



KEGG: Kyoto Encyclopedia of Genes and Genomes

Minoru Kanehisa

*Bioinformatics Center, Institute for Chemical Research,
Kyoto University*

*Human Genome Center, Institute of Medical Science,
University of Tokyo*

June 2010

Background

- High-throughput experiments and large-scale datasets
- Bioinformatics approaches to integration and interpretation

Gene → Genome → Genomics



Post-genome

Biology



Bioinformatics

Informatics




Molecular Building Blocks of Life Genomic and Chemical Spaces

Genomic Space

| | |
|------------|---------------|
| DNA (Gene) | Genome |
| RNA | Transcriptome |
| Protein | Proteome |

Replication
Transcription
Translation

Central dogma  DNA ↔ RNA → Protein

Genetic code

Chemical Space

| | |
|------------|-----------------|
| Metabolite | Metabolome |
| Glycan | Glycome |
| Lipid | Lipidome |
| Compound | Chemical genome |

(Endogenous)

(Exogenous)

Biosynthesis
Biodegradation
Transport

Biosynthetic code

High-throughput Experimental Projects to Uncover Molecular Building Blocks of Life

Genomic space

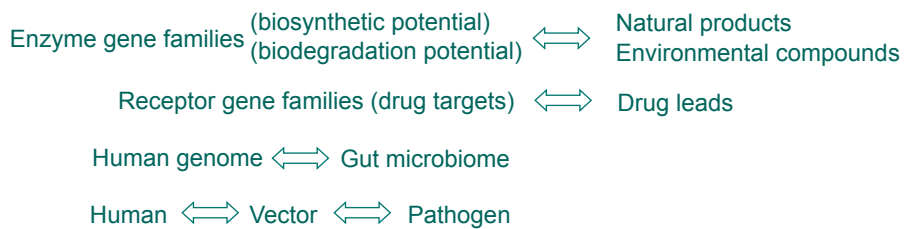
Repertoire of genes and proteins
in individual organisms
(genomics, transcriptomics,
proteomics, etc.)

Repertoire of genes in
environmental samples
(metagenomics)

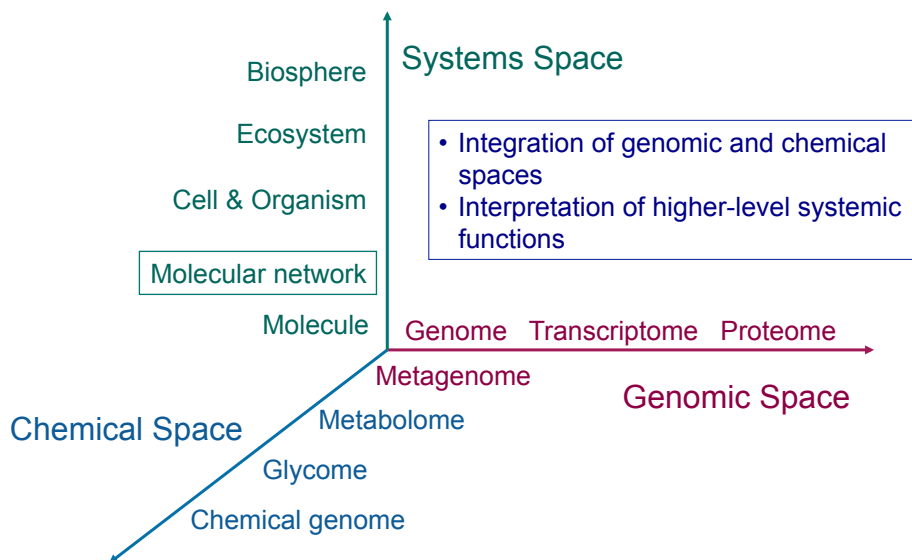
Chemical space

Repertoire of endogenous
molecules
(metabolomics, glycomics,
lipidomics, etc.)

Repertoire of exogenous
molecules
(chemical genomics)

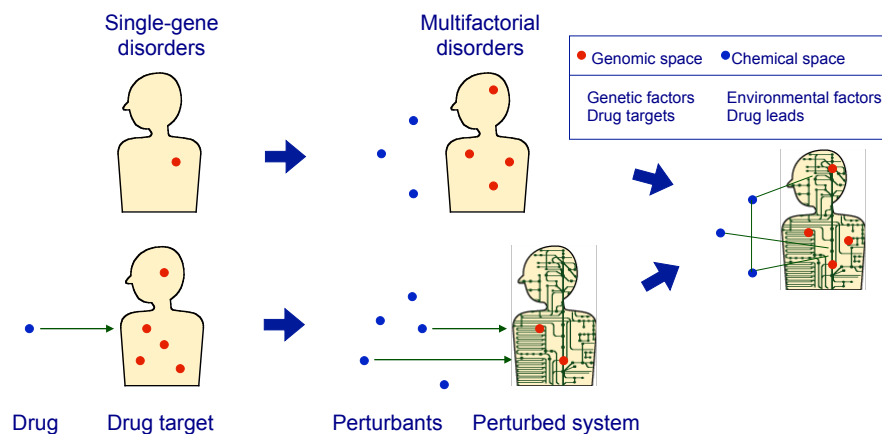


Bioinformatics Approaches to Reconstructing Biological Systems from Building Blocks



Integration of Genomic and Chemical Spaces Medical and Pharmaceutical Implications

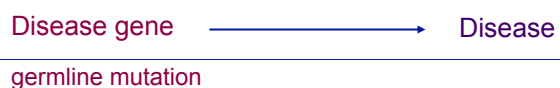
Diseases viewed as perturbed states of molecular systems
Drugs viewed as perturbants to molecular systems



Gene-Disease Associations

Single-gene diseases

Genetic disorders with Mendelian inheritance patterns



Multifactorial (polygenic) diseases

Common diseases such as cancers, heart disease, diabetes, etc.



Cancer causing factors:

somatic mutation, translocation, overexpression, etc.
carcinogen, viral infection, etc.

Network-Disease Associations

Single-gene diseases

Single genetic perturbation \rightarrow Molecular network \rightarrow Disease
germline mutation

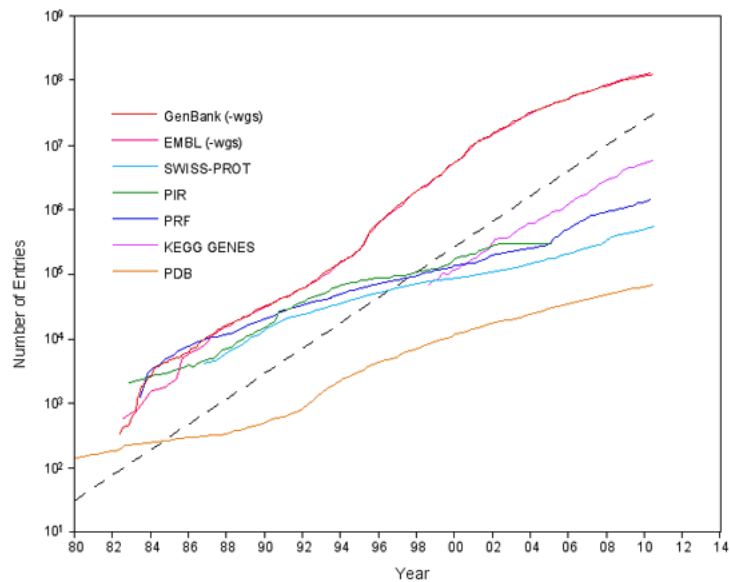
Multifactorial diseases

Multiple genetic perturbations \rightarrow Molecular network \rightarrow Disease
Environmental perturbations \rightarrow Molecular network \rightarrow Disease
somatic/germline mutations, translocation, overexpression, etc.
environmental chemicals, infections, human microbiome, etc.

Infectious diseases

Environmental perturbations \rightarrow Molecular network \rightarrow Disease
pathogens

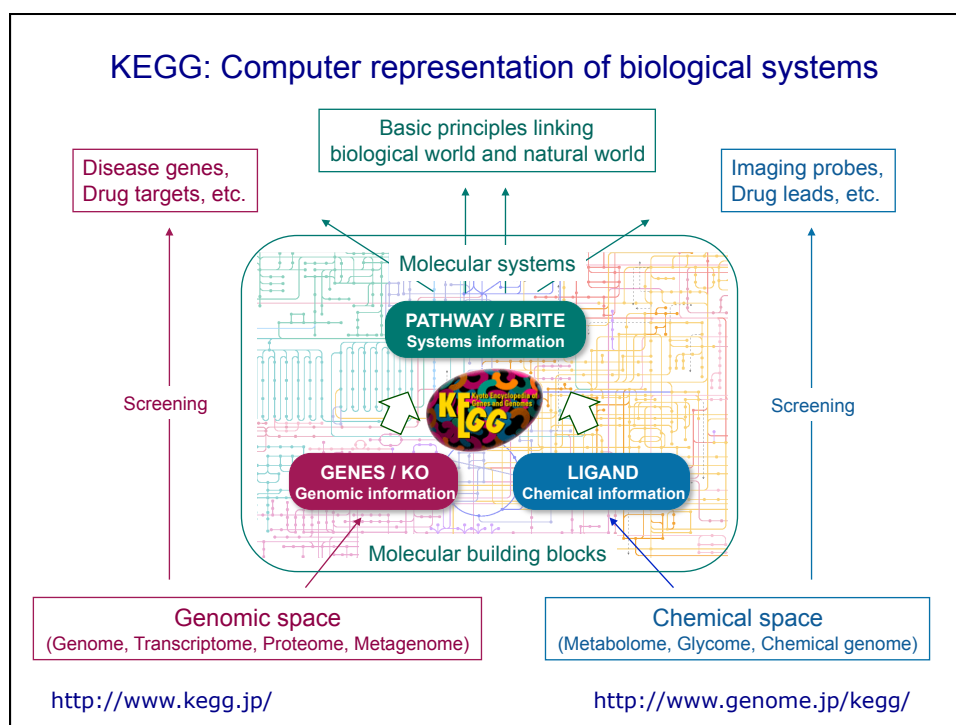
Growth of sequence and 3D structure databases



More details at: http://www.genome.jp/en/db_growth.html

Overview of KEGG

- From building blocks to biological systems
- Integration of genomics and chemistry



KEGG Databases

| Database | Content | Data size |
|----------------|--|----------------|
| KEGG PATHWAY | Pathway maps, reference (total) | 361 (112,000) |
| KEGG BRITE | Functional hierarchies, reference (total) | 89 (29,483) |
| KEGG MODULE | Pathway modules | 537 |
| KEGG DISEASE | Human diseases | 374 |
| KEGG DRUG | Drugs | 9,454 |
| KEGG ORTHOLOGY | KEGG Orthology (KO) groups | 13,597 |
| KEGG GENOME | KEGG Organisms, manual/koala + kaas | 1,255 + 99 |
| KEGG GENES | Genes in high-quality genomes (131 eukaryotes + 1041 bacteria + 83 archaea) | 5,568,602 |
| KEGG SSDB | Best hit relations within GENES | 41,546,146,182 |
| | Bi-directional best hit relations within GENES | 710,431,505 |
| KEGG DGENES | Genes in draft genomes (15 eukaryotes) | 284,078 |
| KEGG EGENES | Genes as EST contigs (84 eukaryotes) | 3,133,980 |
| KEGG COMPOUND | Metabolites and other small molecules | 16,250 |
| KEGG GLYCAN | Glycans | 10,969 |
| KEGG REACTION | Biochemical reactions | 8,173 |
| KEGG RPAIR | Reactant pair chemical transformations | 12,265 |
| KEGG ENZYME | Enzyme nomenclature | 5,184 |

As of June 12, 2010

KEGG Object Identifiers

Prefix + 5-digit number

| Database | Prefix | Example | Release |
|----------------|--------------------|--------------|---------|
| KEGG PATHWAY | map/ko/ec/rn/(org) | hsa04930 | 1995 |
| KEGG BRITE | br/jp/ko/(org) | ko01003 | 2005 |
| KEGG MODULE | M | M00008 | 2007 |
| KEGG DISEASE | H | H00004 | 2008 |
| KEGG DRUG | D | D01441 | 2005 |
| KEGG ORTHOLOGY | K | K04527 | 2002 |
| KEGG GENOME | T | T01000 (hsa) | 2000 |
| KEGG COMPOUND | C | C00031 | 1995 |
| KEGG GLYCAN | G | G00109 | 2003 |
| KEGG REACTION | R | R00259 | 2001 |
| KEGG RPAIR | RP | RP04458 | 2004 |

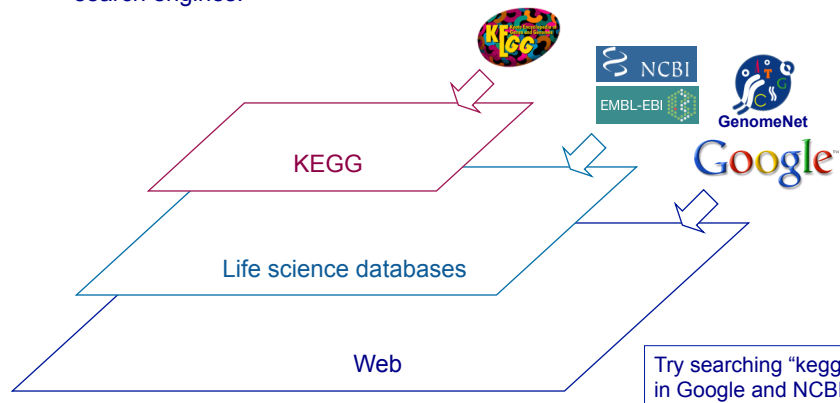
db:name

| | | | |
|-----------------|-----------|-------------|------|
| KEGG GENES | org:gene | hsa:3643 | 1995 |
| KEGG ENZYME | ec:number | ec:2.7.10.1 | 1995 |
| DBGET databases | db:entry | sp:P06213 | 1994 |

org KEGG Organism code
 gene gene identifier (locus_tag, Gene ID, etc.)

