

Flu Vaccines blood array with or without antibiotic pre-treatment

Janis Corona

1/6/2020

This markdown file is on a study from GEO that used blood samples from day 0,1,3,7, and 21 days after treating patients with antibiotics before administering the flu vaccine in 2018. There were two phases to this study and many samples. However, only three patients who were treated with antibiotics and three patients who were not treated with antibiotics were used to analyze the gene expression profiles of the flu vaccine in these blood samples. Then compare the gene expression profiles from the EGCG, hemochromatosis, and myocardial infarction samples. These particular samples don't give the age or gender, but the patients were mixed of female and gender in the age range of 18-45 years old.

The study can be linked to by visiting: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120717> and <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120719>

```
flu <- read.csv('fluSamples.csv', sep=',', header=TRUE, na.strings=c('', ' '))
platform <- read.delim('GPL13158-5065.txt', sep='\t', header=TRUE,
                      comment.char='#', na.strings=c('', ' '))
```

```
PL <- platform[,c('ID', 'Gene.Symbol')]
```

```
Flu <- merge(PL, flu, by.x='ID', by.y='ID_REF')
Flu <- Flu[, -1]
```

```
Flu <- Flu[complete.cases(Flu$Gene.Symbol),]

names0 <- strsplit(as.character(Flu$Gene.Symbol), '//')
names1 <- lapply(names0, '[', 1)
names2 <- as.data.frame(as.character(names1))
colnames(names2) <- 'Gene'

Flu2 <- cbind(names2, Flu)
Flu2 <- Flu2[, -2]
```

```
library(dplyr)

samples <- colnames(Flu2)[2:31]
Samples <- as.vector(samples)

uniqueGenes <- Flu2 %>% group_by(Gene) %>% summarise_at(Samples, mean,
                                                         na.rm=TRUE)
```

```
uniqueGenes <- uniqueGenes[complete.cases(uniqueGenes$Gene),]
write.csv(uniqueGenes, 'uniqueGenes.csv', row.names=FALSE)
```

The flu samples treated with antibiotics and those not treated with antibiotics

```
treated <- uniqueGenes[,c(1:16)]
notTreated <- uniqueGenes[,c(1,17:31)]
```

Separate into days 0, 1, 3, 7, and 21 days after flu vaccine given.

```
treated_0 <- treated[,c(1,2,7,12)]
treated_1 <- treated[,c(1,3,8,13)]
treated_3 <- treated[,c(1,4,9,14)]
treated_7 <- treated[,c(1,5,10,15)]
treated_21 <- treated[,c(1,6,11,16)]

notTreated_0 <- notTreated[,c(1,2,7,12)]
notTreated_1 <- notTreated[,c(1,3,8,13)]
notTreated_3 <- notTreated[,c(1,4,9,14)]
notTreated_7 <- notTreated[,c(1,5,10,15)]
notTreated_21 <- notTreated[,c(1,6,11,16)]
```

The row means of each gene per day of treated and not treated.

```
treated_0$T0_Mean <- rowMeans(treated_0[2:4])
treated_1$T1_Mean <- rowMeans(treated_1[2:4])
treated_3$T3_Mean <- rowMeans(treated_3[2:4])
treated_7$T7_Mean <- rowMeans(treated_7[2:4])
treated_21$T21_Mean <- rowMeans(treated_21[2:4])

notTreated_0$NT0_Mean <- rowMeans(notTreated_0[2:4])
notTreated_1$NT1_Mean <- rowMeans(notTreated_1[2:4])
notTreated_3$NT3_Mean <- rowMeans(notTreated_3[2:4])
notTreated_7$NT7_Mean <- rowMeans(notTreated_7[2:4])
notTreated_21$NT21_Mean <- rowMeans(notTreated_21[2:4])
```

Fold change within the treated group of flu vaccinated with antibiotics.

```
treatedMeans <- cbind(treated_0[5], treated_1[5], treated_3[5],
                      treated_7[5], treated_21[5])
FC1 <- treatedMeans %>% mutate(FC_t1=T1_Mean/T0_Mean) # fold change day1 from day0
FC3 <- FC1 %>% mutate(FC_t3=T3_Mean/T1_Mean) # fold change day3 from day 1
FC7 <- FC3 %>% mutate(FC_t7=T7_Mean/T3_Mean) # fold change day7 from day3
FC21 <- FC7 %>% mutate(FC_t21=T21_Mean/T7_Mean) # fold change day21 from day7
```

