RNAseq

Laura Pauner

6/1/2020

**DEFINICIÓN DE LOS DATOS**

En primer lugar importo los datos situados en la carpeta *‘Datos originales’*:

* targets:

targetsOriginal<-read.table("./datos\_originales/targets.csv", header = TRUE, sep = ",")  
head(targetsOriginal)

## Experiment SRA\_Sample Sample\_Name Grupo\_analisis body\_site  
## 1 SRX567480 SRS626942 GTEX-111CU-0226-SM-5GZXC 1 Thyroid  
## 2 SRX615964 SRS644174 GTEX-111FC-1026-SM-5GZX1 1 Thyroid  
## 3 SRX563960 SRS625636 GTEX-111VG-0526-SM-5N9BW 3 Thyroid  
## 4 SRX564185 SRS625665 GTEX-111YS-0726-SM-5GZY8 1 Thyroid  
## 5 SRX559141 SRS624025 GTEX-1122O-0226-SM-5N9DA 1 Thyroid  
## 6 SRX561718 SRS625313 GTEX-1128S-0126-SM-5H12S 1 Thyroid  
## molecular\_data\_type sex Group ShortName  
## 1 Allele-Specific Expression male NIT 111CU\_NIT  
## 2 RNA Seq (NGS) male NIT 111FC\_NIT  
## 3 RNA Seq (NGS) male ELI 111VG\_ELI  
## 4 Allele-Specific Expression male NIT 111YS\_NIT  
## 5 RNA Seq (NGS) female NIT 1122O\_NIT  
## 6 Allele-Specific Expression female NIT 1128S\_NIT

dim(targetsOriginal)

## [1] 292 9

* counts:

countsOriginal<-read.table("./datos\_originales/counts.csv", header = TRUE, sep = ";", row.names = 1)  
head(countsOriginal)

