

2009/9/26 ゲノムリテラシー講座

# バイオインフォマティクスの基礎： 分子生物学データベース・ 分子系統解析

村上 勝彦

(社)バイオ産業情報化コンソーシアム  
産業技術総合研究所  
生物情報解析研究センター

# はじめに

- 日本バイオインフォマティクス学会主催の**バイオインフォマティクス技術者認定試験の受験を念頭に**

- 出題範囲**

- <http://www.jsbi.org/modules/jsbi/index.php/nintei/index.html>

- 赤線: 本日扱う範囲**

生命科学分野	
生物学基礎	細胞(細胞、細胞内小器官、真核生物、原核生物)、細胞周期(体細胞分裂、減数分裂)、遺伝学(メンデルの法則、連鎖、Hardy-Weinbergの法則)
分子生物学・生化学	セントラルドグマ、複製(DNA)、転写(RNA、遺伝子発現、逆転写、スプライシング)、翻訳(タンパク質、tRNA、リボソーム、遺伝暗号、二次構造、三次構造、四次構造、翻訳後修飾)、シグナル伝達(受容体、シグナル変換)
バイオテクノロジー	ゲノム(ショットガン法)、トランスクリプトーム(マイクロアレイ)、プロテオーム(質量分析、二次元電気泳動、酵母ツーハイブリッド法)、立体構造決定(X線結晶構造解析、NMR、電子顕微鏡)
情報科学分野	
コンピュータシステム	二進数、論理演算(論理積、論理和、排他的論理和)、数値演算(浮動小数点数、丸め誤差、桁落ち)、プログラミング言語(C言語、Perl言語)、ネットワーク(OSI参照モデル、IPアドレス)、マークアップ言語(HTML、XML)
アルゴリズム	データ構造(スタック、キュー、リスト、木構造)、探索(二分探索、ハッシュ表、木探索)、ソート、文字列比較
データベース技術	3層スキーマ構造、データモデル(階層型、リレーショナル型)
確率・統計	確率分布・確率変数(母集団、標本、独立性、ベイズの定理)、データ解析(平均、分散、相関、回帰)、推定・検定(点推定、区間推定、帰無仮説、対立仮説、有意水準)
認識と学習	決定木、k-最近接近傍法、ニューラルネットワーク、サポートベクターマシン、感度、特異度、クロスバリデーション法、クラスタリング、K-平均法、自己組織化マップ(SOM)
バイオインフォマティクス	
分子生物学データベース	文献DB(PubMed)、ゲノムDB、核酸配列DB、アミノ酸DB、モチーフDB(モチーフライブラリー)、立体構造DB、代謝パスウェイDB、多型DB、発現DB、アノテーション、遺伝子オントロジー(Gene ontology)
配列解析	アライメント(動的計画法(dynamic programming)、スコアテーブル、ギャップペナルティ、ローカルアライメント、グローバルアライメント、Smith-Waterman法、ベアウィズアライメント、マルチプルアライメント、累進法(ツリーベース法)、ClustalW、HMM(隠れマルコフモデル))、相同性検索(FASTA、ハッシング、BLAST、有限オートマトン、PSI-BLAST、位置特異的スコア行列(PSSM)、プロファイル比較)、モチーフ解析(正規表現、重み行列)、分子系統解析(オーソログ、パラログ、距離行列法、UPGMA、近隣結合法(N-J法)、最節約法、最尤法、同義置換、非同義置換)、タンパク質機能予測(膜貫通部位予測、細胞内局在部位予測)、RNA二次構造予測
タンパク質立体構造解析	立体構造表現(コンタクトマップ、ラマチャンドランマップ)、構造比較(重ね合わせ、RMSD、構造アライメント、構造モチーフ、構造分類)、タンパク質二次構造予測、立体構造予測(ホモロジーモデリング、フォールド認識、スレッディング、3D-1D法)
ゲノム解析・ゲノム遺伝学	遺伝子発見(ORF(open reading frame)、スプライシング解析、プロモータ解析、偽遺伝子)、ゲノム特徴抽出(繰り返し配列発見、転写因子、SSR(simple sequence repeat)、GC含量、コドン使用頻度)、ゲノム比較(ゲノムアライメント、編集距離、系統プロファイル法、ロゼッタストーン法、遺伝子並び順の保存、遺伝子の水平伝搬、多型マーカー(SNP、マイクロサテライト、VNTR、RFLP、HLAタイプ))
トランスクリプトーム解析・プロテオーム解析	遺伝子発現クラスタリング、遺伝子ネットワーク推定(ブーリアンネットワーク、ベイジアンネットワーク)、タンパク質相互作用解析
パスウェイ解析・システム生物学	ネットワーク解析(スケールフリー、ハブ、ネットワークモチーフ)、動的シミュレーションシステム解析(微分方程式、ロバストネス、フィードバック、フィードフォワード、感度解析、安定性解析、代謝流束解析)

# 分子生物学データベースの データの対象範囲は様々

	原核生物	原生生物	植物	動物	ヒト
DNA	■		■	■	■
タンパク質	■		■	■	■
機能			■		■
立体構造	■				
発現			■	■	■
多型		■			■

特定の分野のデータベースや、広範囲をカバーする  
統合データベースがある

# 医学・生物学分野の 文献データベース

MEDLINE

PubMed

# MEDLINE®

(Medical Literature Analysis and Retrieval System Online)

- 医学および関連分野の論文を収集
- U.S. National Library of Medicine 's® (NLM)と協力機関が編集
- 論文のタイトル、著者、要旨、誌名、ページ番号などを格納
- 約1700万件(2008年まで)、約5,200の学術雑誌
- **Medical Subject Headings (MeSH)**をつけているのが特徴
  - MeSH is the National Library of Medicine's controlled vocabulary thesaurus.
  - 階層構造をもつキーワード群
- 論文そのものではない
  - 要旨は掲載する(2000年以降のみで約80%)
  - 本文、図表は掲載していない
  - 約90%は英語(2000年以降のみ)
- 要旨が載ってないケース
  - 論文そのものに、そもそも無い
  - 要旨が英語で書かれていない

# PubMed

- ・ PubMedは、米国国立医学図書館(NLM)が編纂する医学文献データベースであるMEDLINEに、MEDLINEの対象とならない生命科学分野の文献を追加した文献データベース
- ・ 医学・生物学関係の文献検索に非常に便利
- ・ 1900万件の文献情報(1948年以降)
- ・ 無料で利用できる
- ・ 全文へのリンクがある(雑誌社、PubMedCentral等)

http://www.ncbi.nlm.nih.gov/pubmed/

NCBI **PubMed** A service of the [U.S. National Library of Medicine](#) and the [National Institutes of Health](#) [My NCBI](#) [\[Sign In\]](#) [\[Register\]](#)

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for

[Advanced Search \(beta\)](#)

Text Version

**To get started with PubMed, enter one or more search terms.**  
Search terms may be [topics](#), [authors](#) or [journals](#).

**The NIH Public Access Policy May Affect You**  
**Does NIH fund your work?**  
Then your manuscript must be made available in PubMed Central  
How?  
If you publish in one of [these journals](#), they will take care of the whole process.  
If you publish *anywhere else*, deposit the manuscript in PubMed Central via one of the options described at [publicaccess.nih.gov](#).

PubMed is a service of the [U.S. National Library of Medicine](#) that includes over 18 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s. PubMed includes links to full text articles and other related resources.

[Write to the Help Desk](#)  
[NCBI](#) | [NLM](#) | [NIH](#)  
[Department of Health & Human Services](#)  
[Privacy Statement](#) | [Freedom of Information Act](#) | [Disclaimer](#)

**Entrez PubMed**  
Overview  
Help | FAQ  
Tutorials  
New/Noteworthy [E-Utilities](#)

**PubMed Services**  
Journals Database  
MeSH Database  
Single Citation Matcher  
Batch Citation Matcher  
Clinical Queries  
Special Queries  
LinkOut  
My NCBI

**Related Resources**  
Order Documents  
NLM Mobile  
NLM Catalog  
NLM Gateway  
TOXNET  
Consumer Health  
Clinical Alerts  
ClinicalTrials.gov  
PubMed Central

# MeSH

- MeSH is NLM's controlled vocabulary used for indexing articles for MEDLINE/PubMed. MeSH terminology provides a consistent way to retrieve information that may use different terminology for the same concepts.
- 検索に用いるキーワードを選択する際に、同義語の問題があり、MeSHが役立つ
- MeSHは米国医学図書館が作成した専門用語集で、MEDLINEやPubMedの文献データの検索用インデックス(索引)を作るために用いられている
- 文献の内容を表す用語を10～15個加える
- 用語の関係として、厳密さによる階層構造をもつ

# MeSHと検索条件の翻訳(1)

NCBI PubMed A service of the U.S. National Library of Medicine and the National Institutes of Health  
www.pubmed.gov

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Jour

Search PubMed for tumor Go Clear [Advanced Search](#)  
[Save Search](#)

Limits Preview/Index History Clipboard **Details**

Display Summary Show 20 Sort By Send to

All: 2340726 Review: 256504

20 of 2340726 Page 1 of 117037 Next

Summary形式

- 1: [Sorafenib induced tumor lysis syndrome in an advanced hepatocellular carcinoma patient.](#)  
Huang WS, Yang CH.  
World J Gastroenterol. 2009 Sep 21;15(35):4464-6.  
PMID: 19764104 [PubMed - in process]  
[Related Articles](#) [Free article in PMC](#) | [at journal site](#)
- 2: [Stevens-Johnson syndrome complicating adalimumab therapy in Crohn's disease.](#)  
Salama M, Lawrance IC.  
World J Gastroenterol. 2009 Sep 21;15(35):4449-52.  
PMID: 19764100 [PubMed - in process]  
[Related Articles](#) [Free article in PMC](#) | [at journal site](#)
- 3: [Treatment of malignant gastric outlet obstruction with endoscopically placed self-expandable metal stents.](#)  
Gaidos JK, Draganov PV.  
World J Gastroenterol. 2009 Sep 21;15(35):4465-74.

検索条件の詳細をみる

# MeSHと検索条件の翻訳(2)

The screenshot displays the PubMed search interface. At the top, there are navigation tabs for 'All Databases', 'PubMed', 'Nucleotide', 'Protein', 'Genome', 'Structure', and 'OMIM'. The search bar contains the text 'PubMed' in a dropdown menu, followed by 'for tumor' in the main input field. To the right of the search bar are 'Go' and 'Clear' buttons. Below the search bar, there are buttons for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. A central message box reads: **New Try the new Advanced Search** Incorporating Limits, History, Preview/Index, Details, and Citation Search. On the left side, there is a blue sidebar with links: 'About Entrez', 'Text Version', 'Entrez PubMed', 'Overview', 'Help | FAQ', 'Tutorials', 'New/Noteworthy', and 'E-Utilities'. The 'Query Translation:' section is highlighted in blue and contains the following text: `"tumour"[All Fields] OR "neoplasms"[MeSH Terms]`  
`OR "neoplasms"[All Fields] OR "tumor"[All Fields]`

# 検索結果: MEDLINE形式(1)

Limits Preview/Index History Clipboard Details

Display MEDLINE Show 20 Sort By Send to

All: 1 Review: 0

1: [Sauvé S et al.](#) NMR assignment of the N-termi...[PMID: 19636918]

PMID- 19636918

OWN - NLM

STAT- MEDLINE

DA - 20090728

DCOM- 20090826

IS - 1874-270X (Electronic)

VI - 2

IP - 1

DP - 2008 Jun

TI - NMR assignment of the N-terminal TRAF-like RING zinc finger domain of human FLN29.

