Andy’s Algorithm User Guide

Table of Contents

1. Introduction

Andy’s Algorithm is an all-in-one tutorial, optimization and analysis pipeline designed for the rapid automated image analysis of specific assays such as immunohistochemistry (IHC), H&E, proximity ligation assays (PLA), and 3D colony assays. A step-by-step tutorial is available to guide users through the optimization and analysis process in each pipeline to make batch image analysis simple and adaptable across laboratories. Users are initially prompted to optimize the pipelines using representative images from their experimental cohorts to determine the ideal parameters to apply in the analysis. Upon finalization of the optimization process, the calculated parameters are applied to the experimental images for analysis.

Andy’s Algorithm accepts any image format that can be opened by FIJI and allows users to analyze specific images by giving the option to input a custom name

when selecting the image format. It is preferable to have image names contain the file format as an extension as part of the file name. If the file extension is not available in the file name users can still select for their images by inputting a common name. By design, the algorithm only analyzes single images and not Z-stacks.

Output data is provided as a summary spreadsheet file in the same folder as the analyzed images.

2. Installation

To download, install and run Andy's Algorithms, follow the following steps:

1. Download and install FIJI or update FIJI (version 1.51k or later) [https://fiji.sc]
2. Download Andy's Algorithms (eg DAB_IHC_v2.40.ijm, PLA_v2.40.ijm, HandE_v2.40.ijm and 3D_colony_v2.40.ijm) from https://github.com/andlaw1841/Andy-s-Algorithm (Provided in the related manuscript archive file and will be uploaded subsequent to publication)
3. Go to Plugins > Macros > Install the Algorithm of choice in the menu bar of FIJI
4. Select the Andy's Algorithm preference (eg DAB_IHC_v2.40.ijm) to install the algorithm
5. Go to Plugins > Macros. An option to select the algorithm will now be in the dropdown menu.

The algorithm will be temporally installed into the toolbar of FIJI and closes when FIJI is exited. Simply reinstall the algorithm when you open FIJI again.

3. Method Overview

The steps involved in Andy's Algorithm vary based on the pipeline used to analyze the assays. All the pipelines, excluding the PLA pipeline, follow a sequence of common steps of threshold selection, identification and measurement of total and positive regions within an image. Users are provided the options to apply a Gaussian blur to smooth the edges of the selection before applying a threshold. The thresholding process can be accomplished manually, where the user employs a set value for the threshold, or automatically, where a threshold method is applied and the threshold value for each image is calculated by the method chosen. Upon setting the threshold to identify the region of interest (ROIs), the image is converted to a mask image where binary processes such as watershed (segmentation), fill holes, and edge exclusion can be employed to modify the selections. The selections can then be excluded based on size and circularity by using the particle analysis function. The final selection can then be overlaid on top of the raw image to allow user to visually examine the accuracy of the selections.

3.1 DAB IHC

The DAB IHC pipeline separates blue (hematoxylin) from brown (DAB+) regions to determine the percentage DAB+ relative to the total region identified. The pipeline measures the count and area of both the total selection (hematoxylin and DAB+) and positive selection (DAB+ only), and the intensity of the positive selection. Images are analyzed with the total selection first and subsequently the positive selection.

A basic or enhanced method of selection is available for both the total and positive selection. In the basic method the total selection is determined by converting

the raw image into an 8-bit greyscale image before identifying ROIs via threshold and processed through the normal sequence of steps. For the positive selection, a color deconvolution under the H&E DAB vector is applied on to isolate the DAB+ regions before it is converted into an 8-bit grescale image and following the same steps as the total selection. In the enhanced method, a deuteranope and tritanope color blind filter is applied before a color deconvolution filter with a Feulgen Light Green vector and a H&E DAB vector is used on the total and positive selection, respectively. The application of these filters dilutes the color differences between the hematoxylin and DAB+ regions and allows for more accurate selection of the nuclei, cytoplasm or tissue, while reducing background in the total selection. In the positive selection the tritanope filter enhances the contrast of the DAB+ regions from the hematoxylin allowing an improved isolation of the DAB+ regions when using the H&E DAB color deconvolution. The enhanced method is recommended for images with poor staining whether it is due to overly dilute or concentrated DAB staining and/or hematoxylin counterstaining. Poor IHC staining can lead to inadequately defined nuclear outlines, which can result in inaccurate nuclei segmentation.

