

# Contents

Basic DESeq2 results exploration	1
Introduction	1
Code setup . . . . .	1
PCA	2
Sample-to-sample distances	3
MA plots	3
P-values distribution	6
Adjusted p-values distribution	7
Top features	8
Count plots top features	8
Reproducibility	19
Bibliography	21

## Basic DESeq2 results exploration

Project: PDF report.

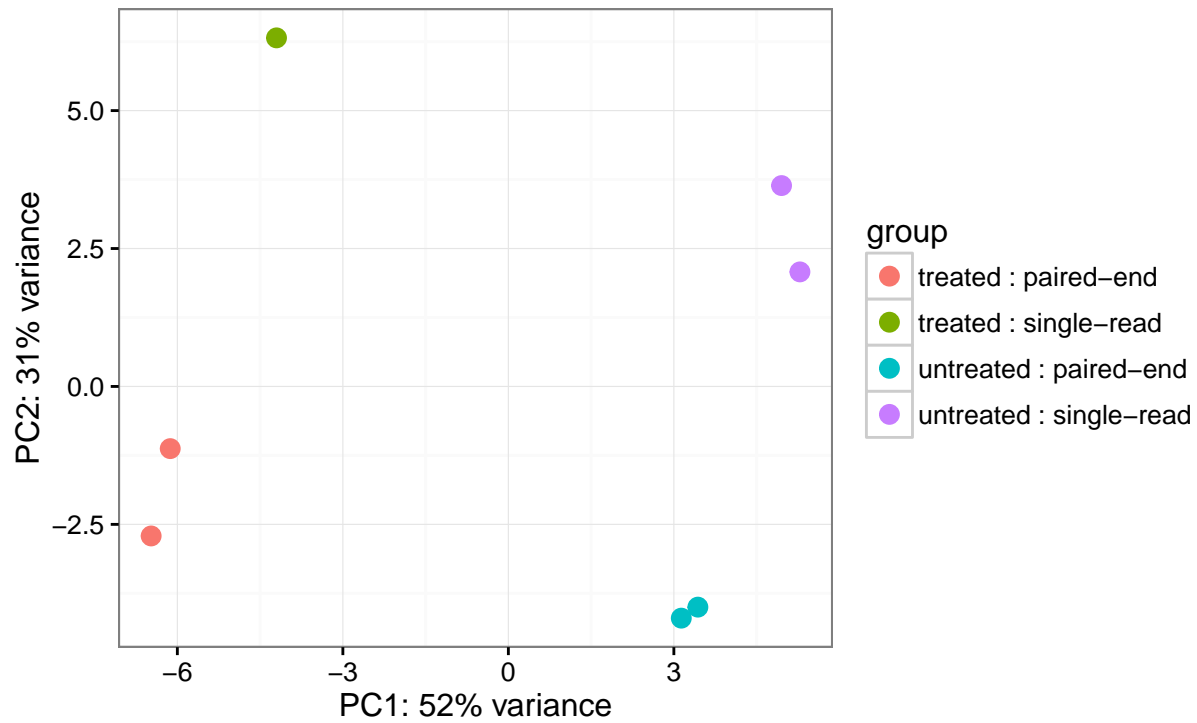
## Introduction

This report is meant to help explore DESeq2 (Love, Huber, and Anders, 2014) results and was generated using the `regionReport` (Collado-Torres, Jaffe, and Leek, 2015) package. While the report is rich, it is meant to just start the exploration of the results and exemplify some of the code used to do so. If you need a more in-depth analysis for your specific data set you might want to use the `customCode` argument. This report is based on the vignette of the DESeq2 (Love, Huber, and Anders, 2014) package which you can find [here](#).

## Code setup

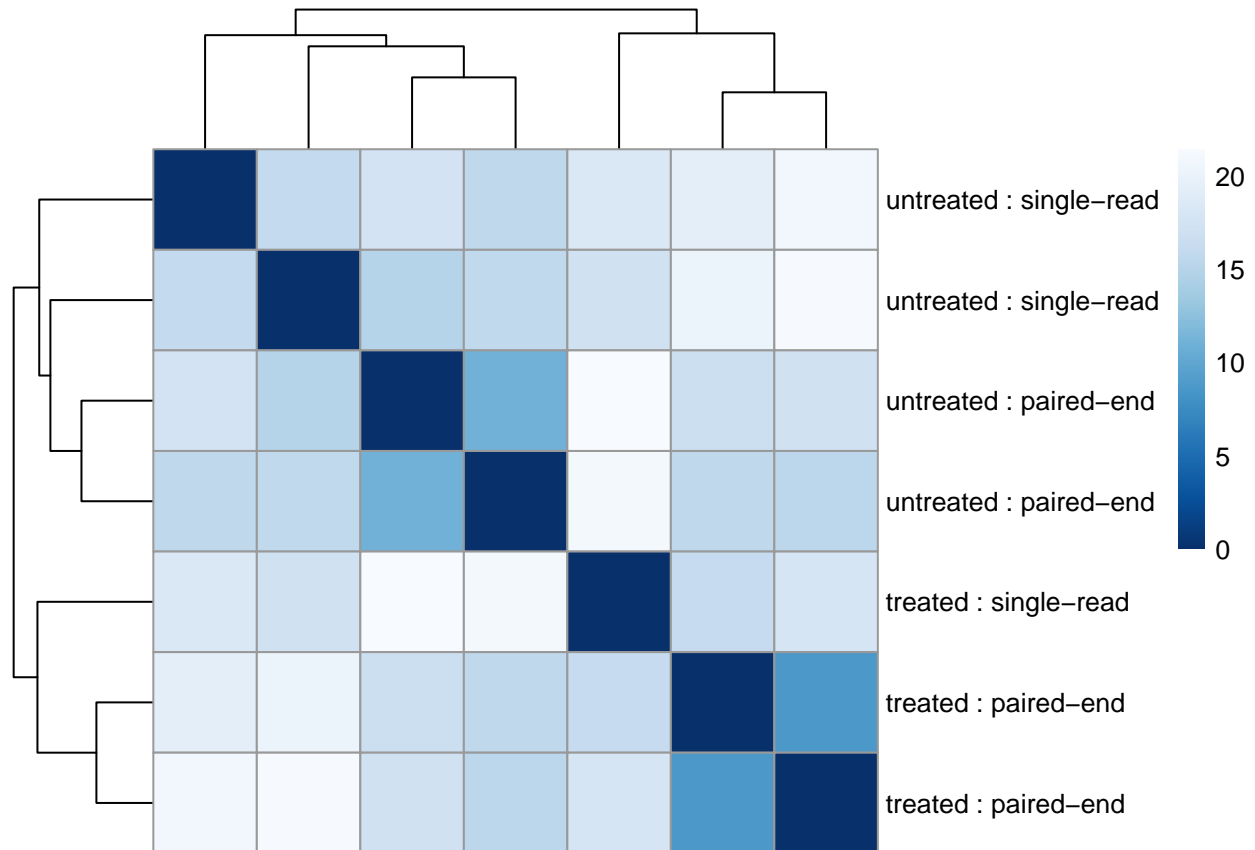
This section contains the code for setting up the rest of the report.

## PCA



The above plot shows the first two principal components that explain the variability in the data using the regularized log count data. If you are unfamiliar with principal component analysis, you might want to check the [Wikipedia entry](#) or this [interactive explanation](#). In this case, the first and second principal component explain 52 and 31 percent of the variance respectively.

## Sample-to-sample distances

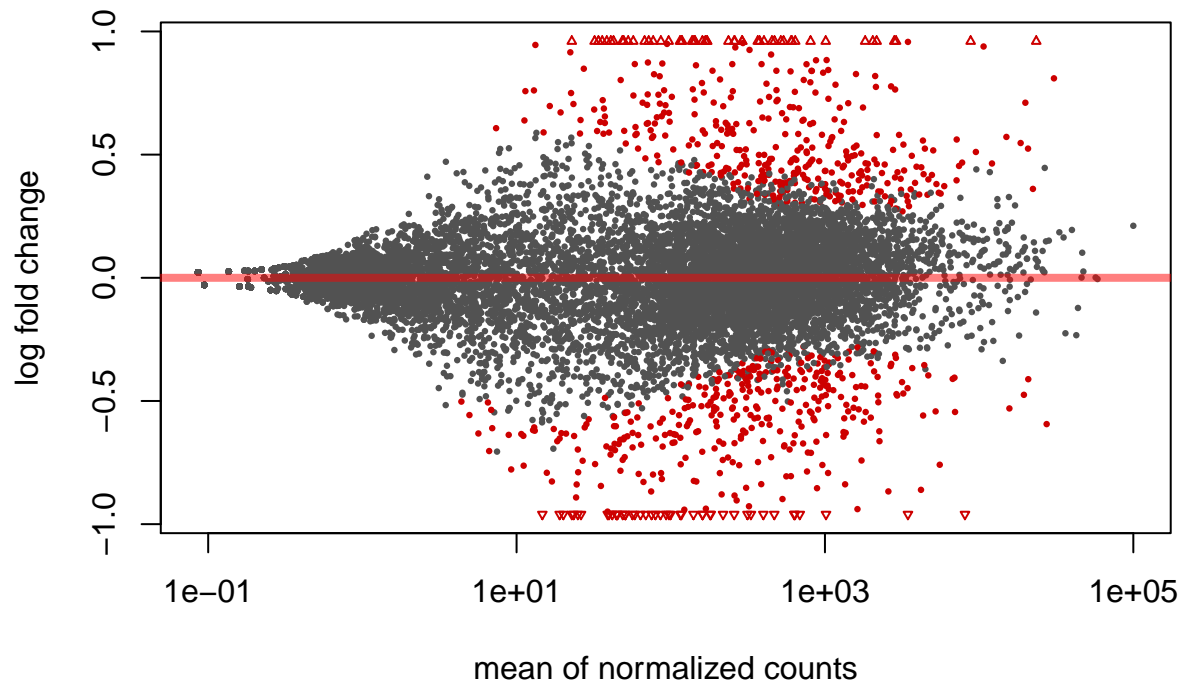


This plot shows how samples are clustered based on their euclidean distance using the regularized log transformed count data. This figure gives an overview of how the samples are hierarchically clustered. It is a complementary figure to the PCA plot.

## MA plots

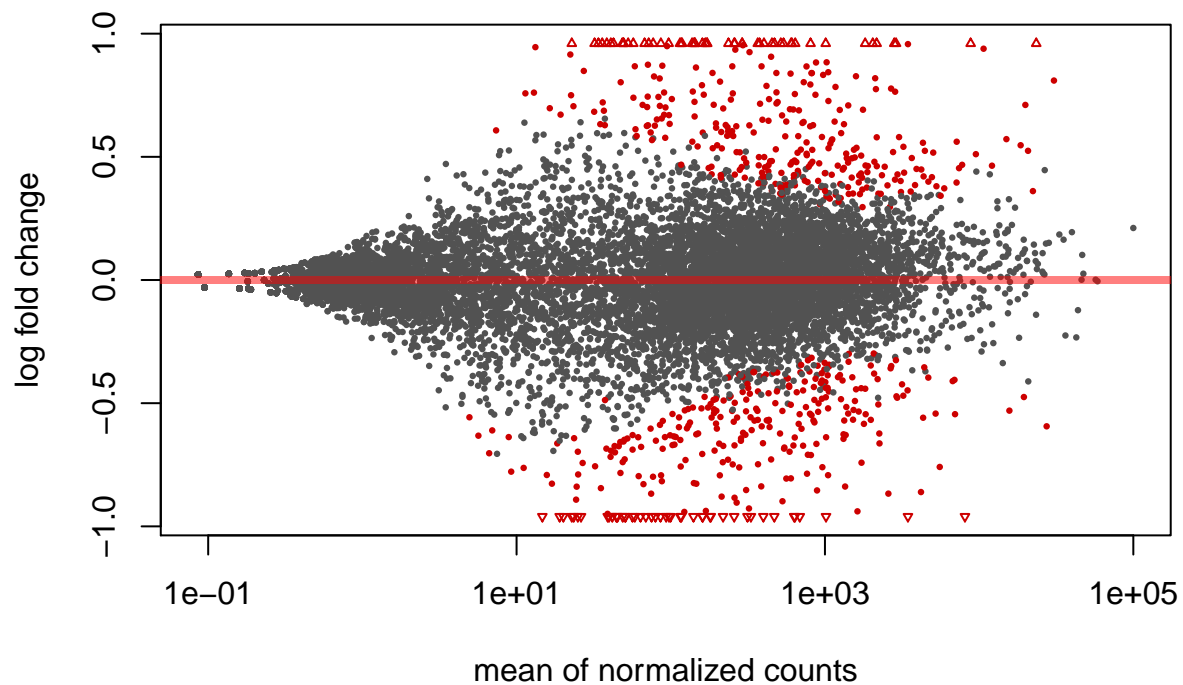
This section contains three MA plots (see [Wikipedia](#)) that compare the mean of the normalized counts against the log fold change. They show one point per feature. The points are shown in red if the feature has an adjusted p-value less than  $\alpha$ , that is, the statistically significant features are shown in red.

