

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



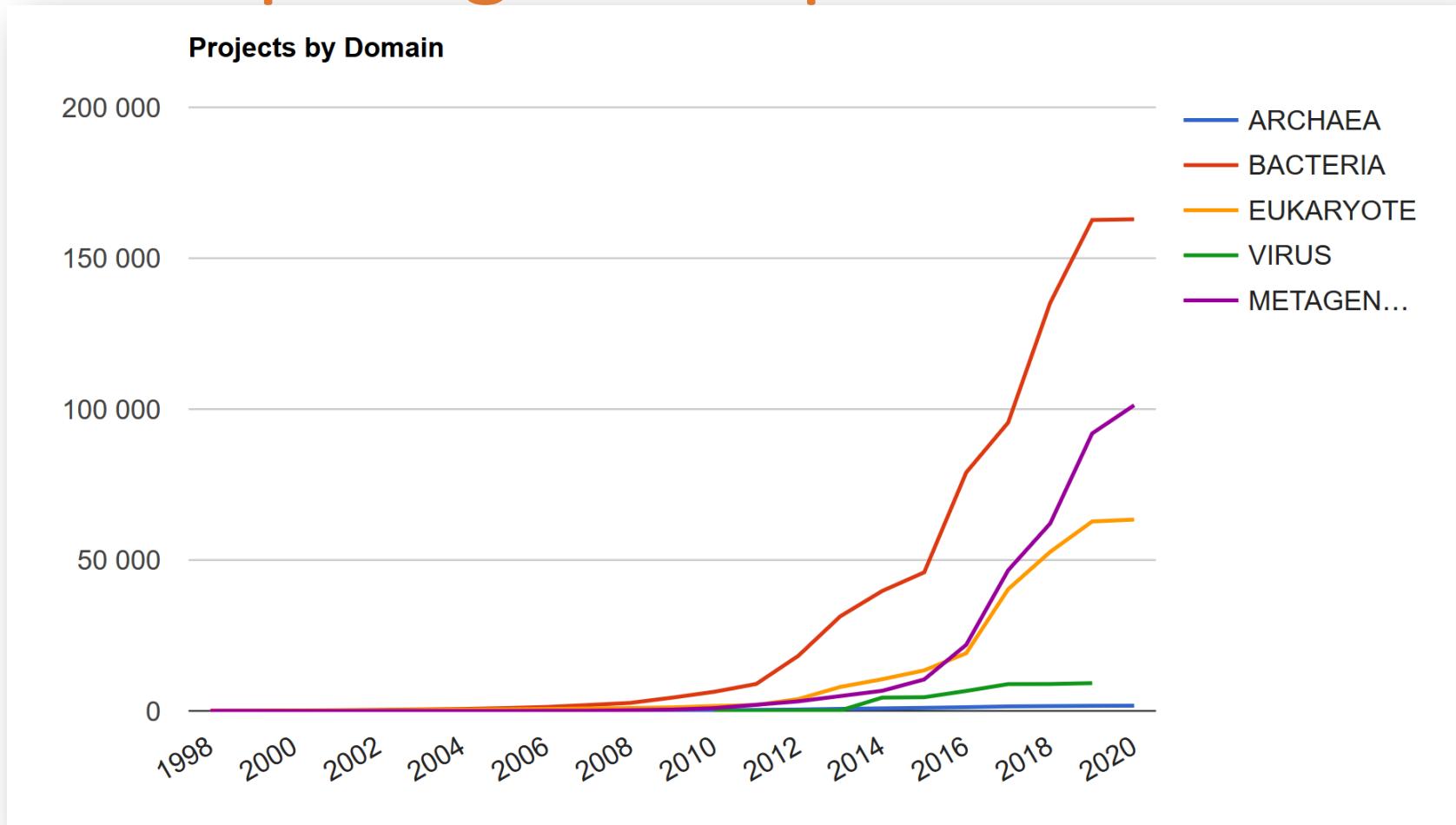
L'ère post-génomique

- 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
- 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 archées, 31736 bactéries et 9173 eukaryotes (www.genomesonline.org, 10/2016)

Exomes et génomes humains séquencés complètement (patients + pop. Générale)



L'ère post-génomique



Centres de bioinformatique

- EBI (European Bioinformatics Institute)



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

- NCBI (National Center for Biotechnology Information)

The screenshot shows the NCBI homepage with a blue header. On the left is the NCBI logo (a stylized 'S' icon followed by the letters 'NCBI'). To the right of the logo is the text 'National Center for Biotechnology Information'. Below the header, there are links for 'National Library of Medicine' and 'National Institutes of Health'. A navigation bar below the header includes links for 'PubMed', 'All Databases', 'BLAST', 'OMIM', 'Books', 'TaxBrowser', and 'Structure'. At the bottom of the page is a search bar with the placeholder 'Search All Databases' and a dropdown menu, followed by a 'Go' button.

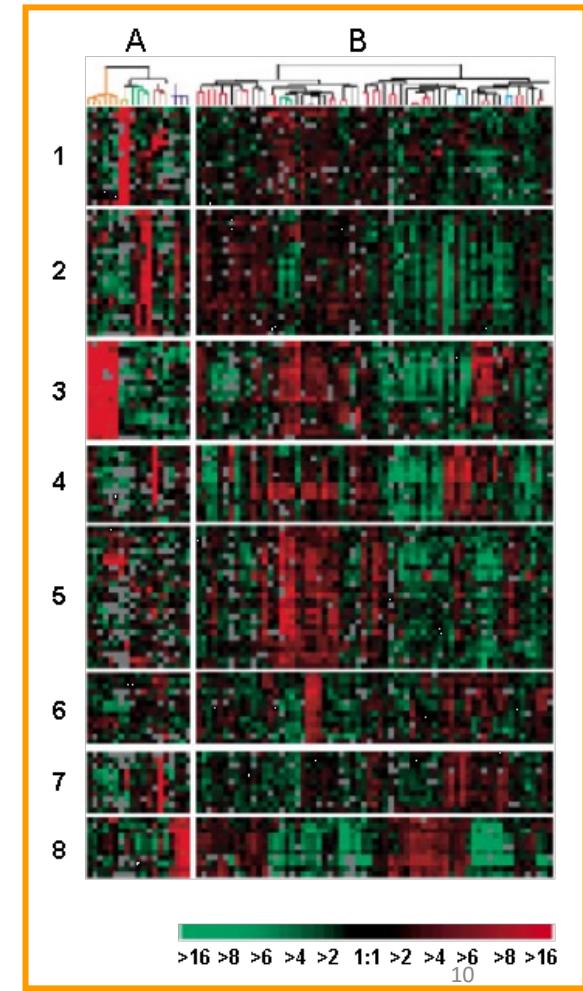
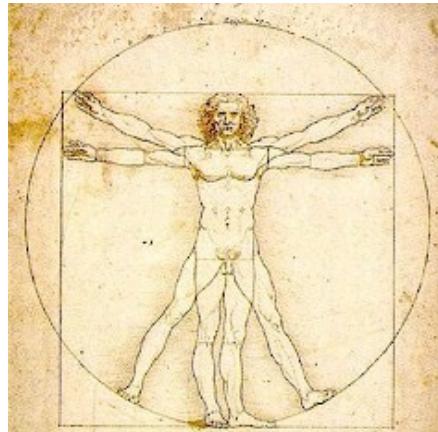
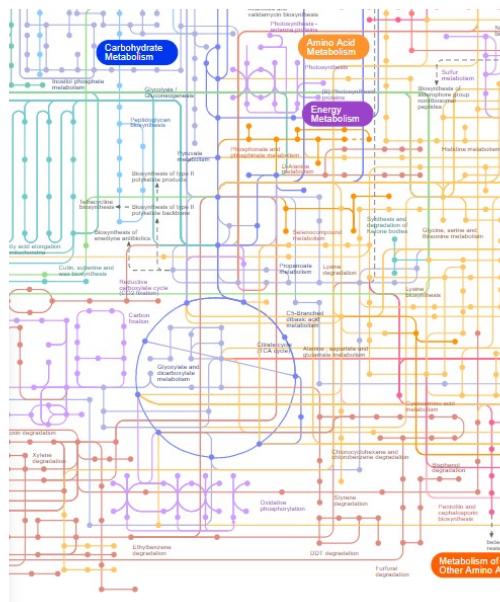
<http://www.ncbi.nlm.nih.gov/>

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

MALWTRLRPLLALLALWPPPPARAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEGPQVGCALELAGGPGA



Banques de séquences nucléiques généralistes



GenBank



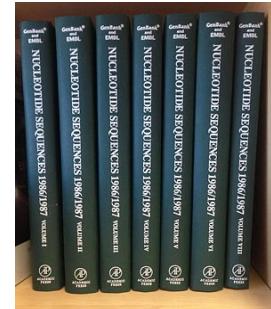
EMBL



DNA
databank of
Japan



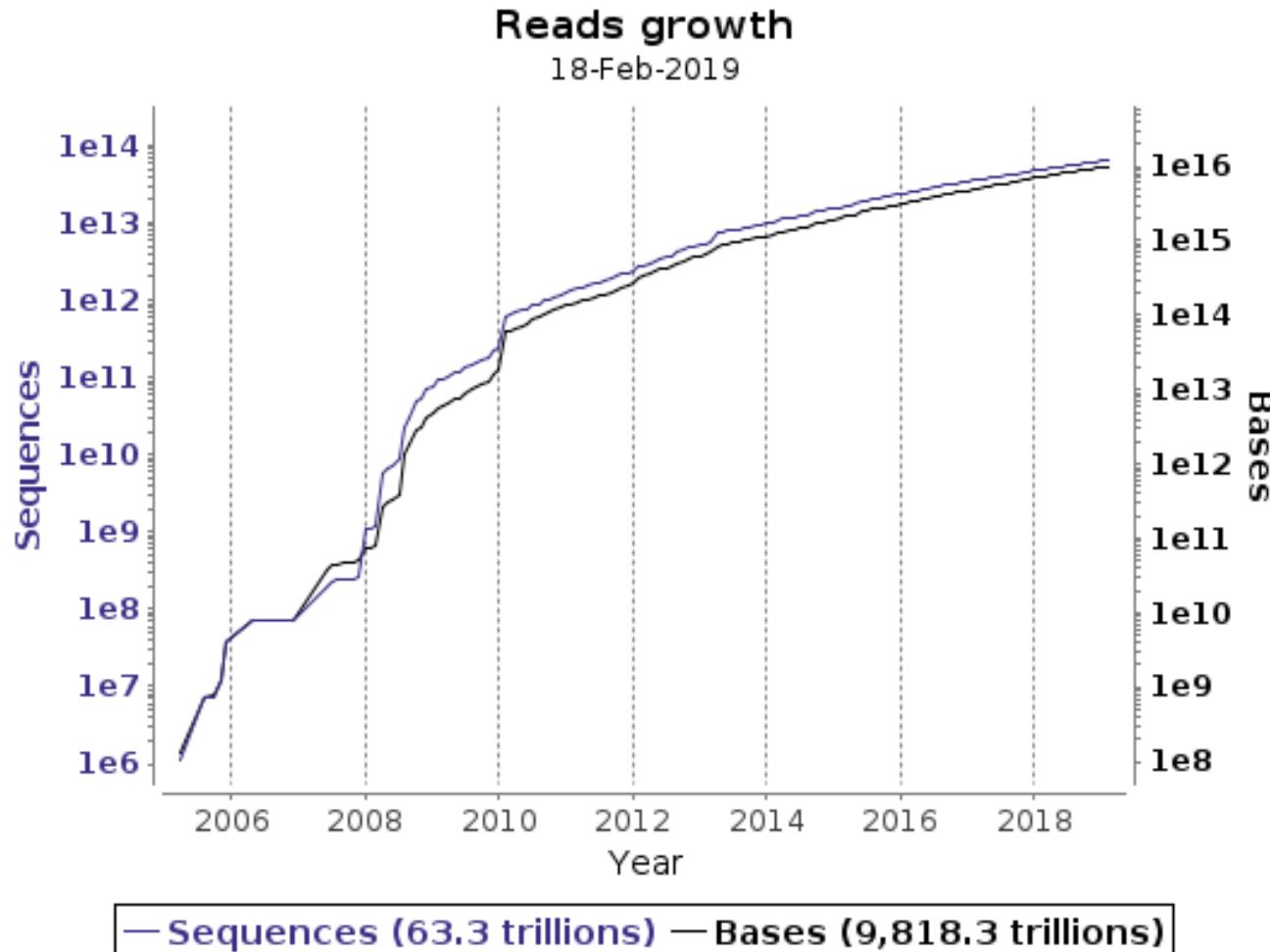
- 3 banques
- Échanges quotidiens des séquences collectées
- Effort d'unification=> format
 - accord entre GenBank et EMBL en 1986
 - accord entre GenBank/EMBL et DDBJ in 1987