3.2 H&E

The H&E pipeline separates the blue/purple (hematoxylin) from the red/pink (eosin) areas. Using the same principles as the DAB IHC, the H&E pipeline follows the same steps to measure the count and area of both the total selection (eosin and hematoxylin) and positive selection (hematoxylin). Images are analyzed with the total selection first followed by the positive selection.

A basic and an enhanced method of selection is available for both the total and positive selection. The basic method converts the raw image to an 8-bit greyscale

image for the total selection and a color deconvolution under an H&E vector is used for the isolation of the positive selection of the hematoxylin. Similar to the DAB IHC, the enhanced method is used to permit more accurate segmentation and selection of ROIs for images with poor staining and color discrepancy between the hematoxylin and eosin. A deuteranope color blind filter and a color deconvolution with FastRed/FastBlue/DAB vector is applied for the total selection to mitigate color discrepancy and tritanope filter with a color deconvolution with H&E DAB vector for positive selection to enhance contrast, respectively.

3.3 3D Colony Assay

The workflow of the 3D Colony Assay pipeline comprises of the general steps involved in both the IHC DAB and H&E pipelines to select for the ROIs. This pipeline measures the count, average area, and average circularity of colonies to determine their growth, proliferation, and invasion within a 3D matrix.

To allow for an accurate selection of colonies the raw images are first processed with either a background subtraction with the rolling ball algorithm or normalize local contrast to remove any uneven illumination within the image. The ROIs are then smoothed with a Gaussian blur, converted to an 8-bit greyscale image, and processed with the general analysis steps.

3.4 PLA

The PLA pipeline is composed of a different method of analysis compared with the other three pipelines. Analysis of images is performed in sets of images composed of nuclei, foci, and cytoplasm (optional), acquired under a different channel. For the algorithm to identify the correct image for analysis, the user is first

required to input a unique name for each channel. The nuclei and cytoplasmic image is processed in the same manner where an enhance-contrast step is used to artificially enhance any dim fluorescence before applying a Gaussian blur to smooth the ROIs, reduce noise, threshold the ROIs, and finally employ a particle analysis with a size and circularity exclusion. The foci image is initially processed with a Gaussian blur to remove any multiple false positives that may occur due to uneven dispersion of bright pixels within individual foci. Foci are then selected using the Find Maxima function based on a user-defined noise tolerance value. Selections are converted to a mask image that are then counted using the particle analysis. Foci within a nuclear or cytoplasmic domain are identified using the foci mask image and setting a nuclear and/or cytoplasmic ROI to determine the raw integrated density within the region.

4. Troubleshooting

4.1 General

Difference between auto-threshold and manual threshold

Auto-threshold selects an algorithm to apply onto an image that uses a different threshold value based on each image. A manual threshold is a set threshold value that is determined by the user that will apply that single threshold value across all images. For more details on auto-threshold refer to https://imagej.net/Auto_Threshold

Applying size and circularity exclusion in particle analysis

To determine the appropriate values for the size and circularity exclusion, uses are advised to use their own representative images to adjust the parameters that suit their images best. If the exclusion is too stringent or lenient, retry the size exclusion

by selecting ‘no’ when prompted with the dialogue box “Are you happy with the selection?” to try different values. A lower and upper size exclusion guide is available adjusted for epithelial tissues.

Magnification	**Lower Size Exclusion (pixel size)**	**Upper Size Exclusion (pixel size)**
10x	40	infinity
20x	100	infinity
40x	150	infinity

Images are not opening

When selecting the file format ensure the image filenames contains the file extension at the end, for example if selecting “tif” for the file format then the image name requires the extension to match, such as Image01.tif. Alternatively, images can also be identified by inputting words/letters/numbers (case sensitive) that is common in the filename of all images within a given folder by choosing “custom” when selecting the image format. In Windows, the file extension can be shown in the filenames by following the steps https://support.microsoft.com/en-au/help/865219/how-to-show-or-hide-file-name-extensions-in-windows-explorer

No values in dialogue box

When the user is prompted to enter a value in the dialogue box and a value within the dialogue box is deleted and no value is entered an error with the following message will display.

Macro Error

Numeric value expected in run() function

Dialog box title: "[Title of dialog where the value is deleted]"

Key: "[Section where value was deleted]"

Value or variable name: "[A variable name]"

This error will only occur if the user deletes a value and leaves it blank before proceeding with the next step. If the error occurs the user will have to close everything that is opened within FIJI and restart the algorithm. To avoid this error ensure that no parameters are empty or is left blank throughout the algorithm.