### MA plot with alpha = 0.1



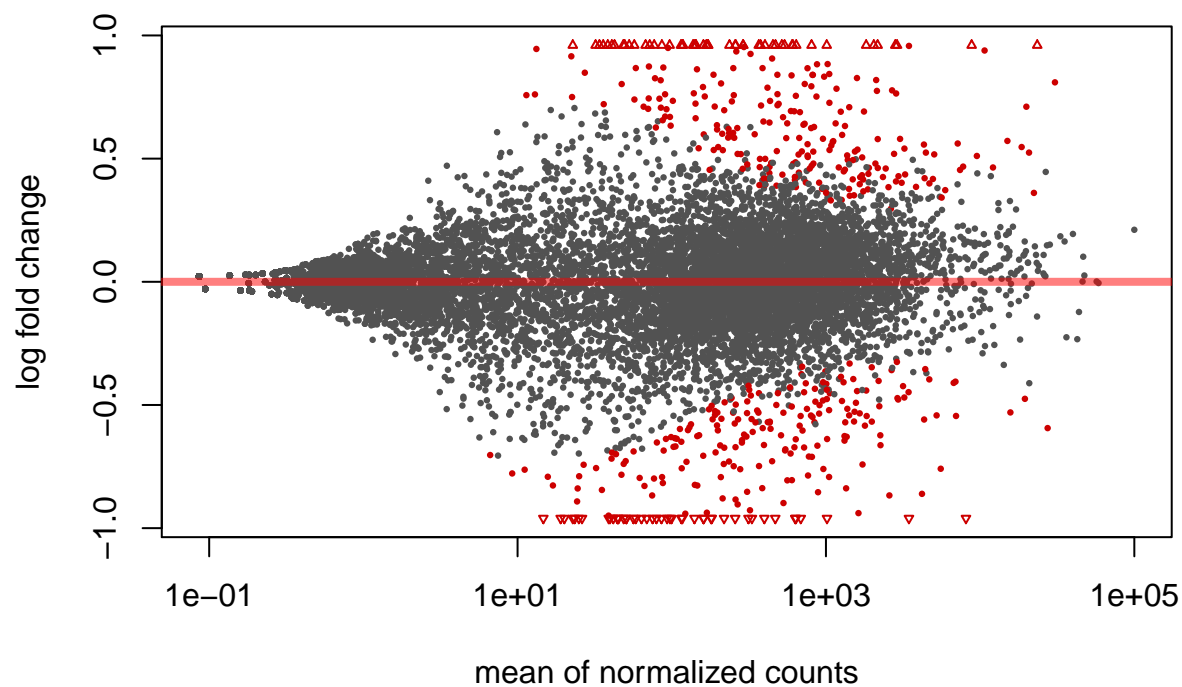
This first plot shows uses `alpha = 0.1`, which is the `alpha` value used to determine which resulting features were significant when running the function `DESeq2::results()`.

### MA plot with alpha = 0.05



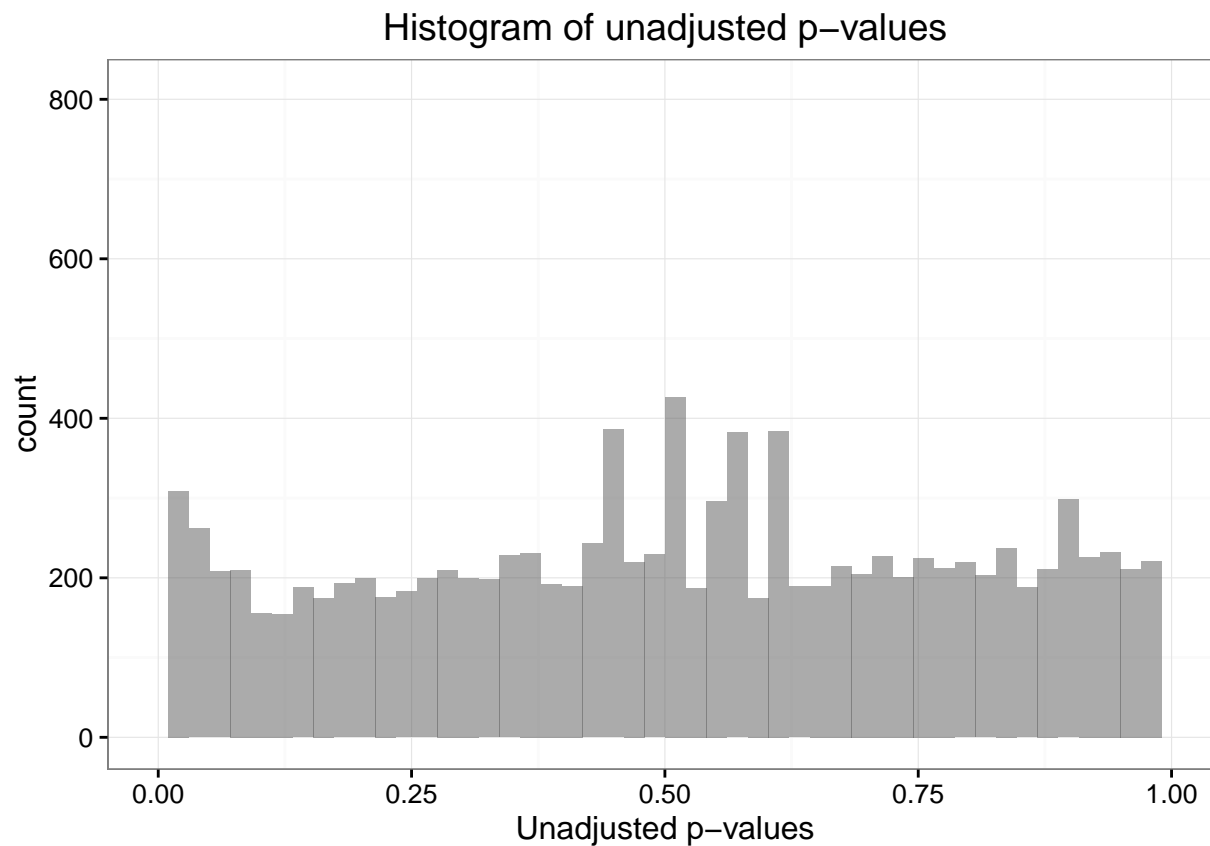
This second MA plot uses `alpha = 0.05` and can be used against the first MA plot to identify which features have adjusted p-values between 0.05 and 0.1.

## MA plot for top 500 features



The third and final MA plot uses an alpha such that the top 500 features are shown in the plot. These are the features that whose details are included in the *top features* interactive table.

## P-values distribution



This plot shows a histogram of the unadjusted p-values. It might be skewed right or left, or flat as shown in the [Wikipedia examples](#). The shape depends on the percent of features that are differentially expressed. For further information on how to interpret a histogram of p-values check [David Robinson's post on this topic](#).

