sgRNA lentiviral production

To generate lentivirus expressing sgRNA libraries, CRISPRi sub-pooled libraries were used (Horlbeck et al., 2016). On day 0, 7.5 x 10^7 Hek293-lentiX cells (Clontech) were seeded on 15-cm tissue culture plates. The next day (Day 1), 20 µg/ml subpooled sgRNA plasmid, 14.1 µg/ml PAX2, 4.2 mg/ml MDG2 and 1.2 µg/ml pAdvantage (3rd generation lentiviral vector packaging systems) were transfected using 75 µl of Lipofectamine 2000 (11668019, ThermoFisher) in Opti-MEM (ThermoFisher). On day 2 media was changed, and on day 3 virus was harvested. A lentivirus precipitation kit (VC100, Alstem Cell Advancements) was used according to the manufacturer's suggestions to concentrate the virus.

To determine MOI, 0.1 x 10^6 cells were seeded in 24-well plates and infected with four titrations of the concentrated virus. Genomic DNA was isolated using QIAamp DNA Micro Kit (56304, Qiagen). The number of genomic viral integration sites was compared with the number of housekeeping genes using a ddPCR—BioRad QX200 AutoDG Droplet Digital PCR System (BioRad). The volume to MOI ratio was calculated using the following formula:

1) Insertion number (from ddPCR) x dilution factor = Transducing Units (TU)
2) (Desired MOI x cell number) / Transducing Units = Virus Volume

The ddPCR primer mix for amplifying upstream of the sgRNA integration region was purchased from BioRad:
Gaagaagaaggtggagagagagacagagacagatccattcgattagtgaacggatcggcactgcgtgcgccaattct gcagacaaatggcagtattcatccacaattttaaaagaaaagggggg (FAM)
The house keeping probe used for comparison was EiF2C1 (Assay ID: dHsaCP2500349 Cat: 10031243, BioRad).

To conduct the screen, library expression, 5 x 10^6 dCas9-PA-mCh-expressing cells were seeded on day 0. The next day, the appropriate virus volume was added to cells to achieve an MOI below five. Two days after infection, sgRNA-expressing cells were sorted using a 407 nm Laser and 450/50 nm filters. Following four days of growth, cells were reseeded in 2-well screening chambers. To maintain sufficient sgRNA representation, cells were maintained at numbers corresponding to a coverage of at least 100 cells per sgRNA.

Activated sample isolation

After screening, cells were detached using trypsin (Sigma), washed once with PBS and filtered using a 50 micron sieve (Corning) to obtain a single-cell suspension. The volume was adjusted to obtain up to 10 million cells per ml using PBS. Cells were kept in the dark on ice until sorting, which was done using a BD FACS Aria cell sorter equipped with 355 nm, 407 nm, 532 nm, and 640 nm laser lines and BD FACSDIVA software to perform aseptic cell sorting. Physical properties (FSC and SSC parameters) of cells were used to identify and exclude debris, dead cells and doublets. All single cells were then selected for GFP expression using the signal from the 488 nm laser line 515/30 nm filters. mCherry signal was identified using the 532 nm laser line and 610/25 nm filter and BFP signal was identified using signal from the 407 nm Laser and 450/50 nm filters. Cells were purified into two populations, which were either GFP+/BFP+/RFP+ or GFP+/BFP+/RFP- for downstream analysis.

Illumina library construction and sequencing

Following FACS sorting, samples were pelleted by centrifugation and subjected to genomic DNA isolation using the QIAamp DNA Micro Kit (56304, Qiagen). To construct the sequencing library, genomic DNA was amplified by two-step PCR. In the first step, Unique Modifier Identifiers (UMI) fused with lentiviral vector integration site (step 1 Fw primer) were mixed with 7i adaptor primer fused with lentiviral vector integration 3'prime integration site (step 1 Rev primer). The mixture was amplified using 5-10 PCR cycles. The second amplification step included a forward primer complementary to the UMI primer fused to 5i (step 2 Fw primer) Illumina adaptor primers and 7i (step 2 Rev primer) and amplified using 25 PCR cycles. DNA concentration were measured using the NEBNext® Library Quant Kit for Illumina (E7630L, NEB). Each 50 µl PCR reaction was composed of 0.5 µM primers, 0.5 µl of Phusion hot-start DNA polymerase (F549S, ThermoFisher) and 2.5 µM dNTPs (N0447S, NEB). After 25 cycles (second PCR step), the PCR products were cleaned using AMPure beads (A63880, Beckman Coulter) according to the manufacturer's protocol.

Fragment size and purity was determined using Agilent TapeStation 2200 and 4200 models, and the desired fragment size of 300 bp was extracted and eluted using a Pippin instrument (Sage Science) with HT 2% Agarose Gel, 100-600 bp (HTC2010). For the Parkin screen, we used 300 v2 Cassettes (15 million reads) on MiSeq (MS-102-2002), whereas, for the TFEB screen, Illumina pair-end sequencing was performed on a NextSeq 550 instrument using a sequencing chip of 300 Mid Output Kit v2.5 (120 million reads, cat 20024905, Illumina). The read length was 200 bp and 7 bp for the indexing primers. Custom sequencing primer were used: (UMI sequence- N, Index sequence- n).

Primer set for step 1:
Fw: 5'-AAGCAGTGGTATCAACGCAGAGTAC**NNNTNNNTNNNTNNNNNNNN**GCACAAAAGGAAACTCACCCT-3'
Rev: 5'-CAAGCAGAAGACGGCATACGAGAT**nnnnnnn**CGACTCGGTGCCACTTTTTC-3'

Primer set for step 2:
Fw: 5'-AATGATACGGCGACCACCGAGATCTACACAAGCAGTGGTATCAACGCAGAGTAC-3'
Rev: 5'-CAAGCAGAAGACGGCATACGAGAT**nnnnnnn**-3'
Sequencing primer: 5'-TTATCAACTTGAAAAAGTGGCACCGAGTCG-3'

UMI extraction and read count generation

The sgRNA abundance analysis was split into four parts: first, the fastq file was demultiplexed according to the run sample sheet using the FASTX Barcode Splitter. Second, using UMI-tools the sequences were extracted and low-quality sequences were trimmed using trimmomatic. Sequences were aligned and mapped to the library data set using Bowtie and Tryhard modules as described previously (Horlbeck et al., 2016). Finally, deduplication grouping and counting were conducting using UMI-tools. The complete Unix based bash file is available on Github.

Differential sgRNA abundance analysis

The differential abundance of sgRNAs between photoactivated-sorted samples and control untreated samples were assessed using the EdgeR package. First, samples were $\log_2$- and count-per-million normalized. Sample variation was determined by covariance-based PCA analysis and read count flooring was established by modeling the noise using coverage as a function of read count. sgRNA enrichment is defined as two standard deviations from the mean of the distribution of non-target-sgRNA controls. For gene aggregation analysis, similarly to a previous paper (Tian et al., 2019), the highest enrichment sgRNA sets were selected by bootstrapping the entire data set. Using EdgeR (Robinson et al., 2010; Dai et al., 2019), the FDR-corrected p-value was calculating by the roast function (Rotation Gene Set Test (Robinson et al., 2010)) following the exactTest function of EdgeR (n = 3 or 4 replicates). Gene set analysis was performed using GSEA 4.0.3 and our whole genome list was ranked according to FC and p-value. The pathway annotation used are the MSigDB Collections (C2:C5) (Reimand et al., 2019).

Experimental approach for validation

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%. 128 sgRNAs targeting 64 high-scoring hits (See also Supplementary Table 2) identified from the primary pooled screen (two sgRNAs per gene) and 2 nontargeting control sgRNAs were individually cloned into the lentiviral mU6-BstXI-BlpI-BFP sgRNA vector (Horlbeck et al., 2016) and confirmed via sequencing.

Nontargeting sgRNA sequences :
Non-targeting control sgRNA 1 - 5’ GCTGCATGGGGCGCGAATCA 3’
Non-targeting control sgRNA 2 - 5’ GTGCACCCGGCTAGGACCGG 3’

To generate virus, 2 x 10^6 Lenti-X 293T cells (Clontech) were seeded in 6-well plates in 1.5 ml DMEM (Life Technologies) supplemented with 10% (v/v) FBS (Gemini Bio Products), 10 mM HEPES (Life Technologies), 1 mM sodium pyruvate (Life Technologies), 1 mM non-essential amino acids (Life Technologies) and 2 mM glutamine (Life Technologies). Cells were transfected the next day in the following manner using Lipofectamine 3000 (Thermofisher): 1.2 µg lentiviral sgRNA plasmid, 0.8 µg psPAX2 packaging vector, 0.3 µg pMD2G packaging vector, 0.8 µg pAdvantage packaging vector, and 5 µl P3000 Reagent were dilute in 150 µl Opti-MEM and incubated 5 minutes at RT; 3.75 µL Lipofectamine 3000 Transfection Reagent (Thermofisher) was diluted into 150 µl Opti-MEM and incubated at room temperature for 5 min, after which the diluted DNA was added, mixed via pipetting, incubated at RT for 40 minutes, then added dropwise to cells. Media was replaced the next day and harvested after two days and centrifuged at 4°C for 10 min at 10,000 x g to pellet cell debris. The supernatant was aliquoted and frozen at -80°C to ensure consistency throughout the validation process.

