

# **PRACTICAL**

唐凯临

2021.6

#### REVIEW: DATA

Platforms (1)

Samples (15) ■ Less...

#### Data table header descriptions

ID

TRANSCRIPT\_TYPE

ACC

ORF

SOURCE BUILD protein\_coding or noncoding or smallRNA

Accession number of spotted sequence

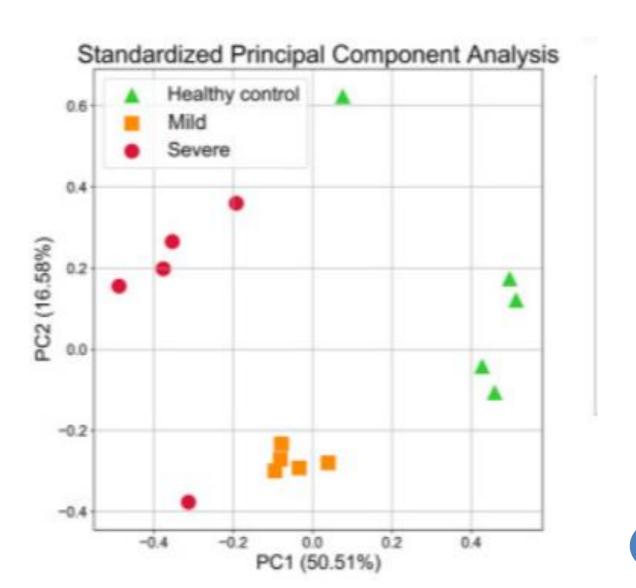
source database Genome Version

```
GPL26963 Agilent-085982 Arraystar human IncRNA V5 microarray
```

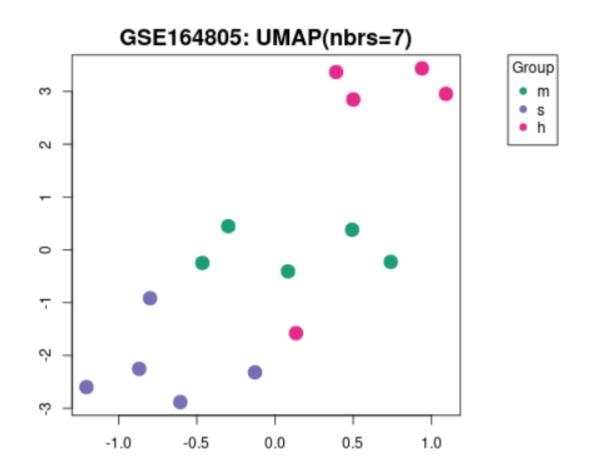
```
GSM5019817 PBMC_healthy_subject1
GSM5019818 PBMC_healthy_subject2
GSM5019819 PBMC_healthy_subject3
GSM5019820 PBMC_healthy_subject4
GSM5019821 PBMC_healthy_subject5
GSM5019822 PBMC_mild_patient_subject1
GSM5019823 PBMC_mild_patient_subject2
GSM5019824 PBMC_mild_patient_subject3
GSM5019825 PBMC_mild_patient_subject4
GSM5019826 PBMC_mild_patient_subject5
GSM5019827 PBMC_severe patient_subject1
GSM5019828 PBMC_severe patient_subject2
GSM5019829 PBMC_severe patient_subject3
GSM5019830 PBMC_severe patient_subject4
GSM5019831 PBMC_severe patient_subject5
```

