

区 分	☐ 中核機関（ ☐ 代表機関／ ☐ 参画機関） ☒ 分担機関（ ☒ 代表機関／ ☐ 参画機関） ☐ 補完課題実施機関
課 題 名	ライフサイエンス知識の階層化・統合化事業
実 施 機 関 名	国立大学法人 京都大学
代 表 研 究 者 名	五斗 進（平成20年度～22年度）、金久 實（平成19年度）

1. 課題開始時における達成目標
すでに世界有数のバイオ情報サービスとなっているゲノムネットを京都大学の事業と位置づけ、化学研究所バイオインフォマティクスセンターにおいて分子情報を中心とした統合データベースを構築する。日本語での統合DB検索システムを提供するとともに、革新的なウェブ技術とKEGGにおいて人手で構築された知識の体系を融合して、ライフサイエンス分野における世界最高水準の知的情報基盤を確立する。
2. 平成22年10月末時点における事業計画に対する成果
(1) 成果概要 本統合データベースプロジェクトはゲノムネット (http://www.genome.jp/) をKEGGと分離して開発・運用するために提案し実施してきた。KEGG (http://www.genome.jp/kegg/) は現時点ではゲノムネットの主要サービスであるが、京都大学と東京大学の金久研究室が別予算で構築しており、KEGG独自のウェブサイト (http://www.kegg.jp/) も存在する。本プロジェクトでは、ゲノムネットを京都大学の事業と位置づけ、DBGET/LinkDBシステムを中心に統合化を行った。具体的には、<u>統合データベース開発・運用</u>で、化合物・医薬品データベースを整備するとともにそれ以外の情報も含めて統合化を進めた。この中でのKEGGは統合化対象データベースの中心という位置づけである。<u>共通基盤技術開発</u>では化学情報解析ツールの開発と化合物・医薬品データベースのための検索システムの開発を行った。化学情報解析ツールの一部はこれまでKEGGの中で開発されていたものを引き継いで進めた。 <u>統合データベース開発・運用</u> 医薬品、化合物、化学反応という観点から以下のデータベースを開発し、運用してきた。 - ゲノムネット医薬品データベース (http://www.genome.jp/kusuri/) 医薬品添付文書情報をKEGGの医薬品情報と統合して検索できるシステムを開発・運用してきた。 - ゲノムネット化合物データベース (http://www.genome.jp/ja/gn_dbget_ja.html) ゲノムネットで開発しているDBGET/LinkDBの仕組みを用いて、主に化合物関連のデータベースを統合検索できるようにした。25のデータベースについてリンク情報による統合化を実現し、そのうち12のデータベースについてはキーワード検索の対象とした。 - 反応物ペアデータベース RPAIR (http://www.genome.jp/dbget-bin/www_bfind?rpair) 化合物データベースの拡張という観点から、KEGGの反応データベースから基質と生成物のペアを抽出したデータベースを開発した。特に、ペアのアライメントを整備した。 - 化学反応パターンデータベース RCLASS (http://www.genome.jp/dbget-bin/www_bfind?rclas) RPAIRデータベースの基質・生成物ペアから反応パターンを抽出して整備した。本データは、反応経路予測ツール PathPred に応用している。 <u>共通基盤技術開発</u>

以下のツールを開発・運用してきた。

- ・ 化合物類似構造検索ツール SIMCOMP (<http://www.genome.jp/tools/simcomp/>)
グラフ比較に基づく化合物類似構造検索アルゴリズムの開発と改良を行い、高速化と高機能化を実現した。
- ・ 化合物部分構造検索ツール SUBCOMP (<http://www.genome.jp/tools/subcomp/>)
高速な化合物部分構造検索アルゴリズムを開発し、高機能化を実現した。
- ・ 化学構造変化に基づく反応予測ツール E-zyme (<http://www.genome.jp/tools/e-zyme/>)
与えられた二つの化合物間の反応を予測し、対応する酵素を探索するシステムの予測精度を改良し、公開した。
- ・ 反応パターンを利用した反応経路予測ツール PathPred (<http://www.genome.jp/tools/pathpred/>)
与えられた化合物を合成する経路や、化合物を分解する経路を予測するシステムを開発し、公開した。
- ・ 糖鎖類似構造検索ツール KCaM (<http://www.genome.jp/tools/kcam/>)
- ・ ゲノムネット統合検索システム DBGET/LinkDB (http://www.genome.jp/ja/gn_dbget_ja.html)
DBGET システムの日本語での利用環境作りとして、入力した日本語キーワードの翻訳機能と検索結果画面の英単語を日本語に翻訳するための支援ツールを開発した。

(2) 進捗及び成果

統合データベース開発運用

研究の最先端と医療の現場、さらには一般社会をつなぐ日本語の医薬品統合データベースであるゲノムネット医薬品データベースを開発し、医療用医薬品に関するデータを平成 19 年 9 月に、一般用医薬品に関するデータを平成 20 年 1 月に公開した (<http://www.genome.jp/kusuri/>) (図 1)。本データベースは日本医薬情報センター (JAPIC) が提供する医薬品添付文書情報 (平成 22 年 11 月現在、医療用医薬品 12,239 件、一般用医薬品 12,042 件、毎月更新) を、医薬品名、成分、薬効などで検索できる。ライフサイエンス辞書を利用した同義語・類義語検索もできる。各添付文書情報には KEGG DRUG 中の対応する医薬品へのリンク付を行っているため、KEGG が持つターゲットや代謝酵素とそれらが働くパスウェイ情報とするとともに、医薬品の開発過程をパスウェイで表現した情報にも容易に到達できる。また、参考文献のうち可能なものについては、文献データベース PubMed、J-STAGE、J-GLOBAL へのリンクを付加している。

ゲノムネット医薬品データベースでは、各医薬品の成分情報を抽出し対応をとっているため、同じ成分を持つ医薬品を類似商品一覧として検索することができる。また、KEGG DRUG の構造情報に基づいた類似構造検索機能を使えば、成分の化学構造が似た商品の検索も可能である。平成 22 年 6 月には医薬品の併用禁忌と併用注意の情報に基づいた相互作用情報を検索できるようにし、それを検索する一つのインタフェースとしておくすり手帖を公開した。おくすり手帖では、ユーザが検索した医薬品をブラウザの Cookie に保存しておくことにより、指定した医薬品間での相互作用があれば表示する仕組みを実装している。

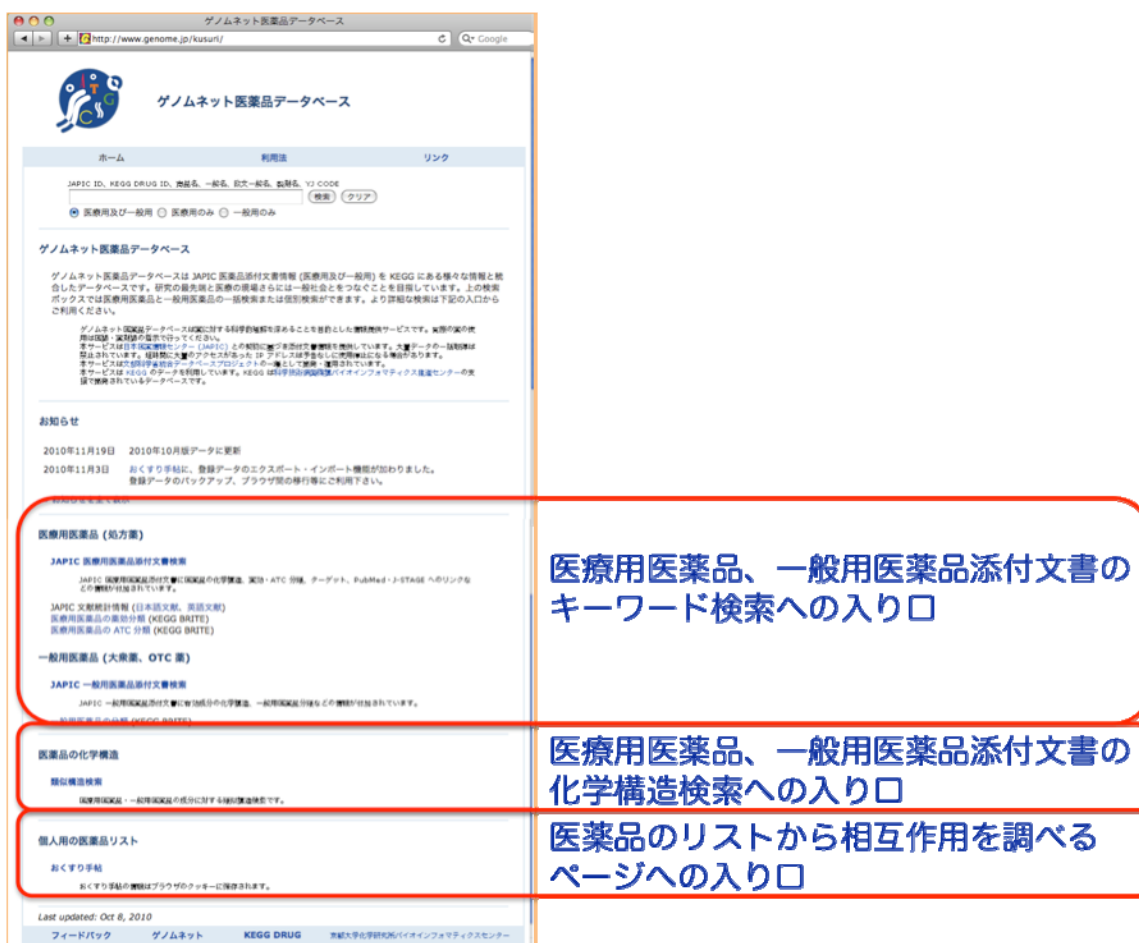


図 1：ゲノムネット医薬品データベース

ゲノムネット医薬品データベースへは平成 22 年 9 月の一ヶ月間で 18,380 ノードから約 88 万のアクセス（ロボットを含む）があった（表 1）。アクセス数は平成 19 年の公開時から順調に増えており、最近では月平均 20,000 ノードから 90 万前後のアクセスがある。アクセス元のドメインを見てみると、大学・教育関係よりも製薬企業からのアクセスが多い傾向がある。また、プロバイダ系のドメインも多い。これはロボットの影響もあるが、多数のアクセスポイントからのアクセスがあるので、不特定多数の個人（一般利用者）がアクセスしていると考えられる。

表 1. ゲノムネット医薬品データベース (<http://www.genome.jp/kusuri/>) のアクセス数

	アクセス数	ノード数		アクセス数	ノード数
2009年4月	952,436	22,346	2010年1月	730,095	21,539
2009年5月	862,742	22,745	2010年2月	729,103	20,744
2009年6月	1,006,818	22,639	2010年3月	888,723	20,752
2009年7月	967,659	27,196	2010年4月	1,142,956	19,971
2009年8月	790,645	23,732	2010年5月	891,593	18,790
2009年9月	979,522	23,017	2010年6月	872,944	19,147
2009年10月	776,307	23,913	2010年7月	782,656	19,852
2009年11月	793,245	21,813	2010年8月	881,187	18,380
2009年12月	701,599	19,253	2010年9月	861,485	21,062

国内外で開発されている化合物データベースを統合的に検索できるゲノムネット化合物データベー

スを開発し http://www.genome.jp/ja/gn_dbget_ja.html から公開した（図2）。本システムでは平成19年までに検索環境を整備した DBGET/LinkDB を化合物・化学反応データベースの統合化に応用している。その際、データベース検索の高機能化を実現するために、DBGET システムのインデクシングプログラムである SEQNEW を改良し、全データベース一括検索できるようにした。また、エントリ取得のためインデクシング方式を最新のものに変更し高速化も実現した。平成22年11月までに12のデータベースをキーワード検索対象として、25のデータベースを LinkDB 対象として組み込んだ。

全データベースへのキーワード検索
環境設定でライフサイエンス辞書を用いた翻訳サポート機能を利用することも可能

統合した化合物・医薬品データベースのうちキーワード検索できるようにした12のデータベースと関連するKEGGのデータベース

区分	データベース	内容
ゲノムネット化合物データベース	PubChem	化合物全庫
	ChEMBL	化合物オントロジー
	KEGG COMPOUND	代謝化合物, 生物体化合物
	LIPIDMAPS	脂質
	LipidBank	脂質
	KNApSack	植物二次代謝化合物
	KEGG DRUG	医薬品
	DrugBank	医薬品
	KEGG GLYCAN	糖鎖
	PDBeChem	低分子立体構造
ゲノムネット医薬品データベース	3DMET	代謝化合物立体構造モデル
	LigandBox	医薬品立体構造モデル
	KEGG REACTION	生体内化学反応
	KEGG RPAIR	反応ペア
	JAPIC	医薬品臨床試験報告書

図2：ゲノムネット化合物データベースを含む統合データベース検索システム

化学反応データベースに関しては、反応の基質と生成物のペアを抽出した RPAIR データベースの化合物アライメント情報を整備した。さらに、そこから抽出される反応パターンの情報を RCLASS データベースとして整備し、平成22年7月に公開した（2010年11月17日版で RPAIR は12,326 エントリ、RCLASS は2,296 エントリ）。

