

グローバルCOEプログラム系統講義「ベーシックサイエンスコース」

**バイオインフォマティクス：**

**生物系研究者における情報リテラシーとしての配列解析**

**Bioinformatics: Sequence analyses as information literacy of life scientists**

嶋田 誠

(藤田保健衛生大学 遺伝子発現機構学)

1日目：2010年 9月28日(火) 17:00-18:30 基礎医学研究棟1階会議室

2日目：2010年10月 4日(月) 18:00-19:30 基礎医学研究棟1階会議室

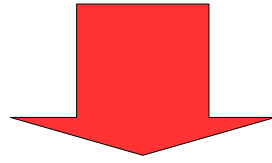
# アウトライン

- [第1回]
  - 概論
    - ねらい
    - バイオインフォマティクス:どこまで把握すべきか?
  - データベース
  - ゲノム関連大規模プロジェクト
- [第2回]
  - バイオインフォマティクス独学のコツ
  - ツール群の分類と見つけ方
  - 実験研究者にとってのプログラミング
    - アルゴリズム
    - データ加工変換
    - パース
  - 配列研究関連ツールの紹介



# ねらい

- 実験系の研究者が、
- 今後、それぞれの研究活動において、
- 適宜、必要なデータやツールを探し出して、使えるようになること。



## 基本的に独学:

- 調べ方
- 他人への尋ね方
  - --> 基本的用語の解説
- 最初の段階で重要な事は何か
  - データ構造の解釈 & データ入出力の仕方
  - --> 将来 in silico で解析しやすい実験データのまとめ方

# バイオインフォマティクス： どこまで把握しておくべきか？

実験室での作業だけでも時間管理が大変なのに。

- バイオインフォマティクス用語
  - やたらと略語が多い？
- データベースやツール群
  - どこになにがあるのか。
- 計算手法・アルゴリズム
  - ソフト任せのブラックボックス状態でいいのか？



# 「どこまで」を考えるうえでのヒント1

- 開発者とユーザーとがいる
  - 開発者：
    - プログラムを組んでソフトやDBを作りだす。
    - アルゴリズムを考える。
    - バイオインフォマティクスそのものの未来を考える。
  - ユーザー：
    - 自分のデータを解析する。
    - 解析のアイデアを練る。-->開発者へ
- ユーザーは開発者にならなくてもいい。
  - 開発者(仲介者)とコミュニケーションをとる能力は養えるのでは。
    - ヘルプ、マニュアル文書、メール、打合せ等
    - ユーザーとして必要な用語や概念だけでも理解しておきたい。

# 「どこまで」を考えるうえでのヒント2

- 研究手法・解析手法における栄枯盛衰
  - DBは激しい。
  - 手法やアルゴリズムにおける原理は様々な分野で繰り返し利用されている。
    - 例) モンテカルロ、パーシモニー、ニューラルネットワーク
- 
- DBは必要な時に必要なものを探し出す技術を。
- 手法やアルゴリズムは原理を(一度は)理解し、アレルギーを起こさないようにしておく。

# 「どこまで」を考えるうえでのヒント3

- 研究は常に新発見を求め、新しきことを試す行為なり。

ならば、

- 既存のツールでは限界がある。
  - 別の処理と組み合わせたり、修正したり、が必要。
  - 開発といっても、全く無から新規を作り出す必要はない。
    - 「車輪の再発明はするな。」
- 「数こそ力」の課題もある。

ならば、

- 繰り返し作業の得意なコンピュータを利用しよう。

# 「どこまで」を考えるうえでのヒント4

- Publicationは論文だけではない。
  - 論文投稿前に、データをデータベースに登録する必要がある。
    - 配列データ、発現データ、立体構造など
  - データベース管理者側も親切にマニュアルを用意している。
    - eg., 次世代シーケンサーによるデータの登録システムのマニュアル  
[http://trace.ddbj.nig.ac.jp/files/manual/DRA\\_submission\\_manual.pdf](http://trace.ddbj.nig.ac.jp/files/manual/DRA_submission_manual.pdf)
    - 技術革新の速さ、研究の多様性から、想定外のことも多い。
    - データの形式を一括変換するのはユーザー(登録者)側。

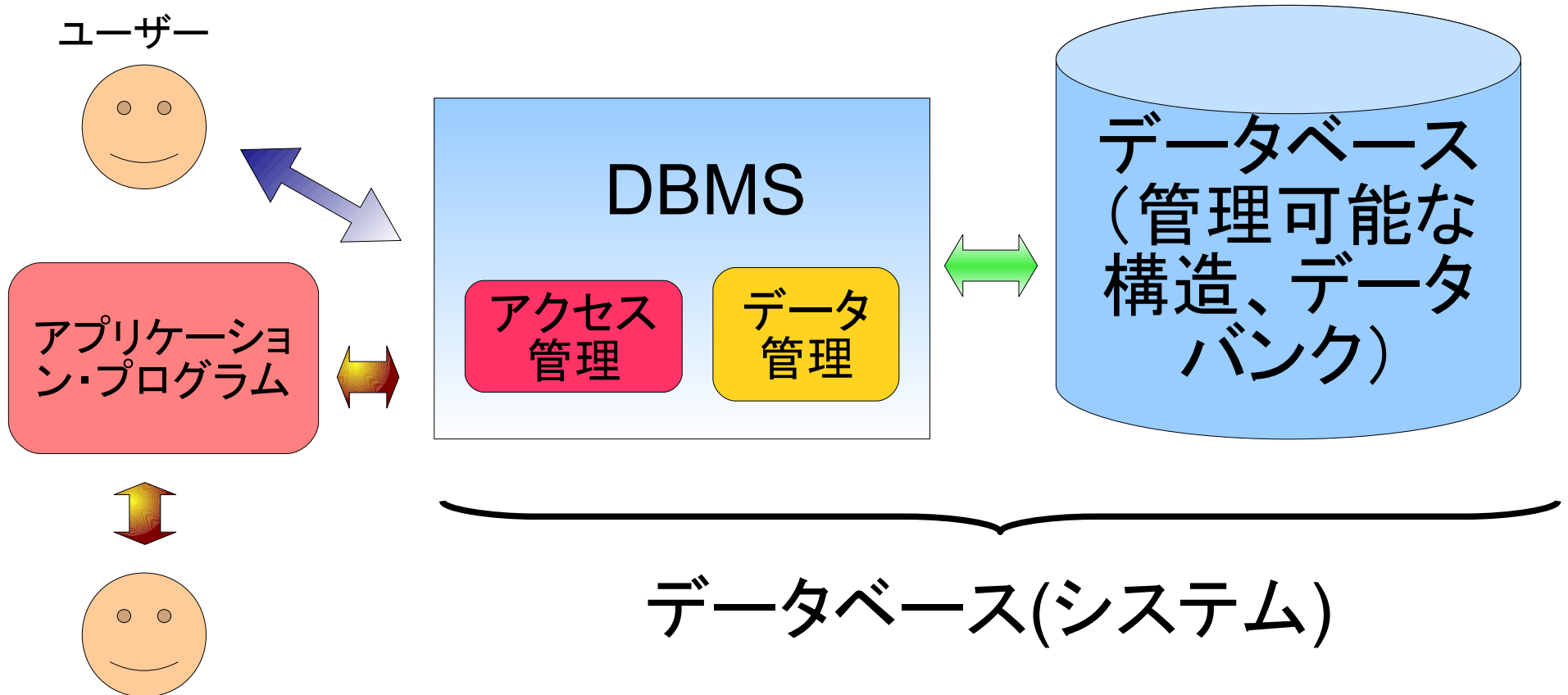


# データベース

- データベース(DB)とは
- 基本的なDB・ブラウザの紹介
  - 文献DBとMeSH
  - 核酸DBとannotation(注釈づけ)
  - タンパク質DBと配列の機能推定
  - ゲノムDBとゲノム・ブラウザ
- 必要なDBを見つけるために

# データベースとは？(問い)

- 「～をデータベース化して」vs.「～をリスト化して」
- データベース管理システム (DBMS)



# さらにDBMS

- DBMSが無いと、
  - プログラムから直接、ファイルに対してデータの保存や読み込みを行う
  - 個々のファイルのデータ構造を予め知っていないとアクセスできない
  - 複数のユーザーやアプリケーションに対応できない
  - 安全性・安定性に限界
- DBMS: 構造の設計から管理・運用を行う
- 現場では意外と聞かないことば
  - リレーショナル・データベース (RDBMS) が一般的

# リレーショナル・データベース (RDB: Relational Data Base) の概要 (1)

- データを表 (テーブル) 形式で表し、複数の表にリレーションシップ (関係付け) を結ぶことができる。
- 通常、元データの表はタブやコンマなどで区切ったテキストファイル形式。

HIT00004499413	-2	NHEJ1	rs10498064	A/T	T	fwd/T	Y	na	inconsistent
HIT0000155588	-1	KIAA0802	rs632423	C/G	C	fwd/B	Y	0.497+/-0.037	inconsistent
HIT0000414818	-1	C13orf18	rs2478044	C/G	G	fwd/T	Y	na	consistent
HIT00002846621	-1	IL16	rs4778639	G/T	T	fwd/B	Y	0.188+/-0.242	inconsistent

Column Name of Table S2

#1 HIT ID: H-InvDB Transcript ID

#2 Intron No: \_th intron in the transcript

#3 SitePos: see Figure 1a

#4 GeneSybl: HUGO gene symbol

#5 rs ID: rs SNP ID in dbSNP

#6 rsAlleles: alleles in dbSNP

#7 Ancestral Allele: ancestral allele estimated by NCBI and shown in dbSNP

#8 Ori/Str(rs): orientation and strand of the SNP, see ftp://ftp.ncbi.nih.gov/snp/database/Illumina\_top\_bot\_strand.note.txt

#9 number of locations SNP mapped: 'Y'='OneMap' or 'N'='MultiMap', see List of Abbreviation

#10 Het: heterozygosity with standard error shown in dbSNP

#11 AncAlvsGT-AG: whether estimated ancestral allele is consistent with GT-AG rule

record (行)

field, column (列)

# リレーショナル・データベース (RDB: Relational Data Base) の概要 (2)

- キー(key)の重要性:
  - 主キー(Primary key): レコードを一意に指定するフィールド(通常非冗長ID)
  - 外部キー(Foreign key): 2つの表を結び付けるキー
    - 応用例(DB間を結び付ける), <http://biodb.jp/>
- SQLと呼ばれるデータベース(問い合わせ)言語が標準化されている。
- 代表的実装例
  - 商用: Oracle Database, Microsoft SQL Server
  - オープンソース: MySQL, PostgreSQL, BerkeleyDB

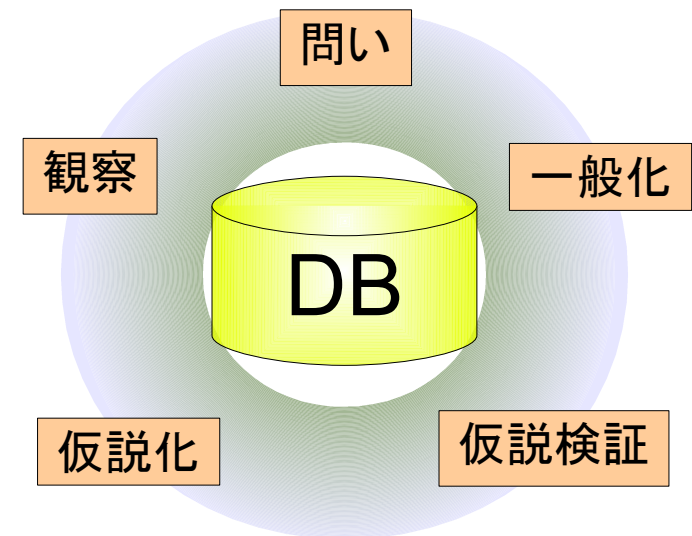
# データベースとは(答え)

- 「様々な目的を考慮して整理整頓されたデータの集まり」である。
- 設計思想をもつ(DBMSにて体现)。
  - 生物情報DBでは普通、
    - 検索機能、
    - web利用が前提、
    - リンク付け、
    - viewerやtool群とセットになっている。
- 始まりは、第二次大戦後の米軍が点在する情報を集約し、一か所にアクセスするだけで、様々な情報が得られるようにした「情報基地」



# DB(生物情報)の生命科学における意義

- データの種類の変化: 計算機利用前提→データ量の急速な増大
  - シークエンサーのデータ産出量 vs. CPU速度
  - 増加率 [2002(ABI3730)以降]: X2/8mo. vs. X1.5/8mo (Moors' Law)
- 研究活動の様々なステップにおいて活用される余地あり。
  - 膨大な量のデータ[DBに蓄積、仮説構築、実験による検証、DBに蓄積]
  - 帰納的(データ→法則) vs. 演繹的(法則→検証)(仮説driven vs. Data driven)
    - 情報リテラシーが研究活動の様々な局面で要求される



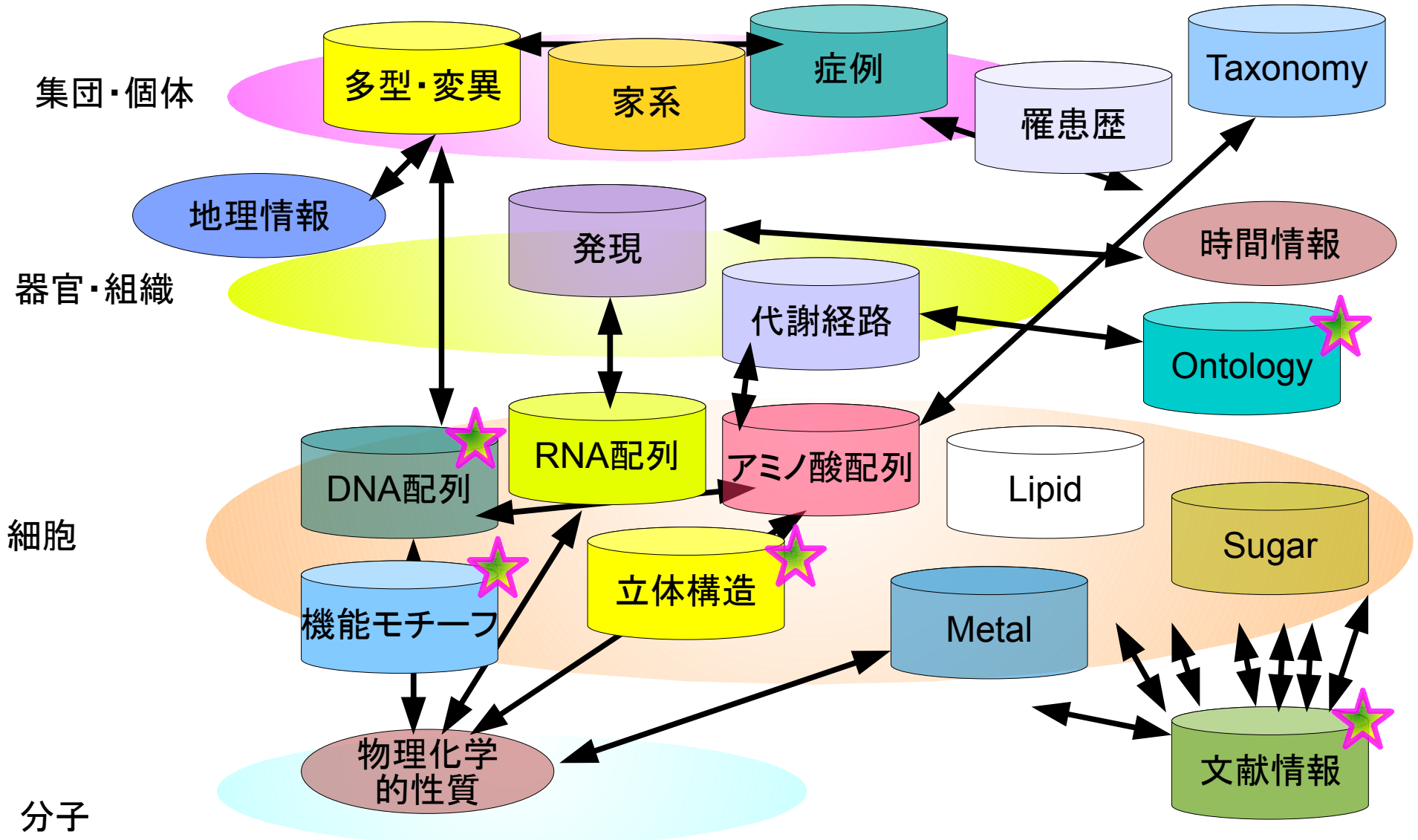
# 個々の研究者とDBとの関係

- 研究活動における公開原則と研究者の姿勢
  - 「配列データはDBに登録しなければ論文受理できない」
    - 研究成果は共通財産という考え→意図的に働きかけた。
      - →研究活動が活性化した。
    - DBによるデータの質の管理：
      - →論文著者には厳しいときも。
  - Curatorの言うことは聞きましょう。
  - きまりは守るためにある。。。





# 生物情報データベース：対象と次元



# DB検索に知っとくと得するキーワード

- 演算子(operator)、引数(オペランド operand)、  
例)  $a + 2$  の例では演算子と引数はそれぞれどれか  
答え) 演算子 = “+”、引数 = “a”, “2”
- ブール演算(Boolean Operation)
  - 論理演算(Logical Operation)ともいわれる。
  - 1(真)か0(偽)かの2通りの入力値に対して1つの値を出力する演算のこと。
  - プログラミングだけでなく、データベースを複合検索(eg., AND検索、NOT検索)する際にも用いられる。
    - データベースごとに約束事が違うので注意。

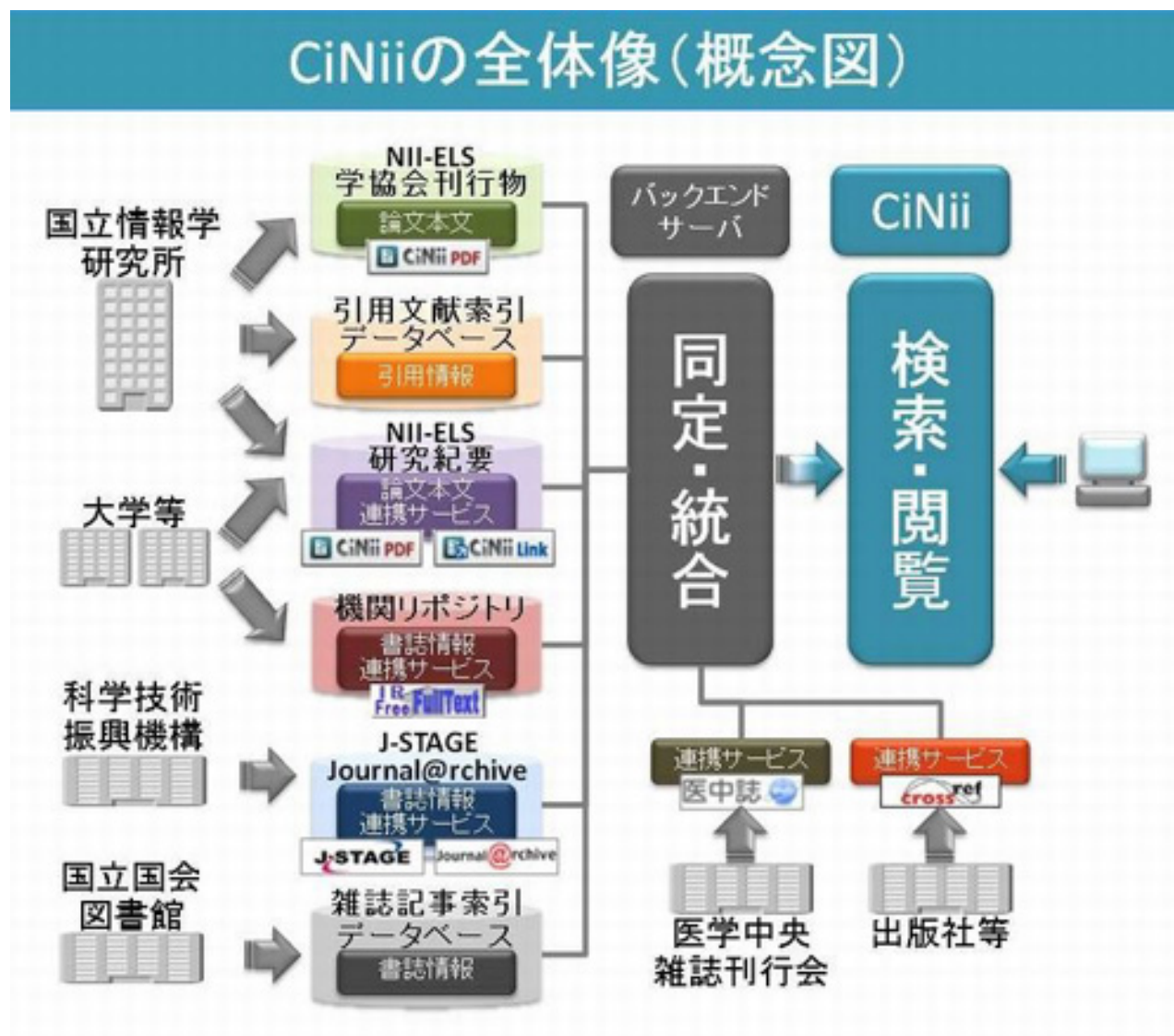
# ブール演算(Boolean Operation)

	式	真理値表(Truth Table)			ベン図(Venn Diagram)
		入力A	入力B	出力	
論理積 (AND)	A and B	0	0	0	
	$A \cdot B$	0	1	0	
	$A \times B$	1	0	0	
	$A \cap B$	1	1	1	
	$A \wedge B$				
論理和 (OR)	A or B	0	0	0	
	$A + B$	0	1	1	
	$A \cup B$	1	0	1	
	$A \vee B$	1	1	1	
否定論理積 (NAND)	A nand B	0	0	1	
	$\overline{A \cdot B}$	0	1	1	
	$\neg(A \wedge B)$	1	0	1	
		1	1	0	
否定論理和 (NOR)	A nor B	0	0	1	
	$\overline{A + B}$	0	1	0	
	$\neg(A \vee B)$	1	0	0	
		1	1	0	
排他的論理和 (EOR、XOR)	A eor B	0	0	0	
	A xor B	0	1	1	
	$\overline{A \cdot B} + A \cdot \overline{B}$	1	0	1	
	$(A \vee B) \wedge \neg(A \wedge B)$	1	1	0	
否定(NOT)	A not B	0		1	
	$\overline{A}$	1		0	
	$\neg A$				

# 文献DBの例

- 生命科学関連DBほとんど生データは英語
  - 2バイト文字、機種依存文字をきらうから。
  - 医学文献の場合:国や文化によって症状、薬が違う。
- 医中誌 (医学中央雑誌刊行会 <http://www.jamas.or.jp/>)
- J-STAGE (<http://www.jstage.jst.go.jp/browse/-char/ja>)
- CiNii (NII論文情報ナビゲータ[サイニイ], <http://ci.nii.ac.jp/>) --> 概念図参照
- PubMed (NCBI, <http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed>)
  - 英文綴り (イギリス式、オクスフォード式、アメリカ式)
  - MeSHについて --> 詳細に
- OMIM (NCBI, <http://www.ncbi.nlm.nih.gov/omim>)
  - human curated

# CiNiiの概念図



# PubMed

## 教則ビデオ

- 1~5分程度に細かく区切っている。
- きれいな英語
- お勧め機能: E-mail Alert
- MeSH本日説明

PubMed home

www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed

緑のgoo 環境goo goo Firefoxを使ってみよう Introduction to Bioinf... その他のブックマーク

NCBI Resources How To My NCBI Sign In

PubMed.gov  
U.S. National Library of Medicine  
National Institutes of Health

Search: PubMed Limits Advanced search Help

hemoglobin Search Clear

### PubMed

PubMed comprises more than 20 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

#### Using PubMed

- [PubMed Quick Start Guide](#)
- [Full Text Articles](#)
- [PubMed FAQs](#)
- [PubMed Tutorials](#)
- [New and Noteworthy](#)

#### PubMed Tools

- [Single Citation Matcher](#)
- [Batch Citation Matcher](#)
- [Clinical Queries](#)
- [Topic-Specific Queries](#)

#### More Resources

- [MeSH Database](#)
- [Journals Database](#)
- [Clinical Trials](#)
- [E-Utilities](#)
- [LinkOut](#)

You are here: NCBI > Literature > PubMed Write to the Help Desk

<b>GETTING STARTED</b>	<b>RESOURCES</b>	<b>POPULAR</b>	<b>FEATURED</b>	<b>NCBI INFORMATION</b>
<a href="#">NCBI Help Manual</a>	<a href="#">Literature</a>	<a href="#">PubMed</a>	<a href="#">GenBank</a>	<a href="#">About NCBI</a>
<a href="#">NCBI Handbook</a>	<a href="#">DNA &amp; RNA</a>	<a href="#">Nucleotide</a>	<a href="#">Reference Sequences</a>	<a href="#">Research at NCBI</a>
<a href="#">Training &amp; Tutorials</a>	<a href="#">Proteins</a>	<a href="#">BLAST</a>	<a href="#">Map Viewer</a>	<a href="#">NCBI Newsletter</a>
	<a href="#">Sequence Analysis</a>	<a href="#">PubMed Central</a>	<a href="#">Genome Projects</a>	<a href="#">NCBI FTP Site</a>

# スペルミスの場合

The screenshot shows a web browser window with the address bar at [www.ncbi.nlm.nih.gov/sites/entrez](http://www.ncbi.nlm.nih.gov/sites/entrez). The search bar contains the text "hemoglobine". Below the search bar, a suggestion box displays "Did you mean: [hemoglobin](#) (130842 items)". The search results section shows "Results: 1 to 20 of 48" and a list of results, with the first result being "[Impact of hemoglobin level on lung cancer survival]". The "Filter your results:" section on the right shows "All (48)", "Review (1)", and "Free Full Text (3)".

hemoglobine - PubMed res... x

www.ncbi.nlm.nih.gov/sites/entrez

緑のgoo 環境goo goo Firefox を使ってみよう Introduction to Bioinf... その他のブックマーク

NCBI Resources How To My NCBI Sign In

PubMed.gov  
U.S. National Library of Medicine  
National Institutes of Health

Search: PubMed  
hemoglobine Search Clear

Display Settings: Summary, 20 per page, Sorted by Recently Added Send to: Filter your results:

★ Did you mean: [hemoglobin](#) (130842 items)

Results: 1 to 20 of 48 << First < Prev Page 1 Next > Last >>

[\[Impact of hemoglobin level on lung cancer survival\]](#) Manage Filters

All (48)  
[Review \(1\)](#)  
[Free Full Text \(3\)](#)

# 英文綴字 の問題 と MeSH

MeSH term:  
Medical  
Subject  
Headings,  
いろいろな表  
現がある単  
語を統一して  
使えるように  
関連付けした  
用語集

The screenshot shows a web browser window displaying a PubMed search results page for the term 'hemoglobin'. The browser's address bar shows the URL: [www.ncbi.nlm.nih.gov/pubmed?term=hemoglobin&cmd=correctspelling](http://www.ncbi.nlm.nih.gov/pubmed?term=hemoglobin&cmd=correctspelling). The page header includes the NCBI logo and navigation links like 'Resources' and 'How To'. The search bar contains the text 'hemoglobin' and a 'Search' button. Below the search bar, there are options for 'Display Settings' (Summary, 20 per page, Sorted by Recently Added) and 'Send to'. The main content area displays 'Results: 1 to 20 of 130842' and lists several search results. The first result is 'Factors predicting clinically significant fatigue in women following treatment for primary breast cancer' by Gerber LH et al. The second result is 'Enhanced Production of L-Arginine by Expression of Vitreoscilla Hemoglobin Using a Novel Expression System in Corynebacterium crenatum' by Xu M et al. The third result is 'Revision total hip replacement: predictors of blood loss, transfusion requirements, and length of hospitalisation' by Mahadevan D et al. The fourth result is 'The Relationship between Lung Function and Metabolic Syndrome in Obese and Non-Obese Korean Adult Males' by Kim SK et al. On the right side of the page, there are filters for 'Filter your results' (All (130842), Review (7994), Free Full Text (23157)), 'Also try:', 'Titles with your search terms', and '9074 free full-text articles in PubMed Central'. At the bottom right, there is a 'Search details' section with a red circle around it, showing the search query: `"haemoglobin"[All Fields] OR "hemoglobins"[MeSH Terms] OR "hemoglobins"[All Fields] OR "hemoglobin"[All Fields]`. A 'Search' button is located below the search details.



