

Coronavirus Liver and Blood Capillary Samples

Janis Corona

2/8/2020

These samples are the headers added from three Gene Expression Omnibus studies at

- ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89166
- ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89160
- ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100509

The first two studies are part of the same study that used human liver tumor samples in vitro to compare the effects of the coronavirus over time. The third study used human microvascular blood capillaries in vitro to study the effects of the coronavirus over time.

In the first two studies that used the liver tumor samples to examine the effects of the coronavirus in vitro, there were four groups inoculated or treated with the active coronavirus and four groups not inoculated with the active coronavirus, and two samples that were treated with heat inactivated coronavirus, and two samples that were treated with active coronavirus and IL-1alpha to see the gene expression changes over one hour's time.

In the the third study that used blood capillaries, there were five samples followed over a 0,12,24,36, and 48 hour time intervals in groups A,B,C,D, and E that compared the time interval values of screening for changes in microarray analysis with a mock group of the same.

This following data is the data of all genes in common between these three studies, cleaned to remove missing values and with the attached gene symbols from the GEO platform for the probe IDs.

```
both <- read.csv('both_clean_liver_capillary_CoV.csv', sep=',', header=TRUE,
                 na.strings=c('', ' '))

dim(both)

## [1] 21754    63

colnames(both)

## [1] "GENE_SYMBOL"
## [2] "LiverTumorSamples.GSM2359851_CoV1"
## [3] "LiverTumorSamples.GSM2359853_CoV2"
## [4] "LiverTumorSamples.GSM2359910_CoV3"
## [5] "LiverTumorSamples.GSM2359913_CoV4"
## [6] "LiverTumorSamples.GSM2359850_ctrl1"
## [7] "LiverTumorSamples.GSM2359852_ctrl2"
```

```
## [8] "LiverTumorSamples.GSM2359911_ctrl3"
## [9] "LiverTumorSamples.GSM2359914_ctrl4"
## [10] "LiverTumorSamples.GSM2359912_IL1"
## [11] "LiverTumorSamples.GSM2359917_IL2"
## [12] "LiverTumorSamples.GSM2359915_inactiveHeatCoV1"
## [13] "LiverTumorSamples.GSM2359916_inactiveHeatCoV2"
## [14] "capillarySamples.GSM2685693_MERS_CoV_0hr_A"
## [15] "capillarySamples.GSM2685694_MERS_CoV_0hr_B"
## [16] "capillarySamples.GSM2685695_MERS_CoV_0hr_C"
## [17] "capillarySamples.GSM2685696_MERS_CoV_0hr_D"
## [18] "capillarySamples.GSM2685697_MERS_CoV_0hr_E"
## [19] "capillarySamples.GSM2685698_ctrl_0hr_A"
## [20] "capillarySamples.GSM2685699_ctrl_0hr_B"
## [21] "capillarySamples.GSM2685700_ctrl_0hr_C"
## [22] "capillarySamples.GSM2685701_ctrl_0hr_D"
## [23] "capillarySamples.GSM2685702_ctrl_0hr_E"
## [24] "capillarySamples.GSM2685703_MERS_CoV_12hr_A"
## [25] "capillarySamples.GSM2685704_MERS_CoV_12hr_B"
## [26] "capillarySamples.GSM2685705_MERS_CoV_12hr_C"
## [27] "capillarySamples.GSM2685706_MERS_CoV_12hr_D"
## [28] "capillarySamples.GSM2685707_MERS_CoV_12hr_E"
## [29] "capillarySamples.GSM2685708_ctrl_12hr_A"
## [30] "capillarySamples.GSM2685709_ctrl_12hr_B"
## [31] "capillarySamples.GSM2685710_ctrl_12hr_C"
## [32] "capillarySamples.GSM2685711_ctrl_12hr_D"
## [33] "capillarySamples.GSM2685712_ctrl_12hr_E"
## [34] "capillarySamples.GSM2685713_MERS_CoV_24hr_A"
## [35] "capillarySamples.GSM2685714_MERS_CoV_24hr_B"
## [36] "capillarySamples.GSM2685715_MERS_CoV_24hr_C"
## [37] "capillarySamples.GSM2685716_MERS_CoV_24hr_D"
## [38] "capillarySamples.GSM2685717_MERS_CoV_24hr_E"
## [39] "capillarySamples.GSM2685718_ctrl_24hr_A"
## [40] "capillarySamples.GSM2685719_ctrl_24hr_B"
## [41] "capillarySamples.GSM2685720_ctrl_24hr_C"
## [42] "capillarySamples.GSM2685721_ctrl_24hr_D"
## [43] "capillarySamples.GSM2685722_ctrl_24hr_E"
## [44] "capillarySamples.GSM2685723_MERS_CoV_36hr_A"
## [45] "capillarySamples.GSM2685724_MERS_CoV_36hr_B"
## [46] "capillarySamples.GSM2685725_MERS_CoV_36hr_C"
## [47] "capillarySamples.GSM2685726_MERS_CoV_36hr_D"
## [48] "capillarySamples.GSM2685727_MERS_CoV_36hr_E"
## [49] "capillarySamples.GSM2685728_ctrl_36hr_A"
## [50] "capillarySamples.GSM2685729_ctrl_36hr_B"
## [51] "capillarySamples.GSM2685730_ctrl_36hr_C"
## [52] "capillarySamples.GSM2685731_ctrl_36hr_D"
## [53] "capillarySamples.GSM2685732_ctrl_36hr_E"
## [54] "capillarySamples.GSM2685733_MERS_CoV_48hr_A"
## [55] "capillarySamples.GSM2685734_MERS_CoV_48hr_B"
## [56] "capillarySamples.GSM2685735_MERS_CoV_48hr_C"
## [57] "capillarySamples.GSM2685736_MERS_CoV_48hr_D"
```

```
## [58] "capillarySamples.GSM2685737_MERS_CoV_48hr_E"
## [59] "capillarySamples.GSM2685738_ctrl_48hr_A"
## [60] "capillarySamples.GSM2685739_ctrl_48hr_B"
## [61] "capillarySamples.GSM2685740_ctrl_48hr_C"
## [62] "capillarySamples.GSM2685741_ctrl_48hr_D"
## [63] "capillarySamples.GSM2685742_ctrl_48hr_E"
```

Lets group the samples that are our columns with descriptive and GEO ID names into their respective groups, get the fold change between the controls from those groups, attach to the original data table, both, as a different names, then order by the genes that have the most fold change then the least fold change. Take the first 100 genes from both lists, combine into one table of 200 genes and the samples with their fold change values ordered, make into a transposed data frame so that the samples are the rows, the stats removed, and the 200 genes are the header columns to save as a machine learning ready file.

