

搞懂KEGG 数据库

王鹏

科技服务技术支持部

KEGG

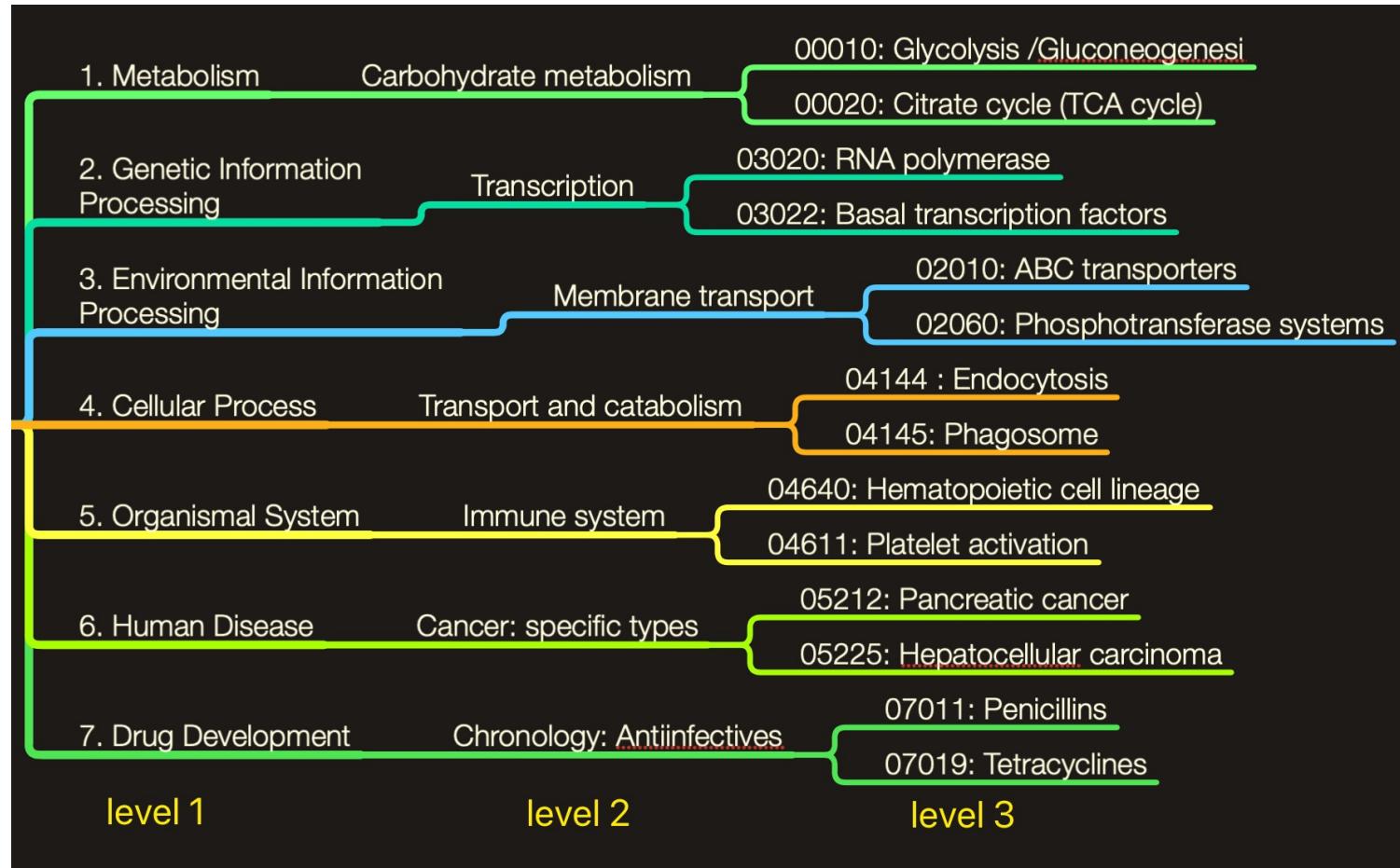
Category	Database	Content	Color
Systems information	KEGG PATHWAY	KEGG pathway maps	
	KEGG BRITE	BRITE hierarchies and tables	
	KEGG MODULE	KEGG modules and reaction modules	
Genomic information	KEGG ORTHOLOGY (KO)	Functional orthologs	
	KEGG GENES	Genes and proteins	
	KEGG GENOME	KEGG organisms and viruses	
Chemical information	KEGG COMPOUND	Small molecules	
	KEGG GLYCAN	Glycans	
	KEGG REACTION / RCLASS	Reactions and reaction class	
	KEGG ENZYME	Enzyme nomenclature	
Health information	KEGG NETWORK	Disease-related network variations	
	KEGG VARIANT	Human gene variants	
	KEGG DISEASE	Human diseases	
	KEGG DRUG / DGROUP	Drugs and drug groups	