## GTEX.111CU.0226.SM.5GZXC GTEX.111FC.1026.SM.5GZX1  
## ENSG00000223972.4 7 0  
## ENSG00000227232.4 401 1064  
## ENSG00000243485.2 4 0  
## ENSG00000237613.2 2 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 1  
## GTEX.111VG.0526.SM.5N9BW GTEX.111YS.0726.SM.5GZY8  
## ENSG00000223972.4 1 4  
## ENSG00000227232.4 474 395  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 1 1  
## GTEX.1122O.0226.SM.5N9DA GTEX.1128S.0126.SM.5H12S  
## ENSG00000223972.4 2 2  
## ENSG00000227232.4 732 631  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.113JC.0126.SM.5EGJW GTEX.117XS.0526.SM.5987Q  
## ENSG00000223972.4 0 3  
## ENSG00000227232.4 331 511  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 0 4  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 2 1  
## GTEX.117YW.0126.SM.5EGGN GTEX.117YX.1226.SM.5H11S  
## ENSG00000223972.4 1 3  
## ENSG00000227232.4 483 529  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.1192W.0126.SM.5EGGS GTEX.1192X.1126.SM.5EGGU  
## ENSG00000223972.4 3 2  
## ENSG00000227232.4 573 674  
## ENSG00000243485.2 1 3  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 3 0  
## GTEX.11DXX.0226.SM.5P9HL GTEX.11DXY.0426.SM.5H12R  
## ENSG00000223972.4 4 0  
## ENSG00000227232.4 825 663  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 3  
## GTEX.11DXZ.0926.SM.5N9CG GTEX.11DYG.0826.SM.5N9GH  
## ENSG00000223972.4 1 4  
## ENSG00000227232.4 768 954  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.11DZ1.2726.SM.5A5KH GTEX.11EI6.0726.SM.59866  
## ENSG00000223972.4 0 1  
## ENSG00000227232.4 525 619  
## ENSG00000243485.2 0 2  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 1  
## GTEX.11EM3.0126.SM.5985K GTEX.11EMC.0226.SM.5EGLP  
## ENSG00000223972.4 2 5  
## ENSG00000227232.4 669 786  
## ENSG00000243485.2 2 0  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 1 1  
## GTEX.11EQ8.0826.SM.5N9FG GTEX.11EQ9.0626.SM.5A5K1  
## ENSG00000223972.4 1 6  
## ENSG00000227232.4 802 640  
## ENSG00000243485.2 1 4  
## ENSG00000237613.2 0 3  
## ENSG00000268020.2 1 1  
## ENSG00000240361.1 0 3  
## GTEX.11GS4.0826.SM.5986J GTEX.11GSO.0626.SM.5A5LW  
## ENSG00000223972.4 0 0  
## ENSG00000227232.4 533 544  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 1 3  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.11I78.0526.SM.5986A GTEX.11LCK.0526.SM.5A5M9  
## ENSG00000223972.4 3 0  
## ENSG00000227232.4 507 315  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 5 1  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 3 1  
## GTEX.11NSD.0126.SM.5987F GTEX.11NUK.1026.SM.5HL5J  
## ENSG00000223972.4 0 5  
## ENSG00000227232.4 408 683  
## ENSG00000243485.2 1 3  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 2  
## GTEX.11NV4.0626.SM.5N9BR GTEX.11O72.2326.SM.5BC7H  
## ENSG00000223972.4 3 0  
## ENSG00000227232.4 1301 633  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 1  
## GTEX.11OF3.0626.SM.5BC4Y GTEX.11P7K.0226.SM.5986Z  
## ENSG00000223972.4 2 4  
## ENSG00000227232.4 518 840  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 3 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 2  
## GTEX.11P81.0126.SM.5HL5Y GTEX.11P82.0226.SM.5HL4O  
## ENSG00000223972.4 0 5  
## ENSG00000227232.4 319 634  
## ENSG00000243485.2 2 2  
## ENSG00000237613.2 2 2  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 2  
## GTEX.11TT1.1126.SM.5P9GV GTEX.11TTK.0826.SM.5N9EG  
## ENSG00000223972.4 2 3  
## ENSG00000227232.4 524 1241  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 2 2  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 1  
## GTEX.11TUW.0226.SM.5LU8X GTEX.11UD2.0626.SM.5GU6L  
## ENSG00000223972.4 4 1  
## ENSG00000227232.4 627 616  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.11VI4.0226.SM.5GU6C GTEX.11XUK.0226.SM.5EQLW  
## ENSG00000223972.4 1 0  
## ENSG00000227232.4 571 419  
## ENSG00000243485.2 2 0  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 2 0  
## GTEX.11ZTS.1126.SM.5LU9X GTEX.11ZTT.1026.SM.5EQKF  
## ENSG00000223972.4 6 3  
## ENSG00000227232.4 547 580  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 3 0  
## GTEX.11ZVC.0126.SM.5986G GTEX.1211K.0726.SM.5FQUW  
## ENSG00000223972.4 0 3  
## ENSG00000227232.4 719 426  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 2  
## GTEX.1212Z.0426.SM.5FQT6 GTEX.12584.0826.SM.5FQSK  
## ENSG00000223972.4 2 1  
## ENSG00000227232.4 452 1064  
## ENSG00000243485.2 0 2  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 1 2  
## ENSG00000240361.1 0 2  
## GTEX.12696.0326.SM.5EGL4 GTEX.1269C.0226.SM.5EGKS  
## ENSG00000223972.4 5 1  
## ENSG00000227232.4 542 925  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 3 0  
## GTEX.12BJ1.0426.SM.5FQSO GTEX.12WSC.0826.SM.5EQ5Q  
## ENSG00000223972.4 1 4  
## ENSG00000227232.4 410 605  
## ENSG00000243485.2 2 1  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 0  
## GTEX.12WSD.0926.SM.5GCNL GTEX.12WSG.0226.SM.5EGIF  
## ENSG00000223972.4 0 3  
## ENSG00000227232.4 1029 369  
## ENSG00000243485.2 0 1  
## ENSG00000237613.2 0 3  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 1 2  
## GTEX.12WSH.0226.SM.5GCOG GTEX.12WSJ.0326.SM.5GCMT  
## ENSG00000223972.4 3 4  
## ENSG00000227232.4 641 892  
## ENSG00000243485.2 0 5  
## ENSG00000237613.2 0 4  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 2  
## GTEX.12WSK.0926.SM.5CVNQ GTEX.12WSL.0626.SM.5GCOY  
## ENSG00000223972.4 2 2  
## ENSG00000227232.4 731 335  
## ENSG00000243485.2 3 0  
## ENSG00000237613.2 2 4  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 3 0  
## GTEX.12WSN.0726.SM.5GCMS GTEX.12ZZX.1226.SM.5EGHS  
## ENSG00000223972.4 1 2  
## ENSG00000227232.4 481 679  
## ENSG00000243485.2 0 2  
## ENSG00000237613.2 1 4  
## ENSG00000268020.2 2 0  
## ENSG00000240361.1 1 1  
## GTEX.12ZZY.0826.SM.5EQMT GTEX.12ZZZ.1226.SM.59HK1  
## ENSG00000223972.4 2 0  
## ENSG00000227232.4 491 424  
## ENSG00000243485.2 2 0  
## ENSG00000237613.2 0 2  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 0  
## GTEX.13111.0226.SM.5EQ55 GTEX.13112.0326.SM.5P9IW  
## ENSG00000223972.4 3 3  
## ENSG00000227232.4 540 600  
## ENSG00000243485.2 3 1  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 4  
## ENSG00000240361.1 1 0  
## GTEX.13113.0126.SM.5LZVX GTEX.1313W.0726.SM.5EGK1  
## ENSG00000223972.4 1 2  
## ENSG00000227232.4 687 890  
## ENSG00000243485.2 0 1  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 2 0  
## GTEX.131XE.0126.SM.5LZVC GTEX.131XF.1826.SM.5EGKG  
## ENSG00000223972.4 1 5  
## ENSG00000227232.4 623 656  
## ENSG00000243485.2 0 1  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 1  
## GTEX.131XG.0226.SM.5IFG1 GTEX.131XH.0526.SM.5DUX7  
## ENSG00000223972.4 0 0  
## ENSG00000227232.4 325 766  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 2 2  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 2 1  
## GTEX.131YS.0726.SM.5P9G9 GTEX.132AR.1126.SM.5P9GA  
## ENSG00000223972.4 1 0  
## ENSG00000227232.4 1487 907  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 2 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 3 0  
## GTEX.132NY.1026.SM.5P9IY GTEX.132QS.0326.SM.5IJFN  
## ENSG00000223972.4 3 1  
## ENSG00000227232.4 781 604  
## ENSG00000243485.2 0 1  
## ENSG00000237613.2 2 0  
## ENSG00000268020.2 1 2  
## ENSG00000240361.1 0 1  
## GTEX.133LE.0326.SM.5P9G4 GTEX.1399R.0126.SM.5IFEV  
## ENSG00000223972.4 4 3  
## ENSG00000227232.4 1175 517  
## ENSG00000243485.2 2 2  
## ENSG00000237613.2 3 1  
## ENSG00000268020.2 2 1  
## ENSG00000240361.1 0 3  
## GTEX.1399T.0126.SM.5KM15 GTEX.1399U.0326.SM.5P9G5  
## ENSG00000223972.4 2 5  
## ENSG00000227232.4 855 934  
## ENSG00000243485.2 2 3  
## ENSG00000237613.2 2 2  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 0 0  
## GTEX.139T6.0326.SM.5J2LY GTEX.139TS.0126.SM.5K7XJ  
## ENSG00000223972.4 1 0  
## ENSG00000227232.4 639 485  
## ENSG00000243485.2 4 0  
## ENSG00000237613.2 5 1  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 4 1  
## GTEX.139UW.0126.SM.5KM1B GTEX.139YR.1226.SM.5IFEU  
## ENSG00000223972.4 2 4  
## ENSG00000227232.4 430 240  
## ENSG00000243485.2 0 1  
## ENSG00000237613.2 0 2  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.13CF3.0926.SM.5LZZC GTEX.13D11.0226.SM.5LZXL  
## ENSG00000223972.4 2 0  
## ENSG00000227232.4 441 426  
## ENSG00000243485.2 1 3  
## ENSG00000237613.2 2 4  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 1 1  
## GTEX.13FH7.0126.SM.5KLZ1 GTEX.13FHO.0926.SM.5N9EW  
## ENSG00000223972.4 5 4  
## ENSG00000227232.4 576 907  
## ENSG00000243485.2 4 3  
## ENSG00000237613.2 3 2  
## ENSG00000268020.2 0 2  
## ENSG00000240361.1 1 2  
## GTEX.13FHP.0926.SM.5L3EC GTEX.13FLV.0226.SM.5J2OF  
## ENSG00000223972.4 2 5  
## ENSG00000227232.4 1517 1572  
## ENSG00000243485.2 0 1  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 2  
## ENSG00000240361.1 0 2  
## GTEX.13FLW.0326.SM.5J2M4 GTEX.13FTW.0626.SM.5IFEX  
## ENSG00000223972.4 2 0  
## ENSG00000227232.4 1245 561  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 3 0  
## GTEX.13FTY.0726.SM.5J2OH GTEX.13FXS.0726.SM.5LZXJ  
## ENSG00000223972.4 1 5  
## ENSG00000227232.4 675 1564  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 3 2  
## ENSG00000268020.2 1 1  
## ENSG00000240361.1 1 0  
## GTEX.13G51.1226.SM.5K7Z3 GTEX.13IVO.0926.SM.5KLZP  
## ENSG00000223972.4 0 4  
## ENSG00000227232.4 878 700  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 1 3  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 0  
## GTEX.13JVG.0926.SM.5IJE1 GTEX.13N11.1026.SM.5K7XQ  
## ENSG00000223972.4 4 2  
## ENSG00000227232.4 568 427  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 3 0  
## GTEX.13N1W.0826.SM.5MR5J GTEX.13N2G.0726.SM.5MR38  
## ENSG00000223972.4 0 2  
## ENSG00000227232.4 851 854  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 1  
## GTEX.13NYB.0726.SM.5MR4J GTEX.13NYC.2426.SM.5MR3K  
## ENSG00000223972.4 1 0  
## ENSG00000227232.4 1380 949  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.13NZ8.0226.SM.5J2OK GTEX.13NZ9.1126.SM.5MR37  
## ENSG00000223972.4 1 0  
## ENSG00000227232.4 1164 1002  
## ENSG00000243485.2 2 1  
## ENSG00000237613.2 2 0  
## ENSG00000268020.2 2 0  
## ENSG00000240361.1 0 1  
## GTEX.13NZA.1026.SM.5MR48 GTEX.13O1R.0826.SM.5J2MB  
## ENSG00000223972.4 1 3  
## ENSG00000227232.4 902 460  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 0 2  
## ENSG00000240361.1 1 0  
## GTEX.13O21.2226.SM.5MR3L GTEX.13O3O.0926.SM.5KM1F  
## ENSG00000223972.4 1 2  
## ENSG00000227232.4 885 754  
## ENSG00000243485.2 1 3  
## ENSG00000237613.2 0 2  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 1  
## GTEX.13O3P.0726.SM.5J2OM GTEX.13O3Q.0626.SM.5IJG1  
## ENSG00000223972.4 2 2  
## ENSG00000227232.4 808 791  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 4 0  
## ENSG00000268020.2 1 1  
## ENSG00000240361.1 0 0  
## GTEX.13O61.0226.SM.5KM52 GTEX.13OVG.0226.SM.5LU93  
## ENSG00000223972.4 2 4  
## ENSG00000227232.4 624 719  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 2 2  
## ENSG00000240361.1 1 2  
## GTEX.13OVI.0826.SM.5KLZ8 GTEX.13OVJ.0626.SM.5J2O2  
## ENSG00000223972.4 2 5  
## ENSG00000227232.4 723 1307  
## ENSG00000243485.2 2 4  
## ENSG00000237613.2 2 3  
## ENSG00000268020.2 3 0  
## ENSG00000240361.1 5 1  
## GTEX.13OW5.0626.SM.5J2N2 GTEX.13OW6.0726.SM.5L3FX  
## ENSG00000223972.4 1 5  
## ENSG00000227232.4 1042 723  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 4 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 2 0  
## GTEX.13OW7.0826.SM.5L3EL GTEX.13OW8.0126.SM.5IJE5  
## ENSG00000223972.4 0 2  
## ENSG00000227232.4 629 910  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 0  
## GTEX.13PDP.1026.SM.5L3FA GTEX.13PL6.1026.SM.5L3E5  
## ENSG00000223972.4 0 0  
## ENSG00000227232.4 1487 561  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 2 0  
## GTEX.13PVQ.0726.SM.5L3GI GTEX.13QBU.0626.SM.5J2OG  
## ENSG00000223972.4 1 7  
## ENSG00000227232.4 760 695  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 0 2  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 1  
## GTEX.13QJC.0826.SM.5RQKC GTEX.13RTK.0326.SM.5RQHS  
## ENSG00000223972.4 0 0  
## ENSG00000227232.4 825 627  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 1 1  
## GTEX.13S86.1126.SM.5RQJX GTEX.13U4I.0526.SM.5LU59  
## ENSG00000223972.4 1 2  
## ENSG00000227232.4 800 636  
## ENSG00000243485.2 2 0  
## ENSG00000237613.2 2 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 0  
## GTEX.13VXT.0626.SM.5SIA1 GTEX.13VXU.0826.SM.5KLZ2  
## ENSG00000223972.4 1 1  
## ENSG00000227232.4 899 440  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 1 0  
## GTEX.13W46.0926.SM.5LU3T GTEX.13X6H.0526.SM.5LU4Q  
## ENSG00000223972.4 2 2  
## ENSG00000227232.4 872 468  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 1  
## GTEX.13X6J.0826.SM.5LU32 GTEX.13YAN.0926.SM.5O9C3  
## ENSG00000223972.4 1 1  
## ENSG00000227232.4 720 960  
## ENSG00000243485.2 0 2  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 2 0  
## GTEX.144GL.1226.SM.5O9A4 GTEX.144GM.0226.SM.5Q5CB  
## ENSG00000223972.4 3 5  
## ENSG00000227232.4 913 658  
## ENSG00000243485.2 1 3  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 3 1  
## GTEX.144GO.0126.SM.5LUAO GTEX.145LU.0426.SM.5O9AH  
## ENSG00000223972.4 1 2  
## ENSG00000227232.4 482 851  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 1 0  
## GTEX.145ME.0126.SM.5S2QM GTEX.145MG.0826.SM.5Q5C2  
## ENSG00000223972.4 1 0  
## ENSG00000227232.4 588 710  
## ENSG00000243485.2 2 0  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 0  
## GTEX.145MH.0426.SM.5LU8T GTEX.145MI.1126.SM.5O9AK  
## ENSG00000223972.4 3 0  
## ENSG00000227232.4 544 890  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 1 1  
## GTEX.146FQ.0726.SM.5LUA7 GTEX.146FR.0326.SM.5SI8U  
## ENSG00000223972.4 1 0  
## ENSG00000227232.4 467 514  
## ENSG00000243485.2 2 1  
## ENSG00000237613.2 2 2  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 1  
## GTEX.14753.0926.SM.5Q5BI GTEX.147F4.0826.SM.5QGRB  
## ENSG00000223972.4 2 0  
## ENSG00000227232.4 582 948  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.147GR.0726.SM.5S2PL GTEX.148VJ.0726.SM.5LU8J  
## ENSG00000223972.4 2 1  
## ENSG00000227232.4 542 564  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 2  
## ENSG00000240361.1 0 1  
## GTEX.1497J.0126.SM.5Q5BK GTEX.14A5H.0726.SM.5Q5DW  
## ENSG00000223972.4 2 2  
## ENSG00000227232.4 751 742  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 3 2  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 1 1  
## GTEX.14A6H.2426.SM.5Q5BO GTEX.14ABY.0926.SM.5Q5DY  
## ENSG00000223972.4 0 1  
## ENSG00000227232.4 1325 775  
## ENSG00000243485.2 2 2  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.14AS3.0226.SM.5Q5B6 GTEX.14ASI.0726.SM.5Q5DC  
## ENSG00000223972.4 0 0  
## ENSG00000227232.4 834 1246  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 1  
## GTEX.14BMU.0226.SM.5S2QA GTEX.14C38.0826.SM.5S2U8  
## ENSG00000223972.4 2 10  
## ENSG00000227232.4 423 879  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 2 0  
## ENSG00000240361.1 1 1  
## GTEX.14E6C.2626.SM.5RQJP GTEX.14ICL.0426.SM.5RQJ3  
## ENSG00000223972.4 0 2  
## ENSG00000227232.4 655 703  
## ENSG00000243485.2 2 2  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.N7MS.2326.SM.2HMLD GTEX.NFK9.0726.SM.2HMJW  
## ENSG00000223972.4 8 9  
## ENSG00000227232.4 493 889  
## ENSG00000243485.2 5 5  
## ENSG00000237613.2 5 3  
## ENSG00000268020.2 7 2  
## ENSG00000240361.1 0 6  
## GTEX.OHPK.2626.SM.2HMK9 GTEX.OHPM.2626.SM.33HC5  
## ENSG00000223972.4 17 12  
## ENSG00000227232.4 484 772  
## ENSG00000243485.2 10 3  
## ENSG00000237613.2 7 3  
## ENSG00000268020.2 6 1  
## ENSG00000240361.1 5 4  
## GTEX.OIZG.0226.SM.2TC5L GTEX.OIZI.0726.SM.2XCEI  
## ENSG00000223972.4 6 0  
## ENSG00000227232.4 313 523  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 1 3  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 3 3  
## GTEX.OXRK.0626.SM.2HMJ5 GTEX.OXRL.2626.SM.2I3F1  
## ENSG00000223972.4 11 8  
## ENSG00000227232.4 873 690  
## ENSG00000243485.2 3 8  
## ENSG00000237613.2 3 8  
## ENSG00000268020.2 1 4  
## ENSG00000240361.1 2 2  
## GTEX.OXRO.1226.SM.48TDL GTEX.OXRP.0326.SM.33HBJ  
## ENSG00000223972.4 1 3  
## ENSG00000227232.4 967 1205  
## ENSG00000243485.2 0 4  
## ENSG00000237613.2 1 3  
## ENSG00000268020.2 0 3  
## ENSG00000240361.1 1 1  
## GTEX.P4PQ.2626.SM.33HC9 GTEX.P4QS.2626.SM.2I3EV  
## ENSG00000223972.4 13 4  
## ENSG00000227232.4 1859 513  
## ENSG00000243485.2 6 9  
## ENSG00000237613.2 10 10  
## ENSG00000268020.2 1 3  
## ENSG00000240361.1 6 4  
## GTEX.P4QT.2626.SM.2I3FM GTEX.P78B.0526.SM.2I5F7  
## ENSG00000223972.4 5 8  
## ENSG00000227232.4 825 548  
## ENSG00000243485.2 4 11  
## ENSG00000237613.2 2 6  
## ENSG00000268020.2 2 1  
## ENSG00000240361.1 2 4  
## GTEX.PLZ4.1226.SM.2I5FE GTEX.PWN1.2626.SM.2I3FH  
## ENSG00000223972.4 5 5  
## ENSG00000227232.4 489 297  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 3 0  
## ENSG00000268020.2 2 1  
## ENSG00000240361.1 1 2  
## GTEX.PX3G.2626.SM.2I3EG GTEX.Q2AG.0826.SM.2HMKF  
## ENSG00000223972.4 8 10  
## ENSG00000227232.4 636 759  
## ENSG00000243485.2 5 7  
## ENSG00000237613.2 2 3  
## ENSG00000268020.2 2 4  
## ENSG00000240361.1 3 5  
## GTEX.Q2AH.0726.SM.2I3EA GTEX.Q2AI.0326.SM.2I3EK  
## ENSG00000223972.4 1 3  
## ENSG00000227232.4 874 442  
## ENSG00000243485.2 8 8  
## ENSG00000237613.2 2 3  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 1 0  
## GTEX.Q734.0526.SM.2I3EH GTEX.QDVJ.0226.SM.2I5FV  
## ENSG00000223972.4 2 15  
## ENSG00000227232.4 749 944  
## ENSG00000243485.2 2 10  
## ENSG00000237613.2 3 2  
## ENSG00000268020.2 2 1  
## ENSG00000240361.1 2 4  
## GTEX.QDVN.0626.SM.2I3FP GTEX.QEG5.0826.SM.2I5GF  
## ENSG00000223972.4 6 4  
## ENSG00000227232.4 511 978  
## ENSG00000243485.2 7 1  
## ENSG00000237613.2 6 3  
## ENSG00000268020.2 1 2  
## ENSG00000240361.1 2 2  
## GTEX.QEL4.0726.SM.3GIJ5 GTEX.QLQ7.0726.SM.2I5G2  
## ENSG00000223972.4 4 6  
## ENSG00000227232.4 511 666  
## ENSG00000243485.2 2 3  
## ENSG00000237613.2 2 3  
## ENSG00000268020.2 4 2  
## ENSG00000240361.1 1 4  
## GTEX.QV31.0726.SM.3GAEG GTEX.QV44.0826.SM.2S1RG  
## ENSG00000223972.4 3 2  
## ENSG00000227232.4 450 412  
## ENSG00000243485.2 4 0  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 2 0  
## ENSG00000240361.1 2 0  
## GTEX.QXCU.0326.SM.2TC63 GTEX.R3RS.0726.SM.3GIJR  
## ENSG00000223972.4 3 1  
## ENSG00000227232.4 453 176  
## ENSG00000243485.2 4 0  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 2 1  
## GTEX.R53T.0526.SM.3GADL GTEX.R55C.0626.SM.2TF4Q  
## ENSG00000223972.4 9 9  
## ENSG00000227232.4 782 302  
## ENSG00000243485.2 2 4  
## ENSG00000237613.2 2 2  
## ENSG00000268020.2 1 2  
## ENSG00000240361.1 2 0  
## GTEX.R55E.0826.SM.2TC5M GTEX.R55G.0726.SM.2TC6J  
## ENSG00000223972.4 3 3  
## ENSG00000227232.4 533 134  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 0 2  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 0  
## GTEX.REY6.0526.SM.2TF5M GTEX.RM2N.0526.SM.2TF4N  
## ENSG00000223972.4 1 3  
## ENSG00000227232.4 590 406  
## ENSG00000243485.2 3 4  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 2 1  
## GTEX.RN64.0626.SM.2TC5V GTEX.RNOR.0926.SM.2TF56  
## ENSG00000223972.4 2 3  
## ENSG00000227232.4 1311 807  
## ENSG00000243485.2 0 4  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 2 1  
## ENSG00000240361.1 2 1  
## GTEX.RTLS.0626.SM.5SI7Z GTEX.RU1J.0226.SM.2TF5Y  
## ENSG00000223972.4 4 4  
## ENSG00000227232.4 1210 415  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 1 3  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 1 2  
## GTEX.RU72.0126.SM.2TF6Z GTEX.RUSQ.1026.SM.2TF6V  
## ENSG00000223972.4 3 8  
## ENSG00000227232.4 323 471  
## ENSG00000243485.2 4 5  
## ENSG00000237613.2 1 4  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 1 2  
## GTEX.RVPV.1226.SM.2TF73 GTEX.RWS6.0626.SM.2XCAS  
## ENSG00000223972.4 5 6  
## ENSG00000227232.4 304 801  
## ENSG00000243485.2 3 1  
## ENSG00000237613.2 0 3  
## ENSG00000268020.2 2 1  
## ENSG00000240361.1 3 2  
## GTEX.RWSA.0826.SM.2XCBF GTEX.S32W.0726.SM.2XCBL  
## ENSG00000223972.4 4 6  
## ENSG00000227232.4 852 296  
## ENSG00000243485.2 4 6  
## ENSG00000237613.2 4 9  
## ENSG00000268020.2 1 3  
## ENSG00000240361.1 3 2  
## GTEX.S341.0226.SM.5S2VG GTEX.S7SE.0726.SM.2XCD7  
## ENSG00000223972.4 2 4  
## ENSG00000227232.4 457 422  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 2 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 2 2  
## GTEX.S7SF.0226.SM.5SI7H GTEX.SE5C.0726.SM.4BRWY  
## ENSG00000223972.4 3 1  
## ENSG00000227232.4 487 622  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 2 1  
## ENSG00000268020.2 0 2  
## ENSG00000240361.1 0 1  
## GTEX.SIU8.0626.SM.2XCDN GTEX.SJXC.0726.SM.2XCFJ  
## ENSG00000223972.4 3 4  
## ENSG00000227232.4 899 548  
## ENSG00000243485.2 2 2  
## ENSG00000237613.2 2 1  
## ENSG00000268020.2 0 2  
## ENSG00000240361.1 1 1  
## GTEX.T2IS.0626.SM.32QP6 GTEX.T5JW.1226.SM.3GACY  
## ENSG00000223972.4 0 1  
## ENSG00000227232.4 976 541  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 2 0  
## GTEX.T6MN.0626.SM.32PM9 GTEX.T6MO.0226.SM.32QOL  
## ENSG00000223972.4 1 3  
## ENSG00000227232.4 892 909  
## ENSG00000243485.2 1 5  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 1 2  
## ENSG00000240361.1 0 1  
## GTEX.T8EM.0226.SM.3DB7C GTEX.TKQ1.0126.SM.33HB3  
## ENSG00000223972.4 8 8  
## ENSG00000227232.4 1688 1388  
## ENSG00000243485.2 6 1  
## ENSG00000237613.2 4 1  
## ENSG00000268020.2 1 1  
## ENSG00000240361.1 6 0  
## GTEX.TMMY.0826.SM.33HB9 GTEX.TSE9.0626.SM.3DB8B  
## ENSG00000223972.4 3 3  
## ENSG00000227232.4 979 1704  
## ENSG00000243485.2 3 1  
## ENSG00000237613.2 2 1  
## ENSG00000268020.2 5 0  
## ENSG00000240361.1 8 0  
## GTEX.U8T8.2326.SM.3DB96 GTEX.VUSG.0426.SM.3GIKD  
## ENSG00000223972.4 3 4  
## ENSG00000227232.4 841 565  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 1  
## GTEX.W5X1.0426.SM.3GILB GTEX.WHPG.0226.SM.3NMB9  
## ENSG00000223972.4 5 0  
## ENSG00000227232.4 560 658  
## ENSG00000243485.2 2 0  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 1 1  
## ENSG00000240361.1 3 1  
## GTEX.WHSB.1626.SM.3LK6J GTEX.WHSE.0626.SM.4RGNF  
## ENSG00000223972.4 8 4  
## ENSG00000227232.4 883 381  
## ENSG00000243485.2 2 1  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 1 1  
## GTEX.WK11.0926.SM.3NMAU GTEX.WL46.0126.SM.3TW8I  
## ENSG00000223972.4 1 0  
## ENSG00000227232.4 895 1133  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 0  
## GTEX.WOFL.0726.SM.3MJG4 GTEX.WYBS.1926.SM.3NM8N  
## ENSG00000223972.4 1 7  
## ENSG00000227232.4 837 730  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 0  
## GTEX.WYJK.1626.SM.3NM9J GTEX.WYVS.0326.SM.3NM9V  
## ENSG00000223972.4 5 6  
## ENSG00000227232.4 980 820  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 4  
## GTEX.X15G.0526.SM.3NMB7 GTEX.X4XY.0826.SM.4E3JM  
## ENSG00000223972.4 26 1  
## ENSG00000227232.4 810 1071  
## ENSG00000243485.2 6 1  
## ENSG00000237613.2 7 3  
## ENSG00000268020.2 4 0  
## ENSG00000240361.1 6 0  
## GTEX.X8HC.0726.SM.46MWG GTEX.XBED.0126.SM.47JY7  
## ENSG00000223972.4 0 3  
## ENSG00000227232.4 879 766  
## ENSG00000243485.2 1 3  
## ENSG00000237613.2 1 4  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 2 4  
## GTEX.XBEW.0126.SM.4AT66 GTEX.XGQ4.0426.SM.4AT4I  
## ENSG00000223972.4 1 0  
## ENSG00000227232.4 473 492  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 2 0  
## ENSG00000268020.2 1 1  
## ENSG00000240361.1 1 0  
## GTEX.XLM4.0726.SM.4AT64 GTEX.XMK1.0626.SM.4B65A  
## ENSG00000223972.4 1 9  
## ENSG00000227232.4 647 568  
## ENSG00000243485.2 0 1  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 0  
## GTEX.XUW1.1026.SM.4BONY GTEX.XUZC.0126.SM.4BOO6  
## ENSG00000223972.4 2 0  
## ENSG00000227232.4 688 445  
## ENSG00000243485.2 1 0  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 1  
## GTEX.XV7Q.0326.SM.4BRVM GTEX.XXEK.1326.SM.4BRV1  
## ENSG00000223972.4 2 3  
## ENSG00000227232.4 837 818  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 0 3  
## ENSG00000268020.2 1 1  
## ENSG00000240361.1 0 2  
## GTEX.XYKS.0826.SM.4BRVF GTEX.Y111.1926.SM.4SOIS  
## ENSG00000223972.4 4 2  
## ENSG00000227232.4 432 810  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 1 2  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 1 1  
## GTEX.Y114.0626.SM.4TT98 GTEX.Y3IK.0526.SM.4WWE3  
## ENSG00000223972.4 3 5  
## ENSG00000227232.4 688 1372  
## ENSG00000243485.2 4 1  
## ENSG00000237613.2 3 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 2 2  
## GTEX.Y5LM.0626.SM.4V6G4 GTEX.Y5V5.0326.SM.5RQJG  
## ENSG00000223972.4 2 1  
## ENSG00000227232.4 941 424  
## ENSG00000243485.2 3 0  
## ENSG00000237613.2 1 2  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 0  
## GTEX.Y5V6.0526.SM.4VBRV GTEX.Y8E4.0126.SM.4VBQ2  
## ENSG00000223972.4 3 11  
## ENSG00000227232.4 482 852  
## ENSG00000243485.2 3 4  
## ENSG00000237613.2 2 3  
## ENSG00000268020.2 2 2  
## ENSG00000240361.1 2 3  
## GTEX.Y9LG.0426.SM.4VBRT GTEX.YB5E.0626.SM.4VDSE  
## ENSG00000223972.4 5 3  
## ENSG00000227232.4 1034 700  
## ENSG00000243485.2 5 2  
## ENSG00000237613.2 5 3  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 3 2  
## GTEX.YB5K.0526.SM.5LUAS GTEX.YEC3.0826.SM.4WWFP  
## ENSG00000223972.4 6 4  
## ENSG00000227232.4 2278 705  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 2 0  
## ENSG00000268020.2 1 1  
## ENSG00000240361.1 3 1  
## GTEX.YEC4.0626.SM.5CVLU GTEX.YF7O.0726.SM.4W213  
## ENSG00000223972.4 1 4  
## ENSG00000227232.4 365 912  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 0 3  
## ENSG00000240361.1 1 1  
## GTEX.YFC4.2626.SM.5P9FQ GTEX.YFCO.0326.SM.4W1ZP  
## ENSG00000223972.4 1 7  
## ENSG00000227232.4 1472 1000  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 0 5  
## ENSG00000268020.2 0 3  
## ENSG00000240361.1 1 5  
## GTEX.YJ89.0726.SM.5P9F7 GTEX.Z9EW.0226.SM.5CVM7  
## ENSG00000223972.4 4 3  
## ENSG00000227232.4 1325 450  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 0 2  
## ENSG00000268020.2 2 0  
## ENSG00000240361.1 1 1  
## GTEX.ZA64.0426.SM.5HL96 GTEX.ZAB5.0726.SM.5P9JG  
## ENSG00000223972.4 2 1  
## ENSG00000227232.4 388 676  
## ENSG00000243485.2 2 1  
## ENSG00000237613.2 1 0  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 2 1  
## GTEX.ZAJG.0726.SM.5HL9A GTEX.ZAK1.0726.SM.5HL8Q  
## ENSG00000223972.4 3 0  
## ENSG00000227232.4 755 555  
## ENSG00000243485.2 2 1  
## ENSG00000237613.2 1 2  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 1  
## GTEX.ZC5H.0626.SM.5LU9K GTEX.ZDYS.0626.SM.5J2N5  
## ENSG00000223972.4 3 5  
## ENSG00000227232.4 644 637  
## ENSG00000243485.2 1 3  
## ENSG00000237613.2 0 4  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 0  
## GTEX.ZE7O.1126.SM.57WC8 GTEX.ZF28.0826.SM.4WKGJ  
## ENSG00000223972.4 1 3  
## ENSG00000227232.4 713 1018  
## ENSG00000243485.2 2 0  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 1  
## GTEX.ZGAY.1026.SM.4WWBR GTEX.ZLFU.0626.SM.4WWBO  
## ENSG00000223972.4 1 8  
## ENSG00000227232.4 630 1058  
## ENSG00000243485.2 0 3  
## ENSG00000237613.2 3 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 2  
## GTEX.ZLV1.0126.SM.4WWBZ GTEX.ZLWG.0526.SM.4WWFB  
## ENSG00000223972.4 2 5  
## ENSG00000227232.4 689 1023  
## ENSG00000243485.2 2 4  
## ENSG00000237613.2 4 3  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 2 0  
## GTEX.ZPCL.0126.SM.4WWC8 GTEX.ZPU1.0426.SM.4WWCA  
## ENSG00000223972.4 1 3  
## ENSG00000227232.4 534 904  
## ENSG00000243485.2 3 1  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 2 2  
## ENSG00000240361.1 4 1  
## GTEX.ZQG8.0926.SM.57WFF GTEX.ZT9W.0226.SM.4YCCZ  
## ENSG00000223972.4 3 18  
## ENSG00000227232.4 554 431  
## ENSG00000243485.2 1 5  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 1 2  
## ENSG00000240361.1 1 2  
## GTEX.ZT9X.0226.SM.51MT2 GTEX.ZTPG.0826.SM.5DUVC  
## ENSG00000223972.4 1 1  
## ENSG00000227232.4 830 524  
## ENSG00000243485.2 2 1  
## ENSG00000237613.2 4 0  
## ENSG00000268020.2 2 0  
## ENSG00000240361.1 2 0  
## GTEX.ZTSS.0226.SM.59877 GTEX.ZTX8.0626.SM.59HKC  
## ENSG00000223972.4 3 3  
## ENSG00000227232.4 619 586  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 2 2  
## ENSG00000268020.2 0 2  
## ENSG00000240361.1 0 0  
## GTEX.ZUA1.0926.SM.4YCDX GTEX.ZV6S.0226.SM.59HJT  
## ENSG00000223972.4 1 5  
## ENSG00000227232.4 739 515  
## ENSG00000243485.2 2 1  
## ENSG00000237613.2 2 2  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.ZV7C.0126.SM.57WDE GTEX.ZVP2.0426.SM.57WC2  
## ENSG00000223972.4 3 5  
## ENSG00000227232.4 990 659  
## ENSG00000243485.2 0 2  
## ENSG00000237613.2 1 5  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 1 2  
## GTEX.ZVT3.0726.SM.5GICN GTEX.ZVZP.1026.SM.5GICI  
## ENSG00000223972.4 1 5  
## ENSG00000227232.4 742 529  
## ENSG00000243485.2 0 2  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 1  
## GTEX.ZVZQ.0626.SM.59HJU GTEX.ZXG5.0926.SM.5NQ8H  
## ENSG00000223972.4 4 1  
## ENSG00000227232.4 1560 740  
## ENSG00000243485.2 3 3  
## ENSG00000237613.2 3 3  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 0  
## GTEX.ZY6K.0226.SM.5SIAY GTEX.ZYFC.0926.SM.5GZWW  
## ENSG00000223972.4 2 1  
## ENSG00000227232.4 587 1051  
## ENSG00000243485.2 1 3  
## ENSG00000237613.2 2 2  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 0  
## GTEX.ZYFD.0826.SM.5NQ9A GTEX.ZYFG.0626.SM.5GZYA  
## ENSG00000223972.4 0 2  
## ENSG00000227232.4 643 502  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 1  
## GTEX.ZYT6.0426.SM.5GID3 GTEX.ZYVF.1126.SM.5E458  
## ENSG00000223972.4 5 2  
## ENSG00000227232.4 772 838  
## ENSG00000243485.2 2 1  
## ENSG00000237613.2 1 4  
## ENSG00000268020.2 0 1  
## ENSG00000240361.1 0 1  
## GTEX.ZYW4.1126.SM.5SI99 GTEX.ZYY3.1926.SM.5GZXS  
## ENSG00000223972.4 1 6  
## ENSG00000227232.4 584 1003  
## ENSG00000243485.2 0 1  
## ENSG00000237613.2 0 2  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 1  
## GTEX.ZZ64.0126.SM.5GZXA GTEX.ZZPU.1326.SM.5GZWS  
## ENSG00000223972.4 7 3  
## ENSG00000227232.4 732 555  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 1 2  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 2 0

