

パスウェイデータベースおよび プロテオームデータベースの紹介

奥田修二郎

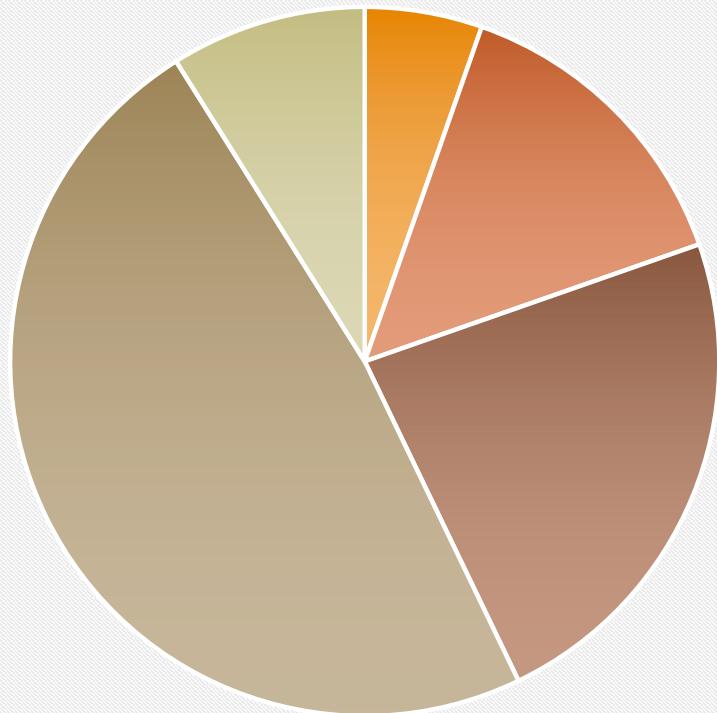
新潟大学大学院医歯学総合研究科

AJACS浜松@浜松医科大学臨床講義棟大講義室

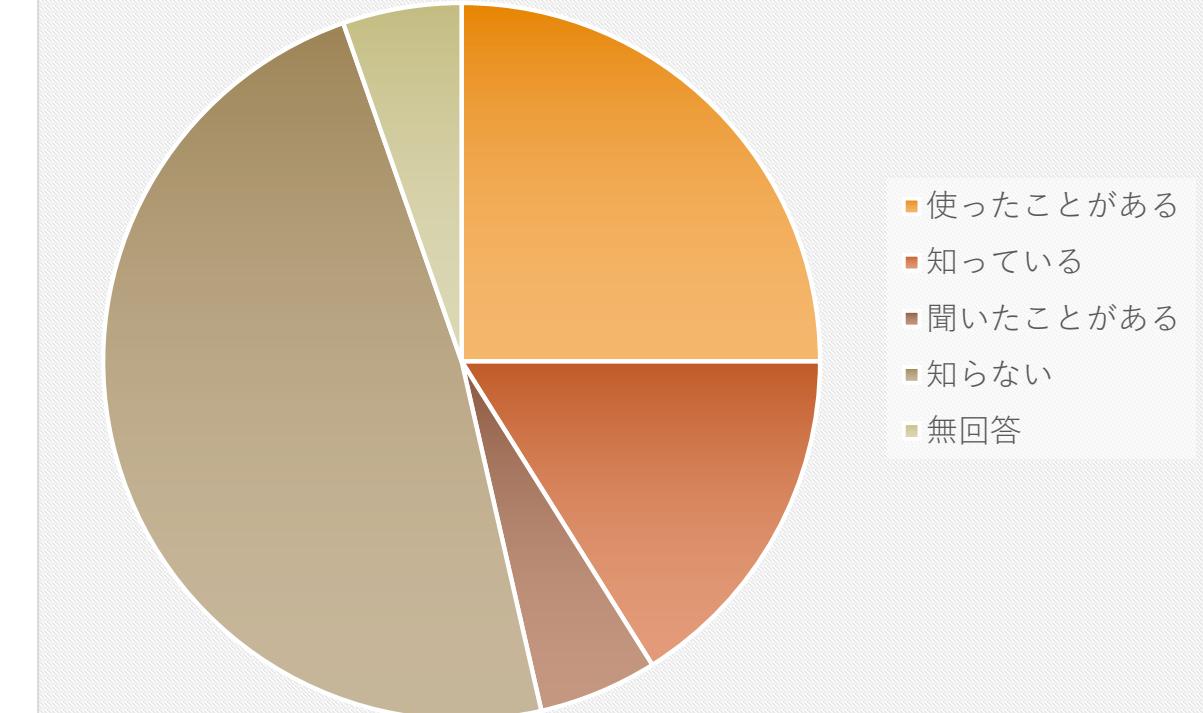
2018年1月16日

事前アンケート

【設問7-1】 パスウェイ関連のデータベース使ったことがありますか。



【設問7-3】 KEGGを使ったことがありますか。



パスウェイデータベースとは

- パスウェイとは、生体内での遺伝子やタンパク質、その他の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。相互作用の知識を集積して可視化、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
 - ベーリングガーマンハイム社（現ロシュ社）の代謝マップ (1965~)
 - 最初は紙媒体での出版。のちに電子化されています
- 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、生命現象の理解が行い易くなります。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

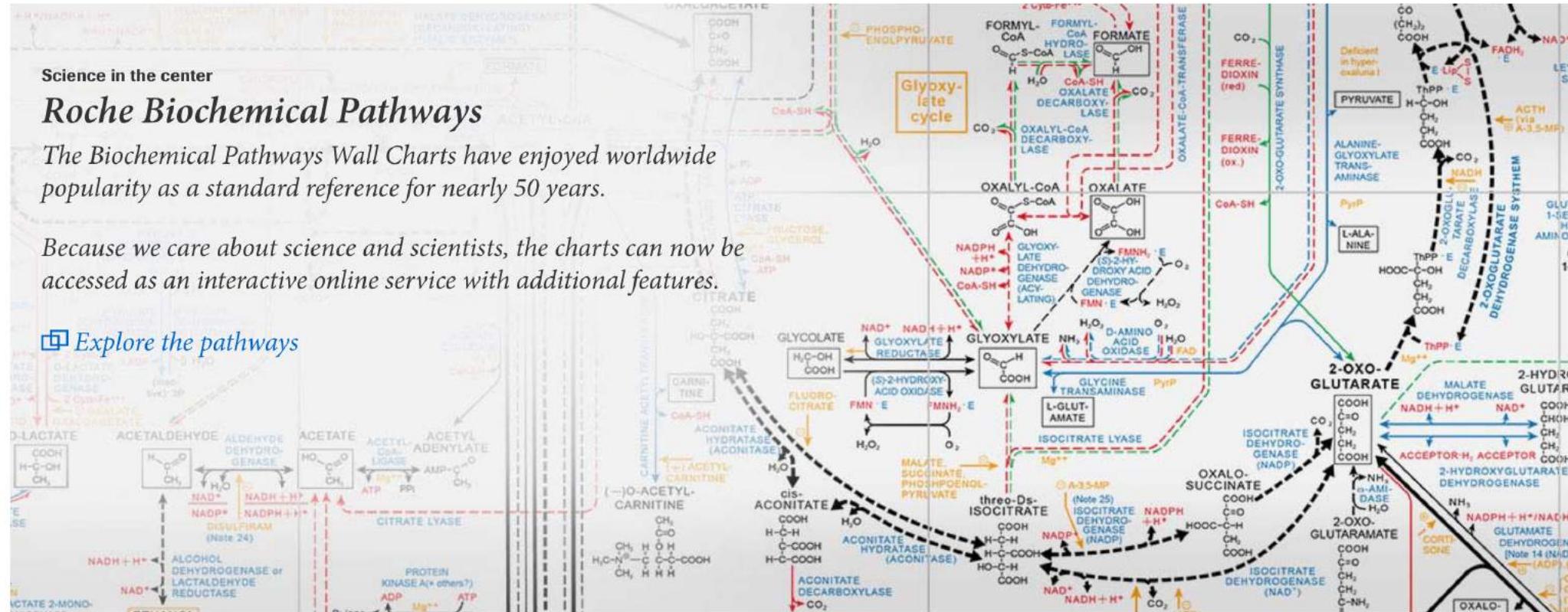
Science in the center

Roche Biochemical Pathways

The Biochemical Pathways Wall Charts have enjoyed worldwide popularity as a standard reference for nearly 50 years.

Because we care about science and scientists, the charts can now be accessed as an interactive online service with additional features.

Explore the pathways



Share Rate

Mapping the paths of life

Biochemical Pathways provide an overview of the chemical reactions of cells in various species and organs. Dr. Michal first compiled the Pathways Chart in 1965 and has been fine-tuning it ever since. Today, and with the collaboration of Roche, the two enormous posters can be found hanging in just about every research institute from Argentina to New Zealand.

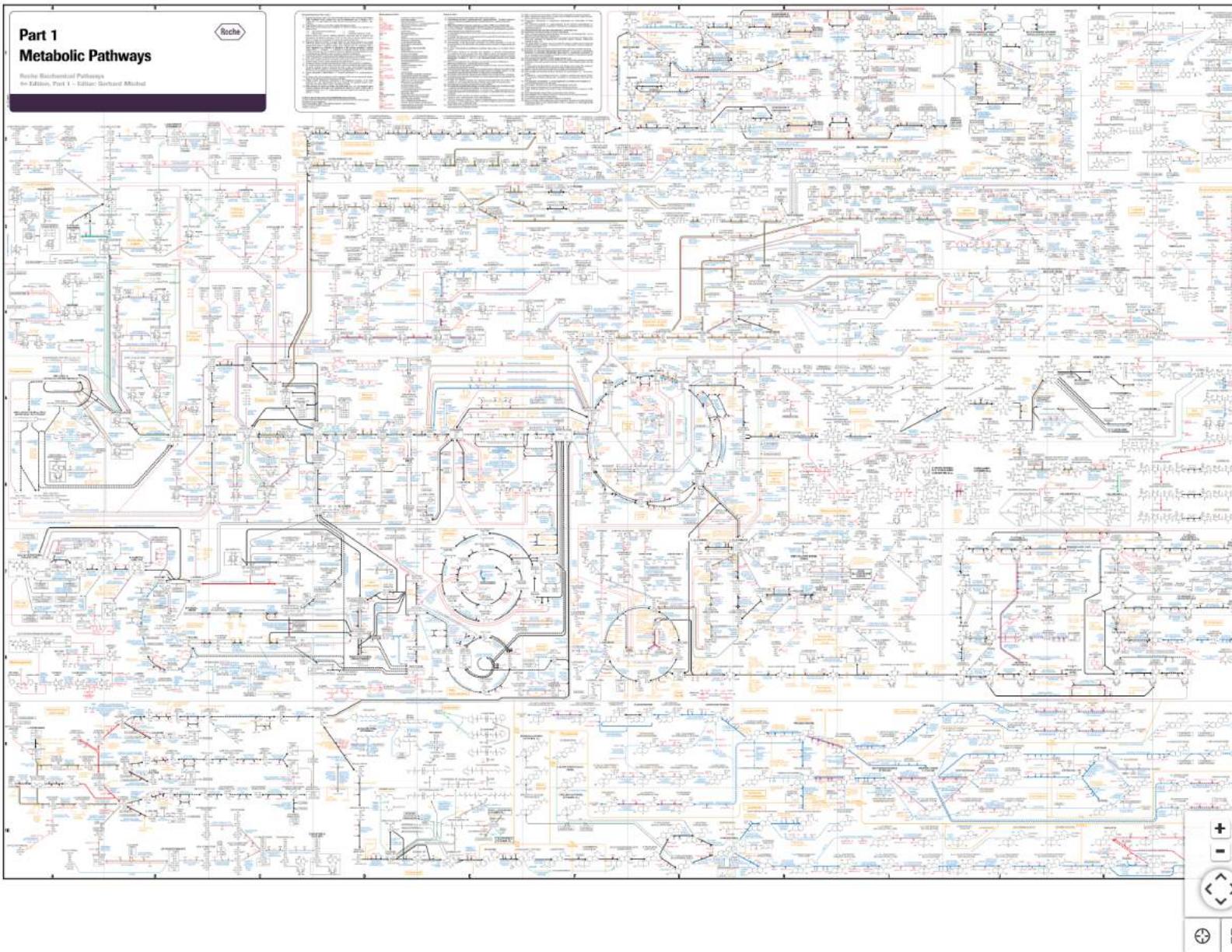
“ You have to be someone with tenacity and patience. And love for science

Dr Gerhard Michal
Editor of the Roche Biochemical Pathways

By the numbers

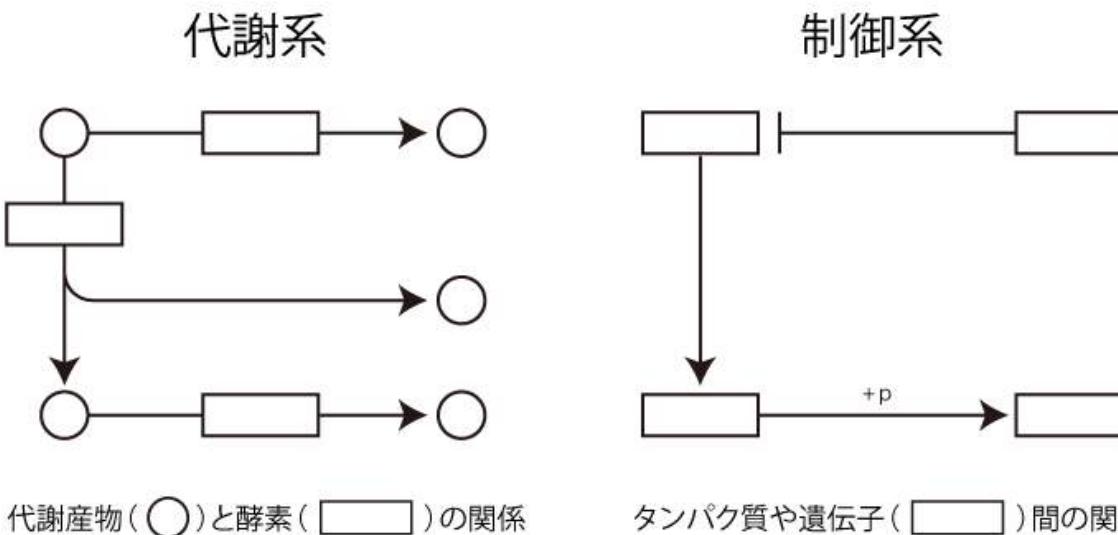
49 years
continuously fine-tuned by the editor himself.

Click here to open the menu



パスウェイデータベースでの表現

- 見やすくするために、ダイアグラムで表現されていることが多いっています。
- 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



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いっています。
 - [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, Reactome, KEGG PATHWAY のブラウザ上での使い方を紹介します。

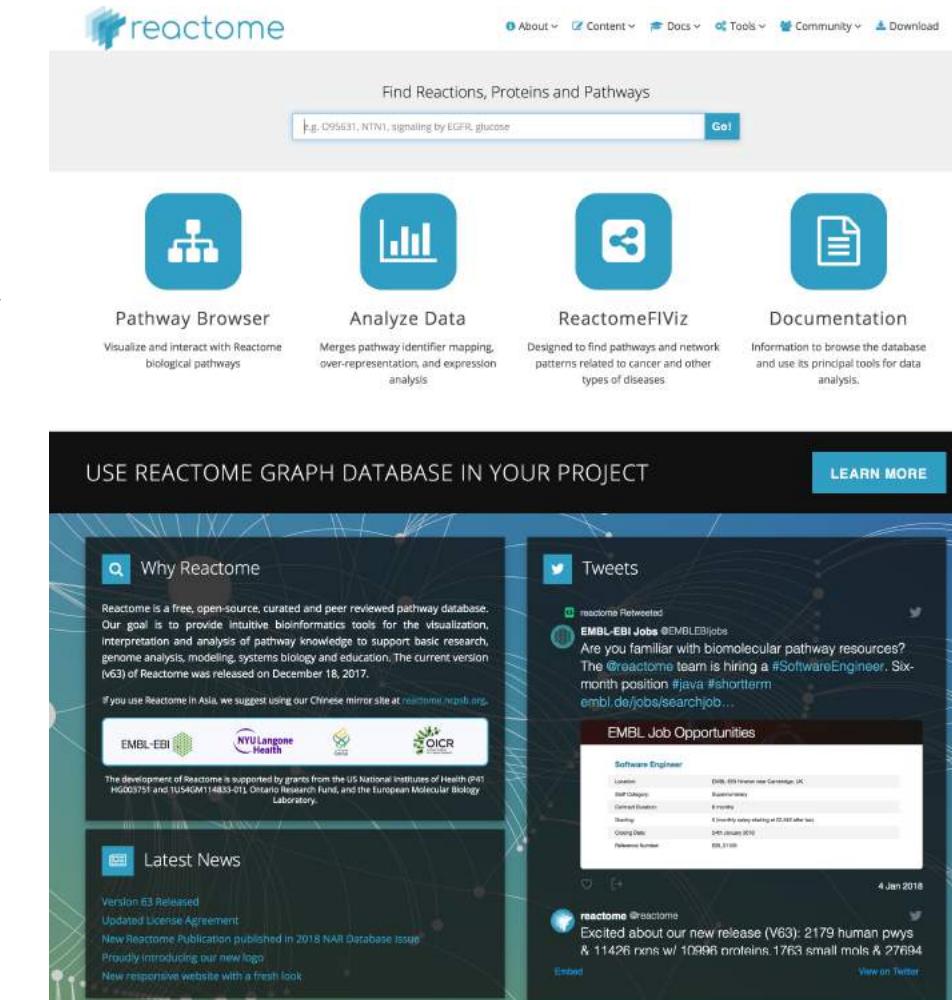
BioCyc

- ウェブサイト：<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
- 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 LOGIN, Why Log in?, Create New Account, Sites, Search, Genome, Metabolism, Analysis, SmartTables, and Help. Below the navigation is a search bar with the placeholder "Search a gene, protein, metabolite or pathway..." and buttons for Quick Search and Gene Search. A message above the search bar says "Searching Escherichia coli K-12 substr. MG1655 (EcoCyc) | change organism database". The main content area features a large diagram titled "RouteSearch" illustrating metabolic pathways. It shows three routes from a glucose derivative to a final product: a "Long Route" (orange circles), a "Loss of Atoms" route (yellow dashed line), and the "Best Route: Atoms Conserved" (blue line). Below the diagram, there's a section titled "BioCyc Database Collection" with a brief description of subscriptions and a "Getting Started" section with usage instructions. To the right, there are sections for "Tools" (with a list of features like Genome browser, Display of individual metabolic pathways, etc.), "BioCyc Databases" (mentioning three tiers: Tier 1, Tier 2, Tier 3), and "Funding Sources" (noting NIH funding). At the bottom, there's a page navigation bar with numbers 1 through 10.

Reactome

- ウェブサイト：
<http://www.reactome.org/>
- 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（[参照](#)）
 - 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



KEGG PATHWAY

- ウェブサイト：<http://www.kegg.jp/>
- 開発：京都大学
- 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] (<http://apps.cytoscape.org/apps/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。

The screenshot shows the KEGG homepage with a colorful logo at the top left. The main menu includes "KEGG Home", "KEGG Database", "KEGG Objects", and "KEGG Software". A red box highlights the "KEGG2" link under the "Data-oriented entry points" section. The central content area features a brief introduction to KEGG and an announcement about the release of KEGG NETWORK.

The screenshot shows a search results page for "KEGG" with a yellow header bar containing links for PATHWAY, BRITE, MODULE, KO, GENES, LIGAND, NETWORK, DISEASE, DRUG, and DBGET. Below the header is a search bar. The main content is a table titled "Data-oriented entry points" with four columns: Category, Entry Point, Content, and DBGET Search. The table lists various KEGG databases and their corresponding DBGET search terms.

Category	Entry Point	Content	DBGET Search
Systems information	KEGG PATHWAY KEGG BRITE KEGG MODULE	KEGG pathway maps BRITE hierarchies and tables KEGG modules	PATHWAY BRITE MODULE
Genomic information	KO (KEGG Orthology) KEGG GENOME KEGG GENES KEGG SSDB	Functional orthologs KEGG organisms (complete genomes) Genes and proteins GENES sequence similarity	ORTHOLOGY GENOME GENES
Chemical information (KEGG LIGAND)	KEGG COMPOUND KEGG GLYCAN KEGG REACTION KEGG ENZYME	Small molecules Glycans Reactions and reaction classes Enzyme nomenclature	COMPOUND GLYCAN REACTION RCLASS ENZYME
Health information	KEGG NETWORK <i>New!</i> KEGG DISEASE KEGG DRUG KEGG ENVIRON KEGG MEDICUS	Disease-related network elements Human diseases Drugs and drug groups Health related substances Japanese drug labels (JAPIC) FDA drug labels (DailyMed)	NETWORK VARIANT DISEASE DRUG DGROUP ENVIRON

対象生物種を見る

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

Subject-oriented entry points

KEGG Cancer	Cancer research
KEGG Pathogen	Pathogen genomes and antimicrobial resistance
KEGG Virus	Virus research
KEGG Plant	Plant research
KEGG Annotation	KO annotation of genes and proteins
KEGG SeqData	KEGG sequence data collection
KEGG RModule	Architecture of metabolic networks

Organism-specific entry points

[KEGG organisms](#) the list of currently available organisms

Select (examples) hsa mmu sce eco bsu syn

Organism group	Pangenome		Genome	
Eukaryotes	Animals Plants Fungi Protists	Genus level pangenomes	Species level pangenomes	KEGG organisms
Prokaryotes	Bacteria Archaea			

See [KEGG GENOME](#) for more details.

