Curso Bioinformática para la medicina personalizada

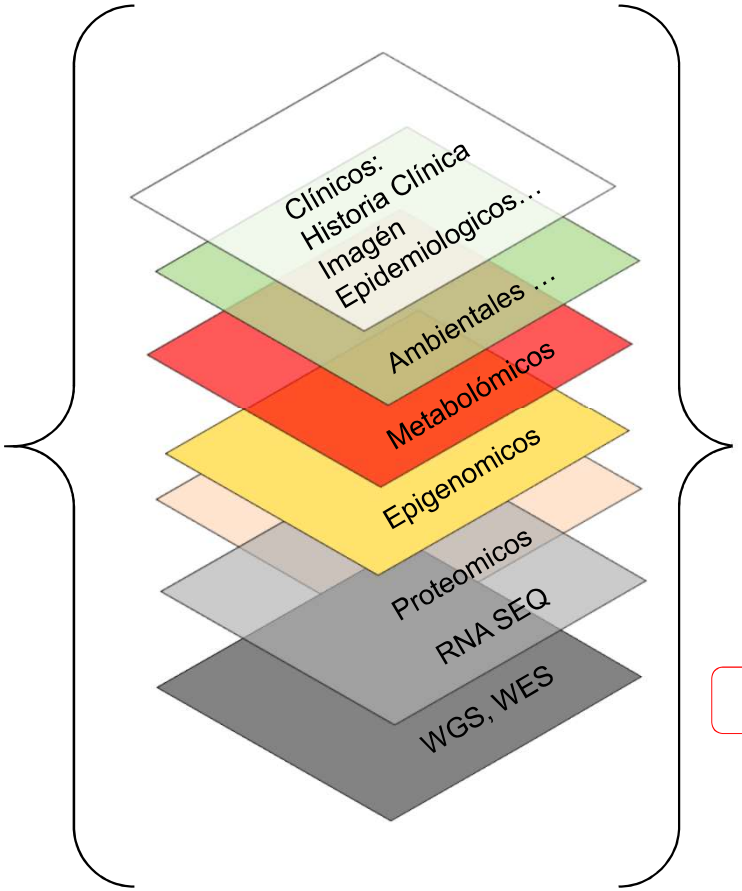
(Formación interna ISCIII)

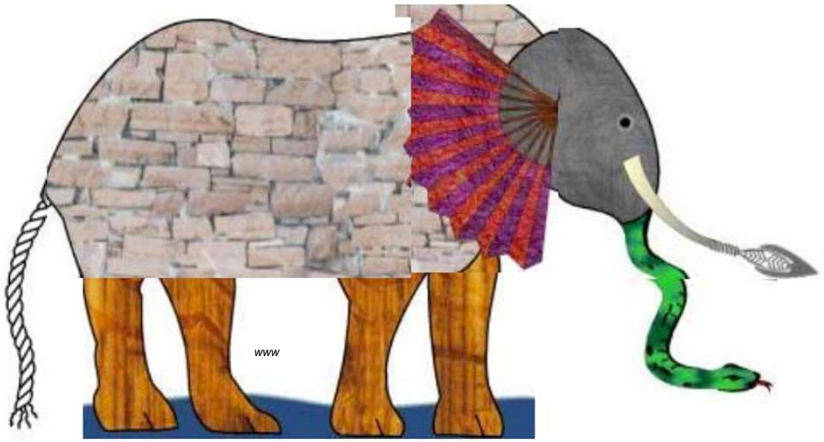
https://github.com/viclopezalonso

Victoria López Alonso
Unidad de Biología Computacional (UFIEC)
28 de octubre de 2019


Ensembl
Galaxy
igv
Integrative
Genomics
Viewer
Bases de datos biomédicas
Pipeline de Análisis de variantes
Priorización de Variantes
isc
Instituto de Salud Carlos III

