

RNA-seq COVID

Code

library(DESeq2)

Obtención de la base de datos de GEO GSE147507

Hide

load("/Users/sofia/Documents/MCI/4 SEMESTRE /BIOINFORMATICA/Actividad 7- Visualization/GSE147507_datos_covid.Rdata")

Hide

datos_covid[1:10,]

	SARS004_mock_3	SARS004_mock_2	SARS004_mock_1	SARS004_CoV2_3	SARS004_CoV2_2
	<int>	<int>	<int>	<int>	<int>
DDX11L1	0	0	0	0	0
WASH7P	29	24	23	34	19
FAM138A	0	0	0	0	0
FAM138F	0	0	0	0	0
OR4F5	0	0	0	0	0
LOC729737	112	119	113	127	84
LOC100132287	0	0	0	0	0
LOC100132062	0	0	0	0	0
LOC100133331	19	21	36	16	23
OR4F29	0	0	0	0	0

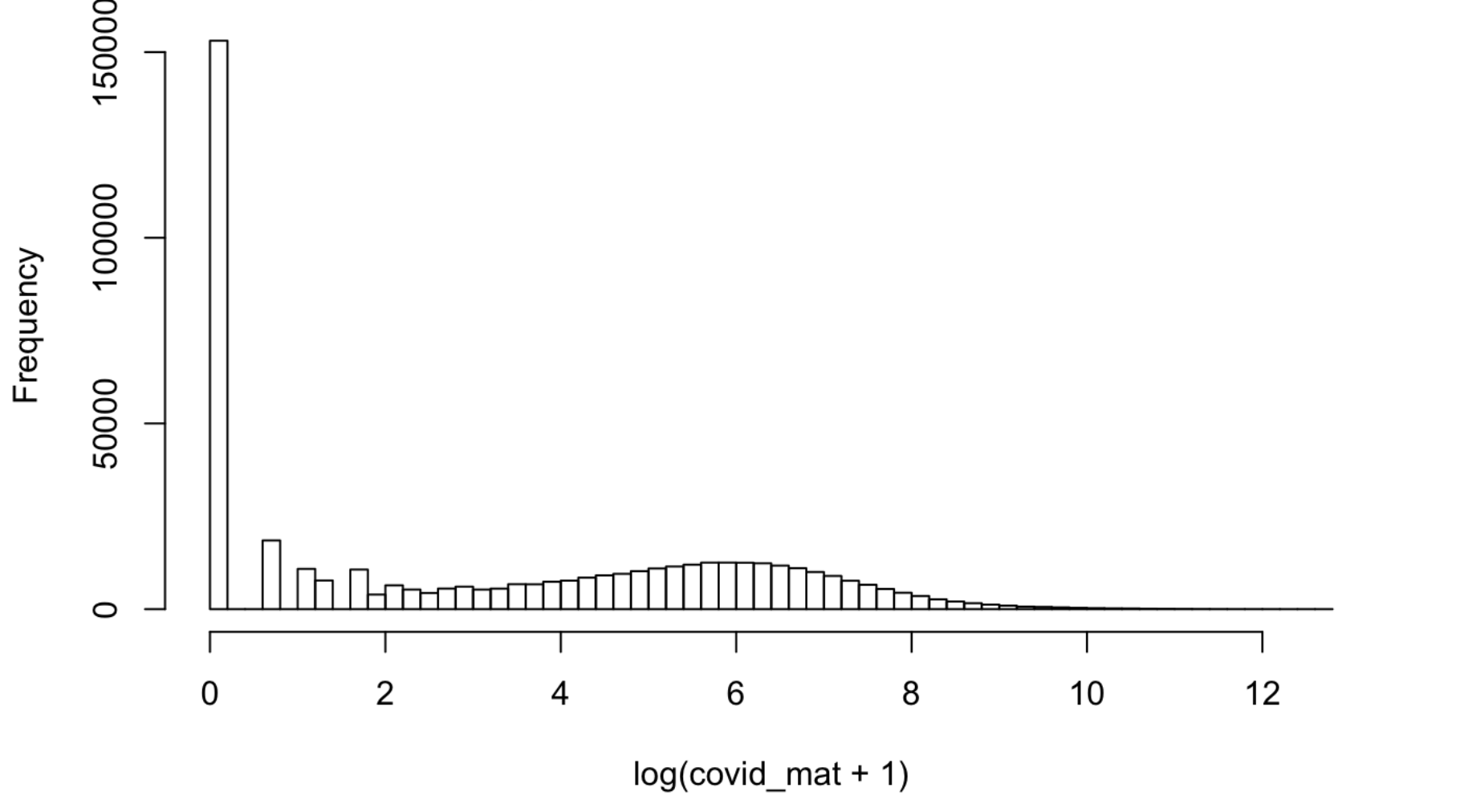
1-10 of 10 rows | 1-6 of 20 columns

Hide

covid_mat<-data.matrix(datos_covid)

Hide

hist(log(covid_mat+1), breaks=60, main="Histogram of covid expression count")



Filter out low expression

Hide

count_mat_filter <- apply(covid_mat, 1, function(x) length(which(x>=16)))
table(count_mat_filter)

Hide

count_mat_filter
0 1 2 3 4 5 6 7 8 9 10 11
8319 498 376 273 221 208 665 231 194 198 243 231
12 13 14 15 16 17 18 19 20
265 277 382 183 268 382 468 866 9842

Hide

dim(covid_mat)

