

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

- [第1回]

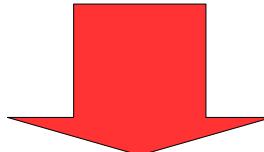
アウトライン

- 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。
 - 開発といつても、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

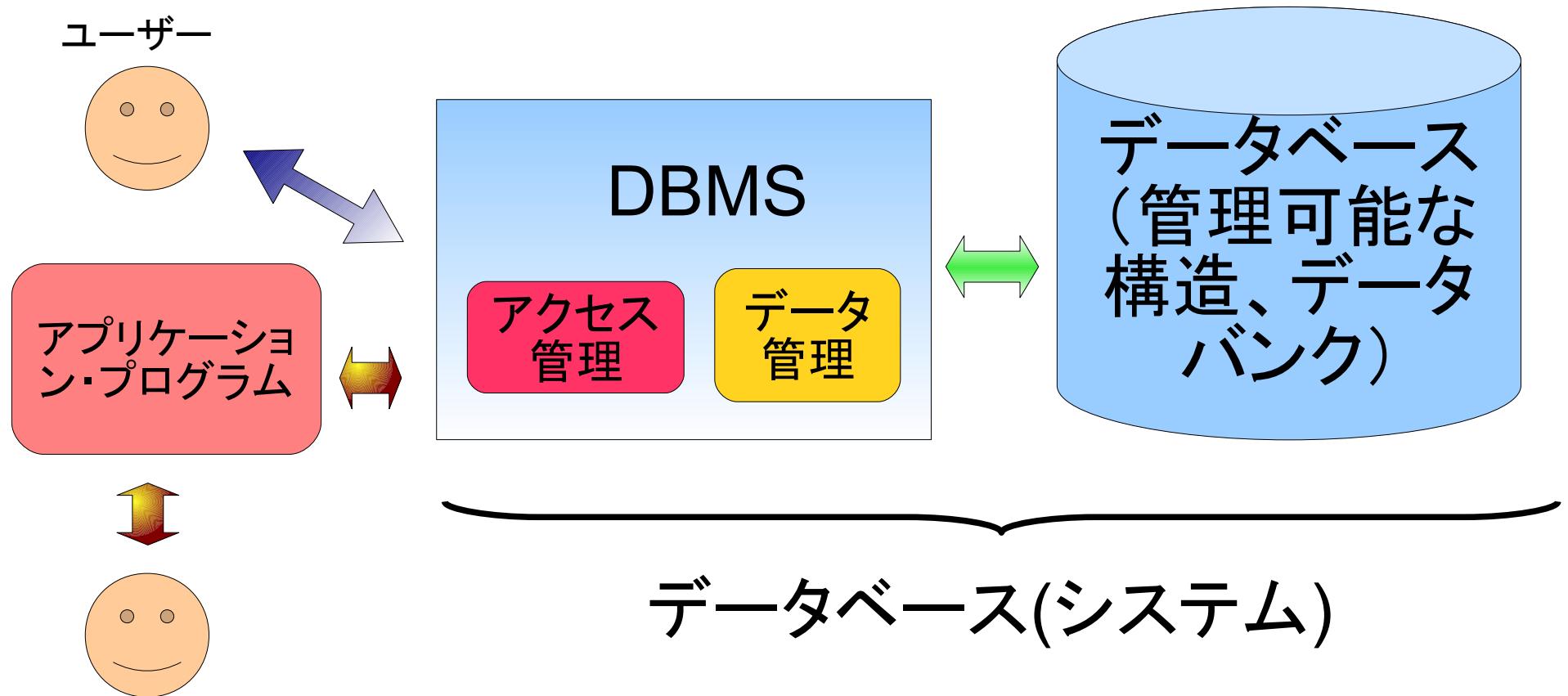
- Publicationは論文だけではない。
 - 論文投稿前に、データをデータベースに登録する必要がある。
 - 配列データ、発現データ、立体構造など
 - データベース管理者側も親切にマニュアルを用意している。
 - eg., 次世代シーケンサーによるデータの登録システムのマニュアル
http://trace.ddbj.nig.ac.jp/files/manual/DRA_submission_manual.pdf
 - 技術革新の速さ、研究の多様性から、想定外のことも多い。
 - データの形式を一括変換するのはユーザー(登録者)側。

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？(問い合わせ)

- ・「～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は
 - 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

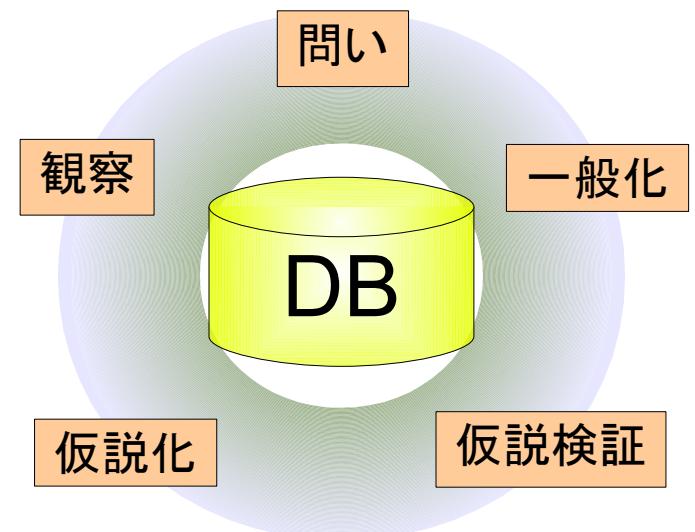
データベースとは(答え)

- 「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にて体現)。
 - 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」



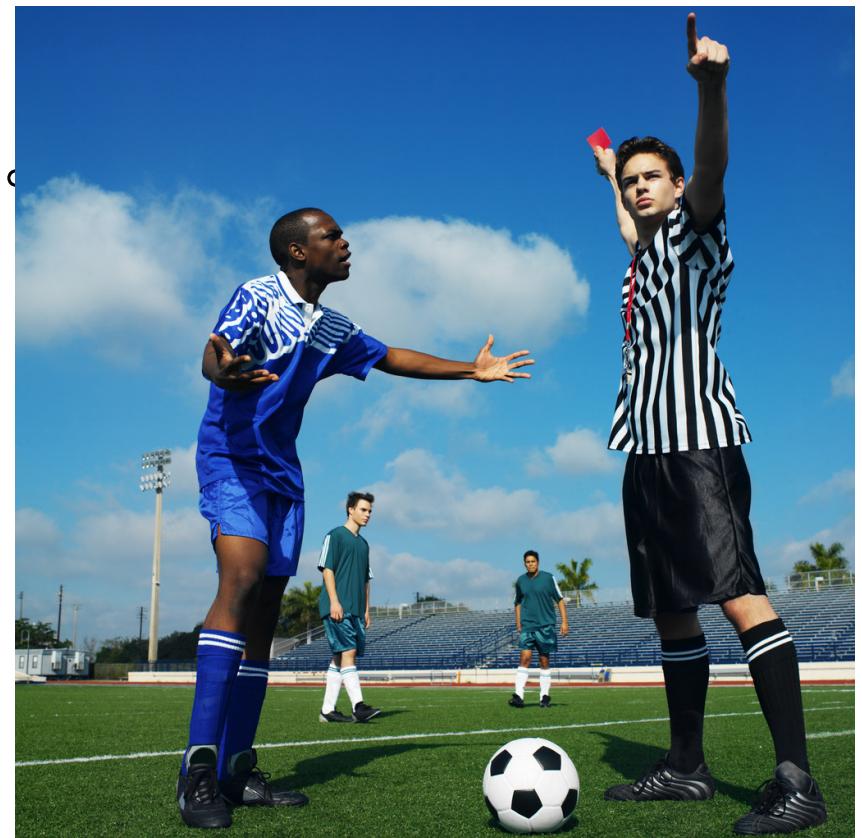
DB(生物情報)の生命科学における意義

- 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

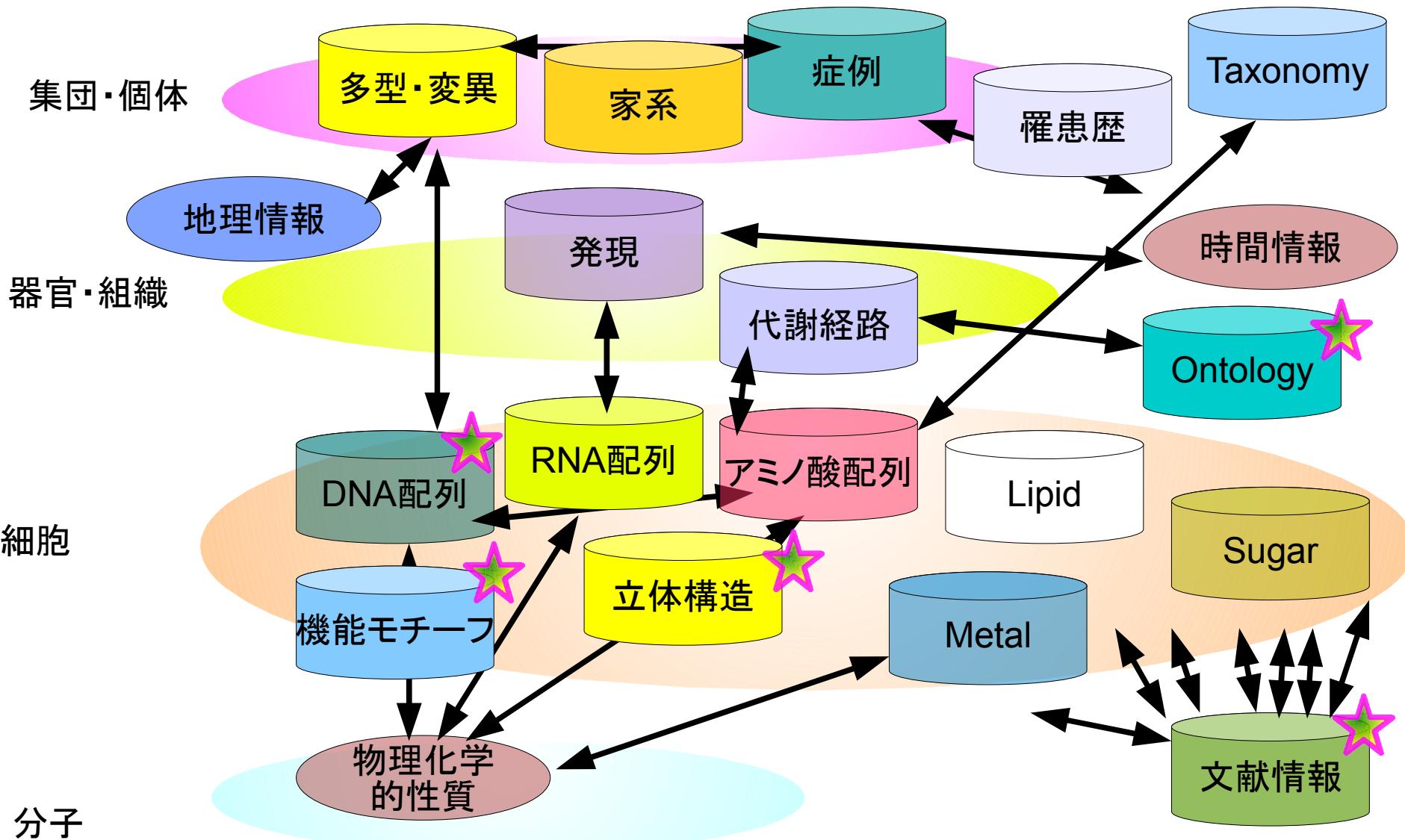


個々の研究者とDBとの関係

- 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方→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ある。
 - 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。

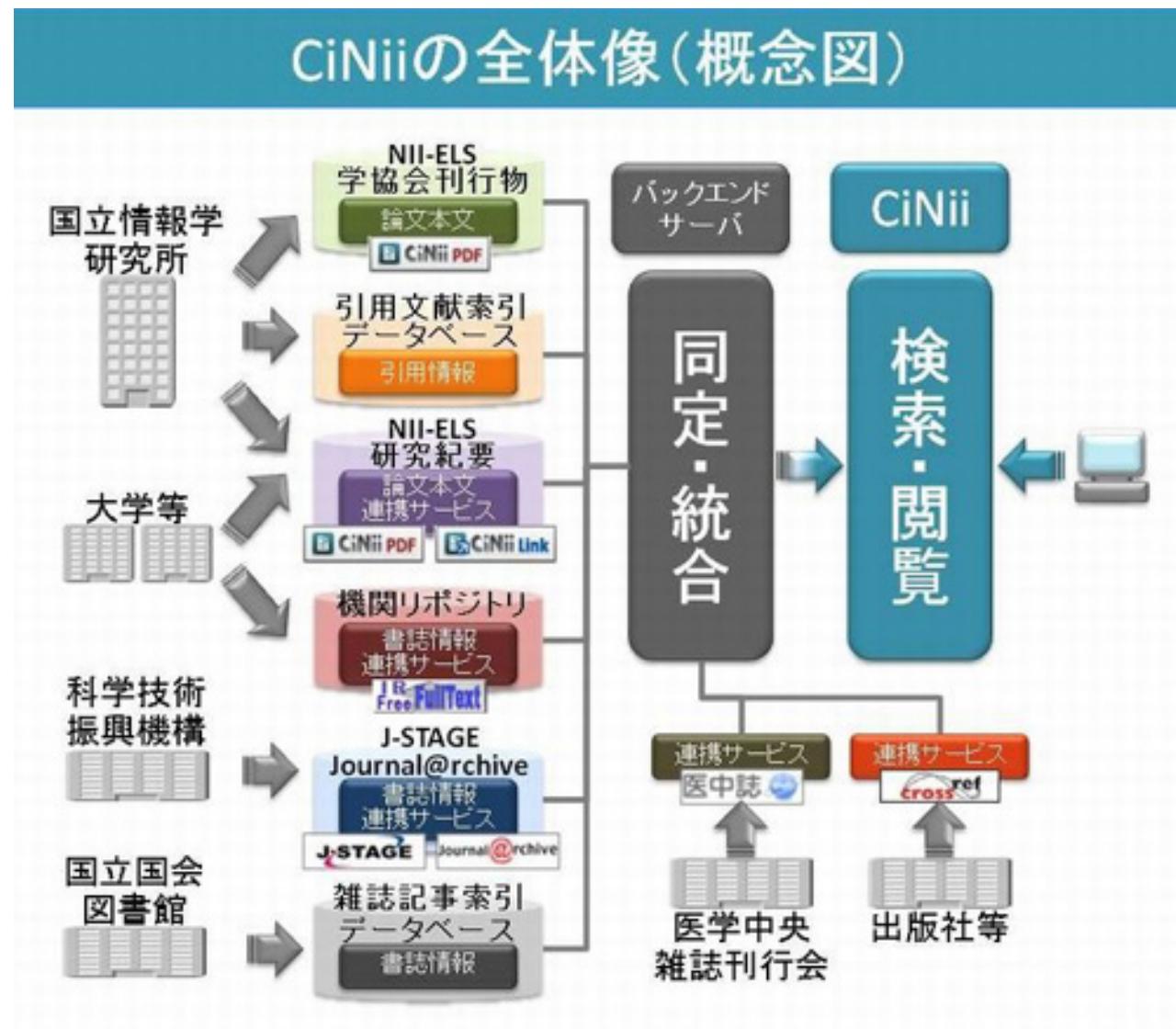
ブール演算(Boolean Operation)

	式	真理値表(Truth Table)			ベン図(Venn Diagram)
		入力A	入力B	出力	
論理積 (AND)	$A \text{ and } B$	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)	$A \text{ or } B$	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)	$A \text{ nand } B$	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)	$A \text{ nor } B$	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)	$A \text{ eor } B$	0	0	0	
	$A \text{ xor } B$	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)	$A \text{ not } B$	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、オックスフォード式、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ある。
- 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

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](#)

