

Support du cours

- Wiki:

<http://genomeast.igbmc.fr/wiki/>

- Go to: Training
- Go to [Introduction to Ensembl Genome Browser/Biomart \(DU Dijon\)](#)

Introduction à Ensembl/Biomart

Stéphanie Le Gras

Jean Muller

Objectifs

- Révision sur les banques/bases de données biologiques
- Connaitre l'existence et l'utilité des principaux "Genome browser"
- Comprendre comment fonctionne le "Genome browser : Ensembl"
- S'initier à
 - la navigation dans Ensembl
 - l'utilisation des outils d'Ensembl
 - l'utilisation de Biomart

Plan

- Introduction
 - Les banques/bases de données biologiques
 - Les “genome browsers”
- Le projet Ensembl
- Comprendre Ensembl
- Navigation dans le “genome browser” Ensembl
- Les outils intégrés à Ensembl
- Utilisation de Biomart

Les banques/BAses de données biologiques

De l'artisanat au haut débit...

1951 première séquence protéique

1967 construction d'arbres phylogénétiques

1970 algorithme de Needleman & Wunsch

1977 séquençage de l'ADN (Méthode Sanger)

premier package bioinformatique (Staden)

1978 bases de données Pir, EMBL, Genbank

1981 algorithme d'alignement local (Smith & Waterman)

1990 programme Blast

1991 étiquettes d'ADNc « EST »

1995 séquençage du génome complet d'une bactérie

1996 séquençage complet du génome de la levure

2001 première version du génome humain

=> Début de l'ère post-génomique



L'ère post-génomique

- 2002 Séquence préliminaire du génome de la souris (Waterston et al., 2002)
- 2004 ENCODE, Identification de tous les éléments fonctionnels du génome humain
- 2005 Roche 454: Séquenceur auto. haut-débit de 2ème génération par pyroséquençage : GS20
- 2007 Illumina/Solexa NGS de 2ème génération par synthèse microfluidique : GAIIx, Applied Biosystems
NGS de 2ème génération par ligation : système SOLiD



L'ère post-génomique

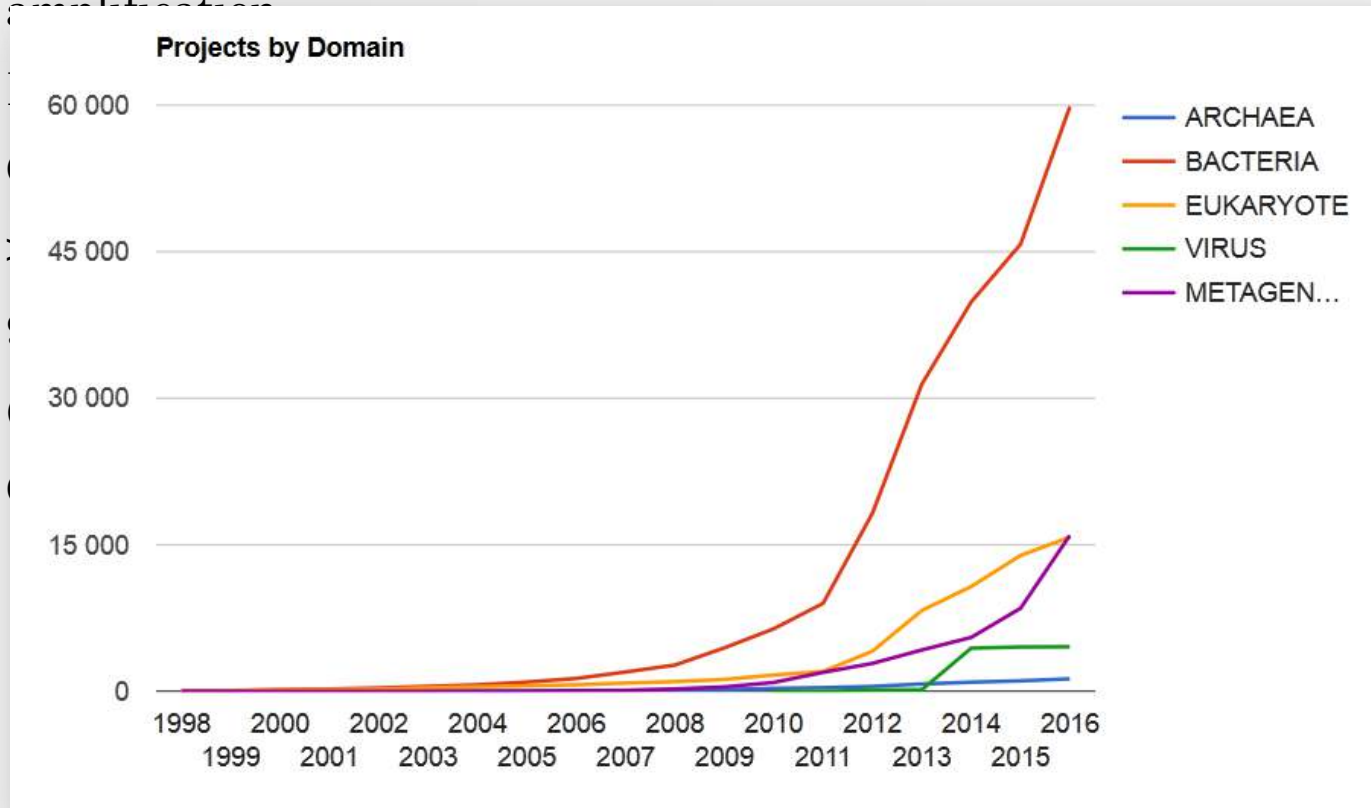
- 2008 Helicos Séquenceur auto. de 2ème génération par synthèse sans pré-amplification
- 2012 ENCODE Encyclopédie des éléments fonctionnels du génome humain
- 2014 Génome à 1000\$ 2 annonces Illumina et Life Technologies
- 2016- >40 000 génomes complets publiés (3 domaines du vivant)
956 archés, 31736 bactéries et 9173 eukaryotes (+10% vs 2014)
(www.genomesonline.org, 10/2016)
Génomes humains séquencés complètement, Exomes de patients



L'ère post-génomique

- 2008 Helicos Séquenceur auto. de 2ème génération par synthèse sans pré-

- 2012
- 2014
- 2016-



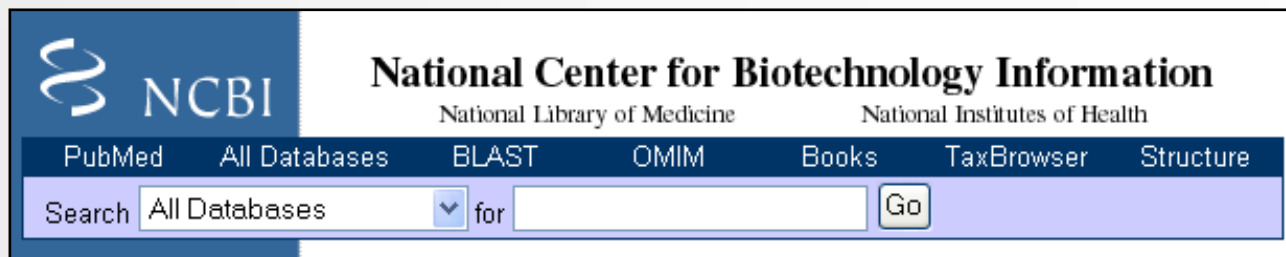
Centres de bioinformatique

- EBI (European Bioinformatics Institute)



<http://www.ebi.ac.uk/>

- NCBI (National Center for Biotechnology Information)



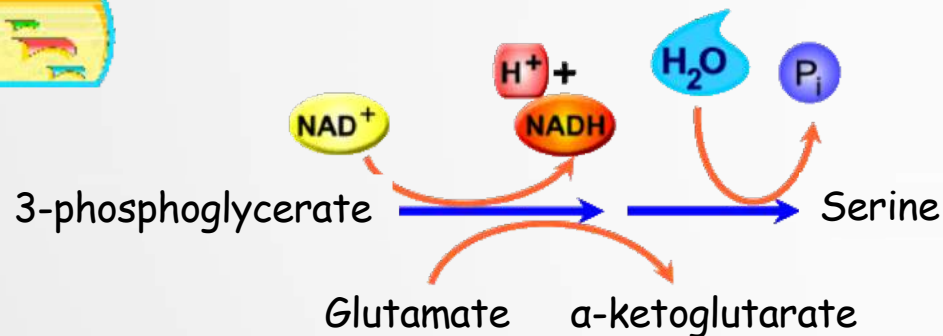
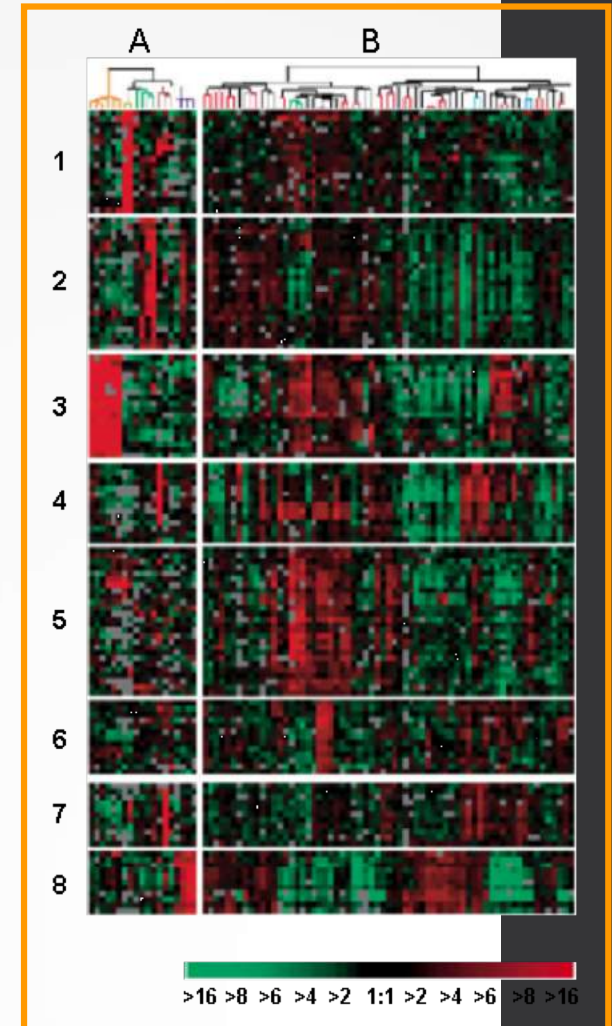
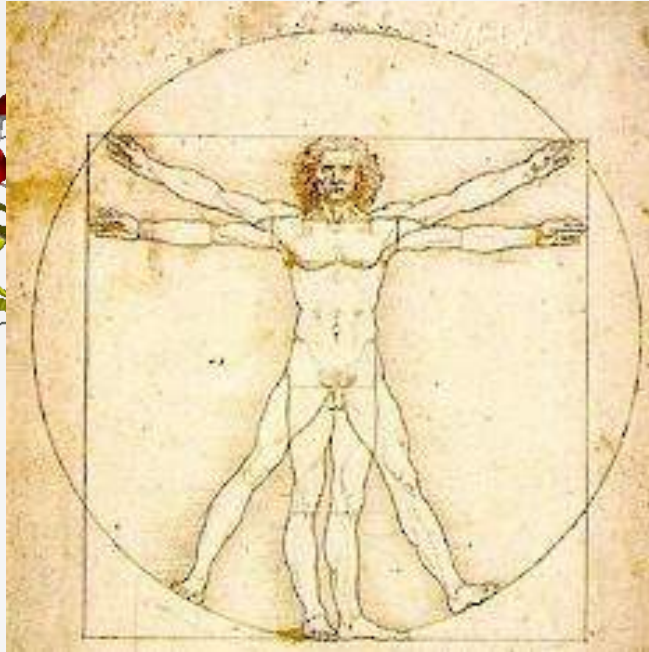
<http://www.ncbi.nlm.nih.gov/>

Banques de données en biologie moléculaire

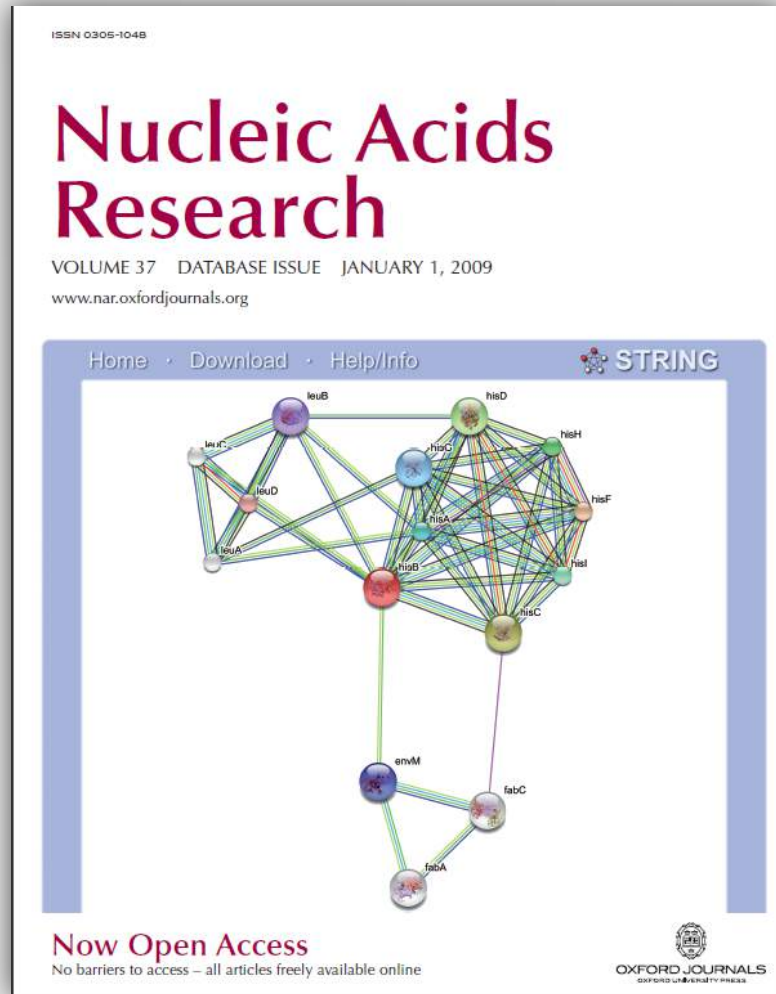
- Rôles des banques
 - Stockage
 - Diffusion (ftp, web...)
 - Organisation et standardisation des données
 - Connectivité avec autres banques
 - Actualisation

Multiplicité des banques

MALWTRLRPLLALLALWPPPPARAFVNQHLGSHLVEALYLVCGERGFYTPKARREVEGPQVGALELAGGPGA



Nucleic Acids Research (NAR) – Database issue



Contents: Volume 37, Database issue, January 2009

- ▣ [Articles](#)
- ▣ [Front-Matter/Back-Matter](#)

NAR Database Categories

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS

[Oxford Journals](#) > [Life Sciences](#) > [Nucleic Acids Research](#) > Database Summary Paper

2009 NAR Database Summary Papers

Nucleotide Sequence Databases
RNA sequence databases
Protein sequence databases
Structure Databases
Genomics Databases (non-vertebrate)
Metabolic and Signaling Pathways
Human and other Vertebrate Genomes
Human Genes and Diseases

- Protein Mutant Database
- General human genetics databases
 - BodyParts3D
 - Comparative Toxicogenomics Database
 - DG-CST
 - GenAtlas
 - GeneCards
 - Genetics Home Reference
 - HAGR - Human Ageing Genomic Resources
 - HCAD - Human Chromosome Aberration Database
 - HERVd - Human Endogenous Retrovirus database
 - HGNC Database
 - Human PAML Browser
 - MSY Breakpoint Mapper
 - MutDB
 - OMIM - Online Mendelian Inheritance in Man
 - SNP2NMD
- General polymorphism databases
- Cancer gene databases
- Gene-, system- or disease-specific databases
- Microarray Data and other Gene Expression Databases
- Proteomics Resources
- Other Molecular Biology Databases
- Organelle databases
- Plant databases
- Immunological databases

OMIM - Online Mendelian Inheritance in Man

NAR Molecular Biology Database Collection entry number 143

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

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA

Contact info@ncbi.nlm.nih.gov

Database Description

Online Mendelian Inheritance in Man (OMIM) is a comprehensive, authoritative and timely knowledgebase of human genes and genetic disorders compiled to support research and education in human genomics and the practice of clinical genetics. Started by Dr Victor A. McKusick as the definitive reference Mendelian Inheritance in Man, OMIM (<http://www.ncbi.nlm.nih.gov/omim>) is now distributed electronically by the National Center for Biotechnology Information (NCBI), where it is integrated with the Entrez suite of databases. Derived from the biomedical literature, OMIM is written and edited at Johns Hopkins University with input from scientists and physicians around the world. Each OMIM entry has a full-text summary of a genetically determined phenotype and/or gene and has numerous links to other genetic databases such as DNA and protein sequence, PubMed references, general and locus-specific mutation databases, approved gene nomenclature, and the highly detailed mapviewer, as well as patient support groups and many others. OMIM is an easy and straightforward portal to the burgeoning information in human genetics.

