

## Домашнее задание №2

### Анализ таблицы экспрессии в R

0. Создаем новый проект в RStudio

1. Анализируем таблицу экспрессии LMNA\_AD\_raw\_counts.txt с помощью DEseq2 (не limma!). Файлы LMNA\_AD\_raw\_counts.txt и conditions.tsv находятся на гугл-диске в папке R\_RNA-seq > HW\_data. Из файла conditions.tsv вам понадобятся только те строки, которые совпадают с новой таблицей каунтов ("control" и "A").

Убираем выбросы, если есть.

```
> col_sums
wt_control_1 wt_control_2 wt_control_3 wt_A2d_1 wt_A2d_2
13085107 11707673 11850531 13764757 13637628
wt_A2d_3 wt_A_term_1 wt_A_term_2 wt_A_term_3 232_control_1
10931497 12655332 335 12857525 12915554
232_control_2 232_control_3 232_A_2d_1 232_A_2d_2 232_A_2d_3
5407278 9217078 15448461 11258146 11983946
232_A_term_1 232_A_term_2 232_A_term_3 482_control_1 482_control_2
5737658 12566719 11468705 10879810 12247422
482_control_3 482_A_2d_1 482_A_2d_2 482_A_2d_3 482_A_term_1
10711265 12801923 12437899 11464566 8045124
482_A_term_2 482_A_term_3
4207971 10838720
> mean(col_sums)
[1] 10745505
> col_sums[col_sums < 7500000]
wt_A_term_2 232_control_2 232_A_term_1 482_A_term_2
335 5407278 5737658 4207971
```

*В анализе были определены и исключены аутлаеры. Согласно коду, значения колонок, которые ниже порога в 7,500,000, были исключены из дальнейшего анализа. Исключены следующие образцы:*

*wt\_A\_term\_2 (значение 335)*

*232\_control\_2 (значение 5,407,278)*

*232\_A\_term\_1 (значение 5,737,658)*

*482\_A\_term\_2 (значение 4,207,971)*

```
head(exprs(es), 20)
wt_control_1 wt_control_2 wt_control_3 wt_A2d_1 wt_A2d_2 wt_A2d_3
1 0 0 0 0 0 0
2 0 0 0 0 0 0
3 0 0 0 0 0 0
4 0 0 0 0 0 0
5 0 0 0 0 0 0
6 0 0 0 0 0 0
7 0 0 0 0 0 0
8 0 0 0 0 0 0
9 0 0 0 0 0 0
10 0 0 0 0 0 0
11 0 0 0 0 0 0
12 0 0 1 0 0 0
13 0 0 0 0 0 0
14 0 0 0 0 0 0
```

15	0	0	0	0	0	0
16	1	0	0	0	0	0
17	0	0	0	0	0	0
18	18	15	15	8	9	6
19	0	0	0	0	0	0
20	0	0	0	0	0	0
	wt_A_term_1	wt_A_term_2	wt_A_term_3	232_control_1	232_control_2	
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	0	0	0	0	0	1
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0	0	0
16	0	0	0	0	0	0
17	0	0	0	0	0	0
18	3	0	2	9		2
19	0	0	0	0		0
20	0	0	0	0		0
	232_control_3	232_A_2d_1	232_A_2d_2	232_A_2d_3	232_A_term_1	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	
5	0	0	0	0	0	
6	0	0	0	0	0	
7	0	0	0	0	0	
8	0	0	0	0	0	
9	0	0	0	0	0	
10	0	0	0	0	0	
11	0	0	0	0	0	
12	0	0	0	0	0	
13	0	0	0	0	0	
14	0	0	0	0	0	
15	0	0	0	0	0	
16	0	0	0	0	0	
17	0	0	0	0	0	
18	12	5	3	3	2	
19	0	0	0	0	0	
20	0	0	0	0	0	
	232_A_term_2	232_A_term_3	482_control_1	482_control_2		
1	0	0	0	0		
2	0	0	0	0		
3	0	0	0	0		
4	0	0	0	0		
5	0	0	0	0		
6	0	0	0	0		
7	0	0	0	0		
8	0	0	0	0		
9	0	0	0	0		
10	0	1	0	0		
11	0	0	0	0		
12	0	0	0	0		
13	0	0	0	0		
14	0	0	0	0		
15	0	0	0	0		
16	0	0	0	0		
17	0	0	0	0		
18	6	7	11	16		
19	0	0	0	0		
20	0	0	0	0		
	482_control_3	482_A_2d_1	482_A_2d_2	482_A_2d_3	482_A_term_1	
1	0	0	0	0	0	
2	0	0	0	0	0	

```

3      0      0      0      0      0
4      0      0      0      0      0
5      0      0      0      0      0
6      0      0      0      0      0
7      0      0      0      0      0
8      0      0      0      0      0
9      0      0      0      0      0
10     0      0      0      0      0
11     0      0      0      0      0
12     0      0      0      3      0
13     0      0      0      0      0
14     0      0      0      0      0
15     0      0      0      0      0
16     0      0      0      0      0
17     0      0      0      0      0
18     7     19     21     18     2
19     0      0      0      0      0
20     0      0      0      0      0
    482_A_term_2 482_A_term_3
1      0      0
2      0      0
3      0      0
4      0      0
5      0      0
6      0      0
7      0      0
8      0      0
9      0      0
10     0      1
11     0      0
12     0      0
13     0      0
14     0      0
15     0      0
16     0      0
17     0      0
18     3      8
19     0      0
20     0      0
head(fData(es))
      Gene_id Symbol
1 ENSMUSG000000102693 <NA>
2 ENSMUSG000000064842 Gm26206
3 ENSMUSG0000000051951 Xkr4
4 ENSMUSG000000102851 Gm18956
5 ENSMUSG000000103377 <NA>
6 ENSMUSG000000104017 <NA>
head(pData(es))
      Condition Cell_type Treatment Day Replicate
wt_control_1 wt_control WT control 0 1
wt_control_2 wt_control WT control 0 2
wt_control_3 wt_control WT control 0 3
wt_A2d_1 wt_A2d WT A 2 1
wt_A2d_2 wt_A2d WT A 2 2
wt_A2d_3 wt_A2d WT A 2 3
> exprs(es)[which(fData(es)$Symbol == "Actb"), ]
      wt_control_1 wt_control_2 wt_control_3 wt_A2d_1 wt_A2d_2
      52554 40699 49100 21478 21831
      wt_A2d_3 wt_A_term_1 wt_A_term_2 wt_A_term_3 232_control_1
      19584 11444 0 11467 40289
232_control_2 232_control_3 232_A_2d_1 232_A_2d_2 232_A_2d_3
      18332 32419 28406 20198 24843
      232_A_term_1 232_A_term_2 232_A_term_3 482_control_1 482_control_2
      5749 11530 12641 37838 52971
482_control_3 482_A_2d_1 482_A_2d_2 482_A_2d_3 482_A_term_1
      46509 24009 21432 23028 9512
      482_A_term_2 482_A_term_3
      4291 12718
> length(unique(table$Symbol)) # Check if we have duplicated Symbol ids
[1] 32018
> head(exprs(es.qnorm.top12K))
      wt_control_1 wt_control_2 wt_control_3 wt_A2d_1

