

# 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	22

## Basic DESeq2 results exploration

Project: DESeq2 PDF report.

## Introduction

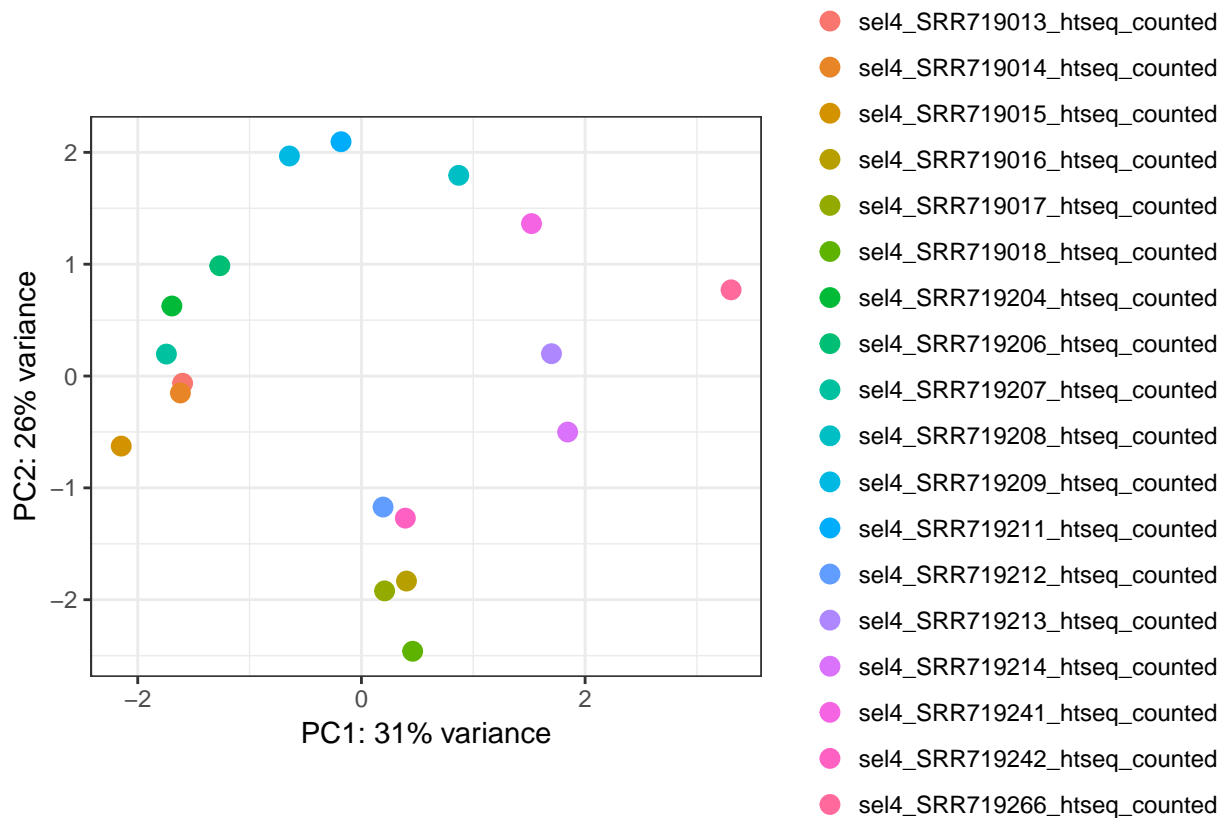
This report is meant to help explore DESeq2 (Love, Huber, and Anders, 2014) results and was generated using the `regionReport` (Collado-Torres, Jaffe, and Leek, 2016) package. While the report is rich, it is meant to just start the exploration of the results and exemplify some of the code used to do 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.

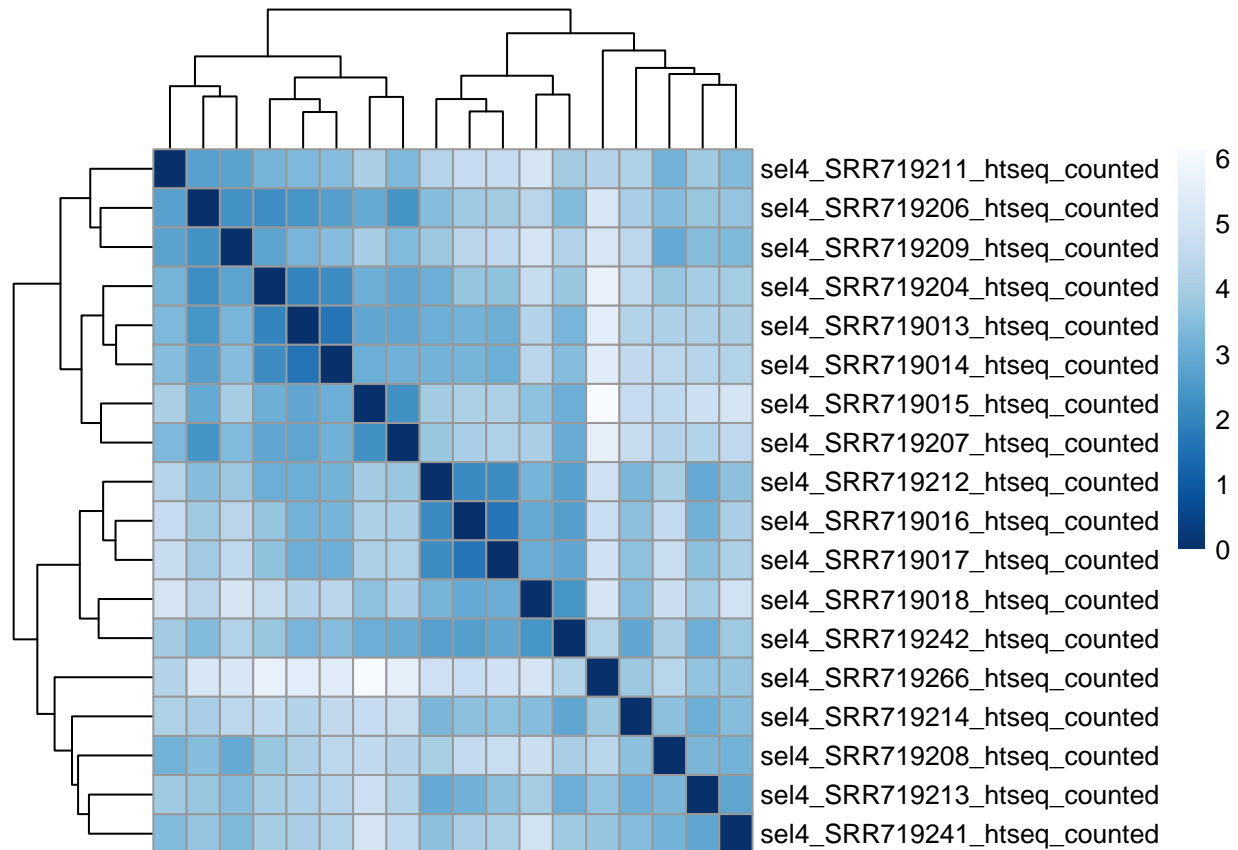
```
## knitrBootstrap and device chunk options
load_install('knitr')
opts_chunk$set(bootstrap.show.code = FALSE, dev = device)
if(!outputIsHTML) opts_chunk$set(bootstrap.show.code = FALSE, dev = device, echo = FALSE)
```

## 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 31 and 26 percent of the variance respectively.

## Sample-to-sample distances

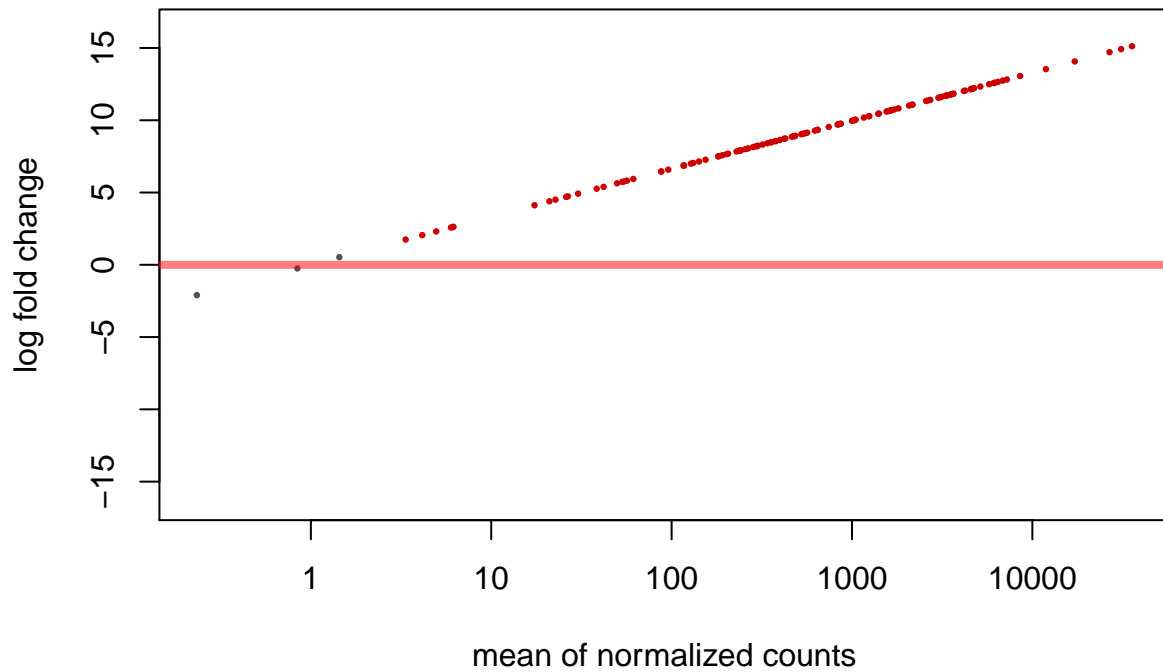


This plot shows how samples are clustered based on their euclidean distance using the regularized log transformed count data. This figure gives an overview of how the samples are hierarchically clustered. It is a complementary figure to the PCA plot.

## MA plots

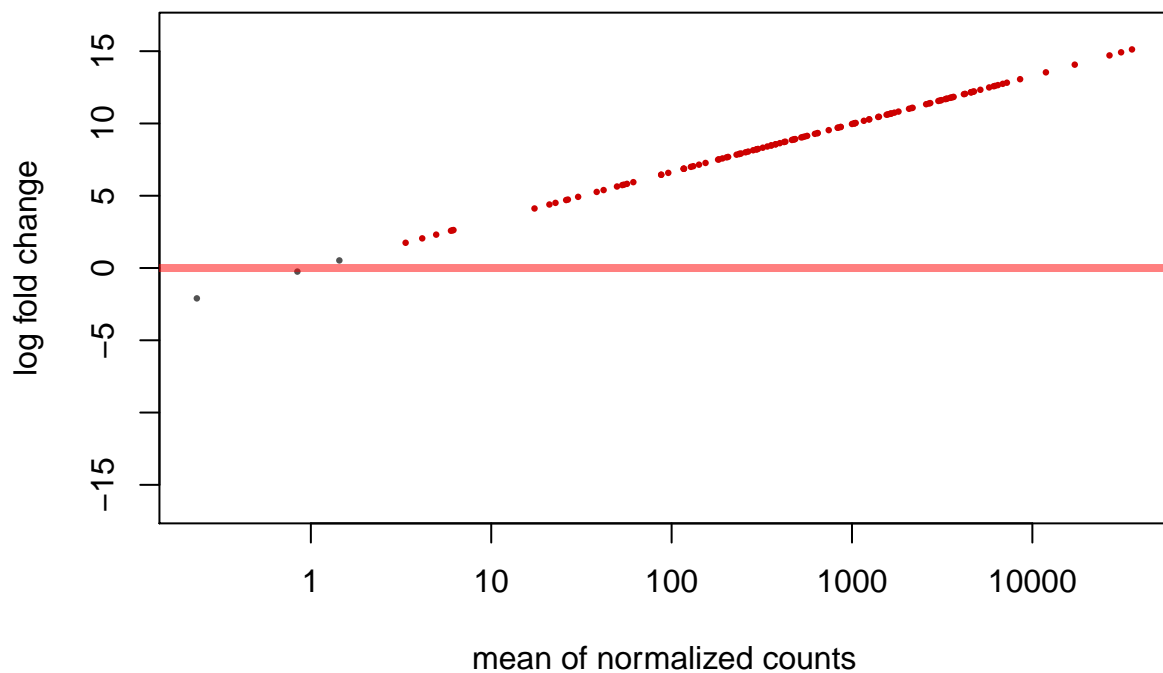
This section contains three MA plots (see Wikipedia) that compare the mean of the normalized counts against the log fold change. The points are shown in red if the feature has an adjusted p-value less than  $\alpha$ , that is, the statistically significant features are shown in red.

