

Vanessa cardui DE

Code

Differential expression in *Vanessa cardui* adults

Experimental set up was prepared with adult females in two food availability conditions: ad libitum and no resources, to analyze the influence that environmental conditions may have on investment in migration or reproduction in imagines

Abdomen tissue

In this section we analyse differential expression in the abdomen tissue

Preparing data We upload count data from the DESeq2 object crated within the nf-core rnaq pipeline. Pipeline was run for the entire experimental dataset therefore needs to be split into different sets/tissues:

Pipeline was run for the entire experimental dataset therefore needs to be split into different sets/tissues:

Hide

```
colDataLRab<-colDataLR[colDataLR$tissue == "A", ] #selecting abdomen samples only
#selecting samples with abdomen tissue based on ID
countLRab<-countLR[,c(1,2,3,4,5,10,11,12,13,14)]
names(colDataLRab)[names(colDataLRab)=="Group3"]<-"tissue"
names(colDataLRab)[names(colDataLRab)=="Group1"]<-"condition"

head(countLR)
```

	AL_ADF_A_11	AL_ADF_A_19	AL_ADF_A_3	AL_ADF_A_7	AL_ADF_A_9	AL_ADF_H_11	AL_ADF_H_3	AL_ADF_H_7
Vcard_DToL00001	0	0	0	0	0	0	0	0
Vcard_DToL00002	0	0	0	0	0	0	0	0
Vcard_DToL00003	0	0	0	0	0	0	0	0
Vcard_DToL00004	0	0	0	0	0	0	0	0
Vcard_DToL00005	0	0	0	0	0	0	0	0
Vcard_DToL00006	0	0	0	0	0	0	0	0
	AL_ADF_H_9	LI_ADF_A_10	LI_ADF_A_12	LI_ADF_A_14	LI_ADF_A_16	LI_ADF_A_8	LI_ADF_H_10	LI_ADF_H_12
Vcard_DToL00001	0	0	0	0	0	0	0	0
Vcard_DToL00002	0	0	0	0	0	0	0	0
Vcard_DToL00003	0	0	0	0	0	0	0	0
Vcard_DToL00004	0	0	0	0	0	0	0	0
Vcard_DToL00005	0	0	0	0	0	0	0	0
Vcard_DToL00006	0	0	0	0	0	0	0	0
	LI_ADF_H_14	LI_ADF_H_16	LI_ADF_H_8					
Vcard_DToL00001	0	0	0					
Vcard_DToL00002	0	0	0					
Vcard_DToL00003	0	0	0					
Vcard_DToL00004	0	0	0					
Vcard_DToL00005	0	0	0					
Vcard_DToL00006	0	2	0					

Hide

```
head(colDataLR)
```

```
DataFrame with 6 rows and 5 columns
  sample    condition    tissue    Group4 sizeFactor
  <character> <character> <character> <character> <numeric>
AL_ADF_A_11 AL_ADF_A_11    AL      A      11  1.117485
AL_ADF_A_19 AL_ADF_A_19    AL      A      19  1.035097
AL_ADF_A_3  AL_ADF_A_3    AL      A      3  1.056977
AL_ADF_A_7  AL_ADF_A_7    AL      A      7  0.883751
AL_ADF_A_9  AL_ADF_A_9    AL      A      9  1.579290
AL_ADF_H_11 AL_ADF_H_11    AL      H      11  0.826602
```

Constructing DESeq2 object, at this point counts are transformed according to the DESeq2 specific algorithm. Current hypothesis includes split into treatment groups only (~condition), filtering out genes with less than 10 total counts

Hide

```
ddsLRab <- DESeqDataSetFromMatrix(countData=countLRab, colData =colDataLRab, design = ~condition)
```

some variables in design formula are characters, converting to factors

Hide

```
#Basic filtering
keep <- rowSums(counts(ddsLRab)) >= 10
ddsLRab <- ddsLRab[keep,]
head(ddsLRab)
```

```
class: DESeqDataSet
dim: 6 10
metadata(1): version
assays(1): counts
rownames(6): Vcard_DToL00007 Vcard_DToL00010 ... Vcard_DToL00013 Vcard_DToL00014
rowData names(0):
colnames(10): AL_ADF_A_11 AL_ADF_A_19 ... LI_ADF_A_16 LI_ADF_A_8
colData names(5): sample condition tissue Group4 sizeFactor
```

DESeq analysis Main valuation, calculation of p-values

Hide

```
#ddsLRab <- DESeq(ddsLRab) #running the main function
resLRab <- results(ddsLRab)
resLRab <- resLRab[order(resLRab$padj),]
head(resLRab)
```

```
log2 fold change (MLE): condition LI vs AL
Wald test p-value: condition LI vs AL
DataFrame with 6 rows and 6 columns
  baseMean log2FoldChange    lfcSE      stat      pvalue      padj
  <numeric>    <numeric>    <numeric>    <numeric>    <numeric>    <numeric>
Vcard_DToL13592    541.427    -1.89555    0.313908    -6.03855    1.55504e-09    1.56422e-05
Vcard_DToL04829     52.077    -2.30664    0.401831    -5.74031    9.45037e-09    4.75306e-05
Vcard_DToL13539     274.921    1.23667    0.269397    4.59051    4.42161e-06    1.48257e-02
Vcard_DToL02600     478.949    -2.46656    0.567757    -4.34440    1.39658e-05    3.27025e-02
Vcard_DToL10357     341.966    -1.37088    0.323629    -4.23596    2.27575e-05    3.27025e-02
Vcard_DToL13220     160.525    4.34707    1.023610    4.24681    2.16838e-05    3.27025e-02
```

Hide

```
summary(resLRab)
```

```
out of 11237 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)      : 10, 0.089%
LFC < 0 (down)   : 12, 0.11%
outliers [1]     : 112, 1%
low counts [2]   : 1066, 9.5%
(mean count < 5)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```

Setting up p-value to 0.05

Hide

```
res05 <- results(ddsLRab, alpha=0.05)
summary(res05)
```

```
out of 11237 with nonzero total read count
adjusted p-value < 0.05
LFC > 0 (up)      : 2, 0.018%
LFC < 0 (down)   : 5, 0.044%
outliers [1]     : 112, 1%
low counts [2]   : 0, 0%
(mean count < 1)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```

Log fold change shrinkage is recommended for visualization and ranking, her we follow main DESeq2 vignette exactly

*Shrinkage of effect size (LFC estimates) is useful for visualization and ranking of genes. To shrink the LFC, we pass the dds object to the function lfcShrink. Below we specify to use the apeglm method for effect size shrinkage (Zhu, Ibrahim, and Love 2018), which improves on the previous estimator.