KEGG Organisms

- KEGG では 3-4 文字の独自の生物種コードを使用している

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
	csab Chlorocebus sabaeus (green monkey)	RefSeq
	rro Rhinopithecus roxellana (golden snub-nosed monkey)	RefSeq
	rbb Rhinopithecus bieti (black snub-nosed monkey)	RefSeq
	cjc Callithrix jacchus (white-tufted-ear marmoset)	RefSeq
	sbq Saimiri boliviensis boliviensis (Bolivian squirrel monkey)	RefSeq
	mmu Mus musculus (mouse)	RefSeq
	rno Rattus norvegicus (rat)	RefSeq
	cge Cricetus griseus (Chinese hamster)	RefSeq
	ngi Nannospalax galili (Upper Galilee mountains blind mole rat)	RefSeq
	hgl Heterocephalus glaber (naked mole rat)	RefSeq
	ccan Castor canadensis (American beaver)	RefSeq
	ocu Oryctolagus cuniculus (rabbit)	RefSeq
	tup Tupai chinensis (Chinese tree shrew)	RefSeq

KEGG Organisms

- 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: <input type="text"/>	<input type="button" value="Go"/>	<input type="button" value="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_000001405.37) BioProject: 168					
Original DB	NCBI, OMIM, HGNC, HPRD, Ensembl, Vega, Pharos					
Statistics	Number of protein genes: 20364 Number of RNA genes: 19178					
Created	2000					

KEGG Pan paniscus (bonobo)

Genome info	Pathway map	Brite hierarchy	Module	Genome map	Blast	Taxonomy
Search genes: <input type="text"/>	<input type="button" value="Go"/>	<input type="button" value="Clear"/>				
Genome information						
T number	T02283					
Org code	pps					
Aliases	9597					
Full name	Pan paniscus (bonobo)					
Definition	Pan paniscus (bonobo)					
Annotation	KOALA					
Taxonomy	TAX: 9597					
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Pan					
Data source	RefSeq (Assembly: GCF_000258655.2) BioProject: 169343					
Original DB	Max-Planck					
Statistics	Number of protein genes: 20581 Number of RNA genes: 4217					
Created	2012					

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



KEGG Organisms: Complete Genomes

Eukaryotes: 429 Bacteria: 4521 Archaea: 263

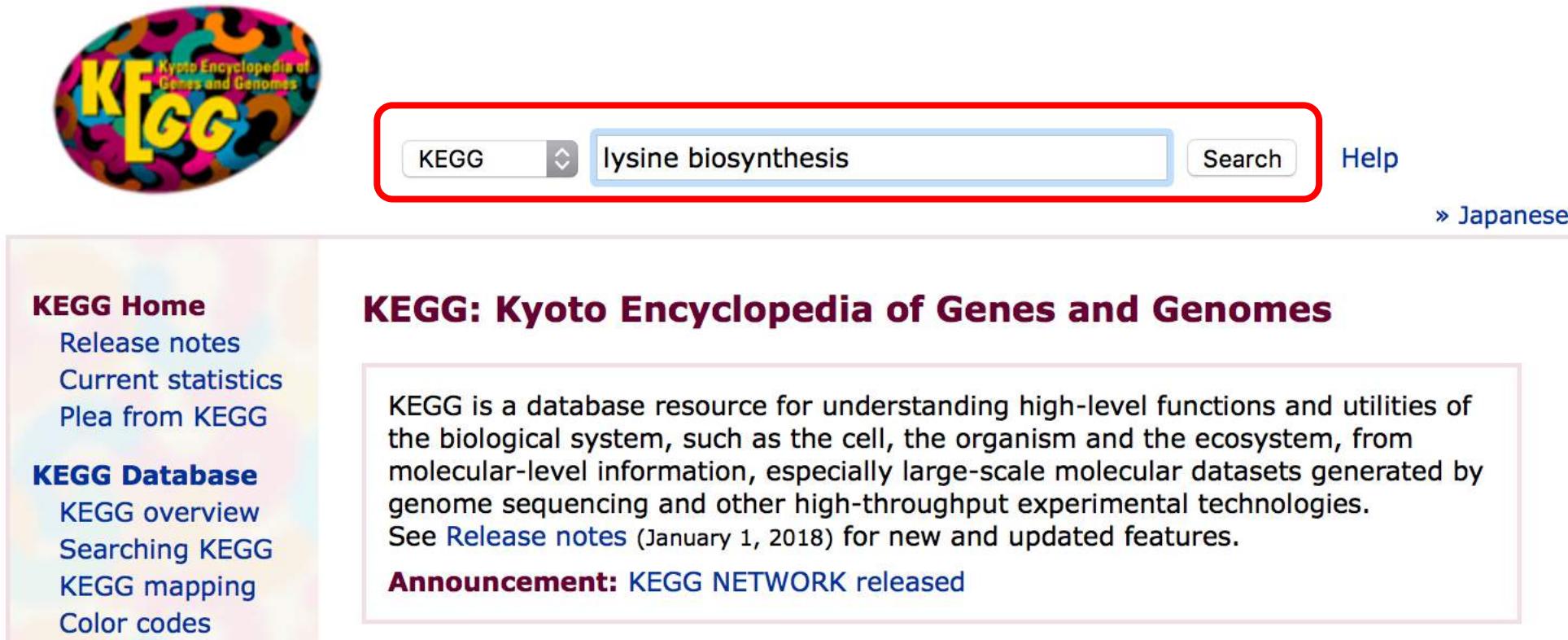
[Genomes | Species | Genus | Viruses | Meta]

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 leucomoanus (northern white-cheeked gibbon)	RefSeq

KEGGパスウェイマップを見る

- 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 left is the KEGG logo. To its right is a search bar with the text "lysine biosynthesis" inside, which is highlighted with a red rectangle. To the right of the search bar are "Search" and "Help" buttons, and a link to "» Japanese". On the far left, there is a sidebar with links to "KEGG Home", "Release notes", "Current statistics", "Plea from KEGG", "KEGG Database", "KEGG overview", "Searching KEGG", "KEGG mapping", and "Color codes". The main content area features the heading "KEGG: Kyoto Encyclopedia of Genes and Genomes" in large red text. Below it is a paragraph describing KEGG as a database resource for understanding biological systems. It also mentions the "KEGG NETWORK released" announcement.

lysine biosynthesis

Search

Help

» Japanese

KEGG Home

Release notes

Current statistics

Plea from KEGG

KEGG Database

KEGG overview

Searching KEGG

KEGG mapping

Color codes

KEGG: Kyoto Encyclopedia of Genes and Genomes

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 experimental technologies. See [Release notes](#) (January 1, 2018) for new and updated features.

Announcement: KEGG NETWORK released

KEGGパスウェイマップ

- 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。

 Search KEGG for lysine biosynthesis Go Clear

Database: KEGG - Search term: lysine biosynthesis

KEGG PATHWAY

map00300
Lysine biosynthesis

map01064
Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid

KEGG MODULE

M00016
Lysine biosynthesis, succinyl-DAP pathway, aspartate => lysine

M00030
Lysine biosynthesis, AAA pathway, 2-oxoglutarate => 2-aminoadipate => lysine

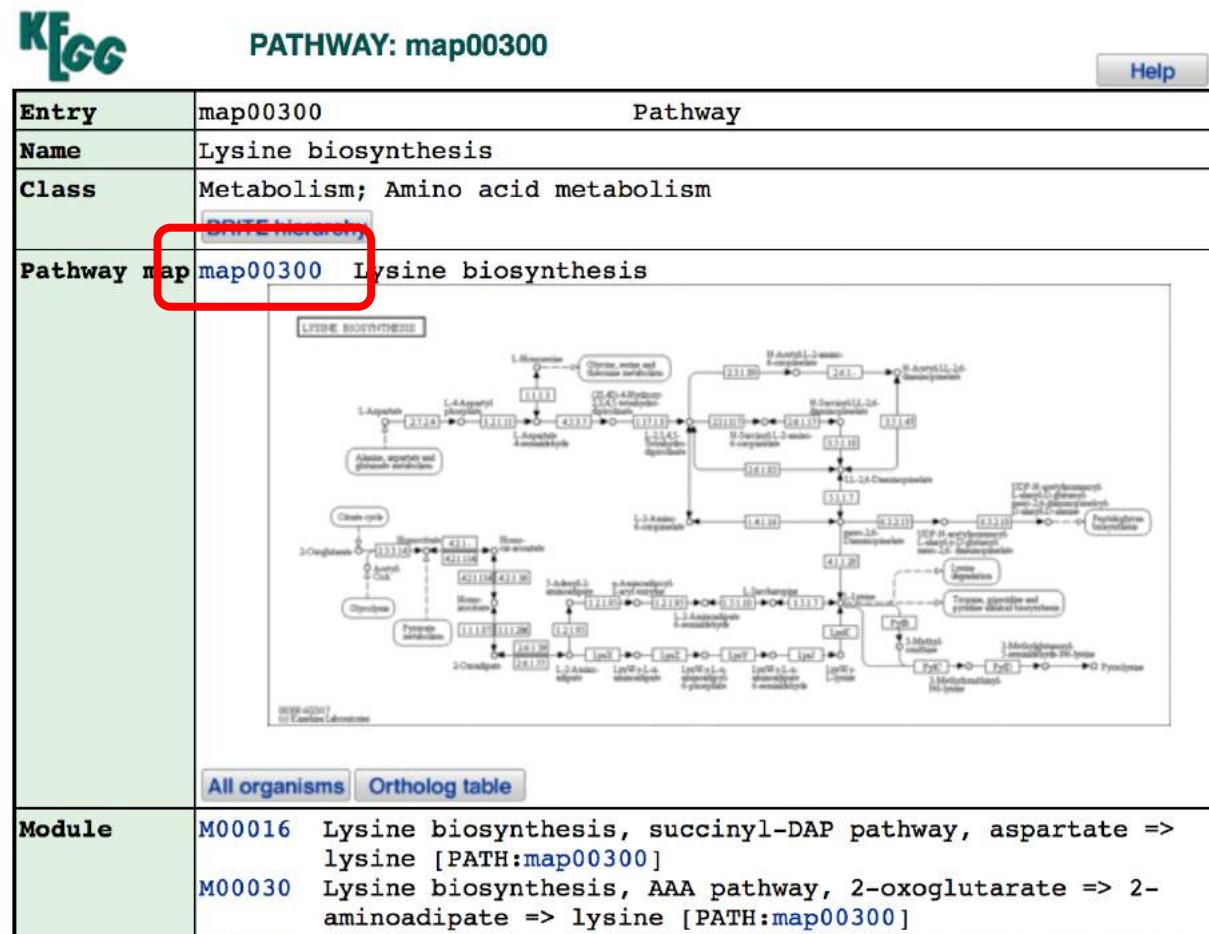
M00031
Lysine biosynthesis, mediated by LysW, 2-aminoadipate => lysine

M00433
Lysine biosynthesis, 2-oxoglutarate => 2-oxoadipate

M00525
Lysine biosynthesis, acetyl-DAP pathway, aspartate => lysine

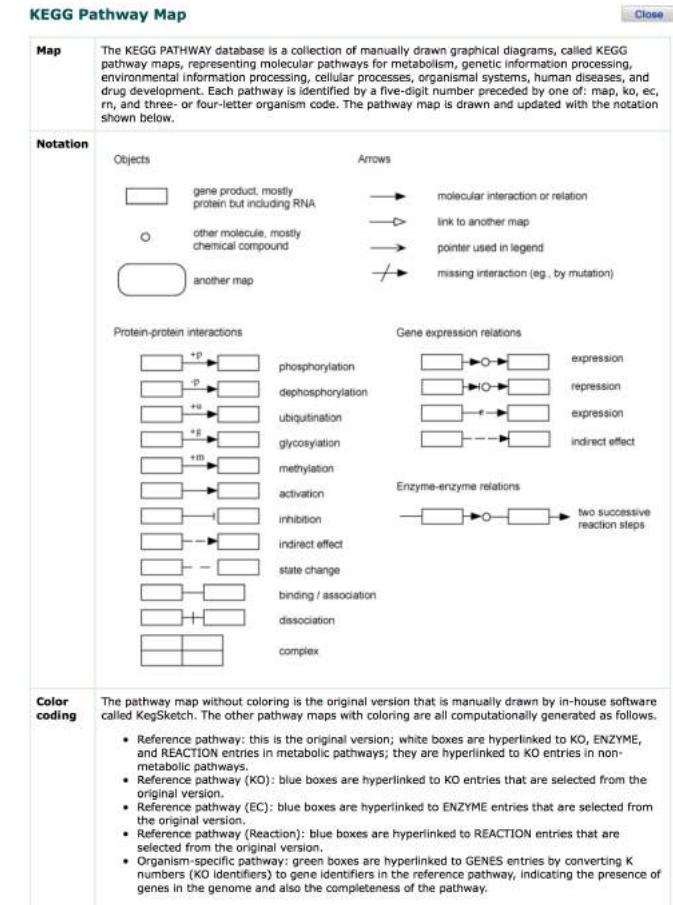
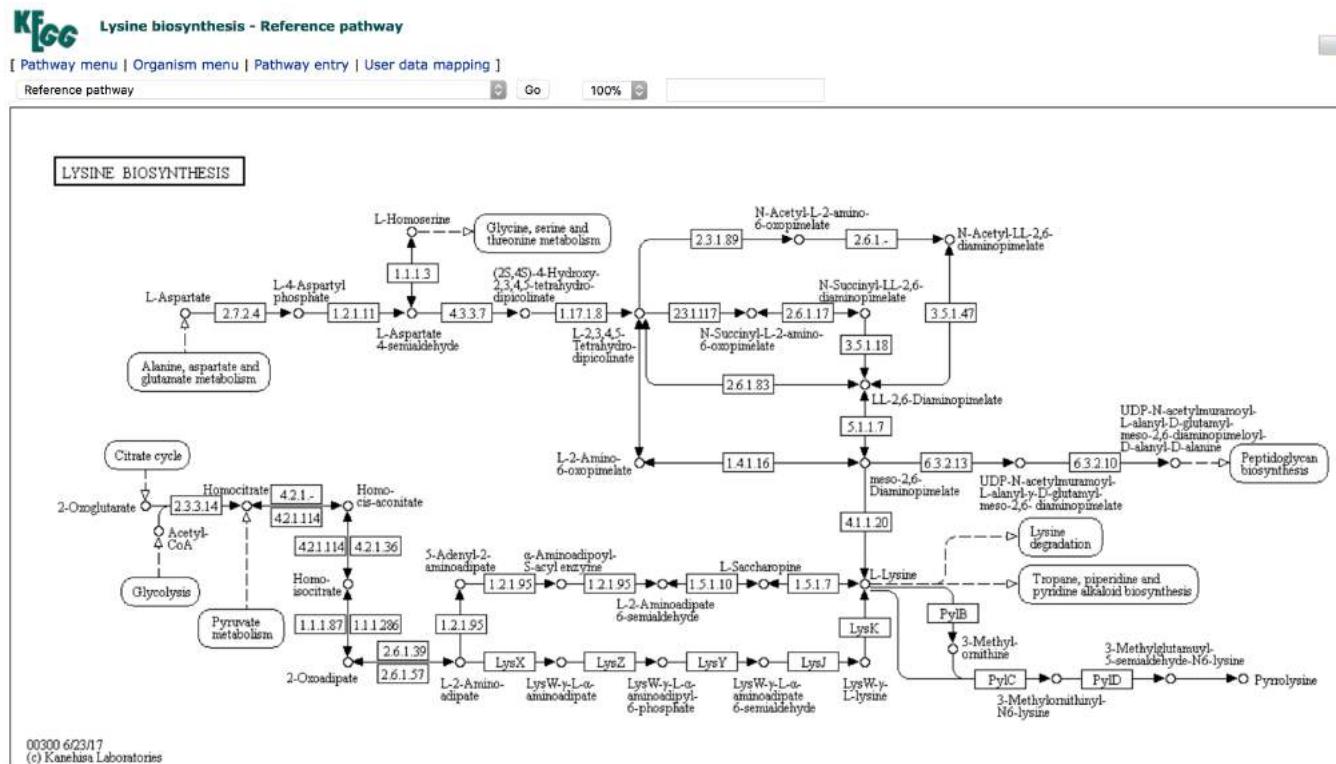
KEGGパスウェイマップ

- 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、ダイアグラム画像を“マップ”と呼んでいます。



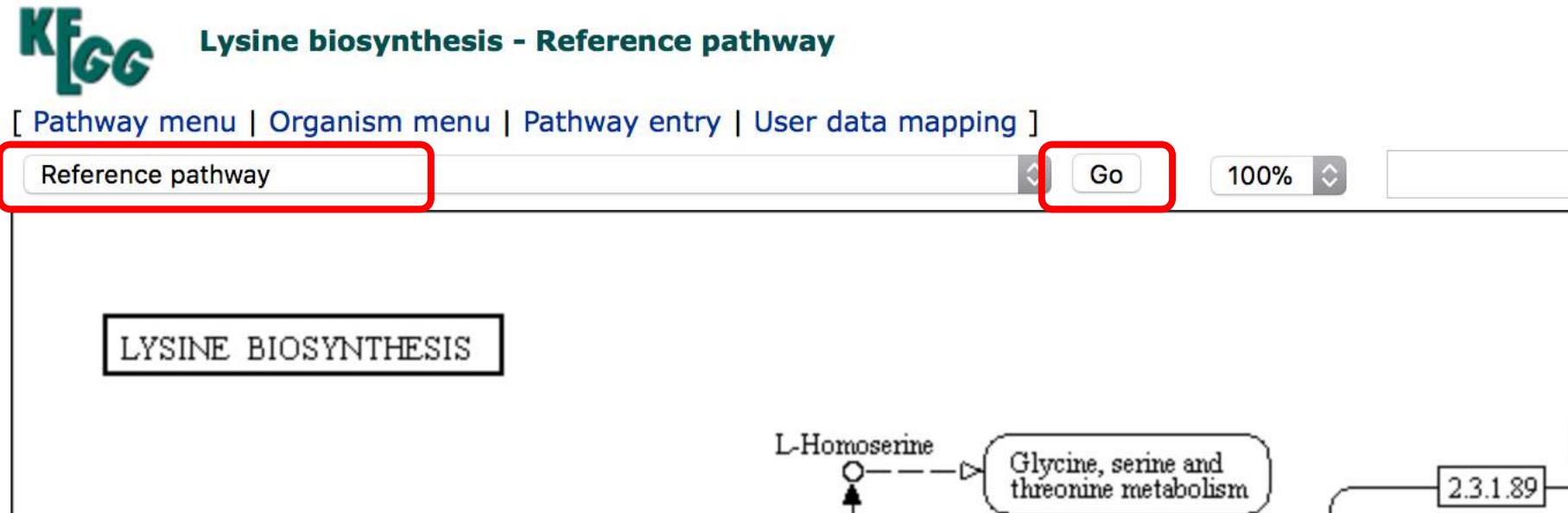
KEGGパスウェイマップ

- この色のついていない白いダイアグラムが、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
 - ボックスが遺伝子やタンパク質などの配列情報、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



好きな生物のパスウェイを見る

- 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



別のアプローチ

- 種、属でまとめたパスウェイ、ドラフトゲノム、メタゲノムの
パスウェイはここからは選べないので、生物種リストのページ
から、種のページ、パスウェイリストへ移動する必要があります。

Kfcc Homo sapiens (human)

Genome info Pathway map Brite hierarchy Module Genome map Blast Taxonomy

Search genes: Go Clear

Prokaryotes

Category	Organism
	eco Escherichia coli K-12 MG1655
	ecj Escherichia coli K-12 W3110
	ecd Escherichia coli K-12 DH10B
	ebw Escherichia coli BW2952
	ecok Escherichia coli K-12 MDS42
	ece Escherichia coli O157:H7 EDL933 (EHEC)
	ecs Escherichia coli O157:H7 Sakai (EHEC)
	ecf Escherichia coli O157:H7 EC4115 (EHEC)
	etw Escherichia coli O157:H7 TW14359 (EHEC)
	elx Escherichia coli O157:H7 Xuzhou21 (EHEC)
	eoj Escherichia coli O26:H11 11368 (EHEC)
	eoi Escherichia coli O111:H- 11128 (EHEC)

Pathway map (Red box)

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
	2009 GenBank
	2012 GenBank
	2009 GenBank
	2009 GenBank