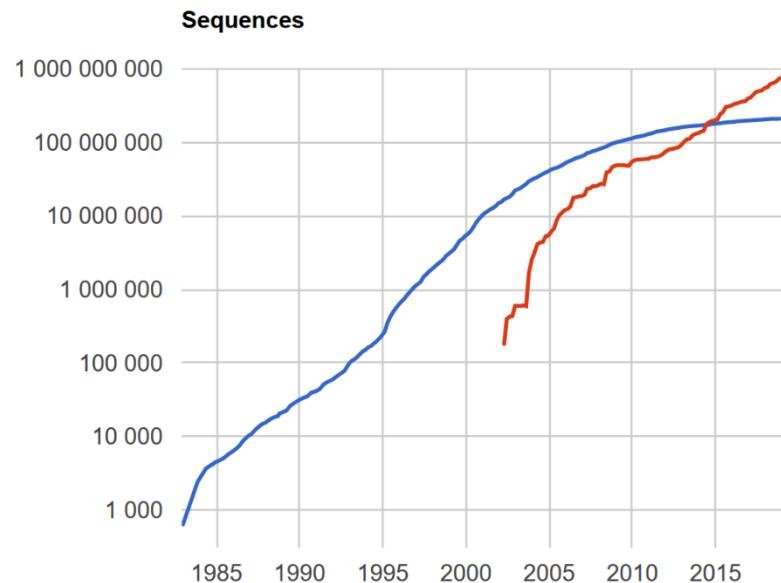
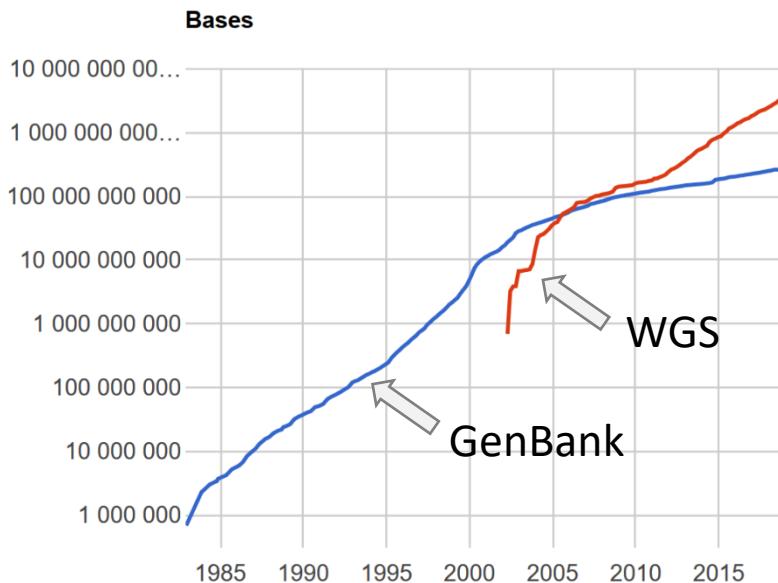
Banques de séquences

- 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



12/2018: 285 milliards de nucléotides, 211 millions d'entrées
Doublement tous les 18 mois

Banques de séquences protéiques généralistes



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

| | | |
|-------------|-------------|-------------|
| 03/2018 | 01/2019 | 02/2020 |
| 106,245,682 | 130,366,644 | 167,278,920 |

Transcrits: 29,869,155
Organismes: 99,842



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

| | | |
|------------|-------------|-------------|
| 10/2016 | 02/2018 | 02/2020 |
| 68,493,254 | 109,414,541 | 179,812,129 |



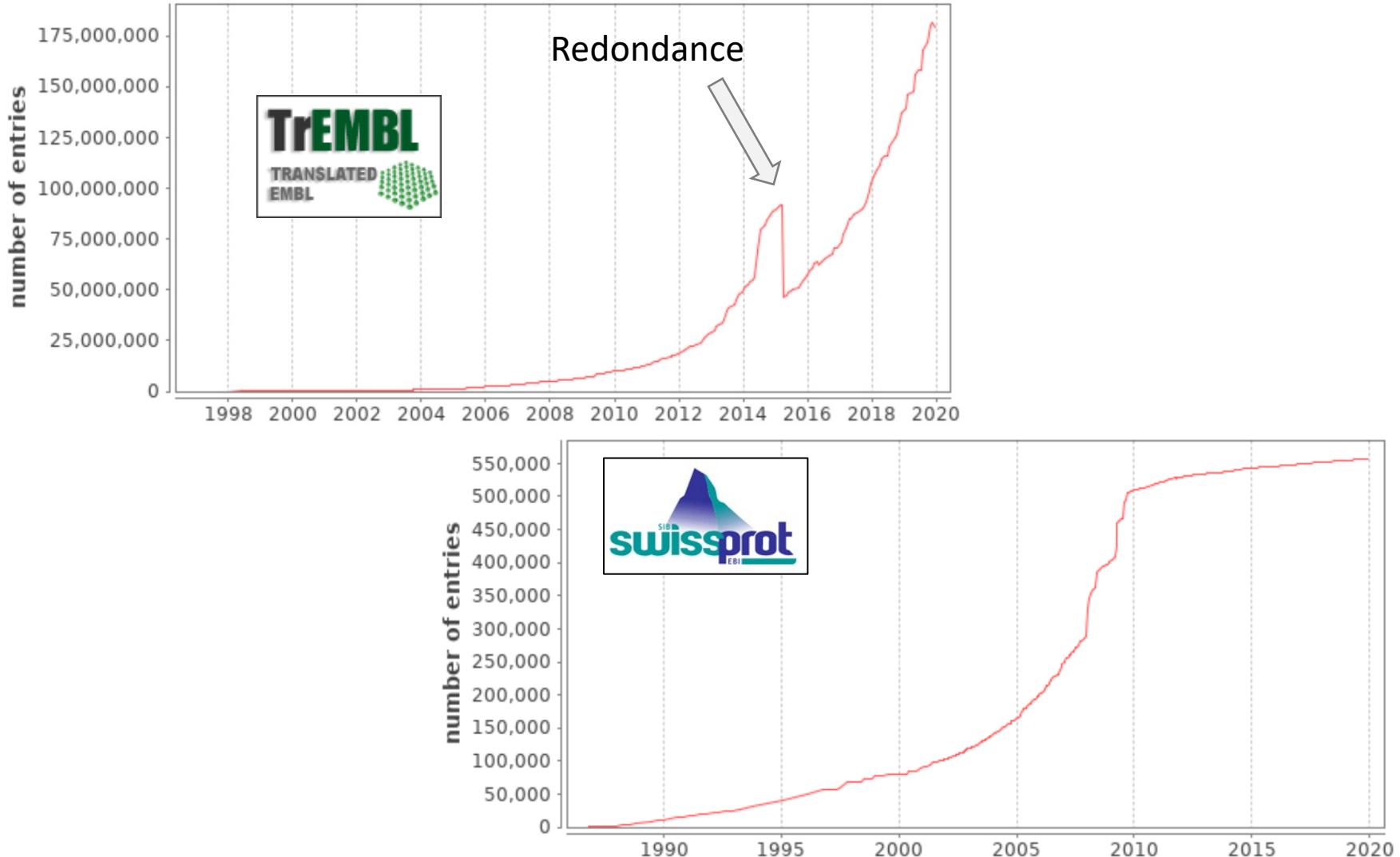
TrEMBL:
179,250,561 entrées

Swiss-Prot:
561,568 entrées

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

| 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



Hétérogénéité de la qualité en fonction de leur origine

La séquence des protéines est prédite!



La qualité des séquences de protéines dépend de la source et est donc très hétérogène

cDNA clonés et séquencés individuellement => protéine
(complets, séquençage multiple, vérification)



HTC (High-Throughput cDNA) => protéine
(full-length mais séquence brute, *indels*, *multiple codons initiateur*)



Structure 3D => protéine
(attention au *substitutions ponctuelles/délétions*)



Séquence génomique procaryote => protéine prédite
(prédiction réalisée par *outils bioinformatiques*, erreurs de codon initiateur de traduction fréquents, *indels en Nter*)



Séquence génomique eucaryote => protéine prédite
(prédiction réalisée par *outils bioinformatiques*, erreurs de prédictions de sites d'épissage fréquents, frameshifts, *indels*)



Hétérogénéité de la qualité en fonction de leur origine

1) Annotations manuelles



Réalisées par des experts, les entrées sont traitées une par une (UniProt/SwissProt)

2) Annotations automatiques



Réalisées par des outils bioinformatiques de prédiction de domaines, de fonctions...

« **by similarity** », « **homologous to** », « **related to** », « **-like** », « **putative** », « **potential** »

Sont produites en haut-débit (ex: annotation de génomes)

Elles sont légions dans les banques ... et en attente d'une validation

3) Absence d'annotations



« **hypothetical protein** »

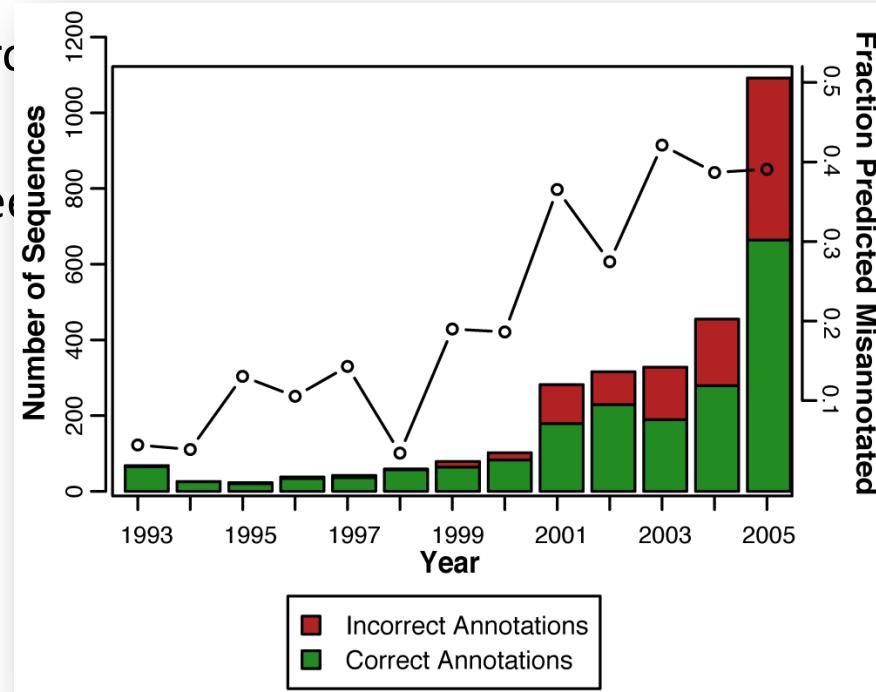
Exemple de l'importance de l'annotation

Exemple 1: DUF domain = Domain of Unknown Function

Exemple 2: FAM20C = Family with sequence similarity 20, member C

Exemple 3: Analyse de 37 familles de protéines

L'augmentation de la **quantité** de données ne signifie pas une augmentation de la **qualité** de ces données.