dim(countsOriginal)

## [1] 56202 292

Una vez importados los datos targets genero un nuevo data frame con 10 muestras aleatorias de cada grupo utilizando las funciones *‘subset()’* y *‘sample\_n()’* tal y como se muestra a continuación:

require(dplyr)

## Loading required package: dplyr

## Warning: package 'dplyr' was built under R version 3.6.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

# NIT:  
muestrasNIT<-subset(targetsOriginal, targetsOriginal$Group=="NIT")  
muestrasNIT<-sample\_n(muestrasNIT, size = 10)  
  
# ELI:  
muestrasELI<-subset(targetsOriginal, targetsOriginal$Group=="ELI")  
muestrasELI<-sample\_n(muestrasELI, size = 10)  
  
# SFI:  
muestrasSFI<-subset(targetsOriginal, targetsOriginal$Group=="SFI")  
muestrasSFI<-sample\_n(muestrasSFI, size = 10)

Genero un dataframe con las 10 muestras de cada grupo:

targets<-rbind(muestrasELI, muestrasNIT, muestrasSFI)  
dim(targets)

## [1] 30 9

head(targets)

## Experiment SRA\_Sample Sample\_Name Grupo\_analisis body\_site  
## 1 SRX615373 SRS644099 GTEX-YFC4-2626-SM-5P9FQ 3 Thyroid  
## 2 SRX628009 SRS648152 GTEX-11NV4-0626-SM-5N9BR 3 Thyroid  
## 3 SRX199272 SRS333099 GTEX-PLZ4-1226-SM-2I5FE 3 Thyroid  
## 4 SRX204036 SRS374975 GTEX-R55G-0726-SM-2TC6J 3 Thyroid  
## 5 SRX222429 SRS389623 GTEX-TMMY-0826-SM-33HB9 3 Thyroid  
## 6 SRX619829 SRS644736 GTEX-11XUK-0226-SM-5EQLW 3 Thyroid  
## molecular\_data\_type sex Group ShortName  
## 1 Allele-Specific Expression female ELI YFC4-\_ELI  
## 2 RNA Seq (NGS) male ELI 11NV4\_ELI  
## 3 RNA Seq (NGS) female ELI PLZ4-\_ELI  
## 4 RNA Seq (NGS) female ELI R55G-\_ELI  
## 5 Allele-Specific Expression female ELI TMMY-\_ELI  
## 6 RNA Seq (NGS) female ELI 11XUK\_ELI

Por último modifico los nombres de las muestras del archivo targets para que coincidan con los nombres de las muestras del archivo counts y selecciono las muestras de counts que aparecen en la variable targets de 30 muestras aleatorias generada anteriormente:

targets$Sample\_Name<-gsub("-",".",targets$Sample\_Name)

counts<-countsOriginal[,targets$Sample\_Name]  
dim(counts)

## [1] 56202 30

head(counts)

## GTEX.YFC4.2626.SM.5P9FQ GTEX.11NV4.0626.SM.5N9BR  
## ENSG00000223972.4 1 3  
## ENSG00000227232.4 1472 1301  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 1  
## GTEX.PLZ4.1226.SM.2I5FE GTEX.R55G.0726.SM.2TC6J  
## ENSG00000223972.4 5 3  
## ENSG00000227232.4 489 134  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 3 2  
## ENSG00000268020.2 2 1  
## ENSG00000240361.1 1 0  
## GTEX.TMMY.0826.SM.33HB9 GTEX.11XUK.0226.SM.5EQLW  
## ENSG00000223972.4 3 0  
## ENSG00000227232.4 979 419  
## ENSG00000243485.2 3 0  
## ENSG00000237613.2 2 1  
## ENSG00000268020.2 5 0  
## ENSG00000240361.1 8 0  
## GTEX.111VG.0526.SM.5N9BW GTEX.YJ89.0726.SM.5P9F7  
## ENSG00000223972.4 1 4  
## ENSG00000227232.4 474 1325  
## ENSG00000243485.2 1 1  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 1 2  
## ENSG00000240361.1 1 1  
## GTEX.13QJC.0826.SM.5RQKC GTEX.14ABY.0926.SM.5Q5DY  
## ENSG00000223972.4 0 1  
## ENSG00000227232.4 825 775  
## ENSG00000243485.2 1 2  
## ENSG00000237613.2 0 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 1 0  
## GTEX.OXRK.0626.SM.2HMJ5 GTEX.13OVG.0226.SM.5LU93  
## ENSG00000223972.4 11 4  
## ENSG00000227232.4 873 719  
## ENSG00000243485.2 3 2  
## ENSG00000237613.2 3 1  
## ENSG00000268020.2 1 2  
## ENSG00000240361.1 2 2  
## GTEX.PWN1.2626.SM.2I3FH GTEX.11TT1.1126.SM.5P9GV  
## ENSG00000223972.4 5 2  
## ENSG00000227232.4 297 524  
## ENSG00000243485.2 0 1  
## ENSG00000237613.2 0 2  
## ENSG00000268020.2 1 0  
## ENSG00000240361.1 2 0  
## GTEX.ZLFU.0626.SM.4WWBO GTEX.1122O.0226.SM.5N9DA  
## ENSG00000223972.4 8 2  
## ENSG00000227232.4 1058 732  
## ENSG00000243485.2 3 1  
## ENSG00000237613.2 1 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 2 0  
## GTEX.ZXG5.0926.SM.5NQ8H GTEX.13FLV.0226.SM.5J2OF  
## ENSG00000223972.4 1 5  
## ENSG00000227232.4 740 1572  
## ENSG00000243485.2 3 1  
## ENSG00000237613.2 3 1  
## ENSG00000268020.2 0 2  
## ENSG00000240361.1 0 2  
## GTEX.14ICL.0426.SM.5RQJ3 GTEX.13OVJ.0626.SM.5J2O2  
## ENSG00000223972.4 2 5  
## ENSG00000227232.4 703 1307  
## ENSG00000243485.2 2 4  
## ENSG00000237613.2 0 3  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 1  
## GTEX.11DXY.0426.SM.5H12R GTEX.13NYC.2426.SM.5MR3K  
## ENSG00000223972.4 0 0  
## ENSG00000227232.4 663 949  
## ENSG00000243485.2 0 0  
## ENSG00000237613.2 0 1  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 3 0  
## GTEX.131XG.0226.SM.5IFG1 GTEX.Y5V6.0526.SM.4VBRV  
## ENSG00000223972.4 0 3  
## ENSG00000227232.4 325 482  
## ENSG00000243485.2 1 3  
## ENSG00000237613.2 2 2  
## ENSG00000268020.2 0 2  
## ENSG00000240361.1 2 2  
## GTEX.12584.0826.SM.5FQSK GTEX.ZYVF.1126.SM.5E458  
## ENSG00000223972.4 1 2  
## ENSG00000227232.4 1064 838  
## ENSG00000243485.2 2 1  
## ENSG00000237613.2 0 4  
## ENSG00000268020.2 2 1  
## ENSG00000240361.1 2 1  
## GTEX.ZLV1.0126.SM.4WWBZ GTEX.139UW.0126.SM.5KM1B  
## ENSG00000223972.4 2 2  
## ENSG00000227232.4 689 430  
## ENSG00000243485.2 2 0  
## ENSG00000237613.2 4 0  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 2 0  
## GTEX.R55E.0826.SM.2TC5M GTEX.Q2AH.0726.SM.2I3EA  
## ENSG00000223972.4 3 1  
## ENSG00000227232.4 533 874  
## ENSG00000243485.2 1 8  
## ENSG00000237613.2 0 2  
## ENSG00000268020.2 0 0  
## ENSG00000240361.1 0 1

Los nombres de las filas del archivo *‘counts’* correspondientes a los transcritos de los genes en estudio contienen el identificador de Ensembl pero también el número de la versión; antes de seguir con el estudio modifico estos nombres para que se correspondan únicamente con el identificador de Ensembl:

row.names(counts)<-gsub("\\..\*", "", row.names(counts))

Una vez preparados los datos de trabajo es posible realizar el análisis de expresión diferencial; existen varias opciones para ello. En este caso el análisis se va a llevar a cabo mediante R con el paquete de Bioconductor *‘DESeq2’*.