KEGG Objects

- KEGG is a computer representation of the biological systems consisting of molecular objects and higher-level objects.
- KEGG objects (database entries) are highly integrated representing biological relationships.
- KEGG objects are linked to/from major life science databases.
- KEGG objects are part of the Web; they can be found by Web search engines.



- KEGG Home
- Introduction
- Overview
- Release notes
- Current statistics
- KEGG Identifiers
- KEGG XML
- KEGG API
- KEGG FTP
- KegTools
- GenomeNet
- DBGET/LinkDB
- Feedback

KEGG: Kyoto Encyclopedia of Genes and Genomes

A grand challenge in the post-genomic era is a complete computer representation of the cell, the organism, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviors from genomic and molecular information. Towards this end we have been developing a bioinformatics resource named KEGG as part of the research projects of the Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo.

- **Main entry point to the KEGG web service**
KEGG2 KEGG Table of Contents Update notes Help
- **Data-oriented entry points**
KEGG PATHWAY Pathway maps and pathway modules Pathway maps
KEGG BRITE Functional hierarchies and ontologies Brite hierarchies
KEGG DISEASE Human diseases Article in NAR DB Issue
KEGG DRUG Drugs
KEGG ORTHOLOGY KO system and ortholog annotation
KEGG GENES Genes and proteins
KEGG GENOME Genomes
KEGG LIGAND Chemical compounds, glycans, and reactions
- **Organism-specific entry points**
KEGG Organisms Select () (example) hsa
- **Other entry points**
KEGG Atlas New interface to navigate pathway maps
KEGG GLYCAN Glycome informatics resource
KEGG COMPOUND Knowledge base for biochemical compounds
KEGG REACTION Knowledge base for biochemical reactions
KEGG PLANT Knowledge base for plant natural products
KAAS KEGG automatic annotation server

Access the KEGG top page:
<http://www.genome.jp/kegg/>
 and enter keywords or any KEGG object identifier in the search box.

Try, for example, **hsa04930** to retrieve the KEGG pathway map for type 2 diabetes.

Web Page Organization

Top page (KEGG)

KEGG Home
Introduction
Overview
Release notes
Current statistics
KEGG Identifiers
KEGG XML
KEGG API
KEGG FTP
KEGTools

GenomeNet
DBGET/LinkDB
Feedback

KEGG: Kyoto Encyclopedia of Genes and Genomes

A grand challenge in the post-genomic representation of the cell, the organism, enable computational prediction of high-level processes and organism behaviors in information. Towards this end we have led research named KEGG as part of the research laboratories in the Bioinformatics Center, Human Genome Center of the University of Kyoto.

Main entry point to the KEGG web service
KEGG2 KEGG Table of Contents

Data-oriented entry points
KEGG PATHWAY Pathway maps and part
KEGG BRITE Functional hierarchies of
KEGG DISEASE Human diseases, drugs
KEGG DRUG Drugs
KEGG ORTHOLOGY KO system and orthology
KEGG GENES Genes and proteins
KEGG GENOME Genomes
KEGG LIGAND Chemical compounds, 1
Systems information

Organism-specific entry points
KEGG Organisms Select (Organism)

Other entry points
KEGG Atlas New interface to navigate
KEGG GLYCAN Glycane information resource
KEGG COMPOUND Knowledge base for bio
KEGG REACTION Knowledge base for drug
KEGG PLANT Knowledge base for plant
KAAS KEGG automatic annotator

Copyright 1996-2010 Kanehisa Laboratories

This icon will bring back to the KEGG top page.

Table of Contents (KEGG2)

KEGG2 PATHWAY BRITE DISEASE DRUG KO GENES GENOME LIGAND DBGET

Search [KEGG2] for [] Go Clear

| Category | Entry Point | Release Info | Search & Compute | DBGET Search |
|----------|--------------|---|---|-----------------------------------|
| | KEGG PATHWAY | New maps Map# changes Updates status | Search objects in pathways Color objects in pathways | PATHWAY |
| | KEGG BRITE | New hierarchies Updates status | Search objects in brite Color objects in brite | BRITE |
| | KEGG DISEASE | Human diseases ATC drug classification | Human diseases ATC drug classification | DISEASE DRUG |
| | KEGG DRUG | KEGG Orthology (KO) | KEGG Orthology (KO) | ORTHOLOGY |
| | KEGG GENES | SSDB search BLAST search FASTA search KAAS automatic annotation KEGG automatic annotation | SSDB search BLAST search FASTA search KAAS automatic annotation KEGG automatic annotation | GENES GENES GENES GENOME |
| | KEGG GENOME | New organisms Update history Update status | KEGG automatic annotation KEGG automatic annotation | GENOME |
| | KEGG LIGAND | KEGG Glycan KEGG Plant | KEGG Glycan KEGG Plant | GLYCAN PLANT |

See Kanehisa et al. (2010) for the new features of KEGG2.

KEGG for specific organisms
KEGG Organisms - the list of currently available organisms
Select (Organism) Go Clear (examples) hsa mmu sco eco bsi syn

KEGG as an integrated web resource

PATHWAY DISEASE
BRITE DRUG ORTHOLOGY GENES LIGAND

KEGG KEGG KEGG KEGG KEGG

These icons will bring back to the KEGG2 page.

DBGET
BLAST/FASTA, KAAS,
SIMCOMP/SUBCOMP,
KCaM, PathPred, etc.

This search box accepts both keywords and KEGG identifiers

This search box accepts keywords only

This icon will bring back to the GenomeNet top page.

GenomeNet

KEGG KEGG2 PATHWAY BRITE DISEASE DRUG DBGET

[English | Japanese]



Search [All databases] for [] Go Clear

GenomeNet
About GenomeNet
Announcements
Release notes
Acknowledgments

DBGET
Overview
DB release info
DB growth curve

KEGG
varDB

Community DBs

Bioinformatics tools
Other tools

Feedback

GenomeNet Database Resources

DBGET: Integrated Database Retrieval System
LinkDB search

KEGG: Kyoto Encyclopedia of Genes and Genomes
KEGG2 - Table of contents
KEGG PATHWAY - Systems information: pathways
KEGG BRITE - Systems information: ontologies
KEGG DISEASE - Disease information resource
KEGG DRUG - Drug information resource
KEGG Organisms - Organism-specific entry points
KEGG GENES - Genomic information
KEGG LIGAND - Chemical information
KEGG GLYCAN - Glycan information resource
KEGG PLANT - Plant information resource

varDB: Antigenic variation database

Community Databases
CYORF - Cyanobacteria annotation database
BSORF - Bacillus subtilis genome database
EXPRESSION - Gene expression profile database

GenomeNet Bioinformatics Tools

Sequence Analysis
BLAST / FASTA - Sequence similarity search
MOTIF - Sequence motif search
CLUSTALW / MAFFT / PRN - Multiple alignment

Genome Analysis
KAAS - KEGG automatic annotation server
EGassembler - EST consensus contigs
GENIES - Gene network prediction
GECS - Gene expression to chemical structure

Chemical Analysis
SIMCOMP / SUBCOMP - Chemical structure search
KCaM - Glycan structure search
PathPred - Reaction pathway prediction
E-zyme - Enzymatic reaction prediction

KEGG release info
KEGG update notes
KEGG pathway maps
BRITE functional hierarchies
KEGG Organisms
[Animals | Plants]



GenomeNet
<http://www.genome.jp/>

GenomeNet Bioinformatics Tools

Sequence Analysis
BLAST / FASTA - Sequence similarity search
MOTIF - Sequence motif search
CLUSTALW / MAFFT / PRN - Multiple alignment

Genome Analysis
KAAS - KEGG automatic annotation server
EGassembler - EST consensus contigs
GENIES - Gene network prediction
GECS - Gene expression to chemical structure

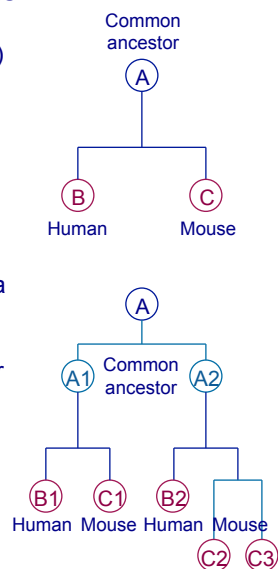
Chemical Analysis
SIMCOMP / SUBCOMP - Chemical structure search
KCaM - Glycan structure search
PathPred - Reaction pathway prediction
E-zyme - Enzymatic reaction prediction

KEGG GENES and ortholog annotation

Orthologs and Paralogs

- Sequence similarity between two genes (or proteins) may imply ortholog or paralog relationship.
- Orthologs are genes in different species evolved from a common ancestral gene by speciation and tend to have the same function.
- Paralogs are generated by gene duplication within a species and often represent diversified functions in a broader functional category.
- Identification of ortholog relationships is the basis for genome annotation (assigning gene functions), and it requires distinction from paralog relationships.

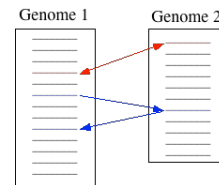
Orthologs: B-C, B1-C1
 Co-orthologs: B2-(C2,C3)
 Inparalogs: C2-C3
 Outparalogs: B1-B2, B1-(C2,C3), B2-C1



Computational Identification of Orthologs

Between two species

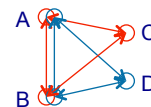
Bi-directional best hit (BBH)
(Reciprocal best hit)



Among multiple species

1. COG

Triangle of BBH relationships among three species



2. KEGG OC

p-Quasi clique among multiple species

Superposition of ABC and ABD



Clique
(completely connected subgraph)

p-Quasi clique is an almost complete subgraph, where the degree of completeness is represented by p.

Genome annotation in KEGG: KO (K number) assignment

KEGG GENES

- Gene information for completely sequenced genomes
- Computationally generated from RefSeq and other public resources
- Partial annotation by automated KOALA
- Manual annotation with KOALA and GFIT tools

KEGG GENES in Oracle

| | |
|----------------|-----------|
| Genomes | 1,255 |
| Genes | 5,631,338 |
| Genes with KO | 1,992,279 |
| KO assignment | 35% |
| Genome/UniProt | 0.2% |
| Genes/UniProt | 50% |

As of June 12, 2010

KEGG ORTHOLOGY (KO)

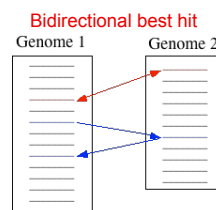
- Manually defined ortholog groups that correspond to KEGG pathway nodes and BRITE hierarchy nodes
- Identified by K numbers

KEGG SSDB

- Sequence similarity scores and best hit relations
- Computationally generated from GENES by pairwise genome comparisons using SSEARCH

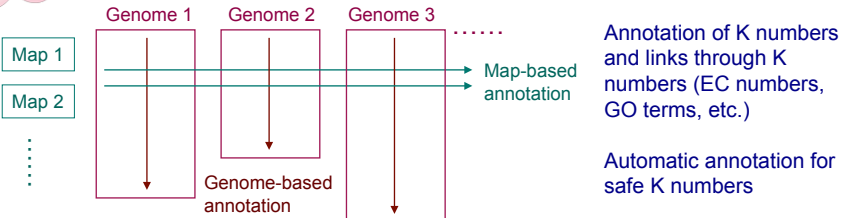
KEGG OC

- Ortholog clusters
- Computationally generated from SSDB by a quasi-clique finding algorithm





