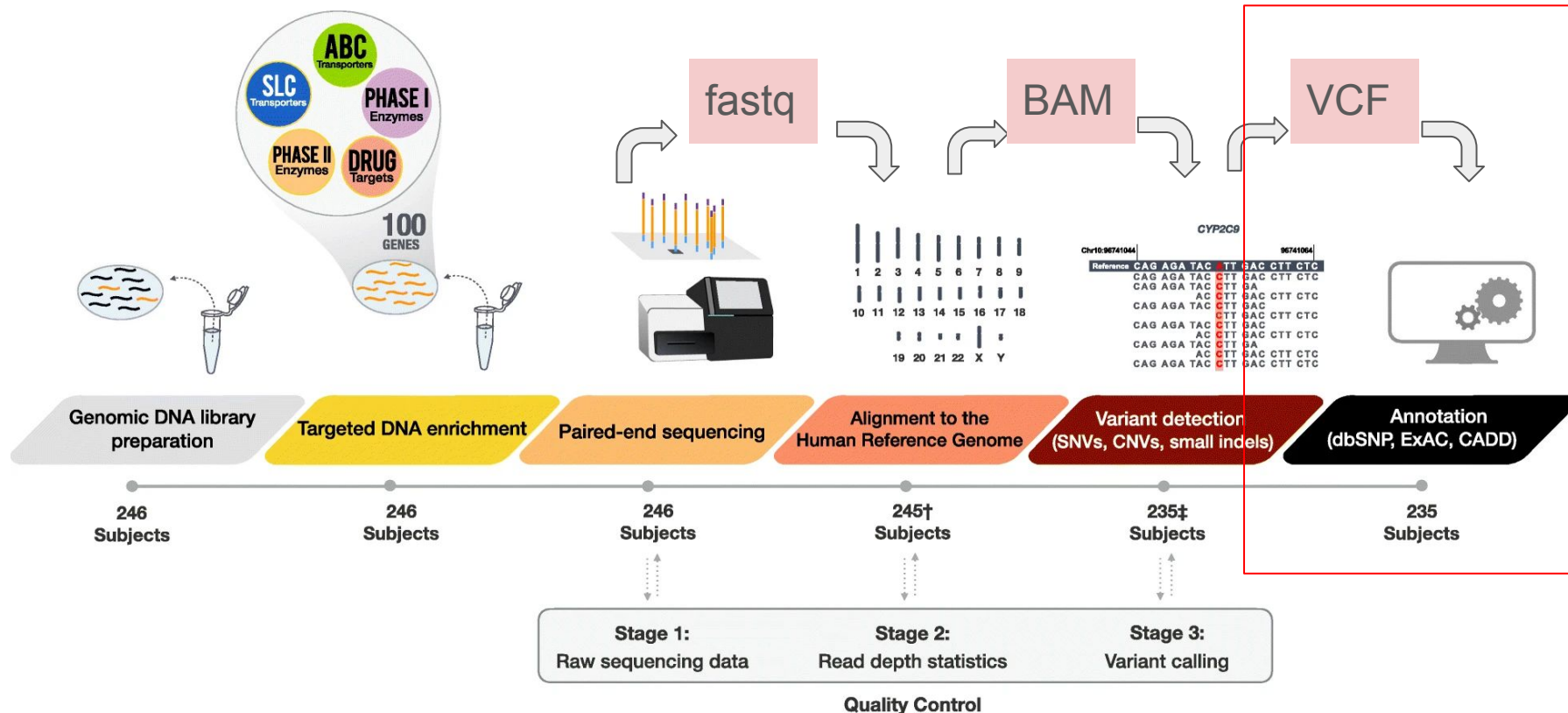




Llamado y anotación funcional de variantes

Evelin González F.
evefeliu@gmail.com

Flujo de trabajo “Targeting Sequencing”