Evolution des bases de données protéiques

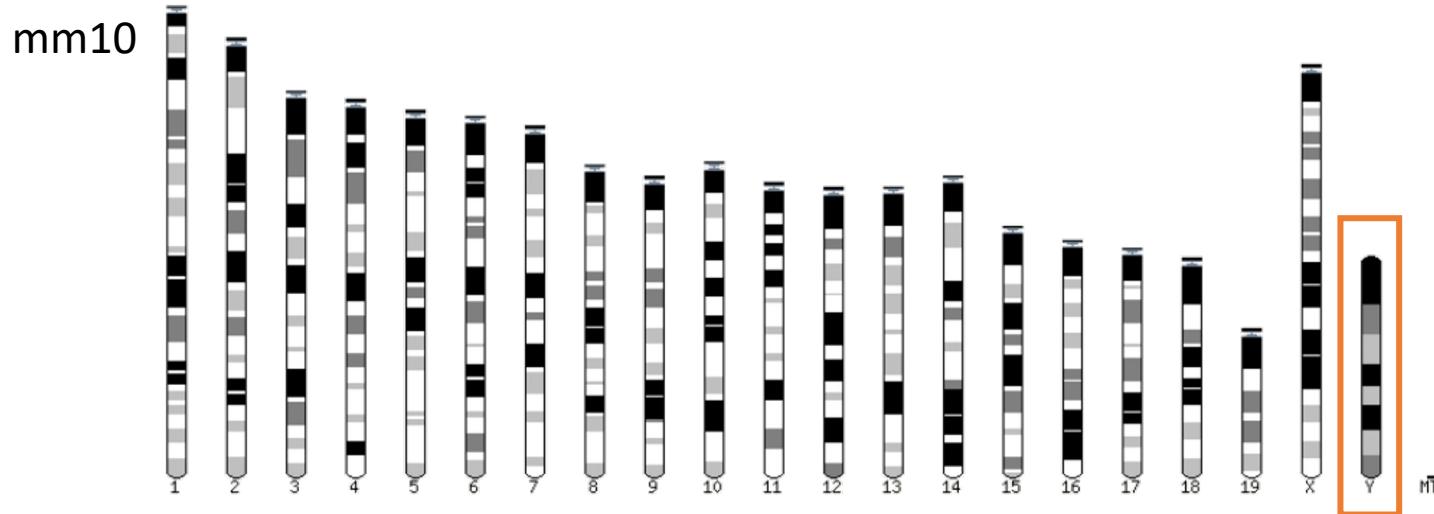
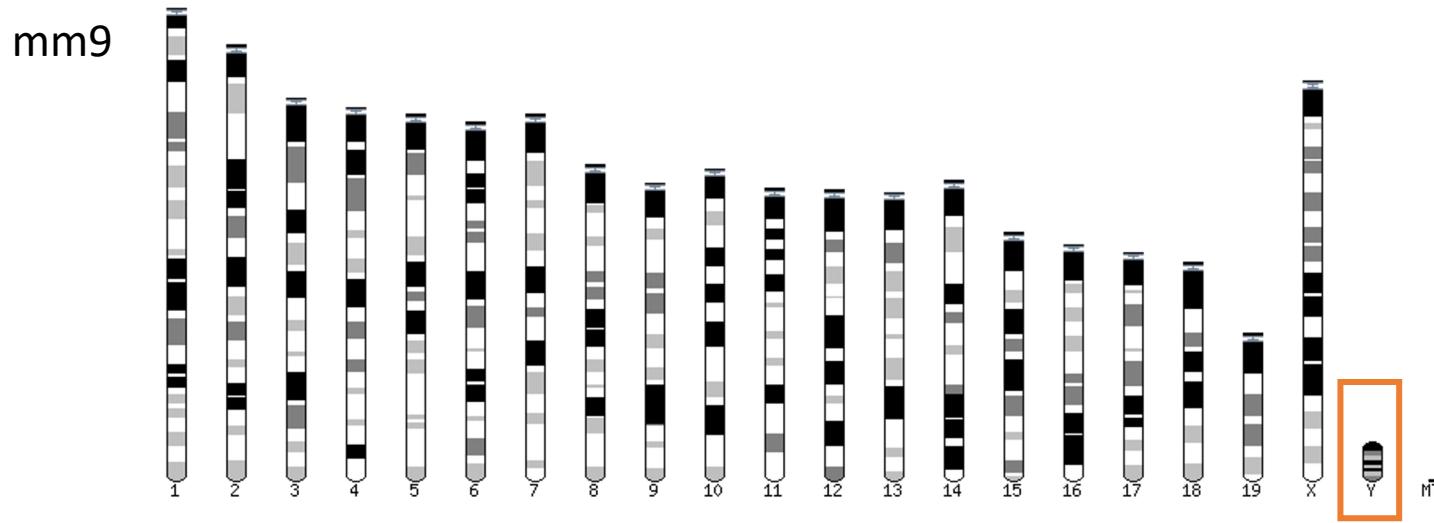
Bases de données majeures
collecte des données individuelles et collectives

Attention à la qualité de ces données
bases avec les Raw data vs Annotation

Ces données seront agrégées sur le génome humain

Genome browsers

Genome builds



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 |
| | hg12 | Jun. 2002 | NCBI Build 30 | Archived |
| | hg11 | Apr. 2002 | NCBI Build 29 | Archived (data only) |
| | hg10 | Dec. 2001 | NCBI Build 28 | Archived (data only) |
| | hg8 | Aug. 2001 | UCSC-assembled | Archived (data only) |
| | hg7 | Apr. 2001 | UCSC-assembled | Archived (data only) |
| | hg6 | Dec. 2000 | UCSC-assembled | Archived (data only) |
| | hg5 | Oct. 2000 | UCSC-assembled | Archived (data only) |
| | hg4 | Sep. 2000 | UCSC-assembled | Archived (data only) |
| | hg3 | Jul. 2000 | UCSC-assembled | Archived (data only) |
| | hg2 | Jun. 2000 | UCSC-assembled | Archived (data only) |
| | hg1 | May 2000 | UCSC-assembled | Archived (data only) |

Genome Browsers – L'outil de référence

- Elément de référence absolue le **génome**
- Agrégateur et générateur d'informations/annotations
 - Prédictions de gènes
 - Protéines
 - Données d'expression
 - Variations
- Synthèse rapide et visuelle de données primordiales

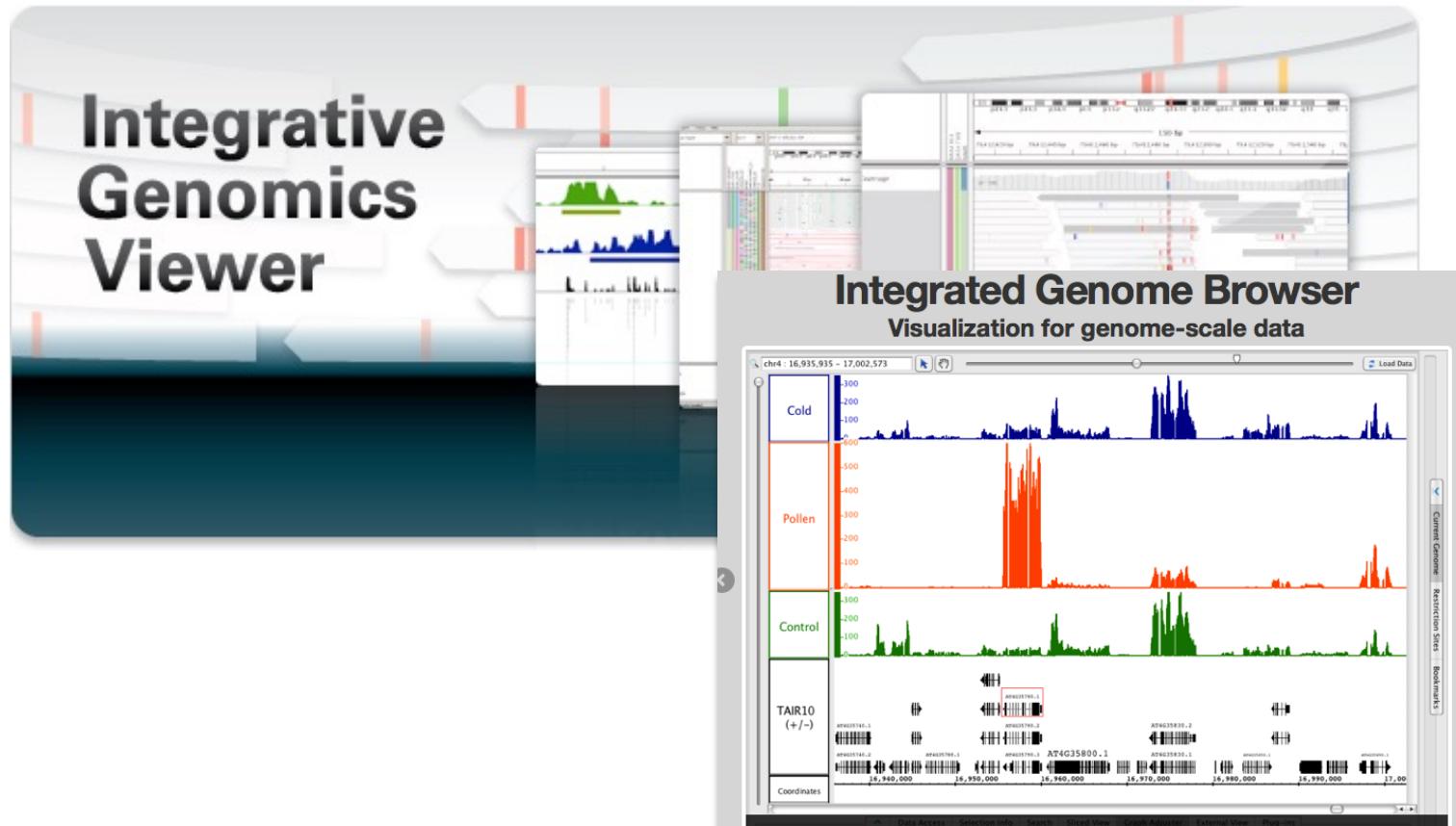
Il y a Genome Browsers...

EBI - Ensembl

UCSC – Genome Browser

NCBI – Map Viewer

Et Genome browsers



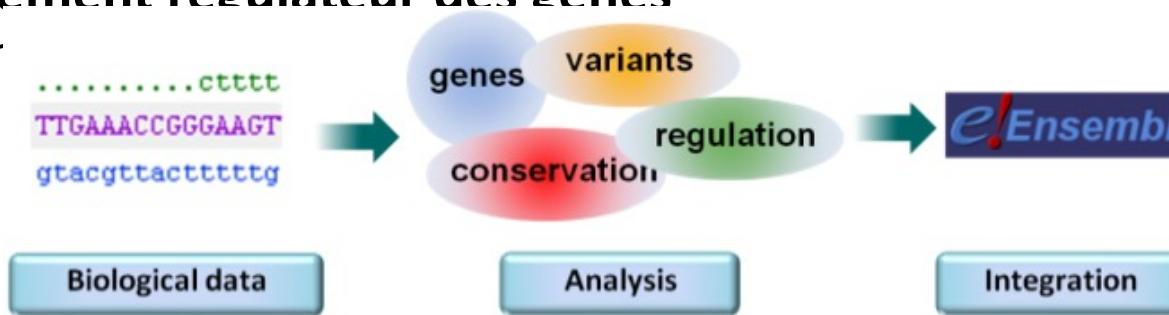
Ensembl

Le projet Ensembl

- Initié en 1999 (avant la première version du génome humain)
- Projet en collaboration entre l'European Bioinformatics Intitute (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
 - Elément régulateur des gènes
 - Ar



- 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) :
<https://ensemblgenomes.org/>
 - 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

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

Login/Register

Search all species...

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 108 (Oct 2022)

- Changes in the default tracks in the Location view: cDNAs EST cluster (UniGene) CCDS to be removed when MANE Select is available
- RNASeq tracks including data from GeneSWiCH consortium for chicken
- Variation data for crab-eating macaque, pike-perch, prairie vole, Japanese quail and collared flycatcher
- Retirement of postGAP tool

More release news on our blog

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.

Note: species that already exist on this site will continue to be updated with the full range of annotations.

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

Rapid Release news on our blog

Other news from our blog

- 02 Dec 2022: Job: Senior Full Stack Developer
- 24 Nov 2022: The first invertebrate-themed Ensembl Rapid Release is out!
- 18 Nov 2022: Geek for a Week : Georgia Argirou

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.