Liver tumor study control and CoV treated. Also, the IL-alpha treated and the inactive CoV treated tables are in this code block.

```
names <- both$GENE_SYMBOL

liverCtrl <- both[,c(6:9)]
row.names(liverCtrl) <- names

liverCoV <- both[,c(2:5)]
row.names(liverCoV) <- names

liverIL <- both[,10:11]
row.names(liverIL) <- names

liverIACoV <- both[,12:13]
row.names(liverIACoV) <- names
```

Get the row means of those liver samples groups each.

```
liverCtrl$CtrlMeanLvr <- rowMeans(liverCtrl)
liverCoV$CoVMeanLvr <- rowMeans(liverCoV)
liverIL$ILMeanLvr <- rowMeans(liverIL)
liverIACoV$IACoVMeanLvr <- rowMeans(liverIACoV)
```

Get the fold change values of those states as a ratio to the control group values.

```
fold1 <-
as.data.frame(cbind(liverCtrl$CtrlMeanLvr,liverCoV$CoVMeanLvr,liverIL$ILMeanL
```

```

vr,
      liverIACoV$IACoVMeanLvr))
row.names(fold1) <- names
colnames(fold1) <- c('CtrlMeanLvr', 'CoVMeanLvr', 'ILMeanLvr', 'IACoVMeanLvr')

fold1$FC_CoV <- fold1$CoVMeanLvr/fold1$CtrlMeanLvr
fold1$FC_IL <- fold1$ILMeanLvr/fold1$CtrlMeanLvr
fold1$FC_IACoV <- fold1$IACoVMeanLvr/fold1$CtrlMeanLvr

```

Most expressed in liver samples by fold change of the Coronavirus, inactive CoronaVirus, and the IL-alpha treated Coronavirus as tables.

```

mostCoV <- fold1[order(fold1$FC_CoV, decreasing = TRUE)[0:100],]
mostIL <- fold1[order(fold1$FC_IL, decreasing = TRUE)[0:100],]
mostIACoV <- fold1[order(fold1$FC_IACoV, decreasing = TRUE)[0:100],]

```

Least expressed in liver samples by fold change of the Coronavirus, inactive CoronaVirus, and the IL-alpha treated Coronavirus as tables.

```

leastCoV <- fold1[order(fold1$FC_CoV, decreasing = FALSE)[0:100],]
leastIL <- fold1[order(fold1$FC_IL, decreasing = FALSE)[0:100],]
leastIACoV <- fold1[order(fold1$FC_IACoV, decreasing = FALSE)[0:100],]

```

Gene Expressions with most changes in the liver samples.

```

changes <- rbind(mostCoV,mostIL,mostIACoV,leastCoV,leastIL,leastIACoV)
Changes <- changes[!duplicated(row.names(changes)),]
length(unique(row.names(Changes)))

## [1] 600

```

Get the magnitude of the fold change genes' row means.

```

Changes$MagnitudeFCs <- abs(rowMeans(Changes[,5:7]))

```

Combine this to the samples data for the liver tumor group.

```

Changes$Gene <- row.names(Changes)
combined1 <- merge(both, Changes, by.x='GENE_SYMBOL', by.y='Gene')

combined2 <- combined1[order(combined1$MagnitudeFCs, decreasing=TRUE),]

CombinedLiver <- combined2[c(0:100,354:453),]

```

Machine Learning data for liver samples with 200 genes in the group of most gene expression changes.

```

names1 <- CombinedLiver$GENE_SYMBOL
names2 <- colnames(CombinedLiver)
row.names(CombinedLiver) <- names1

Combo_lvr_ML <- as.data.frame(t(CombinedLiver))

```

```
colnames(Combo_lvr_ML) <- gsub('-', '_', colnames(Combo_lvr_ML))
Combo1 <- Combo_lvr_ML[c(2:63),] #remove stats of fold change values and gene symbol row
```

Lets add a class field called Class_Type to use machine learning on predicting class with these 200 genes and 62 mixed samples of capillary and liver tumor both inoculated with Coronavirus.

```
a <- rep('liver_CoV', 4)
b <- rep('liver_Ctrl', 4)
c <- rep('liver_CoV_IL', 2)
d <- rep('liver_IA_CoV', 2)
e <- rep('capillary_CoV_0hr', 5)
f <- rep('capillary_Ctrl_0hr', 5)
g <- rep('capillary_Cov_12hr', 5)
h <- rep('capillary_Ctrl_12hr', 5)
i <- rep('capillary_Cov_24hr', 5)
j <- rep('capillary_Ctrl_24hr', 5)
k <- rep('capillary_Cov_36hr', 5)
l <- rep('capillary_Ctrl_36hr', 5)
m <- rep('capillary_Cov_48hr', 5)
n <- rep('capillary_Ctrl_48hr', 5)
```

```
type <- as.data.frame(c(a,b,c,d,e,f,g,h,i,j,k,l,m,n))
colnames(type) <- 'Class_Type'
row.names(type) <- row.names(Combo1)
type
```

