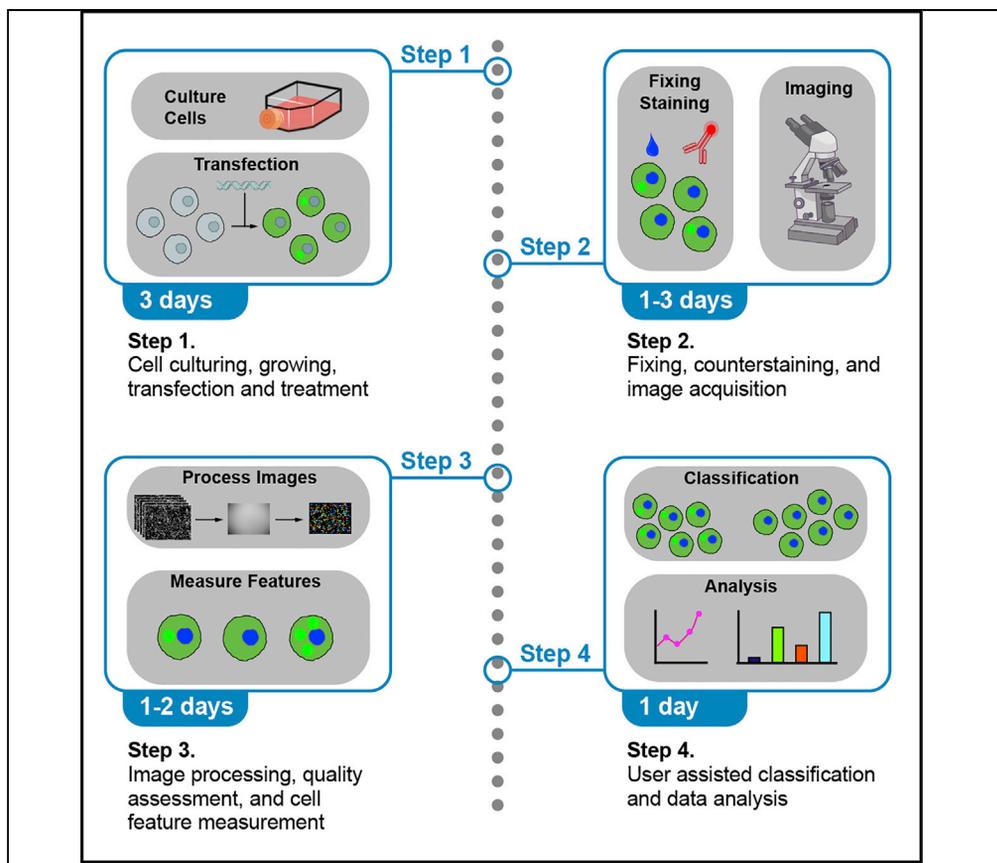


Protocol

Assessment of protein inclusions in cultured cells using automated image analysis



Proteinaceous inclusions are associated with neurodegenerative diseases and cell models are often used to determine genetic and chemical modifiers of their formation. This protocol involves the usage of automated microscopy and machine learning-based image analysis to accurately quantify the levels of protein inclusion formation in cultured cells from fluorescence microscopy images. This protocol is highly scalable and can be applied to a few images or large datasets.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Luke McAlary,
Victoria K.
Shepherd, Mine
Sher, Lauren J. Rice,
Justin J. Yerbury,
Neil R. Cashman,
Steven S. Plotkin

lmcalary@uow.edu.au
(L.M.)
steve@phas.ubc.ca
(S.S.P.)

Highlights
Accurately quantify
cells with protein
inclusions from image
sets

Generate high-
quality microscopy
image sets for
analysis

Applicable to small
datasets or high-
throughput image-
based screens

McAlary et al., STAR Protocols
3, 101748
December 16, 2022 © 2022
<https://doi.org/10.1016/j.xpro.2022.101748>



Protocol

Assessment of protein inclusions in cultured cells using automated image analysis

Luke McAlary,^{1,2,3,4,5,6,*} Victoria K. Shephard,^{1,2} Mine Sher,^{3,4} Lauren J. Rice,^{1,2} Justin J. Yerbury,^{1,2} Neil R. Cashman,⁴ and Steven S. Plotkin^{3,5,7,*}

¹Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, NSW 2522, Australia

²Molecular Horizons and School of Chemistry and Molecular Bioscience, Faculty of Science, Medicine and Health, University of Wollongong, Wollongong, NSW 2522, Australia

³Department of Physics and Astronomy, The University of British Columbia, Vancouver, BC V6T 1Z1, Canada

⁴Djavad Mowafaghian Centre for Brain Health, The University of British Columbia, Vancouver, BC, Canada

⁵Genome Science and Technology Program, The University of British Columbia, Vancouver, BC V6T 1Z1, Canada

⁶Technical contact

⁷Lead contact

*Correspondence: lmcalary@uow.edu.au (L.M.), steve@phas.ubc.ca (S.S.P.)
<https://doi.org/10.1016/j.xpro.2022.101748>

SUMMARY

Proteinaceous inclusions are associated with neurodegenerative diseases and cell models are often used to determine genetic and chemical modifiers of their formation. This protocol involves the usage of automated microscopy and machine learning-based image analysis to accurately quantify the levels of protein inclusion formation in cultured cells from fluorescence microscopy images. This protocol is highly scalable and can be applied to a few images or large datasets. For complete details on the use and execution of this protocol, please refer to McAlary et al. (2022).

BEFORE YOU BEGIN

In this protocol we use transient overexpression of amyotrophic lateral sclerosis (ALS)-associated variants of superoxide dismutase-1 (SOD1) protein fused to enhanced green fluorescent protein (EGFP). ALS-associated mutants of SOD1 form inclusions at a propensity that is dependent on the mutation (McAlary et al., 2019). This protocol was created with the assumption that the user has generated a mammalian expression construct for their protein-of-interest (POI). Below are some considerations that users should consider prior to starting with experimentation.

When examining protein inclusion formation in cultured cells, we recommend transiently overexpressing aggregation-prone proteins to induce inclusion formation. Stable integration of transgenes or tagging of genomic loci results in low expression that does not result in inclusion formation due to high-level expressing cells dying prior to, or during, clonal selection.

If genomic tagging or stable integration is required, inclusion formation can be induced in these lines by using compounds that inhibit protein homeostasis (proteostasis) pathways. However, use of proteostasis impairing drugs can lead to uncertain results when screening for compounds or genes that prevent or reduce inclusion formation, due to the confounding effects between compounds of interest and altered protein homeostasis.

A large palette of fluorescent protein tags exists, as do many affinity tags. Affinity tags have the advantage that they are small and the spectral properties of the sample can be chosen by using



secondary antibodies with different fluorophore conjugations. However, some affinity tags can become inaccessible in inclusions. Fluorescent proteins are considerably more bulky than affinity tags and can alter normal protein function. However, they become incorporated and remain detectable by fluorescence microscopy in inclusions. The user should choose between affinity tags and fluorescent tags, as well as the terminus for tagging, on the basis of their knowledge of their POI and empirical determination that a tag does not affect function.

If selecting a fluorescent protein (FP) as the tag, consult databases (<https://www.fpbases.org/>) (Lambert, 2019) to determine key properties of the fluorophore. Key properties include the oligomerization tendency, brightness, excitation/emission wavelengths, and maturation. True monomeric FPs or those with a low association propensity are desirable to prevent FP-associated oligomerization. The FP should be bright so the user can reduce excitation light levels to avoid bleaching during imaging. The spectral properties of the FP should match the capabilities of microscopes the user has access to. The maturation of the FP should be relatively fast (<1 h) for temporal accuracy in quantifying protein expression.

If the user has interest in co-expression of the POI with other proteins that may affect inclusion formation, it is advised that the user consider the above points but also to choose FPs with excitation and emission spectra adequately distanced from each other. This aids in later quantification by avoiding spectral overlap that can lead to artifacts in data. The user should consult a microscope technician to determine the excitation sources and filter sets available to them.

Installation of software for analysis

⌚ Timing: 1 h

Image analysis software is integral to this method. Here, we detail the basic software that users will require to fulfill the protocol. Importantly, we recognize that a great deal of image acquisition is performed by non-experts in image analysis. Therefore, we provide a simple and accessible method to process and analyze image data that does not require any proprietary software or high levels of user-expertise in coding. Each software package can be installed on either Windows or Mac. We recommend that users contact their information technology experts if they have trouble installing any of the software.

1. Download and install the relevant version of 'Fiji Is Just ImageJ' (FIJI) (Rueden et al., 2017) for your operating system from the ImageJ website (<https://imagej.net/software/fiji/>).
 - a. Update FIJI.
 - i. Go to Help > Update...
 - ii. Follow the prompts until your version of FIJI is up to date.
 - b. Install the StarDist plugin.
 - i. Go to Help > Update...
 - ii. In the ImageJ Updater window click *Manage update sites*.
 - iii. Search the list of plugins and activate both *CSBDeep* and *StarDist* (Schmidt et al., 2018). Click close.
 - iv. In the ImageJ updater window, click apply changes and follow the prompts.
2. Download and install CellProfiler (McQuin et al., 2018) for your operating system (<https://cellprofiler.org/releases>).

Note: Sometimes installation of CellProfiler requires updating JAVA software on your computer. If this is necessary, contact your local information technology department to aid you if you do not know how.

3. Download and install CellProfiler Analyst (Dao et al., 2016) for your operating system (<https://cellprofileranalyst.org/releases>).