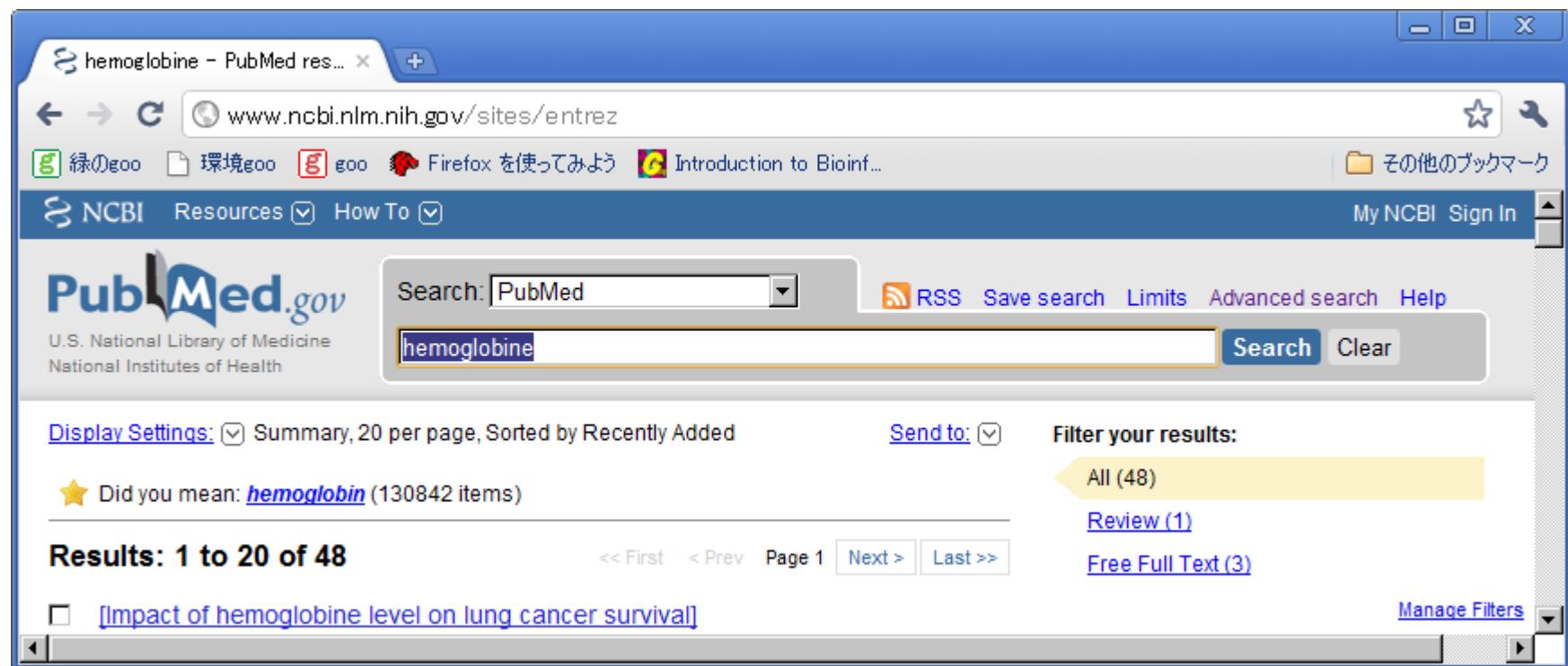
You are here: NCBI > Literature > PubMed

Write to the Help Desk

GETTING STARTED RESOURCES POPULAR FEATURED NCBI INFORMATION

NCBI Help Manual	Literature	PubMed	GenBank	About NCBI
NCBI Handbook	DNA & RNA	Nucleotide	Reference Sequences	Research at NCBI
Training & Tutorials	Proteins	BLAST	Map Viewer	NCBI Newsletter
	Sequence Analysis	PubMed Central	Genome Projects	NCBI FTP Site

スペルミスの場合



A screenshot of a Firefox browser window. The title bar says "hemoglobine - PubMed res...". The address bar shows the URL "www.ncbi.nlm.nih.gov/sites/entrez". The page content is the PubMed search results for the query "hemoglobine". The search bar has "PubMed" selected and "hemoglobine" typed into the search field. Below the search bar, there are "Display Settings" (Summary, 20 per page, Sorted by Recently Added), "Send to" options, and a "Filter your results" section with "All (48)" highlighted. The results list starts with "Results: 1 to 20 of 48" and includes a link to "Impact of hemoglobine level on lung cancer survival".

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

hemoglobin - PubMed resu... 

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

NCBI Resources How To

My NCBI Sign In

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

Search: PubMed hemoglobin RSS Save search Limits Advanced search Help

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

Results: 1 to 20 of 130842 << First < Prev Page 1 Next > Last >>

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:

- [Factors predicting clinically significant fatigue in women following treatment for primary breast cancer.](#)
1. Gerber LH, Stout N, McGarvey C, Soballe P, Shieh CY, Diao G, Springer BA, 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 by Expression of Vitreoscilla Hemoglobin Using a Novel Expression System in Corynebacterium crenatum.](#)
2. Xu M, Rao Z, Xu H, Lan C, Dou W, Zhang X, Xu H, Jin 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.](#)
3. 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.](#)
4. 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 - PubMed resu... + 

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

EST
Structure
Genome
BioSystems
Books
CancerChromosomes
Conserved Domains
dbGaP
dbVar
3D Domains
Epigenomics
Gene
Genome Project
GENSAT
GEO Profiles
GEO DataSets
HomoloGene
Journals
MeSH

RSS Save search Limits Advanced search Help

Display Settings: Summary, 20 per page

Results: 1 to 20 of 130842

[Factors predicting clinically significant primary breast cancer.](#)
1. [Using a Novel Expression System in](#)

Gerber LH, Stout N, McGarvey C, S, 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 by](#)
2. [Using a Novel Expression System in](#)

Xu M, Rao Z, Xu H, Lan C, Dou W, Zhang X, Xu H, Jin 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.](#)
3. [Using a Novel Expression System in](#)

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.](#)
4. [Using a Novel Expression System in](#)

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](#)

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:

Find items

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

Search See more...

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

鐸形赤血球は別
タームとなっ
て
いることに注意。

hemoglobin - MeSH Result... ×

www.ncbi.nlm.nih.gov/mesh?term=hemoglobin

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

My NCBI [Sign In] [Re]

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

NCBI MeSH

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

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

Limits Preview/Index History Clipboard Details

Search PubMed Clear

Display Summary Show 20 Send to

All: 985

Items 1 - 20 of 985 Page 1 of 50 Next

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.

Recent activity

Turn Off

hemoglobin (985)

hemoglobine (48)

Haemoglobin (130842)

Hemoglobin (130842)

"Cold Temperature"[MeSH M... (21)]

» See more

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 NCBI MeSH search results for the term 'hemoglobins'. The search bar at the top contains 'MeSH' and 'hemoglobins'. Below the search bar, there is a 'Search Box with AND' checkbox, which is highlighted with a purple oval. The results list shows 'All: 1' result for 'Hemoglobins'. The detailed description for 'Hemoglobins' states: '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.' It also notes that the term was introduced in 1975. A 'Subheadings' section is present, with a note that the list may not reflect current rules for allowable combinations. A large oval highlights the 'classification' subheading under 'Subheadings'. Below this, there is a list of MeSH terms and subterms, with 'classification' checked. Other terms listed include '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', and 'toxicity'. A 'See Also' section at the bottom lists 'Hemoglobin', 'Eryhem', 'Ferrous Hemoglobin', and 'Hemoglobin, Ferrous'.

下のようにチェックを入れると、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、そのまま[Search PubMed]

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 G Introduction to Bioinf...

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](#)
[Erythrocyruorins](#)
[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](#)
[Hemeproteins](#)
[Globins](#)

Hemoglobins

[Carboxyhemoglobin](#)
[Erythrocyruorins](#)

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]

Hemoglobins - MeSH Res... 

www.ncbi.nlm.nih.gov/mesh

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

NCBI MeSH A service of the N and the

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search MeSH for cold Go Clear

Limits Preview/Index History Clipboard Details

About Entrez "Hemoglobins/classification" [Major]

Text Version

Entrez PubMed 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

My NCBI [Sign In] [Register]

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ます。

1) checkいれてから

The screenshot shows the NCBI MeSH search results for the query "cold". The search interface includes a search bar with "MeSH" selected, a dropdown menu for "Search Box", and a "Display" dropdown set to "Summary". The results list 51 items, with the first few being:

- 1: Cold Temperature**
An absence of warmth or heat or a temperature 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**

A yellow callout bubble on the right side of the screen points to the "Search Box with AND" option in the dropdown menu, with the text "2) Search Box with ANDを選択".

再び別タームを同様に選択

checkいてから

cold - MeSH Results

www.ncbi.nlm.nih.gov/mesh

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

My NCBI [Sign In] [Register]

NCBI MeSH

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

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search MeSH for cold

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 temperature below 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](#)

Send to Text File Printer Clipboard Search Box with AND Search Box with OR Search Box with NOT

of 3 Next Links

Recent activity

Turn Off Clear

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

"Hemoglobins/classification" [PubMed]

cold (51)

Cold Climate

"Cold Climate" (1) MESH

The hemoglobins of fishes living at polar latitudes - current

» See more...

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。

cold - MeSH Results

www.ncbi.nlm.nih.gov/mesh

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

My NCBI [Sign In] [Register]

NCBI MeSH

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

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)[Mejr] AND "Cold Temperature"[Mesh]
OR "Cold Climate"[Mesh]

Search PubMed Clear

Display Summary Show 20 Send to

All: 51

Items 1 - 20 of 51 Page 1 of 3 Next

□ 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

どんなMeSH term
で検索されるか

The hemoglobins of fishes...

www.ncbi.nlm.nih.gov/pubmed?term="Hemoglobins/classification"[Mej] 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"[Mej] AND ("Cold Climate"[Mesh] OR "Cold Temp") **Search** Clear

Display Settings: Abstract Send to:

BenthamDirect www.bentham-direct.org

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. Notothenioidei, 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

"Hemoglobins/classification"[Mej] AND ("Cold Climate"[Mesh] OR "Cold Temperature"[Mesh])

Search See more...

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

Recent activity

修正した検索式

Search for

Go

Clear

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.

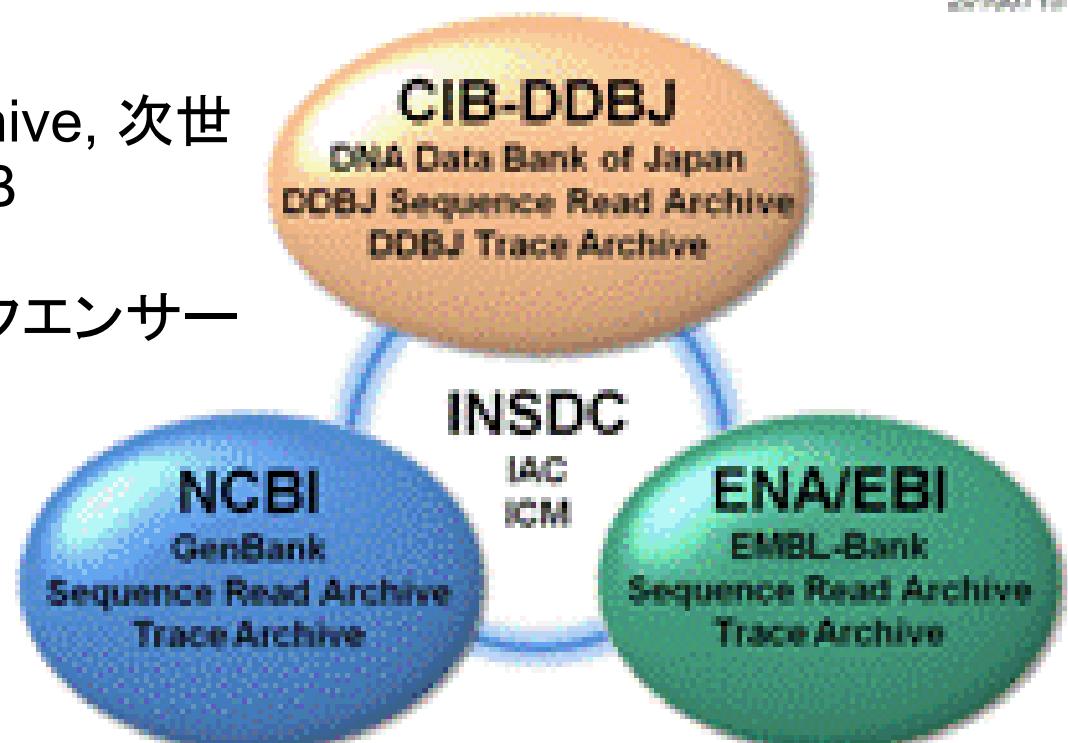
OMIM® and Online Mendelian Inheritance in Man® are registered trademarks of the Johns Hopkins University.

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

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

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					
AUTHORS	1 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				

フラット・ファイル の説明・ Locus field

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

- 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グ番号と同一に。

- 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, 次世代シークエンサーのreadの保存用DB)
 - Trace Archive, 従来型シークエンサーの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	バクテリア
	VRL	ウイルス
	PHG	バクテリオファージ

* GenBankでは無い。

BLASTなど相同意検索の際の参考に。

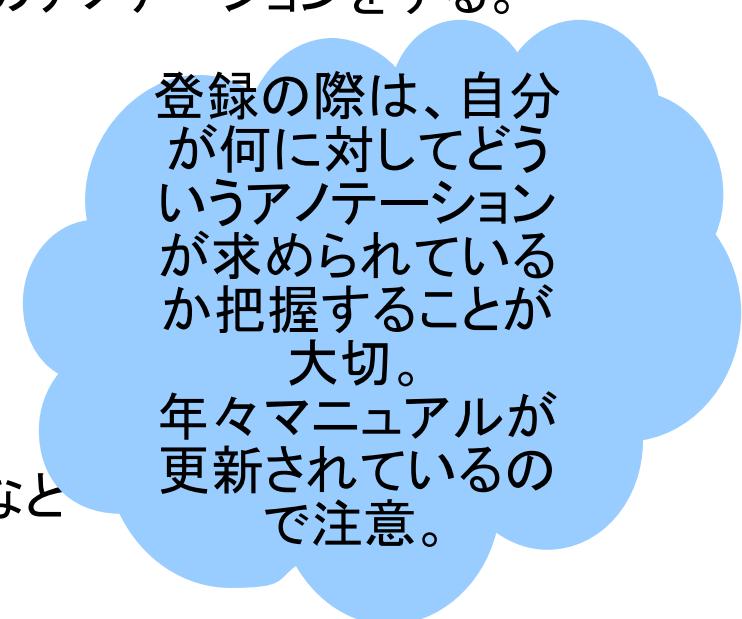
*	PAT	特許出願に含まれる塩基配列データ
	ENV	環境上のサンプルに由来した配列
	SYN	synthetic constructs
	EST	人為的に構成された配列
*	TSA	transcriptome shotgun assemblies, すでに登録された転写物配列をassembleして得られた長い配列データ
	GSS	genome survey sequences short single pass のゲノム配列
	HTC	high throughput cDNA sequences, 配列が finish した後, 生物種による division に移される場合あり。
	HTG	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つのアクセス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 primary literature	Akin to review articles
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: **Alternate assembly or annotation**
 - N: **kN**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: **prediX**tion
 - Z: **NZ_**accession(WGS)上について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

