

Genomeeting 2016

Análisis de datos RNA-Seq



Sesión 1

Bases de Datos de Información Genómica

M.C. Patricia Oliver Ocaño

(patricia.oliver@wintergenomics.com)

LCG. Fernando Pérez Villatoro (fperez@wintergenomics.com)

Instituto Nacional de Medicina Genómica
Octubre 2016

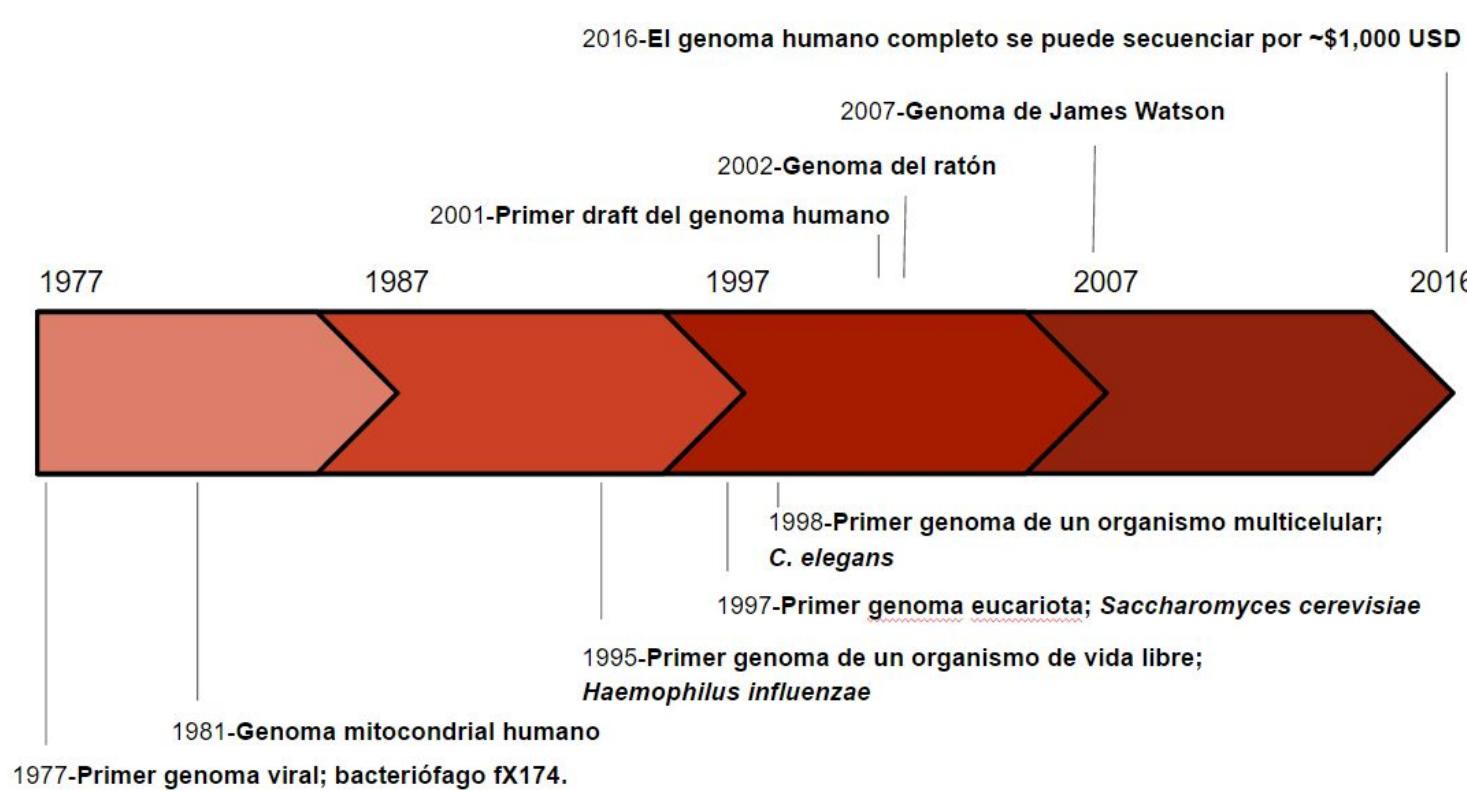
CONTENIDO

- **Genómica, el “Big Data” y la Bioinformática**
- **Introducción General sobre Bases de Datos**
- **Bases de datos en genómica**
- **Uso de ENSEMBL**

CONTENIDO

- **Genómica, el “Big Data” y la Bioinformática**
- Introducción General sobre Bases de Datos
- Bases de datos en genómica
- Uso de ENSEMBL

Timeline de los genomas secuenciados

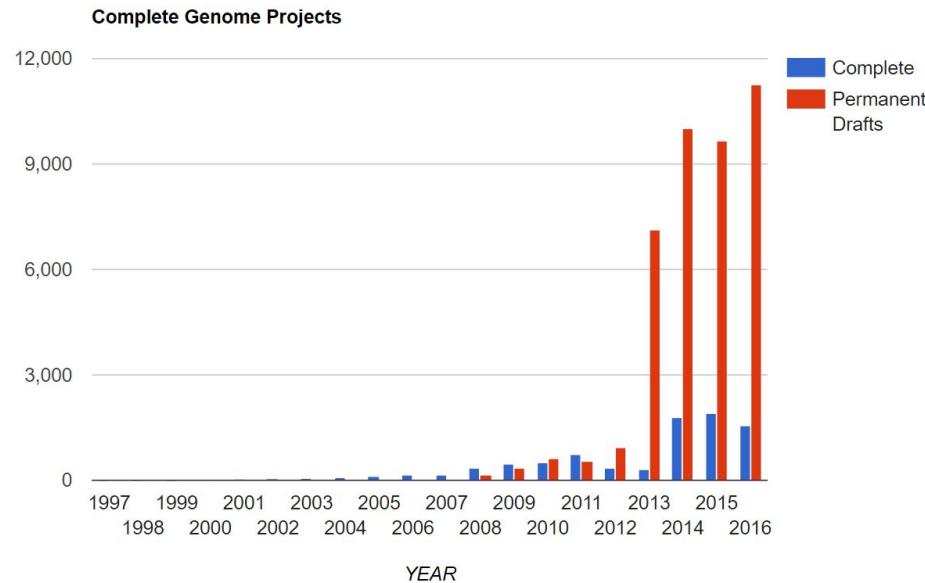


Genómica

Conjunto de ciencias y técnicas dedicadas al estudio integral del funcionamiento, contenido, evolución y origen de los genomas.



Crecimiento exponencial de datos genómicos



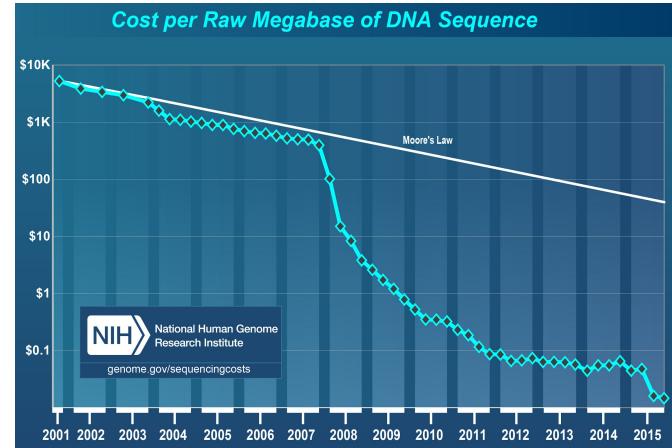
- Billones de datos que requieren ser curados, almacenados y ...

Genomes OnLine Database



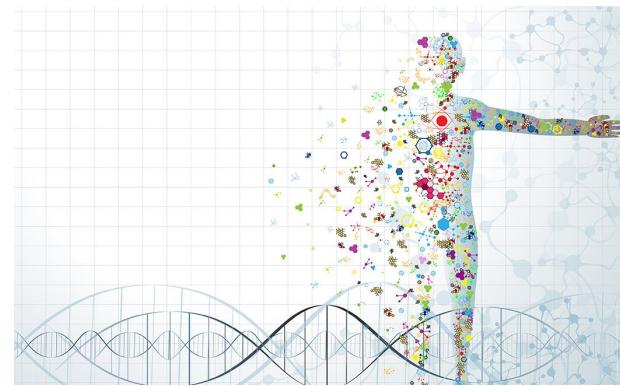
<https://gold.jgi.doe.gov/statistics>

Studies	<u>26.145</u>
Biosamples	<u>16.020</u>
Sequencing Projects	<u>98.395</u>
Analysis Projects	<u>79.681</u>
Organisms	<u>239.934</u>



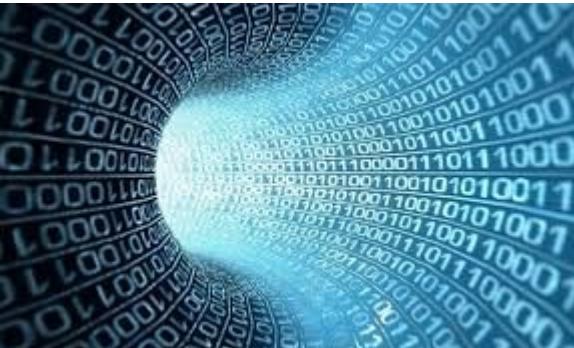
... Hacerlos fácilmente accesibles para continuar usándolos, realizar análisis y aportar nuevos conocimientos

El “Big Data” = Datos a gran escala

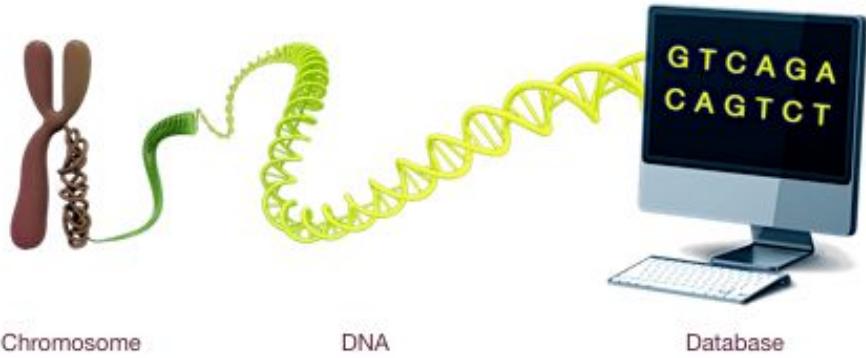


El “Big Data” en la Genómica

- Concepto que hace referencia al almacenamiento de grandes cantidades de datos Genómicos provenientes de tecnologías de secuenciación NGS y a los procedimientos usados para encontrar patrones repetitivos dentro de esos datos.



