

Expresión diferencial de genes con DESeq2

Expresión diferencial

Piero Palacios Bernuy

Contenido

1 Importación de los conteos de genes	1
---------------------------------------	---

1 Importación de los conteos de genes

Importamos las 92 muestras con su respectiva metadata. Esto lo haremos con el paquete `tximport` debido a que tiene una implementación para importar directamente datos provenientes de *RSEM*.

```
sample_table <-  
  read_excel("D:/tesis cafe/SRaRunTablecoffeaarabica.xlsx")  
rownames(sample_table) <- sample_table$Run  
  
dir <- "D:/tesis cafe/DESeq2 coffea/Gene_counts"  
list.files(dir)  
  
[1] "SRR11196520_.genes.results" "SRR11196521_.genes.results"  
[3] "SRR11196522_.genes.results" "SRR11196523_.genes.results"  
[5] "SRR11196524_.genes.results" "SRR11196525_.genes.results"  
[7] "SRR11196526_.genes.results" "SRR11196527_.genes.results"  
[9] "SRR11196528_.genes.results" "SRR11196529_.genes.results"  
[11] "SRR11196530_.genes.results" "SRR11196531_.genes.results"  
[13] "SRR11196532_.genes.results" "SRR11196533_.genes.results"  
[15] "SRR11196534_.genes.results" "SRR11196535_.genes.results"  
[17] "SRR11196536_.genes.results" "SRR11196537_.genes.results"  
[19] "SRR11196538_.genes.results" "SRR11196539_.genes.results"  
[21] "SRR11711678_.genes.results" "SRR11711679_.genes.results"  
[23] "SRR11711680_.genes.results" "SRR11711681_.genes.results"  
[25] "SRR11711682_.genes.results" "SRR11711683_.genes.results"
```

```

[27] "SRR11711684_.genes.results" "SRR11711685_.genes.results"
[29] "SRR11711686_.genes.results" "SRR11711687_.genes.results"
[31] "SRR11711688_.genes.results" "SRR11711689_.genes.results"
[33] "SRR11711690_.genes.results" "SRR11711691_.genes.results"
[35] "SRR11711692_.genes.results" "SRR11711693_.genes.results"
[37] "SRR11711695_.genes.results" "SRR11711706_.genes.results"
[39] "SRR11711717_.genes.results" "SRR11711728_.genes.results"
[41] "SRR11711739_.genes.results" "SRR11711760_.genes.results"
[43] "SRR11711771_.genes.results" "SRR11711782_.genes.results"
[45] "SRR11711793_.genes.results" "SRR11711804_.genes.results"
[47] "SRR11711805_.genes.results" "SRR11711806_.genes.results"
[49] "SRR11711807_.genes.results" "SRR11711808_.genes.results"
[51] "SRR11711809_.genes.results" "SRR11711810_.genes.results"
[53] "SRR11711811_.genes.results" "SRR11711812_.genes.results"
[55] "SRR11711813_.genes.results" "SRR11711814_.genes.results"
[57] "SRR11711815_.genes.results" "SRR11711816_.genes.results"
[59] "SRR11711817_.genes.results" "SRR11711818_.genes.results"
[61] "SRR11711819_.genes.results" "SRR11711820_.genes.results"
[63] "SRR11711821_.genes.results" "SRR11711822_.genes.results"
[65] "SRR11711823_.genes.results" "SRR11711824_.genes.results"
[67] "SRR11711825_.genes.results" "SRR11711826_.genes.results"
[69] "SRR11711827_.genes.results" "SRR11711828_.genes.results"
[71] "SRR11711833_.genes.results" "SRR11711844_.genes.results"
[73] "SRR11711855_.genes.results" "SRR11711866_.genes.results"
[75] "SRR11711877_.genes.results" "SRR11711888_.genes.results"
[77] "SRR11711899_.genes.results" "SRR11711907_.genes.results"
[79] "SRR11711908_.genes.results" "SRR11711909_.genes.results"
[81] "SRR11711910_.genes.results" "SRR11711911_.genes.results"
[83] "SRR11711912_.genes.results" "SRR11711913_.genes.results"
[85] "SRR11711914_.genes.results" "SRR11711915_.genes.results"
[87] "SRR11711916_.genes.results" "SRR11711927_.genes.results"
[89] "SRR11711938_.genes.results" "SRR11711949_.genes.results"
[91] "SRR11711950_.genes.results" "SRR11711951_.genes.results"

files <-
  list.files(file.path(dir), pattern = ".genes.results", full.names = TRUE)
files <-
  files[sapply(rownames(sample_table), function(x)
    grep(x, files, value = FALSE, fixed = TRUE))]

names(files) <- rownames(sample_table)

txi.rsem <- tximport(files,
                      type = "rsem",

```

```

      txIn = F,
      txOut = F)
head(txi.rsem$counts, 6, 6)