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
## 0.0000  0.2258  0.4972  0.4795  0.7287  0.9999    2649
```

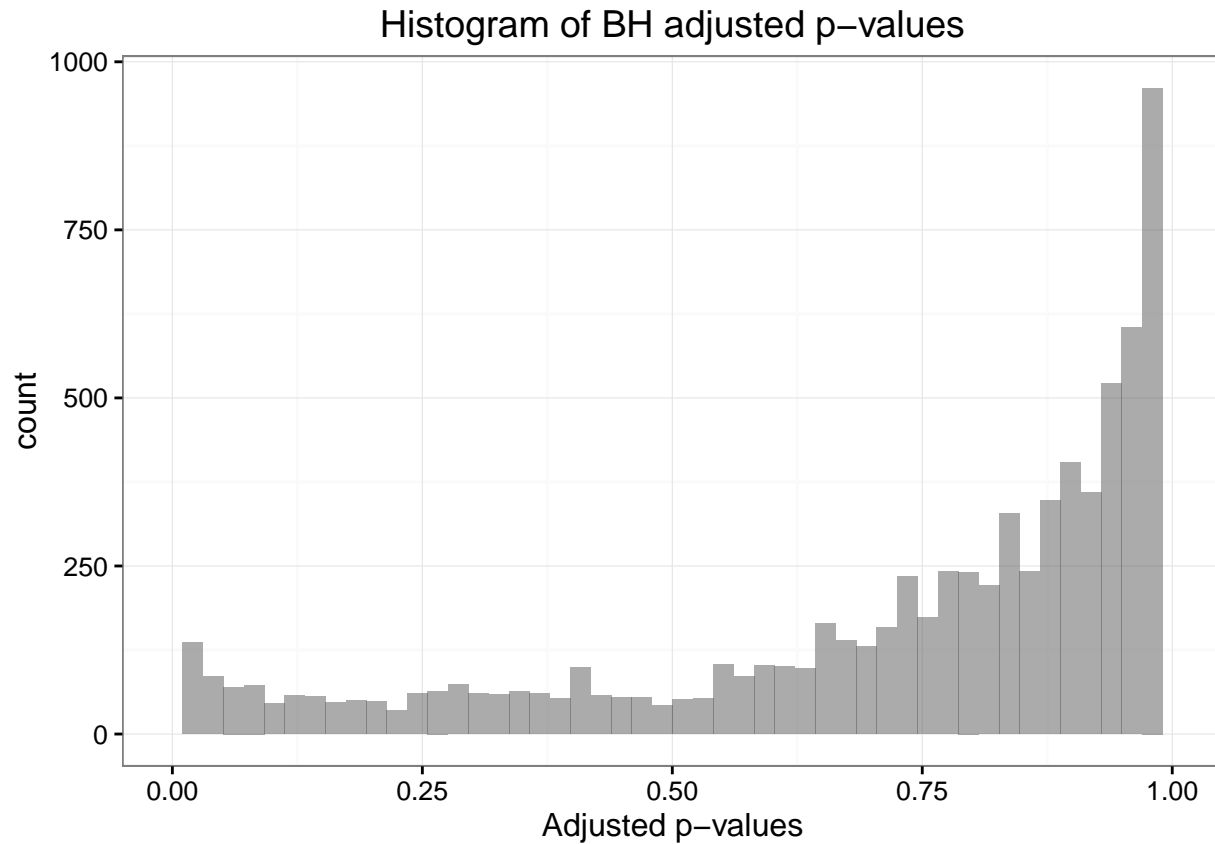
This is the numerical summary of the distribution of the p-values.

Cut	Count
0.0001	319
0.0010	468
0.0100	807
0.0250	1052
0.0500	1375
0.1000	1861
0.2000	2724
0.3000	3673
0.4000	4696
0.5000	5949
0.6000	7396
0.7000	8560
0.8000	9626
0.9000	10667

Cut	Count
1.0000	11821

This table shows the number of features with p-values less or equal than some commonly used cutoff values.

## Adjusted p-values distribution



This plot shows a histogram of the BH adjusted p-values. It might be skewed right or left, or flat as shown in the [Wikipedia examples](#).

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	0.000	0.560	0.827	0.702	0.947	1.000	6046

