

平成 26 年度 研究開発実施報告書

ライフサイエンスデータベース統合推進事業「統合化推進プログラム」研究代表者

[金久 實]

[京都大学 化学研究所・特任教授]

[ゲノムとフェノタイプ・疾患・医薬品の統合データベース]

§1. 研究実施体制

(1) 研究代表者グループ(京都大学)

① 研究代表者: 金久 實 (京都大学化学研究所、特任教授)

② 研究項目

(1) 疾患情報のデータベース

(1-1) KEGG DISEASE 疾患エン트리

(1-2) 病原体ゲノムのシグネチャーモジュール

(2) 医薬品情報のデータベース

(2-1) KEGG DRUG と医薬品添付文書

(2-2) 医薬品グループデータベース

(3) 統合インターフェース

§ 2. 研究実施内容

研究の目的

ヒトゲノム、病原体ゲノム、様々なメタゲノムの解析で、ゲノムのゆらぎや環境のゆらぎを分子レベルで系統的に計測することが可能になってきた。そこから生体システムの機能やヒトの疾患を理解し、医療や創薬への実用化を実現するためには、計算手法とともに知識ベースに基づくバイオインフォマティクス技術が不可欠である。本研究では、個々の遺伝子だけでなく、複数の遺伝子から構成された機能モジュール、さらには遺伝子、タンパク質、環境因子、医薬品等から構成された相互作用ユニットに関する知識をデータベース化し、ゲノムとフェノタイプ・疾患・医薬品をつなぐ統合データベースを開発する。病原体ゲノムと薬剤耐性との関連、ヒトゲノムと薬剤応答との関連などを推定するためのデータベース利用ツールも開発する。

概要

本研究開発では、疾患情報のデータベース、医薬品情報のデータベース、及び統合インターフェースの3つの研究項目を設定し、ゲノムとフェノタイプ・疾患・医薬品の統合データベース **KEGG MEDICUS** の開発を行う。疾患情報のデータベースでは、疾患を遺伝要因と環境要因のリスト(広い意味の相互作用ユニット)で表現した **KEGG DISEASE** 疾患エントリの拡充と高品質化を行う。また新たに病原体ゲノムシーケンスから薬剤耐性や病原性を判定するために、シグネチャーとなる遺伝子や遺伝子セット(モジュール)の知識を蓄積し、実用的価値のあるデータベースとする。医薬品情報のデータベースでは、医薬品の有効成分を網羅的に集積した **KEGG DRUG** を中核に、医療用・一般用医薬品添付文書の統合、医薬品相互作用情報の更新、疾患と関連づけた医薬品分類の開発などを行う。また新たに相互作用の観点から各医薬品をグループ化した **KEGG DGROUP** 医薬品グループデータベースを開発する。統合インターフェースについては、ゲノムシーケンスデータからフェノタイプ・疾患・医薬品に関する知見を得ることができるよう **KEGG Mapper** ツールの改良と拡張を行う。また病原体ゲノムから薬剤耐性や病原性の解釈を支援するツールの開発も行う。

進捗状況

それぞれの開発項目の進捗状況は以下の通りで、研究開発計画書に記載した目標はすべて達成した。(1-1) **KEGG DISEASE** 疾患エントリ数は 1,359 (2014.4.15) から 1,414 (2015.4.1) に増加した。また各疾患に分類情報のフィールドを追加し、医薬品情報のフィールドでは新規に開発中の医薬品グループ表記を導入した。(1-2) 病原体ゲノムのシグネチャーモジュールでは、薬剤耐性について広範な文献調査と予備的な解析を行ったところ、遺伝子セット(モジュール)とともに重要な遺伝子変異をシグネチャーKO (**KEGG Orthology**) と呼ぶ形式で表現可能なことが分かった。シグネチャーモジュールについては 54 種類、シグネチャーKO については β -ラクタマーゼの 31 種類を公開済みである。(2-1) 医薬品添付文書とそこから抽出した相互作用情報を毎月更新して **KEGG DRUG** に統合し、常に最新の医薬品情報を提供している。また添付文書にある適応症情報をもとに疾患との対応づけを行い、医薬品クラスの概念で疾患ごとの分類体系を作成している。(2-2) 医薬品グループデータベース **KEGG DGROUP** の構築を開始し、ATC 分類に基づくグループと添付文書にある相互作用グループを中心に 1,808 エントリを新規作成し公開した。

(3) 統合インターフェースでは BRITE テーブル形式へのマッピング機能を用い、病原体ゲノムから薬剤耐性のシグネチャーを判定する Pathogen Checker ツールを開発した。また KEGG MEDICUS API の改良も行った。

研究成果

本研究では 2 つの新規開発項目があった。まず薬剤耐性の知識ベース構築について、当初計画ではシグネチャーとなる遺伝子セット(モジュール)だけでは不十分で、遺伝子変異のデータベース化を解決しなければならない課題として記載した。一般にはアミノ酸配列のどこがどのように変異しているかが重要であるとの考えがあり、例えば β -ラクタマーゼについては細かな変異情報を蓄積する試みがなされている。本研究では、耐性遺伝子が急速に伝播していることから系統的な関係に着目し、細分化された KEGG Orthology (KO) で系統的に同じ配列グループを表現する方法を考案した。これに耐性のある薬剤の情報を付加してデータベース化することにより、新規ゲノムで配列比較による KO アサインメントを行うことで薬剤耐性を推定する。この考え方を実践したのが Pathogen Checker ツール(現時点では β -ラクタマーゼに限定)である。もう 1 つの新規開発項目であった医薬品グループデータベースについては当初の予定通り進め、上記の耐性と関連づけた薬剤情報や、疾患と関連づけた医薬品クラスの情報などにおいて、有効性を発揮している。またこれらの開発の中でゲノム・疾患・医薬品などの相互関係を BRITE テーブルファイルと呼ぶ形式で表現することとし、これに対応するよう KEGG Mapper ツールの拡張も行った。なお当初計画ではモジュールなどの新規データを既存のエントリに組み込んで提供することを予定していたが、これらは BRITE テーブルファイル形式のまま、ウェブ・FTP での提供を行っている。

