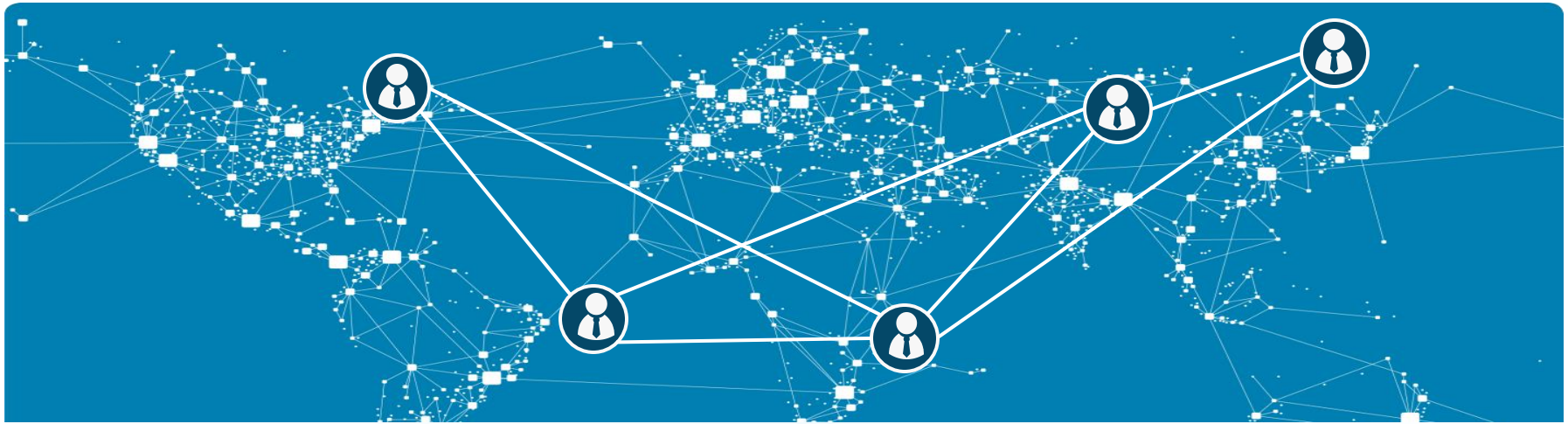
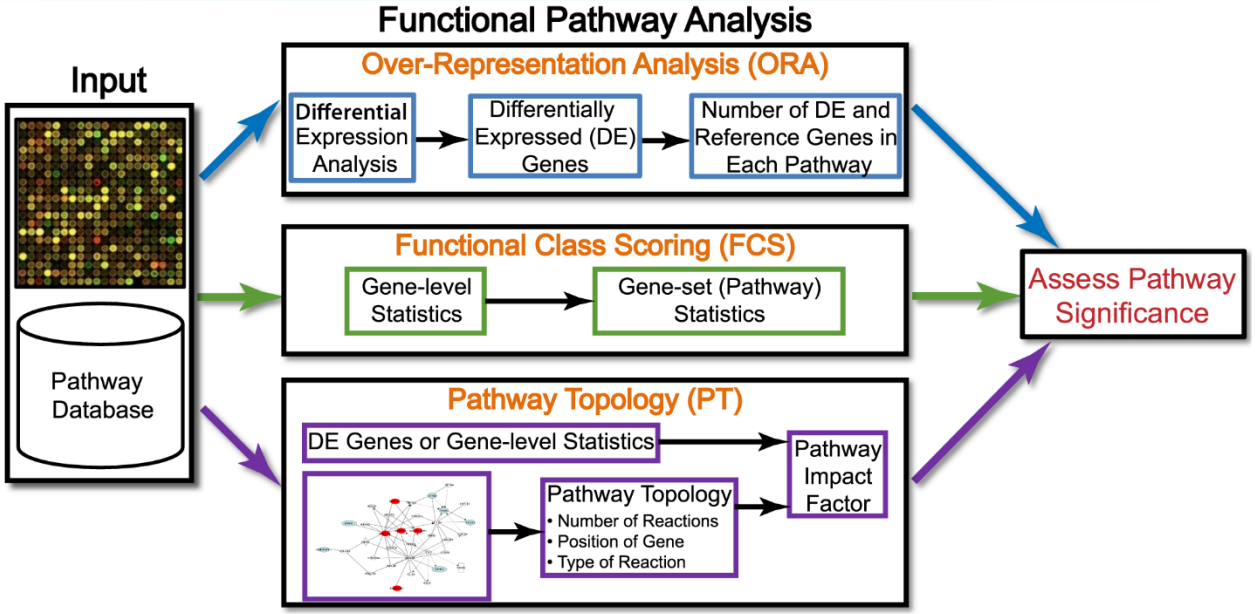


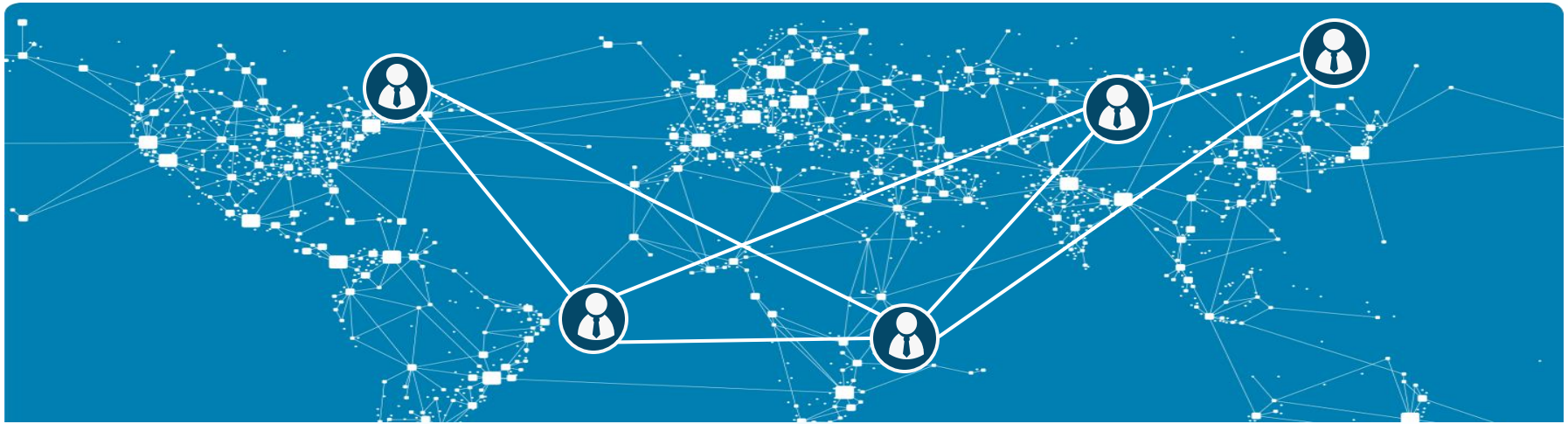
GSEA富集分析

配套视频

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。





GSEA富集分析

基因功能富集分析 – GSEA

给定一个排序的基因表 L 和一个预先定义的基因集 S (比如编码某个代谢通路的产物的基因，基因组上物理位置相近的基因，或同一GO注释下的基因)，GSEA的目的是判断 S 里面的成员 s 在 L 里面是随机分布还是主要聚集在 L 的顶部或底部。这些基因排序的依据是其在不同表型状态下的表达差异，若研究的基因集 S 的成员显著聚集在 L 的顶部或底部，则说明此基因集成员对表型的差异有贡献，也是我们关注的基因集。

