Interim update on Data+ summer project

Students

- Pablo Ortiz
- Lina Yang
- Vivek Sriram

Graduate student mentor

- Robert Ravier

Volunteer data scientist

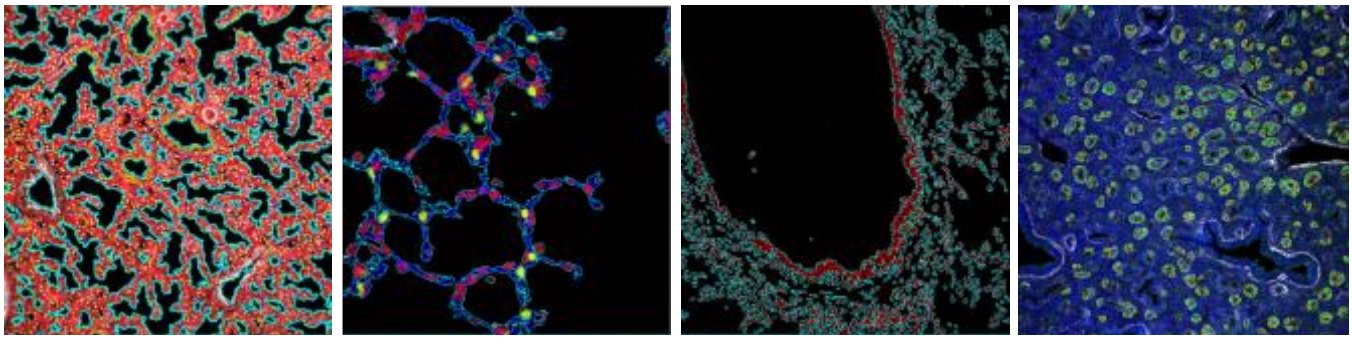
- Ben Neely

Faculty advisors

- Cliburn Chan
- Megan Neely

With help on image analysis pipeline from

- Aronow, Bruce
- Deeptha Girish
- Vineeta Singh



Completed

1. Assembled database of characteristics for 4557 image files. Metadata is being used to group collections of images to process together (e.g. images from E16.5 mouse at 20X with antibodies to X, Y and Z using defined probes) – See Table 1
2. Successfully identify specific features of certain lung images by combining information from the data frame above and image analysis routines from the OpenCV package (Python binidngs) = See Figures 1 and 2 for examples of different image features identified from the same image.

In progress

1. Building up a database of the characteristics of all image segmentation features found in each image such as bitmap mask, xy coordinate of center, label assigned, rule-set used to identify, average R, G, B color intensities. Each image will be processed by multiple rule-sets to identify distinct segments, with rule sets chosen as appropriate based on image metadata in the data frame. For each image, there will be multiple types of image segments identified, and multiple instances of image segment type.
2. Exploratory (visual) analysis of trends over time for the features captured in 1
3. Demo web application where images dragged in will be automatically segmented with legend or mouse-over indicating assigned label (which may be wrong since none of us are lung developmental anatomists, but should be easy for experts to re-label)

Timeline remaining

1. July 20 (2nd group presentation to Data+ program)
2. July 29 Poster presentation and end of Data+ program

For discussion

1. Suggest second meeting or WebEx call with Data+ team/RTI to decide how best to make use of Data+ team efforts and possible integration with LungMAP web site
2. Possibility of paying Lina to continue with the project after Data+ ends to achieve integration with LungMAP (10 hours per week during term time)
3. Ideas for follow up projects at next years Data+

1. img_file
2. X (serial number)
3. image
4. experiment
5. age_label
6. date
7. experiment_id
8. gender
9. label
10. magnification
11. organism_label
12. platform
13. s3downloadkey
14. strain
15. path
16. x_scaling
17. y_scaling
18. researcher
19. site
20. experiment_type
21. release_date
22. color
23. probe_id
24. probe_label
25. target_conditions
26. target_molecules
27. age_group
28. age_group_label
29. cause_of_death
30. crown_rump_length
31. genotype
32. harvest_date
33. health_status
34. local_id
35. race
36. sample_id
37. sex
38. tax_id
39. weight
40. term
41. term_label
42. img_n_channels
43. img_shape
44. img_size
45. img_z_depth

