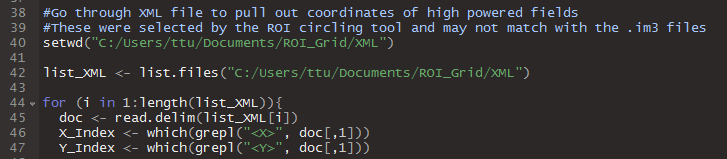
The purpose of this R Script is to identify the nearest high powered field that is actually taken by the Vectra microscope (.im3 files) to a set of boxes drawn by the region of interest circling tool in Phenochart. The region of interest circling tool allows us to systematically grid and sample 100%, 50%, 25%, 11% or 6% of the high powered fields in a tissue. However, the coordinates of those fields do not actually correspond to any of the coordinates on the images taken at a 20x whole slide scan so this script was written to the coordinates of those fields to its nearest neighbor 20x whole slide scan. There are two parts to this script: 1) **the image part** **which is based on using a list of .im3 files** and 2) **the text file part** **which is based on using a merged cell\_seg\_data.txt file or a cell\_seg\_data\_summary.txt file.**

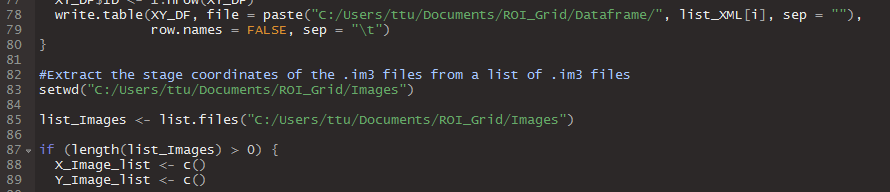
NOTE: This script will only work if 1) the images were taken under Vectra3 (otherwise there would be no Phenochart ROI circling tool to create .xml file) and 2) the images were whole slide scanned at the 20x magnification. **Basically use this script only if you have whole slide scanned at 20x in Vectra3 and you want to retroactively get a percentage of the high powered fields in a region of interest.**

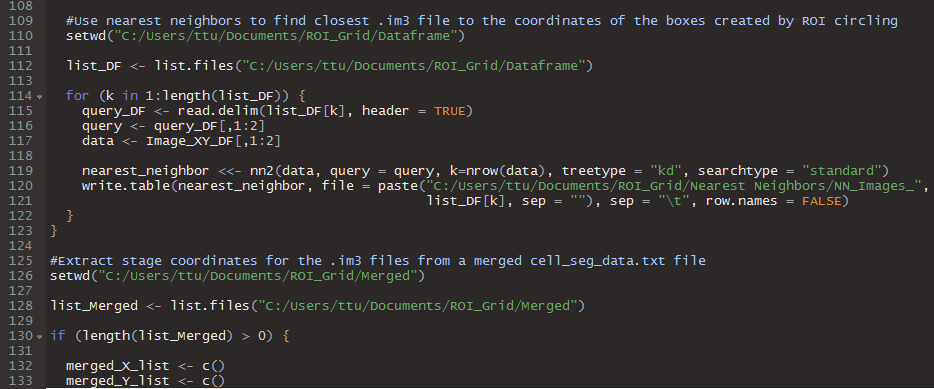
**Classic Workflow:** Scan whole slide 20x images in Vectra3 -> Circle in Phenochart -> ROI\_Grid.R

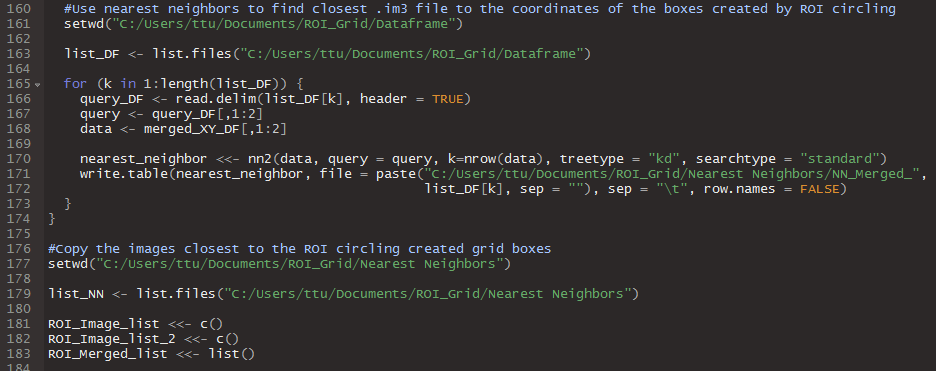
**WARNING: This particular script should be run on only ONE PATIENT AT A TIME! There should be only one file in XML and there should be only one cell\_seg\_data.txt or cell\_seg\_data\_summary.txt file in Merged (if needed) and there should be only one patient’s worth of whole slide scanned HPFs (.im3) in Images.**

