

統合データベース講習会@加賀 生命科学横断検索の利用法

大学共同利用機関法人 情報・システム研究機構
ライフサイエンス統合データベースセンター
特任准教授
川本 祥子

世の中にはたくさんのデータベースが存在



ディレクトリ型のデータベース検索 統合DBのデータベースカタログ

DB型(329)

- カタログ(7)
- データバンク(19)
- プログラム(66)
- プロジェクト(131)
- リソース(8)
- 解析サービス(15)
- 辞典(38)
- 知識モデル(27)
- 注釈(18)

対象(284)

- DNAモチーフ(1)
- DNA制御領域(3)
- DNA多型(25)
- DNA配列(42)
- DNA立体構造(1)
- RNA(5)
- アノテーション(24)
- タンパク質アミノ酸特性(1)
- タンパク質プロテオーム(5)
- タンパク質モチーフ(5)
- タンパク質機能(8)
- タンパク質局在(1)
- タンパク質抗体(2)
- タンパク質配列(10)
- タンパク質変異(1)
- タンパク質立体構造(21)
- マーカー分子(1)
- 遺伝子
- 画像(8)

表示形式: サムネイル | 表示順: 名前順 | 検索

COG - Clusters of Orthologous Groups of proteins
タンパク質オルソログ・グループ

詳細画面

LsdbID	133
Title	COG - Clusters of Orthologous Groups of proteins
URL	http://www.ncbi.nlm.nih.gov/COG
LsdbName	タンパク質オルソログ・グループ
Category	辞書: プログラム 対象: タンパク質, 配列
Catalog	The Bio Notebook Geneexpress NAR catalog MetaDB Bioformat
Organization	NCBI::USA
Desc-J	ゲノムが解析された原核/真核細胞生物のタンパク質を、オルソロググループに分けたデータベース。グループは機能カテゴリー(COアノテーション)ごとに分類されている。真核生物ゲノム上で予測されたオルソログを同様に分類したデータベースはOGGとして別途公開されている。

H20年3月現在の登録状況
国内 350, 国外 50

横断検索は
多数の異種のデータベースの中身
(レコード)を一括して検索する
サービスです。



生命科学データベース横断検索の概要

- 国内外のデータベースを対象にしたGoogleライクなキーワード検索
- 分子データベースから文献や特許が対象に。
- 対象DB数 220DB, 対象文書数 787万

生命科学データベース横断検索

データベース | HELP | 検索結果URL表示

データベース

全てのデータベース (14267)

- 統合DBプロジェクト (20)
- 文献 (101)
 - 蛋白質核酸糖質系 (47)
 - 蛋白質核酸糖質系 (近刊) (2)
 - 文科学「ゲノム特定領域」年次報告書 (16)
 - 文科学「旧ゲノム特定領域」最終報告書 (31)
 - BIOPHYIS: 生物物理 (3)
 - 科学新聞記事 (2)
- 学会要旨 (132)
 - 日本農芸化学会大会要旨 (131)
 - 日本生物物理学会大会発表要旨 (1)
- 特許関連文書 (8230)
 - 日本国特許 (8230)
- 用語解説 (376)
 - 日本語版ウィキペディア (290)
 - Wikipedia (English) (86)
- ゲノム・遺伝子・RNA (1849)
 - 遺伝子発現・転写制御 (10)
- タンパク質 (78)
- 代謝・相互作用・生体反応 (2)
- 糖・脂質 (10)
- 細胞・組織 (0)
- 発生 (0)
- 医療・薬 (3407)
- 食品・栄養 (52)
- 農学・環境 (0)
- 生物圏系・分類 (0)

共通 お気に入り

データベースの詳細情報

日本生物物理学会大会発表要旨

収録データ数: 7114

収録期間:

Synopsis: 生命現象の解明は、自然科学における究極の目的の一つです。物理的手法と生物科学との融合により、この目標に対して新しい研究が進展しています。問題の解決に向けて、研究が進展しています。

検索結果 ALL

OMIM: #145500 HYPERTENSION, ESSENTIAL [omim]