Fold change within the non-treated group of flu vaccinated with antibiotics.

```
nonTreatedMeans <- cbind(notTreated_0[5], notTreated_1[5], notTreated_3[5],
                          notTreated_7[5], notTreated_21[5])

# day1 from day0 non-treated
nFC1 <- nonTreatedMeans %>% mutate(FC_nt1 = NT1_Mean/NT0_Mean)

# day3 from day1 non-treated
nFC3 <- nFC1 %>% mutate(FC_nt3 = NT3_Mean/NT1_Mean)
```

```
# day7 from day3 non-treated
nFC7 <- nFC3 %>% mutate(FC_nt7 = NT7_Mean/NT3_Mean)

# day21 from day7 non-treated
nFC21 <- nFC7 %>% mutate(FC_nt21 = NT21_Mean/NT7_Mean)
```

Fold change between the groups of treated and non-treated flu vaccinated with antibiotics using the initial treated mean for all ratios as the common denominator.

```
BothMeans <- cbind(FC21, nFC21)

FC_both1 <- BothMeans %>% mutate(FCB_1 = NT1_Mean/T0_Mean)
FC_both3 <- FC_both1 %>% mutate(FCB_3 = NT3_Mean/T0_Mean)
FC_both7 <- FC_both3 %>% mutate(FCB_7 = NT7_Mean/T0_Mean)
FC_both21 <- FC_both7 %>% mutate(FCB_21 = NT21_Mean/T0_Mean)

MeansBothAll <- FC_both21[,c(6:9,15:22,1:5,10:14)]
row.names(MeanBothAll) <- treated$Gene
```

Combine all the samples with the fold change within the treated, within the non-treated, between both from day 0, and the means of each group by day.

```
uniqueGenes <- as.data.frame(uniqueGenes)
row.names(uniqueGenes) <- uniqueGenes$Gene
```

```
MeansBothAll$Gene <- row.names(MeanBothAll)
All_FoldChange_Values <- merge(MeanBothAll,uniqueGenes, by.x='Gene',
                               by.y='Gene')
```

```
write.csv(All_FoldChange_Values, 'FoldChange_All.csv', row.names=FALSE)
```

```
colnames(All_FoldChange_Values)
```

```
## [1] "Gene"
## [3] "FC_t3"
## [5] "FC_t21"
## [7] "FC_nt3"
## [9] "FC_nt21"
## [11] "FCB_3"
## [13] "FCB_21"
## [15] "T1_Mean"
## [17] "T7_Mean"
## [19] "NT0_Mean"
## [21] "NT3_Mean"
## [23] "NT21_Mean"
## [25] "GSM3409107_29_day_1"
## [27] "GSM3409004_29_day_7"
## [29] "GSM3409006_30_day_0"
## [31] "GSM3409008_30_day_3"
## [33] "GSM3409005_30_day_21_screening"
## [35] "GSM3409014_05_day_1"

"FC_t1"
"FC_t7"
"FC_nt1"
"FC_nt7"
"FCB_1"
"FCB_7"
"T0_Mean"
"T3_Mean"
"T21_Mean"
"NT1_Mean"
"NT7_Mean"
"GSM3409106_29_day_0"
"GSM3409108_29_day_3"
"GSM3409105_29_day_21_screening"
"GSM3409007_30_day_1"
"GSM3409009_30_day_7"
"GSM3409013_05_day_0"
"GSM3409015_05_day_3"
```

```
## [37] "GSM3409016_05_day_7"          "GSM3409012_05_day_21_screening"
## [39] "GSM3409161_33_day_0_no"       "GSM3409162_33_day_1_no"
## [41] "GSM3409163_33_day_3_no"       "GSM3409111_33._day_7_no"
## [43] "GSM3409160_33_day_21_screening_no" "GSM3409124_36_day_0_no"
## [45] "GSM3409125_36_day_1_no"       "GSM3409126_36_day_3_no"
## [47] "GSM3409127_36_day_7_no"       "GSM3409123_36_day_21_screening_no"
## [49] "GSM3409135_38_day_0_no"       "GSM3409136_38_day_1_no"
## [51] "GSM3409137_38_day_3_no"       "GSM3409138_38_day_7_no"
## [53] "GSM3409134_38_day_21_screening_no"
```