KEGG (Kyoto Encyclopedia of Genes and Genomes , 京都基因和基因组百科全书) 于1995年由Kanehisa 实验室开发 , 目前由两个单位整理维护 , 京都大学和东京大学。

整理的信息分为四类 :
系统信息 ; 遗传信息
化学信息 ; 健康信息
共包含16个数据库

谁干活 (基因) , 在哪里干活 (基因组) , 干什么活 (酶 / 化合物 / 多糖) , 对个体有什么影响 (变异或疾病) ; 合起伙来做事情 (pathway/module)

KEGG PATHWAY



Level 1: 7大类

Level 2 : 58类

Level 3 : 547个

每个编号前会有map、 ko和hsa等前缀

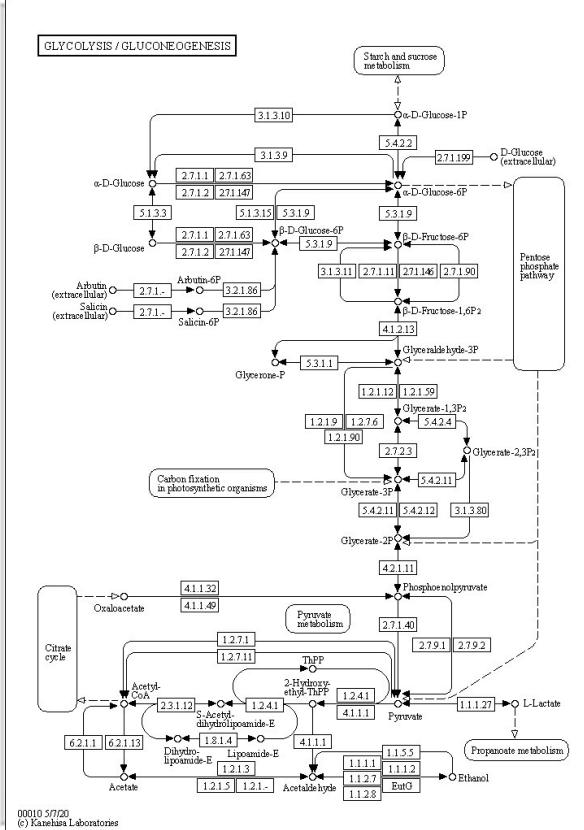