```

	SRR11711678	SRR11711679	SRR11711680	SRR11711681	SRR11711682
CoarCr001	2506.00	4131.00	4188.50	4212.50	4173.0
CoarCr002	9996.58	21086.07	17713.19	18550.78	18443.5
CoarCr003	18.00	21.00	22.00	24.00	24.5
CoarCr004	4.00	7.00	8.50	6.00	7.5
CoarCr005	4.00	7.00	8.50	6.00	7.5
CoarCr006	18.00	21.00	22.00	24.00	24.5
	SRR11711683	SRR11711684	SRR11711685	SRR11711686	SRR11711687
CoarCr001	4216.50	106460.0	4165.50	25.5	3613.50
CoarCr002	21566.75	221322.4	17425.13	113.0	16020.53
CoarCr003	17.00	1192.0	24.50	0.0	21.00
CoarCr004	4.00	317.0	7.00	0.0	4.50
CoarCr005	4.00	317.0	7.00	0.0	4.50
CoarCr006	17.00	1192.0	24.50	0.0	21.00
	SRR11711688	SRR11711689	SRR11711690	SRR11711691	SRR11711692
CoarCr001	3801.00	3718.00	3702.50	3701.00	3806.00
CoarCr002	15074.59	15338.64	12253.89	17175.86	15488.81
CoarCr003	14.00	19.50	18.50	24.00	16.00
CoarCr004	5.50	6.00	6.00	5.00	8.00
CoarCr005	5.50	6.00	6.00	5.00	8.00
CoarCr006	14.00	19.50	18.50	24.00	16.00
	SRR11711693	SRR11711695	SRR11711706	SRR11711717	SRR11711728
CoarCr001	3677.5	105988.0	106109.0	106615.5	4388.00
CoarCr002	12533.9	201294.7	219917.0	201384.8	12700.52
CoarCr003	21.5	1155.0	1149.0	1063.0	27.00
CoarCr004	6.5	273.0	292.5	271.5	7.00
CoarCr005	6.5	273.0	292.5	271.5	7.00
CoarCr006	21.5	1155.0	1149.0	1063.0	27.00
	SRR11711739	SRR11711760	SRR11711771	SRR11711782	SRR11711793
CoarCr001	29346.00	29361.00	29295.00	614.50	3636.00
CoarCr002	83955.17	86965.28	87741.62	1391.57	12066.16
CoarCr003	160.00	140.50	153.50	5.50	26.00
CoarCr004	63.00	58.50	61.50	2.00	10.50
CoarCr005	63.00	58.50	61.50	2.00	10.50
CoarCr006	160.00	140.50	153.50	5.50	26.00
	SRR11711804	SRR11711805	SRR11711806	SRR11711807	SRR11711808
CoarCr001	3514.50	4009.50	5287.50	5306.0	5257.50
CoarCr002	16224.23	14613.34	20870.63	15915.4	16185.21
CoarCr003	27.50	23.00	32.00	29.5	25.00
CoarCr004	10.50	2.50	5.50	7.5	10.00

CoarCr005	10.50	2.50	5.50	7.5	10.00
CoarCr006	27.50	23.00	32.00	29.5	25.00
	SRR11711809	SRR11711810	SRR11711811	SRR11711812	SRR11711813
CoarCr001	5209.50	5203.00	5314.00	5312.00	894.50
CoarCr002	18181.64	12381.07	21709.61	21792.54	791.01
CoarCr003	25.50	26.00	23.00	25.00	6.00
CoarCr004	8.00	4.50	6.00	9.50	1.00
CoarCr005	8.00	4.50	6.00	9.50	1.00
CoarCr006	25.50	26.00	23.00	25.00	6.00
	SRR11711814	SRR11711815	SRR11711816	SRR11711817	SRR11711818
CoarCr001	4079.50	3562.50	3978.00	4097.00	4049.00
CoarCr002	17754.36	16064.47	15481.25	15905.18	11849.59
CoarCr003	17.00	32.50	24.00	21.00	20.50
CoarCr004	6.00	5.50	4.50	7.00	5.50
CoarCr005	6.00	5.50	4.50	7.00	5.50
CoarCr006	17.00	32.50	24.00	21.00	20.50
	SRR11711819	SRR11711820	SRR11711821	SRR11711822	SRR11711823
CoarCr001	4001.00	4067.00	3990.50	4022.50	3989.00
CoarCr002	14373.16	11873.48	15331.26	17725.95	16280.08
CoarCr003	18.50	20.50	19.50	19.50	18.50
CoarCr004	5.00	7.50	4.50	0.00	4.50
CoarCr005	5.00	7.50	4.50	0.00	4.50
CoarCr006	18.50	20.50	19.50	19.50	18.50
	SRR11711824	SRR11711825	SRR11711826	SRR11711827	SRR11711828
CoarCr001	9124.50	10959.00	3612.00	11172.50	10938.0
CoarCr002	32383.59	38480.42	16779.05	36712.26	39774.3
CoarCr003	43.50	52.50	25.00	50.00	47.0
CoarCr004	11.50	10.50	8.00	14.50	12.0
CoarCr005	11.50	10.50	8.00	14.50	12.0
CoarCr006	43.50	52.50	25.00	50.00	47.0
	SRR11711833	SRR11711844	SRR11711855	SRR11711866	SRR11711877
CoarCr001	10994.50	11250.50	10824.00	10907.00	11155.50
CoarCr002	39785.32	40237.02	36332.75	38709.76	40890.28
CoarCr003	62.50	52.00	47.50	53.50	60.50
CoarCr004	14.00	18.50	14.00	13.00	15.50
CoarCr005	14.00	18.50	14.00	13.00	15.50
CoarCr006	62.50	52.00	47.50	53.50	60.50
	SRR11711888	SRR11711899	SRR11711907	SRR11711908	SRR11711909
CoarCr001	101384.5	106414.5	2684.50	4504.00	4643.00
CoarCr002	182608.5	207486.1	12252.43	23159.29	21085.25
CoarCr003	1122.5	1190.0	12.50	22.00	24.00
CoarCr004	277.5	303.5	3.00	7.00	5.00
CoarCr005	277.5	303.5	3.00	7.00	5.00
CoarCr006	1122.5	1190.0	12.50	22.00	24.00
	SRR11711910	SRR11711911	SRR11711912	SRR11711913	SRR11711914