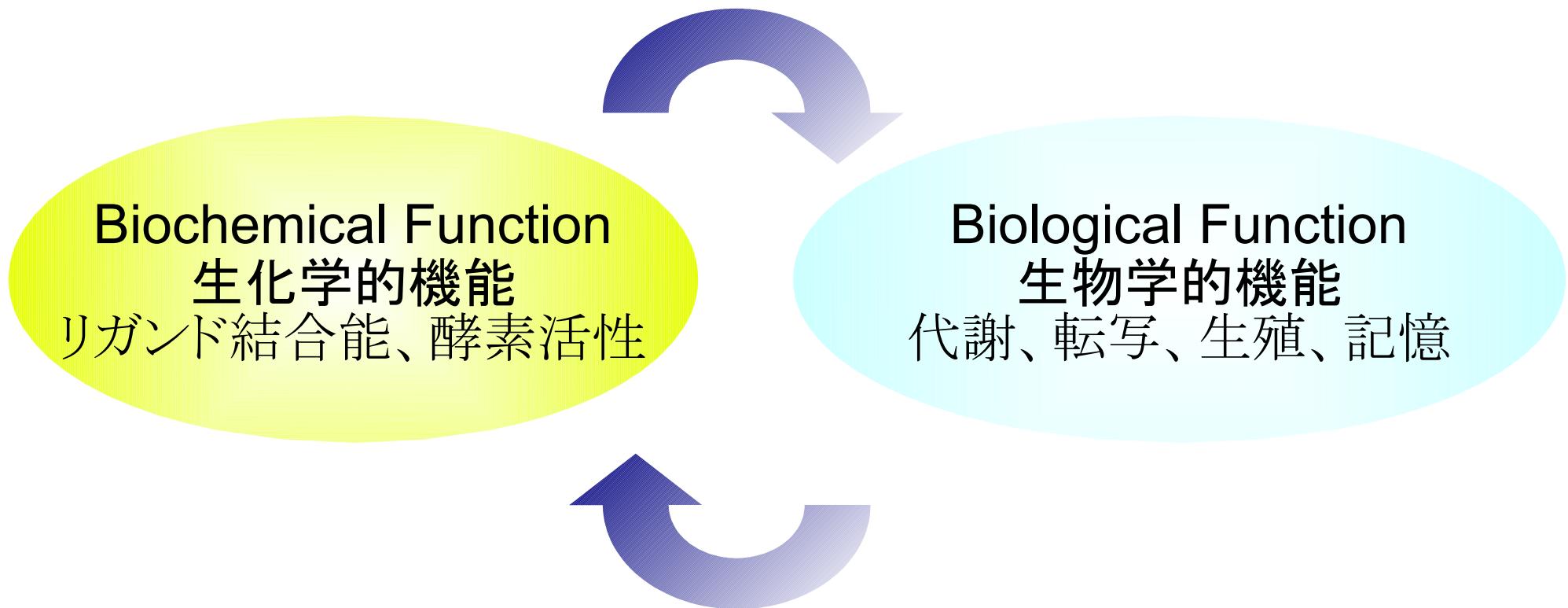
• H-InvDBのID

- 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

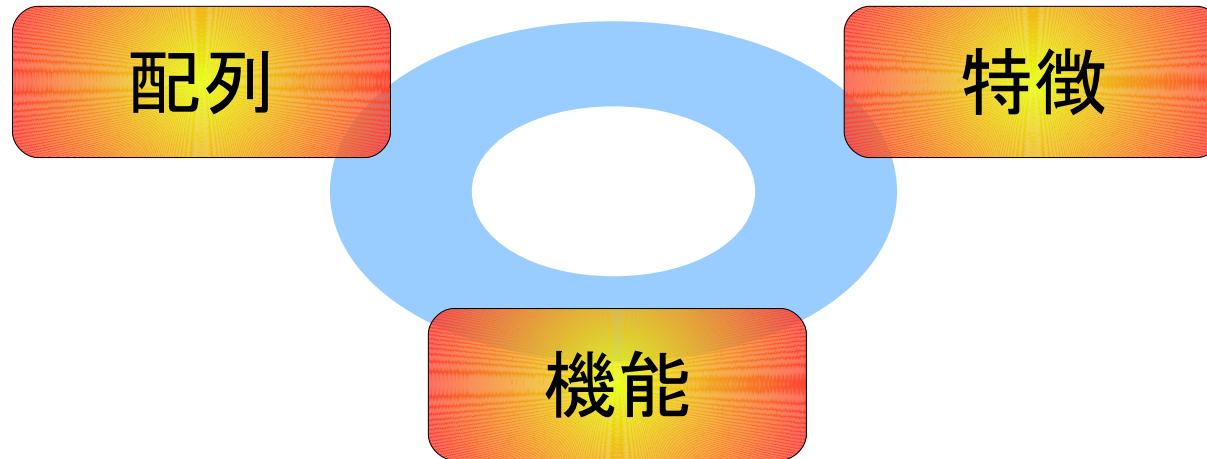
- 「機能」と配列について
- 機能ドメインデータベース
 - UniProt
 - InterPro
- 立体構造データベース
 - PDB

「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 リンクの変更 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">★ Swiss-Prot, which is manually annotated and reviewed.★ TrEMBL, which is automatically annotated and is not reviewed. Includes Complete Proteome Sets.
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations , taxonomy , keywords and more .

Getting started

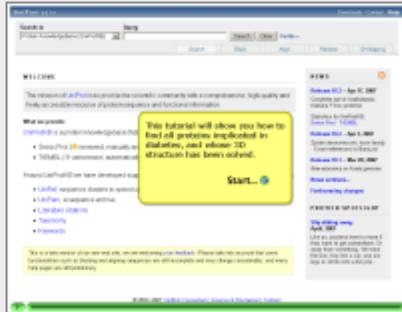
- Text search
- Sequence similarity searches (BLAST)
- Sequence alignments
- Batch retrieval

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



This screenshot shows the UniProt homepage. The top navigation bar includes links for 'Downloads', 'Contact', 'Documentation/Help', 'Search', 'Clear', and 'Fields ». Below this is a search interface with a dropdown menu set to 'Protein Knowledgebase (UniProtKB)'. The main content area features a 'WELCOME' section with a mission statement, followed by a 'What we provide' section in a red-bordered box. The 'What we provide' box contains five entries: UniProtKB (describing Swiss-Prot and TrEMBL), UniRef (sequence clusters), UniParc (sequence archive), and Supporting data (literature citations, taxonomy, keywords, and more). Below this is a 'Getting started' section with a list of tools: Text search, Sequence similarity searches (BLAST), Sequence alignments, and Batch retrieval. To the right, there is a 'NEWS' section with a link to the 'UniProt release 2010_09 – Aug 10, 2010' and a 'SITE TOUR' section with a screenshot of the UniProt interface.

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

PROTEIN SPOTLIGHT

“U2AF”で検索

UniProt ... > U2AF in... > geneu2a... > U2AF ta... > Splicing ... > Splicing ... > blastp o... > Wilms.tu... > +

www.uniprot.org/uniprot/?query=U2AF&sort=score

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

UniProt > UniProtKB

Downloads · Contact · Documentation/Help

Search · Blast · Align · Retrieve · ID Mapping *

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

Customize display · 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
P26368	U2AF2_HUMAN	★	Splicing factor U2AF 65 kDa subunit	U2AF2 U2AF65	Homo sapiens (Human)	475
Q01081	U2AF1_HUMAN	★	Splicing factor U2AF 35 kDa subunit	U2AF1 U2AF35 U2AFBP	Homo sapiens (Human)	240
P36629	U2AF2_SCHPO	★	Splicing factor U2AF 59 kDa subunit	prp2 mis11 SPBC146.07	Schizosaccharomyces pombe (Fission yeast)	517
Q8WU68	U2AF4_HUMAN	★	Splicing factor U2AF 26 kDa subunit	U2AF1L4 U2AF1-RS3 U2AF1L3	Homo sapiens (Human)	220
Q09176	U2AF1_SCHPO	★	Splicing factor U2AF 23 kDa subunit	SPAP8A3.06	Schizosaccharomyces pombe (Fission yeast)	216
Q8BGJ9	U2AF4_MOUSE	★	Splicing factor U2AF 26 kDa subunit	U2af1I4 U2af26	Mus musculus (Mouse)	220
P26369	U2AF2_MOUSE	★	Splicing factor U2AF 65 kDa subunit	U2af2 U2af65	Mus musculus (Mouse)	475
Q24562	U2AF2_DROME	★	Splicing factor U2AF 50 kDa subunit	U2af50 CG9998	Drosophila melanogaster (Fruit fly)	416
P90978	U2AF2_CAEEL	★	Splicing factor U2AF 65 kDa subunit	uaf-1 Y92C3B.2	Caenorhabditis elegans	496
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

UniProt ... [U2AF in...](#) [geneu2a...](#)

www.uniprot.org/uniprot/?query=gene:u2af&sort=score

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

UniProt > UniProtKB Downloads · Contact · Documentation/Help

Search Blast Align Retrieve ID Mapping *

Search in Query

Protein Knowledgebase (UniProtKB) gene:u2af Search Clear 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% | Download

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

Accession	Entry name	Status	Protein names	Gene names	Organism	Length
Q01081	U2AF1_HUMAN	★	Splicing factor U2AF 35 kDa subunit	U2AF1 U2AF35 U2AFBP	Homo sapiens (Human)	240
P26368	U2AF2_HUMAN	★	Splicing factor U2AF 65 kDa subunit	U2AF2 U2AF65	Homo sapiens (Human)	475
P26369	U2AF2_MOUSE	★	Splicing factor U2AF 65 kDa subunit	U2af2 U2af65	Mus musculus (Mouse)	475
Q8BGJ9	U2AF4_MOUSE	★	Splicing factor U2AF 26 kDa subunit	U2af1I4 U2af26	Mus musculus (Mouse)	220
Q8WU68	U2AF4_HUMAN	★	Splicing factor U2AF 26 kDa subunit	U2AF1L4 U2AF1-RS3 U2AF1L3	Homo sapiens (Human)	220
Q15696	U2AFM_HUMAN	★	U2 small nuclear ribonucleoprotein auxiliary	ZRSR2 U2AF1-RS2 U2AF1L2 U2AF1RS2	Homo sapiens (Human)	482
Q94535	U2AF1_DROME	★	Splicing factor U2af 38 kDa subunit	U2af38 CG3582	Drosophila melanogaster (Fruit fly)	264
Q9D883	U2AF1_MOUSE	★	Splicing factor U2AF 35 kDa subunit	U2af1	Mus musculus (Mouse)	239
Q24562	U2AF2_DROME	★	Splicing factor U2AF 50 kDa subunit	U2af50 CG9998	Drosophila melanogaster (Fruit fly)	416
Q15695	U2AF1_HUMAN	★	U2 small nuclear	ZRSR1 U2AF1-RS1 U2AF1L1

“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
Q01081	U2AF1_HUMAN	★	Splicing factor U2AF 35 kDa subunit	U2AF1 U2AF35 U2AFBP	Homo sapiens (Human)	240
P26368	U2AF2_HUMAN	★	Splicing factor U2AF 65 kDa subunit	U2AF2 U2AF65	Homo sapiens (Human)	475
Q8WU68	U2AF4_HUMAN	★	Splicing factor U2AF 26 kDa subunit	U2AF1L4 U2AF1-RS3 U2AF1L3	Homo sapiens (Human)	220
Q15696	U2AFM_HUMAN	★	U2 small nuclear ribonucleoprotein auxiliary ...	ZRSR2 U2AF1-RS2 U2AF1L2 U2AF1RS2	Homo sapiens (Human)	482
Q15695	U2AFL_HUMAN	★	U2 small nuclear ribonucleoprotein auxiliary ...	ZRSR1 U2AF1-RS1 U2AF1L1 U2AF1P U2AF1RS1 U2AFBPL	Homo sapiens (Human)	479
Q9H665	TM149_HUMAN	★	Transmembrane protein 149	TMEM149 U2AF1L4	Homo sapiens (Human)	355
Q7Z780	Q7Z780_HUMAN	★	U2 small nuclear RNA auxiliary factor 1	U2AF1	Homo sapiens (Human)	167
Q701P4	Q701P4_HUMAN	★	U2 snRNP auxiliary factor small subunit	U2AF1 hCG_401287	Homo sapiens (Human)	240
Q71RF1	Q71RF1_HUMAN	★	FP793	U2AF1 hCG_401287	Homo sapiens (Human)	167
B5BU08	B5BU08_HUMAN	★	U2 small nuclear RNA auxiliary factor 1 isofo...	U2AF1	Homo sapiens (Human)	240
B5BU25	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

Names and origin

Protein names	<p>Recommended name: Splicing factor U2AF 26 kDa subunit</p> <p>Alternative name(s): 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</p>
Gene names	Name: U2AF1 L4 Synonyms:U2AF1-RS3, U2AF1 L3
Organism	Homo sapiens (Human) [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 GFI1. Binds to RNA at the AG dinucleotide at the 3'-splice site (By similarity) .
Subunit structure	Interacts with GFI1 and U2AF2 (By similarity) .
Subcellular location	Nucleus (By similarity) Nucleus speckle (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. (Ref1)
Domain	The second zinc finger is necessary for interaction with GFI1 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](#)

Names and origin

Protein names	<p>Recommended name: Transmembrane protein 149</p> <p>Alternative name(s): U2 small nuclear RNA auxiliary factor 1-like 4</p>
Gene names	Name: TMEM149 Synonyms:U2AF1 L4
Organism	Homo sapiens (Human) [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	355 AA.
Sequence status	Complete.
Sequence processing	The displayed sequence is further processed into a mature form.
Protein existence	Evidence at transcript level.

General annotation (Comments)

Function	Membrane; Single-pass type I membrane protein (Potential)
Subunit structure	Membrane
Subcellular location	Alternative splicing Polymorphism
Tissue specificity	Signal Transmembrane Transmembrane helix
Domain	Complete proteome

General annotation (Comments)

Function	Integral to membrane Inferred from electronic annotation. Source: UniProtKB-KW
Subunit structure	
Subcellular location	
Tissue specificity	
Domain	

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

Sequence or UniProt identifier

```
>Q9H665-1
MGPGRCLLTALLLALAPPPEASQYCGRLEYWNPDNKCCSSCLQRFGGPPPCPDYEFRENC
GLNDHGDFVTPPFRKCSSGQCNPDGAELCSPCGGGAVTPTPAAGGGRTPWRCRERPVPAK
GHCPLTPGNPGAPSSRERSSPASSIAWRTPPEPVPRRAWPNFLPLVVLVLLTLAVIAILL
FILLWHLCWPKEKADPYPYPGLVCGVPNTHTPSSSHLSSPGALETGTWKEASLLPLLSR
ELSSLASQPLSRLLDELEVLEELIVLDDPEPGPGGGMAHGTTRHLAARYGLPAAWSTFAY
```

Help

For a sequence similarity search, enter:
• a protein or nucleotide sequence
• a UniProt identifier, e.g.
P00750 or A4_HUMAN or UP1000000001
More...

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 Taxonomy

UniProtKB (6) Homo sapiens (Human) (6) Filter Reset

Graphic overview

Color code for identity 0-100% = 

Accession	Entry name	OQuery hit355	OMatch hit (sqrt scale) 314	Name (Organism)
Query I8UR				
Q9H665	TM149_HUMAN	██████████	██████	Transmembrane protein 149 (Homo sapiens)
Q9H665-2	TM149_HUMAN	██	██	Isoform 2 (Homo sapiens)
B4DEL3	B4DEL3_HUMAN	····	····	cDNA FLJ50454, moderately similar to ... (Homo sapiens)
B4DEU6	B4DEU6_HUMAN	····	····	cDNA FLJ55675, highly similar to Ster... (Homo sapiens)
A2VEC9-2	SSPO_HUMAN	····	····	Isoform 2 (Homo sapiens)
Q9BUK6-3	MSTO1_HUMAN	····	····	Isoform 3 (Homo sapiens)

Detailed BLAST results

Alignments	Accession	Entry name	Status	Protein Names	Organism	Length	Identity	E-Value
██████████	Q9H665	TM149_HUMAN	★	Transmembrane protein 149	Homo sapiens (Human)	355	100.0%	0.0
██	Q9H665-2	TM149_HUMAN	★	Isoform 2	Homo sapiens (Human)	167	100.0%	2.0 × 10 ⁻⁵⁷
····	B4DEL3	B4DEL3_HUMAN	★	cDNA FLJ50454, moderately similar to	Homo sapiens	510	25.0%	1.5 × 10 ⁻¹

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 - Google 検索

InterPro protein sequence ...

www.ebi.ac.uk/interpro/

- InterPro:Home
- Advanced Search
- InterProScan
- BioMart
- Help / Documentation
- About InterPro
- Release Notes
- BioMart Manual
- Tutorial
- Publications
- Contributors
- Web Services
- Downloads
- Protein Focus
- Collagen

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](#).