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Subcloning Efficiency™ DH5α Competent Cells	Thermo Fisher Scientific	Cat# 18265017
Chemicals, peptides, and recombinant proteins		
DMEM/F12 Powder	Thermo Fisher Scientific	Cat# 12500096
Opti-MEM™ Reduced Serum Medium	Thermo Fisher Scientific	Cat# 31985062
Fetal Bovine Serum (FBS) – cannot be substituted if culturing the cell lines used in this manuscript.	Gibco	Cat# 10099-141
L-Glutamine	Gibco	Cat# 25030-149
Paraformaldehyde Powder	Sigma	Cat# P6148
Ethylenediaminetetraacetic acid (EDTA)	Sigma	Cat# E9884
Triton X-100	Sigma	Cat# T8787
Tris-base	Sigma	Cat# 10708976001
n-Propyl gallate	Sigma	Cat# 02370
Glycerol	Sigma	Cat# G5516
Hoechst 33342, Trihydrochloride, Trihydrate - 10 mg/mL Solution in Water	Thermo Fisher Scientific	Cat# H3570
ActinRed™ 555 ReadyProbes™ Reagent (Rhodamine phalloidin)	Thermo Fisher Scientific	Cat# R37112
Critical commercial assays		
TransIT-X2 Transfection Reagent	Mirus Bioscience	Cat# MIR 6000
ViaFect Transfection Reagent	Promega	Cat# E4981
GeneJET Endo-Free Plasmid Maxiprep Kit	Thermo Fisher Scientific	Cat# K0861
GeneJET Plasmid Miniprep Kit	Thermo Fisher Scientific	Cat# K0502
Experimental models: Cell lines		
Human: U-2 OS (Human, Osteosarcoma)	Cell Bank Australia	Cat# 92022711
Murine: NSC-34 (Mouse, Hybrid spinal cord motor neurons fused with mouse neuroblastoma cells N18TG2)	(Cashman et al., 1992)	N/A
Recombinant DNA		
pEGFP-N1-SOD1 vectors	(Turner et al., 2005) This lab	N/A
Software and algorithms		
CellProfiler	(McQuin et al., 2018)	https://cellprofiler.org/releases
CellProfiler Analyst	(Dao et al., 2016)	https://cellprofileranalyst.org/releases
ImageJ (FIJI)	(Rueden et al., 2017)	https://imagej.net/Fiji.html#Downloads
StarDist Plugin	(Schmidt et al., 2018)	https://imagej.net/StarDist
Code for automating StarDist processing via folders	(This work)	https://github.com/PlotkinLab/Stardist-BATCH https://github.com/ljr979/Stardist-BATCH https://doi.org/10.5281/zenodo.7017858

MATERIALS AND EQUIPMENT

Software and microscope

ImageJ

ImageJ is an open source image processing program. We recommend that the user install ImageJ by using the FIJI (FIJI is just imageJ) package. FIJI contains ImageJ with multiple add-ons and plugins that are useful for general image processing. FIJI can be found here (<https://imagej.net/Fiji.html>) or follow the above instructions. For help with general Fiji operation, search and post at forum.image.sc.

StarDist

StarDist is an ImageJ plugin that utilizes a trained neural network to identify nuclei in fluorescence microscopy images. StarDist can be installed through the ImageJ update feature by following the instructions above or here (<https://imagej.net/plugins/stardist>).

CellProfiler

CellProfiler is a powerful and accessible image-processing program that has various features and modules that allows non-experts to easily scale their image analysis. We have had excellent results using version 3.1.9 of the software. CellProfiler and instructions for its installation can be found here (<https://cellprofiler.org/releases>) or follow the above instructions. Tutorials for modules used in this protocol can be found at <https://cellprofiler.org/tutorials>. Help for general operation is at forum.image.sc.

CellProfiler Analyst

CellProfiler Analyst is a program that allows users to train a machine learning-based classifier to recognize and categorize features and characteristics of interest at a single-cell level. CellProfiler Analyst can be found here (<https://cellprofileranalyst.org/releases>). Tutorials can be found at <https://cellprofileranalyst.org/tutorials>. Help for general operation is at forum.image.sc.

An epifluorescent microscope with an automated stage and autofocusing. High content imaging systems work very well for large sample numbers. Talk to your microscope technician for specifics on how to use such systems. Laser scanning confocal microscopes are a viable alternative to wide-field microscopes, however, image acquisition will be slower. If an automated stage or autofocusing method is not available, users can still manually acquire images. Data processing and analysis with the machine learning in our hands generally need ~300 unique cells per category for accurate identification. This number of cells is easily obtainable with manual imaging. Note that this number may change depending upon the strength of the aggregation phenotype.

Note: If users are encountering problems with software not reported on here, an excellent resource for asking for help are the image.sc forums (<https://forum.image.sc/>). These forums contain sub-forums dedicated to many different image analysis software and plugins.

Media for cell culture

Standard Growth Medium		
Reagent	Final concentration	Amount
Dulbecco's modified Eagle medium with F12 (DMEM/F12)	89%	44.5 mL
Fetal Bovine Serum (FBS)	10%	5 mL
L-Glutamine	1%	0.5 mL
Total	N/A	50 mL

Note: Standard growth medium can be stored at 4°C for one month.

Note: If using cells other than those listed here, check the suppliers recommended culturing conditions as formulations can vary depending on the cell type.

1 × PBS 0.5 mM EDTA pH (8.0)		
Reagent	Final concentration	Amount
1 × PBS	99%	49.5 mL
500 mM EDTA pH (8.0)	1%	0.5 mL
Total	N/A	50 mL

Note: This solution can be stored at room temperature for 2 years.

Note: The dissociating PBS is listed here to aid for the dissociation of highly adherent cell types. For some cell lines, use of PBS with EDTA is not necessary and either PBS or serum free medium should be used to wash prior to addition of Trypsin-EDTA solution.

1× PBS 0.1% (v/v) Triton X-100

Reagent	Final concentration	Amount
Triton X-100 (100%)	0.1%	100 µL
1× PBS	1×	99.9 mL
Total	N/A	100 mL

Note: This solution can be stored for 3 months at 4°C.

Alternatives: Plasma membrane permeabilization can also be achieved using Digitonin. However Triton X-100 permeabilizes the nuclear membrane too, making nuclear stains easier to apply.

Hoechst-33342 working solution

Reagent	Final concentration	Amount
Hoechst 33342, Trihydrochloride, Trihydrate - 10 mg/mL Solution in Water	1 µg/mL	1 µL
1× PBS	1×	10 mL
Total	N/A	10 mL

Note: This solution should be used immediately and not stored.

△ **CRITICAL:** Hoechst-33342 is a DNA-binding chemical and as such should be treated as potentially toxic.

Alternatives: Hoechst-33342 is excited with UV light and emits at blue wavelengths. Other DNA-binding fluorescent dyes are available that have excitation and emission properties across the light spectrum. An example is RedDot1, which is a far-red nuclear dye.

Anti-fade solution

Reagent	Final concentration	Amount
Tris-base	20 mM (pH 8.0)	0.24 g
n-propyl gallate	0.5% (w/v)	0.5 g
dH ₂ O	N/A	50 mL
99.9% Glycerol	50% (v/v)	50 mL
Total	N/A	100 mL

Note: When making this solution it is easier to dissolve the n-propyl gallate overnight at 37°C. It is also recommended that the solution be filtered through a 0.45 µm filter prior to storage at –20°C.

Note: Antifade solution is stored at –20°C shielded from light and is stable for 6 months. The glycerol content of this solution can be altered according to the user's needs between 50%–90%. A glycerol content of 90% should be used if the user is obtaining images using glycerin or oil immersion objectives.

Alternatives: Commercial antifade solutions are available. We provide this one due to the user being able to alter the properties of the buffer at their own discretion (e.g., the user can change the refractive index of the solution by varying glycerol concentrations).

4% (w/v) PFA solution		
Reagent	Final concentration	Amount
Paraformaldehyde Powder	40 g/L	4 g
1× PBS	1×	100 mL
Total	N/A	100 mL

△ CRITICAL: PFA is a carcinogenic chemical and is highly toxic. PFA powders and solutions should be used in fume hoods at all times whilst wearing appropriate PPE. PFA waste should be discarded according to local guidelines.

Note: PFA solutions require a protocol of heating and filtration be followed to obtain high quality PFA solution for fixing that is free of particulate matter. In our hands, 4% PFA can be frozen at -20°C for up to a year and can be stored at 4°C for 1 month after it has been thawed. Do not freeze/thaw PFA multiple times.

Alternatives: Cell fixation can also be achieved using ice-cold methanol or ethanol. This method dehydrates cells and substantially alters protein conformation leading to fluorescent proteins being unable to fluoresce. This method is only recommended if the user cannot safely use PFA. Methanol is also toxic and should be used according to local safety guidelines.

STEP-BY-STEP METHOD DETAILS

We begin this protocol by detailing the culturing of cells for transfection. We then describe the fixation and imaging of cells using microscopes. Following this, we describe the processing of images through ImageJ and CellProfiler to accurately identify nuclei and segment cells respectively. Finally, we describe the usage of CellProfiler Analyst to classify cells and determine the number of cells containing inclusions. There are some optional steps in the protocol that depend on how the initial experimental design is performed by the user.

The overall protocol can be seen below in [Figure 1](#).

Cell culture, plating, and transient transfection

⌚ **Timing:** 3 days

We carried out all cell culture according to standard protocols that can be found from cell bank databases (<https://www.lgcstandards-atcc.org/>). NSC-34 and U-2 OS cells were maintained under humidified conditions at $37^{\circ}\text{C}/5\% \text{CO}_2$ in DMEM/F12 supplemented with 10% fetal bovine serum (FBS) and 2 mM L-glutamine.

Note: We recommend that the user chooses a cell line on the basis of its ease of transfection, ability to recapitulate the aggregation phenotype of the POI, and morphology being simple (flat, large, adherent, lack of extensions). If you do not know what cell line might be best, we recommend U-2 OS cells due to their large size, flat morphology, high transfection efficiency, and high transgene expression levels.

1. Thawing cells from cryogenic storage.
 - a. Pre-warm to 37°C :
 - i. 10 mL of serum free culture medium per vial of cells.

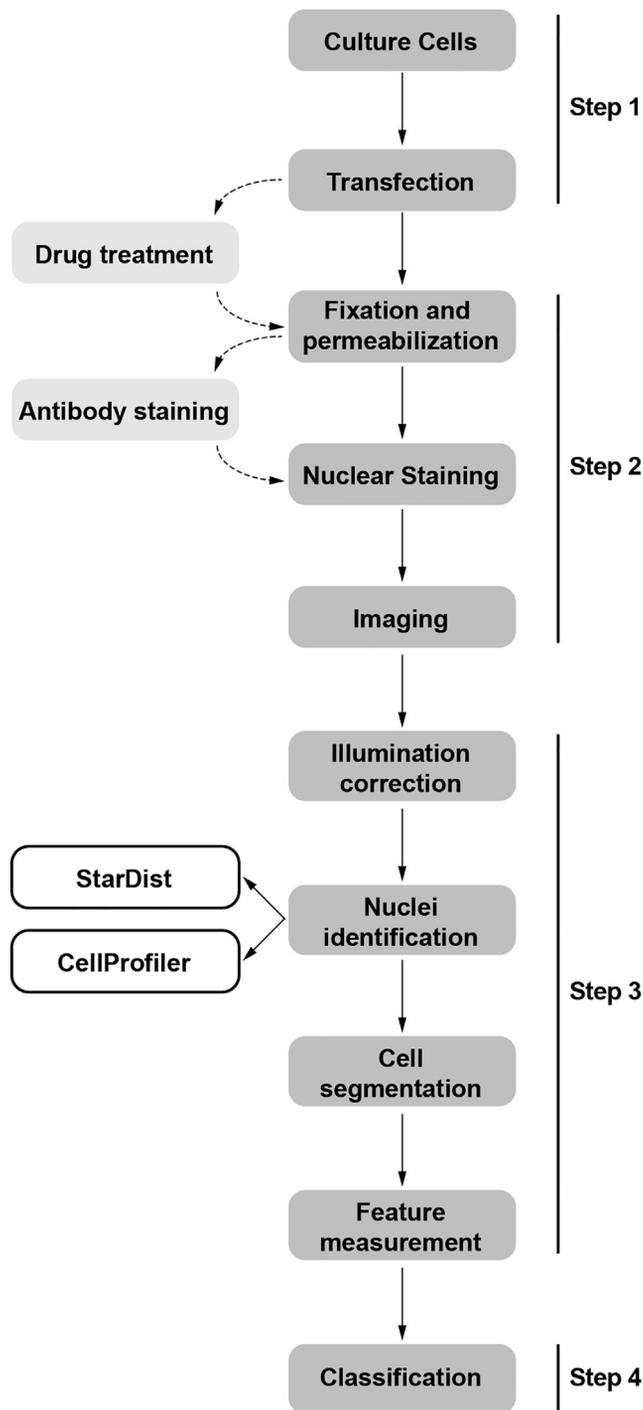


Figure 1. Overview of the workflow for identification of cultured cells containing protein inclusions

Light gray boxes denote optional steps. White boxes denote that the step can be achieved using either CellProfiler or StarDist.