[1] 23719 28

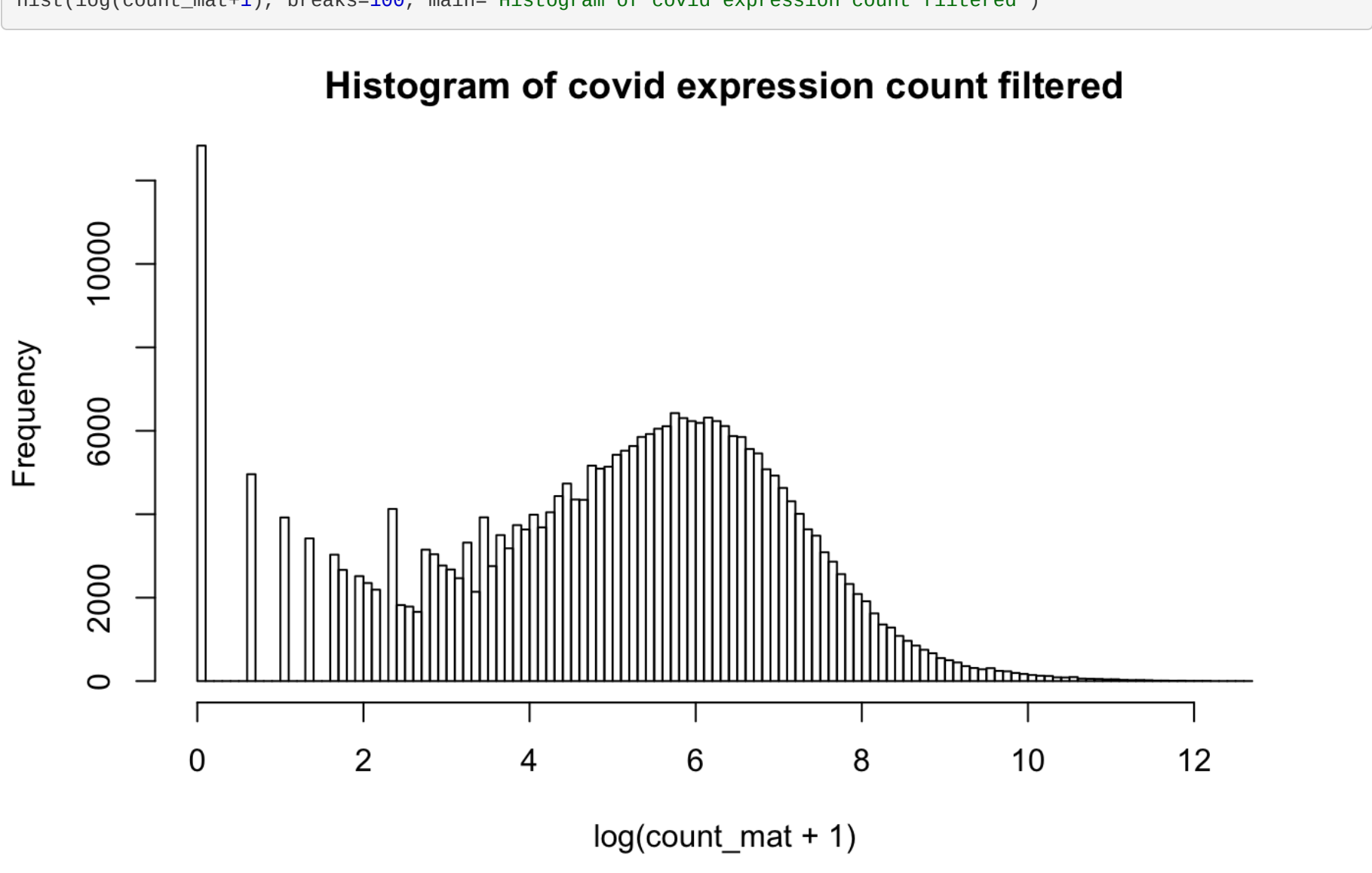
Hide

count_mat <- covid_mat[which(count_mat_filter>=2),]
#count_mat <- covid_mat[apply(covid_mat > exp(3) - 1, 1, any),]
dim(count_mat)

[1] 14893 28

Hide

hist(log(count_mat+1), breaks=100, main="Histogram of covid expression count filtered")



Obtención de genes diferencialmente expresados usando DESeq2

Hide

DESeq_func <- function(matrix_c, classes_c, classes_names=NA){
 classes_1 <- which(classes_c == 1)
 classes_0 <- which(classes_c == 0)
 aux_data <- matrix_c[,c(classes_1,classes_0)]
 aux_desc <- data.frame(condition=c(rep("B",length(classes_1)),rep("A",length(classes_0))), type=rep("paired-end",c(length(classes_1)+length(classes_0))))
 aux_dds <- DESeqDataSetFromMatrix(countData = aux_data, colData = aux_desc, design = ~condition)
 aux_dds <- DESeq(aux_dds)
 aux_results <- as.data.frame(results(aux_dds))
 aux_results <- aux_results[order(aux_results\$pvalue),]
