

パスウェイデータベースの紹介と 疾患研究への応用

統合データベース講習会 AJACS河内
2017/8/24 10:40-12:10

九州大学生体防御医学研究所システムコホート学分野
山西 芳裕
(yamanishi@bioreg.kyushu-u.ac.jp)

本日の内容

- パスウェイデータベース
 - パスウェイデータベースの概要
 - BioCyc の紹介・使い方
 - KEGG PATHWAY の紹介・使い方
 - 疾患研究への応用

はじめに

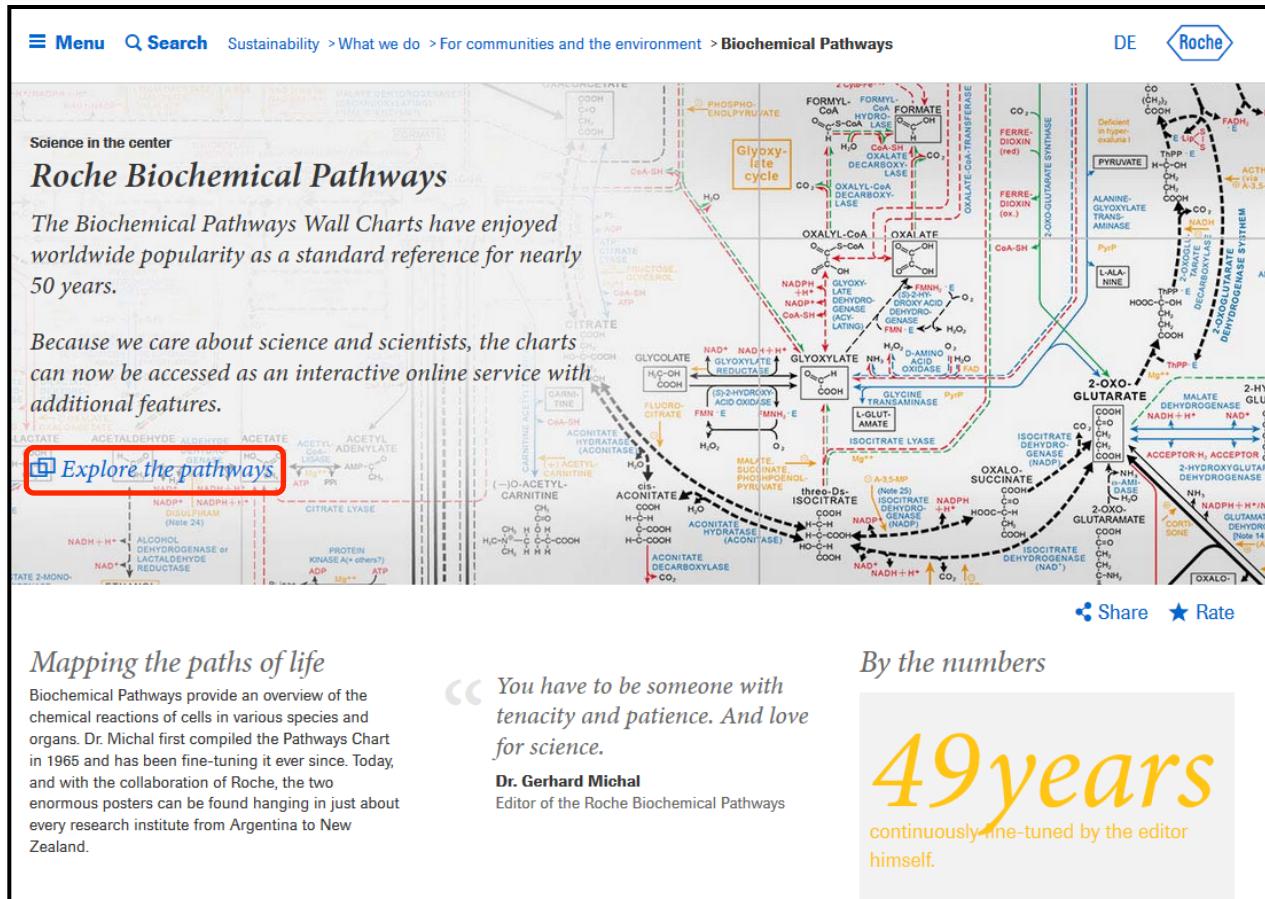
- ・ パスウェイデータベースに関しては、過去の AJACSで繰り返し講習がされています。過去の資料もぜひご参照ください。
- ・ 今回の資料は、過去に五斗進先生(DBCLS)、守屋勇樹先生(DBCLS)、時松敏明先生(DDBJ)が作成した資料をベースにしています。

現在の課題

- オミックスデータが得られるようになり、個々の分子の一覧は明らかになってきた。
- 生命システムは多数の分子が協調して働くことによって成り立っている。
- グローバルな分子間相互作用ネットワーク(パスウェイ)の全貌を明らかにしたい。

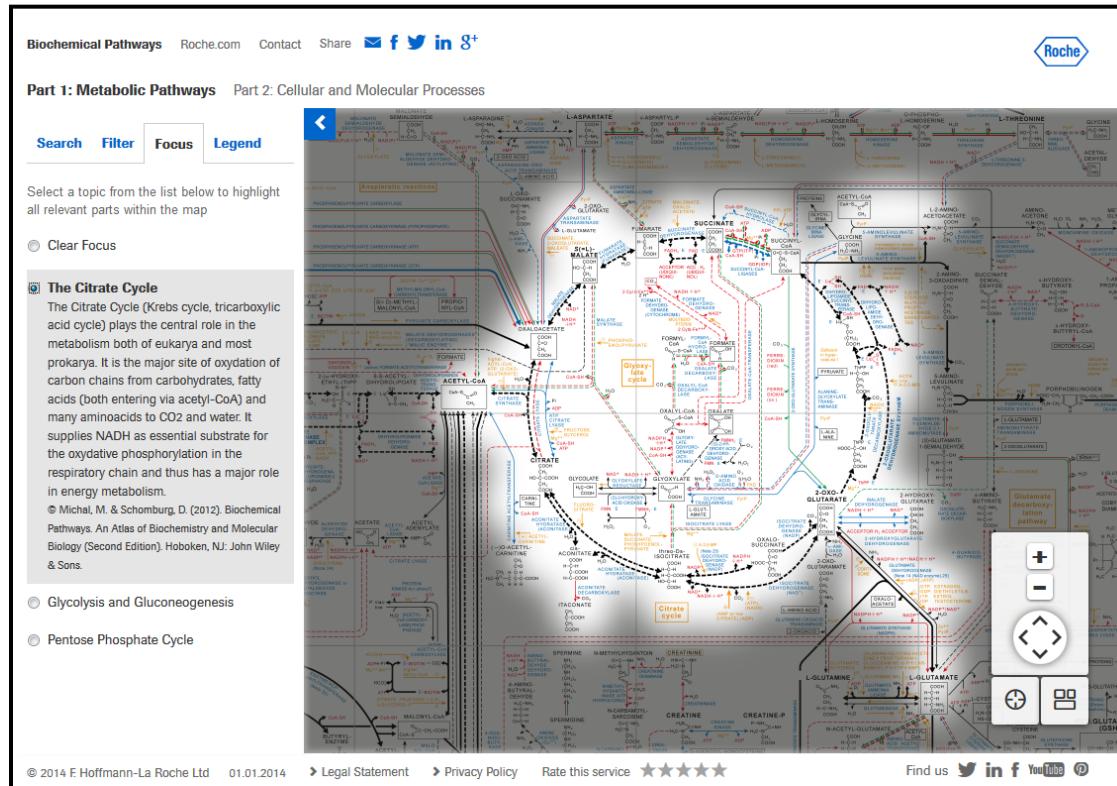
パスウェイデータベースとは

パスウェイとは、生体内での遺伝子、タンパク質、その他化合物等の分子間相互作用を“経路”として表現したものです。パスウェイを集積して可視化、電子化したものがパスウェイデータベースと呼ばれています。歴史的には、代謝経路の表現から始まりました。



ベーリンガーマンハイム社(現ロシュ社)の代謝マップ(1965~)
(紙媒体で出版、のちに電子化)

パスウェイデータベースとは



計算機上で表現することで、様々な可視化が行えるようになり、パスウェイ全体を俯瞰したり、一部の相互作用に注目したり、生命現象を理解しやすくなります
また、データベース化することで網羅的に扱えるようになり、コンピューターでの利用が可能になります

- ・ゲノムアノテーションや種間比較、進化解析
- ・遺伝子発現などのエンリッチメント解析
- ・モデル化、シミュレーション、予測

今回の講習は、パスウェイデータベースへのマッピングを中心に実習します

いろいろなパスウェイデータベース

Pathguide (<http://www.pathguide.org/>)

- パスウェイリソースのリスト約 550 (2013)
 - Availability: 有料か無料か
 - Standards: 標準データ形式(BioPAX, SBML等)に準拠しているかどうか

The screenshot shows the homepage of Pathguide. At the top, there's a navigation bar with links to Home, BioPAX, cBio, and MSKCC. Below the header, the title "Pathguide» the pathway resource list" is displayed. On the left, there's a sidebar with a "Navigation" section containing links to various pathway types like Protein-Protein Interactions, Metabolic Pathways, and Pathway Diagrams. Below that is a "Search" section with dropdown menus for Organisms, Availability, and Standards, along with "Reset" and "Search" buttons. Further down are sections for Analysis (Statistics, Database Interactions), and Contact (with a note that comments, questions, and suggestions are always welcome). The main content area is titled "Complete Listing of All Pathguide Resources". It contains a brief description of the database and a message encouraging users to send feedback if they know of missing resources. A "News" box on the right lists recent updates, including a major update in August 2013 and the addition of visual navigation in May 2010. The main listing table has columns for Database Name, Details, Availability (Free or X), and Standards (BioPAX, SBML, PSI-MI). The table lists numerous databases such as 2P2Idb, 3D-Interologs, ADAN, AHD2.0, AllFuse, aMAZE, ANAP, AnimalTFDB, AntiJen, APID, AS-ALPS, ASD, ASEdb, ASPD, ATDB, AtPID, AtPIN, Bacteriome.org, BIANA, BID, BIND, BioGRID, and BiologicalNetworks.

Database Name (Order: alphabetically / by web popularity)	Details	Availability	Standards
2P2Idb - The Protein-Protein Interaction Inhibition Database	Details	Free	
3D-Interologs - 3D-Interologs	Details	Free	
3DID - 3D interacting domains	Details	Free	
ADAN - Prediction of protein-protein interaction of modular domains	Details	X	
AHD2.0 - Arabidopsis Hormone Database 2.0	Details	Free	
AllFuse - Functional Associations of Proteins in Complete Genomes	Details	X	
aMAZE - Protein Function and Biochemical Pathways Project	Details	X	
ANAP - Arabidopsis Network Analysis Pipeline	Details	Free	
AnimalTFDB - Animal Transcription Factor Database	Details	Free	
AntiJen - AntiJen is a Kinetic, Thermodynamic and Cellular Database	Details	Free	
APID - Agile Protein Interaction DataAnalyzer	Details	Free	
AS-ALPS - Alternative Splicing - induced ALteration of Protein Structure	Details	Free	
ASD - Allosteric Database	Details	Free	
ASEdb - Alanine Scanning Energetics Database	Details	Free	
ASPD - Artificial Selected Proteins/Peptides Database	Details	Free	
ATDB - Animal Toxin Database	Details	Free	
AtPID - Arabidopsis thaliana Protein Interactome Database	Details	Free	
AtPIN - Arabidopsis thaliana Protein Interactome Network	Details	Free	
Bacteriome.org - Bacterial Protein Interaction Database for Escherichia Coli	Details	Free	
BIANA - Biologic Interaction and Network Analysis	Details	Free	
BID - Binding Interface Database	Details	Free	
BIND - Biomolecular Interaction Network Database	Details	Free	PSI-MI
BioGRID - Biological General Repository for Interaction Datasets	Details	Free	PSI-MI
BiologicalNetworks - BiologicalNetworks	Details	Free	

いろいろなパスウェイデータベース

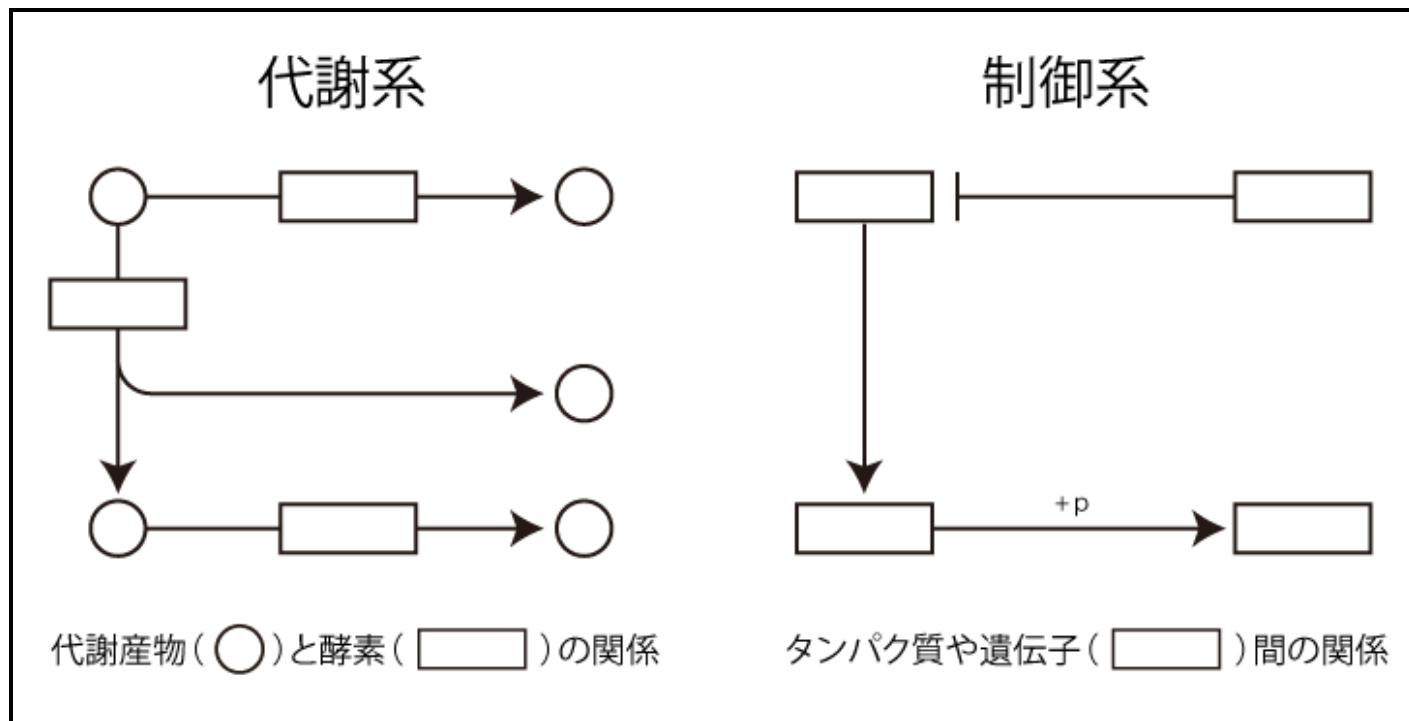
- 歴史的には代謝経路の表現から始まりましたが、現在ではタンパク質間相互作用、シグナル伝達系、遺伝子制御、環境シグナルなど様々な生命現象がパスウェイとして表現されています
- Pathguide での分類
 - タンパク質間相互作用
 - 代謝パスウェイ
 - シグナリングパスウェイ
 - 転写因子・遺伝子制御ネットワーク
 - タンパク質-化合物間相互作用
 - 遺伝的相互作用ネットワーク
 - アミノ酸配列解析
 - その他

パスウェイデータベースの表現

見やすくするために、グラフで表現されていることが多くなっています

•KEGG Pathway での表現例

- 代謝パスウェイでは代謝産物をノード、酵素反応をエッジとして表現されています
- 制御系ではタンパク質や遺伝子、その他の小分子をノード、その関係性(活性化、抑制、リン酸化など)がエッジとして表現されています

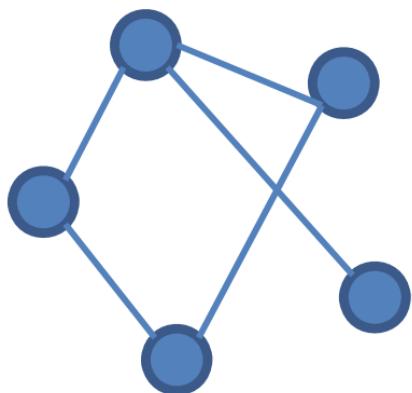


グラフ表現

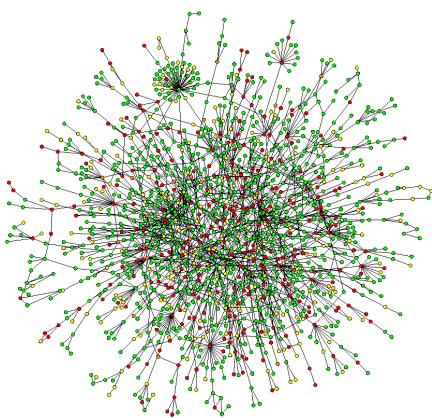
- グラフとは、情報科学において、ノード(頂点)とノード間の連結関係を表すエッジ(辺)で構成されるデータ構造。数学におけるグラフ理論に基づく。
- グラフは $G=(V,E)$ で表され、 V はノードの集合、 E はノードとノードをつなぐエッジの集合。

様々なグラフ構造

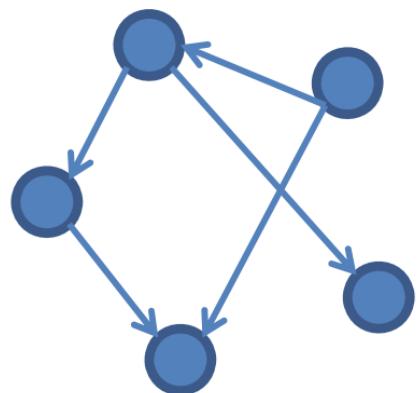
無向グラフ



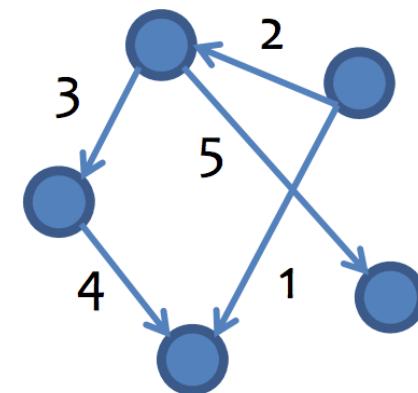
例:タンパク質間相互作用ネットワーク



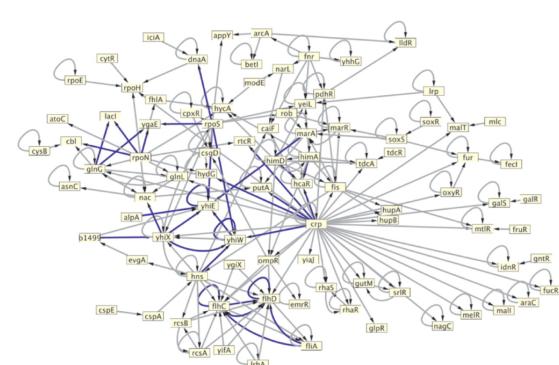
有向グラフ



重み付き有向グラフ

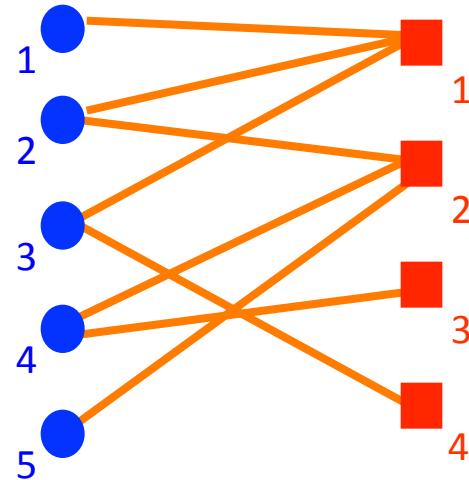


例:遺伝子制御ネットワーク

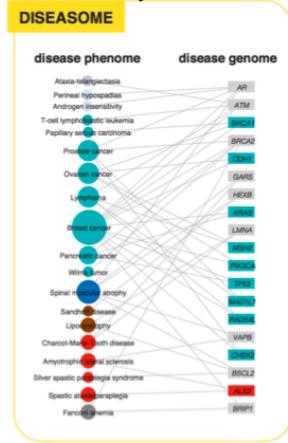


異種ノードのグラフ構造(二部グラフ)

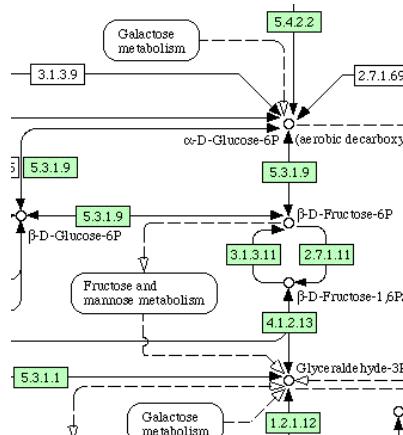
二部グラフとは、1種類目のノード(頂点)と2種類目のノード間の連結関係を表すエッジ(辺)で構成されるデータ構造。



