

生命システムの理解に向けたバイオインフォマティクス

Bioinformatics Towards Understanding Life Systems

第 694 回新潟医学会

日 時 平成 25 年 11 月 16 日 (土) 午後 1 時から
会 場 新潟大学医学部 有壬記念館

司 会 奥田修二郎准教授 (バイオインフォマティクス分野)
演 者 五斗 進 (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)
重水大智 (理化学研究所統合生命医科学研究センター)
菊地正隆 (遺伝子機能解析学分野), 吉沢明康 (島津製作所田中最先端研究所)
山田拓司 (東京工業大学大学院生命理工学研究科生命情報専攻)

1 KEGG: ゲノムと疾患・医薬品をつなぐデータベース

五斗 進

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

KEGG Database for Bridging Genomes and Diseases/Drugs

Susumu GOTO

Bioinformatics Center, Institute for Chemical Research, Kyoto University

要 旨

KEGG (Kyoto Encyclopedia of Genes and Genomes) は生命システムのデータベースとして、ゲノムプロジェクトをはじめとする多くの生命科学基礎研究の参照データベースとして利用されてきた。最近、疾患・医薬品データを従来のゲノムや生命システム情報と統合して使えるようにしており、より一般向けの応用へも可能性を広げている。

Reprint requests to: Susumu GOTO
Bioinformatics Center,
Institute for Chemical Research,
Kyoto University,
Gokasho Uji,
Kyoto 611 - 0011, Japan.

別刷請求先: 〒611-0011 京都府宇治市五ヶ庄
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 五斗 進

キーワード：translational bioinformatics, disease database, drug database, package inserts, drug interaction

KEGG とは

ヒトゲノムプロジェクト以降、次世代シーケンサーを始めとする技術発展により、大量のオミックスデータが得られるようになってきた。その結果、コホート研究など様々な大量データ解析でゲノムと疾患や薬の効きやすさとの関係を解明する研究が進められているが、データの解釈は未だに容易ではない。1995年に開発が始まったKEGG (Kyoto Encyclopedia of Genes and Genomes) プロジェクトでは、代謝系やシグナル伝達系などの生命システムの知識を計算機で扱えるようにし、ゲノムを始めとする大量データの解析に応用でき

るようにしてきた¹⁾。生命システムの知識をデータベース化したものがKEGG PATHWAYとMODULEであり、ゲノムと遺伝子の情報をまとめたものがKEGG GENESである(図1)。さらに、代謝系を中心とするネットワークの重要な構成要素である化合物と反応の情報をまとめたKEGG LIGANDも構築してきた²⁾。このように従来は研究者や教科書の知識であった生命科学や生化学の知識をデータベース化することにより、ゲノムなどのオミックスデータとつないで計算機による解析が可能になった(<http://www.kegg.jp/kegg/>)。

