“AQUISTO” - Principle and workflow

“AQUISTO”[1] describes a methodical approach to quantify stainings in a histological section on a slide. Its aim is to provide a workflow collection or straightforward pipeline for standardized analysis of whole experimental sets. Automatic image acquisition of whole organ sections is followed by the generation of size reduced previews of every section on the slides with their slide labels and overview. These overviews are used to create a database with information about every scanned section (File_Names.csv) and experiment/staining (General_Information.csv) in the folder Registration_Tables. According to the information in these tables the files are renamed and sorted. In the next step a manual selection of the tissue area and if necessary the tissue compartments is done. With these selections the .tif-image files are created from the raw .czi-files. If the selection of a minor tissue (like glomeruli in the kidney or islets in the pancreas) tiles can be created as well. After these preparation steps a histogram analysis can be run, to show differences in staining intensity between the sections. They help finding aberrant samples. Parameters for nuclear recognition and marker segmentation are set in the next step. Then the processing of the DAPI channel and subsequently of all the other channels can be done. This results in a database with information about: Marker positive area and cell count and total cell count for the entire section, as well as for all selected tissue compartments. At last it is possible to run the standard analysis for data summary and the generation of overview graphs. Before running this step it is necessary to fill out the table “Groups.csv” in the analysis folder of the staining.

To get started: Download the required software (as described in this manual) and the folders "Registration Tables" and "AQUISTO_Macros", then execute the file "AQUISTO.R" located in "AQUISTO_Macros/R" and source the script.

Programs

R[2] (https://cran.r-project.org/src/base/R-3/, program version should not have an impact on code functionality 17.09.18)

R-packages: reshape2[3] and ggplot2[4]

For a more visual programming environment for R: RStudio [5](https://www.rstudio.com/products/rstudio/download/)

For image preprocessing and processing: Fiji[6-8] (is just imagej) (https://fiji.sc/#download, 64-bit version, 17.09.18)

Fiji plugins: MorphoLibJ[9,10] (https://imagej.net/MorphoLibJ#Installation, 17.09.18), BioFormats[11-13], StarDist[14]

To ensure always having the most recent versions of the required plugins, please activate the following update sites in Fiji

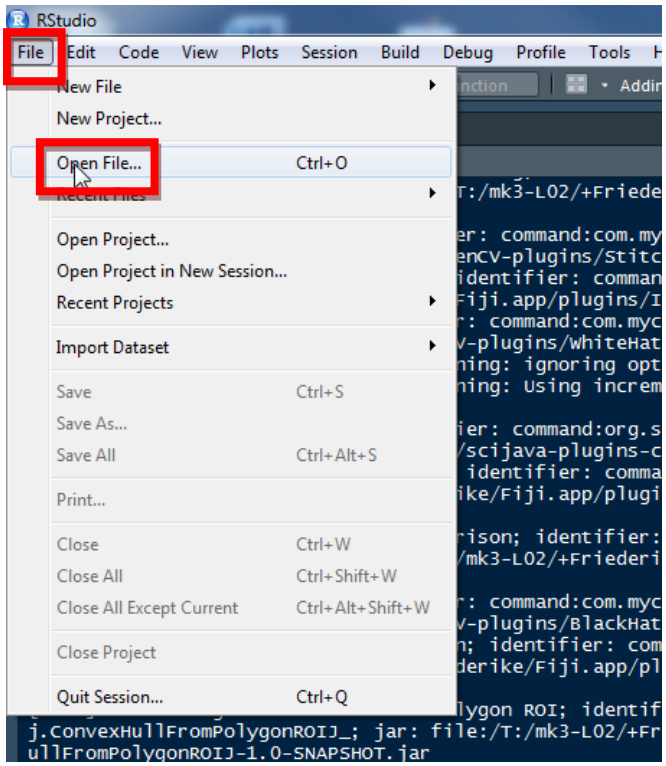
StarDist
CSBDeep
ImageScience
BioFormats
IJPB-plugins

General Preparations

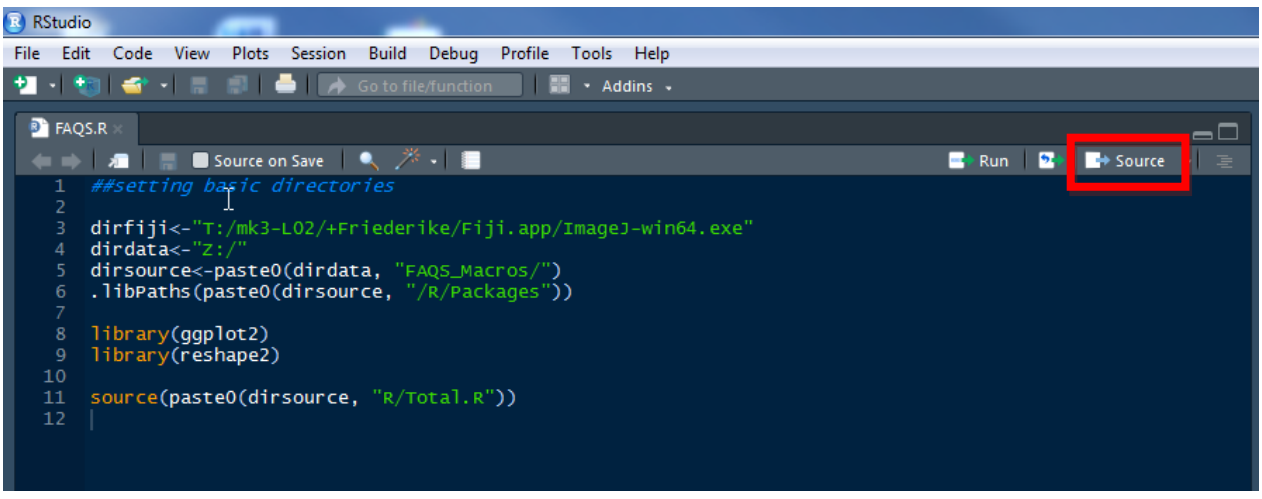
The entire workflow relies on a basic data infrastructure with a base-folder (“AQUISTO”) and subfolders (Experiment, RAW_DATA-new, Registration_Tables, File_Header).