PG - 33-6

AB - Resistance of cancer cells to oncotherapeutics designed to trigger programmed cell death (a.k.a. apoptosis) greatly limits clinical efficacy. The human FLN29 protein may play a role in this process via protein-protein interactions. Here we report the NMR spectral assignment of the N-terminal TRAF2/6-RING-zinc finger-like domain of this protein.

AD - Centre for Biologics Research, Biologics and Genetic Therapies Directorate, Health Canada, 251 Sir Frederick Banting Driveway, A/L 2201E, Tunney's Pasture, Ottawa, ON, Canada.

FAU - Sauve, Simon

AU - Sauve S

FAU - Gingras, Genevieve

AU - Gingras G

FAU - Aubin, Remy A

AU - Aubin RA

FAU - Hodgson, Derek J

AU - Hodgson DJ

FAU - Aubin, Yves

AU - Aubin Y

選択する

タイトル

概要

著者

# 檢索結果：MEDLINE形式(2)

```
FAU - Aubin, Yves
AU - Aubin Y
LA - eng
PT - Journal Article
DEP - 20071228
PL - Netherlands
TA - Biomol NMR Assign
JT - Biomolecular NMR assignments
JID - 101472371
RN - 0 (Carbon Isotopes)
RN - 0 (Intracellular Signaling Peptides and Proteins)
RN - 0 (Nitrogen Isotopes)
RN - 0 (Protons)
RN - 0 (TRAFD1 protein, human)
RN - 0 (Tumor Necrosis Factor Receptor-Associated Peptides and Proteins)
SB - IM
MH - Amino Acid Sequence
MH - Carbon Isotopes/chemistry
MH - Humans
MH - Intracellular Signaling Peptides and Proteins/*chemistry
MH - Magnetic Resonance Spectroscopy/*methods
MH - Molecular Sequence Data
MH - Molecular Weight
MH - Nitrogen Isotopes/chemistry
MH - Protein Structure, Tertiary
MH - Protons
MH - *RING Finger Domains
MH - Tumor Necrosis Factor Receptor-Associated Peptides and Proteins/*chemistry
EDAT- 2009/07/29 09:00
MHDA- 2009/08/27 09:00
CRDT- 2009/07/29 09:00
PHST- 2007/11/15 [received]
PHST- 2007/12/11 [accepted]
PHST- 2007/12/28 [aheadofprint]
AID - 10.1007/s12104-007-9077-1 [doi]
PST - ppublish
SO - Biomol NMR Assign. 2008 Jun;2(1):33-6. Epub 2007 Dec 28.
```

# ゲノムDB

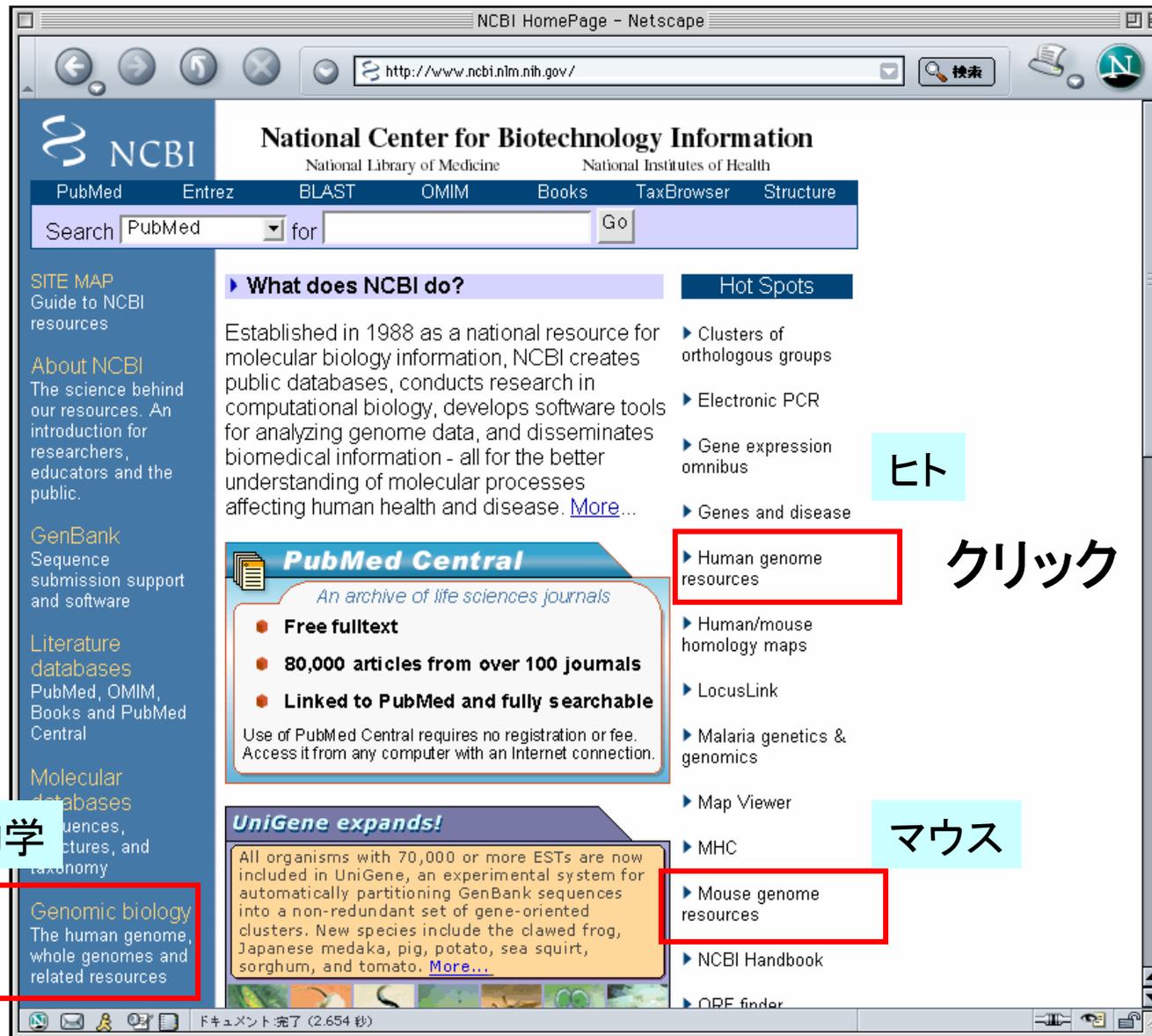
- NCBI Map Viewer (NCBI)
- Ensembl (EMBL-EBI, Sanger Institute)
- UCSC Genome Browser (UCSC)

ゲノムDB 1

NCBI Map Viewer (NCBI)

# NCBI トップページ

http://ncbi.nlm.nih.gov/



ゲノム生物学

ヒト

クリック

マウス

# NCBIのヒトゲノム情報資源

<http://www.ncbi.nlm.nih.gov/Genomes/>

## ▶ Genomic Biology

NCBI provides several genomic biology tools and resources, including organism-specific pages that include links to many web sites and databases relevant to that species. We invite you to explore the links provided on this page.

## ▶ Assembly and Annotation Information

- [The Genome Reference Consortium \(GRC\)](#) **NEW**
- [AGP Resources](#)
- [Annotation Information](#)
- [Assembly Information](#)
- [Genome Glossary](#)
- [NCBI Handbook, Chapter 14: Genome Assembly and Annotation Process](#)

## ▶ Announcements

### Map Viewer - genome annotation updates:

Species	Build	Map Viewer Release
<a href="#">Homo sapiens (human)</a>	37.1	August 4, 2009
<a href="#">Vitis vinifera (wine grape)</a>	IGGP 1	April 7, 2009
<a href="#">Taeniopygia guttata (zebra finch)</a>	1.1	March 5, 2009
<a href="#">Hydra magnipapillata</a>	1.1	January 28, 2009
<a href="#">Physcomitrella patens (moss)</a>	1.1	January 8, 2009
<a href="#">Caenorhabditis elegans (nematode)</a>	WS190	October 10, 2008
<a href="#">Anopheles gambiae (mosquito)</a>	AgamP3.3	October 10, 2008
<a href="#">Arabidopsis thaliana (mouse-ear cress)</a>	8.1	September 29, 2008
<a href="#">Ciona intestinalis</a>	1.1	September 16, 2008
<a href="#">Acyrtosiphon pisum (pea aphid)</a>	1.1	August 6, 2008
<a href="#">Bos taurus (cow)</a>	Btau_4.0	August 5, 2008
<a href="#">Equus caballus (horse)</a>	EquCab2.0	July 21, 2008
<a href="#">Sus scrofa (pig)</a>	Sscrofa5	July 14, 2008
<a href="#">Danio rerio (zebrafish)</a>	Zv7	July 12, 2008

## Genome Resources

- ▶ [Entrez Genome](#)
- ▶ [Fungal Genomes Central](#)
- ▶ [Genome Projects Database](#)
  - [Eukaryotic](#)
  - [Fungi](#)
  - [Insects](#)
  - [Mammals](#)
  - [Microbial](#)
  - [Plants](#)
- ▶ [Map Viewer](#)
- ▶ [Organelles](#)
- ▶ [Plant Genomes Central](#)
- ▶ [Viral Resources](#)
  - [Influenza Virus Resource](#)
  - [Retroviruses](#)
  - [Viral Genomes](#)

## Organism-Specific

-  Genome Resources
-  BLAST
-  Map Viewer
-  Genome Project DB

- ▶ [Aphid](#) **NEW**
- ▶ [Arabidopsis](#)
- ▶ [Aspergillus](#)
- ▶ [Bee](#)
- ▶ [Beetle](#)
- ▶ [Buffalo](#)
- ▶ [Cat](#)



# NCBIのゲノムマップ (ヒト21番染色体)

Map Viewer - Windows Internet Explorer  
 http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?taxid=9606&chr=21

NCBI NCBI Map Viewer

Search [ ] Find Find in This View Advanced Search

BLAST The Human Genome

Homo sapiens (human) Build 37.1 (Current)  
 Chromosome: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 [21] 22 X Y MT

