

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

「ゲノム・疾患・医薬品のネットワークデータベース」

金久 實 | 京都大学化学研究所 特任教授



©2019 金久 實(京都大学) Licensed under CC BY 4.0

§1. 研究開発実施の概要

生命システムには異なるレベルで保存性と多様性が内在する。これまで KEGG データベースでは生物種間で保存された機能に着目し、機能オーソログ (KO) で表現した KEGG パスウェイマップを開発してきた。この知識のレファレンスは様々な生物種ゲノムのデータから高次生命システム機能を解読するためのレファレンスデータベースとして広く利用されている。一方、平成 23 年度からの統合化推進プログラムでは、ゲノムの情報を医療・創薬等で活用するためのトランスレーショナルバイオインフォマティクスリソースとして KEGG MEDICUS の開発を行ってきた。とくに疾患は生体システムを構成する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と環境のゆらぎを列挙する形で KEGG DISEASE データベースを構築し、また医薬品は分子ネットワークに対する別のゆらぎ物質であるとの観点から、各医薬品の標的分子と関連パスウェイを KEGG DRUG データベースに蓄積してきた。そこで明らかになったことは、生物種間の保存性に着目してきた従来の KEGG のアプローチの限界である。ゲノムと疾患あるいは医薬品との関連づけを行うためには、ヒトゲノムの多様性(バリエーション)を含めたデータベース化が必要である。

そこで本研究開発では、疾患や医薬品に関連した分子ネットワークのバリエーションを蓄積した KEGG NETWORK を開発する。分子ネットワークのバリエーションとは、レファレンスとなるネットワーク(例えばシグナル伝達ネットワークや代謝ネットワーク)に対して、ヒトゲノムのバリエーション、ウイルスその他の病原体、環境因子、医薬品などがもたらす分子ネットワークのゆらぎを表現したものである。KEGG の他のデータベースと同様に、ネットワークバリエーションのデータはすべて論文をもとに集約し、ハイスループット実験データは含めない。KEGG NETWORK データベースでは、個々のレファレンスまたはゆらいだネットワークをネットワーク要素(N 番号エントリ)として定義し、多数の関連したネットワーク要素をアライメント表示したネットワークバリエーションマップ(nt 番号エントリ)を開発する。ネットワークバリエーションマップは KEGG Mapper に組み込み、クリニカルシーケンシングデータの解釈や医薬品標的分子の探索など、ゲノム情報有効利用のための新たなレファレンスリソースとして提供する。

本研究は 2017 年 4 月に開始し、がんに関連する KEGG NETWORK を開発して同年 12 月に最初のバージョンを公開した。その後、ウイルス感染症、内分泌疾患、先天性代謝異常症について KEGG NETWORK の開発を行い、2019 年 10 月の時点で 900 のネットワーク要素と 100 のネットワークバリエーションマップを公開した。研究開発提案書に記載した第 3 年次末までに 1000 のネットワーク要素という数値目標は達成できる見込みである。またネットワークバリエーションマップのグラフィカルインターフェースを開発し、KEGG Mapper への組み込みも行った。KEGG NETWORK、DISEASE、DRUG および医薬品添付文書を統合した KEGG MEDICUS に対する利用者数は毎日 10 万から 20 万あり、研究開始時点の 10 倍近くになっている。

§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(研究代表者グループ)

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

有識者会議等は設置していない。

§3. 研究開発の目的、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

正式名称	略称	概要
KEGG MEDICUS		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、生体システムを構成するネットワーク情報として統合され、ゲノムシーケンスデータ等の大量データの解釈と有効利用を促進するリソースである。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
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。