We provide the dds object and the name or number of the coefficient we want to shrink, where the number refers to the order of the coefficient as it appears in resultsNames(dds).*

Hide

```
head(resultsNames(ddsLRab))
```

```
[1] "Intercept"          "condition_LI_vs_AL"
```

Hide

```
resLFC <- lfcShrink(ddsLRab, coef="condition_LI_vs_AL", type="apeglm")

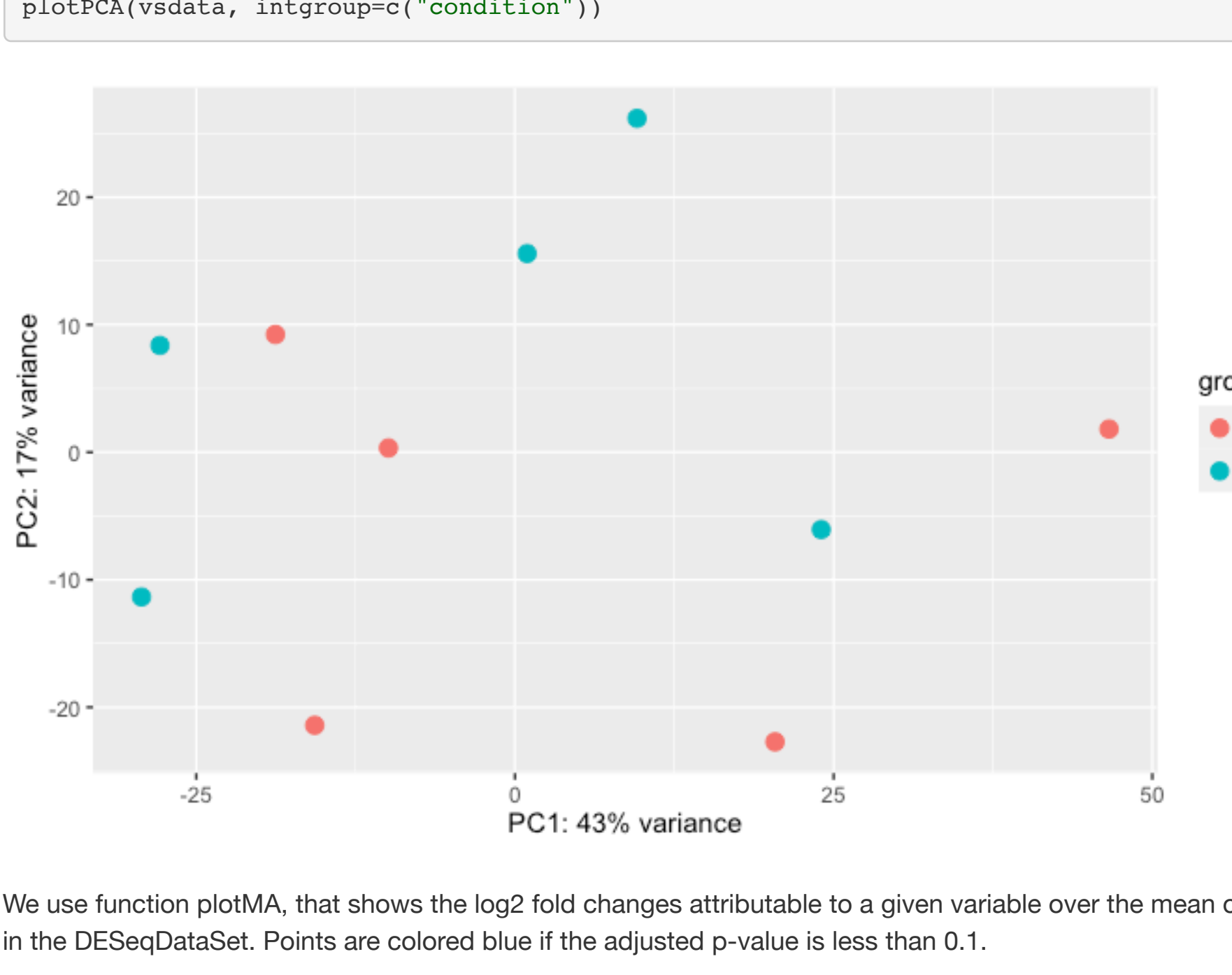
Error in lfcShrink(ddsLRab, coef = "condition_LI_vs_AL", type = "apeglm") :
  type='apeglm' requires installing the Bioconductor package 'apeglm'
```

Visual exploration of the results

We use first 100 significant genes to build a PCA plot:

Hide

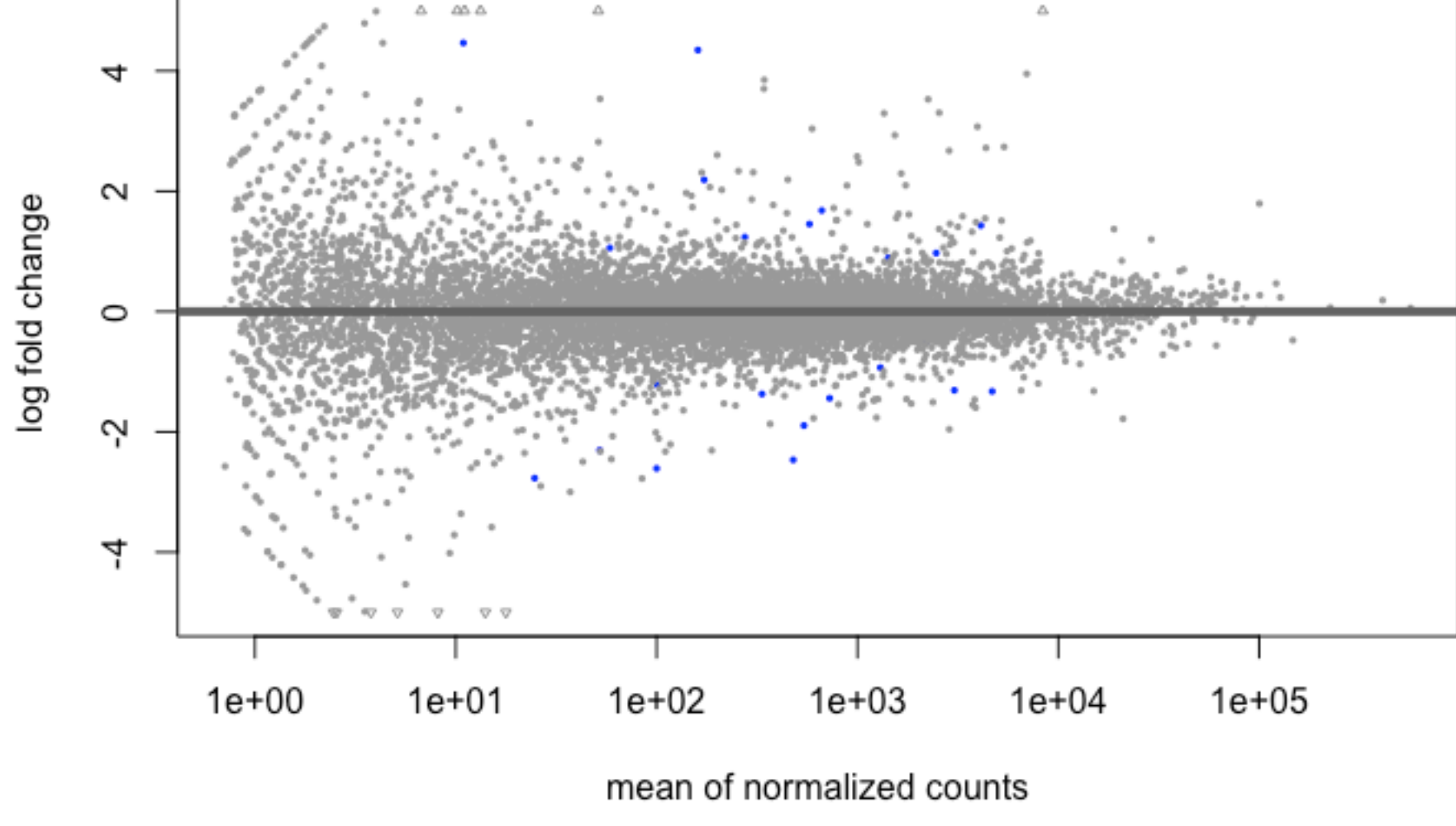
```
vsdata <- vst(ddsLRab, blind=FALSE)
plotPCA(vsdata, intgroup=c("condition"))
```



We use function plotMA, that shows the log2 fold changes attributable to a given variable over the mean of normalized counts for all the samples in the DESeqDataSet. Points are colored blue if the adjusted p-value is less than 0.1.

Hide

```
plotMA(resLRab, ylim=c(-5,5))
```



Checking significant gene candidates:

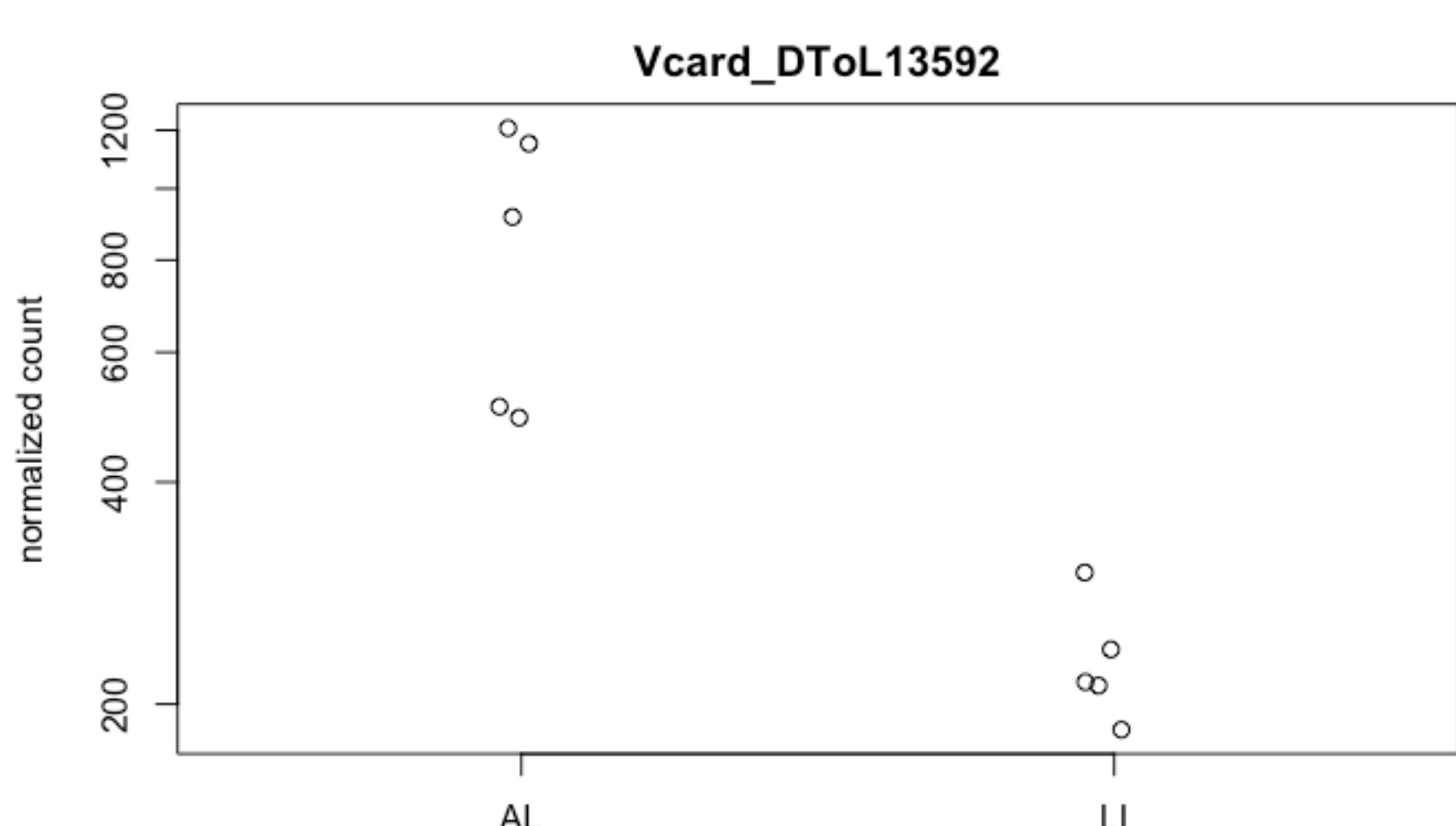
Hide

```
res <- resLRab[order(resLRab$padj),]
head(res, 10)
```

```
log2 fold change (MLE): condition LI vs AL
Wald test p-value: condition LI vs AL
DataFrame with 10 rows and 6 columns
  baseMean log2FoldChange    lfcSE      stat      pvalue      padj
  <numeric>    <numeric>    <numeric>    <numeric>    <numeric>    <numeric>
Vcard_DToL13592    541.4268    -1.89555    0.313908    -6.03855    1.55504e-09    1.56422e-05
Vcard_DToL04829     52.0770    -2.30664    0.401831    -5.74031    9.45037e-09    4.75306e-05
Vcard_DToL13539     274.9208    1.23667    0.269397    4.59051    4.42161e-06    1.48257e-02
Vcard_DToL02600     478.9490    -2.46656    0.567757    -4.34440    1.39658e-05    3.27025e-02
Vcard_DToL10357     334.9660    -1.37088    0.323629    -4.23596    2.27575e-05    3.27025e-02
Vcard_DToL13220     160.5249    4.34707    1.023610    4.24681    2.16838e-05    3.27025e-02
Vcard_DToL16018     24.6975    -2.76963    0.650293    -4.25905    2.05300e-05    3.27025e-02
Vcard_DToL01847     664.3799    1.67992    0.429905    3.90765    9.31980e-05    8.13203e-02
Vcard_DToL04245     4685.0017    -1.32700    0.334322    -3.96925    7.21005e-05    8.13203e-02
Vcard_DToL05076     100.0051    -2.60856    0.677126    -3.85240    1.16968e-04    8.13203e-02
```

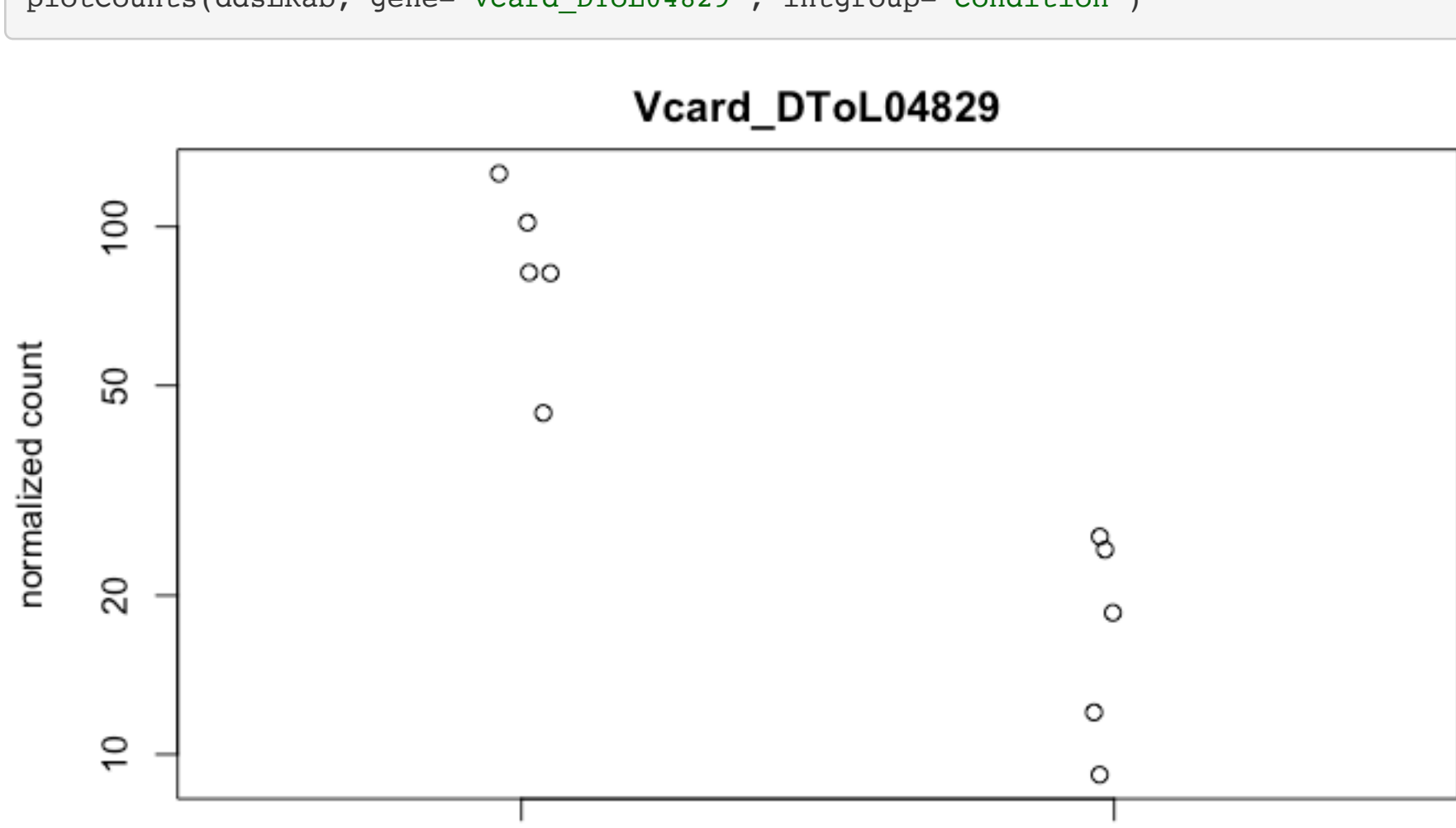
Hide

```
#Plotting these genes
plotCounts(ddsLRab, gene="Vcard_DToL13592", intgroup="condition")
```



