

Input Module Tutorial #3: Using the *NamesAndTypes* Module

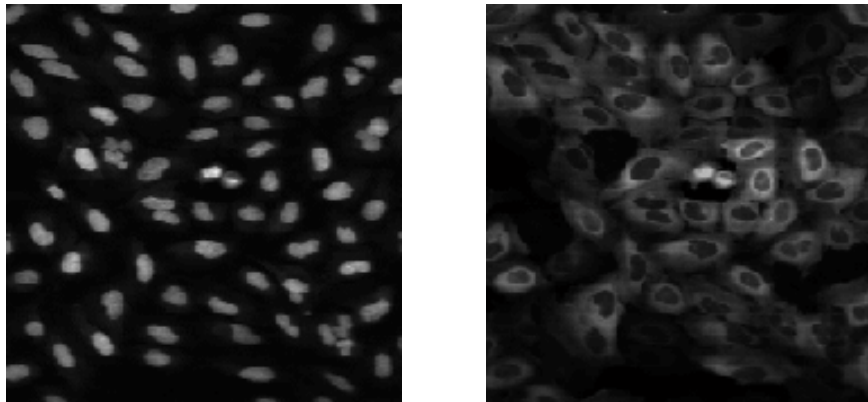
This tutorial describes how to use the **NamesAndTypes** module for properly identifying the images for processing in CellProfiler. Setting this module correctly is an important step for creating a new project in CellProfiler.

The **NamesAndTypes** module gives images and/or channels a meaningful name to a particular image or channel that you can use for reference in later modules in the pipeline. It also lets you define the relationships between images to create image sets.

For the purposes of this tutorial, you can use a set of files from a translocation assay which are available from http://www.cellprofiler.org/linked_files/TranslocationActivity/TranslocationData.zip. You also should understand how to use the **Images** module, and the **Metadata** module if necessary; if not, please refer to the [tutorials](#) for those modules before proceeding to this one.

What is an "image set"?

An *image set* is the collection of channels that represent a single field of view. For example, a fluorescent assay may have samples using DAPI and GFP to label separate cellular sub-compartments (see figure below), and for each site imaged, one DAPI (left) and one GFP image (right) is acquired by the microscope. Sometimes, the two channels are combined into a single color images and other times they are stored as two separate grayscale images, as in the figure:

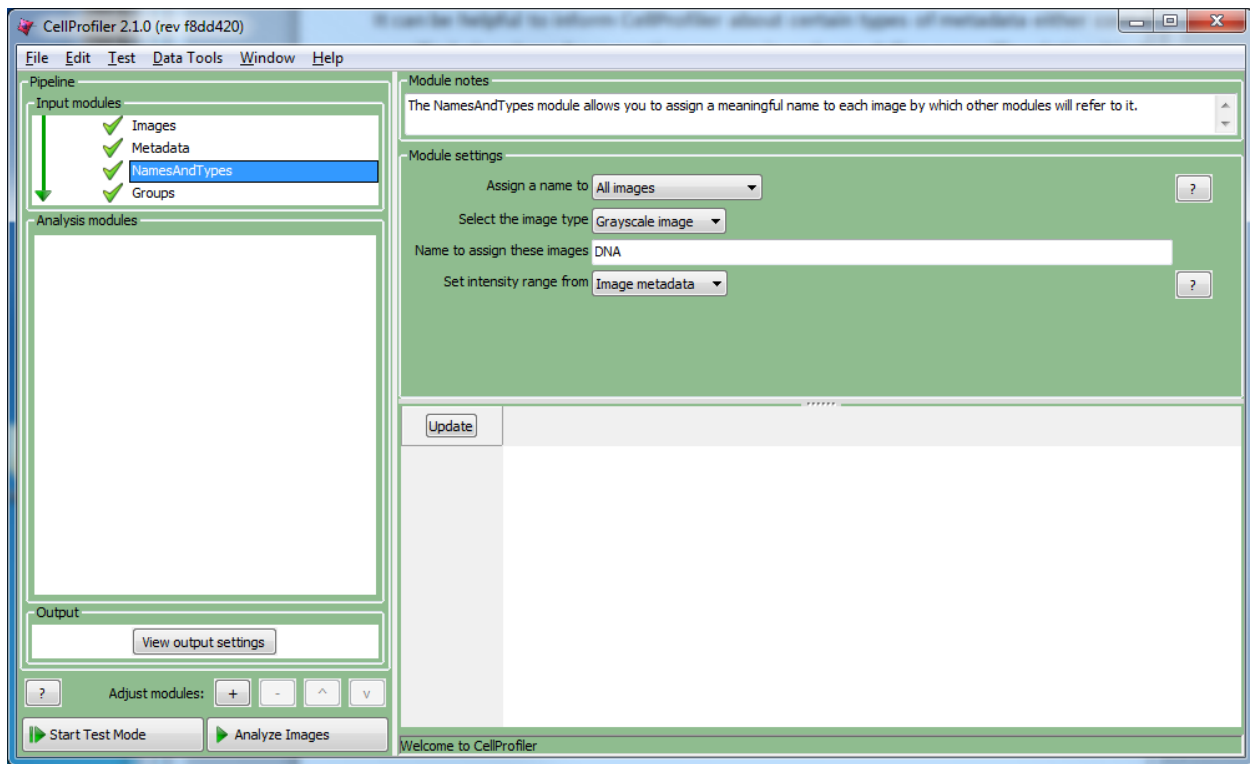


For the purposes of analysis, you want the DAPI and GFP image for a given site to be loaded and processed together. Therefore, the DAPI and GFP image for a given site comprise an image set for that site.

How do I start using this module?

The **NamesAndTypes** module receives the file list produced by the **Images** module. If you used the **Metadata** module to attach metadata to the images, this information is also received by **NamesAndTypes** and available for its use.

The **NamesAndTypes** module is the third Input module, shown in the "Input modules" panel on the upper left. Selecting this module will display a panel, allowing you to assign a unique name (provided by you) to each image channel. All files of a given channel will be referred to by the chosen name within the pipeline, and the data exported by the pipeline will also be labeled according to this name. This simplifies the bookkeeping of your pipeline and results by making the input and output data more intuitive: a large number of images are referred to by a small collection of names, which are hopefully easier for you to recognize.



The default (shown above) is to assign all the images to a single name. This is the simplest choice and the appropriate one if you have only one kind of image and you want to process each image independently in the pipeline rather than associating any of them with each other. CellProfiler will give each image the same name (e.g., “DNA”) and the pipeline will load only one of the “DNA” images per analysis iteration.

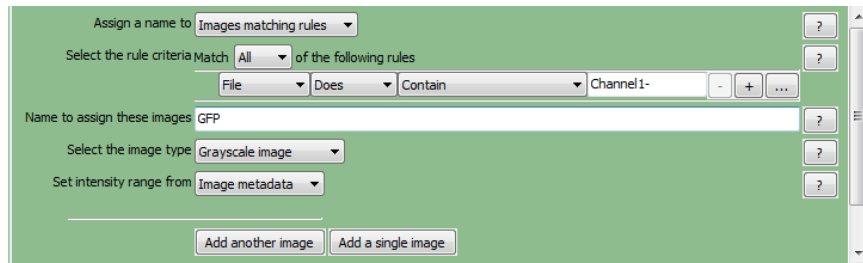
The alternative is to give images one of several names depending on some criteria that you define. This is the appropriate choice if more than one image was acquired from each imaging site and you want to associate them with each other so they can be processed together in the pipeline (e.g., “DNA” and “GFP”). You will be asked for distinctive criteria for each category of image and will be able to assign a name to each category for reference in downstream modules.

The most common way to perform this assignment is by specifying the pattern in the filename that the channel(s) of interest have in common. This is done by defining rules in a similar manner to that of the **Images** module, so refer to the tutorial for that module for help; other attributes of the file may also be used. If you have multiple channels, you then assign the relationship between channels.

For example, in the case mentioned above, the DAPI and GFP images are named in such a way that their identity is apparent, that is, "Channel2-" is contained in the file for the DAPI images, and "Channel1-" in the file name for the GFP images.

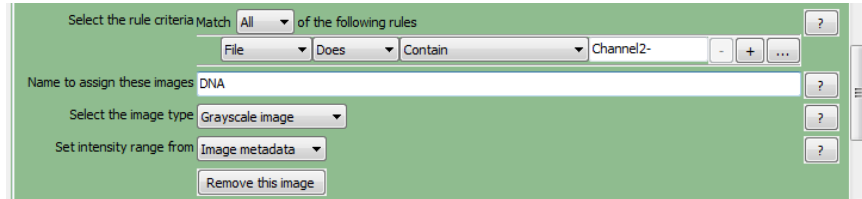
- Select “Images matching rules” from the assignment drop-box.
- From the rules menu, select “File” as the attribute, “Does” and “Contain” as the operator, and “Channel1-“ as the condition.
- Specify a name to refer to these images, in this case, “GFP”.
- You can also specify the type of image to be loaded (the most common options are listed here; the other options are lesser-used and are omitted for brevity):
 - *Grayscale image*: Most modules require grayscale images as input, so we recommend that the images arriving via the **Images** module be of this type. If the input image is color, this option will automatically convert it to grayscale.
 - *Color image*: This option will preserve the image as a three-channel RGB image, which may be appropriate for some applications, e.g., images of immunohistochemical stains.
 - *Binary mask*: This is a black and white image which is helpful for applications requiring masking. Masking an image with a binary image restricts the active pixels of the original image to only those which lie within the white regions of

the mask. This is useful for when you only want certain regions of interest in an image to be quantified and the rest is to be effectively ignored.