The raw .czi-files are saved without a subfolder directly into the folder “RAW_DATA-new” without changing the automatic output names from the Axioscan (e.g. “2020_11_09__0004.czi”), since these files contain information about date and slide number that are necessary for further steps.

The folder “AQUISTO_Macros” contains the required packages, R scripts and ImageJ macros. The R-script “AQUISTO.R” is the only script that requires interaction. The file is also linked to the main folder “AQUISTO”. Open the script by starting R-Studio, then select “File”>”Open File” and choose the script.



To start the program, click “Source” on the top right corner of the script.

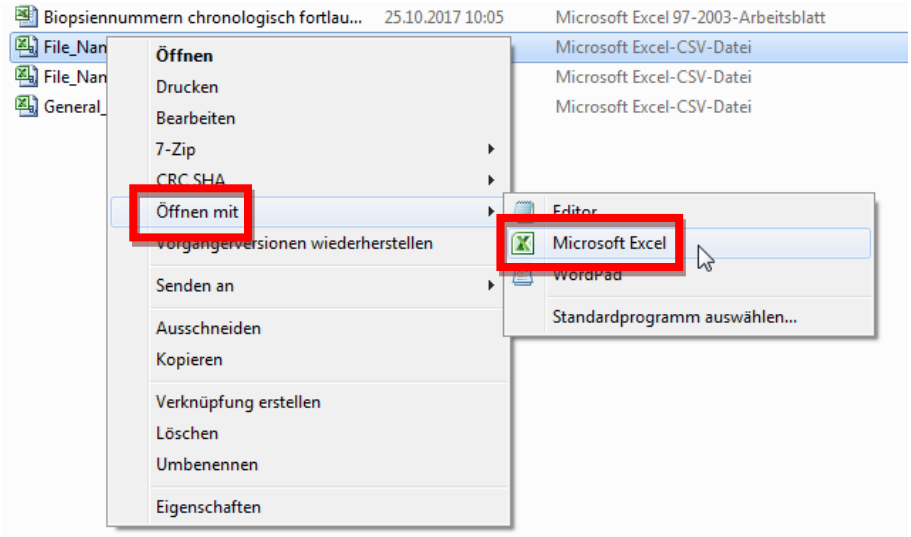


Upon the first run, basic directories are set and additional subfolders are generated within the main “AQUISTO” folders. This is based on the location of the script “AQUISTO.R” that you just executed. A dialog window will pop up, asking about the location of the .exe to start FIJI/ImageJ. It is usually located within the folder “Fiji.app” on your computer. If you cannot see the window, it might still be in the background, check your taskbar or minimize R-Studio, to see it.

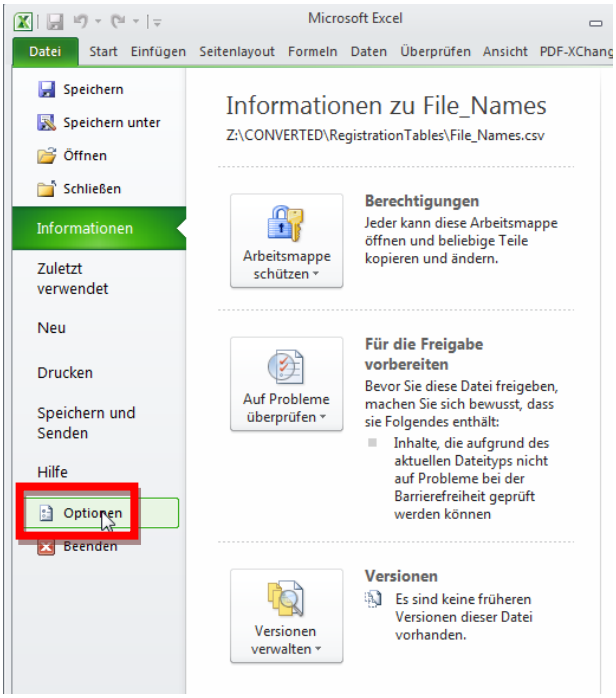
After initialization you can load the raw .czi files directly into the folder “RAW_DATA-new” (not into a subfolder in there). If you want to “hide” older, already processed files from the script, you can put them in a subfolder. If they are in a subfolder, they will not be processed.

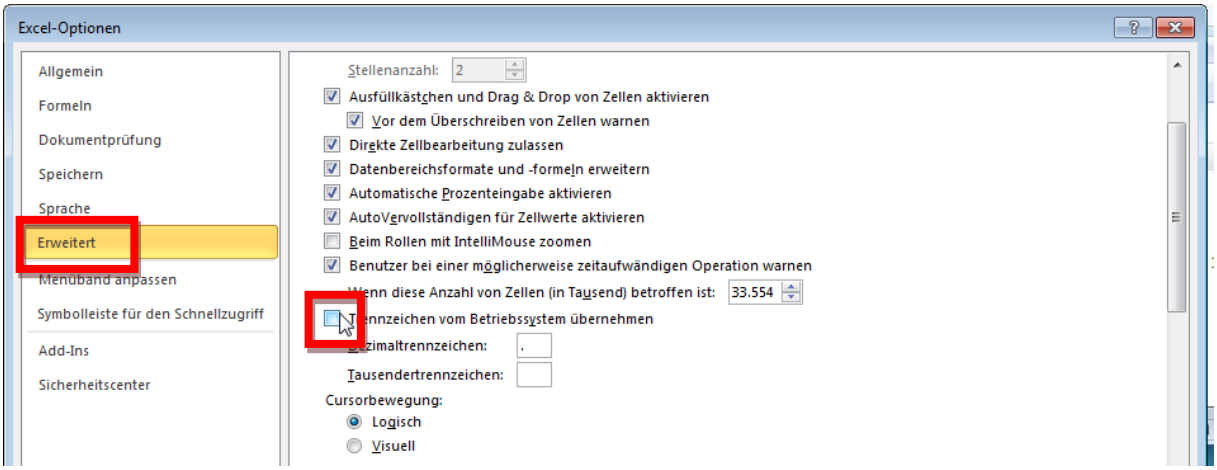
The folder “Registration_Tables” contains the tables “File_Names.csv” (for a database of all scanned sections) and “General_Information.csv” (for the characterization of every experiment and staining). These files are necessary for the automatic renaming and sorting of the files, as well as the generation of experiment-specific subfolders.