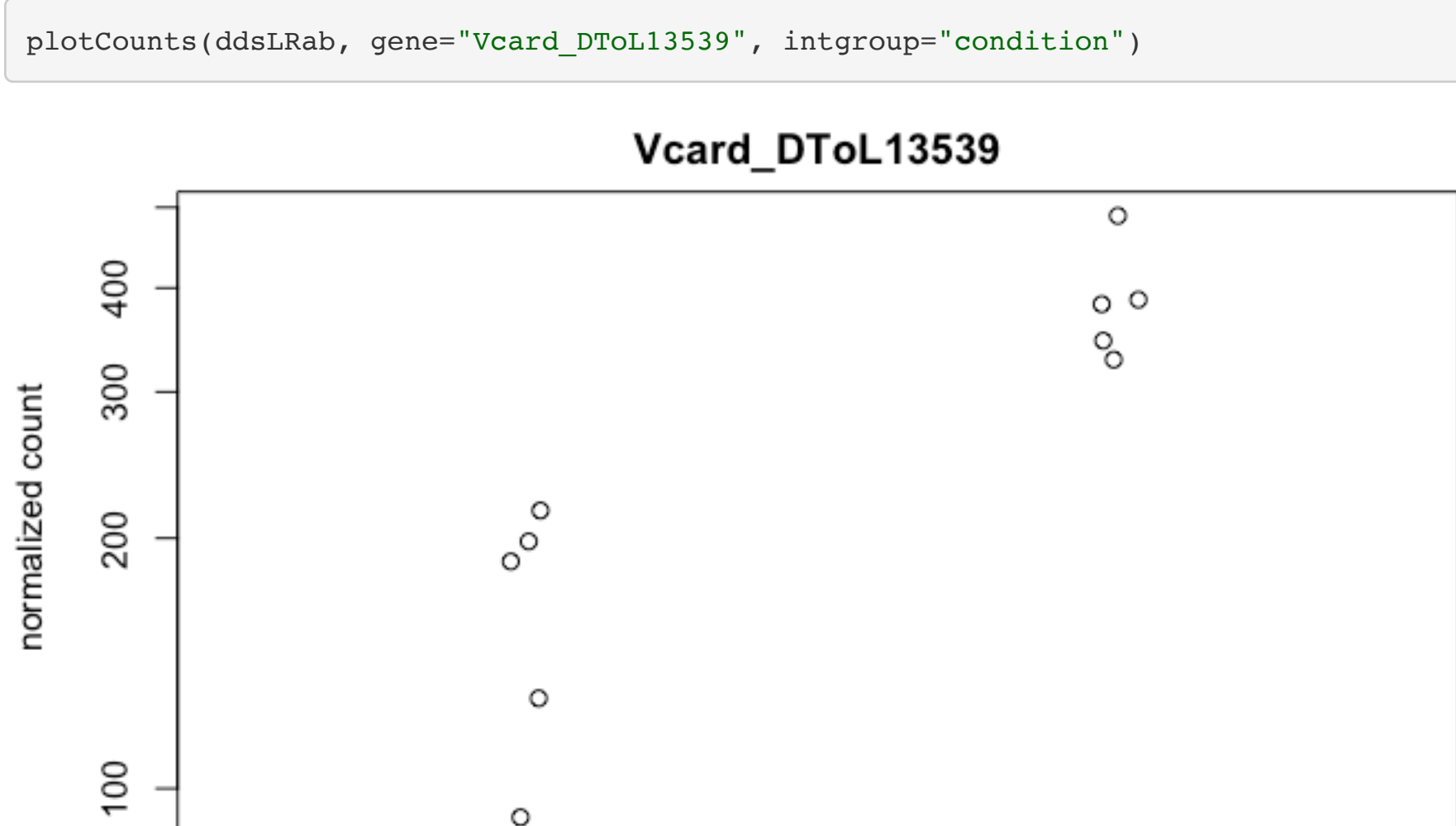
Hide

```
plotCounts(ddsLRab, gene="Vcard_DToL04829", intgroup="condition")
```