#145500 HYPERTENSION, ESSENTIAL; EHT *FIELD* TX A number sign (#) is used with this entry because of genes contribute to the phenotype of essential hypertension. Candidate genes include those encoding angiotensin chromosome 1p36. Susceptibility loci for essential hypertension have been mapped to chromosomes 17 (HTT1; 603918),

<http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=145500>

サブリットデータベース [supdb]

血圧が高いカニズム 血圧とは、心臓から送り出された血液が動脈の壁を押す力、つまり血管にかかる圧力。基準値は、収縮期血圧140mmHg/拡張期血圧90mmHg以上が 高血圧。120/80mmHg未満を正常あるいは適血圧としている。その間には、正常、正常高値、正常前値、高血圧 前症などのレベルがある。慢性的に血圧が高い状態が続く

http://www.j-sup.com/sup_db/?mode=syoujou&id=78

@ヘルスケアインフォメーション: 高血圧症の診断 [healthcare]

高血圧 症の診断 日頃の血圧が130mmHg未満/85mmHg未満である場合は正常血圧としており、血圧値が140/90mmHg以上の場合、高血圧と定義しています。その中間領域の130~139/85~89mmHgについて、正常高値血圧として将来、高血圧 症に移行する可能性があるものと、われわれは注意を要する領域

<http://www.healthcare.city.chuo.yamanashi.jp/database.php?id=28>

RefSeq: NP_115763/WNK lysine deficient protein kinase 4 [Homo sapiens] [refseq]

My NCBI [Sign In] [Register] PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books ons and cellular mechanisms. Familial hyperkalemic hypertension -causing mutants of WNK4 exhibit differences regula al role in hypercalciuria of familial hyperkalemic hypertension caused by gene mutation of WNK4 JOURNAL Am. J. Phy

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=protein&id=NP_115763

蛋白質核酸糖質系: ANPの血圧調節作用 [pne]

蛋白質核酸糖質系 33 14 1988 2490-2495 特集ANP 体液量と血圧調節に働くホルモン！心臓性ナトリウム利尿 Pの分泌は、主として心臓壁の伸張に惹くものであり、高血圧 患者にみられる血漿ANP濃度上昇は、心臓圧の上昇の結果と 考えられる。本態性 高血圧 症において、ANP分泌が障害されていることを示唆する事実

[http://lifesciencedb.jp/dbsearch/Literature/get_pne_cpdf.php?year=1988&number=3314&file=Cj4F7GuzMVB2RthWQ6xg==&search=\"%E9%AB%98%E8%A1%80%E5%9C%A7\"](http://lifesciencedb.jp/dbsearch/Literature/get_pne_cpdf.php?year=1988&number=3314&file=Cj4F7GuzMVB2RthWQ6xg==&search=\)

RefSeq: NP_150234/SA hypertension-associated homolog [Rattus norvegicus] [refseq]

My NCBI [Sign In] [Register] PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books D STS tRNA

microRNA Exon 1: NP_150234 . Reports SA hypertension -a...[gi:65301416] BLink, Conserved Domains, Links 150234 580 aa linear ROD 25-SEP-2007 DEFINITION SA hypertension -associated homolog [Rattus norvegicus]. ACCESSION

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=protein&id=NP_150234

Page 1 of 206

検索結果 14267件中 1 - 50件目を表示

遺伝子リスト=> Human

acyl-CoA synthetase medium-chain family member 3 [GENE:6296][GNP:6296][UNP:][HGNC:10522]

Hypertension, essential, susceptibility to, 2 [GENE:50986][GNP:50986][UNP:][HGNC:]

Hypertension, essential, susceptibility to, 4 [GENE:444980][GNP:444980][UNP:][HGNC:]

Hypertension, essential, susceptibility to, 3 [GENE:387575][GNP:387575][UNP:][HGNC:]

primary pulmonary hypertension 2 [GENE:89873][GNP:89873][UNP:][HGNC:15733]

Hypertension, essential, susceptibility to, 1 [GENE:117191][GNP:117191][UNP:][HGNC:]

hypertension with brachydactyly [GENE:8080][GNP:8080][UNP:][HGNC:]

Renal failure, progressive, with hypertension [GENE:59331][GNP:59331][UNP:][HGNC:]