Permanent link - View in archive site

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

Global Core Biodata Resource

elixir Core Data Resource

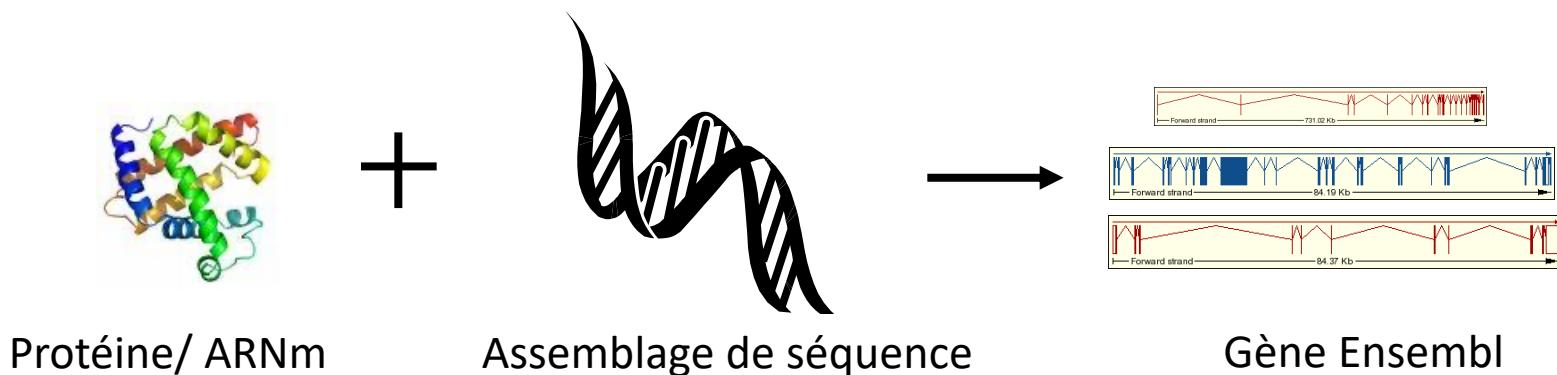
Comprendre ENSEMBL

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

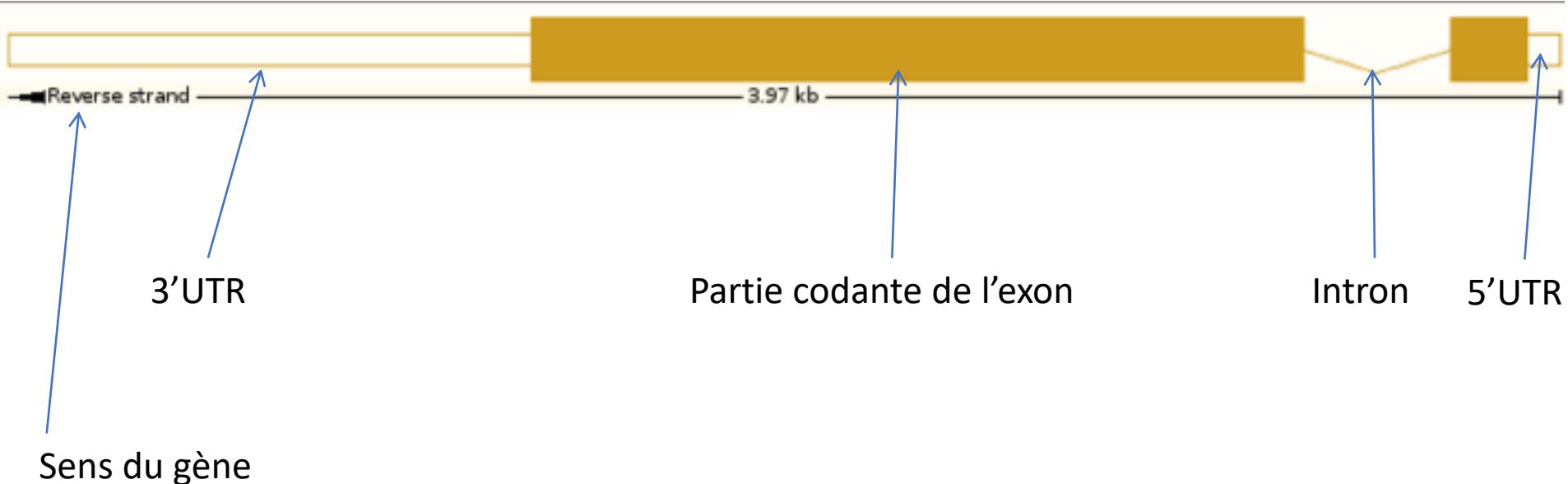
- Les transcrits d'Ensembl sont basés sur les bases de données suivantes :
 - Uniprot/Swiss-Prot (*curation manuelle*)
 - Uniprot/TrEMBL
 - NCBI refSeq (*curation manuelle*)



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/CcdsBrowse.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)

Transcrits Ensembl



Identifiants Ensembl

- ENS**G**### Ensembl Gene ID
- ENST**T**### Ensembl Transcript ID
- ENSP**P**### Ensembl Peptide ID
- ENSE**E**### Ensembl Exon ID
- Ajout d'un suffix pour les autres espèces
 - MUS (*Mus musculus*) pour la souris: ENS**MUS**G###
 - DAR (*Danio rerio*) pour le zebrafish: ENS**DAR**G###
 - etc.

Version (Release)

- ~ tous les 3-4 mois
- Lien vers la dernière version d'Ensembl est toujours : <http://www.ensembl.org>

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.



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

- Lien vers une version particulière d'Ensembl : <http://Oct2022.archive.ensembl.org/index.html>

Ensembl : Archives

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

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

Search all species...

In this section | Archives: Table of assemblies | Help & Documentation | Using this website | Archives | Search documentation | Go

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](#).
- Annotation on the [mouse NCBIm37 assembly](#) is available at our [Ensembl 67 archive](#).
- 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;g=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;g=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 108 - Oct 2022 © EMBL-EBI | Permanent link

About Us

[About us](#)

[Contact us](#)

[Citing Ensembl](#)

Get help

[Using this website](#)

[Adding custom tracks](#)

[Downloading data](#)

Our sister sites

[Ensembl Bacteria](#)

[Ensembl Fungi](#)

[Ensembl Plants](#)

Follow us

[Blog](#)

[Twitter](#)

[Facebook](#)

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

Ensembl : Archives

Archive! Ensembl BioMart | Downloads | Help & Docs | Blog

Login/Register

Search all species... 

Tools **BioMart >**

[All tools](#) Export custom datasets from Ensembl with this data-mining tool

Search

All species for 

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

All genomes  **Favourite genomes** 

 **Human**
GRCh38.p13
[Still using GRCh37?](#)

 **Mouse**
GRCm39

 **Zebrafish**
GRCz11

[View full list of all species](#)

Ensembl Archive Release 104 (May 2021)

- Update to the Ensembl Canonical transcript set.
- Human and mouse gene sets updated to GENCODE 38 and GENCODE M27, respectively.
- Retirement of gene names derived from BAC clones.

[More release news](#)  on our blog

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.

Note: species that already exist on this site will continue to be updated with the full range of annotations.



The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-

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>

Screenshot of the Ensembl Archives website showing a table of assemblies across various species and time periods.

The table displays assembly versions for different species, categorized by presence in the archive (yellow) or absence (grey). The columns represent months from Oct 2022 to Sep 2015, with specific assembly versions labeled (e.g., v108, v107, v106, v105, v104, v103, v102, v101, v99, v98, v97, v96, v95, v94, v93, v92, v91, v90, v89, v88, v87, v86, v85, v84, v83, v82).

| Species | Oct 2022 v108 | Jul 2022 v107 | Apr 2022 v106 | Dec 2021 v105 | May 2021 v104 | Feb 2021 v103 | Nov 2020 v102 | Aug 2020 v101 | Apr 2020 v100 | Jan 2020 v99 | Sep 2019 v98 | Jul 2019 v97 | Apr 2019 v96 | 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 | | | |
|--------------------------------|--------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--|--|--|
| Abingdon island giant tortoise | ASM359739v1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| African ostrich | ASM69969v1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Agassiz's desert tortoise | ASM289641v1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Algerian mouse | SPRET_EU_v1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Alpaca | vicPac1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Alpine marmot | marMar2.1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Amazon molly | Poecilia_formosa-5.1.2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| American beaver | C.can_genome_v1.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| American bison | Bison_UMD1.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| American black bear | ASM334442v1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| American mink | NNQGG.v01 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Angola colobus | Cang.pa_1.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Arabian camel | CamDro2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Arctic ground squirrel | ASM342692v1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Argentine black and white tegu | HLtupMer3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Species | Oct 2022 v108 | Jul 2022 v107 | Apr 2022 v106 | Dec 2021 v105 | May 2021 v104 | Feb 2021 v103 | Nov 2020 v102 | Aug 2020 v101 | Apr 2020 v100 | Jan 2020 v99 | Sep 2019 v98 | Jul 2019 v97 | Apr 2019 v96 | 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 | | | |
| Armadillo | Dasnov3.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Asian bonytongue | fSciFor1.1 | | | | | | | | | | | | | | | ASM162426v1 | | | | | | | | | | | | | | |
| Asiatic black bear | ASM966005v1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Atlantic cod | gadMor3.0 | | | | | | | | | | | | | | gadMor1 | | | | | | | | | | | | | | | |
| Atlantic herring | Ch_v2.0.2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Atlantic salmon | Ssal_v3.1 | | | | | | | | | | | | | | ICSASG_v2 | | | | | | | | | | | | | | | |
| Australian saltwater crocodile | CroPor_comp1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Balloon wrasse | BalGen_V1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Barramundi perch | ASB_HGAPassembly_v1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Beluga whale | ASM228892v3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Bengalese finch | LonStrDom1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Bicolor damselfish | Stegastes_partitus-1.0.2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

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

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

Search all species... 🔍

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 --

Pig breeds
Pig reference genome and 12 additional breeds
[View full list of all species](#)

Favourite genomes ✎

Human
GRCh38.p13
[Still using GRCh37?](#)

Mouse
GRCm39

Zebrafish
GRCz11

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.
Note: species that already exist on this site will continue to be updated with the full range of annotations.

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

[Rapid Release news](#) on our blog

Other news from our blog

- 02 Dec 2022: Job: Senior Full Stack Developer
- 24 Nov 2022: The first invertebrate-themed Ensembl Rapid Release is out!
- 18 Nov 2022: Geek for a Week : Yerorgia Argirou

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.

TGBC GLOBAL CORE BIODATA RESOURCE

elijir Core Data Resource

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

Ensembl Genomes

Bactéries

EnsemblBacteria - HMMER | BLAST | Tools | Downloads | More | Search Ensembl Bacteria...

Search for a gene | Search for a genome
e.g. [hsZ or uridine](#) | Start typing the name of a genome...
e.g. type egs to find Escherichia

Archive sites
The following archive sites are available to access previous versions of data:

- Release 49, December 2020 [eg9-bacteria.ensembl.org](#)
- Release 45, September 2019 [eg45-bacteria.ensembl.org](#)
- Release 40, July 2018 [eg40-bacteria.ensembl.org](#)
- Release 37, October 2017 [eg37-bacteria.ensembl.org](#)

Search for a gene - type the name of a gene or other identifier into the search box above.
 Find a genome - click in the browse a genome box above and start typing your genome name to find matching genomes.
 View full list of all Ensembl Bacteria species
 Access Ensembl Bacteria programmatically

What's New in Release 52
Release 52 of Ensembl Bacteria has no major updates since the previous release. As for releases 49-45, we are defining bacterial genomes as defined by criteria set out by UniProt. See more details about this update in our [blog post](#).

• Genomes
 • A total of 31,332 bacterial and archaeal genomes

• Data
 • Annotation of pathogen-host interaction data ([PhI-base](#)) version 2019-09-16
 • Alignments to Rfam covariance alignments (Rfam 12.2) visible in separate track (Rfam module)

Did you know?
 To access Ensembl Genomes data from any programming language, try our [REST API](#). For full documentation, including examples and a range of genomic data, visit [http://rest.ensembl.org](#).

Ensembl Genomes is developed by EMBL-EBI and is powered by the Ensembl software system for the analysis and visualisation of genomic data. For details of our funding please [click here](#).

EMBL-EBI  

Fungi

EnsemblFungi - HMMER | BLAST | BioMart | Tools | Downloads | More | Search Ensembl Fungi...

Search: All species | Go
e.g. [NAT2 or alcohol](#)

All genomes **Favourite genomes**  
Select a species -  
[View full list of all species](#)

What's New in Release 52
 • Genomes
 • EnsemblFungi has 1506 genomes in total
 • 477 new genomes imported from ENA ([https://www.ebi.ac.uk/ena/browser/home](#))
 • 15 genomes imported from VEuPath DB

• Updated data
 • Updated fungal gene trees
 • Updated protein features for all species using InterProScan with version 86 of InterPro

• Updated BioMarts for all gene and variation data
 • Updated pan-taxonomic gene trees and homologies

Ensembl Rapid Release
New assemblies with gene and protein annotation every two weeks.
Note: species that already exist on this site will continue to be updated with the full range of annotations.

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

Rapid Release news on our blog

Archive sites

Plantes

EnsemblPlants - HMMER | BLAST | BioMart | Tools | Downloads | More | Search Ensembl Plants...

Search: All species | Go
e.g. Carboxy* or chx28

All genomes **Favourite genomes**  TAIR10
Select a species - 
[View full list of all species](#)

Wheat assemblies
Ensembl Plants hosts the latest wheat assembly from the IWGSC (RefSeq v1.0), including:

- The IWGSC RefSeq v1.1 gene annotation, with links to [wheat-expression.com](#) and [KoTeMapper](#)
- 14 wheat cultivars from the [10x genome project](#)
- Alignment of 98,270 high confidence genes from the TGACv1 annotation
- Axon 53K, 800K SNP arrays from [CerealiDB](#), including QTL links in selected cases and Linkage Disequilibrium display. See QTL example [here](#).
- EMS-induced mutations from sequenced TILLING populations of *Cassava* (coding regions) and *Kinnow* (coding regions and promoters).
- Inter-*homologous Variants* (IHVs) between the A, B and D genome components
- Chromosome specific KASP markers were added from the Nottingham BBSRC Wheat Research Centre.
- Whole genome alignments to rice, *Brachypodium* and barley.
- Assembly-to-assembly mapping and gene ID mapping to the previous *TAO4* assembly are available at [https://www.ensembl.org](#).
- Phylogenetic analysis, allowing users to view alignments among multiple wheat components simultaneously.
- Dunum wheat 35K, 90K, 200K and TaBW200K variants
- Chromosome and centromere data can be viewed [here](#).

Archive sites
Archive of release 49 of EnsemblPlants: [eg49-plants.ensembl.org](#) (Dec 2020)
Archive of release 45 of EnsemblPlants: [eg45-plants.ensembl.org](#) (Sep 2019)

Navigation dans Ensembl

Protistes

EnsemblProtists - HMMER | BLAST | BioMart | Tools | More | Search Ensembl Protists...

Search: All species | Go
e.g. PF3D7_0532500 or cyto*

All genomes **Favourite genomes**  WBC07v2
Select a species - 
[View full list of all species](#)

What's New in Release 52
 • Genomes
 • No updated genomes from last release

• Updated data
 • Updated protein features for all species using InterProScan with version 86 of InterPro

• Updated BioMarts for all gene and variation data

• Updated pan-taxonomic gene trees and homologies

Ensembl Rapid Release
New assemblies with gene and protein annotation every two weeks.
Note: species that already exist on this site will continue to be updated with the full range of annotations.