##	Class_Type
## LiverTumorSamples.GSM2359851_CoV1	liver_CoV
## LiverTumorSamples.GSM2359853_CoV2	liver_CoV
## LiverTumorSamples.GSM2359910_CoV3	liver_CoV
## LiverTumorSamples.GSM2359913_CoV4	liver_CoV
## LiverTumorSamples.GSM2359850_ctrl11	liver_Ctrl
## LiverTumorSamples.GSM2359852_ctrl12	liver_Ctrl
## LiverTumorSamples.GSM2359911_ctrl13	liver_Ctrl
## LiverTumorSamples.GSM2359914_ctrl14	liver_Ctrl
## LiverTumorSamples.GSM2359912_IL1	liver_CoV_IL
## LiverTumorSamples.GSM2359917_IL2	liver_CoV_IL
## LiverTumorSamples.GSM2359915_inactiveHeatCoV1	liver_IA_CoV
## LiverTumorSamples.GSM2359916_inactiveHeatCoV2	liver_IA_CoV
## capillarySamples.GSM2685693_MERS_CoV_0hr_A	capillary_CoV_0hr
## capillarySamples.GSM2685694_MERS_CoV_0hr_B	capillary_CoV_0hr
## capillarySamples.GSM2685695_MERS_CoV_0hr_C	capillary_CoV_0hr
## capillarySamples.GSM2685696_MERS_CoV_0hr_D	capillary_CoV_0hr
## capillarySamples.GSM2685697_MERS_CoV_0hr_E	capillary_CoV_0hr
## capillarySamples.GSM2685698_ctrl_0hr_A	capillary_Ctrl_0hr
## capillarySamples.GSM2685699_ctrl_0hr_B	capillary_Ctrl_0hr
## capillarySamples.GSM2685700_ctrl_0hr_C	capillary_Ctrl_0hr

```
## capillarySamples.GSM2685701_ctrl_0hr_D      capillary_Ctrl_0hr
## capillarySamples.GSM2685702_ctrl_0hr_E      capillary_Ctrl_0hr
## capillarySamples.GSM2685703_MERS_CoV_12hr_A  capillary_Cov_12hr
## capillarySamples.GSM2685704_MERS_CoV_12hr_B  capillary_Cov_12hr
## capillarySamples.GSM2685705_MERS_CoV_12hr_C  capillary_Cov_12hr
## capillarySamples.GSM2685706_MERS_CoV_12hr_D  capillary_Cov_12hr
## capillarySamples.GSM2685707_MERS_CoV_12hr_E  capillary_Cov_12hr
## capillarySamples.GSM2685708_ctrl_12hr_A      capillary_Ctrl_12hr
## capillarySamples.GSM2685709_ctrl_12hr_B      capillary_Ctrl_12hr
## capillarySamples.GSM2685710_ctrl_12hr_C      capillary_Ctrl_12hr
## capillarySamples.GSM2685711_ctrl_12hr_D      capillary_Ctrl_12hr
## capillarySamples.GSM2685712_ctrl_12hr_E      capillary_Ctrl_12hr
## capillarySamples.GSM2685713_MERS_CoV_24hr_A  capillary_Cov_24hr
## capillarySamples.GSM2685714_MERS_CoV_24hr_B  capillary_Cov_24hr
## capillarySamples.GSM2685715_MERS_CoV_24hr_C  capillary_Cov_24hr
## capillarySamples.GSM2685716_MERS_CoV_24hr_D  capillary_Cov_24hr
## capillarySamples.GSM2685717_MERS_CoV_24hr_E  capillary_Cov_24hr
## capillarySamples.GSM2685718_ctrl_24hr_A      capillary_Ctrl_24hr
## capillarySamples.GSM2685719_ctrl_24hr_B      capillary_Ctrl_24hr
## capillarySamples.GSM2685720_ctrl_24hr_C      capillary_Ctrl_24hr
## capillarySamples.GSM2685721_ctrl_24hr_D      capillary_Ctrl_24hr
## capillarySamples.GSM2685722_ctrl_24hr_E      capillary_Ctrl_24hr
## capillarySamples.GSM2685723_MERS_CoV_36hr_A  capillary_Cov_36hr
## capillarySamples.GSM2685724_MERS_CoV_36hr_B  capillary_Cov_36hr
## capillarySamples.GSM2685725_MERS_CoV_36hr_C  capillary_Cov_36hr
## capillarySamples.GSM2685726_MERS_CoV_36hr_D  capillary_Cov_36hr
## capillarySamples.GSM2685727_MERS_CoV_36hr_E  capillary_Cov_36hr
## capillarySamples.GSM2685728_ctrl_36hr_A      capillary_Ctrl_36hr
## capillarySamples.GSM2685729_ctrl_36hr_B      capillary_Ctrl_36hr
## capillarySamples.GSM2685730_ctrl_36hr_C      capillary_Ctrl_36hr
## capillarySamples.GSM2685731_ctrl_36hr_D      capillary_Ctrl_36hr
## capillarySamples.GSM2685732_ctrl_36hr_E      capillary_Ctrl_36hr
## capillarySamples.GSM2685733_MERS_CoV_48hr_A  capillary_Cov_48hr
## capillarySamples.GSM2685734_MERS_CoV_48hr_B  capillary_Cov_48hr
## capillarySamples.GSM2685735_MERS_CoV_48hr_C  capillary_Cov_48hr
## capillarySamples.GSM2685736_MERS_CoV_48hr_D  capillary_Cov_48hr
## capillarySamples.GSM2685737_MERS_CoV_48hr_E  capillary_Cov_48hr
## capillarySamples.GSM2685738_ctrl_48hr_A      capillary_Ctrl_48hr
## capillarySamples.GSM2685739_ctrl_48hr_B      capillary_Ctrl_48hr
## capillarySamples.GSM2685740_ctrl_48hr_C      capillary_Ctrl_48hr
## capillarySamples.GSM2685741_ctrl_48hr_D      capillary_Ctrl_48hr
## capillarySamples.GSM2685742_ctrl_48hr_E      capillary_Ctrl_48hr
```

```
Combo2 <- cbind(type, Combo1)
Combo2[1:10, 1:5]
```

##	Class_Type	NEURL3	DUSP1
## LiverTumorSamples.GSM2359851_CoV1	liver_CoV	1429.61750	8491.40875
## LiverTumorSamples.GSM2359853_CoV2	liver_CoV	190.21750	2219.85650
## LiverTumorSamples.GSM2359910_CoV3	liver_CoV	10.004148	11.494585

```
## LiverTumorSamples.GSM2359913_CoV4      liver_CoV      11.245589      12.898250
## LiverTumorSamples.GSM2359850_ctrl1      liver_Ctrl      34.57000      228.18775
## LiverTumorSamples.GSM2359852_ctrl12     liver_Ctrl      17.25750      216.08550
## LiverTumorSamples.GSM2359911_ctrl13     liver_Ctrl      3.708157      7.184185
## LiverTumorSamples.GSM2359914_ctrl14     liver_Ctrl      4.757780      7.113854
## LiverTumorSamples.GSM2359912_IL1       liver_CoV_IL      4.879242      9.576161
## LiverTumorSamples.GSM2359917_IL2       liver_CoV_IL      5.1138565     9.5527540
##                                           ATF3             PCLO
## LiverTumorSamples.GSM2359851_CoV1      3608.28250      17.74792
## LiverTumorSamples.GSM2359853_CoV2      974.76937      613.28583
## LiverTumorSamples.GSM2359910_CoV3      8.337322      3.355859
## LiverTumorSamples.GSM2359913_CoV4      9.441972      2.741117
## LiverTumorSamples.GSM2359850_ctrl1      108.52250      13.80667
## LiverTumorSamples.GSM2359852_ctrl12     97.05125      12.88750
## LiverTumorSamples.GSM2359911_ctrl13     5.373051      3.904719
## LiverTumorSamples.GSM2359914_ctrl14     5.513552      3.621765
## LiverTumorSamples.GSM2359912_IL1       6.500059      3.653289
## LiverTumorSamples.GSM2359917_IL2       6.6111744     3.7909157
```

Write this ML ready file to csv.

```
write.csv(Combo2, 'ML_ready_CoV_14_classes.csv', row.names=TRUE)
```