2. 達成目標及び実施計画

(1) 当初の実実施計画・達成目標

本研究開発では、ゲノムの情報から疾患や医薬品に関する知見を得るための新しいデータベースとして、ヒトゲノムのバリエーション(多様性)を生体システムを構成するネットワーク要素のバリエーションとして蓄積した KEGG NETWORK を開発する。ネットワーク要素とは、シグナル伝達ネットワーク、代謝ネットワークをはじめとした分子間相互作用・反応ネットワークの中で、基本的な単位となる部分を定義したものである。KEGG MEDICUS には疾患情報(KEGG DISEASE)、医薬品情報(KEGG DRUG)、ネットワーク情報(KEGG NETWORK)が統合され、臨床シークエンシングデータの解釈や医薬品標的分子の探索など、ゲノム情報有効利用のための新たなレファレンスリソースとして提供する。実施計画では「ネットワークデータベース」と「ネットワーク解析」の新規開発項目、および「KEGG MEDICUS 統合リソース」の継続開発項目を設定して研究開発を進める。達成目標としては、(i) KEGG NETWORK データベースの最初のバージョンを研究開始後 1 年以内に公開し、第 3 年次末までに、(ii) ネットワーク要素について、ノード間の様々な制御関係を定性的に記述ができる論理表現法を確立してデータ入力、(iii) 遺伝子バリエーションについて、ネットワークノードに対応したグループ化の基準を確立してデータ入力、(iv) ネットワークマッピングツールと可視化ツールの開発・提供を行う。計画の詳細は以下の通りである。

1. ネットワークデータベース

1-1 ネットワーク要素

KEGG NETWORK では分子間のつながりを表現したネットワーク要素を基本的なエン트리(N 番号エン트리)とし、KEGG MODULE の M 番号エントリで用いた論理表現を拡張した形で定義する。ネットワーク要素のノードはヒト遺伝子、そのバリエーション、医薬品、環境因子(化合物や病原体)などである。モジュールと同じように、AND (+) で複合体を、OR (,) でパラログなどを表現できるようにし、モジュールにはないノード間のつながり(エッジ)のタイプも記述できるようにする予定である。例えば Activation, Inhibition, Missing interaction といったタイプである。Missing interaction は遺伝子変異により相互作用そのものがなくなるような場合に用いる。これまで KEGG DISEASE に蓄積してきた疾患と遺伝子の関係、および KEGG DRUG に蓄積してきた医薬品と標的分子の関係の中から、KEGG パスウェイマップと対応づけられるものを選別し、そこからネットワーク要素を定義していく。

1-2 遺伝子バリエーション

遺伝子バリエーションについてはすでに多くのデータベースで詳細な情報が蓄積されている。本研究では疾患や医薬品に関与するバリエーションのみ考慮し、また1つ1つのバリエーションを区別するのではなく、ネットワーク中のノードとして適切になるようグループ化を行う。KEGG パスウェイマップのノードが機能オーソログ KO でグループ化されていること、統合化推進プログラム第二期の研究で医薬品相互作用ネットワークのノードに医薬品グループ DG を導入したことにも通じる考え方である。KEGG GENES では例えばヒトの ALK 遺伝子には hsa:238 という ID がついており(238 は NCBI GeneID にも対応する)、KEGG NETWORK では 238v1、238v2 といった形の ID でバリエーションを区別する。

2. ネットワーク解析

2-1 ネットワークエントリ検索

KEGG NETWORK を用いたネットワーク解析のツールとして、本研究では2つを開発する。1つはテキスト検索ツールで、与えられたキーワードやノード ID(遺伝子 ID など)でヒットした KEGG NETWORK のエントリを表示する。これは KEGG で使用している DBGET 検索システムに組み込み、KEGG MEDICUS 統合インターフェースからも検索できるようにする。

2-2 ネットワークマッピング

もう1つは多数のデータの一括処理で、ネットワークマッピングのツールを開発する。従来からの KEGG パスウェイマッピングでは、ゲノム、トランスクリプトーム、メタボロームなどの大量データを KEGG PATHWAY データベースにマッピングし、再構築できた(もとのデータと対応づけられた)部分をもとに、例えばどのような代謝経路やシグナリング経路があるかを判定し、大量データの意味づけを行う。ネットワークマッピングでも同様に大量データの解釈を支援し、とくにバリエーションを含むゲノムのデータから疾患や医薬品に関する知見が得られるようにする。

3. KEGG MEDICUS 統合リソース

3-1 疾患情報

KEGG MEDICUS には上記のネットワーク情報(KEGG NETWORK)が追加され、既存の疾患情報(KEGG DISEASE)、医薬品情報(KEGG DRUG/DGROUP/ENVIRON)と統合したリソースとする。疾患情報については、第二期の開発で医薬品添付文書の適応症との対応づけが明確になるような形で KEGG DISEASE 疾患エントリの全体的な見直しを行ったが、日本だけでなく米国の医薬品添付文書も用いてこれを継続する。また疾患情報とネットワーク情報の

