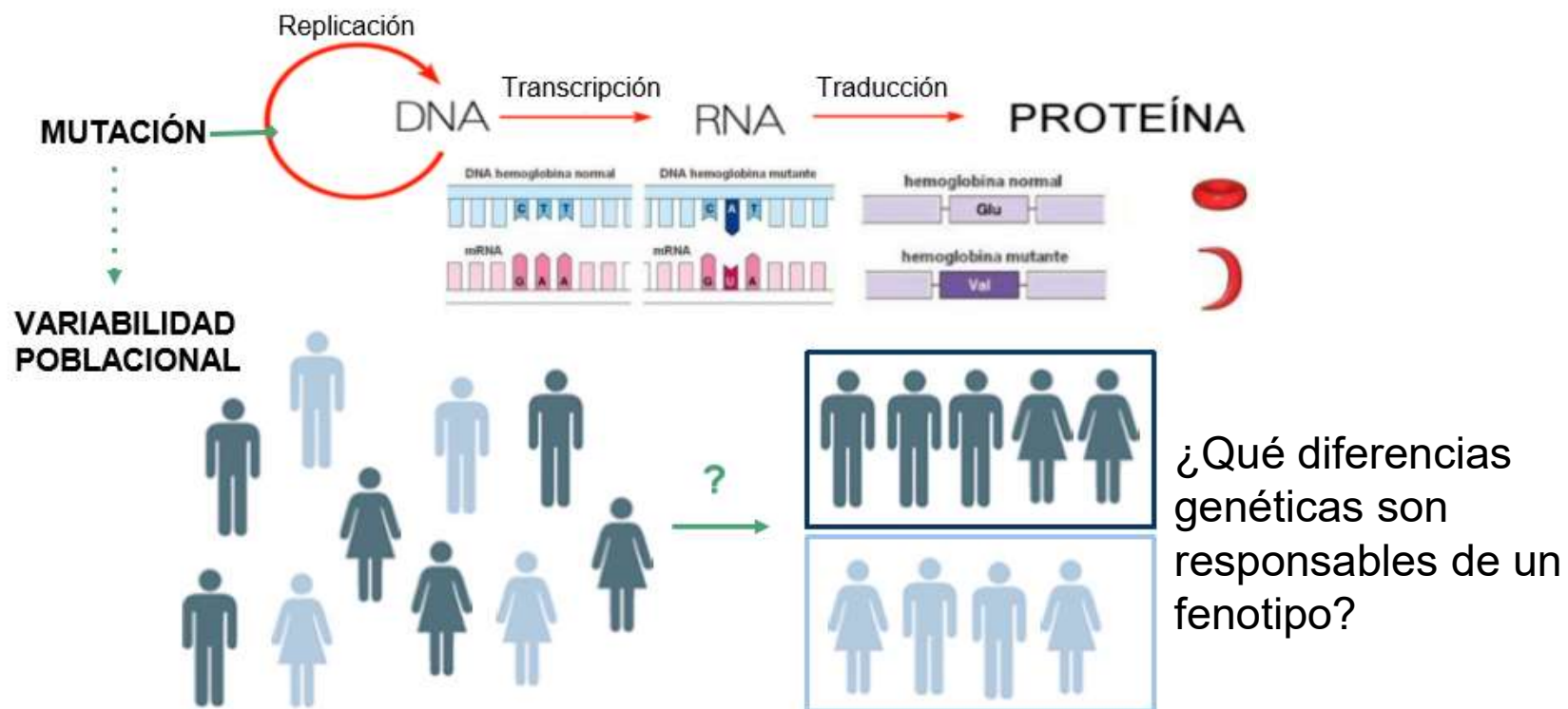




## Variantes Genéticas



## Tipos de variación genética

GCRhg37: Genome Reference Consortium human genome (build 37)

**REFERENCIA** AATGCCAGTAAGTTCGGGCCAGTGTTGTACCAATTCTGCGGCA

**SNV** AATGCCAGTAAGTTCGGG**A**CAGTGTTGTACCAATTCTGCGGCA

**SNP** > 1% en la población

**INDELS**

**INSERCIÓN** AATGCCAGTAAGTTCGGG**CT**CAGTGTTGTACCAATTCTGCGGCA

**DELECIÓN** AATGCCAGTAAG**■**CGGGCCAGTGTTGTACCAATTCTGCGGCA

**SVs**: Variantes estructurales como duplicaciones, grandes deleciones ó inseciones (CNVs), inversiones y translocaciones.



**dbSNP y dbVAR**

## Análisis de la variabilidad genética

**SNPs: 90%** variabilidad genética

Aparece 1 SNP cada **1300 pb**

1 variante  $\longrightarrow$  1 fenotipo ? . Enfermedades mendelianas

Sin impacto

Individualidad: pigmentación, receptores olfativos, ...

Susceptibilidad a enfermedades complejas.

Metabolismo y inmunidad

Respuesta farmacológica

Gran número de variantes por individuo:

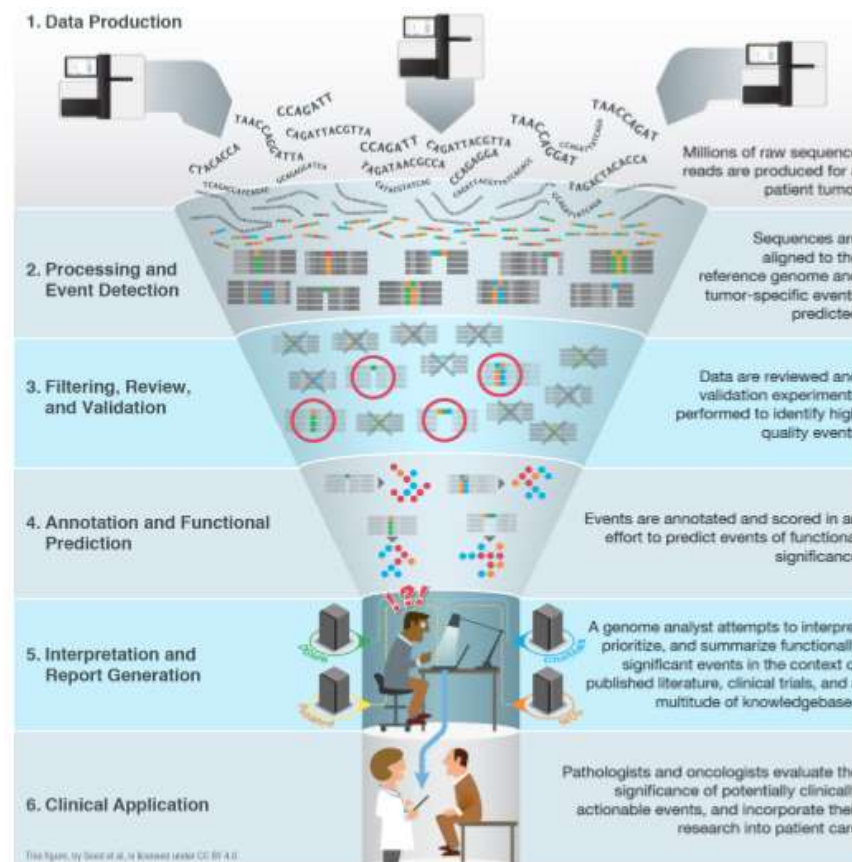
Exoma: 250.000 variantes.

Genoma: > 6 M de variantes.