U2OS cells expressing dCas9 and PA-mCherry were seeded at 20,000 cells per well in 96-well plates on day 0, excluding all exterior wells. On day 1, cells were transduced with virus for 24 hours with 8 µg/ml polybrene at 2 concentrations with 3 replicates per concentration, allowing 10 different viruses, including a control nontargeting sgRNA, to be tested per plate. Cells were checked visually on days 2 and 3 for confluency and blue nuclear signal indicating expression of the sgRNA. If crowded, cells were trypsinized and split to 1-2 96-well plates. Cells were split again on days 4 or 5 as needed into a 96-well imaging plate (Perkin Elmer). A half media change was performed every other day if cells were not being split. On day 7, media was removed, cells were washed 3 times and then left in warm PBS without calcium and magnesium (Thermofisher). Cells were imaged every 60 minutes for 20 hours using a 20X air objective (NA 0.75) on a Nikon Ti-2 CSU-W1 spinning disk system with a photometrics 95B camera operated by Nikon Elements software equipped with temperature regulation and CO_2 control. For every sgRNA, 9 images per well in 3 replicate were acquired. For TFEB translocation response compression, a fixed number of single cell images (n=360) per guide RNA per time per biological repeat were normalized to nontargeted control mean value. To determine if there is a significant difference between the difference values generated for the control replicates on the same plate and the difference values for a guide's replicates on the same plate, we used repeated-measures ANOVA.

Sample size power calculation

To estimate the necessary screening sample size, we conducted power calculations using t-tests of means (pwr.t.test function in R). First the pooled group sample standard deviation was calculated for each sgRNA (Cohen's d): d=sqrt((n1-1)*s1^2+(n2-1)*s2^2)/(n1+n2-2)) where n1 designated the observed number of controls and n2 the observed number of photoactivated and sorted samples. Then, d was passed into the pwr.t.test function with a p-value of 0.05 and increasing power set from 0.2 to 0.8. The pwr.t.test function yields the sample size require for detect a significant deferential abundance.

Shiny AI-PS Application

We created a graphic user interface (GUI) in Shiny (by Rstudio) that performs each step (image segmentation and classification, and creation and testing of model) required to build and test an SVM-based classification model for AI- Photoswitchable Screening. This application can be accessed directly through the website: https://hab-gk-app.shinyapps.io/gk_shiny_app/. Alternatively, the app can be run locally from the source code found at https://github.com/hbaldwin07/GK_shiny_app. Performance is better on local machines than on the network server, so this is the recommended method for those using particularly large datasets or data files (> 10 MB per image). All instructions for running/using the program can be found on the GitHub website.

Author contributions

G.K. led the project and performed most of the experimental work. G.K. conducted and performed all the screens and created the cell lines. K.M. and G.K. created the SVM based code. H.B. and G.K. created the Shiny APP. G.K. created the deployments and CNN code. M.K. and M.W. build the sgRNA libraries. M.K., K.J., M.W. and G.K. conducted NGS analysis and statistics. S.S. and G.K. planed and preformed the secondary analysis validation screen. G.K., S.S. and R.Y. wrote the paper. G.K., R.Y. and J.L.S. designed the study. All authors discussed the results and commented on the manuscript. R.Y. and J.L.S. supervised the project.

Acknowledgments

We thank Nico Tjandra for intellectual contributions. We thank Nick Ader, Eric Bunker, Elyssa Hawk and Sue Smith for helping with cloning, cell lines and Lentivirus production. We thank Catherine Nezich, Hetal Shah, Jose Norbert Vargas and Benoit Kornmann for comments on the manuscript and all the Youle lab and the Lippincott-Schwartz lab members for critical comments. We thank Talya Chooly for supporting the project. Flow cytometry cell sorting and sample isolation was performed at the Flow Cytometry Core, NHLBI. Next generation deep sequencing was performed at the CCR Genomics Core, NCI. We thank the NIH based Nikon team for helping with imaging and integration of our external codes. This work used the computational resources of the NIH HPC Biowulf cluster (https://hpc.nih.gov). This work was supported by the NINDS intramural program. M.K. was supported by NIH/NIGMS grant DP2 GM119139.

Competing interests

The authors declare no competing interests

Fig. 1

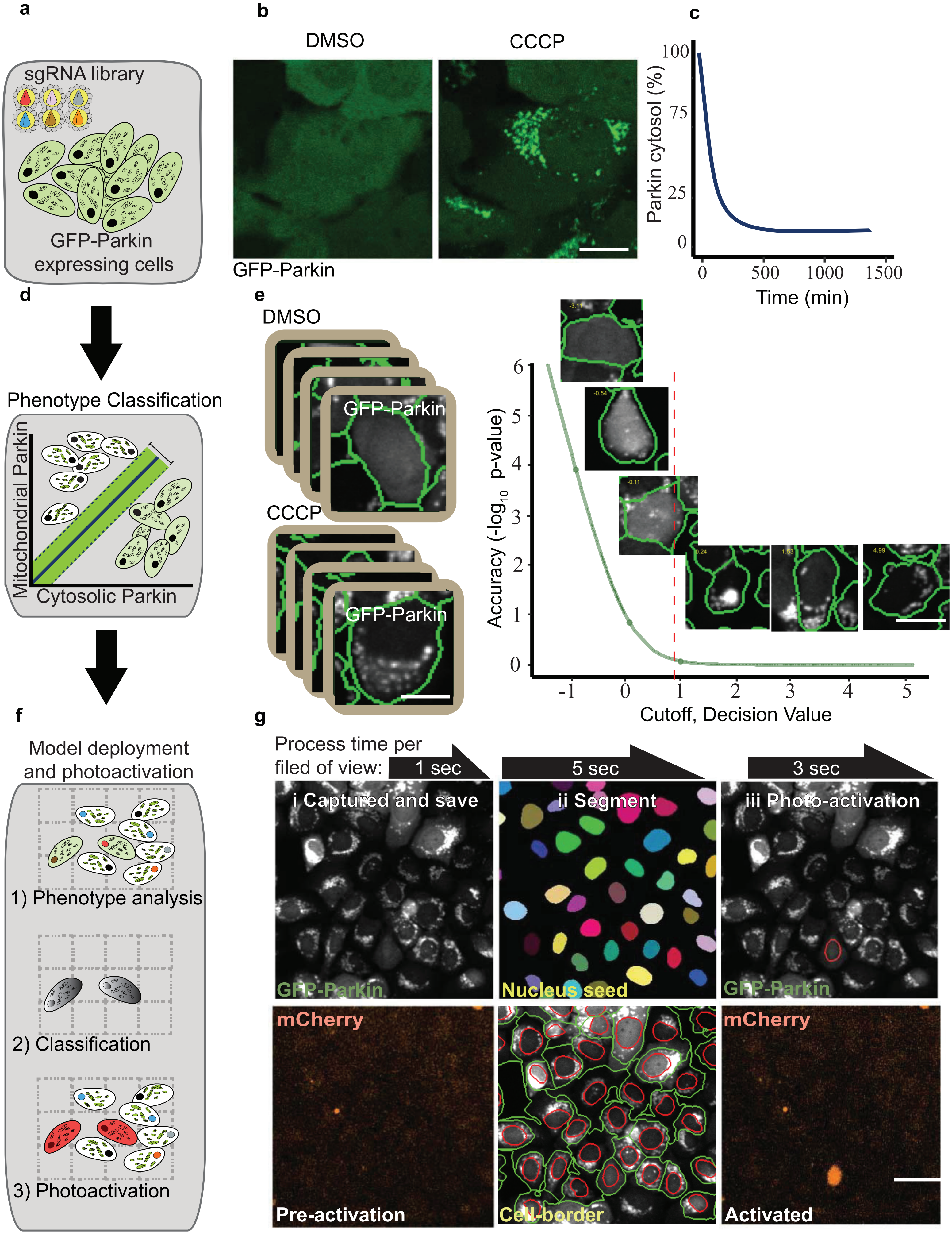
a
sgRNA library
GFP-Parkin
expressing cells
b
DMSO
CCCP
GFP-Parkin
c
Parkin cytosol (%)
100
75
25
0
0
500
1000
1500
Time (min)
d
Phenotype Classification
Mitochondrial Parkin
Cytosolic Parkin
e
DMSO
GFP-Parkin
CCCP
GFP-Parkin
Accuracy (-log10 p-value)
6
5
4
3
2
1
0
-1
0
1
2
3
4
5
Cutoff, Decision Value
-3.11
-0.54
-0.11
0.24
1.53
4.99
f
Model deployment
and photoactivation
1) Phenotype analysis
2) Classification
3) Photoactivation
g
Process time per
filed of view:
1 sec
5 sec
3 sec
i Captured and save
ii Segment
iii Photo-activation
GFP-Parkin
Nucleus seed
GFP-Parkin
mCherry
mCherry
Pre-activation
Cell-border
Activated

Figure 1. **Machine learning genetic screening platform for Parkin localization - proof of principle screen (a)** Pooled sgRNAs infecting target cells. **(b)** GFP-Parkin U2OS cells treated with DMSO or CCCP (2 hr). Scale bar, 5 µm . **(c)** Percent of cytosolic Parkin over time after CCCP 10 µM treatment supplemented with 100 nM Bafilomycin A. **(d, e)** Single-cell image examples for the training set of phenotypes used for SVM classification. Cutoff determined by classification model accuracy (y axis). Cutoff used in this screen is a prediction score of 0.8 (also known as decision value), scale bar, 5 µm **(f, g)** Representative Field Of View (FOV) of GFP-Parkin U2OS cell screening procedure. i, Images were captured and saved on local computer. ii, Cells borders were identified (green circle surrounding cell border, red circle nucleus) following Nucleus segmentation (color labeled nuclei). SVM classification model was deployed and masked (red circle). iii, photoactivation of the SMV identified cell; scale bar, 10 µm

