計画研究：2005 ～ 2009 年度

生命システム解明の基盤データベース構築

●金久 實 [1,2]　◆服部 正泰 [1]　◇片山 俊明 [2]
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　2) 東京大学医科学研究所ヒトゲノム解析センター

＜研究の目的と進め方＞

本研究は、遺伝子･分子レベルの網羅的な解析から、細胞･個体･生態系レベルでの生命システムの全体像を明らかにすることを目指し、新しい情報技術の開発とともに、新しいタイプの基盤データベースを構築することを目的とする。細胞・個体レベルでの生命システム情報は、これまでの KEGG においてすでにデータベース化が行われているので、本研究では生物種間相互作用や環境との相互作用といったより高次レベルの生命システム情報をゲノムの情報と統合し、医療や産業をはじめ、ゲノム情報の有効利用へつなぐ基盤データベースを構築する。同時に支援班との協力の下に、様々な利用・解析ツールを開発して、特定ゲノム 4 領域との間でフィードバック連携をはかり、これら領域の研究推進及び成果の統合化に寄与する。

＜研究開始時の研究計画＞

文献データやなまの実験データを蓄積した従来型のデータベースに対し、本研究のデータベースはこれらのデータから得られる「知識」を蓄積する。その知識をもとに新しい研究を推進することが可能となるように、本データベースの様々な利用ツールを開発する。具体的な内容は以下の通りである。

- 生命システムを構成する部品の情報として、ゲノム情報から直接的に規定される遺伝子とタンパク質はKEGG GENESデータベースに、それ以外のケミカル情報に関連した分子はKEGG LIGANDデータベースに蓄積する。
- KEGG GENESではRefSeq等の公共データベースより、ゲノム配列が決定されたすべての生物種の遺伝子・タンパク質情報を集約し、独自に機能アノテーションを行う。
- 機能アノテーションは、KEGGパスウェイに基づくオーソログ遺伝子分類であるKO (KEGG Orthology) システムを拡張することでそのカバー率を高めていく。また機能アノテーションツール群を改良し、その活用のための支援を行う。
- KEGG LIGANDでは化合物（糖鎖・薬物等を含む）と化学反応に関する情報を文献等より蓄積する。
- 生体内化学反応の分類体系RCを開発し、反応予測や酵素番号づけの自動化を実現する。さらに化合物の化学反応ネットワークと酵素のネットワークの関係から、生体システムと環境との相互作用を解析する方法論を確立する。
- ゲノムネットのWebサービスを通して、KEGGをはじめとしたリソースをプログラムから呼び出し、カスタマイズして利用できるインターフェースKEGG APIの開発を行う。
- 基盤ゲノム領域及び他のゲノム領域に対して、ゲノムデータやEST データの機能アノテーションなど、本研究の成果を生かした支援を、本領域の支援班と連携して行う。
- KEGGの利用講習会やKEGG API入門コース等を京都または東京で定期的に開催して、我が国全体での普及をはかる。
- 高度の専門知識を効率的に集積するために、研究コミュニティと密接に連携し、その知識を集約する「コミュニティデータベース」の枠組みとして、CYORFデータベース等を発展させる。

＜研究期間の成果＞

1. KEGG GENESデータベースの構築

KEGG GENES は全ゲノム配列が決定されたすべての生物種について、その遺伝子セットを RefSeq その他の公共データベースから自動生成し､KEGG 独自のアノテーション（KO アノテーション）を行っているデータベースである。全ゲノム配列が決定された生物種の数は加速度的に増加している。表 1 は本研究期間に、生物種の数､そこに含まれる総遺伝子数､およびオーソロググループ KO の数がどのように増加してきたかを示している。現時点では 533 万の遺伝子の 33% に KO づけがなされており、この割合も年々増加している。後述するように、本研究の成果として、ゲノム数の急増に今後とも十分に対応できる「生命システム解明の基盤データベース構築」が実現できたと考えている。