# MeSH

The screenshot shows a web browser window displaying a PubMed search results page. The search term is 'hemoglobin', and the results are filtered by MeSH terms. A dropdown menu is open, showing a list of MeSH terms including EST, Structure, Genome, BioSystems, Books, CancerChromosomes, Conserved Domains, dbGaP, dbVar, 3D Domains, Epigenomics, Gene, Genome Project, GENSAT, GEO Profiles, GEO DataSets, HomoloGene, Journals, MeSH (highlighted), and NCBI Web Site. The search results list four items, each with a title, authors, journal information, and PMID. The right sidebar contains filters for 'Filter your results' (All (130842), Review (7994), Free Full Text (23157)), 'Also try:', 'Titles with your search terms', '9074 free full-text articles in PubMed Central', 'Find related data', and 'Search details'.

hemoglobin - PubMed resu... x

www.ncbi.nlm.nih.gov/pubmed?term=hemoglobin&cmd=correctspelling

緑のgoo 環境goo goo Firefox を使ってみよう Introduction to Bioinf... その他のブックマーク

NCBI Resources How To My NCBI Sign In

PubMed.gov  
U.S. National Library of Medicine  
National Institutes of Health

Search: MeSH  
hemoglobin

RSS Save search Limits Advanced search Help

Search Clear

Display Settings: Summary, 20 per page, 1

Send to: Filter your results:

All (130842)  
Review (7994)  
Free Full Text (23157)

Manage Filters

Also try:

Titles with your search terms

9074 free full-text articles in PubMed Central

Find related data

Database: Select

Find items

Search details

"haemoglobin"[All Fields] OR  
"hemoglobins"[MeSH Terms] OR  
"hemoglobins"[All Fields] OR  
"hemoglobin"[All Fields]

Search See more...

**Results: 1 to 20 of 130842**

[Factors predicting clinically significant primary breast cancer.](#)  
Gerber LH, Stout N, McGarvey C, Pfalzer LA.  
Support Care Cancer. 2010 Sep 12. [Epub ahead of print]  
PMID: 20835835 [PubMed - as supplied by publisher]  
[Related citations](#)

[Enhanced Production of L-Arginine Using a Novel Expression System in \*S. cerevisiae\*.](#)  
Xu M, Rao Z, Xu H, Lan C, Dou W, Zhang X, Xu H, Jiao J, Xu Z.  
Appl Biochem Biotechnol. 2010 Sep 11. [Epub ahead of print]  
PMID: 20835781 [PubMed - as supplied by publisher]  
[Related citations](#)

[Revision total hip replacement: predictors of blood loss, transfusion requirements, and length of hospitalisation.](#)  
Mahadevan D, Challand C, Keenan J.  
J Orthop Traumatol. 2010 Sep 11. [Epub ahead of print]  
PMID: 20835744 [PubMed - as supplied by publisher]  
[Related citations](#)

[The Relationship between Lung Function and Metabolic Syndrome in Obese and Non-Obese Korean Adult Males.](#)  
Kim SK, Hur KY, Choi YH, Kim SW, Chung JH, Kim HK, Lee MK, Min YK, Kim KW, Kim JH.  
Korean Diabetes J. 2010 Aug;34(4):253-60. Epub 2010 Aug 31.  
PMID: 20835343 [PubMed - in process]  
[Related citations](#)

# MeSH hemoglobin

関連するMeSH  
term: それぞれ  
の定義を比較、  
選択できる。

クリック

鎌形赤血球は別  
タームとなっ  
ていることに注意。

The screenshot shows the MeSH (Medical Subject Headings) website search results for the term 'hemoglobin'. The browser address bar shows 'www.ncbi.nlm.nih.gov/mesh?term=hemoglobin'. The page header includes the NCBI logo and 'MeSH' branding, along with the text 'A service of the National Library of Medicine and the National Institutes of Health'. A navigation bar at the top lists various databases: All Databases, PubMed, Nucleotide, Protein, Genome, Structure, OMIM, PMC, Journals, and Books. The search bar contains 'MeSH' and 'hemoglobin', with 'Go', 'Clear', and 'Save Search' buttons. Below the search bar are buttons for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. A large empty box is present below these buttons. The search results section shows 'All: 985' items, with 'Items 1 - 20 of 985' displayed. The first four results are listed with checkboxes and links:

- 1: [Hemoglobins](#) [Links](#)  
The oxygen-carrying proteins of ERYTHROCYTES. They are found in all vertebrates and some invertebrates. The number of globin subunits in the hemoglobin quaternary structure differs between species. Structures range from monomeric to a variety of multimeric arrangements.  
Year introduced: 1975
- 2: [Hemoglobin A, Glycosylated](#) [Links](#)  
Minor hemoglobin components of human erythrocytes designated A1a, A1b, and A1c. Hemoglobin A1c is most important since its sugar moiety is glucose covalently bound to the terminal amino acid of the beta chain. Since normal glycohemoglobin concentrations exclude marked blood glucose fluctuations over the preceding three to four weeks, the concentration of glycosylated hemoglobin A is a more reliable index of the blood sugar average over a long period of time.  
Year introduced: 1982
- 3: [Hemoglobin, Sickle](#) [Links](#)  
An abnormal hemoglobin resulting from the substitution of valine for glutamic acid at position 6 of the beta chain of the globin moiety. The heterozygous state results in sickle cell trait, the homozygous in sickle cell anemia.  
Year introduced: 1977(1975)
- 4: [Hemoglobin SC Disease](#) [Links](#)  
One of the sickle cell disorders characterized by the presence of both hemoglobin S and hemoglobin C. It is similar to, but less severe than sickle cell anemia

On the right side of the page, there is a 'Recent activity' section with a search icon and a list of recent searches:

- hemoglobin (985)
- hemoglobine (48)
- Haemoglobin (130842)
- Hemoglobin (130842)
- "Cold Temperature"[MeSH M... (21)

A yellow callout bubble points to the first result, 'Hemoglobins', with the text 'クリック' (Click).

# MeSH hemoglobins ページ 上部

Subheadingsを絞ることで、特定の意味で検索語が使われている文献に絞って検索できる。

単に検索語が含まれるだけの文献を排除できる。

The screenshot shows the MeSH website interface. The search bar contains 'MeSH' and the search results are displayed for 'Hemoglobins'. The 'Search Box with AND' dropdown menu is highlighted with a pink circle. The 'Subheadings' section is also highlighted with a pink circle, showing a list of subheadings with checkboxes. The 'Restrict Search to Major Topic headings only' checkbox is checked and highlighted with a pink circle. The 'Do Not Explode this term' checkbox is unchecked. The 'Entry Terms' section lists 'Hemoglobin', 'Eryhem', 'Ferrous Hemoglobin', and 'Hemoglobin, Ferrous'. The 'See Also' section is empty.

下のように入力すると、この状態で自動的にSearch PubMed窓が開くはず。ブラウザが対応していれば。ANDを使わずこの語のみで良いのであれば、そのまま[Search PubMed]

- If making selections (e.g., Subheadings, etc.), use the [Send to Search Box](#) feature to see PubMed records with those specifications.
- Select PubMed under the Links menu to retrieve all records for the MeSH Term.
- Select [NLM MeSH Browser](#) under the Links menu for additional information.

1: **Hemoglobins** Links

The oxygen-carrying proteins of ERYTHROCYTES. They are found in all vertebrates and some invertebrates. The number of globin subunits in the hemoglobin quaternary structure differs between species. Structures range from monomeric to a variety of multimeric arrangements.  
Year introduced: 1975

Subheadings: This list includes those paired at least once with this heading in MEDLINE and may not reflect current rules for allowable combinations.

abnormalities  administration and dosage  adverse effects  agonists  analogs and derivatives  analysis  antagonists and inhibitors  biosynthesis  blood  cerebrospinal fluid  chemical synthesis  chemistry  classification  deficiency  diagnostic use  drug effects  economics  genetics  history  immunology  isolation and purification  metabolism  pathology  pharmacokinetics  pharmacology  physiology  radiation effects  secretion  standards  supply and distribution  therapeutic use  toxicity  ultrastructure

Restrict Search to Major Topic headings only.

Do Not Explode this term (i.e., do not include MeSH terms found below this term in the MeSH tree).

Entry Terms:

- Hemoglobin
- Eryhem
- Ferrous Hemoglobin
- Hemoglobin, Ferrous

See Also:

参考のため、これより下の表示は次のスライドで。

# MeSH hemoglobins ページ 下部

ncbi.nlm.nih.gov/mesh

goo Firefox を使ってみよう Introduction to Bioinf...

その他のブックマーク

Do NOT explode this term (i.e., do not include MESH terms found below this term in the MESH tree).

## Entry Terms:

- Hemoglobin
- Eryhem
- Ferrous Hemoglobin
- Hemoglobin, Ferrous

## See Also:

- [Iron Chelating Agents](#)

[All MeSH Categories](#)

[Chemicals and Drugs Category](#)

[Amino Acids, Peptides, and Proteins](#)

[Proteins](#)

[Blood Proteins](#)

**Hemoglobins**

[Carboxyhemoglobin](#)

[Erythrocrucins](#)

[Fetal Hemoglobin](#)

[Hemoglobin A](#)

[Hemoglobin A, Glycosylated](#)

[Hemoglobin A2](#)

[Hemoglobin Subunits](#)

[alpha-Globins +](#)

[beta-Globins +](#)

[Hemoglobins, Abnormal](#)

[Hemoglobin C](#)

[Hemoglobin E](#)

[Hemoglobin H](#)

[Hemoglobin J](#)

[Hemoglobin M](#)

[Hemoglobin, Sickle](#)

[Methemoglobin](#)

[Oxyhemoglobins](#)

[Sulfhemoglobin](#)

[All MeSH Categories](#)

[Chemicals and Drugs Category](#)

[Amino Acids, Peptides, and Proteins](#)

[Proteins](#)

[Hemoproteins](#)

[Globins](#)

**Hemoglobins**

[Carboxyhemoglobin](#)

[Erythrocrucins](#)

MeSH term  
のカテゴリ  
調べたい概念  
により近い  
MeSH termが  
下部の階層に  
あるならば、な  
るべく下部階  
層の語を使う。

“cold”と検索窓に入力して[Go]

The screenshot shows a web browser window displaying the MeSH (Medical Subject Headings) website. The search bar at the top contains the text "cold". The search results page shows a single result for "Hemoglobins/classification" [Major]. The page includes a navigation menu with options like "All Databases", "PubMed", "Nucleotide", "Protein", "Genome", "Structure", "OMIM", "PMC", "Journals", and "Books". The search results are displayed in a list format, with the first result being "Hemoglobins". The page also includes a sidebar with various links and a footer with "Entry Terms" and "Restrict Search" options.

Search MeSH for cold

Search PubMed Clear

Display Full Show 20 Send to

All: 1

- If making selections (e.g., Subheadings, etc.), use the [Send to Search Box](#) feature to see PubMed records with those specifications.
- Select PubMed under the Links menu to retrieve all records for the MeSH Term.
- Select [NLM MeSH Browser](#) under the Links menu for additional information.

1: **Hemoglobins** [Links](#)

The oxygen-carrying proteins of ERYTHROCYTES. They are found in all vertebrates and some invertebrates. The number of globin subunits in the hemoglobin quaternary structure differs between species. Structures range from monomeric to a variety of multimeric arrangements.  
Year introduced: 1975

[Subheadings:](#) This list includes those paired at least once with this heading in MEDLINE and may not reflect current rules for allowable combinations.

abnormalities  administration and dosage  adverse effects  agonists  analogs and derivatives  analysis  antagonists and inhibitors  biosynthesis  blood  cerebrospinal fluid  chemical synthesis  chemistry  classification  deficiency  diagnostic use  drug effects  economics  genetics  history  immunology  isolation and purification  metabolism  pathology  pharmacokinetics  pharmacology  physiology  radiation effects  secretion  standards  supply and distribution  therapeutic use  toxicity  ultrastructure

Restrict Search to Major Topic headings only.  
 Do Not Explode this term (i.e., do not include MeSH terms found below this term in the MeSH tree).

Entry Terms:

- Hemoglobin
- Eryhem

# MeSH term vs. 自然語

「低温」や「風邪」といった自然語もMeSH検索で概念を絞ることができる。

1) checkいれてから

The screenshot shows the MeSH website interface. The search bar contains 'MeSH' and 'cold'. The search results are displayed in a list format. The first result, 'Cold Temperature', is checked. A dropdown menu is open over the search bar, showing options like 'Search Box with AND', 'Search Box with OR', and 'Search Box with NOT'. A yellow callout box points to the 'Search Box with AND' option.

Search: MeSH for cold

Display: Summary Show: 20

All: 51

Items 1 - 20 of 51

- 1: [Cold Temperature](#)  
An absence of warmth or heat or a temperature below the normal.  
Year introduced: 2009 (1963)
- 2: [Common Cold](#)  
A catarrhal disorder of the upper respiratory tract, which may be viral or a mixed infection. It generally involves a runny nose, nasal congestion, and sneezing.
- 3: [Pulmonary Disease, Chronic Obstructive](#)  
A disease of chronic diffuse irreversible airflow obstruction. Subcategories of COPD include CHRONIC BRONCHITIS and PULMONARY EMPHYSEMA.  
Year introduced: 2002
- 4: [Cold Ischemia](#)  
The chilling of a tissue or organ during decreased BLOOD perfusion or in the absence of blood supply. Cold ischemia time during ORGAN TRANSPLANTATION begins when the organ is cooled with a cold perfusion solution after ORGAN PROCUREMENT surgery, and ends after the tissue reaches physiological temperature during implantation procedures. WARM ISCHEMIA TIME starts then and ends with completion of SURGICAL ANASTOMOSIS.  
Year introduced: 2006
- 5: [Cold Climate](#)  
A climate characterized by COLD TEMPERATURE for a majority of the time during the year.  
Year introduced: POLAR REGIONS was see under COLD CLIMATE 1963-1977
- 6: [Extreme Cold](#)  
Below normal weather temperatures that may lead to serious health problems. Extreme cold is a dangerous situation that can bring on health emergencies in susceptible people.  
Year introduced: 2009
- 7: [Cryopyrin-associated Periodic Syndromes](#)

2) Search Box with ANDを選択

# 再び別タームを同様に選択

cold - MeSH Results

www.ncbi.nlm.nih.gov/mesh

NCBI MeSH

A service of the National Library of Medicine and the National Institutes of Health

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search MeSH for cold [Go] [Clear] [Save Search]

Limits Preview/Index History Clipboard Details

"Hemoglobins/classification"[Majr] AND "Cold Temperature"[Mesh]

Search PubMed [Clear]

Display Summary Show 20 Search Box with OR

All: 51

Items 1 - 20 of 51

- 1: [Cold Temperature](#)  
An absence of warmth or heat or a temper... norm.  
Year introduced: 2009 (1963)
- 2: [Common Cold](#)  
A catarrhal disorder of the upper respiratory tract, which may be viral or a mixed infection. It generally involves a runny nose, nasal congestion, and sneezing.
- 3: [Pulmonary Disease, Chronic Obstructive](#)  
A disease of chronic diffuse irreversible airflow obstruction. Subcategories of COPD include CHRONIC BRONCHITIS and PULMONARY EMPHYSEMA.  
Year introduced: 2002
- 4: [Cold Ischemia](#)  
The chilling of a tissue or organ during decreased BLOOD perfusion or in the absence of blood supply. Cold ischemia time during ORGAN TRANSPLANTATION begins when the organ is cooled with a cold perfusion solution after ORGAN PROCUREMENT surgery, and ends after the tissue reaches physiological temperature during implantation procedures. WARM ISCHEMIA TIME starts then and ends with completion of SURGICAL ANASTOMOSIS.  
Year introduced: 2006
- 5: [Cold Climate](#)  
A climate characterized by COLD TEMPERATURE for a majority of the time during the year.  
Year introduced: POLAR REGIONS was see under COLD CLIMATE 1963-1977
- 6: [Extreme Cold](#)  
Below normal weather temperatures that may lead to serious health problems. Extreme cold is a dangerous situation that can bring on health emergencies in susceptible people.  
Year introduced: 2009
- 7: [Cryopyrin-associated Periodic Syndromes](#)

of 3 Next

Recent activity

- Turn Off Clear
- Q "Hemoglobins/classificati... PubMed (1)
- Q cold (51)
- Q Cold Climate
- Q "Cold Climate" (1) MeSH
- The hemoglobins of fishes living at polar latitudes - current
- » See more...

checkいれてから

# 検索窓を確認

括弧の位置が意図したもの(Cold Temp OR Cold Climate)と違う  
-->手入力で修正

注意:  
MeSH termを検索し、  
検索語を順次追加-->  
最後にPubMed検索  
すること。  
逆に、PubMed検索し  
た後にMeSH検索して  
も、両者の検索語の  
結合は自動的にはで  
きない。

The screenshot shows a web browser window displaying the MeSH (Medical Subject Headings) search results for the term "cold". The browser's address bar shows "www.ncbi.nlm.nih.gov/mesh". The page header includes the NCBI logo and the text "A service of the National Library of Medicine and the National Institutes of Health". The search bar contains "MeSH" and "cold". The search results are displayed in a list format, showing the first 20 items of 51 total results. The first seven items are listed below:

- 1: [Cold Temperature](#) (Links)  
An absence of warmth or heat or a temperature notably below an accustomed norm.  
Year introduced: 2009 (1963)
- 2: [Common Cold](#) (Links)  
A catarrhal disorder of the upper respiratory tract, which may be viral or a mixed infection. It generally involves a runny nose, nasal congestion, and sneezing.
- 3: [Pulmonary Disease, Chronic Obstructive](#) (Links)  
A disease of chronic diffuse irreversible airflow obstruction. Subcategories of COPD include CHRONIC BRONCHITIS and PULMONARY EMPHYSEMA.  
Year introduced: 2002
- 4: [Cold Ischemia](#) (Links)  
The chilling of a tissue or organ during decreased BLOOD perfusion or in the absence of blood supply. Cold ischemia time during ORGAN TRANSPLANTATION begins when the organ is cooled with a cold perfusion solution after ORGAN PROCUREMENT surgery, and ends after the tissue reaches physiological temperature during implantation procedures. WARM ISCHEMIA TIME starts then and ends with completion of SURGICAL ANASTOMOSIS.  
Year introduced: 2006
- 5: [Cold Climate](#) (Links)  
A climate characterized by COLD TEMPERATURE for a majority of the time during the year.  
Year introduced: POLAR REGIONS was see under COLD CLIMATE 1963-1977
- 6: [Extreme Cold](#) (Links)  
Below normal weather temperatures that may lead to serious health problems. Extreme cold is a dangerous situation that can bring on health emergencies in susceptible people.  
Year introduced: 2009
- 7: [Cryopyrin-associated Periodic Syndromes](#) (Links)



# 検索結果 1件の場合

The hemoglobins of fishes... x

www.ncbi.nlm.nih.gov/pubmed?term="Hemoglobins/classification"[Majr]+AND+("Cold Climate"[Mesh]+OR+"Cold Temperature"[Mesh])

NCBI Resources How To My NCBI Sign In

PubMed.gov  
U.S. National Library of Medicine  
National Institutes of Health

Search: PubMed RSS Save search Limits Advanced search Help

"Hemoglobins/classification"[Majr] AND ("Cold Climate"[Mesh] OR "Cold Temperature"[Mesh]) Search Clear

Display Settings: Abstract Send to: BenthamDirect

Curr Protein Pept Sci. 2008 Dec;9(6):578-90.

**The hemoglobins of fishes living at polar latitudes - current knowledge on structural adaptations in a changing environment.**

Verde C, Vergara A, Mazzarella L, di Prisco G.  
Institute of Protein Biochemistry, Naples, Italy. c.verde@ibp.cnr.it

**Abstract**  
Fishes thriving in polar habitats offer many opportunities for comparative approaches to understanding protein adaptations to temperature. Notothenioidae, the dominant suborder in the Antarctic Ocean, have evolved reduction of hemoglobin concentration and multiplicity, perhaps as a consequence of temperature stability and other environmental parameters. In the icefish family, the blood pigment is absent. In contrast, similar to other acanthomorph teleosts, Arctic fish, thriving in a more complex oceanographic system, have maintained higher hemoglobin multiplicity and a highly diversified globin system in response to environmental variability and/or variations in metabolic demands. This review summarises the current knowledge on the structure, function and phylogeny of hemoglobins of fish living in polar habitats. On the basis of crystallographic analysis, a novel guideline to the interpretation of the Root effect in terms of a three-state model is suggested, implying the accessibility of an R/T intermediate quaternary structure, frequently observed in Antarctic fish hemoglobins. The occurrence of bis-histidyl and penta-coordinate states in ferric forms of polar fish hemoglobins suggests additional redox properties.

PMID: 19075748 [PubMed - indexed for MEDLINE]

Publication Types, MeSH Terms, Substances

**Publication Types:**  
Research Support, Non-U.S. Gov't  
Review

**MeSH Terms:**  
Adaptation, Physiological  
Animals  
Arctic Regions  
Cold Climate  
Evolution, Molecular  
Fish Proteins/chemistry\*  
Fish Proteins/classification\*  
Fish Proteins/physiology  
Fishes/physiology\*  
Hemoglobins/chemistry\*  
Hemoglobins/classification\*  
Hemoglobins/physiology

**Related citations**

**Cited by 1 PubMed Central article**  
Correlation between hemichrome stability and the root effect in tetrameric hen [Biophys J. 2009]

**All links from this record**  
Related Citations  
Substance (MeSH Keyword)  
Cited in PMC

**Search details**  
"Hemoglobins/classification"[Majr] AND ("Cold Climate"[Mesh] OR "Cold Temperature"[Mesh])  
Search See more...

**Recent activity**

どんなMeSH term  
で検索されるか

修正した検索式

# OMIM: 古くからあるhuman curated DB

OMIM Home  
www.ncbi.nlm.nih.gov/omim



**OMIM**  
Online Mendelian Inheritance in Man



Johns  
Hopkins  
University

My NCBI  
[\[Sign In\]](#) [\[Register\]](#)

All Databases PubMed Nucleotide Protein Genome Structure PMC OMIM

Search OMIM for

[Limits](#) [Preview/Index](#) [History](#) [Clipboard](#) [Details](#)

Entrez

## OMIM

Search OMIM  
Search Gene Map  
Search Morbid Map

## Help

OMIM Help  
How to Link

## FAQ

Numbering System  
Symbols  
How to Print  
Citing OMIM  
Download

## OMIM Facts

Statistics  
Update Log  
Restrictions on Use

## Allied Resources

Genetic Alliance  
Databases  
HGMD  
Locus-Specific  
Model Organisms  
MitoMap  
Phenotype  
Human/Mouse/Rat  
Homology Maps  
Coriell  
The Jackson

- Enter one or more search terms.
- Use **Limits** to restrict your search by search field, chromosome, and other criteria.
- Use **Index** to browse terms found in OMIM records.
- Use **History** to retrieve records from previous searches, or to combine searches.