Bioinformática



Chromosome

DNA

Database

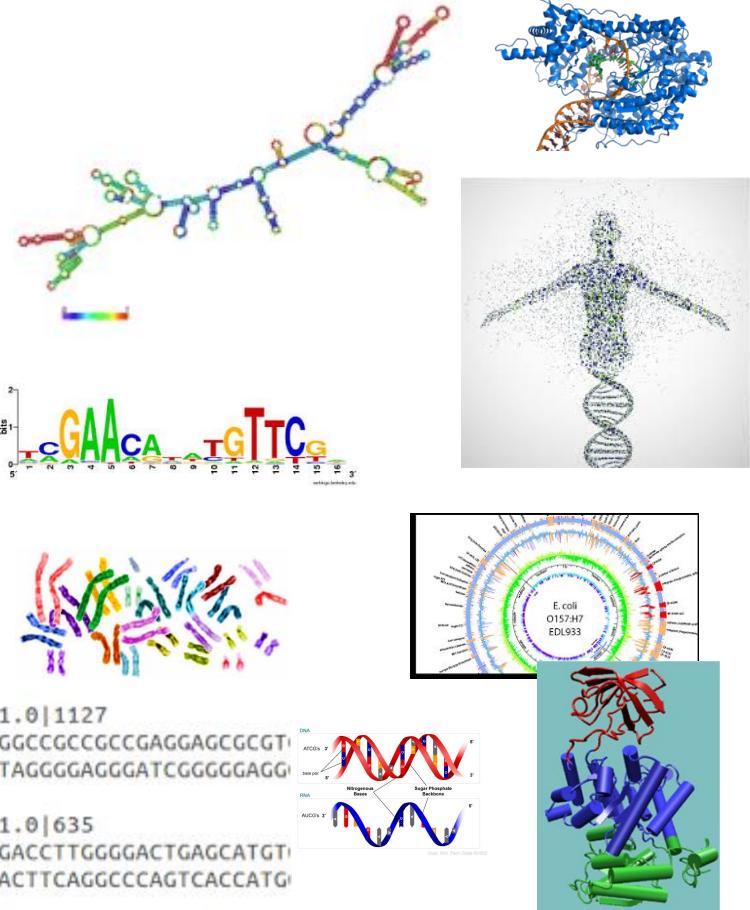
Herramientas para abordar diversos problemas bioinformáticos

- Aplicación de tecnologías computacionales a la gestión y análisis de datos biológicos



Tipos de datos en biología

- Secuencias nucleotídicas (DNA, RNA)
- Secuencias de proteínas
- Genomas
- Patrones de secuencias
- Estructuras
- etc.



Retos en la distribución de los datos

- Existen diferentes tipos de datos biológicos (Interatómicos, Nucleotídicos, Estructurales, Relacionales, Funcionales)
- Diversas tecnologías para obtener datos (NGS, Microarrays, MS, diferentes aplicaciones de NGS: WGS, RNASeq, ChIP-Seq, etc.)
- Se generan distintos formatos para organizar la información (fasta, fastq, vcf, gff, pdb, otros...)



Retos en la distribución de los datos

- Volumen de información (de orden de Tb o más)
- Redundancia de la información (p.ej. información duplicada entre bases de datos)
- Errores que pueden “heredarse” al procesar la información
- Anotaciones “inconsistentes”



CONTENIDO

- Genómica, el “Big Data” y la Bioinformática
- **Introducción General sobre Bases de Datos**
- Bases de datos en genómica
- Uso de ENSEMBL

¿Base de Datos?

Conjunto de datos organizado de tal modo que permitan obtener con rapidez diversos tipos de información (**DLE-RAE**)

Cuerpo organizado de información relacionada (**Oxford Dictionary**).

¿Sistema de Bases de Datos?

Un sistema de software que facilita la creación, mantenimiento y uso de una base de datos electrónica (**Oxford Dictionary**).

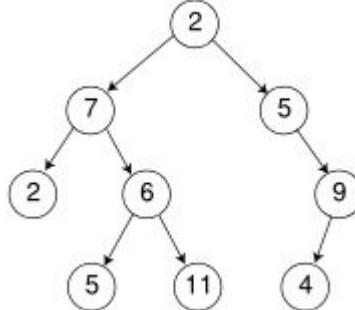


Modelos de Bases de Datos

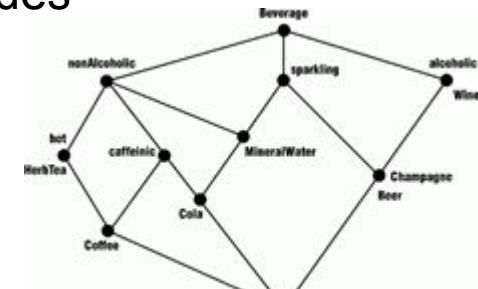
- Relacional

FIELD NAME	DESCRIPTION	LINE(S)	COL(S)	FIELD SIZE	USAGE	LOCATION
F1	Provider Number - Hospital	2	2	6	X	1 - 6
F2	Provider Number - Subprovider	3	2	6	X	7 - 12
F3	Provider Number - Subprovider	3.01	2	6	X	13 - 18
F4	Provider Number - Swing Bed SNF	4	2	6	X	19 - 24
F5	Provider Number - Swing Bed NF	5	2	6	X	25 - 30
F6	Provider Number - Hospital-Based SNF	6	2	6	X	31 - 36
F7	Provider Number - Hospital-Based NF	7	2	6	X	37 - 42
F7A	Provider Number - Hospital-Based ICF/MR	7.01	2	6	X	43 - 48
F8	Provider Number - Hospital-Based Hospice	9	2	6	X	49 - 57
F9	Provider Number - Separately Certified ASC I	11	2	6	X	58 - 60
F10	Provider Number - Hospital-Based Hospice I	12	2	6	X	61 - 66
F11	Provider Number - Hospital-Based ICF/MR I	14	2	6	X	67 - 72
F11A	Provider Number - Hospital-Based FQHC I	14.10	2	6	X	73 - 78
F12	Provider Number - Outpatient CORF I	15	2	6	X	79 - 84
F13	Provider Number - Outpatient CHHC I	15.10	2	6	X	85 - 90

- Jerárquico



- Redes



- Orientadas a objetos, Asociativas, Multidimensionales, etc.

Acceso dinámico de los datos mediante consultas

Clave



Ensembl Gene ID	Ensembl Transcript ID	Description	Chromosome	Associated Gene Name	Gene type
ENSG00000185231	ENST00000327606	melanocortin 2 receptor (adrenocorticotropic hormone) [Source:HGNC Symbol;Acc:HGNC:6930]	18	MC2R	protein_coding
ENSG00000140478	ENST00000434739	golgin A6 family member D [Source:HGNC Symbol;Acc:HGNC:32204]	15	GOLGA6D	protein_coding
ENSG00000161634	ENST00000293371	dermcidin [Source:HGNC Symbol;Acc:HGNC:14669]	12	DCD	protein_coding
ENSG00000254999	ENST00000530758	BRICK1, SCAR/WAVE actin-nucleating complex subunit [Source:HGNC Symbol;Acc:HGNC:23057]	3	BRK1	protein_coding
ENSG00000109501	ENST00000503569	Wolfram syndrome 1 (wolframin) [Source:HGNC Symbol;Acc:HGNC:12762]	4	WFS1	protein_coding
ENSG00000161634	ENST00000456047	dermcidin [Source:HGNC Symbol;Acc:HGNC:14669]	12	DCD	protein_coding
ENSG00000169271	ENST00000302005	heat shock protein family B (small) member 3 [Source:HGNC Symbol;Acc:HGNC:5248]	5	HSPB3	protein_coding
ENSG00000184697	ENST00000396925	claudin 6 [Source:HGNC Symbol;Acc:HGNC:2048]	16	CLDN6	protein_coding
ENSG00000156284	ENST00000399899	claudin 8 [Source:HGNC Symbol;Acc:HGNC:2050]	21	CLDN8	protein_coding
ENSG00000109501	ENST00000226760	Wolfram syndrome 1 (wolframin) [Source:HGNC Symbol;Acc:HGNC:12762]	4	WFS1	protein_coding
ENSG00000184697	ENST00000328796	claudin 6 [Source:HGNC Symbol;Acc:HGNC:2048]	16	CLDN6	protein_coding
ENSG00000150269	ENST00000279791	olfactory receptor family 5 subfamily M member 9 [Source:HGNC Symbol;Acc:HGNC:15294]	11	OR5M9	protein_coding
ENSG00000119801	ENST00000379520	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000119801	ENST00000379519	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000119801	ENST00000261353	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000119801	ENST00000402003	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000159231	ENST00000290354	carbonyl reductase 3 [Source:HGNC Symbol;Acc:HGNC:1549]	21	CBR3	protein_coding
ENSG00000171540	ENST00000306422	orthopedia homeobox [Source:HGNC Symbol;Acc:HGNC:8518]	5	OTP	protein_coding
ENSG00000119801	ENST00000402708	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000245848	ENST00000498907	CCAAT/enhancer binding protein (C/EBP), alpha [Source:HGNC Symbol;Acc:HGNC:1833]	19	CEBPA	protein_coding
ENSG00000132429	ENST00000254765	popeye domain containing 3 [Source:HGNC Symbol;Acc:HGNC:17649]	6	POPDC3	protein_coding
ENSG00000125971	ENST00000357156	dynein, light chain, roadblock-type 1 [Source:HGNC Symbol;Acc:HGNC:15468]	20	DYNLRB1	protein_coding
ENSG00000186977	ENST00000334151	keratin associated protein 19-5 [Source:HGNC Symbol;Acc:HGNC:18940]	21	KRTAP19-5	protein_coding
ENSG00000178591	ENST00000382410	defensin beta 125 [Source:HGNC Symbol;Acc:HGNC:18105]	20	DEFB125	protein_coding