Multidemensionalidad de los datos

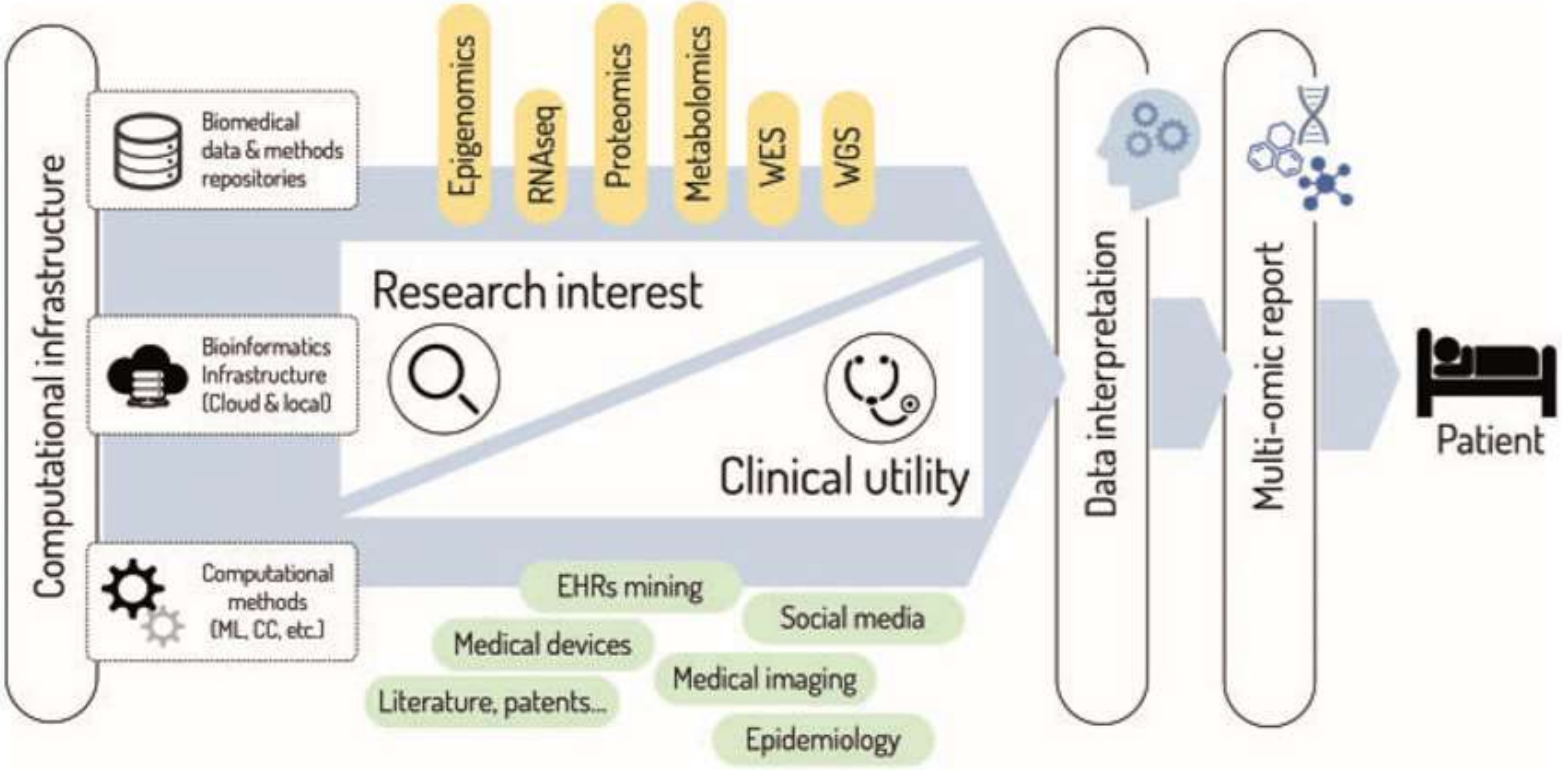




Al incrementar la dimensionalidad incrementa la complejidad

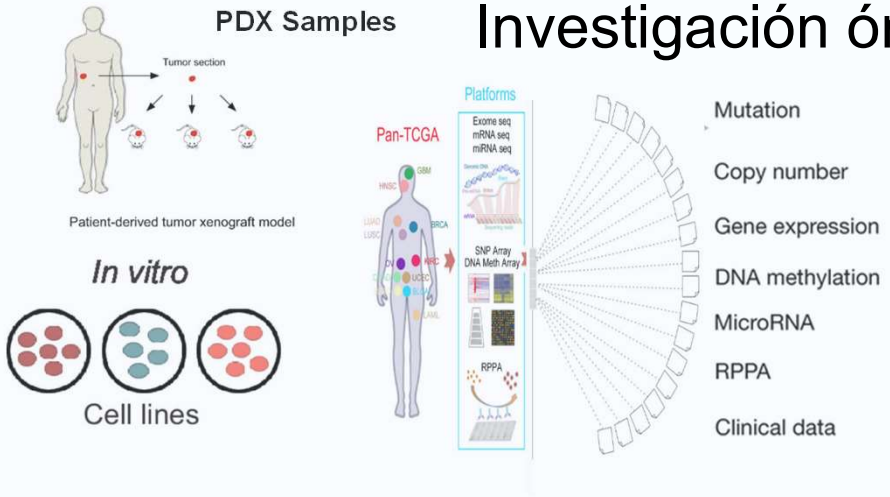
Instituto de Salud Carlos III

La Bioinformática propone herramientas para avanzar en Medicina Personalizada



Briefings in Bioinformatics, 20(3), 2019, 752–766

Instituto de Salud Carlos III



Validación
Regulación

Protocolos de Diagnóstico, Terapia, Prevención …

¿A qué datos/información se pueden acceder a través de repositorios públicos?
¿Cómo están asociados estos omicos con el fenotipo?
¿Los datos están disponibles en formatos que permitan su análisis integrado?
¿Qué plataformas bioinformáticas son accesibles online ?
….