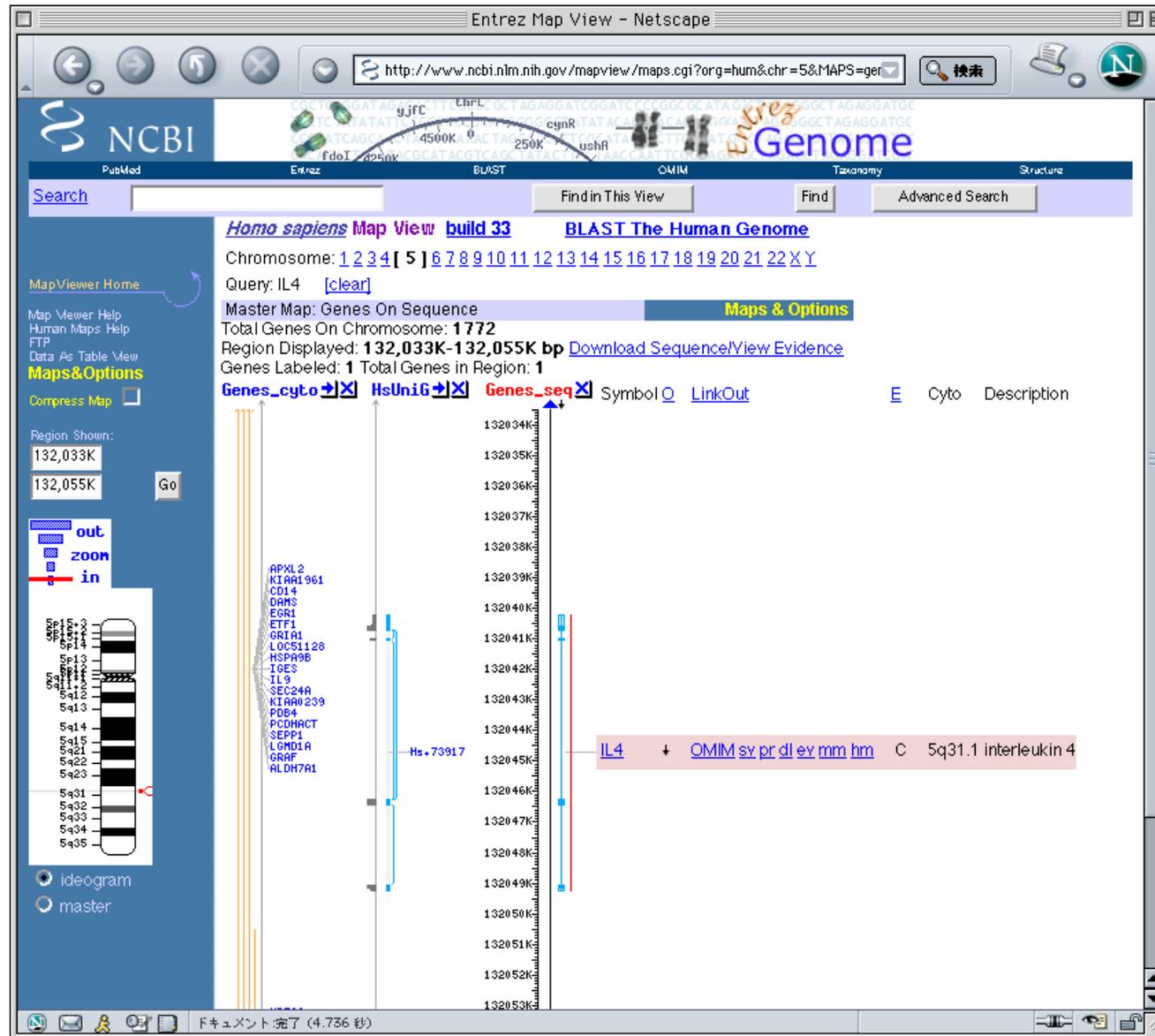
Master Map: Genes On Sequence Summary of Maps Maps & Options

Region Displayed: 0-48M bp Download/View Sequence Evidence

Contig	Gene	Symbol	Links	E	Cyto	Description
		LOC100129030	sv pr dj ev mm	SNP	protein	21 hypothetical LOC100129030
		SAMSN1	OMIM HGNC sv pr dj ev mm hm sts	SNP	best RefSeq	21q11 SAM domain, SH3 domain and nuclear localization signals 1
		RPS26P5	HGNC sv dj ev mm		best RefSeq	21q21.1 ribosomal protein S26 pseudogene 5
		RPL37P3	HGNC sv dj ev mm		best RefSeq	21 ribosomal protein L37 pseudogene 3
		PP1A3	HGNC sv dj ev mm sts		best RefSeq	21 peptidylprolyl isomerase A (cyclophilin A)-like 3
		KRTAP13-2	HGNC sv pr dj ev mm hm sts	SNP	best RefSeq	21q22.1 keratin associated protein 13-2
		KRTAP6-1	HGNC sv pr dj ev mm sts	SNP	best RefSeq	21q22.1 keratin associated protein 6-1
		EXOSC3P1	HGNC sv dj ev mm		best RefSeq	21q22.11 exosome component 3 pseudogene 1
		SYNJ1	OMIM HGNC sv pr dj ev mm hm sts	SNP	best RefSeq	21q22.2 synaptojanin 1
		RIMKB2	HGNC sv dj ev mm		best RefSeq	21q22.12 ribosomal modification protein rimK-like family member B2
		DYRK1A	OMIM HGNC sv pr dj ev mm hm sts	SNP	best RefSeq	21q22.13 dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A
		PCP4	OMIM HGNC sv pr dj ev mm hm sts	SNP	best RefSeq	21q22.2 Purkinje cell protein 4
		FAM3B	OMIM HGNC sv pr dj ev mm hm sts	SNP	best RefSeq	21q22.3 family with sequence similarity 3, member B
		C21orf128	HGNC sv pr dj ev mm sts		best RefSeq	21 chromosome 21 open reading frame 128
		UBASH3A	OMIM HGNC sv pr dj ev mm hm sts	SNP	best RefSeq	21q22.3 ubiquitin associated and SH3 domain containing, A
		HSE1BP	OMIM HGNC sv pr dj ev mm hm sts	SNP	best RefSeq	21q22.3 heat shock transcription factor 2 binding protein
		RPL31P1	HGNC sv dj ev mm		best RefSeq	21q22.3 ribosomal protein L31 pseudogene 1
		TRAPPC10	OMIM HGNC sv pr dj ev mm hm sts	SNP	best RefSeq	21q22.3 trafficking protein particle complex 10
		H2AFZP	HGNC sv dj ev mm		best RefSeq	21q22.3 H2A histone family, member Z, pseudogene
		C21orf90	HGNC sv pr dj ev mm sts		best RefSeq	21 chromosome 21 open reading frame 90

Summary of Maps:  
 Map 1: Ideogram  
 Region Displayed: 21pter-21qter  
 Map 2: Contig  
 Region Displayed: 0-48M bp Download/View Sequence Evidence  
 Total Contigs On Chromosome: 8 [1 not localized]  
 Contigs Labeled: 8 Total Contigs in Region: 8  
 Map 3: Homo sapiens UniGene Clusters  
 Region Displayed: 0-48M bp Download/View Sequence Evidence  
 Total Transcript alignments On Chromosome: 1802  
 UniGene Clusters Labeled: 34 Total Transcript alignments in Region: 1802  
 Histogram Data: Tick Width=97.036bp/pixel, Max Height=3814 transcripts (logarithmic scale)  
 Map 4: Genes On Sequence  
 Region Displayed: 0-48M bp Download/View Sequence Evidence  
 Total Genes On Chromosome: 425  
 Genes Labeled: 20 Total Genes in Region: 425

# NCBIのゲノムマップ (IL4遺伝子)



# ゲノムDB 2

## Ensembl

# Ensembl (<http://www.ensembl.org/>)

Search for:

e.g. gene BRCA2 or AL032821.2.1.143563 or muscular dystrophy

## 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](#))



## All genomes

-- Select a species --

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

Other pre-build species are available in [Ensembl Pre!](#) →



Ensembl is a joint project between [EMBL - EBI](#) and the [EMBL-EBI](#) 

[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 your own 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 Ensembl 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](#)

## What's New in Release 55 (14 July 2009)

- [Human GRCh37](#) (Human)
- [Assembly converter](#) (Human, Mouse)
- [Wallaby 2x genome](#) (Wallaby)
- [Mosquitoes moving to Ensembl Metazoa](#) (Mosquito, *A. aegypti*)
- [User data displays on chromosomes](#) (all species)

## More news...

- 2009-09-18: [Ensembl events in October 2009](#)
- 2009-09-15: [NCBI36 Ensembl Site](#)
- 2009-08-21: [Ensembl events in September 2009](#)

[Go to Ensembl blog](#) →

# Ensembl

- ヨーロッパ分子生物学研究所 (EMBL) のヨーロッパ生命情報学研究所 (EBI) と、ウェルカム財団サンガー研究所による共同プロジェクト
- ゲノム配列のアノテーションと、その解析ツールを開発・提供
- ヒト、チンパンジー、マウス、ゼブラフィッシュなどの脊椎動物、ショウジョウバエ、線虫、酵母、バクテリアまで含む多くの生物のゲノム解析データを公開  
(当初はいくつかの真核生物のみ)
- データベースおよびソフトウェアを無償で提供

# Ensemblの主な機能

- ゲノムブラウザによるデータの閲覧

既知遺伝子あるいは予測された遺伝子の存在、リピート配列、近縁なモデル生物との塩基配列の比較、塩基配列のGC含量など

- 自分のデータを合わせて表示

- ゲノムデータベースに対する情報の検索

キーワード検索、BLASTやSSAHAを用いた相同性検索

- 内部データをまとめて取得

ftpサイトからのデータのダウンロード、BioMartによるデータ絞込み

# Ensembl species

系統樹から  
探す

「生物種リスト」  
(この画面がでる)

原生生物、  
細菌、古細菌

菌類

この画面以外の  
後生動物

**Find a Species**  
[Species tree \(Requires Java\)](#)  
**Ensembl Species**

<b>Aedes</b> <i>Aedes aegypti</i>	<b>Gorilla</b> <i>Gorilla gorilla</i>	<b>Pig</b> (preview - assembly only) <i>Sus scrofa</i>
<b>Alpaca</b> <i>Vicugna pacos</i>	<b>Guinea Pig</b> <i>Cavia porcellus</i>	<b>Pika</b> <i>Ochotona princeps</i>
<b>Anole Lizard</b> <i>Anolis carolinensis</i>	<b>Hedgehog</b> <i>Erinaceus europaeus</i>	<b>Platypus</b> <i>Ornithorhynchus anatinus</i>
<b>Anopheles</b> <i>Anopheles gambiae</i>	<b>Horse</b> <i>Equus caballus</i>	<b>Rabbit</b> <i>Oryctolagus cuniculus</i>
<b>Armadillo</b> <i>Dasypus novemcinctus</i>	<b>Human</b> <i>Homo sapiens</i>	<b>Rat</b> <i>Rattus norvegicus</i>
<b>Bushbaby</b> <i>Otolemur gamettii</i>	<b>Hyrax</b> <i>Procavia capensis</i>	<b>Saccharomyces cerevisiae</b>
<b>Caenorhabditis elegans</b>	<b>Kangaroo rat</b> <i>Dipodomys ordii</i>	<b>Shrew</b> <i>Sorex araneus</i>
<b>Ciona intestinalis</b>	<b>Lamprey</b> (preview - assembly only) <i>Petromyzon marinus</i>	<b>Sloth</b> <i>Choloepus hoffmanni</i>
<b>Ciona savignyi</b>	<b>Lesser hedgehog tenrec</b> <i>Echinops telfairi</i>	<b>Squirrel</b> <i>Sciurus harrisi</i>
<b>Cat</b> <i>Felis catus</i>	<b>Macaque</b> <i>Macaca mulatta</i>	<b>Stickleback</b> <i>Gasterosteus aculeatus</i>
<b>Chicken</b> <i>Gallus gallus</i>	<b>Marmoset</b> (preview - assembly only) <i>Callithrix jacchus</i>	<b>Tarsier</b> <i>Tarsius syrichta</i>
<b>Chimpanzee</b> <i>Pan troglodytes</i>	<b>Medaka</b> <i>Oryzias latipes</i>	<b>Tetraodon</b> <i>Tetraodon nigroviridis</i>
<b>Cow</b> <i>Bos taurus</i>	<b>Megabat</b> <i>Pteropus vampyrus</i>	<b>Tree Shrew</b> <i>Tupaia belangeri</i>
<b>Dog</b> <i>Canis familiaris</i>	<b>Microbat</b> <i>Myotis lucifugus</i>	<b>Wallaby</b> <i>Macropus eugenii</i>
<b>Dolphin</b> <i>Tursiops truncatus</i>	<b>Mouse</b> <i>Mus musculus</i>	<b>Xenopus tropicalis</b>
<b>Elephant</b> <i>Loxodonta africana</i>	<b>Mouse Lemur</b> <i>Microcebus murinus</i>	<b>Zebra Finch</b> <i>Taeniopygia guttata</i>
<b>Fruitfly</b> <i>Drosophila melanogaster</i>	<b>Opossum</b> <i>Marsupialis domestica</i>	<b>Zebrafish</b> <i>Danio rerio</i>
<b>Fugu</b> <i>Takifugu rubripes</i>	<b>Orangutan</b> <i>Pongo pygmaeus</i>	