map : 代谢通路图

ko : 与map类似，有浅蓝标注的表示该酶有链接到其他单元

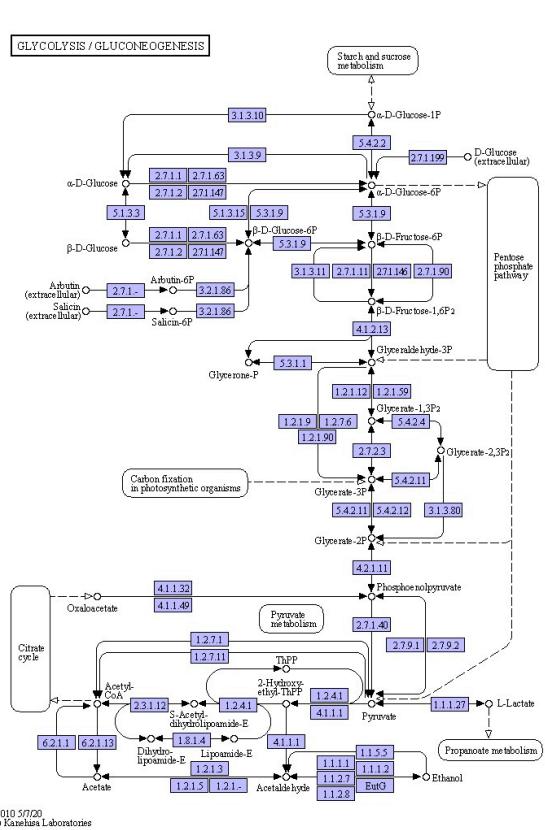
hsa : 该物种的代谢通路图，有浅绿标注的表示该基因或酶在该物种中存在

Pathway不同前缀示例

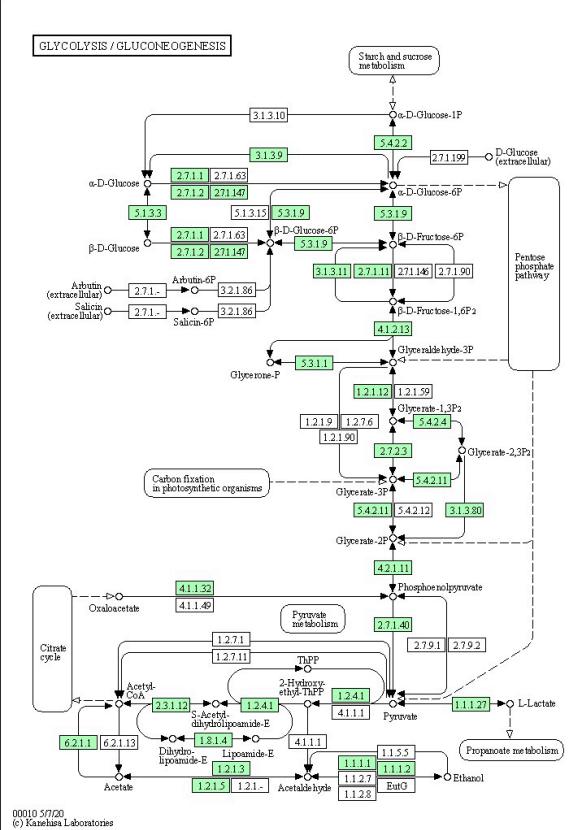
map00010



ko00010

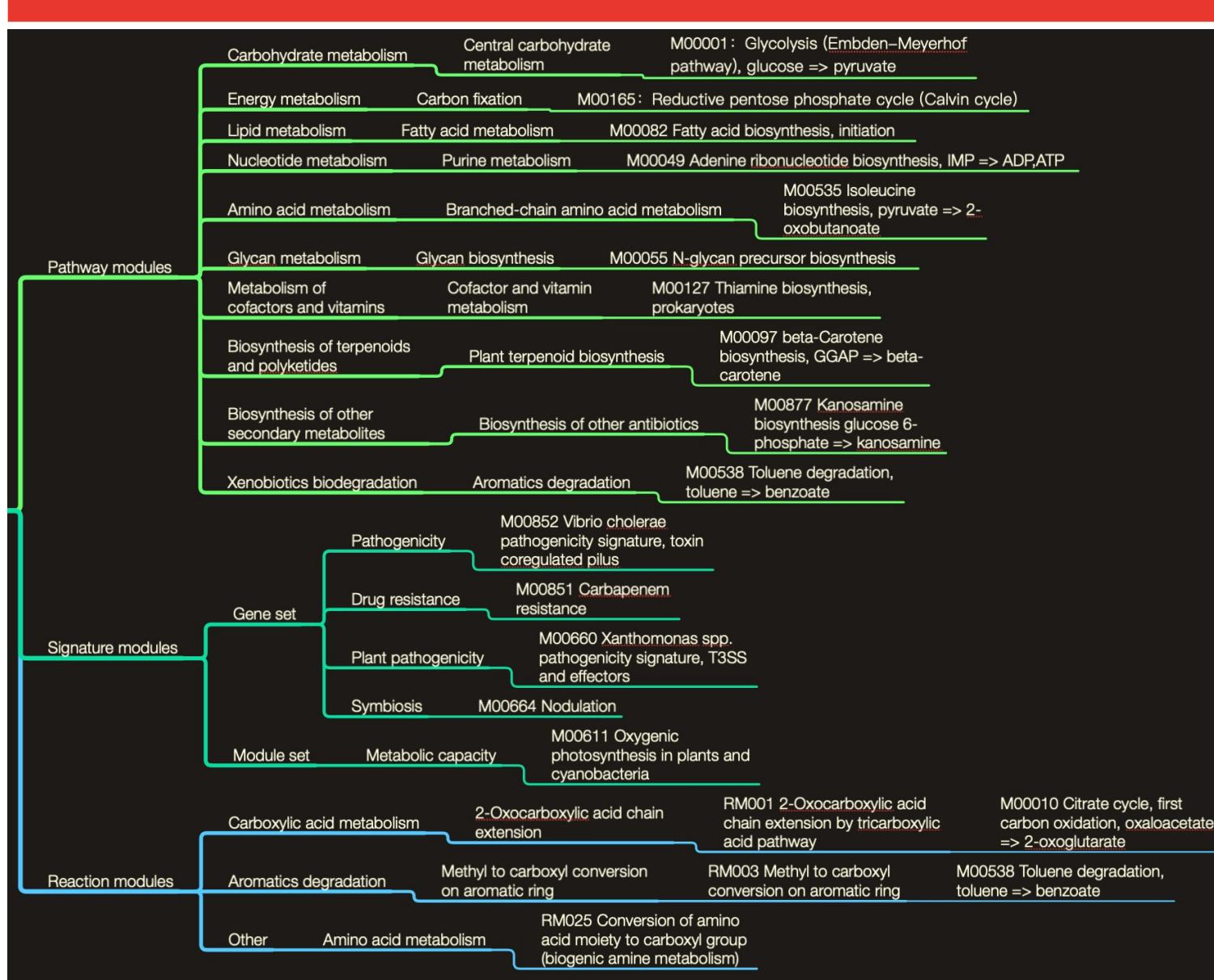


hsa00010



<https://www.genome.jp/pathway/map00010>
<https://www.genome.jp/pathway/ko00010>
<https://www.genome.jp/pathway/hsa00010>