Since most datatables are not actual excel files but „Comma separated files“ (csv) with a point as a decimal placeholder and not a comma it is necessary to change general settings for the way excel files are opened. If the .csv-files open in the text editor, change the default program to open them to excel (right click on a .csv-file, “open with”, “excel”).



To change the way decimal numbers are recognized in excel, open excel, click “Start” in the Taskbar, go to “Options”, “Advanced” and scroll down to the checkbox “use operating system default”. Make sure that the box is not checked and the entry for “Decimals” is a point, not a comma. Separator for thousands has to be set as well, for example by simply putting a space. After confirming the settings (“OK”) close excel without saving and open the file again. The decimals should now be displayed correctly.





Sometimes when closing a .csv file, even right after saving it, Excel warns you that you didn't save the file. Simply ignore this message if you're sure you saved it. The other warning that .csv-files don't support certain formatting settings can also be ignored.

START

("AQUISTO.R")

These instructions describe the complete "AQUISTO" workflow step by step. Depending on the experiment and staining characterization some steps may be obsolete (indicated in the paragraph). Mandatory steps are highlighted in red. Start R-studio, then open the program "AQUISTO.R" located in the folder "*/AQUISTO/".

To start the program, click "Source" on the top right corner of the script.

Preparation

Get slide headers and previews

("Preparation">"Get slide headers and previews", ImageJ-Macro "Header_SplitScenes_Previews.ijm").

In the first step a size reduced preview image (1:16) is generated for every section of the slide and saved together with the information about the location on the slide and the label. The macro batch processes all .czi-files in the folder "RAW_DATA-new", and saves the files in a subfolder in "File_Header". If the subfolder already exists the file will be skipped in a re-run. This step does not require any further user interaction. (~3 minutes per slide).

Add file names to list

("Preparation">"Add file names to list", ImageJ-Macro "Entering_File_Names.ijm")

Before starting the processing of Slidescanner aquired images it is necessary to add the new images to the database of scanned images files. Upon run of the macro the data table is opened and previews with their corresponding macro and label image from the AxioScan are shown. With these pictures it is possible to find the biopsy number/name and order of every section on a slide, as well as check for correct focus of the image. From the files present in the folder "File_Header" missing scenes (that have not been entered in the data table) are automatically added to the bottom of the table.

File_Names

	A	B	C	D	E	F	G	H	I
1	Date	Slide	Scene	Biopsy	Labmember	Experiment	Staining	Converted	Focus
2	2017_03_13	1	1	6698-NK-Wdh	Anne	Austestung	PCNA_GFP	Yes	yes
3	2017_03_13	1	2	6698-Wdh	Anne	Austestung	PCNA_GFP	Yes	yes
4	2017_03_13	1	3	6324-Wdh	Anne	Austestung	PCNA_GFP	Yes	yes
5	2017_03_13	2	1	6698-NK-Wdh	Anne	Austestung	Ki67_GFP	Yes	
6	2017_03_13	2	2	6698-Wdh	Anne	Austestung	Ki67_GFP	Yes	
7	2017_03_13	2	3	6324-Wdh	Anne	Austestung	Ki67_GFP	Yes	

This file - "File_Names.csv" - contains a list of all available converted or to be converted images, including the date, the scene number, the slide number and the corresponding biopsy number, experiment name and labmember. Biopsies that are scanned twice in the same staining and experiment by the same labmember (for example for negative controls) have to be indexed "*BiopsyNr-2*" or "*BiopsyNr-NC*" (if the same biopsy exists twice for a staining, experiment and labmember in the list it is skipped during the renaming process). Make sure the specifications for

Date, **Slide** and **Scene** follow the same pattern as the previous samples. Since Fiji sometimes has a problem dealing with folder structures with spaces **don't separate different words with spaces in the list**. It might be possible that the scenes don't follow the same order as the labeling on the slide, so before assigning the biopsy number to a scene check the order of the sections and how they correspond to the scenes in ZEN. **Labmember** refers to the person responsible for the experiment. Each **Staining** should have a unique name within the experiment, so if you stained for the same marker twice within one experiment, with different antibody dilutions for example, it is necessary to index the staining. In this context "**Staining**" refers to a set of biopsies that should be processed and treated identically. For fluorescence that includes all different markers in a staining and also negative controls, different dilutions or antibodies etc. For comparison between different tissue fixations, make sure you specify in either the biopsy-number or the staining, if you also want to apply different processing and analysis parameters. If you notice that the image was out of focus or you want to skip the scene on the slide since it was an unnecessary biopsy that you don't want to waste any more time with type "out of focus", "no" or any other remark into the column "Focus". Make sure that every biopsy you want to process has "yes" copied into that column. Any biopsy without a "yes" in the column "Focus" will be skipped by the program. Spelling is important in this case, and this column is also case-sensitive: If you type "YES", the biopsy will not be processed!

Leave the colum "converted" empty, it will be filled by the program, once the file has correctly been found and only serves to check later, whether the set was processed entirely.

This step is completely manual (~1 minute/slide).

Add general information about a new staining

("Preparation">"Add general information about a new staining", included in the main R-Script).

Further input about every staining is entered. This includes information about the scanned co-marker channels, the name of the sectioned organ and tissue compartments that should be selected. Minor tissues refer to small tissue compartments that cannot be selected in an overview picture of the section (like glomeruli in the kidney or bronchiole in the lung). Follow the required user instructions for input. Major tissues are always considered to be mutually exclusive tissue compartments. For the kidney that means that: Organ = Kidney, Major tissue compartments = 2, Medulla and Cortex (no part of the medulla is part of the cortex and vice-versa), minor tissue compartments: Glomeruli.

