KEGG: Kyoto Encyclopedia of Genes and Genomes

Minoru Kanehisa
Institute for Chemical Research,
Kyoto University

June 2012

Background

- High-throughput experiments and large-scale datasets
- Bioinformatics approaches to integration and interpretation
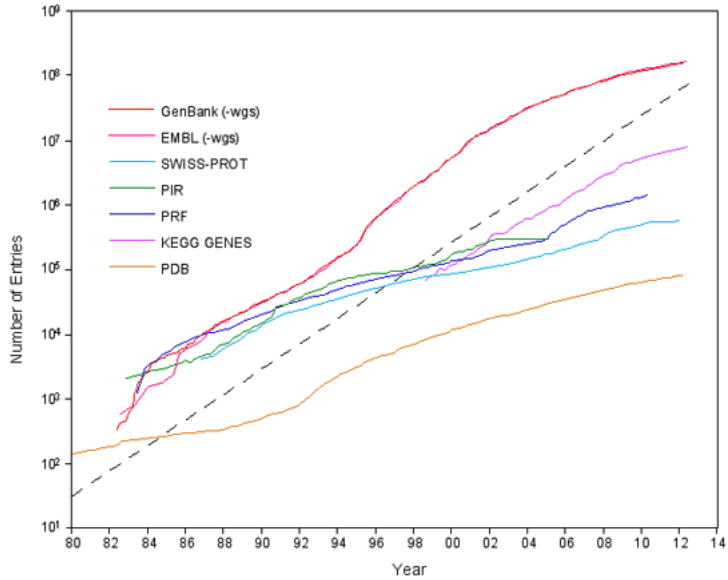- Linking genomes to health and society

Seeds and Needs: Changing Roles of Bioinformatics

Period	Seeds	Needs
1990s	Informatics technologies	Large-scale sequence data in the Human Genome Project
2000s	Bioinformatics technologies (bench)	Large-scale molecular data in all areas of biological sciences
2010s	Bioinformatics technologies (bedside)	Scientific data (personal genome data, etc.) in society

Translational Bioinformatics

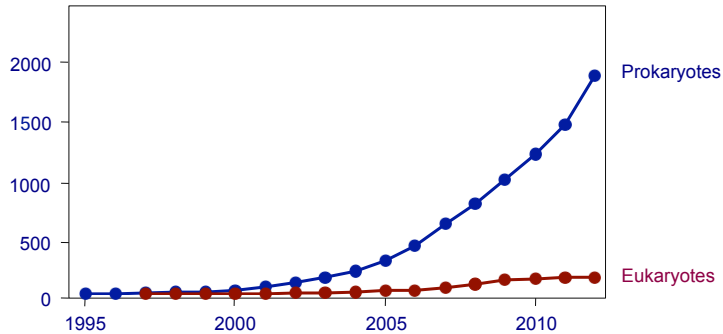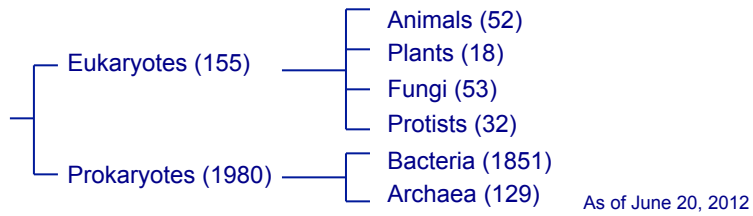|	Role	Example
Research Community	Helping to bring research results into practical applications	Personalized medicine Drug discovery
Society	Helping to understand scientific basis of diseases and drugs	Participatory medicine Self-medication

Growth of sequence and 3D structure databases

Number of Entries vs. Year

Legend:
- GenBank (-wgs)
- EMBL (-wgs)
- SWISS-PROT
- PIR
- PRF
- KEGG GENES
- PDB

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

Complete Genomes in KEGG

Eukaryotes (155)
- Animals (52)
- Plants (18)
- Fungi (53)
- Protists (32)

Prokaryotes (1980)
- Bacteria (1851)
- Archaea (129)

As of June 20, 2012

Prokaryotes

Eukaryotes

More details at: http://www.kegg.jp/kegg/docs/upd_genome.html

3

Genetic and Chemical Blueprints of Life

Genetic Blueprint of Life (Genome)

- DNA sequence
- Information about genes, gene regulatory elements, proteins, RNAs, etc.
- Partial information about transcription network

Chemical Blueprint of Life (Chemical Logic)

- Molecular reaction/interaction network
- Information about small molecules, biopolymers, etc. (biosynthetic codes)
- Information about interactions with the environment

Molecular Building Blocks of Life
Genomic and Chemical Spaces

Genomic Space

DNA (Gene)	Genome
RNA	Transcriptome
Protein	Proteome

Replication
Transcription
Translation

Genetic code

Central dogma   DNA ⇌ RNA ⟶ Protein

Chemical Space

Metabolite	Metabolome
Glycan	Glycome
Lipid	Lipidome

(Endogenous)

| Compound | Chemical genome |

(Exogenous)
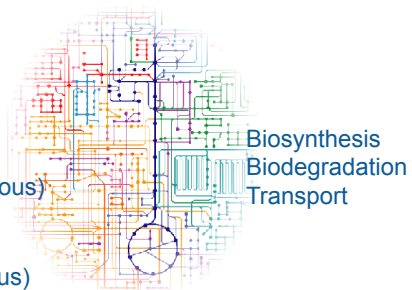
Biosynthesis
Biodegradation
Transport

Biosynthetic code

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

<u>Genomic space</u>

<u>Chemical space</u>

Genes and proteins
in individual organisms
(genomics, transcriptomics,
proteomics)

⟷

Endogenous molecules
(metabolomics, glycomics,
lipidomics, etc.)

⟸

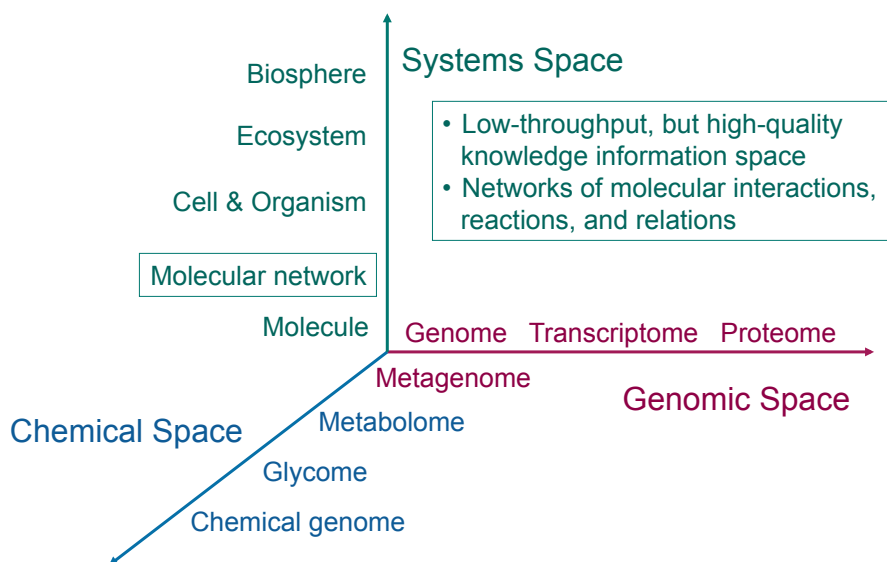Exogenous molecules
(chemical genomics)

Genes and proteins
in environmental samples
(metagenomics, etc.)

⟷

Metabolites
in environmental samples
(meta-metabolomics)

- Integration of genomic and chemical spaces
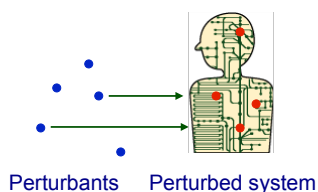- Interpretation of higher-level systemic functions

---

**Bioinformatics Approaches
to Reconstructing Biological Systems from Building Blocks**

Systems Space

Biosphere

Ecosystem

Cell & Organism

Molecular network

Molecule

- Low-throughput, but high-quality
  knowledge information space
- Networks of molecular interactions,
  reactions, and relations

Genome   Transcriptome   Proteome

Metagenome

Genomic Space

Chemical Space

Metabolome
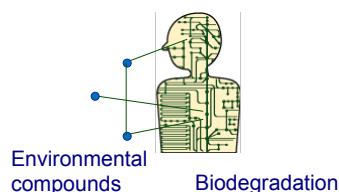
Glycome

Chemical genome

Molecular network-based analysis of
diseases, drugs, and environmental compounds

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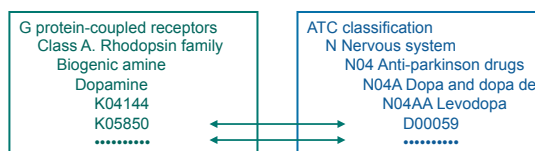- Drugs and environmental compounds viewed as perturbants to the molecular system

Molecular interaction network   Molecular reaction network
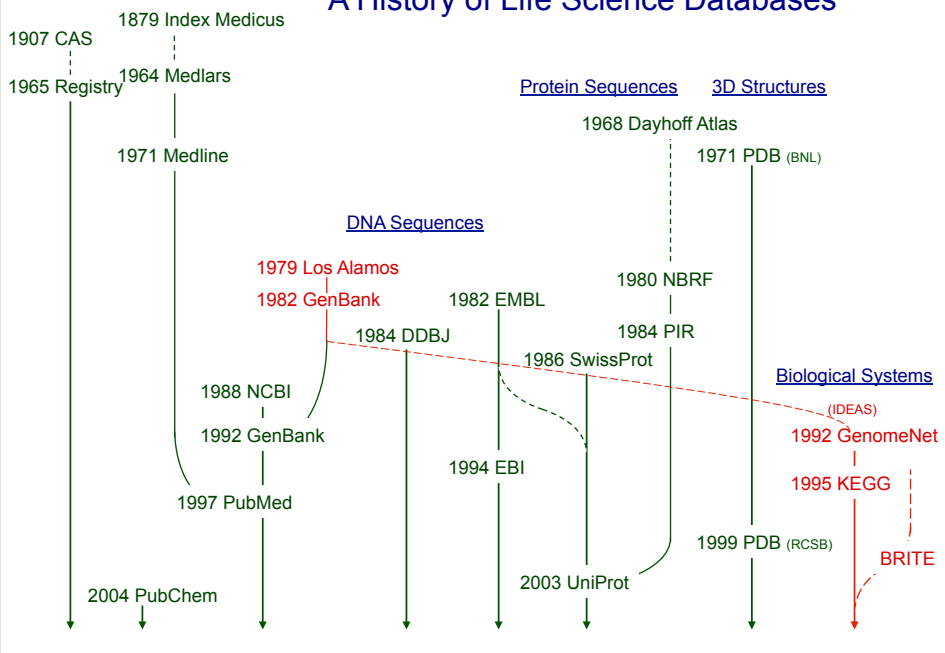
Perturbants   Perturbed system   Environmental
                                       compounds   Biodegradation

Molecular relation network

G protein-coupled receptors   ATC classification
  Class A. Rhodopsin family   N Nervous system
    Biogenic amine   N04 Anti-parkinson drugs
      Dopamine   N04A Dopa and dopa de
        K04144   N04AA Levodopa
        K05850   D00059
        ··········   ··········

Compounds   Literature   A History of Life Science Databases

1879 Index Medicus
1907 CAS

1965 Registry   1964 Medlars   Protein Sequences   3D Structures

                                 1968 Dayhoff Atlas
1971 Medline   1971 PDB (BNL)

                DNA Sequences

            1979 Los Alamos
            1982 GenBank   1982 EMBL   1980 NBRF
              1984 DDBJ   1984 PIR
                            1986 SwissProt   Biological Systems
      1988 NCBI   (IDEAS)
      1992 GenBank   1992 GenomeNet
                            1994 EBI   1995 KEGG
      1997 PubMed
                                       1999 PDB (RCSB)
                                                        BRITE
2004 PubChem   2003 UniProt

Overview of KEGG

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

The KEGG Database

1. Manually created reference knowledge base

   Experimental knowledge on systemic functions of the cell, the organism and the ecosystem is represented in terms of molecular networks (KEGG pathway maps, BRITE functional hierarchies and KEGG modules).

2. Enabling interpretation of genomes and high-throughput data

   A mechanism (KEGG Orthology system) is developed for linking genes in the genome to nodes of the molecular network.

3. Supporting translational bioinformatics

   Disease and drug information is integrated in the KEGG molecular networks (diseases as perturbed states and drugs as perturbants).