The 'FC_t1' through 'FC_t21' are the fold change of the treated days from the highest day ratio to the previous day ratio, ie: FC_t1 is day1/day0 fold change for treated and FC_t21 is day21/day7 treated flu with antibiotics. The 'FC_nt1' through 'FC_nt21' are the same ratios but for the non-treated samples, ie: FC_nt21 is day21/day7 non-treated flu with antibiotics gene expression fold change values. The 'FCB-1' through 'FCB_21' are the fold change values from the day to the initial day of treatment ratios, ie: FCB_21 is the ratio of day21 non-treated to day0 treated. The 'T0_Mean' through 'T21_Mean' are the treated means per gene of those 3 samples of treated flu samples for each day group ie: 21 day group. The 'NT0_Mean' through 'NT21_Mean' are the non-treated means for each day group for the three non-treated samples over 21 days. The additional fields are the actual samples if a comparison would like to be made to find outliers in those gene expression values that might skew the data.

The top five genes expressed the most from day 0 non-treated to day 21 non-treated.

```
top5_nt21 <- All_FoldChange_Values[order(All_FoldChange_Values$FC_nt21,
                                         decreasing=TRUE)[1:5],]
top5_nt21$Gene
```

```
## [1] "SERPING1" "IFI27"      "SIGLEC1"  "RSAD2"    "C1QB"
```

The top five genes expressed the most from day 0 treated to day 21 treated.

```
top5_t21 <- All_FoldChange_Values[order(All_FoldChange_Values$FC_t21,
                                         decreasing=TRUE)[1:5],]
top5_t21$Gene
```

```
## [1] "CYP4F3"      "CAMP"          "LOC100130938" "POM121L10P"    "ZNF385D"
```

The top five genes expressed the most from day 0 treated to day 21 non-treated.

```
top5_b21 <- All_FoldChange_Values[order(All_FoldChange_Values$FCB_21,
                                         decreasing=TRUE)[1:5],]
top5_b21$Gene
```

```
## [1] "IL6"      "CCL2"     "IFNB1"    "IFI27"    "CCL8"
```

The least five genes expressed the most from day 0 non-treated to day 21 non-treated.

```
least5_nt21 <- All_FoldChange_Values[order(All_FoldChange_Values$FC_nt21,
                                             decreasing=FALSE)[1:5],]
least5_nt21$Gene
```

```
## [1] "CAV1" "GJB2" "OLR1" "HTRA1" "IL1A"
```

The least five genes expressed the most from day 0 treated to day 21 treated.

```
least5_t21 <- All_FoldChange_Values[order(All_FoldChange_Values$FC_t21,
                                          decreasing=FALSE)[1:5],]
least5_t21$Gene
```

```
## [1] "CH25H" "SERPINB2" "NR4A3" "HOXA5" "EREG"
```

The least five genes expressed the most from day 0 treated to day 21 non-treated.

```
least5_b21 <- All_FoldChange_Values[order(All_FoldChange_Values$FCB_21,
                                          decreasing=FALSE)[1:5],]
least5_b21$Gene
```

```
## [1] "RPS4Y1" "DDX3Y" "USP9Y" "CYorf15B" "EIF1AY"
```

Read in the Fold Change groups from the hemochromatosis data set.

```
hemochromatosis <- read.csv('all-fc-grops.csv', sep=',', header=TRUE)
hemo <- hemochromatosis[complete.cases(hemochromatosis$X),]
colnames(hemo)[1] <- 'Gene'
```

combine the data

```
combined <- merge(hemo, All_FoldChange_Values, by.x='Gene', by.y='Gene')
interested <- combined[,c(1,18:19,52:55)]
```