The file "General_Information.csv" is necessary a) to give an overview over the existing experiments and stainings and b) for the different preprocessing steps to "know" if they are needed and applicable. For example if glomerular selection is not desired, the creation of tiles of the sample is not necessary. Here it is important to make sure the **Labmember**, **Experiment** and **Staining** are identical to the ones specified in the file "File_Names.csv" (copy-paste it from there to avoid any typing errors). Furthermore the spelling of the specification for the **type of staining, counterstaining** and the **pseudocolors** is critical for the program to work. **The markers must not contain underscores (_).** For the markers only the spelling of **"DAPI"** and **"Brightfield"** play a role in the correct processing.

Type of staining: Peroxidase, Fluorescence, PAS, Sirius red

Counterstaining: Hematoxylin, DAPI

Pseudocolors: Red, Blue, Green, Grays, Cyan, Magenta, Yellow

For the following steps to work, the **assigned marker names MUST NOT contain any spaces or underscores!**

In the number of scanned channels also include the DAPI channel.

Major tissue compartments define distinctive subtissues in an organ, that do not overlap. In the kidney this would be cortex and medulla. Minor tissues are tissues that are too small to be selected in the overview image (Glomeruli in the kidney, Islets in the pancreas). This step can also be done without the assistance of the program, by simply entering the information into the data table manually.

Sort and rename files

("Preparation">"Sort and rename files", R-Script "Sample_Characterization.ijm").

According the information in the files "File_Names" and "General_Information" the previews created in the first step are automatically renamed, relocated to the folder "Experiments" and assigned to a lab member, experiment and staining. Additional folders applying to every staining, like folders for results, images and the analysis are created. Furthermore subfolders indicating the tissue compartments and order of the co-marker channels are generated. Depending on the necessity to select minor tissues and crop these details from the original image the folders "Tiles" and "Cropped_*" are set up

This step is only done if the staining folder does not contain the subfolder "Analysis" and doesn't require any more user input (max. 5 minute total). After filling out both data tables and renaming and sorting all the files the corresponding folders in "File_Headers" should be deleted.

Manual Selection

Major tissue selection

("Manual selection">" Major tissue selection", ImageJ-Macro "Tissue_Selection.ijm").

This is the first step that requires setting a working directory. Choose the experiment folder, the program automatically goes through all the staining subfolders and images. It is advised to use a graphics display to optimize the selection of the tissue.

The size reduced preview is opened and the section area is recognized automatically. To make it easier and more accurate the whole selection is done on a overexposed and blurred image which is automatically thresholded and a selection is automatically created. The overexposed image disappears and you can correct the selection in the 'normal' image. Exclude artifacts like airbubbles, tissue-folds, dye crystals or overexposed tiles by holding 'Alt' and circling the respective area. These parts are now excluded from the selection and will be cut out of the image to not disturb the measurement or give false positive results. To add parts of the tissue press "Shift" while making the selection. Additional tissue compartments can be selected according to the information given in previous steps. These selections are automatically a) within the first selection of the entire tissue area and b) mutually exclusive. After confirming the last selection the selection is automatically recalculated to fit the original size image and saved in the folder "ROIs/Original_Tissues/". Since this

selection will be used to crop the image, it is also recalculated to new coordinates and saved in the folder “ROIs/Tissues/”.

This step required manual input (depending on section quality: ~1 minute per section).

Conditions for this macro to run are

a) Existence of the corresponding preview
b) existence of the folder “**ROIs\Original_Tissues**”
c) non-existence of the “BiopsyNr.zip” (if the selection has already been made, the image is skipped).

If you notice that you made a mistake in a selection and the image is still there (abort the run):

a) Delete all ROIs from the ROIManager
b) Close all open images (don’t save!)
c) Select “OK”, Program will abort
d) Start the Macro again, it will bring you the same biopsy again

If you notice that you made a mistake in a selection and the image is not open anymore or the next image is already open:

a) Delete all ROIs from the ROIManager
b) Close all open images (don’t save!)
c) Select “OK”, Program will abort
d) Go to the folder of you experiment and staining, “/ROIs/Original_Tissues/” and delete the latest .zip-file, that should correspond to the wrong selection
e) Go to the folder of you experiment and staining, “/ROIs/ Tissues/” and delete the latest .zip-file, that should correspond to the wrong selection
f) Start the Macro again, it will bring you the same biopsy again

Preprocessing

Set parameters for tiles

(“Preprocessing”>”Set parameters for tiles”, ImageJ-Macro “Tile_Parameters.ijm”)

If the selection of a minor tissue compartment (glomeruli in the kidney, islets in the pancreas) is required it is necessary to create tiles (split the main image into a number of small squares) for the selection. In some cases the adaption of tissue brightness or pseudocolor of these images facilitates the recognition of these structures. Choose the experiment folder to run the macro over all stainings in an experiment. To set the parameters of this selection there is an assisted manual step in which a) you need to select an area in one of the previews that shows the tissue compartment you want to select and b) opens the original size section of this preview and saves the detail you just selected in the folder “*/Macros/Tile_Parameters”. This detail is then processed in to show different settings of contrast and pseudocolour in all channels for this staining. (Careful, do not look at the screen too closely, it can get very annoying!). Note the settings that bring out the region of interest best (channel, maximum, LUT), click “OK” and enter the values as requested.

The macro only runs if the folder "*/Macros/Tile_Parameters" for this staining doesn't exist yet (it is generated throughout the run). These parameters only have to be set for Immunofluorescent stainings.

Create Images/Tiles

(Preprocessing">"Create Images/Tiles", ImageJ-Macro "Images.ijm).

In the next step the full sized image is opened from the raw .czi-files and cropped to the previously selected tissue area. Every channel is saved as a separate image file in a folder with the identifier name assigned to this section (for fluorescent images) or as RGB (brightfield). The images are scaled to actual resolution (0.325 µm/px for fluorescence and 0.422 µm/px for brightfield). At the same time pixel intensity distributions are saved for every channel and section (saved in the folder "Histograms" and, if applicable, the image is split to tiles, which are processed as set in the previous step (saved in "Tiles").