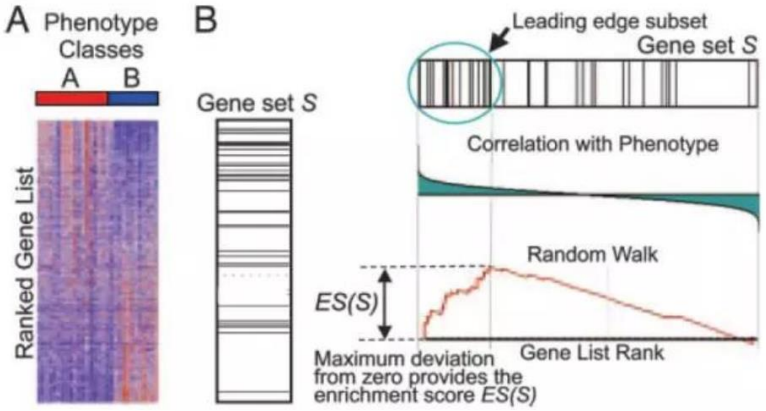
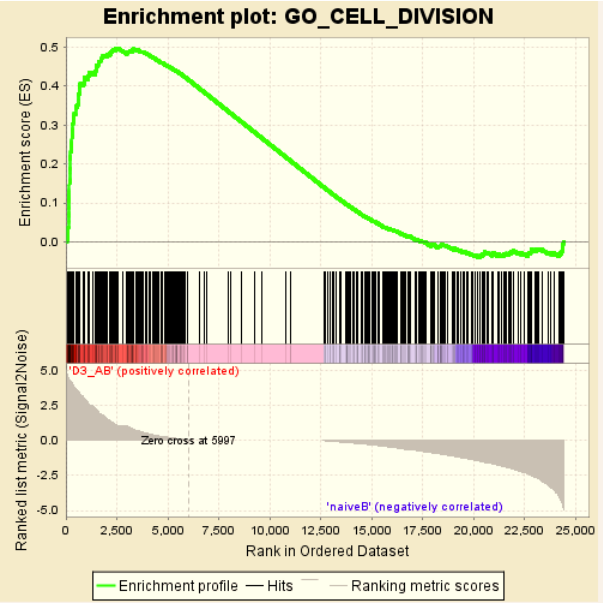


Fig. 1. A GSEA overview illustrating the method. (*A*) An expression data set sorted by correlation with phenotype, the corresponding heat map, and the "gene tags," i.e., location of genes from a set *S* within the sorted list. (*B*) Plot of the running sum for *S* in the data set, including the location of the maximum enrichment score (*ES*) and the leading-edge subset.

一文掌握GSEA，超详细教程

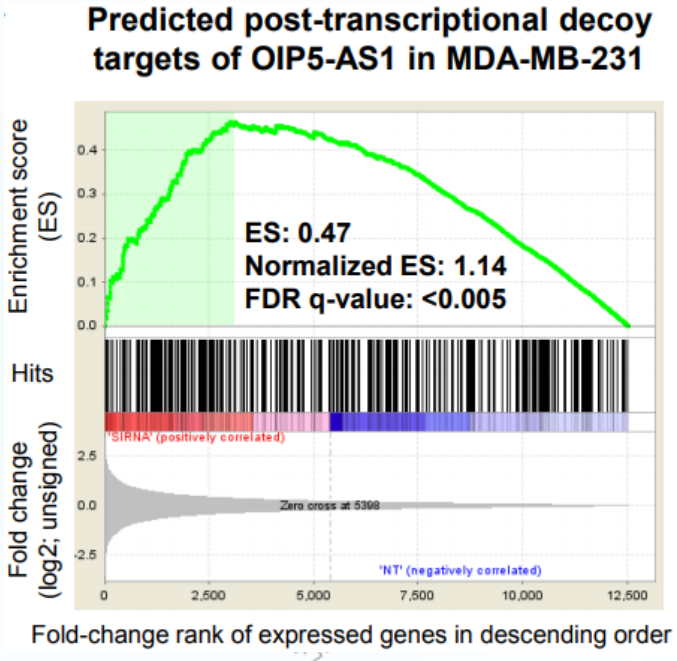
基因功能富集分析 – GSEA

计算富集得分（ES，enrichment score）. ES反应基因集成员 s 在排序列表 L 的两端富集的程度。计算方式是，从基因集 L 的第一个基因开始，计算一个累计统计值。当遇到一个落在 s 里面的基因，则增加统计值。遇到一个不在 s 里面的基因，则降低统计值。每一步统计值增加或减少的幅度与基因的表达变化程度（更严格的是与基因和表型的关联度）是相关的。富集得分ES最后定义为最大的峰值。正值ES表示基因集在列表的顶部富集，负值ES表示基因集在列表的底部富集。



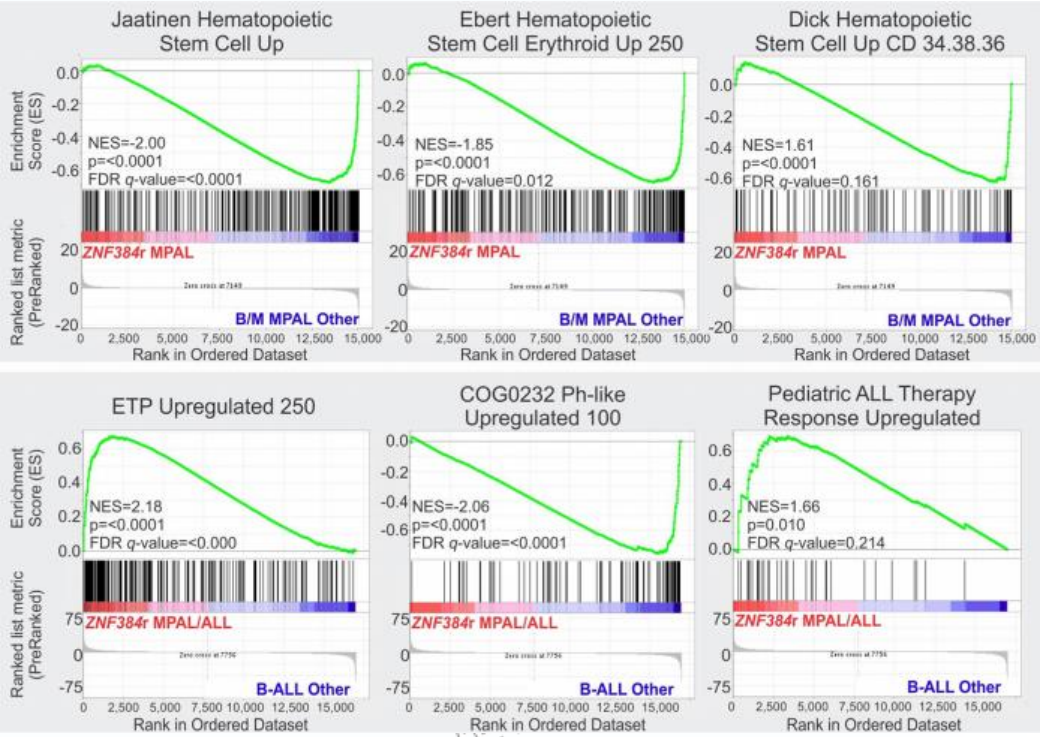
GSEA – unsigned log2 Fold Change

- GSEA analysis of expressed predicted post-transcriptional decoy targets in MDA-MB-231; all expressed genes were sorted by siOIP5-AS1 to NT absolute fold changes, GSEA used weighted enrichment statistics and ratio of classes, with p-values computed using 1k gene-set permutations.