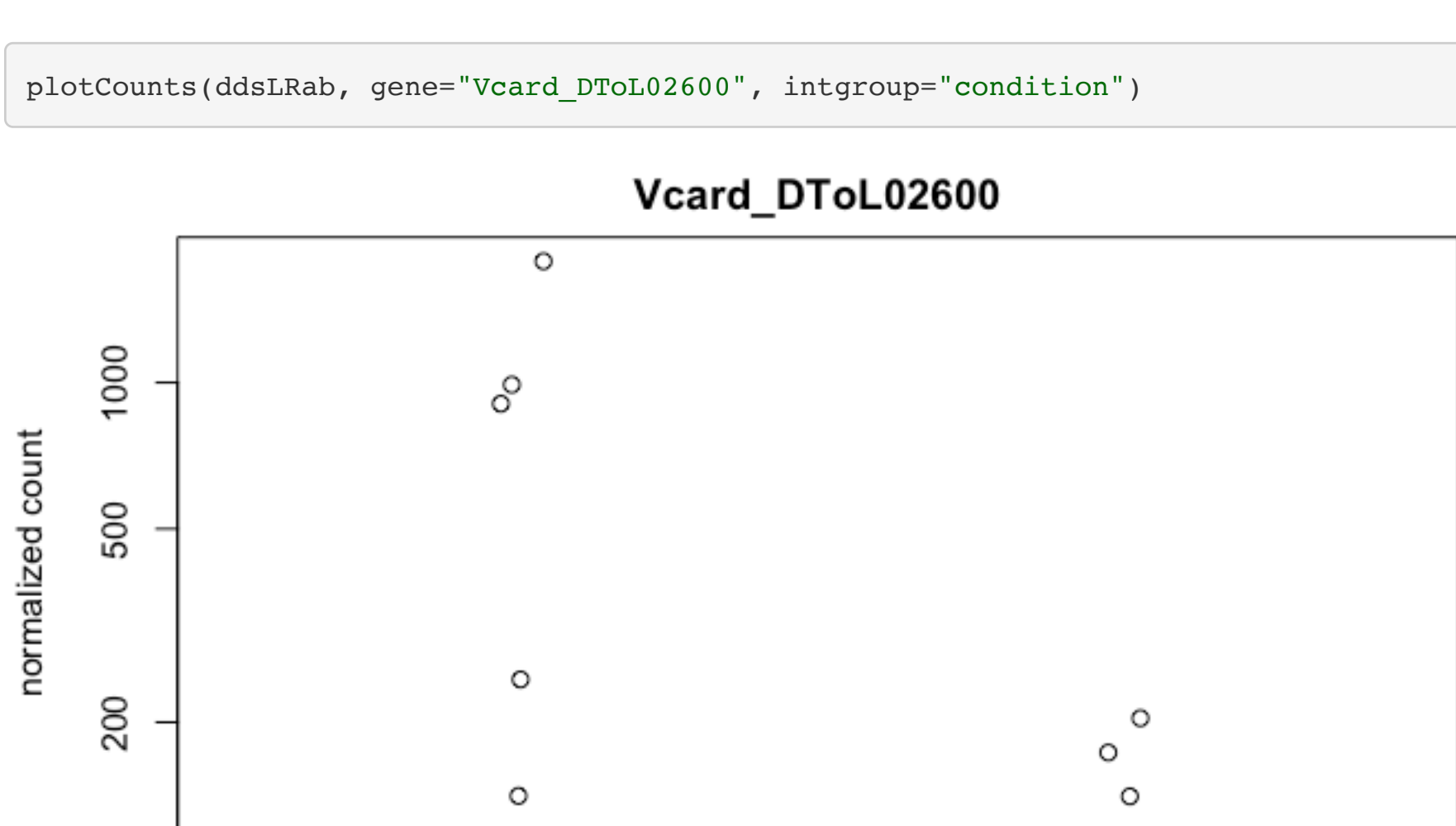
Hide

```
plotCounts(ddsLRab, gene="Vcard_DToL13539", intgroup="condition")
```