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。



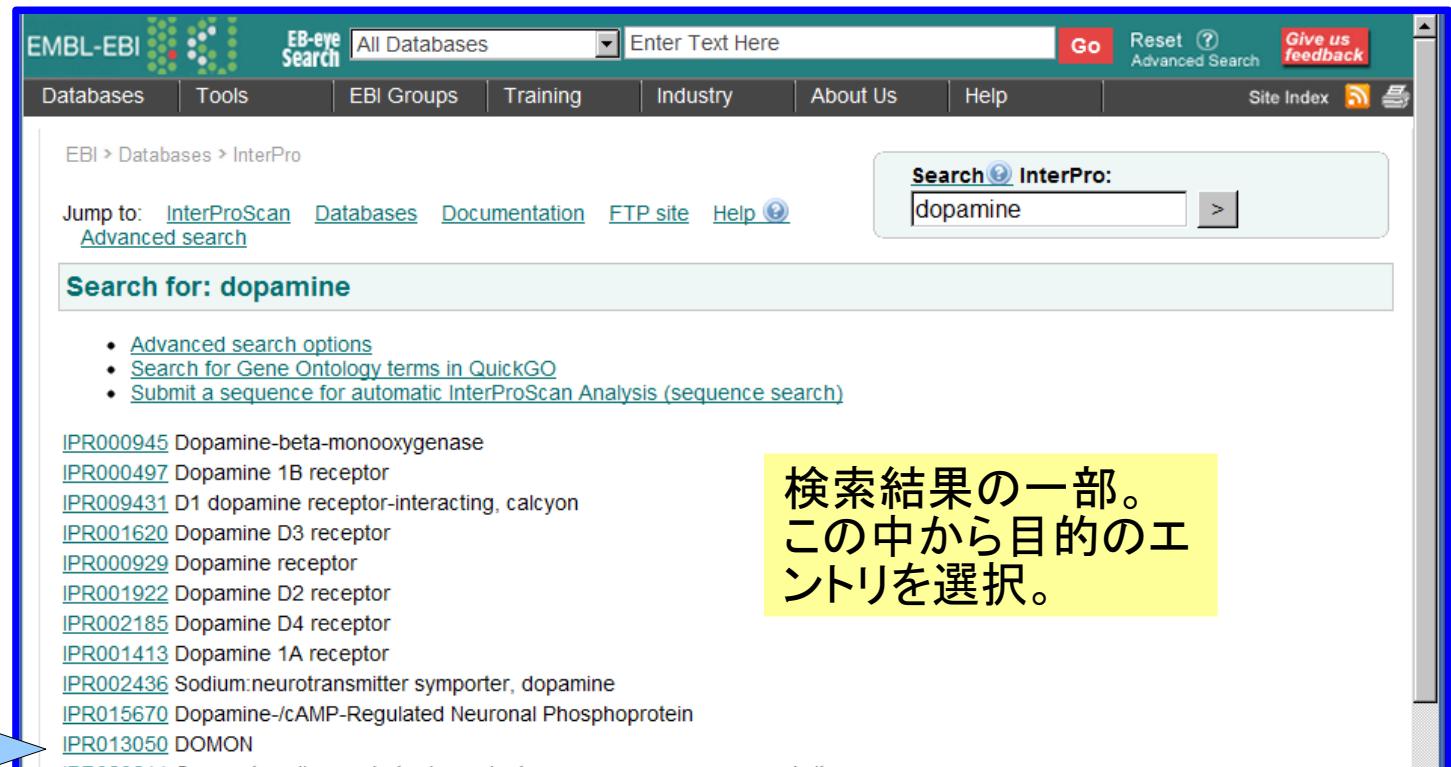
InterPro 28.0

・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**
 - ・リピート
 - ・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)

統合された
データベース

“dopamine”
を検索窓に



The screenshot shows the EMBL-EBI InterPro search results for the query "dopamine". The search bar at the top contains "dopamine". Below the search bar, a list of search options is provided: "Advanced search options", "Search for Gene Ontology terms in QuickGO", and "Submit a sequence for automatic InterProScan Analysis (sequence search)". The main results list includes: IPR000945 Dopamine-beta-monoxygenase, 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, and IPR013050 DOMON. A yellow callout box on the right side of the results list says: "検索結果の一部。この中から目的のエントリを選択。" (A portion of the search results. Select the entry you want from this list.)

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.

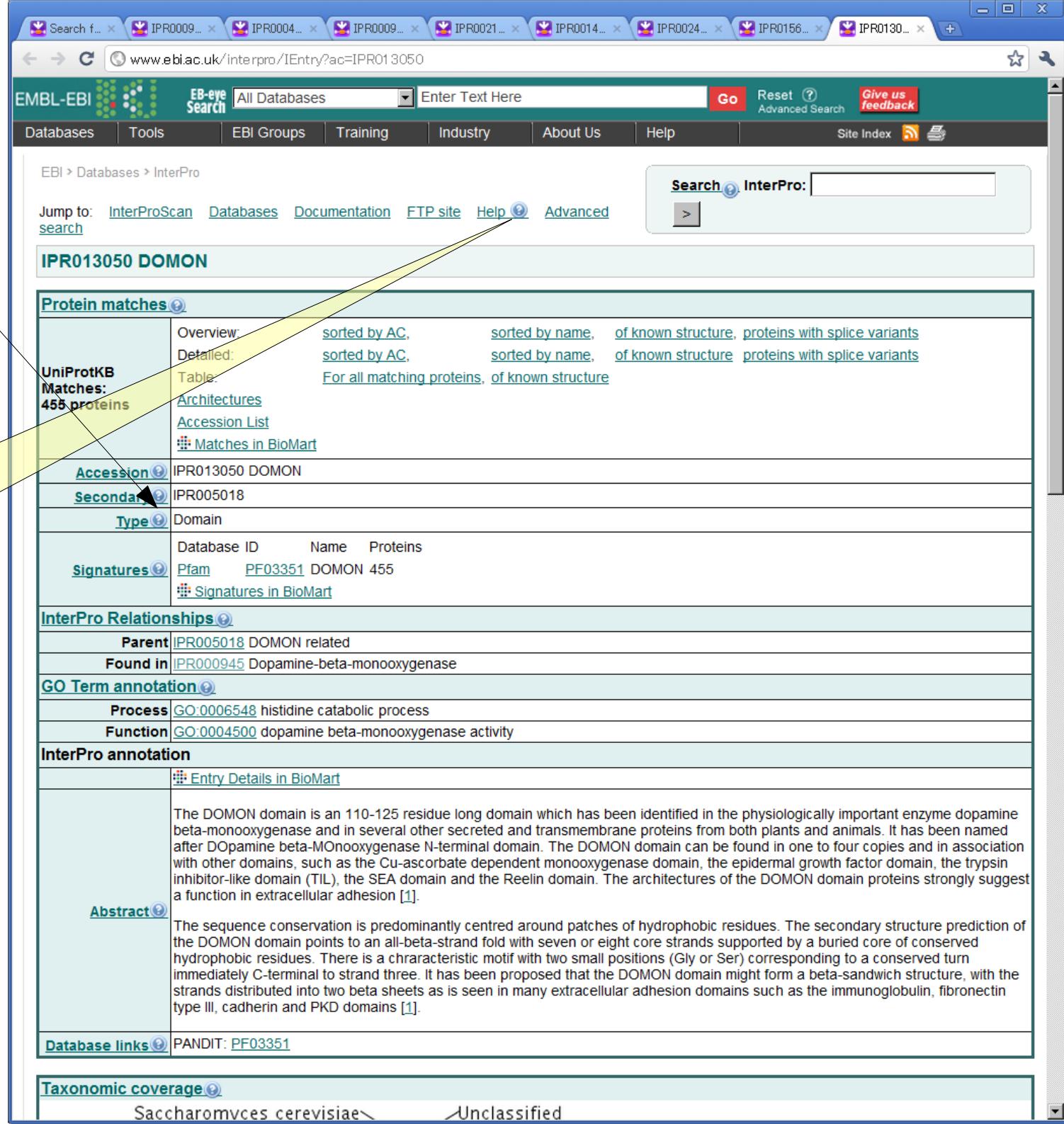
Member database information

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

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clickしてから、各field名の脇(?)をclick。
-->それぞれのfiledの説明が表示される。



www.ebi.ac.uk/interpro/INentry?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

Search InterPro: >

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: Architectures Accession List Matches in BioMart
Accession	IPR013050 DOMON
Secondary	IPR005018
Type	Domain
Signatures	Database ID Name Proteins Pfam PF03351 DOMON 455 Signatures in BioMart

InterPro Relationships

Parent	IPR005018 DOMON related
Found in	IPR000945 Dopamine-beta-monoxygenase

GO Term annotation

Process	GO_0006548 histidine catabolic process
Function	GO_0004500 dopamine beta-monoxygenase activity

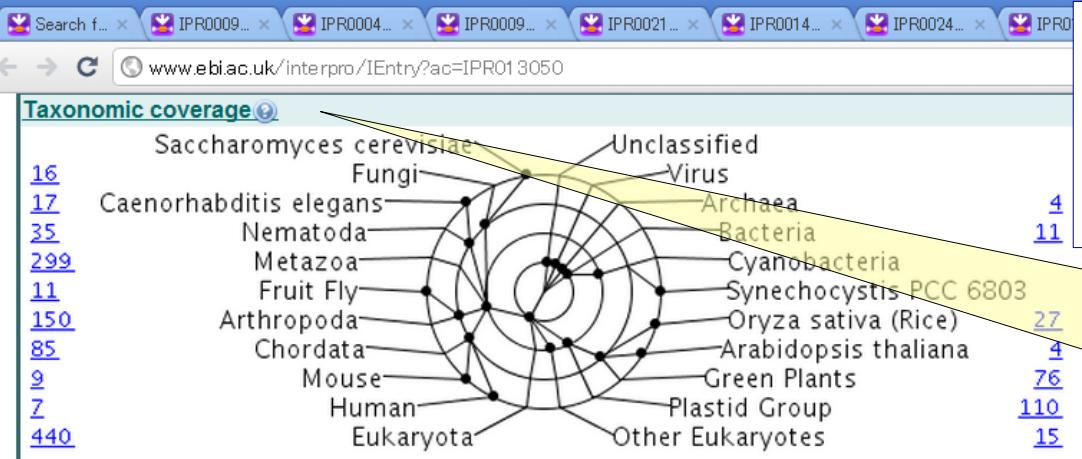
InterPro annotation

Abstract	<p>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 monooxygenase 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].</p> <p>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].</p>
Database links	PANDIT: PF03351

Taxonomic coverage

Saccharomyces cerevisiae	Unclassified
--------------------------	--------------

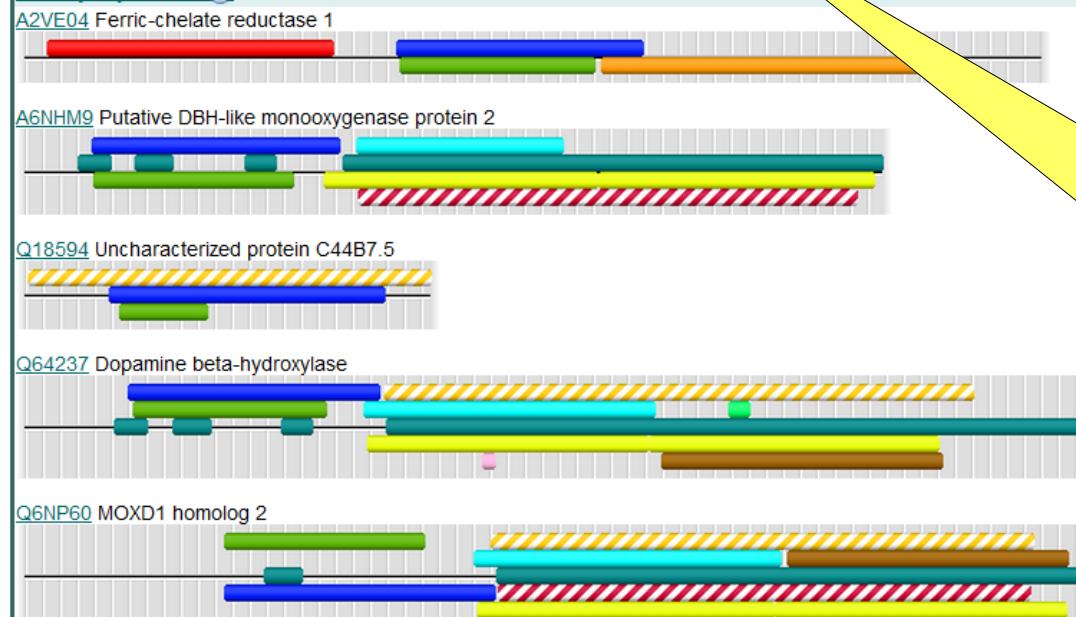
Type = Domainの例 (下部)



Overlapping InterPro entries



Example proteins



[More proteins](#)

Example Proteins Key

InterPro entry accession number/name and structure databases	Colour code
IPR020611 Copper type II, ascorbate-dependent monooxygenase, conserved site	pink
IPR002861 Reeler domain	red
IPR008977 PHM/PNGase F-fold domain	yellow

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が可能。
- <http://www.ebi.ac.uk/Tools/InterProScan/>

InterProScan トップ画面

Search for: dopamine IPR000945 Dopamine-βeta... IPR000497 Dopamine 1B r... InterProScan Signature Re...

www.ebi.ac.uk/Tools/InterProScan/

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 > 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

BlastProDom FPrintScan HHMPfam HMMSmart
 HMMTigr ProfileScan HAMAP patternScan SuperFamily SignalPHMM
 TMHMM HHMPanther Gene3D

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

Upload a file: 選択されていません

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

Search for: dopamine IPR000945 Dopamine-beta... IPR000497 Dopamine 1B r... InterProScan Signature Re... +

www.ebi.ac.uk/Tools/InterProScan/

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 > 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

BlastProDom FPrintScan HHMPiR HHMPfam HMMSmart
 HMMTigr ProfileScan HAMAP patternScan SuperFamily SignalPHMM
 TMHMM HHMPanther Gene3D

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

MGQPNGSAFLIAPNGSHADPHDVTOQRDFWVWVGMGIVMSLIVLAIVFGNVLVITAIKFERLQTVTNYEFTSLACADIVMCLA
VVPFGAAHILMKMWTFEGNFWCEFWLSDVLCVTASETLCLVADRYFAITSPFKYQSLLTKNKARVILLMWIVSGLTSFLPTQ
MHMYRATHQEAINCYANETCCDFTINQAYAIASSIVSFYVPLVIMVFVYSRVEQEAQRLQKIDKSEGFRHVQNLSQLVEQDGRTG
HGLRSSSKFCLEHKALKTLGIIMGTFLLCMLPFTIVNIVHVQDNLIRKEYVILLNWIGVNSGFNPILYCRSPDFRIAFQELL
CLRRSSLKAYGNGYSSNGNTGEQSGYHVEQEKENKLLEDIPGTEDFVGHQGTVPSONIDSPGRNCSTNDLL

Upload a file: 選択されていません

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.