- ii. a culture flask's volume of culture media supplemented with FBS (e.g., 20 mL for a T-75 flask).
- b. Once the media is warmed, remove the cells from the cryogenic storage and place the vial into a 37°C water bath.

△ **CRITICAL:** Once cells begin to defrost in the water bath, work quickly to perform the next steps so that the cells spend the least amount of time in contact with high DMSO concentrations in the cryogenic preservation medium.

- c. Gently swirl the vial during the defrosting cells until there is a small particle of ice remaining.
- d. Spray the vial with ethanol to sterilize it, and transfer the contents to 10 mL of warm serum medium in a 50 mL conical tube.
- e. Centrifuge the conical tube at an appropriate *g*-force for 5 min to pellet your cells (e.g., 300 × *g* for NSC-34 cells, 150 × *g* for U-2 OS cells).
- f. Discard the supernatant and resuspend the pellet in 500 μL of serum-supplemented culture medium.
- g. Transfer the resuspended cells into a culture flask containing an appropriate volume of serum-supplemented culture medium.
- h. Allow 2–3 passages of cells prior to performing experiments with them.

△ **CRITICAL:** All mammalian cell culture should be performed in biosafety cabinets with sterile reagents and equipment.

2. Plating cells.

- a. Pre-warm to 37°C:
 - i. 500 mL 1 × PBS 0.5 mM EDTA.
 - ii. 50 mL of serum free culture medium per T-75 flask of cells.
 - iii. 25 mL of Trypsin-EDTA dissociation solution.
 - iv. An appropriate volume of culture media supplemented with FBS (a good rule is 12 mL per multi-well plate of cells).
- b. View your cells under a microscope to determine the confluency and health.
 - i. Use this formula to calculate volumes for plating OR use a hemocytometer to specifically count cell numbers.

$$x = \frac{(\text{Stock Vessel Confluency}) \times (\text{Surface Area of Stock Vessel})}{(\text{Desired Confluency in Plate}) \times (\text{Surface Area of Single Well in Plate})}$$

$$\text{Volume of Media per Well} = \frac{\text{Resuspension Volume}}{x}$$

- c. Discard growth medium from the cultured cells and wash the cells carefully twice with ~10 mL of 1 × PBS 0.5 mM EDTA.
- d. Wash the cells once with ~10 mL of serum free culture medium.
- e. Add an appropriate volume of Trypsin-EDTA dissociation solution to the cells (e.g., 5 mL for a T-75 flask) and incubate for 5 min.
- f. Check the cells under the microscope to see if they have dissociated from the flask bottom. If they are still adhered, you can lightly tap the flask side to release them.
- g. Pour the cell suspension into a sterile conical 50 mL tube.
- h. Wash the stock vessel with ~10 mL serum free culture medium twice and pour each wash into the tube containing the cell suspension.
- i. Pellet the cell suspension via centrifugation at an appropriate *g*-force for 5 min (e.g., 300 × *g* for NSC-34 cells or 150 × *g* for U-2 OS).
- j. Discard the supernatant and gently but thoroughly resuspend the cells into 1 mL (resuspension volume) of growth media.
- k. Dilute the cells for hemocytometer counting (1 in 5 dilution) here or proceed via confluency calculations.
- l. Dilute the needed volume of cell resuspension into an appropriate volume for plating (this information will depend on the user's preferences).

m. Plate the cells and place them into the incubator.

Note: We recommend that the user plate into wells with relatively large surface areas so that a large pool of cells is available for transfection (i.e., a well of a 6-well plate). This cell pool can then be re-plated following transfection to decrease intra-assay variability that occurs from transfection.

3. Transfecting and re-plating cells.

- a. View plated cells under a microscope to ensure correct confluency and high viability (> 90% viable cells).
- b. Equilibrate ~10 mL Opti-MEM media to room temperature.
- c. Defrost frozen plasmid stocks to room temperature.
- d. Carry out transfection reagent protocol according to the manufacturer's instructions using your predetermined optimized DNA:reagent ratio.

Note: For TransIT-X2 transfection reagent the manufacturer's protocol can be found here (<https://www.mirusbio.com/products/transfection/transit-x2-dynamic-delivery-system>).

- e. Add Transfection mixture to cells and incubate cells for 24 h.
- f. Following transfection incubation for 24 h, re-plate the cells into multi-well plates according to the above protocol for plating cells (be sure to account for changes in cell number and media volume).
- g. The user is free to treat cells with drugs if doing so at this point.

Note: Some cell lines are not amenable to replating following transfection and will not re-adhere to tissue culture vessel. In these cases, transfect well-by-well.

Note: There are many different methods of transfection available. The user should perform a search of literature and commercial sources to determine what transfection method/reagent may be the most appropriate for their experiment. The best option is empirical testing of reagents in the users own hands. Transfections should result in a substantial proportion of the cell population expressing the transgene and minimal toxicity to cells. For example, we find that the TransIT-X2 reagent (Mirus Bio) is good for NSC-34 transfection but results in high levels of toxicity in U-2 OS cells. Therefore, for U-2 OS cells we use ViaFect (Promega) or Lipofectamine LTX reagent (Thermo Fisher Scientific).

Note: Transfection efficiencies between cell lines are not always the same and must be optimized for each plasmid and transfection method/reagent. We highly recommend that the user perform a transfection optimization to ensure the best transfections before imaging the cells. Instructions for transfection optimization are found in transfection reagent manufacturer protocols.

Note: Test transfections and optimization transfections can be performed using plasmids purified via miniprep kits. Once optimizations have been carried out, we recommend using endotoxin-free plasmid preparation kits to ensure optimal transfection and low toxicity of plasmid transfections.

Note: Be sure that your imaging vessel (multiwell plate, chamber slide, dish, etc) has the appropriate bottom thickness for the objective lens(es) you will use and can be held on the microscope stage. Plastic-bottomed plates (usually 1 mm thick bottom) are too thick for high numerical aperture (NA) air objectives and immersion objectives to image the plane at which the cells adhere (expect a 170 μm thick bottom). If in doubt, consult with your microscope technician.

Fixing and counterstaining transfected cells

⌚ Timing: 1 h

This phase of the protocol is aimed at carrying out the fixation of the cells in a manner that retains cellular morphology and provides the least disruption to fluorescent protein qualities. This is achieved by fixing the cells in warm PFA solution in growth media. We find that cold PFA solution, and sometimes warm neat PFA solution, significantly alters the morphology of neuronal cell types. By using warm PFA solution mixed with growth medium, excessive pH and temperature changes are avoided, leading to the maintenance of cellular morphology. The outcome of this step should be fixed and stained cells of a quality that is acceptable for fluorescence microscopy.

Note: If the user thinks that their image acquisition will result in long exposure times to the samples (potentially caused by software-based autofocus or low-quality cameras) they can utilize anti-fade solutions to maintain fluorophore brightness throughout image acquisition. We have provided a homemade recipe in the [materials and equipment](#) section above.

4. Fixing cells.

- a. Pre-warm 4% PFA solution to 37°C.

⚠ CRITICAL: PFA is a toxic chemical and should always be used inside a fume hood. Discard PFA solutions according to your local hazardous waste disposal guidelines.

- b. Add 1 volume of warmed 4% PFA solution per well of the plate containing transfected cells (e.g., if the cells are cultured at 100 μ L per well of a 96-well plate, add 100 μ L of 4% PFA solution to result in a 2% PFA final concentration).
- c. Incubate the cells for 5 min at room temperature.

Note: Cover your cells from light by placing them in an opaque container or wrapping them in foil. This prevents unwanted fluorescent protein bleaching.

- d. Aspirate the medium off the cells and discard appropriately as PFA is present.
- e. Carefully add 1 volume of 4% PFA solution to each well in the plate and incubate at room temperature for 15 min.

Optional: Media can be quickly removed from multi-well plates by flicking the plate upside down into a large waste container. We find that this is quicker and results in less cell detachment compared to aspiration with a pipette.

Optional: PFA fixation can cause increased autofluorescence from samples. Good practice is to quench remaining PFA by addition of 0.1 M Tris (pH 8.0) for 5 min to prevent this.

5. Permeabilizing cells.

- a. Add 1 volume of 1 \times PBS 0.1% Triton X-100 to each well in the plate and incubate with gentle rocking for 5 min.
- b. Wash the cells three times with 1 \times PBS for 5 min per wash. If the user has a low-adherence cell line, washing may be done less often (1 \times).

Note: If performing antibody staining, at this point begin with an optimized protocol for your target antigen before proceeding to counterstaining.

6. Counterstaining.
 - a. Add 1 volume of Hoechst-33342 working solution to each well in the plate and incubate for 5 min with gentle rocking.
 - b. Wash the cells three times with 1 × PBS for 5 min per wash. If the user has a low-adherence cell line, washing may be done less often (1 ×).

Optional: The cytoplasm of cells can be counterstained as well to provide the user with an internal control for non-transfected cells. We typically achieve this using ActinRed™ 555 ReadyProbes™ Reagent. We typically stain cell cytoplasm prior to nuclear counter staining (between steps 2 and 3) according to the manufacturer's directions.

Note: The manufacturer's instructions for the usage of the ActinRed™ 555 ReadyProbes™ Reagent can be found here (<https://www.thermofisher.com/order/catalog/product/R37112>).

Optional: Discard the 1 × PBS and replace with 1 volume of antifade solution.

- c. Store the cells at 4°C protected from light until imaging. Imaging should occur within 3 days of staining.

Imaging of fluorescent cells and image quality determination

⌚ Timing: 1–3 h

This step in the protocol is to help the user establish best practices for fluorescence microscopy and imaging in plate-based formats. This can be achieved either manually by a practiced user at the microscope, or automatically using microscope controlling software with automatic focusing and acquisition. Here, we detail automated acquisition protocols for inverted microscopes. Each user will likely have a different microscope setup available to them. Therefore, we recommend that the user consult their microscope technician if they need help. We also recommend the user perform this step with an epi-fluorescent wide-field microscope, however, a laser scanning or spinning disk confocal microscope can also be used. The outcome of this step should be a set of high-quality fluorescence microscopy images that are acceptable for image analysis. Images that are acceptable for analysis are in focus, do not contain artefacts, and have fluorescence channels of adequate brightness (Buchser et al., 2012). Users should check their files manually for unacceptable images. Alternatively, previous work has also defined methods of automatically and quantitatively determining poor quality images (Bray et al., 2012). As with all quantitative image analysis, it can take a few iterations of acquisition and analysis to find the optimal settings for both.