```

```

ENSMUSG000000064351      14.40848      15.34946      15.10982      15.93796
ENSMUSG000000000031      11.65209      11.58177      11.75861      16.89898
ENSMUSG000000034994      15.34946      15.23267      15.34946      15.47872
ENSMUSG000000064370      14.05342      14.70451      14.70451      15.64180
ENSMUSG000000024661      14.64123      14.45485      14.30586      15.10982
ENSMUSG000000026728      16.27161      16.89898      16.89898      14.36087
      wt_A2d_2 wt_A2d_3 wt_A_term_1 wt_A_term_2
ENSMUSG000000064351      16.27161      15.64180      15.64180      14.57431
ENSMUSG000000000031      16.89898      16.89898      16.89898      16.89898
ENSMUSG000000034994      15.64180      15.93796      15.34946      12.37882
ENSMUSG000000064370      15.47872      15.34946      15.01649      15.41409
ENSMUSG000000024661      15.01649      15.10982      15.23267      14.57431
ENSMUSG000000026728      14.36087      14.45485      13.39250      12.37882
      wt_A_term_3 232_control_1 232_control_2
ENSMUSG000000064351      15.64180      15.23267      15.47872
ENSMUSG000000000031      16.89898      13.59818      13.77769
ENSMUSG000000034994      15.34946      15.47872      15.23267
ENSMUSG000000064370      14.91202      14.57431      15.01649
ENSMUSG000000024661      15.23267      14.70451      14.64123
ENSMUSG000000026728      13.50085      16.89898      16.89898
      232_control_3 232_A_2d_1 232_A_2d_2 232_A_2d_3
ENSMUSG000000064351      15.64180      15.93796      16.27161      16.27161
ENSMUSG000000000031      13.52552      16.89898      16.89898      16.89898
ENSMUSG000000034994      15.34946      15.64180      15.34946      15.64180
ENSMUSG000000064370      14.91202      15.34946      15.47872      15.10982
ENSMUSG000000024661      14.50770      15.23267      15.23267      15.47872
ENSMUSG000000026728      16.89898      14.70451      14.64123      14.70451
      232_A_term_1 232_A_term_2 232_A_term_3
ENSMUSG000000064351      15.64180      15.64180      15.64180
ENSMUSG000000000031      16.89898      16.89898      16.89898
ENSMUSG000000034994      15.23267      15.01649      15.23267
ENSMUSG000000064370      15.34946      15.47872      15.34946
ENSMUSG000000024661      15.10982      15.10982      15.10982
ENSMUSG000000026728      13.57748      13.59818      13.72249
      482_control_1 482_control_2 482_control_3
ENSMUSG000000064351      15.10982      15.34946      15.23267
ENSMUSG000000000031      13.17961      12.92165      13.03620
ENSMUSG000000034994      15.23267      15.47872      15.47872
ENSMUSG000000064370      14.50770      14.64123      14.81176
ENSMUSG000000024661      14.81176      14.81176      15.01649
ENSMUSG000000026728      16.89898      16.89898      16.89898
      482_A_2d_1 482_A_2d_2 482_A_2d_3 482_A_term_1
ENSMUSG000000064351      15.93796      16.27161      15.23267      15.64180
ENSMUSG000000000031      16.89898      16.89898      16.89898      16.89898
ENSMUSG000000034994      15.64180      15.64180      15.93796      15.23267
ENSMUSG000000064370      15.34946      15.47872      14.81176      15.47872
ENSMUSG000000024661      15.23267      15.23267      15.47872      15.10982
ENSMUSG000000026728      14.36087      14.45485      14.50770      13.68171
      482_A_term_2 482_A_term_3
ENSMUSG000000064351      15.47872      15.64180
ENSMUSG000000000031      16.89898      16.89898
ENSMUSG000000034994      15.34946      15.23267
ENSMUSG000000064370      14.91202      15.34946
ENSMUSG000000024661      15.23267      15.10982
ENSMUSG000000026728      13.70342      13.80716

> head(phmap_z$kmeans$cluster, 20) # vector of genes and clusters they belong to
ENSMUSG000000064351 ENSMUSG000000000031 ENSMUSG000000034994
      1      11      2
ENSMUSG000000064370 ENSMUSG000000024661 ENSMUSG000000026728
      11      1      8
ENSMUSG000000068220 ENSMUSG000000023944 ENSMUSG000000064341
      9      2      13
ENSMUSG000000026185 ENSMUSG000000026193 ENSMUSG000000025393
      1      3      9
ENSMUSG000000029304 ENSMUSG000000035783 ENSMUSG000000026208
      2      3      10
ENSMUSG000000026043 ENSMUSG000000004891 ENSMUSG000000018593
      11      9      4
ENSMUSG000000040249 ENSMUSG000000064367
      2      11
> colnames(es.qnorm.top12K.no_out)

```

```

[1] "wt_control_1" "wt_control_2" "wt_control_3" "wt_A2d_1" "wt_A2d_
2"
[6] "wt_A2d_3" "wt_A_term_1" "wt_A_term_2" "wt_A_term_3" "232_con
trol_1"
[11] "232_control_2" "232_control_3" "232_A_2d_1" "232_A_2d_2" "232_A_2
d_3"
[16] "232_A_term_1" "232_A_term_2" "232_A_term_3" "482_control_1" "482_con
trol_2"
[21] "482_control_3" "482_A_2d_1" "482_A_2d_2" "482_A_2d_3" "482_A_t
erm_1"
[26] "482_A_term_2" "482_A_term_3"
> es.design

```

	Condition232_A_2d	Condition232_A_term	Condition232_control
wt_control_1	0	0	0
wt_control_2	0	0	0
wt_control_3	0	0	0
wt_A2d_1	0	0	0
wt_A2d_2	0	0	0
wt_A2d_3	0	0	0
wt_A_term_1	0	0	0
wt_A_term_2	0	0	0
wt_A_term_3	0	0	0
232_control_1	0	0	1
232_control_2	0	0	1
232_control_3	0	0	1
232_A_2d_1	1	0	0
232_A_2d_2	1	0	0
232_A_2d_3	1	0	0
232_A_term_1	0	1	0
232_A_term_2	0	1	0
232_A_term_3	0	1	0
482_control_1	0	0	0
482_control_2	0	0	0
482_control_3	0	0	0
482_A_2d_1	0	0	0
482_A_2d_2	0	0	0
482_A_2d_3	0	0	0
482_A_term_1	0	0	0
482_A_term_2	0	0	0
482_A_term_3	0	0	0
	Condition482_A_2d	Condition482_A_term	Condition482_control
wt_control_1	0	0	0
wt_control_2	0	0	0
wt_control_3	0	0	0
wt_A2d_1	0	0	0
wt_A2d_2	0	0	0
wt_A2d_3	0	0	0
wt_A_term_1	0	0	0
wt_A_term_2	0	0	0
wt_A_term_3	0	0	0
232_control_1	0	0	0
232_control_2	0	0	0
232_control_3	0	0	0
232_A_2d_1	0	0	0
232_A_2d_2	0	0	0
232_A_2d_3	0	0	0
232_A_term_1	0	0	0
232_A_term_2	0	0	0
232_A_term_3	0	0	0
482_control_1	0	0	1
482_control_2	0	0	1
482_control_3	0	0	1
482_A_2d_1	1	0	0
482_A_2d_2	1	0	0
482_A_2d_3	1	0	0
482_A_term_1	0	1	0
482_A_term_2	0	1	0
482_A_term_3	0	1	0
	Conditionwt_A_term	Conditionwt_A2d	Conditionwt_control
wt_control_1	0	0	1
wt_control_2	0	0	1
wt_control_3	0	0	1

```
wt_A2d_1 0 1 0
wt_A2d_2 0 1 0
wt_A2d_3 0 1 0
wt_A_term_1 1 0 0
wt_A_term_2 1 0 0
wt_A_term_3 1 0 0
232_control_1_1 0 0 0
232_control_1_2 0 0 0
232_control_1_3 0 0 0
232_A_2d_1 0 0 0
232_A_2d_2 0 0 0
232_A_2d_3 0 0 0
232_A_term_1 0 0 0
232_A_term_2 0 0 0
232_A_term_3 0 0 0
482_control_1_1 0 0 0
482_control_1_2 0 0 0
482_control_1_3 0 0 0
482_A_2d_1 0 0 0
482_A_2d_2 0 0 0
482_A_2d_3 0 0 0
482_A_term_1 0 0 0
482_A_term_2 0 0 0
482_A_term_3 0 0 0
attr("assign")
[1] 1 1 1 1 1 1 1 1 1
attr("contrasts")
attr("contrasts")$Condition
[1] "contr.treatment"
```

