

Supports du cours

- Wiki:

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

- Aller à: Training / DNA-seq / DU Dijon
- Aller dans [Introduction to Ensembl/Biomart](#) et cliquer sur Hands-on

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

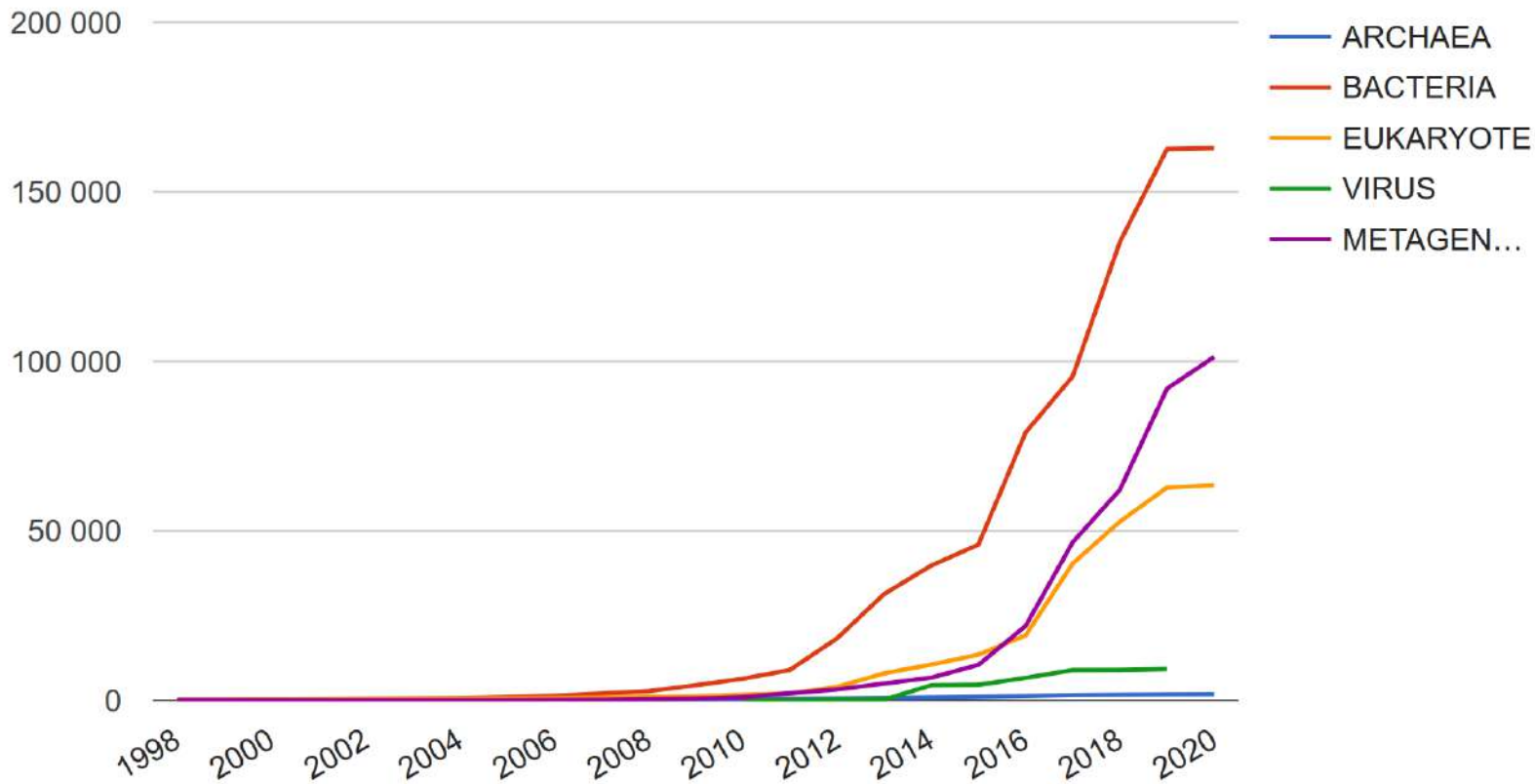
- 2002 Séquence préliminaire du génome de la souris (Waterston et al., 2002)
- 2004 ENCODE, Identification de tous les éléments fonctionnels du génome humain
- 2005 Roche 454: Séquenceur auto. haut-débit de 2ème génération par pyroséquençage : GS20
- 2007 Illumina/Solexa NGS de 2ème génération par synthèse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

Projects by Domain



Centres de bioinformatique

- EBI (European Bioinformatics Institute)



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

- NCBI (National Center for Biotechnology Information)



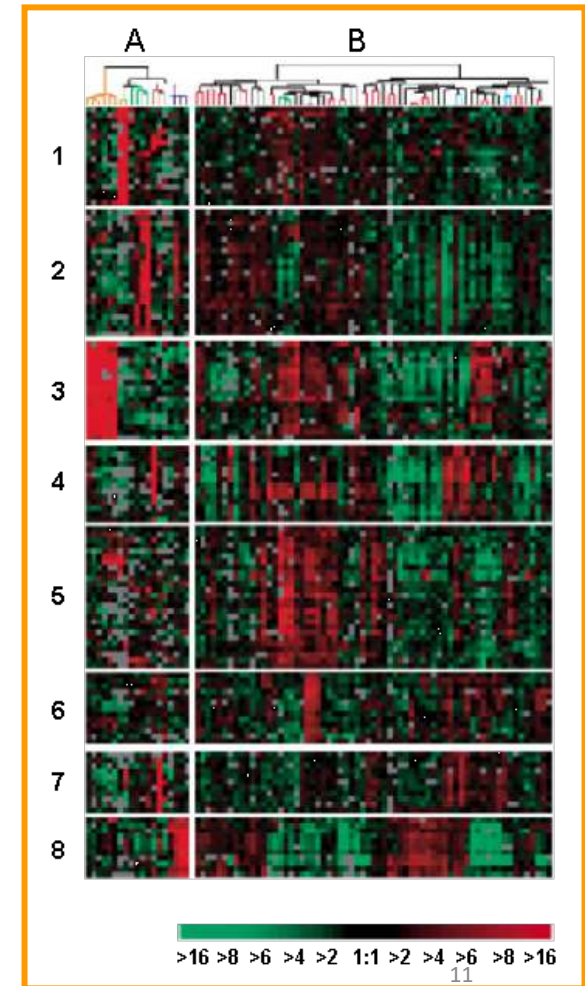
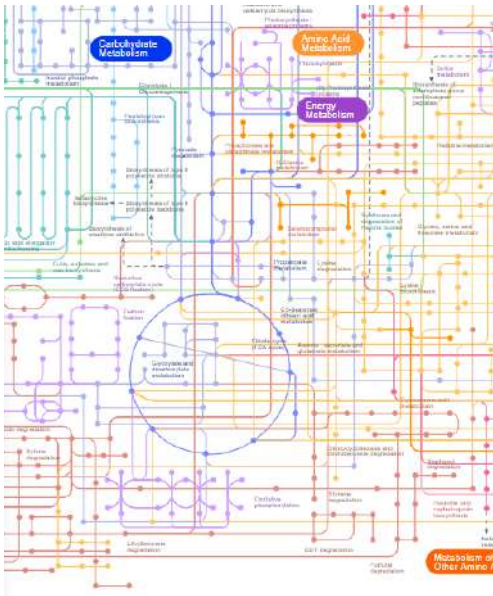
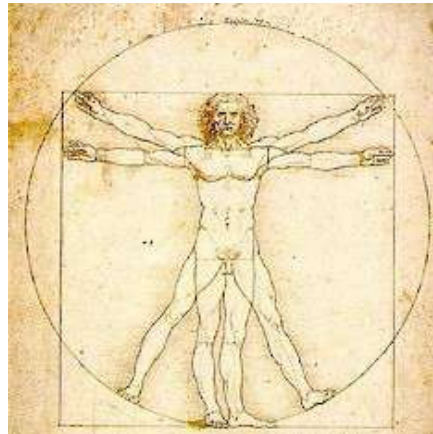
<http://www.ncbi.nlm.nih.gov/>

Banques de données en biologie moléculaire

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

Multiplicité des banques

MALWTRLRPLLALLALWPPPPARAFVNQHLGSHLVEALYLVCGERGFYTPKARREVEGPQVGALELAGGPGA



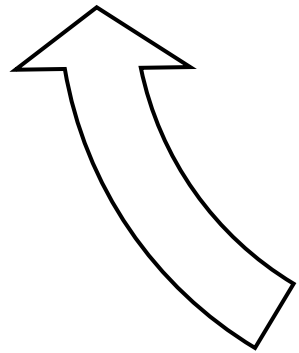
Banques de séquences nucléiques généralistes



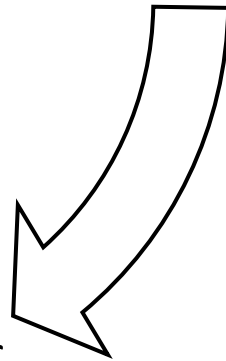
GenBank



EMBL



DNA
databank of
Japan



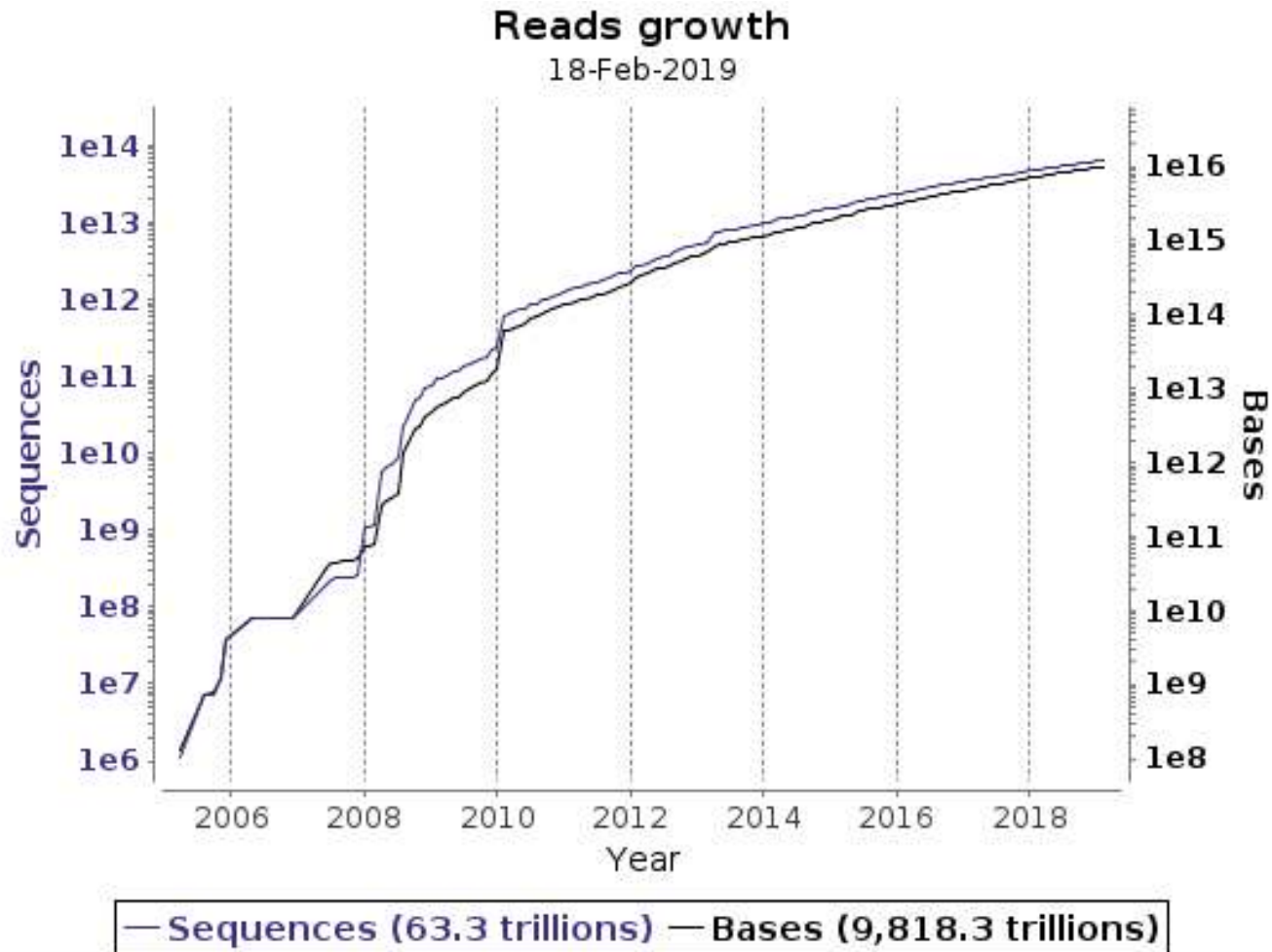
- 3 banques
- Échanges quotidiens des séquences collectées
- Effort d'unification=> format
 - accord entre GenBank et EMBL en 1986
 - accord entre GenBank/EMBL et DDBJ in 1987