7. Microscope and sample preparation.
 - a. Turn on the microscope, light source, camera, and any other associated equipment you require.
 - b. Place samples to be imaged in the same room as the microscope so that they can equilibrate to the same temperature as the microscope (30 min–1 h).
 - c. Ensure the light source is emitting at the lowest setting.

Note: Light sources are most likely to include a mercury lamp or light emitting diodes (LEDs). Emission power settings on a mercury lamp typically control all wavelengths of emission, whereas LED's at different wavelengths may require individual power control. Check with your technician.

- d. Ensure the camera is set to acquire images in the highest bit-depth available (typically 16-bit).
8. Image acquisition optimization.
 - a. Place the sample onto the stage.

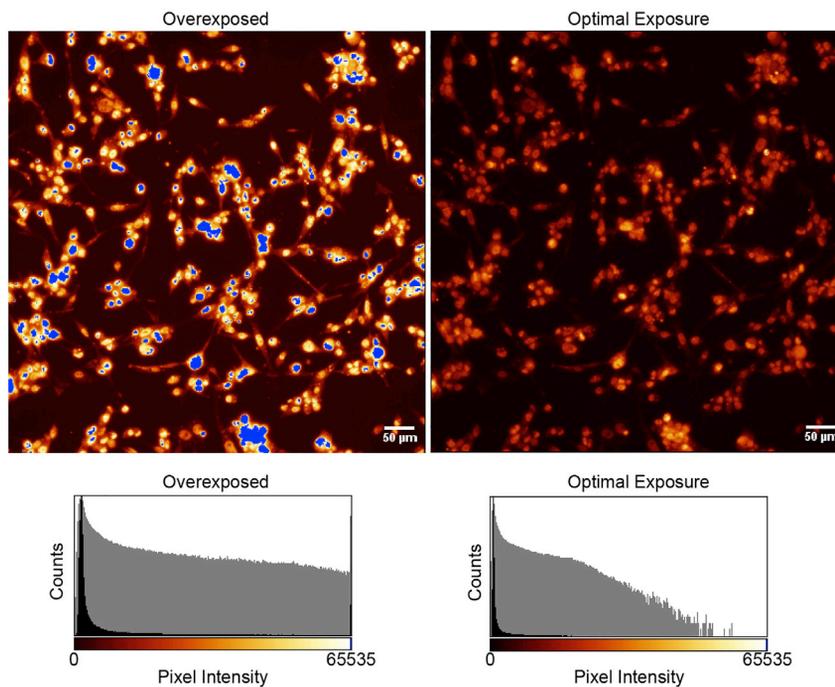


Figure 2. Examples of overexposed and optimally exposed images

(Left) an overexposed image of cells using a glow lookup table (LUT) where blue pixels are overexposed. (Right) an optimally exposed image where cells are easily visible but there are no overexposed pixels. Corresponding pixel intensities are plotted under respective images where gray are log-normalized values. Scale bar is 50 µm.

- b. Move the stage so that you are imaging in the sample well with the highest expected brightness of your FP-tagged POI (i.e., highest expression/inclusion amount).
- c. Find the plane where the cells are in-focus using brightfield or another transmitted light imaging method.

Note: If the cells are too difficult to find by transmitted light, or the microscope does not have the ability to image transmitted light, focus on the cells using the nuclear marker of choice (here we used Hoechst-33342). You may need to turn up the light source emission for your chosen marker. Nuclear counterstains are typically the brightest channel in stained samples.

- d. View the cells in each channel necessary to excite the fluorophores present in the sample and check that no channel has signal exceeding the detection limits of the camera (i.e., no saturated pixels).

Note: Changing the lookup table on your microscope to one that presents overexposed pixels is useful for determining if you are overexposing your sample (Figure 2). If the signal is too high, decrease exposure time or illumination intensity for that channel. If the signal is too low the user can increase excitation power, exposure time, or binning (each has a compromise: potential photobleaching, limitations in kinetics, and lower resolution, respectively). More details on image optimization have been previously published for camera-based systems (Lee and Kitaoka 2018; Salmon and Waters 2011) and point scanning confocal systems (Jonkman et al., 2020).

- e. Move around to different locations in this well of the sample to ensure that as few as possible cells have saturated pixels.

△ **CRITICAL:** It is highly important to make sure that as few as possible cells have signal above the maximum detection of the camera (saturated pixels). Saturated pixels will make data processing and analysis more difficult later for determining cells with inclusions.

Note: If a large set of samples makes the above optimization too time consuming or impossible, look for the brightest well in a particular sample and optimize the excitation and acquisition settings so that the camera is detecting ~60% brightness. This leaves enough collection space for most outlier samples in a large screen. Good practice is to have at least one well of each plate be your brightest sample (e.g., an EGFP only control).

- f. Switch between acquisition channels and ensure that focus offsets are applied to compensate for wavelength effects on focal plane.
- g. Set up the microscope acquisition software to automatically image a proportion (generally > 50% of the well area) of each well of the multi-well plate or sample in it using the settings that were just optimized.

△ **CRITICAL:** High-resolution in-focus images are necessary due to the usage of spatial measurements to accurately determine which cells contain inclusions later in the protocol. Setting a correct offset for each channel ensures that the plane of imaging in each channel is appropriate.

- h. Set up the autofocus method that the microscope uses. This will be either hardware or software-based. For software-based methods, ensure that you are autofocusing using the brightest channel, which is typically the nuclear marker.

Note: Plastic multi-well plates are not uniformly flat across the bottom. Each manufacturer will have a different molding process that will introduce variance in the plate bottom height and plastic thickness. Make sure your autofocus method searches a large enough distance in the axial plane to capture each well if you are using plastic-bottomed plates.

- i. Proceed to image your samples and save your data.

Note: A comprehensive file naming and saving strategy is highly important for the following steps of data export, data import for analysis, and metadata import for analysis. In general, as the complexity of the experiment increases, so should the file naming complexity. A multi-plate experiment carried out across days should have a file naming format similar to the following "DATE_PLATE ID_WELL_REGION_CHANNEL". If this information is not present in the image name, a hierarchy of folders containing the information are acceptable too, e.g., "C:\DATA\PLATE ID\WELL\REGION\CHANNEL". Both of these formats are compatible with regular expression metadata extraction.

9. Manually investigate your images for artifacts post acquisition.

Note: Sometimes images are not acquired properly due to focus issues or artifacts are present (e.g., dust or fibers in a sample) and these images need to be identified and removed from analysis. To do this, we recommend manually checking each image for grossly aberrant features such as particulate matter, incorrect image acquisition, or out focus images (Figure 3). It is good practice to record why an image is removed from the analysis group. Furthermore, in cases where a low number of cells (less than 300) is detected in an individual treatment, reimaging is suggested at a new location that contains no artifacts.

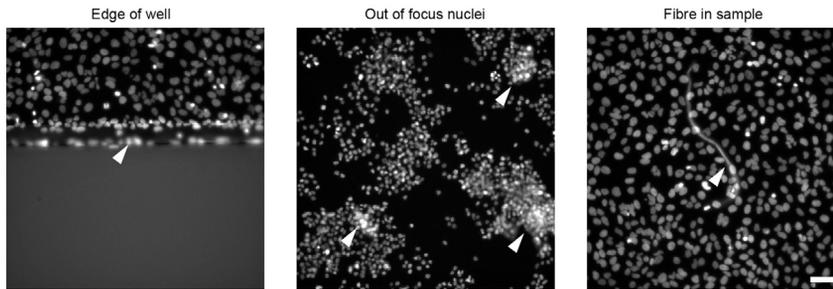


Figure 3. Examples of some common artifacts in image acquisition

Images such as these, as well as the corresponding images in other channels, should be removed from analysis. White arrows point to listed artifacts. Scale bar is 50 μm .

Data export for processing and analysis

⌚ Timing: 15 min per plate

File management is important for the analysis of large image sets. A key component of this management is the naming of files and folders so that they incorporate information about the experiment (metadata). Here, we provide a strategy for naming files so that several levels of metadata may be captured (e.g., date, plate, time-point etc). Other users may encounter differences with their microscope set up. Below we provide naming conventions that will allow for easy capture of metadata later when using CellProfiler.

10. Exporting metadata in a file name.

Note: The key information about an imaging experiment includes the date it was performed, the sample identifier, the well number (if multi-well plates/slides are used), image region, time-point (if time-lapse imaging is carried out), and the imaging channel. The naming convention in Figure 4A shows a naming strategy that incorporates all of this information. It is important to use underscores to separate this information as this makes later steps to capture metadata much easier. The information in this image is captured in CellProfiler using regular expression (a method of specifying a search pattern of text data). In Figure 4B an example of a regular expression is shown that would capture the 'date' from the filename shown in Figure 4A. In Figure 4C we provide two examples of regular expressions that fully capture the information about the experiment from the example filename provided in Figure 4A. One is verbose and the other simplified. The verbose regular expression is making use of specific expressions that capture specific information. Meanwhile, the simplified regular expression makes use of '.', where full stop captures any character and a star symbol dictates the capture of one or more characters. We recommend that users utilize the naming strategy and regular expressions shown here until they are familiar with these concepts.

⚠ CRITICAL: If you acquired 16-bit images, do not export 8-bit images as you will lose the dynamic range of fluorescence intensity data in each pixel. You must maintain your image bit-depth across steps in this protocol.

⚠ CRITICAL: It is essential that large image datasets contain information about the sample in each image. This allows for easy data handling in later steps and for the user to recall information about a specific image to examine if the image processing is accurate.

Note: Each microscope manufacturer has different file types that they use to store imaging data. Batch export options are typically available from most commercial microscopy software. Check with your technician.

