

独立行政法人科学技術振興機構
バイオインフォマティクス推進センター事業

生命情報データベースの高度化・標準化

第Ⅱ期 研究開発課題

(追跡調査報告書)

研究開発課題：ゲノムと環境の統合解析による

生命システムの機能解読 (KEGG)

(平成18年4月～平成23年3月)

代表研究者氏名：金久 實

(京都大学化学研究所 特任教授)

目次

本編

1. 研究開発による成果.....	1
1.1. 研究開発課題の目標及び新規性・独創性	1
1.2. 研究開発開始時の課題を取り巻く研究・技術水準及び分野における 課題の位置づけ	1
1.3. 研究開発終了時の成果概要	2
1.4. 研究開発の達成度	3
2. 研究開発による成果の活用状況や発展状況.....	4
2.1. 構築されたデータベース・ソフトウェア等の活用状況.....	4
2.1.1. データベース・ソフトウェア等の継続状況や発展状況	4
2.1.2. 第三者によるデータベース・ソフトウェア等の活用事例	6
2.1.3. データベース・ソフトウェア等へのアクセス数／ダウンロード数、visit 数 ..	7
2.2. 課題終了後の研究開発成果の継続状況や発展状況.....	8
2.2.1. BIRD 終了から現在まで、BIRD で取り組んだ課題に関連した研究開発成果の 継続状況や発展状況（国内外の研究者との共同研究へ発展した等）	8
2.2.2. BIRD 終了後に発表された論文リスト.....	8
2.2.3. 研究開発成果の現在の国際的な評価・位置づけ	10
2.3. 現在の科学技術研究における研究開発成果の活用状況、発展状況のまとめ.....	10
3. 研究開発による成果の科学技術的、社会的及び経済的な効果.....	11
3.1. バイオインフォマティクス分野及びライフサイエンス分野の研究への貢献.....	11
3.1.1. バイオインフォマティクス分野の研究への貢献	11
3.1.2. ライフサイエンス分野（バイオインフォマティクス以外）や その他、科学技術分野の研究への貢献	12
3.2. 人材育成の面から参加研究者の活動状況	13
3.2.1. 研究開発に参加した研究者のキャリアアップ	13
3.3. 社会的及び経済的な効果	14
3.3.1. 研究開発成果が大学や公的研究機関、企業等で、応用に向けて 継承または発展した例	14
3.3.2. その他、研究開発成果が社会的、経済的な効果・効用につながる兆し、 可能性	14

資料編

1. 論文リスト	15
2. 主要論文の被引用回数.....	20

3.	学会招待講演・基調講演.....	23
4.	新聞発表等	26
5.	特許出願・成立	26
6.	学会賞等の受賞	26
7.	グラントの獲得実績.....	26
8.	書籍等の執筆実績	26
9.	総説の執筆実績	27
10.	参加研究者の活動状況.....	27

本調査報告書は平成 26 年 4 月に作成
表紙の代表研究者所属は平成 25 年 10 月時点

1. 研究開発による成果

1.1. 研究開発課題の目標及び新規性・独創性

ヒトゲノム解読終了後もシーケンシングその他のハイスループット実験技術の進歩により、大量かつ多様なゲノム関連データの蓄積が続いている。本課題は、遺伝子・タンパク質などに関するゲノム情報と生体内物質・生体外物質に関するケミカル情報を分子ネットワークのレベルで統合し、生命システムの構築原理を理解すると同時に、そこから得られる知識を医療・創薬・環境保全に役立てることを目標とした。すでにゲノム解読で国際的に広く利用されている生命システム情報統合データベース (KEGG; Kyoto Encyclopedia of Genes and Genomes) に、新たに生体外物質との関連情報を蓄積し、生命システムの機能を生体内だけでなく、環境との相互作用も含めて理解することを目指した。

これまで生体内分子に限られていた KEGG PATHWAY のネットワーク配線図に生体外分子とのつながりを含め、ゲノム情報とケミカル情報との関連、及び生命システムと環境との相互作用に関する基盤情報をインターフェースとともに提供し、また、細胞・個体・エコシステムといった様々な高次の知識を KEGG BRITE の機能階層に取り入れ、ゲノムの機能解読と有効利用を促進するために KEGG DISEASE 及び KEGG DRUG に疾患と医薬品に関する知識を集積することで、ゲノムの機能解読と有効利用を促進した。

1.2. 研究開発開始時の課題を取り巻く研究・技術水準及び分野における課題の位置づけ

日常的な病気は様々な遺伝因子と環境因子が複雑に絡み合った多因子性疾患である。ヒトの健康や病気を理解するには、これらの因子を含む生体システムが安定した状態にあるのか、揺らいだ状態にあるのかを決めるシステム的な観点でのアプローチが必要である。ヒトゲノムの解読以降、マイクロアレイ実験によるトランスクリプトームの網羅的データ取得が盛んに行われ、データの意味付けのための遺伝子ネットワーク構築やタンパク質相互作用ネットワーク解析、シグナル伝達ネットワーク解析等、ネットワークやグラフ理論を関係付けたインタラクトーム解析が世界的に流行した。その結果多数のパスウェイ情報データベースや相互作用解析手法が開発されてきた。

そのような中で、KEGG は代謝に関するネットワークのデータベースとして位置づけられ発展してきた。本研究開始時点で KEGG は、ゲノム情報とケミカル情報に関する大量データから、生命システムと環境との相互作用を理解し、医療・創薬・環境保全に役立てることを目的とした生命システム情報統合データベースとして位置づけられていた。当時の KEGG はゲノム情報、ケミカル情報、システム情報を統合したデータベースとして、高次生命システム機能に関する知識を人手で集約し、広い意味での分子ネットワークとしてコンピュータ化していた。同時に全ゲノム配列が決定されたすべての生物種に対して KEGG 独自のアノテーションを行い、ゲノムと生命システムの関連づけを行っていた。本研究開

始時に KEGG はすでにゲノムをはじめとした大量データの統合処理と生物学的意味解釈のための標準的なリソースとして国際的に広く利用されていた。

本研究では、KEGG の有用性をさらに高めるためのデータ、データ構造およびシステムの高度化を進めた。

1.3. 研究開発終了時の成果概要

本研究開発期間中に、全ゲノムの高精度配列が決定された生物種の数 は 4 倍となり、またメタゲノムや個人ゲノムのデータも大量に蓄積されるようになった。これら分子レベルの大量データから、細胞・個体・エコシステムレベルでの生命システム機能を解釈し、医療・創薬・環境保全等において有用性を見いだすには計算処理技術の向上とともに、レファレンスとなる知識ベースの整備が不可欠であった。このような知識ベースへのニーズに対応した本研究の主要な成果は以下の通りであった。

- 1) KEGG の中核である KEGG PATHWAY に階層的な概念を導入し、従来型パスウェイマップより上位レベルで代謝の全体像などを提供するグローバルマップと、下位レベルでより細かな機能単位を定義した KEGG MODULE の開発を行った。
- 2) ゲノム情報と生命システムをつなぐ KEGG ORTHOLOGY を着実に改良し、新たなアノテーションツール KOALA (Kegg Orthology And Links Annotation) の開発を行った。これによって、アノテーション手続きの自動化を進め、解読されたすべての生物種に対応するゲノムデータだけでなく、メタゲノムなどの新しいデータにも対応できるようにした。
- 3) ヒト疾患に関する分子レベルの知識集約では、がんをはじめとした疾患パスウェイマップを多数追加し、疾患と病因遺伝子・病原体ゲノムとの関連を KEGG DISEASE 疾患エントリとしてコンピュータ化した。ゆらぎ状態の分子ネットワーク (疾患パスウェイマップ) 表現とともに、分子ネットワークが未知の大多数の疾患情報をコンピュータ化するため、単一遺伝子疾患・多因子性疾患と病因遺伝子との関連、感染症疾患と病原体ゲノムとの関連を蓄積した。
- 4) 化学構造をベースに日米欧の医薬品を網羅的に蓄積した KEGG DRUG では、ターゲット、薬物代謝酵素、薬物間相互作用といった分子ネットワーク情報の集積に重点をおき、また薬効分類をはじめとした KEGG BRITE 機能階層の充実と KEGG DRUG の日本語化を行った。これによって、ターゲット分子、薬物代謝酵素をはじめ、分子間相互作用に関する知識を集約した。
- 5) ゲノムと環境との相互作用の 1 つの側面として、植物・微生物による二次代謝物質合成経路と、微生物による環境物質分解経路の知識集約を行い、またこれら物質の