Conditions for this macro to run are

a) Existence of the corresponding image in "**Raw_Images**"
b) Existence of the tissue selection in "ROIs/Original_Tissues"
c) Non-existence of the image folder in "**Images**"
d) For splitting into tiles: non-existence of file "**Tiles/Biopsy.Nr**"

This step doesn't require further manual input (~10 minutes per section). After the images were created the .czi-files in "RAW_DATA-new" can be archived

Manual Selection

Minor tissue selection

("Manual Selection">"Minor tissue selection", ImageJ-Macro "Tile_tissue_Selection.ijm")

This selection is done manually in the previously generated tiles. Upon start of this step select the experiment folder, the macro automatically processes all stainings in which this step is necessary.

The macro creates a virtual stack of all the tiles of an image for a sample. Select the glomeruli in every tile and add them to the ROI-Manager by pressing "**T**". You can go to the next tile by scrolling down or clicking the arrow at the bar at the bottom of the window. After going through all the tiles click "**OK**" and the macro saves the ROI-Set as "**ROIs\Tiles_*Minortissue*\BiopsyNr.zip**" and opens the tiles for the next biopsy. Cancelling the macro works by removing all the ROIs from the ROI-Manager and then clicking "**OK**", resulting in a loss of the unfinished selection. If you pressed "OK" before actually finishing the selection for the whole biopsy (25 Tiles) delete the .zip-file in your staining folder under "ROIs/Tiles_*Minortissue*" and under "ROIs/*Minortissue*". Since the minor tissue was also added to the .zip in "ROIs/Tissues", open the corresponding .zip file and delete the last ROI within the zip file (*Minortissue)*. You can then abort the program and restart it.

Conditions for this macro to run are

a) Existence of the corresponding tiles
b) Non-existence of the ROI-Set in "**ROIs\Tiles_M*inortissue* \BiopsyNr.zip**" or "**ROIs*Minortissue*\BiopsyNr.zip**"

For AG Hugo: Internal Information ----------
If the server Z:\\Nephro is working slow and every tile takes >5seconds to open it would save a lot of time to transfer a few tile-sets to the computer (Desktop) and continuing locally:

a) Create your folder (e.g.: "Tiles_Experiment") and the subfolder ("Tiles_Staining") containing another subfolder ("Tiles" >> Spelling is crucial)

b) Copy the folder "ROIs" from the Z:\\nephro server ("**Experiments\yourname\experimentname\stainingname**") into the folder "Tiles_Staining" on the PC-Desktop

c) Copy up to 60 folders with the biopsy numbers containing the tiles from the Z:\\nephro server ("Experiments\yourname\experimentname\stainingname\Tiles") into the "Tiles"-folder within "Tiles_Staining" on your desktop, wait for at least 2-3 folders to be transferred before starting the program

d) Run the program and select the folder "Tiles_Experiment" on the desktop

Even when it seems like a big effort to do all this you save a significant amount of idle loading time for every single tile. Even if it's just 3 seconds per tile, that's already a reduction by 30% for every single tile. After the macro reached an end, try and start it again on the same folder, if it still says 'command finished' copy another set of tiles to the desktop (you can also do this in advance) and transfer the new ROI-Sets (ROIs\Tiles_*Minortissue*\) from your desktop to the server.

The ROIs for the tissue compartments and minor tissues are automatically recalculated to fit the full size image. The minor tissues are also combined to one selection and added to the major tissue compartments as one ROI.

This step is manual and requires (depending on the type of selection) 5-10 minutes of manual work.

Fluorescence Processing

Histogram Analysis

("Fluorescence Processing">"Histogram Analysis", R-Script "Histogram_Analysis.ijm")

This R macro uses the data generated during the creation of the full size images to create histogram distribution diagrams for every channel of a staining. Choose the staining folder for which you want the analysis when running the program. The graphs are saved in the folder "Histograms". Statistical outliers in this histogram distribution are automatically designated with the corresponding biopsy number in the graph. These sections can be considered "aberrant" due to artifacts or changes in staining intensity. This step doesn't require further manual input (<1 minute per staining).

Choose samples

("Processing">"Choose samples", ImageJ-Macro "Sample_Selection.ijm")

To set and verify segmentation parameters for the processing and analysis of every section in a staining set sample images are chosen. These samples should cover the range of the staining

intensity shown in the histogram distribution analysis. Start this step and select the staining folder you want to process. Select 5-10 biopsy numbers that cover the range from low to high staining intensity in every channel, maybe including positive and negative controls as well as "normal" samples from different experimental groups. After choosing the biopsy numbers the preview for this biopsy is automatically opened. The pre-selected rectangle is 125x125px big, resulting in a 2000x2000 px detail in the original size image. Move the rectangle to a representative area within this image and press "OK". The next preview is opened. After all selections have been made the program uses the manual selections to crop these details from the original size image and saves them in the folder "Macros/Samples".

This step is split into a manual part (<1 minute per biopsy to be used) and an automatic part (~1 minute per biopsy).

Parameter setting nuclei (via marker controlled watershed or StarDist)

("Fluorescence Processing">"Parameter setting nuclei", ImageJ-Macro "Watershed_Array.ijm" for marker controlled watershed, "StarDist_Array.py for StarDist)

In the first approach nuclear detection relied on the separation of DAPI stained particles by the plugin "MorpholibJ marker-controlled watershed"[9]. In brief, in the DAPI channel the DAPI positive area is defined, edges to background are created and maxima are detected within the image. From every maximum located in a DAPI positive area the surrounding pixels are flooded until reaching an edge or a different flooded area, which is then separated. This results in binary images with separated particles. For optimal nuclear detection three parameters are adapted: To ensure only finding one maximum per nucleus the image is processed with a Gaussian filter to equalize granules (Figure 2d) and the noise tolerance for maximum detection is set. For the recognition of DAPI positive area and its edges the intensity threshold is defined.