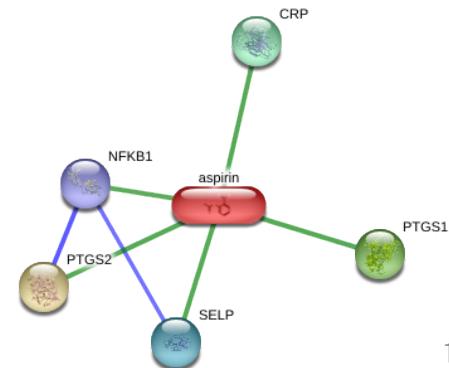
疾患・疾患関連遺伝子
ネットワーク



代謝パスウェイ



薬と標的分子間
相互作用ネットワーク



パスウェイデータベースのデータ形式

計算機での取り扱いを目的として XML (Extensible Markup Language) で記述されていることが多くなっています。

- [KGML](#) (KEGG Markup Language) は分子間の関係とダイアグラムのレイアウトを取り扱うための KEGG 独自のフォーマット
 - [SBML](#) (Systems Biology Markup Language)、[CellML](#)、[CSML](#) (Cell System Markup Language) はパスウェイのシミュレーションやモデリングを行うためのフォーマット
 - [PSI-MI](#) (Proteomics Standards Initiative Molecular Interaction XML Format) はタンパク質間相互作用を記述するためのフォーマット
 - [BioPAX](#) (Biological Pathways Exchange) は様々なパスウェイデータを統合したり、データ交換を行うことを目的として策定された標準化を目指したフォーマット
- これらのデータ形式を扱うことのできるネットワーク可視化ソフトウェアには [Cytoscape](#) や [VisANT](#) などがあります。
- [AJACS58 : Cytoscapeを使ったデータの可視化](#)

どのパスウェイデータベースを研究に使えば良いかは、対象生物や対象パスウェイ、目的によって異なってきます。今回は BioCyc , KEGG PATHWAY のブラウザ上での使い方を紹介します。

本日の内容

- パスウェイデータベース
 - パスウェイデータベースの概要
 - BioCyc の紹介・使い方
 - KEGG PATHWAY の紹介・使い方
 - 疾患研究への応用

BioCyc (<http://biocyc.org/>)

- ・開発:SRIインターナショナル(Stanford Research Institute)
- ・対象:大腸菌からヒトまで、異株を含めて 7,600 種以上
 - ・専門家が手作業で作成した文献ベースのデータ+自動ツール
 - ・代謝パスウェイ、制御系
- ・利用:アカデミックフリー(→ 要 subscription ?)
- ・データ形式:BioPAX

The screenshot shows the BioCyc Database Collection homepage. At the top, there's a navigation bar with links for Sites, Search, Genome, Metabolism, Analysis, SmartTables, and Help. The main content area has a heading "BioCyc Database Collection". It includes a note about subscriptions and a brief description of what BioCyc is. Below that is a "Getting Started" section with usage instructions and a "New User Guide >>" button. To the right, there's a large diagram titled "RouteSearch" illustrating metabolic pathways. The diagram shows two routes: a "Long Route" (orange line) and a "Best Route: Atoms Conserved" (blue line). The "Loss of Atoms" is indicated by a dashed yellow arrow between the two routes. At the bottom of the diagram, there's a "RouteSearch: Search for Paths through the Metabolic Network" section with a "Learn More" button and a page navigation menu from 1 to 12. At the very bottom, there are sections for "Tools" and "BioCyc Databases".

BioCyc 実習1

対象生物種を見てみよう (<http://biocyc.org/biocyc-pgdb-list.shtml>)

各データベースには3つのレベルがあります

Tier 1: 専門家が手作業で作成した文献ベースのデータから構築したデータベース (7 DBs)

Tier 2: 自動ツールで作成したデータを手作業で修正 (>40 DBs)

Tier 3: 自動ツールで作成 (>7,500 DBs)

MetaCycだけは 生物種のデータベースではなく、2,000種以上から構築したリファレンス代謝パスウェイ

News

[News](#)
[Update History](#)

Information

[Introduction to BioCyc](#)
7615 Databases
Download
Pathway Tools Software
Pathway Tools Blog
Publications
Linking to BioCyc
Webinars
Contact Us

Services

[Subscribe to BioCyc](#)
Metabolic Posters
Genome Posters
Software/Database Downloads
Registry
Web Services

LIST OF BIOCYC PATHWAY/GENOME DATABASES

The 7615 databases within the BioCyc collection as of May 5, 2016 (version 20.0) are as follows:

Tier 1 Databases

PGDBs in Tier 1 have received at least one year of literature-based curation by scientists. For example, the EcoCyc DB is the result of more than 20 person-years of effort to enter information from the *E. coli* literature about the metabolism, transport, and genetic regulatory processes of this organism, and to update the genome annotation to reflect newly characterized *E. coli* genes.

Database	Contents	Organization
EcoCyc	<i>Escherichia coli</i> K-12 substr. MG1655	SRI International
MetaCyc	Metabolic pathways and enzymes from over 2063 organisms	SRI International
HumanCyc	250 metabolic pathways	SRI International
AraCyc	<i>Arabidopsis thaliana</i>	S. Rhee, Department of Plant Biology, Carnegie Institution, USA
YeastCyc	<i>Saccharomyces cerevisiae</i>	SGD Curators, Stanford U., USA
LeishCyc	<i>Leishmania major</i> Friedlin	Bio21 Molecular Science and Biotechnology Institute, University of Melbourne
TrypanoCyc	<i>Trypanosoma brucei</i>	French National Institute of Agricultural Research (INRA), University of Glasgow (UoG)

Tier 2 Databases

PGDBs in Tier 2 were generated by the PathoLogic program, which was used to predict their metabolic pathways, their operons (for bacteria only), and their pathway hole fillers (for many of the PGDBs). The resulting PGDBs underwent manual review by a person to remove false-positive pathway predictions that they could detect, and to perform other manual polishing steps such as defining protein complexes. The resulting PGDBs also underwent less than one year of literature-based curation, such as to enter metabolic pathways that had been experimentally elucidated in the organism but that were not inferred by PathoLogic. Most Tier 2 PGDBs have undergone 1-4 months of curation.

Many of the Tier 2 DBs are available for adoption [more] for curation and updating by interested scientists.

Click on the name of a database to navigate to the home page for that database for more information, such as the database authors, and statistics on the database content.

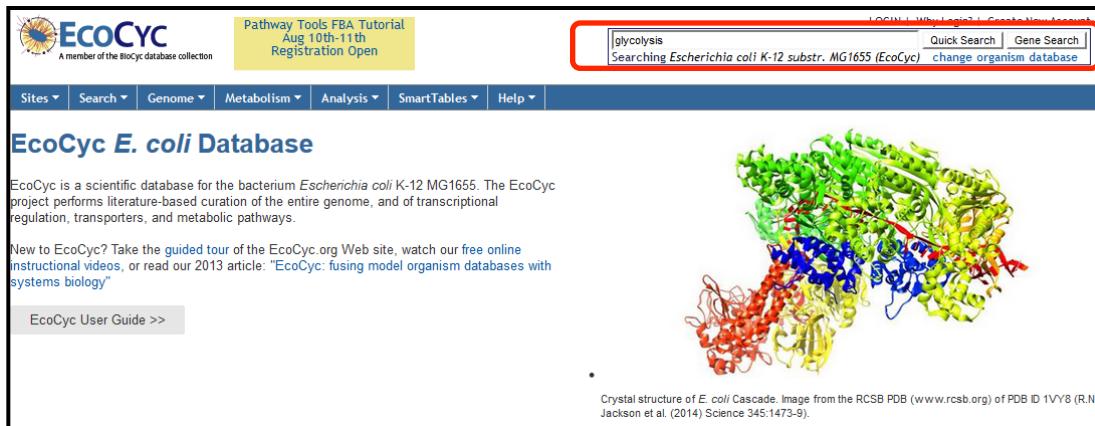
Database	Organization
<i>Agrobacterium fabrum</i> C58	SRI International
<i>Anopheles gambiae</i>	Ezekiel Adebiyi, Olubanke Ogunlana, Covenant University
<i>Aurantimonas manganoxydans</i> SI85-9A1	SRI International
<i>Bacillus entrorhizus</i> Ames	SRI International
<i>Bacillus subtilis</i> 168	SRI International
<i>Bacteroides thetaiotaomicron</i> VPI-5482	SRI International
<i>Candidatus Cardinium hertigii</i>	Cavanilles Institute
<i>Candidatus Evansia muelleri</i>	Cavanilles Institute
<i>Candidatus Portiera aleurodgidarum</i> BT-QVLC	Cavanilles Institute
<i>Caulobacter crescentus</i> CB15	SRI International
<i>Caulobacter crescentus</i> NA1000	SRI International
<i>Chloromyces restiformis</i>	Carmnia Institute of Science (DMM)

BioCyc 実習2

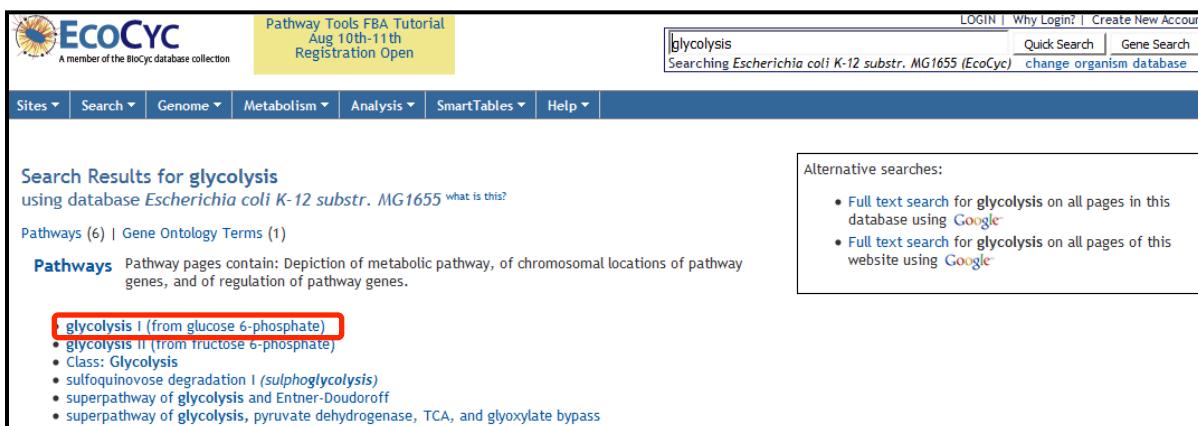
好きな生物のパスウェイを見てみよう

リストから好きな生物種のデータベースへ移動(例: [EcoCyc](#))

右上の検索ボックスで注目している生命現象関連の単語を入力し、Quick Search ボタンをクリック(例: glycolysis)

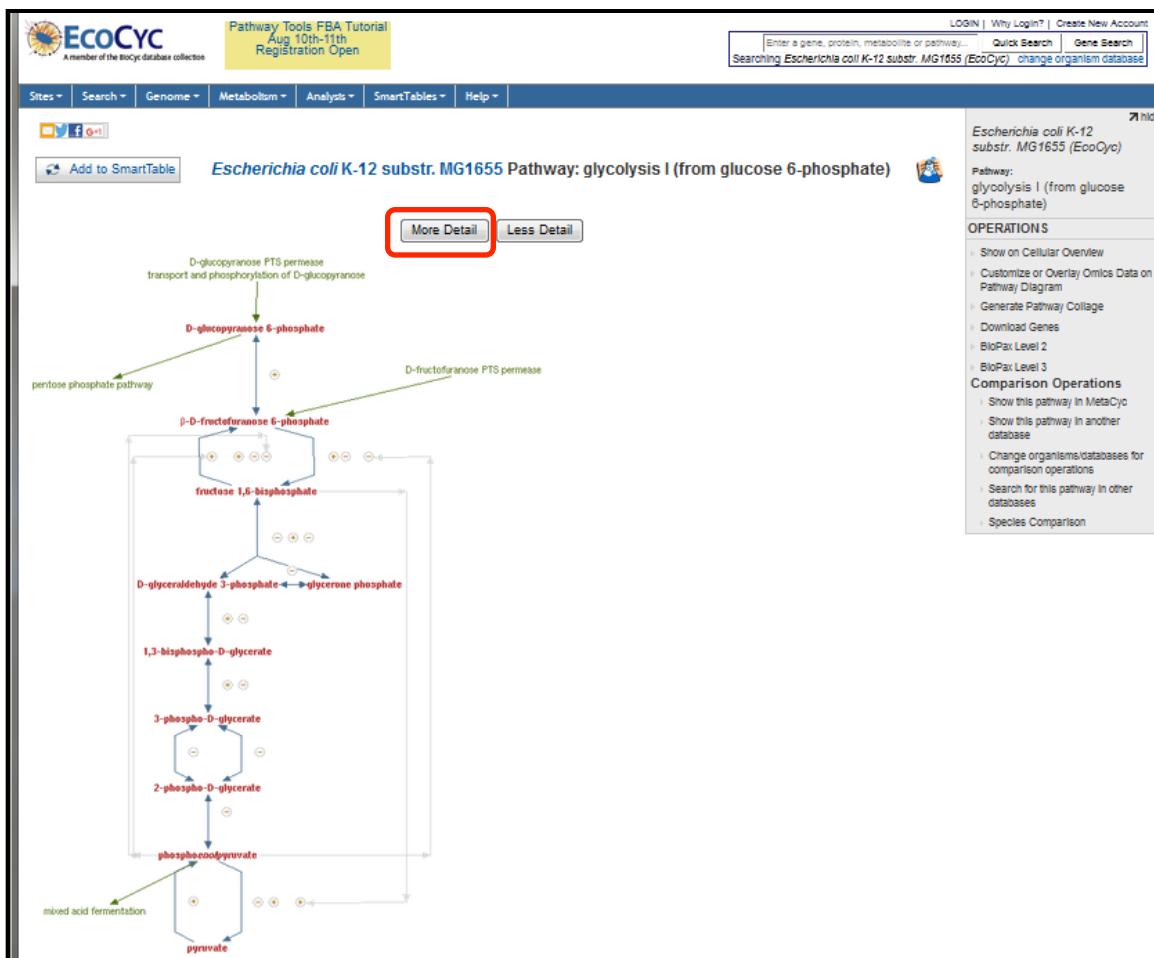


パスウェイのリストが表示されるので、クリック



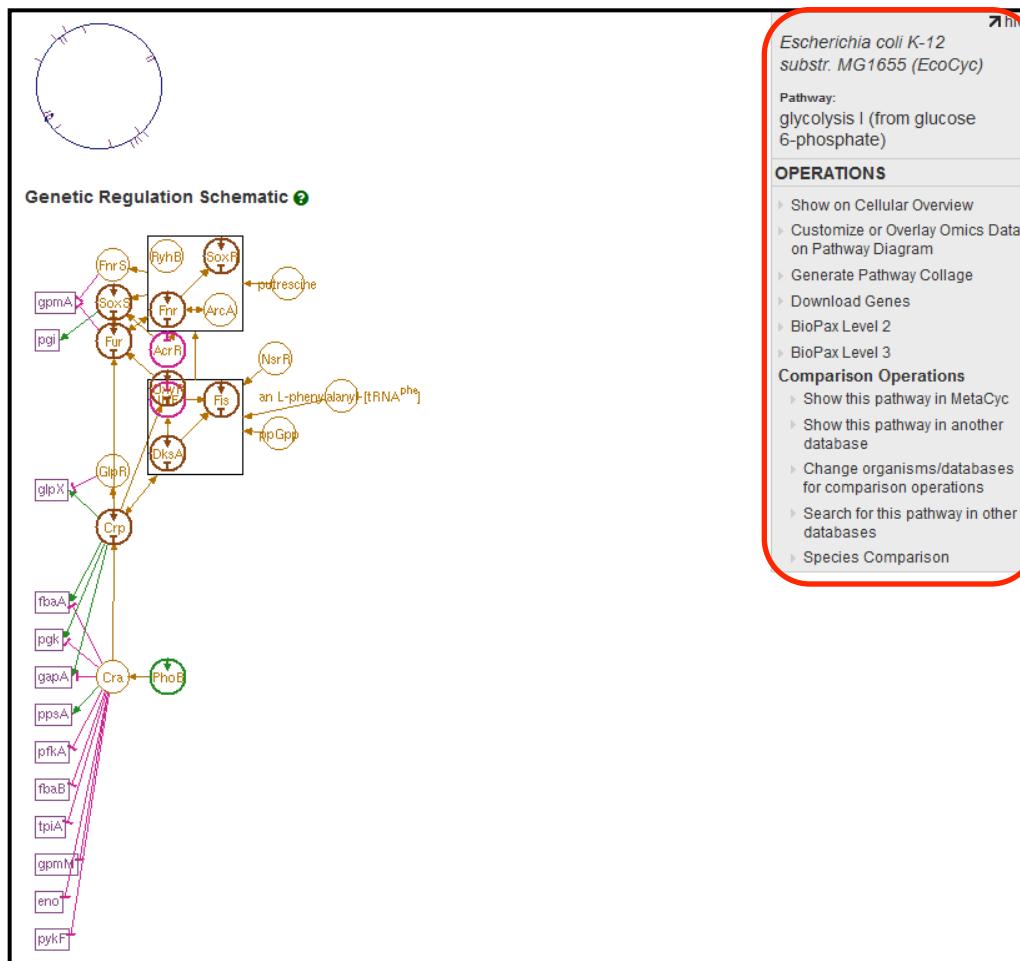
BioCyc 実習2

- 代謝パスウェイが表示されます
 - 代謝反応が青の矢印で、制御関係が灰色の矢印で示されている
 - More Detail ボタンをクリックすると、酵素名、代謝産物の構造などが追加される



BioCyc 実習2

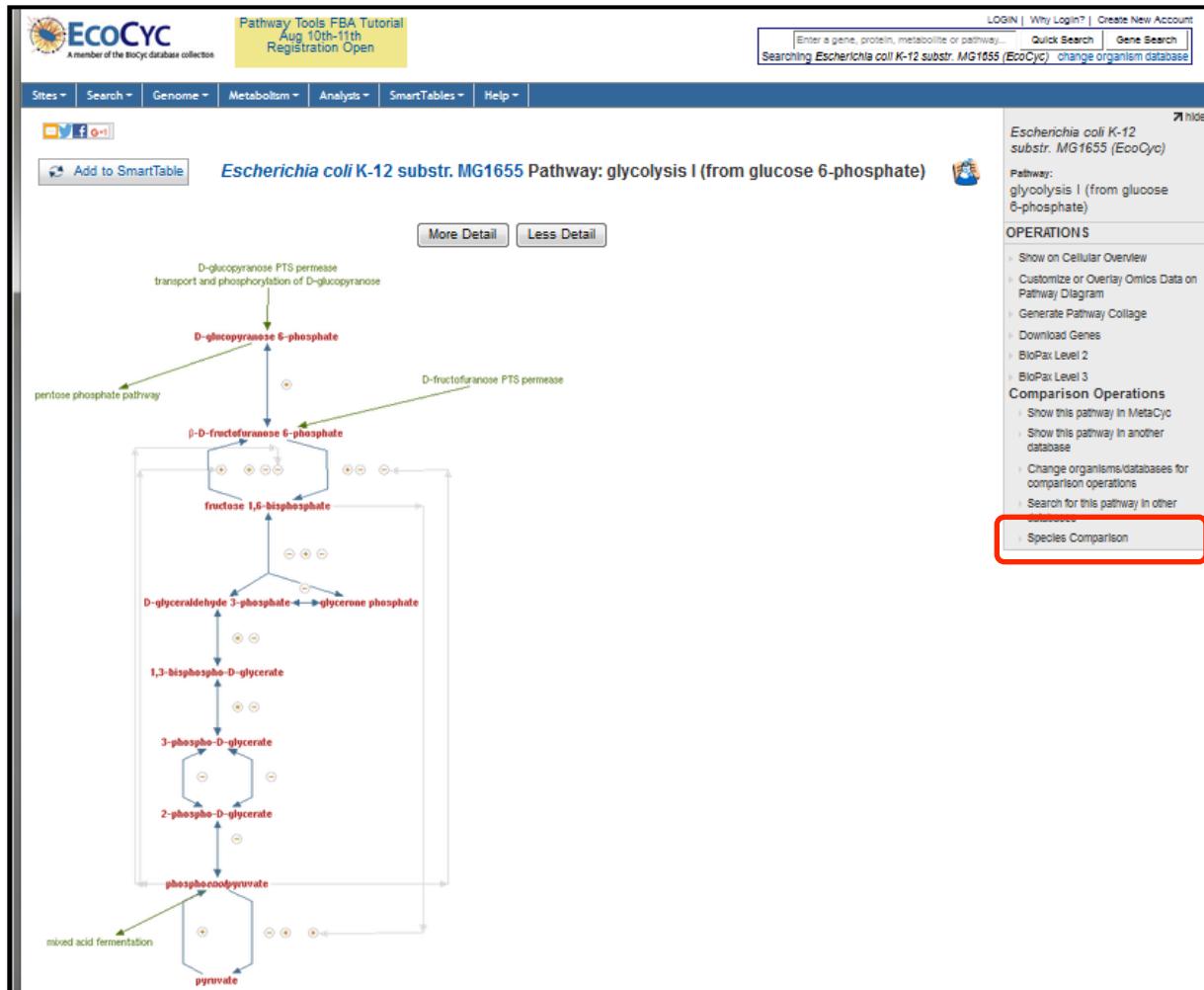
- EcoCyc など一部の生物種では、代謝パスウェイに関する酵素の発現制御ネットワークも表示されます
- 右の Options メニューから、表示のカスタマイズやダウンロードが可能



BioCyc 実習3

種間比較をしてみよう

- 右の Option メニューの Species Comparison をクリック



BioCyc 実習3

- 比較する種の選択画面が表示されるので、比較したい好きな種を入力し、OK をクリック (例: *Bacillus subtilis*)
- パスウェイや遺伝子、オペロン構造などが比較できます

The screenshot shows the 'Specify List of Organism Databases' dialog box from the EcoCyc database. The left panel lists organism databases by name, with 'Bacillus sub' selected and highlighted with a red box. The right panel shows the 'Current Selection' list, which also contains 'Bacillus subtilis' and is also highlighted with a red box. Below these panels are buttons for 'Add →', 'Remove ←', and 'Clear All'. At the bottom, there is a note about saving database lists for next time, and 'OK' and 'Cancel' buttons. A small diagram of a metabolic pathway is visible at the very bottom.