化学構造を KEGG BRITE 機能分類とともに蓄積した。結果として、環境物質については、内分泌攪乱 物質等の情報収集と微生物による環境物質分解経路の知識集約を行うことができた。

6) 大量データ解析ツールとして KEGG Mapper と KEGG Atlas を開発した。

7) KEGG と NCBI の緊密な連携体制を構築した。

これらの結果、KEGG ウェブサイトには月間 15 万～20 万のユニークビジターがあり、2 年ごとに発表している Nucleic Acids Research 誌 Database Issue (2010 年) での KEGG 論文全体の引用回数は年間 1,000 件に達した。本研究により、KEGG はゲノム、メタボローム、メタゲノムをはじめとした大量データを解釈するためのレファレンスとして、我が国を代表する国際標準データベースに発展させることができた。

1.4. 研究開発の達成度

本研究開発計画には大きく 2 つの内容があった。1 つは基盤的な部分でゲノムから高次生命システム機能を解読すること、もう 1 つは応用的な部分でゲノムから医療・創薬・環境保全につながる有用性を見いだすことである。本研究はこの 2 つの部分で、知識のレファレンスとしての KEGG の高度化を行った。いずれも達成度の程度差はあるが、計画内容の項目としてはすべてを達成した。とくに基盤的な部分は、高次のパスウェイマップと BRITE 機能階層の整備、ゲノムアノテーションの質的・量的向上、ゲノム数の増加への対応など、当初の研究計画は十分に達成した。一方、応用的な部分では達成度にかかなりの差が残っているものの、おおむね当初の研究計画は達成した。医薬品情報については当初計画以上の進捗があり、KEGG DRUG は医薬品の全体像を提供できるリソースに発展した。また、疾患情報については実際に KEGG DISEASE エントリ化できた疾患数は少ないが知識集約の枠組み作りとしては達成した。環境物質情報は微生物による分解経路や植物による合成経路との関連で蓄積したが、ヒトの健康や疾患に影響を与える物質の全体像を与えるようなところまでは到達できなかった。

わが国発の重要なデータベースとして、国内外の広い分野の生物学者に利用されており、関連分野の発展に大きな影響を与えた。次世代シーケンサーに代表されるハイスループット実験技術の進歩に伴い、多数の生物種ゲノム、個人ゲノム、メタゲノムなどの配列データを迅速かつ大量に決定することが可能となり、そこに含まれる遺伝子の機能、さらにはその総体としての高次生命システム機能を自動的に解読する情報技術の必要性が高まっていた。KEGG はこのような進歩とニーズに迅速に対応し、ゲノム機能解読の国際標準として広く利用され、ライフサイエンス分野全体に貢献した。また KEGG を利用したソフトウェアや二次的なデータベースも多数存在し、バイオインフォマティクス領域の推進にも貢献した。さらに KEGG は創薬等の応用の方向に展開し、メタゲノムなど考慮したパスウ

エイ解析手法の基盤が作成され、生体を個体レベルを越えた形で捉える基盤を与えた。疾患とパスウェイの関係もさらに発展させ、基礎、応用両面での利用価値を高めた。

急増するゲノム等の大量データから生命のシステムを理解し、医療・創薬・環境保全に役立っており、高次生命機能に関する知識をパスウェイマップやオントロジーで表現し、分子レベルのゲノム情報やケミカル情報と統合したデータベースとして認知された。また、ゲノム解読の最先端リソースとして国際的に既に広く利用された。KEGG DRUG、KEGG DISEASE など、今回の開発を通じてさらに生命科学の発展に大きく寄与することが期待された。

2. 研究開発による成果の活用状況や発展状況

2.1. 構築されたデータベース・ソフトウェア等の活用状況

2.1.1. データベース・ソフトウェア等の継続状況や発展状況

分類	No	データベース	URL	公開状況	備考
終了時の公開データベース					
システム情報	1	KEGG PATHWAY	http://www.genome.jp/kegg/pathway.html	無料公開中	代謝、様々な細胞プロセス、ヒトの病気などをタンパク質間相互作用ネットワークとして表現したパスウェイマップ；文献等から手作業で入力
	2	KEGG BRITE	http://www.genome.jp/kegg/brite.html	無料公開中	生命システムの様々な側面に関する機能分類とオントロジー；文献等から手作業で入力
	3	KEGG MODULE	http://www.genome.jp/kegg/module.html	無料公開中	パスウェイやコンプレックスの機能単位；手作業で入力
ヘルス情報	4	KEGG DISEASE	http://www.genome.jp/kegg/disease/disease_ja.html	無料公開中	病因遺伝子・分子のリスト；文献等から手作業で入力
	5	KEGG DRUG	http://www.genome.jp/kegg/drug/drug_ja.html	無料公開中	日米欧で承認された医薬品の化学構造と関連情報；文献等から手作業で入力
	6	KEGG EDRUG	KEGG ENVIRON に名称変更		

分類	No	データベース	URL	公開状況	備考
ゲノム情報	7	KEGG ORTHOLOGY	http://www.genome.jp/kegg/ko.html	無料公開中	KO オーソロググループ; PATHWAY と BRITE に基づき定義
	8	KEGG GENOME	http://www.genome.jp/kegg/genome.html	無料公開中	ゲノムマップと生物種に関する情報; RefSeq 等の公共データベースより作成
	9	KEGG GENES	http://www.genome.jp/kegg/genes.html	無料公開中	全塩基配列が決定されたゲノムの遺伝子カタログと手作業アノテーション; RefSeq 等の公共データベースより作成
	10	KEGG DGENE	http://www.genome.jp/dbget-bin/www_bfind?dgenes	無料公開中	ドラフトゲノムの遺伝子カタログと自動アノテーション; ウェブのリソースより作成
	11	KEGG EGENES	http://www.genome.jp/dbget-bin/www_bfind?egenes	無料公開中	EST のコンセンサスコンティグとして作られた遺伝子カタログと自動アノテーション; dbEST より作成
	12	KEGG MGENOME	http://www.genome.jp/dbget-bin/www_bfind?mggenome	無料公開中	メタゲノム配列が決定された環境サンプルのデータベース
	13	KEGG MGENE	http://www.genome.jp/dbget-bin/www_bfind?mggenes	無料公開中	メタゲノムの遺伝子カタログと自動アノテーション; NCBI のリソースより作成
ケミカル情報	14	KEGG COMPOUND	http://www.genome.jp/kegg/compound/	無料公開中	代謝物質、環境物質など低分子化合物; 文献等から手作業で入力
	15	KEGG GLYCAN	http://www.genome.jp/kegg/glycan/	無料公開中	糖鎖; 文献等から手作業で入力
	16	KEGG REACTION	http://www.genome.jp/kegg/reaction/	無料公開中	生体内化学反応; ENZYME と PATHWAY から手作業で入力
	17	KEGG RPAIR	http://www.genome.jp/dbget-bin/www_bfind?rpair	無料公開中	化学反応に伴う化学構造変化パターン; REACTION から手作業で入力
	18	KEGG RCLASS	http://www.genome.jp/kegg-bin/get_htext?br08204.keg	無料公開中	メインペアの化学構造変化パターンで定義した化学反応クラス; RPAIR から自動変換後、アノテーションを付加
	19	KEGG ENZYME	http://www.genome.jp/dbget-bin/www_bfind?enzyme	無料公開中	酵素; ExplorEnz から自動変換後、アノテーションを付加

