

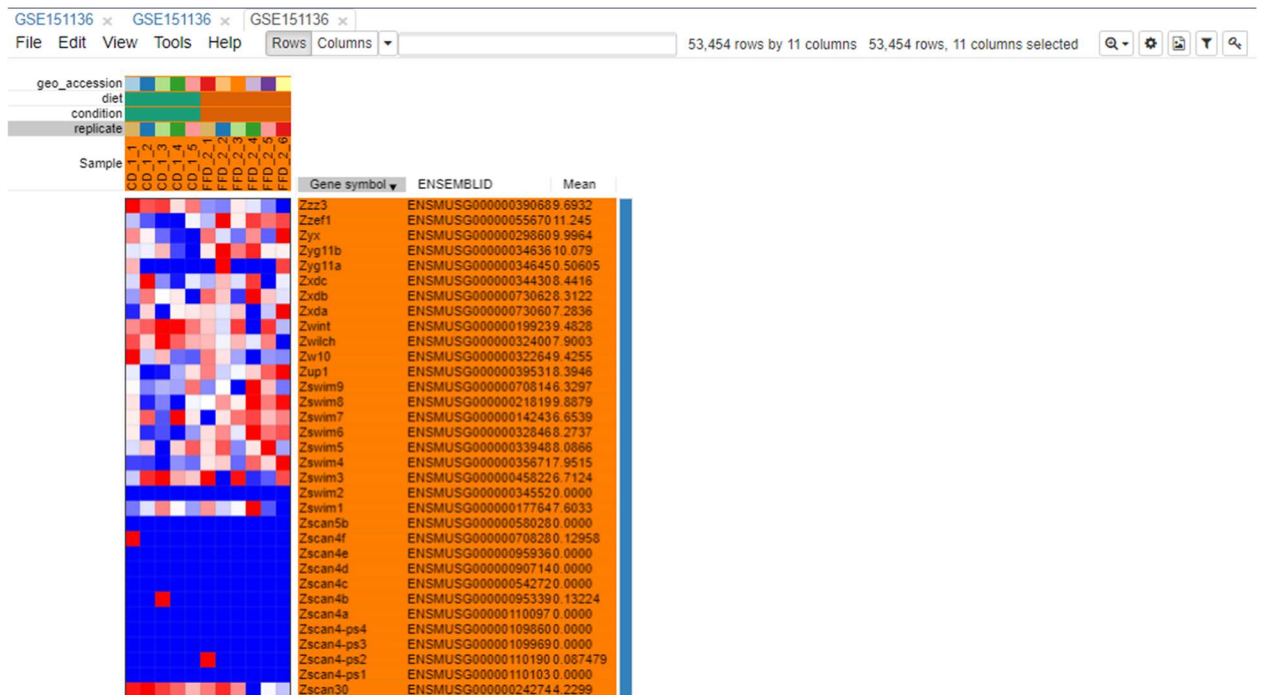
Домашнее задание №1

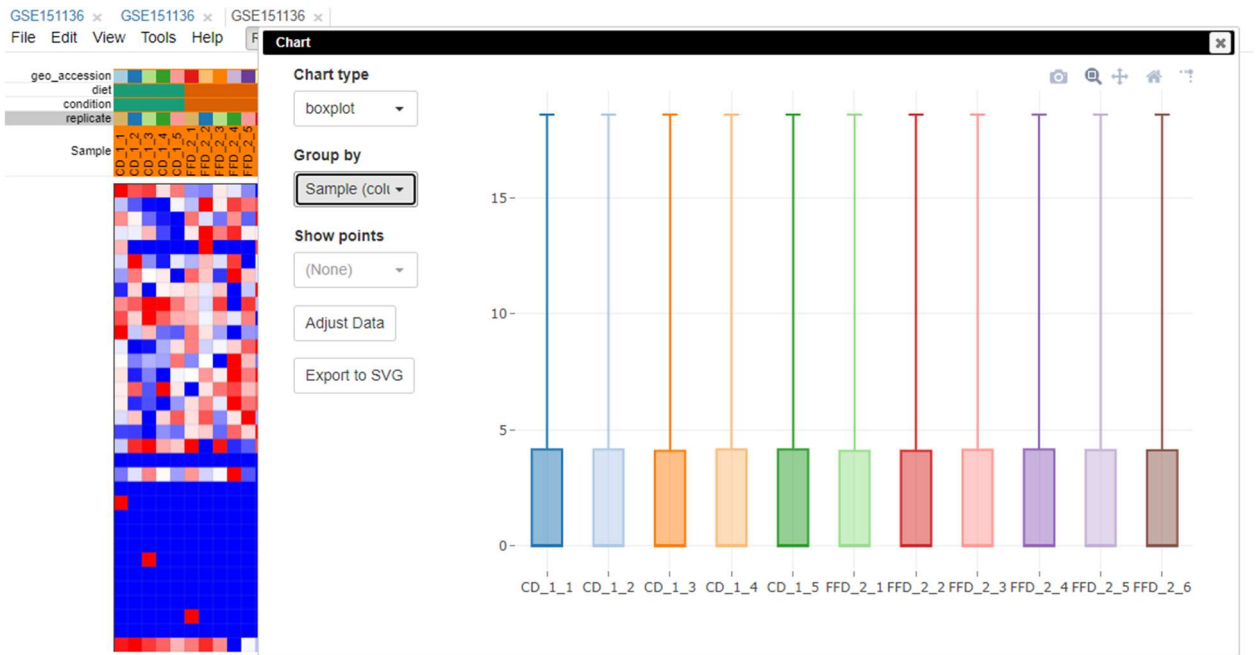
Анализ таблицы экспрессии в phantasus

1. Найти любые данные РНК-сека из GEO (raw counts) на интересующую вас тему. (можно microarray)
2. Сделать анализ экспрессии генов в Phantasus.
3. Сравнить результаты с исходной статьей.
4. Отчет работы в свободном формате в Word или Power Point (Ссылка на статью и данные, описание эксперимента, скрины основных этапов анализа, сравнение результатов со статьей, что совпало или не совпало с оригинальным экспериментом, выводы).