Poseemos dos tipos de datos: - el archivo counts contiene el número de lecturas de cada gen. Es un dataframe con los genes en las filas y las diferentes muestras en las columnas (en este caso se han seleccionado 10 muestras de cada grupo (NIT, SFI y ELI)).

* el archivo targets contiene información acerca de las muestras como por ejemplo el experimento, el nombre de la muestra, el grupo de análisis al que pertenece, la condición del experimento o el sexo del individuo del cual se extrajo la muestra.

Para la utilización del paquete *‘DESeq2’* es fundamental que los nombres de las muestras de las columnas del archivo counts coincidan con los nombres de las filas del archivo targets y que en ambos casos estén en el mismo orden; si no fuera así, funciones que se utilizarán en el análisis más adelante producirían errores.

Genero una matriz con la información de las muestras necesaria para el análisis de expresión diferencial; selecciono las columnas *‘Group’* que contiene la información sobre la condición del experimento y *‘Grupo\_analisis’* que contiene información sobre el grupo de análisis al que pertenece la muestra:

condition<-targets[,"Group"]  
analysisGroup<-targets[,"Grupo\_analisis"]  
coldata<-data.frame(Condition = condition, AnalysisGroup = analysisGroup, row.names = targets$Sample\_Name)  
head(coldata)

## Condition AnalysisGroup  
## GTEX.YFC4.2626.SM.5P9FQ ELI 3  
## GTEX.11NV4.0626.SM.5N9BR ELI 3  
## GTEX.PLZ4.1226.SM.2I5FE ELI 3  
## GTEX.R55G.0726.SM.2TC6J ELI 3  
## GTEX.TMMY.0826.SM.33HB9 ELI 3  
## GTEX.11XUK.0226.SM.5EQLW ELI 3

Compruebo que el nombre de las filas de coldata coincide con el de las columnas de counts:

all(rownames(coldata)==colnames(counts))

## [1] TRUE

Por último guardo las tablas generadas (coldata y counts) en dos archivos de tipo ‘csv’; pueden consultarse en la carpeta *‘datos’*:

write.table(coldata, file = "./Datos/coldata.csv", sep = ",", row.names = FALSE)

write.table(counts, file = "./Datos/counts.csv", sep = ",", row.names = FALSE)

**CONSTRUCCIÓN DEL OBJETO DESeqDataSet**

El paquete *‘DESeq2’* proporciona todas las herramientas necesarias para llevar a cabo el análisis de expresión diferencial de una forma más sencilla que con otros paquetes ya que la mayoría de funciones se han unificado; en primer lugar es necesario generar el objeto de la clase *‘DESeq-DataSet’* con los datos de conteos e información sobre las muestras que actúa como input de la función *‘DESeq’* en el análisis de expresión diferencial que se realizará más adelante.

Como disponemos de la matriz de conteos, la función a utilizar para generar el objeto de clase *‘DESeq-DataSet’* es *‘DESeqDataSetFromMatrix()’*; además de la matriz de conteos, la función tiene como argumentos *‘colData’* para introducir información acerca de las muestras y *‘design’* para introducir la fórmula de diseño, en este caso en base a las distintas condiciones del experimento (Not infiltrated tissues (NIT), Small focal infiltrates (SFI), Extensive lymphoid infiltrates (ELI)).

El procedimiento se muestra a continuación:

library("DESeq2")

## Loading required package: S4Vectors

## Warning: package 'S4Vectors' was built under R version 3.6.3

## Loading required package: stats4

## Loading required package: BiocGenerics

## Loading required package: parallel

##   
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:parallel':  
##   
## clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,  
## clusterExport, clusterMap, parApply, parCapply, parLapply,  
## parLapplyLB, parRapply, parSapply, parSapplyLB

## The following objects are masked from 'package:dplyr':  
##   
## combine, intersect, setdiff, union

## The following objects are masked from 'package:stats':  
##   
## IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':  
##   
## anyDuplicated, append, as.data.frame, basename, cbind, colnames,  
## dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,  
## grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,  
## order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,  
## rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,  
## union, unique, unsplit, which, which.max, which.min

##   
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:dplyr':  
##   
## first, rename

## The following object is masked from 'package:base':  
##   
## expand.grid

## Loading required package: IRanges

## Warning: package 'IRanges' was built under R version 3.6.2

##   
## Attaching package: 'IRanges'

## The following objects are masked from 'package:dplyr':  
##   
## collapse, desc, slice

## Loading required package: GenomicRanges

## Loading required package: GenomeInfoDb

## Warning: package 'GenomeInfoDb' was built under R version 3.6.3

## Loading required package: SummarizedExperiment

## Warning: package 'SummarizedExperiment' was built under R version 3.6.2

## Loading required package: Biobase

## Welcome to Bioconductor  
##   
## Vignettes contain introductory material; view with  
## 'browseVignettes()'. To cite Bioconductor, see  
## 'citation("Biobase")', and for packages 'citation("pkgname")'.

## Loading required package: DelayedArray

## Warning: package 'DelayedArray' was built under R version 3.6.3

## Loading required package: matrixStats

##   
## Attaching package: 'matrixStats'

## The following objects are masked from 'package:Biobase':  
##   
## anyMissing, rowMedians

## The following object is masked from 'package:dplyr':  
##   
## count

## Loading required package: BiocParallel

## Warning: package 'BiocParallel' was built under R version 3.6.2

##   
## Attaching package: 'DelayedArray'

## The following objects are masked from 'package:matrixStats':  
##   
## colMaxs, colMins, colRanges, rowMaxs, rowMins, rowRanges

## The following objects are masked from 'package:base':  
##   
## aperm, apply, rowsum

dds<-DESeqDataSetFromMatrix(countData = counts, colData = coldata, design = ~ Condition)  
dds

## class: DESeqDataSet   
## dim: 56202 30   
## metadata(1): version  
## assays(1): counts  
## rownames(56202): ENSG00000223972 ENSG00000227232 ... ENSG00000210195  
## ENSG00000210196  
## rowData names(0):  
## colnames(30): GTEX.YFC4.2626.SM.5P9FQ GTEX.11NV4.0626.SM.5N9BR ...  
## GTEX.R55E.0826.SM.2TC5M GTEX.Q2AH.0726.SM.2I3EA  
## colData names(2): Condition AnalysisGroup

Observando el archivo *‘counts’* generado veo que hay bastantes genes con un conteo muy bajo (0, 1, 2…); aunque la función *‘DESeq’* realiza un filtrado de los genes con bajo número de lecturas voy a realizar un pre-filtrado eliminando los genes con conteos muy bajos para reducir el tamaño del objeto dds y así reducir el tiempo de cómputo.

keep<-rowSums(counts(dds)) >= 5  
dds<-dds[keep,]  
dds

## class: DESeqDataSet   
## dim: 39032 30   
## metadata(1): version  
## assays(1): counts  
## rownames(39032): ENSG00000223972 ENSG00000227232 ... ENSG00000210195  
## ENSG00000210196  
## rowData names(0):  
## colnames(30): GTEX.YFC4.2626.SM.5P9FQ GTEX.11NV4.0626.SM.5N9BR ...  
## GTEX.R55E.0826.SM.2TC5M GTEX.Q2AH.0726.SM.2I3EA  
## colData names(2): Condition AnalysisGroup

Es importante tener en cuenta que por defecto, R elige un nivel de referencia para los factores basado en el orden alfabético de los mismos; establezco como nivel de referencia la condición ‘NIT’ ya que corresponde a las muestras de tejidos no infiltrados:

dds$Condition<-relevel(dds$Condition, ref = "NIT")

**ANÁLISIS EXPLORATORIO Y VISUALIZACIÓN DE LOS DATOS**

Llevar a cabo un primer análisis de los datos realizando diversas transformaciones de los mismos nos permite comprobar si existe algún tipo de relación entre las muestras.

En el caso de análisis de RNAseq, la varianza de los datos de conteo crece con la media de los datos. Existen dos transformaciones posibles de los datos (VST, transformaciones estabilizadoras de la varianza y rlog, logaritmo regularizado) cuyo objetivo es eliminar la dependencia de la varianza de los datos con respecto a la media. En ambos casos la opción de transformación de los datos utilizada es el logaritmo.

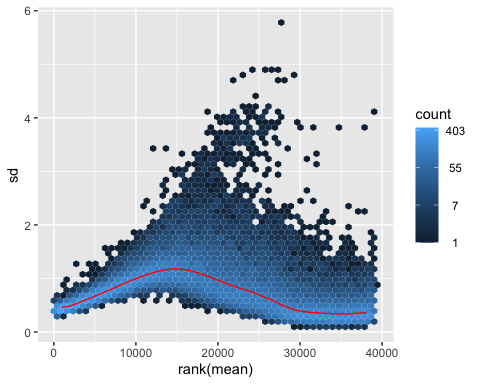
Ambas funciones de transformación devuelven un objeto de la clase *‘DESeqTransform’*; en este caso utilizo la función *‘vst()’* para la tranformación de los datos. Es necesario especificar *‘blind = FALSE’* para que las condiciones del experimento no contribuyan a la varianza.

vsd<-vst(dds, blind = FALSE)  
# Con la función assay se accede a la matriz de los valores normalizados:  
head(assay(vsd),3)

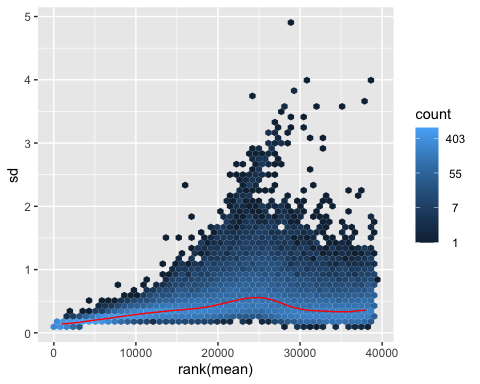
## GTEX.YFC4.2626.SM.5P9FQ GTEX.11NV4.0626.SM.5N9BR  
## ENSG00000223972 4.475773 4.772081  
## ENSG00000227232 9.804040 10.265275  
## ENSG00000243485 4.475773 4.540538  
## GTEX.PLZ4.1226.SM.2I5FE GTEX.R55G.0726.SM.2TC6J  
## ENSG00000223972 4.880599 5.243535  
## ENSG00000227232 8.728661 8.884793  
## ENSG00000243485 4.518145 4.819578  
## GTEX.TMMY.0826.SM.33HB9 GTEX.11XUK.0226.SM.5EQLW  
## ENSG00000223972 4.692415 4.221207  
## ENSG00000227232 9.434413 8.914103  
## ENSG00000243485 4.692415 4.221207  
## GTEX.111VG.0526.SM.5N9BW GTEX.YJ89.0726.SM.5P9F7  
## ENSG00000223972 4.564302 4.753029  
## ENSG00000227232 9.074187 9.791189  
## ENSG00000243485 4.564302 4.488245  
## GTEX.13QJC.0826.SM.5RQKC GTEX.14ABY.0926.SM.5Q5DY  
## ENSG00000223972 4.221207 4.520736  
## ENSG00000227232 9.837381 9.370894  
## ENSG00000243485 4.564997 4.644052  
## GTEX.OXRK.0626.SM.2HMJ5 GTEX.13OVG.0226.SM.5LU93  
## ENSG00000223972 5.280002 4.920785  
## ENSG00000227232 9.765879 9.715577  
## ENSG00000243485 4.783080 4.718282  
## GTEX.PWN1.2626.SM.2I3FH GTEX.11TT1.1126.SM.5P9GV  
## ENSG00000223972 5.103901 4.627557  
## ENSG00000227232 8.859180 8.737171  
## ENSG00000243485 4.221207 4.509013  
## GTEX.ZLFU.0626.SM.4WWBO GTEX.1122O.0226.SM.5N9DA  
## ENSG00000223972 5.146894 4.717665  
## ENSG00000227232 10.087219 9.736852  
## ENSG00000243485 4.794075 4.573117  
## GTEX.ZXG5.0926.SM.5NQ8H GTEX.13FLV.0226.SM.5J2OF  
## ENSG00000223972 4.587229 4.952662  
## ENSG00000227232 9.861242 10.622981  
## ENSG00000243485 4.851844 4.551119  
## GTEX.14ICL.0426.SM.5RQJ3 GTEX.13OVJ.0626.SM.5J2O2  
## ENSG00000223972 4.751053 4.933075  
## ENSG00000227232 9.862922 10.285801  
## ENSG00000243485 4.751053 4.859189  
## GTEX.11DXY.0426.SM.5H12R GTEX.13NYC.2426.SM.5MR3K  
## ENSG00000223972 4.221207 4.221207  
## ENSG00000227232 9.440848 9.520873  
## ENSG00000243485 4.221207 4.221207  
## GTEX.131XG.0226.SM.5IFG1 GTEX.Y5V6.0526.SM.4VBRV  
## ENSG00000223972 4.221207 4.743699  
## ENSG00000227232 8.638220 8.761099  
## ENSG00000243485 4.572900 4.743699  
## GTEX.12584.0826.SM.5FQSK GTEX.ZYVF.1126.SM.5E458  
## ENSG00000223972 4.552656 4.707881  
## ENSG00000227232 10.088857 9.868361  
## ENSG00000243485 4.688927 4.566149  
## GTEX.ZLV1.0126.SM.4WWBZ GTEX.139UW.0126.SM.5KM1B  
## ENSG00000223972 4.686409 4.724906  
## ENSG00000227232 9.473114 9.051059  
## ENSG00000243485 4.686409 4.221207  
## GTEX.R55E.0826.SM.2TC5M GTEX.Q2AH.0726.SM.2I3EA  
## ENSG00000223972 4.845842 4.554146  
## ENSG00000227232 9.383118 9.828047  
## ENSG00000243485 4.583709 5.148925

El efecto que tiene la transformación de los datos sobre la varianza puede verse representando la desviación estándar de los datos transformados frente a la media:

library("vsn")  
ntd<-normTransform(dds, f = log2, pc = 1)  
meanSdPlot(assay(ntd))



meanSdPlot(assay(vsd))



Puede verse en el gráfico anterior que la desviación estándar de los datos se vuelve más constante a lo largo de todo el rango dinámico.