CoarCr001	106265.0	3614.00	4529.50	4494.50	4558.00
CoarCr002	244885.7	14868.62	24555.91	22512.67	24694.78
CoarCr003	1118.5	26.00	26.00	22.50	31.50
CoarCr004	290.5	4.50	5.50	3.50	5.00
CoarCr005	290.5	4.50	5.50	3.50	5.00
CoarCr006	1118.5	26.00	26.00	22.50	31.50
	SRR11711915	SRR11711916	SRR11711927	SRR11711938	SRR11711949
CoarCr001	4711.50	29296.00	29075.00	29185.50	29139.00
CoarCr002	23858.33	82228.25	79439.67	94723.04	86972.71
CoarCr003	24.00	146.50	166.50	170.00	163.00
CoarCr004	7.50	55.00	54.00	52.50	51.50
CoarCr005	7.50	55.00	54.00	52.50	51.50
CoarCr006	24.00	146.50	166.50	170.00	163.00
	SRR11711950	SRR11711951	SRR1196520	SRR11196521	SRR11196522
CoarCr001	3537.50	3531.50	6016.50	8803.00	3435.50
CoarCr002	16120.86	18258.71	89.39	1158.75	1061.25
CoarCr003	27.50	27.50	0.00	0.00	0.00
CoarCr004	8.00	5.50	0.00	0.00	0.00
CoarCr005	8.00	5.50	0.00	0.00	0.00
CoarCr006	27.50	27.50	0.00	0.00	0.00
	SRR1196523	SRR11196524	SRR11196525	SRR11196526	SRR11196527
CoarCr001	3265.50	4206.50	2625.50	5376.00	4267.00
CoarCr002	248.69	436.98	1695.41	120.27	1959.33
CoarCr003	0.00	0.00	0.00	0.00	0.00
CoarCr004	0.00	0.00	0.00	0.00	0.00
CoarCr005	0.00	0.00	0.00	0.00	0.00
CoarCr006	0.00	0.00	0.00	0.00	0.00
	SRR11196528	SRR11196529	SRR11196530	SRR11196531	SRR11196532
CoarCr001	3022.00	7942.00	3182.00	67136.79	3326.50
CoarCr002	1507.18	856.27	146.99	5838.47	505.91
CoarCr003	0.00	0.00	0.00	0.00	0.00
CoarCr004	0.00	0.00	0.00	0.00	0.00
CoarCr005	0.00	0.00	0.00	0.00	0.00
CoarCr006	0.00	0.00	0.00	0.00	0.00
	SRR11196533	SRR11196534	SRR11196535	SRR11196536	SRR11196537
CoarCr001	5741.00	7323.00	3249.00	12193.00	7431.00
CoarCr002	1472.92	748.07	123.83	871.85	359.44
CoarCr003	0.00	0.00	0.00	0.00	0.00
CoarCr004	0.00	0.00	0.00	0.00	0.00
CoarCr005	0.00	0.00	0.00	0.00	0.00
CoarCr006	0.00	0.00	0.00	0.00	0.00
	SRR11196538	SRR11196539			
CoarCr001	6534.50	46723.50			
CoarCr002	3601.83	3758.95			
CoarCr003	0.00	0.00			

```

CoarCr004      0.00      0.00
CoarCr005      0.00      0.00
CoarCr006      0.00      0.00