Статья:

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>
GSE151136





GSE151136 × GSE151136 × GSE151136 ×

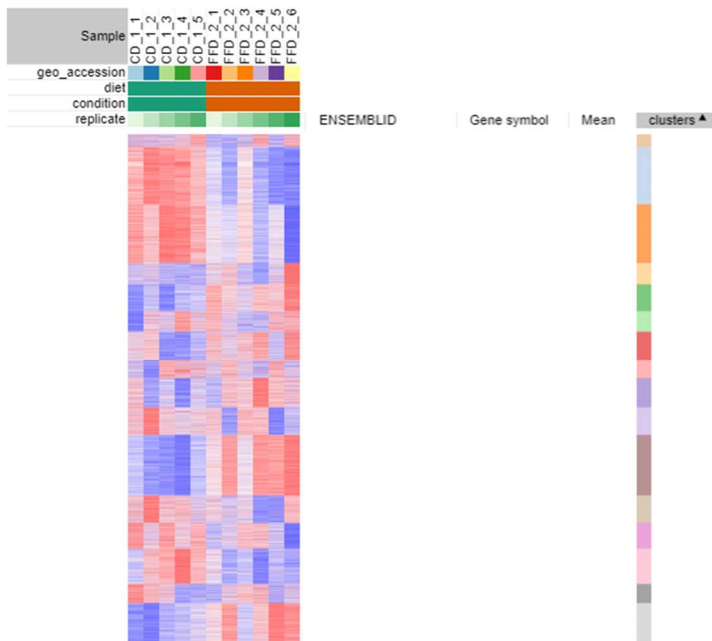
File Edit View Tools Help

Rows Columns 12,000/53,454 rows by 11 columns 0 rows, 11 columns selected

geo_accession diet condition replicate Sample

CD_1_1 CD_1_2 CD_1_3 CD_1_4 CD_1_5 FFD_2_1 FFD_2_2 FFD_2_3 FFD_2_4 FFD_2_5 FFD_2_6

Gene symbol	ENSEMBLID	Mean
Zzz3	ENSMUSG000000390689.6932	
Zzef1	ENSMUSG00000055670.11.245	
Zyx	ENSMUSG000000298609.9964	
Zyg11b	ENSMUSG00000034636.10.079	
Zxdc	ENSMUSG00000034430.8.4416	
Zxdb	ENSMUSG00000073062.8.3122	
Zxda	ENSMUSG00000073060.7.2836	
Zwint	ENSMUSG000000199239.4828	
Zwilch	ENSMUSG000000324007.9003	
Zwr10	ENSMUSG000000322649.4255	
Zup1	ENSMUSG000000395318.3946	
Zswim9	ENSMUSG000000708146.3297	
Zswim8	ENSMUSG000000218199.8879	
Zswim7	ENSMUSG0000000142436.6539	
Zswim6	ENSMUSG000000328468.2737	
Zswim5	ENSMUSG000000339468.0866	
Zswim4	ENSMUSG000000356717.9515	
Zswim3	ENSMUSG000000458226.7124	
Zswim1	ENSMUSG000000177647.6033	
Zscan29	ENSMUSG000000506198.9766	
Zscan26	ENSMUSG000000222289.3256	
Zscan25	ENSMUSG000000704207.3841	
Zscan22	ENSMUSG000000547157.5331	
Zscan21	ENSMUSG000000370178.5829	
Zscan20	ENSMUSG000000618946.4498	
Zscan12	ENSMUSG000000367216.6726	
Zrsr2	ENSMUSG000000313709.4026	
Zrsr1	ENSMUSG000000440687.1725	
Zranb3	ENSMUSG000000360865.7608	
Zranb2	ENSMUSG00000028180.10.332	
Zranb1	ENSMUSG000000309679.6490	
Zpr1	ENSMUSG000000320789.5778	
Znrf3	ENSMUSG000000419617.6433	



MSigDB Home

Human Collections

- About
- Browse
- Search
- Investigate
- Gene Families

Mouse Collections

- About
- Browse
- Search
- Investigate

Help

Compute Overlaps for Selected Genes

Converted 114 submitted identifiers into 114 NCBI (Entrez) genes. [Download symbol conversion report](#)



Collections	# Overlaps Shown	# Gene Sets in Collections	# Genes in Comparison (n)	# Genes in Universe (N)
GO:BP, MH	20	7846	114	42726

