corrFISH guidelines (MATLAB)

We will describe working principles of corrFISH using simulated data and experimental test images. First, simulations will show how corrFISH accurately measures expected transcript abundance in a perfect case scenario without any experimental noise.

Simulated Data and Results

- Use the FOLDER: "simulation" and FILE: "maincorrfishsimulation"
- Add FOLDER to the path with subfolders in MATLAB (not the entire software package)

1. Raw images are provided for two sequential hybridizations in simulated cells with 20 µmx 20 µm area:
   TestSimulHyb1.tif: Image1, hyb1 image with A: 1000 and B: different 1000 transcripts
   TestSimulHyb1.tif: Image2, hyb2 image with A: same 1000 transcripts and C: other 1000
   - corrFISH provides copy number of **Gene A, total transcripts in Hyb1 and Hyb2**
2. *Corrfishcountquantification*: Simulated hybridization images are calculated for transcript quantification. Plots three-dimensional Gaussian fit of cross-correlation result (Figure 1).

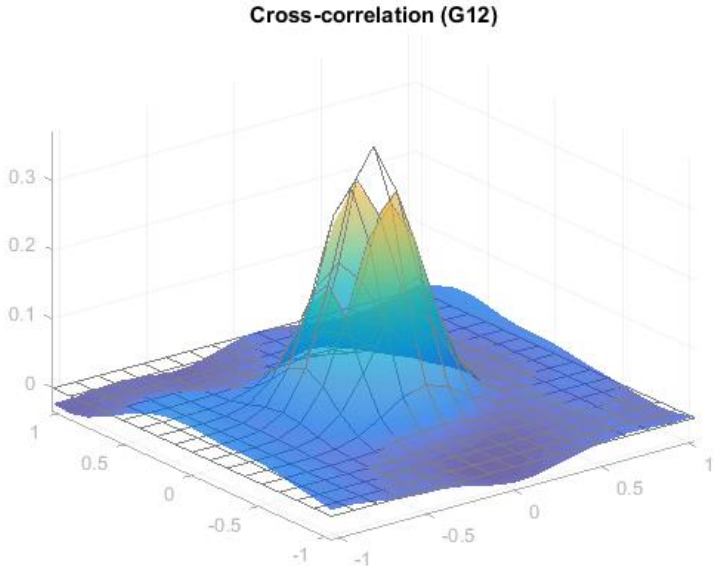

Figure 1. Gaussian fit of cross-correlation result of image 1 and image 2.

3. *Plotresults*: Provides simulated images, correlation result, and quantification of transcripts (**Figure 2**). Program outputs RNA A copy number on the cross-correlation image located at the bottom-left panel. Bar plot shows total RNA abundance in Image 1 (2014 molecules) and Image 2 (1938 molecules), as well as RNA A copy number (1008), which were all close to what we expect from simulated dot counts.