Exoma clínico (6700 genes: 10.000-15.000 variantes)



# NGS: Análisis de Variantes



Good BM, Ainscough BJ, McMichael JF, Su AI, Griffith OL. 2014. Genome Biology. 15(8):438.

```

tophat2 -p 4 \
--library-type fr-unstranded \
--no-novel-juncs \
-G Saccharomyces_cerevisiae.EF4.69.gtf \
-o control \
Saccharomyces_cerevisiae.EF4.69.dna.toplevel.fa \
control.fastq
  
```

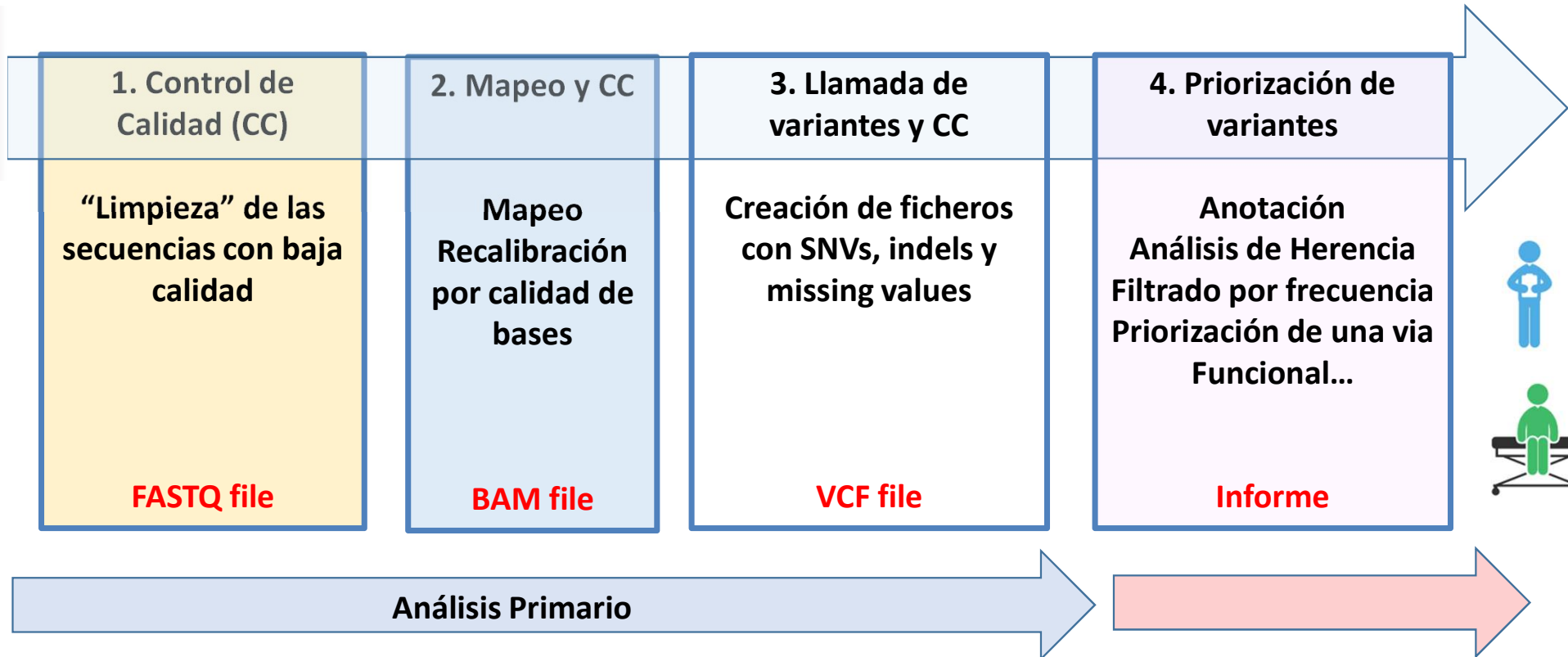
**Galaxy** es una plataforma web de software libre  
 para el análisis de datos en investigación  
 biomédica.

Accesible  
 Reproducible  
 Transparente




<https://wiki.galaxyproject.org/PublicGalaxyServers>

# NGS: Análisis de Variantes





Diagnóstico del gen causante de la enfermedad de un niño afectado de osteopetrosis (proband), padre y madre no afectados a partir de los exomas del trio

1. En carpeta de Datos (forward/reverse orientation of all the fragments)
2. Subir a Galaxy archivos **proband\_R1.fq, proband\_R2.fq** como fastqsangergz



<https://usegalaxy.eu/>

The screenshot shows the Galaxy / Europe web interface. The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', and 'Login or Register'. The left sidebar contains a 'Tools' section with a search bar and a list of tool categories: 'Get Data', 'Send Data', 'Collection Operations', 'Expression Tools', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'GENOMIC FILE MANIPULATION', and 'Convert Formats'. The main content area features a quote from Prof. Stephen Hawking, a 'News' section with several updates, and a 'History' section on the right which is currently empty. An orange arrow points from the URL <https://usegalaxy.eu/> to the 'Tools' section in the sidebar.

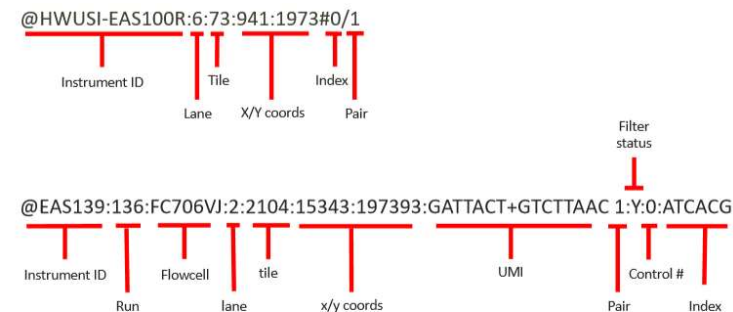


## Visualizar el archivo que has subido Fastq

```

@HWUSI-EAS100R:6:73:941:1973#0/1
CTTTTTTATTTTGTCTGACTGGGTTGATTCAAAA
+
CCCFFFFHHHHGJHIJHIHIIIFHIJJJJIJG
  
```

encabezado "@"  
 Línea de secuencia  
 Línea espaciadora  
 Valores de calidad



### Phred quality score

ASCII code	33	59	64	73	104	126
ASCII code	!"#\$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMN	O	PQRSTUVWXYZ[\]^_`	abcdefghijklmnopqrstuvwxyz{~		
Sanger	0	26	31	40		
Solexa		-5	0	9	40	
Illumina 1.3+			0	9	40	
Illumina 1.5+			3	9	40	
Illumina 1.8+	0	26	31	41		

Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%
50	1 in 100000	99.999%

## Tools



Fastqc



### FASTA/FASTQ

**Combine FASTA and QUAL** into FASTQ

**Manipulate FASTQ** reads on various attributes

**fastp** - fast all-in-one preprocessing for FASTQ files

### FASTQ Quality Control

**FastQC** Read Quality reports

### Mapping



Executed **FastQC** and successfully added 1 job to the queue.

The tool uses this input:

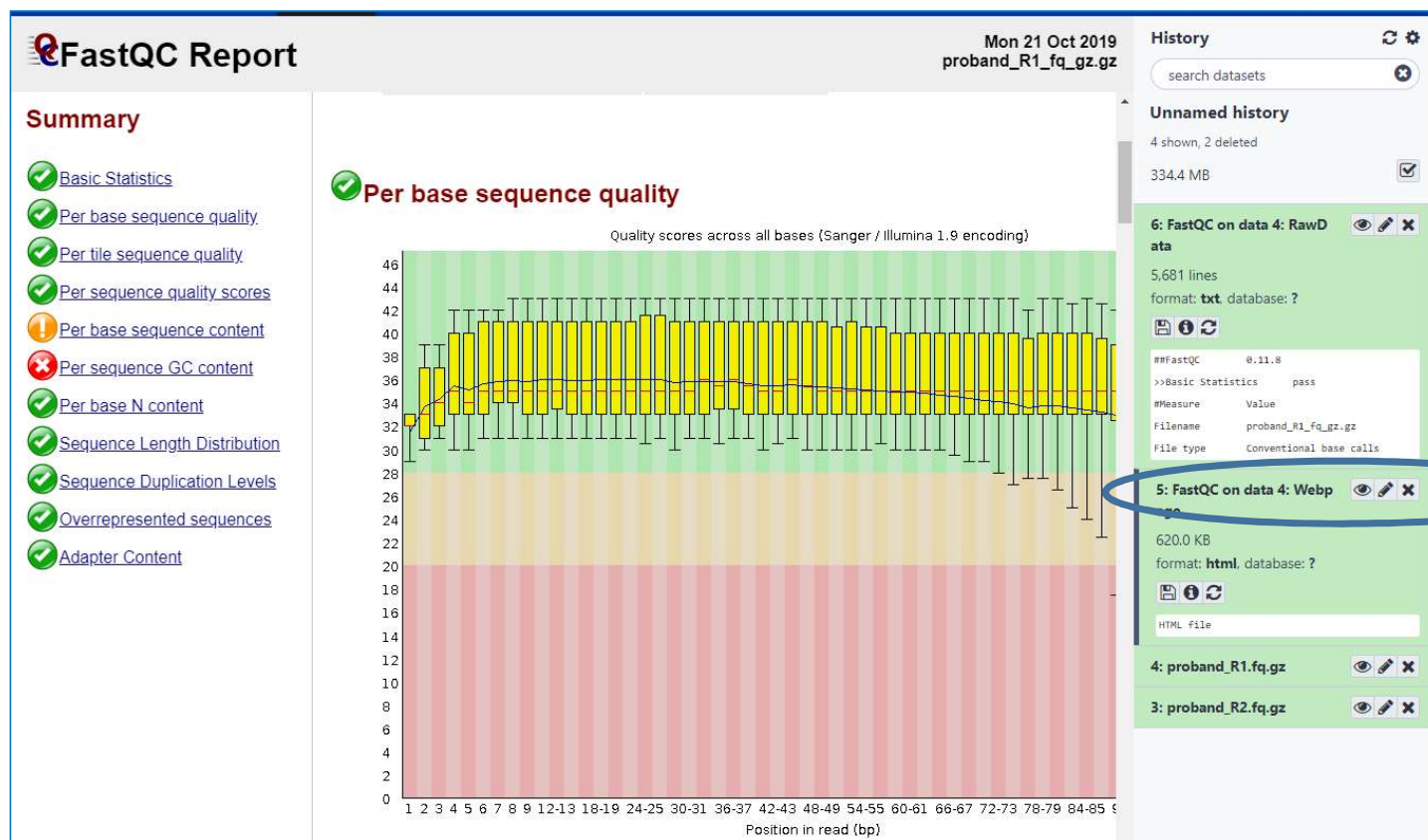
- **4: proband\_R1.fq.gz**

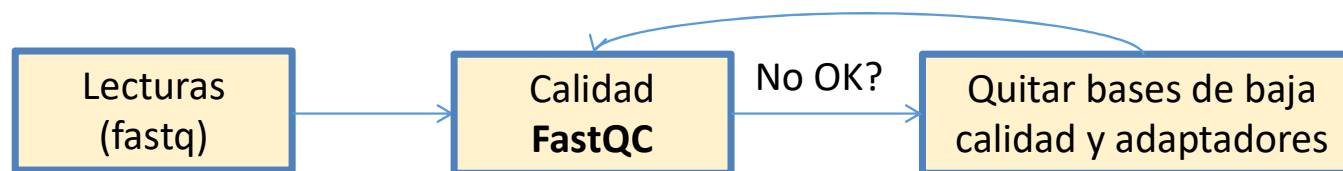
It produces 2 outputs:

- **5: FastQC on data 4: Webpage**
- **6: FastQC on data 4: RawData**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

## Análisis de calidad con FastQC





Phred score  $\sim 20$ , se deberían filtrar estas lecturas, usando **Cutadapt**




- “Minimum length”: 20
- “Quality cutoff”: 20
- Adapter removal (if any), we choose to remove ends ( 5’ and/or 3’) with low-quality, here below 20 in quality).

**Cutadapt** Remove adapter sequences from Fastq/Fasta (Galaxy Version 1.16.5) Versions Options


**Single-end or Paired-end reads?**

single-end


**FASTQ/A file**

  4: proband\_R1.fq.gz 

Should be of datatype "fastq.gz" or "fasta"


**Read 1 Options** 

**3' (End) Adapters**

 Insert 3' (End) Adapters


Sequence of an adapter ligated to the 3' end (paired data: of the first read). The adapter and subsequent bases are trimmed. If a 'S' character is appended (anchoring), the adapter is only found if it is a suffix of the read. To search for a linked adapter, separate the 2 sequences with 3 dots (ADAPTER1..ADAPTER2), see Help below.

**5' (Front) Adapters**

 Insert 5' (Front) Adapters

Sequence of an adapter ligated to the 5' end (paired data: of the first read). The adapter and any preceding bases are trimmed. Partial matches at the 5' end are allowed. If a '^' character is prepended (anchoring), the adapter is only found if it is a prefix of the read. To search for a linked adapter, separate the 2 sequences with 3 dots (ADAPTER1..ADAPTER2), see Help below.

**5' or 3' (Anywhere) Adapters**


 Insert 5' or 3' (Anywhere) Adapters


Sequence of an adapter that may be ligated to the 5' or 3' end (paired data: of the first read). Both types of matches as described under 3' and 5' Adapters are allowed. If the first base of the read is part of the match, the behavior is as with 5' Adapters, otherwise as with 3' Adapters. This option is mostly for rescuing failed library preparations - do not use if you know which end your adapter was ligated to!


**Cut bases from reads before adapter trimming**


0

Remove bases from each read (first read only if paired). If positive, remove bases from the beginning. If negative, remove bases from the end. This is applied "before" adapter trimming. (-u)

**Adapter Options** 

**Filter Options** 

**Read Modification Options** 

**Output Options** 

## Mapecto con **BWA-MEM** utilizando el genoma Humano hg19



Tools

BWA

FASTA/FASTQ

Trim Galore! Quality and adapter trimmer of reads

AB-SOLID DATA

Convert SOLiD output to fastq

Annotation

TB-Profiler Profile Infer strain types and drug resistance markers from sequences

Mapping

Map with minimap2 A fast pairwise aligner for genomic and spliced nucleotide sequences

Map with BWA - map short reads (< 100 bp) against reference genome

Map with BWA-MEM - map medium and long reads (> 100 bp) against reference genome

Map with BWA-MEM

Variant Calling

SnpEff build: database from Genbank or GFF record

snippy Snippy finds SNPs between a haploid reference genome and your NGS sequence reads.

Epigenetics

bwameth Fast and accurate aligner of

Use a built-in genome index

Built-ins were indexed using default options. See 'Indexes' section of help below

Using reference genome

Human (Homo sapiens): hg19 Full

Select genome from the list

Single or Paired-end reads

Paired

Select between paired and single end data

Select first set of reads

4: proband\_R1.fq.gz

Specify dataset with forward reads

Select second set of reads

3: proband\_R2.fq.gz

Specify dataset with reverse reads

Enter mean, standard deviation, max, and min for insert lengths.

-l: This parameter is only used for paired reads. Only mean is required while sd, max, and min will be inferred. Examples: both "250" and "250,25" will work while "250,10" will not. See below for details.

Set read groups information?

Do not set

Specifying read group information can greatly simplify your downstream analyses by allowing combining multiple datasets.

Select analysis mode

1.Simple Illumina mode

Execute

History

search datasets

Unnamed history

4 shown, 2 deleted

334.4 MB

6: FastQC on data 4: RawData

5,681 lines

format: txt database: ?