分類	No	データベース	URL	公開状況	備考
研究開発終了報告書に記載のない関連データベース					
青 報 ゲ ノ ム	20	KEGG SSDB	http://www.kegg.jp/kegg/ssdb/	無料 公開中	ゲノム比較による遺伝子間の配列類似とベストヒット関係; GENES の計算処理で自動生成
青 報 ヘル ス	21	KEGG ENVIRON	http://www.genome.jp/kegg/drug/envIRON.html	無料 公開中	生薬、精油、薬草など有用天然物の化学成分と関連情報; 文献等から手作業で入力

KEGG はゲノム、メタボローム、メタゲノムをはじめとした大量データを解釈するためのレファレンスとして、我が国を代表する国際標準データベースに発展してきた。KEGG では知識をシステム情報、ヘルス情報、ゲノム情報、ケミカル情報の 4 つのカテゴリに分類し、手作業によるデータの選別と入力を中心としてデータベースの開発を進めている。人手によるデータの蓄積だけでなく、自動アノテーション KAAS と手動アノテーション KOALA GFIT)以外にも、データを検索・解析するためのツール群を開発しデータベースの構築効率化を進めている。

2.1.2. 第三者によるデータベース・ソフトウェア等の活用事例

Nature、Science をはじめ多数の学術雑誌で KEGG を活用した研究成果が報告されている (KEGG 論文の年間引用回数は 1500 回以上)。

例えば、種々のタンパク質複合体を形成し、遺伝子転写を調節することができる LIM ドメインタンパク質である FHL2 について、マウス肝臓において FHL2 により調節された遺伝子発現プロファイルをマイクロアレイで調べた研究で、発現量に差がある遺伝子を KEGG を用いて解析している論文 (Ng, C-F ら、J. Cell. Biochem., 2014) や、腹部大動脈瘤の形成メカニズム解明のため、大動脈の頸部標本の 2 倍の発現量の差がある遺伝子を KEGG を用いてパスウェイのマイクロアレイ解析を行った論文 (Biros, E ら、Atherosclerosis, 2014) などの論文がある。

2.1.3. データベース・ソフトウェア等へのアクセス数／ダウンロード数、visit 数

	ツール名	アクセス数、ダウンロード数、visit 数	集計期間	備考
(1)	KEGG	アクセス数合計 943,105,086	2012 年度	
		visitor 数 (月間ユニーク IP 数) 合計 3,805,346 平均 32 万	2012 年度	
		月間アクセス数 7,000 万～1 億	2013 年度	
		月間 visitor 数 (ユニーク IP 数) 平均 47 万	2013 年度	

月別の visitor 数 (ユニーク IP 数)

年月	国外	国内	合計	備考
2010.04	127,968	29,149	157,117	BIRD 最終年度
2010.05	123,482	29,440	152,922	
2010.06	114,967	32,158	147,125	
2010.07	104,534	40,578	145,112	
2010.08	102,312	36,856	139,168	
2010.09	112,247	37,237	149,484	
2010.10	127,921	42,841	170,762	
2010.11	130,946	43,213	174,159	
2010.12	116,895	37,893	154,788	
2011.01	118,516	42,393	160,909	
2011.02	132,802	41,920	174,722	
2011.03	147,165	41,270	188,435	
2011.04	132,626	48,225	180,851	
2011.05	142,516	59,838	202,354	
2011.06	138,211	67,312	205,523	
2011.07	130,020	61,220	191,240	
2011.08	130,947	57,734	188,681	
2011.09	144,581	59,924	204,505	
2011.10	152,500	62,860	215,360	
2011.11	161,276	61,659	222,935	
2011.12	148,116	56,498	204,614	
2012.01	144,869	58,171	203,040	
2012.02	153,818	61,207	215,025	
2012.03	171,721	56,336	228,057	

2012.04	161,877	46,788	208,665	
2012.05	169,424	56,366	225,790	
2012.06	150,975	64,761	215,736	
2012.07	147,509	73,115	220,624	
2012.08	149,754	79,492	229,246	
2012.09	171,655	112,858	284,513	
2012.10	215,571	143,728	359,299	
2012.11	222,244	147,710	369,954	
2012.12	203,876	139,702	343,578	
2013.01	248,777	178,207	424,984	
2013.02	253,195	178,643	431,838	
2013.03	292,338	198,781	491,119	
2013.04	291,716	204,735	496,451	
2013.05	274,236	231,267	505,503	
2013.06	249,227	226,092	475,319	
2013.07	236,096	219,612	455,708	
2013.08	226,833	192,791	419,624	
2013.09	238,112	210,935	449,047	
2013.10	275,964	221,565	497,529	

2. 2. 課題終了後の研究開発成果の継続状況や発展状況

2. 2. 1. BIRD 終了から現在まで、BIRD で取り組んだ課題に関連した研究開発成果の継続状況や発展状況（国内外の研究者との共同研究へ発展した等）

BIRD 終了後も KEGG データベース開発は継続・発展している。KEGG の中で疾患・医薬品に関する新規開発部分（KEGG MEDICUS）は統合化推進プログラムの支援を受けているが、全体的には国からの予算は BIRD の時期の 4 割程度になり、ライセンス収入等の自己努力でデータベースを維持している。

2. 2. 2. BIRD 終了後に発表された論文リスト

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42, D199-205 (2014).

Kanehisa, M.; Automated interpretation of metabolic capacity from genome and metagenome sequences. Quant. Biol. 1, 192-200 (2013).

Kanehisa, M.; Chemical and genomic evolution of enzyme-catalyzed reaction networks. FEBS Lett. 587, 2731-2737 (2013).

Muto, A., Kotera, M., Tokimatsu, T., Nakagawa, Z., Goto, S., and Kanehisa, M.; Modular architecture of metabolic pathways revealed by conserved sequences of reactions. *J. Chem. Inf. Model.* 53, 613-622 (2013).

Kanehisa, M.; Molecular network analysis of diseases and drugs in KEGG. In "Data Mining for Systems Biology" (Mamitsuka, H., DeLisi, C., and Kanehisa, M., eds.), Humana Press, Methods Mol. Biol. 939, 263-275 (2013).

Nakaya, A., Katayama, T., Itoh, M., Hiranuka, K., Kawashima, S., Moriya, Y., Okuda, S., Tanaka, M., Tokimatsu, T., Yamanishi, Y., Yoshizawa, A.C., Kanehisa, M., and Goto, S.; KEGGOC: a large-scale automatic construction of taxonomy-based ortholog clusters. *Nucleic Acids Res.* 41, D353-D357 (2013).

Kotera, M., Yamanishi, Y., Moriya, Y., Kanehisa, M., and Goto, S.; GENIES: gene network inference engine based on supervised analysis. *Nucleic Acids Res.* 40, W162-W167 (2012).

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M.; KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Res.* 40, D109-D114 (2012)

Takarabe, M., Shigemizu, D., Kotera, M., Goto, S., and Kanehisa, M.; Network-based analysis and characterization of adverse drug-drug interactions. *J. Chem. Inf. Model.* 51, 2977-2985 (2011).

Kotera, M., Tokimatsu, T., Kanehisa, M., and Goto, S.; MUCHA: multiple chemical alignment algorithm to obtain building block substructures of orphan metabolites. *BMC Bioinformatics* 12, Suppl 14, S1 (2011).

金久 實「ゲノムと疾患・医薬品の統合データベース」細胞工学 31(3), 374-377 (2012).

Kotera, M., Moriya, Y., Tokimatsu, T., Kanehisa, M., and Goto, S.; KEGG and GenomeNet, new developments, metagenomic analysis. In "Encyclopedia of Metagenomics (Nelson, K.E., ed.), Springer (2012).

Tanabe, M. and Kanehisa, M.; Using the KEGG database resource. In "Current Protocols in Bioinformatics", Unit 1.12, Wiley (2012).