Table 1: Fields in image data frame with information for 4557 images

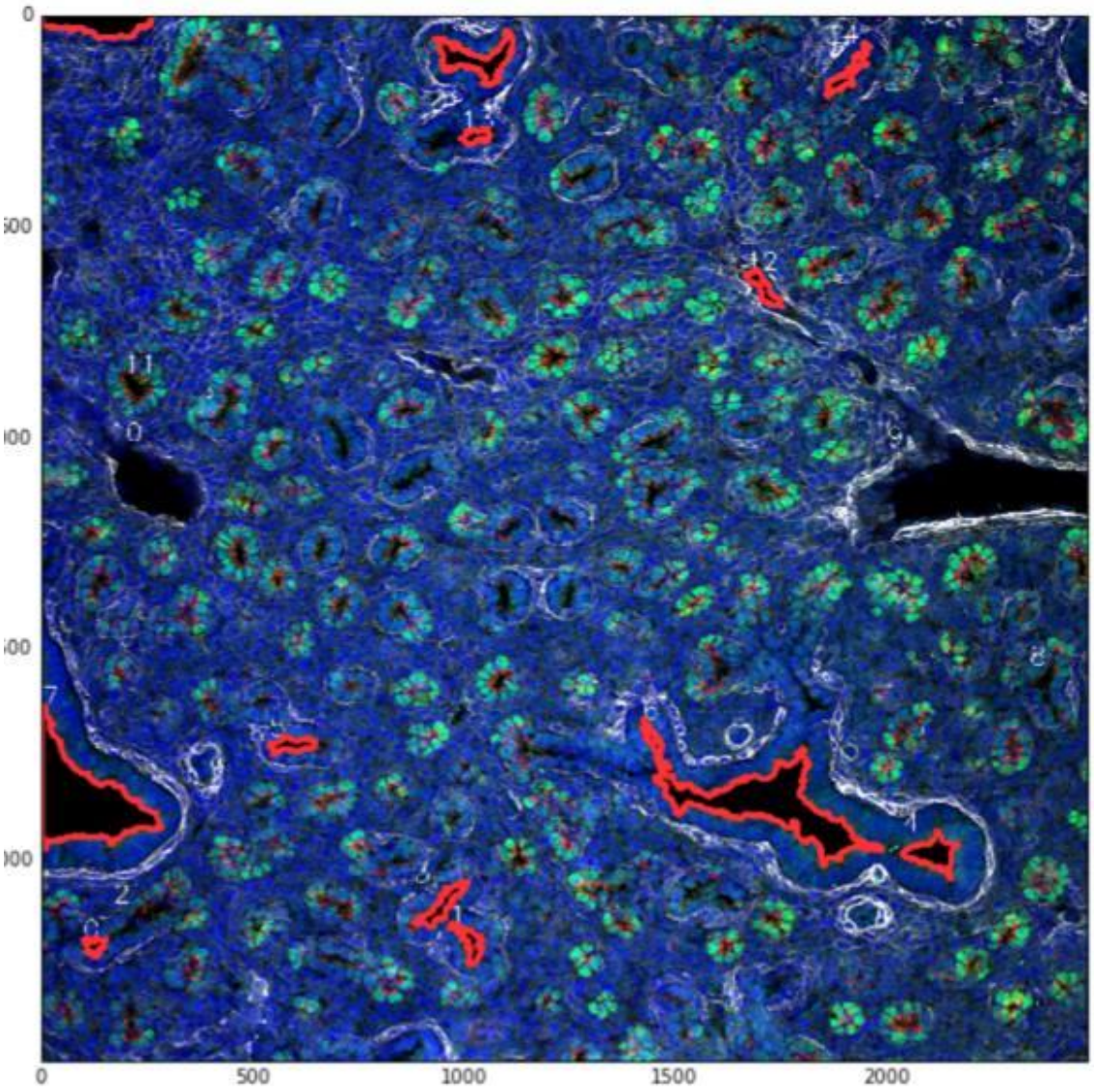


Figure 1: The red highlighted portions in this image correspond to bronchioles. They are defined by the presence of blue mesenchyme inside the white actin filaments.