様々な検索サービス

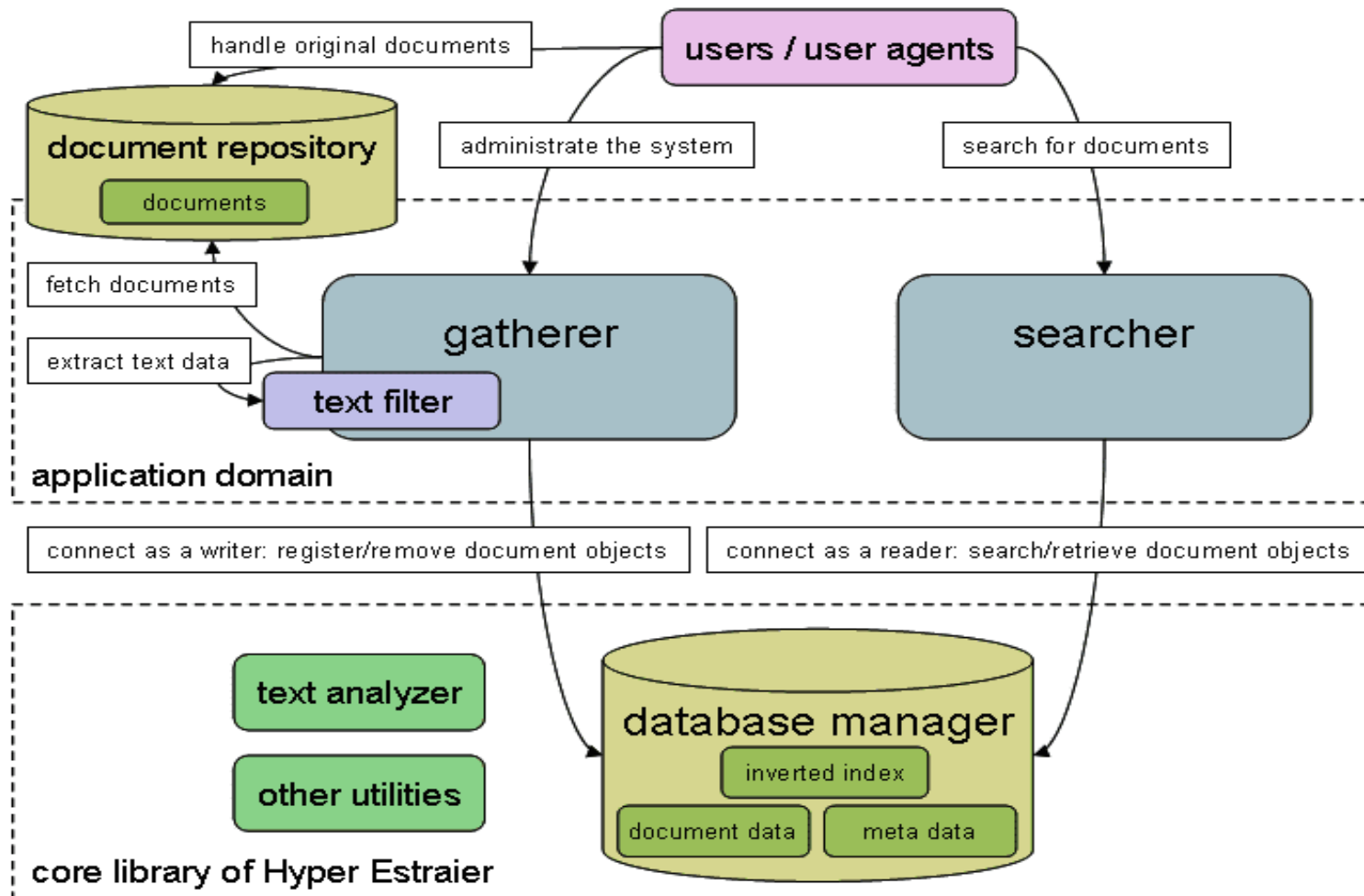
- Google — 検索といえばこれ
- Yahoo — 日本で強いシェア
- Bing — マイクロソフトの検索エンジン
- PubMed — 生命科学の文献検索
- Entrez — NCBIの統合検索プラットフォーム
- 生命科学横断検索 — 生命科学に特化
- その他色々、用途にあわせて使いこなす