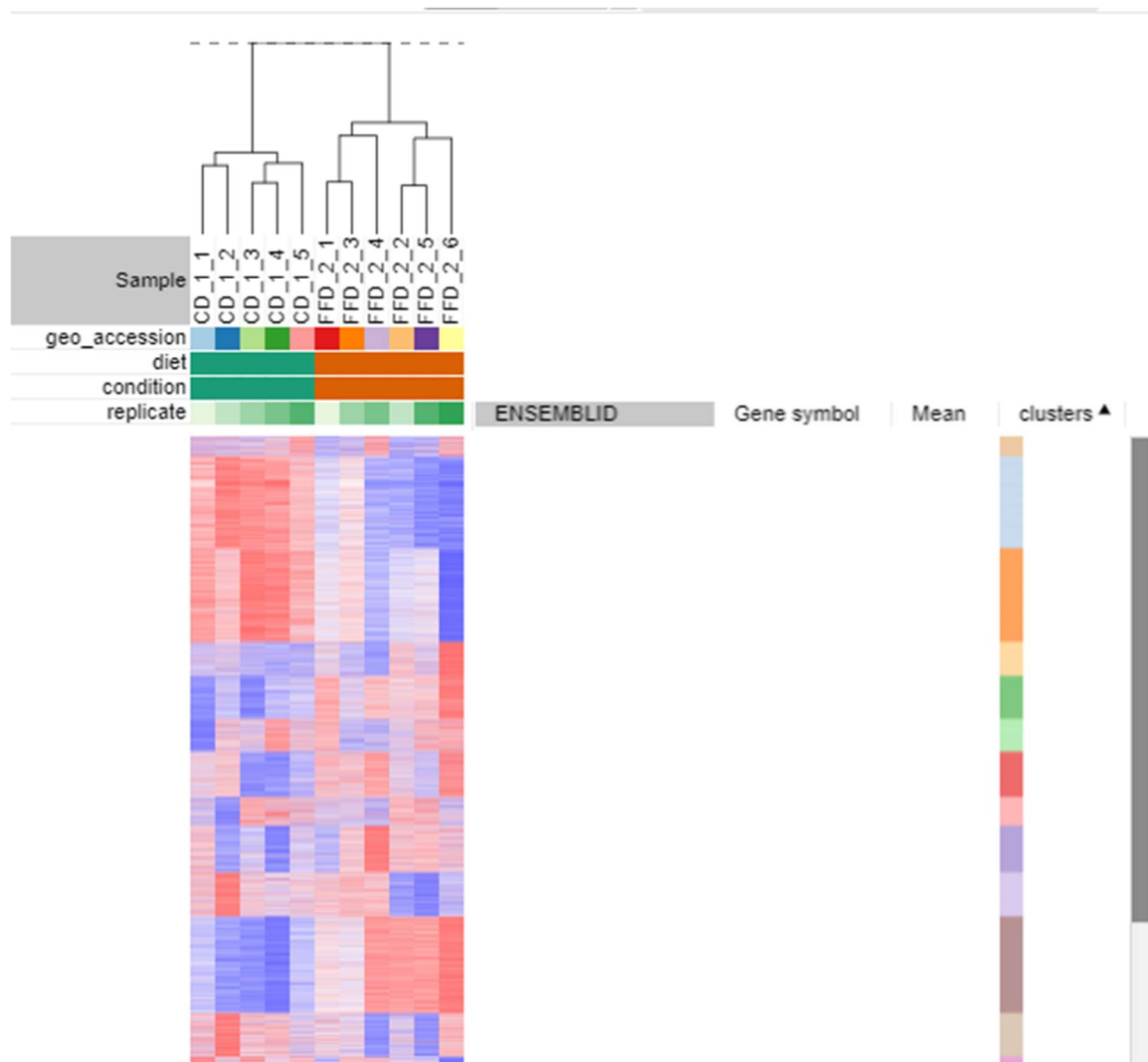
Click the gene set name to see the gene set page. Click the number of genes [in brackets] to download the list of genes.

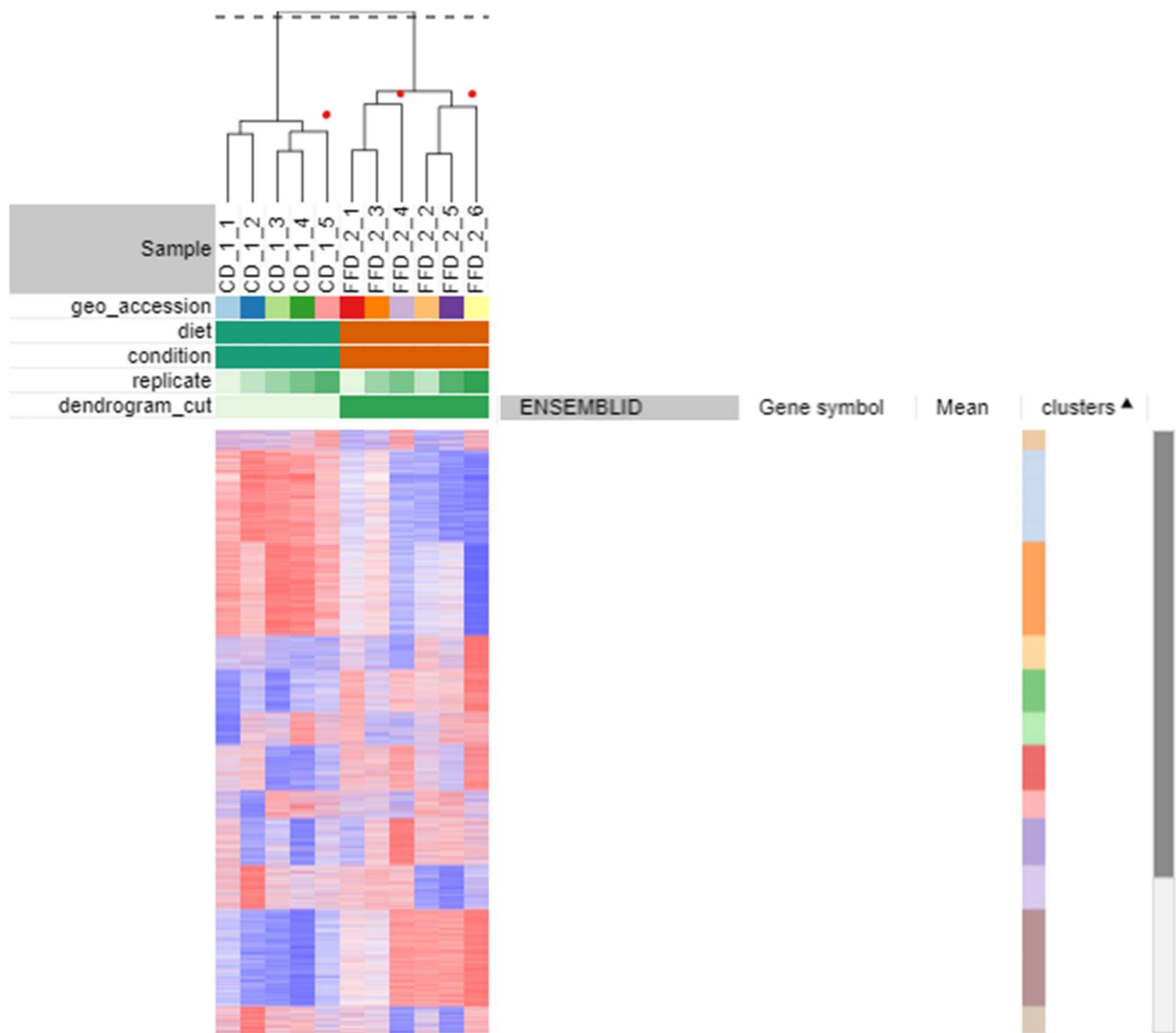
Color bar shading from light green to black, where lighter colors indicate more significant FDR q-values (< 0.05) and black indicates less significant FDR q-values (≥ 0.05).Save to: [Text](#) (as Tab separated values; *.tsv)