今後の見通し

Antimicrobial resistance (AMR) は、どんな先端医療においても感染症をコントロールする際の大きな問題であるとして、とくに米国では CDC やホワイトハウスからレポートが出され、ゲノムシーケンシングの重要性が取り上げられている。個人ゲノム情報を用いた個別化医療よりも先に、病原体ゲノム情報を用いた感染症対策において、ゲノムシーケンシングの実用化が進むかもしれない。一方、既知の薬剤耐性遺伝子については、いくつかの遺伝子ファミリーごとのデータベースは存在するが、網羅的なデータベースは存在しない。本研究ではこれを開発することにより、病原体ゲノムシーケンスデータの解釈を支援する。既知の遺伝子・タンパク質の機能を KO で表現するのは、「限られた実験データを一般化した形で表現することで予測に利用できる知識にする」という KEGG の基本的な考え方である。これまでの研究により、機能を細分化した KO (tight KO) が薬剤耐性遺伝子で有用なことが明らかとなった。これを様々な遺伝子ファミリーに適用し、シグネチャー KO とシグネチャーモジュールの知識ベースを構築できる見通しとなった。

§3. 成果発表等

(3-1) 原著論文発表

- ① 発行済論文数(国内(和文) 0件、国際(欧文) 0件):
- ② 未発行論文数(“accepted”、“in press”等)(国内(和文) 0件、国際(欧文) 1件)
- ③ 論文詳細情報

* 1. Minoru Kanehisa, “KEGG bioinformatics resource for plant genomics and metabolomics”, *Methods Mol. Biol.*, in press.

概要:植物研究者向けの KEGG 紹介論文であるが、一般的な内容として KEGG アノテーションの方法及び Pathogen Checker が利用している BlastKOALA 自動アノテーションの方法も含まれている。

(3-2) データベースおよびウェブツール等の構築と公開

- ① 公開中のデータベース・ウェブツール等
別紙1参照。
- ② 未公開のデータベース・ウェブツール等
なし。

(3-3) データベース・ウェブツール等の活用事例

Web of Science で2014年4月から2015年3月までの1年間に KEGG の論文引用数は1,930件、Google Scholar で2014年1月から12月までの1年間に Kanehisa の論文引用数は4,300件であった。この中でどれだけが KEGG MEDICUS データベースの部分を利用しているかは詳しく調査しなければ分からないが、Google Scholar で KEGG DISEASE あるいは KEGG DRUG をキーワードに2014以降の検索をすると100件以上のヒットがあった。以下に3つだけ例を示す。

- Wu D, et al., Genome-wide association study combined with biological context can reveal more disease-related SNPs altering microRNA target seed sites. *BMC Genomics*. 15:669, 2014.
- Bakir-Gungor B, et al., Identification of possible pathogenic pathways in Behçet's disease using genome-wide association study data from two different populations. *Eur J Hum Genet*. 23(5):678-87, 2015.
- Aksoy BA, et al., Prediction of individualized therapeutic vulnerabilities in cancer from genomic profiles. *Bioinformatics*. 30(14):2051-9, 2014.

(3-4) その他の著作物(総説、書籍など)

なし

(3-5) 国際学会および国内学会発表

- ① 招待講演 (国内 0件、国際 1件)
〈国内〉

〈国際〉

1. Minoru Kanehisa, Automated interpretation of metabolic capacity from complete genomes using KEGG modules, Plant and Animal Genome XXIII Conference, San Diego, 2015年1月11日

② 口頭講演 (国内 0件、国際 0件)
なし

③ ポスター発表 (国内 0件、国際 0件)
なし

(3-6) 知的財産権の出願
なし

(3-7) 受賞・報道等
なし

(3-8) その他(技術移転など)
なし

§4. 研究開発期間中に主催した活動(主催したワークショップ等)

年月日	名称	場所	参加人数	概要
2014年11月25 ~27日	KEGG MEDICUS の展示・デモ	パシフィコ 横浜	100人	日本分子生物学会年 会での展示

以上

別紙1 既公開のデータベース・ウェブツール等

No.	研究開発課題名	データベース、ウェブツール等の名称【公開】	概要【公開】 (150字程度)	URL【公開】	公開開始日	関連論文 (論文リストに記載があれば、その番号でも可)	備考
01	ゲノムとフェノタイプ・疾患・医薬品の統合データベース	KEGG MEDICUS 統合データベース	ヒトゲノム、病原体ゲノム、腸内細菌メタゲノムをはじめとしたゲノムの情報と疾患との関連、および医薬品の作用・副作用との関連が知識ベース化され、ゲノム解読と有効利用を促進する統合リソースです。また日本と米国のすべての医薬品添付文書も統合されており、一般社会にとっても有用なリソースです。	http://www.kegg.jp/kegg/medicus	2010/10/1		
02	ゲノムとフェノタイプ・疾患・医薬品の統合データベース	Pathogen Checker	病原体ゲノムから薬剤耐性を推定するツールです。ゲノムシーケンスデータにKEGGの自動アノテーションを適用して、シグネチャー-KOおよびシグネチャーモジュールと呼ぶ特徴的な遺伝子あるいは遺伝子セットの有無を調べ、どのような薬剤に対して耐性があるかを推定します。	http://www.kegg.jp/kegg/disease/pathogen_checker.html	2015/4/1		