LinkDB では、これまでゲノムネットの DBGET システムで運用している内部データベースが検索の対象であった。本プロジェクトでは、それ以外のデータベース（外部データベース）についてもデータベース間のクロスリファレンス情報として定義されていることに着目し、それらを二項関係ファイルとして取り込んで検索できる枠組みを開発した。取り込むべき外部データベースの最初の対象を、DBGET システムで運用している内部データベースから参照されているデータベースとした。これにより、内部データベースに加え500以上の外部データベースが LinkDB の検索対象となった。同時に、検索コマンドの高機能化として、外部データベースと内部データベースを区別なく検索する機能および等価なリンクを扱うための機能を実装した。また、LinkDB の検索効率を高めるために、塩基配列データベース

(DDBJ, GenBank, EMBL)を対象としたデータの階層化・重複除去作業を行い、DBGET システムに於いて INSDC という一つのデータベースとして検索できるようにした。キーワード検索結果画面からは、上記 3 つのデータベースすべてにリンクが張られており、どのデータベースのエントリも直接検索できる。本 LinkDB システムの基本機能の拡張は平成 20 年までに完了し公開した。

共通基盤技術開発

統合データベースを構築する基盤技術は平成 18 年度までに KEGG プロジェクトである程度確立してきたので、本プロジェクトでは統合データベースを利用するための技術開発を中心に進めた。利用の形態は大きく分けてキーワード検索と類似構造検索や解析・予測などのソフトウェア利用とがある。共通基盤技術開発では、これら 2 つを業務項目とし計画を進めた。

キーワード検索の基本的なものはすでに **DBGET** システムで運用しているので、日本語での利用環境づくりとして、入力した日本語キーワードの翻訳機能と検索画面の英単語を日本語に翻訳するための支援ツールを開発し、平成 19 年 10 月に運用を開始した。また、ゲノムネットで提供する全データベースに対する一括検索機能を開発し、平成 19 年 7 月に運用を開始した（図 2）。その際、ゲノムネットにインストールされていないデータベースに関しては、等価なエン트리間をつなぐリンクを含む **LinkDB** の新バージョンに対する検索機能を実現した。その結果、一部の外部データベースも内部データベースと同様にキーワード検索できるようになった。化合物・医薬品データベースのキーワード検索結果には、**LinkDB** の情報を利用して、対応する **KEGG** のエントリを直接取得できるようにした。また、エントリ検索結果にも **LinkDB** のリンク情報を直接表示するようにし、関連情報を見やすくした（図 3）。

GenomeNet

Search for Go Clear

Database: Chemical substance - Search term: GDP L-Ucose

KEGG COMPOUND

C00325 [KegDraw](#) [Jmol](#)
GDP-L-ucose, GDP-beta-L-ucose

KEGG GLYCAN

G10515 [KegDraw](#)
GDP-L-ucose, (GDP-L-Fuc)1

PUBCHEM

24895171 [PubChem]
6-Deoxy-beta-L-galactosylranosylguanosine 5'-diphosphate; G4401, SIGMA; GDP-Fuc; GDP-Ucose;
Guanosine 5'-diphospho-beta-L-ucose sodium salt

26758842 [PubChem]
GDP-beta-L-ucose, Disodium Salt; Guanosine 5'-phosphate; diphosphate-beta-L-ucose, 2-Na

2782 [PubChem]
2-amino-9-D,4-dihydroxy-5-hydroxy-thiopyran-3,4,5-trihydroxy-6-methyl-tetrahydropyran-2-yl-oxo-
phosphoribosyl phosphonate; (oxymethyl)tetrahydropyran-2-yl-, 1,9-dihydroxy-5-one, GDP-L-ucose;
GUANOSINE 5'-diphospho-beta-L-ucose; FUCOSE; GUANOSINE 5'-trihydr...

3619 [PubChem] [KEGG-link](#)
GDP-L-ucose, GDP-beta-L-ucose, [CPO: C00325]

**PubChem, ChEBI, KEGG, KNApSack 等の検索結果から
対応する KEGG へのリンク**

ChEBI

17009 [ChEBI] [KEGG-link](#)
GDP-L-ucose, [CPO: C00325]

26510 [ChEBI] [KEGG-link](#)
GDP-4-deoxy-beta-L-ucose, [CPO: C05389]

13332 [ChEBI] [KEGG-link](#)
GDP-beta-L-ucose

SDMET

04668 [SDMT] [KEGG-link](#)
GDP-L-ucose, GDP-beta-L-ucose, [CPO: C00325]

KNApSack

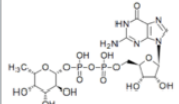
C0007245 [KNApSack] [KEGG-link](#)
GDP-L-ucose, [CPO: C00325]

KEGG

COMPOUND: C00325 Help

Entry	Compound
C00325	GDP-L-ucose
Formula	C ₁₆ H ₂₅ N ₅ O ₁₅ P ₂
Mass	569.0822

Structure



060306
[Mol file] [KEGG file] [DB search] [Jmol] [KegDraw]

Remark
Same as: G10615
[Go to history]

Reaction
R01591: R01519 R04081 R04553 R04446 R05106 R05692 R07060

Pathway
PATH0: ko00051 Fructose and mannose metabolism
PATH0: ko00520 Amino sugar and nucleotide sugar metabolism
PATH0: ko01100 Metabolic pathways

Enzyme
1.1.1.271 2.4.1.65 2.4.1.69 2.4.1.152
2.7.7.30

Other IDs
PubChem: 3619
ChEBI: 17009
KNApSack: C0007245
SDMET: R04668

KCP data Show

All links

Pathway (1062)
KEGG PATHWAY (1041)
KEGG MODULE (1)
Chemical substance (6)
KEGG GLYCAN (1)
PubChem (1)
ChEBI (12)
SDMET (13)
KNApSack (1)
KNApSack (1)
Chemical reaction (3)
KEGG reaction (7)
KEGG REACTION (8)
KEGG PATH (19)
All databases (1302)

図3. 各種化合物データベースへのキーワード検索結果から、対応するKEGGのエントリへの直接リンク（左）と、KEGGのエントリに埋め込まれたLinkDB検索結果（右）。LinkDBにはKEGGのエントリには直接書かれていないリンク（例ではMASSBANKへのリンク）も含めて一覧することができる。

平成 20 年以降は、中間評価および中核機関との調整を考慮し、高速・高機能な新規検索エンジンの開発として、化合物・医薬品により特化したものを検討した。具体的には、構造検索と他の検索を組み合わせることによる高機能化を検討し、構造検索の結果からパスウェイ情報や階層的な機能分類情報へと関連検索できるシステムを実現し、平成 21 年 4 月に公開した（図 4）。

Figure 4 shows the SIMCOMP search interface. The main window displays search results for a steroid compound. A red box highlights the '類似構造から機能の検索' (Search for function from similar structure) button. Below the table, a chemical structure of a steroid is shown with a similarity score of 1.00. To the right, a 'Pathway Search Result' window is open, showing a list of pathways. A red arrow points from the '類似構造から機能の検索' button to the 'Pathway Search Result' window. A green box at the bottom right contains text: '構造検索結果から機能の手がかりを探すためのリンク' (Link to search for function clues from structure search results), 'メタボロームなどで KEGG ない複数の化合物データから' (From multiple compound data not in KEGG, such as metabolomics).

図 4：構造検索結果から関連する機能情報を検索する機能

ゲノムネット利用形態のうちソフトウェア利用については、化合物に関するソフトウェアを集中的に整備した。特に化学研究所バイオインフォマティクスセンターの研究成果をもとにしたソフトウェアを整備し実用化し、以下に列举するツール群を平成 19 年より順次公開している。

- ・ 化合物類似構造検索ツール SIMCOMP (<http://www.genome.jp/tools/simcomp/>)
- ・ 化合物部分構造検索ツール SUBCOMP (<http://www.genome.jp/tools/subcomp/>)
- ・ 化学構造変化に基づく反応予測ツール E-zyme (<http://www.genome.jp/tools/e-zyme/>)
- ・ 反応パターンを利用した反応経路予測ツール PathPred (<http://www.genome.jp/tools/pathpred/>)
- ・ 糖鎖類似構造検索ツール KCaM (<http://www.genome.jp/tools/kcam/>)

SIMCOMP については、サイズの大きな化合物に対する検索効率の悪さが問題となっていたので、化合物構造を比較する際のグラフ表現と比較アルゴリズムについて複数の方法を調査・検討し、平成 20 年 7 月には 10 倍程度の高速化を実現したバージョンを公開した。さらに、ユーザーが与えた化合物データセット間の構造比較を全対全で行うためのインタフェース SIMCOMP2 を開発し、平成 22 年 8 月に公開した (<http://www.genome.jp/tools/simcomp2/>)。

SUBCOMP については、特定の構造に対する問題点を改良したバージョンを平成 20 年 12 月に公開した。また、クエリ化合物を含む構造の検索に加え、クエリ化合物に含まれる構造の検索機能を実装し、平成 21 年 7 月に公開した。光学異性体を区別して検索する機能は SIMCOMP、SUBCOMP ともに必要であることが分かったため、異性体の区別を実装したバージョンを、平成 21 年 4 月に公開した。これにより D 体と R 体の区別だけでなく、cis と trans の区別もできるようになった。SIMCOMP と SUBCOMP は構造データインデクシングプログラムと構造アライメント表示プログラムも含めてパッケージ化し、平成 22 年 10 月に FTP で公開した (<ftp://ftp.genome.jp/pub/tools/simcomp/>)。

E-zyme については、酵素番号の自動割り当てに使用する反応パターンの階層的な表現を用いることによる改良を実現し、平成 21 年 1 月に公開した（図 5）。さらに、入力された反応が複数の基質や生成物を持つ場合に、自動割り当てに必要な基質と生成物の 1 対 1 のペアを抽出する方法を実装し、遺伝子情報へのリンク機能を含めて平成 22 年 10 月にテスト公開した。

The screenshot shows the 'e-zyme result' page. At the top, it displays 'Pair 1 [KCF]' with chemical structures for C17248 and C08409. Below this, the 'EC number assignments' table is shown, with the first row highlighted. To the right of the table, there is a section for 'Reactions having the same RDM' with links to R03214 and R08167. A red box highlights the first row of the table, and a blue box highlights the 'Reactions having the same RDM' section. A text box points to the chemical structures, stating they are extracted from the alignment of substrate (C17248) and product (C08409). Another text box points to the table, stating it shows the assigned EC number and score based on the reaction pattern distribution in the database. A third text box points to the 'Reactions having the same RDM' section, stating it provides a link to reactions with the same pattern.

EC number	Weighted score	Observed freq.	Reactions having the same RDM
2.8.2	88.6	2	R03214 R08167
3.1.6	9.9		
1.8.4	3.0		
1.8.99	2.0		
3.6.2	2.0		
3.6.1	1.0		

Confidence level of prediction results: **RDM**

[GenomeNet | KEGG | LIGAND | e-zyme]

図 5. 新規反応に酵素番号を自動で割り当てるツールE-zymeの実行結果画面。基質と生成物をSIMCOMPでアライメントし、反応パターンを抽出したのち、データベース中から同じようなパターンを持つ反応を探し出して、酵素番号を割り当てる。平成 20 年度はデータベース中の反応パターン分布に基づくスコアリング（Weighted score）を導入し、精度を上げるとともに、データベース中の類似反応を表示するインターフェースを追加した（Reactions having the same RDM）。

PathPred は、始点と終点（または始点のみ）の化合物構造を入力し、その間の反応経路を予測するシステムとして平成 22 年 1 月に公開した（図 6）。PathPred では、環境物質の分解経路で使われている反応から抽出した反応パターン（RCLASS データ）を適用することによって、環境物質分解経路を予測できるようにしている。また、医薬品やターゲットに関連した解析の第一歩として、漢方薬などの成分として重要な植物二次代謝産物の生合成経路で使われている反応から抽出した変換ルールも用意して経路予測に応用できるようにした。また、RCLASS の応用として、類似反応経路検索システムを開発し、平成 22 年 10 月にテスト公開した。

