

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）
「ゲノムとフェノタイプ・疾患・
医薬品の統合データベース」
研究開発期間：平成26年4月～平成29年3月

研究開発終了報告書

研究代表者：金久 實
(京都大学 化学研究所 特任教授)

§1. 研究開発実施の概要

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ことを目的とした。また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・相互作用との関連を推定するツールをはじめ、データベース利用ツールの開発も目的とした。これらの成果は KEGG MEDICUS (<http://www.kegg.jp/kegg/medicus/>) に含めて公開している。

KEGG MEDICUS は統合化推進プログラム第 I 期より開発・運用を行っている統合データベースである。本研究開発計画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、統合インターフェースの 3 つの開発項目を設定して開発を進めてきた。(1) 疾患情報のデータベースでは、病因遺伝子、病原体、環境因子などの知識を集約した KEGG DISEASE データベースの大幅な拡充を行った。また新たに文献情報をもとに病原体ゲノムから薬剤耐性を推定するために特徴的な遺伝子配列パターン (シグネチャー KO) と特徴的な複数遺伝子セット (シグネチャーモジュール) のデータベース化を行った。(2) 医薬品情報のデータベースでは、KEGG DRUG データベースの拡充とともに、日本と米国の医薬品添付文書を毎月更新し、KEGG DRUG との統合を継続して行っている。また医薬品の作用・副作用を標的分子、薬物代謝酵素、その遺伝子多型などを含むネットワークの中で理解するために、医薬品相互作用ユニットの考え方を導入し、医薬品のオーソログに相当する医薬品グループを定義した KEGG DGROUP データベースの新規開発を行った。(3) 統合インターフェースでは、病原体ゲノムから上記のシグネチャーデータセットを用いて薬剤耐性を推定する Pathogen Checker の開発を行った。また第 I 期で作成した医薬品相互作用チェックツールの新しいバージョンを開発し、医薬品グループを用いる形に変更した。さらに KEGG DISEASE を KEGG Mapper の対象データベースの 1 つとして Disease mapping ツールの開発を行った。

本研究では当初の開発項目に掲げた開発計画はすべて達成した。達成項目としての成果以外に、有用なデータベースを構築するアプローチとして 2 つの成果があった。1 つは何をデータソースにするかである。もともと KEGG DISEASE も KEGG DRUG も文献情報に基づき作成してきたデータベースである。一方、医薬品添付文書は実際に使われている薬と、適応症として実際に治療の対象となる疾患が記載されている。KEGG DRUG では塩や水和状態の違いなどを区別してエントリ化しているため添付文書の 4 倍のエントリ数があるが漏れはない。しかし KEGG DISEASE では疾患の定義が適切でなかったり、漏れがあったりするケースが多いことが分かり、添付文書の適応症との対応づけにより全体的な見直しを行っている。社会的に有用なデータベースを構築するには、学术论文だけでなく、添付文書に代表される規制文書の活用が不可欠であると認識した。もう 1 つは相互作用ユニットの考え方に基づく発展形である。本研究では相互作用ユニットを定義するために必要な医薬品グループのデータベース化を行ったが、実際に相互作用ユニットを網羅的にデータベース化したわけではない。また相互作用ユニットは医薬品が関与するものに限定され、疾患を理解するための相互作用ユニットは含まれていない。そこで今後の展開として、ゲノム、疾患、医薬品が関与する広い意味での相互作用ユニットをデータベース化し、そこからグローバルなネットワークの知見を得ることができるようなレファレンスリソースを構築していきたいと考えている。

§2. 研究開発実施体制

1. 研究グループ

(1) 研究代表者グループ

人員構成

氏名	所属機関	役職	研究開発項目	参加時期
金久 實	京都大学化学研究所	特任教授	総括、KEGG MEDICUS 開発	H26.4～ H29.3
古道 美穂	同上	特定研究員	KEGG MEDICUS 開発	H26.4～ H29.3
田辺 麻央	同上	特定研究員	KEGG MEDICUS 開発	H26.4～ H29.3
米納 朋子	同上	特定研究員	KEGG MEDICUS 開発	H26.4～ H29.3
神谷 知美	同上	特定研究員	KEGG MEDICUS 開発	H26.4～ H29.3
池内 志帆	同上	特定研究員	KEGG MEDICUS 開発	H26.4～ H29.3
緒方 博之	同上	教授	ゲノムネットとの連携	H26.4～ H29.3

担当項目

以下のすべての項目を担当した。

1. 疾患情報のデータベース
 - 1-1. KEGG DISEASE 疾患エントリ
 - 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. 有識者会議等

有識者会議等は設置していない。

§ 10.1 に記載した通り有識者を招いた個別のミーティングで対応した。

§3. 研究開発の目的、実施内容及び成果

1. 研究開発の背景

個人ゲノム解析、がんゲノム解析、病原体ゲノム解析、マイクロビオームのメタゲノム解析など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、バイオインフォマティクス研究の重要性がますます高まっている。本研究グループは、平成 23 年度から 25 年度まで 3 年間の統合化推進プログラム第 I 期において、疾患・医薬品・環境物質など社会的ニーズの高いデータを、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、最先端の研究と一般社会との架け橋となる統合データベース KEGG MEDICUS の開発を行った。この開発の背景として、2 つの意味でのトランスレーショナルバイオインフォマティクスの考え方があった。1 つは研究者コミュニティに医療・創薬等で研究成果の実用化を支援するためのリソースを、もう 1 つは一般社会に病気や薬に対する科学的理解を深めるためのリソースを提供するという考え方である。医薬品情報については添付文書という社会で使われている標準情報があり、これを独自に構築してきた医薬品の知識ベースと統合したことで、KEGG MEDICUS は一般社会からも広く利用されるようになり、月間訪問者数は 3 年間で 4 万から 20 万へと増加した。一方、疾患情報についてはそのような標準情報がないため統合化は不十分であった。そこで平成 26 年度から 28 年度までの本研究開発では、疾患情報と前者の意味でのトランスレーショナルバイオインフォマティクスに焦点をあて、すでに社会で使われている情報ではなく、今後使われるであろう情報との統合化、とくに上記の様々なタイプのゲノム情報の有効利用の観点から、ゲノムとフェノタイプ・疾患・医薬品の統合化を行うこととした。