UC San Diego

BROAD
INSTITUTE

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p-value ?	FDRq-value ?
GOBP_TRANSMEMBRANE_TRANSPORT [1446]	The process in which a solute is transported across a lipid bilayer, from one side of a membrane to the other. [GOC:dph, GOC:jid]	37		2.56×10^{-26}	2.01×10^{-22}
GOBP_ORGANIC_ANION_TRANSPORT [438]	The directed movement of organic anions into, out of or within a cell, or between cells, by means of some agent such as a transporter	19		9.01×10^{-18}	3.53×10^{-14}





GSE151136 dataset x

File Edit View Tools Help Rows Columns

This figure displays a heatmap of gene expression data. A dendrogram at the top shows hierarchical clustering of samples. The samples are labeled as follows:

Sample	geo_accession	diet	condition	replicate	dendrogram_cut
CD_1_1	CD_1_1	CD_1_1	CD_1_1	CD_1_1	CD_1_1
CD_1_2	CD_1_2	CD_1_2	CD_1_2	CD_1_2	CD_1_2
CD_1_3	CD_1_3	CD_1_3	CD_1_3	CD_1_3	CD_1_3
CD_1_4	CD_1_4	CD_1_4	CD_1_4	CD_1_4	CD_1_4
CD_1_5	CD_1_5	CD_1_5	CD_1_5	CD_1_5	CD_1_5
FFD_2_1	FFD_2_1	FFD_2_1	FFD_2_1	FFD_2_1	FFD_2_1
FFD_2_2	FFD_2_2	FFD_2_2	FFD_2_2	FFD_2_2	FFD_2_2
FFD_2_3	FFD_2_3	FFD_2_3	FFD_2_3	FFD_2_3	FFD_2_3
FFD_2_4	FFD_2_4	FFD_2_4	FFD_2_4	FFD_2_4	FFD_2_4
FFD_2_5	FFD_2_5	FFD_2_5	FFD_2_5	FFD_2_5	FFD_2_5
FFD_2_6	FFD_2_6	FFD_2_6	FFD_2_6	FFD_2_6	FFD_2_6

The heatmap shows gene expression levels across these samples. A color scale on the right indicates the mean expression level, ranging from blue (low) to red (high). The scale is labeled 'clusters' with an upward arrow.

One-factor design Advanced design

Field condition dendrogram_cut diet genotype

Target level Search selected 1 of 2 ☒ CD_1 ☐ FFD_2

Reference level Search selected 1 of 2 ☐ CD_1 ☒ FFD_2

GSE151136 v.2

File Edit View Tools Help

Rows Columns

12,000 rows by 11 columns 0 rows, 0 columns selected

Q



Sample	CD_1_1	CD_1_2	CD_1_3	CD_1_4	CD_1_5	FFD_2_1	FFD_2_3	FFD_2_4	FFD_2_5	FFD_2_6
geo_accession										
diet										
condition										
replicate										
Comparison										

