

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)

研究開発中間報告書

「ヒトゲノム・病原体ゲノムと疾患・医薬品を
つなぐ統合データベース」

金久 實

京都大学 化学研究所 特任教授



©2023 金久 實(京都大学) Licensed under CC BY 4.0

目次

§1. 研究開発実施内容の概要	4
§2. 研究開発実施体制	5
1. 各グループの担当項目	5
(1) 金久グループ	5
2. 有識者会議等	5
(1) 会議概要	5
(2) 委員一覧	5
(3) 開催歴	5
§3. 研究開発の実施計画、実施内容及び成果	6
1. 研究開発対象のデータベース・ツール	6
(1) データベース	6
(2) ツール等	6
2. 中間評価時までの達成目標及び達成状況	7
(1) 当初計画における中間評価時までの達成目標と達成状況	7
(2) 期間中に追加・削除・変更した実施計画及び達成目標	7
3. 実施内容	8
(1) 中間評価時までに実施した研究開発の内容	8
(2) 若手研究者の多様なキャリアパスの支援	13
§4. 主要なデータベースの利活用状況	14
1. アクセス数	14
(1) 実績	14
(2) 分析	14
2. データベースを利用して得られた研究成果・産業応用の例	14
3. その他	15
§5. 今後の研究実施計画	16
§6. 自己評価	17
§7. 外部発表等	18
1. 原著論文発表	18
(1) 論文数概要	18
(2) 論文詳細情報	18
2. その他の著作物（総説、書籍など）	18

3. 国際学会発表及び主要な国内学会発表.....	19
(1) 概要.....	19
(2) 招待講演	19
(3) 口頭講演	19
(4) ポスター発表	19
4. 知財出願	20
(1) 出願件数	20
(2) 一覧.....	20
5. 受賞・報道等.....	20
(1) 受賞.....	20
(2) メディア報道	20
(3) その他.....	20
§8. 研究開発期間中の活動.....	21
1. 進捗ミーティング	21
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	21

§1. 研究開発実施内容の概要

KEGG MEDICUS はネットワーク情報、疾患情報、医薬品情報を統合したデータベースで、研究者に対してはゲノムの情報から創薬・医療等の社会的価値を見いだすための情報基盤を、医療従事者や一般の人々に対しては医薬品添付文書を入口として疾患・医薬品に関する科学的知識を提供している。本研究開発では「ゲノム」としてこれまでのようにヒトゲノムだけではなく、ウイルスその他の病原体ゲノムも含め、ゲノムと社会をつなぐリソースとしてさらなる高度化や機能拡張を目指している。前期の統合化推進プログラムで開発を始めた **NETWORK** データベースでは、疾患は生体機能を司る分子ネットワークのゆらぎ状態であるとの観点から、ヒト遺伝子バリエーション、ウイルスその他の病原体タンパク質、化学物質などが、シグナル伝達や代謝をはじめとした生体内分子ネットワークにどのようなゆらぎを与え、疾患と関連しているかの知識集約を行ってきた。前期ではがんや神経変性疾患など限られた数の疾患パスウェイマップにバリエーションの詳細情報を付与する観点での開発が中心であったが、今期は通常のパスウェイマップにある様々な分子間相互作用・反応ネットワークについて疾患との対応を調べ、**DISEASE** データベースにネットワーク情報を付与する観点での開発を行っている。ネットワーク情報が付与された疾患エントリの割合は研究開発開始時点での 7%から現在は 30%近くにまで増加した。また医薬品とネットワークの関連については、前期はバリエーションを標的またはマーカーにする場合に限定していたが、今期はネットワーク上で関連があれば通常遺伝子の場合でも医薬品・標的関係を定義し、ネットワークバリエーションマップに表示できるようにした。さらにネットワークデータベースのサブ項目として開発しているウイルスタンパク質のオーソロググループ (VOG) については当初計画はすべて達成し、次のステップの新たな展開があった。VOG はもともと KEGG にある 68 万個のウイルスタンパク質から機能の手がかりを得るために、計算手法でオーソロググループを定義したものであるが、KEGG 生物種全体 5000 万個のタンパク質でもこれらグループに属するものを調べたデータセットを定期的に作成している。そして遺伝子の並びを比較するアルゴリズムを用いて、ウイルスと宿主あるいはその他の生物種で共通に存在する遺伝子クラスターを調べることができるようにした。

