

バイオインフォマティクス・システムバイオロジーと 生命システム情報統合データベース

本 多 渉¹, 田 辺 麻 央², 矢 野 亜 津 子², 金 久 實^{1,2}

生命をシステムとして捉え、コンピュータの中に再現することを目的としたデータベース、それがKEGGである。KEGGは大きく分けて3種類のデータからなっている。ゲノム情報、ケミカル情報、そしてシステム情報であり、それぞれは多数のデータベースから構成されている。本稿ではシステム情報を主に説明する。まずゲノム情報について簡単に解説し、そしてケミカル情報、中でも最近大幅に充実化が進むKEGG DRUGについて詳しく説明する。続いてシステム情報の中でも最もユニークなKEGG PATHWAYを解説し、その後に比較的新しいコンテンツであるKEGG MODULE, KEGG AtlasとKEGG DISEASEについて説明する。

1. はじめに

現在の生物学の研究ではゲノムの配列のみならず、ハイスループットな実験手法の開発と発展に伴い、トランスクリプトーム、プロテオーム、インタラクトーム等、大規模な解析によって生み出されたデータを統合的に取り扱い、生命をシステムとして捉える研究が重要視されている。KEGG (Kyoto encyclopedia of genes and genomes : <http://www.genome.jp/kegg>) は表1に示されている通り、2008年6月現在で19のデータベースから構成されており、人手によって文献等の情報に基づいて発展・更新され続けている¹⁾。KEGGデータベースは遺伝子配列、タンパク質配

列等の「ゲノム情報」、生体内外の化学物質、生体内反応や薬といった「ケミカル情報」、そして分子間相互作用・反応ネットワークや分子・細胞・個体に関する様々なオブジェクトの階層と関係を記述した「システム情報」から構成されている。このうちシステム情報は全て文献に基づき、一つ一つ人手によって入力されたものである。本稿では表1に示されているゲノム情報、ケミカル情報について簡単に説明した後、システム情報について具体的に説明する。実際にweb上の画面を見て、いろいろなところをクリックしながら読まれることをおすすめしたい。

2. ゲノム情報

KEGGのゲノム情報には核酸やタンパク質の配列情報あるいは配列間の類似性情報等が蓄積されている。表1にあるように、ゲノム情報にはKEGG ORTHOLOGY, KEGG GENOME, KEGG GENES等複数のデータベースを含んでいるが、本稿ではGENESについて簡単に説明する。ゲノム全塩基配列が決定された生物種について、その生物種の持つ遺伝子のカタログがGENESであり、現在約332万の遺伝子情報が登録されている。図1にGENESのエントリー例を示した。図1はヒトのアルコールデヒドロゲナーゼ遺伝子ADH1Aのエントリーである。ここに示されているように、遺伝子の塩基配列、アミノ酸配列はもちろん配列中に見いだされるモチーフや、タンパク質の立体

¹ 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 生命知識システム領域 (〒611-0011 京都府宇治市五ヶ庄)

² 東京大学医科学研究所ヒトゲノム解析センター ゲノムデータベース分野 (〒108-8639 東京都港区白金台4-6-1)

Bioinformatics, systems biology and KEGG

Wataru Honda¹, Mao Tanabe², Atsuko Yano² and Minoru Kanehisa^{1,2} (¹Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan; ²Laboratory of Genome Database, Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan)

表1 KEGG データベース

カテゴリ	データベース	内 容
システム情報	KEGG PATHWAY	代謝や細胞プロセスといった分子間相互作用・反応ネットワークダイアグラム
	KEGG BRITE	生命システムの様々な側面に関する機能分類とオントロジー
	KEGG MODULE	パスウェイ上のより小さな機能単位
	KEGG DISEASE	がん, 神経変成疾患, 感染症などの病気
ゲノム情報	KEGG ORTHOLOGY	PATHWAY と BRITE に基づくオーソロググループ
	KEGG GENOME	全塩基配列が決定された生物種のゲノム
	KEGG GENES	全塩基配列が決定された生物種の遺伝子カタログ
	KEGG DGENES	ドラフトゲノムの遺伝子カタログ
	KEGG EGENES	EST コンティグ
	KEGG VGENOME	ウイルスゲノム
	KEGG VGENES	ウイルス遺伝子
	KEGG OGENES	オルガネラ遺伝子
	KEGG SSDB	あらゆる遺伝子間の配列類似性
	ケミカル情報	KEGG COMPOUND
KEGG DRUG		米国および日本で認可された薬
KEGG GLYCAN		糖鎖
KEGG ENZYME		酵素と酵素反応
KEGG REACTION		生体内化学反応
KEGG RPAIR		基質と生成物の間の構造変換パターン

構造, 染色体上の位置, 後に述べるパスウェイへのリンクがあり, またこの遺伝子情報を他のデータベースで参照することもできる. 詳細は解説書を参照していただきたい²⁾.

3. ケミカル情報

ケミカル情報は代謝化合物や糖鎖, 及びそれらが関係する酵素・酵素反応等を含んでおり, 表1に示されているように六つのデータベースからなる. これら六つのデータベースはまとめて KEGG LIGAND と呼ばれており, 以下, それぞれについて解説する. COMPOUND は約1万5,000の代謝化合物・環境物質等の情報を格納しており, 後に詳しく述べる. DRUG も後に詳しく述べるが, 米国および日本で認可された薬を収録し, 構造あるいはターゲットに基づいた分類などがあり, 現在約7,000の薬が登録されている. GLYCAN は約1万の糖鎖の構造に関する情報や, 構造類似糖鎖を検索したり, マイクロアレイの発現データから出現が期待される糖鎖の構造を予測するツールも実装されている. ENZYME には約5,000の酵素と酵素反応に関する情報が登録されており, REACTION には約7,600の生体内化学反応が登録されている. こういった生体内反

応において, 基質と生成物の構造が原子レベルでどのように変換されたのかということ, 計算機を用いて網羅的に解析し, 人手によって修正したものを蓄積したのが RPAIR データベースであり現在約9,600パターン登録されている. これらのケミカル情報に関連した様々な研究も行われている³⁻⁷⁾. 図2に COMPOUND のエントリー例としてオキサロ酢酸を示した. この図から分かるように, 化合物の化学式, 構造はもちろん分子量, その分子が関わる反応やパスウェイ等の情報が得られる. 他のデータベースでこの分子を参照することもできる.

本稿で紹介するデータベースは, 基本的に全て英語で書かれているものであるが, KEGG DRUG と後に述べる KEGG DISEASE は日本語版も用意されている. ここでは DRUG の日本語版について詳しく見てみる. 図3は DRUG のトップページであるが, ここからも分かるように, 薬の一般名, 商品名, 分類, ターゲット情報, 開発の歴史を参照することができる. 薬の商品名では, 日本医薬情報センター (JAPIC) が作成している医薬品添付文書が参照できるようになっている. 図4がその例である. ここでは JAPIC 医療用医薬品データベースからアスピリンを検索した例を示す. 図4の上の図が検索結果で, JAPIC の ID と KEGG DRUG のエントリー名の対応が取れて表示されているのが分かるであろう. ここで JAPIC ID をクリックすれば, 図4下のように薬品名はもちろん, 使用上の注意や禁忌, 薬物動態, 薬効薬理等の詳細を全て参照できるようになっている. さらに, 薬の分類については日本における薬効分類, 米国 USP の薬の分類 (英語), WHO の解剖治療化学分類に加えて, 生薬・漢方方剤の分類が充実している. ここでは例として, トップページの検索ボックスから「葛根湯」を検索した結果を図5に示す. 検索結果が図5の一番上の図である. ここには葛根湯を一般名または商品名に含む薬がリストされている. さらに大杉, 常磐薬品, 大峰堂等から出ている葛根湯エキスを選択してみると, 実際の商品名や成分, 薬効分類を見ることができる. この場合「カクコン」と呼ばれる成分を含むので, さらにそれをクリックすれば, どのような分子から成るかを見ることができ, 先に説明した COMPOUND データベースへのリンクもあり, 成分の分子構造を確認することもできる.

4. KEGG PATHWAY

生命は多くの分子や反応からなる巨大で複雑なシステムであるが, より小さく単純な多数のサブシステムから構成されていると考えることができる. 臓器や組織といった比較的大きなサブシステムから, 個々の細胞, オルガネラ, あるいはクエン酸回路といった, 様々なスケールのサブシステムが考えられるが, 現在 KEGG PATHWAY では分子レベルのサブシステムを提供している. すなわち, タンパ


Entry	124 CDS H.sapiens
Gene name	ADH1A, ADH1
Definition	alcohol dehydrogenase 1A (class I), alpha polypeptide (EC:1.1.1.1)
Orthology	KO: K00001 alcohol dehydrogenase [EC:1.1.1.1]
Pathway	PATH: hsa00010 Glycolysis / Gluconeogenesis PATH: hsa00071 Fatty acid metabolism PATH: hsa00120 Bile acid biosynthesis PATH: hsa00350 Tyrosine metabolism PATH: hsa00624 1- and 2-Methylnaphthalene degradation PATH: hsa00641 3-Chloroacrylic acid degradation PATH: hsa00980 Metabolism of xenobiotics by cytochrome P450 PATH: hsa00982 Drug metabolism - cytochrome P450
Class	BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster
Motif	Pfam: ADH_N ADH_zinc_N PROSITE: ADH_ZINC Motif
Other DBs	OMIM: 103700 NCBI-GI: 4501929 NCBI-GeneID: 124 HGNC: 249 HPRD: 00063 Ensembl: ENSG00000187758 UniProt: A8K3E3 P07327
LinkDB	All DBs
Structure	PDB: 1U3T 1HSO Thumbnails  Jmol
Position	4q21-q23
AA seq	375 aa AA seq DB search MSTAGKVIKCKAAVLWELKKPFSIEEVEVAPPKAHEVRIKMVAVGICGTDHVVSGTMVT PLPVILGHEAAGIVESVGEVTVTKPGDKVIPLAIPQCGKCRICKNPESNYCLKNDVSNP QGTLQDGTSRFTCRRKPIHHFLGISTFSQYTVVDENAVAKIDAASPLEKVCVLIIGCGFSTG YGSAVNVAKVTPGSTCAVFGVGLGSAIMGCKAAGAARI IAVDINKDKFAKAKELGATE CINPQDYKKPIQEVLEKEMTDGGVDFSFVEVIGRLDTMMASLLCCHEACGTSVIVGVPPDSQ