KEGG: Computer Representation of Biological Systems
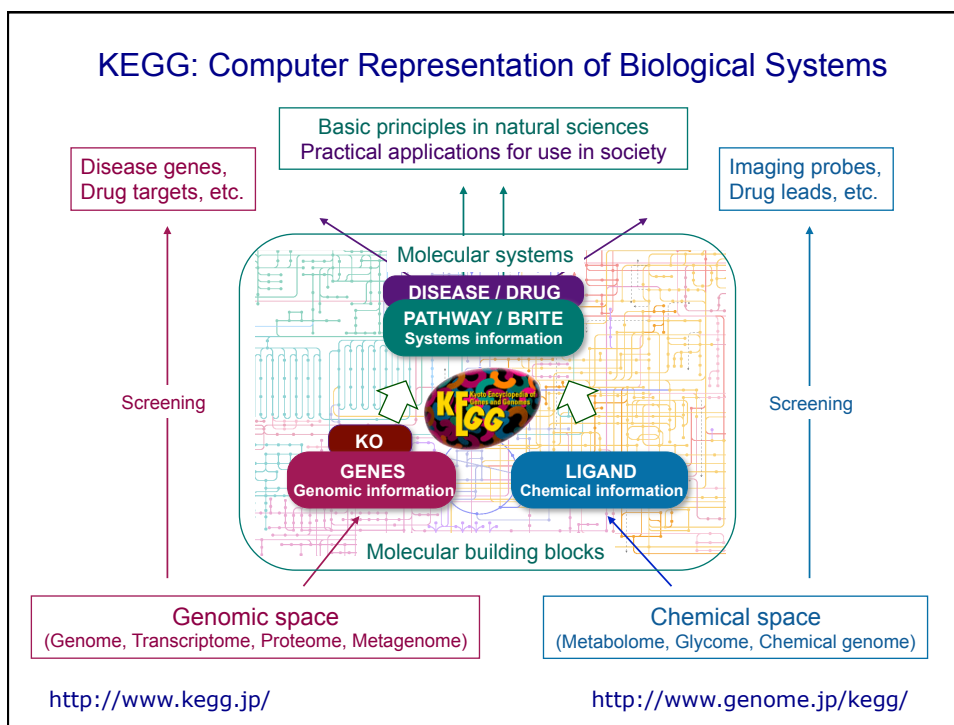
KEGG Databases

Database	Content	Data size
KEGG PATHWAY	Pathway maps, reference (total)	425 (189,042)
KEGG BRITE	Functional hierarchies, reference (total)	137 (61,818)
KEGG MODULE	KEGG modules, reference (total)	536 (134,097)
KEGG DISEASE	Human diseases	1,221
KEGG DRUG	Drugs	9,779
KEGG EDRUG	Crude drugs and health-related substances	840
KEGG ORTHOLOGY	KEGG Orthology (KO) groups	15,713
KEGG GENOME	KEGG Organisms	2,135
KEGG GENES	Genes in high-quality genomes	8,761,904
	(155 eukaryotes + 1847 bacteria + 129 archaea)	
KEGG SSDB	Best hit relations within GENES	64,225,605,182
	Bi-directional best hit relations within GENES	1,200,860,761
KEGG DGENES	Genes in draft genomes (17 eukaryotes)	359,223
KEGG EGENES	Genes as EST contigs (99 eukaryotes)	3,792,883
KEGG MGENES	Genes in metagenomes (139 samples)	14,822,267
KEGG COMPOUND	Metabolites and other small molecules	16,787
KEGG GLYCAN	Glycans	10,979
KEGG REACTION	Biochemical reactions	9,025
KEGG RPAIR	Reactant pair chemical transformations	13,530
KEGG RCLASS	Reaction class	2,521
KEGG ENZYME	Enzyme nomenclature	5,823

As of June 20, 2012

KEGG Object Identifiers

(Prefix + five-digit number)

Database	Content	Prefix	Example
KEGG PATHWAY	Pathway maps	map, ko, ec, rn, *org*	hsa04930
KEGG BRITE	Functional hierarchies	br, jp, ko, *org*	ko01003
KEGG MODULE	KEGG modules	M, *org*_M	M00008
KEGG DISEASE	Human diseases	H	H00004
KEGG DRUG	Drugs	D	D01441
KEGG ENVIRON	Crude drugs, etc.	E	E00048
KEGG ORTHOLOGY	KEGG Orthology (KO) groups	K	K04527
KEGG GENOME	KEGG Organisms	T	T01001 (hsa)
KEGG GENES	Genes in high-quality genomes		hsa:3643
KEGG COMPOUND	Metabolites & small molecules	C	C00031
KEGG GLYCAN	Glycans	G	G00109
KEGG REACTION	Biochemical reactions	R	R00259
KEGG RPAIR	Reactant pairs	RP	RP04458
KEGG RCLASS	Reaction class	RC	RC00046
KEGG ENZYME	Enzyme nomenclature		ec:2.7.10.1

org:gene	KEGG GENES entry identifier	*db:entry*	DBGET database entry identifier
org	KEGG Organism code	*db*	dababase name
gene	gene identifier (locus_tag, GeneID, etc.)	*entry*	entry name

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 inherent 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

Life science databases

Web

Try searching "kegg" in Google and NCBI.

Access the KEGG top page: http://www.kegg.jp/ 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.
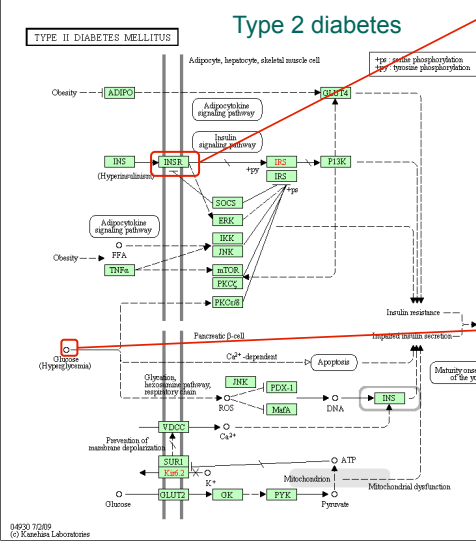
KEGG Pathway Maps
Molecular network representation of experimental knowledge

KEGG Orthology (KO)
Genome-based generalization of experimental knowledge

Insulin receptor

K04527

KO (K number) entries represent manually defined ortholog groups corresponding to the KEGG pathway nodes and the BRITE hierarchy nodes (bottom leaves)

Protein kinases (ko01001)
- Serine/threonine protein kinases
- Tyrosine protein kinases
  - InsR family
- Histidine protein kinases

Enzyme-linked receptors (ko01020)
- Receptor tyrosine kinase
  - RTK class II
- Receptor serine/threonine kinases

Cellular antigens (ko04090)
- CD molecules
- Non-CD molecules

Type II diabetes mellitus (map04930)

Go to KEGG MEDICUS:
http://www.kegg.jp/kegg/medicus.html
and enter a drug name of your interest.

Try, for example, statin (HMG-CoA reductase inhibitor) to retrieve cholersterol lowering drugs.

Lipitor (Atorvastatin)

KEGG Pathway Map
How it works

KEGG DRUG structure map
How it was developed

DRUG: D02258

Entry	D02258 Drug
Name	Atorvastatin calcium hydrate (JP16); Lipitor (TN)
Products	LIPITOR (A-S Medication Solutions LLC), ATORVASTATIN CALCIUM (American Health Packaging), LIPITOR (Cardinal Health), ATORVASTATIN CALCIUM (Dispensing Solutions), LIPITOR (Lake Erie Medical & Surgical Supply DBA Quality Care Products LLC), LIPITOR (PHARMAKON), LIPITOR (Physicians Total Care), ATORVASTATIN CALCIUM (Physicians Total Care), ATORVASTATIN CALCIUM (Ranbaxy Pharmaceuticals), ATORVASTATIN CALCIUM (Rebel Distributors Corp), LIPITOR (Rebel Distributors Corp), LIPITOR (PD-Rx Pharmaceuticals), LIPITOR (PD-Rx Pharmaceuticals), LIPITOR (PD-Rx Pharmaceuticals), LIPITOR (Parke-Davis Division of Pfizer)
Formula	(C33H34FN2O5)2. 3H2O. Ca
Exact mass	1208.4846
Mol weight	1209.3876
Structure	

$$(C_{33}H_{34}FN_2O_5)_2 \cdot 3H_2O \cdot Ca$$

Target	hydroxymethylglutaryl coenzyme A (HMG-CoA) reductase inhibitor [HSA:3156] [KO:K00021]
Pathway	hsa00900 Terpenoid backbone biosynthesis
Metabolism	Enzyme: CYP3A4 [HSA:1576]
Interaction	Genomic biomarker: LDLR [HSA:3949]
	DDI search
Activity	Inhibitor [HMG-CoA reductase]
Remark	Therapeutic category: 2189 ATC code: C10AA05
	BRITE hierarchy
Structure map	map07024 HMG-CoA reductase inhibitors
Other DBs	CAS: 344423-98-9 PubChem: 7849317 LigandBox: D02258 NIKKAJI: J1.648.241B
KCF data	Show

KEGG MEDICUS My Meidcation List
Any drug interactions and
non-drug interactions

My Medication List
(Japanese version only)

1. 'Okusuri-techo' in Japan

Keeps track of prescribed drugs
Used for checking drug allergies and past adverse events
Used for checking drug interactions and overdose of similar drugs

2. Drug interaction database

Molecular to higher-level drug interactions including:
• Known drug-drug interactions (contraindications, warnings, etc.)
• Interactions with OTC drugs, food supplements, etc.
• Personal health status
• Personal genomes

3. User interface

Automatic checking against drug interaction database
Links to KEGG and other web resources
Personal data managed by individuals (no data stored on the server)

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

Common ancestor
A

B   C
Human   Mouse

A

A1 Common A2
    ancestor

B1 C1   B2
Human Mouse Human Mouse
              C2 C3
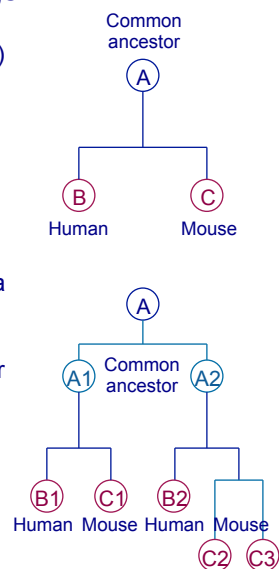
Computational Identification of Orthologs

Between two organisms

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

Genome 1   Genome 2
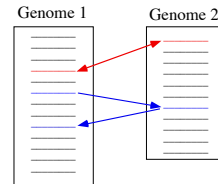
Among multiple organisms

1. COG
   Triangle of BBH relationships among three organisms
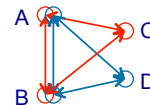
2. KEGG OC
   p-Quasi clique among multiple organisms

A   C
B   D

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

Separate function database
- Experimental evidence is stored in the KEGG ORTHOLOGY (KO) database
- KO entries identified by K numbers are manually defined in the context of molecular networks; i.e., as KEGG pathway nodes and BRITE hierarchy nodes

Ortholog annotation
- KO (K number) assignment; i.e., it establishes links from the KO database
- Gene definitions in the original database (mostly RefSeq) are not rewritten
- But the definitions of KO entries are frequently updated to follow guidelines

Cross-species annotation
- Molecular network-based annotation; i.e., starting from a pathway map, etc. orthologs are searched in all available genomes
- Genome-based annotation is also done

Genome 1   Genome 2   ......

Map 1

Map 2

Map-based
(cross-species)
annotation

Genome-based
annotation
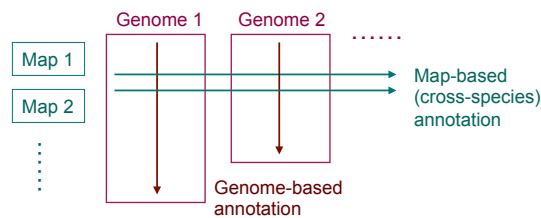
Genomes	2,134
Genes	8,775,231
Genes with KO	3,532,258
KO assignment	40%

KO	15,713
Safe KO	9,157
KOALA automation	58%

As of June 20, 2012

KOALA: KEGG Orthology And Links Annotation

Integrating computationally generated SSDB graph
and manually defined KO groups

RefSeq, etc.　　　　　　　　　　　　　　Literature, etc.