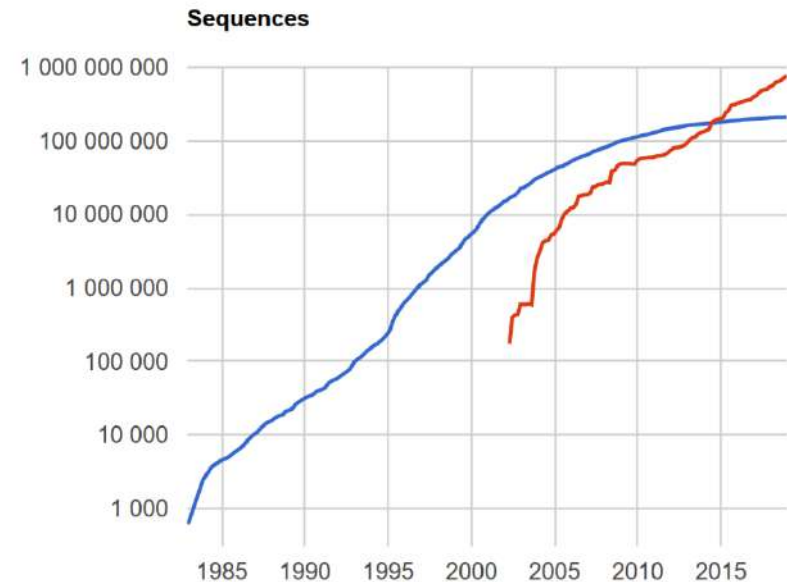
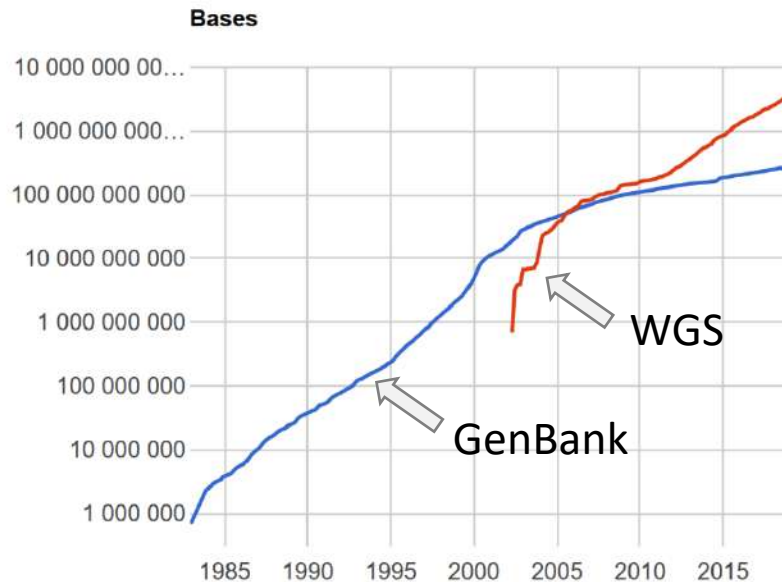
Banques de séquences

- 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

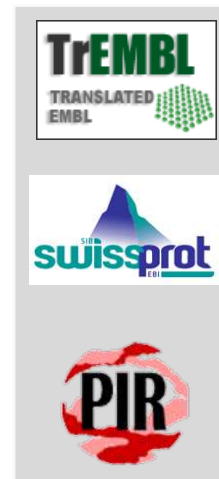
Banques de séquences protéiques généralistes



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

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

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



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

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

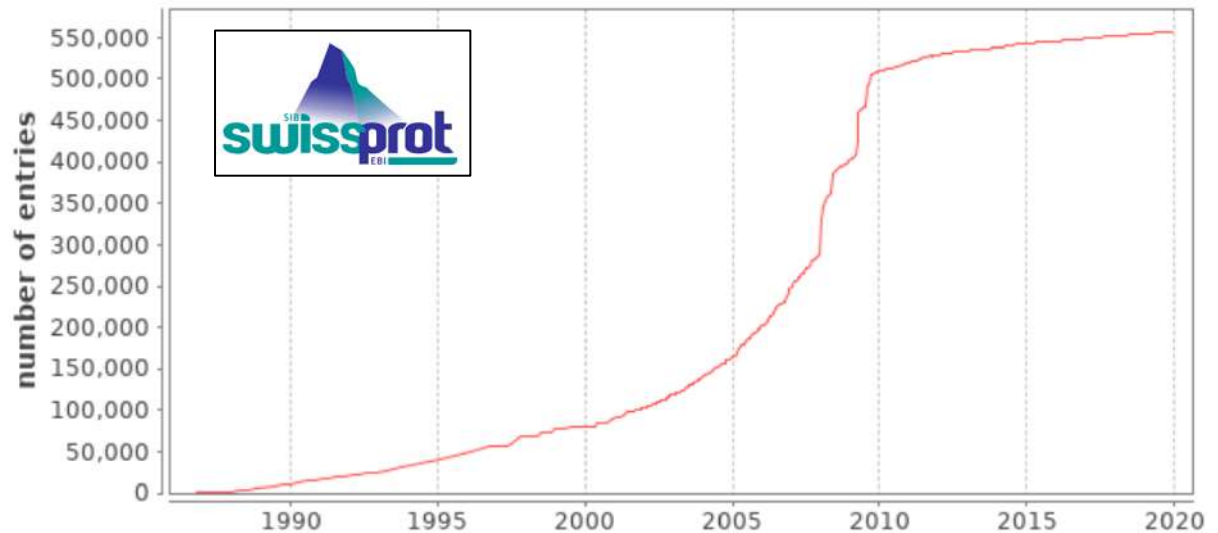
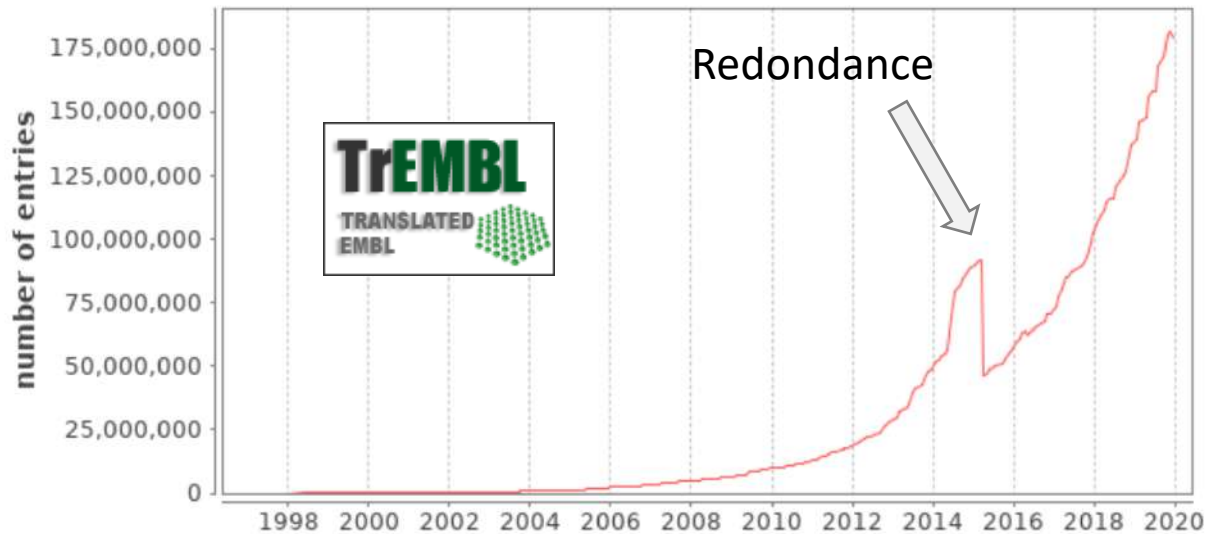
TrEMBL:
179,250,561 entrées

Swiss-Prot:
561,568 entrées

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

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

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 »

