

SignTest__ImprintedGene

Sahar Mozaafari

4/23/2018

Testing for significant difference between maternal and paternal expression for genes across all individuals.

Read in not normalized maternal and paternal expression matrixes. Remove genes that are not expressed in maternal or paternal, and remove those genes expressed in less than 100 reads summed over all maternal and paternal for that gene.

```
# grab genes that are expressed maternally and not paternally.
monly <- as.data.frame((pat0[(which(!names(pat0)%in%names(mat0))))]))
monly$genes <- rownames(monly)
mgenes <- monly$genes
mm <- newmat2[c(mgenes),]
dim(monly)

## [1] 64 2

mcounts <- as.data.frame((apply(mm, 1,function(x) length(which(x>0)))))
mcounts$genes <- rownames(mcounts)
rownames(mcounts) <- NULL

monlygg <- merge(monly, mcounts, by="genes")
monlygenenames <- merge(monlygg, genenames2, by.x="genes", by.y="V13")
monlyimp <- merge(monlygenenames, supp, by.x="V2", by.y="Gene" )
print(monlyimp)

##          V2          genes (pat0[(which(!names(pat0) %in% names(mat0))))])
## 1 IFITM1 ENSG00000185885                                     7731
## (apply(mm, 1, function(x) length(which(x > 0))))              Position
## 1                                     1 chr11:313991-315272
## X.Hg38.
## 1 Maternal

supmtable <- monlygenenames[,c(1,3,4)]
colnames(supmtable)[2] <- "individuals"
write.table(supmtable, "S1Table_mat.txt", quote=F, col.names=T, row.names = F, sep="\t")

ponly <- as.data.frame((mat0[(which(!names(mat0)%in%names(pat0))))]))
dim(ponly)

## [1] 75 1

ponly$genes <- rownames(ponly)
pgenes <- ponly$genes
pp <- newpat2[c(pgenes),]
dim(ponly)

## [1] 75 2

pcounts <- as.data.frame((apply(pp, 1,function(x) length(which(x>0)))))
pcounts$genes <- rownames(pcounts)
```

```

rownames(pcounts) <- NULL
ponlygg <- merge(ponly, pcounts, by="genes")
ponlygenenames <- merge(ponlygg, genenames2, by.x="genes", by.y="V13")
ponlyimp <- merge(ponlygenenames, supp, by.x="V2", by.y="Gene" )
print(ponlyimp)

##          V2          genes (mat0[(which(!names(mat0) %in% names(pat0)))]))
## 1 CDKN1C ENSG00000129757                                     7787
## 2  NDN ENSG00000182636                                     10214
## 3 SNRPN ENSG00000128739                                     10215
## (apply(pp, 1, function(x) length(which(x > 0))))          Position
## 1                                     3  chr11:2883218-2885832
## 2                                     22 chr15:23685407-23687303
## 3                                     9  chr15:24823647-24978582
## X.Hg38.
## 1 Maternal
## 2 Paternal
## 3 Paternal

supptable <- ponlygenenames[,c(1,3,4)]
colnames(supptable)[2] <- "individuals"
write.table(supptable, "S1Table_pat.txt", quote=F, col.names=T, row.names = F, sep="\t")

## Warning in qvalue(as.numeric(as.character(Zall3$signpval))): NAs introduced
## by coercion

##          genes          less0 greater0 signpvalue
## [1,] "ENSG00000204186" "2" "148" "1.58711064069431e-41"
## [2,] "ENSG00000242265" "2" "136" "5.50553799467154e-38"
## [3,] "ENSG00000224078" "2" "131" "1.63687588352002e-36"
## [4,] "ENSG00000100138" "23" "189" "1.24132502395267e-33"
## [5,] "ENSG00000185513" "1" "107" "6.71764364602269e-31"
## [6,] "ENSG00000130844" "36" "184" "4.05268157677564e-25"
## [7,] "ENSG00000139679" "0" "76" "2.64697796016969e-23"
## [8,] "ENSG00000145945" "0" "75" "5.29395592033941e-23"
## [9,] "ENSG00000053918" "79" "1" "1.3400325923359e-22"
## [10,] "ENSG00000168994" "12" "81" "9.82875021147483e-14"
## [11,] "ENSG00000257151" "0" "43" "2.27373675443232e-13"
## [12,] "ENSG00000177432" "0" "29" "3.72529029846191e-09"
## [13,] "ENSG00000196756" "113" "45" "6.20406731749264e-08"
## [14,] "ENSG00000198346" "63" "132" "0.000000869597040247412"
## [15,] "ENSG00000140443" "14" "49" "0.000011123023000449"
## [16,] "ENSG00000100592" "66" "126" "0.0000178070355770431"
## [17,] "ENSG00000271122" "52" "106" "0.0000208698418624478"
## [18,] "ENSG00000253954" "32" "6" "0.0000243425602093339"
## [19,] "ENSG00000128654" "0" "16" "0.000030517578125"
## [20,] "ENSG00000160352" "35" "79" "0.0000460924737125119"
## [21,] "ENSG00000179632" "17" "51" "0.000044534487714936"
## [22,] "ENSG00000146143" "30" "71" "0.0000553307058758015"
## [23,] "ENSG00000172349" "61" "115" "0.0000571463568305444"
## [24,] "ENSG00000214078" "111" "58" "0.0000556233723726144"
## [25,] "ENSG00000159720" "32" "7" "0.000070254784077406"
## [26,] "ENSG00000180185" "68" "29" "0.0000934068710022919"
## [27,] "ENSG00000064666" "127" "72" "0.000117901945332261"
## [28,] "ENSG00000183199" "7" "31" "0.000116167007945478"