### MA plot with alpha = 0.1



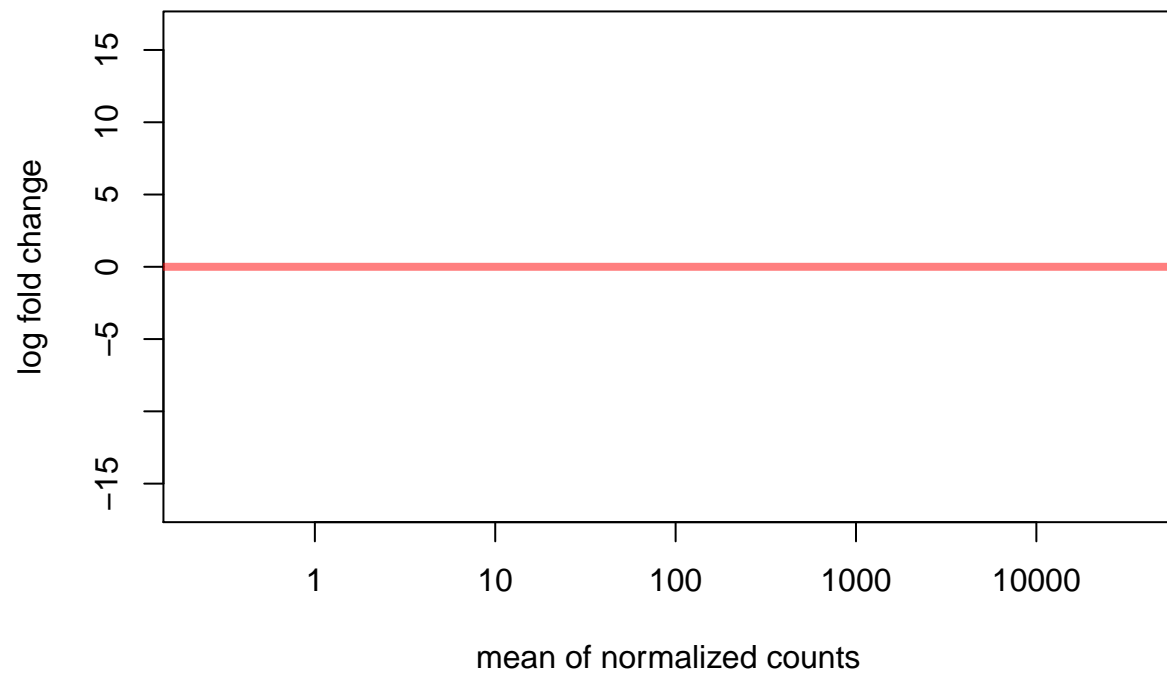
This first plot shows uses `alpha = 0.1`, which is the `alpha` value used to determine which resulting features were significant when running the function `DESeq2::results()`.

### MA plot with alpha = 0.05



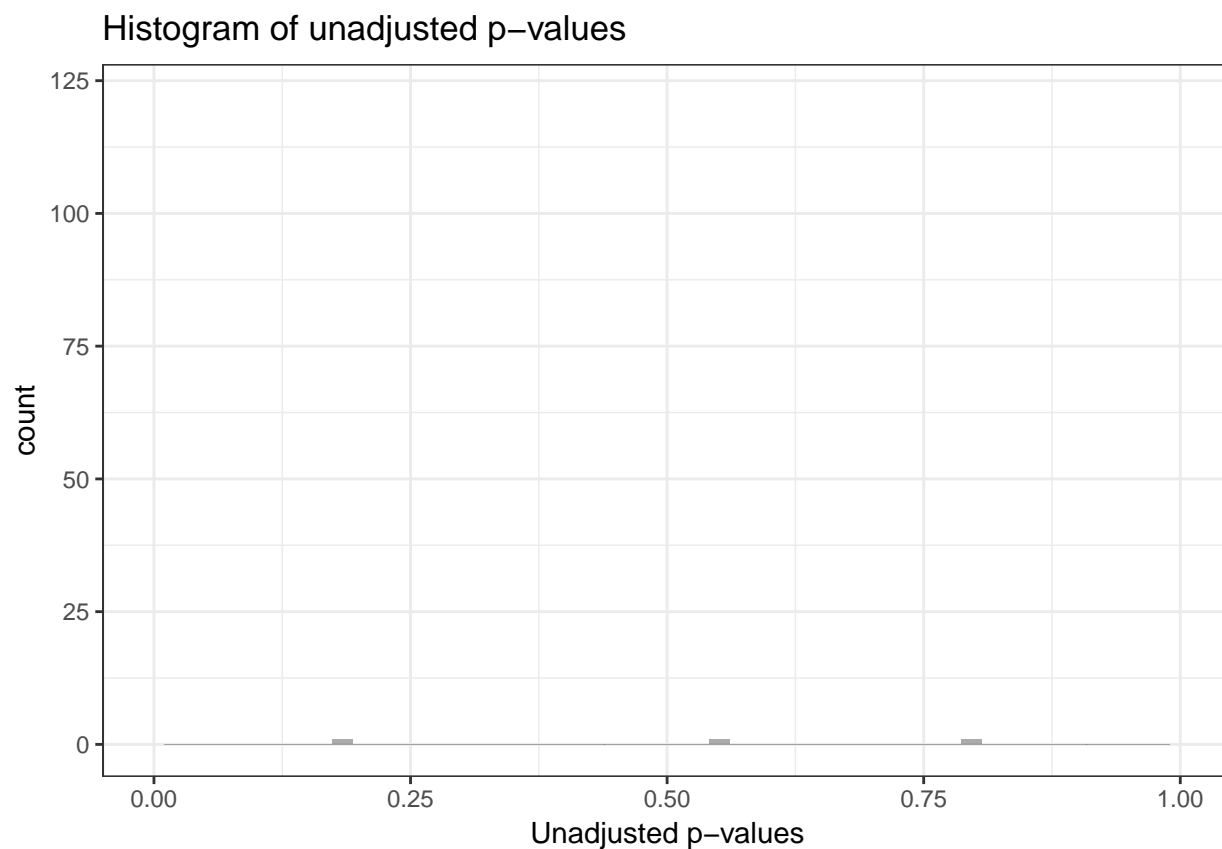
This second MA plot uses `alpha = 0.05` and can be used against the first MA plot to identify which features have adjusted p-values between 0.05 and 0.1.

### MA plot for top 126 features



The third and final MA plot uses an alpha such that the top 126 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.

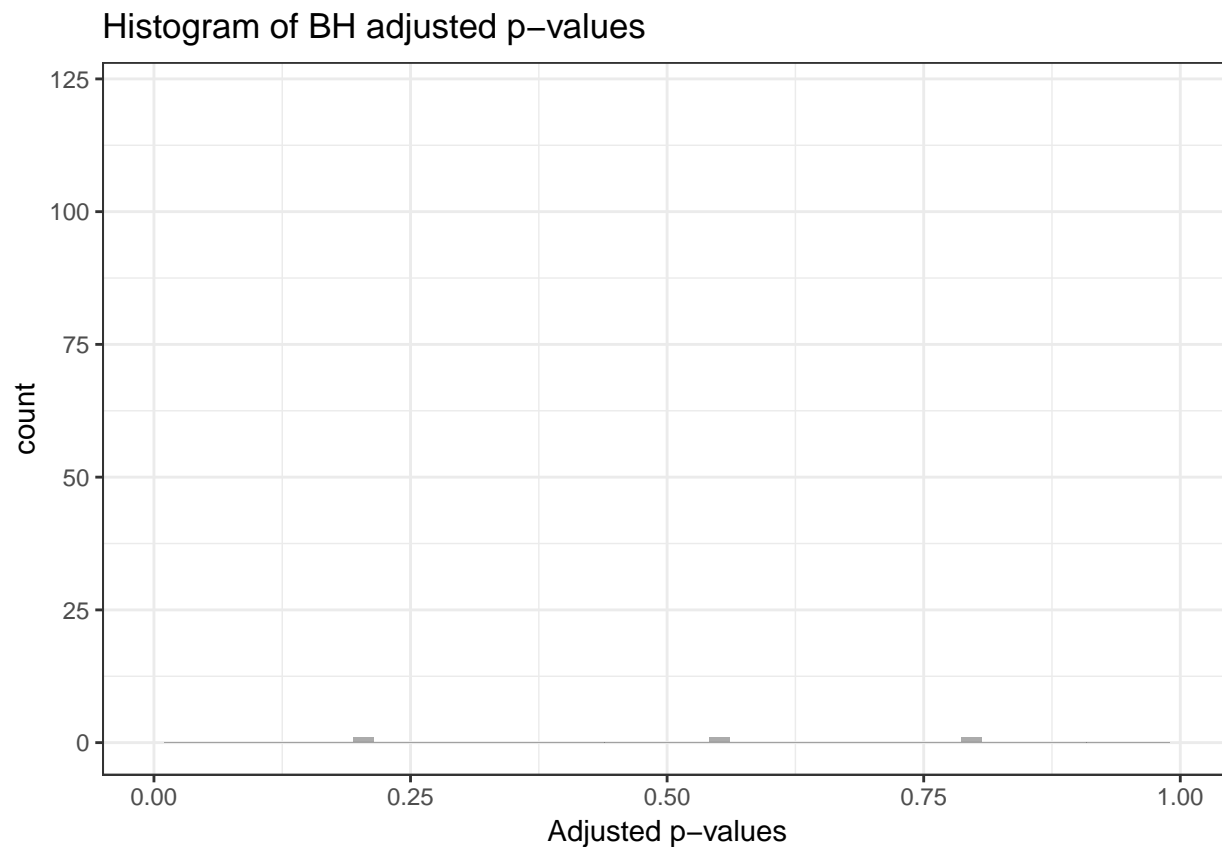
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
## 0.00000 0.00000 0.00000 0.01219 0.00000 0.78909         1
```

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

Cut	Count
0.0001	121
0.0010	122
0.0100	122
0.0250	122
0.0500	122
0.1000	122
0.2000	123
0.3000	123
0.4000	123
0.5000	123
0.6000	124
0.7000	124
0.8000	125
0.9000	125
1.0000	125

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
## 0.00000 0.00000 0.00000 0.01225 0.00000 0.78909      1
```

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

Cut	Count
0.0001	121
0.0010	122
0.0100	122
0.0250	122
0.0500	122
0.1000	122
0.2000	123
0.3000	123
0.4000	123
0.5000	123
0.6000	124
0.7000	124
0.8000	125