Lysine biosynthesis マップ

- 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。

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

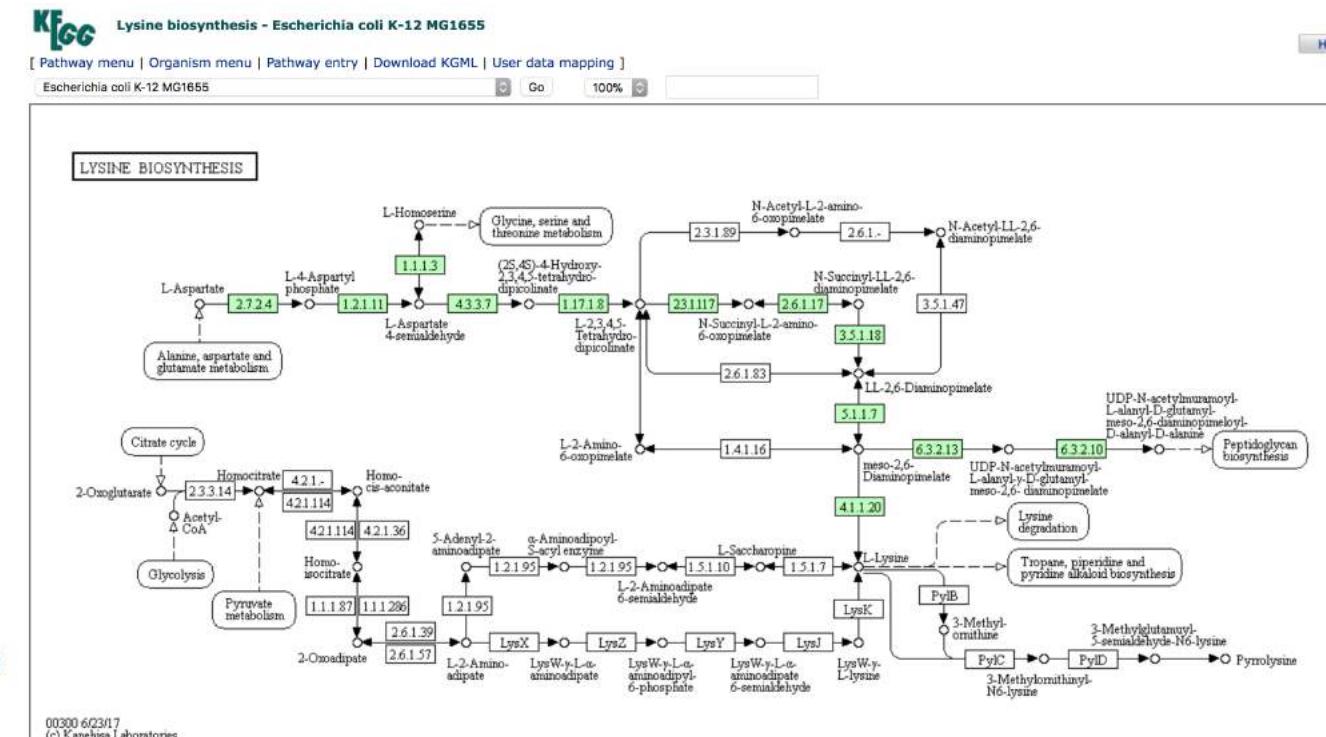
00350 Tyrosine metabolism

00360 Phenylalanine metabolism

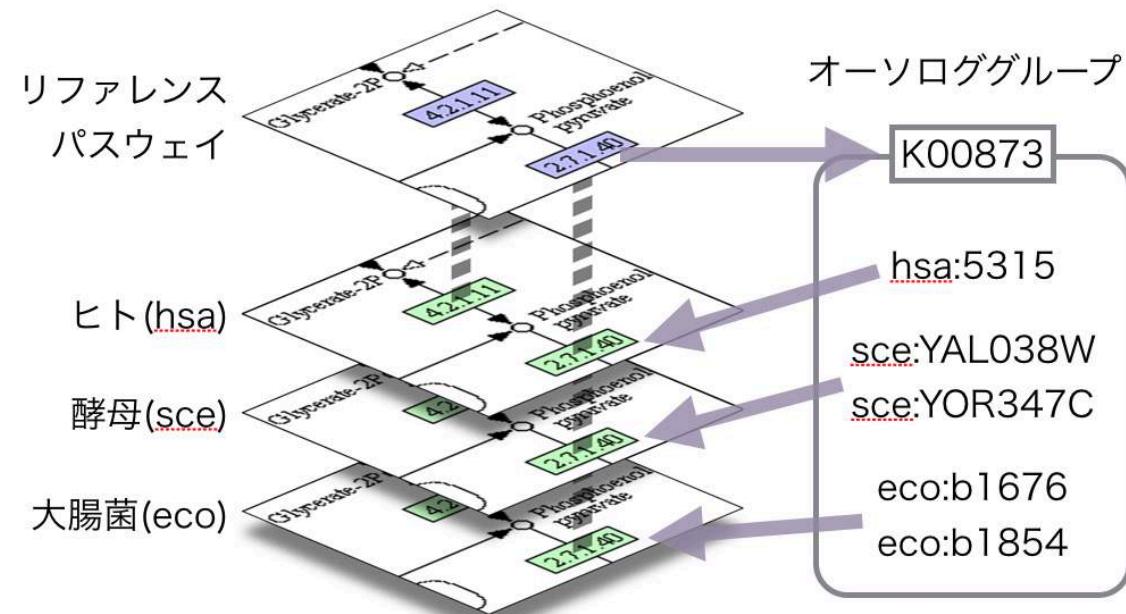
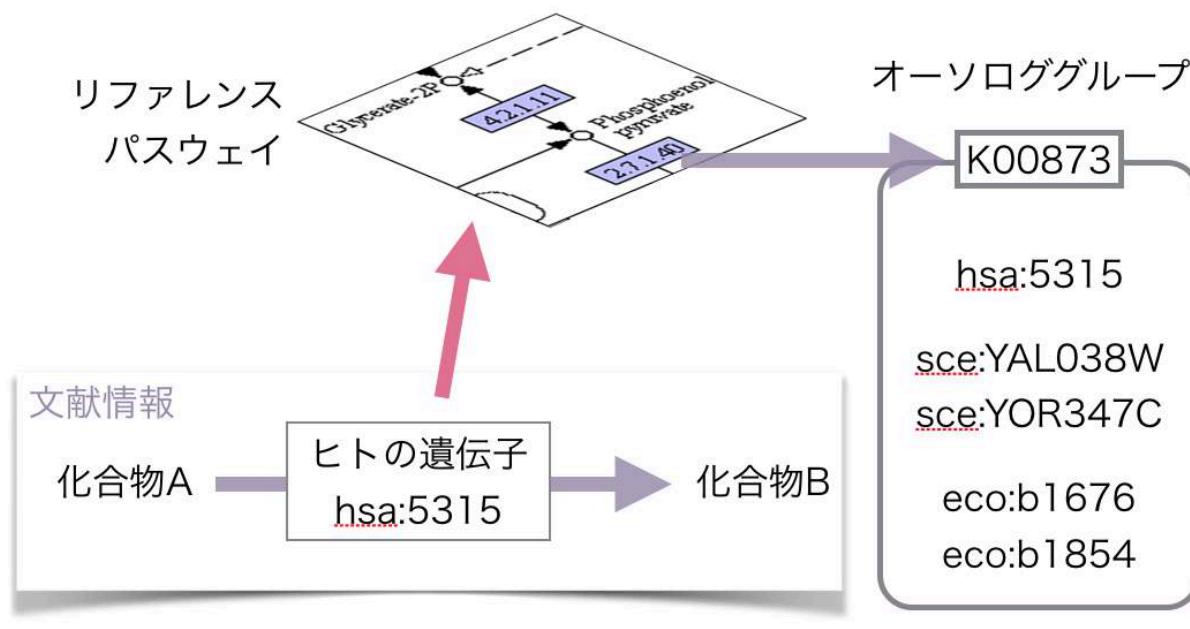
00380 Tryptophan metabolism

00400 Phenylalanine, tyrosine and tryptophan biosynthesis

Metabolism of other amino acids



リファレンスパスウェイと種毎のパスウェイの関係



KOと化合物のネットワークであるリファレンスパスウェイを基に、生物種毎のパスウェイが作られる

Overviewマップを見る

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

1. Metabolism

1.0 Global and overview maps

01 00	Metabolic pathways	[KEGG Atlas]
01110	Biosynthesis of secondary metabolites	[KEGG Atlas]
01120	Microbial metabolism in diverse environments	[KEGG Atlas]
01130	Biosynthesis of antibiotics	[KEGG Atlas]
01200	Carbon metabolism	[KEGG Atlas]
01210	2-Oxocarboxylic acid metabolism	[KEGG Atlas]
01212	Fatty acid metabolism	[KEGG Atlas]
01230	Biosynthesis of amino acids	[KEGG Atlas]
01220	Degradation of aromatic compounds	[KEGG Atlas]

1.1 Carbohydrate metabolism

00010	Glycolysis / Gluconeogenesis
00020	Citrate cycle (TCA cycle)

[Pathway menu | Organism menu | Pathway entry | Hide module list | User data mapping | Image (png) file]

Reference pathway

Go

35%

KEGG module

Energy metabolism

Carbon fixation

- M00165 Reductive pentose phosphate cycle
- M00166 Reductive pentose phosphate cycle
- M00167 Reductive pentose phosphate cycle
- M00168 CAM (Crassulacean acid metabolism)
- M00169 CAM (Crassulacean acid metabolism)
- M00172 C4-dicarboxylic acid metabolism
- M00171 C4-dicarboxylic acid metabolism
- M00170 C4-dicarboxylic acid metabolism
- M00173 Reductive citrate cycle
- M00579 Phosphate acetyl transferase

Nitrogen metabolism

- M00175 Nitrogen fixation

Methane metabolism

- M00567 Methanogenesis, archaeal
- M00174 Methane oxidation

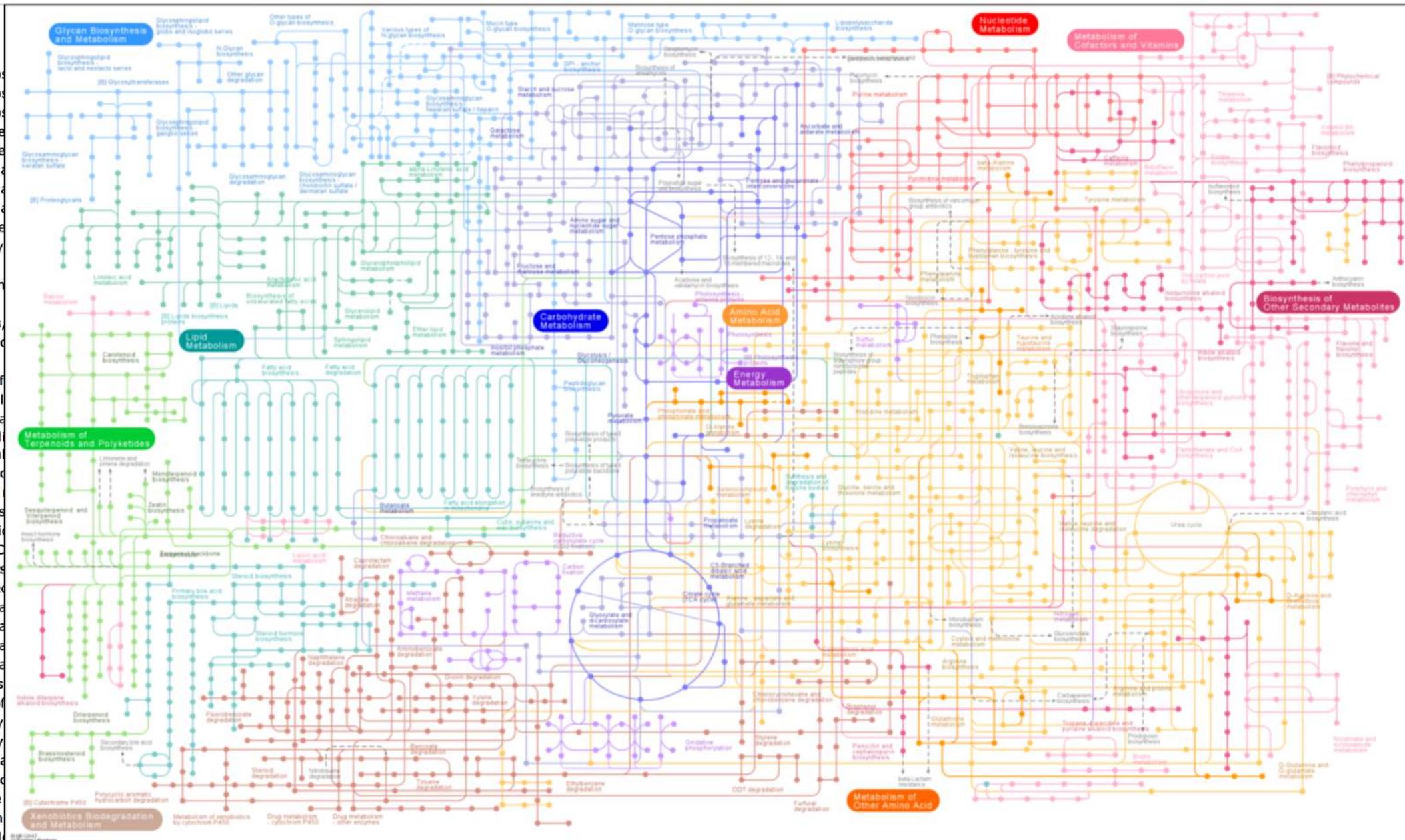
Sulfur metabolism

- M00176 Assimilatory sulfate reduction
- M00596 Dissimilatory sulfate reduction
- M00595 Thiosulfate oxidation

Carbohydrate and lipid metabolism

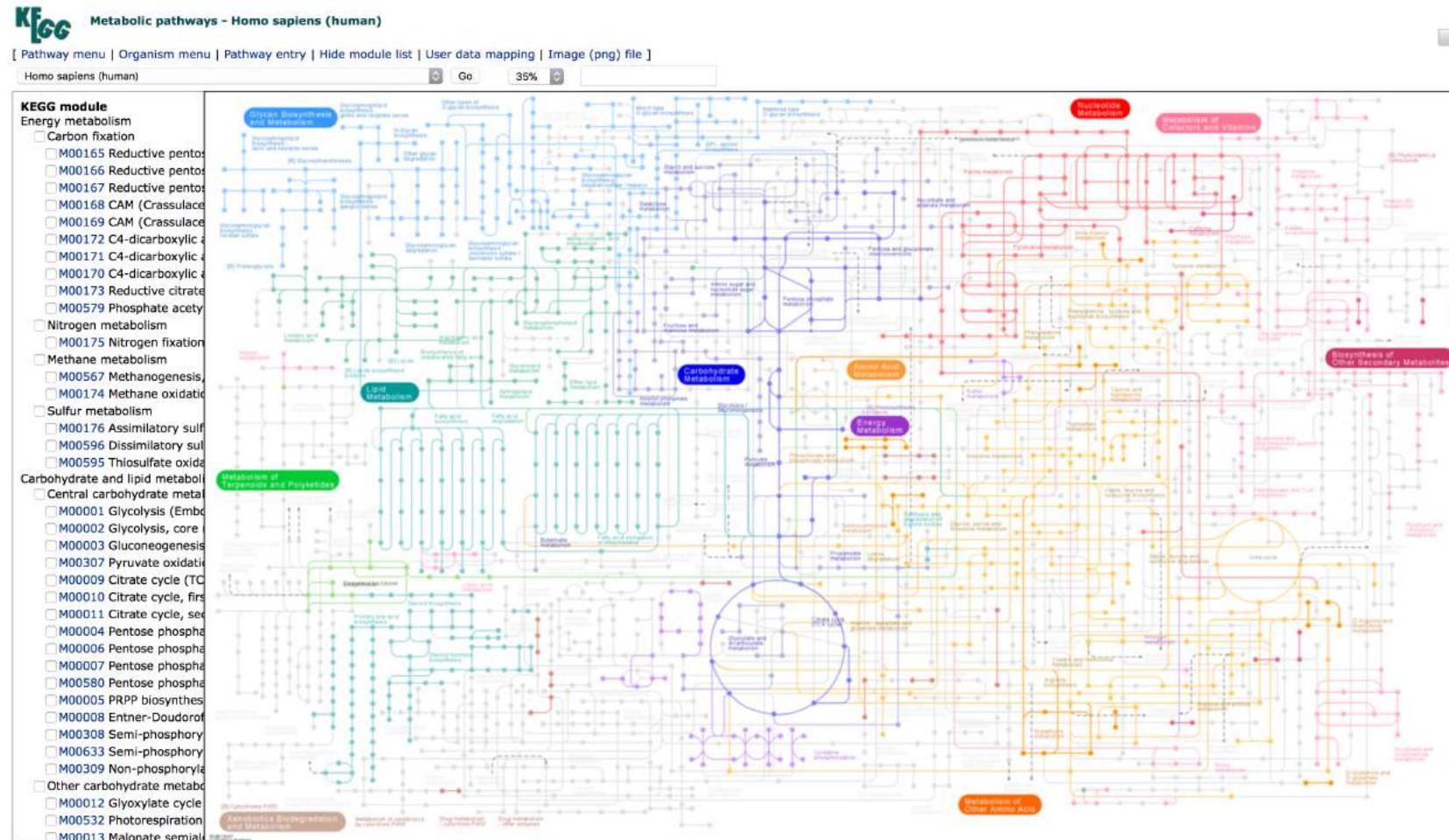
Central carbohydrate metabolism

- M00001 Glycolysis (Embden-Meyerhof-Parnet pathway)
- M00002 Glycolysis, core
- M00003 Gluconeogenesis
- M00307 Pyruvate oxidation
- M00009 Citrate cycle (TCA cycle)
- M00010 Citrate cycle, first step
- M00011 Citrate cycle, second step
- M00004 Pentose phosphate cycle
- M00006 Pentose phosphate cycle
- M00007 Pentose phosphate cycle
- M00580 Pentose phosphate cycle
- M00005 PRPP biosynthesis
- M00008 Entner-Doudoroff pathway
- M00308 Semi-phosphorylated glycolysis
- M00633 Semi-phosphorylated glycolysis
- M00309 Non-phosphorylated glycolysis
- M00012 Glyoxylate cycle
- M00532 Photorespiration
- M00013 Malonate semialdehyde cycle



生物種毎のOverviewマップを見る

- 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



ヒトの疾患パスウェイを見る

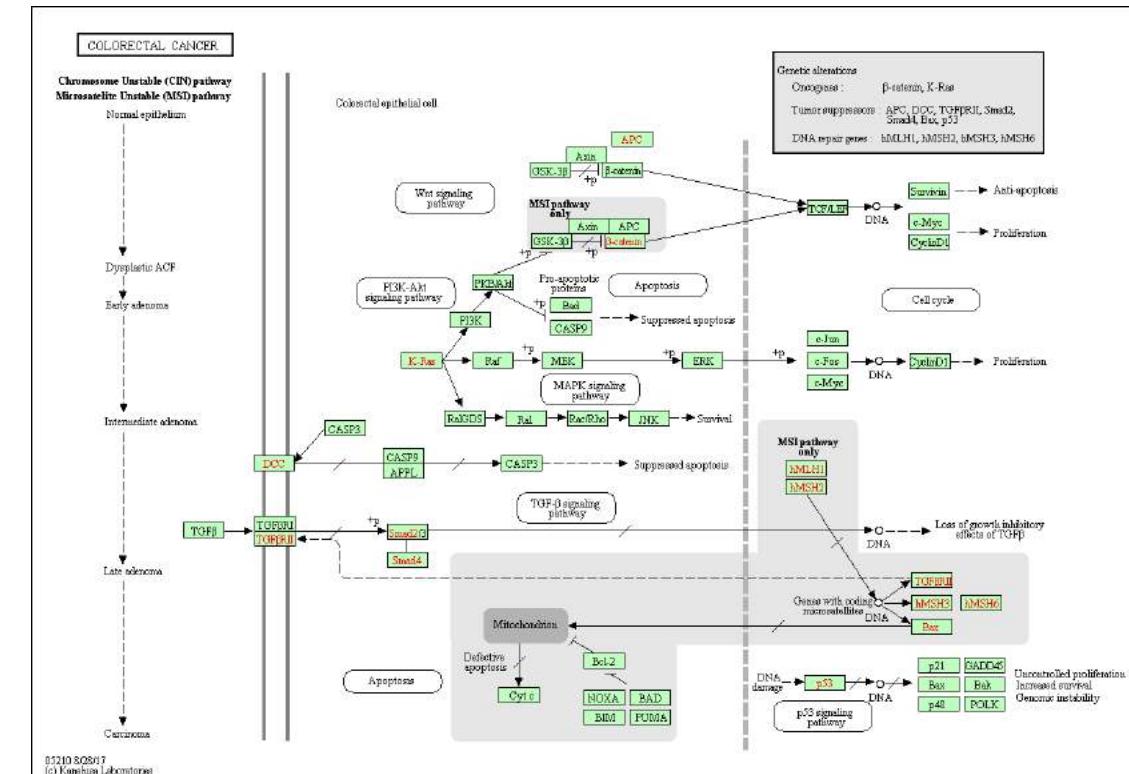
- <http://www.genome.jp/kegg/pathway.html#disease>
- がん、免疫系疾患、神経変性疾患など多因子性の疾患
- 好きな疾患パスウェイをクリック（例：[大腸がん](#)）
 - 赤字の遺伝子が疾患の病因遺伝子を示しています。

6.1 Cancers: Overview

- 05200 Pathways in cancer
- 05230 Central carbon metabolism in cancer
- 05231 Choline metabolism in cancer
- 05202 Transcriptional misregulation in cancer
- 05206 MicroRNAs in cancer
- 05205 Proteoglycans in cancer
- 05204 Chemical carcinogenesis
- 05203 Viral carcinogenesis