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.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

SEQUENCE: Sequence_1 CRC64: FB17B56DC1EAD8A4 LENGTH: 413 aa

Table View Raw Output XML Output Original Sequences SUBMIT ANOTHER JOB

InterPro IPR000276 Family InterPro SRS

7TM GPCR, rhodopsin-like

PR00237  GPCRRHODOPSN

PF00001  7tm_1

PS00237  G_PROTEIN_RECEP_F1_1

InterPro IPR000332 Family InterPro SRS

Adrenergic receptor, beta 2

PR00562  ADRENRCB2AR

PTHR19266-SF42  BETA-2 ADRENERGIC RECEPTOR (BETA-2 ADRENOCEPTOR)

InterPro IPR002233 Family InterPro SRS

Adrenergic receptor

PR01103  ADRENERGICR

InterPro IPR017452 Domain InterPro SRS

GPCR, rhodopsin-like superfamily

PS50262  G_PROTEIN_RECEP_F1_2

noIPR unintegrated

unintegrated

G3DSA:1.20.1070.10  no description

PTHR19266  G-PROTEIN COUPLED RECEPTOR

SignalP  signal-peptide

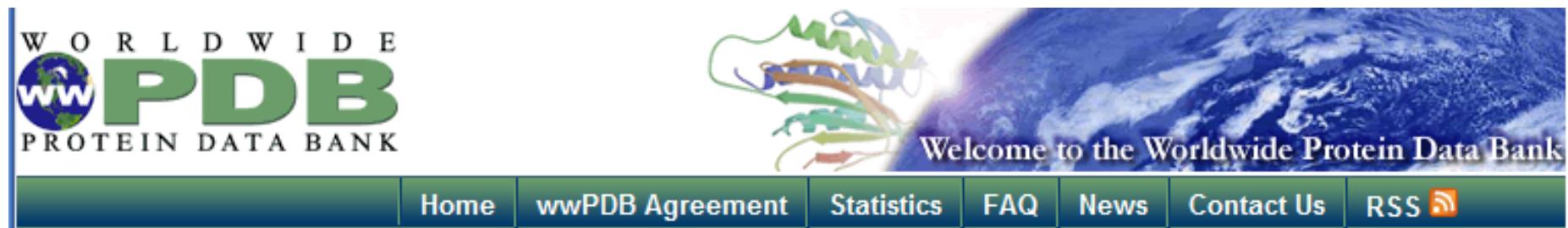
tmhmm  transmembrane_regions

SSF81321  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

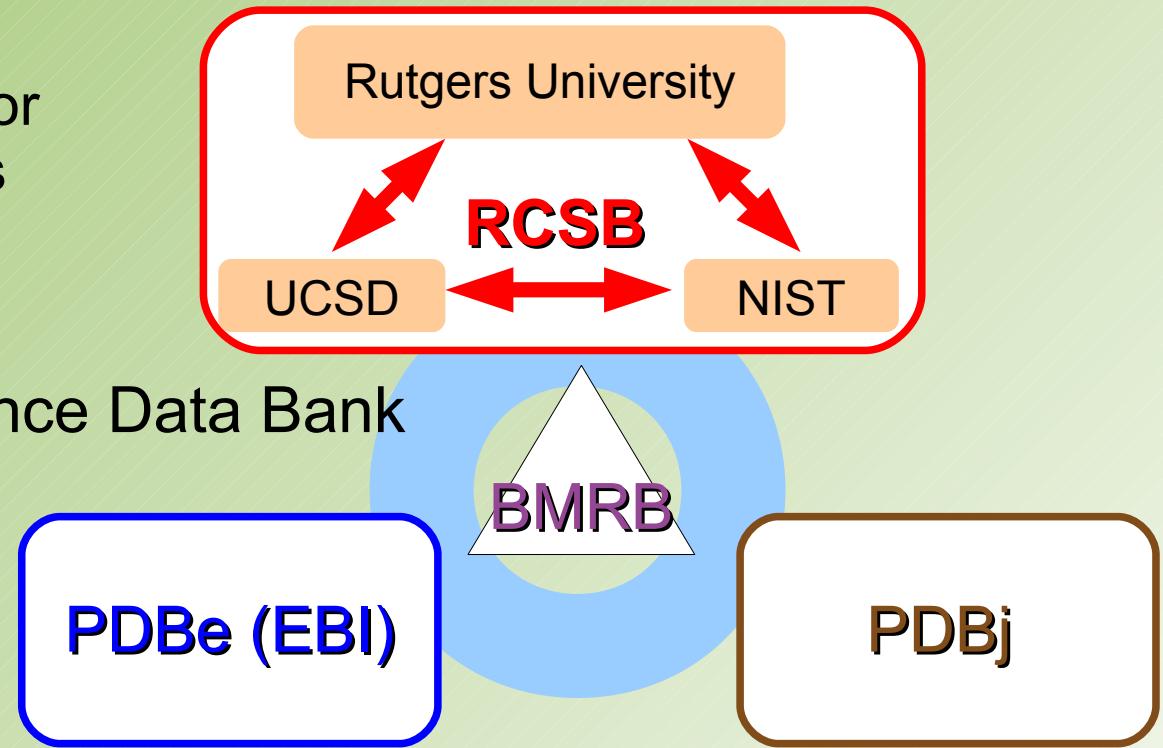
- ・タンパク質および核酸の三次元構造のデータベース
- ・X線結晶解析法やNMR(核磁気共鳴)法実験によって得られた三次元データを登録者本人とアノテータとでそれぞれ審査・検証を経て公開される。
- ・wwPDBにおける共同関係は核酸DBにおける国際塩基配列データベース(INSDC)に似ている。
 - ・データ・アーカイブは唯一で共通である。
 - ・各機関はそれぞれにデータの登録受付とおよびブラウザ、ツール、webサービスを開発し公開する。

wwPDBの組織

Research Collaboratory for
Structural Bioinformatics
(RCSB), USA

Biological Magnetic Resonance Data Bank
(BMRB), USA

Protein Databank in
Europe (PDBe), Europe



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

PDBj

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 >> 検索 >> サービス&ソフトウェア >> 二次データベース >> ダウンロード >> リンク集

ADIT: PDB Deposition ADIT-NMR

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

/V: Graphic Viewer
Protein Globe
ASH
MAFFTash

Structure Prediction >>
CRNPRED
Spanner
SFAS

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

PDB Archive/Snapshot Archive

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  NMRデータ登録 

検索

PDB検索  Mine 日本語ページについて

Accession number
Deposition code

U2AF
PDB ID or Keyword Go

NMRデータ検索 

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・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 R&D

MP900433188.JPG MP900438881.JPG MP900438753.JPG すべ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)によつ
て、3Dデータが異
なるので注意。
サムネール画像の
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。

Google Chrome

service.pdbj.org/mine/Search2?query=U2AF&type=quick&search_go=Go

PDBj
蛋白質データベース

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 >> 検索 >> 検索結果ページ 日本語ページについて PDBj Mineについて 更新情報

ADIT: PDB Deposition
ADIT-NMR

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

(PDB-IDをクリックすると、詳細情報をご覧いただけます)

1 - 10/ 10

クエリ: U2AF PDB ID or Keyword 表示順: 一致件数

リセット 検索

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

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

個別データ表示例

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

PDBj Protein Data Bank Japan

Japanese

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
- MAFFT

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

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

PDB ID or Keyword 検索

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

エントリーID (PDB ID)
2u2f 配列情報 (FASTA形式)
PDBファイルのダウンロード

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
[UNP - U2AF2_HUMAN] Lymph, and Skin
[UNP - U2AF2_HUMAN] Colon carcinoma
[UNP - U2AF2_HUMAN] Hepatocyte
[UNP - U2AF2_HUMAN] Epithelium

由来する組織

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)). EMBO J 18:4523-4534

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。 http://mmcif.pdb.org/ Data parsing tools(C++)がRCSBから用意されている (http://sw-tools.rcsb.org/)。	+	+	+
PDBML	mmCIFに基づくXML形式。 http://pdbml.pdb.org/	+	—	++

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 2H2C
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
CONECT   163  715
CONECT   174  716
CONECT   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



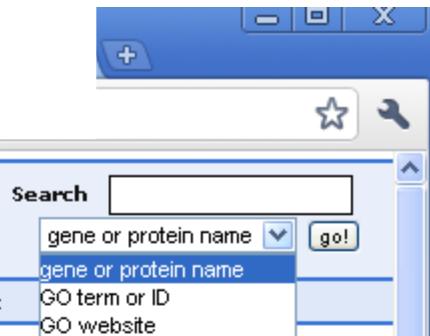
The Gene Ontology

gene ontology

www.geneontology.org

the Gene Ontology

Downloads Tools Documentation Projects About Contact



Search

gene or protein name go!

gene or protein name

GO term or ID

GO website

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:

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 [5P41HG002273-09](#)]. 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)

AmiGO: U2AF1 Assoc... x AmiGO: Cajal body De... x

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

 the Gene Ontology

AmiGO

Search Browse BLAST More Tools Help

Search GO terms genes or proteins exact match

Cajal body

Term information Term lineage External references 135 gene product associations

Term Information

Accession	GO:0015030
Ontology	cellular component
Synonyms	exact: coiled body exact: Gems
Definition	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 coolin 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: PMID:10944589 , PMID:11031238 , PMID:7559785 , http://genetics.cwru.edu/matera3.html]
Comment	None
Subset	None
Community	There have been 0 comments for this term. If you would like to view or participate in the community annotation, please continue to the CONITS page

AmiGOはGOのブラウザ: 例(下部)

A 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** Species: **Candida albicans**

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]

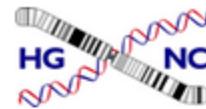
GO:0015030 : Cajal body [135 gene products]

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

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

HUGO Gene Nomenclature Committee



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.

 [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.

GCCTTGGT
CATGGAC
CCGCTCCC

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

See [Column definitions](#) for descriptions of the various data fields. Also see our [custom downloads](#) page for bulk access to our data

HGNC Comparison of Orth...  HUGO 

www.geneontology.org/data/humot_documentation.html



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 MGD 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

www.ncbi.nlm.nih.gov

NCBI Resources How To

My NCBI Sign In

NCBI National Center for Biotechnology Information

Search All Databases

Search Clear

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

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

- NCBI Help Manual
- NCBI Handbook
- Training & Tutorials

RESOURCES

- Literature
- DNA & RNA
- Proteins
- Sequence Analysis
- Genes & Expression
- Genomes & Maps
- Domains & Structures
- Genetics & Medicine

POPULAR

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

FEATURED

- GenBank
- Reference Sequences
- Map Viewer
- Genome Projects
- Human Genome
- Mouse Genome
- Influenza Virus
- Primer-BLAST
- Sequence Read Archive

NCBI INFORMATION

- About NCBI
- Research at NCBI
- NCBI Newsletter
- NCBI FTP Site

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

About Entrez Entrez Genome Help

Submitting Genome Project Genome sequence

Microbial Genome Projects

Genomic BLAST Microbial Eukaryotic

Resources Organelles Viruses

Archaea Chromosome Plasmid DraftAssembly

Bacteria Chromosome Plasmid DraftAssembly

Eukaryota Chromosome Plasmid Organelles

Viruses Phages

Viroids

All Plasmids

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

Total species (6596)

Category	Count
Viruses	2425
Eukaryota	2455
Archaea	102
Bacteria	1534
Plasmids	39

Total records (12475)

Category	Count
Viruses	3682
Eukaryota	1314 chromosomes
Bacteria	2520 chromosomes
Archaea	103 chromosomes
	2448 organelles
	2189 plasmids
	95 plasmids
Viroids	41
Plasmids	40

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

Click! 

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

Search

Search: [Select Group or Organism](#)

for:

Go

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, 2009An update to the human genome assembly and annotation is now... [more](#)**Annotation update released for human genome build 36** Mar 24, 2008An 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)

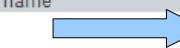
Mammals

(15)

Primates

(3)

Scientific name	Common name	Build	Tools
<i>Homo sapiens</i>	human	Build 37.1	
		Build 36.3	
<i>Macaca mulatta</i>	rhesus macaque	Build 1.1	
<i>Pan troglodytes</i>	chimpanzee	Build 2.1	



Rodents

(2)

Scientific name	Common name	Build	Tools
<i>Mus musculus</i>	laboratory mouse	Build 37.1	
		Build 36.1	
<i>Rattus norvegicus</i>	rat	RGSC v3.4	

▶ 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>		Build 1.1	
<i>Aspergillus fumigatus</i>		Build 2.1	
<i>Aspergillus niger</i>		Build 1.1	
<i>Candida glabrata</i>		Build 1.1	
<i>Cryptococcus neoformans</i>		Build 2.1	
<i>Debaryomyces hansenii</i>		Build 1.1	
<i>Encephalitozoon cuniculi</i>		Build 1.1	
<i>Eremothecium gossypii</i>		Build 3.1	
<i>Gibberella zeae</i>		Build 1.2	
<i>Kluyveromyces lactis</i>		Build 1.1	
<i>Magnaporthe oryzae</i>	rice blast fungus	Build 3.1	
<i>Neurospora crassa</i>		Build 2.1	
<i>Saccharomyces cerevisiae</i>	baker's yeast	Build 2.1	
<i>Scheffersomyces stipitis</i>		Build 1.1	
<i>Schizosaccharomyces pombe</i>	fission yeast	Build 1.1	
<i>Ustilago maydis</i>		Build 1.1	
<i>Yarrowia lipolytica</i>		Build 1.1	

Map viewer Top画面

U2AF1
と入力

Entrez Genome view x Map Viewer x
www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9606

NCBI Map Viewer

PubMed Nucleotide Protein Genome Gene Structure PopSet Taxonomy Help

Search for on chromosome(s) assembly All Find Advanced Search

Map Viewer Map Viewer Home Map Viewer Help Human Maps Help Release Notes NCBI Resources Genome Project TaxPlot Consensus Coding Sequence (CCDS) GRC Human Genome Resources NCBI Handbook RefSeq Trace Archive (Watson) Trace Archive (Venter) Trace FTP (Personal Genomics) Whole Genome Association (WGA) Organism Data in GenBank EST Genomic mRNA Protein WGS Sequencing Projects Helix Kazusa MHC Project NERD

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

BLAST search the human genome

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

Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo; Homo sapiens

August 2009: NCBI released an updated version of the human genome reference genome assembly and updated annotation for all available assemblies. The reference assembly update includes modifications to all chromosomes and adds 9 alternate loci to the reference assembly definition; the updated assembly, named GRCh37, was provided by the [Genome Reference Consortium \(GRC\)](#). The previous version of the reference genome assembly, [NCBI Build 36.3](#), can still be accessed for Map Viewer display and for BLAST. For additional information about changes, statistics, and the status of the CCDS project please refer to:

- [Release Notes](#)
- [Statistics](#)
- [CCDS Project](#)

The NCBI Map Viewer provides graphical displays of features on the human genome sequence assembly as well as cytogenetic, genetic, physical, and radiation hybrid maps. Extensive [documentation](#) is provided to describe the resource features and methods used, tutorials, and statistics.

Map features that can be seen along the sequence include genes, transcripts, [NCBI contigs](#) (the 'Contig' map), the BAC tiling path (the 'Component' map), STSs, FISH mapped clones, ESTs and transcripts from several different organisms, [Gnomon](#) predicted gene models, and more.

You can find genes or markers of interest by submitting a query against the whole genome, or a chromosome at a time. Use the Advanced Search form for more complex queries. Results are indicated both graphically, as tick marks on the ideogram, and in a tabular format. The results table includes links to a chromosome graphical view where the gene or marker can be seen in the context of additional data. For genes, a particularly useful display includes the Gene, Ab initio (e.g., the Gnomon predicted models), and UniGene sequence maps. You can also browse a chromosome by clicking on a chromosome link in the ideogram above. Use the "Maps & Options" window, available on individual chromosome displays, to configure your display.

Please note that other genomes can also be viewed in the NCBI Map Viewer resource. The [Map Viewer Home Page](#) provides a current list. Additional organism-specific web pages are listed on the [Genomic Biology](#) site.