GSE151136 v.2

File Edit View Tools Help

Rows Columns

12,000 rows by 11 columns 0 rows, 0 columns selected

Q

Sample
geo_accession
diet
condition
replicate
Comparison

ENSEMBLID	Gene symbol	Mean	clusters	logFC	AveExpr	t	P.Value	adj.P.Val	B
ENSMUSG00000036437	Npy1r	5.1968	10.000	2.0632	5.1968	9.1705	0.0000012850	0.014039	5.8180
ENSMUSG00000009197	Hspa1a	9.8457	10.000	2.1844	9.8457	8.9591	0.0000016280	0.014039	5.5984
ENSMUSG00000044254	Pcsk9	9.6399	10.000	1.7846	9.6400	8.2792	0.0000035960	0.014278	4.8560
ENSMUSG00000006179	Prss16	5.6234	10.000	1.0375	5.6234	8.1651	0.0000041270	0.014278	4.7258
ENSMUSG00000029027	Dffb	7.3314	10.000	0.58949	7.3314	8.0709	0.0000046290	0.014618	4.6171
ENSMUSG00000048696	Mex3d	7.3497	10.000	0.75887	7.3497	7.9415	0.0000054270	0.015885	4.4658
ENSMUSG00000044763	Tmt10c	9.2834	10.000	0.46024	9.2834	7.8406	0.0000061520	0.016608	4.3465
ENSMUSG00000018623	Mmp7	5.9292	11.000	1.3711	5.9292	7.5856	0.0000084900	0.016608	4.0386
ENSMUSG00000020137	Thap2	8.6294	10.000	0.50929	8.6294	7.5612	0.0000087590	0.016608	4.0087
ENSMUSG0000000416	Cttnbp2	6.0428	10.000	2.2448	6.0428	7.4809	0.0000097120	0.016608	3.9096
ENSMUSG00000040434	Large2	7.5709	10.000	0.42630	7.5709	7.3793	0.0000110790	0.016608	3.7831
ENSMUSG00000020538	Srebf1	10.947	10.000	0.57303	10.947	7.3676	0.0000112490	0.016608	3.7684
ENSMUSG00000020482	Ccdc117	9.4597	10.000	0.59570	9.4597	7.3533	0.0000114610	0.016608	3.7504
ENSMUSG00000030498	Gas2	5.7514	10.000	0.47919	5.7514	7.1424	0.0000151360	0.017983	3.4823
ENSMUSG00000026558	Uck2	10.493	10.000	0.65483	10.493	7.1336	0.0000153140	0.018017	3.4709
ENSMUSG00000009087	Hspa1b	11.808	10.000	1.5808	11.808	7.0808	0.0000164340	0.018396	3.4027
ENSMUSG00000002167	Tcf3	9.6428	10.000	0.64364	9.6428	6.9575	0.0000194020	0.018863	3.2419
ENSMUSG00000003500	Impdh1	8.6688	10.000	0.78111	8.6688	6.9523	0.0000195370	0.018863	3.2352
ENSMUSG00000001739	Cldn15	11.088	10.000	0.94633	11.088	6.9312	0.0000201060	0.018863	3.2074
ENSMUSG00000006728	Cdk4	10.990	10.000	1.0389	10.990	6.9092	0.0000207160	0.018863	3.1784
ENSMUSG000000025264	Tsr2	8.1180	11.000	0.58231	8.1181	6.9058	0.0000208110	0.018863	3.1739
ENSMUSG00000007670	Khrrp	11.240	10.000	0.46137	11.241	6.9012	0.0000209440	0.018863	3.1678
ENSMUSG000000057134	Ado	9.2970	10.000	0.48314	9.2970	6.8708	0.0000218290	0.019261	3.1276
ENSMUSG00000036285	Noa1	9.4389	10.000	0.66603	9.4389	6.8499	0.0000224610	0.019529	3.0999
ENSMUSG00000046865	Fbl	10.406	11.000	0.71635	10.406	6.7868	0.0000244920	0.019858	3.0158
ENSMUSG000000037071	Scd1	8.2140	11.000	1.7764	8.2140	6.7709	0.0000250350	0.019895	2.9945
ENSMUSG00000007029	Vars	11.697	10.000	0.66983	11.698	6.6921	0.0000279200	0.020447	2.8883
ENSMUSG000000086290	Snhg12	7.2567	10.000	0.69496	7.2567	6.6881	0.0000280760	0.020447	2.8829
ENSMUSG000000037216	Lipt1	5.9959	10.000	0.67868	5.9959	6.6443	0.0000298400	0.020447	2.8236
ENSMUSG00000016624	Phf21b	6.9629	10.000	0.63310	6.9629	6.6420	0.0000299340	0.020447	2.8205
ENSMUSG00000024735	Prpf19	11.238	10.000	0.61450	11.238	6.6324	0.0000303420	0.020447	2.8073
ENSMUSG000000037278	Tmem97	10.590	10.000	0.75393	10.590	6.6296	0.0000304600	0.020447	2.8035