```

```
##      V2      signFDR
## [1,] "ZDBF2"      "1.69582771958187e-37"
## [2,] "PEG10"      "2.94133367365327e-34"
## [3,] "SNHG14"      "5.8300062718038e-33"
## [4,] "NHP2L1"      "3.31588947023357e-30"
## [5,] "L3MBTL1"      "1.43556044715505e-27"
## [6,] "ZNF331"      "7.21715044130795e-22"
## [7,] "LPAR6"      "4.04042278634473e-20"
## [8,] "FAM50B"      "7.07073987610333e-20"
## [9,] "KCNQ1"      "1.59091647212323e-19"
## [10,] "PXDC1"      "1.05020196009609e-10"
## [11,] "PWAR6"      "2.20862520191903e-10"
## [12,] "NAP1L5"      "0.00000331706056992213"
## [13,] "SNHG17"      "0.0000509926609903145"
## [14,] "ZNF813"      "0.000663688883931686"
## [15,] "IGF1R"      "0.00792330005065317"
## [16,] "DAAM1"      "0.0118917609462941"
## [17,] "RP11-379H18.1" "0.0131173094294267"
## [18,] "HMG1P38"      "0.0144500142131518"
## [19,] "MTX2"      "0.0171621222245066"
## [20,] "ZNF714"      "0.0234522896008662"
## [21,] "MAF1"      "0.0234522896008662"
## [22,] "PRIM2"      "0.0254420342805986"
## [23,] "IL16"      "0.0254420342805986"
## [24,] "CPNE1"      "0.0254420342805986"
## [25,] "ATP6VOD1"      "0.0300268947146833"
## [26,] "FAHD1"      "0.0383866314099803"
## [27,] "CNN2"      "0.0449922244955432"
## [28,] "HSP90AB3P"      "0.0449922244955432"
```

```
sigsign <- as.data.frame(x)
colnames(sigsign)[5] <- "genename"
i <- 28

plotparent <- function(i, lim=100){
  n <- sigsign$genes[i]
  v <- grep(n, genes)
  gg <- as.data.frame(cbind(as.numeric(as.character(newmat3[v,])), as.numeric(as.character(newpat3[v,]))))
  gg$V3 <- gg$V2-gg$V1
  gg$Parent =gg$V3>0
  gg$Parent[gg$V3>0] <- paste("Paternal", ":", sigsign$greater0[i])
  gg$Parent[gg$V3<0] <- paste("Maternal", ":", sigsign$less0[i])
  gg$Parent[gg$V3==0] <- "Equal"
  gg2 <- gg[which((gg$V1+gg$V2)>5),]
  gg2$Parent[gg2$V3==0] <- paste("Equal", ":", length(which(gg2$V1==gg2$V2)))
  ggplot(gg2, aes(V1, V2, color=Parent,alpha=0.2))+
    geom_point(size=3, alpha=0.5)+
    geom_abline(slope=1, color="grey")+
    labs(title=paste(sigsign$genename[i], "Maternal vs. Paternal Expression"),
         x="Maternal Expression",
         y="Paternal Expression")+
    coord_cartesian(xlim=c(-1,lim), ylim=c(-1,lim))+
    scale_color_manual(values=c( "#800000CC" , "#0F425CCC", "#616530CC" ))+
    theme(legend.position = c(0.6, 0.3),
```

```

        legend.background = element_rect(color = "black", size = 0.5, linetype = 1))

}

pdf("ZDBF2.pdf", height=6, width=6)
plotparent(1)
dev.off()

## pdf
## 2

pdf("PEG10.pdf", height=6, width=6)
plotparent(2)
dev.off()

## pdf
## 2

pdf("SNHG14.pdf", height=6, width=6)
plotparent(3)
dev.off()

## pdf
## 2

plotparent2 <- function(i, lim=100){
n <- sigsign$genes[i]
v <- grep(n, genes)
gg <- as.data.frame(cbind(as.numeric(as.character(newmat3[v,])), as.numeric(as.character(newpat3[v,]))))
gg$V3 <- gg$V2-gg$V1
gg$Parent =gg$V3>0
gg$Parent[gg$V3>0] <- paste("Paternal", ":", sigsign$greater0[i])
gg$Parent[gg$V3<0] <- paste("Maternal", ":", sigsign$less0[i])
gg$Parent[gg$V3==0] <- "Equal"
gg2 <- gg[which((gg$V1+gg$V2)>5),]
gg2$Parent[gg2$V3==0] <- paste("Equal", ":", length(which(gg2$V1==gg2$V2)))
ggplot(gg2, aes(V1, V2, color=Parent,alpha=0.2))+
  geom_point(size=3, alpha=0.5)+
  geom_abline(slope=1, color="grey")+
  labs(title=paste(sigsign$genename[i], "Maternal vs. Paternal Expression"),
       x="Maternal Expression",
       y="Paternal Expression")+
  coord_cartesian(xlim=c(-1,lim), ylim=c(-1,lim))+
  scale_color_manual(values=c("#616530CC", "#800000CC" , "#0F425CCC" ))+
  theme(legend.position = c(0.6, 0.3),
        legend.background = element_rect(color = "black", size = 0.5, linetype = 1))

}

pdf("NHP2L1.pdf", height=6, width=6)
plotparent2(4, lim=150)
dev.off()

## pdf
## 2