生命科学横断検索とは

- 分子や文献のGoogle型の**キーワード検索**
- キーワードの日英翻訳
- 国内のDB統合の第一歩として、データベースは各提供サイトに置いたまま、外部から検索する**横断検索**を実現したもの
- 分子だけでなくPNEや特許などの**文献も同時に検索**
- 検索エンジンとしてフリーソフトウェアの**HyperEstraier**を利用

横断検索エンジン: HyperEstrailer の概略



HyperEstraiierのインデックス

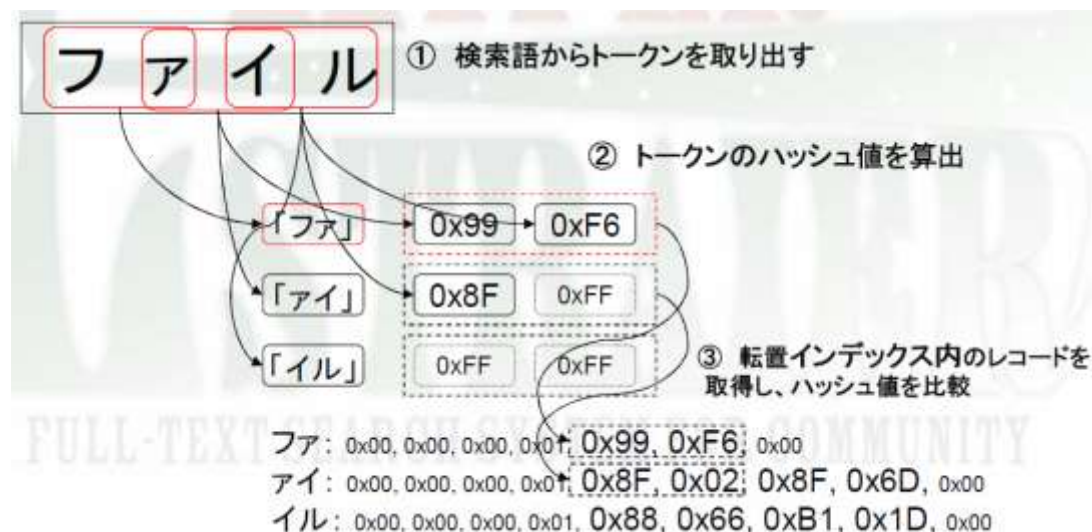
• N-gram

文書IDとオフセット情報をレコードとして保持



• N.M-gram

文書IDと後続M個のトークンのハッシュ値をレコードとして保持



生命科学データベース横断検索システムのサーバ概略図

このシステムは:

生命科学に特化したGoogleライクな検索サービス。検索エンジンにはHyperEstraierを使用。分子DBから文献、特許まで。

Webサーバ

検索インタフェイス
結果表示

質問
答え

利用者

クラスタサーバ

検索ノード

クロウリング
インデックス作成
検索
HyperEstraier

インデックス
インデックス
インデックス

検索ノード

クロウリング
インデックス作成
検索
HyperEstraier

インデックス
インデックス
インデックス

検索ノード

クロウリング
インデックス作成
検索
HyperEstraier

インデックス
インデックス
インデックス

検索ノード

クロウリング
インデックス作成
検索
HyperEstraier

インデックス
インデックス
インデックス

検索ノード

クロウリング
インデックス作成
検索
HyperEstraier

インデックス
インデックス
インデックス

X20ノード

生命科学関連
データベース

遺伝子DB

ゲノムDB

タンパク
DB

遺伝子発現
DB

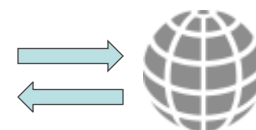
パスウェイ
DB

微生物DB

文献

辞書

etc



クロウリング
データダウンロード
リンク

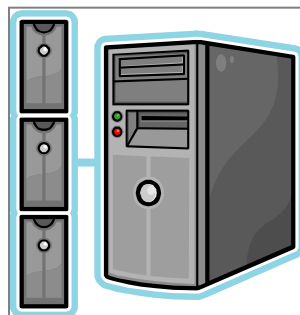
> 200DB



横断検索の連携システム概要

文科省統合DB横断検索サーバー

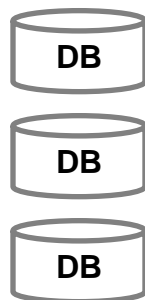
①各機関でHyperEstraierをインストールし検索プログラムを常駐させる。



