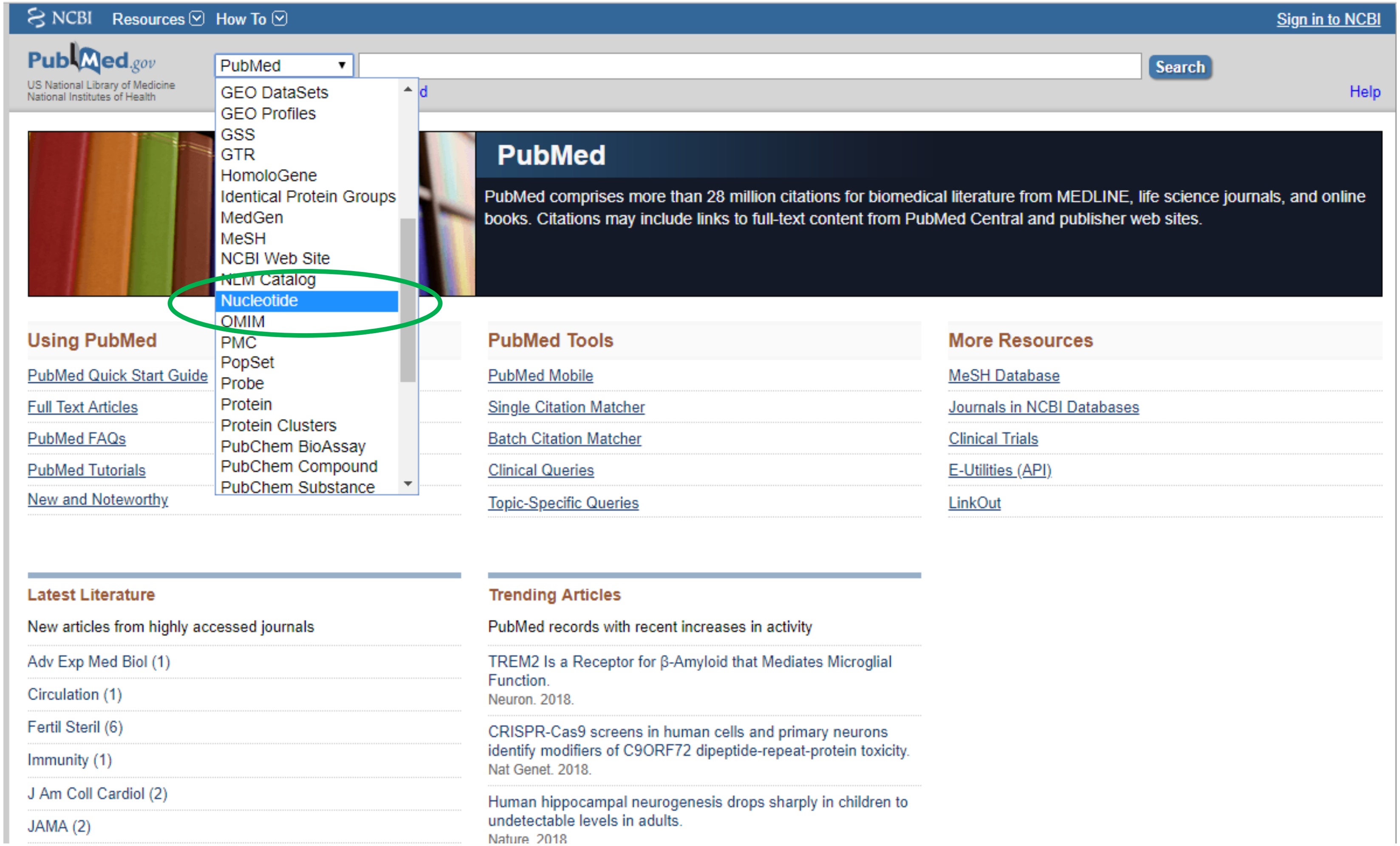
Base de Données
de
Séquences Nucléotidiques

Par Joel sandé

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



PubMed

PubMed comprises more than 28 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Using PubMed
- PubMed Quick Start Guide
- Full Text Articles
- PubMed FAQs
- PubMed Tutorials
- New and Noteworthy

PubMed Tools
- PubMed Mobile
- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources
- MeSH Database
- Journals in NCBI Databases
- Clinical Trials
- E-Utilities (API)
- LinkOut

Latest Literature

New articles from highly accessed journals
- Adv Exp Med Biol (1)
- Circulation (1)
- Fertil Steril (6)
- Immunity (1)
- J Am Coll Cardiol (2)
- JAMA (2)

Trending Articles

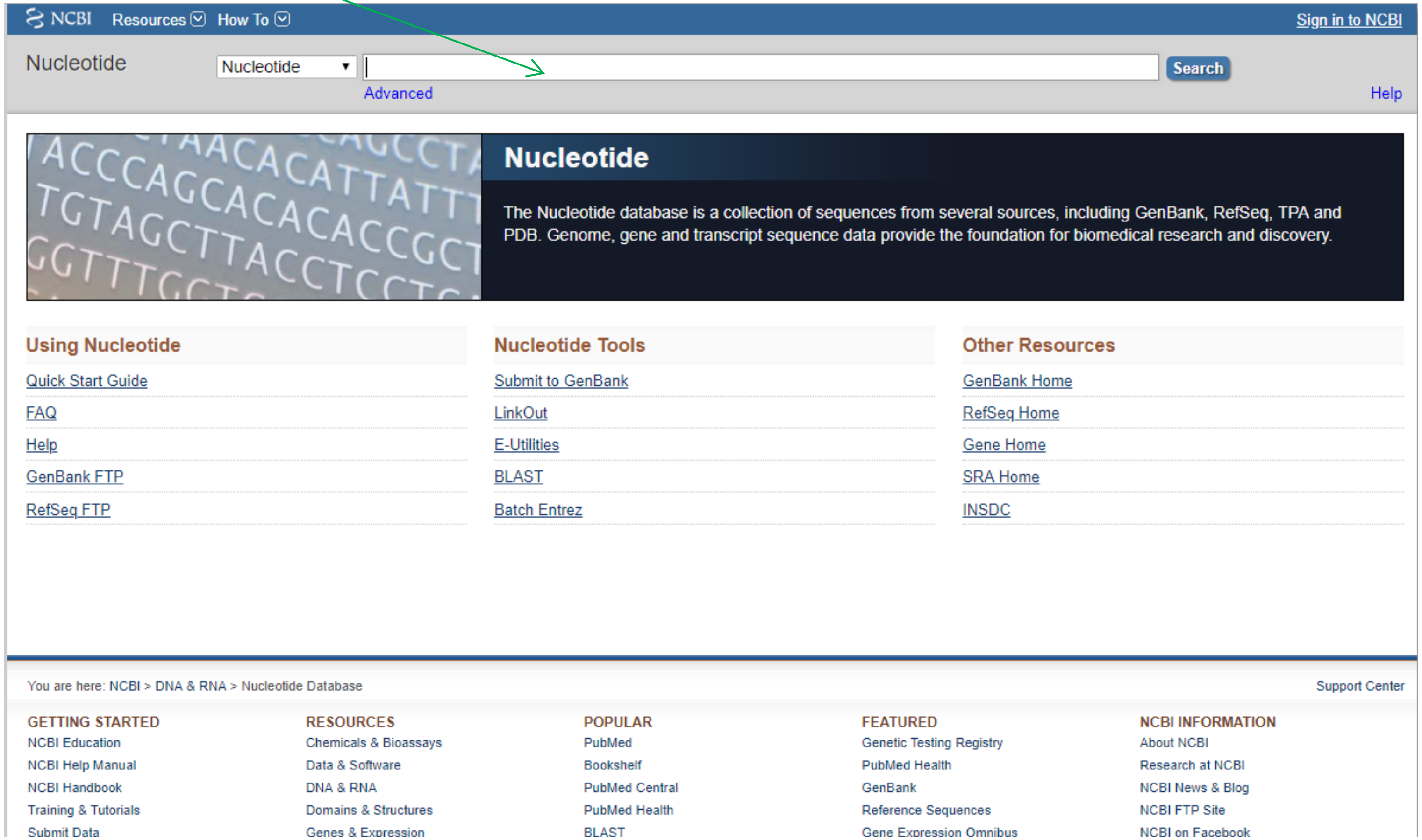
PubMed records with recent increases in activity

TREM2 Is a Receptor for β-Amyloid that Mediates Microglial Function.
Neuron. 2018.

CRISPR-Cas9 screens in human cells and primary neurons identify modifiers of C9ORF72 dipeptide-repeat-protein toxicity.
Nat Genet. 2018.

Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults.
Nature. 2018

Entrez le nom du gène X01714

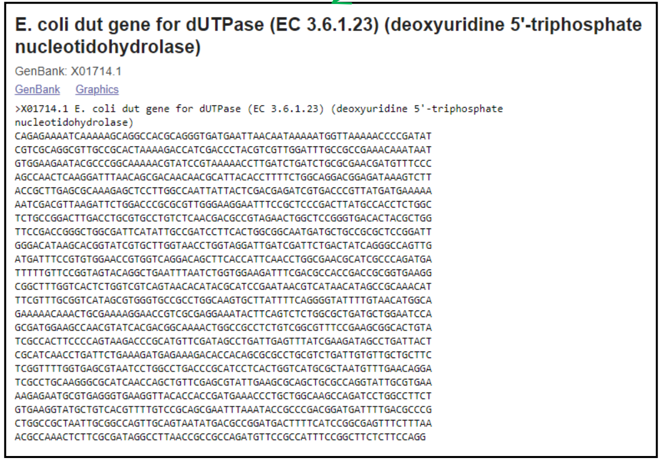


Nucleotide

The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.

Vous verrez les informations sur le gène apparaisant par défaut en format Genbank.

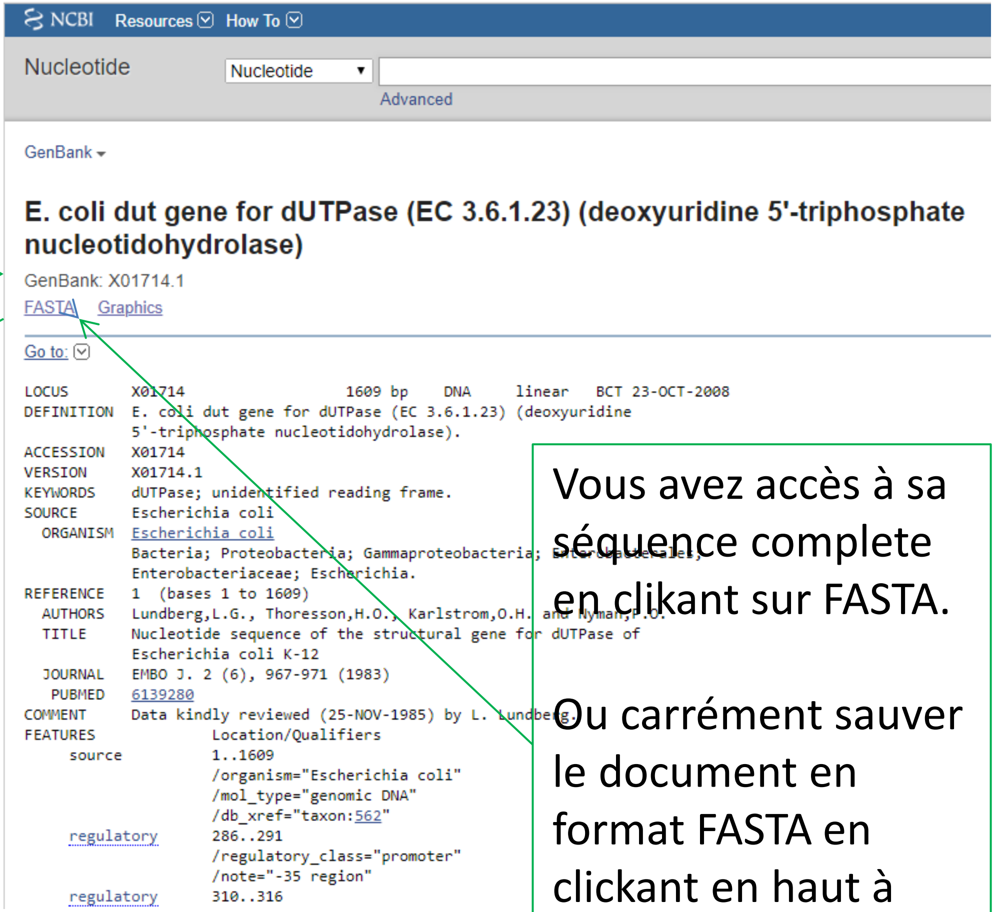
Vous avez toutes les informations sur le gène.


E. coli dut gene for dUTPase (EC 3.6.1.23) (deoxyuridine 5'-triphosphate nucleotidohydrolase)

GenBank: X01714.1

GenBank Graphics

```
>X01714.1 E. coli dut gene for dUTPase (EC 3.6.1.23) (deoxyuridine 5'-triphosphate
nucleotidohydrolase)
CAGAGAAAATCAAAAAGCAGGCCACGCAGGGTGATGAATTAACAATAAAAATGGTTAAAAACCCCGATAT
CGTCGCAGGCGTTGCCGCACTAAAAGACCATCGACCCTACGTCGTTGGATTTGCCGCCGAAACAAATAAT
GTGGAAGAATACGCCCGGCAAAAACGTATCCGTAAAAACCTTGATCTGATCTGCGCGAACGATGTTTCCC
AGCCAACTCAAGGATTTAACAGCGACAACAACGCATTACACCTTTTCTGGCAGGACGGAGATAAAGTCTT
ACCGCTTGAGCGCAAAGAGCTCCTTGGCCAATTATTACTCGACGAGATCGTGACCCGTTATGATGAAAAA
AATCGACGTTAAGATTCTGGACCCGCGCGTTGGGAAGGAATTTCCGCTCCCGACTTATGCCACCTCTGGC
TCTGCCGGACTTGACCTGCGTGCCTGTCTCAACGACGCCGTAGAACTGGCTCCGGGTGACACTACGCTGG
TTCCGACCGGGCTGGCGATTCATATTGCCGATCCTTCACTGGCGGCAATGATGCTGCCGCGCTCCGGATT
GGGACATAAGCACGGTATCGTGCTTGGTAACCTGGTAGGATTGATCGATTCTGACTATCAGGGCCAGTTG
ATGATTTCCGTGTGGAACCGTGGTCAGGACAGCTTCACCATTCAACCTGGCGAACGCATCGCCCAGATGA
TTTTTGTTCCGGTAGTACAGGCTGAATTTAATCTGGTGGAAGATTTCGACGCCACCGACCGCGGTGAAGG
CGGCTTTGGTCACTCTGGTCGTCAGTAACACATACGCATCCGAATAACGTCATAACATAGCCGCAAACAT
TTCGTTTGCGGTCATAGCGTGGGTGCCGCCTGGCAAGTGCTTATTTTCAGGGGTATTTTGTAACATGGCA
GAAAAACAAACTGCGAAAAGGAACCGTCGCGAGGAAATACTTCAGTCTCTGGCGCTGATGCTGGAATCCA
GCGATGGAAGCCAACGTATCACGACGGCAAAACTGGCCGCCTCTGTCGGCGTTTCCGAAGCGGCACTGTA
TCGCCACTTCCCCAGTAAGACCCGCATGTTCGATAGCCTGATTGAGTTTATCGAAGATAGCCTGATTACT
CGCATCAACCTGATTCTGAAAGATGAGAAAGACACCACAGCGCGCCTGCGTCTGATTGTGTTGCTGCTTC
TCGGTTTTGGTGAGCGTAATCCTGGCCTGACCCGCATCCTCACTGGTCATGCGCTAATGTTTGAACAGGA
TCGCCTGCAAGGGCGCATCAACCAGCTGTTCGAGCGTATTGAAGCGCAGCTGCGCCAGGTATTGCGTGAA
AAGAGAATGCGTGAGGGTGAAGGTTACACCACCGATGAAACCCTGCTGGCAAGCCAGATCCTGGCCTTCT
GTGAAGGTATGCTGTCACGTTTTGTCCGCAGCGAATTTAAATACCGCCCGACGGATGATTTTGACGCCCG
CTGGCCGCTAATTGCGGCCAGTTGCAGTAATATGACGCCGGATGACTTTTCATCCGGCGAGTTTCTTTAA
ACGCCAAACTCTTCGCGATAGGCCTTAACCGCCGCCAGATGTTCCGCCATTTCCGGCTTCTCTTCCAGG
```