##FastQC 0.11.8

>>Basic Statistics pass

#Measure Value

Filename proband\_R1.fq.gz

File type Conventional base calls

5: FastQC on data 4: Webpage

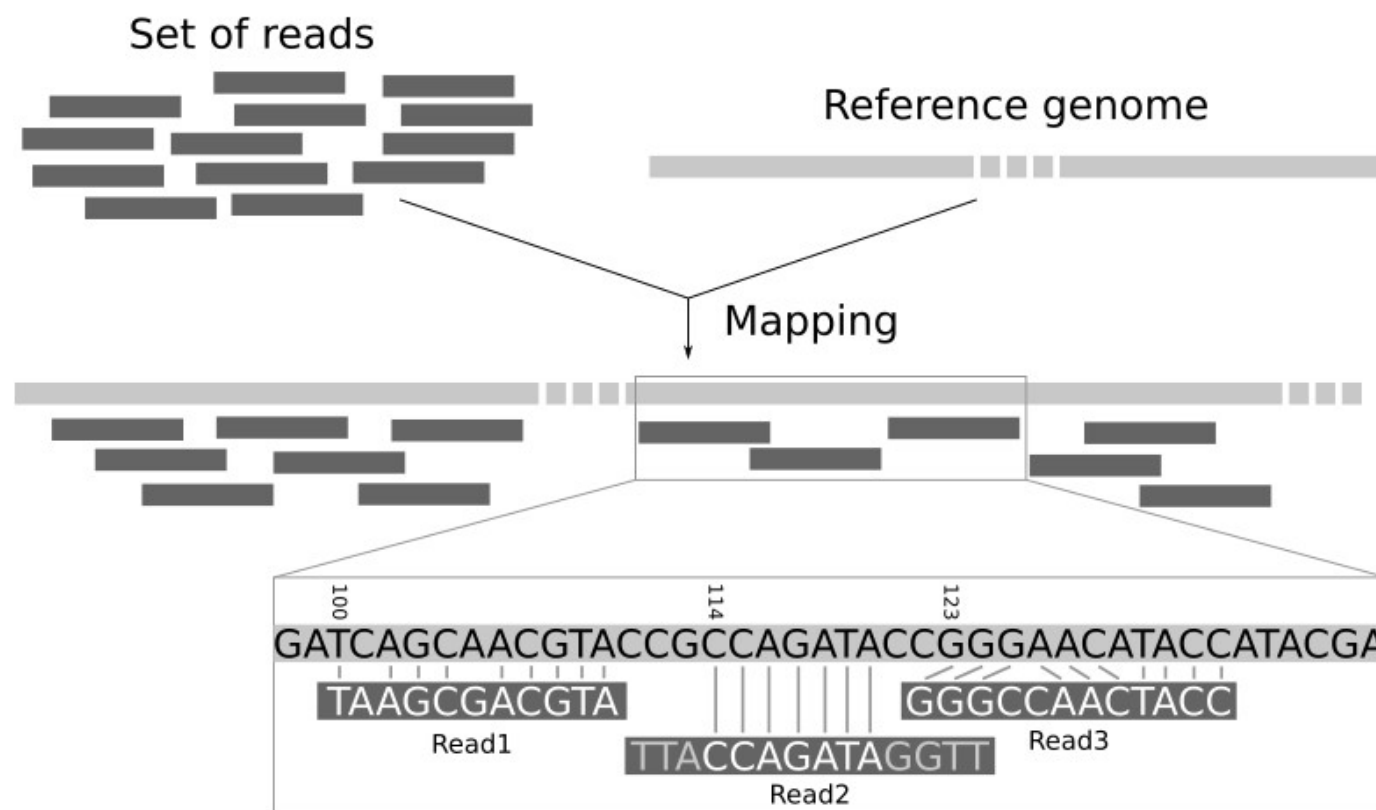
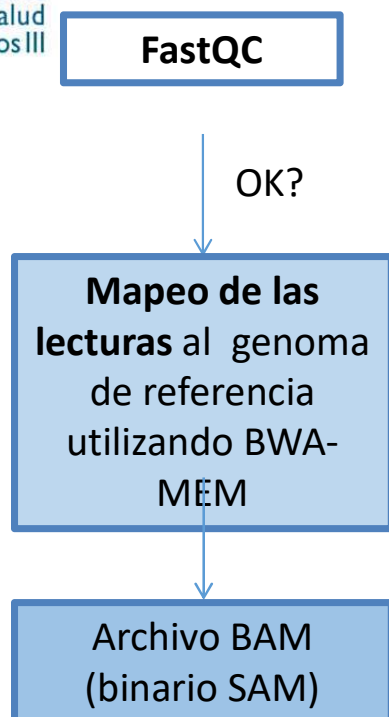
620.0 KB

format: html database: ?

HTML file

4: proband\_R1.fq.gz

3: proband\_R2.fq.gz



Source: Galaxy



SAM = Sequence Alignment Map  
BAM = Binary SAM = compressed SAM

## Encabezado

@HD The header line  
VN: format version  
SO: Sorting order of alignments

@SQ Reference sequence dictionary  
SN: reference sequence name  
LN: reference sequence length  
SP: species

@RG Read group  
ID: read group identifier  
CN: name of sequencing center  
SM: sample name

@PG Program  
PN: program name  
VN: program version

## Alineamiento

Nº	Campo	Tipo	Exp Regular / Rango	Descripcion
1	QNAME	Cadena	[!-?A-~]{1,255}	Nombre de la consulta
2	FLAG	Entero	[0, 2 <sup>16</sup> - 1]	Bandera de opciones
3	RNAME	Cadena	\ *  [! - () + - < > - ~]  [! - ~] *	Referencia del nombre de la secuencia
4	POS	Entero	[0, 2 <sup>29</sup> - 1]	Posición de la primera base más a la izquierda
5	MAPQ	Entero	[0, 2 <sup>8</sup> - 1]	Calidad del Mapeo
6	CIGAR	Cadena	\ *  ([0 - 9] + [MIDNSHPX =]) +	Cadena CIGAR
7	RNEXT	Cadena	\ *   =  [! - () + - < > - ~]  [! - ~] *	Nombre de referencia del siguiente fragmento
8	PNEXT	Entero	[0, 2 <sup>29</sup> - 1]	Posición el siguiente fragmento
9	TLEN	Entero	[-2 <sup>29</sup> + 1, 2 <sup>29</sup> - 1]	Longitud de la plantilla
10	SEQ	Cadena	\ *  [A - Za - z = .] +	Fragmento de secuencia
11	QUAL	Cadena	[! - ~] +	Calidad de la secuencia

<https://samtools.github.io/hts-specs/SAMv1.pdf>



Galaxy / Europe

Tools

sam

**SAM/BAM Manipulation**

- Filter pileup** on coverage and S
- Generate pileup** from BAM dat
- SAM-to-BAM** convert SAM to f
- Samtools flagstat** tabulate descriptive stats for BAM dataset
- samtools mpileup** multi-way pi of variants
- Slice** BAM by genomic regions
- Pileup-to-Interval** condenses p format into ranges of bases
- BAM-to-SAM** convert BAM to :
- Samtools stats** generate statist BAM dataset
- CalMD** recalculate MD/NM tags
- Samtools idxstats** reports stats of the BAM index file
- Samtools sort** order of storing aligned sequences
- Filter SAM or BAM, output SAM or BAM** files on FLAG MAPQ RG LN or by region
- sort** a BAM file

**SAM Tools**

- Samtools stats** generate statistics for BAM dataset
- Samtools flagstat** tabulate descriptive stats for BAM dataset
- Samtools idxstats** reports stats of the BAM index file
- Filter SAM or BAM, output SAM or BAM** files on FLAG MAPQ RG LN or by region
- Split** BAM dataset on readgroups
- BedCov** calculate read depth for a set of genomic intervals
- Samtools depth** compute the depth at each position or region
- Samtools sort** order of storing aligned sequences

Estadística del alineamiento BAM

## Samtools flagstat

Control de calidad de los archivos BAM

**RmDup** Quita duplicados de PCR

*“Filter on bitwise flag”*: **yes**

*“Only output alignments with all of these flag bits set”*:

param-check *“Read is mapped in a proper pair”*

*“Is this paired-end or single end data”*: **BAM is paired-end**

*“Treat as single-end”*: **No**

Visualizar los archivos BAM

**IGV**

## Llamada de variantes

REFERENCE: atcatgacggcaGtagcatat

-----  
READ1: atcatgacggcaGtagcatat

READ2: tgacggcaGtagcatat

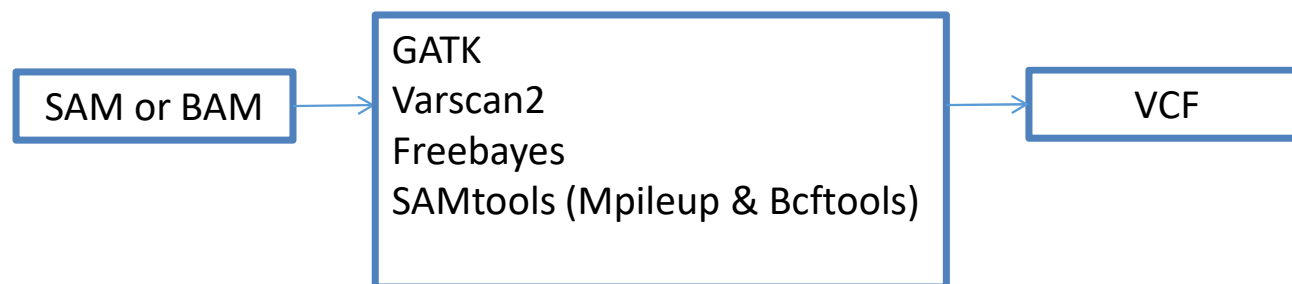
READ3: atcatgacggcaAtagca

READ4: cggcaGtagcatat

READ5: atcatgacggcaGtagc

Podría ser:

- una verdadera variante
- un artefacto experimental, ( un error de preparación de la muestra)
- un error de llamada de la base
- un error de análisis, una desalineación (aunque poco probable en el ejemplo anterior)



FreeBayes bayesian genetic variant detector (Galaxy Version 1.0.2.29-3)

Choose the source for the reference genome

Locally cached

BAM file

8: RmDup on data 7  
7: Map with BWA-MEM on data 3 and data 4 (mapped reads in BAM format)

Using reference genome

Human (Homo sapiens): hg19

Limit variant calling to a set of regions?