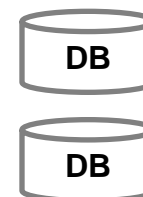
②インターネット上の他機関のサーバのIPとポート番号を指定し、LAN内と同様にインデックスを検索する。

検索インデックスを分散共有

経産省統合DB
(MEDALS)



産総研日本糖鎖科学
統合DB(JCGGDB)



検索してみます

★ お気に入り ★ OTTAVA Google

Twitter / ホ... Gmail - Inbo... Dropbox - Fil... ライフサイ... x

Guest Account | アカウント | ログイン | English

LSDB 文部科学省委託研究開発事業
統合データベースプロジェクト

ホーム データベース 検索 ツール ダウンロード About us

統合ホームページへようこそ

はじめての方へ: サイトの内容をムービー やリーフレット でご紹介しています。
「統合データベースプロジェクト」とは

ポータル

- 生命科学系 データベース カタログ
- 生命科学系 学協会カタログ
- 生命科学系主要プロジェクト一覧
- 生物アイコン
- WingPro (JSTのDBポータル)
- Webリソースポータルサイト (JST解析ツールポータル)

アーカイブ

- 生命科学系データベースアーカイブ
- DBJトレースアーカイブ (遺伝研 DBJ)
- DBJリードアーカイブ (遺伝研 DBJ)

ツール & 解析サービス

- BodyParts3D/Anatomography
- Wired-Marker
- MiGAP (微生物ゲノムアノテーションパイプライン)
- DBCLS Galaxy

検索

- 生命科学データベース横断検索
- 蛋白質核酸酵素 全文検索
- 文科省「ゲノム」研究報告書 全文検索

新着情報

- アグリバイオインフォマティクスのセミナーが開催されます 2010-6-30 (Wed)
- CBRC2010にてデモを行います 2010-6-28 (Mon)
- 第50回生物物理若手の会夏の学校にてランチョンセミナーを行います 2010-6-25 (Fri)

LSDBブログ

- 相同性検索と系統解析 ~BLASTとClustalW~ 2010-07-05 (Mon) 18:39:19
- TogoProの使い方 ~検索と連携~ 2010-07-05 (Mon) 15:34:27
- SAKURAを用いた塩基配列登録の方法 2010-07-05 (Mon) 09:20:28

ニュース

- パーキンソン病類縁疾患 神経障害の仕組み解明 放医研 千葉大 PETで観察

インターネット | 保護モード: 有効 100%

キーワードを入力

検索

- **口蹄疫**を検索してみましょう。

検索

- その他思いつくキーワードを入力。

生命科学データベース横断検索のIF

The screenshot shows the DBCLS cross-database search interface. The interface is divided into several sections:

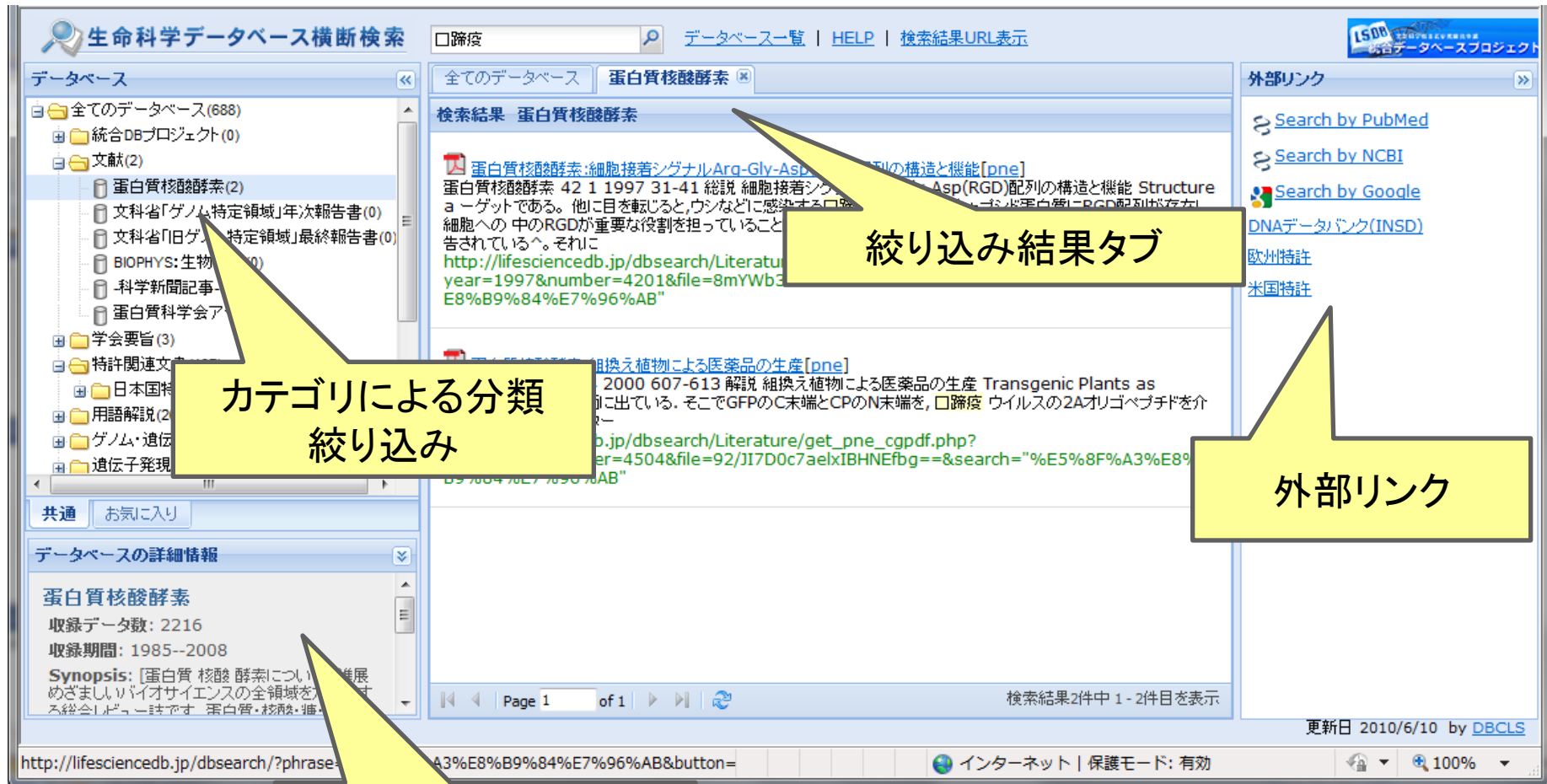
- Search Bar:** Located at the top, containing a search input field and buttons for "データベース一覧" (Database List), "HELP", and "検索結果URL表示" (Display Search Result URL).
- Left Sidebar:** Contains a "データベース" (Database) section with a tree view of databases, including "統合DBプロジェクト(0)", "プロジェクト公開資料(0)", "日本の生命科学データベース政策(0)", "生命科学系主要プロジェクト一覧(0)", "生命科学系データベースアーカイブ(0)", "統合TV(0)", "生物アイコン(0)", "WingPro (JSTのDBポータル)(0)", "MEDALS(閲覧)(0)", and "AgriTogo(0)". Below this is a "文献(2)" (Literature) section with buttons for "共通" (Common) and "お気に入り" (Favorites).
- Main Content Area:** Displays search results. The top section is titled "全てのデータベース" (All Databases) and shows a list of results. The first result is from Wikipedia, titled "日本語ウィキペディア:口蹄疫 - Wikipedia[jpwiki]", with a brief description of the disease and a link to the Wikipedia page. The second result is from NIAH, titled "動物研:家畜の監視伝染病: 3口蹄疫[domesticanimalinfection]", with a brief description of the disease and a link to the NIAH page. Below the results is a "PDB Code Search[hct-pdbnavi]" section, which displays a list of PDB codes and their corresponding structures.
- Right Sidebar:** Contains links to external search engines, including "Search by PubMed", "Search by NCBI", "Search by Google", "DNAデータベース(INSDB)", "欧州特許", and "米国特許".