特定の酵素遺伝子が欠損して生じる先天性代謝異常症などの単一遺伝子疾患は、メンデルの法則に従い、遺伝子型 (genotype) と表現型 (phenotype) の関係は単純である。しかし多くの疾患は多因子性であり、様々な遺伝要因と環境要因が複雑に関与している。我々はこれまで生体を分子ネットワークの相互作用システムとして記述し、疾患は分子ネットワークのゆらぎ状態であるとの観点から、各疾患ごとにゲノムのゆらぎ (先天的・後天的な遺伝子変異など) と環境のゆらぎ (化学物質等だけでなく病原菌や体内細菌叢を含む) を列挙する形で KEGG DISEASE データベースを構築してきた。また医薬品はゆらぎ状態にある分子ネットワークを補正するゆらぎ物質であるとの観点から、各医薬品の標的分子、薬物代謝酵素や薬物トランスポーター、それらの相互作用などを蓄積した KEGG DRUG データベースを構築してきた。OMIM をはじめとした他の疾患データベースや DrugBank をはじめとした他の医薬品データベースなどと比較して、KEGG DISEASE と KEGG DRUG はこのようなネットワークの考えに基づいていること、疾患・医薬品の両者を統合的に開発していることが大きな特色である。この独自のリソース、とくに疾患情報を量的にも質的にも向上させることは国際的にも大きな意義がある。

個々の遺伝子やタンパク質単位ではなく、ネットワーク単位に考えることで、実際にゲノムの情報と疾患との関連づけ、および医薬品の作用・副作用との関連づけがどの程度できるようになるかは今後の大きな課題である。KEGG では主に代謝機能について、必要最小限の機能単位を遺伝子セットで定義した KEGG MODULE データベースの開発が行われてきた。本研究開発では、このモジュールの概念を用いてゲノムとフェノタイプ、とくに病原体ゲノムから薬剤耐性を判定できるような遺伝子セットをモジュールとして定義する。また医薬品の作用や副作用についても、複数の医薬品や遺伝子からなる相互作用ユニットに着目する。例えば、がんの個別化医療におけるマーカー遺伝子と分子標的薬との組合せ、あるいは薬物代謝酵素の誘導・阻害といった相互作用のある医薬品の組合せなどである。ただし現時点では、相互作用ユニットはモジュールのようにゲノム情報に基づく判別を目的とするのではなく、ネットワーク記述のためのものである。KEGG で個々の遺伝子・タンパク質機能が KO (KEGG Orthology) として一般化されているように、個々の医薬品の作用・副作用を医薬品グループ (医薬品の機能オーソログ) として一般化することが重要であり、そのために KEGG DGROUP (Drug Group) データベースの開発を行う

こととした。

2. 研究開発対象のデータベース・ツール

(1) データベース

主要なもの

正式名称	略称	概要
KEGG MEDICUS		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ある。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となっている。

(2) ツール等

正式名称	略称	概要
Pathogen Checker		病原体ゲノムから薬剤耐性を推定するツール。ゲノムシークエンスデータに 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する。
医薬品相互作用チェック		与えられた医薬品リストの中に併用禁忌・併用注意に該当する相互作用があるかを判定するツール。医薬品添付文書に記載された相互作用をすべて抽出し、KEGG DRUG/DGROUP で標準化したデータセットを用いている。

※データベース、ツールの詳細は別紙参照。

3. 達成目標及び実施計画

(1) 当初の実施計画・達成目標

以下は当初の全体開発計画書に記述した研究開発内容である。

本研究開発では、KEGG MEDICUS を構成する疾患情報のデータベースと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つの開発項目を設定する。疾患情報のデータベースは KEGG DISEASE で、従来からの疾患エントリの作成と、新たに追加する病原体ゲノムのシグネチャーモジュールを2つの小項目として開発を行う。医薬品情報のデータベースでは KEGG DRUG（それを補完する KEGG ENVIRON を含む）と医薬品添付文書の統合と、新たに開発する医薬品グループデータベース KEGG DRGOU P を2つの小項目として開発を行う。

1. 疾患情報のデータベース

1-1. KEGG DISEASE 疾患エントリ

KEGG DISEASE データベースではこれまで、がんについては独自に文献調査を行い、がん以外の多因子性疾患や単一遺伝子疾患については NCBI から入手した当時の OMIM データベースを参考にし、感染症疾患については病原体ゲノム配列が既知のものについてエントリ化を行ってきた。本研究開発においては、これらのタイプの疾患について文献情報をもとに、より網羅的な集積を行う。病因遺伝子は未知でも何らかの

分子的記述ができる疾患や、ゲノム配列は決定されていないが病原体は既知の感染症疾患なども含めてデータベース化を行う。同時に ICD-10, MeSH その他の外部リソースとの対応づけを行う過程で、KEGG DISEASE の網羅性をチェックする。KEGG DISEASE 疾患エントリーは病因遺伝子のリスト、発がん物質その他の環境因子のリスト、病原体、診断マーカー、治療薬などのリストで構成されている。今回の開発では病原体の部分に以下に述べるシグネチャーモジュールを付加する。また代表的な治療薬についての情報を充実させ、KEGG DRUG との統合を進める。KEGG DISEASE 疾患エントリーの高品質化は3年間かけて行い、疾患エントリーの数値目標は2000とする。