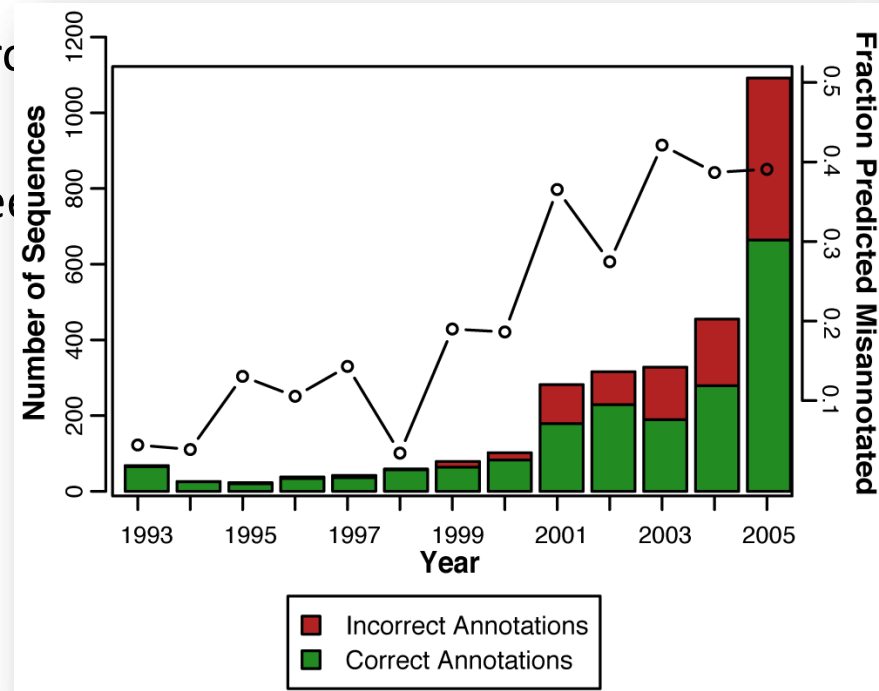
Exemple de l'importance de l'annotation

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

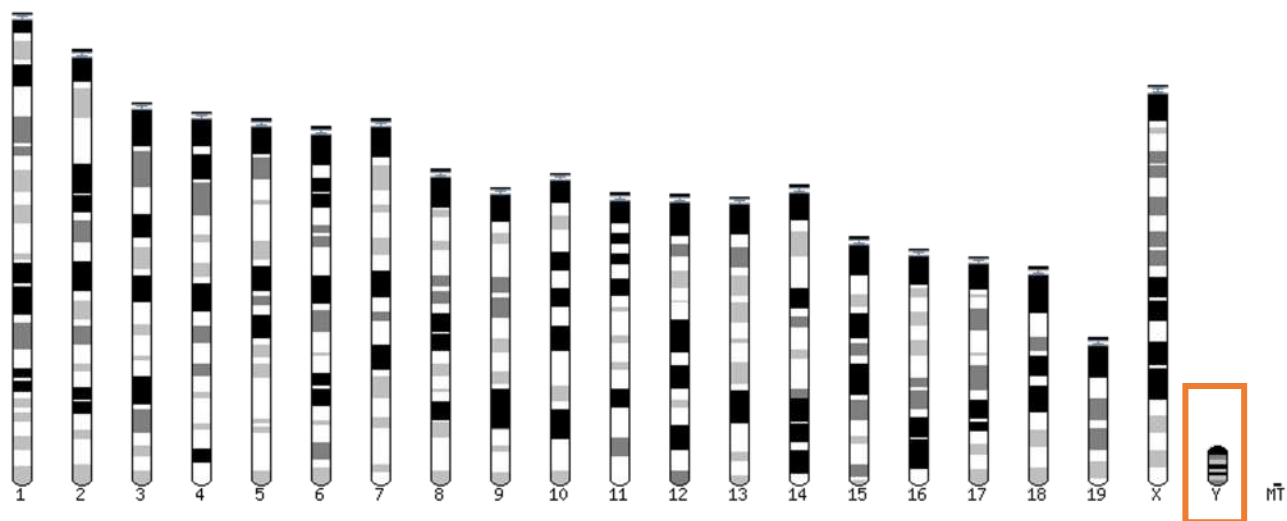
bases avec les Raw data vs Annotation

Ces données seront agrégées sur le génome humain

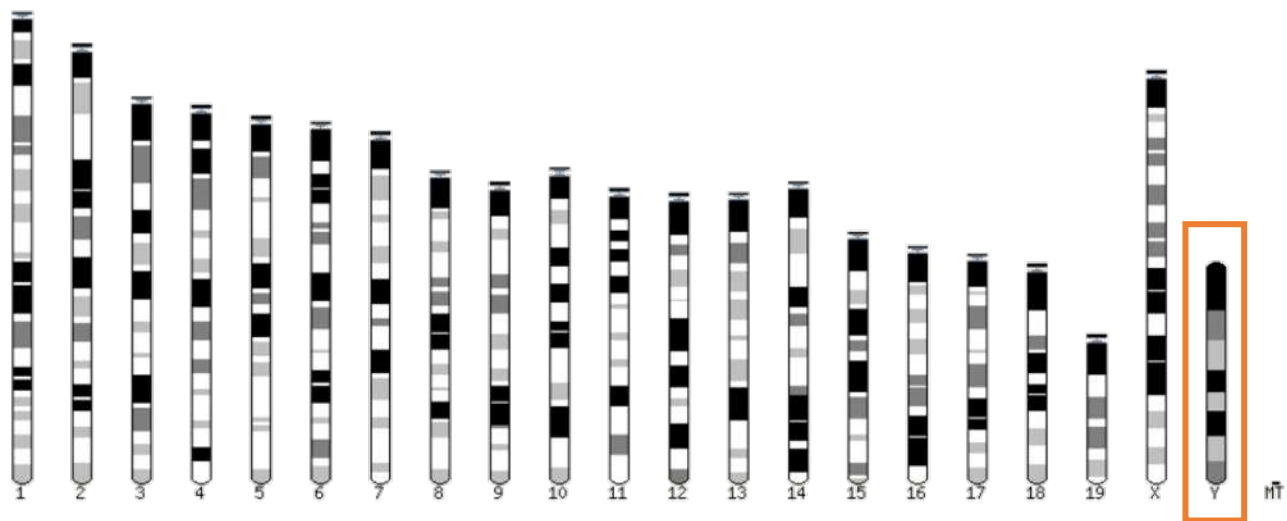
Genome browsers

Genome builds

mm9



mm10



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)

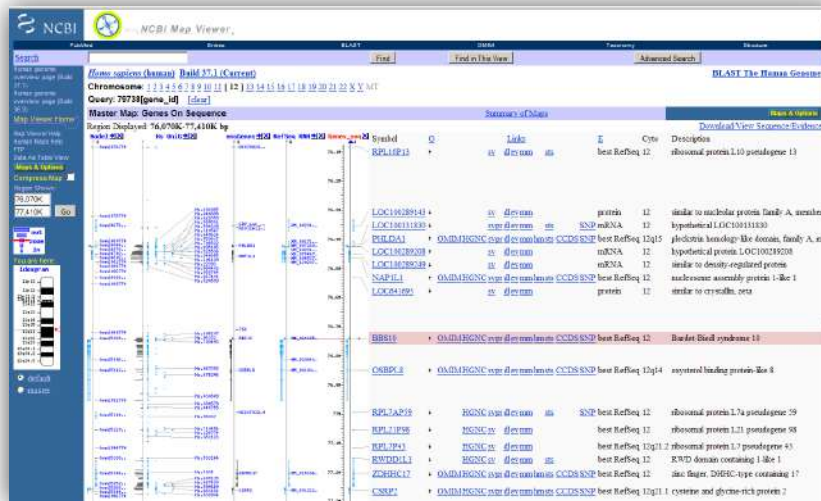
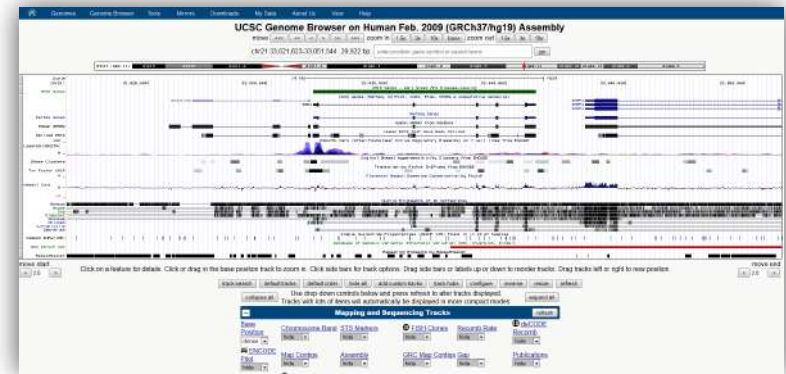
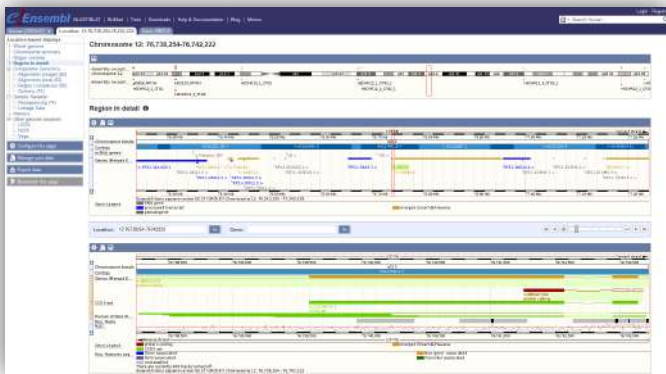
Genome Browsers – L'outil de référence

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

II y a Genome Browsers...

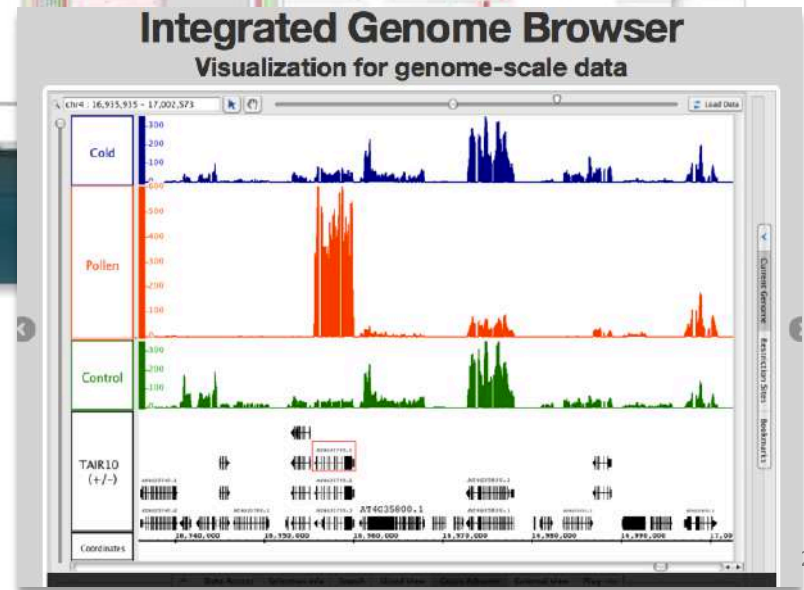
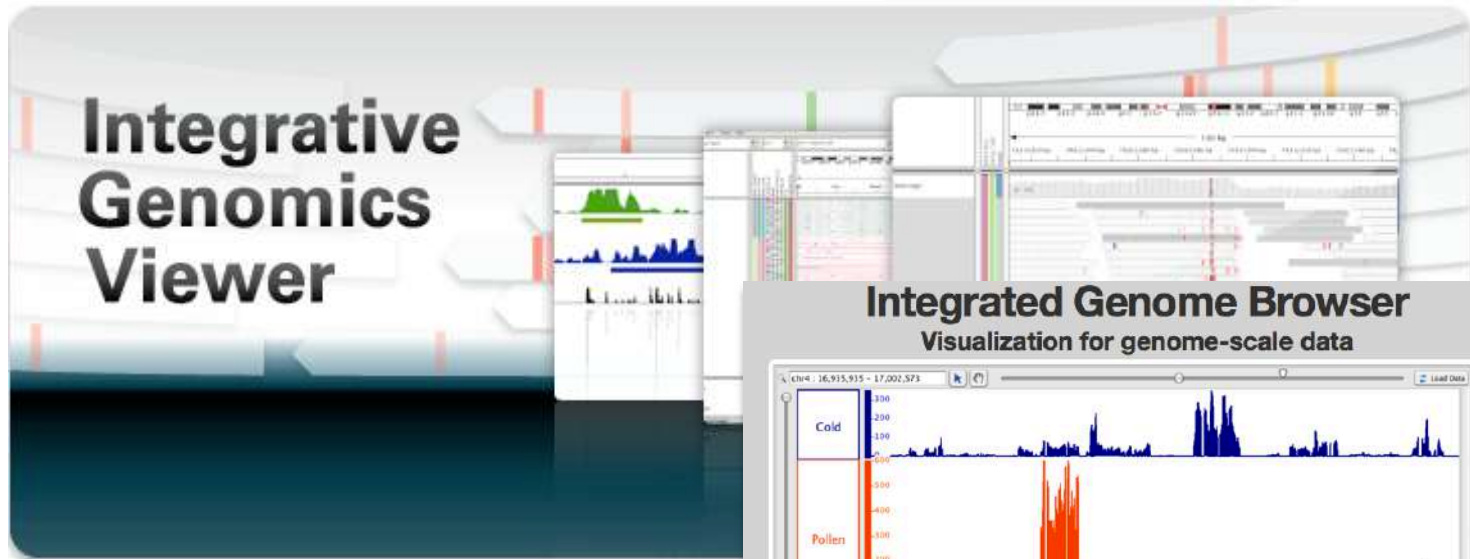
EBI - Ensembl