Make a separate ML ready file with a smaller set of classes to classify by liver or capillary and control or CoronaVirus

```
a <- rep('liver', 4)
b <- rep('liver', 4)
c <- rep('liver', 2)
d <- rep('liver', 2)
e <- rep('capillary', 5)
f <- rep('capillary', 5)
g <- rep('capillary', 5)
h <- rep('capillary', 5)
i <- rep('capillary', 5)
j <- rep('capillary', 5)
k <- rep('capillary', 5)
l <- rep('capillary', 5)
m <- rep('capillary', 5)
n <- rep('capillary', 5)

type <- as.data.frame(c(a,b,c,d,e,f,g,h,i,j,k,l,m,n))
colnames(type) <- 'Class_Type'
row.names(type) <- row.names(Combo1)

Combo3 <- cbind(type, Combo1)

Combo3[1:10, 1:5]

##                               Class_Type      NEURL3      DUSP1
## LiverTumorSamples.GSM2359851_CoV1      liver      1429.61750      8491.40875
```



```
## LiverTumorSamples.GSM2359853_CoV2      liver      190.21750      2219.85650
## LiverTumorSamples.GSM2359910_CoV3      liver      10.004148      11.494585
## LiverTumorSamples.GSM2359913_CoV4      liver      11.245589      12.898250
## LiverTumorSamples.GSM2359850_ctrl11    liver      34.57000      228.18775
## LiverTumorSamples.GSM2359852_ctrl12    liver      17.25750      216.08550
## LiverTumorSamples.GSM2359911_ctrl13    liver      3.708157      7.184185
## LiverTumorSamples.GSM2359914_ctrl14    liver      4.757780      7.113854
## LiverTumorSamples.GSM2359912_IL1      liver      4.879242      9.576161
## LiverTumorSamples.GSM2359917_IL2      liver      5.1138565      9.5527540
##                                     ATF3          PCLO
## LiverTumorSamples.GSM2359851_CoV1      3608.28250      17.74792
## LiverTumorSamples.GSM2359853_CoV2      974.76937      613.28583
## LiverTumorSamples.GSM2359910_CoV3      8.337322      3.355859
## LiverTumorSamples.GSM2359913_CoV4      9.441972      2.741117
## LiverTumorSamples.GSM2359850_ctrl11    108.52250      13.80667
## LiverTumorSamples.GSM2359852_ctrl12    97.05125      12.88750
## LiverTumorSamples.GSM2359911_ctrl13    5.373051      3.904719
## LiverTumorSamples.GSM2359914_ctrl14    5.513552      3.621765
## LiverTumorSamples.GSM2359912_IL1      6.500059      3.653289
## LiverTumorSamples.GSM2359917_IL2      6.6111744      3.7909157
```

```
write.csv(Combo3, 'ML_ready_CoV_2_classes.csv', row.names=TRUE)
```

```
a <- rep('CoV', 4)
b <- rep('Ctrl', 4)
c <- rep('CoV_IL', 2)
d <- rep('IA_CoV', 2)
e <- rep('CoV', 5)
f <- rep('Ctrl', 5)
g <- rep('Cov', 5)
h <- rep('Ctrl', 5)
i <- rep('Cov', 5)
j <- rep('Ctrl', 5)
k <- rep('Cov', 5)
l <- rep('Ctrl', 5)
m <- rep('Cov', 5)
n <- rep('Ctrl', 5)
```

```
type <- as.data.frame(c(a,b,c,d,e,f,g,h,i,j,k,l,m,n))
colnames(type) <- 'Class_Type'
row.names(type) <- row.names(Combo1)
```

```
Combo4 <- cbind(type, Combo1)
```

```
Combo4[1:10, 1:5]
```

```
##                                     Class_Type      NEURL3      DUSP1
## LiverTumorSamples.GSM2359851_CoV1      CoV      1429.61750      8491.40875
## LiverTumorSamples.GSM2359853_CoV2      CoV      190.21750      2219.85650
## LiverTumorSamples.GSM2359910_CoV3      CoV      10.004148      11.494585
```



```
## LiverTumorSamples.GSM2359913_CoV4      CoV  11.245589  12.898250
## LiverTumorSamples.GSM2359850_ctrl1      Ctrl  34.57000  228.18775
## LiverTumorSamples.GSM2359852_ctrl12     Ctrl  17.25750  216.08550
## LiverTumorSamples.GSM2359911_ctrl13     Ctrl  3.708157   7.184185
## LiverTumorSamples.GSM2359914_ctrl14     Ctrl  4.757780   7.113854
## LiverTumorSamples.GSM2359912_IL1       CoV_IL  4.879242   9.576161
## LiverTumorSamples.GSM2359917_IL2       CoV_IL  5.1138565  9.5527540
##                                     ATF3      PCLO
## LiverTumorSamples.GSM2359851_CoV1  3608.28250  17.74792
## LiverTumorSamples.GSM2359853_CoV2   974.76937  613.28583
## LiverTumorSamples.GSM2359910_CoV3    8.337322   3.355859
## LiverTumorSamples.GSM2359913_CoV4    9.441972   2.741117
## LiverTumorSamples.GSM2359850_ctrl1  108.52250  13.80667
## LiverTumorSamples.GSM2359852_ctrl12  97.05125  12.88750
## LiverTumorSamples.GSM2359911_ctrl13  5.373051   3.904719
## LiverTumorSamples.GSM2359914_ctrl14  5.513552   3.621765
## LiverTumorSamples.GSM2359912_IL1    6.500059   3.653289
## LiverTumorSamples.GSM2359917_IL2    6.6111744  3.7909157
```

```
write.csv(Combo4, 'ML_ready_CoV_4_classes.csv', row.names=TRUE)
```

We didn't do any fold change or stat measures on the capillary samples, but we can plot them by using ggplot2 and group the sets by timed intervals for each group A through E and picking a handful of genes to compare over the 0,12,24,36, and 48 hour time intervals for the control group and the Coronavirus inoculated groups.