**Other Metazoa**  
Additional metazoan genomes (initially insect vectors and nematodes) are available from [EnsemblMetazoa](#)

**Plants and Fungi**  
Ensembl currently includes Brewer's Yeast, *Saccharomyces cerevisiae*, for comparative purposes.

**Protists, Bacteria and Archaea**  
Unicellular and non-eukaryotic genomes can be found at [EnsemblProtists](#) and [EnsemblBacteria](#)



# Bacteria Ensembl

 <p><b>Bacillus</b></p> <ul style="list-style-type: none"><li>• Bacillus cereus AH187</li><li>• Bacillus cereus Q1</li><li>• Bacillus anthracis Sterne</li><li>• Bacillus cereus ZK</li><li>• Bacillus cereus NVH 391-98</li><li>• Bacillus clausii</li><li>• Bacillus anthracis Ames ancestor</li><li>• Bacillus anthracis Ames</li><li>• Bacillus licheniformis Novozymes</li><li>• Bacillus cereus ATCC 10987</li><li>• Bacillus cereus AH820</li><li>• Bacillus licheniformis Goettingen</li><li>• Bacillus amyloliquefaciens</li><li>• Bacillus thuringiensis konkukian 97-27</li><li>• Bacillus cereus B4264</li><li>• Bacillus thuringiensis Al Hakam</li><li>• Bacillus subtilis</li><li>• Bacillus cereus ATCC 14579</li><li>• Bacillus weihenstephanensis</li><li>• Bacillus pumilus SAFR-032</li><li>• Bacillus cereus G9842</li><li>• Bacillus halodurans</li></ul>	<ul style="list-style-type: none"><li>• Escherichia coli O157:H7</li><li>• Escherichia coli O7:K1 IAI39</li><li>• Escherichia coli O127:H6 E2348/69</li><li>• Escherichia coli O6 ATCC 700928</li></ul>  <p><b>Mycobacterium</b></p> <ul style="list-style-type: none"><li>• Mycobacterium abscessus</li><li>• Mycobacterium bovis AF2122/97</li><li>• Mycobacterium leprae TN</li><li>• Mycobacterium marinum</li><li>• Mycobacterium paratuberculosis</li><li>• Mycobacterium sp. KMS</li><li>• Mycobacterium vanbaalenii</li><li>• Mycobacterium sp. JLS</li><li>• Mycobacterium ulcerans</li><li>• Mycobacterium bovis BCG</li><li>• Mycobacterium tuberculosis ATCC 25177</li><li>• Mycobacterium gilvum</li><li>• Mycobacterium tuberculosis CDC1551</li><li>• Mycobacterium avium</li><li>• Mycobacterium smegmatis</li><li>• Mycobacterium sp. MCS</li><li>• Mycobacterium tuberculosis H37Rv</li><li>• Mycobacterium leprae Br4923</li></ul>	<ul style="list-style-type: none"><li>• Staphylococcus aureus ET3-1</li><li>• Staphylococcus carnosus</li><li>• Staphylococcus aureus ATCC 700698</li></ul>  <p><b>Streptococcus</b></p> <ul style="list-style-type: none"><li>• Streptococcus pyogenes MGA59429</li><li>• Streptococcus pyogenes M1 SF370</li><li>• Streptococcus suis 98HAH33</li><li>• Streptococcus pyogenes MGA510270</li><li>• Streptococcus sanguinis</li><li>• Streptococcus pneumoniae CGSP14</li><li>• Streptococcus equi zoepidemicus</li><li>• Streptococcus thermophilus ATCC BAA-250</li><li>• Streptococcus suis 05ZYH33</li><li>• Streptococcus uberis</li><li>• Streptococcus agalactiae III</li><li>• Streptococcus pneumoniae D39</li><li>• Streptococcus pneumoniae TIGR4</li><li>• Streptococcus pneumoniae G54</li><li>• Streptococcus pyogenes MGA52096</li><li>• Streptococcus gordonii</li><li>• Streptococcus pyogenes MGA510750</li><li>• Streptococcus pyogenes M6 ATCC BAA-346</li><li>• Streptococcus pyogenes M3 SSI-1</li><li>• Streptococcus pyogenes M1 MGA55005</li><li>• Streptococcus pyogenes Manfredo</li><li>• Streptococcus agalactiae Ia</li><li>• Streptococcus agalactiae V</li><li>• Streptococcus thermophilus CNRZ 1066</li><li>• Streptococcus pyogenes M18 MGA58232</li><li>• Streptococcus equi MGC510565</li><li>• Streptococcus pneumoniae ATCC 700669</li><li>• Streptococcus pyogenes NZ131</li><li>• Streptococcus pneumoniae ATCC BAA-255</li><li>• Streptococcus pyogenes M3 ATCC BAA-595</li><li>• Streptococcus pneumoniae Hungary19A-6</li><li>• Streptococcus mutans</li><li>• Streptococcus thermophilus ATCC BAA-491</li><li>• Streptococcus equi 4047</li><li>• Streptococcus pyogenes M28 MGA56180</li></ul>
 <p><b>Escherichia / Shigella</b></p> <ul style="list-style-type: none"><li>• Escherichia coli DSM 5911</li><li>• Escherichia coli UT89</li><li>• Escherichia coli O1:K1 / APEC</li><li>• Shigella flexneri 2457T</li><li>• Shigella boydii 4</li><li>• Escherichia coli O81 ED1a</li><li>• Shigella sonnei</li><li>• Escherichia coli 55989</li><li>• Shigella boydii 18</li><li>• Shigella dysenteriae</li><li>• Escherichia coli Crooks</li><li>• Escherichia coli O9:H4 HS</li><li>• Escherichia fergusonii</li><li>• Escherichia coli RIMD 0509952</li><li>• Shigella flexneri 301</li><li>• Escherichia coli SMS-3-5</li><li>• Shigella flexneri 8401</li><li>• Escherichia coli O17:K52:H18 UMN026</li><li>• Escherichia coli O157:H7 EC4115</li><li>• Escherichia coli DH10B</li><li>• Escherichia coli K12</li><li>• Escherichia coli O45:K1 S88</li><li>• Escherichia coli SE11</li><li>• Escherichia coli O8 IAI1</li><li>• Escherichia coli O139:H28 E24377A</li><li>• Escherichia coli O6:K15:H31 536</li></ul>	 <p><b>Neisseria</b></p> <ul style="list-style-type: none"><li>• Neisseria meningitidis C</li><li>• Neisseria meningitidis B</li><li>• Neisseria gonorrhoeae NCCP11945</li><li>• Neisseria gonorrhoeae ATCC 700825</li><li>• Neisseria meningitidis 2a</li><li>• Neisseria meningitidis A 4A</li></ul>  <p><b>Pyrococcus</b></p> <ul style="list-style-type: none"><li>• Pyrococcus furiosus</li><li>• Pyrococcus abyssi Orsay</li><li>• Pyrococcus horikoshii</li><li>• Pyrococcus kodakaraensis</li></ul>	 <p><b>Staphylococcus</b></p> <ul style="list-style-type: none"><li>• Staphylococcus aureus NCTC 8325</li><li>• Staphylococcus saprophyticus</li><li>• Staphylococcus aureus JH9</li><li>• Staphylococcus aureus N315</li><li>• Staphylococcus aureus TCH1516</li><li>• Staphylococcus aureus USA300</li><li>• Staphylococcus aureus MW2</li><li>• Staphylococcus epidermidis ATCC 12228</li><li>• Staphylococcus aureus JH1</li><li>• Staphylococcus haemolyticus</li><li>• Staphylococcus aureus Newman</li><li>• Staphylococcus epidermidis ATCC 35984</li><li>• Staphylococcus aureus MSSA476</li><li>• Staphylococcus aureus ATCC 700699</li><li>• Staphylococcus aureus MRSA252</li><li>• Staphylococcus aureus COL</li></ul>

# Ensembl (ヒト) の統計

- 年4回更新

## Summary

Assembly:	GRCh37, Feb 2009
Database version:	55.37
Base Pairs:	3,272,480,989
Golden Path Length:	3,101,804,741
Genebuild by:	Ensembl
Genebuild method:	Full genebuild
Genebuild started:	Mar 2009
Genebuild released:	May 2009
Genebuild last updated/patched:	May 2009

## Gene counts

Known protein-coding genes:	22,258
Novel protein-coding genes:	1,021
Pseudogenes:	8,225
RNA genes:	6,411
Immunoglobulin/T-cell receptor gene segments:	361
Gene exons:	404,376
Gene transcripts:	101,641

## Other

Genscan gene predictions:	68,186
SNPs:	15,094,251

# Ensemblのヒト遺伝子情報 (Gene Summary上部)

e!Ensembl

[Home > Human](#)
[Login / Register](#) | 
 [BLAST/BLAT](#) | 
 [BioMart](#) | 
 [Docs & FAQs](#)

Location: 7:114,055,052-114,331,092 **Gene: FOXP2** Transcript: FOXP2-202

**Gene-based displays**

- Gene summary
- Splice variants (19)
- Supporting evidence
- Sequence
- External references (2)
- Regulation
- Comparative Genomics
  - Genomic alignments (38)
  - Gene Tree (image)
    - Gene Tree (text)
    - Gene Tree (alignment)
  - Orthologues (48)
  - Paralogues (8)
  - Protein families (7)
- Genetic Variation
  - Variation Table
  - Variation Image
- External Data
  - Personal annotation
- ID History
  - Gene history

- [Configure this page](#)
- [Manage your data](#)
- [Export data](#)
- [Bookmark this page](#)

