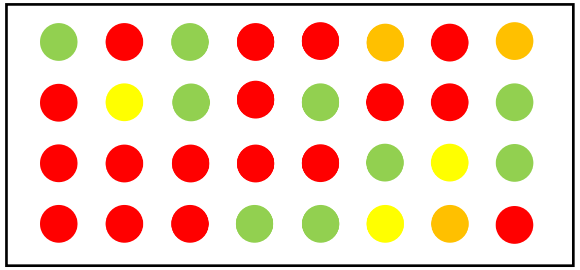
生物学的背景に基づいたマイクロアレイデータの解釈

フィルジェン株式会社　バイオインフォマティクス部
(support@filgen.jp)

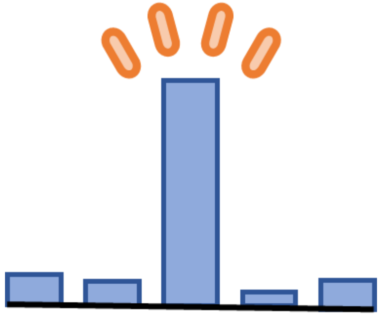
マイクロアレイ解析のワークフロー



画像の取得

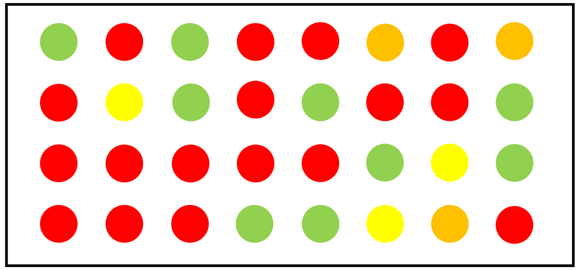
	Sample 1	Sample 2	Sample 3
Gene A	0	0	10
Gene B	5	5	6
Gene C	5	1	7

発現定量



データの可視化、統計解析

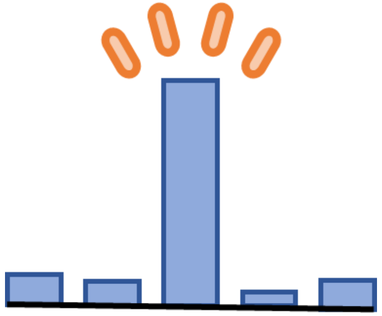
マイクロアレイ解析のワークフロー



画像の取得

	Sample 1	Sample 2	Sample 3
Gene A	0	0	10
Gene B	5	5	6
Gene C	5	1	7

発現定量



解析者の技量に依拠するところが大きい

データの可視化、統計解析

本セミナーで使用するプラットフォーム

Omics Playground

- 定量済みデータをアップロードするだけで、
  データの解釈に必要な100以上のプロットを生成可能

- 生物学で用いられる様々な統計手法をカバー
  （Deseq2, EdgeR, limmaなど）

- マイクロアレイデータの他、
  トランスクリプトミクスデータ、プロテオミクスデータ
  もアップロード可能

- ユーザーどうしで解析結果の共有が可能

データさえあれば、誰でも重要な遺伝子を特定できる

入力ファイル

	Sample A	Sample B	Sample C
Gene 1	5.4	4.1	0.1
Gene 2	0.1	0.3	8.9
Gene3	6.7	5.2	4.4

①expression.csv：
サンプルごとの発現定量データ

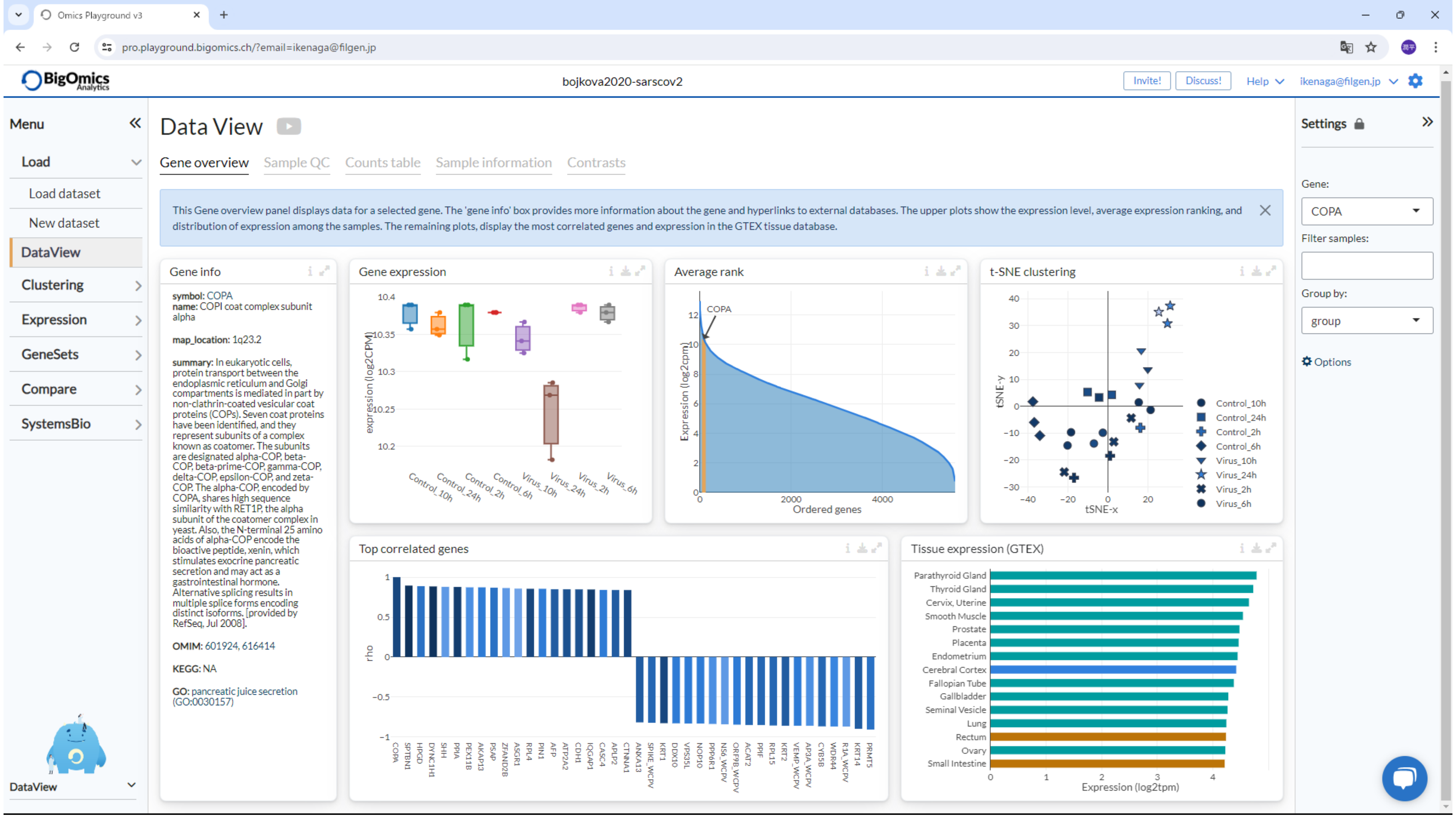
	Sex	Country	Dominant hand
Sample A	male	Japan	right
Sample B	male	USA	left
Sample C	female	Japan	right

②sample.csv：
各サンプルの属性を記述するデータ

	Left vs Right	male vs female
Sample A	right	male
Sample B	left	male
Sample C	right	female

③contrast.csv：
比較解析の内容を記述するデータ

DataView



This Gene overview panel displays data for a selected gene. The 'gene info' box provides more information about the gene and hyperlinks to external databases. The upper plots show the expression level, average expression ranking, and distribution of expression among the samples. The remaining plots, display the most correlated genes and expression in the GTEX tissue database.

Gene info

symbol: COPA
name: COPI coat complex subunit alpha

map_location: 1q23.2

summary: In eukaryotic cells, protein transport between the endoplasmic reticulum and Golgi compartments is mediated in part by non-clathrin-coated vesicular coat proteins (COPs). Seven coat proteins have been identified, and they represent subunits of a complex known as coatomer. The subunits are designated alpha-COP, beta-COP, beta-prime-COP, gamma-COP, delta-COP, epsilon-COP, and zeta-COP. The alpha-COP, encoded by COPA, shares high sequence similarity with RET1P, the alpha subunit of the coatomer complex in yeast. Also, the N-terminal 25 amino acids of alpha-COP encode the bioactive peptide, xenin, which stimulates exocrine pancreatic secretion and may act as a gastrointestinal hormone. Alternative splicing results in multiple splice forms encoding distinct isoforms. [provided by RefSeq, Jul 2008].