Acceso dinámico de los datos

clave



Información asociada al gen



Ensembl Gene ID	Ensembl Transcript ID	Description	Chromosome	Associated Gene Name	Gene type
ENSG00000185231	ENST00000327606	melanocortin 2 receptor (adrenocorticotropic hormone) [Source:HGNC Symbol;Acc:HGNC:6930]	18	MC2R	protein_coding
ENSG00000140478	ENST00000434739	golgin A6 family member D [Source:HGNC Symbol;Acc:HGNC:32204]	15	GOLGA6D	protein_coding
ENSG00000161634	ENST00000293371	dermcidin [Source:HGNC Symbol;Acc:HGNC:14669]	12	DCD	protein_coding
ENSG00000254999	ENST00000530758	BRICK1, SCAR/WAVE actin-nucleating complex subunit [Source:HGNC Symbol;Acc:HGNC:23057]	3	BRK1	protein_coding
ENSG00000109501	ENST00000503569	Wolfram syndrome 1 (wolframin) [Source:HGNC Symbol;Acc:HGNC:12762]	4	WFS1	protein_coding
ENSG00000161634	ENST00000456047	dermcidin [Source:HGNC Symbol;Acc:HGNC:14669]	12	DCD	protein_coding
ENSG00000169271	ENST00000302005	heat shock protein family B (small) member 3 [Source:HGNC Symbol;Acc:HGNC:5248]	5	HSPB3	protein_coding
ENSG00000184697	ENST00000396925	claudin 6 [Source:HGNC Symbol;Acc:HGNC:2048]	16	CLDN6	protein_coding
ENSG00000156284	ENST00000399899	claudin 8 [Source:HGNC Symbol;Acc:HGNC:2050]	21	CLDN8	protein_coding
ENSG00000109501	ENST00000226760	Wolfram syndrome 1 (wolframin) [Source:HGNC Symbol;Acc:HGNC:12762]	4	WFS1	protein_coding
ENSG00000184697	ENST00000328796	claudin 6 [Source:HGNC Symbol;Acc:HGNC:2048]	16	CLDN6	protein_coding
ENSG00000150269	ENST00000279791	olfactory receptor family 5 subfamily M member 9 [Source:HGNC Symbol;Acc:HGNC:15294]	11	OR5M9	protein_coding
ENSG00000119801	ENST00000379520	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000119801	ENST00000379519	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000119801	ENST00000261353	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000119801	ENST00000402003	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000159231	ENST00000290354	carbonyl reductase 3 [Source:HGNC Symbol;Acc:HGNC:1549]	21	CBR3	protein_coding
ENSG00000171540	ENST00000306422	orthopedia homeobox [Source:HGNC Symbol;Acc:HGNC:8518]	5	OTP	protein_coding
ENSG00000119801	ENST00000402708	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000245848	ENST00000498907	CCAAT/enhancer binding protein (C/EBP), alpha [Source:HGNC Symbol;Acc:HGNC:1833]	19	CEBPA	protein_coding
ENSG00000132429	ENST00000254765	popeye domain containing 3 [Source:HGNC Symbol;Acc:HGNC:17649]	6	POPDC3	protein_coding
ENSG00000125971	ENST00000357156	dynein, light chain, roadblock-type 1 [Source:HGNC Symbol;Acc:HGNC:15468]	20	DYNLRB1	protein_coding
ENSG00000186977	ENST00000334151	keratin associated protein 19-5 [Source:HGNC Symbol;Acc:HGNC:18940]	21	KRTAP19-5	protein_coding
ENSG00000178591	ENST00000382410	defensin beta 125 [Source:HGNC Symbol;Acc:HGNC:18105]	20	DEFB125	protein_coding

Acceso dinámico de los datos

clave



Información asociada al gen



Ensembl Gene ID	Ensembl Transcript ID	Description	Chromosome	Associated Gene Name	Gene type
ENSG00000185231	ENST00000327606	melanocortin 2 receptor (adrenocorticotropic hormone) [Source:HGNC Symbol;Acc:HGNC:6930]	18	MC2R	protein_coding
ENSG00000140478	ENST00000434739	golgin A6 family member D [Source:HGNC Symbol;Acc:HGNC:32204]	15	GOLGA6D	protein_coding
ENSG00000161634	ENST00000293371	dermcidin [Source:HGNC Symbol;Acc:HGNC:14669]	12	DCD	protein_coding
ENSG00000254999	ENST00000530758	BRICK1, SCAR/WAVE actin-nucleating complex subunit [Source:HGNC Symbol;Acc:HGNC:23057]	3	BRK1	protein_coding
ENSG00000109501	ENST00000503569	Wolfram syndrome 1 (wolframin) [Source:HGNC Symbol;Acc:HGNC:12762]	4	WFS1	protein_coding
ENSG00000161634	ENST00000456047	dermcidin [Source:HGNC Symbol;Acc:HGNC:14669]	12	DCD	protein_coding
ENSG00000169271	ENST00000302005	heat shock protein family B (small) member 3 [Source:HGNC Symbol;Acc:HGNC:5248]	5	HSPB3	protein_coding
ENSG00000184697	ENST00000396925	claudin 6 [Source:HGNC Symbol;Acc:HGNC:2048]	16	CLDN6	protein_coding
ENSG00000156284	ENST00000399899	claudin 8 [Source:HGNC Symbol;Acc:HGNC:2050]	21	CLDN8	protein_coding
ENSG00000109501	ENST00000226760	Wolfram syndrome 1 (wolframin) [Source:HGNC Symbol;Acc:HGNC:12762]	4	WFS1	protein_coding
ENSG00000184697	ENST00000328796	claudin 6 [Source:HGNC Symbol;Acc:HGNC:2048]	16	CLDN6	protein_coding
ENSG00000150269	ENST00000279791	olfactory receptor family 5 subfamily M member 9 [Source:HGNC Symbol;Acc:HGNC:15294]	11	OR5M9	protein_coding
ENSG00000119801	ENST00000379520	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000119801	ENST00000379519	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000119801	ENST00000261353	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000119801	ENST00000402003	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000159231	ENST00000290354	carbonyl reductase 3 [Source:HGNC Symbol;Acc:HGNC:1549]	21	CBR3	protein_coding
ENSG00000171540	ENST00000306422	orthopedia homeobox [Source:HGNC Symbol;Acc:HGNC:8518]	5	OTP	protein_coding
ENSG00000119801	ENST00000402708	yippee like 5 [Source:HGNC Symbol;Acc:HGNC:18329]	2	YPEL5	protein_coding
ENSG00000245848	ENST00000498907	CCAAT/enhancer binding protein (C/EBP), alpha [Source:HGNC Symbol;Acc:HGNC:1833]	19	CEBPA	protein_coding
ENSG00000132429	ENST00000254765	popeye domain containing 3 [Source:HGNC Symbol;Acc:HGNC:17649]	6	POPDC3	protein_coding
ENSG00000125971	ENST00000357156	dynein, light chain, roadblock-type 1 [Source:HGNC Symbol;Acc:HGNC:15468]	20	DYNLRB1	protein_coding
ENSG00000186977	ENST00000334151	keratin associated protein 19-5 [Source:HGNC Symbol;Acc:HGNC:18940]	21	KRTAP19-5	protein_coding
ENSG00000178591	ENST00000382410	defensin beta 125 [Source:HGNC Symbol;Acc:HGNC:18105]	20	DEFB125	protein_coding

CONTENIDO

- Genómica, el “Big Data” y la Bioinformática
- Introducción General sobre Bases de Datos
- **Bases de datos en genómica**
- Uso de ENSEMBL

Categorías de Bases de Datos Biológicas



- Ácidos nucléicos (RNA, DNA)
- Proteínas
- Motivos
- Dominios de proteínas
- Estructuras
- Genomas
- Genes
- Perfiles de expresión
- SNPs y mutaciones
- Vías metabólicas
- Taxonomía

- **Repositorios**
- **Proyectos**
- **Genome Browsers**

La colección de bases de datos de Biología Molecular de la NAR (*journal Nucleic Acids Research*), reporta que en 2014, había 1552 DB públicamente accesibles online.

Genomics, Proteomics & Bioinformatics

Volume 13, Issue 1, February 2015, Pages 55–63

Biological Databases for Human Research

Dong Zou^a, Lina Ma^{a, b}, Jun Yu^{a, c}, Zhang Zhang^{a, d}

● Repostorios

NCBI(GenBank),
EMBL-EBI, DDBJ,
INSDC,RefSeq

NCBI Resources How To Sign in to NCBI

All Databases Search

NCBI National Center for Biotechnology Information

NCBI Home Resource List (A-Z) All Resources Chemicals & Bioassays Data & Software DNA & RNA Domains & Structures Genes & Expression Genetics & Medicine Genomes & Maps Homology Literature Proteins Sequence Analysis Taxonomy Training & Tutorials Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News | Blog

Submit Deposit data or manuscripts into NCBI databases

Download Transfer NCBI data to your computer

Learn Find help documents, attend a class or watch a tutorial

Develop Use NCBI APIs and code libraries to build applications

Analyze Identify an NCBI tool for your data analysis task

Research Explore NCBI research and collaborative projects

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI Announcements

October 26th NCBI Minute: New BLAST Databases Provide Cleaner Results 11 Oct 2016

On October 26th, NCBI staff will introduce two new BLAST databases: the Multiple Sequence Alignment Viewer 1.1 is now available 06 Oct 2016

The NCBI Multiple Sequence Alignment Viewer (MSA Viewer) has recently been NLM presents "Insider's Guide" webinar on E-utilities and PubMed on October 19th 06 Oct 2016

On Wednesday, October 19, 2016 the [More...](#)

● Proyectos

- ENCODE
- 1000 Genomes
- HapMap
- Etc...

Índice de /1000genomes/ftp/release/20130502/

	Nombre	Tamaño	Fecha de modificación
📁	[directorio principal]		
📄	20140625_related_individuals.txt	1.1 kB	25/06/14 00:00:00
📄	ALL.chr1.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	1.1 GB	27/05/15 00:00:00
📄	ALL.chr1.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	219 kB	27/05/15 00:00:00
📄	ALL.chr10.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	738 MB	27/05/15 00:00:00
📄	ALL.chr10.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	130 kB	27/05/15 00:00:00
📄	ALL.chr11.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	732 MB	27/05/15 00:00:00
📄	ALL.chr11.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	130 kB	27/05/15 00:00:00
📄	ALL.chr12.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	706 MB	27/05/15 00:00:00
📄	ALL.chr12.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	130 kB	27/05/15 00:00:00
📄	ALL.chr13.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	531 MB	27/05/15 00:00:00
📄	ALL.chr13.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	94.4 kB	27/05/15 00:00:00
📄	ALL.chr14.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	483 MB	27/05/15 00:00:00
📄	ALL.chr14.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	86.3 kB	27/05/15 00:00:00
📄	ALL.chr15.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	437 MB	27/05/15 00:00:00
📄	ALL.chr15.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	79.7 kB	27/05/15 00:00:00
📄	ALL.chr16.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	472 MB	27/05/15 00:00:00
📄	ALL.chr16.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	79.1 kB	27/05/15 00:00:00
📄	ALL.chr17.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	414 MB	27/05/15 00:00:00
📄	ALL.chr17.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	77.3 kB	27/05/15 00:00:00
📄	ALL.chr18.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	416 MB	27/05/15 00:00:00
📄	ALL.chr18.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	72.9 kB	27/05/15 00:00:00
📄	ALL.chr19.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	343 MB	27/05/15 00:00:00
📄	ALL.chr19.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	54.2 kB	27/05/15 00:00:00
📄	ALL.chr2.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	1.2 GB	27/05/15 00:00:00
📄	ALL.chr2.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	232 kB	27/05/15 00:00:00
📄	ALL.chr20.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz	326 MB	27/05/15 00:00:00
📄	ALL.chr20.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz.tbi	55.2 kB	27/05/15 00:00:00

DB especializadas en ciertos datos

- PDB, DB de estructuras de proteínas en 3D
- Pfam, familias de proteínas
- ArrayExpress, repositorio para datos de transcriptómica
- dbGaP
- HuGE Navigator

