Pre-Processing

This protocol describes how to take multichannel whole slide scans (*.qptiff files) and extract individual sections and split them into single channel images for processing. This will take a several gigabyte sized file and break it into smaller files that can be easily handled in subsequent steps by most computers. This protocol is a precursor to the "Training," "Validation," and "Analysis" protocols.

NOTE ABOUT FILE NAMES:

Pay special attention to **STEP THREE – 1**. All subsequent processing steps use a file name structure separated by "_" to encode critical metadata about the images. These file names have a root made of 5 parts in the form part1_part2_part3_part4_part5_[other info added by the macros], for example: "WT_GFP_M_0001_FA2_.tif"

For our experiments we used the following metadata categories: "GENOTYPE_REPORTER_SEX_TAG_CONDITION_[other info added by the macros]" but these categories can be customized for the unique needs of your laboratory.

NOTES ABOUT PREPARING SLIDES FOR SCANNING:

- Slides are produced with 3-4 blocks per slide so multiple sections can be stained and imaged at once.
- To preserve antigens, slides are kept at 4^{o}C in the dark, desiccated and under nitrogen.
- It is critical that the entire section is stained uniformly. Take special care during preparation that all slides are treated uniformly.
- Make sure slides have all traces of Pap pen removed with xylene
- Slides need to be re-stained with DAPI 1 ug/ml so the DAPI signal is strong. The DAPI in the mounting medium is not strong enough.
- Optional but needed for injured sections: use Vector® TrueVIEW® Autofluorescence Quenching Kit
- Use Antifade Diamond or VECTASHIELD® Vibrance™ Antifade mounting medium with DAPI
- After staining and curing, store slides at 4^{o}C until imaging.
- Make sure imaged slides have adequate signal intensity (aim for log intensity histogram to be 50% of the dynamic range at least for all channels).
- Make sure there is no signal bleed through
- Include the autofluorescence channel to be used for determining section outlines later (this is optional if using the Vector® TrueVIEW® Autofluorescence Quenching Kit as the autofluorescence signal is very diminished after using this kit)

BEFORE STARTING:

Set the script directory in QuPath by going to Automate→"Set Script Directory"→selected the folder where QuPath scripts are located. Should include (not all used in this protocol):

- Export Training Annotations.groovy
- Adjust BC.groovy
- Export Section Outlines.groovy
- Separate Sections.groovy

Fiji macros can be run using "Plugins" → "Macros" → "Run…" or by dragging the macro file onto the Fiji window and running from the editor (preferred in case small changes or debugging is needed). Some macros work most efficiently when installed so they can be associated with a shortcut key. This can be done FIJI→ "Plugins" → "Macros" → "Install…" and finding the macro to include.

STEP ONE: Make a directory that will contain all the files for the project (Windows)

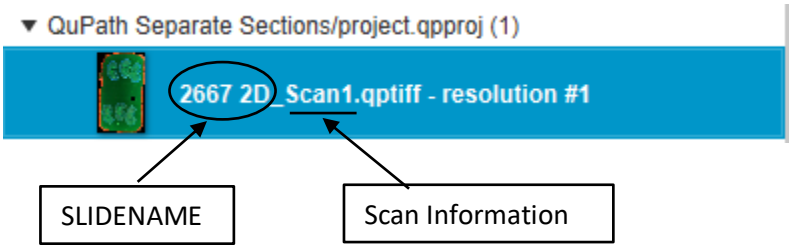
This will include:

1) "**Originals**" (contains original *.qptiff whole slide scan files)
2) "**Separate Slide to Sections Project**" (QuPath project folder)
3) "**Extracted_Decoded_Split_Analyzed**" (Main directory where most of the data is stored). As you work through these protocols this will eventually include:
    a. Separate folders with *.tifs for each channel
    b. QC Montages for qualitative analysis
    c. Folders containing the *.roi outlines for each section and associated folders
    d. Output from UNET segmentations for each channel
    e. Analysis files

