

# Online tools for the study of mutated genes in cancer

Cecilia Mathó, PhD

[mathocecilia@gmail.com](mailto:mathocecilia@gmail.com)

Cancer Genome Analysis Course

November 28<sup>th</sup>, 2023

Montevideo, Uruguay



UNIVERSIDAD  
DE LA REPÚBLICA  
URUGUAY



# .vcf Annotation using ANNOVAR

<https://wannovar.wglab.org/>

The screenshot shows the wANNOVAR homepage with a red header. The header includes the wANNOVAR logo, navigation links for Home, Tutorial, Example, and Related projects, and the WGLAB logo. Below the header, the title "wANNOVAR" is prominently displayed in large white letters. A descriptive text block explains that ANNOVAR is a rapid, efficient tool for annotating functional consequences of genetic variation. It also mentions that wANNOVAR provides easy and intuitive web-based access to the most popular functionalities of the ANNOVAR software. At the bottom of the header, there are three blue buttons labeled "Get Started", "About", and "Contact". Below the header, there are social sharing buttons for "Like" and "Share", followed by a message indicating 10 people like this.

## Basic Information

Email

Sample Identifier

Input File

or Paste Variant Calls

I agree to the [Terms of Use](#). Please note that commercial users would need to obtain a license.

We need an institutional email  
e upload a .vcf and get a list of variants we can load into  
Excel or other program

## Submission 421863

Your submission has been received by the ANNOVAR server at Tue Nov 28 07:30:05 2023.

The results will be generated at </done/421863/QwqaPeb5JnAHZgCP/index.html> after the computation is done.

**WARNING WARNING WARNING:** Many email servers nowadays block suspicious emails with URL inside, so you may NOT receive an email, and all results will be deleted every 24 hours. Therefore, it is best that you record the URL for results (such as bookmarking it), or keep the browser open until results are shown.

### A summary of input file

User input contains 150 lines

All Rights Reserved @Wang Genomics Lab 2010-2019

## Submission ID: 421863

Sample identifier = TEST  
File\_name=TSVC\_variants\_IonXpress\_009.vcf  
File\_format=vcf4  
Reference\_genome=hg19  
Disease\_model=no filtering  
Processed variants=79

### Basic Information

<b>exome summary results</b>	<a href="#">view</a>	<a href="#">CSV file</a>	<a href="#">TXT file</a>
<b>genome summary results</b>	<a href="#">view</a>	<a href="#">CSV file</a>	<a href="#">TXT file</a>

All Rights Reserved @Wang Genomics Lab 2010-2019

Chr	Start	End	Ref	Alt	Func.refGene	Gene.refGene	GeneDetail.refGene	ExonicFunc.refGene	AAChange.refGene
chr1	45797505	45797505	C	G	exonic	MUTYH		nonsynonymous SNV	MUTYH:NM_001350650:exon11:c.G585C;p.Q195H,MUTYH:NM_001350651:e
chr2	47601106	47601106	T	C	exonic	EPCAM		nonsynonymous SNV	EPCAM:NM_002354:exon3:c.T344C:p.M115T
chr3	37053568	37053568	A	G	exonic	MLH1		nonsynonymous SNV	MLH1:NM_000249:exon8:c.A655G;p.I219V,MLH1:NM_001167617:exon8:c.A3
chr5	112162854	112162854	T	C	exonic	APC		synonymous SNV	APC:NM_001127511:exon10:c.T1404C:p.Y468Y,APC:NM_000038:exon12:c.T1
chr5	112164565	112164561	G	A	exonic	APC		synonymous SNV	APC:NM_001127511:exon12:c.G1581A:p.A527A,APC:NM_000038:exon14:c.G
chr5	112175770	112175770	G	A	exonic	APC		synonymous SNV	APC:NM_001127511:exon14:c.G4425A:p.T1475T,APC:NM_000038:exon16:c.C
chr5	112176325	112176325	G	A	exonic	APC		synonymous SNV	APC:NM_001127511:exon14:c.G4980A:p.G1660G,APC:NM_000038:exon16:c.C
chr5	112176559	112176559	T	G	exonic	APC		synonymous SNV	APC:NM_001127511:exon14:c.T5214G:p.S1738S,APC:NM_000038:exon16:c.T
chr5	112176756	112176756	T	A	exonic	APC		nonsynonymous SNV	APC:NM_001127511:exon14:c.T5411A:p.V1804D,APC:NM_000038:exon16:c.T
chr5	112177171	112177171	G	A	exonic	APC		synonymous SNV	APC:NM_001127511:exon14:c.G5826A:p.P1942P,APC:NM_000038:exon16:c.C
chr7	6013049	6013049	C	G	exonic	PMS2		nonsynonymous SNV	PMS2:NM_001322008:exon13:c.G2252C:p.G751A,PMS2:NM_001322010:exon9
chr7	6026775	6026775	T	C	exonic	PMS2		nonsynonymous SNV	PMS2:NM_001322008:exon9:c.A1303G:p.K435E,PMS2:NM_001322010:exon9
chr7	6026988	6026988	G	A	exonic	PMS2		nonsynonymous SNV	PMS2:NM_001322008:exon9:c.C1090T:p.P364S,PMS2:NM_001322010:exon9
chr7	6036980	6036980	G	C	exonic	PMS2		synonymous SNV	PMS2:NM_001322008:exon5:c.C462G:p.S154S,PMS2:NM_001322004:exon6:c.C
chr15	31197976	31197976	C	T	exonic	FAN1		synonymous SNV	FAN1:NM_001146094:exon2:c.C1110T:p.T370T,FAN1:NM_001146095:exon2:c.C
chr15	31202961	31202961	G	A	exonic	FAN1		nonsynonymous SNV	FAN1:NM_001146094:exon4:c.G1520A:p.R507H,FAN1:NM_001146095:exon4

