

# 使用GO::TermFinder进行GO 富集分析


## 一、安装GO::TermFinder

- 操作系统：Ubuntu
- GO::TermFinder：是Perl下的一个模块，可以对一组gene和GO Terms做基因富集分析
- 1. 首先在Ubuntu系统下安装好**Perl解释器**以及Perl的**包管理器CPAN**。安装完毕后确认CPAN能够使用：  
sudo perl -MCPAN -eshell 打开成功后即为成功安装CPAN，键入q退出CPAN
- 2. 在安装GO::TermFinder之前需要**安装一些Perl下的依赖库**，否则安装时会一直报错，具体如下(待补充):
  1. Storable 使用CPAN下install Storable即可安装
  2. CGI 使用CPAN下install CGI 即可安装
  3. GD 首先**需要在Ubuntu系统下**安装zlib、libpng、libgd模块(可能可以仅安装个别，未验证但全部安装肯定可以)，使用sudo apt-get update 以及 sudo apt-get install libgd2-xpm-dev 等命令进行安装，**然后再进入CPAN**下install GD进行安装才能成功，否则直接CPAN下安装会报错
  4. GraphViz 首先**需要在Ubuntu系统下**sudo apt-get install graphviz，**然后再进入CPAN**下install GraphViz进行安装才能成功，否则直接CPAN下安装会报错
- 3. 安装完上述模块后入CPAN下install GO::TermFinder进行安装，如果还出现报错，则需要按照出错提示安装对应的依赖模块才能够成功

## 二、下载对应的ontology files

- GO::TermFinder主要需要两个与Gene Ontology有关的文件
    - 1.**obo文件**，是用来存储所有GO term关系的文件，GOC上分有basic版本和不稳定版本，不稳定版本会经常拓展，但是包括了GO的其他关系
    - 2.**annotations文件**，是用来存储gene 的GO 注释的文件，不同的物种有不同的annotation文件，在GOC上找对应的物种文件进行下载
- 这两个文件均可以从Gene Ontology Consortium的官网上(<http://www.geneontology.org/>)进行下载

下载界面：


Gene Ontology Consortium

[Home](#)
[Documentation](#)
[Downloads](#)
[Tools](#)
[About](#)
[Contact us](#)

# Gene Ontology Consortium

## Enrichment analysis

Your gene IDs here...

biological process

Homo sapiens

Submit

Help

Powered by [PANTHER](#)

## Search GO data

Search for terms and gene products...

Search

## Ontology

[Filter classes](#)

[Download ontology](#)

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

**molecular function**  
molecular activities of gene products

**cellular component**  
where gene products are active

**biological process**  
pathways and larger processes made up of the activities of multiple gene products.

[more](#)

## Annotations

[Download annotations](#) (standard files)

[Filter and download](#) (customizable files <100k lines)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. [more](#)

The mission of the GO Consortium is to develop an up-to-date, comprehensive, **computational model of biological systems**, from the molecular level to larger pathways, cellular and organism-level systems. [more](#)

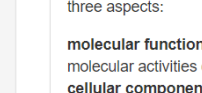
## Search documentation

Search

## What is the Gene Ontology?

- [An introduction to the Gene Ontology](#)
- [What are annotations?](#)
- [Enrichment analysis](#)
- [Downloads](#)

## Statistics



## Other GOC tools

Explore other GOC [tools](#) in the AmiGO software suite.

### 三、获取分析代码并进行分析

- 在CPAN上下载的GO::TermFinder的模块包里面，examples文件夹中有一个其他团队写好的分析代码 **batchGOView.pl** 及其配置文件**GoView.conf**。该perl代码是利用GO::TermFinder模块对基因文件进行批量处理的代码，配置文件内能够对生成的图以及p-value阈值等进行设定。
- 分析步骤
  1. 打开GoView.conf，更改：
    1. annotationFile 把对应值改成下载下来的GO annotations文件的位置
    2. ontologyFile 把对应值改成下载下来的obo文件的位置
    3. aspect 表示GO分析的类别。GO数据库被分成了三个大类，分别cellular\_component, biological\_process 和molecular\_function，所以这里的参数有三种可能：C, P 或者 F。每次只能设置一个。
    4. pvalueCutOff 可以设置p-value的阈值
  2. 更改配置文件完成后，准备gene文件(因为是批量处理，所以可以是多个)，gene以txt文本的形式存储，每行存储一个gene，放在对应的位置。
  3. 此时再次确保 batchGOView.pl、GoView.conf、obo文件、annotations文件、gene.txt文件都已经存在
  4. 运行命令 perl batchGOView.pl GoView.conf genes.txt genes2.txt ..... 进行富集分析
  5. 成功运行后，每个gene.txt文件都会对应生成一个html文件，里面即为富集分析的结果

