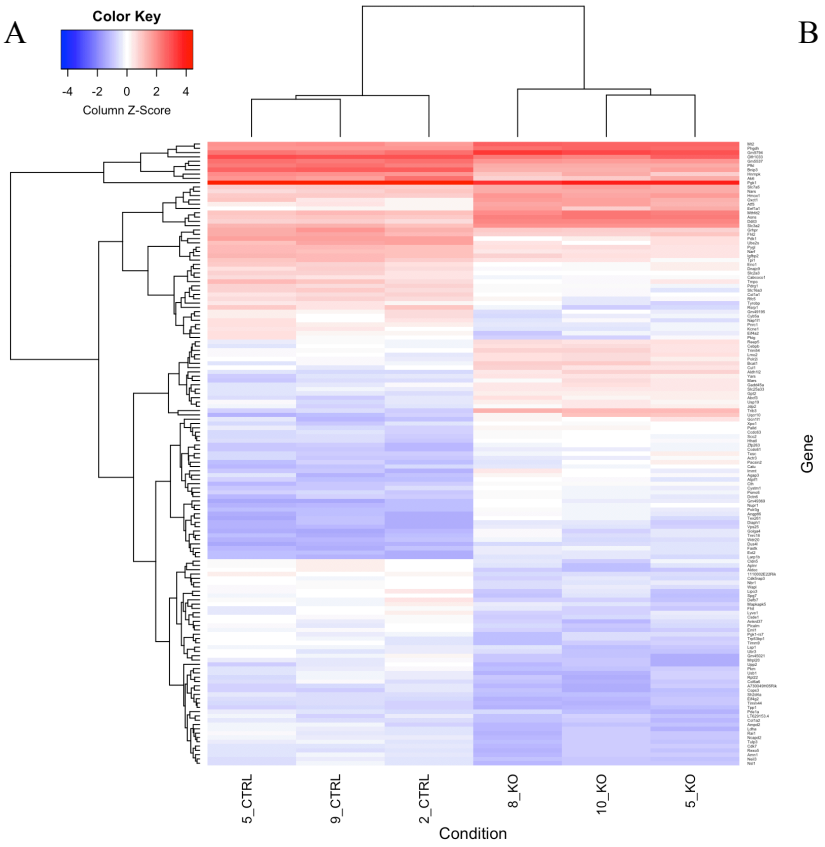
Gene 440-Fuctional Genomics ____/5pts
Eric Kil

Figure 1. Differentially expressed genes (DEG). A) Heatmap clustering of up- and down-regulated genes following *Hif1a* KO in mice cardiac tissue using log_2 fold change compared with control. Gene expression values are colored from blue (low) to red (high) with rows corresponding to gene names (i.e. *Pdk1*, *Bnip3*, *Trib3,* etc.) and columns corresponding to sample condition. B) List of top 10 most significant differentially expressed genes of *Hif1a* KO. For each gene, log_2 fold change as compared to control and p-values are shown.

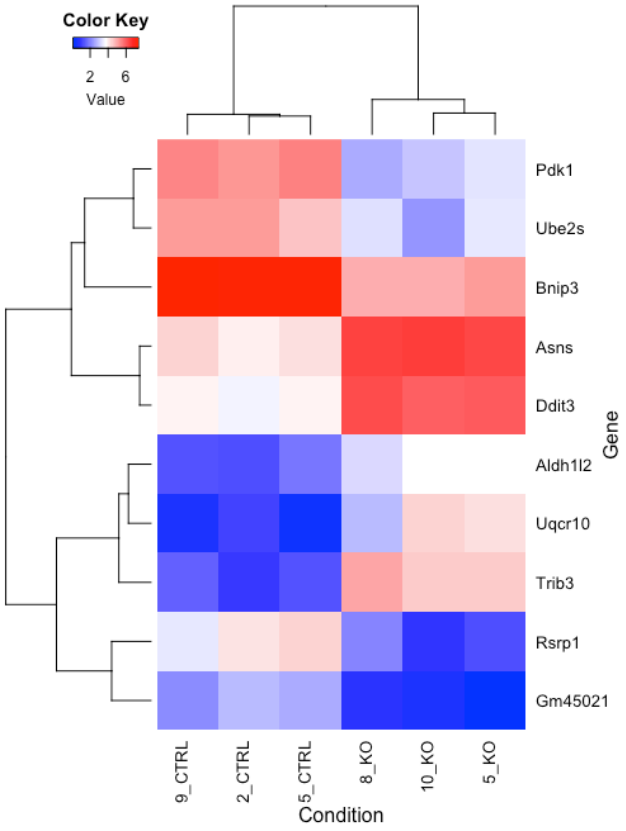
A



B

Gene ID	Log_2 Fold Change	P. value
Trib3	**10.54**	**1.4e-4**
Uqcr10	**8.51**	**1.5e-2**
Ddit3	**4.83**	**1.9e-4**
Asns	**4.63**	**2.5e-4**
Aldh1l2	**4.32**	**1.1e-3**
Ube2s	**-3.93**	**1.8e-2**
Gm45021	**-3.98**	**2.1e-3**
Bnip3	**-4.94**	**1.9e-4**
Pdk1	**-5.54**	**3.5e-3**
Rsrp1	**-6.09**	**5.3e-3**

Figure 2. Zoomed snapshot of the top 10 unique differential gene expression signatures between *Hif1a* knockout mice relative to control. Gene expression values are colored from blue (low) to red (high) with rows corresponding to gene names and columns corresponding to sample condition.