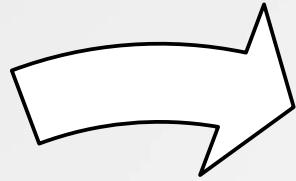
Category: [Human Genes and Diseases](#)

Subcategory: [General human genetics databases](#)

Category: [Human Genes and Diseases](#)

Subcategory: [General polymorphism databases](#)

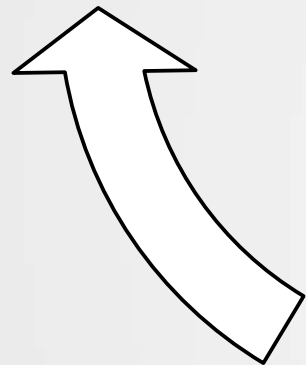
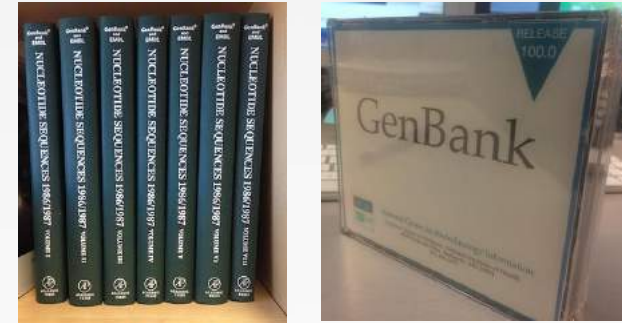
Banques de séquences nucléiques généralistes



GenBank



EMBL



DNA
databank
of Japan

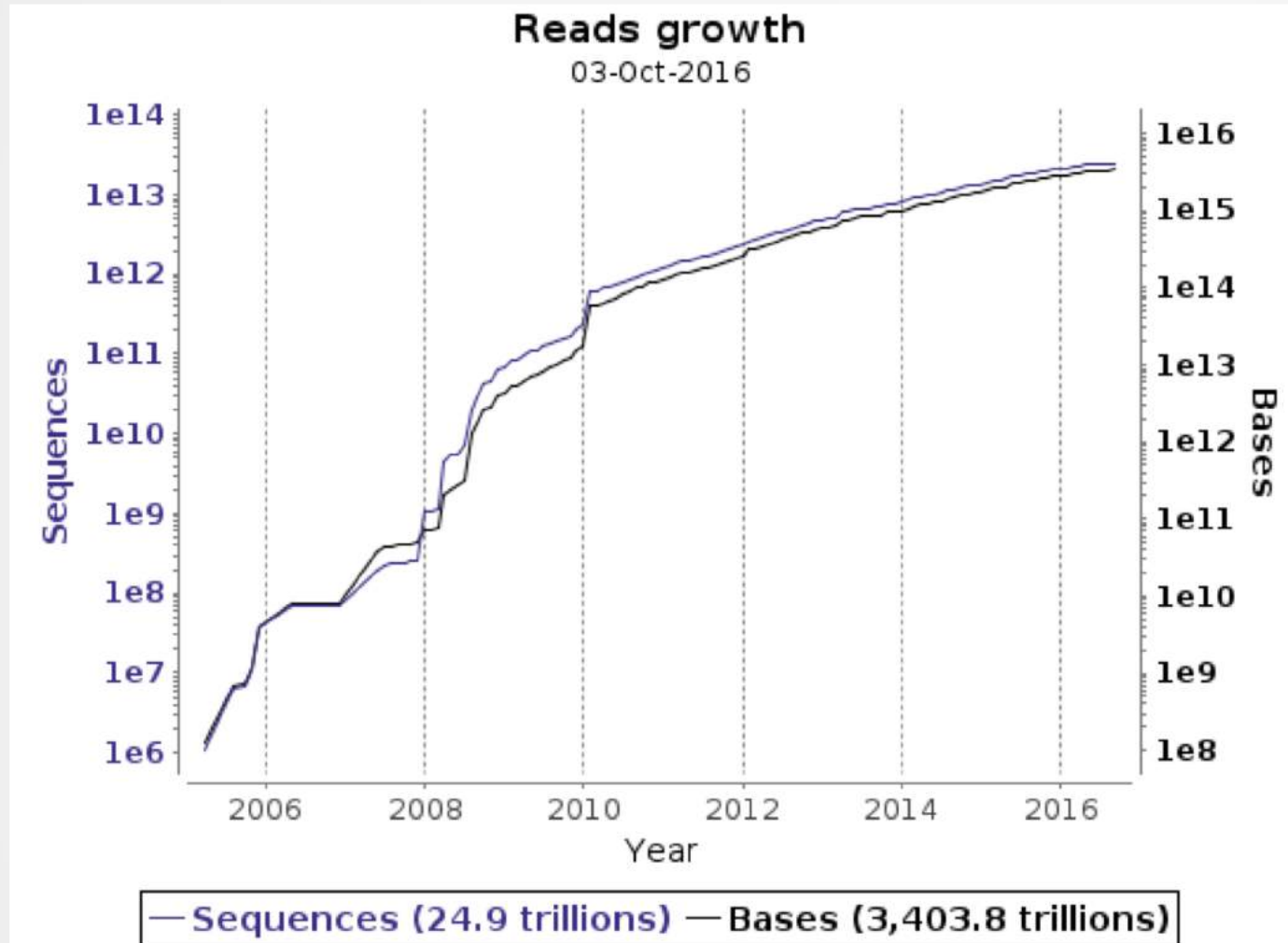


- 3 banques
- Échanges quotidiens des séquences collectées
- Effort d'unification=> format
 - accord entre GenBank et EMBL en 1986
 - accord entre GenBank/EMBL et DDBJ in 1987

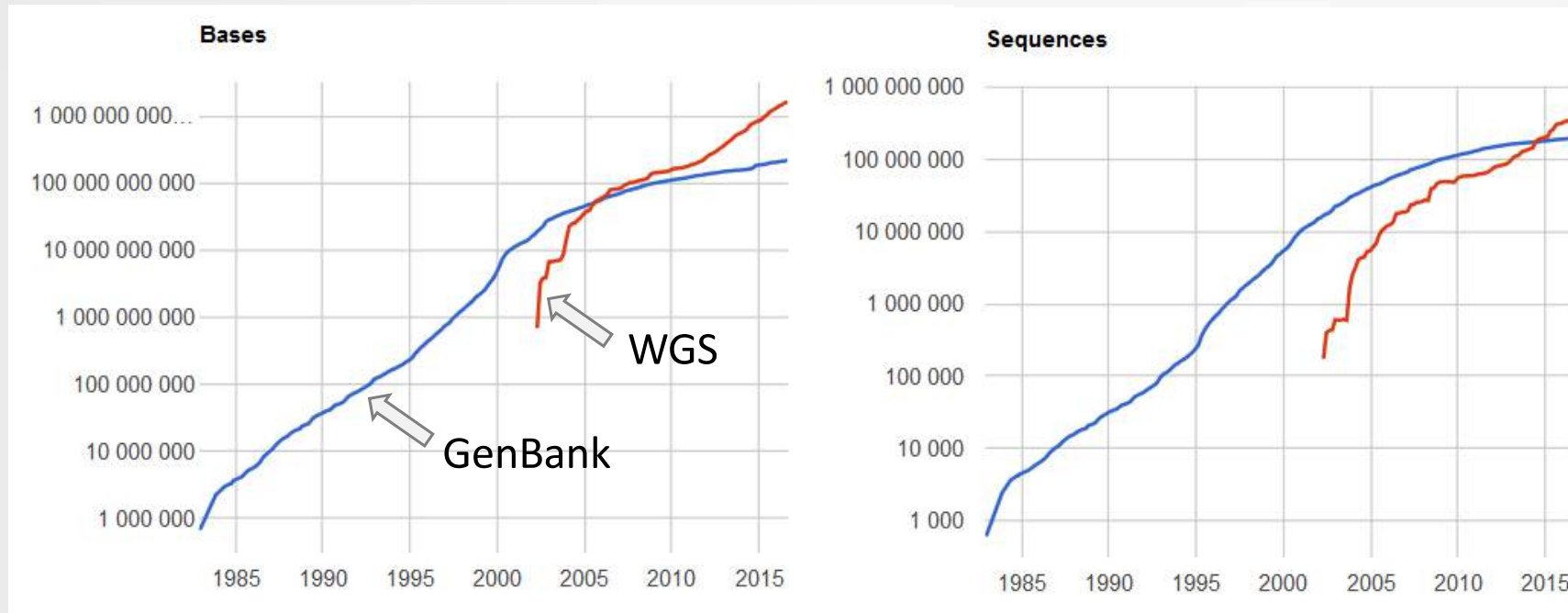
Banques de séquences

- Des banques incontournables
 - dépôt obligatoire dans une des 3 banques avant publication
 - unique moyen d'accès aux séquences
- Alimentation
 - soumission directe par la communauté scientifique (associée ou non à une publication)
 - dépôts de brevets
- Conséquences
 - banques exhaustives
 - banques extrêmement redondantes
 - contiennent des erreurs

Evolution de la banque EMBL



Evolution de la banque GenBank



06/2016: 211 milliards de nucléotides, 195 millions d'entrées

Banques de séquences protéiques généralistes



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

10/2013
41,958,567

09/2014
45,166,402

09/2016
70,427,238

- 2 banques majeures
- Qualité variable/stabilisée
- Exhaustivité / Annotation



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

09/2013
43,362,837

10/2014
85,501,513

10/2016
68,493,254

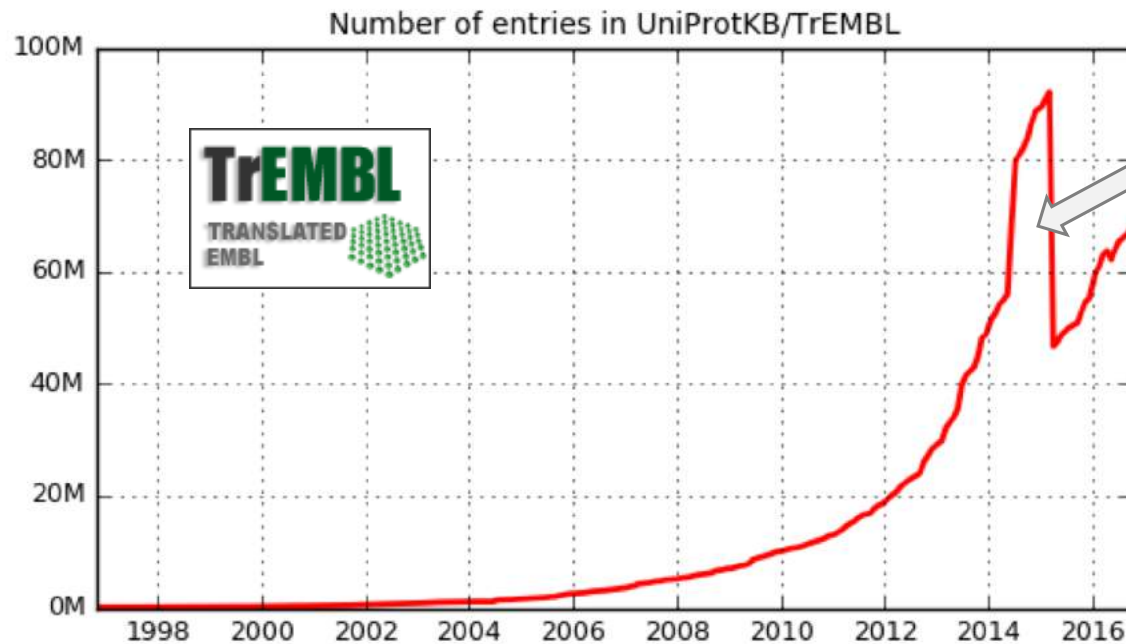


TrEMBL:
67,940,995 entrées

Swiss-Prot:
552 259 entrées

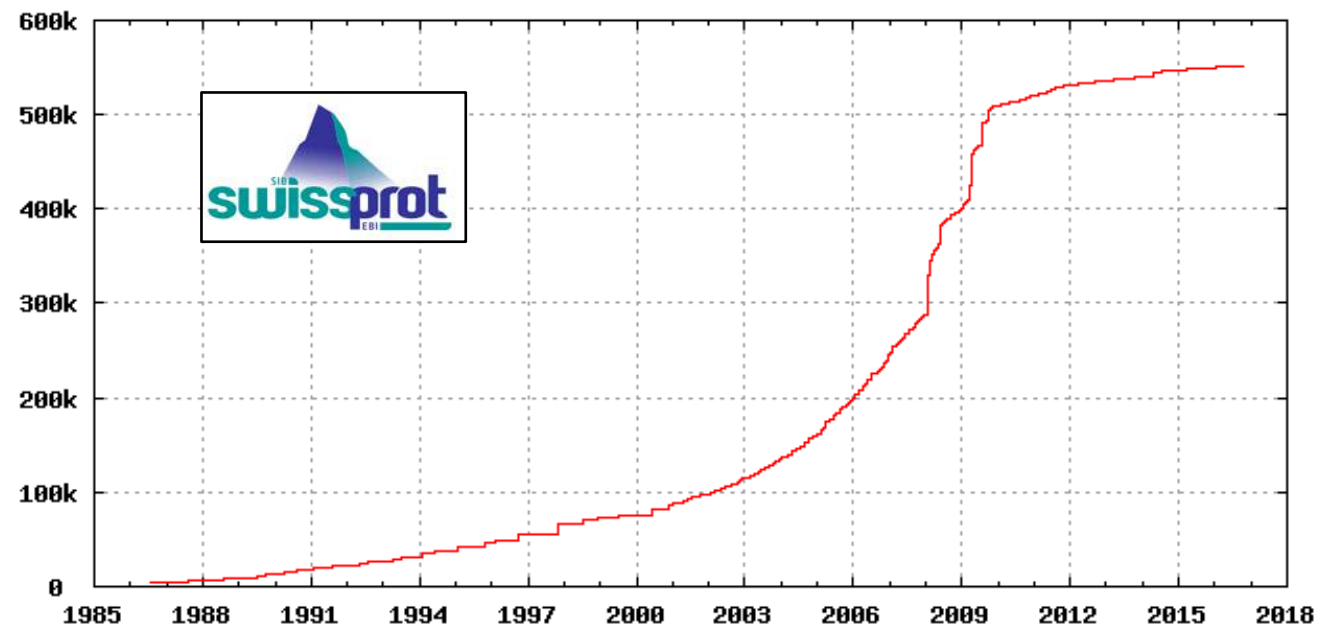
Annotation	UniProt		TrEMBL	
Evidence at protein level	90,921	16,5%	118,013	0,2%
Evidence at transcript level	57,673	10,5%	971,005	1,8%
Inferred from homology	387,632	70,5%	11,091,443	21,1%
Predicted	11,465	2,1%	40,603,140	76,9%
Uncertain	1,955	0,4%	0	0%