対応づけを行い、ネットワークマッピングなどで疾患に関する知見が得られるようにする。

3-2 医薬品情報

医薬品情報については、今後とも日本医薬情報センター (JAPIC) から毎月提供を受けている我が国の医療用・一般用医薬品添付文書を KEGG の D 番号、DG 番号、C 番号、E 番号などで標準化し、KEGG DRUG と統合して公開する。米国の医療用・一般用医薬品添付文書についても同様の標準化を行い、現状通り DailyMed データベースへのリンクとして統合する。また上記の通り日米の医薬品添付文書から適応症を抽出し、KEGG DRUG エントリの DISEASE フィールドに記載する作業を継続する。

3-3 統合インターフェース

新規開発するネットワークエントリ検索システムを既存の KEGG MEDICUS 検索システムに組み込み、KEGG DISEASE, KEGG DRUG などと統合的に検索できるインターフェースとする。また添付文書から抽出・標準化したデータセットに基づく医薬品成分 (有効成分と添加物) 検索ツールと医薬品相互作用チェックのツールの提供も継続する。

KEGG NETWORK に対するネットワークマッピングのツールは KEGG Mapper に組み込んで提供する。KEGG MEDICUS API の開発も継続する。

(2) 期間中に追加・削除・変更した実施計画・達成目標

主な変更点は 2 つある。第 1 は遺伝子バリエーションの取扱いである。KEGG NETWORK で定義した遺伝子バリエーションを dbSNP、ClinVar、COSMIC などの外部データベースに対応づけようとする、対応するデータが登録されていなかったり (著者からの登録に依存したレポジトリであるため)、対応すると思われるデータが大量にあってどれを選ぶべきかが分からない (ハイスループットデータが登録されているため)、といったことがしばしばあることが明らかとなった。当初計画では臨床シークエンスデータのバリエーションアノテーションで得られる dbSNP や ClinVar 等の ID を KEGG の遺伝子バリエーション ID に変換して利用されることを想定していたが、今後は KEGG 独自の遺伝子バリエーション ID が直接利用されるような形で開発を進めていくこととした。これに関連して、当初計画では KEGG NETWORK と dbSNP や ClinVar 等とのリンク情報を RDF 形式で提供する予定であったが、外部データベースの問題点からこの計画は削除した。なお、KEGG DISEASE や KEGG DRUG については従来からゲノムネットワークの LinkDB を通じて RDF データが提供されている。

第 2 は追加項目で、当初計画では明確に記載していなかった「ネットワークバリエーション

マップ」と呼ぶグラフィカルマップに nt 番号の識別子をつけてエントリ化し、KEGG NETWORK の主要なデータセットとして開発を行うこととした。当初の KEGG NETWORK データベースはネットワーク要素 (N 番号エントリ) のコレクションであったが、現在の KEGG NETWORK はネットワークバリエーションマップ (nt番号エントリ) のコレクションという位置づけである。

3. 実施内容

(1) 実施内容

当初の全体計画における達成目標は、上記 2(1) にも記載した通り、研究開始後 1 年以内に (i) KEGG NETWORK データベースの最初のバージョンを公開、第 3 年次末までに、(ii) ネットワーク要素について、ノード間の様々な制御関係を定性的に記述ができる論理表現法を確立してデータ入力、(iv) 遺伝子バリエントについて、ネットワークノードに対応したグループ化の基準を確立してデータ入力、(iv) ネットワークマッピングツールの開発と提供であった。KEGG NETWORK の最初のバージョンは 2017 年 12 月 1 日に公開した。以下に示す通り、当初の全体計画に記載した第 3 年次末の達成目標はすべて達成し、上記 2(2) に記載した追加項目も含め、計画はスケジュール通りに進んでいる。

1. ネットワークデータベース

ネットワークデータベースについて最初に行ったのは、データを登録・編集するためのデータベース管理システムの開発である。2017 年 5 月末までにデータベースのデザインを決定し、これまでの KEGG データベースの管理システムを拡張する形で NETWORK と VARIANT の入力編集システムを稼働させた。その後、内部用のデータベース管理システムから外部公開のためのフラットファイルデータベースを出力できるようにし、NETWORK と VARIANT を DBGET 検索システムに組み込んだ。

表 1. KEGG NETWORK データの増加

リリース	Network element	Network variation map	Gene variant	ネットワークの疾患タイプ
2018 年 1 月	282		118	がん 2017/12/1 公開
2018 年 4 月	296		135	
2018 年 7 月	438		169	内分泌疾患 2018/5/10 公開
2018 年 10 月	448		169	ウイルス感染症 2018/8/1 公開
2019 年 1 月	515	67	171	
2019 年 4 月	690	88	245	先天性代謝異常症 2019/4/1 公開
2019 年 7 月	782	91	307	
2019 年 10 月	891	96	376	