UCSC – Genome Browser



NCBI – Map Viewer

Et Genome browsers



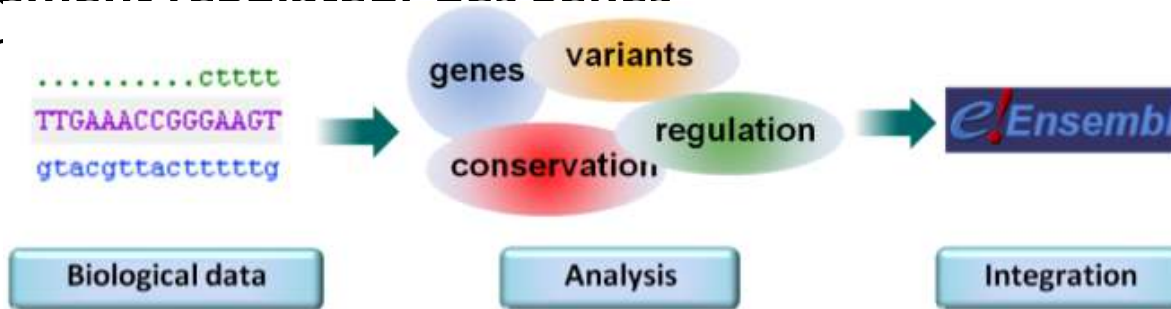
Ensembl

Le projet Ensembl

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

Le projet Ensembl

- Données disponibles :
 - Génomes
 - Données de génomique comparative
 - Variations
 - 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

The screenshot displays the Ensembl genome browser interface. At the top, the Ensembl logo is followed by navigation links: BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar on the right contains the text "Search all species...".

Tools

- BioMart >**
All tools: 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 [v] for [input] [Go]
e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes

-- Select a species -- [v]

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 Release 105 (Dec 2021)

- Updated allele frequency data from the NCBI Allele Frequency Aggregator (ALFA) release 2.
- Update to the Variant Recoder supporting MANE annotation and variant names in external databases
- Dog (Canis lupus familiaris) reference genome has changed from CanFam3.1 to ROS Cfam 1.0 Labrador retriever.
- Support for BCF files

[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

- 19 Jan 2022: [Update to the Ensembl COVID-19 resource](#)
- 12 Jan 2022: [Homology data available in Ensembl Rapid Release](#)
- 17 Dec 2021: [146 new insect genomes on Ensembl Rapid Release](#)

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.

elixir European Data Resource

Ensembl release 105 - Dec 2021 © [EMBL-EBI](#) [Permanent link](#) - [View in archive site](#)

Comprendre ENSEMBL

Les annotations

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

Les annotations

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

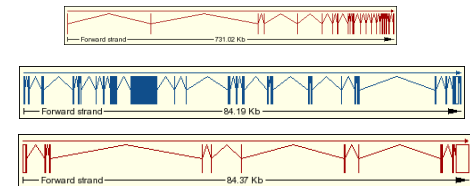


Protéine/ ARNm

+



Assemblage de séquence

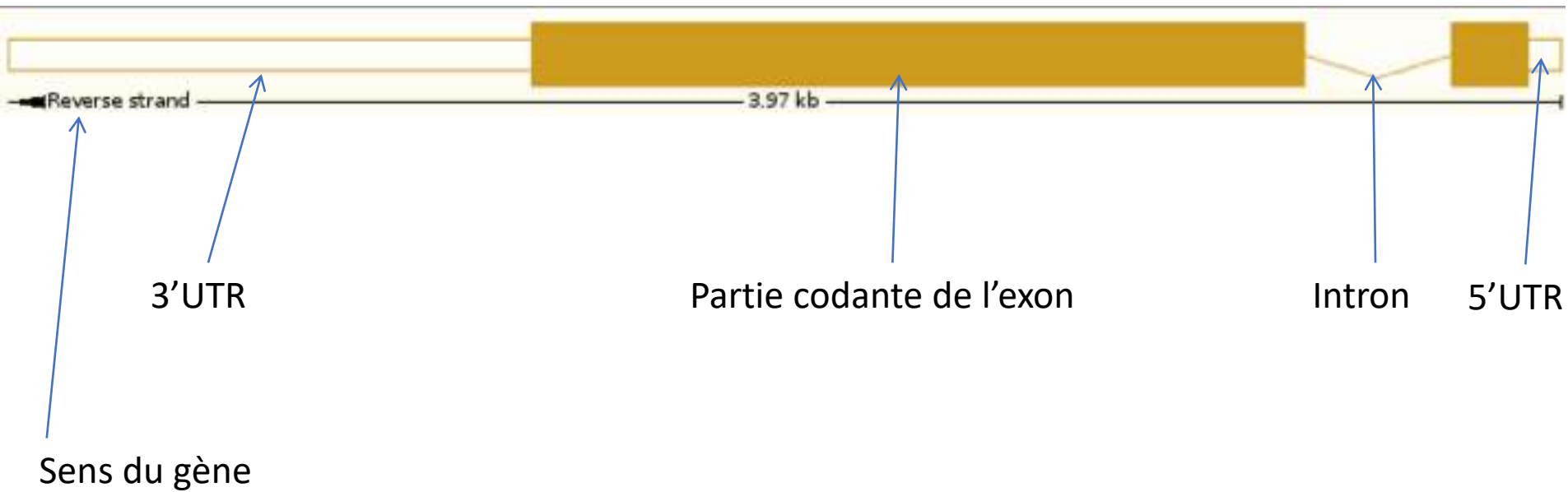


Gène Ensembl

Les annotations

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

Transcrits Ensembl



Identifiants Ensembl

- ENSG### Ensembl Gene ID
 - ENST### Ensembl Transcript ID
 - ENSP### Ensembl Peptide ID
 - ENSE### Ensembl Exon ID
-
- Ajout d'un suffix pour les autres espèces
 - MUS (*Mus musculus*) pour la souris: ENSMUSG###
 - DAR (*Danio rerio*) pour le zebrafish: ENSDARG###
 - 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



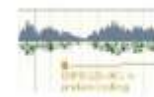
Retrieve gene sequence

```
GGCTGACTTCGGGGGG  
GGCTTGTGGGGGAGC  
GGCTTCTGCTGGGGC  
AAGGGACAGATTGTGA  
GAGCTCTGGAGGGGTT  
GGGATTCGAGGGGGC
```

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](https://www.ebi.ac.uk) 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.



- Lien vers une version particulière d'Ensembl : <http://Dec2021.archive.ensembl.org/index.html>

Ensembl : Archives

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | More ▾ Login/Register

Search all species... 🔍


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

In this section
Archives: Table of assemblies

Search documentation

Ensembl Archives

About Archive Ensembl



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

- Annotation on the **human NCBI36 assembly** is available at our [Ensembl 54 archive](#) site.
- Annotation on the **mouse NCBI m37 assembly** is available at our [Ensembl 67 archive](#) site.
- As from August 2014 we are supporting the **human GRCh37 assembly** at our dedicated [GRCh37 human](#) site. Unlike the other Ensembl archive sites, this will be updated to the latest web interface every Ensembl release and there may be occasional data updates to human.

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

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

Notes

- Ensembl aims to maintain stable identifiers for genes (ENSG), transcripts (ENST), proteins (ENSP) and exons (ENSE) as long

List of currently available archives

- [Ensembl GRCh37](#): Full Feb 2014 archive with BLAST, VEP and BioMart
- [Ensembl 105: Dec 2021](#) - this site
- [Ensembl 104: May 2021](#)
- [Ensembl 103: Feb 2021](#)
- [Ensembl 102: Nov 2020](#)
- [Ensembl 101: Aug 2020](#)
- [Ensembl 100: Apr 2020](#)
- [Ensembl 99: Jan 2020](#)
- [Ensembl 98: Sep 2019](#)
- [Ensembl 97: Jul 2019](#)
- [Ensembl 96: Apr 2019](#)
- [Ensembl 95: Jan 2019](#)
- [Ensembl 94: Oct 2018](#)
- [Ensembl 93: Jul 2018](#)
- [Ensembl 92: Apr 2018](#)
- [Ensembl 91: Dec 2017](#)
- [Ensembl 90: Aug 2017](#)
- [Ensembl 89: May 2017](#)
- [Ensembl 88: Mar 2017](#)
- [Ensembl 87: Dec 2016](#)
- [Ensembl 86: Oct 2016](#)
- [Ensembl 80: May 2015](#)
- [Ensembl 77: Oct 2014](#)
- [Ensembl 75: Feb 2014](#)
- [Ensembl 54: May 2014](#)

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

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

Ensembl : Archives

The screenshot shows the Ensembl website interface. At the top, there is a navigation bar with the 'Archive! Ensembl' logo, links for 'BioMart', 'Downloads', 'Help & Docs', and 'Blog', and a search bar for 'Search all species...'. Below the navigation bar, there are several sections:

- Tools:** A section titled 'Tools' with a link to 'BioMart >' and a sub-link 'All tools'. Below this is a description: 'Export custom datasets from Ensembl with this data-mining tool'.
- Search:** A search box with a dropdown menu set to 'All species' and a 'Go' button. Below the search box, there are examples: 'e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease'.
- All genomes:** A section with a dropdown menu set to '-- Select a species --'. Below this, there are three entries: 'Pig breeds' (with a pig image), 'Human' (with a human image), and 'Mouse' (with a mouse image). There is also a 'View full list of all species' link.
- Favourite genomes:** A section with a pencil icon and three entries: 'Human' (GRCh38.p13), 'Mouse' (GRCm39), and 'Zebrafish' (GRCz11). There is a link 'Still using GRCh37?'.
- Ensembl Archive Release 104 (May 2021):** A section with a list of updates: 'Update to the Ensembl Canonical transcript set.', 'Human and mouse gene sets updated to GENCODE 38 and GENCODE M27, respectively.', and 'Retirement of gene names derived from BAC clones.' There is a link 'More release news' on our blog.
- Ensembl Rapid Release:** A section with a blue header and a 'Go' button. It contains the text: 'New assemblies with gene and protein annotation every two weeks.' and 'Note: species that already exist on this site will continue to be updated with the full range of annotations.'