1-2. 病原体ゲノムのシグネチャーモジュール

KEGG モジュールとはパスウェイ、構造コンプレックス、その他の機能的なユニットに対応する遺伝子セットである。実際には個別生物種の遺伝子情報をオーソログ化して、KEGG ORTHOLOGY (KO) グループに対応した K 番号のセットで表現されている。モジュールの中でゲノムからフェノタイプを推定するためのマーカーとなるような遺伝子セットをシグネチャーモジュールと呼んでいる。本研究開発でとりあげるのは病原性や薬剤耐性を特徴づけるシグネチャーモジュールである。モジュールとなりそうな遺伝子セットを文献から探し、すでに配列決定がされた病原体ゲノムでモジュールの有無を確認してデータベース化を行う。初年度に予備的なデータ入力を行って関連するシステムの整備を行い、2年目以降に本格的なデータ入力を行う。よく知られたモジュールデータを取り込んだ KEGG DISEASE エントリーは初年度中に公開する。

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

KEGG DRUG は日本、米国、欧州で使われている医薬品の有効成分を網羅的に集積したデータベースである。標的分子、薬物代謝酵素とトランスポーター、ゲノムバイオマーカーなどが関与する分子ネットワーク情報をはじめ様々な情報が含まれ、日本の医療用医薬品添付文書から抽出し標準化した相互作用情報と添加物情報も付随している。また日本と米国のすべての医療用・一般用医薬品添付文書、WHO の ATC 分類をはじめ様々な分類情報が、KEGG DRUG の D 番号識別子と対応づけることで統合されている。日本医薬情報センター (JAPIC) とのライセンス契約で提供している日本の添付文書情報は毎月1回更新され、同じタイミングで米国 FDA の添付文書情報も月1回更新している。これと連動して D 番号との対応づけ、相互作用情報等の更新等を行い、また日本や米国で新規医薬品が発売された場合には新規の D 番号エントリーを作成している。

このように KEGG DRUG は内容的に非常に充実したデータベースで、更新体制も確立しているが、不十分な部分として疾患情報とのつながりが残っている。本研究開発において KEGG DRUG で薬効グループの記述と疾患情報へのリンクを充実させ、KEGG DISEASE での代表的な治療薬情報の充実とともに、全体的に整合性のある内容にする。これは下記の医薬品グループデータベース作成がある程度進んだ2年目以降に行う。

2-2. 医薬品グループデータベース

KEGG DRUG にある相互作用のデータは、医療用医薬品添付文書に記載された併用禁忌・併用注意の相互作用の欄から抽出し標準化したものであるが、添付文書の記述は必ずしも十分ではない。これを補うことを主な目的として医薬品グループデータベース KEGG DGROUP を開発する。医薬品グループにはいくつかのタイプがある。最も小さな単位は基本構造は同じで塩や水和状態など化学構造の微妙な違いをまとめた

同一薬効のグループである。また化学構造は多少異なるが同じ標的分子をもつ類似薬効のグループや、同じような作用機序をもつグループがある。さらには CYP 代謝酵素の誘導・阻害を引き起こす医薬品グループや同じ CYP 代謝酵素の基質グループといったグループもある。本研究開発では以下の4つを主なソースとしてこのような医薬品グループを定義したデータベースを作成する。

- (1) ATC 分類最下層に対応した同一基本構造、同一薬効のグループ
- (2) 添付文書の相互作用欄に記載された同一相互作用のグループ
- (3) 米国の商品を分類した FDA EPC と VA NDF-RT にあるグループ
- (4) USP Dictionary に記載された医薬品名の suffix に対応したグループ

これとともに複数の医薬品間の相互作用の解析や、薬物代謝酵素の遺伝子多型と薬物代謝能力との関連解析などで、グループ情報を用いて既知データの一般化を行い、その有効性を確認する。開発スケジュールとしては(1)を3ヶ月以内に公開し、順次その他のグループを作成して2年間で完成させる。

3. 統合インターフェース

KEGG MEDICUS 医薬品情報の統合インターフェースとして、医薬品添付文書情報検索、医薬品相互作用チェック、医薬品成分検索はすでに医療従事者や一般の人々を含め広く利用されている。今後とも利用者からのフィードバックをもとに改良を行う。また医薬品相互作用チェックは KEGG DGROUP を利用する形で改良する。

KEGG MEDICUS 疾患情報の統合インターフェースについては、シグネチャーモジュール追加その他の KEGG DISEASE 高度化とともに新しい検索インターフェースを開発する。

ゲノムやメタゲノムの解釈を支援する KEGG Mapper については、シグネチャーモジュールをはじめとしたモジュール情報を利用するツールの改良と拡張を行う。また KEGG DISEASE で定義された遺伝子・分子リスト（相互作用ユニット）の情報を、ゲノムその他の大量データ解析で有効利用できるように、新たに Disease mapping ツールの開発を行う。

KEGG MEDICUS API では医薬品グループデータベースへの対応、これに基づく医薬品相互作用チェック機能などの改良を行い、データベース間のリンク情報については RDF 形式でも取得できるようにする。

(2) 期間中に追加・削除・変更した実施計画・達成目標

研究開発開始以降に変更を行ったのは以下の2点である。いずれも進捗とともに明らかになった内容を計画に追加した。

1-1. KEGG DISEASE 疾患エントリ

疾患エントリの作成については、OMIM に記載された論文やゲノム配列に関連したものも含め、文献情報をもとに行う計画であった。本研究開発2年を経過した段階で、医薬品添付文書に記載された適応症を網羅的に調べる作業を始めたところ、これまでの疾患の定義の仕方との不整合が明らかになった。論文ベースの疾患の定義と実際に治療の対象となる疾患の定義は必ずしも一致しないのである。KEGG DISEASE データベースをより実用的なものとするために、適応症との対応がつくよう全エントリの見直しを開始し、新規エントリの作成や既存エントリの修正を行うこととした。