GSE151136 v.2

GSE151136 v.2

File Edit View Tools Help

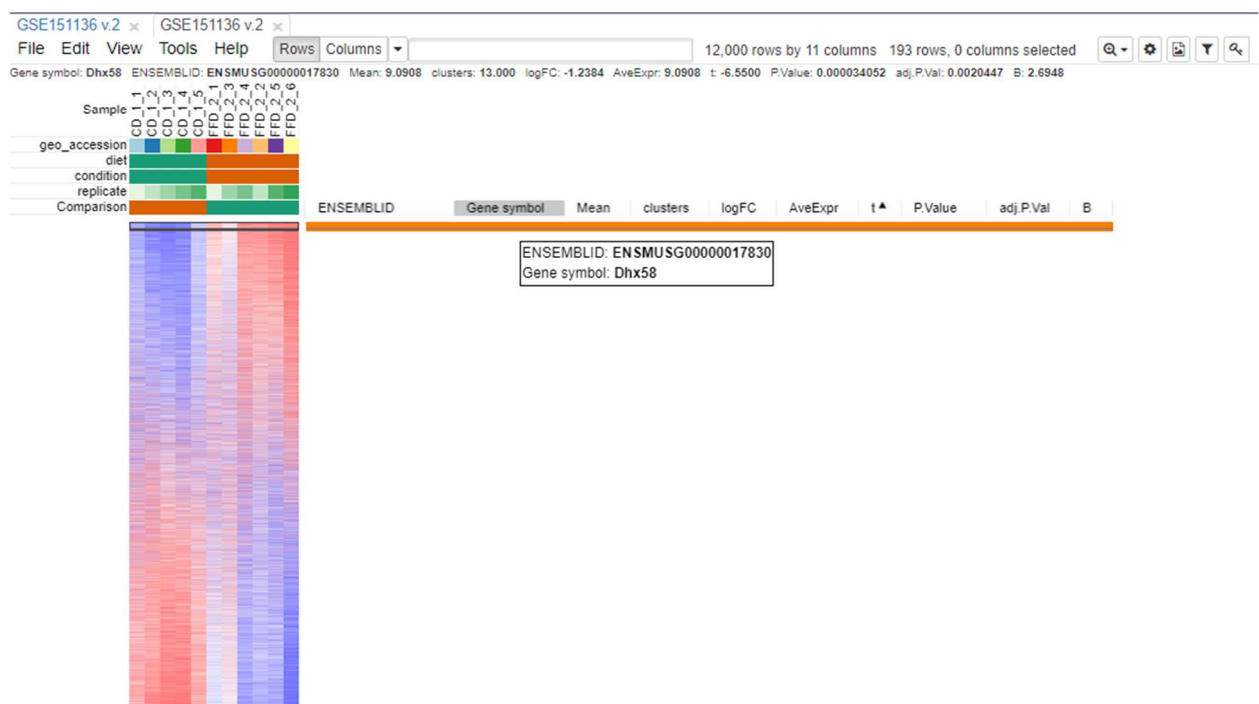
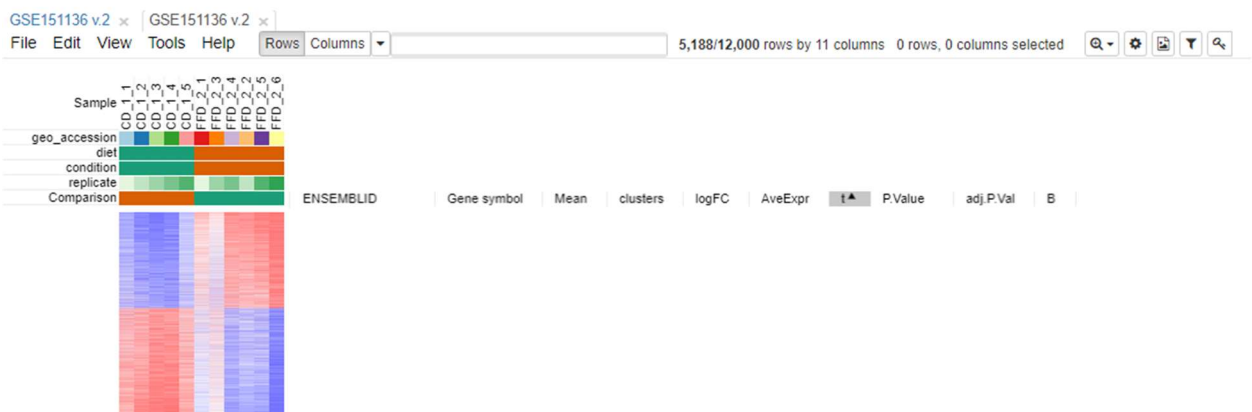
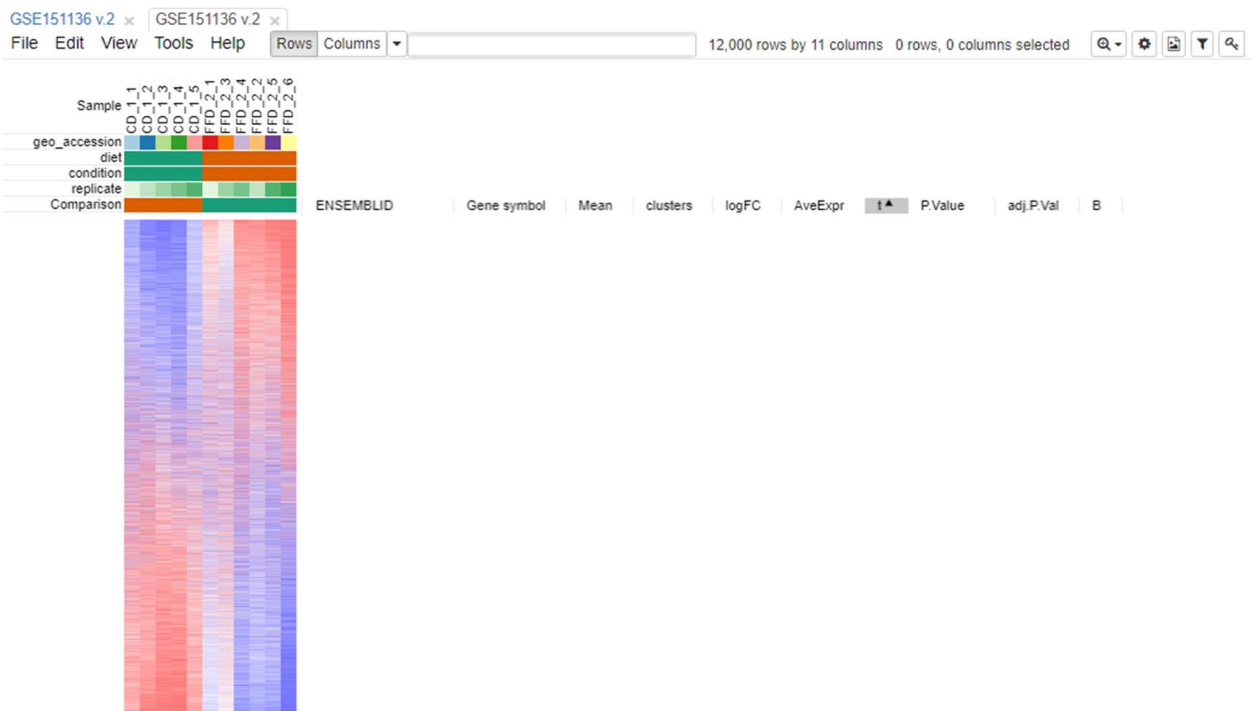
Rows Columns

