Analysis of chicken light/dark dataset. Modeling retina dataset with DESeq2

Code

get transcript IDs and gene IDs from transcriptome fasta

Hide

```
grep ">" GCF_000002315.6_GRCg6a_rna.fna > headers.txt #extract fasta headers

cat headers.txt | cut -d " " -f1 |sed 's/^>//' > tx_id.txt #extract list of transcr
ipt IDs

sed 's/^.*(//' headers.txt | sed 's/).*$//' > gene_id.txt #extract list of gene IDs
(symbols)
```

make a data frame with transcript to gene mappings

Hide

```
tx2gene<-data.frame(read.csv("tx_id.txt",header=FALSE),read.csv("gene_id.txt",heade
r=FALSE))

head(tx2gene)
```

	V1 <fctr>	V1.1 <fctr>
1	NM_001001127.1	EDNRB
2	NM_001001129.1	CYP26A1
3	NM_001001131.1	GHRL
4	NM_001001189.1	CFDP1
5	NM_001001192.1	ST3GAL5
6	NM_001001193.1	AvBD6

6 rows

get the metadata for the samples

Hide

```
sample<-read.table("~/OneDrive/chick_versace/KallistoAnalysis/data/metaData/chicken
FullData_salmon.csv", header=TRUE)
```

each library was sequenced in two lanes, and the PCA indicates that these runs are very similar add a column to identify each library so the two sequencing runs can be combined (increases statistical power)

Hide

```
sample$library<-paste0(sample$animal,sample$tissue,sample$side)

sample
```

sample <fctr>	animal <fctr>	condition <fctr>	tissue <fctr>	side <fctr>	lane <fctr>	library <chr>
WTCHG_323579_209	07A	DARK	TEL	LEFT	run1lane3	07ATELLEFT
WTCHG_323579_210	07A	DARK	TEL	RIGHT	run1lane3	07ATELRIGHT
WTCHG_323579_221	22A	LIGHT6H	TEL	LEFT	run1lane3	22ATELLEFT
WTCHG_323579_222	22A	LIGHT6H	TEL	RIGHT	run1lane3	22ATELRIGHT
WTCHG_323579_223	22A	LIGHT6H	RET	LEFT	run1lane3	22ARETLEFT
WTCHG_323579_224	22A	LIGHT6H	RET	RIGHT	run1lane3	22ARETRIGHT
WTCHG_323579_241	42A	LIGHT24H	TEL	LEFT	run1lane3	42ATELLEFT
WTCHG_323579_242	42A	LIGHT24H	TEL	RIGHT	run1lane3	42ATELRIGHT
WTCHG_323579_243	42A	LIGHT24H	RET	LEFT	run1lane3	42ARETLEFT
WTCHG_323579_244	42A	LIGHT24H	RET	RIGHT	run1lane3	42ARETRIGHT

1-10 of 120 rows Previous **1** 2 3 4 5 6 … 12 Next

In order to test for a difference between right and left sides, need to include individual animal id as a factor. This presents a problem because individual animals can only be in one of the treatment groups, so this produces an error that the model is not full rank.

A workaround is provided in the DESeq2 vignette (see “Group-specific condition effects, individuals nested within groups”); animal.n is a factor that distinguishes the individuals nested within a group, to produce a full rank model

Hide

```
ret_meta<-sample[sample$tissue=="RET",]%>%arrange(condition)

ret_meta
```

sample <fctr>	animal <fctr>	condition <fctr>	tissue <fctr>	side <fctr>	lane <fctr>	library <chr>
WTCHG_323579_275	02A	DARK	RET	LEFT	run1lane3	02ARETLEFT
WTCHG_323579_276	02A	DARK	RET	RIGHT	run1lane3	02ARETRIGHT
WTCHG_323579_279	03A	DARK	RET	LEFT	run1lane3	03ARETLEFT
WTCHG_323579_280	03A	DARK	RET	RIGHT	run1lane3	03ARETRIGHT
WTCHG_323580_211	07A	DARK	RET	LEFT	run1lane4	07ARETLEFT
WTCHG_323580_212	07A	DARK	RET	RIGHT	run1lane4	07ARETRIGHT
WTCHG_323580_215	10A	DARK	RET	LEFT	run1lane4	10ARETLEFT

sample <fctr>	animal <fctr>	condition <fctr>	tissue <fctr>	side <fctr>	lane <fctr>	library <chr>
WTCHG_323580_216	10A	DARK	RET	RIGHT	run1lane4	10ARETRIGHT
WTCHG_323580_219	14A	DARK	RET	LEFT	run1lane4	14ARETLEFT
WTCHG_323580_220	14A	DARK	RET	RIGHT	run1lane4	14ARETRIGHT

1-10 of 60 rows Previous **1** 2 3 4 5 6 Next

Hide

```
ret_meta$animal.n<-factor(rep(rep(1:5,each=2),3))

ret_meta
```

sample <fctr>	animal <fctr>	condition <fctr>	tissue <fctr>	side <fctr>	lane <fctr>	library <chr>	animal.n <fctr>
WTCHG_323579_275	02A	DARK	RET	LEFT	run1lane3	02ARETLEFT	1
WTCHG_323579_276	02A	DARK	RET	RIGHT	run1lane3	02ARETRIGHT	1
WTCHG_323579_279	03A	DARK	RET	LEFT	run1lane3	03ARETLEFT	2
WTCHG_323579_280	03A	DARK	RET	RIGHT	run1lane3	03ARETRIGHT	2
WTCHG_323580_211	07A	DARK	RET	LEFT	run1lane4	07ARETLEFT	3
WTCHG_323580_212	07A	DARK	RET	RIGHT	run1lane4	07ARETRIGHT	3
WTCHG_323580_215	10A	DARK	RET	LEFT	run1lane4	10ARETLEFT	4
WTCHG_323580_216	10A	DARK	RET	RIGHT	run1lane4	10ARETRIGHT	4
WTCHG_323580_219	14A	DARK	RET	LEFT	run1lane4	14ARETLEFT	5
WTCHG_323580_220	14A	DARK	RET	RIGHT	run1lane4	14ARETRIGHT	5

1-10 of 60 rows Previous **1** 2 3 4 5 6 Next

import data for retina samples, dropping samples identified as outliers by St. Andrews (sample 50A)

Hide

```
directory<-"~/OneDrive/chick_versace/quants"

files <- file.path(directory, list.files(directory), "quant.sf")



d_ret <- tximport(files[sample$tissue=="RET"&sample$animal!="50A"], type="salmon",
dropInfReps=TRUE, tx2gene=tx2gene)
```

```
reading in files with read_tsv

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56

summarizing abundance

summarizing counts

summarizing length
```

create the DESeq dataset with the retina data, dropping samples identified as outliers by St. Andrews