LOGIN | Why Login? | Create New Account

Pathway Tools FBA Tutorial
Aug 10th-11th
Registration Open

Enter a gene, protein, metabolite or pathway... Quick Search Gene Search

Searching *Escherichia coli* K-12 substr. MG1655 (EcoCyc) change organism database

Specify List of Organism Databases

To compare the requested object across species, please select from among the following organism databases.

Select One or More Databases:

By Name By Taxonomy By Organism Properties My Lists

A Bacillus sub

Bacillus subtilis BSn5
Bacillus subtilis natto BEST195
Bacillus subtilis PY79
Bacillus subtilis QB928
Bacillus subtilis spizizenii TU-B-10
Bacillus subtilis spizizenii W23
Bacillus subtilis subtilis 168
Bacillus subtilis subtilis 6051-HGW
Bacillus subtilis subtilis BAB-1
Bacillus subtilis subtilis BSP1
Bacillus subtilis subtilis NCIB 3610
Bacillus subtilis subtilis RO-NN-1
Bacillus subtilis XF-1

13 matching databases available

Current Selection

1 databases or taxonomic groups currently selected.

Bacillus subtilis

Add → Remove ← Clear All

Note: if you create a user account and log in, you can save your database lists for next time.

OK Cancel

EC 5.4.2.11 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase: gpmA

EC 5.4.2.12 2,3-bisphosphoglycerate-independent phosphoglycerate mutase: gpmM

glpX glpK glpF

gpmM envC yibQ

BioCyc 実習4

好きな生物の Overview パスウェイを見てみよう

- 上のメニューの Metabolism > [Cellular Overview](#) をクリック

The screenshot shows the EcoCyc website for *Escherichia coli* K-12 MG1655. The top navigation bar includes links for Sites, Search, Genome, Metabolism (with a red box around 'Cellular Overview'), Analysis, SmartTables, and Help. A yellow banner at the top left promotes the 'Pathway Tools FBA Tutorial Aug 10th-11th Registration Open'. The main content area features a sidebar with links like 'EcoCyc E. coli', 'Metabolic Map Poster', and 'Enrichment parameters' (which is highlighted with a red box). The central part of the page displays metabolic pathways and compound structures. At the bottom, there are sections for 'Tools' and 'Sibling Sites', along with a navigation menu.

EcoCyc is a scientific database for *Escherichia coli* K-12 MG1655. The database contains literature-based curation of metabolism, transcriptional regulation, and pathways.

New to EcoCyc? Take the [guided tour](#) of the EcoCyc.org Web site, watch our [free online instructional videos](#), or read our 2013 article: "EcoCyc: fusing model organism databases with systems biology"

[EcoCyc User Guide >>](#)

Metabolomics Data Analysis
Multiple tools are available in this website for metabolomics data analysis.

[Learn More](#)

1 2 3 4 5 6 7 8

Tools

BioCyc provides tools for navigating, visualizing, and analyzing the underlying databases, and for analyzing omics

Sibling Sites

PortEco hosts *E. coli* omics data and analysis tools; the [E. coli Student Portal](#) is a microbiology education

BioCyc 実習4

パスウェイや反応、遺伝子、酵素、化合物の名前や ID で Overview パスウェイをハイライト（例:TCA cycle）

EcoCyc
A member of the BioCyc database collection

Pathway Tools FBA Tutorial
Aug 10th-11th
Registration Open

LOGIN | Why Login? | Create New Account

Enter a gene, protein, metabolite or pathway...
Quick Search Gene Search
Searching *Escherichia coli K-12 substr. MG1655 (EcoCyc)* change organism database

Sites Search Genome Metabolism Analysis SmartTables Help

Base Layer Cellular Overview

OPERATIONS

Overlay Experimental Data (Omics Viewer) ▾

- Upload Data from File
- Enter/Paste Data from Keyboard
- Import Data from GEO
- From Recent Datasets (GEO only)
- From SmartTable

Highlight ▾

Highlight Pathway(s) (highlighted)

- Highlight Reaction(s) ▾
- Highlight Gene(s) ▾
- Highlight Enzyme(s) ▾
- Highlight Compound(s) ▾
- Clear All Highlighting

Show Legend
Generate Bookmark for Current Cellular Overview
Help

Cellular Overview of *Escherichia coli K-12 substr. MG1655 (EcoCyc)*
Pan left/right/up/down the entire diagram by holding the left mouse button, click on an object for more info, right-click (ctrl-click for Mac) for menu

Pathways shown include: Amino Acid Biosynthesis, Nucleosides and Nucleotides Biosynthesis, Cell Structures Biosynthesis, TCA cycle, and various degradation pathways like Fermentation, Respiration, and Detoxification.

BioCyc 実習4

パスウェイや反応、遺伝子、酵素、化合物の名前や ID で Overview パスウェイをハイライト（例:TCA cycle）

EcoCyc
A member of the BioCyc database collection

Pathway Tools FBA Tutorial
Aug 10th-11th
Registration Open

LOGIN | Why Login? | Create New Account

Enter a gene, protein, metabolite or pathway...
Searching Escherichia coli K-12 substr. MG1655 (EcoCyc) change organism database

Sites ▾ Search ▾ Genome ▾ Metabolism ▾ Analysis ▾ SmartTables ▾ Help ▾

Cellular Overview of *Escherichia coli* K-12 substr. MG1655 (EcoCyc)

Pan left/right/up/down the entire diagram by holding the left mouse button, click on an object for more info, right-click (ctrl-click for Mac) for menu

Base Layer

Cellular Overview

Overlays

List Pathway TCA cycle I (prokaryotic)

OPERATIONS

Overlay Experimental Data (Omics Viewer) ▾

- ▶ Upload Data from File
- ▶ Enter/Paste Data from Keyboard
- ▶ Import Data from GEO
- ▶ From Recent Datasets (GEO only)
- ▶ From SmartTable

Highlight ▾

Highlight Pathway(s) ▾

- ▶ By Name or Frame ID
- ▶ By Substring
- ▶ By Curation
- ▶ By Evidence

Highlight Reaction(s) ▾

Highlight Gene(s) ▾

Highlight Enzyme(s) ▾

Highlight Compound(s) ▾

▶ Clear All Highlighting

Show Legend

Generate Bookmark for Current Cellular Overview

Help

The screenshot shows the EcoCyc metabolic pathway overview for *Escherichia coli* K-12 substr. MG1655. The main area displays a complex network of pathways, including Carbohydrate Degradation, Amino Acids Degradation, and the TCA cycle I. A red box highlights a specific node in the TCA cycle I pathway. On the right side, there is an 'OVERLAYS' panel where the 'Pathway TCA cycle I (prokaryotic)' is selected. The 'OPERATIONS' panel contains various options for overlaying experimental data and highlighting specific pathways, reactions, genes, enzymes, and compounds.

BioCyc 実習5

遺伝子発現データをマッピングしてみよう

- NCBI GEO (Gene Expression Omnibus)のデータを直接マッピングできます
- 右のメニューの Import Data from GEO をクリック

The screenshot shows the EcoCyc metabolic pathway diagram for *Escherichia coli* K-12 substr. MG1655. The diagram is a complex network of blue arrows representing biochemical reactions and nodes representing metabolites. Key pathways visible include Carbohydrates Degradation, Amino Acids Degradation, Lipid Biosynthesis, and Cell Structures Biosynthesis. On the right side of the interface, there is a vertical menu titled "OPERATIONS". One of the items in this menu, "Import Data from GEO", is highlighted with a red box. Other menu items include "Overlay Experimental Data (Omics Viewer)", "From Recent Datasets (GEO only)", "From SmartTable", "Highlight" (with sub-options for Pathways, Reactions, Genes, Enzymes, Compounds), and "Show Legend".

BioCyc 実習5

- キーワード検索で GEO のデータを検索し、データを一つ選択(例:O157)
- マッピングオプション例
 - Type of display: Static (single timepoint)を選択
 - Use data form: Ratio of two columns を選択
 - Numerator Column, Denominator Column を一つづつ選択

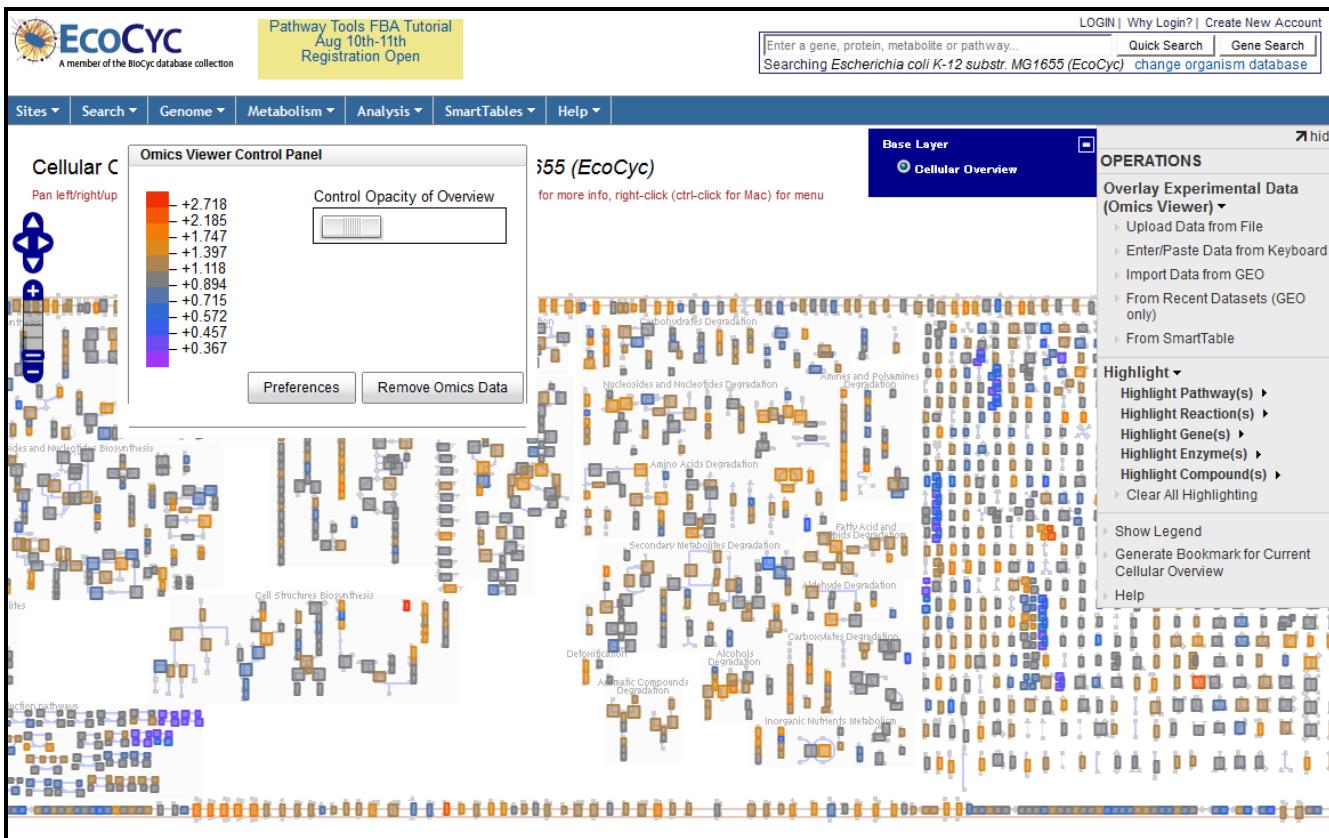
The screenshot shows the BioCyc Pathway Tools FBA Tutorial interface. The main window displays a cellular overview of Escherichia coli K-12 substr. MG1655 (EcoCyc). On the left, there is a pathway diagram with various metabolic pathways highlighted in blue. A central dialog box titled "Omics Viewer Data and Parameters" is open, showing the following details:

- Datasource: GEO
- Organism: Escherichia coli
- Keywords: O157
- Note: A keyword is optional, but without one, the request may not return in a timely fashion.
- Your query matched 2 datasets:
 - GDS4321: [Visvalingam13] Escherichia coli O157:H7 response to cinnamaldehyde, time course
 - GDS2879: [Bansal07] Enterohemorrhagic Escherichia coli response to epinephrine, norepinephrine, and...
- Show Details of Selected Dataset | Download Data to File
- Type of display: Static (single timepoint)
- Use data from: Ratio of two columns
- Numerator Column: GSM999245: E. coli O157:H7 at 2 h control biological rep1
GSM999246: E. coli O157:H7 at 2 h control biological rep2
GSM999247: E. coli O157:H7 at 4 h control biological rep1
GSM999248: E. coli O157:H7 at 4 h control biological rep2
GSM999249: E. coli O157:H7 at 2 h cinnamaldehyde treatment biological rep1
GSM999250: E. coli O157:H7 at 2 h cinnamaldehyde treatment biological rep2
GSM999251: E. coli O157:H7 at 4 h cinnamaldehyde treatment biological rep1
GSM999252: E. coli O157:H7 at 4 h cinnamaldehyde treatment biological rep2
- Submit | Cancel

On the right side of the interface, there is an "OPERATIONS" panel with options like "Overlay Experimental Data (Omics Viewer)" and "Highlight Pathway(s)". Below the operations panel, a reaction detail box is visible, showing EC 1.1.1.35 [3-oxo-(7Z)-hydroxy-(7Z)-hexadecene] with Enzymes: FadJ comp [FadJ]₂ fatty acid oxidase and Reversibility: physiol.

BioCyc 実習5

- 例) 病原性大腸菌O157のシンナムアルデヒド応答、2時間目と4時間目の遺伝子発現の割合
 - この例の場合、オレンジ-赤で示された遺伝子は発現が増え、青-紫で示された遺伝子は発現が減っています
- 余裕のある人は、いろいろなオプション(アニメーション、テーブル出力など)を試してみましょう
もっと詳しく知りたい場合は [BioCyc User's guide](#) を参照しましょう



本日の内容

- パスウェイデータベース
 - パスウェイデータベースの概要
 - BioCyc の紹介・使い方
 - KEGG PATHWAY の紹介・使い方
 - 疾患研究への応用

KEGG PATHWAY (<http://www.kegg.jp/>)

- 開発: 京都大学化学研究所
- 対象: ゲノムの決まった全生物種(異株を含む)4,200 種以上(>300 真核生物、>3,700 真正細菌、>220 古細菌)、真核ドラフトゲノム 25 種、環境メタゲノム 300 サンプル、生体メタゲノム 700 サンプル
 - 専門家が手作業で作成した文献ベースのデータ + 自動ツール
 - リファレンスパスウェイ: 専門家が手作業で文献ベースから作成
 - 生物種パスウェイ
 - 自動ツールでリファレンスパスウェイから作成し、手作業でキュレーション
 - 自動ツールで作成(自動ツールにも段階があります)
 - 代謝パスウェイ、シグナル伝達系、他
- 利用: アカデミックフリー
- データ形式: KGML
 - [KCPAVS KEGG-XML converter](#) などで代謝パスウェイ、シグナル伝達などの多くのパスウェイを標準形式に変換可能
 - [KEGGscape](#) でネットワーク可視化ソフト Cytoscape に読み込み可能

KEGG はデータベースの集合

- KEGG2 をクリック
- KEGG PATHWAY を含むシステム情報データベースの他に、遺伝情報、化学情報、健康情報などのデータベースがリンクしています
- 今回は PATHWAY以外の詳細は省くので、それ以外の詳細は過去の AJACS 資料を参照 ([AJACS50](#), [AJACS51](#) (付録資料), [AJACS53](#), [AJACS54](#))

The screenshot shows the KEGG homepage with a sidebar containing links for Home, Database, Objects, Software, and more. The main content area features a search bar, a logo, and a summary of KEGG's mission to understand biological systems. It also lists various entry points like KEGG2, KEGG PATHWAY, KEGG BRITE, etc.

The screenshot shows the KEGG Table of Contents page. A large blue arrow points from the KEGG2 link on the left to the KEGG PATHWAY category in the table of contents. The table is organized into categories: Systems information, Genomic information, Chemical information, and Health information. Each category lists various KEGG databases with their corresponding search functions.

Category	Entry Point	Search & Compute	DBGET Search
Systems information	KEGG PATHWAY	Search Pathway	PATHWAY
	KEGG BRITE	Search Brite	BRITE
	KEGG MODULE	Reconstruct Module	MODULE
	KEGG Mapper	Map Taxonomy	
Genomic information	KEGG ORTHOLOGY	BlastKOALA	ORTHOLOGY
	KEGG Annotation	GhostKOALA Annotate Sequence	
	KEGG GENOME	KO system	
	KEGG GENES	SSDB search	GENOME
Chemical information	KEGG Organisms	OC viewer† BLASTT / FASTA† KAAS†	GENES DGENES MGENOME† MGENEST†
	KEGG LIGAND	SIMCOMP† / SUBCOMP†	COMPOUND
	KEGG COMPOUND	KCaMT	GLYCAN
	KEGG GLYCAN	PathSearch†	REACTION RPAIR RCLASS
Health information	KEGG REACTION	PathComp†	ENZYME
	Reaction Modules	PathPred† E-zyme†	
	KEGG DISEASE	SIMCOMP† / SUBCOMP†	DISEASE
	[Cancer Pathogen]	KCaMT	DRUG DGROUP
KEGG DRUG	PathSearch†	ENVIRON	
KEGG ENVIRON	PathComp†		
KEGG MEDICUS	PathPred† E-zyme†		
	Pathogen Checker MEDICUS search		
	Drug interaction checker		
	Human diseases		
	Infectious diseases		
	ATC drug classification		

KEGG PATHWAY 実習1

対象生物種を見てみよう

- データベースのテーブルの下、[KEGG organisms](#) をクリック

The screenshot shows the KEGG homepage with two main sections: 'Chemical information' and 'Health information'. Under 'Chemical information', there are links to KEGG LIGAND, KEGG COMPOUND, KEGG GLYCAN, KEGG REACTION, and Reaction Modules. Under 'Health information', there are links to KEGG DISEASE, KEGG DRUG, KEGG ENVIRON, and KEGG MEDICUS. Below these sections, a note states 'KEGG is developed by Kanehisa Laboratories. See [Kanehisa et al. \(2016\)](#) for updates of KEGG.' and '+ Developed and maintained by Kyoto University Bioinformatics Center as part of its GenomeNet service.' At the bottom, there is a search bar with the placeholder 'KEGG organisms - the list of currently available organisms' and a button labeled 'Select Organism'.

- KEGG では 3-4 文字の独自の生物種コードを使用している

The screenshot shows the 'KEGG Organisms: Complete Genomes' page. It features a logo for 'Kyoto Encyclopedia of Genes and Genomes' and a navigation bar with links for 'Genomes', 'Species', 'Genus', 'Draft', and 'Meta'. Below this, a section titled 'Eukaryotes' displays a table with columns for 'Category' and 'Organisms'. The 'Organisms' column lists various species with their common names and source information. A red box highlights the first few entries in the 'Organisms' column: 'hsa Homo sapiens (human) RefSeq', 'ptr Pan troglodytes (chimpanzee) RefSeq', 'pps Pan paniscus (bonobo) RefSeq', 'ggo Gorilla gorilla gorilla (western lowland gorilla) RefSeq', 'pon Pongo abelii (Sumatran orangutan) RefSeq', 'nle Nomascus leucogenys (northern white-cheeked gibbon) RefSeq', 'mcc Macaca mulatta (rhesus monkey) RefSeq', 'mdf Macaca fascicularis (crab-eating macaque) RefSeq', and 'rro Rhinopithecus roxellana (golden snub-nosed monkey) RefSeq'.