OMIM: 601924, 616414

KEGG: NA

GO: pancreatic juice secretion (GO:0030157)

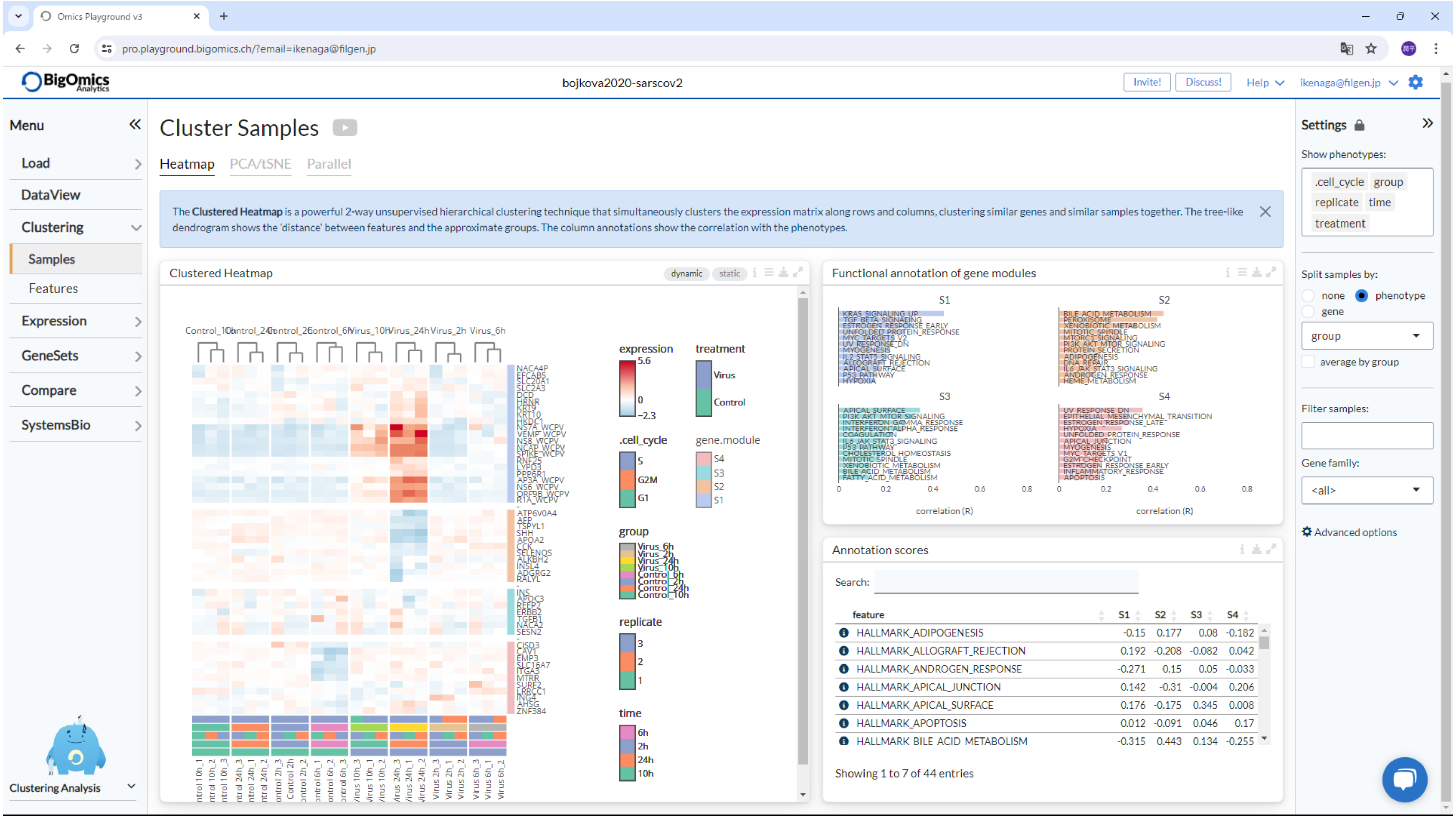
Settings

Gene: COPA

Filter samples:

Group by: group

Clustering



The **Clustered Heatmap** is a powerful 2-way unsupervised hierarchical clustering technique that simultaneously clusters the expression matrix along rows and columns, clustering similar genes and similar samples together. The tree-like dendrogram shows the 'distance' between features and the approximate groups. The column annotations show the correlation with the phenotypes.

Show phenotypes: .cell_cycle group replicate time treatment

Split samples by:
○ none ● phenotype
○ gene

group

☐ average by group

Filter samples:

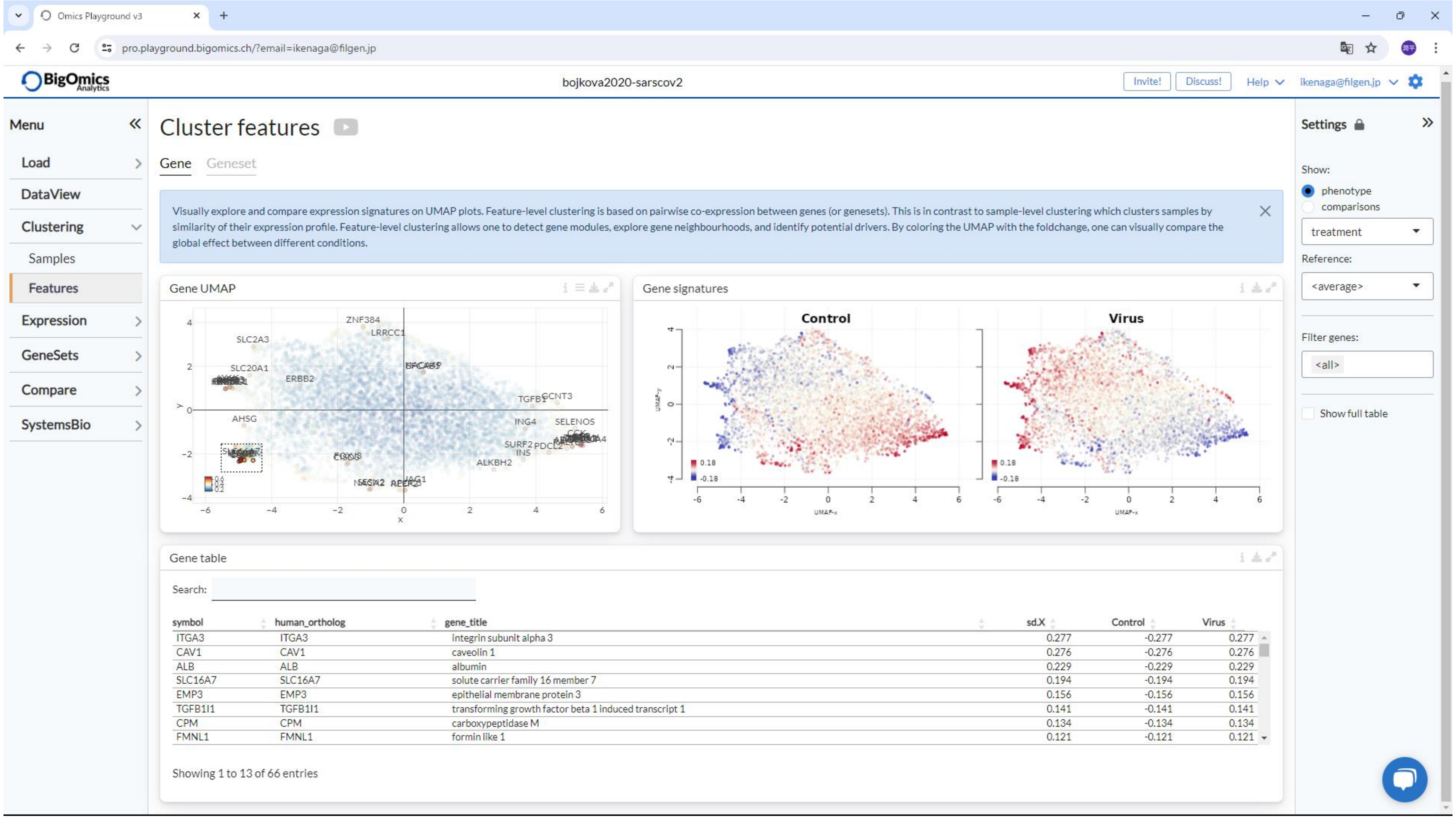
Gene family: <all>

Search:

feature	S1	S2	S3	S4
HALLMARK_ADIPOGENESIS	-0.15	0.177	0.08	-0.182
HALLMARK_ALLOGRAFT_REJECTION	0.192	-0.208	-0.082	0.042
HALLMARK_ANDROGEN_RESPONSE	-0.271	0.15	0.05	-0.033
HALLMARK_APICAL_JUNCTION	0.142	-0.31	-0.004	0.206
HALLMARK_APICAL_SURFACE	0.176	-0.175	0.345	0.008
HALLMARK_APOPTOSIS	0.012	-0.091	0.046	0.17
HALLMARK_BILE_ACID_METABOLISM	-0.315	0.443	0.134	-0.255

Showing 1 to 7 of 44 entries

Clustering



Cluster features

Visually explore and compare expression signatures on UMAP plots. Feature-level clustering is based on pairwise co-expression between genes (or genesets). This is in contrast to sample-level clustering which clusters samples by similarity of their expression profile. Feature-level clustering allows one to detect gene modules, explore gene neighbourhoods, and identify potential drivers. By coloring the UMAP with the foldchange, one can visually compare the global effect between different conditions.