図1 KEGG GENESのエントリー例：ヒトのアルコールデヒドロゲナーゼ ADH1A

ク質や化合物といった分子が、どのように相互作用あるいは反応しているかということがネットワークダイアグラムとしてグラフィカルに表現されている。表2はその階層分類である。第一階層は六つのカテゴリーからなっているが、2, 3, 4の三つは合わせて、1の代謝パスウェイと区別する意味で特に制御パスウェイと呼んでいる。KEGG PATHWAYが提供する個々のパスウェイマップは最下層の第三階層に属しているが、第二階層のマップも全てではないが用意されている。第三階層には現在約350のマップが提供されており、以下で紹介するいくつかの例は全てこの階層のものである。このパスウェイを含めて、KEGG データベースでは、様々な生命に関するタンパク質や薬、

反応や病気等が階層的に分類されている。このようなデータベースがBRITEである。

始めに、代謝パスウェイの例を示す。図6は解糖系・糖新生系のパスウェイマップである。このパスウェイは表2における第一階層「代謝」、第二階層「糖質代謝」の中に属している。まず始めにこのパスウェイ上に描かれているオブジェクトについて、図6を部分的に取り出した図7を用いて解説する。図7では α -D-Glucoseと α -D-Glucose-6Pといった化合物が円で表わされている。このような円をノードと呼ぶ。この二つのノードが矢印で結ばれているが、このような矢印（または直線）をエッジと呼ぶ。エッジは分子の相互作用や反応を表している。この場合エッジ

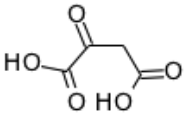
Entry	C00036	Compound
Name	Oxaloacetate; Oxalacetic acid; Oxaloacetic acid; 2-Oxobutanedioic acid; Oxosuccinic acid; keto-Oxaloacetate	
Formula	C4H4O5	
Mass	132.0059	
Structure	 <p>C00036</p> <p>Mol file KCF file DB search Jmol KegDraw</p>	
Reaction	R00217 R00338 R00339 R00340 R00341 R00342 R00343 R00344 R00345 R00346 R00347 R00348 R00350 R00351 R00352 R00353 R00354 R00355 R00357 R00359 R00360 R00361 R00362 R00363 R00373 R00400 R00431 R00477 R00695 R00726 R00930 R00931 R01144 R01257 R01447 R01713 R01731 R03735 R05053 R05758 R07164 R07165	
Pathway	PATH: map00020 Citrate cycle (TCA cycle) PATH: map00251 Glutamate metabolism PATH: map00252 Alanine and aspartate metabolism PATH: map00362 Benzoate degradation via hydroxylation PATH: map00620 Pyruvate metabolism PATH: map00630 Glyoxylate and dicarboxylate metabolism PATH: map00710 Carbon fixation PATH: map00720 Reductive carboxylate cycle (CO2 fixation)	
Enzyme	1.1.1.37 1.1.1.82 1.1.3.3 1.1.99.7 1.1.99.16 1.4.1.21 1.4.3.1 1.4.3.2 1.4.3.15 1.4.3.16 2.1.3.1 2.3.3.1 2.3.3.2 2.3.3.3 2.3.3.5 2.3.3.8 2.6.1.1 2.6.1.12 2.6.1.21 2.6.1.31 2.6.1.35 2.6.1.57 2.6.1.70 2.6.1.78 3.1.1.44 3.5.1.3 3.7.1.1 4.1.1.3 4.1.1.31 4.1.1.32 4.1.1.38 4.1.1.49 4.1.3.6 4.1.3.13 4.1.3.17 4.1.3.34 4.2.1.32 4.2.1.81 4.3.1.16 4.3.1.20 5.3.2.2 6.4.1.1	
Other DBs	CAS: 328-42-7 PubChem: 3338 ChEBI: 30744 PDB-CCD: OAA	

図2 KEGG COMPOUNDのエントリー例：オキサロ酢酸

上には EC 番号 (enzyme commission number: 酵素分類番号) の書かれたボックスが三つ書かれているが、このボックスは反応、酵素、あるいは遺伝子を表す。これらの情報から、この図の意味するところは「 α -D-Glucose は、図中の三つのいずれかの EC 番号を持つ酵素によって α -D-Glucose-6P へと変換される」ということである。このように、ノードとエッジ、およびボックスを用いて代謝化合物と酵素反応の情報を網羅的に描いたのが図6であると言える。この図6においてノードは KEGG COMPOUND データベースのエントリーにリンクされており、クリックすることでその化合物の詳細な情報を参照することができる。同様にボックスは KEGG ENZYME, KEGG REACTION と

いったケミカル情報、あるいは KEGG ORTHOLOGY, KEGG GENES といったゲノム情報へとリンクされており、反応情報はもちろん、酵素の遺伝子情報を参照することもできる。さらに、これらノードの化合物に構造的に類似した化合物をデータベースから検索したり、遺伝子の配列情報をもとに相同タンパク質を検索したりするなどの有用なツールも搭載されているが、詳細は文献を参照していただきたい²⁾。

ここで再び図7を用いてパスウェイについて若干の説明を加えておく。図7では、 α -D-Glucose を α -D-Glucose-6P に変換する酵素として、EC 番号 2.7.1.1, 2.7.1.2, 2.7.1.63 の三つを挙げているが、例えば 2.7.1.1 の酵素



KEGG DRUG

化学構造をベースにした医薬品情報のリソース

KEGG2	病気	薬	GLYCAN	COMPOUND	REACTION	Organisms
-------	----	---	--------	----------	----------	-----------

葛根湯

KEGG DRUG データベース

Entry	D00109 Drug
Name	Aspirin (JP15/USP); Acetylsalicylic acid; Easpirin (TN)
Formula	C ₉ H ₈ O ₄
Mass	180.0423
Structure	 D00109 <input type="button" value="Mol file"/> <input type="button" value="KCF file"/> <input type="button" value="DB search"/> <input type="button" value="Jmol"/> <input type="button" value="KegDraw"/>
Target	cyclooxygenase-1 (COX-1) inhibitor [NSA5742] [EC:1.14.99.1] cyclooxygenase-2 (COX-2) inhibitor [NSA5743] [EC:1.14.99.1]
Activity	Analgesic; Antipyretic; Antiinflammatory
Remark	Same as: C01405 Therapeutic category: 1143 3399 ATC code: N01AD09 B01AC06 B02BA01
Comment	Name previously used: Acetylsalicylic acid Component of Bufferin (TN), Percodan (TN), Darvon compound-65 (TN), E.A.C (TN)
Pathway	PATH: map07110 Benzoic acid family PATH: map07219 Cyclooxygenase inhibitors
Other DBs	CAS: 50-78-2 DailyMed: aspirin
LinkDB	<input type="button" value="All DBs"/>
KCF data	<input type="button" value="Show"/>

例: D00109 (日本語表示)
D00109 (英語表示)

KEGG DRUG は日本および米国で承認されたすべての薬の情報を、化学構造をベースに蓄積したデータベースです。各化学構造は D 番号で識別され、一般名、商品名、効能、ターゲットおよび関連パスウェイなどの情報が提供されています。KEGG DRUG は LIGAND リレーションナルデータベースの一部として構築されています。

- DBGET サーチ (英語)

商品名

英語表示の KEGG DRUG エントリーには代表的な商品名しか記載されていません。米国の商品名は FDA が承認した医薬品の添付文書を蓄積した DailyMed への外部リンクをたどって調べてください。日本語表示の KEGG DRUG エントリーでは、JAPIC が構築している医療用医薬品添付文書データベースに含まれる日本の医薬品の商品名と、個々の添付文書への直接リンクが提供されています。

- JAPIC 医療用医薬品データベース (日本語)
- JAPIC 一般用医薬品データベース (日本語)

薬の分類

KEGG BRITE データベースの一部として、薬の分類 (オントロジー) が提供されています。

- 日本の薬効分類 (日本語)
- 日本の薬効分類 (英語)
- 米国 USP の薬の分類 (英語)
- WHO の解剖治療化学分類 (日本語)
- WHO の解剖治療化学分類 (英語)
- 生薬・漢方方剤の分類 (日本語)
- 生薬・漢方方剤の分類 (英語)

ターゲット情報

薬とターゲットの関係は単に分子と分子の関係だけではありません。KEGG DRUG データベースでは、ターゲット情報は KEGG PATHWAY の一部として記載され、生体システムに対するゆらぎ物質としての薬のはたらきを解析できるようにになっています。ターゲット情報は BRITE デー

図3 KEGG DRUG 日本語版トップページ

はヒトには存在しているが、大腸菌では確認されていない。一方 2.7.1.63 はヒトでも大腸菌でも確認されておらず、特定のバクテリアがコードしている。しかしながらこのマップのように、生物種ごとの違いを考慮することなく、これまでに集積された生化学の知識を全て同一のマップに盛り込むことで表現したものをレファレンス・パスウェイ (Reference pathway) と呼ぶ。KEGG のマップを開いた場合、デフォルトでこのレファレンス・パスウェイが選択されているが、マップ左上にあるプルダウン・メニューから生物種を特定することで、その生物種に特異的な酵素のみを着色することができる。従ってこの場合、ヒトを選べば 2.7.1.1 と 2.7.1.2 の両方が着色されるが、大腸菌を選べば、2.7.1.2 のみが着色されたマップを得ることができる。レファレンス・パスウェイでは生物種を特定しない、つまりボックスに対応する遺伝子が特定されないため、各ボックスは KEGG ENZYME、あるいは KEGG REACTION といったケミカル情報へとリンクされていて、