**Gene: FOXP2 (ENSG00000128573)**

Forkhead box protein P2 (CAG repeat protein 44)(Trinucleotide repeat-containing gene 10 protein) [Source: UniProtKB/Swiss-Prot O15409](#)

**Location** [Chromosome 7: 113,726,382-114,331,092](#) forward strand.

**Transcripts** There are 19 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
FOXP2-001	<a href="#">ENST00000350908</a>	<a href="#">ENSP00000265436</a>	protein_coding
FOXP2-002	<a href="#">ENST00000440349</a>	<a href="#">ENSP00000395552</a>	nonsense_mediated_decay
FOXP2-003	<a href="#">ENST00000324462</a>	<a href="#">ENSP00000319424</a>	protein_coding
FOXP2-004	<a href="#">ENST00000408937</a>	<a href="#">ENSP00000386200</a>	protein_coding
FOXP2-005	<a href="#">ENST00000393498</a>	<a href="#">ENSP00000377135</a>	protein_coding
FOXP2-006	<a href="#">ENST00000412402</a>	<a href="#">ENSP00000405470</a>	nonsense_mediated_decay
FOXP2-008	<a href="#">ENST00000441290</a>	<a href="#">ENSP00000416825</a>	nonsense_mediated_decay
FOXP2-010	<a href="#">ENST00000378237</a>	<a href="#">ENSP00000367492</a>	protein_coding
FOXP2-014	<a href="#">ENST00000393494</a>	<a href="#">ENSP00000377132</a>	protein_coding
FOXP2-015	<a href="#">ENST00000452963</a>	<a href="#">ENSP00000409826</a>	protein_coding
FOXP2-016	<a href="#">ENST00000390668</a>	<a href="#">ENSP00000375084</a>	protein_coding
FOXP2-017	<a href="#">ENST00000360232</a>	<a href="#">ENSP00000353367</a>	protein_coding
FOXP2-018	<a href="#">ENST00000393495</a>	<a href="#">ENSP00000377133</a>	protein_coding
FOXP2-201	<a href="#">ENST00000393489</a>	<a href="#">ENSP00000377129</a>	protein_coding
FOXP2-202	<a href="#">ENST00000393491</a>	<a href="#">ENSP00000377130</a>	protein_coding
FOXP2-203	<a href="#">ENST00000393500</a>	<a href="#">ENSP00000377137</a>	protein_coding
FOXP2-204	<a href="#">ENST00000403100</a>	<a href="#">ENSP00000384090</a>	protein_coding
FOXP2-205	<a href="#">ENST00000403559</a>	<a href="#">ENSP00000385069</a>	protein_coding
FOXP2-206	<a href="#">ENST00000403611</a>	<a href="#">ENSP00000385497</a>	protein_coding

**Transcript and Gene level displays**

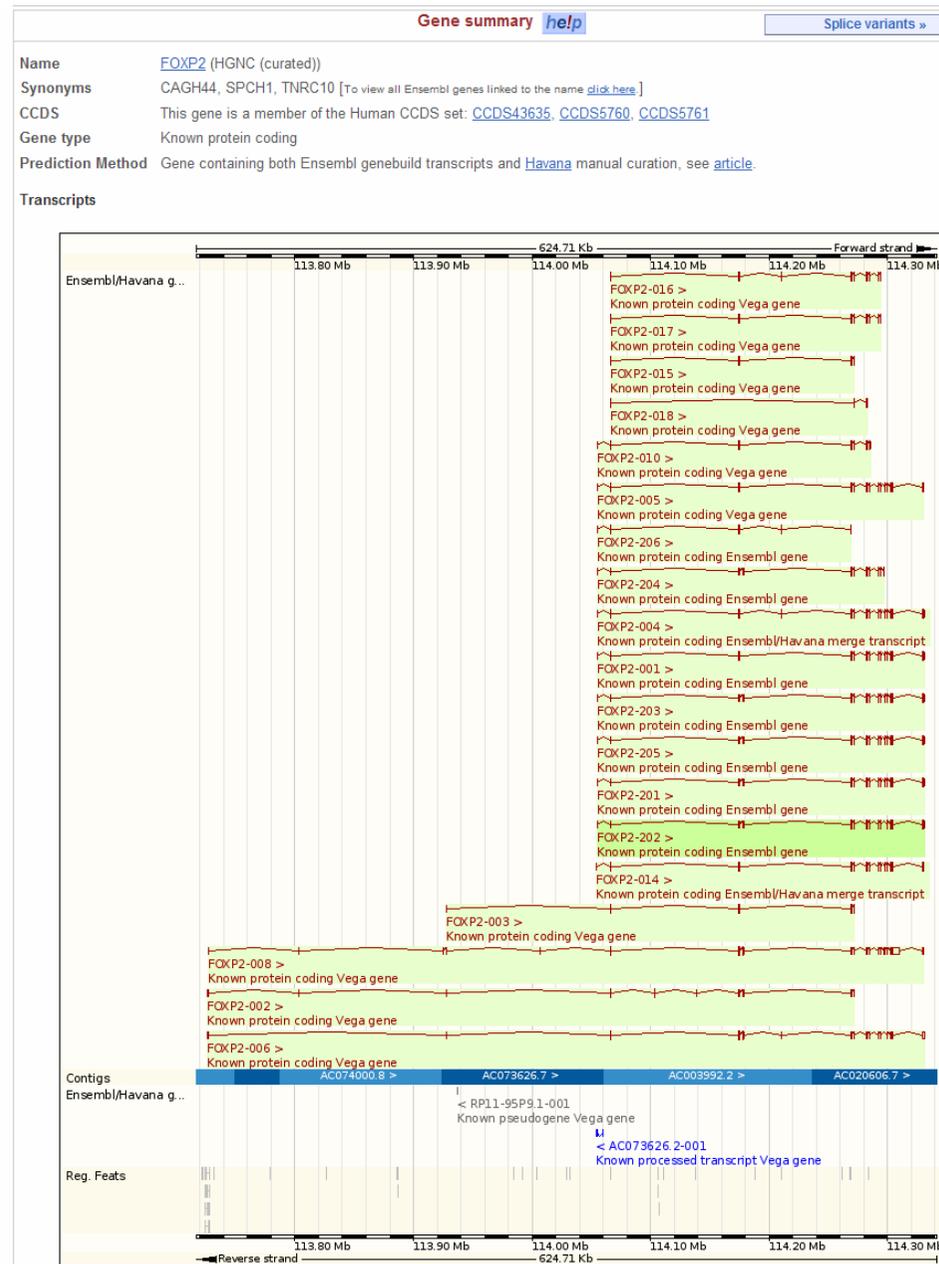
In Ensembl a gene is made up of one or more transcripts. We provide displays at two levels:

- Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation.
- Gene views which provide displays for data associated at the gene level such as orthologues and paralogues, regulatory regions and splice variants.

This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want using the menu at the left hand side of the page. To return to viewing gene level information click on the Gene tab in the menu bar at the top of the page.

[Gene summary](#) [help](#)
[Splice variants »](#)

# Ensemblのヒト遺伝子情報 (Gene Summary下部)



# EnsemblのBioMartによるデータマイニング

MartView (<http://www.ensembl.org/Multi/martview>) というツールにより、大規模なゲノムデータから欲しい情報を抽出できる。以下の3つのステップで実行する

## 1. Start

データセットの選択、生物種の選択

## 2. Filter

データの絞りこみのためのフィルターを選択する

フィルター：染色体領域、Ensembl ID、InterPro ID、Gene Ontology、発現データなど

## 3. Output

出力するデータの選択 (Features, SNPs, Structures, Sequences) 、  
フォーマットの選択 (HTML、テキストあるいはMicrosoft Excel)

# Ensemblトップ画面 (http://www.ensembl.org)

**e!Ensembl**  
Home

Search for:   
Go

e.g. gene BRCA2 or AL032821.2.1.143563 or muscular dystrophy

### 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 pre-build species are available in [Ensembl Pre!](#)

Ensembl is a joint project between **Wellcome Trust Sanger Institute** and the **EMBL-EBI**

Ensembl is 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 your own 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 Ensembl 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, etc.

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

### What's New in Release 55 (14 July 2009)

- [Human GRCh37](#) (Human)
- [Assembly converter](#) (Human, Mouse)
- [Wallaby 2x genome](#) (Wallaby)
- [Mosquitoes moving to Ensembl Metazoa](#) (Mosquito, *A. aegypti*)
- [User data displays on chromosomes](#) (all species)

[More news...](#)

- 2009-09-18: [Ensembl events in October 2009](#)
- 2009-09-15: [NCBI36 Ensembl Site](#)
- 2009-08-21: [Ensembl events in September 2009](#)

[Go to Ensembl blog](#)

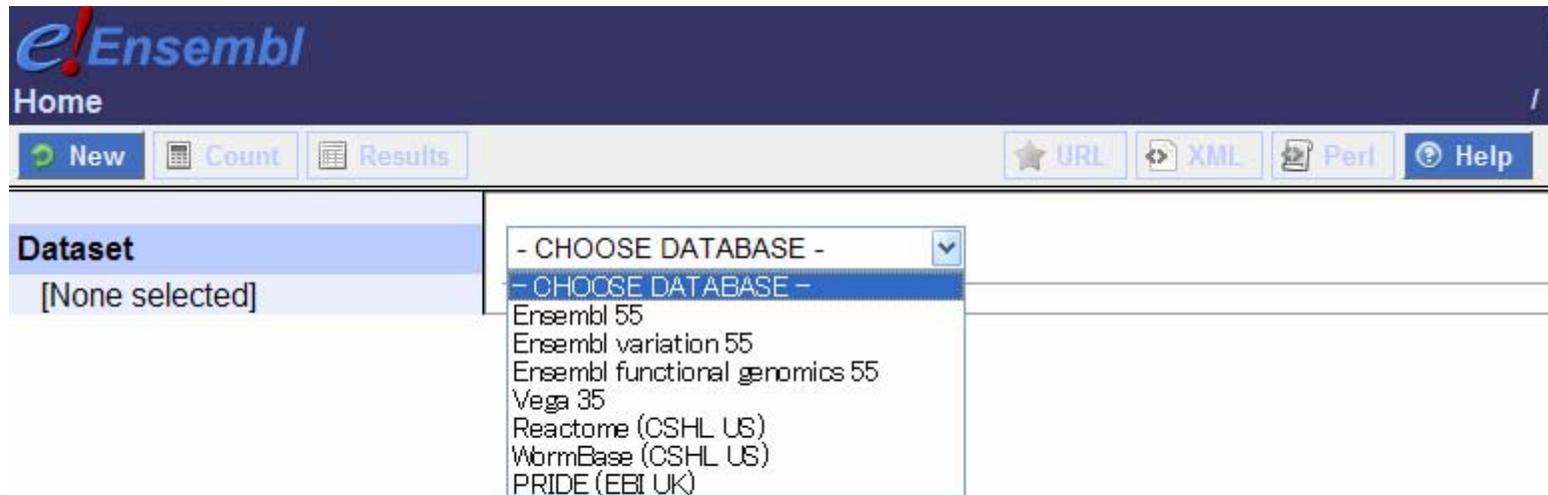
Ensembl release 55 - July 2009 © [WTSI](#) / [EBI](#)

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

# Ensembl BioMart

- 大量のテキストデータから欲しい部分だけ取得
  - ✓取得データを、後でエクセルやプログラミングでさらに解析するため

操作法：データセットと生物種を選択



# BioMart : フィルター (検索条件) の選択

The screenshot displays the Ensembl BioMart interface. The top navigation bar includes 'New', 'Count', and 'Results' buttons. The main content area is divided into several sections:

- Dataset:** A dropdown menu showing 'Ensembl 55' and 'Homo sapiens genes (GRCh37)'. A red box highlights the 'Filters' section, with a red arrow pointing to it from the text 'Filterでクリック'.
- Filters:** A section with '[None selected]' and a red box around the 'Filters' header.
- Attributes:** A section with 'Ensembl Gene ID' and 'Ensembl Transcript ID'.
- Dataset:** A section with '[None Selected]'.