```

```
pdf("L3MBTL1.pdf", height=6, width=6)
plotparent(5, lim=50)
dev.off()
```

```
## pdf
## 2
```

```
pdf("ZNF331.pdf", height=6, width=6)
plotparent2(6)
dev.off()
```

```
## pdf
## 2
```

```
plotparent3 <- function(i, lim=100){
  n <- sigsign$genes[i]
  v <- grep(n, genes)
  gg <- as.data.frame(cbind(as.numeric(as.character(newmat3[v,])), as.numeric(as.character(newpat3[v,]))))
  gg$V3 <- gg$V2-gg$V1
  gg$Parent =gg$V3>0
  gg$Parent[gg$V3>0] <- paste("Paternal", ":", sigsign$greater0[i])
  gg$Parent[gg$V3<0] <- paste("Maternal", ":", sigsign$less0[i])
  gg$Parent[gg$V3==0] <- "Equal"
  gg2 <- gg[which((gg$V1+gg$V2)>5),]
  gg2$Parent[gg2$V3==0] <- paste("Equal", ":", length(which(gg2$V1==gg2$V2))))
  ggplot(gg2, aes(V1, V2, color=Parent,alpha=0.2))+
    geom_point(size=3, alpha=0.5)+
    geom_abline(slope=1, color="grey")+
    labs(title=paste(sigsign$genename[i], "Maternal vs. Paternal Expression"),
         x="Maternal Expression",
         y="Paternal Expression")+
    coord_cartesian(xlim=c(-1,lim), ylim=c(-1,lim))+
    scale_color_manual(values=c("#0F425CCC", "#616530CC", "#800000CC" ))+
    theme(legend.position = c(0.6, 0.3),
          legend.background = element_rect(color = "black", size = 0.5, linetype = 1))
}
pdf("LPAR6.pdf", height=6, width=6)
plotparent3(7, lim=50)
dev.off()
```

```
## pdf
## 2
```

```
pdf("FAM50B.pdf", height=6, width=6)
plotparent3(8, lim=50)
dev.off()
```

```
## pdf
## 2
```

```
pdf("KCNQ1.pdf", height=6, width=6)
plotparent2(9, lim=50)
dev.off()
```

```
## pdf
```

```

##      2
pdf("PXDC1.pdf", height=6, width=6)
plotparent2(10, lim=50)
dev.off()

## pdf
##      2
pdf("PWAR6.pdf", height=6, width=6)
plotparent3(11, lim=50)
dev.off()

## pdf
##      2
pdf("NAP1L5.pdf", height=6, width=6)
plotparent3(12, lim=20)
dev.off()

## pdf
##      2
pdf("SNHG17.pdf", height=6, width=6)
plotparent2(13)
dev.off()

## pdf
##      2
pdf("ZNF813.pdf", height=6, width=6)
plotparent2(14, lim=50)
dev.off()

## pdf
##      2
pdf("IGF1R.pdf", height=6, width=6)
plotparent2(15, lim=50)
dev.off()

## pdf
##      2
pdf("DAAM1.pdf", height=6, width=6)
plotparent2(16)
dev.off()

## pdf
##      2
pdf("RP11-379H18.1.pdf", height=6, width=6)
plotparent2(17, lim=75)
dev.off()

## pdf
##      2
pdf("HMG1P38.pdf", height=6, width=6)
plotparent(18, lim=50)
dev.off()

