10/06/22

WSI

* De momento he hecho la clasificación image-wise sin darme cuenta, 250 imágenes, 185 casos aprox. Es necesario pasar preprocesamiento y descargar todos los datos de GDC haciendo todo case-wise.
* Para ver más progreso a través de los epochs usar undersampling en vez de oversampling, así las epochs serán más cortas

RNASeq

* RNASeq, 178 casos, por lo que no hay datos de todas las modalidades para todos los casos.
* Descargo archivos STAR-COUNT .tsv 366 archivos, 183 no son autorizados

13/06/22

WSI

* Y si para entrenar el clasificador patch-wise uso todas las imágenes de TCGA pero luego para hacer la clasificación case-wise solo uso los diagnostic slide. Otra opción es expandir con la base de datos GTEX
* Son 244 GB, cinco veces más datos, solución, hacer un undersampling al crear las bases de datos para train, y tmb crear otras bases de datos para el test sin undersampling. De esta forma, no debería petar en el training al cargar todos los splits, ni al hacer el test.

RNASeq

* STAR-COUNTS, primary, solid tissue normal RNA SEQ
* Problema, todos son casos positivos de RNASeq, quizá necesito otras secuencias…
* Comprobar que efectivamente todos los datos son positivos.
* Ordenar los case\_ids y sacarle un sentido.

14/06/22

Sample type:

185 Primary tumor 42 Solid tissue normal

185 Casos en total

183 Casos con WSI -> 2 casos sin WSI

178 Casos con RNASeq -> 7 casos sin RNASeq

185 Casos con CNV

185 Casos con DNAmet

178 Casos con MIRNA -> 7 casos sin MIRNA

Si solo hay 42 muestras de solid tissue normal, tengo casos negativos??? EL POSITIVO O NEGATIVO NO ESTÁ EN CLINICAL, ESTÁ EN LA ETIQUETA

15/06/22

WSI

Hay tres IDs importantes. Patient ID, Case ID, y Sample ID. :

* El 10 CV se hace con el Patient ID
* La clasificación se hace con el Case ID
* El simple ID identifíca tu dato concreto

Patient ID = Case ID: TCGA-YB-A89D ¿?

Sample ID: TCGA-YB-A89D-01A

File ID: TCGA-YB-A89D-01A -01-TS1.53D8C086-00CC-4565-8C52-E81E8F8A8409

“In most cases for each sample, various modalities are available”