E. coli dut gene for dUTPase (EC 3.6.1.23) (deoxyuridine 5'-triphosphate nucleotidohydrolase)

GenBank: X01714.1

FASTA Graphics

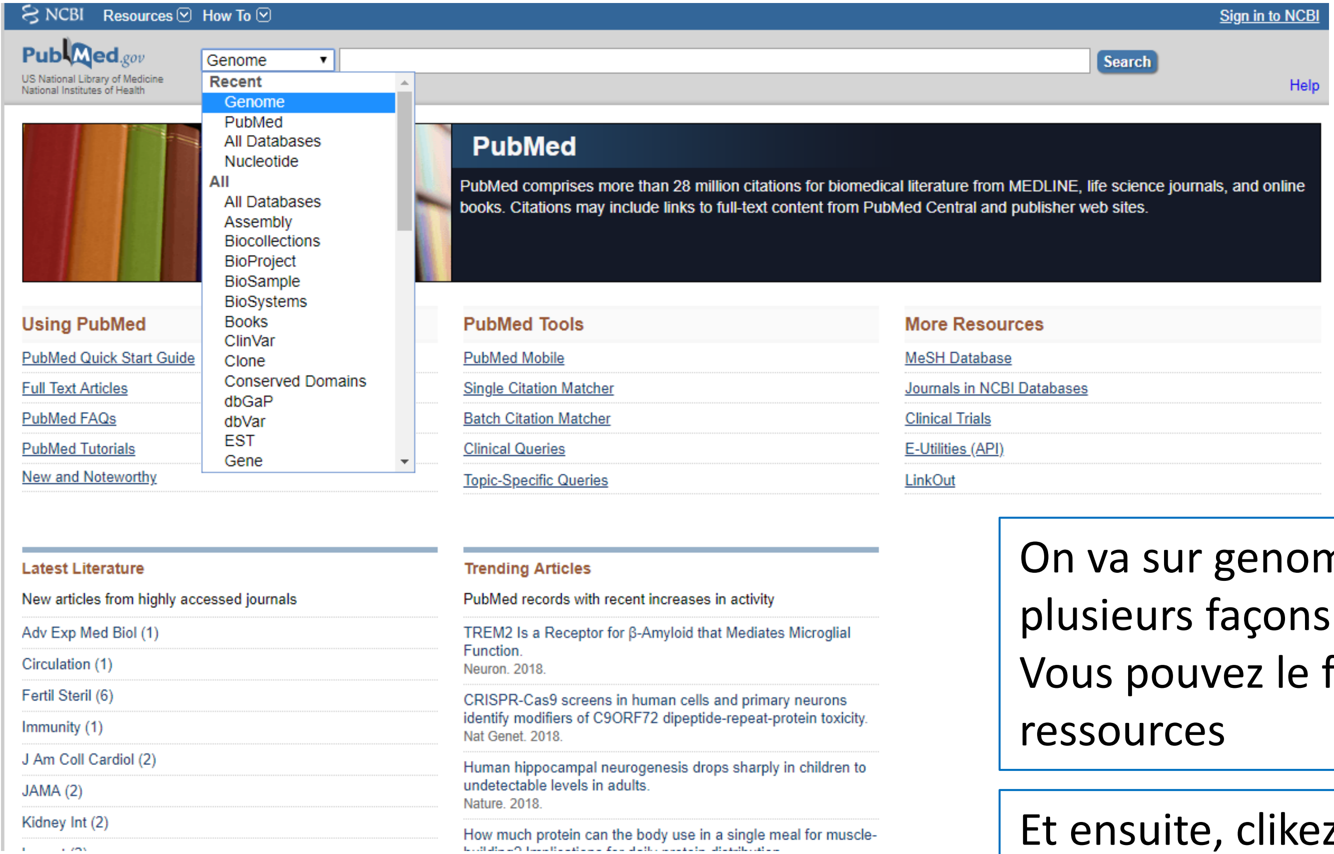
```
LOCUS       X01714                  1609 bp    DNA     linear   BCT 23-OCT-2008
DEFINITION  E. coli dut gene for dUTPase (EC 3.6.1.23) (deoxyuridine
            5'-triphosphate nucleotidohydrolase).
ACCESSION   X01714
VERSION     X01714.1
KEYWORDS    dUTPase; unidentified reading frame.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 1609)
  AUTHORS   Lundberg,L.G., Thoresson,H.O., Karlstrom,O.H. and Nyman,P.O.
  TITLE     Nucleotide sequence of the structural gene for dUTPase of
            Escherichia coli K-12
  JOURNAL   EMBO J. 2 (6), 967-971 (1983)
   PUBMED   6139280
COMMENT     Data kindly reviewed (25-NOV-1985) by L. Lundberg.
FEATURES             Location/Qualifiers
     source          1..1609
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
     regulatory      286..291
                     /regulatory_class="promoter"
                     /note="-35 region"
     regulatory      310..316
```

Vous avez accès à sa séquence complete en clikant sur FASTA.

Ou carrément sauver le document en format FASTA en clickant en haut à droite sur Send to FASTA.

Genome complet d’un organisme

https://www.ncbi.nlm.nih.gov/



PubMed

PubMed comprises more than 28 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Using PubMed

- PubMed Quick Start Guide
- Full Text Articles
- PubMed FAQs
- PubMed Tutorials
- New and Noteworthy

PubMed Tools

- PubMed Mobile
- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources

- MeSH Database
- Journals in NCBI Databases
- Clinical Trials
- E-Utilities (API)
- LinkOut

Latest Literature

New articles from highly accessed journals

- Adv Exp Med Biol (1)
- Circulation (1)
- Fertil Steril (6)
- Immunity (1)
- J Am Coll Cardiol (2)
- JAMA (2)
- Kidney Int (2)

Trending Articles

PubMed records with recent increases in activity

TREM2 Is a Receptor for β-Amyloid that Mediates Microglial Function.
Neuron. 2018.

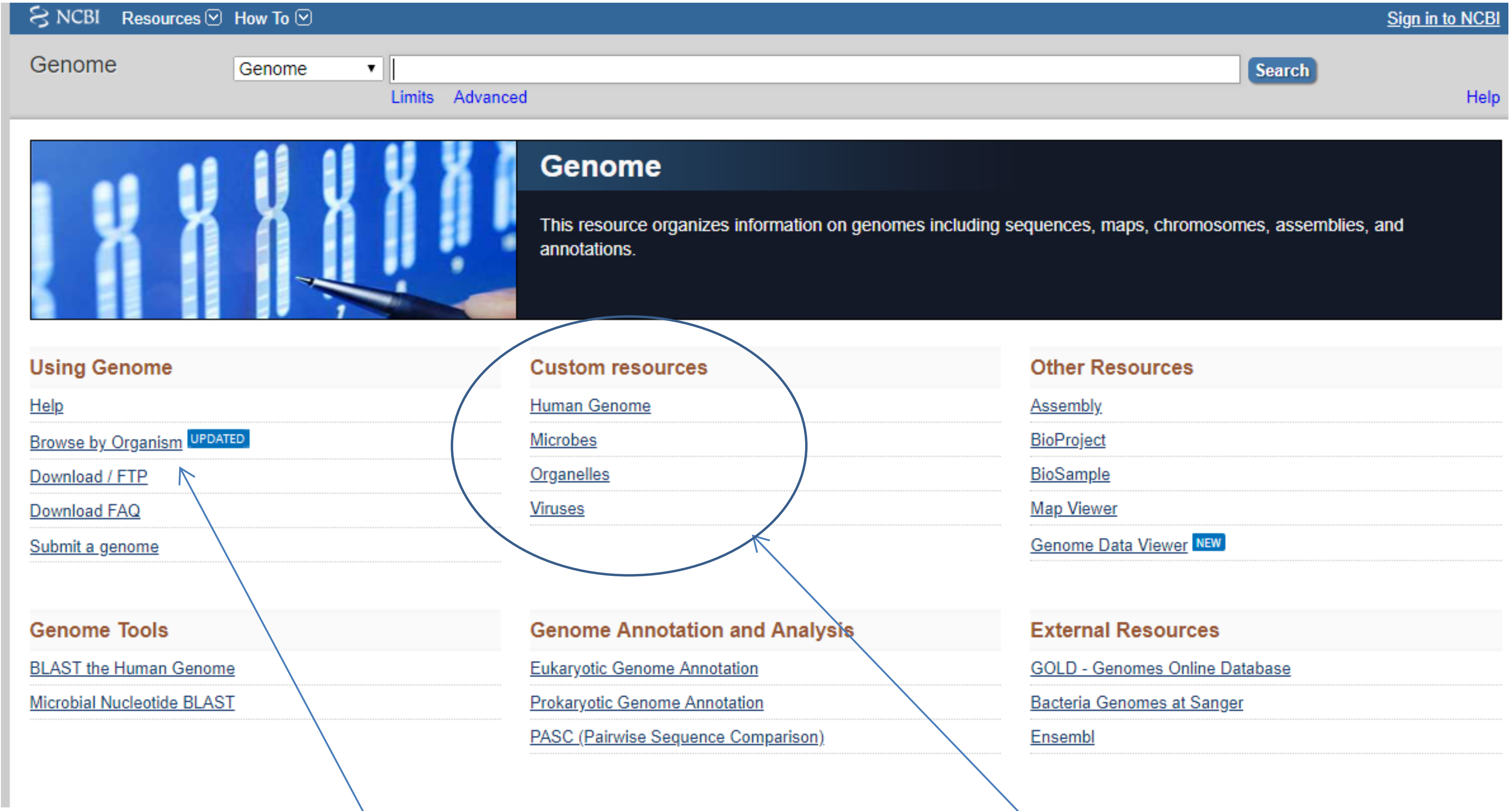
CRISPR-Cas9 screens in human cells and primary neurons identify modifiers of C9ORF72 dipeptide-repeat-protein toxicity.
Nat Genet. 2018.

Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults.
Nature. 2018.

How much protein can the body use in a single meal for muscle-

On va sur genome : Il y a plusieurs façons de faire celà. Vous pouvez le faire Via ressources

Et ensuite, clikez sur Search en haut à droite



Genome

This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.

Using Genome

- Help
- Browse by Organism UPDATED
- Download / FTP
- Download FAQ
- Submit a genome

Custom resources

- Human Genome
- Microbes
- Organelles
- Viruses

Other Resources

- Assembly
- BioProject
- BioSample
- Map Viewer
- Genome Data Viewer NEW

Genome Tools

- BLAST the Human Genome
- Microbial Nucleotide BLAST

Genome Annotation and Analysis

- Eukaryotic Genome Annotation
- Prokaryotic Genome Annotation
- PASC (Pairwise Sequence Comparison)

External Resources

- GOLD - Genomes Online Database
- Bacteria Genomes at Sanger
- Ensembl

Vous pouvez faire la recherche par organisme

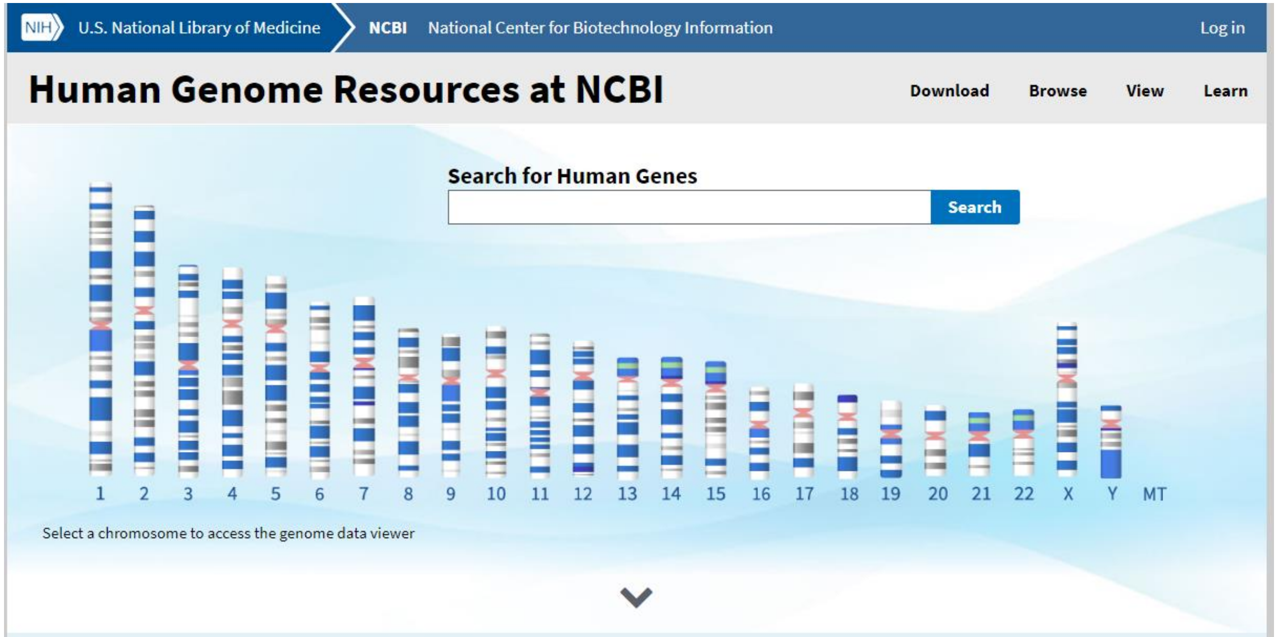
Ou aller directement sur un organisme désiré selon l’espèce désirée

Genome Complet de l’humain

Clickez sur human genome et clikez sur l’un des chromosomes.

En passant, voyez la différence de taille entre le chromosome X de la femme et le chromosome Y de l’homme.

Distes-donc au hommes de respecter la femmes.



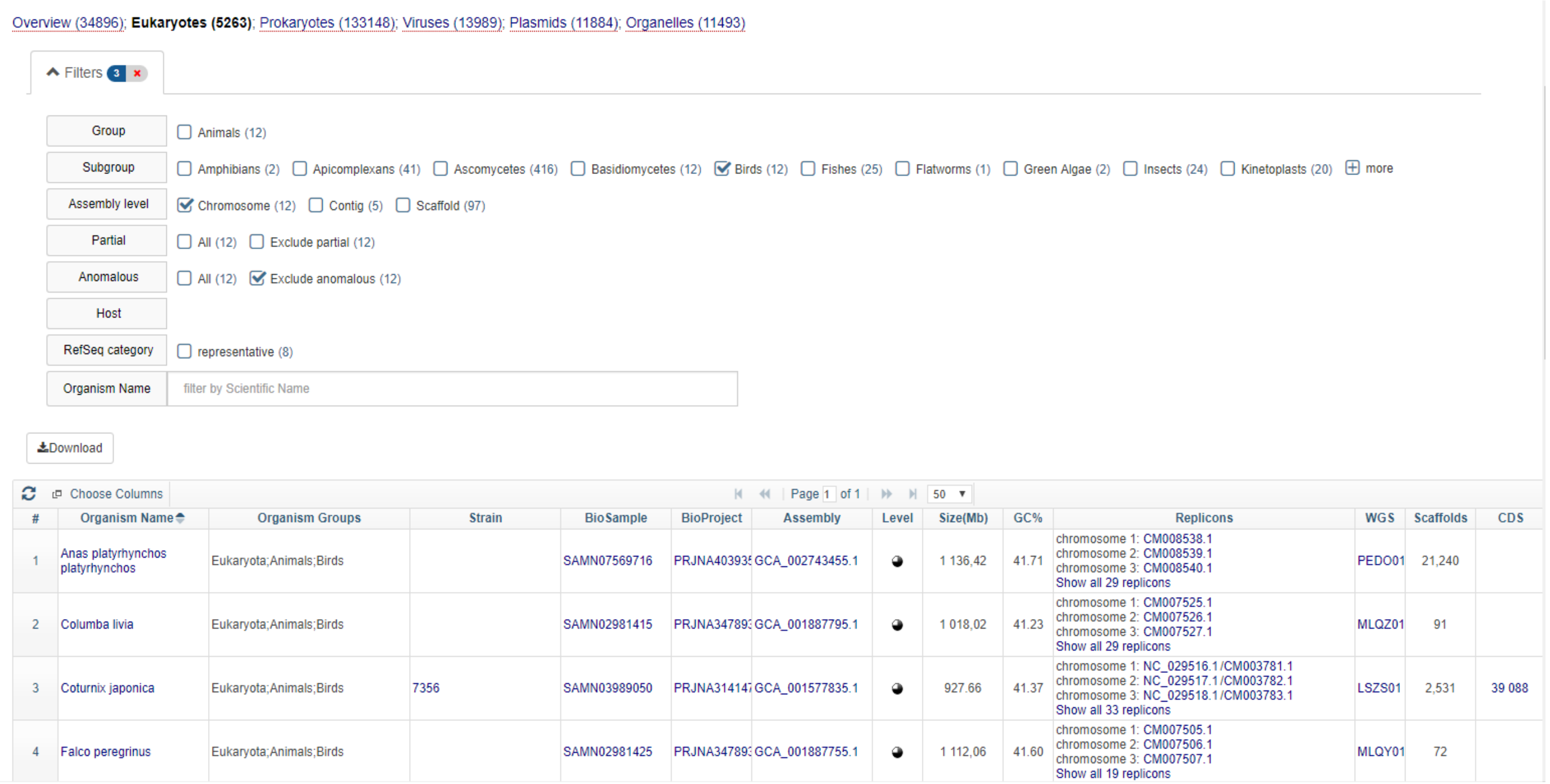
Select a chromosome to access the genome data viewer

Allons chercher le genome d’un Oiseau

- Browse by organism : Voyez-vous ? Il y a plein de bactéries et virus répertoriés dans cette base de données.
- Il est plus facile de trouver le génome d’un micro-organisme que de trouver le genome d’un mammifère.
- On peut pour chaque organisme, la taille du génome, le nombre de chromosomes, le nombre d’organels, de plasmides, …

- Pour trouver le génome d’un Oiseau:
  - Sélectionnez Eucaryote
  - Utilisez l’option Filter
  - Cochez et décochez les options qui vous conviennent. : la page se mettra automatiquement à jours de vos sélections.