Order by most fold change after 7 days and 21 days, then by most fold change in the hemochromatosis males and hemochromatosis females.

```
flu21 <- interested[order(interested$FCB_21, decreasing=TRUE),]
flu7 <- interested[order(interested$FCB_7, decreasing=TRUE),]
maleHemo <- interested[order(interested$FC_malesOverall, decreasing=TRUE),]
femaleHemo <- interested[order(interested$FC_femalesOverall, decreasing=TRUE),]

write.csv(flu7, 'flu7AndHemo.csv', row.names=FALSE)
write.csv(flu21, 'flu21AndHemo.csv', row.names=FALSE)
```

Many interleukins are increased in the flu vaccine after 7 and 21 days without antibiotic treatments, this table can show how the hemochromatosis females and males overall gene expression values were compared to healthy females and males.

```
head(flu21,20)
```

##	Gene	FC_malesOverall	FC_femalesOverall	FCB_1	FCB_3	FCB_7
## 6262	IL6	0.9905198	1.0078106	1.3940525	1.7848902	3.1048767
## 2029	CCL2	0.9358741	0.9617681	0.9927215	1.3904404	1.7176532
## 6107	IFNB1	1.0775397	1.0634954	0.8122579	0.7823874	1.7877768

##	6078	IFI27	0.9781346	1.0099580	1.5668583	1.1451201	0.9983559
##	2041	CCL8	1.0184633	0.9500181	1.0547728	1.1109512	1.5256625
##	13565	TNF	0.9719170	1.0133758	1.6787810	1.5930246	2.1467874
##	2030	CCL20	0.9612723	0.9864661	1.6579669	1.6445958	2.8346409
##	11605	SERPING1	1.1503607	0.9444856	1.1368927	0.8896728	0.8471819
##	4584	FFAR2	1.0342964	0.9586904	1.3279832	1.3238931	1.8816057
##	1461	C15orf48	0.9640186	1.0411939	1.5048385	1.8665492	2.5865373
##	5345	GPR84	1.0058212	1.0722619	1.2693609	1.3145617	1.8714439
##	9709	PLA1A	1.0389679	1.0193292	1.1976270	1.1374683	2.2642583
##	6584	KCNJ2	1.0509034	0.9646049	1.3167536	1.3349880	1.8604059
##	3120	CXCL11	0.9447864	0.8601953	0.8790431	0.8292410	1.0457066
##	3127	CXCL3	0.9833485	1.0166585	1.0127438	1.0092076	2.2426897
##	11764	SIGLEC1	1.0567353	0.9703749	1.0702210	0.9482647	0.9184086
##	13569	TNFAIP6	0.9598734	1.0058044	1.0988208	1.3019843	1.8504047
##	3119	CXCL10	1.0749625	0.9750578	1.2474636	0.9558006	1.5206253
##	6222	IL1B	1.0464863	0.9363755	1.5198337	1.5644490	2.0113713
##	6197	IL12B	1.0256984	1.0793044	0.9597901	1.0565575	2.1763510
##		FCB_21					
##	6262		2.497008				
##	2029		2.339408				
##	6107		2.211414				
##	6078		2.143919				
##	2041		2.139143				
##	13565		2.122594				
##	2030		2.091839				
##	11605		2.001143				
##	4584		1.983356				
##	1461		1.967010				
##	5345		1.891219				
##	9709		1.836358				
##	6584		1.825781				
##	3120		1.774508				
##	3127		1.774431				
##	11764		1.774161				
##	13569		1.756858				
##	3119		1.744018				
##	6222		1.735314				
##	6197		1.701901				

```
head(flu7,20)
```