1-2. 病原体ゲノムのシグネチャーモジュール

当初計画でも薬剤耐性に関与する遺伝子変異の重要性は認識していたが、そのデータをどのように扱うかについては課題として取り上げただけで、明確な計画は記載し

ていなかった。実際にベータラクタマーゼをはじめ様々な薬剤耐性遺伝子の変異アミノ酸配列を収集し解析したところ、遺伝子の系統関係と耐性を示す薬剤に関連があることが明らかとなった。そこでシグネチャーモジュールに加えてシグネチャーKOと呼ぶ表現形式で遺伝子変異グループの知識集約を行うこととした。

4. 実施内容

(1) 実施内容

本研究では KEGG MEDICUS を構成する疾患情報のデータベース、医薬品情報のデータベース、統合インターフェースについて、第 I 期に整備を行ったデータベース及び解析ツール等の拡張（開発項目 1-1, 2-1 及び 3）と、新規開発項目として病原体ゲノムのシグネチャーモジュール（1-2）と医薬品グループデータベース（2-2）の開発を行った。

1. 疾患情報のデータベース

1-1. KEGG DISEASE 疾患エン트리

本研究開始時の 2014 年 4 月に 1,359 であった KEGG DISEASE 疾患エントリの数は、2016 年 12 月末に 1,700 を越え（25%の増加）、本研究終了時には 1,800 程度になる見込みである。上記の通り添付文書に記載された適応症に基づく全エントリの見直しを終了すれば、当初計画で数値目標とした 2,000 エントリーで医薬品との対応づけに必要な疾患エント리는ほぼカバーできると考えている。ICD-10 国際疾病分類、MeSH の疾患分類は定期的にアップデートし、疾患エン트리との対応づけも更新している。また MEDIS（医療情報システム開発センター）が提供する標準病名との対応づけも行っている。なお疾患エントリの網羅性については、これらの分類との比較ではなく添付文書に記載された適応症との対応で行うこととした。医薬品と治療の対象となる疾患の対応を明確にし、より実用的な疾患データベース構築するためである。疾患エン트리作成は田辺を中心に行われた。

1-2. 病原体ゲノムのシグネチャーモジュール

病原体ゲノムから抗微生物薬に対する薬剤耐性を判別するために、初年度に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。そこで当初計画のシグネチャーモジュールに加えてシグネチャーKO の定義を行った。集積した変異アミノ酸配列の数、それに基づくシグネチャーKO の数、及びシグネチャーモジュールの数は表 1 の通りである。

表 1. 薬剤耐性のシグネチャーKO とシグネチャーモジュール

		Number of signatures	Number of sequences
Gene variants (Signature KOs)	beta-Lactamase genes	46	1,461
	Aminoglycoside resistance genes	39	149
	Tetracycline resistance genes	11	46
	Macrolide resistance genes	18	84
	Phenicol resistance genes	8	35
	Trimethoprim resistance genes	6	38
	Quinolone resistance genes	2	68
Gene sets (Signature modules)	beta-Lactam resistance modules	3	
	Vancomycin resistance modules	3	
	CAMP resistance modules	4	

	Multipdrug resistance modules	15	
--	-------------------------------	----	--

これらのデータは BRITE テーブルファイル br08451~br08457 にまとめられており、FASTA 形式のアミノ酸配列ファイル、それに基づく系統樹解析の結果とともに公開している^{3,5)}。本研究項目は金久の総括の下に、古道、神谷、池内が担当した。

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

KEGG DRUG は日米欧の医薬品を網羅したデータベースで、新薬等の新規エントリの追加とともに、既存エントリについて標的分子や薬物代謝に関するネットワーク情報の見直しと、効能の記述の標準化を行った。とくに添付文書をもとに適応となる疾患を、Activity フィールドの Disease サブフィールドとして追加し、医薬品と疾患の関連が明確になるようにしている。また別の観点での医薬品情報と疾患情報の統合のために、代表的な疾患群に対して治療薬をまとめた 25 種類の BRITE テーブルファイルを新たに作成し公開した。日本と米国の医薬品添付文書は毎月定期的に更新して KEGG DRUG との対応づけを行い、相互作用チェックのためのデータ更新も行っている。日本の医療用医薬品については商品一覧のページを充実させ、同一有効成分 (KEGG DRUG D 番号) かつ同一薬効分類番号をもつ商品について、先発品・後発品の区別、薬価、添加物、警告・禁忌、相互作用、適応症、適応菌種などを選択して比較できるようにした。医薬品添付文書と KEGG DRUG の統合は米納を中心に行われた。

2-2. 医薬品グループデータベース

医薬品グループデータベースは医薬品と標的分子の相互作用、医薬品と代謝酵素やトランスポーターとの相互作用、添付文書で併用禁忌・併用注意と記載された医薬品同士の相互作用などの記述を一般化し、添付文書の記載漏れや新たな医薬品の候補などにも適用できるようにすることを目的として開発した。医薬品グループは様々な医薬品ネットワークにおける機能オーソログという言い方もできる。当初計画に記載した通りの開発を行い、医薬品グループデータベースの内容は表 2 の通りとなっている。

表 2. 医薬品グループのデータ数

Group type	Number of groups
Chemical	1,576
Cyp/Transporter	24
Other	361

Chemical group とは塩や水和状態のみ異なる同一基本構造のグループで、多くは ATC 分類の最下層に対応する。Other には Structure group や Target group などが含まれ、必ずしも一意に定義できない場合もあるが、一般に何らかの相互作用ネットワークで同一ノードの意味をもたせている。医薬品グループは KEGG DGROUP データベースとして開発し、KEGG DRUG では Chemical group を Remark フィールドに、それ以外のグループを新たな Class フィールドに記載して対応づけがなされている。本研究項目は金久の総括の下に、米納らが担当した。

