High-throughput colocalization pipeline quantifies efficacy of mitochondrial targeting signals across different protein types

Sierra K. Lear[1,2], Jose A. Nunez[3], & Seth L. Shipman[1,4,5]

[1]Gladstone Institute of Data Science and Biotechnology, San Francisco, CA, USA

[2]Graduate Program in Bioengineering, University of California, San Francisco and Berkeley, CA, USA

[3]Department of Mechanical Engineering, University of California, Santa Barbara, CA, USA

[4]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA

[5]Chan Zuckerberg Biohub – San Francisco, San Francisco, CA, USA

Correspondence to: seth.shipman@gladstone.ucsf.edu

ABSTRACT

Efficient metabolic engineering and the development of mitochondrial therapeutics often rely upon the specific and strong import of foreign proteins into mitochondria. Fusing a protein to a mitochondria-bound signal peptide is a common method to localize proteins to mitochondria, but this strategy is not universally effective with particular proteins empirically failing to localize. To help overcome this barrier, this work develops a generalizable and open-source framework to design proteins for mitochondrial import and quantify their specific localization. By using a Python-based pipeline to quantitatively assess the colocalization of different proteins previously used for precise genome editing in a high-throughput manner, we reveal signal peptide-protein combinations that localize well in mitochondria and, more broadly, general trends about the overall reliability of commonly used mitochondrial targeting signals.

KEYWORDS

subcellular localization, protein engineering, mitochondria, high-throughput imaging, digital image analysis, Python

INTRODUCTION

Synthetic biologists increasingly leverage natural mitochondrial protein import pathways for compartmentalized metabolic engineering[1,2] and the development of molecular therapeutics[3,4]. For metabolic engineering, sequestering enzymes within yeast mitochondria has resulted in a ~300-fold increase in production of high-value biosynthetic compounds[5,6]. In human health, emerging mitochondrial therapeutics address a major unmet need since mutations to mitochondrial DNA (mtDNA) are at the root of numerous incurable diseases that affect over 1 in 5000 individuals[7,8]. In search of cures, researchers are exploring both allotopic expression, where corrected mitochondrial genes are expressed from the nuclear genome and sent to the mitochondria[9], and gene editing, where mutated mtDNA is either depleted by nucleases or corrected by base editors[10–18].

All of these approaches require efficient targeting of proteins of interest (POIs) to the mitochondria. The most common strategy to achieve such localization is by fusing a mitochondrial targeting sequence (MTS), typically a short and positively charged signal peptide, to the N-terminus of the POI. MTSs are recognized by translocases on the outer and inner mitochondrial membrane (TOM/TIM23 complex) that unfold the POI, import it through both mitochondrial membranes, and release the protein into the mitochondrial matrix following cleavage of the N-terminus MTS from the POI[19,20]. Hundreds of putative MTSs have been identified from natural proteins using computational tools[21–24].

However, attachment of an individual MTS to a given POI does not always guarantee efficient import into mitochondria[25,3]. In fact, allotopic expression and gene editing approaches in mammalian mitochondria have been hindered by low or non-specific POI localization. For instance, only a small sub-selection of protein subunits typically encoded in mtDNA are able to be allotopically expressed in mammalian cells[26,27]. Moreover, even when proteins are imported to mammalian mitochondria, they can also accumulate in other organelles[28]. For mitochondrial gene editing, such imprecise localization poses danger; for example, the mitochondrial base editor DdCBE[16] has substantial off-target editing in nuclear DNA[29], highlighting the need for more specific import of genome editing-related proteins to mitochondria.

Previous studies in yeast suggest that length, hydrophobicity, charge, and folding of the MTS or POI can all affect the efficiency of mitochondrial import[30–32,19]. However, research in mammalian cells is much sparser and at present a given MTS-POI combination cannot be assumed to result in reliable mitochondrial import. Instead, researchers often empirically test multiple MTSs before finding one that results in their specific POI localizing in mitochondria[33,15].

To address a relative lack of broad experimental data and help establish a quantitative assessment of mitochondrial localization in mammalian cells, we developed a quantitative and high-throughput imaging-based pipeline to measure POI import into mitochondria. Using this platform, we screened combinations of three commonly used N-terminus MTSs and POIs from five protein families relevant to mitochondrial gene editing to reveal the most reliable MTS-POI combinations.

RESULTS

High-throughput Localization Workflow

To investigate the effect of MTS on mitochondrial import across different POIs, we generated 66 protein cassettes containing a combination of localization signal and POI, followed by a HA tag on the C-terminus (**Fig 1a**). Localization signals included three commonly used MTSs—COX8 (29 amino acid-long peptide derived from human cytochrome c oxidase subunit VIII)[34], Su9 (69 amino acid-long peptide derived from *Neurospora crassa* ATPase subunit 9)[35], and ATG4D (42 amino acid-long peptide derived from a Atg4 cysteine protease)[36]—that were previously used to probe the import of CRISPR nucleases into mammalian mitochondria[15]. The use of no localization signal or a nuclear localization signal (NLS) served as negative controls. Nineteen POIs were chosen across five different protein classes that have been used as components of precise gene editing technologies: Class I CRISPR systems (Cas3/CASCADE)[37], Class II CRISPR/Cas nucleases, RecTs, single-stranded binding proteins (SSBs), and retron reverse transcriptases (RTs). Both Class I and II CRISPR systems can cut DNA at programmable sites to induce editing using double-stranded break (DSB) repair pathways[38]. In contrast, RecTs and SSBs have been used to integrate donor DNA into bacterial and yeast genomes through recombineering[39–42]. Finally, retron RTs allow *in vivo* production of DNA donor editing templates in both prokaryotes and eukaryotes that mediate precise editing using either DSB repair pathways or recombineering[43–46]. As a positive control, we used the construct mito-APEX2, a protein that contains an MTS derived from the mitochondrially imported COX4 fused to APEX2, which has been shown to localize to the mammalian mitochondrial matrix using immunocytochemistry and proteomic mapping that found mito-APEX2 in close proximity to mitochondrial matrix proteins[47].

To engineer mammalian cell lines expressing a given protein cassette, each construct was cloned into a PiggyBac vector under the control of a doxycycline-inducible promoter adjacent to a constitutive puromycin resistance gene. These cassettes were randomly integrated into the genome of HEK293T cells using the PiggyBac transposase system and selected with puromycin (**Fig 1b**). Biological replicates of a given construct were defined as either individual clones derived from a single bulk transposase integration or multiple parallel transposase integrations. We screened the localization of each cassette by seeding cells into 96-well plates, expressed each protein for 24 hours under an inducible promoter, performed immunocytochemistry, and imaged each well using a high-throughput confocal microscope (ImageXpress Micro Confocal High-Content Imaging System). Specifically, each cell line was imaged for nuclei using Hoescht, POI using an antibody against HA, and mitochondria using an antibody against the mitochondrial marker TIM23 (**Fig 1c**). Using this high-throughput method, we found that localization of some cassettes varied qualitatively depending on localization signal. For instance, while the mito-APEX2 and Su9-LbCas12a showed punctate expression that colocalizes with mitochondria, NLS-LbCas12a showed clear nuclear colocalization and LbCas12a with no localization signal showed a diffuse, cytoplasmic phenotype. While these particular lines illustrate the expected localization based on signal, the localization of many other cassettes, such as of ATG4D-LbCas12a, were less predictable or more ambiguous. Thus, we next developed an analytical pipeline to quantify localization within mitochondria or nuclei.