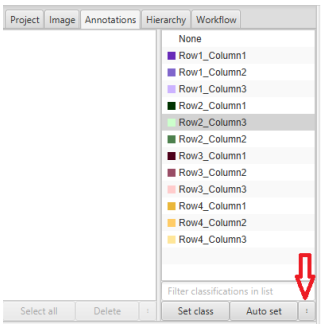
STEP TWO: Separate individual sections into separate pyramidal tiff files (QuPath)

1) Load all raw multichannel image files into folder "Originals" created in **STEP ONE – 1.**
2) Within "**Separate Slide to Sections Project**" (folder created in **STEP ONE – 2**), create a subfolder named "**QuPath Separate Sections**"
3) Within QuPath, generate new project using pathway "Create project" → select folder "QuPath Separate Sections" created in previous step→"Add Images"→"Set image type" to "Fluorescence"→"Choose files"→select all original *.qtiff files contained within "Originals" folder

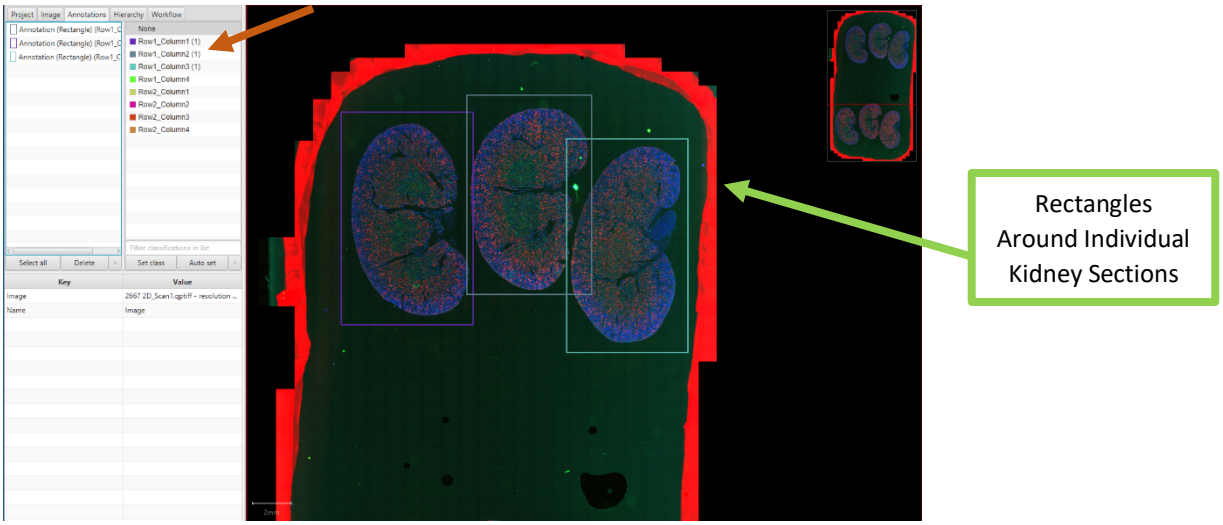
    Note: These files should have the form of "SLIDENAME_Scan1.qptiff – resolution #1" (e.g. "2667 2D_Scan1.qptiff – resolution #1"). The stem before the "_" in the name will be used as the root for separating the images. Anything that follows the "_" will be removed when images are exported below. The "SLIDENAME" will be used to identify the kidney sections in subsequent steps.



4) Use the rectangle tool to define the area of each section, then set the class of each rectangle (right click, "Set class") to define the position on the slide for that section (e.g. "Row1_Column1"). These "classes" will be appended to the file name in **STEP TWO – 6** below and later used to decode the identity of each tissue piece is **STEP THREE - 4**.
    a. To create classes for identification, go to "Annotations" → Three dots next to "Auto set" (see below) → "Add/Remove Class" → "Add Class"