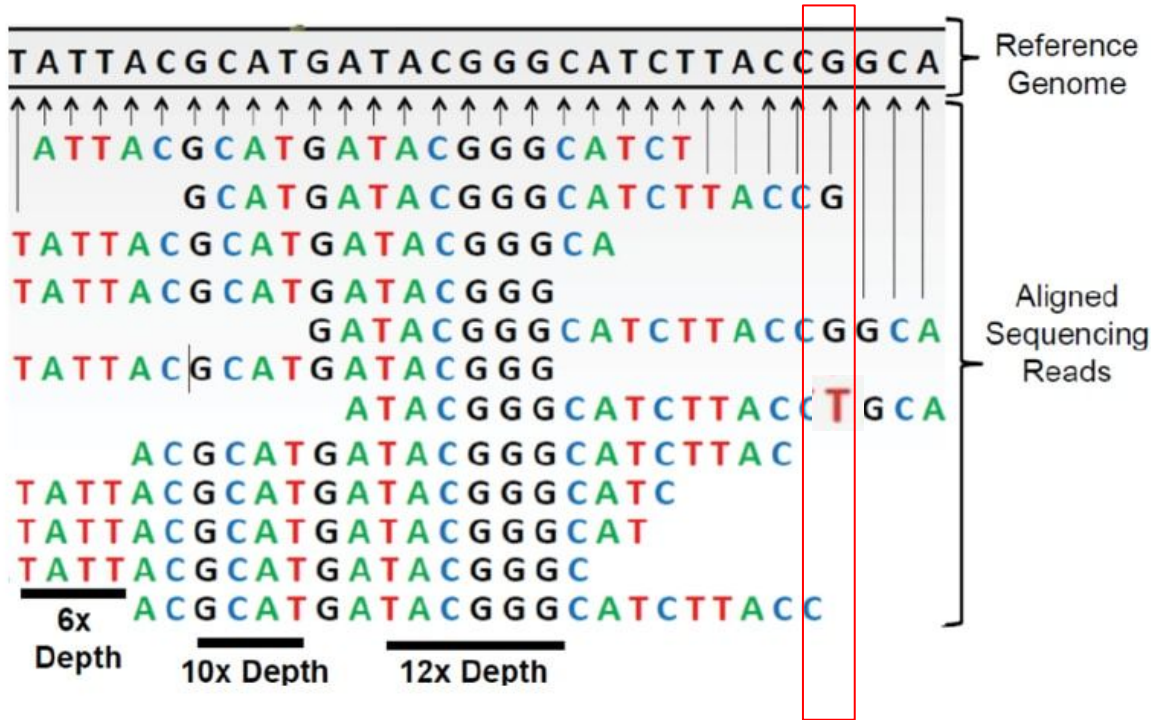
Archivos de entrada y salida

File Type	Full Name	Description	Approximate File Size (Average Coverage 160×)	
			Exome	4800 Genes
FASTQ	Files with consensus assessment of sequence and variation	Raw sequencing data after demultiplexing	50 GB	18 GB
BAM	Binary version of sequence alignment/map	Sequencing data after alignment	16 GB	6 GB
VCF	Variant call file	File containing variants called relative to the reference	9.3 GB	3.5 MB

Abbreviations: GB, gigabytes; MB, megabytes.



Archivos *.bam



BAM: Formato binario para el almacenamiento de datos de secuenciación.

La extensión de archivo (.bam) contiene información sobre lecturas de secuencias después de haber sido estas **alineadas contra un genoma de referencia**.

Utilizado para almacenar datos de secuenciación masiva.

Y si las lecturas presentan cambios con respecto a la referencia ??

Concepto de variante genética (SNPs, INDELs)

Substitution

original

C T G G A G



mutated

C T G G G G

Insertion

original

C T G G A G



mutated

C T G G T G G A G

Deletion

original

C T ~~G~~ G A G



mutated

C T A G

- **Sustitución (SNP):** Cambio de una base por otra en una posición específica (ej. A \rightarrow T).
- **Inserción:** Adición de una o más bases en la secuencia de ADN.
- **Delección:** Pérdida de una o más bases en la secuencia de ADN.

Variant Call Format: VCF

A VCF example

Header

```
##fileformat=VCFv4.1
##fileDate=20110413
##source=VCFtools
##reference=file:///refs/human_NCBI36.fasta
##contig=<ID=1,length=249250621,md5=1b22b98cdeb4a9304cb5d48026a85128,species="Homo Sapiens">
##contig=<ID=X,length=155270560,md5=7e0e2e580297b7764e31dbc80c2540dd,species="Homo Sapiens">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##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">
##ALT=<ID=DEL,Description="Deletion">
##INFO=<ID=SVTYPE,Number=1,Type=String,Description="Type of structural variant">
##INFO=<ID=END,Number=1,Type=Integer,Description="End position of the variant">
```

Body

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE1	SAMPLE2
1	1	.	ACG	A,AT	40	PASS	.	GT:DP	1/1:13	2/2:29
1	2	.	C	T,CT	.	PASS	H2;AA=T	GT	0 1	2/2
1	5	rs12	A	G	67	PASS	.	GT:DP	1 0:16	2/2:20
X	100	.	T		.	PASS	SVTYPE=DEL;END=299	GT:GQ:DP	1:12:.	0/0:20:36