Guardo los gráficos en la carpeta de Imagenes:

png("./Imagenes/meanSDplotNTD.png", width = 1000, height = 800, res = 72, units = "px", pointsize = 12, bg = "white")  
meanSdPlot(assay(ntd))  
dev.off()

## quartz\_off\_screen   
## 2

png("./Imagenes/meanSDplotVSD.png", width = 1000, height = 800, res = 72, units = "px", pointsize = 12, bg = "white")  
meanSdPlot(assay(vsd))  
dev.off()

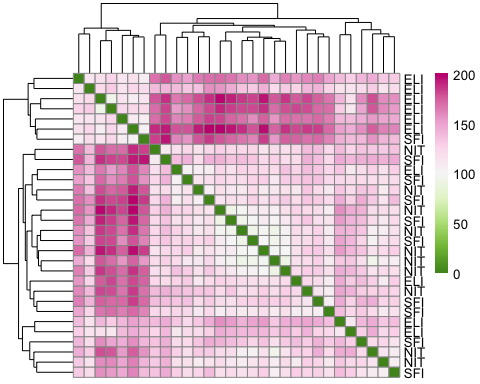
## quartz\_off\_screen   
## 2

Mediante la agrupación de muestras podemos obtener una visión general sobre las similitudes y diferencias entre ellas; se utiliza la función *‘dist()’* para obtener la matriz de distancias:

sampleDists<-dist(t(assay(vsd)))

library(pheatmap)

library("RColorBrewer")  
sampleDistMatrix<-as.matrix(sampleDists)  
rownames(sampleDistMatrix)<-paste(vsd$Condition)  
colnames(sampleDistMatrix)<-NULL  
colors<-colorRampPalette(rev(brewer.pal(9, "PiYG")))(255)  
pheatmap(sampleDistMatrix,  
 clustering\_distance\_rows = sampleDists,  
 clustering\_distance\_cols = sampleDists,  
 col = colors,  
 )



Colores más verdes indican una menor distancia entre muestras y por tanto mayor similitud.

Guardo el gráfico en la carpeta Imagenes:

png("./Imagenes/pheatmapDistMatrix.png", width = 800, height = 800, res = 72, units = "px", pointsize = 12, bg = "white")  
pheatmap(sampleDistMatrix,  
 clustering\_distance\_rows = sampleDists,  
 clustering\_distance\_cols = sampleDists,  
 col = colors,  
 )  
dev.off()

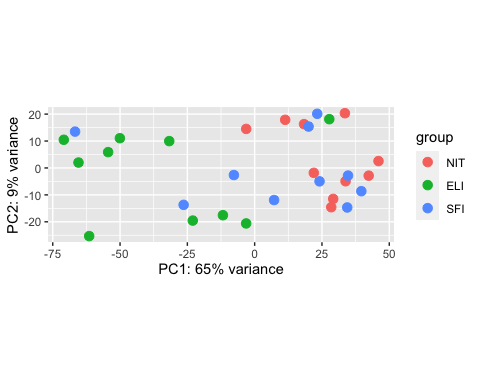
## quartz\_off\_screen   
## 3

El gráfico de componentes principales de las muestras es útil para evaluar el efecto general de las covariables experimentales así como los efectos por lotes. Se representan en este caso los dos primeros componentes:

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.2

DESeq2::plotPCA(vsd, intgroup = c("Condition"))



El primer componente representa el 62% de la variabilidad total de las muestras y como puede observarse en el gráfico anterior, la variabilidad es debida a la condición del tratamiento de infiltración al que se han sometido las muestras.

Guardo el gráfico en la carpeta Imagenes:

png("./Imagenes/PCAplot.png", width = 1000, height = 800, res = 72, units = "px", pointsize = 12, bg = "white")  
DESeq2::plotPCA(vsd, intgroup = c("Condition"))  
dev.off()

## quartz\_off\_screen   
## 2

**ANÁLISIS DE EXPRESIÓN DIFERENCIAL**

La función divide los conteos de cada gen en una muestra por el número total de lecturas de dicha muestra.

El análisis de expresión diferencial se lleva a cabo utilizando la función *‘DESeq()’* que divide los conteos de cada gen en una muestra por el número total de lecturas de dicha muestra. Los resultados pueden obtenerse con la función *‘results()’*.

La función *‘DESeq()’* realiza los siguientes pasos:

* Normalización de los datos mediante la estimación de los tamaños de muestra y los factores de normalización.
* Estimación de la dispersión seleccionando el valor máximo entre las estimaciones de dispersión individuales y la tendencia media de la dispersión.
* Ajuste de los datos a un modelo lineal generalizado (GLM) binomial negativo.
* Comprobación de la expresión diferencial de cada gen mediante la realización del Test paramétrico de Wald.

——**REF 7,22**———-

library("DESeq")

## Loading required package: locfit

## locfit 1.5-9.4 2020-03-24

## Loading required package: lattice

## Warning: package 'lattice' was built under R version 3.6.2

## Welcome to 'DESeq'. For improved performance, usability and  
## functionality, please consider migrating to 'DESeq2'.

##   
## Attaching package: 'DESeq'

## The following objects are masked from 'package:DESeq2':  
##   
## estimateSizeFactorsForMatrix, getVarianceStabilizedData,  
## varianceStabilizingTransformation

dds<-DESeq(dds)

## estimating size factors

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates

## fitting model and testing

## -- replacing outliers and refitting for 302 genes  
## -- DESeq argument 'minReplicatesForReplace' = 7   
## -- original counts are preserved in counts(dds)

## estimating dispersions

## fitting model and testing

La tabla de resultados se obtiene mediante la función *‘results()’*; voy a generar una tabla para cada comparación para que la interpretación de los resultados sea más sencilla:

* Comparación SFI vs NIT:

# SFI vs NIT:  
res\_SFIvsNIT<-results(dds, contrast = c("Condition","SFI","NIT"))

Muestro las primeras líneas de la tabla de resultados obtenida:

head(res\_SFIvsNIT)

## log2 fold change (MLE): Condition SFI vs NIT   
## Wald test p-value: Condition SFI vs NIT   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE  
## <numeric> <numeric> <numeric>  
## ENSG00000223972 2.8345404032717 -1.74684553089361 0.625286699812811  
## ENSG00000227232 747.587653249404 -0.395822251478619 0.240883359149993  
## ENSG00000243485 1.68279424637501 -0.231795081415646 0.660627671403322  
## ENSG00000237613 1.36858179181616 -0.0423639113639465 0.853127539828263  
## ENSG00000268020 0.727212075357395 -0.38691558425816 1.28866030400163  
## ENSG00000240361 1.17226053193609 0.138467136023563 0.836735622911107  
## stat pvalue padj  
## <numeric> <numeric> <numeric>  
## ENSG00000223972 -2.79367133735061 0.0052113413777433 0.267194090331742  
## ENSG00000227232 -1.6432112740181 0.100339228084807 0.860693505693956  
## ENSG00000243485 -0.350870984443114 0.725685140343968 NA  
## ENSG00000237613 -0.0496571841678847 0.960395576423151 NA  
## ENSG00000268020 -0.300246374515213 0.763989234092658 NA  
## ENSG00000240361 0.165484930044951 0.868562310382841 NA

La tabla de resultados contiene la siguiente información:

* baseMean: promedio de las lecturas normalizadas de cada gen; equivale a la intensidad global de la expresión.
* log2FoldChange: tasa de cambio o valor de expresión diferencial en escala logarítmica en base 2. Valores positivos equivalen a genes inducidos y valores negativos equivalen a genes reprimidos.
* lfcSE: error estándar del log2FoldChange.
* stat: valor estadístico principal, en este caso estadístico de Wald.
* pvalue: p-valor; inversamente relacionado con la expresión diferencial; indica si la diferencia observada entre ambos grupos es estadísticamente significativa.
* padj: p-valor ajustado para limitar el número de falsos positivos debido al testeo múltiple.

———**REF sthda.com**———

* Comparación ELI vs NIT:

# ELI vs NIT:  
res\_ELIvsNIT<-results(dds, contrast = c("Condition","ELI","NIT"))

* Comparación ELI vs SFI:

# ELI vs SFI:  
res\_ELIvsSFI<-results(dds, contrast = c("Condition","ELI","SFI"))

Con la función *‘summary()’* obtengo un resumen de la cantidad de genes sobreexpresados e infraexpresados para cada condición; establezco el nivel de significación alfa en 0.05:

summary(res\_SFIvsNIT, alpha = 0.05)

##   
## out of 39028 with nonzero total read count  
## adjusted p-value < 0.05  
## LFC > 0 (up) : 118, 0.3%  
## LFC < 0 (down) : 13, 0.033%  
## outliers [1] : 0, 0%  
## low counts [2] : 12111, 31%  
## (mean count < 3)  
## [1] see 'cooksCutoff' argument of ?results  
## [2] see 'independentFiltering' argument of ?results

summary(res\_ELIvsNIT, alpha = 0.05)

##   
## out of 39028 with nonzero total read count  
## adjusted p-value < 0.05  
## LFC > 0 (up) : 3579, 9.2%  
## LFC < 0 (down) : 1537, 3.9%  
## outliers [1] : 0, 0%  
## low counts [2] : 7571, 19%  
## (mean count < 1)  
## [1] see 'cooksCutoff' argument of ?results  
## [2] see 'independentFiltering' argument of ?results

summary(res\_ELIvsSFI, alpha = 0.05)

##   
## out of 39028 with nonzero total read count  
## adjusted p-value < 0.05  
## LFC > 0 (up) : 1542, 4%  
## LFC < 0 (down) : 584, 1.5%  
## outliers [1] : 0, 0%  
## low counts [2] : 10597, 27%  
## (mean count < 2)  
## [1] see 'cooksCutoff' argument of ?results  
## [2] see 'independentFiltering' argument of ?results

Para ver los resultados de una forma más visual genero un *‘MAplot’* que muestra los cambios en el log2FoldChange en función de la expresión media de los genes (baseMean). Para ello utilizo la función *‘plotMA’*.

Es útil reducir las estimaciones LFC para visualizar y ordenar los genes ya que se elimina el ruido asociado al log2FoldChange de genes con recuentos bajos; para ello utilizo la función *‘lfcShrink()’*. Lo llevo a cabo para las comparaciones de los dos tratamientos de infiltración respecto al control (‘Not Infiltrated Tissues’).

resultsNames(dds)

## [1] "Intercept" "Condition\_ELI\_vs\_NIT" "Condition\_SFI\_vs\_NIT"

library(apeglm)

res\_ELIvsNIT\_LFC<-lfcShrink(dds, coef = "Condition\_ELI\_vs\_NIT", type = "apeglm")

## using 'apeglm' for LFC shrinkage. If used in published research, please cite:  
## Zhu, A., Ibrahim, J.G., Love, M.I. (2018) Heavy-tailed prior distributions for  
## sequence count data: removing the noise and preserving large differences.  
## Bioinformatics. https://doi.org/10.1093/bioinformatics/bty895

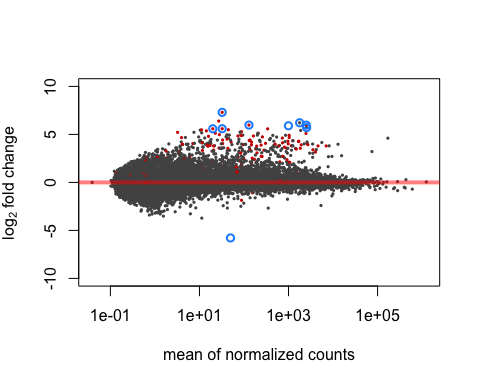
res\_SFIvsNIT\_LFC<-lfcShrink(dds, coef = "Condition\_SFI\_vs\_NIT", type = "apeglm")

## using 'apeglm' for LFC shrinkage. If used in published research, please cite:  
## Zhu, A., Ibrahim, J.G., Love, M.I. (2018) Heavy-tailed prior distributions for  
## sequence count data: removing the noise and preserving large differences.  
## Bioinformatics. https://doi.org/10.1093/bioinformatics/bty895

Obtengo los genes significativos para cada comparación y selecciono los 10 genes con el mayor log2FoldChange en valor absoluto.

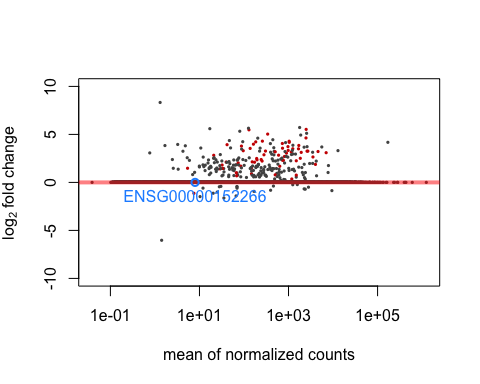
Muestro los gráficos para cada comparación a continuación.

# Condición SFI vs NIT:  
# Selección genes:  
res\_SFIvsNITsig<-subset(res\_SFIvsNIT, res\_SFIvsNIT$padj < 0.05)  
res\_SFIvsNITsigOrd<-res\_SFIvsNITsig[order(abs(res\_SFIvsNITsig$log2FoldChange), decreasing = TRUE),]  
  
# MAplot:  
topGeneSFIvsNIT<-rownames((res\_SFIvsNITsigOrd)[1:10,])  
plotMA(as.data.frame(res\_SFIvsNIT), ylim = c(-10,10))  
with(res\_SFIvsNIT[topGeneSFIvsNIT,], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 1, lwd = 2)  
})

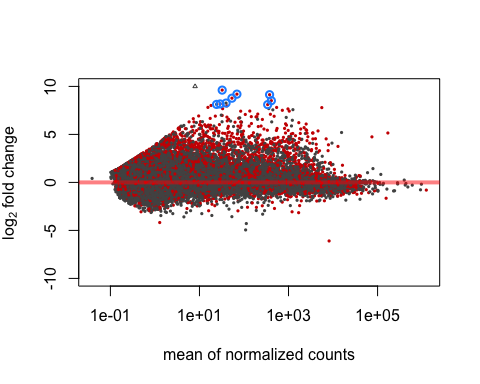


En el siguiente gráfico puede verse señalado el gen con el menor p-valor ajustado en un MAplot con reducción de las estimaciones LFC:

topGeneSFIvsNIT<-rownames(res\_SFIvsNIT\_LFC)[which.min(res\_SFIvsNIT\_LFC$padj)]  
plotMA(as.data.frame(res\_SFIvsNIT\_LFC), ylim = c(-10,10))  
with(res\_SFIvsNIT\_LFC[topGeneSFIvsNIT,], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 1, lwd = 2)  
 text(baseMean, log2FoldChange, topGeneSFIvsNIT, pos = 1, col = "dodgerblue")  
})

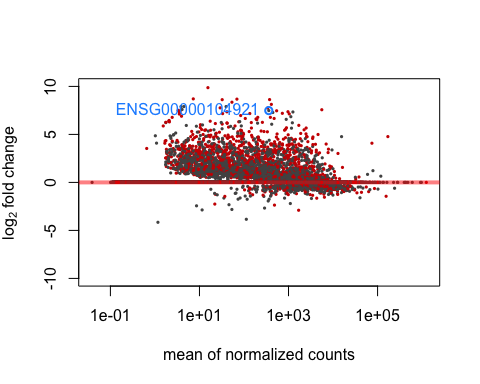


# Condición ELI vs NIT:  
# Selección genes:  
res\_ELIvsNITsig<-subset(res\_ELIvsNIT, res\_ELIvsNIT$padj < 0.05)  
res\_ELIvsNITsigOrd<-res\_ELIvsNITsig[order(abs(res\_ELIvsNITsig$log2FoldChange), decreasing = TRUE),]  
  
# MAplot:  
topGeneELIvsNIT<-rownames((res\_ELIvsNITsigOrd)[1:10,])  
plotMA(as.data.frame(res\_ELIvsNIT), ylim = c(-10,10))  
with(res\_ELIvsNIT[topGeneELIvsNIT,], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 1, lwd = 2)  
})

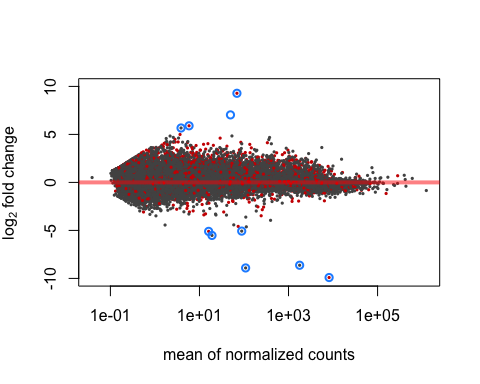


En el siguiente gráfico puede verse señalado el gen con el menor p-valor ajustado en un MAplot con reducción de las estimaciones LFC:

topGeneELIvsNIT<-rownames(res\_ELIvsNIT\_LFC)[which.min(res\_ELIvsNIT\_LFC$padj)]  
plotMA(as.data.frame(res\_ELIvsNIT\_LFC), ylim = c(-10,10))  
with(res\_ELIvsNIT\_LFC[topGeneELIvsNIT,], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 1, lwd = 2)  
 text(baseMean, log2FoldChange, topGeneELIvsNIT, pos = 2, col = "dodgerblue")  
})



# Condición ELI vs SFI:  
# Selección genes:  
res\_ELIvsSFIsig<-subset(res\_ELIvsSFI, res\_ELIvsSFI$padj < 0.05)  
res\_ELIvsSFIsigOrd<-res\_ELIvsSFIsig[order(abs(res\_ELIvsSFIsig$log2FoldChange), decreasing = TRUE),]  
  
# MAplot:  
topGeneELIvsSFI<-rownames((res\_ELIvsSFIsigOrd)[1:10,])  
plotMA(as.data.frame(res\_ELIvsSFI), ylim = c(-10,10))  
with(res\_ELIvsSFI[topGeneELIvsSFI,], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 1, lwd = 2)  
})



En este caso no calculo la reducción de las estimaciones LFC porque no creo que aporte información esencial; podría hacerlo llevando a cabo de nuevo el desarrollo del análisis con un nivel de referencia de la condición diferente.

