

平成26年度
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

京都大学化学研究所

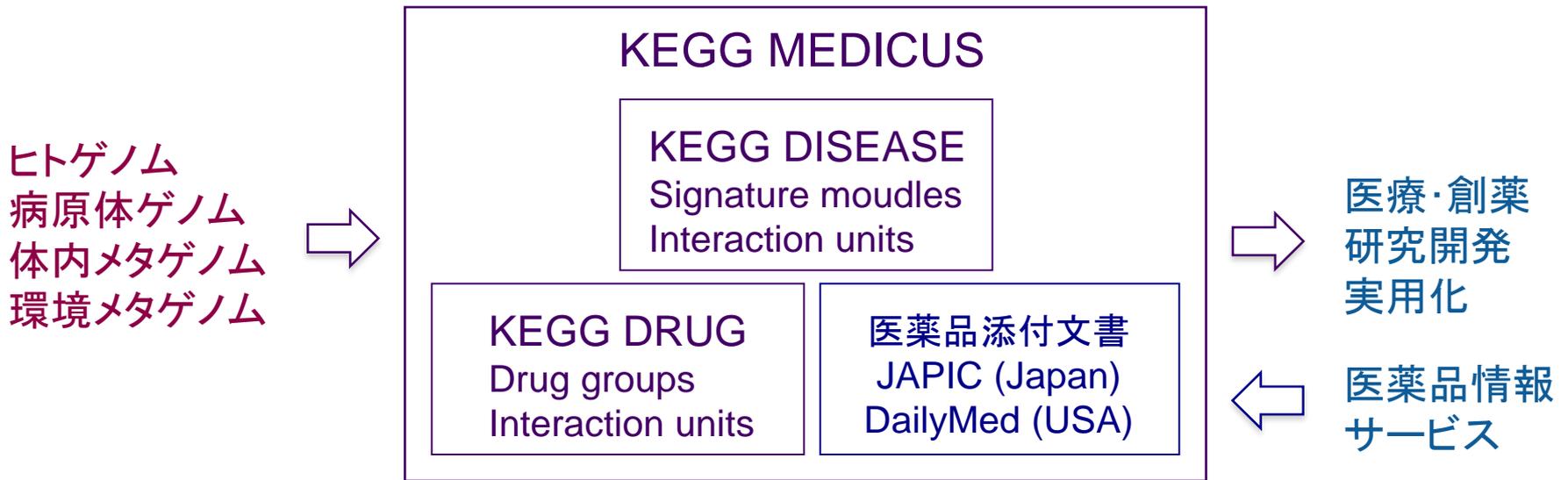
金久 實

2014年6月2日 キックオフミーティング



©2014 金久 實(京都大学) licensed under CC表示2.1日本

研究開発の目標・ねらい



- ゲノム情報を有効利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 KEGG MEDICUS をさらに発展させた開発を行う。
H23-H25: 研究者コミュニティと社会との架け橋
H26-H28: 研究者コミュニティのトランスレーショナル研究を推進
- ゲノムとフェノタイプとの関連、とくに疾患との関連を理解するために、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する。

ゲノムデータの社会的価値

ヒトゲノム	医薬品の作用・副作用（個別化医療）
病原体ゲノム	病原性・薬剤耐性
メタゲノム	？

遺伝子セット

Reprogramming

OCT4 (K09367)
SOX2 (K16796)
MYC (K04377)
KLF4 (K17846)

Breast cancer

BRCA1 (K10605)
BRCA2 (K08775)

Methicillin resistance

mecR1 (K02547)
mecI (K02546)
mecA (K02545)

KEGG module M00625

モジュールと相互作用ユニット

KEGG モジュール

M00095 Mevalonate pathway
K00626 K01641 **K00021** K00869
(K00938,K13273) K01597 K01823

Disease 相互作用ユニット

H00155 Hypercholesterolemia
K12473 LDLR
C00187 Cholesterol
D02258 Atorvastatin

Drug 相互作用ユニット

D02258 Atorvastatin
K00021 HMGCR
K17689 CYP3A4
K05043 SLCO1B
K12473 LDLR

map00900 テルペノイド骨格の生合成

K00021 HMG-CoA 還元酵素

map00100 ステロイドの生合成

C00187 コレステロール

map00120 一次胆汁酸の生合成

map00121 二次胆汁酸の生合成

map04976 胆汁の分泌

K17689 CYP3A4

代謝酵素

標的分子

K05043 SLCO1B

トランスポーター

D02258 アトルバスタチン(リピトール)