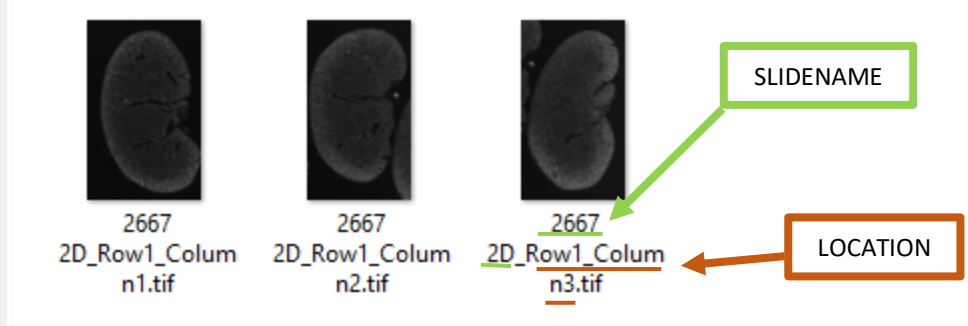
b. It is easiest to use the convention "Row#_Column#"



5) Double check that all sections were accurately labeled. It is difficult to fix this in later steps.
6) Run the QuPath script "Separate Sections.groovy". This script will extract the annotated regions from the image as separate pyramidal *.tif files (the only way these large images can be exported easily). Each file will have the form of "FILEROOT_CLASS.tif" (e.g. "2667 2D_Row1_Column1.tif")

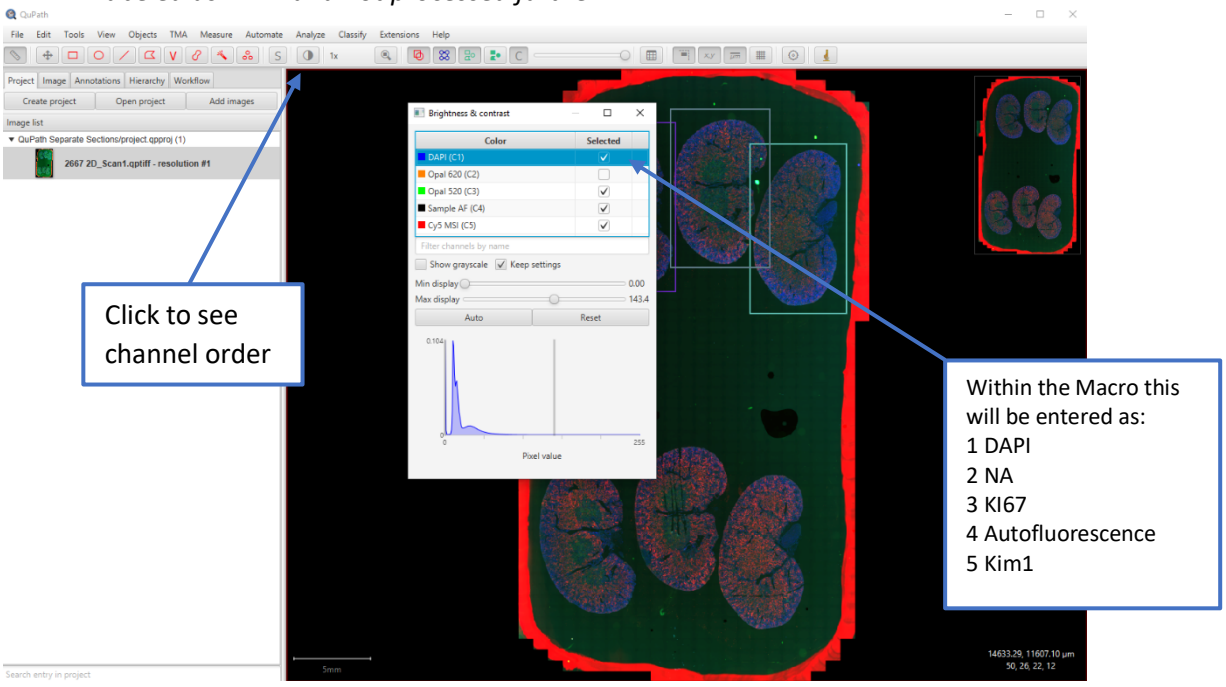
   The pathway to perform this step is: Automate→Shared User Scripts→"Separate Sections.groovy". Then in the editor window: Run→Run for Project→ >> →OK