KEGG GENES　　　　　　　　　　　KEGG PATHWAY / KEGG BRITE

SSEARCH

SSDB graph　　　　　　　KEGG Oracle　　　Manual grouping of KO entries

　　　　　　　　　　　　Genes | KO

GFIT tables

　　　　　　　　　　KOALA tool / GFIT tool

KOALA program

Computational annotation　　Manual annotation　　Manual re-grouping of KO entries

SSDB is a huge graph weighted by similarity scores and directed by best-hit relations

Original membership of each KO group (subgraph of SSDB) is manually defined

---

Tools to check annotation quality
Based on reconstruction of functional units

Ortholog table

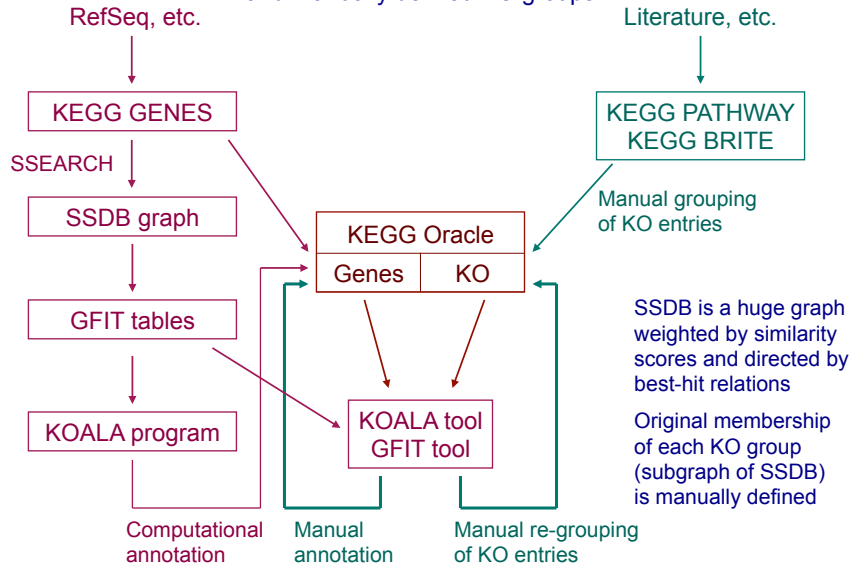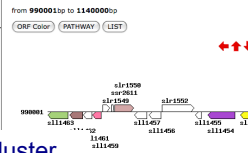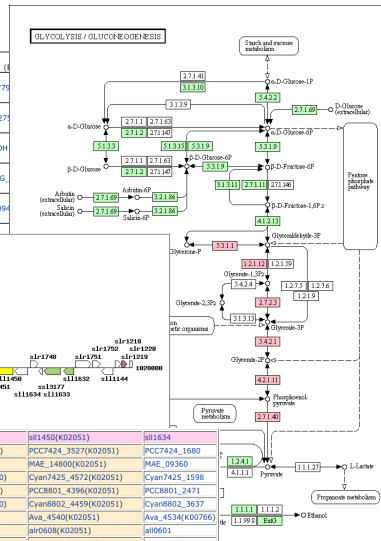| Organism	K01803 (TPI) [1170]	K00134 (GAPDH) [1469]	K00150 (gap2) [127]	K00927 (PGK) [1108]	K01834 (PGAM) [2711]	K01837 (BPGM) [1	
eco	b3919	b1779		b2926	b0755 b3612 b4395	b2779	
ecj	JW3890	JW1413 JW1768		JW2893	JW0738 JW3587 JW4358	JW	
ecd	ECDH10B_4108	ECDH10B_1917		ECDH10B_3101	ECDH10B_0822 ECDH10B_3794 ECDH10B_4553	ECDH	
ebw	BWG_3588	BWG_1592		BWG_2649	BWG_3303 BWG_0607 BWG_4087	BWG_	
ece	Z5464	Z2304 Z2818		Z4265	Z0925 Z5039 Z5997	Z4094	
ecs	ECs4844	ECs2022 ECs2488			ECs4490		
ecf	ECH74115_5374	ECH74115_2503 ECH74115_2022					

KEGG Genome map – Synechocystis sp. PCC6803

from **990001**bp to **1140000**bp

ORF Color | PATHWAY | LIST

Conserved gene cluster

syn	sll1455	sll1454(K00367)	sll1453(K02049)	sll1452(K02049)	sll1451(K02050)	sll1450(K02051)	sll1634
cyc	PCC7424_3464	PCC7424_3462	PCC7424_3524(K02049)	PCC7424_3525(K02049)	PCC7424_3526(K02050)	PCC7424_3527(K02051)	PCC7424_1680
mar	MAE_55270	MAE_53960	MAE_14770(K02049)	MAE_14780(K02049)	MAE_14790(K02050)	MAE_14800(K02051)	MAE_09360
cyn	Cyan7425_3581	Cyan7425_4566		Cyan7425_4570(K02049)	Cyan7425_4571(K02050)	Cyan7425_4572(K02051)	Cyan7425_1598
cyp	PCC8801_4057	PCC8801_2463		PCC8801_4394(K02049)	PCC8801_4395(K02050)	PCC8801_4396(K02051)	PCC8801_2471
cyh	Cyan8802_4096	Cyan8802_3646		Cyan8802_4457(K02049)	Cyan8802_4458(K02050)	Cyan8802_4459(K02051)	Cyan8802_3637
ava	Ava_4546	Ava_4544(K00367)		Ava_4542(K02049)	Ava_4541(K02050)	Ava_4540(K02051)	Ava_4534(K00766)
ana	alr0614	alr0612(K00367)	alr0611(K02049)	alr0610(K02049)	alr0609(K02050)	alr0608(K02051)	all0601
cyb	CYB_0042	CYB_0040(K00367)		CYB_0037(K02049)	CYB_0036(K02050)	CYB_0035(K02051)	CYB_0333
cya	CYA_0174	CYA_0619(K00367)		CYA_0615(K02049)	CYA_0614(K02050)	CYA_0613(K02051)	CYA_0985
tel	tlr1356	tlr1355(K00367)		tlr1352(K02049)	tlr1351(K02050)	tlr1350(K02051)	tll1358
aeh		Mlg_1702(K00360)	Mlg_2587(K02049)	Mlg_1703(K02049)	Mlg_1704(K02050)		Mlg_1665
cps		CPS_4945(K00372)		CPS_3314(K02049)	CPS_3315(K02050)		
tcx		Tcr_1159(K00372)		Tcr_1155(K02049)	Tcr_1154(K02050)		

Public version of KOALA and GFIT tools

Read-only version of KOALA

Read-only version of GFIT

Search against other DBs, e.g., UniProt

SSDB query for orthologs and paralogs

KAAS: KEGG Automatic Annotation Server
http://www.genome.jp/tools/kaas/

Pairwise genome comparison

BLAST search in both directions

Set of aa seq
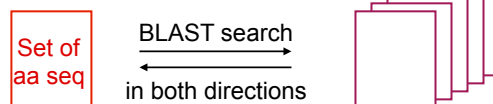
Query genome

Annotated genomes in KEGG

Bidirectional best hit

Genome 1   Genome 2

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

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

K number assignment by a heuristic scoring

$$S_{KO} = S_h - \log_2(mn) - \log_2\left(\sum_{k=N}^{x} {}_x C_k \, p^k (1-p)^{x-k}\right)$$
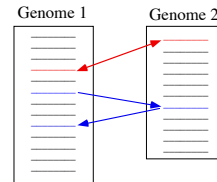
Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M.; KAAS: an automatic genome annotation and pathway reconstruction server. *Nucl. Acids Res.* 35, W182-W185 (2007).

KEGG PATHWAY and BRITE:
Reference knowledge base

Data objects
for computer representation of molecular systems

Data object	Database
Element 　　gene, protein, small molecule, etc.	KEGG GENES KEGG COMPOUND KEGG GLYCAN KEGG DRUG
Pair (binary relation) 　　reactant pair, drug-target relationship, etc.	KEGG RPAIR KEGG BRITE
Graph (wiring diagram) 　　pathway, complex, etc.	KEGG PATHWAY KEGG SSDB
Hierarchical list (tree) 　　hierarchical classification, ontology, etc.	KEGG BRITE
Simple list (membership) 　　disease gene list, etc.	KEGG MODULE KEGG DISEASE KEGG ORTHOLOGY

Set operations
for biological interpretation of large-scale datasets

Large scale data + Reference knowledge	Tool
KEGG pathway mapping (pathway enrichment) elements + KEGG pathway map	**KEGG Mapper**
KEGG pathway mapping (network extension) binary relations + KEGG pathway map	——
BRITE mapping (ontology enrichment) elements + BRITE functional hierarchy	**KEGG Mapper**
BRITE mapping (tree manipulation) binary relations + BRITE functional hierarchy	**KEGG Mapper**
KEGG annotation SSDB graph + KO group	**KOALA**

Knowledge Representation of Systemic Functions
(Molecular interaction, reaction and relation networks)

Pathway map
KEGG PATHWAY
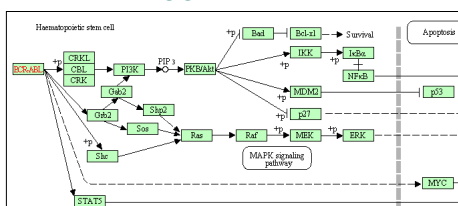
Hierarchical list (ontology)
KEGG BRITE

Membership list
KEGG DISEASE   KEGG MODULE

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

KEGG Pathway Maps

KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 PATHWAY BRITE MODULE DISEASE DRUG KO GENOME GENES LIGAND DBGET

Select prefix Enter keywords

map [Organism] [Go] Help

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps (see new maps and update history) representing our knowledge on the molecular interaction and reaction networks for:

0. Global Map
1. Metabolism
 Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid Glycan
 Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Overview
2. Genetic Information Processing
3. Environmental Information Processing
4. Cellular Processes
5. Organismal Systems
6. Human Diseases

and also on the structure relationships (KEGG drug structure maps) in:

7. Drug Development

Pathway Mapping

KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretation of higher-level systemic functions.

 - Search Pathway – basic pathway mapping tool
 - Search&Color Pathway – advanced pathway mapping tool
 - Color Pathway – selected pathway map coloring tool

0. Global Map

0.1 Metabolism

Metabolic pathways [zoom out]   Launch KEGG Atlas
Biosynthesis of secondary metabolites [zoom out]   Launch KEGG Atlas
Microbial metabolism in diverse environments [zoom out] Launch KEGG Atlas

1. Metabolism

1.1 Carbohydrate Metabolism

Glycolysis / Gluconeogenesis   Enzymes
Citrate cycle (TCA cycle)   Compounds with biological roles
Pentose phosphate pathway
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

425 pathway maps

BRITE Functional Hierarchies

KEGG BRITE Database

Functional hierarchies and binary relationships of biological entities

KEGG2 PATHWAY BRITE MODULE DISEASE DRUG KO GENOME GENES LIGAND DBGET

Functional Hierarchies

KEGG BRITE is a collection of hierarchical classifications (see new hierarchies and update history) representing our knowledge on various aspects of biological systems. In contrast to KEGG PATHWAY, which is limited to molecular interactions and reactions, KEGG BRITE incorporates many different types of relationships.

The hierarchies, especially those with additional tab-delimited fields, can better be examined with a desktop application called KegHier, which works on Mac, Windows, and Linux platforms.

 - Download KegHier

Search BRITE hierarchies for [____] [Go] [Clear]
 ☐ KEGG identifiers only

Brite Mapping

KEGG BRITE mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the BRITE functional hierarchies for biological interpretation of higher-level systemic functions.

 - Search Brite – basic brite mapping tool
 - Search&Color Brite – advanced brite mapping tool
 - Join Brite – selected brite hierarchy manipulation tool

Pathways and Ontologies
 KEGG pathway maps
 BRITE functional hierarchies

Genes and Proteins

 Network hierarchy
 KEGG Orthology (KO) (K numbers)
 KEGG modules (M numbers)

 Protein families (K numbers)
 Enzymes
 Protein kinases
 Peptidases
 Glycosyltransferases
 LPS biosynthesis proteins
 Lipid biosynthesis proteins
 Polyketide biosynthesis proteins
 Prenyltransferases

Diseases and Drugs

 Diseases (H numbers)
 Human diseases [+ gene]
 Infectious diseases [+ genome]
 Human diseases in ICD-10 classification

 Disease genes and pathways
 Cancer stages

 Prescription drugs (D numbers)
 ATC classification [+ target]
 USP drug classification
 Therapeutic category of drugs in Japan
 Japanese Pharmacopoeia
 Target-based classification of drugs
 Antiinfectives
 Antineoplastics

 OTC and crude drugs (D/E/C numbers)

137 brite hierarchies

As of June 20, 2012

KEGG PATHWAY mapping
Linking Genomes to Biological Systems

Human gene list with KO assignment

Gene	KO
3098	K00844
2645	K12407
2821	K01810
5211	K00850
2203	K03841
226	K01623
7167	K01803
2597	K00134
26330	K10705
5230	K00927
2023	K01689
5315	K00873
5313	K12406
5160	K00161
5162	K00162
1737	K00627
1738	K00382
3939	K00016
124	K13951
130	K13952
10327	K00002
217	K00128
218	K00129
55902	K01895
5236	K01835
2538	K01084
83440	K08074
....	