## **REVIEW**

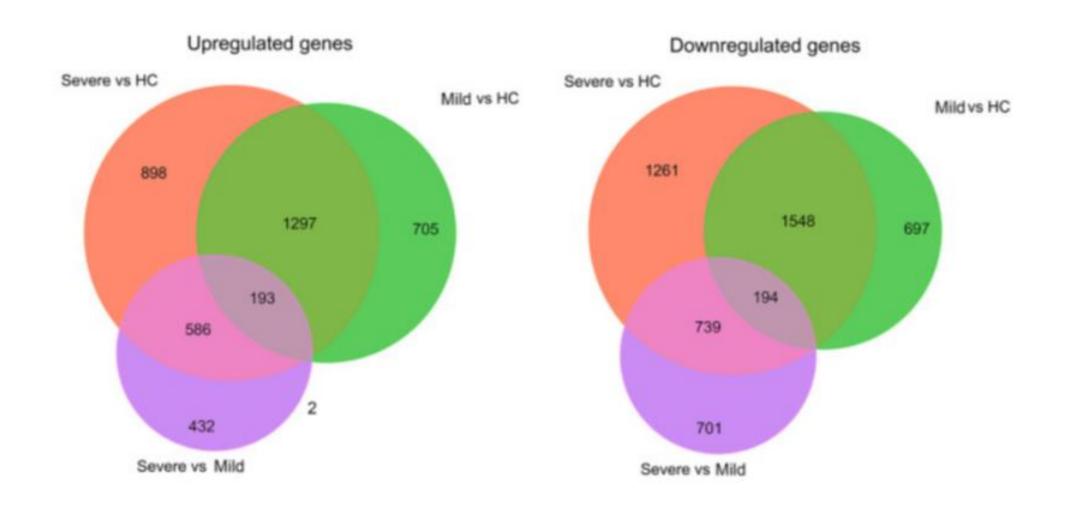
Normalization



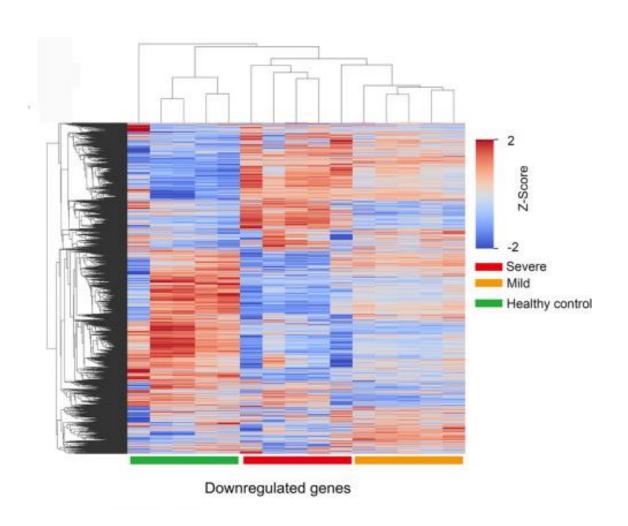
# GEO2R



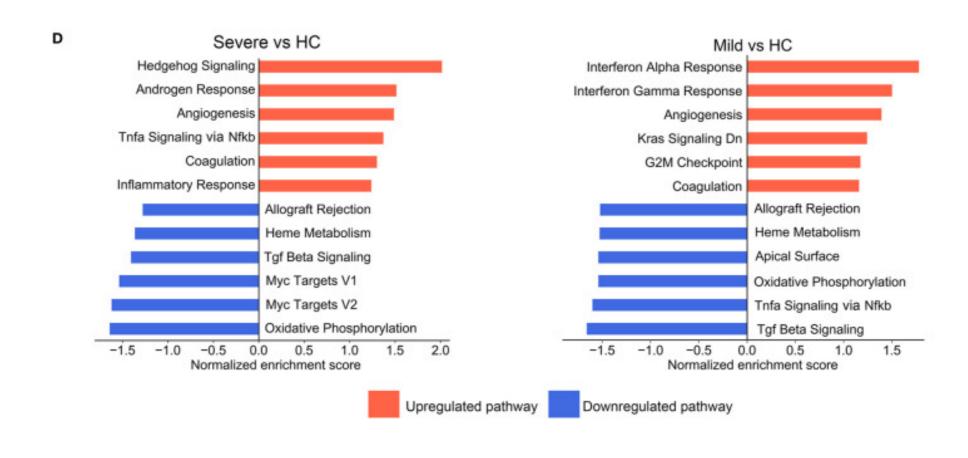
# **DEGS**



# ONLY DEG IS NOT ENOUGH



# ONLY DEG IS NOT ENOUGH



# 常用工具

- 热图绘制
- 功能富集
- 网络构建

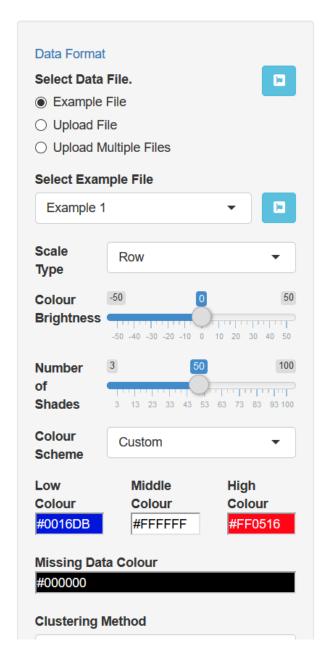
# DATA FORMAT

| JNIQID       | d1-1       | d1-2  | d1-3 | d1-4  | d2-1  | d2-2  | d2-3  | d2-4  | d2-5  | d3-1 | d3-2 | d3-3 | d3-4 |
|--------------|------------|-------|------|-------|-------|-------|-------|-------|-------|------|------|------|------|
| <b>GPNMB</b> | 6.07       | 5.18  | 6.74 | 6.75  | 6.61  | 6.49  | 6.93  | 6.71  | 6.88  | 6.65 | 6.98 | 7.37 | 6.84 |
| ANKRD3       | 6BP2       | 8.32  | 8.11 | 9.13  | 8.92  | 7.37  | 7.3   | 7.25  | 7.29  | 7.26 | 7.52 | 7.93 | 7.96 |
| KRT19        | 11         | 11.18 | 9.89 | 10.61 | 11.18 | 10.83 | 10.79 | 10.58 | 11.26 | 8.64 | 9.1  | 9.27 | 9.13 |
| SLC6A1       | 5.42       | 4.18  | 3.59 | 4.27  | 4.68  | 4.7   | 4.87  | 4.71  | 4.6   | 4.99 | 4.49 | 4.35 | 4.33 |
| <b>GNMT</b>  | 4.68       | 4.2   | 4.11 | 4.38  | 4.76  | 3.99  | 4.39  | 4.72  | 4.49  | 5.99 | 5.15 | 4.37 | 5.4  |
| GLV1-40      | )///IGLV1- | 44    | 8.9  | 8.99  | 9.7   | 9.65  | 9.6   | 9.91  | 9.54  | 9.33 | 9.55 | 9.24 | 9.53 |
| KRT23        | 8.1        | 8.39  | 8.45 | 8.67  | 9.2   | 9.62  | 9.14  | 9.45  | 9.54  | 7.78 | 7.87 | 7.68 | 7.96 |
| KCNN2        | 4.43       | 4     | 4.15 | 4.27  | 4.22  | 4.72  | 4.83  | 4.99  | 4.45  | 4.22 | 4.15 | 4.05 | 4.09 |
| SLCO4C1      | 4.84       | 4.36  | 4.47 | 4.7   | 4.08  | 4.16  | 4.27  | 4.35  | 4.22  | 4.48 | 4.33 | 4.59 | 4.36 |
| FCRL5        | 7.46       | 7.92  | 8.43 | 7.99  | 7.1   | 6.51  | 6.63  | 6.33  | 6.45  | 7.39 | 7.65 | 7.46 | 7.44 |
| PPP1R1A      | 4.31       | 4.42  | 4.26 | 4.43  | 4.21  | 4.31  | 3.88  | 3.99  | 4.51  | 4.98 | 4.57 | 4.18 | 4.59 |
| ACSM2A       | 3.48       | 3.41  | 3.65 | 3.56  | 4.73  | 4.63  | 4.17  | 4.67  | 4.16  | 5.97 | 5.56 | 5.43 | 6.1  |
|              |            |       |      |       |       |       |       |       |       |      |      |      |      |

# HEATMAP

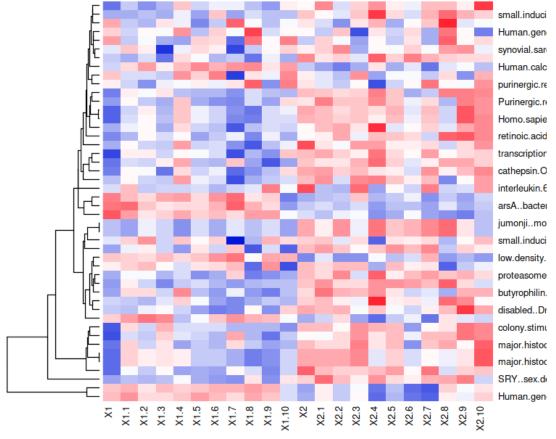
o <a href="http://www.heatmapper.ca/expression/">http://www.heatmapper.ca/expression/</a>

| UNIQID    | NAME                                 | 1            | 1            | 1            | 2            | 2            | 2            |
|-----------|--------------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Hs.9305   | angiotensin receptor-like 1          | 0.533865724  | 2.667389535  | 1.98516763   | 4.832817262  | 3.590801314  | 3.309607441  |
| Hs.181307 | H3 histone, family 3A                | -0.942951315 | -0.500171526 | -0.705453196 | -0.32191231  | -0.14030858  | -0.427467964 |
| Hs.83484  | SRY (sex determining region Y)-box 4 | -3.418094026 | -3.181358856 | -3.276305367 | -2.392780506 | -1.740889988 | -2.248737455 |



Plot Interactive Row Dendrogram Column Dendrogram Table





#### HEATMAP

Open







Visit Cancer Dependency Map or cBioPortal for additional datasets

#### TCGA (Broad GDAC 1/28/2016)

Acute Myeloid Leukemia (LAML)

Adrenocortical carcinoma (ACC)

Bladder Urothelial Carcinoma (BLCA)

Brain Lower Grade Glioma (LGG)

Breast invasive carcinoma (BRCA)