Instituto

The Cancer Genome Atlas — Understanding genomics to improve cancer care

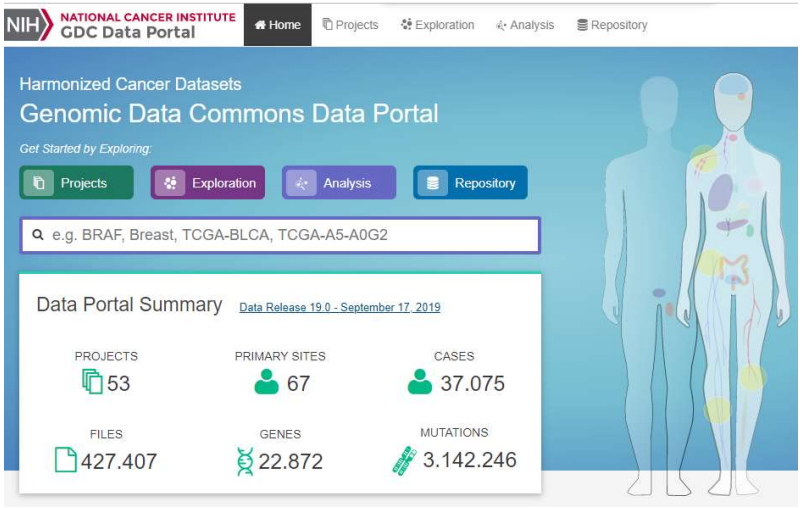
ICGC Data Portal

cBioPortal for Cancer Genomics

http://www.cbioportal.org/

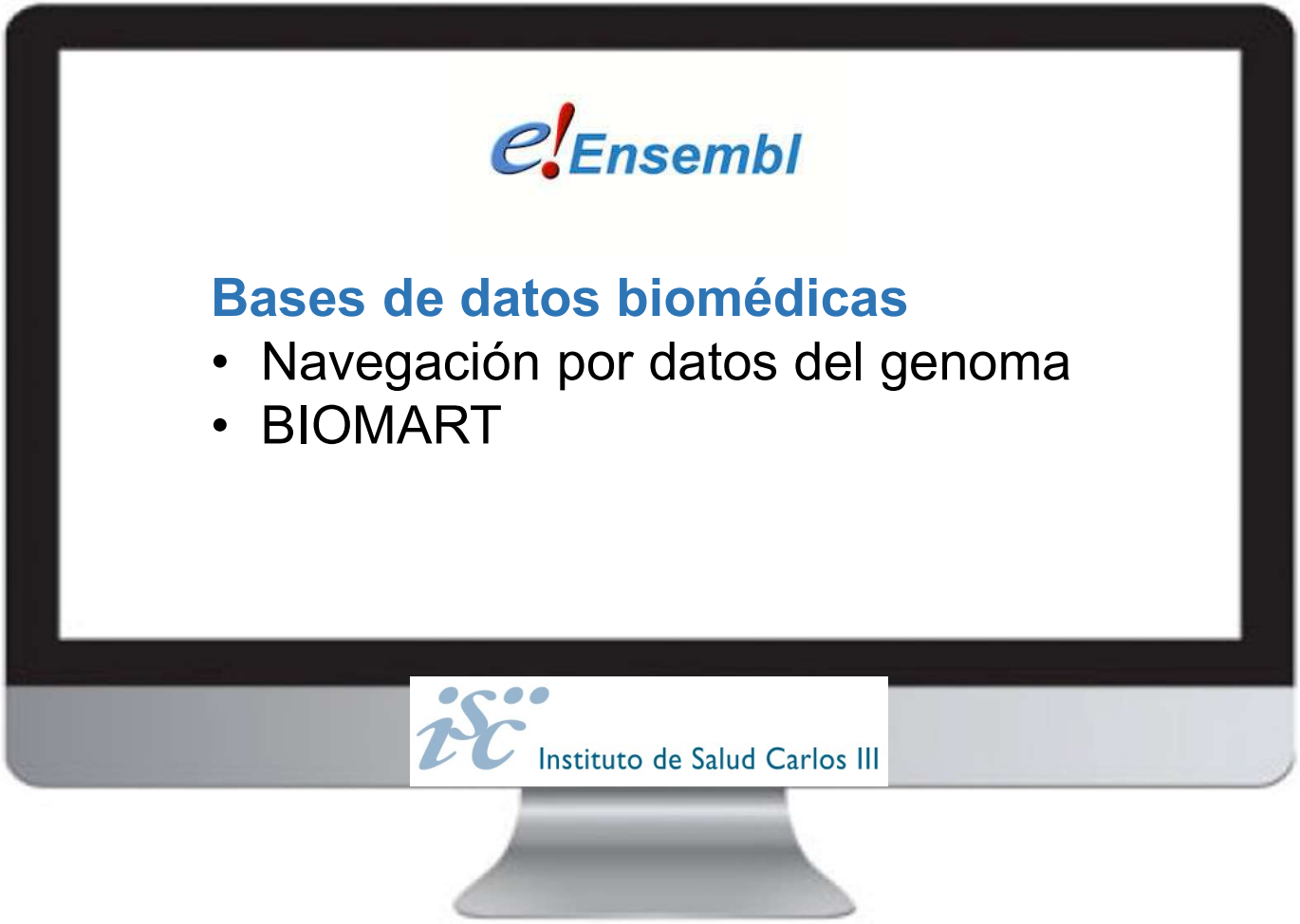
CANCER GENOMICS CLOUD
SEVEN BRIDGES

http://www.cancergenomicscloud.org/

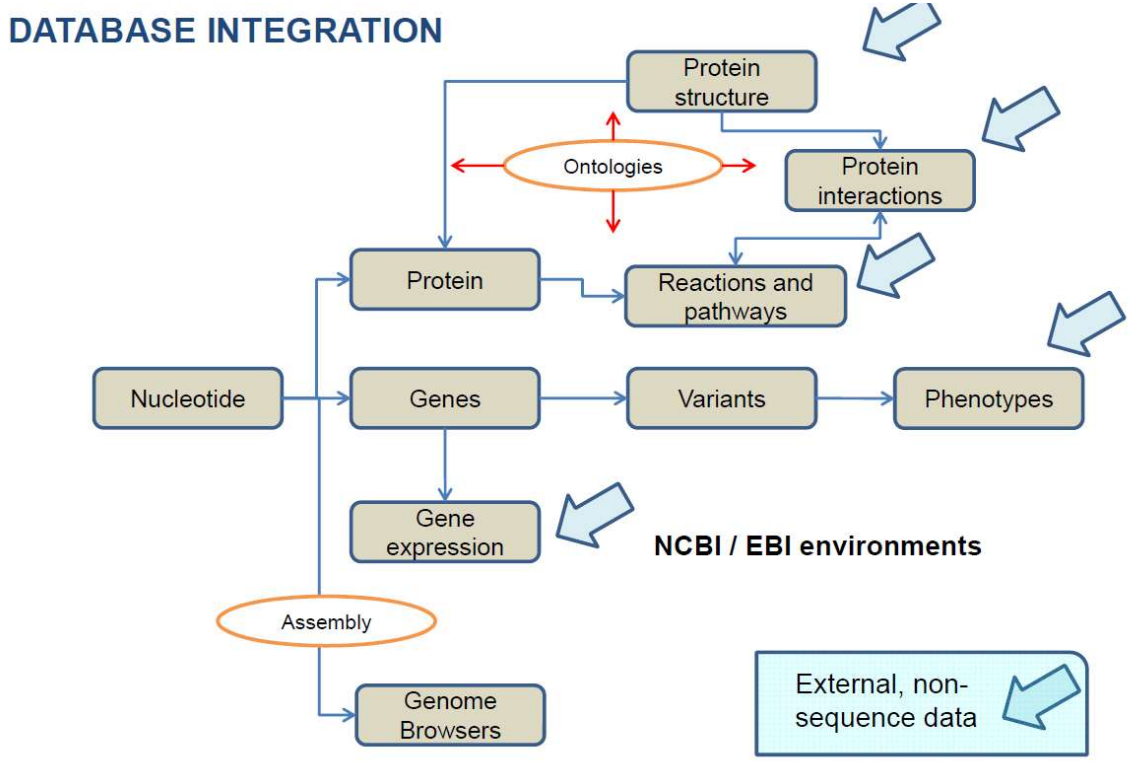


Repositorios para Medicina de Precision en Cancer

CIViC (WashU)
Cancer Genome Interpreter (Barcelona)
OncoKB (MSKCC)
PMKB (Cornell)
JAX-Clinical Knowledgebase (Jackson lab)
MolecularMatch
MyCancerGenome (Vanderbilt)
KnowledgeBase for Precision Oncology (MD Anderson)
CanDL (Ohio State)
COSMIC (Sanger)
…

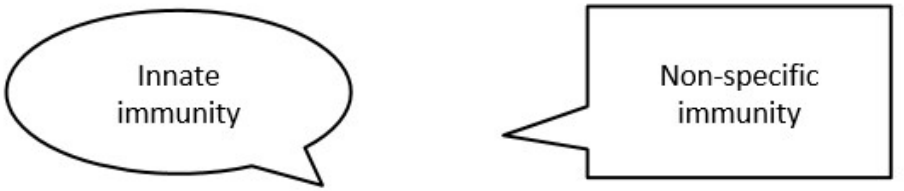

e!Ensembl
Bases de datos biomédicas
• Navegación por datos del genoma
• BIOMART
Instituto de Salud Carlos III