 aux_results
}

Comparar entre Covid e Influenza

Omitiendo las muestras de SRV. Considerando tanto para NHBE como para A549.

Hide

count_mat1<-count_mat[,1:12]
count_mat1<-cbind(count_mat1, count_mat[,17:20])
count_mat1[1:5,]

Hide

	SARS004_mock_3	SARS004_mock_2	SARS004_mock_1	SARS004_CoV2_3	SARS004_CoV2_2
WASH7P	29	24	23	34	19
LOC729737	112	119	113	127	84
LOC100133331	19	21	36	16	23
LOC100288069	13	17	27	19	9
LINC00115	9	12	26	12	7

	SARS004_CoV2_1	Cov002_mock3.indexG3	Cov002_mock2.indexG2	Cov002_mock1.indexG1
WASH7P	44	68	43	33
LOC729737	278	11	3	6
LOC100133331	54	54	23	20
LOC100288069	50	23	18	5
LINC00115	28	29	15	2

	Cov002_CoV2.3.indexG6	Cov002_CoV2.2.indexG5	Cov002_CoV2.1.indexG4	X3_9_mock1_13
WASH7P	65	79	48	15
LOC729737	8	19	18	1
LOC100133331	45	42	36	16
LOC100288069	30	17	14	5
LINC00115	17	13	16	5

	X3_9_mock2_14	X3_9_wt1_15	X3_9_wt2_16
WASH7P	12	3	3
LOC729737	5	0	2
LOC100133331	14	10	3
LOC100288069	15	8	3
LINC00115	11	4	0