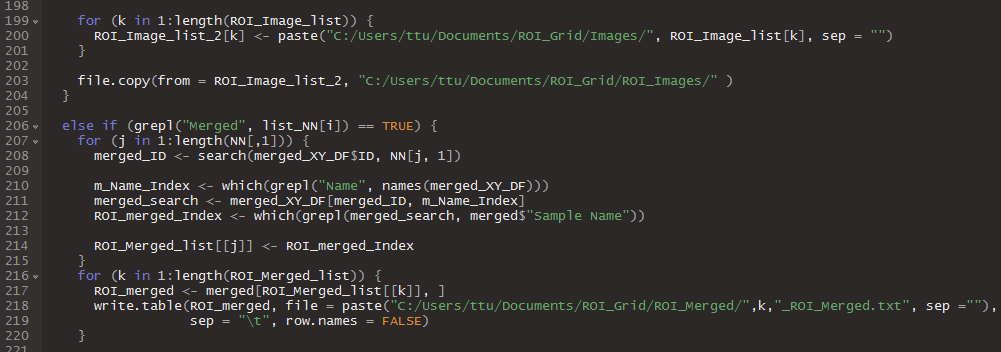
1. The folder **ROI\_Grid** should have nine subfolders: **Dataframe, Deduplicated, Final, Images, Merged, Nearest Neighbors, ROI\_Images, ROI\_Merged,** and **XML**.
   1. Check to make sure the ROI\_Grid folder with the nine subfolders is in the working directory (the stuff inside the **setwd()**, **list.files()**, **file.rename(), file.copy(),** and **paste()** functions). There may be other subfolders but these are the only necessary ones.
      1. If it is not in the working directory, change your working directory in the areas boxed in red below to wherever the ROI\_Grid folder is.
      2. Make sure to keep the slashes **“/”** at the end of the string (the stuff inside the quotes in green) when there is a **paste** function or the directory will be incorrect.
      3. In the final line of the script (write.table()), feel free to change the name of the final output file by changing the content in the string (the stuff in green in the quotes) as long as it includes “.txt” at the end.
         1. Currently it says “”Pt7\_cell\_seg\_summary\_6%.txt”