Available Documentation:

Entrez Genome view > Map Viewer > www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9606&query=U2AF1&qchr=&strain=All

NCBI Map Viewer

PubMed Nucleotide Protein Genome Gene Structure PopSet Taxonomy Help

Search for U2AF1 on chromosome(s) assembly reference Find Advanced Search

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

BLAST search the human genome

多すぎなので、referenceに絞る

Hits: 1 2 3 4 5 6 7 8 9 10 11 12 13

Hits: 14 15 16 17 18 19 20 21 22 X Y MT

Search results for query "U2AF1": 56 hits

Chr Assembly Match Map element Type Maps

11	reference	all matches	Gallus gallus U2 small nuclear...U2AF1, mRNA Gallus gallus mRNA for U2snRNP...U2AF1 gene Gallus gallus mRNA for U2snRNP...U2AF1 gene	NM_204655.2 AJ291765.1 AJ291764.1	TRANSCRIPT	Gga RNA
11	Celera	all matches	Gallus gallus U2 small nuclear...U2AF1, mRNA Gallus gallus mRNA for U2snRNP...U2AF1 gene Gallus gallus mRNA for U2snRNP...U2AF1 gene	NM_204655.2 AJ291765.1 AJ291764.1	TRANSCRIPT	Gga RNA
11	HuRef	all matches	Gallus gallus U2 small nuclear...U2AF1, mRNA Gallus gallus mRNA for U2snRNP...U2AF1 gene Gallus gallus mRNA for U2snRNP...U2AF1 gene	NM_204655.2 AJ291765.1 AJ291764.1	TRANSCRIPT	Gga RNA
21	reference	all matches	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	ENST00000457176 ENST00000398137 ENST00000380276 ENST00000291552 NM_024187.3 AJ291765.1 AJ291763.1 NM_001080268.1 CV865602.1 DN994470.1 AJ627978.1	TRANSCRIPT	ensRNA ensRNA ensRNA ensRNA Mm RNA Gga RNA Gga RNA Bt RNA Ssc RNA Hs RNA Hs RNA

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 U2AF1 on chromosome(s) assembly reference Find Advanced Search

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

Hits: 1 2 3 4 5 6 7 8 9 10 11 12 13

Hits: 14 15 16 17 18 19 20 21 22 8 Y MT

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

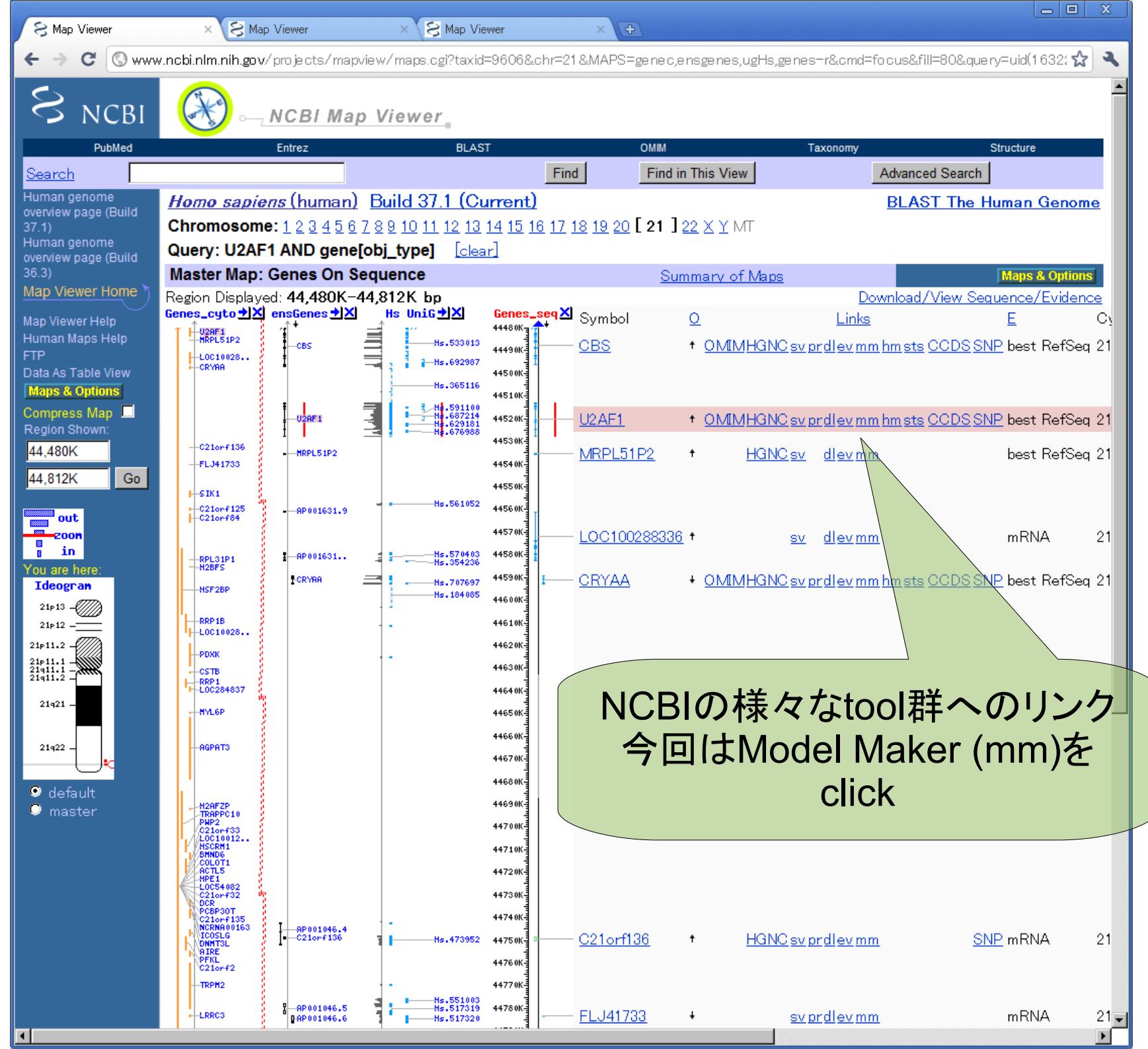
Chr	Match	Map element	Type	Map
11	all matches	NM_204655.2	TRANSCRIPT	Gga RNA
	Gallus gallus U2 small nuclear...U2AF1), mRNA	AJ291765.1	TRANSCRIPT	Gga RNA
	Gallus gallus mRNA for U2snRNP...U2AF1 gene)	AJ291764.1	TRANSCRIPT	Gga RNA
21	all matches	ENST00000457176	TRANSCRIPT	ensRNA
	ENST00000398137	ENST00000398137	TRANSCRIPT	ensRNA
	ENST00000380276	ENST00000380276	TRANSCRIPT	ensRNA
	ENST00000291552	ENST00000291552	TRANSCRIPT	ensRNA
	Mus musculus U2 small nuclear...U2af1), mRNA	NM_024187.3	TRANSCRIPT	Mm RNA
	Gallus gallus mRNA for U2snRNP...U2AF1 gene)	AJ291765.1	TRANSCRIPT	Gga RNA
	Gallus gallus mRNA for U2snRNP...U2AF1 gene)	AJ291763.1	TRANSCRIPT	Gga RNA
	Bos taurus U2 small nuclear...U2AF1), mRNA	NM_001080268.1	TRANSCRIPT	Bt RNA
	PDUts1019H11 Porcine testis...U2AF1), mRNA, mRNA sequence	CV865602.1	TRANSCRIPT	Ssc RNA
	TC115911 Human adult whole...U2AF1), mRNA sequence	DN994470.1	TRANSCRIPT	Hs RNA
	Homo sapiens mRNA for U2 snRNP...U2AF1 gene), splice variant...	AJ627978.1	TRANSCRIPT	Hs RNA
	Homo sapiens U2 small nuclear...U2AF1), transcript variant...	NM_006758.2	TRANSCRIPT	RefSeq RNA
	Homo sapiens U2 small nuclear...U2AF1), transcript variant...	NM_001025204.1	TRANSCRIPT	RefSeq RNA
	Homo sapiens U2 small nuclear...U2AF1), transcript variant...	NM_001025203.1	TRANSCRIPT	RefSeq RNA
	Homo sapiens U2AF1 mRNA for U2 small nuclear RNA auxiliary...	AB451244.1	TRANSCRIPT	Hs RNA
	U2AF1 : U2 small nuclear RNA auxiliary factor 1	U2AF1	GENE	Genes cyto Genes sea
	U2AF1 : ENSG00000160201	U2AF1	GENE	ensGenes

Quick Filter
 Gene
 Transcript :
 all
 RefSeq

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。

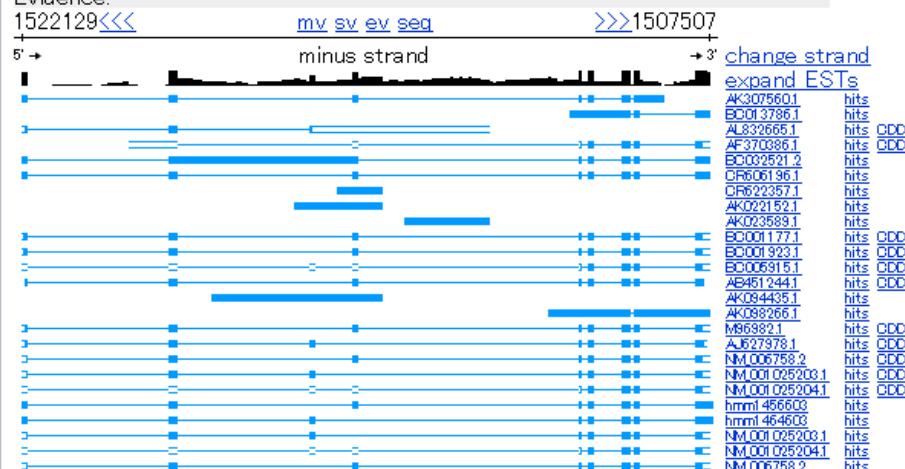
Geneにcheckを入れ
た後に、ここをclick



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

Organism: Homo sapiens (Build 37.1) Chromosome: 21 Contig: [NT_011515.12](#) Locus: [U2AF1](#)

Evidence:



Putative exons (graphic view):



Your model:

[clear](#)

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

Frame1, ORF= CDD Frame2, ORF= CDD Frame3, ORF= CDD



Putative exons (table view): [custom exons](#) [intron bases](#): 2

- 1 1522143 GG|AAG...TC|CGG 1522126-1522046 GAA|AT => [2](#) or [4](#)
- 2 [1](#) <= GG|AAG 1522143-1522002 CAA|GT => [5](#)
- 3 1522129 GC|GTC...GT|CGA 1522073-1522002 CAA|GT => [7](#) or [8](#)
- 4 [1](#) <= TC|CGG 1522126-1522002 CAA|GT => [5](#)
- 5 [2](#) or [4](#) <= AA|ATG 1522045-1522002 CAA|GT => [7](#)
- 6 TT|CCC 1519803-1518866 CAG|GT => [12](#)
- 7 [3](#) or [5](#) <= AG|AGT 1518953-1518866 CAG|GT => [10](#) or [11](#) or [12](#)
- 8 [3](#) <= AG|AGT 1518953-1515004 GCT|GT => [15](#)
- 9 1518037 TC|TTT...TT|TCC 1515405-1514524 TCC|AG...TCC|AA 1514493
- 10 [7](#) <= AG|ACC 1515983-1515917 ACT|GT => [12](#) or [15](#)
- 11 [7](#) <= AG|ACC 1515983-1512191 CTC|AA
- 12 [6](#) or [7](#) or [10](#) <= AG|ACC 1515070-1515004 GCT|GT => [15](#)
- 13 TT|AGA 1513959-1512191 CTC|AA

ユーザーが自由にgene modelを組み立てるtool

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

clear

ORF Finder

Save

Frame1, ORF=75 CDD Frame2, ORF=44 CDD Frame3, ORF=25 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 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 1518003-1518886 CAG GT => 12	
<input checked="" type="checkbox"/> 7	3 or 5 <= AG AGT 1518953-1518886 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 1507800-1507653 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 1507652-1507505 ATT CT	

Disclaimer | Write to the Help Desk

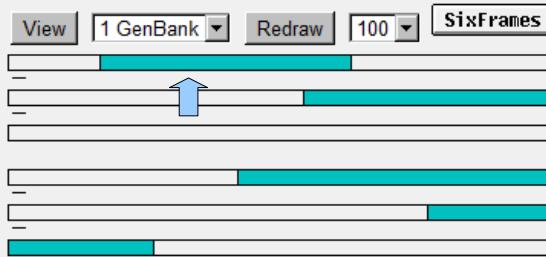
NCBI | NLM | NIH

gene modelの構築：
転写物のevidenceとエ
クソンの組み合わせに
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。

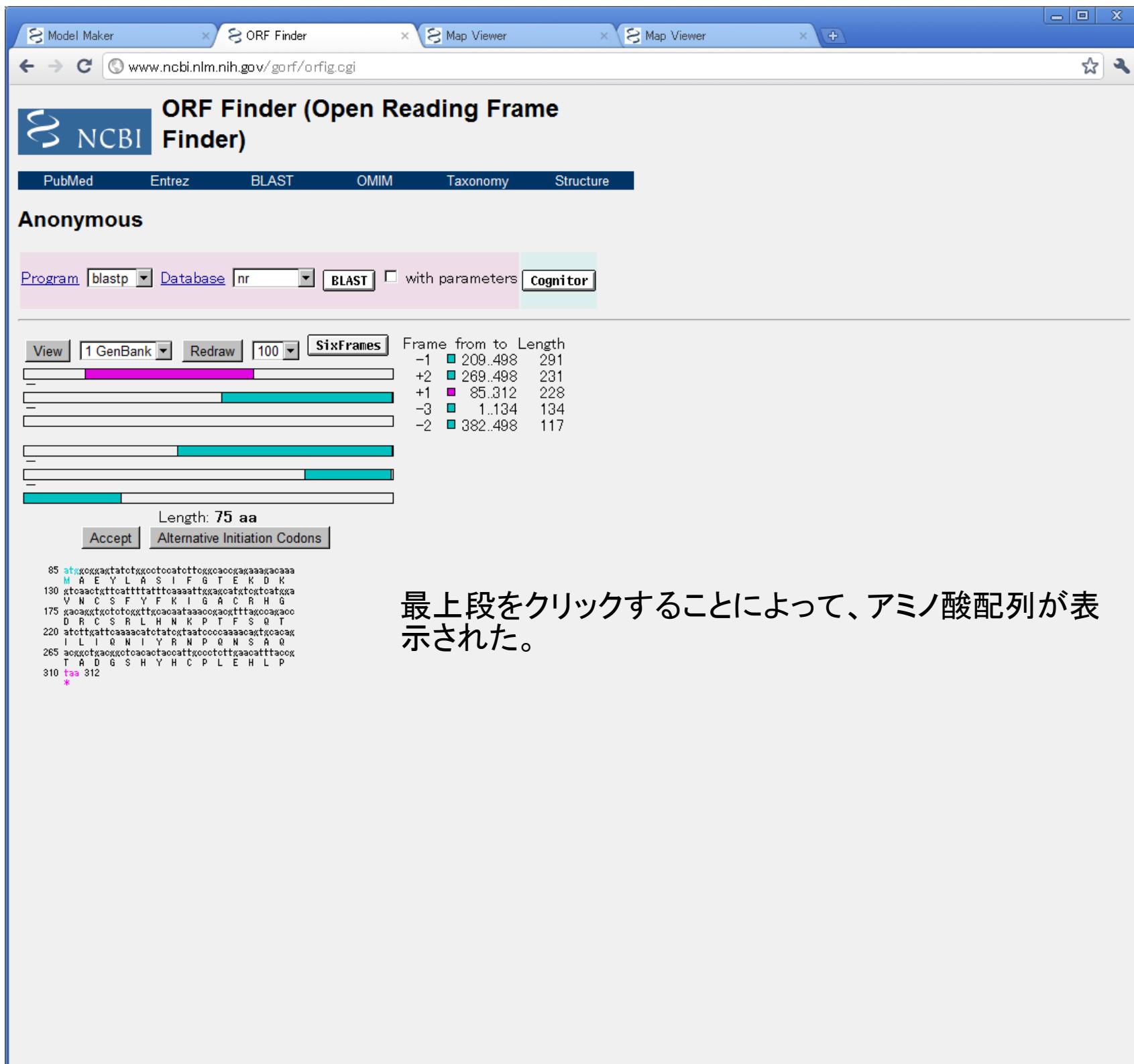
ORF Finder (Open Reading Frame Finder)

PubMed Entrez BLAST OMIM Taxonomy Structure

Anonymous



Frame	from	to	Length
-1	209	498	291
+2	269	498	231
+1	85	312	228
-3	1	134	134
-2	382	498	117



Ensembl (www.ensembl.org)

e!Ensembl West 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

-- Select a species --

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



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



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.



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

Species List in Ensembl

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

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

Ensemblのdocument一覧



Ensembl: genome browserの例

Ensembl genome browser screenshot showing the "Region in detail" view for chromosome 21.

The URL in the browser bar is: uswest.ensembl.org/Homo_sapiens/Location/View?db=core;g=ENSG00000160201;r=21:44513066-44527697

Region in detail view for chromosome 21, band q22.3 (44.50 Mb to 44.55 Mb).

Region overview and **Region in detail** tabs are visible.

Gene Legend:

- processed transcript (blue)
- pseudogene (grey)
- merged Ensembl/Havana (yellow)
- RNA gene (purple)

Location: 21 : 44513066 - 44527697

Region in detail view for the U2AF1 gene (AP001631.1) on the forward strand.