Hide

```
ddsTxi_ret <- DESeqDataSetFromTximport(d_ret,

                                       colData = filter(ret_meta,animal!="50A"),

                                       design = ~1)
```

```
using counts and average transcript lengths from tximport
```

Hide

```
#don't specify the model yet

#design = ~ condition + condition:animal.n + condition:side



##old code for dropping samples (also must drop corresponding metadata)

# ddsTxi <- DESeqDataSetFromTximport(d,

#                                    colData = samples[-2,],

#                                    design = ~ condition)
```

view the metadata columns

Hide

```
colData(ddsTxi_ret)[1:28,]
```

```
DataFrame with 28 rows and 8 columns

              sample   animal condition   tissue     side      lane     library ani
mal.n

            <factor> <factor>  <factor> <factor> <factor>  <factor> <character> <fa
ctor>

1   WTCHG_323579_275      02A      DARK      RET     LEFT run1lane3  02ARETLEFT
1

2   WTCHG_323579_276      02A      DARK      RET    RIGHT run1lane3 02ARETRIGHT
1

3   WTCHG_323579_279      03A      DARK      RET     LEFT run1lane3  03ARETLEFT
2

4   WTCHG_323579_280      03A      DARK      RET    RIGHT run1lane3 03ARETRIGHT
2

5   WTCHG_323580_211      07A      DARK      RET     LEFT run1lane4  07ARETLEFT
3

...              ...      ...       ...      ...      ...       ...         ...
...

24  WTCHG_323579_248      44A  LIGHT24H      RET    RIGHT run1lane3 44ARETRIGHT
2

25  WTCHG_323580_251      49A  LIGHT24H      RET     LEFT run1lane4  49ARETLEFT
3

26  WTCHG_323580_252      49A  LIGHT24H      RET    RIGHT run1lane4 49ARETRIGHT
3

27  WTCHG_323580_259      54A  LIGHT24H      RET     LEFT run1lane4  54ARETLEFT
5

28  WTCHG_323580_301      54A  LIGHT24H      RET    RIGHT run1lane4 54ARETRIGHT
5
```

collapse technical replicates

Hide

```
ddsTxi_ret<-collapseReplicates(ddsTxi_ret,ddsTxi_ret$library,ddsTxi_ret$lane)

colData(ddsTxi_ret)
```

```
DataFrame with 28 rows and 9 columns

                     sample   animal condition   tissue     side      lane     lib
rary animal.n

                   <factor> <factor>  <factor> <factor> <factor>  <factor> <charac
ter> <factor>

02ARETLEFT  WTCHG_323579_275      02A      DARK      RET     LEFT run1lane3  02ARET
LEFT        1

02ARETRIGHT WTCHG_323579_276      02A      DARK      RET    RIGHT run1lane3 02ARETR
IGHT        1

03ARETLEFT  WTCHG_323579_279      03A      DARK      RET     LEFT run1lane3  03ARET
LEFT        2

03ARETRIGHT WTCHG_323579_280      03A      DARK      RET    RIGHT run1lane3 03ARETR
IGHT        2

07ARETLEFT  WTCHG_323580_211      07A      DARK      RET     LEFT run1lane4  07ARET
LEFT        3

...                     ...      ...       ...      ...      ...       ...
...      ...

44ARETRIGHT WTCHG_323579_248      44A  LIGHT24H      RET    RIGHT run1lane3 44ARETR
IGHT        2

49ARETLEFT  WTCHG_323580_251      49A  LIGHT24H      RET     LEFT run1lane4  49ARET
LEFT        3

49ARETRIGHT WTCHG_323580_252      49A  LIGHT24H      RET    RIGHT run1lane4 49ARETR
IGHT        3

54ARETLEFT  WTCHG_323580_259      54A  LIGHT24H      RET     LEFT run1lane4  54ARET
LEFT        5

54ARETRIGHT WTCHG_323580_301      54A  LIGHT24H      RET    RIGHT run1lane4 54ARETR
IGHT        5

                 runsCollapsed

                   <character>

02ARETLEFT  run1lane3,run2lane5

02ARETRIGHT run1lane3,run2lane5

03ARETLEFT  run1lane3,run2lane5

03ARETRIGHT run1lane3,run2lane5

07ARETLEFT  run1lane4,run2lane6
```

```
...                          ...

44ARETRIGHT run1lane3,run2lane5

49ARETLEFT  run1lane4,run2lane6

49ARETRIGHT run1lane4,run2lane6

54ARETLEFT  run1lane4,run2lane6

54ARETRIGHT run1lane4,run2lane6
```

eliminate genes with very low counts

Hide

```
keep <- rowSums(counts(ddsTxi_ret)) >= 10

ddsTxi_ret <- ddsTxi_ret[keep,]
```

make the model matrix, then drop the all zero column corresponding to the dropped sample

Hide

```
m1<-model.matrix(~ condition + condition:animal.n + condition:side,colData(ddsTxi_r
et)) #construct the model matrix

all.zero <- apply(m1, 2, function(x) all(x==0))   #find the column which is all zer
o

idx <- which(all.zero)

m1 <- m1[,-idx]   #remove the column which is all zero from the model matrix
```

fit and test the model, specifying the corrected model matrix

Hide

```
dds<-DESeq(ddsTxi_ret, full= m1)
```

```
using supplied model matrix

estimating size factors

using 'avgTxLength' from assays(dds), correcting for library size

estimating dispersions

gene-wise dispersion estimates

mean-dispersion relationship

final dispersion estimates

fitting model and testing
```

Hide

```
resultsNames(dds)
```

```
 [1] "Intercept"                   "conditionLIGHT24H"           "conditionLIGHT6H"

 [4] "conditionDARK.animal.n2"     "conditionLIGHT24H.animal.n2" "conditionLIGHT6H.
animal.n2"

 [7] "conditionDARK.animal.n3"     "conditionLIGHT24H.animal.n3" "conditionLIGHT6H.
animal.n3"

[10] "conditionDARK.animal.n4"     "conditionLIGHT6H.animal.n4"  "conditionDARK.ani
mal.n5"

[13] "conditionLIGHT24H.animal.n5" "conditionLIGHT6H.animal.n5"  "conditionDARK.sid
eRIGHT"

[16] "conditionLIGHT24H.sideRIGHT" "conditionLIGHT6H.sideRIGHT"
```

Hide

```
res <- results(dds, name="conditionDARK.sideRIGHT",alpha=.2)

summary(res)
```

```
out of 19360 with nonzero total read count

adjusted p-value < 0.2

LFC > 0 (up)       : 0, 0%

LFC < 0 (down)     : 0, 0%

outliers [1]       : 0, 0%

low counts [2]     : 0, 0%

(mean count < 0)

[1] see 'cooksCutoff' argument of ?results

[2] see 'independentFiltering' argument of ?results
```