7) Files will be saved to a subfolder in the project called "Extracted Multichannel Whole Sections"
   a. Move this out of the "QuPath Separate Slide Sections" folder, so that it is another subfolder in "Separate Slides to Sections Project" (*optional, but this will make finding these images easier to find in later steps*)
   b. Examples of the images generated are below:



8) Before closing the QuPath project, we need to record the order of the channels and make sure they are the same for all images.

   a. For the first image in the project click on the brightness and contrast button to see the channels. Make a note of the names and the order of the channels. In practice, we scan all slides for 4 channels (DAPI, Opal 620, Opal 520, Cy5 MSI) and the scanner generates an autofluorescence channel as well called "Sample AF". You will need to refer to notes for each slide to relate the scan wavelengths (for example "Cy5 MSI") to the target in the stain (for example "Kim1" in the example below). The order of the stains will be needed in **STEP THREE**. *Note: in the example below there was no stain in the "Opal 620" wavelength. The generated images contain only background fluorescence and were labeled as "NA" and not processed further.*



   b. Open each additional image in the project and repeat **STEP TWO -- 8a** above to confirm all images have the same order. This is occasionally a problem if scans need to be repeated.

STEP THREE: Generate single channel *.tif files for each image (Fiji):

This step (1) separates individual images by channel, (2) decodes sample information, (3) adds this metadata to the file name, and (4) saves as regular *.tif files for processing

1) Generate a key to converts the slide and position information created in **STEP TWO** to a file name root that contains 5 categories of metadata (see notes above). For example: the section "2667 2D_Row1_Column1.tif" will be renamed with the root: "WT_GFP_M_0001_FA2_".

   This file will be a *.csv file that contains two columns. Column 1 header "Input File Name" and Column 2 Header "Output File Root".

   The parts of the file name are used in all subsequent steps, so it is CRITICAL that they are accurate! DOUBLE CHECK THIS FILE!! It is very time consuming to make corrections at later steps. Incorrect file name conventions will cause errors in subsequent macros.

   a. We generated an Excel file that uses a formula to make the file names from the various categories for each sample.
   b. Double check the spreadsheet for errors

Output

Slide	Input File Root	Row	Column	Genotype	Reporter	Sex	Tag	Condition	Input File Name	Output File Root
2667 2D	2667 2D	1	1	WT	GFP	M	0001	FA2	2667 2D_Row1_Column1.tif	WT_GFP_M_0001_FA2_
2667 2D	2667 2D	1	2	WT	GFP	M	0002	FA2	2667 2D_Row1_Column2.tif	WT_GFP_M_0002_FA2_
2667 2D	2667 2D	1	3	WT	GFP	M	0003	FA2	2667 2D_Row1_Column3.tif	WT_GFP_M_0003_FA2_

   c. Copy the "Input File Name" and "Output File Root" columns and use "Paste values" to paste into a new tab.

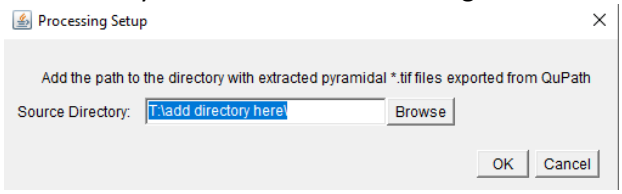
Input File Name	Output File Root
2667 2D_Row1_Column1.tif	WT_GFP_M_0001_FA2_
2667 2D_Row1_Column2.tif	WT_GFP_M_0002_FA2_
2667 2D_Row1_Column3.tif	WT_GFP_M_0003_FA2_

These are copied and pasted into new Excel SHEET

   d. Save the Key file as a *.csv in the project's main directory. SAVE AS →File Type→"CSV(comma defined. (*.csv))"

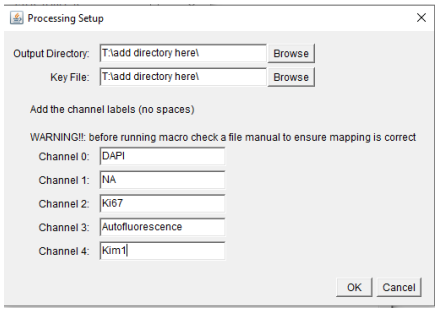
2) Open Fiji
3) Run the macro "Preprocessing_STEP_THREE_Process_Slides_to_Single_Channel.ijm"

   a. In the first dialog box, the "Source Directory:" Browse for "Extracted Multichannel Whole Section" folder created in **STEP TWO – 7.** The macro will search this file to find out how many channels there in the images.