Evolution des bases de données protéiques

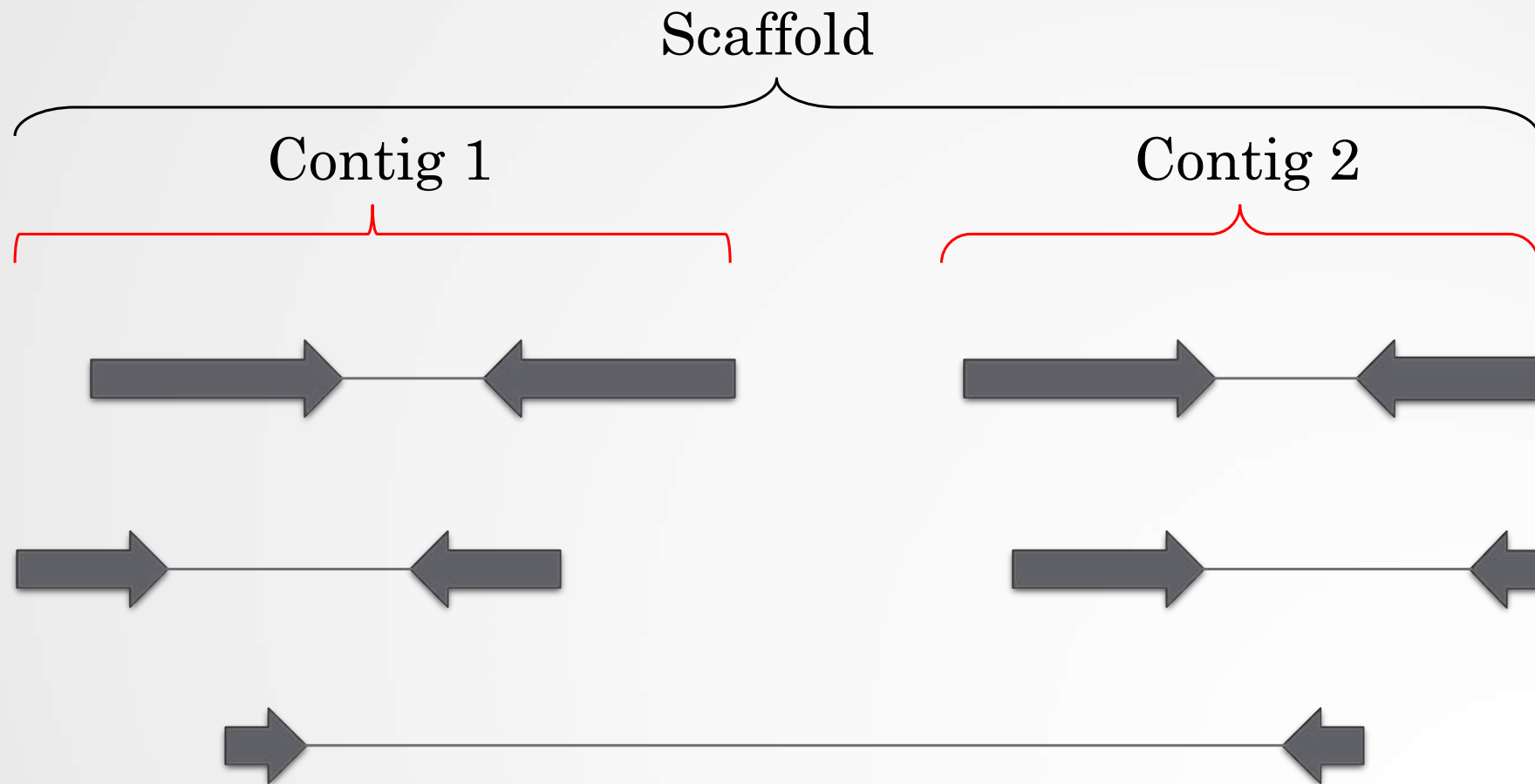


Redondance

Number of entries in UniProtKB/Swiss-Prot



Assemblage de génome



Human Genome Builds

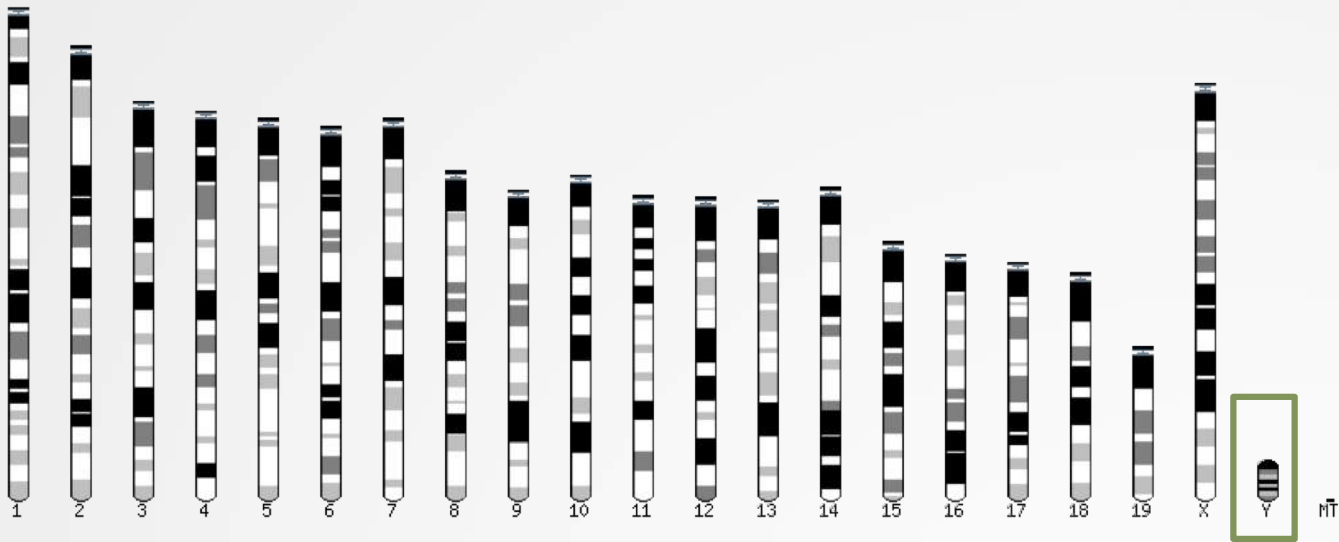
SPECIES	UCSC VERSION	RELEASE DATE	RELEASE NAME	STATUS
MAMMALS				
Human	hg38	Dec. 2013	Genome Reference Consortium GRCh38	Available
	hg19	Feb. 2009	Genome Reference Consortium GRCh37	Available
	hg18	Mar. 2006	NCBI Build 36.1	Available
	hg17	May 2004	NCBI Build 35	Available
	hg16	Jul. 2003	NCBI Build 34	Available
	hg15	Apr. 2003	NCBI Build 33	Archived
	hg13	Nov. 2002	NCBI Build 31	Archived
		Jun. 2002	NCBI Build 30	Archived
		Apr. 2002	NCBI Build 29	Archived (data only)
		Dec. 2001	NCBI Build 28	Archived (data only)
		Aug. 2001	UCSC-assembled	Archived (data only)
		Apr. 2001	UCSC-assembled	Archived (data only)
		Dec. 2000	UCSC-assembled	Archived (data only)
		Oct. 2000	UCSC-assembled	Archived (data only)
	hg4	Sep. 2000	UCSC-assembled	Archived (data only)
		Jul. 2000	UCSC-assembled	Archived (data only)
		Jun. 2000	UCSC-assembled	Archived (data only)
		May 2000	UCSC-assembled	Archived (data only)



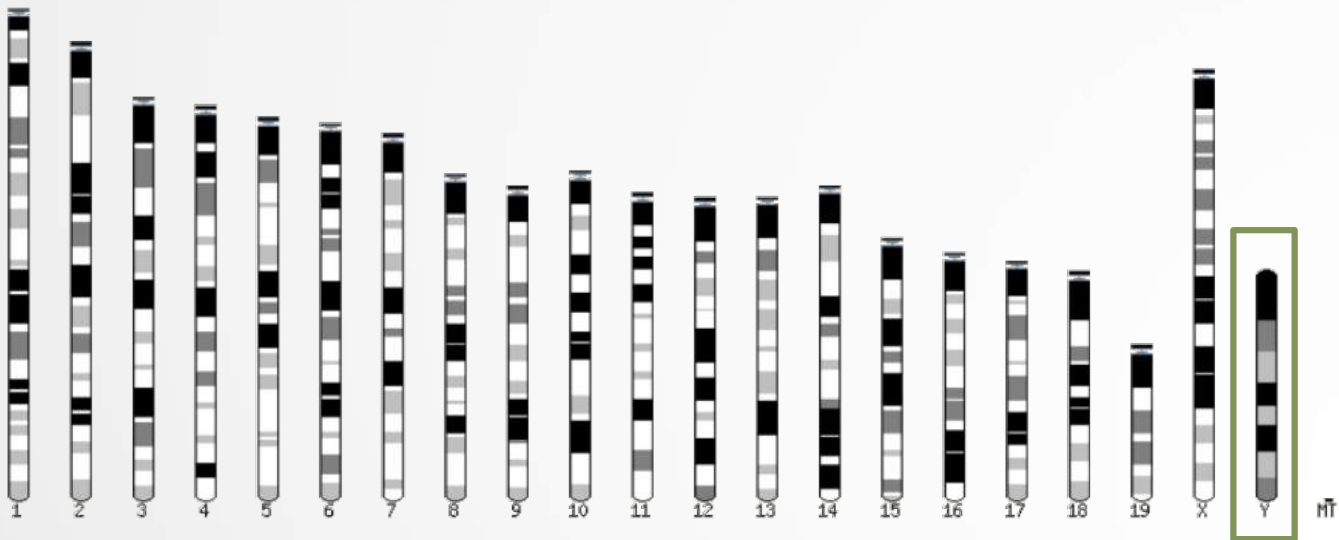
Source: <https://genome.ucsc.edu/FAQ/FAQreleases.html>

Genome builds

mm9



mm10



Genome browsers

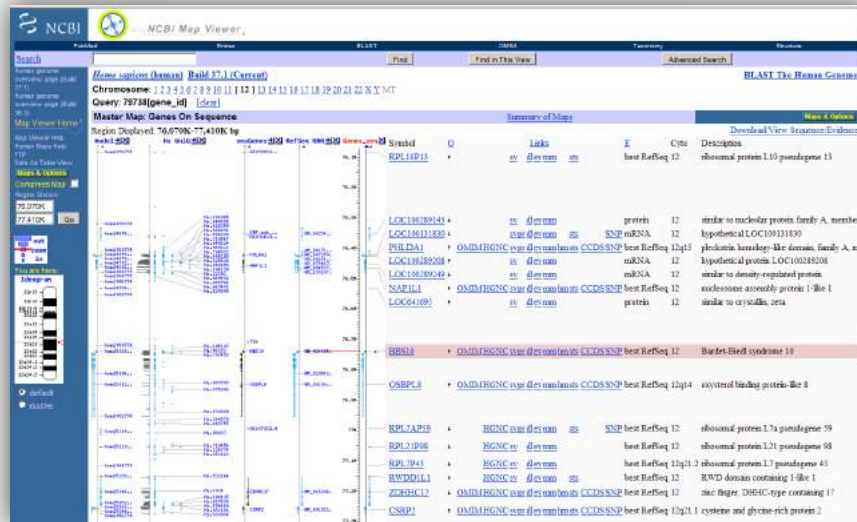
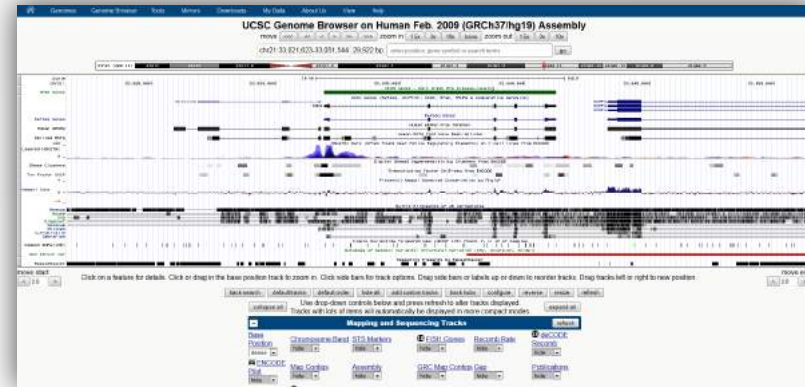
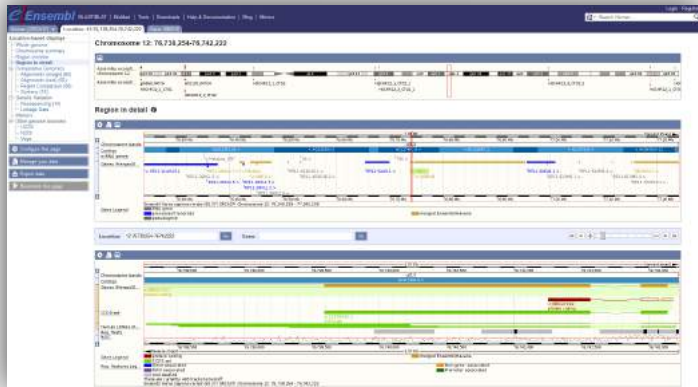
Genome Browsers – L'outil de référence

- Élément de référence absolue le **génom**e
 - Agrégateur et générateur d'informations/annotations
 - Prédiction
s de gènes
 - Protéines
 - Données d'expression
 - Variations
- Synthèse rapide et visuelle de données primordiales

Il y a Genome Browsers...

EBI - Ensembl

UCSC - Genome Browser



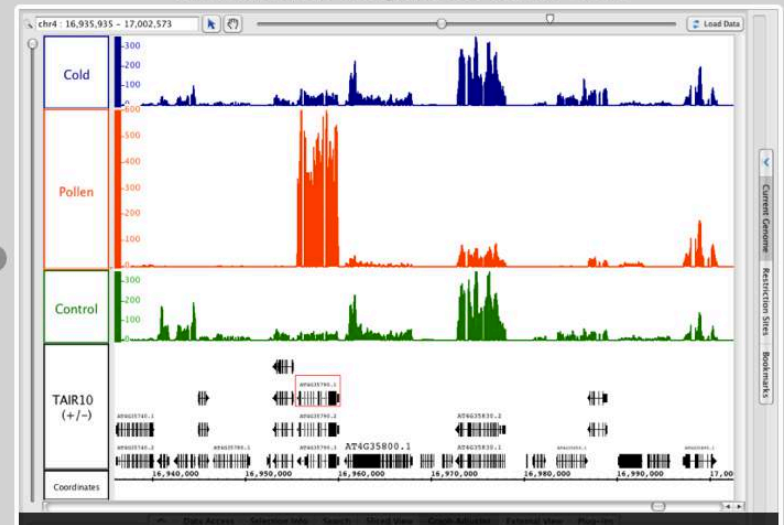
NCBI - Map Viewer

Et Genome browsers



**Integrative
Genomics
Viewer**

Integrated Genome Browser
Visualization for genome-scale data



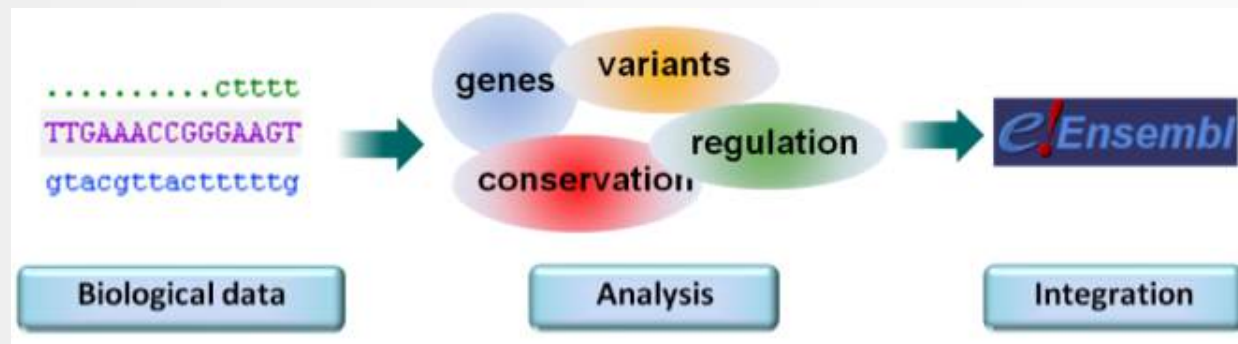
Ensembl

Le projet Ensembl

- Initié en 1999 (avant la première version du génome humain)
- Projet en collaboration entre l'European Bioinformatics Institute (EBI) et le Wellcome Trust Sanger Institute (WTSI)
- Objectif :
 - Annoter automatiquement les génomes
 - Ajouter des données biologiques aux annotations
 - Rendre publique les annotations sur le web
- Ensembl ne produit pas ses propres données d'assemblage de génome!

Le projet Ensembl

- Données disponibles :
 - Génomes
 - Données de génomique comparative
 - Variations
 - Élément régulateur des gènes
 - Annotations externes



- Lancement du site web en juillet 2000 (au début il n'y avait que le génome humain)

Les génomes d'Ensembl

- Espèces de vertébrés dans <http://ensembl.org>
- EnsemblGenomes (avril 2009)
 - Métazoaires : <http://metazoa.ensembl.org>
 - Bactéries : <http://bacteria.ensembl.org>
 - Plantes : <http://plants.ensembl.org>
 - Fungi : <http://fungi.ensembl.org>
 - Protistes : <http://protists.ensembl.org>

L'interface web



BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

Login/Register

Search all species...

Tools

[All tools](#)

BioMart >

Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >

Search our genomes for your DNA or protein sequence

Variant Effect Predictor >

Analyse your own variants and predict the functional consequences of known and unknown variants

Search

All species for

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes

-- Select a species --

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

Favourite genomes



Human
GRCh38.p12

[Still using GRCh37?](#)



Mouse
GRCm38.p6



Zebrafish
GRCz11

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 95 (January 2019)

- New regulatory build for human, incorporating new data from ENCODE
- Update to GENCODE M20 for mouse
- New genomes: donkey, polar bear, black bear, red fox, koala, dingo, tuatara, painted turtle and desert tortoise
- Updated genomes for chicken, cow and horse
- New protein structure variation view

[More release news](#) on our blog

Other news from our blog

- 08 Mar 2019: [Joint REST server for Ensembl and Ensembl Genomes in Ensembl 96](#)
- 07 Mar 2019: [Removal of database patches script from Ensembl repository in Ensembl 96](#)
- 01 Mar 2019: [Getting to know us: Guy from Ensembl Plants](#)

Compare genes across species



Find SNPs and other variants for my gene

```
GGTTATACATTC
CRTRAAAGTCTT
CTTCTAATTCT
GACACATTTTCC
```

Gene expression in different tissues



Retrieve gene sequence

```
GGCTGACTTCGGATGG
GGGCTTGTGCGCGAGCT
GGGCTTCTGCTGCGGCTT
AGGGACAGATTGTGTAM
GACCTCTGAGACGGTGT
CCCACTCAGCTGAGCG
```

Find a Data Display



Use my own data in Ensembl



Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available.

Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.



Comprendre ENSEMBL

Les annotations

- 3 à 6 mois
- Annotation par Ensembl
 - Annotation automatique (Ensembl Genebuild) :
 - Détermination des transcrits dans le génome entier
 - Basées sur des séquences d'ARNm et protéiques extraites des banques de données publiques
 - *Curation* manuelle : au cas par cas. Ex: l'humain, la souris, le rat, le zebrafish + autres vertébrés (produit par le groupe HAVANA du WTSI)
 - Fusion des annotations automatiques et manuelles (Gold)
- + Annotations importées depuis flyBase, WormBase, SGD

Les annotations

- Les transcrits d'Ensembl sont basés sur les bases de données suivantes :
 - Uniprot/Swiss-Prot (*curation* manuelle)
 - Uniprot/TrEMBL
 - NCBI refSeq (*curation* manuelle)

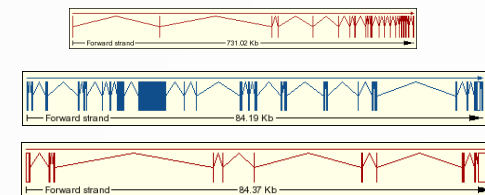


Protéine/ ARNm

+



Assemblage de séquence



Gène Ensembl

Les annotations

- Les annotations des gènes peuvent varier entre les différents genome browsers (Ensembl, UCSC, NCBI)
- CCDS (Consensus CDS) est un jeu de données de gènes codants validés par tous les membres du consortium (EBI, HGNC, MGI, NCBI, WTSI)
 - <http://www.ncbi.nlm.nih.gov/CCDS/CcidsBrowse.cgi>
 - Il faut que l'assemblage du génome soit suffisamment stable pour identifier les gènes dont les positions sont identiques entre les différentes sources (chez humain et souris)

preEnsembl

Pre!Ensembl

BLAST | Downloads | Help & Documentation | Blog

Login/Register

Search all species...

