

Input Module Tutorial #4: Using the *Groups* Module

This tutorial describes how to use the **Groups** module to organize sets of images into groups in CellProfiler. Since image grouping is not necessary for every assay, *setting this module is optional*.

Once images have been identified with the **Images** module, have had metadata associated with them using the **Metadata** module, and have been assigned names by the **NamesAndTypes** module, you have the option of further sub-dividing the image sets into groups that share a common feature. Some downstream modules of CellProfiler are capable of processing groups of images in useful ways (e.g., object tracking within a set of images comprising a time series, illumination correction within a set of images comprising an experimental batch, data export for a set of images comprising a plate).

It is worth noting that CellProfiler will analyze your image sets more efficiently if you do not use this functionality, so we recommend enabling image groups only if your pipeline requires it.

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**, **Metadata**, and **NamesAndTypes** modules; if not, please refer to the [tutorials](#) for those modules before proceeding to this one.

What exactly is an image “group,” anyway?

The key to understanding why grouping may be necessary is that CellProfiler processes the input images sequentially and in the order given by the **NamesAndTypes** module. If you have multiple collections (or “groups”) of images that should be processed independently from each other, CellProfiler will simply finish processing one collection and proceed to the next, ignoring any distinction between them unless told otherwise via the **Groups** module.

To illustrate this idea, here are two examples where the grouping concept can be useful or important:

- If you have time-lapse movie data that is in the form of individual image files, and you are performing object tracking, it is important to indicate to CellProfiler that the end of each movie indicates the end of a distinct data set. Without doing so, CellProfiler will simply take the first frame of the next movie as a continuation of the previous one and attempt to track cells across one movie to the next. If each set of files that comprise a movie is defined using the **Metadata** module, the relevant metadata can be used in this module to insure that object tracking only takes place within each movie.
- If you are performing illumination correction for a screening experiment, we recommend that the illumination function (an image which represents the overall background fluorescence) be calculated on a per-plate basis. Since the illumination function is an aggregate of images from a plate, running a pipeline must yield a single illumination function for each plate (per channel). Running this pipeline multiple times, once for each plate, will give the desired result but would be tedious and time-consuming. CellProfiler can use image grouping to ease this workflow; if plate ID metadata is defined by the **Metadata** module, grouping will enable you to process images that have the same plate metadata together.

A grouping may be defined using any of the metadata categories defined by the **Metadata** module; one metadata tag or several can be used to achieve the desired result. Upon adding a metadata category, two tables will update in panels below showing the resulting organization of the image data for each group.

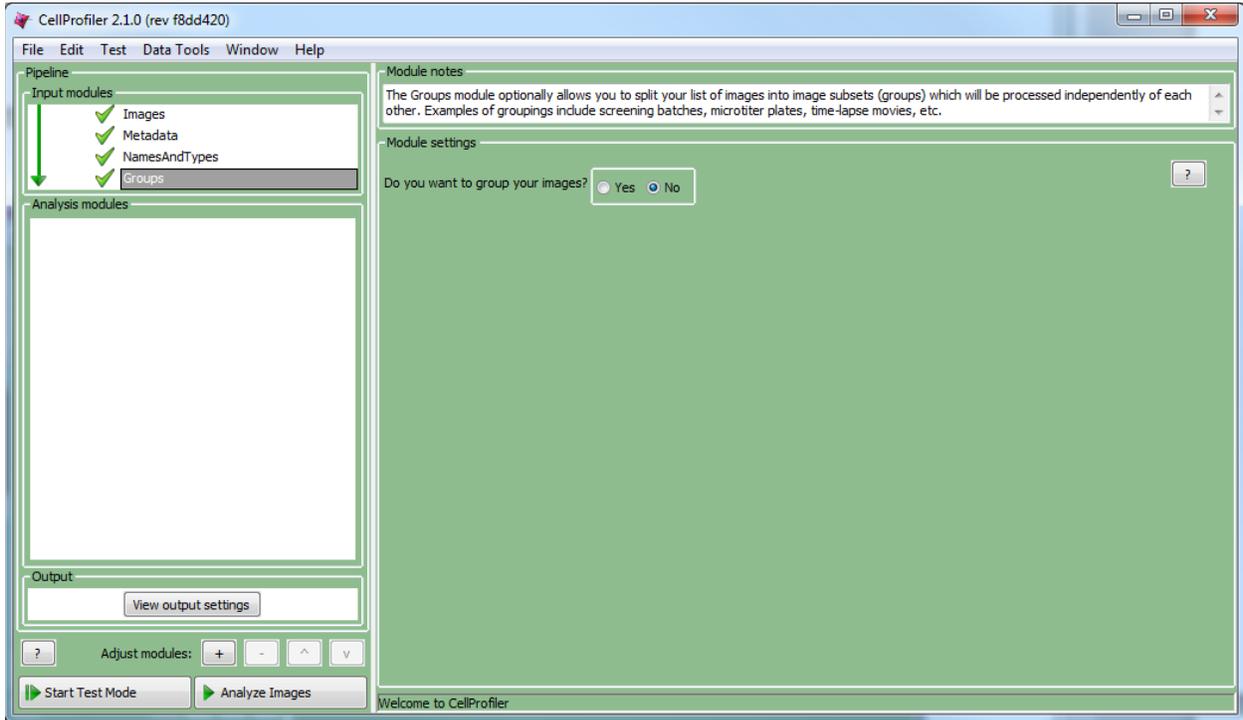
How do I start using this module?