GSEA – 排序好的基因和自定义注释集

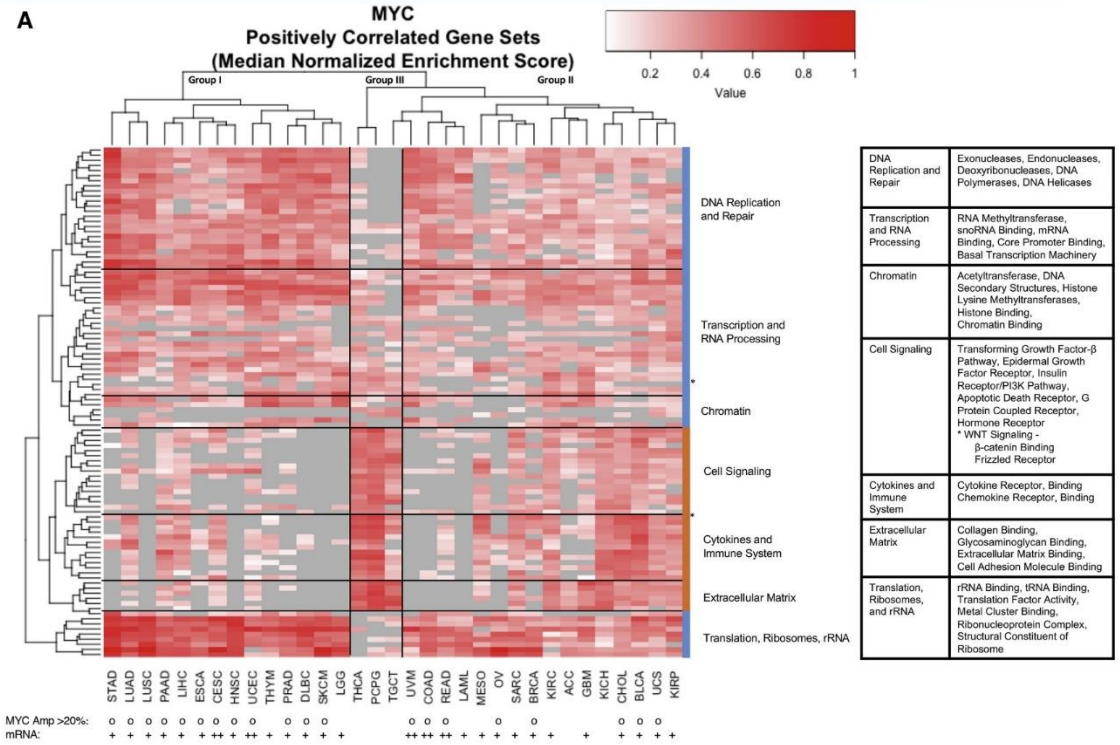
- HSC gene sets are negatively enriched in ZNF384r.
- GSEA of all ZNF384r cases positive enrichment for genes upregulated in ETP-ALL (a stem cell leukaemia), and negative enrichment for genes upregulated in Ph-like ALL in other B-ALL cases.



Nature FigS6

与Myc表达相的基因的GSEA富集分析

- 筛选与Myc/MycN表达显著正相关和负相关的基因
- 相关性系数作为加权值进行GSEA分析
- 901个MF GO条目去冗余为396个MF条目，用做检测基因集(条目基因集重合小于75%)
- 热图每一行为一个条目，每一列为一种癌症，颜色表示normalized enrichment score



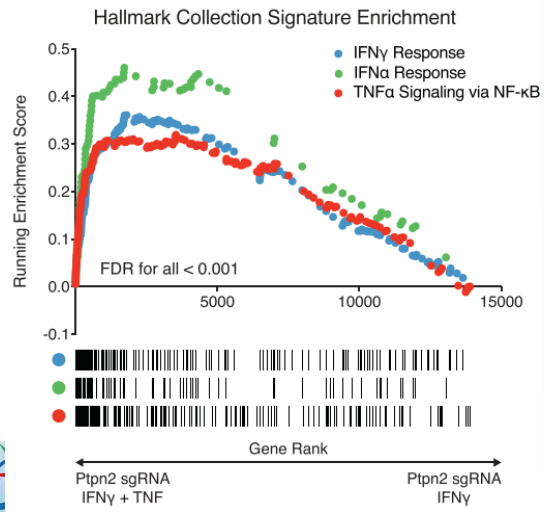
DNA Replication and Repair	Exonucleases, Endonucleases, Deoxyribonucleases, DNA Polymerases, DNA Helicases
Transcription and RNA Processing	RNA Methyltransferase, snoRNA Binding, mRNA Binding, Core Promoter Binding, Basal Transcription Machinery
Chromatin	Acetyltransferase, DNA Secondary Structures, Histone Lysine Methyltransferases, Histone Binding, Chromatin Binding
Cell Signaling	Transforming Growth Factor-β Pathway, Epidermal Growth Factor Receptor, Insulin Receptor/PI3K Pathway, Apoptotic Death Receptor, G Protein Coupled Receptor, Hormone Receptor * WNT Signaling - β-catenin Binding Frizzled Receptor
Cytokines and Immune System	Cytokine Receptor, Binding Chemokine Receptor, Binding
Extracellular Matrix	Collagen Binding, Glycosaminoglycan Binding, Extracellular Matrix Binding, Cell Adhesion Molecule Binding
Translation, Ribosomes, and rRNA	rRNA Binding, tRNA Binding, Translation Factor Activity, Metal Cluster Binding, Ribonucleoprotein Complex, Structural Constituent of Ribosome

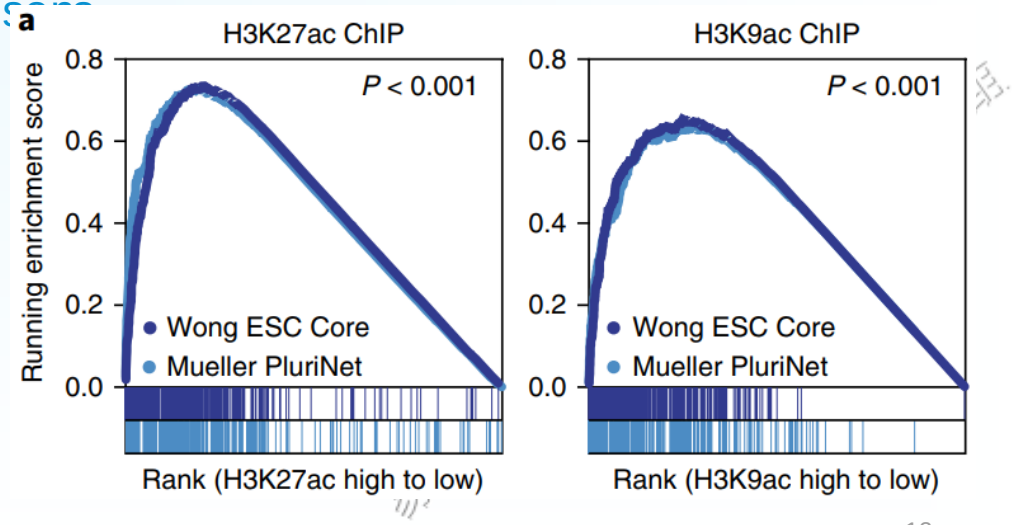
Cell

计算和绘图代码

Epigenetic modification level and GSEA

- Gene set enrichment plot showing that genes associated with high levels of H3K9ac and H3K27ac are enriched for two independently defined pluripotency gene sets: Muller PluriNet and Wong ESC Core
- P values are calculated based on 1,000 permutations by the GSEA algorithm and were not adjusted for multiple comparisons