About Pre Ensembl

The Ensembl pre-build provides displays of genomes that are in the process of being created.

Not all standard Ensembl displays are available.

Genomes are put up when we have done the initial BLAST analysis on a new assembly. We have not completed the gene build. It is provided as an "early view" site for our users.

Owing to the preliminary nature of the data, Pre-Ensembl provides views of the assembly and gene models against the assembly and download of the data. In general a full Ensembl release takes months to complete the data are and

Occasional updates but with data.

Many of features

Download

You can FTP site. These files helpdes favourite

Last update

Pre Species



Aardvark
OryAfe1.0



Budgerigar
MelUnd6.3



Painted turtle
ChrPicBel3.0.1



Pig (map)



Dolphin
Ttru_1.4



Fugu
FUGU5



Pika
OchPri3



Rhinoceros
CerSimSim1



Hamadryas baboon
Pham



Hedgehog
EriEur2



Shrew
SorAra2.0

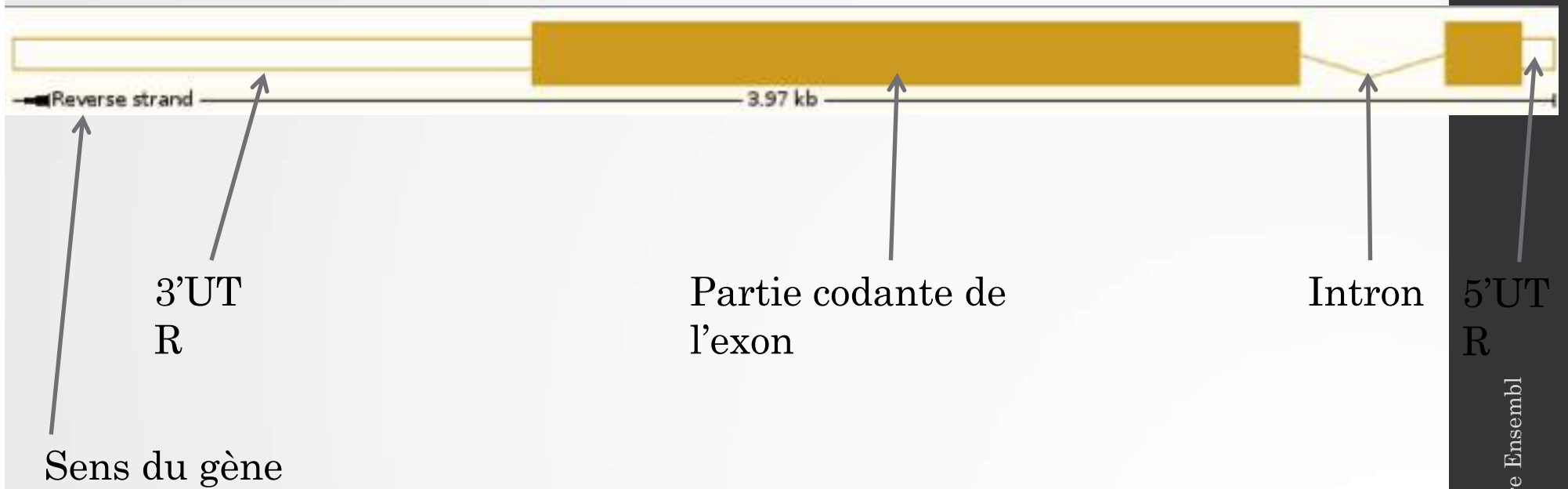


Sperm whale
PhyMac_2.0.2

- Dédié aux génomes en train d'être annotés.
- Moins de fonctionnalités que pour les génomes dont l'annotation est terminée

[Privacy Policy](#) | [Contact Us](#)

Transcrits Ensembl



Identifiants Ensembl


- ENS**G**### Ensembl **Gene** ID
- ENS**T**### Ensembl **Transcript** ID
- ENS**P**### Ensembl **Peptide** ID
- ENS**E**### Ensembl **Exon** ID

- Ajout d'un suffix pour les autres espèces
 - MUS (*Mus musculus*) pour la souris: ENS**MUSG**###
 - DAR (*Danio rerio*) pour le zebrafish: ENS**DARG**###
 - etc.

Version (Release)

- ~ tous les 3 mois
- Lien vers la dernière version d'Ensembl est toujours :
<http://www.ensembl.org>



EMBL-EBI  Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available. Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.




Ensembl release 95 - January 2019 © [EMBL-EBI](#)

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

- Lien vers une version particulière d'Ensembl :
<http://Jan2019.archive.ensembl.org/index.html>

Ensembl : Archives

 BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

Using this website | Annotation and prediction | Data access | API & software | About us


In this section

Archives: Table of assemblies

Search documentation.

Ensembl Archives

About Archive Ensembl



The main Ensembl site (www.ensembl.org) and the mirror sites are updated with the latest data approximately every three months. We maintain the Ensembl Archive sites so that there are stable links to data from a particular release. As of December 2016 these will be available for five years, together with the following longer term archives:

- Annotation on the **human NCBI36 assembly** is available at our [Ensembl 54 archive](#) site.
- Annotation on the **mouse NCBI37 assembly** is available at our [Ensembl 67 archive](#) site.
- As from August 2014 we are supporting the **human GRCh37 assembly** at our dedicated [GRCh37 human](#) site. Unlike the other Ensembl archive sites, this will be updated to the latest web interface every Ensembl release and there may be occasional data updates to human.

Archived databases are also maintained for at least 10 years. Currently all databases are available from 2004. More information is available from our [MySQL database documentation](#). We also maintain data archives from 2004 available from our [FTP site](#).

For all enquiries, please [contact the Ensembl HelpDesk](#).

Notes

- Ensembl aims to maintain stable identifiers for genes (ENSG), transcripts (ENST), proteins (ENSP) and exons (ENSE) as long as possible. Changes within the genome sequence assembly or an updated genome annotation may dramatically change a gene model. In these cases, the old set of stable IDs is retired and a new one assigned. Gene and transcript pages both have an ID History view which maps changes in the ID from the earliest version in Ensembl.
- Protein family identifiers (fam), Ensembl EST gene identifiers (ENSESTG) and Genscan identifiers (GENSCAN) are currently not stable.
- With the exception of the GRCh37 human site **BLAST, BLAT and other tools** are not available from the archive sites.
- **Accounts** are shared between the current site and almost all archives. The exceptions are the older human NCBI36 and the mouse GRCh37 sites where changes in architecture and code make sharing logins impractical.

Linking to the Archive Ensembl sites

The Archive Ensembl sites have the format: <http://<three-letter-month><year>.archive.ensembl.org> for example <http://nov2008.archive.ensembl.org>

In the footer of each current Ensembl page, there is a link called 'Permanent link', which links to the corresponding page in the Ensembl Archive. A similar link on each archive page links back to the current site (i.e. www.ensembl.org).

For example if you are looking at the Alternative Splicing view for human gene BRCA2 on the [main Ensembl site](#) in August 2015, when Ensembl 80 was the current version, the URL would be:

http://www.ensembl.org/Homo_sapiens/Gene/Splice?db=core;q=ENSG00000139618;r=13:31787617-31871809;t=ENST00000380152

and the equivalent archived page URL would be:

http://jul2015.archive.ensembl.org/Homo_sapiens/Gene/Splice?db=core;q=ENSG00000139618;r=13:31787617-31871809;t=ENST00000380152

Unfortunately, owing to the change in site organisation between releases it is not always possible to map pages one-to-one between the current Ensembl site and the older archives. If the link does not take you to the data you expected, trying using the search facility to locate the information.

Ensembl release 95 - January 2019 © [EMBL-EBI](#)

[Permanent link](#)

List of currently available archives

- [Ensembl GRCh37](#): Full Feb 2014 archive with BLAST, VEP and BioMart
- [Ensembl 95: Jan 2019](#) - this site
- [Ensembl 94: Oct 2018](#)
- [Ensembl 93: Jul 2018](#)
- [Ensembl 92: Apr 2018](#)
- [Ensembl 91: Dec 2017](#)
- [Ensembl 90: Aug 2017](#)
- [Ensembl 89: May 2017](#)
- [Ensembl 88: Mar 2017](#)
- [Ensembl 87: Dec 2016](#)
- [Ensembl 86: Oct 2016](#)
- [Ensembl 85: Jul 2016](#)
- [Ensembl 84: Mar 2016](#)
- [Ensembl 83: Dec 2015](#)
- [Ensembl 82: Sep 2015](#)
- [Ensembl 81: Jul 2015](#)
- [Ensembl 80: May 2015](#)
- [Ensembl 79: Mar 2015](#)
- [Ensembl 78: Dec 2014](#)
- [Ensembl 77: Oct 2014](#)
- [Ensembl 76: Aug 2014](#)
- [Ensembl 75: Feb 2014](#)
- [Ensembl 67: May 2012](#)
- [Ensembl 54: May 2009](#)

[Table of archives showing assemblies present in each one.](#)

<http://www.ensembl.org/info/website/archives/index.html>

Ensembl : Archives

Archive! **Ensembl**

BioMart | Downloads | Help & Docs | Blog

Login/Register

Search all species...

Tools

[All tools](#)

BioMart >

Export custom datasets from Ensembl with this data-mining tool

Search

All species for

Go

e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)

All genomes

-- Select a species --

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

Favourite genomes



Human
GRCh38.p12

[Still using GRCh37?](#)



Mouse
GRCm38.p6



Zebrafish
GRCz11

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Archive Release 94 (October 2018)

- New fish: 38 new and updated fish genomes
- GENCODE update 29 for human and M19 for mouse
- Additional pathogenicity predictors for missense variants
- New transcription factor binding motifs from SELEX
- Gene trees using HMMs

[More release news](#) on our blog

Other news from our blog

- 22 Feb 2019: [Cool stuff the Ensembl VEP can do: take a REST!](#)
- 21 Feb 2019: [Job: Applications software developers!](#)
- 11 Feb 2019: [Ensembl insights: Annotating readthrough transcription in Ensembl!](#)

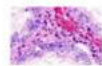
Compare genes across species



Find SNPs and other variants for my gene

GATATACATTCC
CCTAAAGTCTT
CTTCTAATTCT
GACACATTTTCC

Gene expression in different tissues



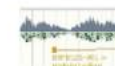
Retrieve gene sequence

```
GCCGACTTCGCGGTGG  
GGGCTTGTGGCCGAGC  
GGGCTCTACTGAGGCT  
AGGGGACAGATTTGTA  
CACCTCTGAGAGGGTT  
CCGACTCCAGGTGGC
```

Find a Data Display



Use my own data in Ensembl



Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available.

Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.



Ensembl release 94 - October 2018 © EMBL-EBI

Les anciennes version d'Ensembl sont conservées pendant 5 ans sauf si elles contiennent la dernière version de l'annotation d'un génome.

Ensembl : Archives

- <http://www.ensembl.org/info/website/archives/assembly.html>

	Jan 2019 v95	Oct 2018 v94	Jul 2018 v93	Apr 2018 v92	Dec 2017 v91	Aug 2017 v90	May 2017 v89	Mar 2017 v88	Dec 2016 v87	Oct 2016 v86	Jul 2016 v85	Mar 2016 v84	Dec 2015 v83	Sep 2015 v82	Jul 2015 v81	May 2015 v80	Mar 2015 v79	Dec 2014 v78	Oct 2014 v77	Aug 2014 v76	Feb 2014 v75
Horse	EquCab3.0	Equ Cab 2																			
Human	GRCh38.p12	GRCh38.p10				GRCh38.p7			GRCh38.p5		GRCh38.p3		GRCh38.p2		GRCh38			GRCh37.p13			
Hyrax	proCap1																				
Indian medaka	Om_v0.7.RACA																				
Japanese medaka HNI	ASM223471v1																				
Japanese medaka HSOK	ASM223469v1																				
Japanese medaka HdrR	ASM223467v1	HdrR																			
Kangaroo rat	Dord_2.0	dipOrd1																			
Koala	phaCin_tgac_v2.0																				
Lamprey	Pmarinus_7.0																				
Leopard	PanPar1.0																				
Lesser Egyptian jerboa	JacJac1.0																				
Lesser hedgehog tenrec	TENREC																				
Long-tailed chinchilla	ChiLan1.0																				
Lyretail cichlid	NeoBri1.0																				

Aide et documentations

- Vidéo Youtube (workshop...)
- FAQ
- Exercices
- Cours en ligne
- Publications :
 - Flicek, P. et al. **Ensembl 2013**. Nucleic Acids Res. Advanced Access (Database Issue).
<http://www.ncbi.nlm.nih.gov/pubmed/23203987>
 - Xosé M. Fernández-Suárez and Michael K. Schuster. **Using the Ensembl Genome Server to Browse Genomic Sequence Data**. UNIT 1.15 in Current Protocols in Bioinformatics, Jun 2010
 - Giulietta M Spudich and Xosé M Fernández Suárez. **Touring Ensembl: A practical guide to genome browsing**. BMC Genomics 2010, 11:295 (11 May 2010)

Naviguer dans ensembl

www.ensembl.org



BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

Login/Register

Search all species...

Tools

[All tools](#)

BioMart >

Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >

Search our genomes for your DNA or protein sequence

Variant Effect Predictor >

Analyse your own variants and predict the functional consequences of known and unknown variants

Search

All species for

Go

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes

-- Select a species --

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

Favourite genomes



Human
GRCh38.p12

[Still using GRCh37?](#)



Mouse
GRCm38.p6



Zebrafish
GRCz11

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 95 (January 2019)

- New regulatory build for human, incorporating new data from ENCODE
- Update to GENCODE M20 for mouse
- New genomes: donkey, polar bear, black bear, red fox, koala, dingo, tuatara, painted turtle and desert tortoise
- Updated genomes for chicken, cow and horse
- New protein structure variation view

[More release news](#) on our blog

Other news from our blog

- 08 Mar 2019: [Joint REST server for Ensembl and Ensembl Genomes in Ensembl 96](#)
- 07 Mar 2019: [Removal of database patches script from Ensembl repository in Ensembl 96](#)
- 01 Mar 2019: [Getting to know us: Guy from Ensembl Plants](#)

Compare genes across species



Find SNPs and other variants for my gene

```
GGTATACATTC
CRTRAAAGTCTT
CTTCTAAATTCT
GACACATTTTCC
```

Gene expression in different tissues



Retrieve gene sequence

```
GGCTACTTCGGGTTG
GGGCTTGTGCGCGAGC
GGGCTTGTGCGCGAGC
AGGGACAGATTGTGAM
GACCTCTGAGAGCGTIT
CCCACTCAGCTGAGCG
```

Find a Data Display



Use my own data in Ensembl



Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available.

Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.



Ensembl Genomes

Bactéries

Protistes

The screenshot shows the Ensembl Bacteria homepage. It features a search bar at the top with the text "Search Ensembl Bacteria". Below the search bar, there are sections for "Search for a gene" and "Search for a genome". A "What's New in Release 21" section is visible on the left, listing updates such as "44,646 genomes (15,162 bacteria and 29,484 archaea) from BSRG accessions". A "Did you know...?" section on the right provides information about the Ensembl database structure.

The screenshot shows the Ensembl Protists homepage. It features a search bar at the top with the text "Search Ensembl Protists". Below the search bar, there are sections for "Search for a gene" and "Search for a genome". A "What's New in Release 21" section is visible on the right, listing updates such as "New genomes" and "Updated data". A "Did you know...?" section on the left provides information about the Ensembl database structure.

Plantes

The screenshot shows the Ensembl Plants homepage. It features a search bar at the top with the text "Search Ensembl Plants". Below the search bar, there are sections for "Search for a gene" and "Search for a genome". A "What's Now in Release 21" section is visible on the left, listing updates such as "Updated gene models for *Oryza sativa* from v2.1.0". A "Did you know...?" section on the right provides information about the Ensembl database structure.

The screenshot shows the Ensembl Genomes homepage. It features a search bar at the top with the text "Search Ensembl Genomes". Below the search bar, there are sections for "Search for a gene" and "Search for a genome". A "What's Now in Release 21" section is visible on the right, listing updates such as "New genomes" and "Updated data". A "Did you know...?" section on the left provides information about the Ensembl database structure.

Fungi

The screenshot shows the Ensembl Fungi homepage. It features a search bar at the top with the text "Search Ensembl Fungi". Below the search bar, there are sections for "Search for a gene" and "Search for a genome". A "What's Now in Release 21" section is visible on the left, listing updates such as "New genomes" and "Updated data". A "Did you know...?" section on the right provides information about the Ensembl database structure.