2. Смотрим на дифф.экспрессию для сравнений 232\_A\_term vs wt\_A\_term ИЛИ 482\_A\_term vs wt\_A\_term. Сколько получилось генов? Сколько апрегулированных, сколько даун? Какие критерии фильтрации вы выбрали?

```
> head(de)
Gene_id Symbol Entrez mean logFC AveE
xpr
ENSMUSG00000037010 ENSMUSG00000037010 Ap1n 30878 6.028574 -1.573569 6.028
574
ENSMUSG00000024030 ENSMUSG00000024030 Abcg1 11307 6.086207 -1.110971 6.086
207
ENSMUSG00000079018 ENSMUSG00000079018 Ly6c1 17067 7.822497 1.108399 7.822
497
ENSMUSG00000003153 ENSMUSG00000003153 Slc2a3 20527 4.129266 -2.967848 4.129
266
ENSMUSG00000006403 ENSMUSG00000006403 Adamts4 240913 7.851564 1.007337 7.851
564
ENSMUSG00000059824 ENSMUSG00000059824 Dbp 13170 6.540170 -1.368297 6.540
170
t P.Value adj.P.Val B
ENSMUSG00000037010 -9.309179 8.224470e-09 0.00009869364 9.689885
ENSMUSG00000024030 -7.433459 3.022772e-07 0.00115104276 6.636648
ENSMUSG00000079018 7.341406 3.651191e-07 0.00115104276 6.472098
ENSMUSG00000003153 -7.317331 3.836809e-07 0.00115104276 6.428831
ENSMUSG00000006403 6.892766 9.317345e-07 0.00182923801 5.650246
ENSMUSG00000059824 -6.862607 9.932555e-07 0.00182923801 5.593823
> head(de2)
Gene_id Symbol Entrez mean logF
C
ENSMUSG00000070348 ENSMUSG00000070348 Ccnd1 12443 9.352436 1.047299
1
ENSMUSG00000030268 ENSMUSG00000030268 Bcat1 12035 8.897308 1.020356
0
ENSMUSG00000059824 ENSMUSG00000059824 Dbp 13170 6.540170 -1.519243
0
ENSMUSG00000036356 ENSMUSG00000036356 Csgalnact1 234356 8.407592 0.868951
5
ENSMUSG00000085148 ENSMUSG00000085148 Mir22hg 100042498 7.995290 1.090902
4
```

```

ENSMUSG00000022369 ENSMUSG00000022369 Mtbp 105837 8.393290 0.743170
8
      AveExpr      t      P.Value      adj.P.Val      B
ENSMUSG00000070348 9.352436 8.101599 7.940233e-08 0.0008280301 7.836188
ENSMUSG00000030268 8.897308 7.815543 1.397071e-07 0.0008280301 7.347243
ENSMUSG00000059824 6.540170 -7.619669 2.070075e-07 0.0008280301 7.004751
ENSMUSG00000036356 8.407592 7.128165 5.680104e-07 0.0015878800 6.117737
ENSMUSG000000085148 7.995290 7.042963 6.788609e-07 0.0015878800 5.959971
ENSMUSG000000022369 8.393290 6.968504 7.939400e-07 0.0015878800 5.821137
> # Take only upregulated DEGs (or only down)
> set1 <- de_filt[logFC > 1, Symbol]
> set2 <- de2_filt[logFC > 1, Symbol]
> venn.diagram(
+   x = list(set1, set2),
+   category.names = c("232_vs_wt", "482_vs_wt"),
+   filename = '#1_venn_diagramm.png',
+   output=TRUE)
[1] 1
> intersect(set1, set2) # Look at the intersection of UP-regulated genes in b
oth comparisons
[1] "Tnc" "Dpys13"
> setdiff(set2, set1) # unique upregulated genes for 482_vs_wt
[1] "Ccnd1" "Bcat1" "Mir22hg" "Sod3" "Leng8" "Srsf2"
[7] "Igfn1" "Maspl" "Ets2" "Enpp2" "Igfbp3" "Cirbp"
[13] "Srsf6" "Sema3a" "Hdac4" "Thbs2" "Ccnc80" "Srsf5"
[19] "Tnfrsf12a"

# Take only downregulated DEGs
set3 <- de_filt[log2FoldChange < -1, Symbol]
set4 <- de2_filt[logFC < -1, Symbol]
venn.diagram(
  x = list(set3, set4),
  category.names = c("232_vs_wt", "482_vs_wt"),
  filename = '#1_venn_diagramm.png',
  output=TRUE)
> intersect(set3, set4) # Look at the intersection of Down-regulated genes in
both comparisons
[1] "Slc2a3" "Apln"
> setdiff(set4, set3) # unique upregulated genes for 482_vs_wt
[1] "Gm5617" "Pde4a"
[3] "Pax7" "Serpine1"
[5] "Syng1" "Rpl19-ps11"
[7] "Pigh" "Zc3h8"
[9] "9530077C05Rik" "Mfsd9"
[11] "Rps17" "Fos"
[13] "Gm6206" "Fam83a"
[15] "Snx16" "Mrps36"
[17] "Cdk15" "Gipr"
[19] "Cys1" "Dnajc28"
[21] "Mmp17" "Ube2t"
[23] "Disp2" "Lonrf3"
[25] "A930005H10Rik" "Lypd6"
[27] "Pip5k1b" "Cyb5d1"
[29] "Oxld1" "Npy4r"
[31] "Agap2" "Lrmda"
[33] "Nmrk1" "Tmem229b"
[35] "Itga2b" "Dusp19"
[37] "Proser2" "Fa2h"
[39] "Ppp1r3b" "Tcea3"
[41] "Crebl2" "Fosl1"
[43] "8430429K09Rik" "Gm11110"
[45] "Asb2" "Cntn3"
[47] "Cc19" "Dbp"

```

При сравнении 232\_A\_term vs wt\_A\_term получилось 10 генов:

de_filt	10 obs. of 10 variables
---------	-------------------------

При сравнении 482\_A\_term vs wt\_A\_term получилось 71 гена:

de2_filt	71 obs. of 11 variables
----------	-------------------------



Из них 7 апрегулированных и 5 даун для (232\_A\_term vs wt\_A\_term) и 21 апрегулированных, 50 даун (482\_A\_term vs wt\_A\_term):

set1	chr	[1:7]	"Ly6c1"	"Adamts4"	"Dio2"	"Tnc"	"My12"	"Armcx4"	"Dpys13"		
set2	chr	[1:21]	"Ccnd1"	"Bcat1"	"Mir22hg"	"Tnc"	"Sod3"	"Leng8"	"dpys13"	"Srsf2"	"Igf...
set3	chr	[1:5]	"Gdpd2"	"slc2a3"	"Apln"	"Ndr1"	"Nos2"				
set4	chr	[1:50]	"Gm5617"	"Pde4a"	"Pax7"	"Serpine1"	"Syng1"	"Rpl19-ps11"	"Pigh"	"Zc3...	

Критерии фильтрации были выбраны по  $\log_2FC$ . Гены, которые имеют положительное значение  $\log_2FC$  (конкретно в пайплайне выбраны значения  $> 1$ ) считаются апрегулированными. Гены с отрицательным значением  $\log_2FC$  ( $< -1$ ) считаются даунрегулированными.

3. Рисуем картинки: PCA, ап- и даун-регулированные гены (хитмап), кластеризацию, молекулярные пути (overrepresentation analysis (ORA, тест-сравнения) и GSEA), что получится.

Какие пути в общем активированы (и в каком условии)? Какие подавлены?

Как я поняла, пути с низкими  $padj$  ( $p$ -value) считаются активированными, а с высоким – подавленными. Получается, активированы 10 путей, подавлено 40:

```
fr_res <- fr[order(NES)][padj < 0.05] # list of pathways
fr_res2 <- fr[order(NES)][padj > 0.05]
```

```
fr_res      10 obs. of 8 variables
fr_res2     40 obs. of 8 variables
```

	pathway	pval	padj	ES	NES	nMoreExtreme	size	leadingEdge
1	HALLMARK_HYPOXIA	0.001156069	0.04306632	-0.4682623	-1.995903	0	183	17988, 1....
2	HALLMARK_XENOBIOTIC_METABOLISM	0.003575685	0.04306632	-0.3807825	-1.581794	2	152	16010, 1....
3	HALLMARK_TNFA_SIGNALING_VIA_NFKB	0.002328289	0.04306632	-0.3650283	-1.550047	1	178	18787, 2....
4	HALLMARK_GLYCOLYSIS	0.004656577	0.04306632	-0.3557660	-1.510716	3	178	13807, 5....
5	HALLMARK_OXIDATIVE_PHOSPHORYLATION	0.003452244	0.04306632	-0.3501976	-1.503415	2	199	109672, ....
6	HALLMARK_ADIPOGENESIS	0.009163803	0.04581901	-0.3318193	-1.421487	7	195	170768, ....
7	HALLMARK_MYOGENESIS	0.007575758	0.04306632	0.2844618	1.439481	0	193	17906, 1....
8	HALLMARK_MITOTIC_SPINDLE	0.007751938	0.04306632	0.3212741	1.622694	0	196	21844, 7....
9	HALLMARK_E2F_TARGETS	0.007518797	0.04306632	0.3511026	1.789605	0	199	20382, 1....
10	HALLMARK_G2M_CHECKPOINT	0.007575758	0.04306632	0.4390376	2.222818	0	198	12443, 2....

	pathway	pval	padj	ES	NES	nMoreExtreme	size	leadingEdge
1	HALLMARK_IL2_STAT5_SIGNALING	0.01323706	0.06016847	-0.3359752	-1.4013285	10	162	17988, 2....
2	HALLMARK_PANCREAS_BETA_CELLS	0.15844544	0.33009467	-0.4574225	-1.2881499	105	19	16392, 1....
3	HALLMARK_MTORC1_SIGNALING	0.05909618	0.18467555	-0.2985354	-1.2806792	50	201	20527, 5....
4	HALLMARK_FATTY_ACID_METABOLISM	0.06968215	0.20494750	-0.3104179	-1.2757670	56	141	13807, 1....
5	HALLMARK_BILE_ACID_METABOLISM	0.15175097	0.33009467	-0.3198631	-1.2287021	116	87	19659, 2....
6	HALLMARK_INTERFERON_GAMMA_RESPONSE	0.11976048	0.29940120	-0.2885831	-1.2086009	99	165	18712, 1....
7	HALLMARK_PEROXISOME	0.17250324	0.33173700	-0.3127447	-1.2013580	132	87	20148, 7....
8	HALLMARK_APICAL_SURFACE	0.22746781	0.40468584	-0.3664524	-1.1854804	158	33	72043, 1....
9	HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	0.15411356	0.33009467	-0.2779419	-1.1840159	132	184	16010, 1....
10	HALLMARK_UV_RESPONSE_DN	0.16666667	0.33173700	-0.2857232	-1.1718514	135	139	18787, 1....
11	HALLMARK_COAGULATION	0.23791349	0.40468584	-0.2924032	-1.1444481	186	95	18787, 1....



