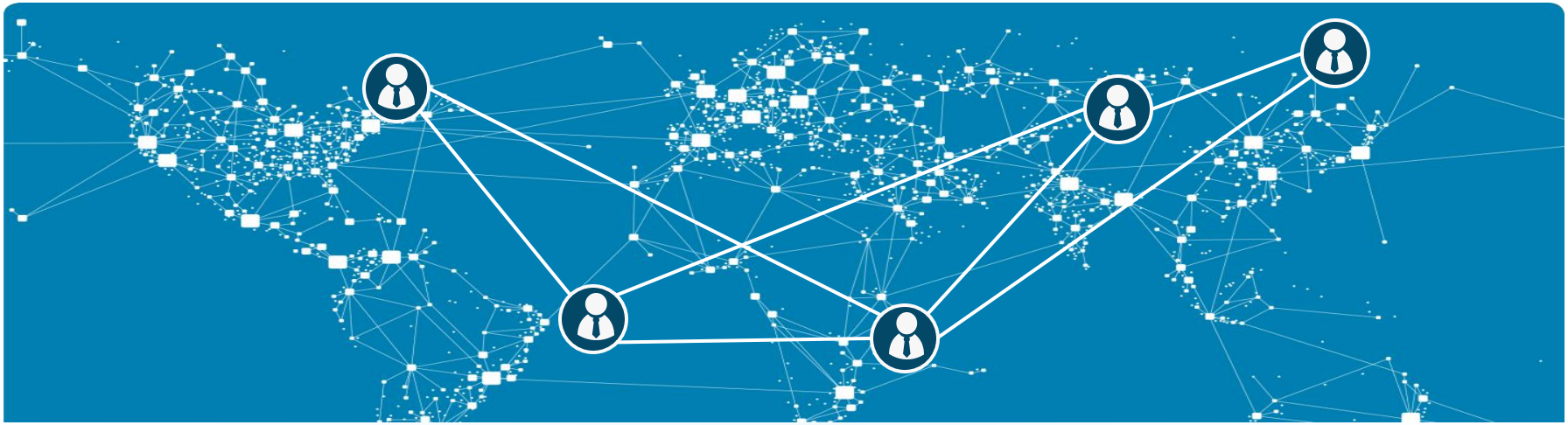
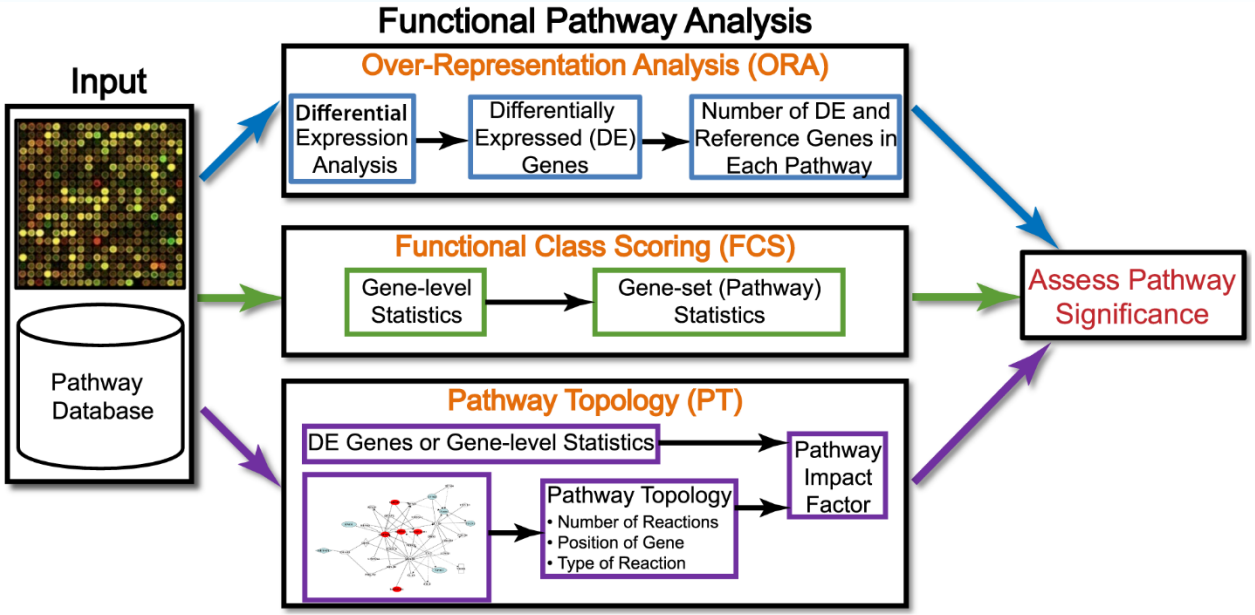