This is the numerical summary of the distribution of the BH adjusted p-values.

Cut	Count
0.0001	194
0.0010	267
0.0100	411
0.0250	516
0.0500	631
0.1000	797
0.2000	1043

Cut	Count
0.3000	1326
0.4000	1615
0.5000	1916
0.6000	2283
0.7000	2927
0.8000	3920
0.9000	5357
1.0000	8424

This table shows the number of features with BH adjusted p-values less or equal than some commonly used cutoff values.

## Top features

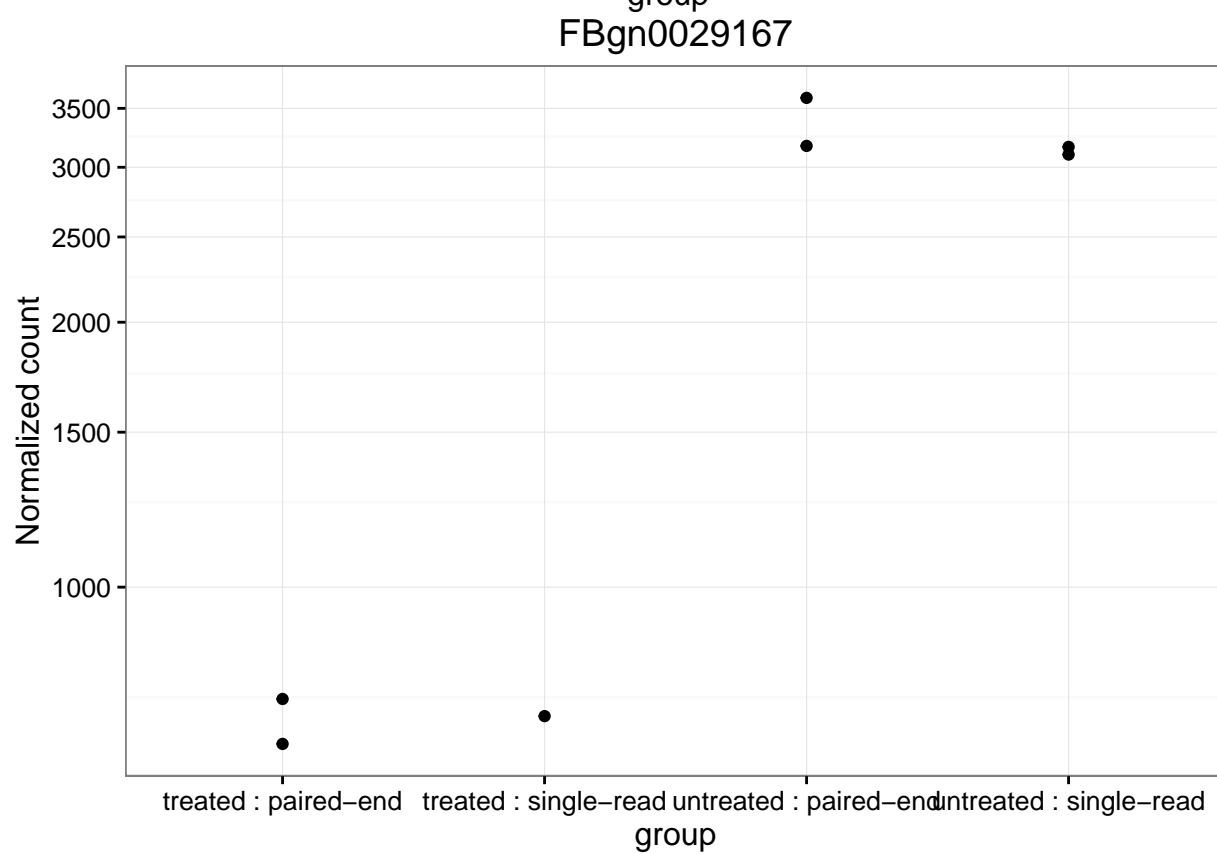
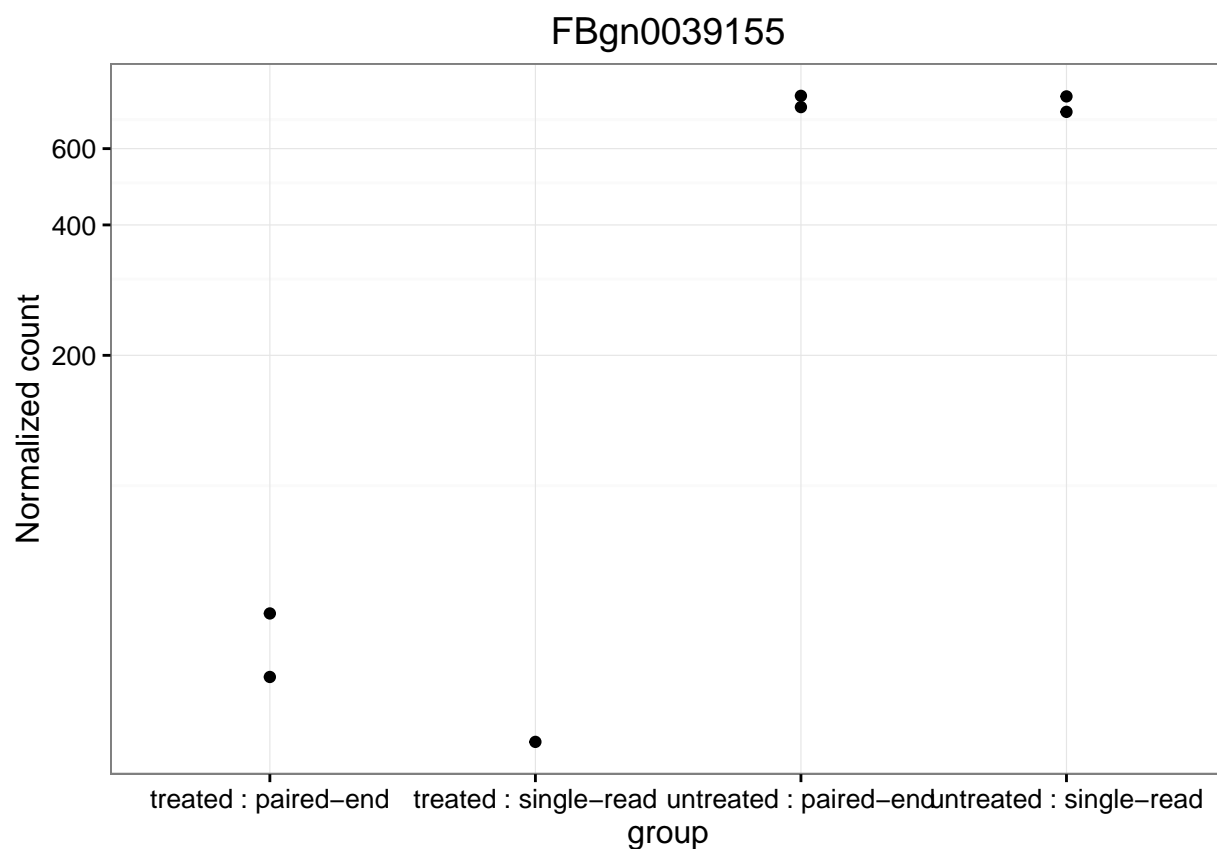
This table shows the top 500 features ordered by their BH adjusted p-values. Since the report is in PDF format, only the top 20 features are shown.

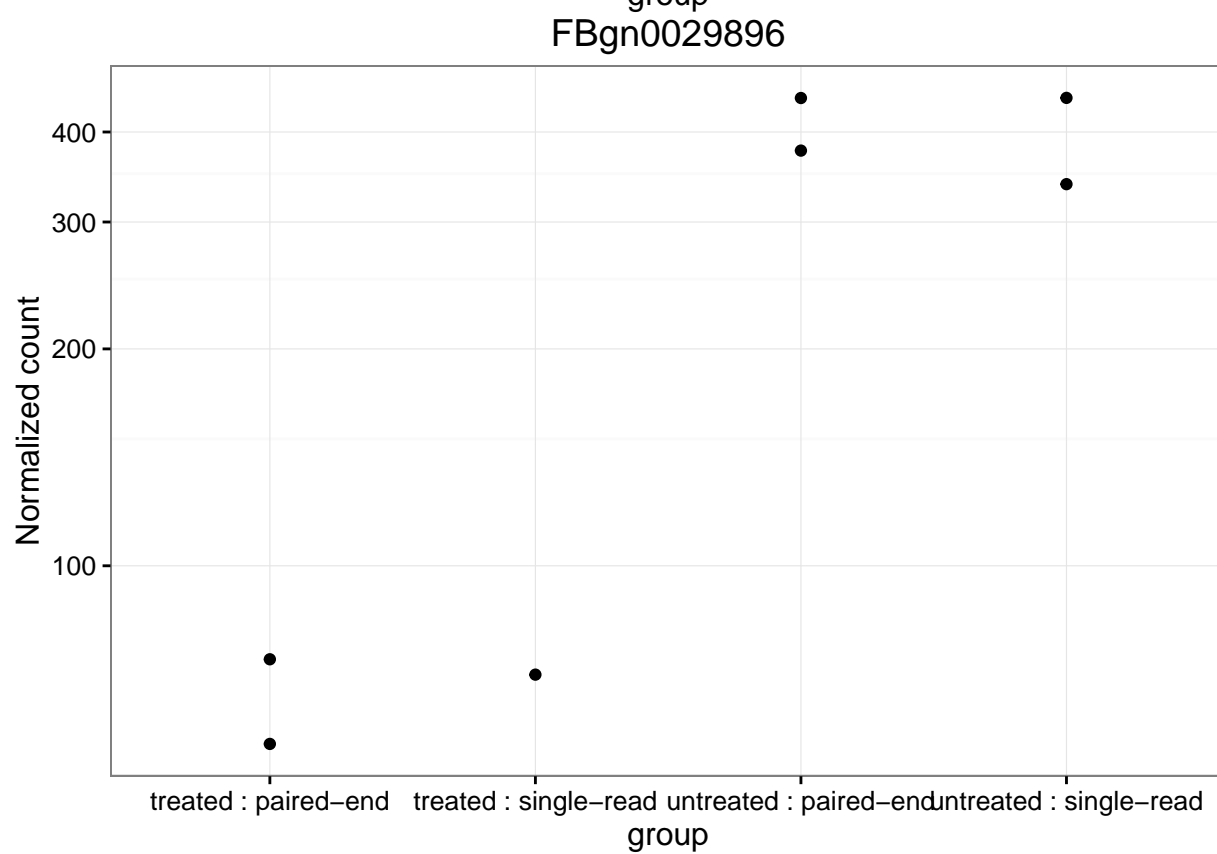
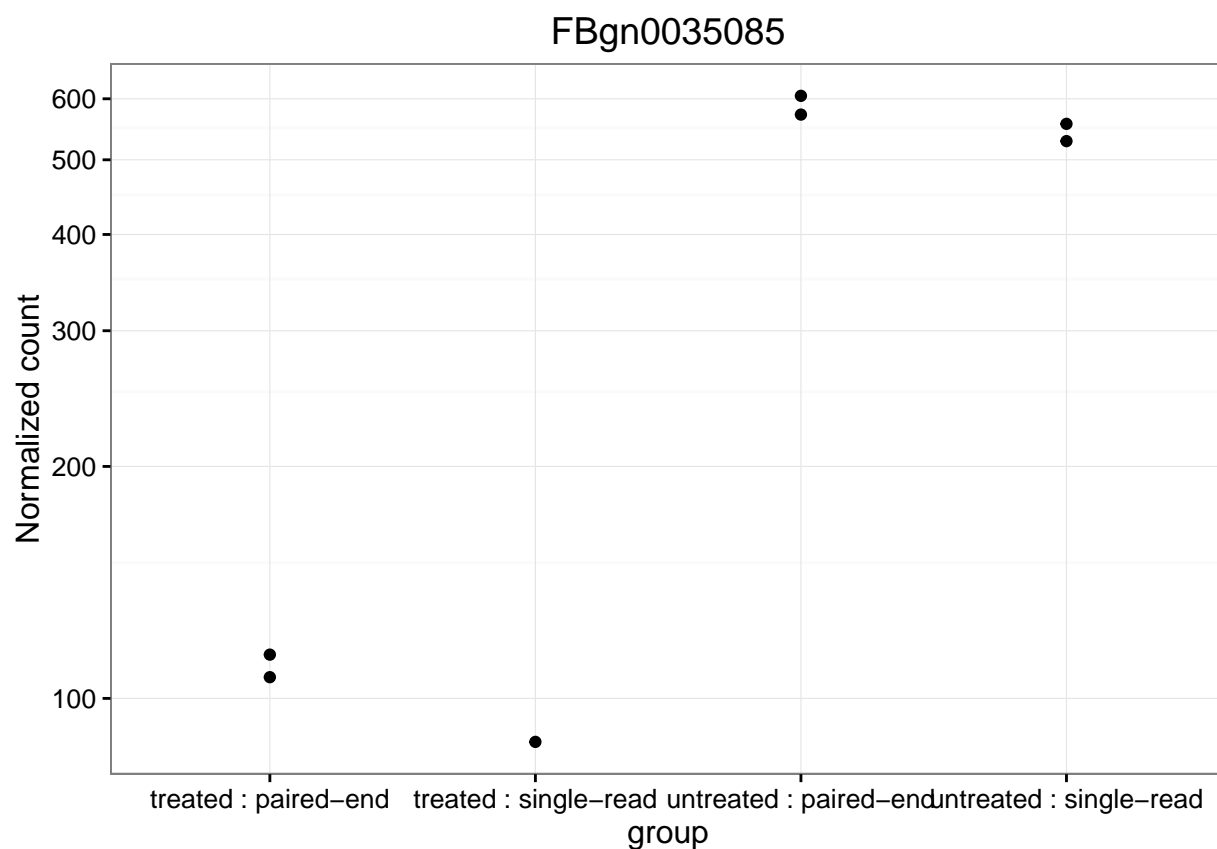
Feature	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
FBgn0039155	453.28	3.71	0.16	23.21	4.013291e-119	3.380797e-115