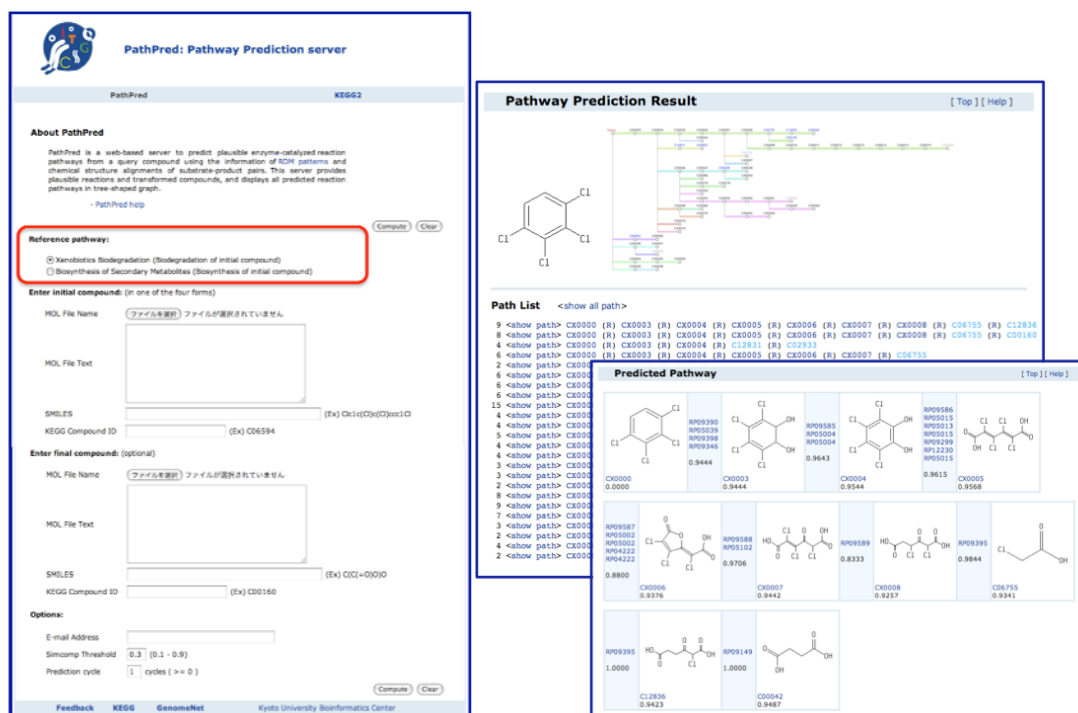


図6. 化合物の構造を入力して、その化合物が代謝される経路を計算するツールPathPredの画面。構造入力画面ではリファレンスとなる反応パターンが実際に使われている経路を選択できるようにしており、現在は微生物による分解系と植物の二次代謝産物合成系に特徴的な反応パターンを選ぶことができる（赤枠で囲った部分）。右の図はテトラクロロベンゼンをクエリ構造として計算した結果と、その中から一番上の経路を表示したもの。ベンゼン環が開裂してコハク酸（C00042）まで分解されていく様子が分かる。ここで、CXではじまる番号がついている化合物（テトラクロロベンゼンから7つ）はデータベースに登録されていないもので、中間代謝物として新たに予測された構造である。

化学情報解析ツール群をまとめた <http://www.genome.jp/tools/> には、平成22年9月の一ヶ月間で3,016ノードから約25万のアクセス（ロボットを含む）があった。

3. 当初目標に対する達成度

分子情報を中心とした統合データベースの構築はゲノムネット医薬品データベース、ゲノムネット化合物データベースとして実現し、当初目的を達成した。特に、ゲノムネット医薬品データベースは公開当初から着実にユーザーを増やし、表1に示したように、最近では月平均20,000ノードから90万前後のアクセスがある。また、ゲノムネット医薬品データベースを中心に日本語のコンテンツを充実化するとともに、日本語での検索システムも公開し、日本語での統合DB検索システムの提供についても当初の目的を達成した。特に、本システムのポスター発表などで得られたコメントからは、医薬品相互作用情報など研究者だけでなく一般ユーザーや薬剤師など医療の現場でも活用されるようになってきていると感じた。

革新的なウェブ技術に関しては、基本的なキーワード検索については中核機関との調整で初年度の拡張にとどまったが、化学情報解析ツールを整備・充実化し、ゲノムネット医薬品データベースと化合物データベースを対象とした解析を可能にした。これらをまとめたウェブサイトへのアクセスも医薬品データベースほどではないが、着実に増えており、当初の目的を達成したと考える。また、ゲノムネット全

体では、一日あたりの総アクセス数は 100 万件程度、アクセスノード数は 15,000 程度あり、国際的にも有数の情報サービスとなっている。

4. 中間評価に対する対応

中間評価への対応および中核機関との調整の結果、平成 20 年以降は、高速・高機能な新規検索エンジンの開発として、化合物・医薬品により特化したものを検討した。具体的には、構造検索と他の検索を組み合わせることによる高機能化を検討し、構造検索の結果からパスウェイ情報や階層的な機能分類情報へと関連検索できるシステムを実現した。

また、JAPIC 以外のデータとして KEGG DRUG を通したパスウェイ情報や遺伝子情報へのリンク、SIDER データベースの副作用情報へのリンクなどを実現した。KEGG のパスウェイ情報や遺伝子情報は、従来の有料化合物・医薬品データベースにはないものであり、構造検索やパスウェイ予測など他のデータベースにはない魅力ある検索方法を提案してきた。

5. 他機関との連携

本プロジェクト発足時当初の計画では、分子情報の統合データベースの構築と共に、革新的なウェブ技術を用いた統合データベース検索システムの構築を目標に掲げていた。しかし、中核機関との緊密な調整により、京都大学では、主に医薬品と化合物のデータベースに特化することとした。検索エンジンに関しても化合物・医薬品により特化したものとした。その結果、ゲノムネット医薬品・化合物データベースの充実化につながるとともに、関連解析ツールの開発も十分進めることができた。また、本プロジェクトで開発したデータベースはすべて中核機関の横断検索システムで検索できるようになっている。

化合物情報には当初、糖鎖情報も含めて開発を進めていたが、産総研糖鎖グループの参画により、糖鎖関連データベースは産総研グループが取りまとめることとなった。産総研グループとは開発当事者間で直接緊密な連携ができたため、これまでにゲノムネットで扱っていた糖鎖情報のやりとりも含めて情報交換がスムーズに進んだ。また、遺伝子やタンパク質関係のデータでは中核機関や補完課題実施機関で構築しているデータベースへのリンク付けを行うための連携も取ることができた。

6. 今後の見通し、計画、展望

医薬品・化合物データベースに関しては、有料サービスが多く、無料サービスの場合でも知財が絡む場合が多い。JAPIC も本プロジェクトのデータベースとしてサービスするにはライセンス契約が必要である。したがって、多様なデータベースの統合を進めるには、本プロジェクトで進めている LinkDB を用いたリンク情報を中心とした形態が有効であると考えられる。

また、その際にはマニュアルでの等価リンクの定義は重要であり、複数のデータベースを単純にキーワード検索できるようにするだけでなく、異なるデータベース間で同じ対象を扱っているデータ間の関係を扱えるようにしていくことは今後とも重要である。

7. 全体総括

分子レベル、特に化合物情報のデータベース統合という点とライフサイエンスの基礎分野への応用という点からは、日本だけでなく海外のデータベースを LinkDB という枠組みで統合することができるとともに様々な解析ツールの提供ができたため、目標は達成した。一方で医療への応用という観点からは、相互作用情報の取り込みなどで医療現場、特に薬剤師などへの情報提供という応用の可能性は示せたものの、副作用情報など収集が難しいデータについては不十分な点も多い。インタビューフォームなども含めた情報の充実については、中核機関（ライフサイエンスデータベースセンター）や他省庁との連携も含めて、今後の課題である。しかし、今後の見通しで述べたように、LinkDB や KEGG といったコアなコンテツをベースにして他のデータベースとの融合を図るという、一つの方向性を示せたことは大きな成果である。

8. 特記事項

本プロジェクトでは、すでに国際的に知名度の高い KEGG を中心とした LinkDB による統合によって、既存のデータベースにある様々な情報に、マニュアルで整備されたライフサイエンスの情報をリンク付けしたことが売りの一つであり、ゲノムネット医薬品データベースが成功している最大の理由である。また、JAPIC 医薬品添付文書情報の検索サービスはライセンス契約が必要なものの、新たな情報とのリンクという点から日本医薬情報センターとは緊密に連絡をとることができた。同センターとは単にデータの受け取りだけではなく、ゲノムネット医薬品データベースにインストールする際に発見したエラーを報告したり、逆にゲノムネットでの不具合を同センター経由でユーザーから受け取ったりと相互に有用な情報交換もできた。また、両者にとって新しい分野のユーザーを開拓するという利点もあった。

9. 委託研究費一覧

	18年度	19年度	20年度	21年度	22年度	計
設備備品費（千円）		35,948	0	0	0	35,948
人件費（千円）		17,228	26,557	28,249	28,277	100,311
業務実施費（千円）		128,642	82,534	47,206	32,086	290,468
一般管理費（千円）		18,182	10,909	7,545	6,037	42,673
合計（千円）		200,000	120,000	83,000	66,400	469,400

整備実績一覧

(1) データ（又はDB）の連結、統合化整備

通 番	データ（又はDB）の名称	公 開 ／ 未 公開	概要（データの種類（生物種）・数量（kB 等）、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述）
1	ゲノムネット医薬品データベース http://www.genome.jp/kusuri/	公開	研究の最先端と医療の現場さらには一般社会をつなぐ日本語の医薬品統合データベース。JAPIC 医薬品添付文書情報（医療用医薬品 12,239 件、一般用医薬品 12,042 件、平成 22 年 11 月現在）を検索可能。KEGG DRUG の構造情報やターゲット情報と統合している。また、文献データベース PubMed、J-STAGE、J-GLOBAL へのリンクも付加している。医療用医薬品は平成 19 年 9 月、一般用医薬品は平成 20 年 1 月より公開している。平成 21 年度からはゲノムネット統合データベースからもキーワード検索できるようにした。JAPIC データの更新に伴い毎月の更新を行なっている。
2	ゲノムネット化合物データベース http://www.genome.jp/ja/gn_dbget_ja.html	公開	平成 19 年までに検索環境を整備した DBGET/LinkDB を化合物・化学反応データベースの統合化に応用したもの。平成 22 年 11 月までに 12 データベースをキーワード検索対象として、25 データベースを LinkDB 対象として組み込んだ。
3	化学反応パターンデータベース RPAIR/RCLASS http://www.genome.jp/dbget-bin/www_bfind?rpair http://www.genome.jp/dbget-bin/www_bfind?rclass	公開	反応の基質と生成物のペアを抽出したデータベース RPAIR の化合物アライメント情報を整備した。また、RPAIR データベースから抽出される反応パターンの情報を RCLASS データベースとして整備し、平成 22 年 7 月に公開した（2010 年 11 月 17 日版で RPAIR は 12,326 エントリ、RCLASS は 2,296 エントリ）。

(2) DB 基盤システム、ツール等開発成果物の整備

通 番	DB 基盤システム、ツール等の名称	公 開 ／ 未 公開	概要（主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述）
4	DBGET/LinkDB: ゲノムネット統合データベース検索システム http://www.genome.jp/ja/gn_dbget_ja.html	公開	平成 18 年度までに DBGET/LinkDB として開発してきたシステムを、日本語支援環境の整備、LinkDB の拡張、新たな検索システムの開発という観点から改良したもの。全データベース一括検索と外部データベースを含む LinkDB 検索を平成 19 年 7 月に、日本語支援環境を平成 19 年 10 月に公開した。
5	類似化合物検索システム SIMCOMP http://www.genome.jp/tools/simcomp/	公開	グラフ比較に基づいた精度の高い類似度計算を実現している。検索速度に問題があったため、平成 19 年度に高速化についての調査を行い、平成 20 年度に高速化と光学異性体の違いを認識するように改良した。平成 22 年度にはパッケージ化して FTP で公開した。
6	部分構造検索システム SUBCOMP	公開	ビットストリングを用いた高速な化合物部分構造検索システム。これまで研究レベルで開

	http://www.genome.jp/tools/subcomp/		発されてきたものを、バグの修正などをした上で、ウェブの検索システムとして平成 20 年度に公開した。平成 21 年度には光学異性体の違いを認識するための改良、および、クエリを含む構造の検索を実装した。平成 22 年度にはパッケージ化して FTP で公開した。
7	化合物間の酵素反応予測システム E-zyne http://www.genome.jp/tools/e-zyne/	公開	化学構造変化に基づく反応予測システム。基質と生成物を与えると、その間の反応パターンを予測、EC 番号との対応付けなどを行う。テンプレートとなる反応パターンの充実が課題であったため、平成 20 年度には反応パターンデータベースを整備するとともに、データベース中の EC 番号と反応パターンの分布を用いて精度の向上を実現したものを平成 21 年 1 月に公開した。
8	分解・合成反応経路予測システム PathPred http://www.genome.jp/tools/pathpred/	公開	化学構造変化に基づく反応経路予測システム。基質（と生成物）を与えると、その基質を分解する経路や、生成物を合成するための経路を予測する。バクテリアが環境物質を分解する経路に特徴的な反応パターン（化学構造変化）や植物の二次代謝産物整合性経路に特徴的な反応パターンを用いて経路探索の効率化を図っている。平成 22 年 1 月に公開した。
9	糖鎖類似構造検索システム KCaM http://www.genome.jp/tools/kcam/	公開	糖鎖に特徴的な木構造のための動的計画法を実装したシステム。平成 19 年度にユーザインタフェースと他のシステムと統一した。