La mayoría de las bases de datos biomédicos están conectadas

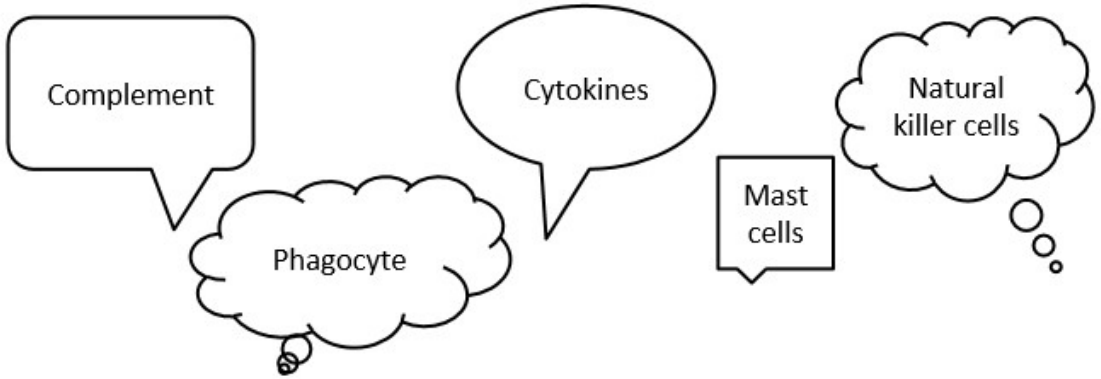


Gene Ontology

Multiple terms for the same thing



Gene descriptions too specific



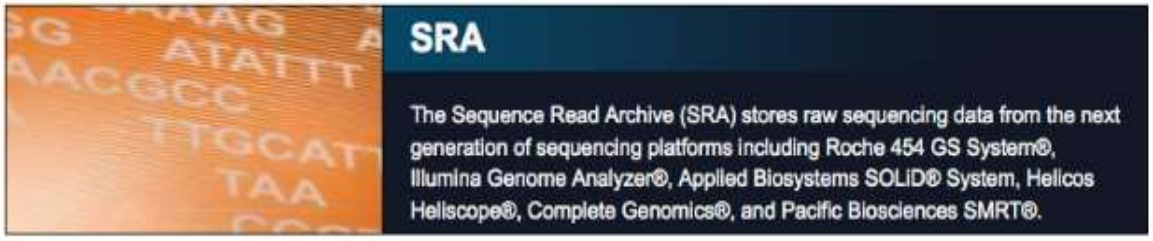
GO:0045087 - innate immune response: defense responses mediated by germline encoded components that directly recognise components of potential pathogens.

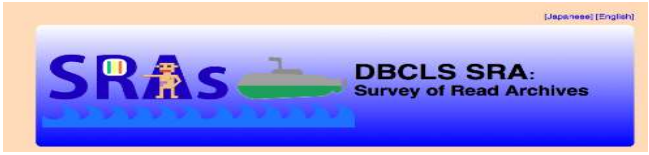
Molecular Function/ Biological Process/ Cellular Component

Bases de datos Biomédicas

Secuencias

Data type	DDBJ	EMBL-EBI	NCBI
Next generation reads	Sequence Read Archive	European Nucleotide Archive (ENA)	Sequence Read Archive
Capillary reads	Trace Archive		Trace Archive
Annotated sequences	DDBJ		GenBank
Samples	BioSample		BioSample
Studies	BioProject		BioProject





Instituto de Salud Carlos III

Bases de datos Biomédicas

Proteinas

Macromolecular Structure Database (MSD)

Instituto de Salud Carlos III

Bases de datos Biomédicas

Expresión

Unigene

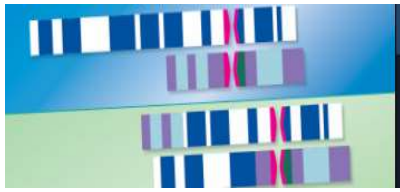
Epigenoma

Instituto de Salud Carlos III

Bases de datos Biomédicas

Variantes

dbSNP
Short Genetic Variations



dbVar

dbVar is NCBI's database of human genomic structural variation — insertions, deletions, duplications, inversions, mobile elements, translocations, and others

HGMD®
The Human Gene Mutation Database
Cardiff

DECIPHER
GRCh37

LOVD
Leiden Open Variation Database

ClinGen
Clinical Genome Resource

Bases de datos Biomédicas

ClinVar Variantes conocidas y su relevancia clínica

NM_007294.3(BRCA1):c.2677A>T (p.Lys893Ter)

FEEDBACK

Interpretation:	**Pathogenic**
Review status:	★★★☆ reviewed by expert panel
Submissions:	3 (Most recent: Aug 29, 2018)
Last evaluated:	Sep 8, 2016
Accession:	VCV000037480.1
Variation ID:	37480
Description:	single nucleotide variant

Fiabilidad

Interpretation (Last evaluated)	Review status (Assertion criteria)	Condition (Inheritance)	Submitter	Supporting information (See all)
Pathogenic (Sep 08, 2016)	reviewed by expert	Breast-ovarian cancer, familial 1 Allele origin: germline	Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) Accession: SCV000299807.2 Submitted: (Sep 13, 2016)	Evidence details Comment: Variant allele predicted to encode a truncated non-functional protein.