§2. 研究開発実施体制

1. 各グループの担当項目

(1) 金久グループ

金久グループ(研究代表者グループ)がすべての研究開発を実施する。

2. 有識者会議等

(1) 会議概要

有識者会議等は設置していない。

(2) 委員一覧

(3) 開催歴

§3. 研究開発の実施計画、実施内容及び成果

1. 研究開発対象のデータベース・ツール

(1) データベース

① 主要なもの

正式名称	略称	概要
KEGG MEDICUS		ゲノムの情報と疾患・医薬品との関連を、生体システムを構成する分子ネットワークを通して統合的に理解し、ヒトゲノム情報および病原体ゲノム情報の有効利用を促進するためのリソースである。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である。

② 上記以外のもの

正式名称	略称	概要

(2) ツール等

正式名称	略称	概要
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。

2. 中間評価時までの達成目標及び達成状況

(1) 当初計画における中間評価時までの達成目標と達成状況

達成目標	達成状況
・DISEASE データベースの GENE フィールドにある疾患遺伝子について、分子ネットワークの記述ができるかを調査し、可能なもののネットワーク要素を定義する。第 3 年次末までに作成する新規ネットワーク要素の数は 600 を目標とする。	第 3 年次 8 月末時点で、ネットワークバリエーションマップと対応づけた (分子ネットワークの記述ができた) DISEASE エントリ の数は研究開発開始時点の 175 から 819 に増加し、割合としても 7% から 30% 近くに増加した。また新規に定義したネットワーク要素の数はバリエーションのみの要素も含め 1,300 ほどであった。
・DISEASE データベースの構成と内容を全体的に見直し、NETWORK データベースと DRUG データベースとの統合利用を前提とした形に再編成する。この作業を第 3 年次までに完了する。	DISEASE データベースの構成の見直しは初年度に行い、NETWORK データベースで入力した関連疾患情報と、DRUG データベースで入力した適応症情報を、それぞれ逆引きとして DISEASE データベースの分子ネットワーク情報と医薬品情報に表示する形に再編成した。この作業は第 2 年次に終了した。
・ウイルスタンパク質とヒタンパク質の相互作用について、ヒト生体内システムの観点からデータ収集を行う。第 3 年次末までにヒトの免疫系に関するデータをデータベース化する。	主に自然免疫系を中心にヒトの免疫系に関するネットワークバリエーションマップを 8 個作成し、ウイルスその他の病原体およびヒト遺伝子バリエーションが関与する疾患が現在 134 個データベース化されている。
・配列類似性と系統関係の計算処理で作成したウイルスタンパク質のオーソログクラスターを研究開始後 1 年をめどに公開し、第 3 年次末までには検索・閲覧ツールとともに改良版を公開する。	ウイルスオーソログクラスターはウイルスオーソロググループ (VOG) と名称変更し、初年度にキーワード検索ツールとともに公開、第 2 年度には VOG クラスターと呼ぶ遺伝子の並びをウイルス間で調べるツールを、第 3 年度にはこれをウイルスと生物種の間で調べるツールを公開した。
・医薬品添付文書サービスに関して、独自に提供している商品一覧 (プログラム名 similar_product) の利用件数を、現在の月間平均 17 万件から 50% 増加させる。	KEGG MEDICUS 全体の利用者増の影響も大きいですが、現在の商品一覧の利用件数は 150% 増 (2.5 倍) に達している。

(2) 期間中に追加・削除・変更した実施計画及び達成目標

とくにありません。

3. 実施内容

(1) 中間評価時まで実施した研究開発の内容

(1) ネットワークデータベースの開発

NETWORK データベースはネットワーク要素(N 番号エン트리)とネットワークバリエーションマップ(nt 番号エン트리)から構成される。ネットワーク要素はシグナル伝達や代謝などに関与する一次元的な分子のつながりで、レファレンスとなる通常のネットワーク要素とゆらいだネットワーク要素があり、後者はさらにヒト遺伝子バリエント、ウイルスその他の病原体遺伝子、環境因子などゆらぎの種類で区別される。図 1 に示した通り、ネットワークバリエーションマップは通常のネットワーク要素(緑)の下にゆらいだネットワーク要素(ヒト遺伝子バリエントは赤)をアライメント表示したもので、どのようなゆらぎがどのような疾患に関連しているか病名が示されている。また医薬品と標的の関係(青)も表示できるようになっている。通常のネットワーク要素は PATHWAY データベースと関連づけられ、パスウェイマップ上でその位置がハイライト表示されるようになっている。