Métazoaires

The screenshot shows the Ensembl Metazoa homepage. It features a search bar at the top with the text "Search Ensembl Metazoa". Below the search bar, there are sections for "Search for a gene" and "Search for a genome". A "What's Now in Release 21" section is visible on the right, listing updates such as "New genomes" and "Updated data". A "Did you know...?" section on the left provides information about the Ensembl database structure.

Le site web Ensembl: page d'accueil

Outils

Recherche



Tools

BioMart >

BLAST/BLAT >

Variant Effect Predictor >

[All tools](#)

Export custom datasets from Ensembl with this data-mining tool

Search our genomes for your DNA or protein sequence

Analyse your own variants and predict the functional consequences of known and unknown variants

Search

All species for

Go

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

Recherche

All genomes

-- Select a species --

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

Liste déroulante
Accès aux
génomés

Favourite genomes

- Human
GRCh38.p12
[Still using GRCh37?](#)
- Mouse
GRCm38.p6
- Zebrafish
GRCz11

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 95 (January 2019)

- New regulatory build for human, incorporating new data from ENCODE
- Update to GENCODE M20 for mouse
- New genomes: donkey, polar bear, black bear, red fox, koala, dingo, tuatara, painted turtle and desert tortoise
- Updated genomes for chicken, cow and horse
- New protein structure variation view

[More release news](#) on our

Other news from our blog

- 08 Mar 2019: [Joint REST server for Ensembl and Ensembl Genomes in Ensembl 96](#)
- 07 Mar 2019: [Removal of database patches script from Ensembl repository in Ensembl 96](#)
- 01 Mar 2019: [Getting to know us: Guy from Ensembl Plants](#)

News

Compare genes across species



Find SNPs and other variants for my gene

GGTATACATTC
CCTBAAAGTCTT
CTTCTAATTCT
GACACATTTCC

Gene expression in different tissues



Retrieve gene sequence

GGCTACTTCGGGTTG
GGGCTGTGCGGAGCT
GGGCTCTGCTGGGCTT
AGGGACAGATTGTG
GACCTGTGAGAGGCTTT
CCACATGCA

Find a Data Display



Use my own data in Ensembl



Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.

Accès aux archives
d'Ensembl



Le site web Ensembl: les génomes

The screenshot shows the Ensembl website interface. At the top, there is a navigation bar with links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located in the top right corner. Below the navigation bar, the main content area is divided into several sections:

- Search Human (Homo sapiens):** A search bar with a dropdown menu for "Search all categories" and a search button. A callout box labeled "Recherche" points to the search bar.
- Genome assembly: GRCh38.p12 (GCA_000001405.27):** A section with a list of links: "More information and statistics", "Download DNA sequence (FASTA)", "Convert your data to GRCh38 coordinates", and "Display your data in Ensembl". A callout box labeled "Informations, statistiques" points to this section.
- Gene annotation:** A section with a list of links: "More about this genebuild", "Download FASTA files for genes, cDNAs, ncRNA, proteins", "Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins", and "Update your old Ensembl IDs". A callout box labeled "Lien vers des exemples" points to this section.
- Comparative genomics:** A section with a list of links: "More about comparative analysis" and "Download alignments (EMF)".
- Regulation:** A section with a list of links: "More about the Ensembl regulatory build and microarray annotation", "Experimental data sources", and "Download all regulatory features (GFF)".
- Variation:** A section with a list of links: "More about variation in Ensembl", "Download all variants (GVF)", and "Variant Effect Predictor".

Each section includes an "Example" link and a small image representing the data type. The "Example region" link is located at the bottom right of the Genome assembly section. The "Example gene" link is located at the bottom right of the Gene annotation section. The "Example variant" link is located at the bottom right of the Variation section. The "Example phenotype" link is located at the bottom right of the Variation section. The "Example structural variant" link is located at the bottom right of the Variation section.

Le site web Ensembl: statistiques des génomes

Human assembly and gene annotation


Assembly

This site provides a data set based on the December 2013 *Homo sapiens* high coverage assembly GRCh38 from the [Genome Reference Consortium](#). This assembly is used by Ensembl in their hg38 database. The data set consists of gene models built from the pairwise alignments of the human proteome as well as from the cDNA2genome model of exonerate.

This release of the assembly has the following characteristics:

- contig length total 3.4 Gb.
- chromosome length total 3.1 Gb

It also includes 261 alt loci scaffolds, [region on chromosome 6](#) (7 alternate sequence representations).

 [Watch a video on YouTube](#) about patches and haplotypes in the Human genome.

Patches

As the GRC maintains and improves the assembly, patches are being introduced. Currently, assembly patches are of two types:

- Novel patch: new sequences that add alternative sequence at a loci and will remain as haplotypes in the next major assembly release by GRC
- Fix patch: sequences that correct the reference sequence and will replace the given region of the reference assembly at the next major assembly release by GRC.

The genome assembly represented here corresponds to GenBank Assembly ID [GCA_000001405.27](#)

Other assemblies

Gene annotation

The Ensembl human gene annotations have been updated using Ensembl's automatic annotation pipeline. The updated annotation incorporates new protein and cDNA sequences which have become publicly available since the last GRCh38 genebuild (December 2013).

In the current release, we continue to display a joint gene set based on the merge between the automatic annotation from Ensembl and the manually curated annotation from Havana. See the statistics table, right, for the corresponding GENCODE version number. The Consensus Coding Sequence (CCDS) identifiers have also been mapped to the annotations. More information about the [CCDS project](#).

Updated manual annotation from Havana is merged into the Ensembl annotation every release. Transcripts from the two annotation sources are merged if they share the same internal exon-intron boundaries (i.e. have identical splicing pattern) with slight differences in the terminal exons allowed. Importantly, all Havana transcripts are included in the final Ensembl/Havana merged (GENCODE) gene set.

- [Detailed information on genebuild](#) (PDF)

Neanderthal genome

A preliminary assembly of the Neanderthal (*Homo sapiens neanderthalensis*) genome is available via the [Neanderthal Genome Browser](#), an Ensembl-powered project based at the Max Planck Institute.

More information

General information about this species can be found in [Wikipedia](#).

Informations générales sur l'assemblage

Statistics

Summary

Assembly	GRCh38.p12 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.27 , Dec 2013
Base Pairs	3,609,003,417
Golden Path Length	3,096,649,726
Annotation provider	Ensembl
Annotation method	Full genebuild
Genebuild started	Jan 2014
Genebuild released	Jul 2014
Genebuild last updated/patched	Jul 2018
Database version	95.38
Gencode version	GENCODE 29

Gene counts (Primary assembly)

Coding genes	20,418 (incl 650 readthrough)
Non coding genes	22,107
Small non coding genes	4,871
Long non coding genes	15,014 (incl 284 readthrough)
Misc non coding genes	2,222
Pseudogenes	15,195 (incl 8 readthrough)
Gene transcripts	206,762

Gene counts (Alternative sequence)

Coding genes	2,958 (incl 46 readthrough)
Non coding genes	1,429
Small non coding genes	278
Long non coding genes	974 (incl 39 readthrough)
Misc non coding genes	177
Pseudogenes	1,754
Gene transcripts	20,652

Other

Genscan gene predictions	51,153
Short Variants	665,695,433
Structural variants	6,013,111

Statistiques

Le site web Ensembl: caryotype

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog Login/Register

Human (GRCh38.p12) ▾ Search all species...

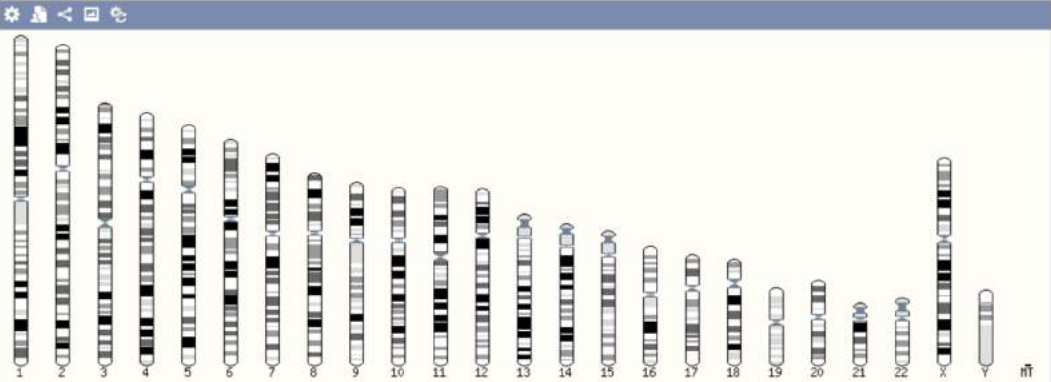
Genome

- Location-based displays
 - Whole genome
 - Chromosome summary
 - Region overview
 - Region in detail
- Comparative Genomics
 - Synteny
 - Alignments (image)
 - Alignments (text)
 - Region Comparison
- Genetic Variation
 - Variant table
 - Resequencing
 - Strain table
 - Linkage Data
- Markers
- Other genome browsers
 - UCSC
 - NCBI
 - Ensembl GRCh37

Configure this page | Custom tracks | Export data | Share this page | Bookmark this page

Whole genome

+ Add features



Click on the image above to jump to a chromosome, or click and drag to select a region

Summary

Assembly	GRCh38.p12 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.27 , Dec 2013
Base Pairs	3,609,003,417
Golden Path Length	3,096,649,726
Annotation provider	Ensembl
Annotation method	Full genebuild
Genebuild started	Jan 2014
Genebuild released	Jul 2014
Genebuild last updated/patched	Jul 2018
Database version	95.38
Gencode version	GENCODE 29

Gene counts (Primary assembly)

Coding genes	20,418 (incl 650 readthrough)
Non coding genes	22,107
Small non coding genes	4,871
Long non coding genes	15,014 (incl 284 readthrough)
Misc non coding genes	2,222
Pseudogenes	15,195 (incl 8 readthrough)
Gene transcripts	206,762

Le site web Ensembl : statistiques par chromosome

The screenshot displays the Ensembl genome browser interface for Human (GRCh38.p12). The main content area shows 'Chromosome 1: 1-1' with a detailed view of the chromosome structure, including assembly exceptions and a chromosome summary. The chromosome summary includes a bar chart showing the distribution of various genomic features across the chromosome, such as Protein Coding Genes, Short Non Coding Genes, Long Non Coding Genes, Pseudogenes, GC Repeats, and Variations. The chromosome statistics table at the bottom provides key metrics for chromosome 1.

Chromosome 1: 1-1

Assembly exceptions, chromosome 1

Assembly exceptions

Chromosome summary

+ Add features

Chromosome 1 Protein Coding Genes Short Non Coding Genes Long Non Coding Genes Pseudogenes GC Repeats Variations

Click on the image above to zoom into that point

Change chromosome: 1 Go

Chromosome Statistics	
Length (bps)	248,956,422
Coding genes	2,050 (incl. 49 readthrough)
...	...

Le site web Ensembl : navigateur de génome

The screenshot displays the Ensembl genome browser interface. At the top, the Ensembl logo is visible along with navigation links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is present with the text "Search all species...". The main header indicates the species is "Human (GRCh38.p12)".

The interface is divided into several sections:

- Location-based displays:** A sidebar on the left lists various display options such as "Whole genome", "Chromosome summary", "Region overview", "Region in detail", "Comparative Genomics", "Genetic Variation", and "Other genome browsers". A gear icon next to "Configure this page" is circled in red.
- Chromosome 17: 63,992,802-64,038,237:** A horizontal track showing the chromosome with a red box highlighting a specific region.
- Region in detail:** A detailed view of the highlighted region, showing various tracks including "Chromosome bands", "Contigs", "Genes", "Regulatory Build", and "Gene Legend". A gear icon next to the "Region in detail" title is also circled in red. The "Gene Legend" includes categories like "Ensembl protein coding", "processed transcript", "RNA gene", "CTCF", "Open Chromatin", "Promoter Flank", "merged Ensembl/Havana", "pseudogene", "Enhancer", "Promoter", and "Transcription Factor Binding Site".
- Location:** A field showing the coordinates "17:63992802-64038237" and a "Go" button.
- 75 way GERP elements:** A track showing constrained elements for 75 eutherian mammals, with a "Constrained elements for 75 eutherian mammals EPO-Low-Coverage" track.
- Human cDNAs:** A track showing human cDNAs from RefSeq/ENA and CCDS_set, with specific entries like "CCDS54158.1" and "CCDS54157.1" highlighted.

Le site web Ensembl : le gène

[eEnsembl](#)
[BLAST/BLAT](#)
[VEP](#)
[Tools](#)
[BioMart](#)
[Downloads](#)
[Help & Docs](#)
[Blog](#)
Login/Register

Human (GRCh38.p12) ▾
Search all species...

Location: 13:32,315,474-32,400,266 Gene: BRCA2

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Gene alleles
- Sequence
- Secondary Structure
- Comparative Genomics
- Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogues
- Ensembl protein families
- Ontologies
- GO: Biological process
- GO: Cellular component
- GO: Molecular function
- Phenotypes
- Genetic Variation
- Variant table
- Variant image
- Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
- Gene history

Gene: BRCA2 ENSG00000139618

Description BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]

Gene Synonyms BRCC2, FACD, FAD, FAD1, FANCD, FANCD1, XRCC11

Location [Chromosome 13: 32,315,474-32,400,266](#) forward strand.
GRCh38:CM000675.2

About this gene This gene has 7 transcripts ([splice variants](#)), [132 orthologues](#), is a member of [1 Ensembl protein family](#) and is associated with [128 phenotypes](#).

Transcripts [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq	Flags
BRCA2-201	ENST00000380152.7	11986	3418aa	Protein coding	CCDS9344.1	P51587	-	TSL:5 GENCODE basic APPRIS P1
BRCA2-206	ENST00000544455.5	10984	3418aa	Protein coding	CCDS9344.1	P51587	NM_000059 NP_000050	TSL:1 GENCODE basic APPRIS P1
BRCA2-202	ENST00000470094.1	842	186aa	Nonsense mediated decay	-	H0YE37	-	CDS 5' incomplete TSL:5
BRCA2-203	ENST00000528762.1	495	64aa	Nonsense mediated decay	-	H0YD86	-	CDS 5' incomplete TSL:4
BRCA2-207	ENST00000614259.1	7950	No protein	Processed transcript	-	-	-	TSL:2
BRCA2-204	ENST00000530893.6	2011	No protein	Processed transcript	-	-	-	TSL:1
BRCA2-205	ENST00000533776.1	523	No protein	Retained intron	-	-	-	TSL:3

Summary

Name [BRCA2](#) (HGNC Symbol)

CCDS This gene is a member of the Human CCDS set: [CCDS9344.1](#)

UniProtKB This gene has proteins that correspond to the following UniProtKB identifiers: [P51587](#)

RefSeq Overlapping RefSeq Gene ID [675](#) matches and has similar biotype of protein_coding

LRG [LRG_293](#) provides a stable genomic reference framework for describing sequence variants for this gene

Ensembl version ENSG00000139618.14

Other assemblies This gene maps to [32,889,611-32,974,403](#) in GRCh37 coordinates.
View this locus in the GRCh37 archive: [ENSG00000139618](#)

Gene type Protein coding

Annotation method Annotation for this gene includes both automatic annotation from Ensembl and [Havana](#) manual curation, see [article](#).

Annotation Attributes overlapping locus [Definitions](#)

[Go to Region in Detail for more tracks and navigation options \(e.g. zooming\)](#)

Le site web Ensembl : le transcript

Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

Human (GRCh38.p12) Location: 13:32,315,474-32,400,266 Gene: BRCA2 Transcript: BRCA2-201

Transcript-based displays

- Summary
- Sequence
 - Exons
 - cDNA
 - Protein
- Protein Information
 - Protein summary
 - Domains & features
 - Variants
 - 3D Protein model
- Genetic Variation
 - Variant table
 - Variant image
 - Haplotypes
 - Population comparison
 - Comparison image
- External References
 - General identifiers
 - Oligo probes
- Supporting evidence
- ID History
 - Transcript history
 - Protein history

Configure this page | Custom tracks | Export data | Share this page | Bookmark this page

Transcript: BRCA2-201 ENST00000380152.7

Description BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]

Gene Synonyms BRCC2, FACD, FAD, FAD1, FANCD, FANCD1, XRCC11

Location [Chromosome 13: 32,315,474-32,400,266](#) forward strand.

About this transcript This transcript has [27 exons](#), is annotated with [51 domains and features](#), is associated with [29608 variations](#) and maps to [1004 oligo probes](#).

Gene This transcript is a product of gene [ENSG00000138618](#) [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq	Flags
BRCA2-201	ENST00000380152.7	11986	3418aa	Protein coding	CCDS9344	P51587	-	TSL:5 GENCODE basic APPRIS P1
BRCA2-206	ENST00000544455.5	10984	3418aa	Protein coding	CCDS9344	P51587	NM_000059 NP_000050	TSL:1 GENCODE basic APPRIS P1
BRCA2-202	ENST00000470094.1	842	186aa	Nonsense mediated decay	-	H0YE97	-	CDS 5' incomplete TSL:5
BRCA2-203	ENST00000528762.1	495	64aa	Nonsense mediated decay	-	H0YD86	-	CDS 5' incomplete TSL:4
BRCA2-207	ENST00000614259.1	7950	No protein	Processed transcript	-	-	-	TSL:2
BRCA2-204	ENST00000530893.6	2011	No protein	Processed transcript	-	-	-	TSL:1
BRCA2-205	ENST00000533776.1	523	No protein	Retained intron	-	-	-	TSL:3

Summary

Statistics Exons: 27, Coding exons: 26, Transcript length: 11,986 bps, Translation length: 3,418 residues

CCDS This transcript is a member of the Human CCDS set: [CCDS9344](#)

Uniprot This transcript corresponds to the following Uniprot identifiers: [P51587](#)

Transcript Support Level (TSL) [TSL:5](#)

Version ENST00000380152.7

Type Protein coding

Annotation Method Transcript where the Ensembl genbuild transcript and the [Vega](#) manual annotation have the same sequence, for every base pair. See [article](#).