Específicos para Especies y Taxa

- Rat Genome DB
- Flybase (*Drosophila* y otras especies)
- Wormbase (*C.elegans* y nemátodos relacionados)
- TAIR, *Arabidopsis*
- SGD, *Sacharomyces* Genome DB
- E.coliHub, *E.coli* DB

NCBI (National Center for Biotechnology Information)

- NCBI está registrado como un repositorio de datos de investigación.
- Parte de la Librería Nacional de Medicina de los Estados Unidos (NLM) del NIH (National Institute of Health).
- Se fundó en 1988 en la ciudad de Bethesda, Maryland.
- Alberga una serie de DB muy relevantes para la biotecnología y la biomedicina como GenBank y PubMed.
- Fuente de herramientas y servicios para los bioinformáticos



NCBI

[NCBI Home](#)[Resource List \(A-Z\)](#)[All Resources](#)[Chemicals & Bioassays](#)[Data & Software](#)[DNA & RNA](#)[Domains & Structures](#)[Genes & Expression](#)[Genetics & Medicine](#)[Genomes & Maps](#)[Homology](#)[Literature](#)[Proteins](#)[Sequence Analysis](#)[Taxonomy](#)[Training & Tutorials](#)[Variation](#)

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News](#) | [Blog](#)

Submit

Deposit data or manuscripts into NCBI databases



Download

Transfer NCBI data to your computer



Learn

Find help documents, attend a class or watch a tutorial



Develop

Use NCBI APIs and code libraries to build applications



Analyze

Identify an NCBI tool for your data analysis task



Research

Explore NCBI research and collaborative projects



Popular Resources

[PubMed](#)[Bookshelf](#)[PubMed Central](#)[PubMed Health](#)[BLAST](#)[Nucleotide](#)[Genome](#)[SNP](#)[Gene](#)[Protein](#)[PubChem](#)

NCBI Announcements

October 26th NCBI Minute: New BLAST Databases Provide Cleaner Results

11 Oct 2016

On October 26th, NCBI staff will introduce two new BLAST databases: the

Multiple Sequence Alignment Viewer 1.1 is now available

06 Oct 2016

The NCBI Multiple Sequence Alignment Viewer (MSA Viewer) has recently been

NLM presents "Insider's Guide" webinar on E-utilities and PubMed on October 19th

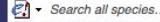
06 Oct 2016

On Wednesday, October 19, 2016, the

[More...](#)

● Genome Browser Ensembl

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Login/Register  

Search: All species  for

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

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.

Popular genomes

 Human GRCh38.p7	 Human GRCh37
 Mouse GRCm38.p4	 Zebrafish GRCz10

[★ Log in to customize this list](#)

All genomes

-- Select a species -- 

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

Other species are available in [Ensembl Pre!](#) and [EnsemblGenomes](#).

Still using Human GRCh37?

[Go to e!GRCh37](#)

Variant Effect Predictor



Gene expression in different tissues



Find SNPs and other variants for my gene

GTATACATTC
CRTBAAAGTTT
CTTCTAATTCT
GRAACATITCC

Retrieve gene sequence

GGCTGACTTCCC GG G TGG
GG G CTT TGTGG GCG GAGC
GG G CTC TGT GCT GCG CCT
A G G G A C A G A T T T G T G A
C A C C T C T G A G C G G T T G
C C C A G T C C A G C G T G G C G

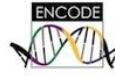
Compare genes across species



Use my own data in Ensembl



ENCODE data in Ensembl



What's New in Ensembl Release 85 (July 2016)

- Update to Ensembl-Havana human GENCODE gene set (release 25)
- 30 new epigenomes from the Roadmap Epigenomics Project
- New zebrafish rnaseq
- Update to Rat Ensembl-Havana gene set
- Mouse: update to Ensembl-Havana GENCODE gene set

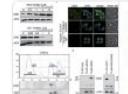
[Full details](#) | [All web updates, by release](#) | [More news on our blog](#)

- 10 Mar 2016: [Ensembl 84 has been released!](#)
- 16 Feb 2016: [Learn about Ensembl – online, live and free!](#)
- 25 Jan 2016: [Sharing feature on the new mobile site \(\[m.ensembl.org\]\(#\)\)](#)

[Go to Ensembl blog](#)

Tweets by @ensembl

e! Ensembl @ensembl
Novel O-GlcNAcylation of H2ASer40 is linked to the evolution of placental animals #UsingEnsembl @SciReports buff.ly/2cRgQ4R



Novel O-GlcNAcylation on Ser40 of...
We report here newly discovered O-linked-N-acetylglucosamine (O-GlcNAc) nature.com



Search: for

e.g. [CPR34](#) or [chitin*](#)

Popular genomes

**Caenorhabditis elegans**

WBcel235

**Anopheles gambiae**

AgamP4

**Drosophila melanogaster**

BDGP6

**Trichoplax adhaerens**

ASM15027v1

**Amphimedon queenslandica**

Aquila

**Nasonia vitripennis**

GCA_000002325.2

[Log in to customize this list](#)

All genomes

-- Select a species --

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

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

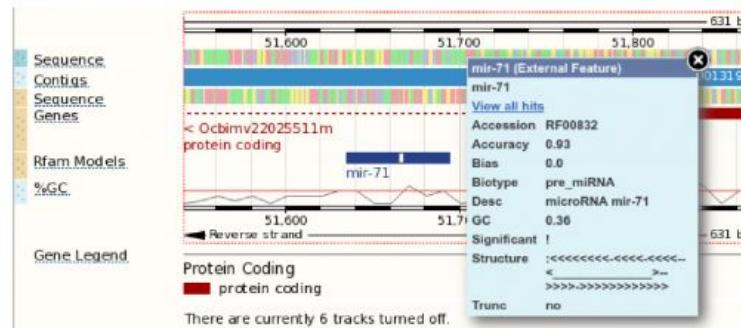


RNA gene alignments

Rfam covariance models (version 12.1) have been aligned against all metazoan genomes, with the exception of tRNA models which are aligned using tRNAscan-SE.

The covariance models are taxonomically filtered before alignment, so that structural RNA features that have never been annotated in metazoan species are not inappropriately aligned. The alignment is performed with cmmscan, from the Infernal software suite.

RNA gene alignments are visible as browser tracks ('Rfam models' and 'tRNA models'); selecting an alignment in the genome browser displays metadata such as the description and the secondary structure.



What's New in Release 32

- Updated genomes
 - *Caenorhabditis japonica* ([WormBase](#) WS250)

Did you know...?

Data in CRAM format can



Search: for

e.g. [Carboxy*](#) or [chx28](#)

Popular genomes



Arabidopsis thaliana

TAIR10



Oryza sativa Japonica

IRGSP-1.0



Triticum aestivum

TGACv1



Hordeum vulgare

ASM32608v1



Zea mays

AGPv4



Physcomitrella patens

ASM242v1

[Log in to customize this list](#)

All genomes

-- Select a species --

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

What's New in Release 32

- New genomes
 - [Beta vulgaris](#)
 - [Brassica napus](#)

Did you know...?

You can search the

[Track](#)

New Bread Wheat Genome Assembly

A new genome assembly of *Triticum aestivum* cv. Chinese Spring is now available in Ensembl Plants. The assembly (TGACv1) and its accompanying annotation was produced by the [Earlham Institute](#), formerly The Centre for Genome Analysis (TGAC), as part of the [Triticeae Genomics for Sustainable Agriculture](#) project.

The assembly has a scaffold N50 of 88 Kbp and a total length of 13.4 Gbp in contigs greater than 500 bp ([read more](#)). The gene model annotation consists of 217,907 loci and 273,739 transcripts. A total of 104,09 protein coding genes (154,798 transcripts) and 10,156 long ncRNAs have been annotated with high confidence ([read more](#)). Approximately 99,000 genes (99% of the total) annotated on the previous IWGSC CSS assembly (MIPS) have been mapped to the new assembly.

The Axiom 35k and 820k SNP marker sets have been provided by [CerealsDB](#) and located on the new assembly ([read more](#)).

Ensembl Plants Archive Site

Alongside release 32 we have launched a new [archive site](#), where we will keep selected previous releases of Ensembl Plants publicly available. The first release available on the archive site is release 31, and includes the previous assemblies for wheat and maize.

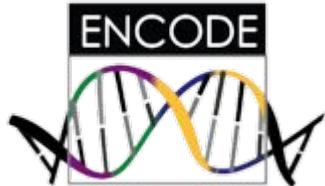
Ensembl Plants is developed in coordination with other plant genomics and bioinformatics groups via the EBI's role in the [transPLANT](#) consortium. The transPLANT project is funded by the [European Commission](#) within its [7th Framework Programme](#), under the thematic area "Infrastructures", contract number [283496](#).



Wheat genomics resources are developed as part of our involvement in the consortium [Triticeae Genomics For Sustainable Agriculture](#). Barley genomics resources are funded through the [UK barley](#)