Figure 2

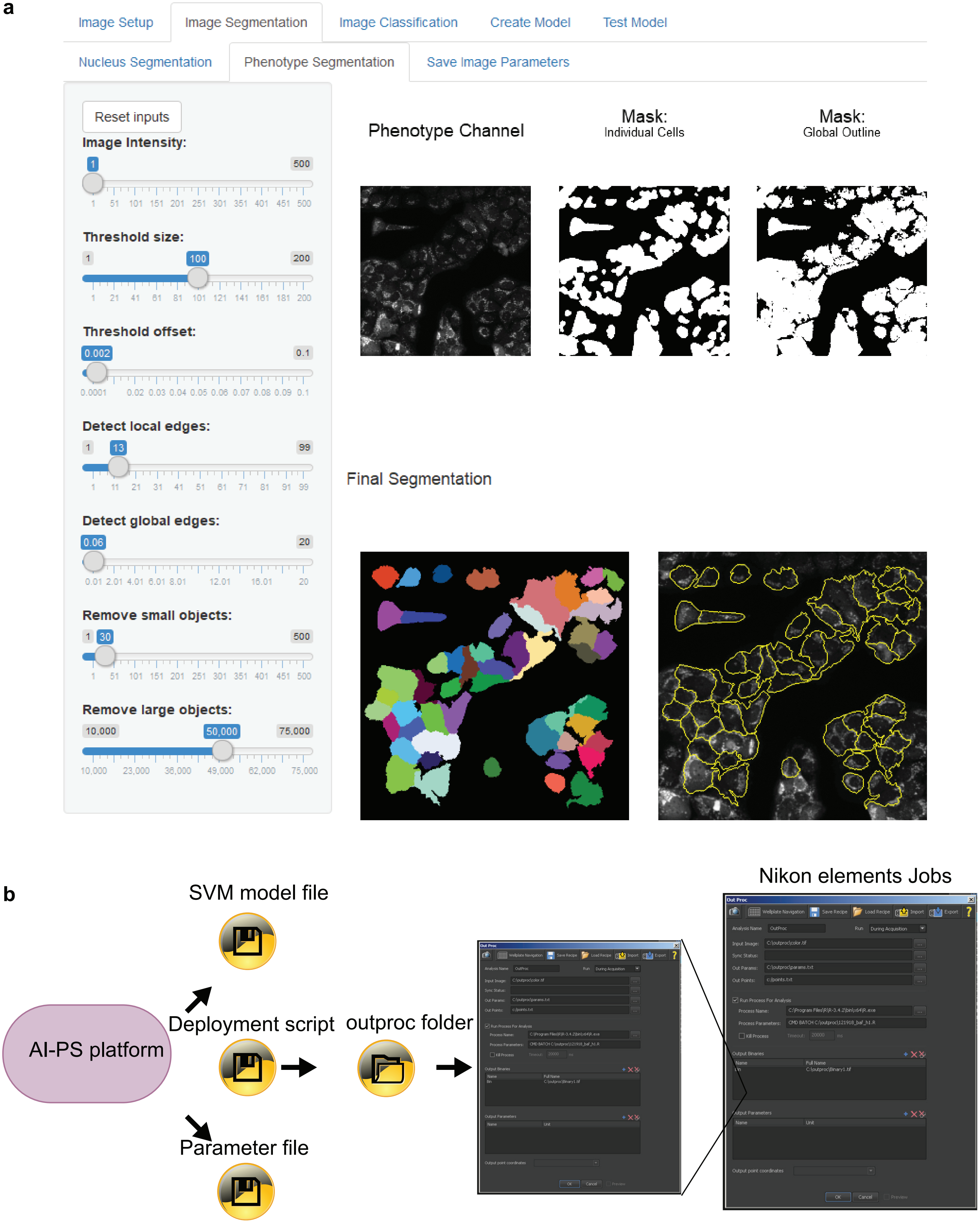

Image Intensity: 1 (1 – 500)

Threshold size: 100 (1 – 200)

Threshold offset: 0.002 (0.0001 – 0.1)

Detect local edges: 13 (1 – 99)

Detect global edges: 0.06 (0.01 – 20)

Remove small objects: 1 – 30 (1 – 500)

Remove large objects: 10,000 – 50,000 (10,000 – 75,000)

Out Proc

Analysis Name: OutProc   Run: During Acquisition

Input Image: C:\outproc\color.tif

Sync Status:

Out Params: C:\outproc\params.txt

Out Points: c:/points.txt

☑ Run Process For Analysis

Process Name: C:\Program Files\R\R-3.4.2\bin\x64\R.exe

Process Parameters: CMD BATCH C:\outproc\121918_baf_h1.R

☐ Kill Process   Timeout: 20000 ms

Output Binaries

Name	Full Name
bin	C:\outproc\Binary1.tif

Output Parameters

Name	Unit

Output point coordinates

OK   Cancel   ☐ Preview

Figure. 2 **Shiny AI-PS Application and output files.** **(a)** AI-PS GUI interface example. https://hab-gk-app.shinyapps.io/gk_shiny_app/ **(b)** Three deployment output files integrated with Nikon elements software.

Figure. 3

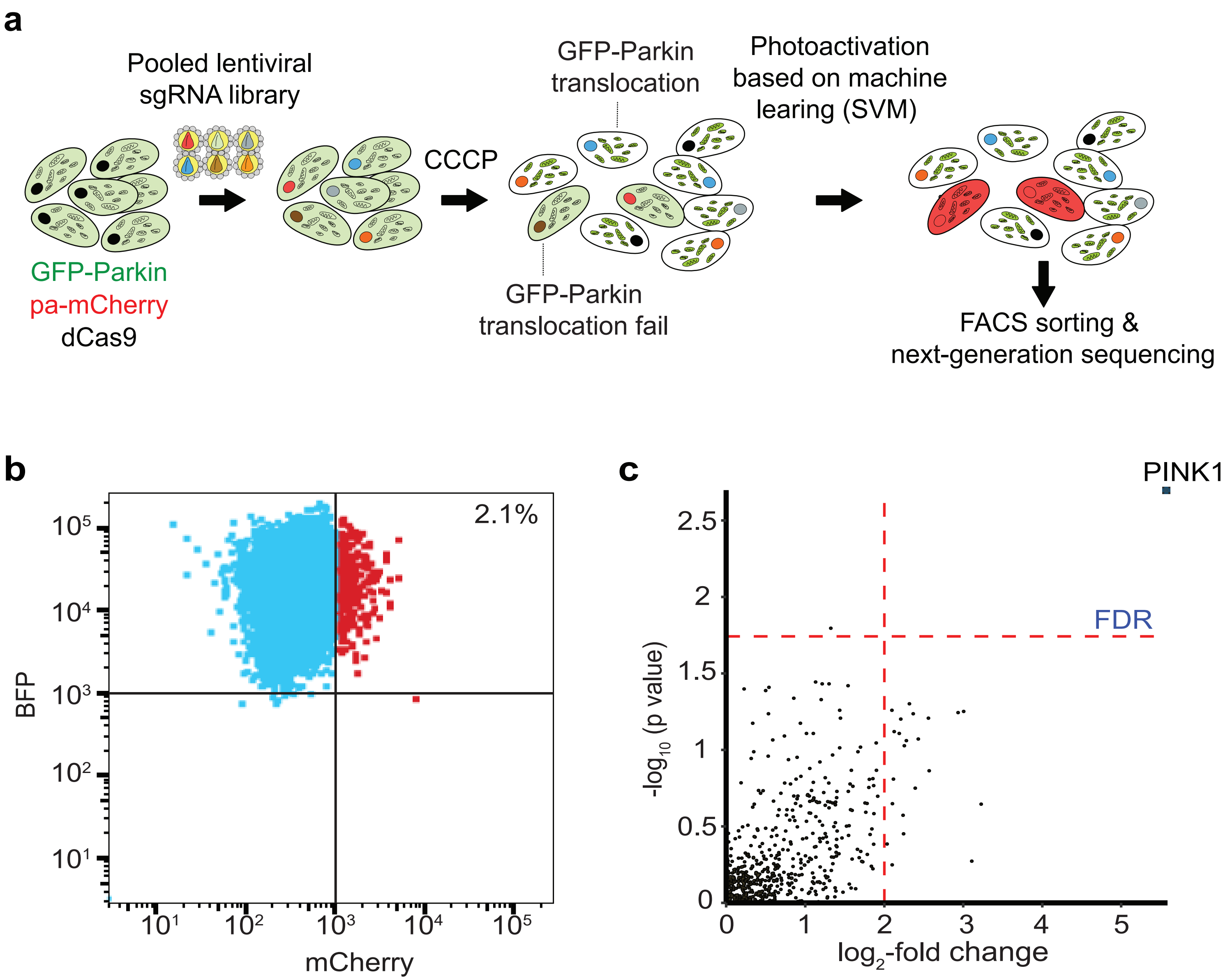
a
Pooled lentiviral sgRNA library
GFP-Parkin
pa-mCherry
dCas9
CCCP
GFP-Parkin translocation
GFP-Parkin translocation fail
Photoactivation based on machine learing (SVM)
FACS sorting & next-generation sequencing
b
2.1%
BFP
10^5
10^4
10^3
10^2
10^1
mCherry
c
PINK1
FDR
-log$_{10}$ (p value)
2.5
2
1.5
1
0.5
0
1
2
3
4
5
log$_2$-fold change

Figure 3. **Validation of the platform with a Parkin localization screen**, targeting Kinases, Phosphatases, and Drug Targets - gRNA pooled library (12,500 sgRNAs targeting 2774 genes). **(a)** Schematic representation of AI-PS platform **(b)** Flow cytometry scatterplot representing the separation of the post-screen photoactivated (mCherry florescence signal x-axis in red) from the inactivated cell population (BFP florescence signal y-axis in cyan). **(c)** Enrichment plot comparing sgRNA abundance in the photoactivated sample follow CCCP treatment to sgRNA abundance prior to treatment. Vertical red line set on $\log_2$-fold change threshold; horizontal red line indicating the Benjamini-Hochberg corrected p-value set on 5%. See also Table S1.