Guardo los MAplots obtenidos en la carpeta Imagenes:

png("./Imagenes/MAplot\_SFIvsNIT.png", width = 800, height = 800, res = 72, units = "px", pointsize = 12, bg = "white")  
plotMA(as.data.frame(res\_SFIvsNIT), ylim = c(-10,10))  
with(res\_SFIvsNIT[topGeneSFIvsNIT,], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 1, lwd = 2)  
})  
dev.off()

## quartz\_off\_screen   
## 2

png("./Imagenes/MAplot\_SFIvsNIT\_LFC.png", width = 800, height = 800, res = 72, units = "px", pointsize = 12, bg = "white")  
plotMA(as.data.frame(res\_SFIvsNIT\_LFC), ylim = c(-10,10))  
with(res\_SFIvsNIT\_LFC[topGeneSFIvsNIT,], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 1, lwd = 2)  
 text(baseMean, log2FoldChange, topGeneSFIvsNIT, pos = 1, col = "dodgerblue")  
})  
dev.off()

## quartz\_off\_screen   
## 2

png("./Imagenes/MAplot\_ELIvsNIT.png", width = 800, height = 800, res = 72, units = "px", pointsize = 12, bg = "white")  
plotMA(as.data.frame(res\_ELIvsNIT), ylim = c(-10,10))  
with(res\_ELIvsNIT[topGeneELIvsNIT,], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 1, lwd = 2)  
})  
dev.off()

## quartz\_off\_screen   
## 2

png("./Imagenes/MAplot\_ELIvsNIT\_LFC.png", width = 800, height = 800, res = 72, units = "px", pointsize = 12, bg = "white")  
plotMA(as.data.frame(res\_ELIvsNIT\_LFC), ylim = c(-10,10))  
with(res\_ELIvsNIT\_LFC[topGeneELIvsNIT,], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 1, lwd = 2)  
 text(baseMean, log2FoldChange, topGeneELIvsNIT, pos = 2, col = "dodgerblue")  
})  
dev.off()

## quartz\_off\_screen   
## 2

png("./Imagenes/MAplot\_ELIvsSFI.png", width = 800, height = 800, res = 72, units = "px", pointsize = 12, bg = "white")  
plotMA(as.data.frame(res\_ELIvsSFI), ylim = c(-10,10))  
with(res\_ELIvsSFI[topGeneELIvsSFI,], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 1, lwd = 2)  
})  
dev.off()

## quartz\_off\_screen   
## 2

**ANOTACIÓN DE LOS RESULTADOS**

Para realizar la anotación de los genes y poder asociar los identificadores de ENSEMBLE con información sobre los mismos utilizo el paquete de anotaciones *‘EnsDb.Hsapiens.v86’* (la idea original era utilizar el paquete de anotaciones *‘org.Hs.eg.db’* pero al obtener un número muy bajo de mapeados se decide probar con otro paquete y el resultado obtenido es algo mejor).

library("EnsDb.Hsapiens.v86")

## Loading required package: ensembldb

## Loading required package: GenomicFeatures

## Warning: package 'GenomicFeatures' was built under R version 3.6.2

## Loading required package: AnnotationDbi

##   
## Attaching package: 'AnnotationDbi'

## The following object is masked from 'package:dplyr':  
##   
## select

## Loading required package: AnnotationFilter

##   
## Attaching package: 'ensembldb'

## The following object is masked from 'package:dplyr':  
##   
## filter

## The following object is masked from 'package:stats':  
##   
## filter

edb<-EnsDb.Hsapiens.v86

Los tipos de claves disponibles y por tanto información a las que se puede acceder son los siguientes:

keytypes(edb)

## [1] "ENTREZID" "EXONID" "GENEBIOTYPE"   
## [4] "GENEID" "GENENAME" "PROTDOMID"   
## [7] "PROTEINDOMAINID" "PROTEINDOMAINSOURCE" "PROTEINID"   
## [10] "SEQNAME" "SEQSTRAND" "SYMBOL"   
## [13] "TXBIOTYPE" "TXID" "TXNAME"   
## [16] "UNIPROTID"

Utilizo la función *‘mapIDs()’* para convertir los IDs y obtener el símbolo del gen, el identificador de Entrez y el nombre del gen para las tres tablas finales de resultados:

# Comparación ELI vs NIT:  
# Símbolo del gen:  
res\_ELIvsNIT$genSYMBOL<-mapIds(edb, keys=row.names(res\_ELIvsNIT), column="SYMBOL", keytype="GENEID", multiVals="first")

## Warning: Unable to map 1678 of 39032 requested IDs.

# Identificador de Entrez:  
res\_ELIvsNIT$ENTREZID<-mapIds(edb, keys=row.names(res\_ELIvsNIT), column="ENTREZID", keytype="GENEID", multiVals="first")

## Warning: Unable to map 18089 of 39032 requested IDs.

# Nombre del gen:  
res\_ELIvsNIT$GENENAME<-mapIds(edb, keys=row.names(res\_ELIvsNIT), column="GENENAME", keytype="GENEID", multiVals="first")

## Warning: Unable to map 1678 of 39032 requested IDs.

# Comparación SFI vs NIT:  
# Símbolo del gen:  
res\_SFIvsNIT$genSYMBOL<-mapIds(edb, keys=row.names(res\_SFIvsNIT), column="SYMBOL", keytype="GENEID", multiVals="first")

## Warning: Unable to map 1678 of 39032 requested IDs.

# Identificador de Entrez:  
res\_SFIvsNIT$ENTREZID<-mapIds(edb, keys=row.names(res\_SFIvsNIT), column="ENTREZID", keytype="GENEID", multiVals="first")

## Warning: Unable to map 18089 of 39032 requested IDs.

# Nombre del gen:  
res\_SFIvsNIT$GENENAME<-mapIds(edb, keys=row.names(res\_SFIvsNIT), column="GENENAME", keytype="GENEID", multiVals="first")

## Warning: Unable to map 1678 of 39032 requested IDs.

# Comparación ELI vs SFI:  
# Símbolo del gen:  
res\_ELIvsSFI$genSYMBOL<-mapIds(edb, keys=row.names(res\_ELIvsSFI), column="SYMBOL", keytype="GENEID", multiVals="first")

## Warning: Unable to map 1678 of 39032 requested IDs.

# Identificador de Entrez:  
res\_ELIvsSFI$ENTREZID<-mapIds(edb, keys=row.names(res\_ELIvsSFI), column="ENTREZID", keytype="GENEID", multiVals="first")

## Warning: Unable to map 18089 of 39032 requested IDs.

# Nombre del gen:  
res\_ELIvsSFI$GENENAME<-mapIds(edb, keys=row.names(res\_ELIvsSFI), column="GENENAME", keytype="GENEID", multiVals="first")

## Warning: Unable to map 1678 of 39032 requested IDs.

Guardo las tablas generadas en formato csv en la carpeta de Resultados:

# Comparación SFI vs NIT:  
write.table(res\_SFIvsNIT, file = "./Resultados/finalSFIvsNIT.csv", sep = ",", row.names = FALSE)  
  
# Comparación ELI vs NIT:  
write.table(res\_ELIvsNIT, file = "./Resultados/finalELIvsNIT.csv", sep = ",", row.names = FALSE)  
  
# Comparación ELI vs SFI:  
write.table(res\_ELIvsSFI, file = "./Resultados/finalELIvsSFI.csv", sep = ",", row.names = FALSE)

Asumiendo que una fracción del 10% de falsos positivos es aceptable, aquellos genes con un p-valor ajustado menor de 0.1 pueden considerarse estadísticamente significativos:

# Comparación SFI vs NIT:  
sum(res\_SFIvsNIT$padj < 0.1, na.rm = TRUE)

## [1] 228

Dentro de los genes estadísticamente significativos los genes sobre e infraexpresados son:

# Comparación SFI vs NIT  
res\_SFIvsNITsig<-subset(res\_SFIvsNIT, res\_SFIvsNIT$padj < 0.1)  
# Infraexpresados:  
head(res\_SFIvsNITsig[order(res\_SFIvsNITsig$log2FoldChange),])

## log2 fold change (MLE): Condition SFI vs NIT   
## Wald test p-value: Condition SFI vs NIT   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE  
## <numeric> <numeric> <numeric>  
## ENSG00000264940 49.7027304902574 -5.78584024705562 1.46756666196535  
## ENSG00000222414 28.3823632565977 -3.75765283726538 0.714203958219986  
## ENSG00000262202 11.2856786031432 -3.71908509005658 0.919740650763231  
## ENSG00000128965 349.963810539654 -2.67841962141558 0.658269405726361  
## ENSG00000006128 299.577227592198 -2.45706929834528 0.62772986725271  
## ENSG00000244693 41.9180637147604 -2.12073863067949 0.554645024853543  
## stat pvalue padj  
## <numeric> <numeric> <numeric>  
## ENSG00000264940 -3.94247184608793 8.06461195338951e-05 0.0277114906700149  
## ENSG00000222414 -5.2613161744869 1.43027839843342e-07 0.000923008153533636  
## ENSG00000262202 -4.0436236964957 5.26313683871688e-05 0.0204650109597714  
## ENSG00000128965 -4.06888061045478 4.72395440674887e-05 0.019565165628321  
## ENSG00000006128 -3.9142144201275 9.0699012129119e-05 0.0283405881447143  
## ENSG00000244693 -3.82359623840399 0.000131519222880951 0.0350557326651295  
## genSYMBOL ENTREZID GENENAME  
## <character> <integer> <character>  
## ENSG00000264940 SNORD3C 780853 SNORD3C  
## ENSG00000222414 RNU2-59P NA RNU2-59P  
## ENSG00000262202 RP11-160E2.6 NA RP11-160E2.6  
## ENSG00000128965 CHAC1 79094 CHAC1  
## ENSG00000006128 TAC1 6863 TAC1  
## ENSG00000244693 CTAGE8 100142659 CTAGE8

# Comparación SFI vs NIT  
# Sobreexpresados:  
head(res\_SFIvsNITsig[order(-res\_SFIvsNITsig$log2FoldChange),])

## log2 fold change (MLE): Condition SFI vs NIT   
## Wald test p-value: Condition SFI vs NIT   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE  
## <numeric> <numeric> <numeric>  
## ENSG00000152266 7.97205512668314 21.2783685319236 3.59355669853249  
## ENSG00000163518 32.4049051469384 7.29669381936176 1.6251889226098  
## ENSG00000100604 1777.09791964318 6.21166110991056 1.20195829088776  
## ENSG00000211946 129.280705989048 5.97059356954862 1.15589527702103  
## ENSG00000242371 2488.88887091322 5.96729539470855 1.0952770731194  
## ENSG00000211937 995.74913036063 5.90097159830283 1.1222133309746  
## stat pvalue padj  
## <numeric> <numeric> <numeric>  
## ENSG00000152266 5.92125582451869 3.19492406209143e-09 8.60105506755633e-05  
## ENSG00000163518 4.48975114083623 7.13064316467868e-06 0.00685585873701124  
## ENSG00000100604 5.16795063273172 2.36674801101311e-07 0.000923008153533636  
## ENSG00000211946 5.16534126251991 2.40000634253388e-07 0.000923008153533636  
## ENSG00000242371 5.44820624950491 5.08803374344633e-08 0.000684874782036594  
## ENSG00000211937 5.25833318445616 1.45366966008643e-07 0.000923008153533636  
## genSYMBOL ENTREZID GENENAME  
## <character> <integer> <character>  
## ENSG00000152266 PTH 5741 PTH  
## ENSG00000163518 FCRL4 83417 FCRL4  
## ENSG00000100604 CHGA 1113 CHGA  
## ENSG00000211946 IGHV3-20 NA IGHV3-20  
## ENSG00000242371 IGKV1-39 NA IGKV1-39  
## ENSG00000211937 IGHV2-5 NA IGHV2-5

# Comparación ELI vs NIT:  
sum(res\_ELIvsNIT$padj < 0.1, na.rm = TRUE)

## [1] 6800

Dentro de los genes estadísticamente significativos los genes sobre e infraexpresados son:

# Comparación ELI vs NIT  
res\_ELIvsNITsig<-subset(res\_ELIvsNIT, res\_ELIvsNIT$padj < 0.1)  
# Infraexpresados:  
head(res\_ELIvsNITsig[order(res\_ELIvsNITsig$log2FoldChange),])

## log2 fold change (MLE): Condition ELI vs NIT   
## Wald test p-value: Condition ELI vs NIT   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE  
## <numeric> <numeric> <numeric>  
## ENSG00000110680 8152.28547152127 -6.0998957533415 1.4479273594484  
## ENSG00000079689 112.825938997516 -4.2873448732989 0.917679794105111  
## ENSG00000266670 1.28148963314164 -4.1810437894763 1.17337832792464  
## ENSG00000223001 7.57603912527461 -4.0212570052417 1.17345841946871  
## ENSG00000006128 299.577227592198 -3.63042430745604 0.629548151376309  
## ENSG00000250620 4.47574109978457 -3.59263203953142 0.995198992872343  
## stat pvalue padj  
## <numeric> <numeric> <numeric>  
## ENSG00000110680 -4.21284653096499 2.52172409581335e-05 0.000565473711891546  
## ENSG00000079689 -4.67193993028883 2.98368228153598e-06 9.64744380877734e-05  
## ENSG00000266670 -3.56325295088015 0.000366287412431207 0.0048035716058767  
## ENSG00000223001 -3.42684234782035 0.000610643482225171 0.00713915072251434  
## ENSG00000006128 -5.76671426882799 8.0831922870992e-09 7.59120335953516e-07  
## ENSG00000250620 -3.60996350002562 0.000306240095165447 0.00417083101038967  
## genSYMBOL ENTREZID GENENAME  
## <character> <integer> <character>  
## ENSG00000110680 CALCA 796 CALCA  
## ENSG00000079689 SCGN 10590 SCGN  
## ENSG00000266670 NA NA NA  
## ENSG00000223001 RNU2-61P NA RNU2-61P  
## ENSG00000006128 TAC1 6863 TAC1  
## ENSG00000250620 RP11-91J3.3 NA RP11-91J3.3