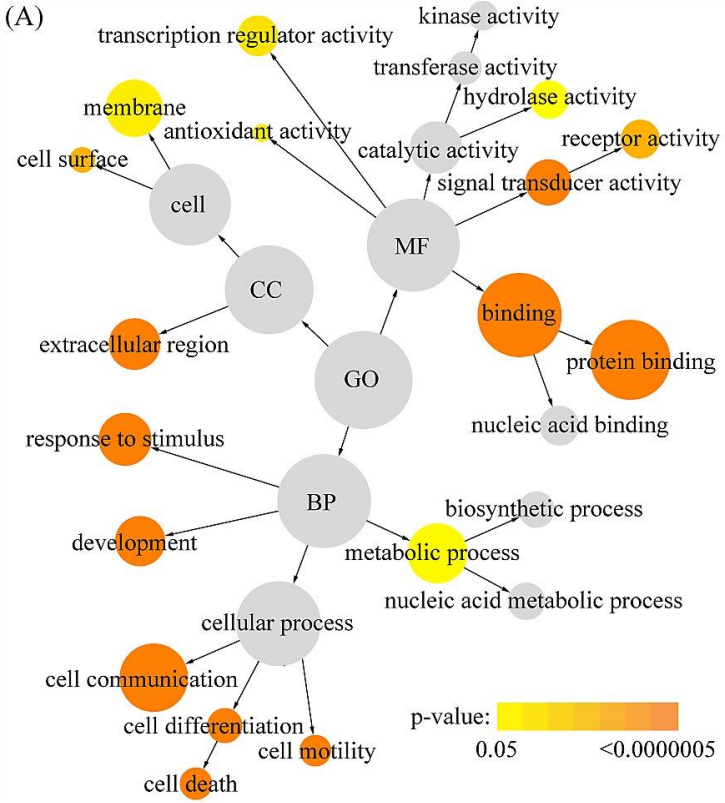
富集分析

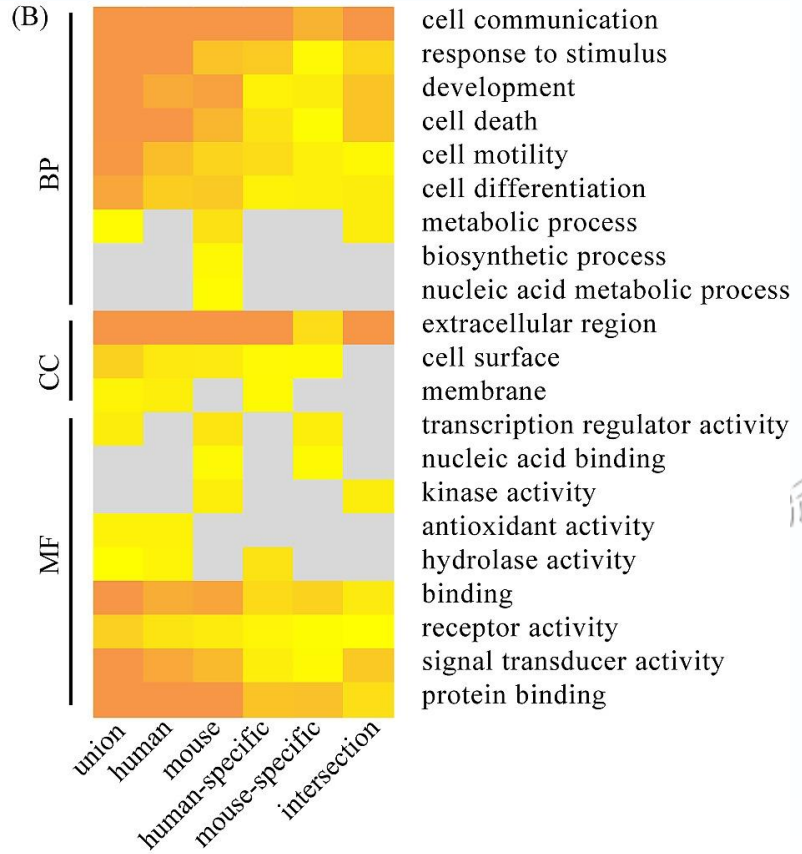
富集分析的意义

- 从数以千计的基因里面查找其倾向参与的调控通路，以指导下一步的研究方向。

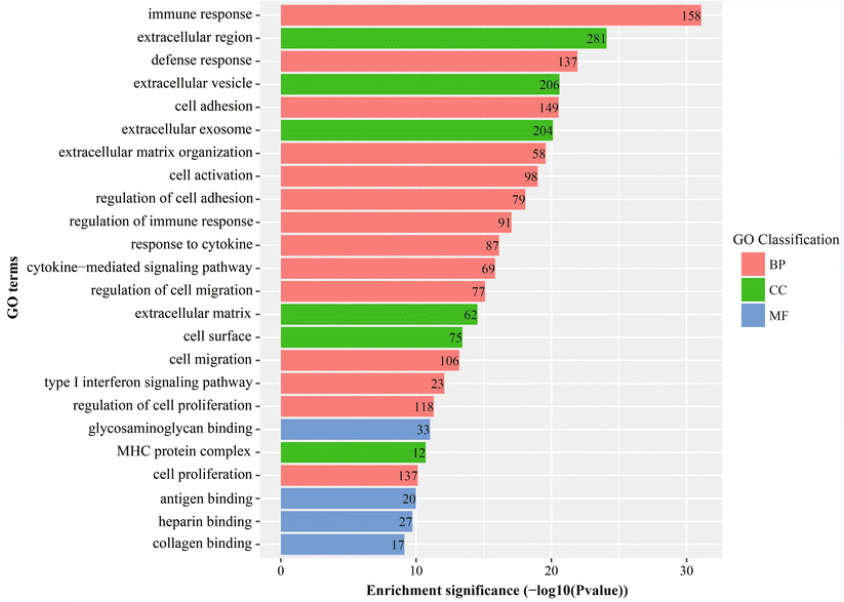


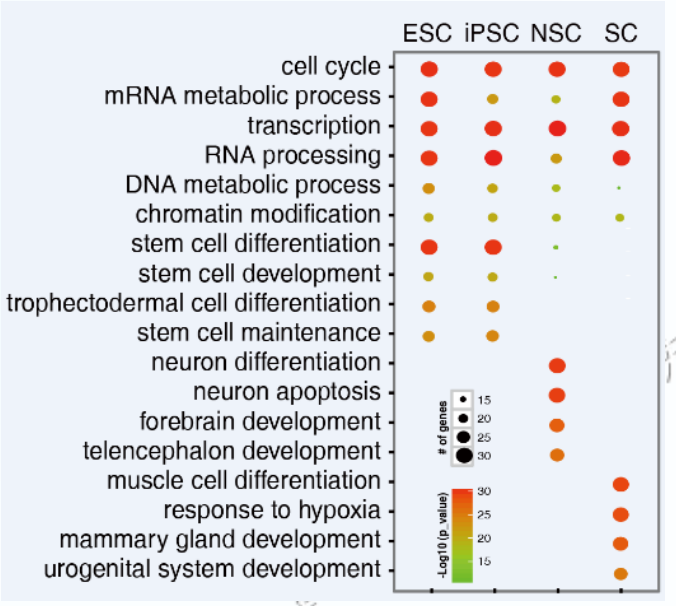
GO富集结果展示





GO富集结果展示



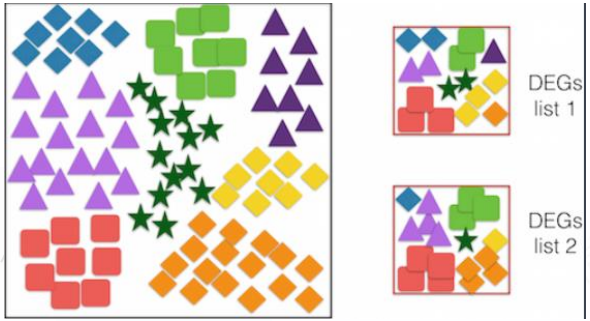


基因功能富集分析 – GO/KEGG/Reactome

基于筛选的差异基因或其它自己定义的一组基因，采用**超几何检验**判断上调或下调基因在哪些GO或KEGG或其它定义的通路富集。假设背景基因数目为 `m` ，背景基因中某一通路pathway中注释的基因有 `n` 个；上调基因有 `k` 个，上调基因中落于通路pathway的数目为 `l` 。简单来讲就是比较 `l/k` 是否显著高于（也可以是低于） `n/m` ，即上调基因中落在通路pathway的比例是否高于背景基因在这一通路的比例。(实际计算时，是算的 `odds ratio` 的差异， `l/(k-l) vs (n-l)/(m-k-n+l))` 。

	User Genes	Genome
In Pathway	l	n-l
Not In Pathway	k-l	m-k-n+l

Inputs



Enrichment analysis

GO/GSEA

WGCNA Cytoscape

Network analysis

GO富集分析的工具 – GOEAST输入

- 输入基因列表即可(13_salmon_de_go/ehbio_salmon..DESeq2.untrt._lowerThan_.trt.id.xls，需要id转换)
- 注释数据每周更新

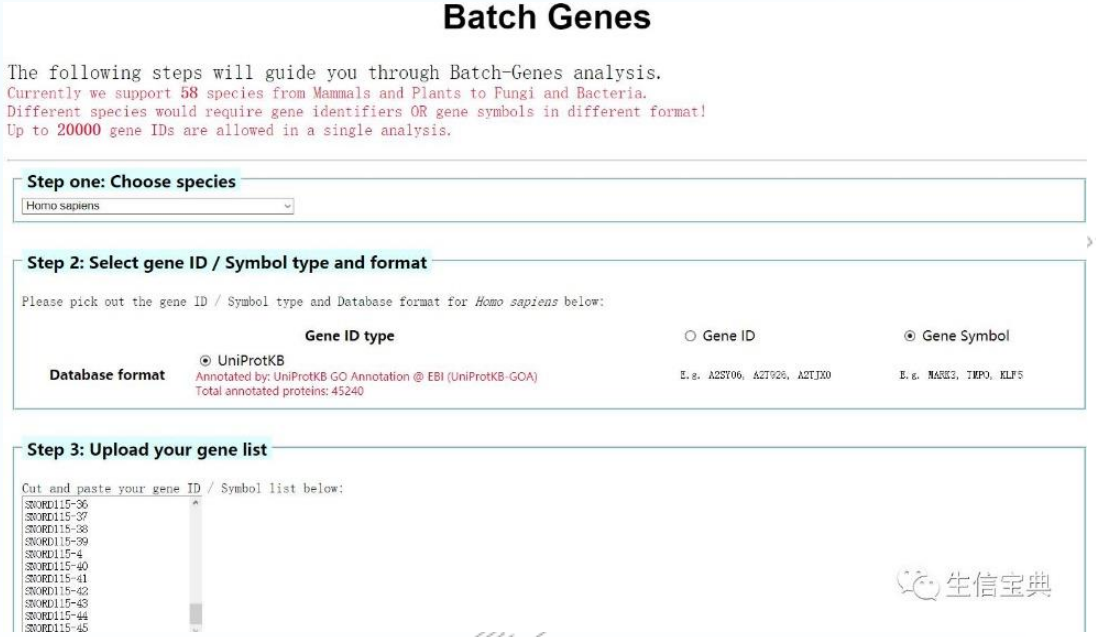
GOEAST News:

2016-07-02: new!
GOEAST server auto update is significantly speeded up, i.e. much shorter maintenance time now.