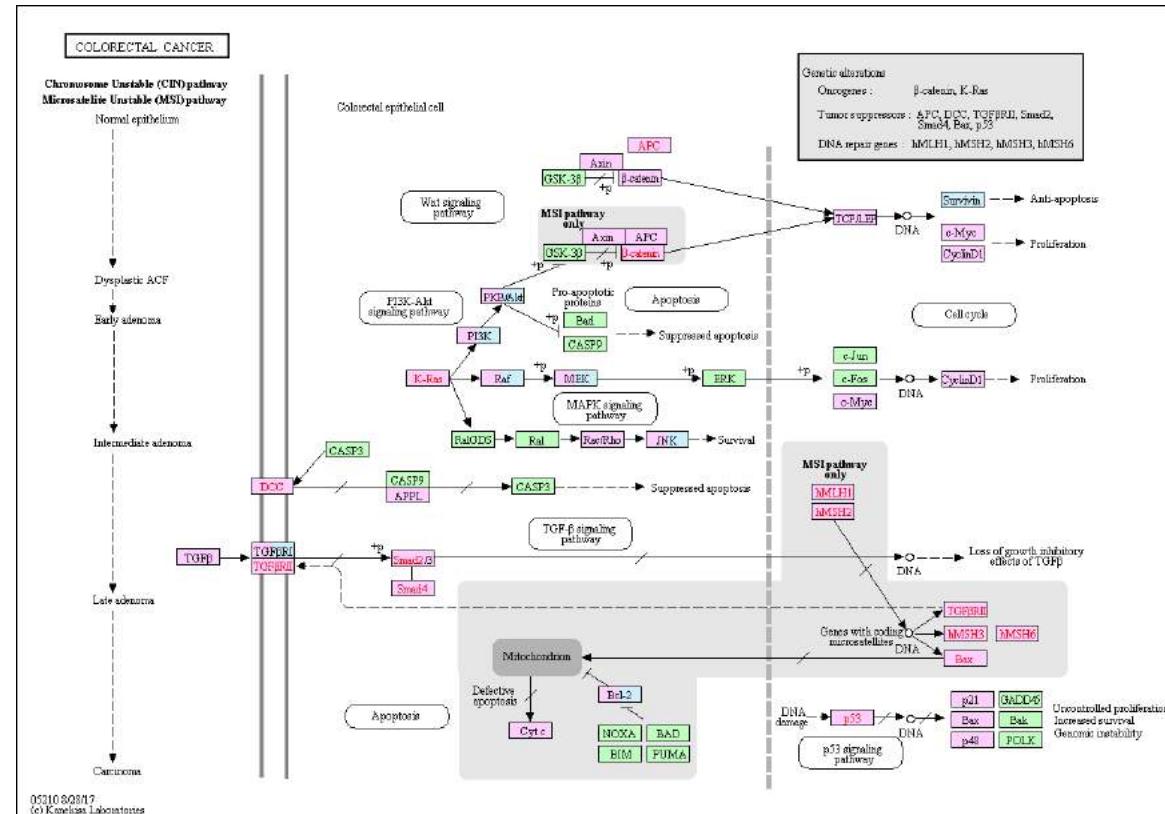
6.2 Cancers: Specific types

- 05210 Colorectal cancer
- 05212 Pancreatic cancer
- 05225 Hepatocellular carcinoma *New!*
- 05226 Gastric cancer *New!*



ヒトの疾患パスウェイ

- 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
- このように、KEGG では正常な状態のパスウェイの他に、病原因子や医薬品、そのターゲットなどの情報もパスウェイとして表現されています



種間比較をする

- <http://www.genome.jp/kegg/kegg2.html>
- KEGG GENOMEのリンクをクリック
- KEGG Mapping for Genome Comparison and Combination のテキストボックスに "eco ecs" と入力して Go をクリック
 - eco : 非病原性大腸菌 k-12
 - ecs : 病原性大腸菌 O157

Data-oriented entry points

Category	Entry Point	Content	DBGET Search
Systems information	KEGG PATHWAY	KEGG pathway maps	PATHWAY
	KEGG BRITE	BRITE hierarchies and tables	BRITE
	KEGG MODULE	KEGG modules	MODULE
Genomic information	KO (KEGG Orthology)	Functional orthologs	ORTHOLOGY
	KEGG GENOME	KEGG organisms (complete genomes)	GENOME
	KEGG GENES	Genes and proteins	GENES
Chemical information (KEGG LIGAND)	KEGG SSDB	GENES sequence similarity	
	KEGG COMPOUND	Small molecules	COMPOUND
	KEGG GLYCAN	Glycans	GLYCAN
	KEGG REACTION	Reactions and reaction classes	REACTION
Health information	KEGG ENZYME	Enzyme nomenclature	RCLASS ENZYME
	KEGG NETWORK <small>New!</small>	Disease-related network elements	NETWORK
	KEGG DISEASE	Human diseases	VARIANT
	KEGG DRUG	Drugs and drug groups	DISEASE
	KEGG ENVIRON	Health related substances	DRUG DGROUP
	KEGG MEDICUS	Japanese drug labels (JAPIC)	ENVIRON
		FDA drug labels (DailyMed)	

KEGG Mapping for Genome Comparison and Combination

An organism group may be defined to compare or combine KEGG organisms, other organisms and environmental samples, enabling the analysis of combined pathway maps for the group.

Define organism group (enter organism codes or T numbers):

eco ecs

Go

(Genome comparison)

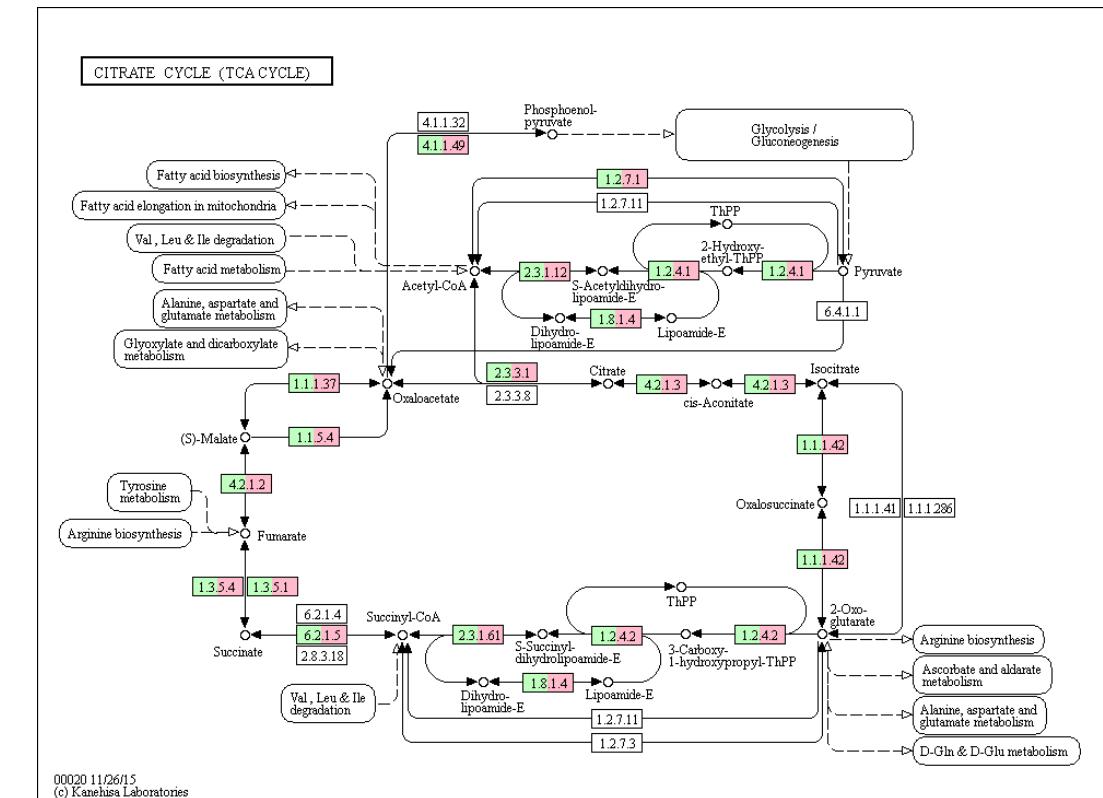
hsa eco Homo sapiens (human) vs. Escherichia coli K-12 strain
such as [Terpenoid backbone biosynthesis](#)
where green for human, red for E.coli, and split for both

種間比較

- 上のメニューの Pathway map をクリック
- 好きなパスウェイをクリック
 - 前者が持っている遺伝子が緑、後者が持っている遺伝子がピンクで表示
 - Overview パスウェイの場合、両者が持っている遺伝子はライトブルーで表示

KEGG Organism group: eco ecs

Category info	Pathway map	Brite hierarchy	Taxonomy				
Search genes:	<input type="text"/> Go Clear						
eco ecs							
T00007 eco Escherichia coli K-12 MG1655							
T00048 ecs Escherichia coli O157:H7 Sakai (EHEC)							
KEGG2	PATHWAY	BRITE	GENES	GENOME	LIGAND	DISEASE	DRUG



Bacterial secretion system マップ

- 03070 : Bacterial secretion system パスウェイでは病原性に関わる III型、VI型分泌装置が O157 側だけが持っているのがわかる

Environmental Information Processing

Membrane transport

02010 ABC transporters

02060 Phosphotransferase system (PTS)

03070 Bacterial secretion system

Signal transduction

02020 Two-component system

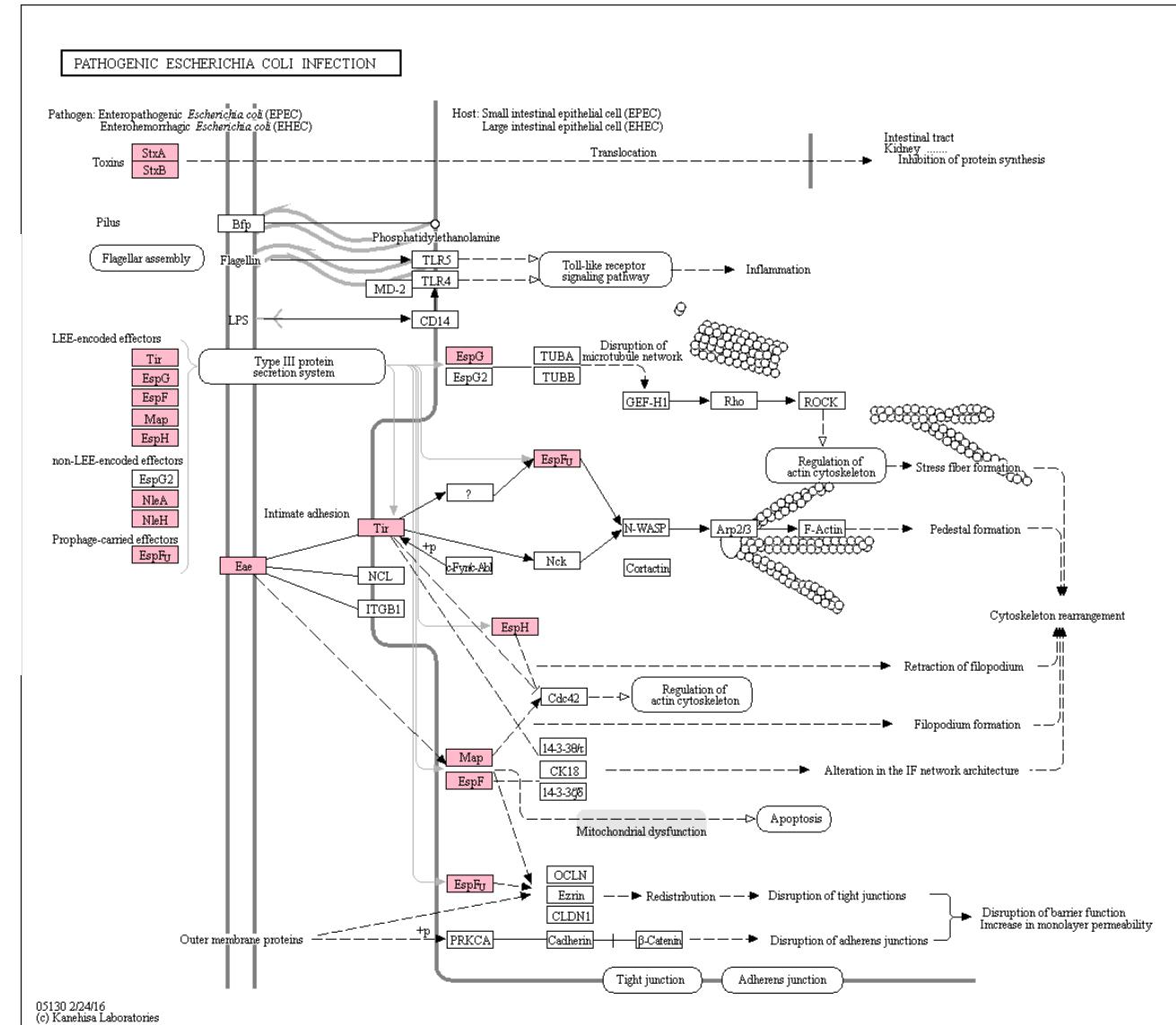
04014 Peptidyl-prolyl cis-trans isomerase

	Type I	Type III	Type II
Outer membrane protein (OMP)	TolC	Needle	Secretin
Membrane fusion protein	HlyD	YscF	GspD
ABC transporter	HlyB	YscO YscP YscX	GspS
Secretin	YscC		GspC GspF GspG
OMP	YscW		GspH GspI GspJ
Inner membrane protein (IMP)	YscJ YscR YscS	GspK GspL GspM	
	YscT YscU YscV	GspE	
ATPase	YscN	Leaderpeptidase	GspO
ATPase-associated protein	YscQ YscL		
	Sec-SRP		
IMP	SecD/F SecE SecG		
	SecY YajC VidC		
ATPase	SecA		
Secretion monitor	SecM		
SRP receptor	FtsY		
Targeting protein	SecB ffh		

Pathogenic Escherichia coli infection

Infectious diseases: Bacterial

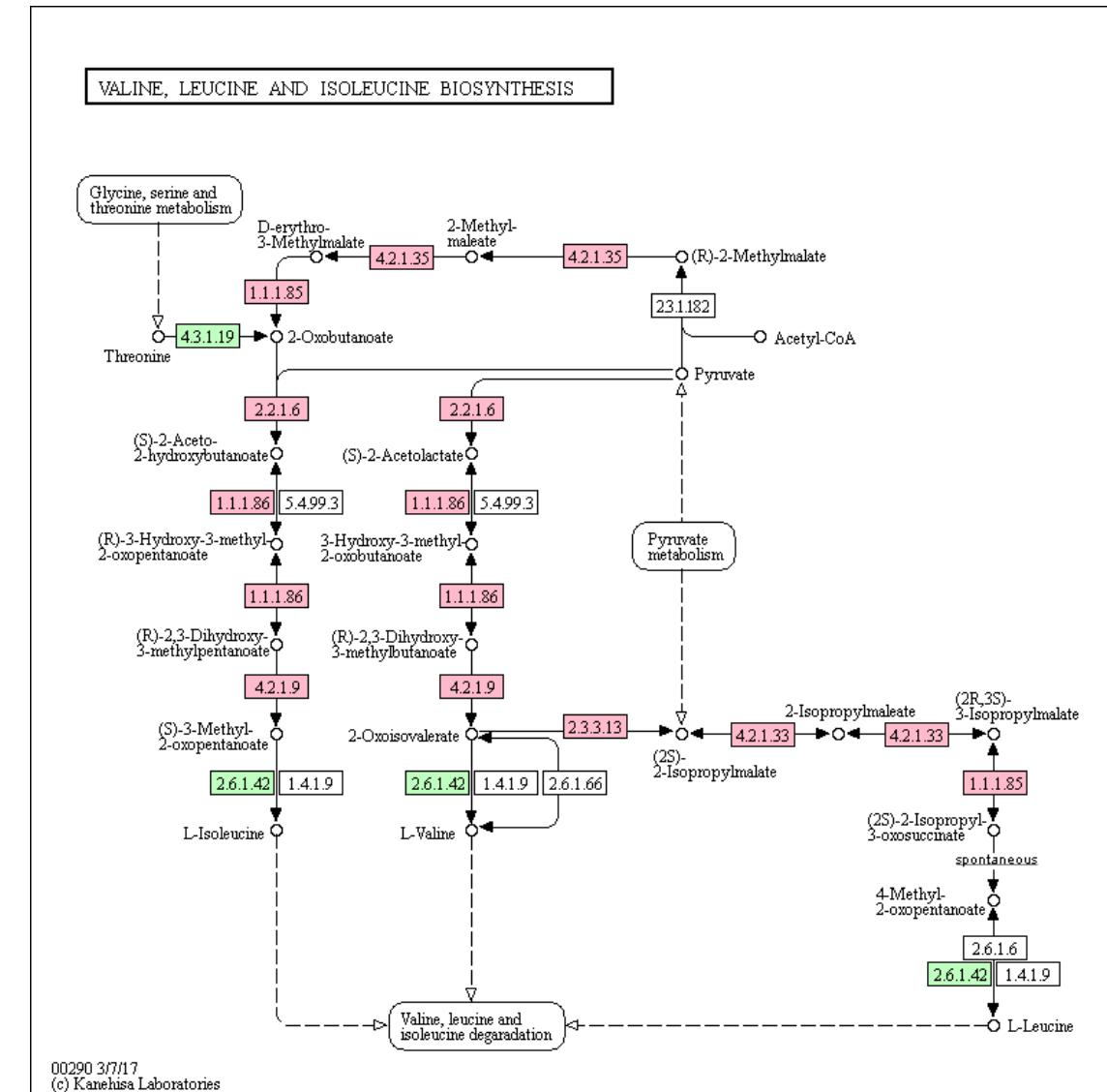
- 05110 Vibrio cholerae infection
- 05120 Epithelial cell signaling in Helicobacter pylori infection
- 05130 Pathogenic Escherichia coli infection
- 05132 Salmonella infection
- 05131 Shigellosis
- 05133 Pertussis
- 05134 Legionellosis
- 05150 Staphylococcus aureus infection
- 05152 Tuberculosis
- 05100 Bacterial invasion of epithelial cells



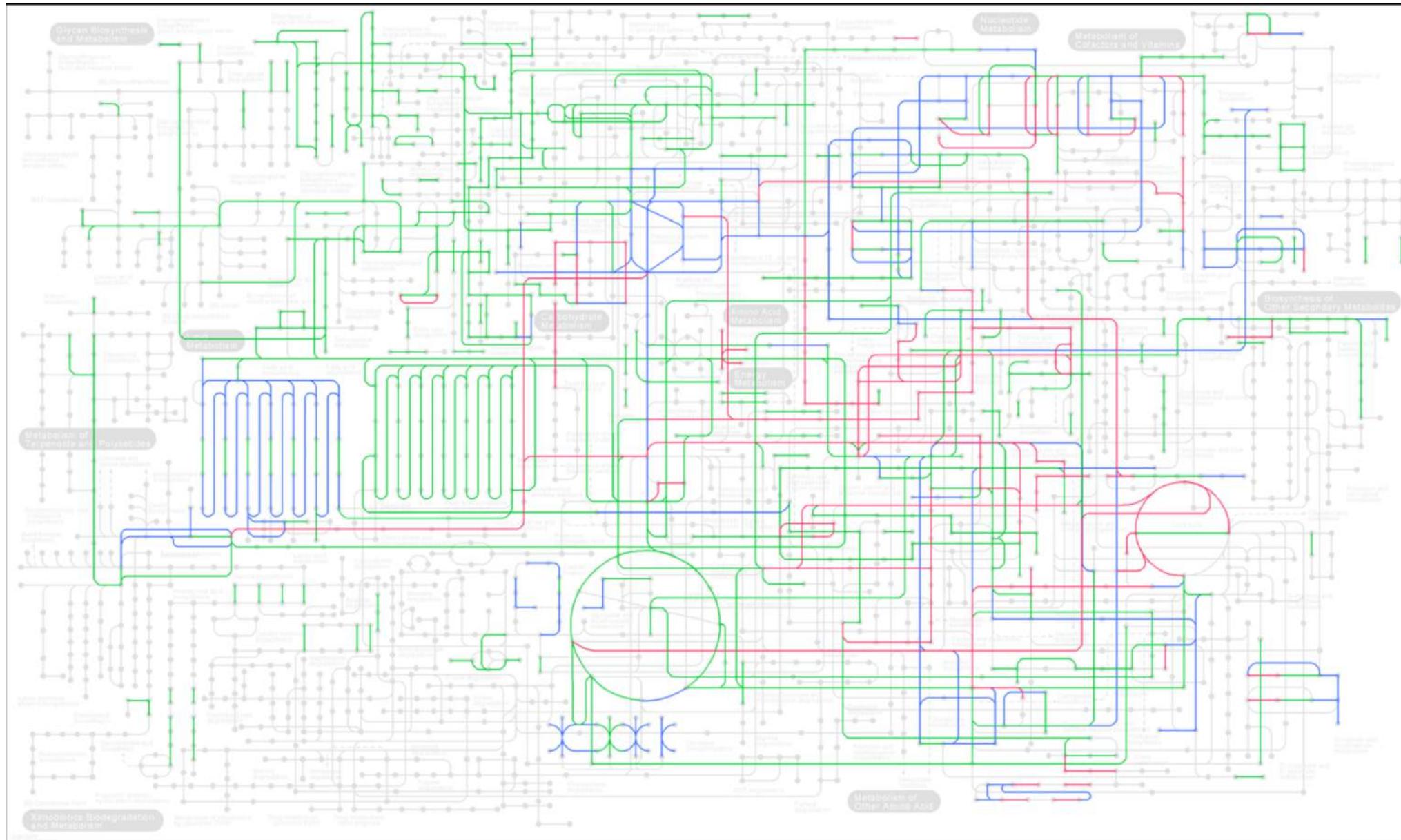
アブラムシとブフネラで種間比較

- 00290 : Valine, leucine and isoleucine biosynthesis マップ
 - アブラムシ : api
 - ブフネラ : buc
- 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

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



Overviewマップの場合



アブラムシ
ブフネラ
共通

サンプルデータのマッピング

- KEGGには遺伝子リストからパスウェイをマッピングするツールが組み込まれています。

- KEGG Mapper**

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

- 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-GeneID, 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 Mapper

A suite of KEGG mapping tools

About KEGG Mapper

Search Pathway
Search&Color Pathway
Color Pathway

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
Draw Phylogram

Annotate Sequence
BlastKOALA

KEGG Atlas
KEGG

About KEGG Mapper

KEGG Mapper is a collection of tools for **KEGG mapping**: KEGG pathway mapping, BRITE mapping, and MODULE mapping. Two pathway mapping tools, "Search Pathway" and "Color Pathway", were made available from the beginning of the KEGG project. The naming of these tools was somewhat misleading since both involved searching and coloring procedures. The latter is now called "Search&Color Pathway" and a new "Color Pathway" is introduced for coloring of a selected pathway map (without searching).