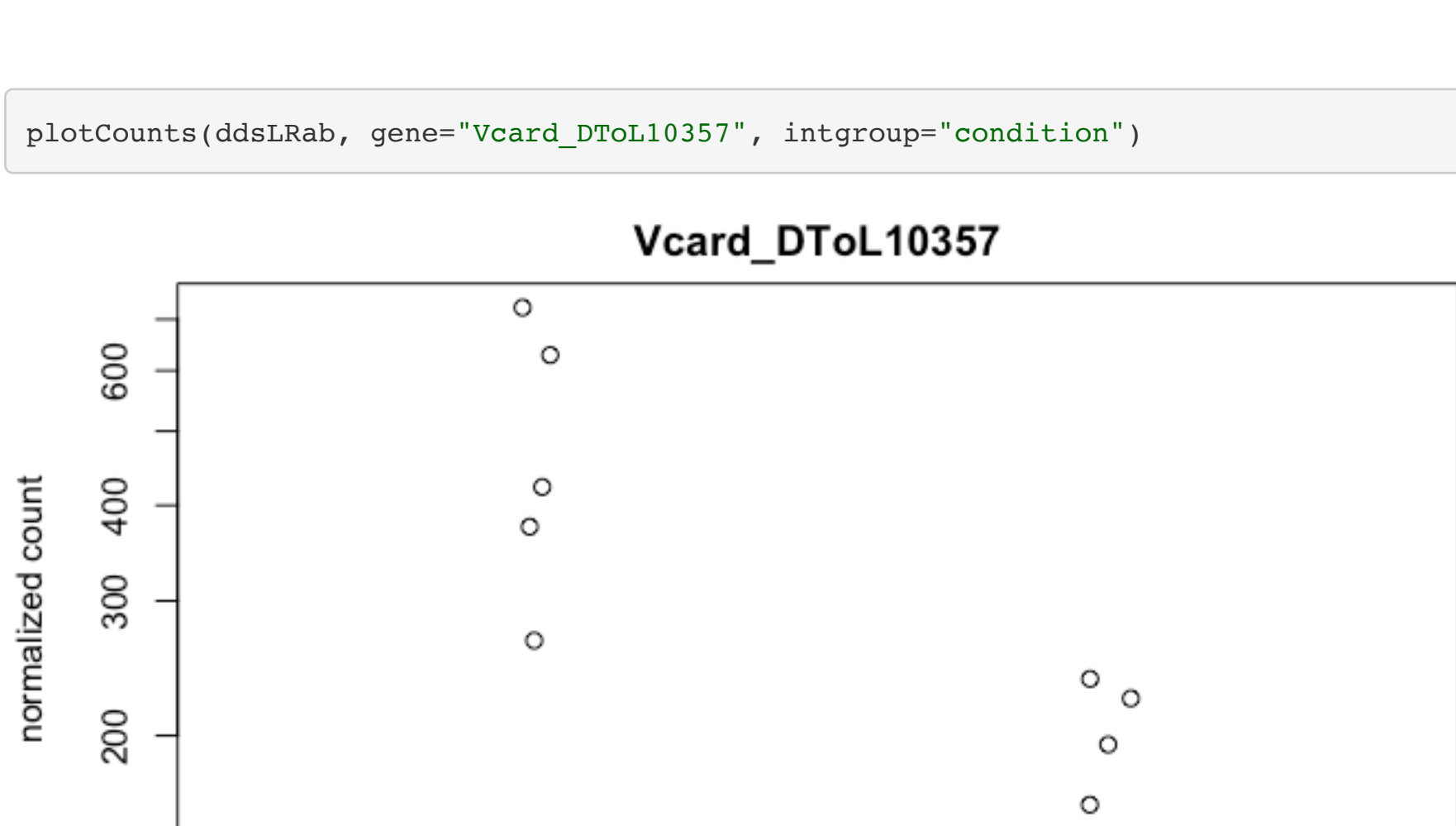
Hide

```
plotCounts(ddsLRab, gene="Vcard_DToL02600", intgroup="condition")
```



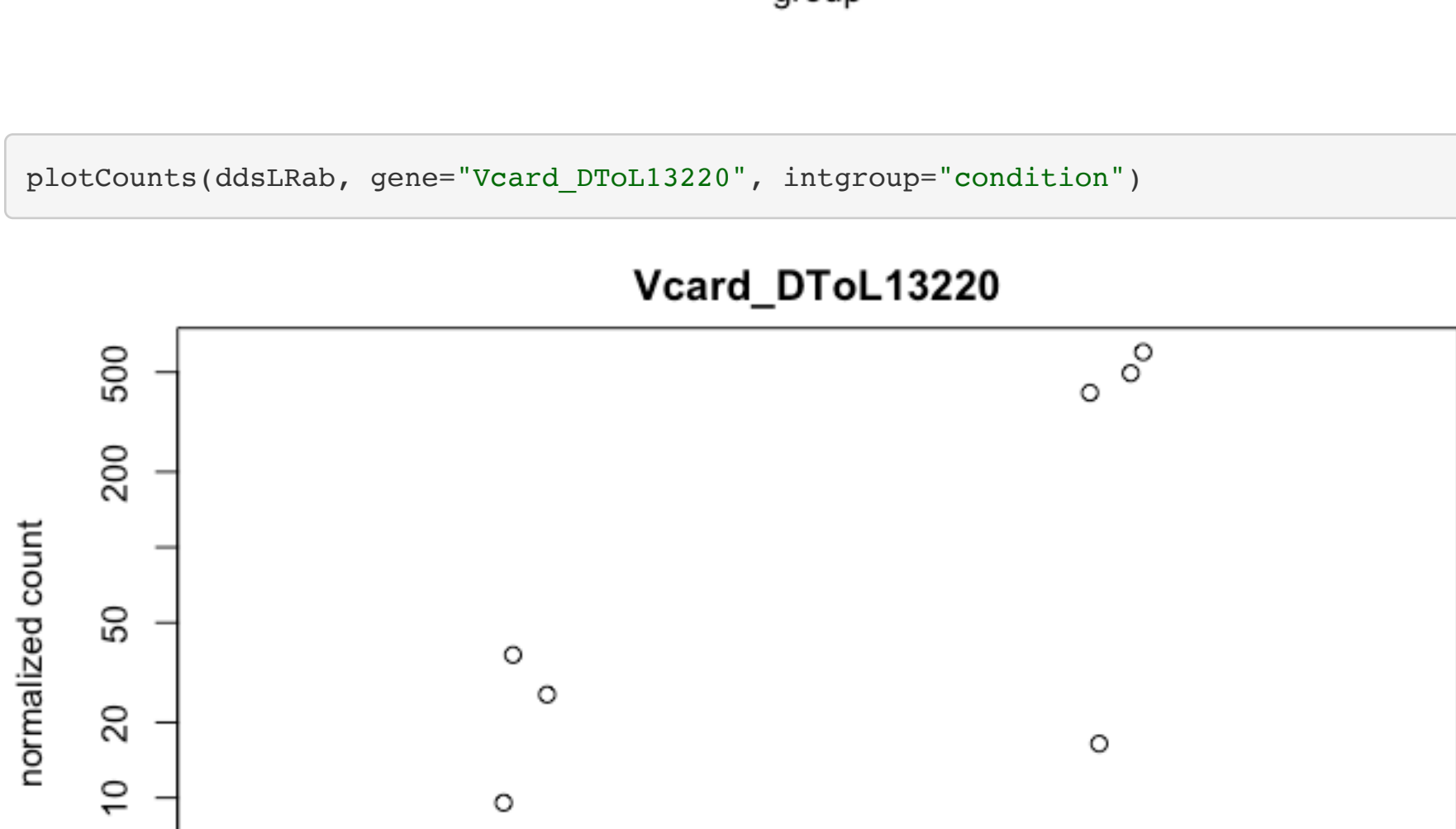
Hide

```
plotCounts(ddsLRab, gene="Vcard_DToL10357", intgroup="condition")
```