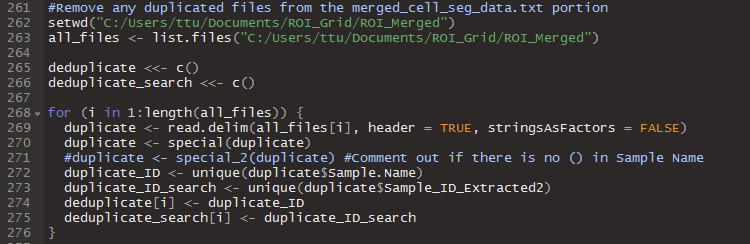


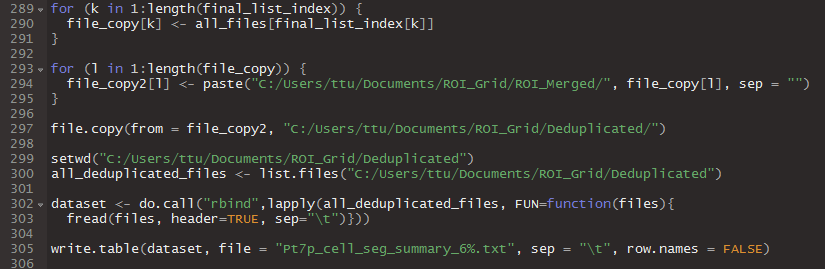




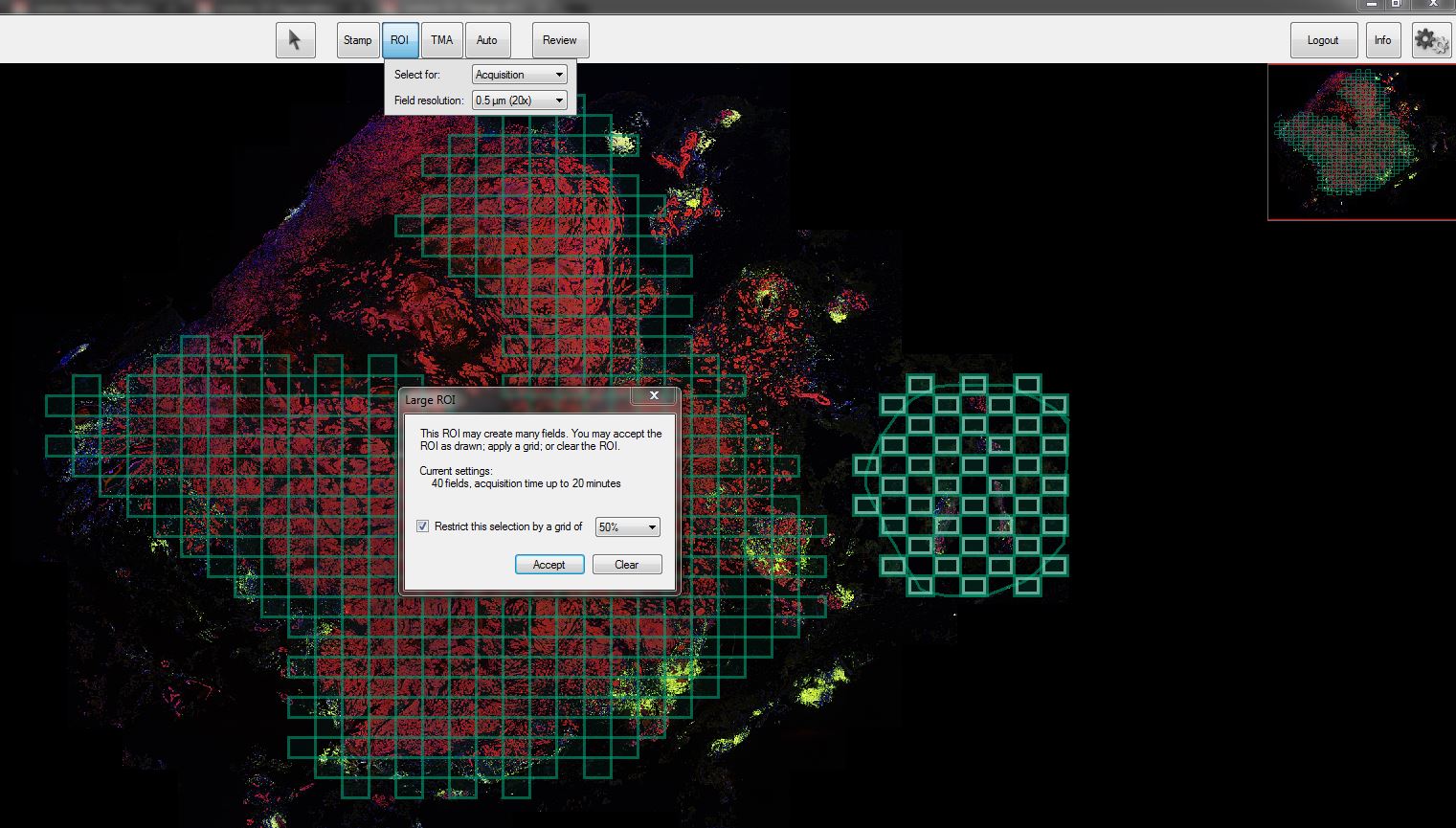


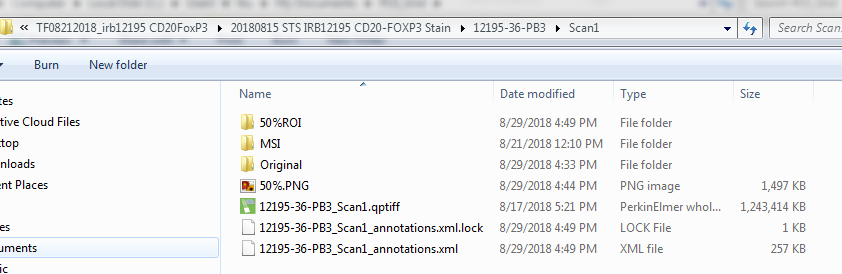




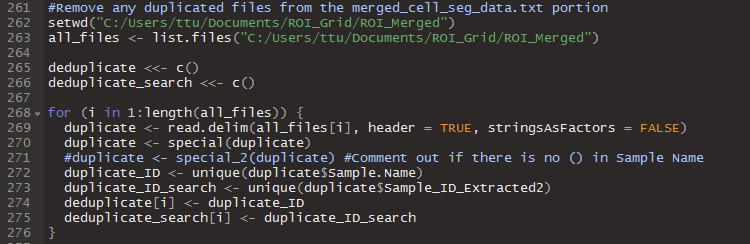


1. The **XML** subfolder should contain **one .xml** file. The **Images** subfolder should contain all the **.im3** images from the whole slide 20x scan. The **Merged** subfolder should contain a **merged cell\_seg\_data.txt file** or **cell\_seg\_data\_summary.txt** file. Before running the **Dataframe, Deduplicated, Final, Nearest Neighbors, ROI\_Images,** and **ROI\_Merged** subfolders should be empty. **The results for the image part will go into the ROI\_Images subfolder and the results for text file part will go into the Dedpulicated subfolder.**
   1. **The .xml file in the subfolder should be converted to a .txt file or else R will not be able to read it. Just click on the file name and delete the “.xml” and replace it with “.txt”.**
   2. The script will still run if the Images or the Merged subfolders are empty. This is important to note because usually only the images part of this script (using the .im3 files list) will be needed. So if you only need to use the image part, just leave Merged empty.
   3. Seriously make sure the listed subfolders that need to be empty are empty before running. **If you run it twice without changing anything, then the “Merged.txt” file will double in size because you merged the already merged file.**
      1. For example if the Deduplicated folder contained File1 and File2 before the first merged, it will contain File1, File2, and Merged.txt after where Merged.txt = File1 + File2. If you merge right away without changing anything, the Split folder will still contain File1, File2, and Merged.txt but Merged.txt now contains File1+File2+Merged.txt.
      2. Also if Nearest Neighbors or Dataframe contains multiple files, it will run them all and the images picked will be completely off
2. Ask Ting-Fang for the details on how to create the XML file from Phenochart. The basics are listed below:
   1. Create a folder named Original in the directory where the Phenochart image (.qptiff) is saved.
   2. Take the original **.xml file and .xmllock file** and **cut and paste** them into the Original folder. This removes any previous notations.
   3. Open the .qptiff file in Phenochart (this software is free to download off the PerkinElmer website).