结果示例：

## Terms for genes.txt

Result Table

Terms from the Process Ontology with p-value as good or better than 0.01

Gene Ontology term	Cluster frequency	Genome frequency of use	Corrected P-value	FDR	False Positives	Genes annotated to the term
<a href="#">cellular lipid metabolism</a>	68 out of 238 genes, 28.6%	996 out of 24863 genes, 4.0%	6.72e-36	0.00%	0.00	<a href="#">Abca1</a> , <a href="#">Ang</a> , <a href="#">Akr1c6</a> , <a href="#">Acat1</a> , <a href="#">Acsm5</a> , <a href="#">Acs15</a> , <a href="#">Acot4</a> , <a href="#">Abhd2</a> , <a href="#">Acot3</a> , <a href="#">Acsm3</a> , <a href="#">Akr1d1</a> , <a href="#">Akr1c13</a> , <a href="#">Acot12</a> , <a href="#">Amacr</a> , <a href="#">Adipor2</a> , <a href="#">Acaa2</a> , <a href="#">Acaa1b</a> , <a href="#">Acsm1</a> , <a href="#">Akr1c12</a> , <a href="#">Abcd3</a> , <a href="#">B3galt1</a> , <a href="#">Abhd3</a> , <a href="#">Bdh2</a> , <a href="#">Aldh5a1</a> , <a href="#">Apoa5</a> , <a href="#">Bcmo1</a> , <a href="#">Apoa1</a> , <a href="#">Angptl3</a> , <a href="#">Acnat1</a> , <a href="#">Apoc3</a> , <a href="#">Agmo</a> , <a href="#">Apoa5</a> , <a href="#">Apoa1</a> , <a href="#">Baat</a> , <a href="#">Abcd2</a> , <a href="#">Akr1c18</a> , <a href="#">Abcb4</a> , <a href="#">Akr1c20</a> , <a href="#">Abhd15</a> , <a href="#">Aldh8a1</a> , <a href="#">Apoabec1</a> , <a href="#">9130409123Rik</a> , <a href="#">Anxa1</a> , <a href="#">Apoab</a> , <a href="#">App</a> , <a href="#">Agpat3</a> , <a href="#">Apoc1</a> , <a href="#">Adh1</a> , <a href="#">Apoa4</a> , <a href="#">Adh7</a> , <a href="#">Akr1c14</a> , <a href="#">Acox1</a> , <a href="#">Adh4</a> , <a href="#">Acoc2</a> , <a href="#">Akr1c19</a> , <a href="#">Acs1</a> , <a href="#">Acadvl</a> , <a href="#">Apoc2</a> , <a href="#">Adh5</a> , <a href="#">Aldh1a1</a> , <a href="#">Apoa</a> , <a href="#">Aig1</a> , <a href="#">Acnat2</a> , <a href="#">Agt</a> , <a href="#">Apoa2</a> , <a href="#">9530008L14Rik</a> , <a href="#">Acat3</a> , <a href="#">Bco2</a>
<a href="#">lipid metabolism</a>	68 out of 238 genes, 28.6%	1130 out of 24863 genes, 4.5%	1.85e-32	0.00%	0.00	<a href="#">Aldh8a1</a> , <a href="#">Abhd15</a> , <a href="#">Akr1c20</a> , <a href="#">Abcb4</a> , <a href="#">Akr1c18</a> , <a href="#">Abcd2</a> , <a href="#">Adh4</a> , <a href="#">Acoc1</a> , <a href="#">Akr1c19</a> , <a href="#">Acoc2</a> , <a href="#">Akr1c14</a> , <a href="#">Adh7</a> , <a href="#">Apoa4</a> , <a href="#">App</a> , <a href="#">Apoc1</a> , <a href="#">Agpat3</a> , <a href="#">Adh1</a> , <a href="#">Anxa1</a> , <a href="#">9130409123Rik</a> , <a href="#">Apoab</a> , <a href="#">Apoabec1</a> , <a href="#">Apoa</a> , <a href="#">Aig1</a> , <a href="#">Adh5</a> , <a