Do not limit

Sets --targets or --region options

Choose parameter selection level

1. Simple diploid calling

Select how much control over the freebayes run you need

✓ Execute

## bcftools norm.

[https://genome.sph.umich.edu/wiki/Variant\\_Normalization](https://genome.sph.umich.edu/wiki/Variant_Normalization)

- “When any REF allele does not match the reference genome base”: ignore the problem (-w)
- “Left-align and normalize indels?”: Yes
- “Perform deduplication for the following types of variant records”: do not deduplicate any records
- Freebayes is not producing any duplicate calls.
- “~multiallelics”: split multiallelic sites into biallelic records (-)
- split the following variant types”: both

# VCF = Variant Call Format

## ## METAINFO

```

##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">

```

## # CABECERA

## VARIANTES

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	NA000001	NA000002	NA000003
20	14370	rs6054257	G	A	29	PASS	NS=3;DP=14;AF=0.5;DB;H2	GT:GQ:DP:HQ	0 0:48:1:51,51	1 0:48:8:51,51	1/1:43:5:..
20	17330	.	T	A	3	q10	NS=3;DP=11;AF=0.017	GT:GQ:DP:HQ	0 0:49:3:58,50	0 1:3:5:65,3	0/0:41:3
20	1110696	rs6040355	A	G,T	67	PASS	NS=2;DP=10;AF=0.333,0.667;AA=T;DB	GT:GQ:DP:HQ	1 2:21:6:23,27	2 1:2:0:18,2	2/2:35:4
20	1230237	.	T	.	47	PASS	NS=3;DP=13;AA=T	GT:GQ:DP:HQ	0 0:54:7:56,60	0 0:48:4:51,51	0/0:61:2
20	1234567	microsat1	GTC	G,GTCT	50	PASS	NS=3;DP=9;AA=G	GT:GQ:DP	0/1:35:4	0/2:17:2	1/1:40:3
...											

## Priorización de Variantes

- VEP
- GEMINI

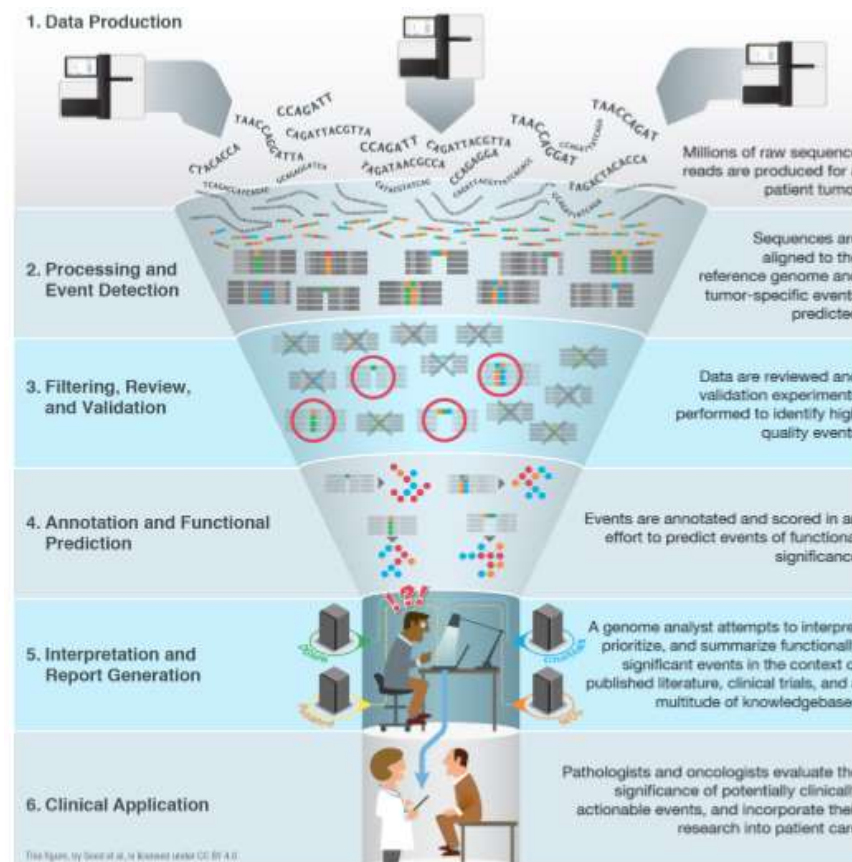
**Ve!P**

**Galaxy**



Integrative  
Genomics  
Viewer

# NGS: Análisis de Variantes



Good BM, Ainscough BJ, McMichael JF, Su AI, Griffith OL. 2014. Genome Biology. 15(8):438.





## Anotación en Bases de datos clínicas

**Guia American College of Medical Genetics and Genomics /Association of Molecular Pathology  
ACMG/AMP 2015**

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data			<b>Significado incierto</b> <b>Benigna</b> <b>Probablemente benigna</b> <b>Probablemente patológica</b> <b>Patológica</b>			Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3					
Segregation data	Nonsegregation with disease BS4					
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in trans with a dominant variant BP2 Observed in cis with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM5		

Asociada con la condición estudiada y modo de herencia





## Frecuencia poblacional

Catálogos de variabilidad genética.

El proyecto 1000 genomas:

<http://www.1000genomes.org>

Exome Sequencing Project (ESP):

<http://evs.gs.washington.edu/EVS>

Exome Aggregation Consortium (ExAC/ gnomAD):

<http://exac.broadinstitute.org>

<http://gnomad.broadinstitute.org/about>

The CIBERER Spanish Variant Server (CSVs):

<http://csvs.clinbioinfospa.es>

MAF: Frecuencia del alelo minoritario —> Incidencia enfermedad

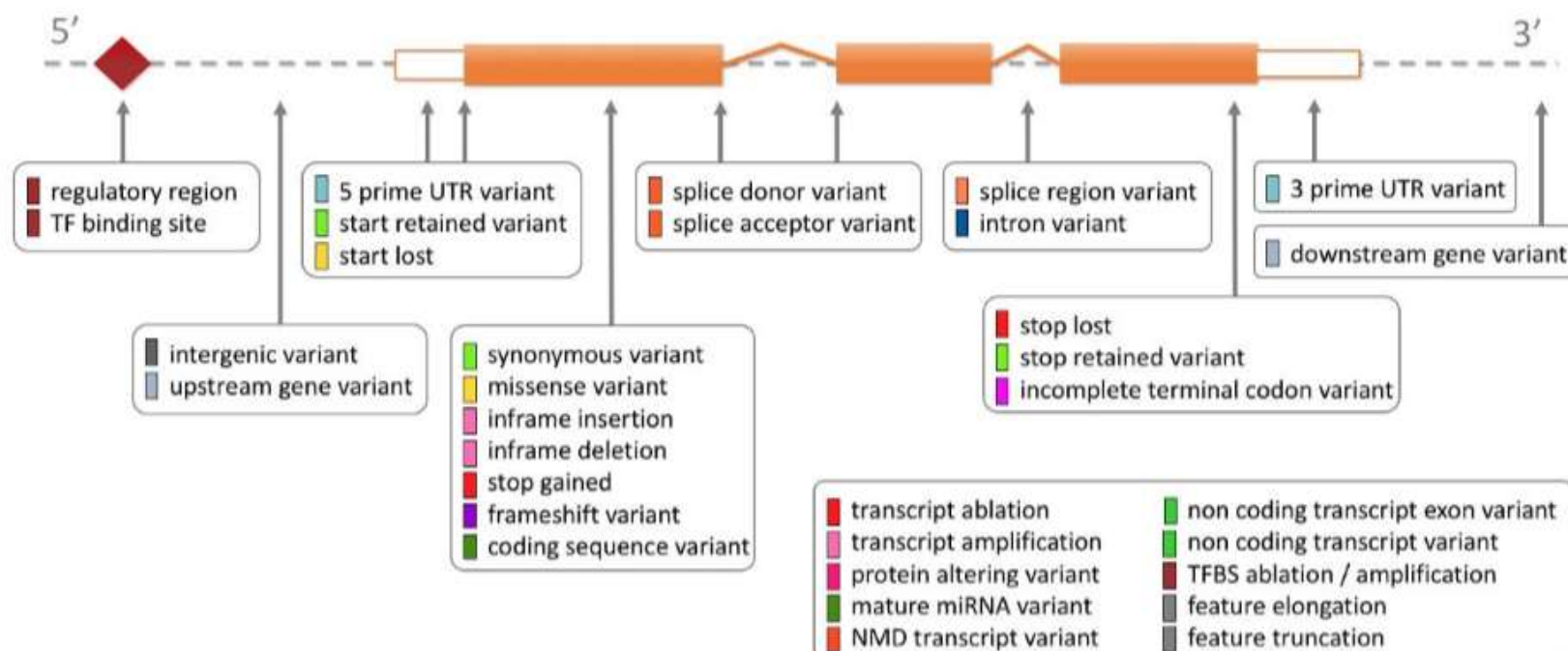
Tipo de herencia  
Penetrancia  
Edad de inicio