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

Rapid Release news on our blog

Archive sites
The following archive sites are available to access previous versions of data:

- Release 49, December 2020 [eg49-protists.ensembl.org](#)
- Release 45, September 2019 [eg45-protists.ensembl.org](#)
- Release 40, July 2018 [eg40-protists.ensembl.org](#)

Métazoaires

EnsemblMetazoa - HMMER | BLAST | BioMart | Tools | More | Search Ensembl Metazoa...

Search: All species | Go
e.g. CP934 or chitin*

All genomes **Favourite genomes**  WBC07v35
Select a species - 
[View full list of all species](#)

What's New in Release 52
 • Updated data
 • Updated species
 • *Cimex lectularius* (Hirudin)

• Updated protein features for all species using InterProScan with version 86 of InterPro

• Updated BioMarts for all gene and variation data

• Updated pan-taxonomic gene trees and homologies

Ensembl Rapid Release
New assemblies with gene and protein annotation every two weeks.
Note: species that already exist on this site will continue to be updated with the full range of annotations.

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

Rapid Release news on our blog

Archive sites
Archive of release 49 of EnsemblMetazoa: [eg49-metazoa.ensembl.org](#) (Dec 2020)
Archive of release 45 of EnsemblMetazoa: [eg45-metazoa.ensembl.org](#) (Sep 2019)
Archive of release 40 of EnsemblMetazoa: [eg40-metazoa.ensembl.org](#) (47

Le site web Ensembl: page d'accueil

Outils



Recherche



Recherche

Search

All species for Go

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

News

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotates 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 108 (Oct 2022)

- Changes in the default tracks in the Location view: cDNAs EST cluster (UniGene) CCDS to be removed when MANE Select is available
- RNASeq tracks including data from GeneSWiCH consortium for chimp
- Variation data for crab-eating macaque, pike-perch, prairie vole, Japanese grey flycatcher
- Retirement of postGAP tool

[More release news](#) on our blog

Liste déroulante Accès aux génomes

All genomes

Select a species --

Pig breeds
Pig reference genome and 12 additional breeds

View full list of species

Favourite genomes

Human GRCh38.p13
Still using GRCh37?

Mouse GRCm39

Zebrafish GRCz11

Find SNPs and other variants for my gene

Gene expression in different tissues

Retrieve gene sequence

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.

The acknowledgements page includes a list of current and previous funding bodies. How to cite Ensembl in your own publications.

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.

Note: species that already exist on this site will continue to be updated with the full range of annotations.

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

[Rapid Release news](#) on our blog

Other news from our blog

- 02 Dec 2022: Job: Senior Full Stack Developer
- 24 Nov 2022: The first invertebrate-themed Ensembl Rapid Release is out!
- 18 Nov 2022: Geek for a Week : Georgia Argirou

Accès aux archives d'Ensembl



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

Le site web Ensembl: les génomes

Recherche

Gene annotation

Lien vers des exemples

Informations, statistiques

Ensembl release 108 - Oct 2022 © EMBL-EBI

Permanent link - View in archive site

About Us

About us

Contact us

Citing Ensembl

Privacy policy

Get help

Using this website

Adding custom tracks

Downloading data

Videos/tutorials

Our sister sites

Ensembl Bacteria

Ensembl Fungi

Ensembl Plants

Ensembl Protists

Follow us

Blog

Twitter

Facebook

Le site web Ensembl: statistiques des génomes

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

Human (GRCh38.p13) ▾

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 UCSC to create their hg38 database. The data set consists of gene models built from the genewise alignments of the human proteome as well as from alignments of human cDNAs using the cDNA2genome model of exonrate.

This release of the assembly has the following properties:

- contig length total 3.4 Gb.
- chromosome length total 3.1 Gb (excluding haplotypes).

It also includes 261 alt loci scaffolds, mainly in the LRC/KIR complex on chromosome 19 (35 alternate sequence representations) and the [MHC 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 an alternative assembly to the main assembly, patches are being introduced. Currently, assembly patches are of two types:

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

Informations générales sur l'assemblage

The Ensembl human genome 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](#).

Statistics

Summary

| | |
|---------------------------------------|---|
| Assembly | GRCh38.p13 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.28 , Dec 2013 |
| Base Pairs | 3,096,649,726 |
| Golden Path Length | 3,096,649,726 |
| Assembly provider | Genome Reference Consortium |
| Annotation provider | Ensembl |
| Annotation method | Full genebuild |
| Genebuild started | Jan 2014 |
| Genebuild released | Jul 2014 |
| Genebuild last updated/patched | Jul 2022 |
| Database version | 108.38 |
| Gencode version | GENCODE 42 |

Gene counts (Primary assembly)

| | |
|-------------------------|-------------------------------|
| Coding genes | 19,813 (excl 651 readthrough) |
| Non coding genes | 25,972 |
| Small non coding genes | 4,864 |
| Long non coding genes | 18,887 |
| Misc non coding genes | 2,221 |
| Pseudogenes | 15,241 |
| Gene transcripts | 252,477 |

Gene counts (Alternative sequence)

| | |
|-------------------------|-----------------------------|
| Coding genes | 3,028 (excl 26 readthrough) |
| Non coding genes | 1,682 |
| Small non coding genes | 297 |
| Long non coding genes | 1,198 |
| Misc non coding genes | 187 |
| Pseudogenes | 1,796 |
| Gene transcripts | 21,630 |

Other

| | |
|---------------------------------|-------------|
| Genscan gene predictions | 51,756 |
| Short Variants | 715,081,156 |
| Structural variants | 7,097,115 |

Le site web Ensembl: caryotype

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

Human (GRCh38.p13) ▾

Login/Register

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

Add features

Add/remove tracks Custom tracks Share Export image Reset configuration

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

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

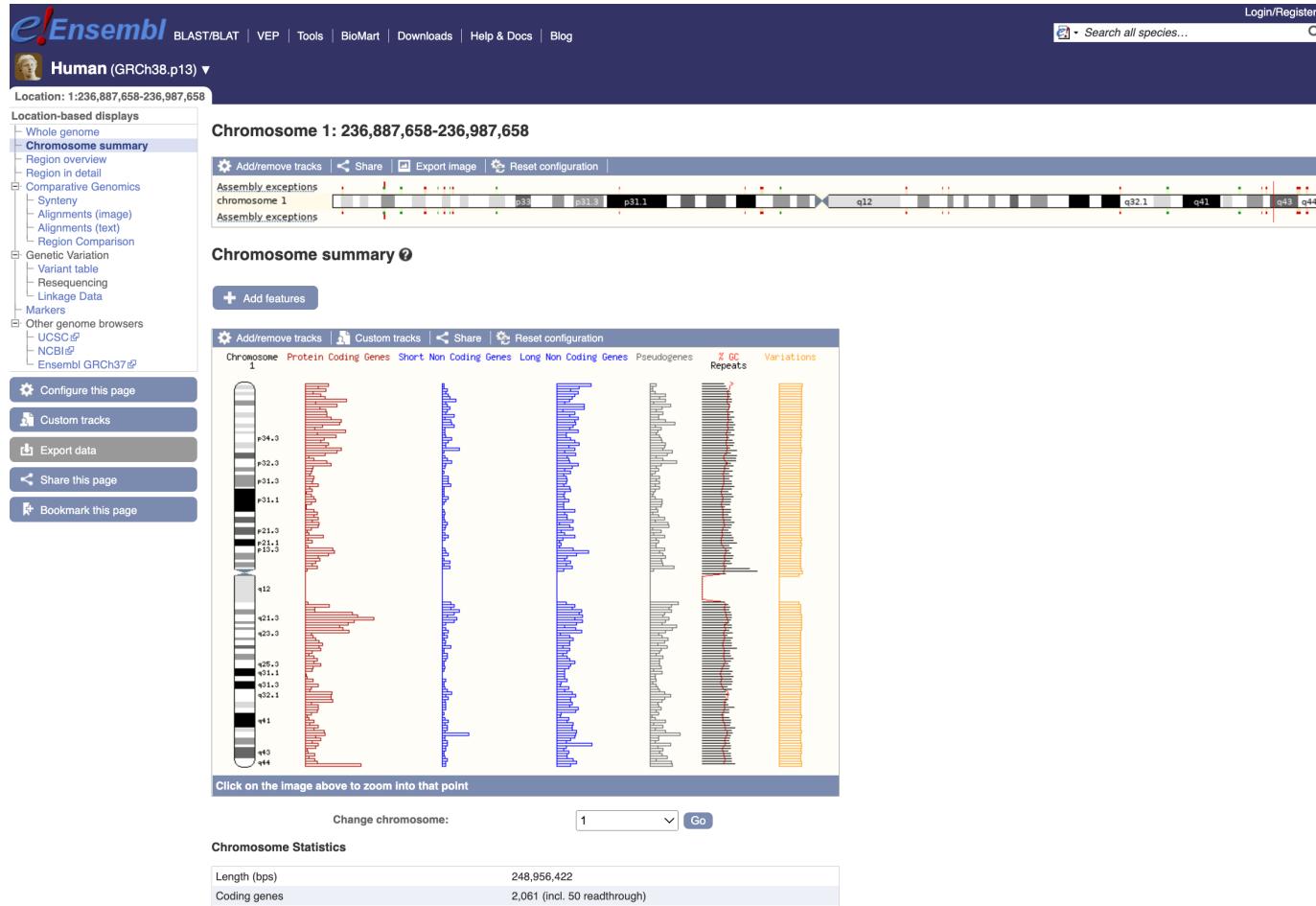
Summary

| | |
|--------------------------------|--|
| Assembly | GRCh38.p13 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.28, Dec 2013 |
| Base Pairs | 3,096,649,726 |
| Golden Path Length | 3,096,649,726 |
| Assembly provider | Genome Reference Consortium |
| Annotation provider | Ensembl |
| Annotation method | Full genebuild |
| Genebuild started | Jan 2014 |
| Genebuild released | Jul 2014 |
| Genebuild last updated/patched | Jul 2022 |
| Database version | 108.38 |
| Gencode version | GENCODE 42 |

Gene counts (Primary assembly)

| | |
|------------------------|-------------------------------|
| Coding genes | 19,813 (excl 651 readthrough) |
| Non coding genes | 25,972 |
| Small non coding genes | 4,864 |
| Long non coding genes | 18,887 |
| Micro non coding genes | 2,221 |

Le site web Ensembl : statistiques par chromosome



Le site web Ensembl : navigateur de génome

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

Location: Human (GRCh38.p13) ▾

Gene: BRCA2

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
- Linkage Data
- Markers
- Other genome browsers
- UCSC
- NCBI
- Ensembl GRCh37

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Chromosome 13: 32,315,086-32,400,268

Add/remove tracks | Share | Export image | Reset configuration | Assembly exceptions

Chr. 13 p13 p11.2 q12.11 q12.3 q13.3 q14.11 q14.2 q14.3 q21.1 q21.33 q31.1 q31.3 q32.1 q33.1 q33.3 q34

Region in detail

Add/remove tracks | Share | Resize image | Export image | Reset configuration | Reset track order | Switch image | Scroll: ← → Track height: □ Drag>Select: ← → Forward strand

Chromosome bands

Contigs

Genes (Merged Ensembl/Havana)

Regulatory Build

AC002525.1 > AL137143.8 > AL137143.8 > AL138692.26 > AL445212.9 > AL137247.14 > AL353665.13 > ZB4467.1 > Z75889.1 > AL138620.1

51.90 Mb 32.00 Mb 32.10 Mb 32.20 Mb 32.30 Mb 32.40 Mb 32.50 Mb 32.60 Mb 32.70 Mb 32.80 Mb

q13.1

59381 Y> EEF1D P3 > FRY AS1 Metazoa SRP < ENSG00000279314

ENSG00000289817 > N4BP2L2 < ZAR1L IFIT1P1 > N4BP2L2 < ENSG00000212293 < N4BP2L1 ATP8A2P2 > PDS5B < RNY1P4 ENSG000000277

SG00000287904

Regulation Legend

merged Ensembl/Havana

CTCF

open chromatin

enhancer

transcription factor binding

processed transcript

RNA gene

Regulation Legend

Location: 13:32315086-32400268 Go Gene: Go

Add/remove tracks | Custom tracks | Share | Resize image | Export image | Reset configuration | Reset track order | Drag>Select: ← → Forward strand

Chromosome bands

91-way GERP elements

MANE Select Transcripts

Genes (Comprehensive set from GENCODE.42)

85.18 kb

32.32Mb 32.33Mb 32.34Mb 32.35Mb 32.36Mb 32.37Mb 32.38Mb 32.39Mb

q13.1

Constrained elements for 91 eutherian mammals EPO-Extended

BRCA2-201 - ENST00000380152 > protein coding

BRCA2-206 - ENST00000544455 > protein coding

BRCA2-204 - ENST00000530893 > protein coding

BRCA2-201 - ENST00000380152 > protein coding

BRCA2-208 - ENST00000665585 > nonsense mediated decay

BRCA2-212 - ENST00000702020 > retained intron

BRCA2-214 - ENST00000700202 > protein coding

Le site web Ensembl : le gène

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

Human (GRCh38.p13) ▾

Location: 13:32,315,086-32,400,268

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
- Ontologies
 - GO: Molecular function
 - GO: Biological process
 - GO: Cellular component
- Phenotypes
- Genetic Variation
 - Variant table
 - Variant image
 - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
- Gene history

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

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

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

Location: Chromosome 13: 32,315,086-32,400,268 forward strand. GRCh38:CM000675.2

About this gene: This gene has 15 transcripts (splice variants), 174 orthologues and is associated with 182 phenotypes.