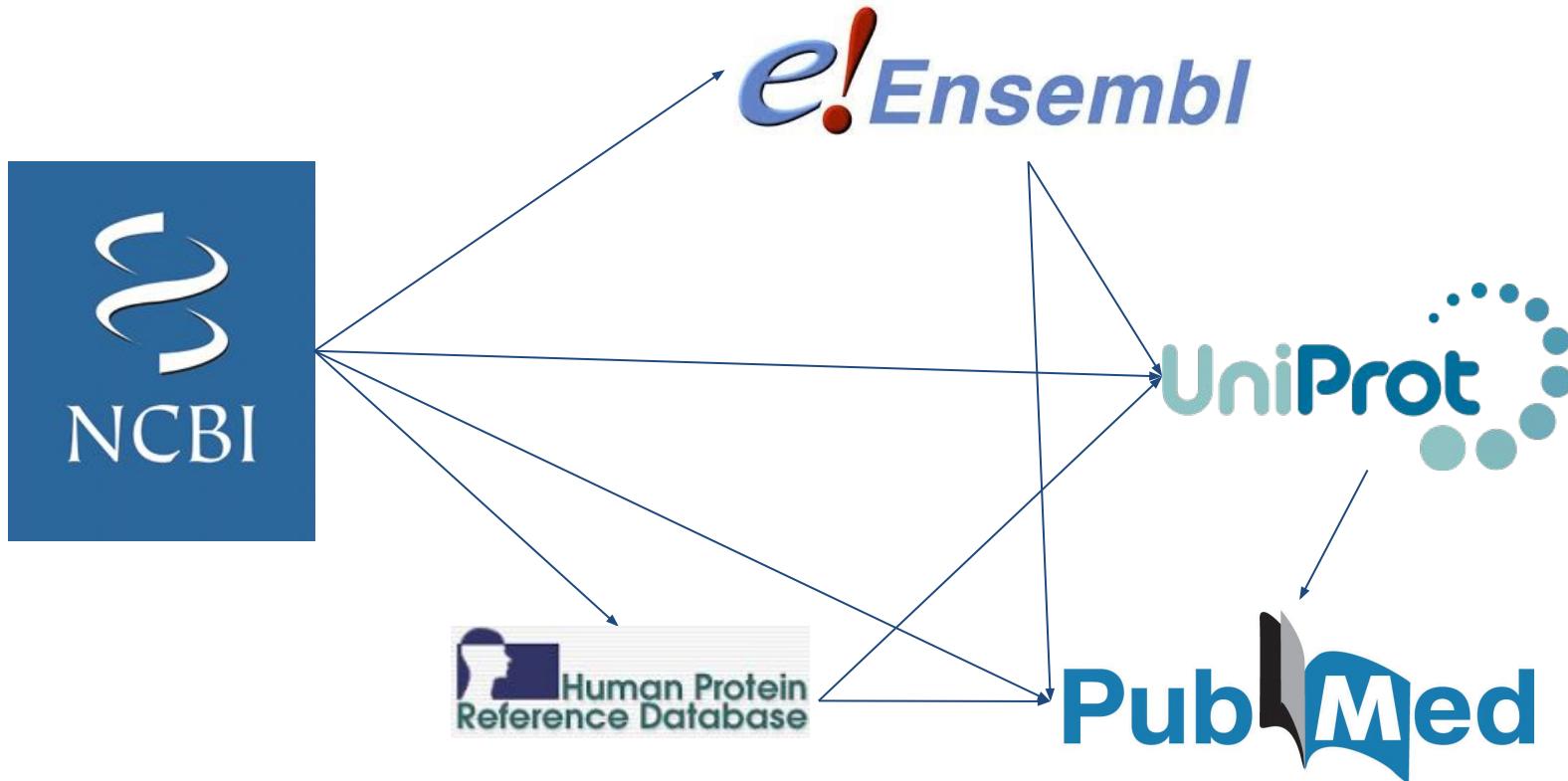


- **ClinVar** es un archivo de acceso público que reporta las relaciones entre las variaciones humanas y los fenotipos, basándose en evidencia de soporte



- **ENCODE (Encyclopedia of DNA Elements)** es un proyecto de investigación que se lleva cabo por [National Human Genome Research Institute](#) (NHGRI) de los Estados Unidos desde September 2003.
- Para dar seguimiento al Proyecto Genoma Humano (Investigación Genómica), tiene como objetivo identificar todos los elementos funcionales en el genoma humano.
- Consiste en un consorcio mundial de grupos de investigación, y los datos generados se pueden acceder a través de DB públicas.

Bases de datos interconectadas



Bases de datos interconectadas

i.nlm.nih.gov

NCBI Resources How To Sign in to NCBI

All Databases ▾ BRCA2

All Databases

- Assembly
- BioProject
- BioSample
- BioSystems
- Books
- ClinVar
- Clone
- Conserved Domains
- dbGaP
- dbVar
- EST
- Gene
- Genome
- GEO DataSets
- GEO Profiles
- GSS
- GTR
- HomoloGene
- MedGen

Submit or manuscripts to NCBI

Download Transfer NCBI data to your computer

Learn Find help documents, attend a class or watch a tutorial

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

Develop Use NCBI APIs and code libraries to build applications

Analyze Identify an NCBI tool for your data analysis task

Research Explore NCBI research and collaborative projects

NCBI Announcements

October 26th NCBI Minute: New BLAST Databases Provide Cleaner Results
11 Oct 2016

On October 26th, NCBI staff will introduce two new BLAST databases: the

Multiple Sequence Alignment Viewer 1.1 is now available
06 Oct 2016

The NCBI Multiple Sequence Alignment Viewer (MSA Viewer) has recently been

NLM presents "Insider's Guide" webinar on E-utilities and PubMed on October 19th
06 Oct 2016

On Wednesday, October 19, 2016, the

More...

Bases de datos interconectadas

i.nlm.nih.gov

NCBI Resources How To Sign in to NCBI

NCBI National Center for Biotechnology Information Gene BRCA2 Search

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News](#) | [Blog](#)

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Popular Resources

PubMed
Bookshelf
PubMed Central
PubMed Health
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

NCBI Announcements

October 26th NCBI Minute: New BLAST Databases Provide Cleaner Results
11 Oct 2016
On October 26th, NCBI staff will introduce two new BLAST databases: the Multiple Sequence Alignment Viewer 1.1 is now available

06 Oct 2016
The NCBI Multiple Sequence Alignment Viewer (MSA Viewer) has recently been

NLM presents "Insider's Guide" webinar on E-utilities and PubMed on October 19th
06 Oct 2016
On Wednesday, October 19, 2016, the

[More...](#)

Bases de datos interconectadas

NCBI Resources How To Sign in to NCBI

Gene Gene BRCA2 Search Help

Gene sources Create RSS Create alert Advanced

Send to: [Email](#) [Print](#) [Copy](#) [Report](#) [Help](#)

Gene sources 20 per page Sort by Relevance

Categories Alternatively spliced Annotated genes Protein-coding

Sequence content CCDS Ensembl RefSeq RefSeqGene

Status Current Chromosome locations more... Clear all Show additional filters

Search results Items: 1 to 20 of 200 See also 2 discontinued or replaced items.

Send to: First < Prev Page 1 of 10 Next > Last >>

Filters: Manage Filters

Results by taxon

Top Organisms [Tree]

- Homo sapiens (1)
- Pan troglodytes (1)
- Pan paniscus (1)
- Gorilla gorilla (1)
- Pongo abelii (1)
- All other taxa (195)

More...

Find related data Database: Select

Find Items

Search details

BRCA2[sym] AND alive[prop]

Search See more...

Recent activity

Turn Off Clear

BRCA2[sym] AND (alive[prop]) (200) Gene

BRCA2 AND (alive[prop]) (1697) Gene

BCRA[sym] AND (alive[prop]) (9) Gene

Name/Gene ID	Description	Location	Aliases	MIM
BRCA2 ID: 675	BRCA2, DNA repair associated [Homo sapiens (human)]	Chromosome 13, NC_000013.11 (32315480..32399672)	BRCC2, BROVCA2, FACD, FAD, FAD1, FANCD, FANCD1, GLM3, PNCA2, XRCC11	600185
Brca2 ID: 12190	breast cancer 2, early onset [Mus musculus (house mouse)]	Chromosome 5, NC_000071.6 (150522021..150570147)	Fancd1, RAB163	
Brca2 ID: 360254	BRCA2, DNA repair associated [Rattus norvegicus (Norway rat)]	Chromosome 12, NC_005111.4 (503660..544754)		
Brca2 ID: 37916	Breast cancer 2, early onset homolog [Drosophila melanogaster (fruit fly)]	Chromosome 2R, NT_033778.4 (24526172..24529581)	Dmel_CG30169, 30169, BRCA2, BcDNA:SD25109, CG13583, CG13584, CG30169, Dmbrca2, Dmel_CG30169, brca2, dmbrca2	
BRCA2 ID: 474180	BRCA2, DNA repair associated [Canis lupus familiaris (dog)]	Chromosome 25, NC_006607.3 (7734450..7797815, complement)		
BRCA2 ID: 374139	breast cancer 2 [Gallus gallus (chicken)]	Chromosome 1, NC_006088.4 (174560560..174597292, complement)		
brca2 ID: 566758	breast cancer 2, early onset [Danio rerio (zebrafish)]	Chromosome 15, NC_007126.6 (32053981..32070682)	fancd1	
Brca2 ID: 100689197	BRCA2, DNA repair associated [Cricetulus griseus (Chinese hamster)]		I79_011834	
BRCA2 ID: 721981	BRCA2, DNA repair associated [Macaca mulatta (Rhesus monkey)]	Chromosome 17, NC_027909.1 (11850054..11915799)		
BRCA2 ID: 108785046	BRCA2, DNA repair associated [Nanorana parkeri]			
BRCA2 ID: 108785046	BRCA2, DNA repair associated [Bos taurus (cow)]	Chromosome 12, NC_000012.11 (150522021..150570147)		

Bases de datos interconectadas

NCBI Resources How To Sign in to NCBI

Gene Gene BRCA2 Advanced Search Help

Full Report Send to: Hide sidebar >>

BRCA2 BRCA2, DNA repair associated [*Homo sapiens* (human)]

Gene ID: 675, updated on 9-Oct-2016

Summary

Official Symbol: BRCA2 provided by HGNC
Official Full Name: BRCA2, DNA repair associated provided by HGNC
Primary source: HGNC:HGNC:1101
See related: Ensembl:ENSG0000013961; HPRD:02554; MIM:600185; Vega:OTTHUMG00000017411
Gene type: protein coding
RefSeq status: REVIEWED
Organism: Homo sapiens
Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as: FAD; FAD1; GLM3; BRCC2; FANCD; RNCA2; FANCD1; XRCC11; BROVCA2
Summary: Inherited mutations in BRCA1 and this gene, BRCA2, confer increased lifetime risk of developing breast or ovarian cancer. Both BRCA1 and BRCA2 are involved in maintenance of genome stability, specifically the homologous recombination pathway for double-strand DNA repair. The BRCA2 protein contains several copies of a 70 aa motif called the BRC motif, and these motifs mediate binding to the RAD51 recombinase which functions in DNA repair. BRCA2 is considered a tumor suppressor gene, as tumors with BRCA2 mutations generally exhibit loss of heterozygosity (LOH) of the wild-type allele. [provided by RefSeq, Dec 2008]
Orthologs: mouse all

Table of contents

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- Phenotypes
- Variation
- Pathways from BioSystems
- Interactions
- General gene information
- Markers, Homology, Gene Ontology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences
- Additional links
- Locus-specific Databases



Interconexión con Ensembl

Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (GRCh38.p7) ▾ Location: 13:32,315,474-32,400,266 Gene: BRCA2

Login/Register

Gene-based displays

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

Gene: BRCA2 ENSG00000139618

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

Synonyms
XRCC11, FACD, FAD1, PNCA2, FAD, BRCC2, FANCD1, GLM3, BROVCA2, FANCD

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

About this gene
This gene has 7 transcripts ([splice variants](#)), [64 orthologues](#), is a member of [1 Ensembl protein family](#) and is associated with [63 phenotypes](#).

[Show transcript table](#)

Summary

Name
[BRCA2](#) (HGNC Symbol)

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

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

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

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

ENSG00000139618.14

This gene maps to [32,889,611-32,974,403](#) in GRCh37 coordinates.

View this locus in the GRCh37 archive: [ENSG00000139618](#)

Known protein coding

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

Gene type
This gene corresponds to the following database identifiers:

Annotation method
Alternative genes
Annotation Attributes

[Configure this page](#)

[Custom tracks](#)

[Export data](#)

Interconexión con HPRD DB

You are at: HPRD >> Query >> BRCA2

BRCA2

Molecular Class: Transcription regulatory protein
Molecular Function: Transcription regulator activity
Biological Process: Regulation of nucleobase, nucleoside, nucleotide and r

Diagram showing five red dots connected by vertical lines to a horizontal line, representing phosphorylation sites or PTMs.