2017 年 6 月から田辺麻央が中心となってデータ入力を開始した。まず 17 種類のがんについて論文をもとに、ヒト遺伝子バリエントによるシグナル伝達と転写制御ネットワークのバリエント(ゆらいだネットワーク)を定義し、6 種類のがんウイルスについてもウイルスタンパク質によるネットワークバリエントを定義してデータベース化を行った。ウイルスタンパク質については古道美穂が

KO (KEGG Orthology) の定義などを行った。がんを対象とした KEGG NETWORK の最初のバージョンは 2017 年 12 月 1 日に KEGG MEDICUS (<https://www.kegg.jp/kegg/medicus/>) の一部として公開した。第 2 年次にはがんウイルス以外の主要なウイルスとして 5 種類を追加し、がんウイルス感染症については一応の区切りをつけた。また第 2 年次には内分泌疾患と先天性代謝異常症についてもネットワークバリエーションのデータベース化を開始し、これは現在も続いている。3 ヶ月ごとの KEGG リリース時点でのデータ数は表 1 に示した通りである。

表 1 にあるネットワークバリエーションマップとは関連するネットワーク要素をグラフィカルにアライメント表示したもので、例えばネットワーク要素 N00001 の EGF-EGFR-RAS-ERK signaling pathway をレファレンスとして、これに様々な遺伝子バリエーション(変異、融合、過剰発現、その他)やウイルスタンパク質が引き起こすネットワークバリエーションの全体像が分かるようにしたものである。例えば図 1 は、レファレンスのネットワーク要素(緑)に対して、ALK 融合遺伝子(赤)と KSHV (Kaposi sarcoma-associated herpesvirus) K1 タンパク質(紫)によるバリエーションのネットワーク要素をアライメント表示して比較できるようにしたもので、両者が類似のゆらぎを与えていることが分かる。³⁾ この例のように、疾患に関連するヒト遺伝子バリエーションだけでなく、ウイルスその他の病原体、さらには様々な環境因子がどのように生体内の分子ネットワークに影響を与えているか、その共通性と多様性を探るためのリソースにすることが、KEGG NETWORK 構築の大きな目標である。

N00001	EGF	→ EGFR	→ GRB2	→ SOS	→ RAS	→ RAF	→ MEK	→ ERK	→ CCND1
N00007				EML4-ALK	→ RAS	→ RAF	→ MEK	→ ERK	→ CCND1
N00160				kshvK1	→ RAS	→ RAF	→ MEK	→ ERK	
N00023	EGF	→ EGFR	→ PLCG	→ (Ca2+,DAG)	→ PKC	→ RAF	→ MEK	→ ERK	→ CCND1
N00025		EML4-ALK	→ PLCG	→ (Ca2+,DAG)	→ PKC	→ RAF	→ MEK	→ ERK	→ CCND1
N00147	EGF	→ EGFR	→ PLCG	→ IP3	→ Ca2+	→ CALM	→ CN	→ NFAT	
N00180		kshvK1	→ PLCG2	→ IP3	→ Ca2+	→ CALM	→ CN	→ NFAT	
N00033	EGF	→ EGFR	→ PI3K	→ PIP3	→ AKT	→ BAD			
N00047		EML4-ALK	→ PI3K	→ PIP3	→ AKT	→ BAD			
N00159		kshvK1	→ PI3K	→ PIP3	→ AKT	→ MTOR			
N00053	Cytokine	→ Receptor	→ JAK	→ STAT	→ PIM1				
N00105		EML4-ALK	→ JAK3	→ STAT3,STAT5					

図 1. ネットワークバリエーションマップの例

ネットワークバリエーションマップは可視化ツールとして KEGG NETWORK 最初の公開時点から存在していたが、例えば ERK signaling に対して nt06210 のような ID をつけ、DBGET 検索システムにも組み込んだのは 2018 年 11 月で、表 1 にはその次のリリースからのデータ数が表示されている。ネットワーク要素の N 番号エントリはテキストデータで、現時点では表 2 に示したノードのタイプと表 3 に示したエッジのタイプが定義されている。図 1 にもある通り、ネットワーク