Show:
● phenotype
○ comparisons

treatment

Reference:
<average>

Filter genes:
<all>

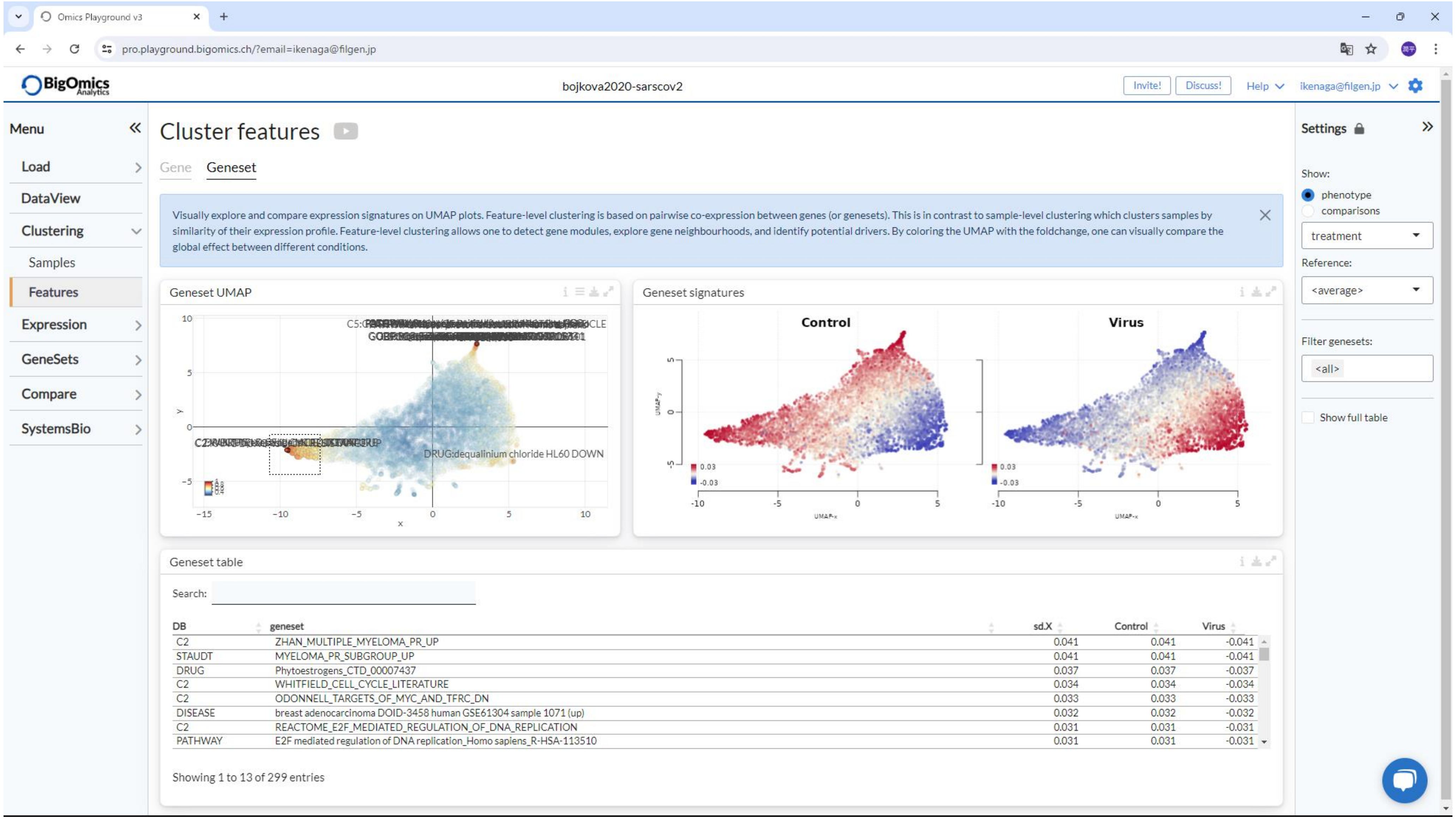
☐ Show full table

Search:

symbol	human_ortholog	gene_title	sd.X	Control	Virus
ITGA3	ITGA3	integrin subunit alpha 3	0.277	-0.277	0.277
CAV1	CAV1	caveolin 1	0.276	-0.276	0.276
ALB	ALB	albumin	0.229	-0.229	0.229
SLC16A7	SLC16A7	solute carrier family 16 member 7	0.194	-0.194	0.194
EMP3	EMP3	epithelial membrane protein 3	0.156	-0.156	0.156
TGFB1I1	TGFB1I1	transforming growth factor beta 1 induced transcript 1	0.141	-0.141	0.141
CPM	CPM	carboxypeptidase M	0.134	-0.134	0.134
FMNL1	FMNL1	formin like 1	0.121	-0.121	0.121

Showing 1 to 13 of 66 entries

Clustering



Visually explore and compare expression signatures on UMAP plots. Feature-level clustering is based on pairwise co-expression between genes (or genesets). This is in contrast to sample-level clustering which clusters samples by similarity of their expression profile. Feature-level clustering allows one to detect gene modules, explore gene neighbourhoods, and identify potential drivers. By coloring the UMAP with the foldchange, one can visually compare the global effect between different conditions.

Show:
● phenotype
○ comparisons
treatment

Reference:
<average>

Filter genesets:
<all>

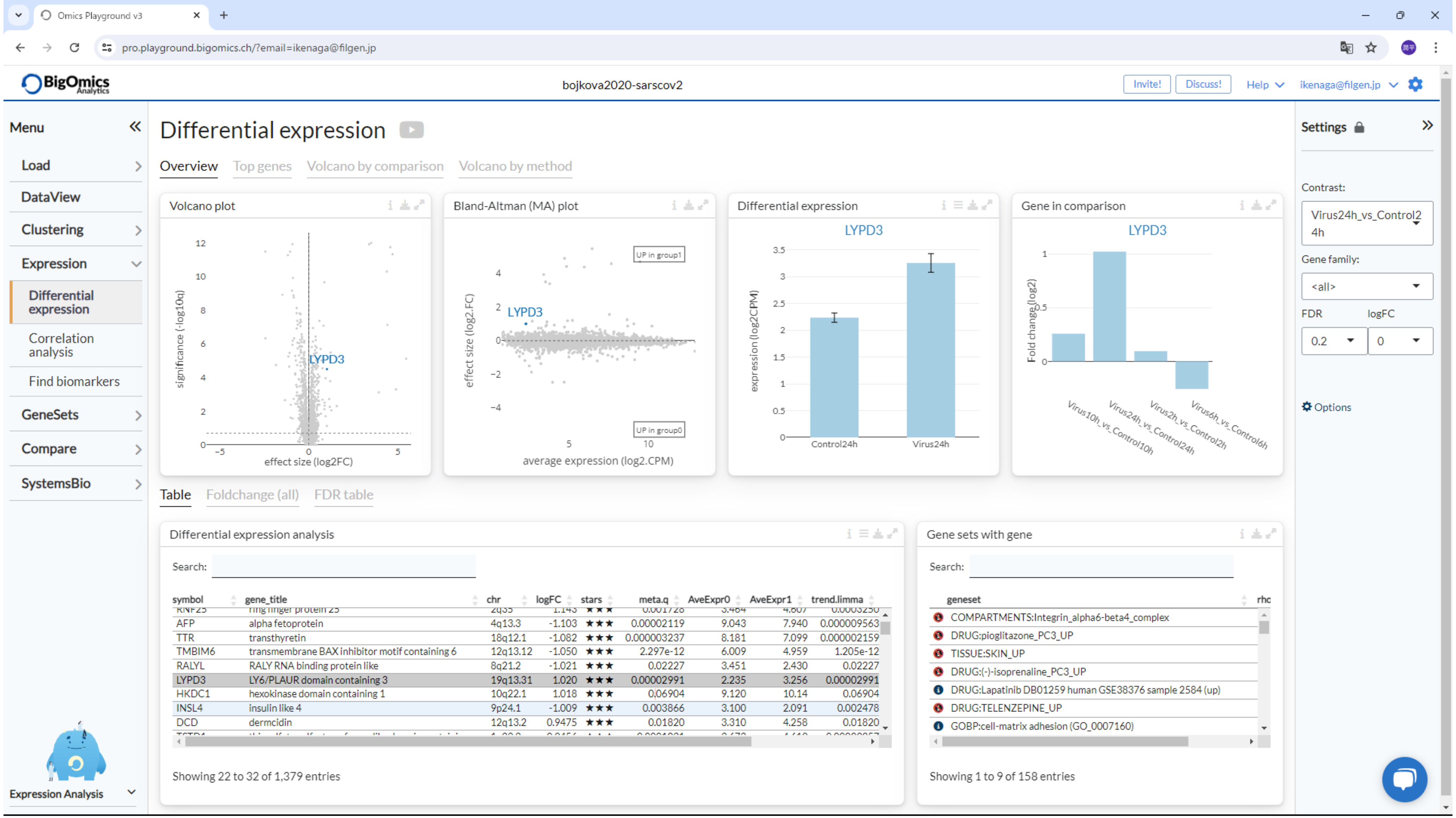
☐ Show full table

Search:

DB	geneset	sd.X	Control	Virus
C2	ZHAN_MULTIPLE_MYELOMA_PR_UP	0.041	0.041	-0.041
STAUDT	MYELOMA_PR_SUBGROUP_UP	0.041	0.041	-0.041
DRUG	Phytoestrogens_CTD_00007437	0.037	0.037	-0.037
C2	WHITFIELD_CELL_CYCLE_LITERATURE	0.034	0.034	-0.034
C2	ODONNELL_TARGETS_OF_MYC_AND_TFRC_DN	0.033	0.033	-0.033
DISEASE	breast adenocarcinoma DOID-3458 human GSE61304 sample 1071 (up)	0.032	0.032	-0.032
C2	REACTOME_E2F_MEDIATED_REGULATION_OF_DNA_REPLICATION	0.031	0.031	-0.031
PATHWAY	E2F mediated regulation of DNA replication_Homo sapiens_R-HSA-113510	0.031	0.031	-0.031

Showing 1 to 13 of 299 entries

Expression



Differential expression

Contrast: Virus24h_vs_Control24h

Gene family: <all>

FDR: 0.2

logFC: 0

Differential expression analysis

Search:

symbol	gene_title	chr	logFC	stars	meta.q	AveExpr0	AveExpr1	trend.limma
RNF25	ring finger protein 25	2q35	1.143	★★★	0.001728	3.464	4.607	0.0003250
AFP	alpha fetoprotein	4q13.3	-1.103	★★★	0.00002119	9.043	7.940	0.000009563
TTR	transthyretin	18q12.1	-1.082	★★★	0.000003237	8.181	7.099	0.000002159
TMBIM6	transmembrane BAX inhibitor motif containing 6	12q13.12	-1.050	★★★	2.297e-12	6.009	4.959	1.205e-12
RALYL	RALY RNA binding protein like	8q21.2	-1.021	★★★	0.02227	3.451	2.430	0.02227
LYPD3	LY6/PLAUR domain containing 3	19q13.31	1.020	★★★	0.00002991	2.235	3.256	0.00002991
HKDC1	hexokinase domain containing 1	10q22.1	1.018	★★★	0.06904	9.120	10.14	0.06904
INSL4	insulin like 4	9p24.1	-1.009	★★★	0.003866	3.100	2.091	0.002478
DCD	dermcidin	12q13.2	0.9475	★★★	0.01820	3.310	4.258	0.01820