Key features shown in the "Region in detail" view:

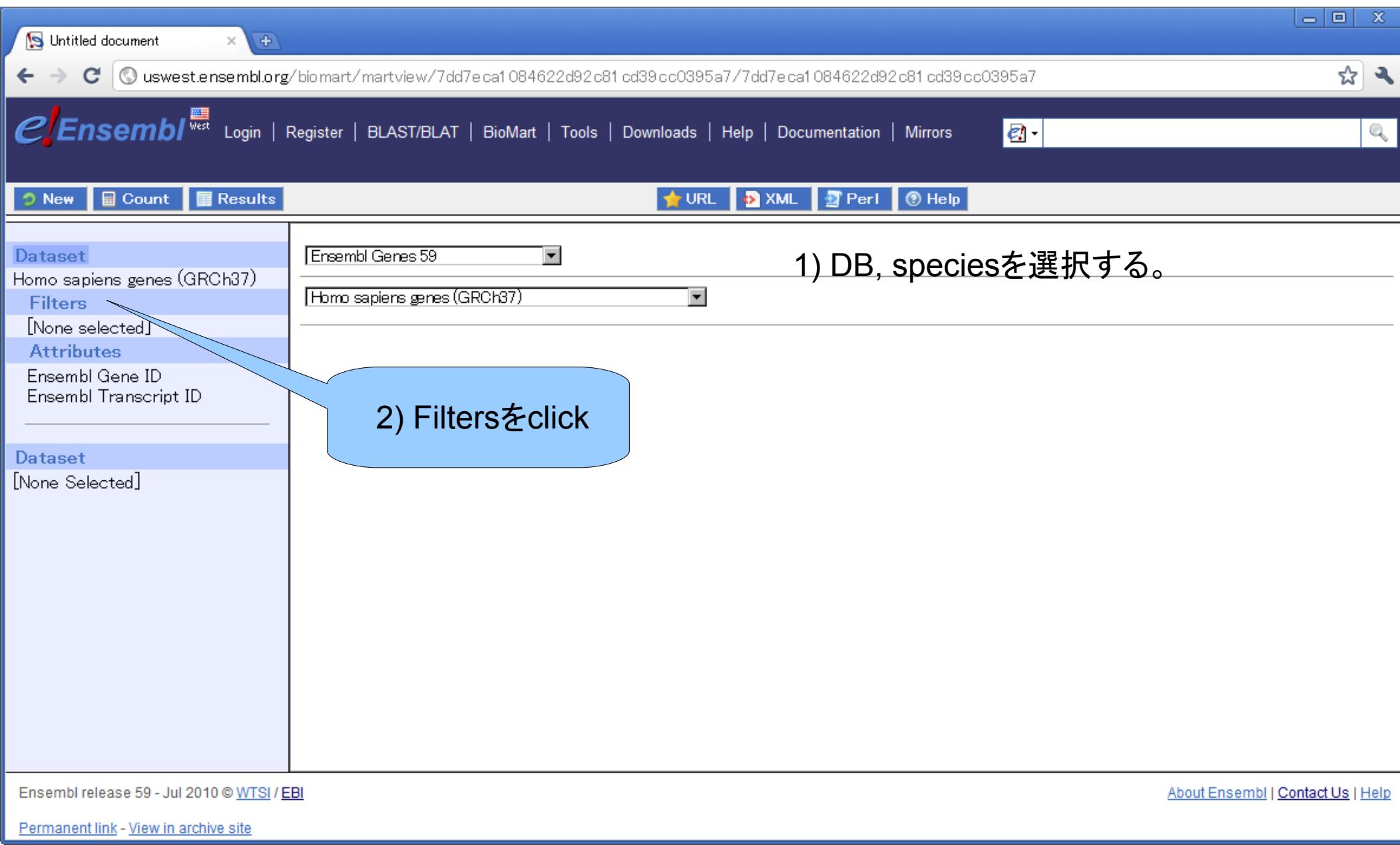
- Genomic tracks for chromosome bands, contigs, and Ensembl/Havana genes.
- Gene tracks for AP001628.1, AP001630.1, and AP001046.1.
- ncRNA gene tracks for AP001626.1, AP001627.1, WDR4, C2orf105, CBS, U2AF1, AP001631.10, PDE9A, NDUFV3, PKNOX1, AP001626.2, AP001628.7, AP001630.5, AP001628.6, AP001629.1, and AP001630.1.
- All Structural variants track.
- Gene Legend: processed transcript (blue), pseudogene (grey), merged Ensembl/Havana (yellow), RNA gene (purple).

Region in detail view for the U2AF1 gene (AP001631.1) on the forward strand.

Key features shown in the "Region in detail" view:

- Genomic tracks for chromosome bands, contigs, and Ensembl/Havana genes.
- Gene tracks for AP001631.1.
- Exon-intron structure of the U2AF1 gene, showing three protein coding exons (U2AF1-001, U2AF1-002, U2AF1-003) and a non-coding exon (U2AF1-000).

BioMart top page



Untitled document

uswest.ensembl.org/biomart/martview/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7

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

New Count Results URL XML Perl Help

Dataset
Homo sapiens genes (GRCh37)

Filters
[None selected]

Attributes
Ensembl Gene ID
Ensembl Transcript ID

Dataset
[None Selected]

Ensembl Genes 59

Homo sapiens genes (GRCh37)

1) DB, speciesを選択する。

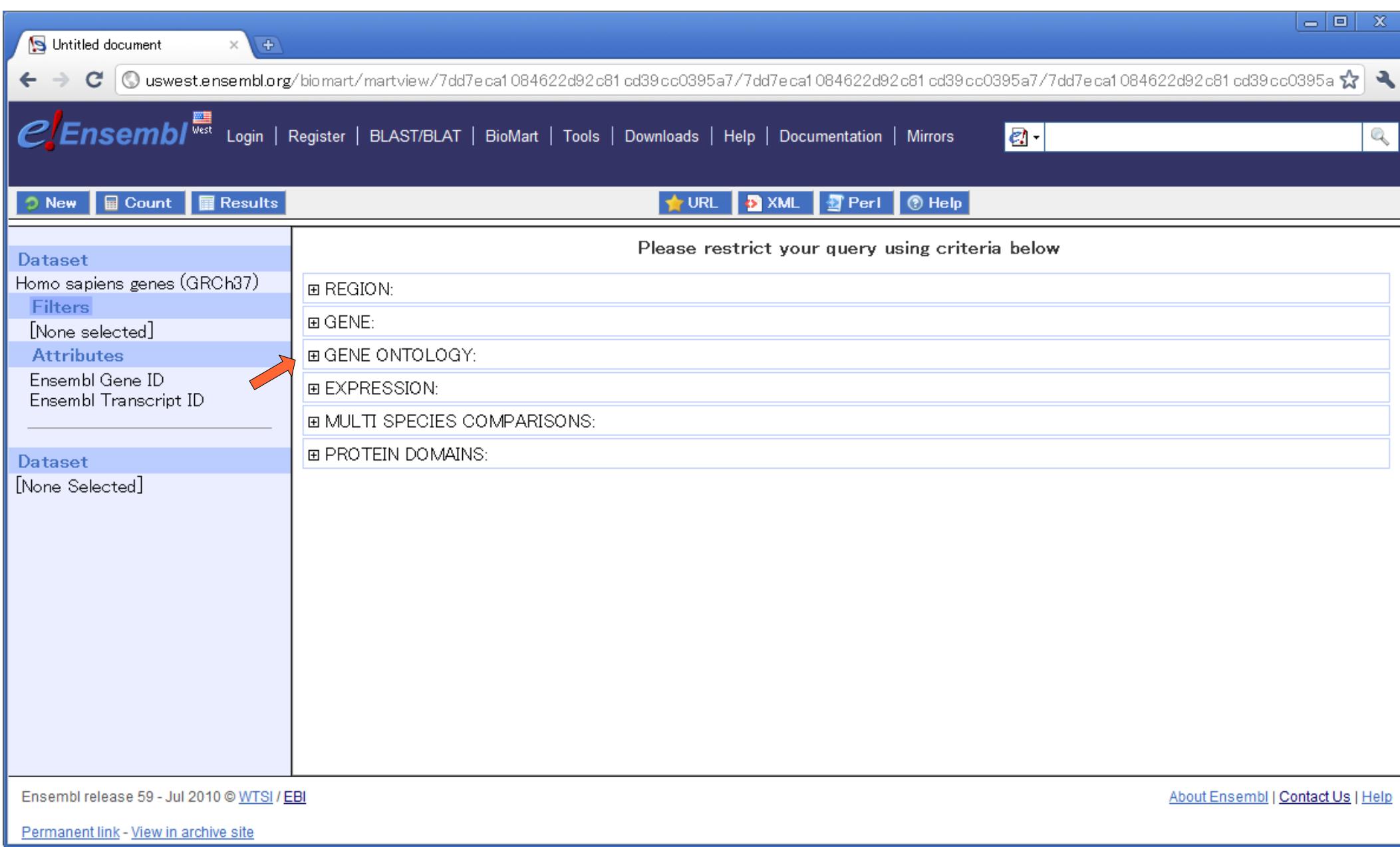
2) Filtersをclick

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. On the left, there is a sidebar with sections for 'Dataset' (Homo sapiens genes (GRCh37)), 'Filters' (None selected), and 'Attributes' (Ensembl Gene ID, Ensembl Transcript ID). A red arrow points to the 'Attributes' section. On the right, the main query form is displayed with the heading 'Please restrict your query using criteria below'. The criteria sections include: REGION, GENE, GENE ONTOLOGY, EXPRESSION, MULTI SPECIES COMPARISONS, and PROTEIN DOMAINS. At the bottom of the page, there is footer text: 'Ensembl release 59 - Jul 2010 © WTSI / EBI', 'About Ensembl | Contact Us | Help', and 'Permanent link - View in archive site'.

Untitled document

uswest.ensembl.org/biomart/martview/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7

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

New Count Results URL XML Perl Help

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

Untitled document

uswest.ensembl.org/biomart/martview/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7

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

New Count Results URL XML Perl Help

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]

Please restrict your query using criteria below

1) 絞り込む

2) Countをclick

3) Attributesをclick

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

Untitled document

uswest.ensembl.org/biomart/martview/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7

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

New Count Results Resultsをclick URL XML Perl Help

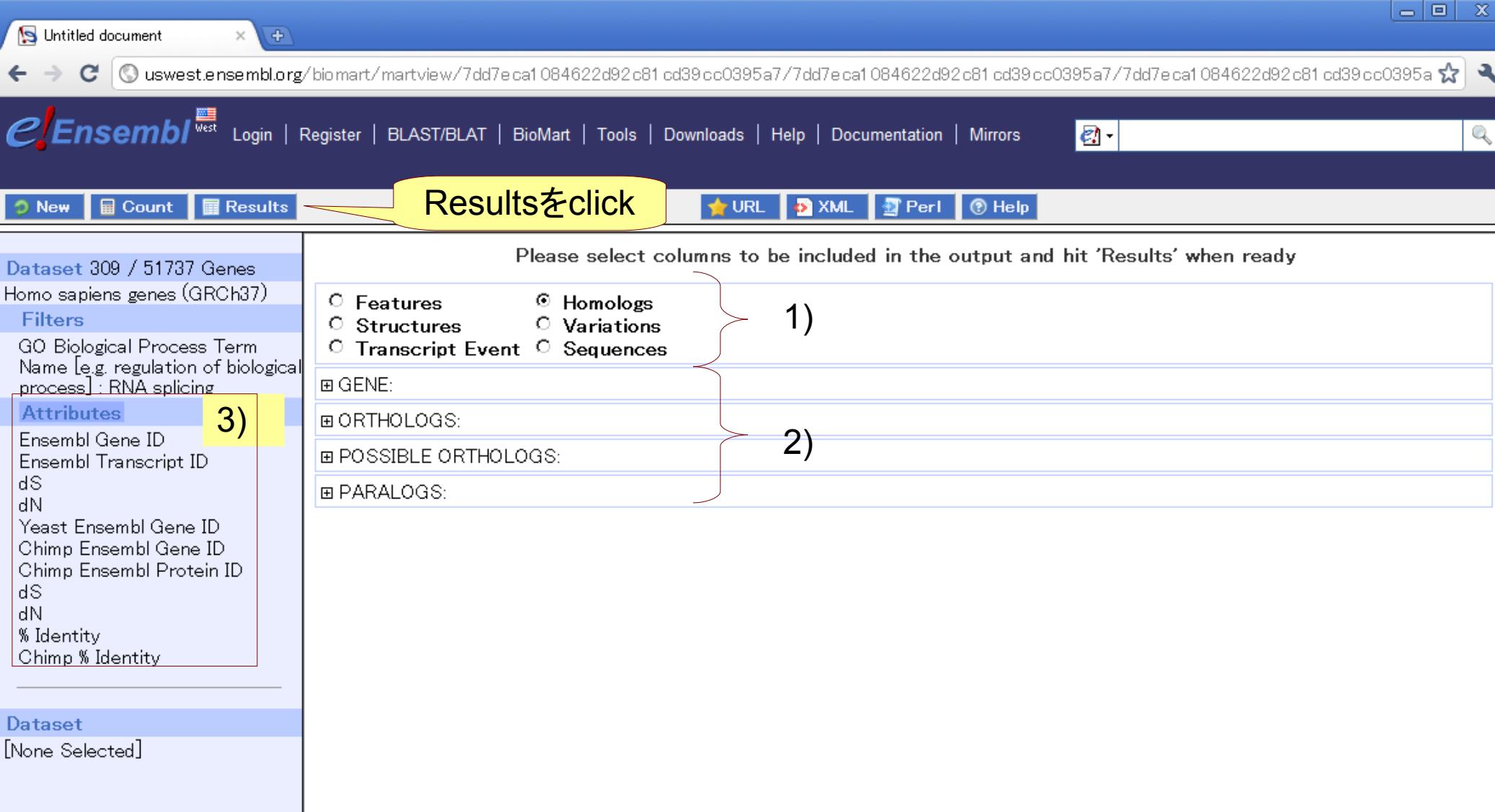
Dataset 309 / 51737 Genes
Homo sapiens genes (GRCh37)
Filters
GO Biological Process Term
Name [e.g. regulation of biological process] : RNA splicing
Attributes 3)
Ensembl Gene ID
Ensembl Transcript ID
dS
dN
Yeast Ensembl Gene ID
Chimp Ensembl Gene ID
Chimp Ensembl Protein ID
dS
dN
% Identity
Chimp % Identity

Please select columns to be included in the output and hit 'Results' when ready

Features Homologs
 Structures Variations
 Transcript Event Sequences

GENE:
 ORTHOLOGS:
 POSSIBLE ORTHOLOGS:
 PARALOGS:

1)
2)



BioMart: Result表示例

Untitled document

uswest.ensembl.org/biomart/martview/7dd7eca1084622d92c81cd39cc0395a7/7dd7eca1084622d92c81cd39cc0395a7

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

New Count Results

URL XML Perl Help

HTML TSV(タブ区切り) CSV(コンマ区切り) XML

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
Yeast Ensembl Gene ID
dS
dN
Chimp Ensembl Gene ID
dS
dN
Chimp Ensembl Protein ID
% Identity
Chimp % Identity

Dataset
[None Selected]