KEGG MODULE



主要分成三部分

1) pathway modules : 代谢通路
中基因集的功能单元

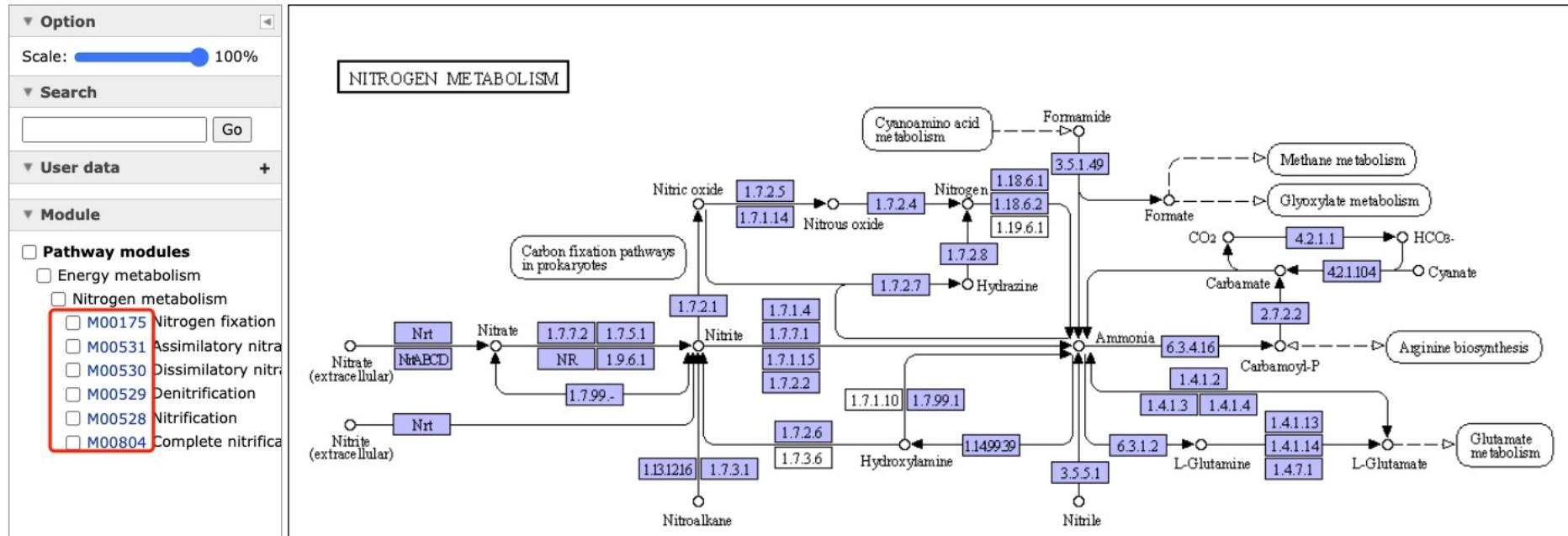
2) signature modules : 涉及表
型特征的基因集的功能单元

3) reaction modules : 代谢通路
中连续反应的功能单元

目前KEGG共分类了446个功能
单元

Pathway与Module

<https://www.genome.jp/pathway/ko00910>



- Pathway是代谢通路，module是功能单元
- 一个pathway可包含多个module（如上图ko00910，包含硝化，反硝化，异化和同化硝酸盐还原反应）
- 一个module可以跨越多个pathway（如M00175，可参与ko00190氮代谢，也参与ko00195光合作用）

BRITE

Functional Hierarchies

KEGG BRITE is a collection of hierarchical classification systems capturing functional hierarchies of various biological objects, especially those represented as [KEGG objects](#). They are represented as BRITE hierarchy files, also called [hierarchical text \(htext\) files](#), supplemented with BRITE table files using html tables. BRITE hierarchy files may contain tab-delimited attributes that are added manually or computationally. BRITE table files focus more on such multi-column attributes rather than hierarchy relationships. In contrast to KEGG PATHWAY, which is limited to molecular interactions and reactions, KEGG BRITE incorporates many different types of relationships including:

- [1. Genes and Proteins](#)
- [2. Compounds and Reactions](#)
- [3. Drugs](#)
- [4. Diseases](#)
- [5. Organisms and Cells](#)

<https://www.genome.jp/kegg/brite.html#gene>

KEGG BRITE is the reference database for brite mapping in [KEGG Mapper](#).