Transcripts: Hide transcript table

Show/hide columns (1 hidden)

| Transcript ID | Name | bp | Protein | Biotype | CCDS | UniProt Match | RefSeq Match | Flags |
|-------------------|-----------|-------|------------|-------------------------|------------|---------------|--------------|---|
| ENST00000380152.8 | BRCA2-201 | 11954 | 3418aa | Protein coding | CCDS9344 | P51587 | NM_000059.4 | MANE Select Ensembl Canonical GENCODE basic APPRIS P1 TSL:5 |
| ENST00000680887.1 | BRCA2-210 | 11880 | 3418aa | Protein coding | CCDS9344 | A0A7P0T9D7 | - | APPRIS P1 |
| ENST00000544456.6 | BRCA2-206 | 11854 | 3418aa | Protein coding | CCDS9344 | P51587 | - | GENCODE basic APPRIS P1 TSL:1 |
| ENST00000700202.1 | BRCA2-214 | 2673 | 890aa | Protein coding | - | - | - | CDS 5' incomplete |
| ENST00000530893.6 | BRCA2-204 | 2011 | 481aa | Protein coding | A0A590UJ17 | - | - | TSL:1 CDS 3' incomplete |
| ENST00000614259.2 | BRCA2-207 | 11763 | 2649aa | Nonsense mediated decay | A0A7P0TAP7 | - | - | TSL:2 |
| ENST00000665585.1 | BRCA2-208 | 2598 | 438aa | Nonsense mediated decay | A0A590UJU6 | - | - | CDS 5' incomplete |
| ENST00000700201.1 | BRCA2-213 | 2103 | 129aa | Nonsense mediated decay | - | - | - | - |
| ENST00000470094.1 | BRCA2-202 | 842 | 186aa | Nonsense mediated decay | H0YE37 | - | - | TSL:5 CDS 5' incomplete |
| ENST00000666593.1 | BRCA2-209 | 523 | 58aa | Nonsense mediated decay | A0A590UJ24 | - | - | CDS 5' incomplete |
| ENST00000528762.1 | BRCA2-203 | 495 | 64aa | Nonsense mediated decay | H0YD86 | - | - | TSL:4 CDS 5' incomplete |
| ENST00000700203.1 | BRCA2-215 | 2532 | No protein | Retained intron | - | - | - | - |
| ENST00000700200.1 | BRCA2-212 | 860 | No protein | Retained intron | - | - | - | - |
| ENST00000700199.1 | BRCA2-211 | 553 | No protein | Retained intron | - | - | - | - |
| ENST00000533776.1 | BRCA2-205 | 523 | No protein | Retained intron | - | - | - | TSL:3 |

Summary:

Name: BRCA2 (HGNC Symbol)

MANE: This gene contains MANE Select ENST00000380152, ENSP00000369497

UniProtKB: This gene has proteins that correspond to the following UniProtKB identifiers: P51587

RefSeq: This Ensembl/Gencode gene contains transcript(s) for which we have selected identical RefSeq transcript(s). If there are other RefSeq transcripts available they will be in the External references table

CCDS: This gene is a member of the Human CCDS set: CCDS9344.1

LRG: LRG_293 provides a stable genomic reference framework for describing sequence variants for this gene

Ensembl version: ENSG00000139618.18

Other assemblies: This gene maps to 32,889,223-32,974,405 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000139618

Gene type: Protein coding

Login/Register

Search all species...

54

Le site web Ensembl : le transcript

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

Human (GRCh38.p13) ▾

Location: 13:32,315,086-32,400,268 Gene: BRCA2 Transcript: BRCA2-201

Transcript-based displays

- Summary
- Sequence
 - Exons
 - cDNA
 - Protein
- Protein Information
 - Protein summary
 - Domains & features
 - Variants
 - PDB 3D protein model
 - AlphaFold predicted model
- Genetic Variants
 - 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

Login/Register

Search all species...

Transcript: ENST00000380152.8 BRCA2-201

Description BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:[HGNC:1101](#)]

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

Location Chromosome 13: 32,315,086-32,400,268 forward strand.

About this transcript This transcript has 27 exons, is annotated with 68 domains and features, is associated with 35622 variant alleles and maps to 958 oligo probes.

Gene This transcript is a product of gene ENSG00000139618.[Hide transcript table](#)

| Transcript ID | Name | bp | Protein | Biotype | CCDS | UniProt Match | RefSeq Match | Flags |
|-------------------|-----------|-------|------------|-------------------------|----------|---------------|--------------|---|
| ENST00000380152.8 | BRCA2-201 | 11954 | 3418aa | Protein coding | CCDS9344 | P51587 | NM_000059.4 | MANE Select Ensembl Canonical GENCODE basic APPRIS P1 TSL:5 |
| ENST00000680887.1 | BRCA2-210 | 11880 | 3418aa | Protein coding | CCDS9344 | A0A7P0T9D7 | - | APPRIS P1 |
| ENST00000544455.6 | BRCA2-206 | 11854 | 3418aa | Protein coding | CCDS9344 | P51587 | - | GENCODE basic APPRIS P1 TSL:1 |
| ENST00000700202.1 | BRCA2-214 | 2673 | 890aa | Protein coding | - | - | - | CDS 5' incomplete |
| ENST00000530893.6 | BRCA2-204 | 2011 | 481aa | Protein coding | - | A0A590UJ77 | - | TSL:1 CDS 3' incomplete |
| ENST00000614259.2 | BRCA2-207 | 11763 | 2649aa | Nonsense mediated decay | - | A0A7P0TAP7 | - | TSL:2 |
| ENST00000665585.1 | BRCA2-208 | 2598 | 438aa | Nonsense mediated decay | - | A0A590UJU6 | - | CDS 5' incomplete |
| ENST00000700201.1 | BRCA2-213 | 2103 | 129aa | Nonsense mediated decay | - | - | - | - |
| ENST00000470094.1 | BRCA2-202 | 842 | 186aa | Nonsense mediated decay | - | H0YE37 | - | TSL:5 CDS 5' incomplete |
| ENST00000666593.1 | BRCA2-209 | 523 | 58aa | Nonsense mediated decay | - | A0A590UJ24 | - | CDS 5' incomplete |
| ENST00000528762.1 | BRCA2-203 | 495 | 64aa | Nonsense mediated decay | - | H0YD86 | - | TSL:4 CDS 5' incomplete |
| ENST00000700203.1 | BRCA2-215 | 2532 | No protein | Retained intron | - | - | - | - |
| ENST00000700200.1 | BRCA2-212 | 860 | No protein | Retained intron | - | - | - | - |
| ENST00000700199.1 | BRCA2-211 | 553 | No protein | Retained intron | - | - | - | - |
| ENST00000533776.1 | BRCA2-205 | 523 | No protein | Retained intron | - | - | - | TSL:3 |

Summary

Export image

BRCA2-201 ENST00000380152 > protein coding

Statistics Exons: 27, Coding exons: 26, Transcript length: 11,954 bps, Translation length: 3,418 residues

MANE This MANE Select transcript contains [ENSP00000369497](#) and matches to [NM_000059.4](#) and [NP_000050.3](#)

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

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

Transcript Support Level (TSL) TSL:5

Version ENST00000380152.8

Type Protein coding

Annotation Method Transcript where the Ensembl genebuild transcript and the Havana 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 108 - Oct 2022 © EMBL-EBI

Permanent link - View in archive site

55

Naviguer dans Ensembl : Partie pratique

Visualiser ses propres données

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog [Login/Register](#)

Tools [BioMart >](#) [BLAST/BLAT >](#) [Variant Effect Predictor >](#)

[All tools](#) Export custom datasets from Ensembl with this data-mining tool

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

All genomes Favourite genomes Human GRCh38.p13 Still using GRCm37? Mouse GRCm39 Zebrafish GRCz11 [View full list of all species](#)

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 108 (Oct 2022)

- Changes in the default tracks in the Location view: cDNAs EST cluster (UniGene) CCDS to be removed when MANE Select is available
- RNASeq tracks including data from GeneSWiCH consortium for chicken
- Variation data for crab-eating macaque, pike-perch, prairie vole, Japanese quail and collared flycatcher
- Retirement of postGAP tool

[More release news](#) on our blog

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks. Note: species that already exist on this site will continue to be updated with the full range of annotations. The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes such as Darwin Tree of Life, the Vertebrate Genome Project.

Other news from our blog

- 02 Dec 2022: [Job: Senior Full Stack Developer](#)
- 24 Nov 2022: [The first invertebrate-themed Ensembl Rapid Release](#)
- 18 Nov 2022: [Geek for a Week : Yerjia Argiroiu](#)

Visualiser ses propres données

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.

GLOBAL CORE DATA RESOURCE

Ensembl release 108 - Oct 2022 © EMBL-EBI

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

About Us

About us

Contact us

Get help

Using this website

Adding custom tracks

Our sister sites

Ensembl Bacteria

Ensembl fungi

Follow us

Blog

Twitter

Navigation dans Ensembl

57

LES OUTILS

Les outils

BLAST/BLAT

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

Search all species...

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 108 (Oct 2022)

- Changes in the default tracks in the Location view: cDNAs EST cluster (UniGene) CCDS to be removed when MANE Select is available
- RNASeq tracks including data from GeneSWiCH consortium for chicken
- Variation data for crab-eating macaque, pike-perch, prairie vole, Japanese quail and collared flycatcher
- Retirement of postGAP tool

[More release news](#) on our blog

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.
Note: species that already exist on this site will continue to be updated with the full range of annotations.

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

[Rapid Release news](#) on our blog

All genomes
-- Select a species --

Pig breeds
Pig reference genome and 12 additional breeds

[View full list of all species](#)

Favourite genomes

Human GRCh38.p13
Mouse GRCm39
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.

GLOBAL CORE BIO DATA RESOURCE elixir Core Data Resource

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:

- Déetecter 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

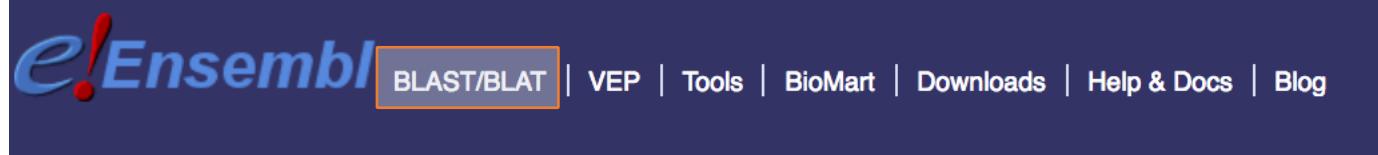
| Programmes | Requête | Banque | Comparaison | Exemples d'utilisation |
|------------|---------------------------------|---------------------------------|-------------|---|
| Blastn | ADN | ADN | nucléique | Recherche d'ARN structuraux, d'éléments régulateurs |
| Blastp | Protéine | protéines | protéique | 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

Blast



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 1981 alignments, 23 hits [RawResult] view ►