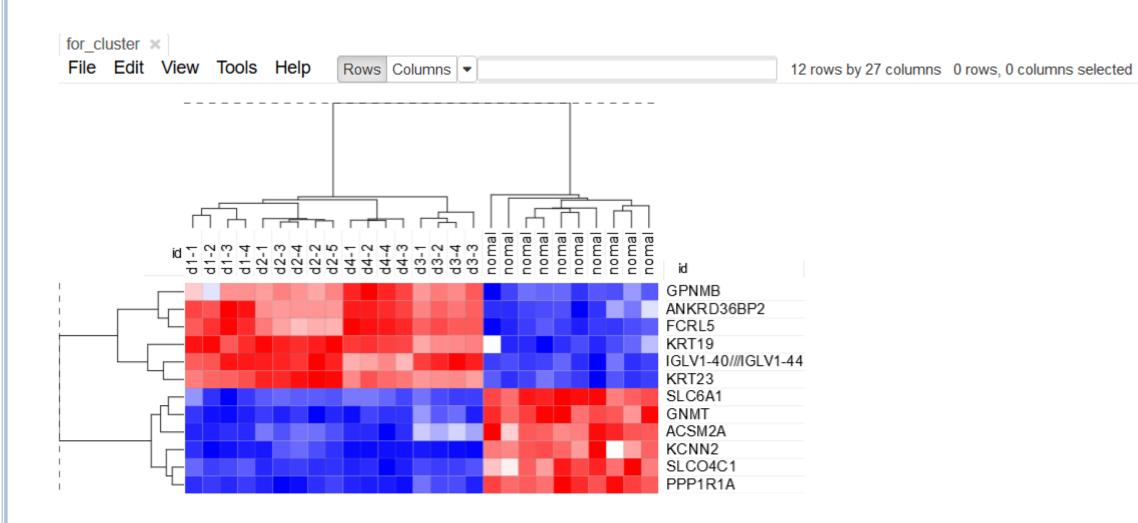
Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC)

Cholangiocarcinoma (CHOL)

Colon adenocarcinoma (COAD)

Colonrectal adenocarcinoma (COADREAD)

Esophageal carcinoma (ESCA)



#### FUNCTION ANNOTATION METHOD

## • Step1:

cluster genes on the basis of similarity measures Select up-regulated or down-regulated genes

#### • Step2:

Select a reference function frame

#### • Step3:

Determine the statistical significance of the mapping between cluster (differential genes group) and a functional category.

# 功能富集

| 工具名称       | 工具分类        | 网址   |
|------------|-------------|--|
| DAVID      | 在线工具        | https://david.ncifcrf.gov/   |
| PANTHER    | 在线工具        | http://www.pantherdb.org/  |
| GenMAPP    | 在线工具        | http://www.genmapp.org/  |
| GoMiner    | 在线工具        | https://discover.nci.nih.gov/gominer/index.jsp                         |
| GSEA       | Java软件、R语言包 | http://software.broadinstitute.org/gsea                                |
| Goplot     | R语言包        | https://CRAN.R-project.org/package=Goplot                              |
| topGO      | R语言包        | http://www.bioconductor.org/packages/release/bioc/html/topGO.html      |
| Sigpathway | R语言包        | http://www.bioconductor.org/packages/release/bioc/html/sigPathway.html |
| ClueGO     | Cytoscape插件 | http://apps.cytoscape.org/apps/cluego                                  |
| CluePedia  | Cytoscape插件 | http://apps.cytoscape.org/apps/cluepedia                               |
| BiNGO      | Cytoscape插件 | http://apps.cytoscape.org/apps/bingo                                   |

#### **Shortcut to DAVID Tools**

#### Functional Annotation

Gene-annotation enrichment analysis, functional annotation clustering, BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and more

#### Gene Functional Classification

Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. More

#### Gene ID Conversion

Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. More

#### Gene Name Batch Viewer

Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched detailed information. More

#### **Hot Links**



Submit papers for a Special Issue: "DNA or RNA-Mediated Innate Immune Response" of the International Journal of Molecular Sciences

#### DAVID Forum

Forum for DAVID users to ask questions, suggest a new functions for DAVID and help other users by answering their questions.

#### ► FAQ

Frequently Asked Questions

#### LHRI Publications

Publications of the Laboratory of Human Retrovirology and Immunoinformatics, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research

DAVID Publications

Recommending: A paper published in Nature Protocols describes step-by-step procedure to use DAVID!

#### Welcome to DAVID 6.8

2003 - 2020

The Database for Annotation, Visualization and Integrated Discovery (DAVID ) v6.8 comprises a full Knowledgebase update to the sixth version of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

- Identify enriched biological themes, particularly GO terms
- ☑ Discover enriched functional-related gene groups
  - Cluster redundant annotation terms
- ▼ Visualize genes on BioCarta & KEGG pathway maps
- ☑ Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes not in the list
- ✓ List interacting proteins
- Explore gene names in batch
- Link gene-disease associations
- Highlight protein functional domains and motifs

→ Darling of the melot of literature.

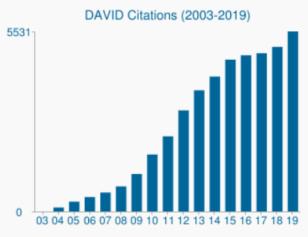
#### What's Important in DAVID?

- Cite DAVID
- IDs of Affy Exon and Gene arrays supported

Search

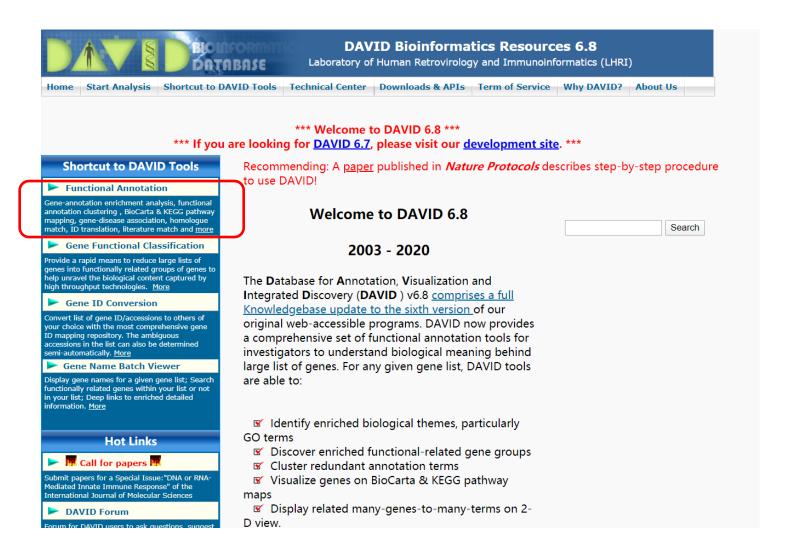
- Novel Classification Algorithms
- · Pre-built Affymetrix and Illumina <u>backgrounds</u>
- <u>User's customized gene background</u>
- Enhanced calculating speed

#### Statistics of DAVID



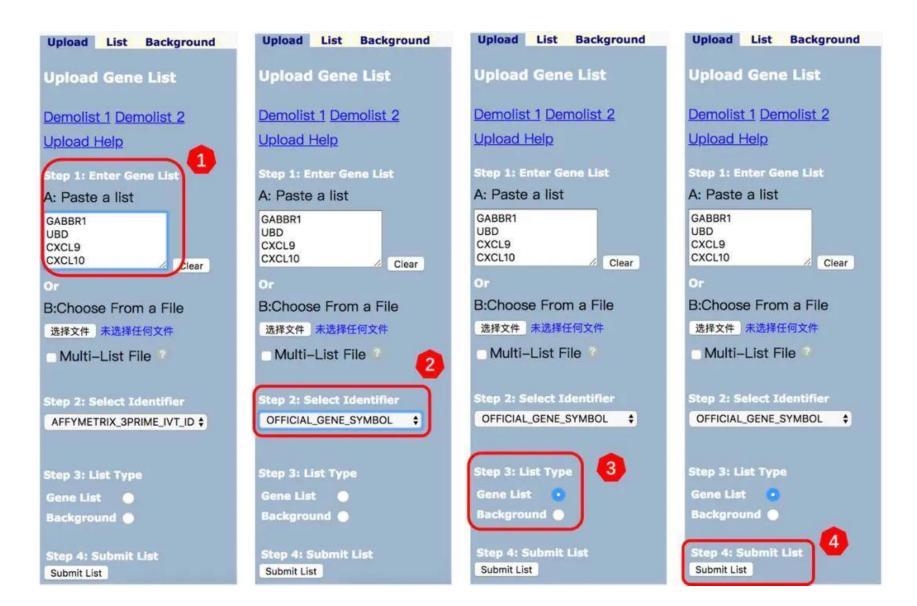
- > 38,000 Citations
- Average Daily Usage: ~2,700 gene lists/sublists from ~900 unique researchers.