Kotera, M., Hirakawa, M., Tokimatsu, T., Goto, S., and Kanehisa, M.; The KEGG databases and tools facilitating omics analysis: latest developments involving human diseases and pharmaceuticals. In "Next Generation Microarray Bioinformatics" (Wang, J., Tan, A.C., and Tian, T., eds.), Humana Press, Methods Mol. Biol. 802, 19-39 (2012).

Tokimatsu, T., Kotera, M., Goto, S., and Kanehisa, M.; KEGG and GenomeNet resources for predicting protein function from omics data including KEGG PLANT resource. In "Protein Function Prediction for Omics Era" (Kihara, D., ed.), pp. 271-288, Springer (2011).

2.2.3. 研究開発成果の現在の国際的な評価・位置づけ

BIRD 終了に伴い KEGG に対する国の支援が大幅に減少したことに対して国際的に大きな反響（憂慮する声）があった。例えば *Nucleic Acids Research* 誌 2012 Database Issue (vol. 40, pages D1-D8) を参照。KEGG が国際的に最も重要なライフサイエンスデータベースの 1 つとして認められていることの反映であり、これは現在でも変わらない。論文引用回数においても、同誌 Database Issues に発表した 8 つの論文の総引用回数は 8,000 回を越え、最近は加速度的に増えていることから、KEGG の重要性は年々高まっていると思われる。

2.3. 現在の科学技術研究における研究開発成果の活用状況、発展状況のまとめ

1991 年に京都大学化学研究所金久研究室ではゲノムネットサービスを立ち上げ、東京大学医科学研究所 ヒトゲノム解析センターと協力しながら、電子メールによる BLAST、FASTA の配列検索と DBGET キーワード検索を中心とした分子生物学データベース検索サービスを開始した。1994 年から、GenBank、PIR、PDB 等、欧米各国の研究機関で構築されていたデータベースをミラーし検索サービスを提供していたが、同時に独自のデータベース開発も始め、1994 年から LinkDB、1995 年から KEGG としての提供をゲノムネットとして行い、順次発展した。KEGG では独自のデータベース開発とデータ検索や閲覧のための各種ツール群の開発を行い、ゲノムネットでは KEGG も含めて世界各国のデータベース検索サービスおよび各種データ解析ツール群の開発を行っている。KEGG では当初、代謝系を中心とした PATHWAY データベースと関連するゲノム情報、ケミカル情報のデータベースとして開発が始まったが、医療や医薬品開発への応用も視野に入れた拡張と改良も進めてきている。2002 年には PATHWAY のカテゴリーとして「ヒト疾患」を追加し、アルツハイマー病などの疾患に係わるタンパク質相互作用ネットワークのデータベースを開発し始めた。現在は、がん、免疫系疾患、神経変性疾患、循環器疾患、代謝疾患、感染症疾患の 6 カテゴリーに分類された 30 以上のパスウェイから構成されている。また、2008 年には DISEASE データベースも開発を始め、パスウェイとして表現される疾患以外にも、疾患の内容、病因遺伝子、環境要因、診断マーカー、治療薬などの情報を文献情報とともにデータベース化している。

一方、医薬品に関しても、化学構造を中心とした DRUG データベースと医薬品開発の歴史をパスウェイの形で表したものを 2005 年から提供している。さらに 2007 年からは日本で市販されている医療用医薬品と一般用医薬品の添付文書情報を KEGG の PATHWAY、DRUG、GENES などとリンクづけし日本語での検索ができるようにしている。KEGG では主に研究者向けのデータベース開発を行ってきたが、日本語での検索および医療の現場において使われるデータを扱うようにしたことによって、薬局や病院などの医療現場での利用も増えてきた。また、一般の利用も増えており、利用者が処方された医薬品同士の相

相互作用を検索できるようにした「おくすり手帖」のサービスもはじめている。疾患・医薬品データに副作用情報なども加え高度化すると同時に、環境物質や生薬を含めた社会的ニーズの高いデータを、ゲノム情報・生命システム情報と統合化することにより、今後は、一般の人々にも有用なリソースの提供をめざしていく予定である。その手はじめとして KEGG MEDICUS を公開している。

3. 研究開発による成果の科学技術的、社会的及び経済的な効果

3.1. バイオインフォマティクス分野及びライフサイエンス分野の研究への貢献

3.1.1. バイオインフォマティクス分野の研究への貢献

特許は取得していないが、KEGG は公開されており、研究者に広く利用され、日本を代表する国際標準データベースとなっている。KEGG はコンピュータ化が遅れている高次生命システム機能に焦点を当て、独自のパスウェイデータベースで国際的な地位を獲得している。わが国発の重要なデータベースとして、国内外の広い分野の生物学者に利用されており、関連分野の発展に大きな影響を与え続けている。次世代シーケンサーに代表されるハイスループット実験技術の進歩に伴い、多数の生物種ゲノム、個人ゲノム、メタゲノムなどの配列データを迅速かつ大量に決定することが可能となり、そこに含まれる遺伝子の機能、さらにはその総体としての高次生命システム機能を自動的に解読する情報技術の必要性が高まっている。KEGG はこのような進歩とニーズに迅速に対応し、ゲノム機能解読の国際標準として広く利用され、ライフサイエンス分野全体に貢献している。また KEGG を利用したソフトウェアや二次的なデータベースも多数存在し、バイオインフォマティクス領域の推進にも貢献している。さらに KEGG は創薬等の応用の方向に展開し、メタゲノムなど考慮したパスウェイ解析手法の基盤が作成され、生体を個体レベルを越えた形で捉える基盤を与えている。疾患とパスウェイの関係もさらに発展させ、基礎、応用両面での利用価値を高めている。特に、国際的には米国 NCBI がライフサイエンス分野の統合データベースとして不動の地位を確立しており、文献情報データベース PubMed が存在することも大きな強みとなっている。KEGG は、NCBI にある大量データの生物学的意味解釈に参考文献として利用されていることから、相互に膨大な数のリンクづけがなされている。欧州の UniProt と KEGG GENES の間では相互にリンクづけがなされている。また代謝系から出発した KEGG は、酵素の EC 番号づけを行っている IUBMB/IUPAC 生化学命名委員会にも参加し、生化学・分子生物学データの標準化にも貢献している。KEGG は、国内外の同種のデータベースの中では、最も内容と機能が充実し、動向をリードするような研究開発である。ゲノムと疾患、化合物に関わるデータベースの開発と構築、公開は基礎研

究および創薬研究において大きな意義があり、その拡充とスピードが大いに期待される。

また、本課題終了後も研究開発成果を継続・発展させることにより、とくに統合化推進プログラムの中でトランスレーショナルバイオインフォマティクスの開発を行うことにより、世界で最も先進的なバイオインフォマティクスリソースとして貢献を続けている。

3.1.2. ライフサイエンス分野（バイオインフォマティクス以外）やその他、科学技術分野の研究への貢献

急増するゲノム等の大量データから生命のシステムを理解し、医療・創薬・環境保全に役立っており、高次生命機能に関する知識をパスウェイマップやオントロジーで表現し、分子レベルのゲノム情報やケミカル情報と統合したデータベースとして認知されている。また、ゲノム解読の最先端リソースとして国際的に既に広く利用されている。KEGG DRUG、KEGG DISEASE など、今回の開発を通じてさらに生命科学の発展に大きく寄与することが期待される。

国際協力と国際競争をうまく融合させる形で、我が国を代表する国立遺伝学研究所のDDBJと大阪大学蛋白質研究所のPDBJ、およびKEGGの3つのデータベース間の緊密な連携体制作りにも着手しており、今後の日本のデータベースの大きな柱になるに違いない。全ゲノムの高精度配列が決定された生物種の数急増し、またメタゲノムや個人ゲノムのデータも大量に蓄積されるようになった。これら分子レベルの大量データから、細胞・個体・エコシステムレベルでの生命システム機能を解釈し、医療・創薬・環境保全等での有用性を見いだすには、計算処理技術の向上とともに、レファレンスとなる知識ベースの整備が不可欠である。KEGGはこのような知識ベースへのニーズに以下のように対応している。