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>

Table of Assemblies

This table lists the assembly per species in each of the archive sites.

Key: New species or Species present in archive Species not in this version of Ensembl

	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
Abingdon Island giant tortoise	ASM359739v1																
African ostrich	ASM69896v1																
Agassiz's desert tortoise	ASM289641v1																
Algerian mouse	SPRET_EIJ_v1																
Alpaca	vicPact																
Alpine marmot	marMar2.1																
Amazon molly	Poecilia_formosa-5.1.2																
American beaver	C.san_genome_v1.0																
American bison	Bison_UMD1.0																
American black bear	ASM334442v1																
American mink	NNQG3.v01																
Angola colobus	Ceng_pa_1.0																
Arabian camel	CamDro2																
Arctic ground squirrel	ASM342682v1																
Arasatins bleak and white tegu	HltupMor3																
Armadillo	Dasnov3.0																
Asian bonytongue	ISdFor1.1									ASM162426v1							
Asiatic black bear	ASM966005v1																
Atlantic cod	gadMor3.0			gadMor1													
Atlantic herring	Ch_v2.0.2																
Atlantic salmon	IC3ASG_v2																
Australian saltwater crocodile	CroPor_comp1																
Ballan wrasse	BallGen_V1																
Barramundi perch	ASB_HGAPassembly_v1																
Beluga whale	ASM228892v3																

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

The screenshot shows the Ensembl genome browser homepage. At the top, there is a navigation bar with the Ensembl logo and links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located on the right with the text "Search all species...". Below the navigation bar, there are four main tool categories: Tools (with a link to "All tools"), BioMart (for exporting custom datasets), BLAST/BLAT (for searching genomes), and Variant Effect Predictor (for analyzing variants). A central search box allows users to search by species or genomic coordinates, with examples like "BRCA2" or "rat 5:62797383-63627669". Below this, there are sections for "All genomes" (with a species selector) and "Favourite genomes" (listing Human, Mouse, and Zebrafish). On the right side, there is a "Ensembl Release 105 (Dec 2021)" section with a list of updates, followed by an "Ensembl Rapid Release" section highlighting new assemblies. At the bottom, there are several utility boxes for comparing genes, finding SNPs, gene expression, and retrieving sequences. The footer contains information about EMBL-EBI funding and the Elixir logo.

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

Tools [All tools](#)

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

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

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

Search

All species for

e.g. **BRCA2** or **rat 5:62797383-63627669** 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 Release 105 (Dec 2021)

- Updated allele frequency data from the NCBI Allele Frequency Aggregator (ALFA) release 2
- Update to the Variant Recoder supporting MANE annotation and variant names in external databases
- Dog (*Canis lupus familiaris*) reference genome has changed from CanFam3.1 to ROS Clam 1.0 Labrador retriever
- Support for BCF files

[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

- 19 Jan 2022: [Update to the Ensembl COVID-19 resource](#)
- 12 Jan 2022: [Homology data available in Ensembl Rapid Release](#)
- 17 Dec 2021: [146 new insect genomes on Ensembl Rapid Release](#)

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.

elixir Life Data Resource

Ensembl release 105 - Dec 2021 © [EMBL-EBI](#) [Permanent link](#) - [View in archive site](#)

Ensembl Genomes

Bactéries

The screenshot shows the Ensembl Bacteria homepage. At the top, there is a search bar for species and a navigation menu with links for Home, BLAST, Tools, Downloads, and More. Below the search bar, there are sections for 'Search for a gene' and 'Search for a genome'. A 'What's New in Release 52' section highlights updates such as the addition of 13,252 bacterial and archaeal genomes and the inclusion of protein features for all species using InterProScan. A 'Did you know...' box provides information on how to access Ensembl Genomes data from programming languages like Perl and Python. The 'Archive sites' section lists previous releases from December 2020 to July 2019.

Plantes

The screenshot displays the Ensembl Plants homepage. It features a search bar for species and a navigation menu. The 'What's New in Release 52' section reports the addition of 98,276 high confidence genes from the T2ACv1 assembly and the inclusion of 15 wheat outgroups from the 12x pan-genome project. It also mentions the addition of KASP markers and the inclusion of 18,000 pan-taxonomic gene trees and homologs. The 'Archive sites' section lists releases from December 2020 to July 2019.

Protistes

The screenshot shows the Ensembl Protists homepage. It includes a search bar for species and a navigation menu. The 'What's New in Release 52' section highlights updates such as the addition of 10,000 new protein features for all species using InterProScan and the inclusion of 10,000 pan-taxonomic gene trees and homologs. The 'Archive sites' section lists releases from December 2020 to July 2019.

Fungi

The screenshot displays the Ensembl Fungi homepage. It features a search bar for species and a navigation menu. The 'What's New in Release 52' section reports the addition of 477 new genomes imported from JGI and the inclusion of 15 genomes imported from MycoBank. It also mentions the addition of protein features for all species using InterProScan and the inclusion of 10,000 pan-taxonomic gene trees and homologs. The 'Archive sites' section lists releases from December 2020 to July 2019.

Métazoaires

The screenshot shows the Ensembl Metazoa homepage. It includes a search bar for species and a navigation menu. The 'What's New in Release 52' section highlights updates such as the addition of 10,000 new protein features for all species using InterProScan and the inclusion of 10,000 pan-taxonomic gene trees and homologs. The 'Archive sites' section lists releases from December 2020 to July 2019.

Le site web Ensembl: page d'accueil

The screenshot shows the Ensembl website homepage. At the top, there is a navigation bar with links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located in the top right corner. Below the navigation bar, there are several sections: Tools, BioMart, BLAST/BLAT, and Variant Effect Predictor. A central search box is present, followed by a section for All genomes and Favourite genomes. On the right side, there are sections for Ensembl Release 105 (Dec 2021), Ensembl Rapid Release, and Other news from our blog. At the bottom, there are several utility boxes for comparing genes, finding SNPs, gene expression, and retrieving gene sequences. The footer contains information about EMBL-EBI and a permanent link to the archive site.

Outils

Recherche

Recherche

Liste déroulante Accès aux génomes

Recherche

News

Accès aux archives d'Ensembl

Le site web Ensembl: statistiques des génomes

The screenshot displays the Ensembl website interface for the Human genome (GRCh38.p13). The page is titled "Human assembly and gene annotation" and is divided into several sections:

- Assembly:** Provides information about the December 2013 *Homo sapiens* high coverage assembly GRCh38 from the Genome Reference Consortium. It includes details about the data set, properties of the assembly (contig length, chromosome length), and the inclusion of alternate sequence representations.
- Patches:** Discusses the introduction of assembly patches as the GRC maintains and improves the assembly.
- Gene annotation:** Explains that Ensembl's automatic annotation pipeline has been updated to incorporate new protein and cDNA sequences. It also mentions the merge of manual annotations from Havana and the current release's joint gene set.
- Neanderthal genome:** Mentions the availability of the Neanderthal (*Homo sapiens neanderthalensis*) genome via the Neanderthal Genome Browser.

On the right side, there are two main sections:

- Statistics Summary:** A table providing key metrics for the assembly, such as Base Pairs, Golden Path Length, and Assembly provider.
- Gene counts:** Two tables showing the number of coding genes, non-coding genes, and pseudogenes for both the primary assembly and an alternative sequence.

Navigation and search options are visible at the top, including links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog, along with a search bar and a Login/Register link.

Statistiques

Informations générales sur l'assemblage

Le site web Ensembl: caryotype

The screenshot shows the Ensembl genome browser interface for the Human (GRCh38.p13) genome. The main display area features a caryotype of the human genome, with chromosomes arranged in pairs from 1 to 22, plus the X and Y chromosomes. Below the caryotype is a summary table providing key genomic statistics.

Summary

Assembly	GRCh38.p13 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.28 v5, 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	Aug 2021
Database version	105.38
Gencode version	GENCODE 39

Gene counts (Primary assembly)

Coding genes	20,465 (incl 653 readthrough)
Non coding genes	24,849
Small non coding genes	4,865
Long non coding genes	17,763 (incl 308 readthrough)
Misc non coding genes	2,221
Pseudogenes	15,217 (incl 6 readthrough)

Le site web Ensembl : statistiques par chromosome

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

Human (GRCh38.p13)

Location: 1:204,088,063-204,188,063

Chromosome 1: 204,088,063-204,188,063

Assembly exceptions chromosome 1

Chromosome summary

+ Add features

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

Click on the image above to zoom into that point

Change chromosome: Go

Chromosome Statistics

Length (bps)	248,956,422
Coding genes	2,057 (incl. 47 readthrough)

Le site web Ensembl : navigateur de génome

The screenshot displays the Ensembl genome browser interface for Human (GRCh38.p13). The main view shows a genomic region on Chromosome 1, specifically the coordinates 204,088,063 to 204,188,063. The interface includes a navigation sidebar on the left with options like 'Whole genome', 'Chromosome summary', and 'Region in detail'. The main content area is titled 'Chromosome 1: 204,088,063-204,188,063' and features a 'Region in detail' section. This section displays a genomic track with various annotations, including genes (e.g., *g32.1*, *AC114402.2*, *AC096446.2*, *AL92148.8*, *AL992114.13*, *AL906486.28*, *AL512306.16*), regulatory elements (CTCF, Open Chromatin, Promoter Flank, Enhancer, Promoter, Transcription Factor Binding Site), and gene legends. A red box highlights a specific region within the track. Below the main track, there is a search bar for 'Location' and 'Gene', and a zoomed-in view of the region showing '91 way GERP elements', 'Human cDNAs (RefSeq/ENA)', and 'EST cluster (Unigene)'. The interface also includes a 'Configure this page' button in the sidebar, which is circled in red.

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 Jobs ▼

Gene: BRCA2 ENSG00000139618

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

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

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

About this gene
This gene has 10 transcripts (splice variants), 175 orthologues and is associated with 171 phenotypes.