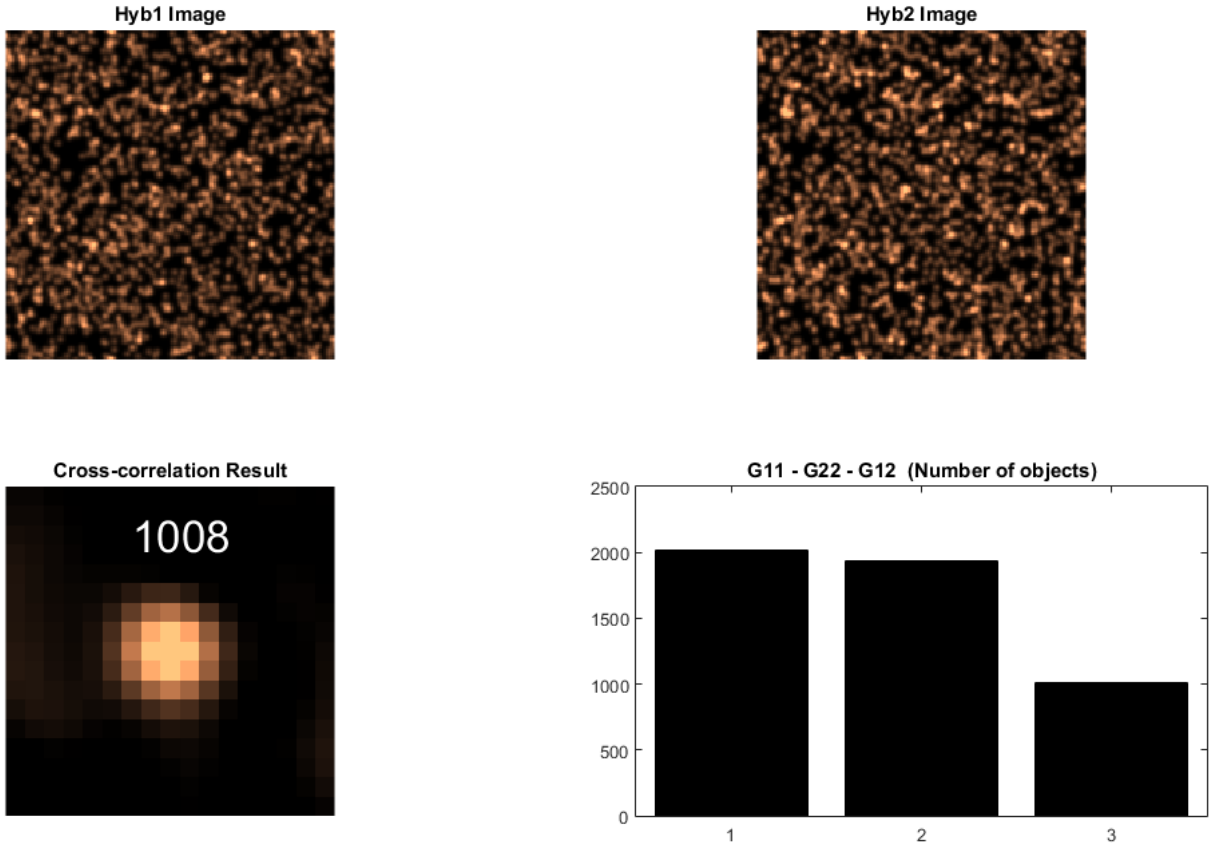

Figure 2. Example output image. (Top) Hyb1 and Hyb2 simulated images containing 1000 common A transcripts within total 2000 transcripts. (Bottom) Cross correlation result provides common A transcript counts to be 1008. Autocorrelations provided total number of transcripts in hyb1 image (2014) and in hyb 2 (1938). Cross-correlation of the two yielded colocalized transripts across two hybs, that is RNA A abundance of 1008 molecules.

Next, corrFISH will be used to decode experimental data in mammalian cells. Compared to the simulations, there are additional factors contributing to these images: Cell autofluorescence, out of focus light, and non-specific binding. Since auto-fluorescence and out of focus light are conserved across hybridizations, it might contaminate corrFISH calculations. Thus, we have used image subtraction to clean up cell background and deconvolution to avoid out of focus light. Non-specific binding should be random from one hyb to another, which should cancel their contribution in the correlations. Simply, they will not colocalize across sequential hybridizations. We will keep the analysis to one section of a cell and go over step-by-step image processing approaches used in the corrFISH quantitation.

Experimental Data and Results

- Use the FOLDER: “experiment” and FILE: “maincorrfishexperiment”

4. Raw images are provided for two sequential hybridizations in NIH3T3 cells:
   TestImageHyb1.tif: Image1, hyb1 image with A, B, C, D, E genes
   TestImageHyb2.tif: Image2, hyb2 image with A, F, G, H, I genes
   TestImageBackground.tif: Imagebckraw, the cell background

   - corrFISH provides copy number of **Gene A**
5. *corrfishbcksubtraction*: Cell autofluorescence background (image without FISH probes) is subtracted from hybridization images 1 and 2.

6. *corrfishdeconvolution*: Background subtracted image is deconvolved to remove out-of-focus light.

7. *Corrfishcountquantification*: Final hybridization images are computed for transcript quantification. Plots three-dimensional Gaussian fit of cross-correlation result (Figure 3).