表1. KEGG GENESデータベースの増加

年月日	生物種	遺伝子	KO
2006.1.1	348	1,852,171	7,983
2007.1.1	486	2,461,000	9,286
2008.1.1	694	3,225,934	10,686
2009.1.1	899	4,214,060	11,417
2009.12.1	1,145	5,336,897	12,848

2. KO (KEGG Orthology) システムの拡張

KEGG GENES のアノテーションでは、RefSeq 等オリジナルデータベースに記述された機能情報は Definition 行としてそのまま残し、後述する KOALA ツールで各遺伝子に KO 識別子（K 番号）を割り当て､その定義とともに Orthology 行に記載している。KO (KEGG Orthology) とは KEGG パスウェイの各ノードまたは BRITE 機能階層の最下層ノードに対応したオーソロググループを手作業で定義したものである。従って、ゲノム中の各遺伝子に K 番号を付与することで、KEGG パスウェイや BRITE 機能階層へのマッピング（エンリッチメント）が可能となる。すなわち KO システムは、ゲノムの情報からパスウェイ等高次生命システム情報を解読するための架け橋となっている。

本研究開始時には、KO システムは代謝・遺伝情報処理・環境情報処理・細胞プロセス・ヒト疾患に関する KEGG パスウェイのみから作られていたが、これに様々なタンパク質ファミリーを表現した BRITE 機能階層を追加することで、KO システムの大幅な拡張が実現した。GO (Gene Ontology) との比較でみると、オントロジーが biological process だけでなく molecular function も含む形に拡張されたことになる。表 2 に現時点で、KO 定義のもととなる KEGG パスウェイと BRITE 機能階層の数を示した。

表 2. KEGG における生命システムの機能階層
（括弧内はパスウェイマップ数 + BRITE 階層ファイル数）

Metabolism
Carbohydrate Metabolism (15)
Energy Metabolism (8+1)
Lipid Metabolism (17+1)
Nucleotide Metabolism (2)
Amino Acid Metabolism (13)
Metabolism of Other Amino Acids (9)
Glycan Biosynthesis and Metabolism (15+3)
Biosynthesis of Polyketides and Nonribosomal Peptides (9)
Metabolism of Cofactors and Vitamins (12)
Biosynthesis of Secondary Metabolites (30)
Xenobiotics Biodegradation and Metabolism (26)
Overview / Enzyme Families (0+4)
Genetic Information Processing
Transcription (3+1)
Translation (2+2)
Folding, Sorting and Degradation (4+3)
Replication and Repair (6+3)
Environmental Information Processing
Membrane Transport (4+2)
Signal Transduction (14+1)
Signaling Molecules and Interaction (4+8)
Cellular Processes
Transport and Catabolism (3)
Cell Motility (3+2)
Cell Growth and Death (6)
Cell Communication (4)
Circulatory System (2)
Endocrine System (7)
Immune System (14)
Nervous System (3)
Sensory System (2)
Development (2)
Behavior (3)
Human Diseases
Cancers (15)
Immune Disorders (6)
Neurodegenerative Diseases (5)
Circulatory Diseases (4)
Metabolic Disorders (3)
Infectious Diseases (4)

3. KOALA (KEGG Orthology And Links Annotation) ツール

KEGG ではオーソログ関係を決めるために、SSEARCH プログラムでアミノ酸レベルのゲノム比較を行い、ゲノムペアごとに全遺伝子間の配列類似性スコアとベストヒット関係を保持した KEGG SSDB データベースが維持されている。これをもとに GFIT ツールでゲノムごとに手作業のアノテーションを行うのが従来のやり方であった。

本研究ではこれまでのノウハウをコンピュータ化した新しいアノテーションツール KOALA (KEGG Orthology And Links Annotation) を開発した。その最大の特色は、従来のようにゲノム単位のアノテーション（縦方向アノテーション）だけでなく、パスウェイマップや BRITE 階層ファイル単位のアノテーション（横方向アノテーション）ができる点である（図 1)。すなわちパスウェイや階層のあるノードにマップされるべきオーソログ遺伝子を、すべての生物種から探してアノテーションを行うのである。また、KOALA には GFIT 作業を自動化した機能があり、間違いの少ない、オーソロググループとしてまとまりのいい安全な K 番号は、新規ゲノムに対して自動アサイメントを行う。K 番号のグルーピングは常に見直しを行っており、グルーピングをきれいにすることで、K 番号の大半を自動アサイメントできる見込みがついた。