KOALA: KEGG Orthology And Links Annotation



| GRP | ORG | KEGG ID | KO (KOALA) | KO (GENES) | Orth (GENES) | OC | Score | Memo | Ann |
|-------|-----|------------------------|------------|------------|--------------|---------------------------|-------|------|-----|
| E Ani | hsa | > A hsa:5664(448) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 18611 | | A |
| E Ani | ptr | > A ptr:457790(442) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 16345 | | A |
| E Ani | mcc | > A mcc:698770(654) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 10464 | | A |
| E Ani | mmu | > A mmu:19165(448) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 17031 | | A |
| E Ani | rno | > A rno:81751(448) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 16204 | | A |
| E Ani | cfa | > A cfa:490382(717) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 7851 | | A |
| E Ani | bta | > A bta:282010(449) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 16358 | | A |
| E Ani | ssc | > A ssc:780410(448) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 18558 | | A |
| E Ani | mdo | > A mdo:100026453(426) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 15132 | | A |
| E Ani | oaa | > A oaa:100076299(455) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 14136 | | A |
| E Ani | gga | > A gga:374188(451) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 10574 | | A |
| E Ani | xla | > A xla:397713(449) | K04522 | K04522 | PSEN2 | 62906/60206/Animals.27398 | 11327 | | A |

| KEGG ID | KO | Orth | Definition |
|---------------|--------|-------|------------------------------|
| hsa:5663 | K04505 | PSEN1 | presenilin 1 (EC:3.4.23.-) |
| hsa:5664 | K04522 | PSEN2 | presenilin 2 (Alzheimer dis) |
| ptr:457790 | K04522 | PSEN2 | presenilin 2 (Alzheimer dis) |
| mcc:698770 | K04522 | PSEN2 | presenilin 2 |
| mmu:19164 | K04505 | PSEN1 | presenilin 1 (EC:3.4.23.-) |
| mmu:19165 | K04522 | PSEN2 | presenilin 2 |
| rno:81751 | K04505 | PSEN1 | presenilin 1 (EC:3.4.23.-) |
| cfa:403408 | K04505 | PSEN1 | presenilin 1 (EC:3.4.23.-) |
| cfa:490382 | K04522 | PSEN2 | presenilin 2 (Alzheimer dis) |
| bta:282010 | K04522 | PSEN2 | presenilin 2 (Alzheimer dis) |
| bta:282705 | K04505 | PSEN1 | presenilin 1 (EC:3.4.23.-) |
| ssc:780410 | K04522 | PSEN2 | presenilin 2 |
| ssc:780411 | K04505 | PSEN1 | presenilin 1 (EC:3.4.23.-) |
| mdo:100019628 | K04505 | PSEN1 | similar to presenilin 1 |
| mdo:100026453 | K04522 | PSEN2 | similar to presenilin 2 |
| gga:100076299 | | | similar to presenilin 2 (Alz |
| gga:373977 | K04505 | PSEN1 | presenilin 1 (Alzheimer dis) |
| gga:374188 | K04522 | PSEN2 | presenilin 2 (Alzheimer dis) |
| xla:397713 | K04522 | PSEN2 | presenilin-beta |
| xla:399258 | K04505 | PSEN1 | presenilin-alpha |

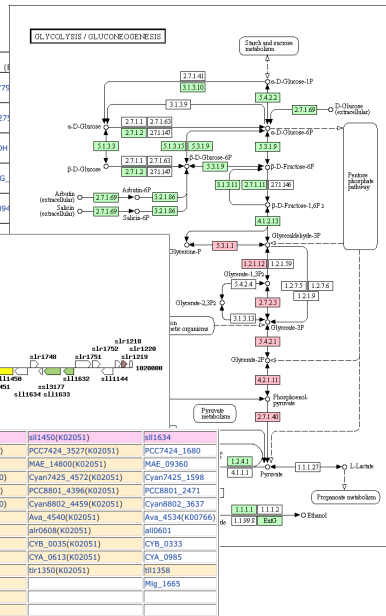
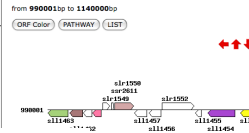
GFIT link KOALA's Current OC links suggestion assignment

Tools to check annotation quality Based on reconstruction of functional units

Ortholog table

| Organism | K01803 (TPS) [1170] | K00134 (GAPDH) [1469] | K00150 (gap2) [127] | K00927 (PGK) [1108] | K01834 (PGAM) [2711] | K01837 (BPGM) [1] |
|----------|---------------------|-----------------------|---------------------|---------------------|----------------------|-------------------|
| ECO | b3919 | b1779 | b2926 | b0759 | b3612 | b2778 |
| ECJ | JW3890 | JW1413 | JW2893 | JW0738 | JW3587 | JW2757 |
| ECG | ECDH108_4108 | ECDH108_1917 | ECDH108_3101 | ECDH108_0822 | ECDH108_3794 | ECDH108_4533 |
| ECW | BWG_3588 | BWG_1592 | BWG_2649 | BWG_3303 | BWG_0657 | BWG_4087 |
| ECO | Z5464 | Z2304 | Z4265 | Z0925 | Z9039 | Z4034 |
| ECG | EC54844 | EC52022 | EC52488 | EC54490 | | |
| ECF | ECH74115_5374 | ECH74115_2503 | ECH74115_2022 | | | |

KEGG Genome map - *Synechocystis* sp. PCC6803



Conserved gene cluster

| byn | bjl1485 | bjl1454(K02049) | bjl1451(K02050) | bjl1450(K02051) | bjl1434 |
|-----|------------------|------------------|-----------------------|-----------------------|-----------------------|
| byn | bjl1485 | bjl1454(K02049) | bjl1451(K02050) | bjl1450(K02051) | bjl1434 |
| ccr | PCC7424_3464 | PCC7424_3462 | PCC7424_3525(K02049) | PCC7424_3526(K02050) | PCC7424_3527(K02051) |
| mar | MAE_55270 | MAE_53960 | MAE_14780(K02049) | MAE_14790(K02050) | MAE_14800(K02051) |
| cyh | Cyan7425_3581 | Cyan7425_4566 | Cyan7425_4570(K02049) | Cyan7425_4571(K02050) | Cyan7425_4572(K02051) |
| cyp | PCC8801_4057 | PCC8801_2463 | PCC8801_4394(K02049) | PCC8801_4395(K02050) | PCC8801_4396(K02051) |
| cyh | Cyan8802_4096 | Cyan8802_3646 | Cyan8802_4457(K02049) | Cyan8802_4458(K02050) | Cyan8802_4459(K02051) |
| ava | Ava_4546 | Ava_4544(K02049) | Ava_4541(K02050) | Ava_4540(K02051) | Ava_4534(K02052) |
| ana | ana0614 | ana0612(K02049) | ana0610(K02049) | ana0608(K02050) | ana0606(K02051) |
| cyb | CYB_0042 | CYB_0040(K02049) | CYB_0037(K02049) | CYB_0036(K02050) | CYB_0035(K02051) |
| cya | CYA_0174 | CYA_0619(K02049) | CYA_0615(K02049) | CYA_0614(K02050) | CYA_0613(K02051) |
| tel | tel1356 | tel1355(K02049) | tel1352(K02049) | tel1351(K02050) | tel1350(K02051) |
| beh | Mig_1702(K02049) | Mig_2587(K02049) | Mig_1703(K02049) | Mig_1704(K02050) | Mig_1665 |
| eps | EPS_4945(K02049) | EPS_3314(K02049) | EPS_3315(K02050) | EPS_3316(K02051) | |
| ter | Ter_1159(K02049) | Ter_1158(K02049) | Ter_1156(K02049) | Ter_1154(K02050) | |

Public version of KEGG annotation tools

Orthology: K04522

Entry: K04522 KO

Name: PSEN2, PS2

Definition: presenilin 2 [EC:3.4.23.-]

Pathway: ko04330 Notch signaling pathway
ko05010 Alzheimer's disease

Disease: H00056 Alzheimer's disease (AD)

Class: Metabolism; Enzyme Families; Peptidases [BR:h001002]
Environmental Information Processing; Signal Transduction; Notch signaling pathway [PATH:ko04330]
Human Diseases; Neurodegenerative Diseases; Alzheimer's disease [PATH:ko05010]

Genes: HSA: 5664(PSEN2)
PTR: 457790(PSEN2)
MCC: 698770(PSEN2)
MMU: 19165(Psen2)
RNO: 81751(Psen2)
CPA: 490382(PSEN2)
BTA: 283010(PSEN2)
SSC: 780410(PSEN2)
ECB: 100054506
MDO: 100026453
OAA: 100076299
GGA: 374188(PSEN2)
TGU: 100232599
XLA: 397113(psen2)
XTR: 549935(psen2)
DRE: 50016(psen2)

Reference: PMID:10497236

Authors: Steiner H, Duff K, Capell A, Rom Yu X, Picciano M, Fechteler K, C Bødder M, Tgaita I, Iwatsubo T, A loss of function mutation of p beta-peptide production and not J Biol Chem 274:28669-73 (1999)

Position: Iq31-q42

AA seq: 448 aa

Motif: Pfm: Presenilin HGRP

Other DBs: NCBI-GI: 156105679
NCBI-geneid: 5664
OMIM: 600759
HGNC: 9509
HPD: 02460
Ensembl: ENSG00000143801
UniProt: P49810

BLAST Search: Enter query sequence: >hso:5664 PSEN2; presenilin 2 (Alzheimer disease 4); K04522 presenilin 2 [EC:3.4.23.-] (A)

SSDB query for orthologs and paralogs

Conserved gene cluster search (useful for bacterial genomes)

Read-only version of GFIT

Ortholog table via pathway entry

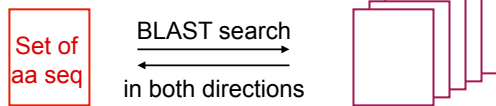
DB search against UniProt, etc. may be done on the fly

Read-only version of KOALA

KAAS: KEGG Automatic Annotation Server

<http://www.genome.jp/tools/kaas/>

Pairwise genome comparison



Query genome

Annotated genomes in KEGG

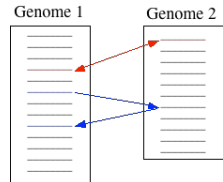
BLAST result screening by bi-directional best hit rate (BHR)

$$BHR = R_f \times R_r > 0.95$$

K number assignment by a heuristic scoring

$$S_{KO} = S_h - \log_2(mn) - \log_2 \left(\sum_{k=N}^x C_k p^k (1-p)^{x-k} \right)$$

Bidirectional best hit



Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M.; KAAS: an automatic genome annotation and pathway reconstruction server. *Nucl. Acids Res.* 35, W182-W185 (2007).



KEGG PATHWAY and BRITE: Reference knowledge base

Data objects for computer representation of molecular systems

Element

gene, protein, small molecule, etc.

Pair (binary relation)

protein-protein interaction, drug-target relationship, etc.

Graph (wiring diagram)

pathway, complex, etc.

Simple list (membership)

pathway, complex, etc.

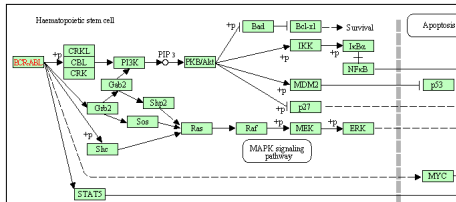
Hierarchical list

hierarchical classification, ontology, etc.

Knowledge Representation of Systemic Functions

Molecular network (pathway map)

KEGG PATHWAY



Simple list (membership)

KEGG DISEASE

| Gene | BCR-ABL (translocation) [HSA:613 25] EVI1 (overexpression) [HSA:2122] AML1 (translocation) [HSA:861] p16/INK4 (mutation) [HSA:1029] p53 (mutation) [HSA:7157] RB1 (mutation) [HSA:5925] |
|------------|--|
| Carcinogen | 1,3-Butadiene [CPD:C16450] Rubber industry |
| Marker | BCR-ABL (translocation) [HSA:613 25] WT1 [HSA:7490] |
| Drug | Imatinib mesylate (Gleevec) [DR:D01441] Hydroxyurea [DR:D00341] Interferon-alpha [DR:D00745 D02745 D03305 D04552 D04553] |

Hierarchical list (ontology)

KEGG BRITE

```

▼ ▼ ▼
▼ Cancers
  ► Cancers of the nervous system
  ► Cancers of the digestive system
  ▼ Cancers of haematopoietic and lymphoid tissues
    H00003 Acute myeloid leukemia (AML) [PATH:h0a05221]
    H00001 Acute lymphoblastic leukemia (ALL) (Precursor B
    H00002 Acute lymphoblastic leukemia (ALL) (Precursor T
    H00004 chronic myeloid leukemia (CML) [PATH:h0a05220]
    H00005 Chronic lymphocytic leukemia (CLL)
    H00007 Hodgkin lymphoma
    H00006 Hairy-cell leukemia
    H00008 Burkitt lymphoma
    H00009 Adult T-cell leukemia
    H00010 Multiple myeloma
    H00011 Lymphoplasmacytic lymphoma
    H00012 Polycythemia vera
  ► Cancers of the breast and female genital organs
  ► Cancers of soft tissues and bone
  ► Skin cancers
  ► Cancers of the urinary system and male genital organs
  ► Cancers of endocrine organs
  ► Head and neck cancers
  ► Cancers of the lung and pleura

► Immune System Diseases
► Nervous System Diseases
► Circulatory System Diseases
► Metabolic Diseases
    
```

Data source: review articles, other publications, specialists' websites, etc.