KEGG Enzymes

[Brite menu | Download htext | Download json | Help]

Change brite type

▼ Option □ One-click mode □ Row border □ shading

▼ Search Go

▼ ID search Go

▼ Join Pathway

▼ Related Brite table br01800 Sequence data for EC

▼ Option □ One-click mode □ Row border □ shading

▼ Search Go

▼ ID search Go

▼ Join Pathway

▼ Related Brite table br01800 Sequence data for EC

▼ 1. Oxidoreductases
1.1 Acting on the CH-OH group of donors
1.1.1 With NAD+ or NADP+ as acceptor
1.1.1.1 alcohol dehydrogenase
K00001 E1.1.1.1; adh; alcohol dehydrogenase [EC:1.1.1.1]
K00121 frmA; ADH5; adhc; S-(hydroxymethyl)glutathione dehydrogenase / alcohol dehydrogenase [EC:1.1.1.284 1.1.1.1]
K04072 adhE; acetaldehyde dehydrogenase / alcohol dehydrogenase [EC:1.2.1.10 1.1.1.1]
K11440 gbsB; choline dehydrogenase [EC:1.1.1.1]
K13951 ADH1_7; alcohol dehydrogenase 1/7 [EC:1.1.1.1]
K13952 ADH6; alcohol dehydrogenase 6 [EC:1.1.1.1]
K13953 adhP; alcohol dehydrogenase, propanol-preferring [EC:1.1.1.1]
K13954 yiaY; alcohol dehydrogenase [EC:1.1.1.1]
K13980 ADH4; alcohol dehydrogenase 4 [EC:1.1.1.1]
K18857 ADH1; alcohol dehydrogenase class-P [EC:1.1.1.1]
1.1.1.2 alcohol dehydrogenase (NAD+)
1.1.1.3 homoserine dehydrogenase
1.1.1.4 (R,R)-butanediol dehydrogenase
1.1.1.6 glycerol dehydrogenase
1.1.1.7 propanediol-phosphate dehydrogenase