Figure 4

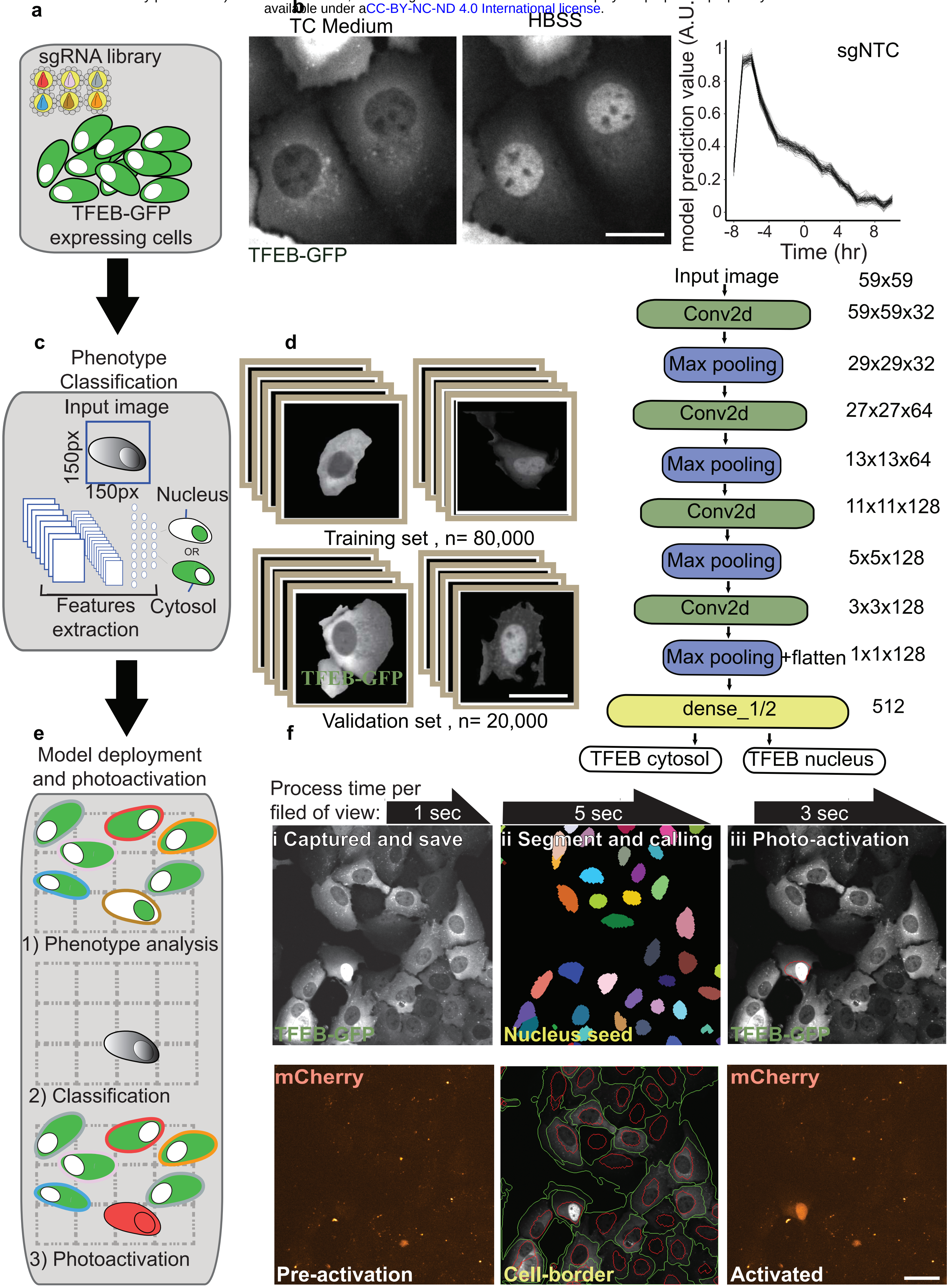
a
sgRNA library
TFEB-GFP
expressing cells
b
TC Medium
HBSS
TFEB-GFP
model prediction value (A.U.)
1
0.8
0.6
0.4
0.2
0
-8
-4
0
4
8
Time (hr)
sgNTC
c
Phenotype
Classification
Input image
150px
150px
Nucleus
OR
Features
extraction
Cytosol
d
Training set , n= 80,000
TFEB-GFP
Validation set , n= 20,000
Input image
59x59
Conv2d
59x59x32
Max pooling
29x29x32
Conv2d
27x27x64
Max pooling
13x13x64
Conv2d
11x11x128
Max pooling
5x5x128
Conv2d
3x3x128
Max pooling
+flatten
1x1x128
dense_1/2
512
TFEB cytosol
TFEB nucleus
e
Model deployment
and photoactivation
1) Phenotype analysis
2) Classification
3) Photoactivation
f
Process time per
filed of view:
1 sec
5 sec
3 sec
i Captured and save
ii Segment and calling
iii Photo-activation
TFEB-GFP
Nucleus seed
TFEB-GFP
mCherry
mCherry
Pre-activation
Cell-border
Activated

Figure 4. **Deep learning genetic screening platform for TFEB localization. (a)** Pooled sgRNAs infecting target cells. **(b)** Nuclear TFEB-GFP in cells treated with HBSS (1 hr) or cytosolic in cells in complete medium (cm); scale bar, 5 μm. **(c, d)** Learning set composed of 100,000 single-cell images was used for convolutional neural network (CNN) classification. ImageNet like CNN architecture is composed of four sets of convoluted processes followed by the max pooling procedure. The phenotype decision is based on probability value. Low probability value is assigned to cells with cytosolic TFEB-GFP and high probability values for cells containing nuclear TFEB-GFP. **(e)** Illustration of optical decision making coupled with photoactivation. **(f)** Representative Field Of View (FOV) of the TFEB-GFP U2OS cell screening procedure. Images were captured and saved on a local computer. i. Cell borders were identified (green circle surrounding cell border, red circle nucleus) following ii nuclear segmentation (color labeled nuclei). iii CNN classification model was deployed and masked (red circle) and photoactivation of the CNN identified cell; scale bar, 10 μm.

Figure. 5

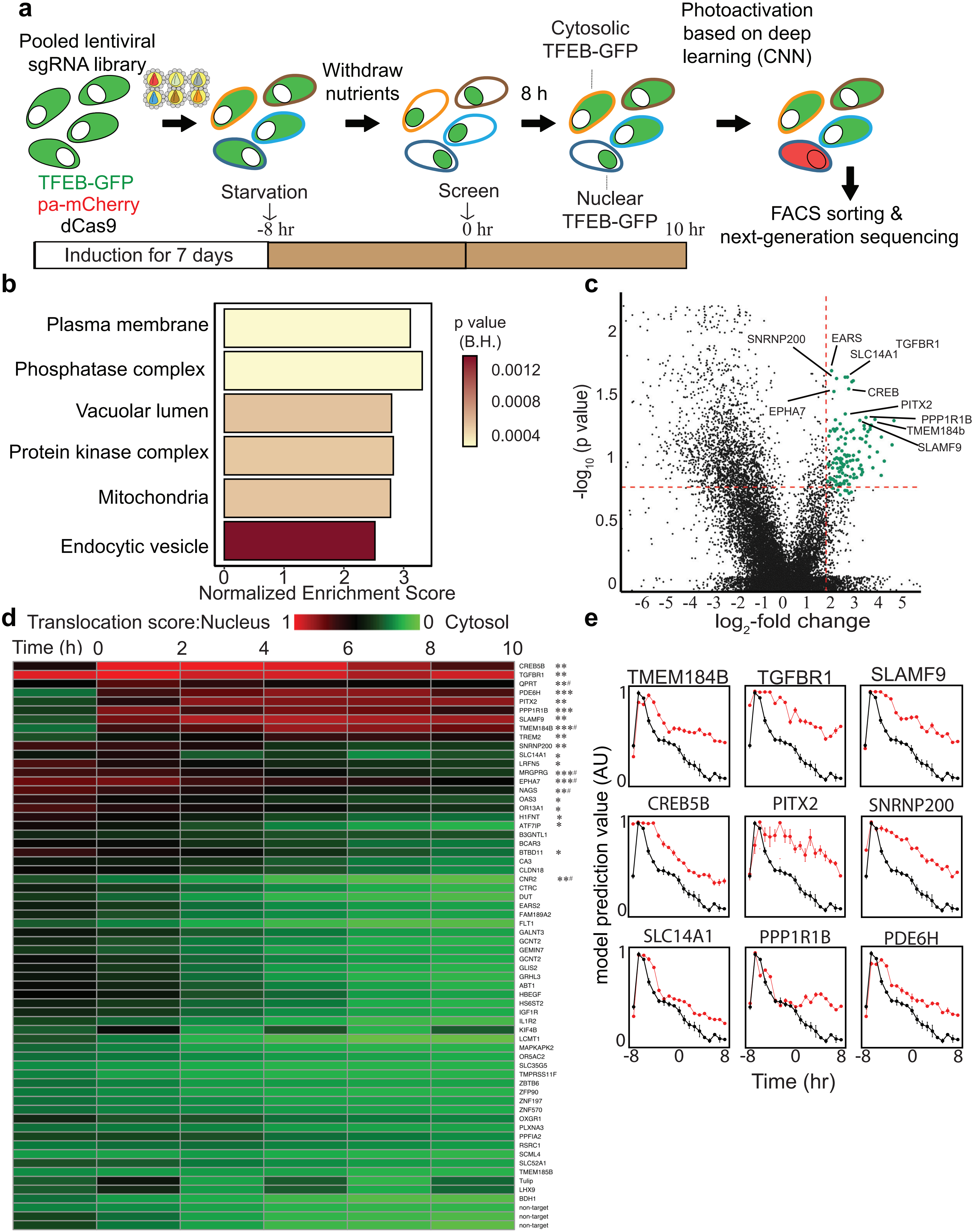
a
Pooled lentiviral sgRNA library
TFEB-GFP
pa-mCherry
dCas9
Withdraw nutrients
Starvation
-8 hr
Screen
0 hr
8 h
Cytosolic TFEB-GFP
Nuclear TFEB-GFP
10 hr
Photoactivation based on deep learning (CNN)
FACS sorting & next-generation sequencing
Induction for 7 days
b
Plasma membrane
Phosphatase complex
Vacuolar lumen
Protein kinase complex
Mitochondria
Endocytic vesicle
0
1
2
3
Normalized Enrichment Score
p value (B.H.)
0.0012
0.0008
0.0004
c
SNRNP200
EARS
TGFBR1
SLC14A1
CREB
PITX2
EPHA7
PPP1R1B
TMEM184b
SLAMF9
$-\log_{10}$ (p value)
0
0.5
1
1.5
2
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5
$\log_2$-fold change
d
Translocation score:Nucleus 1
0 Cytosol
Time (h) 0 2 4 6 8 10
CREB5B **
TGFBR1 **
QPRT **#
PDE6H ***
PITX2 **
PPP1R1B ***
SLAMF9 **
TMEM184B ***#
TREM2 **
SNRNP200 **
SLC14A1 *
LRFN5 *
MRGPRG ***#
EPHA7 ***#
NAGS **#
OAS3 *
OR13A1 *
H1FNT *
ATF7IP *
B3GNTL1
BCAR3
BTBD11 *
CA3
CLDN18
CNR2 **#
CTRC
DUT
EARS2
FAM189A2
FLT1
GALNT3
GCNT2
GEMIN7
GCNT2
GLIS2
GRHL3
ABT1
HBEGF
HS6ST2
IGF1R
IL1R2
KIF4B
LCMT1
MAPKAPK2
OR5AC2
SLC35G5
TMPRSS11F
ZBTB6
ZFP90
ZNF197
ZNF570
OXGR1
PLXNA3
PPFIA2
RSRC1
SCML4
SLC52A1
TMEM185B
Tulip
LHX9
BDH1
non-target
non-target
non-target
e
TMEM184B
TGFBR1
SLAMF9
CREB5B
PITX2
SNRNP200
SLC14A1
PPP1R1B
PDE6H
model prediction value (AU)
1
0
-8 0 8
Time (hr)

