

GO(Gene Ontology)

网址：<http://www.geneontology.org/>

报告人：邓志罗

什么是ontology

- 本体论(ontology)是哲学概念，它是研究存在的本质的哲学问题。通俗地讲，“存在”即是事物固有的属性集合，人们可能用不同的词汇、不同的语句、甚至不同的语言来描述同一事物，但该事物不会因为人们的不同描述而改变。
- 信息工程中，经常研究词汇之间的相关性，就是要找出人们对同一事物的不同描述，或者从含有相关词汇的语句中，找出他们共同描述的事物。

GO(Gene Ontology)是：

- 基因本体联合会(Gene Ontology Consortium)所建立的数据库，旨在建立一套适用于各种物种的，对基因和蛋白质功能进行限定和描述的，并能随着研究不断深入而更新的语义(terms)词汇标准。
- GO是一类综合数据库，他将全世界所有与基因有关的研究结果，分类汇总。主要分为三个方面的汇总：生物过程、分子功能及细胞组分。它将基因结合这三个方面的内容进行注释。

GO语义的分类

分子功能(*Molecular Function*)

描述在个体分子生物学上的活性，如催化活性或结合活性。

生物学过程(*Biological Process*)

由分子功能有序地组成的，具有多个步骤的一个过程。

细胞组件(*Cellular Component*)