<https://www.genome.jp/brite/ko01000>

KO

KO Database of Molecular Functions

The **KO (KEGG Orthology)** database is a database of molecular functions represented in terms of functional orthologs. A functional ortholog is manually defined in the context of KEGG molecular networks, namely, KEGG pathway maps, BRITE hierarchies and KEGG modules. Each node of the network, such as a box in the KEGG pathway map, is given a KO identifier (called K number) as a functional ortholog defined from experimentally characterized genes and proteins in specific organisms, which are then used to assign orthologous genes in other organisms based on sequence similarity. The granularity of "function" is context-dependent, and the resulting KO grouping may correspond to a group of highly similar sequences within a limited organism group or it may be a more divergent group.

The KO system is a network-based classification of KOs shown below:

KEGG Orthology (KO)

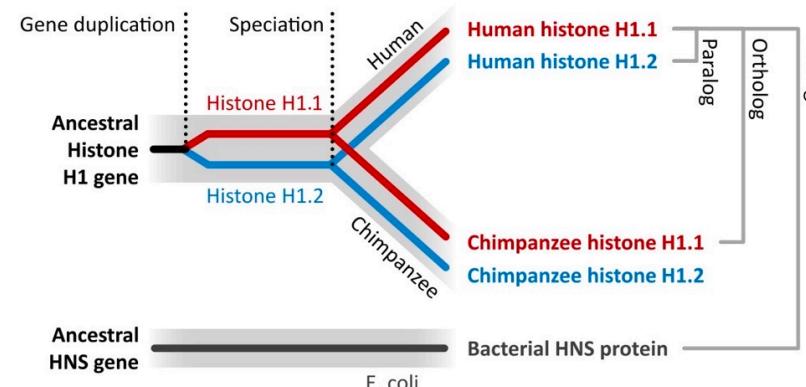
It consists of six top categories (09100 to 09160) for KEGG pathway maps and one top category (09180) for BRITE hierarchies, as well as one top category (09190) for those KOs that are not yet included in either of them. The category numbers for these top categories and the second-level categories under metabolism (09101 to 09112) are used to define color coding of functions (see [KEGG Color Codes](#)).

Major efforts have been made to associate each KO entry with experimental evidence of functionally characterized sequence data as shown in the SEQUENCE subfield of the REFERENCE field. In many cases such data are not found in the complete genomes of KEGG organisms. The addendum category of the GENES database allows functionally characterized individual protein sequences to be included in KEGG. As a byproduct of these efforts, sequence data have also been associated with EC numbers in [Enzyme Nomenclature](#).

KO : KEGG Orthology (直系同源基因)

旁系同源：基因经过复制，发生功能分化

直系同源：物种分化后存在不同物种的相同基因



KO

KEGG Orthology (KO)

[Brite menu | Download htext | Download json | Help]