B SNP

Alignment	VCF representation
1234	POS REF ALT
ACGT	2 C T
ATGT	
^	

C Insertion

12345	POS REF ALT
AC-GT	2 C CT
ACTGT	
^	

D Deletion

1234	POS REF ALT
ACGT	1 ACG A
A--T	
^^	

E Replacement

1234	POS REF ALT
ACGT	1 ACG AT
A-TT	
^^	

VCF

```
##fileformat=VCFv4.2
##contig=<ID=2,length=51304566>
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT SAMPLE1 SAMPLE2 SAMPLE3 SAMPLE4 SAMPLE5 SAMPLE6 SAMPLE7
2 81170 . C T . . AC=9;AN=7424 GT:DP:GQ 0/0:4:12 0/0:3:9 0/1:1:3 0/1:9:24 1/0:4:12 0/0:5:15 0/0:4:12
2 81171 . G A . . AC=6;AN=7446 GT:DP:GQ 0/1:4:12 0/0:3:9 0/0:1:3 0/0:9:24 0/1:4:12 0/1:5:15 0/0:4:12
2 81182 . A G . . AC=5;AN=7506 GT:DP:GQ 0/0:5:15 0/0:4:12 0/0:5:15 0/0:9:24 0/0:4:12 0/0:4:12 0/0:4:12
2 81204 . T G . . AC=2;AN=7542 GT:DP:GQ 1/0:5:15 0/0:9:27 0/0:10:30 0/0:15:39 0/0:9:27 1/0:13:39 0/1:14:42
```

- **CHROM**: Cromosoma de la variante.
- **POS**: Posición de la variante.
- **REF**: Base de referencia.
- **ALT**: Base alternativa.
- **QUAL**: Calidad de la llamada.
- **INFO**: Información adicional (por ejemplo, profundidad de cobertura y frecuencia alélica).

¿Qué es la Anotación de Variantes?

La anotación de variantes es el proceso de agregar información biológica, clínica y funcional a variantes genéticas identificadas mediante secuenciación.

Tipos de Información Anotada:

- **Frecuencia alélica** : Bases de datos poblacionales como gnomAD.
- **Impacto funcional**: Predicciones de efectos en proteínas: REVEL, SIFT, PolyPhen.
- **Relevancia clínica**: Clasificación patogénica: ClinVar, OncoKB.
- **Información genómica**: Posición en genes, exones/intrones: ANNOVAR.

Clasificación de variantes

AATCGTGTACGTAC
GTAGTTCATCTTCA
GCTGCATTCGGATT



Anotación



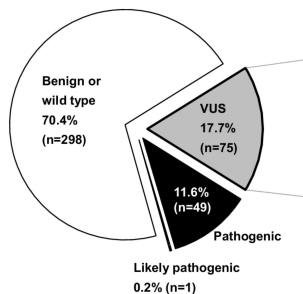
Clasificación



- Patogénica
- Probablemente patogénica
- Variante de significado incierto (VUS)
- Probablemente benigna
- Benigna

● Patogénica

- Evidencia de que **afecta la función de la proteína/gen**
- Reportada en bases clínicas (ClinVar, OncoKB, literatura)
- Asociada a enfermedad
- No siempre explica por sí sola el fenotipo