```

```
txi.rsem$length[txi.rsem$length == 0] <- 1
```

Una vez importado, formaremos un objeto del tipo *DESeq* con el siguiente diseño: \sim *Temperatura*. El análisis exploratorio nos dará indicios sobre si añadir los cultivares y los lugares de los laboratorios como parte del modelo.

```

dds <-
  DESeqDataSetFromTximport(txi.rsem, colData = sample_table, design = ~ temp)

dds_coll <-
  collapseReplicates(dds, groupby = dds$Replicate, run = dds$Run)

rownames(colData(dds_coll)) <- dds_coll$Run

```

Podemos ver los *size factors* que junto a la longitud promedio de los transcriptos (*avgTxLength*) sirven para la normalización de los conteos:

```

nm<-assays(dds_coll)[["avgTxLength"]]

sf <- estimateSizeFactorsForMatrix(counts(dds_coll)/nm)

## Estimación de los factores normalizantes
dds_coll<-estimateSizeFactors(dds_coll)

head(normalizationFactors(dds_coll))

```

	SRR11196521	SRR11196522	SRR11196523	SRR11196524	SRR11196525
CoarCr001	0.7086190	0.5121491	0.8505704	0.5349571	0.8344866
CoarCr002	0.7188326	0.5165273	0.8664913	0.5403310	0.8472706
CoarCr003	0.2115752	0.1511570	0.2561005	0.1583571	0.2495985
CoarCr004	0.1935059	0.1382476	0.2342285	0.1448328	0.2282818
CoarCr005	0.1935059	0.1382476	0.2342285	0.1448328	0.2282818
CoarCr006	0.2115752	0.1511570	0.2561005	0.1583571	0.2495985
	SRR11196526	SRR11196527	SRR11196528	SRR11196529	SRR11196530
CoarCr001	1.0464147	0.7913775	0.8593875	0.9429600	0.5735467
CoarCr002	1.0691366	0.8083066	0.8745002	0.9670485	0.5791604
CoarCr003	0.3169022	0.2395161	0.2581858	0.2876854	0.1696938
CoarCr004	0.2898375	0.2190605	0.2361357	0.2631159	0.1552013
CoarCr005	0.2898375	0.2190605	0.2361357	0.2631159	0.1552013

CoarCr006	0.3169022	0.2395161	0.2581858	0.2876854	0.1696938
	SRR11196531	SRR11196532	SRR11196533	SRR11196534	SRR11196535
CoarCr001	0.6330437	0.5941399	0.8457193	0.7841580	0.9032135
CoarCr002	0.6445404	0.6018519	0.8588150	0.7962573	0.9220486
CoarCr003	0.1903985	0.1768965	0.2530399	0.2345955	0.2730795
CoarCr004	0.1741377	0.1617889	0.2314294	0.2145601	0.2497574
CoarCr005	0.1741377	0.1617889	0.2314294	0.2145601	0.2497574
CoarCr006	0.1903985	0.1768965	0.2530399	0.2345955	0.2730795
	SRR11196536	SRR11196537	SRR11196538	SRR11196539	SRR11711824
CoarCr001	0.8999838	0.7220310	0.6965846	0.28770437	2.485832
CoarCr002	0.9135520	0.7343564	0.7084316	0.28976970	2.399430
CoarCr003	0.2690607	0.2167022	0.2090392	0.08468317	36.223801
CoarCr004	0.2460819	0.1981950	0.1911865	0.07745090	44.171488
CoarCr005	0.2460819	0.1981950	0.1911865	0.07745090	44.171488
CoarCr006	0.2690607	0.2167022	0.2090392	0.08468317	36.223801
	SRR11711813	SRR11711805	SRR11196520	SRR11711686	SRR11711678
CoarCr001	2.815966	2.361025	0.7763431	1.959477	1.880729
CoarCr002	2.718090	2.278962	0.7909489	1.891376	1.815366
CoarCr003	41.034552	34.405108	0.2337946	28.548611	27.401302
CoarCr004	50.037742	41.953764	0.2138276	34.813933	33.414834
CoarCr005	50.037742	41.953764	0.2138276	34.813933	33.414834
CoarCr006	41.034552	34.405108	0.2337946	28.548611	27.401302
	SRR11711907	SRR11711782	SRR11711728	SRR11711684	
CoarCr001	1.904944	2.566903	2.249489	1.235534	
CoarCr002	1.838740	2.477684	2.171302	1.192594	
CoarCr003	27.754104	37.405181	32.779786	18.001121	
CoarCr004	33.845063	45.612069	39.971838	21.951674	
CoarCr005	33.845063	45.612069	39.971838	21.951674	
CoarCr006	27.754104	37.405181	32.779786	18.001121	

A partir de aquí existen dos caminos debido a que la matriz del diseño experimental no tiene *full rank*; es decir, sus columnas son combinaciones lineares de otras. Estos dos caminos son:

- Modelar el lugar y los cultivares como efectos batch además de efectos batch desconocidos si sólo es de interés la variable temperatura. $\sim \sum_{batches} + Temperatura$
- Agrupar los cultivares y temperaturas como interacciones y modelar efectos batch desconocidos si nos interesa el comportamiento de los cultivares a diferentes temperaturas. $\sim \sum_{batches} + Temperatura \times Cultivar$