指基因产物位于何种细胞器或基因产物组中(如糙面内质网，核糖体，蛋白酶体等)，即基因产物在什么地方起作用。

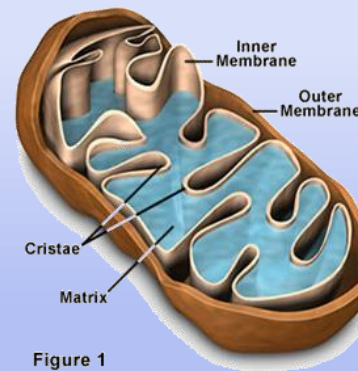
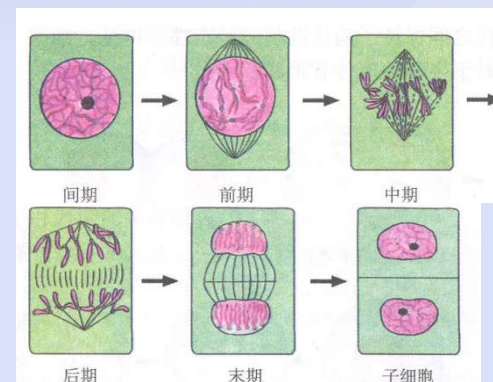
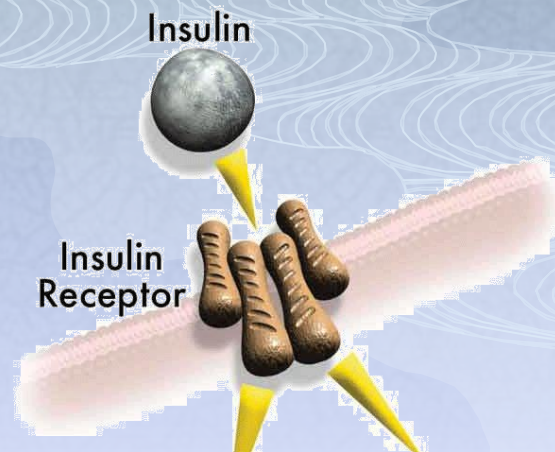
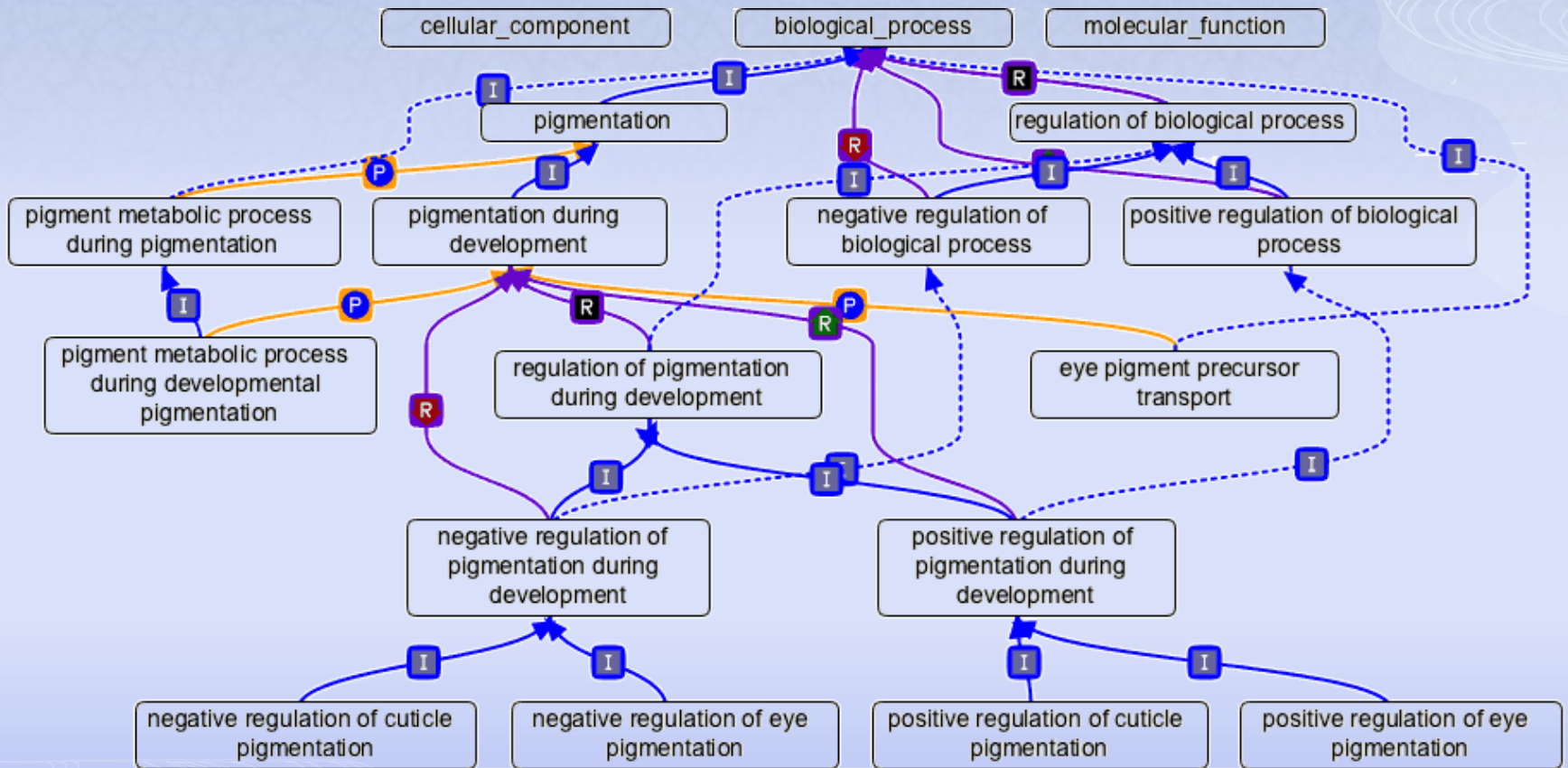


Figure 1



GO项目诞生原因

- 解决生物学定义混乱的现象，使各种数据库中基因产物功能描述相一致
- 使得在不同生物数据库中的查询具有极高的一致性
- 允许在各种水平查询基因产物的特性
- 致力于各种基因组数据库的标准化。