バリエーションマップでは表 2 の右端にある色づけと表 3 の右端にあるシンボル表示がなされ、さらにがんのタイプやその他の疾患情報も表示して比較解析することができるようになっている。

表 2. KEGG NETWORK のノード

Node	Identifier	Coloring
Human reference gene/protein	hsa ID	Green
Human gene variant	variant ID	Red
Viral or other pathogen gene/protein	K number	Purple
Metabolite and other chemical compound	C number	None
Drug	D number	Blue

表 3. KEGG NETWORK のエッジ

Type	Edge	Text	Symbol
Signaling network	Activation (single- or multi-step)	->	→
	Inhibition (single- or multi-step)	-	⇩
	Complex formation	--	—
	Missing interaction	//	∕
	Expression (single-step)	=>	⇒
	Repression (single-step)	=	⇨
Metabolic network	Enzyme-enzyme relation	>>	⇒
	Substrate-enzyme relation	--	—
	Enzyme-product relation	->	→
	Missing reaction	//	∕

2. ネットワーク解析

ネットワーク解析のツール群としては、テキスト検索、ネットワーク可視化、ネットワークマッピングを計画した。まずテキスト検索では既存の 2 つの検索システムに組み込む形で実現した。1 つは NETWORK と VARIANT のフラットファイルデータベースを DBGET 検索システムに組み込んで検索できるようにした。NETWORK データベースではネットワーク要素の N 番号エン트리とともに、ネットワークバリエーションマップの nt 番号エントリのテキスト情報も検索できるようになっている。もう 1 つは KEGG MEDICUS 検索システムへの組み込みで、これはフラットファイルではなくリレーショナルデータベースを直接検索する。KEGG NETWORK のページでは NETWORK または VARIANT に対してキーワード入力をする形になっているが、実際は統合的な検索がなされ、キーワードがヒットした DISEASE、DRUG/DGROUP、ENVIRON/COMPOUND のエン트리も同時に分かるようになっている。

ネットワーク可視化でも 2 つのツールを開発した。1 つは図 1 に示したネットワークバリエーションマップのインターフェースである。KEGG MEDICUS 検索システムでは、NETWORK 検索結果の画面からヒットしたネットワーク要素を選択してネットワークバリエーションマップの形で表示す

することもできるようになっている。これをもとに利用者が並べ替えやアライメントなどの編集をして、ネットワークバリエーションマップをカスタマイズできる機能も今年度中に公開する予定である。もう 1 つの可視化ツールはがんに特化したもので、KEGG のパスウェイマップにあるがんのシグナル伝達グローバルマップ (hsa05200) にネットワーク要素の情報を付加する形で **Cancer Network Viewer** を開発し公開した。しかしながらその後 KEGG NETWORK データベースは、個々のネットワーク要素のコレクションではなく、関連するネットワーク要素をアライメント表示したネットワークバリエーションマップのコレクションという位置づけに発展したため、アライメント情報が含まれない **Cancer Network Viewer** は廃止する予定である。

ネットワークマッピングとはヒト遺伝子データのセットからどのような生体システムの分子ネットワークが関連しているか、またどの遺伝子バリエーションがどのようにネットワークにゆらぎを与え、疾患に関連しているか、といった解析を行う方法論である。具体的にはヒト遺伝子 ID、およびヒト遺伝子バリエーション ID のデータセットを KEGG NETWORK のネットワークバリエーションマップに対してマッピングを行い、ヒットしたノードが赤枠で色づけて表示されるツールを開発した。このツールは KEGG Mapper の一部として公開した。⁵⁾

3. KEGG MEDICUS 統合リソース

これは継続開発項目で、これまでの統合化推進プログラムで開発してきた KEGG DISEASE 疾患データベース、KEGG DRUG 医薬品データベース、および日本医薬情報センター (JAPIC) から毎月提供を受けている医薬品添付文書の統合化を維持・発展させている。米国 FDA の医薬品添付文書もリンク情報として統合している。KEGG DISEASE データベースについては新規エントリの作成と既存エントリの見直しを常時行っている。とくに日本または米国の医薬品添付文書に記載された適応症 (日本語版と英語版で異なることがあり得る) と対応がつかないように疾患エントリを定義して、DRUG データベースの DISEASE フィールドにこれを記載し、DISEASE データベースの DRUG フィールドにはその逆引きの医薬品が表示されるようにした。本研究での新たな改良点として、疾患エントリ間の階層的関係 (全体的な病名と細分化された病名など) を示すために、疾患エントリにサブグループとスーパーグループの名称が表示されリンクされるようにした。また国際疾病分類 (International Classification of Diseases) の新バージョン ICD-11 への対応も行った。