KEGG PATHWAY Database

Collection of KEGG pathway maps

Global Map

Metabolism Map (2)

Metabolism

- Carbohydrate Metabolism (15)
- Energy Metabolism (8)
- Lipid Metabolism (16)
- Nucleotide Metabolism (2)
- Amino Acid Metabolism (13)
- Metabolism of Other Amino Acids (9)
- Glycan Biosynthesis and Metabolism (15)
- Metabolism of Cofactors and Vitamins (12)
- Metabolism of Terpenoids and Polyketides (20)
- Biosynthesis of Secondary Metabolites (21)
- Xenobiotics Biodegradation and Metabolism (25)
- Overview (9)

Genetic Information Processing

- Transcription (3)
- Translation (2)
- Folding, Sorting and Degradation (5)
- Replication and Repair (6)

Environmental Information Processing

- Membrane Transport (3)
- Signal Transduction (14)
- Signaling Molecules and Interaction (4)

<http://www.genome.jp/kegg/pathway.html>

Cellular Processes

- Transport and Catabolism (4)
- Cell Motility (3)
- Cell Growth and Death (7)
- Cell Communication (4)

Organismal Systems

- Immune System (15)
- Endocrine System (7)
- Circulatory System (2)
- Excretory System (4)
- Nervous System (3)
- Sensory System (3)
- Development (2)
- Environmental Adaptation (4)

Human Diseases

- Cancers (15)
- Immune Disorders (6)
- Neurodegenerative Diseases (5)
- Cardiovascular Disease (4)
- Metabolic Disorders (3)
- Infectious Diseases (7)

Drug Development

- Chronology: antibiotics (8)
- Chronology: antineoplastics (5)
- Chronology: nervous system agents (9)
- Chronology: other drugs (9)
- Target based classification (15)
- Structure based classification (5)
- Skeleton based classification (8)

As of June 12, 2010



KEGG BRITE Database Collection of BRITE functional hierarchies

Pathways and Ontologies
Pathways (1)
Functional hierarchies (1)

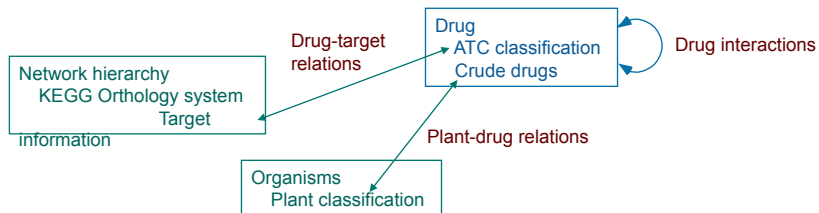
Genes and Proteins
Network hierarchy (1)
Protein families: metabolism (9)
Protein families: genetic information processing (10)
Protein families: environmental information processing (6)
Protein families: cellular processes (7)

Compounds and Reactions
Compounds (5)
Reactions (3)
Compound interactions (6)

Drugs and Diseases
Drugs (10)
Diseases (2)

Cells and Organisms
Organisms (4)

and associated binary relations



<http://www.genome.jp/kegg/brite.html>

As of June 12, 2010

Drug-target relationship in KEGG PATHWAY and KEGG BRITE

TYPE II DIABETES MELLITUS

ATC Classification

- A ALIMENTARY TRACT AND METABOLISM
 - A01 STOMATOLOGICAL PREPARATIONS
 - A02 DRUGS FOR ACID RELATED DISORDERS
 - A03 DRUGS FOR FUNCTIONAL GASTROINTESTINAL DISORDERS
 - A04 ANTIEMETICS AND ANTINAUSEANTS
 - A05 BILE AND LIVER THERAPY
 - A06 LAXATIVES
 - A07 ANTIDIARRHEALS, INTESTINAL ANTIINFLAMMATORY/ANTIINFECTIVE AGENTS
 - A08 ANTIOSBESITY PREPARATIONS, EXCL. DIET PRODUCTS
 - A09 DIGESTIVES, INCL. ENZYMES
 - A10 DRUGS USED IN DIABETES
 - A10A INSULINS AND ANALOGUES
 - A10B BLOOD GLUCOSE LOWERING DRUGS, EXCL. INSULINS
 - A10BA Biguanides
 - A10BB Sulfonamides, urea derivatives
 - A10BB01 Glibenclamide
 - D00336 Glibenclamide (JP15/INN); Glyburide (USP) **Target**
 - A10BB02 Chlorpropamide
 - D00271 Chlorpropamide (JP15/USP/INN)
 - A10BB03 Tolbutamide
 - D00180 Tolbutamide (JP15/USP/INN)
 - D06177 Tolbutamide sodium, sterile

Search BRITE hierarchies

KEGG Orthology (KO) [BR:hsa00001]
Environmental Information Processing
Membrane Transport
02000 Transporters [BR:hsa02000]
6833 ABCc8, ABC16, HNF1, HI, HRINS, MRP8, PHLI, SUR, SUR1, 02010 ABC transporters [PATH:hsa02010]
6833 ABCc8, ABC16, HNF1, HI, HRINS, MRP8, PHLI, SUR, SUR1, Signaling Molecules and Interaction
04000 Receptors and channels [BR:hsa04000]
6833 ABCc8, ABC16, HNF1, HI, HRINS, MRP8, PHLI, SUR, SUR1, Human Diseases
Metabolic Diseases
04930 Type II diabetes mellitus [PATH:hsa04930]
6833 ABCc8, ABC16, HNF1, HI, HRINS, MRP8, PHLI, SUR, SUR1, Transporters [BR:hsa02000]
ABC Transporters, Eukaryotic Type
ABCC (CFTR/MRP) subfamily
ABCC8, 9 subgroups
6833 ABCc8, ABC16, HNF1, HI, HRINS, MRP8, PHLI, SUR, SUR1, Receptors and channels [BR:hsa04000]
Ion Channels
Related to ATPase-linked transporters
High-affinity sulfonyleurea receptor (SUR)
6833 ABCc8, ABC16, HNF1, HI, HRINS, MRP8, PHLI, SUR, SUR1,

Target gene classification

sulfonyleurea receptor 1 agonist [HSA:6833]
sulfonyleurea receptor 1 agonist [HSA:6833]
sulfonyleurea receptor 1 agonist [HSA:6833]
sulfonyleurea receptor 1 agonist [HSA:6833]

Drug classification

Try brite join operations in the KEGG BRITE page.

Target information in the context of pathway

Pathway mapping and BRITE mapping

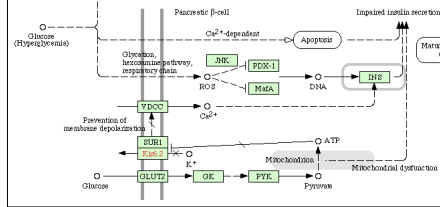
Linking genomes to biological systems and the environment

KO assignment

| Gene | KO |
|-------|--------|
| 10000 | K04456 |
| 10003 | K01301 |
| 10004 | K01301 |
| 10005 | K01068 |
| 10007 | K02564 |
| 10008 | K04897 |
| 10020 | K00885 |
| 10020 | K01791 |
| 10021 | K04957 |
| 10023 | K03069 |
| 10026 | K05290 |
| 10038 | K00774 |
| 10039 | K00774 |
| 10050 | K08193 |
| 10051 | K06675 |
| 10052 | K07616 |
| 10056 | K01890 |
| 10057 | K05668 |
| 10058 | K05661 |
| 10059 | K01528 |
| | |

Genes to K numbers

KEGG pathway mapping



Try examples of pathway mapping and brite mapping from KEGG2 page.

BRITE hierarchy mapping

- ▶ ABC Transporters, Eukaryotic Type
 - ▶ ABCA (ABCI) subfamily (proteins)
 - ▶ ABCB (MDR/TAP) subfamily
 - ▶ ABCC (CFTR/MRP) subfamily
 - ▶ ABCC1, 2, 3, 5, 6, 13 subgroups
 - ▶ ABCC8, 9 subgroups
 - R05032 ABCC9; ATP-binding cassette, subfamily C (CFTR)
 - R05033 ABCC9; ATP-binding cassette, subfamily C (CFTR)
 - R05071 ABCC11; ATP-binding cassette, subfamily C (CFTR)
 - R05072 ABCC11; ATP-binding cassette, subfamily C (CFTR)
 - ▶ ABCC-ABC subgroups
 - ▶ Others
 - ▶ ABCD (ALD) subfamily
 - ▶ ABCG (White) subfamily
- ▶ Other subfamilies
- ▶ ABC Transporters, Prokaryotic Type
- ▶ Major Facilitator Superfamily (MFS)

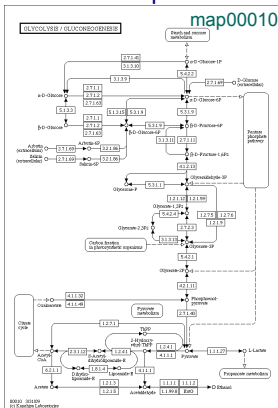
K numbers to C/D/G/R numbers

- ▶ 1 Agents affecting nervous system and sensory organs
- ▶ 2 Agents affecting individual organs (drugs)
- ▶ 3 Agents affecting metabolism
 - ▶ 31 Vitamins
 - ▶ 32 Nutrients, tonics
 - ▶ 33 Blood and body fluid agents
 - ▶ 34 Dialysis agents
 - ▶ 39 Other agents affecting metabolism
 - ▶ 391 Liver disease agents
 - ▶ 392 Antidotes
 - ▶ 393 Habitual intoxication agents
 - ▶ 394 Gout preparations
 - ▶ 395 Enzyme preparations
 - ▶ 396 Antidiabetic agents
 - ▶ 3961 Sulfonylureas
 - D00219 Acetohexamide (CF14/USP); Dymelor (TM)
 - D00071 Chlorpropamide (CF14/USP); Diabinese (TM)
 - D00338 Glipizide (USP); Glucotrol (TM)
 - D00336 Glibenclamide (CF14); Glimeclor (TM)
 - D00379 Tolazamide (CF14/USP); Tolinase (TM)
 - D00580 Tolbutamide (CF14/USP); Orinase (TM)
 - D00593 Glimepiride (JAN/USP); Amaryl (TM)
 - D01599 Gliclazide (JAN); Glimicron (TM)
 - D01799 Glicypirramide (JAN); Diamulin-S (TM)
 - ▶ 3962 Biguanides
 - ▶ 3969 Others

Convention of map number prefix

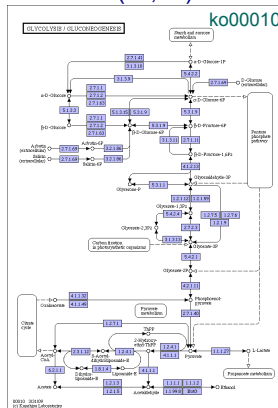
Reference pathway

map



Each box is manually associated with KO identifier (K number), EC number, and reaction identifier (R number)

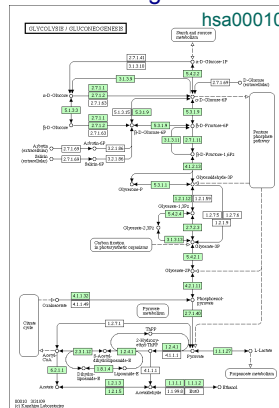
ko (ec, m)



Blue boxes represent selection of K numbers (prefix ko), as well as EC numbers (prefix ec) and R numbers (prefix m)

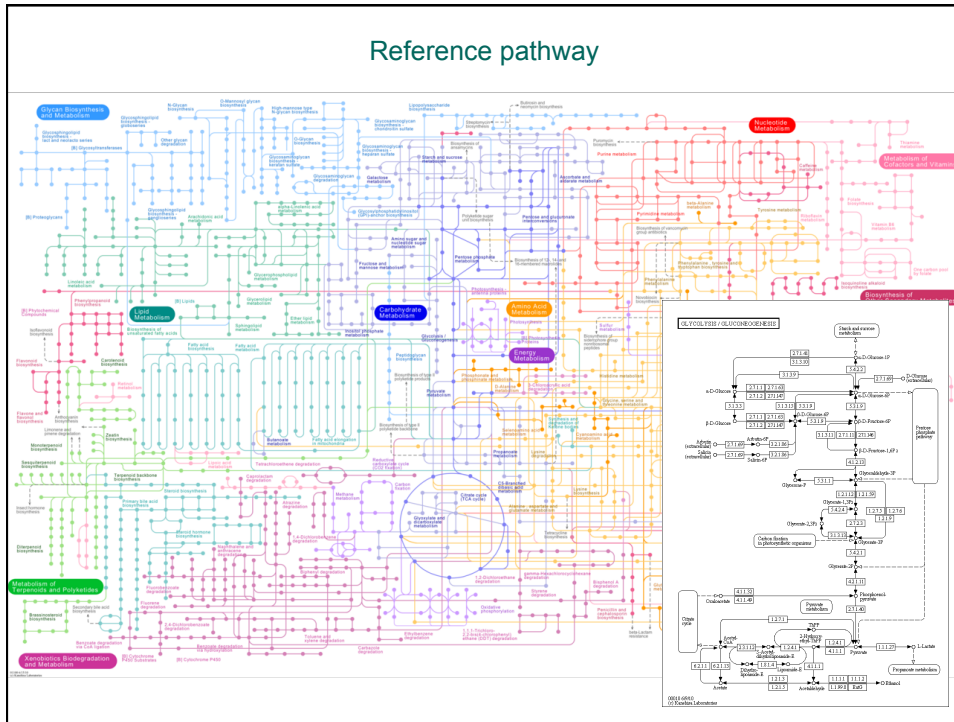
Organism-specific pathway

org

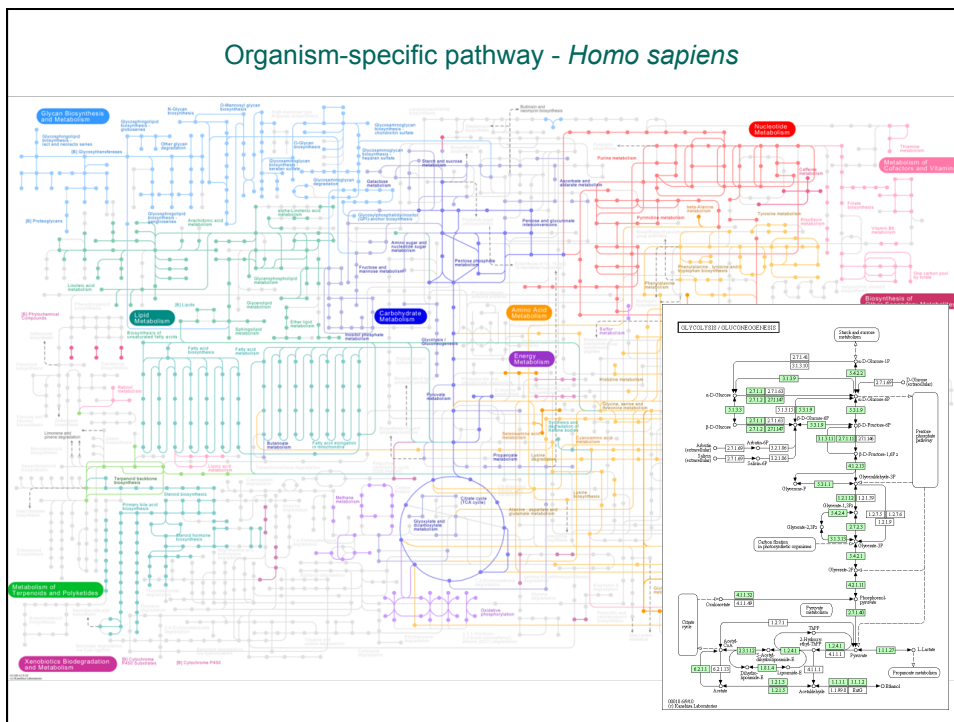


Green boxes correspond to gene identifiers in an organism that are computationally converted from K numbers