2019-02-18: new!
Gene Ontology information updated.

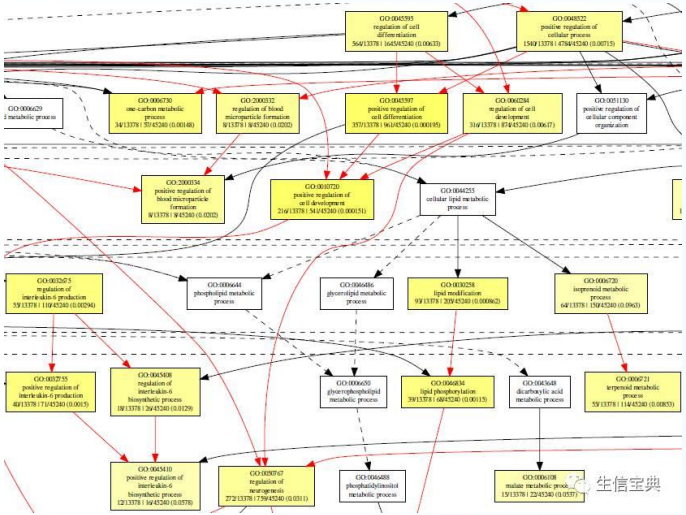
2019-02-11: new!
Gene Ontology information updated.

2019-02-04:
Gene Ontology information updated.



去东方最好用的在线富集分析工具

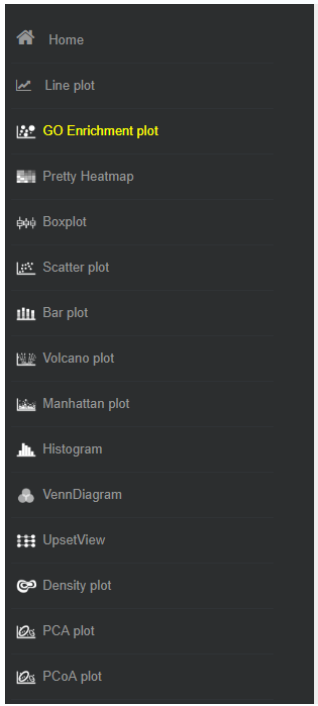
GO富集分析的工具 – GOEAST输出



GOID	Ontology	Term	Level	q	m	t	k	geneIDs	symbols	log_odds_ratio	p
GO:0006730	biological_process	one-carbon metabolic process	4	34	57	45240	13378	C9JZV7 // C9K0S0 // D
GO:0044699	biological_process	single-organism process	1	5781	18620	45240	13378	C9JZV7 // C9K0S0 // D
GO:0044763	biological_process	single-organism cellular process	2	4988	16172	45240	13378	C9JZV7 //
GO:0007154	biological_process	cell communication	2	2169	6843	45240	13378	C9J1Z8 // P84085 // Q16864
GO:0007165	biological_process	signal transduction	5	1955	6136	45240	13378	C9J1Z8 // P84085 // Q16864
GO:0023052	biological_process	signaling	1	2100	6613	45240	13378	C9J1Z8 // P84085 // Q16864 // Q9Y5
GO:0044700	biological_process	single organism signaling	2	2100	6613	45240	13378	C9J1Z8 // P84085 /
GO:0050896	biological_process	response to stimulus	1	3251	10438	45240	13378	C9J1Z8 // P84085 // Q1
GO:0005515	molecular_function	protein binding	1	3299	10399	45240	13378	P84085 // Q16864 // O43852 //
GO:0005737	cellular_component	cytoplasm	3	4711	15220	45240	13378	P84085 // Q16864 // H0Y875 // O43
GO:0005794	cellular_component	Golgi apparatus	6	611	1835	45240	13378	P84085 // O43852 // H7C1T7 // Q9UB
GO:0012505	cellular_component	endomembrane system	2	1521	4648	45240	13378	P84085 // H0Y875 // O43852
GO:0044444	cellular_component	cytoplasmic part	4	3314	10694	45240	13378	P84085 // Q16864 // H0Y875
GO:0071944	cellular_component	cell periphery	2	2059	6559	45240	13378	P84085 // Q14315 // Q96CN9 //
GO:0065008	biological_process	regulation of biological quality	1	1379	4097	45240	13378	Q16864 //
GO:0007166	biological_process	cell surface receptor signaling pathway	5	856	2561	45240	13378	Q16864 //
GO:0007167	biological_process	enzyme linked receptor protein signaling pathway	5	373	1007	45240	13378	
GO:0007169	biological_process	transmembrane receptor protein tyrosine kinase signaling pathway	5	271	732	4
GO:0032991	cellular_component	macromolecular complex	1	2221	6950	45240	13378	Q16864 // F6X838 // H
GO:0043234	cellular_component	protein complex	1	1900	5720	45240	13378	Q16864 // F6X838 // H7C1T7 //
GO:0098796	cellular_component	membrane protein complex	4	595	1695	45240	13378	Q16864 // F6X838 // H
GO:0005773	cellular_component	vacuole	5	421	1132	45240	13378	Q16864 // P30048 // Q7L804 // Q01085 // P
GO:0005774	cellular_component	vacuolar membrane	9	242	544	45240	13378	Q16864 // Q7L804 // P15313 // Q9U
GO:0031090	cellular_component	organelle membrane	4	909	2684	45240	13378	Q16864 // O[illegible]NX63 // H7C1T7 //
GO:0044437	cellular_component	vacuolar part	7	268	619	45240	13378	Q16864 // Q7L804 // [illegible]5553 [illegible]971
GO:0098588	cellular_component	bounding membrane of organelle	4	638	1759	45240	13378	Q16864 // H7C1T7
GO:0098805	cellular_component	whole membrane	1	525	1447	45240	13378	Q16864 // H7C1T7 // Q9UBF2 // O00

GO富集分析的工具 – GOEAST输出绘图

易生信



GOID	Ontology	Term	Level	q	m	t	k	geneIDs	symbols	log_odds_ratio	p
GO:0006730	biological_process	one-carbon metabolic process		4	34	57	45240	13378	C9JZV7 // C9K0S0 // D7UEQ7 // F8W7N8 // H0Y8B3 // Q96HN2 // B4DN45 // P31153 // P13995 // E5RFE7 // E5RFL2 // E5RG43 // E5RG81 // E5RGU8 // E5RH81 // E5RHP7 // E5RIF9 // E5RII2 // E5RJF6 // E5RJI8 // H0YBE2 // P00915 // Q8N1Q1 // P00918 // P07451 // Q9NZL9 // Q9Y2D0 // Q14749 // Q00266 // Q6UB35 // F8W148 // P23280 // P42898 // Q5SNW6	AHCYL2 // AHCYL2 // AHCYL2 // AHCYL2 // AHCYL2 // AHCYL2 // MAT2A // MAT2A // MTHFD2 // CA1 // CA1 // CA1 // CA1 // CA1 // CA1 // CA1 // CA1 // CA1 // CA1 // CA1 // CA1 // CA1 // CA13 // CA2 // CA3 // MAT2B // CA5B // GNMT // MAT1A // MTHFD1L // CA6 // CA6 // MTHFR // MTHFR	1.012309306035810.00148115106666474	
GO:0044699	biological_process	single-organism process		1	5781	18620	45240	13378	C9JZV7 // C9K0S0 // D7UEQ7 // F8W7N8 // H0Y8B3 // Q96HN2 // C9J1Z8 // P84085 // Q16864 // Q43852 // C9IZ34 // C9J6R3 // C9J9X8 // C9JXA0 // Q9BYV8 // Q9NX63 // Q14315 // Q9Y5L2 // C9J029 // C9J381 // C9K0R9 // E7EQS0 // H7C511 // H7C5T1 // P20839 // Q13568 // P41159 // Q9UL63 // B4DDV6 // C9JP85 // Q16656 // P03999 // G3V4Q1 // J3KPG0 // Q43316 // Q9HCM2 // Q00592 // H7C509 // H7C5S9 // Q99835 // Q7KZF4 // Q9ULQ0 // Q9Y5L0 // Q86WB0 // O14639 // Q9ULC5 // P08913 // P08588 // Q5VV63 // Q95817 // A8MTZ0 // Q86Y37 // P55210 // Q5SVL1 // Q5SVL2 // Q6PJP8 // Q16690 // Q04743 // A6NI74 // A6NNW6 // J3KNX1 // P56159 // Q5VW52 // Q9HCL2 // G8JLE8 // P34947 // P78417 // Q9H4Y5 // Q9Y2H2 // Q7Z418 // Q9BTE3 // P50539 // Q8WY41 // Q53EL6 // Q8NEN9 // P16233 // E7EX42 // E9PJC9 // E9PMA6 // E9PMX6 // E9PR20 // E9PRT6 // H0YDY1 // P54315 // F5H1V6 // P54317 // Q17RR3 // P30048 // P49683 // Q7L804 // Q5T481 // Q9UQ13 // Q05940 // Q9UQE7 // O75940 // Q8WY21 // Q9UPU3 // E2GH20 // H0YEE5 // Q5JRY3 // Q5JRY4 // Q5JRY5 // Q5VVR5 // Q5VVR7 // Q5VVR8 // Q9NQB0 // Q9BXT4 // Q01085 // Q5SQQ9 // Q8TDX5 // H7C2M3 // H7C345 // Q9NUZ1 // F5H6T1 // P61160 // B4DXW1 // F5H3P5 // F8WE84 // F8WEW2 // P61158 // C9J080 // C9J299 // P35612 // P18089 // P51826 // A6NMY3 // Q8TCU4 // Q8N944 // Q9H1A4 // Q86SG2 // Q9H6X2 // P09525 // B7Z8K7 // C9JB56 // H7C4P1 // P42331 // E9PEM0 // H0YCT6 // Q9NR80 // C9J1Q0 // Q03989 //											

