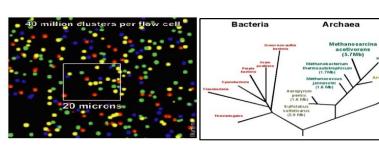


TAACCCTAACCCTAACCCTAACCCTA CCTAACCCTAACCCTAACCCTAACCC CCCTAACCCTAACCCTAACCCTAAC AACCCTAACCCTAACCCTAACCCTA ACCCTAACCCCAACCCCAACCCCAAC CTACCCTAACCCTAACCCTAACCCTA ACCCTAACCCTAACCCTAACCCTAACCCTAA

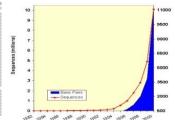


# Ontology, and Identification of Molecular Pathways

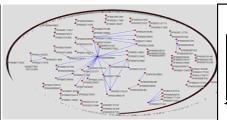
北京大学生物信息学中心 魏丽萍 Liping Wei, Ph.D.

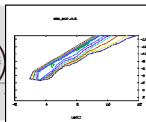
Center for Bioinformatics, Peking University





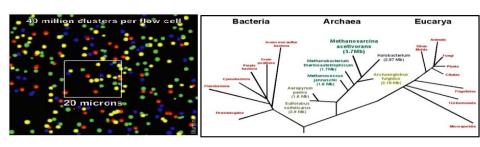








TAACCCTAACCCTAACCCTAACCCTA CCTAACCCTAACCCTAACCCTAACCC CCCTAACCCTAACCCTAACCCTAAC AACCCTAACCCTAACCCTAACCCTA ACCCTAACCCCAACCCCAACCCCAAC CTACCCTAACCCTAACCCTAACCCTA ACCCTAACCCTAACCCTAACCCTAACCCTAA

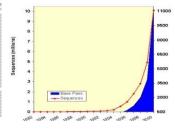


# Unit 2: KEGG Pathway Database

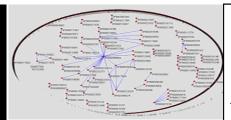
北京大学生物信息学中心 魏丽萍 Liping Wei, Ph.D.

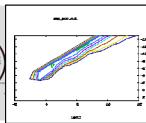
Center for Bioinformatics, Peking University











# What is a biological pathway?

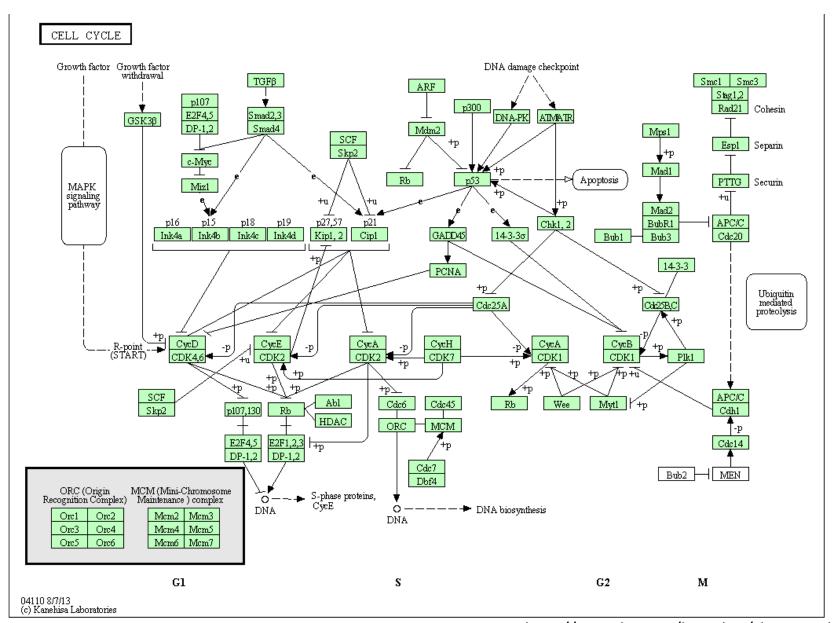
A series of actions among molecules in a cell that lead to a certain product or a change in a cell.