##	Gene	FC_malesOverall	FC_femalesOverall	FCB_1	FCB_3	FCB_7	
##	6262	IL6	0.9905198	1.0078106	1.3940525	1.7848902	3.104877
##	2030	CCL20	0.9612723	0.9864661	1.6579669	1.6445958	2.834641
##	1461	C15orf48	0.9640186	1.0411939	1.5048385	1.8665492	2.586537
##	9709	PLA1A	1.0389679	1.0193292	1.1976270	1.1374683	2.264258
##	3127	CXCL3	0.9833485	1.0166585	1.0127438	1.0092076	2.242690
##	6197	IL12B	1.0256984	1.0793044	0.9597901	1.0565575	2.176351
##	13565	TNF	0.9719170	1.0133758	1.6787810	1.5930246	2.146787
##	6221	IL1A	0.9460817	0.8910832	1.1353118	1.2229545	2.082487
##	6222	IL1B	1.0464863	0.9363755	1.5198337	1.5644490	2.011371
##	3118	CXCL1	1.0525977	0.9904077	0.9094089	1.0847690	1.918277
##	4584	FFAR2	1.0342964	0.9586904	1.3279832	1.3238931	1.881606
##	5345	GPR84	1.0058212	1.0722619	1.2693609	1.3145617	1.871444

```
## 6584 KCNJ2 1.0509034 0.9646049 1.3167536 1.3349880 1.860406
## 13569 TNFAIP6 0.9598734 1.0058044 1.0988208 1.3019843 1.850405
## 3126 CXCL2 1.0014514 0.9804473 1.0709207 1.2335290 1.805172
## 6107 IFNB1 1.0775397 1.0634954 0.8122579 0.7823874 1.787777
## 7823 MIR155HG 0.9522879 0.9761173 1.0005117 1.0901737 1.752125
## 2029 CCL2 0.9358741 0.9617681 0.9927215 1.3904404 1.717653
## 10448 PTX3 1.1115180 0.9912605 0.9653456 1.0699951 1.635240
## 10672 RASGEF1B 0.9786717 1.0053300 1.2054556 1.2796025 1.623240
## FCB_21
## 6262 2.497008
## 2030 2.091839
## 1461 1.967010
## 9709 1.836358
## 3127 1.774431
## 6197 1.701901
## 13565 2.122594
## 6221 1.434711
## 6222 1.735314
## 3118 1.488941
## 4584 1.983356
## 5345 1.891219
## 6584 1.825781
## 13569 1.756858
## 3126 1.520795
## 6107 2.211414
## 7823 1.523876
## 2029 2.339408
## 10448 1.314918
## 10672 1.504673
```

Lets map out the interleukins over 7 days after being vaccinated with the flu virus.

```
library(ggplot2)
library(tidyr)
```