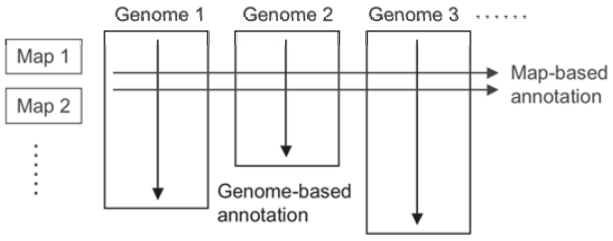

図1. KOALA による縦方向・横方向アノテーション

KOALA からは GFIT ツールへのリンクの他に、染色体上での遺伝子の並びの情報 (バクテリアのオペロンなど) をアノテーションに利用する Gene cluster ツールや、機能単位の K 番号を並べてモジュールやコンプレックスの情報を利用する Ortholog table ツールへのリンクがつけられている。また KEGG SSDB から計算のみでオーソログクラスターを生成する KEGG OC の情報も KOALA で参照できるようになっており、作業の効率化に役立っている。

KOALA の有用性は検証済みであり、国内外の研究グループに広く提供することで、ゲノムやメタゲノムの解読に貢献できる。しかし KOALA は SSEARCH 計算と SSDB データベース、および Oracle の GENES アノテーションデータベースが必要なため、ウェブでの公開サービスには向かない。そこで､内部用（Oracle）から外部用（PostgreSQL）のデータベースコピーを作成し、研究グループごとに自分のゲノムのみ閲覧・編集できるようなシステムを年度末までに開発する。本研究終了後はこれを京都大学化学研究所の共同利用・共同研究の枠組みの中で提供したいと考えている、

4. その他のゲノムアノテーションツール

KOALA より以前に本研究で開発した KAAS (KEGG Automatic Annotation Server) は、BLAST 計算で KEGG GENES に対するゲノム比較を行い、独自のスコアリングアルゴリズムで K 番号づけを自動的に行う。さらにこれをもとにパスウェイマッピングおよび BRITE マッピングを行う。表 3 にある URL で公開しているが、ゲノムネットの計算サービスで最もよく利用されているツールの 1 つになっている。

一方､大量の EST データセットから EST コンセンサスコンティグを自動作成する EGassember ツールも開発し、これは東大ヒトゲノム解析センターの公開サービスとして提供している。コンセンサスコンティグを遺伝子とみなして KAAS で自動アノテーショ

ンを行うと、EST からパスウェイ等の高次機能解読が可能となる。EGassembler ツールは、植物など全ゲノム配列が不足している生物種の KEGG EGENES データベース作成にも利用されている。

さらに KAAS を補うツールとして GENIES を開発した。これはカーネル法という本研究室で取り組んできた最先端の情報技術に基づくもので、多様な大量データ（例えば、マイクロアレイ発現データ、系統プロファイル、ゲノム上の位置情報）をで統合して遺伝子ネットワークを予測する。KAAS による KEGG パスウェイマッピングで色がつかない（遺伝子との対応がつかない）ボックスが残ることがしばしばあるが、そのような missing element を埋めることに有効である。これらのプログラムは表 3 の URL で運用中である。

表3. 開発したゲノム・ESTアノテーションツール

ツール名	URL
KAAS	http://www.genome.jp/tools/kaas/
EGassembler	http://egassember.hgc.jp/
GENIES	http://www.genome.jp/tools/genies/

5. ケミカルアノテーション

ケミカル情報に関しては、まず本研究開始直後に糖鎖に関する解析手法の開発とデータベース化が進み、糖鎖構造・遺伝子・パスウェイを統合した KEGG GLYCAN を以下の URL で公開した。これは、今では糖鎖研究において世界的に最も権威のある標準リソースとして広く利用されている。