Cut	Count
0.9000	125
1.0000	125

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

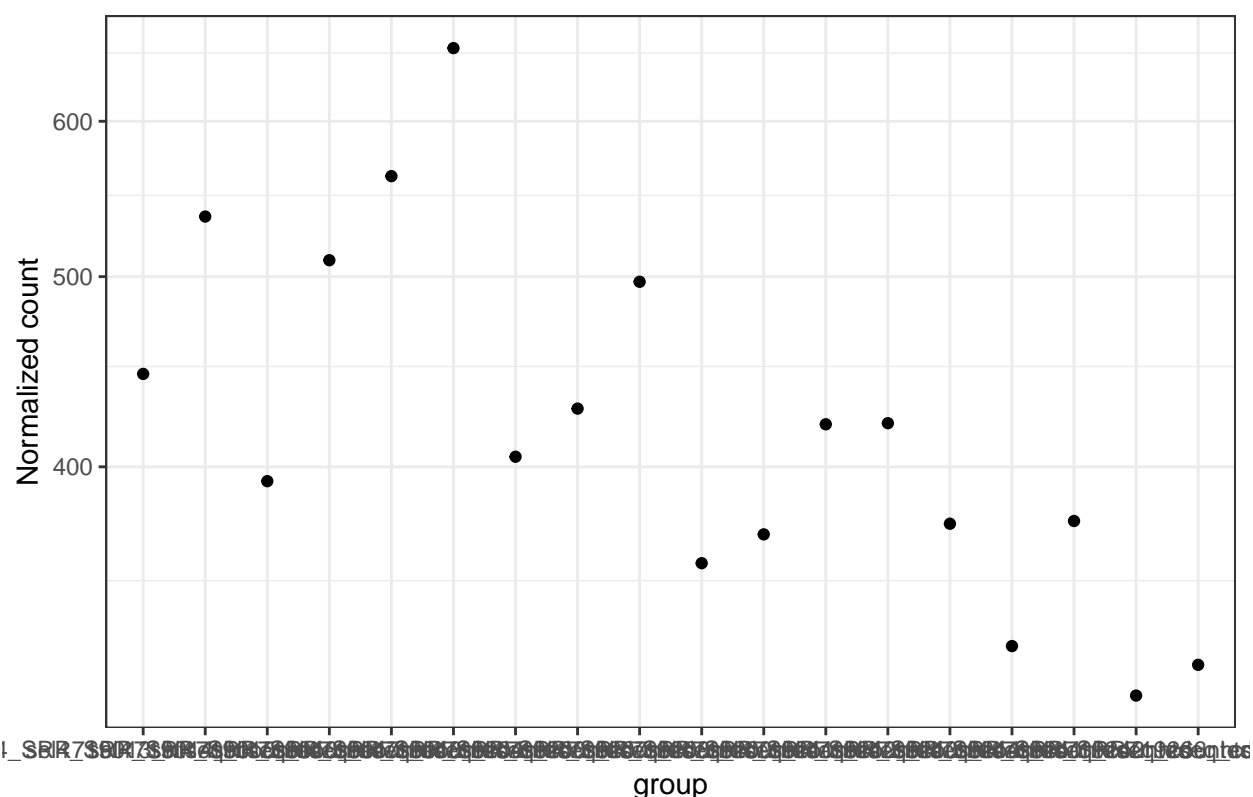
Feature	baseMean	log2FoldChange	lfcSE	stat	pvalue
augustus_masked-NW_015503979.1-processed-gene-10.4	427.11	8.74	0.07	117.03	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-16.0	3659.20	11.84	0.06	185.43	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-16.1	265.66	8.05	0.14	58.39	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-16.2	154.19	7.27	0.12	59.78	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-20.0	205.71	7.68	0.13	57.54	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-25.13	3370.57	11.72	0.08	146.73	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-29.0	356.88	8.48	0.12	69.01	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-29.1	237.09	7.89	0.15	52.12	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-38.0	339.32	8.41	0.13	64.62	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-39.0	41.97	5.39	0.14	39.00	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-44.0	2167.84	11.08	0.11	103.37	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-7.0	376.82	8.56	0.08	113.50	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-7.1	837.07	9.71	0.06	152.44	0.000000e+00
augustus_masked-NW_015503979.1-processed-gene-9.0	999.96	9.97	0.07	145.94	0.000000e+00
maker-NW_015503979.1-augustus-gene-10.0	229.12	7.84	0.14	55.53	0.000000e+00
maker-NW_015503979.1-augustus-gene-12.0	1644.28	10.68	0.08	132.86	0.000000e+00
maker-NW_015503979.1-augustus-gene-12.11	473.50	8.89	0.08	108.35	0.000000e+00
maker-NW_015503979.1-augustus-gene-12.13	534.20	9.06	0.10	88.59	0.000000e+00
maker-NW_015503979.1-augustus-gene-12.2	1028.35	10.01	0.12	84.97	0.000000e+00
maker-NW_015503979.1-augustus-gene-14.0	4747.61	12.21	0.10	126.99	0.000000e+00

## Count plots top features

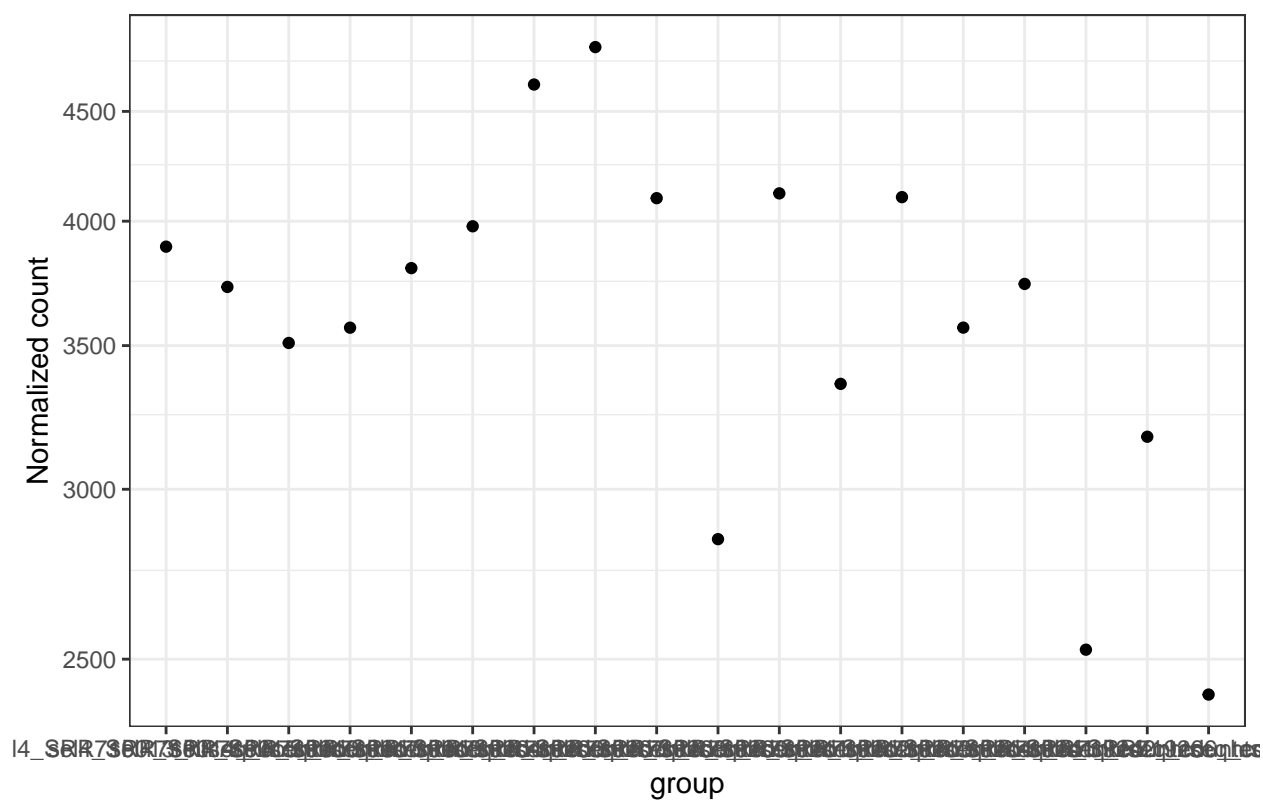
This section contains plots showing the normalized counts per sample for each group of interest. Only the best 20 features are shown, ranked by their BH adjusted p-values. The Y axis is on the log10 scale and the feature name is shown in the title of each plot.



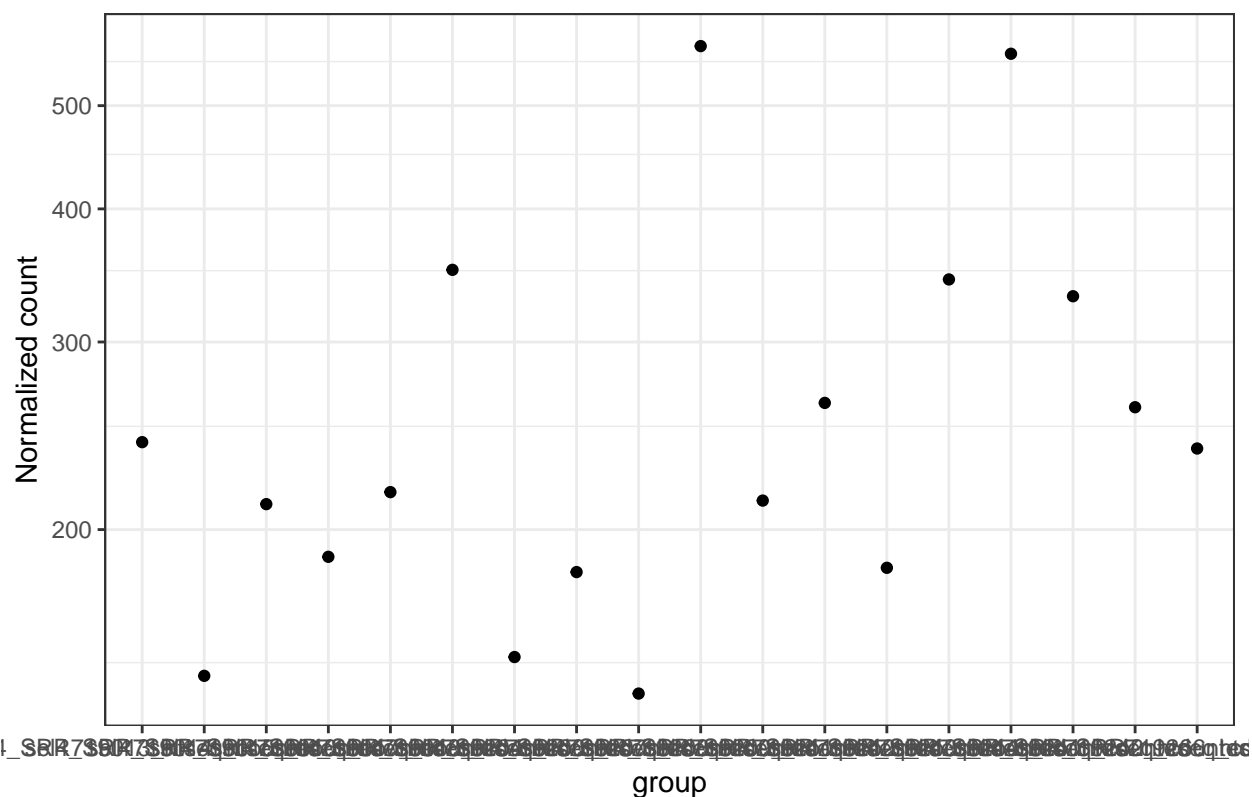
augustus\_masked-NW\_015503979.1-processed-gene-10.4



