

## Differential expression

In this lab, we will be conducting a two-sample test for each gene/probe on the array to identify differentially expressed genes/probes between ketogenic rats and control diet rats. This small data set was run on the rat RAE230A Affymetrix array. The objective of the study was to determine differences in mRNA levels between brain hippocampi of animals fed a ketogenic diet (KD) and animals fed a control diet. “KD is an anticonvulsant treatment used to manage medically intractable epilepsies”, so differences between the 2 groups of rats can provide biological insight into the genes that are regulated due to the treatment.

We are going to identify those genes/probes that are differentially expressed between the 2 rat diet groups and plot the results with a couple of different visual summaries.

- 1.) Download the GEO rat ketogenic brain data set and save as a text file.

### Done

- 2.) Load into R, using read.table() function and header=T/row.names=1 arguments.

```
>Lab5File <- "C:/Users/fermi/Documents/Fall 2020/Lab 5/rat_KD.txt"
>dat<-read.table(Lab5File, header=T, row.names=1)
```

- 3.) First log<sub>2</sub> the data, then use the Student's t-test function in the notes to calculate the changing genes between the control diet and ketogenic diet classes. (Hint: use the names() function to determine where one class ends and the other begins).

```
>controls<-dat[,1:6]
>keto<-dat[,7:11]
```

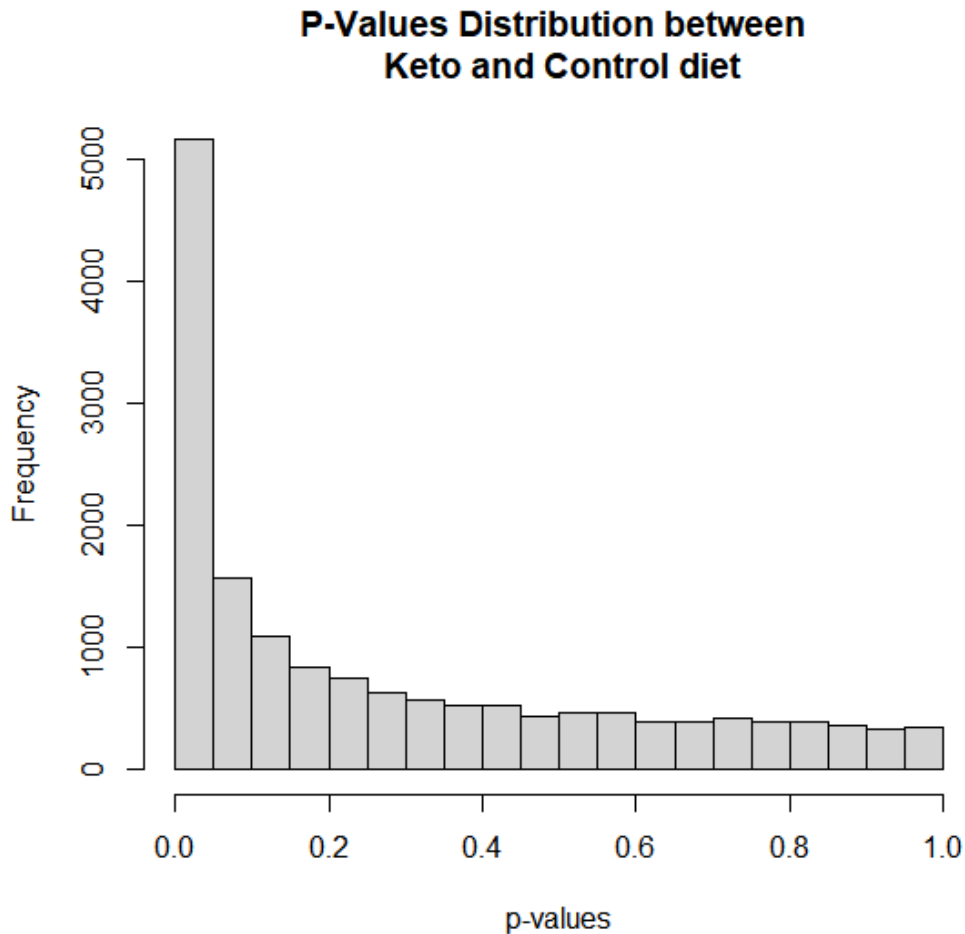
```
> dat2<-log2(dat)
```

```
>t.test.all.genes<- function(x,s1,s2){
  x1<-x[s1]
  x2<-x[s2]
  x1<-as.numeric(x1)
  x2<-as.numeric(x2)
  t.out<-t.test(x1,x2,alternative="two.sided",var.equal=T)
  out<-as.numeric(t.out$p.value)
  return(out)
}
```

```
>pv<-apply(dat2,1,t.test.all.genes,s1=colnames(keto),s2=colnames(controls))
```