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

本研究ではゲノム情報とケミカル情報を融合したケミカルアノテーションに関する情報技術開発を行ってきた。具体的にはゲノムやトランスクリプトームの情報から生体内物質の化学構造を予測する技術である。糖鎖の場合、ゲノムまたはトランスクリプトーム中の糖転移酵素のレパートリーは、個体が潜在的にもつか、または実際に発現している糖鎖構造のレパートリーと深い関連がある。本研究では米国糖鎖コンソーシアム CFG（Consortium for Functional Glycomics）のマイクロアレイ遺伝子発現プロフィールデータから糖鎖構造を予測する試みを行い、表 4 にある GECS ツールを開発した。

表4. 開発したケミカルアノテーションツール

ツール名	URL
GECS	http://www.genome.jp/tools/gecs/
E-zyme	http://www.genome.jp/tools/e-zyme/

低分子の代謝化合物についても、糖鎖の場合と同様に、ゲノム中の酵素遺伝子のレパートリーと生体内の化合物化学構造のレパートリーを関連づけることができるはずである。ただし関与する反応は糖転移酵素反応だけといった単純な形ではない。そこで生体内化学反応に関する知識の体系化を以下の手続きで行ってきた。まず KEGG ENZYME（IUBMB/IUPAC 生化学命名委員会の EC 番号情報）および KEGG PATHWAY にある既知すべての反応は KEGG REACTION に蓄積する。一般に複数の基質と複数の生成物からなる反応を、基質・生成物のペアに分解し、反応の前後でどのような化学構造変化があったかを RDM パターンと呼ぶ反応モチーフで表現する。反応ペアと RDM パターンは KEGG RPAIR に蓄積する。本研究では反応ペアのデータを全面的に見直し、RDM パターンのアノテーションを行った。これを用いて、化合物化学構造のペア（複数でもよい）から反応を予測し、EC 番号を割り当てる E-zyme ツールの新バージョンを開発した。

6. 化学構造変換ネットワーク

ゲノムの情報から生体内で合成され得る化合物の化学構造を予測する方法は、逆に生体外の環境物質が分解され得るかどうかをゲノムの情報と関連づけて予測する方法にもつながっている。実際、本研究においても RDM パターンを用いて、微生物による環境物質等の分解経路予測を行った。さらに図 2 に示したように、合成経路予測は植物や微生物のゲノムから天然物を予測し、さらには医薬品開発へとつないでいく方向が考えられる。これをめざして本研究では、植物ゲノムと EST の情報を KEGG GENES に、植物の二次代謝経路を KEGG PATHWAY に、合成経路に基づく植物二次代謝物質の分類を KEGG BRITE に整備し統合した KEGG PLANT インターフェースを開発し、以下の URL で公開した。

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

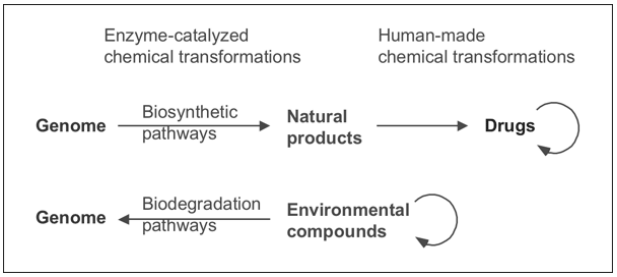

図 2. 化学構造変換ネットワークによるゲノム･天然物･医薬品･環境物質の関連解析

一方、多くの医薬品は天然物などのリード化合物や既存の医薬品を起点に構造展開を行うことで、すなわち基本骨格を維持しながら化学構造を変化させることで、新製品の開発が行われてきた。このような医薬品開発の歴史を化学構造変化のパターンとして眺めると、酵素反応による化学構造変化パターンと統合して知識ベース化する可能性が見えてくる。本研究ではその準備段階として市販されている医薬品の化学構造をコア部分と周辺部分に分解し、コアごとに周辺部分の変化と薬効等との関連解析を行った。