外部発表実績一覧

(1) セミナー、研究会等イベント開催

通 番	タイトル	発表者 (代表者)	年月日	開催場所	イベント名 称	概要（対象者（層、参加人数）、出席者の主な 反応等）
10	ゲノムネットデータベース講習会	五斗進他	平成 20 年 1 月 30 日、31 日	東京大学	ゲノムネッ トデータベ ース講習会	PC を用いた実習形式での講習会。ホームページ上で一般から 20 名の参加者を募った。大学、公的機関の研究所、企業から幅広く集まった。
11	ゲノムネットデータベース講習会	五斗進他	平成 21 年 1 月 29 日、30 日	東京大学	ゲノムネッ トデータベ ース講習会	PC を用いた実習形式での講習会。ホームページ上で一般から 30 名の参加者を募った。大学、公的機関の研究所、企業から幅広く集まった。
12	ゲノムネットデータベース講習会	五斗進他	平成 22 年 2 月 4 日、5 日	東京大学	ゲノムネッ トデータベ ース講習会	PC を用いた実習形式での講習会。ホームページ上で一般から 30 名の参加者を募った。大学、公的機関の研究所、企業から幅広く集まった。

(2) プレス発表、取材対応

通 番	タイトル	発表媒体	年月日	特記事項
13	ゲノムネット医薬品・化合物統合 DB を公開	Medical Tribune	2009 年 3 月 26 日	

(3) 展示会等出展

通 番	タイトル	展示会等の名称	年月日	特記事項
14	医薬品情報統合データベースの開発	BMB2007	2007 年 12 月 11 日	
15	ゲノムネット化合物・医薬品データベース	統合データベースプロジェクト シンポジウム 2009	2009 年 6 月 12 日	
16	酵素反応式からの基質／生成物ペアの予測	BMB2009	2009 年 12 月 10 日	
17	ゲノムネット化合物・医薬品データベースと関連ツールの開発	BMB2009	2009 年 12 月 10 日	
18	Structural clustering of plant secondary metabolite to estimate compound classes reflecting their biosynthetic pathway	GIW2009	2009 年 12 月 14 日	
19	Reclassification of oxidoreductase subclasses based on the relationships of reductants and oxidants	GIW2009	2009 年 12 月 15 日	
20	ゲノムネット化合物データベースの最新状況	第 51 回日本植物生理学会年会	2010 年 3 月 21 日	
21	ゲノムネットにおける医薬品・化合物データベースの開発	統合データベースプロジェクト シンポジウム 2010	2010 年 10 月 5 日	

(4) 学会等への口頭発表

通 番	タイトル	発表者	学会等の名称	年月日	特記事項
22	ゲノムネットの化合物情報データベース	五斗進	BMB2008	2008 年 12 月 10 日	
23	E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs	Yamanishi, Y., Hattori, M., Kotera, M., Goto, S. and Kanehisa M.	ISMB2009	2009 年 6 月 30 日	
24	植物研究のための代謝化合物統合データベース	時松敏明、寺島悦子、木下広美、五斗進、金久實	BMB2009	2009 年 12 月 10 日	
25	Tools for investigating mechanisms of antigenic variation: new extensions to varDB	Hayes, C. N., Diez, D., Joannin, N., Kanehisa, M., Wahlgren, M., Wheelock, C. E. and Goto, S.	GIW2009	2009 年 12 月 16 日	

(5) 学術雑誌等への論文寄稿

通 番	タイトル	著者名	雑誌等の名称	掲載巻、号、ページ	特記事項
26	医薬品の統合データベース	金久 實、伊藤真純、奥田修 二郎、五斗進、太田福子	蛋白質 核酸 酵素	52(12):1486-1491 (2006)	
27	varDB: a pathogen-specific sequence database of protein families involved in antigenic variation	Hayes, C. N., Diez, D., Joannin, N., Honda, W., Kanehisa, M., Wahlgren, M., Wheelock, C. E. and Goto, S.	Bioinformatics	24:2564-2565 (2008)	
28	PathPred: an enzyme-catalyzed metabolic pathway prediction server	Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Goto, S. and Kanehisa, M.	Nucleic Acids Research	38:W138-W143 (2010)	
29	SIMCOMP/SUBCOMP: chemical structure search servers for network analyses	Hattori, M., Tanaka, N., Kanehisa, M. and Goto, S.	Nucleic Acids Research	38:W652-W656 (2010)	
30	varDB: a database of antigenic variant sequences – current status and future prospects	Diez, D., Hayes, C. N., Joannin, N., Normark, J., Kanehisa, M., Wahlgren, M., Wheelock, C. E. and Goto, S.	Acta Tropica	114:144-151 (2010)	

区 分	☐ 中核機関（ ☐ 代表機関／ ☐ 参画機関） ☒ 分担機関（ ☒ 代表機関／ ☐ 参画機関） ☐ 補完課題実施機関
課 題 名	統合医科学データベース構築方式の開発
実 施 機 関 名	国立大学法人 東京医科歯科大学
代 表 研 究 者 名	田中 博

1. 課題開始時における達成目標
①がん疾患、神経疾患を対象にした臨床・疾患データベースの統合化技術開発 疾患データベースの統合化に必要な技術要件をまとめる。その後疾患データベースを統合するための標準様式と標準API(Application Programming Interface)のモデルを開発・整備する。特に、分子から病理、臨床情報の多階層病態データを関連づけるオミックス縦断解析の標準手法を開発する。 ②疾患データベース（がん、神経難病を中心として）の高度化と実証的統合に関する研究 東京医科歯科大学の統合臨床オミックスデータベース300症例、大阪大学のパーキンソン病臨床情報データベース500症例の実証的統合に加え、Genome Medicine Database of Japan (GeMDBJ)などのがん、神経疾患に特化してデータベースの統合を進め、統合データベース検索システムをプロトタイプ的に構築し、公開を行う。特に疾患の区別を越えて横断的に検索する検索システムの要件を明らかにする。 ③オントロジー、ターミノロジー、シソーラスの整備 がん疾患、神経疾患を典型例として、オントロジー、ターミノロジー、シソーラスを整備する。 ④倫理面の検討 ・臨床疾患データベースを統合公開するにあたり必須となる倫理規定案を作成し、中核機関へ提案する。
2. 平成22年10月末時点における事業計画に対する成果
（1）成果概要 ①がん疾患、神経疾患を対象にした臨床・疾患データベースの統合化技術開発 ・統合技術開発における具体的要求と課題を検討し、要件定義を行った。 ・縦断的統合検索機能の開発・高度化を行った。(URL:http://omics.tmd.ac.jp/) ②疾患データベース（がん、神経難病を中心として）の高度化と実証的統合に関する研究 ・多階層統合医科学データベースを構築するための要件定義を行った。 ・横断的統合検索機能（IBMD: Integrated Biomedical Database）の開発・高度化、およびプロトタイプシステムを構築し運営を行った。(URL:http://ibmd.tmd.ac.jp/) ・実証的統合として統合臨床オミックスデータベース(iCOD)447症例、Genome Medicine Database of Japan (GeMDBJ)440症例を対象データベースとして、がん疾患およそ700症例の統合を完了した。 ③オントロジー、ターミノロジー、シソーラスの整備

オントロジー情報の高度化として、大腸癌、口腔癌、肝臓癌、パーキンソン病（パーキンソン症候群）、頭頸部癌において、語数 247 語、関連数 190 のオントロジー情報を整備した。これをベースに横断的検索における検索キーワードが直接ヒットしないとき、オントロジーを参照して、意味的変換を行い同意語による検索を行えることを確認した。

④倫理面の検討

- ・倫理規定ガイドラインの整備に関連して、各データベースに対するアクセスレベルを設定し、一般公開版、登録版として運営を行った。
- ・データベース統合の是非・程度、オミックス情報公開の是非・程度、データベース使用の体制、責任分界などについて明記した倫理規定ガイドラインをまとめた。

（２）進捗及び成果

がん疾患、神経疾患を対象にした臨床・疾患データベースの統合化技術開発として、まず初めに我が国のライフサイエンス研究を支える社会情報基盤としての多階層の統合医科学データベースを統合構築するために必要な統合技術（疾患データベース情報モデル、疾患データベースオントロジー、セマンティクスなど）の要件分析、標準化、公開倫理などの検討に着手し、要件定義を行なった。要件定義を元に、ユーザインタフェースの高度化、オントロジー情報の高度化を行い、臨床医科学情報を疾患横断的に統合検索が可能となる横断的統合検索機能（IBMD: Integrated Biomedical Database）を開発し、まずはプロトタイプとして運営を開始した。現在、横断的統合検索用 API をウェブ上に公開する準備を進めている。

疾患データベース（がん、神経難病を中心として）の高度化と実証的統合に関する研究としては、各疾患データベース（がん疾患、神経疾患）の高度化によって得られた統合技術開発における具体的要求と課題の検討を進め、要件定義を行なった。また、分子情報レベルから臨床情報レベルの分子階層ネットワーク縦断的な統合検索機能として、統合臨床オミックスデータベース（iCOD: Integrated Clinical Omics Database）をベースに、二次元三層マップ、正則化正準相関分析を開発した。今年度中にモジュール化した上で成果物としてウェブ上でダウンロードできる形で公開の予定である。

倫理規定ガイドラインの整備については、(1)一般論としてデータベース統合についてその影響や特質について検討 (2)具体的な医科学データベースの統合を進める場合に想定される問題を、オミックスデータベースと臨床データベースについて検討 (3)法的側面を整備するため、罰則規定や訴訟の影響について、事例等を踏まえつつ検討 (4)社会的な適合性も損なわないよう、成立／改定や施行のプロセスについても策定 という４段階を経た。上記の検討の結果より、DB 統合の是非・程度をはじめとして、正確性、安全性、透明性（３要素）の確保、また、匿名性などを最重要項目としてそれぞれの項目について検討を行った。これに加え、中核機関より、より実際のデータベース利用を考慮したガイドラインを作成するよう指示を受け、データベース提供者、データベース管理者、データベース利用者毎のデータ管理（使用）体制、責任分界などを追記した。現在、データ種類別アクセスレベルの設定などについて大阪大学と積極的な意見交換を行いながら最終案の作成を進めており、その後中核機関を通じて本ガイドラインを社会全体へ提案の予定である。

また、本プロジェクトの認知・広報活動として、関連情報との連携を推進した。まず、国内の主要な疾患データベースの精査を行い、次に各データベースの構造などを把握した上で本プロジェクトとの関連性を検討し、連携が妥当だと判断した機関について、まずはホームページ上での掲載を行った。平成

2 2 年 1 0 月 末 時 点 で は 、 日 本 肝 癌 研 究 会 、 大 腸 癌 研 究 会 と の 連 携 を 行 っ て お り 、 今 後 も 国 内 外 を 問 わ ず 連 携 を 進 め る 予 定 で あ る。
3. 当初目標に対する達成度 提案時の目標のとおり、分子階層ネットワーク縦断的・疾患横断的な臨床医科学情報を相互的有機的に関連させる統合化方式を開発・実装した。 横断的統合検索機能については、開発当初の提案では方法論の開発と2疾患への適応にとどまっていたが、計画開始とともに中核機関から範囲を広げるような提案がありその方向で開発を進めていた。しかし中間評価において、範囲を過度に拡張するのではなくロールモデルに徹するよう指示があり、最終到達目標としてプロトタイプ確立と必要要件の列挙にとどめた。その意味ではほぼ目標に到達したといえる。また、公開症例数についてもがん疾患、神経疾患のそれぞれで当初目標を超える症例数を公開済みである。 縦断的統合検索機能については、分子レベルから臨床レベルの情報を検索し、症例データ及び分子・臨床レベルの情報を統合的に表示する機能の高度化として、統合臨床オミックスデータベース（iCOD）を開発ベースとして、分子・病理・臨床の多階層の患者情報を関連付ける二次元三層マップ方式や正規化正準相関分析の開発を完了しており、提案時の目標を十分に達成したといえる。 倫理規定ガイドラインについては、現在も東京医科歯科大学と大阪大学の間で各項目に関する意見交換を行っており、年度末までには作業を完了させ、中核機関へ提案の予定である。
4. 中間評価に対する対応 横断的統合検索機能では中核機関の指示もあって疾病領域における代表的疾患をほぼ網羅するように対象疾患を一時拡大したが、中間評価結果を受けて再びがんおよび神経疾患（パーキンソン病）に限定し、ロールモデルの開発に重点を置いた。ロールモデルの開発では、当初計画に先行しプロトタイプの開発・運営を行った。 また、本来提案時で目指した分子から臨床に互る縦断的統合検索機能では、統合臨床オミックスデータベース（iCOD）のモデル的確立を目指した開発に重点を置いた。さらに患者情報を可及的に広い範囲で一般公開し、またデータベース利用法の簡便な習得を可能にするインターフェースを充実させた。 疾病情報の標準化に関しては国際的な医療情報標準化会議に出席し、標準的なオントロジーに基づいた概念を可及的に使用するよう変更した。また、標準的な医療オントロジーを試作し、国際的にも提案した。
5. 他機関との連携 本プロジェクトは中核機関の協力の下、東京医科歯科大学、大阪大学が連携し開発を進めた。また、中核機関の全体戦略に従い、対象疾患の範囲や重点課題について臨機応変に対応し、開発・運営を行った。その他、東京大学グループとのパーキンソン病データベースの連携、倫理規定ガイドラインについても中核機関の指示を仰いだ上で検討を行った。 本プロジェクトにおいて、東京医科歯科大学と大阪大学で連携を行った。統合化要素技術の開発、標準化において、開発担当者を含め定期的な打ち合わせを行った。倫理規定ガイドラインの作成においては倫理規定検討会を開催し、アンケートや調査結果をもとに論点の整理、項目の追記・修正など、意見交換を積極的に行った。