Transcripts
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_000069.4	MANE Select v0.95 Ensembl Canonical GEN
ENST00000680687.1	BRCA2-210	11880	3418aa	Protein coding	CCDS9344	-	-	APPRIS P1
ENST00000544455.6	BRCA2-206	11854	3418aa	Protein coding	CCDS9344	P51587	-	GENCODE basic APPRI
ENST00000530893.6	BRCA2-204	2011	481aa	Protein coding	-	A0A590UJ17	-	TSL:1 CDS 3' incomp
ENST00000614259.2	BRCA2-207	11763	2649aa	Nonsense mediated decay	-	-	-	TSL:2
ENST00000665585.1	BRCA2-208	2598	438aa	Nonsense mediated decay	-	A0A590UJU6	-	CDS 5' incomp
ENST00000470094.1	BRCA2-202	842	186aa	Nonsense mediated decay	-	H0YE37	-	TSL:5 CDS 5' incomp
ENST00000666593.1	BRCA2-209	523	58aa	Nonsense mediated decay	-	A0A590UJ24	-	CDS 5' incomp
ENST00000528762.1	BRCA2-203	495	64aa	Nonsense mediated decay	-	H0YD86	-	TSL:4 CDS 5' incomp
ENST00000533776.1	BRCA2-205	523	No protein	Retained intron	-	-	-	TSL:3

Summary

Name
BRCA2 (HGNC Symbol)

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

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

RefSeq
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

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

Ensembl version
ENSG00000139618.17

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

Annotation method
Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article

Annotation Attributes
overlapping locus Definitions

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

Le site web Ensembl : le transcript

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

Human (GRCh38.p13) Search all species...

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

Transcript-based displays

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

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

Transcript: ENST00000380152.8 BRCA2-201

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

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

Location [Chromosome 13: 32,315,508-32,400,268 forward strand.](#)

About this transcript This transcript has [27 exons](#), is annotated with [58 domains and features](#), is associated with [35198 variant alleles](#) and maps to [935 oligo probes](#).

Gene This transcript is a product of gene [ENSG00000139618.17](#) [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 v0.95 Ensembl Canonical GENCODE basic APPRIS P1
ENST00000680887.1	BRCA2-210	11880	3418aa	Protein coding	CCDS9344	-	-	APPRIS P1
ENST00000544455.6	BRCA2-206	11854	3418aa	Protein coding	CCDS9344	P51587	-	GENCODE basic APPRI
ENST00000530893.6	BRCA2-204	2011	481aa	Protein coding	-	A0A590UJ17	-	TSL:1 CDS 3' inco
ENST00000614259.2	BRCA2-207	11763	2649aa	Nonsense mediated decay	-	-	-	TSL:2 CDS 5' incomp
ENST00000665585.1	BRCA2-208	2598	438aa	Nonsense mediated decay	-	A0A590UJU6	-	CDS 5' incomp
ENST00000470094.1	BRCA2-202	842	186aa	Nonsense mediated decay	-	H0YE37	-	TSL:5 CDS 5' inco
ENST00000666593.1	BRCA2-209	523	56aa	Nonsense mediated decay	-	A0A590UJ24	-	CDS 5' incomp
ENST00000528762.1	BRCA2-203	495	64aa	Nonsense mediated decay	-	H0YD86	-	TSL:4 CDS 5' inco
ENST00000533776.1	BRCA2-205	523	No protein	Retained intron	-	-	-	TSL:3

Summary

BRCA2-201 > protein coding

84.76 kb Forward strand

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

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

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

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

Version ENST00000380152.8

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 105 - Dec 2021 © [EMBL-EBI](#) [Permanent link](#) - [View in archive site](#)

Naviguer dans Ensembl : Partie pratique

Visualiser ses propres données

The screenshot shows the Ensembl genome browser homepage. At the top, there is a navigation bar with links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located on the right with the text "Search all species...". Below the navigation bar, there are several tool links: Tools, BioMart, BLAST/BLAT, and Variant Effect Predictor. A search box is present with a dropdown menu for "All species" and a "Go" button. Below the search box, there are sections for "All genomes" and "Favourite genomes" with a dropdown menu for "Select a species". The "Favourite genomes" section lists Human (GRCh38.p13), Mouse (GRCm39), and Zebrafish (GRCz11). On the right side, there is a "Ensembl Release 105 (Dec 2021)" section with a list of updates. Below that is an "Ensembl Rapid Release" section with a "Go" button. At the bottom, there are several quick links: "Compare genes across species", "Find SNPs and other variants for my gene", "Gene expression in different tissues", "Retrieve gene sequence", "Find a Data Display", and "Use my own data in Ensembl". A callout box with an orange border points to the "Use my own data in Ensembl" button, containing the text "Visualiser ses propres données".

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 105 (Dec 2021)

- Updated allele frequency data from the NCBI Allele Frequency Aggregator (ALFA) release 2
- Update to the Variant Recoder supporting MANE annotation and variant names in external databases
- Dog (*Canis lupus familiaris*) reference genome has changed from CanFam3.1 to ROS Clam 1.0 Labrador retriever
- Support for BCF files

[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

- 19 Jan 2022: [Update to the Ensembl COVID-19](#)
- 12 Jan 2022: [Homology data available in Ensembl Release](#)
- 17 Dec 2021: [146 new insect genomes on Ensembl Release](#)

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.

elixir European Data Infrastructure

Ensembl release 105 - Dec 2021 © [EMBL-EBI](#)

Navigation dans Ensembl

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

Visualiser ses propres données

LES OUTILS

Les outils

BLAST/BLAT

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

Tools

[All tools](#)

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

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

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

Search

All species for

e.g. **BRCA2** or **rat 5:62797383-63627669** 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](#)

Other news from our blog

- 19 Jan 2022: [Update to the Ensembl COVID-19 resource](#)
- 12 Jan 2022: [Homology data available in Ensembl Rapid Release](#)
- 17 Dec 2021: [146 new insect genomes on Ensembl Rapid Release](#)

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.

Elixir
Life Data Resource

Ensembl release 105 - Dec 2021 © [EMBL-EBI](#) [Permanent link](#) - [View in archive site](#)

Blast



- Recherche de similarité

- 1 séquence (**Query**) comparée à des milliers ou des millions de séquences (**base de données**) par comparaison 2 à 2.

- But:

- Détecter des séquences proches
- Annotation simple (domaines protéiques, localisation génomique, nombre d'exons)

Les différentes comparaisons

BLAST : Basic Local Alignment Search Tool

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

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

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

Les différentes comparaisons

BLAT (BLAST-Like Alignment Tool)

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

Blast



```
MADTQYILPNDIGVSSLDCREAFRLLSPTERLYAYHLSRAAWYGGLAVLLQTSPEAPYIYALLSRLFRAQDP
DQLRQHALAEGLTEEEYQAFLVYAAGVYSNMGNYSFGDTKFVPNLPKEKLERVILGSEAAQQHPPEEVRGLW
QTCGELMFSLEPRLRHLGLGKEGITTYFSGNCTMEDAKLAQDFLDSQNLSAYNTRLFKEVDGEGKPYEVR
ASVLGSEPSLDSEVTSKLLKSYEFRGSPFQVTRGDYAPILQKVVEQLEKAKAYAANSHQGQMLAQYIESFTQG
SIEAHKRGSRFWIQDKGPIVESYIGFIESYRDPFGSRGEFEGFVAVVVKAMSAKFERLVASAEQLLKELPWP
PTFEKDKFLTPDFTSLDVLTFFAGSGIPAGINIPNYDDLQTEGFKNVSLGNVLAVAYATQREKLTFLLEDDK
DLYILWKGPSFDVQVGLHELLGHGSGKLFVQDEKGA FNFDQETVINPETGEQIQSWYRSGETWDSKFSTIAS
SYEECRAESVGLYLCLHPQVLEIFGFEGADAEDVIYVNWLNMV RAGLLALEFYTPEAFNWRQAHMQARFVIL
RVLLEAGEGLVTITPTTGSDGRPDARVRLDRSKIRSVGKPALERFLRRLQVLKSTGDVAGGRALYEGYATVT
DAPPECFLTTLRDTVLLRKE SRKLIVQPNTRLEGS DVQLLEYEASAAGLIRSFSE RFPEDGPELEEILTQLAT
ADARFWKGPSEAPSGQA
```

new SETUP CONFIG RESULTS DISPLAY refresh Online Help

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_IESYdDXDJ Retrieve

Alignment Display Options:

Locations vs. Karyotype Locations vs. Query

Summary Table

1: unnamed (737 letters) Vs. LATESTGP

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

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

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

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

Query= unnamed (737 letters)

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

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

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

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

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

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

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

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

Query: 477 INPETGEOIQSWYRSGETWDSKFSSTIASSYEECRAESVGLYLCLHPOVLEIFGFEGADAF 536
 INPE EQIQSWYRS +TWDKSFSTI SSYEECRAESVGLYLCLHPOVLE FGFEGADAE
 Sbjct: 76090065 INPEMREIQSWYRSKMTWDSKFSSTIVSSYEECRAESVGLYLCLHPOVLETFGFEGADAE 76089886

Query: 537 DVIYVNWLNVMVRAGLLALEFYTPEAFNWRQAHMOARFVILRVLEAGEGLVITPTTGSD 596
 +VI VNWLNVM AGLLALEFYTPEA NW+QAH++AR VILRVL EAGEGL TITPT GSD
 Sbjct: 76089885 EVISVNWLNVMVGAGLLALEFYTPEASNWQAHIRARIVILRVLPEAGEGLGITPTAGSD 76089706

Query: 597 GRPDARVLRDRSKIRSVGKPALERFLRRLLOVLKSTGDVAGGRALYEGYATVTDAPPECFI 656
 GRP+A+VRLDRSKI+SVG PALERFLRR STGDVAGG LYE YA V DAPPE FL
 Sbjct: 76089705 GRPEAQVLRDRSKIOSVGNPALERFLRRCW---STGDVAGGWILYERYAAVADAPPEGFL 76089535

Query: 657 TLRDTVLLRKESRKLIVQPNTRLEGS DVOLLEYEASAAGLIRSFSEFPEDGPELEELIT 716
 TLRD VLLRKES KLIVQPN RLEGS DVOLLEYE SAAGLIRSFSE FPEGD ELE+ILT
 Sbjct: 76089534 TLRDRVLLRKESWKLIVQPNTRLEGS DVOLLEYEVAAGLIRSFSEHFPEDGLELEDIL 76089355

Query: 717 QLATADARFWKGPSEAPSGOA 737
 QLATADA+F KGPSEAPSGOA
 Sbjct: 76089354 QLATADAOF*KGPEAPSGOA 76089292

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

Query: 224 PSLDSEVTSKLLKSYEFRGSPFQVTRGDYAPILQKVVEQLEKAKAYAANSHQQLAQYIE 283
 P L + SKLKS EFRGSPFQVT G+Y PILQKVVEQLEKAK YAANSHQ QMLAQYIE
 Sbjct: 76090816 PGLRGD--SKLKS*EFRGSPFQVWGNVMPILQKVVEQLEKAKTYAANSHQEQMLAQYIE 76090643

Query: 284 SFTQGSIEAHKGRSFRWFIQDKGPVIVESYIGFIESYRDPFGSRGFEFVAVVNKAMSAKF 343
 SFTQGS EAHK+GSRFWI DKGPIVESYI FI+SYRD FGSRG EGFVAVVNKAMSAK
 Sbjct: 76090642 SFTQGSIEAHKGRSFRWI*DKGPVIVESYIEFIQSYRDSFGSRGVCGEFVAVVNKAMSAKF 76090463