Run the program and choose the staining folder as a directory. To create an array, showing different combinations of parameters and find possibly fitting results for the sample click "OK" (~2 minutes automatic). A stack of images is created with different parameters for threshold (z-stack, 16 to 28 in steps of 2), Gaussian sigma (columns, 1 to 5 in steps by 1) and noise tolerance (rows, 2 to 10 in steps by 2). Outlines of the detected particles are shown as an overlay over the original DAPI-image. Go through the layers of the stack and find the setting that shows the best detection and separation of nuclei, note the parameters for threshold, sigma and noise. Confirm ("OK") and enter these parameters as requested. The program shows the output with these parameters for all sample images created in the previous step. Verify that these parameters work equally well for all images.

Note: a decrease in noise tolerance leads to more separated nuclei, a decrease in the Gaussian sigma leads to more separated nuclei, a decrease of threshold leads to smaller and fewer nuclei that are also more likely to be separated. Adapt the parameters until they reliably detect and separate nuclei in all sample details. Confirm this by choosing "Yes" if the processing was successful. To repeat the sample-processing with different settings select "No". Try to avoid the generation of a new array. If you don't know where to start: Treshold = 24, Sigma = 1.5, Noise = 5.

Information about the chosen parameters for DAPI is saved as a text file (*\Macros\Processing). This is a manual step, that depending on staining quality and routine requires 20 minutes of manual time per staining set.

If you want to apply a pretrained deep learning network integrated in the StarDist plugin, use this part to check if it works on your nucleus staining. Select the choice “StarDist”, the program will run on the DAPI-channel details and show an array of the output. If the output looks reasonable, it is safe to assume it will work on the whole experimental set.

Parameter setting co-channels

(“Fluorescence Processing”>”Parameter setting co-channel”, ImageJ-Macro “Channel_Array.ijm”)

Strategies to segment positive signal from background in the co-marker channels rely largely on signal intensity and intracellular localization of the marker. A Gaussian or median filter reduces noise within marker positive areas and equalizes staining intensity. Background subtraction may significantly enhance the signal to noise ratio but can also enhance the intensity of staining artifacts. A general histogram adaption, setting the background pixel intensity to 0, also increases signal to background ratio while only having a linear impact on the general intensity distribution. For segmentation a fixed value can be applied if staining quality is very consistent over the entire set. We also found that applying a threshold adapted to median pixel intensity and intensity standard deviation provides good results. However, this method is restricted to stainings with less than 50 % of co-marker positive area and a limited variance of co-marker positivity between the sections. Once the co-marker channel is segmented to a binary image postprocessing steps can enhance nuclear coverage (dilation of the signal), especially for cytoplasmatic markers, or exclude small artifacts (removal of small bright and dark outliers).

Run the program and choose the staining folder as a directory. To create an array, showing different combinations of parameters and find possibly fitting results for the sample click “OK” (~2 minutes automatic). Find parameters that reliably detect marker-positive areas in the image and note them. Confirm (“OK”) and enter the parameters as requested. The program shows the output with these parameters for all sample images created in the previous step. Verify that these parameters work equally well for all images. Confirm this by choosing “Yes” if the processing was successful. To repeat the sample-processing with different settings select “No”. Try to avoid the generation of a new array.

Information about the chosen parameters for co-channel segmentation is saved as a text file (*\Macros\Processing). This is a manual step, that depending on staining quality requires 20 minutes of manual time per co-channel.

I am currently working on an iterative, error minimizing approach, in which (in the sample images), the segmentation is done supervised, to create the selection of what “should” be positive and negative (“ground truth”). Roughly 3 million different parameter combinations are then applied to every sample detail that was selected earlier and compared to the “ground truth”, to minimize false positive and false negative area in the processed image. These iterations, although optimized for efficiency, are still time consuming (~8h for one co-channel) but provide the possibly best solutions for processing, without requiring machine learning or deep learning approaches. This iterative approach will be included at a later timepoint.

Nucleus Detection

(“Fluorescence Processing”>”Channel Processing”, select the DAPI channel, ImageJ-Macro “Nucleus_Detection.ijm” with marker controlled watershed, “stardist_script_aquisto.px” for StarDist)

Before processing the co-marker channels the DAPI channel for nuclear detection ALWAYS has to be processed first. Choose the staining folder you want to process as a directory. With the parameter set in a previous step the DAPI channel of every section in a staining set are processed equally. For this the image is opened, and a marker-controled watershed is performed according to the specifications. The resulting binary image is translated to a single selection, defining the coordinates of all nuclei within a section. This ROI-Set is saved for the section in the folder "Results/Nuclear_ROIs".

This step doesn't require further manual input (~5 minutes per section).

If the option for StarDist was selected (only available if there are no processing parameters for the marker controlled watershed in the DAPI channel), the pretrained model for nucleus recognition processes the images by tiles and the resulting binary image is translated to a single selection, defining the coordinates of all nuclei within a section. This ROI-Set is saved for the section in the folder "Results/Nuclear_ROIs".

The process with StarDist takes up to 1h per section and does not require further manual input.

Channel processing

("Fluorescence Processing">"Channel Processing", ImageJ-Macro "Channel_Detection.ijm").

Run the program, choose the staining folder as a directory and select the co-channels you want to process (multiple selections possible). Only process the co-channels, after processing the DAPI channel and – if applicable – selecting the minor tissues for every sample.

For the co-channels, after applying filters and background adaption, the image is segmented according to the parameters set in a previous step. The ROI-Set for the tissue selection and tissue compartments is loaded and overall positive area is measured for every tissue compartment in the segmented image ("Results/Total_Area"). Subsequently the ROI for the detected particles in the DAPI channel is loaded and used to create a black and white image with segmented particles. Every single particle is analyzed for the different tissue compartments for overlap with the segmented co-channel image ("Results/Nuclear_Coverage"). If minor tissue compartments were selected every single entity is measured separately in the same way as described for tissue compartments (total co-marker positive area, co-marker overlapping with every nucleus). Additionally for every particle area and circularity is assessed and the information in the segmented image is saved, to ensure the possibility of backtracing of marker-positive areas in the original image ("Results/Processed_Overview/").