Change brite type

▼ Option

One-click mode
 Row border shading

▼ Search

Go

▼ ID search

Go

▼ Join

09100 Metabolism

09120 Genetic Information Processing

09130 Environmental Information Processing

09140 Cellular Processes

09150 Organismal Systems

09160 Human Diseases

09180 Brite Hierarchies ←

09181 Protein families: metabolism

09182 Protein families: genetic information processing

09183 Protein families: signaling and cellular processes

09185 Viral protein families

09184 RNA family

09190 Not Included in Pathway or Brite ←

09191 Unclassified: metabolism

09192 Unclassified: genetic information processing

09193 Unclassified: signaling and cellular processes

▼ 09100 Metabolism

▼ 09101 Carbohydrate metabolism

00010 Glycolysis / Gluconeogenesis [PATH:ko00010]
K00844 HK; hexokinase [EC:2.7.1.1]
K12407 GCK; glucokinase [EC:2.7.1.2]
K00845 glk; glucokinase [EC:2.7.1.2]
K25026 glk; glucokinase [EC:2.7.1.2]
K01810 GPI; pgi; glucose-6-phosphate isomerase [EC:5.3.1.9]
K06859 pgii; glucose-6-phosphate isomerase, archaeal [EC:5.3.1.9]
K13810 tal-pgi; transaldolase / glucose-6-phosphate isomerase [EC:2.2.1.2 5.3.1.9]
K15916 pgi-pmi; glucose/mannose-6-phosphate isomerase [EC:5.3.1.9 5.3.1.8]
K24182 PFK9; 6-phosphofructokinase [EC:2.7.1.11]
K00850 pfkA, PFK; 6-phosphofructokinase 1 [EC:2.7.1.11]
K16370 pfkB; 6-phosphofructokinase 2 [EC:2.7.1.11]
K21071 pfk, pfp; ATP-dependent phosphofructokinase / diphosphate-dependent phosphof
K00918 pfkC; ADP-dependent phosphofructokinase/glucokinase [EC:2.7.1.146 2.7.1.147]
K00895 pfp, PFP; diphosphate-dependent phosphofructokinase [EC:2.7.1.90]
K03841 FBP, fbp; fructose-1,6-bisphosphatase I [EC:3.1.3.11]
K02446 glpX; fructose-1,6-bisphosphatase II [EC:3.1.3.11]
K11532 glpX-SEBP; fructose-1,6-bisphosphatase II / sedoheptulose-1,7-bisphosphatase
K01086 fbp-SEBP; fructose-1,6-bisphosphatase I / sedoheptulose-1,7-bisphosphatase [
K04041 fbp3; fructose-1,6-bisphosphatase III [EC:3.1.3.11]
K01623 ALDO; fructose-bisphosphate aldolase, class I [EC:4.1.2.13]
K11645 fbaB; fructose-bisphosphate aldolase, class I [EC:4.1.2.13]

<https://www.genome.jp/entry/K00844>

GENES

GENES Database

KEGG GENES is a collection of genes and proteins in complete genomes of cellular organisms and viruses generated from publicly available resources, mostly from NCBI RefSeq and GenBank, and annotated by KEGG in the form of KO (KEGG Orthology) assignment. The collection is supplemented with a KEGG original collection of functionally characterized proteins from published literature. Protein sequences and RNA sequences of all GENES entries are subject to SSDB computation and KO assignment by KOALA tools (see [annotation statistics](#)).

Category	Content	Data source	Organism code	Gene identifier
KEGG organisms (Complete genomes)	Genes and proteins in cellular organisms	RefSeq or GenBank	<org>	GeneID or Locus_tag
Viruses	Genes and proteins in viruses	RefSeq	vg	GeneID
	Mature peptides in viruses		vp	GeneID+no
Addendum	Functionally characterized proteins	KEGG	ag	ProteinID, etc

<org> three- or four-letter organism code for cellular organisms

The Addendum category is a PubMed-based collection of protein sequences whose functions are experimentally characterized. They are used to define new KOs that are not covered by complete genomes (see [KO database](#)).

The viral peptide (vp) category is a collection of mature peptides processed from genome-encoded polyproteins, which are not usually found as separate entries in the public databases such as NCBI and UniProt. Viral mature peptides appear in KEGG pathway maps and as drug targets and are given KOs.

Each GENES entry is identified by the combination of organism code and gene identifier in the form of

org:gene

such as [hsa:351](#) for human amyloid beta gene.

理清关系

- 最小的功能单元是**基因 (GENES , 物种名:基因编号 , mmu:100861531)**
- 基因的产物为**酶 (ENZYME , EC*****)** , 酶具有催化功能生成化学物质 (**COMPOUNDS , C*******)
- 基因存在于不同的**物种 (GENOME , T*****)** 中 , 行使相似功能的基因聚成一类 , 为**KO (K*****)**
- 很多KO彼此关联 , 共同行使特定类型的功能模块 , 为**MODULE (M*****)**
- 很多KO彼此关联 , 组成某一特定途径的通路图 , 为**PATHWAY (map/ko/物种缩写*****)**



KO和ko到底什么区别 ?

Pathway和Module的区别是什么 ?

KEGG API

<https://www.kegg.jp/kegg/rest/keggapi.html>

KEGG API

KEGG API is a REST-style Application Programming Interface to the KEGG database resource.