3. 統合インターフェース

KEGG MEDICUS 医薬品情報検索、疾患情報検索の改良とともに以下の開発を行った。

- ① 病原体ゲノムシーケンスデータから薬剤耐性を推定するために、新たに集積した変異アミノ酸配列データ、シグネチャーKO、シグネチャーモジュールを用いた

- Pathogen Checker ツールを開発した^{3,5)}。
- ② KEGG DGROUP 医薬品グループデータベースを組み込む形で、医薬品相互作用チェックツールの新しいバージョンを開発した。現時点ではこのチェックツールは添付文書情報のみに基づいているが、オプションで CYP 代謝酵素の多型による Poor metabolizer 等がとくに影響を受けると推定される医薬品や相互作用を表示するように改良中である。
 - ③ KEGG Mapper の新しいツールとして Join Brite Table を開発した。例えば、代表的な疾患群に対する治療薬のテーブルに KEGG DRUG のターゲット情報を追加して表示するといった利用ができる。
 - ④ KEGG Mapper の新しいツールとして、Search Disease を開発し、ヒト遺伝子 ID または KO を用いて、疾患遺伝子と関連パスウェイを調べることができるようにした。
 - ⑤ KEGG MEDICUS API の改良を行い、KEGG DRUG、ATC 分類、薬効分類、日本の医薬品コード (YJ コード)、米国の医薬品コード (NDC) などの関係を取得する link オペレーションで RDF 対応を行った。
- 本研究項目は金久の総括の下に、古道、神谷、池内が担当した。

(2) データベースの利便性に関する利用者ニーズと具体的な対応

§ 10.1 に記載した通り主に医薬品情報について医療関係者とくに薬剤師の方々のニーズを収集した。個々のニーズに細かく対応したわけではないが、KEGG MEDICUS の特徴を生かす形で全体的なニーズに対応した。例えば医薬品の商品一覧ページや日本・米国・欧州での新薬承認状況の一覧など、他のサイトでは得られない独自の情報を提供できるようにした。

(3) 持続的なデータベース運用体制の構築に向けた取り組み

持続的なデータベース運用体制の構築は国際的な課題である。米国の医薬品データベースである Drug Central の諮問委員会や、Human Frontier Science Program Organization (HFSP) が世界の主要なデータベースを招いて開催した International Workshop: Data Resources in the Life Sciences に参加し sustainability の議論を行った。とくに HFSP のワークショップでは Funding agencies の国際的な連携で Core databases (KEGG も含まれる) のサポート体制作りを目指すこととなり、Ensuring global support of core data resources for the life sciences と題した報告を Nature 誌に投稿予定とのことである。

(4) 統合化推進プログラムの他のチームや DBCLS との連携

糖鎖研究チームとは国内外の学会（国内招待講演 2、国際招待講演 4）や単糖のシンボル表現 (Varki, A., et al., “Symbol nomenclature for graphical representations of glycans” Glycobiology vol. 25, no. 12, pp. 1323-1324, 2015) などで連携している。

(5) データ産出を行う研究組織や研究室、プロジェクトとの連携

データ産出を行う研究組織や研究室、プロジェクト等との直接的な連携はないが、データ産出の次のステップとして必須のデータ解釈に、KEGG MEDICUS は広く利用されている。

(6) 人材の育成

とくになし。

(7) その他

KEGG プロジェクト支援のために設立された NPO 法人バイオインフォマティクス・ジャパンが NPO 法人システム薬学研究機構との共催で毎年開催している「新たな創薬パラダイムの創出ー日本発の創薬技術・新薬のシンポジウムー」により、医薬品業界、薬学研究コミュニティとの連携ができています。

§4. 主要なデータベースの利活用状況

1. アクセス数

(1) 実績

名称	種別	平成 25(2013) 年度	平成 26(2014) 年度	平成 27(2015) 年度	平成 28(2016) 年度
KEGG MEDICUS (年間合計値)	訪問者数	1,672,656	2,324,456	2,545,418	1,496,835
	訪問数	3,873,973	6,180,045	6,734,191	3,487,850
	ページ数	6,983,200	10,866,271	12,599,680	7,734,384

表 1-1 研究開発対象の主要なデータベースの利用状況(年度別)

名称	種別	平成 25(2013) 年度	平成 26(2014) 年度	平成 27(2015) 年度	平成 28(2016) 年度
KEGG MEDICUS (月間平均値)	訪問者数	201,586	295,216	324,072	231,366
	訪問数	322,831	515,004	561,183	387,539
	ページ数	581,933	905,523	1,049,973	859,376

表 2-2 研究開発対象の主要なデータベースの利用状況(月間平均)

(注)年間合計値の訪問者数は年間を通してユニークな IP アドレス数

平成 28 年度の年間合計値は 4 月から 12 月まで

(2) 分析

- 平成 26, 27 年度と比較して平成 28 年度は（とくに 7 月以降に）アクセスが大幅に減少した。これは Google で医薬品の検索をすると、以前のように KEGG の添付文書のページが上位には出なくなったためである。KEGG MEDICUS 検索で従来からの添付文書ページに加えて新たに作成した商品単位のページへのリンクを入れたことが要因かもしれない。ただ Google のランキングのやり方は常に変化している印象があるので、実際のところはよく分からない。
- 別の見方をすると、Google でたまたま上位にあったページをクリックしただけのアクセスがなくなったわけで、より積極的に KEGG MEDICUS を利用しているユーザ数が月間 20 万はいることが分かった。