Allons chercher le genome d’un Oiseau



Overview (34896); **Eukaryotes (5263)**; Prokaryotes (133148); Viruses (13989); Plasmids (11884); Organelles (11493)

Group: ☐ Animals (12)

Subgroup: ☐ Amphibians (2) ☐ Apicomplexans (41) ☐ Ascomycetes (416) ☐ Basidiomycetes (12) ☑ Birds (12) ☐ Fishes (25) ☐ Flatworms (1) ☐ Green Algae (2) ☐ Insects (24) ☐ Kinetoplasts (20) ⊞ more

Assembly level: ☑ Chromosome (12) ☐ Contig (5) ☐ Scaffold (97)

Partial: ☐ All (12) ☐ Exclude partial (12)

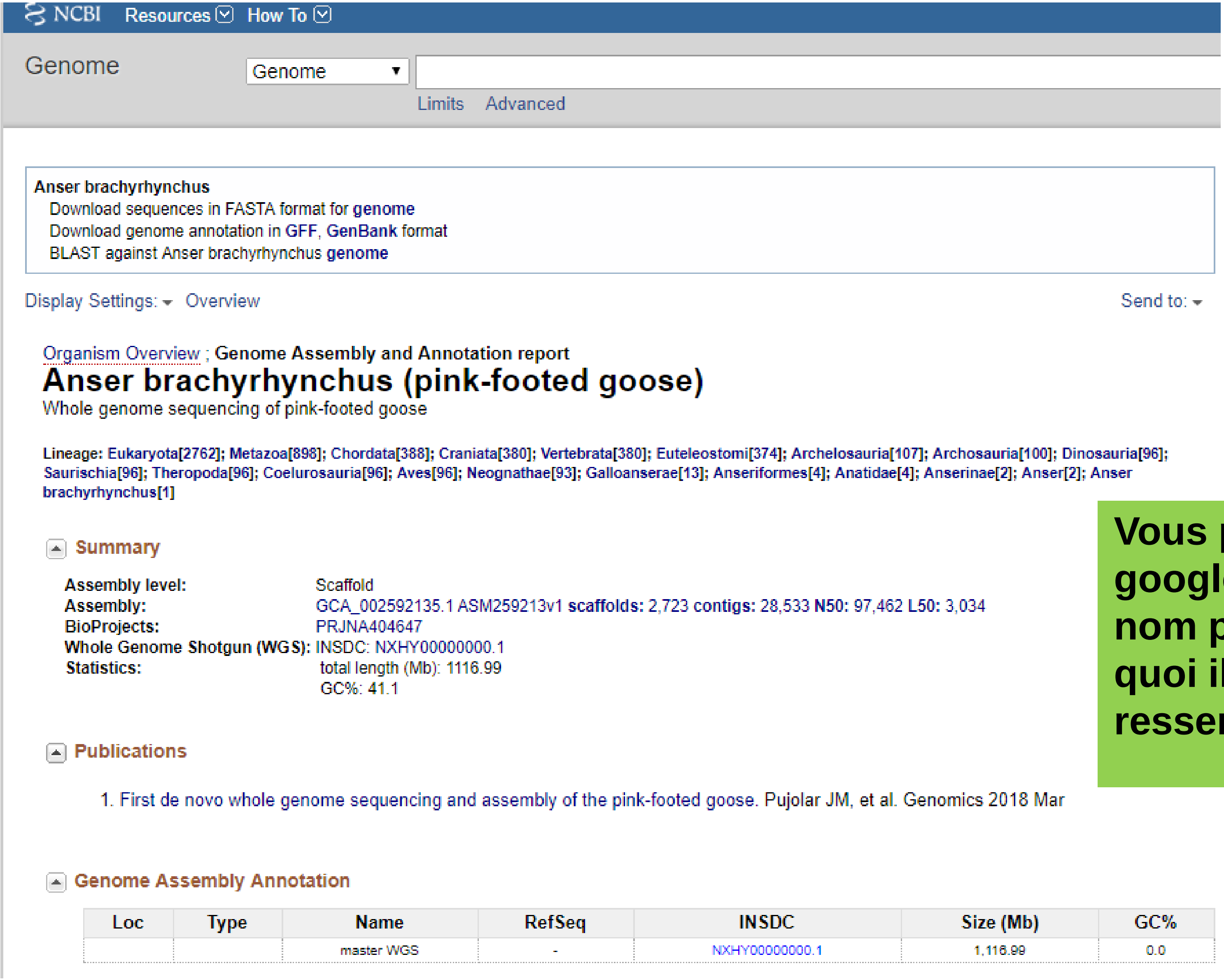
Anomalous: ☐ All (12) ☑ Exclude anomalous (12)

Host

RefSeq category: ☐ representative (8)

Organism Name: filter by Scientific Name

#	Organism Name	Organism Groups	Strain	BioSample	BioProject	Assembly	Level	Size(Mb)	GC%	Replicons	WGS	Scaffolds	CDS
1	Anas platyrhynchos platyrhynchos	Eukaryota;Animals;Birds		SAMN07569716	PRJNA40393	GCA_002743455.1	●	1 136,42	41.71	chromosome 1: CM008538.1 chromosome 2: CM008539.1 chromosome 3: CM008540.1 Show all 29 replicons	PEDO01	21,240	
2	Columba livia	Eukaryota;Animals;Birds		SAMN02981415	PRJNA34789	GCA_001887795.1	●	1 018,02	41.23	chromosome 1: CM007525.1 chromosome 2: CM007526.1 chromosome 3: CM007527.1 Show all 29 replicons	MLQZ01	91	
3	Coturnix japonica	Eukaryota;Animals;Birds	7356	SAMN03989050	PRJNA31414	GCA_001577835.1	●	927.66	41.37	chromosome 1: NC_029516.1 /CM003781.1 chromosome 2: NC_029517.1 /CM003782.1 chromosome 3: NC_029518.1 /CM003783.1 Show all 33 replicons	LSZS01	2,531	39 088
4	Falco peregrinus	Eukaryota;Animals;Birds		SAMN02981425	PRJNA34789	GCA_001887755.1	●	1 112,06	41.60	chromosome 1: CM007505.1 chromosome 2: CM007506.1 chromosome 3: CM007507.1 Show all 19 replicons	MLQY01	72	



Anser brachyrhynchus
Download sequences in FASTA format for **genome**
Download genome annotation in **GFF**, **GenBank** format
BLAST against Anser brachyrhynchus **genome**

Organism Overview ; **Genome Assembly and Annotation report**

Anser brachyrhynchus (pink-footed goose)

Whole genome sequencing of pink-footed goose

Lineage: Eukaryota[2762]; Metazoa[898]; Chordata[388]; Craniata[380]; Vertebrata[380]; Euteleostomi[374]; Archelosauria[107]; Archosauria[100]; Dinosauria[96]; Saurischia[96]; Theropoda[96]; Coelurosauria[96]; Aves[96]; Neognathae[93]; Galloanserae[13]; Anseriformes[4]; Anatidae[4]; Anserinae[2]; Anser[2]; Anser brachyrhynchus[1]

Summary

Assembly level:	Scaffold
Assembly:	GCA_002592135.1 ASM259213v1 **scaffolds:** 2,723 **contigs:** 28,533 **N50:** 97,462 **L50:** 3,034
BioProjects:	PRJNA404647
Whole Genome Shotgun (WGS):	INSDC: NXHY00000000.1
Statistics:	total length (Mb): 1116.99 GC%: 41.1