Below the main interface, there are three smaller panels illustrating the process:

- The first panel shows the 'Count' button highlighted with a red box and a red arrow pointing to it from the text 'Countをクリック'.
- The second panel shows the 'Transmembrane domains : Only' filter selected, with a red arrow pointing to it from the text 'Transmembraneドメインを持つ'.
- The third panel shows the resulting dataset: 'Dataset 5977 / 44285 Genes' and 'Homo sapiens genes (GRCh37)'. A red circle highlights the dataset name, with a red arrow pointing to it from the text '検索の結果'.

# BioMart: 属性 (取得情報) の選択

**Attributes**をクリック

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

Features  Homologs  
 Structures  Sequences  
 Variations

取得するAttributes (属性) を選

反映される

**Results**をクリック  
検索結果を表示する

Dataset 5977 / 44285 Genes  
Homo sapiens genes (GRCh37)

Filters  
[None selected]  
Transmembrane domains : Only

Attributes  
Ensembl Gene ID  
Ensembl Transcript ID  
Description  
% GC content  
Transcript count

Dataset  
[None Selected]

GENE:  
Ensembl  
 Ensembl Gene ID  
 Ensembl Transcript ID  
 Associated Gene Name  
 Associated Transcript Name  
 Associated Gene DB  
 Associated Transcript DB  
 Description  
 Chromosome Name  
 Gene Start (bp)  
 Gene End (bp)  
 Strand  
 Band  
 Transcript Start (bp)  
 Transcript End (bp)  
 Transcript count  
 % GC content  
 Gene Biotype  
 Transcript Biotype  
 Source  
 Status (gene)  
 Status (transcript)

# BioMart : 出力データとフォーマットを選択

好きな形式で保存する

The screenshot shows the BioMart interface with a red box highlighting the export and view options. The interface includes a top navigation bar with 'New', 'Count', and 'Results' buttons. Below this, there are buttons for 'URL', 'XML', 'Perl', and 'Help'. The main content area is divided into several sections:

- Dataset:** 5977 / 44285 Genes, Homo sapiens genes (GRCh37)
- Filters:** [None selected], Transmembrane domains : Only
- Attributes:** Ensembl Gene ID, Ensembl Transcript ID, Description, % GC content, Transcript count
- Dataset:** [None Selected]

The export options are:

- Export all results to: File (dropdown), TSV (dropdown),  Unique results only
- Go (button)
- Email notification to: (text input)

The view options are:

- View: 10 (dropdown), rows as HTML (dropdown),  Unique results only

The main data table is as follows:

Ensembl Gene ID	Ensembl Transcript ID	Description	% GC content	Transcript count
<a href="#">ENSG00000177951</a>	<a href="#">ENST00000382762</a>	BET1-like protein (Vesicle transport protein GOS15)(Golgi SNARE with a size of 15 kDa)(GOS-15)(GS15) [Source:UniProtKB/Swiss-Prot;Acc:Q9NYM9]	48.36	4
<a href="#">ENSG00000177951</a>	<a href="#">ENST00000332865</a>	BET1-like protein (Vesicle transport protein GOS15)(Golgi SNARE with a size of 15 kDa)(GOS-15)(GS15) [Source:UniProtKB/Swiss-Prot;Acc:Q9NYM9]	48.36	4
<a href="#">ENSG00000212913</a>	<a href="#">ENST00000391600</a>	Putative uncharacterized protein ENSP00000375442 [Source:UniProtKB/TrEMBL;Acc:B7WP02]	54.56	1
<a href="#">ENSG00000215783</a>	<a href="#">ENST00000400891</a>	Putative uncharacterized protein ENSP00000383684 Fragment [Source:UniProtKB/TrEMBL;Acc:A8MWP5]	54.30	1
<a href="#">ENSG00000196143</a>	<a href="#">ENST00000359695</a>	Olfactory receptor 11H12 [Source:UniProtKB/Swiss-Prot;Acc:P62892]	43.43	1

ゲノムDB 3

UCSC Genome Browser  
(UCSC)

# UCSCゲノムブラウザ

- 標準ゲノム配列の座標の上に、各種のアノテーション情報を表示するウェブサイト
- 表示しているアノテーション情報は、known genes, predicted genes, ESTs, mRNAs, CpG islands, assembly gaps and coverage, chromosomal bands, mouse homologiesなど
- ヒトなどの脊索動物、ハエ、線虫、酵母まで(植物、細菌等無し)
- カリフォルニア大学サンタクルーズ校のゲノムバイオインフォマティクスグループが運営

# UCSCゲノムブラウザの主なサービス

- Genome Browser (ヒトやモデル生物)
- ENCODE
- Blat
- Table Browser
- Gene Sorter
- In Silico PCR
- Genome Graphs
- Galaxy
- VisiGene
- Proteome Browser

# UCSCゲノムブラウザ

## http://genome.ucsc.edu/

The image shows a screenshot of the UCSC Genome Bioinformatics website. At the top, the title "UCSC Genome Bioinformatics" is displayed in a yellow banner. Below this is a blue navigation bar with links: "Genomes", "Blat", "Tables", "Gene Sorter", "PCR", "VisiGene", "Proteome", "Session", "FAQ", and "Help". The "Genomes" link is circled in red. A red arrow points from the "Genomes" link to the "Human (*Homo sapiens*) Genome Browser Gateway" section below. This section contains a search form with the following fields: "clade" (dropdown menu with "Mammal" selected), "genome" (dropdown menu with "Human" selected), "assembly" (dropdown menu with "Mar. 2006" selected), "position or search term" (text input field containing "chrX:151,073,054-151,383,976"), and "image width" (text input field containing "800"). A "submit" button is circled in red.

UCSC Genome Bioinformatics

[Genomes](#) - [Blat](#) - [Tables](#) - [Gene Sorter](#) - [PCR](#) - [VisiGene](#) - [Proteome](#) - [Session](#) - [FAQ](#) - [Help](#)

[Genome Browser](#)

[ENCODE](#)

[Blat](#)

[Table Browser](#)

[Gene Sorter](#)

[In Silico PCR](#)

### About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides a portal to the ENCODE project.

We encourage you to explore these sequences with our tools. The [Genome Browser](#) zooms and scrolls over chromosomes, showing the work of annotators worldwide. The [Gene Sorter](#) shows expression, homology and other information on groups of genes that can be related in many ways. [Blat](#) quickly maps your sequence to the genome. The [Table Browser](#) provides convenient access to the underlying database. [VisiGene](#) lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. [Genome Graphs](#) allows you to upload and display genome-wide data sets.

### Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).  
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade	genome	assembly	position or search term	image width	
Mammal	Human	Mar. 2006	chrX:151,073,054-151,383,976	800	<input type="submit" value="submit"/>

# UCSC Genome Browser (ヒト)

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

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

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

position/search chrX:151,073,054-151,383,976 jump clear size 310,923 bp. configure

chrX (q28) 22.2 q21.1 22.3 q23 24 Xq25 Xq28

Scale 100 kb  
chrX: 151100000 151150000 151200000 151250000 151300000 151350000

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

GABRA3 RefSeq Genes CS266382.1

RefSeq Genes

Mammalian Gene Collection Full ORF mRNAs BC028629

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

Vertebrate Multiz Alignment & Conservation (44 Species)  
Placental Mammal Basewise Conservation by PhyloP

Mammal Cons 2 -0.3

Multiz Alignments of 44 Vertebrates

Rhesus Tarsier Mouse Dog Elephant Opossum Platypus Chicken Lizard X\_tropicalis Stickleback

Simple Nucleotide Polymorphisms (dbSNP build 138)

SNP Genotyping Arrays

Affy SNP 6.0 Affy SNP 6.0 SV Illumina 1M-Duo

RepeatMasker

move start Click on a feature for details. Click or drag in the base position track to zoom in. Click gray/blue move end  
< 2.0 > bars on left for track options and descriptions. < 2.0 >

default tracks hide all add custom tracks configure reverse refresh

collapse all Use drop-down controls below and press refresh to alter tracks displayed. expand all  
Tracks with lots of items will automatically be displayed in more compact modes.

## UCSCゲノムブラウザで見られる項目(Track)

- ・ マッピング(STS,FISH等)とシーケンシング
- ・ 表現型と疾患
- ・ 遺伝子と遺伝子予測
- ・ mRNAとEST
- ・ 発現
- ・ 発現制御
- ・ 比較ゲノム
- ・ ゲノム変異、リピート
- ・ パイロットENCODE(領域と遺伝子)
- ・ パイロットENCODE(転写産物)
- ・ パイロットENCODE(ChIP)
- ・ パイロットENCODE(クロマチン構造)
- ・ パイロットENCODE(比較ゲノムとゲノム変異)

# Table Browser

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

## Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data.

clade:  genome:  assembly:

group:  track:

table:

region:  genome  ENCODE  position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format:

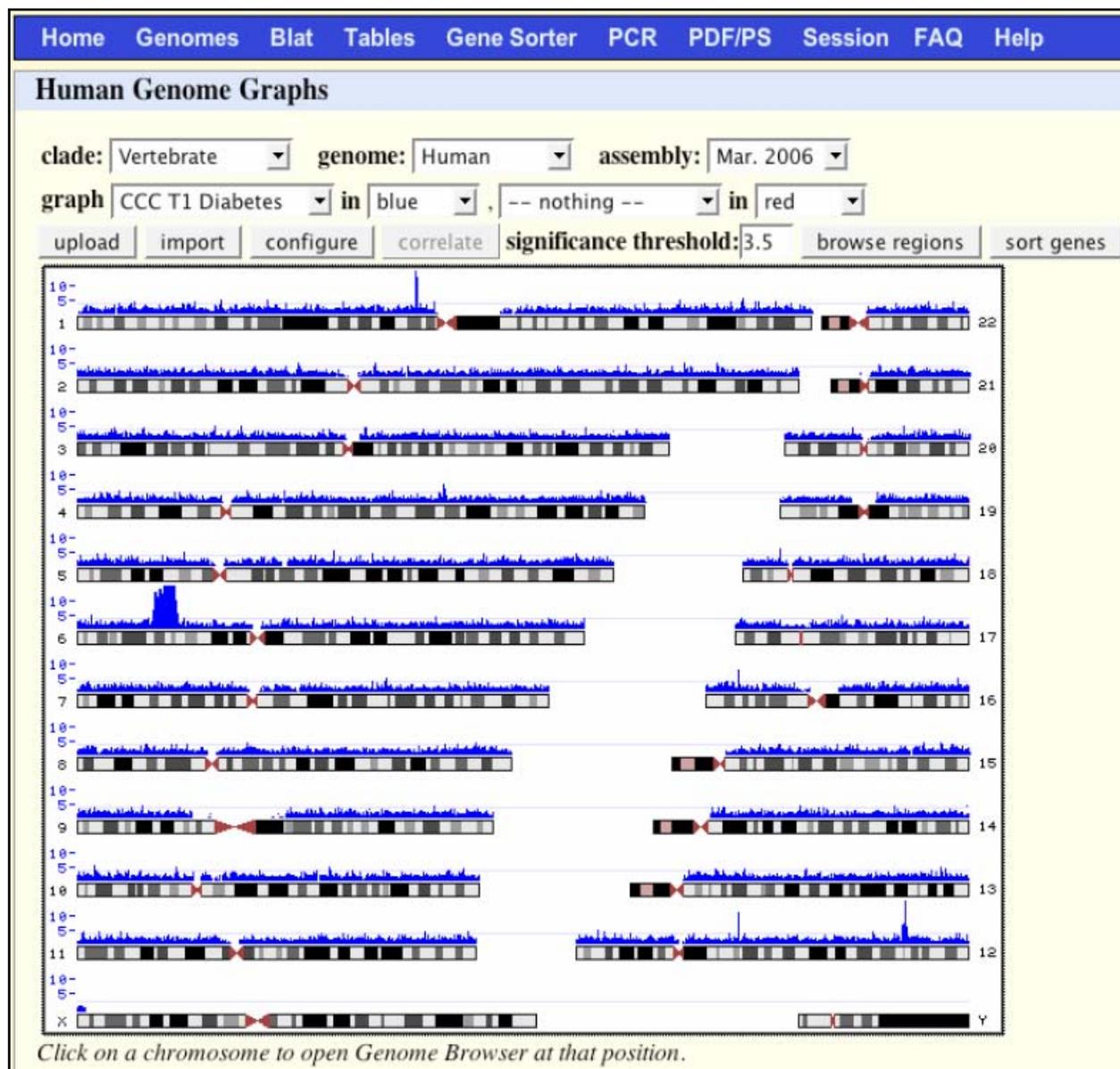
output file:

file type returned:  plain text  gzip compressed

To reset **all** user cart settings (including custom tracks), [click here](#).

BioMartのように  
データをフィルタリング + 取得できる

# Genome Graph: ゲノム座標上の数値データ表示 ゲノムワイド相関解析によるP値の表示



# Gene Sorter: 遺伝子発現データ 類似した発現パターンを持つ遺伝子の表示

UCSC Human Gene Sorter																				
genome <input type="text" value="Human"/> assembly <input type="text" value="Mar. 2006"/> search <input type="text" value="uc001mah.1"/> <input type="button" value="Go!"/> sort by <input type="text" value="Expression (GNF Atlas2)"/> <input type="button" value="configure"/> <input type="button" value="filter (now off)"/> display <input type="text" value="50"/> output <input type="text" value="sequence"/> <input type="text" value="text"/>																				
#	Name	VisiGene	fetal brain	whole brain	amygdala	thymus	bone marrow	PB-CD4+ Tcells	skin	adipocyte	pancreatic islets	heart	lung	kidney	liver	ovary	testis	BLASTP E-Value	Genome Position	Description
1	HBG1	151637																5e-82	chr11 5,226,870	A-gamma globin
2	HBM	n/a																1e-19	chr16 156,370	hemoglobin mu chain
3	SLC4A1	n/a																n/a	chr17 39,691,779	solute carrier family 4, anion exchanger, men
4	EPB42	185276																n/a	chr15 41,288,746	erythrocyte membrane protein band 4.2
5	ALAS2	184128																n/a	chrX 55,063,174	aminolevulinate, delta-, synthase 2 isoform a
6	HEMGN	176972																n/a	chr9 99,734,620	hemogen
7	RHCE	180378																n/a	chr1 25,590,638	Rhesus blood group, CcEe antigens isoform
8	ERAF	n/a																n/a	chr16 31,447,164	erythroid associated factor
9	NR_001589	n/a																9e-28	chr11 5,220,579	NR_001589
10	HBQ1	176942																5e-20	chr16 170,756	theta 1 globin
11	ALS2CR2	66341																n/a	chr2 202,039,166	amyotrophic lateral sclerosis 2 (juvenile)
12	ARG1	173628																n/a	chr6 131,941,609	arginase, type I
13	HBD	n/a																1e-59	chr11 5,211,534	delta globin
14	CTSE	175242																n/a	chr1 204,491,404	cathepsin E isoform a preproprotein
15	TAL1	152775																n/a	chr1 47,461,290	T-cell acute lymphocytic leukemia 1
16	DKFZp547J0510	170063																n/a	chr10 44,187,920	CDNA FLJ42650 fis, clone BRACE302747
17	MMRN2	178862																n/a	chr10 88,696,464	multimerin 2
18	DKFZp434N178	n/a																n/a	chr14 103,716,634	Homo sapiens cDNA: FLJ22753 fis, clone K
19	KLF1	152233																n/a	chr19 12,857,627	Kruppel-like factor 1 (erythroid)
20	SPTA1	181196																n/a	chr1 156,885,125	spectrin, alpha, erythrocytic 1
21	MYL4	179055																n/a	chr17 42,648,735	atrial/embryonic alkali myosin light chain
22	UNQ464/PRO809	183640																n/a	chr17 73,663,931	Protein UNQ464/PRO809 precursor.
23	NR_002762	n/a																n/a	chr1 31,233,783	Homo sapiens PRO0611 mRNA, complete c

# VisiGene

images of *in situ* hybridization in mouse and frog

UCSC VisiGene

DRD2 search Zoom: out in full fit Help

64 images match

Mutant Mouse *Drd2* Penk1 Th. Click image to zoom in, drag or arrow keys to move. Caption is below.

+/+ -/-

A A'

D2R

B B'

Enk

C C'

TH

D D'

BrdU

+/+ -/-

A A'

D2R

B B'

Enk

C C'

TH

D D'

BrdU

Mutant Mouse *Drd2* Penk1 Th

source: MGI Reference: [Mash1 regulates neurogenesis in the ventral telencephalon.](#)  
Year: 1999 Contributors: Casarosa S., Fode C., Guillemot F.,

# さらなる学習資料

- JSTのミニコースのサイト
  - NCBI、Ensembl、UCSCゲノムブラウザ
    - NCBIは本家の英語版を翻訳
    - チュートリアル式解説
    - 日付は2008年10月
  - <http://www-bird.jst.go.jp/minicourses/>

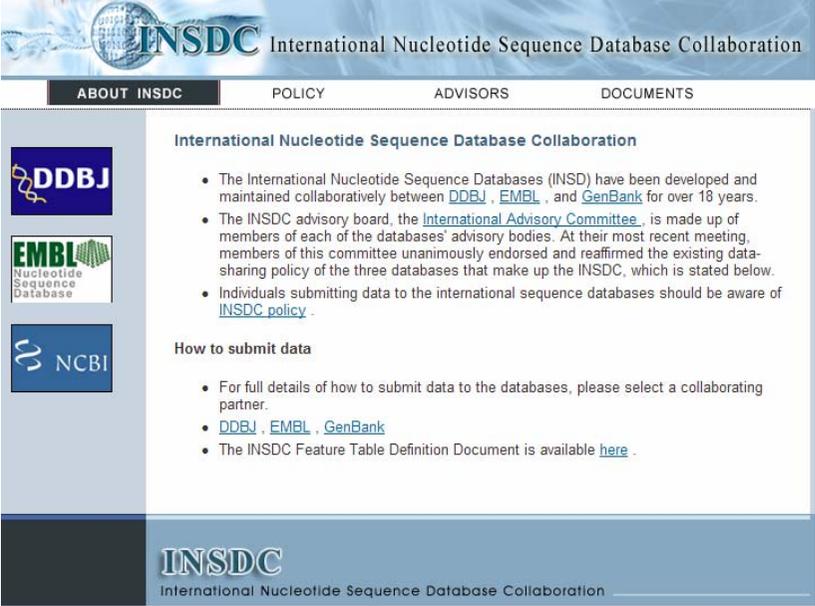
# DNA配列のデータベース

DDBJ/EMBL/GenBank

# International Nucleotide Sequence Database (INSD)

- 「DDBJ/EMBL/GenBank 国際塩基配列データベース」
- 3機関
  - 日本DNAデータバンク(DDBJ)
  - 欧州 (EMBL-EBI)
  - 米国 (NCBI)

通称として、3つのうち何れかの名で呼ばれる



The screenshot shows the INSDC website. The header includes the INSDC logo and the text "International Nucleotide Sequence Database Collaboration". Below the header is a navigation menu with links for "ABOUT INSDC", "POLICY", "ADVISORS", and "DOCUMENTS". The main content area features the INSDC logo and a list of bullet points describing the collaboration. The left sidebar contains logos for DDBJ, EMBL Nucleotide Sequence Database, and NCBI. The footer includes the INSDC logo and the text "International Nucleotide Sequence Database Collaboration".

International Nucleotide Sequence Database Collaboration

ABOUT INSDC POLICY ADVISORS DOCUMENTS

International Nucleotide Sequence Database Collaboration

- The International Nucleotide Sequence Databases (INSD) have been developed and maintained collaboratively between [DDBJ](#), [EMBL](#), and [GenBank](#) for over 18 years.
- The INSDC advisory board, the [International Advisory Committee](#), is made up of members of each of the databases' advisory bodies. At their most recent meeting, members of this committee unanimously endorsed and reaffirmed the existing data-sharing policy of the three databases that make up the INSDC, which is stated below.
- Individuals submitting data to the international sequence databases should be aware of [INSDC policy](#).

How to submit data

- For full details of how to submit data to the databases, please select a collaborating partner.
- [DDBJ](#), [EMBL](#), [GenBank](#)
- The INSDC Feature Table Definition Document is available [here](#).

INSDC  
International Nucleotide Sequence Database Collaboration

# DNA Data Bank of Japan (DDBJ)

- 塩基配列データの登録を受け付けており、登録された配列データに対してアクセッション番号を発行する
- アクセッション番号による検索、キーワード検索、相同性検索、生物分類データベース検索、パターンマッチング検索、多重整列検索、`clustalw`, `fastdnaml`, `phylip`などの解析サービスを提供している

# DDBJ (<http://www.ddbj.nig.ac.jp/>)

The screenshot shows the homepage of the DNA Data Bank of Japan (DDBJ). The header features the DDBJ logo and navigation links for Accession, DNA, Protein, Taxonomy, and Site Search. A search bar is present for accession numbers. The main navigation bar includes links for HOME, registration, search/analysis, FTP/Web API, documents, statistics, and contact, along with an English language option. The left sidebar contains a menu with categories like Introduction, User Guide, Q&A, Registration, Search, System Analysis, and Genomics. The main content area features a banner for DDBJ with a description of its role as one of the three major international DNA data banks. Below the banner are sections for Hot Topics and Maintenance Information, each with a list of recent updates and links. At the bottom, there are two yellow boxes providing details on registration and FTP/Web API services.