2. データベースを利用して得られた研究成果事例

- Hiroaki Iwata, Ryusuke Sawada, Sayaka Mizutani, Masaaki Kotera and Yoshihiro Yamanishi, "Large-scale prediction of beneficial drug combinations using drug efficacy and target profiles", *Journal of Chemical Information and Modeling*, vol. 28, No. 12, pp. 2705-2716, 2015 (DOI: 10.1021/acs.jcim.5b00444).
- Warren A. Kibbe, Cesar Arze, Victor Felix, Elvira Mitraka, Evan Bolton, Gang Fu, Christopher J. Mungall, Janos X. Binder, James Malone, Drashti Vasant, Helen Parkinson, and Lynn M. Schriml, "Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data", *Nucleic Acids Research*, vol. 43, No. D1, pp. D1071-D1078, 2015 (DOI: 10.1093/nar/gku1011).
- Oleg Ursu, Jayme Holmes, Jeffrey Knockel, Cristian G. Bologna, Jeremy J. Yang, Stephen L. Mathias, Stuart J. Nelson, and Tudor I. Oprea, "DrugCentral: online drug compendium", *Nucleic Acids Research, Journal of Chemical Information and Modeling*, vol. 45, No. D1, pp. D932-D939, 2017 (DOI: 10.1093/nar/gkw993).

3. その他

KEGG の論文 (NAR DB issues) を引用している論文、すなわち KEGG データベースを利用したと思われる論文を Web of Science で調査したところ 2016 年は 2078 件、2015 年は 1917 件、2014 は 1602 件で、3 年間合計 5597 件のヒットがあった。そのうち Nature 誌に出た論文は以下の 19 件であった。医療に関連したものも多い。

1. Bulut-Karslioglu A, et al. (2016) Inhibition of mTOR induces a paused pluripotent state. *Nature* 19 (7631) 119-123 (DOI: 10.1038/nature20578)
2. Adam M, et al. (2016) Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* 538 (7625) 336-343 (DOI: 10.1038/nature19840)
3. Paez-Espino D, et al. (2016) Uncovering Earth's virome. *Nature* 536 (7617) 425-430 (DOI: 10.1038/nature19094)
4. Pedersen HK, et al. (2016) Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* 535 (7612) 376-381 (DOI: 10.1038/nature18646)
5. Brito IL, et al. (2016) Mobile genes in the human microbiome are structured from global to individual scales. *Nature* 535 (7612) 435-439 (DOI: 10.1038/nature18927)
6. Hamann E, et al. (2016) Environmental Breviatea harbour mutualistic *Arcobacter* epibionts. *Nature* 534 (7606) 254-258 (DOI: 10.1038/nature18297)
7. Westermann AJ, et al. (2016) Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* 529 (7587) 496-501 (DOI: 10.1038/nature16547)
8. Bai Y, et al. (2015) Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* 528 (7582) 364-369 (DOI: 10.1038/nature16192)
9. Forslund K, et al. (2015) Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* 528 (7581) 262-266 (DOI: 10.1038/nature15766)
10. Brown CT, et al. (2015) Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523 (7559), 208-211 (DOI: 10.1038/nature14486)
11. Shungin D, et al. (2015) New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518 (7538), 187-196 (DOI: 10.1038/nature14132)
12. Plata G, et al. (2015) Long-term phenotypic evolution of bacteria. *Nature* 517 (7534), 369-372 (DOI: 10.1038/nature13827)
13. Hsiao A, et al. (2014) Members of the human gut microbiota involved in recovery from *Vibrio cholerae* infection. *Nature* 515 (7527), 423-426 (DOI: 10.1038/nature13738)
14. Pickard JM, et al. (2014) Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. *Nature* 514 (7524), 638-641 (DOI: 10.1038/nature13823)
15. Oh J, et al. (2014) Biogeography and individuality shape function in the human skin metagenome. *Nature* 514 (7520), 59-64 (DOI: 10.1038/nature13786)
16. Rodgers JT, et al. (2014) mTORC1 controls the adaptive transition of quiescent stem cells from G0 to G(Alert). *Nature* 510 (7505), 393-396 (DOI: 10.1038/nature13255)
17. Moroz LL, et al. (2014) The ctenophore genome and the evolutionary origins of neural systems. *Nature* 510 (7503), 109-114 (DOI: 10.1038/nature13400)
18. Dohm JC, et al. (2014) The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* 505 (7484), 546-549 (DOI: 10.1038/nature12817)
19. David LA, et al. (2014) Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505 (7484), 559-563 (DOI: 10.1038/nature12820)

§5. 研究開発期間中に得られた科学・技術や産業に対する波及効果

抗微生物薬（抗菌薬、抗ウイルス薬、抗真菌薬、抗寄生虫を含む）に対して微生物がもつ耐性 Antimicrobial resistance (AMR) は古くからの課題である。2013年にCDCのレポート、2014年にホワイトハウスやWHOのレポートが出され、とくに感染症をコントロールできなければ先端医療も実現できないとの観点から、新たな研究開発が進められるようになってきている。その中ではゲノムシーケンシングの重要性も取り上げられている。本研究計画はこれらのレポートを知らずに立案したものであったが、§8に記載した通りAMRに関する知識ベースを構築したことで、AMRの研究分野に貢献し、産業への波及効果も期待できると考えている。

