**Project: Drug Response Prediction based on multi-Omics data**

**Sample or Clinical Information**

Clinical information have been attached in the excel.

**Data Read with R**

All the data have been saved as Rdata including miRNA-seq, mRNA-seq and methylation array.

**mRNA-seq data**

Each row is a gene and each column is a cancer or normal samples. The values are mRNA expression level. Log2 transfer is recommended before further statistical analysis.

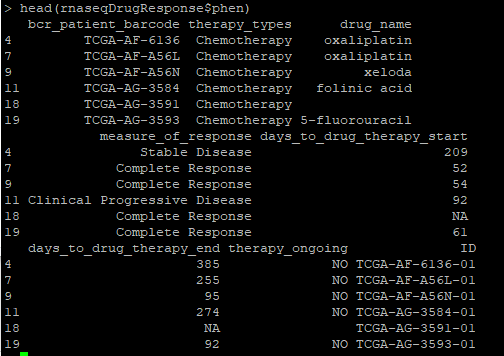
For mRNA-seq data, a R ‘list’ called *rnaseqDrugResponse* was saved.

load("Pancancer.rnaseqDrugResponse.RData")

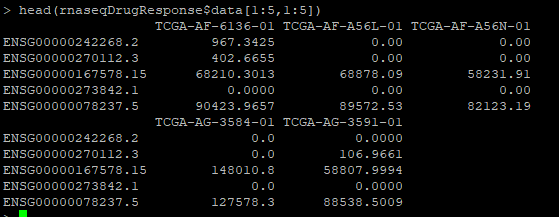
rnaseqDrugResponse$phen

rnaseqDrugResponse$data

head(rnaseqDrugResponse$phen)



As above shown, measure\_of\_response is what we need for our drug response analysis.

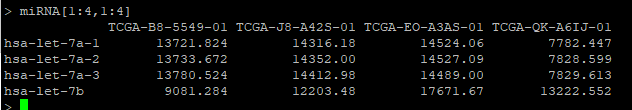


As above show, it is RNA-seq data for all the samples. The first three code is for patient ID, for example: TCGA-AF-6136. The last -01 or -11 is for sample type. -01 indicates it is a cancer sample while -11 indicates a normal sample. The patient id is same as “bcr\_patient\_barcode” in phenotype data frame.

**miRNA-seq dataset**

Same as DNA methylation dataset, each row is a gene and each column is a cancer or normal samples. The values are miRNA expression level. Log2 transfer is recommended before further statistical analysis.

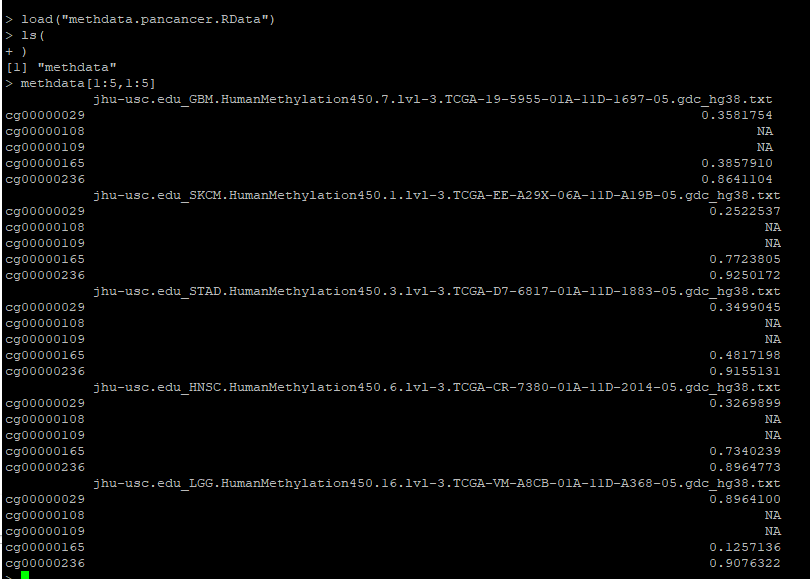
load("TCGA-Pancancer.miRNAseq.RData")



**DNA methylation dataset**

The whole RData is 30G. In the file, each row is a CpG sites in human genome and each column is a cancer or normal samples. The value (Beta) is methylation level for this CpG in the corresponding sample. Beta is the percentage level of the methylation and ranges from 0 to 1 in which 0 indicates totally non-methylation and 1 represent 100% methylation.

load("methdata.pancancer.RData")



**Image data**

We need to discuss how to deal with this data. It is a huge dataset.

Update: 2019/10/20

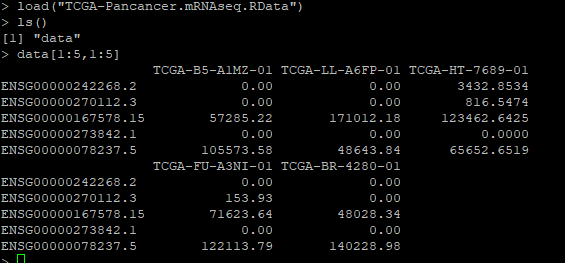
As the discussion during ASHG2019, Dr. Xiong and Juyang want to use RNA-seq data from drug-naïve patients for background/counterfactual effect estimation/simulation. Therefore, I collected all RNA-seq data for Juyang again.

The data were saved in mxiong@HGCNT43

/home/hgc/mxiong/drugResponse/TCGA-Pancancer.mRNAseq.RData

Size=2.7G

Please check the RData structure:



Don’t hesitate to contact me if you have any question about the data.

Shicheng