FBgn0029167	2165.04	2.08	0.10	20.10	6.684454e-90	2.815492e-86
FBgn0035085	366.83	2.23	0.14	16.26	1.888618e-59	5.303239e-56
FBgn0029896	257.90	2.21	0.16	13.91	5.854593e-44	1.232977e-40
FBgn0034736	118.41	2.57	0.18	13.88	8.067448e-44	1.359204e-40
FBgn0040091	610.60	1.43	0.12	11.91	1.114552e-32	1.564831e-29
FBgn0000071	180.01	-2.14	0.18	-11.81	3.498482e-32	4.210174e-29
FBgn0011260	140.36	-1.96	0.17	-11.61	3.723734e-31	3.921092e-28
FBgn0034434	76.91	2.39	0.21	11.26	2.045205e-29	1.914312e-26
FBgn0001226	686.72	-1.53	0.14	-11.10	1.310312e-28	1.103807e-25
FBgn0038832	167.59	2.04	0.19	10.92	9.049372e-28	6.930174e-25
FBgn0026562	23460.08	1.90	0.18	10.58	3.761064e-26	2.640267e-23
FBgn0051092	93.33	-1.95	0.19	-10.41	2.332072e-25	1.511183e-22
FBgn0035189	115.49	-2.12	0.20	-10.37	3.376356e-25	2.031602e-22
FBgn0031191+FBgn0027279	2043.81	0.96	0.09	10.30	6.734072e-25	3.587332e-22