Pathway mapping tools

Search Pathway - basic pathway mapping tool

- Search against all pathway maps in one category (ko, org, etc.)
- Multiple search objects may be specified in one line
- Mapped objects are marked in red

Search&Color Pathway - pathway mapping tool with coloring options

- Search against all pathway maps in one category (ko, org, etc.)
- Each search object is specified in one line together with color attributes (background color and foreground color)
- Mapped objects can be marked in any color
- NCBI and UniProt identifiers may be used

Color Pathway - selected pathway map coloring tool

- One selected pathway map is colored in multiple ways
- Each object is specified in one line together with color attributes, i.e., multiple background and foreground colors in multiple columns
- Results shown as multiple snapshots of image maps
- When numerical values are given as attributes, they are displayed by color gradation or 3D bar graph

KEGG Mapper

- 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:

Examples: Homo sapience pathway

7167 red,blue
C00118 pink

Alternatively, enter the file name containing the data:

ファイルを選択 ファイル未選択

If necessary, change default bgcolor: pink

Include aliases

チンパンジーの遺伝子をマッピング

- テキストボックスに「例」のIDを記入
- Search against: にチンパンジーの生物種コードを入れる
 - コードがわからないので、org ボタンをクリック
 - ポップアップウィンドウでに 種名を入力すると、下のボックスに候補が出るので、選択したあと Select をクリック
 - チンパンジーのコード "ptr" が入力されていることを確認

例

453039 red
104003784 coral
453645 gray,red
453565 blue,yellow
450453 #fbfb88
463861 #88ffbb

Find three- or four-letter KEGG organism code

SelectCancel

chimp
chimpanzee) Pan troglodytes (chimpanzee) [ptr]

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

Primary ID: KEGG identifiers (Outside IDs for organism-specific p

Enter objects one per line followed by bgcolor, fgcolor:

```
453039 red
104003784 coral
453645 gray,red
453565 blue,yellow
450453 #fbfb88
463861 #88ffbb
```

Examples: Select

Alternatively, enter the file name containing the data:

ファイルを選択 ファイル未選択

数値データをマッピング

- [Color Pathway](http://www.kegg.jp/kegg/tool/map_pathway3.html) (http://www.kegg.jp/kegg/tool/map_pathway3.html) をクリック
- 右のサンプル Numerical values to colorを選択
 - 中身は配列 ID と数値の対応リスト
- Select KEGG pathway map: でパスウェイを指定 (hsa05200)
- Option: で Numerical values converted to color gradationを選択
- Exec ボタンをクリック

Select KEGG pathway map: hsa05200

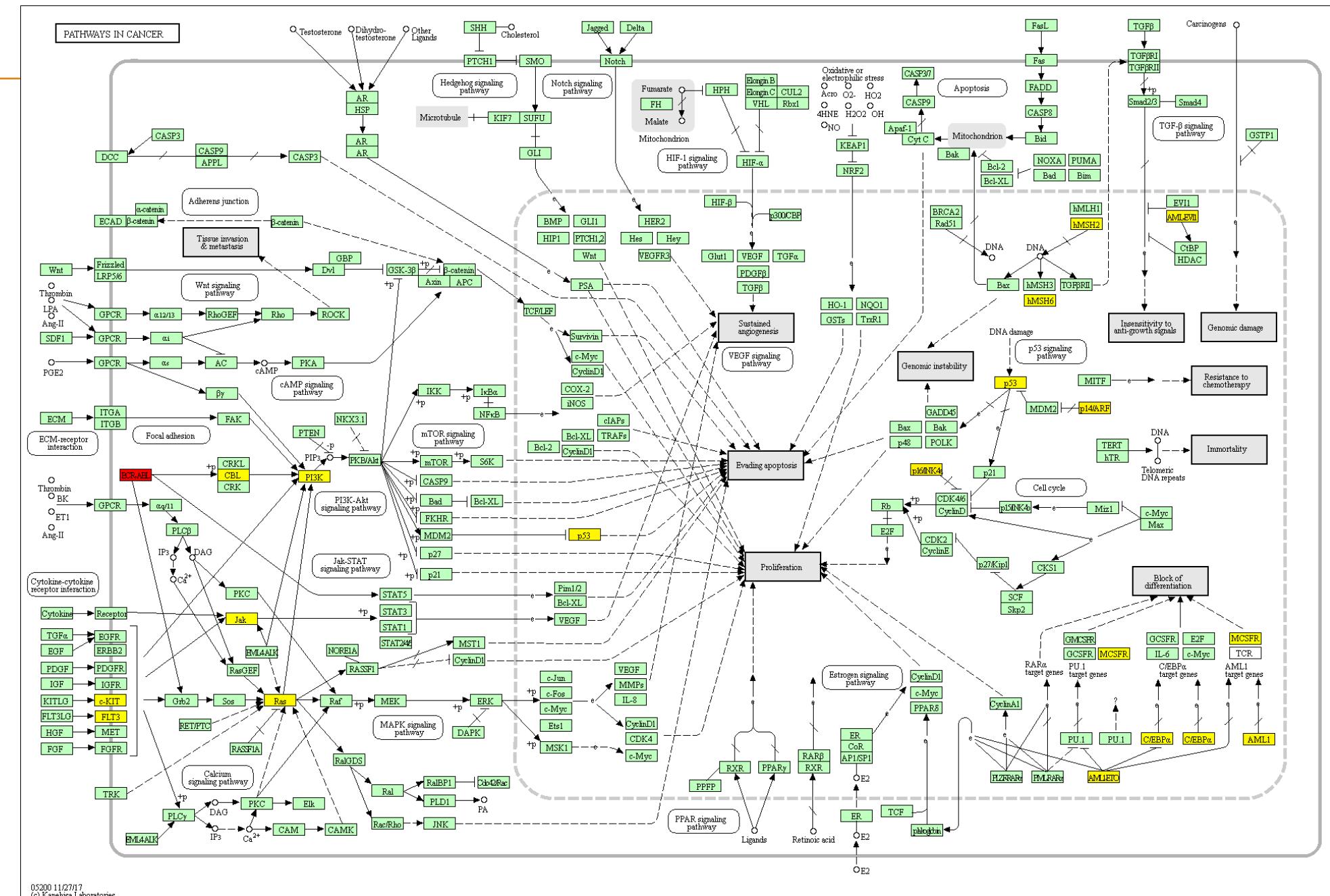
Enter data:

```
#hsa COSMIC
hsa:25    678
hsa:861   26
hsa:867   13
hsa:1029  44
hsa:1050  4
hsa:1436  3
hsa:2146  9
hsa:2322  7
hsa:2623  2
```

Examples:
Numerical values to color

Alternatively, enter the file name containing the data:

ファイルを選択 ファイル未選択



3Dグラフマッピング

- 右のサンプル Numerical values to colorを選択する

Select KEGG pathway map: hsa05200

Enter data:

```
#hsa COSMIC
hsa:25 678
hsa:861 26
hsa:867 13
hsa:1029 44
hsa:1050 4
hsa:1436 3
hsa:2146 9
hsa:2322 7
hsa:2623 2
```

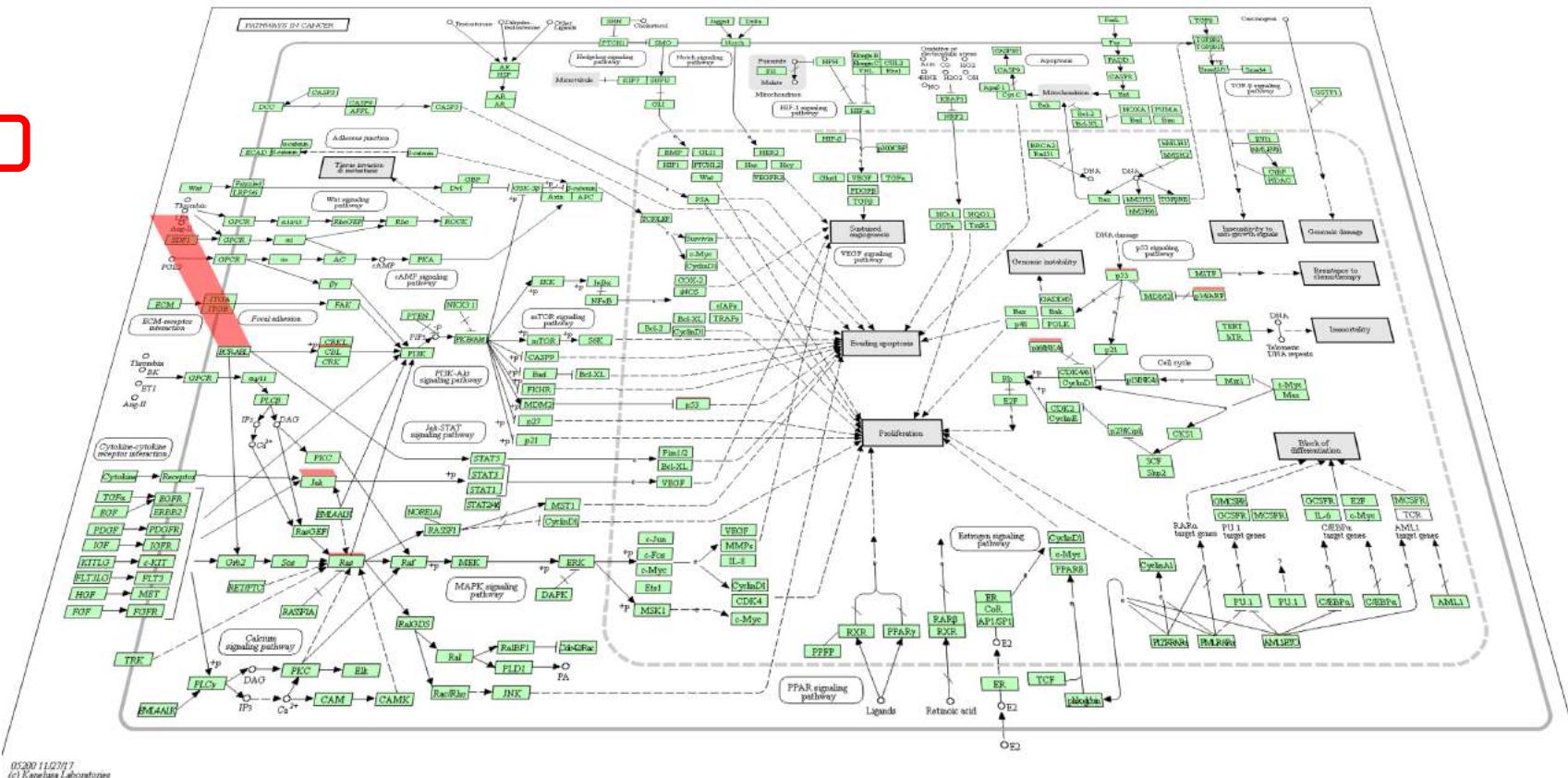
Numerical values to graph

Alternatively, enter the file name containing the data:

Option: Color specification
 Numerical values converted to color gradation
Range: minimum-maximum negative-zero-positive
minimum #ffff00 maximum #ff0000

Numerical values converted to 3D bar graph

Use uncolored diagram



KEGGデータベースにはない遺伝子のマッピング

- KEGG に登録されている配列データと類似性を計算し、自動で遺伝子機能を推定、パスウェイへのマッピングを行う。
- KAAS (<http://www.genome.jp/tools/kaas/>)
- BlastKOALA, GhostKOALA



KAAS - KEGG Automatic Annotation Server
for ortholog assignment and pathway mapping

Request

About KAAS
KAAS (KEGG Automatic Annotation Server) provides functional annotation of genes by BLAST or GHOST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways.

- KAAS Help

Complete or Draft Genome
KAAS works best when a complete set of genes in a genome is known. Prepare query amino acid sequences and use the BBH (bi-directional best hit) method to assign orthologs.

- KAAS job request (BBH method)

Partial Genome
KAAS can also be used for a limited number of genes. Prepare query amino acid sequences and use the SBH (single-directional best hit) method to assign orthologs.

- KAAS job request (SBH method)
- KAAS interactive

Metagenomes
When the query consists of large numbers of sequences and / or sequences from mixture of species such as those from metagenome sequencing project, we recommend the GHOSTX search and SBH method.

- KAAS job request (SBH method for amino acid sequence query)

Example of Results

KO assignment

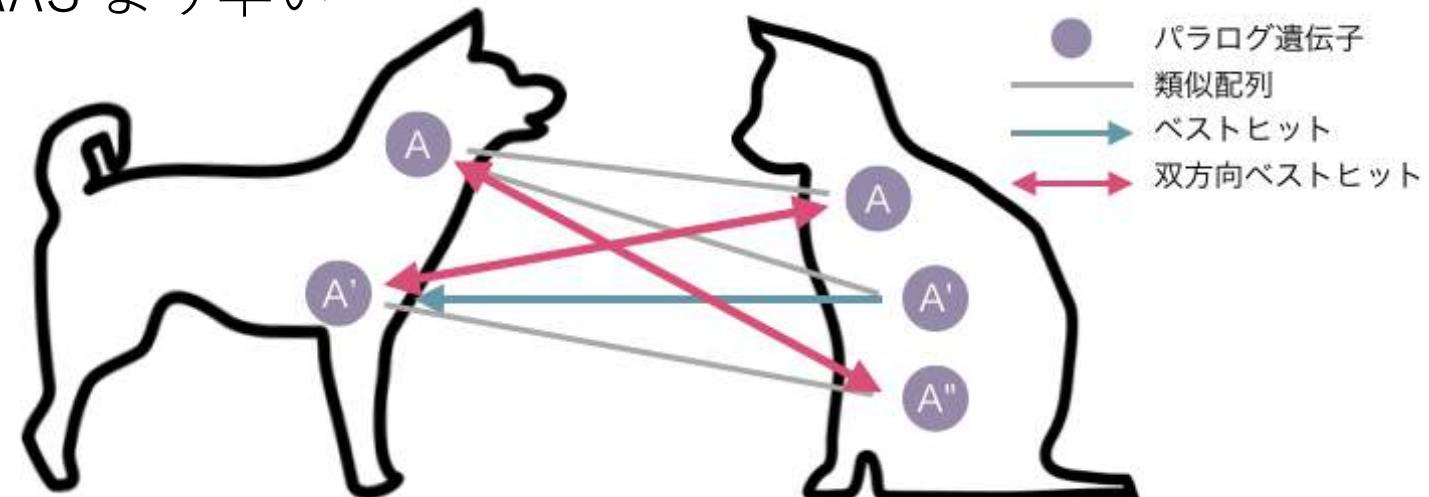
Query gene : KO assignment
ko0570411
query_2001
query_2002 x00093
query_2003 x00872
query_2004 x03753
query_2005
query_2006
query_2007 x03310
query_2008 x00518
query_2009 x03321
query_2010 x03324
query_2011
query_2012
query_2013
query_2014 x04043
query_2015 x03308
query_2016

KEGG pathway mapping

The diagram illustrates a complex network of metabolic pathways. Key nodes include Glycolysis, Gluconeogenesis, and various enzymes like Pyruvate kinase, Pyruvate dehydrogenase, and Phosphofructokinase. Arrows indicate the flow of metabolites between these pathways, showing the interconnectedness of biological processes.

配列相同意検索

- [KAAS](http://www.genome.jp/tools/kaas/) (<http://www.genome.jp/tools/kaas/>)
 - 配列類似性の計算は [BLAST](#), [GhostX](#), [GhostZ](#) ベースの 3 つ
 - GhostX は BLAST より精度は劣るが 100 倍早い
 - GhostZ は GhostX より精度は劣るが 2 倍早い
 - 種間で両方向ベストヒットを利用して遺伝子機能を推定（片方向も可能）
- [BlastKOALA](#), [GhostKOALA](#)
 - 配列類似性の計算は BLAST, GhostX ベース
 - クエリーからデータベースへの片方向の計算なのと、データベースを圧縮しているぶん KAAS より早い



BlastKOALAを使ったパスウェイマッピング

- [Annotate Sequence by BlastKOALA](#)
(http://www.kegg.jp/kegg/tool/annotate_sequence.html)
- Example: の sequence.txt をコピー & ペースト、もしくはダウンロードしてファイルを選択
- Family/Genus ボタンをクリック
 - サンプルが Buchnera の仲間なので、KEGG の Buchnera データを使う
 - 新たに開いたウィンドウで、Buchnera を探し、Taxonomy番号をクリック

KEGG Mapper – Annotate Sequence by BlastKOALA

About KEGG Mapper
Search Pathway
Search&Color Pathway
Color Pathway

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

Upload query amino acid sequences in FASTA format

Enter FASTA sequences

```
YLINFFYKTLKIKGTPIQIQFKDNENPYVKNNK  
>cl|NZ_ACFK0100001.1_prot_WP_009874555.1_575 [gene=C5S_RS0103170]  
[protein=/inference=EXISTENCE: similar to AA sequence:SwissProt:P57663.1]  
[protein_id=WP_009874555.1] [location=complement(641210..641791)]  
MLNISKKNIIFILFFIISLILFNWKYFSLVNKENLESLKYEIIKKINKKSKNLYEVENFIQ  
NTSI  
YGTLTALSLLAKKYVECNLKDALLQLNNSLKYTKNEENLKNLLKINIAKIQIQQNNNKAMNL  
ETIQNHN  
WKNIIEHMKGDIFININNKKEAIKSWKSLFIEDSNASKEIINMKLNELKEQN
```

Or upload file:

Select GENES family/genus dataset to be searched

Select Example:
Query data: [sequence.txt](#)
Select Buchnera (32199)

Exec Clear

Erwinia	551	Erwinia amylovora
		Erwinia billingiae
		Erwinia sp. Ejp617
		Erwinia gerundensis
Buchnera	32199	Buchnera aphidicola
Wigglesworthia	51228	Wigglesworthia glossinidia
		Pantoea ananatis
		Pantoea vagans
		Pantoea sp. At-Qh

計算中

- Exec ボタンをクリック
- 数分待つ



BlastKOALA

Job Request

Request accepted

Your job ID is 2fed9b298f4efd7e9a9eebed200f16a1fa687edb

Started at Mon Jan 8 15:09:48 JST 2018

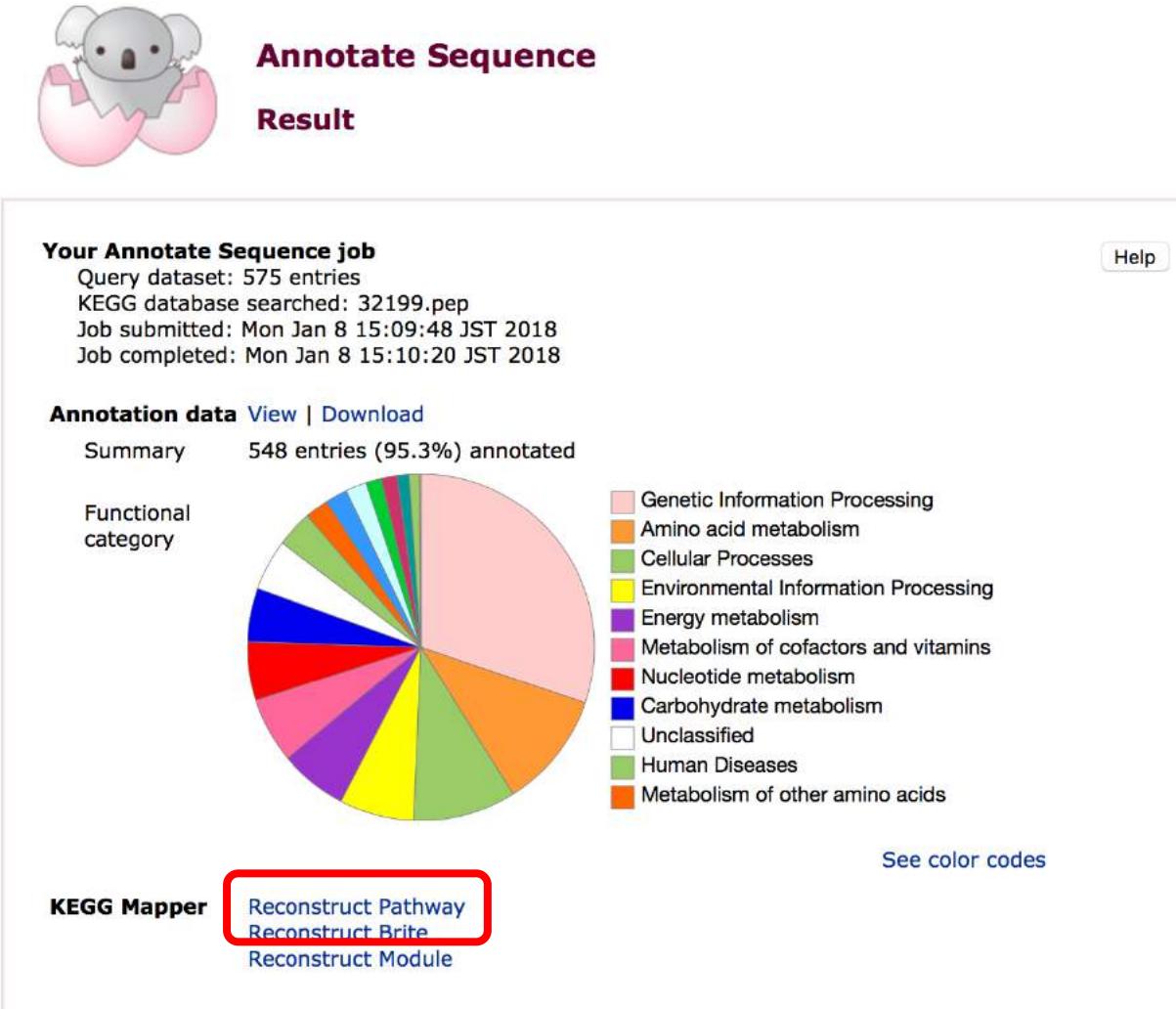
This page reloads automatically every 5 seconds