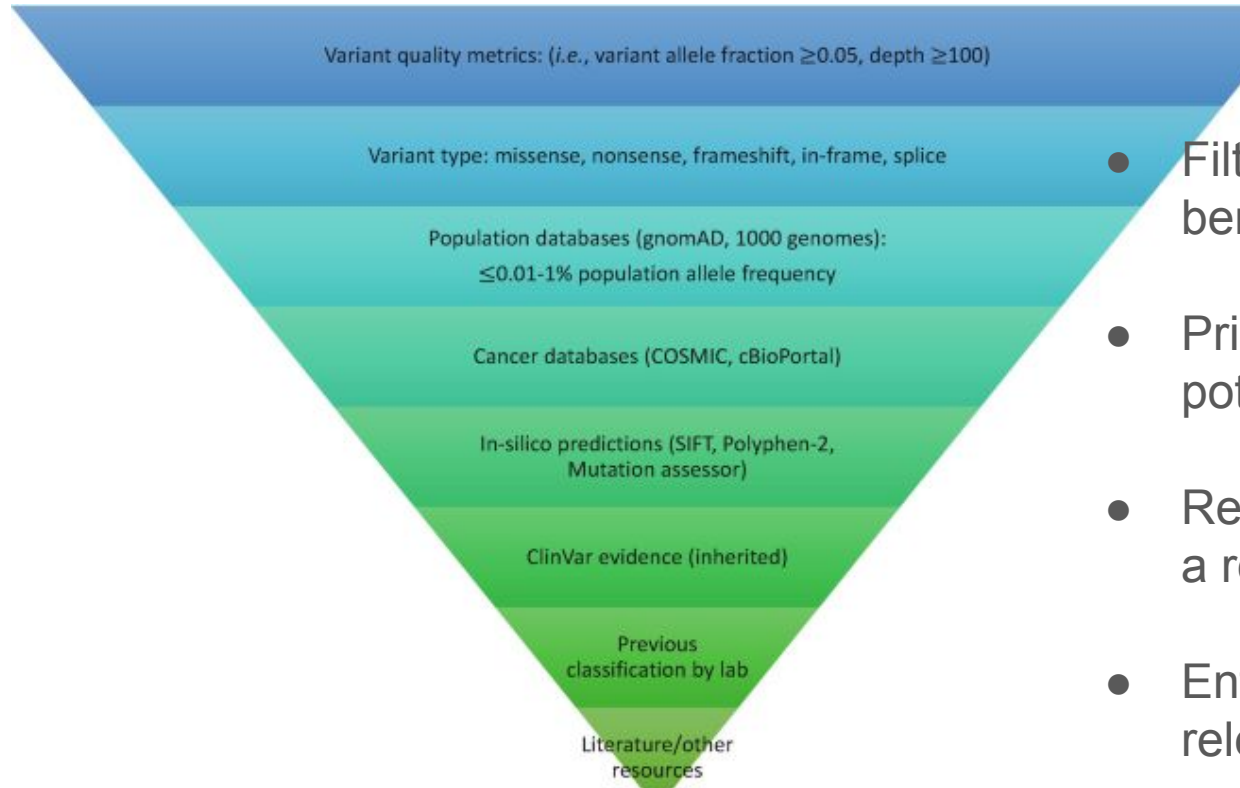
● VUS

- Impacto funcional incierto
- Evidencia insuficiente o contradictoria
- No se usa para decisiones clínicas

● Benigna

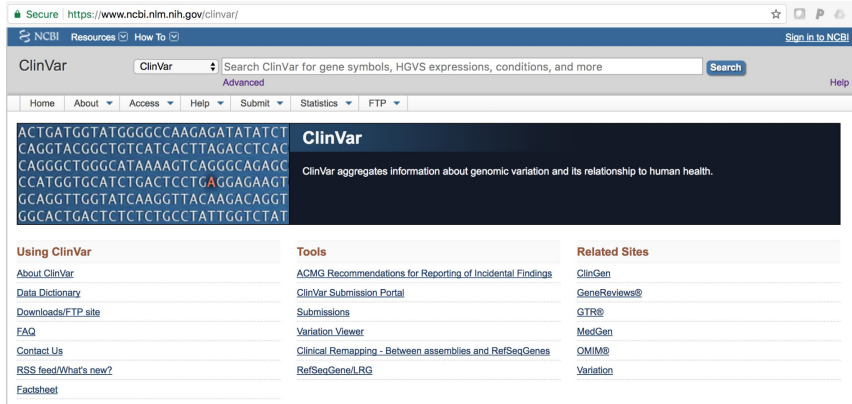
- No altera la función proteica
- Frecuente en población general
- Sin asociación con enfermedad

Priorización de variantes



- Filtra variantes comunes y benignas.
- Prioriza variantes potencialmente patogénicas.
- Reduce el número de variantes a revisar.
- Enfoca el análisis en variantes relevantes para el cáncer

ClinVar



How to search ClinVar

<https://www.youtube.com/watch?v=A8G3ej83ZgU>

Base de datos pública que proporciona información sobre la relación entre las **variaciones genéticas** y la **salud humana**, específicamente en el contexto de diversas enfermedades, **incluido el cáncer**.