FBgn0033913	2816.22	1.10	0.11	10.30	6.813545e-25	3.587332e-22
FBgn0023479	1823.54	1.36	0.13	10.28	8.376665e-25	4.150884e-22
FBgn0033764	41.75	-2.23	0.22	-10.23	1.480711e-24	6.929728e-22
FBgn0003501	161.22	-1.76	0.18	-10.05	9.406230e-24	4.170425e-21
FBgn0039419	806.81	1.11	0.11	9.78	1.336255e-22	5.628306e-20

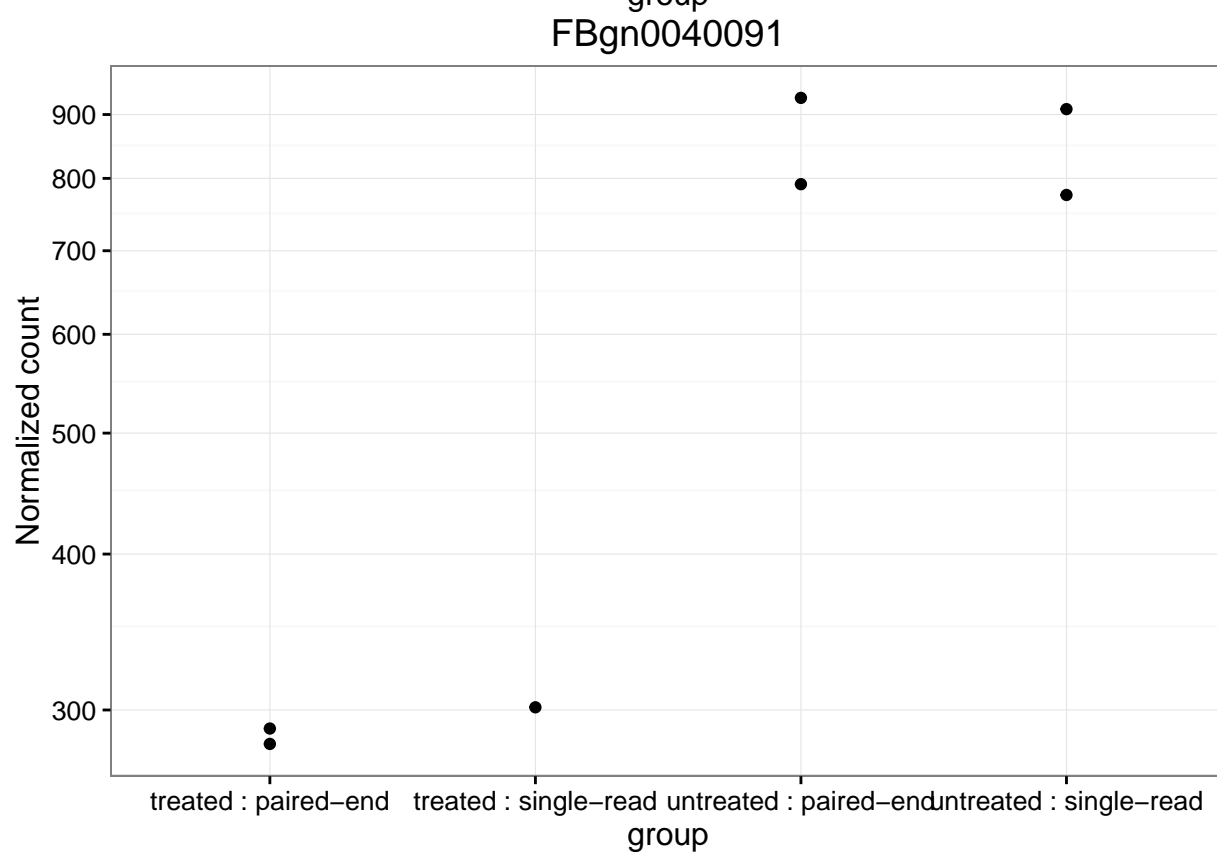
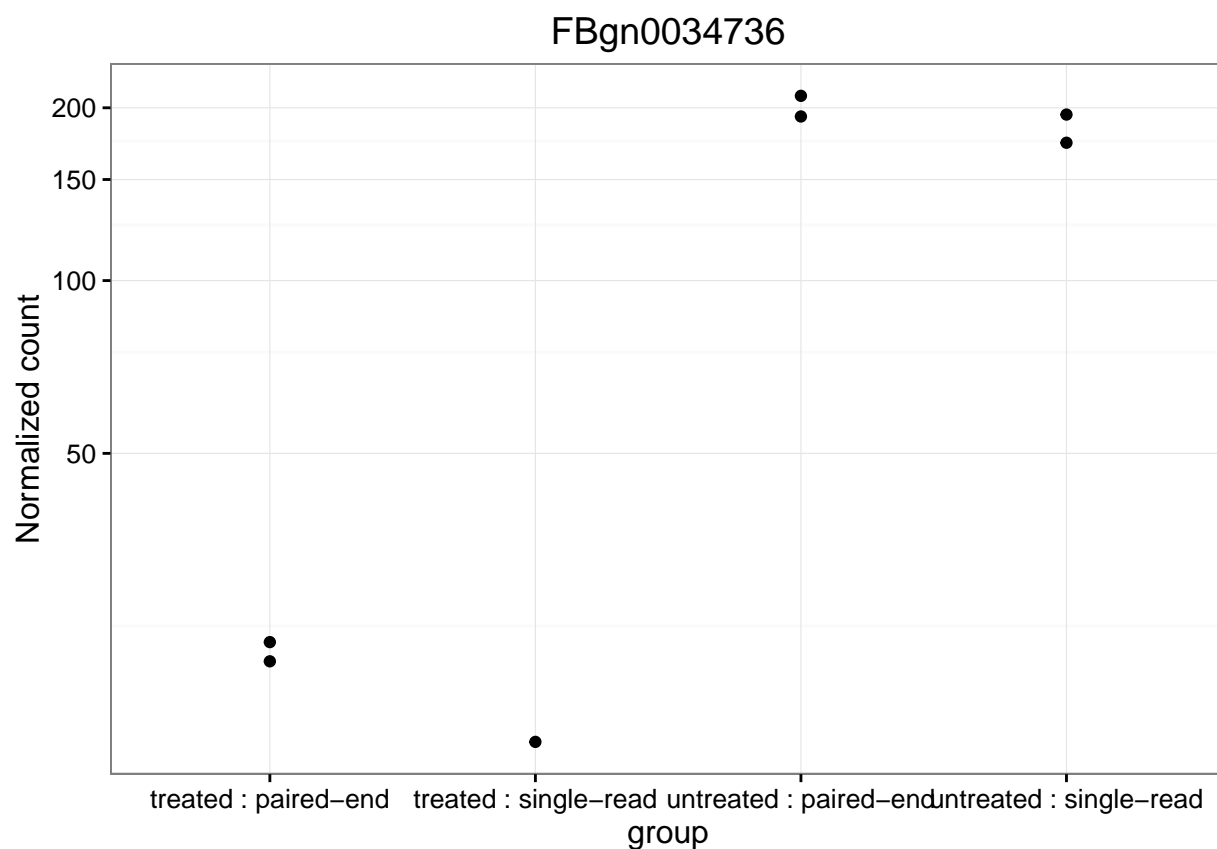
## Count plots top features

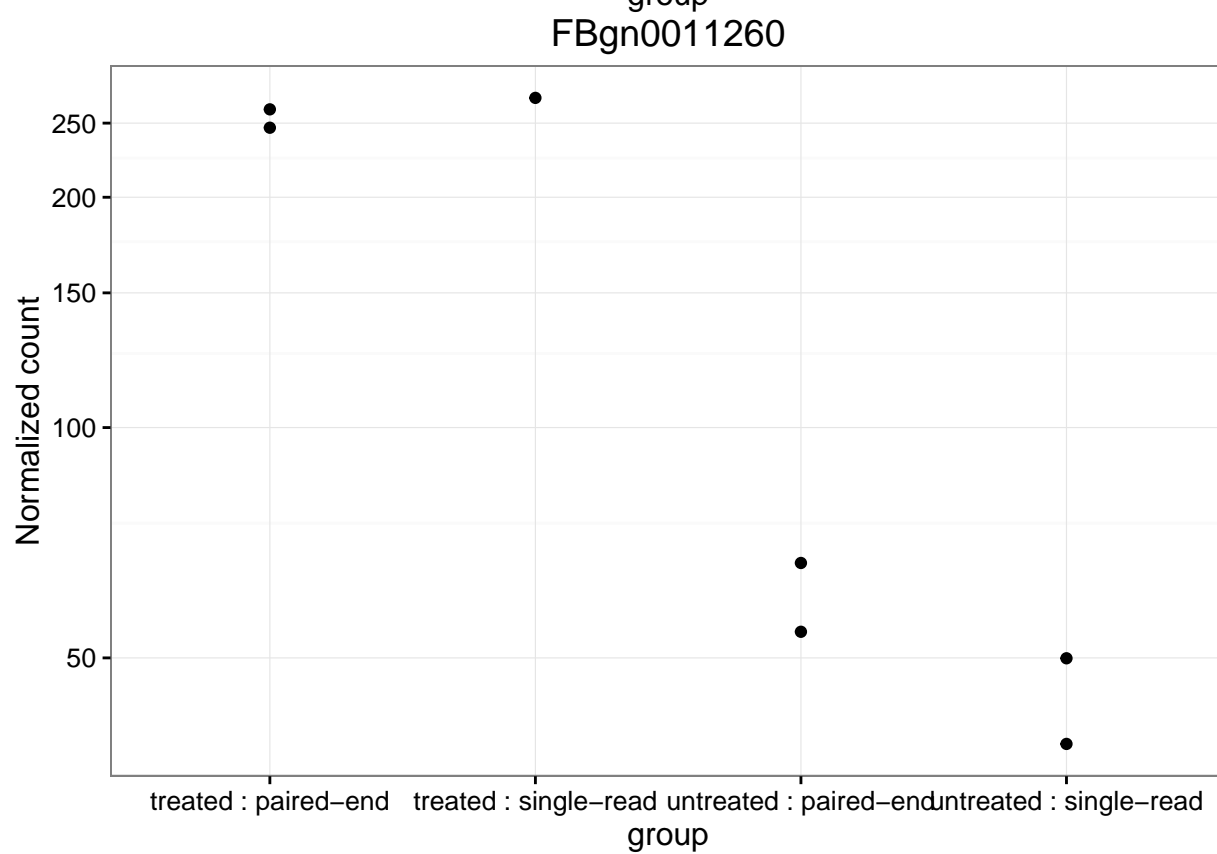
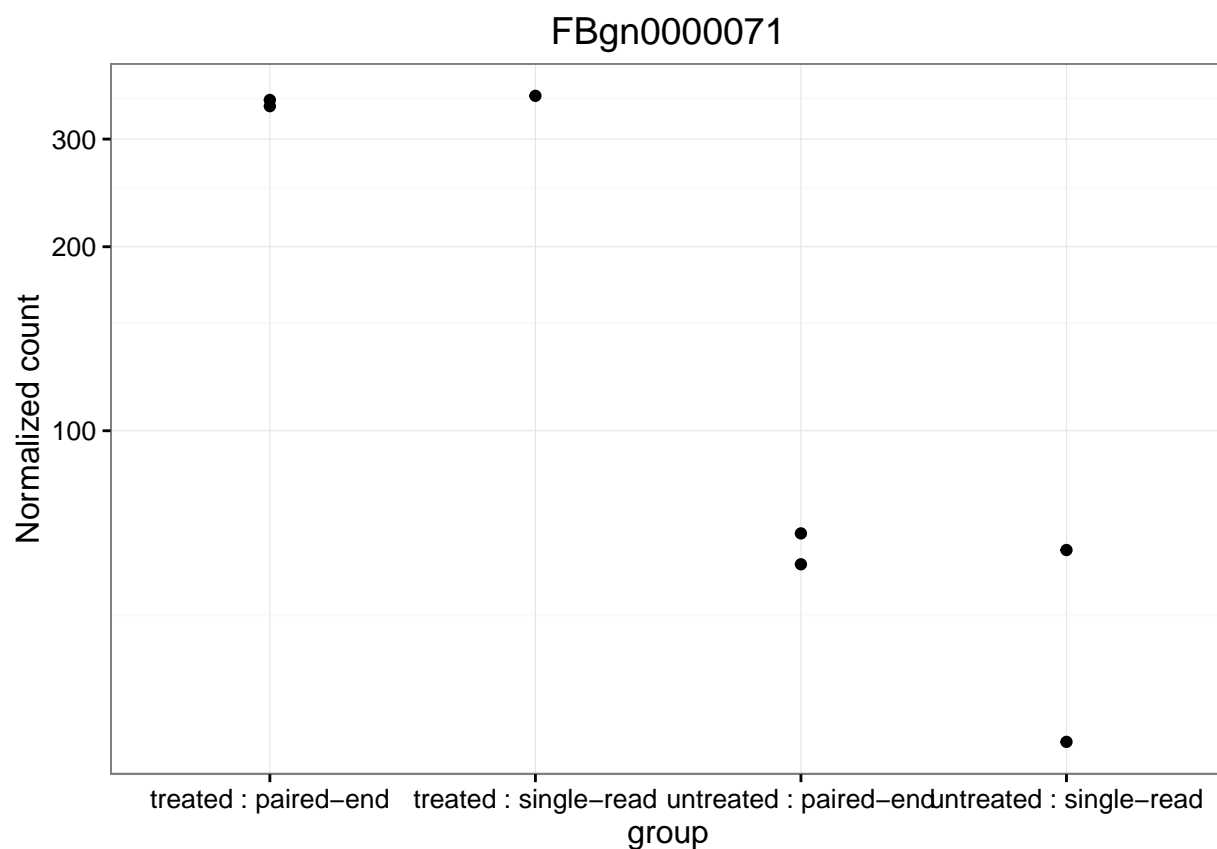
This section contains plots showing the normalized counts per sample for each group of interest. Only the best 20 features are shown, ranked by their BH adjusted p-values. The Y axis is on the log10 scale and the feature name is shown in the title of each plot.

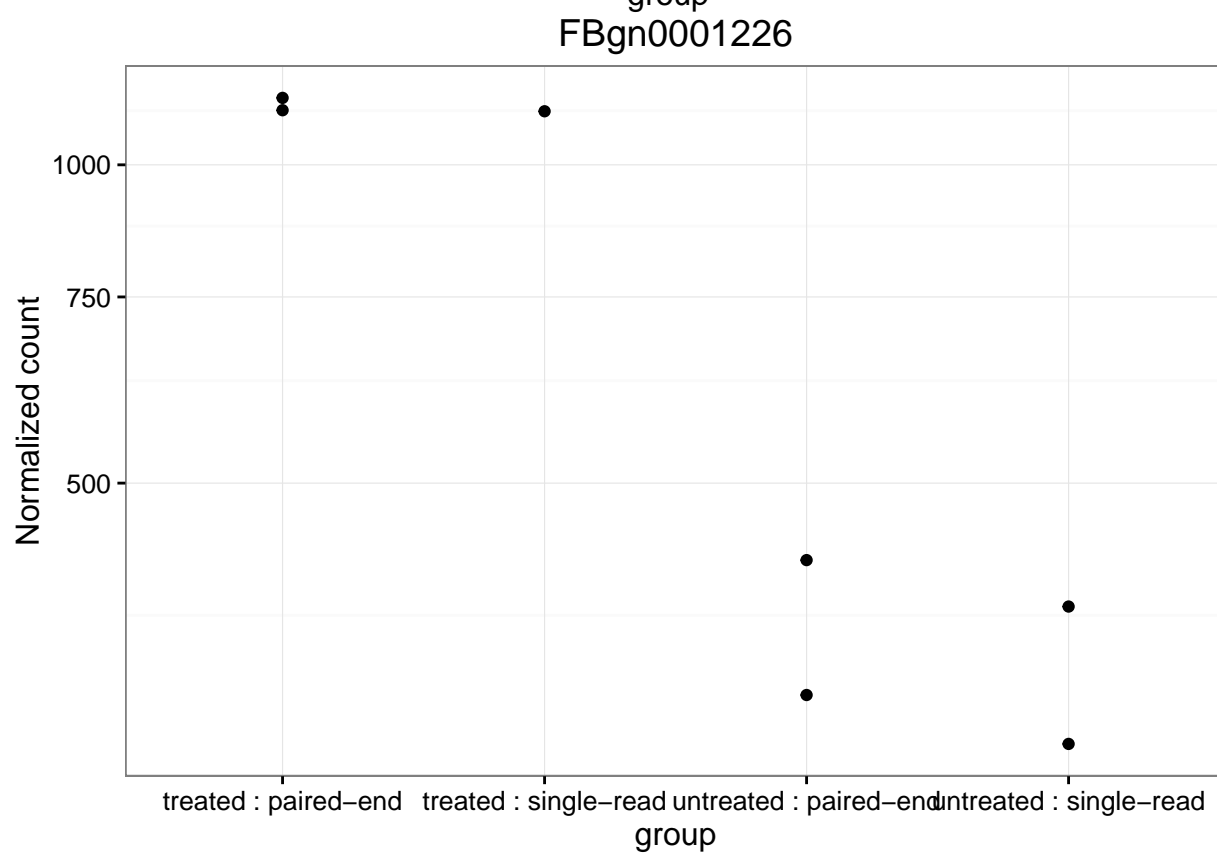
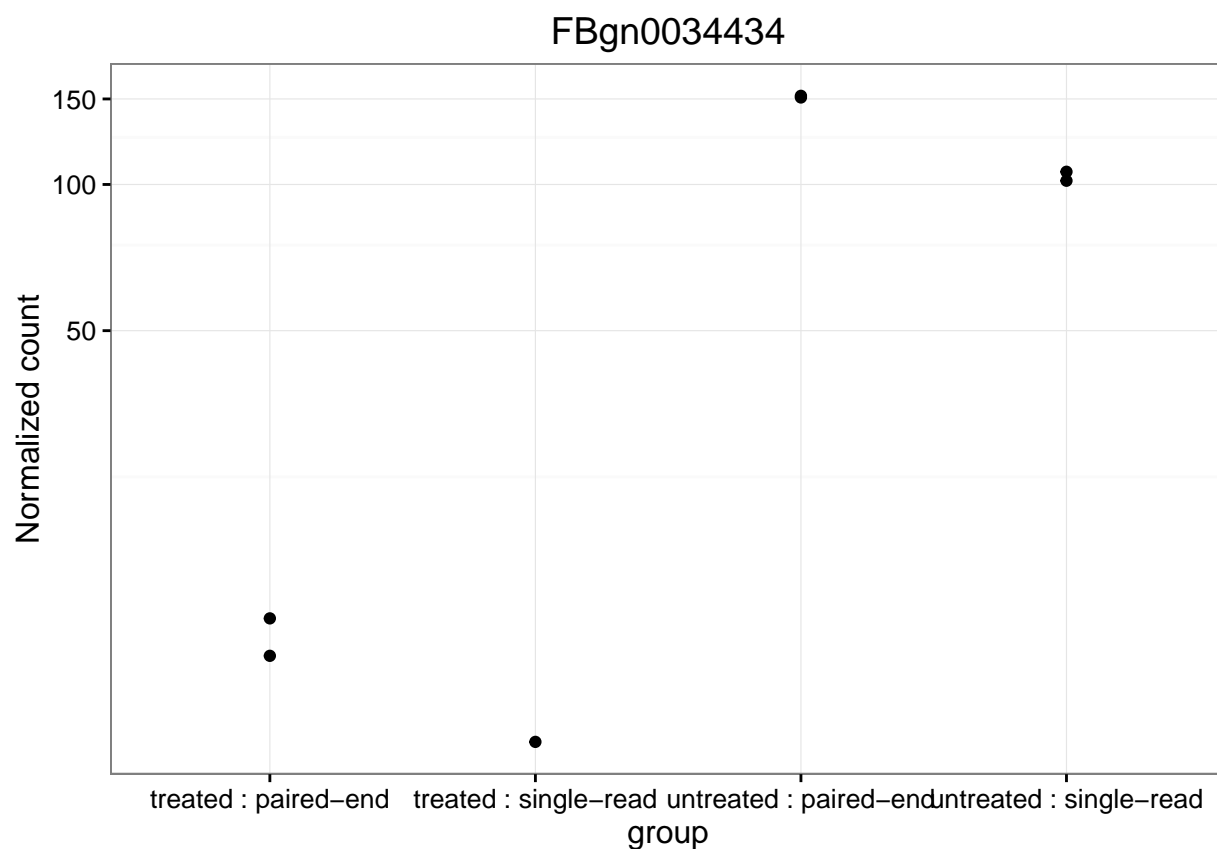