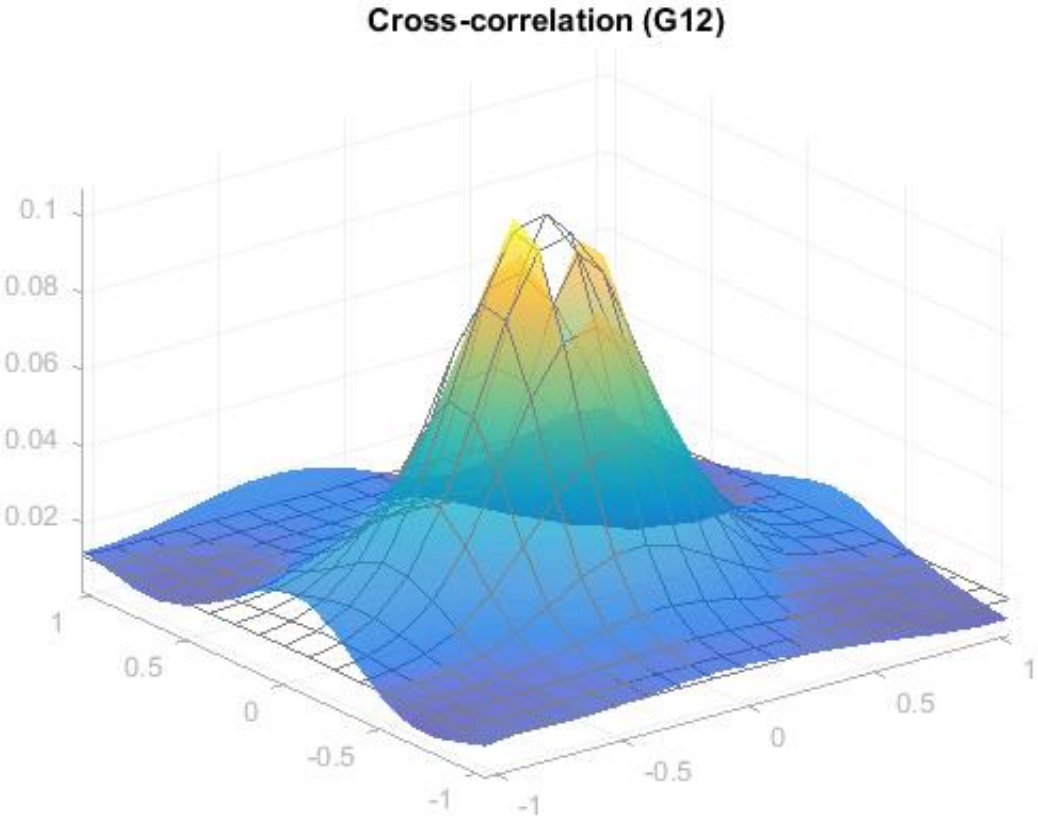

Figure 3. Gaussian fit of cross-correlation result of image 1 and image 2.

8. *Plotresults*: Provides an integrated plot of the summary of image processing steps (**Figure 4**). Program outputs RNA A copy number on the cross-correlation image located at the bottom-right panel. It is recommended to wait a few seconds to fit this plot onto the full screen size.