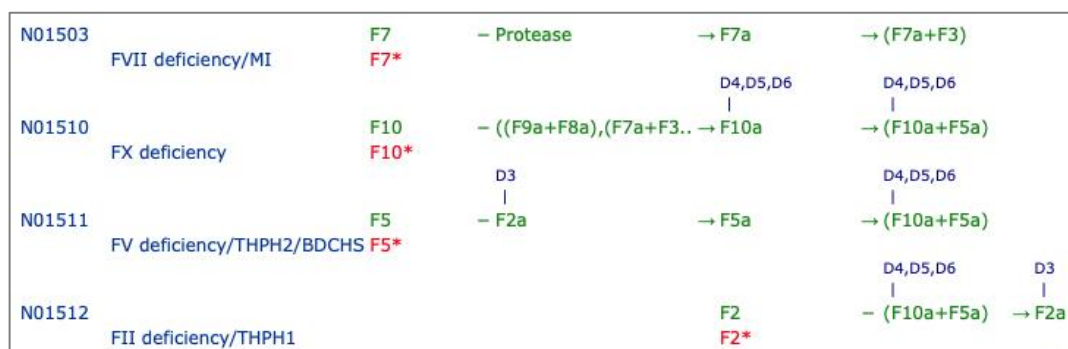


図1. ネットワークバリエーションマップ nt06514 Coagulation cascade の一部

初年度に行ったネットワーク要素表現法の単純化により、遺伝子バリエントの下流は loss of function では省略、gain of function では上向きの矢印表示だけにした。すでにレファレンスネットワークに記載されているものとの重複をなくしたわけである。このように遺伝子バリエントのみから構成されるネットワーク要素には N 番号は表示されず、VARIANT エン트리としての情報だけ提供している。

表 1. KEGG MEDICUS データ数の推移

	2022/4/1	2023/4/1	2024/4/1	2024/9/1
KEGG NETWORK (N)	1,408	1,349+702	1,396+1,099	1,408+1,334
(nt)	133	151	140	141
KEGG VARIANT	456	928	1,328	1,457
KEGG DISEASE (H)	2,551	2,627	2,701	2,792
(nt linked)	175	417	739	819
KEGG DRUG (D)	11,873	12,101	12,368	12,484
KEGG DGROU (DG)	2,384	2,426	2,462	2,481

(N) – Network element

(nt) – Network variation map

表1は KEGG MEDICUS データ数の推移である。ネットワーク要素数で+をつけた数は遺伝子バリエーションのみで構成される要素数で、当初計画ではこの数も含めていたのでここに記載した。ネットワークバリエーションマップ一覧のページ(KEGG NETWORKのトップページ)では、右側の Disease view に疾患パスウェイマップを出発点とした前期開発分、左側の Pathway view に疾患エン트리を出発点とした今期開発分をまとめている。表1で nt 番号エントリが 151 から 140 に減少しているのは、前期開発分にあった重複を除き今期開発分にまとめたため、実際には今期開発分のバリエーションマップは常に改良や拡張がなされている。

当初のネットワークバリエーションマップはネットワーク要素を構成する遺伝子シンボルや矢印などの記号を画像化してアライメント表示を行っていたが、アライメントデータがほしいという要望があったこともあり、現在はテキスト情報を html テーブルの形で表示している。また医薬品のネットワーク要素について、当初はヒト遺伝子バリエーションを標的またはマーカーとするものに限定して NETWORK データベースの中でネットワーク要素の定義を行っていたが、第3年目以降はこの制限をなくし DRUG データベースの NETWORK サブフィールドで管理している。さらに細かな改良点として complex formation の表記法を取り入れることにより、例えば細胞周期などでの複合体の構造変化による制御ネットワークも表現できるようにした。