全てのボックスがクリックできる。一方プルダウン・メニューから特定の生物種を選択すると、ボックスに対応する遺伝子が特定されることになるので、その生物種に特異的なボックスのみがクリックできるようになり、それらはゲノム情報へとリンクされることになる。

次に図8は第一階層「細胞プロセス」、第二階層「免疫系」に属する T 細胞シグナル伝達経路のパスウェイマップである。このマップでは今度はボックスがノードとなっている。遺伝子産物を意味するこれらボックスの中には、酵素であれば EC 番号を、それ以外であれば遺伝子名等を表示している。従ってこのボックスは KEGG ORTHOLOGY、あるいは KEGG GENES といったデータベースへとリンクされており、遺伝子の詳細な情報を得ることができる。また代謝マップの時と同様に、マップ左上にあるプルダウン・メニューから、対象となる生物種を選択することで、その生物種が持っている遺伝子にのみ色を付けることも可能である。この図ではエッジ上に「+p」や「+u」と

JAPIC 医療用 医薬品データベース

アスピリン 名称検索 薬効検索 全文検索

JAPIC ID	KEGG ID	名称	薬効
00054394	D00109	アスピリン; アスピリンアセチルサリチル酸; アスピリン製剤	3399 抗血小板剤
00054394	D00109	アスピリン; アスピリン	1143 解熱鎮痛消炎剤
00051357	D00109	アスピリン; アスピリン	1143 3399 解熱・鎮痛・消炎剤, 川崎病用剤
00050771	D00109	ニチアスピリン; アスピリンアセチルサリチル酸; アスピリン製剤	3399 抗血小板剤
00050646	D00109	アスピリン; アスピリンアセチルサリチル酸; アスピリン製剤	3399 抗血小板剤
00050583	D00109	ゼンアスピリン; アスピリン; アスピリン製剤	3399 抗血小板剤
00050115	D00109	アスピリン; アスピリン; アスピリン製剤	3399 抗血小板剤
00050033	D03960	ファモター; アスピリンジヒドロキシアルミニウムアミンオセテート炭酸マグネシウム; 制酸緩衝アスピリン製剤	3399 抗血小板剤
00049529	D00109	アスピリン	3399 抗血小板剤

JAPIC 医療用 医薬品データベース

商品名: アスピリン
 一般名: アスピリンアセチルサリチル酸
 欧文一般名: Aspirin
 製剤名: アスピリン製剤
 薬効分類名: 抗血小板剤
 薬効分類番号: 3399
 KEGG DRUG: D00109 -類似商品一覧
 JAPIC ID: 00054394 -PDF文書

販売和名: アスピリン製剤100mg「マイラン」
 欧文商標名: Aspirin Tab.100mg「Mylan」
 製造会社: マイラン製薬
 承認番号: 21600AMZ00625
 YJコード: 3399007H1072
 日本標準商品分類番号: 873399
 薬価: 6.1円/錠
 後発品フラグ: 後発品
 JAPIC ID: 00054394-001 -個別表示

本文情報
 すべて開く
 注意: 次の患者には投与しないこと
 本剤の成分又はサリチル酸系製剤に対し過敏症の既往歴のある患者
 消化性潰瘍のある患者 (プロスタグランジン合成抑制作用により、胃の血流量が減少し、消化性潰瘍を悪化させることがある。) (ただし、「1.慎重投与」の項参照)
 出血傾向のある患者 (血小板機能異常がおこることがあるため、出血傾向を助長するおそれがある。)
 アスピリン喘息 (非ステロイド性消炎鎮痛剤等による喘息発作の誘発) 又はその既往歴のある患者 (重篤なアスピリン喘息発作を誘発させることがある。)
 出産予定日12週以内の妊婦 (「6.妊婦、産婦、授乳婦等への投与」の項参照)
 低出生体重児、新生児又は乳児 (「7.小児等への投与」の項参照)

使用上の注意:
 薬物動態:
 薬効薬理:
 低用量アスピリンは、選択的に血小板におけるプロスタグランジン類の生合成を阻害することにより、血小板凝集抑制作用を示す(4)。

有効成分に関する理化学的知見:
 取扱い上の注意:
 包装:
 主要文献

図4 JAPIC 医療用 医薬品データベース

いった表記があり、それぞれリン酸化、ユビキチン化を意味するが、詳細は右上にある Help ボタンをクリックすることで参照することができる。図中にあるように、角の丸い長方形が破線で書かれたエッジで結ばれており、他のパスウェイマップへのリンクを示しているの、様々なマップ間のつながりも理解することができるであろう。

5. KEGG MODULE

次に KEGG MODULE について説明する。パスウェイマップは解糖系、TCA サイクル等、機能単位として明瞭な分子と反応のネットワークであるが、そのマップ内においてさらに小さな機能単位を考えることができるであろう。図9に例を示した。図9左側は、先ほど示した解糖系・糖新生系のマップであるが、点線の枠で囲まれた内部の遺伝子群は協同的に働くことが知られている。こういった遺伝子群をモジュールとして考え、様々なパスウェイから人手で抽出し、蓄積したものが KEGG MODULE データベースである。図9右側にあるように、KEGG MODULE のエントリーは、そのモジュール内の遺伝子、反応、化合

物等の情報を含んでいる。KEGG MODULE は現在約 600 あり、図10に示すように KEGG BRITE に階層的に分類されている。それぞれのエントリーをクリックすることで、モジュールに属する遺伝子群がマップ上で色付けされて見ることができる。

6. KEGG Atlas

KEGG Atlas はこれまでに蓄積されたパスウェイマップを、より大きな視点からグローバルに見ることを目的として開発されたマップビューワーである⁸⁾。KEGG Atlas では、現在二つのマップが提供されており、それぞれ代謝マップ (metabolism map) とがんのマップ (Cancer map) である。がんのマップは伝統的なパスウェイマップの形式を踏襲しているが、代謝マップでは新しいインターフェースも用意されている。また、現在はまだリンクは張られていないが、既に web 上に書かれているように Cell map, Body map といった他の Atlas も追加されていく予定である。

代謝マップでは、これまでに約 120 の代謝パスウェイのマップが統合されて、1 枚の巨大なマップとして表現され

Drug Data Search Result Top

Number of entries in a page

Items: 1 - 4 of 4

No	エントリ	一般名	商品名
1	D06698	葛根湯エキス (JP15) 葛根湯 カクコントウエキス カクコントウ	葛根湯エキス (大杉, 常盤薬品, 大峰堂) 葛根湯エキス顆粒 (ジェービーエス, ツムラ, 建林松鶴堂, 阪本, 劑盛堂, 東亜薬品, 太虎精堂, 本草, 帝國漢方, 松浦) 葛根湯エキス粗粒 (クラシエ, 三和生薬, 東洋薬行, 小太郎漢方, 康和薬通)

DRUG: D06698

エントリ	Formula Drug
D06698	葛根湯エキス (JP15); 葛根湯; カクコントウエキス; カクコントウ; Kakkonto extract (JP15)
商品名	オースギ (常盤薬品工業), 葛根湯 (クラシエ製薬), 葛根湯 (ジェービーエス製薬), 葛根湯 (ツムラ), 葛根湯 (建林松鶴堂), 葛根湯 (康和薬通), 葛根湯 (阪本漢法製薬), 葛根湯 (三和生薬), 葛根湯 (小太郎漢方製薬), 葛根湯 (松浦薬業), 葛根湯 (太虎精堂製薬), 葛根湯 (大杉製薬), 葛根湯 (大峰堂薬品工業), 葛根湯 (帝國漢方製薬), 葛根湯 (東亜薬品), 葛根湯 (東亜薬品), 葛根湯 (東亜薬品), 葛根湯 (東亜薬品), 葛根湯 (東亜薬品), 葛根湯 (東亜薬品), 葛根湯 (東洋薬行), 葛根湯 (中華製薬)
成分	カクコン [DR:D06693], マオウ [DR:D06791], タイソウ [DR:D06758], シャクヤク [DR:D06739 D06850], カンソウ [DR:D04365 D06828], ショウキョウ [DR:D06744 D06852], ゲイヒ [DR:D06712 D06841] または, ケイシ [DR:D06898]
効能	漢方 感冒, 鼻かぜなど
注釈 1	薬効 200

COMPOUND: D06693

エントリ	Crude Drug
D06693	カクコン (JP15); 葛根; Pueraria root (JP15)
成分	デンブun [DR:D06507] [CPD:C00369], ダイゼイン [CPD:C10216], ダイジン [CPD:C10208], プエラリン [CPD:C10241], プエラリンキシロシド, パルミチン酸メチル, スペリン酸ジメチル
ソース	Pueraria lobata [TAX:3893]

COMPOUND: C10216

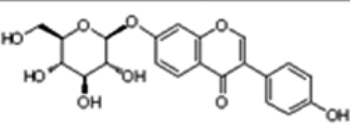
エントリ	Compound
C10216	ダイジン
名称	ダイジン
組成式	C21H20O9
質量	416.1107
構造式	 <p>C10216</p> <p>Mol file KCF file DB search Jmol KegDraw</p>
反応	R07720 R07721
マップ	PATH: map00943 Isoflavonoid biosynthesis
酵素	2.3.1.115 2.4.1.170
リンク	CAS: 552-66-9 PubChem: 12402 ChEBI: 42202 PDB-CCD: DZN