GENCODE basic gene This transcript is a member of the [Gencode basic](#) gene set.

Ensembl release 95 - January 2019 © [EMBL-EBI](#) [Permanent link](#) - [View in archive site](#)

Naviguer dans Ensembl : Partie pratique

Visualiser ses propres données

Tools [All tools](#)

BioMart > Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT > Search our genomes for your DNA or protein sequence

Variant Effect Predictor > Analyse your own variants and predict the functional consequences of known and unknown variants

Search

All species for

Go

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes

-- Select a species --

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

Favourite genomes

- Human** GRCh38.p12
[Still using GRCh37?](#)
- Mouse** GRCm38.p6
- Zebrafish** GRCz11

Compare genes across species

Find SNPs and other variants for my gene

Gene expression in different tissues

Retrieve gene sequence

Find a Data Display

Use my own data in Ensembl

EMBL-EBI Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available. Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.

Visualiser ses propres données

LES OUTILS

Les outils

BLAST/BLAT

data-mining tool

Search from Ensembl with this

BLAST/BLAT >

Search our genomes for your DNA or protein sequence

Variation Effect Predictor >

Analyse your own variants and predict the functional consequences of known and unknown variants

Search

All species for

Go

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes

-- Select a species --

- View full list of all Ensembl species
- Edit your favourites

Favourite genomes



Human
GRCh38.p12

Still using GRCh37?



Mouse
GRCm38.p6



Zebrafish
GRCz11

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 95 (January 2019)

- New regulatory build for human, incorporating new data from ENCODE
- Update to GENCODE M20 for mouse
- New genomes: donkey, polar bear, black bear, red fox, koala, dingo, tuatara, painted turtle and desert tortoise
- Updated genomes for chicken, cow and horse
- New protein structure variation view

[More release news](#) on our blog

Other news from our blog

- 08 Mar 2019: [Joint REST server for Ensembl and Ensembl Genomes in Ensembl 95](#)
- 07 Mar 2019: [Removal of database patches script from Ensembl repository in Ensembl 96 r](#)
- 01 Mar 2019: [Getting to know us: Guy from Ensembl Plants](#)

Compare genes across species



Find SNPs and other variants for my gene

```
GTATATACATTC  
CCTRAAAGTCTT  
CTTCTAATTCT  
GAAACATTTTC
```

Gene expression in different tissues



Retrieve gene sequence

```
GCCTACTCCGGGTTG  
GGACTTGTGCGCGAGC  
GGGCTCTGCTGCGGCT  
AGGGACAGATTTGTGA  
GACTCTGAGAGGTTT  
CCAGTCCAGCTTGGC
```

Find a Data Display



Use my own data in Ensembl



Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available. Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.



Blast



- Recherche de similarité
 - 1 séquence (*Query*) comparée à des milliers ou des millions de séquences (*base de données*) par comparaison 2 à 2.
- But:
 - Détecter des séquences proches
 - Annotation simple (domaines protéiques, localisation génomique, nombre d'exons)

Les différentes comparaisons

BLAST : Basic Local Alignment Search Tool

Altschul *et al.* Basic local alignment search tool. *J. Mol. Biol.* 1990

Altschul *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997

Programmes	Requête	Banque	Comparaison	Exemples d'utilisation
Blastn	ADN	ADN	nucléique	Recherche d'ARN structuraux, d'éléments régulateurs
Blastp	Protéine	protéines	protéique	Recherche de protéines homologues
Tblastn	Protéine	ADN (traduit dans les 6 cadres)	protéique	Recherche de similarités entre une protéine et une séquence génomique mal annotée
Blastx	ADN (traduit dans les 6 cadres)	protéines	protéique	Recherche des phases de lecture dans une séquence codante
Tblastx	ADN (traduit dans les 6 cadres)	ADN (traduit dans les 6 cadres)	protéique	Avantages de tblastn et blastx mais très long

Les différentes comparaisons

BLAT (BLAST-Like Alignment Tool)

- An mRNA/DNA and cross-species protein sequence analysis tool to quickly find sequences of $\geq 95\%$ similarity of length ≥ 40 bases.
- was developed by Jim Kent at the University of California Santa Cruz (UCSC) in the early 2000s to assist in the assembly and annotation of the human genome.
- The target database of BLAT is not a set of GenBank sequences, but instead an index derived from the assembly of the entire genome. **Blat works by keeping an index of an entire genome in memory.**
- By default, the index consists of all non-overlapping 11-mers for DNA and 4-mers for protein.
- Kent, W.J.. BLAT -- The BLAST-Like Alignment Tool. *Genome Research* 2002

MADTQYILPNDIGVSSLDCREAFRLLSPTERLYAYHLSRAAWYGGLAVLLQTSPEAPYIYALLSRLFRAQDPDQ
LRQHALAEGLTEEEYQAFLVYAAGVYSNMGNYSFGDTKFVPNLPKEKLERVILGSEAAQQHPPEVRGLWQTCG
ELMFSLEPRLRHLGLGKEGITTYFSGNCTMEDAKLAQDFLDSQNL SAYNTRLFKEVDGEGKPYEVR LASVLGS
EPSLDSEVTSKLSYEFRGSPFQVTRGDYAPILQKVVEQLEKAKAYAANSHQGQMLAQYIESFTQGSIEAHKRG
SRFWIQDKGPIVESYIGFIESYRDPFGSRGEFEGFVAVVNKAMSAKFERLVASAEQLLKELPWPPTFEKDKFLT
PDFTSLDVLT FAGSGI PAGINI PNYDDL RQTEGFKNVSLGNVLAVAYATQREKLT FLEEDDKDLYILWKGPSFD
VQVGLHELLGHGSGKLFVQDEKGA FNFDQETVINPETGEQIQSWYRSGETWDSKFSTIASSYEECRAESVGLYL
CLHPQVLEIFGFEGADAEDVIYVNWLN MVRAGLLALEFYTPEAFNWRQAHMQARFVILRVLLEAGEGLVTITPT
TGSDGRPDARVRLDRSKIRSVGK PALERFLRRLQVLKSTGDVAGGRALYEGYATVTDAPPECFLT LRDTVLLRK
ESRKLIVQPNTRLEGS DVQLLEYEASAAGLIRSFSEFPEDGPELEEILTQLATADARFWKGPSEAPSGQA

new SETUP CONFIG RESULTS DISPLAY refresh Online Help

new SETUP CONFIG RESULTS DISPLAY refresh Online Help

new SETUP CONFIG RESULTS DISPLAY refresh Online Help

Retrieve result for ID:

BLA_IESYdDXDJ Retrieve

Alignment Display Options:

Locations vs. Karyotype Locations vs. Query

Summary Table

1: unnamed (737 letters) Vs. LATESTGP

Homo_sapiens 1961 alignments, 23 hits [RawResult] view ▶

refresh Online Help

Summary

▶ setup

- Homo_sapiens
- Genomic sequence
- TBLASTN
- Low sensitivity

▶ configure

- -E: 10
- -B: 100
- -filter: seg
- -W: 4
- -hitdist: 40
- -matrix: BLOSUM80
- -T: 16

▶ results

▶ display

⊙ Not yet initialised

We would like to hear your impressions or feedback, especially regarding functionality that you would like to see provided in the future. Many thanks for your time. [Feedback](#)

Content-type: text/plain

TBLASTN 2.OMP-WashU [04-May-2006] [linux26-x64-I32LPF64 2006-05-10T17:22:28]

Copyright (C) 1996-2006 Washington University, Saint Louis, Missouri USA. All Rights Reserved.

Reference: Gish, W. (1996-2006) http://blast.wustl.edu

Query= unnamed (737 letters)

WARNING: Precomputed values for Lambda, K and H are unavailable for the BLOSUM80 scoring matrix, when used with gap penalties +9 and +2. Unless overridden on the command line, the values computed for ungapped alignments will be used instead, but the reported E-values and P-values may be much too low.

Database: Homo_sapiens.GRCh37.dna.toplevel.fa 297 sequences; 32,036,512,383 total letters.

WARNING: Use of the hspsepSmax parameter should be considered with long database sequences, to improve the biological relevance of the HSP groups that are assembled and to improve the statistical discrimination of these groups from random background.

Searching...10...20...30...40...50...60...70...80...90...100% done

WARNING: hspmax=1000 was exceeded by 37 of the database sequences, causing the associated cutoff score, S2, to be transiently set as high as 73.

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Probability P(N)	Smallest Sum N
9 dna:chromosome chromosome:GRCh37:9:1:141213431:1 REF	-3	1765	0.	6
11 dna:chromosome chromosome:GRCh37:11:1:135006516:1 REF	+3	763	3.2e-292	9
4 dna:chromosome chromosome:GRCh37:4:1:191154276:1 REF	+3	1542	5.5e-250	4
20 dna:chromosome chromosome:GRCh37:20:1:63025520:1 REF	-1	131	0.0035	9
16 dna:chromosome chromosome:GRCh37:16:1:90354753:1 REF	+1	120	0.014	10
12 dna:chromosome chromosome:GRCh37:12:1:133851895:1 REF	-2	126	0.060	11
19 dna:chromosome chromosome:GRCh37:19:1:59128983:1 REF	-1	128	0.069	9
22 dna:chromosome chromosome:GRCh37:22:1:51304566:1 REF	+1	130	0.10	10
GL000199.1 dna:supercontig supercontig:GRCh37:GL000199.1:... +3		149	0.11	2
14 dna:chromosome chromosome:GRCh37:14:1:107349540:1 REF	+2	167	0.21	8
1 dna:chromosome chromosome:GRCh37:1:1:249250621:1 REF	-1	134	0.25	8
GL000220.1 dna:supercontig supercontig:GRCh37:GL000220.1:... -3		124	0.26	4
5 dna:chromosome chromosome:GRCh37:5:1:180915260:1 REF	+1	127	0.33	9
GL000224.1 dna:supercontig supercontig:GRCh37:GL000224.1:... -2		126	0.49	2
7 dna:chromosome chromosome:GRCh37:7:1:159138663:1 REF	-3	129	0.88	9
21 dna:chromosome chromosome:GRCh37:21:1:48129895:1 REF	-2	131	0.98	9
GL000237.1 dna:supercontig supercontig:GRCh37:GL000237.1:... -2		89	0.98	5
GL000202.1 dna:supercontig supercontig:GRCh37:GL000202.1:... +1		111	0.995	3
GL000218.1 dna:supercontig supercontig:GRCh37:GL000218.1:... -1		145	0.996	5
15 dna:chromosome chromosome:GRCh37:15:1:102531392:1 REF	+2	134	0.999	12
6 dna:chromosome chromosome:GRCh37:6:1:171115067:1 REF	-2	118	0.9991	13
3 dna:chromosome chromosome:GRCh37:3:1:198022430:1 REF	-3	118	0.9998	11
GL000206.1 dna:supercontig supercontig:GRCh37:GL000206.1:... -3		92	0.99992	6

>9 dna:chromosome chromosome:GRCh37:9:1:141213431:1 REF Length = 141,213,431

Score = 1765 (578.9 bits), Expect = 0., Sum P(6) = 0. Identities = 220/261 (84%), Positives = 230/261 (88%), Frame = -3

Query: 477 INPETGEIQISWYRSGETWDSKFSTIASSYEECRAESVGLYLCLHPQVLEIFGFEGADAF 536
INPE EQIQSWYRS +TWDSKFSTI SSYEECRAESVGLYLCLHPQVLE FGFEGADAE
Sbjct: 76090065 INPEMREIQISWYRSMKTWDSKFSTIVSSYEECRAESVGLYLCLHPQVLETFGFEGADAE 76089886

Query: 537 DVIYVNWLNLMVRAGLLALEFYTPPEAFNWRQAHMQARFVILRVLLEAGEGLVITITPTGSD 596
+VI VNWLNLMV AGLLALEFYTPPEA NW+QAH++AR VILRVL EAGEGL TITPT GSD
Sbjct: 76089885 EVISVNWLNLMVGAGLLALEFYTPPEASNWQQAHIRARIVILRVLPEAGEGLGITITPTAGSD 76089706

Query: 597 GRPDAVRVLRDRSKIRSVGKPALERFLRRLOVLKSTGVDVAGGRALYEGYATVTDAPPECFI 656
GRP+A+VRLDRSKI+SVG PALERFLRR STGVDVAGG LYE YA V DAPPE FI
Sbjct: 76089705 GRPEAQVRLDRSKIQSVGNPALERFLRRCW---STGVDVAGGWILYERYAAVADAPPEGFI 76089535

Query: 657 TLRDVTLLRKESRKLIVQPNTRLEGSVDVQLLEYEASAAGLIRSFSEFPEDGPELEEIILT 716
TLRD VLLRKES KLIVQPN RLEGSVDVQLLEYE SAAGLIRSFSE FPEDG ELE+ILT
Sbjct: 76089534 TLRDRVLLRKESWKLIVQPNIRLEGSVDVQLLEYEVSAAAGLIRSFSEHFPEDELELDILT 76089355

Query: 717 QLATADARFWKGPSEAPSGQA 737
QLATADA+F KGPSEAPSGQA
Sbjct: 76089354 QLATADAQF*KGPSEAPSGQA 76089292

Score = 1700 (557.6 bits), Expect = 0., Sum P(6) = 0. Identities = 212/252 (84%), Positives = 221/252 (87%), Frame = -2

Query: 224 PSLDSEVTSKLSYEFGRGSPFQVIRGDYAPILQKVVVEQLEKAKAYAANSHQGMQLAQYIE 283
P L + SKLKS EFRGSPFQVT G+Y PILQKVVVEQLEKAK YAANSHQ QMLAQYIE
Sbjct: 76090816 PGLRGD--SKLKS*EFRGSPFQVTWGNYPILQKVVVEQLEKAKTYAANSHQEQMLAQYIE 76090643

Query: 284 SFTQGSIEAHRKGRSFRWIQDKGPIVESYIGFIESYRDPFGSRGFEFEGFVAVVNKAMSAKF 343
SFTQGS EAHK+GSRFWI DKGPIVESYI FI+SYRD FGSRG EGFVAVVNKAMSAKF
Sbjct: 76090642 SFTQGSTEAHKKGRSFRWI*DKGPIVESYIEFIQSYRDSFGSRGCEGFVAVVNKAMSAKF 76090463

Query: 344 ERLVASAEQLLKELEWPPFTFEKDKFLTPDFTSLDVLTFAGSGIPAGINIPNYDDLKQTEG 403
E LV SAEQLLKELEW P FEKDKFLTPDFTS+DVLTFAGSGI AGINI NY+DL+QTEG
Sbjct: 76090462 EWLVSVAEQLLKELEWSPAFKDKFLTPDFTSVDVLTFAGSGIAAGINISNYNDLQKTEG 76090283

Query: 404 FKNVSLGNVLAVAYATQREKLTFLLEDDKDLIYLWKGPSFDVQVGLHELLGHGSGKLFVQ 463
FKNVSLGNVLAV ATQ EKLT LEE DKDLYI+ GPSFDVQVGLHELLG+GSGKL Q
Sbjct: 76090282 FKNVSLGNVLAVV*ATQWEKLTIVLEESDKDLYIVLWKGPSFDVQVGLHELLGYGSGKLIQ 76090103

Query: REF 464 DEKGFANFDQET 475
DEKGFANFDQET
Sbjct: 76090102 DEKGFANFDQET 76090067

[new](#) [SETUP](#) [CONFIG](#) **[RESULTS](#)** [DISPLAY](#)

Retrieve result for ID:

Alignment Display Options:

Locations vs. Karyotype Locations vs. Query
 Summary Table

1: unnamed (737 letters) Vs. LATESTGP

Homo_sapiens 1961 alignments, 23 hits [\[RawResult\]](#)

[refresh](#) [Online Help](#)

Summary

▶ setup

- [Homo_sapiens](#)
- [Genomic sequence](#)
- [TBLASTN](#)
- [Low sensitivity](#)

▶ configure

- [-E: 10](#)
- [-B: 100](#)
- [-filter: seg](#)
- [-W: 4](#)
- [-hitdist: 40](#)
- [-matrix: BLOSUM80](#)
- [-T: 16](#)