KEGG DRUG データベースでは日本、米国、欧州で使用されている医薬品の有効成分がすべて網羅されており、それぞれでの新薬も直ちに KEGG DRUG に登録される体制ができています。また同じ新薬が日本、米国、欧州でいつ承認されたか、承認日の比較表といったユニークなデータも提供している。ただ KEGG DRUG で最も重視しているのは、ターゲット (標的分子) とそれが

関与するパスウェイの情報を付与することである。遺伝子バリエーション(の産物)をターゲットまたはマーカーにした医薬品はすでに多数存在し、今後も個別化医療あるいは精密医療の観点から医薬品開発の主流になると考えられることから、ターゲットにバリエーションを含めて記載することにした。

本研究で新規開発中の KEGG NETWORK/VARIANT は KEGG DISEASE と KEGG DRUG、さらには KEGG PATHWAY や KEGG GENES と以下のように統合されている。まず NETWORK の nt 番号エントリ(ネットワークバリエーションマップ)とそれを構成する N 番号エントリ、DISEASE の H 番号エントリおよび PATHWAY の map 番号エントリ間には相互にリンクづけがなされている。例えば Non-small cell lung cancer として H00014 の DISEASE エントリ、hsa05233 の疾患パスウェイマップ、nt06266 のネットワークバリエーションマップでそれぞれ異なる表現がされているが、それらは相互にリンクされ統合されている。Melanoma の H00038、hsa05218、nt06268 も同様である。DRUG についてはバリエーションをターゲットまたはマーカーにした医薬品情報を蓄積しており、表 4 はがんの分子標的薬の例である。非小細胞肺癌治療薬の Gefitinib は第 1 世代の EGFR 阻害薬、Osimertinib はその薬剤耐性に対する第 3 世代の EGFR 阻害薬である。メラノーマ治療薬の Trametinib は BRAF の変異(673v1)をマーカーとしてその下流の MEK をターゲットにした医薬品である。これらについては Drug と Target の関係がネットワーク要素として定義されており、ネットワークバリエーションマップに重ねて表示することができるようになっている。

表 4. がんの分子標的薬の例

Drug	KEGG DRUG	Target	KEGG VARIANT	Biomarker	KEGG NETWORK	KEGG DISEASE
Gefitinib	D01977	EGFR	1956v2	EGFR exon 19 deletion or exon 21 L858R mutation	nt06266	H00014
Osimertinib	D10766	EGFR	1956v3	EGFR T790M mutation	nt06266	H00014
Trametinib	D10176	MEK	673v1	BRAF V600E mutation	nt06268	H00038

日本および米国で使用されている医薬品の有効成分と添加物は KEGG の D 番号、DG 番号、C 番号、E 番号で標準化している。また医療用医薬品については有効成分の D 番号とともに、Chemical structure group の DG 番号、Class タイプの DG 番号での標準化も行っている。KEGG MEDICUS の日本語版医薬品添付文書のページには毎日 10 万から 20 万の訪問者があるが、単に添付文書を見るだけでなく、商品一覧のインターフェースも頻繁に利用されている。商品一覧は上記標準化に基づき関連商品を比較できるようにしたもので、1) 同一の D 番号と薬効分類番号、2) 同一の D 番号、3) 同一の Chemical structure group DG 番号、4) 同一の Class DG 番号(例えば、HMG-CoA 還元酵素阻害薬)、の 4 つのレベルを提供している。

この標準化に基づき医薬品相互作用チェックツール、医薬品成分検索インターフェースの提供も維持している。その他の統合インターフェースとしては、すでに述べた通り、DBGET 検索システムへの組み込み、KEGG MEDICUS 統合インターフェースへの組み込み、可視化ツールとしてネットワークバリエーションマップの開発と提供を行い、ネットワークマッピングツールを開発し KEGG Mapper の一部として公開した。また NETWORK と VARIANT は KEGG API の中にも組み込まれている。

(2) 「対象とするデータベース条件」のうち採択時に未達成であった項目の対応状況
なし。

(3) 統合化推進プログラムの他のチームや DBCLS との連携

とくに連携はないが、DBCLS からの要望に対応し、ゲノムネットで公開している KEGG MEDICUS FTP サイト(<ftp://ftp.genome.jp/pub/kegg/medicus/>)の改善を行った。