This step doesn't require further manual input (~5 minutes per section and co-marker).

Brightfield Processing

Histogram Analysis

("Brightfield Processing">"Histogram Analysis", R-Script "Histogram_Analysis.ijm")

This R macro uses the data generated during the creation of the full size images to create histogram distribution diagrams for every channel of a staining. Choose the staining folder for which you want the analysis when running the program. The graphs are saved in the folder "Histograms". Statistical outliers in this histogram distribution are automatically designated with the corresponding biopsy

number in the graph. These sections can be considered “aberrant” due to artifacts or changes in staining intensity. This step doesn’t require further manual input (<1 minute per staining).

Choose samples

(“Brightfield Processing”>”Choose samples”, ImageJ-Macro “Sample_Selection.ijm”)

To set and verify segmentation parameters for the processing and analysis of every section in a staining set sample images are chosen. These samples should cover the range of the staining intensity shown in the histogram distribution analysis. Start this step and select the staining folder you want to process. Select 5-10 biopsy numbers that cover the range from low to high staining intensity in every channel, maybe including positive and negative controls as well as “normal” samples from different experimental groups. After choosing the biopsy numbers the preview for this biopsy is automatically opened. The pre-selected rectangle is 125x125px big, resulting in a 2000x2000 px detail in the original size image. Move the rectangle to a representative area within this image and press “OK”. The next preview is opened. After all selections have been made the program uses the manual selections to crop these details from the original size image and saves them in the folder “Macros/Samples”.

This step is split into a manual part (<1 minute per biopsy to be used) and an automatic part (~1 minute per biopsy).

Parameter setting co-channels

(“Brightfield Processing”>”Parameter setting co-channel”, ImageJ-Macro “Channel_Array.ijm”)

Strategies to segment positive signal from background in the co-marker channels rely largely on signal intensity and intracellular localization of the marker. A Gaussian or median filter reduces noise within marker positive areas and equalizes staining intensity. Background subtraction may significantly enhance the signal to noise ratio but can also enhance the intensity of staining artifacts. A general histogram adaption, setting the background pixel intensity to 0, also increases signal to background ratio while only having a linear impact on the general intensity distribution. For segmentation a fixed value can be applied if staining quality is very consistent over the entire set. We also found that applying a threshold adapted to median pixel intensity and intensity standard deviation provides good results. However, this method is restricted to stainings with less than 50 % of co-marker positive area and a limited variance of co-marker positivity between the sections. Once the co-marker channel is segmented to a binary image postprocessing steps can enhance nuclear coverage (dilation of the signal), especially for cytoplasmatic markers, or exclude small artifacts (removal of small bright and dark outliers).

Run the program and choose the staining folder as a directory. To create an array, showing different combinations of parameters and find possibly fitting results for the sample click “OK” (~2 minutes automatic). Find parameters that reliably detect marker-positive areas in the image and note them. Confirm (“OK”) and enter the parameters as requested. The program shows the output with these parameters for all sample images created in the previous step. Verify that these parameters work equally well for all images. Confirm this by choosing “Yes” if the processing was successful. To repeat the sample-processing with different settings select “No”. Try to avoid the generation of a new array.

Information about the chosen parameters for co-channel segmentation is saved as a text file (*\Macros\Processing). This is a manual step, that depending on staining quality requires 20 minutes of manual time per co-channel. Currently only the colour deconvultion of PAS- and Sirius Red staining is possible in this version of AQUISTO. Further histochemical stainings, including DAB, will be added in the future.

I am currently working on an iterative, error minimizing approach, in which (in the sample images), the segmentation is done supervised, to create the selection of what "should" be positive and negative ("ground truth"). Roughly 3 million different parameter combinations are then applied to every sample detail that was selected earlier and compared to the "ground truth", to minimize false positive and false negative area in the processed image. These iterations, although optimized for efficiency, are still time consuming (~8h for one co-channel) but provide the possibly best solutions for processing, without requiring machine learning or deep learning approaches. This iterative approach will be included at a later timepoint.

Processing

("Brightfield Processing">"Channel Processing", ImageJ-Macro "Channel_Detection.ijm").

Run the program and choose the staining folder as a directory. Only process the co-channels, after – if applicable – selecting the minor tissues for every sample.

For the co-channels, after applying colour deconvolution, filters and background adaption, the image is segmented according to the parameters set in a previous step. The ROI-Set for the tissue selection and tissue compartments is loaded and overall positive area is measured for every tissue compartment in the segmented image ("Results/Total_Area"). If minor tissue compartments were selected every single entity is measured separately in the same way as described for tissue compartments (total co-marker positive area). Additionally the information in the segmented image is saved, to ensure the possibility of backtracing of marker-positive areas in the original image ("Results/Processed_Overview/").

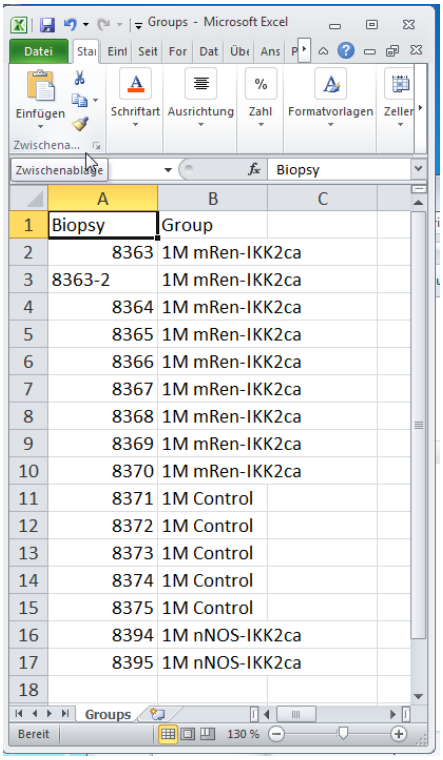
This step doesn't require further manual input (~5 minutes per section and co-marker).

Analysis

Standard analysis

("Analysis">"Standard Analysis", R Script "Slidescan_Analysis.R").