ALTERNATE NAMES **DISEASES** **PTMs & SUBSTRATES**
SUMMARY SEQUENCE INTERACTIONS EXTERNAL LINKS

General

HPRD ID:	02554	Molecular Weight (Da):	384201
Gene Symbol:	BRCA2	Gene Map Locus:	13q12.3

Localization

Primary	Nucleus GO	Alternate	Cytoplasm GO Extracellular GO
---------	----------------------------	-----------	---

Domains and Motifs

Domains	Motifs
---------	--------

Expression

Site of Expression
Mammary gland
Serum

CONTENIDO

- Genómica, el “Big Data” y la Bioinformática
- Introducción General sobre Bases de Datos
- Bases de datos en genómica
- **Uso de ENSEMBL**

Genome Browser Ensembl

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Login/Register

Search all species...

Search: All species for

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

Browse a Genome

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

Popular genomes

Human GRCh38.p7	Human GRCh37
Mouse GRCm38.p4	Zebrafish GRCz10

[★ Log in to customize this list](#)

All genomes

-- Select a species --

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

Other species are available in [Ensembl Pre!](#) and [Ensembl Genomes](#).

Still using Human GRCh37?

Go to [e!GRCh37](#)

Variant Effect Predictor

Gene expression in different tissues

Find SNPs and other variants for my gene

G_TA_TA_CA_TT_C
C_RT_BA_AG_TT_T
C_TT_CT_AT_TT_T
G_RA_AC_TT_TC_C

Retrieve gene sequence

GGCTGACTTCCGGGTTGG
GGCGCTTGTGGGCCGAACG
GGGCCTCTGCTGCGCCCT
AAGGAGACAGATTGTGAA
CACCTCTGAGACGCGTTG
CCCAAGTCACAGCGTGCGC

Compare genes across species

Use my own data in Ensembl

ENCODE data in Ensembl

ENCODE

What's New in Ensembl Release 85 (July 2016)

- Update to Ensembl-Havana human GENCODE gene set (release 25)
- 30 new epigenomes from the Roadmap Epigenomics Project
- New zebrafish rnaseq
- Update to Rat Ensembl-Havana gene set
- Mouse: update to Ensembl-Havana GENCODE gene set

[Full details](#) | [All web updates](#), by release | [More news on our blog](#)

- 10 Mar 2016: [Ensembl 84 has been released!](#)
- 16 Feb 2016: [Learn about Ensembl – online, live and free!](#)
- 25 Jan 2016: [Sharing feature on the new mobile site \(\[m.ensembl.org\]\(#\)\)](#)

[Go to Ensembl blog](#)

Tweets by @ensembl

e! Ensembl @ensembl
Novel O-GlcNAcylation of H2ASer40 is linked to the evolution of placental animals #UsingEnsembl @SciReports [buff.ly/2cRgQ4R](#)

Novel O-GlcNAcylation on Ser40 of...

We report here newly discovered O-linked-N-acetylglucosamine (O-GlcNAc) nature.com

Utilidades en Ensembl:



Herramientas de
análisis/consultas

Acceso a
información plana

Genome Browser

Herramientas de análisis:

Name	Description
Variant Effect Predictor 	Analyse your own variants and predict the functional consequences of known and unknown variants via our Variant Effect Predictor (VEP) tool.
BLAST/BLAT	Search our genomes for your DNA or protein sequence.
File Chameleon	Convert Ensembl files for use with other analysis tools
Assembly Converter	Map (liftover) your data's coordinates to the current assembly.
ID History Converter	Convert a set of Ensembl IDs from a previous release into their current equivalents.

<http://www.ensembl.org/info/docs/tools/index.html>

Herramientas de consultas:

Name	Description
<u>BioMart</u>	Use this data-mining tool to export custom datasets from Ensembl.
<u>Ensembl Perl API</u>	Programmatic access to all Ensembl data using simple Perl scripts
<u>Ensembl Virtual Machine</u>	VirtualBox virtual Machine with Ubuntu desktop and pre-configured with the latest Ensembl API plus Variant Effect Predictor (VEP). NB: download is >1 GB
<u>Ensembl REST server</u>	Access Ensembl data using your favourite programming language

<http://www.ensembl.org/info/docs/tools/index.html>

La base de datos de ENSEMBL se localiza en: <http://www.ensembl.org/>



Pantalla principal

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Login/Register

Search all species... 

Search: All species for

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

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.p7	 Human GRCh37
 Mouse GRCm38.p4	 Zebrafish GRCz10

[Edit favourites](#)

All genomes

-- Select a species -- 

Still using Human GRCh37? 

Variant Effect Predictor 

Gene expression in different tissues 

Find SNPs and other variants for my gene 

Retrieve gene sequence 

Compare genes across species 

Use my own data in Ensembl

ENCODE data in Ensembl

What's New in Ensembl Release 86 (October 2016)

- Mouse Strains
- Chicken new assembly and gene set
- Macaque new assembly and genebuild
- Mouse lemur new assembly and genebuild
- Zebrafish: update to Ensembl-Havana merged gene set

[Full details](#) | [All web updates](#) | [by release](#) | [More news on our blog](#)

- 17 Oct 2016: [Ensembl helpdesk maintenance](#)
- 05 Oct 2016: [Ensembl 86 has been released!](#)
- 24 Aug 2016: [Haplotype data from the 1000 Genomes Project available in Ensembl](#)

[Go to Ensembl blog](#)

Tweets by @ensembl

e! Ensembl @ensembl
Vitamin K epoxide reductase complex subunit 1 (VKORC1) is involved in blood coagulation
[@news4GO](#) #Geneoftheweek [buff.ly/2dPLkT](#)

Gene: VKORC1 [ensembl.org](#)
Interaction: [InterPro](#), [MGI](#), [MINT](#), [Reactome](#), [SIFMA](#)
Location: [GenBank](#), [NCBI](#), [Ensembl](#), [UCSC](#), [ENSEMBL](#), [OMIM](#), [Phenotype](#)
About this gene

Pantalla principal

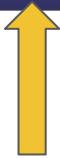
The screenshot shows the top navigation bar of the Ensembl website. On the left is the Ensembl logo. To its right are several menu items: BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. On the far right are links for Login/Register and a search bar with the placeholder "Search all species...". A yellow arrow points upwards from the text "Herramientas de análisis/consultas" towards the top of the image.

!Ensembl

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Login/Register

Search all species...



Herramientas de análisis/consultas

Pantalla principal



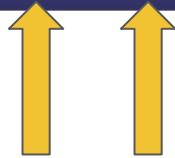
The screenshot shows the top navigation bar of the Ensembl website. On the left is the Ensembl logo. To its right are links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. On the far right are links for Login/Register and a search bar with the placeholder "Search all species...". Two yellow arrows point upwards from the text below to the "Tools" and "Downloads" links in the navigation bar.

e!Ensembl

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Login/Register

Search all species...



Atajos a algunas herramientas populares

Existen muchos atajos hacia herramientas:

The screenshot shows the Ensembl homepage. At the top, there is a navigation bar with links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. On the right side of the top bar is a search bar labeled "Search all species..." with a magnifying glass icon. Below the top bar, there is a search form with a dropdown menu set to "All species" and a text input field. A yellow arrow points from the "e!Ensembl" logo on the left towards this search form. To the right of the search form is a section titled "What's New in Ensembl Release 86 (October 2016)" which lists several new features. Below this section are three columns of links: "Still using Human GRCh37?", "Variant Effect Predictor (Ve!P)", "Gene expression in different tissues", "Find SNPs and other variants for my gene", "Retrieve gene sequence", and "Compare genes across species".

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

Search: All species for [Go](#)

e.g. BRCA2 or rs62797383-63627669 or rs699 or coronary heart disease

Still using Human GRCh37? [Go to e!GRCh37](#)

Variant Effect Predictor [Ve!P](#)

Gene expression in different tissues

Find SNPs and other variants for my gene

GT_TTACATT_C C_RTAAGTCT_T CTTCT_AATT_G GRAACATTTC_C

Retrive gene sequence

Compare genes across species

What's New in Ensembl Release 86 (October 2016)

- Mouse Strains
- Chicken new assembly and gene set
- Macaque new assembly and genebuild
- Mouse lemur new assembly and genebuild
- Zebrafish: update to Ensembl-Havana merged gene set

[Full details](#) | [All web updates, by release](#) | [More news on our blog](#)

- 17 Oct 2016: [Ensembl helpdesk maintenance](#)
- 05 Oct 2016: [Ensembl 86 has been released!](#)
- 24 Aug 2016: [Haplotype data from the 1000 Genomes Project available in Ensembl](#)

[Go to Ensembl blog](#)

Existen muchos atajos hacia herramientas:



VEP

EBI and the Wellcome Trust Sanger Institute to develop a software system which produces and analyses eukaryotic genomes.

Wellcome Trust. Our [acknowledgements page](#) includes a list of additional current and previous funding



Sección inferior de la página

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

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](#)

Práctica con BIOMART

Ejercicio

- 1.- ¿Cuántos y cuáles genes hay en el cromosoma Y?
- 2.- ¿Cuántos y cuáles codifican proteínas?



<http://www.biomart.org>

Biomart es un proyecto que tiene dentro de sus objetivos permitir el acceso de información a la comunidad científica.

El software de Biomart permite el acceso a diferentes bases de datos:

CENTER FOR MATHEMATICAL MODELING AND CENTER FOR GENOME REGULATION (CMM), CHILE	SalmonDB	Genomic information for Atlantic salmon, rainbow trout, and related species
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS), FRANCE	Cildb	Database for eukaryotic cilia and centriolar structures, integrating orthology relationships for 33 species with high throughput studies and OMIM
	Paramecium DB	Paramecium genome database
COLD SPRING HARBOR LABORATORY (CSHL), US	Gramene	Agriculturally important grass genomes
	WormBase	<i>C. elegans</i> and related nematode genomic information
DEPARTMENT OF GENETICS, UNIVERSITY OF CAMBRIDGE, CAMBRIDGE, UK	DAPPER	Mass spec identified protein interaction networks in Drosophilacell cycle regulation
	Ensembl Genomes	Genome databases for fungi, metazoa, plants and protists
	HGNC	Repository of human gene nomenclature and associated resources
EUROPEAN BIOINFORMATICS INSTITUTE (EBI), UK	InterPro	Integrated database of predictive protein "signatures" used for the classification and automatic annotation of proteins and genomes
	PRIDE	Repository for protein and peptide identifications
	UniProt	Protein sequence and functional information
	WormBase ParaSite	Parasitic worms (helminths) are responsible for more than a billion human infections globally and have a devastating impact on livestock and agriculture.

Ejercicio

- 1.- ¿Cuántos y cuáles genes hay en el cromosoma Y?
- 2.- ¿Cuántos y cuáles codifican proteínas?

e!