Figure 5. **Whole-genome TFEB-GFP localization screen. (a)** Screen workflow, TFEB-GFP cells were transduced with pooled sgRNA libraries targeting the whole genome for seven days. Following eight hours of starvation, AI-PS screening platform was initiated. **(b)** GSEA pathway analysis annotated using the gene sets derived from the GO Cellular Component database of the Molecular Signatures Database. On the x-axis is the GSEA normalised enrichment score and the color of the bars represents the GSEA calculated FDR probabilities. **(c)** Volcano sgRNA plot comparing sgRNA abundance in the photoactivated samples following eight hours of starvation to sgRNA abundance in the unsorted sample. Vertical red line is set at the $\log_2$-fold change threshold; horizontal red line is set at the Benjamini-Hochberg corrected p-value of 15%. Genes selected for secondary screening are shown in green. See also Supplementary Table 2. **(d)** Top candidates from the primary screen were selected for secondary screening. The TFEB-GFP localization CNN model probability value was used for measuring the perturbation effect on TFEB-GFP localization over time during starvation. Heat map including all genes included in secondary screen; low probability values are shown in green for cytosolic TFEB-GFP, high probability values (red) for nuclear TFEB-GFP. n=3, *p < 0.05, **p < 0.01, or ***p < 0.0001 obtained using repeated measures Anova test. p value, one sgRNA was significant. e. TFEB-GFP translocation dynamics observed during starvation for selected gene candidates, non-targeted sgRNA in black, gRNA target the designated protein in red. Quantification is displayed as mean ± s.e.m. from 3 independent experiments.

Figure. 6

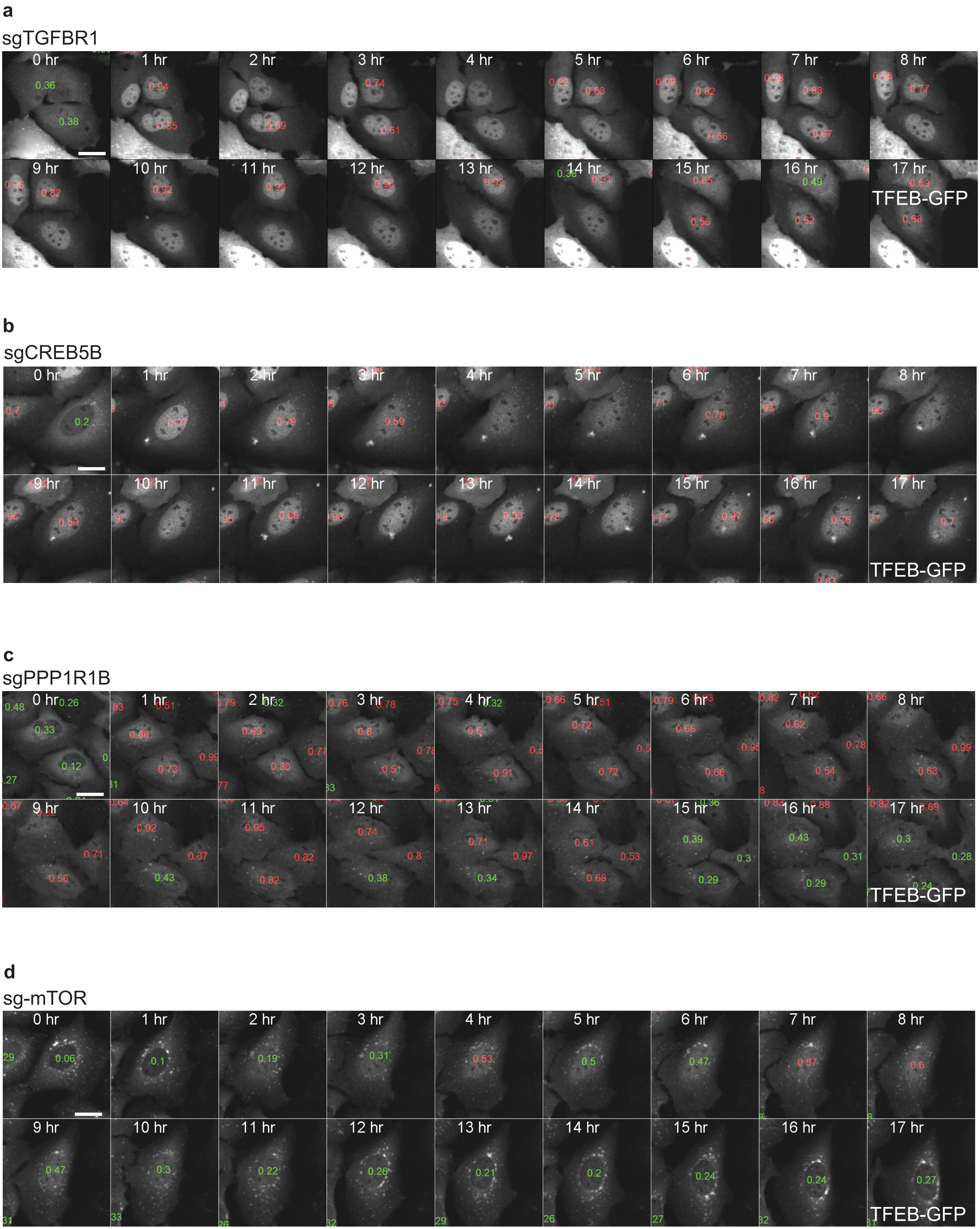
a
sgTGFBR1
0 hr
0.36
0.38
1 hr
0.94
0.55
2 hr
0.59
3 hr
0.74
0.61
4 hr
5 hr
0.92
0.83
6 hr
0.98
0.82
0.56
7 hr
0.98
0.88
0.67
8 hr
0.98
0.77
9 hr
0.96
0.82
10 hr
0.92
11 hr
0.93
12 hr
0.96
13 hr
0.92
0.38
14 hr
0.91
15 hr
0.65
0.56
16 hr
0.49
0.62
17 hr
0.59
TFEB-GFP
0.68
b
sgCREB5B
0 hr
0.7
0.2
1 hr
0.97
2 hr
0.79
3 hr
0.59
4 hr
5 hr
6 hr
0.78
7 hr
0.9
8 hr
9 hr
0.59
10 hr
11 hr
0.86
12 hr
13 hr
0.93
14 hr
15 hr
0.87
16 hr
0.75
17 hr
0.7
TFEB-GFP
c
sgPPP1R1B
0.48
0 hr
0.26
0.33
0.12
1 hr
0.51
0.86
0.99
0.73
0.79
2 hr
0.32
0.69
0.77
0.85
0.76
3 hr
0.78
0.8
0.78
0.51
0.75
4 hr
0.32
0.6
0.91
0.66
5 hr
0.51
0.72
0.72
0.79
6 hr
0.68
0.66
0.82
7 hr
0.62
0.62
0.78
0.54
0.66
8 hr
0.99
0.63
0.57
9 hr
0.71
0.56
10 hr
0.92
0.87
0.43
11 hr
0.95
0.82
0.82
12 hr
0.74
0.8
0.38
13 hr
0.71
0.97
0.34
14 hr
0.61
0.53
0.68
15 hr
0.36
0.39
0.3
0.29
0.83
16 hr
0.88
0.43
0.31
0.29
0.82
17 hr
0.69
0.3
0.28
0.24
TFEB-GFP
d
sg-mTOR
0 hr
0.06
1 hr
0.1
2 hr
0.19
3 hr
0.31
4 hr
0.53
5 hr
0.5
6 hr
0.47
7 hr
0.57
8 hr
0.6
9 hr
0.47
10 hr
0.3
11 hr
0.22
12 hr
0.26
13 hr
0.21
14 hr
0.2
15 hr
0.24
16 hr
0.24
17 hr
0.27
TFEB-GFP

Figure 6. **Selected sgRNA targeting genes validated in the TFEB-GFP secondary screening.** low probability values are shown in green for cytosolic TFEB-GFP, high probability values (red) for nuclear TFEB-GFP. TFEB-GFP-expressing U2OS cells treated with the designated sgRNA were starved in HBSS for 18 hours, and images acquired every hour. scale bar 5 µm.