Hide

aux_classes <- rep(1, times=ncol(count_mat1))
aux_classes[grepl(pattern="X3", x=colnames(count_mat1))] <- 0

aux_classes

[1] 1 1 1 1 1 1 1 1 1 1 1 0 0 0

Hide

colnames(count_mat1)

[1] "SARS004_mock_3" "SARS004_mock_2" "SARS004_mock_1" "SARS004_CoV2_3"
[5] "SARS004_CoV2_2" "SARS004_CoV2_1" "Cov002_mock3.indexG3" "Cov002_mock2.indexG2"
[9] "Cov002_mock1.indexG1" "Cov002_CoV2.3.indexG6" "Cov002_CoV2.2.indexG5" "Cov002_CoV2.1.indexG4"
[13] "X3_9_mock1_13" "X3_9_mock2_14" "X3_9_wt1_15" "X3_9_wt2_16"

Hide

grep(pattern = "X3", x=colnames(count_mat1))

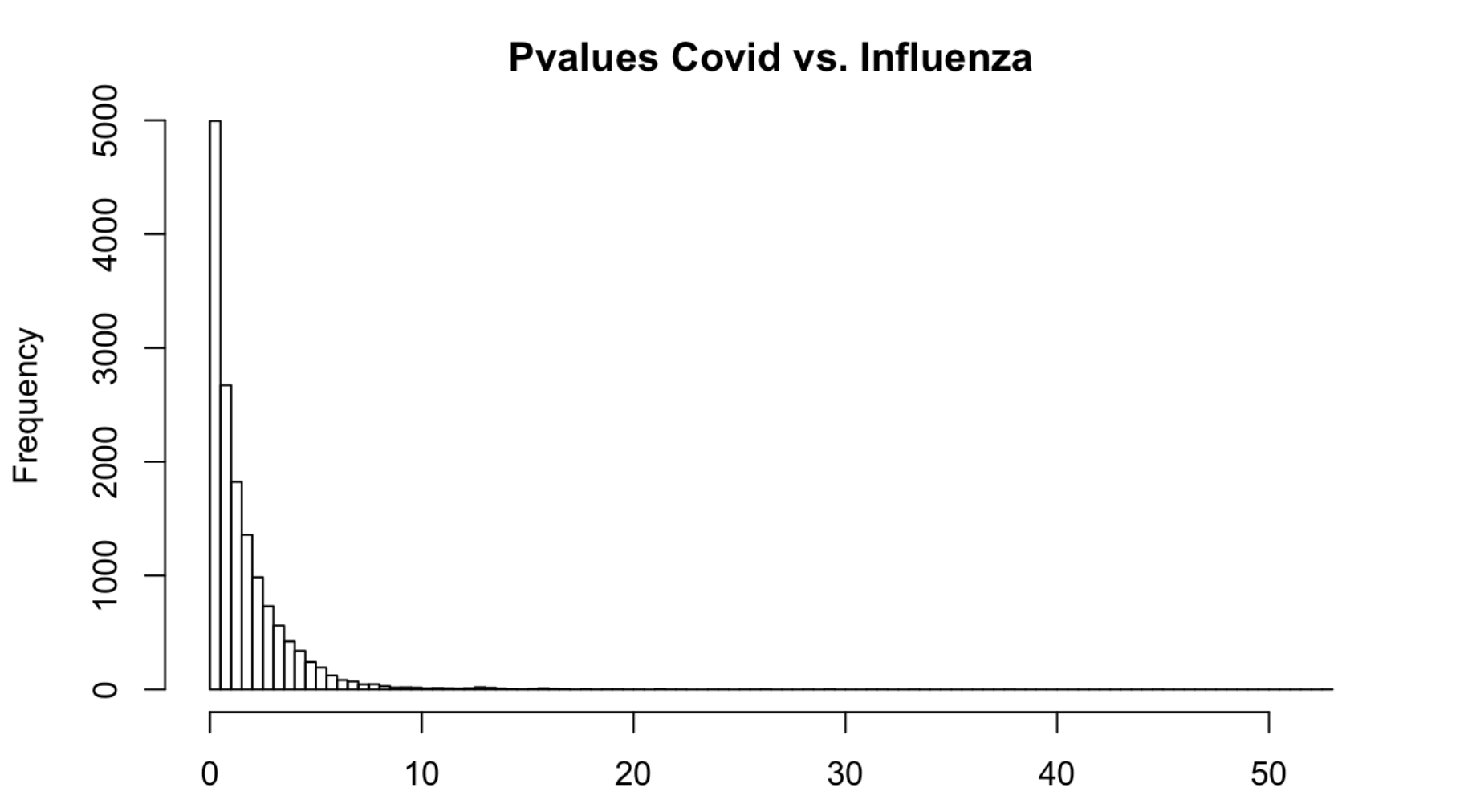
[1] 13 14 15 16

Hide

count_results <- DESeq_func(matrix_c = count_mat1, classes_c = aux_classes)

Hide

hist(log10(count_results\$pvalue)*-1, breaks=100, main="Pvalues Covid vs. Influenza")



Ahora lo repetimos pero sin considerar los "mock treated"

Hide

count_mat2<-count_mat[,4:6]