Category	Organisms	Source
	hsa Homo sapiens (human)	RefSeq
	ptr Pan troglodytes (chimpanzee)	RefSeq
	pps Pan paniscus (bonobo)	RefSeq
	ggo Gorilla gorilla gorilla (western lowland gorilla)	RefSeq
	pon Pongo abelii (Sumatran orangutan)	RefSeq
	nle Nomascus leucogenys (northern white-cheeked gibbon)	RefSeq
	mcc Macaca mulatta (rhesus monkey)	RefSeq
	mdf Macaca fascicularis (crab-eating macaque)	RefSeq
	rro Rhinopithecus roxellana (golden snub-nosed monkey)	RefSeq

KEGG PATHWAY 実習1

・生物種コードのリンクをクリックすると、種の情報が表示されます

•Annotation

- manual : 手作業によるアノテーション(ヒト(hsa)等)
- KOALA : SSEARCH ベースの自動ツールによるアノテーション(ゴリラ(ggo)等)
- BlastKOALA : BLAST ベースの自動ツールによるアノテーション(ドラフトゲノム)
- GhostKOALA : GhostX ベースの自動ツールによるアノテーション(メタゲノム)

KEGG Homo sapiens (human)

Genome info Pathway map Brite hierarchy Module Genome map Blast Taxonomy

Search genes: Go Clear

Genome information

T number	T01001
Org code	hsa
Aliases	HUMAN, 9606
Full name	Homo sapiens (human)
Definition	Homo sapiens (human)
Annotation	manual
Taxonomy	TAX: 9606
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Data source	RefSeq (Assembly: GCF_000001405.31)
Original DB	NCBI, OMIM, HGNC, HPRD, Ensembl
Statistics	Number of protein genes: 20392 Number of RNA genes: 19362

KEGG Gorilla gorilla gorilla (western lowland gorilla)

Genome info Pathway map Brite hierarchy Module Genome map Blast Taxonomy

Search genes: Go Clear

Genome information

T number	T02442
Org code	ggo
Aliases	GORGO, 9595
Full name	Gorilla gorilla gorilla (western lowland gorilla)
Definition	Gorilla gorilla gorilla (western lowland gorilla)
Annotation	KOALA
Taxonomy	TAX: 9595
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Gorilla
Data source	RefSeq (Assembly: GCF_000151905.1)
Original DB	Sanger
Statistics	Number of protein genes: 22000 Number of RNA genes: 24

KEGG PATHWAY データベースリスト

•データベースリスト

- Genomes : 主に NCBI Refseq、GenBank に登録された生物種
- Species : 異株を一つにまとめたデータベース
- Genus : Genus レベルでまとめたデータベース
- Draft : Genomes に入っていない真核生物
- Meta : メタゲノム

The screenshot shows the KEGG Organisms: Complete Genomes page. At the top left is the KEGG logo. In the center, the title "KEGG Organisms: Complete Genomes" is displayed. Below it, statistics are shown: Eukaryotes: 342, Bacteria: 3778, Archaea: 229. A navigation bar at the bottom contains links: Genomes (highlighted with a red box), Species, Genus, Draft, and Meta.

Eukaryotes

Category	Organisms	Source
	hsa Homo sapiens (human)	RefSeq
	ptr Pan troglodytes (chimpanzee)	RefSeq
	pps Pan paniscus (bonobo)	RefSeq
	ggo Gorilla gorilla gorilla (western lowland gorilla)	RefSeq
	pon Pongo abelii (Sumatran orangutan)	RefSeq
	nle Nomascus leucogenys (northern white-cheeked gibbon)	RefSeq
	mcc Macaca mulatta (rhesus monkey)	RefSeq
	mcf Macaca fascicularis (crab-eating macaque)	RefSeq
	rro Rhinopithecus roxellana (golden snub-nosed monkey)	RefSeq

KEGG PATHWAY 実習2

パスウェイマップを見てみよう

・トップページ 上方の検索ボックスで "lysine biosynthesis" や "glycolysis" やなどの生命現象関連の単語を入力し、Search ボタンをクリック

The screenshot shows the KEGG homepage. At the top, there is a logo for 'Kyoto Encyclopedia of Genes and Genomes' (KEGG) featuring a colorful circular design. Below the logo is a search bar with the word 'lysine' typed into it. A red box highlights the search bar and the 'Search' button to its right. To the right of the search bar, there are links for 'Help' and '» Japanese'. On the left side of the page, there is a sidebar with links to 'KEGG Home', 'Release notes', 'Current statistics', 'Plea from KEGG', 'KEGG Database', and 'KEGG overview'. The main content area is titled 'KEGG: Kyoto Encyclopedia of Genes and Genomes' and contains a brief description of what KEGG is: a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput.

・KEGG データベース全体でヒットしたエントリーが全てリストアップされ、KEGG PATHWAY にヒットがあれば、一番上に表示されます

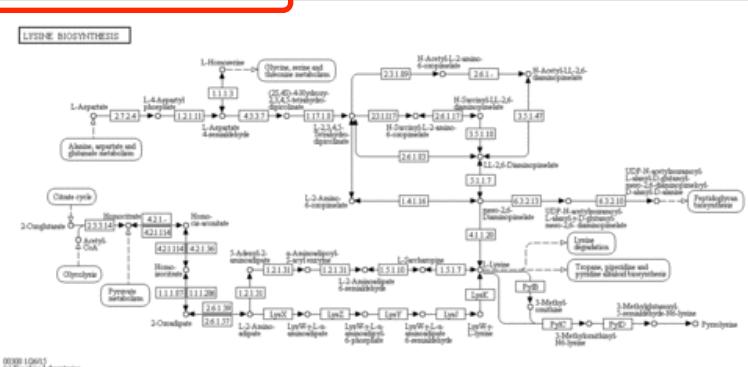
The screenshot shows the search results for 'lysine' on the KEGG website. At the top, there is a search bar with 'KEGG' selected as the database and 'lysine' as the search term. Below the search bar, the text 'Database: KEGG - Search term: lysine' is displayed. The results are categorized under 'KEGG PATHWAY'. The first result listed is 'map00300 Lysine biosynthesis', which is highlighted with a red box. Other results listed are 'map00310 Lysine degradation' and 'map01064 Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid'. At the bottom of the list, there is a section titled 'KEGG BRITE'.

KEGG PATHWAY 実習2

各パスウェイの情報が表示されます

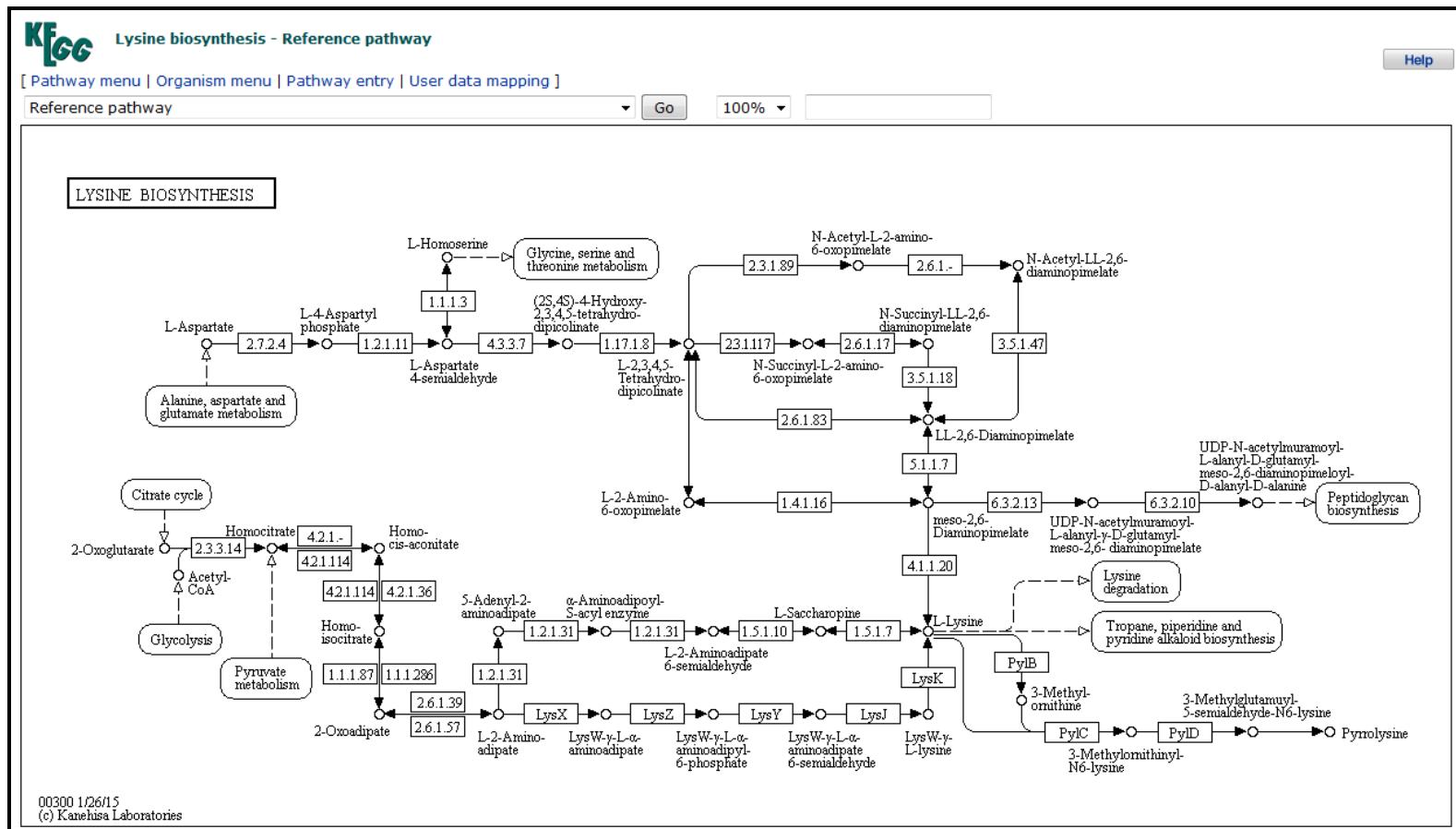
- KEGGにおけるパスウェイの最小単位で、グラフ画像を "マップ" と呼んでいます

KEGG PATHWAY: map00300 Help

Entry	map00300	Pathway
Name	Lysine biosynthesis	
Class	Metabolism; Amino acid metabolism	
BRITE hierarchy		
Pathway map	map00300 Lysine biosynthesis	
		
All organisms Ortholog table		
Module	M00016 Lysine biosynthesis, succinyl-DAP pathway, aspartate => lysine [PATH: map00300] M00030 Lysine biosynthesis, AAA pathway, 2-oxoglutarate => 2-aminoadipate => lysine [PATH: map00300] M00031 Lysine biosynthesis, mediated by LysW, 2-aminoadipate => lysine [PATH: map00300] M00433 Lysine biosynthesis, 2-oxoglutarate => 2-oxoadipate [PATH: map00300] M00525 Lysine biosynthesis, acetyl-DAP pathway, aspartate => lysine [PATH: map00300] M00526 Lysine biosynthesis, DAP dehydrogenase pathway, aspartate => lysine [PATH: map00300] M00527 Lysine biosynthesis, DAP aminotransferase pathway, aspartate => lysine [PATH: map00300]	

KEGG PATHWAY 実習2

- この色のついていない白いグラフが、専門家が手作業で文献ベースから作成したリファレンスパスウェイになります
- 代謝パスウェイの場合、長方形の箱が酵素反応（酵素、酵素遺伝子）、丸が代謝産物、環境物質などの化合物
- 各図形の説明は右上の Help から見られます



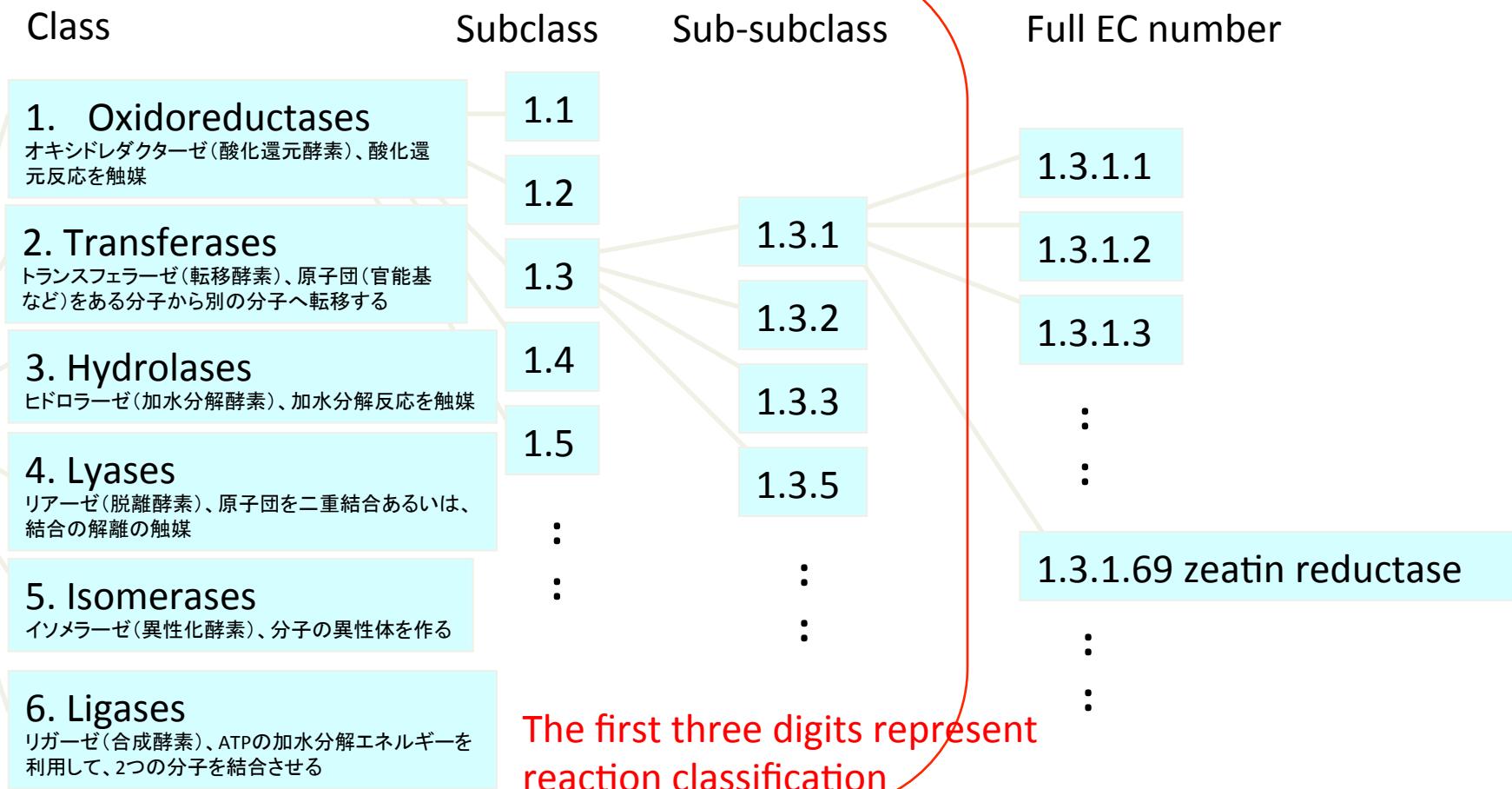
Helpページ

KEGG Pathway Maps

[Close](#)

Map	The KEGG PATHWAY database is a collection of manually drawn graphical diagrams, called KEGG pathway maps, representing molecular pathways for metabolism, genetic information processing, environmental information processing, other cellular processes, human diseases, and drug development. Each pathway is identified by a five-digit number preceded by one of: map, ko, ec, rn, and three- or four-letter organism code. The pathway map is drawn and updated with the notation shown below.																																				
Notation	<table><thead><tr><th>Objects</th><th>Arrows</th><th></th></tr></thead><tbody><tr><td></td><td></td><td>gene product, mostly protein but including RNA</td></tr><tr><td></td><td></td><td>other molecule, mostly chemical compound</td></tr><tr><td></td><td></td><td>another map</td></tr></tbody></table> <table><thead><tr><th>Protein-protein interactions</th><th>Gene expression relations</th></tr></thead><tbody><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr></tbody></table>			Objects	Arrows				gene product, mostly protein but including RNA			other molecule, mostly chemical compound			another map	Protein-protein interactions	Gene expression relations																				
Objects	Arrows																																				
		gene product, mostly protein but including RNA																																			
		other molecule, mostly chemical compound																																			
		another map																																			
Protein-protein interactions	Gene expression relations																																				
		phosphorylation																																			
		dephosphorylation																																			
		ubiquitination																																			
		glycosylation																																			
		methylation																																			
		activation																																			
		inhibition																																			
		indirect effect																																			
		state change																																			
		binding / association																																			

長方形の箱の中に書かれている4桁の番号はEC番号(酵素番号)を示す。酵素反応の分類の一つで、酵素の識別子としても使われる。



EC番号(酵素番号)の分類基準

クラス	サブクラス	サブサブクラス
1:酸化還元酵素	基質の分類	補酵素の分類
2:転移酵素	転移基の分類	より詳細な転移基の分類
3:加水分解酵素	分解される結合の分類	より詳細な結合の分類
4:付加脱離酵素	切断される結合の分類	生成物の分類
5:異性化酵素	異性化反応の分類	基質の分類
6:合成酵素	生成される結合の分類	生成物の分類

例:アセトアルデヒド脱水素酵素(EC: 1.2.1.10)

KEGG PATHWAY 実習3

好きな生物のパスウェイを見てみよう

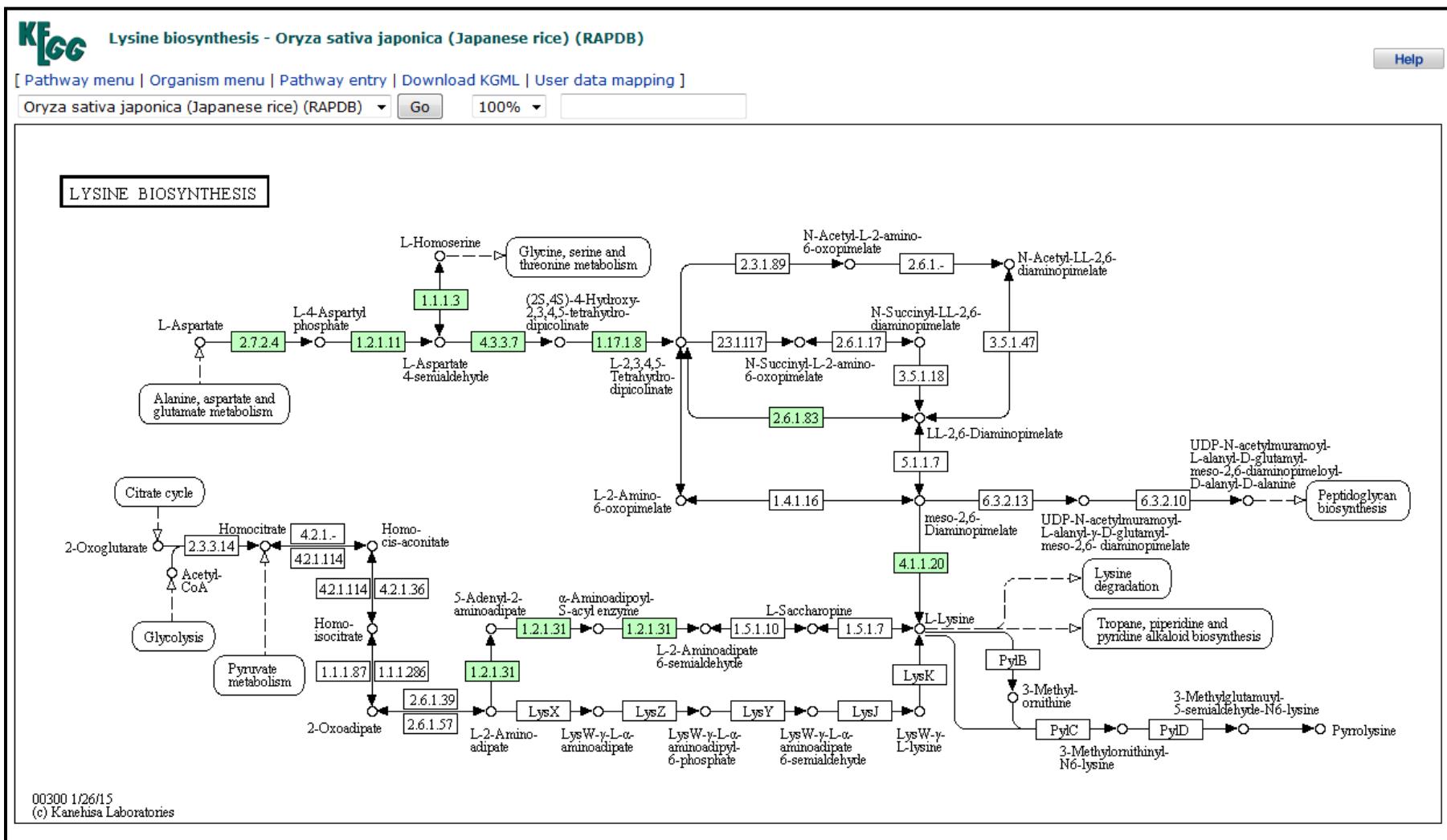
・プルダウンメニューから好きな生物を選択して Go をクリック

- リストが多すぎて選びにくいので
 - < Sort below by alphabet > を選択して Go をクリックでリストをソート
 - < Set personalized menu > を選択して Go をクリックでポップアップウィンドウからリストの絞り込み