Reference pathway



Organism-specific pathway - *Homo sapiens*





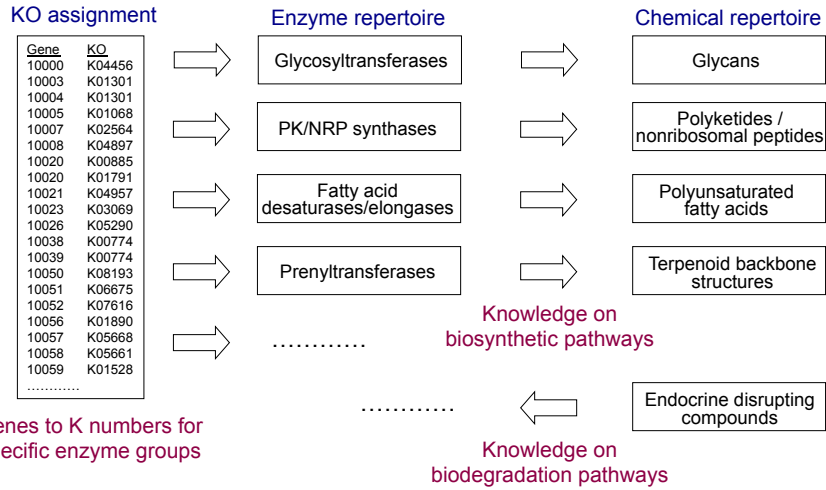
KEGG LIGAND and chemical bioinformatics

Bioinformatics for Small Molecules

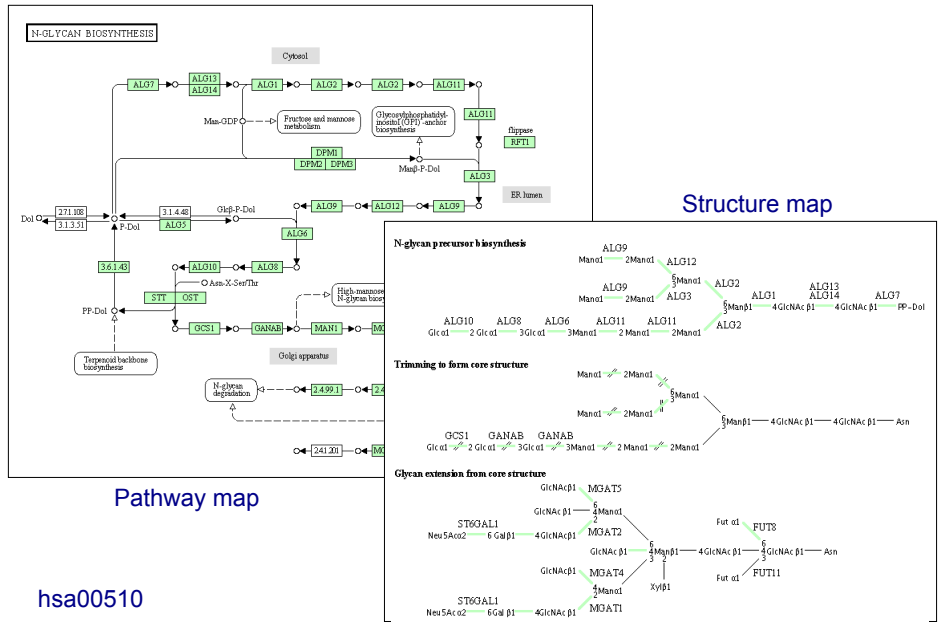
1. Chemical structure similarity
 - Comparison of bit-represented vectors (fingerprints)
 - Comparison of graph objects
2. Chemical building blocks
 - Conserved substructures as building blocks of compounds/drugs
 - Variable substructures as building blocks of reactivity/efficacy
3. Network modules
 - Genomic module, e.g. operon
 - Chemical module, e.g. overall reaction
4. Predictive methods
 - Interaction prediction, e.g. toxicity
 - Reaction prediction, e.g. metabolic fate
5. Examples
 - Plant/fungi/bacterial genomes and secondary metabolites via biosynthetic pathways
 - Bacterial genomes and environmental compounds via biodegradation pathways

Prediction of biosynthetic/biodegradation potentials

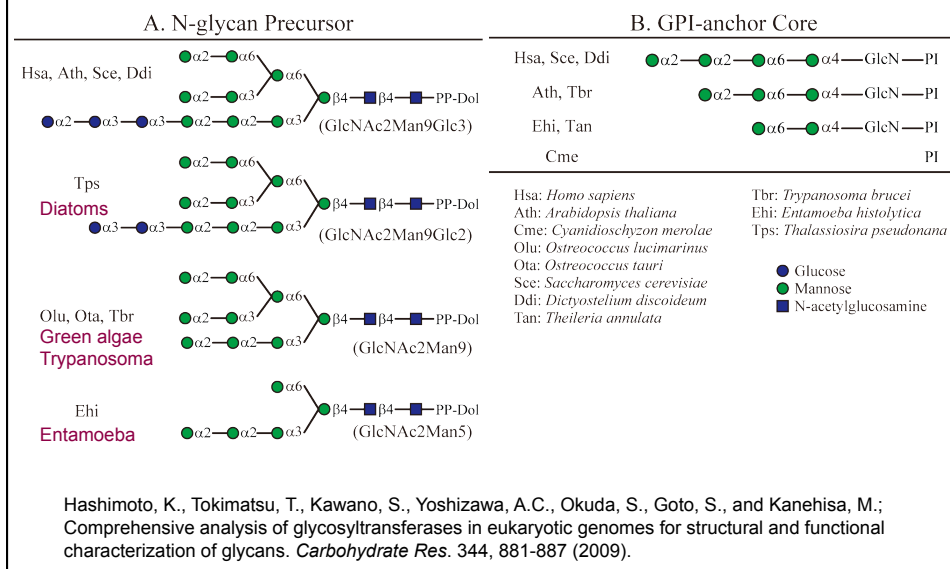
Linking genomes to endogenous/exogenous molecules



N-Glycan biosynthesis - Homo sapiens (human)



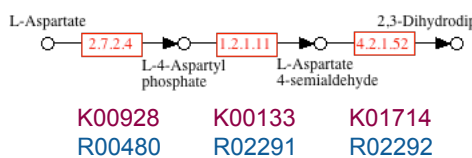
Truncated glycan structures in parasitic protists and algae predicted from genomic information



Dual Aspect of Metabolic Network

1. Genomic information network

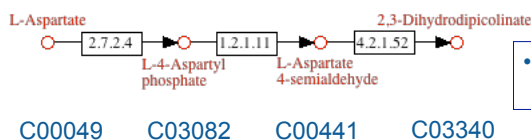
Network of enzyme genes (or enzymes)



- EC numbers are displayed as node names, but they are not used as identifiers in KEGG
- Genes and proteins are identified by K numbers
- Reactions are identified by R numbers

2. Chemical information network

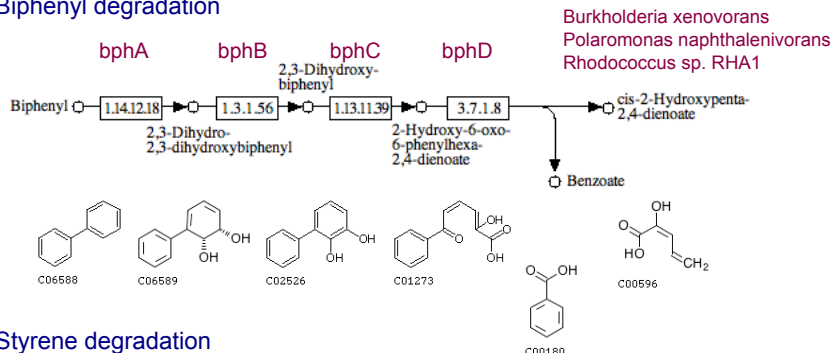
Network of small molecules (or chemical structure transformations)



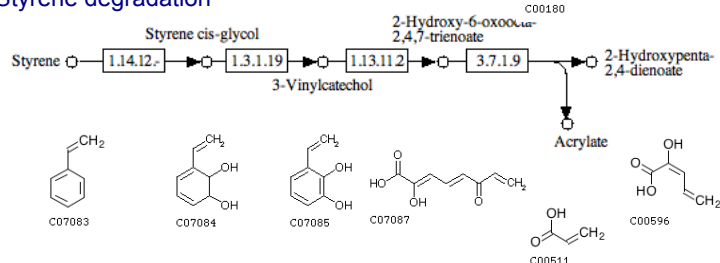
- Compounds are identified by C numbers

Chemical structure transformation network

Biphenyl degradation

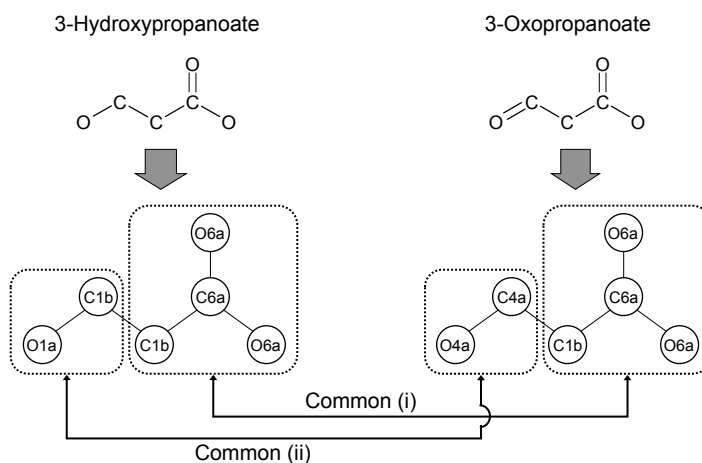


Styrene degradation



Chemical structure comparison based on atom typing

SIMCOMP @ <http://www.genome.jp/tools/simcomp/>



Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M.: Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.* 125, 11853-11865 (2003).

Hattori, M., Tanaka, N., Kanehisa, M., and Goto, S.: SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. *Nucleic Acids Res.* Epub 2010 May 11 (2010).

KEGG atom types

Carbon 23 types

| | | |
|------------------|-----|-----------------------------------|
| Alkane | C1a | R-CH3 |
| | C1b | R-CH2-R |
| | C1c | R-CH(R)-R |
| | C1d | R-C(R)2-R |
| Cyclic alkane | C1x | ring-CH2-ring |
| | C1y | ring-CH(R)-ring |
| | C1z | ring-CH(R)2-ring |
| Alkene | C2a | R=CH2 |
| | C2b | R=CH-R |
| | C2c | R=C(R)2 |
| Cyclic alkene | C2x | ring-CH=ring |
| | C2y | ring-C(R)=ring ring-C(=R)-ring |
| Alkyne | C3a | R≡CH |
| | C3b | R≡C-R |
| Aldehyde | C4a | R-CH=O |
| Ketone | C5a | R-C(=O)-R |
| Cyclic ketone | C5x | ring-C(=O)-ring |
| Carboxylic acid | C6a | R-C(=O)-OH |
| Carboxylic ester | C7a | R-C(=O)-O-R |
| Aromatic ring | C8x | ring-CH=ring |
| | C8y | ring-C(R)=ring |
| Undefined C | C0 | |

Nitrogen 16 types

| | | |
|---------------|-----|----------------|
| Amine | N1a | R-NH2 |
| | N1b | R-NH-R |
| | N1c | R-N(R)2 |
| | N1d | R-N(R)3+ |
| Cyclic amine | N1x | ring-NH-ring |
| | N1y | ring-N(R)-ring |
| Imine | N2a | R=N-H |
| | N2b | R=N-R |
| Cyclic imine | N2x | ring-N=ring |
| | N2y | ring-N(R)+ring |
| Cyan | N3a | R≡N |
| Aromatic ring | N4x | ring-NH-ring |
| | N4y | ring-N(R)-ring |
| | N5x | ring-N=ring |
| | N5y | ring-N(R)+ring |
| Undefined N | N0 | |

Oxygen 18 types

| | | |
|-----------------|-----|-------------------|
| Hydroxy | O1a | R-OH |
| | O1b | N-OH |
| | O1c | P-OH |
| | O1d | S-OH |
| Ether | O2a | R-O-R |
| | O2b | P-O-R |
| | O2c | P-O-P |
| | O2x | ring-O-ring |
| Oxo | O3a | N=O |
| | O3b | P=O |
| | O3c | S=O |
| Aldehyde | O4a | R-CH=O |
| Ketone | O5a | R-C(=O)-R |
| | O5x | ring-C(=O)-ring |
| Carboxylic acid | O6a | R-C(=O)-OH |
| Ester | O7a | R-C(=O)-O-R |
| | O7x | ring-C(=O)-O-ring |
| Undefined O | O0 | |

Sulfur 7 types

| | | |
|-------------|-----|---------------|
| Thiol | S1a | R-SH |
| Thioether | S2a | R-S-R |
| | S2x | ring-S-ring |
| Disulfide | S3a | R-S-S-R |
| | S3x | ring-S-S-ring |
| Sulfate | S4a | R-SO3 |
| Undefined S | S0 | |

Phosphorus 2 types

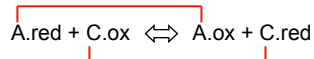
| | | |
|----------------------------|-----|-----|
| Attached to other elements | P1a | P-R |
| Attached to oxygen | P1b | P-O |

Other elements 2 types

| | | |
|----------|---|--------------|
| Halogens | X | F, Cl, Br, I |
| Others | Z | |

Reactant pairs extracted from enzymatic reactions

EC1. Oxidoreductases



EC2. Transferases



EC3. Hydrolases



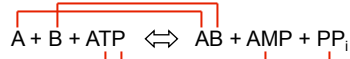
EC4. Lyases



EC5. Isomerases



EC6. Ligases



EC (Enzyme Commission) number

- EC numbers represent classification of enzymatic reactions (and enzymes)
- Manually assigned by IUBMB-IUPAC Biochemical Nomenclature Committee
- Published experimental evidence on the enzyme and the reaction is required
- Four numbers separated by periods, representing reaction specificity (class, sub-class, and sub-subclass numbers) and substrate specificity (serial number).