Reference pathway map
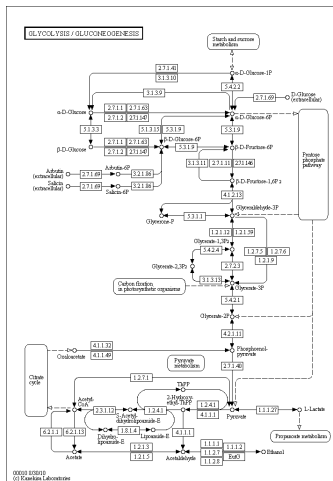
Glycolysis / Gluconeogenesis
map00010

+ **=**

Human pathway map

Glycolysis / Gluconeogenesis
hsa00010

19

Reference pathway

Metabolic pathways (global map)
map01100
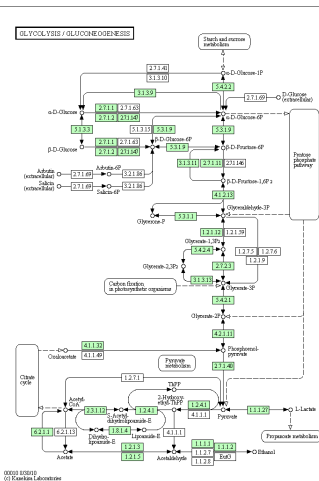
Glycolysis / Gluconeogenesis
map00010
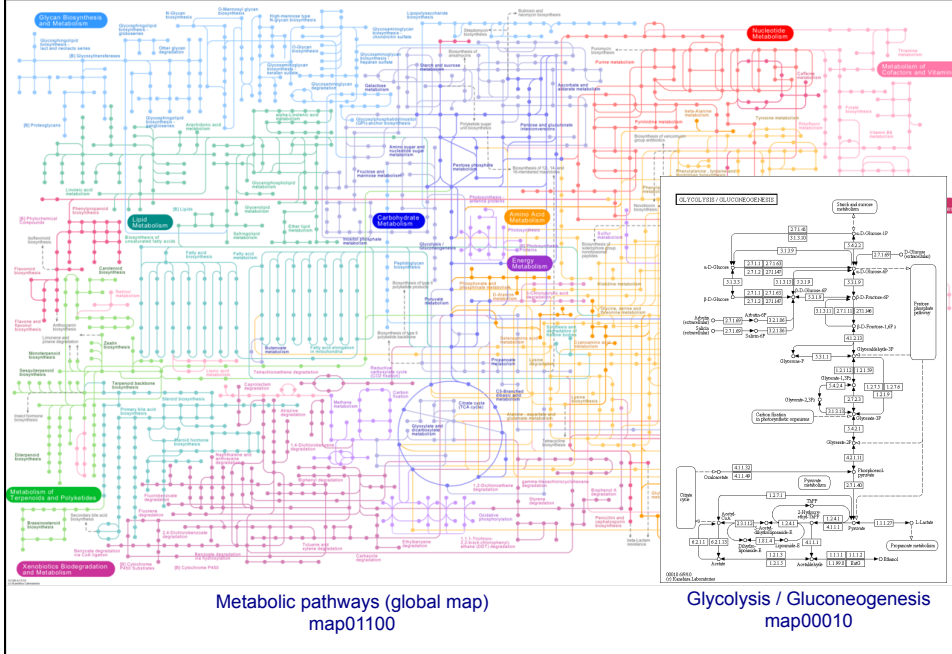
Organism-specific pathway – *Homo sapiens* (human)

Metabolic pathways (global map)
hsa01100

Glycolysis / Gluconeogenesis
hsa00010
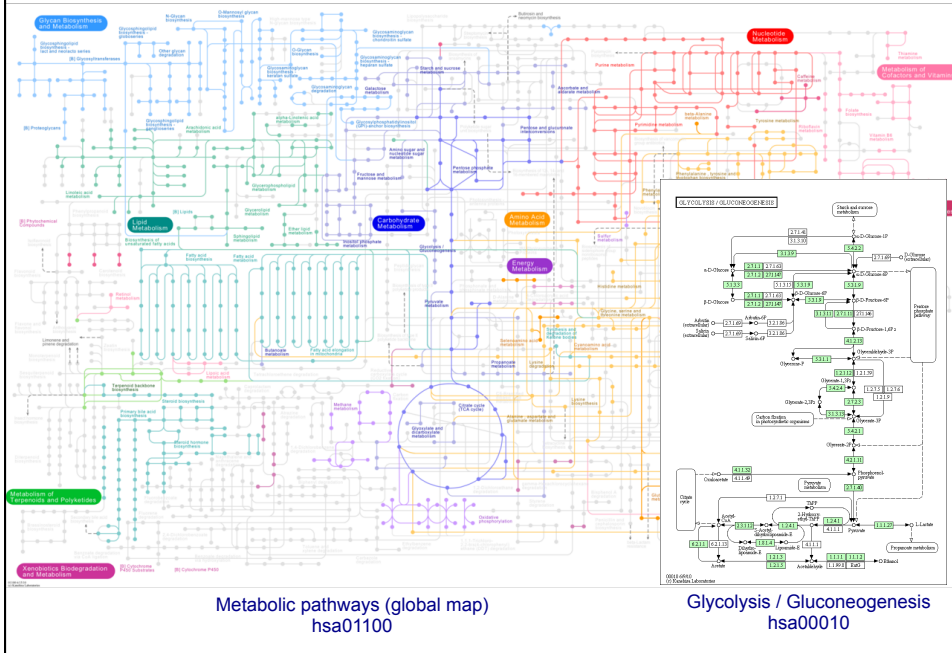
Organism-specific pathway - *Arabidopsis thaliana*

Metabolic pathways (global map)
ath01100

Glycolysis / Gluconeogenesis
ath00010
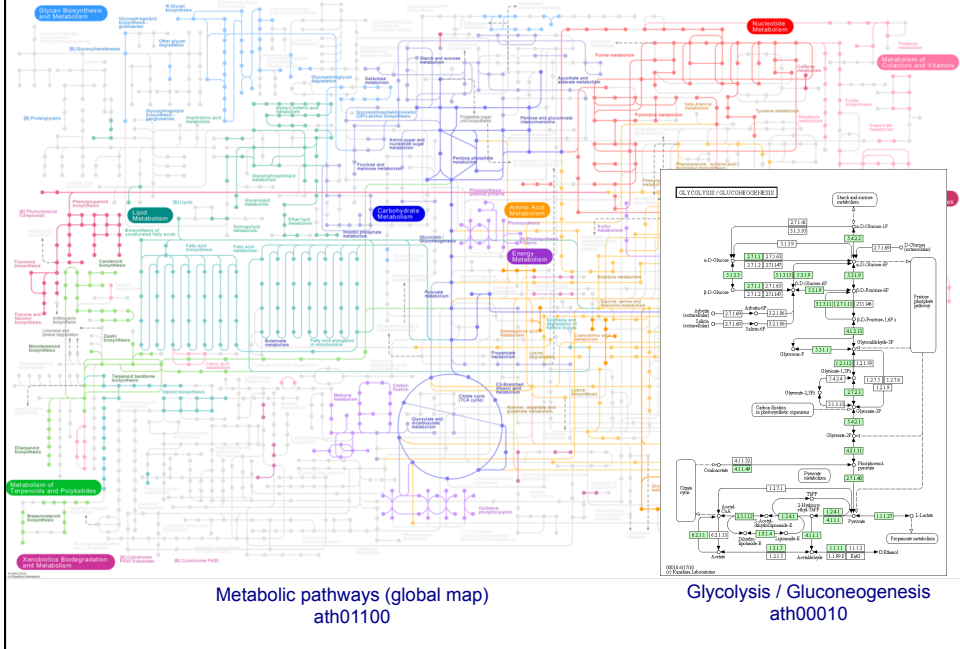
Metagenome analysis: Combined pathway of human and gut microbiome

hsa
T30003
hsa+T30003

Metabolic pathways (global map)

Terpenoid backbone biosynthesis
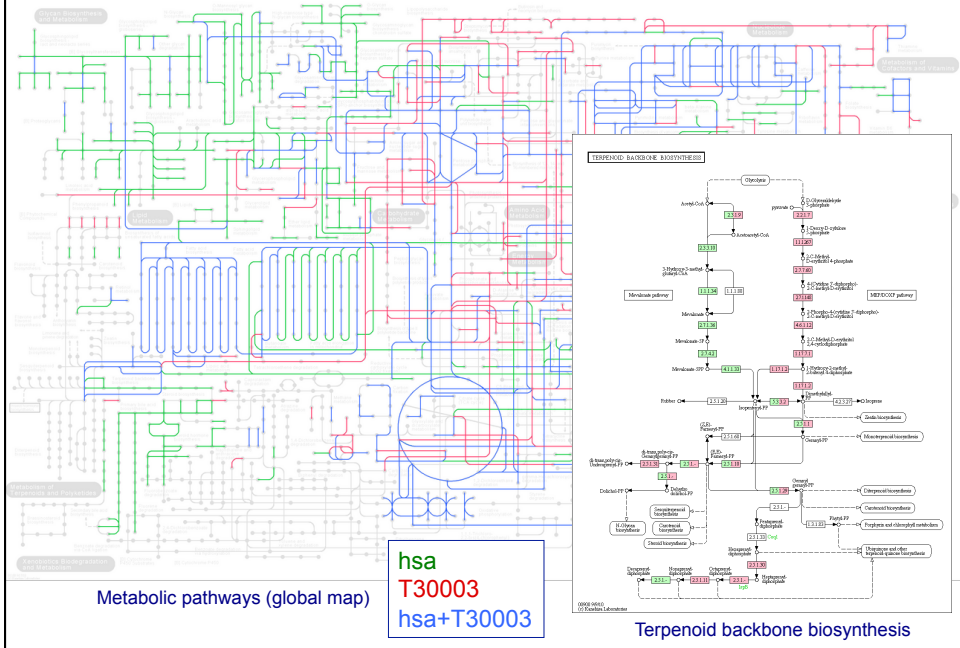
KEGG BRITE mapping
Genome-based extension of KO groups

Reference BRITE hierarchy **+** Human gene list with KO assignment

G Protein-Coupled Receptors

[Brite menu | Organism menu | Download htext]

Reference hierarchy (KO) ⬩ Go

▼ ▼ ▼ ☐ One-click mode

▼ **Class A. Rhodopsin family**
  ▼ Biogenic amine [Fig]
    ▼ Acetylcholine (muscarinic) [Table]
      K04129 CHRM1; muscarinic acetylcholine receptor M1
      K04131 CHRM3; muscarinic acetylcholine receptor M3
      K04133 CHRM5; muscarinic acetylcholine receptor M5
      K04130 CHRM2; muscarinic acetylcholine receptor M2
      K04132 CHRM4; muscarinic acetylcholine receptor M4
      K04134 CHRMN; muscarinic acetylcholine receptor
    ▶ Adrenaline [Table]
    ▼ Dopamine [Table]
      K04144 DRD1; dopamine receptor D1
      K05840 DRD5; dopamine receptor D5
      K04148 DRD1N, DopR; dopamine D1-like receptor
      K04145 DRD2; dopamine receptor D2
      K04146 DRD3; dopamine receptor D3
      K04147 DRD4; dopamine receptor D4
      K14049 DRD2N; dopamine D2-like receptor
    ▶ Histamine [Table]
    ▶ Serotonin [Table]
    ▶ Octopamine
    ▶ Trace amine
  ▼ Neuropeptide
    ▶ Tachykinin [Table]
    ▶ Neurotensin
    ▶ Ghrelin
    ▶ Galanin [Table]

br:ko04030

= Human BRITE hierarchy

G Protein-Coupled Receptors - Homo sapiens (human)

[Brite menu | Organism menu | Download htext]