Main types of biological pathways:

Metabolic pathways

Gene regulation pathways

Signal transduction pathways



Copyright © Peking University http://www.kegg.jp/kegg-bin/show\_pathway?hsa04110

# Pathway databases

**KEGG PATHWAY** 

BioCarta

BioCyc

Protein ANalysis Through Evolutionary Relationships (PANTHER)

Pathway Interaction Database (PID)

Reactome

## **KEGG PATHWAY**

http://www.kegg.jp/kegg/



KEGG: Kyoto Encyclopedia of Genes and Genomes



### **Data-oriented entry points**

KEGG PATHWAY	KEGG pathway maps [Pathway list]		
KEGG BRITE	BRITE functional hierarchies [Brite list]		
<b>KEGG MODULE</b>	KEGG modules [Module list]		
<b>KEGG DISEASE</b>	Human diseases [Cancer   Infectious disease]		
KEGG DRUG	Drugs [ATC drug classification]		
<b>KEGG ORTHOLOGY</b>	Ortholog groups [KO system]		
<b>KEGG GENOME</b>	Genomes [KEGG organisms]		
<b>KEGG GENES</b>	Genes and proteins Release history		
<b>KEGG COMPOUND</b>	Small molecules [Compound classification]		
<b>KEGG REACTION</b>	Biochemical reactions [Reaction modules]		

## KEGG PATHWAY



## **KEGG Pathway Maps**

[ Brite menu | Download htext ]

KEGG pathway maps

- v v v
- One-click mode
- Metabolism
- Genetic Information Processing
- Environmental Information Processing
- Cellular Processes
- Organismal Systems
- Human Diseases
- Drug Development

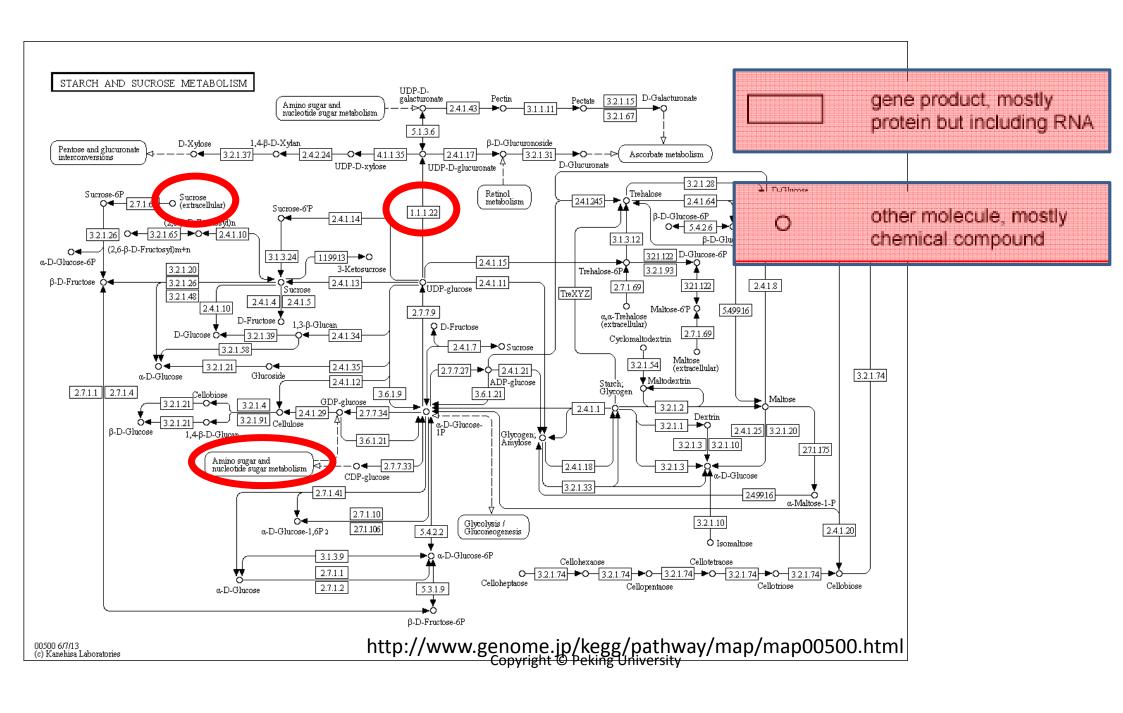
http://www.kegg.jp/kegg-bin/get\_htext?br08901.keg