href="#">Aldh1a1</a> , <a href="#">Apoc2</a> , <a href="#">Acadvl</a> , <a href="#">Acs1</a> , <a href="#">Bco2</a> , <a href="#">Acat3</a> , <a href="#">9530008L14Rik</a> , <a href="#">Apoa2</a> , <a href="#">Agt</a> , <a href="#">Acnat2</a> , <a href="#">Acs15</a> , <a href="#">Acat1</a> , <a href="#">Acsm5</a> , <a href="#">Akr1c6</a> , <a href="#">Ang</a> , <a href="#">Abca1</a> , <a href="#">Amacr</a> , <a href="#">Acot12</a> , <a href="#">Akr1c13</a> , <a href="#">Akr1d1</a> , <a href="#">Acot3</a> , <a href="#">Acsm3</a> , <a href="#">Acot4</a> , <a href="#">Abhd2</a> , <a href="#">Bcmo1</a> , <a href="#">Apoa</a> , <a href="#">Aldh5a1</a> , <a href="#">Bdh2</a> , <a href="#">B3galt1</a> , <a href="#">Abhd3</a> , <a href="#">Acsm1</a> , <a href="#">Acaa1b</a> , <a href="#">Akr1c12</a> , <a href="#">Abcd3</a> , <a href="#">Acaa2</a> , <a href="#">Adipor2</a> , <a href="#">Baat</a> , <a href="#">Agmo</a> , <a href="#">Apoa5</a> , <a href="#">Apoc3</a> , <a href="#">Angptl3</a> , <a href="#">Acnat1</a> , <a href="#">Apoa1</a>
<a href="#">carboxylic acid metabolism</a>	64 out of 238 genes, 26.9%	1213 out of 24863 genes, 4.9%	5.77e-27	0.00%	0.00	<a href="#">App</a> , <a href="#">Acmsd</a> , <a href="#">Aif1</a> , <a href="#">Aass</a> , <a href="#">Anxa1</a> , <a href="#">Acoc1</a> , <a href="#">Acoc2</a> , <a href="#">Akr1c14</a> , <a href="#">Apoa4</a> , <a href="#">Adhfe1</a> , <a href="#">Akr1c18</a> , <a href="#">Abcd2</a> , <a href="#">Akr1c20</a> , <a href="#">Amdhd1</a> , <a href="#">Apoa2</a> , <a href="#">Ahcy</a> , <a href="#">Agt</a> , <a href="#">Acnat2</a> , <a href="#">Agi</a> , <a href="#">Agxt2</a> , <a href="#">Abat</a> , <a href="#">Acat3</a> , <a href="#">9530008L14Rik</a> , <a href="#">Acadvl</a> , <a href="#">Apoc2</a> , <a href="#">Acs1</a> , <a href="#">Agi1</a> , <a href="#">Acot3</a> , <a href="#">Acsm3</a> , <a href="#">Acot4</a> , <a href="#">Abhd2</a> , <a href="#">Aldh4a1</a> , <a href="#">Arg1</a> , <a href="#">Amacr</a> , <a href="#">Acot12</a> , <a href="#">Akr1d1</a> , <a href="#">Aldh6a1</a> , <a href="#">Adi1</a> , <a href="#">Bckdha</a> , <a href="#">Akr1c6</a> , <a href="#">Acs15</a> , <a href="#">Bbox1</a> , <a href="#">Acat1</a> , <a href="#">Acsm5</a> , <a href="#">Acnat1</a> , <a href="#">Angptl3</a> , <a href="#">Apoa1</a> , <a href="#">Ass1</a> , <a href="#">Aldh1l1</a> , <a href="#">Agt</a> , <a href="#">Baat</a> , <a href="#">Apoa5</a> , <a href="#">Adssl1</a> , <a href="#">Agphd1</a> , <a href="#">Abhd3</a> , <a href="#">Acsm1</a> , <a href="#">Acaa1b</a> , <a href="#">Abcd3</a> , <a href="#">1500003O03Rik</a> , <a href="#">Adipor2</a> , <a href="#">Acaa2</a> , <a href="#">Aldh5a1</a> , <a href="#">Aadat</a> , <a href="#">Bdh2</a>