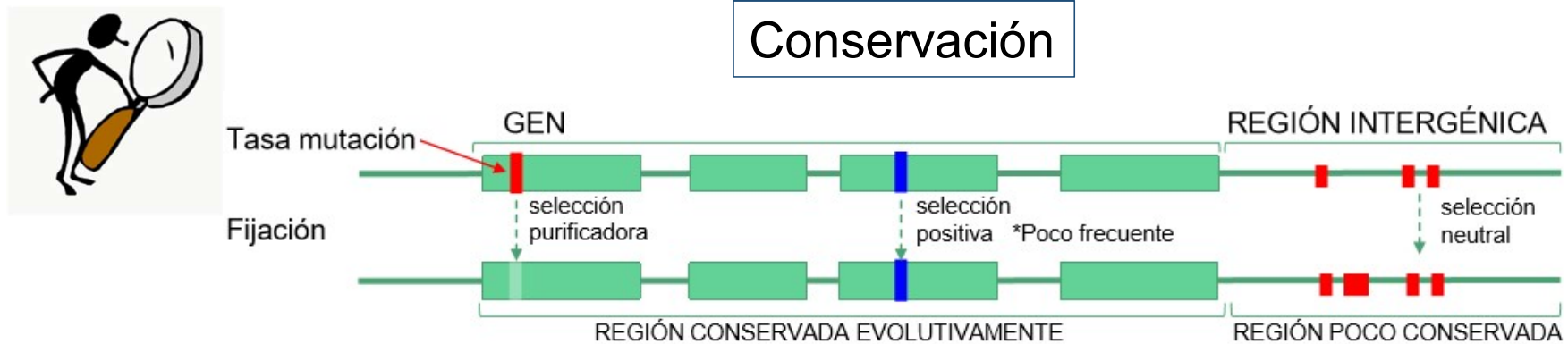


## Consecuencia: Sequence Ontology (SO)

Standardised variant consequence terms as defined by The Sequence Ontology  
<http://www.sequenceontology.org>



LoF (Loss of Function): Provocan pérdida de función: splicing, STOP, ...



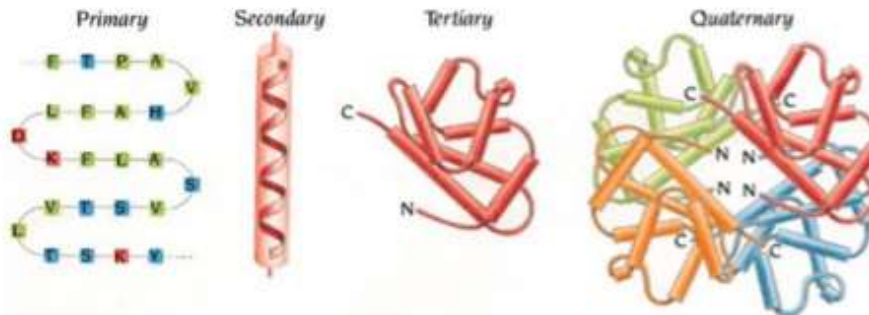
Regiones con funciones importantes para la célula acumulan poca variabilidad: más conservadas

**GERP:** Genomic Evolutionary Rate Profiling. Método para estimar la presión selectiva de una posición específica mediante máxima verosimilitud. A mayor valor, mayor presión selectiva.

**PhastCons:** Calcula la conservación por ventanas usando un modelo filogenético HMM (hidden Markov model). - [0 - 1]: Representa la probabilidad de que una región esté sometida a su mayor conservación posible.

**PhyloP:** Calcula el -log p-valor bajo una hipótesis nula de evolución neutral. - En humanos [-14, 3]: scores positivos indican conservación, negativos indican una aceleración.

## Patogenicidad



La **función** de una proteína depende de su **estructura** cuaternaria final.

Su estructura depende de las **propiedades fisicoquímicas** de los **aminoácidos** que la componen.

**SIFT** [0 - 1] sorts intolerant from tolerant Probabilidad de tolerancia basada en la homología de proteínas entre alineamientos.

**Polyphen** [0 - 1] Polymorphism Phenotyping v2 Usa **parámetros fisicoquímicos** y anotación **estructural** para calcular, mediante métodos bayesianos, la patogenicidad de un cambio de aminoácido.

**CADD**: Combined Annotation Dependent Depletion. Usando **información anotada disponible**, evalúa en un ranking el impacto de la variante.

ScaledCADD 10 = 10%

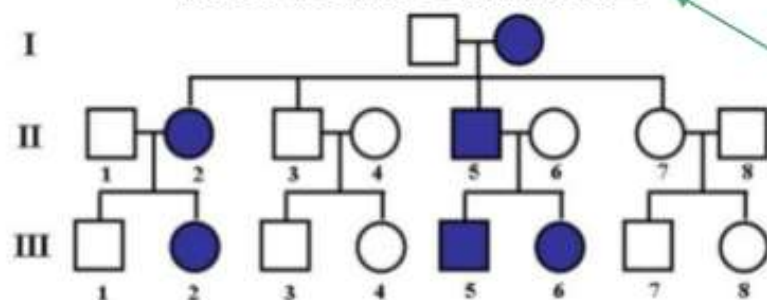
ScaledCADD 20 = 1%

**dbNSFP** Predicción funcional de variantes no sinónimas del genoma

Valor de SIFT	Predicción
< 0.05	<b>DAMAGING</b>
>= 0.05	<b>TOLERATED</b>
Valor de Polyphen	Predicción
> 0.908	<b>Probably damaging</b>
0.446 - 0.908	<b>Possibly damaging</b>
< 0.446	<b>Benign</b>

## Modos de herencia: Genotipo

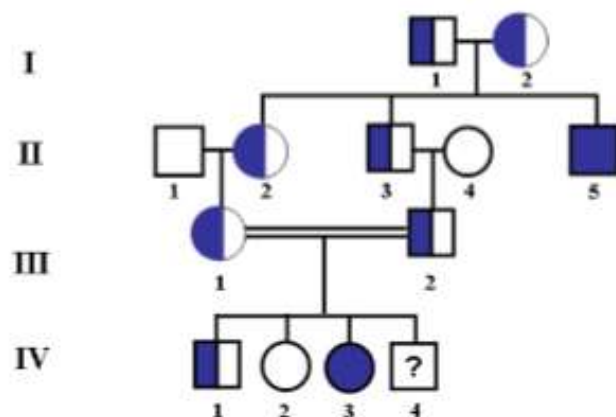
### HERENCIA DOMINANTE



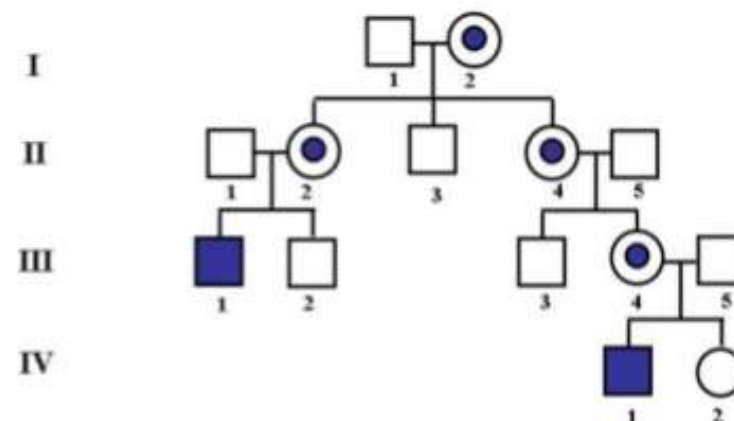
HERENCIA LIGADA AL X (DOMINANTE)

HERENCIA LIGADA AL X (RECESIVA)

### HERENCIA RECESIVA



### HETEROCIGOSIS COMPUESTA

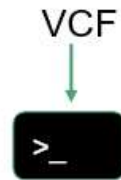


HERENCIA LIGADA AL Y

Sólo varones afectados, transmisión de padre a hijos.