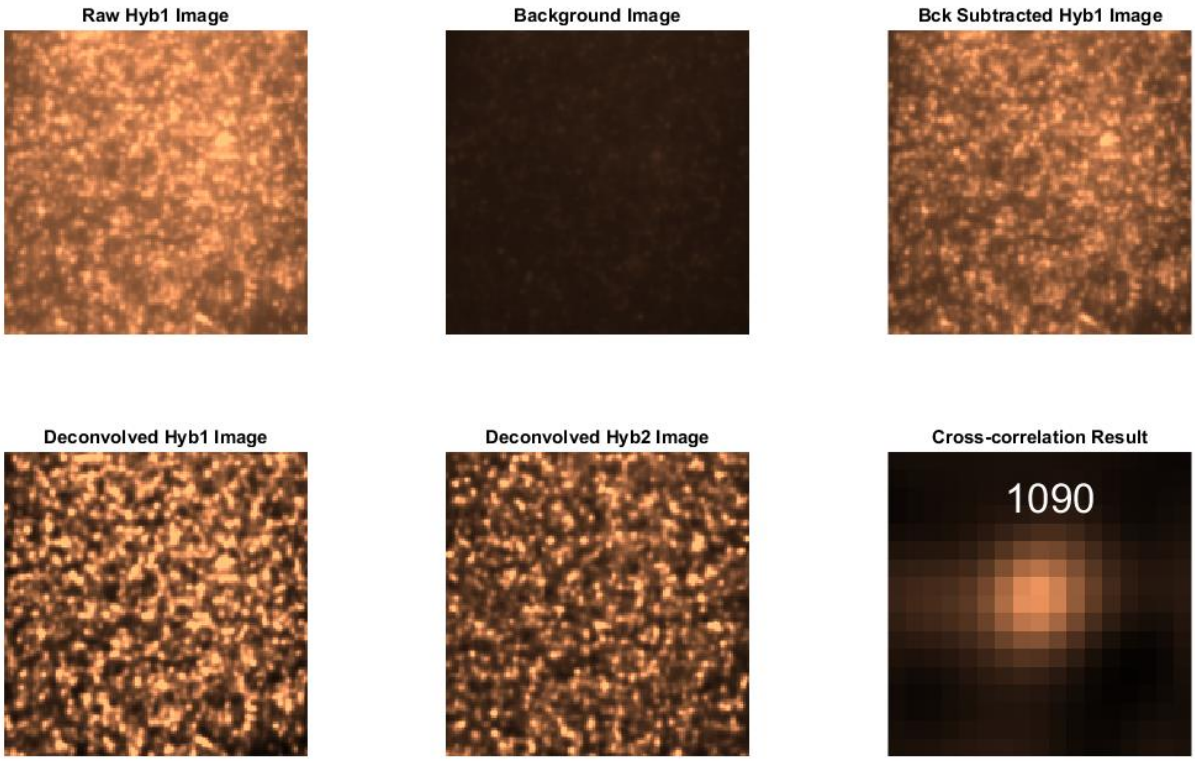

Figure 4. Example output image. (Top) Raw hyb1 image, raw cell background, image subtracted hyb1 image. (Bottom) Deconvolved and background subtracted hyb1 image. Similarly, deconvolved hyb2 image. Center region of the hyb1 * hyb2 image cross-correlation. The peak value is used to compute the RNA A copy number (For instance, 1090).

After demonstrating detailed analysis on a single section, we will discuss how we perform z-by-z corrFISH analysis in cultures. For correlation work, the best focus optical image of each hyb should perfectly match for best results. Thus, we implemented a biplane approach to sum subsequent planes and increment by one plane at a time. Here, we provide an example from NIH3T3 cells from two hybs. Four different optical sections span the cell body to capture most of the transcripts. Transcripts are then combined from each optical section and spatial subregion to calculate the gene expression per cell.

Optical Section Data and Results

- Use the FOLDER: “culture” and FILE: “maincorrfishculture”

9. Background subtracted and deconvolved images are provided for two sequential hybridizations in NIH3T3 cells:
   TestCultureCellHyb1Cell1.tif: Cell1, Image1, hyb1 image with A, B, and C genes
   TestCultureCellHyb2Cell1.tif: Cell1, Image2, hyb2 image with A, D, and E genes
   TestCultureCellHyb1Cell2.tif: Cell2, Image1, hyb1 image with A, B, and C genes
   TestCultureCellHyb2Cell2.tif: Cell2, Image2, hyb2 image with A, D, and E genes
   TestCultureCellHyb1Cell3.tif: Cell3, Image1, hyb1 image with A, B, and C genes
   TestCultureCellHyb2Cell3.tif: Cell3, Image2, hyb2 image with A, D, and E genes
   TestCultureCellHyb1Cell4.tif: Cell4, Image1, hyb1 image with A, B, and C genes
   TestCultureCellHyb2Cell4.tif: Cell4, Image2, hyb2 image with A, D, and E genes

   - corrFISH provides copy number of **Gene A**
10. *corrbyplanefishanalyze*: Optimal cell optical layers were registered using a custom approach. Biplanes were created to make corrFISH robust. Incrementing up to 4 optical layers covers the entire cell.