```

```

## pdf
## 2

plotparent4 <- function(i, lim=100){
  n <- sigsign$genes[i]
  v <- grep(n, genes)
  gg <- as.data.frame(cbind(as.numeric(as.character(newmat3[v,])), as.numeric(as.character(newpat3[v,]))))
  gg$V3 <- gg$V2-gg$V1
  gg$Parent =gg$V3>0
  gg$Parent[gg$V3>0] <- paste("Paternal", ":", sigsign$greater0[i])
  gg$Parent[gg$V3<0] <- paste("Maternal", ":", sigsign$less0[i])
  gg$Parent[gg$V3==0] <- "Equal"
  gg2 <- gg[which((gg$V1+gg$V2)>5),]
  gg2$Parent[gg2$V3==0] <- paste("Equal", ":", length(which(gg2$V1==gg2$V2)))
  ggplot(gg2, aes(V1, V2, color=Parent,alpha=0.2))+
    geom_point(size=3, alpha=0.5)+
    geom_abline(slope=1, color="grey")+
    labs(title=paste(sigsign$genename[i], "Maternal vs. Paternal Expression"),
         x="Maternal Expression",
         y="Paternal Expression")+
    coord_cartesian(xlim=c(-1,lim), ylim=c(-1,lim))+
    scale_color_manual(values=c( "#616530CC", "#0F425CCC" ))+
    theme(legend.position = c(0.6, 0.3),
          legend.background = element_rect(color = "black", size = 0.5, linetype = 1))
}

pdf("MTX2.pdf", height=6, width=6)
plotparent4(19, lim=50)
dev.off()

## pdf
## 2

pdf("ZNF714.pdf", height=6, width=6)
plotparent2(20, lim=75)
dev.off()

## pdf
## 2

pdf("MAF1.pdf", height=6, width=6)
plotparent2(21, lim=150)
dev.off()

## pdf
## 2

pdf("PRIM2.pdf", height=6, width=6)
plotparent2(22, lim=50)
dev.off()

## pdf
## 2

```

```
pdf("IL16.pdf", height=6, width=6)
plotparent2(23, lim=200)
dev.off()
```

```
## pdf
## 2
```

```
pdf("CPNE1.pdf", height=6, width=6)
plotparent2(24, lim=200)
dev.off()
```

```
## pdf
## 2
```

```
pdf("ATP6V0D1.pdf", height=6, width=6)
plotparent2(25)
dev.off()
```

```
## pdf
## 2
```

```
pdf("FAHD1.pdf", height=6, width=6)
plotparent2(26, lim=150)
dev.off()
```

```
## pdf
## 2
```

```
pdf("CNN2.pdf", height=6, width=6)
plotparent2(27, lim=400)
dev.off()
```

```
## pdf
## 2
```

```
pdf("HSP90AB3P.pdf", height=6, width=6)
plotparent2(28, lim=20)
dev.off()
```

```
## pdf
## 2
```