Рисуем enrichmentPlot для Adipogenesis, Myogenesis или похожих (можно сделать анализ путей по базе данных Hallmark, там есть Adipogenesis, Myogenesis:

```
hall_df <- msigdb(species = "Mus musculus", category = "H")
pathways <- split(hall_df$entrez_gene, hall_df$gs_name)
)
```

```
> dds
class: DESeqDataSet
dim: 15345 27
metadata(1): version
assays(4): counts mu H cooks
rownames(15345): ENSMUSG000000109644 ENSMUSG00000007777 ...
               ENSMUSG00000055670 ENSMUSG00000039068
rowData names(50): baseMean baseVar ... deviance maxCooks
colnames(27): wt_control_1 wt_control_2 ... 482_A_term_2 482_A_term_3
colData names(6): Condition Cell_type ... Replicate sizeFactor
> unique(dds$Condition)
[1] wt_control wt_A2d wt_A_term 232_control 232_A_2d
[6] 232_A_term 482_control 482_A_2d 482_A_term
9 Levels: 232_A_2d 232_A_term 232_control 482_A_2d ... wt_control
> de <- results(dds, contrast = c("Condition", "232_A_term", "wt_A_term"), co
oksCutoff = F)
> head(de)
log2 fold change (MLE): Condition 232_A_term vs wt_A_term
Wald test p-value: Condition 232_A_term vs wt_A_term
DataFrame with 6 rows and 6 columns
```

	baseMean	log2FoldChange	lfcSE	stat
	<numeric>	<numeric>	<numeric>	<numeric>
ENSMUSG000000109644	3.39691	-0.468344	0.701545	-0.667589
ENSMUSG00000007777	144.77678	-0.036268	0.256597	-0.141343
ENSMUSG00000086714	2.56349	0.106899	0.827136	0.129240
ENSMUSG000000043644	8.38648	-1.016225	0.526836	-1.928919
ENSMUSG00000020831	5.14315	0.124832	0.715380	0.174497
ENSMUSG000000107002	155.10079	-0.302158	0.259364	-1.164993

```

      pvalue      padj
      <numeric> <numeric>
ENSMUSG000000109644 0.5043960      1
ENSMUSG00000007777 0.8875993      1
ENSMUSG00000086714 0.8971681      1
ENSMUSG000000043644 0.0537409      1
ENSMUSG00000020831 0.8614749      1
ENSMUSG000000107002 0.2440219      1
> head(de)
      ID      baseMean log2FoldChange      lfcSE      stat
      <char>      <num>      <num>      <num>      <num>
1: ENSMUSG000000109644 3.396913 -0.46834390 0.7015452 -0.6675890
2: ENSMUSG00000007777 144.776782 -0.03626802 0.2565966 -0.1413425
3: ENSMUSG00000086714 2.563494 0.10689868 0.8271362 0.1292395
4: ENSMUSG000000043644 8.386479 -1.01622478 0.5268365 -1.9289188
5: ENSMUSG00000020831 5.143146 0.12483163 0.7153798 0.1744970
6: ENSMUSG000000107002 155.100790 -0.30215760 0.2593644 -1.1649927
      pvalue      padj
      <num>      <num>
1: 0.50439597      1
2: 0.88759934      1
3: 0.89716814      1
4: 0.05374094      1
5: 0.86147489      1
6: 0.24402194      1
> tail(de)
      ID      baseMean log2FoldChange      lfcSE      stat
      <char>      <num>      <num>      <num>      <num>
1: ENSMUSG00000073062 26.41154 0.138517208 0.5474413 0.25302658
2: ENSMUSG00000034430 296.49830 -0.058167732 0.3186137 -0.18256506
3: ENSMUSG00000034636 822.21841 0.043795366 0.2468185 0.17743958
4: ENSMUSG00000029860 5649.36869 0.142074082 0.2384161 0.59590804
```

```

5: ENSMUSG00000055670 773.00379 -0.004808387 0.2386892 -0.02014497
6: ENSMUSG00000039068 431.55900 0.079078551 0.2355799 0.33567608
      pvalue padj
      <num> <num>
1: 0.8002477 1
2: 0.8551393 1
3: 0.8591631 1
4: 0.5512367 1
5: 0.9839277 1
6: 0.7371151 1
> head(fData(es))
      Gene_id Symbol Entrez
ENSMUSG00000109644 ENSMUSG00000109644 0610005C13Rik 71661
ENSMUSG00000007777 ENSMUSG00000007777 0610009B22Rik 66050
ENSMUSG00000086714 ENSMUSG00000086714 0610009E02Rik 100125929
ENSMUSG00000043644 ENSMUSG00000043644 0610009L18Rik 66838
ENSMUSG00000020831 ENSMUSG00000020831 0610010K14Rik 104457
ENSMUSG00000107002 ENSMUSG00000107002 0610012G03Rik 106264
> head(de)
      ID baseMean log2FoldChange lfcSE stat
      <char> <num> <num> <num> <num>
1: ENSMUSG00000035095 29.29220 2.284777 0.3752406 6.088833
2: ENSMUSG00000112963 33.56477 1.845490 0.3228939 5.715467
3: ENSMUSG00000037033 28.47265 2.170924 0.3918672 5.539948
4: ENSMUSG00000007682 112.88245 1.320590 0.2416715 5.464399
5: ENSMUSG00000031391 164.79042 1.396476 0.2616869 5.336437
6: ENSMUSG00000026228 21.30095 1.959766 0.3958156 4.951210
      pvalue padj Gene_id Symbol Entrez
      <num> <num> <char> <char> <char>
1: 1.137370e-09 0.000003954962 ENSMUSG00000035095 Fam167a 219148
2: 1.094031e-08 0.000024725109 ENSMUSG00000112963 Gm6093 619715
3: 3.025613e-08 0.000054874469 ENSMUSG00000037033 Clca3b 229927
4: 4.644774e-08 0.000069981263 ENSMUSG00000007682 Dio2 13371
5: 9.479096e-08 0.000116851401 ENSMUSG00000031391 L1cam 16728
6: 7.375339e-07 0.000714354227 ENSMUSG00000026228 Htr2b 15559
> # Look at the specific gene
> de[Symbol == "Gm6093"]
      ID baseMean log2FoldChange lfcSE stat
      <char> <num> <num> <num> <num>
1: ENSMUSG00000112963 33.56477 1.84549 0.3228939 5.715467
      pvalue padj Gene_id Symbol Entrez
      <num> <num> <char> <char> <char>
1: 1.094031e-08 0.00002472511 ENSMUSG00000112963 Gm6093 619715
> head(go_deg)
      ID
GO:0071073 GO:0071073
GO:0098664 GO:0098664
GO:0014831 GO:0014831
GO:0007210 GO:0007210
GO:0051000 GO:0051000
GO:1903727 GO:1903727
      Description
GO:0071073 positive regulation of phospholipid biosynthetic process
GO:0098664 G protein-coupled serotonin receptor signaling pathway
GO:0014831 gastro-intestinal system smooth muscle contraction
GO:0007210 serotonin receptor signaling pathway
GO:0051000 positive regulation of nitric-oxide synthase activity
GO:1903727 positive regulation of phospholipid metabolic process
      GeneRatio BgRatio pvalue p.adjust qvalue geneID
GO:0071073 1/4 12/28943 0.001657487 0.03080871 0.00283121 Htr2b
GO:0098664 1/4 12/28943 0.001657487 0.03080871 0.00283121 Htr2b
GO:0014831 1/4 14/28943 0.001933534 0.03080871 0.00283121 Htr2b
GO:0007210 1/4 15/28943 0.002071536 0.03080871 0.00283121 Htr2b
GO:0051000 1/4 18/28943 0.002485457 0.03080871 0.00283121 Htr2b
GO:1903727 1/4 18/28943 0.002485457 0.03080871 0.00283121 Htr2b
      Count
GO:0071073 1
GO:0098664 1
GO:0014831 1
GO:0007210 1
GO:0051000 1
GO:1903727 1