§6. 今後の展開

本研究開発では、疾患情報としてKEGG DISEASEの大幅な拡張、医薬品情報としてKEGG DRUGの改良とKEDD DGROUPの新規開発を行い、疾患と遺伝子の関連、及び医薬品と標的・代謝遺伝子との関連などを蓄積した。これらの関連データは基本的に2つのノードからなるネットワーク要素（2項関係）である。本研究ではさらにモジュール（遺伝子セット）や相互作用ユニット（遺伝子以外も含むセット）といった形で、2つ以上のノードからなるネットワーク要素もとりあげた。例えば薬剤耐性のシグネチャーモジュールや、医薬品・適応疾患・マーカー遺伝子のユニットである。本研究の次の展開として、これらネットワーク要素を組み合わせることで、ゲノム・疾患・医薬品に関する統合的な知見を得るためのリソースを開発することができると考えている。そのためKEGGのパスウェイマップから疾患・医薬品に関する生体システムのネットワーク要素を（恐らくモジュールのような形式で）定義すること、ネットワーク要素のノードとして遺伝子・ゲノムのvariation (mutation や polymorphism など) を考慮できるようにすることが必要である。手作業で定義されたネットワーク要素の集合を新たなレファレンスデータベースとして提供することで、ゲノムをはじめとした様々なデータの解釈に利用できるようになり、幅広い研究分野や産業に貢献できるだろう。

§7. 自己評価

疾患エン트리数として当初の 1,359 を 2,000 にするという数値目標には 200 ほど足りていない。疾患の数は細かく分ければいくらかでも増やすことは可能であるが、分子メカニズムの観点から、例えば遺伝子変異の細かなタイプまで区別するようなことはせず、また医薬品添付文書の適応症との対応関係のために、できるだけ適切な疾患エントリを作成してきた。数値的には不十分であったが、内容的には疾患エントリは大幅に拡充されている。また疾患エントリ以外のすべての項目では当初計画以上の成果があった。

§6にも記載した通り **Antimicrobial resistance (AMR)** のリソース作りは本研究の目的の1つでもあった。そのためにベータラクタマーゼをはじめ様々な薬剤耐性遺伝子のアミノ酸配列データを文献情報をもとに蓄積した。薬剤耐性遺伝子のデータベースとして最も有用であった **Lahey Clinic** のベータラクタマーゼデータベースは維持が困難となり **NCBI** に移管されたが、従来のような **curation** は行われなくなった。他にも更新が途絶えた **ARDB** や、今でも更新されている **CARD**, **Resfams** などもあるが、データが散在している。本研究では抗微生物薬の医薬品情報、適応となる疾患情報とともに、これらに対する薬剤耐性遺伝子のアミノ酸配列、生物種情報、そして薬剤耐性のシグネチャーKOとシグネチャーモジュールを含む統合的なリソースを開発し、**KEGG Pathogen** ページで **Pathogen Checker** ツールとともに公開した。また関連する **BRITE** テーブルファイルや **FASTA** アミノ酸配列は **KEGG MEDICUS FTP** サイト（および **NBDC** のアーカイブ）でも提供し、誰でも自由に活用できるようになっている。

§8. 外部発表等

1. 原著論文発表

(1) 論文数概要

種別	国内外	件数
発行済論文	国内 (和文)	0 件
	国際 (欧文)	5 件
未発行論文 (accepted, in press 等)	国内 (和文)	0 件
	国際 (欧文)	2 件

(2) 論文詳細情報

1. Kiyoko F. Aoki-Kinoshita and Minoru Kanehisa, "Glycomic analysis using KEGG GLYCAN", In "Glycoinformatics" (Luetke, T. and Frank, M., eds.), Humana Press, Methods in Molecular Biology, vol. 1273, pp. 97-107, 2015 (DOI: 10.1007/978-1-4939-2343-4_7)
2. Minoru Kanehisa, "KEGG bioinformatics resource for plant genomics and metabolomics", In "Plant Bioinformatics: Methods and Protocols" (Edwards, D., ed.), Humana Press, Methods in Molecular Biology, vol. 1374, pp. 55-70, 2016 (DOI: 10.1007/978-1-4939-3167-5_3)
3. Minoru Kanehisa, Yoko Sato, Masayuki Kawashima, Miho Furumichi and Mao Tanabe, "KEGG as a reference resource for gene and protein annotation", Nucleic Acids Research, vol. 44, no. D1, pp. D457-D462, 2016 (DOI: 10.1093/nar/gkv1070).
4. Minoru Kanehisa, Yoko Sato and Kanae Morishima, "BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences", Journal of Molecular Biology, vol. 428, no. 4, pp. 726-731, 2016 (DOI: 10.1016/j.jmb.2015.11.006)
5. Minoru Kanehisa, Miho Furumichi, Mao Tanabe Yoko Sato and Kanae Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs", Nucleic Acids Research, vol. 45, no. D1, pp. D353-D361, 2017 (DOI: 10.1093/nar/gkw1092)
6. Minoru Kanehisa, "KEGG GLYCAN", In "A Practical Guide to Using Glycomics Databases" (Aoki-Kinoshita, K.F., ed.), Springer (in press)
7. Minoru Kanehisa, "Enzyme annotation and metabolic reconstruction using KEGG", In "Protein Function Prediction: Methods and Protocols" (Kihara, D., ed.), Humana Press, Methods in Molecular Biology (in press)

2. その他の著作物 (総説、書籍など)

1. Kanehisa, M.; KEGG resource for deciphering the genome. Kyoto University Research Activities Vol. 5, No. 2, p. 6, September 2015.
2. Kanehisa, M.; KEGG MEDICUS drug information service. Kyoto University Research Activities Vol. 5, No. 2, p. 17, September 2015.