図5 ドラッグデータ例: 葛根湯

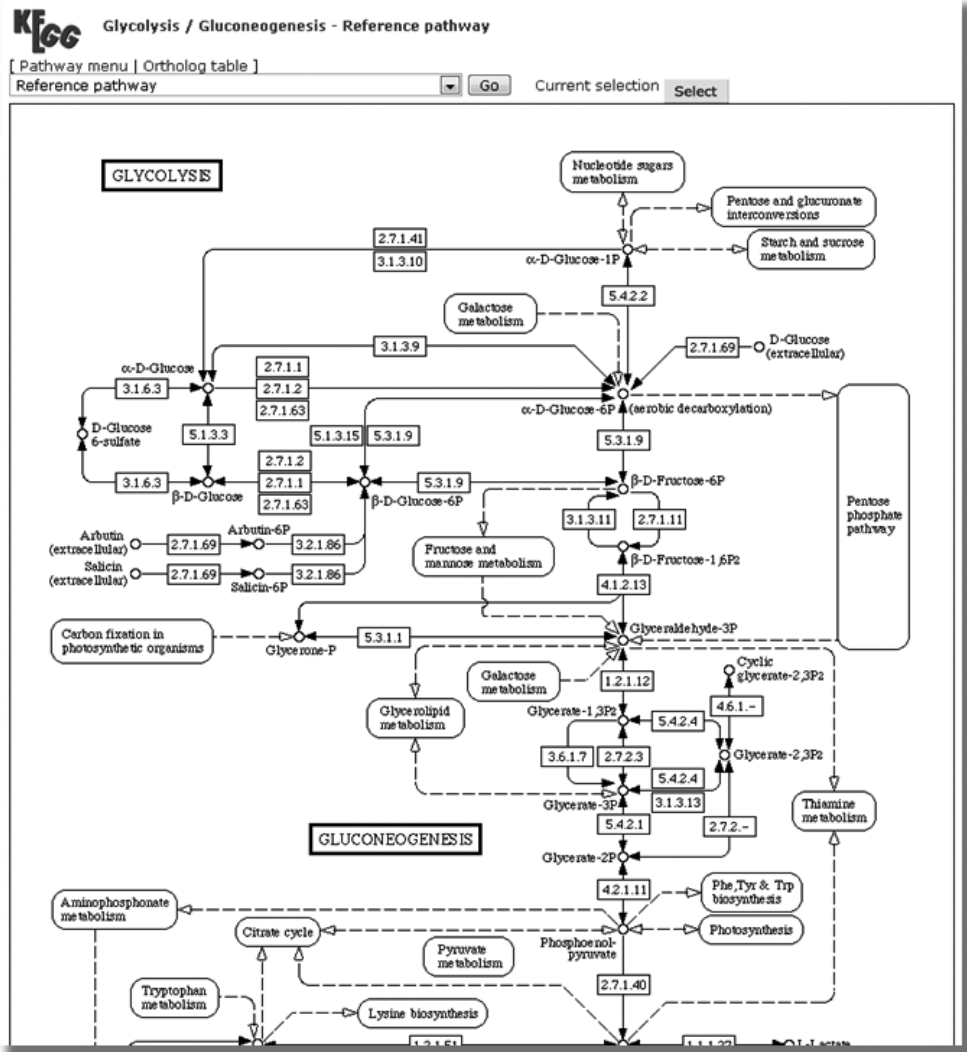


図6 解糖系・糖新生系のパスウェイマップ

ている (図 11)。マップ上でマウスを動かすと、カーソルに合ったノードあるいはエッジがハイライトされるので、どのオブジェクトをポイントしているかが分かるであろう。このマップではノードは代謝パスウェイと同じく化合物を表現しているが、エッジは一つ以上の化合物及び酵素の集合によってなされる一連の反応で、枝分かれのないものを表している。これをネットワークエレメントと呼ぶ。従ってこのマップでは、個々の代謝パスウェイ上で見られた全ての化合物が描かれているわけではなく、省略されているものも存在する。ネットワークエレメントの一つをクリックすることで、どのような酵素を含んでいるか、あるいはどのような途中の化合物が省略されているのか等ということを確認することができる。マップ左上には拡大・縮小ボタンがついており、またマウスのホイールを回転させることでも拡大・縮小が自由にでき、マウスをドラッグすることによってマップを画面上で移動させることもでき

る。

また図 11 に示されている通り、ページ左上にあるボックスに KEGG の生物種コードを入力することで、その生物種に特有なパスウェイにのみ色がついた図を得ることができる。KEGG の生物種コードが分からない場合は organism ボタンをクリックすることで検索、選択することも可能である。右上にある検索ボックスを使えば、C00001 といった化合物のエントリー名あるいは ENO1 といった個別の遺伝子名を入力してマップ上で確認することもできる。また手持ちの化合物リストや遺伝子のリストをもとにこのマップ上に写像し、代謝経路全体としてどのように分布しているかを視覚的に確認することもできる (図 12)。このマップで生物種を特定すると、その生物種に特異的なノード及びエッジが緑色に着色され、その生物種では確認されていないノード及びエッジが薄いグレーにした図が得られ、代謝マップ上でどの部分がその生物種に固有なのか

表2 KEGG PATHWAYの階層分類

第一階層	第二階層
1. 代謝	<ul style="list-style-type: none"> ➢糖質代謝 ➢エネルギー代謝 ➢脂質代謝 ➢ヌクレオチド代謝 ➢アミノ酸代謝 ➢その他のアミノ酸代謝
2. 遺伝情報処理	<ul style="list-style-type: none"> ➢転写 ➢翻訳
3. 環境情報処理	<ul style="list-style-type: none"> ➢膜輸送 ➢シグナル伝達
4. 細胞プロセス	<ul style="list-style-type: none"> ➢細胞運動 ➢細胞増殖と細胞死 ➢細胞間コミュニケーション ➢内分泌系 ➢免疫系
5. ヒトの病気	<ul style="list-style-type: none"> ➢がん ➢免疫疾患 ➢神経変性疾患
6. 薬の開発	<ul style="list-style-type: none"> ➢糖鎖の生合成と代謝 ➢ポリケチド・非リボソームペプチド生合成 ➢補酵素・ビタミン代謝 ➢二次代謝物質の生合成 ➢非生物物質の分解と代謝 ➢局在化と分解 ➢複製と修復 ➢シグナリング分子とその相互作用 ➢神経系 ➢感覚系 ➢発生 ➢行動 ➢代謝疾患 ➢感染症 ➢その他の薬開発の歴史 ➢ターゲットによる構造分類 ➢骨格による構造分類

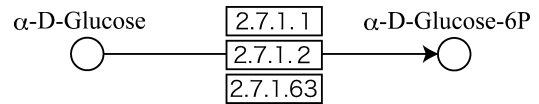


図7 バスウェイマップのオブジェクト

がわかる。さらに、上述の写像した化合物を示すノード、及び酵素を含んでいるエッジは赤色に着色される。図12からも分かるように何色に着色するかはノード・エッジごとにユーザーが選択することができるようになっている。他にも、メタゲノムのデータ、あるいはマイクロアレイデータの解析結果等を用いることによって様々な生物種の違いを、代謝経路全体を比較することで理解する等の応用が考えられる。現在代謝マップのインターフェースは大幅に改訂しているの、近い将来にはさらに充実し、使い易いものになる予定である。

次に、がんのマップについて説明する。このマップは、後に詳しく見るが、KEGG DISEASEに登録されている14のがんのマップを統合し、1枚の図として表現したものである(図13)。ページ左上のプルダウン・メニューでは、