When the values are a ratio like this, it is easier to see the larger changes as in 9 compared to a low change like 0.0005, but this just means that compared to the control samples the inoculated Coronavirus had 9 times the gene expression values or had downregulated or suppressed gene expression values to 1/5000th the amount of the normal range of gene expression values respectively. ***

It makes sense to use some genes we already know have a higher magnitude of change, and we have a column for that in the CombinedLiver table called MagnitudeFCs that was already sorted from largest to smallest when made. We'll just select the first five of those genes to compare in these capillary samples over time.

```
mostChanged <- CombinedLiver[1:5,c(1,71)]
mostSuppressed <- CombinedLiver[196:200,c(1,71)]
row.names(mostChanged)

## [1] "NEURL3" "DUSP1" "ATF3" "PCLO" "LHB"

row.names(mostSuppressed)
```

```

## [1] "RASSF7"          "LOC100335030" "C2orf78"          "DEFB1"           "ZNF610"

capillary <- merge(mostChanged, CombinedLiver, by.x='GENE_SYMBOL',
by.y='GENE_SYMBOL')
capillary1 <- merge(mostSuppressed, CombinedLiver, by.x='GENE_SYMBOL',
by.y='GENE_SYMBOL')
capillaries <- rbind(capillary, capillary1)
Capillaries <- capillaries[,c(1,15:64)]
row.names(Capillaries) <- Capillaries$GENE_SYMBOL

Capillaries2 <- as.data.frame(t(Capillaries))
Capillaries2 <- Capillaries2[-1,]
row.names(Capillaries2) <-
gsub('capillarySamples.', '', row.names(Capillaries2))
row.names(Capillaries2) <- gsub('GSM[0-9][0-9][0-9][0-9][0-9][0-9][0-9]_', '', row.names(Capillaries2))
row.names(Capillaries2) <- gsub('MERS_', '', row.names(Capillaries2))

CoV <- grep('CoV', row.names(Capillaries2))
ctrl <- grep('ctrl', row.names(Capillaries2))

Capillaries2$Class <- 'CoV or ctrl'

Capillaries2[CoV,11] <- 'Coronavirus'
Capillaries2[ctrl,11] <- 'control'

A <- grep('_A', row.names(Capillaries2))
B <- grep('_B', row.names(Capillaries2))
C <- grep('_C', row.names(Capillaries2))
D <- grep('_D', row.names(Capillaries2))
E <- grep('_E', row.names(Capillaries2))

Capillaries2$Group <- 'group'

Capillaries2[A,12] <- 'A'
Capillaries2[B,12] <- 'B'
Capillaries2[C,12] <- 'C'
Capillaries2[D,12] <- 'D'
Capillaries2[E,12] <- 'E'

hr0 <- grep('0hr', row.names(Capillaries2))
hr12 <- grep('12hr', row.names(Capillaries2))
hr24 <- grep('24hr', row.names(Capillaries2))
hr36 <- grep('36hr', row.names(Capillaries2))
hr48 <- grep('48hr', row.names(Capillaries2))

Capillaries2$TimeInterval <- 'time'

Capillaries2[hr0,13] <- '0 hr'

```

```

Capillaries2[hr12,13] <- '12 hr'
Capillaries2[hr24,13] <- '24 hr'
Capillaries2[hr36,13] <- '36 hr'
Capillaries2[hr48,13] <- '48 hr'

```

Capillaries2

##		ATF3	DUSP1	LHB	NEURL3	PCLO	C2orf78
##	CoV_0hr_A	10.290996	13.901505	9.093671	7.498965	6.468670	5.936451
##	CoV_0hr_B	9.780412	13.852933	9.128886	7.446590	6.476326	6.023280
##	CoV_0hr_C	9.574148	13.740905	8.662565	7.495092	6.463702	5.909934
##	CoV_0hr_D	9.848204	13.863910	8.729845	7.364854	6.572579	5.986079
##	CoV_0hr_E	10.114265	13.968982	8.738079	7.436675	6.625125	5.862801
##	ctrl_0hr_A	10.173291	14.398733	8.776979	7.267773	6.410249	6.016610
##	ctrl_0hr_B	10.132629	14.237192	9.610252	7.015544	6.467982	5.931786
##	ctrl_0hr_C	10.308477	14.382699	9.557514	7.440873	6.478339	5.853411
##	ctrl_0hr_D	9.888505	14.404923	9.602914	7.102155	6.489396	5.855096
##	ctrl_0hr_E	9.892730	14.120833	9.577233	6.936543	6.417981	5.942510
##	CoV_12hr_A	9.819186	11.303627	10.795022	12.768140	6.256646	5.865277
##	CoV_12hr_B	10.011539	11.577456	11.006393	12.402078	6.327304	5.922540
##	CoV_12hr_C	9.783105	11.626722	11.000187	12.532051	6.322820	5.988353
##	CoV_12hr_D	9.849858	11.692725	10.938022	12.668202	6.261566	5.901327
##	CoV_12hr_E	9.617070	12.220997	10.553798	12.273302	6.333485	5.898912
##	ctrl_12hr_A	7.825226	11.488735	9.286872	6.829080	6.528833	5.954890
##	ctrl_12hr_B	7.872041	11.710908	8.845744	6.784278	6.532492	5.862801
##	ctrl_12hr_C	7.989155	11.468072	9.267328	6.908973	6.489589	5.973506
##	ctrl_12hr_D	8.028735	11.407001	9.820556	6.615987	6.412667	5.928711
##	ctrl_12hr_E	8.154875	11.042038	9.741747	6.833504	6.431864	5.862305
##	CoV_24hr_A	12.356029	14.922819	10.081784	13.713997	6.330955	5.981748
##	CoV_24hr_B	12.393863	14.762759	10.125434	13.516154	6.376025	5.938335
##	CoV_24hr_C	12.347139	15.286795	9.988253	13.718699	6.398037	6.047165
##	CoV_24hr_D	12.280955	15.019195	9.580647	13.576554	6.441270	6.099102
##	CoV_24hr_E	12.187980	14.989084	9.508540	13.545156	6.428447	6.044217
##	ctrl_24hr_A	7.914153	12.020454	8.926344	6.410178	6.709108	6.033401
##	ctrl_24hr_B	8.179651	11.901618	9.159004	6.560892	6.642224	5.909934
##	ctrl_24hr_C	8.062523	11.831523	9.275883	6.568645	6.626875	5.938570
##	ctrl_24hr_D	8.045721	11.806023	9.273915	6.631753	6.618130	5.997365
##	ctrl_24hr_E	8.204591	11.942950	9.125971	6.562419	6.697083	5.919444
##	CoV_36hr_A	12.318180	14.508819	9.861903	13.561144	6.548994	6.087523
##	CoV_36hr_B	12.173687	14.540567	9.802688	13.549071	6.644295	5.999396
##	CoV_36hr_C	12.163530	14.579001	9.872176	13.449660	6.515321	5.933150
##	CoV_36hr_D	12.142847	14.514722	9.982218	13.712803	6.671993	6.001650
##	CoV_36hr_E	12.297080	14.374295	10.201176	13.521456	6.372439	6.051522
##	ctrl_36hr_A	7.940682	11.742312	9.321322	6.730998	6.678982	6.220477
##	ctrl_36hr_B	7.783378	11.751105	9.056790	6.621403	6.643658	6.125131
##	ctrl_36hr_C	7.810544	11.798201	9.151894	6.822530	6.618794	6.092605
##	ctrl_36hr_D	7.731982	11.876269	9.024215	6.685524	6.629717	6.098892
##	ctrl_36hr_E	7.734450	11.888111	8.922796	6.630443	6.673160	6.028326
##	CoV_48hr_A	11.205349	13.523040	9.679922	12.333937	6.529574	6.049019
##	CoV_48hr_B	11.253111	13.511350	9.912659	12.481150	6.549794	5.980606