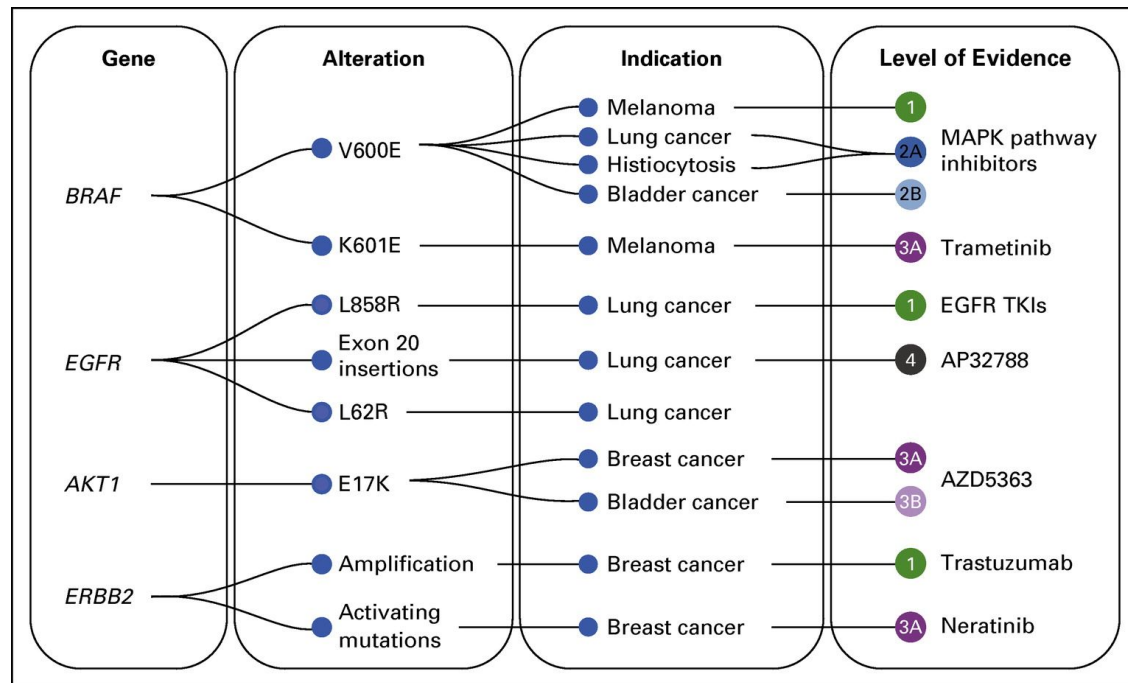
ClinVar clasifica las variantes en categorías como **"patogénica"**, **"probablemente patogénica"**, **"de significado incierto"**, **"probablemente benigna"** y **"benigna"**, lo que ayuda a los clínicos y a los investigadores a interpretar los hallazgos genéticos en relación con el cáncer.

Ejemplo: Variantes Patogénicas y VUS en Pacientes con Cáncer de Mama

Tumor_Sample_Barcode	Hugo_Symbol	Variant Classification	aaChange	CLNSIG	REVEL
77	BRCA1	Frame_Shift_Ins	p.L655Ffs*10	Pathogenic	.
62	BRCA2	Missense_Mutation	p.T703A	Uncertain_significance	0.085
47	BRCA2	Missense_Mutation	p.K797Q	Uncertain_significance	0.281
61	BRCA2	Missense_Mutation	p.N991S	.	0.029
78	BRCA2	Missense_Mutation	p.D1737V	Uncertain_significance	0.244
59	BRCA2	Missense_Mutation	p.R2502C	Conflicting_interpretations_of_pathogenicity	0.519
75	BRCA2	Missense_Mutation	p.I2718T	Conflicting_interpretations_of_pathogenicity	0.357
59	BRCA2	Missense_Mutation	p.I2718T	Conflicting_interpretations_of_pathogenicity	0.357
87	BRCA2	Missense_Mutation	p.I2718T	Conflicting_interpretations_of_pathogenicity	0.357

En este ejemplo de pacientes con cáncer de mama se observan variantes patogénicas y variantes de significado incierto (VUS) reportadas en ClinVar. La variante en **BRCA1** corresponde a una inserción con cambio de marco de lectura (*frameshift*), clasificada como **patogénica**, lo que sugiere un impacto funcional claro sobre la proteína. En contraste, varias variantes *missense* en **BRCA2** se clasifican como **VUS**, ya que no existe suficiente evidencia clínica o funcional para determinar su efecto. El score **REVEL**, un predictor in silico del impacto funcional de variantes missense, puede ayudar a priorizar estas variantes: por ejemplo, una variante con **REVEL > 0.5** sugiere mayor probabilidad de efecto dañino, aunque esto por sí solo no es suficiente para reclasificarla como patogénica. Este tipo de resultados ilustra cómo la interpretación de variantes integra evidencia clínica, poblacional y computacional para su clasificación final.