Showing 22 to 32 of 1,379 entries

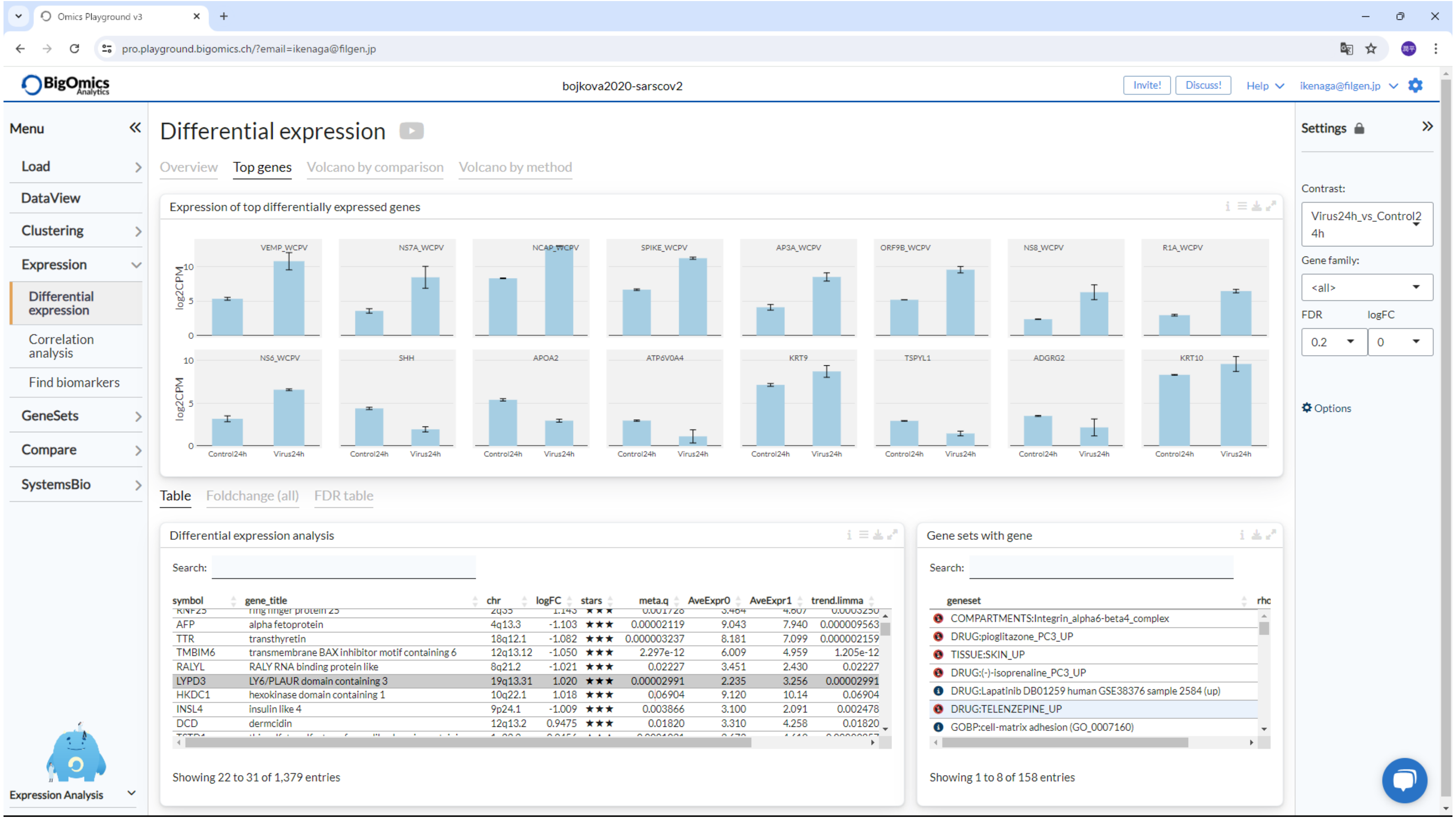
Gene sets with gene

Search:

geneset
COMPARTMENTS:Integrin_alpha6-beta4_complex
DRUG:pioglitazone_PC3_UP
TISSUE:SKIN_UP
DRUG:(-)-isoprenaline_PC3_UP
DRUG:Lapatinib DB01259 human GSE38376 sample 2584 (up)
DRUG:TELENZEPINE_UP
GOBP:cell-matrix adhesion (GO_0007160)

Showing 1 to 9 of 158 entries

Expression



Contrast: Virus24h_vs_Control24h

Gene family: <all>

FDR 0.2 logFC 0

Differential expression analysis

Search:

symbol	gene_title	chr	logFC	stars	meta.q	AveExpr0	AveExpr1	trend.limma
AFP	alpha fetoprotein	4q13.3	-1.103	★★★	0.00002119	9.043	7.940	0.000009563
TTR	transthyretin	18q12.1	-1.082	★★★	0.000003237	8.181	7.099	0.000002159
TMBIM6	transmembrane BAX inhibitor motif containing 6	12q13.12	-1.050	★★★	2.297e-12	6.009	4.959	1.205e-12
RALYL	RALY RNA binding protein like	8q21.2	-1.021	★★★	0.02227	3.451	2.430	0.02227
LYPD3	LY6/PLAUR domain containing 3	19q13.31	1.020	★★★	0.00002991	2.235	3.256	0.00002991
HKDC1	hexokinase domain containing 1	10q22.1	1.018	★★★	0.06904	9.120	10.14	0.06904
INSL4	insulin like 4	9p24.1	-1.009	★★★	0.003866	3.100	2.091	0.002478
DCD	dermcidin	12q13.2	0.9475	★★★	0.01820	3.310	4.258	0.01820

Showing 22 to 31 of 1,379 entries

Gene sets with gene

Search:

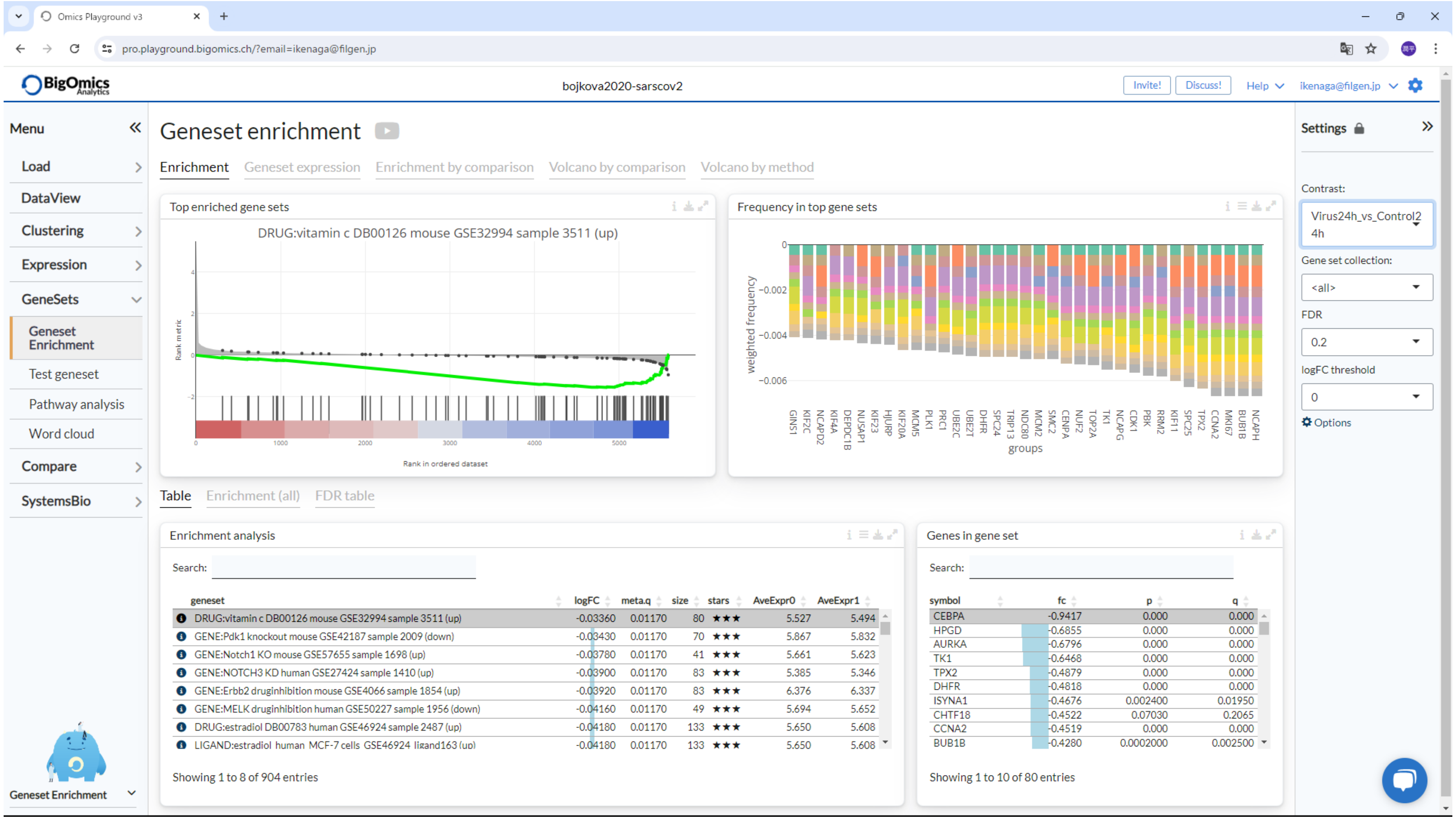
geneset
COMPARTMENTS:Integrin_alpha6-beta4_complex
DRUG:pioglitazone_PC3_UP
TISSUE:SKIN_UP
DRUG:(-)-isoprenaline_PC3_UP
DRUG:Lapatinib DB01259 human GSE38376 sample 2584 (up)
DRUG:TELENZEPINE_UP
GOBP:cell-matrix adhesion (GO_0007160)