Query: 344 ERLVASAEQLLKELPWPPTFEKDKFLTPDFTSLDVLTFAGSGIPAGINIPNYDDLRTQEG 403
 E LV SAEQLLKELPW P FEKDKFLTPDFTS+DVLTFAGSGI AGINI NY+DL+QTEG
 Sbjct: 76090462 EWLVSASAEQLLKELPWSPAFKDKFLTPDFTSVDVLT FAGSGIAAGINISNYNDLKQTEG 76090283

Query: 404 FKNVSLGNVLA VAYATQREKLTFL EEDDKDLYILWKGPSFDVQVGLHELLGHGSGKLFVQ 463
 FKNVSLGNVLAV ATQ EKLT LEE DKDLYI+ GPSFDVQVGLHELLG+GSGKL Q
 Sbjct: 76090282 FKNVSLGNVLAVV*ATQWEKLTVLEESDKDLYIVLMGSPFDVQVGLHELLGYGSGKLIEQ 76090103

Query: 464 DEKGAFNFDQET 475
 DEKGAFNFDQET
 Sbjct: 76090102 DEKGAFNFDQET 76090067

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 ▶**

Summary

▶ setup

- Homo_sapiens
- Genomic sequence
- TBLASTN
- Low sensitivity

▶ configure

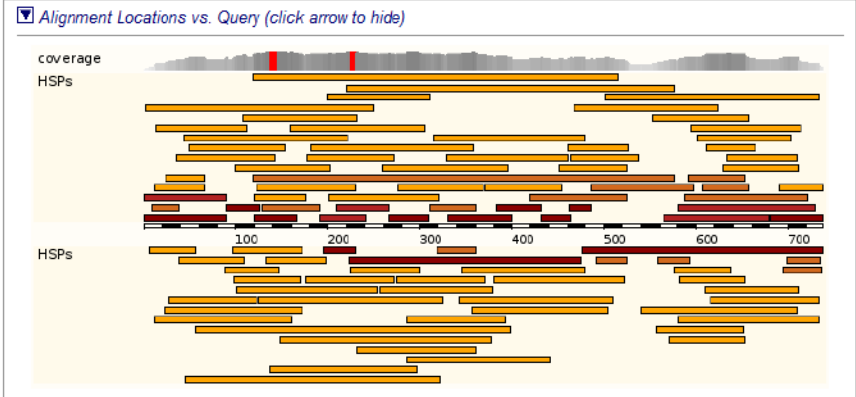
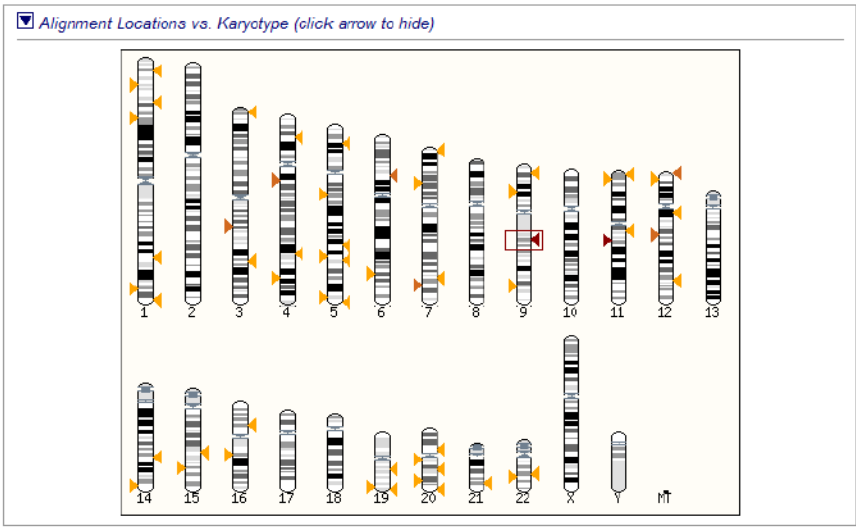
- -E: 10
- -B: 100
- -filter: seg
- -W: 4
- -hitdist: 40
- -matrix: BLOSUM80
- -T: 16

▶ results

▶ display

ⓘ Not yet initialised

Displaying unnamed sequence alignments vs Homo_sapiens LATESTGP database
 Showing top 100 alignments of 1961, sorted by Raw Score
 [refresh](#)



- ### Summary
- setup**
 - Homo_sapiens
 - Genomic sequence
 - TBLASTN
 - Low sensitivity
 - configure**
 - E: 10
 - B: 100
 - filter: seg
 - W: 4
 - hitdist: 40
 - matrix: BLOSUM80
 - T: 16
 - results**
 - display**
 - Not yet initialised

Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort (Use the 'ctrl' key to select multiples) [refresh](#)

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

Les outils

Annotation de variants

The screenshot displays the Ensembl genome browser interface. At the top, the navigation bar includes links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located on the right with the placeholder text "Search all species...".

The main content area is divided into several sections:

- Tools:** A row of tool links: [All tools](#), [BioMart >](#) (with subtext: "Export custom datasets from Ensembl with this data-mining tool"), [BLAST/BLAT >](#) (with subtext: "Search our genomes for your DNA or protein sequence"), and [Variant Effect Predictor >](#) (with subtext: "Analyse your own variants and predict the functional consequences of known and unknown variants").
- Search:** A search box with a dropdown menu set to "All species" and a "Go" button. Below it, an example search string is provided: "e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease".
- All genomes:** A dropdown menu labeled "-- Select a species --". Below it, there are featured genome entries: "Pig breeds" (with a pig icon and subtext: "Pig reference genome and 12 additional breeds"), "Human" (with a human icon, subtext: "GRCh38.p13", and a link "Still using GRCh37?"), "Mouse" (with a mouse icon, subtext: "GRCm39"), and "Zebrafish" (with a zebrafish icon, subtext: "GRCz11"). A link "View full list of all species" is also present.
- Ensembl Release 105 (Dec 2021):** A section with a blue header containing a list of updates:
 - Updated allele frequency data from the NCBI Allele Frequency Aggregator (ALFA) release 2.
 - Update to the Variant Recoder supporting MANE annotation and variant names in external databases.
 - Dog (*Canis lupus familiaris*) reference genome has changed from CanFam3.1 to ROS Clam 1.0 Labrador retriever.
 - Support for BCF files.A link "More release news" is provided.
- Ensembl Rapid Release:** A section with a blue header containing the text: "New assemblies with gene and protein annotation every two weeks." A note states: "Note: species that already exist on this site will continue to be updated with the full range of annotations." A "Go" button is present. Below, it says: "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." A link "Rapid Release news" is provided.
- Other news from our blog:** A list of recent news items:
 - 19 Jan 2022: [Update to the Ensembl COVID-19 resource](#)
 - 12 Jan 2022: [Homology data available in Ensembl Rapid Release](#)
 - 17 Dec 2021: [146 new insect genomes on Ensembl Rapid Release](#)
- Footer:** A row of six tool icons with descriptions: "Compare genes across species", "Find SNPs and other variants for my gene", "Gene expression in different tissues", "Retrieve gene sequence", "Find a Data Display", and "Use my own data in Ensembl". Below this, the EMBL-EBI logo and text state: "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." A link "Our acknowledgements page" is provided. The Elixir logo is also present.

At the bottom of the page, the text "Ensembl release 105 - Dec 2021 © EMBL-EBI" is on the left, "Outils : visualisation de ses données" is in the center, and "Permanent link - View in archive site" is on the right.

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)

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 105?](#)

VEP interfaces

Web interface




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

[Documentation](#)

Launch VEP

Command line tool



- More options and flexibility
- For large volumes of data


[Documentation](#)

[Clone from GitHub](#)

[Download \(zip\)](#)

[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](#)

Variant Effect Predictor

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

Search all species...

VEP

Web Tools

- Web Tools
- BLAST/BLAT
- Variant Effect Predictor**
- Linkage Disequilibrium Calculator
- Variant Recoder
- File Chameleon
- Assembly Converter
- ID History Converter
- VCF to PED Converter
- Data Slicer
- Post-GWAS

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

Variant Effect Predictor

New job Clear form

Species: x

Assembly: GRCh38.p13
[Change species](#)

If you are looking for VEP for Human GRCh37, please go to [GRCh37 website](#).

Name for this job (optional):

Input data:

Either paste data:

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

Or upload file: Aucun fichier choisi

Or provide file URL:

Transcript database to use:

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

Additional configurations:

Identifiers

Variants and frequency data

Variants and frequency data

Find co-located known variants:

Variant synonyms:

Frequency data for co-located variants: 1000 Genomes global minor allele frequency

Variant Effect Predictor

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

Search all species...

VEP

Web Tools

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

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Variant Effect Predictor results

Job details

Summary statistics

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

Consequences (all)

- upstream_gene_variant: 41%
- intron_variant: 20%
- frameshift_variant: 13%
- NMD_transcript_variant: 9%
- downstream_gene_variant: 5%
- TF_binding_site_variant: 4%
- non_coding_transcript_variant: 4%
- 3_prime_UTR_variant: 2%
- missense_variant: 2%
- regulatory_region_variant: 2%

Coding consequences

- frameshift_variant: 88%
- missense_variant: 13%

Results preview

Navigation (per variant) | Filters | Download

Page: 1 of 1 | Show: All variants | Filter: Uploaded variant is defined | 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	ENSG00000240381	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	ENSG00000240381	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	ENST00000358150.10	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

The screenshot shows the Ensembl website interface. At the top, the navigation bar includes 'BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog'. The 'Downloads' tab is highlighted. Below the navigation bar, there are tabs for 'Using this website', 'Annotation and prediction', 'Data access', 'API & software', and 'About us'. The 'API & software' tab is selected, leading to the 'Accessing Ensembl Data' page. The page title is 'Accessing Ensembl Data'. The main content area is divided into four sections: 'Small quantities of data', 'Fast programmatic access', 'Complete datasets and databases', and 'Complex cross-database queries'. Each section contains text, icons, and links. The 'Small quantities of data' section shows a 'Export data' button and a list of nucleotide sequences. The 'Fast programmatic access' section mentions 'REST servers'. The 'Complete datasets and databases' section mentions 'FTP site' and 'MySQL dumps'. The 'Complex cross-database queries' section mentions 'BioMart data-mining tool'. At the bottom of the page, there is a footer with 'Ensembl release 105 - Dec 2021 © EMBL-EBI', a 'Permanent link' button, and four columns of links: 'About Us', 'Get help', 'Our sister sites', and 'Follow us'.

Accessing Ensembl Data

Ensembl data is available through a number of routes - which you choose depends on the amount and type of data you wish to fetch. Please note that Ensembl coordinates always have a one-based start.

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!

Fast programmatic access

For fast access in any programming language, we recommend using our [REST servers](#). Various REST endpoints provide access to vast amounts of Ensembl data.

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.

