

1. Clinical data which includes patients in the rows and clinical information in the columns. Mutation data has the MAF file containing multiple data frames. RNA data contains 3D data frames in which each data frame contains multiple data frames.
2. Clinical data uses a dataframe that has patients on the row and clinical information on the columns. Mutation data uses multiple dataframes in the form of a MAF file. RNA Data uses a 3D dataframe that contains multiple data frames. Protein data also uses a dataframe.
3. No, because this data is usually coded differently due to HIPAA privacy concerns.
4. In the clinical data section
5. Silent mutations will produce the same amino acid despite a change to the codon whereas a non silent mutation can produce a different amino acid when there is a change to a codon. @data contains the non-silent mutations and @maf.silent contains the silent mutations.
6. @dataS/P
7. Gene counts for TP53 can be found in the proteomics data frame
8. This can be found in the MAF file under the @data classification
9. The unit of measurement for this gene expression level.
10. The unit of measurement is the log2 fold change of protein abundance relative to a pooled reference.
11. CPTAC is more limited with only 10 types of cancer as opposed to 33 for TCGA. This is because the CPTAC has close to 129 steps making the pipeline more arduous for processing samples.
12. Not every gene is found in RNA-seq data and proteomic data because not all genes are transcribed and translated.
13. For the “old” group the survival rate is close to 5% at 2000 days. The “middle” and “young” group both appear to be at 12.5%.
14. IDH1 is the 5th most commonly mutated while FLT3 is the most commonly mutated.
15. This plot is called a mutation co-occurrence plot. FLT3 and NPM1 show statistical significance for co-occurrence.
16. This is called a co-lollipop plot. Many of the non-survivors have a missense mutation at PTPc. PTEN\_C2 and PTPc are co-occurring genes.
17. The survivor group is most likely with a relative protein abundance of -1.73 .