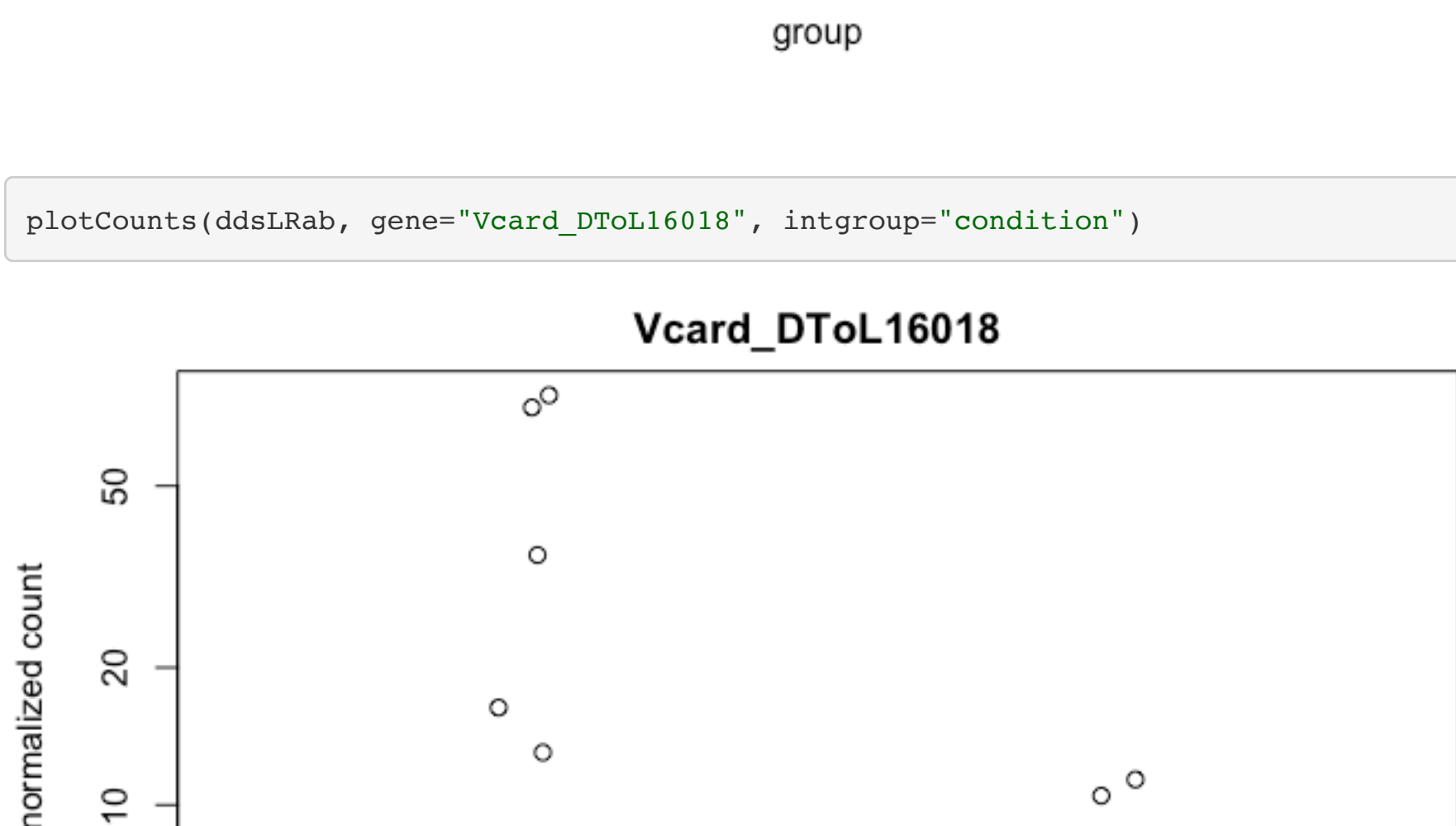
Hide

```
plotCounts(ddsLRab, gene="Vcard_DToL13220", intgroup="condition")
```



Hide

```
plotCounts(ddsLRab, gene="Vcard_DToL16018", intgroup="condition")
```



Checking functions of the differentially expressed genes

Hide

```
#Is
grep "Vcard_DToL13592"-RA "zinc-binding domain present in Lin-11 Isl-1
grep "Vcard_DToL04829"-RA "insect pheromone-binding family A10/O8-0"
grep "Vcard_DToL13539"-RA "insect pheromone-binding family A10/O8-0"
grep "Vcard_DToL02600"-RA "Scp / Tpa-1 / Aps / FR-1 / Sc7 family of extracellular domains. glipr2
grep "Vcard_DToL10357"-RA "Leucine Rich repeat chp
grep "Vcard_DToL13220"-RA "Leucine Rich repeat chp
grep "Vcard_DToL16018"-RA "Protein of unknown function (DUF1676) Osi9
```

```
Vcard_DToL13592-RA --
Vcard_DToL04829-RA "zinc-binding domain present in Lin-11 Isl-1
Vcard_DToL13539-RA "insect pheromone-binding family A10/O8-0"
Vcard_DToL02600-RA "Scp / Tpa-1 / Aps / FR-1 / Sc7 family of extracellular domains. glipr2
Vcard_DToL10357-RA "Leucine Rich repeat chp
Vcard_DToL13220-RA "Leucine Rich repeat chp
Vcard_DToL16018-RA "Protein of unknown function (DUF1676) Osi9
```