(1a) ウイルスタンパク質のオーソロググループの開発

本研究では生体内の分子ネットワークに対するゆらぎ物質の1つとして重要なウイルスタンパク質について、その全体像を把握するため計算手法によるオーソロググループ VOG (Virus Ortholog Group)の開発を行っている。ウイルス遺伝子やウイルスタンパク質に関して実験データに基づく知識は非常に限られているため、手作業で定義している KO(KEGG Orthology)の機能情報を補完すること、また KO のために開発されている様々ツール群を VOG でも活用できるようにすることが目的である。なお VOG は当初計画では VOC(Virus Ortholog Cluster)と呼んでいたが、クラスターは遺伝子クラスターのように染色体上で保存された遺伝子の並びとして使っており、KO クラスターや VOG クラスターを考えるためにも、この名称変更を行った。

KEGG では隔月の NCBI RefSeq リリースからウイルスゲノムのデータを取得し、GENES データベースの vgenes (vg) データセットを作成している。GENES データベースでは SSEARCH プログラムで全対全のアミノ酸配列比較をし、各生物種の各遺伝子ごとに他生物種でのベストヒット配列リスト(GFIT データと呼んでいる)を作成して KEGG の内部アノテーションで利用している。本研究では約 14,000 のウイルスゲノムに 68 万のタンパク質が含まれている vgenes に限定して GFIT データからクラスタリングを行った。配列類似性の尺度としては、アライメントされた部分 Overlap のアミノ酸一致度 Identity を、比較する配列長の違いで補正した

$$\text{Modified identity} = \text{Identity} * \min(1, \text{Overlap} * 2 / (\text{Length1} + \text{Length2}))$$

を用いている。初年度には VOG 作成法を確立し、第2年度以降は類似度の閾値 30%、50%、70%の3種類のデータを公開し、RefSeq の更新にあわせて定期的にデータ更新を行っている(表2参照)。なおこの研究開発項目はすべて金久が担当した。

表 2. RefSeq Release 225 (July 2024)から生成した VOG (Virus ortholog group)

配列類似度の閾値	30%	50%	70%
グループの数	50,667	76,303	87,377
グループに含まれるタンパク質数	605,465	550,999	494,665
ウイルスタンパク質の総数	676,394		

(2) 疾患データベースの開発

DISEASE データベースの高品質化は本研究計画の中核となる部分である。初年度には疾患エン트리(H 番号エン트리)の全体的な見直しを行い、KEGG ID が付与されず不完全なデータしか入っていなかった環境要因についてその項目(Carcinogen と Env_factor フィールド)を削除した。疾患エンtringの主要項目は Gene、Network、Drug の 3 つのフィールドであり、これらは以下のように構築している。Gene フィールドの疾患遺伝子については OMIM や MEDGEN を参考にして文献情報を調べ、手作業で入力し常に見直しも行っている。ただし疾患遺伝子といってもバリエーションに関する記述は translocation といったテキストによる記載のみで、実際にバリエーションエンtringを作成しているわけではない。特定のバリエーションエンtringの作成は NETWORK データベース構築の際に行っているが、変異の詳細を蓄積するのではなく、ゆらいだネットワークとして対応する H 番号へのリンクづけのために行っている。これはネットワークバリエーションマップ編集の際に入力し、疾患エンtringのネットワーク情報(Network フィールド)はその逆引きを表示している。同様に治療薬の情報(Drug フィールド)は医薬品データベース構築の際に、添付文書に基づく適応症から手作業でリンクづけした H 番号の逆引きである。このように疾患データベースの高品質化は KEGG MEDICUS 全体の高品質化と深く関連しており、表 1 において疾患エンtring数の増加よりも、ネットワークバリエーションマップと関連づけられた(nt linked)エンtring数の増加が、本研究の進捗状況を最もよく反映している。

すでに述べた通り、前期の NETWORK データベースは疾患パスウェイマップに詳細情報を付与する観点で開発を行った。疾患パスウェイマップで疾患に関与する遺伝子バリエーションは赤字で表現されているが、実際のリンク先は通常遺伝子であり、NETWORK データベース構築の中でバリエーションエンtringを作成しネットワークバリエーションマップを作成した。このやり方では必然的に疾患の種類として、がんや神経変性疾患など限られたものになっていたので、本研究では疾患エンtringに詳細情報を付与する観点で、より幅広い疾患を対象としている。表1で 2022/4/1 が前期開発分のデータ数であり、本研究開始後にバリエーションの数とネットワーク情報が付与された疾患エンtring数が大きく伸びていることが分かる。