Publications

1. First de novo whole genome sequencing and assembly of the pink-footed goose. Pujolar JM, et al. Genomics 2018 Mar

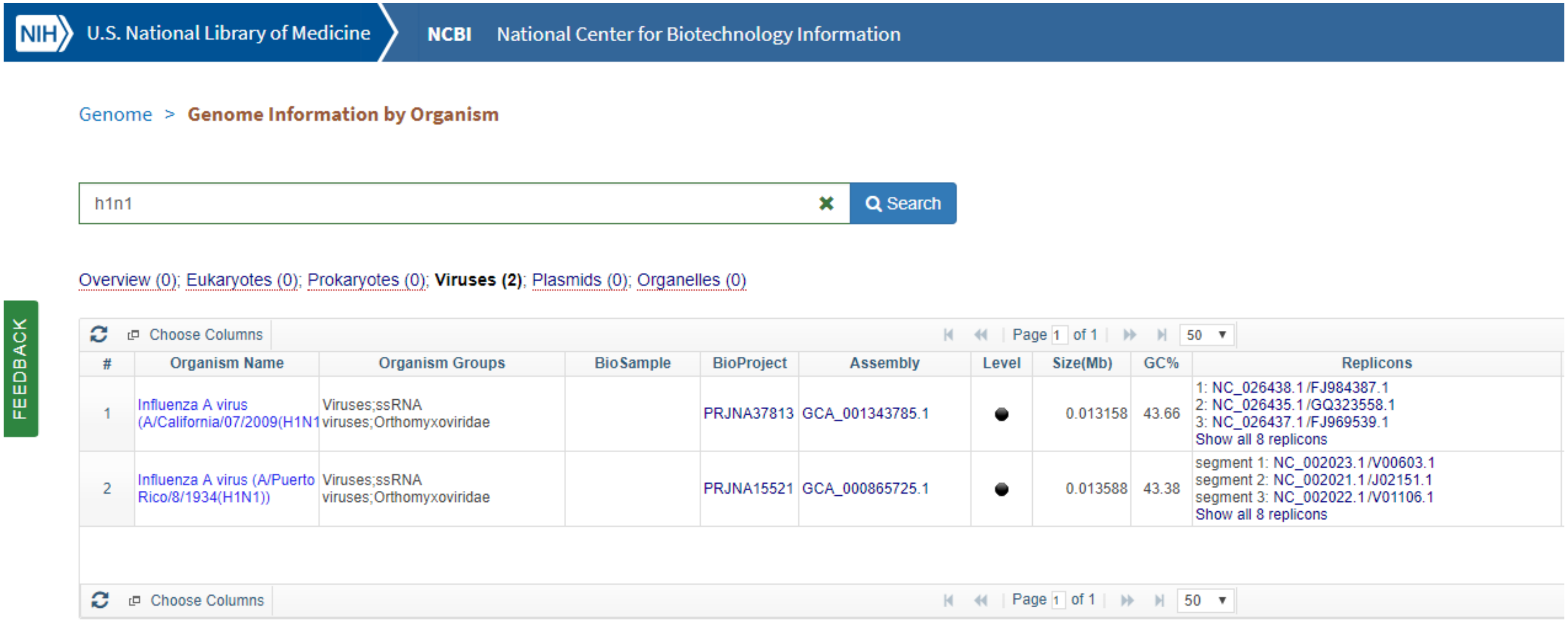
Genome Assembly Annotation

Loc	Type	Name	RefSeq	INSDC	Size (Mb)	GC%
		master WGS	-	NXHY00000000.1	1,116.99	0.0

Vous pouvez googler son nom pour voir à quoi il ressemble.

Anser brachyrhynchus
ou
pink-footed goose

Faisons la même expérience pour le Virus de la grippe H1N1



Genome > **Genome Information by Organism**

h1n1

Overview (0); Eukaryotes (0); Prokaryotes (0); **Viruses (2)**; Plasmids (0); Organelles (0)

#	Organism Name	Organism Groups	BioSample	BioProject	Assembly	Level	Size(Mb)	GC%	Replicons
1	Influenza A virus (A/California/07/2009(H1N1	Viruses;ssRNA viruses;Orthomyxoviridae		PRJNA37813	GCA_001343785.1	●	0.013158	43.66	1: NC_026438.1 /FJ984387.1 2: NC_026435.1 /GQ323558.1 3: NC_026437.1 /FJ969539.1 Show all 8 replicons
2	Influenza A virus (A/Puerto Rico/8/1934(H1N1))	Viruses;ssRNA viruses;Orthomyxoviridae		PRJNA15521	GCA_000865725.1	●	0.013588	43.38	segment 1: NC_002023.1 /V00603.1 segment 2: NC_002021.1 /J02151.1 segment 3: NC_002022.1 /V01106.1 Show all 8 replicons

Il y en a 2 répertoriés dans la base de donnée NCBI https://www.ncbi.nlm.nih.gov/genome/10290?genome_assembly_id=306186

Influenza A virus (A/California/07/2009(H1N1))
Download sequences in FASTA format for genome, protein
Download genome annotation in GFF, GenBank or tabular format
All 7 reference or representative genomes for species:
Browse the list

Organism Overview ; **Genome Assembly and Annotation report**

Influenza A virus (A/California/07/2009(H1N1))

Lineage: Viruses[9563]; ssRNA viruses[2030]; ssRNA negative-strand viruses[523]; Orthomyxoviridae[10]; Alphainfluenzavirus[1]; Influenza A virus[1]; H1N1 subtype[1]; Influenza A virus (A/California/07/2009(H1N1))[1]

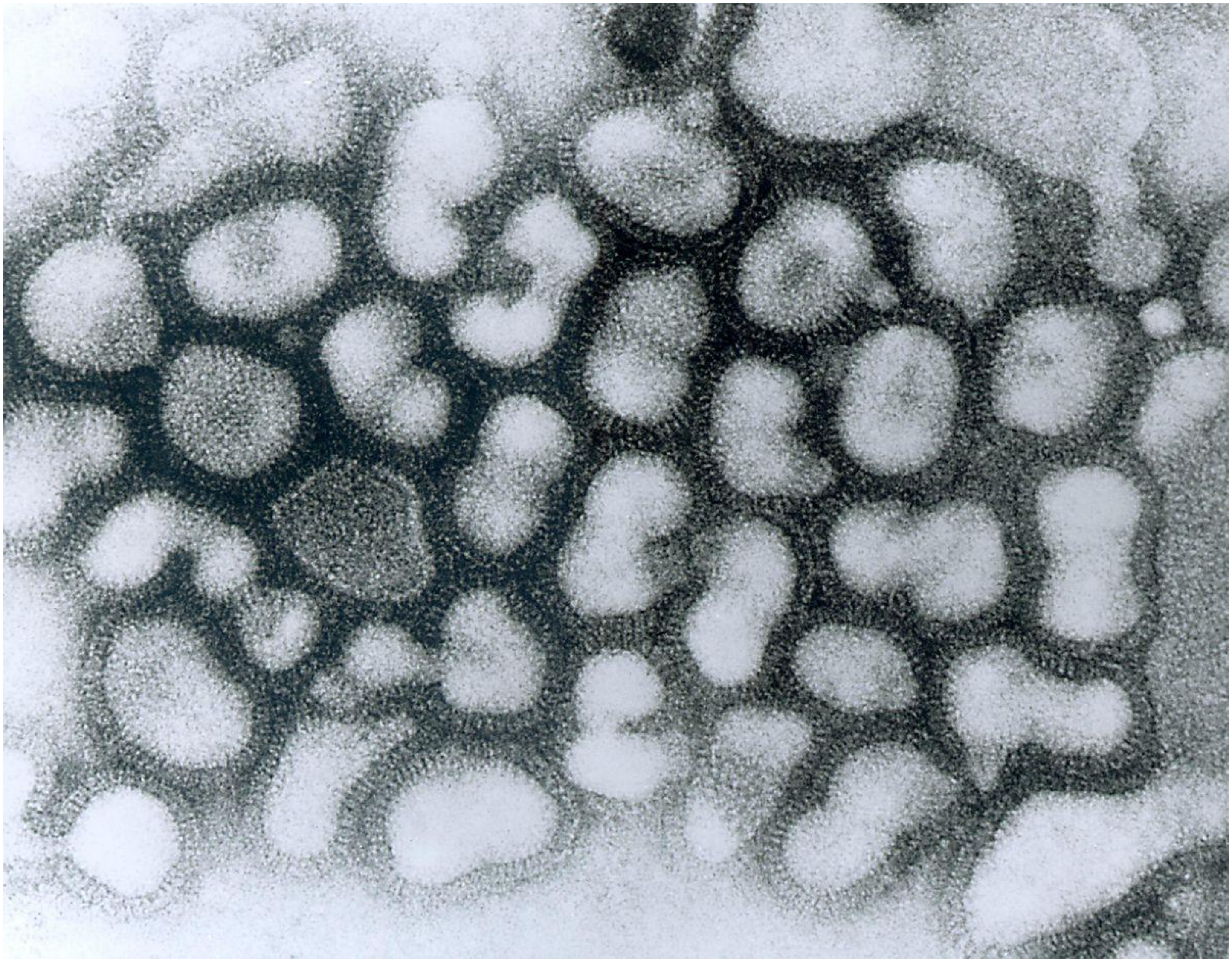
Summary

Submitter: 2009 H1N1 Flu Outbreak Sequencing Centers
Assembly level: Complete Genome
Human Pathogen
Assembly: GCA_001343785.1 ViralMultiSegProj274766 **scaffolds:** 8 **contigs:** 8 **N50:** 2,151 **L50:** 3
BioProjects: PRJNA274766, PRJNA37813
Statistics: total length (Mb): 0.013158
protein count: 11
GC%: 43.6638

Publications

1. New genetic variants of influenza A(H1N1)pdm09 detected in Cuba during 2011-2013. Arencibia A, et al. Infect Genet Evol 2015 Jun
2. Host Adaptation and the Alteration of Viral Properties of the First Influenza A/H1N1pdm09 Virus Isolated in Japan. Ainai A, et al. PLoS One 2015
3. Molecular genetic analysis of the Influenza A(H1N1)pdm09 virus from lethal and recovered cases in Russia from 2009 to 2014: Deletions in the nucleoprotein. Yatsyshina S, et al. Infect Genet Evol 2015 Aug

More



Enseml Genome browser

Search: All species ▾ f

e.g. BRCA2 or rat 5:62797383-63

Browse a Genome

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.