De acuerdo al análisis exploratorio de datos y a los patrones de agrupamiento de decidió optar por el primer enfoque tratando a los cultivares como réplicas biológicas.

```

library(sva)
library(rafalib)

dat <- counts(dds_coll, normalized=TRUE)
idx <- rowMeans(dat) > 1
dat <- dat[idx,]
mod.1 <- model.matrix(~ temp, colData(dds_coll))
mod0.1 <- model.matrix(~ 1, colData(dds_coll))
svseq.1 <- svaseq(dat, mod.1, mod0.1, n.sv = 2)

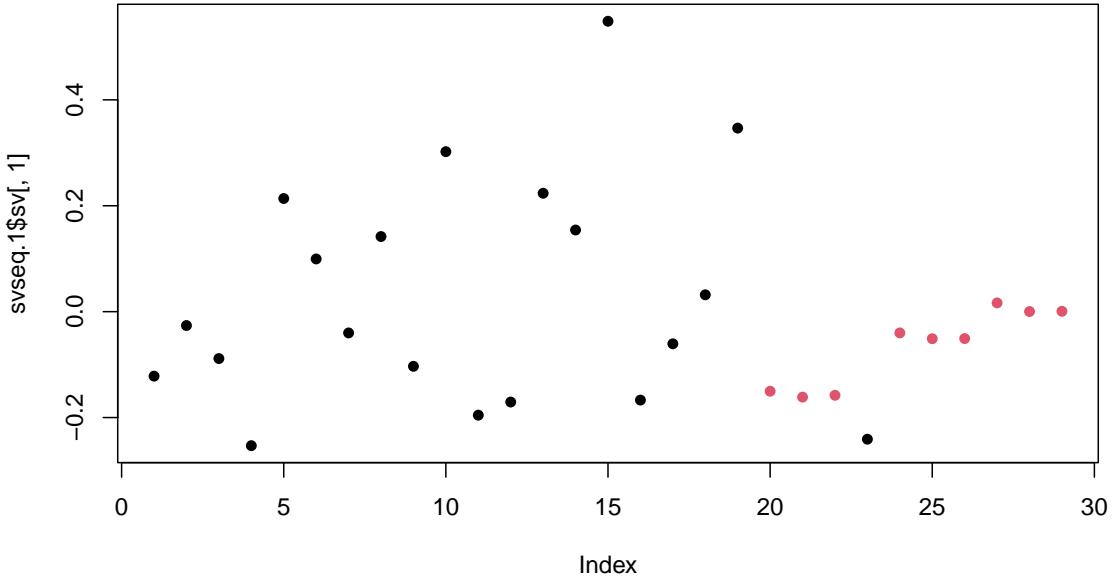
```

Number of significant surrogate variables is: 2
 Iteration (out of 5):1 2 3 4 5

```

plot(svseq.1$sv[, 1],
      col = as.fumeric(dds_coll$geo_loc_name_country),
      pch = 16)

```

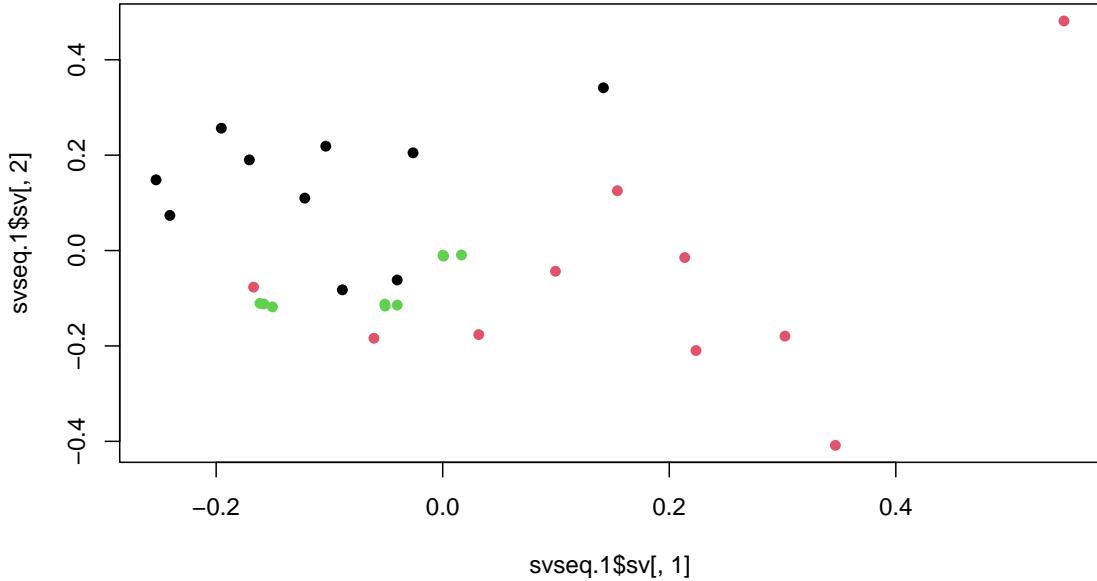


```

plot(svseq.1$sv[, 1],
      svseq.1$sv[, 2],
      col = as.fumeric(dds_coll$Cultivar),