Showing 1 to 8 of 158 entries

Expression


Omics Playground v3
pro.playground.bigomics.ch/?email=ikenaga@filgen.jp
BigOmics Analytics
bojkova2020-sarscov2
Invite!
Discuss!
Help
ikenaga@filgen.jp
Menu
Load
DataView
Clustering
Expression
Differential expression
Correlation analysis
Find biomarkers
GeneSets
Compare
SystemsBio
Biomarker Board
Biomarker Selection
Feature selection
Feature-set ranking
Variable importance
splsda
xgboost.lin
xgboost
randomForest
glmnet.a1
glmnet.a0
ftest
correlation
Biomarker expression
NCAP_WCPV
CAV1
CCDC22
CCK
GCNT3
SFXN4
ITGA3
ORF9B_WCPV
expression
Heatmap
Decision tree
NCAP_WCPV
Virus_24h (n = 100, err = 0.0%)
Virus_10h (n = 100, err = 0.0%)
CAV1
Control_6h (n = 100, err = 0.0%)
CCDC22
Control_10h (n = 100, err = 0.0%)
CCK
Virus_6h (n = 100, err = 0.0%)
GCNT3
Control_24h (n = 100, err = 0.0%)
SFXN4
Virus_2h (n = 100, err = 0.0%)
Settings
Predicted target:
group
Filter samples:
Feature set:
<all>
Compute

Genesets



Contrast: Virus24h_vs_Control24h

Gene set collection: <all>

FDR: 0.2

logFC threshold: 0

Enrichment analysis

Search:

geneset	logFC	meta.q	size	stars	AveExpr0	AveExpr1
DRUG:vitamin c DB00126 mouse GSE32994 sample 3511 (up)	-0.03360	0.01170	80	★★★	5.527	5.494
GENE:Pdk1 knockout mouse GSE42187 sample 2009 (down)	-0.03430	0.01170	70	★★★	5.867	5.832
GENE:Notch1 KO mouse GSE57655 sample 1698 (up)	-0.03780	0.01170	41	★★★	5.661	5.623
GENE:NOTCH3 KD human GSE27424 sample 1410 (up)	-0.03900	0.01170	83	★★★	5.385	5.346
GENE:Erbb2 druginhibition mouse GSE4066 sample 1854 (up)	-0.03920	0.01170	83	★★★	6.376	6.337
GENE:MELK druginhibition human GSE50227 sample 1956 (down)	-0.04160	0.01170	49	★★★	5.694	5.652
DRUG:estradiol DB00783 human GSE46924 sample 2487 (up)	-0.04180	0.01170	133	★★★	5.650	5.608
LIGAND:estradiol human MCF-7 cells GSE46924 ligand163 (up)	-0.04180	0.01170	133	★★★	5.650	5.608

Showing 1 to 8 of 904 entries

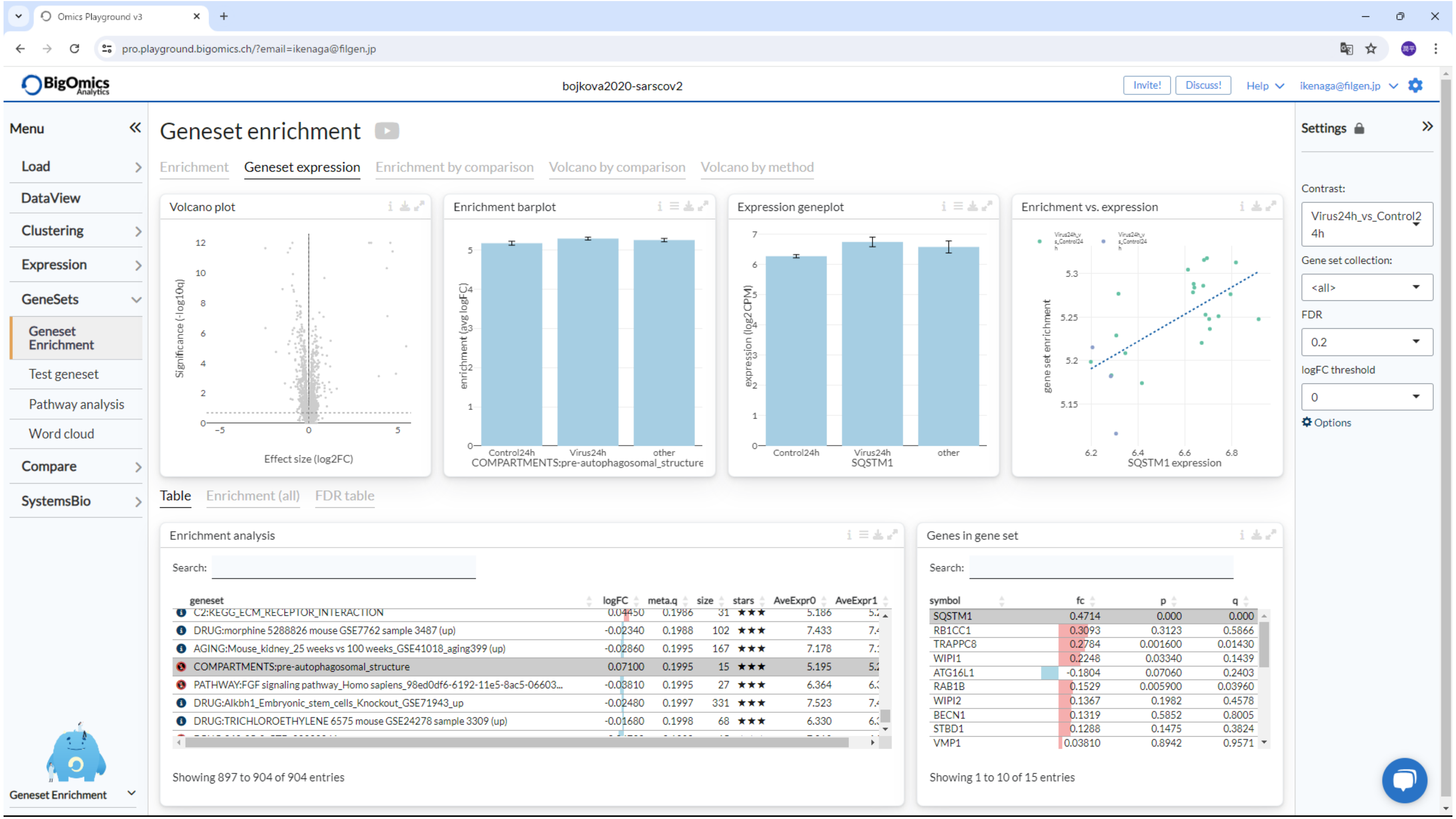
Genes in gene set

Search:

symbol	fc	p	q
CEBPA	-0.9417	0.000	0.000
HPGD	-0.6855	0.000	0.000
AURKA	-0.6796	0.000	0.000
TK1	-0.6468	0.000	0.000
TPX2	-0.4879	0.000	0.000
DHFR	-0.4818	0.000	0.000
ISYNA1	-0.4676	0.002400	0.01950
CHTF18	-0.4522	0.07030	0.2065
CCNA2	-0.4519	0.000	0.000
BUB1B	-0.4280	0.0002000	0.002500

Showing 1 to 10 of 80 entries

Genesets



Contrast: Virus24h_vs_Control24h

Gene set collection: <all>

FDR: 0.2

logFC threshold: 0

Enrichment analysis

Search:

geneset	logFC	meta.q	size	stars	AveExpr0	AveExpr1
C2:KEGG_ECM_RECEPTOR_INTERACTION	0.04450	0.1986	31	★★★	5.186	5.2
DRUG:morphine 5288826 mouse GSE7762 sample 3487 (up)	-0.02340	0.1988	102	★★★	7.433	7.4
AGING:Mouse_kidney_25 weeks vs 100 weeks_GSE41018_aging399 (up)	-0.02860	0.1995	167	★★★	7.178	7.1
COMPARTMENTS:pre-autophagosomal_structure	0.07100	0.1995	15	★★★	5.195	5.2
PATHWAY:FGF signaling pathway_Homo sapiens_98ed0df6-6192-11e5-8ac5-06603...	-0.03810	0.1995	27	★★★	6.364	6.3
DRUG:Alkbh1_Embryonic_stem_cells_Knockout_GSE71943_up	-0.02480	0.1997	331	★★★	7.523	7.4
DRUG:TRICHLOROETHYLENE 6575 mouse GSE24278 sample 3309 (up)	-0.01680	0.1998	68	★★★	6.330	6.3