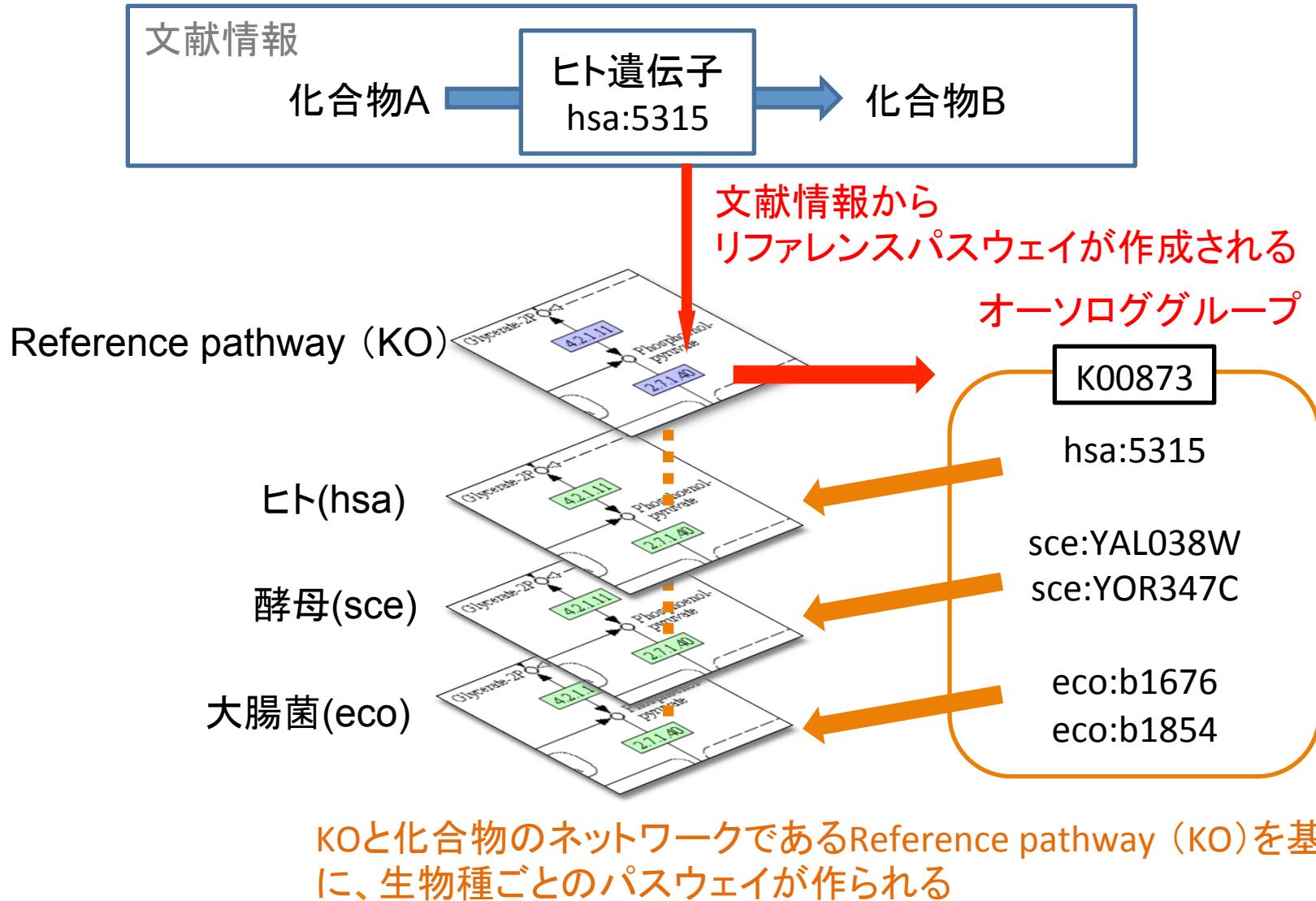
The screenshot shows two windows side-by-side. On the left is the 'Lysine biosynthesis - Reference pathway' diagram from KEGG. It's a complex metabolic map showing the biosynthesis of Lysine from various precursors like L-Aspartate and L-Homoserine. Enzymes are labeled with EC numbers (e.g., 2.7.2.4, 1.1.1.3, 2.3.1.89) and some with protein names (e.g., LysX, LysY, LysZ). Pathways are grouped into boxes: 'LYSINE BIOSYNTHESIS', 'Citrate cycle', 'Glycolysis', 'Pyruvate metabolism', 'Alanine, aspartate and glutamate metabolism', 'Glycine, serine and threonine metabolism', 'Lysine degradation', 'Tropane, piperidine and pyridine alkaloid biosynthesis', and 'Peptidoglycan biosynthesis'. On the right is a 'Select organisms' dialog box from Mozilla Firefox. It has a title bar 'Select organisms - Mozilla Firefox' and a URL 'www.kegg.jp/kegg/misc/kegg2_select.html'. Below the title is a heading 'Select organisms shown in the pathway menu'. There are checkboxes for 'GENES' (checked), 'EGENES' (unchecked), and categories: 'All organisms', 'Eukaryotes', 'Prokaryotes', 'Animals', 'Plants', 'Bacteria', and 'Archaea'. A radio button 'Another category' is selected, with 'monocots' typed into the input field. A red box highlights this input field. At the bottom are 'Select' and 'Favorites (enter KEGG organism codes)' buttons.

KEGG PATHWAY 実習3

・一部のボックスが緑色で塗られる、その生物(またはサンプル)の持つ遺伝子を示しています



リファレンスパスウェイと種毎のパスウェイの関係



オントロジー

- 生命科学における情報（遺伝子、タンパク質、機能など）を有効活用するためには、同じものを統一された語彙で表現する必要がある。
- 異なる分野間、異なるデータベース間、異なる生物種間で、データの結合や、横断比較を行うことが可能になる。

遺伝子のオントロジーのデータベース

- GO (Gene Ontology)
- KO (KEGG Orthology)

GO (Gene Ontology)

- 遺伝子の機能の記述に関して生物学分野における共通語彙の作成を目指した用語、用例辞書
- 同じ遺伝子に対して同じ名前(ID)を用いる
- 同じ機能に対して同じ用語を用いる
- GOターム(GOで定義された用語)は、親子関係はグラフ構造になっており、意味的な包含関係を表す

GOの特徴

- Biological process
 - 細胞内の生理機能に関する用語
- Cellular component
 - 細胞内小器官などの細胞内局在性に関する用語
- Molecular function
 - 生化学的、分子生物学的な分子機能に関する用語

KO (KEGG Orthology)

- 京都大学化学研究所が開発しているKEGGデータベースの一部で、遺伝子の識別子
- 同じ遺伝子に対して同じ名前(ID)を用いる
- 同じ機能に対して同じ用語を用いる
- パスウェイ上または機能の階層で同一の場所にマップされる遺伝子をオーソログ(KO - KEGG Orthology)として定義
- KOターム(KOで定義された用語)は、階層構造になっており、意味的な包含関係を表す形になっている

KO (KEGG Orthology)



KEGG ORTHOLOGY (KO) Database

Linking genomes to pathways by ortholog annotation

[KEGG2](#) [PATHWAY](#) [BRITE](#) [MODULE](#) [KO](#) [GENOME](#) [GENES](#) [SSDB](#) [Organisms](#)

Enter K numbers (Example) K00161 K00162 K00163 K00627 K00382

[Filter](#)

[Ortholog table](#)

[Pathway mapping](#)

[Brite mapping](#)

[Get title](#)

[Get entry](#)

[Clear](#)

KEGG Orthology (KO) System

The KEGG reference pathways maps, BRITE functional hierarchies, and KEGG modules are represented in a general way to be applicable to all organisms. The **KEGG Orthology (KO)** system is the basis for this representation, consisting of manually defined ortholog groups that correspond to KEGG pathway nodes, BRITE hierarchy nodes, and KEGG module nodes.

- [KEGG Orthology \(KO\)](#)

Once genes are assigned the KO identifiers, or the K numbers, by the genome annotation procedure described below, the organism-specific pathway maps, BRITE functional hierarchies, and KEGG modules are automatically generated (see [KEGG mapping](#) for details).



[Search](#)

[ORTHOLOGY](#)

for

[Go](#)

[Clear](#)



[bfm mode](#)

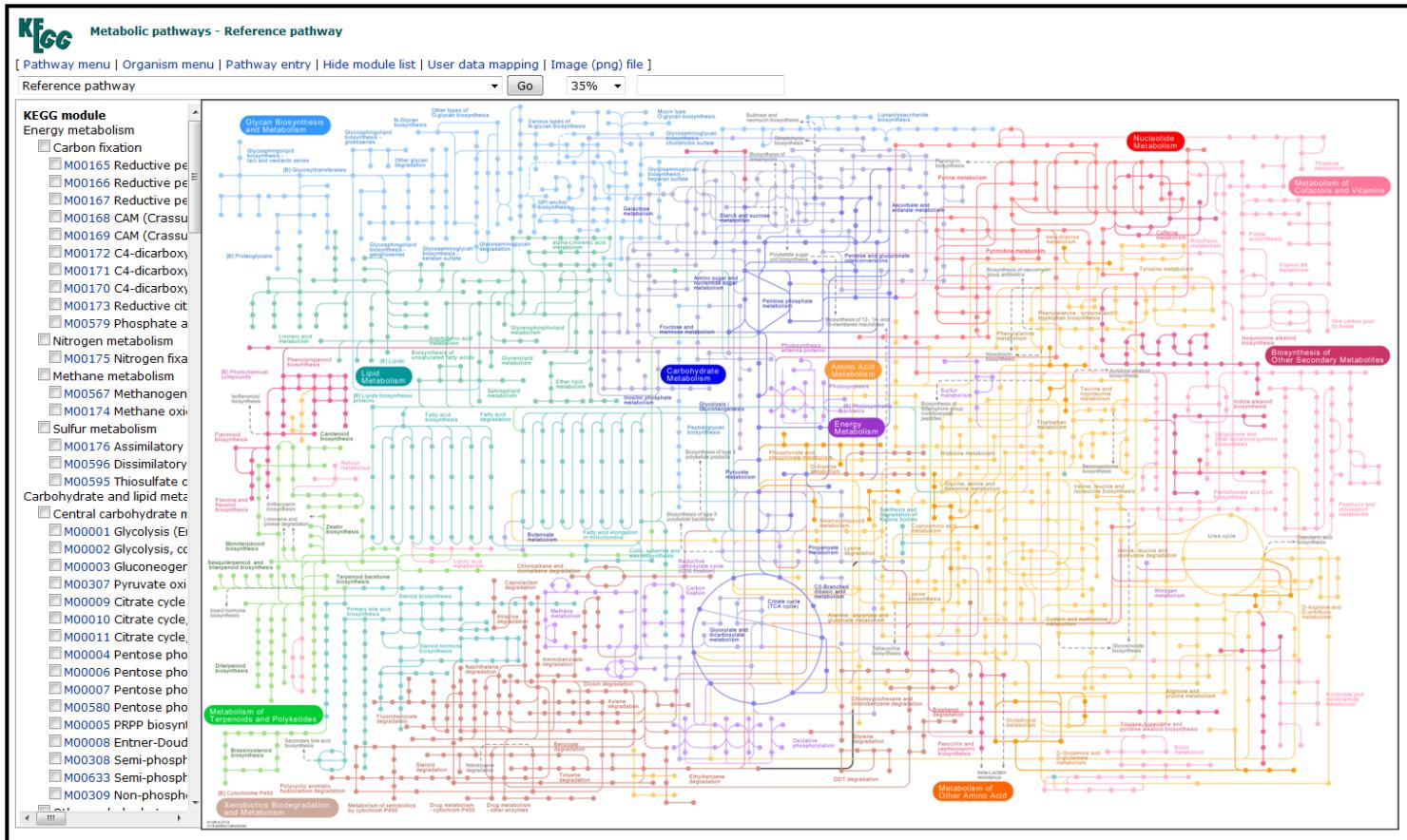


[bgm mode](#)

KEGG PATHWAY 実習4

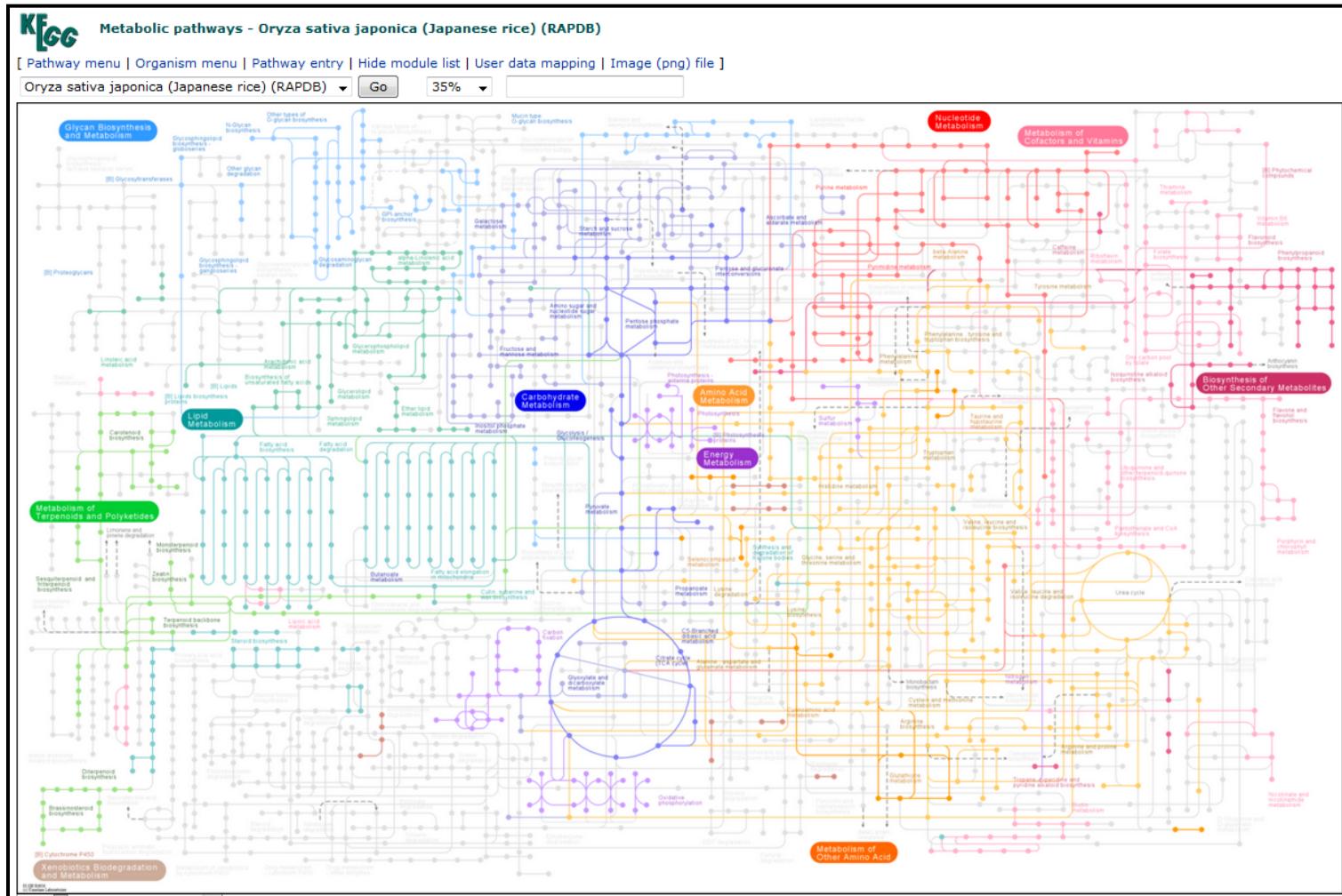
Overview マップを見てみよう

- PATHWAY top page の 1.0 Global and overview maps の [Metabolic pathways](#) をクリック
 - 右の [KEGG Atlus] は Java で動くビューワーで、自由度が少し高い分、動作が重たい
- 左にモジュールのリスト(KEGG におけるパスウェイの小さい機能単位)、右にマップが表示
- 機能単位毎にパスウェイを強調表示できる



KEGG PATHWAY 実習4

- 生物種毎の Overview マップを見てみよう
 - プルダウンメニューから生物を選択し、Go をクリック
 - 生物の持っていない経路は灰色になる



KEGG PATHWAY 実習5

アブラムシとブネラで、00290 : Valine, leucine and isoleucine biosynthesis の主観比較パスウェイを見てみよう

- アブラムシ : api
- ブネラ : buc

KEGG GENOMES

KEGG Mapping for Genome Comparison and Combination

KEGG for pangenomes or genome comparison/combination

KEGG species list - the list of KEGG organisms grouped into species
KEGG genus list - the list of KEGG organisms grouped into genus (and also family in eukaryotes)

Enter multiple organism codes:

(examples) [hsa eco](#) (genome comparison)
[hsa+T30003](#) (genome and metagenome combination)



KEGG Organism group: api+buc

Category info Brite hierarchy Taxonomy

Search genes:

api+buc

T01076	api	Acythosiphon pisum (pea aphid)
T00036	buc	Buchnera aphidicola APS (Acythosiphon pisum)

KEGG2 PATHWAY BRITE GENES GENOME LIGAND DISEASE DRUG

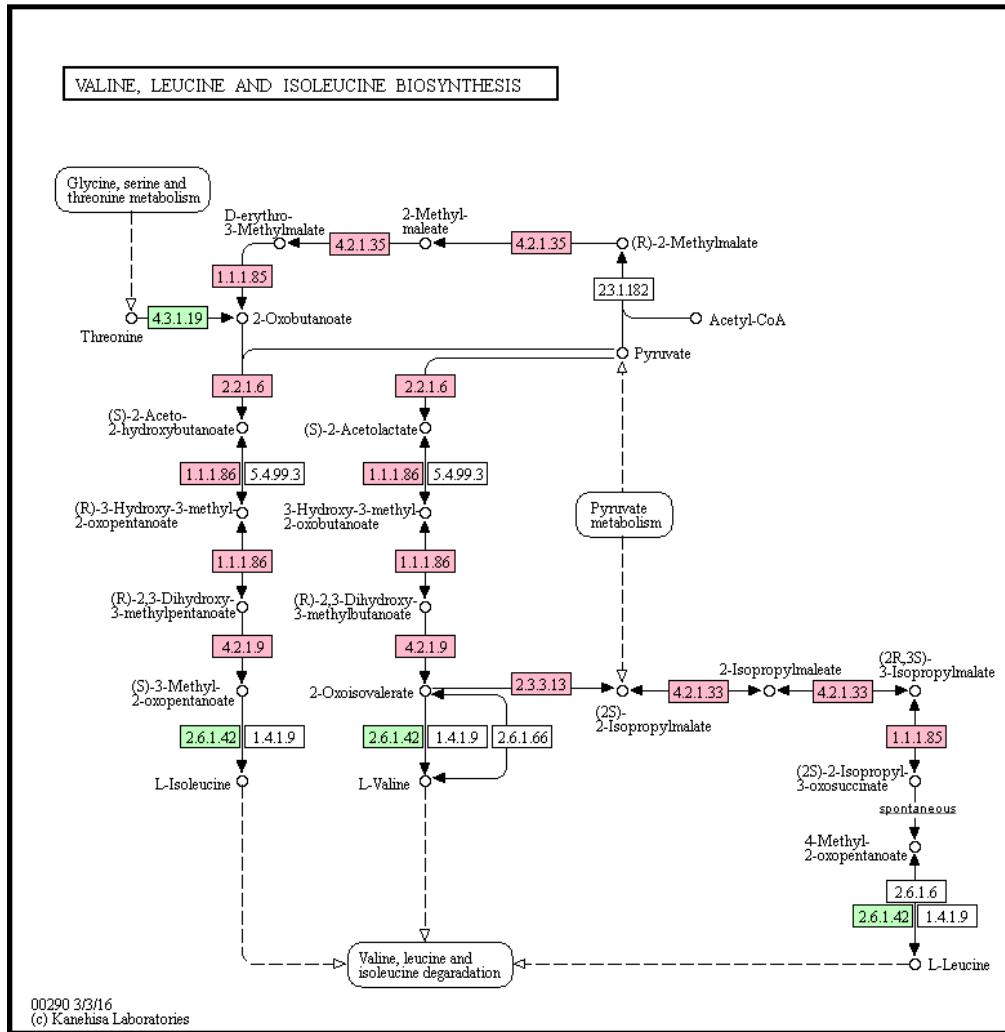


Lipid metabolism
00061 Fatty acid biosynthesis
00062 Fatty acid elongation
00071 Fatty acid degradation
00072 Synthesis and degradation of ketone bodies
00073 Cutin, suberin and wax biosynthesis
00100 Steroid biosynthesis
00120 Primary bile acid biosynthesis
00121 Secondary bile acid biosynthesis
00140 Steroid hormone biosynthesis
00561 Glycerolipid metabolism
00564 Glycerophospholipid metabolism
00565 Ether lipid metabolism
00600 Sphingolipid metabolism
00590 Arachidonic acid metabolism
00591 Linoleic acid metabolism
00592 alpha-Linolenic acid metabolism
01040 Biosynthesis of unsaturated fatty acids
Nucleotide metabolism
00230 Purine metabolism
00240 Pyrimidine metabolism
Amino acid metabolism
00250 Alanine, aspartate and glutamate metabolism
00260 Glycine, serine and threonine metabolism
00270 Cysteine and methionine metabolism
00280 Valine, leucine and isoleucine degradation
00290 Valine, leucine and isoleucine biosynthesis
00300 Lysine biosynthesis
00310 Lysine degradation
00220 Arginine biosynthesis
00330 Arginine and proline metabolism
00340 Histidine metabolism
00350 Tyrosine metabolism
00360 Phenylalanine metabolism
00380 Tryptophan metabolism
00400 Phenylalanine, tyrosine and tryptophan biosynthesis
Metabolism of other amino acids
00410 beta-Alanine metabolism
00430 Taurine and hypotaurine metabolism
00440 Phosphonate and phosphinate metabolism
00450 Selenocompound metabolism
00460 Cyanoamino acid metabolism
00471 D-Glutamine and D-glutamate metabolism
00472 D-Arginine and D-ornithine metabolism
00473 D-Alanine metabolism
00480 Glutathione metabolism