Hide

```
res <- results(dds, name="conditionLIGHT24H.sideRIGHT",alpha=.2)

summary(res)
```

```
out of 19360 with nonzero total read count

adjusted p-value < 0.2

LFC > 0 (up)       : 1, 0.0052%

LFC < 0 (down)     : 0, 0%

outliers [1]       : 0, 0%

low counts [2]     : 0, 0%

(mean count < 0)

[1] see 'cooksCutoff' argument of ?results

[2] see 'independentFiltering' argument of ?results
```

Hide

```
res %>% as.data.frame() %>% rownames_to_column(var = "gene") %>% filter(padj<.2) %
>% arrange(padj)
```

ge…	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>

ge… <chr>	baseMean <dbl>	log2FoldChange <dbl>	lfcSE <dbl>	stat <dbl>	pvalue <dbl>	padj <dbl>
OGT	15191.37	0.1645068	0.0368204	4.467818	7.90215e-06	0.1529856

1 row

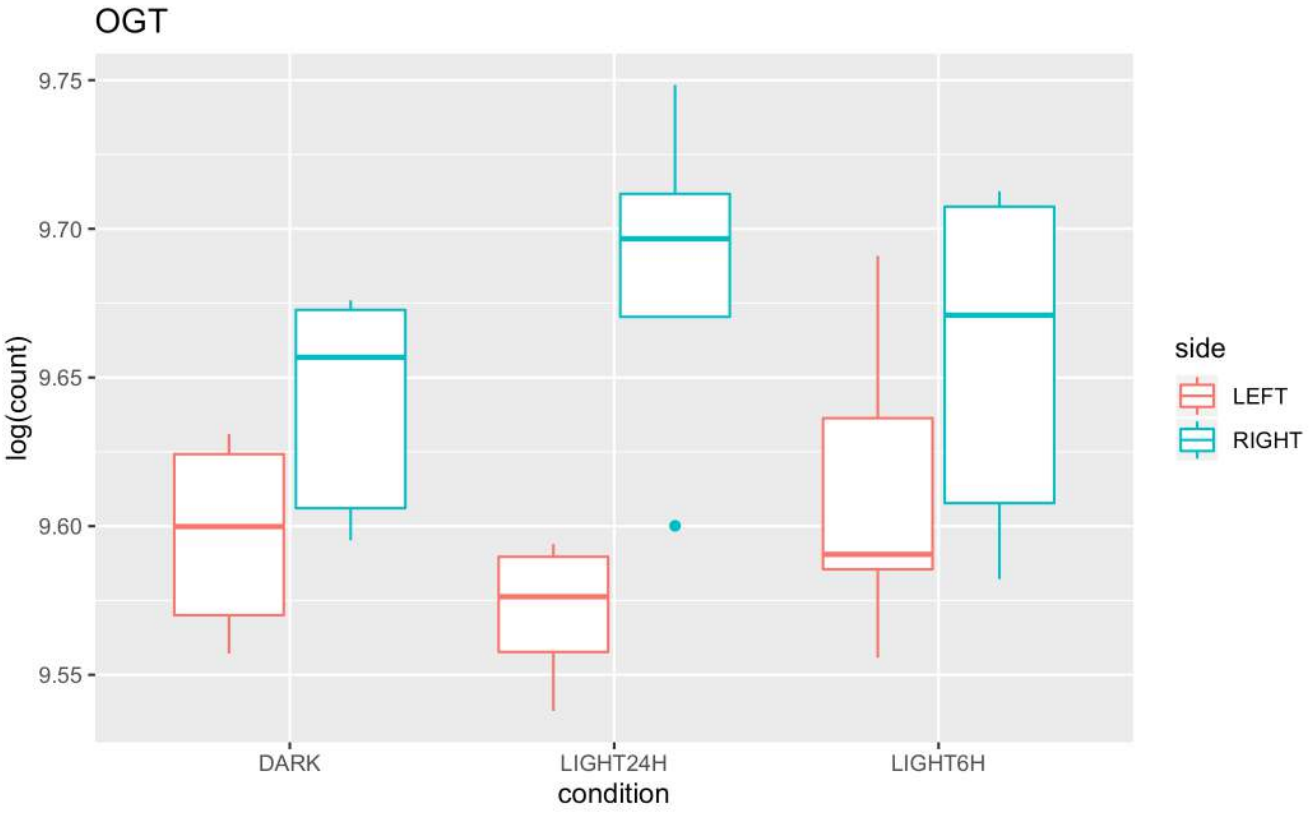
Hide

```
dat=plotCounts(dds,gene="OGT",intgroup = c("condition","side","animal"),returnData=
TRUE)

ggplot(dat,aes(condition,log(count),colour=side))+

  geom_boxplot()+

  ggtitle("OGT")
```



Hide

```
res <- results(dds, name="conditionLIGHT6H.sideRIGHT",alpha=.2)

summary(res)
```

```
out of 19360 with nonzero total read count

adjusted p-value < 0.2

LFC > 0 (up)       : 9, 0.046%

LFC < 0 (down)     : 145, 0.75%

outliers [1]       : 0, 0%

low counts [2]     : 2252, 12%

(mean count < 2)

[1] see 'cooksCutoff' argument of ?results

[2] see 'independentFiltering' argument of ?results
```

Hide

```
res %>% as.data.frame() %>% rownames_to_column(var = "gene") %>% filter(padj<.05) %
>% arrange(padj)
```

gene <chr>	baseMean <dbl>	log2FoldChange <dbl>	lfcSE <dbl>	stat <dbl>	pvalue <dbl>	
HBM	50.158094	-3.0933635	0.54712894	-5.653811	1.569290e-08	0.00026
APCDD1	180.206098	-1.2724214	0.24418279	-5.210938	1.878881e-07	0.00139
SLC28A3	29.082383	-3.2941912	0.63826248	-5.161186	2.453906e-07	0.00139
ANGPT2	2022.841982	-0.4106739	0.08353712	-4.916066	8.830077e-07	0.00205
COL1A1	70.276910	-2.0429784	0.41770844	-4.890920	1.003660e-06	0.00205
LOC423752	63.809341	-2.1334677	0.42885104	-4.974846	6.529968e-07	0.00205
MOCS2	218.877264	-1.4807400	0.30368800	-4.875859	1.083358e-06	0.00205
PDGFRB	50.483806	-2.1544075	0.43625041	-4.938465	7.873978e-07	0.00205
SSPN	18.888788	-2.7269900	0.54531667	-5.000746	5.710899e-07	0.00205
ADGRL4	33.559819	-2.3151860	0.48213610	-4.801935	1.571401e-06	0.00268