Essential parameters

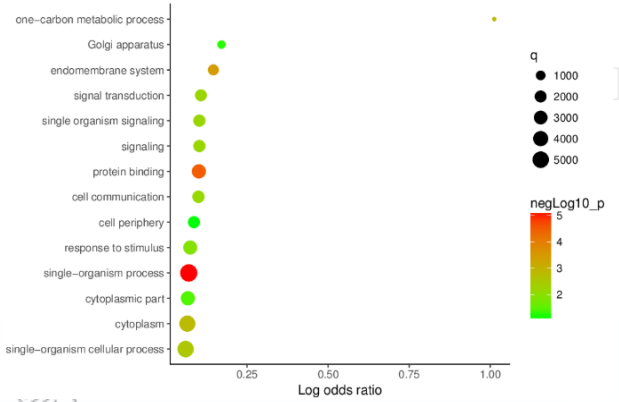
X-axis type: Discrete variable

X-axis variable *: log_odds_ratio

Y-axis variable *: Term

Color variable *: p -Log10

Point size variable *: q



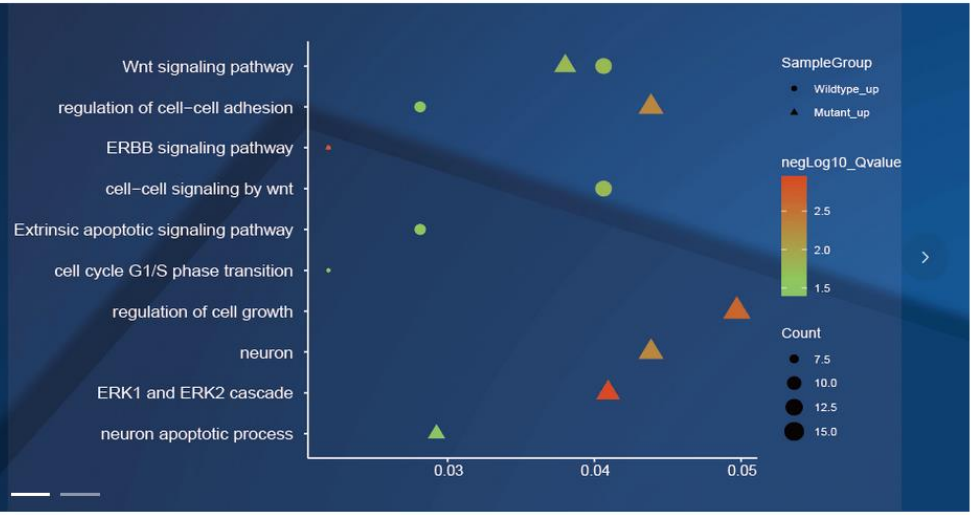
新版绘图网站

Enrichment plot

Data matrix (tab separated text file)

SampleGroup	GeneRatio	Qvalue	Count	Description
Wildtype_up	7/320	0.001836081	7	ERBB signaling pathway
Wildtype_up	13/320	0.01680096	13	Wnt signaling pathway
Wildtype_up	13/320	0.0171473	13	cell-cell signaling by wnt
Wildtype_up	7/320	0.035797856	7	cell cycle G1/S phase transition
Wildtype_up	9/320	0.024164034	9	Extrinsic apoptotic signaling pathway
Wildtype_up	9/320	0.024164034	9	regulation of cell-cell adhesion
Mutant_up	7/320	0.001836081	7	ERBB signaling pathway
Mutant_up	14/342	0.001086508	14	ERK1 and ERK2 cascade
Mutant_up	13/342	0.01680096	13	Wnt signaling pathway
Mutant_up	17/342	0.002228511	17	regulation of cell growth
Mutant_up	15/342	0.005075219	15	regulation of cell-cell adhesion
Mutant_up	10/342	0.040284925	10	neuron apoptotic process
Mutant_up	15/342	0.005075219	15	neuron

Set X-axis variable GeneRatio
Set Y-axis variable Description
Set Color variable Qvalue
Set Neg log10 transform variable Qvalue
Set Point size variable Count
Set Shape variable SampleGroup



Demo1 Demo2

高颜值免费在线绘图平台

ImageGP New

Venn 网络展示富集结果

Matrix (two-column mode matrix)

Gene_enrichment

Gene	Pathway
CDKN2B	P53_PATHWAY
HSPA4L	P53_PATHWAY
SDC1	P53_PATHWAY
HMOX1	P53_PATHWAY
RRAD	P53_PATHWAY
ZNF365	P53_PATHWAY
SOCS1	P53_PATHWAY
JAG2	P53_PATHWAY
TRIB3	P53_PATHWAY
TNFSF9	P53_PATHWAY
JUN	P53_PATHWAY

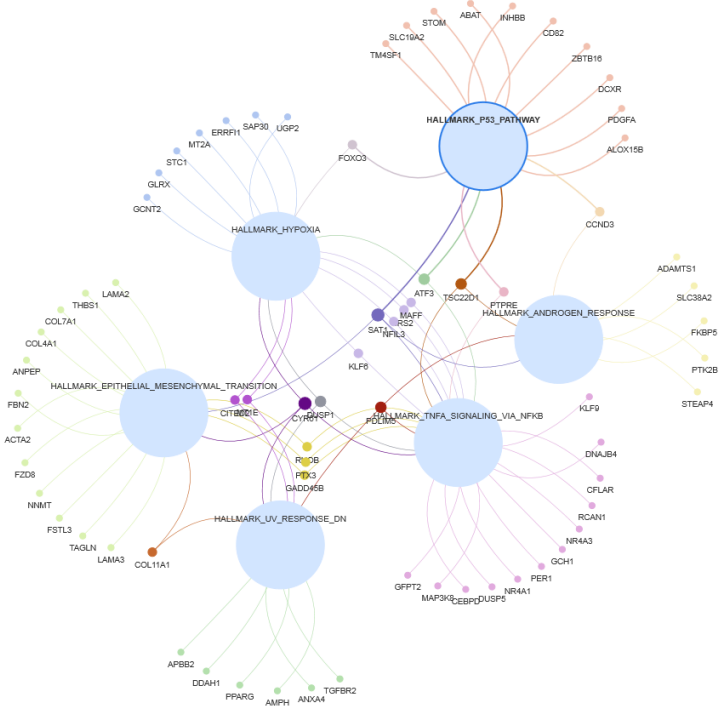
Parameters

Set colors for each set:

+ Open the

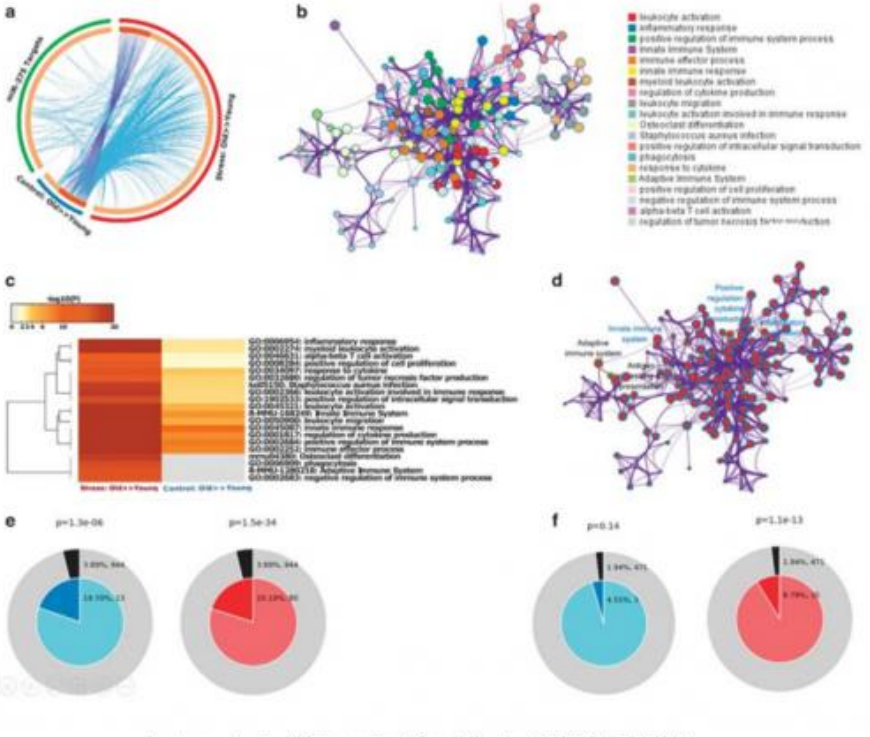
Showing mode:

- Show all elements
- Show only common elements and spe

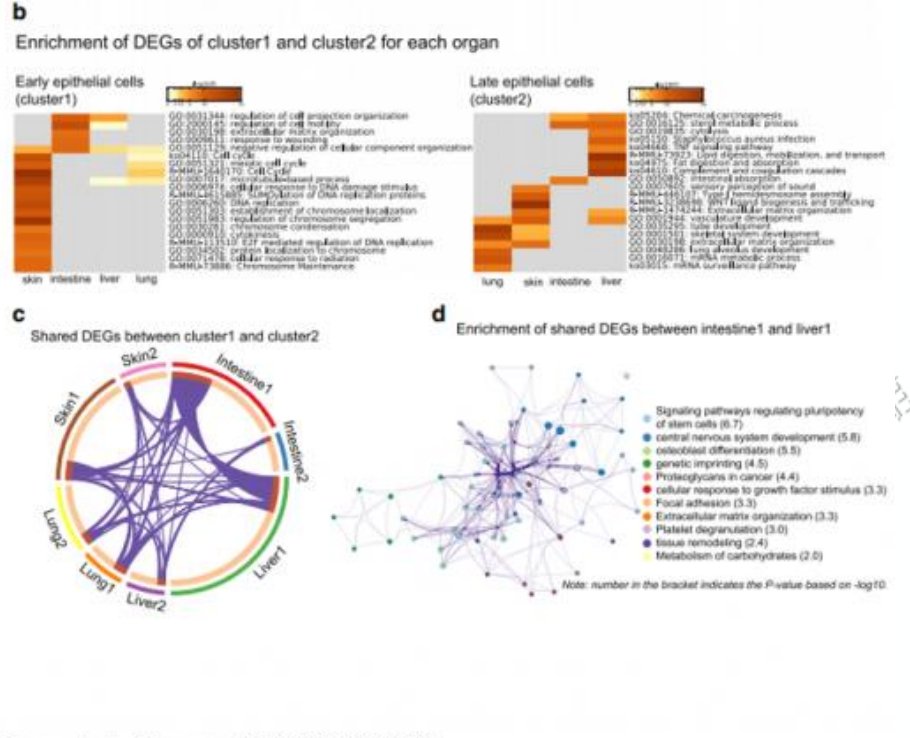


[工具链接](#)　　[使用教程](#)

Metascape富集分析以及蛋白互作网络



Lotan et al. Molecular Psychiatry (2018) 78:865



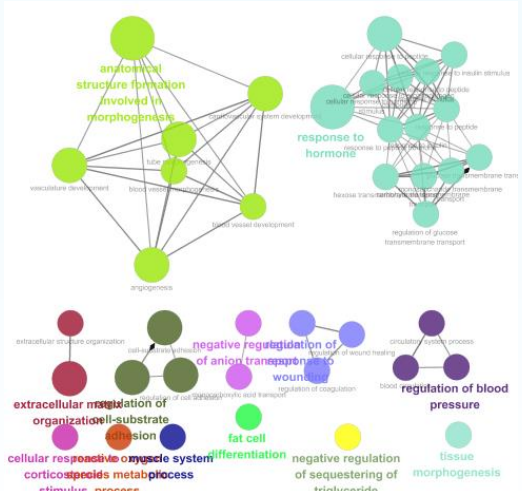
Dong et al. Genome Biol (2018) 19:31

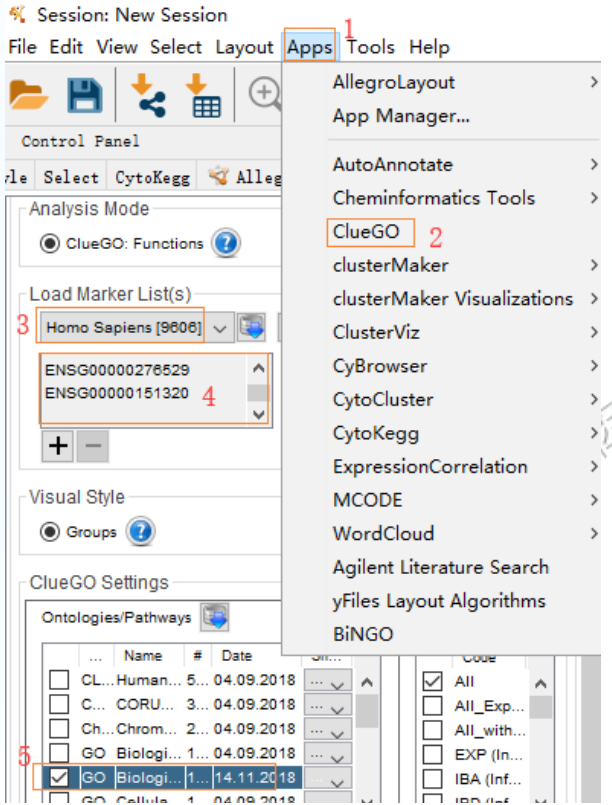
工具网址

这个只需一步就可做富集分析的网站还未发表就被CNS等引用超过350次

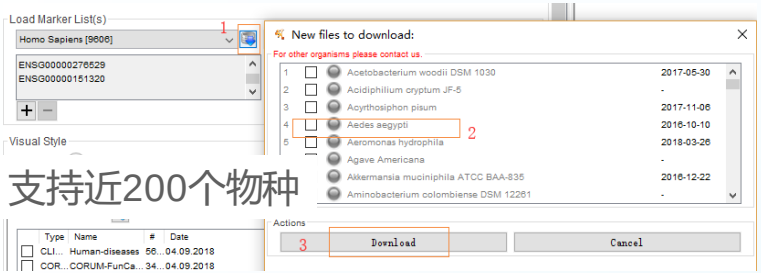
GO富集分析的工具 – clueGO，操作简单

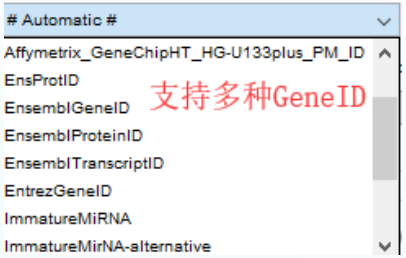
- 按图示输入基因名字（名字格式支持多种），点Start即可
  （clueGo/ehbio_salmon..DESeq2.untrt._lowerThan_.trt.id.xls）