▶ results

▶ display

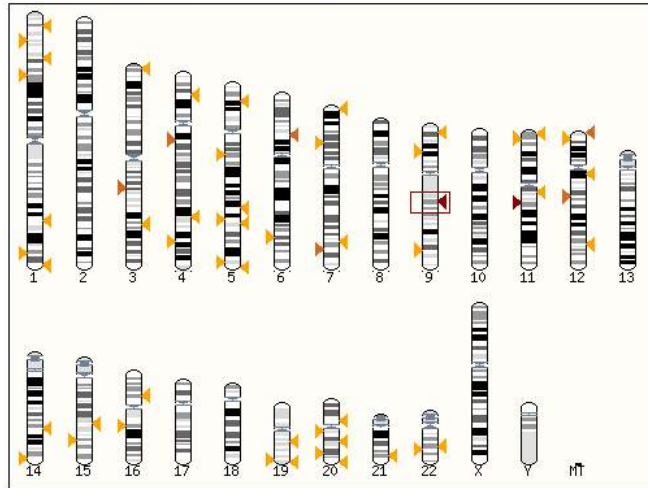
ⓘ *Not yet initialised*

Displaying unnamed sequence alignments vs Homo_sapiens LATESTGP database

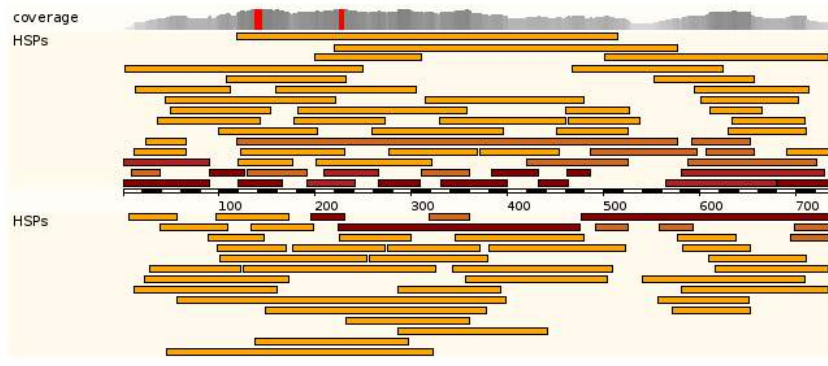
Showing top 100 alignments of 1961, sorted by Raw Score

refresh

Alignment Locations vs. Karyotype (click arrow to hide)



Alignment Locations vs. Query (click arrow to hide)



Summary

setup

- Homo_sapiens
- Genomic sequence
- TBLASTN
- Low sensitivity

configure

- -E: 10
- -B: 100
- -filter: seg
- -W: 4
- -hitdist: 40
- -matrix: BLOSUM80
- -T: 16

results

display

Not yet initialised

Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort

(Use the 'ctrl' key to select multiples)

refresh

Query	Subject	Chromosome	Supercontig	Clone	Contig	Lrg	Stats	Sort By
off	_off_	_off_	_off_	_off_	_off_	_off_	_off_	>Lrg
Name	Name	Name	Name	Name	Name	Name	Score	<Score
Start	Start	Start	Start	Start	Start	Start	E-val	>Score

Links	Query	Chromosome	Stats
	Start End Ori	Name Start End Ori	Score E-val %ID Length
[A] [S] [G] [C]	477 737 +	Chr:9 76089292 76090065 -	1765 0. 84.29 261
[A] [S] [G] [C]	224 475 +	Chr:9 76090067 76090816 -	1700 0. 84.13 252
[A] [S] [G] [C]	119 577 +	Chr:4 65296878 65298248 +	1542 5.5e-250 49.70 497
[A] [S] [G] [C]	581 729 +	Chr:4 65298493 65298930 +	854 5.5e-250 74.83 151
[A] [S] [G] [C]	1 90 +	Chr:11 66249672 66249941 +	763 3.2e-292 100.00 90
[A] [S] [G] [C]	330 399 +	Chr:11 66260186 66260395 +	552 3.2e-292 95.71 70
[A] [S] [G] [C]	565 679 +	Chr:11 66264763 66265104 +	531 3.2e-292 63.71 124
[A] [S] [G] [C]	1 90 +	Chr:4 65296627 65296899 +	529 5.5e-250 76.09 92
[A] [S] [G] [C]	588 721 +	Chr:11 66271972 66272364 +	487 1.7e-276 55.63 142
[A] [S] [G] [C]	681 737 +	Chr:11 66276549 66276719 +	477 3.2e-292 100.00 57
[A] [S] [G] [C]	120 166 +	Chr:11 66254008 66254148 +	391 1.8e-273 97.87 47
[A] [S] [G] [C]	420 526 +	Chr:11 66262674 66262961 +	384 3.2e-292 53.57 112
[A] [S] [G] [C]	486 597 +	Chr:11 66263006 66263296 +	377 1.7e-276 51.72 116
[A] [S] [G] [C]	266 309 +	Chr:11 66258962 66259093 +	375 3.2e-292 97.73 44
[A] [S] [G] [C]	209 266 +	Chr:11 66258657 66258854 +	370 3.2e-292 75.76 66
[A] [S] [G] [C]	384 432 +	Chr:11 66260513 66260650 +	310 5.1e-263 83.67 49
[A] [S] [G] [C]	90 126 +	Chr:11 66252641 66252751 +	272 3.2e-292 89.19 37
[A] [S] [G] [C]	432 463 +	Chr:11 66261009 66261104 +	270 1.7e-276 96.88 32
[A] [S] [G] [C]	192 242 +	Chr:11 66255385 66255576 +	268 1.3e-266 64.06 64
[A] [S] [G] [C]	196 230 +	Chr:9 76090801 76090905 -	257 0. 88.57 35
[A] [S] [G] [C]	129 191 +	Chr:11 66254628 66254813 +	248 3.2e-292 56.06 66

[A] [S] [G] [C]

477 737 +

Chr:9

76089292 76090065 -

1765 0.

84.29 261

[A] [S] [G] [C]

[A]align

Query location : unnamed 477 to 737 (+)
 Database location : 9 76089292 to 76090065 (-)
 Genomic location : 9 76089292 to 76090065 (-)

Alignment score : 1765
 E-value : 0.
 Alignment length : 261
 Percentage identity: 84.29

Query: 477 INFETGEQIQSWYRSGETWDSKFSIIASSYEECRASVGLYLCLHPQVLEIFGFEADG 536
 INFPE EQIQSWYRS +TWDSKFSII SSYEECRASVGLYLCLHPQVLE FGFEGADG
 Sbjct: 76090065 INFEMREIQIQSWYRSKMTWDSKFSIIVSYEECRASVGLYLCLHPQVLETFGFEGDAE 76099886

Query: 597 DVIIYVNWLMVVRAGLLALEFYTFEAFNRQAHMQARFVILRVLLEAGEGLVITPTGSD 596
 +VI VNWLMGV AGLLALEFYTFEA NW+QAH++AR VILRVL EAGEGL IPTFT GSD
 Sbjct: 76089888 EVISVNWLMVVRAGLLALEFYTFEASNWQAHIRARIVILRVLFEAGEGLTFPTAGSD 76089706

Query: 597 GRFDARVLRDSKIRSVGKPALERFLRLQVLRKSTGVAGGRALVEGYVITDAPPECF 656
 GRP+a+VRLRSKI+VUG PALERFLR STGVAGG LYE Y V DAPPE FL
 Sbjct: 76089708 GRFEAQVLRDSKISVGNFALERFLRCW---STGVAGGWLYEYEAADVAPPEFL 76089935

Query: 657 TLRDVLRLKESRKLIVQPNTRELEGSVQLLEYEASAAGLIRSFERFPEDGPELEEILT 716
 TLRD VLLRKS KLIWQPN RLEGSVQLLEYE SAAGLIRSE FPDG ELE+ILT
 Sbjct: 76089934 TLRDVRLLRKSRLIVQPNTRELEGSVQLLEYEVSAAGLIRFSEHFPELLELEILT 76089935

Query: 717 QLATADARFWKGFSEAPSGQA 737
 QLATADA+ F KGFSEAPSGQA
 Sbjct: 76089934 QLATADAQF+KGFSEAPSGQA 76089292

[S]equenc

Query location : unnamed 477 to 737 (+)
 Database location : 9 76089292 to 76090065 (-)
 Genomic location : 9 76089292 to 76090065 (-)

Alignment score : 1765
 E-value : 0.
 Alignment length : 261
 Percentage identity: 84.29

THIS STYLE: Matching bases for selected HSP
 THIS STYLE: Matching bases for other HSPs in selected hit

>unnamed
 MADTIQYILPNDIGVSSLDCREAFRLLSPTERLYAYHLSRAAWYGLLAVLLQTSPEAPYIY
 ALLSRLFRAPDPQLRQHAALEGLTEEEYQAFVLVYAAAGVYSNMGNYKSFQDKTFVNLFPK
 EKLERVILGSEAAQQHPEEVRLVQLTCGELMFLSLEPRLRHLGLGKEGITTYSFGNCTMED
 AKLAQDFLDSQNLASYNRLRKFVEDGKGFYEVRLASVLGSEPSLDSEVTSKLSYEFGR
 GSPFQVIRGDYAPILQKVVEQLKAKAYAAANSHQGQMLAQYIESFTQGSIEAHKRGSRFW
 IQDKGPIVESYIGFIESYRDPFGSRGFEFVAVVNKAMSAKFERLVASAEQLLKELPWP
 PTFEKDRFLTPDFSLDVLTFAGSGIPAGINIPNYDDLRQTEGFKNVSLGNVLAVAYATQ
 REKLTFLIEDDDKLDLYLLWKPFLVQVGLHELLGHGSGKLFVQDEKGFANFDQETVINPE
 TGEQIQSWYRSGETWDSKFSIIASSYEECRASVGLYLCLHPQVLEIFGFEADAEVVIY
 VNWLMVVRAGLLALEFYTFEAFNRQAHMQARFVILRVLLEAGEGLVITPTFTGSDGRPD
 ARVLRDSKIRSVGKPALERFLRLQVLRKSTGVAGGRALVEGYVITDAPPECFITLRD
 TVLLRKSRLIVQPNTRELEGSVQLLEYEASAAGLIRSFERFPEDGPELEEILTQLAT
 ADARFWKGFSEAPSGQA

[G]Sequence

[C]ontig view

Chromosome 9: 76,087,292-76,092,065

Region in detail

Chromosome bands
 Contigs
 Genes (GENCODE)
 Gene Legend
 Location: 9:76087292-76092065
 Genes
 Chromosome bands
 ST max: GENCODE
 Contig
 BLAT/LAST hits
 Genes (GENCODE)
 Genes (ENSEMBL)
 Gene Legend
 There are currently 374 tracks turned off

Query location : unnamed 477 to 737 (+)
 Database location : 9 76089292 to 76090065 (-)
 Genomic location : 9 76089292 to 76090065 (-)

Alignment score : 1765
 E-value : 0.
 Alignment length : 261
 Percentage identity: 84.29

5' Flanking sequence : 300 (bp)
 3' Flanking sequence : 300 (bp)
 Coordinate system : Chromosome
 Orientation : Forward relative to selected alignment
 Alignment markup : All alignments Both orientations
 Feature markup : Ensembl exons Both orientations
 Line numbering : No numbers

update

THIS STYLE: Location of selected alignment
 THIS STYLE: Location of other alignments
 THIS STYLE: Location of Exons
 >chr9:76089292-76092065:76090065:1
 TGGATGTTCTTACCTTCGCTGGCTCGGATCGCTGGCGGATGAACATCTCCAATFACA
 ATGACTTAAACACACAGAAAGCTTTAAGAACCTTCACTGAGAACCTTCCTGCTGCTG
 TCTAAGCCACCGAGTGGGAGAGCTCACCTTCTGGAGGAGTGCACAGGACTGTGACA
 TCGCTGTATGGGGCCCTCTTTCGATGACAGTGGCTGCACAGACTGCTGGGCTATG
 GCAGCGCAAGCTCATCGAGCAGGATGAAGAGGAGCATTCACTTTGACCAGGAGACCC
 ATCAATCCAGAGTGAAGAGCAGGATTCAGAGCTGGATCGAGCATGAAGACTGGGAC
 AGCAAGTTCACACCAATTTCTTCCACTGACAAAGATGCTGGCGGCTGAGAGCTTGGGCTC
 TACTCTGTCTCCACCGCAGTGGCTGGAGACTTTGGCTTTGAGGGGCTGATCGAGAG
 GAGGTGATCTCTGACCTGCTCAACATGCTTGGGCTGGGCTGGGCTGCTGGGATTC
 TACACACTCGAGGCTCCAACTGGCAGCGGCCATATACGGGCGGATTTGTGATCTCG
 AGAGTCTTACAGAGGCTGGCGAGGACTGGTACACTCACTCCACCGCAGGCTCGGAT
 GGCCCGCCAGGCGAGTCCCTTCCACCGCAGCAAGATCCAGTCTGGGCGCACCT
 GGCTAGAGGCTTCTGGGAGGCTGGCTCCAGAGGGGATTTGGCGGAGGCTGGGAGC
 CTGTAGAGAGGATTCAGCAGTCTGATGGCGCCCTGAGGCTTCTCACCTCCAGG
 GACAGGCTGCTGCTGGATAGGAACTTTCGAGAGCTGATGCTTCAGCGCAACTTGGCTT
 GAGGCTGACAGTGGGCTGCGGAAATACAGAGTCTGAGCTTGGCTCATGAGAGC
 TTCTTGAGCAATTCGAGAGGATGAGGCTGAGTTCGAGGACTCTTACACAGCTGAGA
 ACAGCGGATCCCAACTTCTGAAAGGCCCCAGTGGGCGGCTTGGCGAGCTTGTGGA
 AGAGTGTGGCTTCTCCCACTGCATCAACCAAGGCTCAGTGGGCTGCAITC
 GTGTGTATTAGGGGCTGGGAGGGGGAGGGGAGGGGAGGAGGCTTGGACTTGGTAC
 TACTCAGCTGAGGTTGGTACACTAACCTTCACTTGTGAGCACTTCCAGGCTGGC
 AAGTCTCCCTCTGTGACTTATTATATGACTGCCATATGTGASGTGACCAAGAC
 CTGTCCACCTCCGCTACCGATGAGGAATGGAGCTTGTGAGAAATGAG

LES OUTILS : BLAST

PARTIE PRATIQUE

Les outils

Annotation
de variants

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog Login/Register

Tools [All tools](#)

BioMart > Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT > Search our genomes for your DNA or protein sequence

Variant Effect Predictor > Analyse your own variants and predict the functional consequences of known and unknown variants

Search




All species for

e.g. BRCA2 or rat 5:62797363-63627669 or rs599 or coronary heart disease

All genomes

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

Favourite genomes

-  **Human**
GRCh38.p12
[Still using GRCh37?](#)
-  **Mouse**
GRCm38.p6
-  **Zebrafish**
GRCz11

Compare genes across species 

Find SNPs and other variants for my gene 


Gene expression in different tissues 

Retrieve gene sequence

```
GCCTGACTTCGGGTTGG
GGGCTGTGCGCCGAGCC
GGGCTGTGCTGGGCTT
AGGGACAGATTCTGTG
CACCTGTGGAGCGGTTI
CCCACTGCAGCTGGCC
```

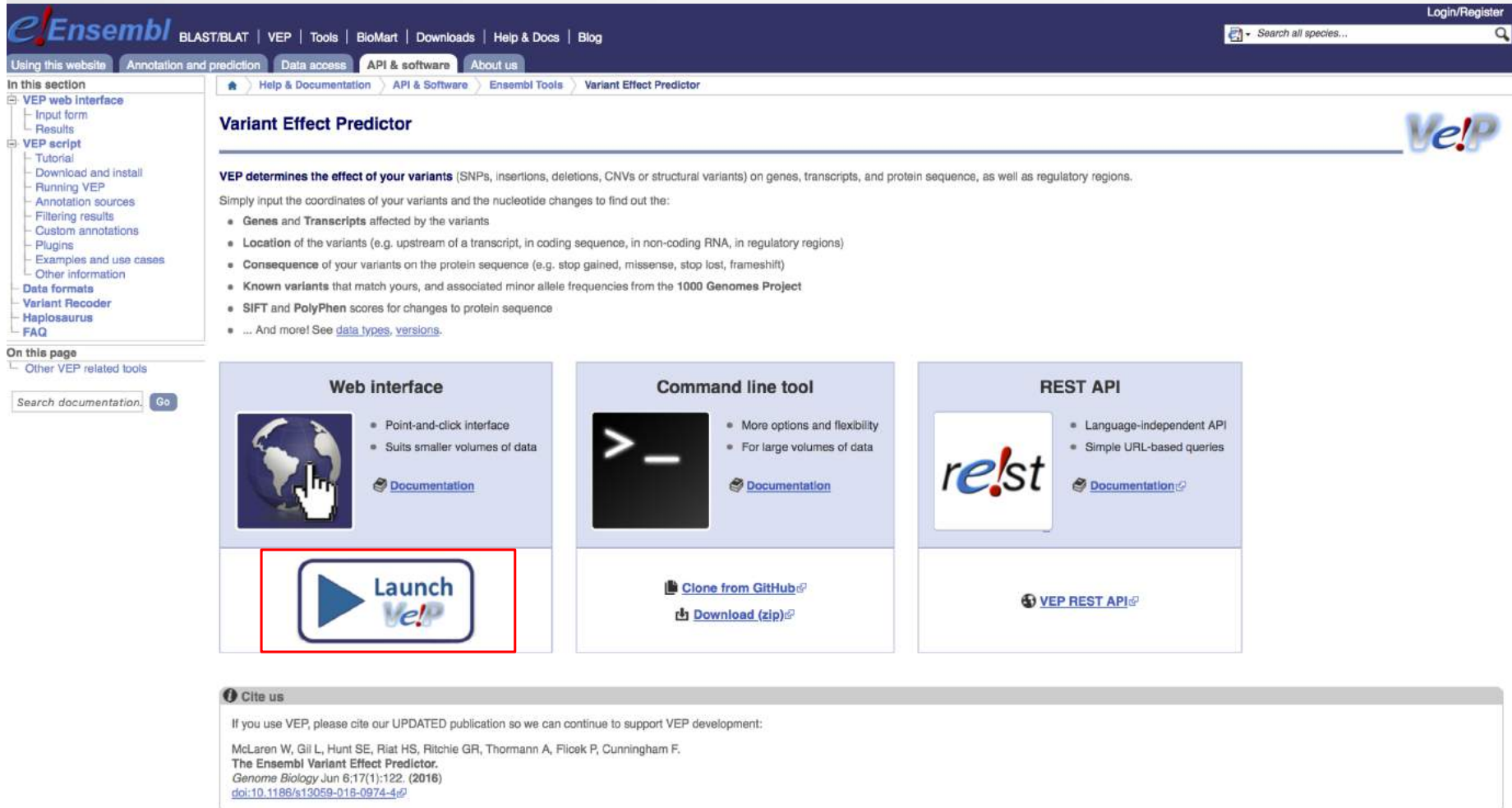
Find a Data Display 

Use my own data in Ensembl 

EMBL-EBI  Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available. Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.