# **FUNCTIONAL ANNOTATION**



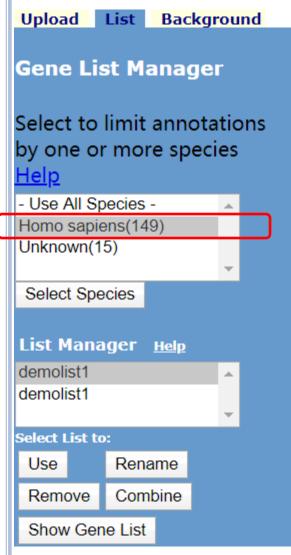
Upload List Background **Upload Gene List** AFFYMETRIX 3PRIME IVT ID AFFYMETRIX EXON ID AGILENT CHIP ID AGILENT ID AGILENT OLIGO ID APHIDBASE ID BEEBASE ID BEETLEBASE ID BGD ID CGNC ID CRYPTODB ID DICTYBASE ID ENSEMBL GENE ID ENSEMBL TRANSCRIPT ID ENTREZ GENE ID FLYBASE GENE ID GENBANK ACCESSION GENOMIC GI ACCESSION GENPEPT ACCESSION LOCUS TAG AFFYMETRIX 3PRIME\_IVT ID ▼

#### FUNCTIONAL ANNOTATION



#### \*\*\* Welcome to DAVID 6.8 \*\*\*

\*\*\* If you are looking for <u>DAVID 6.7</u>, please visit our <u>development site</u>. \*\*\*



#### **Annotation Summary Results**

Help and Tool Manual

Current Gene List: demolist1

**Current Background: Homo sapiens** 

145 DAVID IDs
Check Defaults 

✓

Clear All

- Disease (1 selected)
- Functional\_Categories (3 selected)
- Gene\_Ontology (3 selected)
- General\_Annotations (0 selected)
- Literature (0 selected)
- Main\_Accessions (0 selected)
- Pathways (3 selected)
- Protein\_Domains (3 selected)
- Protein\_Interactions (0 selected)
- Tissue\_Expression (0 selected)

\*\*\*Red annotation categories denote DAVID defined defaults\*\*\*

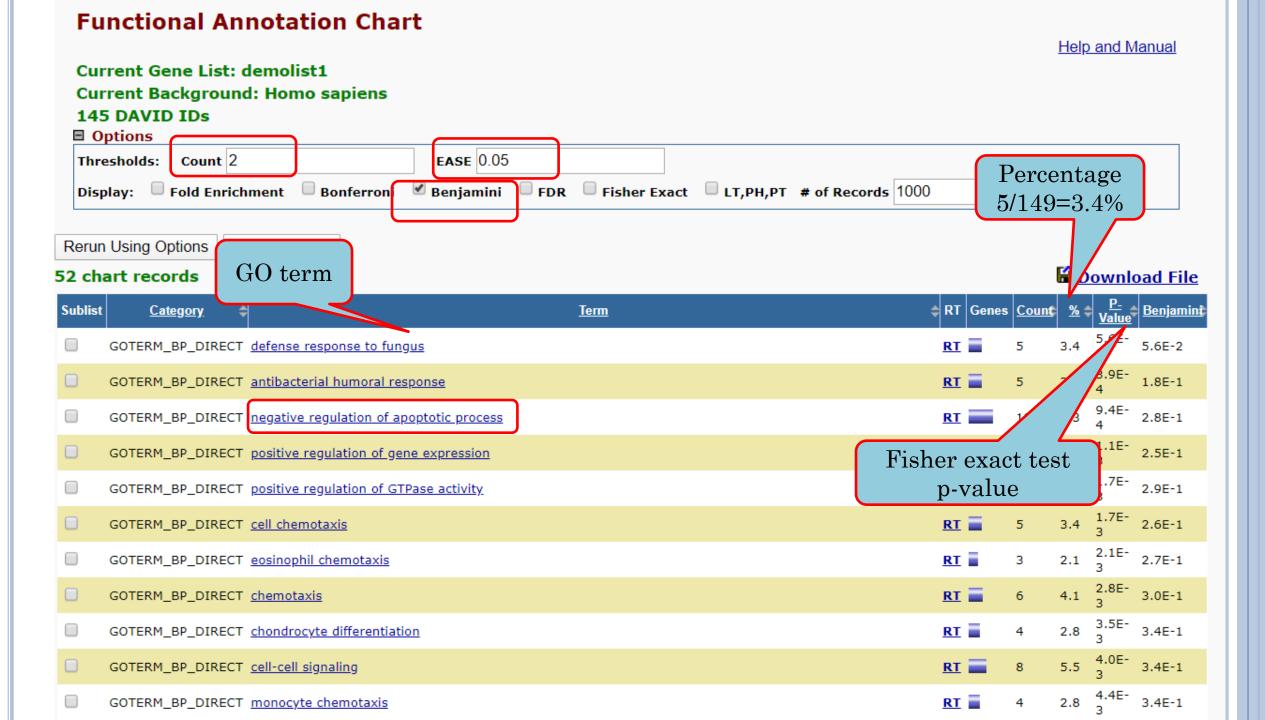
#### Combined View for Selected Annotation