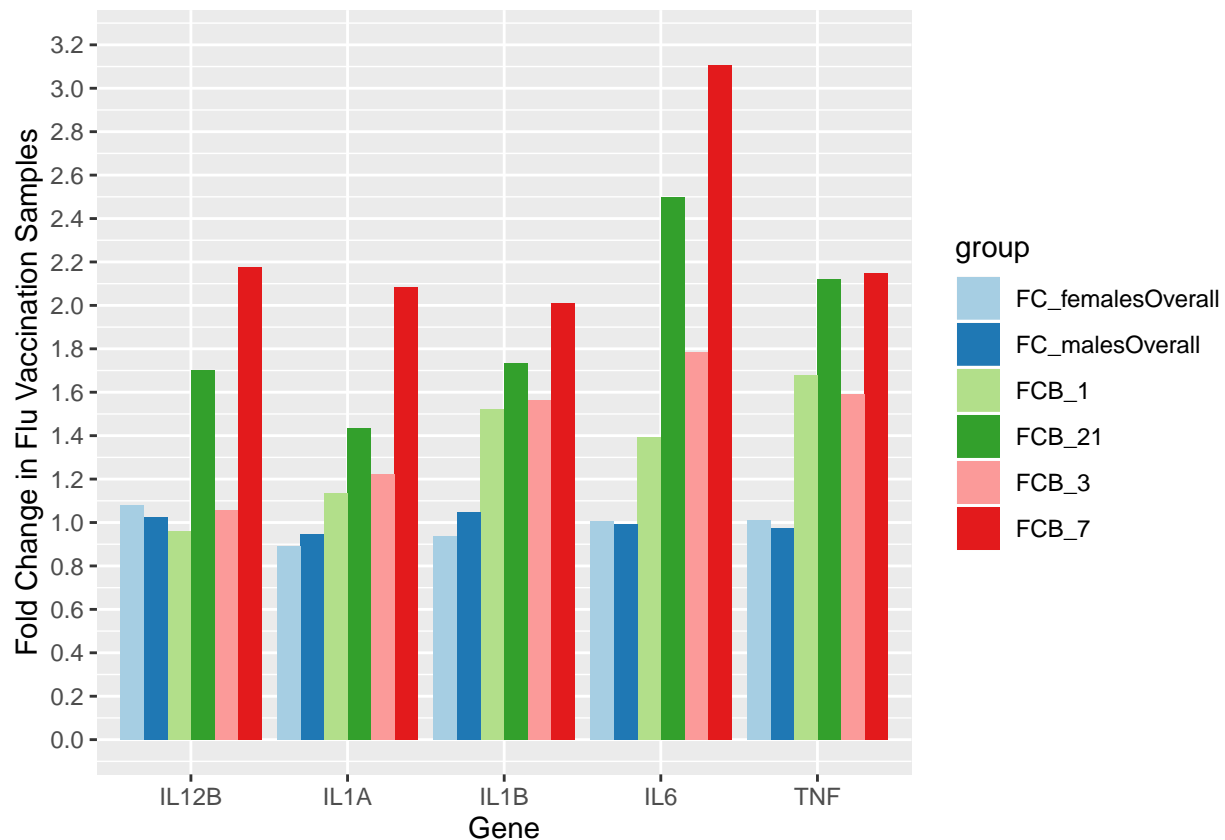
```
flu <- flu7[c(1,6:9),c(1:7)]
```

```
flu_5 <- gather(flu, 'group', 'foldChange', 2:7)
flu_5$foldChange <- round(flu_5$foldChange, 3)
head(flu_5)
```

```
##      Gene      group foldChange
## 1  IL6 FC_malesOverall 0.991
## 2 IL12B FC_malesOverall 1.026
## 3  TNF FC_malesOverall 0.972
## 4  IL1A FC_malesOverall 0.946
## 5  IL1B FC_malesOverall 1.046
## 6  IL6 FC_femalesOverall 1.008
```

```
ggplot(data = flu_5, aes(x=Gene, y=foldChange, fill=group)) +
  geom_bar(stat='identity', position=position_dodge()) +
  scale_y_continuous(breaks = seq(0, 3.2, by=.2), limits=c(0,3.2)) +
```

```
scale_fill_brewer(palette='Paired') +
ylab('Fold Change in Flu Vaccination Samples')
```



From the above plot interleukins and tumor necrosis factor gene expression values that are high in flu vaccination blood samples after 7 days is shown compared to patients who have hemochromatosis or too much iron production in their blood.

Now read in the Epigallocatechin or EGCG data set and compare the above genes with the females. This table also has the heart disease MI male means in it.

```
egcg <- read.csv('BothStudies.csv', sep=',', header=TRUE)
colnames(egcg)[1] <- 'Gene'
```