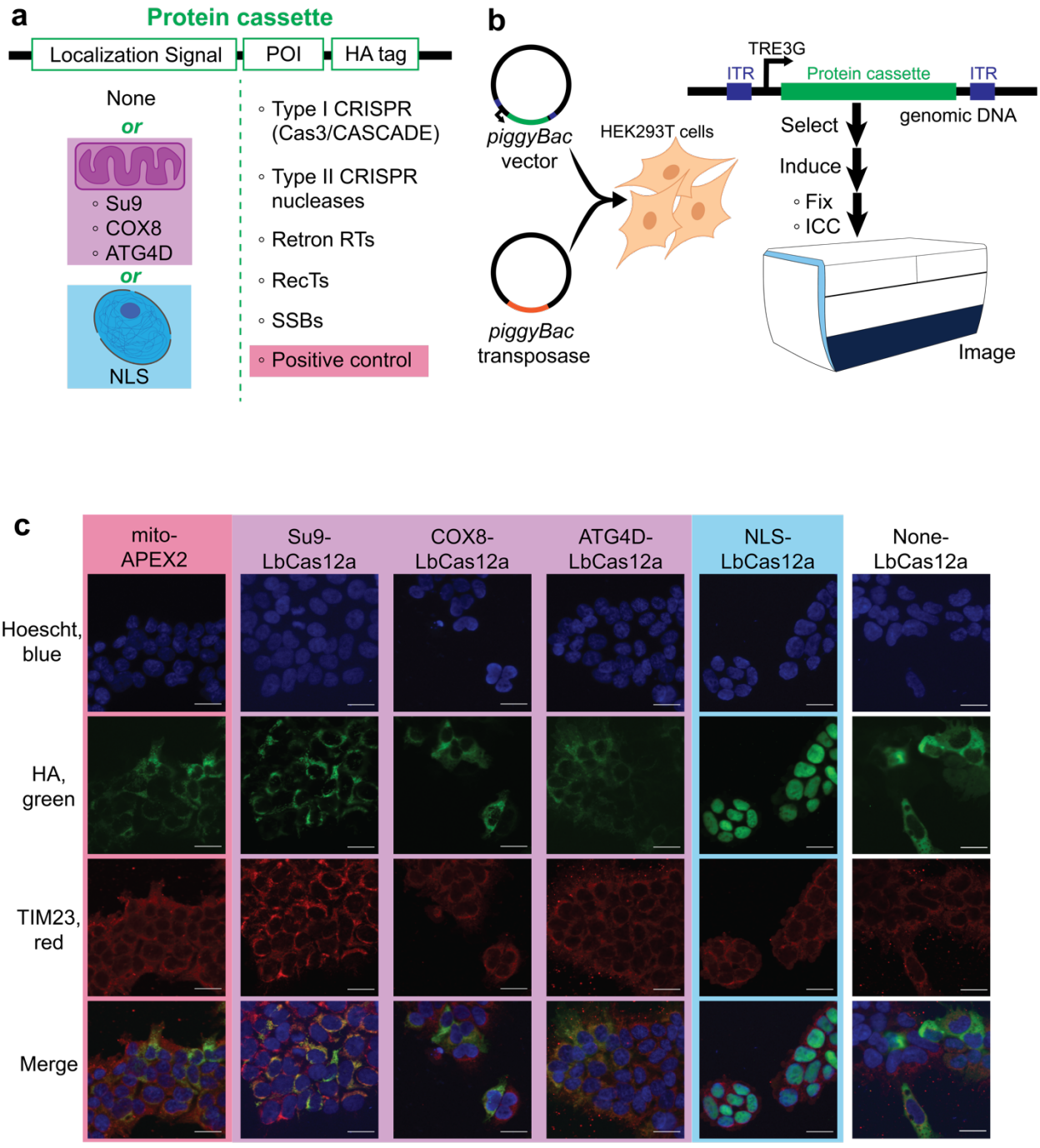

Figure 1. High-throughput localization workflow. **(a)** 66 genetic cassettes containing a combination of localization signal and POI, followed by a C-terminus HA tag, were synthesized. Localization signals include three N-terminus MTSs (Su9, COX8, and ATG4D), a nuclear localization signal (NLS), or no localization tag (None). POIs were chosen from five different protein families—Class I CRISPR/Cas proteins, Class II CRISPR nucleases, retron RTs, RecTs, and SSBs. A positive control of mito-APEX2 was also included. **(b)** Each cassette was randomly integrated into the genome of HEK293T cells using a piggyBac transposase system. Following co-transfection of the cassette in a piggyBac vector and a plasmid constitutively expressing piggyBac transposase, cells were selected for at least one week using the antibiotic puromycin. To image each cell line, expression of a given cassette was induced for 24 hours using doxycycline before being fixed and stained prior

to imaging using a high-throughput confocal microscope (ImageXpress Micro Confocal High-Content Imaging System). **(c)** Engineered cell lines were stained with Hoescht (blue), an antibody against HA (green), and an antibody against the mitochondrial marker TIM23 (red). Shown are representative images from the positive control (red background) and LbCas12a fused to one of three MTSs (purple), NLS (blue), or no localization tag (white).

Automated, quantitative, and open-source analysis pipeline

To better compare localization differences exhibited by MTS-POI combinations, we developed an unbiased Python-based analysis pipeline to quantify the mitochondrial and nuclear import between our dozens of cassettes. Crucially, we found that expression and localization were variable between individual cells of a given condition so our analysis pipeline is built to quantify colocalization at the level of single cells. Images corresponding to the nuclei and mitochondria for each biological replicate cell line from each condition were fed into a Cellpose-based machine learning model[48,49] to label individual cells (**Fig 2a,b**). Next, cells were filtered using Otsu thresholding to remove any cells with no detectable protein expression (**Fig 2c**). Specifically, Otsu thresholding was applied on each image to determine the pixel intensity threshold separating POI signal from background fluorescence. This analysis also revealed that some images had such low fluorescence that signal was effectively indistinguishable from noise. To ensure these specific images did not bias the final colocalization scores, any images in which Otsu thresholding did not separate signal from noise in each cell, as defined by the majority of filtered cells in an image failing to show a non-Gaussian intensity distribution typical of true fluorescent signal, were discarded (**Fig S1**). In some cases, only a few images for a cell line were eliminated, although—in cases where a clonal or transfected line suffered from minimal cassette expression—the entire biological replicate was removed from analysis.