   b. In the second dialog box, enter the rest of the information needed for processing:

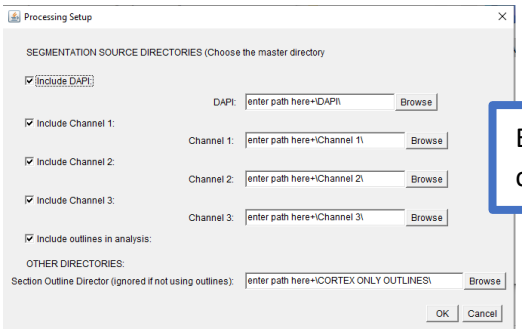
i. The “Output Directory” will be “Extracted_Decoded_Split_Analyzed” created in **STEP ONE – 3.**
ii. The “Key File” will be the one generated in **STEP THREE – 1**.
iii. Enter the stain name for each channel. One of these should be DAPI.



3) Check that the images generated have the appropriate per pixel scale (Fiji→Analyze→Set Scale…)

STEP FOUR: Review images for artifacts (Fiji)

1) Open Fiji
2) Run the macro: “Preprocessing_and_Analysis_Make_QC_Montages.ijm”. This macro will generate montage images to review before further downstream processing to identify any staining or scanning issues.
    a. Along the left side of the user interface, select the channels that you want include in the processing (minimum of one channel). You can select DAPI and up to 3 channels additional channels. The information in the path boxes will be ignored if the box is unchecked.
    b. Leave the “Include outlines in analysis” **unchecked**. Outlines will be generated in the “Analysis” protocol.
    c. For each selected channel enter the path to folders with single channel *.tif files. These path folders were created in **STEP THREE—2**



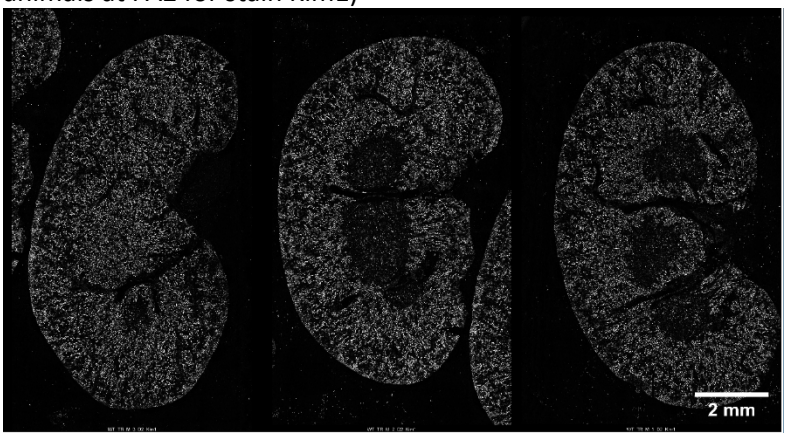
Example Input for creating QC Mosiacs

    d. Click OK to run

3) The macro will generate montages for each channel and put them in a folder called “QC_montages([timestamp]). For each channel, by default ,this macro will sort and group the images by the first and fifth parts of the file name (for example for a file name of structure genotype_reporter_sex_tag_condition_channel…other info.zip it will sort by “genotype” and by “condition,” that is the image "WT_GFP_M_0001_FA2_Ki67.tif” will be grouped in the same

montage as all other images with “WT” in the first position and “FA2” in the second position. You can adjust this by editing the macro code on line 139 and 140.

4) Carefully review the output for any problems. These are large JPEG files but should have enough resolution to identify issues. The file name is printed below each image. (example below, WT animals at FA2 for stain Kim1)



5) The QC_montage folder can be renamed to make it clear what it contains (for example, “QC_Montages Pre Segmentation”)