##	CoV_48hr_C	10.944238	13.316529	9.850259	12.260608	6.708351	6.203350	
##	CoV_48hr_D	10.983900	13.358680	9.899124	11.961065	6.599227	6.142515	
##	CoV_48hr_E	11.255318	13.367467	10.184716	12.498831	6.679909	6.017946	
##	ctrl_48hr_A	7.932607	12.174103	9.024090	6.757483	6.552106	6.016387	
##	ctrl_48hr_B	7.668012	12.449589	8.985262	6.882334	6.700790	5.961156	
##	ctrl_48hr_C	7.755264	12.365436	9.031925	6.844960	6.455577	6.003225	
##	ctrl_48hr_D	7.725891	12.415529	8.940948	6.872043	6.533961	5.986534	
##	ctrl_48hr_E	7.656200	12.455840	8.994880	6.868035	6.479109	6.091125	
##		DEFB1	LOC100335030	RASSF7	ZNF610			Class Group
##	CoV_0hr_A	7.622403	8.114815	6.822339	7.227249	Coronavirus		A
##	CoV_0hr_B	7.778352	8.270884	6.700971	7.428115	Coronavirus		B
##	CoV_0hr_C	7.352772	8.308822	6.966246	7.481717	Coronavirus		C
##	CoV_0hr_D	7.118591	8.246936	6.875601	7.426565	Coronavirus		D
##	CoV_0hr_E	7.179596	8.214737	6.952108	7.420473	Coronavirus		E
##	ctrl_0hr_A	7.331124	8.187333	6.964110	7.403619	control		A
##	ctrl_0hr_B	7.872561	8.288294	6.910892	7.464914	control		B
##	ctrl_0hr_C	7.584274	8.147007	6.994899	7.269174	control		C
##	ctrl_0hr_D	7.650353	8.185575	7.033153	7.359418	control		D
##	ctrl_0hr_E	7.847725	8.156999	7.056144	7.447995	control		E
##	CoV_12hr_A	8.059324	7.868801	6.594891	8.147592	Coronavirus		A
##	CoV_12hr_B	7.843962	7.735502	6.692292	8.178197	Coronavirus		B
##	CoV_12hr_C	7.985707	7.820541	6.586602	8.121462	Coronavirus		C
##	CoV_12hr_D	7.995350	7.717794	6.571151	8.224417	Coronavirus		D
##	CoV_12hr_E	7.418830	7.832236	6.801301	8.247149	Coronavirus		E
##	ctrl_12hr_A	8.009080	8.069778	6.880927	7.524340	control		A
##	ctrl_12hr_B	8.055858	7.986788	6.895230	7.537055	control		B
##	ctrl_12hr_C	7.952662	7.941232	6.903801	7.575149	control		C
##	ctrl_12hr_D	8.277781	8.181431	6.745836	7.641494	control		D
##	ctrl_12hr_E	8.235038	8.169354	6.554766	7.782781	control		E
##	CoV_24hr_A	7.571666	8.536174	6.747448	8.938377	Coronavirus		A
##	CoV_24hr_B	7.629778	8.509799	6.697222	8.733307	Coronavirus		B
##	CoV_24hr_C	7.260831	8.425157	6.692222	9.019246	Coronavirus		C
##	CoV_24hr_D	7.080690	8.584440	6.757083	9.012278	Coronavirus		D
##	CoV_24hr_E	7.058040	8.736338	6.698612	9.092390	Coronavirus		E
##	ctrl_24hr_A	7.563681	8.155811	6.789623	7.914991	control		A
##	ctrl_24hr_B	7.729397	8.138252	6.704019	7.747434	control		B
##	ctrl_24hr_C	7.946592	8.095640	6.643618	7.788730	control		C
##	ctrl_24hr_D	7.625182	8.078397	6.716886	7.876311	control		D
##	ctrl_24hr_E	7.486914	8.005603	6.680610	7.754600	control		E
##	CoV_36hr_A	8.078504	8.482066	6.721733	8.534288	Coronavirus		A
##	CoV_36hr_B	7.874253	8.583444	6.770072	8.659518	Coronavirus		B
##	CoV_36hr_C	7.801655	8.552423	6.862404	8.572155	Coronavirus		C
##	CoV_36hr_D	7.980233	8.606779	6.846841	8.654873	Coronavirus		D
##	CoV_36hr_E	7.610937	8.589262	6.822339	8.589262	Coronavirus		E
##	ctrl_36hr_A	7.737160	8.018402	6.849095	7.846848	control		A
##	ctrl_36hr_B	7.704646	8.042240	6.771987	7.839025	control		B
##	ctrl_36hr_C	7.652111	8.124825	6.842196	7.974192	control		C
##	ctrl_36hr_D	7.564977	8.045657	6.925022	7.847725	control		D
##	ctrl_36hr_E	7.592967	8.041310	6.908012	7.860321	control		E
##	CoV_48hr_A	7.328972	8.062943	6.505043	7.678934	Coronavirus		A

