March 10

“

Figure out if we need at least one gene mutation label for each tile name.

* Try first with tile labels only from tiles had KRAS mutation.
* Then, if not working properly, use other gene name to represent tiles without KRAS mutation.

Need to pull out the KRAS mutation Patient ID ---- also Image ID ---- Tile ID list / dict / table. In order to generate this label file.

”

The work should use ALL LUAD tiles in the beginning to set as 100% test instead of splitting the data into 15% test and 15% valid. Then, this should run through model 1 again to get labels with LUAD. Furthermore, proceed to the gene mutation steps. The reason for this is because there will be more dataset available.

But for now, just trying to find out how/what should we set for the label file.

1. Try: -DONE.

Tile labels are composed with tiles only had KRAS mutation. (non-KRAS mutation tiles are omit)

1. Get patients ID that have value of 1 for KRAS mutation from <LUAD\_Patients\_KRAS\_mutation.csv>
2. Grab all Slides contain that specific patient ID as starter for filenames.
3. Generate a label file where the first column is the slide ID and the second the mutation name.

- Slides directory:

/nfs6/deeppath\_scratch/Pre\_processing\_2nd\_Feb21/

Now, proceed with the LUAD results from rerun (Feb). only 15% test used. - 100% .

1. # sort the LUAD tiles identified as LUAD intro a train, valid a test set for mutation analysis