WormBase

Gene Ontology widely adopted

FlyBase

AgBase



A Comparative mapping resource
GRAMENE



TIGR
THE INSTITUTE FOR GENOMIC RESEARCH



The Wellcome Trust
Sanger Institute

UniProt
the universal protein resource

Gene Ontology Consortium 组织成员

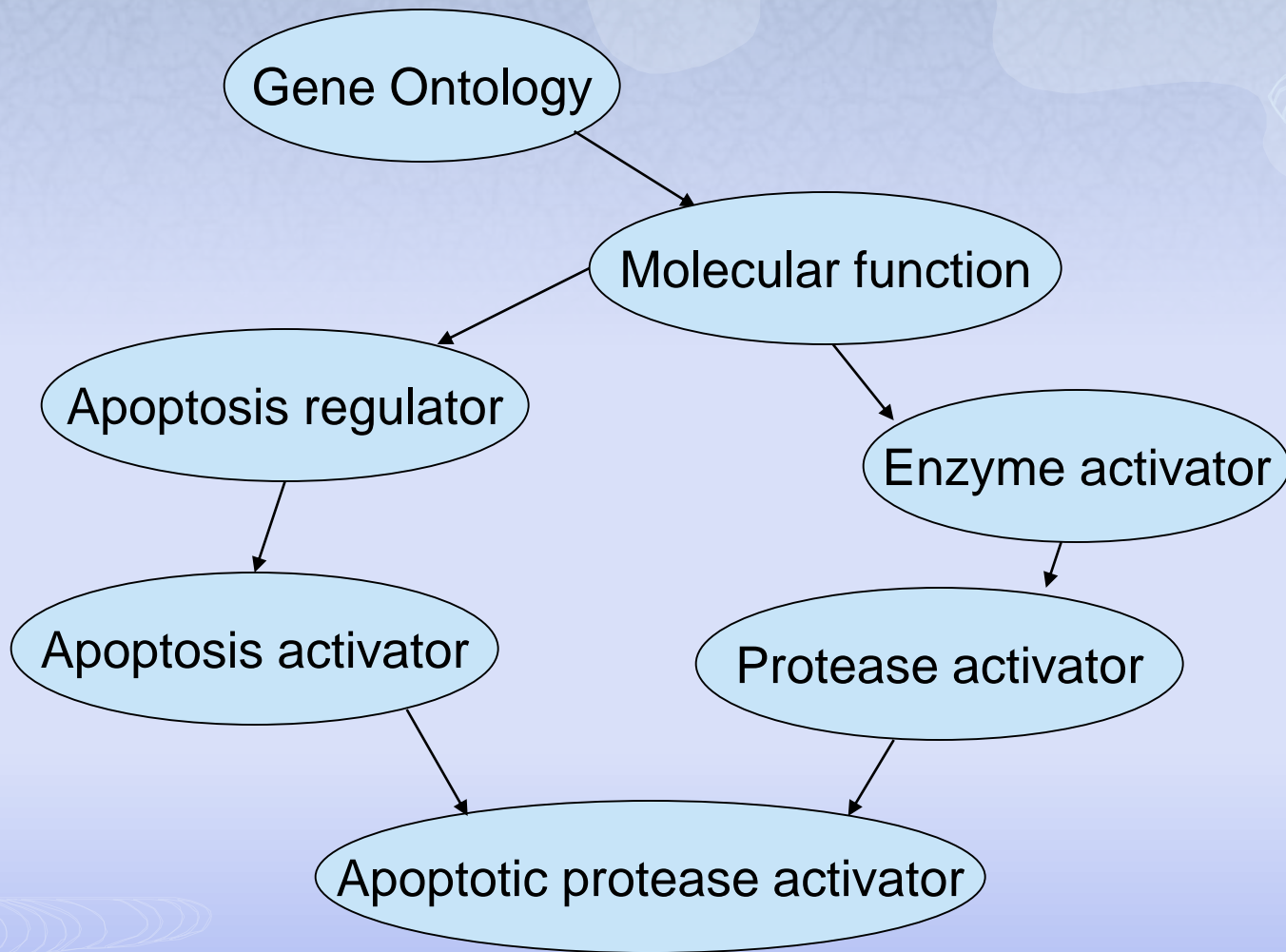


EMBL-EBI



The Scope of GO

- ❑ Gene products
- ❑ Attributes of sequence such as intron/exon parameters
- ❑ Protein domains or structural features
- ❑ Protein-protein interactions
- ❑ Environment, evolution and expression
- ❑ Anatomical or histological features above the level of cellular components, including cell types



GO Is Not a

- ❑ GO is not a database of gene sequences, nor a catalog of gene products
- ❑ GO is not a dictated standard, mandating nomenclature across databases
- ❑ GO is not a way to unify biological databases

存储的数据类型及更新情况

- GO的所有数据都是免费获得的。GO数据有三种格式：flat（每日更新）、XML(每月更新)和MySQL（每月更新）。这些数据格式都可以在GO ftp的站点上下载。XML和MySQL文件是被储存于独立的GO数据库中。
- 如果需要找到与某一个GO术语相关的基因或基因产物，可以找到一个相应表格，搜寻到这种注解的编号，并且可以链接到与之对应的位于不同数据库的基因相关文件。

数据库容量及数据更新

- 如上所述，GO的三层结构是分子功能、生化途径和细胞组件。GO包含的大部分为平板格式文件（GO flat file），由每一种本体论中定义的文件为文本文件，而包含本体论和定义两种格式的是OBO格式的平板文件，XML作为可以用于三种本体论和所有定义的文件格式也有提供。这些文件都在每月的1日更新，GO每月将给出月份更新报告。

查询界面及使用

Welcome to the Gene Ontology website!