Combine the egcg only with the table of genes most expressed in flu vaccine samples.

```
combined2 <- merge(flu7, egcg, by.x='Gene', by.y='Gene')
combined2 <- combined2[order(combined2$FCB_7, decreasing=TRUE),]

combinedGenes <- combined2[c(1,5:8),]

FC_egcg <- combinedGenes %>% mutate(FC_egcg = Post_EGCG_Means/preEGCG_Means)
FC_MI <- FC_egcg %>% mutate(FC_MI = MI_Male_Means/HealthyMale_Means)

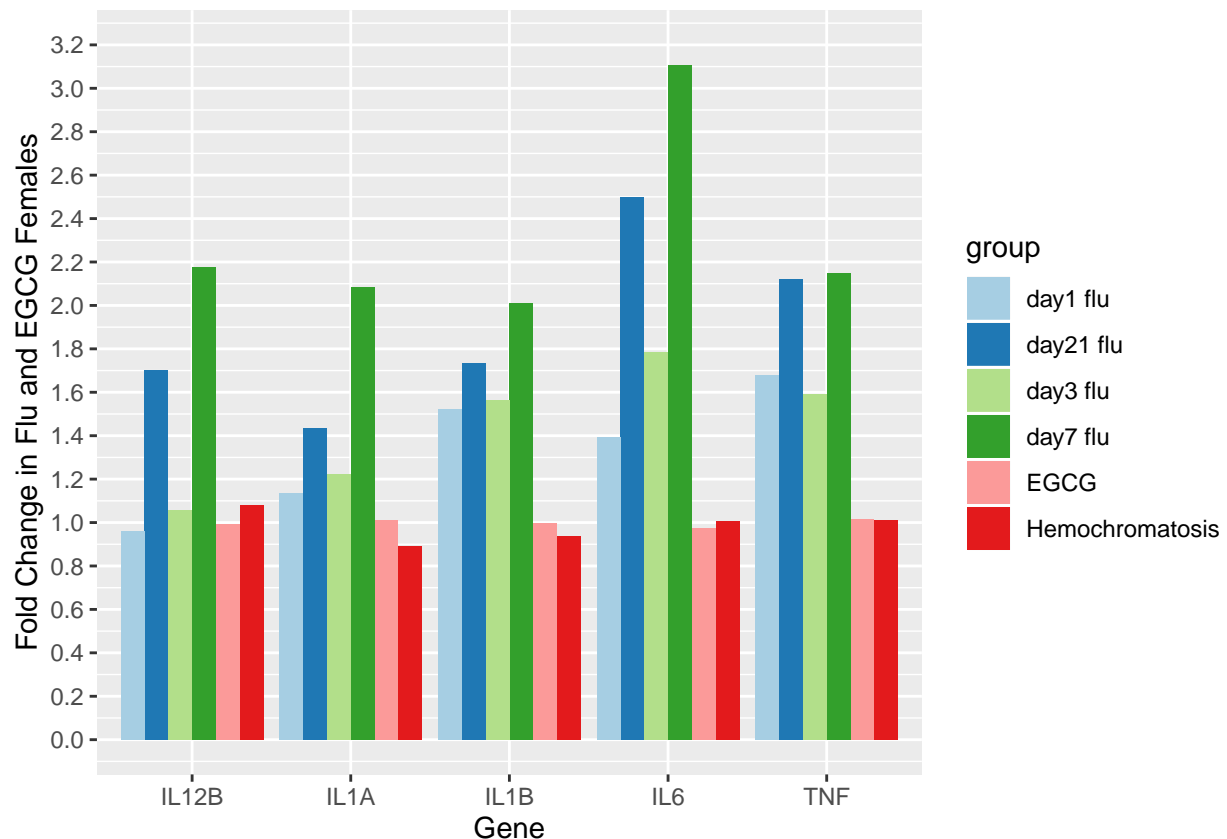
FC_EGCG <- FC_MI[,c(1,3:7,13)]
colnames(FC_EGCG) <- c('Gene','Hemochromatosis','day1 flu', 'day3 flu',
                      'day7 flu', 'day21 flu', 'EGCG')
```



```
flu_egcg <- gather(FC_EGCG, 'group', 'foldChange', 2:7)
flu_egcg$foldChange <- round(flu_egcg$foldChange, 3)
head(flu_egcg)
```

```
##      Gene      group foldChange
## 1  IL6 Hemochromatosis    1.008
## 2 IL12B Hemochromatosis    1.079
## 3   TNF Hemochromatosis    1.013
## 4  IL1A Hemochromatosis    0.891
## 5  IL1B Hemochromatosis    0.936
## 6   IL6      day1 flu    1.394
```

```
ggplot(data = flu_egcg, aes(x=Gene, y=foldChange, fill=group)) +
  geom_bar(stat='identity', position=position_dodge()) +
  scale_y_continuous(breaks = seq(0, 3.2, by=.2), limits=c(0,3.2)) +
  scale_fill_brewer(palette='Paired') +
  ylab('Fold Change in Flu and EGCG Females')
```



The above table shows the interleukins and tumor necrosis factor gene expression fold change values in females who have hemochromatosis and separately who are overweight and using EGCG compared to flu over 21 days.