- 1) KEGGの中核であるKEGG PATHWAYに階層的な概念を導入し、従来型パスウェイマップより上位レベルで代謝の全体像などを提供するグローバルマップと、下位レベルでより細かな機能単位を定義したKEGG MODULEの開発を行っている。
- 2) ゲノムと生命システムをつなぐKEGG ORTHOLOGY(KO)を着実に改良し、新たなアノテーションツールKOALAの開発を行った。
- 3) ヒト疾患に関する分子レベルの知識集約では、がんをはじめとした疾患パスウェイマップを多数追加し、疾患と病因遺伝子・病原体ゲノムとの関連をKEGG DISEASE疾患エントリとしてコンピュータ化した。
- 4) 化学構造をベースに日米欧の医薬品を網羅的に蓄積したKEGG DRUGでは、ターゲット、薬物代謝酵素、薬物間相互作用といった分子ネットワーク情報の集積に重点をおき、また薬効分類をはじめとしたKEGG BRITE機能階層の充実とKEGG DRUGの日本語化を行った。

- 5) ゲノムと環境との相互作用の 1 つの側面として、植物・微生物による二次代謝物質合成経路と、微生物による環境 物質分解経路の知識集約を行い、またこれら物質の化学構造を KEGG BRITE 機能分類とともに蓄積した。
- 6) 大量データ解析ツールとして KEGG Mapper と KEGG Atlas を開発した。
- 7) KEGG と NCBI の緊密な連携体制を構築した。

これらの結果、KEGG ウェブサイトには月間 15 万～20 万のユニークビジターがあり、2 年ごとに発表している NAR 誌 DB 特集号での KEGG 論文全体の引用回数は年間 1,000 件に達している。

2012 年の Nucleic Acids Res.の論文 KEGG for integration and interpretation of large-scale molecular data sets は 2013 年の一年ですでに 79 件引用されている。また、2010 年出版の Nucleic Acids Res.論文、KEGG for representation and analysis of molecular networks involving diseases and drugs は 2012 年 12 月末までに海外中心にのべ 516 回引用され、2008 年出版の Nucleic Acids Res.論文 KEGG for linking genomes to life and the environment は同じく海外中心に平成 12 年末までにのべ 768 回引用されている。

一方で、トランスレーショナルバイオインフォマティクスのリソースは、研究者が研究成果を実用化するために有用だけでなく、医療従事者や一般の人々が疾患・医薬品に関する科学的知識を得るためにも有用であり、最先端研究と社会との架け橋の役割をしている。ウェブのアクセス統計にある通り、国内からのビジター数は BIRD 終了時の 4 万から現在の 20 万以上へ 5 倍以上に増加した。そのほとんどは Google 等のウェブ検索エンジンからの直接アクセスで、社会で広く利用されていることを示している。

3.2. 人材育成の面から参加研究者の活動状況

3.2.1. 研究開発に参加した研究者のキャリアアップ

本研究開発課題に参加した者で、その後新たなポジションを得た者は 5 名いた。

3.3. 社会的及び経済的な効果

3.3.1. 研究開発成果が大学や公的研究機関、企業等で、応用に向けて継承または発展した例

KEGG を利用した二次的なデータベースやウェブサービスは世界中で数多く存在する。また 100 社近い企業が KEGG のライセンスを取得し、製品開発やサービス提供に利用している。

3.3.2. その他、研究開発成果が社会的、経済的な効果・効用につながる兆し、可能性

次世代シーケンサーの普及により、高精度全ゲノム配列が決定された生物種の数 は 2011 年 3 月 31 日の時点では 1,446 であったが、2013 年 11 月 13 日の現時点では 2,892 へと倍増している。またメタゲノムといった新しいタイプのデータも多数出現している。研究開発成果は継続・発展させているので、これらに十分対応できている。また、日本で市販されている医療用医薬品と一般用医薬品の添付文書情報を KEGG の PATHWAY、DRUG、GENES などとリンクづけし、日本語での検索ができるようにしている。研究者向けのデータベース開発だけでなく、日本語での検索および医療の現場において使われるデータを扱うことによって、薬局や病院などの医療現場での利用も増えている。また、一般の利用も増えており、利用者が処方された医薬品同士の相互作用を検索できるようにした「おくすり手帖」のサービスもはじめている。疾患・医薬品データに副作用情報なども加え高度化すると同時に、環境物質や生薬を含めた社会的ニーズの高いデータを、ゲノム情報・生命システム情報と統合化することにより、今後は、一般の人々にも有用なリソースの提供へと発展して行く勢いである。

資料編

1. 論文リスト

1. # Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199-205 (2014).
2. Kanehisa, M.; Automated interpretation of metabolic capacity from genome and metagenome sequences. *Quant. Biol.* 1, 192-200 (2013).
3. # Kanehisa, M.; Chemical and genomic evolution of enzyme-catalyzed reaction networks. *FEBS Lett.* 587, 2731-2737 (2013).
4. # Muto, A., Kotera, M., Tokimatsu, T., Nakagawa, Z., Goto, S., and Kanehisa, M.; Modular architecture of metabolic pathways revealed by conserved sequences of reactions. *J. Chem. Inf. Model.* 53, 613-622 (2013).
5. Kanehisa, M.; Molecular network analysis of diseases and drugs in KEGG. In "Data Mining for Systems Biology" (Mamitsuka, H., DeLisi, C., and Kanehisa, M., eds.), Humana Press, *Methods Mol. Biol.* 939, 263-275 (2013).
6. Nakaya, A., Katayama, T., Itoh, M., Hiranuka, K., Kawashima, S., Moriya, Y., Okuda, S., Tanaka, M., Tokimatsu, T., Yamanishi, Y., Yoshizawa, A.C., Kanehisa, M., and Goto, S.; KEGG OC: a large-scale automatic construction of taxon
7. Kotera, M., Yamanishi, Y., Moriya, Y., Kanehisa, M., and Goto, S.; GENIES: gene network inference engine based on supervised analysis. *Nucleic Acids Res.* 40, W162-W167 (2012).
8. # Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M.; KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Res.* 40, D109-D114 (2012)
9. # Takarabe, M., Shigemizu, D., Kotera, M., Goto, S., and Kanehisa, M.; Network-based analysis and characterization of adverse drug-drug interactions. *J. Chem. Inf. Model.* 51, 2977-2985 (2011).
10. Kotera, M., Tokimatsu, T., Kanehisa, M., and Goto, S.; MUCHA: multiple chemical alignment algorithm to obtain building block substructures of orphan metabolites. *BMC Bioinformatics* 12, Suppl 14, S1 (2011).
11. Mizutani, S., Tanaka, M., Wheelock, C., Kanehisa, M., and Goto, S.; Phylogenetic analysis of lipid mediator GPCRs. *Genome Informatics* 24, 116-126 (2010).
12. Nishimura, Y., Tokimatsu, T., Kotera, M., Goto, S., and Kanehisa, M.; Genome-wide

analysis of plant UGT family based on sequence and substrate information. *Genome Informatics* 24, 127-138 (2010).