DISEASE データベースでは主要フィールド以外に、疾患名の階層化、ICD-11 疾患分類との対応づけ、独自の疾患カテゴリ分類も継続して行っている。疾患エンtringにはもともと疾患をどのレベルで定義するか階層が内在し、総称名の下に複数の個別名があるといったことがしばしばある。このような親子関係は内部用の疾患エンtring編集画面で定義されており、公開されている疾患エンtringではこれらはサブグループ名として表示され、その逆引きとしてスーパーグループ名も表示されるようになっている。

(3) 医薬品データベースの開発

医薬品データベースはすでに高品質化が達成されており、KEGG MEDICUS の中で最も広く利用されているデータベースである。医薬品添付文書をはじめ実際に社会で使われているデータに基づいているため、社会的価値が高い実用的なデータベースでもある。医薬品データベースは、日米欧での医薬品有効成分を蓄積した DRUG データベース、様々な観点で医薬品をグループ化した DGROUP データベース、それに外部から導入している医薬品添付文書のデータから構成されている。DRUG データベースでは有効成分ごとに医薬品エン트리(D 番号エン트리)を作成しており、日本(PMDA)、米国(FDA)、欧州(EMA)で承認された新薬、および日本(JAN)、米国(USAN)、欧州(INN)に登録された新薬名(開発中のものを含む)に対応する有効成分を直ちに取り込む更新体制は確立している。また各エン 트리に対して、医薬品の標的分子(Target フィールド)、薬物代謝酵素と薬物トランスポーターの基質(Enzyme フィールド)および阻害・誘導(Interaction フィールド)、効能とくに適応疾患(Efficacy および Disease フィールド)といった独自のアノテーションを付与する品質管理を継続して行っている。標的分子については従来から Pathway フィールドにおいて関連するパスウェイマップへのリンクづけを行っていたが、本研究において新たにネットワークバリエーションマップにおける医薬品・標的関係のネットワーク要素を Network フィールドで定義できるようにした。標的分子はヒト遺伝子産物(遺伝子 ID)、ヒト遺伝子バリエーションの産物(バリエーション ID)、切断されたペプチド鎖(C 番号)、病原体タンパク質(K 番号)、その他様々なものが含まれている。

DGROUP データベースは医薬品を Chemical、Structure、Target、Metabolism、Class の 5 つの観点でまとめたグループである。Chemical は塩や水和状態などの違い以外は同一の化学構造をもち、有効成分として実質的に同じ化学物質のグループである。Structure は基本骨格、Target は標的分子、Metabolism は薬物代謝が類似のグループで、医薬品相互作用(併用禁忌や併用注意)を定義する際に使用している。さらに、Class は医薬品添付文書のページで関連する商品一覧を提示するための実用的なグループであり、効能・効果が類似の幅広いグループを定義できるようになっている。これら医薬品グループエン 트리(DG 番号エン 트리)の追加・更新も継続して行っている。

医薬品添付文書については従来からの更新体制を継続し、商品ごとに有効成分と添加物について D 番号、DG 番号、C 番号などの対応づけを行っている。日本の医療用および一般用医薬品添付文書は、毎月 1 度、日本医薬情報センター(JAPIC)からデータの提供を受けて内部データベースに登録している。同じタイミングで米国 FDA の National Drug Code (NDC) database から米国の医療用医薬品のリストを取得し、NIH/NLM の DailyMed データベースへのリンクや検索用データを内部データベースに登録している。実際の添付文書データは内部データベースには登録せず、DailyMed を参照する形にしている。日本の医療用医薬品添付文書からは併用禁忌・併用注意に伴う相互作用を網羅的に抽出し、D/DG/C 番号と ATC 番号で標準化したデータセットを作成している。さらに日本の商品については、厚生労働省の薬価基準収載品目リストを常に最新のものと更新しており、個別医薬品コード(YJ コード)と D 番号の対応づけを可能にしている。