count_mat2<-cbind(count_mat2, count_mat[,19:12])
count_mat2<-cbind(count_mat2, count_mat[,19:20])
count_mat2[1:5,]

Hide

	SARS004_CoV2_3	SARS004_CoV2_2	SARS004_CoV2_1	Cov002_CoV2.3.indexG6
WASH7P	34	19	44	65
LOC729737	127	84	279	8
LOC100133331	16	23	54	45
LOC100288069	19	9	50	30
LINC00115	12	12	28	17

	Cov002_CoV2.2.indexG5	Cov002_CoV2.1.indexG4	X3_9_wt1_15	X3_9_wt2_16
WASH7P	79	48	3	3
LOC729737	10	10	0	2
LOC100133331	42	36	10	3
LOC100288069	17	14	8	3
LINC00115	13	16	4	0

Hide

aux_classes2 <- rep(1, times=ncol(count_mat2))
aux_classes2[grepl(pattern="X3", x=colnames(count_mat2))] <- 0

aux_classes2

[1] 1 1 1 1 1 1 0 0

Hide

colnames(count_mat2)

[1] "SARS004_CoV2_3" "SARS004_CoV2_2" "SARS004_CoV2_1" "Cov002_CoV2.3.indexG6"
[5] "Cov002_CoV2.2.indexG5" "Cov002_CoV2.1.indexG4" "X3_9_wt1_15" "X3_9_wt2_16"

Hide

grep(pattern = "X3", x=colnames(count_mat2))