治療薬

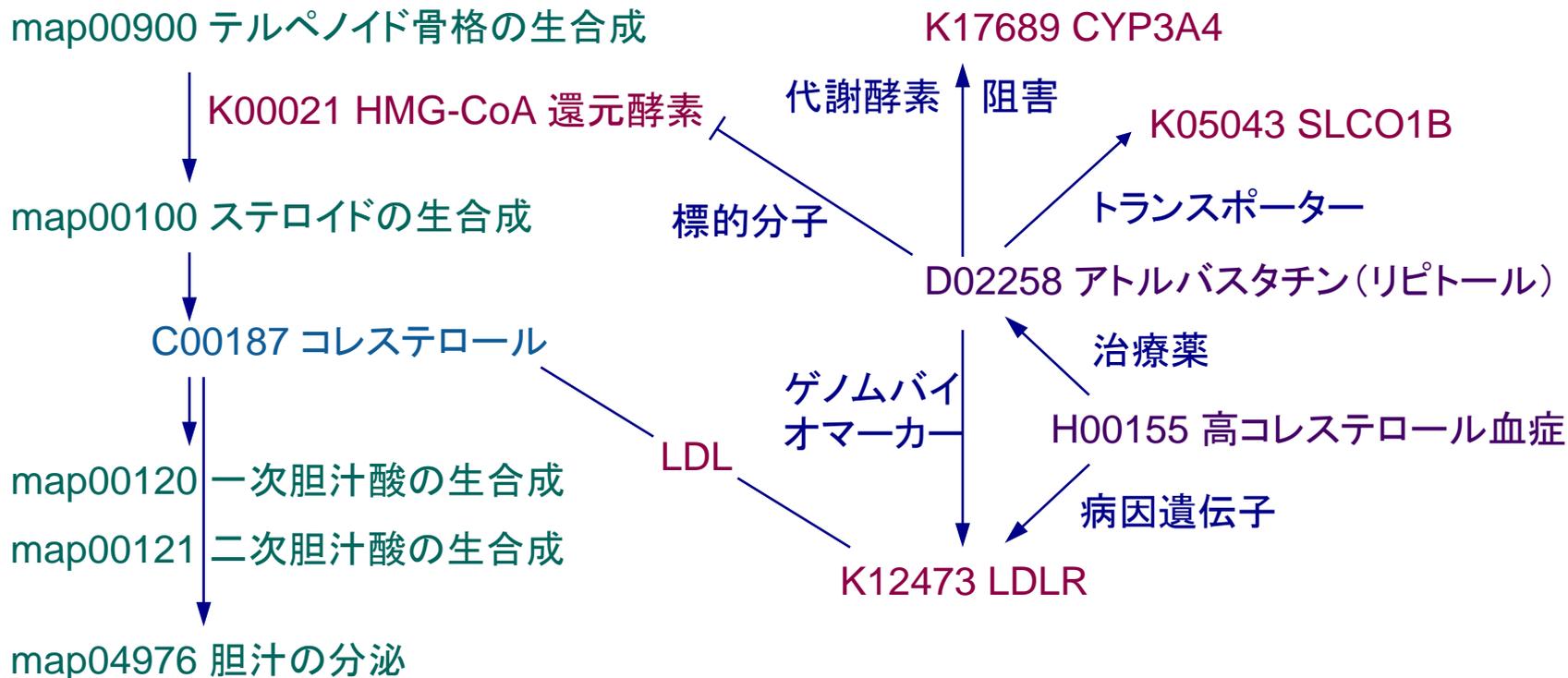
ゲノムバイ
オマーカー

H00155 高コレステロール血症

病因遺伝子

LDL

K12473 LDLR



疾患情報のデータベース: 疾患エントリ

(1-1) ゲノム(ヒトゲノム、病原体ゲノム、体内細菌叢メタゲノムなど)と疾患をつなぐ KEGG DISEASE の内容を充実させる

KEGG DISEASE 疾患エントリの内容

ENTRY	H番号識別子
NAME	疾患名、別名を含む
DESCRIPTION	疾患の簡単な説明文
CATEGORY	疾患カテゴリ
PATHWAY	KEGGパスウェイへのリンク
GENE	既知の病因遺伝子
CARCINOGEN	発癌物質
ENV_FACTOR	その他の環境因子
PATHOGEN	病原体
GENE	病原性・薬剤耐性のモジュールあるいは遺伝子
MICROBIOME	疾患に関連したマイクロビオームの特徴
MARKER	代表的な診断マーカー
DRUG	代表的な治療薬
COMMENT	コメント
DBLINKS	外部データベースへのリンク
REFERENCE	文献情報

