<https://cancergenome.nih.gov> go to this link, click “Launch Data Portal” on the right side

Accept the warning

Projects→on the left side→Primary Site choose Colorectal→Cancer Program choose TCGA→Disease Type choose Colon Adenocarcinoma→Data Category(choose 2 More) then choose Simple Nucleotide Variation→Experimental Strategy choose RNA-Seq.

On the right side, click the first row Exp(459).

Click Files(2493), then we only need the FPKM-UQ files

Click the green cart symbol(arrow), choose add all files to the cart, click the Cart on the top, download the Metadata and click the Download symbol arrow to download the Manifest.

From the manifest file, programming to get the filename to match with the patient id(patient id could get from the entity\_submitter\_id(this guy from the metadata). Eg. "entity\_submitter\_id": "TCGA-AD-6889-01A-11R-1928-07" the 6889 is what we need.

Use manifest to download all the data, search ENSG00000141510(this matches TP53 gene, sometimes could be ENSG00000141510.16) in the files, get the numbers matches with that, get this number matches with each file name and then matches these guys with the entity\_submitter\_id, finally match these numbers with the patient\_id.

Match these numbers to the original excel form column 1 patient id, put the numbers into column “tp53 expression”.