Nature cell biology

GSEA分析软件安装（4.3）

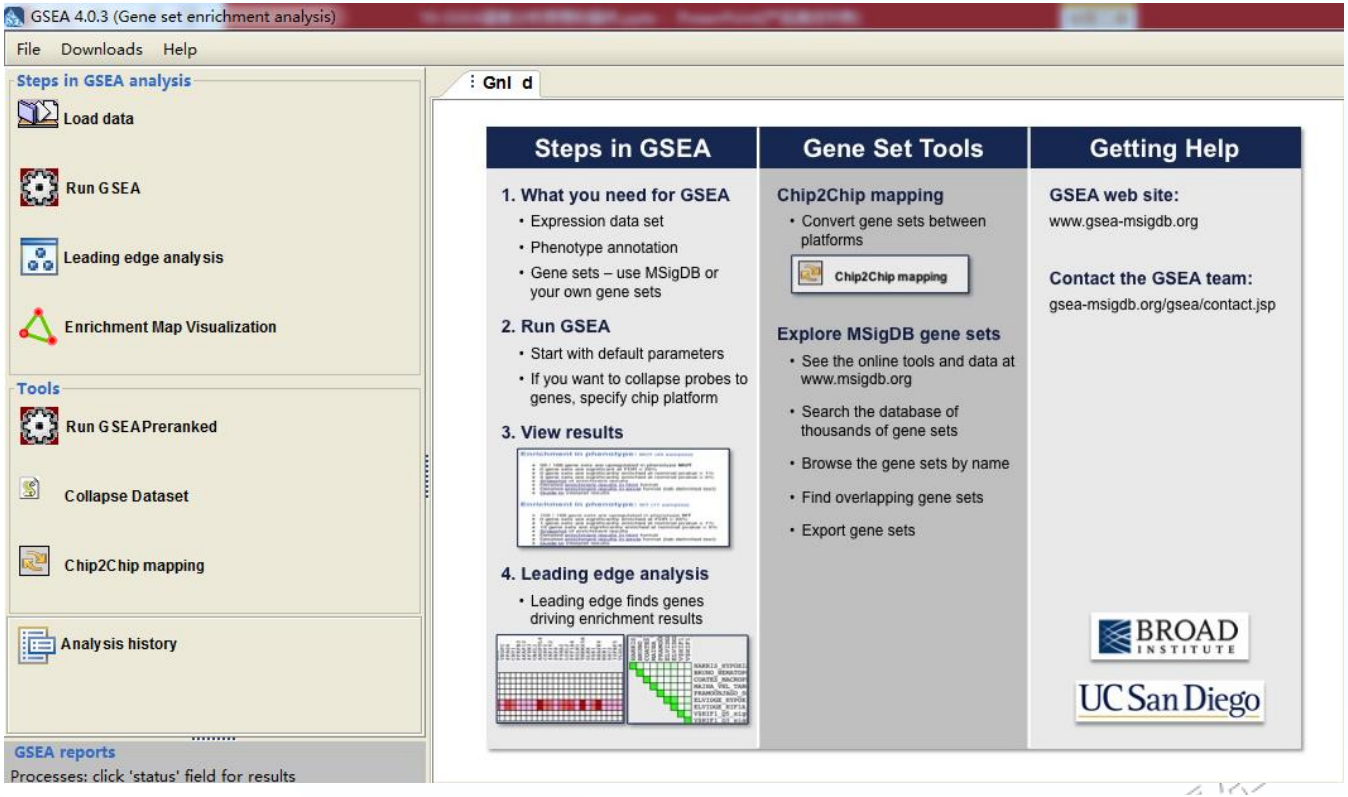
易生信

GSEA v4.3.2 Mac App	Download and unzip the Mac App Archive then double-click the GSEA application to run it. You can move the app to the Applications folder or anywhere else.	download GSEA_MacApp_4.3.2.app.zip
GSEA v4.3.2 for Windows	Download and run the installer. A GSEA shortcut will be created on the Desktop; double-click it to run the application. **64-bit Windows is required**	download GSEA_Win_4.3.2-installer.exe
GSEA v4.3.2 for Linux	Download and unzip the Archive. See the included readme.txt for further instructions. **64-bit Linux is required**	download GSEA_Linux_4.3.2.zip
GSEA v4.3.2 for the command line (all platforms)	Download and unzip the Archive. See the included readme.txt for further instructions. **Requires separate Java 11 installation.**	download GSEA_4.3.2.zip
GenePattern GSEA Module	Use GSEA from within GenePattern (a powerful and flexible analysis platform developed at the Broad Institute and UCSD) in concert with a large suite of other analytics.	
MSigDB XML Browser	The MSigDB XML Browser (formerly part of the main GSEA application) is available from our Development Snapshot builds area. We are no longer releasing official versions of this application.	
Revised GSEA R script	The original GSEA R script from 2005 was revised in 2019 to run on current versions of R. This updated version is available on GitHub. The original script is available from our Archived Downloads page. Note that neither of these GSEA R scripts are actively supported by the GSEA-MSigDB team; we recommend use of the GSEA software provided above.	
Development Snapshot builds	Development Snapshot builds of the above. These are created by our automated build system from ongoing development and may change at any time with little or no QA. **Intended for advanced users only.**	
Older software versions	Older versions of our software are available from our Archived Downloads page.	

GSEA 4.3

需要先安装Java

GSEA启动界面



GSEA中文教程

输入数据 – 常规基因表达矩阵或排序矩阵

id	untrt_N61311	untrt_N052611	untrt_N080611	untrt_N061011	trt_N
FN1	245667.656692696	427435.076783868	221687.512734298	371144.22	
DCN	212953.139271322	360796.228240728	258977.304900524	408573.06	
CEMIP	40996.3399994438	137783.098561546	53813.9227818064	91066	
CCDC80	137229.15270918	232772.172791659	86258.132071261	212237.323123	
IGFBP5	77812.654803177	288609.203033488	210628.865357085	168067.42	
COL1A1	146450.413011744	127367.25201392	152281.498327756	140861.06	
GREM1	124246.414782713	137527.206977703	217280.290691803	11250	
MT-RNR2	63352.8844134643	116052.899291032	177452.362713352	77960	
FTL	234852.946532585	197585.09713336	287309.903014121	180266.109021	
THBS1	37003.7089409061	51260.1709570089	34506.8160753971	36896	
COL1A2	231083.819966544	222832.051447838	235896.470203805	26135	
COL3A1	107753.028584038	107063.096741461	92552.8083665902	11753	
ACTB	55781.4419623163	70102.1163071557	53735.144967013	57956.476	

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
DHFRP1	5.218
URGCP-MRPS24	5.177
MCHR1	4.902
AC007923.1	4.862

GSEA官方测试数据

易生信

DATASET	DESCRIPTION	RELEVANT DATA *(save link to download)*
Gender	Transcriptional profiles from male and female lymphoblastoid cell lines Results of C1 GSEA analysis of this dataset Results of C2 GSEA analysis of this dataset	Gender_hgu133a.gct Gender_collapsed.gct Gender.cls
p53	Transcriptional profiles from p53+ and p53 mutant cancer cell lines Results of C2 GSEA analysis of this dataset	P53_hgu95av2.gct P53_collapsed.gct P53.cls
Diabetes	Transcriptional profiles of smooth muscle biopsies of diabetic and normal individuals Results of C2 GSEA analysis of this dataset	Diabetes_hgu133a.gct Diabetes_collapsed.gct Diabetes.cls