Interfaz general de Biomart:

Pantalla de inicio de Biomart

The screenshot shows the Biomart interface. At the top, there is a navigation bar with several buttons: 'New' (highlighted in blue), 'Count', 'Results', 'URL', 'XML', 'Perl', and 'Help'. Below the navigation bar, there is a large search input field labeled '- CHOOSE DATABASE - ▾'. To the left of this input field, there is a sidebar with a blue header labeled 'Dataset' and the text '[None selected]'. The main body of the interface is currently empty.

1.- Elegir base de datos:

Screenshot of a web-based dataset selection interface.

The interface has a header with several buttons: New, Count, Results, URL, XML, Perl, and Help.

The main area is divided into two sections:

- Dataset**: A light blue sidebar containing the text "[None selected]".
- Main Area**: A white space containing a dropdown menu labeled "- CHOOSE DATABASE - ▾".

A large yellow arrow points from the text "1.- Elegir base de datos:" in the previous slide towards the "- CHOOSE DATABASE -" dropdown menu in this screenshot.

1.- Elegir base de datos:

Screenshot of a bioinformatics tool interface showing the selection of a dataset.

The top navigation bar includes buttons for New, Count, Results, URL, XML, Perl, and Help.

The left sidebar shows the selected dataset: Ensembl Genes 86, with [None selected] below it.

The main panel displays a dropdown menu titled "Dataset" containing a list of available datasets. A yellow arrow points to the option "Homo sapiens genes (GRCh38.p7)", which is highlighted in blue.

The list of datasets includes:

- CHOOSE DATASET -
- CHOOSE DATASET -
- Danio rerio genes (GRCz10)
- Gallus gallus genes (Gallus_gallus-5.0)
- Homo sapiens genes (GRCh38.p7)** (highlighted)
- Mus musculus genes (GRCM38.p4)
- Rattus norvegicus genes (Rnor_6.0)
- Ailuropoda melanoleuca genes (ailMe1)
- Anas platyrhynchos genes (BGI_duck_1.0)
- Anolis carolinensis genes (AnoCar2.0)
- Astyanax mexicanus genes (AstMex102)
- Bos taurus genes (UMD3.1)
- Caenorhabditis elegans genes (WBcel235)
- Callithrix jacchus genes (C_jacchus3.2.1)
- Canis familiaris genes (CanFam3.1)
- Cavia porcellus genes (cavPor3)
- Chlorocebus sabaeus genes (ChlSab1.1)

Interfaz general de Biomart:

Pantalla de ejecución

The screenshot shows the Biomart execution interface. On the left, there are three sections: 'Dataset' (selected), 'Filters' (None selected), and 'Attributes' (Ensembl Gene ID, Ensembl Transcript ID). To the right, orange curly braces group these items under their respective labels: 'Base de datos seleccionada', 'Filtro para acotar las búsquedas en la base de datos seleccionada', and 'Atributos a desplegar en los resultados'. The top navigation bar includes 'New', 'Count', 'Results', 'URL', 'XML', 'Perl', and 'Help'.

Dataset

Homo sapiens genes
(GRCh38.p7)

Filters

[None selected]

Attributes

Ensembl Gene ID
Ensembl Transcript ID

Base de datos seleccionada

Filtro para acotar las búsquedas en la base de datos seleccionada

Atributos a desplegar en los resultados

2.- Elegir el filtro adecuado: Cromosoma Y

New Count Results URL XML Perl Help

Dataset
Homo sapiens genes
(GRCh38.p7)

Filters
[None selected]

Attributes
Ensembl Gene ID
Ensembl Transcript ID

Dataset
[None Selected]

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

REGION:
 GENE:
 PHENOTYPE:
 GENE ONTOLOGY:
 MULTI SPECIES COMPARISONS:
 PROTEIN DOMAINS AND FAMILIES:
 VARIANT:



Atributos en la base de datos para filtrar

2.- Elegir el filtro adecuado: Cromosoma Y

New Count Results URL XML Perl Help

Dataset
Homo sapiens genes
(GRCh38.p7)

Filters
[None selected]

Attributes
Ensembl Gene ID
Ensembl Transcript ID

Dataset
[None Selected]

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

REGION: 

GENE:

PHENOTYPE:

GENE ONTOLOGY:

MULTI SPECIES COMPARISONS:

PROTEIN DOMAINS AND FAMILIES:

VARIANT:

El filtro único a seleccionar es cromosoma, pertenece a región.

2.- Elegir el filtro adecuado: Cromosoma Y

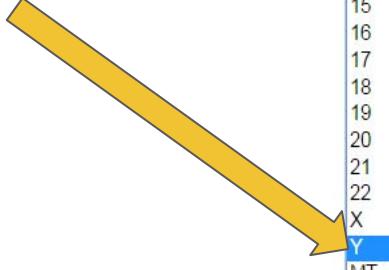
New Count Results URL XML Perl Help

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

REGION:

Chromosome

El atributo único a seleccionar es cromosoma Y



0
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
MT

New Count Results URL XML Perl Help

3.- Elegir el atributo adecuado: Genes

New Count Results URL XML Perl Help

Dataset
Homo sapiens genes (GRCh38.p7)

Filters
Chromosome: Y

Attributes
Ensembl Gene ID
Ensembl Transcript ID

Dataset
[None Selected]

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

Missing non coding genes in your mart query output, please check the following [FAQ](#)

Features Variant (Germline)
 Structures Variant (Somatic)
 Homologues Sequences

GENE: 
 EXTERNAL:
 PROTEIN DOMAINS AND FAMILIES:

El atributo único para desplegar y requerido es gen. Pertenece a GENE.

New Count Results

URL XML Perl Help

Dataset
Homo sapiens genes
(GRCh38.p7)

Filters

Chromosome: Y

Attributes

Ensembl Gene ID

Associated Gene Name

Dataset

[None Selected]

Ensembl

- Ensembl Gene ID
- Ensembl Transcript ID
- Ensembl Protein ID
- Ensembl Exon ID
- Description
- Chromosome Name
- Gene Start (bp)
- Gene End (bp)
- Strand
- Band
- Transcript Start (bp)
- Transcript End (bp)
- Transcription Start Site (TSS)
- Transcript length (including UTRs and CDS)
- Transcript Support Level (TSL)
- GENCODE basic annotation

- APPRIS annotation
- Associated Gene Name
- Associated Gene Source
- Associated Transcript Name
- Associated Transcript Source
- Transcript count
- % GC content
- Gene type
- Transcript type
- Source (gene)
- Source (transcript)
- Status (gene)
- Status (transcript)
- Version (gene)
- Version (transcript)

Se selecciona ID de gen y el nombre
de gen asociado.

4.- Cuantificar los genes (registros) que cumplen los filtros

Dataset 523 / 63305 genes

Homo sapiens genes
(GRCh38.p7)

Filters

Chromosome: Y

Attributes

Ensembl Gene ID

Associated Gene Name

Dataset [None Selected]

Ensembl

Ensembl Gene ID

Ensembl Transcript ID

Ensembl Protein ID

Ensembl Exon ID

Description

Chromosome Name

Gene Start (bp)

Gene End (bp)

Strand

Band

Transcript Start (bp)

Transcript End (bp)

Transcription Start Site (TSS)

Transcript length (including UTRs and CDS)

Transcript Support Level (TSL)

GENCODE basic annotation

APPRIIS annotation

Associated Gene Name

Associated Gene Source

Associated Transcript Name

Associated Transcript Source

Transcript count

% GC content

Gene type

Transcript type

Source (gene)

Source (transcript)

Status (gene)

Status (transcript)

Version (gene)

Version (transcript)

Cuantificación de acuerdo a los filtros seleccionados

5.- Identificar los genes (registros) que cumplen los filtros

The screenshot shows a web-based application for gene filtering. At the top, there are tabs for 'New', 'Count', and 'Results' (which is selected). To the right are links for 'URL', 'XML', 'Perl', and 'Help'. On the left, a sidebar displays the dataset information: 'Dataset 523 / 63305 Genes' (Homo sapiens genes, GRCh38.p7), 'Filters' (set to Chromosome: Y), and 'Attributes' (Ensembl Gene ID, Associated Gene Name). Below this is a 'Dataset' section with '[None Selected]'. The main content area has two columns. The left column lists various Ensembl attributes as checkboxes, with 'Associated Gene Name' checked. A yellow arrow points from the 'Associated Gene Name' checkbox in this list towards the central text. The right column lists other attributes as checkboxes, none of which are checked.

Ensembl	
<input type="checkbox"/> Ensembl Gene ID	<input type="checkbox"/> APPRIS annotation
<input type="checkbox"/> Ensembl Transcript ID	<input checked="" type="checkbox"/> Associated Gene Name
<input type="checkbox"/> Ensembl Protein ID	<input type="checkbox"/> Associated Gene Source
<input type="checkbox"/> Ensembl Variant ID	<input type="checkbox"/> Associated Transcript Name
<input type="checkbox"/> Description	<input type="checkbox"/> Associated Transcript Source
<input type="checkbox"/> Chromosome Name	<input type="checkbox"/> Transcript count
<input type="checkbox"/> Gene Start (bp)	<input type="checkbox"/> % GC content
<input type="checkbox"/> Gene End (bp)	<input type="checkbox"/> Gene type
<input type="checkbox"/> Strand	<input type="checkbox"/> Transcript type
<input type="checkbox"/> Band	<input type="checkbox"/> Source (gene)
<input type="checkbox"/> Transcript Start (bp)	<input type="checkbox"/> Source (transcript)
<input type="checkbox"/> Transcript End (bp)	<input type="checkbox"/> Status (gene)
<input type="checkbox"/> Transcription Start Site (TSS)	<input type="checkbox"/> Status (transcript)
<input type="checkbox"/> Transcript length (including UTRs and CDS)	<input type="checkbox"/> Version (gene)
<input type="checkbox"/> Transcript Support Level (TSL)	<input type="checkbox"/> Version (transcript)
<input type="checkbox"/> GENCODE basic annotation	

Revisar el resultado de
los filtros y atributos
seleccionados

5.- Identificar los genes (registros) que cumplen los filtros

New Count Results URL XML Perl Help

Dataset 523 / 63305 Genes
Homo sapiens genes (GRCh38.p7)

Filters Chromosome: Y

Attributes Ensembl Gene ID
Associated Gene Name

Dataset [None Selected]

Export all results to File TSV Unique results only Go

Email notification to

View All rows as HTML Unique results only

Ensembl Gene ID	Associated Gene Name
ENSG00000252667	RNA5P523
ENSG00000279287	AC012005.2
ENSG00000251953	RNA5P522
ENSG00000225609	CDY20P
ENSG00000235004	USP9YP30
ENSG00000227633	RBMY2YP
ENSG00000252766	RNU6-255P
ENSG00000187191	DAZ3
ENSG00000280301	AC006328.1
ENSG00000235981	AC023274.4

Ejercicio

1.- ¿Cuántos y cuáles genes hay en el cromosoma Y?

R: 523, revisar: <http://www.ensembl.org/biomart/martview/96b52c161ec9c169a77d0a3639b40f7b>

2.- ¿Cuántos y cuáles codifican proteínas?