```

```
pch = 16)
```



```
dds.sva.1 <- dds_coll
dds.sva.1$SV1 <- svseq.1$sv[,1]
dds.sva.1$SV2 <- svseq.1$sv[,2]

dds.sva.1$temp<-factor(
  dds.sva.1$temp,
  levels = c("23_19_C",
             "25_20_C",
             "30_26_C",
             "37_28_C",
             "42_30_C"),
  labels = c(
    "23_25_30_C",
    "23_25_30_C",
    "23_25_30_C",
    "37_28_C",
    "42_30_C"
  )
)
```

```

design(dds.sva.1) <- ~ SV1 + SV2 + temp

dds.sva.1 <- DESeq(dds.sva.1)

```

Necesitamos definir los siguientes contrastes:

- 37/28 vs 23/19, 25/20 y 30/26
- 42/30 vs 23/19, 25/20 y 30/26
- 42/30 vs 37/28

```
levels(dds.sva.1$temp)
```

```
[1] "23_25_30_C" "37_28_C"     "42_30_C"
```

```
resultsNames(dds.sva.1)
```

```
[1] "Intercept"           "SV1"
[3] "SV2"                 "temp_37_28_C_vs_23_25_30_C"
[5] "temp_42_30_C_vs_23_25_30_C"
```

```

## Michael Love:
## A design with ~batch + type, uses a per-gene fixed effect
## to account for the differences between batch (similar to
## the terms that are used to model differences in type).

```

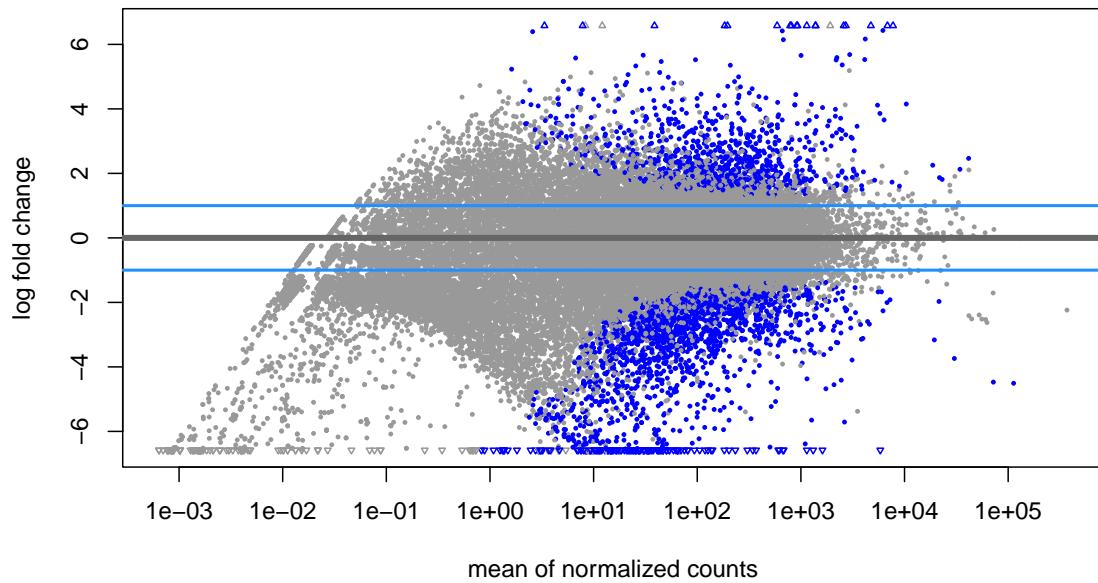
```

res_37_vs_control.1 <- results(
  dds.sva.1,
  alpha = 0.05,
  lfcThreshold = 1,
  contrast = list("temp_37_28_C_vs_23_25_30_C")
)

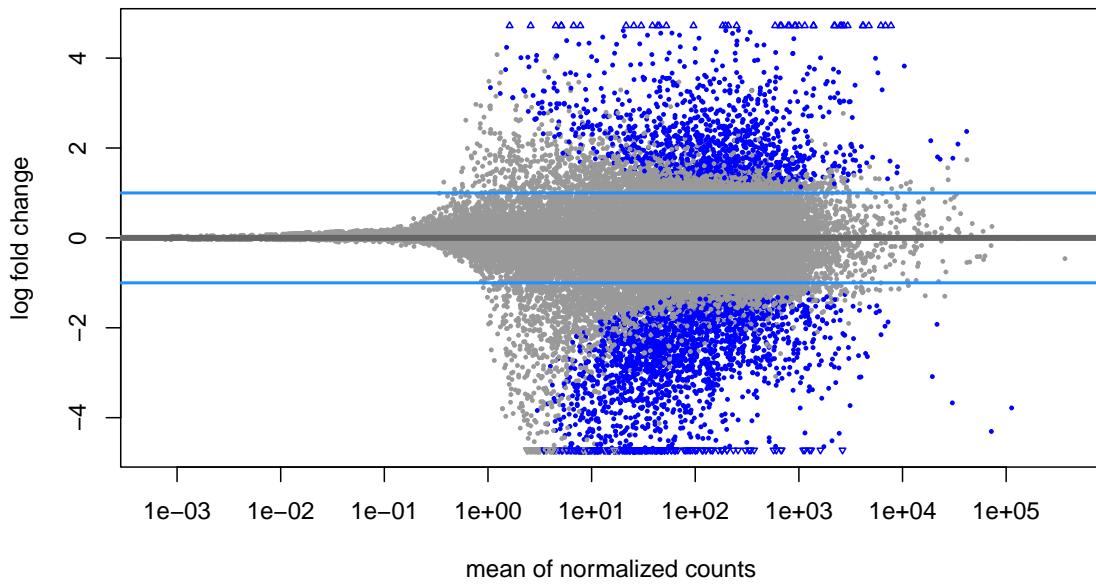
res_37_vs_control.1.lfc <- lfcShrink(dds.sva.1,
                                      type = "apeglm",
                                      lfcThreshold = 1,
                                      coef = 4)

plotMA(res_37_vs_control.1)
abline(h = c(-1, 1), col = "dodgerblue", lwd = 2)