12,000 rows by 11 columns 0 rows, 0 columns selected

Q

Sample
geo_accession
diet
condition
replicate
Comparison

ENSEMBLID	Gene symbol	Mean	clusters	logFC	AveExpr	t	P.Value	adj.P.Val	B
-----------	-------------	------	----------	-------	---------	---	---------	-----------	---



- Browse
- Search
- Investigate
- Gene Families
- Mouse Collections
- About
- Browse
- Search
- Investigate
- Help



Converted 193 submitted identifiers into 175 NCBI (Entrez) genes. [Download symbol conversion report](#)

Collections	# Overlaps Shown	# Gene Sets in Collections	# Genes in Comparison (n)	# Genes in Universe (N)
GO:BP, H	20	7697	175	42722

Click the gene set name to see the gene set page. Click the number of genes [in brackets] to download the list of genes.

Color bar shading from light green to black, where lighter colors indicate more significant FDR q-values (< 0.05) and black indicates less significant FDR q-values (≥ 0.05).

Save to: Text (as Tab separated values; *.tsv)

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p-value	FDRq-value
GOBP_BIOLOGICAL_PROCESS_INVOLVED_IN_IN_INTERSPECIES_INTERACTION_BETWEEN_ORGANISMS [1759]	Any process evolved to enable an interaction with an organism of a different species. [GOC:cc]	47		2.83×10^{-25}	2.18×10^{-21}
GOBP_DEFENSE_RESPONSE [1855]	Reactions, triggered in response to the presence of a foreign body or the occurrence of an injury, which result in restriction of damage to the organism attacked or prevention/recovery from the infection caused by the attack. [GOC:go_curators]	45		1.75×10^{-22}	4.66×10^{-19}
GOBP_DEFENSE_RESPONSE_TO_OTHER_ORGANISMS [1235]	Reactions triggered in response to the presence	38		1.82×10^{-22}	4.66×10^{-19}



GSE151136 v.2 × GSE151136 v.2 ×

File Edit View Tools Help Rows Columns 12,000 rows by 11 columns 317 rows, 0 columns selected

Sample
geo_accession
diet
condition
replicate
Comparison

ENSEMBLID

Submit to Enrichr

Please select rows.

Column with gene symbols
Gene symbol

Ambiguous genes strategy
omit

Preview data
Rpl7 Ad11 Zfp931 Rps24 Rps15 Taf6l Elf3l Gsr Raver1 Ccne2 Anapc5 Rpl27a Cenpv Gg

Result will open in a new window automatically.

Cancel Submit



Login | Register

Transcription Pathways Ontologies Diseases/Drugs Cell Types Misc Legacy Crowd

Description GSE151136 v.2 (317 genes)