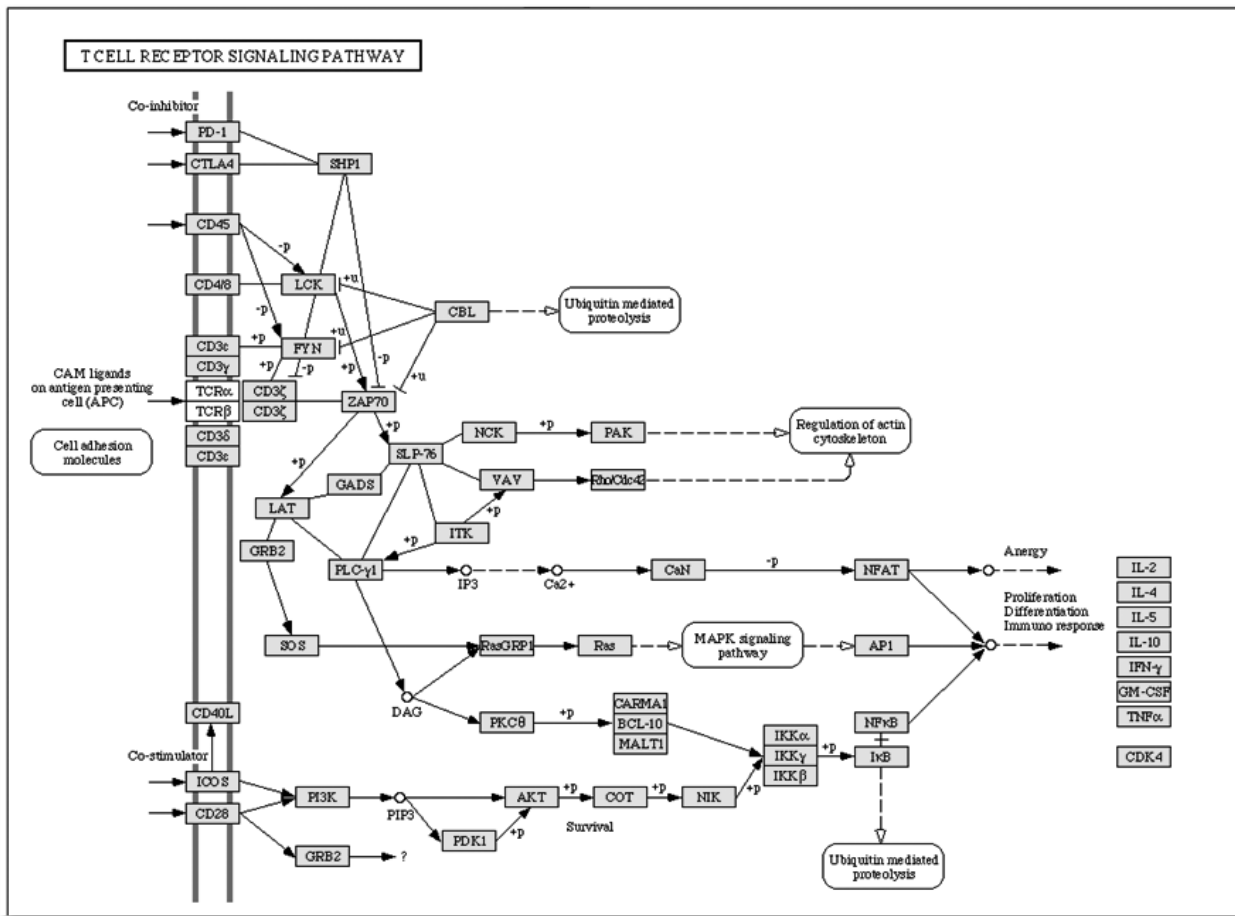


図8 細胞プロセスマップ例：T細胞シグナル伝達経路

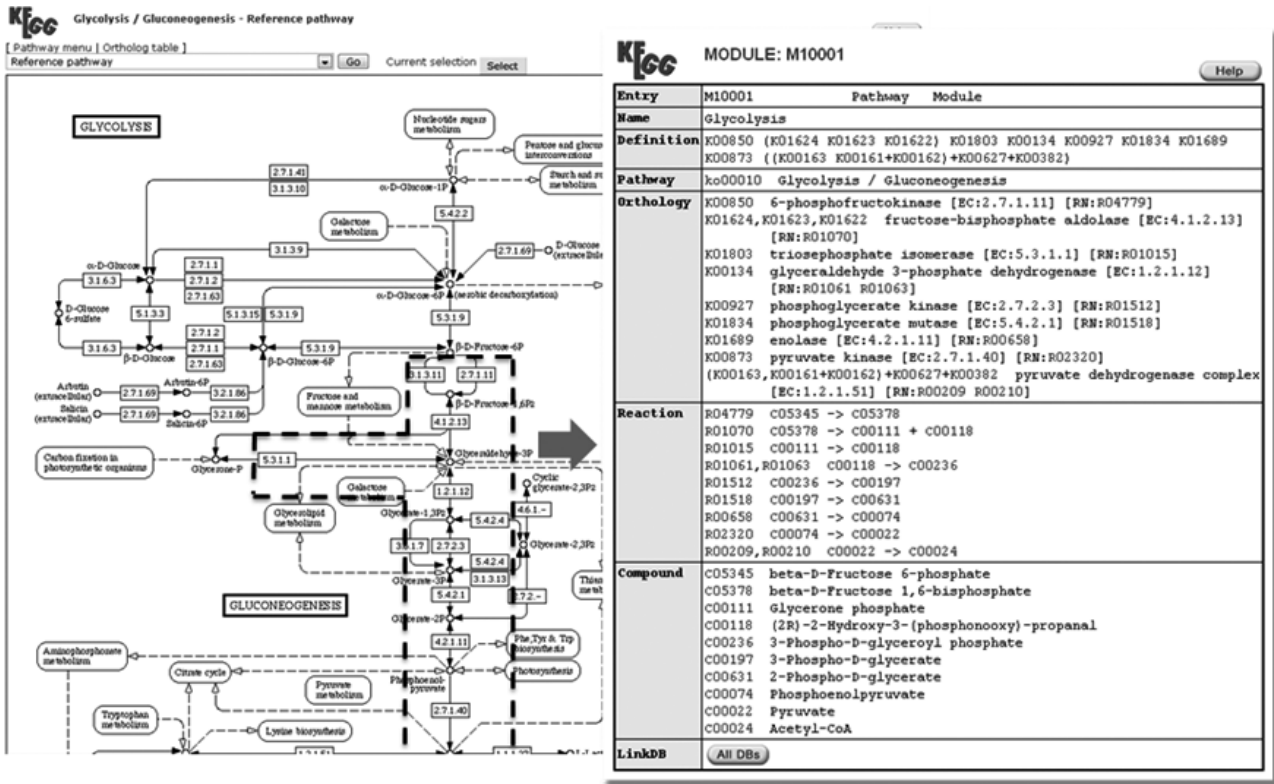


図9 モジュール例：解糖系・糖新生系のマップ上のモジュール

今度は生物種の代わりに、がんの種類を選択することができるようになってきている。例えばリスト内にある慢性骨髄性白血病 (Chronic myeloid leukemia) を選択すると、このタイプのがんに特に関連していると思われる遺伝子のボックスが紫に、このタイプのがんで変異を起こしていることが確認されている遺伝子が赤字に着色された図14が得られる。このように、複数のがんの違いや共通性を個別に視覚的に確認することができる。ページ下には各がんのマップへのリンクもある。いくつかのがんを選択してみると、がんに関連しているシグナル伝達の分子群はモジュール構造を取っていることも見て取れる。

7. KEGG DISEASE

KEGG DISEASEは大きく二つのコンテンツからなり、病気の分類と病気のパスウェイ (14のがん, 6の神経変性疾患, 3の代謝疾患, 3の感染症)が現在提供されている。

病気の分類は「病原体と感染症 (Pathogens and infectious diseases)」と「ヒトの病気 (Human diseases)」からなる。「病原体と感染症」では、図15に左側に示されている通り、現在知られているヒト、動物、及び植物の病原体がKEGG BRITEに階層的にリストアップされている。それぞれの病原体エントリは、各生物種の情報へとリンクされている。ここではコレラをクリックした例が図15右側

である。ここに示されているようにその病原体由来の病気名や病原体ゲノム情報へのリンクを参照することができる。さらに感染時のメカニズムが明らかになっているものについてはパスウェイへのリンクが示されている。また「ヒトの病気」では、よく研究されている病気について、関連している遺伝子、診断用マーカー、薬などの情報を参照することができる (図16)。

次に病気のパスウェイの例を示す。図17は大腸がん (Colorectal cancer) のマップである。このマップは先に見せた細胞プロセスのマップとほぼ同じ作りになっているが、より多くの情報を含んでいる。まず、マップの右上にはこのタイプのがんに関連していることが分かっている遺伝子のうち、変異を起こすことが実験で確認されているもののリストががん遺伝子、がん抑制遺伝子、DNA修復遺伝子のようにカテゴリーに分けて書かれている。またこれらの変異遺伝子はマップ上で認識し易いように赤い文字で表示されている。さらにこのマップは二つの軸に基づいて表現されている。横軸は、シグナル伝達方向であり、この場合左から右へとシグナルが伝達されていくことが分かる。縦軸は時間軸であり、マップ左側に記述されているように上から下に行くほど、がんが進行していることを意味する。従って、この図から初期の大腸がんではAPC、βカテンンといった遺伝子に変異が起きているが、がんの進行

KEGG Pathway Modules	
[1st Level 2nd Level 3rd Level 4th Level Text Search]	
▼ Metabolism	
▼ Central metabolism	
M10001	Glycolysis [PATH:ko00010]
M10002	Glycolysis, core module [PATH:ko00010]
M10003	Gluconeogenesis [PATH:ko00010]
M10004	Pentose phosphate pathway [PATH:ko00030]
M10005	Pentose phosphate pathway and PRPP biosynthesis [PATH:ko00030]
M10006	Pentose phosphate pathway, oxidative phase [PATH:ko00030]
M10007	Pentose phosphate pathway, non-oxidative phase [PATH:ko00030]
M10008	Entner-Doudoroff pathway [PATH:ko00030]
M10009	Citrate cycle [PATH:ko00020]
M10010	Citrate cycle, module 1 [PATH:ko00020]
M10011	Citrate cycle, module 2 [PATH:ko00020]
M10012	Glyoxylate cycle [PATH:ko00630]
M10013	Glyoxylate biosynthesis, glycolate => glyoxylate [PATH:ko00630]
M10014	Glyoxylate biosynthesis, glycine => glyoxylate [PATH:ko00260]
M10015	Malonate semialdehyde pathway [PATH:ko00640]
M10016	Glucuronate pathway (uronate pathway) [PATH:ko00040]
▼ Amino acid metabolism	
M10017	Glutamate biosynthesis, oxoglutarate => glutamate (glutamate [PATH:ko00251]
M10018	Glutamine biosynthesis, glutamate => glutamine [PATH:ko00251]
M10019	Glutamate biosynthesis, oxoglutarate => glutamate (glutamate [PATH:ko00251]
M10020	Proline biosynthesis, glutamate => proline [PATH:ko00220]
M10021	Aspartate biosynthesis, oxaloacetate => aspartate [PATH:ko00252]
M10022	Asparagine biosynthesis, aspartate => asparagine [PATH:ko00252]
M10023	Lysine biosynthesis, aspartate => lysine [PATH:ko00300]
M10024	Methionine biosynthesis, aspartate => homoserine => methionine [PATH:ko00260]
M10025	Threonine biosynthesis, aspartate => homoserine => threonine [PATH:ko00260]
M10026	Alanine biosynthesis, pyruvate => alanine [PATH:ko00252]
M10027	Valine biosynthesis, pyruvate => valine [PATH:ko00290]
M10028	Leucine biosynthesis, pyruvate => leucine [PATH:ko00290]
M10029	Isoleucine biosynthesis, pyruvate => isoleucine [PATH:ko00290]
M10030	Serine biosynthesis, glycerate-3P => serine [PATH:ko00260]
M10031	Glycine biosynthesis, serine => glycine [PATH:ko00260]
M10032	Cysteine biosynthesis, serine => cysteine [PATH:ko00272]
M10033	Chorismate biosynthesis, phosphoenolpyruvate + erythrose-4P => [PATH:ko00400]
M10034	Tryptophan biosynthesis, chorismate => tryptophan [PATH:ko00400]
M10035	Phenylalanine biosynthesis, chorismate => phenylalanine [PATH:ko00400]
M10036	Tyrosine biosynthesis, chorismate => tyrosine [PATH:ko00400]
M10037	Histidine biosynthesis, PRPP => histidine [PATH:ko00340]
M10038	Glutamine degradation, glutamine => glutamate + NH3 [PATH:ko00251]
M10039	Selenocysteinyl-tRNA biosynthesis [PATH:ko00450]
M10040	GABA (gamma-Aminobutyrate) shunt [PATH:ko00251]
M10041	Ornithine biosynthesis, glutamate => ornithine [PATH:ko00220]
M10042	Urea cycle [PATH:ko00220]
M10043	Arginine degradation, arginine => citrulline [PATH:ko00330]
M10044	Aspartate degradation, aspartate => fumarate [PATH:ko00252]