プロジェクト外部機関との連携では、国立がん研究センター研究所が挙げられる。がんデータベースの統合化モデルを効果的に提示するため、国立がん研究センター研究所の GeMDBJ を要素データベースとして連携し、データの連携方法、インターフェースの実装方法といった要件定義を行った。

学会との連携においては、医療情報学会と連携し、医療情報学連合大会で統合データベースのセッションを設けて認知・広報活動を行ったほか、オミックス医療研究会主催のオミックスシンポジウム、その他多くの国内・国際学会で認知活動を推進した。

6. 今後の見通し、計画、展望

横断的統合検索機能ではロールモデルの提供を最重要目標として設定していたため、要素データベースを多数統合した上での実運用、およびその運用でのロールモデルの再検証という面においては、今後推進していく必要のある課題である。また、新たに要素データベースを追加する過程の中で、現状のオントロジー、ターミノロジー、シソーラスで十分な対応が可能かどうかの見直しを行う必要がある。それに加え、アカウント数やアクセス数、ダウンロード数などを参考にさらなる認知・広報活動を行っていく、アンケート結果などを参考にインターフェース等の改良を検討するなどの継続的課題がある。

縦断的統合検索機能としては、統合臨床オミックスデータベース (iCOD) をベースとして開発した二次元三層マップや正則化正準相関分析などを対象疾患の制約を離れて一般化し、全疾患に拡張できるよう標準化した医療データモデルとして完成させ、他のデータベースでも利用可能なようにする。本学の難治疾患研究所が全国共同研究拠点になるに従い、情報センター（現情報基盤部門）だけではなく、研究所の共同利用資源として iCOD を公開し、さらなる全国利用を進めていく。また、臨床学会との共同関係を強めて、分野横断的な臨床統合データベースとしての普及を図る。

7. 全体総括

現状の医科学データベースは、疾患についての原因遺伝子や多型情報などの分子情報のみであったり、また臨床症例情報だけで関連する分子情報を欠いていたりするものが多く、分子情報から病理組織情報さらには臨床情報など、すべての階層に互った「多階層統合的な」疾患情報を提供している疾患データベースは極めて少数である。また、そのデータベースはそれぞれの研究者がその領域での目的達成のためのみに保持している場合が多く、領域や階層が異なるデータベースやその利用者については考慮されていない。

このように、従来散在していた個々の医科学データベースの可用性の向上・高度利用・戦略的統合を可能とし、我が国におけるライフサイエンスの医科学研究の情報基盤を形成すべく、分子階層ネットワーク縦断的・疾患横断的な臨床医科学情報を相互的有機的に関連させる統合化方式を開発・実装した。ロールモデルの提供という最重点目標の達成は、まだまだ統合された要素データベースが少なく、実運用という意味では不十分である面を差し引いても、一定の評価が得られると考える。

特に統合臨床オミックスデータベース (iCOD) をベースとして開発した二次元三層マップ、正則化正準相関分析は、分子から臨床の多階層にわたる医科学研究の支援に用いることのでき、医薬産業の活性化と国際競争力の強化、将来医療実現の礎となり得ることから考えても、高い価値があるだろうと考えられる。

8. 特記事項

医生物学分野におけるデータベース統合は国際的な流れであり、網羅的分子情報、臨床情報、環境情報などについて統合がすすめられてきているが、多階層オミックス情報を軸として、俯瞰的に網羅的分子情報を疾患情報、臨床情報、環境情報と階層的かつ疾患横断的に統合したデータベースは存在せず、我々が開発した分子階層ネットワーク縦断的・疾患横断的な臨床医科学情報を相互的有機的に関連させたロールモデルの提供は世界的に類を見ない。高度データベース活用の取り組みとして、BIRN (Biomedical Information Research Network) の例があるが、全体を俯瞰した統合データベースとはいえない。がんではPDQ (米国NCI) があるが、知識ベースである。

9. 委託研究費一覧

	18年度	19年度	20年度	21年度	22年度	計
設備備品費 (千円)		2,126	0	0	0	2,126
人件費 (千円)		11,493	16,820	12,140	11,713	52,166
業務実施費 (千円)		31,836	14,998	10,587	4,923	62,344
一般管理費 (千円)		4,545	3,182	2,273	1,664	11,664
合 計 (千円)		50,000	35,000	25,000	18,300	128,300

整備実績一覧

(1) データ（又はDB）の連結、統合化整備

通番	データ（又はDB）の名称	公開／未公開	概要（データの種別（生物種）・数量（kB 等）、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述）
1	Integrated Biomedical DataBase (IBMD) http://ibmd.tmd.ac.jp/	公開	要素データベース、データ公開数： ・がん 約700例 (iCOD、GeMDBJ) ・神経疾患 約700例 (パーキンソン病臨床情報データベース) 公開開始日：2007年（プロトタイプとして） アクセス数：月500アクセス前後 進捗状況、今後の計画： 上記を要素データベースとしてロールモデルを開発。各要素データベースに対する横断検索、検索結果について詳細情報の確認（外部リンクを含む）が可能。今年度末までに横断検索用のAPIを公開の予定。
2	Integrated Clinical Omics Database (iCOD) http://omics.tmd.ac.jp/	公開	生物種：ヒト データ数：肝臓癌、大腸癌、口腔癌の症例情報（477例） 肝臓癌、大腸癌、口腔癌検体の遺伝子発現解析結果（129例） データ内容：臨床情報（基本情報、病歴・生活歴、臨床検査、画像診断、治療、予後等） 分子情報（マイクロアレイ遺伝子発現情報） アクセス数：月20000アクセス前後 進捗状況、今後の計画：分子レベルから臨床情報レベルまでの縦断的統合検索機能の開発モデルとして整備。二次元三層マップ、正規化正準相関分析等を開発。今年度末までにモジュールを公開の予定。

(2) DB基盤システム、ツール等開発成果物の整備

通番	DB基盤システム、ツール等の名称	公開／未公開	概要（主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述）
1	統合医科学データベース プロトタイプシステム	公開	各疾患データベースの構造の相違を吸収し、要素データベースとして疾患横断的に統合検索が可能となるシステム。また、疾患オントロジー情報を基にしたセマンティック検索が可能である。
2	横断的検索機能 API	未公開	疾患データベースを要素データベースとして IBMD に統合するためのAPI。プロトタイプシステムの構築に利用。

3	二次元三層マップ モジュール	未公開	二次元三層マップは、各層（臨床層、病理層、分子層）で使用するデータの種類と計算方法を指定し、二次元マップ上に配置する。ある層のある点を選択すると、それに該当する別の層での位置と関係性が見えるため、それぞれの相関や関係性を視覚的に認識することができる。
4	正則化正準相関分析 モジュール	未公開	臨床・病理情報を遺伝子発現データと同一マップ上に分布させ、それらの相関を調べることができる。

（３）その他の成果物（（１）、（２）に該当しないもの）

通番	名称	公開／未公開	概要
1	倫理規定ガイドライン	未公開	データベース統合の是非・程度、オミックス情報公開の是非・程度、データベース使用の体制、責任分界などについて明記。

外部発表実績一覧

（１）セミナー、研究会等イベント開催

通番	タイトル	発表者 (代表者)	年月日	開催場所	イベント名称	概要（対象者（層、参加人数）、出席者の主な反応等）
1	東京医科歯科大学バイオ医療オミックス情報学人材養成プログラム成果発表会－国際シンポジウム「次世代疾患オミックスの新展開」	田中 博	2010年2月2日	みらいホール (桜木町)	シンポジウム	参加者約200人
2	オミックス医療情報学寄附講座開設記念講演会	田中 博	2009年11月19日	歯学部附属病院特別講堂	講演会	参加者約120人
3	東京医科歯科大学・中国医科大学 シンポジウム	田中 博	2009年10月14日	歯学部附属病院特別講堂	シンポジウム	参加者約100人
4	「病気に挑む生命科学」学際生命科学が先端医学の未来を切り拓く！	田中 博	2009年6月2日	歯学部特別講堂	シンポジウム	参加者約120人
5	オミックス医療研究会「International Symposium of Translational Informatics」	田中 博	2009年3月24-25日	品川プリンス	シンポジウム	参加者約200人
6	オミックス医療研究会・創薬 PGx 分科会 “Pharmacogenomics Symposium”	田中 博	2009年3月17-18日	東京工業大学	シンポジウム	参加者約300人
7	オミックス医療研究会、バイオマーカー分科	田中 博	2009年1月24日	パンパシフィ	シンポジウム	参加者約100人

	会「バイオマーカーの最前線 ～次世代診断と創薬にむけて～」			ック横浜ベイホテル東急		
8	第28回医療情報学連合大会「ゲノムインフォマティクス」セッション 医療情報におけるオミックス情報のインパクト ― 統合データベースを踏まえて ―	田中 博	2008年11月25日	パシフィコ横浜	シンポジウム	参加者約100人
9	第28回医療情報学連合大会「オントロジーと疾患統合データベース」セッション 疾患知識表現としてのオントロジーと疾患統合データベース	田中 博	2008年11月24日	パシフィコ横浜	シンポジウム	参加者約100人
10	オミックス医療研究会オミックスシンポジウム「オミックス医療の実現に向けて―統合的医療データベース iCOD―」	田中 博	2008年10月20日	一橋記念講堂	シンポジウム	参加者約250人
11	20年度医療IT推進協議会シンポジウム「日本版EHRと医療の質の評価」	田中 博	2008年9月27日	阪急ホテル	シンポジウム	参加者約250人
12	オミックス医療研究会「オミックス医療が拓く未来2008」～個別化医療実現に向けた橋渡し研究戦略を考える～	田中 博	2008年7月3-4日	学術総合センター	シンポジウム	参加者約300人
13	オミックス医療研究会「オミックス医療が拓く未来2007」	田中 博	2007年10月22-23日	一橋記念講堂	シンポジウム	参加者約250人
14	第4回ゲノム・オミックス医療情報シンポジウム	田中 博	2007年6月2日	東京医科歯科大学	シンポジウム	参加者約250人

(2) プレス発表、取材対応

通番	タイトル	発表媒体	年月日	特記事項
1	「オミックス情報が医療を変える」	NHK放送 「私も一言！夕方ニュース」	2008年7月29日	
2	「病態と分子情報同時登録のDB」	朝日新聞掲載	2008年8月1日	

(3) 学会等への口頭発表

通番	タイトル	発表者	学会等の名称	年月日	特記事項
1	Current State and Futures of Systems Pathology, Translational Research Excellence Conference (TRX10)	Tanaka H.	Brisbane Convention & Exhibition Centre, Brisbane, Australia	Oct11-13, 2010	

2	Systems Pathology - A new concept for understanding disease -	Tanaka H .	CBI-KSBSB Joint Conference, Haeundae Grand Hotel, Busan, Korea	Nov 4-6, 2009	
3	Systems Pathology	Tanaka H .	Personalized medicine 2009, Bioinformatics Center, Hangzhou, China	May 29-31, 2009	
4	Present Situation and Future Trends of Omics-based medicine and systems pathology	Tanaka H .	Conf. on Biomedical Electronics and Biomedical Informatics, Moscow, Russia	Aug, 2009	
5	Systems pathology- A new concept for understanding disease -	Tanaka H .	Chinese Association of Biophysics, China	Jul, 2009	

(4) 学術雑誌等への論文寄稿

通 番	タイトル	著者名	雑誌等の名称	掲載巻、号、ページ	特記事項
1	iCOD: an integrated clinical omics database based on the systems-pathology view of disease.	Shimokawa K., Mogushi K., Shoji S., Hiraishi A., Mizushima H., Tanaka H.	BMC Genomics, 2010	in press	
2	Omics-based medicine and systems pathology.	Tanaka H.	Methods Inf Med	49:173-185.	
3	Structures of protein interaction network and their implications on drug design.	Hase T., Tanaka H., Suzuki Y., Nakagawa S., Kitano H.	PLoS Computational Biology	5(10) e1000550, 2009	
4	Genomic sequence variation markup language (GSVML).	Nakaya, J.†, Kimura, M., Hiroi, K., Ido, K., Yang, W., Tanaka, H.	Int J Med Inform	79:130-142.	
5	Aurora kinase B is a predictive factor for the aggressive recurrence of hepatocellular carcinoma after curative hepatectomy.	Tanaka S., Arii S., Yasen M., Mogushi K., Nguyen T., Zhao C., Imoto I., Eishi Y., Inazawa J., Miki Y., Tanaka H.	British Journal of Surgery	95:611-619, 2008	