## **KEGG Pathway Maps**

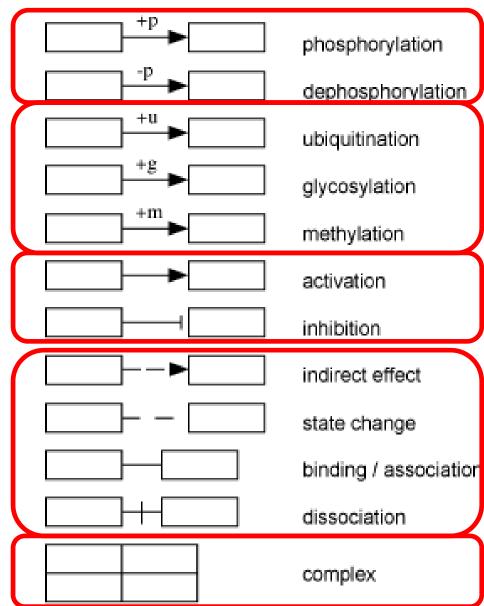
#### [ Brite menu | Download htext ]

KEGG pathway maps One-click mode Metabolism Global and overview maps 01100 Metabolic pathways 01110 Biosynthesis of secondary metabolites 01120 Microbial metabolism in diverse environments 01200 Carbon metabolism 01210 2-Oxocarboxylic acid metabolism 01212 Fatty acid metabolism 01230 Biosynthesis of amino acids 01220 Degradation of aromatic compounds Carbohydrate metabolism 00010 Glycolysis / Gluconeogenesis 00020 Citrate cycle (TCA cycle) 00030 Pentose phosphate pathway 00040 Pentose and glucuronate interconversions 00051 Fructose and mannose metabolism 00052 Galactose metabolism Ascorbate and aldarate metabolism 00500 Starch and sucrose metabolism 00520 Amino sugar and nucleotide sugar metabolism 00620 Pyruvate metabolism 00630 Glyoxylate and dicarboxylate metabolism Copyright © Peking University

Go

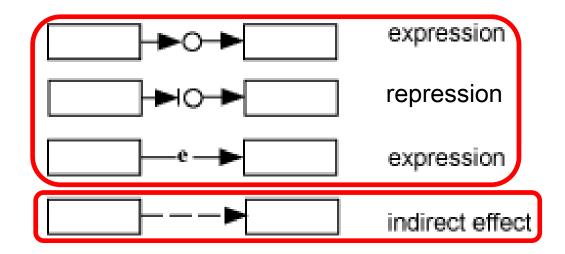


# Protein-protein interactions



http://www.kegg.jp/kegg/document/help\_pathway.html

### Gene expression relations



### Enzyme-enzyme relations



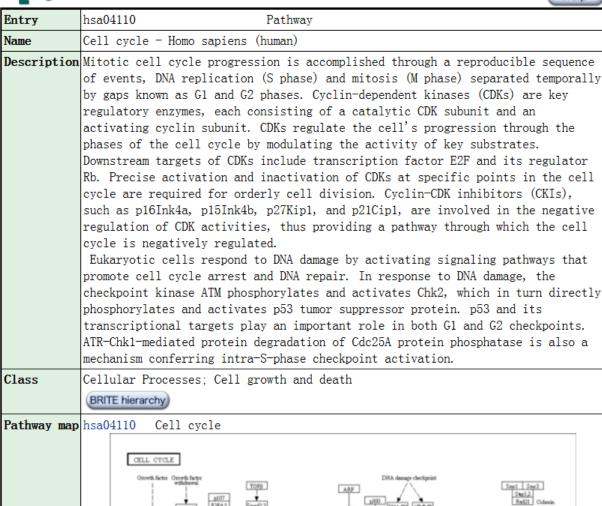
http://www.kegg.jp/kegg/document/help\_pathway.html



