

Cluster Analysis of Wasps

Lindsay Rutter

April 26, 2016

Introduction

```
> rm(list=ls())
> listcond = c("DU", "DU", "DR", "DU", "DR", "DR", "DR", "DR", "DU", "DU", "DR", "DU")
> load("All_wasp.rda")
> #listcond = rep(c("DR", "DU"), each= 6)
> y = DGEList(counts=countTable[,c(1:12)], group=listcond)
> colnames(y[[1]]) = listcond = c("DU.1", "DU.2", "DR.1", "DU.3", "DR.2", "DR.3", "DR.4", "DR.5", "DU.4", "DU.5")
> keep <- rowSums(cpm(y)>1) >= 6
> y <- y[keep, keep.lib.sizes=FALSE]
> y <- calcNormFactors(y)
> y = estimateCommonDisp(y)
> y = estimateTagwiseDisp(y)
> de = exactTest(y, pair=c("DR", "DU"))
> tt = topTags(de, n=nrow(y))
> head(tt$table)
```

	logFC	logCPM	PValue	FDR
85120	-7.374978	1.8914565	7.312417e-13	2.399423e-08
101844	-6.749977	5.6499009	1.382769e-11	2.268641e-07
86302	6.023108	1.6695042	3.663845e-11	4.007391e-07
39871	4.371550	0.5919034	5.162840e-07	4.235206e-03
33361	-4.598356	3.7215474	1.325933e-06	8.701570e-03
101598	-3.652124	2.1664939	3.142334e-06	1.718490e-02

```
> length(which((tt$table)$FDR < 0.05))
```

```
[1] 10
```

```
> nc = cpm(y, normalized.lib.sizes=TRUE)
> rn = rownames(tt$table)
> # Sorted in order of lowest FDR from DE comparison
> head(nc[rn,order(listcond)],5)
```

	DR.1	DR.2	DR.3	DR.4	DR.5	DR.6
85120	2.789968	7.6843366	2.679773	6.706472	10.5641651	6.778788
101844	146.958269	283.7456294	1.399515	89.564482	54.4869451	13.356537
86302	0.000000	0.2345052	0.000000	0.000000	0.0000000	0.000000
39871	0.000000	0.0000000	0.000000	0.000000	0.2927783	0.000000
33361	21.131313	50.4214880	17.600917	3.299965	13.3078296	39.661250

	DU.1	DU.2	DU.3	DU.4	DU.5	DU.6
85120	0.0000000	0.000000	0.000000	0.000000	0.000000	0.000000
101844	0.4270689	0.000000	1.200535	1.343442	0.000000	2.355310
86302	3.9332515	8.887964	4.034935	7.542084	2.350748	3.662323

```
39871  2.5624131 1.286659 2.383311 1.724980 1.070042 1.505248
33361  1.0360374 0.000000 0.000000 4.581138 0.000000 0.000000
```

```
> # just for plotting purposes
> deg = rn[tt$table$FDR < .05] # Only 7
> write.csv(tt$table, file="TopDEG_Perm1.csv")

> topInfo = cbind(nc[rn,order(listcond)], tt$table)
> for (i in 1:100){
+   gene = topInfo[i,1:12]
+   rep = 6
+   fact = 2
+   dat = data.frame(x=rep(1:fact, each=rep),y=t(gene),z=rep(1:rep, times = fact))
+   colnames(dat)=c("x","y","rep")
+   dat$x=as.factor(dat$x)
+   levels(dat$x)=c("DR","DU")
+   genePlot = ggplot(dat, aes(x, y)) + geom_point(aes(colour = factor(x)), shape = 20, size=5) + scale_x_discrete()
+   jpeg(file = paste(getwd(), "/Perm1/", "Gene_", i, ".jpg", sep=""), height = 700, width = 700)
+   print(genePlot)
+   dev.off()
+ }

> rm(list=ls())
> listcond = c("DU","DR","DU","DR","DR","DU","DR","DR","DU","DR","DU","DU")
> load("All_wasp.rda")
> #listcond = rep(c("DR","DU"),each= 6)
> y = DGEList(counts=countTable[,c(1:12)], group=listcond)
> colnames(y[[1]]) = listcond = c("DU.1","DR.1","DU.2","DR.2","DR.3","DU.3","DR.4","DR.5","DU.4","DR.6","DU.5","DR.6")
> keep <- rowSums(cpm(y)>1) >= 6
> y <- y[keep, keep.lib.sizes=FALSE]
> y <- calcNormFactors(y)
> y = estimateCommonDisp(y)
> y = estimateTagwiseDisp(y)
> de = exactTest(y, pair=c("DR","DU"))
> tt = topTags(de, n=nrow(y))
> head(tt$table)

      logFC    logCPM      PValue      FDR
112442  7.179668 1.7294038 1.373485e-11 2.269622e-07
83441   -5.900662 1.9752286 1.383367e-11 2.269622e-07
60736   -3.213624 0.8525100 5.034825e-06 5.506924e-02
44738    3.345472 1.7837084 9.029528e-06 7.407148e-02
68913    2.784105 0.7869036 5.970078e-05 3.917923e-01
102684   3.488239 1.7064023 1.514007e-04 8.279852e-01

> length(which((tt$table)$FDR < 0.05))

[1] 2

> nc = cpm(y, normalized.lib.sizes=TRUE)
> rn = rownames(tt$table)
> # Sorted in order of lowest FDR from DE comparison
> head(nc[rn,order(listcond)],5)

      DR.1      DR.2      DR.3      DR.4      DR.5      DR.6      DU.1
112442 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 3.606359
```

```

83441  7.052894 5.6155213 3.837879 8.683087 10.7694117 3.4007264 0.000000
60736  1.558228 1.3106435 2.845521 2.220537  2.7104629 2.5480369 0.000000
44738  0.000000 0.0000000 0.000000 1.755122  0.6761066 0.3878901 8.167033
68913  0.000000 0.3551879 0.000000 0.000000  0.3018333 1.0031641 1.318114

```

```

      DU.2    DU.3      DU.4    DU.5      DU.6
112442 2.373868 2.574546 10.3445041 8.346265 5.2045302
83441  0.000000 0.000000  0.000000 0.000000 0.4804699
60736  0.000000 0.000000  0.6233571 0.000000 0.5846277
44738  1.864968 6.390767  7.1981627 2.582096 4.9592554
68913  2.394822 1.753778  1.6121305 2.827113 2.0159575

```

```

> # just for plotting purposes
> deg = rn[tt$table$FDR < .05] # Only 7
> write.csv(tt$table, file="TopDEG_Perm2.csv")

> topInfo = cbind(nc[rn,order(listcond)], tt$table)
> for (i in 1:100){
+   gene = topInfo[i,1:12]
+   rep = 6
+   fact = 2
+   dat = data.frame(x=rep(1:fact, each=rep),y=t(gene),z=rep(1:rep, times = fact))
+   colnames(dat)=c("x","y","rep")
+   dat$x=as.factor(dat$x)
+   levels(dat$x)=c("DR","DU")
+   genePlot = ggplot(dat, aes(x, y)) + geom_point(aes(colour = factor(x)), shape = 20, size=5) + scale_x_discrete()
+   jpeg(file = paste(getwd(), "/Perm2/", "Gene_", i, ".jpg", sep=""), height = 700, width = 700)
+   print(genePlot)
+   dev.off()
+ }

```