図10 KEGG MODULEの階層分類

に伴って Bax, p53 といった遺伝子に変異が起きていくということが分かる。また横軸のシグナル伝達の終点として、がんの原因である細胞の増殖、あるいはアポトーシスの阻害といった説明が書かれていることも特徴である。

次に、病気のパスウェイの別の例として病原体の感染についてのマップを説明する。図18, 19はコレラの感染についてのマップである。図18に示されている通り、ヒトのタンパク質のみでなく、病原体のタンパク質も同時に描かれている。この図の場合、左上のプルダウン・メニューからヒトが選択されている場合にはヒトのタンパク質に色がついているが、病原体を選択することで病原体のタンパク質を着色することもできる。さらにこのマップでは、生物種を選ぶプルダウン・メニューの隣に感染ステージを選

択するメニューも用意されている。ここで Colonization を選択したものが図19である。この図は、感染時のコレラ側のパスウェイである。この図では先に説明したがんのパスウェイのように、二つの軸に基づいている。横軸はシグナル伝達の方向であるが、縦軸はマップ右側に示されているように、コレラのライフサイクルのステージを表現している。すなわち、Stationary phase では MshA 等の遺伝子が発現しているが、Exponential phase では AcfA 等の遺伝子を発現しているということが分かる。こういった病気のパスウェイを用いた研究も行われている⁸⁾。

8. 最後 に

実験技術の向上によって、これからもさらに多くのゲノ

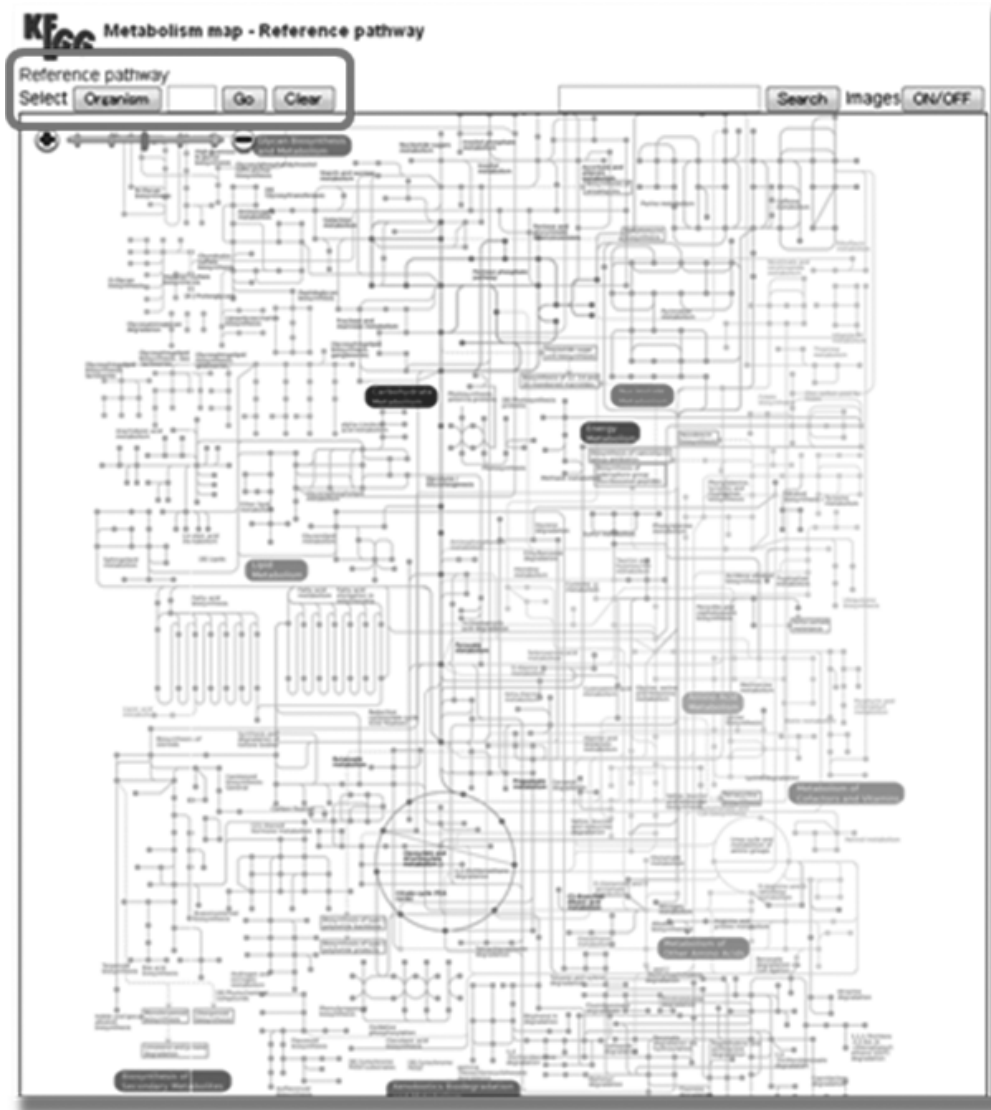


図 11 KEGG Atlas 代謝マップ

ム情報，ケミカル情報等が蓄積されていくことになるが，それらをいかに統合してシステムとしての生命の全体像を明らかにするかということは興味深い課題であろう。本稿で解説した KEGG のシステム情報のように，あらゆるデータを基盤に生命システムを再現していくデータベースが重要になることは間違いない。

文 献

- 1) Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., & Yamanishi, Y. (2008) *Nucleic Acids Res.*, **36**, D480–D484.
- 2) 金久 實 (編) (2002) ゲノムネットのデータベース利用法 [第 3 版] 共立出版, 東京.
- 3) Hashimoto, K., Goto, S., Kawano, S., Aoki-Kinoshita, K.F., Ueda, N., Hamajima, M., Kawasaki, T., & Kanehisa, M. (2006) *Glycobiology*, **16**, 63R–70R.
- 4) Kawano, S., Hashimoto, K., Miyama, T., Goto, S., & Kanehisa, M. (2005) *Bioinformatics*, **21**, 3976–3982.
- 5) Minowa, Y., Araki, M., & Kanehisa, M. (2007) *J. Mol. Biol.*, **368**, 1500–1517.
- 6) Kotera, M., Okuno, Y., Hattori, M., Goto, S., & Kanehisa, M. (2004) *J. Am. Chem. Soc.*, **126**, 16487–16498.
- 7) Oh, M., Yamada, T., Hattori, M., Goto, S., & Kanehisa, M. (2007) *J. Chem. Inf. Model.*, **47**, 1702–1712.
- 8) Okuda, S., Yamada, T., Hamajima, M., Itoh, M., Katayama, T., Bork, P., Goto, S., & Kanehisa, M. (2008) *Nucleic Acids Res.*, **36**, W423–W426.
- 9) Limviphuvadh, V., Tanaka, S., Goto, S., Ueda, K., & Kanehisa, M. (2007) *Bioinformatics*, **23**, 2129–2138.

with an Ajax-based viewer and a cancer map with the traditional KEGG map viewer.

Metabolism map
Cancer map *New!*

Other global maps are being developed or planned including:

- Cell map
- Body map
- Brain map

The new **KEGG metabolism map** is created as an SVG file by manually combining about 120 existing metabolic pathway maps. Each node (circle) is a chemical compound identified by the C number. Each line (curved or straight) connecting two nodes is manually defined as a segment lacking branches in the existing maps, named NetElement, and identified by the N number. Each NetElement corresponds to one to several KO's (such as this) in the reference pathway view, or one to several genes (such as this) in an organism-specific view.

KEGG Atlas Mapping

The new KEGG metabolism map allows the user to view and compare the entire metabolism, such as by mapping genomic, transcriptomic, metagenomic, or metabolomic data. The following interface allows mapping of genes/compounds as colored segments/nodes in the KEGG metabolism map.

Select

Reference pathway

Organism

Enter objects one per line, optionally followed by color:

(Example)

Reference pathway:
 K01689 #ff99cc
 K00873 black
 C00024 yellow
 C00189

H. sapiens (hsa):
 ENO1 red
 PKLR #3300FF

Alternatively, enter the file name containing the data:

NetElements: Partial match Complete match

Default color:

Last updated: April 1, 2008

[Feedback](#)
[KEGG](#)
[GenomeNet](#)
[Kanehisa Laboratories](#)

図 12 KEGG Atlas 代謝マップにデータをマッピング

[Pathway menu]