## OMIM® - Online Mendelian Inheritance in Man

Welcome to OMIM®, Online Mendelian Inheritance in Man®. OMIM is a comprehensive, authoritative, and timely compendium of human genes and genetic phenotypes. The full-text, referenced overviews in OMIM contain information on all known mendelian disorders and over 12,000 genes. OMIM focuses on the relationship between phenotype and genotype. It is updated daily, and the entries contain copious links to other genetics resources.

This database was initiated in the early 1960s by Dr. Victor A. McKusick as a catalog of mendelian traits and disorders, entitled Mendelian Inheritance in Man (MIM). Twelve book editions of MIM were published between 1966 and 1998. The online version, OMIM, was created in 1985 by a collaboration between the National Library of Medicine and the William H. Welch Medical Library at Johns Hopkins. It was made generally available on the internet starting in 1987. In 1995, OMIM was developed for the World Wide Web by NCBI, the National Center for Biotechnology Information.

OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada Hamosh.

NLM's Profiles in Science -- The McKusick Papers [More...](#)

NOTE: OMIM is intended for use primarily by physicians and other professionals concerned with genetic disorders, by genetics researchers, and by advanced students in science and medicine. While the OMIM database is open to the public, users seeking information about a personal medical or genetic condition are urged to consult with a qualified physician for diagnosis and for answers to personal questions.

OMIM® and Online Mendelian Inheritance in Man® are registered trademarks of the Johns Hopkins University.

# 核酸DBと遺伝子アノテーション

- 国際塩基配列データベース (INSDC)
- データ構造
- データのカテゴリ
- INSDC核酸DBでのID命名法
- 遺伝子アノテーション
- 遺伝子推定法とID命名法

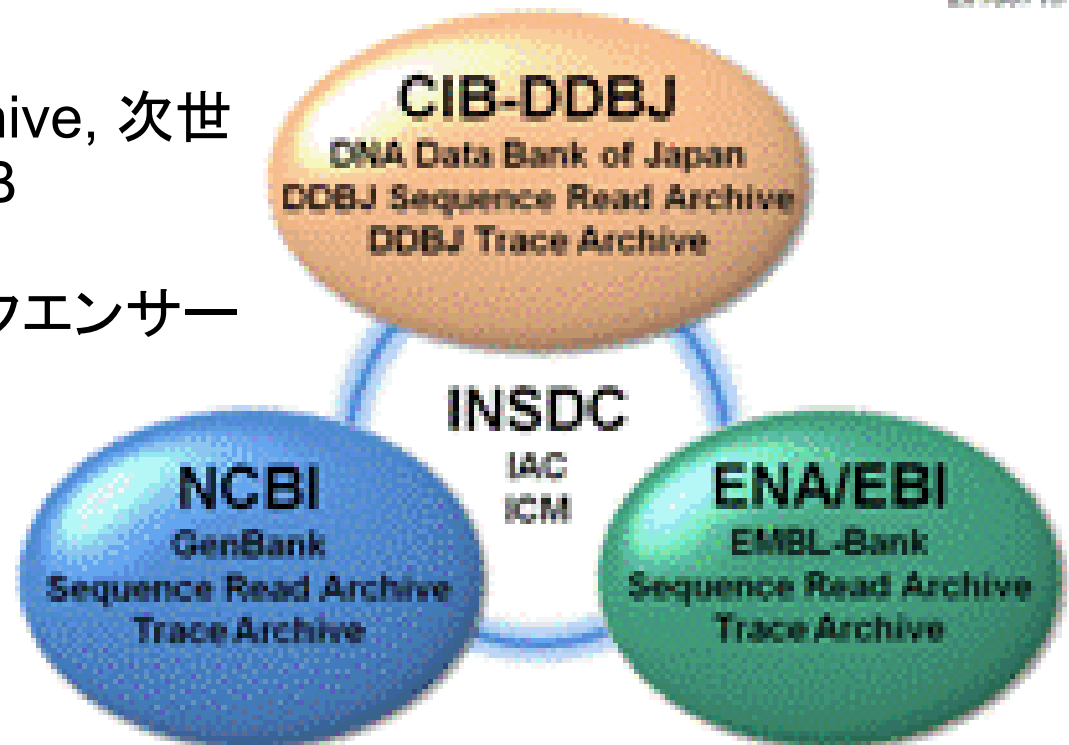
# •国際塩基配列データベース(INSDC)

## •3センターの違い:

### •データアクセス(共通)とデータ登録方法(別々)

### •図:それぞれの機関3行になっている意味

- 1行目:従来からの核酸DB
- 2行目:SRA (Sequence Read Archive, 次世代シーケンサーのreadの保存用DB
  - DDBJ SRA(DRA)
- 3行目:Trace Archive, 従来型シーケンサーのreadの保存用DB
  - DDBJ Trace Archive (DTA)



# データ構造 核酸DB フラット・ファイル

```
LOCUS       AK307560                1114 bp    mRNA    linear    HTC 12-JAN-2008
DEFINITION  Homo sapiens cDNA, FLJ97508.
ACCESSION   AK307560
VERSION     AK307560.1  GI:164692527
KEYWORDS    HTC; HTC_FLI; oligo capping.
SOURCE      Homo sapiens (human)
            ORGANISM      Homo sapiens
                        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                        Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
                        Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Wakamatsu,A., Yamamoto,J., Kimura,K., Ishii,S., Watanabe,K.,
            Sugiyama,A., Murakawa,K., Kaida,T., Tsuchiya,K., Fukuzumi,Y.,
            Kumagai,A., Oishi,Y., Yamamoto,S., Ono,Y., Komori,Y., Yamazaki,M.,
            Kisu,Y., Nishikawa,T., Sugano,S., Nomura,N. and Isogai,T.
TITLE       NEDO human cDNA sequencing project
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1114)
AUTHORS     Isogai,T. and Yamamoto,J
TITLE       Direct Submission
JOURNAL     Submitted (11-JAN-2008) Contact:Takao Isogai Reverse Proteomics
            Research Institute; 1-9-11 Kaji-cho, Chiyoda-ku, Tokyo 101-0044,
            Japan E-mail : flj-cdna@nifty.com
COMMENT     Human cDNA sequencing project focused on splicing variants of mRNA
            in NEDO functional analysis of protein and research application
            project supported by Ministry of Economy, Trade and Industry,
            Japan; cDNA selection for complete cds sequencing: Reverse
            Proteomics Research Institute (REPRORI), Hitachi, Ltd., Japan
            (Hitachi) and Japan Biological Informatics Consortium, Japan
            (JBIC); cDNA complete cds sequencing: JBIC; cDNA library
            construction: Helix Research Institute supported by Japan Key
            Technology Center, Japan (HRI); cDNA 5'- & 3'-end sequencing:
            Research Association for Biotechnology, Japan, Biotechnology
            Center, National Institute of Technology and Evaluation, Japan and
            HRI; cDNA mapping to human genome: Central Research Laboratory,
            Hitachi; evaluation and annotation: REPRORI.
FEATURES             Location/Qualifiers
     source            1..1114
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="NETRP2000337"
                        /cell_type="neutrophils"
                        /clone_lib="NETRP2"
                        /note="cloning vector: pME18SFL3;
                        primary culture, neutrophils"
ORIGIN
1 agtgtcgacg gcagcggcgg cggcgggtgg gaaatggcgg agtatctggc ctccatcttc
```

フィールド

フラットファイルの  
形式はEMBLだけ  
少し異なる。

識別子

内容

識別子はDB内で統一  
-->キーとして重要  
核酸DBの主キーは  
アクセッション

# フラット・ファイル の説明・ Locus field

LOCUS	AK307560	1114 bp	mRNA	linear	HTC 12-JAN-2008
DEFINITION	Homo sapiens cDNA, FLJ97508.				
ACCESSION	AK307560				
VERSION	AK307560.1 GI:164692527				
KEYWORDS	HTC; HTC_FLI; oligo capping.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				

Locus 名

配列長

分子タイプ

Division

分子形態

最終公開日

- Locus名:
  - 各DB内でユニークな名前、DBごとに独自の命名法をとる。DDBJでは1996年7月以降アクセッション番号と同一に。
- 分子タイプ:
  - DNA, RNA, mRNA, rRNA, tRNA, cRNA のいずれか。
  - DDBJで登録する場合、source feature の /mol\_type qualifier に記載する。
- 分子形態:
  - 線状 (linear) か、環状 (circular) か。ただし、circularはエントリが完全長の場合。
- division:
  - 配列登録者が選ぶものではない。
  - 生物種に基づくdivision
  - データ形式に基づくdivision
  - データ形式として良く使われる略語がdivisionではない場合もあるので注意
    - WGS(whole genome shotgun),
    - TPA(Third Party Annotation),
    - MGA(Mass sequence for Genome Annotation)
    - SRA (Sequence Read Archive, 次世代シーケンサーのreadの保存用DB
    - Trace Archive, 従来型シーケンサーのreadの保存用DB
  - これらの略語は相同性検索の際、ターゲットを絞るために知っておくとよい。

# Division

\*

HUM	ヒト
PRI	霊長類 (ヒトを除く)
ROD	齧歯類
MAM	哺乳類 (上記を除く)
VRT	脊椎動物 (MAMを除く)
INV	無脊椎動物
PLN	植物・真菌類 など
BCT	細菌
VRL	ウイルス
PHG	バクテリオファージ

\* GenBankでは無い。

BLASTなど相同性検索の際の参考に。

PAT	特許出願に含まれる塩基配列データ
ENV	環境上のサンプルに由来した配列
SYN	synthetic constructs 人為的に構成された配列
EST	expressed sequence tags
* TSA	transcriptome shotgun assemblies, すでに登録された転写物配列をassembleして得られた長い配列データ
GSS	genome survey sequences short single pass のゲノム配列
HTC	high throughput cDNA sequences, 配列が finish した後, 生物種による division に移される場合あり。
HTG	high throughput genomic sequences, ゲノムプロジェクトに由来, 以下の 3 phase に分類されKEYWORD 行に記載。finish した後は生物種による division に移る。
	phase0: piece contig が構築される以前の配列
	phase1: 構築された piece contig の向きや順序が未確定の配列
	phase2: piece contig の向きや順序が確定した unfinished の配列
STS	sequence tagged sites
* CON	Contig / Constructed, ゲノムプロジェクトのように個々に登録された一連の配列データをデータバンク側が結合し, 1つのアクセッション番号を付与した長大なデータ。

# INSDC核酸DBでのID

- INSDC核酸DB
  - 通常division
    - アルファベット1文字 + 5桁数字または、アルファベット2文字 + 6桁数字
    - アルファベット部分はDBおよびdivisionによって割り振りが決まっている。
    - 参照: <http://www.ddbj.nig.ac.jp/sub/prefix.htm>
  - WGS data
    - アルファベット4文字 + 8桁数字
  - MGA data
    - アルファベット5文字 + 7桁数字



# 遺伝子アノテーション

## •配列決定後、配列に対して有用な記述を追加する作業

### •個々の塩基配列データでは通常、配列決定者が最低限のアノテーションをする。

#### • 例) 配列全体に対して

- 名称、遺伝子シンボル、定義など

#### • 例) 配列中の部分領域に対して

- [ゲノム配列に対し]遺伝子、反復配列など
- [遺伝子配列に対し]遺伝子構造、多型など
- [アミノ酸配列に対し]ドメイン、活性部位、修飾部位など

### •ゲノム配列アノテーション・3段階

#### • 反復配列探索

- 前処理として反復配列をマスクする (eg., RepeatMasker)

#### • 遺伝子探索(cf., 次のスライド)

- 計算機処理によって遺伝子(候補)領域を抽出する

#### • 機能探索

- 機能に関係することが知られている配列(タンパク質機能ドメイン)を抽出する

登録の際は、自分が何に対してどういうアノテーションが求められているか把握することが大切。  
年々マニュアルが更新されているので注意。

# 転写物のアノテーション

- cDNAおよび成熟mRNAは、ゲノム配列とマッピングさせることによって、そのまま遺伝子アノテーションになる。
- cDNAのおもな種類
  - Expressed Sequence Tags (ESTs)
    - cDNAクローンをランダムに選び末端400-600bp程度をone-passで配列決定
    - 比較的不正確(約2%のエラー)な傾向
    - 網羅的に発現位置を決める目的で使用
    - 核酸DBにESTというdivisionがある
  - Full-length cDNA(完全長cDNA)
    - 配列の正確性向上。
    - ASパターンも予測可能。
    - 長い遺伝子ではデータが少なくなる傾向

H-InvDB	完全長cDNA配列にアノテーションをつけることを出発点とし、アミノ酸配列、AS、発現などの関連情報と連動したDB
RefSeq	非冗長的に、遺伝子、転写物、タンパク質、それぞれの代表配列をモデル生物で決め、整備したDB
Ensembl	ゲノムデータの自動アノテーションとブラウザ表示を出発点とし、アミノ酸配列、AS、発現などの関連情報と連動したDB

# primary配列DBとannotated DB

<u>GenBank</u>	<u>RefSeq</u>
Not curated	Curated
Author submits	NCBI creates from existing data
Only author can revise	NCBI revises as new data emerge
Multiple records for same loci common	Single records for each molecule of major organisms
Records can contradict each other	
No limit to species included	Limited to model organisms
Data exchanged among INSDC members	Exclusive NCBI database
Akin to <b>primary literature</b>	Akin to <b>review articles</b>
Proteins identified and linked	Proteins and transcripts identified and linked
Access via NCBI Nucleotide databases	Access via Nucleotide & Protein databases

# 遺伝子探索のおもな戦略

- 転写産物比較
  - 最も信頼度が高く、ゲノム配列と同時に転写物配列も網羅的に読むプロジェクトも多い
- 比較ゲノム
  - 他の生物ゲノムと比較し、保存性の高い領域を遺伝子候補とする
- *ab initio*法
  - 既知の知見によって遺伝子の特徴を備えた配列を遺伝子候補とする
  - 信頼性は最も低い
  - “*ab initio*”とは「始めから」の意。ゲノムのアッセンブル法でもこの語がつかわれるので混乱注意。

# RefSeq、H-InvDBにおけるIDの付け方

## •RefSeqのID

- アルファベット2文字+アンダーバー+数字
- アルファベット1文字目:
  - A: **A**lternate assembly or annotation
  - N: k**N**own (A以外のゲノム配列とキュレータによりレビューされたRNA&protein)
  - X: prediction(predi**X**on)
  - Z: NZ\_accession(WGS)上についたproteinのアノテーション
- アルファベット2文字目:
  - M: **m**RNA
  - R: noncoding **R**NA
  - P: **p**rotein
  - Z: genomic, whole genome shotgun (WGS) sequence data
  - W: genomic, Intermediate assemblies of BAC or WGS sequence data
  - T: genomic, Intermediate assemblies of BAC and/or WGS sequence data
  - S: genomic, unplaced scaffolds, etc
  - G: genomic, incomplete
  - C: genomic, complete

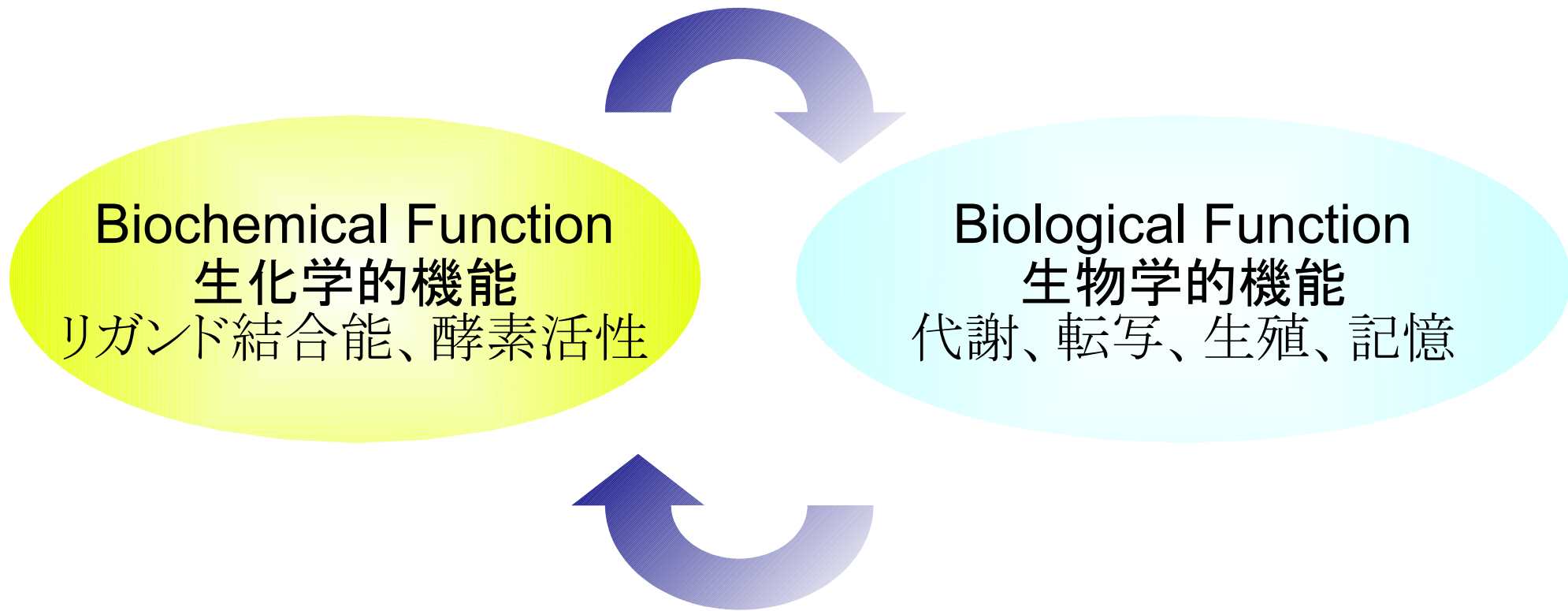
## •H-InvDBのID

- HIT (H-Invitational transcript): HIT + 9桁の数字 + version番号 例)HIT000000001.1
- HIX (H-Invitational cluster): HIX + 7桁の数字 + version番号 例)HIX0000001.1
- HIP (H-Invitational protein): HIP + 9桁の数字 + version番号 例)HIP000000001.1
- HIF (H-Invitational gene family/group): HIF + 7桁の数字 例)HIF0000001

# タンパク質データベース と機能推定

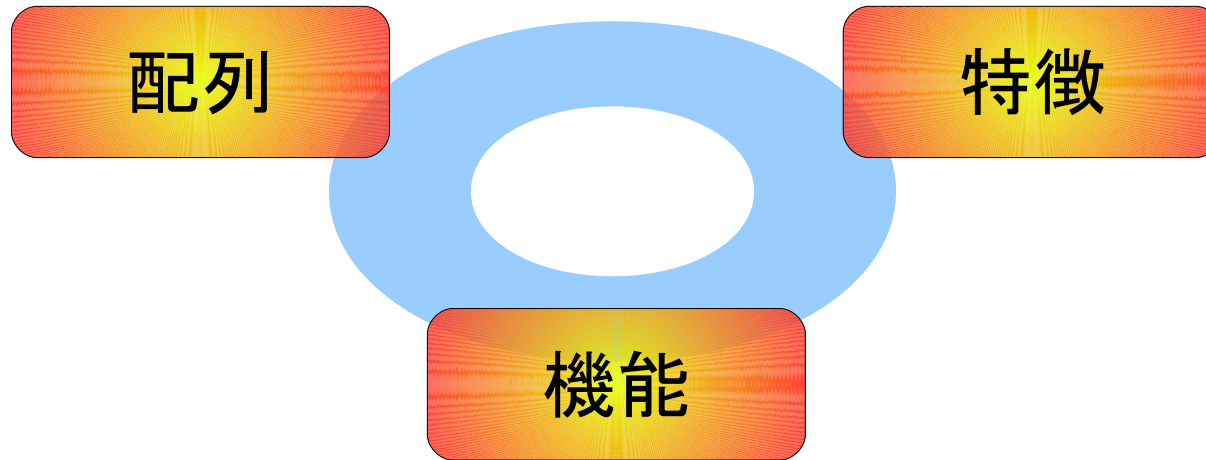
- 「機能」と配列について
- 機能ドメインデータベース
  - UniProt
  - InterPro
- 立体構造データベース
  - PDB

# 「機能」という語の多義性



作業のscopeは何かが重要

# 機能と配列と特徴



- 配列に隠された機能の単位(ドメイン)。
- 機能ドメインは進化的に保存されやすい。
- 保存領域(Sequence signature)の長さ・規模は様々
  - 翻訳後修飾部位、モチーフ、ドメイン
- 配列上の特徴が機能ドメインゆえであることが多い。
- 決まった立体構造を通じて機能する



# UniProt

UniProt - Google 検索 UniProt

www.uniprot.org

HotMail の無料サービス Windows Media Windows M リンクの変更 その他のブックマーク

UniProt Downloads · Contact · Documentation/Help

Search Blast Align Retrieve ID Mapping

Search in Protein Knowledgebase (UniProtKB) Query Search Clear Fields »

## WELCOME

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

### What we provide

UniProtKB	Protein knowledgebase, consists of two sections: <ul style="list-style-type: none"><li>★ Swiss-Prot, which is manually annotated and reviewed.</li><li>★ TrEMBL, which is automatically annotated and is <b>not</b> reviewed.</li></ul> Includes <a href="#">Complete Proteome Sets</a> .
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	<a href="#">Literature citations</a> , <a href="#">taxonomy</a> , <a href="#">keywords</a> and <a href="#">more</a> .

### Getting started

- [Text search](#)
- [Sequence similarity searches \(BLAST\)](#)
- [Sequence alignments](#)
- [Batch retrieval](#)

## NEWS

### UniProt release 2010\_09 – Aug 10, 2010

‘De-merge’ of multi-gene entries derived from a single species in UniProtKB/Swiss-Prot

- > [Statistics for UniProtKB: Swiss-Prot · TrEMBL](#)
- > [Forthcoming changes](#)
- > [News archives](#)

## SITE TOUR



Learn how to make best use of the tools and data on this site.