7. KEGG APIの開発

本研究では KEGG を個別目的にカスタマイズして利用できる環境作りとして、KEGG API の開発も継続して行ってきた。これは KEGG に対する検索や解析をライブラリとして提供し、データベース作成者と末端利用者をつなぐソフトウェア開発者が自分のプログラムに組み込んでを利用できるようにしたものである。従来からの SOAP サーバーに加えて、本研究では REST サーバーの提供も開始した。

8. 支援活動

支援活動については、本特定領域の個別の研究グループに対して cDNA データのアノテーション支援などを行い、また以下の公開行事を開催した。

(1) らん藻ゲノムアノテーションワークショップ

2006 年 8 月 22 ～ 23 日に京大化研バイオインフォマティクスセンターで我が国のシアノバクテリア研究コミュニティとともに「らん藻ゲノムアノテーションワークショップ」を開催し、CYORF アノテーションツールの講習を行った。参加者は 31 名で、米国からも参加があり CyanoBIKE と CYORF の連携が始まっ

た。

(2) プログラミング実習

2006 年 10 月 27 日と 11 月 10 日に「プログラミング実習」を東大医科研ヒトゲノム解析センターにおいて開催し、プログラミング言語（Perl, Ruby など）の概要の講習と、計算機を用いて簡単なプログラミングができるようになるまでの実習を行ったた。延べ参加者数は 12 名。ユニークな参加者数は 8 名。

(3) 2006年度KEGGデータベース利用講習会

2006 年 11 月 17 日、24 日、12 月 1 日に「KEGG データベース利用講習会」を東大医科研ヒトゲノム解析センターにおいて開催し、KEGG データベース及びゲノムネットサービスの利用法について実際に計算機を操作しながら習得した。延べ参加者数は 26 名。ユニークな参加者数は 16 名。

(4) 2007年度第1回KEGGデータベース利用講習会

2007 年 8 月 30 日と 31 日に「第 1 回 KEGG データベース利用講習会」を京大化研バイオインフォマティクスセンターにおいて開催し、EGassembler, KAAS, GENIES を中心とした実習を行った。参加者は 35 名。

(5) 2007年度第2回KEGGデータベース利用講習会

2007 年 11 月 15 日と 16 日に「第 2 回 KEGG データベース利用講習会」を東大医科研ヒトゲノム解析センターにおいて開催し、医薬品や化合物を中心とした実習を行った。参加者は 21 名。

(6) UniProtデータベース利用講習会

2007 年 10 月 2 日に京大化研バイオインフォマティクスセンターで、4 日と 5 日に東大医科研ヒトゲノム解析センターで「UniProt データベース利用講習会」を開催した。これは KEGG と UniProt の連携の一貫として行ったものである。参加者は京都が 10 名。東京が延べ 22 名。

(7) 2008年度KEGGデータベース利用講習会

2009 年 1 月 29 日と 30 日に東大医科研ヒトゲノム解析センターで KEGG データベース利用講習会を開催した。参加者は 35 名。

(8) 2009年度KEGGデータベース利用講習会

2010 年 2 月 4 日と 5 日に東大医科研ヒトゲノム解析センターで KEGG データベース利用講習会を開催。

＜国内外での成果の位置づけ＞

KEGG はすでに米国 NCBI や欧州 EBI 等と並び、世界で最もよく利用されるバイオ情報サービスとなっている。本研究期間にとくに米国 NCBI との連携が進んだ。NCBI RefSeq から KEGG GENES が作られ、その情報が Entrez Gene に取り込まれて KEGG PATHWAY へのリンクがつけられている。今年度には KEGG PATHWAY の遺伝子・化合物リストを取り込んだ NCBI BioSystems が公開された。NCBI では KEGG のような知識集約作業は行っていないが、大量データの高度なコンピュータ処理がなされており、相補的な両者の間には膨大な数（1 千万件近く）の相互リンクがつけられている。今後は疾患情報を中心に OMIM も含めた連携体制を検討している。