<a href="#">organic acid metabolism</a>	65 out of 238 genes, 27.3%	1261 out of 24863 genes, 5.1%	7.11e-27	0.00%	0.00	<a href="#">Akr1c20</a> , <a href="#">Amdhd1</a> , <a href="#">Akr1c18</a> , <a href="#">Abcd2</a> , <a href="#">Acoc1</a> , <a href="#">Acoc2</a> , <a href="#">Akr1c14</a> , <a href="#">Apoa4</a> , <a href="#">Adhfe1</a> , <a href="#">App</a> , <a href="#">Acmsd</a> , <a href="#">Aif1</a> , <a href="#">Abcg2</a> , <a href="#">Aass</a> , <a href="#">Anxa1</a> , <a href="#">Agi1</a> , <a href="#">Apoc2</a> , <a href="#">Acadvl</a> , <a href="#">Acs1</a> , <a href="#">Agxt2</a> , <a href="#">Abat</a> , <a href="#">Acat3</a> , <a href="#">9530008L14Rik</a> , <a href="#">Apoa2</a> , <a href="#">Ahcy</a> , <a href="#">Agt</a> , <a href="#">Agi</a> , <a href="#">Acnat2</a> , <a href="#">Bbox1</a> , <a href="#">Acs15</a> , <a href="#">Acat1</a> , <a href="#">Acsm5</a> , <a href="#">Bckdha</a> , <a href="#">Akr1c6</a> , <a href="#">Amacr</a> , <a href="#">Acot12</a> , <a href="#">Akr1d1</a> , <a href="#">Aldh6a1</a> , <a href="#">Adi1</a> , <a href="#">Acot3</a> , <a href="#">Acsm3</a> , <a href="#">Acot4</a> , <a href="#">Abhd2</a> , <a href="#">Aldh4a1</a> , <a href="#">Arg1</a> , <a href="#">Aldh5a1</a> , <a href="#">Bdh2</a> , <a href="#">Aadat</a> , <a href="#">Agphd1</a> , <a href="#">Abhd3</a> , <a href="#">Acaa1b</a> , <a href="#">Acsm1</a> , <a href="#">Abcd3</a> , <a href="#">1500003O03Rik</a> , <a href="#">Acaa2</a> , <a href="#">Adipor2</a> , <a href="#">Agt</a> , <a href="#">Baat</a> , <a href="#">Apoa5</a> , <a href="#">Adssl1</a> , <a href="#">Angptl3</a> , <a href="#">Acnat1</a> , <a href="#">Apoa1</a> , <a href="#">Ass1</a> , <a href="#">Aldh1l1</a>
<a href="#">fatty acid metabolism</a>	38 out of 238 genes, 16.0%	345 out of 24863 genes, 1.4%	4.84e-26	0.00%	0.00	<a href="#">Bdh2</a> , <a href="#">Aldh5a1</a> , <a href="#">Agi1</a> , <a href="#">Adipor2</a> , <a href="#">Acaa2</a> , <a href="#">Acs1</a> , <a href="#">Abcd3</a> , <a href="#">Acaa1b</a> , <a href="#">Acsm1</a> , <a href="#">Abhd3</a> , <a href="#">Apoc2</a> , <a href="#">Acadvl</a> , <a href="#">9530008L14Rik</a> , <a href="#">Apoa5</a> , <a href="#">Acat3</a> , <a href="#">Baat</a> , <a href="#">Acnat2</a> , <a href="#">Agt</a> , <a href="#">Apoa2</a> , <a href="#">Angptl3</a> , <a href="#">Acnat1</a> , <a href="#">Acsm5</a> , <a href="#">Akr1c20</a> , <a href="#">Acat1</a> , <a href="#">Acs15</a> , <a href="#">Abcd2</a> , <a href="#">Akr1c6</a> , <a href="#">Akr1c18</a> , <a href="#">Apoa4</a> , <a href="#">Acot12</a> , <a href="#">Akr1c14</a> , <a href="#">Acoc2</a> , <a href="#">Acoc1</a> , <a href="#">Anxa1</a> , <a href="#">Abhd2</a> , <a href="#">Acot4</a> , <a href="#">Acsm3</a> , <a href="#">Acot3</a>