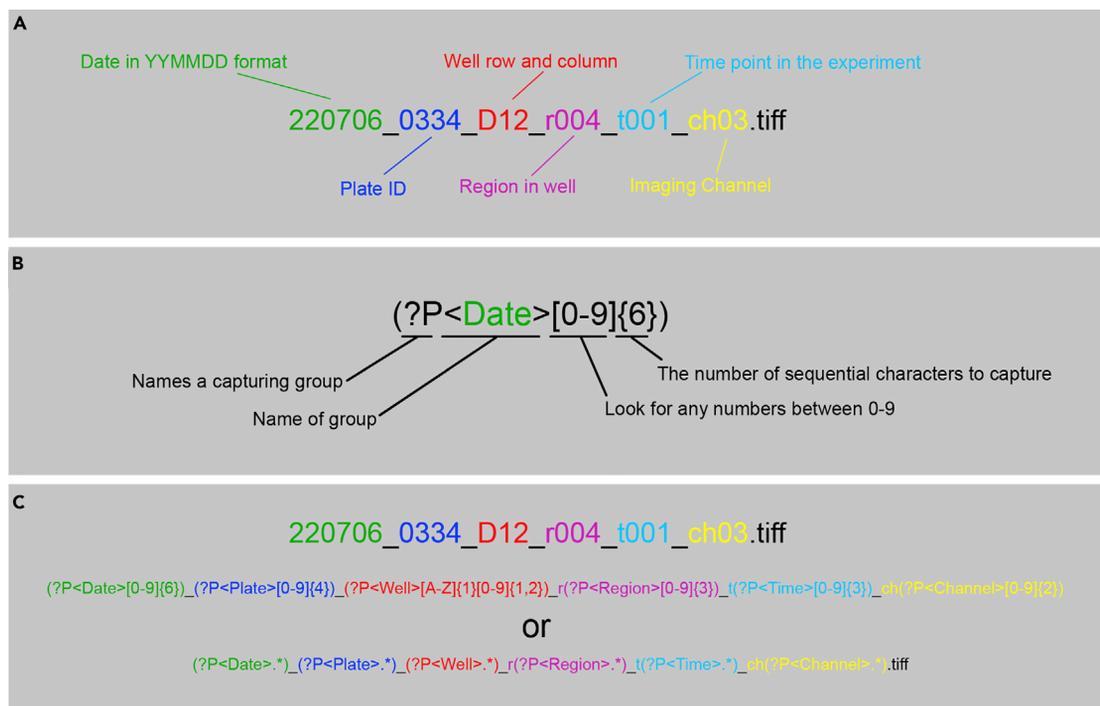


Figure 4. Example of file naming for effective capture of metadata via regular expression

(A) Middle of the panel is a filename that is written in DATE_PLATE_WELL_REGION_TIME_CHANNEL.tiff, where colors code for the respective information they give to the file.

(B) An example of a regular expression used to capture metadata from a filename, in this case the date.

(C) Two example methods of capturing metadata from the example filename. If users follow this file naming strategy, they should be able to simply copy and paste either of these regular expressions into CellProfiler to capture experiment metadata.

Generation of illumination corrected images

⌚ Timing: 30 min per plate

Every microscope and sample will have unique features and characteristics that result in uneven sample illumination. This is often observed as a gradient of intensity across an image. Correcting for uneven sample illumination is important for more accurate measurement of sample features such as fluorescence intensity. Some microscopes allow for onboard correction of uneven illumination (check with your technician), however, best practice is to perform illumination correction post-acquisition using a large dataset of images gathered from the same sample plate. Here we detail how to use CellProfiler software to generate illumination corrected images for each channel imaged. The outcome of this step should be illumination corrected images from the user's data (Figure 5).

11. Make a copy of your images for processing and analysis. This is due to raw data needing to be protected from accidental alterations or deletion.
12. Open CellProfiler.
13. Drag and drop the folder(s) containing your copied exported images into the CellProfiler UI.

Note: We recommend processing illumination correction in a plate-by-plate manner as often the variation may be unique to the plate itself. We find the simplest method to do this is by running illumination correction on individual plates, which is achieved by saving data in a format where images from separate plates have their own folders.

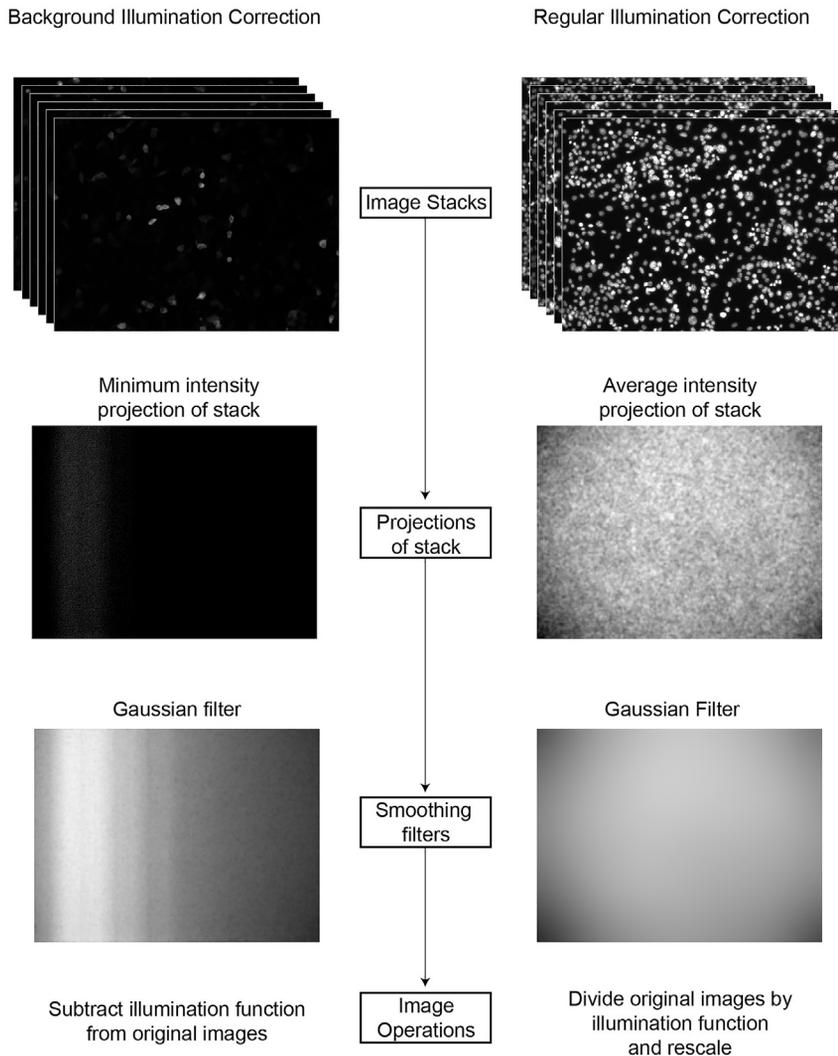


Figure 5. Example of the two methods of performing illumination correction

14. Extract metadata to your images in the *Metadata* module of CellProfiler ([Figure 6](#) shows the main CellProfiler window).

Note: If you have exported your images in a format that retained information from the acquisition parameters, you should be able to extract metadata via the use of regular expression (See step 7 and [Figure 4](#) for more details and the example pipeline [supplemental information](#)). Help windows within CellProfiler provide more information about regular expressions.

15. Establish the channel identifiers under the *NamesAndTypes* module.
 - a. Assign a name to "Images matching rules".
 - b. Alter the rules to capture your specific channels. For example, we capture the GFP channel of our file naming convention (P1_B2_1_GFP) by assigning the rule "File Does Contain GFP". This option will depend on the user's own naming convention.
16. Set up your image groups under the *Groups* module so that plates and channels are processed individually. Metadata Category "Plate".

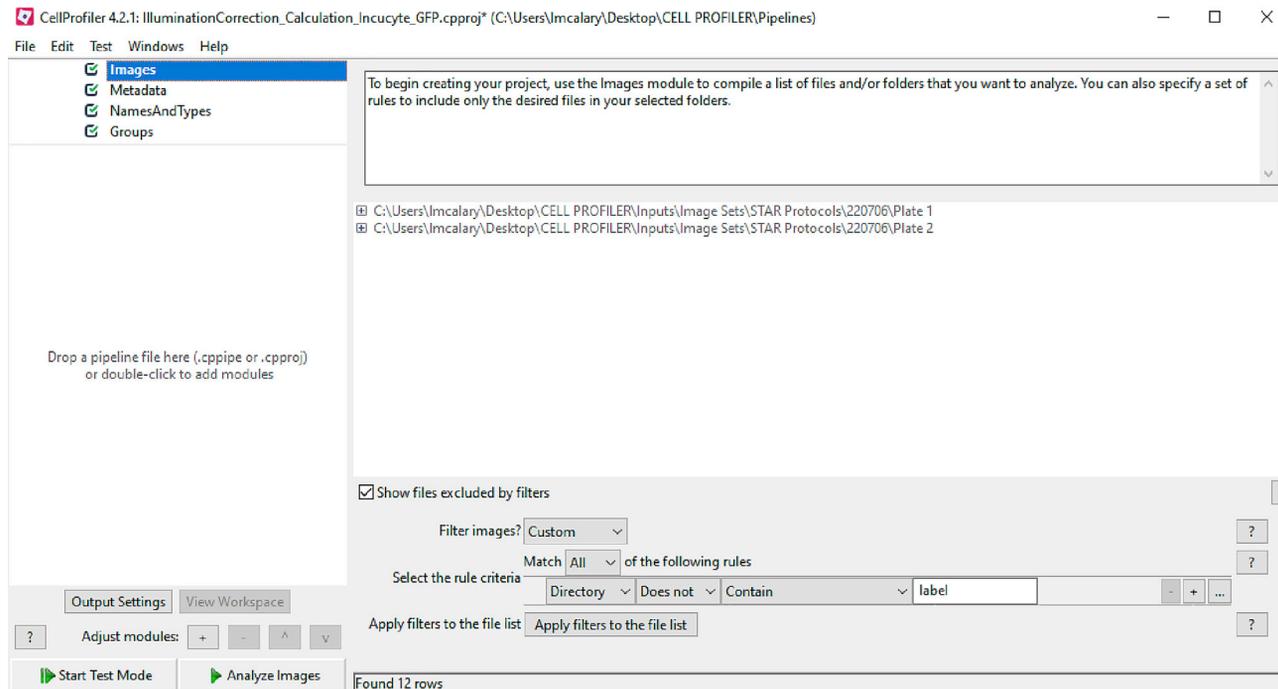


Figure 6. CellProfiler main window

The main window of CellProfiler contains four main modules that should be configured to make later steps easier. Please see [Data S1](#) for a downloadable example pipeline that shows the usage of each of these modules.

Optional: You do not have to group channels or images according to the above step (11) if you process one set of images (one channel from one plate) at a time. We sometimes find it quicker and easier to create illumination functions by only loading one channel at a time into CellProfiler rather than loading all images at one time. This is especially useful if you are working on outdated or slow computer hardware.

Note: The above steps (8–13) are basic steps that we recommend you follow for each CellProfiler Pipeline that you create.

17. Calculate the Illumination Functions for your image sets.
 - a. In the main CellProfiler window, in the empty space below the *Groups* module, right-click and go to "Image Processing > *CorrectIlluminationCalculate*".
 - b. In the *CorrectIlluminationCalculate* module, select your input images and choose between "Regular" or "Background" calculation.

△ CRITICAL: It is important that you understand where the majority of uneven illumination may be coming from. Images with very few bright objects are more likely to have uneven illumination as a background feature, whereas images with many bright objects are likely to have uneven illumination stemming from the objects. In the case of few bright objects, a "Background" calculation is appropriate, and in the case of many bright objects a "Regular" calculation is appropriate (Figure 5). For more information on illumination correction please see the tutorial here (<https://cellprofiler.org/examples>).