The Gene Ontology project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. The project provides [a controlled vocabulary of terms](#) for describing gene product characteristics and [gene product annotation data](#) from GO Consortium members, as well as [tools to access and process this data](#). [Read more about the Gene Ontology...](#)

Search the Gene Ontology Database

Search for genes, proteins or GO terms using [AmiGO](#):



gene or protein name GO term or ID

[AmiGO](#) is the official GO browser and search engine. [Browse the Gene Ontology with AmiGO](#).

查询结果

61 results for **p450** in genes or proteins fields **symbol, full name(s) and synonyms**

Current filters

Ontology: biological process

Species: _

Filter search results ?

Filter Gene Products

Gene Product Type	Data source	Species
All	All	Colwellia psychre...
gene	ASAP	Danio rerio
gene product	AspGD	Dictyostelium dis...
mirna	CGD	Drosophila melano...

Filter Gene Products by Associations

Ontology	Evidence Code
All	All
biological process	IMR
cellular component	IRD
molecular function	IC

Set filters

Remove all filters

Results are sorted by **relevance**. To change the sort order, click on the column headers.

1 2 View all results

Select all Clear all Perform an action with this page's selected gene products... Go!

rel ↓

Symbol, full name

<input type="checkbox"/>	Cpr Cytochrome p450 reductase Query matches synonyms NADPH-cytochrome p450 oxidoreductase, and 2 more	view associations BLAST	protein f
<input type="checkbox"/>	Cyp12a4 Query matches synonyms p450 , and 1 more	view associations BLAST	protein f
<input type="checkbox"/>	Cyp18a1 Cytochrome p450 -18a1 Query matches synonyms Cyt- p450 -17C, and 3 more	view associations BLAST	protein f
<input type="checkbox"/>	Cyp4ae1 Cytochrome p450 -4ae1 Query matches synonym p450	view associations BLAST	protein f



evidence code

- ❑ EXP: Inferred from Experiment (实验推论)
- ❑ IDA: Inferred from Direct Assay (化验)
- ❑ IPI: Inferred from Physical Interaction (双杂交、免疫共沉淀)
- ❑ IMP: Inferred from Mutant Phenotype (变异显形)
- ❑ IGI: Inferred from Genetic Interaction (干扰、功能互补等)
- ❑ IEP: Inferred from Expression Pattern (芯片、northern, southern等)
- ❑ ISS: Inferred from Sequence or Structural Similarity (序列或者结构相似)
- ❑ ISO: Inferred from Sequence Orthology (同源序列)

- ❑ ISA: Inferred from Sequence Alignment (序列比对)
- ❑ ISM: Inferred from Sequence Model (模板序列)
- ❑ IGC: Inferred from Genomic Context (操纵子结构, 通路分析等)
- ❑ RCA: Inferred from Reviewed Computational Analysis (计算分析)
- ❑ TAS: Traceable Author Statement (作者陈述)
- ❑ NAS: Non-traceable Author Statement (无可追溯作者)
- ❑ IC: Inferred by Curator (专家推断)
- ❑ ND: No biological Data available (功能未知)
- ❑ IEA: Inferred from Electronic Annotation (电子注释)
- ❑ NR: Not Recorded (无记录)

用途

GO的应用前景很广阔，不可能一一列出，现在已用到的包括：

1. 整合来自于不同生物的蛋白组信息。
2. 判定蛋白结构域的功能。
3. 找到在疾病/衰老中异常表达的基因的功能类似性。
4. 预测与一种疾病相关的基因。
5. 分析在发育中同时表达的基因。
6. 建立起自动的能从文献中获取基因功能信息的工具。

Thank you!