Figure S1.

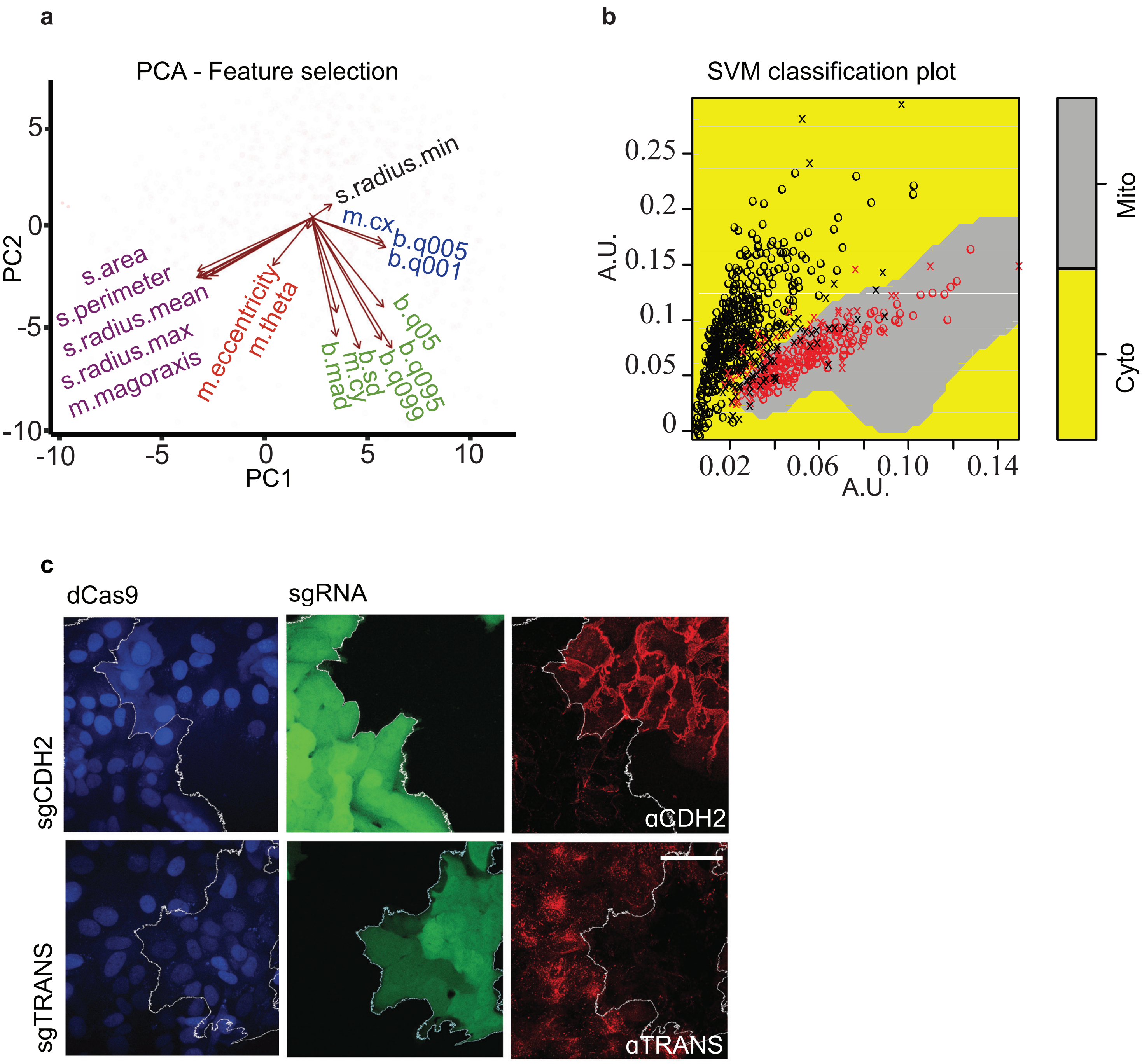
a
PCA - Feature selection
PC2
PC1
5
0
-5
-10
-10
-5
0
5
10
s.radius.min
m.cx
b.q005
b.q001
s.area
s.perimeter
s.radius.mean
s.radius.max
m.magoraxis
m.eccentricity
m.theta
b.q05
b.q095
b.q099
b.sd
m.cy
b.mad
b
SVM classification plot
A.U.
A.U.
0.25
0.2
0.15
0.1
0.05
0
0.02
0.06
0.10
0.14
Mito
Cyto
c
dCas9
sgRNA
sgCDH2
sgTRANS
αCDH2
αTRANS

Figure S1. **Support vector machine classification build for Parkin screen. (a)** PCA analysis of 18 feature-predictors calculated using the R function computeFeatures from the EBImage library. **(b)** 2D representation of non-linear hyperplane separation of mitochondrial Parkin phenotype vs. cytosolic phenotype. Mitochondrial Parkin predicted cells in red, Cytosolic Parkin in black. The variable o for correct classification, and x for misclassification. **(c)** dCas9-expressing U2OS cells treated with sgRNA targeting either TRANS or CDH2 and immunostained using TRANS or CDH2 antibodies, scale bar 20 µm.

Figure S2.

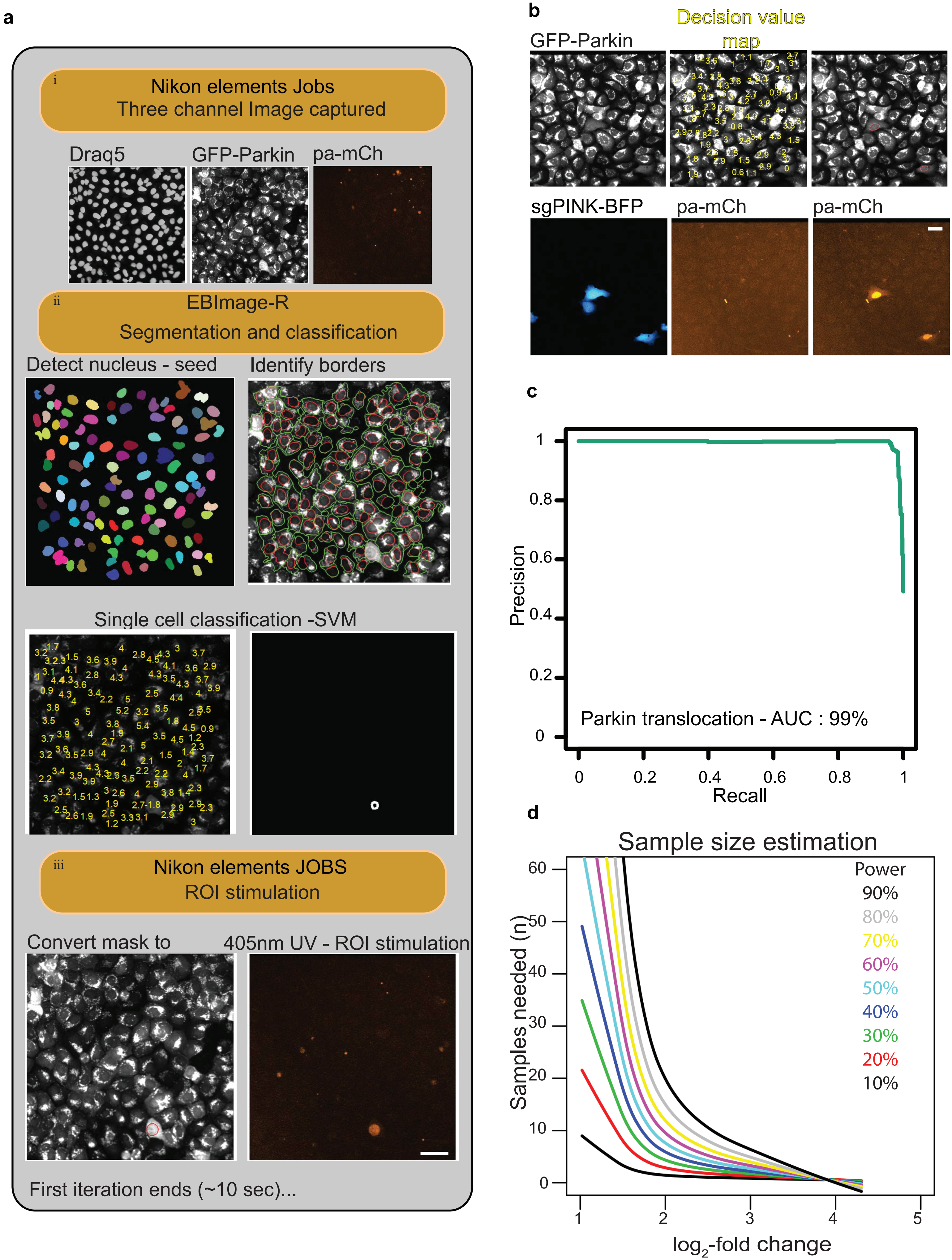
a
i
Nikon elements Jobs
Three channel Image captured
Draq5
GFP-Parkin
pa-mCh
ii
EBImage-R
Segmentation and classification
Detect nucleus - seed
Identify borders
Single cell classification -SVM
iii
Nikon elements JOBS
ROI stimulation
Convert mask to
405nm UV - ROI stimulation
First iteration ends (~10 sec)...
b
Decision value map
GFP-Parkin
sgPINK-BFP
pa-mCh
pa-mCh
c
Precision
Recall
0
0.2
0.4
0.6
0.8
1
Parkin translocation - AUC : 99%
d
Sample size estimation
Samples needed (n)
0
10
20
30
40
50
60
log$_2$-fold change
1
2
3
4
5
Power
90%
80%
70%
60%
50%
40%
30%
20%
10%