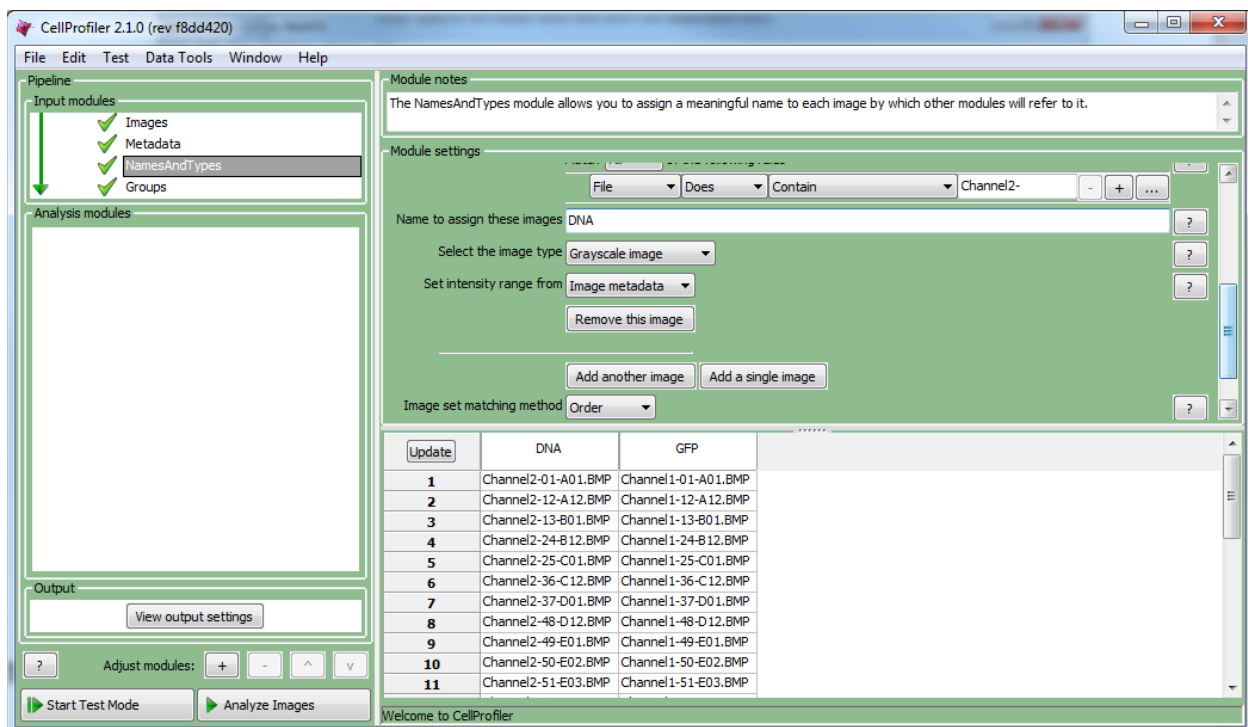
To add another channel, click the “Add another image” button and another group of settings will appear.

- Using the example above, select “Images matching rules” from the assignment drop-box for this new group of settings.
- From the rules menu, select “File” as the attribute, “Does” and “Contain” as the operator, and “Channel2-“ as the condition.
- Specify a name to refer to these images, in this case, “DNA”.
- Again, you can select the image type to be loaded for this channel.
- If you need to remove an image, press the “Remove this image” button.



How do I know whether I’m setting everything correctly?

At this point, it is helpful to see whether the images are matched up correctly to form the image sets you would expect. To do this, press the “Update” button below the divider to display a table of results using the current settings. Each row corresponds to a unique image set, and the columns correspond to the name you specified for CellProfiler to identify the channel. You can press this button as many times as needed to display the most current image sets obtained. When you complete your pipeline and perform an analysis run, CellProfiler will process the image sets in the order shown.



Worth noting is the setting labeled “Image set matching method”, which is responsible for how you want to match the image from one channel with the images from other channels. This setting controls how CellProfiler picks which images should be matched together when analyzing all of the images from one site.