Export all results to File TSV Unique results only Go

Email notification to

View 150 rows as HTML Unique results only

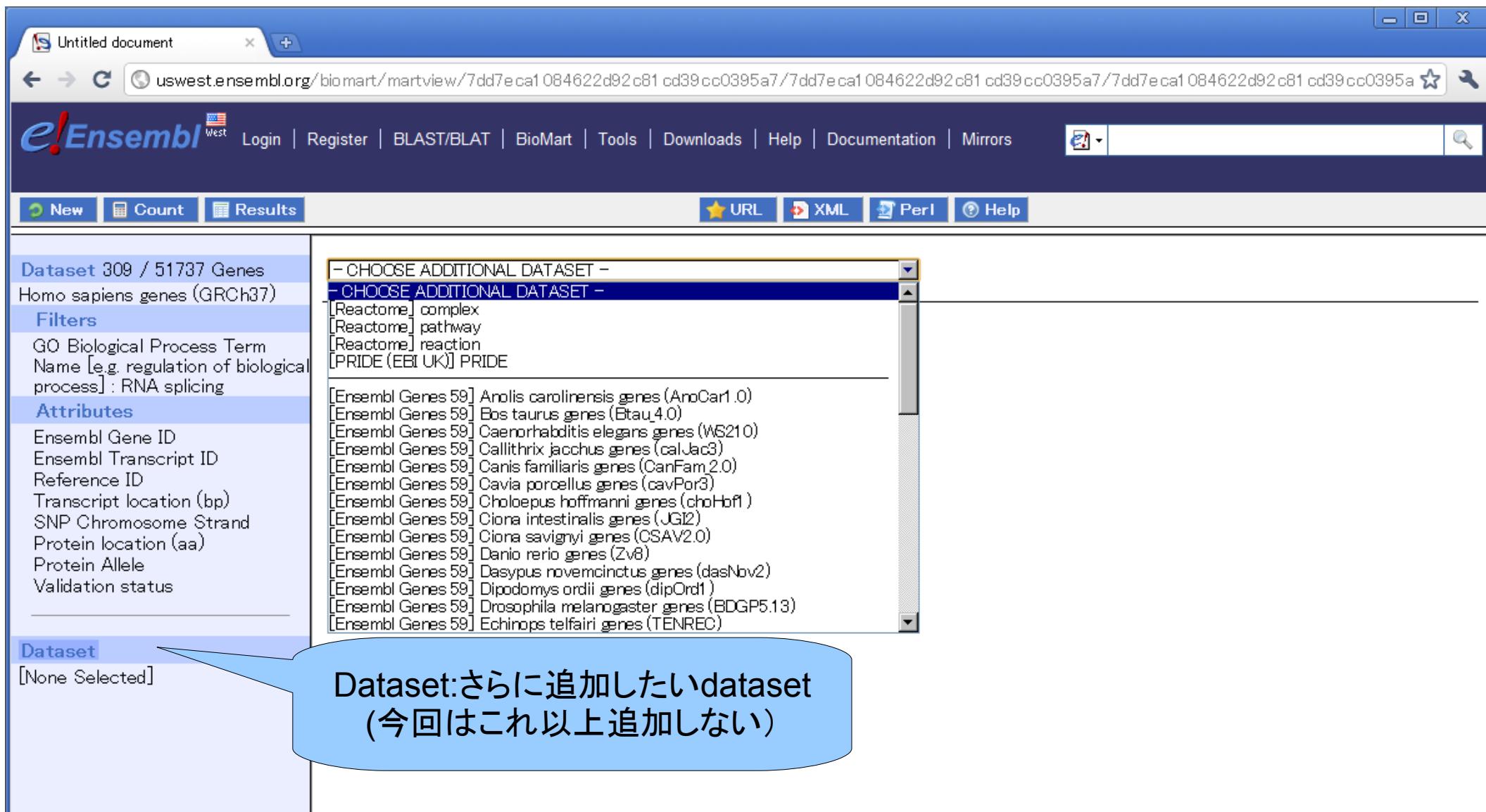
Ensembl Gene ID	Ensembl Transcript ID	Yeast Ensembl Gene ID	dS	dN	Chimp Ensembl Gene ID	dS	dN	Chimp Ensembl Protein ID	% Identity	Chimp % Identity
ENSG00000215699	ENST00000374449				ENSPTRG00000024234	0.01220	0.00000	ENSPTRP00000052680	100	100
ENSG00000215699	ENST00000374449				ENSPTRG00000024234	0.01220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



The screenshot shows the Ensembl BioMart interface. On the left, there is a sidebar with sections for 'Dataset' (309 / 51737 Genes, Homo sapiens genes (GRCh37)), 'Filters' (GO Biological Process Term: RNA splicing), and 'Attributes' (Ensembl Gene ID, Ensembl Transcript ID, Reference ID, Transcript location (bp), SNP Chromosome Strand, Protein location (aa), Protein Allele, Validation status). At the bottom of the sidebar, there is a 'Dataset' section with the text '[None Selected]'. A blue callout bubble points to this section with the Japanese text 'Dataset:さらに追加したいdataset (今回はこれ以上追加しない)'. On the right, there is a dropdown menu titled '- CHOOSE ADDITIONAL DATASET -' containing a list of datasets: 'Reactome complex', 'Reactome pathway', 'Reactome reaction', 'PRIDE (EBI UK) PRIDE', and a long list of Ensembl Genes datasets for various species: Anolis carolinensis, Bos taurus, Caenorhabditis elegans, Callithrix jacchus, Canis familiaris, Cavia porcellus, Chloepus hoffmanni, Ciona intestinalis, Ciona savignyi, Danio rerio, Dasypus novemcinctus, Dipodomys ordii, Drosophila melanogaster, and Echinops telfairi.

UCSC Genome Browser

genome.ucsc.edu

The screenshot shows the UCSC Genome Bioinformatics website at genome.ucsc.edu/index.html?org=Human&db=hg18&hgSID=1. The 'Genomes' menu is highlighted with a red box. Three green callout bubbles provide Japanese translations for the 'Genome Browser' features:

- 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 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 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

Genomes - Blat - Tables - Gene Sorter - PCR - VisiGene - Proteome - Session - FAQ - Help

Genome Browser

ENCODE

Neandertal

Blat

Table Browser

Gene Sorter

In Silico PCR

Genome Graphs

Galaxy

VisiGene

Proteome Browser

Utilities

Downloads

Release Log

Custom Tracks

Microbial Genomes

Mirrors

Archives

Training

About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Bioinformatics website. This site contains the genome assembly and a large collection of genomic data, also provides portals to other databases.

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 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 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.

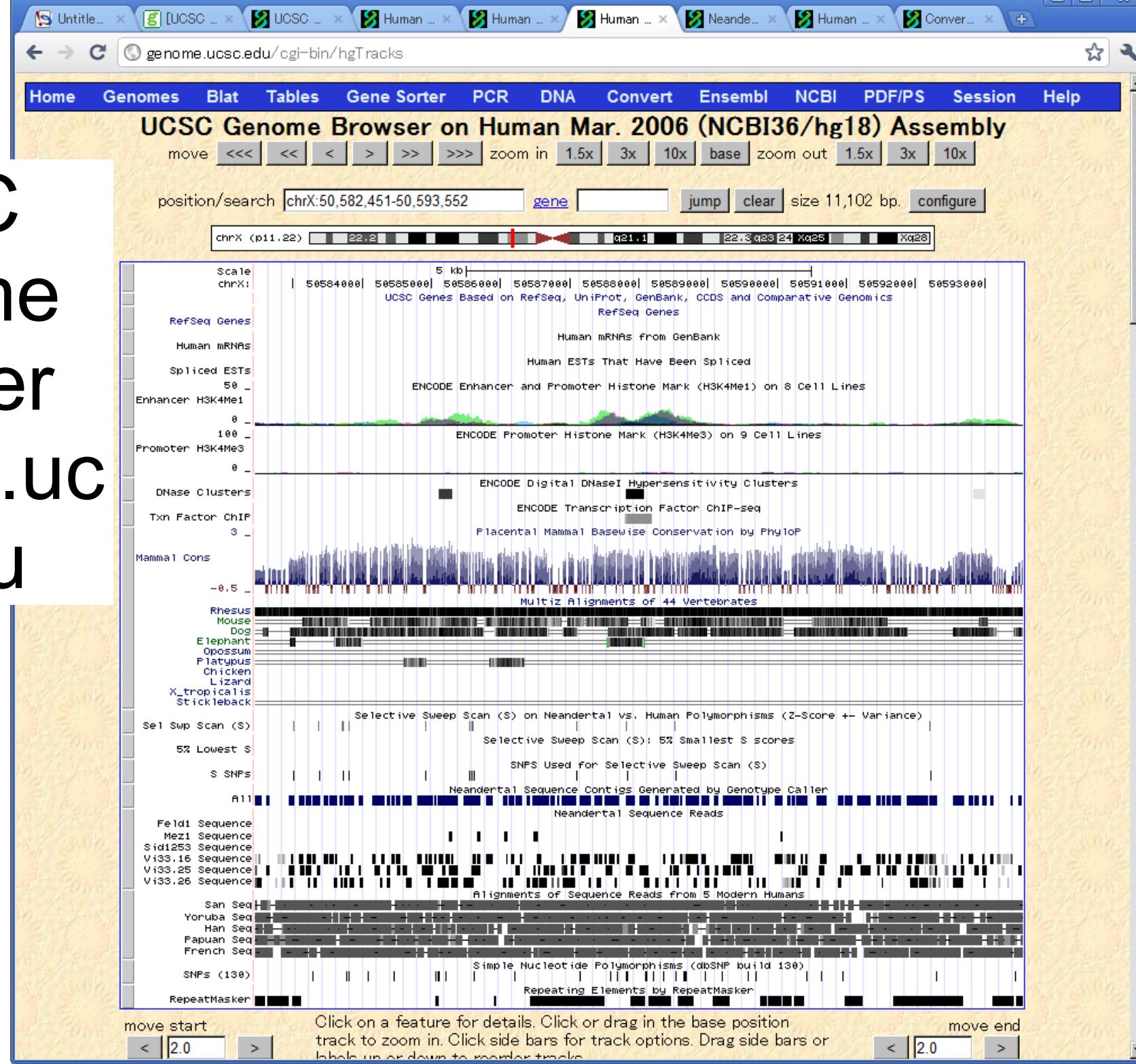
Thanks to Tim Dreszer and the rest of the team for implementing and testing this feature.

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



GenomeBrowser で見られる項目

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

Tracks: [hide all](#) [show all](#) [default](#) Groups: [collapse all](#) [expand all](#)

Control track and group visibility more selectively below.

+ [Mapping and Sequencing Tracks](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [Phenotype and Disease Associations](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [Genes and Gene Prediction Tracks](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [mRNA and EST Tracks](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [Expression](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [Regulation](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [Comparative Genomics](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [Neandertal Assembly and Analysis](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [Variation and Repeats](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [Pilot ENCODE Regions and Genes](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [Pilot ENCODE Transcription](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [Pilot ENCODE ChIP](#) [hide all](#) [show all](#) [default](#) [submit](#)

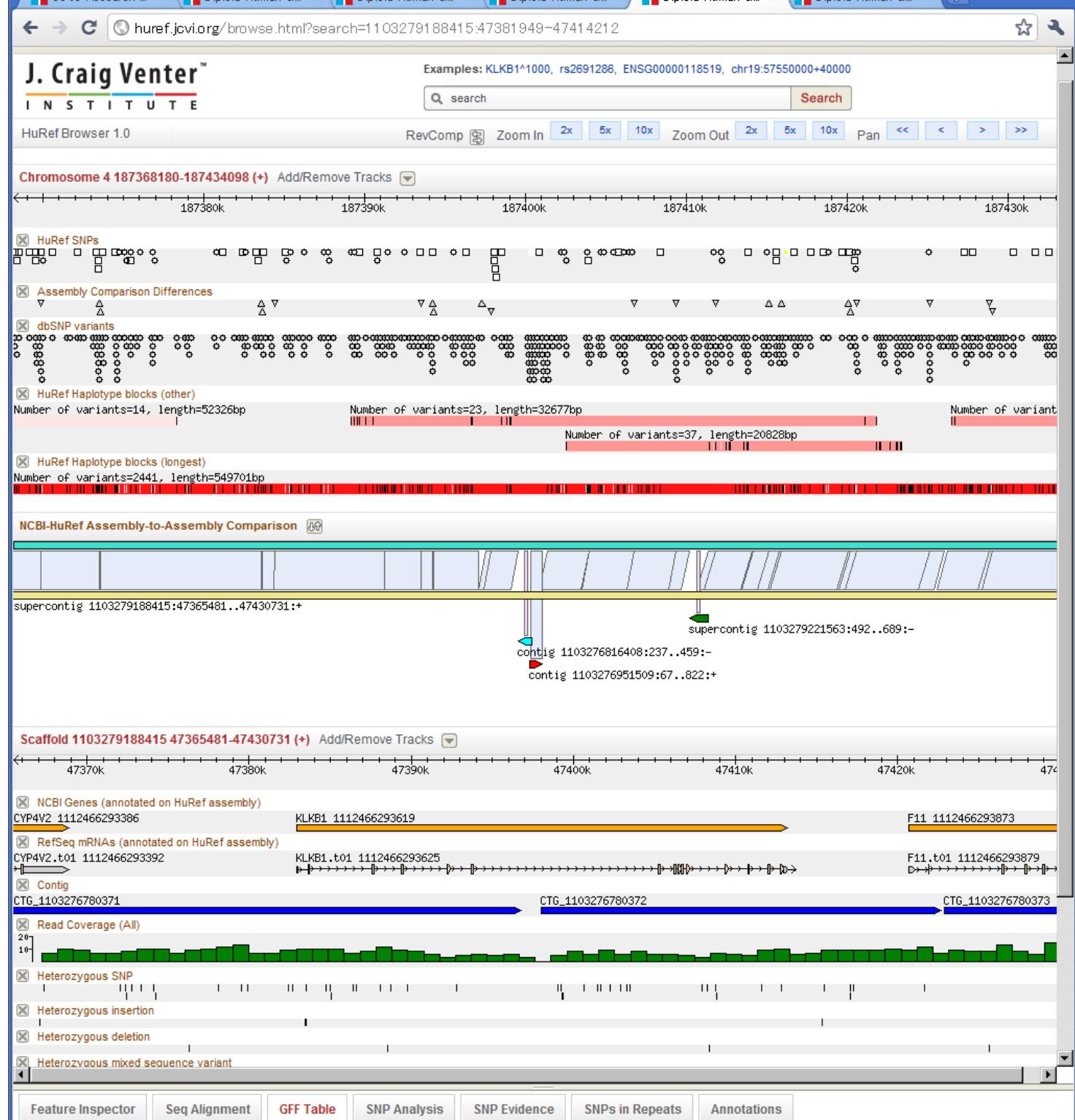
+ [Pilot ENCODE ChIP-seq](#) [hide all](#) [show all](#) [default](#) [submit](#)

+ [Pilot ENCODE Comparative Genomics and Variation](#) [hide all](#) [show all](#) [default](#) [submit](#)

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

List of UCSC genome releases

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

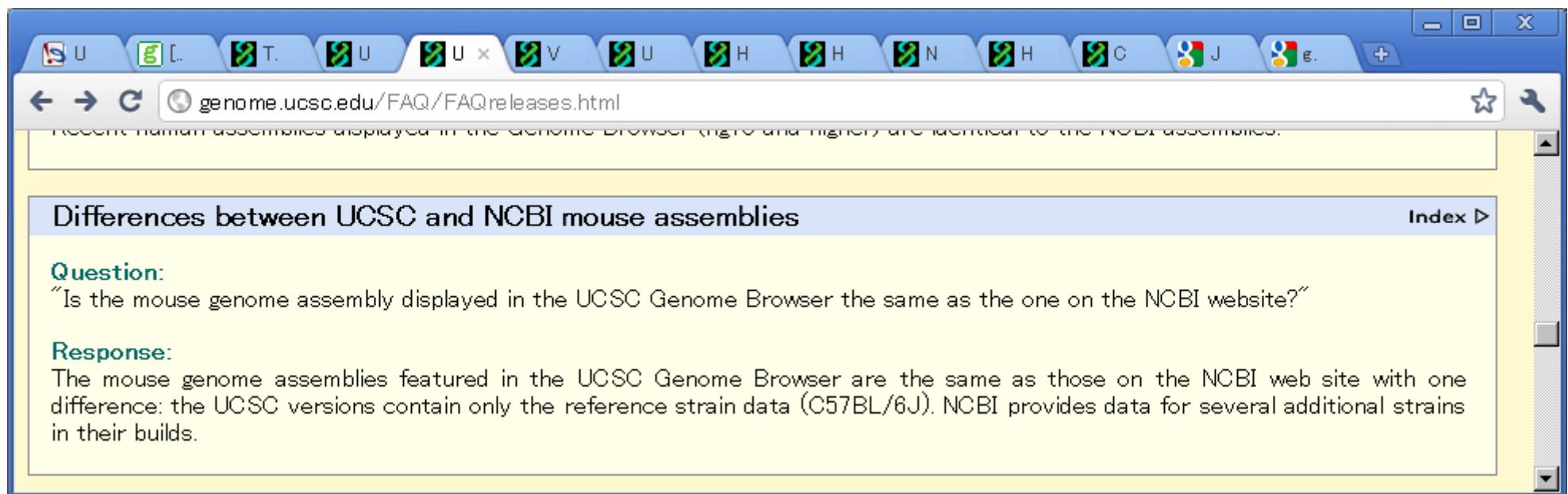
Question:

"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
VERTEBRATES				
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



A screenshot of a web browser window. The address bar shows the URL genome.ucsc.edu/FAQ/FAQreleases.html. The page content is titled "Differences between UCSC and NCBI mouse assemblies". It contains a "Question" section asking if the mouse genome assembly in the UCSC Genome Browser is the same as the one on the NCBI website, and a "Response" section stating that the UCSC versions contain only the reference strain data (C57BL/6J), while NCBI provides data for several additional strains in their builds.

Recent human assemblies displayed in the Genome Browser (light blue) are identical to the NCBI assemblies.

Differences between UCSC and NCBI mouse assemblies

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/lsdb.cgi?pg=1>