Favourite genomes

Human GRCh38.p10

Mouse GRCm38.p5

Zebrafish GRCz10

Edit favourites

All genomes

-- Select a species -- ▾

View full list of all Ensembl species

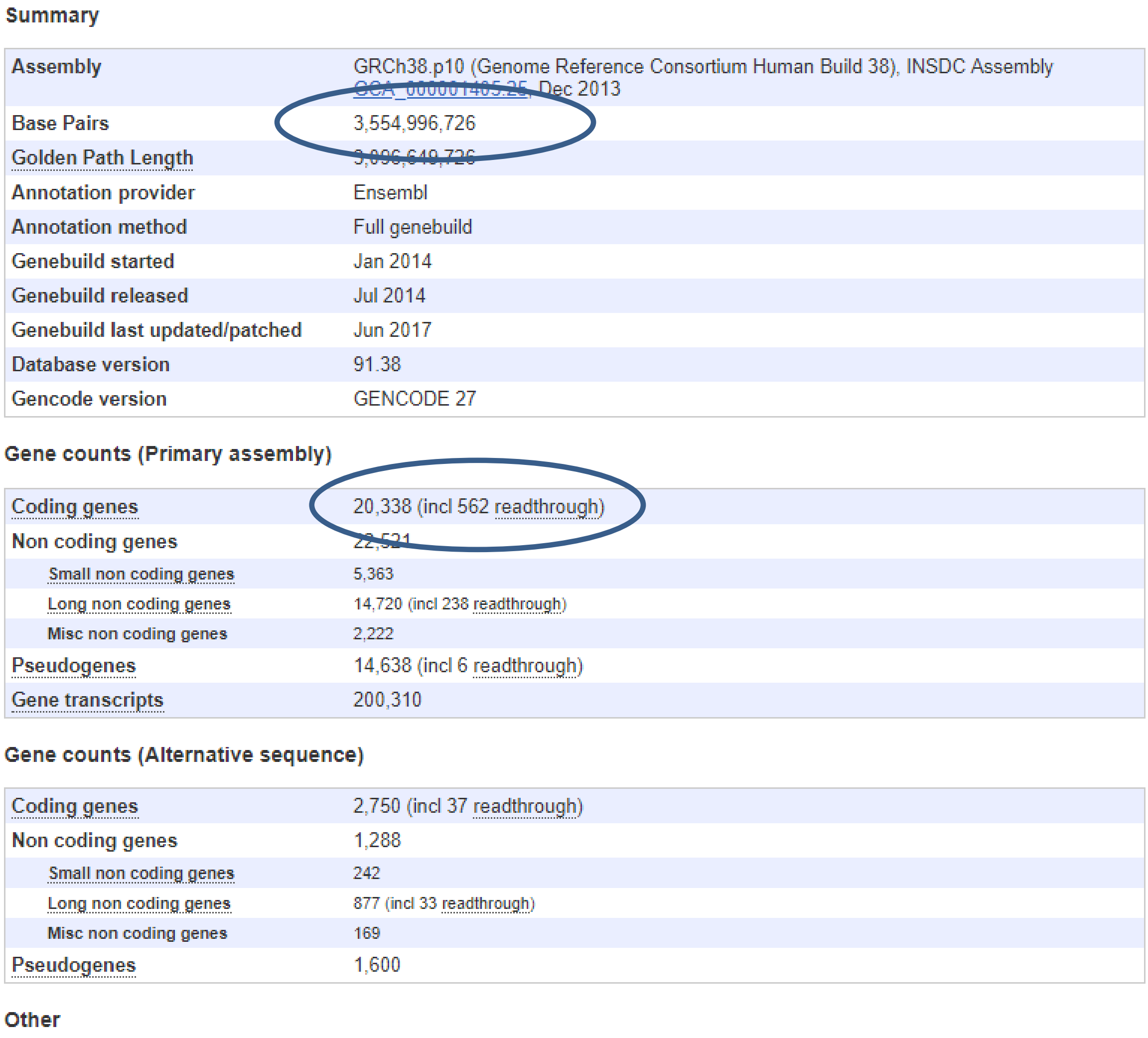
Other species are available in Ensembl *Pre!* and EnsemblGenomes

Bien qu'il y a une panoplie de genomes, nous allons nous interesser au genome Humain.

Genome Humain

http://useast.ensembl.org/Homo_sapiens/Info/Annotation

On peut voir que le génome humain est doté de plus de 3 000 000 000 de pairs de bases, avec 20 000 gènes codant et 22 000 gènes non codant



Summary

Assembly	GRCh38.p10 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.25, Dec 2013
Base Pairs	3,554,996,726
Golden Path Length	3,096,649,726
Annotation provider	Ensembl
Annotation method	Full genebuild
Genebuild started	Jan 2014
Genebuild released	Jul 2014
Genebuild last updated/patched	Jun 2017
Database version	91.38
Gencode version	GENCODE 27

Gene counts (Primary assembly)

Coding genes	20,338 (incl 562 readthrough)
Non coding genes	22,521
Small non coding genes	5,363
Long non coding genes	14,720 (incl 238 readthrough)
Misc non coding genes	2,222
Pseudogenes	14,638 (incl 6 readthrough)
Gene transcripts	200,310

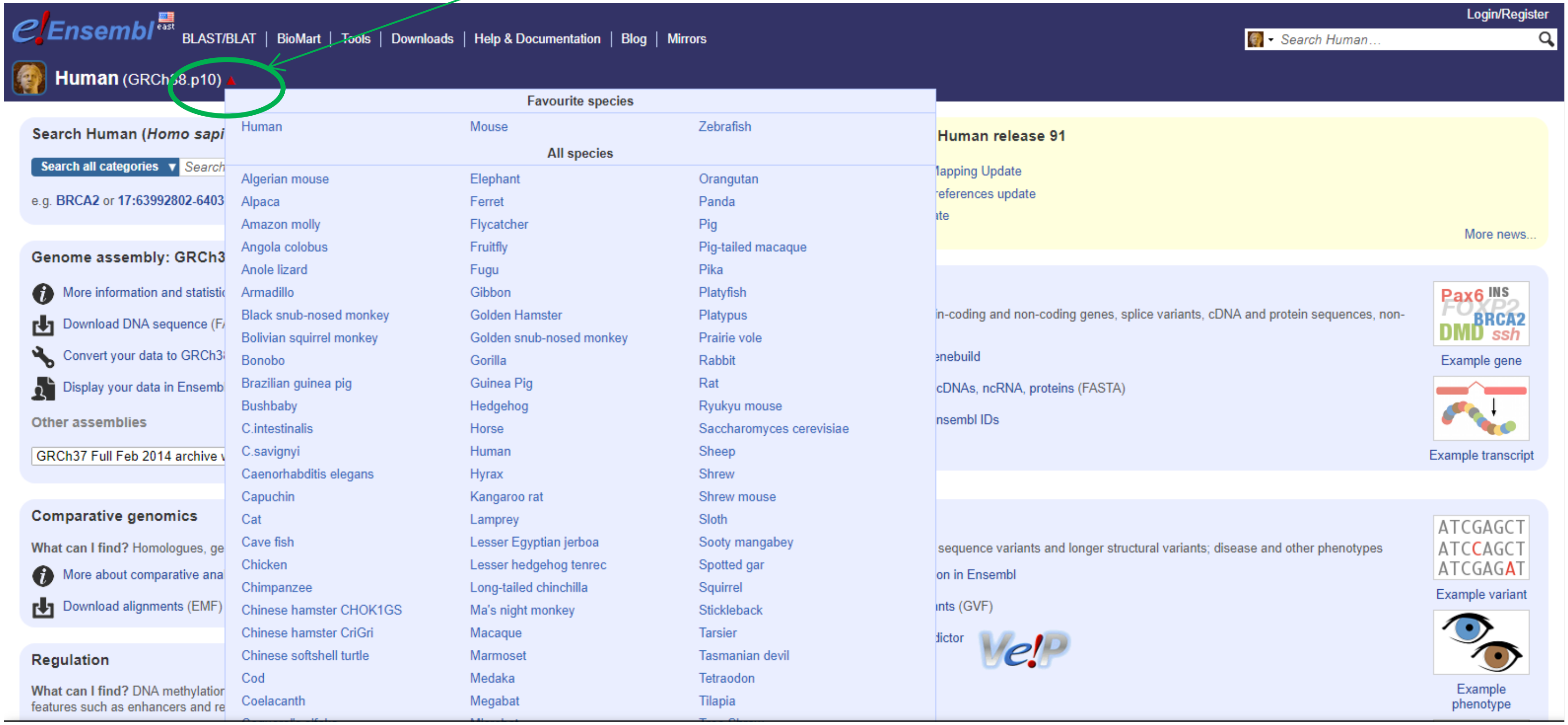
Gene counts (Alternative sequence)

Coding genes	2,750 (incl 37 readthrough)
Non coding genes	1,288
Small non coding genes	242
Long non coding genes	877 (incl 33 readthrough)
Misc non coding genes	169
Pseudogenes	1,600

Other

Genome Humain

Les autres génomes sont accéssibles en tout temps.



