

## CoV liver and 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](https://ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89166)
- [ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89160](https://ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89160)
- [ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100509](https://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

```

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')
CombinedLiver <- combined1[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

```

Write this ML ready file to csv.

```
write.csv(Combo1, 'ML_ready_CoV.csv', row.names=TRUE)
```