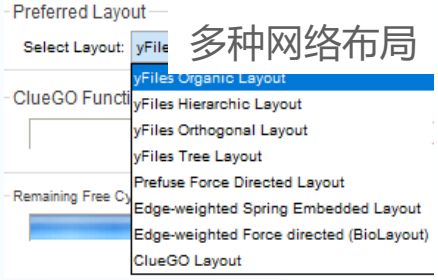


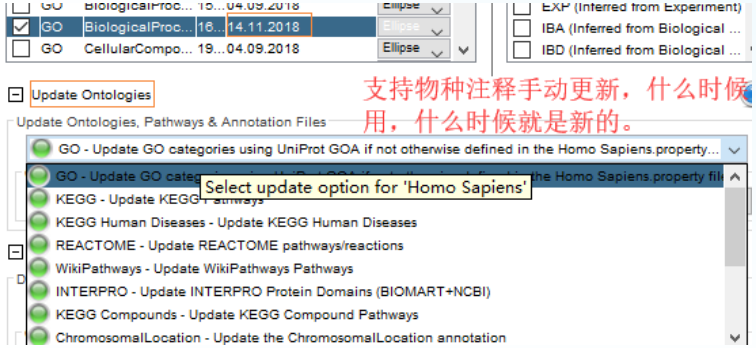


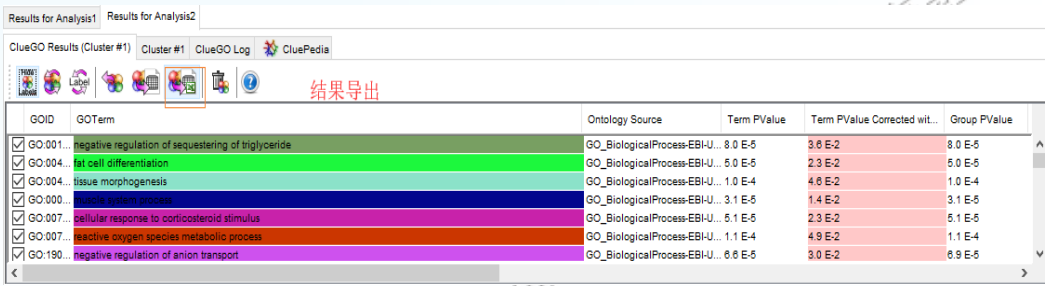
GO富集分析的工具 – clueGO，更新方便



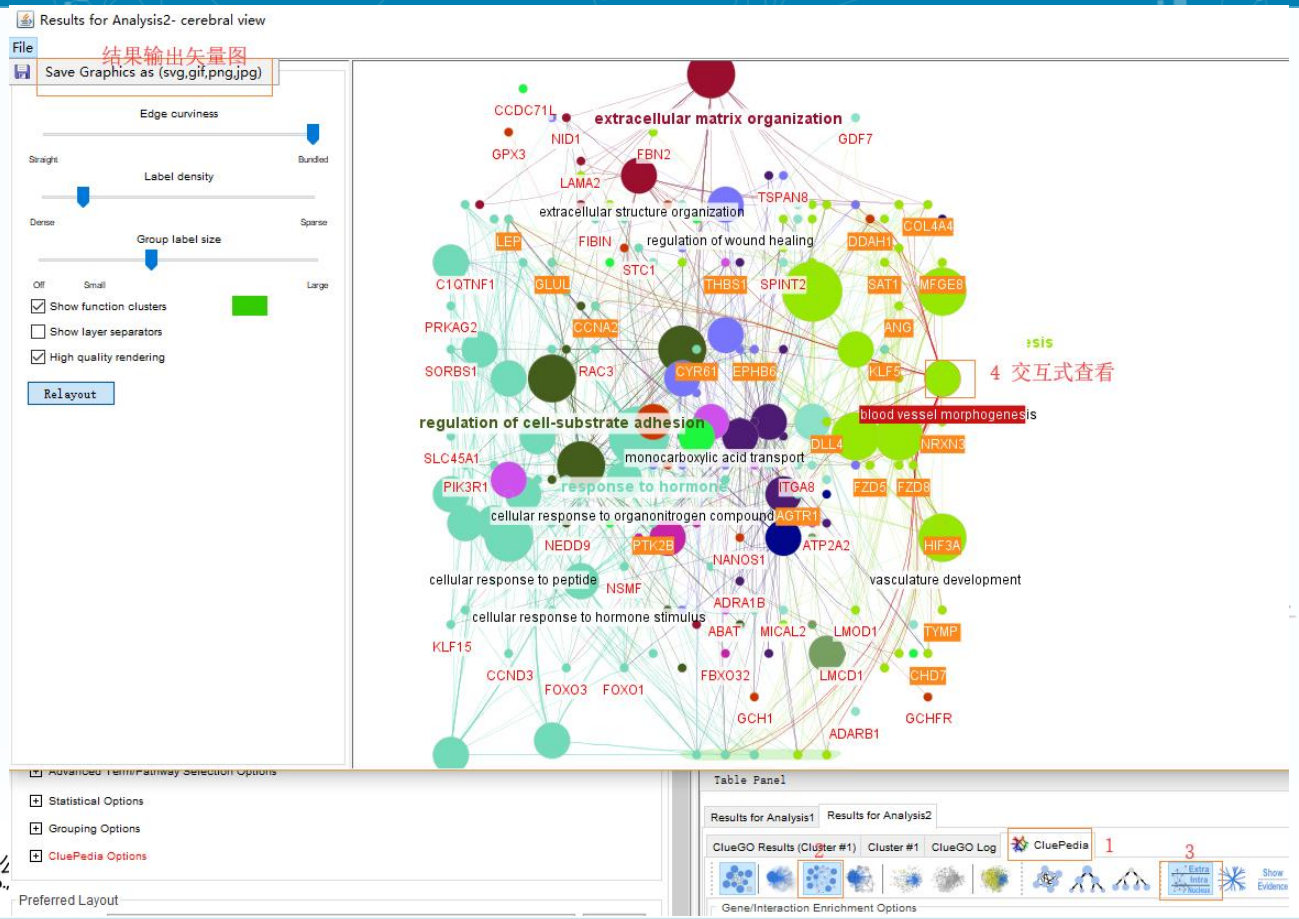




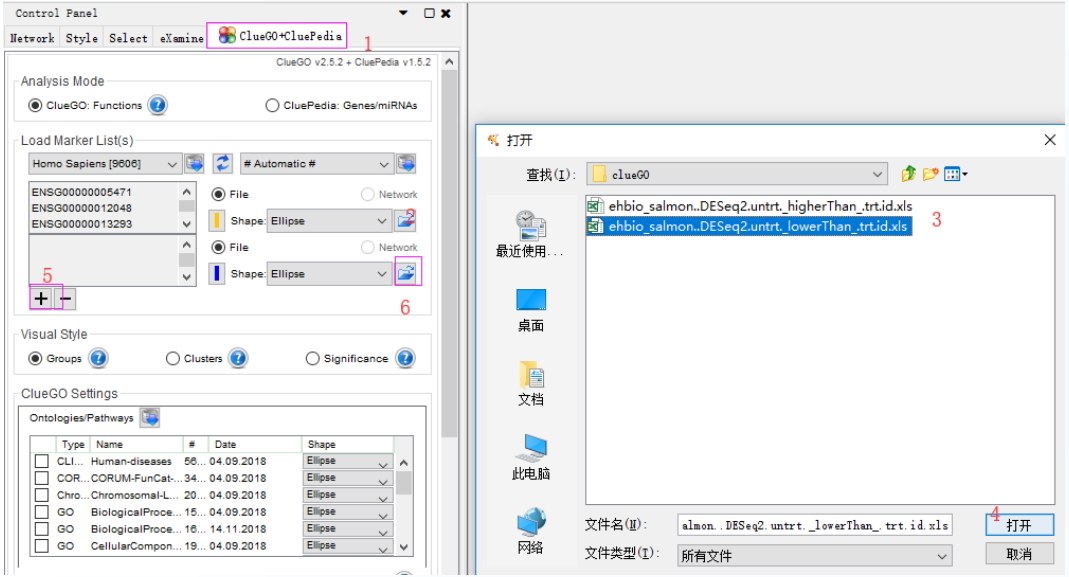


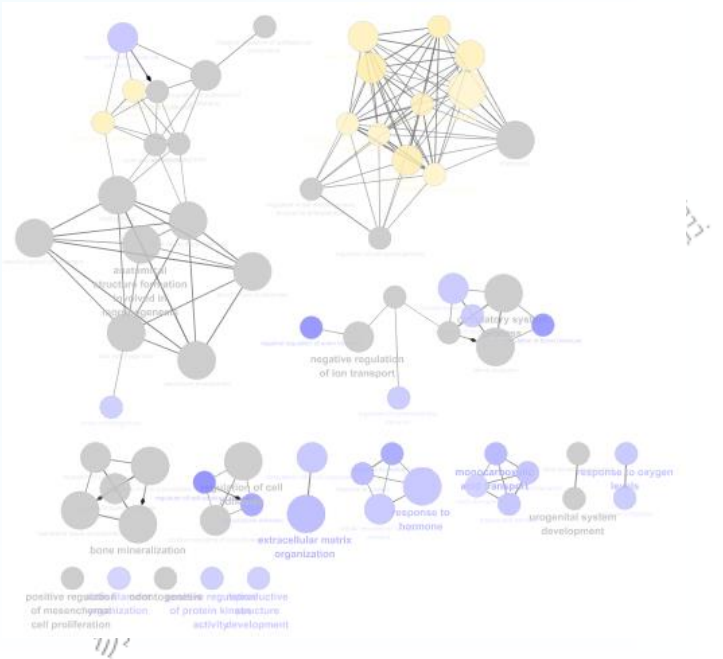


GO富集分析的工具 – clueGO，交互探索

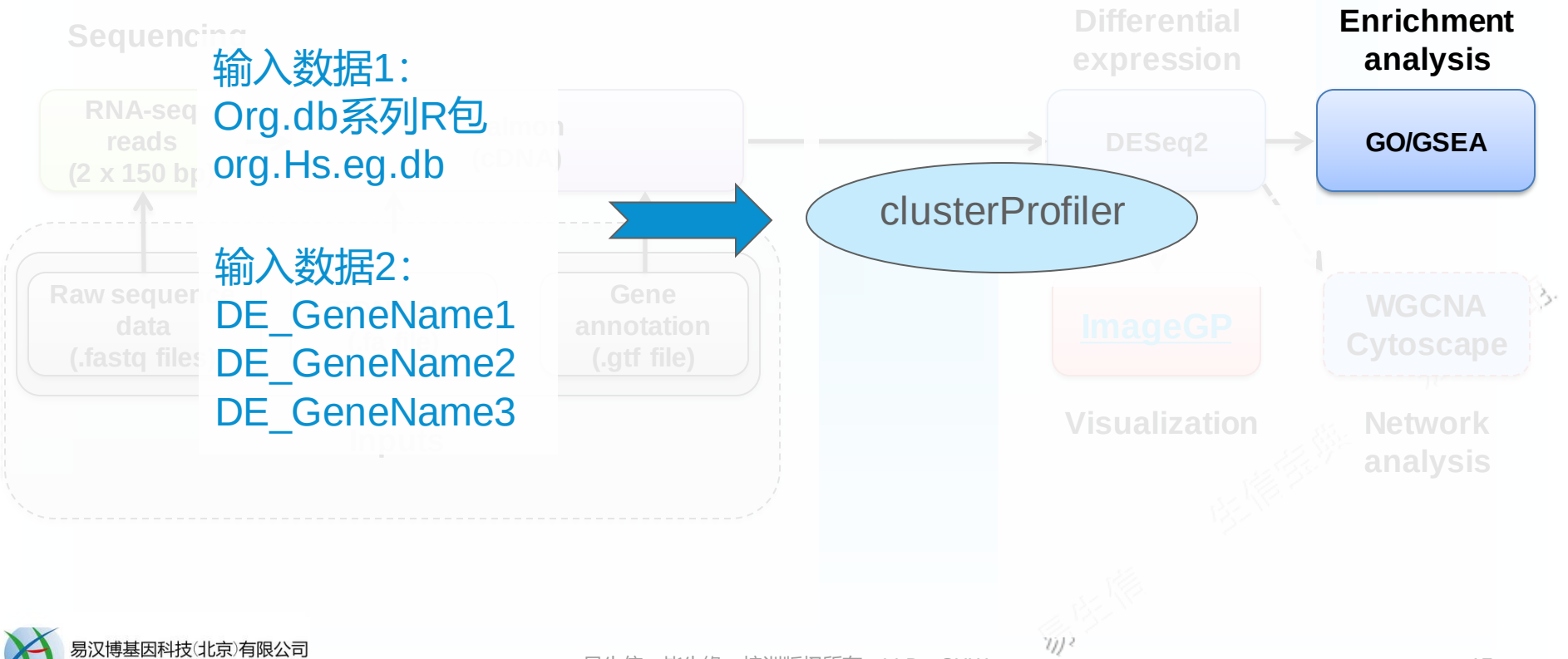


GO富集分析工具 – clueGO，多个数据集比较



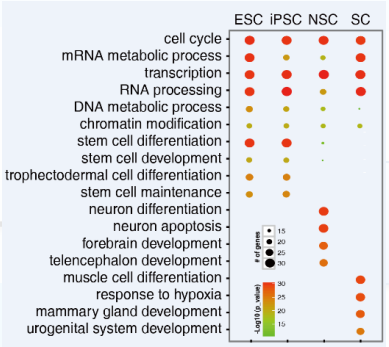


基因功能富集分析 – R包 clusterProfiler



基因功能富集分析

输出结果 （绘制富集分析泡泡图需要哪些列？）

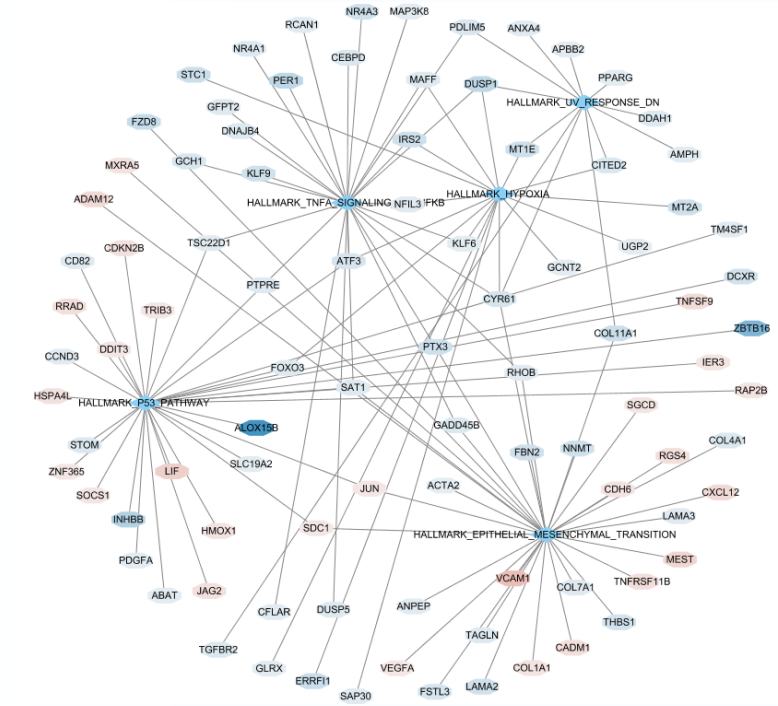


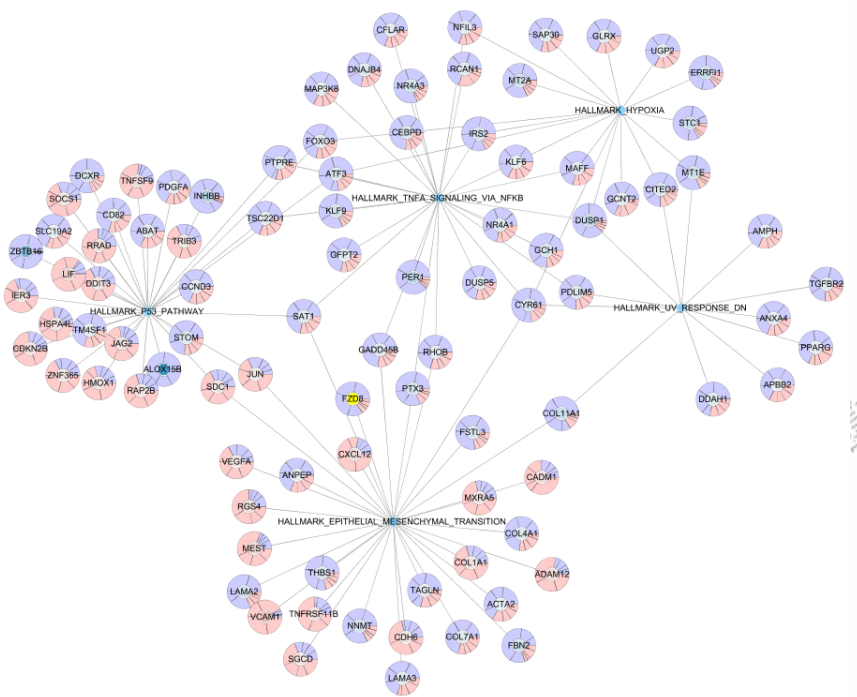
Enrichment analysis

GO/GSEA

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	geneID	Count
GO:0048	blood vessel morphogenesis	29/380	492/16992	2.17E-06	0.003360933	FZD8/TGFBR2/SAT1	29
GO:0030	extracellular matrix organiza	22/380	315/16992	2.60E-06	0.003360933	APBB2/CRISPLD2/CI	22
GO:0043	extracellular structure organi	22/380	316/16992	2.74E-06	0.003360933	APBB2/CRISPLD2/CI	22
GO:1901	cellular response to peptide	21/380	320/16992	1.15E-05	0.008437395	KLF15/ERRFI1/IRS2/	21
GO:0045	fat cell differentiation	16/380	204/16992	1.47E-05	0.009026139	ZBTB16/CEBPD/FOX	16
GO:1901	organonitrogen compound c	22/380	362/16992	2.35E-05	0.010478057	MAOA/SAMHD1/AB	22
GO:0072	reactive oxygen species met	17/380	236/16992	2.42E-05	0.010478057	GPX3/FOXO3/FOXC	17
GO:0008	regulation of blood pressure	14/380	168/16992	2.56E-05	0.010478057	ABAT/KCNK6/ADRA	14
GO:0001	urogenital system developm	20/380	315/16992	2.97E-05	0.010942076	ZBTB16/KLF15/ADA	20

分析结果导入Cytoscape绘制网络图





具体操作见视频： Cytoscape最新视频教程 | 一文学会Cytoscape