4.2 IHC DAB

Poor segmentation for hematoxylin or DAB+ nuclei

We recommend that users employ Andy's algorithms in experiments where fixation, staining, image acquisition settings, exposure, contrast and brightness are consistent across a cohort of samples to enable accurate and reproducible batch analysis. Immunohistochemistry should be performed across a cohort of samples using the same antibody batch, timings and protocol in any given experiment. If large variations in fixation, staining, image acquisition settings, exposure, contrast and brightness exist then average parameters will not perform ideally on outlier images.

Images with overly dilute or concentrated DAB staining and/or hematoxylin counterstaining can lead to poorly defined nuclear outlines, resulting in difficulties in nuclei segmentation. The staining condition should be optimized to allow for well-defined nuclei that can be easily segmented and isolated. However if difficulties in nuclei segmentation still persist it is recommended to use the enhanced method for both total and positive selection, do not apply a Gaussian blur, and perform a manual

threshold in order to fine-tune to nuclear segmentation. It is important for users to reduce the threshold value so it minimizes background selection resulting in greater nuclear segmentation. If nuclei segmentation is still inaccurate then use “Area” when interpreting the results.

4.3 PLA

Assigning each name to a channel

When assigning a unique name to a specific channel ensure that the name only appears in the filename of that series of image (case sensitive).

For example:

Foci_01_ch00.tif

Nucleus_01_ch01.tif

Foci_02_ch00.tif

Nucleus_02_ch01.tif

The unique name that can be assigned for foci images can be “ch00” or “Foci”, while nuclei images can be “ch01” or “Nucleus”.

Images not opening

Ensure that the unique name matches the specific channel (case sensitive) and are single images and not Z-stacks. Image sets must be grouped together in sequential order as the algorithm reads in sets of two or three images at a time, depending on if there is a cytoplasmic image or not.

For example:

Cytoplasm_01_ch02.tif

Foci_01_ch00.tif

Nucleus_01_ch01.tif

Cytoplasm_02_ch02.tif

Foci_02_ch00.tif

Nucleus_02_ch01.tif

The algorithm will analyze the first three images as a set before analyzing the next three. If sets contain additional or missing images then an error will appear before the analysis starts to notify the user to remove or replace the defective image set.

Images not opening (PC)

When running the PLA Algorithm in Windows, hidden files (such as desktop.ini) in the image folder can cause an error message to display with either 'There are no images open' or 'Index (X) out of range in line 4278'. To correct this, in the image folder click on the View tab and select the Advanced settings. Check the "Show hidden files, folder or drives" and uncheck the "Hide protected operating system files" boxes and click OK. This will reveal the hidden files that are within the image folder and you can temporarily move those hidden files to another folder before you proceed with the PLA image analysis. Once the analysis is complete you can move the hidden files back into the folder and go back to the Advanced settings in the View tab and uncheck the "Show hidden files, folder or drives" and check the "Hide protected operating system files" boxes.

5. Outputs

5.1 DAB IHC output

Image Name	Total Count	Positive Count	Percent Count	Total Area	Positive Area	Percent Area	Average Intensity

5.2 H&E output

Image Name	Total Count	Positive Count	Percent Count	Total Area	Positive Area	Percent Area

5.3 3D colony assay output

Image Name	Colony Counts	Total Area of All Colony	Average Area Per Colony	Average Circularity Per Colony	Average Aspect Ratio Per Colony

5.4 PLA output

Image name	Nucleus Count	Total Nuclei Area	Average Nuclei Area	Total Signal	Nuclear Signal	NonNuclear Signal	Percent Nuclear Signal	Average Signal per Nucleus
Cytoplasm Area	Average Cytoplasmic Area	Cytoplasmic Signal	Non Cytoplasmic Signal	Percent Cytoplasmic Signal	Average Signal per Cytoplasm	Intracellular Signal	Extracellular Signal	

6. Glossary

6.1 IHC Glossary

Average Intensity	The average mean grey value of all positive ROI, that ranges from 0-255 where 0 is darkest (black) and 255 is brightest (white).
Percent Area	Percentage DAB positive area relative to the total area measured (Area of positive selection divided by area of total selection multiplied by 100)
Percent Count	Percentage DAB positive count relative to the total count measured (Area of positive selection divided by area of total selection multiplied by 100)
Positive Area	The area of the DAB positive selection (positive ROI), which can be visualized in the “positive selection mask" image
Positive Count	The count of the DAB positive selection (positive ROI), which can be identified in the "positive ROI" zip file
Positive Mask Image	Black and white binary image where the positive count and area is measured from
Positive Overlay Image	Pseudo-color image with the positive ROI overlaid on top of the raw image
ROI	Region of Interest