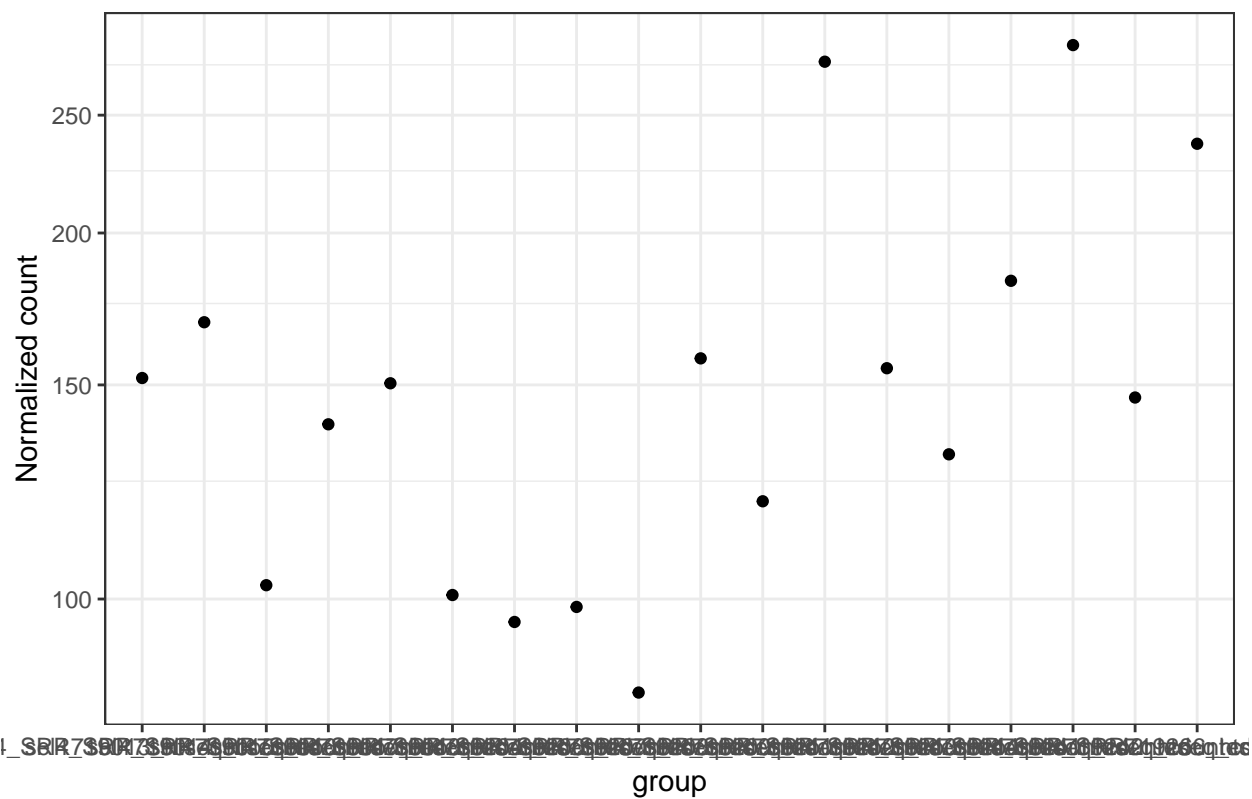
augustus\_masked-NW\_015503979.1-processed-gene-16.0



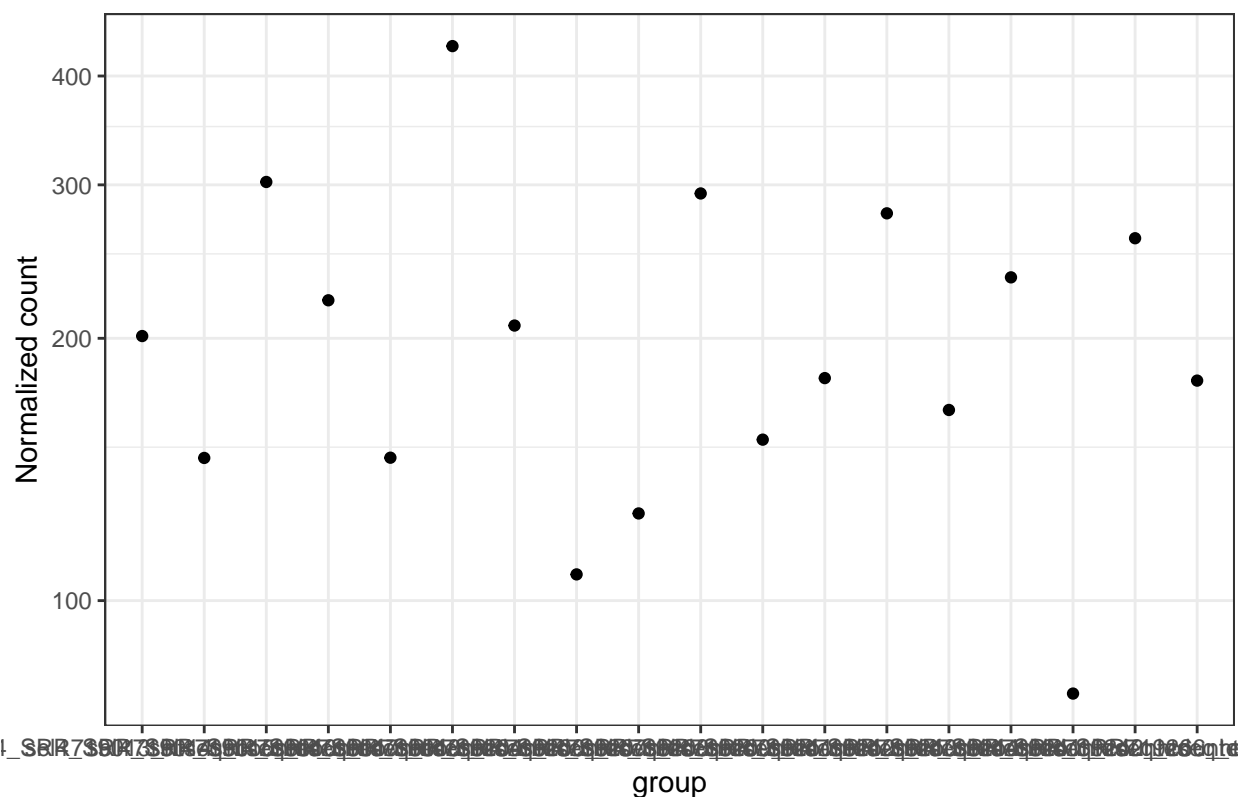
augustus\_masked-NW\_015503979.1-processed-gene-16.1



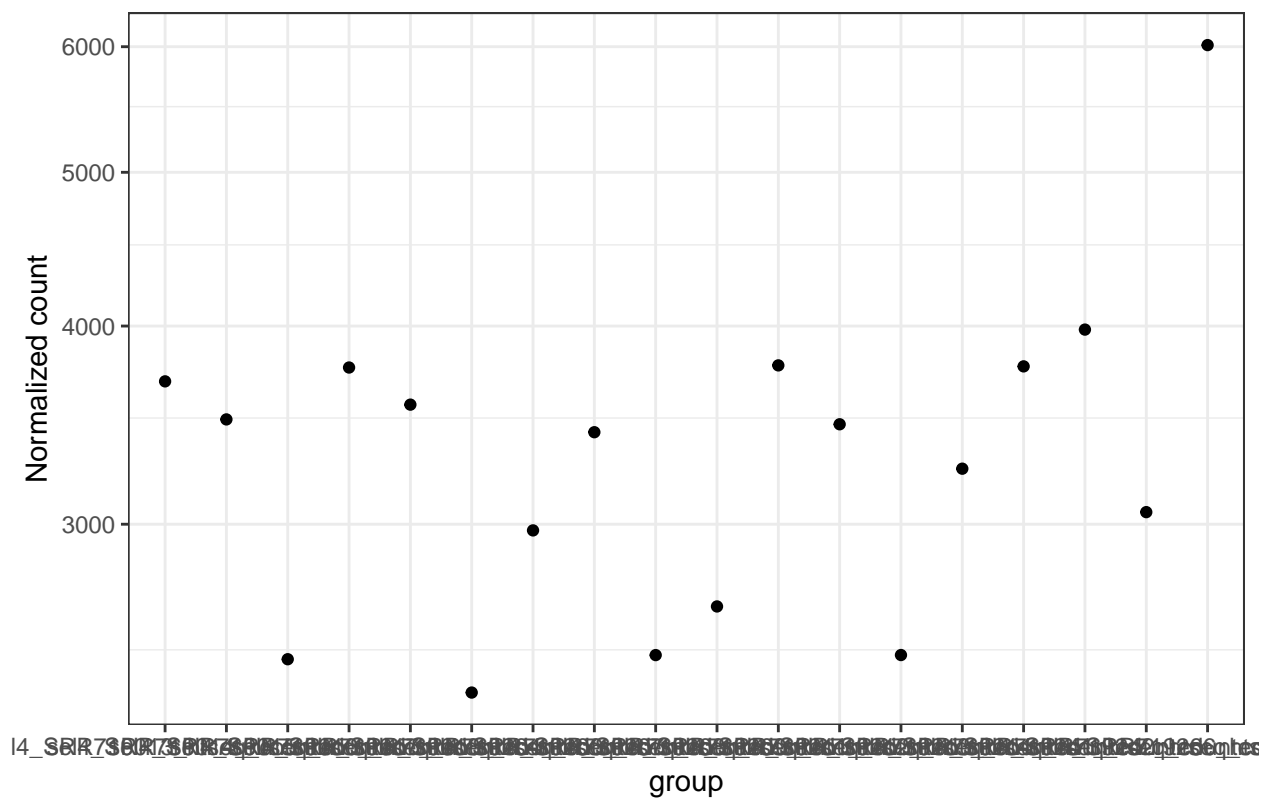
augustus\_masked-NW\_015503979.1-processed-gene-16.2



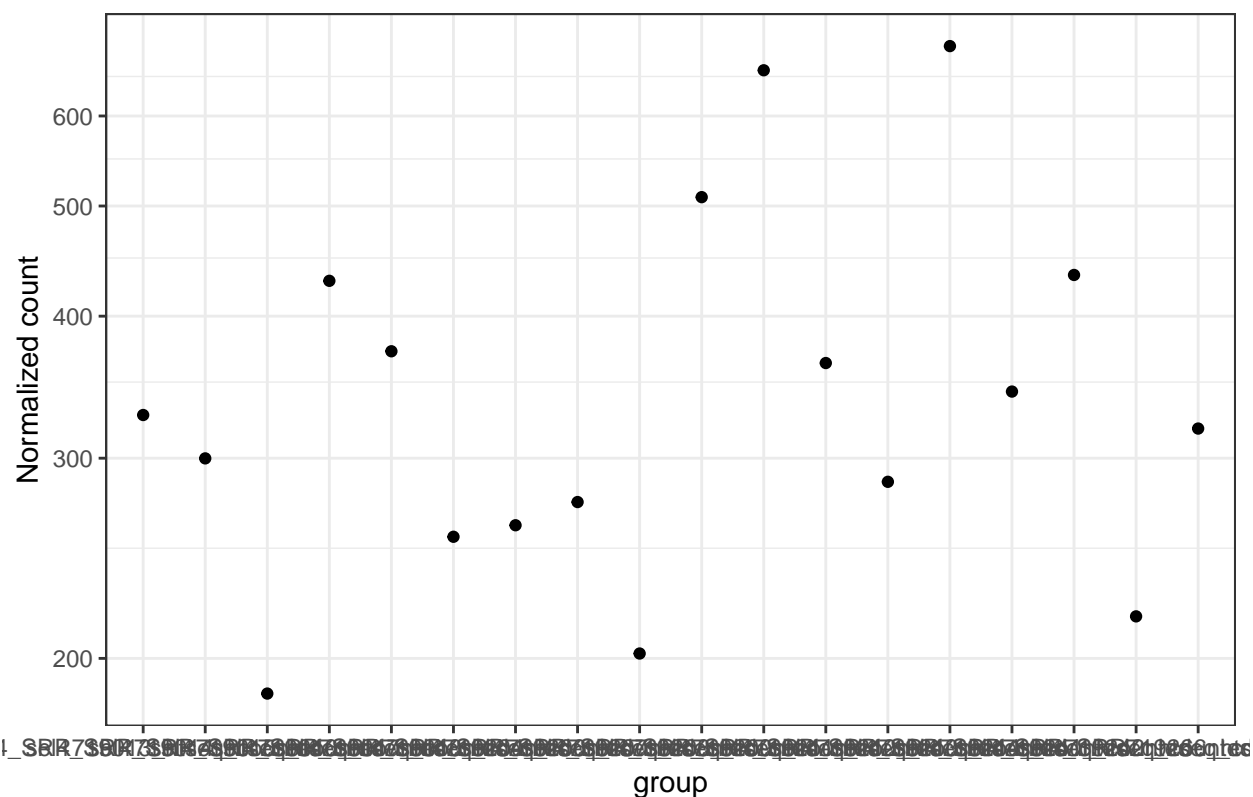
augustus\_masked-NW\_015503979.1-processed-gene-20.0