§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	2016年度	2017年度	2018年度	2019年度 (9月末時点)
ユニーク IP 数	252,491	695,248	1,341,147	2,258,817
ページ数	419,254	1,205,103	2,321,565	4,377,074

(2) 分析

ここに示したのは KEGG MEDICUS へのアクセス(KEGG 全体ではない)である。研究開始以前と比較してアクセス数は 10 倍になっているが、これは Google 等の検索エンジンから医薬品添付文書のページに入るトラフィックが急増したためである。とくに 2018 年 9 月のクラリベイト・アナリティクス引用栄誉賞により KEGG の認知度が高まったことで 10 月以降の増加が大きい。

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

本研究で新規開発している KEGG NETWORK については、まだ研究成果・産業応用の例はない。KEGG ではこれまですべてシーズ先行型の開発を行っており、KEGG NETWORK についても、広く活用されるまでにはある程度時間がかかると考えている。

3. その他

KEGG NETWORK を報告した論文 3 は、2018 年 10 月 13 日にオンライン出版され、2019 年 10 月 22 日現在、Web of Science で 90 件、Google Scholar で 178 件の引用がある。ただしこれらは KEGG 全体の引用で、とくに KEGG NETWORK を利用したものは見当たらなかった。

§5. 今後の展開

KEGG データベースは、もともとゲノム解読の情報技術(KEGG マッピングなど)を提供し、基礎的な生命科学研究を推進するためのリソースとして開発してきた。一方 2011 年度からの統合化推進プログラムにおいては応用的側面を重視し、ゲノム情報を社会で活用するための KEGG MEDICUS の開発を行ってきた。本研究での KEGG NETWORK の導入により、KEGG MEDICUS 関連の方法論の開発は一段落がつき、今後はデータの質と量を向上させて KEGG MEDICUS を継続的に発展させ社会的価値を高めたいと考えている。残された課題は基礎的な部分にあり(論文 4 参照)、KEGG で知識を集約してきた分子間相互作用・反応ネットワークと生物種グループとの関連をより明確に定義し、ゲノムの進化とともにネットワークの進化といったものを解析できるようなリソースにしたいと考えている。

§6. 自己評価

KEGG NETWORK は全くゼロの状態から開発を始めたが、これまでに築きあげてきた KEGG データベースの維持・更新体制を活用できたことにより、本研究の第 3 年次までの目標はすべて達成できた。

本研究が統合化推進プログラム全体にどの程度貢献したかについては、新たな知見を見いだす情報基盤の確立や科学技術イノベーションの創出といった目標については、貢献できたと考えている。しかし、そこに至る方法論についてはシーズ先行の独自路線であり、データ利用者との連携や他のデータベースとの連携などは、プログラムの趣旨には一致していないかもしれない。

§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)	1 件
	国際(欧文)	6 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, "Enzyme annotation and metabolic reconstruction using KEGG", In "Protein Function Prediction: Methods and Protocols" (Kihara, D., ed.), Humana Press, Methods in Molecular Biology, vol. 1611, pp. 135-145, 2017 (DOI: 10.1007/978-1-4939-7015-5_11).
2. Minoru Kanehisa, "Inferring antimicrobial resistance from pathogen genomes in KEGG", In "Data Mining for Systems Biology, Second Edition" (Mamitsuka, H., ed.), Humana Press, Methods in Molecular Biology, vol. 1807, pp. 225-239, 2018 (DOI: 10.1007/978-1-4939-8561-6_17).
3. Minoru Kanehisa, Yoko Sato, Miho Furumichi, Kanae Morishima and Mao Tanabe, "New approach for understanding genome variations in KEGG", Nucleic Acids Research, vol. 47, No.D1, pp. D590-D595, 2019 (DOI: 10.1093/nar/gky962).
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。今回はとくに新しい方法論でゲノムのバリエーションをデータベース化している KEGG NETWORK について報告した。
4. Kanehisa, M; Toward understanding the origin and evolution of cellular organisms. Protein Science, vol. 28, No. 11, pp. 1947-1951, 2019 (DOI: 10.1002/pro.3715)
5. Kanehisa, M. and Sato, Y.; KEGG Mapper for inferring cellular functions from protein sequences. Protein Science, vol. 29, 2020 (DOI: 10.1002/pro.3711).