## PROTEIN SPOTLIGHT

# “U2AF”で検索

UniProtKB Search Results for U2AF

Search in: Protein Knowledgebase (UniProtKB) | Query: U2AF | Search | Clear | Fields >

1 - 25 of 629 results for U2AF in UniProtKB sorted by score descending

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50% | Download

Page 1 of 26 | Next >

Results

> Show only reviewed (366) (UniProtKB/Swiss-Prot) or unreviewed (263) (UniProtKB/TrEMBL) entries  
> Restrict term "u2af" to gene name (1), protein name (142)

Accession	Entry name	Status	Protein names	Gene names	Organism	Length
<input type="checkbox"/> P26368	U2AF2_HUMAN	★	Splicing factor U2AF 65 kDa subunit	U2AF2 U2AF65	Homo sapiens (Human)	475
<input type="checkbox"/> Q01081	U2AF1_HUMAN	★	Splicing factor U2AF 35 kDa subunit	U2AF1 U2AF35 U2AFBP	Homo sapiens (Human)	240
<input type="checkbox"/> P36629	U2AF2_SCHPO	★	Splicing factor U2AF 59 kDa subunit	prp2 mis11 SPBC146.07	Schizosaccharomyces pombe (Fission yeast)	517
<input type="checkbox"/> Q8WU68	U2AF4_HUMAN	★	Splicing factor U2AF 26 kDa subunit	U2AF1L4 U2AF1-RS3 U2AF1L3	Homo sapiens (Human)	220
<input type="checkbox"/> Q09176	U2AF1_SCHPO	★	Splicing factor U2AF 23 kDa subunit	SPAP8A3.06	Schizosaccharomyces pombe (Fission yeast)	216
<input type="checkbox"/> Q8BGJ9	U2AF4_MOUSE	★	Splicing factor U2AF 26 kDa subunit	U2af1I4 U2af26	Mus musculus (Mouse)	220
<input type="checkbox"/> P26369	U2AF2_MOUSE	★	Splicing factor U2AF 65 kDa subunit	U2af2 U2af65	Mus musculus (Mouse)	475
<input type="checkbox"/> Q24562	U2AF2_DROME	★	Splicing factor U2AF 50 kDa subunit	U2af50 CG9998	Drosophila melanogaster (Fruit fly)	416
<input type="checkbox"/> P90978	U2AF2_CAEEL	★	Splicing factor U2AF 65 kDa subunit	uaf-1 Y92C3B.2	Caenorhabditis elegans	496
<input type="checkbox"/> Q94535	U2AF1_DROME	★	Splicing factor U2af 38 kDa subunit	U2af38 CG3582	Drosophila melanogaster (Fruit fly)	264

デフォルトでは妥当性の高い順

# Gene(1)で絞って表示

UniProtKB

Search Blast Align Retrieve ID Mapping \*

Search in **Query**  
Protein Knowledgebase (UniProtKB)    [Fields >](#)

1 - 25 of 71 results for **gene:u2af** in UniProtKB sorted by **score** descending

[Browse by taxonomy, keyword, gene ontology, enzyme class or pathway](#) | [Reduce sequence redundancy to 100%, 90% or 50%](#) |

Customize display Page 1 of 3 | [Next >](#)

### Results

> Show only [reviewed \(30\)](#) ★ (UniProtKB/Swiss-Prot) or [unreviewed \(41\)](#) ☆ (UniProtKB/TrEMBL) entries  
> Show only exact matches for **u2af**

<input type="checkbox"/>	Accession	Entry name	Status	Protein names	Gene names	Organism	Length
<input type="checkbox"/>	<a href="#">Q01081</a>	U2AF1_HUMAN	★	Splicing factor U2AF 35 kDa subunit	U2AF1 U2AF35 U2AFBP	Homo sapiens (Human)	240
<input type="checkbox"/>	<a href="#">P26368</a>	U2AF2_HUMAN	★	Splicing factor U2AF 65 kDa subunit	U2AF2 U2AF65	Homo sapiens (Human)	475
<input type="checkbox"/>	<a href="#">P26369</a>	U2AF2_MOUSE	★	Splicing factor U2AF 65 kDa subunit	U2af2 U2af65	Mus musculus (Mouse)	475
<input type="checkbox"/>	<a href="#">Q8BGJ9</a>	U2AF4_MOUSE	★	Splicing factor U2AF 26 kDa subunit	U2af14 U2af26	Mus musculus (Mouse)	220
<input type="checkbox"/>	<a href="#">Q8WU68</a>	U2AF4_HUMAN	★	Splicing factor U2AF 26 kDa subunit	U2AF1L4 U2AF1-RS3 U2AF1L3	Homo sapiens (Human)	220
<input type="checkbox"/>	<a href="#">Q15696</a>	U2AFM_HUMAN	★	U2 small nuclear ribonucleoprotein auxiliary	ZRSR2 U2AF1-RS2 U2AF1L2 U2AF1RS2	Homo sapiens (Human)	482
<input type="checkbox"/>	<a href="#">Q94535</a>	U2AF1_DROME	★	Splicing factor U2af 38 kDa subunit	U2af38 CG3582	Drosophila melanogaster (Fruit fly)	264
<input type="checkbox"/>	<a href="#">Q9D883</a>	U2AF1_MOUSE	★	Splicing factor U2AF 35 kDa subunit	U2af1	Mus musculus (Mouse)	239
<input type="checkbox"/>	<a href="#">Q24562</a>	U2AF2_DROME	★	Splicing factor U2AF 50 kDa subunit	U2af50 CG9998	Drosophila melanogaster (Fruit fly)	416
<input type="checkbox"/>	<a href="#">Q15695</a>	U2AF1_HUMAN	★	U2 small nuclear	ZRSR1 U2AF1-RS1 U2AF1L1	Homo sapiens (Human)	475

# “taxonomy:9606”を検索窓に追加

UniProt > UniProtKB Downloads · Contact · Documentation/Help

Search Blast Align Retrieve ID Mapping \*

Search in Protein Knowledgebase (UniProtKB) Query gene:u2af taxonomy:9606 Search Clear Fields >>

11 results for **gene:u2af** AND **taxonomy:“Homo sapiens (Human) [9606]”** in UniProtKB sorted by **score** descending

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50% | [Download](#)

Customize display Page 1 of 1

## Results

> Show only reviewed (6) ★ (UniProtKB/Swiss-Prot) or unreviewed (5) ☆ (UniProtKB/TrEMBL) entries  
> Show only entries from a complete proteome set (6)

Accession	Entry name	Status	Protein names	Gene names	Organism	Length
<input type="checkbox"/> <a href="#">Q01081</a>	U2AF1_HUMAN	★	Splicing factor U2AF 35 kDa subunit	U2AF1 U2AF35 U2AFBP	Homo sapiens (Human)	240
<input type="checkbox"/> <a href="#">P26368</a>	U2AF2_HUMAN	★	Splicing factor U2AF 65 kDa subunit	U2AF2 U2AF65	Homo sapiens (Human)	475
<input type="checkbox"/> <a href="#">Q8WU68</a>	U2AF4_HUMAN	★	Splicing factor U2AF 26 kDa subunit	<b>U2AF1L4</b> U2AF1-RS3 U2AF1L3	Homo sapiens (Human)	220
<input type="checkbox"/> <a href="#">Q15696</a>	U2AFM_HUMAN	★	U2 small nuclear ribonucleoprotein auxiliary ...	ZRSR2 U2AF1-RS2 U2AF1L2 U2AF1RS2	Homo sapiens (Human)	482
<input type="checkbox"/> <a href="#">Q15695</a>	U2AFL_HUMAN	★	U2 small nuclear ribonucleoprotein auxiliary ...	ZRSR1 U2AF1-RS1 U2AF1L1 U2AF1P U2AF1RS1 U2AFBPL	Homo sapiens (Human)	479
<input type="checkbox"/> <a href="#">Q9H665</a>	TM149_HUMAN	★	Transmembrane protein 149	<b>TMEM149</b> <b>U2AF1L4</b> ?	Homo sapiens (Human)	355
<input type="checkbox"/> <a href="#">Q7Z780</a>	Q7Z780_HUMAN	☆	U2 small nuclear RNA auxiliary factor 1	U2AF1	Homo sapiens (Human)	167
<input type="checkbox"/> <a href="#">Q701P4</a>	Q701P4_HUMAN	☆	U2 snRNP auxiliary factor small subunit	U2AF1 hCG_401287	Homo sapiens (Human)	240
<input type="checkbox"/> <a href="#">Q71RF1</a>	Q71RF1_HUMAN	☆	FP793	U2AF1 hCG_401287	Homo sapiens (Human)	167
<input type="checkbox"/> <a href="#">B5BU08</a>	B5BU08_HUMAN	☆	U2 small nuclear RNA auxiliary factor 1 isofo...	U2AF1	Homo sapiens (Human)	240
<input type="checkbox"/> <a href="#">B5BU25</a>	B5BU25_HUMAN	☆	U2 small nuclear RNA auxiliary factor 2 isofo...	U2AF2	Homo sapiens (Human)	471

# 個別データ(ブラウザ表示)例

www.uniprot.org/uniprot/Q8WU68

### Names and origin

Protein names	<i>Recommended name:</i> <b>Splicing factor U2AF 26 kDa subunit</b> <i>Alternative name(s):</i> U2 auxiliary factor 26 U2 small nuclear RNA auxiliary factor 1-like protein 4 U2(RNU2) small nuclear RNA auxiliary factor 1-like protein 3 Short name=U2 small nuclear RNA auxiliary factor 1-like protein 3 Short name=U2AF1-like protein 3
Gene names	Name: <b>U2AF1L4</b> Synonyms:U2AF1-RS3, U2AF1L3
Organism	<b>Homo sapiens (Human)</b> [Complete proteome]
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo

### Protein attributes

Sequence length	220 AA.
Sequence status	Complete.
Sequence processing	The displayed sequence is further processed into a mature form.
Protein existence	Evidence at protein level.

### General annotation (Comments)

Function	RNA-binding protein that function as a pre-mRNA splicing factor. Plays a critical role in both constitutive and enhancer-dependent splicing by mediating protein-protein interactions and protein-RNA interactions required for accurate 3'-splice site selection. Acts by enhancing the binding of U2AF2 to weak pyrimidine tracts. Also participates in the regulation of alternative pre-mRNA splicing. Activates exon 5 skipping of PTPRO during T cell activation; an event reversed by GF11. Binds to RNA at the AG dinucleotide at the 3'-splice site (By similarity).
Subunit structure	Interacts with GF11 and U2AF2 (By similarity).
Subcellular location	<b>Nucleus</b> (By similarity). <b>Nucleus speckle</b> (By similarity).
Tissue specificity	Isoform 2 is widely expressed. Isoform 3 is highly expressed in heart, brain and lung, lower expressed in thymus and much lower expressed in peripheral blood leukocytes. (Ref.1)
Domain	The second zinc finger is necessary for interaction with GF11 and for alternative pre-mRNA splicing events (By similarity).

Names · Attributes · General annotation · Ontologies · Alt products · Sequence annotation · Sequences · References · Cross-refs · Entry info · Documents

www.uniprot.org/uniprot/Q9H665

<i>Recommended name:</i> <b>Transmembrane protein 149</b> <i>Alternative name(s):</i> U2 small nuclear RNA auxiliary factor 1-like 4
Name: <b>TMEM149</b> Synonyms:U2AF1L4
<b>Homo sapiens (Human)</b> [Complete proteome]
9606 [NCBI]
Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo
355 AA.
Complete.
The displayed sequence is further processed into a mature form.
Evidence at transcript level.

### General annotation (Comments)

Membrane; Single-pass type I membrane protein (Potential)
Membrane
Alternative splicing
Polymorphism
Signal
Transmembrane
Transmembrane helix
Complete proteome
integral to membrane
Inferred from electronic annotation. Source: UniProtKB-KW

General annotation · Ontologies · Alt products · Sequence annotation · Sequences · References · Cross-refs · Entry info · Documents

field(項目)

# テキスト表示例 (部分)

ID U2AF4\_HUMAN Reviewed; 220 AA.  
AC Q8WU68; A6NKI8; Q56UU3;  
DT 13-NOV-2007, integrated into UniProtKB/Swiss-Prot.  
DT 13-NOV-2007, sequence version 2.  
DT 10-AUG-2010, entry version 81.  
DE RecName: Full=Splicing factor U2AF 26 kDa subunit;  
DE AltName: Full=U2 auxiliary factor 26;  
DE AltName: Full=U2 small nuclear RNA auxiliary factor 1-like protein 4;  
DE AltName: Full=U2(RNU2) small nuclear RNA auxiliary factor 1-like protein 3;  
DE Short=U2 small nuclear RNA auxiliary factor 1-like protein 3;  
DE Short=U2AF1-like protein 3;  
GN Name=U2AF1L4; Synonyms=U2AF1-RS3, U2AF1L3;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;  
OC Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 2), AND TISSUE SPECIFICITY.  
RX PubMed=17312947; DOI=10.1080/10425170600807744;  
RA Chen F., Ji C., Dou T., Zheng N., Qiu R., Peng J., Fang W., Feng C.,  
RA Xie Y., Mao Y.;  
RT "Cloning and characterization of a novel splice variant of human  
RT U2AF1L3 gene."  
RL DNA Seq. 17:282-286(2006).

アクセッション番号は、統合  
などのため別名が複数ある

# Q9H665-1でBLASTし、結果をヒトに絞って表示

6 hits for **blastp** blast on **UNIPROTKB** sorted by **score** descending

Browse by [taxonomy](#), [keyword](#), [gene ontology](#), [enzyme class](#) or [pathway](#) | [Map to UniProtKB](#), [UniRef](#) or [UniParc](#) [Download](#)

Customize display [Filter](#) · [Overview](#) · [Results](#) · [Job information](#) Page 1 of 1

### Filter

Dataset: UniProtKB (6) Taxonomy: Homo sapiens (Human) (6) [Filter](#) [Reset](#)

### Graphic overview

Color code for identity 0-100% =

Accession	Entry name	QQuery hit355	OMatch hit (sqrt scale)1314	Name (Organism)
<input type="checkbox"/> Query18UR				
<input type="checkbox"/> Q9H665	TM1 49_HUMAN			Transmembrane protein 149 (Homo sapiens)
<input type="checkbox"/> Q9H665-2	TM1 49_HUMAN			Isoform 2 (Homo sapiens)
<input type="checkbox"/> B4DEL3	B4DEL3_HUMAN			cDNA FLJ50454, moderately similar to ... (Homo sapiens)
<input type="checkbox"/> B4DEU6	B4DEU6_HUMAN			cDNA FLJ55675, highly similar to Ster... (Homo sapiens)
<input type="checkbox"/> A2VEC9-2	SSPO_HUMAN			Isoform 2 (Homo sapiens)
<input type="checkbox"/> Q9BUK6-3	MSTO1_HUMAN			Isoform 3 (Homo sapiens)

### Detailed BLAST results

Alignments	Accession	Entry name	Status	Protein Names	Organism	Length	Identity	E-Value
<input type="checkbox"/>	Q9H665	TM1 49_HUMAN	★	Transmembrane protein 149	Homo sapiens (Human)	355	100.0%	0.0
<input type="checkbox"/>	Q9H665-2	TM1 49_HUMAN	★	Isoform 2	Homo sapiens (Human)	167	100.0%	2.0×10 <sup>-57</sup>
<input type="checkbox"/>	B4DEL3	B4DEL3_HUMAN	★	cDNA FLJ50454, moderately similar to	Homo sapiens	510	25.0%	1.5×10 <sup>-1</sup>

[Filter](#) · [Overview](#) · [Results](#) · [Job information](#)

自分自身TM149とそのisoformはあるが、U2AF4は無い。両者は配列では無関係。

# InterPro

interpro - Google 検索 x InterPro protein sequence ... x

www.ebi.ac.uk/interpro/

EBI > Databases > InterPro

### InterPro protein sequence analysis & classification

InterPro is an integrated database of predictive protein "signatures" used for the classification and automatic annotation of proteins and genomes. InterPro classifies sequences at superfamily, family and subfamily levels, predicting the occurrence of functional domains, repeats and important sites. InterPro adds in-depth annotation, including GO terms, to the protein signatures.

Current release: **28.0 11th August 2010** (see [Release Notes](#) for further details)

Search InterPro:

Do a sequence search of InterPro, via [InterProScan](#)

Extract large datasets by querying our [BioMart](#)

You can access our data programmatically, via [Web Services](#)

If you have any questions or feedback please [contact us](#).

InterPro 28.0

UniProt Universal Protein Resource	proSite	ProDom	PRINTS Protein Fingerprint Database	ProDom	SMART	TIGR THE INSTITUTE FOR GENOMIC RESEARCH tigr fams
HAMAP	PANTHER Classification System	PIRSF	Superfamily	Gene3D Domain Architecture Classification	PDBe PROTEIN DATA BANK EUROPE	CATH Protein Structure Classification
SCOP	SWISS-MODEL	MOD/BASE				

•タンパク質(おもに機能ドメイン)DBの統合DBである。

- ファミリー分類
- 機能ドメイン
- リpeat
- 翻訳後修飾など機能サイト
- 機能推定に必要な情報

•20837エントリーのデータ。  
(InterPro release 28.0)

検索窓にどのようなIDや語、あるいはそのリストを入れたらよいか、指示が表示される。

統合されたデータベース



# “dopamine” を検索窓に



EMBL-EBI EB-eye Search All Databases Enter Text Here Go Reset ? Advanced Search Give us feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

EBI > Databases > InterPro

Jump to: [InterProScan](#) [Databases](#) [Documentation](#) [FTP site](#) [Help](#)  
[Advanced search](#)

Search InterPro:  >

**Search for: dopamine**

- [Advanced search options](#)
- [Search for Gene Ontology terms in QuickGO](#)
- [Submit a sequence for automatic InterProScan Analysis \(sequence search\)](#)

[IPR000945](#) Dopamine-beta-monooxygenase  
[IPR000497](#) Dopamine 1B receptor  
[IPR009431](#) D1 dopamine receptor-interacting, calcyon  
[IPR001620](#) Dopamine D3 receptor  
[IPR000929](#) Dopamine receptor  
[IPR001922](#) Dopamine D2 receptor  
[IPR002185](#) Dopamine D4 receptor  
[IPR001413](#) Dopamine 1A receptor  
[IPR002436](#) Sodium:neurotransmitter symporter, dopamine  
[IPR015670](#) Dopamine-/cAMP-Regulated Neuronal Phosphoprotein  
[IPR013050](#) DOMON

検索結果の一部。  
この中から目的のエントリを選択。

## Contents and coverage of the current release

InterPro protein matches are now calculated for all UniProtKB and UniParc proteins. The following statistics are for all UniProtKB proteins. InterPro release 28.0 contains [20837](#) entries, representing:

Active site [89](#)  
Binding site [64](#)  
Conserved site [586](#)  
Domain [5608](#)  
Family [13023](#)  
PTM [16](#)  
Region [1188](#)  
Repeat [263](#)

Last entry: [IPR022869](#)

32320 publications in PUBMED are referenced from InterPro.

## Member database information

Signature Database	Version	Signatures	Integrated Signatures
PANTHER	6.1	<a href="#">30128</a>	<a href="#">2260</a>
Pfam	24.0	<a href="#">11912</a>	<a href="#">11431</a>
PIRSF	2.72	<a href="#">3221</a>	<a href="#">2751</a>
PRINTS	40.0	<a href="#">2000</a>	<a href="#">1962</a>
ProDom	2006.1	<a href="#">1894</a>	<a href="#">1008</a>
PROSITE patterns	20.52	<a href="#">1308</a>	<a href="#">1292</a>
PROSITE profiles	20.52	<a href="#">860</a>	<a href="#">837</a>
SMART	6.0	<a href="#">809</a>	<a href="#">804</a>
TIGRFAMs	9.0	<a href="#">3808</a>	<a href="#">3795</a>
GENE3D	3.0.0	<a href="#">2147</a>	<a href="#">1025</a>
SUPERFAMILY	1.69	<a href="#">1538</a>	<a href="#">1094</a>
HAMAP	180510	<a href="#">1656</a>	<a href="#">733</a>
PfamB	24.0	<a href="#">142303</a>	<a href="#">0</a>

検索結果のページ  
下部には、DB全体の  
統計情報が表示  
されている。

# Type = Domainの 例(上部)

Help(?)をクリックしてから、各field名の脇(?)をクリック。  
-->それぞれのfieldの説明が表示される。

www.ebi.ac.uk/interpro/IEntry?ac=IPR013050

EMBL-EBI EB-eye Search All Databases Enter Text Here Go Reset Advanced Search Give us feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

EBI > Databases > InterPro

Jump to: [InterProScan](#) [Databases](#) [Documentation](#) [FTP site](#) [Help](#) [Advanced search](#)

**IPR013050 DOMON**

**Protein matches**

UniProtKB Matches: 455 proteins

Overview: [sorted by AC](#), [sorted by name](#), [of known structure](#), [proteins with splice variants](#)  
Detailed: [sorted by AC](#), [sorted by name](#), [of known structure](#), [proteins with splice variants](#)  
Table: [For all matching proteins](#), [of known structure](#)

[Architectures](#)  
[Accession List](#)  
[Matches in BioMart](#)

<a href="#">Accession</a>	IPR013050 DOMON
<a href="#">Secondary</a>	IPR005018
<a href="#">Type</a>	Domain

**Signatures**

Database ID	Name	Proteins
<a href="#">Pfam</a> PF03351	DOMON	455

[Signatures in BioMart](#)

**InterPro Relationships**

<b>Parent</b>	<a href="#">IPR005018</a> DOMON related
<b>Found in</b>	<a href="#">IPR000945</a> Dopamine-beta-monoxygenase

**GO Term annotation**

<b>Process</b>	<a href="#">GO:0006548</a> histidine catabolic process
<b>Function</b>	<a href="#">GO:0004500</a> dopamine beta-monoxygenase activity

**InterPro annotation**

[Entry Details in BioMart](#)

**Abstract**

The DOMON domain is an 110-125 residue long domain which has been identified in the physiologically important enzyme dopamine beta-monoxygenase and in several other secreted and transmembrane proteins from both plants and animals. It has been named after Dopamine beta-Monoxygenase N-terminal domain. The DOMON domain can be found in one to four copies and in association with other domains, such as the Cu-ascorbate dependent monoxygenase domain, the epidermal growth factor domain, the trypsin inhibitor-like domain (TIL), the SEA domain and the Reelin domain. The architectures of the DOMON domain proteins strongly suggest a function in extracellular adhesion [1].

The sequence conservation is predominantly centred around patches of hydrophobic residues. The secondary structure prediction of the DOMON domain points to an all-beta-strand fold with seven or eight core strands supported by a buried core of conserved hydrophobic residues. There is a characteristic motif with two small positions (Gly or Ser) corresponding to a conserved turn immediately C-terminal to strand three. It has been proposed that the DOMON domain might form a beta-sandwich structure, with the strands distributed into two beta sheets as is seen in many extracellular adhesion domains such as the immunoglobulin, fibronectin type III, cadherin and PKD domains [1].

**Database links**

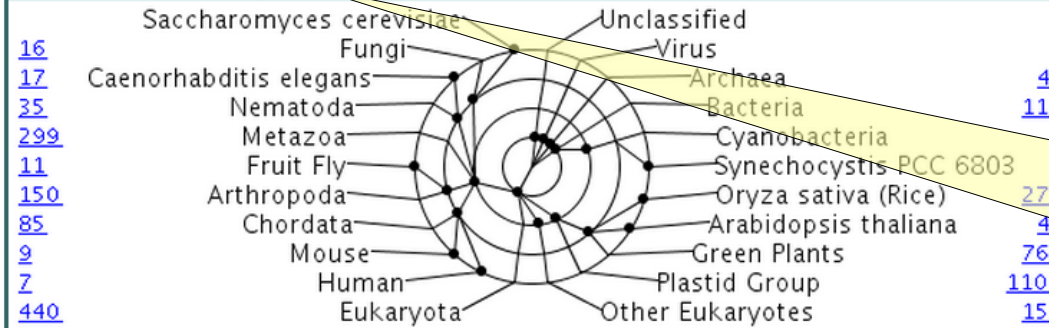
PANDIT: [PF03351](#)

**Taxonomic coverage**

Saccharomvces cerevisiae \ /Unclassified

# Type = Domainの例 (下部)

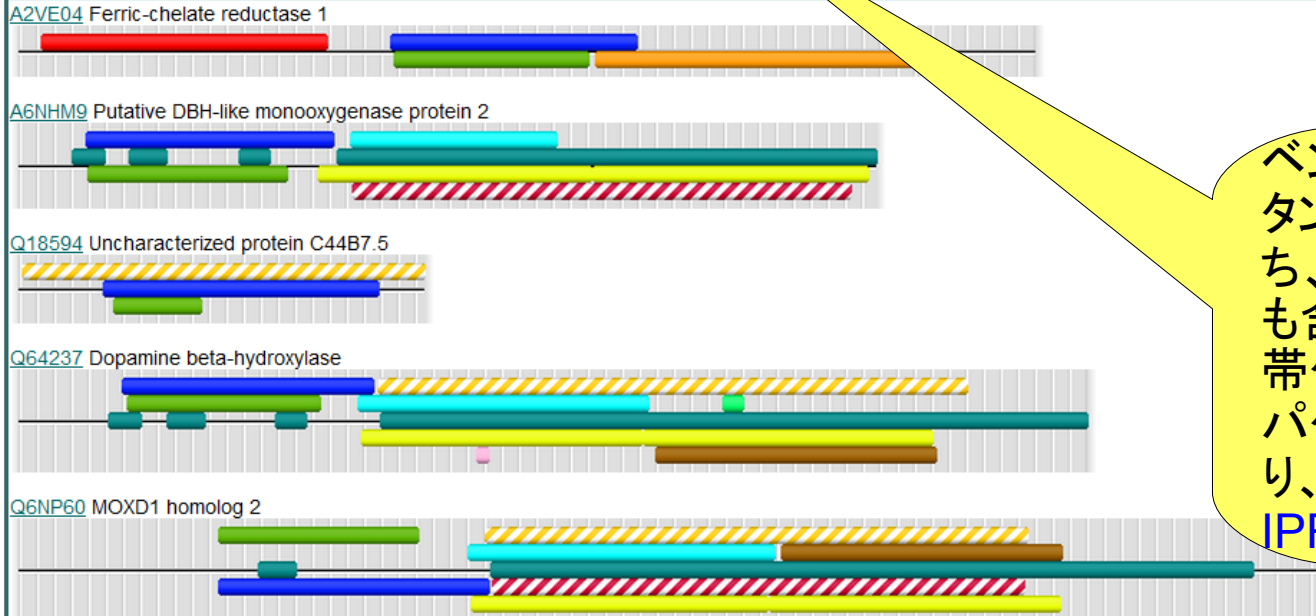
## Taxonomic coverage



## Overlapping InterPro entries



## Example proteins



[More proteins](#)

## Example Proteins Key

InterPro entry accession number/name and structure databases	Colour code
<a href="#">IPR020611</a> Copper type II, ascorbate-dependent monooxygenase, conserved site	
<a href="#">IPR002861</a> Reeler domain	
<a href="#">IPR008977</a> PHM/PNGase F-fold domain	

系統ごとに該当するエントリ(DOMON domain, IPR013050)が含まれているタンパク質の数が表示されている。最も内側の同心円が最も深い分類群を表す。Eukaryota(440)+Archaea(4)+Bacteria(11)=455

ベン図: IPR01350に含まれる455タンパク質(ピンク部分)のうち、434(紫部分)はIPR005018にも含まれる。  
帯グラフ: 両者に共通する434タンパク質は平均109アミノ酸が重なり、33アミノ酸は重ならずにIPR005018のみに見られる。

# InterProを用いた配列解析ツール InterProScan

- InterProに登録された各種機能ドメインを検索する。
- 機能未知配列中の各種機能関連sequence signatureを抽出して、特徴を発見する。
- ツールをダウンロードしてローカルで走らせることも可能。
- <http://www.ebi.ac.uk/Tools/InterProScan/>

# InterProScan トップ画面

The screenshot displays the InterProScan website interface. At the top, there is a search bar with the text "Search for: dopamine" and a navigation menu with options like "Databases", "Tools", "EBI Groups", "Training", "Industry", "About Us", and "Help". The main content area is titled "InterProScan Sequence Search" and includes a "Please Note" section about database updates. Below this is a "Download Software" section with a table of applications to run, all of which are checked. The main search form is titled "Enter or Paste a PROTEIN Sequence in any format:" and includes a "Submit Job" button and a "Reset" button. The footer contains copyright information and contact details.

EMBL-EBI EB-eye Search All Databases Enter Text Here Go Reset Advanced Search Give us feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

■ InterPro home  
■ Text Search  
■ InterProScan  
■ Databases  
■ Documentation  
■ FTP Site

■ InterProScan Help  
Help  
FAQ  
README

■ InterProScan Programmatic Access

■ InterPro BioMart

■ Database Information  
UniProt  
UniParc

EBI > Tools > Protein Functional Analysis

### InterProScan Sequence Search

This form allows you to query your sequence against InterPro. For more detailed information see the documentation for the perl stand-alone InterProScan package ([Readme file](#) or [FAQ's](#)), or the InterPro [user manual](#) or [help pages](#).