KEGG PATHWAY 実習5

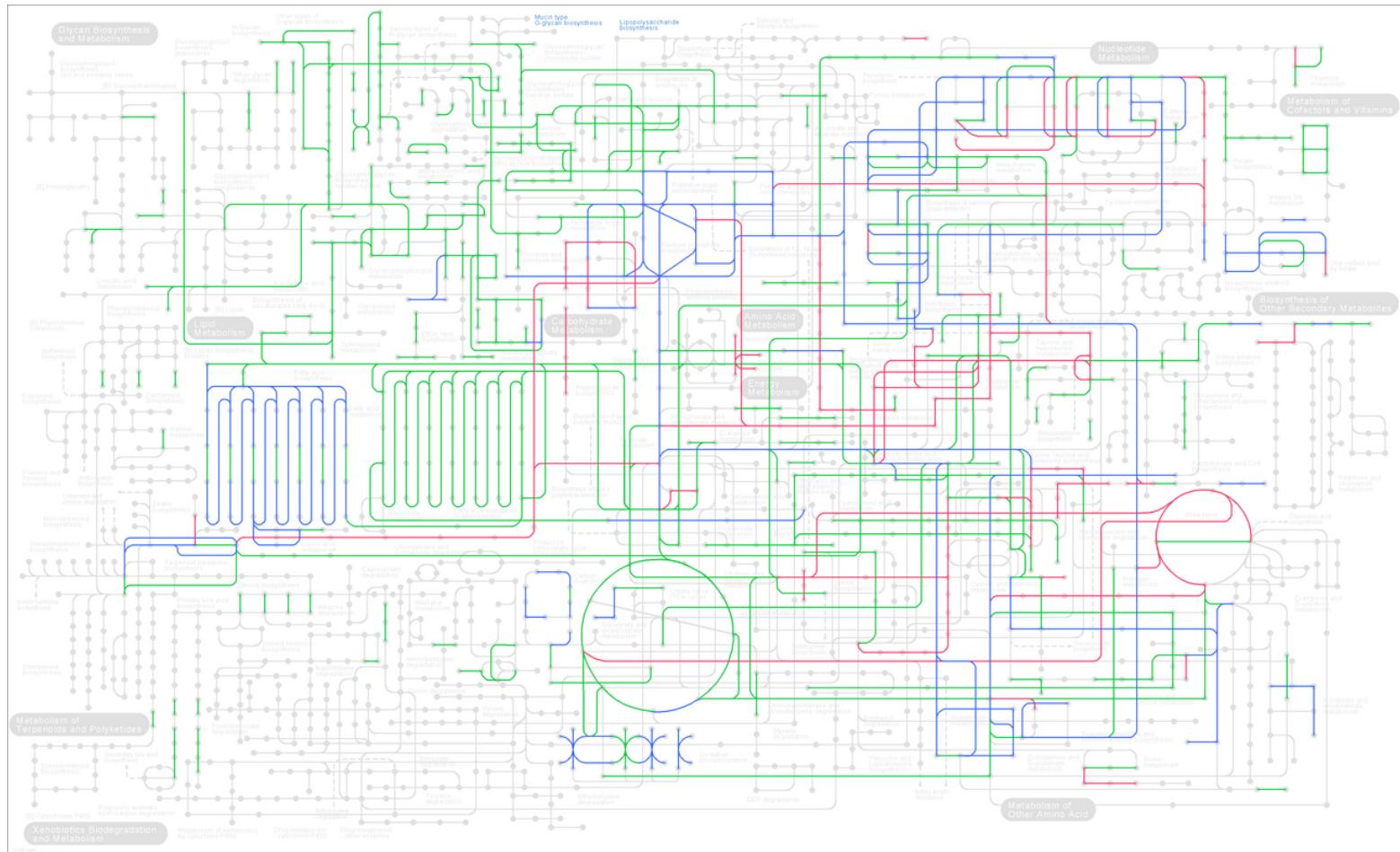
共生生物間のパスウェイ補完によってアミノ酸合成が可能になっていることがわかる

- アブラムシ : api(緑)、ブフネラ : buc(ピンク)
- 3生物以上でも比較可能



KEGG PATHWAY 実習5

Overviewマップで、共生生物間の補完関係を見ることができる



アブラムシ ブフネラ 共通

KEGG PATHWAY 実習6

サンプル・データをマッピングしてみよう

KEGG には遺伝子・化合物リストからパスウェイにマッピングするツールが組み込まれています

- [KEGG Mapper](#)

- Pathway mapping tool の2番目の [Search&Color Pathway](#) をクリック

- Search against : データベースコード
- Primary ID : ID 種類 (KEGG ID, NCBI-GeneID, NCBI-ProteinID, UniProt)
- テキストエリア : 要素のリスト(遺伝子、タンパク質、化合物)
 - [配列 ID or 代謝産物 ID] 塗りつぶし色[線の色]
 - 配列 ID は KEGG gene ID, NCBI-GenelD, NCBI-ProteinID, UniProt ID
 - 代謝産物 ID は KEGG Compound ID (C番号)のみ
 - 線の色はオプション
 - 色は16進数表記か基本的なカラーネームで記述

KEGG PATHWAY 実習6

- テキストエリア右の Example を選択して Exec ボタンをクリックすると、ヒットしたパスウェイのリストが表示されます（カッコの中はヒットした要素の数）

KEGG Mapper – Search&Color Pathway

Search against: hsa Enter: map, ko, ec, rn, hsadd, or org

Primary ID: KEGG identifiers (Outside IDs for organism-specific pathways only)

Enter objects one per line followed by bgcolor, fgcolor:

```
7167 red,blue
C00118 pink
```

Examples: Homo sapience pathway

Alternatively, enter the file name containing the data:

If necessary, change default bgcolor: pink

Include aliases

Use uncolored diagrams

Display objects not found in the search

Search pathways containing all the objects (AND search)

Exec **Clear**

Pathway Search Result

Sort by the pathway list

Show all objects

- hsa01230 Biosynthesis of amino acids - Homo sapiens (human) (2)
- hsa00051 Fructose and mannose metabolism - Homo sapiens (human) (2)
- hsa01100 Metabolic pathways - Homo sapiens (human) (2)
- hsa00562 Inositol phosphate metabolism - Homo sapiens (human) (2)
- hsa01200 Carbon metabolism - Homo sapiens (human) (2)
- **hsa00010 Glycolysis / Gluconeogenesis - Homo sapiens (human) (2)**
- hsa00750 Vitamin B6 metabolism - Homo sapiens (human) (1)
- hsa00052 Galactose metabolism - Homo sapiens (human) (1)
- hsa00030 Pentose phosphate pathway - Homo sapiens (human) (1)
- hsa00730 Thiamine metabolism - Homo sapiens (human) (1)
- hsa00900 Terpenoid backbone biosynthesis - Homo sapiens (human) (1)

GLYCOLYSIS / GLUCONEOGENESIS

Starch and sucrose metabolism

Pentose phosphate pathway

Carbon fixation in photosynthetic organisms

Citrate cycle

Pyruvate metabolism

Propanoate metabolism

00010 6/16/16
(c) Kanehisa Laboratories

Reactome(<http://www.reactome.org/>)

The screenshot shows the Reactome Pathway Database homepage. The top navigation bar includes links for About, Content, Documentation, Tools, Community, Download, and Contact. A search bar is located in the top right. Below the header are six main buttons: 'Browse Pathways', 'Analyze Data', 'Reactome FIViz app', 'User Guide', 'Data Download', and 'Contact Us'. To the right of these buttons is a 'Tweets' section displaying two tweets from the official Reactome Twitter account (@reactome). The first tweet is about a webinar on June 27, 2016. The second tweet is a general update. At the bottom of the page, there is a 'About Reactome' section with a brief description and logos for OICR, NYU Langone Medical Center, CSHL, and EMBL-EBI. A note at the bottom states that the development of Reactome is supported by grants from the US National Institutes of Health, Ontario Research Fund, and the European Molecular Biology Laboratory.

- 開発：
 - EMBLE-EBI 他
- 対象：
 - ヒトを中心に脊椎動物、酵母、植物、19種
- ヒト：
 - 専門家が手作業で作成した文献ベースのデータ
- その他：
 - 計算機での推定
 - 代謝パスウェイ、シグナル伝達系、他
- 利用：フリー
- データ形式：
 - BioPAX, SBML

PMN (<http://www.plantcyc.org/>)

- 開発:S. Rhee, [Department of Plant Biology, Carnegie Institution](#), USA
- 対象:350以上の植物情報からなる植物リファレンスパスウェイ+シロイヌナズナなど22種+外部6種
 - SRI International のPathway Tools Software を使用
 - 専門家が手作業で作成した文献ベースのデータ+自動ツール
 - 代謝パスウェイ、制御系
- 利用:アカデミックフリー
- データ形式:BioPAX

The screenshot shows the homepage of the Plant Metabolic Network (PMN). The header features a green navigation bar with links for Home, About PMN, PMN Data, and Help. Below the header is a search bar with a placeholder "Enter terms then hit Search..." and a magnifying glass icon. The main content area has a white background. On the left, there's a logo for "PMN" with a stylized flower icon and the text "Plant Metabolic Network". In the center, the title "PLANT METABOLIC PATHWAY DATABASES" is displayed in bold capital letters. Below the title, a paragraph states: "The PMN currently houses one multi-species reference database called PlantCyc and 22 species/taxon-specific databases." To the right, a "RECENT NEWS" section highlights the "PMN 11.0 RELEASED!" update from June 2, 2016, by admin. It includes a brief description of the release and a link to the new version. At the bottom of the page, there are links for "PlantCyc [More Information] [Content Statistics]" and "PMN Single-species/taxon Databases".

本日の内容

- パスウェイデータベース
 - パスウェイデータベースの概要
 - BioCyc の紹介・使い方
 - KEGG PATHWAY の紹介・使い方
 - 疾患研究への応用

KEGG DISEASE: 疾患データベース



KEGG DISEASE

分子ネットワークのゆらぎ状態としての疾患情報リソース

» English

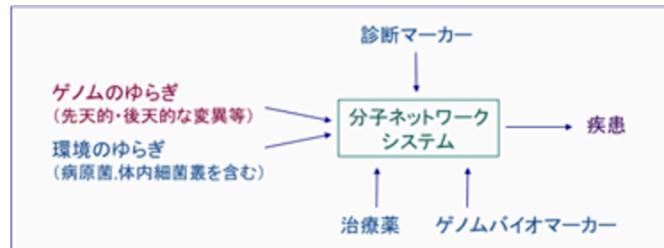
Menu PATHWAY BRITE DISEASE DRUG ENVIRON Cancer Pathogen MEDICUS

DISEASE を H番号、名称、概要、カテゴリ、パスウェイ、 病因遺伝子で検索

検索

KEGG DISEASE データベース

KEGG では、病気とは生体システムを司る分子ネットワークがゆらいだ状態であると、また医薬品は分子ネットワークへのゆらぎ物質であるとみなしています。単一遺伝子疾患、多因子性疾患、感染症疾患など様々な病気を以下を見方で統一的に記述しています。



ゆらいだ分子ネットワークに関する知識は疾患パスウェイマップとして表現され、KEGG PATHWAY データベースの一部として提供されています（例えば、慢性骨髄性白血病のパスウェイマップ [hsa05220](#)）。しかしながら、分子ネットワークの詳細が分かっているケースはまれであり、分子ネットワークのゆらぎの原因となる遺伝要因や環境要因（病原体を含む）に関する知識を蓄積したのが、KEGG DISEASE データベースです。各疾患エントリは H 番号で識別され、病因遺伝子、環境因子、診断マーカー、治療薬を含む遺伝子・分子リストで表現されています（例えば、慢性骨髄性白血病の疾患エントリ [H00004](#)）。遺伝要因既知の疾患と病原体ゲノム既知の感染症疾患有すべてこの形式で表現し、KEGG DISEASE の構築を行っています。

- ヒト疾患 [+ 遺伝子]
- 感染症疾患 [+ ゲノム]
- ICD-10 による疾患分類
- 感染症法による感染症分類
- 特定疾患 (難病)

ヒト疾患の階層分類

KfC ヒト疾患

[Brite menu | Download htext]

ヒト疾患

▼ ▼ ▼ One-click mode

▼ がん

- ▶ 眼、脳および中枢神経系のがん
- ▶ 消化器系のがん
- ▶ 造血組織およびリンパ系組織のがん
- ▶ 軟部組織および骨のがん
- ▶ 皮膚がん
- ▶ 乳房および女性生殖器のがん
- ▶ 男性生殖器のがん
- ▶ 泌尿器系のがん
- ▶ 内分泌器官のがん
- ▶ 頭頸部がん
- ▼ 肺および胸膜のがん
 - H00014 非小細胞肺癌 [PATH:hsa05223 hsa05206]
 - H00013 小細胞肺癌 [PATH:hsa05222 hsa05206]
 - H00015 悪性胸膜中皮腫

▶ 免疫系疾患

▶ 神経系疾患

▶ 循環器系疾患

▶ 呼吸器系疾患

▶ 内分泌代謝疾患

▶ 消化器系疾患

▶ 泌尿器系疾患

▶ 生殖系疾患

疾患エントリの例： 非小細胞肺癌

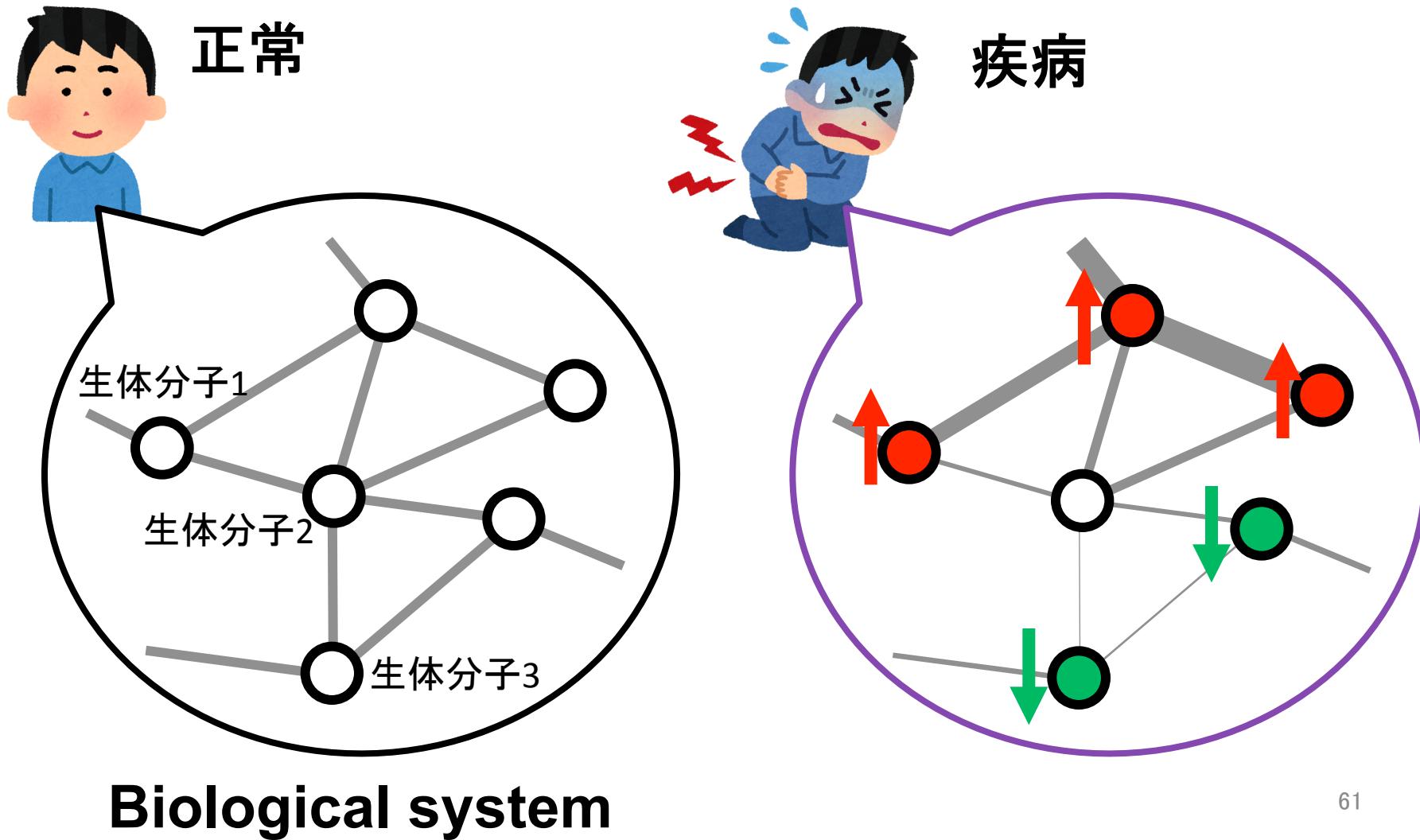
疾患に関する様々な情報を得
ることができる。

- 疾患関連パスウェイ
- 病因遺伝子
- 発がん物質
- 診断マーカー
- 環境因子
- 治療薬

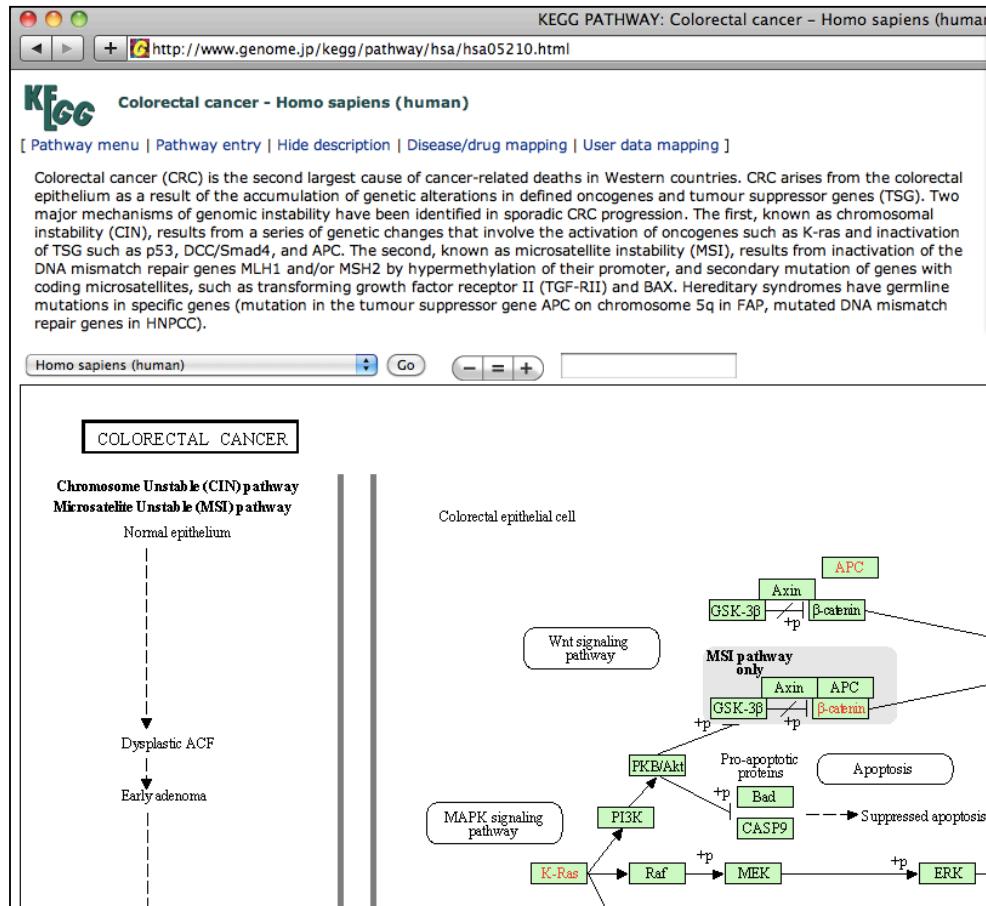
Kegg DISEASE: 非小細胞肺癌 Help

エントリ	H00014
名称	非小細胞肺癌; Non-small cell lung cancer
概要	肺がんは、工業化の進んだ国々において、男女ともにがんによる死亡の主要な原因となっている。非小細胞肺癌(NSCLC)は、肺がんのおよそ85%を占めており、主に扁平上皮癌(SCC)、腺癌(AC)、大細胞癌とから構成される。分子メカニズムの変化には、がん遺伝子であるK-RAS、EGFR、EML4-ALKの活性化と、がん抑制遺伝子であるp53、p16INK4a、RAR-beta、RASSF1などの不活性化が含まれる。K-RASの点突然変異は、GTPaseを不活性化させ、p21-RASタンパクは恒常に増殖シグナルを核に伝える。EGFRの突然変異や過剰発現は細胞の増殖を促進する。EML4とALKの融合はALKを恒常に活性化させ、これにより細胞の増殖、浸潤、アポトーシスの抑制が引き起こされる。p53の不活性化突然変異は細胞の増殖スピードを増大させ、アポトーシスを低下させる。p16INK4aはCDK4やCDK6と競合的に結合することによりCDK-cyclin-D複合体の形成を妨げる。p16INK4aの発現の低下はNSCLCに共通に見られる特徴である。RAR-betaはビタミンA依存に転写活性を増大させる核内受容体である。RASSF1Aは、RASエフェクターであるNore-1とヘテロダイマーを形成することができる。よって、RASSF1Aが失われると、RASの働きは増殖を促進する方に転換されるだろう。
カテゴリ	がん
階層分類	ヒト疾患 [BR:jp08402] がん 肺および胸膜のがん H00014 非小細胞肺癌 ICD-10による疾患分類 [BR:jp08403] 2. 新生物 (C00-D48) C30-C39 呼吸器及び胸腔内臓器の悪性新生物 C34 気管支及び肺の悪性新生物 H00014 非小細胞肺癌
BRITE hierarchy	
パスウェイ	hsa05223 非小細胞肺癌 hsa05206 microRNA とがん
病因遺伝子	K-ras (mutation) [HSA:3845] [KO:K07827] EGFR (mutation) [HSA:1956] [KO:K04361] RAR-beta (promoter hypermethylation) [HSA:5915] [KO:K08528] FHIT (mutation) [HSA:2272] [KO:K01522] RASSF1 (promoter hypermethylation) [HSA:11186] [KO:K09850] p16/INK4A (mutation) [HSA:1029] [KO:K06621] p53 (mutation/deletion) [HSA:7157] [KO:K04451] EML4-ALK (translocation) [HSA:27436 238] [KO:K15420 K05119] TFG-ALK (translocation) [HSA:10342 238] [KO:K09292 K05119] SLC34A2-ROS1 (translocation) [HSA:10568 6098] [KO:K14683 K05088]
発癌物質	Arsenic and arsenic compounds [CPD:C06269] Arsenic in drinking-water Asbestos [CPD:C16442] Chromium[VI] Mustard gas (Sulfur mustard)