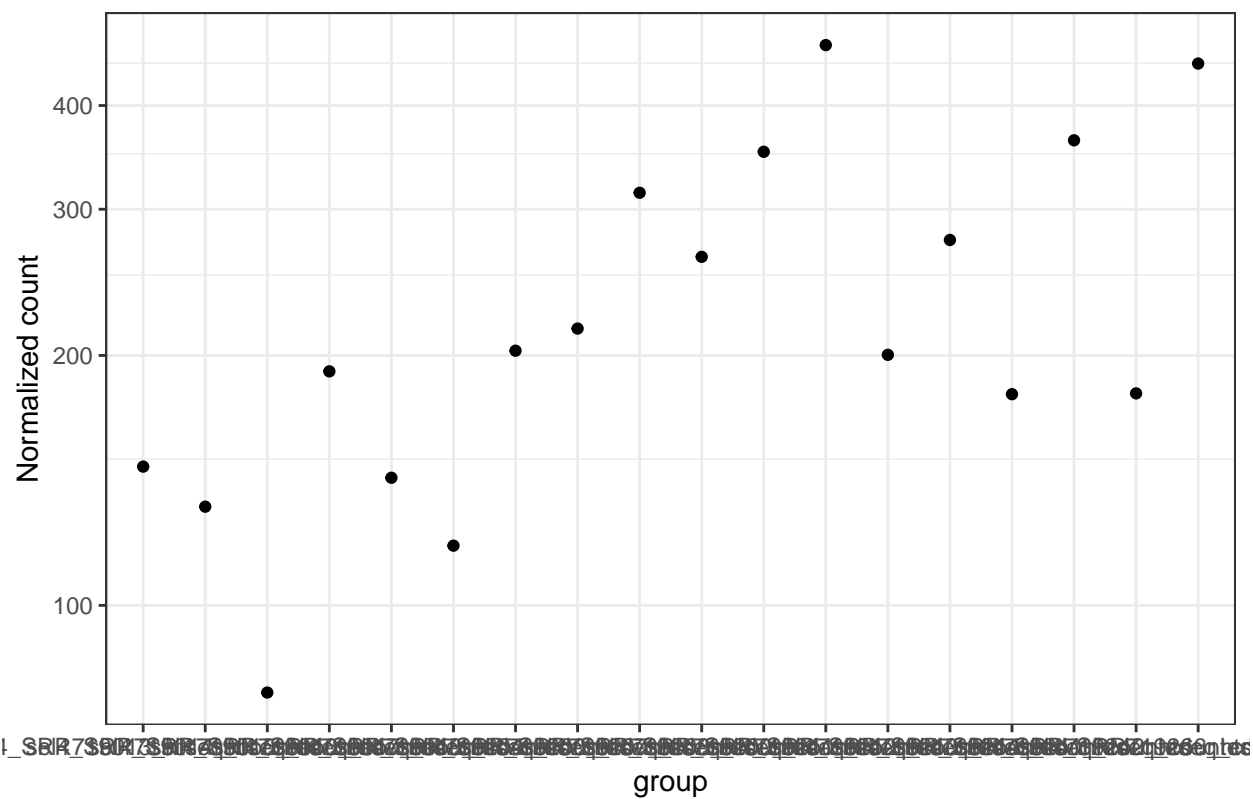
augustus\_masked-NW\_015503979.1-processed-gene-25.13



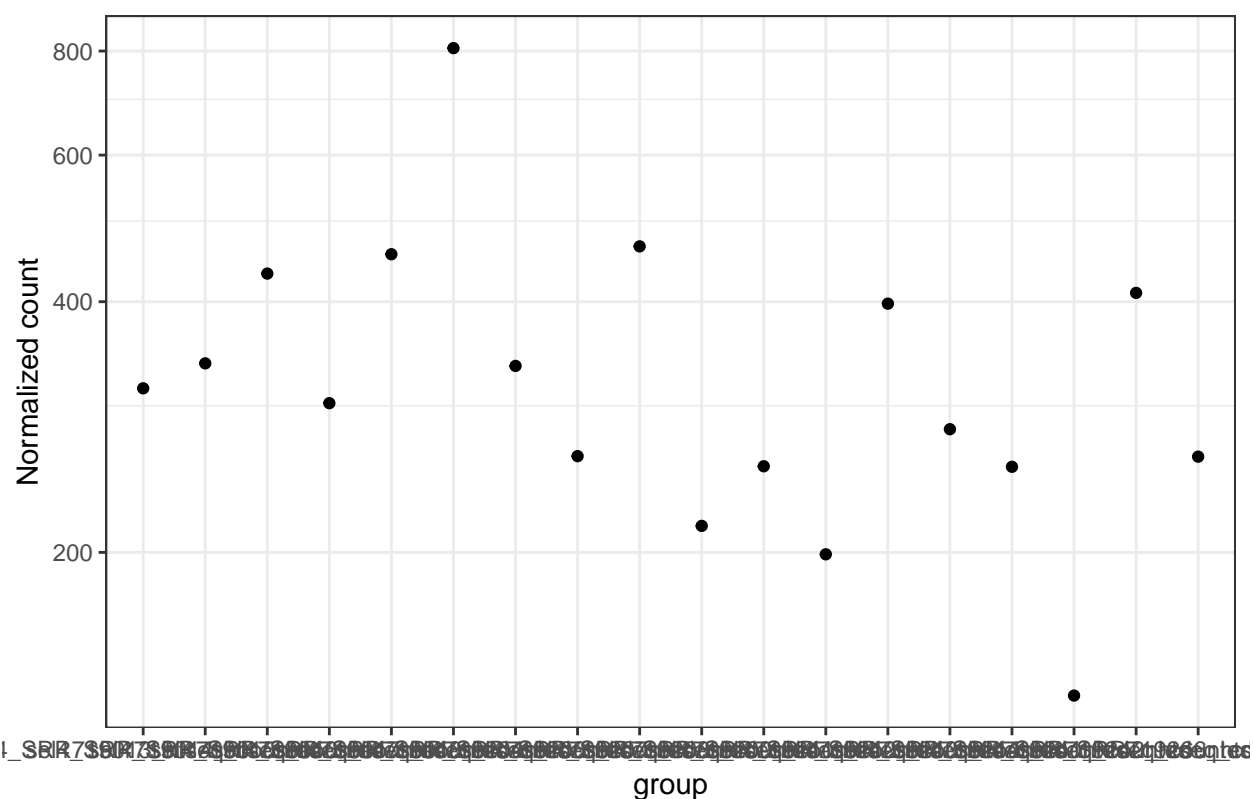
augustus\_masked-NW\_015503979.1-processed-gene-29.0



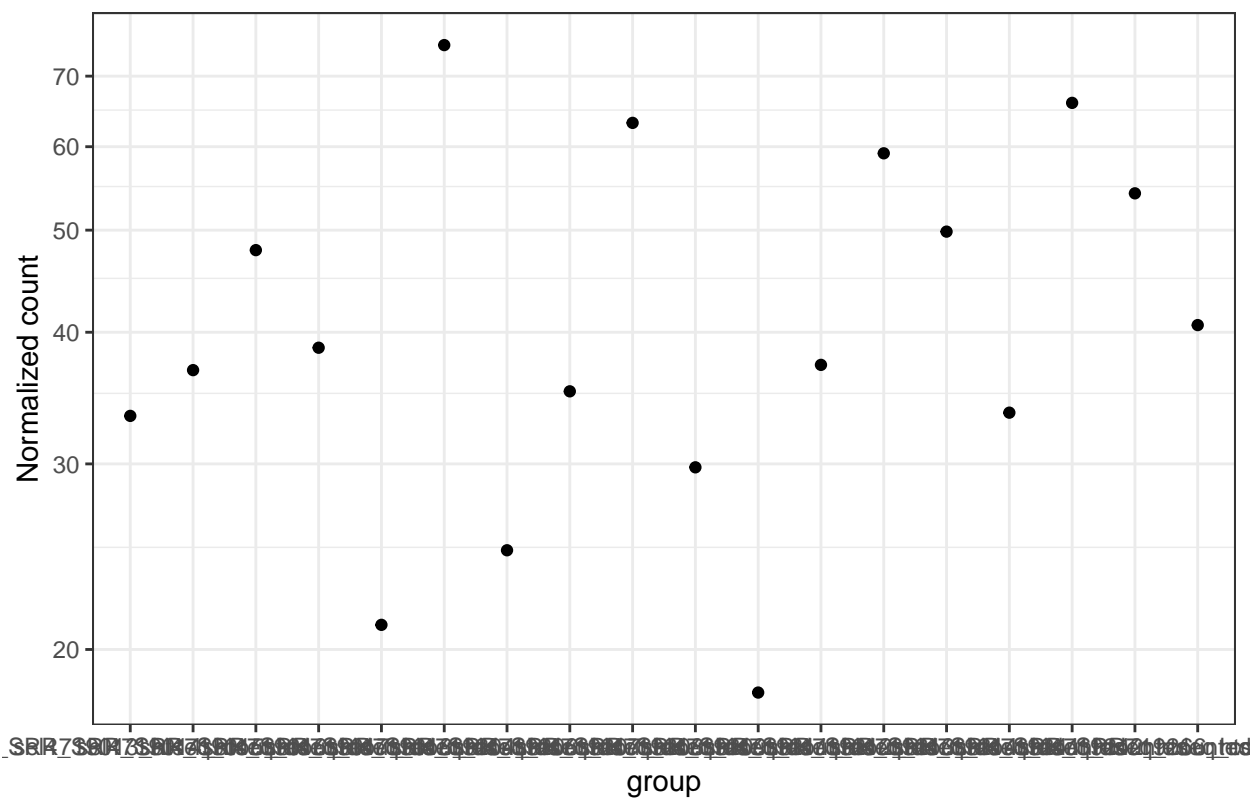
augustus\_masked-NW\_015503979.1-processed-gene-29.1



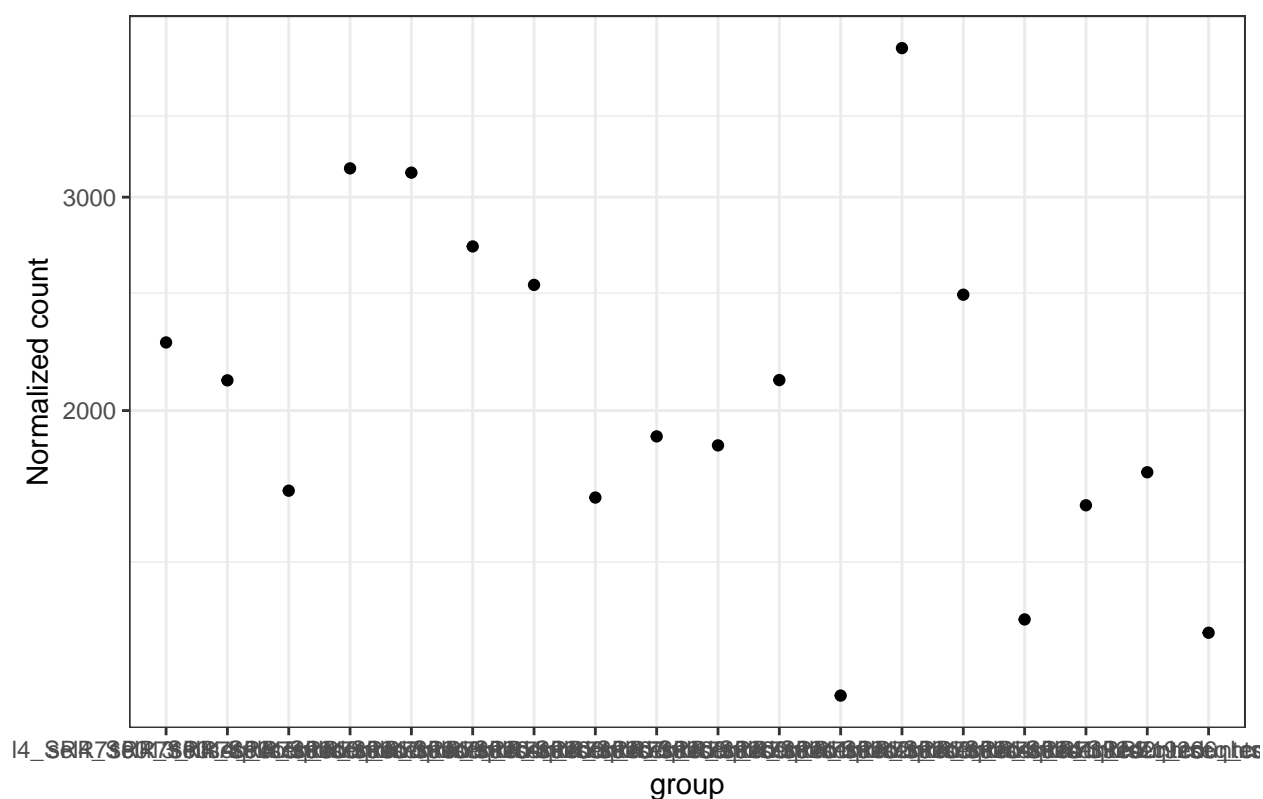
augustus\_masked-NW\_015503979.1-processed-gene-38.0