Showing 897 to 904 of 904 entries

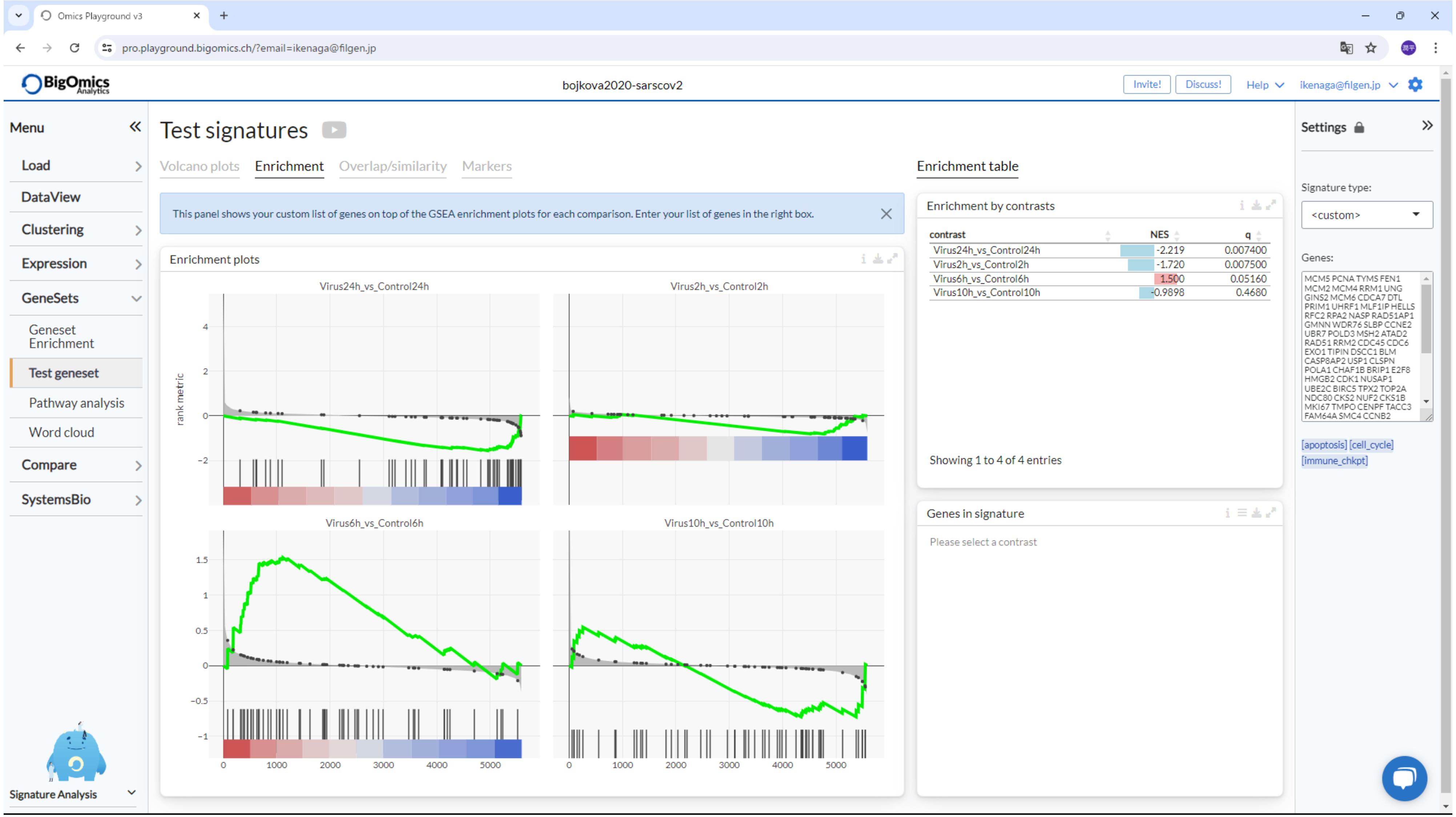
Genes in gene set

Search:

symbol	fc	p	q
SQSTM1	0.4714	0.000	0.000
RB1CC1	0.3093	0.3123	0.5866
TRAPPC8	0.2784	0.001600	0.01430
WIPI1	0.2248	0.03340	0.1439
ATG16L1	-0.1804	0.07060	0.2403
RAB1B	0.1529	0.005900	0.03960
WIPI2	0.1367	0.1982	0.4578
BECN1	0.1319	0.5852	0.8005
STBD1	0.1288	0.1475	0.3824
VMP1	0.03810	0.8942	0.9571

Showing 1 to 10 of 15 entries

Genesets



This panel shows your custom list of genes on top of the GSEA enrichment plots for each comparison. Enter your list of genes in the right box.

Enrichment by contrasts

contrast	NES	q
Virus24h_vs_Control24h	-2.219	0.007400
Virus2h_vs_Control2h	-1.720	0.007500
Virus6h_vs_Control6h	1.500	0.05160
Virus10h_vs_Control10h	-0.9898	0.4680

Showing 1 to 4 of 4 entries

Genes in signature

Please select a contrast

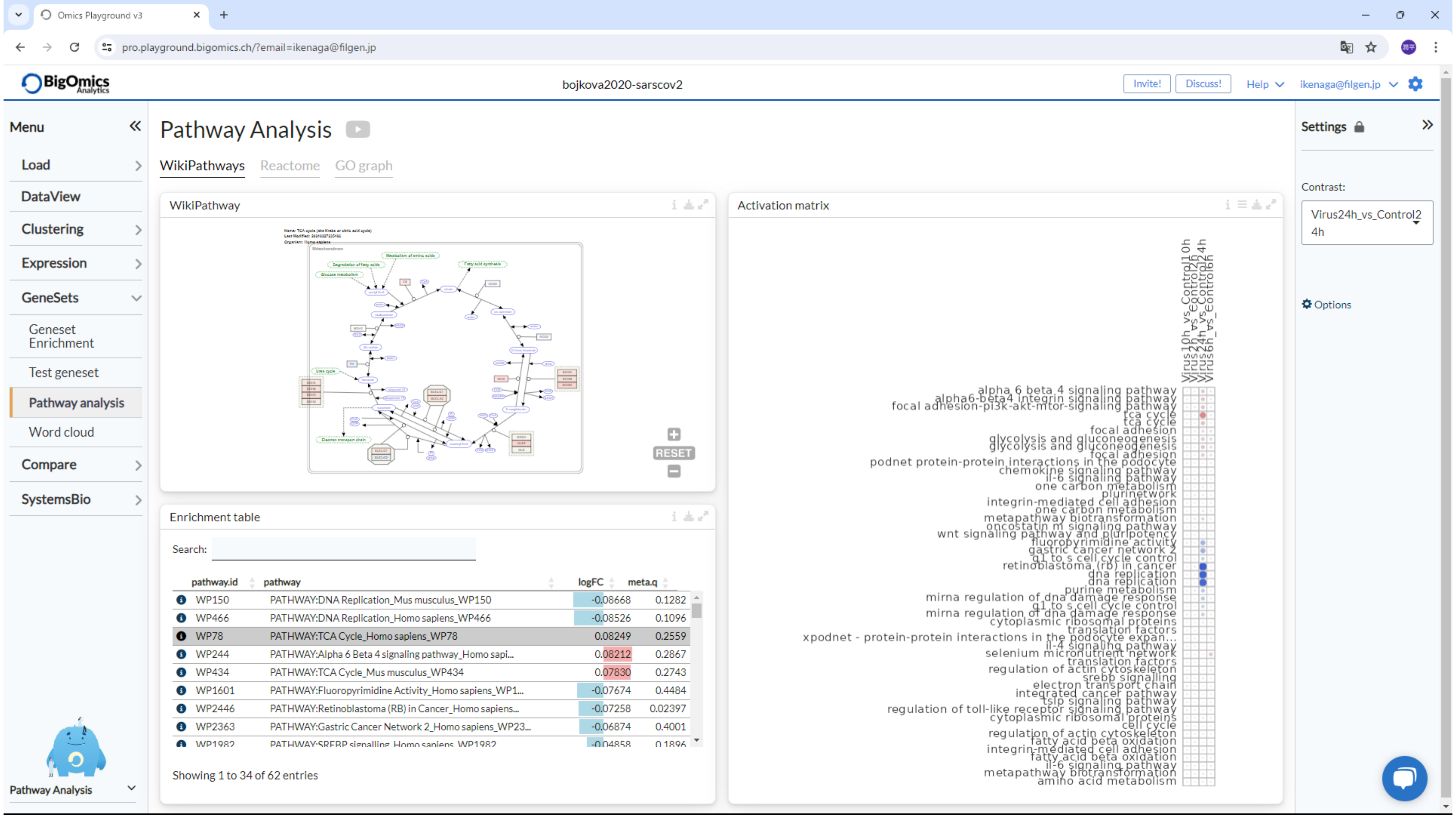
Signature type: <custom>

Genes:

MCM5 PCNA TYMS FEN1
MCM2 MCM4 RRM1 UNG
GINS2 MCM6 CDCA7 DTL
PRIM1 UHRF1 MLF1IP HELLS
RFC2 RPA2 NASP RAD51AP1
GMNN WDR76 SLBP CCNE2
UBR7 POLD3 MSH2 ATAD2
RAD51 RRM2 CDC45 CDC6
EXO1 TIPIN DSCC1 BLM
CASP8AP2 USP1 CLSPN
POLA1 CHAF1B BRIP1 E2F8
HMGB2 CDK1 NUSAP1
UBE2C BIRC5 TPX2 TOP2A
NDC80 CKS2 NUF2 CKS1B
MKI67 TMPO CENPF TACC3
FAM64A SMC4 CCNB2

[apoptosis] [cell_cycle] [immune_chkpt]