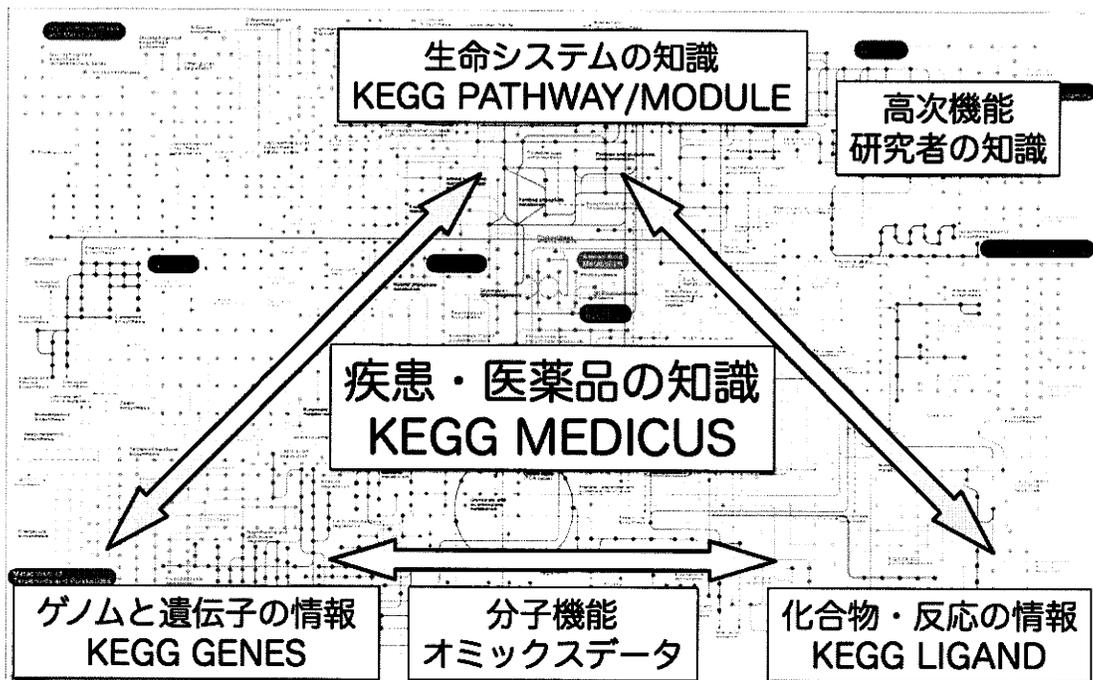


図1 KEGG データベースの概要. 背景の図は代謝系の全体像 (http://www.kegg.jp/kegg-bin/show_pathway?map01100 より).

KEGG MEDICUS

KEGGの開発当初は医学への応用に関する情報は余り含まれていなかったが、病原微生物ゲノムの配列決定および神経変性疾患や癌に関する分子ネットワークの同定に伴ない、データベース化できる情報が徐々に増えてきた。KEGG MEDICUS (<http://www.kegg.jp/kegg/medicus/>) は、このような背景から従来のKEGGの知識に加え、疾患・医薬品・環境物質などをゲノム情報を基盤とした生体システム情報として統合したリソースである³⁾⁴⁾。病因遺伝子やそのパスウェイとともに疾患情報をまとめたKEGG DISEASE、構造ベースで医薬品情報をまとめたKEGG DRUG、生薬・精油など健康促進の効果を示す物質をまとめたKEGG ENVIRONという3つのデータベースから構成されている。さらに、日本医薬品情報センターが提供する医療用・一般用医薬品の添付文書情報も統合的に検索できるようにしている⁵⁾。

疾患情報をまとめたKEGG DISEASEでは疾患の概要を独自に記述し、国際標準の分類体系ICD-10、遺伝病データベースOMIMとの対応を

とっている。さらに、疾患を分子ネットワークに対するゲノムのゆらぎと環境のゆらぎの影響としてまとめているのが特徴である(図2)。ここでゲノムのゆらぎとは先天的・後天的なゲノム上の変異などで、それにより疾患を発症することが分かっているものを指す。また、環境のゆらぎとは病原微生物や細菌叢の変化を指す。これらの変異や変化が分子ネットワーク上のどこに影響を及ぼすかという観点で疾患を見ることができる。さらに、治療薬や診断マーカー、ゲノムバイオマーカーなどの情報も同様の観点から見るための枠組みを提供する。

KEGG DRUGは日米欧で市販されている医療用医薬品・一般用医薬品を構造ベースで統合したデータベースである。効能・効果やATC分類コード以外に、標的タンパク質と代謝酵素の情報をKEGG GENESでつなぐことにより、標的タンパク質がどのような分子ネットワークで働いているかを調べることができるのが大きな特徴である。また、薬が開発されてきた過程を構造変化の情報とともにまとめたパスウェイも提供されている。さらに、添付文書情報との対応関係を取り、そこ