Homo sapiens (human) ⬩ Go

▼ ▼ ▼ ☐ One-click mode

▼ **Class A. Rhodopsin family**
  ▼ Biogenic amine [Fig]
    ▼ Acetylcholine (muscarinic) [Table]
      1128 CHRM1, HM1, M1, M1R, MGC30125; cholinergic receptor, muscarinic 1
      1131 CHRM3, EGBRS, HM3; cholinergic receptor, muscarinic 3
      1133 CHRM5, HM5, MGC41838; cholinergic receptor, muscarinic 5
      1129 CHRM2, FLJ43243, HM2, MGC120006, MGC120007; cholinergic receptor, muscarinic 2
      1132 CHRM4, HM4, M4R; cholinergic receptor, muscarinic 4
    ▶ Adrenaline [Table]
    ▼ Dopamine [Table]
      1812 DRD1, DADR, DRD1A; dopamine receptor D1
      1816 DRD5, DBDR, DRD1B, DRD1L2, MGC10601; dopamine receptor D5
      1813 DRD2, D2DR, D2R; dopamine receptor D2
      1814 DRD3, D3DR, ETM1, FET1, MGC149204, MGC149205; dopamine receptor D3
      1815 DRD4, D4DR; dopamine receptor D4
    ▶ Histamine [Table]
    ▶ Serotonin [Table]
    ▶ Octopamine
    ▶ Trace amine
  ▼ Neuropeptide
    ▶ Tachykinin [Table]
    ▶ Neurotensin
    ▶ Ghrelin
    ▶ Galanin [Table]

br:hsa04030

Dopamine receptor orthologs

D1-like (Gs coupled)   D2-like (Gi coupled)

	D1	D5	D1N	D2	D3	D4	
Organism	K04144 (DRD1) [26]	K05840 (DRD5) [20]	K04148 (DRD1N) [24]	K04145 (DRD2) [23]	K04146 (DRD3) [16]	K04147 (DRD4) [15]	K14049 (DRD2N) [27]
hsa	1812	1816		1813	1814	1815	
ptr	744598	473340		451553	470885	449578	
pon	100447185	100444332		100442948	100454395	100462264	
mcc	697796	712862		694181		699697	
mmu	13488	13492		13489	13490	13491	
rno	24316	25195		24318	29238	25432	
cfa	489110	488824		403701	487984		
aml	100480500	100477987		100468533	100474054		
bta	281125	526221		281126	537043		
ssc	100144487 100627297	100524631		100624857	100513049	100512329	
ecb	100058846			100062195	100061160		
mdo	100026854	100017852		100026172	100014112		
oaa		100092962					
gga	427633 428961	428795		428252	770757	423016	
mgp	100548455	100544442		100303660	100549401	100546496	
tgu	100229034 100227653	100219893		100191007	100190794	100228686	
acs	100560309	100566713		100557445	100556183	100565088	
xla				378584			
xtr	100485463	100495174		100493063		100486482	
dre	568126 792634	563567 100001636 100002285		378719 282557 378718	282554	503564 503565 503566	
bfo	BRAFLDRAFT_175493			BRAFLDRAFT_82865			
spu	762451 581753 759572		579147				
dme			Dmel_CG9652				Dmel_CG33517
dpo			Dpse_GA21941				Dpse_GA14269
dan			Dana_GF18779				Dana_GF20457
der			Dere_GG21429				Dere_GG18014
dpe			Dper_GL23301				Dper_GL12926
dse			Dsec_GM25874				Dsec_GM22697
dsi			Dsim_GD20445				
dwi			Dwil_GK10900				Dwil_GK16340
dya			Dyak_GE10051				Dyak_GE17345
dgr			Dgri_GH19686				Dgri_GH24329
dmo			Dmoj_GI24543				Dmoj_GI14816
dvi			Dvir_GJ22610				Dvir_GJ19479
aga			AgaP_AGAP004613	AgaP_AGAP004034			AgaP_AGAP004453
aag			AaeL_AAEL003920 AaeL_AAEL003902				AaeL_AAEL014373
cqu							Cpij_CPIJ015294
ame			406111				408995
nvi			100122833				100118577
tca			660195				657486
api			100166029				100166556
phu			Phum_PHUM434420				Phum_PHUM286710
isc			IscW_ISCW001496				IscW_ISCW015254
cel			F15A8.5				K09G1.4 T14E8.3
cbr			CBG14463				CBG19317 CBG10841
pmy							Bm1_52070
tsp							Tsp_06111
smm							Smp_127310

deuterostome — vertebrate — mammal

insect

nematode

KEGG Module

M00095: Mevalonate pathway

KEGG C5 isoprenoid biosynthesis, mevalonate pathway

[Module menu | Organism menu | Module entry | KO/Gene/Compound list]

Entry	hsa_M00095
Name	C5 isoprenoid biosynthesis, mevalonate pathway
Definition	K00626 K01641 K00021 K00869 (K00938,K13273) K01597 K01823
	(Ortholog table) (Taxonomy)
Type	Pathway module
Organism	Homo sapiens (human)
Pathway	hsa00900
Reaction	R00238 C00024 -> C00332
	R01978 C00332 + C00024 -> C00356
	R02082 C00356 -> C00418
	R02245 C00418 -> C01107
	R03245 C01107 -> C01143
	R01121 C01143 -> C00129
	R01123 C00129 -> C00235

Homo sapiens (human) ▼ Go

K00626 | K01641 | K00021 | K00869 | K00938 K13273 | K01597 | K01823

C00024 — R00238 — C00332 — R01978 — C00356 — R02082
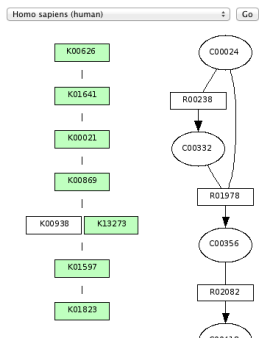
M00096: Non-mevalonate pathway

KEGG C5 isoprenoid biosynthesis, non-mevalonate pathway

[Module menu | Organism menu | Module entry | KO/Gene/Compound list]

Entry	eco_M00096
Name	C5 isoprenoid biosynthesis, non-mevalonate pathway
Definition	K01662 K00099 (K00991,K12506) K00919 (K01770,K12506) K03526 K03527 K01823
	(Ortholog table) (Taxonomy)
Type	Pathway module
Organism	Escherichia coli K-12 MG1655
Pathway	eco00900
Reaction	R05636 C00118 + C00022 -> C11437
	R05688 C11437 -> C11434
	R05633 C11434 -> C11435
	R05634 C11435 -> C11436
	R05637 C11436 -> C11453
	R08689 C11453 -> C11811
	R05884,R08209 C11811 -> C00129
	R01123 C00129 -> C00235

Escherichia coli K-12 MG1655 ▼ Go

K01662 | K00099 | K00991 K12506 | K00919 | K01770 K12506 | K03526 | K03527 | K01823

C00118 C00022 — R05636 — C11437 — R05688 — C11434 — R05633 — C11435
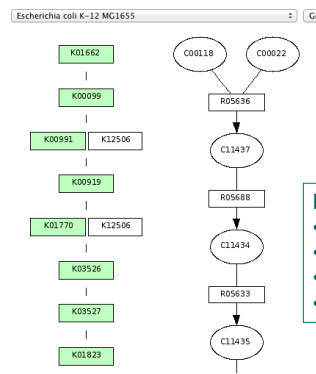
KEGG MODULE
- Pathway module
- Structural complex
- Functional set
- Signature module

---

Extension of KEGG Object Identifiers

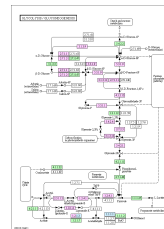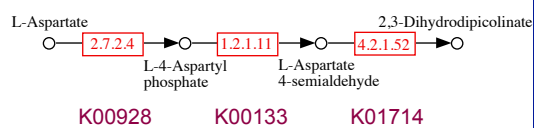| Molecular network	Manually created		Computationally generated		
	Reference		Org-specific	Disease/drug	Other
Pathway (metabolic)	map00010	ko00010 ec00010 rn00010	hsa00010	hsadd00010	
(non-metabolic)	map02010	ko02010	hsa02010	hsadd02010	
Brite (gene/protein)	ko02000		hsa02000	hsa02000_dd	
(non-gene/protein)	br08303				br08303_target
Module	M00001		hsa_M00001		
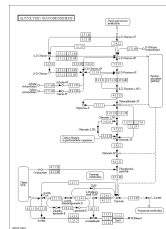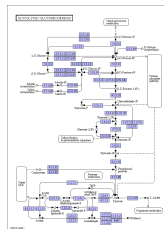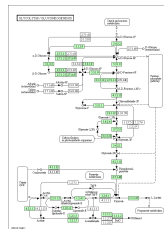
map   ko   hsa   hsadd

KEGG LIGAND:
Integration of genomics and chemistry

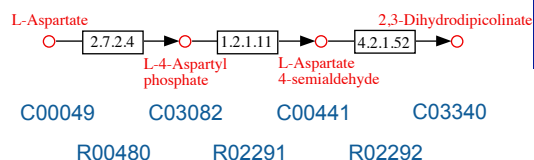## Dual Aspect of Metabolic Network

1. Genomic information network

Network of enzyme genes (or enzymes)

L-Aspartate   2,3-Dihydrodipicolinate

○ — [2.7.2.4] ► ○ — [1.2.1.11] ► ○ — [4.2.1.52] ► ○

L-4-Aspartyl   L-Aspartate
phosphate   4-semialdehyde

K00928   K00133   K01714

- Genes and proteins are identified by K numbers
- EC numbers are displayed as node names, but they are not used as identifiers in KEGG
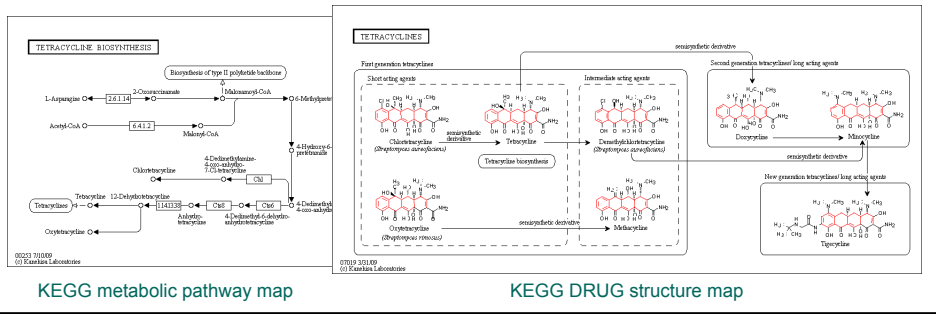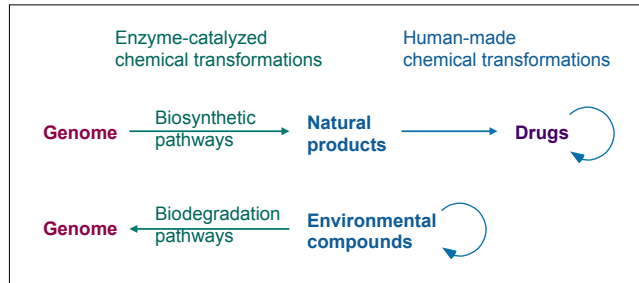
2. Chemical information network

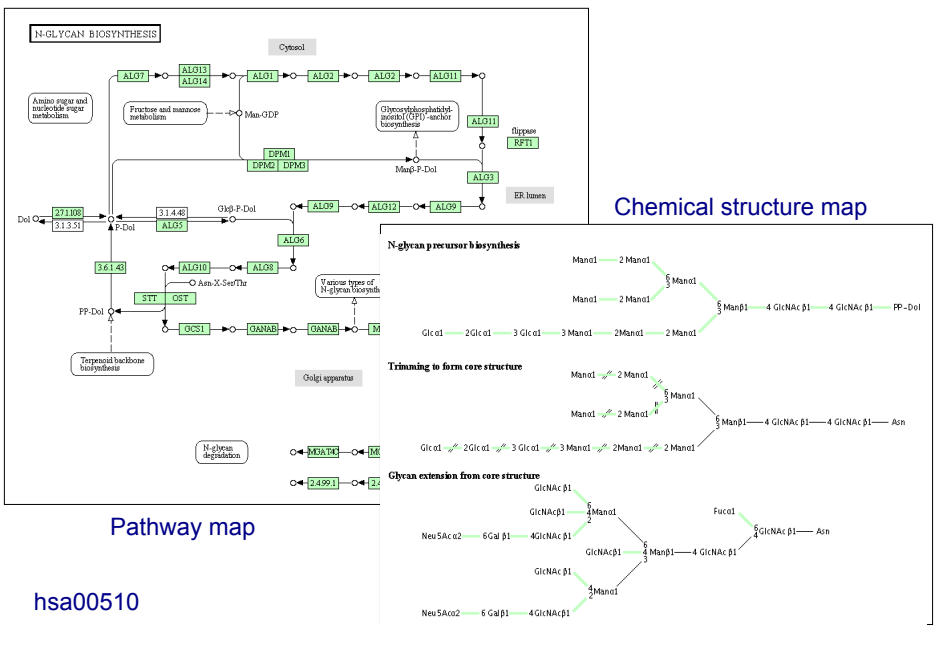Network of small molecules
Chemical structure transformation network