To make the database more accessible for users and give an overview about possibly interesting changes between groups and tissue compartments we used statistical computing with R. For the results from every section to be grouped in a reasonable way the algorithm requires the table "*\Analysis\Groups.csv" to be filled out, assigning every section to an experimental group (fill out before starting the run).



Run the program and choose the staining folder as a working directory. . The order of experimental groups can be set as well. For the co-marker positive area and co-marker intensity in the tissue compartments a summary table is generated, combining the results from all co-markers. This is repeated for the minor tissue compartment, with a separate table for every co-marker and selection area and statistical values like mean, standard deviation, minimum and maximum for every section. The combined lists for every single entity within a section are available in a subfolder. To analyze the contents of the DAPI particle database particles are defined as valid nuclei by parametric exclusion. For this distribution histograms of particle area and circularity are generated and limits can be set. For nuclei in a mouse kidney section an area between 6 and 45 µm² and a circularity of >0.5 were reliable predictors for correct detection of valid nuclei. In the next step nuclei are classified as co-marker positive by co-marker overlap. To set these thresholds, the staining intensity, intracellular co-marker localization and processing strategies, especially postprocessing steps like signal dilation are important. Experience shows that changing the settings here might shift the results up or down, without actually changing relative changes towards controls. Also for nuclear stainings, such as PCNA or TUNEL we found that a coverage around 60% delievers good results, simply because the staining intensity decreases towards the outer rim of the nuclei, leaving only the core actually positively segmented after thresholding and resulting in a relatively low nuclear coverage (considering it is supposed to be a nuclear marker). However for cytoplasmic markers like a-8-integrin the minimum coverage can be set as high as 95%, since by dilating the signal and closing holes the general coverage of positive cells is assumed to be very high. Distribution histograms about co-marker coverage distributions are set up to support this decision. With these parameters, a data table listing the indices of every excluded particle and of particles positive for each co-marker is created. This file contains very long arrays of numbers assigning single particles to their classification as correct nuclei and co-marker positivity. With this information the fraction of co-marker positive nuclei can be calculated, but the data table can also serve as the source for backtracing events in the original image. Summary tables listing the total number of nuclei and co-marker positive nuclei in every tissue compartment are created. For minor tissue compartments additional tables with statistical

information about standard deviation and tables listing numbers of nuclei and fraction of co-marker positive nuclei for every single entity are made. Through combination of the total area of a tissue compartment and the number of nuclei within, the nuclear density can be calculated. Overview graphs are automatically generated to visualize the results summarized in the data tables, while grouping the results to different tissue compartments and experimental groups.

This step requires manual input (“Groups.csv” table, for immunofluorescence only: parameters for nuclear identity, marker positivity, ~2 minutes per staining). The entire calculation can take up to 10 minutes of automatic computing time.

Data interpretation

The standard output of this workflow is a vast amount of data that needs to be interpreted carefully and always in context of controls and the question at hand. The graphs that are automatically generated for covered area, positive nuclei and morphological parameters are supposed to give an overview about possibly interesting developments. Still it is advised to recheck the data (preferably with positive and negative controls) in the images with the coordinates of the marker positive areas ("Results/Processed_Overview"). The output of this whole workflow includes:

Total area of tissue compartments and single glomeruli

Number of nuclei in major and minor tissue compartments (if DAPI was stained)

From these tables the density of nuclei in major and minor tissue compartments can be calculated

% of co-marker positive area in major and minor tissue compartments

% of co-marker positive cells in major and minor tissue compartments (if DAPI was stained)

From these tables the absolute numbers of co-marker positive cells in the tissue compartments can be calculated.

Data Hygiene

If you are satisfied with the whole analysis and processing results, to reduce the data-volume of your experiment you can delete the tiles. If another processing run might be necessary keep the images. For the archive the staining folder containing only the subfolders "Analysis", "ROIs" and "Macros" can be zipped into a .zip-file ("**LabMember_Experiment_Staining.zip")**. This way, only a fraction of the maximum file size is preserved, while full reproducibility of the data from the raw .czi files is possible.

Citations

1 Kessel, F. *et al.* New automatic quantification method of immunofluorescence and histochemistry in whole histological sections. *Cell Signal* **62**, 109335, doi:10.1016/j.cellsig.2019.05.020 (2019).

2 Computing, R. F. f. S. (ed R Core Team) (https://www.R-project.org/, Vienna, Austria, 2017).

3 Wickham, H. Reshaping Data with thereshapePackage. *Journal of Statistical Software* **21**, 1-20, doi:10.18637/jss.v021.i12 (2007).

4 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

5 RStudio, I. (ed RStudio Team) (http://www.rstudio.com/, Boston, MA, 2019).

6 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).

7 Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol Reprod Dev* **82**, 518-529, doi:10.1002/mrd.22489 (2015).

8 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675, doi:10.1038/nmeth.2089 (2012).

9 Legland, D., Arganda-Carreras, I. & Andrey, P. MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics* **32**, 3532-3534, doi:10.1093/bioinformatics/btw413 (2016).

10 (ed Ignacio Arganda-Carreras; David Legland; Curtis Rueden; Dmitry Mikushin; Jan Eglinger; Johannes Schindelin; Stefan Helfrich; Cédric Charière Fiedler) (2019).

11 Besson, S. *et al.* Bringing Open Data to Whole Slide Imaging. *Digit Pathol (2019)* **2019**, 3-10, doi:10.1007/978-3-030-23937-4_1 (2019).

12 Linkert, M. *et al.* Metadata matters: access to image data in the real world. *J Cell Biol* **189**, 777-782, doi:10.1083/jcb.201004104 (2010).

13 Goldberg, I. G. *et al.* The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biol* **6**, R47, doi:10.1186/gb-2005-6-5-r47 (2005).

14 Schmidt U., W. M., Broaddus C., Myers G. Cell Detection with Star-Convex Polygons. *Medical Image Computing and Computer Assisted Intervention*, doi:10.1007/978-3-030-00934-2_30 (2018).