Your result will be displayed in the following page

[http://www.kegg.jp/kegg-bin/blastkoala_result?
id=2fed9b298f4efd7e9a9eebed200f16a1fa687edb&passwd=mTYL46&mode=mapper](http://www.kegg.jp/kegg-bin/blastkoala_result?id=2fed9b298f4efd7e9a9eebed200f16a1fa687edb&passwd=mTYL46&mode=mapper)

計算結果

- Reconstruct Pathway から遺伝子がマッピングされたパスウェイを見ることができる



Overviewマップの再構築結果

Pathway Reconstruction Result

Show all objects

Metabolism

Global and overview maps

- 0_100 Metabolic pathways (205)
- 0_110 Biosynthesis of secondary metabolites (97)

01120 Microbial metabolism in diverse environments (52)

01130 Biosynthesis of antibiotics (76)

01200 Carbon metabolism (30)

01210 2-Oxocarboxylic acid metabolism (10)

01212 Fatty acid metabolism (5)

01230 Biosynthesis of amino acids (62)

Carbohydrate metabolism

00010 Glycolysis / Gluconeogenesis (15)

00020 Citrate cycle (TCA cycle) (5)

00030 Pentose phosphate pathway (12)

00040 Pentose and glucuronate interconversions (1)

00051 Fructose and mannose metabolism (7)

00052 Galactose metabolism (1)

00500 Starch and sucrose metabolism (2)

00520 Amino sugar and nucleotide sugar metabolism (9)

00620 Pyruvate metabolism (6)

00630 Glyoxylate and dicarboxylate metabolism (2)

00640 Propanoate metabolism (3)

00650 Butanoate metabolism (2)

00660 C5-Branched dibasic acid metabolism (2)

00562 Inositol phosphate metabolism (2)

Energy metabolism

00190 Oxidative phosphorylation (27)

00195 Photosynthesis (8)

00710 Carbon fixation in photosynthetic organisms (7)

00720 Carbon fixation pathways in prokaryotes (4)

00680 Methane metabolism (8)

00920 Sulfur metabolism (9)

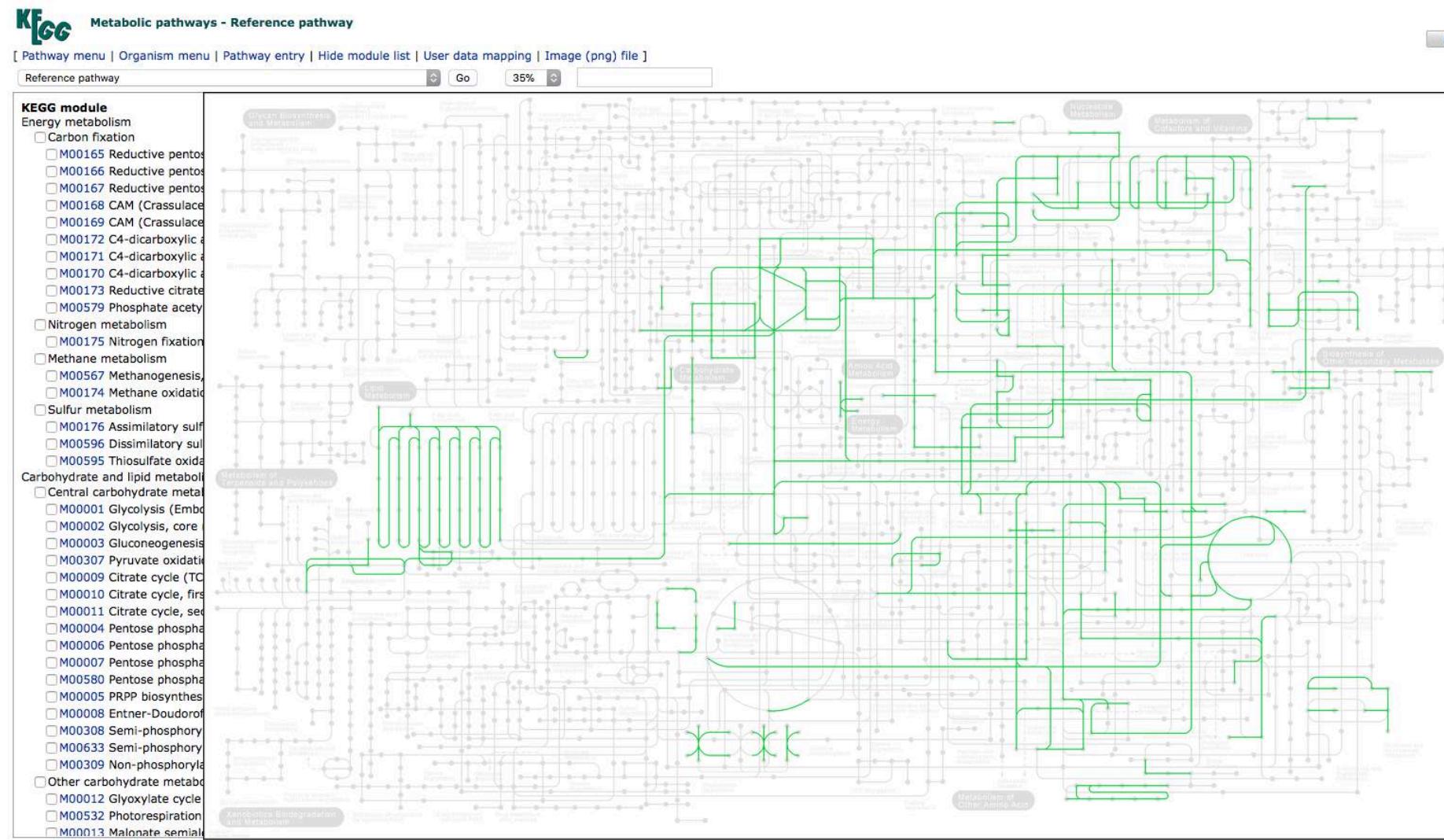
Lipid metabolism

00061 Fatty acid biosynthesis (5)

00564 Glycerophospholipid metabolism (1)

01040 Biosynthesis of unsaturated fatty acids (2)

Nucleotide metabolism



プロテオームデータベース

- プロテオームデータを閲覧できるデータベースとしてはペプチド検出の手法によって、質量分析に基づくものと抗体によるものとがある。
- 質量分析
 - ProteomicsDB (<https://www.proteomicsdb.org/>)
 - Human Proteome Map (<http://www.humanproteomemap.org/>)
- 抗体
 - The Human Protein Atlas (<https://www.proteinatlas.org/>)
- また、プロテオームデータそのものが保存されているリポジトリとしてのプロテオームデータベースも存在する。
- リポジトリ
 - PRIDE Archive (<https://www.ebi.ac.uk/pride/archive/>)
 - jPOST repository (<https://repository.jpostdb.org/>)

ProteomicsDB

- <https://www.proteomicsdb.org/>
- 網羅的なヒトプロテオームを大量の質量分析データから同定した結果がデータベース化されている。

TUM SAP Proteomics DB powered by SAP HANA

HOME HUMAN PROTEINS PEPTIDES CHROMOSOMES ANALYTICS API PROJECTS FAQ ABOUT US NEWS

Status Human Proteome

Coverage:	80%
Proteins:	15721 of 19629
Isoforms:	11353 of 86771
Unique Peptides (Isoform):	113944
Unique Peptides (Gene):	455289
Spectra:	43237800

Repository

Registered Users:	745
Projects:	79
Experiments:	705
Files:	22810
Data Volume:	8.85 TB

Recently Published

- Klaeger_Science_2017
- Cellzome_Cellsurface
- Cellzome_2D-TPP_pano

Welcome to ProteomicsDB!

ProteomicsDB is a joint effort of the Technische Universität München (TUM) and SAP SE. It is dedicated to expedite the identification of the human proteome and its use across the scientific community.

Browse proteins
Explore the human proteome protein by protein.

Browse chromosomes
Explore the human proteome chromosome by chromosome.

Latest News

Fri, Dec 01, 2017
Target landscape of 243 clinical kinase inhibitors available in ProteomicsDB
ProteomicsDB enables the interactive exploration of the target space of 243 clinically relevant kinase inhibitors.
[more ...](#)

Mon, Aug 14, 2017
Hardware maintenance finished
ProteomicsDB has been successfully moved to the new hardware.
[more ...](#)

Mon, Mar 13, 2017
Ongoing hardware maintenance
Due to hardware maintenance work, short periods of downtime might occur.
[more ...](#)

News Archive

タンパク質検索

- HUMAN PROTEINSタブをクリック
- 適当なタンパク質名（ここでは「APC」）を入力し、リターンキーを押す
- トップヒットしたエントリをクリック

The screenshot shows a web-based protein search interface. At the top, there's a navigation bar with links for HOME, HUMAN PROTEINS (which is the active tab), PEPTIDES, CHROMOSOMES, ANALYTICS, API, PROJECTS, and FACETS. Below the navigation bar is a search bar with three tabs: Proteins (selected), Orphans, and APC. A red box highlights the APC tab. To the right of the tabs is a magnifying glass icon. Underneath the search bar is a section titled 'PROTEIN LIST' with a 'Download' button and a dropdown menu for 'Proteins per page' set to 15, with options for 25, 50, and 100. The main area displays a table with columns: Protein, Uniprot Accession, Uniprot Identifier, Database Source, and Description. The first two rows for 'APC' are highlighted with red boxes: the first row for 'APC_HUMAN' and the second row for its isoform 'P25054-2'. The table continues with entries for 'FAM123B' and 'FAM123C'.

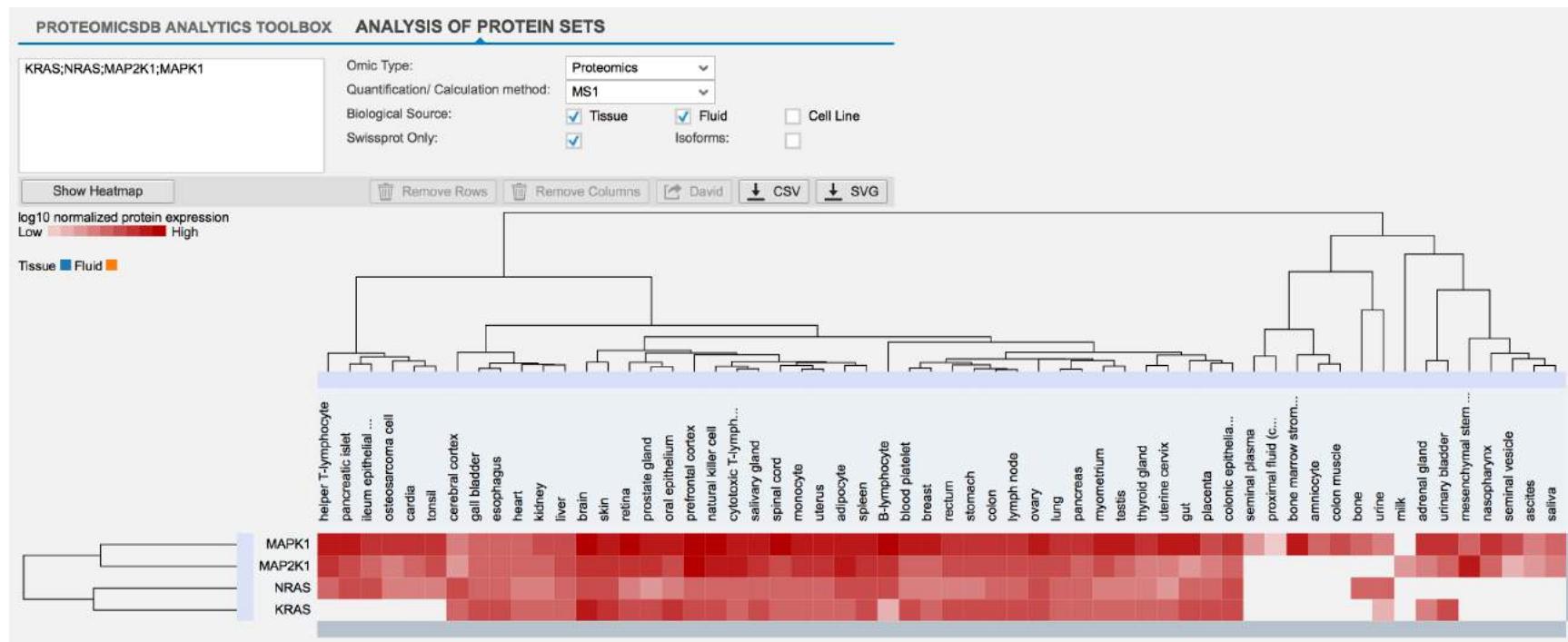
Protein	Uniprot Accession	Uniprot Identifier	Database Source	Description
● APC	APC_HUMAN	P25054	sp	Adenomatous polyposis coli protein
● APC	APC_HUMAN	P25054-2	sp	Isoform Short of Adenomatous polyposis coli protein
● FAM123B	AMER1_HUMAN	Q5JTC6	sp	APC membrane recruitment protein 1
● FAM123C	AMER3_HUMAN	Q8N944	sp	APC membrane recruitment protein 3
● FAM123D	AMER4_HUMAN	Q5ITC6	sp	Isoform 3 of APC membrane recruitment protein 4

APCタンパク質の詳細情報



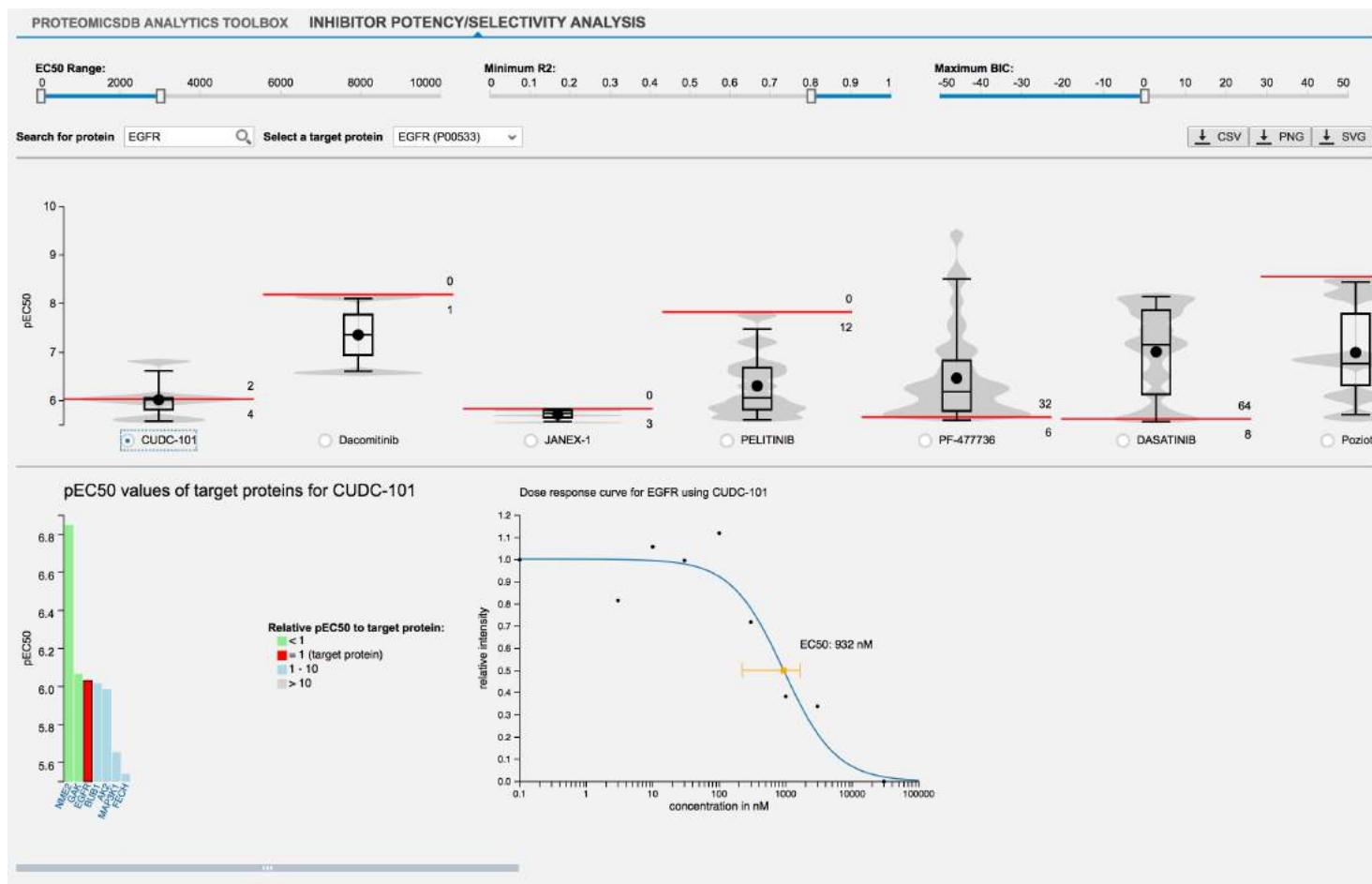
データ解析

- ANALYTICSタブを選択
- ヒートマップのアイコンをクリック
- タンパク質名を入力（例：KRAS;NRAS;MAP2K1;MAPK1）



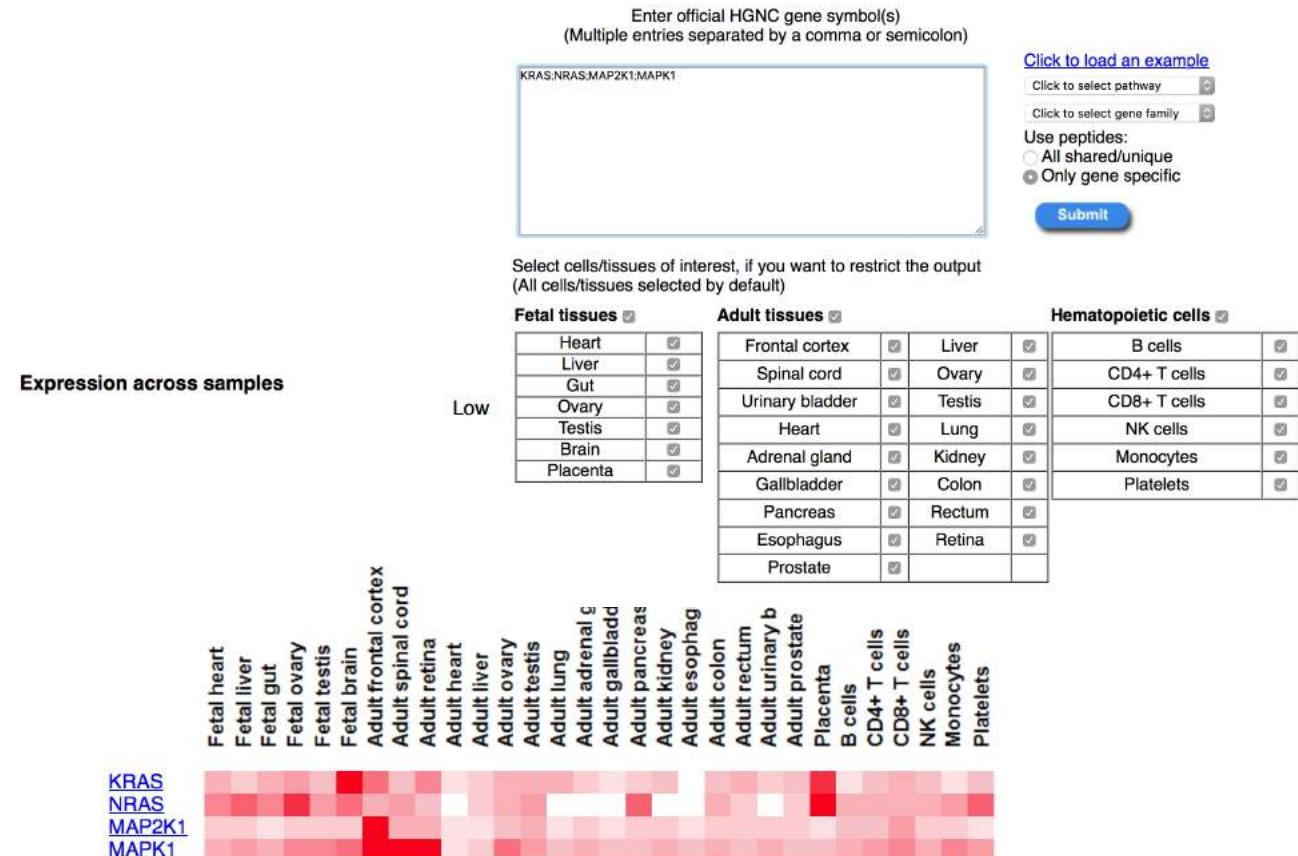
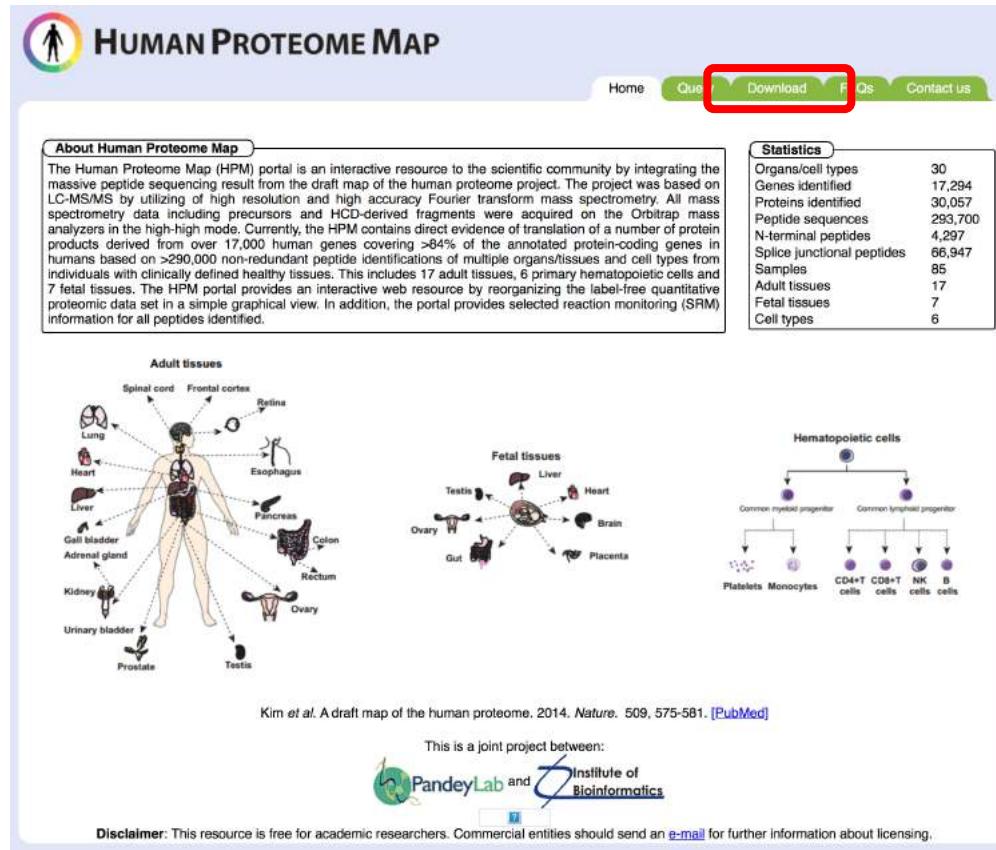
データ解析