L-Aspartate   2,3-Dihydrodipicolinate

○ — [2.7.2.4] ► ○ — [1.2.1.11] ► ○ — [4.2.1.52] ► ○

L-4-Aspartyl   L-Aspartate
phosphate   4-semialdehyde

C00049   C03082   C00441   C03340

R00480   R02291   R02292

- Compounds are identified by C numbers
- Reactions are identified by R numbers

Chemical structure transformation networks
Linking genomes to natural products, drugs, and xenobiotics

Enzyme-catalyzed chemical transformations

Human-made chemical transformations

Genome — Biosynthetic pathways → **Natural products** → **Drugs**

Genome ← Biodegradation pathways — **Environmental compounds**

KEGG metabolic pathway map

KEGG DRUG structure map

N-Glycan biosynthesis - Homo sapiens (human)

Chemical structure map

N-glycan precursor biosynthesis

Trimming to form core structure

Glycan extension from core structure

Pathway map

hsa00510

Reaction Class defined by RDM patterns

KEGG REACTION Database

Reaction

acetyl-CoA + L-glutamate <=> CoA + N-acetyl-L-glutamate

C00024 + C00025 <=> C00010 + C00624   R00259

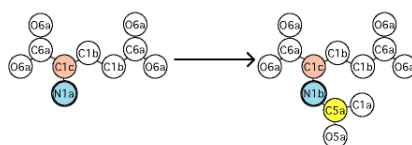**KEGG RPAIR Database**

L-glutamate   N-acetyl-L-glutamate
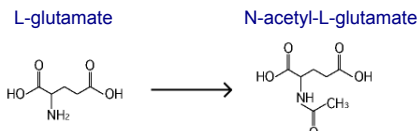
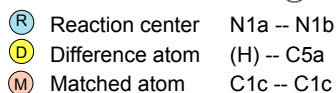Reactant pair   RP04458

Reactant pair in the KEGG atom type representation

KEGG RCLASS Database

KEGG atom changes for R, D, M atoms (RDM patterns)

R Reaction center   N1a -- N1b
D Difference atom   (H) -- C5a   RC00064
M Matched atom   C1c -- C1c

http://www.genome.jp/kegg/reaction/

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
	C7x	ring-C(=O)-O-ring
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	

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	

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(=0)-OH
Ester	O7a	R-C(=O)-O-R
	O7x	ring-C(=O)-O-ring
Undefined O	O0	

Phosphorus 2 types

Attatched to other elements	P1a	P-R
Attatched to oxygen	P1b	P-O

Other elements 2 types

Halogens	X	F, Cl, Br, I
Others	Z	

Correlation of genomes and chemistry

KO Module M00010 **RC Module RM001**

CITRATE CYCLE (TCA CYCLE)

CITRATE CYCLE (TCA CYCLE)
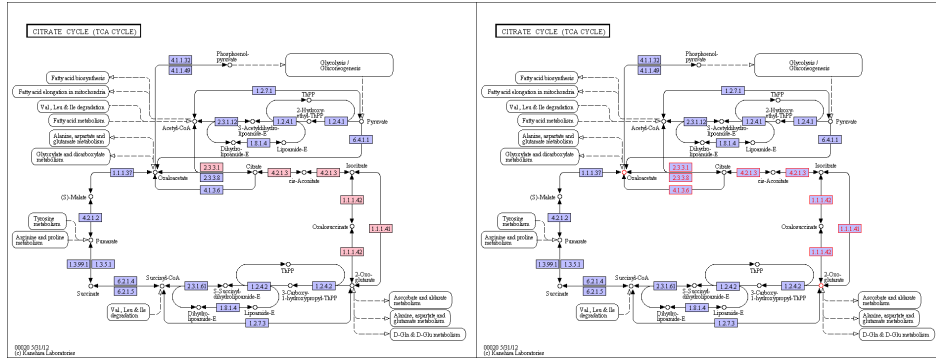
K01647 (K01681,K01682) (K00031,K00030)   RC00067 RC00498 RC00618 (RC00084+RC00626,RC00114)

GENES database   REACTION database
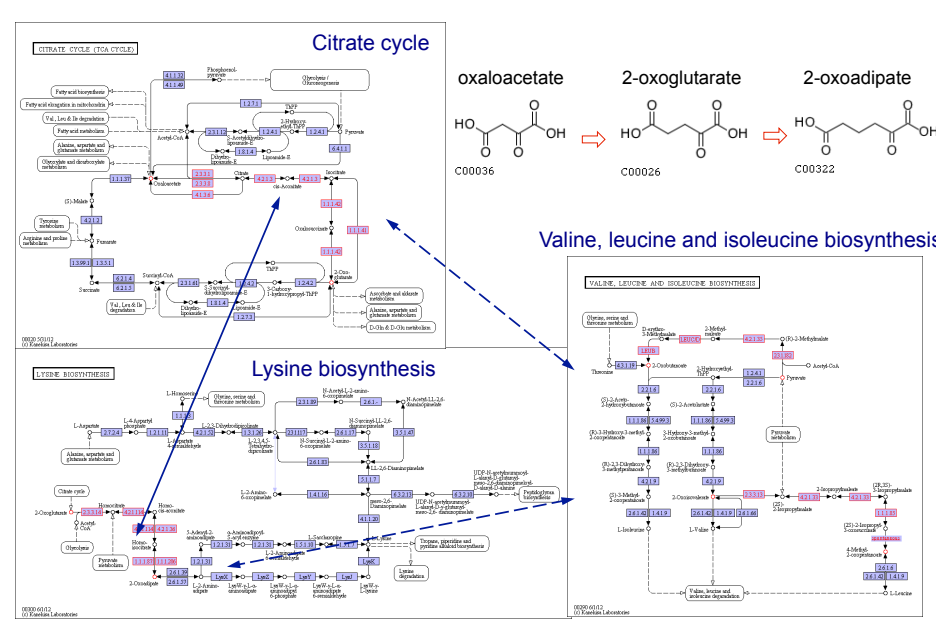⇩   ⇩
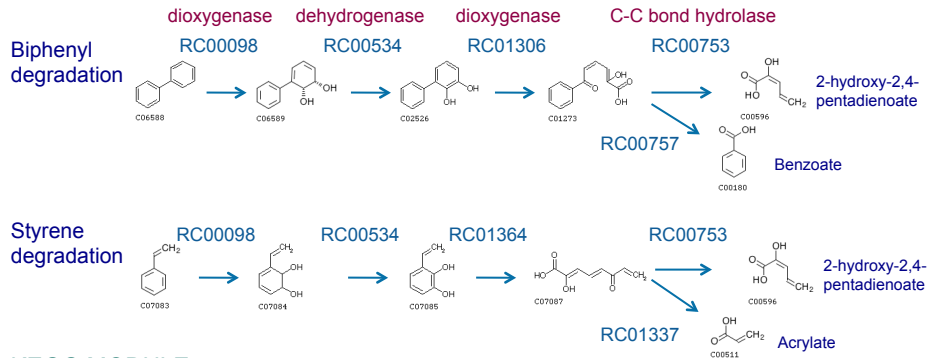KO system   KEGG PATHWAY   RC system
⇩   ⇩
KO module   ⟷   RC module
(KEGG MODULE)   (Reaction sequence motif)

Reaction Sequence Motif
RM001 Carboxylic acid chain elongation

Citrate cycle

CITRATE CYCLE (TCA CYCLE)

oxaloacetate   2-oxoglutarate   2-oxoadipate

C00036   C00026   C00322

Valine, leucine and isoleucine biosynthesis

VALINE, LEUCINE AND ISOLEUCINE BIOSYNTHESIS

LYSINE BIOSYNTHESIS

Lysine biosynthesis

Biodegradation Reaction Sequence Motif
RM004 Meta-cleavage of aromatic ring, type 1

dioxygenase   dehydrogenase   dioxygenase   C-C bond hydrolase

Biphenyl degradation

RC00098 → RC00534 → RC01306 → RC00753 → 2-hydroxy-2,4-pentadienoate

RC00757 → Benzoate

Styrene degradation

RC00098 → RC00534 → RC01364 → RC00753 → 2-hydroxy-2,4-pentadienoate

RC01337 → Acrylate

KEGG MODULE
M00543 Biphenyl degradation

(K08689+K15750+K05710+K00529) K08690 K00462 K10222

bxe Burkholderia xenovorans
pna Polaromonas naphthalenivorans
rha Rhodococcus sp. RHA1

bphA	bphB	bphC	bphD
K08689	K08690	K00462	K10222
R05263	R05239	R03462	R02606
RC00098	RC00534	RC01306	RC00753
			RC00757

---

bxe00621 Dioxin degradation pathway
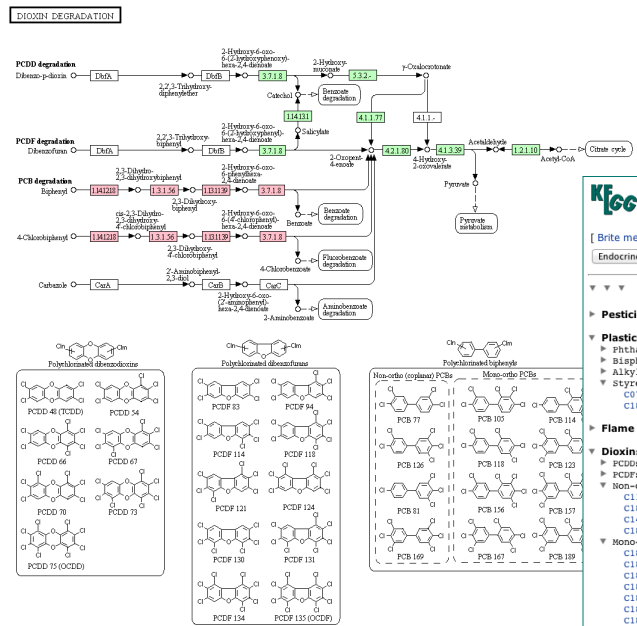
br08006

KEGG Endocrine Disrupting Compounds

[Brite menu | Download htext]

Endocrine disrupting compounds — Go

One-click mode

▶ Pesticides and herbicides
▼ Plasticizers and plastics
  ▶ Phthalates
  ▶ Bisphenols
  ▶ Alkylphenols
  ▼ Styrenes
    C07083 Styrene
    C18148 Octachlorostyrene
▶ Flame retardants
▼ Dioxins and dioxin-like compounds
  ▶ PCDDs
  ▶ PCDFs
  ▼ Non-ortho (coplanar) PCBs
    C11057 3,4,3',4'-Tetrachlorobiphenyl (PCB 77)
    C18114 3,4,4',5-Tetrachlorobiphenyl (PDB 81)
    C14573 3,3',4,4',5-Pentachlorobiphenyl (PCB 126)
    C18115 3,4,5,3',4'-Hexachlorobiphenyl (PCB 169)
  ▼ Mono-ortho PCBs
    C18116 2,3,4,3',4'-Pentachlorobiphenyl (PCB 105)
    C18117 2,3,4,4',5-Pentachlorobiphenyl (PCB 114)
    C18118 2,3',4,4',5-Pentachlorobiphenyl (PCB 118)
    C18119 2,3',4,4',5'-Pentachlorobiphenyl (PCB 123)
    C18120 2,3,3',4,4'-Hexachlorobiphenyl (PCB 156)
    C18121 2,3,4,4',5-Hexachlorobiphenyl (PCB 157)
    C18122 2,3',4,4',5,5'-Hexachlorobiphenyl (PCB 167)
    C18123 2,3,3',4,4',5,5'-Heptachlorobiphenyl (PCB 189)