Pathogenic
Likely pathogenic
Uncertain ignificance
Likely Benign
Benign

Información Adicional

NM_007294.3(BRCA1):c.2677A>T (p.Lys893Ter)

Allele ID:	46036
Variant type:	single nucleotide variant
Variant length:	1 bp
Cytogenetic location:	17q21.31
Genomic location:	17: 41244871 (GRCh37) GRCh37 UCSC 17: 43092854 (GRCh38) GRCh38 UCSC

HGVS:

Nucleotide	Protein	Molecular consequence
NC_000017.10:g.41244871T>A		
NC_000017.11:g.43092854T>A		
NM_007294.3:c.2677A>T	NP_009225.1:p.Lys893Ter	nonsense

... more HGVS

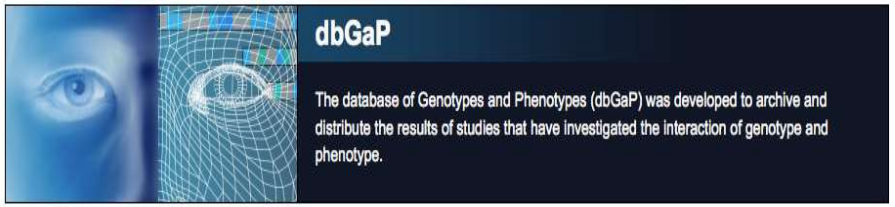
Bases de datos Biomédicas

OMIM® - Online Mendelian Inheritance in Man

Enfermedades humanas genéticas.
Genes humanos con fenotipo asociado
Constante revisión por la comunidad clínica
Descripción detallada por caso publicado

Total number of phenotypes* for which the molecular basis is known	6,526
Total number of genes with phenotype-causing mutation	4,171

Number of genes with 1 phenotype	2,884
Number of genes with 2 phenotypes	778
Number of genes with 3 phenotypes	273
Number of genes with 4+ phenotypes	236



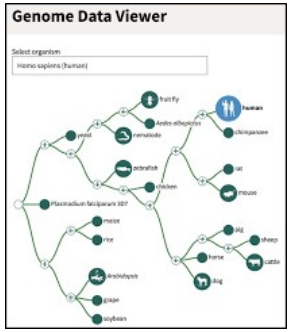
Enfermedades Raras

HPO: Human Phenotype Ontology

Navegadores del Genoma

http://www.ensembl.org/index.html

https://genome-euro.ucsc.edu/cgi-bin/hgGateway?redirect=manual&source=genome.ucsc.edu



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

Instituto de Salud Carlos III

Genome Reference Consortium

Ensembl Release 98 (September 2019), GENCODE 32, dbSNP152 …

e!

e!G

GRCh38

45 especies

- Vertebrates

- Other representative species

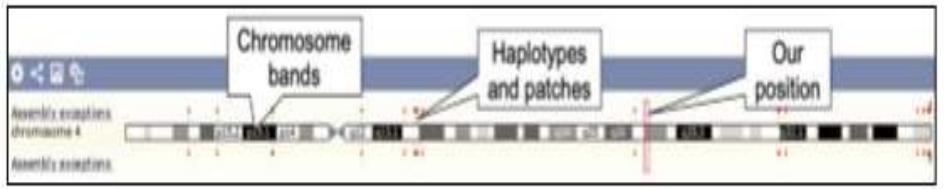
- Bacteria  
- Fungi  
- Protists  
- Metazoa  
- Plants  

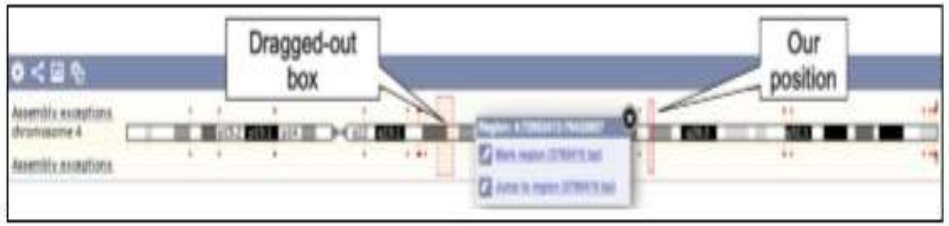
Instituto de Salud Carlos III

- Entrar en Ensembl homepage (http://www.ensembl.org/).
- Visualizar el kariotipo
- Búsqueda por un gen