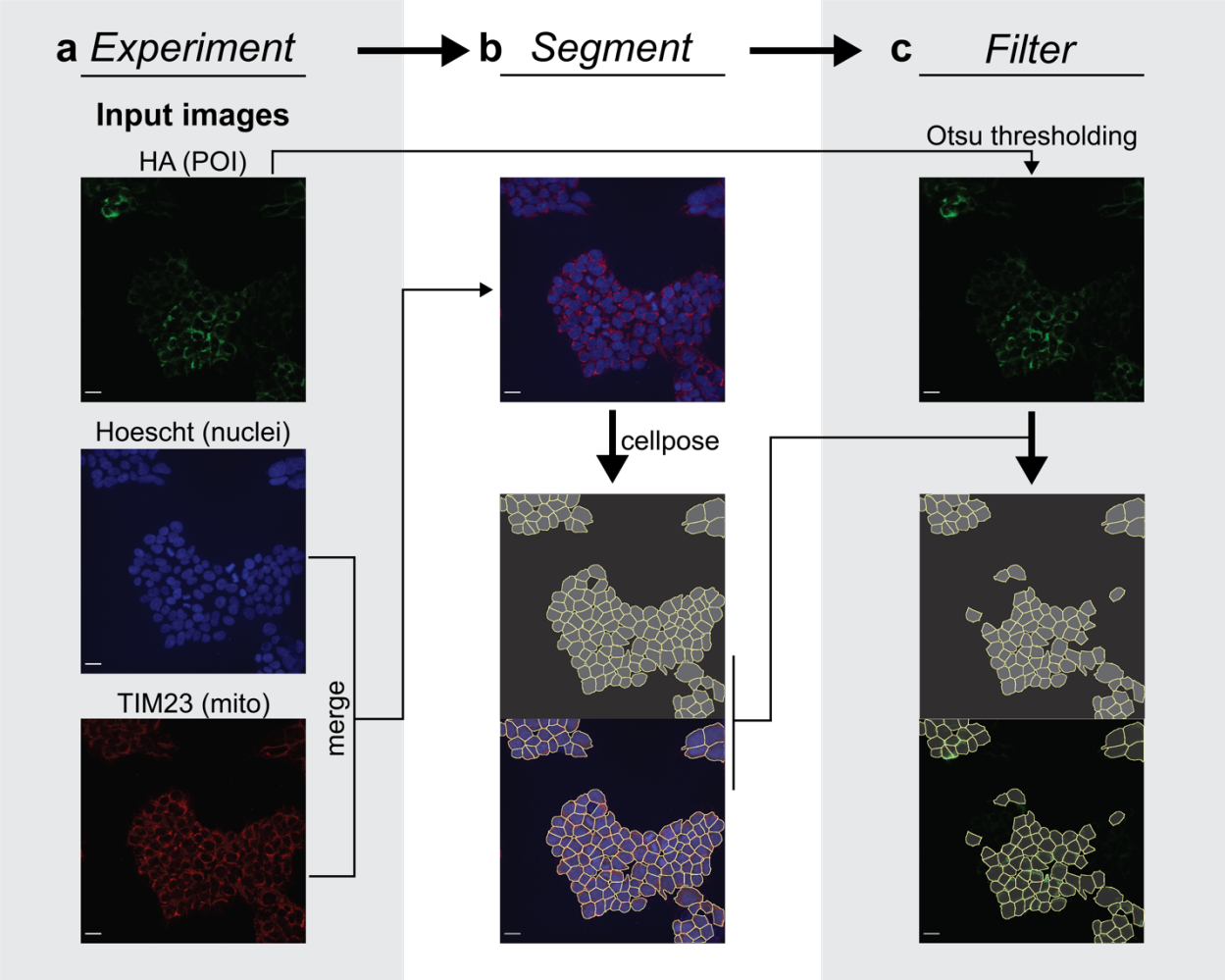

Figure 2. Machine learning algorithm enables automated labeling of cells containing cassette protein. Scale bar = 25 µm. **(a)** Unprocessed images from each fluorescent channel are fed into a Python-based analysis pipeline to segment individual cells. Shown are representative images from the cell line Su9-LbCas12a (top, HA; middle; Hoescht, bottom; TIM23). **(b)** The Hoescht and TIM23 channels are merged (top) prior to being fed into a custom, retrained neural network (arrow), resulting in automated labeling of each cell found within the image. Images below arrow; top image shows Cellpose-generated mask of segmented cells, where each yellow line indicates a cell boundary. Bottom image shows mask overlaid atop merged Hoescht/TIM23 channels. **(c)** Cellpose-segmented cells are filtered to keep only cells which express the cassette protein. Otsu thresholding is applied to the HA channel (top) to determine a threshold separating true fluorescent signal from background noise. Segmented cells containing at least 50 pixels with signal are kept to perform further colocalization measurements. Images below arrow; top image shows segmented cells following filtering. Bottom image shows filtered mask overlaid atop HA channel.

After selecting a population of filtered cells to further analyze for a given protein cassette, colocalization between the cassette protein and either mitochondria (**Fig 3a,b**) or nuclei (**Fig 3d,e**) was measured on a per cell basis using Pearson's correlation coefficient (PCC)[50]. Using this method, PCC scores vary between -1 (anti-correlated) to +1 (highly correlated). High colocalization scores indicate that a protein is collocated with a given organelle while low colocalization scores suggest little to no specific colocalization between a POI and a given organelle occurred.

For our analysis, we considered individual cells that survived quality filters from a single transfection or clone as technical replicates and summarized the overall colocalization score for a single biological replicate of each protein cassette by taking the median of all the individual cell colocalization scores for mitochondria (**Fig 3c**) or nuclei (**Fig 3f**). We replicated our experiments using at least three different transfections or five clonal lines as biological replicates.

We generally found low variability within our biological replicates, suggesting that protein import is a fairly reliable phenomenon. The positive control mito-APEX2 obtained an average score of 0.63 +/- 0.03 (mean +/- std. dev), a high PCC value that strongly implies mito-APEX2 is imported into the mitochondrial matrix. Similar to our previous qualitative assessments (**Fig 1c**), we found that colocalization scores, even across a single POI, vary depending on localization signals. The colocalization scores of LbCas12a fused to the Su9 or COX8 MTSs were not statistically different from mito-APEX2, suggesting mitochondrial import had occurred. In comparison, when LbCas12a was instead fused to ATG4D, no localization signal, or NLS, colocalization scores dropped significantly, indicating less mitochondrial import occurred (**Fig 3g**). Moreover, when comparing nuclear colocalization scores (**Fig 3h**), all cell lines except NLS-LbCas12a showed a consistent, low nuclear colocalization score, suggesting little to no nuclear import. As expected, only the cell line fused to a NLS had a high colocalization score indicating high nuclear import. These findings suggest that our workflow is able to compare import efficiencies across different combinations of MTS and POI for multiple organelles.