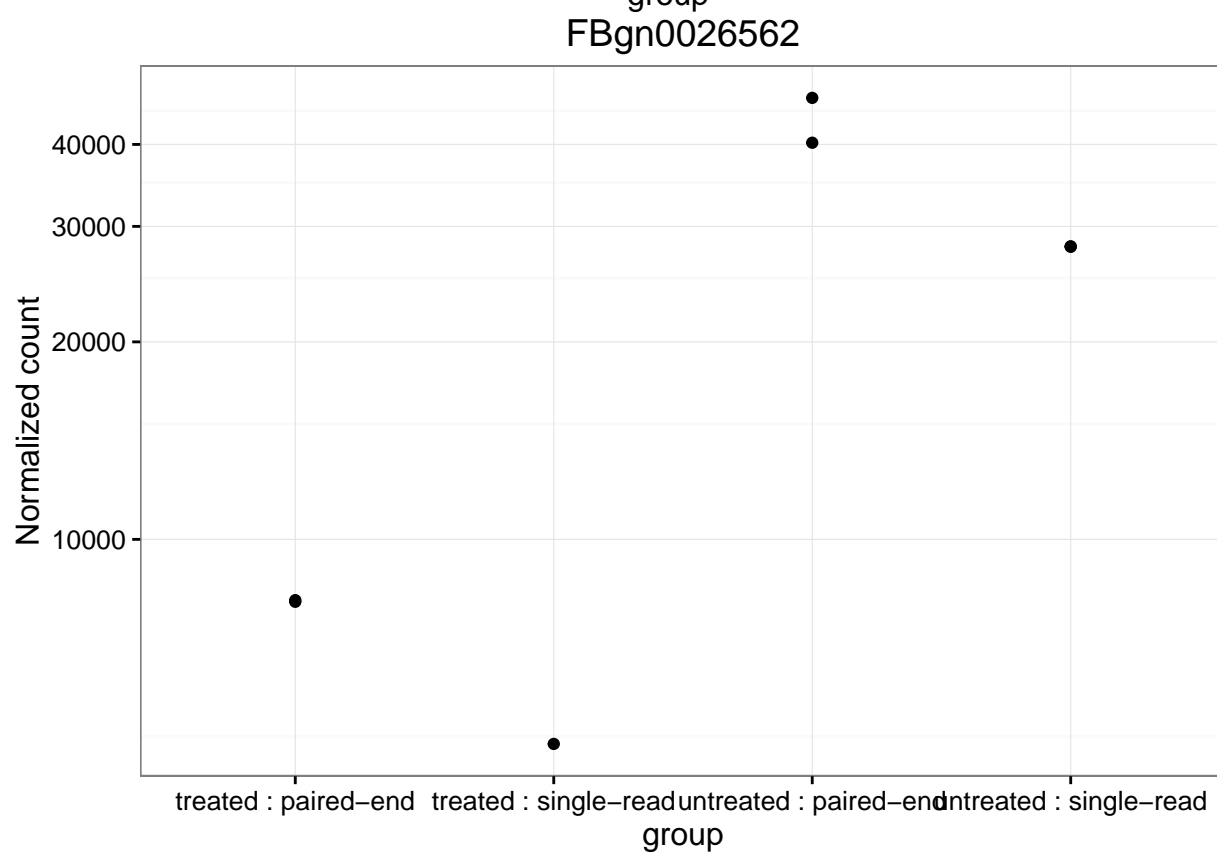
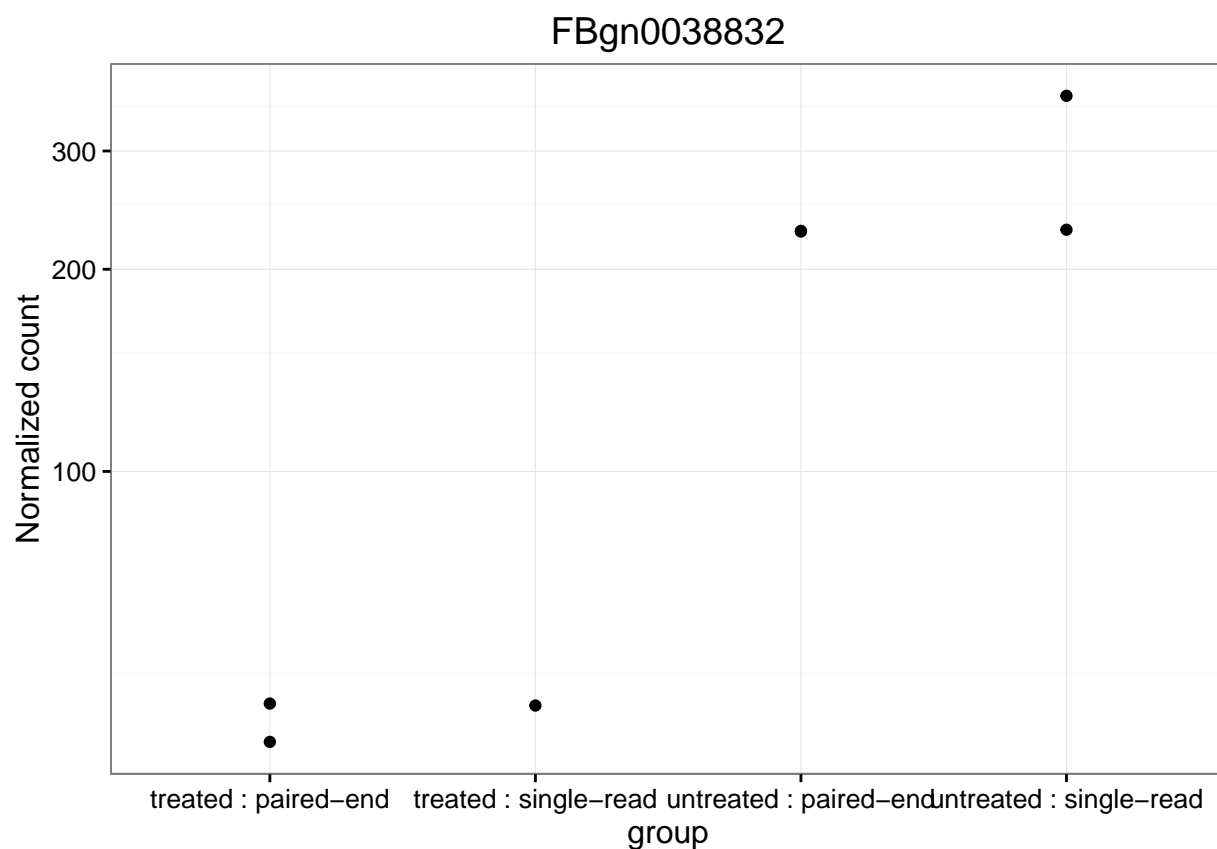


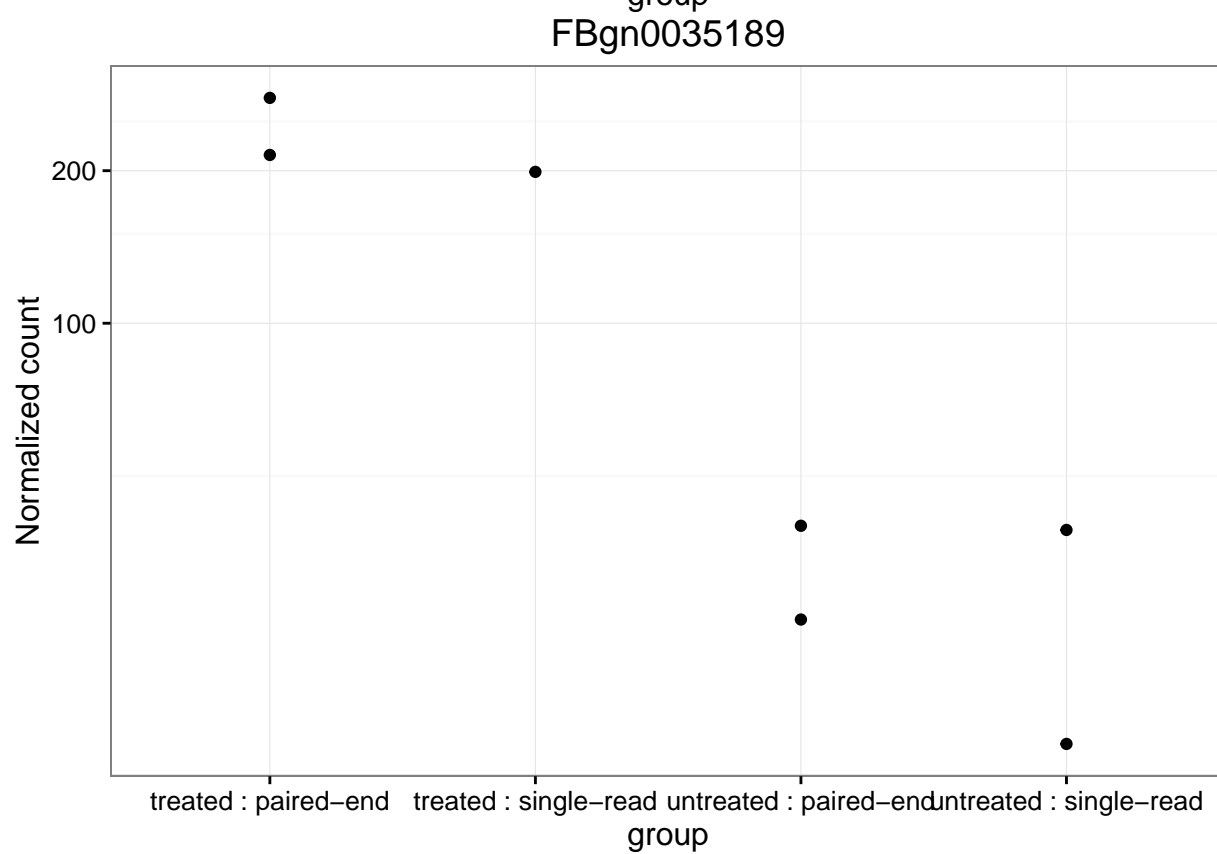
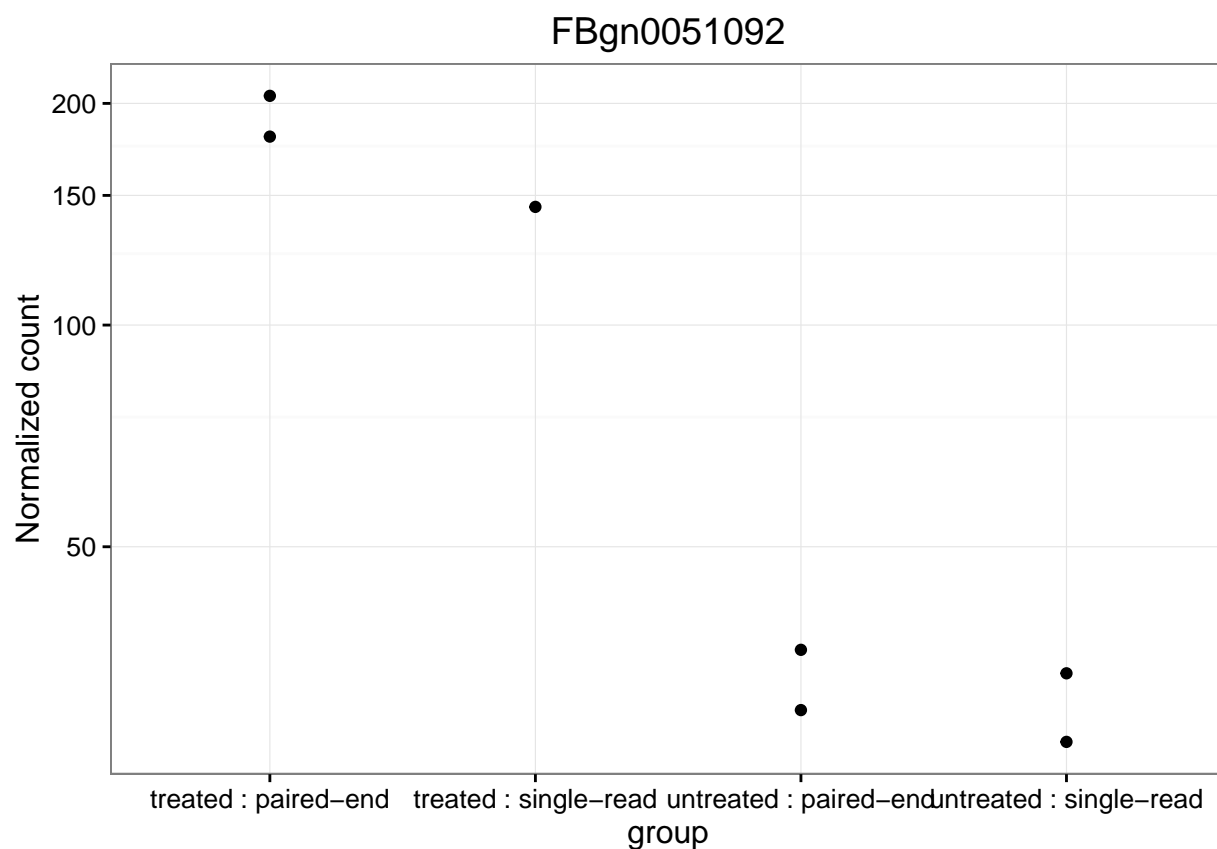


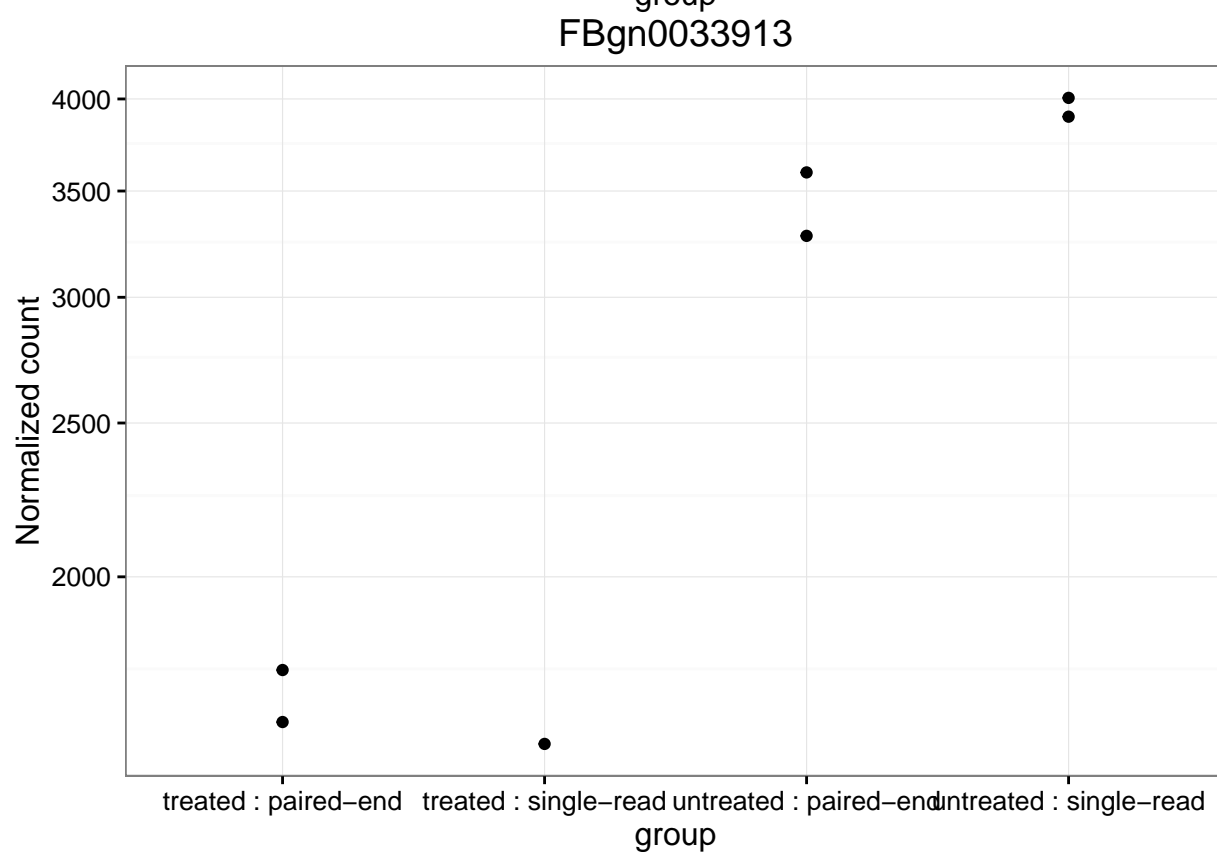
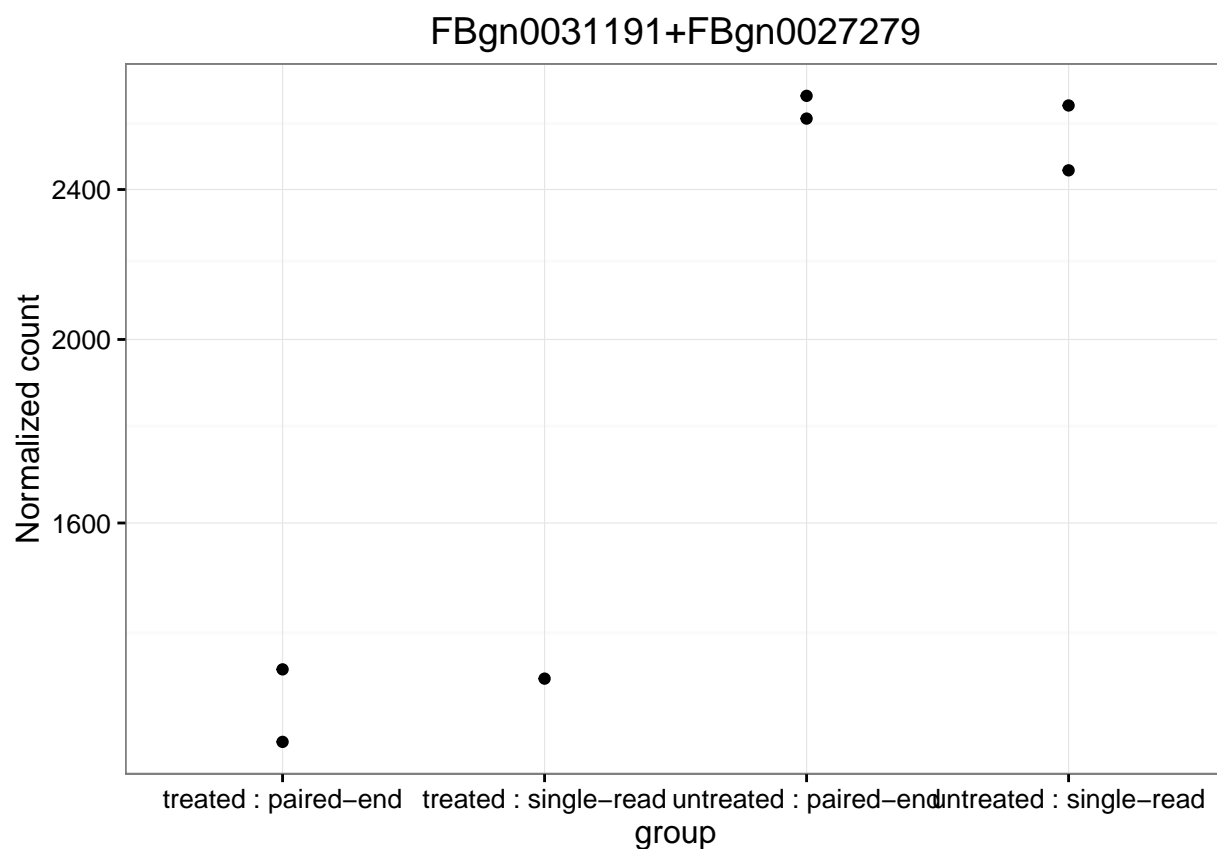




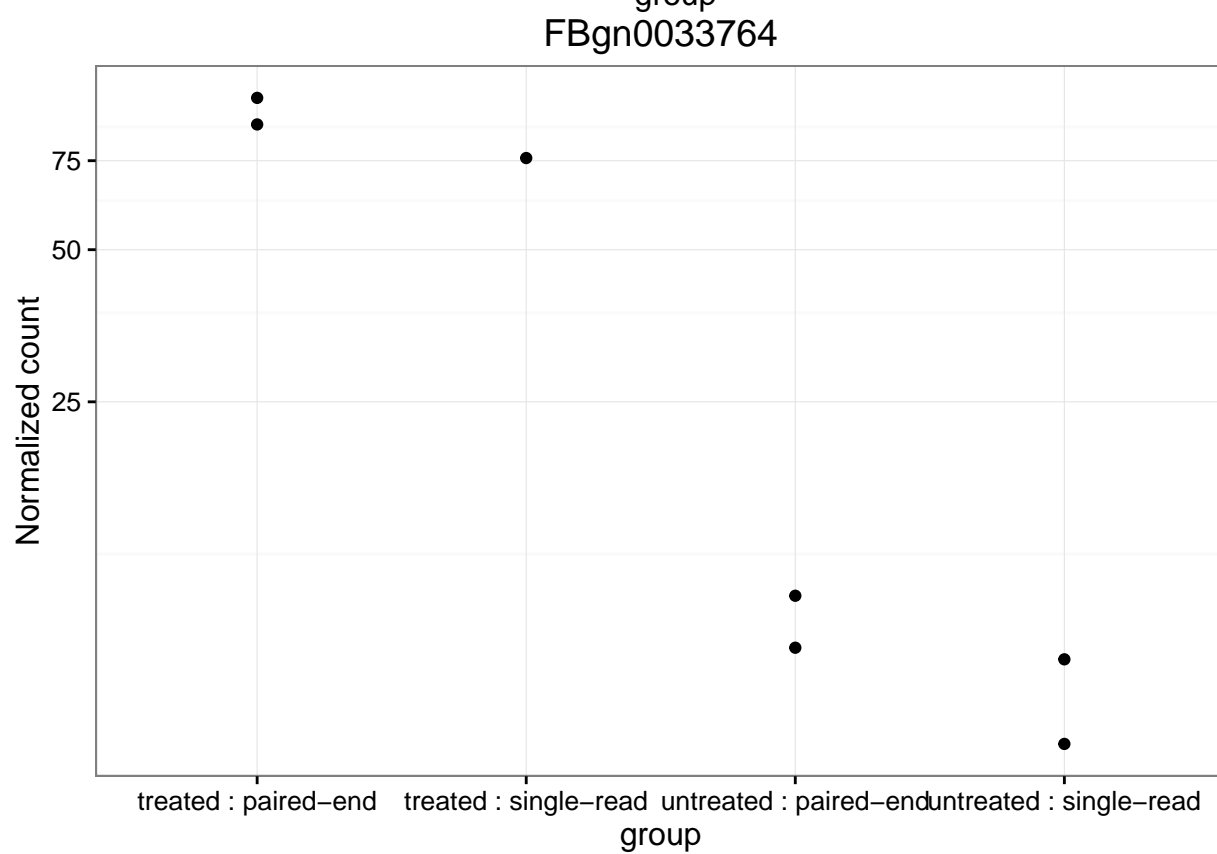
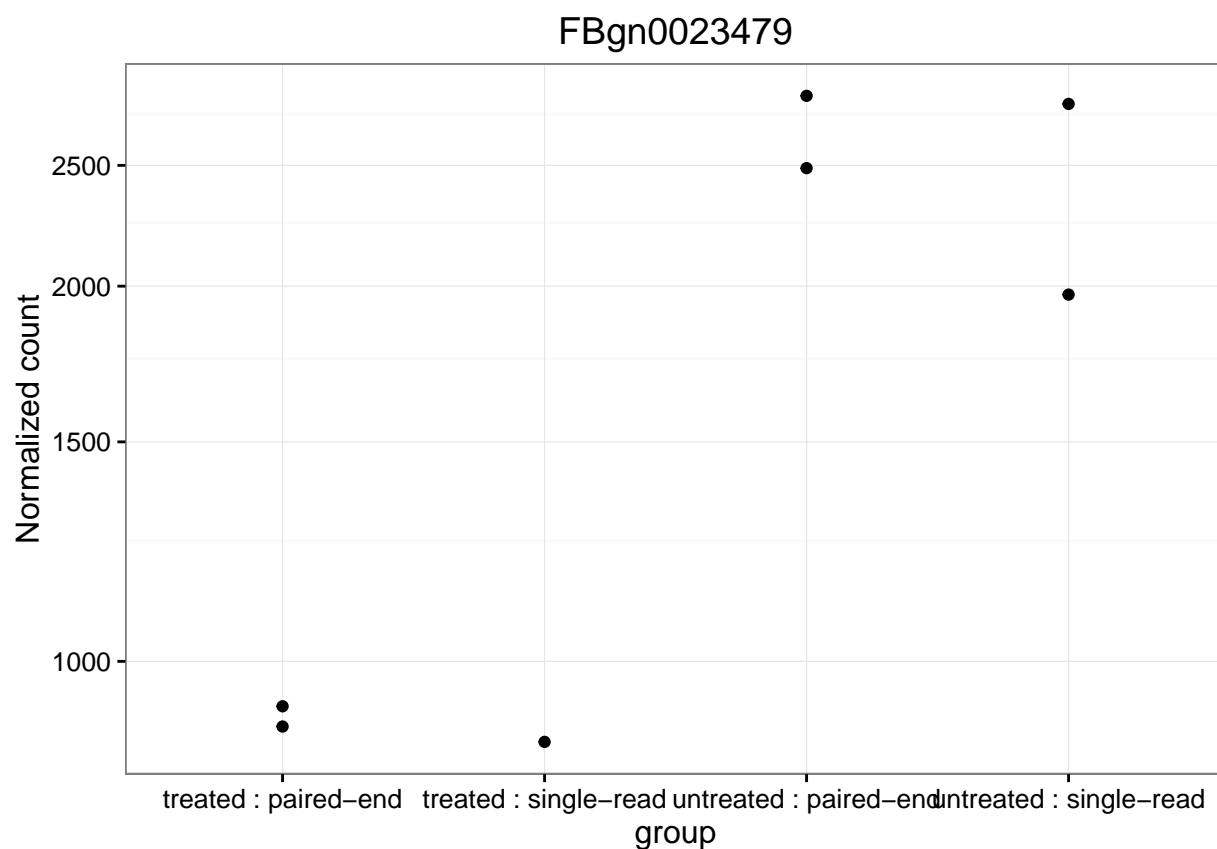


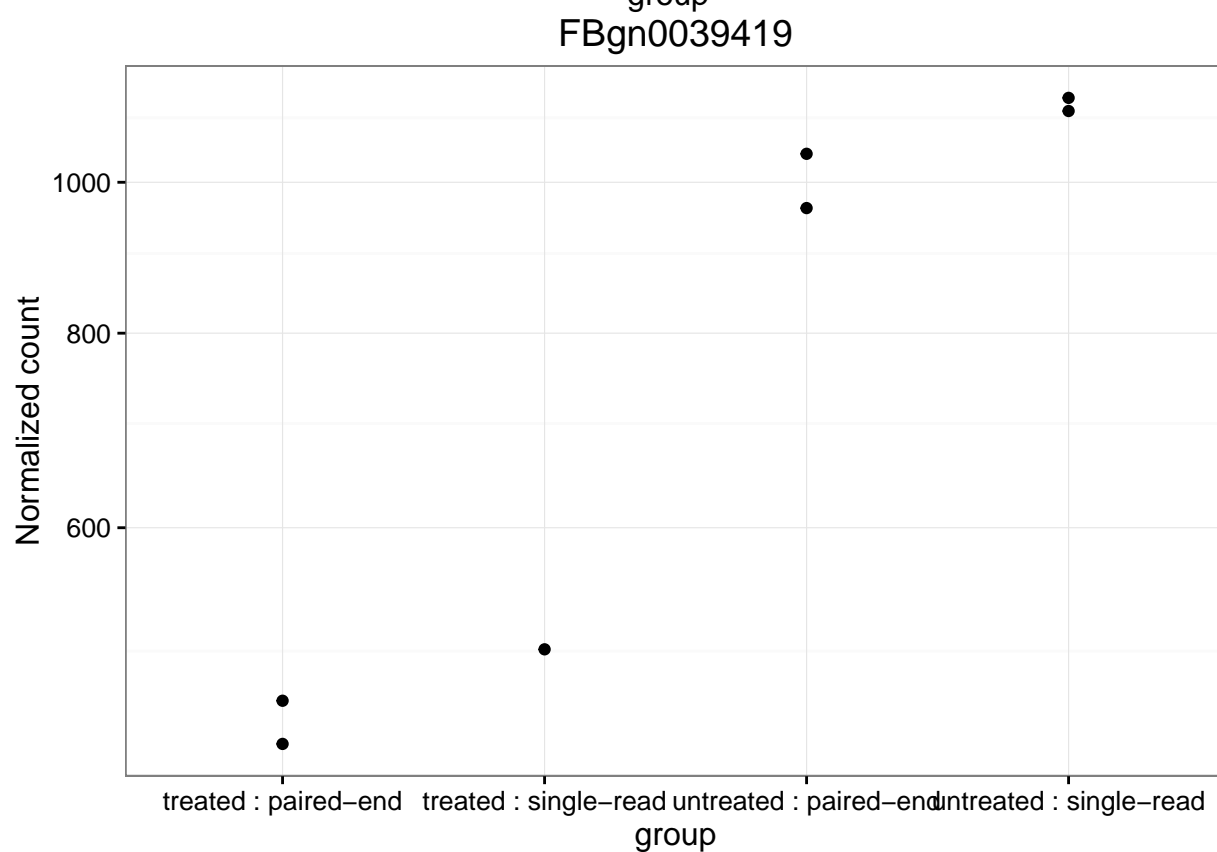
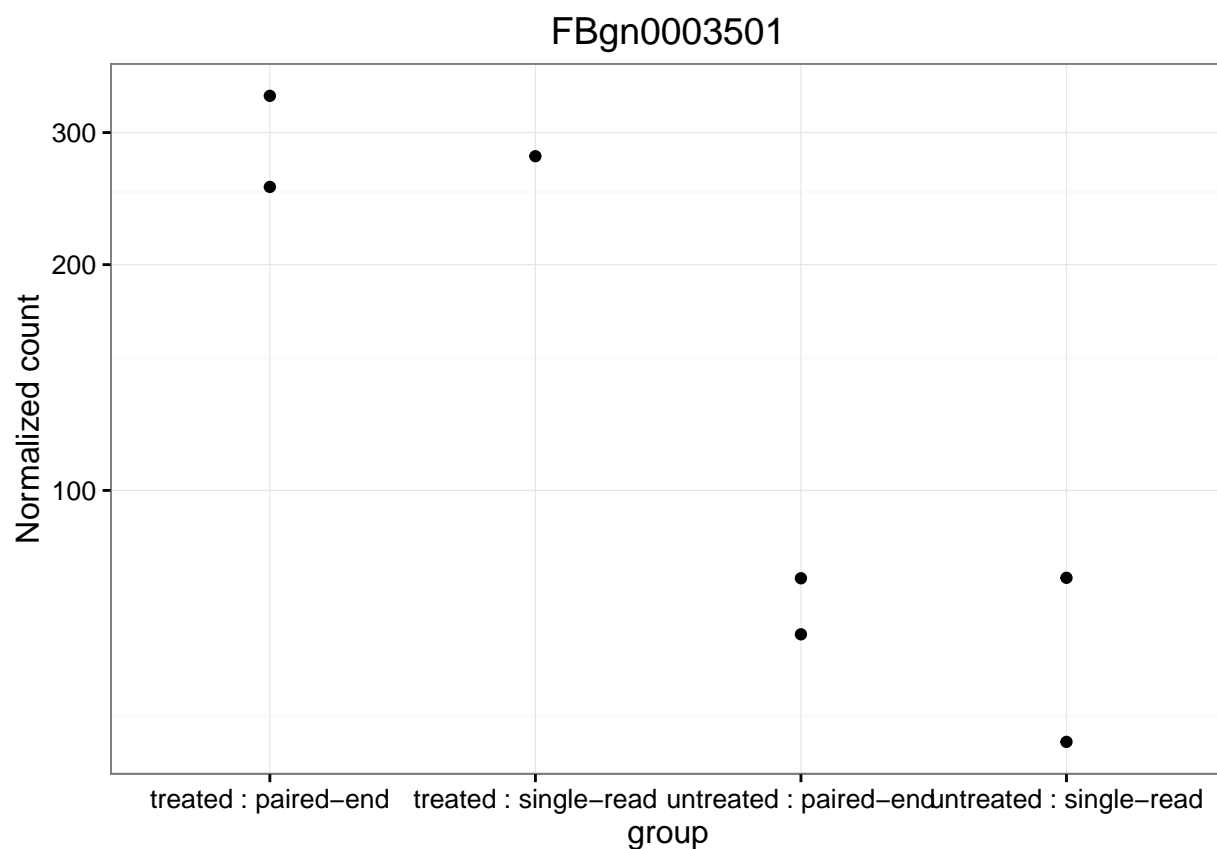












## Reproducibility

The input for this report was generated with DESeq2 (Love, Huber, and Anders, 2014) using version 1.11.36 and the resulting features were called significantly differentially expressed if their BH adjusted p-values were less than  $\alpha = 0.1$ . This report was generated in path /Users/lcollado/Dropbox/JHSPH/Code/regionReportSupp using the following call to DESeq2Report():