- c. If doing regular calculation, input the following options:
 - i. Dilate objects > No.
 - ii. Rescale illumination function > Yes.

- iii. Individual calculation or all > All First Cycle.
 - iv. Smoothing Filter > Gaussian.
 - v. Method to calculate filter size > Manual (you will need to choose a size large enough that little to no texture is apparent in the final illumination function. On a 2048 × 2048 pixel image, we typically use a filter size of 1024 pixels.
- d. If doing background calculation, input the following options:
- i. Block size > A value that captures background without an object in it.
 - ii. Rescale illumination function > No.
 - iii. Individual calculation or all > All First Cycle.
 - iv. Smoothing Filter > Gaussian.
 - v. Method to calculate filter size > Manual (you will need to choose a size large enough that little to no texture is apparent in the final illumination function.
18. Saving the illumination function image.
- a. In the main CellProfiler window, in the empty space below the *Groups* module, right-click and go to "File Processing > *SaveImages*".
 - b. In the *SaveImages* module, make sure you maintain your original image bit-depth and save the image from the first cycle.
 - i. Select the type of image to save > Image.
 - ii. Select method for constructing file names > From image filename.
 - iii. Select image name for file prefix > The name of your image that you defined earlier in the *NamesAndTypes* module.
 - iv. Output file location > Can be anywhere you want, but we typically export illumination functions to a subfolder within the folder that houses the images to be analyzed.

Optional: You can either create a new pipeline for applying the illumination correction or add the following to your current pipeline. Creating a new pipeline can sometimes be easier if the user is dealing with multiple datasets from different experiments that require different parameters to be applied.

19. Apply the illumination function.
- a. In the main CellProfiler window, in the empty space below the *Groups* module, right-click and go to "Image Processing > *CorrectIlluminationApply*".
 - b. Your input image will be the images you want to correct, and your illumination function is the image that was generated from the *CorrectIlluminationCalculate* module.
 - c. Subtract the illumination function if you generated a background function. Divide if you generated a regular illumination function.
 - d. Create a *SaveImages* module for each channel being illumination corrected and save the illuminated corrected images to a folder for later use.

StarDist identification of nuclei

⌚ Timing: 3 h

Accurate identification of nuclei in cell imaging datasets is essential for accurate downstream segmentation of cells and their compartments. There are many nuclei identification options available to the user, some of which are in CellProfiler software. The *IdentifyPrimaryObjects* module in CellProfiler is adequate for the identification of nuclei that are spaced apart and mostly uniform in fluorescence and size (instructions for the use of this module can be found on the CellProfiler <https://cellprofiler.org/examples>). Here, we detail the usage of the StarDist plugin for ImageJ (Schmidt et al., 2018), which is a much more powerful tool for accurate nuclei identification. StarDist is capable of accurately identifying and segmenting nuclei that are partly overlapping. The outcome of this step should be a set of images with highly accurate nuclei identification.

20. Import a small subset of illumination corrected nuclear channel images into ImageJ.
21. Activate the StarDist plugin.
 - a. Plugins > StarDist > StarDist 2D.
22. Optimize StarDist parameters for best detection.
 - a. Model = Versatile > fluorescent nuclei.
 - b. Output Type = Both.
 - c. Run StarDist. A new image and the ImageJ ROI-manager should open.
 - d. In the ImageJ ROI-manager, go to "Show All". This option shows the selected nuclei regions of interest on your raw image.
 - e. Assess the accuracy of the StarDist plugin by looking for cases where it did not identify nuclei or it combined two nuclei into one nucleus. Do this for at least 300 nuclei across your image set. Acceptable accuracy here is > 95%.
 - f. If the nuclei identification is substandard, alter other parameters to enhance accuracy. This will need to be empirically determined by the user. See the published work for details (Schmidt et al., 2018).
23. Once you have determined settings appropriate for accurate identification of nuclei using StarDist. You can run the process in batch mode using code provided in this work or you can generate your own code.
24. Download the StarDist-BATCH ImageJ macro from either '<https://github.com/ljr979/StarDist-BATCH>' or '<https://github.com/PlotkinLab/StarDist-BATCH>' and unzip the file to find the python file.
 - a. Drag and drop the file onto your ImageJ main bar to open it.
 - i. Make sure your language is set to python by using the 'Language' option at the top of the new window.
 - b. In the macro code, lines 9–11 can be altered to the parameters determined in the steps in 18 above, to result in the most accurate nuclei identification.
 - c. For image identification and export, there are two lines that require input from the user as directed.
 - i. Line 17:

```
input_folder = os.path.expanduser('C:\Data\Images\Experiment01\TileScan\')
```

Replace the example directory here with your own directory. Note that you should replace the backslashes with forward slashes and add an additional forward slash to the end. For example, 'C:/Data/Images/Experiment01/TileScan/'.

- ii. Line 20:

```
file_identifier = 'Nuclear channel image identifier goes here'
```

For example, file_identifier = 'DAPI'.

△ CRITICAL: It is very important that the user makes the correct inputs as directed above. If the slashes are not replaced, the script will not output any label images.

- d. Once you have told the macro where your images are located and what the nuclear channel identifier is, press the 'Run' button to execute the macro.
 - e. The output should be new images that are named as 'label_(your image name)'. For example, and image named "P1_B2_1_DAPI" will be "label_P1_B2_1_DAPI" after StarDist batch processing. For an example see [Figure 7](#) and the supplementary example pipeline.

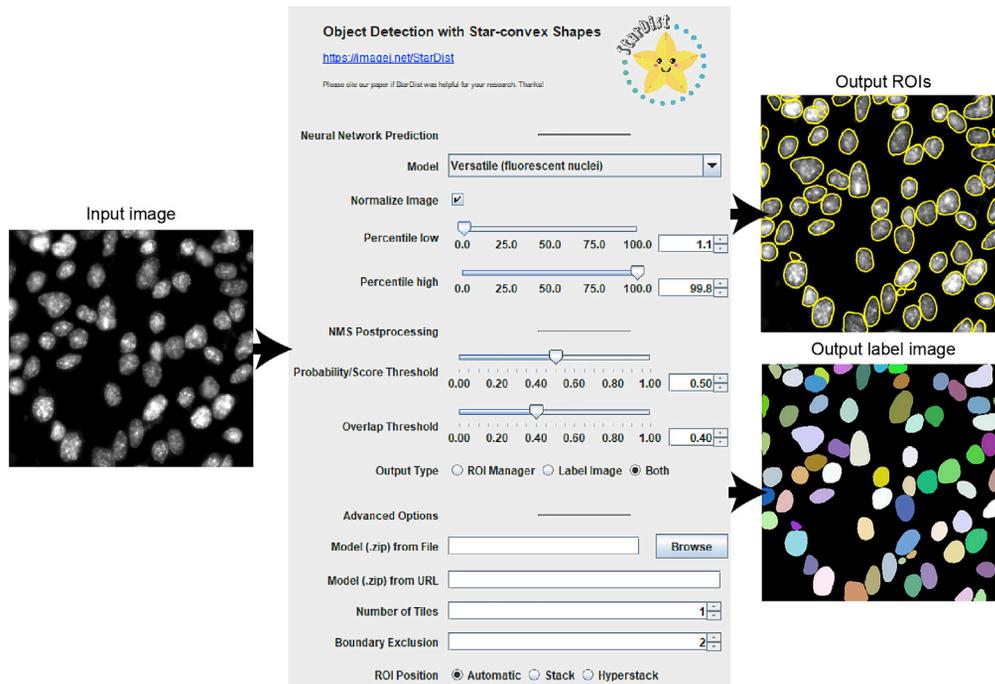


Figure 7. Example of StarDist identification of nuclei from fluorescent imaging

Images are loaded into the StarDist plugin and processed according to the settings shown in the middle panel. The output from this is regions of interest (ROIs) and a label image showing identified nuclei. Note that the automation proceeds through the ImageJ macro functions and this window is not viewed when batch processing.

Note: If the user wishes to change the prefix name given to the StarDist-processed images, they should alter the text **label** and **%s** here from line 40:

```
io.save(label, os.path.join(output_folder, "label_%s" % (name)).
```

Generating a CellProfiler pipeline for measurement of features associated with inclusion formation

⌚ Timing: 3 h

This step in the protocol is the generation of the main CellProfiler pipeline that both segments cells into nuclear and cytoplasmic compartments, as well as generating the feature measurements for the classification of cells. At this point, the user should have illumination corrected images for each channel of analysis and StarDist identified nuclei images. The main outcome of this step is accurate segmentation of cells, generation of excel spreadsheets with individual cell measurements, and generation of files necessary for input into CellProfiler Analyst. A general workflow is presented in Figure 8. See Data S1 to download an example pipeline for windows systems.

Optional: As stated above, you can identify nuclei using the CellProfiler *IdentifyPrimaryObjects* module. See here for more information (<https://cellprofiler.org/examples>).

Note: The exact parameters that are best to implement at each step in the pipeline are guidelines here. The user will have to empirically determine the parameters that best capture the cells within their image sets. Unless stated, all parameters for each module are default.