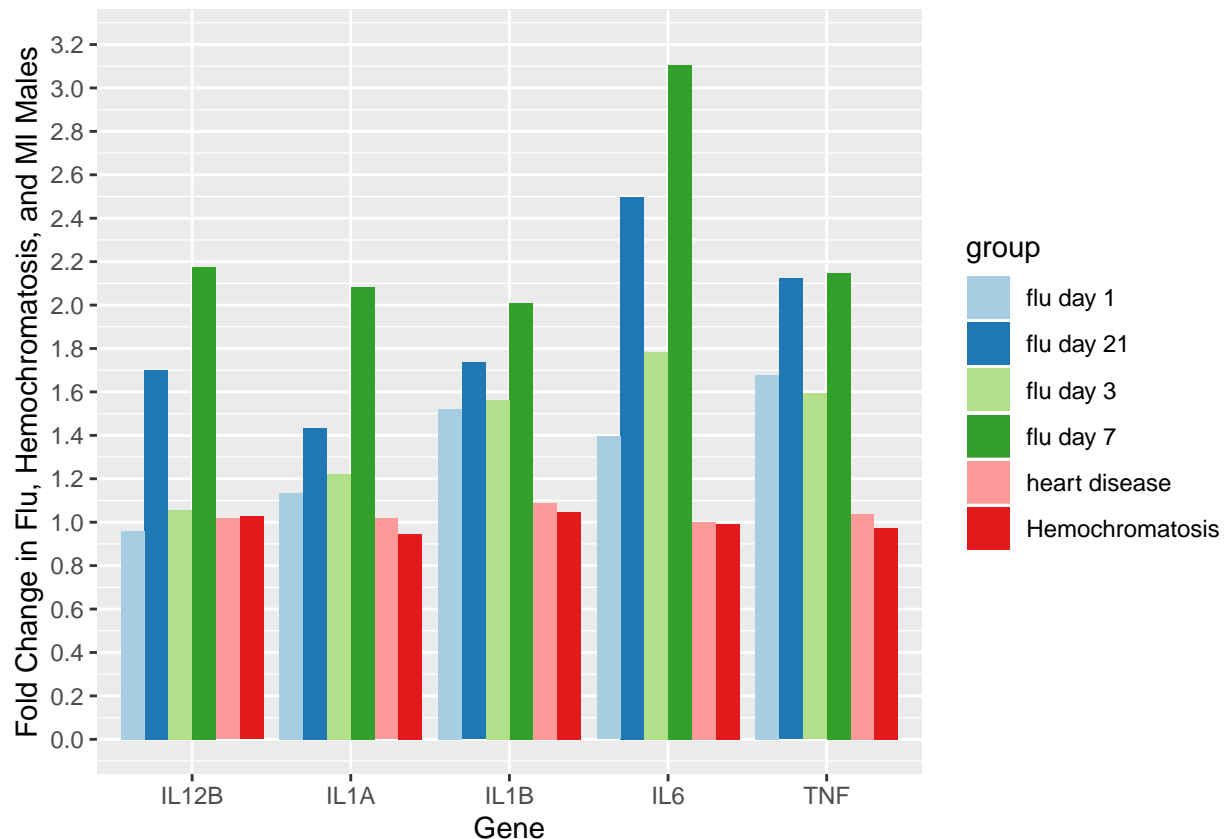
Now compare the males with hemochromatosis and males with heart disease.

```
MI <- FC_MI[,c(1,2,4:7,14)]
colnames(MI) <- c('Gene', 'Hemochromatosis', 'flu day 1', 'flu day 3',
                  'flu day 7', 'flu day 21', 'heart disease')
```

```
hd <- gather(MI, 'group', 'foldChange', 2:7)
hd$foldChange <- round(hd$foldChange, 3)
head(hd)
```

```
##      Gene      group foldChange
## 1  IL6 Hemochromatosis      0.991
## 2 IL12B Hemochromatosis      1.026
## 3   TNF Hemochromatosis      0.972
## 4  IL1A Hemochromatosis      0.946
## 5  IL1B Hemochromatosis      1.046
## 6   IL6      flu day 1      1.394
```

```
ggplot(data = hd, aes(x=Gene, y=foldChange, fill=group)) +
  geom_bar(stat='identity', position=position_dodge()) +
  scale_y_continuous(breaks = seq(0, 3.2, by=.2), limits=c(0,3.2)) +
  scale_fill_brewer(palette='Paired') +
  ylab('Fold Change in Flu, Hemochromatosis, and MI Males')
```



The above plot shows a bar chart of the same genes heavily expressed in flu samples after 7 days as compared to males who have heart disease called myocardial infarction (MI) and males who have hemochromatosis or iron toxicity in their blood.