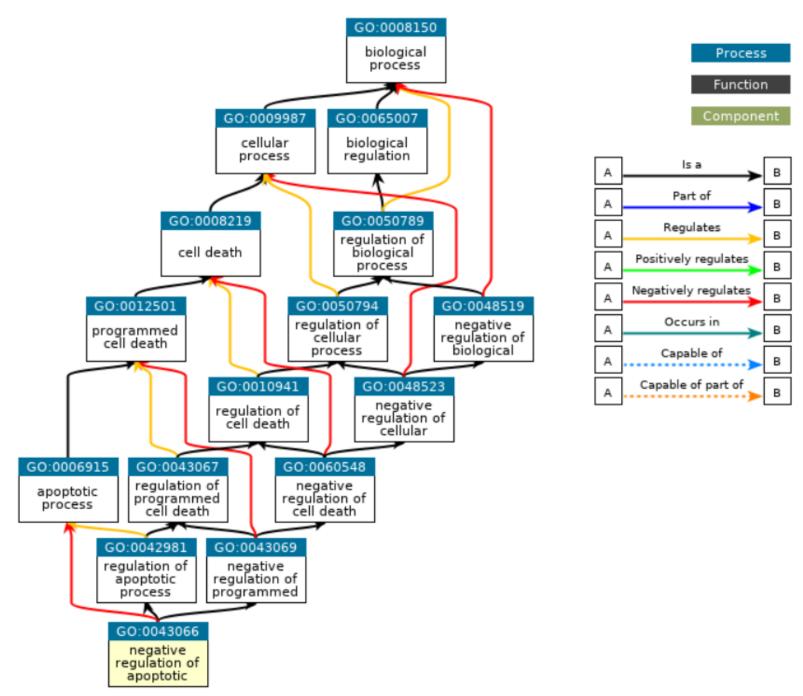
Functional Annotation Clustering

Functional Annotation Chart

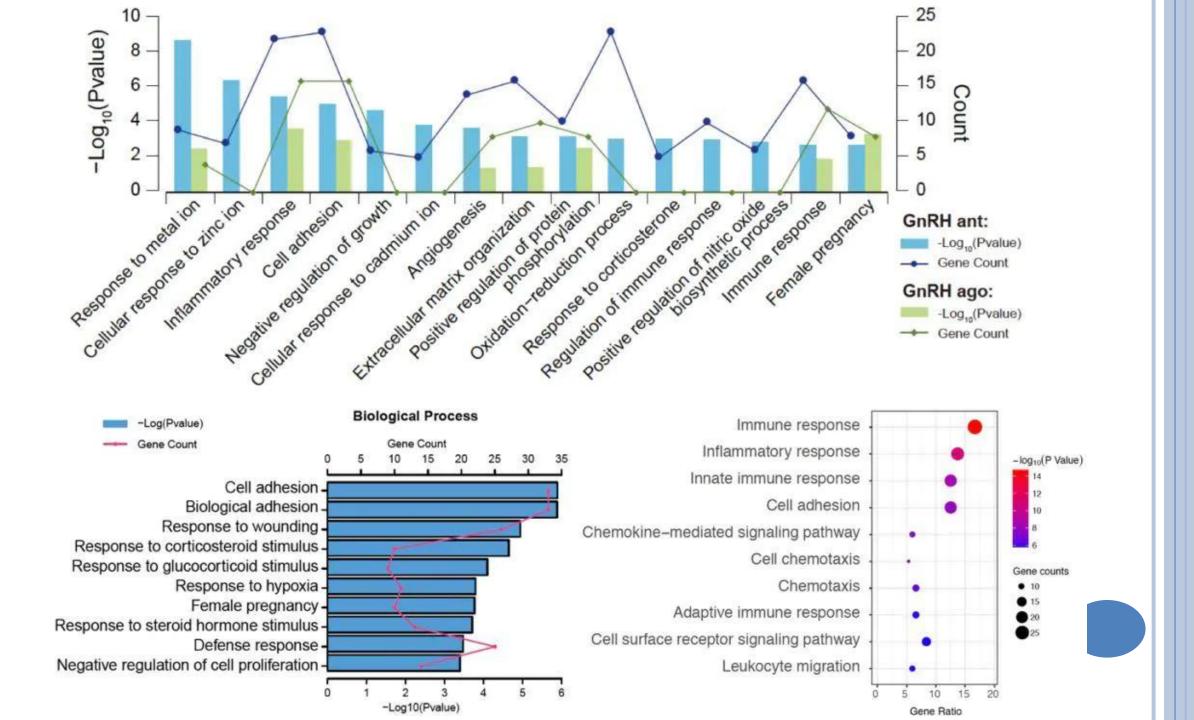
GO

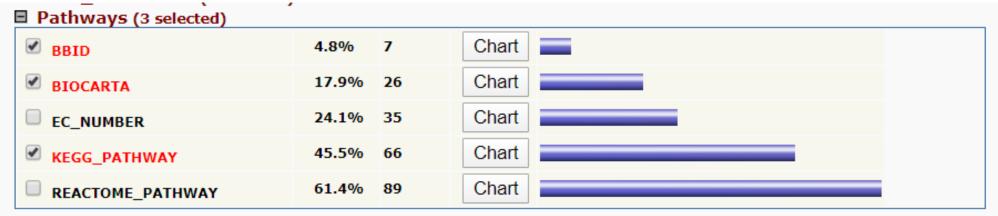






QuickGO - https://www.ebi.ac.uk/QuickGO





- Protein\_Domains (3 selected)
- Protein\_Interactions (0 selected)
- Tissue\_Expression (0 selected)