13. 金久 實「ゲノムと疾患・医薬品の統合データベース」細胞工学 31(3), 374-377 (2012).
14. 金久 實、伊藤真純、奥田修二郎、五斗 進、太田福子「医薬品の統合データベース」蛋白質 核酸 酵素 52(12), 1486-1491 (2007).
15. Kotera, M., Moriya, Y., Tokimatsu, T., Kanehisa, M., and Goto, S.; KEGG and GenomeNet, new developments, metagenomic analysis. In "Encyclopedia of Metagenomics (Nelson, K.E., ed.), Springer (2012).
16. Tanabe, M. and Kanehisa, M.; Using the KEGG database resource. In "Current Protocols in Bioinformatics", Unit 1.12, Wiley (2012).
17. Kotera, M., Hirakawa, M., Tokimatsu, T., Goto, S., and Kanehisa, M.; The KEGG databases and tools facilitating omics analysis: latest developments involving human diseases and pharmaceuticals. In "Next Generation Microarray Bioi
18. Tokimatsu, T., Kotera, M., Goto, S., and Kanehisa, M.; KEGG and GenomeNet resources for predicting protein function from omics data including KEGG PLANT resource. In "Protein Function Prediction for Omics Era" (Kihara, D., ed.),
19. Erguner, B., Hattori, M., Goto, S., and Kanehisa, M.; Characterizing common substructures of ligands for GPCR protein subfamilies. *Genome Informatics* 24, 31-41 (2010). (査読あり)
20. Kotera, M., Kobayashi, T., Hattori, M., Tokimatsu, T., Goto, S., Mihara, H., and Kanehisa, M.; Comprehensive genomic analysis of sulfur-relay pathway genes. *Genome Informatics* 24, 104-115 (2010). (査読あり)
21. # Yamanishi, Y., Kotera, M., Kanehisa, M., and Goto, S.; Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 26, i246-i254 (2010). (査読あり)
22. # Hattori, M., Tanaka, N., Kanehisa, M., and Goto, S.; SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. *Nucleic Acids Res.* 38, W652-W656 (2010). (査読あり)
23. # Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Kotera, M., Goto, S., and Kanehisa, M.; PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res.* 38, W138-W143 (2010). (査読あり)
24. # * 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.* 38, D355-D360 (2010). (査読あり)
25. Takarabe, M., Shigemizu, D., Kotera, M., Goto, S., and Kanehisa, M.; Characterization and classification of adverse drug interactions. *Genome*

- Informatics 22, 167-175 (2009). (査読あり)
26. Yamanishi, Y., Hattori, M., Kotera, M., Goto, S., and Kanehisa, M.; E-zyne: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs. *Bioinformatics* 25, i79-i86 (2009). (査読あり)
 27. * 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). (査読あり)
 28. # 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. *Carbohydr. Res.* 344, 881-887 (2009). (査読あり)
 29. Shimizu, Y., Hattori, M., Goto, S., and Kanehisa, M.; Generalized reaction patterns for prediction of unknown enzymatic reactions. *Genome Informatics* 20, 149-158 (2008). (査読あり)
 30. 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). (査読あり)
 31. # Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M.; Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24, i232-i240 (2008). (査読あり)
 32. # 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). (査読あり)
 33. Kojima, K.K. and Kanehisa, M.; Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture. *Mol. Biol. Evol.* 25, 1395-1404 (2008). (査読あり)
 34. # * 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). (査読あり)
 35. # * 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). (査読あり)
 36. Kadowaki T, Wheelock CE, Adachi T, Kudo T, Okamoto S, Tanaka N, Tonomura K, Tsujimoto G, Mamitsuka H, Goto S, Kanehisa M., Identification of endocrine disruptor biodegradation by integration of structure-activity relationship with pathway analysis., *Environ Sci Technol.* 41(23), 7997-8003, 2007

37. 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). (査読あり)
38. 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). (査読あり)
39. # * 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). (査読あり)
40. # 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). (査読あり)
41. # 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). (査読あり)
42. # 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). (査読あり)
43. Okamoto, S., Yamanishi, Y., Ehira, S., Kawashima, S., Tonomura, K., and Kanehisa, M.; Prediction of nitrogen metabolism-related genes in *Anabaena* by kernel-based network analysis. *Proteomics* 7, 900-909 (2007). (査読あり)
44. Okuda, S., Kawashima, S., Kobayashi, K., Ogasawara, N., Kanehisa, M., and Goto, S.; Characterization of relationships between transcriptional units and operon structures in *Bacillus subtilis* and *Escherichia coli*. *BMC Genomics* 8,
45. Hashimoto, K., Yoshizawa, A.C., Saito, K., Yamada, T., and Kanehisa, M.; The repertoire of desaturases for unsaturated fatty acid synthesis in 397 genomes. *Genome Informatics* 17(1), 173-183 (2006). (査読あり)
46. Honda, W., Kawashima, S., and Kanehisa, M.; Metabolite antigens and pathway incompatibility. *Genome Informatics* 17(1), 184-194 (2006). (査読あり)
47. Tanaka, M., Yamada, T., Itoh, M., Okuda, S., Goto, S., and Kanehisa M.; Analysis of the differences in metabolic network expansion between prokaryotes and eukaryotes. *Genome Informatics* 17(1), 230-239 (2006). (査読あり)
48. # 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). (査読あり)

49. 本多 渉、田辺麻央、矢野亜津子、金久 實「バイオインフォマティクス・システムバ イオロジーと KEGG」生化学 80(12), 1094-1111 (2008).
50. 橋本浩介、五斗 進、金久 實「KEGG GLYCAN データベースと糖鎖インフォマティクス」蛋白質 核酸 酵素 53(12), 1698-1702 (2008).
51. 橋本浩介、五斗 進、金久 實「KEGG: 生命システム情報統合データベース」実験医学 26(7), 1142-1147 (2008).

: 本追跡調査において研究代表者が主要な論文として指定したもの（上限 30 報）

* : 研究開発期間終了後の終了報告書において研究代表者が主要な論文として指定したもの

2. 主要論文の被引用回数

論文	国内外別 件数		分野別件数																						出版年別件数							
	国内 件数	海外 件数	農学	生物学・ 生化学	化学	臨床 医学	コンピ ュータ サイエ ンス	経済 学・ 経営 学	工学	環 境・ 生態 学	地球 科学	免疫 学	材料 科学	数学	微生物 学	分子 生物学・ 遺伝 学	複合 領域	神経 科学・ 行動 科学	薬理 学・ 毒物 学	物理 学	植物 学・ 動物 学	精神 医学・ 心理 学	社会 科学・ 一般	宇宙 科学	2006 年	2007 年	2008 年	2009 年	2010 年	2011 年	2012 年	
8. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M.; KEGG for integration and interpretation of large-scale molecular datasets. Nucleic Acids Res. 40, D109–D114 (2012)	10	69	2	17	3	16	8	0	0	2	0	1	0	0	5	20	0	0	3	1	1	0	0	0	—	—	—	—	—	—	—	79
9. Takarabe, M., Shigemizu, D., Kotera, M., Goto, S., and Kanehisa, M.; Network-based analysis and characterization of adverse drug-drug interactions. J. Chem. Inf. Model. 51, 2977–2985 (2011).	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	—	—	—	—	—	0	2	
21. Yamanishi, Y., Kotera, M., Kanehisa, M., and Goto, S.; Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. Bioinformatics 26, i246–i254 (2010). (査読あり)	3	18	0	8	1	4	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	—	—	—	0	7	14	
22. Hattori, M., Tanaka, N., Kanehisa, M., and Goto, S.; SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. Nucleic Acids Res. 38, W652–W656 (2010). (査読あり)	0	3	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	—	—	—	0	1	2	
23. Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Kotera, M., Goto, S., and Kanehisa, M.; PathPred: an enzyme-catalyzed metabolic pathway prediction server. Nucleic Acids Res. 38, W138–W143 (2010). (査読あり)	3	9	0	4	2	0	3	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	—	—	—	—	1	4	7	
24.* 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. 38, D355–D360 (2010). (査読あり)	35	481	4	148	24	99	45	0	2	1	0	4	0	11	28	103	2	7	12	2	10	1	2	0	—	—	—	—	57	231	228	
28. 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. Carbohydr. Res. 344, 881–887 (2009). (査読あり)	4	7	0	5	1	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	—	—	—	1	4	2	4	
31. Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M.; Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 24, i232–i240 (2008). (査読あり)	6	39	0	12	6	4	10	0	0	1	0	0	0	0	0	3	0	0	7	1	0	0	0	0	—	—	1	6	6	16	16	
32. 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). (査読あり)	11	90	2	32	2	12	13	0	1	1	0	0	0	0	4	19	0	0	5	0	7	0	0	0	—	—	3	22	33	23	20	
33. Kojima, K.K. and Kanehisa, M.; Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture. Mol. Biol. Evol. 25, 1395–1404 (2008). (査読あり)	0	5	0	3	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	—	—	1	1	0	2	1	
34. * 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).(査読あり)	73	695	3	203	26	103	96	0	6	8	1	6	1	11	70	157	0	4	15	1	29	1	0	2	—	—	55	237	150	189	137	
35. * 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). (査読あり)	2	22	0	7	1	2	0	0	0	0	0	0	0	0	0	3	0	0	2	0	7	1	1	0	—	—	4	3	7	4	6	
39. * 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). (査読あり)	5	22	1	10	2	3	4	0	0	1	0	0	0	0	0	3	0	0	0	0	1	0	0	0	—	0	7	6	6	3	5	