Variant Effect Predictor



e/Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog Login/Register

Using this website: Annotation and prediction | Data access | **API & software** | About us Search all species...

Help & Documentation | API & Software | Ensembl Tools | Variant Effect Predictor

Variant Effect Predictor

VEP determines the effect of your variants (SNPs, insertions, deletions, CNVs or structural variants) on genes, transcripts, and protein sequence, as well as regulatory regions.

Simply input the coordinates of your variants and the nucleotide changes to find out the:

- **Genes and Transcripts** affected by the variants
- **Location** of the variants (e.g. upstream of a transcript, in coding sequence, in non-coding RNA, in regulatory regions)
- **Consequence** of your variants on the protein sequence (e.g. stop gained, missense, stop lost, frameshift)
- **Known variants** that match yours, and associated minor allele frequencies from the **1000 Genomes Project**
- **SIFT and PolyPhen** scores for changes to protein sequence
- ... And more! See [data types](#), [versions](#).

Web interface

- Point-and-click interface
- Suits smaller volumes of data

[Documentation](#)

Launch VEP

Command line tool

- More options and flexibility
- For large volumes of data

[Documentation](#)

[Clone from GitHub](#)

[Download \(zip\)](#)

REST API

- Language-independent API
- Simple URL-based queries

[Documentation](#)

[VEP REST API](#)

Cite us

If you use VEP, please cite our UPDATED publication so we can continue to support VEP development:

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F.
The Ensembl Variant Effect Predictor.
Genome Biology Jun 6;17(1):122. (2016)
[doi:10.1186/s13059-016-0974-4](https://doi.org/10.1186/s13059-016-0974-4)

Variant Effect Predictor

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog Login/Register

Search all species...

Variant Effect Predictor

[New job](#) [Clear form](#)

Species:
Assembly: GRCh38.p12 (If you are looking for VEP for Human GRCh37, please go to [GRCh37 website](#).)

Name for this job (optional):

Input data:

Either paste data:

Examples: [Ensembl default](#), [VCF](#), [Variant identifiers](#), [HGVS notations](#)

Or upload file:

Or provide file URL:

Transcript database to use:

- Ensembl/GENCODE transcripts
- Ensembl/GENCODE basic transcripts
- RefSeq transcripts
- Ensembl/GENCODE and RefSeq transcripts

Additional configurations:

- Identifiers** Additional identifiers for genes, transcripts and variants
- Variants and frequency data** Co-located variants and frequency data
- Additional annotations** Additional transcript, protein and regulatory annotations
- Predictions** Variant predictions, e.g. SIFT, PolyPhen
- Filtering options** Pre-filter results by frequency or consequence type

Recent jobs

You have no jobs currently running or recently completed.

Ensembl release 95 - January 2019 © [EMBL-EBI](#)

Variant Effect Predictor

Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog Login/Register

Search all species...

VEP ▾

Web Tools

- Web Tools
 - BLAST/BLAT
 - Variant Effect Predictor
 - VEP analysis of pasted data
 - Linkage Disequilibrium Calculator
 - File Chameleon
 - Assembly Converter
 - ID History Converter
 - VCF to PED Converter
 - Data Slicer

Configure this page
Custom tracks
Export data
Share this page
Bookmark this page

Variant Effect Predictor results

Job details ▾

Summary statistics ▾

Category	Count
Variants processed	6
Variants filtered out	0
Novel / existing variants	5 (83.3) / 1 (16.7)
Overlapped genes	18
Overlapped transcripts	88
Overlapped regulatory features	55

Consequences (all)

- TF_binding_site_variant: 74%
- regulatory_region_variant: 8%
- non_coding_transcript_exon_variant: 4%
- intron_variant: 4%
- upstream_gene_variant: 3%
- feature_truncation: 2%
- frameshift_variant: 1%
- downstream_gene_variant: 1%
- NMD_transcript_variant: 1%
- Others

Coding consequences

- frameshift_variant: 67%
- coding_sequence_variant: 20%
- stop_lost: 13%

Results preview

Navigation (per variant) Download

Page: 1 of 2 | Show: 5 All variants Filters: Uploaded variant is defined Add

All: [VCF VEP TXT](#)
BioMart: [Variants](#) [Genes](#)

Show/hide columns (6 hidden)

Uploaded variant	Location	Allele	Consequence	Impact	Symbol	Gene	Feature type	Feature
1_160283_duplication	1:160282-471362	duplication	upstream_gene_variant	MODIFIER	RNU6-1100P	ENSG00000222623	Transcript	ENST00000410691
1_160283_duplication	1:160282-471362	duplication	non_coding_transcript_exon_variant, intron_variant	MODIFIER	AL732372.1	ENSG00000236601	Transcript	ENST00000412666
1_160283_duplication	1:160282-471362	duplication	non_coding_transcript_exon_variant, intron_variant	MODIFIER	AP006222.1	ENSG00000228463	Transcript	ENST00000424587
1_160283_duplication	1:160282-471362	duplication	coding_sequence_variant, 5_prime_UTR_variant, 3_prime_UTR_variant	MODIFIER	OR4F29	ENSG00000284733	Transcript	ENST00000426406
1_160283_duplication	1:160282-471362	duplication	non_coding_transcript_exon_variant, intron_variant	MODIFIER	AL732372.2	ENSG00000237094	Transcript	ENST00000431321
1_160283_duplication	1:160282-471362	duplication	non_coding_transcript_exon_variant	MODIFIER	WBP1LP7	ENSG00000269732	Transcript	ENST00000437905
1_160283_duplication	1:160282-471362	duplication	non_coding_transcript_exon_variant, intron_variant	MODIFIER	AL732372.2	ENSG00000237094	Transcript	ENST00000440163
1_160283_duplication	1:160282-471362	duplication	non_coding_transcript_exon_variant, intron_variant	MODIFIER	AP006222.1	ENSG00000228463	Transcript	ENST00000441866
1_160283_duplication	1:160282-471362	duplication	non_coding_transcript_exon_variant, intron_variant	MODIFIER	AP006222.1	ENSG00000228463	Transcript	ENST00000442116
1_160283_duplication	1:160282-471362	duplication	non_coding_transcript_exon_variant, intron_variant	MODIFIER	AP006222.1	ENSG00000228463	Transcript	ENST00000448958

Outils de récupération de données

Download a sequence or region



Click on the 'Export data' button in the lefthand menu of most pages to export:

- FASTA sequence
- GTF or GFF features

...and more!

Customise your download



Custom datasets can be retrieved using the BioMart data-mining tool.

You may find exploring this web-based query tool easier than extracting information direct from our databases.

Fetch data programmatically



Write your own Perl scripts to retrieve small-to-medium datasets. All our data, as well as added functionality, is available through the Ensembl Perl API.

Use the API to retrieve gene and transcript sets, fetch alignments between sequences, compare allele frequencies and much more!

You can also use our [REST API](#) to retrieve data to process in the programming language of your choice.

Download databases & software



All of our data and software, including pipelines and web code, is available free.

- [Download data via FTP](#)
- [Ensembl pipeline on GitHub](#)
- [Set up your own Ensembl website](#)

About Us

[About us](#)
[Contact us](#)
[Citing Ensembl](#)
[Privacy policy](#)
[Disclaimer](#)

Get help

[Using this website](#)
[Adding custom tracks](#)
[Downloading data](#)
[Video tutorials](#)
[Variant Effect Predictor \(VEP\)](#)

Our sister sites

[Ensembl Bacteria](#)
[Ensembl Fungi](#)
[Ensembl Plants](#)
[Ensembl Protists](#)
[Ensembl Metazoa](#)

Follow us

[Blog](#)
[Twitter](#)
[Facebook](#)

Outils de récupération de données

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog Login/Register

Using this website | Annotation and prediction | **Data access** | API & software | About us

In this section: [Help & Documentation](#) | [Accessing Ensembl Data](#) | [FTP Download](#)

[Downloading with rsync](#)

Search documentation:

FTP Download

You can download via a browser from our [FTP site](#), use a script, or even use [rsync](#) from the command line.

API Code

If you do not have access to git, you can obtain our latest API code as a gzipped tarball:

[Download complete API for this release](#)

Note: the API version needs to be the same as the databases you are accessing, so please use git to obtain a previous version if querying older databases.

Database dumps

Entire databases can be downloaded from our FTP site in a variety of formats. Please be aware that some of these files can run to many gigabytes of data.

Looking for [MySQL dumps](#) to install databases locally? See our [web installation instructions](#) for full details.

Each directory on [ftp.ensembl.org](#) contains a [README](#) file, explaining the directory structure.

Multi-species data

Database	MySQL	EMF	MAF	BED	XML	Ancestral Alleles
Comparative genomics	MySQL	EMF	MAF	BED	XML	Ancestral Alleles
BioMart	MySQL	-	-	-	-	-
Stable ids	MySQL	-	-	-	-	-

Single species data

Popular species are listed first. You can customise this list via our [home page](#).

Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Other annotations	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig
Human <i>Homo sapiens</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
Zebrafish <i>Danio rerio</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	GVF	VCF	VEP	-	-	BAM/BigWig
Agassiz's desert tortoise <i>Gopherus agassizii</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	-	-	VEP	-	-	BAM/BigWig
Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	-	-	VEP	-	-	-
Alpaca <i>Vicugna pacos</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF JSON	MySQL	-	-	VEP	-	-	-
Amazon molly	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	TSV RDF	MySQL	-	-	VEP	-	-	BAM/BigWig

BIOmart

Le projet BioMart

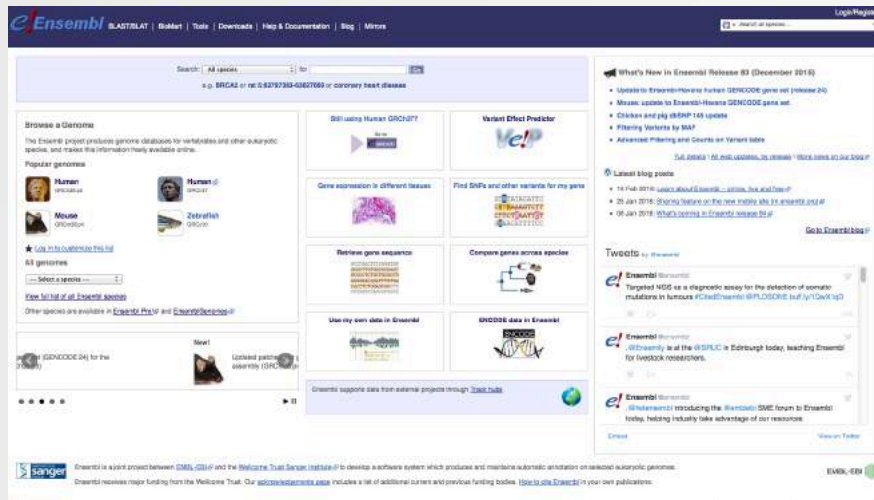
- <http://www.biomart.org/>
- Développé conjointement par :
 - EBI
 - Cold Spring Harbor Laboratory (CSHL)
- Arek Kasprzyk : « BioMart can access diverse databases from a single interface »
- Créer un système générique de stockage et de gestion de données
- « Data-agnostic » : manipulation de n'importe quel type de donnée avec le même software
- Applicable à
 - Tout type de données descriptives (y compris des données biologiques)
 - de grands volumes de données

Les “Marts”

The image displays two web interfaces. The top interface is the Ensembl BioMart tool, showing a search bar for species, a navigation menu with options like 'New', 'Count', and 'Results', and a selection area for columns to include in the output (e.g., Features, Structures, Homologs, Variation). The bottom interface is the ICGC Data Portal, featuring the ICGC logo, the text 'ICGC Data Portal', and three main navigation buttons: 'Cancer Projects', 'Advanced Search', and 'Data Repository'. A search bar at the bottom of the ICGC portal contains the text 'eg. BRAF, KRAS G12D, DO35108, MU7870, TCGA-06-5858'.

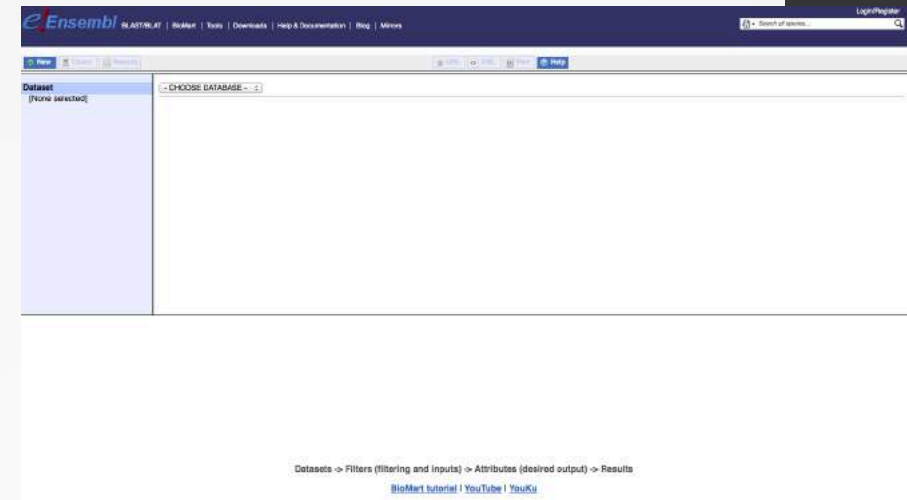
Accéder aux données d'Ensembl

Site web



- 😊 Simple d'utilisation
- 😊 Facile à comprendre
- 😞 Une seule requête à la fois

Outil de fouille: BioMart



- 😊 Requête complexe
- 😊 Rapide
- 😞 Requiert une formation

BioMart/Ensembl

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 95 (January 2019)

- New regulatory build for human, incorporating new data from ENCODE
- Update to GENCODE M20 for mouse
- New genomes: donkey, polar bear, black bear, red fox, koala, dingo, tuatara, painted turtle and desert tortoise
- Updated genomes for chicken, cow and horse
- New protein structure variation view

[More release news](#) on our blog

Other news from our blog

- 08 Mar 2019: [Joint REST server for Ensembl and Ensembl Genomes in Ensembl 96](#)
- 07 Mar 2019: [Removal of database patches script from Ensembl repository in Ensembl 96](#)
- 01 Mar 2019: [Getting to know us: Guy from Ensembl Plants](#)

- Accès à :
 - Annotation génomique (gènes, SNPs)
 - Annotation fonctionnelle
 - Expression

BioMart/Ensembl

The screenshot displays the Ensembl BioMart interface. At the top, the Ensembl logo is on the left, and navigation links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog are in the center. On the right, there is a 'Login/Register' link and a search bar labeled 'Search all species...'. Below the navigation bar, there are buttons for 'New', 'Count', and 'Results', along with output format options: 'URL', 'XML', 'Perl', and 'HTML'. The main content area is divided into two columns. The left column contains a 'Dataset' section with 'Human genes (GRCh38.p12)' selected, a 'Filters' section with '[None selected]', and an 'Attributes' section with 'Gene stable ID' and 'Transcript stable ID'. Below this is another 'Dataset' section with '[None Selected]'. The right column features two dropdown menus. The top one is set to 'Ensembl Genes 95' and the bottom one to 'Human genes (GRCh38.p12)'. Two callout boxes with arrows point to these dropdowns. The top callout box, titled 'Selection de la Base de donnée :', lists 'Genes', 'Variation', 'Regulation', and 'Mouse strain'. The bottom callout box, titled 'Sélection du jeu de données (génom)', points to the 'Human genes (GRCh38.p12)' dropdown.

Selection de la Base de donnée :

- Genes
- Variation
- Regulation
- Mouse strain

Sélection du jeu de données (génom)

Dataset

Human genes (GRCh38.p12)

Filters

[None selected]

Attributes

Gene stable ID

Transcript stable ID

Dataset

[None Selected]

Ensembl Genes 95

Human genes (GRCh38.p12)

• 4 arguments :

- Attributes (entêtes des colonnes dans les résultats)
- Filters (Utilisé pour restreindre les résultats)
- Values (identifiants utilisés pour filtrer)
- Mart (selection des jeux de données)

Biomart : Partie pratique

Comparaison des browsers

- Différences majeures entre Ensembl vs UCSC/NCBI
 - NCBI vs ensembl (UCSC?) – à l'origine de l'assemblage
 - Utilisation d'un pipeline automatique pour la création des jeux de données
 - Utilisation:
 - Visuel: ensembl/UCSC vs NCBI
 - Web: ensembl vs UCSC/NCBI
 - Rapidité/confort: UCSC vs ensembl/NBI
 - Organisation: ensembl/UCSC? Vs NCBI