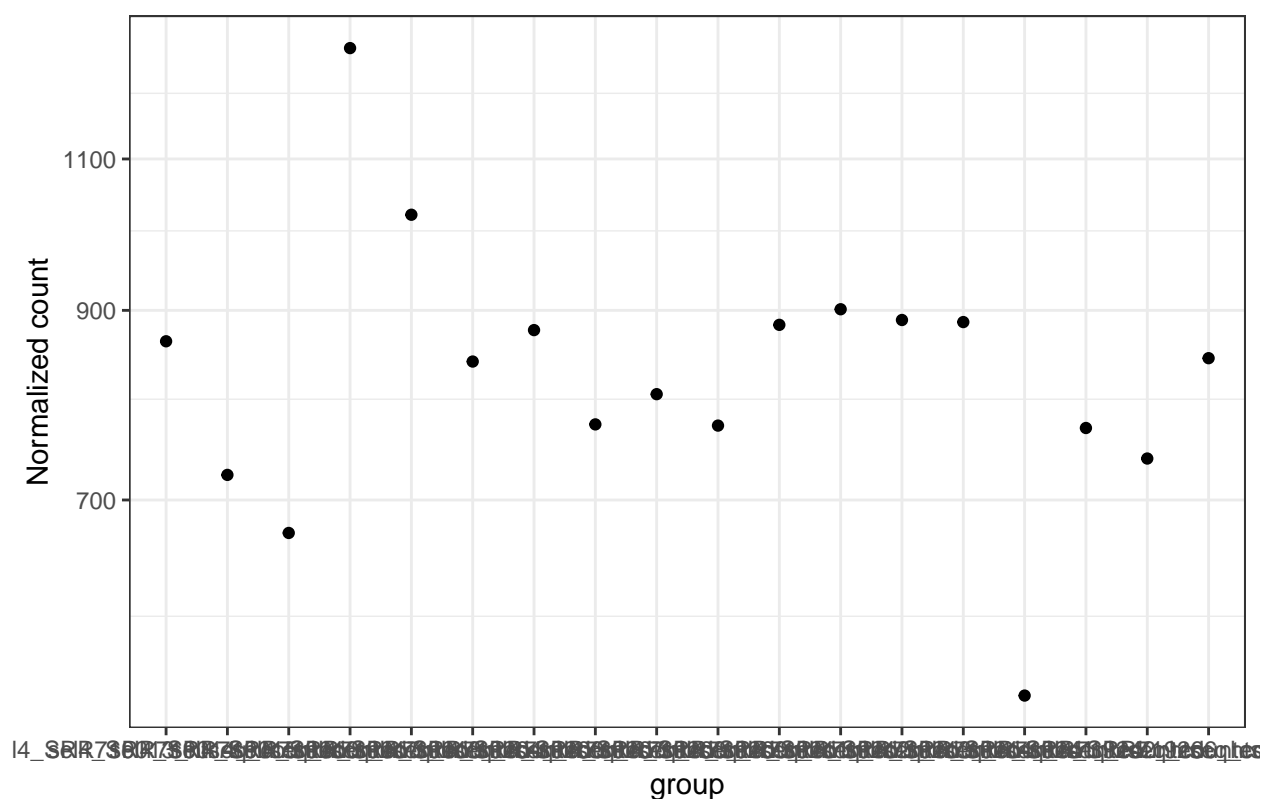
augustus\_masked-NW\_015503979.1-processed-gene-39.0



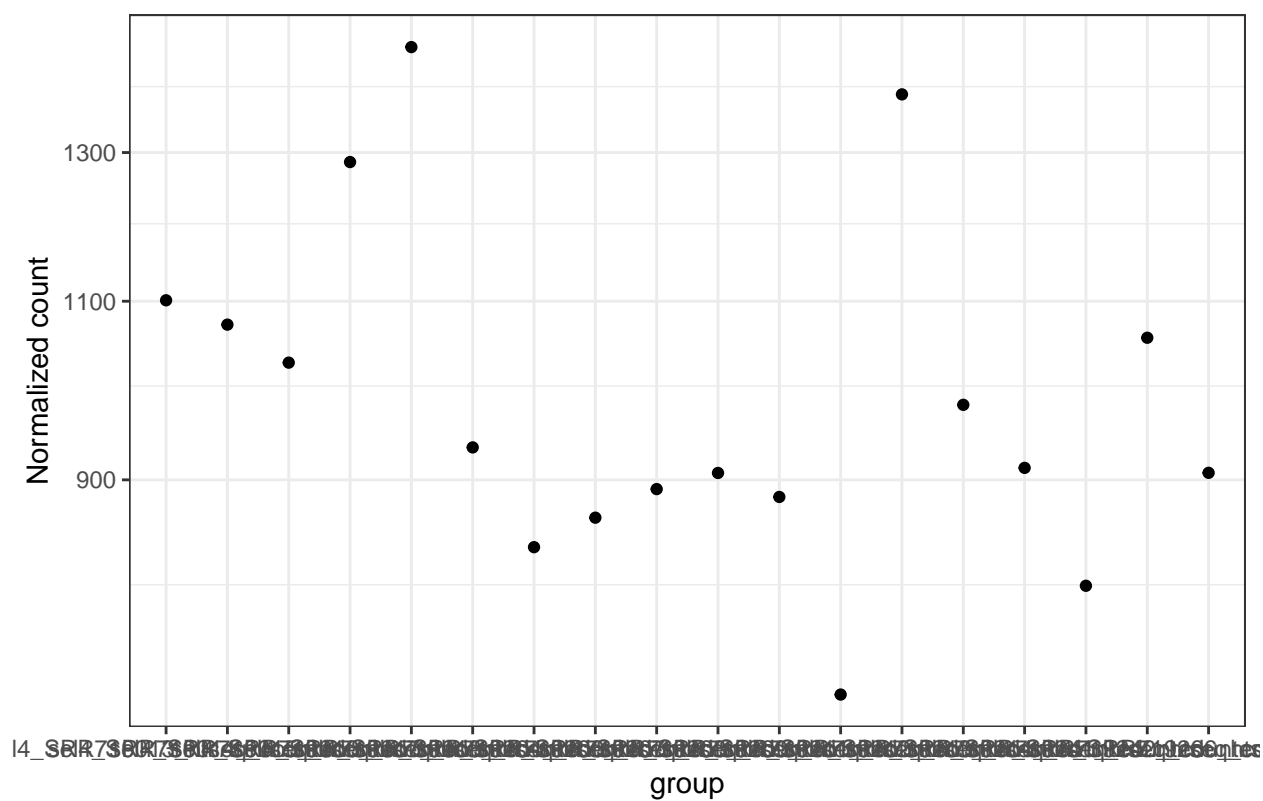
augustus\_masked-NW\_015503979.1-processed-gene-44.0



augustus\_masked-NW\_015503979.1-processed-gene-7.1



augustus\_masked-NW\_015503979.1-processed-gene-9.0



A scatter plot showing the relationship between the number of individuals (N) on the x-axis and the number of species (S) on the y-axis. The plot includes a grid and 10 data points represented by black dots. The data points are approximately as follows:

Sample	Number of Individuals (N)	Number of Species (S)
1	10	10
2	20	15
3	30	12
4	40	15
5	50	25
6	60	10
7	70	12
8	80	20
9	90	8
10	100	10

group

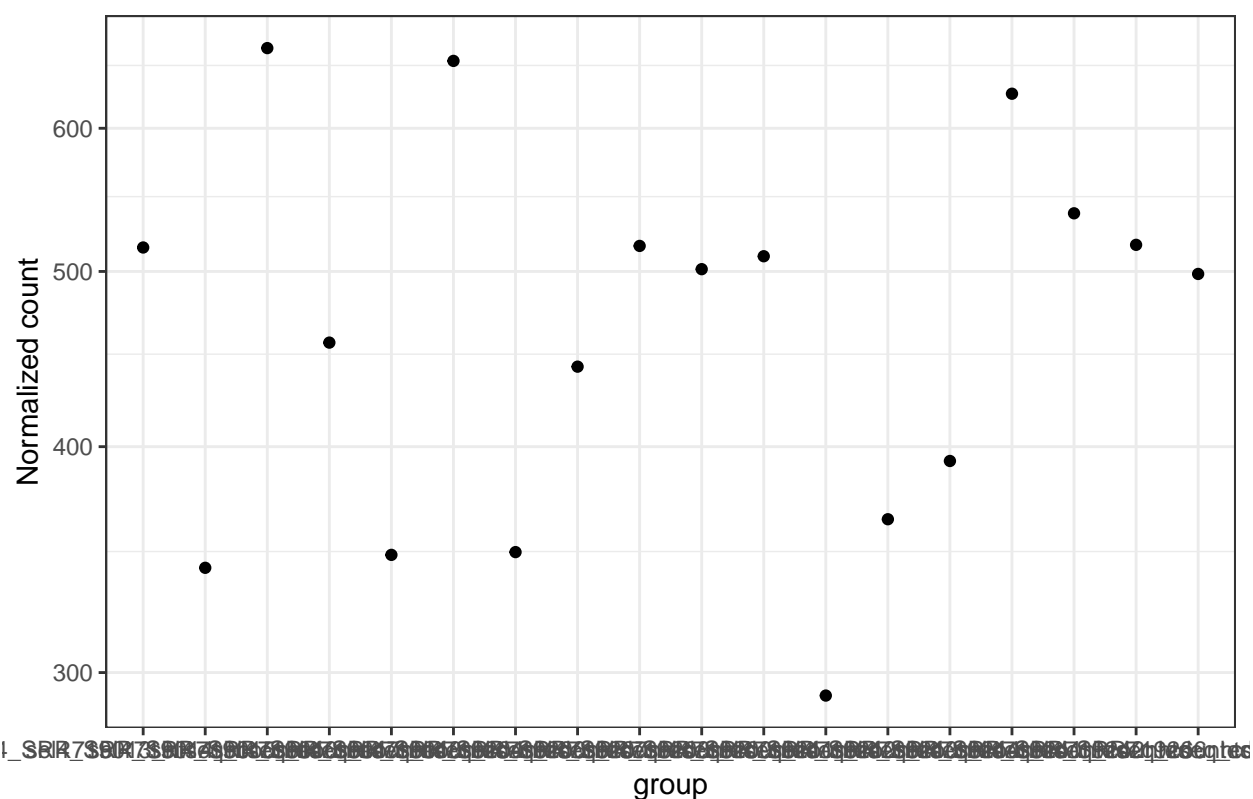
A scatter plot with 15 data points on a grid. The points are distributed across the grid, with a notable concentration in the upper right quadrant. The points are represented by small black dots.