```

```

> head(go_deg$Description, 20)
[1] "positive regulation of phospholipid biosynthetic process"
[2] "G protein-coupled serotonin receptor signaling pathway"
[3] "gastro-intestinal system smooth muscle contraction"
[4] "serotonin receptor signaling pathway"
[5] "positive regulation of nitric-oxide synthase activity"
[6] "positive regulation of phospholipid metabolic process"
[7] "G protein-coupled receptor internalization"
[8] "regulation of phospholipid biosynthetic process"
[9] "response to amine"
[10] "activation of phospholipase C activity"
[11] "protein kinase C-activating G protein-coupled receptor signaling pathwa
y"
[12] "phasic smooth muscle contraction"
[13] "positive regulation of monooxygenase activity"
[14] "desensitization of G protein-coupled receptor signaling pathway"
[15] "negative adaptation of signaling pathway"
[16] "adaptation of signaling pathway"
[17] "regulation of nitric-oxide synthase activity"
[18] "positive regulation of phospholipase C activity"
[19] "positive regulation of telomerase activity"
[20] "cGMP-mediated signaling"
> # Also you can copy genes and go to MSigDB or another online tool
> paste0(genes, collapse = " ")
[1] "Fam167a Gm6093 Clca3b Htr2b Parm1"
> # Myogenesis:
> hall_df <- msigdb(species = "Mus musculus", category = "H")
> hall_df
# A tibble: 7,384 × 18
  gs_cat gs_subcat gs_name      gene_symbol entrez_gene ensembl_gene
  <chr>   <chr>     <chr>      <chr>          <int>   <chr>
1 H      ""       HALLMARK_ADIP... Abca1          11303 ENSMUSG0000...
2 H      ""       HALLMARK_ADIP... Abcb8          74610 ENSMUSG0000...
3 H      ""       HALLMARK_ADIP... Acaa2          52538 ENSMUSG0000...
4 H      ""       HALLMARK_ADIP... Acadl          11363 ENSMUSG0000...
5 H      ""       HALLMARK_ADIP... Acadm          11364 ENSMUSG0000...
6 H      ""       HALLMARK_ADIP... Acads          11409 ENSMUSG0000...
7 H      ""       HALLMARK_ADIP... Acly           104112 ENSMUSG0000...
8 H      ""       HALLMARK_ADIP... Aco2           11429 ENSMUSG0000...
9 H      ""       HALLMARK_ADIP... Acox1          11430 ENSMUSG0000...
10 H     ""       HALLMARK_ADIP... Adcy6          11512 ENSMUSG0000...
# i 7,374 more rows
# i 12 more variables: human_gene_symbol <chr>,
#   human_entrez_gene <int>, human_ensembl_gene <chr>, gs_id <chr>,
#   gs_pm_id <chr>, gs_geoid <chr>, gs_exact_source <chr>,
#   gs_url <chr>, gs_description <chr>, taxon_id <int>,
#   ortholog_sources <chr>, num_ortholog_sources <dbl>
# i Use `print(n = ...)` to see more rows
> fr[order(padj)] # Look at this

```

	pathway	pval	padj
	<char>	<num>	<num>
1:	HALLMARK_E2F_TARGETS	0.007518797	0.04306632
2:	HALLMARK_G2M_CHECKPOINT	0.007575758	0.04306632
3:	HALLMARK_GLYCOLYSIS	0.004656577	0.04306632
4:	HALLMARK_HYPOXIA	0.001156069	0.04306632
5:	HALLMARK_MITOTIC_SPINDLE	0.007751938	0.04306632
6:	HALLMARK_MYOGENESIS	0.007575758	0.04306632
7:	HALLMARK_OXIDATIVE_PHOSPHORYLATION	0.003452244	0.04306632
8:	HALLMARK_TNFA_SIGNALING_VIA_NFKB	0.002328289	0.04306632
9:	HALLMARK_XENOBIOTIC_METABOLISM	0.003575685	0.04306632
10:	HALLMARK_ADIPOGENESIS	0.009163803	0.04581901
11:	HALLMARK_IL2_STAT5_SIGNALING	0.013237064	0.06016847
12:	HALLMARK_ANGIOGENESIS	0.028938907	0.12057878
13:	HALLMARK_APOPTOSIS	0.047368421	0.16786571
14:	HALLMARK_KRAS_SIGNALING_UP	0.048192771	0.16786571
15:	HALLMARK_MYC_TARGETS_V1	0.050359712	0.16786571
16:	HALLMARK_MTORC1_SIGNALING	0.059096176	0.18467555
17:	HALLMARK_FATTY_ACID_METABOLISM	0.069682152	0.20494750
18:	HALLMARK_ESTROGEN_RESPONSE_LATE	0.103030303	0.28619529
19:	HALLMARK_ALLOGRAFT_REJECTION	0.110526316	0.29085873
20:	HALLMARK_INTERFERON_GAMMA_RESPONSE	0.119760479	0.29940120
21:	HALLMARK_APICAL_JUNCTION	0.136363636	0.32467532

22:	HALLMARK_BILE_ACID_METABOLISM	0.151750973	0.33009467
23:	HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	0.154113557	0.33009467
24:	HALLMARK_PANCREAS_BETA_CELLS	0.158445441	0.33009467
25:	HALLMARK_PEROXISOME	0.172503243	0.33173700
26:	HALLMARK_UV_RESPONSE_DN	0.166666667	0.33173700
27:	HALLMARK_INFLAMMATORY_RESPONSE	0.205665025	0.38086116
28:	HALLMARK_APICAL_SURFACE	0.227467811	0.40468584
29:	HALLMARK_COAGULATION	0.237913486	0.40468584
30:	HALLMARK_NOTCH_SIGNALING	0.242811502	0.40468584
31:	HALLMARK_COMPLEMENT	0.256902761	0.40944477
32:	HALLMARK_ESTROGEN_RESPONSE_EARLY	0.262044653	0.40944477
33:	HALLMARK_P53_PATHWAY	0.290322581	0.43988270
34:	HALLMARK_KRAS_SIGNALING_DN	0.398742138	0.58638550
35:	HALLMARK_HEME_METABOLISM	0.411971831	0.58853119
36:	HALLMARK_HEDGEHOG_SIGNALING	0.553054662	0.76813148
37:	HALLMARK_UV_RESPONSE_UP	0.575091575	0.77715078
38:	HALLMARK_CHOLESTEROL_HOMEOSTASIS	0.611570248	0.80469769
39:	HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	0.638620690	0.81874447
40:	HALLMARK_IL6_JAK_STAT3_SIGNALING	0.720000000	0.87804878
41:	HALLMARK_TGF_BETA_SIGNALING	0.718978102	0.87804878
42:	HALLMARK_ANDROGEN_RESPONSE	0.763427110	0.88770594
43:	HALLMARK_WNT_BETA_CATENIN_SIGNALING	0.758865248	0.88770594
44:	HALLMARK_INTERFERON_ALPHA_RESPONSE	0.911688312	0.96988118
45:	HALLMARK_PI3K_AKT_MTOR_SIGNALING	0.879069767	0.96988118
46:	HALLMARK_SPERMATOGENESIS	0.869170984	0.96988118
47:	HALLMARK_UNFOLDED_PROTEIN_RESPONSE	0.901869159	0.96988118
48:	HALLMARK_DNA_REPAIR	0.998778999	0.99877900
49:	HALLMARK_MYC_TARGETS_V2	0.996282528	0.99877900
50:	HALLMARK_PROTEIN_SECRETION	0.990825688	0.99877900