NCBI 以外にも UniProt や ChEBI など欧州の代表的データベースと協力関係にあり、相互にリンクづけがなされている。また、米国糖鎖コンソーシアム CFG との連携による KEGG GLYCAN や、IUPAC/IUBMB 生化学命名委員会および ExplorEnz との連携による KEGG ENZYME をはじめ、様々なグループとも協力関係にある。

表5. KEGGとNCBIの連携体制

KEGG	NCBI
GENES	RefSeq (Gene)
COMPOUND	PubChem
DRUG	PubChem
PATHWAY	BioSystems
DISEASE	BioSystems (予定)

学問的な成果では、まず糖鎖科学において合成経路の観点からゲノムの情報を糖鎖構造予測などに有効利用する考え方と方法論を示した。これを図 2 に示したように、生体内での合成経路と分解経路および人手による合成経路を化学構造変換ネットワークとして一般化し、ゲノムと天然物・医薬品・環境物質との関連解析へと発展させた。このような本研究の先駆性が上記の国際連携の基礎となっている。

KEGG はその３つの柱のうち、パスウェイ情報とケミカル情報については、国際標準のデータベースとしての地位を獲得している。しかしもう１つのゲノム情報については、国際的評価はまだそれほど高くない。しかしながら本研究により、ゲノム情報においても KEGG が国際標準となる道筋ができてきた。現在 KEGG では 1000 種類の生物種に含まれる 500 万の遺伝子・タンパク質に対し、その 1/3 に機能アノテーションをつけている。一方、タンパク質配列データを網羅的に集めている UniProt には KEGG の 600 倍 (60 万) の生物種があるが、タンパク質の数は 2 倍 (1000 万) しかない。全ゲノム配列が決定された生物種のデータしか集めないという KEGG の基本方針でも、すでに既知のタンパク質の半分はカバーしており、KEGG OC クラスタリングの結果では、タンパク質ファミリーの 9 割はすでにカバーしている。近い将来に KEGG と UniProt は量的に差がなくなるだろう。UniProt はごく一部（5%）の手動アノテーション（SwissProt）と大量の自動アノテーション（TrEMBL）が混在しており、全体の品質が落ちてしまった印象がある。KEGG では個々のゲノムにアノテーションを行う立場ではなく、生物界すべてのゲノムに一括アノテーションを行う立場をとっている。KOALA はこのような考え方を実用化したものである。一般にはゲノム数の増加がアノテーションの負荷を高め作業を困難にしていると思われているが、KEGG では状況は全く逆で、ゲノム数の増加が生物種間のつながりを分かり易くし、アノテーションが容易になる方向へ向かっている。また KEGG PATHWAY と KEGG BRITE の拡充により KO の数が増え、アノテーションのカバー率も上がっている。KEGG が量的にも質的にもゲノムアノテーションの国際標準となり得る状況になってきたと考えている。

＜達成できなかったこと、予想外の困難、その理由＞

KEGG のウェブサイトには１日あたり 15,000 のユニーク IP から数百万件（ロボットを含む）のアクセスがある。トムソン社の ISI Web of Science によると Nucleic Acids Research Database Issue に掲載された 2006 年の KEGG 論文（登録番号 0601301835）と 2008 年の KEGG 論文（登録番号 0801200923）は、2009 年 12 月現在それぞれ 580 回と 250 回の引用がある。また Nature 誌をはじめ high impact factor journals に KEGG を利用した論文がしばしば見られる。さらにゲノムネットのフィードバックシステムで寄せられる意見や、利用講習会等での反響から、国内でも個人レベルのユーザは KEGG を頻繁に利用し、また KEGG に対する期待も大きいようである。しかしながら基盤ゲノム領域あるいは特定ゲノム 4 領域研究参加者による KEGG