[[Top](#) | [KEGG API](#) | [MEDICUS Extension](#) | [KEGG WebLinks](#) | [KEGG Database Entry Format](#)]

General form

URL form

```
http://rest.kegg.jp/<operation>/<argument>[/<argument2>/<argument3> ...]  
<operation> = info | list | find | get | conv | link | ddi
```

Database name

```
<database> = KEGG databases (Table 1), KEGG MEDICUS extension databases (Table 2)  
and Outside databases integrated in KEGG (Table 3)  
  
= pathway | brite | module | ko | genome | <org> | vg | vp | ag | compound |  
glycan | reaction | rclass | enzyme | network | variant | disease |  
drug | dgroup | genes | ligand | kegg | <medicus> | <outside>  
  
<org> = KEGG organism code  
  
<medicus> = disease_ja | drug_ja | dgroup_ja | compound_ja | brite_ja |  
atc | jtc | ndc | yj  
  
<outside> = pubmed | ncbi-geneid | ncbi-proteinid | uniprot | pubchem | chebi
```

list

<http://rest.kegg.jp/list/<database>>

<database> = pathway | brite |
module | ko | genome | <org> | vg | vp
| ag | compound | glycan | reaction |
rclass | enzyme | network | variant |
disease | drug | dgroup | organism |
<medicus>

所有的酶

<http://rest.kegg.jp/list/enzyme>

所有的化合物

<http://rest.kegg.jp/list/compound>

列出某一物种的所有基因：

<http://rest.kegg.jp/list/hsa>

<http://rest.genome.jp/list/hsa>

查找某一物种的pathway：

<http://rest.kegg.jp/list/pathway/hsa>

所有的物种：

<http://rest.kegg.jp/list/genome>

<http://rest.kegg.jp/list/organism>

所有的通路

<http://rest.kegg.jp/list/pathway>

所有的模块

<http://rest.kegg.jp/list/module>

所有的KO

<http://rest.kegg.jp/list/ko>

find

<http://rest.kegg.jp/find/<database>/<query>>

<database> = pathway | brite | module | ko | genome | genes | <org> | vg | vp | ag | ligand | compound | glycan | reaction | rclass | enzyme | network | variant | disease | drug | dgroup | <medicus>

查找包含某些关键词的基因：

<http://rest.kegg.jp/find/genes/acid+methyl>

查找包含某些关键词的通路：

<http://rest.kegg.jp/find/pathway/acid>

查找某一化合物的编号：

<http://rest.kegg.jp/find/compound/C7H10O5/formula>

get

`http://rest.kegg.jp/get/<dbentries>[/<option>]`

`<database>` = pathway | brite | module | ko | genome |
`<org>` | vg | vp | ag | compound | glycan | reaction |
rclass | enzyme | network | variant | disease | drug |
dgroup | disease_ja | drug_ja | dgroup_ja |
compound_ja
`<option>` = aaseq | ntseq | mol | kcf |
image | conf | kgml | json

获取某一具体基因的蛋白序列：

<http://rest.kegg.jp/get/hsa:10458/aaseq>

获取某一具体基因的核苷酸序列

<http://rest.kegg.jp/get/hsa:10458/ntseq>

获取某一化合物的化学分子式：

<http://rest.kegg.jp/get/C00002/image>

获取某一物种的代谢通路图：

<http://rest.kegg.jp/get/hsa05130/image>

获取某一通路/KO/基因等的具体信息：

<http://rest.kegg.jp/get/ko00010>

<http://rest.kegg.jp/get/K00844>

<http://rest.kegg.jp/get/hsa:3098>

link

`http://rest.kegg.jp/link/<target_db>/<source_db>`
`<target_db> = <database>`
`<source_db> = <database>`

`<database> = pathway | brite | module | ko | genome |`
`<org> | vg | vp | ag | compound | glycan | reaction |`
`rclass | enzyme | network | variant | disease | drug |`
`dgroup | atc | jtc | ndc | yj | pubmed`

某一物种中所有基因对应的pathway：
<http://rest.kegg.jp/link/hsa/pathway>

特定基因对应的pathway
<http://rest.kegg.jp/link/pathway/hsa:10458>

某一pathway对应的KO：
<http://rest.kegg.jp/link/ko/ko00010>

某一同源基因（KO）对应的具体基因：
<http://rest.kegg.jp/link/genes/K00500>

某一通路对应的化合物信息：
<http://rest.kegg.jp/link/cpd/map00010>



TCGA ANALYSIS Thank You!



官方网站



官方微 信

www.berrygenomics.com