# Comparación ELI vs NIT  
# Sobreexpresados:  
head(res\_ELIvsNITsig[order(-res\_ELIvsNITsig$log2FoldChange),])

## log2 fold change (MLE): Condition ELI vs NIT   
## Wald test p-value: Condition ELI vs NIT   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE  
## <numeric> <numeric> <numeric>  
## ENSG00000152266 7.97205512668314 18.0648796436458 3.60559772163245  
## ENSG00000163518 32.4049051469384 9.61405105474491 1.62208548045644  
## ENSG00000162897 69.3806789720014 9.19226655479561 1.60826554955337  
## ENSG00000181617 375.600649445937 9.13522332836419 1.46137350321582  
## ENSG00000170054 53.6349169768901 8.7730124107132 1.62103690025448  
## ENSG00000100721 411.773745845798 8.51386643983817 1.24516492040985  
## stat pvalue padj  
## <numeric> <numeric> <numeric>  
## ENSG00000152266 5.01023160050889 5.43645703069055e-07 2.36564833530505e-05  
## ENSG00000163518 5.92696942952699 3.08576393869704e-09 3.43043177651405e-07  
## ENSG00000162897 5.71563978184347 1.0929203742749e-08 9.57781835517068e-07  
## ENSG00000181617 6.25112150197176 4.07515679729473e-10 7.20272516852189e-08  
## ENSG00000170054 5.41197576029018 6.23331284989696e-08 4.01857081087312e-06  
## ENSG00000100721 6.83754119657963 8.05639253553146e-12 2.88025188136767e-09  
## genSYMBOL ENTREZID GENENAME  
## <character> <integer> <character>  
## ENSG00000152266 PTH 5741 PTH  
## ENSG00000163518 FCRL4 83417 FCRL4  
## ENSG00000162897 FCAMR 83953 FCAMR  
## ENSG00000181617 FDCSP 260436 FDCSP  
## ENSG00000170054 SERPINA9 327657 SERPINA9  
## ENSG00000100721 TCL1A 8115 TCL1A

# Comparación ELI vs SFI:  
sum(res\_ELIvsSFI$padj < 0.1, na.rm = TRUE)

## [1] 3364

Dentro de los genes estadísticamente significativos los genes sobre e infraexpresados son:

# Comparación ELI vs SFI  
res\_ELIvsSFIsig<-subset(res\_ELIvsSFI, res\_ELIvsSFI$padj < 0.1)  
# Infraexpresados:  
head(res\_ELIvsSFIsig[order(res\_ELIvsSFIsig$log2FoldChange),])

## log2 fold change (MLE): Condition ELI vs SFI   
## Wald test p-value: Condition ELI vs SFI   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE  
## <numeric> <numeric> <numeric>  
## ENSG00000110680 8152.28547152127 -9.90925933696272 1.44788603725929  
## ENSG00000134443 108.836307389791 -8.91160519811308 2.73708307047921  
## ENSG00000100604 1777.09791964318 -8.62945921781587 1.20764926927797  
## ENSG00000157005 19.1186272334118 -5.52456730283538 1.17105536708671  
## ENSG00000132437 16.0775871547864 -5.10119886655333 1.01179966506251  
## ENSG00000105388 88.8879223419104 -5.07272889103988 1.09864315593058  
## stat pvalue padj  
## <numeric> <numeric> <numeric>  
## ENSG00000110680 -6.84394978745703 7.70389342406504e-12 1.09530104756645e-07  
## ENSG00000134443 -3.25587677415755 0.00113042801021034 0.0247640373423196  
## ENSG00000100604 -7.14566674062183 8.95601031481142e-13 2.54664153301663e-08  
## ENSG00000157005 -4.71759701386206 2.38646610505473e-06 0.000665285918600306  
## ENSG00000132437 -5.0417083961361 4.61394015463382e-07 0.000247542242069835  
## ENSG00000105388 -4.61726709319291 3.88826688232387e-06 0.000898885112186011  
## genSYMBOL ENTREZID GENENAME  
## <character> <integer> <character>  
## ENSG00000110680 CALCA 796 CALCA  
## ENSG00000134443 GRP 2922 GRP  
## ENSG00000100604 CHGA 1113 CHGA  
## ENSG00000157005 SST 6750 SST  
## ENSG00000132437 DDC 1644 DDC  
## ENSG00000105388 CEACAM5 1048 CEACAM5

# Comparación ELI vs SFI  
# Sobreexpresados:  
head(res\_ELIvsSFIsig[order(-res\_ELIvsSFIsig$log2FoldChange),])

## log2 fold change (MLE): Condition ELI vs SFI   
## Wald test p-value: Condition ELI vs SFI   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE  
## <numeric> <numeric> <numeric>  
## ENSG00000162897 69.3806789720014 9.28317295006343 1.60827037711234  
## ENSG00000264940 49.7027304902574 7.032077176519 1.46657195907518  
## ENSG00000226423 5.83942590376567 5.89561317739225 1.0777566726781  
## ENSG00000229613 3.85875153304181 5.66788918877105 1.16996041117353  
## ENSG00000231682 3.66226223497985 4.99179335607858 1.53630092629605  
## ENSG00000255582 2.0038740004151 4.91767625562539 1.35501776816411  
## stat pvalue padj  
## <numeric> <numeric> <numeric>  
## ENSG00000162897 5.77214694878073 7.82678119860506e-09 2.65809950392936e-05  
## ENSG00000264940 4.79490769818989 1.6274974978255e-06 0.000511132010447052  
## ENSG00000226423 5.47026367532696 4.49366533682284e-08 8.0356848210944e-05  
## ENSG00000229613 4.8445136558816 1.26922188881938e-06 0.000434823185645532  
## ENSG00000231682 3.2492288917078 0.00115718324039963 0.0251371317347314  
## ENSG00000255582 3.62923377919116 0.00028426372107411 0.0107344474219685  
## genSYMBOL ENTREZID GENENAME  
## <character> <integer> <character>  
## ENSG00000162897 FCAMR 83953 FCAMR  
## ENSG00000264940 SNORD3C 780853 SNORD3C  
## ENSG00000226423 AC093642.4 NA AC093642.4  
## ENSG00000229613 LINC01501 NA LINC01501  
## ENSG00000231682 AC097713.4 NA AC097713.4  
## ENSG00000255582 OR10G2 26534 OR10G2

Genero las tablas con los genes sobre e infraexpresados para cada comparación; la guardo en la carpeta de Resultados:

# Comparación SFI vs NIT:  
SFIvsNIT\_ord<-res\_SFIvsNITsig[order(res\_SFIvsNITsig$log2FoldChange),]  
SFIvsNIT\_down<-SFIvsNIT\_ord[SFIvsNIT\_ord$log2FoldChange < 0,]  
  
SFIvsNIT\_ord<-res\_SFIvsNITsig[order(-res\_SFIvsNITsig$log2FoldChange),]  
SFIvsNIT\_up<-SFIvsNIT\_ord[SFIvsNIT\_ord$log2FoldChange > 0,]  
  
write.table(SFIvsNIT\_down, file = "./Resultados/SFIvsNIT\_down.csv", sep = ",", row.names = FALSE)  
write.table(SFIvsNIT\_up, file = "./Resultados/SFIvsNIT\_up.csv", sep = ",", row.names = FALSE)

# Comparación ELI vs NIT:  
ELIvsNIT\_ord<-res\_ELIvsNITsig[order(res\_ELIvsNITsig$log2FoldChange),]  
ELIvsNIT\_down<-ELIvsNIT\_ord[ELIvsNIT\_ord$log2FoldChange < 0,]  
  
ELIvsNIT\_ord<-res\_ELIvsNITsig[order(-res\_ELIvsNITsig$log2FoldChange),]  
ELIvsNIT\_up<-ELIvsNIT\_ord[ELIvsNIT\_ord$log2FoldChange > 0,]  
  
write.table(ELIvsNIT\_down, file = "./Resultados/ELIvsNIT\_down.csv", sep = ",", row.names = FALSE)  
write.table(ELIvsNIT\_up, file = "./Resultados/ELIvsNIT\_up.csv", sep = ",", row.names = FALSE)

# Comparación ELI vs SFI:  
ELIvsSFI\_ord<-res\_ELIvsSFIsig[order(res\_ELIvsSFIsig$log2FoldChange),]  
ELIvsSFI\_down<-ELIvsSFI\_ord[ELIvsSFI\_ord$log2FoldChange < 0,]  
  
ELIvsSFI\_ord<-res\_ELIvsSFIsig[order(-res\_ELIvsSFIsig$log2FoldChange),]  
ELIvsSFI\_up<-ELIvsSFI\_ord[ELIvsSFI\_ord$log2FoldChange > 0,]  
  
write.table(ELIvsSFI\_down, file = "./Resultados/ELIvsSFI\_down.csv", sep = ",", row.names = FALSE)  
write.table(ELIvsSFI\_up, file = "./Resultados/ELIvsSFI\_up.csv", sep = ",", row.names = FALSE)

**MAPA DE CALOR**

Genero un Heatmap con agrupamiento de los genes más variables; selecciono los 20 genes con la mayor varianza entre las muestras. Utilizo los recuentos transformados de VST.

library(genefilter)

##   
## Attaching package: 'genefilter'

## The following objects are masked from 'package:matrixStats':  
##   
## rowSds, rowVars

library(RColorBrewer)  
library(gplots)

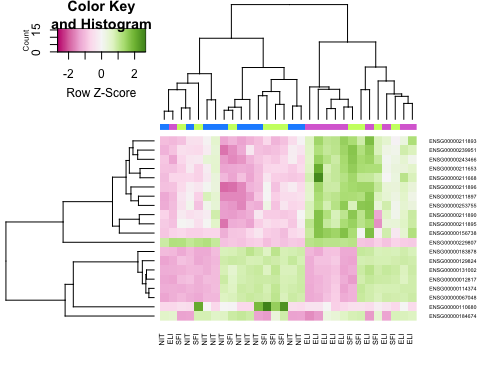
##   
## Attaching package: 'gplots'

## The following object is masked from 'package:IRanges':  
##   
## space

## The following object is masked from 'package:S4Vectors':  
##   
## space

## The following object is masked from 'package:stats':  
##   
## lowess

topVarGenes<-head(order(rowVars(assay(vsd)), decreasing = TRUE), 20)  
colors<-colorRampPalette(rep(brewer.pal(9,"PiYG")))(255)  
sidecols<-c("dodgerblue", "orchid", "darkolivegreen1")[vsd$Condition]  
mat<-assay(vsd)[topVarGenes,]  
mat<-mat - rowMeans(mat)  
colnames(mat)<-vsd$Condition  
heatmap.2(mat,  
 Rowv = TRUE,  
 Colv = TRUE,  
 scale = "row",  
 col = colors,  
 cexRow = 0.5,  
 cexCol = 0.7,  
 key = TRUE,  
 keysize = 2,  
 density.info = "histogram",  
 ColSideColors = sidecols,  
 tracecol = NULL)



Guardo el Heatmap en la carpeta de Imagenes:

png("./Imagenes/Heatmap.png", width = 1200, height = 800, res = 100, units = "px", pointsize = 12, bg = "white")  
heatmap.2(mat,  
 Rowv = TRUE,  
 Colv = TRUE,  
 scale = "row",  
 col = colors,  
 cexRow = 0.5,  
 cexCol = 0.7,  
 key = TRUE,  
 keysize = 2,  
 density.info = "histogram",  
 ColSideColors = sidecols,  
 tracecol = NULL)  
dev.off()

## quartz\_off\_screen   
## 2

**COMPARACIONES MÚLTIPLES**

Como se han realizado tres comparaciones distintas, es interesante saber qué genes se han seleccionado como diferencialmente expresados en cada comparación y ver así si algún gen se ha seleccionado en más de una de ellas.

Para ello voy a utilizar la función *‘intersect()’* del paquete *‘dplyr’*.

* Genes en común entre las comparaciones “SFI vs NIT” y “ELI vs NIT”:

names\_SFIvsNIT\_ELIvsNIT<-intersect(rownames(res\_SFIvsNITsig), rownames(res\_ELIvsNITsig))  
length(names\_SFIvsNIT\_ELIvsNIT)

## [1] 207

Entre las comparaciones “SFI vs NIT” y “ELI vs NIT” se comparten 60 genes significativos. Algunos de estos genes son:

DEG\_SFIvsNIT\_ELIvsNIT<-res\_SFIvsNITsig[names\_SFIvsNIT\_ELIvsNIT,]  
head(DEG\_SFIvsNIT\_ELIvsNIT)

## log2 fold change (MLE): Condition SFI vs NIT   
## Wald test p-value: Condition SFI vs NIT   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE  
## <numeric> <numeric> <numeric>  
## ENSG00000186891 172.303629411351 2.39964265940953 0.560099816873192  
## ENSG00000236782 7.50957097047946 -1.33975458638659 0.341523009900575  
## ENSG00000162631 29.0258409261583 1.81208556140429 0.533244531637789  
## ENSG00000116748 39.7965729342509 2.57301642601205 0.690822665558684  
## ENSG00000231360 20.9085432118606 3.36432014455637 0.869012684692209  
## ENSG00000143297 1701.86535506089 4.28761793336113 0.915880834712703  
## stat pvalue padj  
## <numeric> <numeric> <numeric>  
## ENSG00000186891 4.28431252273166 1.83305100128021e-05 0.0109186980413467  
## ENSG00000236782 -3.92288234627771 8.74958845121381e-05 0.0277114906700149  
## ENSG00000162631 3.39822624310597 0.000678243051282536 0.0882076385680055  
## ENSG00000116748 3.72456862562724 0.000195649486039033 0.043467177571388  
## ENSG00000231360 3.87142812046289 0.000108199564778349 0.0306614787726098  
## ENSG00000143297 4.68141462388618 2.84902047394407e-06 0.00382432690588632  
## genSYMBOL ENTREZID GENENAME  
## <character> <integer> <character>  
## ENSG00000186891 TNFRSF18 8784 TNFRSF18  
## ENSG00000236782 RP11-96L14.7 NA RP11-96L14.7  
## ENSG00000162631 NTNG1 22854 NTNG1  
## ENSG00000116748 AMPD1 270 AMPD1  
## ENSG00000231360 NA NA NA  
## ENSG00000143297 FCRL5 83416 FCRL5

* Genes en común entre las comparaciones “SFI vs NIT” y “ELI vs SFI”:

names\_SFIvsNIT\_ELIvsSFI<-intersect(rownames(res\_SFIvsNITsig), rownames(res\_ELIvsSFIsig))  
length(names\_SFIvsNIT\_ELIvsSFI)

## [1] 61

Entre las comparaciones “SFI vs NIT” y “ELI vs NIT” se comparten 48 genes significativos. Algunos de estos genes son:

DEG\_SFIvsNIT\_ELIvsSFI<-res\_SFIvsNITsig[names\_SFIvsNIT\_ELIvsSFI,]  
head(DEG\_SFIvsNIT\_ELIvsSFI)

## log2 fold change (MLE): Condition SFI vs NIT   
## Wald test p-value: Condition SFI vs NIT   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE  
## <numeric> <numeric> <numeric>  
## ENSG00000162433 502.322238034883 0.963715395548693 0.266439019714623  
## ENSG00000160856 899.044150322298 3.56316736364676 0.880979319238278  
## ENSG00000132704 726.714500543193 4.62004016253617 1.06065041946934  
## ENSG00000163534 971.658177885823 4.29478189701522 1.0198780123033  
## ENSG00000163568 281.007818880693 3.35663174680662 0.826795847469006  
## ENSG00000127074 134.973795352845 3.47096633003384 0.957649796561774  
## stat pvalue padj  
## <numeric> <numeric> <numeric>  
## ENSG00000162433 3.61702049715131 0.000298013712740662 0.057718181012168  
## ENSG00000160856 4.04455279010134 5.24230955135923e-05 0.0204650109597714  
## ENSG00000132704 4.35585568791613 1.32548086476721e-05 0.00914955650266619  
## ENSG00000163534 4.21107411396766 2.54159339080457e-05 0.0124404064861545  
## ENSG00000163568 4.05980721490313 4.91132495377004e-05 0.0197339968776781  
## ENSG00000127074 3.62446307877428 0.000289562380344954 0.056900064534792  
## genSYMBOL ENTREZID GENENAME  
## <character> <integer> <character>  
## ENSG00000162433 AK4 205 AK4  
## ENSG00000160856 FCRL3 115352 FCRL3  
## ENSG00000132704 FCRL2 79368 FCRL2  
## ENSG00000163534 FCRL1 115350 FCRL1  
## ENSG00000163568 AIM2 9447 AIM2  
## ENSG00000127074 RGS13 6003 RGS13