```



```
plotMA(res_37_vs_control.1.lfc)
abline(h = c(-1, 1), col = "dodgerblue", lwd = 2)
```



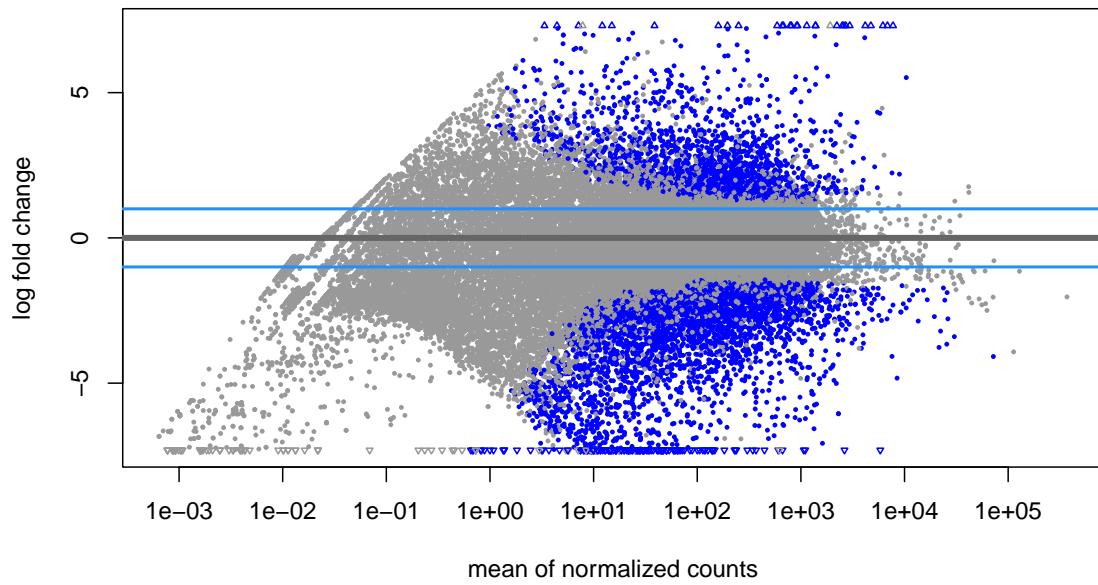
```

res_42_vs_control.1 <- results(
  dds.sva.1,
  alpha = 0.05,
  lfcThreshold = 1,
  contrast = list("temp_42_30_C_vs_23_25_30_C")
)

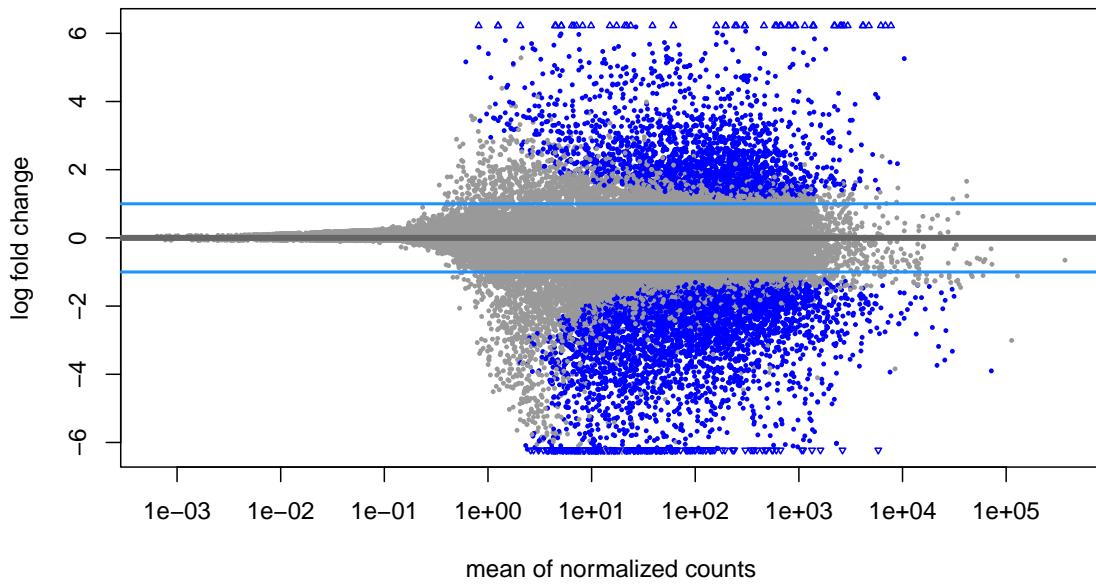
res_42_vs_control.1.lfc <- lfcShrink(dds.sva.1,
                                      type = "apeglm",
                                      lfcThreshold = 1,
                                      coef = 5)

plotMA(res_42_vs_control.1)
abline(h = c(-1, 1), col = "dodgerblue", lwd = 2)