Genesets



Contrast: Virus24h_vs_Control24h

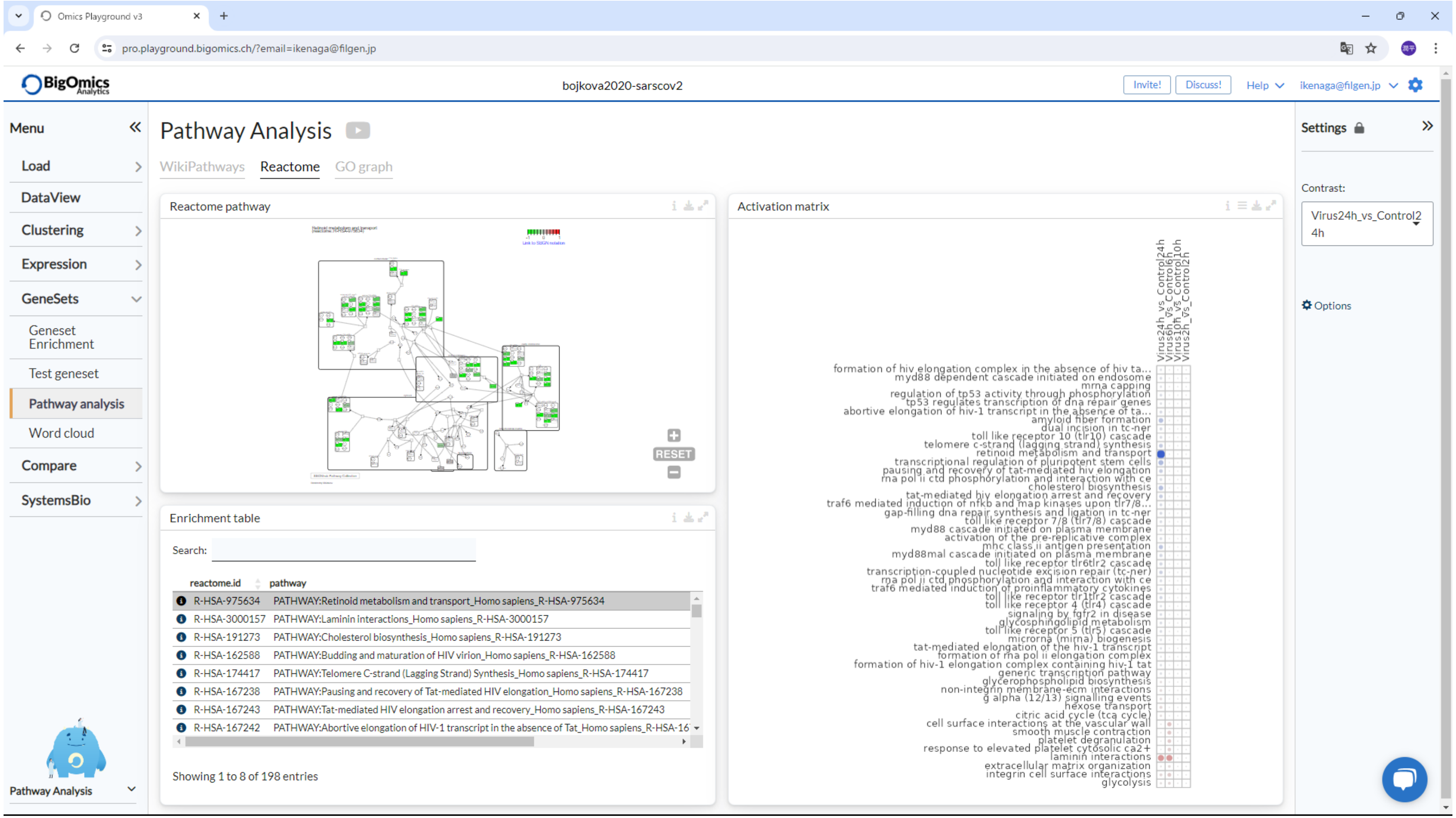
Enrichment table

Search:

	pathway.id	pathway	logFC	meta.q
●	WP150	PATHWAY:DNA Replication_Mus musculus_WP150	-0.08668	0.1282
●	WP466	PATHWAY:DNA Replication_Homo sapiens_WP466	-0.08526	0.1096
●	WP78	PATHWAY:TCA Cycle_Homo sapiens_WP78	0.08249	0.2559
●	WP244	PATHWAY:Alpha 6 Beta 4 signaling pathway_Homo sapi...	0.08212	0.2867
●	WP434	PATHWAY:TCA Cycle_Mus musculus_WP434	0.07830	0.2743
●	WP1601	PATHWAY:Fluoropyrimidine Activity_Homo sapiens_WP1...	-0.07674	0.4484
●	WP2446	PATHWAY:Retinoblastoma (RB) in Cancer_Homo sapiens...	-0.07258	0.02397
●	WP2363	PATHWAY:Gastric Cancer Network 2_Homo sapiens_WP23...	-0.06874	0.4001
●	WP1982	PATHWAY:SREBP signalling_Homo sapiens_WP1982	-0.04858	0.1896

Showing 1 to 34 of 62 entries

Genesets



Contrast: Virus24h_vs_Control24h

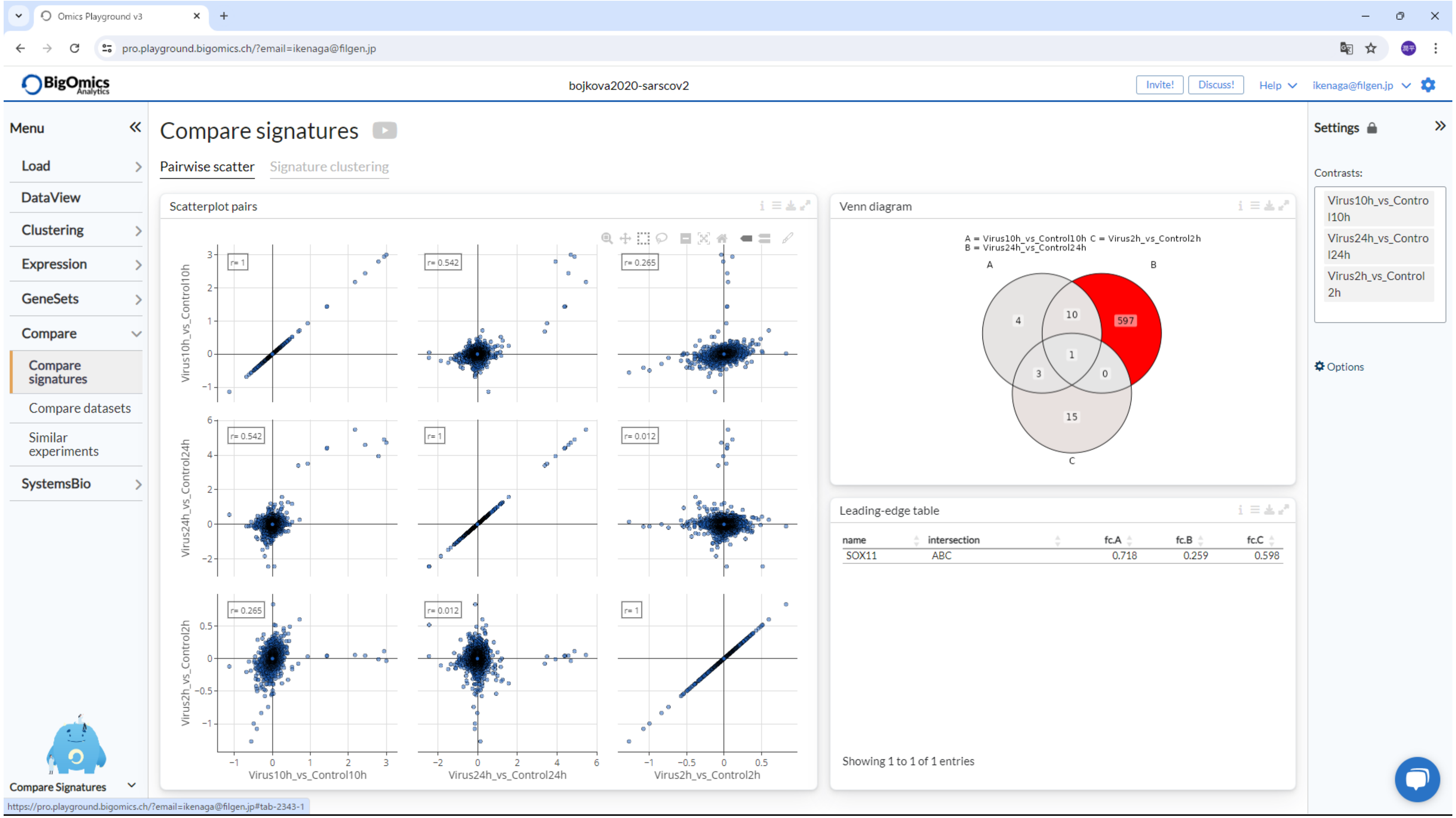
Enrichment table

Search:

reactome.id	pathway
R-HSA-975634	PATHWAY:Retinoid metabolism and transport_Homo sapiens_R-HSA-975634
R-HSA-3000157	PATHWAY:Laminin interactions_Homo sapiens_R-HSA-3000157
R-HSA-191273	PATHWAY:Cholesterol biosynthesis_Homo sapiens_R-HSA-191273
R-HSA-162588	PATHWAY:Budding and maturation of HIV virion_Homo sapiens_R-HSA-162588
R-HSA-174417	PATHWAY:Telomere C-strand (Lagging Strand) Synthesis_Homo sapiens_R-HSA-174417
R-HSA-167238	PATHWAY:Pausing and recovery of Tat-mediated HIV elongation_Homo sapiens_R-HSA-167238
R-HSA-167243	PATHWAY:Tat-mediated HIV elongation arrest and recovery_Homo sapiens_R-HSA-167243
R-HSA-167242	PATHWAY:Abortive elongation of HIV-1 transcript in the absence of Tat_Homo sapiens_R-HSA-16