* Genes en común entre las comparaciones “ELI vs NIT” y “ELI vs SFI”:

names\_ELIvsNIT\_ELIvsSFI<-intersect(rownames(res\_ELIvsNITsig), rownames(res\_ELIvsSFIsig))  
length(names\_ELIvsNIT\_ELIvsSFI)

## [1] 2842

Entre las comparaciones “ELI vs NIT” y “ELI vs SFI” se comparten 3130 genes significativos. Algunos de estos genes son:

DEG\_ELIvsNIT\_ELIvsSFI<-res\_ELIvsNITsig[names\_ELIvsNIT\_ELIvsSFI,]  
head(DEG\_ELIvsNIT\_ELIvsSFI)

## log2 fold change (MLE): Condition ELI vs NIT   
## Wald test p-value: Condition ELI vs NIT   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE  
## <numeric> <numeric> <numeric>  
## ENSG00000234711 2.76093666578235 2.23728605105609 0.898335032124382  
## ENSG00000187642 52.3398275837924 1.54764144121386 0.323132647651639  
## ENSG00000169962 196.292195650886 1.64664371701704 0.365966263529208  
## ENSG00000107404 5401.78280567775 -0.345832008281088 0.108657058290417  
## ENSG00000235098 409.905283832095 1.29886646161389 0.32735553400239  
## ENSG00000157911 670.43157648879 -0.323267622753572 0.138944445833449  
## stat pvalue padj  
## <numeric> <numeric> <numeric>  
## ENSG00000234711 2.49048069044503 0.0127570427622809 0.0690434065618649  
## ENSG00000187642 4.78949265096337 1.67203521410345e-06 5.95902247465583e-05  
## ENSG00000169962 4.49944129040087 6.81322925895342e-06 0.000193457586386222  
## ENSG00000107404 -3.18278456754049 0.00145866100645286 0.0139021308464143  
## ENSG00000235098 3.96775470917931 7.25529520470622e-05 0.00131865304699747  
## ENSG00000157911 -2.326596222069 0.0199867658095659 0.0946138488014975  
## genSYMBOL ENTREZID GENENAME  
## <character> <integer> <character>  
## ENSG00000234711 TUBB8P11 NA TUBB8P11  
## ENSG00000187642 PERM1 84808 PERM1  
## ENSG00000169962 TAS1R3 83756 TAS1R3  
## ENSG00000107404 DVL1 102466740 DVL1  
## ENSG00000235098 ANKRD65 441869 ANKRD65  
## ENSG00000157911 PEX10 5192 PEX10

* Genes en común entre las 3 comparaciones::

names\_SFIvsNIT\_ELIvsNIT\_ELIvsSFI<-intersect(intersect(rownames(res\_SFIvsNITsig), rownames(res\_ELIvsNITsig)), rownames(res\_ELIvsSFIsig))  
length(names\_SFIvsNIT\_ELIvsNIT\_ELIvsSFI)

## [1] 49

Entre las 3 comparaciones se comparten 3130 genes significativos. Algunos de estos genes son:

DEG\_SFIvsNIT\_ELIvsNIT\_ELIvsSFI<-res\_ELIvsNITsig[names\_SFIvsNIT\_ELIvsNIT\_ELIvsSFI,]  
head(DEG\_SFIvsNIT\_ELIvsNIT\_ELIvsSFI)

## log2 fold change (MLE): Condition ELI vs NIT   
## Wald test p-value: Condition ELI vs NIT   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE  
## <numeric> <numeric> <numeric>  
## ENSG00000160856 899.044150322298 5.95816641904795 0.880757802255994  
## ENSG00000132704 726.714500543193 7.51080421195647 1.06032412810726  
## ENSG00000163534 971.658177885823 7.65430727006191 1.01952238246902  
## ENSG00000163568 281.007818880693 5.5591791993058 0.826142152527766  
## ENSG00000127074 134.973795352845 7.43206306191085 0.9534651350066  
## ENSG00000117322 1038.43574725233 7.42558152055317 1.01835464968555  
## stat pvalue padj  
## <numeric> <numeric> <numeric>  
## ENSG00000160856 6.76481820971277 1.33476687770677e-11 4.07700007179929e-09  
## ENSG00000132704 7.08349835004101 1.40560067455792e-12 7.62441427970117e-10  
## ENSG00000163534 7.50773833088897 6.01575977459831e-14 8.82354436103357e-11  
## ENSG00000163568 6.72908310306677 1.70735526978577e-11 5.02010319090937e-09  
## ENSG00000127074 7.79479268726423 6.45143093045147e-15 1.84516789548122e-11  
## ENSG00000117322 7.2917441117847 3.05967884824915e-13 2.6739043401324e-10  
## genSYMBOL ENTREZID GENENAME  
## <character> <integer> <character>  
## ENSG00000160856 FCRL3 115352 FCRL3  
## ENSG00000132704 FCRL2 79368 FCRL2  
## ENSG00000163534 FCRL1 115350 FCRL1  
## ENSG00000163568 AIM2 9447 AIM2  
## ENSG00000127074 RGS13 6003 RGS13  
## ENSG00000117322 CR2 1380 CR2

Guardo las tablas de los genes significativos compartidos por los distintos tipos de comparaciones en la carpeta de Resultados:

write.table(DEG\_SFIvsNIT\_ELIvsNIT, file = "./Resultados/DEG\_SFIvsNIT\_ELIvsNIT.csv", sep = ",", row.names = FALSE)  
  
write.table(DEG\_SFIvsNIT\_ELIvsSFI, file = "./Resultados/DEG\_SFIvsNIT\_ELIvsSFI.csv", sep = ",", row.names = FALSE)  
  
write.table(DEG\_ELIvsNIT\_ELIvsSFI, file = "./Resultados/DEG\_ELIvsNIT\_ELIvsSFI.csv", sep = ",", row.names = FALSE)  
  
write.table(DEG\_SFIvsNIT\_ELIvsNIT\_ELIvsSFI, file = "./Resultados/DEG\_SFIvsNIT\_ELIvsNIT\_ELIvsSFI.csv", sep = ",", row.names = FALSE)

Genero los diagramas de Venn con los genes significativos y también con los genes ‘down-regulated’ y ‘up-regulated’ para ver los genes compartidos por las comparaciones realizadas:

library("VennDiagram")

## Loading required package: grid

## Loading required package: futile.logger

library("RColorBrewer")

# Genes significativos:  
mycol<-brewer.pal(3, "Pastel2")  
venn.diagram(x=list(rownames(res\_SFIvsNITsig), rownames(res\_ELIvsNITsig),   
 rownames(res\_ELIvsSFIsig)),  
 category.names = c("SFI vs NIT", "ELI vs NIT", "ELI vs SFI"),  
 filename = "./Imagenes/significativeVennDiagram.png",   
 output = TRUE,   
 imagetype = "png",   
 height = 480,   
 width = 480,   
 resolution = 300,   
 compression = "lzw",   
 main = "Significative genes Venn Diagram",  
 main.cex = 0.5,  
 lwd = 2,   
 lty = "blank",  
 fill = mycol,  
 cex = 0.5,   
 fontfamily = "sans",   
 cat.cex = 0.5,   
 cat.default.pos = "outer",   
 cat.pos = c(-27, 27, 135),   
 cat.dist = c(0.055, 0.055, 0.085),   
 cat.fontfamily = "sans",   
 cat.col = mycol,   
 rotation = 1)

## [1] 1

# Genes 'down-regulated':  
mycol<-brewer.pal(3, "Pastel2")  
venn.diagram(x=list(rownames(SFIvsNIT\_down), rownames(ELIvsNIT\_down),   
 rownames(ELIvsSFI\_down)),  
 category.names = c("SFI vs NIT", "ELI vs NIT", "ELI vs SFI"),  
 filename = "./Imagenes/downRegulatedVennDiagram.png",   
 output = TRUE,   
 imagetype = "png",   
 height = 480,   
 width = 480,   
 resolution = 300,   
 compression = "lzw",  
 main = "Down-regulated genes Venn Diagram",  
 main.cex = 0.5,  
 lwd = 2,   
 lty = "blank",  
 fill = mycol,  
 cex = 0.5,   
 fontfamily = "sans",   
 cat.cex = 0.3,   
 cat.default.pos = "outer",   
 cat.pos = c(-27, 0, 180),   
 cat.dist = c(0.055, 0.055, 0.085),   
 cat.fontfamily = "sans",   
 cat.col = mycol,   
 rotation = 1)

## [1] 1

# Genes 'up-regulated':  
mycol<-brewer.pal(3, "Pastel2")  
venn.diagram(x=list(rownames(SFIvsNIT\_up), rownames(ELIvsNIT\_up),   
 rownames(ELIvsSFI\_up)),  
 category.names = c("SFI vs NIT", "ELI vs NIT", "ELI vs SFI"),  
 filename = "./Imagenes/upRegulatedVennDiagram.png",   
 output = TRUE,   
 imagetype = "png",   
 height = 480,   
 width = 480,   
 resolution = 300,   
 compression = "lzw",  
 main = "Up-regulated genes Ven Diagramm",  
 main.cex = 0.5,  
 lwd = 2,   
 lty = "blank",  
 fill = mycol,  
 cex = 0.5,   
 fontfamily = "sans",   
 cat.cex = 0.5,   
 cat.default.pos = "outer",   
 cat.pos = c(-27, 27, 135),   
 cat.dist = c(0.055, 0.055, 0.085),   
 cat.fontfamily = "sans",   
 cat.col = mycol,   
 rotation = 1)

## [1] 1

**SIGNIFICACIÓN BIOLÓGICA**

Una vez obtenidos los genes diferencialmente expresados en las distintas comparaciones, los resultados deben interpretarse biológicamente para lo que se lleva a cabo un análisis del conjunto de genes denominado análisis de enriquecimiento básico.

Voy a utilizar el paquete de Bioconductor *‘clusterProfiler’*.

library("clusterProfiler")

## Warning: package 'clusterProfiler' was built under R version 3.6.2

##

## Registered S3 method overwritten by 'enrichplot':  
## method from  
## fortify.enrichResult DOSE

## clusterProfiler v3.14.3 For help: https://guangchuangyu.github.io/software/clusterProfiler  
##   
## If you use clusterProfiler in published research, please cite:  
## Guangchuang Yu, Li-Gen Wang, Yanyan Han, Qing-Yu He. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology. 2012, 16(5):284-287.

##   
## Attaching package: 'clusterProfiler'

## The following object is masked from 'package:DelayedArray':  
##   
## simplify

El paquete permite realizar un análisis estadístico y la visualización de perfiles funcionales para genes y grupos de genes.

library(org.Hs.eg.db)

##

# Comparación SFI vs NIT:  
# Genero el vector con los genes:  
geneList\_SFIvsNIT<-as.vector(res\_SFIvsNIT$log2FoldChange)  
names(geneList\_SFIvsNIT)<-res\_SFIvsNIT$ENTREZID  
geneList\_SFIvsNIT<-sort(geneList\_SFIvsNIT, decreasing = TRUE)  
  
# Selecciono los genes con un valor absoluto de log2FoldChange mayor que 2:  
geneSFIvsNIT<-names(geneList\_SFIvsNIT)[abs(geneList\_SFIvsNIT) > 2]  
  
# Análisis de enriquecimiento:  
egoSFIvsNIT<-enrichGO(geneSFIvsNIT, OrgDb = "org.Hs.eg.db", ont = "BP", pAdjustMethod = "BH", pvalueCutoff = 0.05, readable = TRUE)  
  
# Guardo el análisis de enriquecimiento generado en la carpeta Resultados:  
write.table(egoSFIvsNIT, file = "./Resultados/egoSFIvsNIT.csv", sep = ",", row.names = FALSE)  
  
# Gráficos:  
png("./Imagenes/barplotSFIvsNIT.png", width = 800, height = 800, res = 72, units = "px", pointsize = 6, bg = "white")  
barplot(egoSFIvsNIT, showCategory = 15, font.size = 4, title = "SFI vs NIT Biolog. process")  
dev.off()

## quartz\_off\_screen   
## 2

png("./Imagenes/cnetplotSFIvsNIT.png", width = 1200, height = 820, res = 72, units = "px", pointsize = 12, bg = "white")  
cnetplot(egoSFIvsNIT, showCategory = 2, node\_label = "all", foldChange = geneList\_SFIvsNIT)  
dev.off()

## quartz\_off\_screen   
## 2

# Comparación ELI vs NIT:  
# Genero el vector con los genes:  
geneList\_ELIvsNIT<-as.vector(res\_ELIvsNIT$log2FoldChange)  
names(geneList\_ELIvsNIT)<-res\_ELIvsNIT$ENTREZID  
geneList\_ELIvsNIT<-sort(geneList\_ELIvsNIT, decreasing = TRUE)  
  
# Selecciono los genes con un valor absoluto de log2FoldChange mayor que 2:  
geneELIvsNIT<-names(geneList\_ELIvsNIT)[abs(geneList\_ELIvsNIT) > 2]  
  
# Análisis de enriquecimiento:  
egoELIvsNIT<-enrichGO(geneELIvsNIT, OrgDb = "org.Hs.eg.db", ont = "BP", pAdjustMethod = "BH", pvalueCutoff = 0.05, readable = TRUE)  
  
# Guardo el análisis de enriquecimiento generado en la carpeta Resultados:  
write.table(egoELIvsNIT, file = "./Resultados/egoELIvsNIT.csv", sep = ",", row.names = FALSE)  
  
# Gráficos:  
png("./Imagenes/barplotELIvsNIT.png", width = 800, height = 800, res = 72, units = "px", pointsize = 12, bg = "white")  
barplot(egoELIvsNIT, showCategory = 15, font.size = 4, title = "ELI vs NIT Biolog. process")  
dev.off()

## quartz\_off\_screen   
## 2

png("./Imagenes/cnetplotELIvsNIT.png", width = 1200, height = 820, res = 72, units = "px", pointsize = 12, bg = "white")  
cnetplot(egoELIvsNIT, showCategory = 3, node\_label = "all", foldChange = geneList\_ELIvsNIT)  
dev.off()

## quartz\_off\_screen   
## 2

# Comparación ELI vs SFI:  
# Genero el vector con los genes:  
geneList\_ELIvsSFI<-as.vector(res\_ELIvsSFI$log2FoldChange)  
names(geneList\_ELIvsSFI)<-res\_ELIvsSFI$ENTREZID  
geneList\_ELIvsSFI<-sort(geneList\_ELIvsSFI, decreasing = TRUE)  
  
# Selecciono los genes con un valor absoluto de log2FoldChange mayor que 2:  
geneELIvsSFI<-names(geneList\_ELIvsSFI)[abs(geneList\_ELIvsSFI) > 2]  
  
# Análisis de enriquecimiento:  
egoELIvsSFI<-enrichGO(geneELIvsSFI, OrgDb = "org.Hs.eg.db", ont = "BP", pAdjustMethod = "BH", pvalueCutoff = 0.05, readable = TRUE)  
  
# Guardo el análisis de enriquecimiento generado en la carpeta Resultados:  
write.table(egoELIvsSFI, file = "./Resultados/egoELIvsSFI.csv", sep = ",", row.names = FALSE)  
  
# Gráficos:  
png("./Imagenes/barplotELIvsSFI.png", width = 800, height = 800, res = 72, units = "px", pointsize = 12, bg = "white")  
barplot(egoELIvsSFI, showCategory = 15, font.size = 12, title = "ELI vs SFI Biological process")  
dev.off()

## quartz\_off\_screen   
## 2

png("./Imagenes/cnetplotELIvsSFI.png", width = 1200, height = 820, res = 72, units = "px", pointsize = 12, bg = "white")  
cnetplot(egoELIvsSFI, showCategory = 3, node\_label = "all", foldChange = geneList\_ELIvsSFI)  
dev.off()

## quartz\_off\_screen   
## 2