```
## DESeq2Report(dds = dds, project = "PDF report", intgroup = c("condition",  
##   "type"), outdir = "DESeq2-example", output = "DESeq2Report",  
##   device = "pdf", theme = theme_bw(), output_format = "pdf_document")
```

Date the report was generated.

```
## [1] "2016-04-01 01:22:52 EDT"
```

Wallclock time spent generating the report.

```
## Time difference of 26.562 secs
```

R session information.

```
## Session info -----
```

```
## setting  value  
## version  R version 3.3.0 alpha (2016-03-23 r70368)  
## system    x86_64, darwin13.4.0  
## ui        X11  
## language  (EN)  
## collate   en_US.UTF-8  
## tz        America/New_York  
## date      2016-04-01
```

```
## Packages -----
```

## package	* version	date	source
## acepack	1.3-3.3	2014-11-24	CRAN (R 3.3.0)
## annotate	1.49.1	2016-02-06	Bioconductor
## AnnotationDbi	1.33.7	2016-01-29	Bioconductor
## bibtex	0.4.0	2014-12-31	CRAN (R 3.3.0)
## Biobase	* 2.31.3	2016-01-14	Bioconductor
## BiocGenerics	* 0.17.3	2016-01-29	Bioconductor
## BiocParallel	1.5.21	2016-03-23	Bioconductor
## biomaRt	2.27.2	2016-01-14	Bioconductor
## Biostings	2.39.12	2016-02-21	Bioconductor
## bitops	1.0-6	2013-08-17	CRAN (R 3.3.0)
## BSgenome	1.39.4	2016-02-21	Bioconductor
## bumpHunter	1.11.5	2016-03-29	Bioconductor
## cluster	2.0.3	2015-07-21	CRAN (R 3.3.0)
## codetools	0.2-14	2015-07-15	CRAN (R 3.3.0)
## colorspace	1.2-6	2015-03-11	CRAN (R 3.3.0)
## DBI	0.3.1	2014-09-24	CRAN (R 3.3.0)

##	derfinder	1.5.30	2016-03-25	Bioconductor
##	derfinderHelper	1.5.3	2016-03-23	Bioconductor
##	DESeq	1.23.1	2016-01-14	Bioconductor
##	DESeq2	* 1.11.36	2016-03-31	Bioconductor
##	devtools	* 1.10.0	2016-01-23	CRAN (R 3.3.0)
##	digest	0.6.9	2016-01-08	CRAN (R 3.3.0)
##	doRNG	1.6	2014-03-07	CRAN (R 3.3.0)
##	DT	* 0.1	2015-06-09	CRAN (R 3.3.0)
##	evaluate	0.8.3	2016-03-05	CRAN (R 3.3.0)
##	foreach	1.4.3	2015-10-13	CRAN (R 3.3.0)
##	foreign	0.8-66	2015-08-19	CRAN (R 3.3.0)
##	formatR	1.3	2016-03-05	CRAN (R 3.3.0)
##	Formula	1.2-1	2015-04-07	CRAN (R 3.3.0)
##	genefilter	1.53.3	2016-03-23	Bioconductor
##	geneplotter	1.49.0	2016-01-14	Bioconductor
##	GenomeInfoDb	* 1.7.6	2016-01-29	Bioconductor
##	GenomicAlignments	1.7.20	2016-02-25	Bioconductor
##	GenomicFeatures	1.23.27	2016-03-25	Bioconductor
##	GenomicFiles	1.7.9	2016-02-22	Bioconductor
##	GenomicRanges	* 1.23.25	2016-03-31	Bioconductor
##	ggplot2	* 2.1.0	2016-03-01	CRAN (R 3.3.0)
##	gridExtra	2.2.1	2016-02-29	CRAN (R 3.3.0)
##	gtable	0.2.0	2016-02-26	CRAN (R 3.3.0)
##	highr	0.5.1	2015-09-18	CRAN (R 3.3.0)
##	Hmisc	3.17-2	2016-02-21	CRAN (R 3.3.0)
##	htmltools	0.3.5	2016-03-21	CRAN (R 3.3.0)
##	htmlwidgets	0.6	2016-02-25	CRAN (R 3.3.0)
##	httr	1.1.0	2016-01-28	CRAN (R 3.3.0)
##	IRanges	* 2.5.40	2016-03-11	Bioconductor
##	iterators	1.0.8	2015-10-13	CRAN (R 3.3.0)
##	knitcitations	1.0.7	2015-10-28	CRAN (R 3.3.0)
##	knitr	* 1.12.3	2016-01-22	CRAN (R 3.3.0)
##	knitrBootstrap	1.0.0	2016-03-24	Github (jimhester/knitrBootstrap@cdaa4a9)
##	labeling	0.3	2014-08-23	CRAN (R 3.3.0)
##	lattice	0.20-33	2015-07-14	CRAN (R 3.3.0)
##	latticeExtra	0.6-28	2016-02-09	CRAN (R 3.3.0)
##	locfit	1.5-9.1	2013-04-20	CRAN (R 3.3.0)
##	lubridate	1.5.0	2015-12-03	CRAN (R 3.3.0)
##	magrittr	1.5	2014-11-22	CRAN (R 3.3.0)
##	markdown	0.7.7	2015-04-22	CRAN (R 3.3.0)
##	Matrix	1.2-4	2016-03-02	CRAN (R 3.3.0)
##	matrixStats	0.50.1	2015-12-15	CRAN (R 3.3.0)
##	memoise	1.0.0	2016-01-29	CRAN (R 3.3.0)
##	munsell	0.4.3	2016-02-13	CRAN (R 3.3.0)
##	nnet	7.3-12	2016-02-02	CRAN (R 3.3.0)
##	pasilla	* 0.11.0	2016-03-31	Bioconductor
##	pheatmap	* 1.0.8	2015-12-11	CRAN (R 3.3.0)
##	pkgmaker	0.22	2014-05-14	CRAN (R 3.3.0)
##	plyr	1.8.3	2015-06-12	CRAN (R 3.3.0)
##	qvalue	2.3.2	2016-01-14	Bioconductor
##	R6	2.1.2	2016-01-26	CRAN (R 3.3.0)
##	RColorBrewer	* 1.1-2	2014-12-07	CRAN (R 3.3.0)
##	Rcpp	0.12.4	2016-03-26	CRAN (R 3.3.0)
##	RCurl	1.95-4.8	2016-03-01	CRAN (R 3.3.0)

##	RefManagerR	0.10.12	2016-03-31	CRAN (R 3.3.0)
##	regionReport	* 1.5.29	2016-04-01	Bioconductor
##	registry	0.3	2015-07-08	CRAN (R 3.3.0)
##	reshape2	1.4.1	2014-12-06	CRAN (R 3.3.0)
##	RJSONIO	1.3-0	2014-07-28	CRAN (R 3.3.0)
##	rmarkdown	0.9.5	2016-02-22	CRAN (R 3.3.0)
##	rngtools	1.2.4	2014-03-06	CRAN (R 3.3.0)
##	rpart	4.1-10	2015-06-29	CRAN (R 3.3.0)
##	Rsamtools	1.23.6	2016-03-23	Bioconductor
##	RSQLite	1.0.0	2014-10-25	CRAN (R 3.3.0)
##	rtracklayer	1.31.9	2016-03-29	Bioconductor
##	S4Vectors	* 0.9.44	2016-03-28	Bioconductor
##	scales	0.4.0	2016-02-26	CRAN (R 3.3.0)
##	stringi	1.0-1	2015-10-22	CRAN (R 3.3.0)
##	stringr	1.0.0	2015-04-30	CRAN (R 3.3.0)
##	SummarizedExperiment	* 1.1.22	2016-03-10	Bioconductor
##	survival	2.38-3	2015-07-02	CRAN (R 3.3.0)
##	VariantAnnotation	1.17.22	2016-03-31	Bioconductor
##	XML	3.98-1.4	2016-03-01	CRAN (R 3.3.0)
##	xtable	1.8-2	2016-02-05	CRAN (R 3.3.0)
##	XVector	0.11.7	2016-02-13	Bioconductor
##	yaml	2.1.13	2014-06-12	CRAN (R 3.3.0)
##	zlibbioc	1.17.1	2016-03-19	Bioconductor

## Bibliography

This report was created with **regionReport** (Collado-Torres, Jaffe, and Leek, 2015) using **rmarkdown** (Allaire, Cheng, Xie, McPherson, et al., 2016) while **knitr** (Xie, 2014) and **DT** (Xie, 2015) were running behind the scenes. **pheatmap** (Kolde, 2015) was used to create the sample distances heatmap. Several plots were made with **ggplot2** (Wickham, 2009).

Citations made with **knitcitations** (Boettiger, 2015). The **BibTeX** file can be found [here](#).

- [1] J. Allaire, J. Cheng, Y. Xie, J. McPherson, et al. **rmarkdown**: Dynamic Documents for R. R package version 0.9.5. 2016. URL: <https://CRAN.R-project.org/package=rmarkdown>.
- [2] C. Boettiger. **knitcitations**: Citations for ‘Knitr’ Markdown Files. R package version 1.0.7. 2015. URL: <https://CRAN.R-project.org/package=knitcitations>.
- [3] L. Collado-Torres, A. E. Jaffe and J. T. Leek. **regionReport**: Generate HTML reports for exploring a set of regions. <https://github.com/leekgroup/regionReport> - R package version 1.5.29. 2015. URL: <http://www.bioconductor.org/packages/regionReport>.
- [4] R. Kolde. **pheatmap**: Pretty Heatmaps. R package version 1.0.8. 2015. URL: <https://CRAN.R-project.org/package=pheatmap>.
- [5] M. I. Love, W. Huber and S. Anders. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome Biology* 15 (12 2014), p. 550. DOI: 10.1186/s13059-014-0550-8.
- [6] H. Wickham. **ggplot2**: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009. ISBN: 978-0-387-98140-6. URL: <http://ggplot2.org>.
- [7] Y. Xie. **DT**: A Wrapper of the JavaScript Library ‘DataTables’. R package version 0.1. 2015. URL: <https://CRAN.R-project.org/package=DT>.
- [8] Y. Xie. “knitr: A Comprehensive Tool for Reproducible Research in R”. In: *Implementing Reproducible Computational Research*. Ed. by V. Stodden, F. Leisch and R. D. Peng. ISBN 978-1466561595. Chapman and Hall/CRC, 2014. URL: <http://www.crcpress.com/product/isbn/9781466561595>.