の利用は不十分であった。本研究では特定グループのニーズに基づく開発ではなく、幅広く国内の研究基盤、さらには国際的な研究基盤となるような汎用的なデータベース・ソフトウェアツールの開発を行ってきた。同時に個々のニーズに対応できるよう講習会等の開催も行ってきた。このような方策が特定ゲノム4領域の研究推進にもつながると考えたからである。その是非については今後の評価を待ちたい。

＜今後の課題、展望＞

今後の課題、展望については3つある。第1はすでに述べた通り、ゲノムの急増に対応しつつ、KEGG GENESのアノテーション精度を高めていくことである。1000種類の大腸菌ゲノム、あるいは10,000種類の脊椎動物ゲノムなど、膨大なシークエンシングプロジェクトが計画されている。これらを手作業でアノテーションすることは不可能であり、いかに高品質のコンピュータ処理ができるかが課題であり、国際競争力の観点からも重要である。アノテーションに必須のオーソログ情報をパスウェイのコンテキストに依存した形で手作業で決め、アノテータの知識をコンピュータ化したKOALAで自動処理するKEGGのやり方は、今後のゲノム数の増加に十分対応できる。

第2はメタゲノミクスへの対応である。米国のHMP (Human Microbiome Project) や欧州のMetaHIT (Metagenomics of the Human Intestinal Tract) はとくにヒトの腸内細菌叢などに関するプロジェクトであり、ヒトゲノム情報を補完し、ヒト生体システムおよび疾患メカニズムの理解に有用な情報を提供すると考えられる。これら国際プロジェクトとも連携し、メタゲノムシークエンシングデータから個々のバクテリアごとの生体システム、バクテリアコミュニティとしての生体システム、さらにはこれらと相互作用するヒトの生体システムを解読するための方法論の開拓を行っていく。

第3は疾患情報のデータベース化である。遺伝子との関連に関する疾患情報を記述したデータベースは多数存在するが、これらは読んで理解することを目的としたものである。KEGGでは分子ネットワークと疾患の関連に着目して、計算可能な疾患情報データベースの構築を行っており、パスウェイマップと遺伝子・分子リストという形の表現法を用いている。現時点ではまだ100程度の疾患しか公開していないが、これを早急に充実させる予定である。これにより疾患に関与する分子ネットワークを他のデータと統合して、とくにゲノムやメタゲノムのシークエンシングデータやその他のハイスループット実験データと統合して解析することが可能になるだろう。

＜研究期間の全成果公表リスト＞

論文

1. 0912011029
   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., in press (2010).
2. 0908261758
   Yamanishi, Y., Hattori, M., Kotera, M., Goto, S., and Kanehisa, M.; E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs. Bioinformatics 25, i79-i86 (2009).
3. 0908261742
   Hashimoto, K., Tokimatsu, T., Kawano, S., Yoshizawa, A.C., Okuda, S., Goto, S., and Kanehisa, M.; Comprehensive analysis of glycosyltransferases in eukaryotic genomes for structural and functional characterization of glycans. Carbohydrate Res. 344, 881-887 (2009).
4. 0903121042
   Shigemizu, D., Araki, M., Okuda, S., Goto, S., and Kanehisa, M.; Extraction and analysis of chemical modification patterns in drug development. J. Chem. Inf. Model. 49, 1122-1129 (2009).
5. 0901161008
   Shimizu, Y., Hattori, M., Goto, S., and Kanehisa, M.; Generalized reaction patterns for prediction of unknown enzymatic reactions. Genome Informatics 20, 149-158 (2008).
6. 0901161004
   Takarabe, M., Okuda, S., Itoh, M., Tokimatsu, T., Goto, S., and Kanehisa, M.; Network analysis of adverse drug interactions. Genome Informatics 20, 252-259 (2008).
7. 0901131629
   Okuda, S., Yamada, T., Hamajima, M., Itoh, M., Katayama, T., Bork, P., Goto, S., and Kanehisa, M.; KEGG Atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res. 36, W423-W426 (2008).
8. 0801200923
   Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y.; KEGG for linking genomes to life and the environment. Nucleic Acids Res. 36, D480-D484 (2008).
9. 0801200915
   Hashimoto, K., Yoshizawa, A.C., Okuda, S., Kuma, K., Goto, S., and Kanehisa, M.; The repertoire of desaturases and elongases reveals fatty acid variations in 56 eukaryotic genomes. J. Lipid Res. 49, 183-191 (2008).
10. 0801200905
    Kadowaki, T., Wheelock, C.E., Adachi, T., Kudo, T., Okamoto, S., Tanaka, N., Tonomura, K., Tsujimoto, G., Mamitsuka, H., Goto, S., and Kanehisa, M.; Identification of endocrine disruptor biodegradation by integration of structure-activity relationship with pathway analysis. Environ. Sci. Technol. 41, 7997-8003 (2007).
11. 0708091351
    Limviphuvadh, V., Tanaka, S., Goto, S., Ueda, K., and Kanehisa, M.; The commonality of protein interaction networks determined in Neurodegenerative disorders (NDDs). Bioinformatics 23, 2129-2138 (2007).
12. 0708091346
    Itoh, M., Nacher, J.C., Kuma, K.I., Goto, S., and Kanehisa, M.; Evolutionary history and functional implications of protein domains and their combinations in eukaryotes. Genome Biol. 8, R121 (2007).
13. 0708091337
    Fujita, M., Mihara, H., Goto, S., Esaki, N., and Kanehisa, M.; Mining prokaryotic genomes for unknown amino acids: a stop-codon-based approach. BMC Bioinformatics 8, 225 (2007).
14. 0704271738
    Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., and Kanehisa, M.; KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182-W185 (2007).