11. *corrfishsubcellcountculture*: Cell images were spatially aligned and subregions of cell were independently correlated with corresponding regions across the sequential hybridizations using *Corrfishcountquantification* function. Subregions are then summed to compute transcript abundance.

12. *Plotresults*: Provides optical section images for hybs 1 and 2 (**Figure 5-8**). Corresponding subregions were plotted as a gene expression map. Sum of all regions provided transcript counts per optical section.

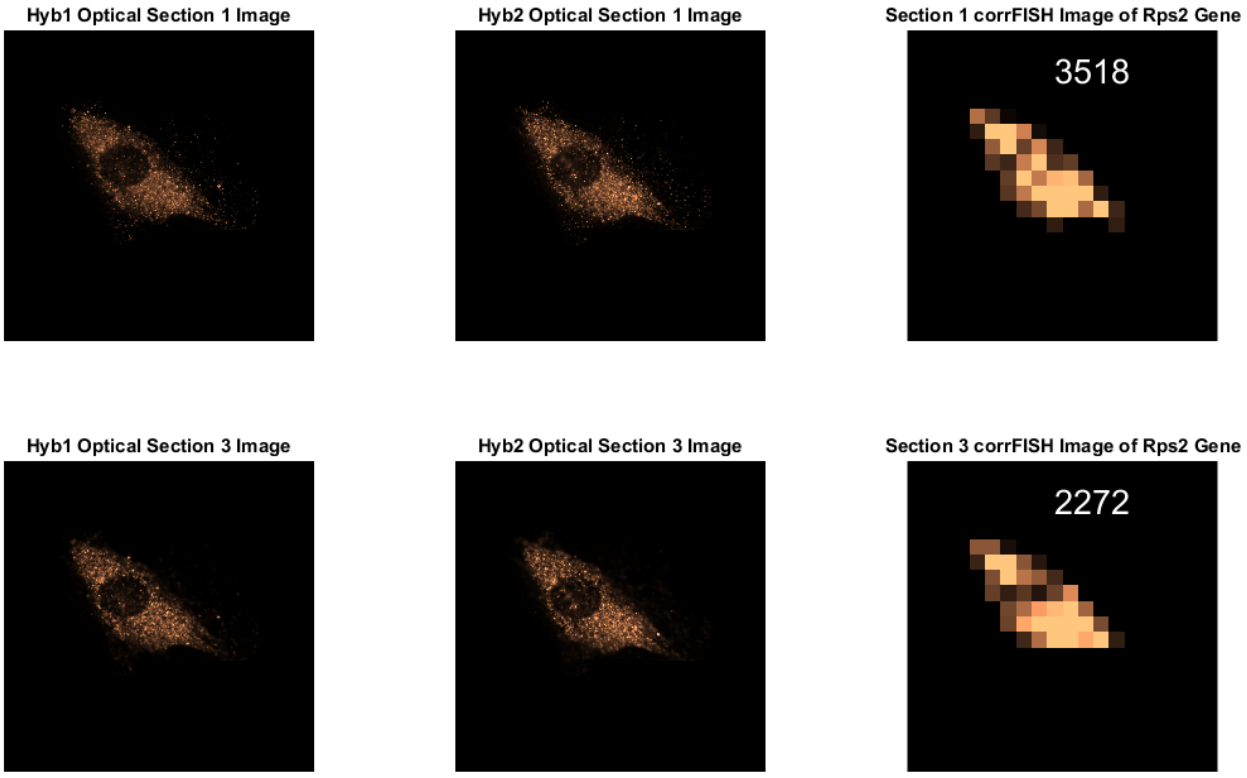

Figure 5. Example output image for cell 1. (Top) Biplane images from hyb1 and hyb2 at optical section 1. Corresponding subregions were calculated and the results of which were summed to compute transcript abundance per optical section 1. (Bottom) Biplane images of the same hybs at optical section 3. Subregions were processed to calculate gene expression. All the transcripts from different optical sections were then summed to profile transcript amount per entire cell.

Similar analyses can be performed on the cell 2 (**Figure 6**), cell 3 (**Figure 7**), and cell 4 (**Figure 8**).