	ES	NES	nMoreExtreme	pathway	pval	padj
	<num>	<num>	<num>	size	leadingEdge	
				<int>	<list>	
1:	0.3511026	1.7896048	0	199	20382, 1....	
2:	0.4390376	2.2228177	0	198	12443, 2....	
3:	-0.3557660	-1.5107162	3	178	13807, 5....	
4:	-0.4682623	-1.9959028	0	183	17988, 1....	
5:	0.3212741	1.6226939	0	196	21844, 7....	
6:	0.2844618	1.4394813	0	193	17906, 1....	
7:	-0.3501976	-1.5034154	2	199	109672, ....	
8:	-0.3650283	-1.5500472	1	178	18787, 2....	
9:	-0.3807825	-1.5817942	2	152	16010, 1....	
10:	-0.3318193	-1.4214875	7	195	170768, ....	
11:	-0.3359752	-1.4013285	10	162	17988, 2....	
12:	0.4333398	1.5699259	8	29	50706, 1....	
13:	0.2669392	1.2881149	8	138	12443, 1....	
14:	0.2642801	1.2855559	7	154	83397, 1....	
15:	0.2414536	1.2446868	6	201	20382, 2....	
16:	-0.2985354	-1.2806792	50	201	20527, 5....	
17:	-0.3104179	-1.2757670	56	141	13807, 1....	
18:	0.2393259	1.1808365	16	163	12443, 1....	
19:	0.2443721	1.1729504	20	136	18752, 1....	
20:	-0.2885831	-1.2086009	99	165	18712, 1....	
21:	0.2314111	1.1444352	23	159	18710, 2....	
22:	-0.3198631	-1.2287021	116	87	19659, 2....	
23:	-0.2779419	-1.1840159	132	184	16010, 1....	
24:	-0.4574225	-1.2881499	105	19	16392, 1....	
25:	-0.3127447	-1.2013580	132	87	20148, 7....	
26:	-0.2857232	-1.1718514	135	139	18787, 1....	
27:	-0.2792674	-1.1423613	166	138	18787, 2....	
28:	-0.3664524	-1.1854804	158	33	72043, 1....	
29:	-0.2924032	-1.1444481	186	95	18787, 1....	
30:	0.3174015	1.1559545	75	30	12443, 5....	
31:	-0.2669855	-1.1078850	213	153	18787, 1....	
32:	-0.2610784	-1.1029114	222	169	16010, 2....	
33:	-0.2563893	-1.0941224	251	188	17988, 1....	
34:	-0.2569116	-1.0256299	316	112	22784, 1....	
35:	-0.2418029	-1.0235220	350	171	12266, 1....	
36:	0.2561487	0.9279888	171	29	16728, 1....	
37:	-0.2323494	-0.9578692	470	144	13807, 1....	
38:	0.2162544	0.9228399	147	72	27279, 7....	
39:	-0.2568051	-0.8926515	462	47	14778, 5....	
40:	-0.2305971	-0.8525022	539	66	21938, 1....	

```

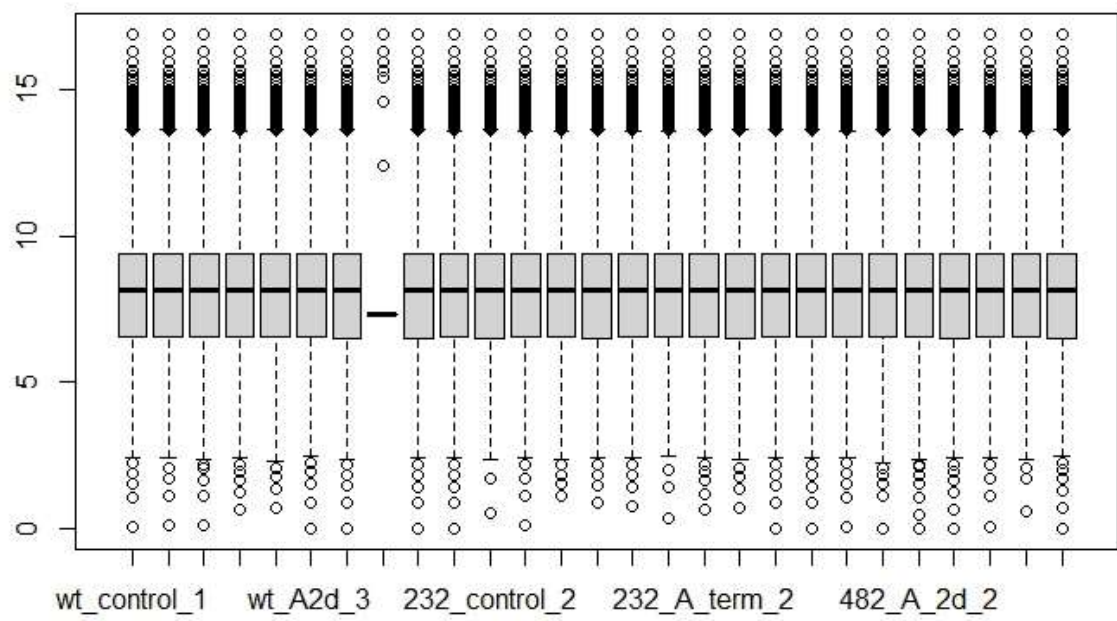
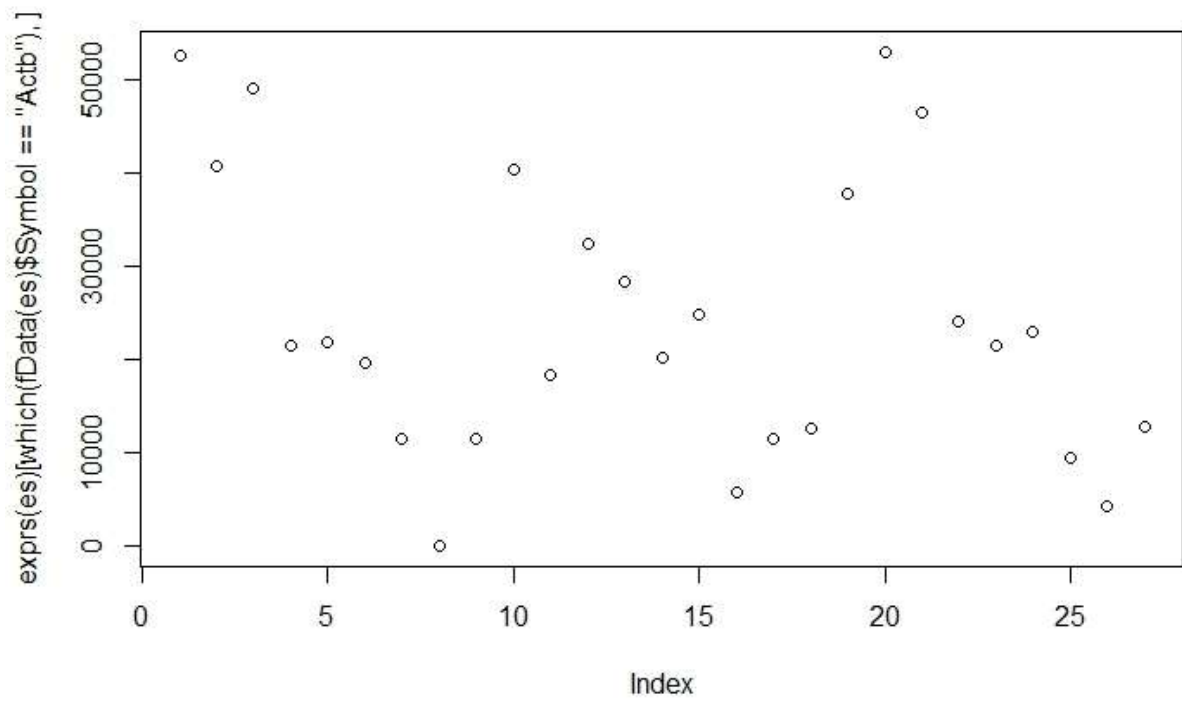
41: 0.2141964 0.8627972 196 52 21825, 2....
42: -0.2158032 -0.8408036 596 93 17988, 1....
43: -0.2462529 -0.8122235 534 37 14677, 1....
44: -0.1914741 -0.7333196 701 83 50908, 5....
45: 0.1865667 0.8416651 188 98 21844, 1....
46: -0.2013454 -0.7688409 670 80 56734, 1....
47: 0.1771465 0.8210669 192 110 233908, ....
48: -0.1506855 -0.6213884 817 147 18578, 6....
49: 0.1476339 0.6032417 267 58 17217, 1....
50: 0.1501991 0.6739619 215 92 69938, 7....
      ES      NES nMoreExtreme size leadingEdge
> str(collapsedPathways)
List of 2
 $ mainPathways : chr [1:8] "HALLMARK_HYPOXIA" "HALLMARK_OXIDATIVE_PHOSPHORY
LATION" "HALLMARK_XENOBIOTIC_METABOLISM" "HALLMARK_E2F_TARGETS" ...
 $ parentPathways: Named chr [1:10] NA "HALLMARK_HYPOXIA" NA NA ...
 ..- attr(*, "names")= chr [1:10] "HALLMARK_HYPOXIA" "HALLMARK_TNFA_SIGNALIN
G_VIA_NFKB" "HALLMARK_OXIDATIVE_PHOSPHORYLATION" "HALLMARK_XENOBIOTIC_METABOL
ISM" ...

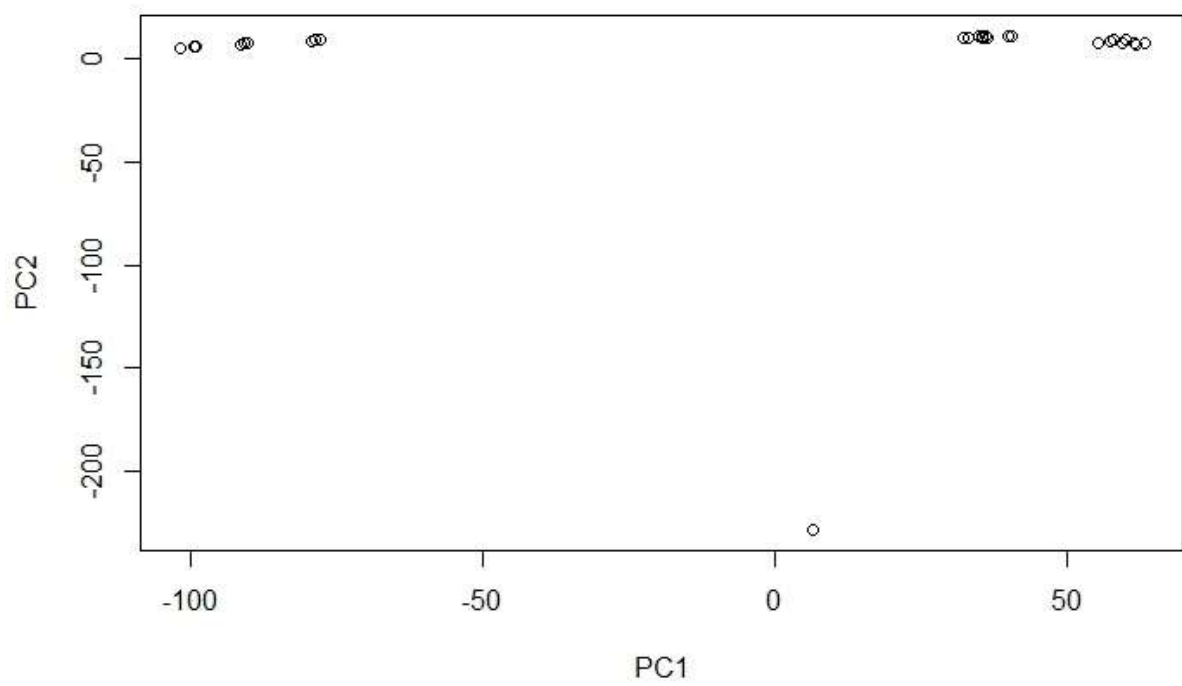
```