(4) 解析ツールの開発

この開発項目では、開発項目(2a)ウイルスタンパク質オーソロググループ(VOG)に関する検索・閲覧・解析ツールの開発を行ってきた。第 2 年度初めに VOG を公開した際には KEGG Virus のトップページでキーワード検索ツールも利用できるようにした。VOG は類似度の閾値 30%, 50%, 70%で自動的に作成され、それぞれ 3, 5, 7 から始まる 6 桁の数字で ID をつけているが、これは更新のたびに変わり得る ID で、実験データに基づき手作業で定義している KO につけられた ID (K 番号)と同じ取扱いはできない。しかし KO と同様に VOG をメンバーシップ同定のために用いること、例えば与えられたウイルスタンパク質と 30%の閾値で同じグループに属するタンパク質は何か、またウイルスは何かといったことは有用な情報となる。そこでこれまで KO のために開発してきたツール群を VOG でも利用できるように拡張した。その1つが類似遺伝子をもつ生物種やウイルスの系統関係を調べるための Taxonomy mapping ツール、もう1つはゲノム上での類似遺伝子の並びを調べる Gene cluster ツールである。これらは KEGG の各ウイルス遺伝子エントリページにつけられた Vog ボタンと Vog cluster ボタンから参照できるようにしている。

以上は当初計画で予定していた VOG の検索・閲覧・解析ツールであるが、当初計画にはなかった拡張としてウイルスと生物種の間での遺伝子のやりとりを調べるツール開発も行っている。そのためウイルスタンパク質のみで定義した VOG データセットに対して、KEGG の全生物種のタンパク質を1つずつ比較判定して、VOG に追加したデータセット作成を行っている。生物種の追加メンバーシップがある場合は、キーワード検索ツールや Vog ボタンからの一覧に+が表示され、詳細を見ることができるようになっている。さらに個々の遺伝子のやりとりだけでなく、遺伝子クラスターとしてのやりとりを調べることができるツールも開発中である。KEGG ではシンテニー領域を調べる独自の方法として、ゲノム上の遺伝子の並びを K 番号の並びとして表現し、配列アライメントの手法として開発した Goad-Kanehisa アルゴリズムを適用したツールを公開している。VOG はウイルス遺伝子の 90%につけられており、ウイルス遺伝子の並びを VOG の並びとして表現することで、ウイルスと生物種の間での遺伝子クラスターのやりとりを網羅的に調べることができる。図 2 に結果の例を示した。

vtax	2681611	73	Lambdavirus lambda
ecs	T00048	5280	Escherichia coli O157:H7 Sakai (EHEC)
sort by [score vtax ecs]			
vg	NC_001416	1 => NC_001416	26 Score: 21 Ident: 1.00 (21/21)
ecs	1540	=> 1562	
K22014	K21512	K26964	K26963 K21511 K26962 K26961 K26965 K26966 K26967
2703523	2703524	2703525	2703526 2703527(2) 2703529 2703482 2703483 2703484 2703485
304116	301881	300882	300826 300375 301487 300808 304918 304583 303463
ECs_1629	ECs_1630	ECs_1631	ECs_1632 ECs_1633 ECs_1634 ECs_1635 ECs_1636 ECs_1637 ECs_1638
K22014	K21512	K26964	K26963 K21511 K26962 K26961 K26965 K26966 K26967
K26968	K26969	K26970	K26972 K26973 K26974 K26975 K26977 K07804
2703486	2703487	2703488(2)	2703511 2703512 2703513 2703514(2) 2703516 2703517 2703518(3)
303464	300002	302562	300302 300268 300075 300073 300086 301683 300002
ECs_1639	ECs_1640	ECs_1641(2)	ECs_1643 ECs_1644 ECs_1645 ECs_1646(2) ECs_1648 ECs_1649 ECs_1650
K26968	K26969	K26970	K26972 K26973 K26974 K26975 K26977
K26980			
2703503			
300810			
ECs_1651			
K26980			
[KO list]			

図 2. ラムダファージと大腸菌 O157 を VOG で比較した例、KO も表示している

(2) 若手研究者の多様なキャリアパスの支援

本プログラム参加者である孟令杰 (Lingjie Meng) さんは当初は京都大学化学研究所 特定研究員として、京都大学の学位を取得後の 2024 年 4 月からは同特定助教として本プログラムで雇用している。2025 年 4 月からは次のステップとして別プログラムに移る予定である。

§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

表 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)

種別	2021 年度 (参考)	2022 年度	2023 年度	2024 年度 (8 月末時点)
訪問数 (Visit)	1,964,636	2,862,801	3,032,459	2,877,716
訪問者数 (Unique)	3,744,592	6,088,038	6,762,179	6,591,965
閲覧ページ数 (Views)	6,570,287	10,985,920	14,667,077	15,023,660