Annotations on the screenshot highlight specific parts of the interface:

- 検索ワード入力、実行部分** (Search word input, execution part): Points to the search bar at the top.
- 検索結果表示部** (Search result display part): Points to the main content area showing search results.
- 遺伝子名辞書検索結果表示部** (Gene name dictionary search result display part): Points to the "PDB Code Search[hct-pdbnavi]" section.

At the bottom of the page, there is a status bar showing the URL "http://lifesciencedb.jp/dbsearch/?phrase=%E5%8F%A3%E8%B9%84%E7%", the date "更新日 2010/6/10 by DBCLS", and the browser status "インターネット | 保護モード: 有効".

生命科学データベース横断検索のIF



生命科学データベース横断検索のIF

検索結果URL表示

生命科学データベース横断検索

□ 検索結果 URL 表示

データベース

- 全てのデータベース (688)
 - 統合DBプロジェクト (0)
 - 文献 (2)
 - 蛋白質核酸酵素 (2)
 - 文科省「ゲノム特定領域」年次報告 (1)
 - 文科省「旧ゲノム特定領域」年次報告 (1)
 - BIOPHYS: 生物物理 (0)
 - 科学新聞記事 (0)
 - 蛋白質科学会アーカイブ (0)
 - 学会要旨 (3)
 - 特許関連文書 (137)
 - 日本国特許 (137)
 - 用語解説 (266)
 - ゲノム・遺伝子・RNA (56)
 - 遺伝子発現・転写制御 (0)

共通 お気に入り

データベースの詳細情報

蛋白質核酸酵素

収録データ数: 2216
収録期間: 1985-2008

Synopsis: [蛋白質 核酸 酵素について] 進展めざましいライフサイエンスの全領域をカバーする総合データベースです。蛋白質・核酸・糖・脂質など

検索結果 蛋白質核酸酵素

1 総説 細胞接着シグナルArg-Gly-Asp(RGD)配列の構造と機能 [pne]
感染する口蹄疫 ウイルスのキャプシド蛋白質にRGD配列が存在し、
ことよすで述べたが、口蹄疫 ウイルスの結晶構造の解析結果も報
structure/get_pne_cgpdf.php?
b3RlC7PRA8cm4reJGw==&search="%E5%8F%A3%

外部リンク

- Search by PubMed
- Search by NCBI
- Search by Google
- DNAデータバンク(INSDB)
- 欧州特許
- 米国特許

データベース一覧

DB名	DB提供機関	ノード...	収録...	収録期間	更新日
プロジェクト公開資料	DBCLS	DBCLS	163	2009-05-13	2010-03-17
日本の生命科学データベース...	DBCLS	DBCLS	87	2009-05-13	2010-06-22
生命科学系主要プロジェクト一覧	DBCLS	DBCLS	67	2009-06-05	2010-03-17
生命科学系データベースアーカイブ	DBCLS	DBCLS	366	2009-06-05	2010-03-17
統合TV	DBCLS	DBCLS	317	2010	2010-03-17
生物アイコン	DBCLS	DBCLS	636	2010	2010-03-17
WingPro (JSTのDBポータル)	WingPro	DBCLS	1715	2010	2010-03-17
MEDALS便覧	JIBC and AIST	DBCLS	157	2008-2010	2010-05-24
AgriTops	農業生物資源研究...	DBCLS	10	2008-2010	2010-06-07
蛋白質核酸酵素	共立出版	DBCLS	2216	1985-2008	2010-03-17
文科省「ゲノム特定領域」年次...	DBCLS	DBCLS	331	--	2010-03-17
文科省「旧ゲノム特定領域」最...	文科省	DBCLS	610	2009-07-03	2010-03-17
BIOPHYS: 生物物理	日本生物物理学会	DBCLS	2659	2010	2010-03-17
科学新聞記事	The Science News	DBCLS	182	2010	2010-03-17

2010/6/10 by DBCLS

100%

検索結果の絞り込み

Try

- iPS細胞で検索してみましょう

Try

- 検索結果から文献を見えます

横断検索を使って

Try

- 問題:iPS細胞の樹立に利用された因子の名前を調べましょう

横断検索を使って

- 問題:iPS細胞の樹立に利用された因子の名前を調べましょう
- 正解例:Oct3/4・Sox2・Klf4・c-Myc
- 山中因子、山中カクテルなどと呼ばれています。