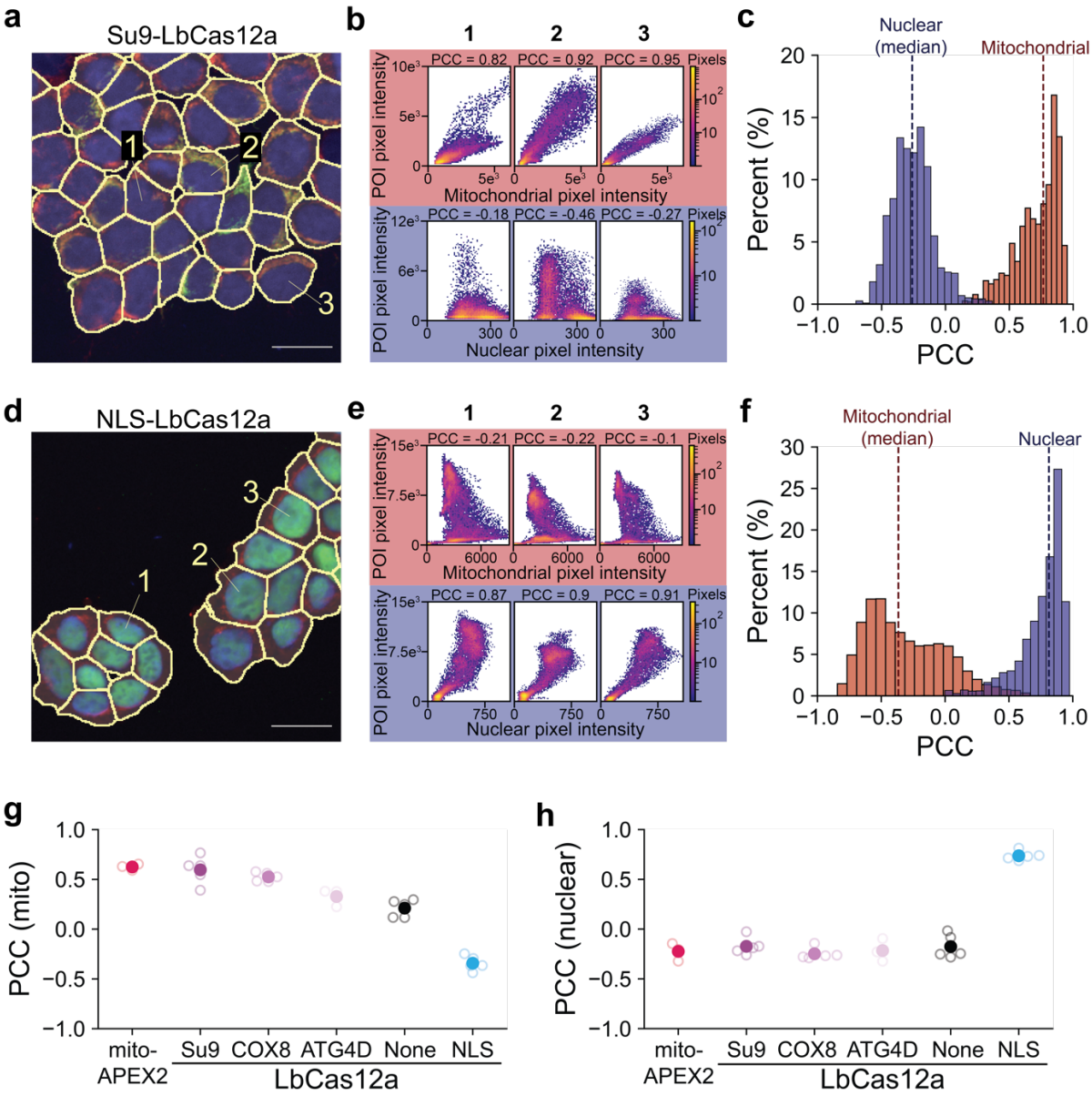

Figure 3. Computational workflow quantifies the colocalization of cassette proteins with mitochondria or nuclei. **(a)** Representative image of a cell line (Su9-LbCas12a) with clear mitochondrial expression of its cassette protein, with mask in yellow overlaid on top. Numbers refer to three representative cells for which data is shown in (b). Scale bar = 25 μm. **(b)** Heatmaps depicting the relationship between Su9-LbCas12a pixel intensity and organellar pixel intensity (mitochondria on top in red; nuclei on bottom in blue) for each pixel within a representative cell from (a) (left; cell #1, middle; cell #2, right; cell #3). Color depicts the number of pixels. The strength of the linear relationship between pixel intensities, or colocalization, within each cell is calculated using PCC, and the result depicted on top its respective heatmap. **(c)** Histogram depicting the all the colocalization scores for all the cells for one clonal line expressing Su9-LbCas12a. Mitochondrial PCC scores are shown in red, while nuclear PCC scores are shown in blue. Dotted lines depict the median colocalization score for mitochondria (red) and nuclei (blue). **(d)** Representative image of a cell line (NLS-LbCas12a) with clear nuclear expression of its cassette protein, with mask in yellow overlaid on top. Numbers refer to three representative cells. Scale bar = 25 μm. **(e)** Heatmaps depicting the relationship between NKS-LbCas12a pixel intensity and organellar pixel intensity (mitochondria on top in red; nuclei on bottom in blue) for each pixel within a representative cell from (d) (left; cell #1, middle; cell #2, right; cell #3). Color depicts the number of pixels. The

strength of the linear relationship between pixel intensities, or colocalization, within each cell is calculated using PCC, whose result is on top its respective heatmap. **(f)** Histogram depicting the all the colocalization scores for all the cells for one clonal line expressing NLS-LbCas12a. Mitochondrial PCC scores are shown in red, while nuclear PCC scores are shown in blue. Dotted lines depict the median colocalization score for mitochondria (red) and nuclei (blue). **(g)** Mitochondrial colocalization of positive control mito-APEX2 (red) and LbCas12a fused to different localization tags, as measured using the described experimental and analytic workflow. There is a significant effect of localization signal (one-way ANOVA, $P<0.0001$), where ATG4D ($P=0.0021$), no signal ($P<0.0001$), and NLS ($P<0.0001$) are all significantly different from mito-APEX2, but Su9 ($P=0.9831$) and COX8 ($P=0.376$) are not (Dunnett's corrected). Open circles are biological replicates; closed circles are average of all biological replicates. **(h)** Nuclear colocalization of mito-APEX2 (red) and LbCas12a fused to different localization tags measured using the described experimental and analytical workflow. There is a significant effect of localization signal (one-way ANOVA, $P<0.0001$), where NLS ($P<0.0001$) is significantly different from mito-APEX2, but Su9 ($P=0.8587$), COX8 ($P=0.9930$), ATG4D ($P=0.9998$), and no signal ($P=0.0.8825$) are not (Dunnett's corrected). Open circles are biological replicates; closed circles are average of all biological replicates. Additional statistical details in Supplementary Table 1.

MTS selection strongly influences mitochondrial import

After validating the analytical pipeline, we used this workflow to quantify the mitochondrial and nuclear import of all 66 different protein cassettes (**Fig 4a; S2a,b**). Interestingly, mitochondrial colocalization scores did not cluster bimodally into high and low scores. Instead, scores were distributed continuously, suggesting that different cassettes have varying capabilities to drive POIs to the mitochondria.

Previous studies have found that both the specific MTS and hydrophobicity of individual POIs can influence the probability that a MTS-POI will localize in the mitochondria [31,19,33]. To determine if localization signal influences mitochondrial import across the four protein families tested, we investigated how the mitochondrial import rank of cassettes clustered based on localization signal (**Fig 4b**). We fit a linear mixed effects model with a fixed effect of localization signal and a random intercept within each protein family, to specifically test the statistical significance of each localization signal while accounting for the variability inherent within each protein of interest. As expected, we found that cassettes clustered based on specific localization signals; while fusion to an MTS resulted in higher rankings across the board, having a NLS led to clear clustering at the bottom of the rankings. However, there was a clear ranking priority to how well each MTS performed compared to each other, with Su9 and COX8 driving high mitochondrial import, while ATG4D performed significantly worse.

Given ATG4D's uneven performance across different POIs, we next further analyzed how colocalization scores varied based on localization scores within different protein families (**Fig 4c-g**). Interestingly, all three MTSs were able to drive high mitochondrial import of Class I CRISPR-related proteins (**Fig 4c**), whereas both COX8 and ATG4D appeared less capable than Su9 of importing the larger Class II CRISPR nucleases (**Fig 4d**). Of the two retron RTs tested, only Su9 led to mitochondrial localization of retron-Eco1 RT, whereas COX8 and ATG4D were able to localize retron Eco2 RT (**Fig 4e**). Finally, while both COX8 and Su9 were able to import all RecTs and SSBs tested into the mitochondria (**Fig 4f,g**), ATG4D instead appeared to misdirect these proteins to a different location, based on an unusual punctate pattern which did not colocalize with mitochondria (**Fig S2c**).