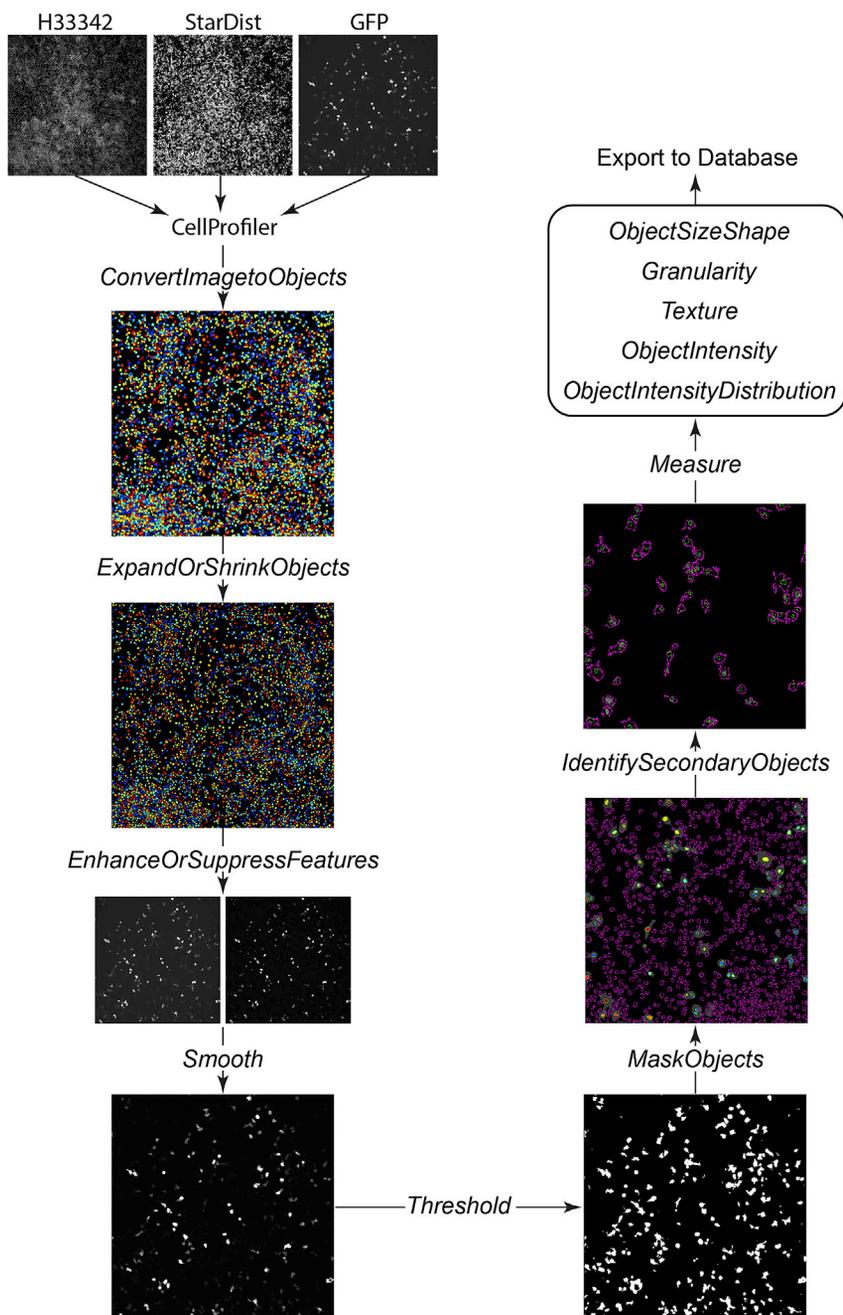


Figure 8. General workflow of image processing in CellProfiler

Note: The following step-by-step protocol assumes that StarDist output images are being used for nuclei detection.

25. Load Images into CellProfiler.
 - a. The images should be illumination corrected images from each channel and the StarDist identified nuclei images.
26. Setup the *Metadata* module for your image set (see steps 8–11 or the example pipeline for more information).
 - a. Capture Plate number, well number, sample identifier (e.g., mutant/drug) etc.

27. Set up *NamesAndTypes* module.
 - a. At a minimum, the user should have a nuclear channel, aggregating protein channel, and the StarDist images as separate names/types.

Note: Since StarDist batch processing results in files named similar to the input images, using “Does not contain” rules help delineate between original nuclear channel images and StarDist images.

Optional: The *Groups* module can be set up here but is not necessary for the following steps.

28. Add a *ConvertImageToObjects* module to the pipeline and change the below settings.
 - a. Input image = StarDist image.
 - b. Convert to Boolean image = No.
29. In some instances, StarDist will detect micro-nuclei in your images. These need to be removed from the analysis as they may confound the identification of cells. To do this add a *MeasureObjectSizeShape* module followed by a *FilterObjects* module.
 - a. Input objects = nuclei.
 - b. Filtering Mode = measurements.
 - c. Filtering Method = Limits.
 - d. Category = AreaShape.
 - e. Measurement = Area.
 - f. Filter only using a minimum value. You will need to empirically determine the value that removes micronuclei, but not actual nuclei in your experiments.
 - g. Keep removed objects as a separate set? No.
30. Add an *ExpandOrShrinkObjects* module to the pipeline and shrink the nuclei by at least one pixel. You may need to empirically determine the most appropriate number of pixels for shrinking that prevents the subsequent masking step from not removing untransfected cells.
31. Add an *EnhanceOrSuppressFeatures* module to the pipeline. This is to subtract unwanted background from the cells.
 - a. Input image = raw image of aggregating protein.
 - b. Enhance speckles using an empirically determined feature size. Generally you want to keep the size slightly above the size of the object you want to enhance, which in this case are entire cells.
32. Add a *Smooth* module to the pipeline.
 - a. Input image = image from *EnhanceOrSuppressFeatures*.
 - b. Smoothing method = Gaussian filter.
 - c. Smooth the image enough that there are no jagged edges on cells. We typically do 4 or 2 pixels smooths.
33. Add a *Threshold* module to the pipeline. This step creates a binary image for overlaying onto identified nuclei.
 - a. Input image = image from *Smooth*.
 - b. Threshold strategy = Global.
 - c. Threshold method = Minimum Cross-Entropy.
 - d. Log transform = yes.
34. Add a *MaskObjects* module to the pipeline. This step lets us identify truly transfected nuclei and enhance subsequent cytoplasm identification.
 - a. Objects to be masked = Shrunken Nuclei (from step 24).
 - b. Mask using an image.
 - c. Masking image = image from *Threshold* module.
 - d. Invert mask = no.
 - e. Handling of objects that are partially masked = remove.
 - f. Numbering of resulting objects = renumber.
35. Add an *IdentifySecondaryObjects* module to the pipeline. This step identifies the cell cytoplasm for each nuclei.
 - a. Input image = image from *EnhanceOrSuppressFeatures* (step 25).

- b. Input objects = objects from MaskObjects.
 - c. Method for identification = propagation.
 - d. Threshold Strategy = Global.
 - e. Thresholding method = Minimum Cross-Entropy.
 - f. Log transform = yes.
 - g. Discard secondary objects touching image border = yes.
36. At this point cell identification should be accurate. The user can screen their pipeline on a small set of images by adding an *OverlayOutlines* module followed by a *SaveImages* module.
37. Add the following modules to the pipeline and make sure they are measuring objects from *IdentifySecondaryObjects* (step 32) in your aggregating protein image (the original illumination corrected images that are NOT processed – i.e., smoothed or masked). To determine how to set the parameters within these settings, we recommend that users measure the size of aggregates using the CellProfiler measurement tool or ImageJ measurement tool. Generally setting spatial measurements to the average size of aggregates within cells is adequate for machine learning to differentiate features.
- a. MeasureGranularity, MeasureTexture, MeasureObjectIntensityDistribution, MeasureObjectSizeShape, MeasureObjectIntensity.

Note: Some of these measurement modules will contain options to measure Zernike features. Measurement of these features can be very computationally intensive. If the user is running this pipeline on a desktop PC for large datasets, we recommend turning the measurement of Zernike features off. If a small dataset is being analyzed (<10,000 images) this feature can remain on for enhanced classification accuracy.

38. Add an *ExportToDatabase* module and configure this.
- a. Database type = SQLite.
 - b. Create CellProfiler Analyst Properties file = Yes.
 - c. Objects to be used for locations = Objects from *IdentifySecondaryObjects*.
 - d. Classification type = Object.
 - e. Export measurements for all objects to the database = Select > objects from *IdentifySecondaryObjects*.
 - f. Export object relationships = no.
 - g. Auto-scale thumbnail pixel intensities = yes.
39. Run this pipeline on a subset of images that incorporates the breadth of expected aggregation phenotypes and evaluate accuracy. This is so the pipeline can be tested and enhanced for the larger image sets that a user might have.
40. Run the pipeline on your entire data set.

Using CellProfiler analyst to classify cells with inclusions

⌚ Timing: 1–3 h

This step details the usage of Analyst to train a classification model to recognize cells containing inclusions. Training of classifiers using machine learning requires careful assessment of the phenotypes that each cell class (no inclusions vs. inclusions) is presenting.

- 41. Load the properties file created from your CellProfiler pipeline into CellProfiler Analyst.
- 42. Open the Classifier module.
- 43. Fetch 50–100 random cells from the experiment.
- 44. Turn off the overlays of nuclear stains and StarDist so you can effectively see your aggregation protein channel in the thumbnails.
 - a. Click on the menu's for each channel and turn them off except for your channel of interest.

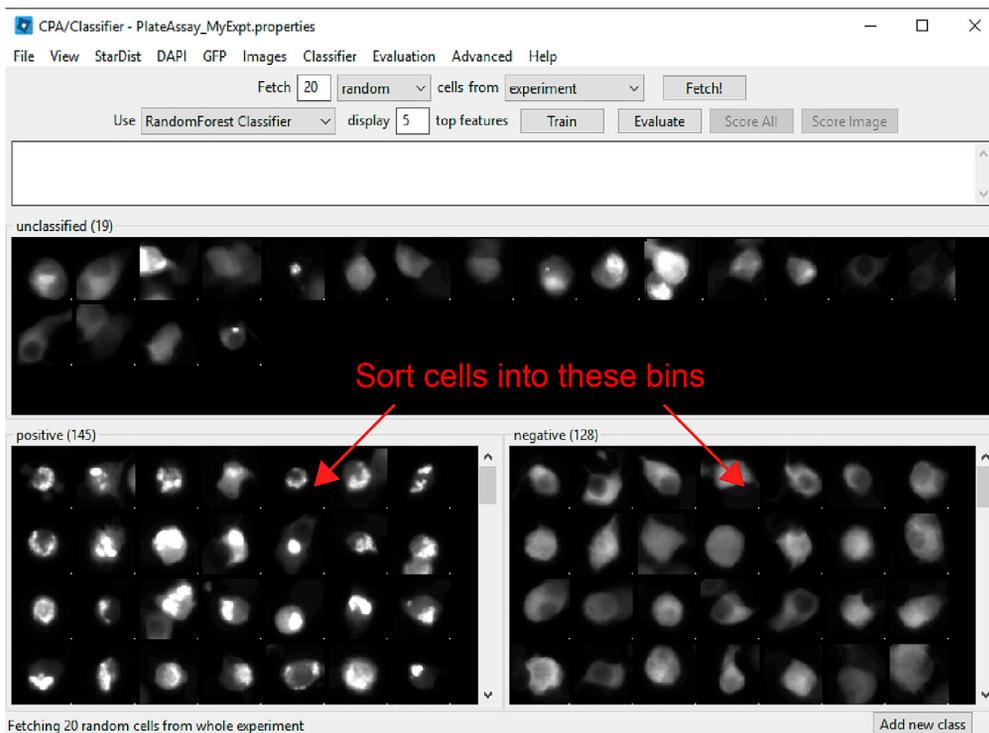


Figure 9. Classifier window for sorting cells into their respective classes

45. Sort cells by dragging and dropping from your requested set into classes on the basis of whether they contain inclusions or not (Figure 9). You can add more classes if needed by clicking the button in the bottom right corner.
 - a. Sort approximately 20 cells into each class. Fetch more cells if you run out of a certain phenotype.
46. Train using a classifier of your choice. We find that the random forest classification is the most accurate, however, fast gentle boosting can be fast and highly accurate if the training set is adequate in size and quality.
 - a. To train the classifier, click on “train” in the window and wait for it to calculate.
47. Request a number of cells from your *inclusions* class and sort them into their respective bins.
48. Request a number of cells from your *no inclusions* class and sort them into their respective bins.
49. Train the classifier again as necessary (see next step).

Note: Make sure to regularly save training sets if an error occurs. The option to save a training set is in the ‘file’ menu.