横断検索を使って

- それぞれの遺伝子について次のような表にまとめましょう。

Gene Symbol	Gene Name	Species	解説	Entrez Gene ID	RefSeq ID
POU5F1	POU class 5 homeobox 1	Homo Sapiens	POUホメオドメインを持つ転写....	5460	NM_001173531.1
SOX2	SRY (sex determining region Y)-box 2	Homo Sapiens	OXファミリー転写因子のひと....	6657	NM_003106.2
KLF4	Kruppel-like factor 4 (gut)	Homo Sapiens	Transcription factor which act..	9314	NM_004235.4
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	Homo Sapiens	多機能な核内リン酸化タンパ....	4609	NM_002467.4

医薬品、化合物の検索

GenomeNet

KEGG KEGG2 PATHWAY BRITE DRUG DBGET



環境設定 ヘルプ

Search for

ゲノムネット

ゲノムネットとは
お知らせ
謝辞

KEGG

KEGGの概要
リリース情報

統合データベース

統合DBの概要
DBGETの概要
リリース情報
データベース増加図

医薬品データベース

利用法

研究支援データベース

計算ツール

その他のツール

フィードバック

ゲノムネット医薬品データベース

ゲノムネットでは、[文部科学省統合データベースプロジェクト](#)の一環として、研究の最先端と医療の現場さらには一般社会とをつなぐ日本語の医薬品統合データベースを開発しました。具体的には JAPIC 医薬品添付文書情報を KEGG DRUG と統合したデータベースです。以下のリンクから検索システムをご利用ください。

[JAPIC 医療用医薬品データベース](#)

[JAPIC 一般用医薬品データベース](#)

ゲノムネット医薬品データベースは薬に対する科学的理解を深めることを目的とした情報提供サービスです。実際の薬の使用は医師・薬剤師の指示で行ってください。

本サービスは[日本医薬情報センター](#)との契約に基づき提供しています。大量データの一括取得は禁止されています。短時間に大量のアクセスがあった IP アドレスは予告なしに使用停止になる場合があります。

関連サイト

- [JAPIC iyakuSearch](#)
- [PMDA 情報提供サービス](#)
- [FDA Drug Information](#)
- [DailyMed](#)
- [KEGG DRUG 日本語版](#)

• 統合DBプロジェクト 京都大学

糖鎖関連の検索

jCGGDB

日本糖鎖科学統合データベース

[Japanese](#) [English](#)

[ホーム](#) | [設立・概要](#) | [参画機関募集](#) | [運営事務局](#) | [データベース](#) | [ログ統計](#) | [リンク](#)

糖鎖関連遺伝子

- ・糖転移酵素・糖分解酵素
- ・糖ヌクレオチド輸送体
- ・硫酸転移酵素
- ・レクチン
- ・合成と分解パスウェイ
- ・ライブラリー化

糖鎖欠損動物・糖鎖改変動物

- ・フェノタイプ解析
- ・ライブラリー化

糖鎖関連疾患

- ・先天代謝異常
- ・癌転移
- ・感染
- ・アレルギー

糖鎖機能

- ・タンパク質の機能調節
- ・細胞間コミュニケーション

日本糖鎖科学
総合データベース
(糖鎖のポータルサイト)

jCGGDB

糖鎖大量合成

- ・ケミカル
- ・ライブラリー化

糖鎖構造の分析・検出

- ・質量分析計
- ・2-D / 3-D 糖鎖マッピング法
- ・糖鎖抗体
- ・レクチン

糖タンパク質

- ・生物種別、組織別のタンパク質およびその糖鎖構造

糖鎖関連分化マーカー

- ・癌、免疫、再生医療、受精



NEWS

- ・アクセスログの統計を公開しました。[ログ統計](#)(2008/12/17)
- ・産総研のGlycan Mass Spectral Databaseを公開しました。[GMDB](#)(2008/12/4)
- ・GlycoForumのサイトを横断検索に加えました。(2008/10/24)
- ・2008年8月19日(火)つくば国際会議場で行われます第28回日本糖質学会年会で統合DBプロジェクトの活動を報告する予定です。