测试数据

输入数据 – 样品分组信息cls文件 (定性分组)

易生信

2. 样品分组信息

- 第一行：三个数分别表示：34个样品，2个分组，最后一个数字1是固定的；
- 第二行：以 `#` 开始， `tab` 键分割，分组信息（有几个分组便写几个，多个分组在比较分析时，后面需要选择待比较的任意2组）；（样品分组中 `NGT` 表示正常耐糖者， `DMT` 表示糖尿病患者，自己使用时替换为自己的分组名字）
- 第三行：样本对应的组名。样本分组信息的第三行，同一组内的**不同重复一定要命名为相同的名字，可以是分组的名字**。例如相同处理的不同重复在自己试验记录里一般是Treat6h_1、Treat6h_2、Treat6h_3，但是在这里一定都要写成一样的值 `Treat6h` 。与表达矩阵的样品列按**位置一一对应**，名字相同的代表样品属于同一组。如果是样本分组信息，上图中的 `0` 和 `1` 也可以对应的写成 `NGT` 和 `DMT` ，更直观。但是，如果想把分组信息作为连续表型值对待，这里就**只能**提供**数字**。

```
34 2 1
# NGT DMT
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

输入数据 – 样品分组信息cls文件 (数量性状)

易生信

- 第一行: #numeric 固定写法
- 第二行: #YSX (度量指标的名字，如病人的血压、体重等)
- 第三行: 具体度量值，与表达矩阵中样品顺序一致
- 度量指标可以有多个，单次只对一个选择的指标就行分析
- **Metric for ranking genes：pearson、Cosine、Manhattan 或 Euclidean**

```
#numeric
#YSX
0 0 0 1 1 1 6 6 6 24 24 24 48 48 48
#Time
0 0 0 1 1 1 6 6 6 24 24 24 48 48 48
```

输入数据 – 基因注释gmt文件

易生信

官网提供的 `gmt` 文件有两种类型， `*.symbols.gmt` 中基因以 `symbols` 号命名， `*.entrez.gmt` 中基因以 `entrez id` 命名。注意根据表达矩阵的基因名字命名方式选择合适的基因集。

All gene sets	Current MSigDB gene sets, gene symbols	msigdb.v6.2.symbols.gmt
	Current MSigDB gene sets, Entrez IDs	msigdb.v6.2.entrez.gmt
	Current MSigDB xml file	msigdb_v6.2.xml
h: hallmark gene sets	hallmark gene sets, gene symbols	h.all.v6.2.symbols.gmt
	hallmark gene sets, Entrez IDs	h.all.v6.2.entrez.gmt
c1: positional gene sets	positional gene sets, gene symbols	c1.all.v6.2.symbols.gmt
	positional gene sets, Entrez IDs	c1.all.v6.2.entrez.gmt

`gmt` 格式是多列注释文件，第一列是基因所属基因集的名字，可以是通路名字，也可以是自己定义的任何名字。第二列，官方提供的格式是URL，可以是任意字符串。后面是基因集内基因的名字，有几个写几列。列与列之间都是 `TAB` 分割。