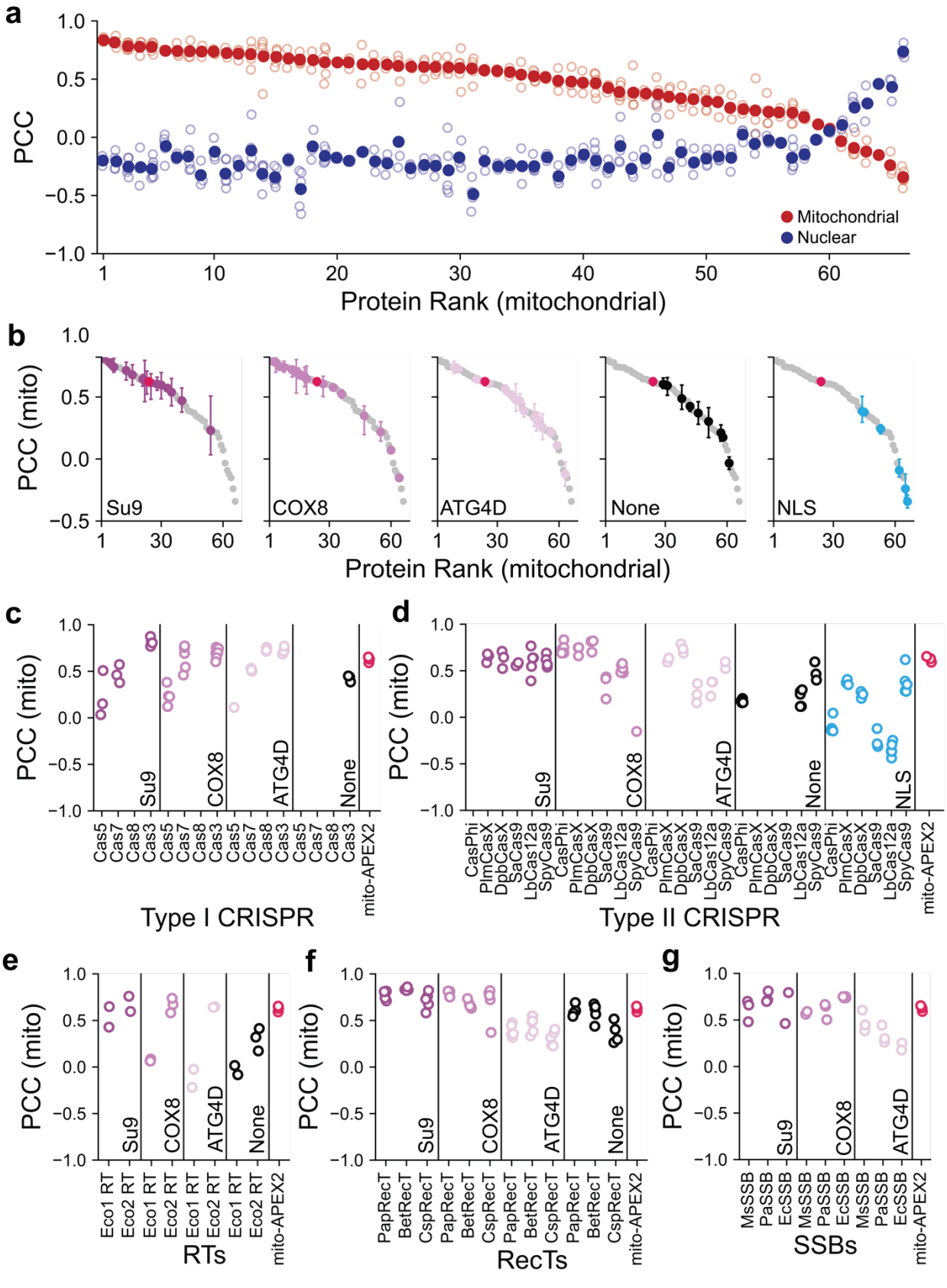

Figure 4. MTS selection strongly influences mitochondrial import. **(a)** Mitochondrial (red) and nuclear (blue) colocalization scores of 66 different protein cassettes. Proteins were ranked based on average mitochondrial colocalization score, from highest mitochondrial PCC to lowest. Open circles are biological replicates; closed circles are average of all biological replicates. **(b)** Clustering of cassettes based on specific localization signal. The entire continuum of all mitochondrial colocalization scores from (a) is shown in each subplot in gray, while the positive control mito-APEX2 is shown in each subplot in red. The clustering of cassettes with a specific localization signal is shown in another color on top (from left to right subplots: Su9, COX8, ATG4D, none, and NLS). There is a significant effect of localization signal (linear mixed effects model, $P<0.0001$), where ATG4D is significantly different from both Su9 ($P<0.0001$) and COX8 ($P<0.0001$) but not None (0.092559) during follow-up (Kenward-Roger corrected). Meanwhile, there is no significant difference between Su9 and COX8 ($P=0.88139$). Closed circles indicate average of all biological replicates; error bars indicate standard deviation. **(c)-(g)** Mitochondrial colocalization scores broken down by localization signal within a given protein family (**c**; Class I CRISPR Cas3/CASCADE, **d**; Class II CRISPR nucleases, **e**; retron RTs, **f**; RecTs, **g**; SSBs). Specific proteins are listed in order of amino acid length, from shortest on the left to longest on the right. As a reference for mitochondrial PCC scores suggestive of mitochondrial import, scores for positive control mito-APEX2 are shown to the right of each figure (red). Open circles are individual biological replicates.

DISCUSSION

Importing non-canonical proteins into mammalian mitochondria is a critical step to overcome particular challenges in metabolic engineering and for developing therapeutics to treat genetic diseases of the mtDNA. Here, we design a high-throughput imaging-based workflow to quickly screen the subcellular localization of a tagged protein. This method enabled us to determine the best MTS-POI combination across three commonly used MTSs and five different protein classes that are components of gene editing technologies. The results from these screens should help drive more robust mitochondrial gene editing by enabling researchers to test mitochondrial gene editing using other nucleases or proteins beyond SpyCas9, which has been reported to cause mitochondrial dysfunction when imported to mitochondria[15], or testing MTSs that yield more specific mitochondria import to prevent off-target effects in nuclear DNA[29].

This work also reveals broader trends about which MTS to use across multiple protein types. By testing three common MTSs on a diverse set of POIs, we find that Su9 and COX8 consistently performed the best while ATG4D performed the most unevenly and misdirected several POIs to an alternate location. In addition, unlike COX8 and ATG4D, Su9 was able to import specific proteins, such as large Class II nucleases and Eco1 RT, into the mitochondria.

As a tool for the field, our Python-based workflow implemented in an annotated Jupyter notebook can be reused for future experiments, including more generally to other screens using different proteins, organelles of interest, or cell lines. Although we used a high-throughput confocal microscope in our own workflow, other confocal microscopes could easily be used, depending on the necessary throughput or number of samples. Computationally, the analytic pipeline uses Python, an open-source programming language, and relies on pixel-based colocalization analyses and quality check steps to eliminate cells not expressing a protein of interest that can be universally applied regardless of cell line or differing expression levels or phenotypes[50]. The only experiment-specific alteration to the pipeline would be to apply a different neural network to segment individual cells, depending on the cell line and seeding density used. However, Cellpose already offers a collection of ready-to-use neural networks or—if further refinement is necessary—an intuitive GUI that enables users to retrain and create their own neural network in fewer than 30

minutes[49]. Therefore, others should be able to easily apply this analytical framework to quantify colocalization within their own fluorescent images.

Methods

Constructs and strains

Protein cassettes were constructed by amplifying localization signals and POI nucleotide sequences using PCR from synthesized gBlocks (IDT) or existing plasmids. Complete protein cassettes were cloned into a PiggyBac integrating plasmid for doxycycline-inducible human protein expression (TetOn-3G promoter) using Gibson assembly. Alternatively, some cassettes were synthesized into the same custom PiggyBac integrating plasmid by Twist Bioscience (see Supplementary Table 2).

Stable mammalian cells for imaging were generated using the standard Lipofectamine 3000 transfection protocol (Invitrogen) and a PiggyBac transposase system. T12.5 flasks with 50-70% confluent HEK293T cells were transfected using 1.6 µg POI cassette expression plasmid and 0.8 µg PiggyBac transposase plasmid (pCMV-hyPBase). Stable cell lines were selected using puromycin for at least one week.

Clonal lines were generated by growing individual cells into separate cell populations. Specifically, stable cell lines were serially diluted to a final concentration of 2.5 cells per mL media then seeded into a 96-well plate using 100 µL/well. Wells that received a single cell had media refreshed weekly until a clonal line proliferated to ~40% confluency, at which point a clonal line was passaged to a larger flask for further experiments.

Immunocytochemistry

96-well glass bottom plates with #1 cover glass (Cellvis, catalog # P96-1-N) were coated with a mixture of 50% poly-D-lysine (ThermoFisher Scientific, catalog #A3890401) and DPBS (ThermoFisher Scientific, catalog #14040133) for 30 minutes at room temperature. Wells were washed three times with distilled water and left out to dry for at least 2 hours prior to seeding.

Cells were seeded at a density of 10,000 cells per well. The following day, doxycycline was added at a final concentration of 1 µg/mL to induce expression of the protein cassette. At 24 hours post-induction, cell nuclei were stained using a final concentration of 10 µM Hoescht for at least 5 minutes prior to fixation.