Cancer map Zoom up

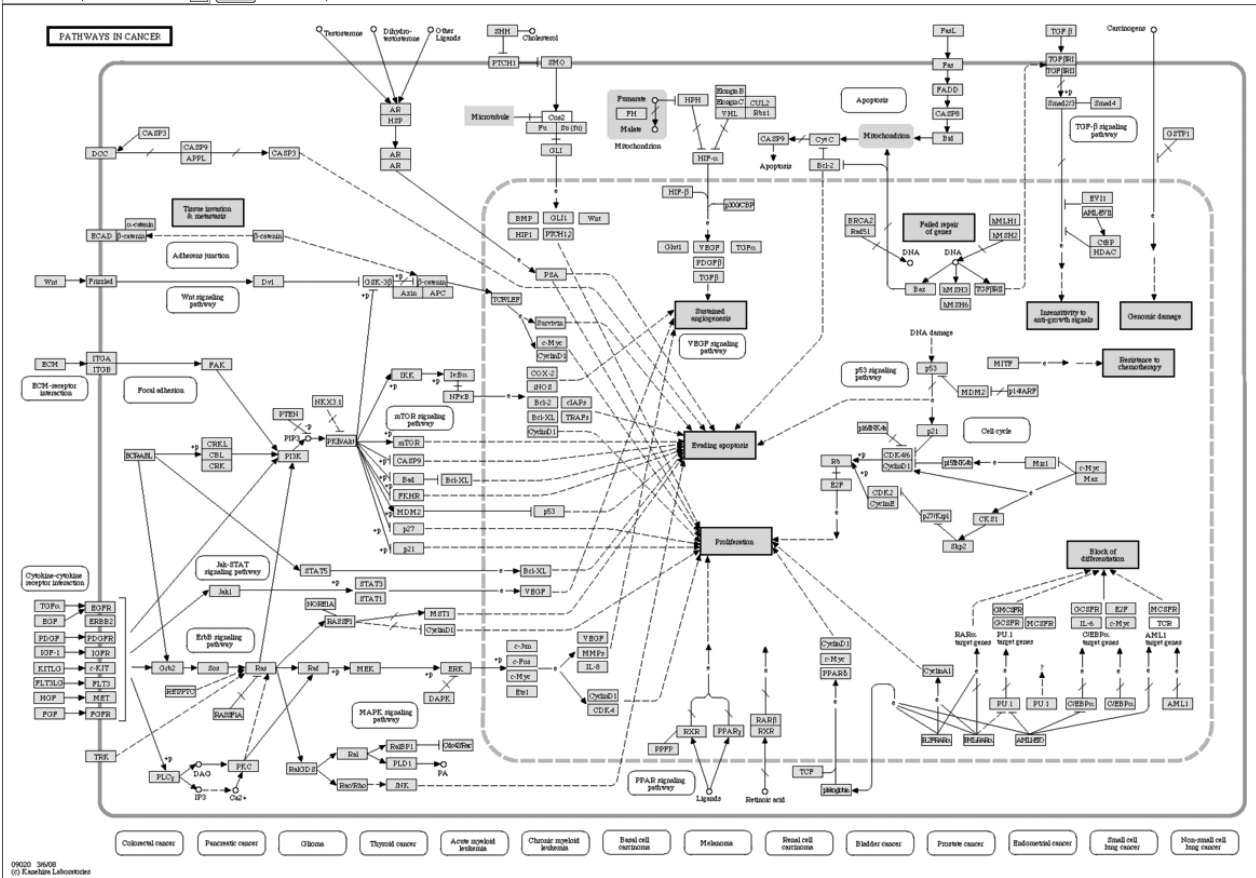


図 13 KEGG Atlas がんのマップ

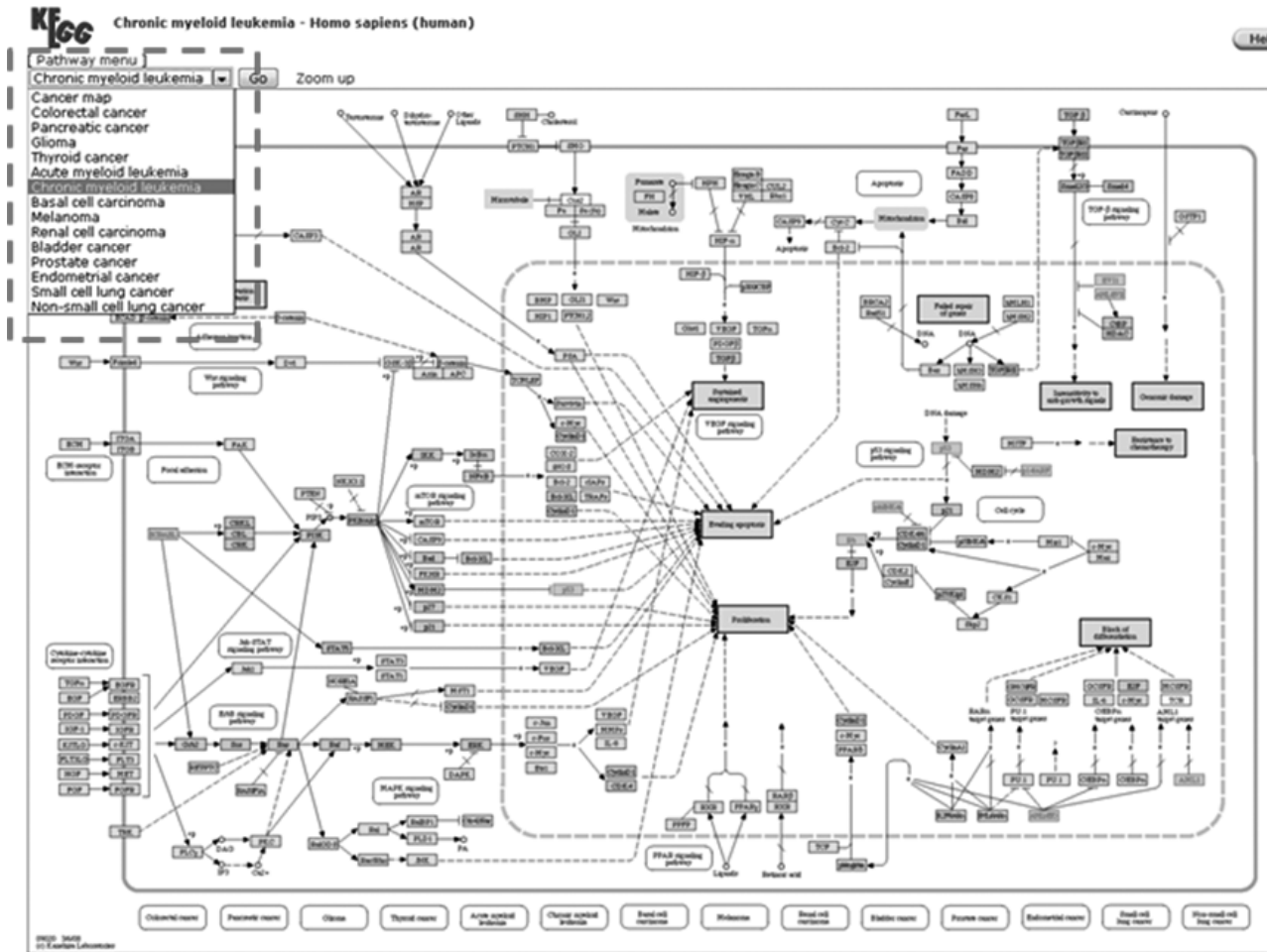


図14 KEGG Atlas 慢性骨髄性白血病の例

図15 KEGG DISEASE 病原体とその感染症

KEGG Human Diseases

▼ ▼ ▼

▼ **Cancers**

- ▼ **Cancers of the Nervous System**
 - H00042 Glioma
 - H00043 Neuroblastoma
 - Schwannoma
 - Meningioma
- ▼ **Cancers of the Digestive System**
 - H00017 Esophageal cancer
 - H00018 Gastric cancer
 - H00034 Carcinoid
 - H00020 Colorectal cancer [PATH:h
 - H00044 Cancer of the anal canal
 - H00048 Hepatocellular carcinoma
 - H00047 Gallbladder cancer
 - H00046 Cholangiocarcinoma
 - H00019 Pancreatic cancer [PATH:h
- ▼ **Cancers of Haematopoietic and Lymphoid T**
 - H00003 Acute myeloid leukemia (A
 - H00001 Acute lymphoblastic leuke
 - H00002 Acute lymphoblastic leuke
 - H00004 Chronic myeloid leukemia
 - H00005 Chronic lymphocytic leuke
 - H00007 Hodgkin lymphoma
 - H00006 Hairy-cell leukemia
 - H00008 Burkitt lymphoma
 - H00009 Adult T-cell leukemia
 - H00010 Multiple myeloma
 - H00011 Lymphoplasmacytic lymphom
 - H00012 Polycythemia vera
- ▼ **Cancers of the Breast and Female Genital C**
 - H00031 Breast cancer
 - H00027 Ovarian cancer
 - H00028 Choriocarcinoma
 - H00026 Endometrial cancer [PATH:h

KEGG DISEASE: H00042

Entry	H00042	Disease
Name	Glioma	
Category	Cancer	
Pathway	hsa05214 Glioma	
Gene	EGFR (amplification, overexpression) [HSA:1956] MDM2 (amplification, overexpression) [HSA:4193] PTEN (mutation) [HSA:5728] p16/INK4A (deletion) [HSA:1029] PDGF-A (overexpression) [HSA:5154] PDGF-B (overexpression) [HSA:5155] PDGFR-alpha (overexpression, amplification) [HSA:5156] PDGFR-beta (overexpression, amplification) [HSA:5159] CDK4 (amplification) [HSA:1019] p53 (mutation) [HSA:7157] RB1 (loss) [HSA:5925]	
Carcinogen	X- and gamma-radiation	
Marker	p53 (mutation) [HSA:7157] PTEN (mutation) [HSA:5728] EGFR (amplification) [HSA:1956]	
Drug	Carmustine [DR:D00254] Lomustine [DR:D00363] Procarbazine [DR:D00478] Vincristine [DR:D02197] Temozolomide [DR:D06067]	
Comment	ICD-O: 9401/3, Tumor type: Anaplastic astrocytoma ICD-O: 9440/3, Tumor type: Glioblastoma	
Reference	PMID:15639402 (gene, tumor type)	
Authors	Soni D, et al.	
Title	Genetics of glioblastoma multiforme: mitogenic signaling and cell cycle pathways converge.	
Journal	J Clin Neurosci 12:1-5 (2005)	
Reference	PMID:11253051 (gene)	
Authors	Holland EC.	
Title	Gliomagenesis: genetic alterations and mouse models.	
Journal	Nat Rev Genet 2:120-9 (2001)	
Reference	PMID:12154354 (gene)	
Authors	Zhu Y, Parada LF.	