```
Pathway_description     Anystring     Gene1   Gene2   Gene3
Pathway_description2    Anystring     Gene4   Gene2   Gene3   Gene5
```

易汉博基
EHBIO Gene

MsigDB

易生信

H **hallmark gene sets** are coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes.

C1 **positional gene sets** for each human chromosome and cytogenetic band.

C2 **curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.

C3 **motif gene sets** based on conserved cis-regulatory motifs from a comparative analysis of the human, mouse, rat, and dog genomes.

C4 **computational gene sets** defined by mining large collections of cancer-oriented microarray data.

C5 **GO gene sets** consist of genes annotated by the same GO terms.

C6 **oncogenic gene sets** defined directly from microarray gene expression data from cancer gene perturbations.

C7 **immunologic gene sets** defined directly from microarray gene expression data from immunologic studies.

Update human MSigDB

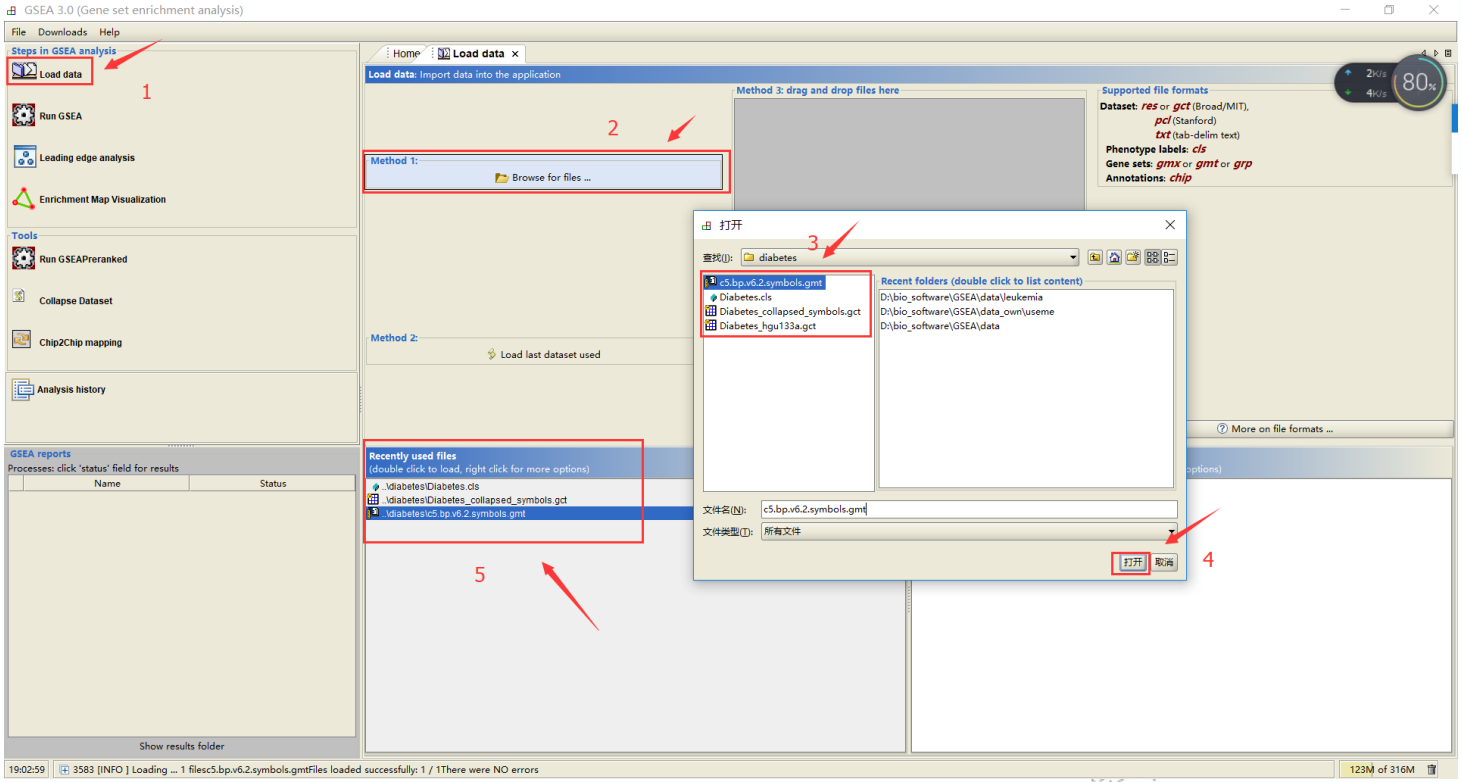
See the license terms page for details about the license for MSigDB. Please note that the license terms vary for different versions of MSigDB and that certain gene sets have special access terms.

Individual Human Gene Set GMTs	Human MSigDB v2023.2.Hs GMTs for each individual collection are available from the Human Collections page.	
Individual Human Gene Annotations Files	Individual gene annotations chip files for Human MSigDB v2023.2.Hs (as well as older versions).	view chip files
Human Gene Set GMT file set (ZIPped)	The Human MSigDB v2023.2.Hs release including all individual Gene Set collections as GMT files, provided as a single ZIP bundle.	download zip file
Human Gene Annotations file set (ZIPped)	The complete Human MSigDB v2023.2.Hs gene annotations file set, provided as a single ZIP bundle.	download zip file
Human MSigDB SQLite database (ZIPped)	The Human MSigDB v2023.2.Hs contents and metadata in the form of a (ZIPped) SQLite database. See our documentation for more details on the contents and usage. Our XML file should be considered deprecated in favor of this SQLite database and will be removed in a future release.	download database
Human Gene Set JSON file set (ZIPped)	The complete Human MSigDB v2023.2.Hs JSON file set, provided as a single ZIP bundle. These JSON files contain the Human gene sets using HUGO (HGNC) gene symbols along with some useful metadata.	download zip file
Human MSigDB XML file (ZIPped)	The Human MSigDB v2023.2.Hs XML file (ZIPped) containing all the Human MSigDB gene sets (this has been deprecated in favor of the SQLite database and will be removed in a future release).	download zip file

Update mouse MSigDB

Individual Mouse Gene Set GMTs	Mouse MSigDB v2023.2.Mm GMTs for each individual collection are available from the Mouse Collections page.	
Individual Mouse Gene Annotations Files	Individual gene annotations chip files for Mouse MSigDB v2023.2.Mm (as well as older versions).	view chip files
Mouse Gene Set file set (ZIPped)	The Mouse MSigDB v2023.2.Mm release including all individual collections as GMT files, provided as a single ZIP bundle.	download zip file
Mouse Gene Annotations file set (ZIPped)	The Mouse MSigDB v2023.2.Mm gene annotations file set, provided as a single ZIP bundle.	download zip file
Mouse MSigDB SQLite database (ZIPped)	The Mouse MSigDB v2023.2.Mm contents and metadata in the form of a (ZIPped) SQLite database. See our documentation for more details on the contents and usage. Our XML file should be considered deprecated in favor of this SQLite database and will be removed in a future release.	download zip file
Mouse Gene Set JSON file set (ZIPped)	The Mouse MSigDB v2023.2.Mm JSON file set, provided as a single ZIP bundle. These JSON files contain the Mouse gene sets using MGI gene symbols along with some useful metadata.	download zip file
Mouse MSigDB XML file (ZIPped)	The Mouse MSigDB v2023.2.Mm XML file (ZIPped) containing all the Mouse MSigDB gene sets (this has been deprecated in favor of the SQLite database and will be removed in a future release).	download zip file

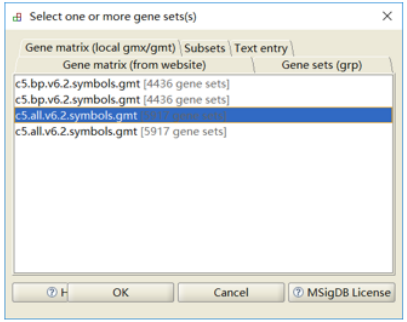
GSEA表达矩阵、分组信息、注释信息同时导入

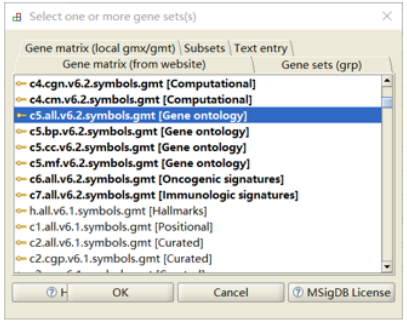


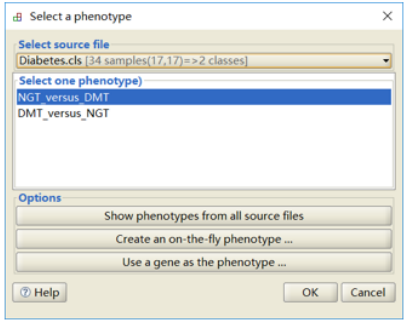
GSEA选择合适的参数 – 导入的文件