For fixation, media was aspirated from each well and replaced with a solution of 4% paraformaldehyde (PFA) created fresh by fixing a 1 mL 16% (w/v) PFA ampule (ThermoFisher Scientific, catalog #28906) with 3 mL PBS. Cells were fixed for 30 minutes at room temperature prior to three 5-minute washes with PBS. Following fixation, cells were permeabilized and blocked for an hour at room temperature using blocking buffer made fresh with the following ingredients: PBS containing 10% donkey serum (Sigma-Aldrich, catalog #D9663), 10% Triton X-100 (Sigma-Aldrich, catalog #X100), and 100 mg BSA (Sigma-Aldrich, catalog #A9418) per 10 mL solution. Next, cells were incubated overnight at 4°C in blocking buffer with the antibodies anti-HA tag conjugated to DyLight 550 (ThermoFisher Scientific, catalog #26183-D550) and anti-TIM23 (Abcam, catalog #ab230253) each added at a 1:100 dilution. After performing three more 5-minute washes, cells were incubated with a secondary antibody goat anti-rabbit conjugated to DyLight

650 (ThermoFisher Scientific, catalog #84546) at a 1:500 dilution in blocking buffer for 3 hours. Following secondary antibody incubation, three more 5-minute washes were performed prior to the addition of 30 uL antifade mountant (ThermoFisher Scientific, catalog #S36967) per well.

Plates were wrapped in aluminum foil to avoid light and either stored temporarily at 4°C or at -20°C for longer-term storage prior to imaging.

Imaging

Stained cells were imaged using an ImageXpress Micro Confocal High-Content Imaging System (Molecular Devices) using a 40X water immersion objective by taking a 7-layer Z-stack, with each layer spaced 0.3 µm apart, at four different sites per well.

Colocalization Image Analysis Pipeline

A colocalization image analysis pipeline was made using jupyter-notebook in Python 3[51,52], and uses the following packages: numpy, pandas, scipy, skimage, tdqm, and tifffile. Additionally, the pipeline requires the Cellpose code library[48,49] along with these additional packages: numba, opencv, and pytorch. Using the Cellpose GUI also requires PyQt and pyqtgraph[53].

TIFF files consisting of merged nuclear and mitochondrial channels were created using a custom function and fed into a neural network retrained according to the instructions for the Cellpose GUI[49]. Briefly, the "CP" model from the Cellpose model zoo was initially used to segment all images. Afterwards, about five images with poor initial segmentation were chosen for manual annotation. The CP model was then retrained using the corrected labels and the new model was re-run on all images.

To remove segmented cells that did not contain expression of the cassette protein, Otsu thresholding was performed on the HA channel to determine a pixel intensity threshold separating signal from background for each image. Only segmented cells containing at least 50 pixels of cassette protein signal, referred to as filtered cells, were kept for further analysis.

Two additional functions to ensure quality-check steps were also implemented. First, any image containing fewer than six filtered cells was automatically removed from further analysis. Second, since the overall expression of a cassette protein can vary between different cell lines, a function was written to ensure that the filtering step effectively distinguished between cells that did or did not express a cassette protein. Individual cells containing noise, rather than signal, exhibit a Gaussian distribution of protein cassette pixel intensities. In contrast, cells with signal tend to exhibit non-normal or skewed pixel intensity distributions. Thus, for every image, a "non-Gaussian" test was performed on each filtered cell by testing for normality. If over 60% of filtered cells failed the "non-Gaussian" test, then this result suggests that the majority of filtered cells within the image do not contain true expression of the cassette protein, thus that specific image would be removed from further analysis.

Afterwards, a custom function was built to calculate PCC between the HA channel pixel intensities and either the mitochondrial or nuclear channel pixel intensities for every filtered cell related to a given biological replicate. Due to the skew present in most PCC distribution, these results were summarized by taking the median of all the filtered cells for a given biological replicate.

Statistics

ANOVA and post-hoc analyses were performed using GraphPad Prism v9.4.1 For linear mixed effects modeling, we used *R* v4.2.3 using the *lme4* and *lmerTest* packages (post-hoc analyses used the Kenward-Roger degrees of freedom correction method implemented in the package *pbkrtest*).

ABBREVIATIONS

MTS: mitochondrial targeting sequence
POI: protein of interest
mtDNA: mitochondrial DNA
CRISPR: clustered regularly interspaced short palindromic repeats

Conflict of interest statement

There are no conflicts of interests.

Acknowledgements

This work was supported by the National Institute of General Medical Sciences (DP2GM140917), the National Institute of Biomedical Imaging and Bioengineering (R21EB031393), and the UCSF Program for Breakthrough Biomedical Research (partially funded by the Sandler Foundation). S.K.L. was supported by an NSF Graduate Research Fellowship (2034836). S.L.S. is a Chan Zuckerberg Biohub – San Francisco Investigator. We thank David Darevsky for his advice on coding and statistics.

References

(1) Agrimi, G. Role of Mitochondrial Carriers in Metabolic Engineering. *J. Bioprocess. Biotech.* **2014**, *04* (05). https://doi.org/10.4172/2155-9821.1000164.

(2) Huttanus, H. M.; Feng, X. Compartmentalized Metabolic Engineering for Biochemical and Biofuel Production. *Biotechnol. J.* **2017**, *12* (6), 1700052. https://doi.org/10.1002/biot.201700052.

(3) Verechshagina, N. A.; Konstantinov, Yu. M.; Kamenski, P. A.; Mazunin, I. O. Import of Proteins and Nucleic Acids into Mitochondria. *Biochem. Mosc.* **2018**, *83* (6), 643–661. https://doi.org/10.1134/S0006297918060032.

(4) Di Donfrancesco, A.; Massaro, G.; Di Meo, I.; Tiranti, V.; Bottani, E.; Brunetti, D. Gene Therapy for Mitochondrial Diseases: Current Status and Future Perspective. *Pharmaceutics* **2022**, *14* (6), 1287. https://doi.org/10.3390/pharmaceutics14061287.

(5) Farhi, M.; Marhevka, E.; Masci, T.; Marcos, E.; Eyal, Y.; Ovadis, M.; Abeliovich, H.; Vainstein, A. Harnessing Yeast Subcellular Compartments for the Production of Plant Terpenoids. *Metab. Eng.* **2011**, *13* (5), 474–481. https://doi.org/10.1016/j.ymben.2011.05.001.

(6) Avalos, J. L.; Fink, G. R.; Stephanopoulos, G. Compartmentalization of Metabolic Pathways in Yeast Mitochondria Improves the Production of Branched-Chain Alcohols. *Nat. Biotechnol.* **2013**, *31* (4), 335–341. https://doi.org/10.1038/nbt.2509.

(7) Taylor, R. W.; Turnbull, D. M. Mitochondrial DNA Mutations in Human Disease. *Nat. Rev. Genet.* **2005**, *6* (5), 389. https://doi.org/10.1038/nrg1606.

(8) Ng, Y. S.; Turnbull, D. M. Mitochondrial Disease: Genetics and Management. *J. Neurol.* **2016**, *263*, 179–191. https://doi.org/10.1007/s00415-015-7884-3.

(9) Artika, I. M. Allotopic Expression of Mitochondrial Genes: Basic Strategy and Progress. *Genes Dis.* **2020**, *7* (4), 578–584. https://doi.org/10.1016/j.gendis.2019.08.001.

(10) Bacman, S. R.; Williams, S. L.; Pinto, M.; Peralta, S.; Moraes, C. T. Specific Elimination of Mutant Mitochondrial Genomes in Patient-Derived Cells by MitoTALENs. *Nat. Med.* **2013**, *19* (9), 1111–1113. https://doi.org/10.1038/nm.3261.

(11) Gammage, P. A.; Rorbach, J.; Vincent, A. I.; Rebar, E. J.; Minczuk, M. Mitochondrially Targeted ZFNs for Selective Degradation of Pathogenic Mitochondrial Genomes Bearing Large-Scale Deletions or Point Mutations. *EMBO Mol. Med.* **2014**, *6* (4), 458–466. https://doi.org/10.1002/emmm.201303672.