Showing 1 to 8 of 198 entries

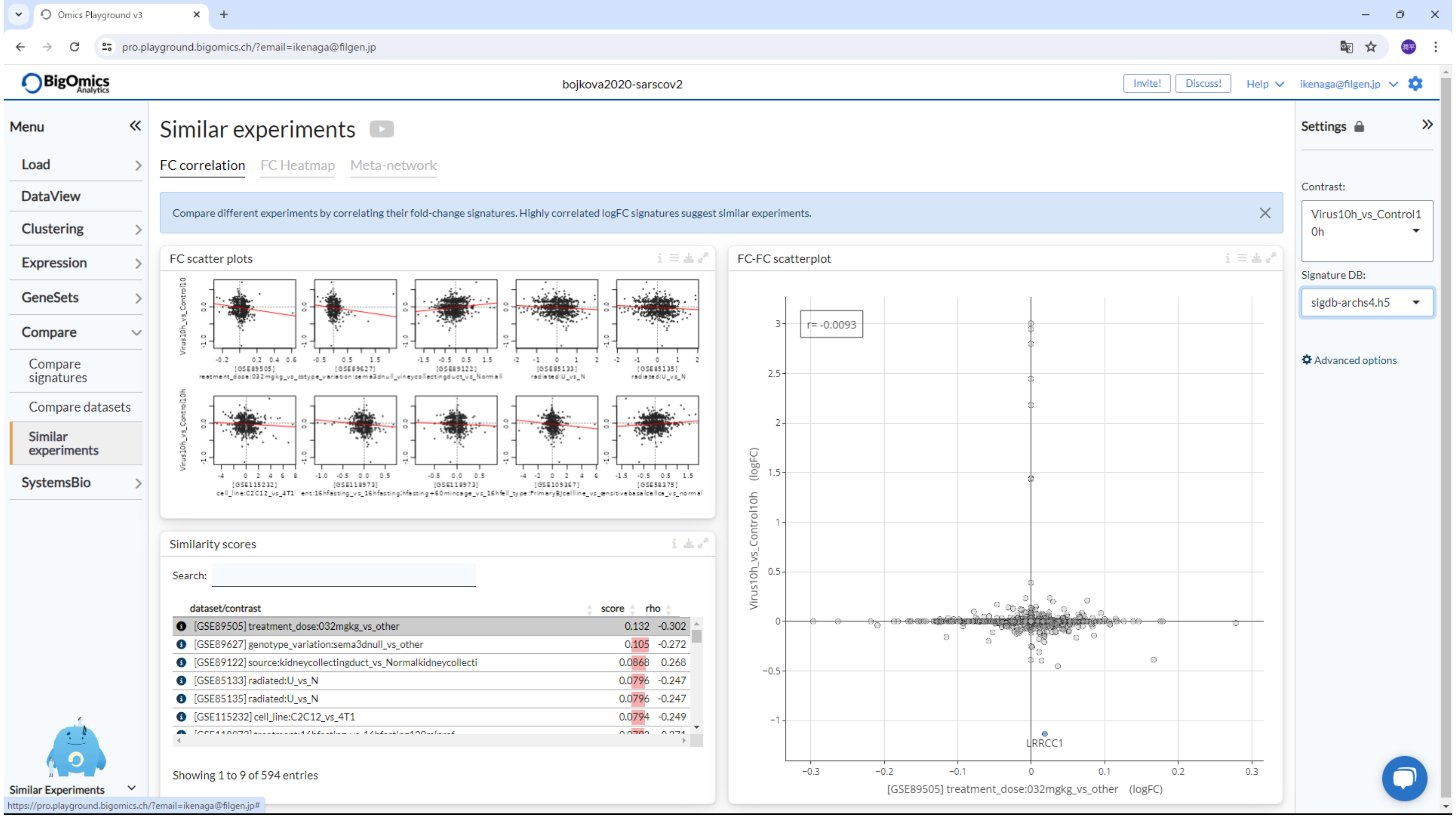
Compare



name	intersection	fc.A	fc.B	fc.C
SOX11	ABC	0.718	0.259	0.598

Showing 1 to 1 of 1 entries

Compare



Similar experiments

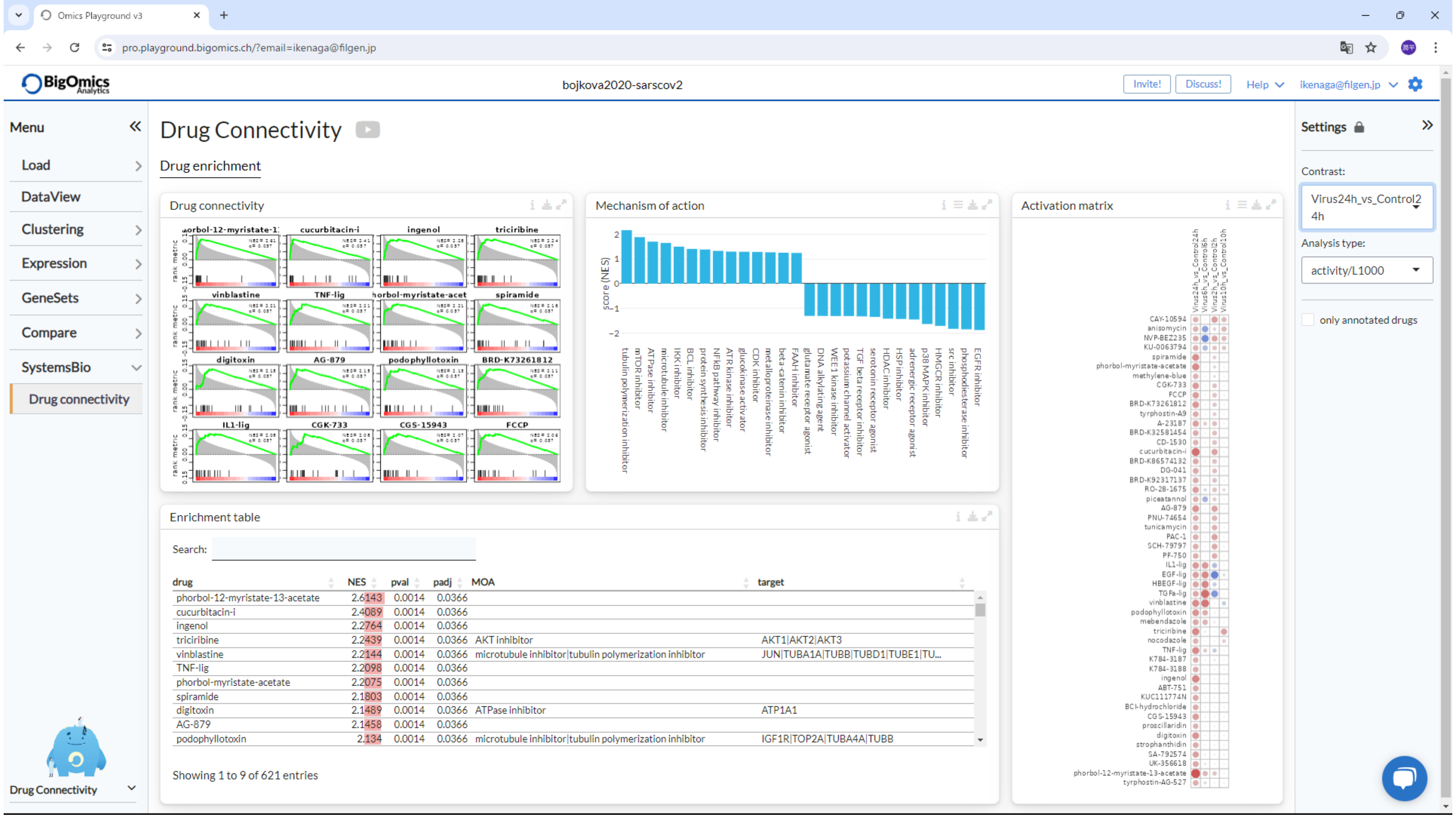
Compare different experiments by correlating their fold-change signatures. Highly correlated logFC signatures suggest similar experiments.

Search:

	dataset/contrast	score	rho
ℹ	[GSE89505] treatment_dose:032mgkg_vs_other	0.132	-0.302
ℹ	[GSE89627] genotype_variation:sema3dnull_vs_other	0.105	-0.272
ℹ	[GSE89122] source:kidneycollectingduct_vs_Normalkidneycollecti	0.0868	0.268
ℹ	[GSE85133] radiated:U_vs_N	0.0796	-0.247
ℹ	[GSE85135] radiated:U_vs_N	0.0796	-0.247
ℹ	[GSE115232] cell_line:C2C12_vs_4T1	0.0794	-0.249

Showing 1 to 9 of 594 entries

Systembio



Contrast: Virus24h_vs_Control24h

Analysis type: activity/L1000

☐ only annotated drugs

Search:

drug	NES	pval	padj	MOA	target
phorbol-12-myristate-13-acetate	2.6143	0.0014	0.0366		
cucurbitacin-i	2.4089	0.0014	0.0366		
ingenol	2.2764	0.0014	0.0366		
triciribine	2.2439	0.0014	0.0366	AKT inhibitor	AKT1|AKT2|AKT3
vinblastine	2.2144	0.0014	0.0366	microtubule inhibitor|tubulin polymerization inhibitor	JUN|TUBA1A|TUBB|TUBD1|TUBE1|TU...
TNF-lig	2.2098	0.0014	0.0366		
phorbol-myristate-acetate	2.2075	0.0014	0.0366		
spiramide	2.1803	0.0014	0.0366		
digitoxin	2.1489	0.0014	0.0366	ATPase inhibitor	ATP1A1
AG-879	2.1458	0.0014	0.0366		
podophyllotoxin	2.134	0.0014	0.0366	microtubule inhibitor|tubulin polymerization inhibitor	IGF1R|TOP2A|TUBA4A|TUBB

Showing 1 to 9 of 621 entries

お問い合わせ先：フィルジェン株式会社

TEL: 052-624-4388 (9:00～17：00)

FAX: 052-624-4389

E-mail: support@filgen.jp