成果報告票 2.(2) 参考資料

成果物表示画面

1

IBMD表示画面

IBMD (http://ibmd.tmd.ac.jp/)

横断検索

3

詳細情報 (外部リンク)

詳細情報

4

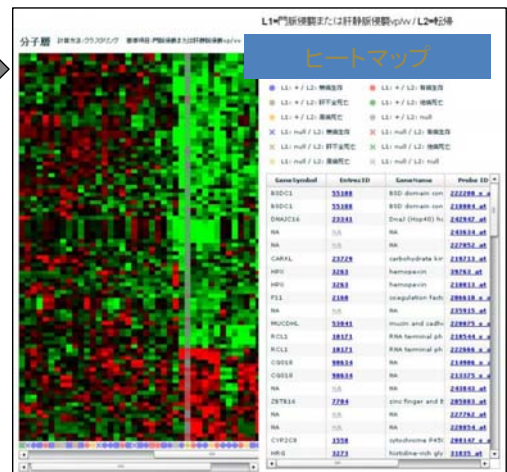
iCOD表示画面

5

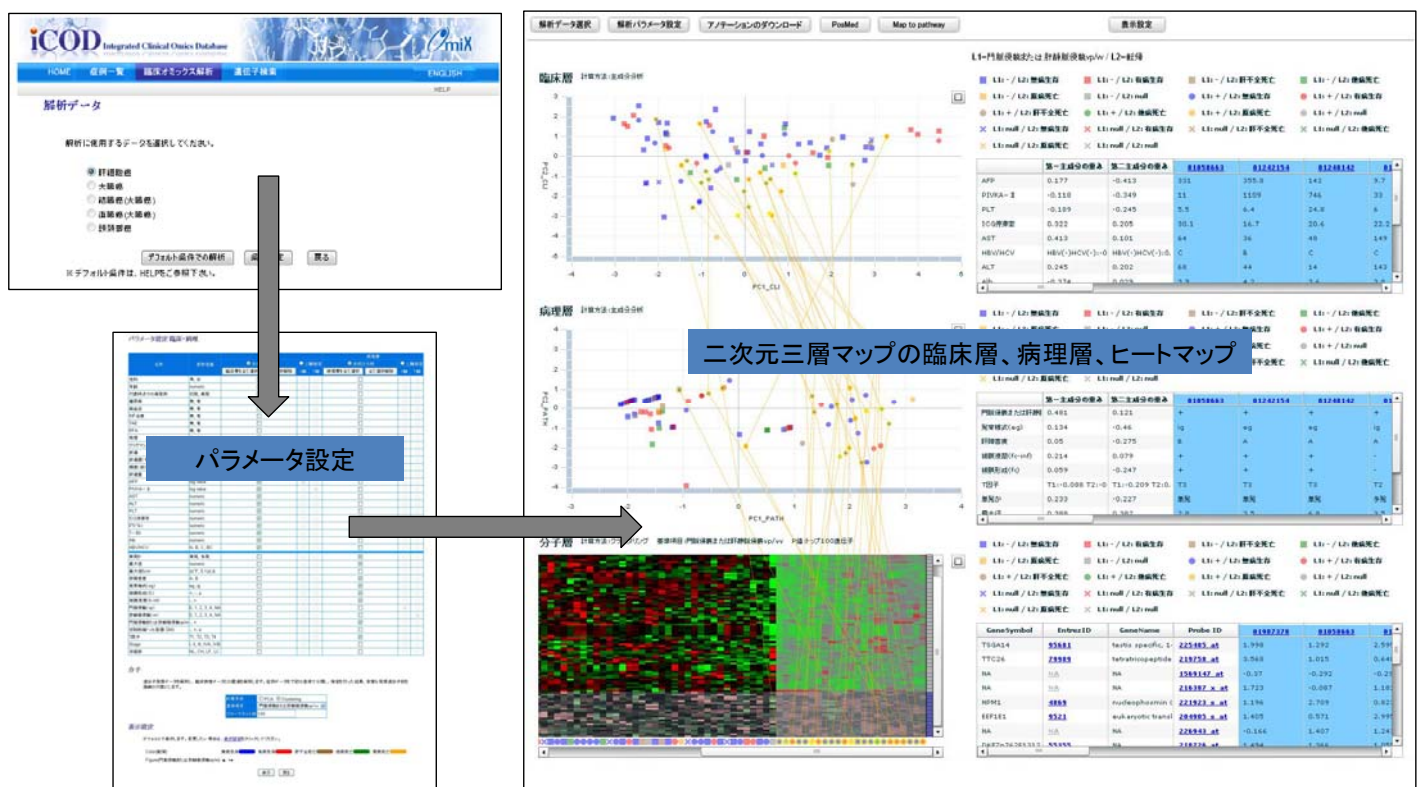
iCOD(<http://omics.tmd.ac.jp/>)

The image displays the iCOD (Integrated Clinical Omics Database) website interface. The main navigation bar includes links for HOME, 症例一覧 (Case List), 臨床オミックス解析 (Clinical Omics Analysis), and 遺伝子検索 (Gene Search). The 症例一覧 section shows a list of cases with filters for disease (肝臓癌, 大腸癌, 膵臓癌, 食道癌) and time points (2010年, 2011年, 2012年). The 臨床オミックス解析 section displays a 3D visualization of omics data. The 遺伝子検索 section includes a search bar and a list of search results. The bottom of the page features a footer with the TMD University logo and contact information.

症例一覽



臨床オミックス解析



遺伝子検索

iCOD Integrated Clinical Omics Database

HOME Case Archive Clinical Omics Data Analysis Gene Search

Results 1 - 25 of 68 next>> 25

Gene search

Diagnosis: Hepatocellular carcinoma

Gene or ProbeSet id: RAS

Search

1: **LOC100133308**
Official Symbol: LOC100133308, Name: Ras suppressor protein 1 pseudogene
ProbeSet id: 1567672_s_at, 1561482_at

2: **RAC1**
Official Symbol: RAC1, Name: Ras-related GTP binding protein Rac1
ProbeSet id: 1567457_at, 1567458_s_at, 200640_at, 200641_s_at

3: **RAC2**
Official Symbol: RAC2, Name: ras-related GTP binding protein Rac2
ProbeSet id: 207419_s_at, 213603_s_at

4: **RAC3**
Official Symbol: RAC3, Name: ras-related GTP binding protein Rac3
ProbeSet id: 206103_at

5: **RADL**
Official Symbol: RADL, Name: Ras association and DIL domains
ProbeSet id: 226693_s_at

6: **RAPH1**
Official Symbol: RAPH1, Name: Ras association (RalGDS/AF-6) and pleckstrin homology domains 1
ProbeSet id: 1562482_at, 225186_at, 225188_at, 225189_s_at, 231075_x_at

7: **RASA1**
Official Symbol: RASA1, Name: RAS p21 protein activator (GTPase activating protein) 1
ProbeSet id: 202677_at, 210621_s_at

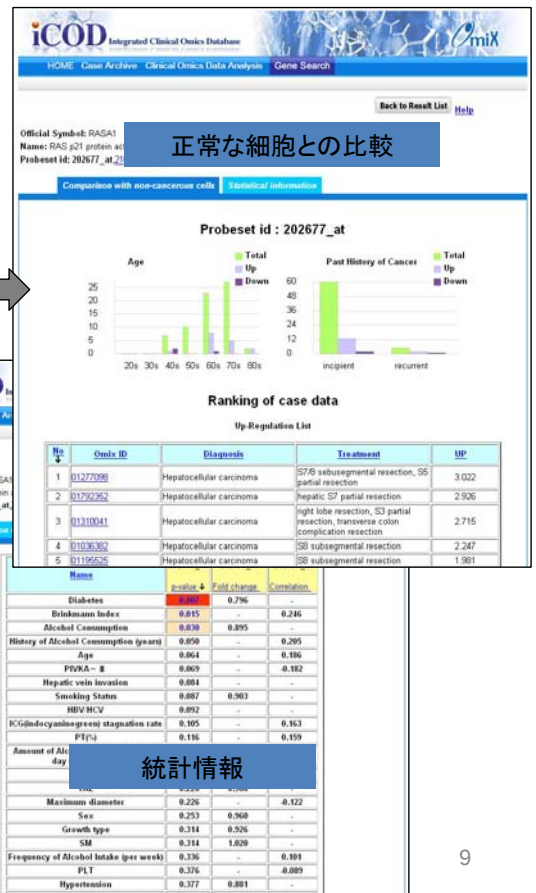
8: **RASA2**
Official Symbol: RASA2, Name: RAS p21 protein activator 2

iCOD Integrated Clinical Omics Database

HOME Case Archive Clinical Omics Data Analysis Gene Search

Official Symbol: RASA1
Name: RAS p21 protein activator (GTPase activating protein) 1
ProbeSet id: 202677_at

Compare



2010年11月4日

統合医科学データベース 倫理規定ガイドライン

東京医科歯科大学、大阪大学

目次

1. はじめに
 - 1.1. [統合医科学データベースの目的]
 - 1.2. [遵守事項]
 - 1.3. [定義]
2. 第一部 データベースの統合について
 - 1.4. 第一条 [データベース統合の是非]
 - 1.5. 第二条 [データベース統合の程度]
3. 第二部 医科学データベースについて
 - 1.6. 第三条 [オミックス情報公開の是非・程度]
 - 1.7. 第四条 [臨床情報の情報収集について]
 - 1.8. 第五条 [臨床データベースの公開について]
 - 1.9. 第六条 [臨床データベースの二次利用について]
 - 1.10. 第七条 [臨床データベースの多次利用について]
4. 第三部 データベース使用の体制
 - 1.11. 第八条 [データベース管理機関の基準等]
 - 1.12. 第九条 [データベース提供機関の基準等]
 - 1.13. 第十条 [データ使用機関(データ使用者)の基準等]
 - 1.14. 第十一条 [倫理審査委員会]
5. 第四部 責任分界
 - 1.15. 第十二条 [統合医科学データベース関係者の責任分界]
6. 第五部 データベースの法的背景について
 - 1.16. 第十三条 [知的所有権について]
 - 1.17. 第十四条 [罰則規定について]
 - 1.18. 第十五条 [訴訟について]
7. 第四部 データベースの社会的背景について
 - 1.19. 第十六条 [データベース運用上の倫理規定の国内外でのコンセンサスのプロセス]
 - 1.20. 第十七条 [倫理規定ガイドラインの見直し]

1. はじめに

1.1. [統合医科学データベースの目的]

よりよい医療の礎としての統合医科学データベースの形成を目指すにあたり、その情報の取扱いについて、法的あるいは社会的通念、医療の良識からみて倫理上問題となる行為および事態を信義誠実の原則に基づいて防止することで、情報の適正かつ円滑な利用、情報統合を実現し、良識ある医科学研究の支援に用いることのできる統合医科学データベースを構築運用することを目的とする。

1.2.