## CoV_48hr_B	7.354622	8.098275	6.544158	7.742997	Coronavirus	B
## CoV_48hr_C	7.478970	8.024786	6.488793	7.725653	Coronavirus	C
## CoV_48hr_D	7.584650	7.979318	6.643906	7.822518	Coronavirus	D
## CoV_48hr_E	7.345740	8.128902	6.493502	9.374216	Coronavirus	E
## ctrl_48hr_A	7.551584	8.234894	6.812821	7.882805	control	A
## ctrl_48hr_B	7.634577	8.221036	6.918835	8.076742	control	B
## ctrl_48hr_C	7.580736	8.234032	6.851086	8.033925	control	C
## ctrl_48hr_D	7.753263	8.230602	6.913527	7.983884	control	D
## ctrl_48hr_E	7.731435	8.270651	6.878047	7.998738	control	E
##	TimeInterval					
## CoV_0hr_A	0 hr					
## CoV_0hr_B	0 hr					
## CoV_0hr_C	0 hr					
## CoV_0hr_D	0 hr					
## CoV_0hr_E	0 hr					
## ctrl_0hr_A	0 hr					
## ctrl_0hr_B	0 hr					
## ctrl_0hr_C	0 hr					
## ctrl_0hr_D	0 hr					
## ctrl_0hr_E	0 hr					
## CoV_12hr_A	12 hr					
## CoV_12hr_B	12 hr					
## CoV_12hr_C	12 hr					
## CoV_12hr_D	12 hr					
## CoV_12hr_E	12 hr					
## ctrl_12hr_A	12 hr					
## ctrl_12hr_B	12 hr					
## ctrl_12hr_C	12 hr					
## ctrl_12hr_D	12 hr					
## ctrl_12hr_E	12 hr					
## CoV_24hr_A	24 hr					
## CoV_24hr_B	24 hr					
## CoV_24hr_C	24 hr					
## CoV_24hr_D	24 hr					
## CoV_24hr_E	24 hr					
## ctrl_24hr_A	24 hr					
## ctrl_24hr_B	24 hr					
## ctrl_24hr_C	24 hr					
## ctrl_24hr_D	24 hr					
## ctrl_24hr_E	24 hr					
## CoV_36hr_A	36 hr					
## CoV_36hr_B	36 hr					
## CoV_36hr_C	36 hr					
## CoV_36hr_D	36 hr					
## CoV_36hr_E	36 hr					
## ctrl_36hr_A	36 hr					
## ctrl_36hr_B	36 hr					
## ctrl_36hr_C	36 hr					
## ctrl_36hr_D	36 hr					
## ctrl_36hr_E	36 hr					

```
## CoV_48hr_A      48 hr
## CoV_48hr_B      48 hr
## CoV_48hr_C      48 hr
## CoV_48hr_D      48 hr
## CoV_48hr_E      48 hr
## ctrl_48hr_A      48 hr
## ctrl_48hr_B      48 hr
## ctrl_48hr_C      48 hr
## ctrl_48hr_D      48 hr
## ctrl_48hr_E      48 hr

write.csv(Capillaries2, 'FC_10_capillaries_CoV.csv', row.names=TRUE)
```

The above table has 10 genes as the columns with the added Class (Coronavirus or control), Group (A,B,C,D,E), and TimeInterval (0,12,24,36,48 hours) fields to filter by and plot.

Lets make these group tables for the corona virus and see how they compare over time.

```
library(dplyr)

A_group <- filter(Capillaries2, Group=='A' & Class == 'Coronavirus')
B_group <- filter(Capillaries2, Group=='B' & Class == 'Coronavirus')
C_group <- filter(Capillaries2, Group=='C' & Class == 'Coronavirus')
D_group <- filter(Capillaries2, Group=='D' & Class == 'Coronavirus')
E_group <- filter(Capillaries2, Group=='E' & Class == 'Coronavirus')
```

Lets use the tidyr package to put the 10 genes into one Gene field.

```
library(tidyr)
```

We will do this for the A_group table and ignore the Group and Class fields, because we made it only the A group of the Coronavirus class.

```
A_group2 <- A_group[,c(1,3,5,7,9,11:13)]
A_tidy <- gather(A_group2, 'Gene', 'GeneExpression', 1:5)

## Warning: attributes are not identical across measure variables;
## they will be dropped

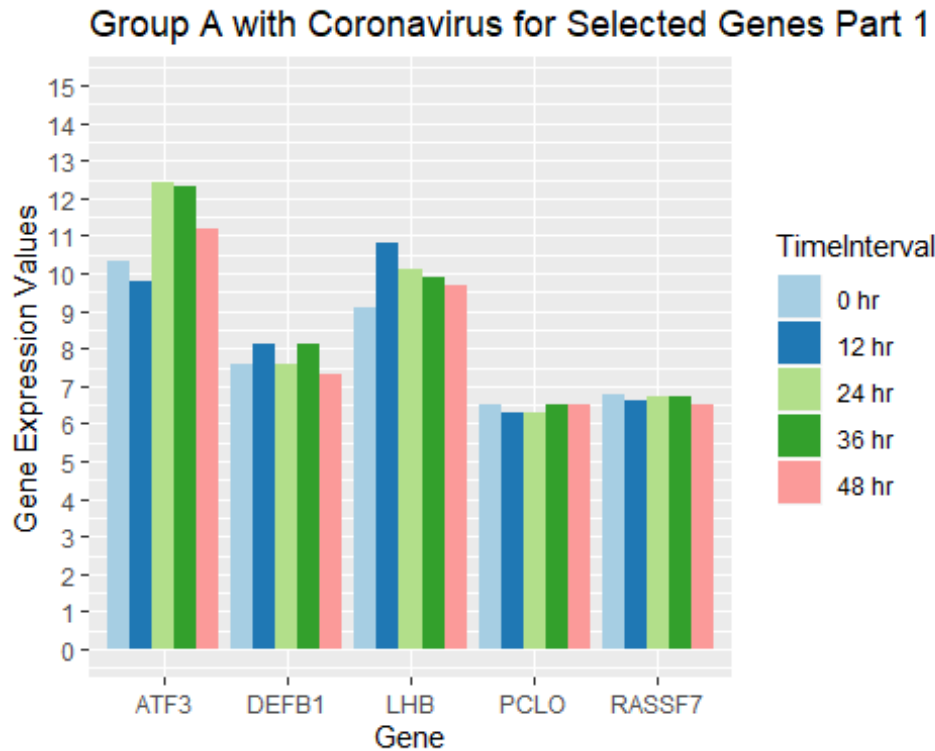
A_tidy$GeneExpression <- round(as.numeric(A_tidy$GeneExpression),1)
A_tidy$TimeInterval <- as.factor(A_tidy$TimeInterval)
A_tidy$Gene <- as.factor(A_tidy$Gene)
A_tidy
```

##	Class	Group	TimeInterval	Gene	GeneExpression
## 1	Coronavirus	A	0 hr	ATF3	10.3
## 2	Coronavirus	A	12 hr	ATF3	9.8
## 3	Coronavirus	A	24 hr	ATF3	12.4
## 4	Coronavirus	A	36 hr	ATF3	12.3
## 5	Coronavirus	A	48 hr	ATF3	11.2

## 6	Coronavirus	A	0 hr	LHB	9.1
## 7	Coronavirus	A	12 hr	LHB	10.8
## 8	Coronavirus	A	24 hr	LHB	10.1
## 9	Coronavirus	A	36 hr	LHB	9.9
## 10	Coronavirus	A	48 hr	LHB	9.7
## 11	Coronavirus	A	0 hr	PCLO	6.5
## 12	Coronavirus	A	12 hr	PCLO	6.3
## 13	Coronavirus	A	24 hr	PCLO	6.3
## 14	Coronavirus	A	36 hr	PCLO	6.5
## 15	Coronavirus	A	48 hr	PCLO	6.5
## 16	Coronavirus	A	0 hr	DEFB1	7.6
## 17	Coronavirus	A	12 hr	DEFB1	8.1
## 18	Coronavirus	A	24 hr	DEFB1	7.6
## 19	Coronavirus	A	36 hr	DEFB1	8.1
## 20	Coronavirus	A	48 hr	DEFB1	7.3
## 21	Coronavirus	A	0 hr	RASSF7	6.8
## 22	Coronavirus	A	12 hr	RASSF7	6.6
## 23	Coronavirus	A	24 hr	RASSF7	6.7
## 24	Coronavirus	A	36 hr	RASSF7	6.7
## 25	Coronavirus	A	48 hr	RASSF7	6.5

```
library(ggplot2)
```

```
ggplot(data = A_tidy, aes(x=Gene, y=GeneExpression, fill=TimeInterval)) +
  geom_bar(stat='identity', position=position_dodge()) +
  scale_y_continuous(breaks = seq(0, 15, by=1), limits=c(0,15)) +
  scale_fill_brewer(palette='Paired') +
  ggtitle('Group A with Coronavirus for Selected Genes Part 1') +
  xlab('Gene') +
  ylab('Gene Expression Values')
```