Favourite species

Human
Mouse
Zebrafish

All species

Algerian mouse
Alpaca
Amazon molly
Angola colobus
Anole lizard
Armadillo
Black snub-nosed monkey
Bolivian squirrel monkey
Bonobo
Brazilian guinea pig
Bushbaby
C.intestinalis
C.savignyi
Caenorhabditis elegans
Capuchin
Cat
Cave fish
Chicken
Chimpanzee
Chinese hamster CHOK1GS
Chinese hamster CriGri
Chinese softshell turtle
Cod
Coelacanth

Elephant
Ferret
Flycatcher
Fruitfly
Fugu
Gibbon
Golden Hamster
Golden snub-nosed monkey
Gorilla
Guinea Pig
Hedgehog
Horse
Human
Hyrax
Kangaroo rat
Lamprey
Lesser Egyptian jerboa
Lesser hedgehog tenrec
Long-tailed chinchilla
Ma's night monkey
Macaque
Marmoset
Medaka
Megabat

Orangutan
Panda
Pig
Pig-tailed macaque
Pika
Platyfish
Platypus
Prairie vole
Rabbit
Rat
Ryukyu mouse
Saccharomyces cerevisiae
Sheep
Shrew
Shrew mouse
Sloth
Sooty mangabey
Spotted gar
Squirrel
Stickleback
Tarsier
Tasmanian devil
Tetraodon
Tilapia

Search Human (*Homo sapi*

Search all categories ▾ Search

e.g. BRCA2 or 17:63992802-6403

Genome assembly: GRCh3

More information and statistic

Download DNA sequence (F

Convert your data to GRCh3

Display your data in Ensemb

Other assemblies

GRCh37 Full Feb 2014 archive

Comparative genomics

What can I find? Homologues, ge

More about comparative ana

Download alignments (EMF)

Regulation

What can I find? DNA methylatior features such as enhancers and re

Human release 91

lapping Update

eferences update

ite

More news…

n-coding and non-coding genes, splice variants, cDNA and protein sequences, non-

enebuild

cDNAs, ncRNA, proteins (FASTA)

nsembl IDs

Example gene

Example transcript

sequence variants and longer structural variants; disease and other phenotypes

on in Ensembl

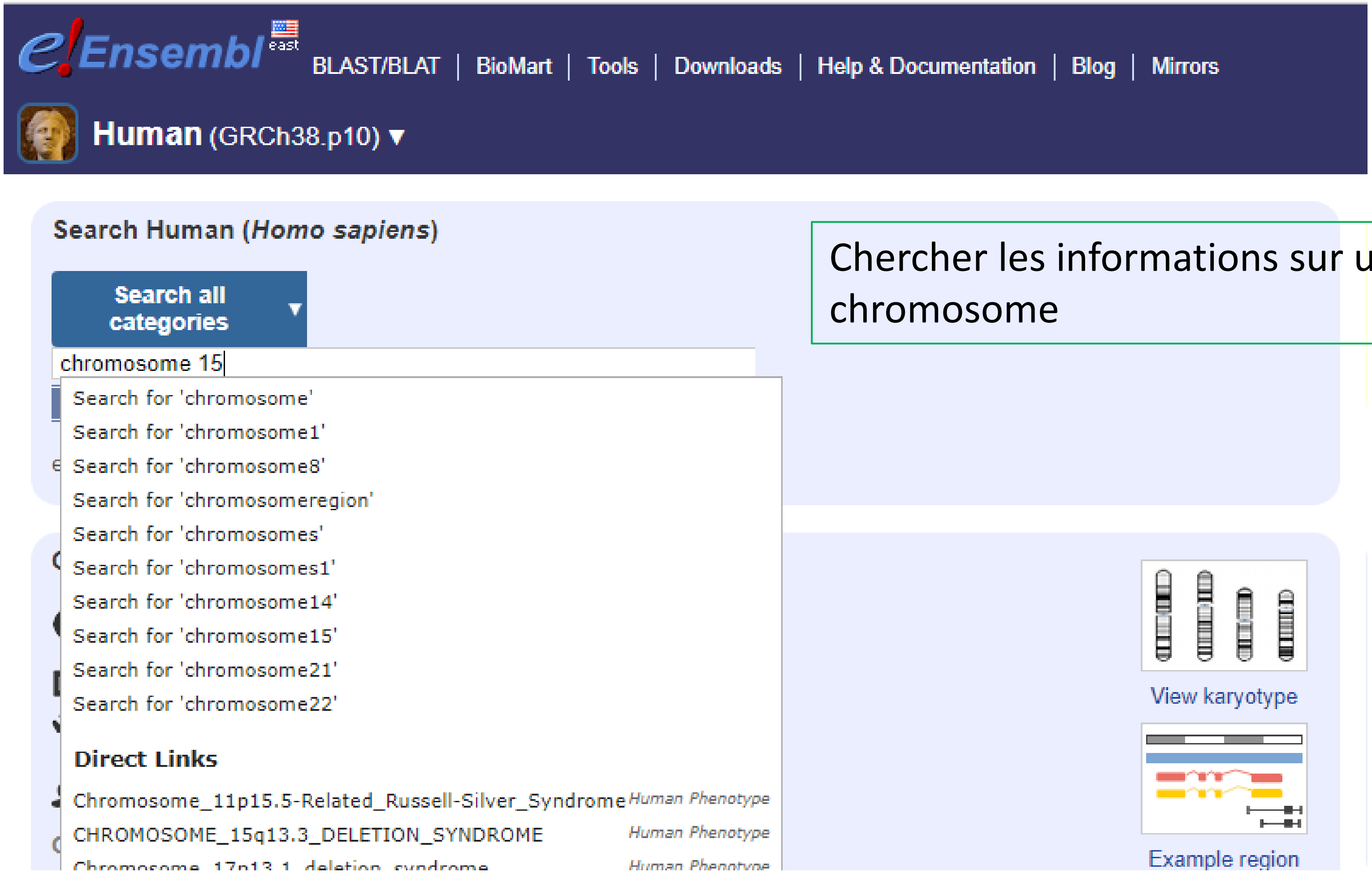
nts (GVF)

dictor

Example variant

Example phenotype

Genome Humain chromosome 15



Search Human (*Homo sapiens*)

chromosome 15

Search for 'chromosome'
Search for 'chromosome1'
Search for 'chromosome8'
Search for 'chromosomeregion'
Search for 'chromosomes'
Search for 'chromosomes1'
Search for 'chromosome14'
Search for 'chromosome15'
Search for 'chromosome21'
Search for 'chromosome22'

Direct Links

Chromosome_11p15.5-Related_Russell-Silver_Syndrome	*Human Phenotype*
CHROMOSOME_15q13.3_DELETION_SYNDROME	*Human Phenotype*
Chromosome_17p13.1_deletion_syndrome	*Human Phenotype*

Chercher les informations sur un chromosome

Human genome : Chromosome 15

Genome Humain : chromosome 15

Plus bas, vous verrez que le chromosome est doté de 100 000 000 de paires de bases.

BioMart

- L’outil Biomart a été mis au point pour repondre à des questions : Data mining.

Data set : Human gene.
Filters : pas mal de choses à filtrer
…

Voici de très bon tutoriels sur BioMart

https://www.youtube.com/watch?v=QvGT2G0-hYA

Usage de Perl

Pour ceux qui veulent rouler du code Perl, voici un lien qui vous sera super interessant.

http://useast.ensembl.org/info/docs/api/variation/variation_tutorial.html#structural