Point ID	X (approx)	Y (approx)
1	1.0	9.0
2	2.0	8.0
3	3.0	7.0
4	4.0	6.0
5	5.0	9.0
6	6.0	5.0
7	7.0	6.0
8	8.0	4.0
9	9.0	7.0
10	10.0	8.0
11	11.0	5.0
12	12.0	6.0
13	13.0	4.0
14	14.0	3.0
15	15.0	2.0

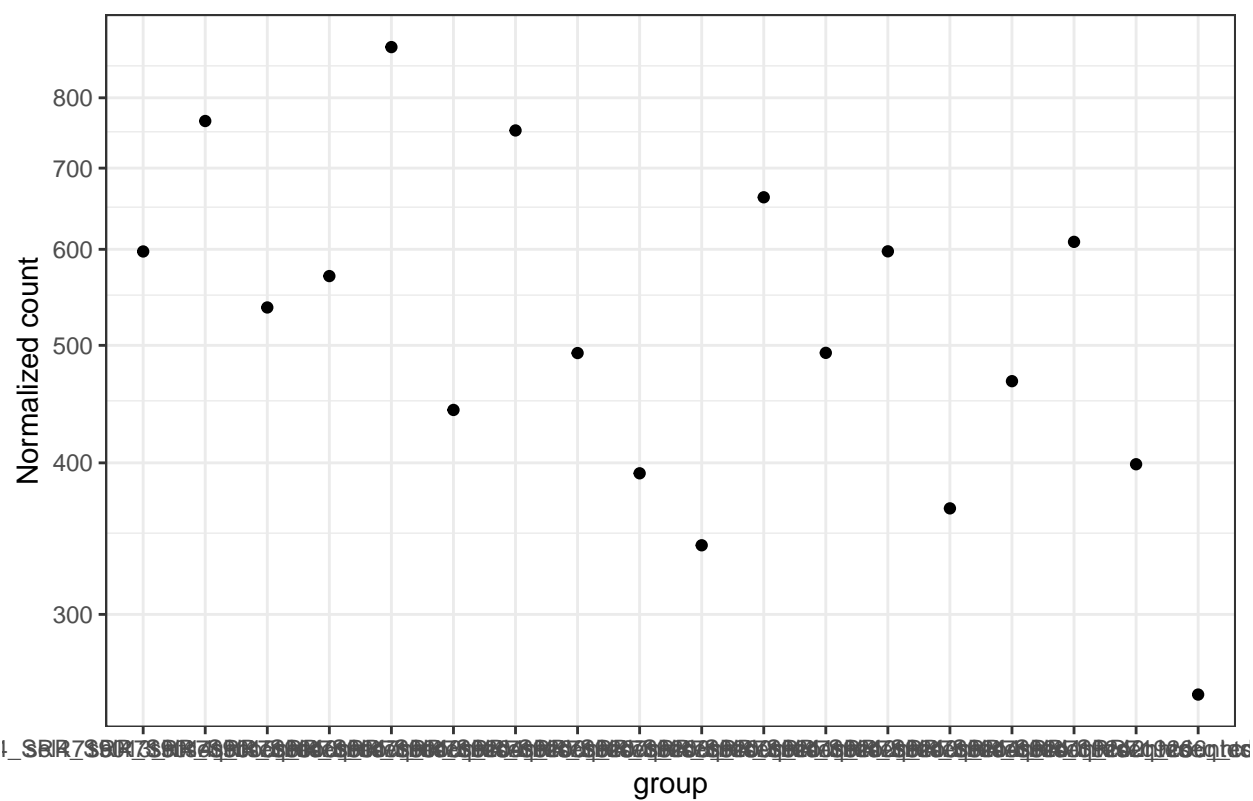
group



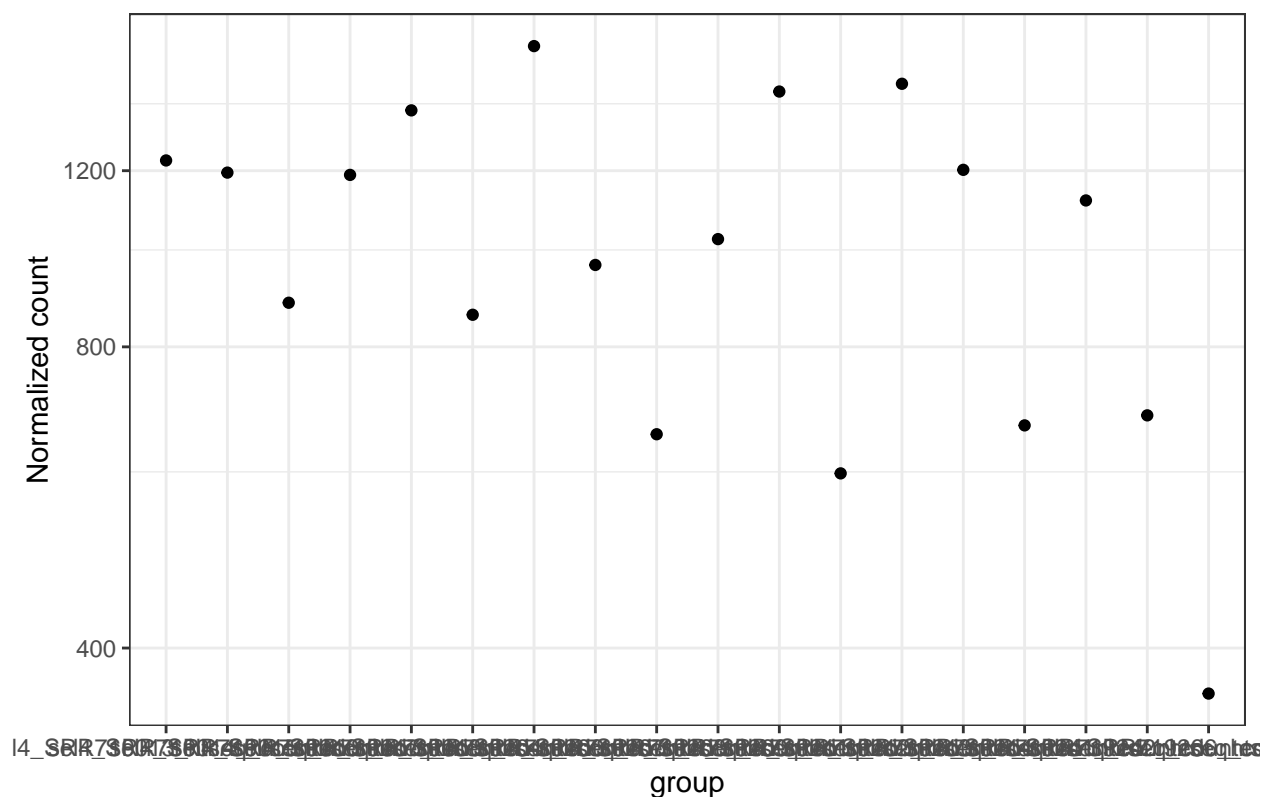
maker-NW\_015503979.1-augustus-gene-12.11



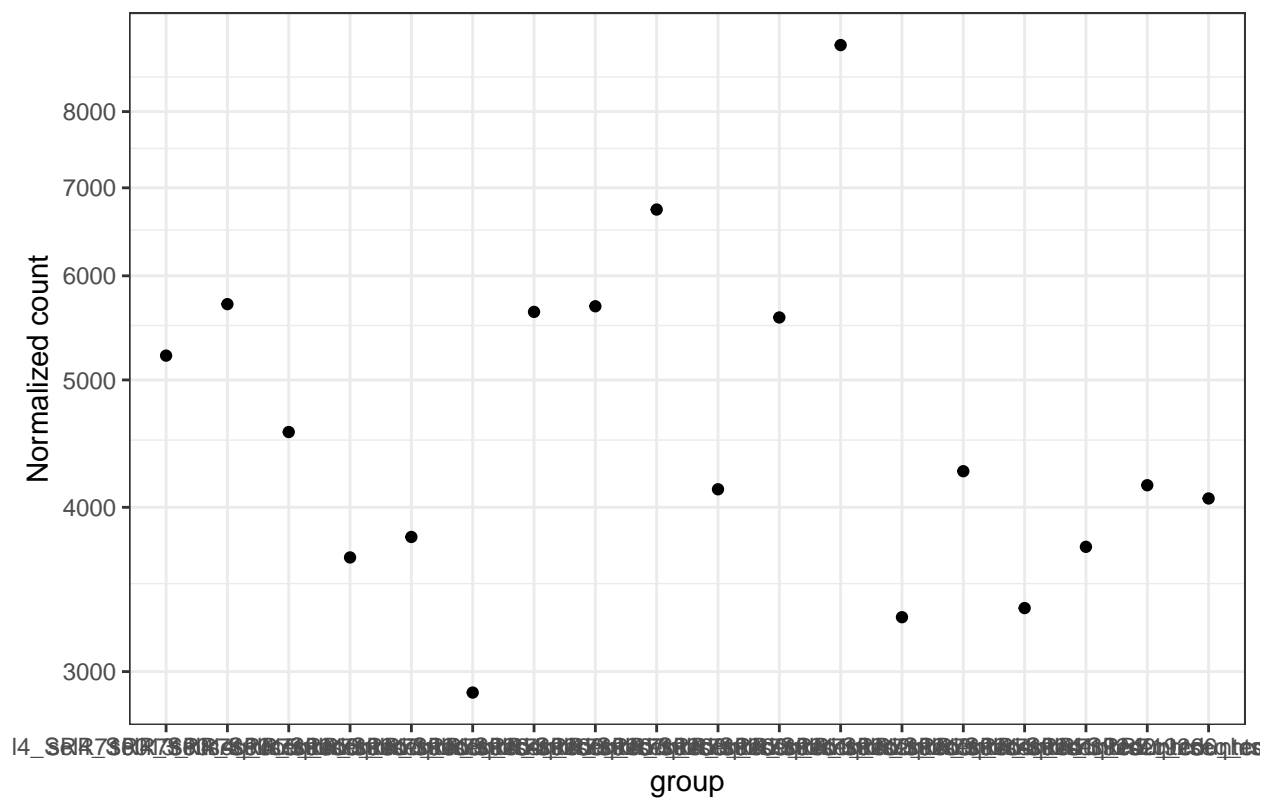
maker-NW\_015503979.1-augustus-gene-12.13



maker-NW\_015503979.1-augustus-gene-12.2



maker-NW\_015503979.1-augustus-gene-14.0



## Reproducibility

The input for this report was generated with DESeq2 (Love, Huber, and Anders, 2014) using version 1.18.1 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 /mnt/c/Genome\_Analysis/Results/HTSEQ\_RESULTS using the following call to DESeq2Report():

```
## DESeq2Report(dds = dds, project = "DESeq2 PDF report", intgroup = "condition",  
##   outdir = "/mnt/c/Genome_Analysis/Results/DESEQ2_RESULTS/",  
##   output = "deseq2_pdf", device = "pdf", theme = theme_bw(),  
##   output_format = "pdf_document")
```

Date the report was generated.

```
## [1] "2018-05-23 12:52:19 DST"
```

Wallclock time spent generating the report.

```
## Time difference of 20.796 secs
```

R session information.

```
## Warning: running command 'timedatectl' had status 1
```

```
## Warning: Your system is mis-configured: '/etc/localtime' is not a symlink
```

```
## Warning: It is strongly recommended to set environment variable TZ to 'localtime' (or equivalent)
```

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

```
## setting value  
## version R version 3.4.4 (2018-03-15)  
## system x86_64, linux-gnu  
## ui X11  
## language (EN)  
## collate en_US.UTF-8  
## tz localtime  
## date 2018-05-23
```