疾患情報のデータベース：病原性・薬剤耐性のモジュール

(1-2) ゲノムから病原性や薬剤耐性等のフェノタイプを特徴づける遺伝子セット (モジュール)を集積する

EHEC pathogenicity signature, Shiga toxin
M00363 = K11006+K11007

	Organism	1 K K11006 (stxA)[19]	2 K K11007 (stxB)[19]
B.Gam	M ece	Z3344(stx1A) Z1464(stx2A)	Z1465(stx2B) Z3343(stx1B)
B.Gam	M ecs	ECs2974 ECs1205	ECs1206 ECs2973
B.Gam	M ecf	ECH74115_2906 ECH74115_3533	ECH74115_3532 ECH74115_2905
B.Gam	M etw	ECSP_2723(stx2A) ECSP_3253(stx2A)	ECSP_3252(stx2B) ECSP_2722(stx2c)
B.Gam	M elx	CDCO157_2739 CDCO157_1154	CDCO157_1155 CDCO157_2738
B.Gam	M eoj	ECO26_1599	ECO26_1600
B.Gam	M eoi	ECO111_3361 ECO111_2430	ECO111_2429 ECO111_3362
B.Gam	M eoh	ECO103_5197 ECO103_2845	ECO103_2844 ECO103_5198
B.Gam	M eso	O3O_09165	O3O_09170
B.Gam	M esm	O3M_16110	O3M_16105
B.Gam	M esl	O3K_16135	O3K_16130
B.Gam	M sdy	SDY_1389(stxA)	SDY_1390(stxB)

Methicillin resistance
M00625 = K02547 K02546 K02545

	Organism	1 K K02547 (mecR1)[14]	2 K K02546 (mecI)[14]	3 K K02545* (mecA)[14]
B.Fir	M sau	SA0039(mecR1)	SA0040(mecI)	SA0038(mecA)
B.Fir	M sav	SAV0042(mecR1)	SAV0043(mecI)	SAV0041(mecA)
B.Fir	M saw	SAHV_0041(mecR1)	SAHV_0042(mecI)	SAHV_0040(mecA)
B.Fir	M sah	SaurJH1_0030	SaurJH1_0031	SaurJH1_0029
B.Fir	M saj	SaurJH9_0030	SaurJH9_0031	SaurJH9_0029
B.Fir	M sar	SAR0040(mecR1)	SAR0041(mecI)	SAR0039(mecA)
B.Fir	M suk	SAA6008_00040(m	SAA6008_00041(m	SAA6008_00039(m
B.Fir	M sut	SAT0131_00035	SAT0131_00036	SAT0131_00034(r
B.Fir	M suf	SARLGA251_00270	SARLGA251_00280	SARLGA251_00260
B.Fir	M sauz	SAZ172_0073(mec	SAZ172_0074(mec	SAZ172_0072(mec
B.Fir	M suy	SA2981_0040(mec	SA2981_0041	SA2981_0039(mec
B.Fir	M sauc	CA347_40(mecR1)	CA347_41(mecI)	CA347_39(mecA)
B.Fir	M ser	SERP2520(mecR1)	SERP2519(mecI)	SERP2521(mecA)
B.Fir	M mcl	MCCL_plsB0022	MCCL_plsB0021	MCCL_plsB0023

M番号 – KEGG module の識別子
K番号 – KEGG Orthology (KO) の識別子
M番号はK番号の論理式で表現

医薬品情報のデータベース: 2つの統合

(2-1) 医薬品情報、疾患情報、ゲノム情報の統合を進め、医薬品添付文書との統合を継続する

KEGG DRUG 有効成分

- D番号識別子
- 一般名、商品名
- 化学構造その他の構造情報
- 様々な医薬品分類情報
- 商品一覧(先発品・後発品)
- 効能・効果、適応症
- 標的分子と関連パスウェイ
- 代謝酵素、トランスポーター
- CYP誘導・阻害など
- ゲノムバイオマーカー
- 投与の条件(抗悪性腫瘍薬等)
- 投与注意(poor metabolizer等)