#### **Functional Annotation Chart**

Help and Manual

**Current Gene List: List\_4** 

**Current Background: Homo sapiens** 

1140 DAVID IDs

**Options** 

Rerun Using Options

Create Sublist

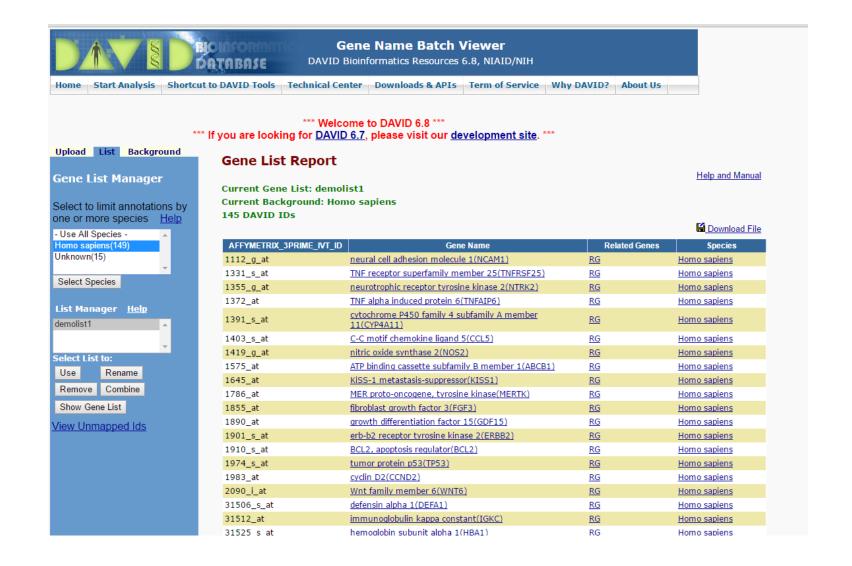
#### 24 chart records

#### **Download File**

| Sublist | <u>Category</u> | <u>Term</u>                         | <b>‡</b> RT | Genes    | Count \$ | <u>%</u> | P-Value \$ | <u>Benjamini</u> |
|---------|-----------------|-------------------------------------|-------------|----------|----------|----------|------------|------------------|
|         | KEGG_PATHWAY    | ECM-receptor interaction            | RT          |          | 37       | 3.2      | 1.8E-21    | 2.8E-19          |
|         | KEGG_PATHWAY    | Focal adhesion                      | <u>RT</u>   | =        | 49       | 4.3      | 6.2E-16    | 5.1E-14          |
|         | KEGG_PATHWAY    | TGF-beta signaling pathway          | RT          |          | 18       | 1.6      | 3.9E-5     | 2.0E-3           |
|         | KEGG_PATHWAY    | Pathways in cancer                  | <u>RT</u>   | =        | 40       | 3.5      | 1.7E-4     | 6.5E-3           |
|         | KEGG_PATHWAY    | Complement and coagulation cascades | RT          |          | 14       | 1.2      | 4.6E-4     | 1.4E-2           |
|         | KEGG_PATHWAY    | Basal cell carcinoma                | RT          | 1        | 12       | 1.1      | 7.5E-4     | 1.9E-2           |
|         | KEGG_PATHWAY    | Regulation of actin cytoskeleton    | RT          |          | 28       | 2.5      | 7.7E-4     | 1.7E-2           |
|         | KEGG_PATHWAY    | Cell adhesion molecules (CAMs)      | <u>RT</u>   | <b>=</b> | 19       | 1.7      | 2.3E-3     | 4.3E-2           |
|         | KEGG_PATHWAY    | Chondroitin sulfate biosynthesis    | RT          |          | 7        | 0.6      | 2.4E-3     | 3.9E-2           |
|         | KEGG_PATHWAY    | Hedgehog signaling pathway          | RT          | i        | 11       | 1.0      | 3.1E-3     | 4.7E-2           |
|         | KEGG_PATHWAY    | Hematopoietic cell lineage          | RT          |          | 14       | 1.2      | 3.8E-3     | 5.1E-2           |
|         | KEGG_PATHWAY    | Gap junction                        | <u>RT</u>   | <b>=</b> | 13       | 1.1      | 1.3E-2     | 1.5E-1           |
|         | KEGG_PATHWAY    | Vascular smooth muscle contraction  | <u>RT</u>   |          | 15       | 1.3      | 1.5E-2     | 1.6E-1           |
|         | VECC DATINAN    | Dileted configuration               | DT          | =        | 12       |          | 1 75 0     | 1751             |

#### FOCAL ADHESION → mDia1 RhoGAP RhoA **▶** ROCK RhoGEF Regulation of actin cytoskeleton MLCP PIP5K +p Cell motility Zyxin MLCK Stress fiber / FA $\widehat{\tau_{+p}}$ Actinin VASP Actin formation PIP2 polymerization Filamin TGA V inculin 988862444668 ECM Actin TTGB Filopodia Lamellipodia formation Parvin Talin PDK1 Paxillin ECM-receptor SrcFA turnover interaction GSK-3B TLK Akt/PKB β-Catenin Calpain +p PKC +p Wnt signaling pathway Cdc42 PTEN Caveolin Cell proliferation +p Г-р -p PI3K-Akt signaling pathway FAK PI3K ► Vav PAK Rac → CycD PIP3 DNA Phosphatidyl inositol signaling system Fyn Cell cycle **▼** DOCK1 GRF2 p130Cas Crk Rap1 JNK c-Jun C3G Cytokine-cytokine receptor MAPK signaling pathway B-Raf interaction Cell survival ► MEK1 ► Shc F GF 🕶 RTK Grb2 Sos ► Ha-Ras Raf-1 ERK1/2 Elk1 **►** cIAPs DNA Bcl-2

## GENE NAME BATCH VIEWER



# ID CONVERSION



#### **DAVID Bioinformatics Resources 6.8**

Laboratory of Human Retrovirology and Immunoinformatics (LHRI)

\*\*\* Welcome to DAVID 6.8 \*\*\*

\*\*\* If you are looking for <u>DAVID 6.7</u>, please visit our <u>development site</u>. \*\*\*

#### **Gene Accession Conversion Tool**

<u>Help</u>

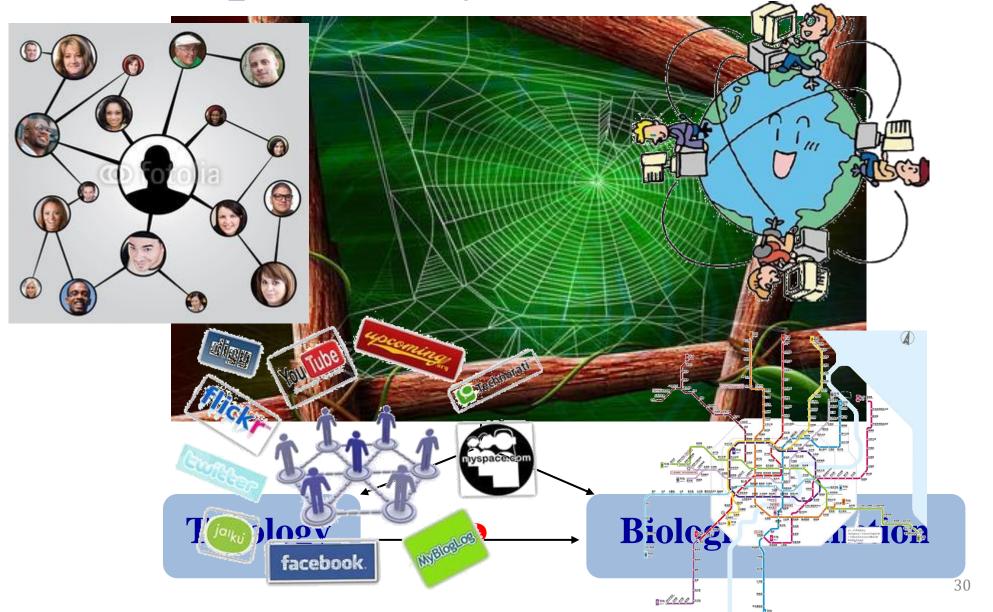
**Gene Accession Conversion Statistics** 

■ Download File

| Conversion                           | Summary           |            | Submit Converted List to DAVID as a Gene List Submit Converted List to DAVID as a Backgrou |                 |  |  |  |  |  |  |
|--------------------------------------|-------------------|------------|--|-----------------|--|--|--|--|--|--|
| ID Count                             | In DAVID DB       | Conversion |  |                 |  |  |  |  |  |  |
| 149                                  | Yes               | Successful | From   | То              | Species  | David Gene Name  |  |  |  |  |
| 0                                    | Yes               | None       | 34702_f_at   | 387590          | Homo<br>sapiens  | transmembrane phosphatase with tensin homology pseudogene<br>1(TPTEP1) |  |  |  |  |
| 0                                    | No                | None       | 27166 at   | 23498           | Homo   | 3 hydrovanthranilate 3 4 diagoganaco(HAAO)                             |  |  |  |  |
| 0                                    | Ambiguous Pending |            | 37166_at   | 23490           | sapiens  | 3-hydroxyanthranilate 3,4-dioxygenase(HAAO)                            |  |  |  |  |
|                                      | e User IDs: 149   |            | 40790_at   | 8553            | Homo<br>sapiens  | basic helix-loop-helix family member e40(BHLHE40)                      |  |  |  |  |
|                                      | f Ambiguous Gen   |            | 35213 at   | 11193           | Homo   | WW domain his ding protein 4/WDD4)                                     |  |  |  |  |
| ID Count Possible Source Convert All |                   |            | 33213_at   | 11193           | sapiens  | WW domain binding protein 4(WBP4)                                      |  |  |  |  |
|                                      | Sources For Aml   |            | 1901_s_at  | 2064            | Homo<br>sapiens  | erb-b2 receptor tyrosine kinase 2(ERBB2)                               |  |  |  |  |
| Ambiguous ID Possiblity Convert      |                   |            | 37898_r_at   | 7033            | Homo<br>sapiens  | trefoil factor 3(TFF3)   |  |  |  |  |
|                                      |                   |            | 33963_at   | 566             | Homo<br>sapiens  | azurocidin 1(AZU1)   |  |  |  |  |
|                                      |                   |            | 679_at   | 1511            | Homo<br>sapiens  | cathepsin G(CTSG)  |  |  |  |  |
|                                      |                   |            | 32717_at   | 9148            | Homo<br>sapiens  | neuralized E3 ubiquitin protein ligase 1(NEURL1)                       |  |  |  |  |
|                                      |                   |            | 40735_at   | 3034            | Homo<br>sapiens  | histidine ammonia-lyase(HAL)   |  |  |  |  |
|                                      |                   | 606_at     | 8153   | Homo<br>sapiens | Rho family GTPase 2(RND2)                              |  |  |  |  |  |
|                                      |                   |            | 35069_at   | 65121           | Homo<br>sapiens  | PRAME family member 1(PRAMEF1)   |  |  |  |  |
|                                      |                   | 35373_at   | 5295   | Homo<br>sapiens | phosphoinositide-3-kinase regulatory subunit 1(PIK3R1) |  |  |  |  |  |
|                                      |                   |            | 38128 at   | 51471           | Homo   | N-acetyltransferase 8B (putative, gene/pseudogene)(NAT8B)              |  |  |  |  |