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

## package	* version	date	source
## acepack	1.4.1	2016-10-29	CRAN (R 3.4.4)
## annotate	1.56.2	2018-05-22	Bioconductor
## AnnotationDbi	1.40.0	2018-05-22	Bioconductor
## assertthat	0.2.0	2017-04-11	CRAN (R 3.4.4)
## backports	1.1.2	2017-12-13	CRAN (R 3.4.4)
## base	* 3.4.4	2018-03-16	local
## base64enc	0.1-3	2015-07-28	CRAN (R 3.4.4)
## bibtex	0.4.2	2017-06-30	CRAN (R 3.4.4)
## Biobase	* 2.38.0	2018-05-22	Bioconductor
## BiocGenerics	* 0.24.0	2018-05-23	Bioconductor
## BiocParallel	1.12.0	2018-05-22	Bioconductor
## BiocStyle	* 2.6.1	2018-05-23	Bioconductor
## biomaRt	2.34.2	2018-05-23	Bioconductor
## Biostrings	2.46.0	2018-05-23	Bioconductor
## bit	1.1-13	2018-05-15	CRAN (R 3.4.4)
## bit64	0.9-7	2017-05-08	CRAN (R 3.4.4)
## bitops	1.0-6	2013-08-17	CRAN (R 3.4.4)
## blob	1.1.1	2018-03-25	CRAN (R 3.4.4)

##	bookdown	0.7	2018-02-18	CRAN (R 3.4.4)
##	BSgenome	1.46.0	2018-05-23	Bioconductor
##	bumphunter	1.20.0	2018-05-23	Bioconductor
##	calibrate	* 1.7.2	2013-09-10	CRAN (R 3.4.4)
##	checkmate	1.8.5	2017-10-24	CRAN (R 3.4.4)
##	cluster	2.0.7-1	2018-04-09	CRAN (R 3.4.4)
##	codetools	0.2-15	2016-10-05	CRAN (R 3.4.4)
##	colorspace	1.3-2	2016-12-14	CRAN (R 3.4.4)
##	compiler	3.4.4	2018-03-16	local
##	crosstalk	1.0.0	2016-12-21	CRAN (R 3.4.4)
##	curl	3.2	2018-03-28	CRAN (R 3.4.4)
##	data.table	1.11.2	2018-05-08	CRAN (R 3.4.4)
##	datasets	* 3.4.4	2018-03-16	local
##	DBI	1.0.0	2018-05-02	CRAN (R 3.4.4)
##	DEFormats	1.6.1	2018-05-23	Bioconductor
##	DelayedArray	* 0.4.1	2018-05-22	Bioconductor
##	derfinder	1.12.6	2018-05-23	Bioconductor
##	derfinderHelper	1.12.0	2018-05-23	Bioconductor
##	DESeq2	* 1.18.1	2018-05-22	Bioconductor
##	devtools	* 1.13.5	2018-02-18	CRAN (R 3.4.4)
##	digest	0.6.15	2018-01-28	CRAN (R 3.4.4)
##	doRNG	1.6.6	2017-04-10	CRAN (R 3.4.4)
##	DT	* 0.4	2018-01-30	CRAN (R 3.4.4)
##	edgeR	3.20.9	2018-05-23	Bioconductor
##	evaluate	0.10.1	2017-06-24	CRAN (R 3.4.4)
##	foreach	1.4.4	2017-12-12	CRAN (R 3.4.4)
##	foreign	0.8-70	2018-04-23	CRAN (R 3.4.4)
##	Formula	1.2-3	2018-05-03	CRAN (R 3.4.4)
##	genefilter	1.60.0	2018-05-22	Bioconductor
##	geneplotter	1.56.0	2018-05-22	Bioconductor
##	GenomeInfoDb	* 1.14.0	2018-05-22	Bioconductor
##	GenomeInfoDbData	1.0.0	2018-05-22	Bioconductor
##	GenomicAlignments	1.14.2	2018-05-23	Bioconductor
##	GenomicFeatures	1.30.3	2018-05-23	Bioconductor
##	GenomicFiles	1.14.0	2018-05-23	Bioconductor
##	GenomicRanges	* 1.30.3	2018-05-22	Bioconductor
##	ggplot2	* 2.2.1	2016-12-30	CRAN (R 3.4.4)
##	graphics	* 3.4.4	2018-03-16	local
##	grDevices	* 3.4.4	2018-03-16	local
##	grid	3.4.4	2018-03-16	local
##	gridExtra	2.3	2017-09-09	CRAN (R 3.4.4)
##	gtable	0.2.0	2016-02-26	CRAN (R 3.4.4)
##	highr	0.6	2016-05-09	CRAN (R 3.4.4)
##	Hmisc	4.1-1	2018-01-03	CRAN (R 3.4.4)
##	htmlTable	1.11.2	2018-01-20	CRAN (R 3.4.4)
##	htmltools	0.3.6	2017-04-28	CRAN (R 3.4.4)
##	htmlwidgets	1.2	2018-04-19	CRAN (R 3.4.4)
##	httpuv	1.4.3	2018-05-10	CRAN (R 3.4.4)
##	httr	1.3.1	2017-08-20	CRAN (R 3.4.4)
##	IRanges	* 2.12.0	2018-05-22	Bioconductor
##	iterators	1.0.9	2017-12-12	CRAN (R 3.4.4)
##	jsonlite	1.5	2017-06-01	CRAN (R 3.4.4)
##	knitcitations	1.0.8	2017-07-04	CRAN (R 3.4.4)
##	knitr	* 1.20	2018-02-20	CRAN (R 3.4.4)

## knitrBootstrap	1.0.1	2017-07-19	CRAN (R 3.4.4)
## labeling	0.3	2014-08-23	CRAN (R 3.4.4)
## later	0.7.2	2018-05-01	CRAN (R 3.4.4)
## lattice	0.20-35	2017-03-25	CRAN (R 3.4.4)
## latticeExtra	0.6-28	2016-02-09	CRAN (R 3.4.4)
## lazyeval	0.2.1	2017-10-29	CRAN (R 3.4.4)
## limma	3.34.9	2018-05-23	Bioconductor
## locfit	1.5-9.1	2013-04-20	CRAN (R 3.4.4)
## lubridate	1.7.4	2018-04-11	CRAN (R 3.4.4)
## magrittr	1.5	2014-11-22	CRAN (R 3.4.4)
## markdown	0.8	2017-04-20	CRAN (R 3.4.4)
## MASS	* 7.3-50	2018-04-30	CRAN (R 3.4.4)
## Matrix	1.2-14	2018-04-09	CRAN (R 3.4.4)
## matrixStats	* 0.53.1	2018-02-11	CRAN (R 3.4.4)
## memoise	1.1.0	2017-04-21	CRAN (R 3.4.4)
## methods	* 3.4.4	2018-03-16	local
## mime	0.5	2016-07-07	CRAN (R 3.4.4)
## munsell	0.4.3	2016-02-13	CRAN (R 3.4.4)
## nnet	7.3-12	2016-02-02	CRAN (R 3.4.4)
## parallel	* 3.4.4	2018-03-16	local
## pheatmap	* 1.0.10	2018-05-19	CRAN (R 3.4.4)
## pillar	1.2.2	2018-04-26	CRAN (R 3.4.4)
## pkgmaker	0.22	2014-05-14	CRAN (R 3.4.4)
## plyr	1.8.4	2016-06-08	CRAN (R 3.4.4)
## prettyunits	1.0.2	2015-07-13	CRAN (R 3.4.4)
## progress	1.1.2	2016-12-14	CRAN (R 3.4.4)
## promises	1.0.1	2018-04-13	CRAN (R 3.4.4)
## qvalue	2.10.0	2018-05-23	Bioconductor
## R6	2.2.2	2017-06-17	CRAN (R 3.4.4)
## RColorBrewer	* 1.1-2	2014-12-07	CRAN (R 3.4.4)
## Rcpp	0.12.17	2018-05-18	CRAN (R 3.4.4)
## RCurl	1.95-4.10	2018-01-04	CRAN (R 3.4.4)
## RefManageR	1.2.0	2018-04-25	CRAN (R 3.4.4)
## regionReport	* 1.12.2	2018-05-23	Bioconductor
## registry	0.5	2017-12-03	CRAN (R 3.4.4)
## reshape2	1.4.3	2017-12-11	CRAN (R 3.4.4)
## rlang	0.2.0	2018-02-20	CRAN (R 3.4.4)
## rmarkdown	1.9	2018-03-01	CRAN (R 3.4.4)
## RMySQL	0.10.15	2018-05-13	CRAN (R 3.4.4)
## rngtools	1.3.1	2018-05-15	CRAN (R 3.4.4)
## rpart	4.1-13	2018-02-23	CRAN (R 3.4.4)
## rprojroot	1.3-2	2018-01-03	CRAN (R 3.4.4)
## Rsamtools	1.30.0	2018-05-23	Bioconductor
## RSQLite	2.1.1	2018-05-06	CRAN (R 3.4.4)
## rstudioapi	0.7	2017-09-07	CRAN (R 3.4.4)
## rtracklayer	1.38.3	2018-05-23	Bioconductor
## S4Vectors	* 0.16.0	2018-05-23	Bioconductor
## scales	0.5.0	2017-08-24	CRAN (R 3.4.4)
## shiny	1.1.0	2018-05-17	CRAN (R 3.4.4)
## splines	3.4.4	2018-03-16	local
## stats	* 3.4.4	2018-03-16	local
## stats4	* 3.4.4	2018-03-16	local
## stringi	1.2.2	2018-05-02	CRAN (R 3.4.4)
## stringr	1.3.1	2018-05-10	CRAN (R 3.4.4)

```
## SummarizedExperiment * 1.8.1      2018-05-22 Bioconductor
## survival              2.42-3      2018-04-16 CRAN (R 3.4.4)
## tibble                1.4.2       2018-01-22 CRAN (R 3.4.4)
## tools                 3.4.4       2018-03-16 local
## utils                  * 3.4.4     2018-03-16 local
## VariantAnnotation      1.24.5     2018-05-23 Bioconductor
## withr                 2.1.2       2018-03-15 CRAN (R 3.4.4)
## xfun                   0.1         2018-01-22 CRAN (R 3.4.4)
## XML                   3.98-1.11   2018-04-16 CRAN (R 3.4.4)
## xml2                   1.2.0      2018-01-24 CRAN (R 3.4.4)
## xtable                 1.8-2      2016-02-05 CRAN (R 3.4.4)
## XVector                0.18.0     2018-05-22 Bioconductor
## yaml                   2.1.19     2018-05-01 CRAN (R 3.4.4)
## zlibbioc               1.24.0     2018-05-22 Bioconductor
```

Pandoc version used: 1.19.2.1.

## Bibliography

This report was created with `regionReport` (Collado-Torres, Jaffe, and Leek, 2016) using `rmarkdown` (Allaire, Xie, McPherson, Luraschi, et al., 2018) while `knitr` (Xie, 2014) and `DT` (Xie, 2018) were running behind the scenes. `pheatmap` (Kolde, 2018) was used to create the sample distances heatmap. Several plots were made with `ggplot2` (Wickham, 2009).

Citations made with `knitcitations` (Boettiger, 2017). The BibTeX file can be found [here](#).

[1] J. Allaire, Y. Xie, J. McPherson, J. Luraschi, et al. `rmarkdown`: Dynamic Documents for R. R package version 1.9. 2018. URL: <https://CRAN.R-project.org/package=rmarkdown>.

[1] C. Boettiger. `knitcitations`: Citations for ‘Knitr’ Markdown Files. R package version 1.0.8. 2017. URL: <https://CRAN.R-project.org/package=knitcitations>.

## No encoding supplied: defaulting to UTF-8.

[1] R. Kolde. `pheatmap`: Pretty Heatmaps. R package version 1.0.10. 2018. URL: <https://CRAN.R-project.org/package=pheatmap>.

## No encoding supplied: defaulting to UTF-8.

[1] H. Wickham. `ggplot2`: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009. ISBN: 978-0-387-98140-6. URL: <http://ggplot2.org>.

[1] Y. Xie. `DT`: A Wrapper of the JavaScript Library ‘DataTables’. R package version 0.4. 2018. URL: <https://CRAN.R-project.org/package=DT>.

[1] 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>.