The genes above for Part 1 of the group A samples of coronavirus in blood capillaries show some variation in gene expression values for some of these genes that had the most change in the liver tumor samples. Starting at the initial hour up to 48 hours after being inoculated in vitro, there is an increase then decrease for ATF3 and LHB genes, while a decrease then increase close to initial value with PCLO and slightly with RASSF7. For DEFB1, it has a cyclical increase, decrease, increase, then decrease to stabilize closer to the initial gene expression value.

Now let's find the other five genes in the group A set of ten genes found to have the most change in the liver tumor samples, and examined here in the blood capillary samples.

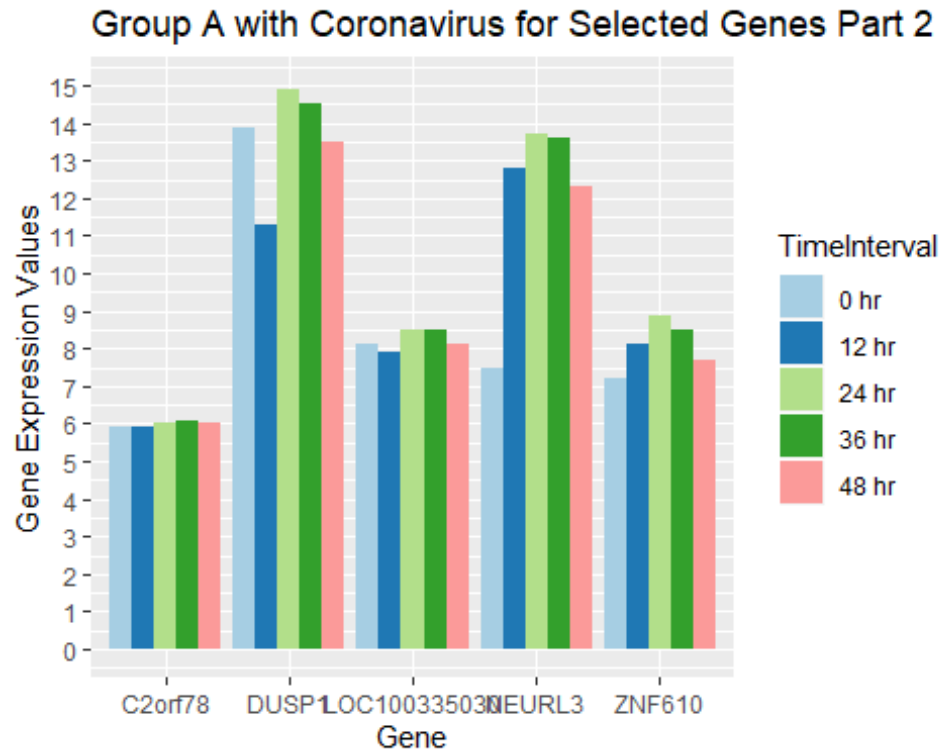
```
A_group3 <- A_group[,c(2,4,6,8,10,11:13)]
A_tidy1 <- gather(A_group3, 'Gene', 'GeneExpression', 1:5)

## Warning: attributes are not identical across measure variables;
## they will be dropped

A_tidy1$GeneExpression <- round(as.numeric(A_tidy1$GeneExpression), 1)
A_tidy1$TimeInterval <- as.factor(A_tidy1$TimeInterval)
A_tidy1$Gene <- as.factor(A_tidy1$Gene)
A_tidy1
```

##	Class	Group	TimeInterval	Gene	GeneExpression
## 1	Coronavirus	A	0 hr	DUSP1	13.9
## 2	Coronavirus	A	12 hr	DUSP1	11.3
## 3	Coronavirus	A	24 hr	DUSP1	14.9
## 4	Coronavirus	A	36 hr	DUSP1	14.5
## 5	Coronavirus	A	48 hr	DUSP1	13.5
## 6	Coronavirus	A	0 hr	NEURL3	7.5
## 7	Coronavirus	A	12 hr	NEURL3	12.8
## 8	Coronavirus	A	24 hr	NEURL3	13.7
## 9	Coronavirus	A	36 hr	NEURL3	13.6
## 10	Coronavirus	A	48 hr	NEURL3	12.3
## 11	Coronavirus	A	0 hr	C2orf78	5.9
## 12	Coronavirus	A	12 hr	C2orf78	5.9
## 13	Coronavirus	A	24 hr	C2orf78	6.0
## 14	Coronavirus	A	36 hr	C2orf78	6.1
## 15	Coronavirus	A	48 hr	C2orf78	6.0
## 16	Coronavirus	A	0 hr	LOC100335030	8.1
## 17	Coronavirus	A	12 hr	LOC100335030	7.9
## 18	Coronavirus	A	24 hr	LOC100335030	8.5
## 19	Coronavirus	A	36 hr	LOC100335030	8.5
## 20	Coronavirus	A	48 hr	LOC100335030	8.1
## 21	Coronavirus	A	0 hr	ZNF610	7.2
## 22	Coronavirus	A	12 hr	ZNF610	8.1
## 23	Coronavirus	A	24 hr	ZNF610	8.9
## 24	Coronavirus	A	36 hr	ZNF610	8.5
## 25	Coronavirus	A	48 hr	ZNF610	7.7

```
ggplot(data = A_tidy1, aes(x=Gene, y=GeneExpression, fill=TimeInterval)) +
  geom_bar(stat='identity', position=position_dodge()) +
  scale_y_continuous(breaks = seq(0, 15, by=1), limits=c(0,15)) +
  scale_fill_brewer(palette='Paired') +
  ggtitle('Group A with Coronavirus for Selected Genes Part 2') +
  xlab('Gene') +
  ylab('Gene Expression Values')
```



The above genes in part 2 of the group A Coronavirus samples over 48 hours, shows that the gene expression values increase up to 24 hours then decrease to 48 hours for most of the genes above.