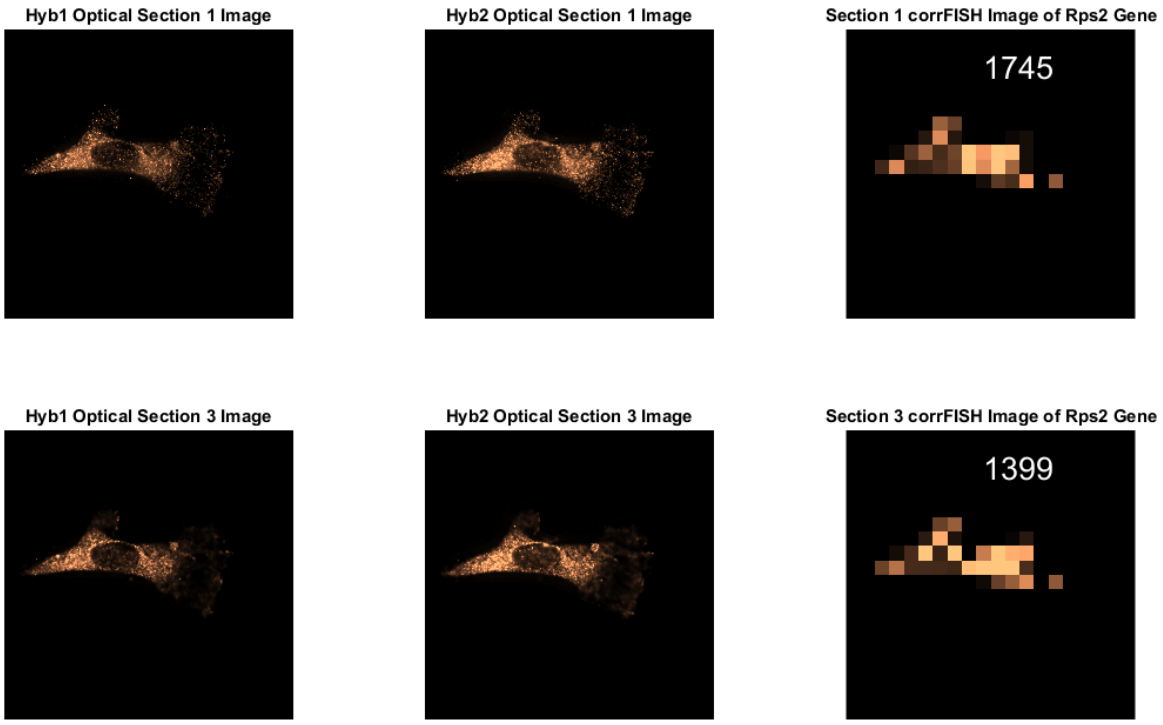

Figure 6. Output image for cell 2.