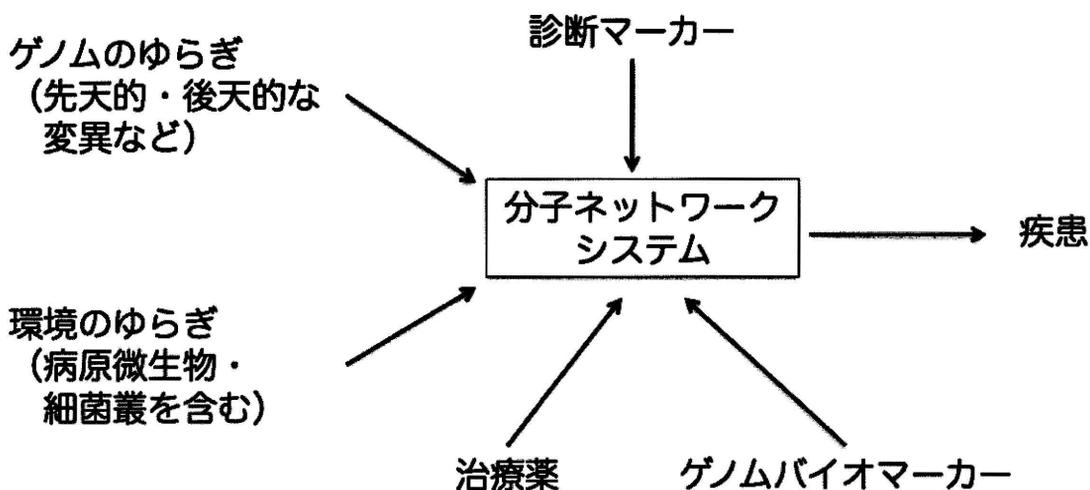


図2 KEGG DISEASEにおける疾患情報とゲノム情報・環境情報を結ぶ分子ネットワークシステム (<http://www.kegg.jp/kegg/disease.html> より)。

に記述されている相互作用(併用禁忌や併用注意)を抽出して検索できるようにしている。KEGG ENVIRON では、主に植物由来の天然物で複数の化学成分からなる未精製物質を、その成分、効能、由来生物種の情報とともにまとめたものである。

データベースに対する検索としては、キーワード検索以外にも化合物構造を MDL/MOL ファイル形式や SMILES 形式で入力することによる類似構造検索と部分構造検索を提供している (<http://www.genome.jp/tools/simcomp/>, <http://www.genome.jp/tools/subcomp/>)。

ま と め

KEGG MEDICUS では、疾患、医薬品、環境物質などをゲノム情報を基盤とした生体システム情報として統合したリソースを提供し、KEGG で従来から提供してきた生命科学や生化学の知識とつなぐことによって様々な解析ができるようにしている。このようなデータベース化は、医薬品のターゲット予測や副作用予測にも応用可能で、既にいくつかの報告もある(例えば⁶⁾⁷⁾)。大量のデータと知識とを融合して機械学習の手法を応用して予測するものである。ターゲット予測についてはウェブベースで利用できるものもあり、著者らのグループでは DINIES (Drug - target Interaction Network Inference Engine based on Supervised analysis, <http://www.genome.jp/tools/dinies/>) サーバーを開発して公開している。

また、基礎研究や科学者向けのデータベースとしてだけでなく、一般向けのリソースとしても応用可能である。KEGG でもお薬手帳のウェブ版を開発して公開している (http://www.kegg.jp/kegg/medicus/medicus2_ja.html)。医薬品に関しては薬剤師による専門知識の提供が重要であるが、処方薬に関する知識を利用者がデータベース検索によって得ることができればリテラシー教育に

もなる。KEGG MEDICUS はそのようなリソースとして一般ユーザーからのアクセスも増えている。

謝 辞

KEGG プロジェクトリーダーの金久實教授をはじめ、KEGG データベースの構築に関与してきた皆さんに感謝します。KEGG データベースは京都大学化学研究所のスーパーコンピュータシステムを利用して開発、サービスされています。

文 献

- 1) Kanehisa M and Goto S: KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28: 27 - 30, 2000.
- 2) Goto S, Nishioka T and Kanehisa M: LIGAND: chemical database of enzyme reactions. *Nucleic Acids Res* 28: 380 - 382, 2000.
- 3) Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M and Tanabe M: Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42: D199 - D205, 2014.
- 4) Kanehisa M, Goto S, Furumichi M, Tanabe M and Hirakawa M: KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38: D355 - D360, 2010.
- 5) 金久 實, 伊藤真純, 奥田修二郎, 五斗 進, 太田福子: 医薬品の統合データベース. *蛋白質 核酸 酵素* 52: 1486 - 1491, 2006.
- 6) Yamanishi Y, Kotera M, Kanehisa M and Goto S: Drug - target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 26: i246 - i254, 2010.
- 7) Mizutani S, Pauwels E, Stoven V, Goto S and Yamanishi Y: Relating drug - protein interaction network with drug side - effects. *Bioinformatics* 28: i522 - i528, 2012.