1-10 of 54 rows Previous **1** 2 3 4 5 6 Next

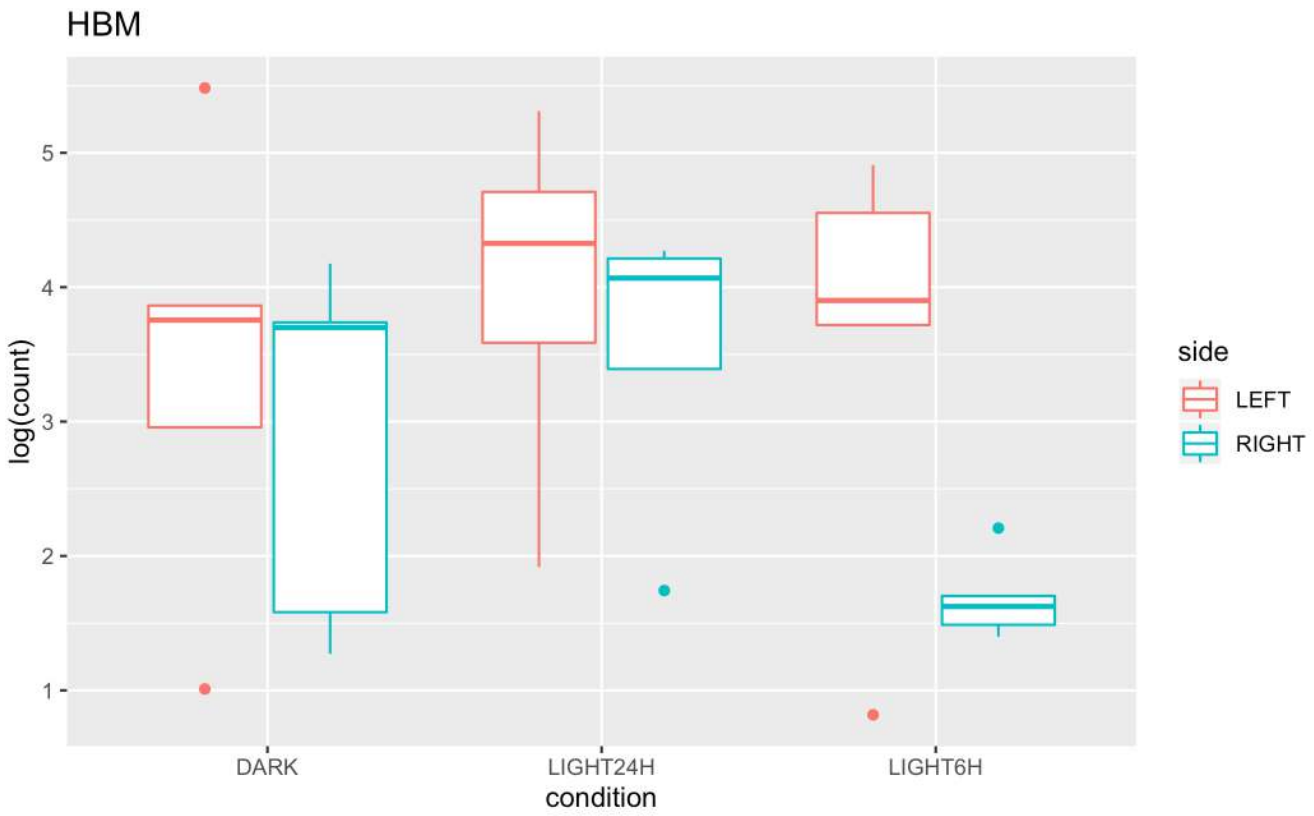
Hide

```
dat=plotCounts(dds,gene="HBM",intgroup = c("condition","side","animal"),returnData=
TRUE)

ggplot(dat,aes(condition,log(count),colour=side))+

  geom_boxplot()+

  ggtitle("HBM")
```



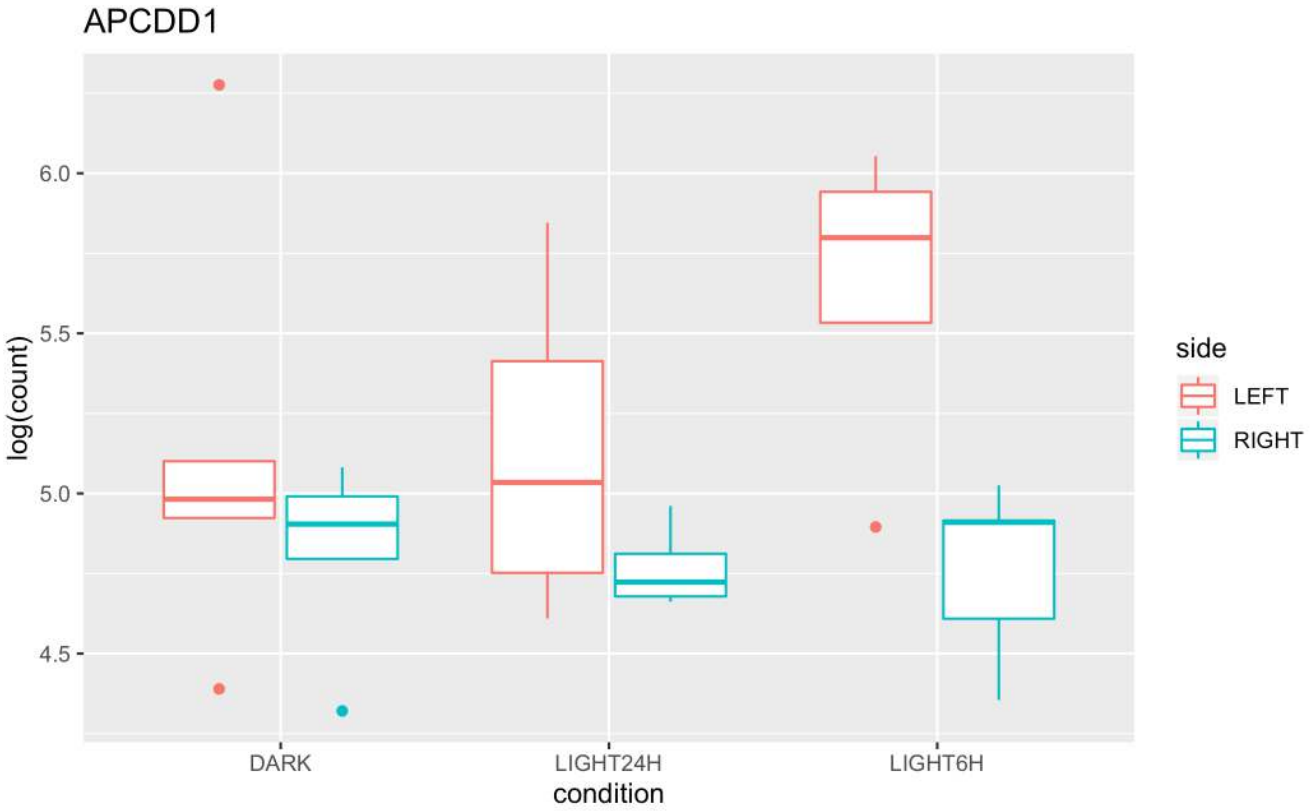
Hide

```
dat=plotCounts(dds,gene="APCDD1",intgroup = c("condition","side","animal"),returnDa
ta=TRUE)

ggplot(dat,aes(condition,log(count),colour=side))+

  geom_boxplot()+

  ggtitle("APCDD1")
```