4.) Plot a histogram of the p-values and report how many probesets have a  $p < .05$  and  $p < .01$ . Then divide an alpha of 0.05 by the total number of probesets and report how many probesets have a p-value less than this value. This is a very conservative p-value thresholding method to account for multiple testing called the Bonferroni correction that we will discuss in upcoming lectures.

```
> hist(pv, xlab="p-values", main= "P-Values Distribution between \n Keto and Control diet")
```



```
> sum(pv<.05)
[1] 5160
```

```
> sum(pv<.01)
[1] 2414
```

**There are 5160 probosets with a  $p\text{-value} < 0.05$ , and 2414 probosets with a  $p\text{-value} < .01$ .**

```
> critical<-.05/15923
```

```
> sum(pv<critical)
[1] 12
```

**12 probosets have a p-value less than the value of .05/(total number of probosets)**

5.) Next calculate the mean for each gene, and calculate the fold change between the groups (control vs. ketogenic diet). Remember that you are on a  $\log_2$  scale.

```
> controlMeans<-rowMeans(controls)
> ketoMeans<-rowMeans(keto)
```

**#FC for log2 scale = mean(x) - mean(y)**

```
FC<-controlMeans-ketoMeans
```

6.) What is the maximum and minimum fold change value, please report on the linear scale? Now report the probesets with a p-value less than the Bonferroni threshold you used in question 4 **and**  $|\text{fold change}| > 2$ . Remember that you are on a  $\log_2$  scale for your fold change and I am looking for a linear  $|\text{fold}|$  of 2.

```
> 2^(max(FC))
Maximum fold change value is = 55.15521
```

```
> 2^min(FC)
Minimum fold change value is = 0.08240443
```

```
> threshold<-which(pv<critical)
> FCp<-which(abs(FC)>2)
```

**#Meeting Bonferroni threshold:**

1367553\_x\_at 1368071\_at 1370239\_at 1370240\_x\_at 1370355\_at 1371102\_x\_at  
1371245\_a\_at 1388608\_x\_at 1373040\_at 1374641\_at 1390092\_at 1376005\_at

**#Meeting  $|\text{fold}| > 2$ :**

1367553\_x\_at 1387011\_at 1387408\_at 1387696\_a\_at 1387827\_x\_at 1370239\_at  
1370240\_x\_at 1371102\_x\_at 1371245\_a\_at 1371272\_at 1388358\_at 1388608\_x\_at  
1372087\_at 1388804\_at 1373938\_at 1374132\_at 1375213\_at 1375608\_at  
1375758\_at 1394198\_at

**#Meeting BOTH:**

6 total probosets meeting both threshold and fold change: 1367553\_x\_at,  
1370239\_at, 1370240\_x\_at, 1371102\_x\_at, 1371245\_a\_at, 1388608\_x\_at

7.) Go to NetAffx or another database source if you like and identify gene information for the probesets that came up in #6. What is the general biological function that associates with these probesets?

**The general biological function from these probesets are related to hemoglobin and beta globin, with most corresponding to hemoglobin subunits.**