ChEA 2022

MYC 28411283 ChIP-Seq MDA231-LM2-4175

MYC 19030024 ChIP-ChIP MESC's Mouse

MYC 18358816 ChIP-ChIP MESC's Mouse

MYC 18555785 ChIP-Seq MESC's Mouse

MYC 19079543 ChIP-ChIP MESC's Mouse

ENCODE and ChEA Consensus TFs from

MYC ENCODE

MYC ChEA

MAX ENCODE

TAF1 ENCODE

TAF7 ENCODE

ARCHS4 TFs Coexp

EIF3K human tf ARCHS4 coexpression

PHB2 human tf ARCHS4 coexpression

RFXANK human tf ARCHS4 coexpression

TP53 human tf ARCHS4 coexpression

CPSF4 human tf ARCHS4 coexpression

TF Perturbations Followed by Expression

MYCN OE MOUSE GSE6077 CREEDSID GENE

EZH2 SHRNA PROE HUMAN GSE59089 RNAS

GATA1 KD MOUSE GSE2527 CREEDSID GENE

RBM10 OE HEK293 HUMAN GSE44976 RNAS

MYC OE U2OS HUMAN GSE59819 RNASEQ D

TRRUST Transcription Factors 2019

MYC mouse

NFYA mouse

TCF4 mouse

PRDM1 human

HOXA1 human

FANTOM6 lncRNA KD DEGs

LINC00339-ASO_G0218510_AD_06-DEGs Do

ERVK3-1-ASO_G0142396_AD_07-DEGs Down

LINC00657-ASO_G0260032_AD_06-DEGs Do

AC092295.7-ASO_G0233527_02-DEGs Down

LINC00517-ASO_G0259091_01-DEGs Up

IncHUB lncRNA Co-Expression

LINC02067

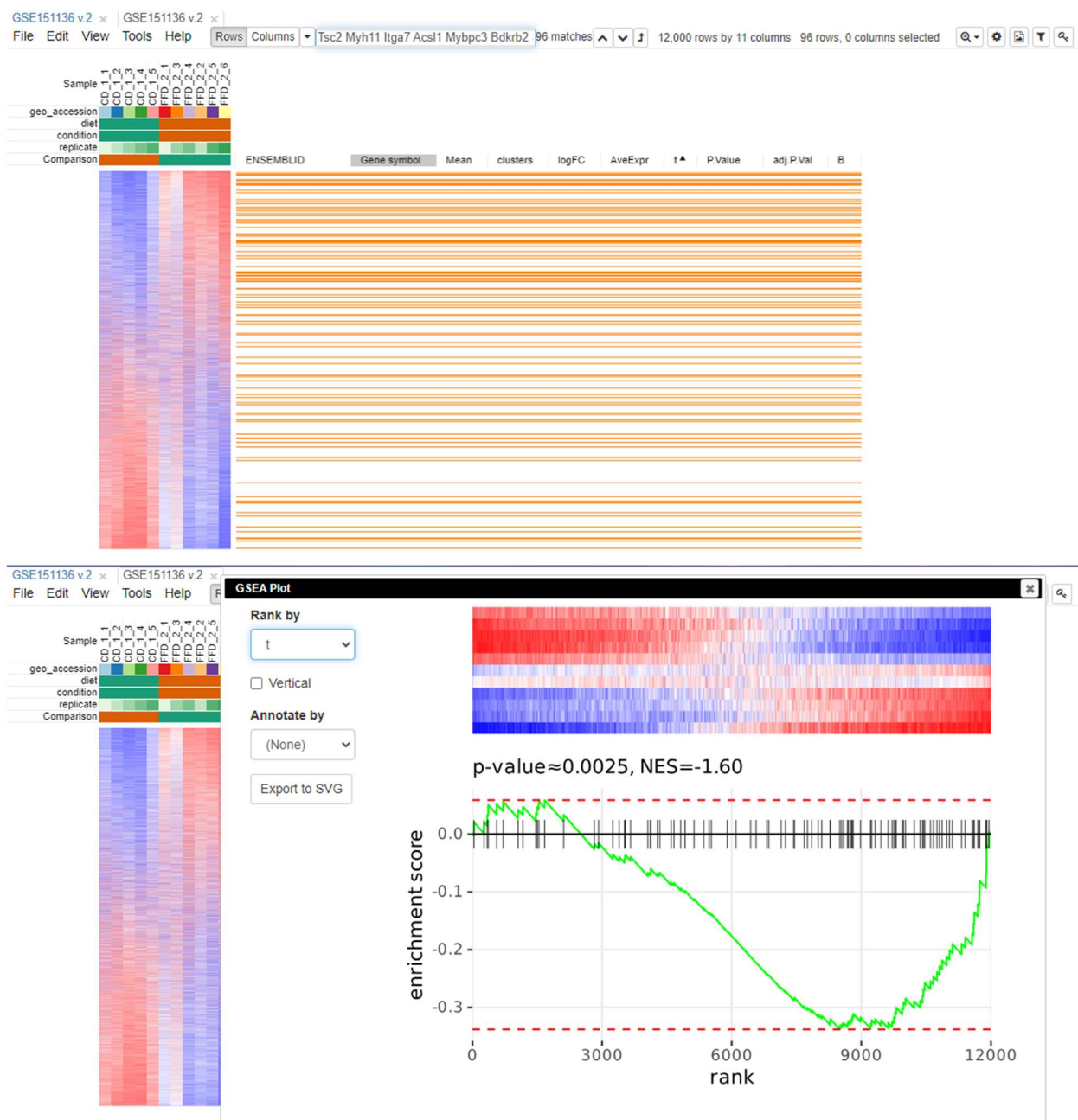
Enrichr Submissions TF-Gene Cooccurrence

PHB2

TRANSFAC and JASPAR PWMs

NRF1 (mouse)

<https://maayanlab.cloud/Enrichr/enrich?dataset=27de9e6622e94f82fc58a64e2019866b>



Из статьи видно, что анализ дифференциальной экспрессии генов эпителиальных клеток толстой кишки показал множественные эффекты пищевой целлюлозы на профили транскрипции. Кроме того, отмечается, что целлюлоза вызывала отчетливую кластеризацию при сравнении сигнатурных генов разных типов эпителиальных клеток.

В анализе Phantastus использовались следующие фильтры для отбора ДЭГ:
 Минимальное количество экспрессируемых генов;
 \log_2 fold change > 1 или < -1 ;
 скорректированное p -значение < 0.05 .

В результате было выявлено 317 дифференциально экспрессированных генов, которые удовлетворяли заданным критериям отбора.

При сравнении с результатами анализа в Phantasus, видим, что некоторые генные наборы (gene sets) из Gene Ontology Biological Process (GOBP) и Hallmark совпадают с эффектами, описанными в статье. Например, генные наборы, связанные с биосинтезом белков, трансляцией, обработкой РНК и сборкой рибосом, имеют значительное пересечение с результатами исследования статьи.

Исходя из приведённых данных Pathways в GSEA и Enrichr:

(GO:BP_CYTOPLASMIC_TRANSLATION)

Обычно ап-регулирован в условиях, где повышена потребность в синтезе белков, таких как рост клеток или восстановление после повреждений.

(GO:BP_PEPTIDE_BIOSYNTHETIC_PROCESS)

Часто ап-регулирован в условиях активного роста или стресса, где требуется синтез новых белков.

(GO:BP_AMIDE_BIOSYNTHETIC_PROCESS)

Может быть ап-регулирован в условиях, требующих повышенного синтеза протеинов и амидных соединений.

(GO:BP_PEPTIDE_METABOLIC_PROCESS)

Может варьироваться в зависимости от условий, но часто ап-регулирован при усиленном клеточном метаболизме.

(GO:BP_AMIDE_METABOLIC_PROCESS)

Как и метаболизм пептидов, часто ап-регулирован в условиях интенсивного клеточного метаболизма.

На графике GSEA plot молекулярный путь из базы данных MSigDb (H_myogenesis_mus).

Путь H_myogenesis_mus преимущественно активирован в условиях CD (target level) по сравнению с FFD (reference level), поскольку большинство генов имеют положительные значения log2FoldChange.

Таким образом, результаты анализа в Phantasus подтверждают некоторые из основных выводов статьи относительно влияния пищевой

целлюлозы на программы транскрипции в эпителиальных клетках толстой кишки во время воспаления.