(12) Reddy, P.; Ocampo, A.; Suzuki, K.; Luo, J.; Bacman, S. R.; Williams, S. L.; Sugawara, A.; Okamura, D.; Tsunekawa, Y.; Wu, J.; Lam, D.; Xiong, X.; Montserrat, N.; Esteban, C. R.; Liu, G.-H.; Sancho-Martinez, I.; Manau, D.; Civico, S.; Cardellach, F.;

del Mar O'Callaghan, M.; Campistol, J.; Zhao, H.; Campistol, J. M.; Moraes, C. T.; Izpisua Belmonte, J. C. Selective Elimination of Mitochondrial Mutations in the Germline by Genome Editing. *Cell* **2015**, *161* (3), 459–469. https://doi.org/10.1016/j.cell.2015.03.051.

(13) Jo, A.; Ham, S.; Lee, G. H.; Lee, Y.-I.; Kim, S.; Lee, Y.-S.; Shin, J.-H.; Lee, Y. Efficient Mitochondrial Genome Editing by CRISPR/Cas9. *BioMed Res. Int.* **2015**, *2015*, 305716. https://doi.org/10.1155/2015/305716.

(14) Hashimoto, M.; Bacman, S. R.; Peralta, S.; Falk, M. J.; Chomyn, A.; Chan, D. C.; Williams, S. L.; Moraes, C. T. MitoTALEN: A General Approach to Reduce Mutant MtDNA Loads and Restore Oxidative Phosphorylation Function in Mitochondrial Diseases. *Mol. Ther.* **2015**, *23* (10), 1592–1599. https://doi.org/10.1038/mt.2015.126.

(15) Antón, Z.; Mullally, G.; Ford, H. C.; van der Kamp, M. W.; Szczelkun, M. D.; Lane, J. D. Mitochondrial Import, Health and MtDNA Copy Number Variability Seen When Using Type II and Type V CRISPR Effectors. *J. Cell Sci.* **2020**, *133* (18), jcs248468. https://doi.org/10.1242/jcs.248468.

(16) Mok, B. Y.; de Moraes, M. H.; Zeng, J.; Bosch, D. E.; Kotrys, A. V.; Raguram, A.; Hsu, F.; Radey, M. C.; Peterson, S. B.; Mootha, V. K.; Mougous, J. D.; Liu, D. R. A Bacterial Cytidine Deaminase Toxin Enables CRISPR-Free Mitochondrial Base Editing. *Nature* **2020**. https://doi.org/10.1038/s41586-020-2477-4.

(17) Hussain, S.-R. A.; Yalvac, M. E.; Khoo, B.; Eckardt, S.; McLaughlin, K. J. Adapting CRISPR/Cas9 System for Targeting Mitochondrial Genome. *Front. Genet.* **2021**, *12*. https://doi.org/10.3389/fgene.2021.627050.

(18) Bi, R.; Li, Y.; Xu, M.; Zheng, Q.; Zhang, D.-F.; Li, X.; Ma, G.; Xiang, B.; Zhu, X.; Zhao, H.; Huang, X.; Zheng, P.; Yao, Y.-G. Direct Evidence of CRISPR-Cas9-Mediated Mitochondrial Genome Editing. *The Innovation* **2022**, *3* (6). https://doi.org/10.1016/j.xinn.2022.100329.

(19) Chacinska, A.; Koehler, C. M.; Milenkovic, D.; Lithgow, T.; Pfanner, N. Importing Mitochondrial Proteins: Machineries and Mechanisms. *Cell* **2009**, *138* (4), 628–644. https://doi.org/10.1016/j.cell.2009.08.005.

(20) Wiedemann, N.; Pfanner, N. Mitochondrial Machineries for Protein Import and Assembly. *Annu. Rev. Biochem.* **2017**, *86* (1), 685–714. https://doi.org/10.1146/annurev-biochem-060815-014352.

(21) Fukasawa, Y.; Tsuji, J.; Fu, S.-C.; Tomii, K.; Horton, P.; Imai, K. MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites. *Mol. Cell. Proteomics MCP* **2015**, *14* (4), 1113–1126. https://doi.org/10.1074/mcp.M114.043083.

(22) Savojardo, C.; Martelli, P. L.; Fariselli, P.; Casadio, R. TPpred3 Detects and Discriminates Mitochondrial and Chloroplastic Targeting Peptides in Eukaryotic Proteins. *Bioinformatics* **2015**, *31* (20), 3269–3275. https://doi.org/10.1093/bioinformatics/btv367.

(23) Almagro Armenteros, J. J.; Salvatore, M.; Emanuelsson, O.; Winther, O.; von Heijne, G.; Elofsson, A.; Nielsen, H. Detecting Sequence Signals in Targeting Peptides Using Deep Learning. *Life Sci. Alliance* **2019**, *2* (5), e201900429. https://doi.org/10.26508/lsa.201900429.

(24) Bayne, A. N.; Dong, J.; Amiri, S.; Farhan, S. M. K.; Trempe, J.-F. MTSviewer: A Database to Visualize Mitochondrial Targeting Sequences, Cleavage Sites, and Mutations on Protein Structures. bioRxiv November 16, 2022, p 2021.11.25.470064. https://doi.org/10.1101/2021.11.25.470064.

(25) Van Steeg, H.; Oudshoorn, P.; Van Hell, B.; Polman, J. E.; Grivell, L. A. Targeting Efficiency of a Mitochondrial Pre-Sequence Is Dependent on the Passenger Protein. *EMBO J.* **1986**, *5* (13), 3643–3650. https://doi.org/10.1002/j.1460-2075.1986.tb04694.x.

(26) Oca-Cossio, J.; Kenyon, L.; Hao, H.; Moraes, C. T. Limitations of Allotopic Expression of Mitochondrial Genes in Mammalian Cells. *Genetics* **2003**, *165* (2), 707–720. https://doi.org/10.1093/genetics/165.2.707.

(27) Perales-Clemente, E.; Fernández-Silva, P.; Acín-Pérez, R.; Pérez-Martos, A.; Enríquez, J. A. Allotopic Expression of Mitochondrial-Encoded Genes in Mammals: Achieved Goal, Undemonstrated Mechanism or Impossible Task? *Nucleic Acids Res.* **2011**, *39* (1), 225–234. https://doi.org/10.1093/nar/gkq769.

(28) Bader, G.; Enkler, L.; Araiso, Y.; Hemmerle, M.; Binko, K.; Baranowska, E.; De Craene, J.-O.; Ruer-Laventie, J.; Pieters, J.; Tribouillard-Tanvier, D.; Senger, B.; di Rago, J.-P.; Friant, S.; Kucharczyk, R.; Becker, H. D. Assigning Mitochondrial Localization of Dual Localized Proteins Using a Yeast Bi-Genomic Mitochondrial-Split-GFP. *eLife* **2020**, *9*, e56649. https://doi.org/10.7554/eLife.56649.

(29) Wei, Y.; Li, Z.; Xu, K.; Feng, H.; Xie, L.; Li, D.; Zuo, Z.; Zhang, M.; Xu, C.; Yang, H.; Zuo, E. Mitochondrial Base Editor DdCBE Causes Substantial DNA Off-Target Editing in Nuclear Genome of Embryos. *Cell Discov.* **2022**, *8* (1), 1–4. https://doi.org/10.1038/s41421-022-00391-5.

(30) Galanis, M.; Devenish, R. J.; Nagley, P. Duplication of Leader Sequence for Protein Targeting to Mitochondria Leads to Increased Import Efficiency. *FEBS Lett.* **1991**, *282* (2), 425–430. https://doi.org/10.1016/0014-5793(91)80529-C.

(31) Claros, M. G.; Perea, J.; Shu, Y.; Samatey, F. A.; Popot, J.-L.; Jacq, C. Limitations to in Vivo Import of Hydrophobic Proteins into Yeast Mitochondria. *Eur. J. Biochem.* **1995**, *228* (3), 762–771. https://doi.org/10.1111/j.1432-1033.1995.0762m.x.