2. その他の著作物(総説、書籍など)

1. Minoru Kanehisa, "KEGG GLYCAN", In "A Practical Guide to Using Glycomics Databases" (Aoki-Kinoshita, K.F., ed.), pp. 177-193, Springer, 2017.
2. 金久 實, "KEGG データベース", バイオサイエンスとインダストリー (B&I), vol. 76, no. 1, pp. 58-62, 2018.

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	3 件
	国際	3 件
口頭発表	国内	3 件
	国際	2 件
ポスター発表	国内	3 件
	国際	0 件

(2) 招待講演

〈国内〉

1. 金久 實、KEGG: 創薬・医療への展開、医薬創成情報科学専攻 10 周年記念シンポジウム、京都大学薬学研究科、2017 年7月 3 日
2. 金久 實、ゲノム・疾患・医薬品の統合リソース KEGG MEDICUS の利用法、MedXML コンソーシアム Seagaia Meeting、宮崎、2018 年 5 月 19 日
3. 金久 實、KEGG for molecular network-based understanding of human diseases、日本バイオインフォマティクス学会年会、東京、2019 年 9 月 9 日

〈国際〉

1. Minoru Kanehisa, From gene variants to network variants: a new database for understanding diseases and drugs, BioNetVisA Workshop in Basel Computational Biology Conference, Basel, Switzerland, September 12, 2017
2. Minoru Kanehisa, "Genome annotation for inferring phenotypic features", WDCM GCM2.0 SIGs Meeting, Beijing, China, June 26, 2018
3. Minoru Kanehisa, "From KEGG Orthology to modules and human diseases", NIBB Conference: Quest for Orthologs, Okazaki, Japan, August 1, 2019

(3) 口頭講演

〈国内〉

1. 金久 實、KEGG NETWORK Version 1: がんのネットワークバリエント、トーゴーの日シンポジウム 2017、東京、2017 年 10 月 5 日

2. 金久 實、ネットワークによるゲノム・疾患・医薬品の統合、トーゴの日シンポジウム 2018、東京、2018年10月5日
3. 金久 實、KEGG NETWORK: 疾患に関連したネットワークバリエーションのデータベース、第41回日本分子生物学会年会シンポジウム、横浜、2018年11月28日

〈国際〉

1. Minoru Kanehisa, “Enzyme annotation in KEGG”, IUPAC-IUBMB Joint Commission on Biochemical Nomenclature, Cambridge, UK, May 31, 2018
2. Minoru Kanehisa, “Toward understanding the origin and evolution of cellular organisms”, European Bioinformatics Institute, Hinxton, UK, September 26, 2019

(4) ポスター発表

〈国内〉

1. 田辺 麻央、金久 實、がんのシグナル伝達ネットワークバリエーションのデータベース、トーゴの日シンポジウム 2017、東京、2017年10月5日
2. 田辺 麻央、金久 實、ウイルスによるネットワークバリエーションのデータベース、トーゴの日シンポジウム 2018、東京、2018年10月5日
3. 田辺 麻央、松浦 有里子、金久 實、内分泌代謝疾患のネットワークバリエーションマップ、トーゴの日シンポジウム 2018、東京、2019年10月5日

〈国際〉

特になし。

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. * Clarivate Analytics Citation Laureate, Minoru Kanehisa, September 20, 2019
2. * Protein Society Carl Brändén Award, Minoru Kanehisa, March 18, 2019

(2) メディア報道

1. 上記 1 に関する報道多数

(3) その他

特になし。

§8. 研究開発期間中の活動

1. 進捗ミーティング

KEGG チーム内のミーティングは毎週行っている。

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
2017年12月13～15日	KEGG MEDICUS の展示 デモ	神戸ポートアイランド	150人	ConBio2017(分子生物学会ほか年会)での展示
2018年1月12日	公開シンポジウム「第9回新たな創薬パラダイムの創出」	東京大学医学科学研究所1号館講堂	92人	日本発の創薬技術・新薬に関するシンポジウム、NPO バイオインフォマティクスジャパンと連携
2018年11月28～30日	KEGG MEDICUS の展示 デモ	パシフィコ横浜	100人	分子生物学会年会での展示とデモ
2019年12月3～6日	KEGG MEDICUS の展示 デモ(予定)	福岡国際会議場		分子生物学会年会での展示とデモ

以上