Localización

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





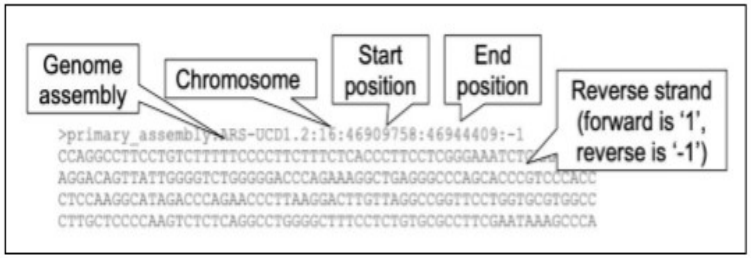
Instituto de Salud Carlos III

Gen

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

Genes rojos y amarillos: codifican proteinas
Genes grises, azules y púrpura: non-codificantes

Click en Secuencia



Click en Muestra SNPs

Click en Ontologias

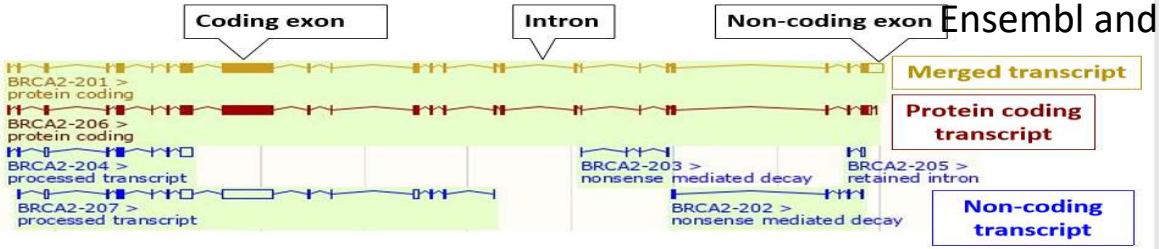
Click en External Referencias: Expression Atlas

Instituto de Salud Carlos III

- Click [Show transcript table]
- Clic en un transcrito.

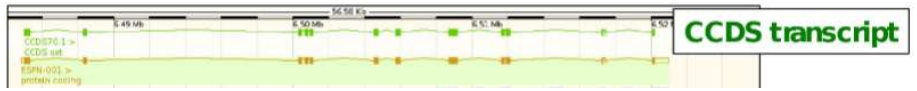
Transcript

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



Ensembl and HAVANA

Pink or green



Consensus Coding Sequence Set

The CCDS set is built by consensus among Ensembl, the National Center for Biotechnology Information (NCBI), and the HUGO Gene Nomenclature Committee (HGNC) for human or Mouse Genome Informatics (MGI)

Anotación de los transcritos

- Transcript support level (TSL): grado de certeza que tiene el transcrito (1 = mejor)
- GENCODE : transcritos anotados completamente de 5' a 3'
- APPRIS principal isoforma combinando información estructural y funcional, P1-P5 (1 = mejor), si es ALT es que está conservada en tres especies.

- MANE: Matched Annotation from NCBI and EMBL-EBI, clinically most relevant

Instituto de Salud Carlos III

Transcript

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

Click en Exons

Click en Protein Summary

Click en Variant Table

Filtrar por Clin Sig. (significado clínico)

Entrar en ClinVar

Variant

Click en General identifiers

Instituto de Salud Carlos III

Gen

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

Regulación

Marcas Epigenéticas
Metilación del ADN
Modificación de Histonas
Sitios de union de Factores de Transcripción (potenciadores y silenciadores)
Cromatina abierta/cerrada

148 cells/tissues

Custom tracks

Ensembl BioMart

Herramienta web (con API) para realizar consultas complejas y filtrar las diversas bases de datos de Ensembl (genes de Ensembl, cepas de ratón, variación de Ensembl y regulación de Ensembl).

http://www.ensembl.org/biomart/martview/

Base de datos

Ensembl Genes, Mouse Strains, Ensembl Variation, or Ensembl Regulation

Filtros

Input y Filtros sobre cromosomas, regiones genómicas, homología entre especies o funciones biológicas. También es posible introducir nuestra propia lista de genes

Phenotype: Amyotrophic lateral sclerosis

Atributos

Output deseado

Genes/términos GO
Variantes