[遵守事項]

統合医科学データベースの管理者、利用者、および統合対象であるデータベースの提供者は、収録されている情報とそれにかかわる行為に広範な影響と責任があることを認識した上で、次に掲げる事項を厳守する。

1.3. [定義]

本ガイドラインにおいて、次の各号に掲げる用語の意義は、それぞれ当該各号に定めるところによる。

1、統合医科学データベース

疾患の分子情報から臨床病情報に至る各階層の病態情報を縦断的に連携する疾患病態の統合的情報提示や、様々な疾患の病態情報を横断的に検索することが可能な臨床医学的データベース

2、要素データベース

統合医科学データベースに収集され、統合された臨床医科学データベース

3、オミックス情報

ゲノムやプロテオーム情報などの生命分子の網羅的な情報

4、HIPPA (Health Insurance Portability and Accountability Act)

米国における医療保険の相互運用性と説明責任に関する法令

5、二次利用

統合医科学データベースで提供されているデータの組織的、大量的利用、および発展した目的のためのデータ再編を必要とする利用

6、多次利用

二次利用として利用されたデータがさらに第三者に提供、利用されること

2. 第一部 データベースの統合について

1.4. 第一条 [データベース統合の是非]

『データベース統合によって要素データベース作成者（データベース提供者）、収録データ元としての個人（データ提供者）のいずれもが、統合前よりも不利益を蒙る事があるとはならない。統合医科学データベース作成者は、上記不利益の無いように努めなければならない。』

1.5. 第二条 [データベース統合の程度]

『データベース統合にあたっては、情報連結によって生じる影響が、法的あるいは社会的通念から倫理上問題となる範囲におよぶ事があるとはならない。また、元情報の内容に変更がある情報統合を行ってはならない。データベースの統合は、統合医科学データベースの目的の範囲内で信義誠実の原則に基づいて行わなければならない。』

3. 第二部 医科学データベースについて

1.6. 第三条 [オミックス情報公開の是非・程度]

『オミックス情報の共有によって、要素データベース作成者（要素データベース提供者）、収録データ元としての個人（データ提供者）、また該当情報に何らかの関係をもつ一般人のいずれもが、新たに不利益を蒙る事があるとはならない。公開できるオミックス情報の程度は、将来にわたって個人を特定されることがない程度、要素データベース作成者（要素データベース提供者）、収録データ元としての個人に不利益を与えることがない程度としなければならない。』

オミックス情報および臨床情報については、ヒトゲノムバリエーションデータベース共有指針（http://gwas.lifesciencedb.jp/gwasdb/db_policy.html）を参考に、以下のデータカテゴリーとレベルによりアクセス制限を行う。

データ種類別アクセスレベル

カテゴリー			レベル	1：一般公開	2：アカウント登録	3：問い合わせ
データの 種類	オミックス情報	先天的分子情報	CNV	SNP個別データ（患者の同意書でその測定が特定されているSNPに関して）	SNP個別データ（raw data）	
		疾患オミックス情報	発現			
	臨床情報	臨床・病情報	臨床検査結果、処方歴、簡易病歴、画像（解像度低）、画像読影所見、肉眼所見、組織所見	画像（解像度高、DICOM形式）		
		臨床個別情報	統計情報のみ	症例原データ、病理詳細レポート、生活歴、手術報告、退院時サマリ	詳細生活歴、詳細病歴	
データ使用について			ウェブサイトにおいて、閲覧可能。データを大量取得する場合はアカウント登録と同様の申請を要する。	氏名、職名、連絡先、使用目的、e-mailアドレス（原則、所属機関から発行されたアドレス）を申請する。	共同研究の申し入れを行う。	

1. 7. 第四条【臨床情報の情報収集について】

内容

『臨床情報収集は、統合医科学データベースの目的の範囲内に限る。』

是非

『臨床情報の情報収集は、統合医科学データベースの目的に限って認められる。いかなる理由があろうと、営利目的については認められない。』

程度

『統合医科学データベースの目的の範囲内で必要最小限の情報に限る。』

ポリシー

『臨床データベースの情報収集に際してはデータの提供元にも統合先にも責任ある情報管理の執行能力が求められる。また、統合に先立って両者間に十分な認識共有と準備段階を持たなければならない。』

1.8. 第五条【臨床データベースの公開について】

内容

『臨床データベースの公開に当たっては、個人を特定できない内容としなければならない。具体的には、HIPPA（Health Insurance Portability and Accountability Act）に倣い、下記の項目については公開してはいけない。統合医科学データベース管理者には、個人を特定できないようにする義務がある。』

程度

『公開する情報は、個人を特定できない程度の情報粒度でなければならない。この情報粒度については、一定の監査期間を設けて継続的にチェックしなければならない。』

匿名化

『匿名化については、ある一定の基準に適合するように匿名化するように努めなければならない。一定の基準については、当該機関の倫理審査委員会などの第三者機関において検討された社会的コンセンサスのある内容としなければならない。第三者機関は、統合医科学データベースと一切の利害関係があってはならない。

匿名化が不可能となった場合、個人が特定できる状態となった場合には、統合医科学データベース管理者は、すぐにデータの公開をやめなければいけない。』

- 名前
- 居住地：郵便番号、都道府県、市区町村、番地、およびそれらと同等の居住者地域別分類
- 日付（情報の対象に関連する日付。例えば生年月日、入院日、退院日、遭遇日、手術日。）
- 電話番号
- ファックス番号
- Eメールアドレス
- 年金番号（住民票コード）
- カルテ番号
- 健康保険被保険者番号
- 口座番号
- 証明書番号、免許証番号
- 車両登録番号、シリアル番号（ナンバープレートを含む）
- 装置識別子番号、シリアル番号
- ウェブURL
- IPアドレス番号

- 生体認証（指紋、声紋）
- 顔全体の画像、およびそれに類似した画像
- その他ユニークな識別番号、特徴、コード

1. 9. 第六条【臨床データベースの二次利用について】

『統合データベースの目的の範囲内で、責任を負うことのできる利用者に限って、二次利用を行うことができる。また、医薬品の開発等、人類の福祉等に寄与する利用については統合医科学データベース管理者との協議、承認の下に利用が可能である。二次利用者は、データが多次利用される事のないよう、データの管理義務を負う。』

1. 10. 第七条【臨床データベースの多次利用について】

『臨床データベースの多次利用については、これを認めない。』

4. 第三部 データベース使用の体制

1. 11. 第八条【データベース管理機関の基準等】

『データベース管理機関は次に掲げる要件を満たすものとする。』

- ・情報セキュリティについての十分な知識を有し、要素データベースについて十分な管理が可能な施設、組織体制であること。
- ・データベース提供機関およびデータ使用機関（データ使用者）に対し、本ガイドラインについての十分な説明が可能であること。
- ・データ提供機関およびデータ使用機関（データ使用者）からの各種相談に対する窓口を設けること。

1. 12. 第九条【データベース提供機関の基準等】

『データベース提供機関は次に掲げる要件を満たすものとする。』

- ・アクセスレベル2以上のデータを含む臨床データベースの取り扱いにおいては、提供機関として臨床情報の取り扱いに関する技術的及び倫理的な事項に関する規則が定められていること。
- ・情報セキュリティについての十分な知識を有すること。
- ・本ガイドラインを十分に理解し、各条項を遵守できること。

1. 13. 第十条 [データ使用機関(データ使用者)の基準等]

『データ使用機関（データ使用者）は次に掲げる要件を満たすものとする。』

・アクセスレベル2以上のデータを含む臨床データベースの取り扱いにおいては、使用機関（使用者）として臨床情報の取り扱いに関する技術的及び倫理的な事項に関する規則が定められていること。

- ・臨床情報を使用するに足りる十分な技術的能力を有すること。
- ・本ガイドラインを十分に理解し、各条項を遵守できること。

1. 14. 第十一条 [倫理審査委員会]

『共同研究の申し入れ（アクセスレベル3のデータ利用）を行う使用機関においては、次に掲げる業務を行うため、倫理審査委員会を設置するものとする。』

・使用計画の科学的妥当性及び倫理的妥当性を総合的に審査できるよう、生物学、医学及び法律に関する専門家、生命倫理に関する意見を述べるにふさわしい識見を有する者並びに一般の立場に立って意見を述べられる者から構成されていること。

- ・当該使用機関が属する法人に所属する者以外の者が二名以上含まれていること。
- ・当該使用計画を実施する研究者、使用責任者との間に利害関係を有する者及び使用責任者の三親等以内の親族が審査に参画しないこと。
- ・倫理審査委員会の活動の自由及び独立が保障されるよう適切な運営手続が定められていること。
- ・倫理審査委員会の構成、組織及び運営並びにその議事の内容の公開その他使用計画の審査に必要な手続に関する規則が定められ、かつ、当該規則が公開されていること。

5. 第四部 責任分界

1. 15. 第十二条 [統合医科学データベース関係者の責任分界]

『統合医科学データベースへの要素データベースとしてのデータ提供は、統合医科学データベース管理者が研究目的のみで患者症例情報を利用するために行われるものであり、データベース提供者から適切な提供がなされる限り、その後の情報保護に関する責任はデータベース提供者から離れることになり、提供を受けた統合医科学データベース管理者に生ずる。ただし、データ提供を行う以前にデータベース提供者によって個人の同定につながる疾患特異的データの削除を行う必要があり、その作業を怠った場合についてはデータ提供者が責任を負う。

また、特に明記されていない場合、統合医科学データベース利用者の過失によるものであ

る場合は、統合医科学データベース使用者が全ての責任を負う。』

6. 第五部 データベースの法的背景について

1. 16. 第十三条【知的所有権について】

『知的所有権を取得する場合については、当事者間で協議を行い平和な解決を目指す。』

1. 17. 第十四条【罰則規定について】

『本規定の違反行為を行った者に対し、国内国外を問わず、統合医科学データベース管理者は利用資格の停止、取消等の措置をとることができる。また、本規定の違反行為を行った者は、法的措置が講じられる。』

1. 18. 第十五条【訴訟について】

『本規定の違反行為を行った者に対しては、民法あるいは刑法上の訴追がある。
日本国外に在住する者が本規定の違反行為を行った場合については、日本政府を経由して当該国に交渉して情報や証拠を提供し、該当国による処罰を要請する。』

7. 第四部 データベースの社会的背景について

1. 19. 第十六条【データベース運用上の倫理規定の国内外でのコンセンサスのプロセス】

『原則として、倫理規定の定着までに導入期と形成期と定着期の3段階制をとる。形成期が完了しなければ、コンセンサスは得られていないものとみなす。コンセンサスを得られていない限りは、統合医科学データベースの実運用を行ってはならない。』

1. 20. 第十七条【倫理規定ガイドラインの見直し】

『統合医科学データベース倫理規定ガイドラインについては社会的背景を考慮し、定期的に見直しを行うものとする。』

区 分	☐ 中核機関（ ☐ 代表機関／ ☐ 参画機関） ☒ 分担機関（ ☐ 代表機関／ ☒ 参画機関） ☐ 補完課題実施機関
課 題 名	「統合医科学データベース構築方式の開発」 （統合医科学データベース構築方式の開発：神経疾患を中心として）
実 施 機 関 名	国立大学法人 大阪大学
代 表 研 究 者 名	山本 洋一

1. 課題開始時における達成目標
①疾患・臨床医科学データベースに特化した統合技術開発に関する研究 ・多階層の統合医科学データベースを統合構築するために必要な統合技術の開発 ②疾患データベースの高度化と実証的統合に関する研究 ・具体的な定量目標として、最終年度までに 500 データのパーキンソン病患者の臨床情報の公開を目指す
2. 平成22年10月末時点における事業計画に対する成果
（1）成果概要 名称：パーキンソン病臨床情報データベース URL：http://pd-db.ais.cmc.osaka-u.ac.jp/ 特徴：臨床情報を 700 データ公開。複数疾患を横断的に統合する試験的实施として検索用 API を実装。また、ソースファイルを含めた技術ファイルも同サイトにて公開している。
（2）進捗及び成果 ・大阪大学医学部神経内科・脳卒中科を中心に 5 つの関連施設からパーキンソン病患者の臨床情報データを収集する「Secure Study システム」の一部を「パーキンソン病臨床情報データベース」として公開するプログラムを開発した（複数機関から提供されたデータの統合）。 ・倫理的課題については、東京医科歯科大と共同で、統合医科学データベース倫理規定ガイドラインを作成している。 ・「パーキンソン病臨床情報データベース」には、東京医科歯科大が開発した統合検索エンジンを紹介した検索用の API を実装しており、複数疾患の横断的統合を行った。当研究においては、「がん」と「パーキンソン病」という共通点の少ない疾患であったため、この 2 つのデータベースの統合の成果を問うことは難しいが、統合の仕組み自体は、疾患の種類に依存することなく汎用に設計しているので、今後、別の疾患を同じ枠組みで統合することが可能と考えている。 ・利用者が、検索結果（統計情報）を視覚的に容易に理解できるように、グラフ表示、経時間変化表示を取り入れた（添付）。 ・「パーキンソン病臨床情報データベース」を、日本神経学会、NPO 法人 SCCRE のホームページとリンクづけを行い、研究成果を共有できるようにした。 ・2010 年 10 月末の時点で当初の計画を超える 700 のデータを公開している。 ・これらの成果物は、同様の臨床データベースを構築する際に活用できるように、ソースファイルを含めた技術資料を公開している。