**DDBJ : DNA Data Bank of Japan**

DDBJ (日本DNAデータバンク) は  
欧州と米国の対応機関 (EBIおよびNCBI) と密接に協力しながら  
DDBJ/EMBL/GenBank国際塩基配列データベースを構築している三大国際DNAデータバンクのひとつです

**Hot Topics**

- ▶ 2008年7月7日 [DADリリース 44.0 公開](#) **NEW**
- ▶ 2008年6月24日 [DDBJリリース 74 完成](#)
- ▶ 2008年6月20日 [ショウジョウバエ EST 190,096 エントリの新規公開](#)

**メンテナンス情報**

- ▶ 2008年7月14日 [supemig 緊急保守作業のため所外からのアクセス停止中 \(アクセス再開 7/25 12:00\)](#)
- ▶ 2008年6月20日 [getentry での CON エントリの配列表示の不具合についてのお詫び](#)

**塩基配列の登録・更新**

- **塩基配列の登録**  
塩基配列の登録手順を御案内します。
- **登録データの修正・更新**  
登録データの修正をされる方は、最初にお読みください。

**FTP・Web API**

- **FTP** (<ftp.ddbj.nig.ac.jp>)  
DDBJリリースなどのデータファイルをダウンロードできます。
- **Web API**  
DDBJが提供するサービスをユーザーが作成したプログラムから御利用頂けます。

# DDBJ のデータ公開形式 (flat file) の説明

- 全世界の研究者が実験によって決定したDNA (または RNA) の塩基配列データを, DDBJ/EMBL/GenBank 国際 DNA データバンクが, 三者間で定めたデータ構築規範に沿って収集・編集する
- 特許から収集した配列を含む
  - 日本のJapan Patent Office (JPO), 欧州の European Patent Office (EPO), 米国の United States Patent and Trademark Office (USPTO), 韓国の Korean Intellectual Property Office (KIPO) がら
- 塩基配列データベースは, データの単位である「エン트리」の集合として構成されています。DDBJ に登録されたそれぞれのエント리는, DDBJ の定めるフォーマットにしたがった「フラットファイル (flat file) の形式」で公開されています。フラットファイルには, 塩基配列のほか, 配列の登録者, 関連文献, 由来生物種, Feature 情報などが表示されています。Feature 情報とは, その塩基配列のもつ生物学的特徴や遺伝子の機能, 特性等に関する情報を記述したもので, 記述形式は, DDBJ/EMBL/GenBank Feature Table Definition (国際塩基配列データベースが定める記述形式) によって定められています。

# DDBJの主な入力情報

- Locus
  - ローカス名、配列長、分子タイプ、分子形態、区分 (division)、最終公開日
- Definition
  - 生物名、遺伝子名、遺伝子産物名など
- Accession番号
- Version番号
- Source
  - 配列データが由来する生物の学名
- Organism
  - 由来生物の系統関係
- Reference
  - 引用文献
- Features
  - 登録配列の生物学的な特徴

試験に出る！

# DDBJエントリーの例 (D12345)

LOCUS HUM000TB03 299 bp mRNA linear EST 11-JUL-2006  
DEFINITION Homo sapiens HepG2 3'-directed MboI cDNA, clone: tb03.  
ACCESSION D12345  
VERSION D12345.1  
KEYWORDS EST (expressed sequence tag); 3'-end sequence (3'-EST).  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;  
Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 299)  
AUTHORS Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.  
and Matsubara,K.  
TITLE Large scale cDNA sequencing for analysis of quantitative aspects  
of gene expression  
JOURNAL Nature Genet. 2, 173-179 (1992)  
COMMENT PROJECT="Matsubara"

3'-EST sequences are presented as sense strand.

FEATURES Location/Qualifiers

source 1..299  
/cell\_line="HepG2"  
/cell\_type="hepatocyte"  
/clone="tb03"  
/db\_xref="taxon:9606"  
/mol\_type="mRNA"  
/organism="Homo sapiens"  
/sex="male"  
/tissue\_type="liver"

BASE COUNT 74 a 67 c 78 g 75 t

ORIGIN

1 atcatcgccn ncattaaagc agccggtgta aatggtgagc cttttggcc tggcttgttt  
61 gcaaaggccc tggccaacgt caacattggg agcctcatct gcaatgtagg ggccggtgga  
121 cctgctccag cagctggtgc tnaccagcag gaggtcctgc ccccttact gntgctgtcc  
181 agctgaggag agaaagtgga agcaaagaaa gaagaatccg aggagtctga tgatgacatg  
241 ggctttggtc ttttgacta aacctcttt ataacatggt caataaaaag ntgaacttt

//

- 分子タイプ
- 分子形態
- 区分

# 分子タイプ、分子形態、および区分

- **分子タイプ**: source feature の /mol\_type qualifier に示される分子種に基づいて DNA, RNA, mRNA, rRNA, tRNA, cRNA のいずれか
- **分子形態**: 塩基配列の分子形態が, 線状 (linear) か, 環状 (circular) か

## ●区分

■ a: 由来生物種に基づく分類

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

# Division (区分)

区分	説明	区分	説明
PAT	特許出願に含まれる塩基配列データ Japan Patent Office (JPO), European Patent Office (EPO), United States Patent and Trademark Office (USPTO), Korean Intellectual Property Office (KIPO) が収集・処理したデータ	HTG	high throughput genomic sequences ゲノムプロジェクトに由来し, 頻繁に update されることが期待されるデータで, 段階に応じて以下の 3 phase に分類されます。その配列がいずれの phase であるかは KEYWORD 行に記載されます。配列が更新されると phase が上がり, finish した後は生物種による division に移されます。  phase0piece 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 phase2piece contig の向きや順序が確定した unfinished の配列
ENV	PCR, DGGE, あるいは, その他の方法で直接, 分子を単離した環境上のサンプルに由来した配列		
SYN	synthetic constructs 人為的に構成された配列		
EST	expressed sequence tags short single pass の cDNA 配列		
TSA	transcriptome shotgun assemblies 再構成された (assembled) mRNA 配列		
GSS	genome survey sequences short single pass のゲノム配列		
HTC	high throughput cDNA sequences EST 以外の大規模 cDNA 配列プロジェクトに由来し, 更新が期待されるデータ。 配列が finish した後, 生物種による division に移される場合があります。	CON	Contig / Constructed ゲノムプロジェクトのように個々に登録された一連の配列データをデータバンク側が結合し, 1つのアクセッション番号を付与した長大なデータ。CON division に格納されるエントリには長大な配列を構成する配列のアクセッション番号が記載され, 塩基配列は記載されません。また, すでに登録されている個々のデータのエントリは, 他の division に保存されています。 CON division への直接のデータ登録は受け付けていません。まず, CON エントリを構成する個々の配列データを登録して頂きます。その上で CON エントリの構築が適当と判断された場合に, データバンク側で CON エントリを構築します。CON エントリを構成する個々の配列データが更新された場合には, CON エントリもバンク側
STS	sequence tagged sites Genome sequencing の tag となる配列。chromosome, map, PCR_condition 等の情報が必要です。		
UNA	未注釈データ 最近では UNA division は使用していません。		

# Feature (特徴), Location (配列上の位置) および Qualifier (さらに特定)

- (1) 由来生物の特徴を記述 (source)
- (2) 配列の中の一定の領域がもつ生物学的機能を記述  
e.g. CDS, rRNA, etc.
- (3) 配列の差違や変更を記述  
e.g. variation, conflict,  
etc.

```
FEATURES             Location/Qualifiers
source               1..450
                    /chromosome="12"
                    /clone="GT200015"
                    /clone_lib="lambda gt11 human liver cDNA (GeneTech.
                    No.20)"
                    /map="12p13"
                    /mol_type="mRNA"
                    /organism="Homo sapiens"
                    /tissue_type="liver"
CDS                 86..>450
                    /codon_start=1
                    /gene="GAPD"
                    /product="glyceraldehyde-3-phosphate dehydrogenase"
                    /protein_id="BAA12345.1"
                    /transl_table=1
                    /translation="MAKIK IGINGFGRIGRLVARVALQSDDELVAVNDPFITTDYMT
                    YMFKYDVTYHGQWKHHEVKVKDSKTLFLFGEKEYTVFGCRNPKEIPWGETSAEFVVEYTG
                    VFTDKDKAYAQLKGGAKKY"
BASE COUNT          102 a             119 c             131 g             98 t
ORIGIN
1  cccacgcgctc cggtcgcatc gcacttctag ctctcgacc cgcgatctca tcctctctc
61  cgccttagttc agatcgaaat cgcaaattgc gaagattaag atcgggatca atgggttcg
121  gaggatcggg aggcctcctg ccaggctggc cctgcagagc gacgacgtcg agctcgtcc
181  cgtcaacgac cccttcacat ccaccgacta catgacatac atgttcaagt atgacactgt
241  gcacggccag tggaaagcatc atgaggttaa ggtgaaggac tccaagacc ttctcttcg
301  tgagaaggag gtcaccctgt tcggctgcag gaaccctaag gagatcccat gsgstgagc
361  tagcgtgag  tttgttctgg agtacactgg tgttttact gacaaggaca aggccgttc
421  tcaacttaag ggtggtgcta agaagctctg
//
```

# 生物種ごとのデータ量ランキング (r.73, 2008)

No.	生物種名		塩基数(bp)	エントリー数
1	<i>Homo sapiens</i>	ヒト	12,940,268,373	11,340,513
2	<i>Mus musculus</i>	マウス	8,328,256,226	7,229,907
3	<i>Rattus norvegicus</i>	ラット	5,776,202,897	1,305,967
4	<i>Zea mays</i>	トウモロコシ	4,348,956,388	3,172,609
5	<i>Bos taurus</i>	ウシ	3,885,847,023	2,077,318
6	<i>Danio rerio</i>	ゼブラフィッシュ	2,853,996,623	1,585,181
7	<i>Sus scrofa</i>	ブタ	2,520,642,158	2,121,844
8	marine metagenome	(海洋メタゲノム)	2,149,478,094	2,642,942
9	<i>Strongylocentrotus purpuratus</i>	エゾパフンウニ	1,352,594,339	227,949
10	<i>Xenopus tropicalis</i>	アフリカツメガエル	1,139,151,982	1,419,723
11	<i>Oryza sativa Japonica Group</i>	イネ(日本晴)	1,109,025,101	1,185,515
12	<i>Pan troglodytes</i>	チンパンジー	982,792,454	212,723
13	<i>Arabidopsis thaliana</i>	シロイヌナズナ	907,118,588	2,204,713
14	<i>Drosophila melanogaster</i>	ショウジョウバエ	906,763,756	772,576
15	<i>Gallus gallus</i>	ニワトリ	866,657,525	802,982
16	<i>Vitis vinifera</i>	ブドウ	838,213,240	530,135
17	Unknown.	(未知)	837,143,796	1,467,573
18	<i>Macaca mulatta</i>	アカゲザル	793,208,409	76,622
19	<i>Canis lupus familiaris</i>	イヌ	705,802,525	1,225,310
20	<i>Triticum aestivum</i>	パンコムギ	661,074,571	1,103,832