Reactant pair transformation patterns (RDM patterns)

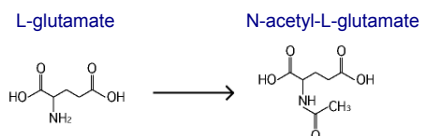
KEGG REACTION Database

Reaction

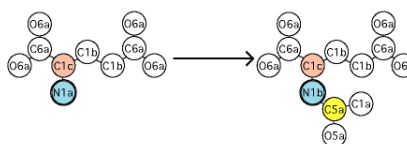


KEGG RPAIR Database

Reactant pair



Reactant pair in the KEGG atom type representation



RDM patterns

KEGG atom changes for R, D, M atoms

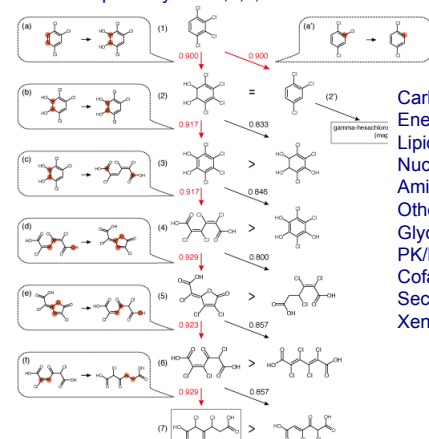
| | | |
|---|-----------------|------------|
| R | Reaction center | N1a -- N1b |
| D | Difference atom | (H) -- C5a |
| M | Matched atom | C1c -- C1c |

Kotera, M., Okuno, Y., Hattori, M., Goto, S., and Kanehisa, M.; Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions. *J. Am. Chem. Soc.* 126, 16487-16498 (2004).

Prediction of xenobiotics degradation pathways

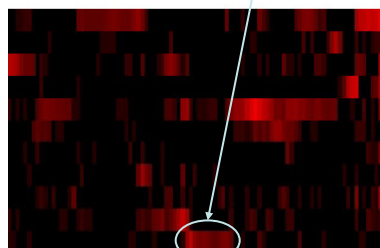
PathPred @ <http://www.genome.jp/tools/pathpred/>

Predicted pathways for 1,2,3,4-tetrachlorobenzene



Carbohydrate
Energy
Lipid
Nucleotide
Amino Acid
Other aa
Glycan
PK/NRP
Cofactors
Sec.metab.
Xenobiotics

RDM patterns for xenobiotics biodegradation pathways

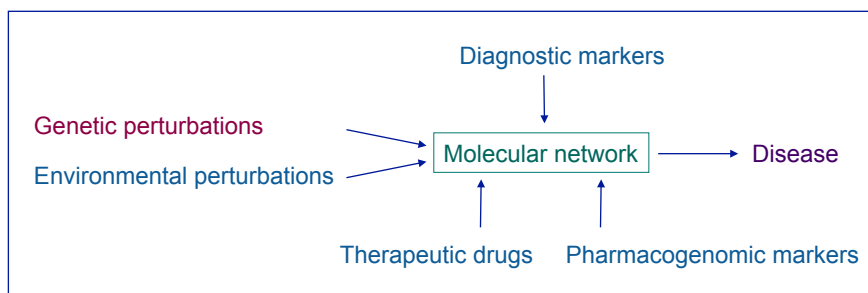


1,901 RDM patterns
Chemical structure transformation patterns in known enzymatic reactions

Oh, M., Yamada, T., Hattori, M., Goto, S., and Kanehisa, M.; Systematic analysis of enzyme-catalyzed reaction patterns and prediction of microbial biodegradation pathways. *J. Chem. Info. Model.* 47, 1702-1712 (2007).
Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Kotera, M., Goto, S., and Kanehisa, M.: PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res.* Epub 2010 Apr 30 (2010).

KEGG DISEASE and DRUG: Linking genomes to practical values

Developing 'Computable' Disease Information Resource



KEGG disease pathway maps

- molecular network representation
- limited number of diseases

KEGG DISEASE entries

- simple list representation
- diseases with known genes/genomes

Diseases with known genetic factors

- Cancers
- Nervous system diseases
- Immune system diseases
- Circulatory system diseases
- Urinary system diseases
- Metabolic diseases
- Genetic disorders

Infectious diseases with known pathogen genomes

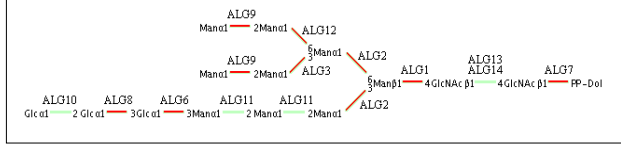
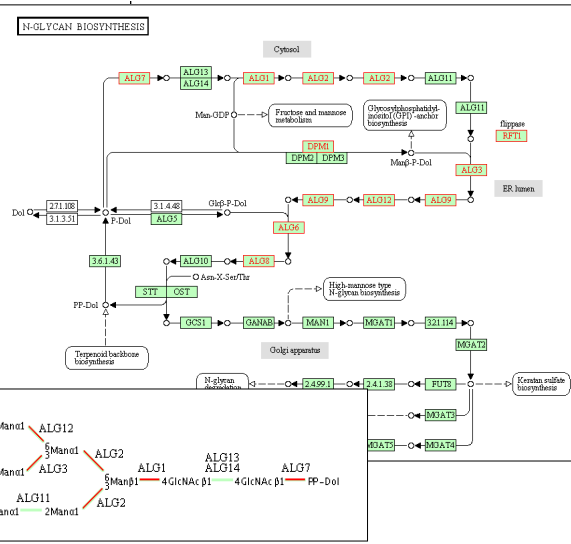
- Bacterial infections
- Viral infections
- Mycosis
- Parasitic infections

Congenital disorders of glycosylation (CDG)

KEGG DISEASE: H00118 Single-gene disease with various types

| | | |
|--------------------|--|---------|
| Entry | H00118 | Disease |
| Name | Congenital disorders of glycosylation (CDG) type I | |
| Description | Congenital disorders of glycosylation (CDGs) are a group of autosomal recessive disorders caused by defects in various genes for N-glycan biosynthesis. CDG types caused by one of the enzyme genes involved in the synthesis of dolichol bound N-glycan precursor. | |
| Category | Inherited metabolic disease | |
| Pathway | hsa00510 N-Glycan biosynthesis hsa00051 Fructose and mannose metabolism | |
| Gene | (CDG-1a) PM2; phosphomannomutase 2 [HSA:5373] (CDG-1b) MPI; mannanose phosphate isomerase [HSA:4351] (CDG-1c) ALG6; α1-3 glucosyltransferase [HSA:29929] (CDG-1d) ALG3; α1-3 mannosyltransferase [HSA:18195] (CDG-1e) DPM1; P-dolichol mannosyltransferase [HSA:18133] (CDG-1f) MPOU1; mannanose-P-dolichol utilization defect 1 [HSA:18133] (CDG-1g) ALG2; α1-6 mannosyltransferase [HSA:79887] (CDG-1h) ALG2; α1-3/6 glucosyltransferase [HSA:79887] (CDG-1i) SMPAT1; ALG7; P-dolichol N-acetylglucosaminophospho-1-4 mannosyltransferase [HSA:56825] (CDG-1j) ALG9; α1-2 mannosyltransferase [HSA:79796] (CDG-1k) DOLK; dolichol kinase [HSA:27845] (CDG-1n) RFT1; oligosaccharyl-lipid flippase [HSA:91869] | |
| Reference | PMID:1651948 | |
| Authors | Sparks SE | |
| Title | Inherited disorders of glycosylation. | |
| Journal | Mol Genet Metab 87:1-7 (2006) | |
| Reference | PMID:12756558 | |
| Authors | Marquardt T, Denecke J | |
| Title | Congenital disorders of glycosylation: review of their molecular presentations and specific therapies. | |
| Journal | Eur J Pediatr 162:359-79 (2003) | |
| Reference | PMID:12905814 (CDG type Ia) | |
| Authors | Neumann LM, von Maer A, Kunze J, Blankenstein O, Marquardt T | |
| Title | Congenital disorder of glycosylation type Ia in a mucopolysaccharidosis-like atypical phenotype and homozygosity of the N216I mutation. | |
| Journal | Eur J Pediatr 162:718-3 (2003) | |

KEGG pathway map for N-Glycan biosynthesis

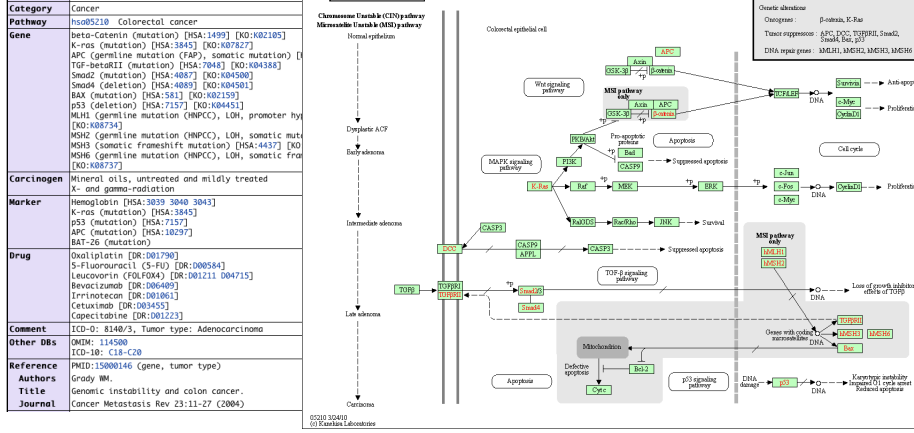


Colorectal cancer

KEGG DISEASE: H00020 Cancers

55 DISEASE entries
14 pathway maps
1 global map

KEGG disease pathway map for colorectal cancer



Chronic myeloid leukemia (CML)

Entry H00004 Disease

Name Chronic myeloid leukemia (CML)

Description Chronic myelogenous leukemia (CML) originates in a pluripotent hematopoietic stem cell of the bone marrow and is characterized by greatly increased numbers of granulocytes in the blood. Myeloid and other hematopoietic cell lineages are involved in the process of clonal proliferation and differentiation. On the cellular level, CML is associated with a specific chromosome abnormality, the t(9; 22) reciprocal translocation that forms the Philadelphia (Ph) chromosome. The Ph chromosome is the result of a molecular rearrangement between the c-ABL proto gene on chromosome 9 and the BCR (breakpoint cluster region) gene on chromosome 22. The BCR gene codes for an 8-kb mRNA and a novel 230-kDa protein with tyrosine kinase activity that is greater than the normal c-ABL protein. The progression to blast crisis requires other abnormalities include mutations in TP53, RB genes such as EVI1. Additional chromosome 4 t(3; 21)(q26;q22), which generates AML1-EVI1.

Category Cancer

Pathway H000230 Chronic myeloid leukemia

Gene BCR-ABL (translocation) [HSA:613.25] EVI1 (overexpression) [HSA:2122] AML1 (translocation) [HSA:182] p16/INK4A (mutation) [HSA:1823] p53 (mutation) [HSA:7157] RB1 (mutation) [HSA:2922]

Carcinogen 1,3-Butadiene [EPD:116450] Rubber industry

Marker BCR-ABL (translocation) [HSA:613.25] WT1 [HSA:7490]

Drug Imatinib mesylate (Gleevec) [DR:001441] Hydroxyurea [DR:000311] Interferon- α [DR:000745 DR:02745 DR:03305 DR:03306]

Comment ICD-O: 8475/3, Tumor type: Chronic myeloid

Reference PMID:15719031 (gene, tumor type)

Authors Ren R.