**Please Note:**  
Forthcoming changes: During the first week of January, the Pfam database will be scanned using HMMER3.03b.

[Download Software](#)

RESULTS YOUR EMAIL  
interactive

APPLICATIONS TO RUN  Clear all  Check all

<input checked="" type="checkbox"/> BlastProDom	<input checked="" type="checkbox"/> FPrintScan	<input checked="" type="checkbox"/> HMMPIR	<input checked="" type="checkbox"/> HMMPfam	<input checked="" type="checkbox"/> HMMSmart
<input checked="" type="checkbox"/> HMMTigr	<input checked="" type="checkbox"/> ProfileScan	<input checked="" type="checkbox"/> HAMAP	<input checked="" type="checkbox"/> patternScan	<input checked="" type="checkbox"/> SuperFamily
<input checked="" type="checkbox"/> TMHMM	<input checked="" type="checkbox"/> HMMPanther	<input checked="" type="checkbox"/> Gene3D		<input checked="" type="checkbox"/> SignalPHMM

Enter or Paste a PROTEIN Sequence in any format: [Help](#)

Upload a file: ファイルを選択 選択されていません [Submit Job](#) [Reset](#)

**PLEASE NOTE:** Interactive job results are stored for 24 hours, email job results are stored for one week. If you plan to use these services during a course please contact us using the email below.

[Terms of Use](#) | [EBI Funding](#) | [Contact EBI](#) | © European Bioinformatics Institute 2010. EBI is an Outstation of the [European Molecular Biology Laboratory](#).

# 配列を ペースト して、実行

The screenshot shows the InterProScan web interface. The browser tabs include 'Search for: dopamine', 'IPR000945 Dopamine-beta...', 'IPR000497 Dopamine 1B r...', and 'InterProScan Signature Re...'. The URL is 'www.ebi.ac.uk/Tools/InterProScan/'. The page header features the EMBL-EBI logo, 'EB-eye Search', and a search bar with 'All Databases' selected. The main content area is titled 'InterProScan Sequence Search' and includes a 'Please Note' section about Pfam database changes. Below this is a 'Download Software' section and a form for running applications. The form has a 'RESULTS' dropdown set to 'interactive' and a 'YOUR EMAIL' field. Under 'APPLICATIONS TO RUN', several tools are checked, including BlastProDom, FPrintScan, HMMPfam, HMMSmart, HMMTigr, ProfileScan, HAMAP, patternScan, SuperFamily, SignalPHMM, TMHMM, HMMPanther, and Gene3D. A text area contains a protein sequence: 'MGQPGNGSAFLLAPNGSHAPDHDVTOQRDEVWVGMGI VMSL I VLA I V F E G N M L V I T A T A K F E R L Q T V T N Y E I T S L A C A D L V M G L A V Y P F G A A H I L M K M W T E G N F W C E F W T S I D V L C V T A S I E T L C V I A V D R Y E A I T S P F K Y G S L L T K N K A R V I I L M W W I V S G L T S E L P T Q M H W Y R A T H Q E A I N C Y A N E T C C D F E F T N Q A Y A I A S S I V S F Y V P L V I M V E V Y S R V F E A K R Q L Q K I D K S E G R F H V Q N L S Q V E Q D G R T G H G L R R S S K F C L K E H K A L K T L G I I M G T E T L C W L P F E I V N I V H V I Q D N L I R K E Y Y I L L N W I G Y V N S G F N P L I Y C R S P D E R T A F Q E L L C L R R S S L K A Y G N G Y S S N G N T G E Q S G Y H V E Q E K E N K L L C E D L P G T E D F V G H Q G T V P S D N I D S P G R N C S T N D S L L'. Below the text area is a file upload section with the text 'Upload a file: ファイルを選択 選択されていません' and 'Submit Job' and 'Reset' buttons. A 'PLEASE NOTE' section at the bottom states: 'PLEASE NOTE: Interactive job results are stored for 24 hours, email job results are stored for one week. If you plan to use these services during a course please contact us using the email below.' The footer contains 'Terms of Use | EBI Funding | Contact EBI | © European Bioinformatics Institute 2010. EBI is an Outstation of the European Molecular Biology Laboratory.'

# InterPro 実行結果

Search for: dopamine | IPR000945 Dopamine-beta... | IPR000497 Dopamine 1B r... | InterProScan

www.ebi.ac.uk/Tools/es/cgi-bin/iprscan/iprscan.cgi?tool=iprscan&jobid=iprscan-20100922-0343386565

EMBL-EBI | EB-eye Search | All Databases | Enter Text Here | Go | Reset | Advanced Search | Give us feedback

Databases | Tools | EBI Groups | Training | Industry | About Us | Help | Site Index

- Help
- General Help
- Formats
- Gaps
- Matrix
- References
- InterProScan Help

### InterProScan Results

Table View | Raw Output | XML Output | Original Sequences | SUBMIT ANOTHER JOB

SEQUENCE: Sequence\_1 CRC64: FB17B56DC1EAD8A4 LENGTH: 413 aa

<b>InterPro</b> <a href="#">IPR000276</a> Family <b>InterPro</b> 	<b>7TM GPCR, rhodopsin-like</b>	
	<a href="#">PR00237</a>	GPCRRHODOPSN
	<a href="#">PF00001</a>	7tm_1
	<a href="#">PS00237</a>	G_PROTEIN_RECEP_F1_1
<b>InterPro</b> <a href="#">IPR000332</a> Family <b>InterPro</b> 	<b>Adrenergic receptor, beta 2</b>	
	<a href="#">PR00562</a>	ADENRGCB2AR
	<a href="#">PTHR19266:SF42</a>	BETA-2 ADRENERGIC RECEPTOR (BETA-2 ADRENOCEPTOR)
<b>InterPro</b> <a href="#">IPR002233</a> Family <b>InterPro</b> 	<b>Adrenergic receptor</b>	
	<a href="#">PR01103</a>	ADRENERGICR
<b>InterPro</b> <a href="#">IPR017452</a> Domain <b>InterPro</b> 	<b>GPCR, rhodopsin-like superfamily</b>	
	<a href="#">PS50262</a>	G_PROTEIN_RECEP_F1_2
noIPR unintegrated	<b>unintegrated</b>	
	<a href="#">G3DSA:1.20_1070.10</a>	no description
	<a href="#">PTHR19266</a>	G-PROTEIN COUPLED RECEPTOR
	SignalP	signal-peptide
	tmhmm	transmembrane_regions
	<a href="#">SSF81321</a>	Family A G protein-coupled receptor-like

Table View | Raw Output | XML Output | Original Sequences | SUBMIT ANOTHER JOB

Terms of Use | EBI Funding | Contact EBI | © European Bioinformatics Institute 2010. EBI is an Outstation of the European Molecular Biology Laboratory.

# タンパク質立体構造のデータベース



<http://www.wwpdb.org/>

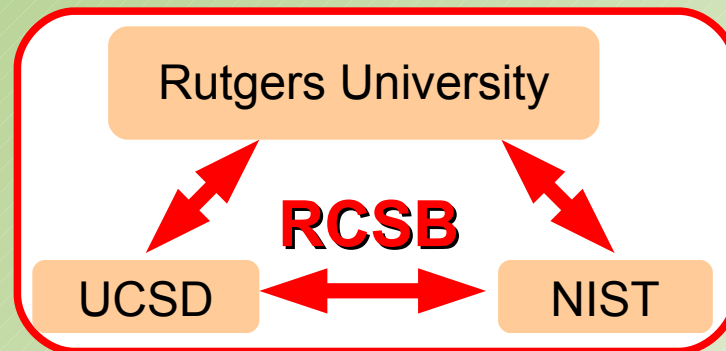


# PDBとは

- タンパク質および核酸の三次元構造のデータベース
- X線結晶解析法やNMR(核磁気共鳴)法実験によって得られた三次元データを登録者本人とアナノデータとでそれぞれ審査・検証を経て公開される。
- wwPDBにおける共同関係は核酸DBにおける国際塩基配列データベース(INSDC)に似ている。
  - データ・アーカイブは唯一で共通である。
  - 各機関はそれぞれにデータの登録受付とおよびブラウザ、ツール、webサービスを開発し公開する。

# wwPDBの組織

Research Collaboratory for  
Structural Bioinformatics  
(RCSB), USA



Biological Magnetic Resonance Data Bank  
(BMRB), USA

BMRB

Protein Databank in  
Europe (PDBe), Europe

PDBe (EBI)

PDBj

日本蛋白質構造データバンク  
(PDBj), Japan

# PDBデータの検索・取得

- <http://www.pdbj.org/>の「検索」を使う。
- 詳しい使い方は以下(今年8月名古屋大情報文化学部棟での講習会資料)参照のこと。
- <http://www.pdbj.org/workshop/201008/Kinjo.pdf>
- ただし、PDBのIDが既知である場合の検索の仕方で説明してある。未知の場合でもキーワードで検索可能。

# PDBデータの キーワード 検索1

情報リッチな画面であって  
も、大概のDBは検索窓が  
トップページにあるので、  
落ち着いて探してみよう。

www.pdbj.org/index\_j.html

English Japanese simplified Chinese traditional Chinese Korean 統計情報 ヘルプ FAQ お問い合わせ

トップページ  
データ登録 >>  
ADIT: PDB Deposition  
ADIT-NMR  
検索 >>  
Search PDB (Mine/xPSSS)  
Latest Released Search  
Sequence-Navigator  
Structure-Navigator  
SeSAW  
Ligand Binding Sites (GIRAF)  
EM Navigator  
Search NMR Data (BMRB)  
Status Search  
サービス&ソフトウェア >>  
JV: Graphic Viewer  
Protein Globe  
ASH  
MAFFTash  
Structure Prediction >>  
CRNPRED  
Spanner  
SFAS  
二次データベース >>  
eF-site/eF-seek/eF-surf  
eProtS  
ProMode  
Molecule of the Month  
ダウンロード >>  
PDB Archive/Snapshot Archive  
リンク集

日本蛋白質構造データバンク(PDBj: Protein Data Bank Japan)は、JST-BIRDの支援を受け、米国RCSB、BMRB、および欧州PDBeと協力して、生体高分子の立体構造データベースを国際的に統一化されたアーカイブとして運営するとともに、様々な解析ツールを提供しております。

データ登録  
データ登録のご案内 >>  
PDB登録 AP! Audio Dep Input Tool NMRデータ登録 AP!NMR

検索  
PDB検索 Mine PDBj  
Mine日本語ページについて  
U2AF  
PDB ID or Keyword Go  
NMRデータ検索 BMRB  
Accession number  
Deposition code  
Go  
詳細条件検索 >>

最新情報  
2010/8/25  
第48回生物物理学会年会(2010年9月20日(月)~22日(水)、会場:東北大学・川内キャンパス)にて、20日にランチョンセミナー、21日・22日に講習会を開催いたします。(講習会のご案内)  
2010/8/19  
PDF形式の構造検証レポートについて(詳細...)  
2010/8/10  
2010年8月9日(月)、10日(火)に、「ライフサイエンス・データベース講習会 in 名大」を名古屋大学情報文化学部棟にて開催いたしました。(詳細・セミナー資料)  
2010/7/13  
PDBj が開発しておりますフリーの分子グラフィックス・ビューア: **JVの最新版(jv3.8)**がリリースされました。displayatom on/off のコマンド機能が追加され、多様な表示が容易にできるようになっております。どうぞ、[こちら](#)からダウンロードしてお使いください。  
2010/6/30  
NMR距離制限情報ファイルバージョン2 が公開されました。(詳細...)  
2010/5/19  
JVの最新版(jv3.7.1)が公開されました。  
2010/5/7  
PDBj WebサーバおよびPDBj FTPサーバ(PDBMLplusを含む各種PDBデータ

68139 entries available on 22 Sep., 2010 00:00(UTC) / 09:00(JST)

WORLDWIDE PDB PROTEIN DATA BANK

蛋白質構造百科事典 eProtS Encyclopedia of Protein Structures

Protein Globe

DBCLS Database Center for Life Science

Tanpaku.org

National Project on Protein Structural and Functional Analyses

Bio Info R&D

MP900433188.JPG MP900438881.JPG MP900438753.JPG 全てのダウンロードを表示

# PDBデータの キーワード 検索2

分子名称が同一であっても、著者や実験手法(つまりエントリー)によって、3Dデータが異なるので注意。  
サムネイル画像の向きが違うだけかもしれない。

entry:エントリー、エントリ。ひとつの分子種につき、ひとつの実験で得られた立体構造情報。

The screenshot shows the PDBj Mine search results page for the keyword 'U2AF'. The page is in Japanese and displays a list of search results. The search criteria are 'U2AF' and 'PDB ID or Keyword'. The results are sorted by '一致件数' (Number of matches) and show 10 results. The first three results are visible in the image.

分子ID	分子名称	タイトル	著者	実験手法	登録日	公開日
1jmt	SPLICING FACTOR U2AF 35 KDA SUBUNIT, SPLICING FACTOR U2AF 65 KDA SUBUNIT	X-ray Structure of a Core U2AF65/U2AF35 Heterodimer	Kielkopf, C.L., Rodionova, N.A., Green, M.R., Burley, S.K.	X-RAY DIFFRACTION	2001-07-19	2001-09-19
2hzc	Splicing factor U2AF 65 kDa subunit	Crystal structure of the N-terminal RRM of the U2AF large subunit	Thickman, K.R., Sickmier, E.A., Kielkopf, C.L.	X-RAY DIFFRACTION	2006-08-08	2006-08-29
2u2f	SPLICING FACTOR U2AF 65 KD SUBUNIT	SOLUTION STRUCTURE OF THE SECOND RNA-BINDING DOMAIN OF HU2AF65	Ito, T., Muto, Y., Green, M.R., Yokoyama, S., RIKEN Structural Genomics/Proteomics Initiative (RSGI)	SOLUTION NMR	1999-05-26	1999-08-20
1u2f	SPLICING FACTOR U2AF 65 KD SUBUNIT	SOLUTION STRUCTURE OF THE FIRST RNA-BINDING DOMAIN OF HU2AF65	Ito, T., Muto, Y., Green, M.R., Yokoyama, S., RIKEN Structural Genomics/Proteomics Initiative (RSGI)	SOLUTION NMR	1999-05-26	1999-08-20

# 個別データ 表示例

service.pdbj.org/mine/Detail2?PDBID=2U2F&PAGEID=Summary

Japanese 統計情報 ヘルプ FAQ お問い合わせ

トップページ

データ登録 >>

- ADIT: PDB Deposition
- ADIT-NMR

検索 >>

- Search PDB (Mine/xPSSS)
- Latest Released Search
- Sequence-Navigator
- Structure-Navigator
- SeSAW
- Ligand Binding Sites (GIRAF)
- EM Navigator
- Search NMR Data (BMRB)
- Status Search

サービス&ソフトウェア >>

- JV: Graphic Viewer
- Protein Globe
- ASH
- MAFFTash
- Structure Prediction >>

  - CRNPRED
  - Spanner
  - SFAS

二次データベース >>

- eF-site/eF-seek/eF-surf
- eProtS
- ProMode
- Molecule of the Month

ダウンロード >>

- PDB Archive/Snapshot Archive

リンク集

日本蛋白質構造データバンク(PDBj: Protein Data Bank Japan)は、JST-BIRDの支援を受け、米国RCSB、BMRB、および欧州PDBeと協力して、生体高分子の立体構造データベースを国際的に統一化されたアーカイブとして運営するとともに、様々な解析ツールを提供しております。

**Mine** 概要[2u2f] [日本語ページについて](#) [PDBj Mineについて](#) [更新情報](#)

概要 構造情報 実験情報 機能情報 相同蛋白質 ダウンロード/画面表示 外部データベース

2u2f 配列情報 (FASTA形式) [PDBファイルのダウンロード](#)

エントリーID (PDB ID) 1u2f

関連構造のPDB ID 1u2f

分子名称 SPlicing Factor U2AF 65 KD SUBUNIT

タイトル SOLUTION STRUCTURE OF THE SECOND RNA-BINDING DOMAIN OF HU2AF65

機能のキーワード SPlicing, U2 snRNP, RBD, RNA-binding protein, RIKEN Structural Genomics/Proteomics Initiative, RSGI, Structural Genomics

由来する生物種 Homo sapiens (human)

細胞内の位置 [UNP - U2AF2\_HUMAN] Nucleus

由来する組織 [UNP - U2AF2\_HUMAN] Lymph, and Skin [UNP - U2AF2\_HUMAN] Colon carcinoma [UNP - U2AF2\_HUMAN] Hepatocyte [UNP - U2AF2\_HUMAN] Epithelium

ポリマー鎖の合計数 1

分子量の合計 9096.5 (詳細は [構造情報のページ](#))

著者 Ito, T., Muto, Y., Green, M.R., Yokoyama, S., RIKEN Structural Genomics/Proteomics Initiative (RSGI) (登録日: 1999-05-26, 公開日: 1999-08-20)

引用文献 Ito, T., Muto, Y., Green, M.R., Yokoyama, S. Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF(65)).

3次元構造ビューア [jV3 / Jmol](#)  
jV3とJmolにはJava(TM)Plug-in 1.5以上が必要です。

x軸周りに90度回転 250X250 500X500

y軸周りに90度回転 250X250 500X500

MP900433188.JPG MP900438881.JPG MP900438753.JPG [すべてのダウンロードを表示](#)

# PDBのentry dataには3種類の書式がある

書式	説明	完全性	人による判別	プログラムでの処理
PDB	Flat file形式。情報に欠落があり、厳密な解析には向かない。	—	++	—
mmCIF	国際結晶学連合(IUCr)が定めた(Crystallographic Information File) CIF形式を拡張した。 <a href="http://mmcif.pdb.org/Data_parsing_tools(C++)">http://mmcif.pdb.org/Data_parsing_tools(C++)</a> がRCSBから用意されている( <a href="http://sw-tools.rcsb.org/">http://sw-tools.rcsb.org/</a> )。	+	+	+
PDBML	mmCIFに基づくXML形式。 <a href="http://pdbml.pdb.org/">http://pdbml.pdb.org/</a>	+	—	++

どの形式でもダウンロードできる。

# PDB書式の例(ヘッダ一部分抜粋)

field名 ▼ 同一field内で2行以上になるときには行番号がついている。

```
HEADER      RNA BINDING PROTEIN                      08-AUG-06    2HZC
TITLE       CRYSTAL STRUCTURE OF THE N-TERMINAL RRM OF THE U2AF LARGE
TITLE       2 SUBUNIT
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: SPLICING FACTOR U2AF 65 KDA SUBUNIT;
COMPND      3 CHAIN: A;
COMPND      4 FRAGMENT: RRM 1;
COMPND      5 SYNONYM: U2 AUXILIARY FACTOR 65 KDA SUBUNIT, U2 SNRNP
COMPND      6 AUXILIARY FACTOR LARGE SUBUNIT, HU2AF65;
COMPND      7 ENGINEERED: YES
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE      3 ORGANISM_COMMON: HUMAN;
SOURCE      4 ORGANISM_TAXID: 9606;
SOURCE      5 GENE: U2AF2, U2AF65;
```

▼ 同一field内で2行以上になるときには行番号がついている。

```
JRNL        AUTH    K.R.THICKMAN,E.A.SICKMIER,C.L.KIELKOPF
JRNL        TITL    ALTERNATIVE CONFORMATIONS AT THE RNA-BINDING
JRNL        TITL 2  SURFACE OF THE N-TERMINAL U2AF(65) RNA RECOGNITION
JRNL        TITL 3  MOTIF.

REMARK      1 REFERENCE 1
REMARK      1 AUTH    E.A.SICKMIER,K.E.FRATO,S.PARANAWITHANA,H.SHEN,
REMARK      1 AUTH 2  M.R.GREEN,C.L.KIELKOPF
REMARK      1 REF     MOL.CELL                      V.  23    49 2006
REMARK      2 RESOLUTION.    1.47 ANGSTROMS.
```

'REMARK 数字' で一field扱い



# PDB書式の例(分子データ部分抜粋)

アミノ酸番号

```
ATOM      1  N   GLY A 143          1.397  10.715  63.542  1.00 79.91          N
ANISOU    1  N   GLY A 143          17305   4118   8949   3151  -2891  -3871          N
ATOM      2  CA  GLY A 143          0.525  10.264  62.474  1.00 76.48          C
ANISOU    2  CA  GLY A 143          14905   5306   8858   3443  -2947  -2851          C

HETATM   925  O   HOH A 492          21.491  -8.646  69.133  1.00 45.06          O
ANISOU   925  O   HOH A 492          5806   4899   6418   309   -542  -1086          O
HETATM   926  O   HOH A 493          11.582  -0.505  36.484  1.00 38.57          O
ANISOU   926  O   HOH A 493          5534   4232   4885   -492  -1658  -847          O
HETATM   927  O   HOH A 494          28.316  -8.557  58.005  1.00 54.25          O
ANISOU   927  O   HOH A 494          7247   5534   7837   2015  1449  -1110          O
HETATM   928  O   HOH A 495          27.091  -6.522  50.635  1.00 41.14          O
ANISOU   928  O   HOH A 495          5567   4645   5421   851   1342   163          O
CONNECT  163   715
CONNECT  174   716
CONNECT  207   716

MASTER          324      0      3      4      6      0      7      6  927      1   28      7
END
```

The ATOM records present the atomic coordinates for standard amino acids and nucleotides.

The ANISOU records present the anisotropic temperature factors

詳細は<http://www.wwpdb.org/docs.html>

# 名称、用語、概念の共通化に関する データベース

Gene Ontology (GO):

<http://www.geneontology.org/>

HGNC

<http://www.genenames.org/>

# Gene Ontologyとは

- “Ontology” = “on”存在 + “logy”
  - 元々、哲学のひとつ存在論: entity (実体) の存在のあり方や他の実体との関係性 (階層性、分類・体系性) について問う学問。
  - 情報学において、「概念化の明示的な仕様」と定義。
    - 同表記異義語の問題を解決、
    - 文章: 単語の集まり --> 意味のある実体として、コンピュータ処理が可能となった。(発展; セマンティック・ウェブ)
- Gene Ontology Project: 生物種やDBさらに分野の枠を超え、遺伝子 (産物) 関連用語を標準化。
  - 異祖同機能の問題解決、
  - --> DB間でのリンクや統合が容易に。
  - おもなゲノム関連研究機関が参加。
- GOは網羅的な遺伝子解析結果集計に多用。

# GO termの構造

- GO term全体に階層性を持つ。
- GOは3つのdomainをカバーする。
  - cellular component, 遺伝子産物細胞内外分布
  - molecular function, 遺伝子産物の機能
  - biological process, 生体内における役割

例:

Accession	GO:0015030
Ontology	cellular component

# GO geneontology.org

The screenshot shows the Gene Ontology website homepage. At the top, there is a navigation bar with links for Downloads, Tools, Documentation, Projects, About, and Contact. A search bar is located in the top right corner, with a dropdown menu showing options like 'gene or protein name', 'GO term or ID', and 'GO website'. The main content area features a large heading 'Welcome to the Gene Ontology website!' followed by a paragraph describing the project's goal of standardizing gene and gene product attributes. Below this is a section titled 'Search the Gene Ontology Database' with a search input field and a 'GO!' button. The search options are set to 'gene or protein name'. To the right, there are two sidebars: 'Quick Links' with various tools and download links, and 'News' with recent updates and announcements.

GO The Gene Ontology

www.geneontology.org

the Gene Ontology

Search

gene or protein name

gene or protein name

GO term or ID

GO website

Downloads Tools Documentation Projects About Contact

## Welcome to the Gene Ontology website!

The Gene Ontology project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. The project provides [a controlled vocabulary of terms](#) for describing gene product characteristics and [gene product annotation data](#) from GO Consortium members, as well as [tools to access and process this data](#). [Read more about the Gene Ontology...](#)

### Search the Gene Ontology Database

Search for genes, proteins or GO terms using [AmiGO](#):

GO!

gene or protein name  GO term or ID

[AmiGO](#) is the official GO browser and search engine. [Browse the Gene Ontology with AmiGO](#).

The Gene Ontology project very much encourages input from the community into both the content of the GO and annotation using GO. We are very happy to work with others to ensure that the GO is both complete and accurate, and we also very much encourage communities to submit GO annotations for inclusion in the GO database. [Please contact us](#).

The Gene Ontology Consortium is supported by a P41 grant from the National Human Genome Research Institute (NHGRI) [grant [SP41HG002273-09](#)]. [See the full list of funding sources](#). The Gene Ontology Consortium would like to acknowledge

#### Quick Links

- Tools
- AmiGO browser
- OBO-Edit ontology editor
- Ontology downloads
- Annotation downloads
- Database downloads
- Documentation
- GO FAQ
- GO on SourceForge
- Contact GO

#### News

- GO on Twitter
- BBOP services down over the weekend (LBL power outage) (7 days ago) [News item](#)
- Two positions available at Virginia Bioinformatics Institute (14 days ago) [News item](#)
- Announcing the new

# AmiGOはGOのブラウザ: 例(上部)

The screenshot shows the AmiGO web browser interface. The browser's address bar displays the URL: `amigo.geneontology.org/cgi-bin/amigo/term-details.cgi?term=GO:0015030&session_id=70amigo1285210575&`. The page header includes the text "the Gene Ontology" and the "AmiGO" logo. A navigation bar contains links for "Search", "Browse", "BLAST", "More Tools", and "Help". Below this is a search bar with the text "Search GO" and a search button labeled "送信". The main content area is titled "Cajal body" and features a menu with options: "Term information", "Term lineage", "External references", and "135 gene product associations". The "Term Information" section is expanded, showing the following details:

<b>Accession</b>	GO:0015030
<b>Ontology</b>	cellular component
<b>Synonyms</b>	<b>exact:</b> coiled body <b>exact:</b> Gems
<b>Definition</b>	A class of nuclear body, first seen after silver staining by Cajal in 1903, enriched in small nuclear ribonucleoproteins, and certain general RNA polymerase II transcription factors; ultrastructurally, they appear as a tangle of coiled, electron-dense threads roughly 0.5 micrometers in diameter; involved in aspects of snRNP biogenesis; the protein coilin serves as a marker for Cajal bodies. Some argue that Cajal bodies are the sites for preassembly of transcriptosomes, unitary particles involved in transcription and processing of RNA. [source: <a href="#">PMID:10944589</a> , <a href="#">PMID:11031238</a> , <a href="#">PMID:7559785</a> , <a href="http://genetics.cwru.edu/matera3.html">http://genetics.cwru.edu/matera3.html</a> ]
<b>Comment</b>	None
<b>Subset</b>	None
<b>Community</b>	There have been <a href="#">0 comments</a> for this term. If you would like to view or participate in the community annotation, please continue to the <a href="#">GONUTS page</a>

# AmiGOはGOのブラウザ: 例(下部)

amigo.geneontology.org/cgi-bin/amigo/term-details.cgi?term=GO:0015030&session\_id=70amigo1285210575&

## Term Lineage

[Switch to viewing term parents, siblings and children](#)

### Filter tree view ?

Filter Gene Product Counts

Data source

All  
ASAP  
AspGD  
CGD

Species

Candida albicans  
Carboxydotherrmus ...  
Clostridium perfr...  
Colwellia psychre...