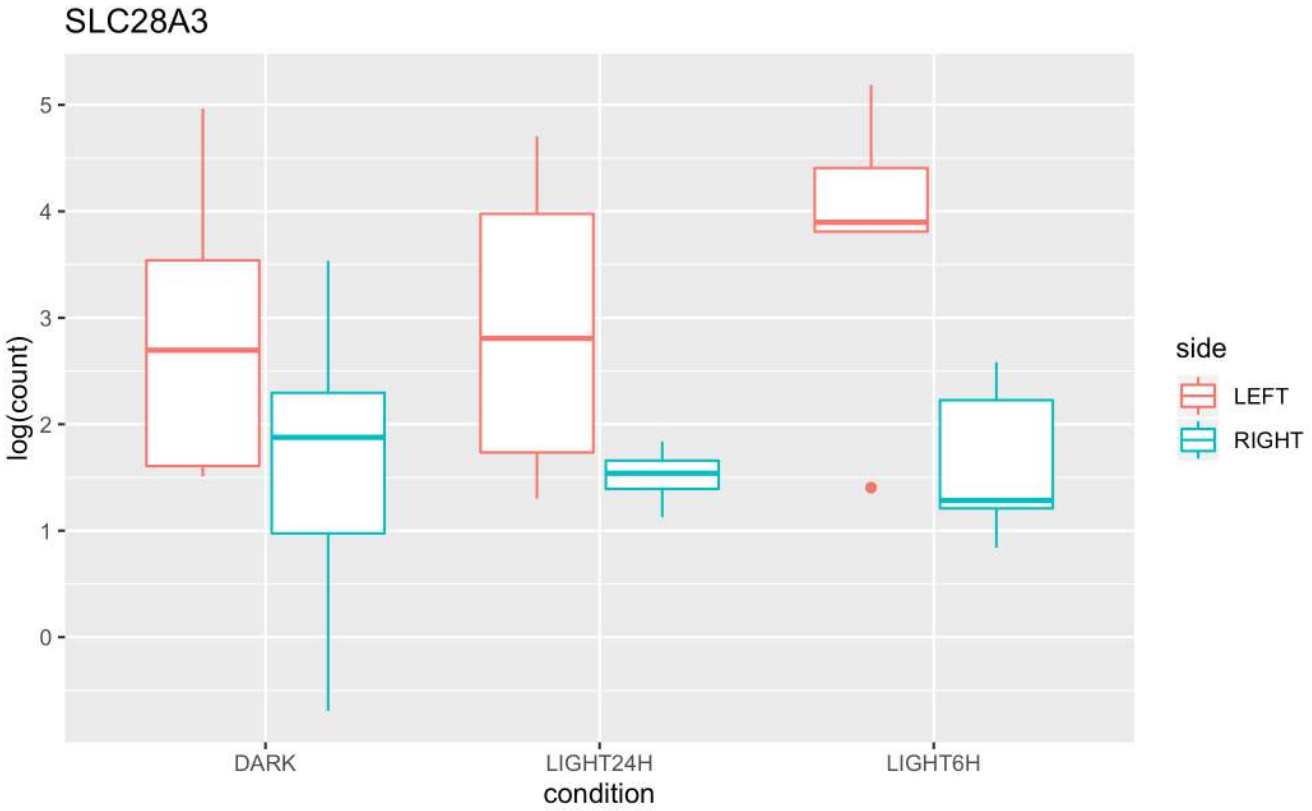
Hide

```
dat=plotCounts(dds,gene="SLC28A3",intgroup = c("condition","side","animal"),returnD
ata=TRUE)

ggplot(dat,aes(condition,log(count),colour=side))+

  geom_boxplot()+

  ggtitle("SLC28A3")
```



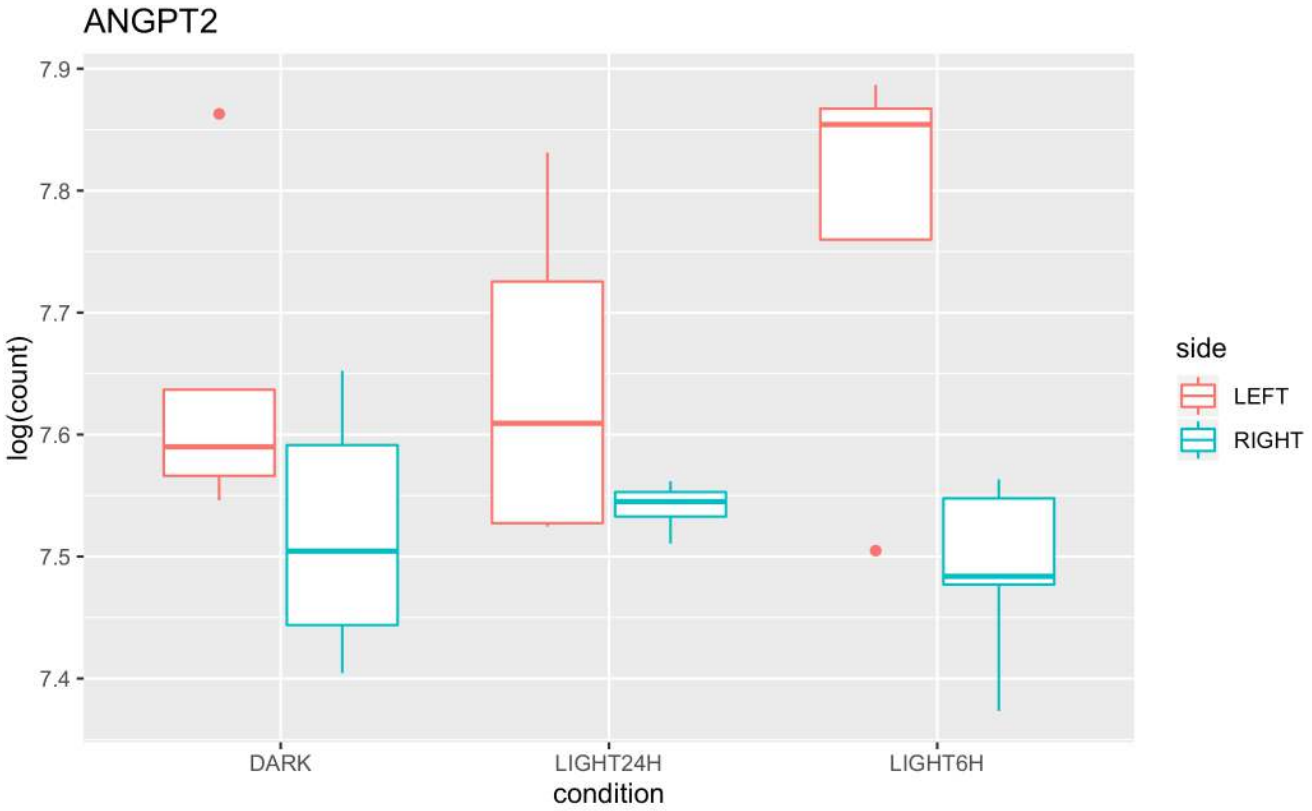
Hide

```
dat=plotCounts(dds,gene="ANGPT2",intgroup = c("condition","side","animal"),returnDa
ta=TRUE)

ggplot(dat,aes(condition,log(count),colour=side))+

  geom_boxplot()+

  ggtitle("ANGPT2")
```