#### PATHWAY: hsa04110

Help

Pathway
entry

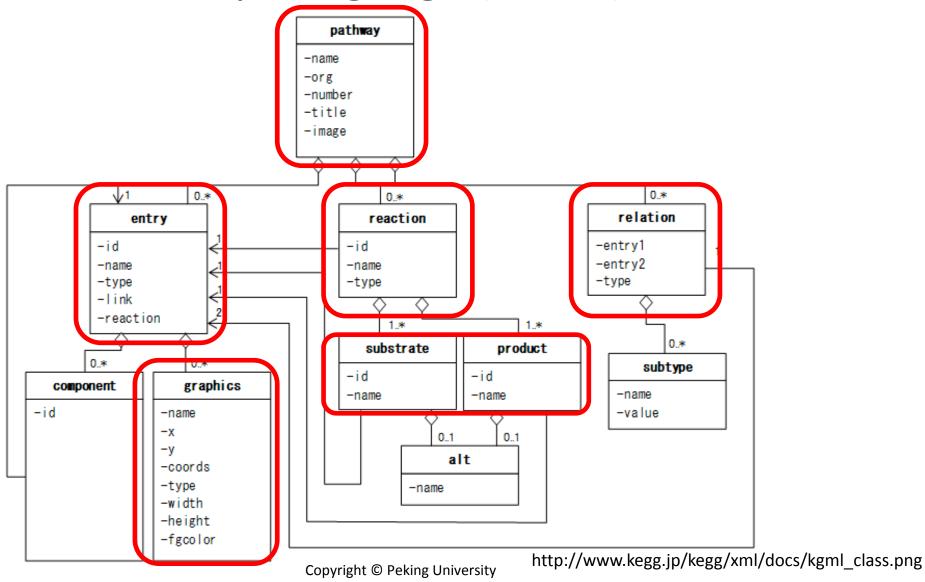


http://www.kegg.jp/dbget-bin/www\_bget?pathway+hsa04110 Copyright © Peking University

## Flat-file format

```
ENTRY
           hsa04110
                                        Pathway
NAME
           Cell cycle - Homo sapiens (human)
DESCRIPTION Mitotic cell cycle progression is accomplished through a reproducible sequence of events, D
phase) and mitosis (M phase) separated temporally by gaps known as G1 and G2 phases. Cyclin-dependent k
key regulatory enzymes, each consisting of a catalytic CDK subunit and an activating cyclin subunit. CD
ll's progression through the phases of the cell cycle by modulating the activity of key substrates. Dow
CDKs include transcription factor E2F and its regulator Rb. Precise activation and inactivation of CDK
ts in the cell cycle are required for orderly cell division. Cyclin-CDK inhibitors (CKIs), such as p16I
7Kip1, and p21Cip1, are involved in the negative regulation of CDK activities, thus providing a pathway
cell cycle is negatively regulated.
             Eukaryotic cells respond to DNA damage by activating signaling pathways that promote cell
NA repair. In response to DNA damage, the checkpoint kinase ATM phosphorylates and activates Chk2, whic
phosphorylates and activates p53 tumor suppressor protein. p53 and its transcriptional targets play an
both G1 and G2 checkpoints. ATR-Chk1-mediated protein degradation of Cdc25A protein phosphatase is als
erring intra-S-phase checkpoint activation.
CLASS
           Cellular Processes; Cell growth and death
PATHWAY MAP hsa04110 Cell cycle
MODULE
           hsa M00284 Origin recognition complex [PATH:hsa04110]
            hsa M00285 MCM complex [PATH:hsa04110]
            hsa M00297 DNA-PK complex [PATH:hsa04110]
            hsa M00381 SCF-SKP2 complex [PATH:hsa04110]
           hsa M00389 APC/C complex [PATH:hsa04110]
```