```



```
plotMA(res_42_vs_control.1.lfc)
abline(h = c(-1, 1), col = "dodgerblue", lwd = 2)
```



```
levels(dds.sva.1$temp)
```

```
[1] "23_25_30_C" "37_28_C"     "42_30_C"
```

```
dds.sva.1$temp <- relevel(dds.sva.1$temp, "37_28_C")
```

```
dds.sva.1 <- nbinomWaldTest(dds.sva.1)
```

```
resultsNames(dds.sva.1)
```

```
[1] "Intercept"                 "SV1"
[3] "SV2"                      "temp_23_25_30_C_vs_37_28_C"
[5] "temp_42_30_C_vs_37_28_C"
```

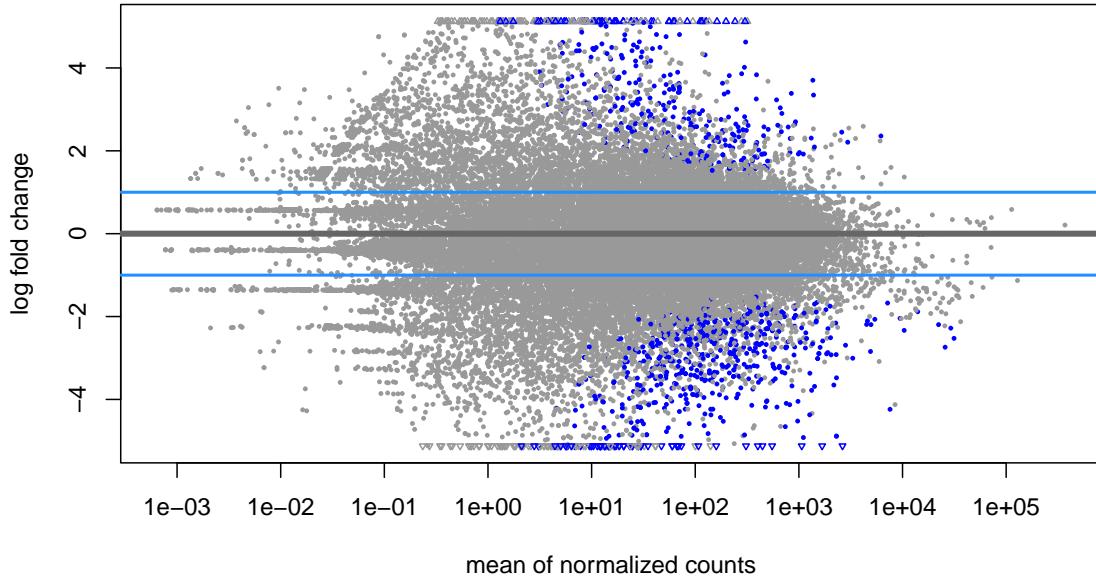
```
res_42_vs_37.1 <- results(
  dds.sva.1,
  alpha = 0.05,
  lfcThreshold = 1,
  contrast = list("temp_42_30_C_vs_37_28_C"))
```

```

)
res_42_vs_37.1.lfc <- lfcShrink(dds.sva.1,
                                    type = "apeglm",
                                    coef = 5,
                                    lfcThreshold = 1)

plotMA(res_42_vs_37.1)
abline(h = c(-1, 1), col = "dodgerblue", lwd = 2)

```



```

plotMA(res_42_vs_control.1.lfc)
abline(h = c(-1, 1), col = "dodgerblue", lwd = 2)

```