- ボックスプロットのアイコンをクリック
- Search for proteinにEGFRを入力し、リターンキーを押す



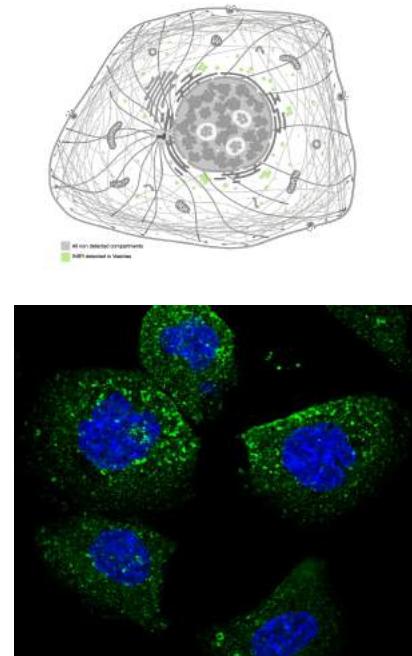
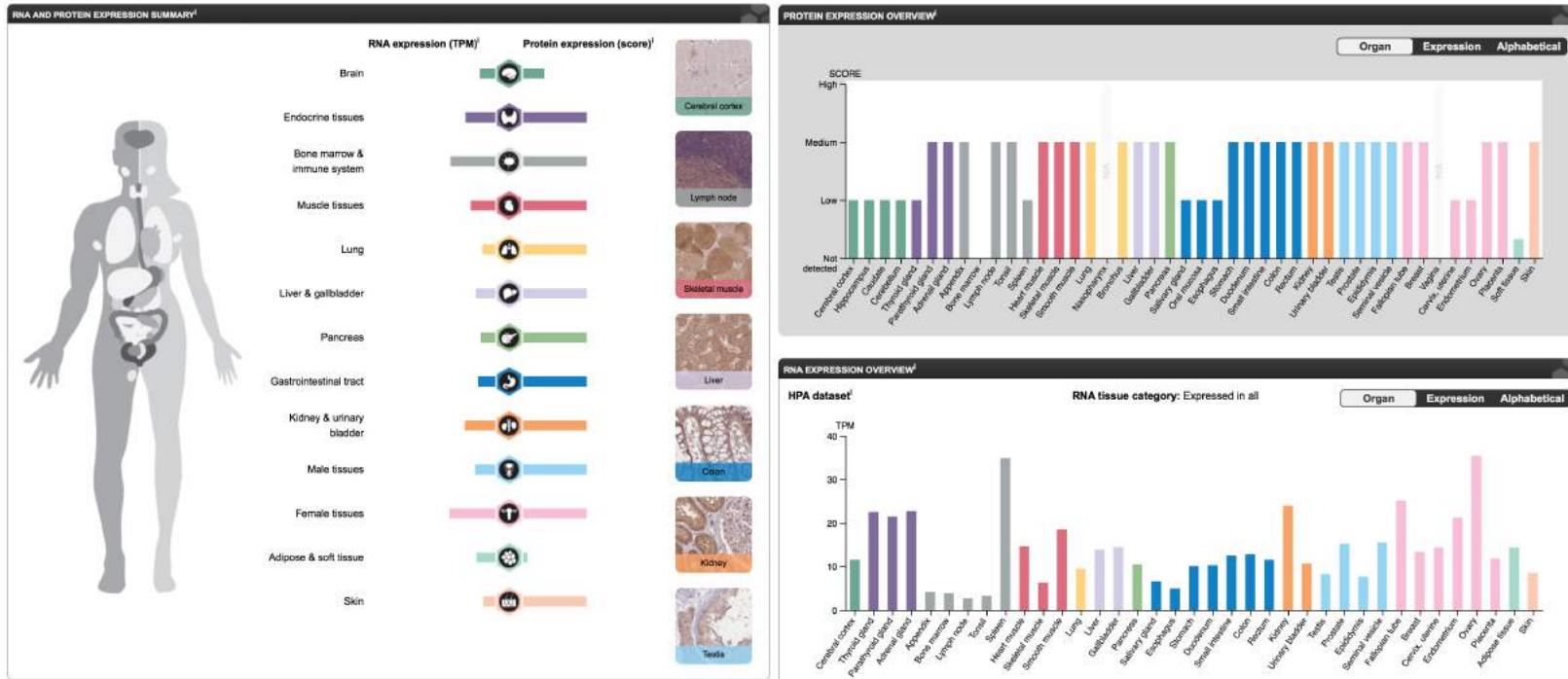
Human Protein Map

- <http://www.humanproteomemap.org/>
- Queryタブを選択
- 先ほどと同じリスト (KRAS;NRAS;MAP2K1;MAPK1) を入力



The Human Protein Atlas

- <https://www.proteinatlas.org/>
- 抗体ベースで検出されたタンパク質の情報がデータベース化されている。
- また、RNAseqのデータも登録されているため、RNAの発現とタンパク質の発現とを同時に観察することが出来る。
- 組織画像も豊富に登録されている。
- 例：insulin

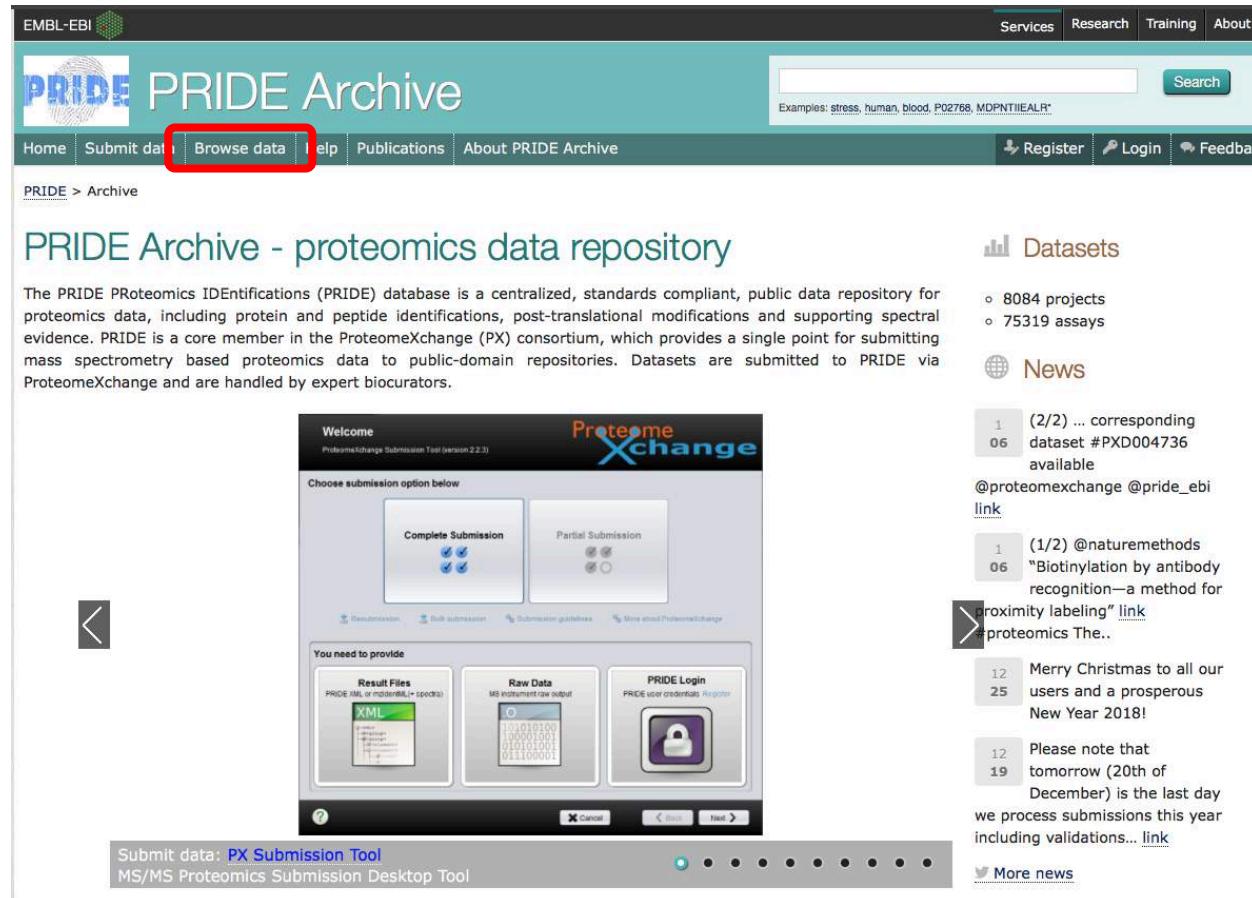


リポジトリデータベース

- 質量分析データを利用した論文では、生データをリポジトリデータベースに登録し、アクセッション番号を論文内に記載することが推奨されている。
- リポジトリデータベース
 - PRIDE Archive (<https://www.ebi.ac.uk/pride/archive/>)
 - Massive
(<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>)
 - PASSEL (<http://www.peptideatlas.org/>)
 - jPOST repository (<https://repository.jpostdb.org/>)

PRIDE Archive

- <https://www.ebi.ac.uk/pride/archive/>
- EBIが運営するプロテオームデータのリポジトリ



The screenshot shows the PRIDE Archive homepage. At the top, there is a navigation bar with links for Home, Submit data, Browse data, Help, Publications, About PRIDE Archive, Register, Login, and Feedback. The 'Browse data' link is highlighted with a red box. Below the navigation bar, there is a search bar with placeholder text 'Examples: stress, human, blood, P02768, MDPNTIIELR*' and a 'Search' button. The main content area has a teal header 'PRIDE Archive - proteomics data repository'. Below the header, there is a brief introduction about the PRIDE database and its submission process. To the right of the introduction, there are two sections: 'Datasets' (listing 8084 projects and 75319 assays) and 'News' (listing several recent news items). At the bottom of the page, there is a 'Submit data' section featuring the 'PX Submission Tool' and 'MS/MS Proteomics Submission Desktop Tool'.

PRIDE Archive - proteomics data repository

The PRIDE PRoteomics IDEntifications (PRIDE) database is a centralized, standards compliant, public data repository for proteomics data, including protein and peptide identifications, post-translational modifications and supporting spectral evidence. PRIDE is a core member in the ProteomeXchange (PX) consortium, which provides a single point for submitting mass spectrometry based proteomics data to public-domain repositories. Datasets are submitted to PRIDE via ProteomeXchange and are handled by expert biocurators.

Welcome
ProteomeXchange Submission Tool (version 2.2.3)

Choose submission option below

Complete Submission Partial Submission

You need to provide

Result Files (PRIDE XML or mzML+ spectra) Raw Data (MS instrument raw output) PRIDE Login (PRIDE user credentials: Register)

Submit data: [PX Submission Tool](#)
[MS/MS Proteomics Submission Desktop Tool](#)

Datasets

- 8084 projects
- 75319 assays

News

1 06 (2/2) ... corresponding dataset #PXD004736 available @proteomexchange @pride_ebi link

1 06 (1/2) @naturemethods "Biotinylation by antibody recognition—a method for proximity labeling" link proteomics The..

12 25 Merry Christmas to all our users and a prosperous New Year 2018!

12 19 Please note that tomorrow (20th of December) is the last day we process submissions this year including validations... link

[More news](#)

PRIDE Archive

- 現在約4500件のデータが収納されている
- FieldをSpecies、ContainsをHomo sapiensでフィルターする

PRIDE > Archive > Search results

Search results

4584 Search results

Filter your results

Field

Species

Contains

-Any-

Add filter

Page 1 2 3 4 5 ... 459 Showing 1 - 10 of 4584 results Page size 10 20 50 100

Sort by: Accession Title Relevance ▾ Publication date

PXD006605 ✓

Insights In The Molecular Regulation Of Monolignol-derived Product Biosynthesis In The Growing Hemp Hypocotyl

Species: Cannabis

Project description: Lignin and lignans are both deriving from the monolignol pathway (More)

Made public: 2018-01-08

Biological Dataset

PXD008078

Nudt21 controls cell fate by connecting alternative polyadenylation to chromatin signaling

Species: Mus musculus (Mouse)

Project description: Cell fate transitions involve rapid changes of gene expression p (More)

Made public: 2018-01-08

Biological Dataset

PRIDE Archive

- ヒトに関するMSデータ
- さらにSubmission typeをCompleteでさらにフィルターを追加

Search results

1866 Search results + 1 filters

Filter your results

Page 1 2 3 4 5 ... 187 Showing 1 - 10 of 1866 results Page size 10 20 50 100

Sort by: Accession Title Relevance ▾ Publication date

Field

Species

Contains

-Any-

Add filter

PXD008620

Characterization of pericardial fluid proteome through fractionation with magnetic nanoparticles (NPs@EDTA)

Species: **Homo sapiens (Human)**

Project description: Despite the invasive nature of its collection, being in close co (More)

Made public: 2018-01-09

Technical Dataset

Biomedical Dataset

Current active filters

Remove all

Species:

Homo sapiens (Human)

PXD003459

The interactome of the human RNA Pol II

Species: **Homo sapiens (Human)**

Project description: We aim to describe the proteins interacting with the human RNA p (More)

Made public: 2018-01-09

Biological Dataset

PXD008270

Tyrosine-1 of RNA polymerase II CTD controls global termination of gene transcription in mammals

Species: **Homo sapiens (Human)**

Project description: The aim of the project was to analyse the functional role of Tyr (More)

Made public: 2018-01-09

Biological Dataset

PartialとComplete submission

- Complete submissionとは、データの再解析にとって必要とされる情報・データが揃っていることを保証した状態のデータをリポジトリに投稿すること。
- 質量分析の生データ (Rawファイル)、スペクトルのピークデータ (Peakファイル)、ペプチド検索結果データ (Resultsファイル) が揃っている
- Raw-Peak-Resultsの対応関係が取れている。
- Species、Modification、Instrumentの情報が揃っている。
- などの諸条件を満たす必要がある。
- Human Proteome Project (HPP) などはComplete submissionでのデータ投稿を必須条件としている。

Complete submissionデータ

Search results

407 Search results + 2 filters

Filter your results

Field
Species
Contains
-Any-
Add filter

Page 1 2 3 4 5 ... 41 Showing 1 - 10 of 407 results Page size 10 20 50 100

Sort by: Accession Title Relevance ▾ Publication date

PXD006572 ✓

Aging-related proteome alterations in B cells may predispose for chronic lymphocytic leukemia - cytoplasmic proteins of young B cells

Species: Homo sapiens (Human)

Project description: Chronic lymphocytic leukemia (CLL), the most common type of leuk (More)

Made public: 2018-01-05

Submission type: COMPLETE

PXD006799 ✓

Oxygen-Sensitive Remodeling of Central Carbon Metabolism by Archaic eIF5B

Species: Homo sapiens (Human)

Project description: The eukaryotic translation initiation factor 5B (eIF5B) is a hom (More)

Made public: 2018-01-03

Submission type: COMPLETE

Biological Dataset Biomedical Dataset

PXD002600 ✓

Protein expression in TKI-resistant RCC cells.

Species: Homo sapiens (Human)

Project description: Resistance to tyrosine kinase inhibitors is an essential issue c (More)

Current active filters

Remove all

Species:

Homo sapiens (Human) X

Submission Type contains:

Complete submissionデータ

- 選択したプロテオームデータのペプチドとタンパク質の情報へのリンク：Peptide tableとProtein table

Project PXD006572 ✓

Download Project Files

Project Protein Table (link)
Project Peptide Table (link)

Visualize in PRIDE Inspector

Summary

Title
Aging-related proteome alterations in B cells may predispose for chronic lymphocytic leukemia - cytoplasmic proteins of young B cells

Description
Chronic lymphocytic leukemia (CLL), the most common type of leukemia in adults, is still incurable despite the development of novel therapeutic strategies. This reflects the incomplete understanding of the pathophysiology of this disease. In order to get more detailed insights into CLL development, we performed a comprehensive proteome analysis of primary human CLL cells and B cells from young and age-matched healthy

[Read more](#)

Sample Processing Protocol
Non-coagulated healthy whole blood samples were collected in CPDA tubes (Greiner Bio-One GmbH, Austria) and diluted 1:2 with PBS buffer. The diluted blood suspension was subsequently carefully overlaid on Ficoll Paque (GE Healthcare, Bio-Sciences AB, Uppsala, Sweden) followed by centrifugation at 720g for 20min at room temperature with slow acceleration and very slow deceleration. PBMCs were collected from the Ficoll

Species	Tissue
Homo sapiens (Human)	blood

Cell Type
B cell

Instrument	Software
Q Exactive	Not available

Modification	Quantification
Carbamidomethyl residues	Not available
isobaric at 128.058578	
Da	
Acetyl	
Oxidation	

Peptide table

- 検出されたペプチド配列の情報

Peptides in Project PXD006572

Visualize in PRIDE Inspector

155272 Search results

Filter your results

Field
Modification

Contains
-Any-

Page 1 2 3 4 5 ... 15528		Showing 1 - 10 of 155272 results					Page size 10 20 50 100		
#	Peptide Sequence	Submitted Id	Assay	Search Engine Score	Experimental m/z	Charge	PRIDE Cluster Quality		
1	-AAEAADLGLGAAVPVELR.-	Q9Y5Q8	78322	Mascot:score 50.85	882.4761353	2			
2	-EAFVVEEWVK.-	P35606	78322	Mascot:score 35.91	617.814209	2			
3	-EAFVVEEWVK.-	P35606	78322	Mascot:score 30.46	617.8162231	2			
4	-LPEAAFLAR.-	P35606	78322	Mascot:score 52.33	493.78125	2			
5	-QLAELAISK.-	P35606	78322	Mascot:score 46.07	486.2860107	2			
6	-QLAELAISK.-	P35606	78322	Mascot:score 65.4	486.2863464	2			
7	-TFEV CDLPVR .-	P35606	78322	Mascot:score 38.71	617.8032837	2			
8	-TFEV CDLPVR .-	P35606	78322	Mascot:score 40.2	617.8049316	2			
9	-TYLPSQVSR.-	P35606	78322	Mascot:score 28.34	525.2791748	2			
10	-VFNYNTLER.-	P35606	78322	Mascot:score 42.71	577.7894897	2			