Title Mechanisms of BCR-ABL in the pathogenesis of Nat Rev Cancer 5:172-83 (2005)

Reference PMID:10403855 (gene)

Authors Faderl S, Tajima M, Estrov Z, O'Brien S, Niu

Title The biology of chronic myeloid leukemia. N Engl J Med 341:164-72 (1999)

Reference PMID:14382876 (gene)

Authors Colabretto B, Ferruoti D.

Title The biology of CML blast crisis. Blood 103:4010-22 (2004)

Reference PMID:16155477 (carcinogen)

Authors Graff J, Sathikumar N, Macaluso M, Maldon

Title Chemical exposures in the synthetic rubber mortality. J Occup Environ Med 47:916-32 (2005)

Reference PMID:9489905 (Carcinogen)

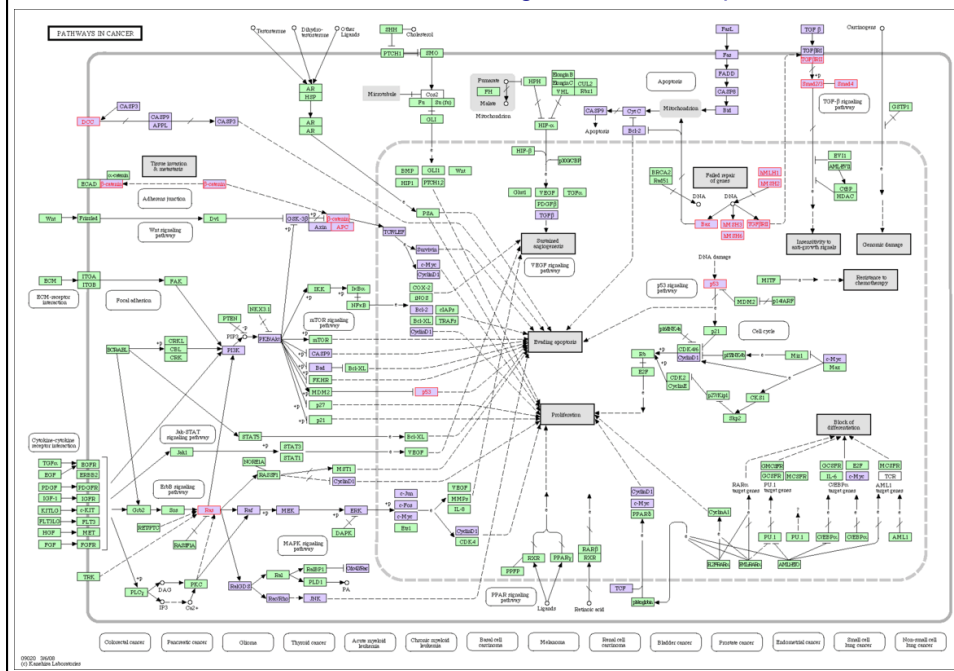
Authors Blair A, Kawanishi N.

Title Reactive chemicals and cancer. Cancer Causes Control 8:473-90 (1997)

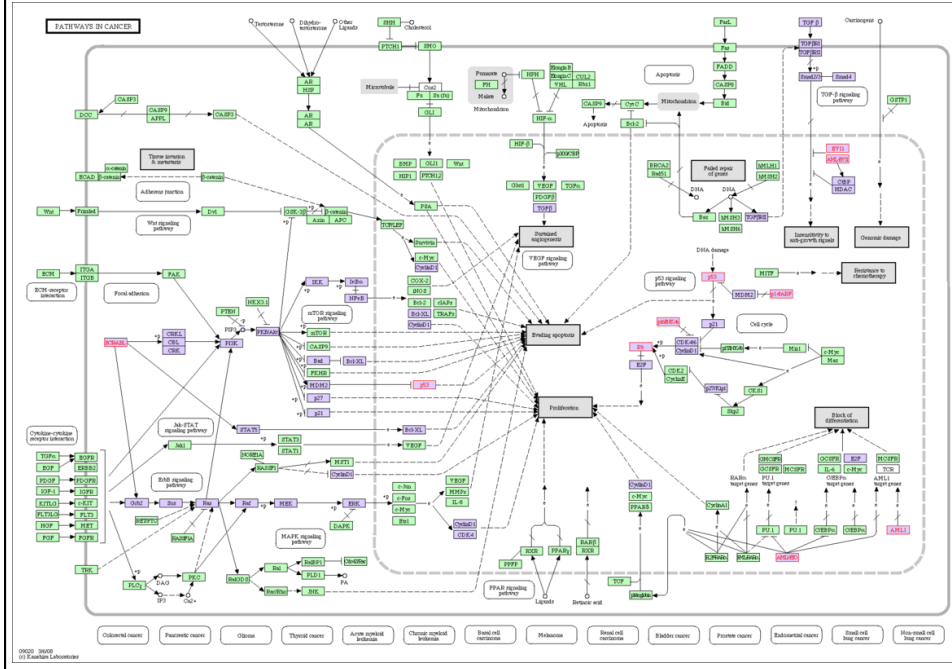
Reference PMID:16873426 (Carcinogen)

KEGG disease pathway map for chronic myeloid leukemia

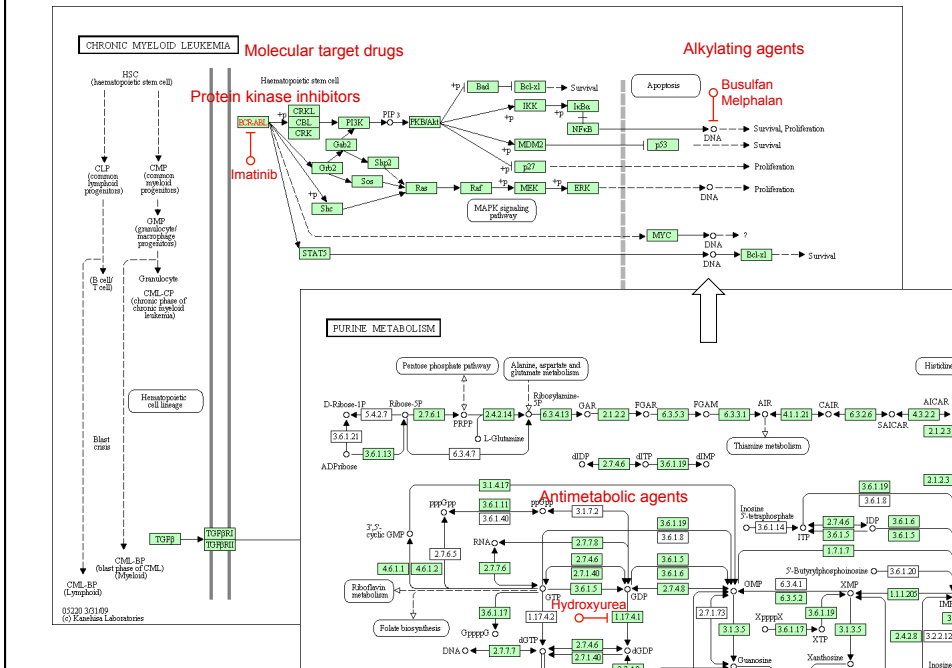
Colorectal cancer in the global cancer map



Chronic myeloid leukemia in the global cancer map



Drugs for chronic myeloid leukemia



KEGG DRUG as a Chemical Structure Database

Chemical structures and/or chemical components of:


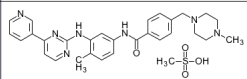
- All prescription drugs in Japan
- All OTC drugs in Japan
- Most prescription drugs in USA
- Many prescription drugs in Europe

Structure activity relationships organized in:

- ATC classification of drugs in Japan, USA, and Europe
- Therapeutic category of drugs in Japan
- Classification of OTC drugs in Japan
- TCM drugs in Japan
- Crude drugs in Japan

| | |
|-----------------------|-------|
| Number of entries | |
| KEGG DRUG: | 9,454 |
| Crude drugs | 531 |
| TCM formulas: | 229 |
| (As of June 12, 2010) | |

Imatinib mesylate (Gleevec)

| | |
|---|--|
|  DRUG: D01441 | |
| Entry | D01441 Drug |
| Name | Imatinib mesilate (JAN); Imatinib mesylate; Gleevec (TN); Glivec (TN) |
| Formula | C29H31N7O2, CH5O3 |
| Mass | 589.2471 |
| Structure |  Mod file KCF file DB search Jmol KegDraw |
| Target | Bcr-Abl tyrosine kinase inhibitor [HSA:25] [PATH:hs05228]; PDGFR tyrosine kinase inhibitor [HSA:5156] [PATH:hs04068]; PDGFRB tyrosine kinase inhibitor [HSA:5159] [PATH:hs04068]; c-KIT tyrosine kinase inhibitor [HSA:3815] [PATH:hs04068]; [PATH:hs04648] [PATH:hs05221] |
| Metabolism | CYP3A4 [HSA:1576] |
| Interaction | Biomarker: c-KIT expression [HSA:3815] |
| Activity | Antineoplastic |
| Remark | Therapeutic category: 4291 ATC code: L01XE01 (GITE therapy) |
| Comment | Treatment of chronic myelogenous leukemia (CML) and gastrointestinal stromal tumors (GIST) |
| Pathway | PATH: msp07045 Antineoplastics - protein kinases inhibitors |
| Other DBs | CAS: 228127-57-1 PubChem: 7845584 ChEBI: 31690 DrugBank: DB00019 LigandBox: D01441 NKAJ: J1_337_1443 |

ATC classification

Anatomical Therapeutic Chemical (ATC) classification [BR:br08303]
L ANTINEOPLASTIC AND IMMUNOMODULATING AGENTS
L01 ANTINEOPLASTIC AGENTS
L01X OTHER ANTINEOPLASTIC AGENTS
L01XE Protein kinase inhibitors
L01XE01 Imatinib
D01441 Imatinib mesilate (JAN)

KEGG DRUG as a Molecular Network Database


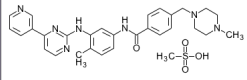
1. Molecular interaction network involving:

- Target molecules
- Drug metabolizing enzymes
- Drug transporters
- Adverse drug interactions
- Pharmacogenomic markers

2. Chemical structure transformation network for:

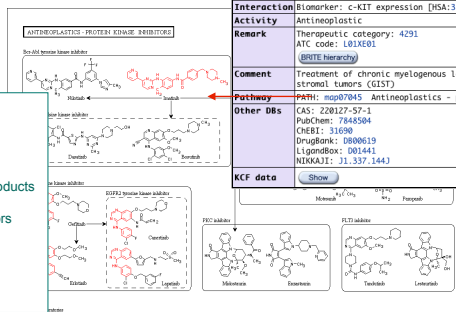
- Biosynthetic pathways of natural products (KEGG metabolic pathway maps)
- History of drug development (KEGG DRUG structure maps)

Imatinib mesylate (Gleevec)

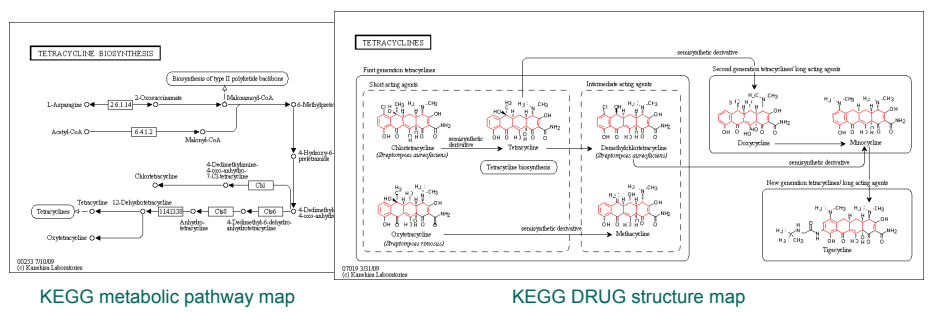
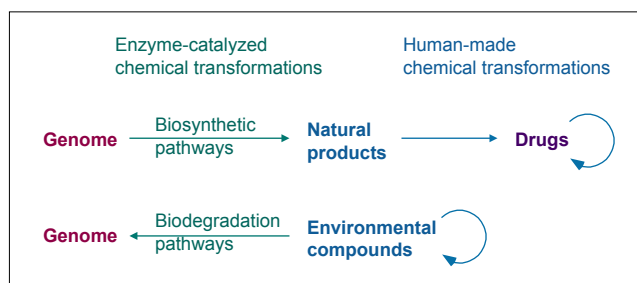
| | |
|---|--|
|  DRUG: D01441 | |
| Entry | D01441 Drug |
| Name | Imatinib mesilate (JAN); Imatinib mesylate; Gleevec (TN); Glivec (TN) |
| Formula | C29H31N7O2, CH5O3 |
| Mass | 589.2471 |
| Structure |  Mod file KCF file DB search Jmol KegDraw |
| Target | Bcr-Abl tyrosine kinase inhibitor [HSA:25] [PATH:hs05228]; PDGFR tyrosine kinase inhibitor [HSA:5156] [PATH:hs04068]; PDGFRB tyrosine kinase inhibitor [HSA:5159] [PATH:hs04068]; c-KIT tyrosine kinase inhibitor [HSA:3815] [PATH:hs04068]; [PATH:hs04648] [PATH:hs05221] |
| Metabolism | CYP3A4 [HSA:1576] |
| Interaction | Biomarker: c-KIT expression [HSA:3815] |
| Activity | Antineoplastic |
| Remark | Therapeutic category: 4291 ATC code: L01XE01 (GITE therapy) |
| Comment | Treatment of chronic myelogenous leukemia (CML) and gastrointestinal stromal tumors (GIST) |
| Pathway | PATH: msp07045 Antineoplastics - protein kinases inhibitors |
| Other DBs | CAS: 228127-57-1 PubChem: 7845584 ChEBI: 31690 DrugBank: DB00019 LigandBox: D01441 NKAJ: J1_337_1443 |
| KCF data | Show |