3. 国際学会発表及び主要な国内学会発表

(1) 概要

種別	国内外	件数
招待講演	国内	4 件
	国際	4 件
口頭発表	国内	2 件
	国際	1 件
ポスター発表	国内	6 件

	国際	0件
--	----	----

(2) 招待講演

〈国内〉

1. 金久 實、ゲノムと生命システムをつなぐ KEGG の分子ネットワーク、千葉工大フォーラム「生命システム原材料の起源と進化」、津田沼、2015年6月19日
2. 金久 實、病原性・薬剤耐性と糖鎖、第34回日本糖質学会年会糖鎖情報学セッション、東京、2015年8月1日
3. 金久 實、KEGG 分子ネットワークによる生命システム再構築、第38回日本分子生物学会年会ワークショップ、神戸、2015年12月3日
4. 金久 實、A database of drug interaction networks for assessing drug safety、第10回スーパーコンピュータ「京」と創薬・医療の産学連携セミナー、大阪、2016年1月22日

〈国際〉

1. Minoru Kanehisa, Automated interpretation of metabolic capacity from complete genomes using KEGG modules, Plant and Animal Genome XXIII Conference, San Diego, USA, 2015/1/11
2. Minoru Kanehisa, Computational assessment of antimicrobial resistance from pathogen genome sequences for managing infectious complications, Computation as a driver of translational medicine (5th Edition of the International School of Functional Genomics), Catania, Italy, 2015/9/28
3. Minoru Kanehisa, Linking genomes to drug interaction networks, International Conference on Pharmaceutical Bioinformatics (ICPB 2016), Pattaya, Thailand, 2016/1/25
4. Minoru Kanehisa, Chemical classifications for biology and medicine, 251st American Chemical Society National Meeting Symposium on “Chemistry, Data & the Semantic Web: An Important Triple to Advance Science”, San Diego, USA, 2016/3/15

(3) 口頭講演

〈国内〉

1. 金久 實、ゲノムと疾患・医薬品をつなぐ相互作用ユニットの知識ベース、トーゴの日シンポジウム2014、東京、2014年10月5日
2. 金久 實、ゲノム・疾患・医薬品のネットワーク解析、トーゴの日シンポジウム2016、東京、2016年10月5日

〈国際〉

1. Minoru Kanehisa, Kyoto Encyclopedia of Genes and Genome, International Workshop on Data Resources in the Life Sciences, Human Frontier Science Program Organization, Strasbourg, France, 2016/11/18

(4) ポスター発表

〈国内〉

1. 田辺 麻央、矢野 亜津子、金久 實、病原体ゲノムの薬剤耐性シグネチャーモジュール、トーゴの日シンポジウム2014、東京、2014年10月5日
2. 米納 朋子、古道 美穂、石黒 真理、永吉 ふみ、金久 實、医薬品相互作用ネットワーク解析のための医薬品グループ、トーゴの日シンポジウム2014、東京、2014年10月5日
3. 田辺 麻央、矢野 亜津子、森嶋 佳苗、金久 實、薬剤耐性のシグネチャーKO とシグネチャーモジュール、トーゴの日シンポジウム2015、東京、2015年10月5-6日

4. 米納 朋子、石黒 真理、永吉 ふみ、金久 實、医薬品階層に基づく医薬品相互作用ネットワーク解析、トーゴの日シンポジウム 2015、東京、2015 年 10 月 5-6 日
5. 田辺 麻央、金久 實、抗微生物薬と抗悪性腫瘍薬に対する薬剤耐性メカニズムの比較解析、トーゴの日シンポジウム 2016、東京、2016 年 10 月 5-6
6. 金 昭、金久 實、ヒト疾患に関するウイルスゲノムのデータベース、トーゴの日シンポジウム 2016、東京、2016 年 10 月 5-6

〈国際〉

なし

4. 知財出願

なし

5. 受賞・報道等

(1) 受賞

なし

(2) メディア報道

なし

(3) その他

1. Thomson Reuters Highly Cited Researchers 2014、金久 實
2. Thomson Reuters Highly Cited Researchers 2015、金久 實

§9. 研究開発期間中の活動

1. 進捗ミーティング

年月日	名称	場所	参加人数	目的・概要
2014年 8月28日	医薬品研究会	東大医科研 京大化研	19人	遺伝子多型と医薬品応答 について、NIHSの専門家を まじえて議論
2015年 8月25日	医薬品研究会	東大医科研 京大化研	17人	病院における薬剤情報システ ムを知るために病院薬剤師の 方より講義
2015年 12月24日	医薬品相互作用検索機 能検討会	東大医科研	5人	上記に基づき医薬品相互作 用検索機能の改良を検討
2016年 2月23日	添付文書検索システム 機能検討会	東大医科研	5人	上記に基づき添付文書検索 システムの改良を検討
2016年 12月6日	医薬品研究会	東大医科研 京大化研	23人	元厚労省の方より医薬品の申 請から承認までの講義

2. 主催したワークショップ、シンポジウム、アウトリーチ活動等

年月日	名称	場所	参加人数	目的・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
2015年 10月14～16日	KEGG MEDICUS の展示・デモ	パシフィコ 横浜	50人	BioJapan2015での展示
2015年 12月1～3日	KEGG MEDICUS の展示・デモ	神戸国際展 示場	30人	日本分子生物学会年会での 展示
2016年 11月30日～12 月2日	KEGG MEDICUS の展示・デモ	パシフィコ 横浜	150人	日本分子生物学会年会での 展示

以上

別紙1 研究開発対象のデータベース等

No.	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系データベースアーカイブ	NBDCヒトデータベース	NBDC RDFポータル	関連文献 (論文リストに記載があれば、その番号でも可)
1	KEGG MEDICUS		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	継続・発展	データベース等	提供済	対象外		
2	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.html	2015/4/1	新規	ツール等	対象外	対象外	対象外	
3	医薬品相互作用チェック		与えられた医薬品リストの中に併用禁忌・併用注意に該当する相互作用があるかを判定するツールです。医薬品添付文書に記載された相互作用をすべて抽出し、KEGG DRUG/DGROUPで標準化したデータセットを用いています。	http://www.kegg.jp/medicus-bin/ddi_manager	2016/4/1	新規	ツール等	対象外	対象外	対象外	