分析结果导入Cytoscape绘制网络图

cytoscape/go/ehbio.DESeq2.all.DE.entrez.all.Hallmark.xls.fornetwork.txt

GO富集分析结果 (网络文件,network)

ID	geneID
HALLMARK_P53_PATHWAY	CDKN2B
HALLMARK_P53_PATHWAY	HSPA4L
HALLMARK_P53_PATHWAY	SDC1
HALLMARK_P53_PATHWAY	HMOX1
HALLMARK_P53_PATHWAY	RRAD
HALLMARK_P53_PATHWAY	ZNF365

cytoscape/go/ehbio.DESeq2.all.DE.entrez.all.Hallmark.xls.fornetwork.attr

ID类型 (属性文件,table)

ID	ehbioType
HALLMARK_P53_PATHWAY	Pathway
CDKN2B	Gene
HSPA4L	Gene
SDC1	Gene
HMOX1	Gene
RRAD	Gene
ZNF365	Gene
SOCS1	Gene

cytoscape/go/ehbio_salmon.DESeq2.log2fc_ranked.symbol

基因表达 (属性文件)

Symbol	log2FoldChange
TBC1D3H	20.191
BORCS7-ASMT	20.024
AL669918.1	6.611
EEF1E1-BLOC1S5	5.894
SLC2A3P1	5.844
AC092143.1	5.539
AC092647.5	5.473
AC107982.1	5.401
AC009086.2	5.377
LINC00906	5.295

cytoscape/go/ehbio_salmon.DESeq2.normalized.symbol.txt

基因表达 (属性文件,table,用于绘制环形图)

id	untrt_N61311	untrt_N052611	untrt_N080611	untrt_N061011	trt_N61311	trt_N052611	trt_N080611	trt_N061011
FN1	245667.656692696	427435.076783868	221687.512734298	371144.225585293	240187.240865362	450103.208915937	280226.18545873	376518.228420769
DCN	212953.139271322	360796.228240728	258977.304900524	408573.064041333	210002.176051172	316009.136823338	225547.389086547	393843.741517662
CEMIP	40996.3399994438	137783.098561546	53813.9227818064	91066.8027335387	62301.1248409426	223111.851744364	212724.837607357	157919.470880358
CCDC80	137229.15270918	232772.172791659	86258.132071261	212237.323123414	136730.761365303	226070.893618867	124634.556387443	236237.808313438