View Options

Tree view  Full  Compact

Set filters

Remove all filters

all : all [447732 gene products]

+ GO:0005575 : cellular\_component [309658 gene products]

+ GO:0005623 : cell [219966 gene products]

+ GO:0044464 : cell part [219929 gene products]

+ GO:0005622 : intracellular [159346 gene products]

+ GO:0044424 : intracellular part [150601 gene products]

+ GO:0043229 : intracellular organelle [122505 gene products]

+ GO:0043231 : intracellular membrane-bounded organelle [102549 gene products]

+ GO:0005634 : nucleus [38516 gene products]

+ GO:0044428 : nuclear part [12329 gene products]

+ GO:0031981 : nuclear lumen [7717 gene products]

+ GO:0005654 : nucleoplasm [5126 gene products]

+ GO:0044451 : nucleoplasm part [3933 gene products]

+ GO:0016604 : nuclear body [672 gene products]

+ **GO:0015030 : Cajal body [135 gene products]**

+ GO:0005654 : nucleoplasm [5126 gene products]

+ GO:0044451 : nucleoplasm part [3933 gene products]

+ GO:0016604 : nuclear body [672 gene products]

Actions...

Last action: Reset the tree

Graphical View

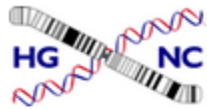
View in tree browser

Download...

OBO

RDF/XML

GraphViz dot



## HUGO Gene Nomenclature Committee


[About HGNC](#)
[Gene Search](#)
[Guidelines](#)
[Gene Submission](#)
[Downloads](#)
[Home](#)

*Giving unique and meaningful names to every human gene*

[HGNC Activities](#)
[Useful Links](#)
[HCOP Search Tool](#)
[Public Engagement of Science](#)
[FAQs](#)
[International Advisory Committee](#)
[Gene Families/Groupings](#)
[BioMart](#)

### Current Staff

Project coordinator

[Dr Elspeth Bruford](#) (on maternity leave)

Gene nomenclature advisors:

[Dr Matt Wright](#)

[Dr Susan Gordon](#)

[Dr Ruth Seal](#)

Bioinformatics support:

[Dr Michael Lush](#)

**FEEDBACK** - we welcome your feedback, please click [here](#) to leave your comments and/or suggestions.

 [Quick Gene Search](#)


[Advanced Gene Search](#)

We have approved over **29,000** human gene symbols and names. Each symbol is unique and we ensure that each gene is only given one approved gene symbol. Search the HGNC database for your gene.

GCCTGGG  
CATTGGAC  
CGGCTCCC

[Request a Gene Symbol](#) - online request form

Obtaining a gene symbol before publication will avoid any possible conflicts with existing symbols and will ensure that your gene is promptly recorded in our database and others. Any information that you provide will be treated in the strictest confidence. For bulk data submissions please email us at [hgnc@genenames.org](mailto:hgnc@genenames.org) prior to submission.



[Gene Families and Groupings](#)

We strongly encourage the use of a stem (or root) symbol as a basis for a hierarchical series that allows the easy identification of other related members in both database searches and the literature. Please contact us as soon as possible with new members of gene families, as some symbols may be reserved in our database.



[New on genenames.org](#)

Our [Gene Search](#) page now includes a "[list search](#)" option.

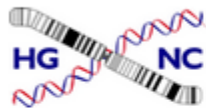
Our [Downloads](#) page has been reformatted to include statistics.



# 表示例

Core Data		Database Links			
<a href="#">Approved Symbol +</a>	<b>HMGA1</b>	<a href="#">RefSeq IDs +</a>			
<a href="#">Approved Name +</a>	high mobility group AT-hook 1	NM_145899	<a href="#">GenBank</a>	<a href="#">EMBL</a>	<a href="#">DDBJ</a> <a href="#">UCSC</a>
<a href="#">HGIC ID +</a>	HGNC:5010	<a href="#">Accession Numbers +</a>			
<a href="#">Status +</a>	Approved	AF176039	<a href="#">GenBank</a>	<a href="#">EMBL</a>	<a href="#">DDBJ</a> <a href="#">UCSC</a>
<a href="#">Chromosome +</a>	6p21	<a href="#">Mouse Genome Database ID +</a>			
<a href="#">Previous Symbols +</a>	HMG1Y	MGI:96160	<a href="#">MGD ID</a>		
<a href="#">Previous Names +</a>	"high-mobility group (nonhistone chromosomal) protein isoforms I and Y"	<a href="#">Rat Genome Database ID (mapped data supplied by RGD) +</a>			
<a href="#">Aliases +</a>		RGD:628699	<a href="#">RGD ID</a>		
<a href="#">Name Aliases +</a>		<a href="#">Entrez Gene ID +</a>			
<a href="#">Locus Type +</a>	gene with protein product	3159	<a href="#">Gene</a>	<a href="#">Map Viewer</a>	
		<a href="#">CCDS IDs +</a>			
		CCDS4788.1	<a href="#">CCDS ID</a>		
		<a href="#">Pubmed IDs +</a>			
		8414980, 11406267	<a href="#">PMID</a>	<a href="#">CiteXplore</a>	
		<a href="#">VEGA IDs +</a>			
		OTTHUMG00000014539	<a href="#">VEGA GeneView</a>		
		<a href="#">Ensembl ID (mapped data supplied by Ensembl) +</a>			
		ENSG00000137309	<a href="#">Ensembl GeneView</a>	<a href="#">UCSC</a>	
		<a href="#">OMIM ID (mapped data supplied by NCBI) +</a>			
		600701	<a href="#">OMIM</a>		
		<a href="#">UCSC ID (mapped data supplied by UCSC) +</a>			
		uc003oiz.2	<a href="#">UCSC Index</a>		
		<a href="#">UniProt ID (mapped data supplied by UniProt) +</a>			
		P17096	<a href="#">UniProt</a>	<a href="#">UCSC</a>	
		<p><a href="#">GENATLAS</a> <a href="#">GeneCards</a> <a href="#">GeneClinics</a> <a href="#">GeneTests</a> <a href="#">GoPubMed</a></p> <p><a href="#">HCOP</a> <a href="#">H-InvDB</a> <a href="#">Treefam</a> <a href="#">wikigenes</a></p>			
		<a href="#">COSMIC</a>			
		<a href="#">Specialist Database Links</a>			
		<a href="#">Locus Specific Database Links</a>			

# HCOP



## HGNC Comparison of Orthology Predictions



About HGNC

Gene Search

Guidelines

Gene Submission

Downloads

Home

*Giving unique and meaningful names to every human gene*

[HGNC Comparison of Orthology Predictions Search](#)

### Overview

The HGNC Comparison of Orthology Predictions (HCOP) search is a tool that integrates and displays the orthology assertions predicted for a specified human gene, or set of human genes, by [Ensembl](#), [Evola](#), [HGNC](#), [Homologene](#), [Inparanoid](#), [MGI](#), [OMA](#), [OPTIC](#), [Treefam](#), [UCSC](#) and [ZFIN](#). An indication of the reliability of a prediction is provided by the number of databases which concur and by the presence or absence of synteny between the relevant chromosomes, where known. HCOP was originally designed to show orthology predictions between human and mouse, but now with the addition of rat, cow, chicken, dog, platypus, fruitfly, chimp, zebrafish, C. elegans and S. cerevisiae data there are currently 12 genomes available for comparison in HCOP.

### Using the search

Orthology assertions can be obtained for a gene by searching with either its approved symbol, Entrez Gene ID, HGNC or MGI ID, or RefSeq ID. The species of the query gene and of the predicted orthologs can optionally be specified. The results provide basic data about the query and its predicted homologs as well as a list of databases that support the assertion and links to further information.

The consensus orthology assertions for multiple genes can be viewed simultaneously by searching with a list of query terms, separated by commas, newlines or spaces.

Searches are case insensitive. The HGNC: or MGI: prefix, and the version number of a RefSeq ID (the last . and following digits) are not required. Only approved nomenclature is shown. Where no name or symbol is given, the relevant nomenclature committee has not yet named this gene.

### Synteny

Synteny is assessed on a whole-chromosome basis using the MGI version of the Oxford Grid. A human and a mouse chromosome are stated as syntenic if homology between them would not create a singleton. According to this method the following chromosomes are syntenic:

Human Chromosome	Syntenic Mouse Chromosomes	Human Chromosome	Syntenic Mouse Chromosomes
1	1 3 4 5 8 11	2	1 2 5 6 10 11 12 17 18
3	3 6 9 14 16 17	4	3 5 6 8
5	1 11 13 15 17 18	6	1 4 9 10 13 17
7	5 6 9 11 12 13	8	1 3 4 8 14 15
9	2 4 13 19	10	2 6 7 8 10 13 14 18 19

# ゲノムDB・ゲノムbrowser

- NCBI
  - NCBI Map viewer
  - Model Maker (mm)
- Ensembl
  - Ensembl Genome browser
  - BioMart
- UCSC
  - UCSC genome browser
- J.Craig Venter Institute (JCVI) HuRef
- バージョンについて

# NCBI home

National Center for Biotechnology Information

Search  All Databases

### Resources

- NCBI Home
- All Resources (A-Z)
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

## Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[More about the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [RSS](#)

### PubMed Central

Free Full Text. Over 1,500,000 articles from over 450 journals. Linked to PubMed and fully searchable.

1 2 3 4

### How To...

- Determine conserved syntenic regions between the genomes of two organisms
- Find a homolog for a gene in another organism
- Obtain the full text of an article
- Design PCR primers and check them for specificity

[See all ...](#)

### Popular Resources

- BLAST
- Bookshelf
- Gene
- Genome
- Nucleotide
- OMIM
- Protein
- PubChem
- PubMed
- PubMed Central
- SNP

### NCBI News

[MyNCBI supports OpenID and InCommons IDs](#)  
22 Sep 2010

A list of supported organizations can be found on the MyNCBI

[Personalized settings in My NCBI](#)  
30 Aug 2010

Personalized preferences such as record format and results display

[More...](#)

You are here: NCBI [Write to the Help Desk](#)

GETTING STARTED	RESOURCES	POPULAR	FEATURED	NCBI INFORMATION
<a href="#">NCBI Help Manual</a>	<a href="#">Literature</a>	<a href="#">PubMed</a>	<a href="#">GenBank</a>	<a href="#">About NCBI</a>
<a href="#">NCBI Handbook</a>	<a href="#">DNA &amp; RNA</a>	<a href="#">Nucleotide</a>	<a href="#">Reference Sequences</a>	<a href="#">Research at NCBI</a>
<a href="#">Training &amp; Tutorials</a>	<a href="#">Proteins</a>	<a href="#">BLAST</a>	<a href="#">Map Viewer</a>	<a href="#">NCBI Newsletter</a>
	<a href="#">Sequence Analysis</a>	<a href="#">PubMed Central</a>	<a href="#">Genome Projects</a>	<a href="#">NCBI FTP Site</a>
	<a href="#">Genes &amp; Expression</a>	<a href="#">Gene</a>	<a href="#">Human Genome</a>	
	<a href="#">Genomes &amp; Maps</a>	<a href="#">Bookshelf</a>	<a href="#">Mouse Genome</a>	
	<a href="#">Domains &amp; Structures</a>	<a href="#">Protein</a>	<a href="#">Influenza Virus</a>	
	<a href="#">Genetics &amp; Medicine</a>	<a href="#">OMIM</a>	<a href="#">Primer-BLAST</a>	
		<a href="#">Genome</a>	<a href="#">Sequence Read Archive</a>	

www.ncbi.nlm.nih.gov を待機しています...

# NCBI > genome

Genome Home | Map Viewer | www.ncbi.nlm.nih.gov/sites/genome

NCBI Entrez Genome

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | OMIM | PMC | Journals | Books

Search Genome for [ ] Go Clear

Limits | Preview/Index | History | Clipboard | Details

The Genome database provides views for a variety of genomes, complete chromosomes, sequence maps with contigs, and integrated genetic and physical maps. The database is organized in six major organism groups: [Archaea](#), [Bacteria](#), [Eukaryotae](#), [Viruses](#), [Viroids](#), and [Plasmids](#) and includes complete chromosomes, organelles and plasmids as well as draft genome assemblies.

**Total species (6596)**

Viruses	Eukaryota	Archaea	Plasmids
2425	2455	102	39
	Bacteria	Viroids	
	1534	41	

**Total records (12475)**

Viruses	Eukaryota	Bacteria	Archaea
3682	1314 chromosomes 2448 organelles 43 plasmids	2520 chromosomes 2189 plasmids	103 chromosomes 95 plasmids

Genome Sequencing Milestone Reached! There are now 1000 complete Prokaryotic Genomes available in Entrez Genome. See the full list of [complete bacterial and archaeal genomes](#). [Microbial Resources](#) are available for search, retrieval, and analysis of all genomes.

**Microbial Genomes Resources** presents public data from prokaryotic genome sequencing projects. The sequence collection contains data from finished genomes as well as draft assemblies. The analytical tools include specialized [BLAST](#) with microbial genomes, newly developed [Concise Protein BLAST](#), annotation tools and many more.

**Identification of SNPs** in two Salmonella enterica serovar Enteritidis PT13a pathotypes that point to epidemiological trends. There has been an increasing number of infections leading to salmonellosis by Salmonella enterica serovar Enteritidis in the United States. Using preliminary genomic sequence data of Salmonella enterica subsp. I serovar Enteritidis PT4 from the Sanger Institute as a starting point, researchers at The Egg Safety and Quality Research Unit (ESQRU) of the U.S. Department of

**Related resources**

- [Entrez Genome Project](#) complete and incomplete large-scale sequencing projects
- [Entrez Protein Clusters](#) a collection of related protein sequences
- [Eukaryotic genome projects and sequences](#)
- [Genomes of Bacillus anthracis](#) reference genome and related sequences
- [Influenza Virus Resource](#) sequence database and analyses
- [Microbial Genomes](#) reference sequences and resources
- [Organelle](#) reference sequences and tools
- [Plant Genomes Central](#) major plant genome projects
- [SARS Coronavirus Resource](#) sequence data and analyses
- [Salmonella SNPs](#) SNP data in two Salmonella enterica pathotypes
- [Viruses](#) reference sequences and tools
- [WGS Projects](#) Whole Genome Shotgun sequencing

**Tools and Analysis**

- [gMap](#) comparative analysis of microbial genomes
- [Map Viewer](#) genome browser for eukaryotic genomes
- [TaxPlot](#)

Click!

The Map Viewer provides a wide variety of genome mapping and sequencing data. [More...](#)

**Search**

Search:

for:

- Tools Legend**
- Search or Browse the Genome
  - BLAST
  - Clone Finder
  - Go to region on a chromosome
  - Genome Resources page

**News**

**Human build 37 released** Aug 3, 2009  
 An update to the human genome assembly and annotation is now... [more](#)

**Annotation update released for human genome build 36** Mar 24, 2008  
 An annotation update for the human genome (NCBI Build 36.3) ... [more](#)

[Show all](#)

- Related Resources**
- NCBI Home
  - NCBI Web Search
  - NCBI Site map
  - Genome Browser agreement
  - Genome Biology
  - Taxonomy
  - Entrez (Global Query)
  - BLAST
  - Map Viewer FTP

- Small Genomes**
- Bacteria
  - Organelles
  - Viruses

Vertebrates (19)			
Mammals (15)			
Primates (3)			
Scientific name	Common name	Build	Tools
<i>Homo sapiens</i>	human	<a href="#">Build 37.1</a> <a href="#">Build 36.3</a>	
<i>Macaca mulatta</i>	rhesus macaque	<a href="#">Build 1.1</a>	
<i>Pan troglodytes</i>	chimpanzee	<a href="#">Build 2.1</a>	
Rodents (2)			
Scientific name	Common name	Build	Tools
<i>Mus musculus</i>	laboratory mouse	<a href="#">Build 37.1</a> <a href="#">Build 36.1</a>	
<i>Rattus norvegicus</i>	rat	<a href="#">RGSC v3.4</a>	
▶ Monotremes (1)			
▶ Marsupials (1)			
▶ Other Mammals (8)			
▶ Other Vertebrates (4)			
▶ Invertebrates (14)			
▶ Protozoa  (18)			
▶ Plants  (46)			
Fungi   (17)			
Scientific name	Common name	Build	Tools
<i>Aspergillus clavatus</i>		<a href="#">Build 1.1</a>	
<i>Aspergillus fumigatus</i>		<a href="#">Build 2.1</a>	
<i>Aspergillus niger</i>		<a href="#">Build 1.1</a>	
<i>Candida glabrata</i>		<a href="#">Build 1.1</a>	
<i>Cryptococcus neoformans</i>		<a href="#">Build 2.1</a>	
<i>Debaryomyces hansenii</i>		<a href="#">Build 1.1</a>	
<i>Encephalitozoon cuniculi</i>		<a href="#">Build 1.1</a>	
<i>Eremothecium gossypii</i>		<a href="#">Build 3.1</a>	
<i>Gibberella zeae</i>		<a href="#">Build 1.2</a>	
<i>Kluyveromyces lactis</i>		<a href="#">Build 1.1</a>	
<i>Magnaporthe oryzae</i>	rice blast fungus	<a href="#">Build 3.1</a> <a href="#">Build 2.1</a>	
<i>Neurospora crassa</i>		<a href="#">Build 2.1</a>	
<i>Saccharomyces cerevisiae</i>	baker's yeast	<a href="#">Build 2.1</a>	
<i>Scheffersomyces stipitis</i>		<a href="#">Build 1.1</a>	
<i>Schizosaccharomyces pombe</i>	fission yeast	<a href="#">Build 1.1</a>	
<i>Ustilago maydis</i>		<a href="#">Build 1.1</a>	
<i>Yarrowia lipolytica</i>		<a href="#">Build 1.1</a>	

# Map viewer Top画面

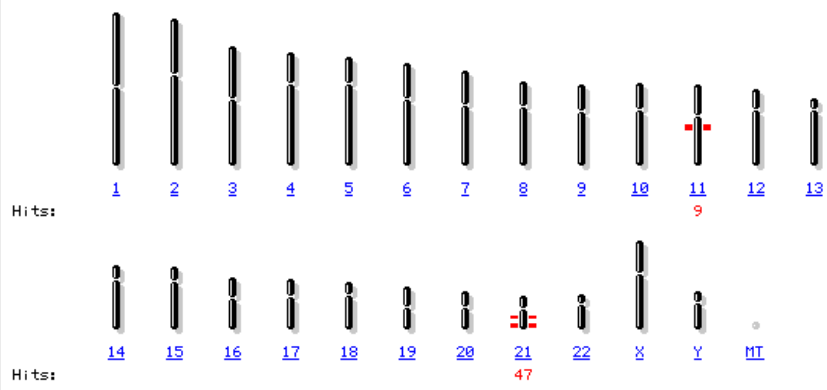
U2AF1  
と入力

The screenshot shows the NCBI Map Viewer interface. At the top, there are navigation tabs for PubMed, Nucleotide, Protein, Genome, Gene, Structure, PopSet, Taxonomy, and Help. Below these is a search bar with the text "Search for" and a dropdown menu set to "on chromosome(s)". A "Find" button and an "Advanced Search" link are also present. The main content area displays the "Homo sapiens (human) genome view" for "Build 37.1 statistics". It features a graphical ideogram of the human genome with chromosomes numbered 1 through 22, X, Y, and MT. Below the ideogram, there is a "Lineage" section with a tree structure: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo; Homo sapiens. A text block dated "August 2009" provides information about the updated version of the human genome reference assembly, GRCh37, and mentions the Genome Reference Consortium (GRC). Below this, there are links for "Release Notes", "Statistics", and "CCDS Project". The page also includes sections for "Organism Data in GenBank" (EST, Genomic, mRNA, Protein, WGS) and "Sequencing Projects" (Helix, Kazusa, MHC Project, NEDC). At the bottom, there is a section for "Available Documentation" and a note about viewing other genomes.

PubMed Nucleotide Protein Genome Gene Structure PopSet Taxonomy Help

Search for  on chromosome(s)  assembly  Find

**Homo sapiens (human) genome view**  
[Build 37.1 statistics](#) [Switch to previous build](#)



- All
- reference
- ALT\_REF\_LOCI\_1
- ALT\_REF\_LOCI\_2
- ALT\_REF\_LOCI\_3
- ALT\_REF\_LOCI\_4
- ALT\_REF\_LOCI\_5
- ALT\_REF\_LOCI\_6
- ALT\_REF\_LOCI\_7
- ALT\_REF\_LOCI\_8
- ALT\_REF\_LOCI\_9
- CRA\_TCAgchr7v2
- Celera
- HuRef

BLAST search the human genome

多すぎなので、referenceに絞る

Search results for query "U2AF1": 56 hits

Chr	Assembly	Match	Map element	Type	Maps
11	reference	<a href="#">all matches</a> Gallus gallus U2 small nuclear...U2AF1), mRNA Gallus gallus mRNA for U2snRNP...U2AF1 gene) Gallus gallus mRNA for U2snRNP...U2AF1 gene)	<a href="#">NM_204655.2</a> <a href="#">AJ291765.1</a> <a href="#">AJ291764.1</a>	TRANSCRIPT	<a href="#">Gga RNA</a> <a href="#">Gga RNA</a> <a href="#">Gga RNA</a>
11	Celera	<a href="#">all matches</a> Gallus gallus U2 small nuclear...U2AF1), mRNA Gallus gallus mRNA for U2snRNP...U2AF1 gene) Gallus gallus mRNA for U2snRNP...U2AF1 gene)	<a href="#">NM_204655.2</a> <a href="#">AJ291765.1</a> <a href="#">AJ291764.1</a>	TRANSCRIPT	<a href="#">Gga RNA</a> <a href="#">Gga RNA</a> <a href="#">Gga RNA</a>
11	HuRef	<a href="#">all matches</a> Gallus gallus U2 small nuclear...U2AF1), mRNA Gallus gallus mRNA for U2snRNP...U2AF1 gene) Gallus gallus mRNA for U2snRNP...U2AF1 gene)	<a href="#">NM_204655.2</a> <a href="#">AJ291765.1</a> <a href="#">AJ291764.1</a>	TRANSCRIPT	<a href="#">Gga RNA</a> <a href="#">Gga RNA</a> <a href="#">Gga RNA</a>
21	reference	<a href="#">all matches</a> ENST00000457176 ENST00000398137 ENST00000380276 ENST00000291552 Mus musculus U2 small nuclear...U2af1), mRNA Gallus gallus mRNA for U2snRNP...U2AF1 gene) Gallus gallus mRNA for U2snRNP...U2AF1 gene) Bos taurus U2 small nuclear...U2AF1), mRNA PDUts1019H11 Porcine testis...U2AF1), mRNA, mRNA sequence TC115911 Human adult whole...U2AF1), mRNA sequence Homo sapiens mRNA for U2 snRNP...U2AF1 gene), splice	<a href="#">ENST00000457176</a> <a href="#">ENST00000398137</a> <a href="#">ENST00000380276</a> <a href="#">ENST00000291552</a> <a href="#">NM_024187.3</a> <a href="#">AJ291765.1</a> <a href="#">AJ291763.1</a> <a href="#">NM_001080268.1</a> <a href="#">CV865602.1</a> <a href="#">DN994470.1</a> <a href="#">AJ627978.1</a>	TRANSCRIPT	<a href="#">ensRNA</a> <a href="#">ensRNA</a> <a href="#">ensRNA</a> <a href="#">ensRNA</a> <a href="#">Mm RNA</a> <a href="#">Gga RNA</a> <a href="#">Gga RNA</a> <a href="#">Bt RNA</a> <a href="#">Ssc RNA</a> <a href="#">Hs RNA</a> <a href="#">Hs RNA</a>

Quick Filter

Gene

Transcript :

all

RefSeq



# どうして、2か所？

Entrez Genome view | Map Viewer

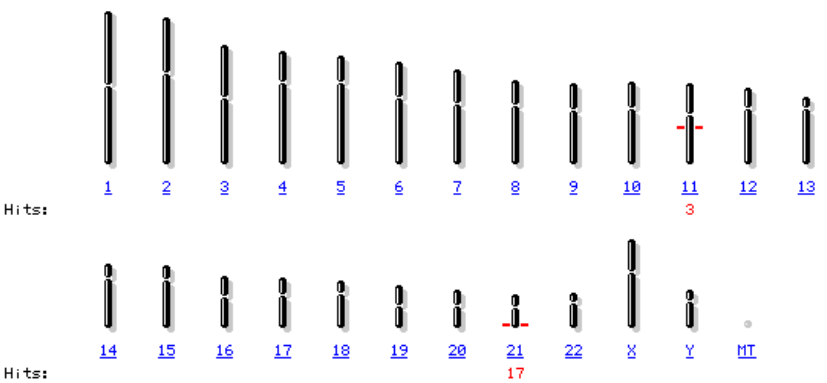
www.ncbi.nlm.nih.gov/projects/mapview/map\_search.cgi?taxid=9606&query=U2AF1&qchr=&st