Figure S2. **AI-PS processing stages and model assessments. (a)** AI-PS procedure iteration on-the-fly. i. Three-channel images are acquired and available for Image processing and analysis. ii. The Draq5 channel is used for nuclear detection and the GFP-Parkin image is used to identify cell borders (right panel, red circle masks nucleus and the green circle masks the cell borders). Post-training single-cell phenotypes are scored by the SVM model (bottom left panel, decision value in yellow), where the mask is marking the predicted cells (bottom right panel, white circle). iii. Mask images are uploaded back to Nikon Elements software and converted to ROI (left panel, red circle). The ROI is stimulated by 405 nm laser for 50 ms with 100% laser power (right panel, 568 nm detector). **(b)** For proof-of-principle screening, the cell population expressing GFP-Parkin contains 10% of cells infected with sgRNA targeting Pink1 (sgRNA in blue for BFP) and treated with 1 μM CCCP; upper panel, representative GFP-Parkin images; middle panel, single-cell phenotype prediction score (decision value in yellow), score < 0.8 for cytosolic GFP-Parkin, score > 0.8 for mitochondrial GFP-Parkin; right panel, SVM-called positive cells shown on 568 nm emission detector. **(c)** Precision-Recall curve from 9,546 single cell images obtained during 12 hours live-image acquisition with 1-min intervals with 1 μM CCCP treatment supplemented with 100 nM Bafilomycin A. The accuracy is computed from the integral area under the Precision-Recall curve (AUC, Area Under the Curve). **(d)** The sample size was estimated for observing a significant effect for a $\log_2$-fold change of 1.5. The $\log_2$-fold change was modeled based on the non-targeting negative control distribution. The sample size was estimated from the GFP-Parkin screen. The effect size is estimated by the $\log_2$-fold change of the photoactivated sample vs. the control sample. The estimation is made using data gathered from 4 biological repeats of the GFP-Parkin screen. For more details see Methods.

Figure S3.

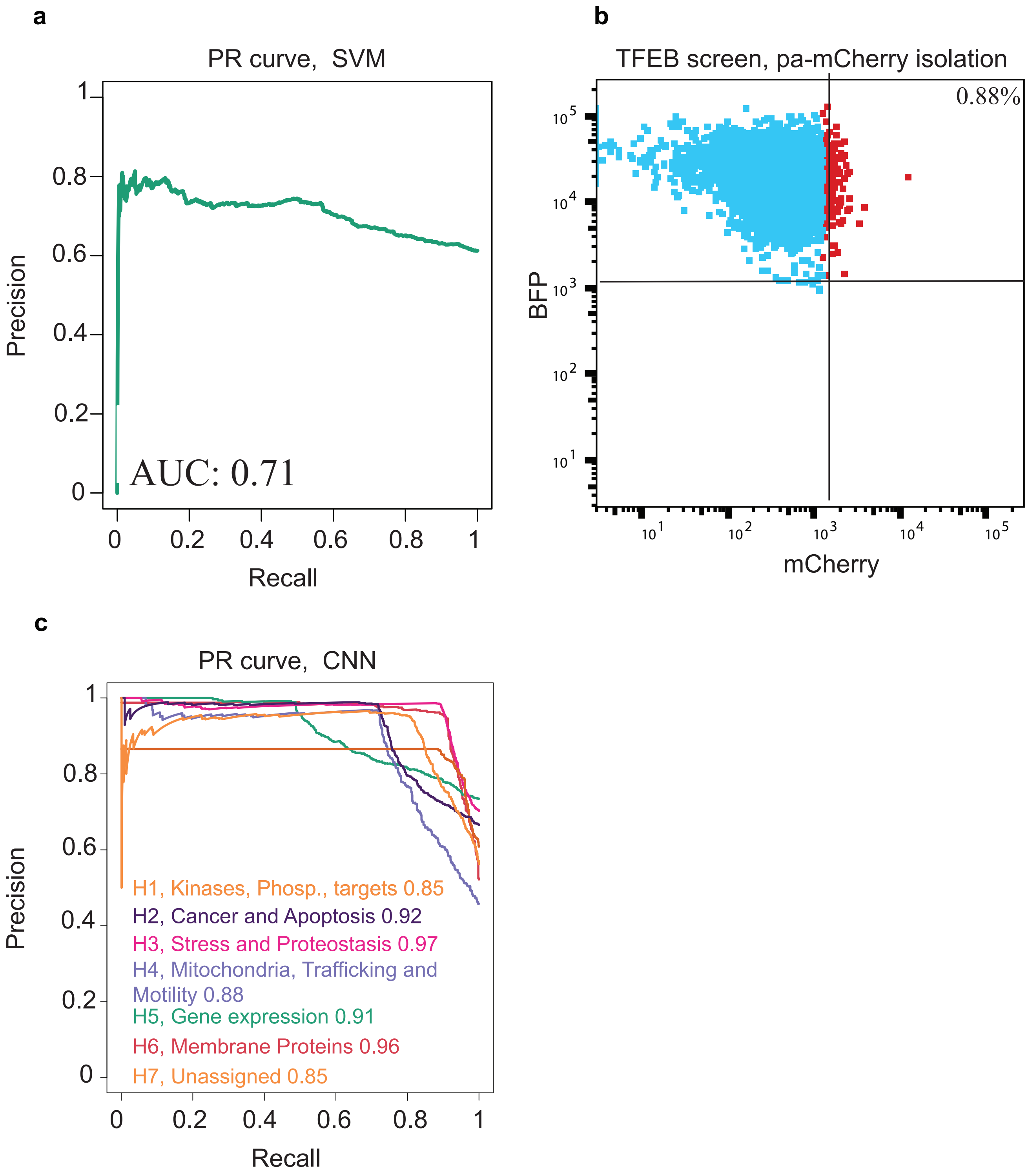
a
PR curve, SVM
Precision
Recall
0
0.2
0.4
0.6
0.8
1
AUC: 0.71
b
TFEB screen, pa-mCherry isolation
0.88%
BFP
mCherry
10^1
10^2
10^3
10^4
10^5
c
PR curve, CNN
Precision
Recall
H1, Kinases, Phosp., targets 0.85
H2, Cancer and Apoptosis 0.92
H3, Stress and Proteostasis 0.97
H4, Mitochondria, Trafficking and Motility 0.88
H5, Gene expression 0.91
H6, Membrane Proteins 0.96
H7, Unassigned 0.85

Figure S3. **TFEB-GFP screen performance.** **(a)** TFEB-GFP phenotype classification performance by SVM, Precision-Recall curve from 7,848 single cell images obtained from HBSS starved cells. Images collection began 8 hours after starvation initiated and continued for another 10 hours. Accuracy is computed from the integral area under the Precision-Recall Curve (AUC, Area Under the Curve). **(b)** Flow cytometry scatterplot representing the separation of the post-screening photoactivated mCherry signal (x-axis in red) from the inactivated cell population (BFP fluorescence signal, y-axis in cyan). **(c)** 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.

Figure S4.

sgNTC

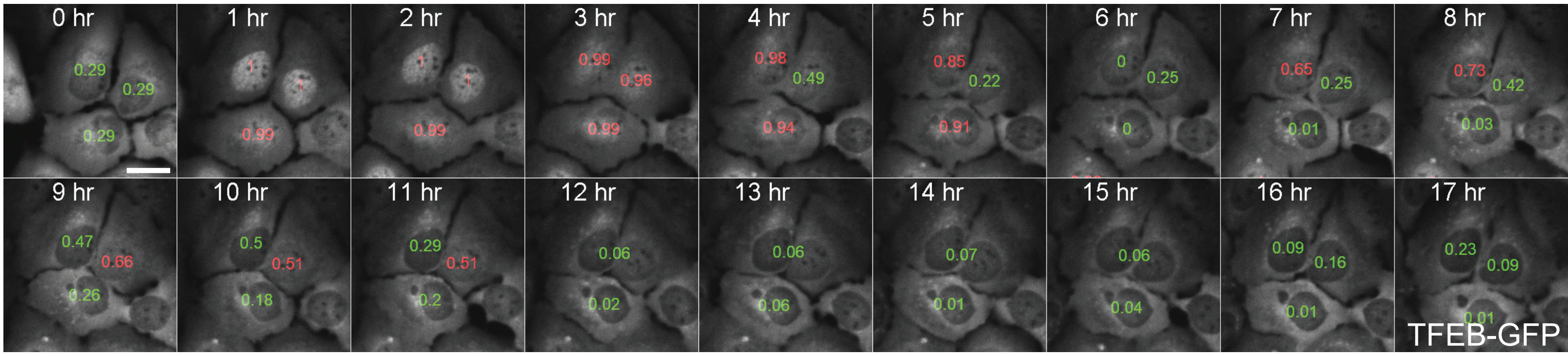

Figure S4. **Live-cell images of TFEB-GFP U2OS cells under starvation conditions.** TFEB-GFP-expressing U2OS cells were starved for 18 hours and images acquired every hour. low probability values are shown in green for cytosolic TFEB-GFP, high probability values (red) for nuclear TFEB-GFP. sgNTC, nontargeting sgRNA control, scale bar 5 µm