15. 0704271736
Oh, M., Yamada, T., Hattori, M., Goto, S., and Kanehisa, M.; Systematic analysis of enzyme-catalyzed reaction patterns and prediction of microbial biodegradation pathways. J. Chem. Inf. Model. 47, 1702-1712 (2007).
16. 0704271733
Minowa, Y., Araki, M., and Kanehisa, M.; Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. J. Mol. Biol. 368, 1500-1517 (2007).
17. 0702011804
Yoshizawa, A.C., Kawashima, S., Okuda, S., Fujita, M., Itoh, M., Moriya, Y., Hattori, M., and Kanehisa, M.; Extracting sequence motifs and the phylogenetic features of SNARE-dependent membrane traffic. Traffic 7, 1104-1118 (2006).
18. 0702011755
Schwartz, J.M. and Kanehisa, M.: Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Bioinformatics 7, 186 (2006).
19. 0608081011
Masoudi-Nejad, A., Tonomura, K., Kawashima, S., Moriya, Y., Suzuki, M., Itoh, M., Kanehisa, M., Endo, T., and Goto, S.; EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. Nucleic Acids Res. 34, W459-W462 (2006).
20. 0602081729
Itoh, M., Goto, S., Akutsu, T., and Kanehisa, M.; Fast and accurate database homology search using upper bounds of local alignment scores. Bioinformatics 21, 912-921 (2005),
21. 0602081724
Yamanishi, Y., Vert, J.-P., and Kanehisa, M.; Supervised enzyme network inference from the integration of genomic data and chemical information. Bioinformatics 21, i468-i477 (2005).
22. 0601301835
Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M.; From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 34, D354-357 (2006).
23. 0601301825
Kawano, S., Hashimoto, K., Miyama, T., Goto, S., and Kanehisa, M.; Prediction of glycan structures from gene expression data based on glycosyltransferase reactions. Bioinformatics 21, 3976-3982 (2005).
24. 0601301821
Hizukuri, Y., Yamanishi, Y., Nakamura, O., Yagi, F., Goto, S., and Kanehisa, M.; Extraction of leukemia specific glycan motifs in human by computational glycomics. Carbohydr. Res. 340, 2270-2278 (2005).
25. 0601301811
Hashimoto, K., Goto, S., Kawano, S., Aoki-Kinoshita, K.F., Ueda, N., Hamajima, M., Kawasaki, T., and Kanehisa, M.; KEGG as a glycome informatics resource. Glycobiology 16, 63R-70R (2005).