## Software



ANNOVAR  
snpEff  
GEMINI



...  
Variant Effect Prediction VEP



Bystro

...

### Tener presente:

Los resultados pueden variar de una herramienta a otra

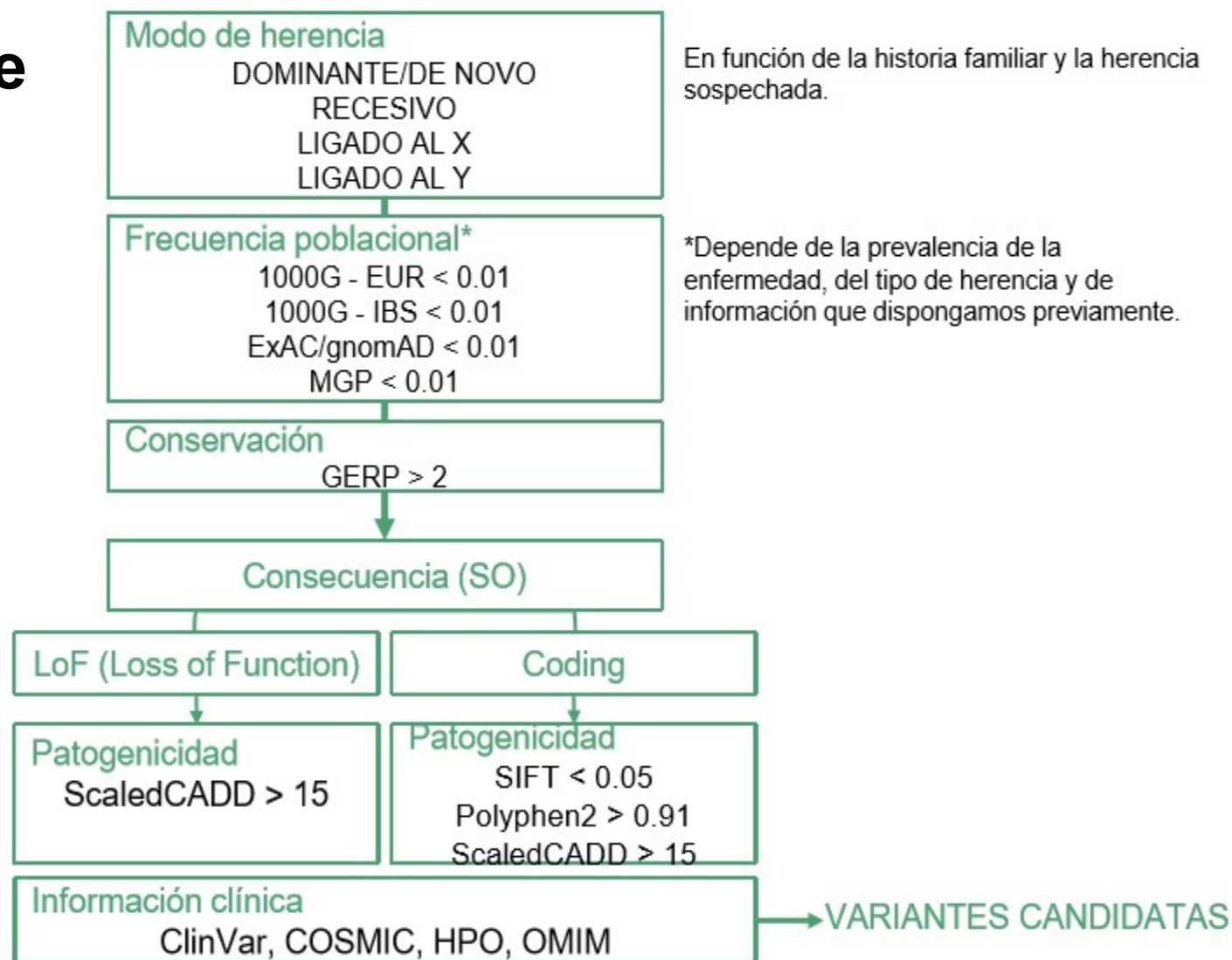
Falsos positivos

Errores en el efecto que describen de la variante...

(frameshift en lugar de stop loss...)



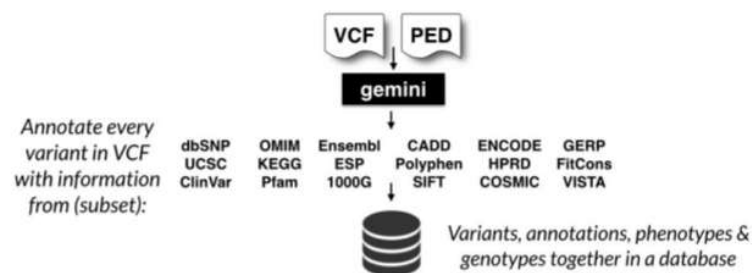
# Priorización de Variantes



!! Algunas variantes LoF tampoco tienen CADD asociado



## GEMINI (Genome MINing) in Galaxy



3 archivos VCF (father, mother, proband) anotados con SnpEff (Merge Galaxy)  
 archivo PED “pedigrí”

#family_id	name	paternal_id	maternal_id	sex	phenotype
FAM	father	0	0	1	1
FAM	mother	0	0	2	1
FAM	proband	father	mother	1	2

family\_trio.sqlite  
 (en datos del curso)  
 Subirlo con extensión  
 sqlite y anotado en  
 hg19

### GEMINI inheritance pattern

“Additional constraints expressed in SQL syntax”

impact\_severity != 'LOW’




**Ouput Custom. Set of columns to include in the variant report table:**

clinvar\_sig, clinvar\_disease\_name, clinvar\_gene\_phenotype, rs\_ids

# Anotación de Variantes

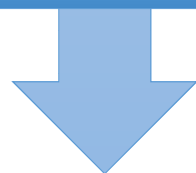
**Datos de  
Variantes**



- Affected gene, transcript and protein sequence
- Pathogenicity 
- Frequency data 
- Regulatory consequences
- Splicing consequences
- Literature citations 

IMPORTACIÓN DE  
VARIANTES  
DE DIVERSAS BASES DE  
DATOS

CONTROL DE CALIDAD



dbSNP  
Short Genetic Variations



DGVar<sup>archive</sup>

DECIPHER v5.1  
GRCh37

PubMed

gnomAD  
genome aggregation database

European Variation Archive

1000 Genomes  
A Deep Catalog of Human Genetic Variation

Genetics

COSMIC  
Catalogue of Somatic Mutations in Cancer

nextgen

GWAS  
Catalog

OMIM<sup>®</sup>

orphanet

UniProt



GANT

GEFOS

HGMD<sup>®</sup>  
The Human Gene Mutation Database  
Cardiff

MAGIC

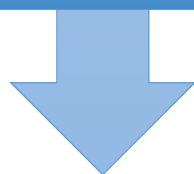
Illumina<sup>®</sup>

affymetrix  
part of Thermo Fisher Scientific

IMPORTACIÓN DE  
VARIANTES  
DE DIVERSAS BASES DE  
DATOS

CONTROL DE CALIDAD

ANOTACION CON  
DATOS DE POBLACION



1000 Genomes

A Deep Catalog of Human Genetic Variation

ClinVar



Europe PMC

OMIM<sup>®</sup>



gnomAD

genome aggregation database



UK10K

Rare Genetic Variants in Health and Disease



NHLBI Trans-Omics for Precision Medicine



## Predicción:

- Consecuencia de las variantes
- Predicción de la Función de proteínas
- Datos de desequilibrio de ligamiento
- Conservación de la variante entre especies





[http://grch37.ensembl.org/Homo\\_sapiens/Tools/VEP](http://grch37.ensembl.org/Homo_sapiens/Tools/VEP)

Subir el fichero probandch8.vcf\_bgzip

Añadir que incluya el dato de fenotipo

Filtrar por IMPACT is HIGH

Consequence is stop\_gained

Additional annotations ▾ Additional transcript, protein and regulatory annotations

**Transcript annotation**

Transcript biotype: ☒

Exon and intron numbers: ☐

Transcript support level: ☒ Find out more about the affected transcripts

APPRIS: ☒

Identify canonical transcripts: ☐

Upstream/Downstream distance (bp): 5000

miRNA structure: ☐

**Protein annotation**

Protein domains: ☐ Visualise the variant on a 3D protein structure

**Regulatory data**

Get regulatory region consequences: ☒ Get regulatory region consequences

**Phenotype data**

Phenotypes: ☐ Get phenotypes

Navigation (per variant)