(2) 分析

アクセス数として KEGG MEDICUS は KEGG 全体の 7～8 割を占めており、研究者以外の利用、とくに Google 等のウェブ検索から医薬品添付文書サービスを利用している方が多い。Google のインデクシングのやり方などにより月ごとのばらつきもあるが、2024 年度の利用状況は今のところ横ばい傾向となっている。

2. データベースを利用して得られた研究成果・産業応用の例

以下は KEGG MEDICUS に限定することが困難であるので、KEGG 全体の利用について述べる。Web of Science で引用回数が 1,000 回以上ある KEGG 論文 17 件を引用した論文は 67,502 件あった(2024 年8月31日現在)。このうち最新の Hot Papers と指定されたものが 16 件あり、そのうち Nature 系列誌 4 件を含む 9 件は以下の通りであった。

1. Zuffa, S; Schmid, R; (...); Dorrestein, PC. microbeMASST: a taxonomically informed mass spectrometry search tool for microbial metabolomics data. Nature Microbiology 2024 Feb;9(2):336-345.
2. Wang, ZY; Fern, C; (...); Kennes, C. Effect of electron acceptors on product selectivity and carbon flux in carbon chain elongation with Megasphaera hexanoica. Science Total Environment. 2024 Feb 20;912:169509.
3. Tang, DD; Chen, MJ; (...); Wang, YW. SRplot: A free online platform for data visualization and graphing. PLoS One 2023 Nov 9;18(11):e0294236.
4. Chklovski, A; Parks, DH; (...); Tyson, GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nature Methods 2023 Aug;20(8):1203-1212.

5. Satam, H; Joshi, K; (...); Malonia, SK. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology (Basel)* 2023 Jul 13;12(7):997.
6. Zhao, L; Zhang, H; (...); Liang, QQ. Network pharmacology, a promising approach to reveal the pharmacology mechanism of Chinese medicine formula. *Journal of Ethnopharmacology* 2023 Jun 12;309:116306.
7. Bartman, CR; Weilandt, DR; (...); Rabinowitz, JD. Slow TCA flux and ATP production in primary solid tumours but not metastases. *Nature* 2023 Feb;614(7947):349-357.
8. Shen, SQ; Zhan, CS; (...); Luo, J. Metabolomics-centered mining of plant metabolic diversity and function: Past decade and future perspectives. *Molecular Plant* 2023 Jan 2;16(1):43-63.
9. Hosomi, K; Saito, M; (...); Kunisawa, J. Oral administration of *Blautia wexlerae* ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota. *Nature Communications* 2022 Aug 18;13(1):4477.

3. その他

KEGG を利用した論文でパスウェイマップ等の図を転載して使う際には Copyright permission を取得する必要がある。その申請件数は学術雑誌での KEGG の利用状況を反映しており以下の通りであった。

年度	2021 年度	2022 年度	2023 年度	2024年度 (8 月まで)
申請件数	1,465	1,997	2,229	675

§5. 今後の研究実施計画

今後の研究実施計画は基本的に当初計画に記載していた通りである。第 5 年次までの達成目標として、後期 2 年間で新規ネットワーク要素(バリエーションのみの要素を含む)の数を 400 増加させることに加え、分子ネットワーク情報が付与された疾患数の割合を現在の 3 割から 4 割に上昇させることを目指す。NETWORK データベース、DISEASE データベース、DRUG データベースの一体的な開発により、とくにゆらいだネットワーク要素と疾患の関連、医薬品と適応症およびネットワーク上の標的との関連を通して DISEASE データベースの高品質化を行う。ウイルスオーソロググループ(VOG)については、どのような VOG クラスター(遺伝子の並び)がウイルスと生物種の間で共通に存在するかのデータを計算手法のみで生成し、データベースと解析ツールを完成させる。VOG クラスター情報は当初計画にはなかった拡張であり、トランスポゾンやウイルスの内在化といったことも含め、ウイルスタンパク質と生物種タンパク質の関連解析ができるリソースとする。