IGFBP5	77812.654803177	288609.203033488	210628.865357085	168067.422122145	96021.7394882009	217439.213375834	162677.384733272	168387.360789806
COL1A1	146450.413011744	127367.25201392	152281.498327756	140861.067252416	62358.6411369943	53800.4702664978	69160.9695135734	51044.057104341
GREM1	124246.414782713	137527.206977703	217280.290691803	112502.63397527	70740.8431693814	85790.5487090611	255895.396887076	63291.6851498017
MT-RNR2	63352.8844134643	116052.899291032	177452.362713352	77960.2682345508	69491.9895908365	124660.170063963	146696.670402103	81818.4318120244
FTL	234852.946532585	197585.09713336	287309.903014121	180266.109021257	157839.859256742	143825.3394449	161717.185405949	120886.517890885
THBS1	37003.7089409061	51260.1709570089	34506.8160753971	36896.2611897521	73328.6312858911	182993.78531324	171753.212688675	170443.866045759
COL1A2	231083.819966544	222832.051447838	235896.470203805	261359.045589391	154771.267200789	192284.416295536	158748.405656781	183571.03362711
COL3A1	107753.028584038	107063.096741461	92552.8083665902	117530.436257088	66851.7831545688	64612.7710105132	46012.4185864905	60375.2554960024
ACTB	55781.4419623163	70102.1163071557	53735.144967013	57956.4766940436	86717.4766968344	121231.813000858	147983.126119645	96333.806445312
MT-CO1	214874.929060859	224656.774233706	341607.297448607	259413.453599984	250572.507177251	262173.621622476	292252.084432335	270444.335732747
FSTL1	42437.0406844558	60334.1747718511	41642.5009992771	56419.534542056	86325.4188546704	78378.726713487	65956.053254028	126447.627675394

Sequencing costs a lot and gains more

扫码关注生信宝典，学习更多生信知识

扫码关注宏基因组，获取专业学习资料

易生信，没有难学的生信知识