疾患はパスウェイの異常？



疾患パスウェイ



- 癌に関する遺伝子の分子間ネットワーク図
- 病原因子となっている遺伝子に赤でマーク

- Disease/drug mapping
- 疾患遺伝子とドラッグターゲットのマッピング

疾患パスウェイの例：非小細胞肺癌

- ・ ガンに関与する遺伝子のネットワーク図
- ・ 赤字の遺伝子が疾患の病因遺伝子を示す



Non-small cell lung cancer - Homo sapiens (human)

[Pathway menu | Organism menu | Pathway entry | Download KGML | Show description | User data mapping]
 Homo sapiens (human) Go 100%

NON-SMALL CELL LUNG CANCER

Alveolar and bronchiolar epithelial cells
Bronchial epithelial cells

Alveolar and bronchiolar epithelial cell
/Bronchial epithelial cell

Atypical adenomatous hyperplasia

Gefitinib (Iressa)
Erlotinib (Tarceva)

Bronchial dysplasia

Primary adenocarcinoma

Primary squamous cell carcinoma

Metastatic adenocarcinoma

Metastatic squamous cell carcinoma

FHIT → ? → Reduced apoptosis
Cell-cycle progression

Ras signaling pathway
RASSF1 → **MST1** → Reduced apoptosis
NORE1A → **CyclinD1** → Proliferation

ErbB signaling pathway
K-Ras → **PI3K** → **PIP3** → **PDK1** → **PKB/Akt** → +p → **BAD**, **CASP9**, **Forkhead** → Antiapoptosis

Grb2 → **Sos** → **Ras** → **Raf** → +p → **MEK** → +p → **ERK** → DNA → **CyclinD1** → Proliferation

ER → **PLCγ** → **Ca²⁺** → **DAG** → **PKC** → Calcium signaling pathway

PI3K → **PIP3** → **PDK1** → +p → **PKB/Akt** → +p → **BAD**, **CASP9**, **Forkhead** → Antiapoptosis

JAK3 → +p → **STAT3δ** → Cell survival

Genetic alterations
Oncogene : K-Ras, EGFR, EML4-ALK
Tumor suppressors : RARβ, FHIT, RASSF1, INK4a/ARF, p53

Crizotinib (Xalkori)

p18^{INK4a} → Cell cycle
CDK4/6 → +p → **Rb** → DNA → G1/S progression

DNA damage → **p53** → DNA → **CyclinD1** → Proliferation
DNA damage → **p53** → DNA → **p21**, **GADD45**, **Bax**, **Bak**, **p43**, **POLK** → Uncontrolled proliferation, Increased survival, Genomic instability

p53 signaling pathway
Cell cycle

疾患パスウェイの例：非小細胞肺癌



Non-small cell lung cancer - Homo sapiens (human) + Disease/drug

[Pathway menu | Show description | Disease/drug list | User data mapping]

Homo sapiens (human) + Disease/drug

Go

100%

ピンクのボックスは何らかの疾患で病因遺伝子となっている遺伝子を示す
ライトブルーのボックスは何らかの疾患で医薬品のターゲットとなっている遺伝子

NON-SMALL CELL LUNG CANCER

Alveolar and bronchiolar epithelial cells
Bronchial epithelial cells

Atypical adenomatous hyperplasia
Bronchial dysplasia

Primary adenocarcinoma
Primary squamous cell carcinoma

Metastatic adenocarcinoma
Metastatic squamous cell carcinoma

Alveolar and bronchiolar epithelial cell /Bronchial epithelial cell

FHIT → Reduced apoptosis
Cell-cycle progression

Ras signaling pathway

ErbB signaling pathway

Overexpression

Genetic alterations

Oncogene : K-Ras, EGFR, EML4-ALK
Tumor suppressors : RAR β , FHIT, RASSF1, INK4a/ARF, p53

Retinoic acid

MST1 → Reduced apoptosis

CyclinD1 → Proliferation

PKB/Akt → Antiapoptosis
BAD, CASP9, Forkhead

PIP₃

PDK1

+p

ER

Ca²⁺

PIP₃

PDK1

+p

JAK3

+p

STAT3S

Raf → MAPK signaling pathway

MEK → ERK

+p

PIP₃

PDK1

+p

BAD, CASP9, Forkhead

PI3K-Akt signaling pathway

+p

STAT3S

→ Cell survival

RAR β → Tumour progression

Cell cycle

p19^{Ink4a}

CDK4/6

CyclinD1

+p

Rb

+

E2F

DNA

→

CyclinD1

→ Proliferation

MAPK signaling pathway

Cell cycle

p53 signaling pathway

Cell cycle

Uncontrolled proliferation
Increased survival
Genomic instability

p21, GADD45

Bax, Bak

p43, POLK

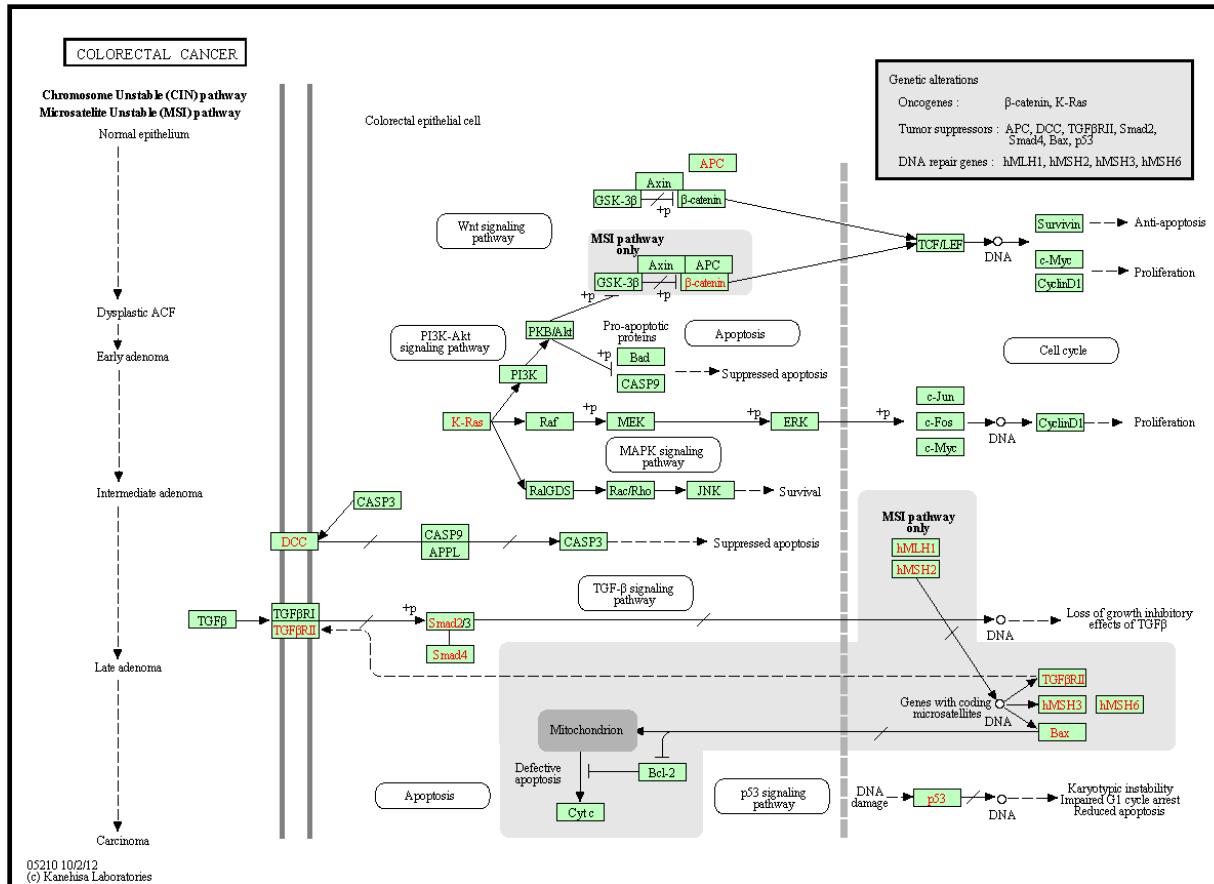
KEGG DISEASE 実習1

ヒトの疾患パスウェイを見てみよう

- がん、免疫系疾患、神経変性疾患など多因子性の疾患

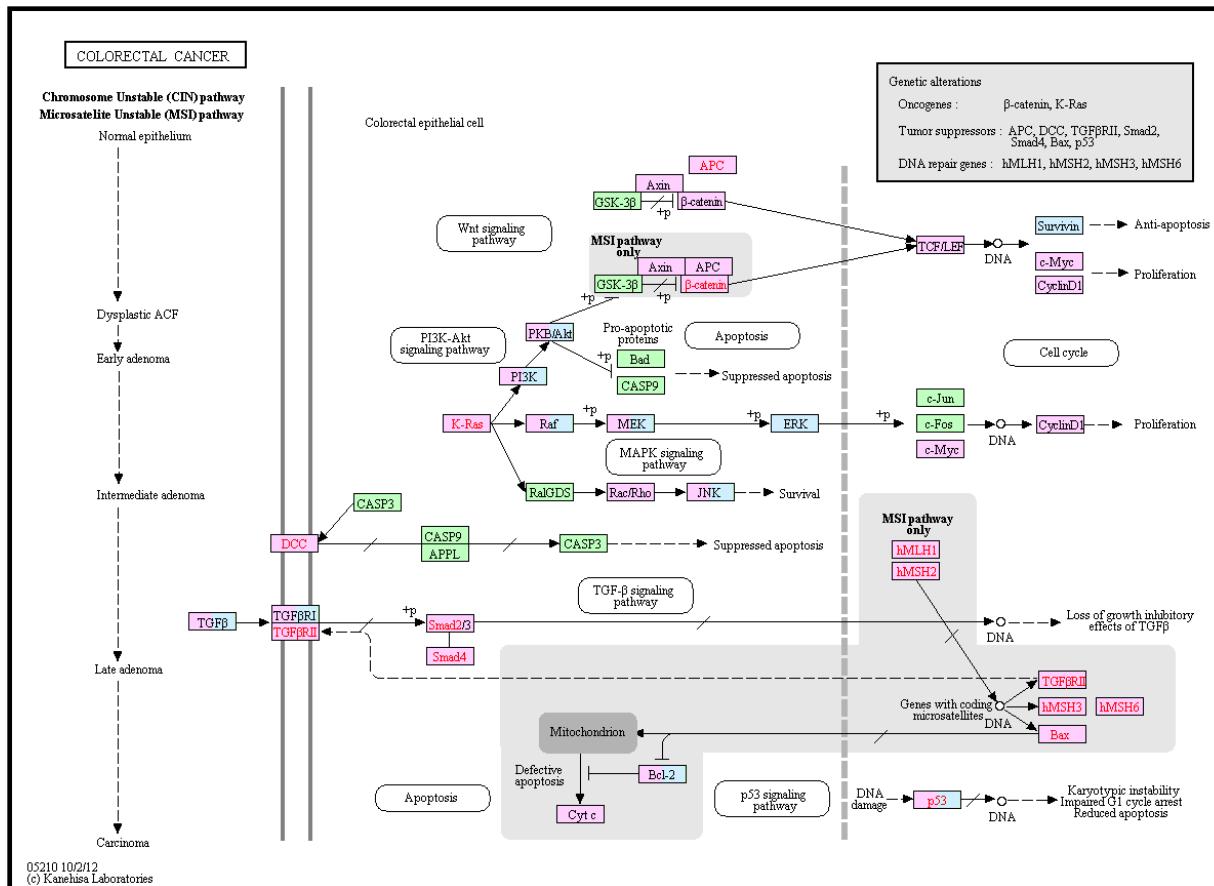
- 好きな疾患パスウェイをクリック(例: [大腸がん](#))

- 赤字の遺伝子が疾患の病因遺伝子を示しています



KEGG DISEASE 実習1

- プルダウンメニューから [Homo sapiens \(human\) + Disease/drug](#) を選択
 - ピンクのボックスは何らかの疾患で病因遺伝子となっている遺伝子を示しています
 - ライトブルーのボックスは何らかの疾患で医薬品のターゲットとなっている遺伝子を示しています



KEGG DISEASE 実習2

- 例えば、小細胞肺がんで発現が高い遺伝子と低い遺伝子をパスウェイにマッピングすることを考える。
- 以下のURLにアクセスし、遺伝子リストを得る。

<http://www.bioreg.kyushu-u.ac.jp/lab0/systemcohort/pathwayDB/>

Supplements

Pathway DB introduction

Differentially expressed genes in small cell lung cancer

- [H00013 gene high.txt \(a text file\)](#)
- [H00013 gene low.txt \(a text file\)](#)



hsa:6217
hsa:346653
hsa:8842
hsa:429
hsa:8972
hsa:102724473
hsa:145773
hsa:3772
hsa:2259
hsa:5995
hsa:3856
hsa:26047
hsa:114088
hsa:163115
hsa:6547
hsa:476
hsa:35
hsa:55859
hsa:9075
hsa:10231
hsa:284486
hsa:5122
hsa:5629

KEGG DISEASE 実習2

- KEGG Pathwayのページの下の方のPathway Mappingのセクションに行く
- Search Pathwayをクリック

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

Last updated: August 7, 2017

KEGG

GenomeNet

Kanehisa Laboratories

KEGG DISEASE 実習2

- Examplesで、Homo sapiens pathwayを選ぶ
- Enter objectsのフォームに、さきほどの遺伝子リストを入力(コピー&ペースト)
- Execで実行

KEGG Mapper – Search Pathway

About KEGG Mapper

- ▶ Search Pathway
- Search&Color Pathway
- Color Pathway
- Color Pathway WebGL

Search Brite

- Search&Color Brite
- Join Brite
- Join Brite Table

Search Module

- Search&Color Module

Search Disease

Reconstruct Pathway

- Reconstruct Brite
- Reconstruct Module
- Map Taxonomy

Convert ID

Annotate Sequence

BlastKOALA

Search against: hsa Enter: map, ko, ec, rn, hsadd, or org

Enter objects:

hsa:7167 hsa:GPI cpd:C00118
ALDOA 1.2.1.12 C00236

Examples: Homo sapiens pathway

Alternatively, enter the file name containing the data:

ファイルを選択 選択されていません

Filter1 Filter2 (to extract K/C/G/D/R/RP/RC numbers)

Include aliases

Display objects not found in the search

Search pathways containing all the objects (AND search)

Exec **Clear**

Pathway Search Result

Following object(s) was/were not found hsa:102724473 hsa:1033 hsa:10592 hsa:10763 hsa:1100 hsa:145773 hsa:148213 hsa:157313 hsa:1611 hsa:162681 hsa:163115 hsa:1837 hsa:1996 hsa:2000 hsa:27075 hsa:284486 hsa:285671 hsa:29106 hsa:29114 hsa:2922 hsa:3148 hsa:3149 hsa:3150 hsa:415 hsa:429 hsa:4321 hsa:4648 hsa:4745 hsa:4922 hsa:5083 hsa:50861 hsa:51046 hsa:51047 hsa:57544 hsa:57549 hsa:58516 hsa:5865 hsa:5995 hsa:64919 hsa:7001 hsa:7033 hsa:709 hsa:83856 hsa:84275 hsa:8445 hsa:84707 hsa:8577 hsa:8618 hsa:8715 hsa:8927 hsa:9118

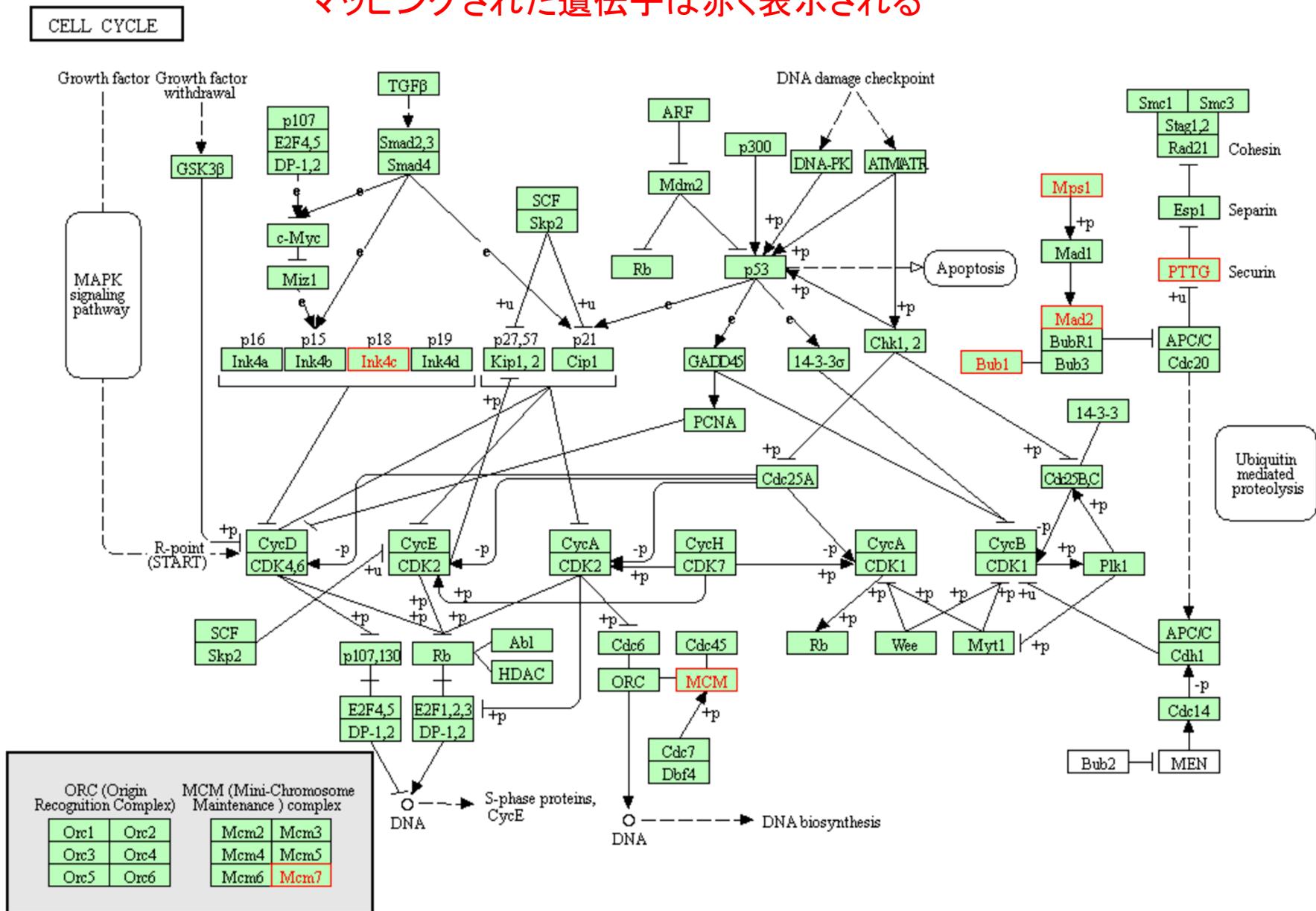
[Sort by the pathway list](#)