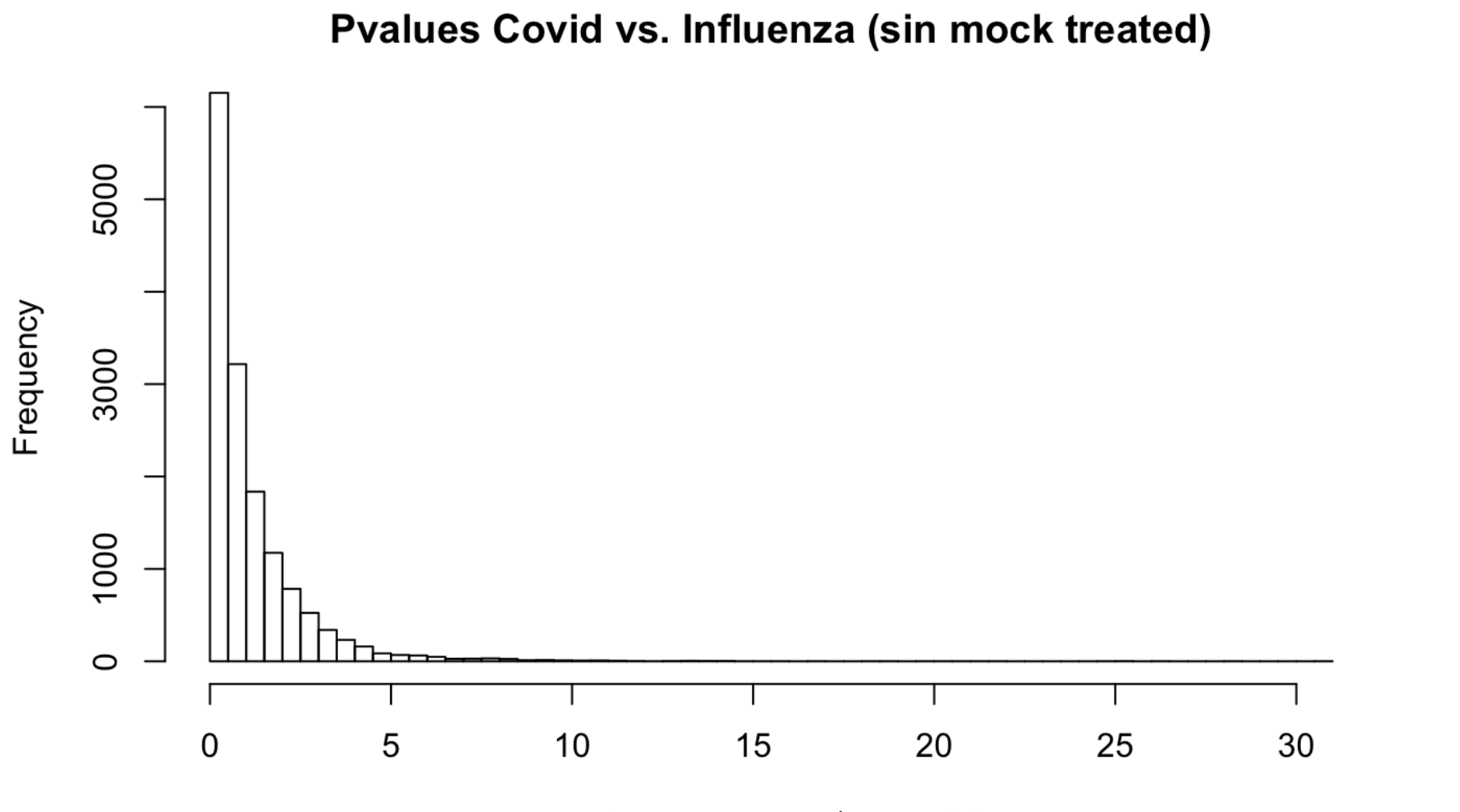
[1] 7 8

Hide

count_results2 <- DESeq_func(matrix_c = count_mat2, classes_c = aux_classes2)

Hide

hist(log10(count_results2\$pvalue)*-1, breaks=100, main="Pvalues Covid vs. Influenza (sin mock treated)")



Droga para los genes significativos

Hide

genes<-rownames(count_mat)[which(count_results2\$pvalue<=1e-6&&abs(count_results2\$log2FoldChange)>=0.5)]

Drug query:

Hide

BiocManager::install("rDGIdb")

Hide

library(rDGIdb)
drugs<-queryDGIdb(genes)

Querying DGIDB...done!

Hide

resultsummary(drugs)[10:20,c(1,2,9)]

Gene	Drug	GuideToPharmacologyInteractions
<chr>	<chr>	<db>
10 HTR1D	ERGOTAMINE	1
11 HDAC1	PANOBINOSTAT	1
12 HDAC1	ROMIDEPSIN	1
13 MTOR	DACTOLISIB	1
14 MTOR	RIDAFOROLIMUS	1
15 HDAC1	BELINOSTAT	1
16 HDAC1	MOCETINOSTAT	1
17 SCN1D	TRIAMTERENE	0
18 GABRD	ESTAZOLAM	0
19 GABRD	TEMAZEPAM	0