OncoKB: Relevancia Clínica de Variantes Oncológicas

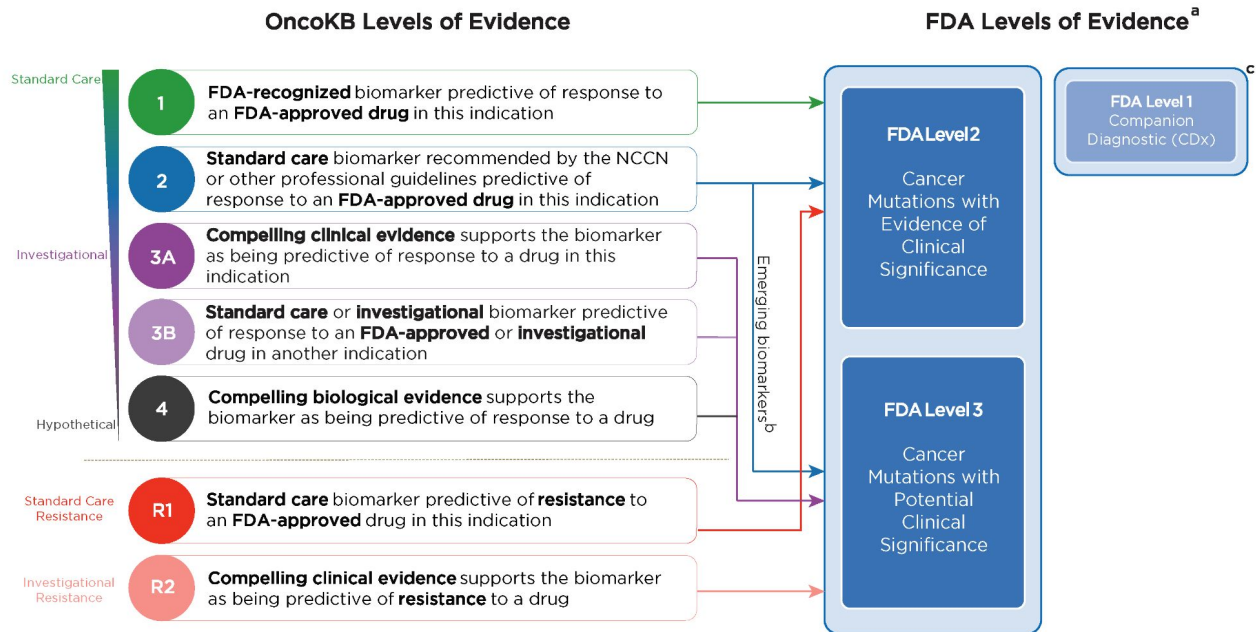


¿Qué es OncoKB?

OncoKB es una base de datos de conocimiento clínico que proporciona anotaciones precisas sobre variantes genéticas en cáncer y su relevancia clínica, incluyendo **información sobre terapias dirigidas y ensayos clínicos.**

OncoKB ayuda a determinar si una variante genética es relevante para el tratamiento, mejorando la toma de decisiones clínicas y personalizadas para los pacientes con cáncer.

Mapecto entre los Niveles de Evidencia de OncoKB™ y los Niveles de Evidencia de la FDA



Footnotes

^a The FDA levels of evidence presented here are modified from the FDA white paper titled "CDRH's approach to tumor profiling next-generation sequencing tests".

^b **Emerging biomarkers** are defined as those alterations listed as a NCCN guideline category 2A biomarker based on limited clinical data, for example early Phase I and Phase II clinical studies with limited patient data/responses. They qualify as OncoKB Level 2, but map to FDA Level 3.

^c Since OncoKB does not include any companion diagnostic claims prescriptive for a specific therapeutic product, by definition, no variants in OncoKB are considered FDA Level 1.

FDA-Approved Oncology Therapies

Content current as of 10/21/2024

The following US Food and Drug Administration (FDA)-approved oncology drugs post June 1998 are categorized by drug class and mechanism of action. Each drug is further classified as to whether it qualifies as a targeted therapy or precision oncology therapy based on [Suehnholz et al., Cancer Discovery 2023](#) (definitions below).

[Show definitions](#)

FDA-Approved
Precision Oncology Therapies
103 therapies

FDA-Approved
Targeted Therapies
193 therapies

FDA-Approved
Oncology Therapies
228 therapies

Select drug(s)
Select class of agent(s)
Select mechanism of action
Search biomarker

Showing 103 therapies: (102 Targeted therapies, 103 Precision oncology therapies, 83 therapies with a biomarker that can be identified by a DNA/NGS-based detection method) [Download Table](#)

Year of drug's first FDA-approval	FDA-approved drug(s)	FDA drug label listed biomarker(s)	Class of agent(s)	Mechanism of action or drug target	Targeted therapy	Precision oncology therapy	Can a DNA/NGS-based method be used for biomarker detection?
2024	Zolbetuximab	HER2- and CLDN18.2+	Monoclonal antibody	CDLN18.2-directed cytolytic antibody	✓	✓	N
2024	Adagrasib + Cetuximab	KRAS G12C	Small molecule inhibitor and monoclonal antibody combination	KRAS G12C inhibitor + Anti-EGFR antibody	✓	✓	Y
2024	Inavolisib + Fulvestrant + Palbociclib	PIK3CA Oncogenic Mutations, HR+/HER2-	Small molecule inhibitor and hormone therapy combination	α-specific PI3K inhibitor + selective estrogen receptor degrader + CDK4/6 inhibitor	✓	✓	Y
2024	Lazertinib + Amivantamab	EGFR L858R, Exon 19 Deletions	Small molecule kinase inhibitor and monoclonal antibody combination	EGFR tyrosine kinase inhibitor + EGFR-MET bispecific antibody	✓	✓	Y
2024	Vorasidenib	IDH1 R132H/C/G/S/L, IDH2 R172K/M/V/S/G	Small molecule inhibitor	IDH1/IDH2 inhibitor	✓	✓	Y
2024	Afamitresgene autoleucl	HLA-A*02:01P, -A*02:02P, -A*02:03P, or -A*02:06P positive and MAGE-A4 antigen positive	T cell receptor (TCR) therapy	MAGE-A4-directed TCR therapy	✓	✓	N