(32) Wilcox, A. J.; Choy, J.; Bustamante, C.; Matouschek, A. Effect of Protein Structure on Mitochondrial Import. *Proc. Natl. Acad. Sci.* **2005**, *102* (43), 15435–15440. https://doi.org/10.1073/pnas.0507324102.

(33) Chin, R. M.; Panavas, T.; Brown, J. M.; Johnson, K. K. Optimized Mitochondrial Targeting of Proteins Encoded by Modified MRNAs Rescues Cells Harboring Mutations in MtATP6. *Cell Rep.* **2018**, *22* (11), 2818–2826. https://doi.org/10.1016/j.celrep.2018.02.059.

(34) Rizzuto, R.; Simpson, A. W. M.; Brini, M.; Pozzan, T. Rapid Changes of Mitochondrial Ca2+ Revealed by Specifically Targeted Recombinant Aequorin. *Nature* **1992**, *358* (6384), 325–327. https://doi.org/10.1038/358325a0.

(35) Hartl, F.-U.; Pfanner, N.; Nicholson, D. W.; Neupert, W. Mitochondrial Protein Import. *Biochim. Biophys. Acta BBA - Rev. Biomembr.* **1989**, *988* (1), 1–45. https://doi.org/10.1016/0304-4157(89)90002-6.

(36) Betin, V. M. S.; MacVicar, T. D. B.; Parsons, S. F.; Anstee, D. J.; Lane, J. D. A Cryptic Mitochondrial Targeting Motif in Atg4D Links Caspase Cleavage with Mitochondrial Import and Oxidative Stress. *Autophagy* **2012**, *8* (4), 664–676. https://doi.org/10.4161/auto.19227.

(37) Csörgő, B.; León, L. M.; Chau-Ly, I. J.; Vasquez-Rifo, A.; Berry, J. D.; Mahendra, C.; Crawford, E. D.; Lewis, J. D.; Bondy-Denomy, J. A Compact Cascade-Cas3 System for Targeted Genome Engineering. *Nat. Methods* **2020**, *17* (12), 1183–1190. https://doi.org/10.1038/s41592-020-00980-w.

(38) Makarova, K. S.; Wolf, Y. I.; Iranzo, J.; Shmakov, S. A.; Alkhnbashi, O. S.; Brouns, S. J. J.; Charpentier, E.; Cheng, D.; Haft, D. H.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Scott, D.; Shah, S. A.; Siksnys, V.; Terns, M. P.; Venclovas, Č.; White, M. F.; Yakunin, A. F.; Yan, W.; Zhang, F.; Garrett, R. A.; Backofen, R.; van der Oost, J.; Barrangou, R.; Koonin, E. V. Evolutionary Classification of CRISPR–Cas Systems: A Burst of Class 2 and Derived Variants. *Nat. Rev. Microbiol.* **2020**, *18* (2), 67–83. https://doi.org/10.1038/s41579-019-0299-x.

(39) DiCarlo, J. E.; Conley, A. J.; Penttilä, M.; Jäntti, J.; Wang, H. H.; Church, G. M. Yeast Oligo-Mediated Genome Engineering (YOGE). *ACS Synth. Biol.* **2013**, *2* (12), 741–749. https://doi.org/10.1021/sb400117c.

(40) Barbieri, E. M.; Muir, P.; Akhuetie-Oni, B. O.; Yellman, C. M.; Isaacs, F. J. Precise Editing at DNA Replication Forks Enables Multiplex Genome Engineering in Eukaryotes. *Cell* **2017**, *171* (6), 1453-1467.e13. https://doi.org/10.1016/j.cell.2017.10.034.

(41) Wannier, T. M.; Nyerges, A.; Kuchwara, H. M.; Czikkely, M.; Balogh, D.; Filsinger, G. T.; Borders, N. C.; Gregg, C. J.; Lajoie, M. J.; Rios, X.; Pál, C.; Church, G. M. Improved Bacterial Recombineering by Parallelized Protein Discovery. *Proc. Natl. Acad. Sci.* **2020**, *117* (24), 13689–13698. https://doi.org/10.1073/pnas.2001588117.

(42) Filsinger, G. T.; Wannier, T. M.; Pedersen, F. B.; Lutz, I. D.; Zhang, J.; Stork, D. A.; Debnath, A.; Gozzi, K.; Kuchwara, H.; Volf, V.; Wang, S.; Rios, X.; Gregg, C. J.; Lajoie, M. J.; Shipman, S. L.; Aach, J.; Laub, M. T.; Church, G. M. Characterizing the Portability of Phage-Encoded Homologous Recombination Proteins. *Nat. Chem. Biol.* **2021**, *17* (4), 394–402. https://doi.org/10.1038/s41589-020-00710-5.

(43) Kong, X.; Wang, Z.; Zhang, R.; Wang, X.; Zhou, Y.; Shi, L.; Yang, H. Precise Genome Editing without Exogenous Donor DNA via Retron Editing System in Human Cells. *Protein Cell* **2021**, *12* (11), 899–902. https://doi.org/10.1007/s13238-021-00862-7.

(44) Schubert, M. G.; Goodman, D. B.; Wannier, T. M.; Kaur, D.; Farzadfard, F.; Lu, T. K.; Shipman, S. L.; Church, G. M. High-Throughput Functional Variant Screens via in Vivo Production of Single-Stranded DNA. *Proc. Natl. Acad. Sci.* **2021**, *118* (18), e2018181118. https://doi.org/10.1073/pnas.2018181118.

(45) Zhao, B.; Chen, S.-A. A.; Lee, J.; Fraser, H. B. Bacterial Retrons Enable Precise Gene Editing in Human Cells. *CRISPR J.* **2022**, *5* (1), 31–39. https://doi.org/10.1089/crispr.2021.0065.

(46) Lopez, S. C.; Crawford, K. D.; Lear, S. K.; Bhattarai-Kline, S.; Shipman, S. L. Precise Genome Editing across Kingdoms of Life Using Retron-Derived DNA. *Nat. Chem. Biol.* **2022**, *18* (2), 199–206. https://doi.org/10.1038/s41589-021-00927-y.

(47) Rhee, H.-W.; Zou, P.; Udeshi, N. D.; Martell, J. D.; Mootha, V. K.; Carr, S. A.; Ting, A. Y. Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science* **2013**, *339* (6125), 1328–1331. https://doi.org/10.1126/science.1230593.

(48) Stringer, C.; Wang, T.; Michaelos, M.; Pachitariu, M. Cellpose: A Generalist Algorithm for Cellular Segmentation. *Nat. Methods* **2021**, *18* (1), 100–106. https://doi.org/10.1038/s41592-020-01018-x.

(49) Pachitariu, M.; Stringer, C. Cellpose 2.0: How to Train Your Own Model. *Nat. Methods* **2022**, *19* (12), 1634–1641. https://doi.org/10.1038/s41592-022-01663-4.

(50) Dunn, K. W.; Kamocka, M. M.; McDonald, J. H. A Practical Guide to Evaluating Colocalization in Biological Microscopy. *Am. J. Physiol. Cell Physiol.* **2011**, *300* (4), C723-742. https://doi.org/10.1152/ajpcell.00462.2010.

(51) Van Rossum, G.; Drake, F. L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, 2009.

(52) Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; Ivanov, P.; Avila, D.; Abdalla, S.; Willing, C.; Jupyter development team. Jupyter Notebooks – a Publishing Format for Reproducible Computational Workflows; Loizides, F., Scmidt, B., Eds.; IOS Press, 2016; pp 87–90. https://doi.org/10.3233/978-1-61499-649-1-87.

(53) Summerfield, M. Rapid GUI Programming with Python and Qt: The Definitive Guide to PyQt Programming. **2007**.