Required fields

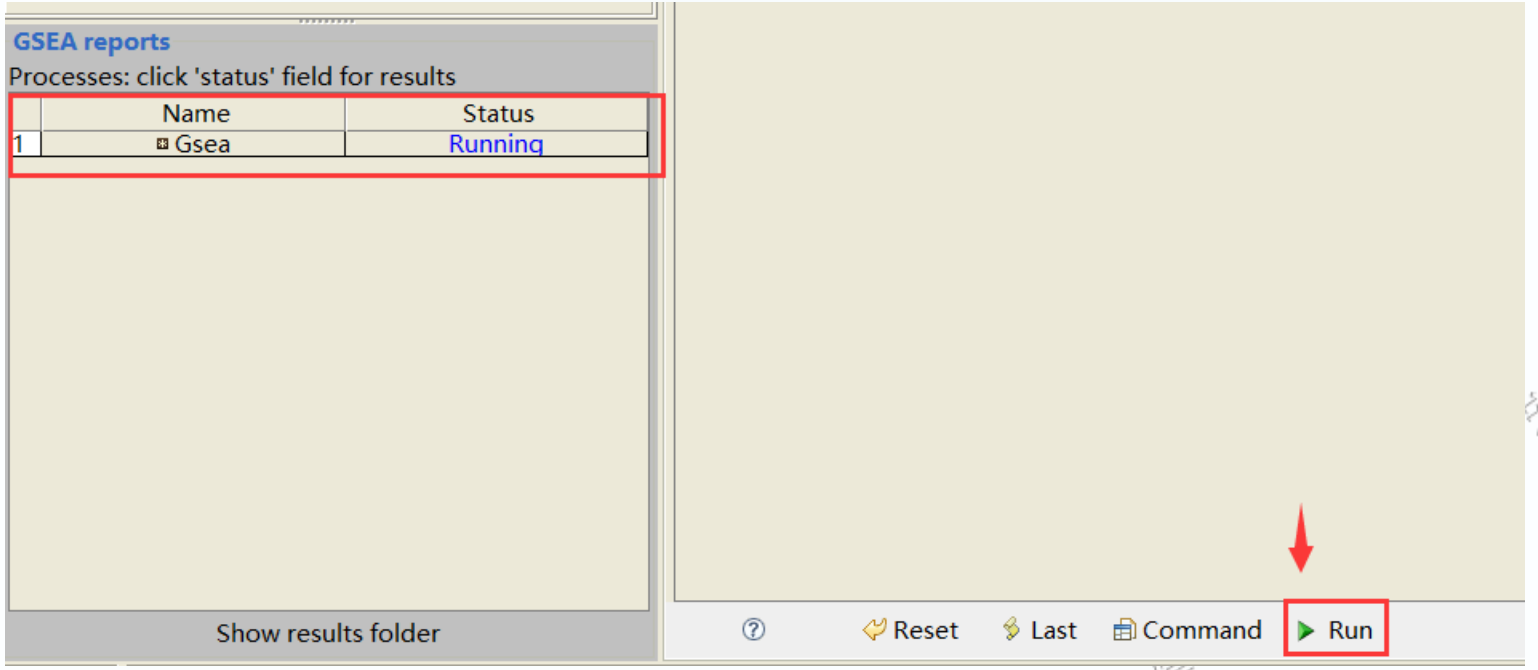
Expression dataset	Diabetes_collapsed_symbols [15056x34 (ann: 15056,34,chip na)]
Gene sets database	ftp.broadinstitute.org://pub/gsea/gene_sets_final/c5.all.v6.2.symbols.gmt
Number of permutations	1000
Phenotype labels	D:\bio_software\GSEA\data\diabetes\Diabetes.cls#NGT_versus_DMT
Collapse dataset to gene symbols	false
Permutation type	phenotype
Chip platform	







GSEA run


GSEA reports
Processes: click 'status' field for results
Name
Status
1
Gsea
Running
Show results folder
Reset
Last
Command
Run

GSEA结果解读

易生信

Enrichment in phenotype: NGT (17 samples)

- 1697 / 3953 gene sets are upregulated in phenotype **NGT**
- 36 gene sets are significant at FDR < 25%
- 19 gene sets are significantly enriched at nominal pvalue < 1%
- 114 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Enrichment in phenotype: DMT (17 samples)

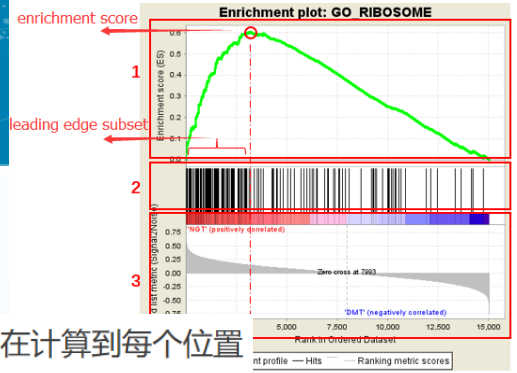
- 2256 / 3953 gene sets are upregulated in phenotype **DMT**
- 0 gene sets are significantly enriched at FDR < 25%
- 13 gene sets are significantly enriched at nominal pvalue < 1%
- 97 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

GSEA结果总结表

Table: Gene sets enriched in phenotype DMT (17 samples) [plain text format]

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
1	GO_TRIGLYCERIDE_RICH_LIPOPROTEIN_PARTICLE	Details ...	15	-0.64	-1.88	0.004	1.000	0.584	1520	tags=47%, list=10%, signal=52%
2	GO_SMOOTH_ENDOPLASMIC_RETICULUM	Details ...	27	-0.54	-1.86	0.000	1.000	0.656	2503	tags=41%, list=17%, signal=49%
3	GO_CADHERIN_BINDING	Details ...	26	-0.61	-1.85	0.002	0.866	0.709	2817	tags=50%, list=19%, signal=61%
4	GO_MUSCLE_CELL_CELLULAR_HOMEOSTASIS	Details ...	15	-0.60	-1.77	0.004	1.000	0.917	2516	tags=47%, list=17%, signal=56%
5	GO_PLASMA_LIPOPROTEIN_PARTICLE_CLEARANCE	Details ...	17	-0.63	-1.75	0.006	1.000	0.962	508	tags=35%, list=3%, signal=36%
6	GO_REGULATION_OF_CARDIAC_MUSCLE_CELL_ACTION_POTENTIAL	Details ...	15	-0.58	-1.74	0.010	1.000	0.965	635	tags=33%, list=4%, signal=35%
7	GO_REGULATION_OF_CGMP_METABOLIC_PROCESS	Details ...	22	-0.51	-1.73	0.002	1.000	0.968	756	tags=32%, list=5%, signal=33%
8	GO_CELL_CELL_ADHERENS_JUNCTION	Details ...	44	-0.49	-1.73	0.012	1.000	0.974	3465	tags=45%, list=23%, signal=59%
9	GO_POSITIVE_REGULATION_OF_LIPID_CATABOLIC_PROCESS	Details ...	23	-0.56	-1.72	0.014	1.000	0.981	1317	tags=35%, list=9%, signal=38%
10	GO_POSITIVE_REGULATION_OF_TRANSLATIONAL_INITIATION	Details ...	18	-0.59	-1.72	0.004	1.000	0.981	1832	tags=39%, list=12%, signal=44%
11	GO_REGULATION_OF_CGMP_BIOSYNTHETIC_PROCESS	Details ...	16	-0.55	-1.71	0.006	1.000	0.988	2045	tags=44%, list=14%, signal=51%
12	GO_STRUCTURAL_CONSTITUENT_OF_MUSCLE	Details ...	32	-0.62	-1.70	0.025	1.000	0.989	3104	tags=59%, list=21%, signal=75%
13	GO_MHC_PROTEIN_BINDING	Details ...	22	-0.57	-1.70	0.014	0.985	0.989	1747	tags=36%, list=12%, signal=41%
14	GO_SARCOMERE_ORGANIZATION	Details ...	22	-0.64	-1.69	0.035	0.985	0.991	3104	tags=59%, list=21%, signal=74%
15	GO_STRUCTURAL_CONSTITUENT_OF_EYE_LENS	Details ...	19	-0.57	-1.69	0.020	0.987	0.993	2990	tags=53%, list=20%, signal=66%
16	GO_MYOFIBRIL_ASSEMBLY	Details ...	38	-0.57	-1.67	0.026	1.000	0.995	3204	tags=50%, list=21%, signal=63%
17	GO_ACTOMYOSIN_STRUCTURE_ORGANIZATION	Details ...	62	-0.48	-1.67	0.016	0.977	0.995	3270	tags=42%, list=22%, signal=53%
18	GO_NUCLEAR_NUCLEOSOME	Details ...	31	-0.50	-1.67	0.030	0.938	0.997	3353	tags=45%, list=22%, signal=58%
19	GO_MULTICELLULAR_ORGANISMAL_MOVEMENT	Details ...	33	-0.51	-1.66	0.025	0.948	0.998	3137	tags=42%, list=21%, signal=53%
20	GO_REGULATION_OF_STEROID_HORMONE_SECRETION	Details ...	18	-0.54	-1.66	0.014	0.943	0.998	1984	tags=39%, list=13%, signal=45%
21	GO_REGULATION_OF_LIPID_TRANSPORT		81	-0.39	-1.64	0.006	1.000	0.999	3634	tags=40%, list=24%, signal=52%
22	GO_REGULATION_OF_ANION_TRANSPORT		114	-0.37	-1.64	0.002	1.000	0.999	3567	tags=39%, list=24%, signal=50%
23	GO_RESPONSE_TO_HEAT		71	-0.41	-1.64	0.012	1.000	0.999	3626	tags=37%, list=24%, signal=48%
24	GO_DEATH_RECEPTOR_ACTIVITY		17	-0.58	-1.62	0.026	1.000	0.999	3404	tags=59%, list=23%, signal=76%

GSEA结果解释



- 第一部分是 `Enrichment score` 折线图：显示了当分析沿着排名列表按排序计算时，ES值在计算到每个位置时的展示。最高峰处的得分 (垂直距离0.0最远)便是基因集的ES值。