3. 当初目標に対する達成度
上記に記載したように、当初の目標は達成している。当初より当該研究はアクセス数の増大を目指す研究ではなく、臨床情報データベースの特殊性から、公開に向けての課題を整理し公開のしくみを確立することに主眼が置かれている。前項2で述べたように、パーキンソン病臨床情報データベースの公開・運用は実現されており、よって、この臨床情報データベースの公開の仕組みを確立するという目標は十分に達成したと言える。 2010年 ログイン数合計 336回 (2.7人/日)
4. 中間評価に対する対応
臨床情報データベースを統合医科学データベースとして統合し、公開するロールモデルを確立することを優先事項として求められたので、臨床情報の公開に至る標準的手法を確立し成果物を公表した。
5. 他機関との連携
代表機関の多階層のデータベースの構築・統合について、パーキンソン病臨床情報データベースをモデルとして、(A)複数機関から提供されたデータの統合 (B)異なる疾患データベースの統合を実施した。また、統合医科学データベース倫理規定ガイドラインを共同で作成している。代表機関との打ち合わせの実績については別紙に記載した。
6. 今後の見通し、計画、展望
今後、各疾患で規模は様々だが医療情報をデータベース化していく方向性にある。その目的は、多くの患者の経過を把握することで、より良い医療を構築していくエビデンスをつくることにある。そのためには、研究対象者を保護した上で、いかに効率よくデータを収集し解析できるかが重要である。以下論点を整理した。 1) 検体試料を使わない臨床情報データベース作成についての患者の同意について 「臨床研究に関する倫理指針」「疫学研究に関する倫理指針」の解釈および実運用については、各施設により違うのが現状である。社会的重要性に鑑みて、患者の同意の省略とそれに代わる方法について国民的議論を展開し、一定の方向性が必要と考える。 2) 費用の捻出 臨床情報データベースには、構築のための初期費用と継続のための維持費用が発生する。主にサーバーなど機器費用と管理する人件費である。長期に渡り継続し、データを収集することが重要であることから、公的資金については長期に使用できることが必須である。他の方法として、企業と共同研究し資金を調達することである。臨床情報を使った研究について、産学共同で行うことについても、国民的議論が必要である。 3) 患者への研究結果の還元 今後臨床情報データベースを利用した研究の結果を、患者に還元することで、患者がその重要性が認識すれば、さらに研究が円滑に実施されることになる。この積み重ねが、今後の臨床情報データベースの方向性を決めることに繋がるのではないかと考える。 臨床情報DBは、その特殊性から、問題点を多く持っている。しかし、患者さんの治療への還元という視点からは、その整備・利用は、医療に直接貢献できる可能性がある。医療情報は、1つの社会的資源でもあり、その利用について国民の理解を求めることがこの分野の発展に繋がるものと考え。今後のJSTライフサイエンスDB統合推進事業においても、公共性の高いアクセス数の高いデータベ

ースの支援だけでなく、医療に直接還元できる臨床情報データベースに日本としてどのように取り組むかは、極めて重要なことである。現状においては、臨床情報データベースが乱立していく方向にあり、記載内容が統一されず、入力医師の負担ばかりが増大する傾向にある。JST ライフサイエンス DB 統合推進事業には、全体を俯瞰し、統合する中心的存在としての機能を期待する。今後の方向性として、電子カルテと連動して必要な臨床情報を抽出し、データベース化する仕組みができれば、飛躍的な進歩が期待できると考える。

7. 全体総括

当初の目標は達成し、臨床情報の収集から公開までのデータベースのロールモデルを示すことができた。1 疾患の臨床情報のみからなるデータベースで、その一部を公開している例はない。当該研究に取り組み始めた頃に比べて IT 技術も格段に進歩し、使いやすい条件が整ってきた。我々の確立した手法を応用することで、より低コストで疾患単位のデータベースの作成が可能になってきたと思う。

8. 特記事項

1 疾患の臨床情報のみからなるデータベースで、公開している例はない。臨床情報のデータベースを各疾患で構築していくことは、世界的に試みられていることではあるが、持続できないのが現状である。データ入力者の協力は必須であり、なにかメリットがないとデータ収集はできない。我々のデータベースでは、入力者も自施設のデータを解析できるように工夫している。

9. 委託研究費一覧

	1 8 年度	1 9 年度	2 0 年度	2 1 年度	2 2 年度	計
設備備品費（千円）		5, 838	0	300	0	6, 138
人 件 費（千円）		4, 183	5, 809	1, 884	2, 027	13, 903
業務実施費（千円）		35, 434	21, 463	15, 998	9, 155	82, 050
一般管理費（千円）		4, 545	2, 727	1, 818	1, 118	10, 209
合 計 （千円）		50, 000	30, 000	20, 000	12, 300	112, 300

整備実績一覧

(1) データ（又はDB）の連結、統合化整備

通番	データ（又はDB）の名称	公開／未公開	概要（データの種別（生物種）・数量（kB 等）、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述）
1	パーキンソン病臨床情報データベース http://pd-db.ais.cmc.osaka-u.ac.jp/	公開	複数の医療機関から収集したパーキンソン病特定疾患臨床個人調査票の記載内容に投薬情報を加えた臨床情報を、患者の個人情報を除いて経年的に収集し、700データを公開している。 統計情報のグラフ表示、データの経年変化表示機能等、利用者の使い勝手を考慮した閲覧機能に加え、医科歯科大が本事業で開発した統合検索機能を用いた検索を可能とするAPIも実装されている。 公開年度 平成21年 平成22年 ログイン数合計 336回（2.7回／日）

(2) DB基盤システム、ツール等開発成果物の整備

通番	DB基盤システム、ツール等の名称	公開／未公開	概要（主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述）
1	パーキンソン病臨床情報データベース（仕様書・ソース）	公開	（1）に述べたパーキンソン病臨床情報データベースのプログラム（ソースコード）および仕様書を公開している。特定疾患臨床個人調査票は、パーキンソン病を含め56の疾患について難病情報センターのWebサイトにて一般公開されている。調査票の形式には共通点が多くあり、それらの疾患のDB開発時に本ソースコードを利用することで、低コストでの開発が見込まれる。また本プログラムは、医科歯科大が本事業で開発した統合検索エンジンからの検索に対応したAPIを実装しており、このプログラムで公開されたDBは異なる疾患を横断した統合検索の対象とすることができる。
2	Secure Study システム仕様書	未公開	（1）に述べたパーキンソン病臨床情報データを複数の医療機関から収集するSecure Study システム仕様書の公開を予定している。プログラム自体は本事業外で開発されたものであるためソース等の公開はできないが、「複数機関からセキュリティ・倫理的な問題を解決した上で臨床情報を収集するシステムの設計思想の一例を示す」という観点から、本仕様書の公開は、今後の臨床DB普及の促進のために意義があると思われる。現時点では、Webページに公開した問い合わせ先を通して希望があった場合には、本仕様書を提供する準備が整っている。今後、この形態を続けるか、自由にダウンロードできる形にするかを検討の上、本年度中に形態を最終決定して公開する。
3	パーキンソン病公開用臨床データ変換システム	未公開	前項で述べたSecure Study システムを用いて収集した臨床データから、インターネット上でオープンにアクセスできるDB上に公開して良い形のデータに変換し、1で述べたパーキンソン病臨床情報データベースに格納するシステムである。本事業で開発されたシステムであるので、ソースを含めた公開が可能であるが、変換元であるSecure Study システムのソースが非公開であることを勘案し、仕様書のみ、またはソースコードを含めた公開のいずれの形態を取るかを検討の上、本年度中に公開する。

パーキンソン病臨床情報データベース 画面抜粋

PD臨床情報DB - Windows Internet Explorer

http://pd-db.ais.cmc.osaka-u.ac.jp/

ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

PhishWall gredでチェック プロキシ使用 (http://pd-db.ais.cmc.osaka-u.ac.jp)

PD臨床情報DB

ホーム(F) フィード(F) メールを読む 印刷(P) ページ(P) セーフティ(S) ツール(T) ヘルプ(H)

ログイン



The Clinical Database of Parkinson's Disease

パーキンソン病臨床情報データベース

パーキンソン病臨床情報データベースについて

文部科学省の「ライフサイエンス分野の統合データベース整備事業」における「統合医科学データベース構築方式の開発」(<http://ibmd.tmd.ac.jp/SemanticWeb/>) (分担機関：東京医科歯科大学) で大阪大学は神経疾患を担当し、我が国のライフサイエンス研究を支える社会基盤としての多層階の統合医科学データベースを構築することを目指しています。

大阪大学が中心となって構築しているパーキンソン病データベースは、大阪大学医学部附属病院神経内科・脳卒中科ならびに関連病院の神経内科に受診している患者さんのデータを蓄積しています。その内容は、特定疾患の申請のために保健所などに提出される「パーキンソン病関連疾患臨床個人調査票」の各項目に加え、服用薬剤の種類と投与量などを患者データとして入力するということです。

これらのデータは、特定非営利活動法人、臨床研究教育支援センター (SCCRE) のサーバーコンピュータに厳重に管理、保存されています。データ内の個人情報と臨床情報とを分離し、匿名化された臨床情報の一部をこの公開サイトに公開しております。各施設における、患者さんからの同意取得および倫理委員会の認可の進捗にともない、公開できるデータ数が増えることで、より信頼性の高いデータベースの構築が期待できます。ここには無作為に抽出したパーキンソン病患者さんの700ほどのデータが入っており、前述の通り個人情報は削除されています。基本情報 (性別・年齢)・臨床情報 (重症度、発症年齢、初発症状など)・治療情報 (処方された薬剤など) の検索条件を入力することにより該当するデータが表示されます。今後データベースの解析結果を日常臨床を通してパーキンソン病患者さんに還元して、より良い医療を追求していきたいと考えています。

大阪大学医学部 神経内科学
データベース開発：(株)日立製作所

お知らせ

- ・統合データベース関連公開資料は現在ダウンロード可能です(2010/10/19)。
- ・データについて(2010/10/19)。
- ・初発症状特異性選択が可能なので、総数と各項目のデータ数の和が整合しません。
- ・Q/F (L) 受容体作動薬変異系(非変異系)の総データ数(1)は「薬剤の効果あり・なし・不明」の合計データ数(2)が多くなる場合があります。(1)は「変異(非変異)系薬剤使用のデータ数」であるのに対し、(2)は「変異(非変異)系薬剤(数種類)のいずれの薬剤使用合計数となりますので、一人の患者さんが複数種類の作動薬を使用している場合(2)のデータ数が多くなり、このような表記となります。
- ・データベースを更新しました(2010/10/04)。
- ・検索条件に「年齢」を使用する場合は「10」～「99」の範囲で設定して下さい(2010/10/04)。
- ・統合データベース関連公開資料は現在準備中です(2010/10/04)。

リンク

- SCCRE特定非営利活動法人 臨床研究・教育支援センター
- 大阪大学医学部 神経内科学講座
- 日本神経学会
- DBCLS(ライフサイエンス統合データベースセンター)
- 統合データベースプロジェクト(文部科学省)
- 統合医科学データベース(東京医科歯科大・阪大Gr)
- 旧ログイン画面

今後順次追加予定です。

公開資料

本データベースのプログラムや技術情報などを公開しております。ダウンロードしてご利用下さい。

ソースプログラムと設計資料(514MB)

お問い合わせ

大阪大学医学部神経内科学 データベース担当者係

メールアドレス: pd-db@neuro.med.osaka-u.ac.jp

セキュリティポリシーの概要と連絡先

検索条件入力 - Windows Internet Explorer

http://pd-bais.cmc.osaka-u.ac.jp/loginServlet

検索条件入力

基本情報
性別
年齢
臨床情報
治療情報
年度情報

検索結果入力

検索条件: [性別]男, [初発症状]初発症状(筋強剛) 28データ/701データ

パーキンソン病(28データ)

性別
男性 28データ
女性 0データ

発症年齢
～19歳
20～39歳
40～64歳
65歳～

発症お進捗性
ある 28データ
ない 0データ

初発症状
振戦
動作緩慢
筋強剛
姿勢反射の障害
その他

臨床所見
静止時振戦
ごくわずか
軽度

症状内訳グラフ表示

臨床疾患データベース
臨床所見グラフ

静止時振戦
28データ/701データ

静止時振戦
振戦
動作緩慢
筋強剛
姿勢反射の障害
その他

静止時振戦
振戦
動作緩慢
筋強剛
姿勢反射の障害
その他

終年情報一覧

静止時振戦の経
該当人数: 5人

患者ID 1年目 2年目 3年目 4年目 5年目 6年目 7年目

1年目 2年目 3年目 4年目 5年目 6年目 7年目

1度 2度 3度 4度 5度

1年目 2年目 3年目 4年目 5年目 6年目 7年目

1度 2度 3度 4度 5度

重年度グラフ

ホーム(M) フィード(F) メールを読む

臨床疾患データベース

Hoehn & Yahrの臨床重症度分類グラフ

検索条件: [性別]男, [初発症状]初発症状(筋強剛) 28データ/701データ

Hoehn & Yahrの臨床重症度分類

データ数

1度 2度 3度 4度 5度

重症度

戻る

薬剤使用割合グラフ

ホーム(M) フィード(F) メールを読む

臨床疾患データベース

薬剤使用割合グラフ

検索条件: [性別]男, [初発症状]初発症状(筋強剛) 28データ/701データ

薬剤使用割合

データ数

0 5 10 15 20 25 30 35

薬剤

L-DOPA薬剤
塩酸アマタジン
抗コリン薬
塩酸セリグリン
ドロキシドパ
ドパミン受容体動作薬

戻る

臨床疾患データベース

シノニム検索条件入力

姿勢

シノニム検索

検索

戻る

臨床疾患データベース

統計一覧

検索条件 : 姿勢保持障害, 典型的姿勢, 前屈姿勢

シノニム検索条件 : 前屈姿勢

パーキンソン病(689データ)

性別	
└ 男性	301データ
└ 女性	388データ
発症年齢	
└ ~19歳	0データ
└ 20~39歳	19データ
└ 40~64歳	384データ
└ 65歳~	286データ
発症は進行性で	
└ ある	677データ
└ ない	12データ
初発症状	
└ 振戦	455データ
└ 動作緩慢	150データ
└ 筋強剛	44データ
└ 姿勢反射の障害	41データ
└ その他	17データ