50. Repeat the above steps until you have obtained sufficient classification accuracy. Assess accuracy by requesting 100 cells of each class 3 times and scoring the percent that are correct. We only accept an accuracy of > 95% in both classes for our work.
51. Once the classifier is accurate, score all images in the data set by clicking ‘Score All’. This may take some time to calculate.
52. Export the data from the table and process in a program of your choice. We use excel followed by GraphPad PRISM but python scripts can be particularly helpful for this output type.

EXPECTED OUTCOMES

Depending on the experimental needs of the user, the outcome will be different. Overall, the user should obtain accurate classification of aggregates/inclusions in their cell populations.

For examples of this method in action to detect protein inclusions in cells, see [Figures 1, 2, and 3](#) for processed and analyzed data in [McAlary et al. \(2022\)](#). For an example of cell thumbnails, see supplementary figure 1 in [McAlary et al. \(2022\)](#).

QUANTIFICATION AND STATISTICAL ANALYSIS

Results should be quantified by the user according to best practices. Controls, both positive and negative should be present in each replicate of experiment undertaken. We typically use GraphPad PRISM software for our statistical analysis. Statistical analysis will depend on the type of assay the user is performing.

LIMITATIONS

Whilst this method is a robust and accurate way to measure protein inclusion formation in cells, some limitations apply. Namely, for the acquisition of large image datasets, the user will need access to an automated microscope. Lack of access does not preclude use of this protocol, but it does decrease the throughput by which a user can generate data.

Another limitation is that the cell type itself may be difficult to properly segment in the pipeline. We find that U-2 OS cells are far easier to accurately segment and identify than NSC-34 cells due to the morphological differences between cells (flat and large for U-2 OS vs rounded and small for NSC-34). Users will have to determine the best parameters for their specific experiments.

TROUBLESHOOTING

Problem 1

Poor transfection efficiency:

A good transfection efficiency is helpful in obtaining enough cells for analysis. Step 3.

Potential solution

Experiment with different transfection reagents for your cell line of choice or, if possible, change to a more easily transfected cell type. Perform transfection optimization using a greater range of DNA and transfection reagent ratios.

Problem 2

Dim fluorescent images:

Images must be bright enough for accurate classification of cells containing inclusions. Step 8.

Potential solution

First check your microscope light source to see if it is working properly. If the microscope light source is working properly, try binning the image, increasing the exposure time of imaging on the microscope, and/or increasing excitation light intensity. If the fluorescence remains dim at this point, consider using immunostaining to enhance the signal by using antibodies against your chosen fluorescent protein.

Problem 3

Inaccurate StarDist identification of nuclei:

Accurate nuclei identification is necessary for downstream processing of image sets. Step 20.

Potential solution

If using the default StarDist settings, try rescaling your images to be either larger or smaller. StarDist works optimally with nuclei that are approximately 90 pixels in diameter. This means if your nuclei are larger or smaller in pixel dimensions than this optimal size range, they may be over-segmented or under-segmented. You can rescale the StarDist output images accordingly once they are generated

by using the *Rescale* module in CellProfiler. For further information, please refer to the StarDist FAQ (<https://stardist.net/docs/faq.html>).

Problem 4

Images are out of focus:

In-focus images are necessary for the accurate segmentation and measuring of cell cytoprofiles. Step 7.

Potential solution

There are likely two main reasons that images of cells would be out of focus. The first is that the microscope focus strategy is not correctly established. The user should check that the focus strategy is capable of accurately capturing in focus images for their samples. Making focus strategies more accurate may make them take longer to image, but saves time from re-imaging later.

The second reason is that the optical section of the objective may not be appropriate for the cell type being imaged. Different cell types have different morphological characteristics (e.g., NSC-34 cells are round and grow sometimes in clumps). Using a lower magnification objective can increase the size of the optical section at the cost of XY resolution. Alternatively, the user can perform imaging with a confocal microscope and collect z-stack images (this will be slower, but more accurate for generating in-focus images).

Problem 5

Trouble deriving metadata from image files:

Proper collection of image metadata (plate identity, well identity, fluorescence channel etc) is necessary for the accurate assessment of output from CellProfiler Analyst. Step 10.

Potential solution

An inability to collect metadata from image files is likely the result of a poor file naming strategy. Simple but informative file names should be generated, with appropriate separation of metadata information. We find that DATE_PLATEID_WELL_REGION_CHANNEL is an easy to use format. The below regular expression can be used to capture this naming type for the example filename here:

File name: 220607_0034_B10_s23_ch00.

Regular Expression: `^(?P<Date>[0-9]{6})_(?P<Plate>[0-9]{4})_(?P<Well>[A-Z]{1}[0-9]{1,2})_s(?P<Region>[0-9]{2})_ch(?P<Channel>[0-9]{2})`.

Of course, users should familiarize themselves with regular expression if they are to perform more complex operations in future that require more information (z-slice, time-point).

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, [Steven S Plotkin] (steve@phas.ubc.ca).

Materials availability

Plasmid constructs, CellProfiler pipelines, and information are available from the technical contact, [Luke McAlary] (lmcalary@uow.edu.au).

Data and code availability

The protocol does not include all datasets generated. An example CellProfiler pipeline is provided as a [supplemental information](#). Other CellProfiler pipelines are available upon reasonable request

to the [lead contact](#) or technical contact. The python script for batch processing of StarDist nuclei identification is available at <https://github.com/ljr979/Stardist-BATCH> or <https://github.com/PlotkinLab/Stardist-BATCH>.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xpro.2022.101748>.

ACKNOWLEDGMENTS

L.M. is supported by the Bill Gole MND postdoctoral Fellowship (MNDRA), Canadian Institute of Health Research Transitional Operating Grant 2682 while at UBC Vancouver in the lab of S.S.P., and graciously acknowledges funding from a FightMND IMPACT Grant. J.J.Y. is supported by an NHMRC Investigator Grant and acknowledges funding from the MNDRIA and FightMND. S.S.P. is supported by Canadian Institute of Health Research Transitional Operating Grant 2682 and Alberta Innovates Research Team Program Grant PTM13007. We would like to acknowledge the support of the microscope technical staff at the Illawarra Health and Medical Research Institute (IHMRI), Molecular Horizons (UOW, Australia), and the Djavad Mowafaghian Centre for Brain Research (UBC, Canada).

AUTHOR CONTRIBUTIONS

Conceptualization: L.M., J.J.Y., N.R.C., S.S.P.; **Methodology:** L.M., V.K.S., M.S., L.J.R.; **Validation:** L.M., V.K.S., M.S.; **Formal analysis:** L.M., V.K.S., M.S.; **Investigation:** L.M., V.K.S., M.S.; **Resources:** L.M., J.J.Y., N.R.C., S.S.P.; **Data curation:** L.M., V.K.S., M.S.; **Writing - original draft:** L.M.; **Writing - review & editing:** L.M., V.K.S., M.S., L.J.R., J.J.Y., N.R.C., S.S.P.; **Visualization:** L.M.; **Supervision:** L.M., J.J.Y., N.R.C., S.S.P.; **Project administration:** L.M., J.J.Y., N.R.C., S.S.P.; **Funding acquisition:** L.M., J.J.Y., N.R.C., S.S.P.

DECLARATION OF INTERESTS

The authors have no declarations of interest related to this work.

REFERENCES

- Bray, M.A., Fraser, A.N., Hasaka, T.P., and Carpenter, A.E. (2012). Workflow and metrics for image quality control in large-scale high-content screens. *J. Biomol. Screen* 17, 266–274. <https://doi.org/10.1177/1087057111420292>.
- Buchser, W., Collins, M., Garyantes, T., Rajarshi, G., Haney, S., Lemmon, V., Li, Z., and Trask, O.J. (2012). Assay development guidelines for image-based high content screening, high content analysis and high content imaging. In *Assay Guidance Manual*. <http://www.ncbi.nlm.nih.gov/books/NBK100913/>.
- Cashman, N.R., Durham, H.D., Blusztajn, J.K., Oda, K., Tabira, T., Shaw, I.T., Dahrouge, S., and Antel, J.P. (1992). Neuroblastoma x spinal cord (NSC) hybrid cell lines resemble developing motor neurons. *Dev. Dyn.* 194, 209–221. <https://doi.org/10.1002/aja.1001940306>.
- Dao, D., Fraser, A.N., Hung, J., Ljosa, V., Singh, S., and Carpenter, A.E. (2016). CellProfiler Analyst: interactive data exploration, analysis and classification of large biological image sets. *Bioinformatics* 32, 3210–3212. <https://doi.org/10.1093/bioinformatics/btw390>.
- Jonkman, J., Brown, C.M., Wright, G.D., Anderson, K.I., and North, A.J. (2020). Guidance for quantitative confocal microscopy. *Nat. Protoc.* <https://doi.org/10.1038/s41596-020-0307-7>.
- Lambert, T.J. (2019). FPbase: a community-editable fluorescent protein database. *Nat. Methods* 16, 277–278. <https://doi.org/10.1038/s41592-019-0352-8>.
- Lee, J.Y., and Kitaoka, M. (2018). A beginner's guide to rigor and reproducibility in fluorescence imaging experiments. *Mol. Biol. Cell* 29, 1519–1525. <https://doi.org/10.1091/mbc.E17-05-0276>.
- McAlary, L., Plotkin, S.S., Yerbury, J.J., and Cashman, N.R. (2019). Prion-like propagation of protein misfolding and aggregation in amyotrophic lateral sclerosis. *Front. Mol. Neurosci.* 12, 262. <https://doi.org/10.3389/fnmol.2019.00262>.
- McAlary, L., Shephard, V.K., Wright, G.S.A., and Yerbury, J.J. (2022). A copper chaperone-mimetic polytherapy for SOD1-associated amyotrophic lateral sclerosis. *J. Biol. Chem.* 298, 101612. <https://doi.org/10.1016/j.jbc.2022.101612>.
- McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., et al. (2018). CellProfiler 3.0: next-generation image processing for biology. *PLoS Biol.* 16, e2005970. <https://doi.org/10.1371/journal.pbio.2005970>.
- Rueden, C.T., Schindelin, J., Hiner, M.C., DeZonia, B.E., Walter, A.E., Arena, E.T., and Elieciari, K.W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinf.* 18, 529. <https://doi.org/10.1186/s12859-017-1934-z>.
- Salmon, W.C., and Waters, J.C. (2011). CCD cameras for fluorescence imaging of living cells. *Cold Spring Harb. Protoc.* 2011, 790–802. <https://doi.org/10.1101/pdb.top113>.
- Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. MICCAI 2018. Lecture Notes in Computer Science, Vol. 11071, A. Frangi, J. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds. (Springer). https://doi.org/10.1007/978-3-030-00934-2_30.
- Turner, B.J., Atkin, J.D., Farg, M.A., Zang, D.W., Rembach, A., Lopes, E.C., Patch, J.D., Hill, A.F., and Cheema, S.S. (2005). Impaired extracellular secretion of mutant superoxide dismutase 1 associates with neurotoxicity in familial amyotrophic lateral sclerosis. *J. Neurosci.* 25, 108–117. <https://doi.org/10.1523/JNEUROSCI.4253-04.2005>.