NCBI NCBI Map Viewer

PubMed Nucleotide Protein Genome Gene Structure PopSet Taxonomy Help

Search for  on chromosome(s)  assembly  Find Advanced Search

Homo sapiens (human) genome view  
Build 37.1 statistics Switch to previous build



1) 略語の意味: Gga=Gallus gallus  
を予め知っておくと助けになる。

2) みると、ヒト11q13  
とニワトリのU2AF1  
がマップされている。

Search results for query "U2AF1": 20 hits (assembly reference)

Chr	Match	Map element	Type	Maps
11	<a href="#">all matches</a>			
	Gallus gallus U2 small nuclear...U2AF1), mRNA	<a href="#">NM_204655.2</a>	TRANSCRIPT	<a href="#">Gga_RNA</a>
	Gallus gallus mRNA for U2snRNP...U2AF1 gene)	<a href="#">AJ291765.1</a>	TRANSCRIPT	<a href="#">Gga_RNA</a>
	Gallus gallus mRNA for U2snRNP...U2AF1 gene)	<a href="#">AJ291764.1</a>	TRANSCRIPT	<a href="#">Gga_RNA</a>
21	<a href="#">all matches</a>			
	ENST00000457176	<a href="#">ENST00000457176</a>	TRANSCRIPT	<a href="#">ensRNA</a>
	ENST00000398137	<a href="#">ENST00000398137</a>	TRANSCRIPT	<a href="#">ensRNA</a>
	ENST00000380276	<a href="#">ENST00000380276</a>	TRANSCRIPT	<a href="#">ensRNA</a>
	ENST00000291552	<a href="#">ENST00000291552</a>	TRANSCRIPT	<a href="#">ensRNA</a>
	Mus musculus U2 small nuclear...U2af1), mRNA	<a href="#">NM_024187.3</a>	TRANSCRIPT	<a href="#">Mm_RNA</a>
	Gallus gallus mRNA for U2snRNP...U2AF1 gene)	<a href="#">AJ291765.1</a>	TRANSCRIPT	<a href="#">Gga_RNA</a>
	Gallus gallus mRNA for U2snRNP...U2AF1 gene)	<a href="#">AJ291763.1</a>	TRANSCRIPT	<a href="#">Gga_RNA</a>
	Bos taurus U2 small nuclear...U2AF1), mRNA	<a href="#">NM_001080268.1</a>	TRANSCRIPT	<a href="#">Bt_RNA</a>
	PDUts1019H11 Porcine testis...U2AF1), mRNA, mRNA sequence	<a href="#">CV865602.1</a>	TRANSCRIPT	<a href="#">Ssc_RNA</a>
	TC115911 Human adult whole...U2AF1), mRNA sequence	<a href="#">DN994470.1</a>	TRANSCRIPT	<a href="#">Hs_RNA</a>
	Homo sapiens mRNA for U2 snRNP...U2AF1 gene), splice variant...	<a href="#">AJ627978.1</a>	TRANSCRIPT	<a href="#">Hs_RNA</a>
	Homo sapiens U2 small nuclear...U2AF1), transcript variant...	<a href="#">NM_006758.2</a>	TRANSCRIPT	<a href="#">RefSeq_RNA</a>   <a href="#">Hs_RNA</a>
	Homo sapiens U2 small nuclear...U2AF1), transcript variant...	<a href="#">NM_001025204.1</a>	TRANSCRIPT	<a href="#">RefSeq_RNA</a>   <a href="#">Hs_RNA</a>
	Homo sapiens U2 small nuclear...U2AF1), transcript variant...	<a href="#">NM_001025203.1</a>	TRANSCRIPT	<a href="#">RefSeq_RNA</a>   <a href="#">Hs_RNA</a>
	Homo sapiens U2AF1 mRNA for U2 small nuclear RNA auxiliary...	<a href="#">AB451244.1</a>	TRANSCRIPT	<a href="#">Hs_RNA</a>
	U2AF1 : U2 small nuclear RNA auxiliary factor 1	<a href="#">U2AF1</a>	GENE	<a href="#">Genes cyto</a>   <a href="#">Genes seq</a>
	U2AF1 : ENSG00000160201	<a href="#">U2AF1</a>	GENE	<a href="#">ensGenes</a>

Quick Filter

Gene

Transcript :

all

RefSeq

Filter

Geneにcheckを入れた後に、ここをクリック

Map Viewer

www.ncbi.nlm.nih.gov/projects/mapview/maps.cgi?taxid=9606&chr=21&MAPS=gene,c,ensgenes,ugHs,genes-r&cmd=focus&fill=80&query=uid(1632)

NCBI NCBI Map Viewer

PubMed Entrez BLAST OMIM Taxonomy Structure

Search Find Find in This View Advanced Search

Human genome overview page (Build 37.1)  
Human genome overview page (Build 36.3)  
Map Viewer Home

Map Viewer Help  
Human Maps Help  
FTP  
Data As Table View  
Maps & Options  
Compress Map  
Region Shown: 44,480K  
44,812K Go

**Homo sapiens (human) Build 37.1 (Current)** [BLAST The Human Genome](#)

Chromosome: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 [ 21 ] 22 X Y MT

Query: U2AF1 AND gene[obj\_type] [clear]

Master Map: Genes On Sequence [Summary of Maps](#) [Maps & Options](#)

Region Displayed: 44,480K-44,812K bp [Download/View Sequence/Evidence](#)

Genes\_cyto → X ensGenes → X Hs UniG → X Genes\_seq → X

Symbol	Q	Links	E	C
CBS	↑	OMIMHGNC sv prd lev mm hm sts CCDS SNP	best RefSeq	21
U2AF1	↑	OMIMHGNC sv prd lev mm hm sts CCDS SNP	best RefSeq	21
MRPL51P2	↑	HGNC sv d lev mm	best RefSeq	21
LOC100288336	↑	sv d lev mm	mRNA	21
CRYAA	↓	OMIMHGNC sv prd lev mm hm sts CCDS SNP	best RefSeq	21
C21orf136	↑	HGNC sv prd lev mm	SNP mRNA	21
FLJ41733	↓	sv prd lev mm	mRNA	21

U2AF1 MRPL51P2 CBS  
LOC100288336 CRYAA  
C21orf136 FLJ41733  
SIK1  
C21orf125 C21orf84  
RPL31P1 H2BFS  
HSF2BP  
RRP1B LOC100288336  
PDXK  
CSTB RRP1  
LOC284837  
MYL6P  
AGPAT3  
H2AFZP TRAPPC10 PMP2  
C21orf33 LOC10012... HSCRM1  
SHND6 COLOT1 ACTL5 HPE1  
LOC54082 C21orf32 DGR  
PCBP30T C21orf135 NCRNA00163  
IGOSL6 DMNT3L AIRE PFKL  
C21orf2  
TRPM2  
LRRC3

44480K  
44490K  
44500K  
44510K  
44520K  
44530K  
44540K  
44550K  
44560K  
44570K  
44580K  
44590K  
44600K  
44610K  
44620K  
44630K  
44640K  
44650K  
44660K  
44670K  
44680K  
44690K  
44700K  
44710K  
44720K  
44730K  
44740K  
44750K  
44760K  
44770K  
44780K

Hs.533013  
Hs.692987  
Hs.365116  
Hs.591100  
Hs.687214  
Hs.629181  
Hs.676988  
Hs.561052  
Hs.570403  
Hs.354236  
Hs.707697  
Hs.184085  
Hs.473952  
Hs.551003  
Hs.517319  
Hs.517320

out  
zoom  
in

You are here:  
Ideogram  
21p13  
21p12  
21p11.2  
21p11.1  
21q11.1  
21q11.2  
21q21  
21q22

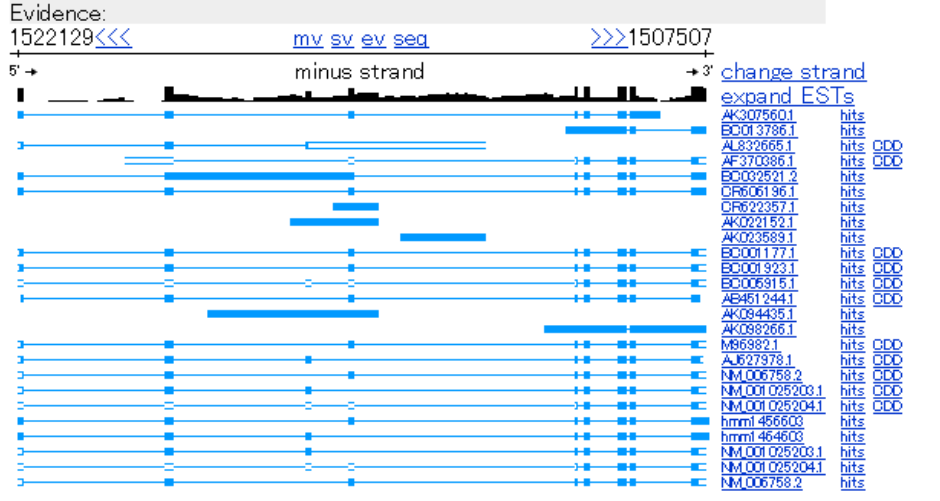
default  
master

NCBIの様々なtool群へのリンク  
今回はModel Maker (mm)を  
click

**Model Maker** (Make Your Own Model by selecting an evidence exon "set" and/or add/remove individual putative exons for inclusion in your model) [help](#) [legend](#)

ユーザーが自由にgene modelを組み立てるtool

Organism: Homo sapiens (Build 37.1) Chromosome: 21 Contig: [NT\\_011515.12](#) Locus: [U2AF1](#)



Your model: [clear](#)

[ORF Finder](#) [Save](#)

Frame1, ORF=  [CDD](#)  Frame2, ORF=  [CDD](#)  Frame3, ORF=  [CDD](#)

Putative exons (table view): [custom exons](#) [intron bases:](#)

<input type="checkbox"/>	1	1522143	GG AAG...TC CGG	1522126-1522046	GAA AT =>	2 or 4
<input type="checkbox"/>	2		1 <= GG AAG	1522143-1522002	CAA GT =>	5
<input type="checkbox"/>	3	1522129	GC GTC...GT CGA	1522073-1522002	CAA GT =>	7 or 8
<input type="checkbox"/>	4		1 <= TC CGG	1522126-1522002	CAA GT =>	5
<input type="checkbox"/>	5		2 or 4 <= AA ATG	1522045-1522002	CAA GT =>	7
<input type="checkbox"/>	6		TT CCC	1518803-1518866	CAG GT =>	12
<input type="checkbox"/>	7		3 or 5 <= AG AGT	1518953-1518866	CAG GT =>	10 or 11 or 12
<input type="checkbox"/>	8		3 <= AG AGT	1518953-1515004	GCT GT =>	15
<input type="checkbox"/>	9	1518037	TC TTT...TT TCC	1515405-1514524	TCC AG...TCC AA	1514493
<input type="checkbox"/>	10		7 <= AG ACC	1515983-1515917	ACT GT =>	12 or 15
<input type="checkbox"/>	11		7 <= AG ACC	1515983-1512191	CTC AA	
<input type="checkbox"/>	12		6 or 7 or 10 <= AG ACC	1515070-1515004	GCT GT =>	15
<input type="checkbox"/>	13		TT AGA	1513959-1512191	CTC AA	

Model Maker | Map Viewer | Map Viewer

www.ncbi.nlm.nih.gov/projects/mapview/modelmaker.cgi?taxid=9606&contig=NT\_011515.12&gene=U2AF1

Putative exons (graphic view):

Your model: 3-7-10-12-15-16

ORF Finder

Frame1, ORF=75 CDD Frame2, ORF=44 CDD Frame3, ORF=25 CDD

Putative exons (table view):

Exon	Coordinates	Sequence	Reading Frames
<input type="checkbox"/>	1 1522143	GG AAG...TC CGG	1522126-1522046 GAA AT => 2 or 4
<input type="checkbox"/>	2	1 <= GG AAG	1522143-1522002 CAA GT => 5
<input checked="" type="checkbox"/>	3 1522129	GC GTC...GT CGA	1522073-1522002 CAA GT => 7 or 8
<input type="checkbox"/>	4	1 <= TC CGG	1522126-1522002 CAA GT => 5
<input type="checkbox"/>	5	2 or 4 <= AA ATG	1522045-1522002 CAA GT => 7
<input type="checkbox"/>	6	TT CCC	1519803-1518866 CAG GT => 12
<input checked="" type="checkbox"/>	7	3 or 5 <= AG AGT	1518953-1518866 CAG GT => 10 or 11 or 12
<input type="checkbox"/>	8	3 <= AG AGT	1518953-1515004 GCT GT => 15
<input type="checkbox"/>	9 1518037	TC TTT...TT TCC	1515405-1514524 TCC AG...TCC AA 1514493
<input checked="" type="checkbox"/>	10	7 <= AG ACC	1515983-1515917 ACT GT => 12 or 15
<input type="checkbox"/>	11	7 <= AG ACC	1515983-1512191 CTC AA
<input checked="" type="checkbox"/>	12	6 or 7 or 10 <= AG ACC	1515070-1515004 GCT GT => 15
<input type="checkbox"/>	13	TT AGA	1513959-1512191 CTC AA
<input type="checkbox"/>	14 1510941	AA TTT...TG GGG	1510494-1509206 GGG GT => 18 or 19
<input checked="" type="checkbox"/>	15	8 or 10 or 12 <= AG GTG	1510294-1510245 GAG GT => 16
<input checked="" type="checkbox"/>	16	15 <= AG GAG	1510087-1509989 AAG GT => 17
<input type="checkbox"/>	17	16 <= AG TTT	1509339-1509206 GGG GT => 18 or 19
<input type="checkbox"/>	18	14 or 17 <= AG AGA	1509114-1509022 GAA GT => 20 or 21
<input type="checkbox"/>	19	14 or 17 <= AG AGA	1509114-1508519 GCA GG...GAT AA 1507514
<input type="checkbox"/>	20	18 <= AG GCA	1507800-1507853 TGA GC => 22
<input type="checkbox"/>	21	18 <= AG GCA	1507800-1507508 AGT A TT 1507507
<input type="checkbox"/>	22	20 <= AG GCA	1507800-1507505 ATT CT => 23
<input type="checkbox"/>	23	22 <= GA GCC	1507852-1507505 ATT CT

Disclaimer | Write to the Help Desk  
NCBI | NLM | NIH

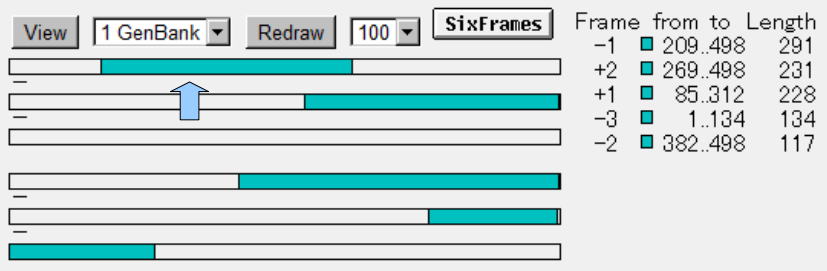
gene modelの構築:  
転写物のevidenceとエクソンの組み合わせによるフレームの合致などの制限の中で、ユーザーが選ぶことができる。



# ORF Finder (Open Reading Frame Finder)

PubMed Entrez BLAST OMIM Taxonomy Structure

Anonymous

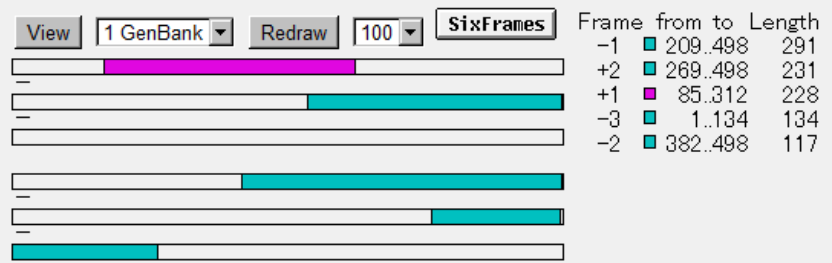


# NCBI ORF Finder (Open Reading Frame Finder)

PubMed Entrez BLAST OMIM Taxonomy Structure

## Anonymous

Program:  Database:  BLAST  with parameters



Length: 75 aa

Accept Alternative Initiation Codons

```

85 atgxcgkxktatctgxcotocattcttgcaccgkgaagacaaa
M A E Y L A S I F G T E R D K
130 gtcaactgttcattttattcaaaattgkagcatgctcatgkka
V N C S F Y F K I G A C R H G
175 gacagxtgctctcgttgacacataaaocagtttagccagacc
D R C S R L H N K P T F S Q T
220 atcttgattcaaaacatctatctaatccccaaaacagtgacag
I L I Q N I Y R N P Q N S A Q
265 acgctgacgctcacaactaccattgxcctcttgaacatttaocg
T A D G S H Y H C P L E H L P
310 *aa 312
*

```

最上段をクリックすることによって、アミノ酸配列が表示された。

# Ensembl (www.ensembl.org)



Login | Register | BLAST/BLAT | BioMart | Tools | Downloads | Help | Documentation | Mirrors

Home

Search:  for



e.g. [human gene BRCA2](#) or [rat X:100000..200000](#) or [coronary heart disease](#)

## Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Click on a link below to go to the species' home page.

**Popular genomes** ([Log in to customize this list](#))



**Human**  
GRCh37



**Mouse**  
NCBIM37



**Zebrafish**  
Zv8

## All genomes

[View full list of all Ensembl species](#)

Other species are available in [Ensembl Pre!](#) and [EnsemblGenomes](#)

## New to Ensembl?

Did you know you can:

- [Learn how to use Ensembl](#)  
with our video tutorials and walk-throughs
- [Add custom tracks](#)  
using our new Control Panel
- [Upload and analyse your data](#)  
and save it to your Ensembl account
- [Search for a DNA or protein sequence](#)  
using BLAST or BLAT
- [Fetch only the data you want](#)  
from our public database, using the Perl API
- [Download our databases via FTP](#)  
in FASTA, MySQL and other formats
- [Mine Ensembl with BioMart](#) ←  
and export sequences or tables in text, html, or Excel format

Still got questions? Try our [FAQs](#) or [glossary](#)

## What's New in Release 59 (5 August 2010)

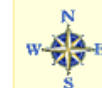
- [New human regulation data and displays](#) (Human)
- [COSMIC somatic mutations](#) (Human)
- [Changes to masthead](#) (all species)
- [Details of data updates, API changes, etc](#)

[More news...](#)

## Latest blog posts

- 2010-09-24: [Ensembl Events in October 2010](#)
- 2010-09-21: [Changing to LASTZ](#)
- 2010-09-09: [A.mel vs A.mel](#)

[Go to Ensembl blog →](#)



[Navigation tips](#)  
from our blog



Ensembl is a joint project between [EMBL - EBI](#) and the [Wellcome Trust Sanger Institute](#) to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes.



Ensembl receives major funding from the Wellcome Trust. Our [acknowledgements page](#) includes a list of additional current and previous funding bodies.

# Species List in Ensembl

Ensembl Genome Bro... x Species List x Ensembl genome bro... x FTP Download x Ensembl Genome Bro... x

uswest.ensembl.org/info/about/species.html


























## Find a Species

The main Ensembl site focuses on vertebrate genomes - [scroll down](#) for links to our sister sites covering invertebrates, plants, bacteria, etc.

[Species tree \(Requires Java\).](#)

**Ensembl Species**

このページでは51種。Ensemble genome (bacteria, protists, fungi, plants and invertebrate metazoa)では234種

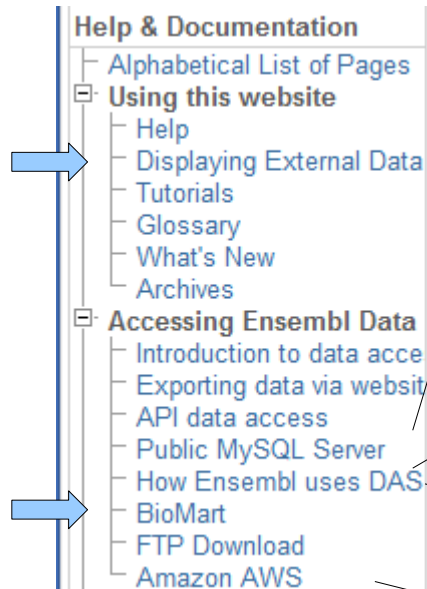
 <b>Alpaca</b> <i>Vicugna pacos</i>	 <b>Guinea Pig</b> <i>Cavia porcellus</i>	 <b>Pig</b> <i>Sus scrofa</i>
 <b>Anole Lizard</b> <i>Anolis carolinensis</i>	 <b>Hedgehog</b> <i>Erinaceus europaeus</i>	 <b>Pika</b> <i>Ochotona princeps</i>
 <b>Armadillo</b> <i>Dasypus novemcinctus</i>	 <b>Horse</b> <i>Equus caballus</i>	 <b>Platypus</b> <i>Ornithorhynchus anatinus</i>
 <b>Bushbaby</b> <i>Otolemur gamettii</i>	 <b>Human</b> <i>Homo sapiens</i>	 <b>Rabbit</b> <i>Oryctolagus cuniculus</i>
 <b>Caenorhabditis elegans</b>	 <b>Hyrax</b> <i>Procavia capensis</i>	 <b>Rat</b> <i>Rattus norvegicus</i>
 <b>Ciona intestinalis</b>	 <b>Kangaroo rat</b> <i>Dipodomys ordii</i>	 <b>Saccharomyces cerevisiae</b> <i>Saccharomyces cerevisiae</i>
 <b>Ciona savignyi</b>	 <b>Lamprey</b> ( <a href="#">preview - assembly only</a> ) <i>Petromyzon marinus</i>	 <b>Shrew</b> <i>Sorex araneus</i>
 <b>Cat</b> <i>Felis catus</i>	 <b>Lesser hedgehog tenrec</b> <i>Echinops telfairi</i>	 <b>Sloth</b> <i>Choloepus hoffmanni</i>
 <b>Chicken</b> <i>Gallus gallus</i>	 <b>Macaque</b> <i>Macaca mulatta</i>	 <b>Squirrel</b> <i>Spermophilus tridecemlineatus</i>

**Help & Documentation**

- Alphabetical List of Pages
- Using this website
  - Help
  - Displaying External Data
  - Tutorials
  - Glossary
  - What's New
  - Archives
- Accessing Ensembl Data
  - Introduction to data acce
  - Exporting data via websit
  - API data access
  - Public MySQL Server
  - How Ensembl uses DAS
  - BioMart
  - FTP Download
  - Amazon AWS
- Ensembl Documentation
  - Gene Annotation
  - Microarray Probeset Map
  - Variation
  - Comparative Genomics
  - Regulatory Build
  - eHive production system
  - API Documentation
  - DAS (Distributed Annotat
  - Web code
  - Known Bugs
- About Ensembl
  - About the Ensembl Proje
  - Species List**
  - Release Cycle
  - Mirror sites
  - Scientific Publications
  - Outreach
  - Contact Us
  - Job Vacancies

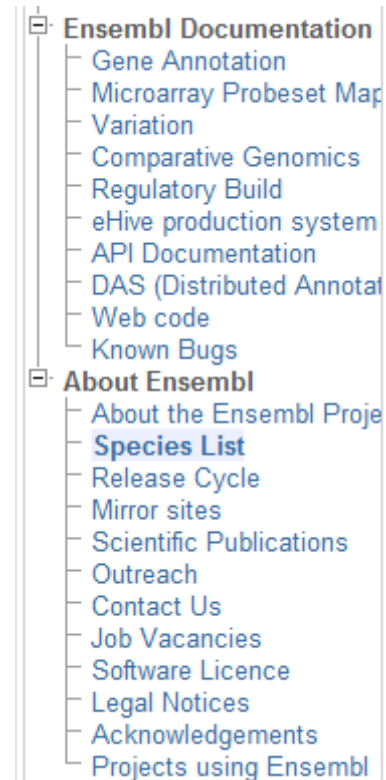


# Ensemblのdocument一覧



- Help & Documentation
  - Alphabetical List of Pages
  - Using this website
    - Help
    - Displaying External Data
    - Tutorials
    - Glossary
    - What's New
    - Archives
  - Accessing Ensembl Data
    - Introduction to data access
    - Exporting data via website
    - API data access
    - Public MySQL Server
    - How Ensembl uses DAS
    - BioMart
    - FTP Download
    - Amazon AWS

- API: Application Programming Interface webアプリケーションの機能の一部を外部のプログラムから利用できるようにするための手順や書式を定めた仕様。
- EnsemblのMySQLサーバーにアクセス。
- DAS: Distributed Annotation System
- AmazonのCloud (EC2) serviceを介したデータの共有。



- Ensembl Documentation
  - Gene Annotation
  - Microarray Probeset Mapping
  - Variation
  - Comparative Genomics
  - Regulatory Build
  - eHive production system
  - API Documentation
  - DAS (Distributed Annotation System)
  - Web code
  - Known Bugs
- About Ensembl
  - About the Ensembl Project
  - Species List
  - Release Cycle
  - Mirror sites
  - Scientific Publications
  - Outreach
  - Contact Us
  - Job Vacancies
  - Software Licence
  - Legal Notices
  - Acknowledgements
  - Projects using Ensembl

# Ensembl: genome browserの例

The screenshot displays the Ensembl genome browser interface. The browser tabs include 'Ensembl gen...', 'Species List', and 'FTP Download'. The address bar shows the URL: `uswest.ensembl.org/Homo_sapiens/Location/View?db=core;g=ENSG00000160201;r=21:44513066-44527697`.