Biosynthetic Reaction Sequence Motif
RM022 Hydroxylation and methylation reaction sequence
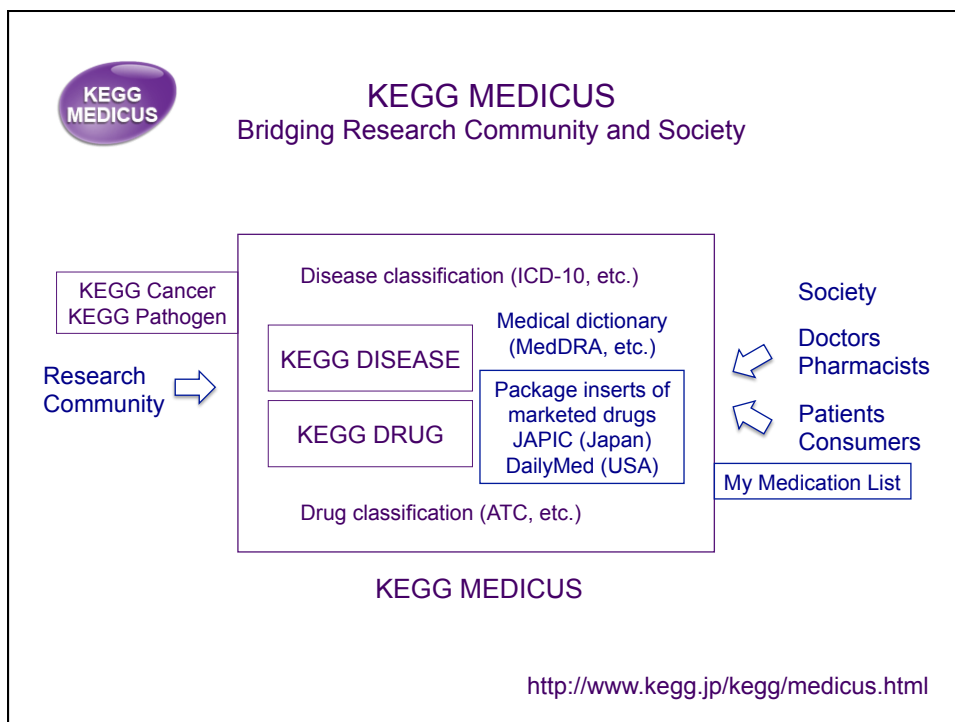
Phenylpropanoid biosynthesis

monooxygenase	methyltransferase	monooxygenase	methyltransferase
RC00046	RC00392	RC00046	RC00392

4-Coumarate	Caffeate	Ferulate	5-Hydroxyferulate	Sinapate
C00811	C01197	C01494	C05619	C00482
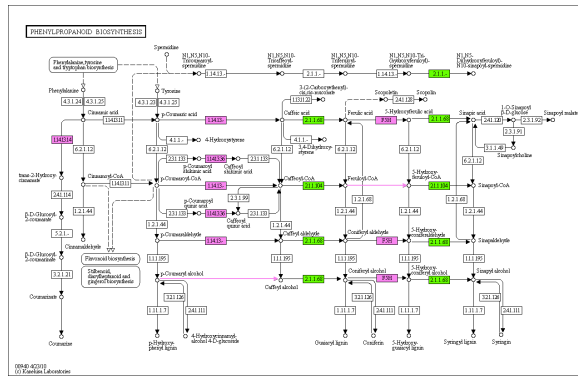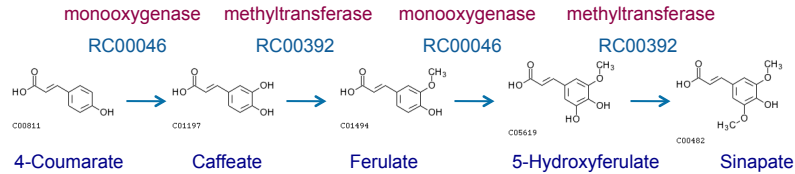
RC00046
RC00392

Isoquinoline alkaloid biosynthesis

D09307 Colchicum seed; 犬泊夫藍
Colchicine

Emetine
D03868 Ipecac; 吐根

Monooxygenase
Methyltransferase

Papaverine
D03444 Opium; 阿片

D06725 Calumba   D06686 Corydalis tuber; 延胡索
Corydaline

Magnoflorine
Berberine   D00092 Coptis rhizome; 黄連
D06689 Phellodendron bark; 黄柏

D00092 Coptis rhizome; 黄連
D06689 Phellodendron bark; 黄柏
D06720 Magnolia bark; 厚朴
D06785 Sinomenium stem; 防已
D06808 Epimedium herb; 淫羊藿

D06686 Corydalis tuber; 延胡索

Codeine

D03444 Opium; 阿片
Morphine

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

KEGG MEDICUS:
Bringing genomic revolution to society

KEGG MEDICUS
Bridging Research Community and Society

KEGG Cancer
KEGG Pathogen

Disease classification (ICD-10, etc.)

Medical dictionary
(MedDRA, etc.)

KEGG DISEASE

Research
Community

KEGG DRUG

Package inserts of
marketed drugs
JAPIC (Japan)
DailyMed (USA)

Drug classification (ATC, etc.)

Society

Doctors
Pharmacists

Patients
Consumers

My Medication List

KEGG MEDICUS

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

Diseases and Drugs Viewed As Molecular Network Information

- Diseases are viewed as perturbed states of the molecular network.
- Genetic factors, environmental factors, and other causative agents are perturbants to the molecular network.
- Drugs are different types of perturbants to the molecular network.

Network-Disease Associations

Genetic perturbations (germline/somatic mutations, etc.) → Molecular network system

Environmental perturbations (incl. pathogens, microbiome) → Molecular network system

Diagnostic markers → Molecular network system

Therapeutic drugs → Molecular network system

Genomic biomarkers → Molecular network system

Molecular network system → Disease
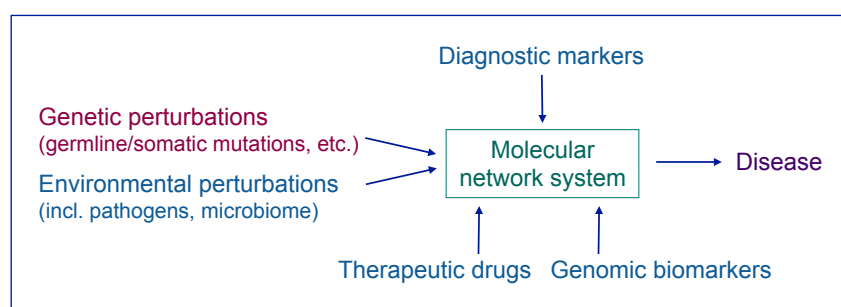
Contents of the KEGG DRUG database

- D number identifier; Generic name; Representative trade names
- Links to FDA approved drug labels in DailyMed (and Japanese labels in JAPIC)
- Chemical structure or component
- Target molecules in the context of KEGG pathway maps
- Drug metabolizing enzymes and transporters
- Other interacting molecules including genomic biomarkers, CYP inducers/ inhibitors, etc.
- Adverse drug-drug interaction search
- Efficacy and links to KEGG DISEASE
- Drug classification in BRITE hierarchies such as ATC codes
- History of drug development represented as a KEGG DRUG structure map
- Links to outside databases

31

Disease gene and drug target as part of KEGG pathway

Chronic myeloid leukemia (H00004)

DISEASE: H00004

Entry	H00004 Disease
Name	Chronic myeloid leukemia (CML)
Description	Chronic myelogenous leukemia (CML) originates in a pluripotent hematopoetic stem cell of the bone marrow and is characterized by greatly increased numbers of granulocytes in the blood. Myeloid and other haematopoetic 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-oncogene on chromosome 9 and the BCR (breakpoint cluster region) gene on chromosome 22. The BCR/ABL fusion gene encodes p210 BCR/ABL, an oncoprotein, which, unlike the normal p145 c-Abl, has constitutive tyrosine kinase activity and is predominantly localized in the cytoplasm. While fusion of c-ABL and BCR is believed to be the primary cause of the chronic phase of CML, progression to blast crisis requires other molecular changes. Common secondary abnormalities include mutations in TP53, RB, and p16/INK4A, or overexpression of genes such as EVI1. Additional chromosome translocations are also observed, such as t(3;21) (q26;q22), which generates AML1-EVI1.
Category	Cancer
	BRITE hierarchy
Pathway	hsa05220 Chronic myeloid leukemia
Gene	BCR-ABL (translocation) [HSA:613 25] [KO:K08878 K06619] EVI1 (overexpression) [HSA:2122] [KO:K04462] AML1 (translocation) [HSA:861] [KO:K08367] p16/INK4A (mutation) [HSA:1027] [KO:K06621] p53 (mutation) [HSA:7157] [KO:K04451] RB1 (mutation) [HSA:5925] [KO:K06618]
Carcinogen	1,3-Butadiene [CPD:C16450] Rubber industry
Marker	BCR-ABL (translocation) [HSA:613 25] WT1 [HSA:7490]

Disease gene

CHRONIC MYELOID LEUKEMIA

Gleevec (D01441)

DRUG: D01441   Help

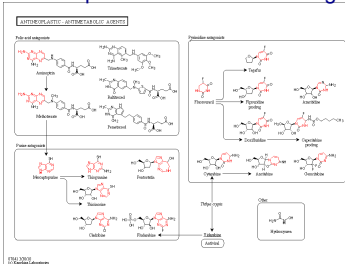|	
Entry	D01441 Drug
Name	Imatinib mesilate (JAN); Imatinib mesylate (USAN); Gleevec (TN); Glivec (TN)
Products	GLEEVEC (Novartis Pharma), GLEEVEC (Physicians Total Care)
Formula	C29H31N7O. CH4SO3
Exact mass	589.2471
Mol weight	589.7084
Structure	D01441 Mol file KCF file DB search Jmol KegDraw
Target	Bcr-Abl tyrosine kinase inhibitor [HSA:25] [KO:K06619]; PDGFRA tyrosine kinase inhibitor [HSA:5156] [KO:K04363]; PDGFRB tyrosine kinase inhibitor [HSA:5159] [KO:K05089]; c-KIT tyrosine kinase inhibitor [HSA:3815] [KO:K05091]
Pathway	hsa04010 MAPK signaling pathway hsa04020 Calcium signaling pathway hsa04060 Cytokine-cytokine receptor interaction hsa04640 Hematopoietic cell lineage hsa05200 Pathways in cancer hsa05220 Chronic myeloid leukemia
Metabolism	Enzyme: CYP3A4 [HSA:1576], CYP3A5 [HSA:1577], CYP2D6 [HSA:1565], CYP2C9

Drug target
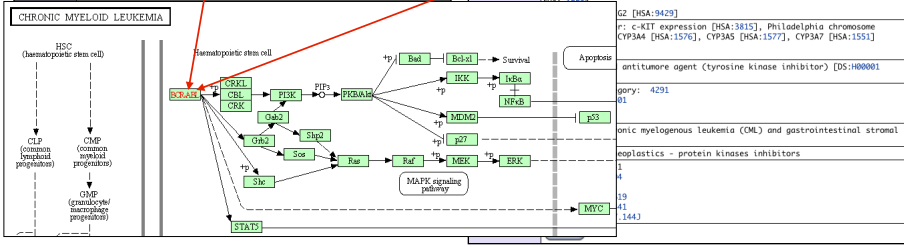
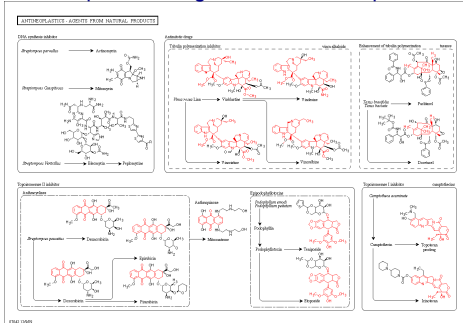# KEGG DRUG strucure map: History of drug development

Antineoplastics – Protein kinase inhibitors
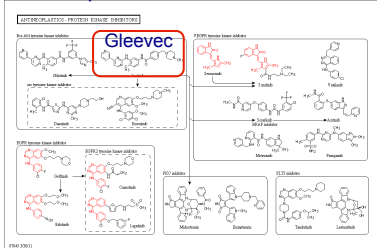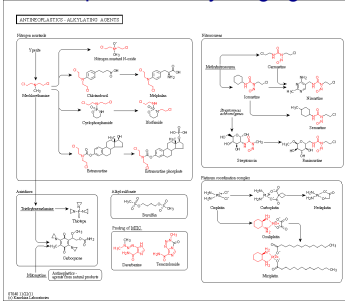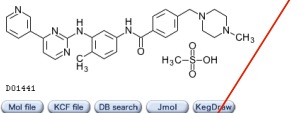
Gleevec

Antineoplastics – Antimetabolic agents

Antineoplastics – Agents from natural products

Antineoplastics – Alkylating agents

Drug classifications in BRITE hierarchies

Drug Interaction Search

Many interactions involving drug metabolizing enzyme CYP3A4

- Drug-drug interaction data associated with contraindications and warnings in package inserts
- Complementation and prediction by using drug hierarchy

- Non-drug data such as grape fruit juice will also be included

KEGG pathway map for Alzheimer's disease (hsa05010)

Four known disease genes
APP (amyloid beta protein)
APOE (apolipoprotein E)
PSEN1 (presenilin 1)
PSEN2 (presenilin 2)