Using this module assumes that you have already opened CellProfiler and adjusted the following Input modules:

- Used the **Images** module to produce a list of images to analyze.
- Used the **Metadata** module to produce metadata defining the distinct sub-divisions between groups of images.
- Used the **NamesAndTypes** module to assign names to individual channels and create image sets.

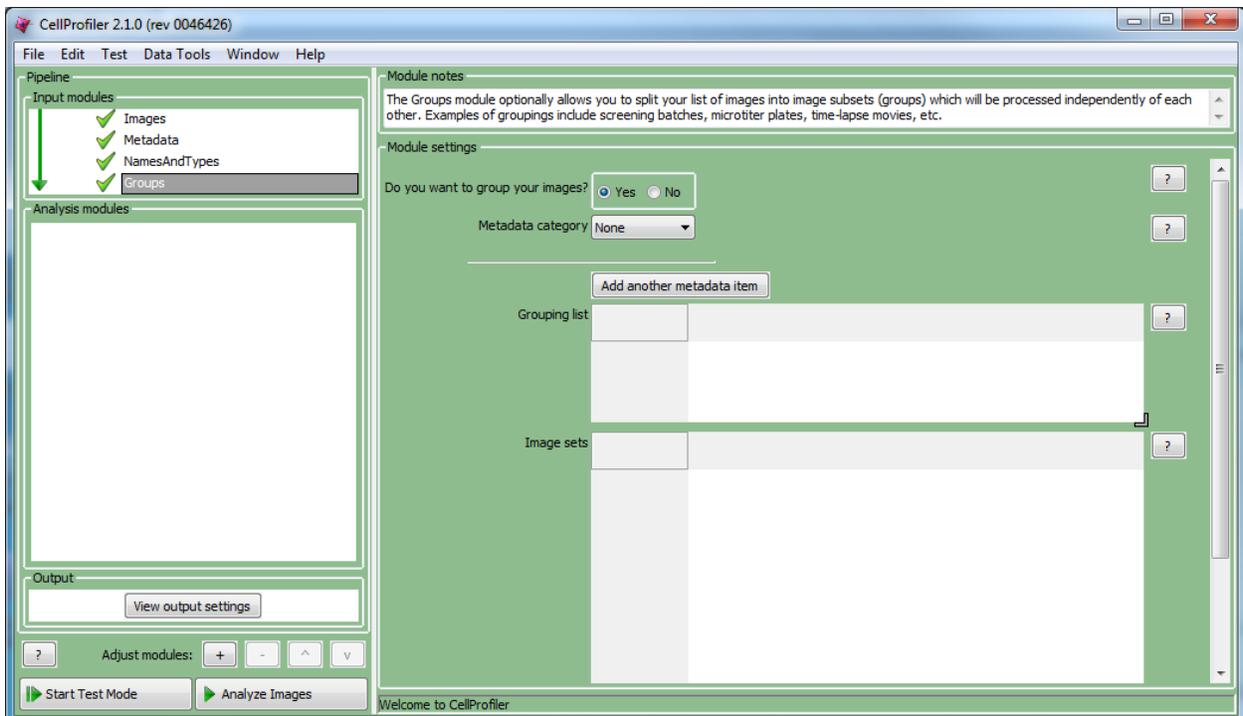
If you have not made the needed adjustments, you can refer to the tutorials for the **Images**, **Metadata**, and **NamesAndTypes** modules if you need help.

The **Groups** module is the fourth Input module, shown in the “Input modules” panel on the upper left. Selecting this module will display a panel, allowing you to select whether you want to create groups or not.