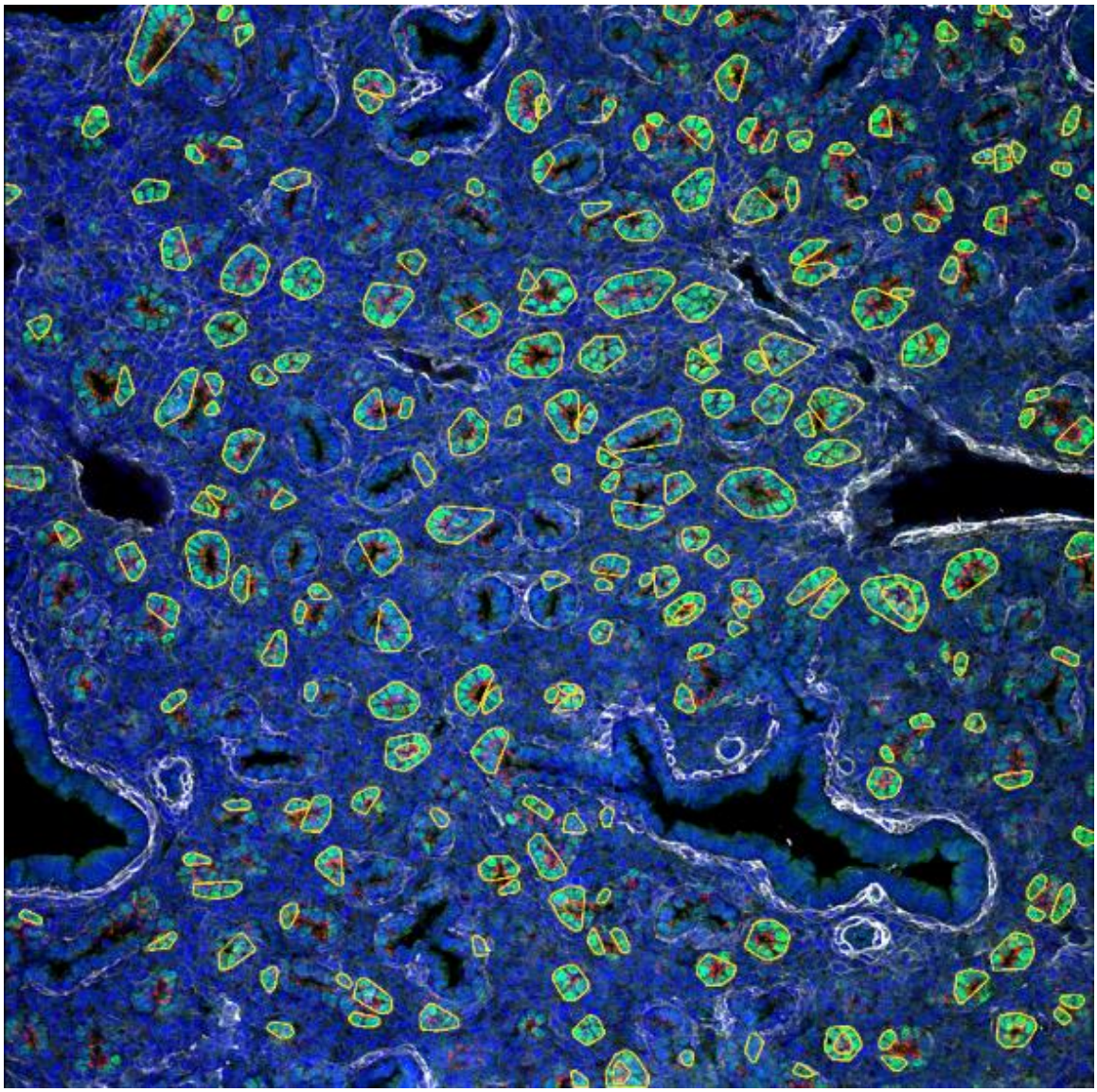


Figure 2: The circled green/red blobs correspond to acini, clusters of cells that serve as precursors to alveoli.