# NETWORK VISULAZATION

# **Graph Theory**



# SMALL WORLD

•Six degrees of separation

# Steps ≤ 6





# 常用网络参数

| Degree                       | How many links a node has to other node. Find hubs.                    |
|------------------------------|--|
| Betweenness centrality(BC)   | Find important node as bottlenecks in a network.                       |
| The edge BC                  | Reflect edges role instead of nodes.                                   |
| The closeness centrality(CC) | Indicate which node is close to the topological center of the network. |

# CYTOSCAPE?

• http://www.cytoscape.org/

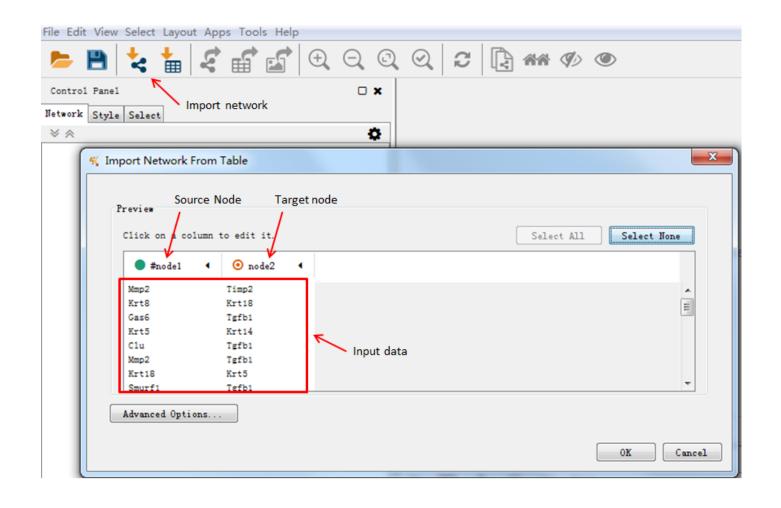


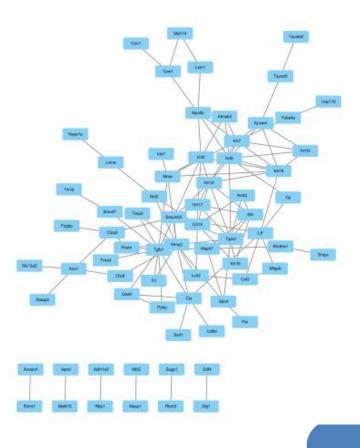
Cytoscape is an open source software platform for visualizing complex networks and integrating these with any type of attribute data. A lot of *Apps* are available for various kinds of problem domains,