KEGG DRUG structure maps

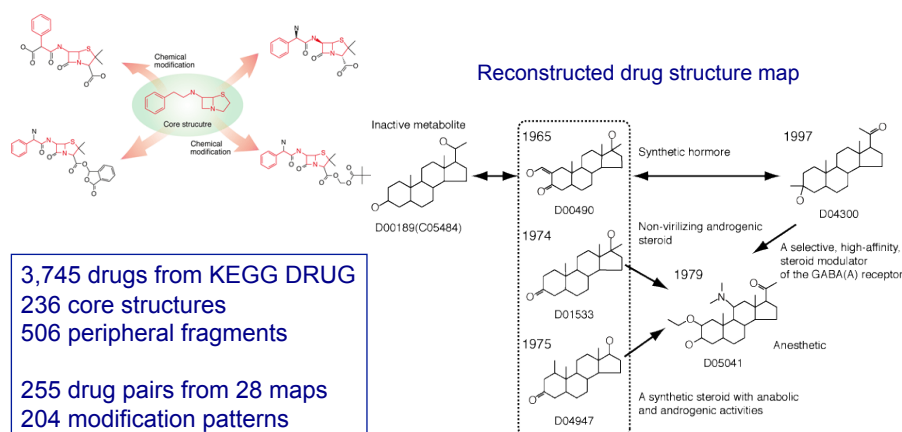
Drug Development
Chronology: Antibiotics
Chronology: Antineoplastics
07040 Antineoplastics - alkylating agents
07041 Antineoplastics - antimetabolic agents
07042 Antineoplastics - agents from natural products
07043 Antineoplastics - hormones
07045 Antineoplastics - protein kinases inhibitors
Chronology: Nervous System Agents
Chronology: Other Drugs
Target Based Classification
Structure Based Classification
Skeleton Based Classification



Chemical structure transformation networks Linking genomes to natural products and drugs



Chemical modification patterns in drug development Towards understanding chemical architecture of marketed drugs



Shigemizu, D., Araki, M., Okuda, S., Goto, S., and Kanehisa, M.; Extraction and analysis of chemical modification patterns in drug development. *J. Chem. Info. Model.* 49, 1122-1129 (2009).

Isoquinoline alkaloids

Structurally optimized drugs

| Phytochemical compound | drug | name | score | ATC |
|---|---------|--------------------------|-------|-----------------|
| Phenylpropanoids and related compounds | | | | |
| Polyketides | | | | |
| Terpenoids | | | | |
| Alkaloids | | | | |
| Alkaloids derived from ornithine | | | | |
| Alkaloids derived from lysine | | | | |
| Alkaloids derived from nicotinic acid | | | | |
| Alkaloids derived from tyrosine | | | | |
| Isoquinoline alkaloids | | | | |
| C03480 Tyramine | D027820 | Dopamine | 0.91 | C01CA04 |
| C03788 4-(2-Aminoethyl)-1,2-benzenedio | D027820 | Dopamine | 1.0 | C01CA04 |
| C05533 Papaverine | D02218 | Papaverine hydrochloride | 1.0 | A03AD01 G04BE02 |
| C00757 Berberine | D01250 | Berberine chloride | 1.0 | |
| C03462 (+)-alpha-Hydrastine | D02172 | Noscapine hydrochloride | 0.94 | R05DA07 |
| C03992 alpha-Narcotine | D02172 | Noscapine hydrochloride | 1.0 | R05DA07 |
| C03953 Narcotine | D02172 | Noscapine hydrochloride | 0.97 | R05DA07 |
| C13798 (+)-Hydrastine | D02172 | Noscapine hydrochloride | 0.94 | R05DA07 |
| C01518 Morphine | D03823 | Morphine | 1.0 | N02AA01 |
| C08172 Codeine | D03185 | Codeine | 1.0 | R05DA04 |
| C08615 Pynamine | D02722 | Dimethyltubocurarine | 0.89 | M03AA04 |
| C09391 Cepharanthine | D01035 | Cepharanthine (JAN) | 1.0 | |
| C03651 (+)-Tetrandrine | D01035 | Cepharanthine (JAN) | 0.89 | |
| C03352 (+)-Berberine | D02207 | Tubocurarine chloride | 0.98 | M03AA02 |
| C03352 Berberine | D02722 | Dimethyltubocurarine | 0.89 | M03AA04 |
| C08415 Daphnandrine | D01035 | Cepharanthine (JAN) | 0.96 | |
| C03481 Hemanadine | D01035 | Cepharanthine (JAN) | 0.86 | |
| C03981 Macleodine | D03655 | Tubocurarine chloride | 0.85 | M03AA02 |
| C08623 Rodissine | D02722 | Dimethyltubocurarine | 0.88 | M03AA04 |
| C11141 (+)-Atherospermoline | D02207 | Tubocurarine chloride | 0.91 | M03AA02 |
| C07347 Tubocurarine | D03655 | Tubocurarine chloride | 1.0 | M03AA02 |
| C01782 Colchicine | D03620 | Colchicine | 1.0 | M04AC01 |
| C08520 Galanthamine | D04292 | Galantamine | 1.0 | N06DA04 |
| C09390 Cephaeline | D03385 | Emetine hydrochloride | 0.97 | P01AX02 |
| C09421 Emetine | D03385 | Emetine hydrochloride | 1.0 | P01AX02 |

- Alkaloids derived from typtophan and anthranilic
 - Alkaloids derived from histidine
 - Alkaloids derived by amination
 - Others
- C01C Cardiac stimulants
 - A03A Drugs for functional bowel disorders
 - R05D Cough suppressants
 - N02A Opioid analgesics
 - M03A Muscle relaxants
 - M04A Antigout preparations
 - N06D Anti-dementia drugs
 - P01A Antiprotozoals

Crude drugs

| Phytochemical compound | drug | name | name |
|---|--------|--------------------|--------|
| Phenylpropanoids and related compounds | | | |
| Polyketides | | | |
| Terpenoids | | | |
| Alkaloids | | | |
| Alkaloids derived from ornithine | | | |
| Alkaloids derived from lysine | | | |
| Alkaloids derived from nicotinic acid | | | |
| Alkaloids derived from tyrosine | | | |
| Isoquinoline alkaloids | | | |
| C12328 Reticuline | D02746 | Magnolia flower | シソイ |
| C05533 Papaverine | D02444 | Opium, powdered | アヘン |
| C03339 (+)-Annonaine | D02720 | Magnolia bark | コウボク |
| C09387 Bulbocapnine | D06686 | Corydalis tuber | エンゴサク |
| C03581 Magnoflorine | D00392 | Coptis rhizome | オウレン |
| C03581 Magnoflorine | D06689 | Phellodendron bark | オウバク |
| C03581 Magnoflorine | D02720 | Magnolia bark | コウボク |
| C03581 Magnoflorine | D06808 | Epimedium herb | インヨウカク |
| C03581 Magnoflorine | D02783 | Sinomenium stem | ボウイ |
| C02792 Berberine | D06396 | Coptis rhizome | オウレン |
| C00757 Berberine | D06689 | Phellodendron bark | オウバク |
| C01789 Columbamine | D06728 | Calumba | コロンボ |
| C05315 Palmatine | D00392 | Coptis rhizome | オウレン |
| C05315 Palmatine | D06689 | Phellodendron bark | オウバク |
| C05315 Palmatine | D06728 | Calumba | コロンボ |
| C09553 Jatrorhizine | D00392 | Coptis rhizome | オウレン |
| C09553 Jatrorhizine | D06689 | Phellodendron bark | オウバク |
| C04118 6,8,13,13a-Tetrahydrocolumbamine | D06686 | Corydalis tuber | エンゴサク |
| C05175 6,7,12b,13a-Tetrahydro-4H-bis[1,8] | D06686 | Corydalis tuber | エンゴサク |
| C02134 Allocryptopine | D06686 | Corydalis tuber | エンゴサク |
| C05189 Protopine | D06686 | Corydalis tuber | エンゴサク |
| C08392 alpha-Narcotine | D03444 | Opium, powdered | アヘン |
| C01518 Morphine | D03444 | Opium, powdered | アヘン |
| C05179 Salutaridine | D06785 | Sinomenium stem | ボウイ |
| C09543 Sinomenine | D02783 | Sinomenium stem | ボウイ |
| C03381 Cepharanthine | D02737 | Stephaura cepi | シロハク |
| C03357 Berberine | D02737 | Stephaura cepi | シロハク |
| C07847 Tubocurarine | D03305 | Curate | 辛温發表薬 |
| C07382 Colchicine | D02307 | Colchicum seed | 清熱薬 |
| C03389 Cephaeline | D03868 | Ipecac | 健胃・止瀉薬 |
| C09420 Emetamine | D03868 | Ipecac | 補気強壯薬 |
| C09421 Emetine | D03868 | Ipecac | 駆瘀血薬 |
| C09612 Psychoflorine | D03868 | Ipecac | 利水薬 |
| C09612 Psychoflorine | D03868 | Ipecac | 去湿健胃薬 |
| C09612 Psychoflorine | D03868 | Ipecac | 鎮咳去痰薬 |

- Alkaloids derived from typtophan and anthranilic
 - Alkaloids derived from histidine
 - Alkaloids derived by amination reactions
 - Others
- 辛温發表薬
 - 清熱薬
 - 健胃・止瀉薬
 - 補気強壯薬
 - 駆瘀血薬
 - 利水薬
 - 去湿健胃薬
 - 鎮咳去痰薬

KEGG MEDICUS

Molecular network based information resource for human diseases and drugs

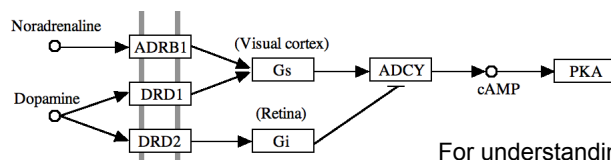
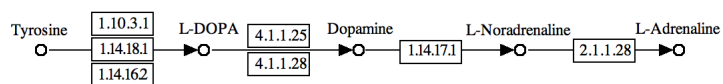
| | KEGG DISEASE | KEGG DRUG |
|-----------------|---|--|
| URL | http://www.genome.jp/kegg/disease/ | http://www.genome.jp/kegg/drug/ |
| Content | Diseases with known genetic factors and infectious diseases with known pathogen genomes, represented as lists of disease genes, environmental factors, markers, drugs, etc. | Chemical structures and/or components of all approved drugs, together with information of interactions with targets, metabolizing enzymes, transporter, genomic biomarkers, etc. |
| Pathway | KEGG pathway maps for human diseases: cancers, immune system diseases, neurodegenerative diseases, cardiovascular diseases, metabolic disorders, and infectious diseases | KEGG pathway maps for drug metabolism and secondary metabolite and antibiotics biosynthesis KEGG DRUG structure maps for the history of drug development |
| BRITE hierarchy | Disease classifications including: Human diseases ICD-10 disease classification | Drug classifications including: ATC classification (WHO) Therapeutic category of drugs (Japan) OTC drug classification in Japan Crude drugs and TCM drugs in Japan |

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M.; KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 38, D355-D360 (2010).

Summary

Three types of molecular networks in KEGG PATHWAY

1. Molecular reaction network
2. Chemical structure transformation network



3. Molecular interaction network

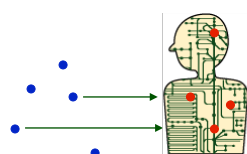
For understanding, for example, how chemical substances are biosynthesized, biodegraded, and used as signaling molecules

| | |
|------------------------|-----|
| Metabolic pathway maps | 1+2 |
| Drug structure maps | 2 |
| All other maps | 3 |

Molecular network-based approaches to diseases, drugs, and environmental compounds

- Diseases viewed as perturbed states of the molecular system
- Drugs and environmental compounds viewed as perturbants to the molecular system

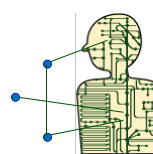
Molecular interaction network



Perturbants Perturbed system

- Chemical perturbants (environmental factors of diseases, drugs)
- Genetic perturbants (genetic factors of diseases)

Molecular reaction network



Chemical structure transformations

- EDCs, etc. Biodegradation pathway in bacteria
- Phytoestrogens, etc. Biosynthetic pathway in plants

Changing roles of bioinformatics: Basic research to practical values

- Capturing knowledge on molecular systems both in normal and perturbed (disease) states **KEGG PATHWAY**
KEGG BRITE
KEGG DISEASE
- Capturing knowledge on drugs and environmental compounds as perturbants to molecular systems **KEGG DRUG**
KEGG COMPOUND
- Generalizing knowledge on genes and proteins as ortholog groups **KEGG ORTHOLOGY**
KEGG GENES
- Generalizing knowledge on chemical transformations in enzymatic reactions by RDM patterns **KEGG RCLASS**
KEGG REACTION
- Knowledge-based analysis of human disease mechanisms
- Drug discovery from the genomes of plants and microorganisms
- Knowledge-based prediction of xenobiotics degradation pathway and responsible genes