Page 1 2 3 4 5 ... 15528

Showing 1 - 10 of 155272 results

Page size 10 20 50 100

Protein table

- 検出されたタンパク質の情報

Proteins in Project PXD006572

Visualize in PRIDE Inspector

29921 Search results

Filter your results

Field
Modification

Contains
-Any-

Add filter

Showing 1 - 10 of 29921 results							Page size	10	20	50	100	
#	Submitted Id	Assay	Curated Id	Cross Refs	Protein Description	Modifications	Ambiguity members					
1	O00566		78322	O00566	 U3 small nucleolar ribonucleoprotein protein MPP10 (M phase phosphoprotein) More		O00566					
2	P21918		78322	P21918	 D(1B) dopamine receptor (D(5) dopamine receptor) (D1beta dopamine receptor) More		P21918					
3	P35606		78322	P35606	 Coatomer subunit beta' (Beta'-coat protein) (Beta'-COP) (p102) More		Carbamidomethyl	Oxidation	P35606			
4	P53621		78322	P53621	 Coatomer subunit alpha (Alpha-coat protein) (Alpha-COP) (HEP-COP) (HEPCOP) More		Carbamidomethyl	Oxidation	P53621			
5	Q86VX2		78322	Q86VX2	 COMM domain-containing protein 7		Q86VX2					
6	Q9NX08		78322	Q9NX08	 COMM domain-containing protein 8		Acetyl	Q9NX08				

PRIDE Inspector

- PRIDE Inspectorというソフトウェアを使ってデータを見ることも出来る。

Download Project Files

Project Protein Table

Project Peptide Table

Visualize in PRIDE Inspector

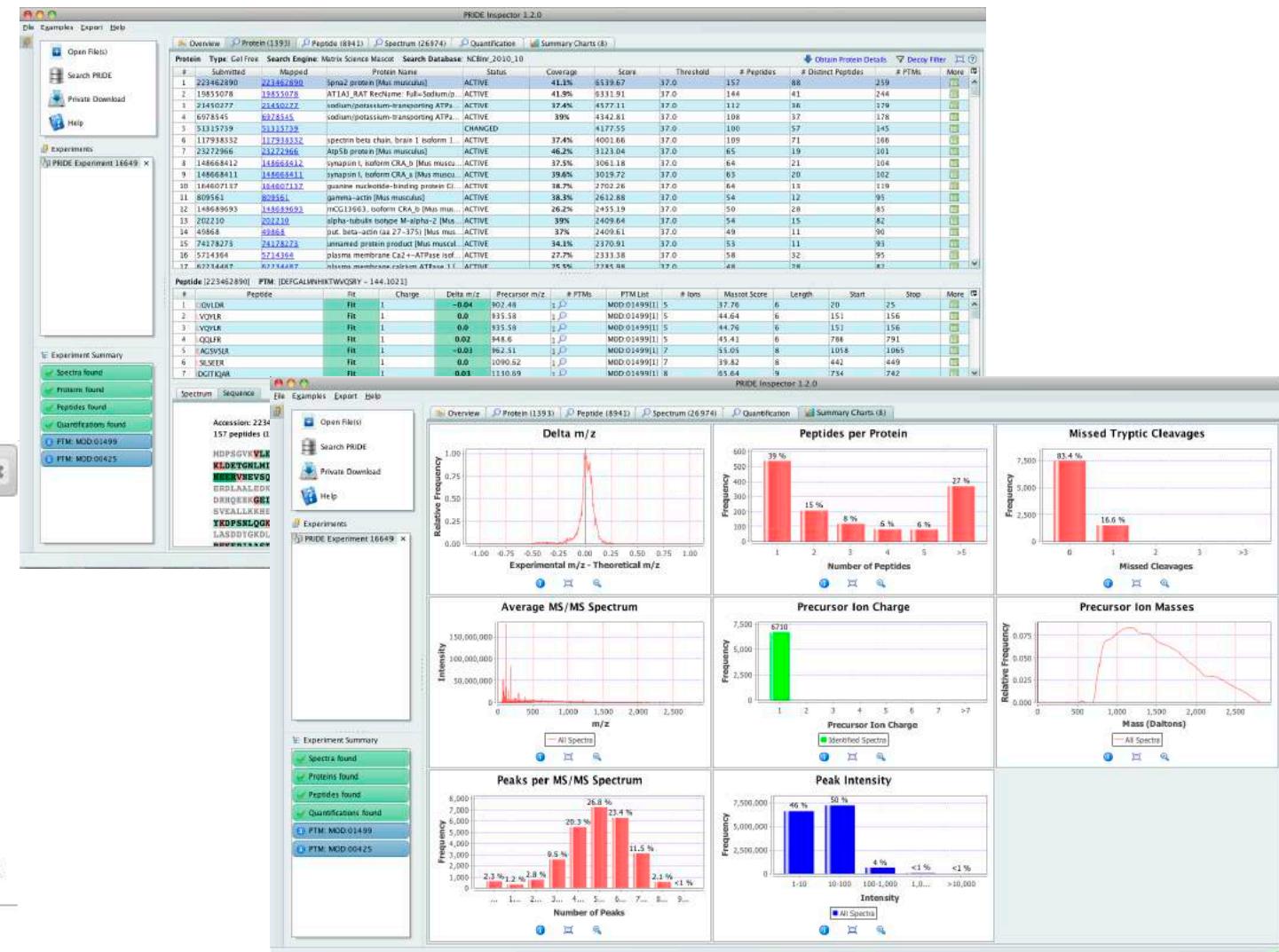
Open in PRIDE Inspector

Follow the next three steps to open your selected project or assay in PRIDE Inspector:

- Download, uncompress and open PRIDE Inspector
- Click in the magnifier on the left top corner, paste the project or assay that you would like to open in the search box, in this case **PXD006572**, and hit search
- Click in the corresponding "Download" button to download the files and visualize them

PRIDE Inspector documentation

Download



jPOST プロジェクト

jPOSTプロジェクトはNational Bioscience Database Center, Japan Science and Technology Agency (NBDC-JST)のサポートの下、開発されている。



➤ **jPOST プロジェクト開始 (2015年4月1日)**
(<https://www.jpost.org/>)

- ✓ リポジトリサイト構築
- ✓ 再解析パイプライン開発
- ✓ 再解析データに基づいたデータベース開発

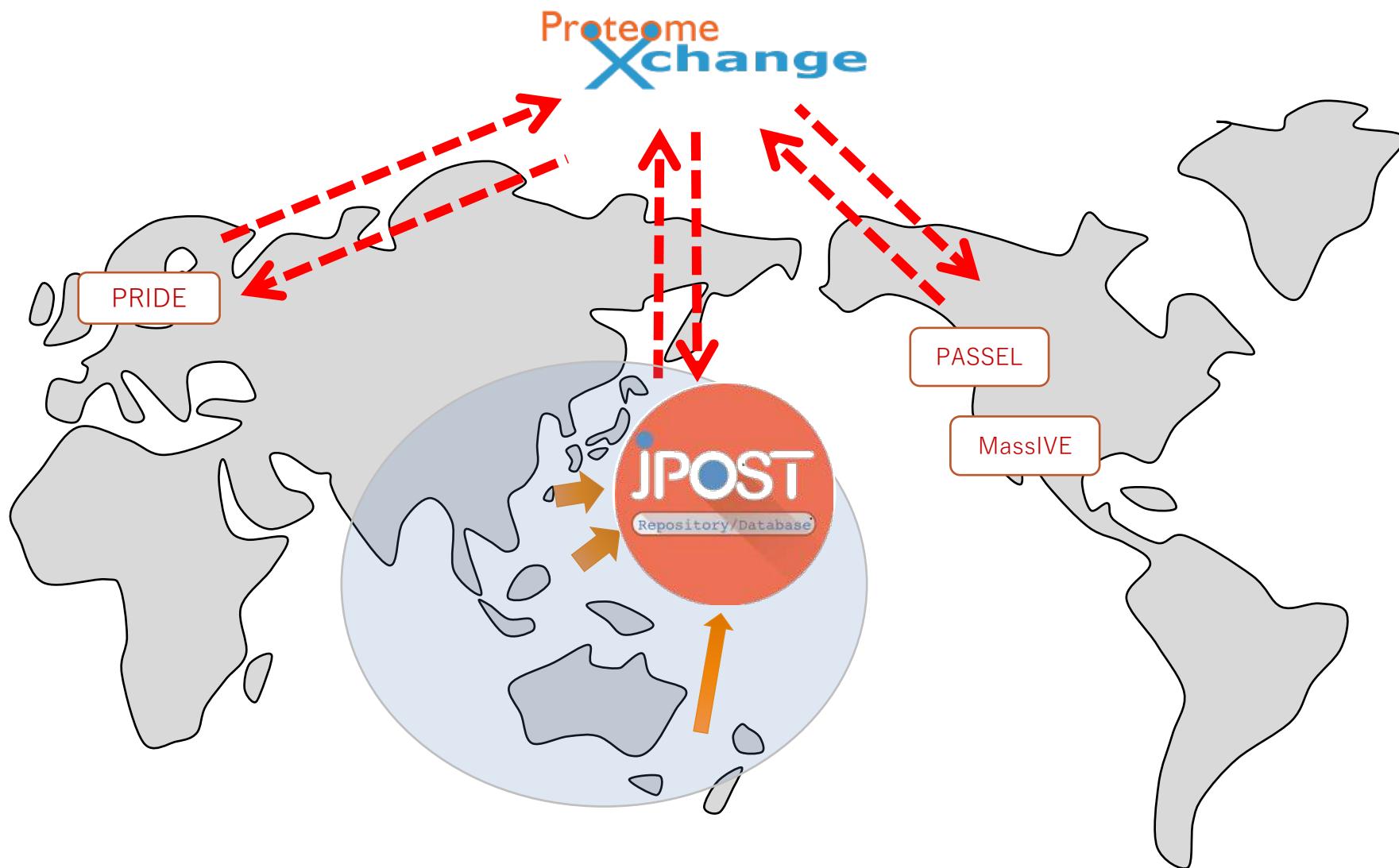


➤ **jPOST リポジトリ正式公開 (2016年5月2日)**
(<https://repository.jpostdb.org/>)

- ✓ PX partnershipに正式承認 (2016年7月6日)

➤ **jPOST データベース開始予定 (2018年3月)**

ProteomeXchange



jPOST repository

<https://repository.jpostdb.org>

The screenshot shows the jPOST repository homepage. At the top is a navigation bar with the jPOST logo, Repository, Submit, Help, Sign in, and Sign up buttons. Below the navigation is a section titled "About jPOSTrepo" with a brief description of the repository's purpose and a reference to a paper by Okuda et al. (2017). To the right of this is a 3D illustration of a globe with a stack of papers and a database icon. Below the "About" section is a "Statistics" area showing 270 registered projects, 29,756 files totaling 7.0 TB, and 33 species. It includes a line chart of project growth from September 2017 to January 2018 and a pie chart of species distribution. The "Data list" section features a search bar for free words or ontology keywords, filters for project type (All, Mass spectrometry, Gel electrophoresis, Antibody), and buttons for Search and Reset. A table at the bottom lists two projects: "Rat liver treated by PCN" and "Schizophrenia brain proteomic analysis".

2016年5月2日運用開始

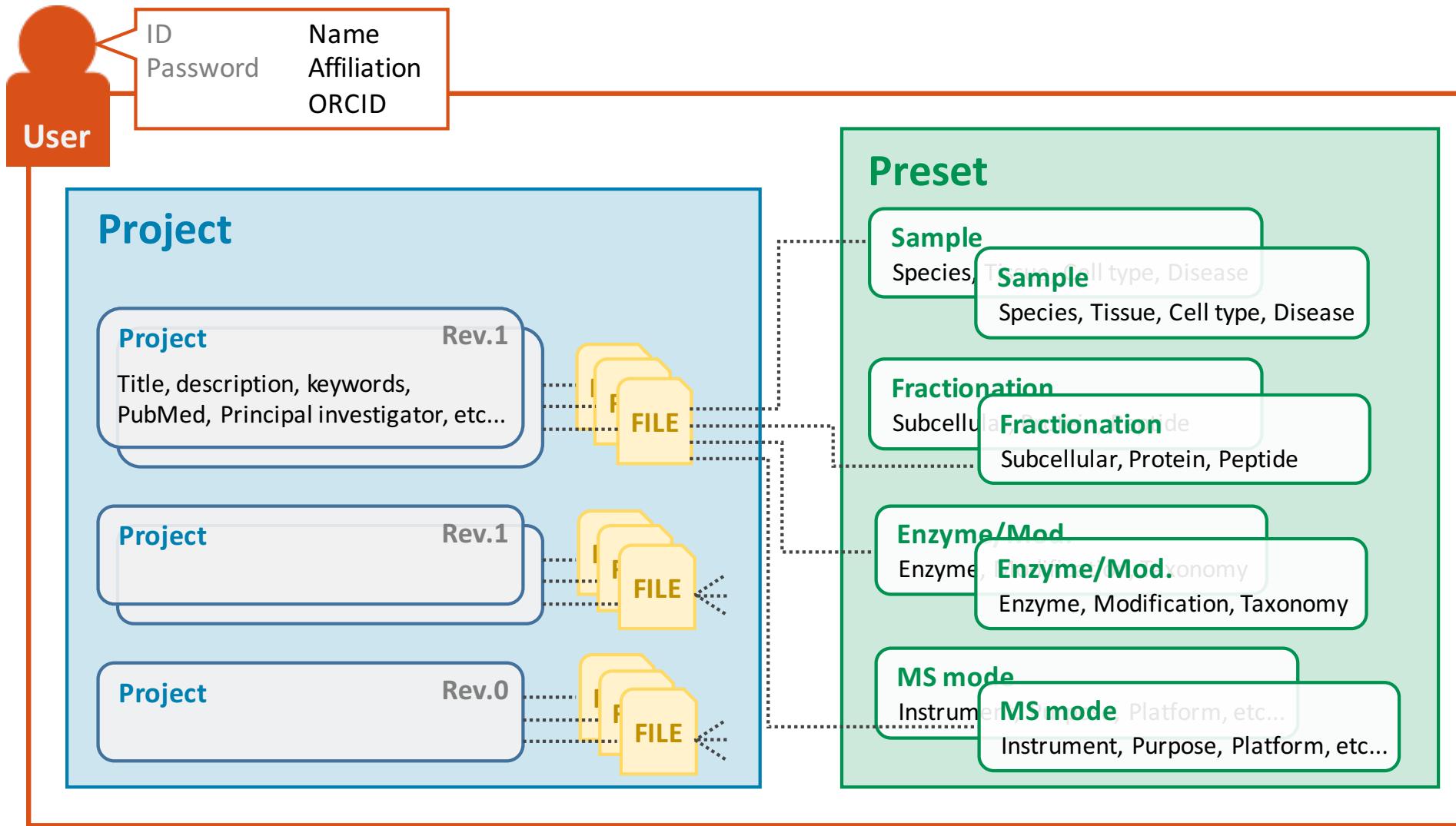
主な特徴

- オントロジーによる詳細メタ情報入力
- メタ情報とファイルとの柔軟な対応付け
- わかりやすいユーザ・インターフェース
- データ公開までの期間の設定
(embargo)
- 未公開状態でのReviewerのアクセス
- 超高速ファイルアップロード
- ウェブブラウザのみで完結

現在の登録状況

270 プロジェクト (142 オープン)
29,756 ファイル
7.0 TB
33 species

メタ情報とファイルの柔軟な対応付け



再利用可能な実験プロトコル用プリセット

New Sample profile

Title*

Species* (select)

Tissue (select)

Cell type (select)

Disease (select)

Note

New Fractionation profile

Title*

Subcellular
Fraction: Replicate:
[+ Add Subcellular](#)

Protein
Fraction: Replicate:
[+ Add Protein](#)

Peptide
Fraction: Replicate:
[+ Add Peptide](#)

Search from EBI Ontology Lookup Service 

or Enter free word 

(select)

Confirm 

New Sample profile

Title* **MOCK-FLAG-IP**

Species* **Arabidopsis thaliana (Mouse-ear cress)** 

Tissue (select) 

Search from EBI Ontology Lookup Service 

or Enter free word 

(select)

Close 

プロファイルのファイルへの適応

Select profiles for files

Sample Fractionation

Add files

Drop files here or click to upload.

Apply profile

Remove selected **Reset** **File name**

Select File Name Sample Fractionation Enzyme/Mod. MS mode Location

(No file found for this project)

The screenshot displays a software interface for managing biological samples and their associated files. At the top, there are tabs for 'Sample' and 'Fractionation'. Below these, a 'Select profiles for files' section is shown. Two sample entries are listed:

S0000000002

Title: MOCK-FLAG-IP
Species: Rattus norvegicus (Rat)
Tissue: **SELECTED**
Cell type: embryonic stem cell
Disease: breast cancer
Note:

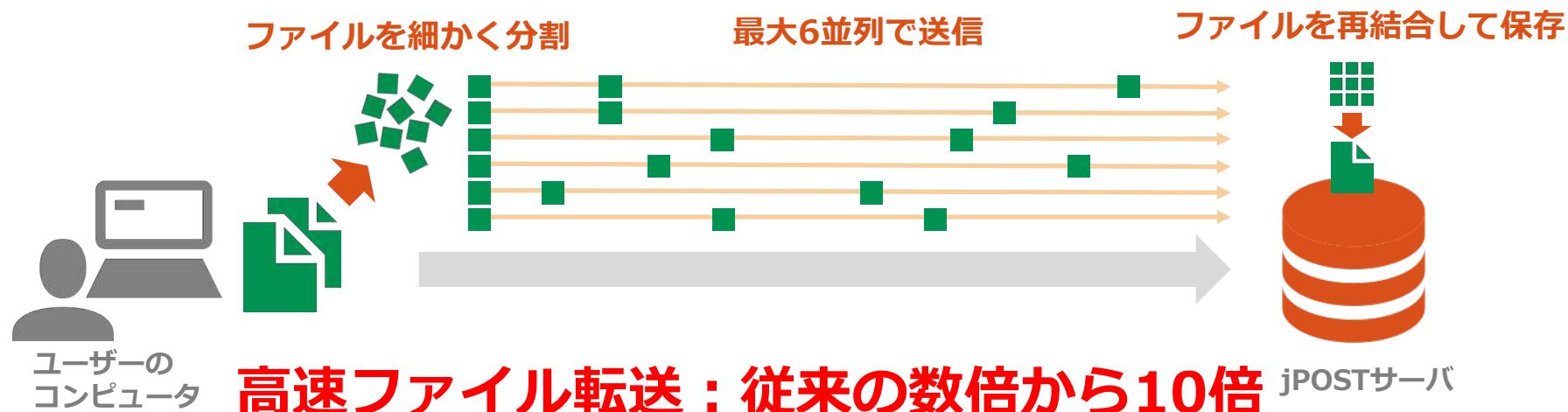
A red arrow points from the 'Sample' tab to the second sample entry. In the 'Add files' section, a dashed blue box contains a list of 16 files, with the number '16' highlighted in red. A large blue arrow points down to the 'Drop files here or click to upload.' area. Below this, a 'Profile' section shows a list of 10 profiles, with the number '10' highlighted in red. At the bottom, there is a table with columns: Select, File Name, Sample, Fractionation, Enzyme/Mod., MS mode, and Location. A message '(No file found for this project)' is displayed in the center of the table area.

ファイル転送の高速化

■ 通常のファイルアップロード



■ jPOSTのファイルアップロード



データ登録デモ

jPOST REPOSITORY

Repository Submit Help My page Sign out

About jPOST Repository

The jPOST Repository is a new data repository of sharing MS raw/processed data. It consists of a newly-developed, high-speed file upload process, flexible file management system and easy-to-use interfaces. Users can release their "raw/processed" data via this site with a unique identifier number for the paper publication. Users also can suspend (or "embargo") their data until their paper is published. The file transfer from users' computer to the jPOST server is very fast (roughly ten times faster than usual file transfer) and uses only web browsers – it does not require installing any additional software.



100MB/25秒
4MB/s



Statistics

20 projects are registered

105 files amount total

1 species.



Date	Project	Opened project
2016.05	0	0
2016.06	1	1
2016.07	2	2
2016.08	3	3
2016.09	4	4

Data list

Search project Search Reset

1 - 6 / 6 1

jPOST ID	Project title	Description	Type	Publication	Principal investigator	Announcement date

jPOST repository デモサイト

- <https://rep-demo.jpostdb.org/>
- 正式なアクセスション番号の取得をしない以外は、ほぼ本環境と同じ
- 一定期間でデータは消去

The screenshot shows the homepage of the jPOST repository demo site. At the top, there is a navigation bar with the jPOST DEMO REPOSITORY logo, Repository, Submit, Help, Sign in, and Sign up buttons. Below the navigation bar, there is a section titled "About jPOSTrepo" with a brief description of the repository's purpose and features. There is also a reference section with a citation and a link to PubMed. On the right side of the page, there is a graphic illustration of a globe with various data storage icons (folders, database, files) floating around it.

jPOST DEMO REPOSITORY

Repository Submit ? Help [Sign in](#) [Sign up](#)

About jPOSTrepo

jPOSTrepo (Japan Proteome Standard Repository) is a new data repository of sharing MS raw/processed data. It consists of a newly-developed, high-speed file upload process, flexible file management system and easy-to-use interfaces. Users can release their "raw/processed" data via this site with a unique identifier number for the paper publication. Users also can suspend (or "embargo") their data until their paper is published. The file transfer from users' computer to our repository server is very fast (roughly ten times faster than usual file transfer) and uses only web browsers – it does not require installing any additional software.

Reference

Okuda, S. et al. jPOSTrepo: an international standard data repository for proteomes. Nucl. Acids Res. 45 (D1): D1107-D1111 (2017). doi: 10.1093/nar/gkw1080
[pubmed]

Statistics

0 project is registered. 0 is opened.