OncoKB: FDA-Approved Oncology Therapies

<https://www.oncokb.org/oncology-therapies>

Búsqueda de variantes accionables

Welcome to OncoKB™

MSK's Precision Oncology Knowledge Base

An FDA-Recognized Human Genetic Variant Database*

Ingresamos el Gen seguido
del cambio en la proteína



894

Genes

7893

Alterations

141

Cancer Types


147

Drugs

BRCA1 L655Ffs*10



BRCA1/L655Ffs*10 

Not annotated by OncoKB™ 

The BRCA1 L655Ffs*10 is a truncating mutation in a tumor suppressor gene, and therefore is likely oncogenic. Click here to see more annotation details.

Is

1 Level 1

FDA-approved drugs

53 Genes

2 Level 2

Standard care

29 Genes

3 Level 3

Clinical evidence

36 Genes

4 Level 4

Biological evidence

29 Genes

5 Level R1/R2

Resistance

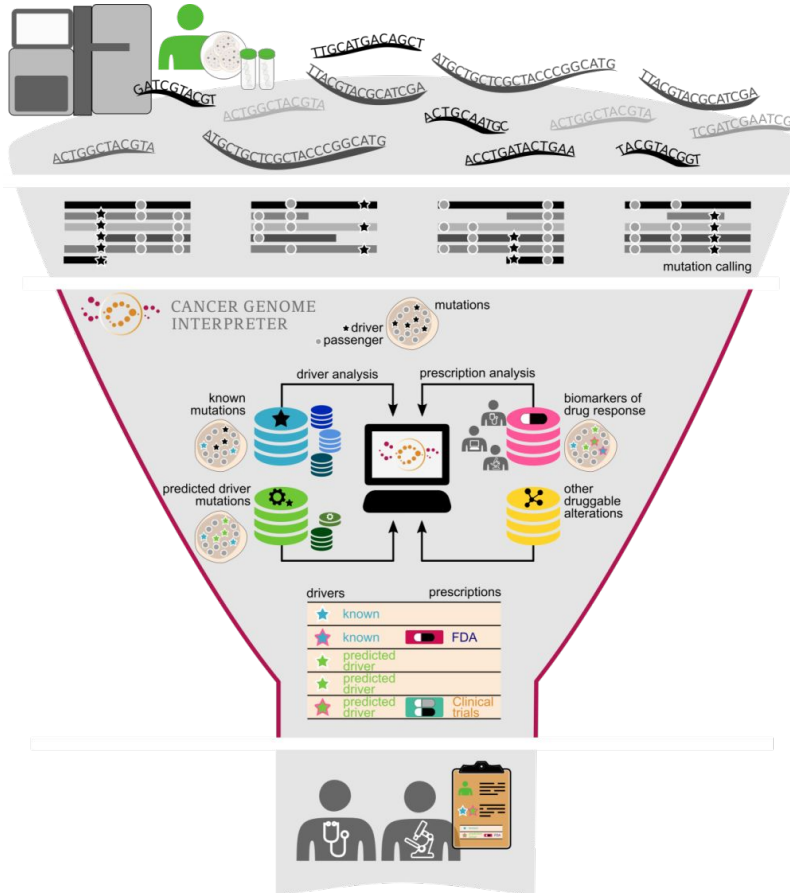
12 Genes

Powered by the clinical expertise of Memorial Sloan Kettering Cancer Center

When using OncoKB™, please cite: [Suehnholz et al., Cancer Discovery 2023](#) and [Chakravarty et al., JCO PO 2017](#).

*FDA recognition of OncoKB™ is for the content that is clearly marked

Cancer Genome Interpreter (CGI)



¿Qué es CGI?

Cancer Genome Interpreter (CGI) está diseñado para apoyar la identificación de **alteraciones tumorales** que **impulsan** la enfermedad y/o que pueden ser **accionables terapéuticamente**. CGI se basa en métodos computacionales, como la mutagénesis de saturación in silico de genes cancerígenos (BoostDM y OncodriveMut), así como en el conocimiento recopilado del dominio público para anotar las alteraciones en un tumor según varios niveles de evidencia.

- CGI puede identificar variantes germinales que afectan genes relacionados con el cáncer y proporcionar información sobre su impacto clínico.
- Incluye variantes en genes como **BRCA1**, **BRCA2**, **TP53**, y otros genes asociados con el riesgo hereditario de cáncer.