- 第二部分，用线条标记了基因集合中成员出现在基因排序列表中的位置，黑线代表排序基因表中的基因存在于当前分析的功能注释基因集。**leading edge subset** 就是（0,0）到绿色曲线峰值ES出现对应的这部分基因。
- 第三部分是排序后所有基因rank值得分布，热图红色部分对应的基因在 `NGT` 中高表达，蓝色部分对应的基因在 `DMT` 中高表达，每个基因对应的信噪比（`Signal2noise`，前面选择的排序值计算方式）以灰色面积图显展示。

在上图中，我们一般关注ES值，峰出现在排序基因集的前端还是后端（ES值大于0在前端，小于0在后端）以及 `Leading edge subset` （即对富集贡献最大的部分，领头亚集）；在ES图中出现领头亚集的形状，表明这个功能基因集在某处理条件下具有更显著的生物学意义；对于分析结果中，我们一般认为 `|NES|>1, NOM p-val<0.05, FDR q-val<0.25` 的通路是显著富集的。

GSEA参数解释

易生信

Basic fields — Hide

Analysis name	my_analysis
Enrichment statistic	weighted
Metric for ranking genes	Signal2Noise
Gene list sorting mode	real
Gene list ordering mode	descending
Max size: exclude larger sets	500
Min size: exclude smaller sets	15
Save results in this folder	D:\bio_software\GSEA\data\diabetes

GSEA参数解释 – rank metrics

易生信

Basic fields Hide

Analysis name	my_analysis
Enrichment statistic	weighted
Metric for ranking genes	Signal2Noise

- 如果表型是分组信息，GSEA在计算分组间的差异值时支持5种统计方式，分别是 `signal2noise`、`t-Test`、`ratio_of_class`、`diff_of_class` (log2转换后的值计算倍数)和 `log2_ratio_of_class`。下面公式很清楚。

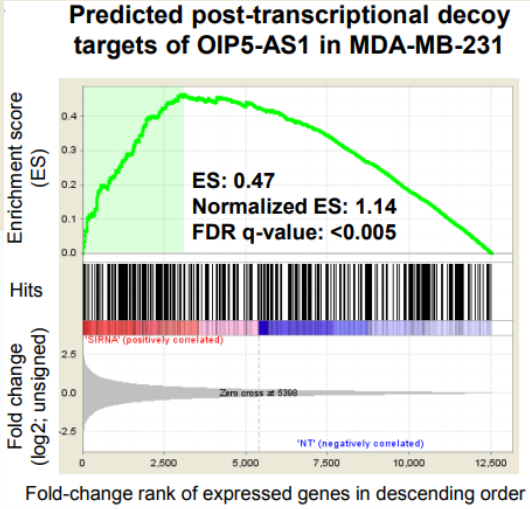
(1) $\frac{\mu_A - \mu_B}{\sigma_A + \sigma_B}$ (2) $\frac{\mu_A - \mu_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}}$ (3) $\frac{\mu_A}{\mu_B}$ (4) $\mu_A - \mu_B$ (5) $\log 2\left(\frac{\mu_A}{\mu_B}\right)$

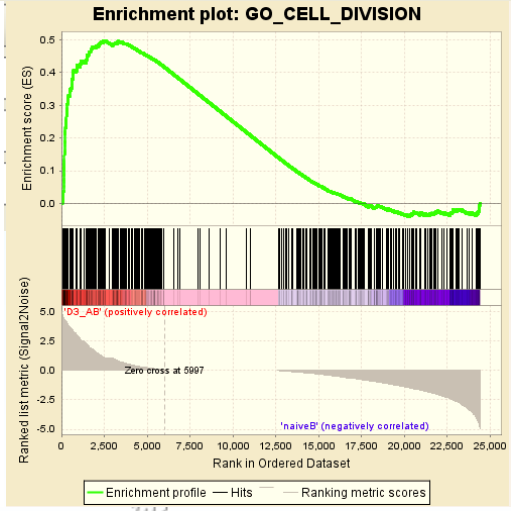
- 如果表型是连续数值信息（定量表型）：GSEA通过表型文件（`cls`）和表达数据集文件（`gct`），使用 `pearson相关性`、`Cosine`、`Manhattan` 或 `Euclidean` 指标之一计算两个配置文件之间的相关性。（注意：若是分组表型文件想转换为定量表型，cls文件中分类标签应该指定为数字)

GSEA参数解释 – sort mode

Basic fields — Hide

Analysis name	my_analysis
Enrichment statistic	weighted
Metric for ranking genes	Signal2Noise
Gene list sorting mode	real
Gene list ordering mode	
Max size: exclude larger sets	
Min size: exclude smaller sets	
Save results in this folder	





GSEA输出 – 矢量图，图形数目

易生信

Advanced fields Hide

Collapsing mode for probe sets => 1 gene	Max_probe
Normalization mode	meandiv
Randomization mode	no_balance
Alternate delimiter	
Create GCT files	false
Create SVG plot images	false
Omit features with no symbol match	true
Make detailed gene set report	true
Median for class metrics	false
Number of markers	100
Plot graphs for the top sets of each phenotype	20
Seed for permutation	timestamp
Save random ranked lists	false
Make a zipped file with all reports	false

GSEA分析自定义数据

易生信

	A	B	C	D	E	F	G	H	I
1	10-formyltetrahydrofolate biosynthetic process	GO:0009257	AT1G50480	AT2G12280	AT2G16370	AT4G34570			
2	(1->3)-beta-D-glucan biosynthetic process	GO:0006075	AT1G05570	AT1G06490	AT2G13680	AT2G31960	AT2G36850	AT3G07160	AT3G14570
3	1-aminocyclopropane-1-carboxylate biosynthetic process	GO:0042218	AT1G01480	AT2G22810	AT4G37770	AT5G65800			
4	1-deoxy-D-xylulose 5-phosphate biosynthetic process	GO:0052865	AT4G15560						
5	1-methylguanosine metabolic process	GO:0080179	AT5G47680						
6	2,4,6-trinitrotoluene catabolic process	GO:0046256	AT1G17170	AT1G17180					
7	2'-deoxyribonucleotide metabolic process	GO:0009394	AT3G46940						
8	[2Fe-2S] cluster assembly	GO:0044571	AT5G06410	AT5G65720					
9	2-methylguanosine metabolic process	GO:0080180	AT3G26410						
10	3-keto-sphinganine metabolic process	GO:0006666	AT3G06060	AT5G19200					
11	3'-UTR-mediated mRNA destabilization	GO:0061158	AT1G32360	AT1G66810	AT1G68200	AT2G35430	AT3G12680	AT3G19360	AT3G20250
12	3'-UTR-mediated mRNA stabilization	GO:0070935	AT2G17975						
13	4,4-dimethyl-9beta,19-cyclopropylsterol oxidation	GO:0080064	AT4G12110	AT4G22753	AT4G22756				
14	4-alpha-methyl-delta7-sterol oxidation	GO:0080065	AT1G07420	AT2G29390					
15	5-carbamoylmethyl uridine residue modification	GO:0080178	AT1G13870	AT5G13680					
16	5-phosphoribose 1-diphosphate biosynthetic process	GO:0006015	AT1G10700	AT1G32380	AT2G35390	AT2G42910	AT2G44530		
17	5S class rRNA transcription by RNA polymerase III	GO:0042791	AT1G58766						
18	7,8-dihydroneopterin 3'-triphosphate biosynthetic process	GO:0035998	AT3G07270						
19	7-methylguanosine cap hypermethylation	GO:0036261	AT1G45231						
20	7-methylguanosine metabolic process	GO:0008618	AT1G03110						

单基因GSEA

易生信

- 拿单个基因 （一般是感兴趣的基因）作为分组方式，探索与给定的单个基因相关的 （可以是表达相关，也可以是其它相关）基因富集在哪些调控通路和分子功能。
- 分组方法有两种，一种是定性分组，一种是定量相关。
- 具体见推文：链接

Sequencing costs a lot and gains more

易生信

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