► 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

① Not yet initialised

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

Content-type: text/plain

TBLASTN 2.0MP-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....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 | High | Smallest | |
|---|---------|------|----------|-----|
| | | | Frame | Sum |
| | | | P(N) | 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 | |
| 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:... 14 dna:chromosome chromosome:GRCh37:14:1:107349540:1 REF | +3 | 149 | 0.11 | 2 |
| 1 dna:chromosome chromosome:GRCh37:1:1:249250621:1 REF | +2 | 167 | 0.21 | 8 |
| GL000220.1 dna:supercontig supercontig:GRCh37:GL000220.1:... 5 dna:chromosome chromosome:GRCh37:5:1:180915260:1 REF | -1 | 134 | 0.25 | 8 |
| GL000224.1 dna:supercontig supercontig:GRCh37:GL000224.1:... 7 dna:chromosome chromosome:GRCh37:7:1:159138663:1 REF | -3 | 127 | 0.33 | 9 |
| 21 dna:chromosome chromosome:GRCh37:21:1:48129895:1 REF | -2 | 131 | 0.88 | 9 |
| GL000237.1 dna:supercontig supercontig:GRCh37:GL000237.1:... GL000202.1 dna:supercontig supercontig:GRCh37:GL000202.1:... GL000218.1 dna:supercontig supercontig:GRCh37:GL000218.1:... 15 dna:chromosome chromosome:GRCh37:15:1:102531392:1 REF | -2 | 89 | 0.98 | 5 |
| 6 dna:chromosome chromosome:GRCh37:6:1:171115067:1 REF | -2 | 111 | 0.995 | 3 |
| 3 dna:chromosome chromosome:GRCh37:3:1:198022430:1 REF | -3 | 145 | 0.996 | 5 |
| GL000206.1 dna:supercontig supercontig:GRCh37:GL000206.1:... 92 | +2 | 134 | 0.999 | 12 |
| | -2 | 118 | 0.9991 | 13 |
| | -3 | 118 | 0.9998 | 11 |
| | -3 | 92 | 0.99992 | 6 |

| | |
|--|----------------------|
| Query: >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 INPETGEOIOSWYRSGETWDWSKFSTIASSYEECRAESVGLYLCILHPOVLEIFGFEGADAE | 536 |
| Sbjct: 76090065 INPE EQIQSWYRS +TWDSKFSTI SSYEECRAESVGLYLCILHPOVLE FGFGEGADAE | 76089886 |
| Query: 537 DVIVVNWLNMVRAGLIALEFYTPAFAFNWQRAHMQARFVILRVLIEAGEGLVITPTTGSD | 596 |
| Sbjct: 76089885 +VIV VNWLNMVY AGLIALEFYTPAFA NW+QAH+AR VILRVL EAGEGL TITPT GSD | 76089706 |
| EVISVNWLNMVGAGLLIALEFYTPAFAFNWQRAHIRARIVLVLPEAGEGLGTITPTAGSD | |
| Query: 597 GRPDARVRLDRSKIRSVGKPALERFLRRLQVLKSTGDVAGGRALYEGYATVDAPPECFL | 656 |
| GRPA+A+VRLDRSKI+SVG PALERFLRR STGDVAGG LYE YA V DAPPE FL | |
| Sbjct: 76089705 GRPEAQVRLDRSKIQSVPNALERFLRRCW---STGDVAGGWTLERYAAVADAPPGEFL | 76089535 |
| Query: 657 TLRDTVLLRKESRKLIQPNTRLEGSDVOLLEYEASAAGLIRSFSERFPEDGPELEEILT | 716 |
| TLRD VLLRKES KLIVQPN RLEGSDVOLLEYE SAAGLIRSFS FFPEDG ELE+ILT | |
| Sbjct: 76089534 TLRDRVLLRKESWKLIQPNIRLEGSDVOLLEYEVSAAGLIRSFS EHFPEDGLEDILT | 76089355 |
| Query: 717 QLATADAFRWKGPSAEPSGOA 737 | |
| QLATADA+F KGPSAEPSGOA | |
| Sbjct: 76089354 QLATADAOF*KGPSEAPSGOA 76089292 | |
| Score = 1700 (557.6 bits), Expect = 0., Sum P(6) = 0. | |
| Identities = 212/252 (84%), Positives = 221/252 (87%), Frame = -2 | |
| Query: 224 PSLDSEVTSKLKSYSYERGSPFQVTRGDYAPILQKVVEQLEKAKAYAANSHQGQMLAQYIE | 283 |
| P L + SKLKS EFRGSQFQVFI G+Y PILQKVVEQLEKAK YAANSHQ QMLAQYIE | |
| Sbjct: 76090816 PGLRGD--SKLKS*EFRGSQFQVFIWGNYMPILQKVVEQLEKAKTYAANSHQEQMLAQYIE | 76090643 |
| Query: 284 SFTQGSIEAHKRGSRFWIQDKGPIVESYIGFIESYRDPFGSRGEFEGFVAVVNKAMSAKF | 343 |
| SFTQGS EAHK-GSRFWI DKGPIVESYI FI+SYRD FGSRG EGTVAVVNKAMSAKF | |
| Sbjct: 76090642 SFTQGSTEAHKKGSRFWI*DKGPIVESYIEFIQSYRDSFGSRGVCEGFVAVVNKAMSAKF | 76090463 |
| Query: 344 ERLVASAEQLLKELPWPPTFEKDKFLTPDFTSLDVLT FAGSGIPAGINIPVNDDLRTQTEG | 403 |
| E LV SAEQLLKELPW P FEKDKFLTPDFTS+DVLTFAGSGI AGINI NY+DL+QTEG | |
| Sbjct: 76090462 EWLVVSAEQLLKELPWSPAFEKDKFLTPDFTSVDVLT FAGSGIAAGINISNYNDLKQTEG | 76090283 |
| Query: 404 FKNVSLGNVLAVAYATQREKLT FLEEDDKDLYILWKGPSFDVQVGLHELLGHGSGKLFVQ | 463 |
| FKNVSLGNVLAV ATQ EKLT LEE DKDLYI+ GPSFDVQVGLHELLG+GSGKL Q | |
| Sbjct: 76090282 FKNVSLGNLVAVV*ATQWEKLTVEESDKDLYIVLMGPSFDVQVGLHELLGYGSGKLFIEQ | 76090103 |
| Query: 464 DEKGAFNFDQET 475 | |
| DEKGAFNFDQET | |
| Sbjct: 76090102 DEKGAFNFDQET 76090067 | |

new **SETUP** **CONFIG** **RESULTS** **DISPLAY** **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

Retrieve result for ID:

BLA_IESTYdDXDJ **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 ►**

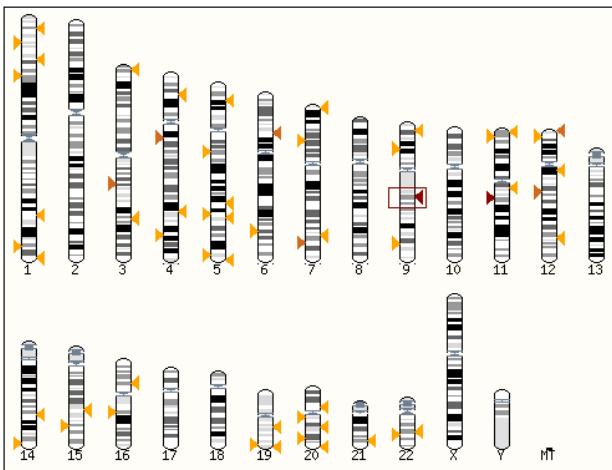
new SETUP CONFIG RESULTS DISPLAY

Displaying unnamed sequence alignments vs Homo_sapiens LATESTGP database

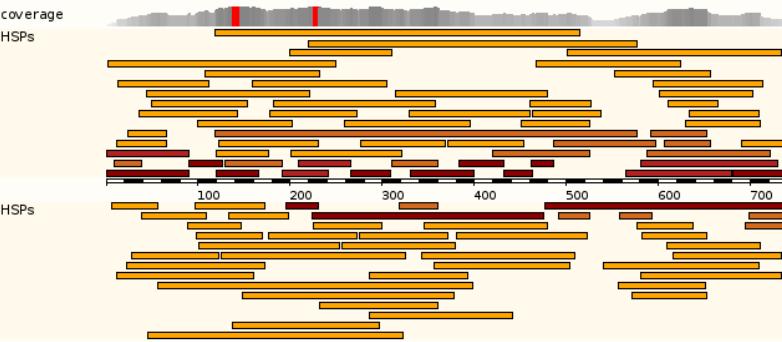
Showing top 100 alignments of 1961, sorted by Raw Score

refresh

Alignment Locations vs. Karyotype (click arrow to hide)



Alignment Locations vs. Query (click arrow to hide)



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

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 | <Lrg |
| 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]lign

```

Query location      : unnamed        477 to    727 (+)
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 INFETEQIQISWRSGETWD8KFSTIASSYEECRAESVGLYLCILHQPVLIEFGPEGADAE 536
          EQIWIPEQIWIPEW+TDWSKFSTI SYYECECRAESVGLYLCILHQPVLIEFGPEGADAE
Sbjct: 76090065 INFEMRFQIQISWRSMTWD8KFSTIWSYEECRAESVGLYLCILHQPVLIEFGPEGADAE 76089886

Query:    537 DIVIVNWLMVLMGAGLALLEFTPEAFNQAHQRMOHFWFILRLVLEAGEGLVLTITGGSD 596
          +VI VNWLNLHM AGILLALEFTPEA NW+QAH+AR VILRVL EAGEGL TITFGD GSD
Sbjct: 76089885 EVISVNWLNMVGAGLALLEFTPEASNWQOAHIRARIVILRLVLPAGEGLGTTFTAGSD 76089706

Query:    597 GREDPARDLDRSKRSVGRKPALERFLRLIQLKSLTGWDAGGRALVEYGA+WTDAPEPCL 656
          GR+E+A+VLDRSK-I+SVG PALERFLR - STGDVAG LYE Y W DAPPE FL
Sbjct: 76089705 GRFEAQWFLDRSKIQJSVGKPALERFLRRCW+ - STGDVAGWGLTLYEP AAVADAPPEGFL 76089535

Query:    657 TLRDVLLRKESKWLIIQPNTRLEGSDUQVLEYEASAAGLIRSE+TFPEDGPELEELIT 716
          TLRD VLLRKES KWLII QLKPWN RLEGSDUQVLEYE SAAGLIRSE FPDG ELE+ILT
Sbjct: 76089534 TLRDVLLRKESKWLIIQPNTRLEGSDUQVLEYEASAAGLIRSE FSHFPDGELELDIT 76089355

Query:    717 QLATADARFWKGPFSEAPSQQA 737
          QLATADAF+ KGPFSEAPSQQA
Sbjct: 76089354 QLATADAF+ KGPFSEAPSQQA 76089292

```

[S]equence

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

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

[G]Sequence

[C]ontig view (?)

```
Query location      : unnamed      477 to      727 (+)
Database location   : 9          76089292 to 76090065 (-)
Genomic location    : 9          76089292 to 76090065 (-)
```

Alignment score : 1768
E-value : 0.
Alignment length : 261
Percentage identity: 84.2

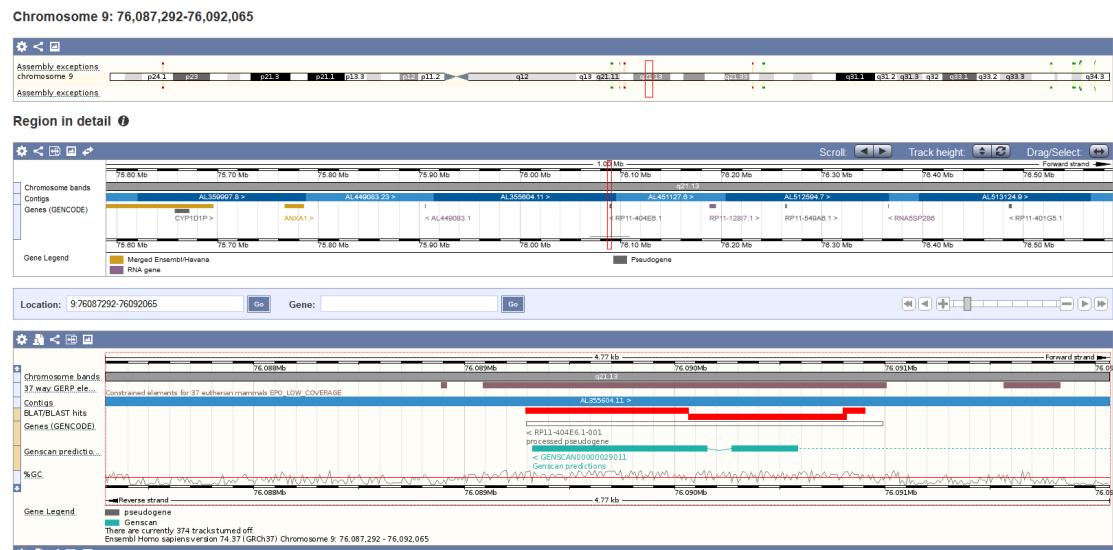
| | |
|----------------------|--|
| 5' Flanking sequence | <input type="text" value="300"/> (bp) |
| 3' Flanking sequence | <input type="text" value="300"/> (bp) |
| Coordinate system | <input type="button" value="Chromosome"/> |
| Orientation | <input type="button" value="Forward relative to selected alignment"/> |
| Alignment markup | <input type="button" value="All alignments"/> <input type="checkbox"/> Both orientations |
| Feature markup | <input type="button" value="Ensembl exons"/> <input type="checkbox"/> Both orientations |
| Line numbering | <input type="button" value="No numbers"/> |

THIS STYLE: Location of selected alignment

THIS STYLE: Location of other alignments

THIS STYLE: Location of Exons

>chromosome:GRCh37:9:76088992:76090365:-



Les outils

Annotation de variants

The Ensembl homepage features a dark blue header with the Ensembl logo and a search bar labeled "Search all species...". Navigation links include BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A "Login/Register" link is also present.

| | | | |
|---|---|---|---|
| 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 –

Pig breeds
Pig reference genome and 12 additional breeds

[View full list of all species](#)

Favourite genomes

Human GRCh38.p13
Still using GRCh37?
Mouse GRCm39
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 108 (Oct 2022)

- Changes in the default tracks in the Location view: cDNAs EST cluster (UniGene) CCDS to be removed when MANE Select is available
- RNASeq tracks including data from GeneSWiCH consortium for chicken
- Variation data for crab-eating macaque, pike-perch, prairie vole, Japanese quail and collared flycatcher
- Retirement of postGAP tool

[More release news](#) on our blog

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.