*Myogenesis: Значение  $P_{adj} < 0.05$  и  $NES > 0$ , что указывает на апрегуляцию пути myogenesis в 232\_term.*

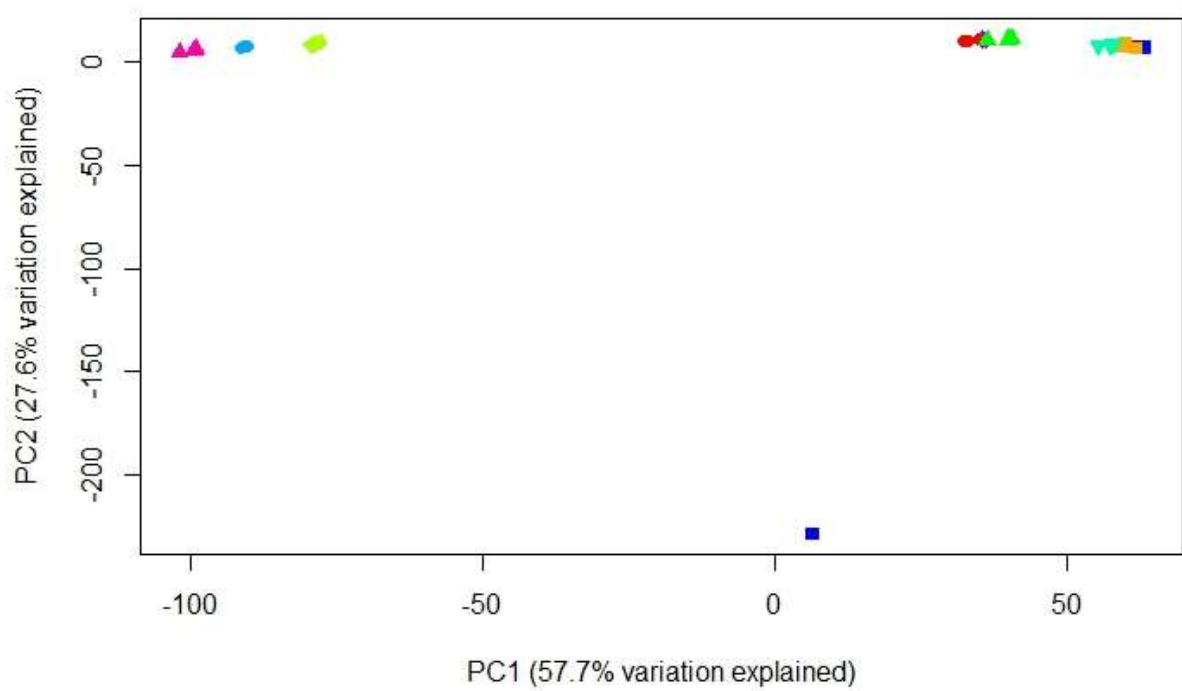
*Adipogenesis: Значение  $P_{adj} < 0.05$  и  $NES > 0$ , что также указывает на апрегуляцию пути adipogenesis в 232\_term.*

*Таким образом, оба пути (myogenesis и adipogenesis) апрегулированы в условиях 232\_term.*

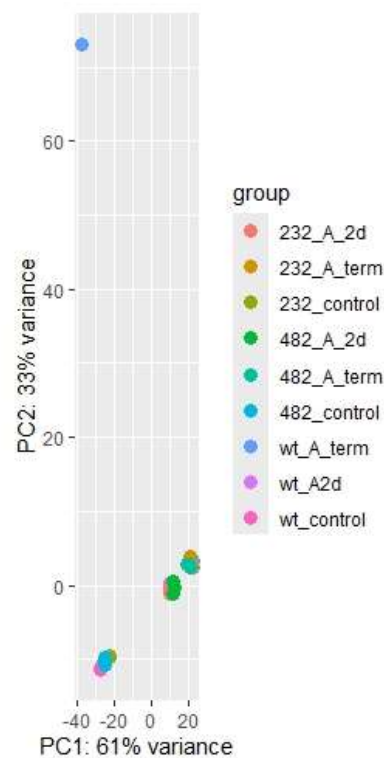
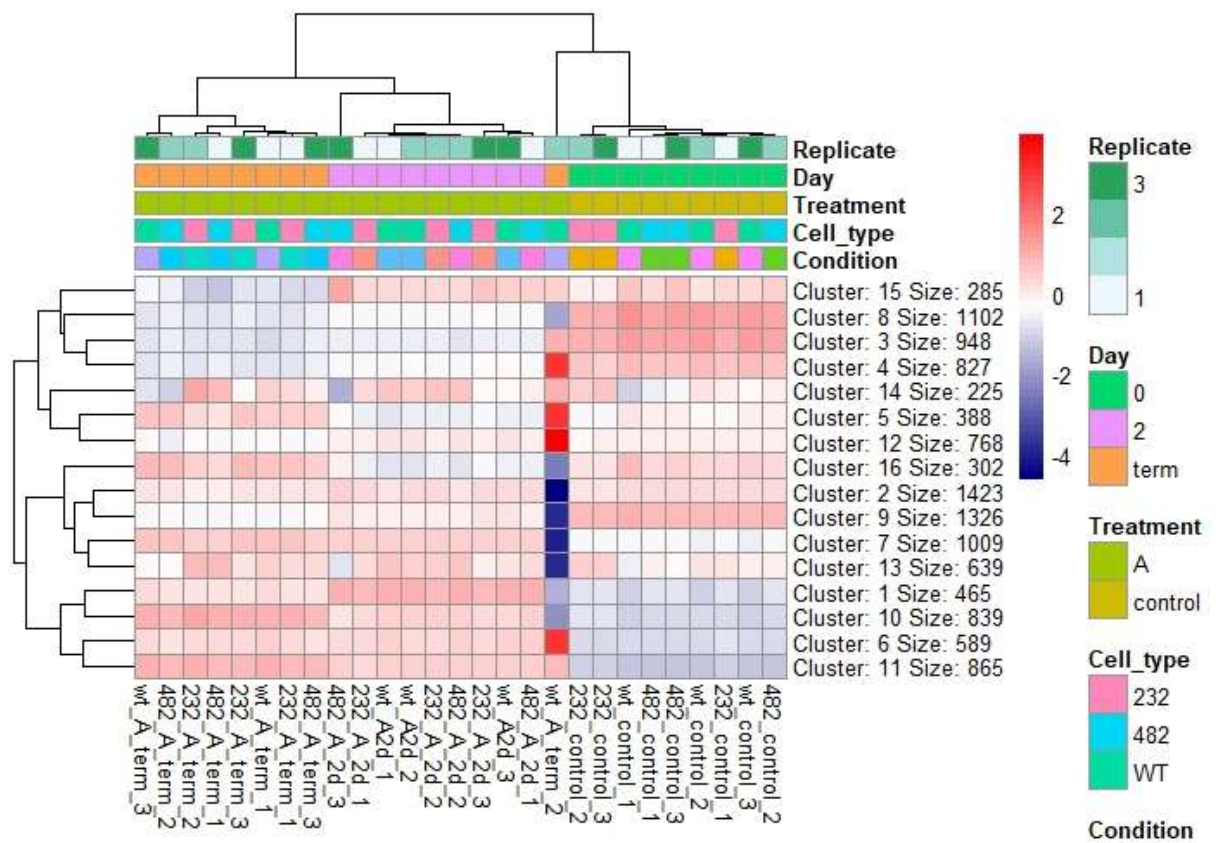


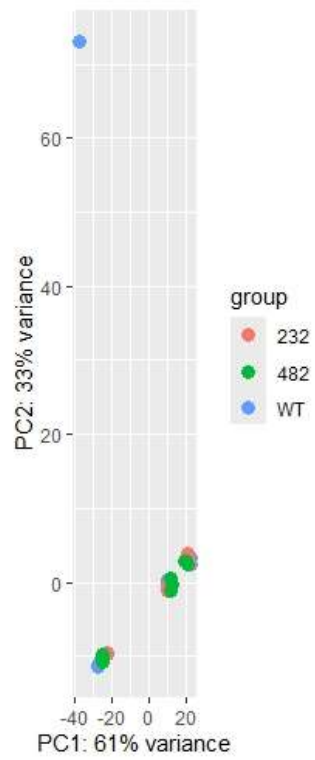
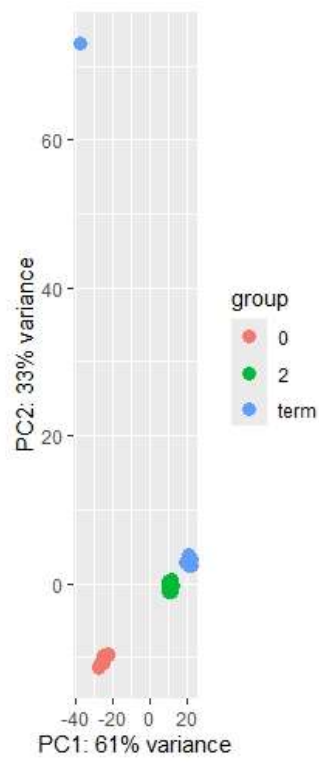