Welcome Letter

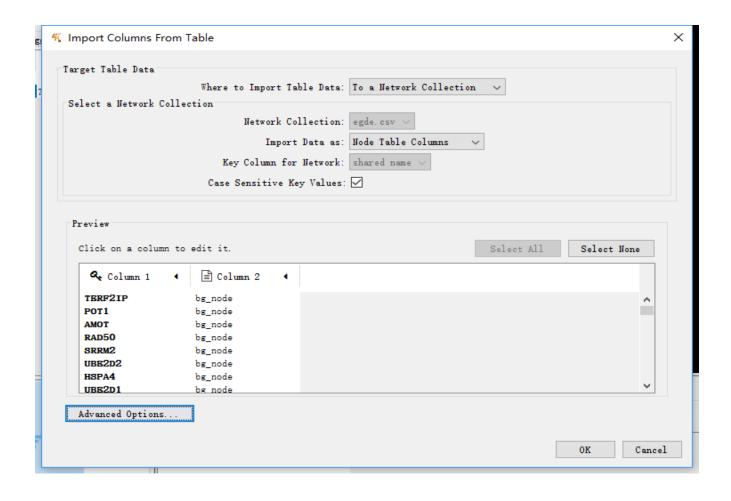
Release Notes

# FILE -> IMPORT -> NETWORK

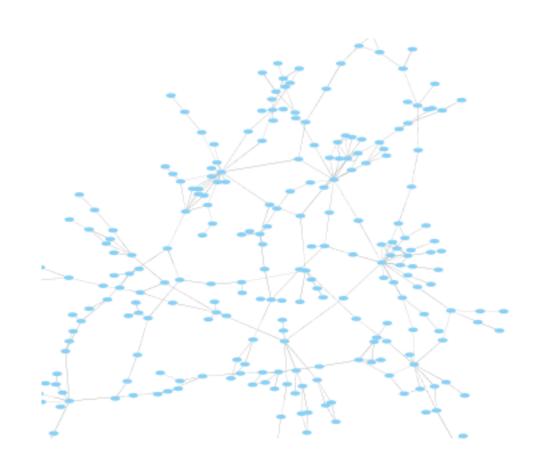


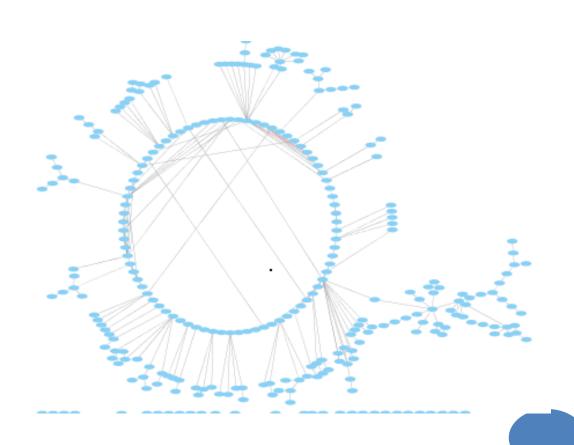


# FILE>IMPORT>TABLE>FILE

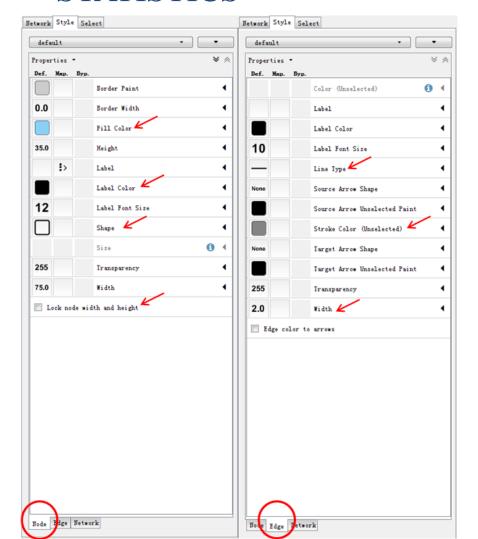


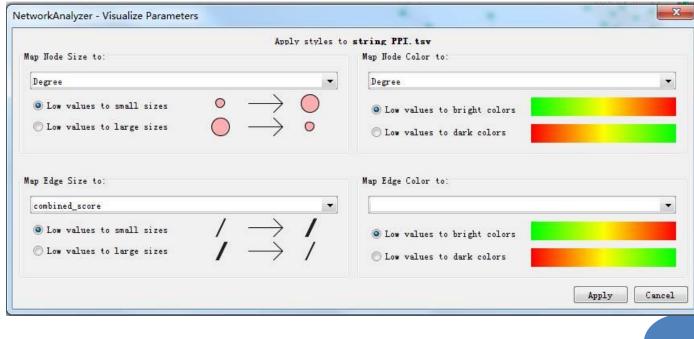
# Layout > yFiles layouts > organic > circular



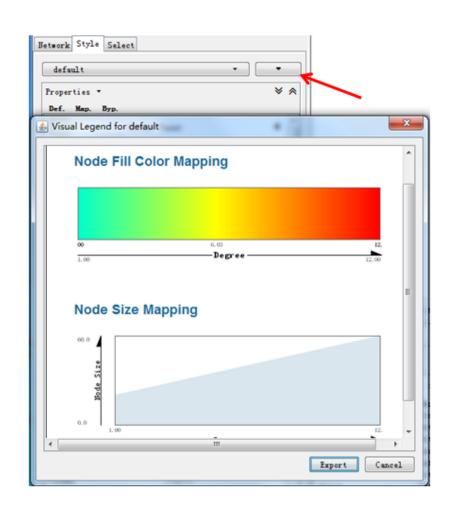


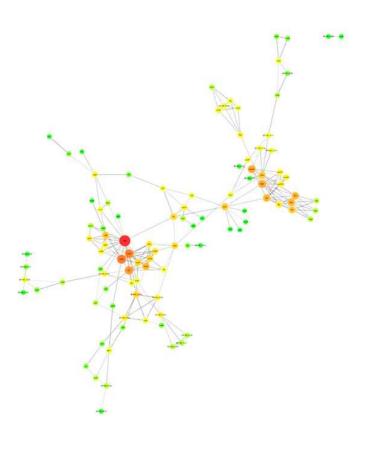
# Tools -> NetworkAnlyzer -> Network Anlysis Tools -> NetworkAnlyzer -> Generate style from Statistics



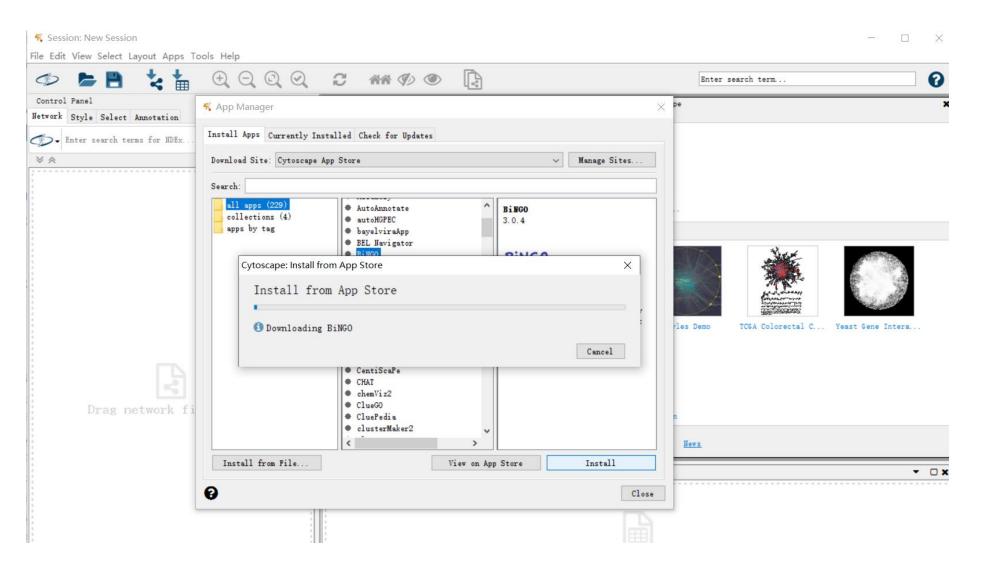


# -> CREAT LEGEND





# PLUG-INS



#### **SUMMARY**

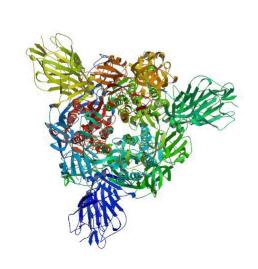
Database



# 序列分析的内容

- 蛋白质家族分析
- 相似序列搜索
- 系统发育分析
- 蛋白质理化性质
- 序列模式、结构域、基序搜索
- 翻译后修饰预测
- 二级结构、三级结构预测
- 定位、跨膜预测
- 功能、生化途径

- Who
- What
- Where
- How



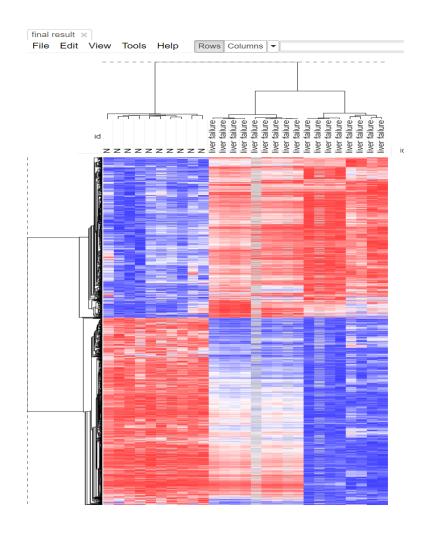
# 序列比对, 进化

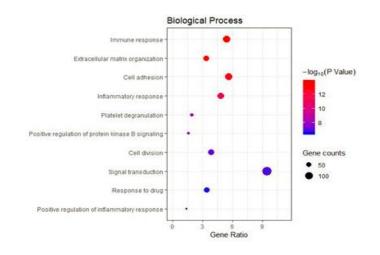
• Blast

• MSA

• tree

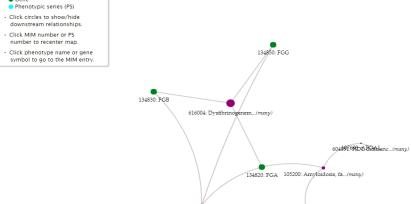
# OMICS





Key:

• Phenotype



202400: Afibrinogenemia...(many)

109700: B2M 241600: Immunodeficienc...(B2M)