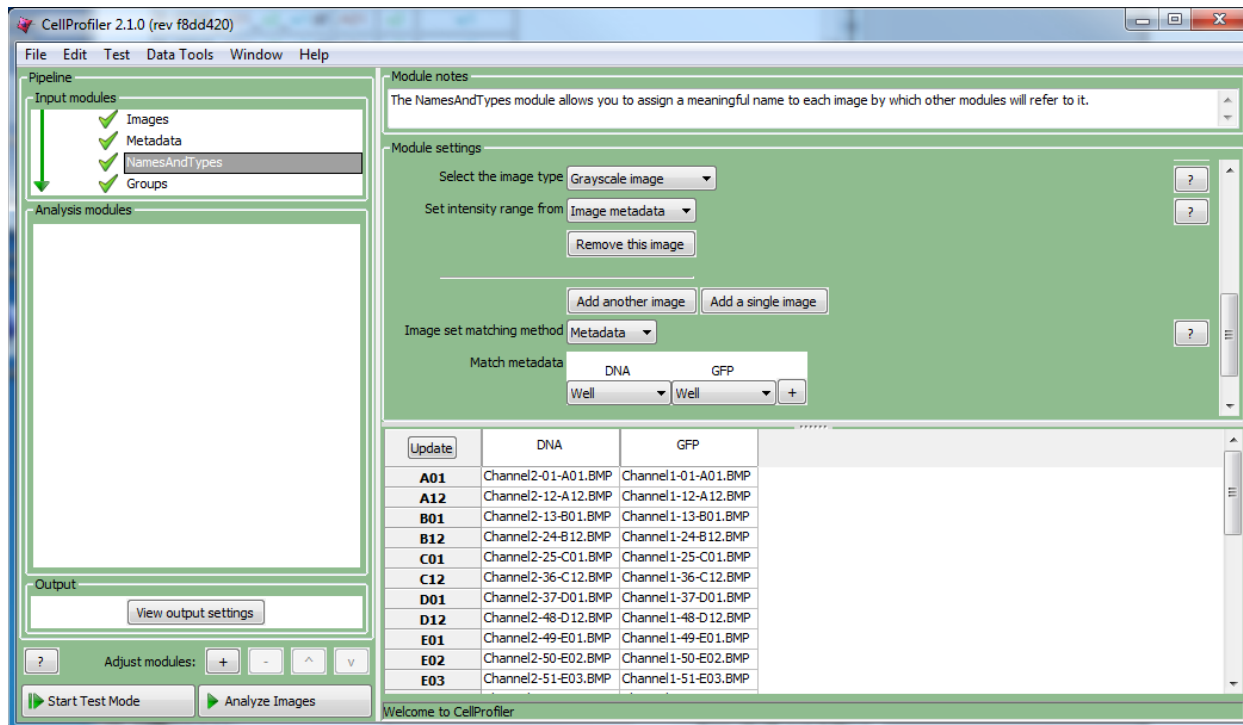
The default option is “Order,” in which CellProfiler will order the images in each channel alphabetically by their path and file name. CellProfiler will then match the first image from one channel to the first image from another channel. This approach is sufficient for most applications, but will match the wrong images if any of the files are missing or misnamed. The image set list will then get truncated according to the channel with fewer files.

A more powerful approach is the “Metadata” option, in which CellProfiler will match files with the same metadata values. This option is more complex to use than “Order” but is more flexible and less prone to inadvertent errors.

- Using this method assumes that you have extracted metadata from your images. If you have not, and would like to use this method, refer to the tutorial for the **Metadata** module on how to do so.
- Selecting “Metadata” for this option will reveal a set of drop-down menus, one column for each named channel.
- Set the drop-downs to pair up the metadata tags for each image set. In the example shown below, the well tags are specified for the “DNA” and “GFP” channels, indicating that the unique well tag for the DNA channel should be matched to the corresponding well tags for the GFP channel.



- This can be done for as many metadata correspondences as you need to specify a unique image set. For example, if you have multiple plates, each with a unique plate ID, using the well tags alone is not sufficient to specify a unique image set, since a given well tag will match many plates. To solve this, you would select “Plate” for the two channels, press **+** to add another row, and select “Well” for the new row. This selection will match the unique plate IDs first, and then match well tags within each plate.
- Pressing the “Update” button will show the image set matching using this method; see figure below.



What should I have when I'm done with this module?

The **NamesAndTypes** module is the last of the required input modules. After this module, you can choose any of the names that you defined in the **NamesAndTypes** module using drop-down lists in downstream analysis modules that require an image as input.