Rscript - Heatmap Generation

##fig 1 big clustering

```
#create sorted differential gene expression list
> upreganddownregsort <- upreganddownreg[order(upreganddownreg$FoldChange),]
#check if any duplicates in Gene.name column and the index of these duplicates
> duplicated(upreganddownregsort$Gene.name)
#remove indeces with a duplicate value
> keepfrombiglist <- upreganddownregsort[-c(4,8,10,12,17,34,60,79,85,109,130),]
#check to see if all values are FALSE
> View(keepfrombiglist)
> duplicated(keepfrombiglist$Gene.name)
#rename column names to something more concise/readable
> names(keepfrombiglist)[names(keepfrombiglist) == "tpm_CTRL_2"] <- "2_CTRL"
> names(keepfrombiglist)[names(keepfrombiglist) == "tpm_CTRL_5"] <- "5_CTRL"
> names(keepfrombiglist)[names(keepfrombiglist) == "tpm_CTRL_9"] <- "9_CTRL"
> names(keepfrombiglist)[names(keepfrombiglist) == "tpm_KO_8"] <- "8_KO"
> names(keepfrombiglist)[names(keepfrombiglist) == "tpm_KO_5"] <- "5_KO"
> names(keepfrombiglist)[names(keepfrombiglist) == "tpm_KO_10"] <- "10_KO"
#keep data corresponding to tpm values and gene name
> bigdf <- keepfrombiglist[c(2,3,4,5,6,7,14)]
> bigdf2 <- keepfrombiglist[c(2,3,4,5,6,7)]
#create a vector of string elements with Gene names
> bigvector <- bigdf[,7]
#make this vector of strings the rownames for dataframe without Gene.name column (df2)
> rownames(bigdf2) <- bigvector
> View(bigdf2)
#make data into a matrix for heatmap building
> bigmat_data <- data.matrix(bigdf2[,1:ncol(bigdf2)])
#scalar to help form heatmap
> bigdf2 <- scale(bigmat_data)
#heatmap command
> heatmap.2(bigmat_data, scale = "none", xlab = "Condition", cexRow = .3, cexCol = 1,
        ylab = "Gene", keysize = 0.5, lhei=c(1,5), lwid=c(1,3), col = bluered(100), trace = "none",
        offsetRow = 0, offsetCol = 0, density.info = "none")
```

##fig 2 small clustering

```
#identify top 5 upregulated genes from total significant calls; artificial fold change threshold set to 4
> top5 <- final[final$p.value< 0.05 & final$FoldChange>4,]
#identify top 5 downregulated genes; artificial fold change threshold set to -3.92
> bottom5 <- final[final$p.value< 0.05 & final$FoldChange< -3.92,]
#combine these two dataframes by rows
> select <- rbind(top5, bottom5)
#add annotations to these genes
> selectwanno <- merge(select, annotations, by = "target_id")
#make sure all gene names are given and no repeats
> View(selectwanno)
#if any repeats, delete the index that corresponds to that repeat
> keep <- selectwanno[-c(7),]
#change name of columns to something more concise/readable
```

```
> names(keep)[names(keep) == "tpm_CTRL_2"] <- "2_CTRL"
> names(keep)[names(keep) == "tpm_CTRL_5"] <- "5_CTRL"
> names(keep)[names(keep) == "tpm_CTRL_9"] <- "9_CTRL"
> names(keep)[names(keep) == "tpm_KO_8"] <- "8_KO"
> names(keep)[names(keep) == "tpm_KO_5"] <- "5_KO"
> names(keep)[names(keep) == "tpm_KO_10"] <- "10_KO"
#make sure all variables, columns, rows look good
> View(keep)
#take columns of tpm data specifically for matrix building, with gene names
> df <- keep[c(2,3,4,5,6,7,14)]
#make a separate dataframe of tpm data without gene names
> df2 <- keep[c(2, 3, 4,5,6,7)]
#take gene names column from first dataframe to build a vector with string elements
> avector <- df[,7]
#make the vector the names of the rows for more informative heatmap clustering
> rownames(df2) <- avector
#create matrix with tpm data
> mat_data <- data.matrix(df2[,1:ncol(df2)])
#appropriate scalar
> df2 <- scale(mat_data)
#generate the heatmap
> heatmap.2(mat_data, scale = "none", xlab = "Condition", cexRow = .8, cexCol = .8, keysize = 1,
        ylab = "Gene", srtCol = 50, col = bluered(100), trace = "none",
        density.info = "none" )
```