**Principal Components Plot**

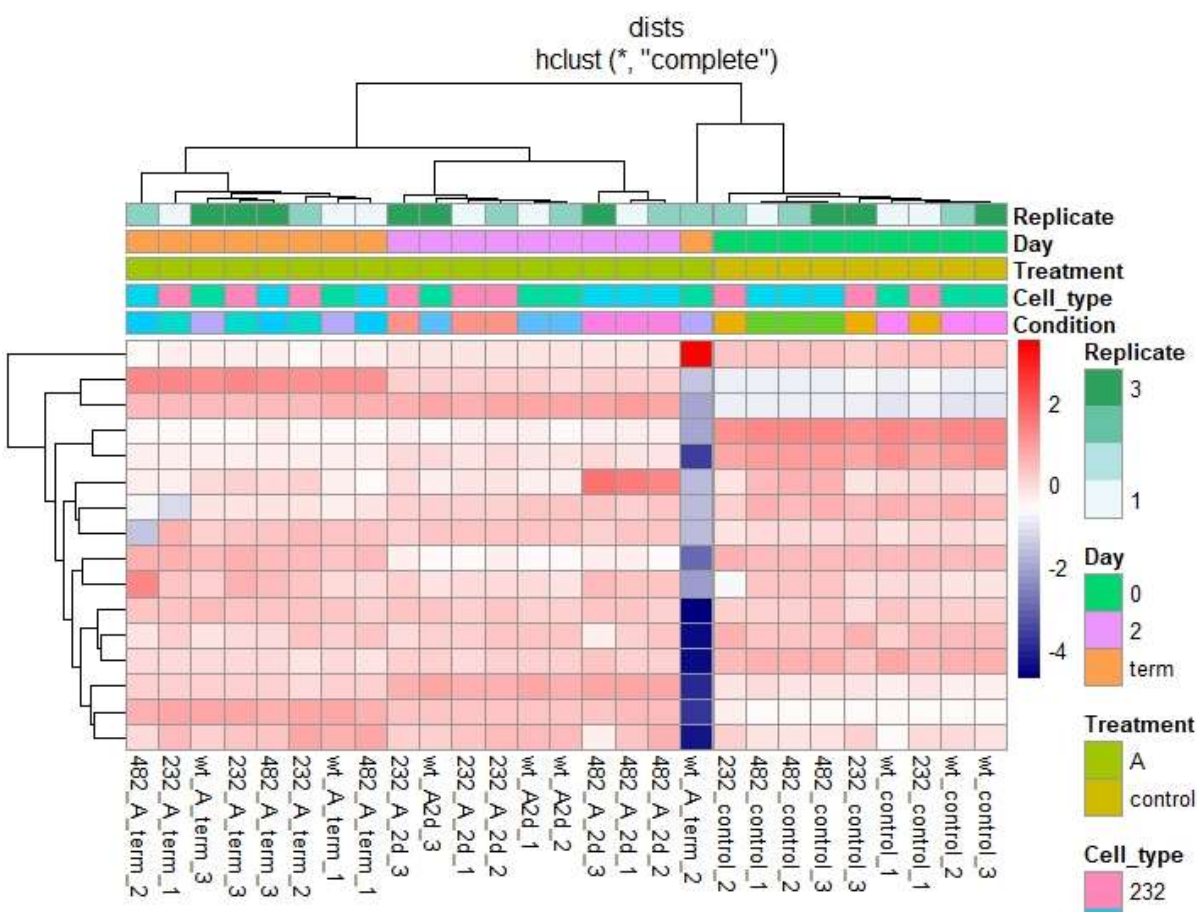
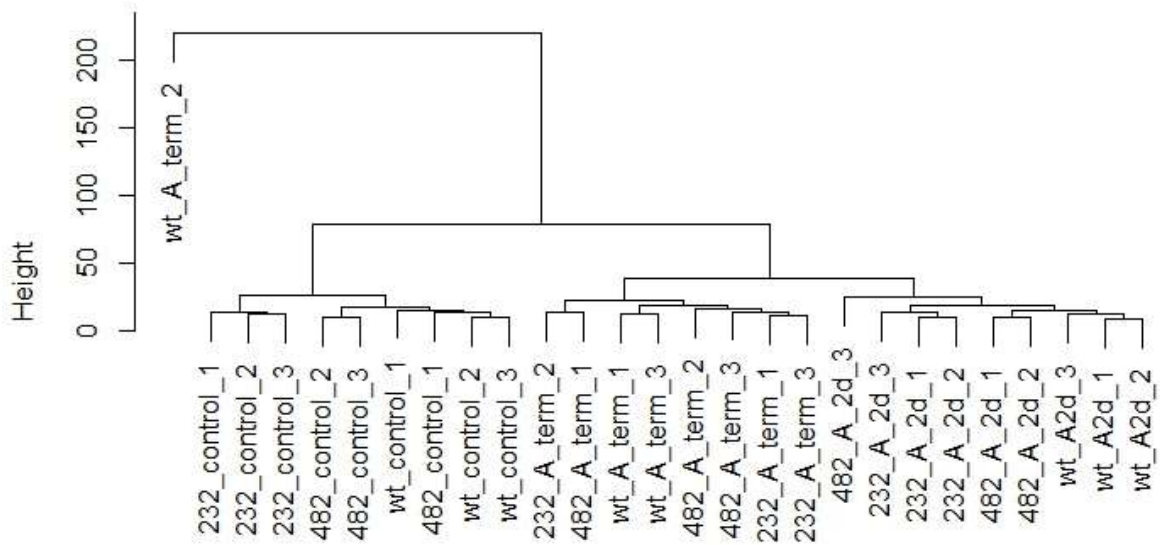




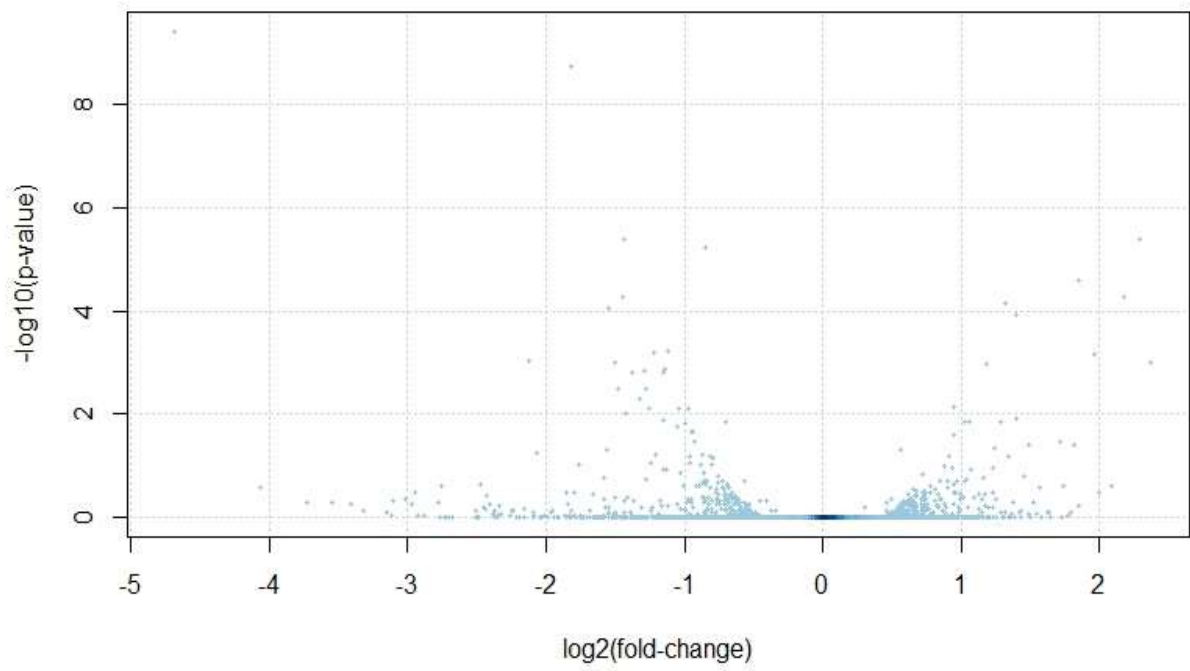




### Cluster Dendrogram



Volcano plot



Volcano plot