   4. Click the Login button on the top right hand corner of the image and type in whatever (it really does not matter what you login as).
   5. Click on the ROI button. Under “Select for” it does not really matter if you use “Acquisition” or “Review” and “Field Resolution” is irrelevant as well. The important thing is that this gives you the drawing tool which you can use to circle whatever region interests you
   6. Once you complete the circle a window will pop up (assuming the ROI you drew is large enough; sometimes if it is too small you just get one large useless box. If that happens just circle a larger area). Click “Restrict this selection by a grid of” and choose the percentage you want (100, 50, 25, 11, 6). Then Accept.
   7. Close Phenochart. **The new .xml should have been created in the same directory as the .qptiff file. That is the file that needs to converted to a text file and placed in the XML folder under ROI\_Grid (yellow box below)**

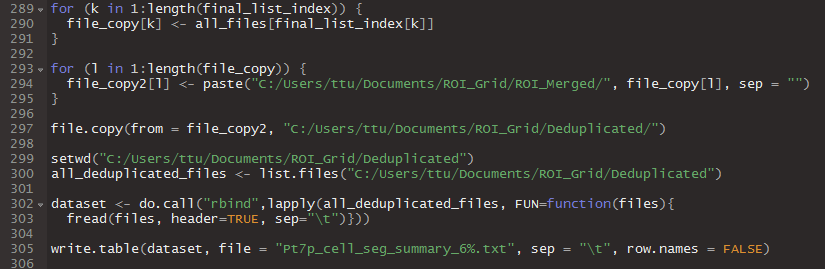




1. If the merged cell\_seg\_data.txt or cell\_seg\_data\_summary.txt file has parentheses in the filename, then uncomment the following line boxed in red below:
   1. Duplicate <- special\_2(duplicate)
   2. The “#” at the beginning of the line creates a comment which is ignored by the computer. The text that has been commented out will appear in blue (at least it does in my view of RStudio). To uncomment the line, just delete the “#” at the beginning of the line. The text should turn white.



1. To change the name of the final merged file for the text part of the script, you can change the string (the stuff in green in between the quotes) in the red box in the figure below. Just make sure it ends with “.txt”.
   1. The images copied over for the image part of the script will have the exact same name as the .im3 file you put into the Images folder



1. Click “Source” in the upper right corner of the scripting window in RStudio to run the script. The result may throw some sort of error at the end if the Merged subfolder is empty but you can just ignore that. The important thing is to make sure ROI\_Images contains approximately the percentage of images you want.
   1. If you are using Merged subfolder then you should also check to make sure the number of items in the Deduplicated subfolder is approximately the percentage of images that you want.