論文	国内外別件数		分野別件数																						出版年別件数						
	国内件数	海外件数	農学	生物学・生化学	化学	臨床医学	コンピュータサイエンス	経済学・経営学	工学	環境・生態学	地球科学	免疫学	材料科学	数学	微生物学	分子生物学・遺伝学	複合領域	神経科学・行動科学	薬理学・毒物学	物理学	植物学・動物学	精神医学・心理学	社会科学・一般	宇宙科学	2006年	2007年	2008年	2009年	2010年	2011年	2012年
40. 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). (査読あり)	13	136	0	25	0	19	18	0	0	4	1	0	1	2	27	39	0	0	1	0	12	0	0	0	—	0	9	16	25	38	61
41. 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). (査読あり)	7	16	0	9	4	0	4	0	0	1	0	0	0	0	2	2	0	0	0	0	1	0	0	0	—	1	4	8	7	1	2
42. 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). (査読あり)	2	25	0	13	4	1	2	0	0	1	0	0	0	0	5	0	0	0	0	0	1	0	0	0	—	1	11	8	1	3	3
48. 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). (査読あり)	5	25	0	11	2	0	1	0	0	0	0	0	0	0	0	8	0	0	1	0	6	0	0	0	0	7	5	9	4	3	2

- ・本追跡調査において研究代表者が主要な論文として指定したもの（上限 30 報）について、トムソン・ロイター社 Web of Science で調査した。調査対象は、2012 年 12 月まで。被引用情報が取得できたものののみ記載した。
- ・論文番号は「資料編 1. 論文リスト」に対応している。
- ・＊：研究開発期間終了後の終了報告書において研究代表者が主要な論文として指定したもの。
- ・[国内外別件数] は、被引用文献の国内外別件数。被引用文献の著者の所属国のうち JAPAN が一つでもあれば国内としている。国内外の合計が全被引用文献数となる。
- ・[分野別件数] は、被引用文献の分野別件数。被引用文献が論文の場合のみカウントしている。
- ・[出版年別件数] は、被引用文献の出版年別件数。「—」は、データなしを表す。

論文	Web of Science 国内外別件数(2012 年 12 月まで)			Google Scholar	
	国内件数	海外件数	合計	～2014 年 4 月 24 日	～2012 年
1. Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42, D199–205 (2014).	—	—	—	12	—
3. Kanehisa, M.; Chemical and genomic evolution of enzyme-catalyzed reaction networks. FEBS Lett. 587, 2731–2737 (2013).	—	—	—	2	—
4. Muto, A., Kotera, M., Tokimatsu, T., Nakagawa, Z., Goto, S., and Kanehisa, M.; Modular architecture of metabolic pathways revealed by conserved sequences of reactions. J. Chem. Inf. Model. 53, 613–622 (2013).	—	—	—	7	—
8. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M.; KEGG for integration and interpretation of large-scale molecular datasets. Nucleic Acids Res. 40, D109–D114 (2012)	10	69	79	979	504
9. Takarabe, M., Shigemizu, D., Kotera, M., Goto, S., and Kanehisa, M.; Network-based analysis and characterization of adverse drug-drug interactions. J. Chem. Inf. Model. 51, 2977–2985 (2011).	1	1	2	13	3
21. Yamanishi, Y., Kotera, M., Kanehisa, M., and Goto, S.; Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. Bioinformatics 26, i246–i254 (2010). (査読あり)	3	18	21	88	42

22. Hattori, M., Tanaka, N., Kanehisa, M., and Goto, S.; SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. Nucleic Acids Res. 38, W652–W656 (2010). (査読あり)	0	3	3	17	9
23. Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Kotera, M., Goto, S., and Kanehisa, M.; PathPred: an enzyme-catalyzed metabolic pathway prediction server. Nucleic Acids Res. 38, W138–W143 (2010). (査読あり)	3	9	12	44	28
24.* 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. 38, D355–D360 (2010). (査読あり)	35	481	516	1200	905
28. 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. Carbohydr. Res. 344, 881–887 (2009). (査読あり)	4	7	11	21	18
31. Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M.; Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 24, i232–i240 (2008). (査読あり)	6	39	45	199	116
32. 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). (査読あり)	11	90	101	214	178
33. Kojima, K.K. and Kanehisa, M.; Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture. Mol. Biol. Evol. 25, 1395–1404 (2008). (査読あり)	0	5	5	13	8
34. * 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).(査読あり)	73	695	763	1712	1426
35. * 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). (査読あり)	2	22	24	44	31
39. * Limviphuwadh, 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). (査読あり)	5	22	27	47	43
40. 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). (査読あり)	13	136	149	477	255
41. 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). (査読あり)	7	16	23	46	38
42. 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). (査読あり)	2	25	27	54	39
48. 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). (査読あり)	5	25	30	39	36

- ・ 情報科学・計算機科学分野における研究開発成果では、会議予稿集での発表に引用されることが多いため、Google Scholar での被引用件数を参考情報として添付した。
- ・ Google での調査方法
 - 対象論文のタイトルを検索し、当該論文を「引用元」とする件数を取得した。
 - 最新の被引用件数と、それから 2013 年以降の被引用件数を引いたものを 2012 年までとした。ただし、年が明確でないものは引かれていないため、2012 年までの被引用件数が実際よりも多くなっている可能性がある。
- ・ 論文番号は「資料編 1. 論文リスト」に対応している。
- ・ *：研究開発期間終了後の終了報告書において研究代表者が主要な論文として指定したもの。

3. 学会招待講演・基調講演

1. Kanehisa, M.; Dual aspect of metabolism revealed by KEGG orthology and reaction classification systems. Bioinformatics Research Infrastructure for the Life Sciences: The First 10 Years of UniProt, Basel, Switzerland (2012/9/8).
2. Kanehisa, M.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Practical Course "Bioinformatics: Computer Methods in Molecular Biology", Trieste, Italy (2012/6/27)
3. Kanehisa, M.; KEGG MEDICUS and LinkDB. FDA Workshop on Disease Outbreak Detection in the Genomics Era, Arlington, USA (2012/3/1).
4. Kanehisa, M.; Bench to bedside: changing roles of bioinformatics. 10th InCoB - 1st ISCB Asia Joint Conference, Kuala Lumpur, Malaysia (2011/11/30).
5. Kanehisa, M.; Translational bioinformatics: from genomes to health and society. Cold Spring Harbor Asia Conference on Bioinformatics of human and animal genomics, Suzhou, China (2011/11/14).
6. Kanehisa, M.; Changing roles of bioinformatics. ICR Symposium to Celebrate the Bioinformatics Center's 10 Year Anniversary, Kyoto, Japan (2011/8/29).
7. Kanehisa, M.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Practical Course "Bioinformatics: Computer Methods in Molecular Biology", Trieste, Italy (2011/6/22).
8. Kanehisa, M.; Use of chemical logic to decipher the genome in KEGG. IUB International Workshop on The Eukaryotic Genome, Brno, Czech (2010/10/16).
9. Kanehisa, M.; Molecular network based annotation of genomes and metagenomes. Biocuration 2010: Conference of the International Society for Biocuration, Tokyo, Japan (2010/10/11).
10. Kanehisa, M.; Developing a computable disease information resource in KEGG. Cold Spring Harbor Asia Conference on Computational Biology, Suzhou, China (2010/9/29).
11. Kanehisa, M.; Molecular network analysis for linking genomes to natural products, drugs and xenobiotics. ABRC seminar, Taipei, Taiwan (2010/7/12).
12. Kanehisa, M.; Molecular network analysis for linking plant genomes to natural products and drugs. 2nd Banff Conference on Plant Metabolism, Banff, Canada (2010/6/25).
13. Kanehisa, M.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Practical Course Bioinformatics: Computer Methods in Molecular Biology, Trieste, Italy (2010/6/21).