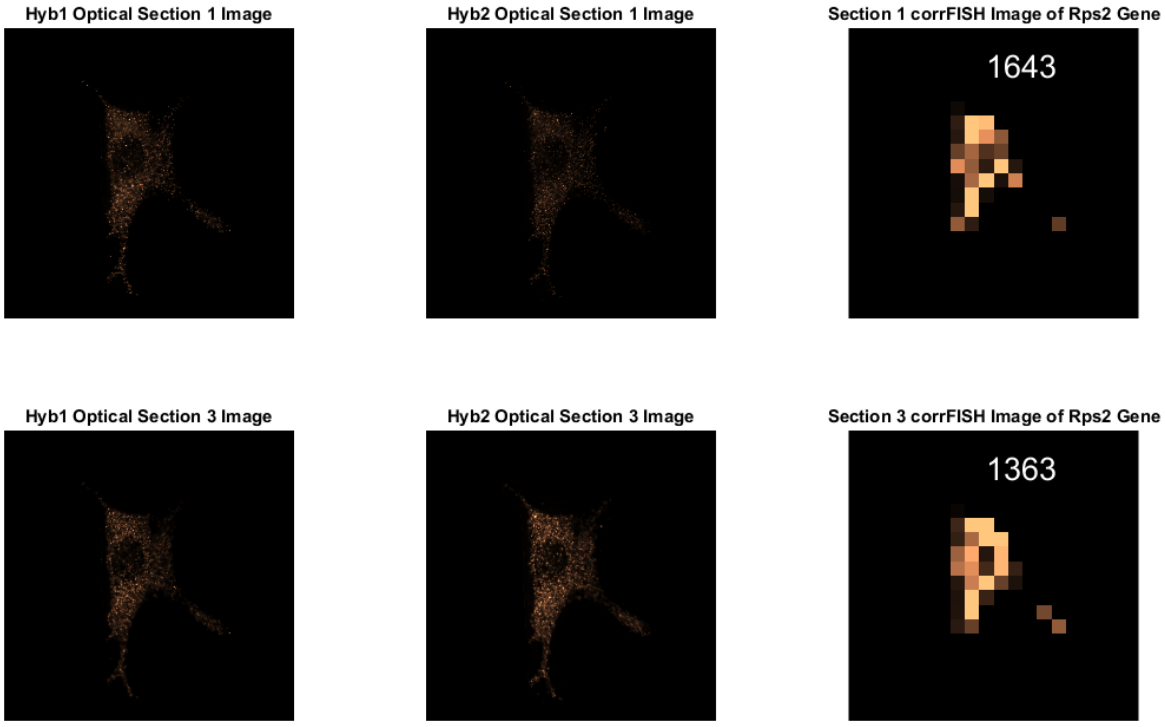

Figure 7. Output image for cell 3.

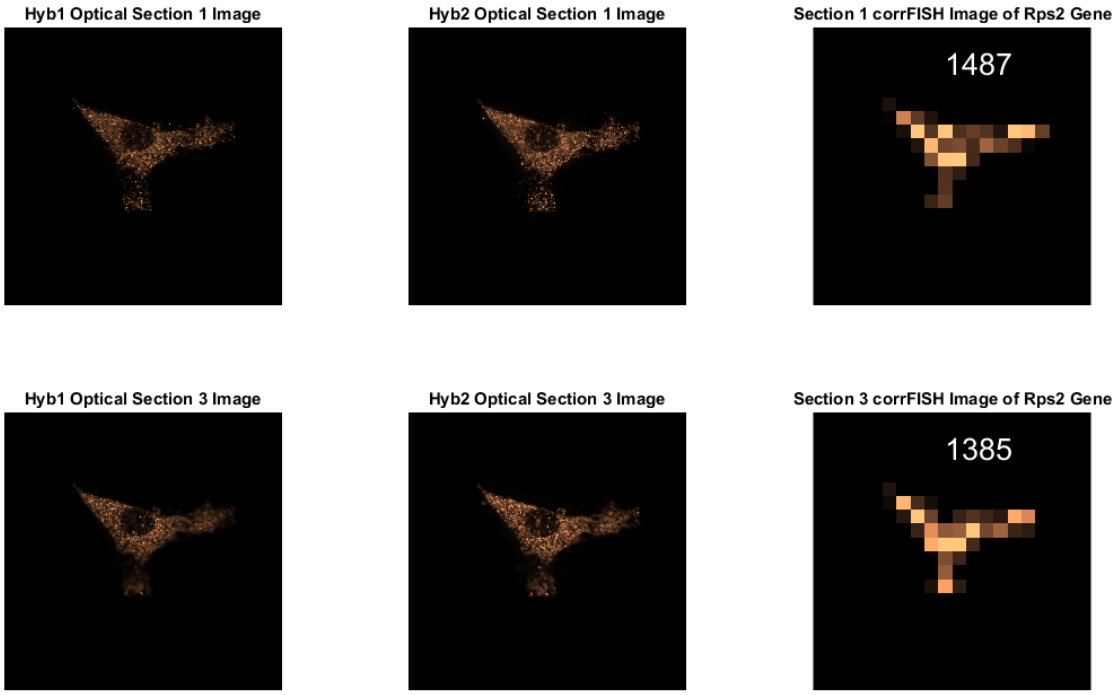

Figure 8. Output image for cell 4.

NOTES:

This code was written in MATLAB 2015a.

Since there are 3 separate packages (simulation, experiment, culture) that utilize some common functions, please make sure to add list to the path of the MATLAB **one folder at a time**. Otherwise, you might encounter "Too many input arguments" error.