- > point mutations and indels
- > CNAs
- > translocations
- > cancer type



tumor(s)
input data



Preprocessing



format recognition
remapping
standardisation



Identification of potentially oncogenic alterations



Compendium of Cancer Genes

Compilation of evidences of cancer genes in different cancer types



Validated Oncogenic Mutations

Annotations of cancer mutations from our own database and others (OncoKB, ClinVar)



BoostDM

Machine learning approach to identify driver mutations



OncodriveMUT

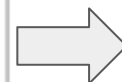
Rule based approach to identify driver mutations

Predicted oncogenic alterations
and annotations from databases

driver

passenger

Annotations



ClinVar
OncoKB
CGI



Identification of genomic biomarkers of drug response



Cancer Biomarker Databases

Annotation of cancer biomarkers from our own database and others (OncoKB, CIVIC)
Annotated according to different levels of evidence



Annotated effect and level of evidence of
genomic biomarkers of drug response


Responsive

Resistant

- ⬆ A Clinical guidelines
- ⬆ B Clinical evidence
- ⬆ C Early trials or case reports
- ⬆ D Pre-clinical evidences

Plataforma Web: ¿Cómo Utilizar Cancer Genome Interpreter (CGI)?

This resource is intended for purely research purposes. It should not be used for medical or professional advice. [Understand](#)



CANCER GENOME
INTERPRETER

HOMEANALYSISABOUTFAQCONTACT

evfeliu@gmail.com

March 2024: ClinVar and CIVIC databases have been updated to include recent data releases

Alterations

Insert one mutation per line, use the same format for all the mutations.

BRAF:V600E
NPL_BMS157:p.T315T
chr3:g.178936891G>A
chr17:7578412 A G
ENB82:AMP
TP53:DEL
PHL1_BARA

(protein change format, gene symbol)
(protein change format, RefSeq/Ensembl transcript)
(nucleotide change format, HGVS)
(nucleotide change format, GTF)
(copy number alteration, gene amplification)
(copy number alteration, gene deletion)
(translocation)

View CGI example results

Add file +

Cargamos archivo VCF

Cancer type

Search

(CANCER) Any cancer type

- (HEMATO) Hematologic malignancies
- (SOLID) Solid tumors
 - (TX) Torax
 - (BRCA) Breast adenocarcinoma
 - (L) Lung
 - (MESO) Mesothelioma
 - (PL) Pleura
 - (US) Urinary system

Seleccionamos tipo de cáncer ->BRCA

Reference genome

- hg38
- hg19

Genoma de referencia hg38

Click en Run

Run

Warning! Max file size is 32MB. If your file exceeds this limit, try compressing it into a ".gz" file.

CGI para la predicción de mutaciones conductoras



HOME ANALYSIS ABOUT FAQ CONTACT

evefeliu@gmail.com

This analysis will be removed after 6 months

ANALYSIS (3145)

DETAILS

Status **Done**
Cancer Type: **BRCA**
Date: **2024-11-11 00:36:22**

Reference genome: **hg38**
Number of samples: **1**
Total mutations: **50**
Driver mutations: **2**

Samples without drivers: **0**
Biomarkers of the tumor: **13**
Biomarkers in other tumors: **68**

Download results Remove this job and all the data Share the job

Details

License: CC BY-NC

ALTERATIONS

PRESCRIPTIONS

Mutations

Show entries with: ☒ Mutations identified as drivers ☒ Mutations with oncogenic annotations ☒ Other mutations

Sample ID	Gene	Protein Change	Oncogenicity	Mutation	Consequence	Oncogenic annotation	Transcript
Search here...							
input01	BRCA2	A2951T	driver	chr13:32379413 G>A	missense variant		ENST00000380152
input01	BRCA1	L702FX	driver	chr17:43093425-43093426 ->A	frameshift variant		ENST00000471181
input01	BRCA2	N289H	passenger	chr13:32332343 A>C	missense variant		ENST00000380152
input01	BRCA2	N372H	passenger	chr13:32332592 A>C	missense variant		ENST00000380152
input01	BRCA2	K797Q	passenger	chr13:32336744 A>C	missense variant		ENST00000380152
input01	BRCA2	N991D	passenger	chr13:32337326 A>G	missense variant		ENST00000380152
input01	BRCA2	T1915M	passenger	chr13:32340099 C>T	missense variant		ENST00000380152
input01	BRCA2	R2108H	passenger	chr13:32340678 G>A	missense variant		ENST00000380152
input01	BRCA2	V2466A	passenger	chr13:32355250 T>C	missense variant		ENST00000380152
input01	BRCA2	I2490T	passenger	chr13:32356461 T>C	missense variant		ENST00000380152
input01	BRCA1	S1634G	passenger	chr17:43071077 T>C	missense variant		ENST00000471181