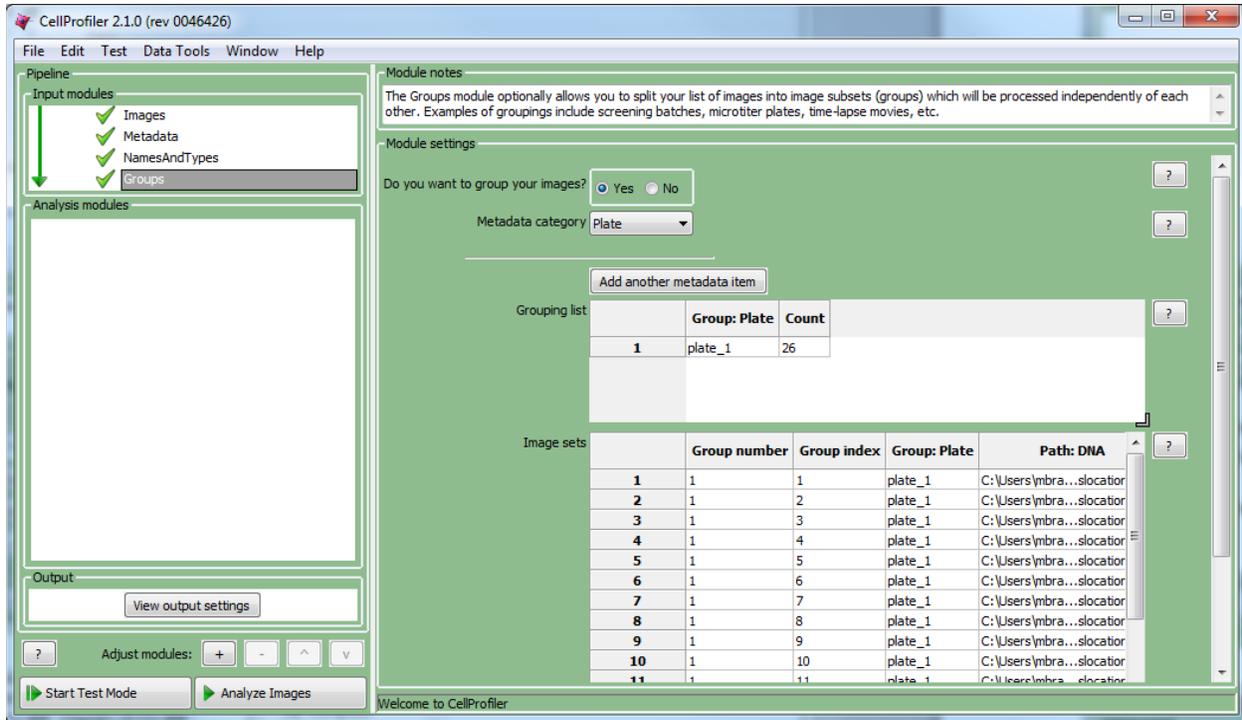


If creating groups of images is not relevant to your analysis, you can leave this module in the default setting, and continue on to the Analysis modules.

If you do have relevant image groups to create, select the button for grouping your images; a new group of settings will appear below:



By selecting a metadata tag from the drop-down for the “Metadata category” setting, the **Groups** module will sub-divide and assemble the image sets according to their unique metadata value. For example, selecting “Plate” from the drop-down will produce a display like the following:



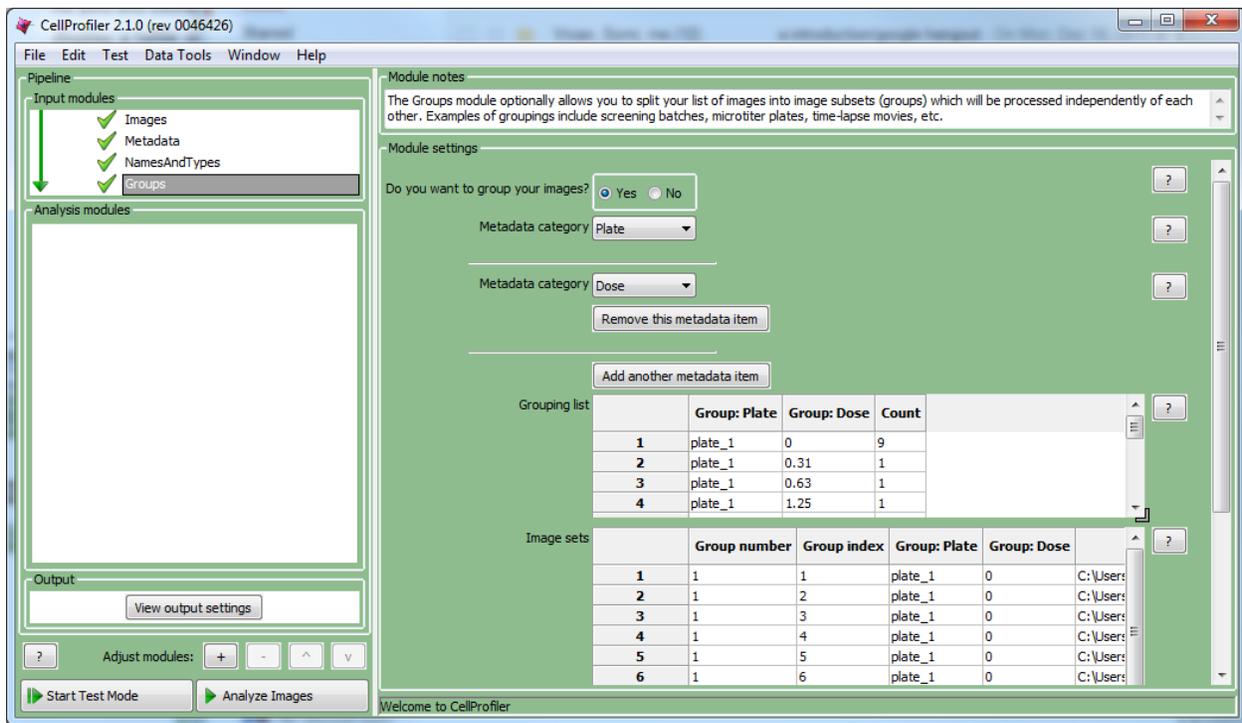
How do I know whether I’m setting everything correctly?

The two tables at the bottom provide the following information:

- The *grouping list* (top table) shows the unique values of the selected metadata under the “Group” column; each of the unique values comprises a group. The “Count” column shows the number of image sets included in a given group; this is useful as a "sanity check" to make sure that the expected numbers of images are present. In this example, grouping by per-plate metadata shows that there are 26 image sets in the “plate_1” group. If the assay were from a 384-well assay with 2 sites per well consisting of 3 plates, you would expect to see 3 rows (each from the 3 unique plate IDs), with $384 \text{ wells} \times 2 \text{ sites/well} = 768$ image sets in each.
- The *image set list* (bottom table) shows the file name and location of each of the image sets that comprise the groups. In this example, the table has 26 rows, one for each of the DNA and GFP image sets defined by the **NamesAndTypes** module. If you were grouping by per-plate metadata from a 384-well assay with 2 sites per well consisting of 3 plates, you would expect to see a table consisting of $3 \text{ plates} \times 384 \text{ wells/plate} \times 2 \text{ sites/well} = 2304$ rows.

You may specify multiple metadata tags to group with by clicking the “Add” button. This would be necessary if a combination of metadata is required in order to define a group. Upon adding a metadata category, the two tables will update automatically in the panels below showing the resulting organization of the image data for each group.

For example, if you wanted the processing to be organized not only per-plate but also per-dosage on each plate, you could also specify the “Dose” metadata in addition to the “Plate”, to get a grouping like the one below:



Here, eleven image groups are now defined, for each of the plate/dosage combinations; note that some have only one image set defined in the group, others have eight and still others have nine. Despite the uneven number of combinations, this is an acceptable grouping.

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

The final product of the **Groups** module is a list defining subsets of image sets that will be processed independently of the other subsets.

- If no groups are defined, the Analysis modules in the rest of the pipeline will be applied to *all* images in exactly the same way.
- If groups *are* defined in the **Groups** module, then organizationally (and transparently to you), CellProfiler will begin the analyses with the first image set of the group, end with the last image set of the group, and then proceed to the next group.

Certain modules are designed to take advantage of this form of book-keeping; see the in-module help for more details:

- **CorrectIlluminationCalculate:** The aggregation used for creating an illumination correction image is performed for each group separately.
- **ExportToDatabase:** Some measurements are exported to a database only after each group is processed.
- **ExportToSpreadsheet:** This module can write out separate files for groups of images, and name them accordingly.
- **MakeProjection:** The aggregation used for creating an image projection is performed for each group separately.
- **RunImageJ:** Scripts, commands, and macros can be run on a group of images that you have defined.
- **SaveImages:** Images can be saved to disk upon the last image set of a group, which is useful for the **CorrectIlluminationCalculate**, **MakeProjection** and **Tile** modules.
- **Tile:** If tiling images across processing cycles, the image groups define the image sets that will be contained within each montage.
- **TrackObjects:** Object tracking will begin and end on the image sets defined by your groups. Some tracking methods require calculations to be performed at the end of each group.
- **UntangleWorms:** A worm model can be trained and saved to disk for each group in your image sets