## KEGG Markup Language (KGML) format



## KGML format

```
<pathway name="path:hsa04110" org="hsa" number="04110"
         title="Cell cycle"
         image="http://www.kegg.jp/kegg/pathway/hsa/hsa04110.png"
         link="http://www.kegg.jp/kegg-bin/show_pathway?hsa04110">
    <entry id="1" name="hsa:1029" type="gene"</pre>
        link="http://www.kegg.jp/dbget-bin/www bget?hsa:1029">
        <graphics name="CDKN2A, ARF, CDK4I, CDKN2, CMM2, INK4, INK4A, MLM, MTS-1, MTS1, P14, P14ARF, P16, P16</pre>
NK4, P16INK4A, P19, P19ARF, TP16" fgcolor="#000000" bgcolor="#BFFFBF"
             type="rectangle" x="532" y="124" width="46" height="17"/>
    Kentry id="2" name="hsa:51343" type="gene"
        link="http://www.kegg.jp/dbget-bin/www bget?hsa:51343">
        <graphics name="FZR1, CDC20C, CDH1, FZR, FZR2, HCDH, HCDH1" fgcolor="#000000" bgcolor="#BFFFBF"</pre>
             type="rectangle" x="919" y="536" width="46" height="17"/>
    (/entry)
    <entry id="3" name="hsa:4171 hsa:4172 hsa:4173 hsa:4174 hsa:4175 hsa:4176" type="gene"</pre>
        link="http://www.kegg.jp/dbget-bin/www bget?hsa:4171+hsa:4172+hsa:4173+hsa:4174+hsa:4175+hsa:4176">
        <graphics name="MCM2, BM28, CCNL1, CDCL1, D3S3194, MITOTIN, cdc19..." fgcolor="#000000" bgcolor="#BFFF</pre>
             type="rectangle" x="553" y="556" width="46" height="17"/>
    </entry>
    <entry id="4" name="hsa:23594 hsa:23595 hsa:4998 hsa:4999 hsa:5000 hsa:5001" type="gene"</pre>
        link="http://www.kegg.jp/dbget-bin/www_bget?hsa:23594+hsa:23595+hsa:4998+hsa:4999+hsa:5000+hsa:5001">
        <graphics name="ORC6, ORC6L..." fgcolor="#000000" bgcolor="#BFFFBF"</pre>
```

```
<reaction id="85" name="rn:R00026" type="irreversible">
   <substrate id="222" name="cpd:C00185"/>
   cproduct id="193" name="cpd:C00221"/>
</reaction>
<reaction id="86" name="rn:R03920" type="reversible">
   <substrate id="176" name="cpd:C02336"/>
   cproduct id="194" name="cpd:C05345"/>
</reaction>
<reaction id="93" name="rn:R00802" type="irreversible">
   <substrate id="186" name="cpd:C00089"/>
   cproduct id="191" name="cpd:C00267"/>
   cproduct id="176" name="cpd:C02336"/>
</reaction>
<reaction id="95" name="rn:R00802" type="irreversible">
   <substrate id="186" name="cpd:C00089"/>
   cproduct id="191" name="cpd:C00267"/>
   cproduct id="176" name="cpd:C02336"/>
</reaction>
```

```
<relation entry1="49" entry2="47" type="GErel">
    <subtype name="expression" value="--&gt;"/>
</relation>
<relation entry1="49" entry2="32" type="GErel">
   <subtype name="expression" value="--&gt;"/>
</relation>
<relation entry1="47" entry2="23" type="PPrel">
    <subtype name="activation" value="--&gt;"/>
</relation>
<relation entry1="36" entry2="48" type="PPrel">
    <subtype name="inhibition" value="--|"/>
</relation>
<relation entry1="36" entry2="49" type="PPrel">
    <subtype name="inhibition" value="--|"/>
</relation>
<relation entry1="37" entry2="49" type="PPrel">
    <subtype name="activation" value="--&gt;"/>
</relation>
```