図 16 KEGG DISEASE ヒトの病気

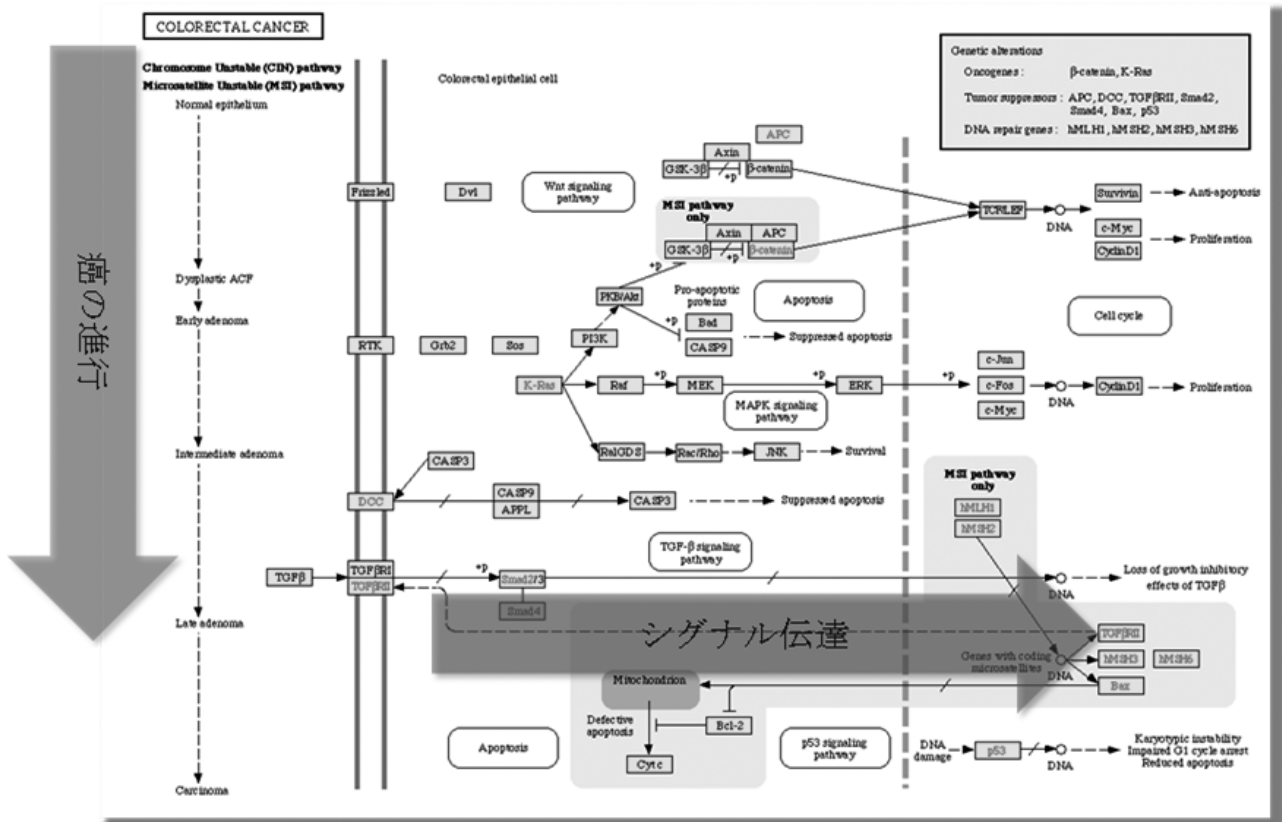


図 17 病気のパスウェイ例：大腸がん

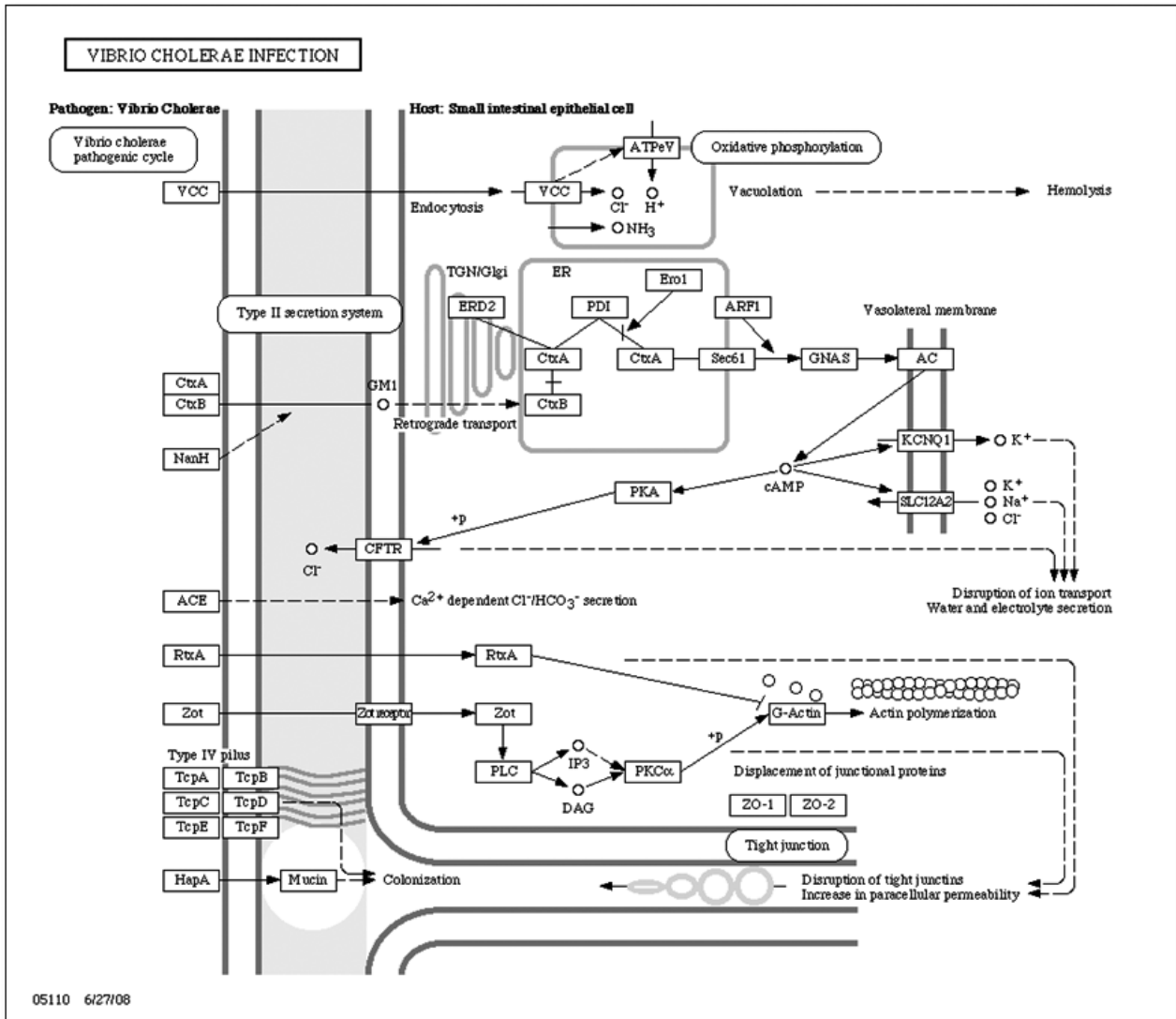


図 18 病気のパスウェイ例：コレラ感染・ヒト内

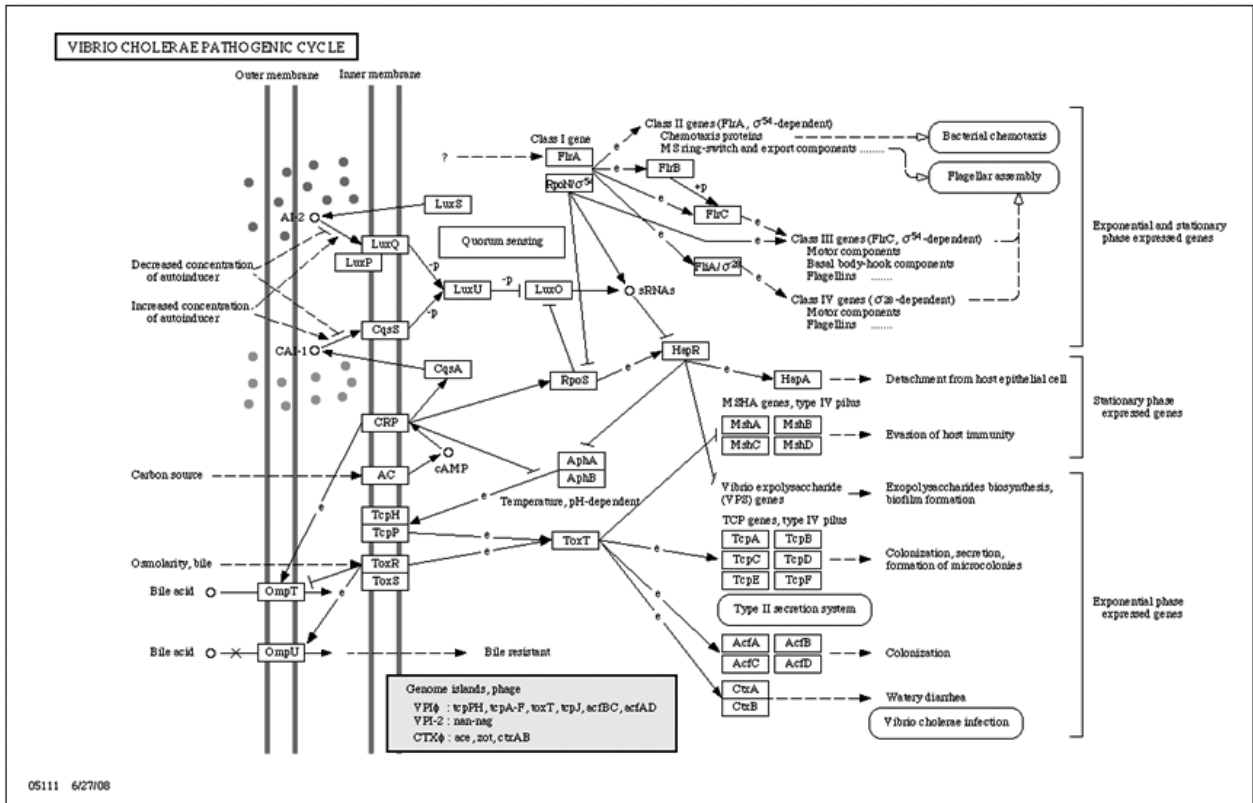


図 19 病気のパスウェイ例：コレラ感染・コレラ内