**Region in detail**

Chromosome 21 bands: p13, p11.2, q11.2, q21.1, q21.2, q21.3, q22.11, q22.2, q22.3. The q22.3 band is highlighted with a red box.

« Region overview | **Region in detail** help | Alignments (image) »

1.00 Mb scale bar: 44.10 Mb to 44.90 Mb. Forward strand.

Chromosome bands: q22.3

Contigs: AP001628.1 >, AP001630.1 >, AP001046.1 >

Ensembl/Havana g...: AP001626.1, AP001627.1, WDR4, C21orf105, CBS, U2AF1, AP001631.10, SKI, C21orf84, RPL31, PDE9A, AP001628.7, NDUFV3, PKNOX1, CRYAA, AP001046.4, C21orf125, H2BF, AP001626.2, AP001628.6, AP001630.5, MRPL5IP2, AP001631.9, AP001046.5, AP001046.6, HSF2BP, AP001048.4

ncRNA gene: AP001629.1, AP001630.1

All Structural varia...: (empty)

Gene Legend: processed transcript (blue), pseudogene (grey), merged Ensembl/Havana (yellow), RNA gene (purple)

Ensembl Homo sapiens version 59.37d (GRCh37) Chromosome 21: 44,020,382 - 45,020,381

Export Image

Location: 21 : 44513066 - 44527697 Go > [Navigation controls]

**14.63 Kb** scale bar: 44,514,000 to 44,526,000. Forward strand.

Chromosome bands: q22.3

Contigs: AP001631.1 >

Ensembl/Havana g...: < U2AF1-001 protein coding, < U2AF1-002 protein coding, < U2AF1-003

# BioMart top page

The screenshot shows the BioMart web interface. The browser address bar displays the URL: `uswest.ensembl.org/bio mart/martview/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7`. The page header includes the Ensembl logo and navigation links: Login, Register, BLAST/BLAT, BioMart, Tools, Downloads, Help, Documentation, and Mirrors. Below the header is a toolbar with buttons for New, Count, Results, URL, XML, Perl, and Help. The main content area is divided into a left sidebar and a main panel. The sidebar contains sections for Dataset (Homo sapiens genes (GRCh37)), Filters ([None selected]), and Attributes (Ensembl Gene ID, Ensembl Transcript ID). The main panel features two dropdown menus: the top one is set to 'Ensembl Genes 59' and the bottom one to 'Homo sapiens genes (GRCh37)'. A blue callout bubble points to the 'Filters' section in the sidebar with the text '2) Filtersをクリック'. Another blue callout bubble points to the top dropdown menu with the text '1) DB, speciesを選択する。'. The footer contains the text 'Ensembl release 59 - Jul 2010 © WTSI / EBI' and links for 'About Ensembl', 'Contact Us', and 'Help'. A 'Permanent link - View in archive site' link is also present.

1) DB, speciesを選択する。

2) Filtersをクリック

Ensembl release 59 - Jul 2010 © WTSI / EBI

[About Ensembl](#) | [Contact Us](#) | [Help](#)

[Permanent link - View in archive site](#)

# BioMart:Filters;欲しい情報に絞り込む1

The screenshot shows the Ensembl BioMart interface. The browser address bar displays the URL: [uswest.ensembl.org/biomart/martview/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7](http://uswest.ensembl.org/biomart/martview/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7). The Ensembl logo is visible in the top left, and navigation links for Login, Register, BLAST/BLAT, BioMart, Tools, Downloads, Help, Documentation, and Mirrors are in the top right. Below the navigation bar, there are buttons for New, Count, Results, URL, XML, Perl, and Help. The main content area is titled "Please restrict your query using criteria below" and contains several filter categories: REGION, GENE, GENE ONTOLOGY, EXPRESSION, MULTI SPECIES COMPARISONS, and PROTEIN DOMAINS. On the left side, there is a sidebar with "Dataset" (Homo sapiens genes (GRCh37)), "Filters" ([None selected]), "Attributes" (Ensembl Gene ID, Ensembl Transcript ID), and another "Dataset" ([None Selected]). An orange arrow points to the "Attributes" section in the sidebar. At the bottom, there is a footer with "Ensembl release 59 - Jul 2010 © WTSI / EBI" and "About Ensembl | Contact Us | Help". A "Permanent link - View in archive site" link is also present.

Dataset  
Homo sapiens genes (GRCh37)

Filters  
[None selected]

Attributes  
Ensembl Gene ID  
Ensembl Transcript ID

Dataset  
[None Selected]

Please restrict your query using criteria below

- REGION:
- GENE:
- GENE ONTOLOGY:
- EXPRESSION:
- MULTI SPECIES COMPARISONS:
- PROTEIN DOMAINS:

Ensembl release 59 - Jul 2010 © [WTSI](#) / [EBI](#) [About Ensembl](#) | [Contact Us](#) | [Help](#)

[Permanent link - View in archive site](#)

# BioMart:Filters;欲しい情報に絞り込む2

uswest.ensembl.org/bio mart/martview/7dd7eca1 084622d92c81 cd39cc0395a7/7dd7eca1 084622d92c81 cd39cc0395a7

Ensembl West Login | Register | BLAST/BLAT | BioMart | Tools | Downloads | Help | Documentation | Mirrors

New Count Results URL XML Perl Help

Please restrict your query using criteria below

**1) 絞り込む**

**2) Countをクリック**

**3) Attributesをクリック**

Dataset: Homo sapiens genes (GRCh37)

Filters: GO Biological Process Term Name [e.g. regulation of biological process]: RNA splicing

Attributes: Ensembl Gene ID, Ensembl Transcript ID

Dataset: [None Selected]

REGION:

GENE:

GENE ONTOLOGY:

Evidence code (GO Biological process) IC IDA IEA IEP IGI

GO Biological Process Term Accession [e.g. GO:0050789]

GO Biological Process Term Name [e.g. regulation of biological process]: RNA splicing

Evidence code (GO Cellular component) IC IDA IEA IEP IMP

GO Cellular Component Term Accession [e.g. GO:0005622]

Ensembl release 59 - Jul 2010 © WTSI / EBI

About Ensembl | Contact Us | Help

Permanent link - View in archive site

# BioMart:Attributes;出力したい項目に絞り込む

The screenshot shows the BioMart web interface. The browser address bar displays the URL: [uswest.ensembl.org/biomart/martview/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7](http://uswest.ensembl.org/biomart/martview/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7). The page title is "Untitled document".

The main navigation bar includes links for "New", "Count", "Results", "URL", "XML", "Perl", and "Help". A yellow callout bubble points to the "Results" button with the text "Resultsをクリック".

The left sidebar shows the "Dataset 309 / 51737 Genes" section, with "Homo sapiens genes (GRCh37)" selected. The "Attributes" section is highlighted with a yellow box and a red bracket labeled "3)", listing various attributes such as "Ensembl Gene ID", "Ensembl Transcript ID", "dS", "dN", "Yeast Ensembl Gene ID", "Chimp Ensembl Gene ID", "Chimp Ensembl Protein ID", "dS", "dN", "% Identity", and "Chimp % Identity".

The main content area displays the instruction: "Please select columns to be included in the output and hit 'Results' when ready". Below this, there are several sections for selecting columns:

- 1)** A group of radio buttons for "Features", "Structures", "Transcript Event", "Homologs", "Variations", and "Sequences".
- 2)** A group of checkboxes for "GENE:", "ORTHOLOGS:", "POSSIBLE ORTHOLOGS:", and "PARALOGS:".

# BioMart: Result表示例

HTML  
TSV(タブ区切り)  
CSV(コンマ区切り)  
XML

uswest.ensembl.org/bio mart/ martview/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7

Ensembl Login Register BLAST/BLAT BioMart Tools Downloads Help Documentation Mirrors

New Count Results URL XML Perl Help

Export all results to    Unique results only

Email notification to

View  rows as   Unique results only

Ensembl Gene ID	Ensembl Transcript ID	Yeast Ensembl Gene ID	dS	dN	Chimp Ensembl Gene ID	dS	dN	Chimp Ensembl Protein ID	% Identity	Chimp % Identity
<a href="#">ENSG00000215699</a>	<a href="#">ENST00000374449</a>				<a href="#">ENSPTRG00000024234</a>	0.01220	0.00000	<a href="#">ENSPTRP00000052680</a>	100	100
<a href="#">ENSG00000215699</a>	<a href="#">ENST00000374449</a>				<a href="#">ENSPTRG00000024234</a>	0.01220	0.00000	<a href="#">ENSPTRP00000052680</a>	100	100
<a href="#">ENSG00000215699</a>	<a href="#">ENST00000374449</a>				<a href="#">ENSPTRG00000024234</a>	0.01220	0.00000	<a href="#">ENSPTRP00000052680</a>	100	100
<a href="#">ENSG00000215699</a>	<a href="#">ENST00000374449</a>				<a href="#">ENSPTRG00000024234</a>	0.01220	0.00000	<a href="#">ENSPTRP00000052680</a>	100	100
<a href="#">ENSG00000215699</a>	<a href="#">ENST00000374449</a>				<a href="#">ENSPTRG00000024234</a>	0.01220	0.00000	<a href="#">ENSPTRP00000052680</a>	100	100
<a href="#">ENSG00000215699</a>	<a href="#">ENST00000374449</a>				<a href="#">ENSPTRG00000024234</a>	0.01220	0.00000	<a href="#">ENSPTRP00000052680</a>	100	100
<a href="#">ENSG00000215699</a>	<a href="#">ENST00000374449</a>				<a href="#">ENSPTRG00000024234</a>	0.01220	0.00000	<a href="#">ENSPTRP00000052680</a>	100	100
<a href="#">ENSG00000169217</a>	<a href="#">ENST00000305596</a>				<a href="#">ENSPTRG00000007995</a>	0.01000	0.00910	<a href="#">ENSPTRP00000013666</a>	83	97
<a href="#">ENSG00000169217</a>	<a href="#">ENST00000305596</a>				<a href="#">ENSPTRG00000007995</a>	0.01000	0.00910	<a href="#">ENSPTRP00000013666</a>	83	97
<a href="#">ENSG00000169217</a>	<a href="#">ENST00000305596</a>				<a href="#">ENSPTRG00000007995</a>	0.01000	0.00910	<a href="#">ENSPTRP00000013666</a>	83	97
<a href="#">ENSG00000169217</a>	<a href="#">ENST00000305596</a>				<a href="#">ENSPTRG00000007995</a>	0.01000	0.00910	<a href="#">ENSPTRP00000013666</a>	83	97
<a href="#">ENSG00000169217</a>	<a href="#">ENST00000305596</a>				<a href="#">ENSPTRG00000007995</a>	0.01000	0.00910	<a href="#">ENSPTRP00000013666</a>	83	97
<a href="#">ENSG00000169217</a>	<a href="#">ENST00000305596</a>				<a href="#">ENSPTRG00000007995</a>	0.01000	0.00910	<a href="#">ENSPTRP00000013666</a>	83	97
<a href="#">ENSG00000101161</a>	<a href="#">ENST00000266079</a>	<a href="#">YBR055C</a>			<a href="#">ENSPTRG00000013765</a>	0.01680	0.00000	<a href="#">ENSPTRP00000023718</a>	99	100
<a href="#">ENSG00000101161</a>	<a href="#">ENST00000266079</a>	<a href="#">YBR055C</a>			<a href="#">ENSPTRG00000013765</a>	0.01680	0.00000	<a href="#">ENSPTRP00000023718</a>	99	100
<a href="#">ENSG00000101161</a>	<a href="#">ENST00000266079</a>	<a href="#">YBR055C</a>			<a href="#">ENSPTRG00000013765</a>	0.01680	0.00000	<a href="#">ENSPTRP00000023718</a>	99	100
<a href="#">ENSG00000101161</a>	<a href="#">ENST00000266079</a>	<a href="#">YBR055C</a>			<a href="#">ENSPTRG00000013765</a>	0.01680	0.00000	<a href="#">ENSPTRP00000023718</a>	99	100
<a href="#">ENSG00000101161</a>	<a href="#">ENST00000266079</a>	<a href="#">YBR055C</a>			<a href="#">ENSPTRG00000013765</a>	0.01680	0.00000	<a href="#">ENSPTRP00000023718</a>	99	100

Ensembl release 59 - Jul 2010 © WTSI / EBI

About Ensembl | Contact Us | Help

Permanent link - View in archive site

# さらなるdataset追加

Dataset 309 / 51737 Genes  
Homo sapiens genes (GRCh37)

**Filters**  
GO Biological Process Term  
Name [e.g. regulation of biological process] : RNA splicing

**Attributes**  
Ensembl Gene ID  
Ensembl Transcript ID  
Reference ID  
Transcript location (bp)  
SNP Chromosome Strand  
Protein location (aa)  
Protein Allele  
Validation status

**Dataset**  
[None Selected]

- CHOOSE ADDITIONAL DATASET -  
- CHOOSE ADDITIONAL DATASET -  
[Reactome] complex  
[Reactome] pathway  
[Reactome] reaction  
[PRIDE (EBI UK)] PRIDE

[Ensembl Genes 59] Anolis carolinensis genes (AnoCar1.0)  
[Ensembl Genes 59] Bos taurus genes (Btau.4.0)  
[Ensembl Genes 59] Caenorhabditis elegans genes (WS210)  
[Ensembl Genes 59] Callithrix jacchus genes (calJac3)  
[Ensembl Genes 59] Canis familiaris genes (CanFam.2.0)  
[Ensembl Genes 59] Cavia porcellus genes (cavPor3)  
[Ensembl Genes 59] Choloepus hoffmanni genes (choHof1)  
[Ensembl Genes 59] Ciona intestinalis genes (JGI2)  
[Ensembl Genes 59] Ciona savignyi genes (CSAV2.0)  
[Ensembl Genes 59] Danio rerio genes (Zv8)  
[Ensembl Genes 59] Dasyurus novemcinctus genes (dasNov2)  
[Ensembl Genes 59] Dipodomys ordii genes (dipOrd1)  
[Ensembl Genes 59] Drosophila melanogaster genes (BDGP5.13)  
[Ensembl Genes 59] Echinops telfairi genes (TENREC)

Dataset: さらに追加したいdataset  
(今回はこれ以上追加しない)



# UCSC Genome Browser genome.ucsu.edu



表形式で  
BioMartのよ  
うに、欲しい  
項目に絞り込  
んだ上で、表  
示、ダウン  
ロード

遺伝子を任意の  
関係(発現、相同  
性、位置など)で  
並び変えて、表  
示、ダウンロード

遺伝子発現部位  
の顕微鏡像  
マウスやアフリカ  
ツメガエルのプロ  
ジェクトより

## News

News Archives ▶

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the [genome-announce](#) mailing list.

### 24 August 2010 - New Drag and Reorder Functionality Released

We are pleased to announce new functionality in the Genome Browser. It is now possible to rearrange the order that tracks appear in the browser image directly from the browser image itself. To reorder tracks, click-and-hold the side label or gray mini-button of a single track and drag the highlighted track to a new position within the image.

Because reordering is now available in the browser image itself, we have removed the "enable track reordering" option from the configuration page.

Thanks to Tim Dreszer and the rest of the team for implementing and testing this feature.

**20 August 2010 - New ENCODE Integrated Regulation Super-track Released:** We are pleased to announce the release of the ENCODE Integrated Regulation super-track. [Read more.](#)

**18 August 2010 - Cat Genome Browser Available:** We have released a Genome Browser for the latest assembly of Cat (*Felis catus*). [Read more.](#)

Conditions of Use

# UCSC Genome Browser genome.uc su.edu

genome.ucsc.edu/cgi-bin/hgTracks

Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session Help

## UCSC Genome Browser on Human Mar. 2006 (NCBI36/hg18) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chrX:50,582,451-50,593,552 gene jump clear size 11,102 bp. configure

chrX (p11.22) 22.2 q21.1 22.3 q23 24 Xq25 Xq28

Scale 5 kb

chrX: 50584000 50585000 50586000 50587000 50588000 50589000 50590000 50591000 50592000 50593000

UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics

RefSeq Genes

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

Spliced ESTs

50

Enhancer H3K4Me1

0

100

Promoter H3K4Me3

0

ENCODE Enhancer and Promoter Histone Mark (H3K4Me1) on 8 Cell Lines

ENCODE Promoter Histone Mark (H3K4Me3) on 9 Cell Lines

DNase Clusters

ENCODE Digital DNaseI Hypersensitivity Clusters

Txn Factor ChIP

3

ENCODE Transcription Factor ChIP-seq

Mammal Cons

-0.5

Placental Mammal Basewise Conservation by PhyloP

Multiz Alignments of 44 Vertebrates

Rhesus

Mouse

Dog

Elephant

Opossum

Platypus

Chicken

Lizard

X\_tropicalis

Stickleback

Se1 Sup Scan (S)

Selective Sweep Scan (S) on Neandertal vs. Human Polymorphisms (Z-score +/- Variance)

5% Lowest S

Selective Sweep Scan (S): 5% Smallest S scores

S SNPs

SNPs Used for Selective Sweep Scan (S)

All

Neandertal Sequence Contigs Generated by Genotype Caller

Neandertal Sequence Reads

Fe1d1 Sequence

Mez1 Sequence

Sid1253 Sequence

Vi33.16 Sequence

Vi33.25 Sequence

Vi33.26 Sequence

Alignments of Sequence Reads from 5 Modern Humans

San Seq

Yoruba Seq

Han Seq

Papuan Seq

French Seq

SNPs (130)

Simple Nucleotide Polymorphisms (dbSNP build 130)

RepeatMasker

Repeating Elements by RepeatMasker

move start < 2.0 > Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. move end < 2.0 >

# UCSC

## GenomeBrowser で見られる項目

Configure Tracks on UCSC Genome Browser: Human Mar. 2006 (NCBI36/hg18)

Tracks:    Groups:

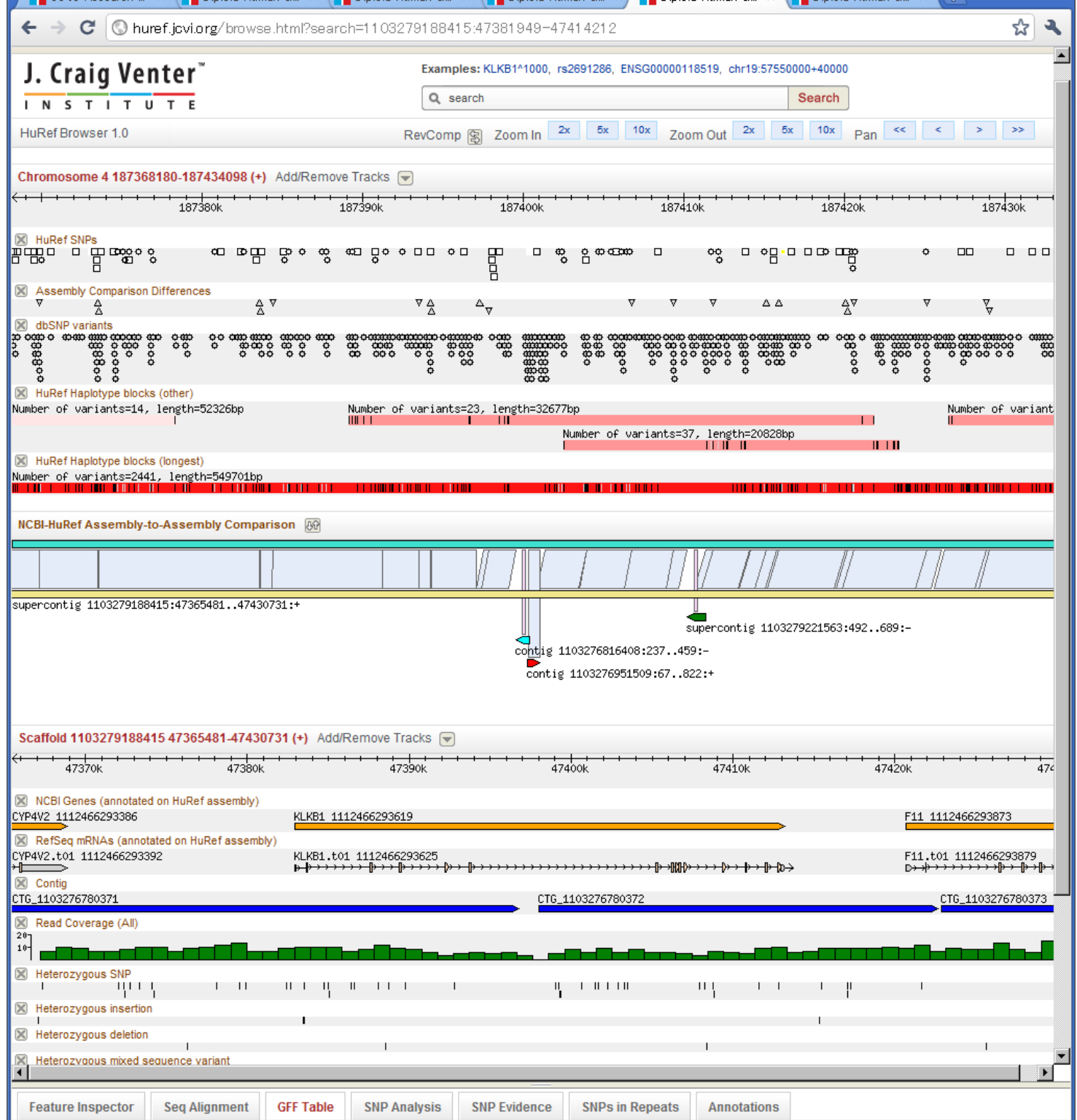
Control track and group visibility more selectively below.

- Mapping and Sequencing Tracks
- Phenotype and Disease Associations
- Genes and Gene Prediction Tracks
- mRNA and EST Tracks
- Expression
- Regulation
- Comparative Genomics
- Neandertal Assembly and Analysis
- Variation and Repeats
- Pilot ENCODE Regions and Genes
- Pilot ENCODE Transcription
- Pilot ENCODE Chromatin Immunoprecipitation
- Pilot ENCODE Chromatin Structure
- Pilot ENCODE Comparative Genomics and Variation

ヒトではhg18が最も  
項目数が多い

# JCVI HuRef Browser huref.jcvi.org/

INDIVIDUAL HUMAN  
DIPLOID GENOME



# genome versionとは

<http://genome.ucsc.edu/FAQ/FAQreleases.html>

List of UCSC genome releases

## Question:

"How do UCSC's release numbers correspond to those of other organizations, such as NCBI?"

## Response:

Ensemblも  
GRChを使用

SPECIES	UCSC VERSION	RELEASE DATE	RELEASE NAME	STATUS
<b>VERTEBRATES</b>				
Human	hg19	Feb. 2009	Genome Reference Consortium GRCh37	Available
	hg18	Mar. 2006	NCBI Build 36.1	Available
	hg17	May 2004	NCBI Build 35	Available
	hg16	Jul. 2003	NCBI Build 34	Available
	hg15	Apr. 2003	NCBI Build 33	Archived
	hg13	Nov. 2002	NCBI Build 31	Archived
	hg12	Jun. 2002	NCBI Build 30	Archived
	hg11	Apr. 2002	NCBI Build 29	Archived
	hg10	Dec. 2001	NCBI Build 28	Archived
	hg8	Aug. 2001	UCSC-assembled	Archived
	hg7	Apr. 2001	UCSC-assembled	Archived
	hg6	Dec. 2000	UCSC-assembled	Archived
	hg5	Oct. 2000	UCSC-assembled	Archived
	hg4	Sep. 2000	UCSC-assembled	Archived
	hg3	Jul. 2000	UCSC-assembled	Archived
	hg2	Jun. 2000	UCSC-assembled	Archived (data set only)
	hg1	May 2000	UCSC-assembled	Archived (data set only)
Cat	felCat4	Dec. 2008	NHGRI catChrV17e	Available
	felCat3	Mar. 2006	Broad Institute Release 3	Available
Chicken	galGal3	May 2006	WUSTL Gallus-gallus-2.1	Available
	galGal2	Feb. 2004	WUSTL Gallus-gallus-1.0	Available
Chimp	panTro2	Mar. 2006	CGSC Build 2 Version 1	Available
	panTro1	Nov. 2003	CGSC Build 1 Version 1	Available
Cow	bosTau4	Oct. 2007	Baylor College of Medicine HGSC Btau_4.0	Available

Recent human assemblies displayed in the genome browser (hg19 and hg18) are identical to the NCBI assemblies.

## Differences between UCSC and NCBI mouse assemblies [Index](#)

**Question:**  
"Is the mouse genome assembly displayed in the UCSC Genome Browser the same as the one on the NCBI website?"

**Response:**  
The mouse genome assemblies featured in the UCSC Genome Browser are the same as those on the NCBI web site with one difference: the UCSC versions contain only the reference strain data (C57BL/6J). NCBI provides data for several additional strains in their builds.

# 必要なDBを見つけるために

- HGNC Useful Links
  - <http://www.genenames.org/useful.html>
- WINGpro (データベースポータルサイト) by JST
  - <http://wingpro.lifesciencedb.jp/>
- 統合DBプロジェクト
  - <http://lifesciencedb.jp/>
  - 生命科学系 データベース カタログ
    - <http://lifesciencedb.jp/lbdb.cgi?pg=1>