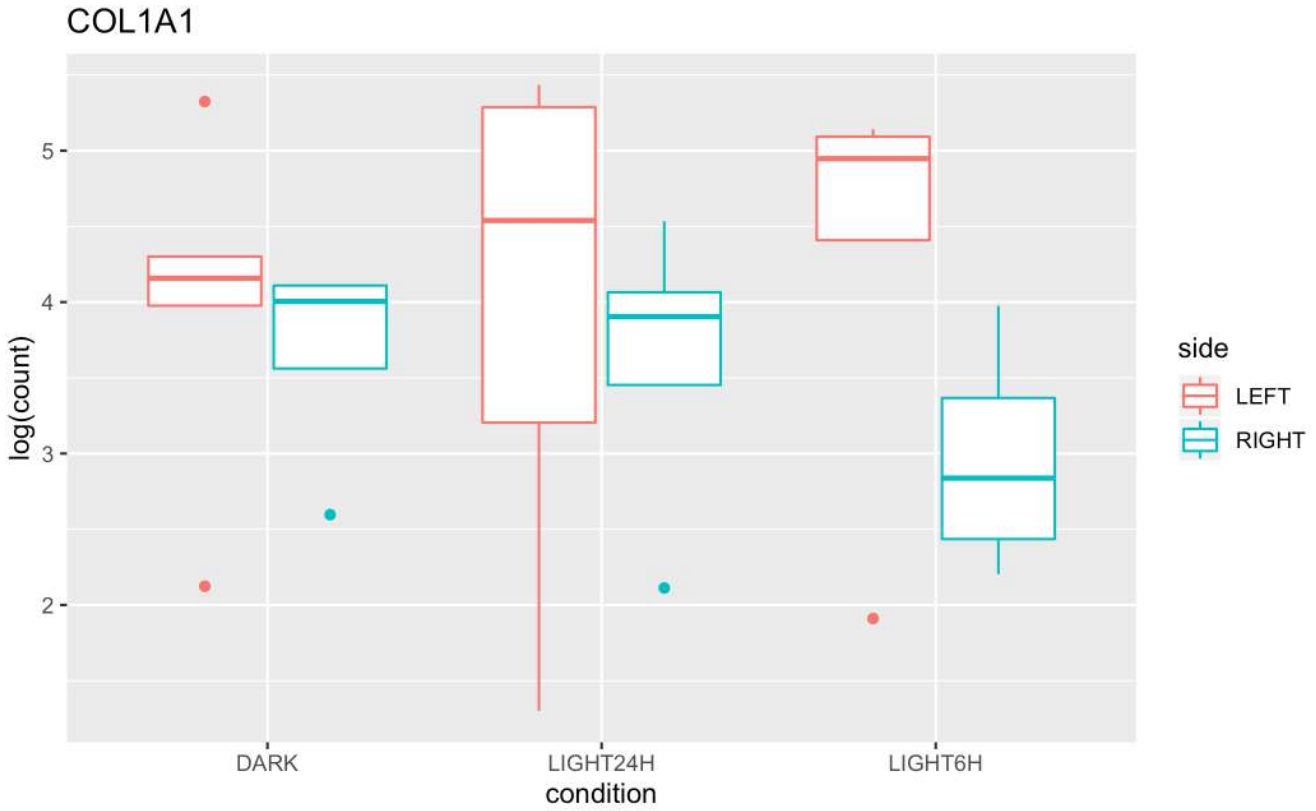
Hide

```
dat=plotCounts(dds,gene="COL1A1",intgroup = c("condition","side","animal"),returnDa
ta=TRUE)

ggplot(dat,aes(condition,log(count),colour=side))+

  geom_boxplot()+

  ggtitle("COL1A1")
```