Total Area	The area of the total selection (total ROI), which can be visualized in the “total selection mask" image
Total Count	The count of the total selection (total ROI), which can be identified in the "total ROI" zip file
Total Mask Image	Black and white binary image where the total count and area is measured from
Total Overlay Image	Pseudo-color image with the total ROI overlaid on top of the raw image

6.2 H&E Glossary

Percent Area	Percentage hematoxylin positive area relative to the total area measured (Area of positive selection divided by area of total selection multiplied by 100)
Percent Count	Percentage hematoxylin positive count relative to the total count measured (Area of positive selection divided by area of total selection multiplied by 100)
Positive Area	The area of the hematoxylin positive selection (positive ROI), which can be visualized in the “positive selection mask" image
Positive Count	The count of the hematoxylin positive selection (positive ROI), which can be identified in the "positive ROI" zip file
Positive Mask Image	Black and white binary image where the positive count and area is measured from
Positive Overlay Image	Pseudo-color image with the positive ROI overlaid on top of the raw image
ROI	Region of Interest
Total Area	The area of the total selection (total ROI), which can be visualized in the “total selection mask" image
Total Count	The count of the total selection (total ROI), which can be identified in the "total ROI" zip file
Total Mask Image	Black and white binary image where the total count and area is measured from
Total Overlay Image	Pseudo-color image with the total ROI overlaid on top of the raw image

6.3 3D Colony Assay Glossary

Average Area Per Colony	The average area of each colony (total area of all colony divided by colony counts)
Average Aspect Ratio Per Colony	The average aspect ratio of each colony based on the major axis divided by the minor axis
Average Circularity Per Colony	The average circularity of each colony that ranges from 0-1 where 1 is a perfect circle and 0 is an elongated polygon
Cell Mask	Black and white binary image where the total colony count and area is measured from

Cell Overlay	Pseudo-color image with the total ROI overlaid on top of the raw image
Colony Counts	The count of total number of colonies, which can be identified in the "cells ROI" zip file
ROI	Region of Interest
Total Area of All Colony	The area of the total selection, measured based on the "cell masks" image

6.4 PLA Glossary

Average Cytoplasmic Area	The average area of each cytoplasm based on the number of nuclei identified (cytoplasm area divided by nucleus count)
Average Nuclei Area	Average area of each nuclei (total nuclei area divided by nucleus count)
Average Signal per Cytoplasm	Average foci in each cytoplasmic region (cytoplasmic signal divided by nucleus count)
Average Signal per Nucleus	Average foci in each nucleus (nuclear signal divided by nucleus count)
Cytoplasm Area	The total area of all cytoplasmic region measured from the "cytoplasm mask" image
Cytoplasm Mask	Black and white binary image where the total cytoplasmic area is measured from
Cytoplasm Overlay	Pseudo-color image with the cytoplasmic ROI overlaid on top of the raw cytoplasm image
Cytoplasmic Signal	Total number of foci identified within cytoplasmic regions
Exctracellular Signal	Total number of foci identified outside of both nuclear and cytoplasmic region (total signal minus intracellular signal)
Foci Overlay	Pseudo-color image with the foci ROI overlaid on top of the raw foci image
Intracellular Signal	Total number of foci identified within either nuclear or cytoplasmic regions (nuclear signal plus cytplasmic signal)
Non Cytoplasmic Signal	Total number of foci identified outside of cytoplasmic regions
NonNuclear Signal	Total number of foci identified outside of all nuclei
Nuclear Signal	Total number of foci identified within all nuclei
Nuclei Mask	Black and white binary image where the total nuclei area and count is measured from
Nuclei Overlay	Pseudo-color image with the nuclear ROI overlaid on top of the raw nucleus image
Nucleus Count	Total number of nuclei identified in the image based on the "nuclei ROI" zip file
Percent Cytoplasmic Signal	Percentage of total foci that are cytoplasmic (cytoplasmic signal divided by total signal multiplied by 100)
Percent Nuclear Signal	Percentage of total foci that are nuclear (nuclear signal divided by total signal multiplied by 100)
ROI	Region of Interest

Total Nuclei Area	Total area of all nuclei measured from the "nuclei mask" image
Total Signal	Total number of foci identified, which can be identified in the "all foci ROI" zip file