[Show all objects](#)

- [hsa01100 Metabolic pathways - Homo sapiens \(human\) \(10\)](#)
- [hsa04110 Cell cycle - Homo sapiens \(human\) \(6\)](#)
- [hsa01200 Carbon metabolism - Homo sapiens \(human\) \(4\)](#)
- [hsa05166 HTLV-I infection - Homo sapiens \(human\) \(3\)](#)
- [hsa04514 Cell adhesion molecules \(CAMs\) - Homo sapiens \(human\) \(3\)](#)
- [hsa05200 Pathways in cancer - Homo sapiens \(human\) \(3\)](#)
- [hsa04114 Oocyte meiosis - Homo sapiens \(human\) \(3\)](#)
- [hsa04971 Gastric acid secretion - Homo sapiens \(human\) \(2\)](#)
- [hsa04973 Carbohydrate digestion and absorption - Homo sapiens \(human\) \(2\)](#)
- [hsa04151 PI3K-Akt signaling pathway - Homo sapiens \(human\) \(2\)](#)
- [hsa04210 Apoptosis - Homo sapiens \(human\) \(2\)](#)
- [hsa05206 MicroRNAs in cancer - Homo sapiens \(human\) \(2\)](#)
- [hsa01230 Biosynthesis of amino acids - Homo sapiens \(human\) \(2\)](#)
- [hsa04371 Apelin signaling pathway - Homo sapiens \(human\) \(2\)](#)
- [hsa03060 Protein export - Homo sapiens \(human\) \(2\)](#)

Cell cycleパスウェイの例

マッピングされた遺伝子は赤く表示される



KEGG DRUG: 医薬品データベース



KEGG DRUG

分子ネットワークへのゆらぎ物質としての医薬品情報リソース

» English

Menu PATHWAY BRITE DISEASE DRUG ENVIRON Cancer Pathogen MEDICUS

DRUG を D番号、一般名(英名を含む)、商品名、 成分で検索

DGROUP を DG番号、名称で検索

KEGG DRUG データベース

KEGG DRUG は、日本、米国、欧州の医薬品情報を化学構造と成分の観点から一元的に集約し、ターゲットや代謝酵素をはじめとした分子間相互作用ネットワークの情報を附加したデータベースです。日本の医薬品は医療用(処方薬)だけでなく一般用(大衆薬、OTC薬)もすべて網羅し、個々の製品情報(添付文書情報)へのリンクがつけられています。また化学薬品だけでなく、生薬・漢方方剤もデータベース化されています。

KEGG DRUG のエントリは D 番号で識別され、化学薬品は化学構造をベースに、混合物や生薬等は成分をベースに、エントリ化がなされています。各エントリには以下の情報が蓄積されています(例えば、慢性骨髄性白血病治療薬グリベック D01441 を参照)。

- 一般名(化学構造を反映した名称、詳細は下記参照)
- 日本の商品名(JAPIC が提供する添付文書へのリンクがつけられています)
- 米国の商品名(DailyMed が提供する添付文書へのリンクがつけられています)
- 化学構造、混合物などの化学成分、ペプチドやポリケチドなどの配列
- KEGG DGROUP で定義された医薬品クラス
- 薬効分類番号、ATC コード、KEGG DGROUP で定義された Chemical グループ
- 効能、適応疾患
- ターゲット分子(関連する KEGG PATHWAY 情報が提供されています)
- 薬物代謝に関与する酵素、トランスポーター
- その他の相互作用分子として、ゲノムバイオマーカー、CYP誘導・阻害薬など
- 薬物間相互作用(添付文書より抽出した薬物間相互作用のデータを検索できます)
- 薬開発の歴史など医薬品化学者の知識を化学構造変化パターンで表現した構造マップ
- BRITE 機能階層としての医薬品分類情報
- 外部データベースへのリンク

医薬品分類

KEGG では医薬品を様々な観点で分類し体系化して、BRITE データベースの一部として提供しています。まず、国や国際機関が提供する分類(生薬等は KEGG 独自の分類)に KEGG DRUG のD番号エントリを割り当てたものは以下の通りです。ここで追加情報を含むバージョンは、KEGG が提供するターゲット・薬物代謝酵素の情報や、JAPIC が提供する添付文書の情報を、医薬品分類の階層にのせて検索・閲覧できるようにしたもので、情報は追加のカラムまたは追加の階層に表示されます。

区分	分類	追加情報				備考
医療用医薬品 (日米欧)	日本の薬効分類	ターゲット	代謝酵素	添付文書	商品	JP17 (H28.3.7)
	WHO の ATC 分類	ターゲット	代謝酵素	添付文書	商品	ATC/DDD 2017
	米国 USP の分類	ターゲット	代謝酵素			v6.0 (March 2014)
一般用医薬品 (日本のみ)	一般用医薬品のリスク区分			添付文書	H28.11.1	
	一般用医薬品の分類			添付文書		
生薬等 (日本のみ)	漢方方剤・生薬の効能分類	KEGG ENVIRON 参照			局方 (JP17)	
	生薬の由来に基づく分類	局外生規 (non-JP) (H28.4.1)				
添加物 (日本のみ)	医薬品添加物				薬添規 1998 (H24.12.4)	

これらを補完するために KEGG では様々な分類を開発中で、現時点では以下が提供されています。

区分	分類	備考
ターゲット分類	ターゲットに基づく医薬品分類	
	抗感染症薬	
	抗悪性腫瘍薬	
	抗菌薬	
	抗ウイルス薬	
	抗真菌薬	
	抗寄生虫薬	
	糖尿病治療薬	
	脂質異常症治療薬	
	骨粗鬆症治療薬	
治療薬分類	循環器用薬	適応症を含む
	精神性疾患用薬	
	腎臓病用薬	

KEGG DRUG 実習1

- KEGG DRUGのページで、好きな医薬品名を入力(例えば、ゲフィチニブ)
- 検索をクリック



KEGG DRUG

分子ネットワークへのゆらぎ物質としての医薬品情報リソース

» English

Menu PATHWAY BRITE DISEASE DRUG ENVIRON Cancer Pathogen MEDICUS

DRUG を D番号、一般名(英名を含む)、商品名、 成分で検索

ゲフィチニブ

検索

DGROUP を DG番号、名称で検索

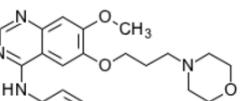
検索

医薬品エントリの例： ゲフィチニブ

医薬品に関する様々な情報を得
ることができる。

- 化学構造
- 効能
- 標的分子
- 標的パスウェイ
- 代謝酵素
- 薬物間相互作用

Kegg DRUG: ゲフィチニブ Help

エントリ	D01977	Drug
一般名	ゲフィチニブ (JAN); Gefitinib (JAN/USAN/INN)	
商品名	イレッサ (アストラゼネカ)	商品一覧
米国の商品	IRESSA (AstraZeneca Pharmaceuticals LP)	
組成式	C22H24ClFN4O3	
質量	446.1521	
分子量	446.9024	
構造式	 D01977	Mol file KCF file DB search Jmol KegDraw
クラス	抗悪性腫瘍薬 DG01918 チロシンキナーゼ阻害薬 DG01917 受容体型チロシンキナーゼ阻害薬 Cyp 基質 DG01644 CYP2D6により代謝される薬剤 DG01633 CYP3Aにより代謝される薬剤 Cyp 阻害 DG01645 CYP2D6阻害薬	
コード	薬効分類: 4291 ATCコード: L01XE02	
効能 疾患	抗悪性腫瘍薬, 受容体型チロシンキナーゼ阻害薬 非小細胞肺癌 (EGFR遺伝子変異陽性) [DS:H00014]	
ターゲット パスウェイ	EGFR [HSA:1956] [KO:K04361] hsa04010 MAPK シグナル伝達経路 hsa04012 ErbB シグナル伝達経路 hsa04020 カルシウムシグナル伝達経路 hsa04060 サイトカインとサイトカインレセプターの相互作用 hsa05200 がんのパスウェイ hsa05223 非小細胞肺癌	
代謝	Enzyme: CYP3A4 [HSA:1576], CYP2D6 [HSA:1565]	
相互作用	Genomic biomarker: EGFR [HSA:1956] CYP inhibition: CYP2D6 [HSA:1565] Transporter inhibition: ABCG2 [HSA:9429]	DDI search
構造マップ	map07045 抗腫瘍薬 - プロテインキナーゼ阻害剤	
他のマップ	map01521 EGFR チロシンキナーゼ阻害薬耐性	
階層分類	医療用医薬品の薬効分類 [BR:jp08301] 4 組織細胞機能用医薬品 42 腫瘍用薬	

化合物データベース ~Introduction~

- 歴史は長いが、有償の文化
 - Chemical Abstracts (1907～) [<http://www.cas-japan.jp/>]
 - ACS(米国化学会)の下部組織、CASによる発行(有償)
 - 化学関連文献・特許、化学構造・反応など総合的な情報を含む
 - CAS登録番号(CAS RN)は化合物IDとして事実上の標準
 - 冊子体、CD、オンライン検索がある
 - 現在、SciFinderによるオンライン検索が主体
 - The Combined Chemical Dictionary (CRC Press)
 - 下記4種の辞典を収載
 - Dictionary of Natural Products (DNP)
 - Dictionary of Organic Compounds
 - Dictionary of Inorganic/Organometallic Compounds
 - Dictionary of Drugs
 - 冊子体、CD(以上DNPのみ)、オンライン検索がある
 - CHEMnetBASEによるオンライン検索 (<http://www.chemnetbase.com/>)
 - 日本では、いずれも化学情報協会(JAICI)の取扱い
- 遺伝子データベースの発展とともに、化合物データベースにおいても様々な無償のデータベースが作られる

化合物データベース

KEGG LIGAND	www.genome.jp/kegg/ligand.html	京大・金久研。代謝産物を中心とした化合物、代謝反応など。
KNAPSAcK	kanaya.naist.jp/KNAPSAcK/	奈良先端大・金谷研。天然物・生物情報の対応が軸。植物中心
J-GLOBAL	jglobal.jst.go.jp/advancedsearch/#t=4	JST。日化辞(有機化合物DB)のWeb検索が可能(旧・日化辞Web)。
DNP	dnp.chemnetbase.com/	CRC出版。天然物DB。詳細閲覧は有償。
UNPD	pxuxxjpku.edu.cn/UNPD/	北京大。天然物DB。件数多いが、出典不明なものも。
ChEBI	www.ebi.ac.uk/chebi/	EBI。化合物分類DB。分類体系は、現状化学構造寄り。
PubChem	pubchem.ncbi.nlm.nih.gov/	米国・NIH。化合物など。他の化合物DBのレポジトリ的性質。件数膨大。
ChemSpider	www.chemspider.com/	400以上の化合物DBのIDを串刺し検索できるサイト。データ更新遅め。
HMDB	www.hmdb.ca/	カナダ・アルバータ大。ヒトメタボロームDB。
ECHDB	www.ecmdb.ca/	同上。大腸菌メタボロームDB。
YMDB	www.ymdb.ca/	同上。酵母メタボロームDB。
DrugBank	www.drugbank.ca/	同上。薬物のDB。
KEGG DRUG	www.kegg.jp/kegg/drug/	京大・金久研。医薬品DB。構造分類、添付文書とのリンクなども。
LIPIDMAPS	www.lipidmaps.org/	米国脂質学会。脂質DB。ポリフェノールなど含む。
Flavonoid Database	metabolomics.jp/wiki/Category:FL	metabolomics.jp(遺伝研・有田研)内にあるWikiベースのフラボノイドDB。Flavonoid Viewer のデータをベースにWiki化。

化合物構造を表記する書式

- 行列表記
 - mol (MDL/molfile)
 - 結合を結合表としてあらわす。二次元構造式の標準形式
 - sdf (sdf file)
 - mol 形式を内包し1ファイルに複数化合物を記述可能にしたもの
 - 詳しくは CTfile format (最新版は BIOVIAから) を参照
- 線形表記
 - InChI、InChIKey
 - IUPACとNISTが開発、現在はIUPAC管理、IUPACとInChI Trustで維持
 - standard InChIでは、構造式から一意な InChI文字列生成可能
 - ほとんどの情報がヒト可読
 - InChIKey(hashed InChI)はWeb検索用、人には読めず、非常に稀に重複可能性あり
 - SMILES
 - 分子の二次元構造を文字列として記述
 - 情報をコンパクトに保存
 - 原子座標の羅列と違い、ユーザーにも理解しやすい
 - canonical SMILESで一つの構造に対して唯一となるSMILES表記を定めることができる

KEGG LIGAND データベース

(<http://www.genome.jp/kegg/ligand.html>)

KEGG LIGANDの
サブカテゴリ

Chemical Substances and Reactions

KEGG LIGAND contains our knowledge on the universe of chemical substances and reactions that are relevant to life. It is a composite database consisting of COMPOUND, GLYCAN, REACTION, RPAIR, RCLASS, and ENZYME databases, whose entries are identified by C, G, R, RP, RC, and EC numbers, respectively. ENZYME is derived from the IUBMB/IUPAC Enzyme Nomenclature, but the others are internally developed and maintained.

Database	Identifier	Content	Specialized entry point
COMPOUND	C number	Chemical compound structures	KEGG COMPOUND
GLYCAN	G number	Glycan structures	KEGG GLYCAN
REACTION	R number	Biochemical reactions	KEGG REACTION
RPAIR	RP number	Reactant pair alignments	
RCLASS	RC number	Reaction class	
ENZYME	EC number	Enzyme nomenclature	

Search LIGAND for Go Clear

bfind mode bget mode

化合物、糖鎖、反応
に特化した入口

ID番号のIdentifierと
各DBに含まれるデータの説明

KEGGにおける「Chem(o)-」の部分、すなわち生化学的な情報全般を扱う
今回は、COMPOUNDを紹介します。

KEGG COMPOUND 実習

- 例: phenylalanine -

The figure consists of three screenshots of the KEGG LIGAND Database interface, each showing a different step in the search process for Phenylalanine.

Screenshot 1 (Left): Shows the main KEGG LIGAND Database search page. A red box highlights the search bar at the top where "Phenylalanine" is entered. Below the search bar is a table of search filters:

Database	Identifier	Content	Specialized entry point
COMPOUND	C number	Chemical compound structures	KEGG COMPOUND
GLYCAN	G number	Glycan structures	KEGG GLYCAN
REACTION	R number	Biochemical reactions	KEGG REACTION
PAIR	RP number	Reactant pair alignments	
RCLASS	RC number	Reaction class	
ENZYME	EC number	Enzyme nomenclature	

Below the table are two search buttons: "bfind mode" and "bget mode".

Screenshot 2 (Middle): Shows the search results for "Phenylalanine" in the LIGAND database. The results list various forms of Phenylalanine, including L-Phenylalanine, D-Phenylalanine, and their alpha and beta isomers. The first result, C00079 (L-Phenylalanine), is highlighted with a red box.

Screenshot 3 (Right): Shows the search results for "Phenylalanine" in the COMPOUND database. The results list numerous metabolites containing Phenylalanine, such as L-Phenylalanine, N-Acetyl-L-phenylalanine, and various forms of D-Phenylalanine. The first result, C00079 (L-Phenylalanine), is also highlighted with a red box.

KEGG COMPOUND 実習

～ KEGG COMPOUND Entry ～

Firefox - KEGG COMPOUND: C00079

www.kegg.jp/dbget-bin/www_bget?cpd:C00079

COMPOUND: C00079

Entry C00079 Compound

Name L-Phenylalanine; (S)-alpha-Amino-beta-phenylpropionic acid

Formula C9H11NO2

Exact mass 165.079

Mol weight 165.1891

Structure

Remark Same as: D00021
BRITE hierarchy

Reaction R00686 R00687 R00688 R00689 R00690 R00691 R00692 R00693
R00694 R00695 R00697 R00698 R00699 R01375 R01376 R01795
R03680 R06744 R07211 R08435 R08463 R08652 R08690 R09535
R09578 R09579

Pathway ko00360 Phenylalanine metabolism
ko00400 Phenylalanine, tyrosine and tryptophan biosynthesis
ko00940 Phenylpropanoid biosynthesis
ko00960 Tropane, piperidine and pyridine alkaloid biosynthesis
ko00966 Glucosinolate biosynthesis
ko00970 Aminocycl-(RNA biosynthesis
map01060 Biosynthesis of plant secondary metabolites
map01061 Biosynthesis of phenylpropanoids
map01063 Biosynthesis of alkaloids derived from shikimate pathway
map01064 Biosynthesis of alkaloids derived from ornithine, lysine and
nicotinic acid
map01070 Biosynthesis of plant hormones
ko01100 Metabolic pathways
ko01110 Biosynthesis of secondary metabolites
ko02010 ABC transporters
ko04974 Protein digestion and absorption
ko04978 Mineral absorption

Enzyme 1.4.1.20 1.4.3.2 1.11.1.7 1.11.1.21
1.13.12.9 1.14.13.124 1.14.16.1 2.3.1.53
2.6.1.1 2.6.1.5 2.6.1.9 2.6.1.28
2.6.1.57 2.6.1.58 2.6.1.64 2.6.1.70
2.7.7.54 3.4.17.23 4.1.1.28 4.1.1.53
4.2.1.51 4.2.1.91 4.3.1.24 4.3.1.25
5.1.1.11 6.1.1.20

Other DBs CAS: 63-91-2
PubChem: 3379
ChEBI: 17295
KNApSAck: C00001386
PDB-CCD: PHE
3DMET: B01151
NIKKAJI: J9.175H

KCF data Show

* Japanese version

DBGET integrated database retrieval system

構造表示、
構造情報(mol, KCF)、
構造検索、構造表示(&編集)、

各種データベースへのリンク
REACTION, PATHWAY, ENZYME

外部データベースへのリンク
PubChem, ChEBI, etc.

LinkDB情報
GenomeNet内部でリンクされているDBや
対応関係のとれる外部DBを表示

ゲノムネットケミカル情報解析ツール

KEGG

varDB

研究支援データベース

計算ツール

その他のツール

FTP

フィードバック

KEGG BRITE - システム情報: オントロジー

KEGG Organisms - 生物種ごとの入口

KEGG GENES - ゲノム情報

KEGG LIGAND - ケミカル情報

Reaction Ontology: 反応パターンと分類

varDB: 抗原変異データベース

研究支援データベース

CYORF - シアノバクテリア

BSORF - 枯草菌

EXPRESSION - 遺伝子発現プロファイル

ゲノムネット計算ツール

配列解析

BLAST / FASTA - ホモロジー検索

MOTIF - モチーフ検索

CLUSTALW / MAFFT / PRRN - 複数アライメント

ゲノム情報解析

OC Viewer - オーソロジックラスター

KAAS - KEGG自動アノテーションサーバー

EGassembler - ESTコンティング作成

GENIES - 遺伝子ネットワーク予測

DINIES - 医薬品-標的ネットワーク予測 *New!*

ケミカル情報解析

SIMCOMP / SUBCOMP - 化合物構造検索

KCam - 糖鎖構造検索

PathComp - 可能な反応経路の計算

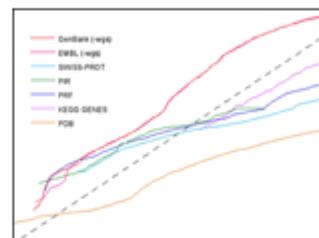
PathSearch - 類似反応経路検索

PathPred - 分解・合成反応経路予測

E-zyme - 化合物間の酵素反応予測



データベース間のリンク



データベース増加図

化学構造比較、パスウェイ予測など

京都大学化学研究所バイオインフォマティクスセンター

<http://www.genome.jp/> (英語)

<http://www.genome.jp/ja/> (日本語)

GenomeNet 実習

GenomeNetからの化合物データベース一括検索

統合データベースを選択しキーワードを
入力するとメニューに表示されている
データベース全てに対する検索となる。

英 語: <http://www.genome.jp/>
日本語: <http://www.genome.jp/ja/>

The screenshot shows two browser windows. The left window is the main GenomeNet homepage with various links like 'GenomeNet', 'DBGET', 'KEGG', 'varDB', etc. The right window is a search results page titled 'DBGET Search Result: All databases...'. It shows a search for 'Phenylalanine' across multiple databases. The results include entries from KEGG, PubChem, and CHEBI. A large blue callout box highlights the search interface, stating: '統合データベースを選択しキーワードを入力するとメニューに表示されているデータベース全てに対する検索となる。' (When you select a unified database and enter a keyword, it will perform a search across all databases listed in the menu.)