14. Kanehisa, M.; Predicting biodegradation potential and degradation pathways of EDCs. American Society for Microbiology 110th General Meeting / Division Q Symposium, San Diego, USA (2010/5/26).
15. Kanehisa, M.; From metagenomes to pathways and diseases. First International Conference on Human Metagenomics, Shenzhen, China (2010/3/3).
16. Kanehisa, M.; Representation and analysis of molecular networks involving diseases and drugs. 20th International Conference on Genome Informatics, Yokohama, Japan (2009/12/16).
17. Kanehisa, M.; Linking genomic knowledge to natural products and drugs. International Symposium on Traditional Medicine in Toyama, Toyama, Japan (2009/11/26).
18. Kanehisa, M.; KEGG: Kyoto Encyclopedia of Genes and Genomes. 5th International Conference of the Brazilian Association for Bioinformatics and Computational Biology, Angra dos Reis, Brazil (2009/10/18).
19. Kanehisa, M.; Re-evaluation of the EC number system from a genomic perspective. 4th International Symposium on Experimental Standard Conditions on Enzyme Characterizations, Rüdesheim, Germany (2009/9/16).
20. Kanehisa, M.; Linking genomic knowledge to natural products and drugs. 238th American Chemical Society National Meeting / Division of Computers in Chemistry Symposium, Washington DC, USA (2009/8/17).
21. Kanehisa, M.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Practical Course Bioinformatics: Computer Methods in Molecular Biology, Trieste, Italy (2009/6/24).
22. Kanehisa, M.; Linking genomic knowledge to diseases and drugs. 38th Annual Meeting of the Brazilian Biochemistry and Molecular Biology Society, Águas de Lindóia, Brazil (2009/5/19).
23. Kanehisa, M.; KEGG for knowledge based analysis of molecular systems in human diseases. Frontiers in Multi-Scale Systems Biology, Atlanta, USA (2008/10/20).
24. Kanehisa, M.; Bioinformatics approaches to integrated analysis of genomic and chemical information. Metabolomics 2008: 4th Annual Conference of the Metabolomics Society, Boston, USA (2008/9/6).
25. Kanehisa, M.; KEGG for plant genomics and metabolomics. ICPM 2008: 5th International Conference on Plant Metabolomics, Yokohama, Japan (2008/7/16).
26. Kanehisa, M.; KEGG for medical and pharmaceutical applications. APBC 2008: 6th Asia-Pacific Bioinformatics Conference, Kyoto, Japan (2008/1/17).
27. Kanehisa, M.; Bioinformatics for drugs and diseases. JSBi Annual Meeting, Tokyo, Japan (2007/12/17).

28. Kanehisa, M.; Bioinformatics approach to chemical genomics. 6th Georgia Tech - ORNL International Conference on Bioinformatics, Atlanta, USA (2007/11/16).
29. Kanehisa, M.; Genomic and chemical annotations in KEGG. 2nd International Biocuration Meeting, San Jose, USA (2007/10/27).
30. Kanehisa, M.; From genomes to life and the environment. BioBase Anniversary Meeting, Braunschweig, Germany (2007/10/16).
31. Kanehisa, M.; Bioinformatics, Chemoinformatics, and Pharmacoinformatics. CBI 2007, Hiroshima, Japan (2007/10/4).
32. Kanehisa, M.; Genome to life and environment. InCoB 2007: 6th International Conference on Bioinformatics, HongKong, China (2007/8/29).
33. Kanehisa, M.; Bioinformatics methods to integrate genomic and chemical information. Summer School in Beijing, Beijing, China (2007/8/13).
34. Kanehisa, M.; Linking genomes to biological systems and environments. Two Decades of Predictive Biology, Boston, USA (2007/3/17).
35. Kanehisa, M.; Deciphering biosynthetic codes of glycans and other biological molecules. Gordon Research Conference on Glycobiology, Ventura, USA (2007/3/7).
36. Kanehisa, M.; Bioinformatics education integrating basic sciences and practical applications. International Symposium on Bioinformatics Education and Research, Tokyo, Japan (2006/12/17).
37. Kanehisa, M.; KEGG BRITE for linking genomes to biological systems and environments. Société Française de Génétique Annual Meeting on Genetics meets Systems Biology: Theory and Practice, Paris, France (2006/11/14).
38. Kanehisa, M.; KEGG BRITE for sharing knowledge in the glycomics community. Frontiers in Glycomics: Bioinformatics and Biomarkers in Disease, Bethesda, USA (2006/9/11).
39. Kanehisa, M.; Community annotation for sharing knowledge. ASM/NCBI Microbial Genome Annotation Workshop, Bethesda, USA (2006/8/28).
40. Kanehisa, M.; KEGG PATHWAY and BRITE for linking genomes to biological systems. 20th IUBMB / 11th FAOBMB Congress, Kyoto, Japan (2006/6/23).

4. 新聞発表等

1. 薬情報サイト開設に関する新聞報道（著作権の関係により非公開）
2. トムソン社 Biology & Biochemistry 分野での Fast Breaking Paper（引用回数の伸びが最も速い論文）に選ばれ、インタビュー記事（2007 年 2 月）
3. トムソン社 ScienceWatch の Emerging Research Fronts にインタビュー記事（2008 年 2 月）
4. GenomeWeb 社の BioInform にインタビュー記事（2008 年 1 月）

5. 特許出願・成立

該当なし

6. 学会賞等の受賞

該当なし

7. グラントの獲得実績

1. 科学技術振興機構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「ゲノム情報に基づく疾患・医薬品・環境物質データの統合」（平成 23～25 年度）

8. 書籍等の執筆実績

1. Mamitsuka, H., DeLisi, C., and Kanehisa, M. (eds); “Data Mining for Systems Biology”. Humana Press (2013).

9. 総説の執筆実績

1. 金久 實「ゲノムと疾患・医薬品の統合データベース」細胞工学 31(3), pp. 374-377 (2012).
2. 五斗 進、金久 實「KEGG とゲノムネットにおける医薬品と生命情報の統合」竹田・志鷹真由子、梅山秀明編「次世代創薬テクノロジー／実践：インシリコ創薬の最前線」遺伝子医学 MOOK 14, pp. 168-174 (2009).
3. 本多 渉、田辺麻央、矢野亜津子、金久 實「バイオインフォマティクス・システムバイオロジーと KEGG」生化学 80(12), 1094-1111 (2008).
4. 橋本浩介、五斗 進、金久 實「KEGG GLYCAN データベースと糖鎖インフォマティクス」蛋白質 核酸 酵素 53(12), 1698-1702 (2008).
5. 橋本浩介、五斗 進、金久 實「KEGG: 生命システム情報統合データベース」実験医学 26(7), 1142-1147 (2008).
6. 金久 實、伊藤真純、奥田修二郎、五斗 進、太田福子「医薬品の統合データベース」蛋白質 核酸 酵素 52(12), 1486-1491 (2007).
7. 荒木通啓、金久 實「創薬科学のための KEGG」化学工業 57(3), 42-46 (2006).

10. 参加研究者の活動状況（個人情報が含まれるため非公開）