§6. 自己評価

KEGG は国際的に広く利用されているデータベースで、ウェブアクセスでは月間 1,000 万の訪問数がある。本研究課題の KEGG MEDICUS は、生命の基本原理を探究するといった基礎的な側面をもつ KEGG 本体部分に対し、ヒト疾患の理解や医薬品開発支援などで社会貢献を行うといった応用的側面を重視した部分で、アクセス数の 2/3 程度がこの部分であり、とくに国内からの利用率が高い。ライフサイエンス研究全体を活性化させるための情報基盤を構築するという本推進事業の目的に貢献していると考ええる。

一方 KEGG はこれまで一貫してシーズ先行の独自開発路線をとっており、KEGG MEDICUS 構築においても、データ利用者との連携や他データベースとの連携などは不十分であるかもしれない。しかしながらすべてを内部で管理・開発することで高品質のデータベースが実現できているのであり、データ利用者や他データベースが利用しやすいインターフェース・ツール等も常に改良していることから、広い意味では本プログラムの趣旨に一致していると考えている。

§7. 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)	2 件
未発行論文 (accepted, in press 等)	国内(和文)	0 件
	国際(欧文)	0 件

(2) 論文詳細情報

1. Minoru Kanehisa, Miho Furumichi, Yoko Sato, Masayuki Kawashima and Mari Ishiguro-Watanabe, “KEGG for taxonomy-based analysis of pathways and genomes”, Nucleic Acids Research, vol. 51, No. D1, pp. D587-D592, 2023 (DOI: 10.1093/nar/gkac963).

概要: Nucleic Acids Research の Database Issue で 2 年ごとに出している KEGG 論文の最新版。今回は生物種やウイルスの系統樹情報を統合する形でゲノムやパスウェイの解析ができるリソースとしての報告。

2. Zhao Jin, Yoko Sato, Masayuki Kawashima and Minoru Kanehisa, “KEGG tools for classification and analysis of viral proteins”, Protein Science, vol. 32, No. 12, e4820, 2023 (DOI: 10.1002/pro.4820).

概要: 本研究で開発したウイルスオーソロググループ(VOG)を含む KEGG Virus リソースの紹介。

2. その他の著作物(総説、書籍など)

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	2 件
	国際	1 件
口頭発表	国内	0 件
	国際	0 件
ポスター発表	国内	0 件
	国際	0 件

(2) 招待講演

〈国内〉

1. 金久 實、ゲノム・情報・医療・ビジネス、聖ラ・サールの日講演会、ラ・サール学園、2022 年 5 月 17 日
2. 金久 實、KEGG がつなぐゲノムと社会、鹿児島大学第一内科開講記念会、城山ホテル鹿児島、2022 年 6 月 16 日

〈国際〉

1. Minoru Kanehisa, Establishing a self-sustaining database for a sustainable society, HFSP Symposium at ISMB/ECCB 2023, Lyon, France, July 26, 2023

(3) 口頭講演

〈国内〉

〈国際〉

(4) ポスター発表

〈国内〉

〈国際〉

4. 知財出願

(1) 出願件数

種別		件数
特許出願	国内	0 件
	国外	0 件
その他の知的財産出願		0 件

(2) 一覧

- ① 国内出願
- ② 海外出願
- ③ その他の知的財産権

5. 受賞・報道等

(1) 受賞

- 1. 小林財団 小林賞、金久 實、2024 年 2 月 29 日

(2) メディア報道

- 1. 南日本新聞、鹿児島県ゆかりのノーベル賞候補、2022 年 10 月 3 日
- 2. テレビ大阪ニュース、生命科学分野の研究を評価、2024 年 2 月 29 日

(3) その他

- 1. Clarivate Analytics Highly Cited Researchers 2022、金久 實、2022 年 11 月 15 日
- 2. Clarivate Analytics Highly Cited Researchers 2023、金久 實、2023 年 11 月 15 日

§8. 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
-----	----	----	------	-------

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
-----	----	----	------	-------

以上

別紙 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		ゲノムの情報と疾患・医薬品との関連を、生体システムを構成する分子ネットワークを通して統合的に理解し、ヒトゲノム情報および病原体ゲノム情報の有効利用を促進するためのリソースである。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である。	https://www.kegg.jp/kegg/medicus/	2010年10月1日	継続・発展	データベース等	公開済			論文1
2	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。KEGG MEDICUSにある医薬品添付文書に記載された相互作用をすべて抽出し、標準化したデータセットを用いている。	https://www.kegg.jp/medicus-bin/ddi_manager	2016年4月1日	継続・発展	ツール等	対象外			