Check plots for a few genes that were tested by qPCR

FXYD6 CHRDL1 RGR

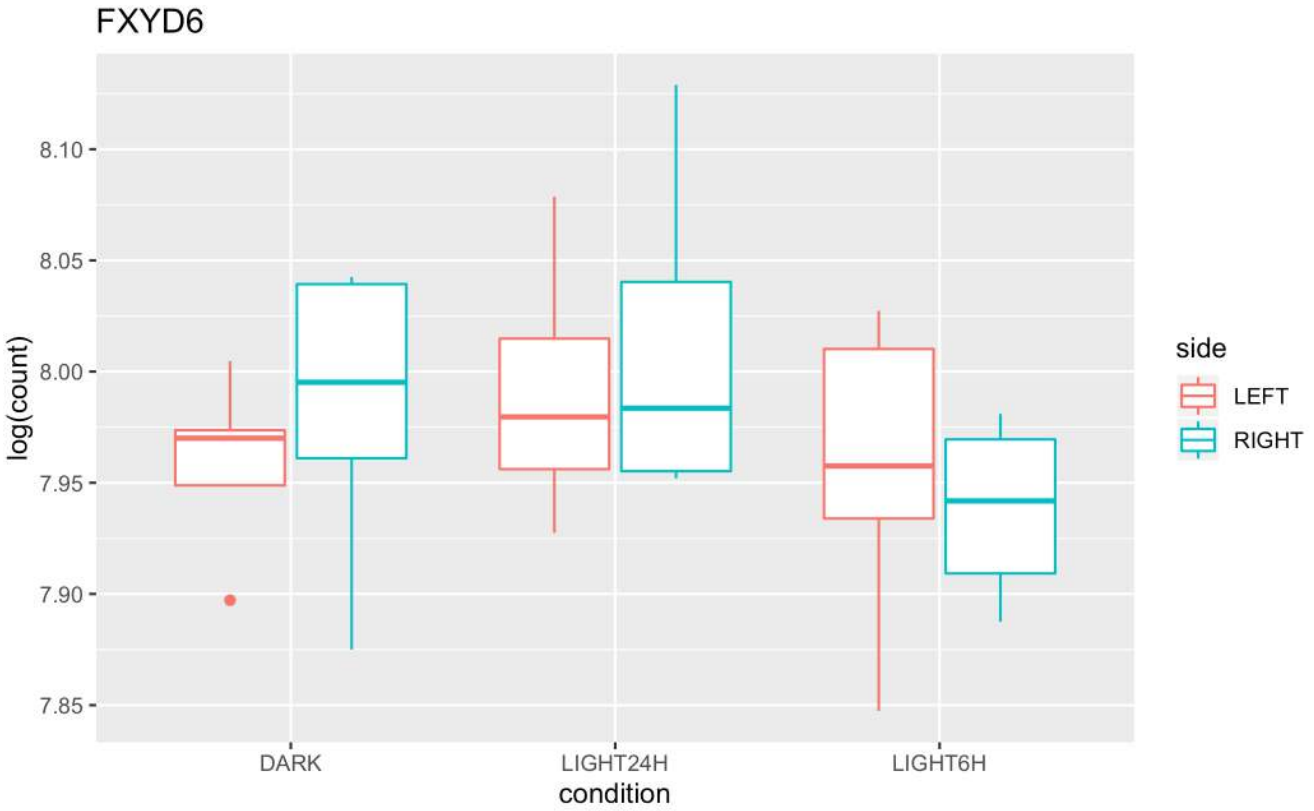
Hide

```
dat=plotCounts(dds,gene="FXYD6",intgroup = c("condition","side","animal"),returnDat
a=TRUE)

ggplot(dat,aes(condition,log(count),colour=side))+

  geom_boxplot()+

  ggtitle("FXYD6")
```



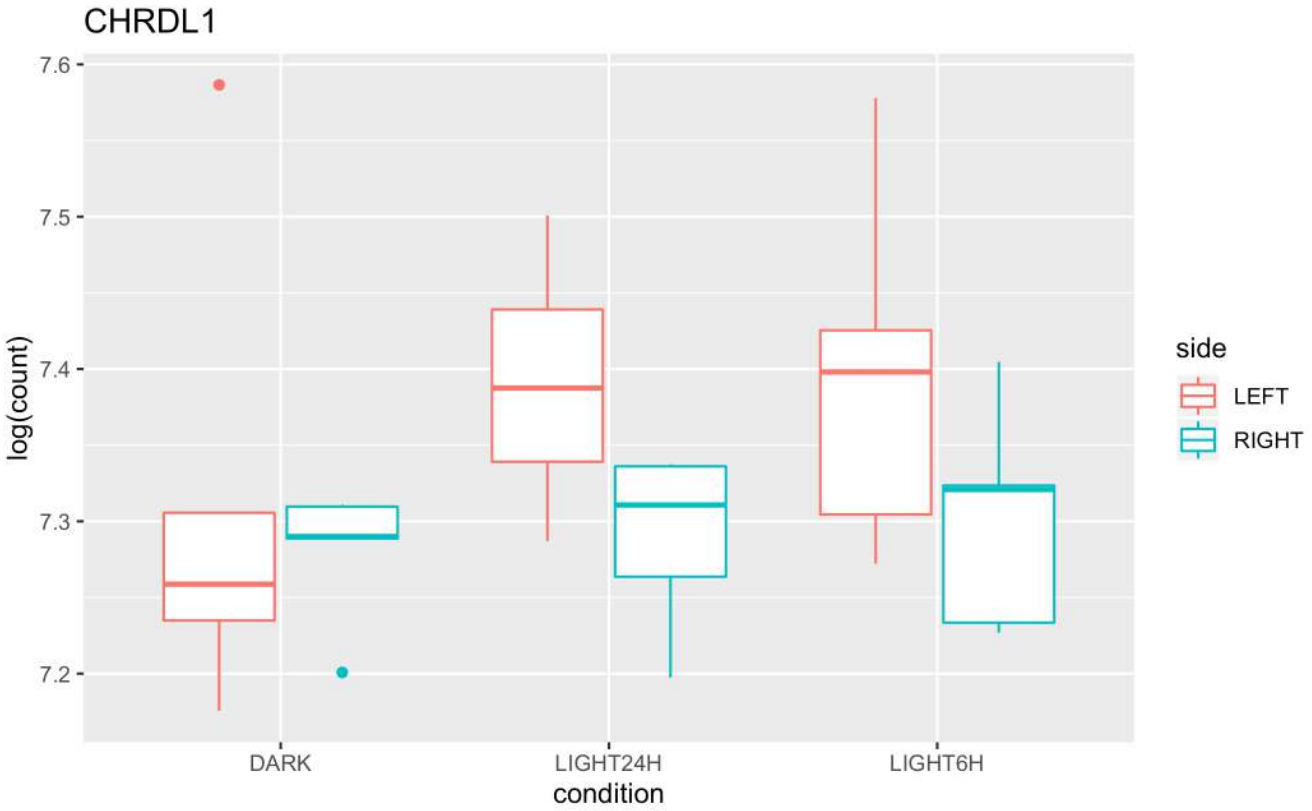
Hide

```
dat=plotCounts(dds,gene="CHRDL1",intgroup = c("condition","side","animal"),returnDa
ta=TRUE)

ggplot(dat,aes(condition,log(count),colour=side))+

  geom_boxplot()+

  ggtitle("CHRDL1")
```