Two therapeutic drugs
NMDA receptor antagonist
  Memantine hydrochloride
  (MEMARY)
AChE inhibitor (not in the map)
  Donepezil hydrochloride
  (ARICEPT)
  Galantamine hydrobromide
  (REMINYL)
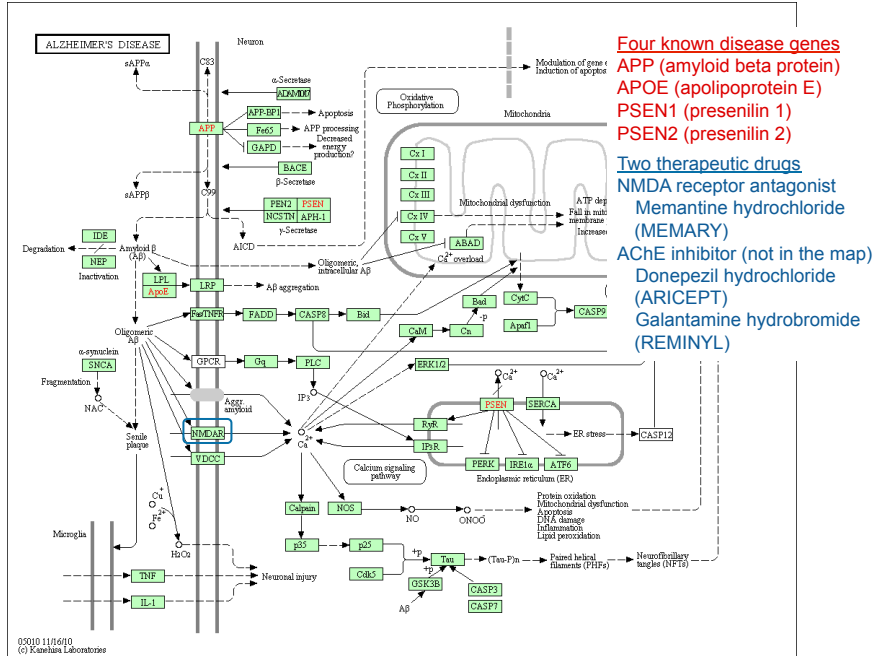
Alzheimer's disease + disease genes and drug targets (hsadd05010)

Causative genes of other neurodegenerative diseases

Leber optic atrophy

Parkinson's disease
Lewy body dementia

Spinocerebellar ataxia

Amyotrophic lateral sclerosis
Progressive supranuclear palsy
Pick's disease

Related neurodegenerative diseases
▼ Nervous system diseases
  ▼ Neurodegenerative diseases
    H00056 Alzheimer's disease [PATH:hsa05010]
    H00066 Lewy body dementia
    H00057 Parkinson's disease [PATH:hsa05012]
    H00058 Amyotrophic lateral sclerosis [PATH:hsa05014]
    H00059 Huntington's disease [PATH:hsa05016]
    H00060 Dentatorubropallidoluysian atrophy (DRPLA)
    H00062 Kennedy's disease
    H00063 Spinocerebellar ataxia [PATH:hsa04730]
    H00061 Prion diseases [PATH:hsa05020]
    H00064 Ataxia telangiectasia
    H00065 Alexander disease
    H00075 Refsum disease [PATH:hsa04146]
    H00067 Friedreich ataxia
    H00068 Leber optic atrophy [PATH:hsa00190]
    H00076 Cockayne syndrome
    H00077 Progressive supranuclear palsy
    H00078 Pick's disease
    H00074 Canavan disease [PATH:hsa00250]

Parkinson's disease + disease genes and drug targets (hsadd05012)

Adverse drug interactions with doperminergic agents

Antiobesity drug
DAT (SLC6A3) inhibitor

Antihypertensive
VMAT (SLC18A1,2) inhibitor
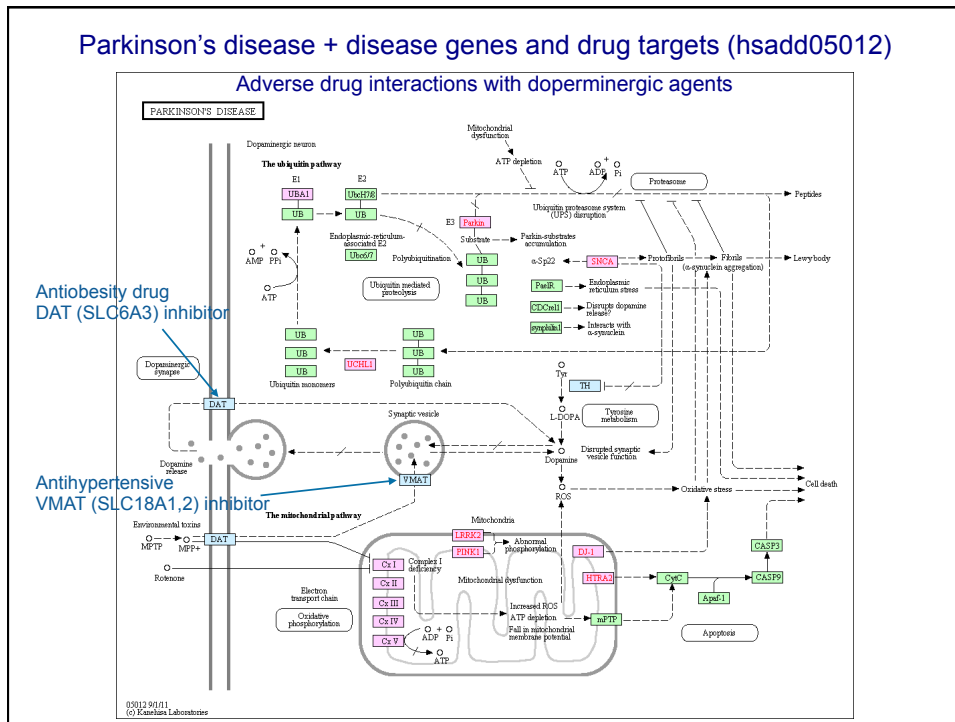
KEGG CANCER Resource

- KEGG DISEASE entries for 55 different cancers
- KEGG pathway maps for 14 specific cancers
- An overview map for signaling pathways (Pathways in cancer)
- An overview map for transcription networks (Transcriptional misregulation in cancer)
- Drug information in BRITE hierarchies (Antineoplastics and other drug classifications)
- Drug information in KEGG DRUG structure maps (Chronology: Antineoplastics)
- Environmental factor information in BRITE hierarchies (Carcinogens, etc.)
- Additional data from outside resouces:

Somatic mutations in cancer genomes (from COSMIC)
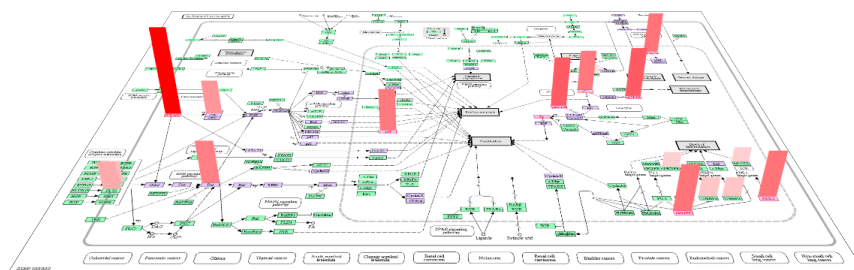
Chronic myeloid leukemia in the overview map

Summary and Supplements

KEGG: Reference knowledge base for use in practice and in society

- Capturing experimental knowledge on molecular systems in both normal and perturbed (disease) states

- Capturing knowledge on genetic and environmental factors of diseases, as well as drugs and health-related substances as perturbants to molecular systems

- Generalizing knowledge on genes and proteins by KEGG Orthology

- Generalizing knowledge on biochemical reactions by reaction class

KEGG PATHWAY
KEGG BRITE
KEGG MODULE
KEGG DISEASE
KEGG DRUG
KEGG ENVIRON
KEGG ORTHOLOGY
KEGG GENES
KEGG COMPOUND
KEGG REACTION
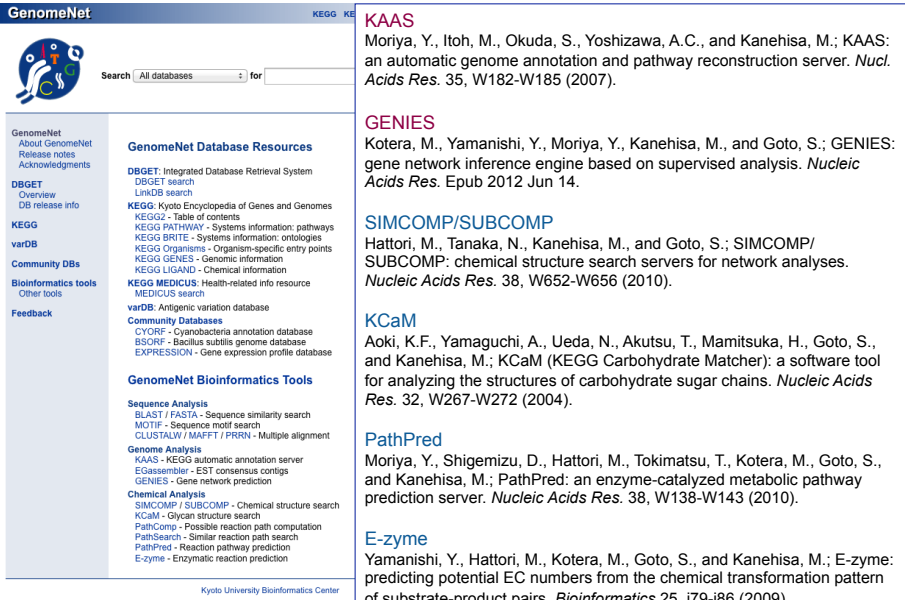KEGG RCLASS

For research community

- Helping to bring research results into practical applications, such as personalized medicine and drug discovery

For society

- Helping to understand scientific basis of diseases and drugs with practical tools, such as My Medication List

Computational tools at GenomeNet
http://www.genome.jp/

KAAS
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).

GENIES
Kotera, M., Yamanishi, Y., Moriya, Y., Kanehisa, M., and Goto, S.; GENIES: gene network inference engine based on supervised analysis. *Nucleic Acids Res.* Epub 2012 Jun 14.

SIMCOMP/SUBCOMP
Hattori, M., Tanaka, N., Kanehisa, M., and Goto, S.; SIMCOMP/ SUBCOMP: chemical structure search servers for network analyses. *Nucleic Acids Res.* 38, W652-W656 (2010).

KCaM
Aoki, K.F., Yamaguchi, A., Ueda, N., Akutsu, T., Mamitsuka, H., Goto, S., and Kanehisa, M.; KCaM (KEGG Carbohydrate Matcher): a software tool for analyzing the structures of carbohydrate sugar chains. *Nucleic Acids Res.* 32, W267-W272 (2004).

PathPred
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.* 38, W138-W143 (2010).

E-zyme
Yamanishi, Y., Hattori, M., Kotera, M., Goto, S., and Kanehisa, M.; E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs. *Bioinformatics* 25, i79-i86 (2009).

Some papers on chemical bioinformatics

Small molecules
- 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).
- 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).
- 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).

Glycans
- Hashimoto, K., Goto, S., Kawano, S., Aoki-Kinoshita, K.F., Ueda, N., Hamajima, M., Kawasaki, T., and Kanehisa, M.; KEGG as a glycome informatics resource. *Glycobiology* 16, 63R-70R (2006).
- Kawano, S., Hashimoto, K., Miyama, T., Goto, S., and Kanehisa, M.; Prediction of glycan structures from gene expression data based on glycosyltransferase reactions. *Bioinformatics* 21, 3976-3982 (2005).
- Hashimoto, K., Tokimatsu, T., Kawano, S., Yoshizawa, A.C., Okuda, S., Goto, S., and Kanehisa, M.; Comprehensive analysis of glycosyltransferases in eukaryotic genomes for structural and functional characterization of glycans. *Carbohydrate Res.* 344, 881-887 (2009).

Polyketides & nonribosomal peptides
- Minowa, Y., Araki, M., and Kanehisa, M.; Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. *J. Mol. Biol.* 368, 1500-1517 (2007).

Fatty acids
- Hashimoto, K., Yoshizawa, A.C., Okuda, S., Kuma, K., Goto, S., and Kanehisa, M.; The repertoire of desaturases and elongases reveals fatty acid variations in 56 eukaryotic genomes. *J. Lipid Res.* 49, 183-191 (2008).

Drugs
- Shigemizu, D., Araki, M., Okuda, S., Goto, S., and Kanehisa, M.; Extraction and analysis of chemical modification patterns in drug development. *J. Chem. Inf. Model.* 49, 1122-1129 (2009).
- Takarabe, M., Shigemizu, D., Kotera, M., Goto, S., and Kanehisa, M.; Network-based analysis and characterization of adverse drug-drug interactions. *J. Chem. Inf. Model.* 51, 2977-2985 (2011).