Page: 1 of 1 | Show: 1 5 10 50 All variants

Filters

Impact is HIGH

Consequence is stop\_gained

Clear filters

Match all of the above rules

Update

Uploaded variant is defined

Add

Download

All: VCF VEP TXT

Filtered: VCF VEP TXT

BioMart: Variants Genes

New job

Show/hide columns (13 hidden)

Feature type	Feature	Biotype	Exon	cDNA position	CDS position	Protein position	Amino acids	Codons	Existing variant	Feature strand	SIFT	PolyPhen	AF	Clinical significance	Phenotype or disease	Pubmed	Associated phenotypes
Transcript	ENST00000285379.5	protein_coding	3/7	521	291	97	W/*	TGG/TGA	rs1173332839	1	-	-	-	-	-	-	OSTEOPETROSIS AUTOSOMAL RECESSIVE 3 (ENSG00000104267, MIM morbid) OSTEOPETROSIS AUTOSOMAL RECESSIVE TYPE 3 (ENSG00000104267, DDG2P) Osteopetrosis with renal tubular acidosis (ENSG00000104267, Orphanet)
Transcript	ENST00000321602.8	protein_coding	2/4	285	50	17	S/*	TCA/TGA	rs201723492	-1	-	-	-	-	-	-	-

# Anotación de Variantes

Variant coordinates (Ensembl default)	1    881907    881906    -/C    + 5    140532    140532    T/C    + 12   1017956   1017956   T/A    + 2    946507    946507    G/C    + 14   19584687   19584687   C/T    -
HGVS notation	ENST00000285667.3:c.1047_1048insC 5:g.140532T>C NM_153681.2:c.7C>T ENSP00000439902.1:p.Ala2233Asp NP_000050.2:p.Ile2285Val
VCF	#CHROM POS ID REF ALT 20 14370 rs6054257 G A 20 17330 . T A 20 1110696 rs6040355 A G,T 20 1230237 . T .
Variant IDs	rs41293501 COSM327779 rs146120136 FANCD1:c.475G>A rs373400041

<http://varnomen.hgvs.org/recommendations/general/>



Identifiers Additional identifiers for genes, transcripts and variants

**Identifiers**

Gene symbol: ☒

CCDS: ☐

Protein: ☐

UniProt: ☐

HGVS: ☐

Which identifiers do you want to see?

Variants and frequency data Co-located variants and frequency data

**Variants and frequency data**

Find co-located known variants: ☒ Yes

Frequency data for co-located variants:
 

- ☒ 1000 Genomes global minor allele frequency
- ☐ 1000 Genomes continental allele frequencies
- ☐ ESP allele frequencies
- ☐ gnomAD (exomes) allele frequencies

PubMed IDs for citations of co-located variants: ☐

Include flagged variants: ☐

Find out if your variants already exist in our database

Get frequency data

Predictions Variant predictions, e.g. SIFT, PolyPhen

**Pathogenicity predictions**

SIFT:

PolyPhen:

dbNSFP: ☒ Disabled ☐ Enabled

CADD: ☐

Condel: ☒ Disabled ☐ Enabled

LoFtool: ☐

**Splicing predictions**

dbSNV: ☐

MaxEntScan: ☐

**Conservation**

BLOSUM62: ☐

Ancestral allele: ☐

Pathogenicity predictions

Filtering options Pre-filter results by frequency or consequence type

**Filters**

Filter by frequency:
 

- ☒ No filtering
- ☐ Exclude common variants
- ☐ Advanced filtering

Return results for variants in coding regions only: ☐

Restrict results:

NB: Restricting results may exclude biologically important data!

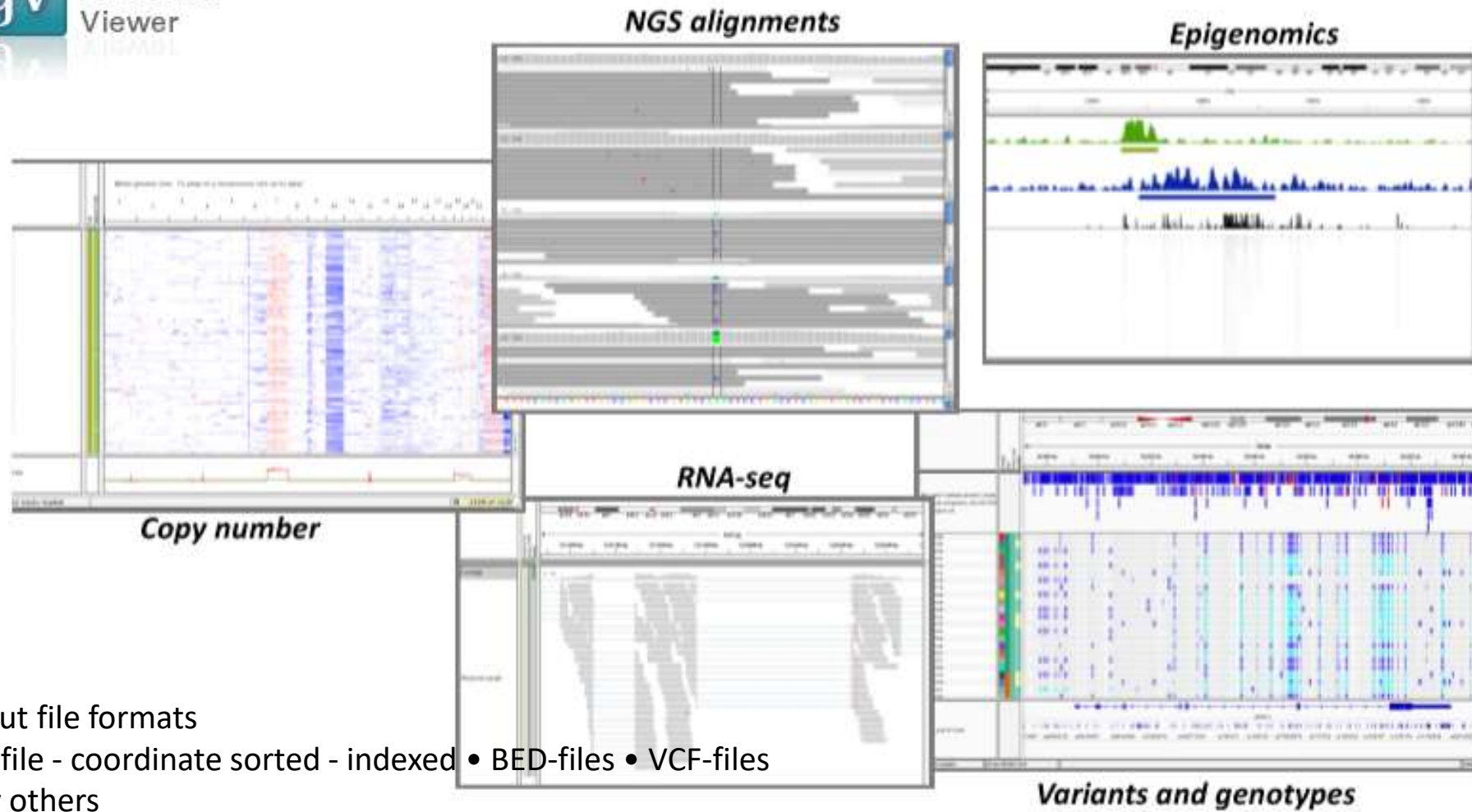
Choose to only see variants with specific allele frequencies

Advanced options Settings to optimise VEP

Run



Integrative  
Genomics  
Viewer



IGV input file formats

- BAM-file - coordinate sorted - indexed
- BED-files
- VCF-files
- Many others

<http://software.broadinstitute.org/software/igv/>



1. Inspeccionar alineamientos y cobertura de archivos BAM
2. SNVs
3. InDels

Por defecto, IGV tiene cargado Human (hg19).

Puedes cargar pistas adicionales (genes Ensembl, dbSNP...) desde Archivo ->

Cargar desde el servidor

Navega por cromosomas ó genes. En este caso vamos al gen Ca2

- Por defecto, IGV tiene cargado Human (hg19).
- Puedes cargar pistas adicionales (genes Ensembl, dbSNP...) desde Archivo ->  
Cargar desde el servidor
- Navega por cromosomas ó genes. En este caso vamos al gen Ca2