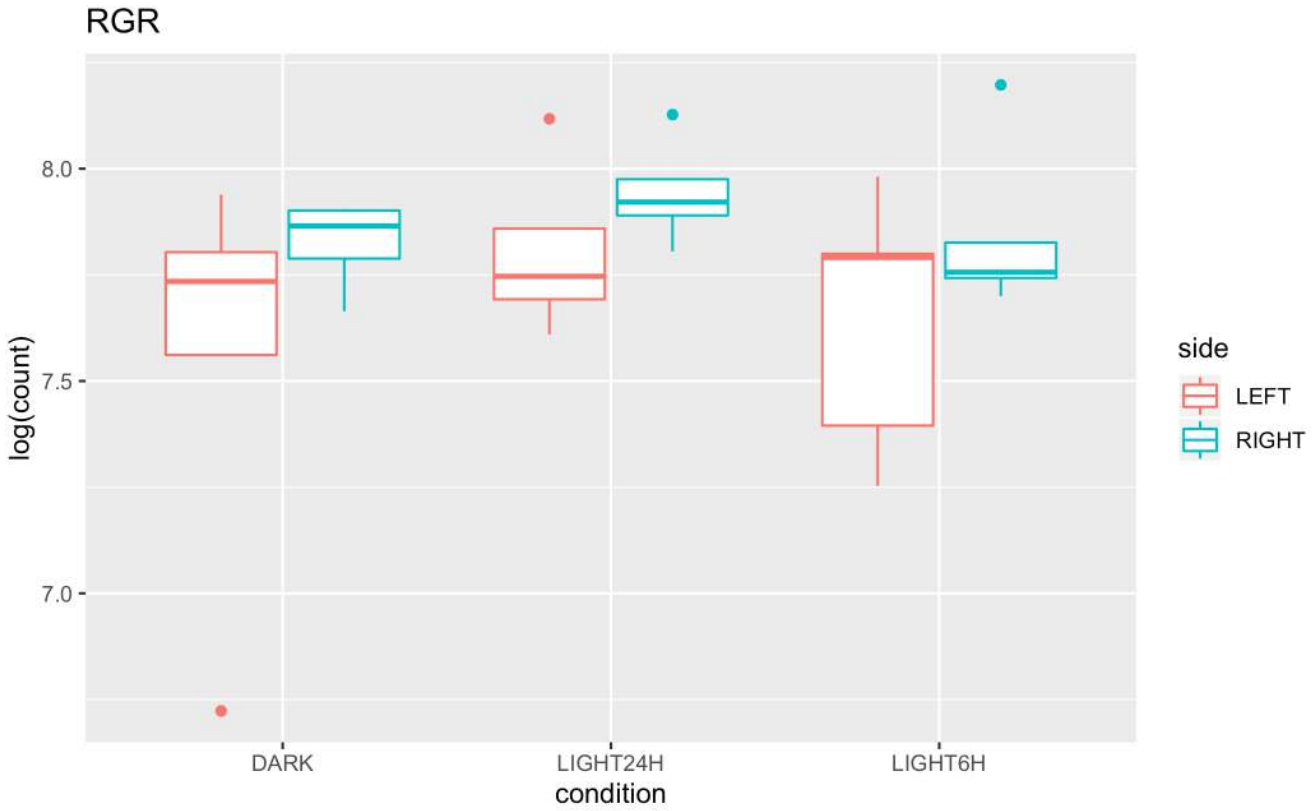
Hide

```
dat=plotCounts(dds,gene="RGR",intgroup = c("condition","side","animal"),returnData=
TRUE)

ggplot(dat,aes(condition,log(count),colour=side))+

  geom_boxplot()+

  ggtitle("RGR")
```



Hide

```
sessionInfo()
```

```
R version 3.6.1 (2019-07-05)

Platform: x86_64-apple-darwin15.6.0 (64-bit)

Running under: macOS Catalina 10.15.3



Matrix products: default

BLAS:   /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLi
b.framework/Versions/A/libBLAS.dylib

LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib



locale:

[1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8



attached base packages:

[1] parallel  stats4    stats     graphics  grDevices utils     datasets  methods
base



other attached packages:

 [1] apeglm_1.6.0                ggplot2_3.2.1               tibble_2.1.3

 [4] rjson_0.2.20                readr_1.3.1                 tximport_1.12.3

 [7] dplyr_0.8.4                 biomaRt_2.40.5              DESeq2_1.24.0

[10] SummarizedExperiment_1.14.1 DelayedArray_0.10.0         BiocParallel_1.18.1

[13] matrixStats_0.55.0          Biobase_2.44.0              GenomicRanges_1.36.1

[16] GenomeInfoDb_1.20.0         IRanges_2.18.3              S4Vectors_0.22.1

[19] BiocGenerics_0.30.0



loaded via a namespace (and not attached):

 [1] bitops_1.0-6           bit64_0.9-7            RColorBrewer_1.1-2     progress_
1.2.2

 [5] httr_1.4.1             numDeriv_2016.8-1.1    tools_3.6.1            backports
```

```
_1.1.5

 [9] R6_2.4.1               rpart_4.1-15           Hmisc_4.3-1            DBI_1.1.0

[13] lazyeval_0.2.2         colorspace_1.4-1       nnet_7.3-12            withr_2.
1.2

[17] tidyselect_1.0.0       gridExtra_2.3          prettyunits_1.1.1      bit_1.1-1
5.2

[21] compiler_3.6.1         htmlTable_1.13.3       labeling_0.3           scales_1.
1.0

[25] checkmate_2.0.0        mvtnorm_1.0-12         genefilter_1.66.0      stringr_
1.4.0

[29] digest_0.6.23          foreign_0.8-75         XVector_0.24.0         base64enc
_0.1-3

[33] jpeg_0.1-8.1           pkgconfig_2.0.3        htmltools_0.4.0        bbmle_1.
0.23.1

[37] htmlwidgets_1.5.1      rlang_0.4.4            rstudioapi_0.11        RSQLite_
2.2.0

[41] farver_2.0.3           acepack_1.4.1          RCurl_1.98-1.1         magrittr_
1.5

[45] GenomeInfoDbData_1.2.1 Formula_1.2-3          Matrix_1.2-18          Rcpp_1.0.
3

[49] munsell_0.5.0          lifecycle_0.1.0        stringi_1.4.5          MASS_7.3-
51.5

[53] zlibbioc_1.30.0        plyr_1.8.5             grid_3.6.1             blob_1.2.
1

[57] bdsmatrix_1.3-4        crayon_1.3.4           lattice_0.20-38        splines_
3.6.1

[61] annotate_1.62.0        hms_0.5.3              locfit_1.5-9.1         knitr_1.2
8

[65] pillar_1.4.3           geneplotter_1.62.0     XML_3.99-0.3           glue_1.3.
1

[69] latticeExtra_0.6-29    data.table_1.12.8      png_0.1-7              vctrs_0.
2.2

[73] gtable_0.3.0           purrr_0.3.3            assertthat_0.2.1       emdbook_
1.3.11

[77] xfun_0.12              xtable_1.8-4           coda_0.19-3            survival_
3.1-8
```

```
[81] AnnotationDbi_1.46.1   memoise_1.1.0          cluster_2.1.0
```