## Browse pathways

#### 1. Metabolism

#### 1.0 Global and overview maps

Metabolic pathways	[KEGG Atlas]	KEGG modules
Biosynthesis of secondary metabolites	[KEGG Atlas]	KEGG reaction modules
Microbial metabolism in diverse environments	[KEGG Atlas]	
Carbon metabolism	[KEGG Atlas]	
2-Oxocarboxylic acid metabolism	[KEGG Atlas]	
Fatty acid metabolism New!	[KEGG Atlas]	
Biosynthesis of amino acids	[KEGG Atlas]	
Degradation of aromatic compounds	[KEGG Atlas]	

#### 1.1 Carbohydrate metabolism

Glycolysis / Gluconeogenesis Enzymes Citrate cycle (TCA cycle) Pentose phosphate pathway

Copyright © Peking University

Pentose and glucuronate interconversions

Fructose and mannose metabolism

Galactose metabolism

Ascorbate and aldarate metabolism

Starch and sucrose metabolism

Amino sugar and nucleotide sugar metabolism

Pyruvate metabolism

Glyoxylate and dicarboxylate metabolism

Propanoate metabolism

Compounds with biological roles

http://www.kegg.jp/kegg/pathway.html

## Search and color pathways



#### **KEGG Mapper - Search&Color Pathway**

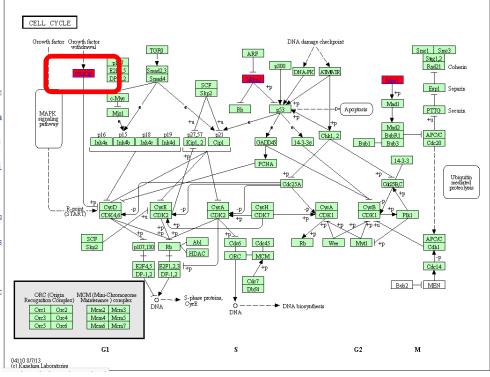


#### **Pathway Search Result**

Search&Color N Sort by the pathway list

#### Show all objects

- hsa04110 Cell cycle Homo sapiens (human) (3)
- hsa05200 Pathways in cancer Homo sapiens (hum
- hsa05169 Epstein-Barr virus infection Homo sa
- hsa05215 Prostate cancer Homo sapiens (human)
- hsa04151 PI3K-Akt signaling pathway Homo sapi
- hsa05219 Bladder cancer Homo sapiens (human)
- hsa05206 MicroRNAs in cancer Homo sapiens (hu
- hsa05205 Proteoglycans in cancer Homo sapiens
- hsa04012 ErbB signaling pathway Homo sapiens
- hsa04062 Chemokine signaling pathway Homo sap
- hsa04120 Ubiquitin mediated proteolysis Homo



http://www.kegg.jp/kegg/tool/map\_pathway2.html



#### [ Brite menu | Organism menu | Download htext ]

Reference hierarchy (KO)

- Metabolism
- Genetic Information Processing
- Environmental Information Processing
- Cellular Processes
- Organismal Systems
- Human Diseases



#### [ Brite menu | Organism menu | Download htext ]

Reference hierarchy (KO)





▼ ▼ ▼ ▼ ■ One-click mode

- ▼ Metabolism
  - Overview
  - Carbohydrate metabolism
  - Energy metabolism
  - Lipid metabolism
  - ► Nucleotide metabolism
  - Amino acid metabolism
  - Metabolism of other amino acids
  - Glycan biosynthesis and metabolism
  - ► Metabolism of cofactors and vitamins
  - Metabolism of terpenoids and polyketides
  - Biosynthesis of other secondary metabolites
  - Xenobiotics biodegradation and metabolism
  - ► Enzyme families