Note: species that already exist on this site will continue to be updated with the full range of annotations.

Go

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

[Rapid Release news](#) on our blog

Other news from our blog

- 02 Dec 2022: [Job: Senior Full Stack Developer](#)
- 24 Nov 2022: [The first invertebrate-themed Ensembl Rapid Release is out!](#)
- 18 Nov 2022: [Geek for a Week : Georgia Argirou](#)

Compare genes across species

Find SNPs and other variants for my gene

Gene expression in different tissues

Retrieve gene sequence

```
GCCTGATTCCTCCGGGTGAG  
GCGTGGGGGGGGGGGGGGGGGG  
GGGGGGGGGGGGGGGGGGGGGG  
GGGGGGGGGGGGGGGGGGGGGG  
AAGGGACAGATTTTGAA  
CACCCCTGGGGGGGGGGGGGG  
CCCCGGGGGGGGGGGGGGGGGG
```

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.

The image shows the logos for the Global Core Data Resource (GCDR) and eLife, indicating their involvement in the project.

Ensembl release 108 - Oct 2022 © EMBL-EBI

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

About Us

About us

Contact us

Get help

Using this website

Adding custom tracks

Our sister sites

Ensembl Bacteria

Ensembl Fungi

Follow us

Blog

Twitter

69

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

Help & Documentation API & Software Ensembl Tools Ensembl Variant Effect Predictor (VEP)

Ensembl Variant Effect Predictor (VEP)

Ve!P

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), see [variant consequences](#)
- Known variants that match yours, and associated minor allele frequencies from the [1000 Genomes Project](#)
- SIFT and PolyPhen-2 scores for changes to protein sequence
- ... And more! See [data types](#), [versions](#).

★ [What's new in release 108?](#)

VEP interfaces

Web interface



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

[Documentation](#)

Command line tool



- More options and flexibility
- For large volumes of data

[Documentation](#)

[Clone from GitHub](#)

[Download \(zip\)](#)

[Pull Docker image from DockerHub](#)

REST API



- Language-independent API
- Simple URL-based queries

[Documentation](#)

[VEP REST API](#)

Publication

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

Cite us

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

VEP related tools

Variant Effect Predictor

The screenshot shows the Ensembl Variant Effect Predictor (VEP) interface. At the top, there's a navigation bar with links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. On the right, there are buttons for Login/Register and a search bar labeled "Search all species...".

The main content area is titled "Variant Effect Predictor". It features a sidebar on the left with "Web Tools" including BLAST/BLAT, Variant Effect Predictor (selected), Linkage Disequilibrium Calculator, Variant Recoder, File Chameleon, Assembly Converter, ID History Converter, VCF to PED Converter, Data Slicer, Configure this page, Custom tracks, Export data, Share this page, and Bookmark this page.

The main form starts with a "New job" button and a "Species" dropdown set to "Homo_sapiens". Below it is a note about assembly: "Assembly: GRCh38.p13" and a link to "GRCh37 website". There's also a "Change species" button.

Below the species selection are fields for "Name for this job (optional)" and "Input data". The "Input data" section includes a text area for "Either paste data" with examples like Ensembl default, VCF, Variant identifiers, HGVS notations, and SPDI. It also has fields for "Or upload file" (with a "Choisir un fichier" button) and "Or provide file URL".

The "Transcript database to use:" section contains radio buttons for Ensembl/Gencode transcripts (selected), Ensembl/Gencode basic transcripts, RefSeq transcripts, and Ensembl/Gencode and RefSeq transcripts.

The "Additional configurations:" section includes several expandable panels: "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), and "Advanced options" (Additional enhancements).

A large green "Run" button is located at the bottom of the configuration area.

Recent jobs

You have no jobs currently running or recently completed.

Variant Effect Predictor

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

Login/Register

Search all species...

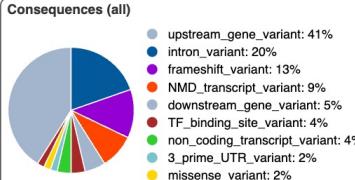
Variant Effect Predictor results 

Job details 

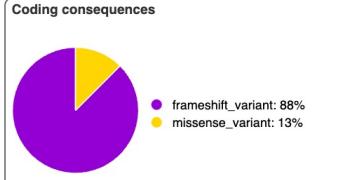
Summary statistics 

| Category | Count |
|--------------------------------|-------|
| Variants processed | 3 |
| Variants filtered out | 0 |
| Novel / existing variants | - |
| Overlapped genes | 5 |
| Overlapped transcripts | 46 |
| Overlapped regulatory features | 1 |

Consequences (all)



Coding consequences



Results preview

Navigation (per variant)  Filters  Download 

Page:  1 of 1  | Show: 1 All variants  is  Add

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

New job

Show/hide columns (13 hidden)

| Uploaded variant | Location | Allele | Consequence | Symbol | Gene | Feature type | Feature | Scroll to see more columns » |
|--|-----------------|--------|-------------------------|---------|---------------------------------|--------------|------------------------------------|--------------------------------|
| 1_65568_A/C  | 1_65568_65568 | C | downstream_gene_variant | OR4G11P | ENSG00000240361 | Transcript | ENST00000492842.2 | transcribed_unprocessed_pseudo |
| 1_65568_A/C  | 1_65568_65568 | C | missense_variant | OR4F5 | ENSG00000186092 | Transcript | ENST00000641515.2 | protein_coding |
| 1_65568_A/C  | 1_65568_65568 | C | downstream_gene_variant | OR4G11P | ENSG00000240361 | Transcript | ENST00000642116.1 | processed_transcript |
| 2_265023_C/T  | 2_265023_265023 | T | intron_variant | ACP1 | ENSG00000143727 | Transcript | ENST00000272065.10 | protein_coding |
| 2_265023_C/T  | 2_265023_265023 | T | intron_variant | ACP1 | ENSG00000143727 | Transcript | ENST00000272067.10 | protein_coding |
| 2_265023_C/T  | 2_265023_265023 | T | upstream_gene_variant | SH3YL1 | ENSG00000035115 | Transcript | ENST00000356150.10 | protein_coding |
| 2_265023_C/T  | 2_265023_265023 | T | upstream_gene_variant | SH3YL1 | ENSG00000035115 | Transcript | ENST00000402632.5 | protein_coding |
| 2_265023_C/T  | 2_265023_265023 | T | upstream_gene_variant | SH3YL1 | ENSG00000035115 | Transcript | ENST00000403657.5 | protein_coding |
| 2_265023_C/T  | 2_265023_265023 | T | upstream_gene_variant | SH3YL1 | ENSG00000035115 | Transcript | ENST00000403658.5 | protein_coding |
| 2_265023_C/T  | 2_265023_265023 | T | upstream_gene_variant | SH3YL1 | ENSG00000035115 | Transcript | ENST00000403712.6 | protein_coding |

Outils de récupération de données

Screenshot of the Ensembl website showing data download options.

The page title is "Accessing Ensembl Data".

The "Downloads" menu item is highlighted with a red box.

Section "Small quantities of data":
Many of the pages displaying Ensembl genomic data offer an [export](#) option, suitable for small amounts of data, e.g. a single gene sequence.
Click on the 'Export data' button in the lefthand menu of most pages to export:
• FASTA sequence
• GTF or GFF features
...and more!

Section "Fast programmatic access":
For fast access in any programming language, we recommend using our REST server. Various REST endpoints provide access to vast amounts of Ensembl data.

Section "Complete datasets and databases":
Many datasets, e.g. all genes for a species, are available to download in a variety of formats from our [FTP site](#).
Entire databases are also available via FTP as MySQL dumps.

Section "Complex cross-database queries":
More complex datasets can be retrieved using the [BioMart](#) data-mining tool.

All data produced by the Ensembl project is [freely available](#) for your own use.

Ensembl release 108 - Oct 2022 © EMBL-EBI

[Permanent link](#)

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

- [!\[\]\(cb464274fb50af7beb213caa4f8057c1_img.jpg\) Blog](#)
- [!\[\]\(8928d25829778359a6457593aae6f826_img.jpg\) Twitter](#)
- [!\[\]\(e638b3073557a3b86fd3542ad25abdc9_img.jpg\) Facebook](#)

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 three separate web interfaces for biological data marts:

- Ensembl BioMart:** A screenshot showing the interface for selecting columns from a dataset (Homo sapiens genes, GRCh37.p13). It includes a sidebar for filters and attributes, and a top navigation bar with links like BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors.
- UniProt BioMart:** A screenshot showing the interface for choosing a database. It features the UniProt logo and a search bar at the top, with a sidebar for dataset selection.
- ICGC Data Portal:** A screenshot showing the main search interface. It features a logo with a red and blue circular emblem, a search bar with placeholder text "eg. BRAF, KRAS G12D, DO35108, MU7870, TCGA-06-5858", and three buttons for "Cancer Projects", "Advanced Search", and "Data Repository".

Accéder aux données d'Ensembl

Site web

The screenshot shows the Ensembl homepage with a search bar at the top. Below it, there's a section for browsing genomes, a 'What's New' section for release 83, and a 'Tweets' section from the Ensembl Twitter account (@ensembl). The page also features links to BLAST, BioMart, and other Ensembl services.

Outil de fouille: BioMart

The screenshot shows the BioMart interface. At the top, there's a search bar and a 'Dataset' dropdown menu set to '[None selected]'. The main area is currently empty, indicating no specific dataset has been chosen yet.

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

- Requête complexe
- Rapide
- Requiert une formation

BioMart/Ensembl

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

Login/Register

BioMart

Tools All tools

BioMart > Export custom datasets from this

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

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 108 (Oct 2022)

- Changes in the default tracks in the Location view: cDNAs EST cluster (UniGene) CCDS to be removed when MANE Select is available
- RNASeq tracks including data from GeneSWiCH consortium for chicken
- Variation data for crab-eating macaque, pike-perch, prairie vole, Japanese quail and collared flycatcher
- Retirement of postGAP tool

More release news on our blog

All genomes -- Select a species --

Pig breeds Pig reference genome and 12 additional breeds

View full list of all species

Favourite genomes Human (GRCh38.p13), Mouse (GRCm39), Zebrafish (GRCz11)

Still using GRCh37?

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.

Note: species that already exist on this site will continue to be updated with the full range of annotations.

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

Rapid Release news on our blog

Other news from our blog

- 02 Dec 2022: Job: Senior Full Stack Developer
- 24 Nov 2022: The first invertebrate-themed Ensembl Rapid Release is out!
- 18 Nov 2022: Geek for a Week : Georgina Argirou

Compare genes across species

EMBL-EBI Ensembl creates Our acknowledgements

Ensembl release 108 - Oct 2022 © EMBL

About Us

About us

Contact us

elixir Core Data Resource

Link - View in archive site

• Accès à :

- Annotation génomique (gènes, SNPs)
- Annotation fonctionnelle
- Expression

78

BioMart/Ensembl

The screenshot shows the Ensembl BioMart interface. At the top, there is a navigation bar with links to BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. Below the navigation bar, there are three buttons: New, Count, and Results. On the right side of the header, there are links for URL, XML, Perl, and Help. The main area is divided into sections: Dataset, Filters, and Attributes. The Dataset section currently displays "Human genes (GRCh38.p13)". The Filters section shows "[None selected]". The Attributes section lists Gene stable ID, Gene stable ID version, Transcript stable ID, and Transcript stable ID version. A dropdown menu labeled "Ensembl Genes 108" is open, showing "Human genes (GRCh38.p13)" as an option. Orange arrows point from the text "Selection de la Base de donnée :" and "Sélection du jeu de données (génome)" to the "Ensembl Genes 108" dropdown menu.

- Selection de la Base de donnée :
- Genes
 - Variation
 - Regulation
 - Mouse strain

Sélection du jeu de données (génome)

In order to maintain service for all users, BioMart browser sessions running for more than 5 minutes are terminated. If you have queries that you think will run longer than this, please choose to have the results emailed to you.

Note that queries that run for longer than 6 hours will be terminated even when submitted this way. If this happens please reformat your query or contact us for details on how to approach this.

BioMart/Ensembl

The screenshot shows the Ensembl BioMart search interface. On the left, there's a sidebar with sections for Dataset (Human genes (GRCh38.p13)), Filters ([None selected]), Attributes (Gene stable ID, Gene stable ID version, Transcript stable ID, Transcript stable ID version), and another Dataset section ([None Selected]). The main area has dropdown menus for 'Dataset' (Ensembl Genes 108) and 'Values' (Human genes (GRCh38.p13)). A large orange box highlights the 'Attributes', 'Filters', 'Values', and 'Dataset' sections from the sidebar. An orange arrow points from the top of this box towards the highlighted area in the main search interface.

- 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)

Note that queries that run for longer than 6 hours will be terminated even when submitted this way. If this happens please reformat your query or contact us for details on how to approach this.

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