JAPIC 医薬品添付文書

- 添付文書
(独自に処理したデータ)
- 併用禁忌・併用注意の相互作用
- 添加物

KEGG DISEASE

ヒトゲノム
病原体ゲノム

医薬品情報のデータベース: 医薬品のオーソロググループ

(2-2) 様々な意味合いで(とくに相互作用の観点から)医薬品機能が同一とみなした医薬品グループのデータベースを開発する

薬1	標的分子 代謝酵素 トランスポーター ゲノムバイオマーカー	遺伝子多型
薬2	標的分子 代謝酵素 トランスポーター ゲノムバイオマーカー	遺伝子多型
薬3	

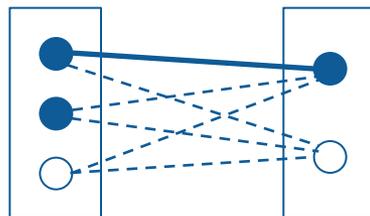
相互作用

PD: pharmacodynamic

PK: pharmacokinetic

グループの例

- 基本化学構造
- 標的分子
- CYP基質
- CYP誘導・阻害



- 添付文書がある医薬品
- 添付文書がない医薬品
- 添付文書記載の相互作用
- 推定される相互作用

グループ1 グループ2

統合インターフェース

(3) KEGG MEDICUS を利用して、ゲノムの情報から疾患・医薬品に関する知見を得るための KEGG Mapper ツールを開発する

医薬品リスト



KEGG DRUG



医薬品間相互作用
を検出(開発済み)

病原体ゲノム



KEGG MODULE



病原性、毒性
薬剤耐性を推定

個人ゲノム
腸内細菌メタゲノム



KEGG DISEASE
KEGG MODULE



疾患の理解
医薬品の開発

CYP遺伝子多型



KEGG DRUG



医薬品の作用・
副作用を推定

課題と対策

- トランスレーショナル研究の情報基盤を開発するだけでなく、我々自身でどこまで社会的価値を生み出す部分に関与するか
 - KEGG お薬手帳の開発は本プログラムでは行わない
個別化医療と感染症対策で有用なリソース開発を行う
- ゲノムと医薬品の相互作用情報を個別化医療で活用できるようにするためには、遺伝子多型の情報が必要
 - 遺伝子と多型のタイプのみ蓄積し、詳細は ClinVar など他のデータベースへのリンクとする
- 病原菌の薬剤耐性については、遺伝子の有無によるモジュールだけでは不十分で、遺伝子多型との関連が必要
 - 多型遺伝子の同定は行うが、その詳細には立ち入らない
NCBI が CDC や FDA と協力して進めている Pathogen Project と連携し、モジュールの有効性を検証

KEGG MEDICUS のデータ一覧

名称	件数	概要
KEGG DISEASE	1,359	単一遺伝子疾患・多因子性疾患と遺伝要因・環境要因との関連、感染症疾患と病原体ゲノムとの関連を蓄積した疾患データベース、本研究で新たに病原体ゲノムから病原性や薬剤耐性に関与する遺伝子セットを定義しシグネチャーモジュールとして追加
KEGG DRUG	10,084	日米欧で使われている医薬品の有効成分を網羅的に集積し、標的分子や代謝酵素など分子ネットワーク情報を付与した医薬品データベース、本研究でゲノムバイオマーカーの情報を充実
KEGG DGROUP		同一効能、類似効能、その他の医薬品グループのデータベース、本研究で新規に開発
KEGG ENVIRON	849	生薬その他健康に関与する天然物を蓄積したデータベース、各国での承認薬に限定した KEGG DRUG を補完
JAPIC 添付文書	24,913	日本の医療用・一般用医薬品添付文書（財団法人日本医薬情報センターとのライセンス契約によりウェブ公開のみ可能、FTP 公開や再配布は禁止されている）

2014年4月15日現在

- すべてはウェブで公開 (<http://www.kegg.jp/kegg/medicus/>)
- JAPIC 添付文書以外は FTP でも公開 (<ftp://ftp.genome.jp/pub/kegg/medicus/>)
- 本研究終了時には FTP データをすべて NBDC アーカイブに提供