1.- Aplicar el nuevo filtro

New Count Results URL XML Perl Help

Dataset 523 / 63305 Genes
Homo sapiens genes
(GRCh38.p7)

Filters

Chromosome: Y

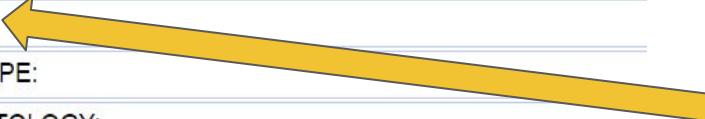
Attributes

Ensembl Gene ID
Associated Gene Name

Dataset

[None Selected]

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

REGION:
 GENE: 
 PHENOTYPE:
 GENE ONTOLOGY:
 MULTI SPECIES COMPARISONS:
 PROTEIN DOMAINS AND FAMILIES:
 VARIANT:

**El filtro de “codificador”
aplica a GENE**

1.- Aplicar el nuevo filtro

New Count Results

URL XML Perl Help

Input microarray probes/probesets ID list [max 500 advised] All files no probeset ID(s) [e.g. GSE_1_an]

Dataset
Homo sapiens genes (GRCh38.p7)

Filters
Chromosome: Y
Gene type: protein_coding

Attributes
Ensembl Gene ID
Associated Gene Name

Dataset
[None Selected]

Transcript count >=

Gene type Ningún archivo seleccionado

Mt_rRNA
Mt_tRNA
non_coding
polymorphic_pseudogene
processed_pseudogene
processed_transcript
protein_coding
pseudogene
ribozyme

Transcript type

3prime_overlapping_ncRNA
antisense
bidirectional_promoter_lncRNA
IG_C_gene
IG_C_pseudogene
IG_D_gene
IG_J_gene

1.- Aplicar el nuevo filtro

Dataset 53 / 63305 Genes
Homo sapiens genes
(GRCh38.p7)

Filters

Chromosome: Y
Gene type: protein_coding

Attributes

Ensembl Gene ID
Associated Gene Name

Dataset
[None Selected]

Input microarray probes/probesets ID list [max 500 advised]
All microarray probeset ID(s) (e.g. 115_1_at)

Transcript count >=

Gene type Ningún archivo seleccionado

Mt_rRNA
Mt_tRNA
non_coding
polymorphic_pseudogene
processed_pseudogene
processed_transcript
protein_coding
pseudogene
ribozyme



2.- Cuantificar los genes “codificantes”

The screenshot shows a web-based bioinformatics application for gene analysis. The top navigation bar includes buttons for 'New', 'Count' (which has a yellow arrow pointing to it), and 'Results'. Other buttons include 'URL', 'XML', 'Perl', and 'Help'. The main interface displays a dataset of 53 genes out of 63,305, specifically Homo sapiens genes (GRCh38.p7). On the left, there are sections for 'Filters' (set to Chromosome: Y and Gene type: protein_coding) and 'Attributes' (listing Ensembl Gene ID and Associated Gene Name). The 'Dataset' section indicates [None Selected]. The right side of the interface features a search bar for 'Input microarray probes/probesets ID list (max 500 advised)' and a file selection area for 'Ally NCBI probeset ID(s) (e.g. 115 2 49)'. Below these are filter options for 'Transcript count >=' and 'Gene type', with 'Gene type' being checked. A large dropdown menu lists various gene types, with 'protein_coding' highlighted in grey. A red box highlights the 'protein_coding' option in the dropdown. The text 'Cuantificamos los genes' is overlaid in red at the bottom left of the dropdown area.

Cuantificamos los genes

- Transcript count \geq
- Gene type

- Mt_rRNA
- Mt_tRNA
- non_coding
- polymorphic_pseudogene
- processed_pseudogene
- processed_transcript
- protein_coding**
- pseudogene
- ribozyme

3.- Identificar los genes “codificantes”

New Count Results Results

Star URL XML Perl Help

Dataset 53 / 63305 Genes
Homo sapiens genes (GRCh38.p7)

Filters

Chromosome: Y
Gene type: protein_coding

Attributes

Ensembl Gene ID
Associated Gene Name

Dataset [None Selected]

Export all results to File HTML ▾ Unique results only Go

Email notification to

View All rows as HTML Unique results only

Ensembl Gene ID	Associated Gene Name
ENSG00000279287	AC012005.2
ENSG00000187191	DAZ3
ENSG00000280301	AC006328.1
ENSG00000279115	AC006386.1
ENSG00000099721	AMELY
ENSG00000188120	DAZ1
ENSG00000183753	BPY2
ENSG00000129824	RPS4Y1
ENSG00000279274	AC012005.1
ENSG00000205916	DAZ4
ENSG00000279664	AC009491.2
ENSG00000092377	TBL1Y
ENSG00000114374	USP9Y
ENSG00000226941	RBMY1J

Ejercicio

1.- ¿Cuántos y cuáles genes hay en el cromosoma Y?

R: 523, revisar: <http://www.ensembl.org/biomart/martview/96b52c161ec9c169a77d0a3639b40f7b>

2.- ¿Cuántos y cuáles codifican proteínas?

R: 53, revisar: <http://www.ensembl.org/biomart/martview/96b52c161ec9c169a77d0a3639b40f7b>



Acceso a información plana

Existen cuatro métodos de acceso a la información

[Login/Register](#)

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

[Download a sequence or region](#)



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

- FASTA sequence
- GTF or GFF features
- ...and more!

[Customise your download](#)



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

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

[Fetch data programmatically](#)



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

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

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

[Download databases & software](#)



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

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

Existen cuatro métodos de acceso a la información

The screenshot shows the Ensembl homepage with a dark blue header. The header includes the Ensembl logo, navigation links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors, and a search bar with a magnifying glass icon and placeholder text 'Search all species...'. On the right side of the header is a 'Login/Register' link. Below the header, there's a large callout box with a white background and a thin gray border. The title of the callout is 'Download a sequence or region'. Inside, there's an icon of a computer monitor displaying a genome browser interface. To the left of the monitor is a small box containing a 'Export data' button and a sequence of DNA bases: CAGAATGAT, AAATGTTCT, AAAAGAGCA, CTGTCATGC, ATAAAAAGAA, AGTGATAC. A red arrow points from this box towards the monitor icon. Below the monitor icon, text reads: 'Click on the 'Export data' button in the lefthand menu of most pages to export:' followed by a list: '• FASTA sequence', '• GTF or GFF features', and '...and more!'. At the bottom left of the callout box, there's a small note: 'Ensembl release 86 - Oct 2016 © WTSI / EMBL-EBI'.

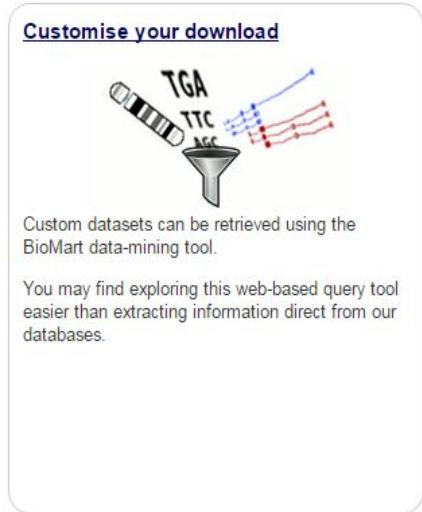
1. A través de las consultas realizadas en el “Genome Browser”

Existen cuatro métodos de acceso a la información

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Login/Register

 Search all species... 



2. Biomart

Existen cuatro métodos de acceso a la información

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Login/Register

 Search all species... 

3. Consultas con *scripts* desde línea de comandos.

[Fetch data programmatically](#)



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

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

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

Existen cuatro métodos de acceso a la información

e!Ensembl Login/Register

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

 Search all species... 

4. Sitio de descargas FTP

Download databases & software



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

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

[Permanent link](#)

Uso del sitio FTP desde el navegador:

The screenshot shows the Ensembl website's navigation bar with links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar at the top right includes a dropdown for species selection and a magnifying glass icon. Below the navigation bar, a breadcrumb trail shows the path: Home > Help & Documentation > Accessing Ensembl Data > FTP Download. The main content area is titled "FTP Download". It contains several sections: "Using this website" (with "FTP site" underlined), "Annotation and prediction", "Data access", "API & software", and "About us". A sidebar on the left lists "In this section" with "Downloading with rsync" selected. A search bar and a "Go" button are also present. The main content includes a "FTP Download" section with instructions for downloading via browser, script, or rsync, an "API Code" section with a link to download complete API code as a tarball, a note about API version compatibility, a "Database dumps" section with information about formats and MySQL dumps, and a "Custom data sets" box with a note about Biomart.

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Login/Register

Search all species...

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

In this section
└ Downloading with rsync

FTP Download

Search documentation... Go

FTP Download

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

API Code

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

[Download complete API for this release](#)

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

Database dumps

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

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

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

Custom data sets

If you want to filter or customise your download, please try [Biomart](#), a web-based querying tool.

Uso del sitio FTP desde el navegador:

Índice de /pub/

	Nombre	Tamaño	Fecha de modificación
📁	[directorio principal]		
📁	assembly/		24/5/07 0:00:00
📁	assembly_mapping/		18/10/16 10:22:00
📄	bamExample.bam	2.2 MB	7/2/11 0:00:00
📄	current_assembly_chain	0 B	4/10/16 15:02:00
📄	current_bamcov	0 B	4/10/16 15:02:00
📄	current_bed	0 B	4/10/16 15:02:00
📄	current_compara	0 B	4/10/16 15:02:00
📄	current_data_files	0 B	4/10/16 15:02:00
📄	current_embl	0 B	4/10/16 15:02:00
📄	current_emf	0 B	4/10/16 15:02:00
📄	current_fasta	0 B	4/10/16 15:02:00
📄	current_genbank	0 B	4/10/16 15:02:00
📄	current_gff3	0 B	4/10/16 15:02:00
📄	current_gtf	0 B	4/10/16 15:02:00
📄	current_maf	0 B	4/10/16 15:02:00
📄	current_mysql	0 B	4/10/16 15:02:00
📄	current_rdf	0 B	4/10/16 15:02:00
📄	current_README	0 B	4/10/16 15:02:00
📄	current_regulation	0 B	4/10/16 15:02:00
📄	current_solr_search	0 B	4/10/16 15:02:00
📄	current_tsv	0 B	4/10/16 15:02:00
📄	current_variation	0 B	4/10/16 15:02:00
📄	current_virtual_machine	0 B	4/10/16 15:02:00
📄	current_xml	0 B	4/10/16 15:02:00
📁	data_files/		28/9/16 9:21:00



PREGUNTAS

> Turning data into forefront knowledge



WINTER GENOMICS

www.wintergenomics.com

Contacto

Manizales No.906, Colonia Lindavista.
México, Distrito Federal.

Tel/fax: (52) (55) 5119-0240
(52) (55) 5119-5624

E-mail: contacto@wintergenomics.com