#### [ Brite menu | Organism menu | Download htext ]

Reference hierarchy (KO)



Go

▼ ▼ ▼ ▼ ■ One-click mode

#### ▼ Metabolism

#### Overview

#### Carbohydrate metabolism

- ▶ 00010 Glycolysis / Gluconeogenesis [PATH:ko00010]
- ▶ 00020 Citrate cycle (TCA cycle) [PATH:ko00020]
- ▶ 00030 Pentose phosphate pathway [PATH:ko00030]
- ▶ 00040 Pentose and glucuronate interconversions [PATH:ko00040]
- ▶ 00051 Fructose and mannose metabolism [PATH:ko00051]
- ▶ 00052 Galactose metabolism [PATH:ko00052]
- ▶ 00053 Ascorbate and aldarate metabolism [PATH:ko00053]
- ▶ 00500 Starch and sucrose metabolism [PATH:ko00500]
- ▶ 00520 Amino sugar and nucleotide sugar metabolism [PATH:ko00520]
- ▶ 00620 Pyruvate metabolism [PATH:ko00620]
- ▶ 00630 Glyoxylate and dicarboxylate metabolism [PATH:ko00630]
- ▶ 00640 Propanoate metabolism [PATH:ko00640]
- ▶ 00650 Butanoate metabolism [PATH:ko00650]
- ▶ 00660 C5-Branched dibasic acid metabolism [PATH:ko00660]
- ▶ 00562 Inositol phosphate metabolism [PATH:ko00562]



[ Brite menu | Organism menu | Download htext ] Reference hierarchy (KO) Go One-click mode Metabolism Overview Carbohydrate metabolism 00010 Glycolysis / Gluconeogenesis [PATH:ko00010] ▶ 00020 Citrate cycle (TCA cycle) [PATH:ko00020] 00030 Pentose phosphate pathway [PATH:k000030] 00040 Pentose and glucuronate interconversions [PATH:ko00040] ▶ 00051 Fructose and mannose metabolism [PATH:ko00051] 00052 Galactose metabolism [PATH:ko00052] 00053 Ascorbate and aldarate metabolism [PATH:k000053] ▼ 00500 Starch and sucrose metabolism [PATH:ko00500] K01212 sacC; levanase [EC:3.2.1.65] K00692 sacB; levansucrase [EC:2.4.1.10] K02808 PTS-Scr-EIIA, scrA; PTS system, sucrose-specific IIA component [EC:2.7.1.69] K02809 PTS-Scr-EIIB, scrA; PTS system, sucrose-specific IIB component [EC:2.7.1.69] K02810 PTS-Scr-EIIC, scrA; PTS system, sucrose-specific IIC component K01193 E3.2.1.26, sacA; beta-fructofuranosidase [EC:3.2.1.26] K00696 E2.4.1.14; sucrose-phosphate synthase [EC:2.4.1.14]

K00695 E2.4.1.13; sucrose synthase [EC:2.4.1.13]
K00689 E2.4.1.5; dextransucrase [EC:2.4.1.5]
K05341 E2.4.1.4; amylosucrase [EC:2.4.1.4]

K12047 MGAM; maltase-glucoamylase [EC:3.2.1.20 3.2.1.3]
K01187 E3.2.1.20, malZ; alpha-glucosidase [EC:3.2.1.20]
K12316 GAA; lysosomal alpha-glucosidase [EC:3.2.1.20]

## **Summary Questions**

What is a pathway and can you name a few important pathway databases?

What is the structure of a KEGG pathway map and can you name any types of interactions?

Can you browse KEGG pathways and search and color pathways on KEGG website?

What is KO?

# 生物信息学:导论与方法 Bioinformatics: Introduction and Methods

Ge Gao 高歌 & Liping Wei 魏丽萍 Center for Bioinformatics, Peking University