Figure S5

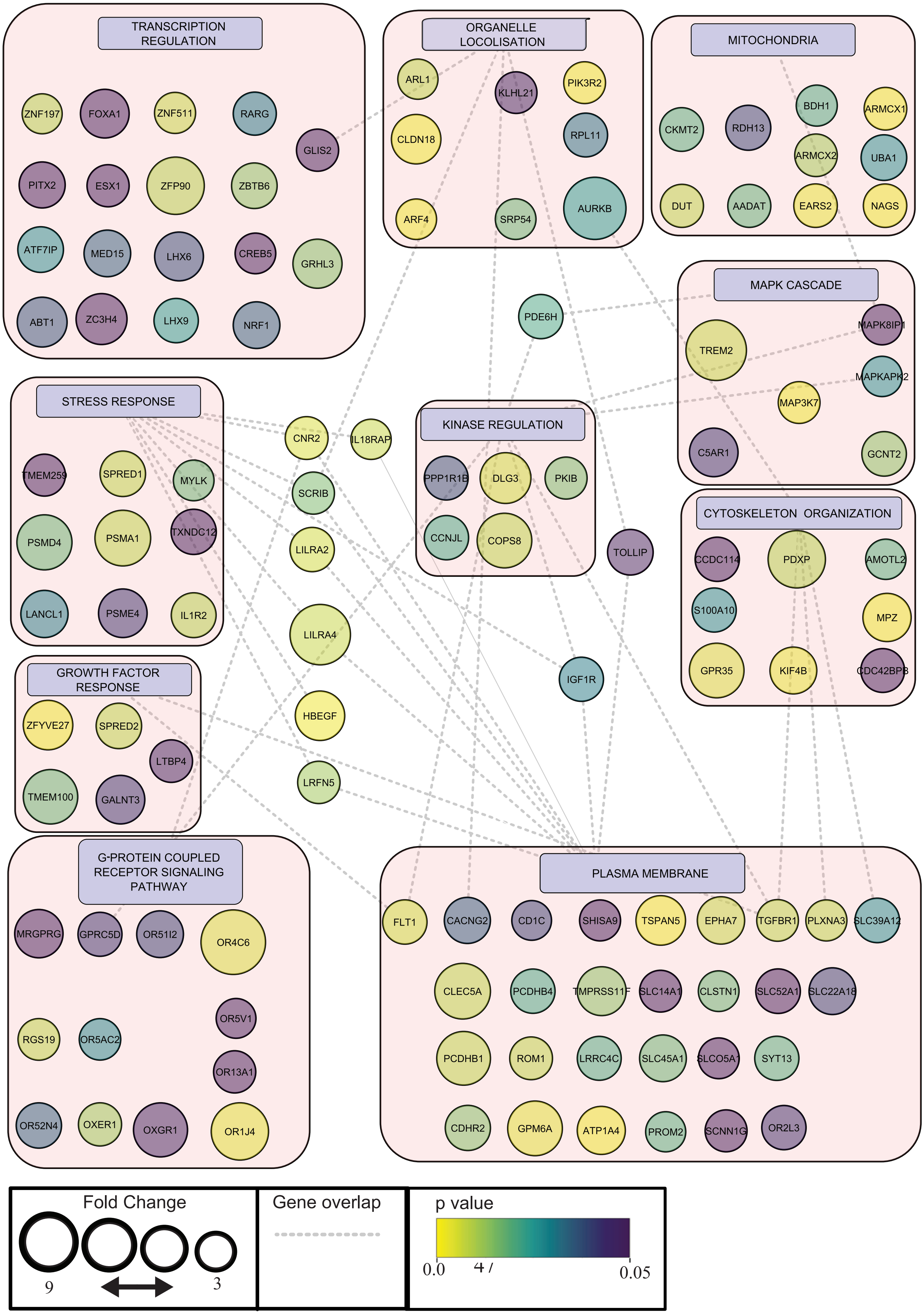
TRANSCRIPTION REGULATION
ZNF197
FOXA1
ZNF511
RARG
GLIS2
PITX2
ESX1
ZFP90
ZBTB6
ATF7IP
MED15
LHX6
CREB5
GRHL3
ABT1
ZC3H4
LHX9
NRF1
ORGANELLE LOCOLISATION
ARL1
KLHL21
PIK3R2
CLDN18
RPL11
ARF4
SRP54
AURKB
MITOCHONDRIA
BDH1
ARMCX1
CKMT2
RDH13
ARMCX2
UBA1
DUT
AADAT
EARS2
NAGS
MAPK CASCADE
PDE6H
MAPK8IP1
TREM2
MAPKAPK2
MAP3K7
C5AR1
GCNT2
STRESS RESPONSE
CNR2
IL18RAP
TMEM259
SPRED1
MYLK
SCRIB
PSMD4
PSMA1
TXNDC12
LILRA2
LANCL1
PSME4
IL1R2
LILRA4
KINASE REGULATION
PPP1R1B
DLG3
PKIB
CCNJL
COPS8
TOLLIP
CYTOSKELETON ORGANIZATION
CCDC114
PDXP
AMOTL2
S100A10
MPZ
GPR35
KIF4B
CDC42BPB
IGF1R
GROWTH FACTOR RESPONSE
ZFYVE27
SPRED2
LTBP4
TMEM100
GALNT3
HBEGF
LRFN5
G-PROTEIN COUPLED RECEPTOR SIGNALING PATHWAY
MRGPRG
GPRC5D
OR51I2
OR4C6
RGS19
OR5AC2
OR5V1
OR13A1
OR52N4
OXER1
OXGR1
OR1J4
PLASMA MEMBRANE
FLT1
CACNG2
CD1C
SHISA9
TSPAN5
EPHA7
TGFBR1
PLXNA3
SLC39A12
CLEC5A
PCDHB4
TMPRSS11F
SLC14A1
CLSTN1
SLC52A1
SLC22A18
PCDHB1
ROM1
LRRC4C
SLC45A1
SLCO5A1
SYT13
CDHR2
GPM6A
ATP1A4
PROM2
SCNN1G
OR2L3
Fold Change
9
3
Gene overlap
p value
0.0
4 /
0.05

Figure S5. **Gene-set clustering of 64 candidates enriched in TFEB-GFP translocation screen.** Cytoscape analysis of enriched genes, the circle size represents fold-change enrichment and p-value is color-coded inside the circle. The dashed lines indicate cluster overlap.

Video 1 | **Example of AI-PS platform Parkin screen proof of principle**, Nikon NIS elements JOB module screen capture during GFP-Parkin (in Green) acquisition. Machine learning deployed for automatic detection of sgRNA targeting PINK-1 (in blue) according to cell phenotype (red circle on top of GFP-Parkin Image). The detected cell is photoactivated (in yellow-red).

Video 2 | **Live-cell images of TFEB-GFP U2OS cells under starvation conditions.** TFEB-GFP-expressing U2OS cells were starved for 18 hours and images acquired every hour. Single-cell CNN prediction scores are marked in red for nuclear TFEB and in green for cytosolic TFEB. Dynamic Bar chart indicates the cumulative distribution of TFEB translocation in the represented cell population.

Video 3 | **Example of AI-PS platform for TFEB screen**, Nikon NIS elements JOB module screen capture during TFEB-GFP (in green) acquisition. Top Image TFEB-GFP in green, red circle for the phenotype automatic detected cells. Bottom left corner, R based image segmentation, machine learning prediction and Mask generation. Bottom right, photoactivation in red (pa-mCh). Three examples are shown here.

Video 4 | **Live-cell images of TFEB-GFP U2OS cells expressing sgRNA targeting TGFBR1 under starvation conditions.** sgTGFBR1-TFEB-GFP-expressing U2OS cells were starved for 18 hours and images acquired every hour. Single cell CNN prediction scores are marked in red for nuclear TFEB and in green for cytosolic TFEB. Dynamic Bar chart indicates the cumulative distribution of TFEB translocation in the represented cell population.

Video 5 | **Live-cell images of TFEB-GFP U2OS cells expressing sgRNA targeting CREB5b under starvation conditions.** sgCREB5b-TFEB-GFP-expressing U2OS cells were starved for 18 hours and images acquired every hour. Single cell CNN prediction scores are marked in red for nuclear TFEB and green for cytosolic TFEB. Dynamic Bar chart indicates the cumulative distribution of TFEB translocation in the represented cell population.

Video 6 | **Live-cell images of TFEB-GFP U2OS cells expressing sgRNA targeting PPP1R1B under starvation conditions.** sgPPP1R1B-TFEB-GFP-expressing U2OS cells were starved for 18 hours and images acquired every hour. Single cell CNN prediction scores are marked in red for nuclear TFEB and in green for cytosolic TFEB. Dynamic Bar chart indicates the cumulative distribution of TFEB translocation in the represented cell population.

Video 7 | **Live-cell images of TFEB-GFP U2OS cells expressing sgRNA targeting mTOR under starvation conditions.** Sg-mTOR-TFEB-GFP-expressing U2OS cells were starved for 18 hours and images acquired every hour. Single cell CNN prediction scores are marked in red for nuclear TFEB and in green for cytosolic TFEB. Dynamic Bar chart indicates the cumulative distribution of TFEB translocation in the represented cell population.

Table S1. **Parkin translocation screen using an sgRNA library subpool targeting all kinases, phosphatases and the druggable genome.** The rows in the Parkin screen gene enrichment spreadsheet are the proteins selected by the screen. Log_2 fold change and corresponded p-values calculated from the gene abundance analysis as described in Methods and related to Figure 3.

Table S2. **TFEB translocation whole genome screen.** The rows in the TFEB screen gene enrichment spreadsheet are the proteins selected by the screen. Log_2 fold change and corresponded p-values calculated from the gene abundance analysis as described in Methods and related to Figure 5.