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 105 - Dec 2021 © 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

- [Blog](#)
- [Twitter](#)
- [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"

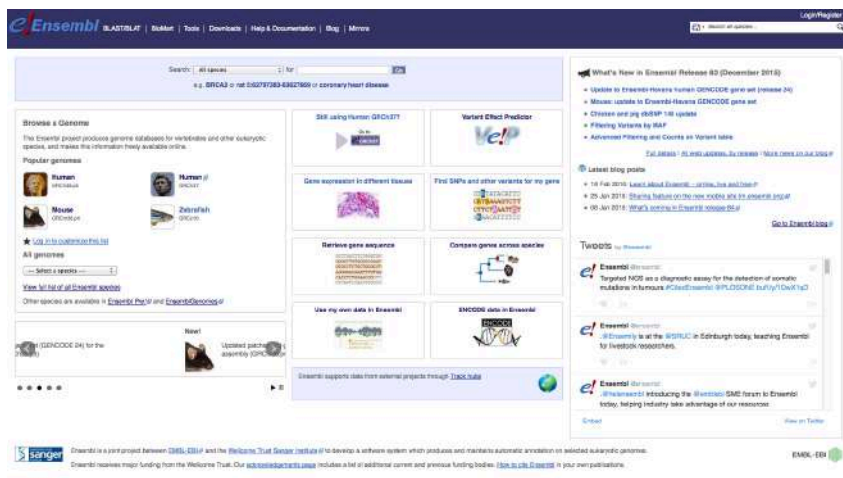
The screenshot shows the Ensembl genome browser interface. At the top, there is a navigation bar with links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar on the right is labeled "Search all species...". Below the navigation bar, there are tabs for "New", "Count", and "Results". The main content area displays a "Dataset" section with "Homo sapiens genes (GRCh37.p13)" selected. There are also sections for "Filters" (with "[None selected]"), "Attributes", and "Ensembl Gene ID". A message in the center says "Please select columns to be included in the output and hit 'Results' when ready". Below this message, there are radio button options for "Features", "Structures", "Homologs", and "Variation".

The screenshot shows the UniProt bioMart interface. At the top, there is a navigation bar with links for "Services", "Research", "Training", and "About us". The main header features the UniProt and bioMart logos. Below the header, there are tabs for "New", "Count", and "Results". A dropdown menu for "Dataset" is currently set to "[None selected]". To the right of the dataset dropdown is a dropdown menu labeled "- CHOOSE DATABASE -".

The screenshot shows the ICGC Data Portal homepage. At the top center is the ICGC logo, which consists of a stylized DNA double helix. To the right of the logo is the text "ICGC Data Portal". Below the logo and text, there are three prominent buttons: "Cancer Projects" (orange), "Advanced Search" (blue), and "Data Repository" (teal). At the bottom of the page, there is a large search bar containing the text "eg. BRAF, KRAS G12D, DO35108, MU7870, TCGA-06-5858".

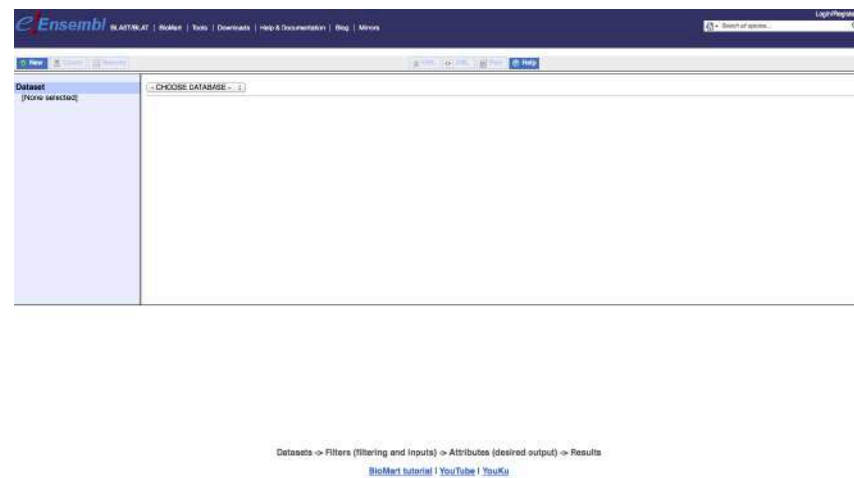
Accéder aux données d'Ensembl

Site web









The screenshot shows the Ensembl website homepage. It features a search bar at the top, navigation links (BLAST/BLAT, BiMart, Tools, Downloads, Help & Documentation, Blog, Mirrors), and a search bar. Below the search bar, there are several sections: 'Browse & Genome' with a 'Popular genomes' list (Human, Mouse, Zebrafish), 'What's New in Ensembl Release 80 (December 2016)', 'Latest blog posts', and 'Tools & Resources'. The footer includes the Ensembl logo and a note about the project's funding.

Outil de fouille: BioMart



The screenshot shows the BioMart tool interface. It features a search bar at the top, navigation links (BLAST/BLAT, BiMart, Tools, Downloads, Help & Documentation, Blog, Mirrors), and a search bar. Below the search bar, there is a 'Dataset' section with a dropdown menu for 'CHOOSE DATABASE' and a 'None selected' option. The footer includes the Ensembl logo and a note about the project's funding.

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

-  Requête complexe
-  Rapide
-  Requiert une formation

BioMart/Ensembl

BioMart

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 105 (Dec 2021)

- Updated allele frequency data from the NCBI Allele Frequency Aggregator (ALFA) release 2
- Update to the Variant Recoder supporting MANE annotation and variant names in external databases
- Dog (*Canis lupus familiaris*) reference genome has changed from CanFam3.1 to ROS Clam 1.0 Labrador retriever
- Support for BCF files

[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

- 19 Jan 2022: [Update to the Ensembl COVID-19 resource](#)
- 12 Jan 2022: [Homology data available in Ensembl Rapid](#)

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

BioMart/Ensembl

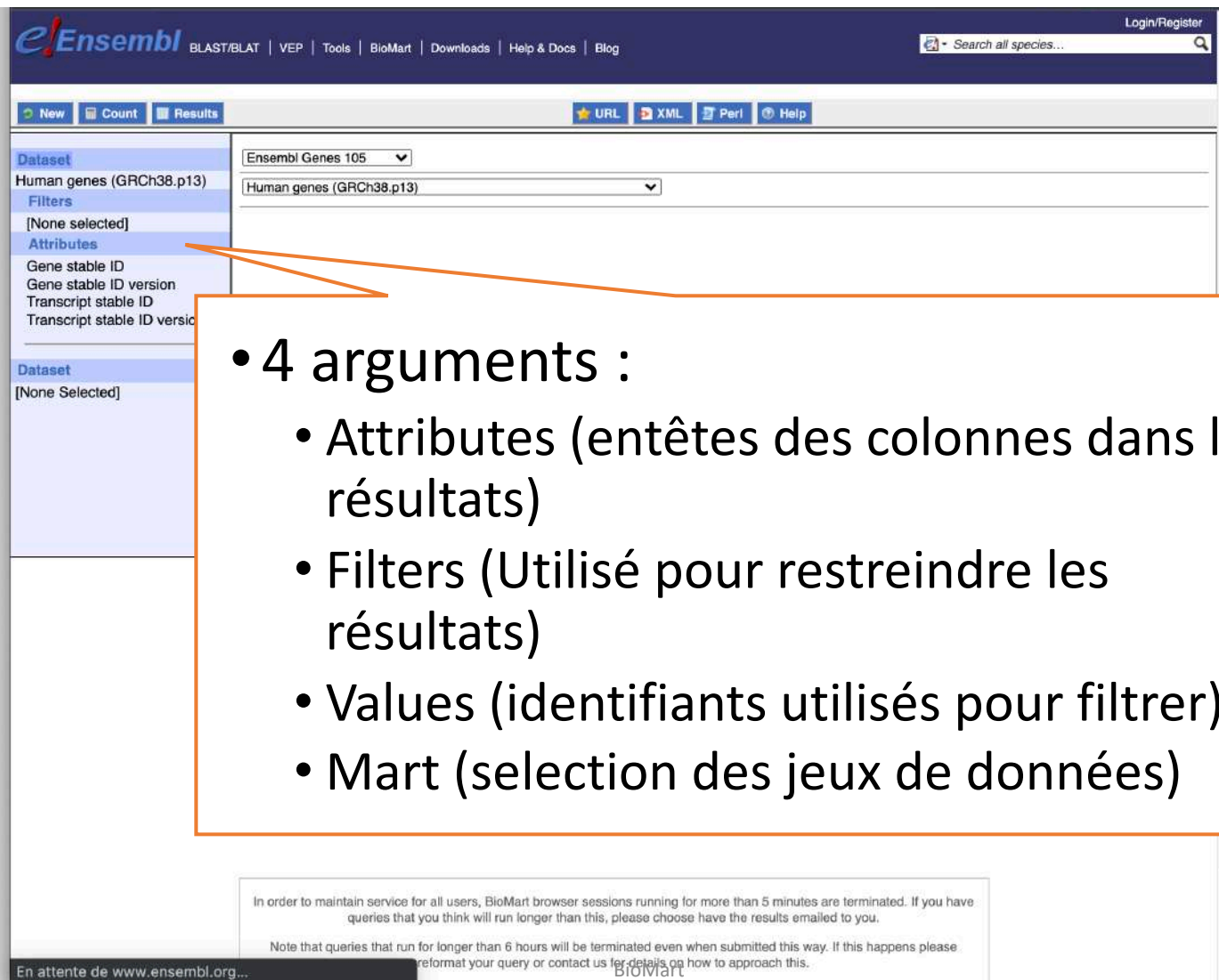
The screenshot displays the Ensembl BioMart interface. At the top, the Ensembl logo is visible along with navigation links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. Below the navigation bar, there are buttons for 'New', 'Count', and 'Results', and a row of icons for 'URL', 'XML', 'Perl', and 'Help'. The main content area is divided into two columns. The left column contains a 'Dataset' dropdown menu currently set to 'Ensembl Genes 105', a 'Filters' section with '[None selected]', and an 'Attributes' section listing 'Gene stable ID', 'Gene stable ID version', 'Transcript stable ID', and 'Transcript stable ID version'. Below this is another 'Dataset' section with '[None Selected]'. The right column shows a list of available datasets, with 'Human genes (GRCh38.p13)' selected. Two callout boxes with orange borders and arrows point to the 'Ensembl Genes 105' dropdown and the 'Human genes (GRCh38.p13)' entry. A third callout box points to the 'Human genes (GRCh38.p13)' entry. At the bottom of the page, there is a footer with the text 'En attente de www.ensembl.org...' and a disclaimer: '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 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.'

Selection de la Base de donnée :

- Genes
- Variation
- Regulation
- Mouse strain

Sélection du jeu de données (génomique)

BioMart/Ensembl



The screenshot shows the Ensembl BioMart interface. At the top, there is a navigation bar with links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located on the right. Below the navigation bar, there are buttons for 'New', 'Count', and 'Results'. The main content area is divided into two sections. The left section is titled 'Dataset' and shows 'Human genes (GRCh38.p13)' with a dropdown menu. Below this, there is a 'Filters' section with '[None selected]' and an 'Attributes' section with a list of attributes: 'Gene stable ID', 'Gene stable ID version', 'Transcript stable ID', and 'Transcript stable ID version'. The right section is titled 'Dataset' and shows '[None Selected]'. An orange callout box points to the 'Attributes' section. At the bottom of the page, there is a footer with the text 'En attente de www.ensembl.org...' and a small BioMart logo.

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

Biomart : Partie pratique

Comparaison des browsers

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