# .vcf Annotation using VEP

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

Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

Login/Register

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

In this section

- VEP web interface
  - Input form
  - Results
- VEP command line
  - Tutorial
  - Download and install
  - Running VEP
  - Annotation sources
  - Filtering results
  - Custom annotations
  - Plugins
  - Examples and use cases
  - Other information
- Data formats
- Variant Recoder
- Haplosaurus
- VEP FAQ

On this page

- VEP interfaces
- Publication
- VEP related tools

Help & Documentation API & Software Ensembl Tools Ensembl Variant Effect Predictor (VEP)

## Ensembl Variant Effect Predictor (VEP)

Ve!P

**VEP determines the effect of your variants** (SNPs, insertions, deletions, CNVs or structural variants) on genes, transcripts, and protein sequence, as well as regulatory regions.

Simply input the coordinates of your variants and the nucleotide changes to find out the:

- Genes and Transcripts affected by the variants
- Location of the variants (e.g. upstream of a transcript, in coding sequence, in non-coding RNA, in regulatory regions)
- Consequence of your variants on the protein sequence (e.g. stop gained, missense, stop lost, frameshift), see [variant consequences](#)
- Known variants that match yours, and associated minor allele frequencies from the [1000 Genomes Project](#)
- SIFT and PolyPhen-2 scores for changes to protein sequence
- ... And more! See [data types, versions](#).

★ [What's new in release 110?](#)

VEP interfaces

VEP ▾

Web Tools

- BLAST/BLAT
- Variant Effect Predictor
- Linkage Disequilibrium Calculator
- Variant Recoder
- File Chameleon
- Assembly Converter
- ID History Converter
- VCF to PED Converter
- Allele Frequency Calculator
- Data Slicer
- Variation Pattern Finder

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

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

## Variant Effect Predictor ?

New job

Recent jobs ▾

Refresh

Show/hide columns (1 hidden)

Filter

Analysis

Jobs

Submitted at (GMT)

Variant Effect Predictor



VEP analysis of prueba in Homo\_sapiens

Done

[View results]

16/11/2023, 23:57



Ensembl GRCh37 release 110 - July 2023 © EMBL-EBI

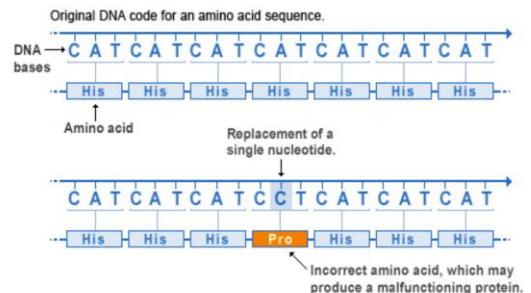
This is how the output looks like:

#Uploaded_variation	Location	Allele	Conseq	IMPACT	SYMBOL	Gene	Feature	Feature	BIOTYPE
376108;376109;COSM303886;COSM5879387;rs1057519766	13:28602340-28602340	T	missense_MODERATE	FLT3		2322	Transcript	NM_0041	protein_coding
375879;375880;375881;COSM1583129;COSM3356083;COSM4422:2:25457242-25457242		T	missense_MODERATE	DNMT3A		1788	Transcript	NM_0013	protein_coding
375879;375880;375881;COSM1583129;COSM3356083;COSM4422:2:25457242-25457242		T	missense_MODERATE	DNMT3A		1788	Transcript	NM_0013	protein_coding
375879;375880;375881;COSM1583129;COSM3356083;COSM4422:2:25457242-25457242		T	missense_MODERATE	DNMT3A		1788	Transcript	NM_0225	protein_coding
375879;375880;375881;COSM1583129;COSM3356083;COSM4422:2:25457242-25457242		T	missense_MODERATE	DNMT3A		1788	Transcript	NM_1537	protein_coding
375879;375880;375881;COSM1583129;COSM3356083;COSM4422:2:25457242-25457242		T	missense_MODERATE	DNMT3A		1788	Transcript	NM_1756	protein_coding
COSM4383936;COSM4383937;COSM5794149	4:106190830-106190830	A	missense_MODERATE	TET2		54790	Transcript	NM_0011	protein_coding
13998;COSM1319222;COSM158604;COSM255167;COSM4170215:170837543-170837543		TCTG	frameshift	HIGH	NPM1	4869	Transcript	NM_0013	protein_coding
13998;COSM1319222;COSM158604;COSM255167;COSM4170215:170837543-170837543		TCTG	frameshift	HIGH	NPM1	4869	Transcript	NM_0013	protein_coding
13998;COSM1319222;COSM158604;COSM255167;COSM4170215:170837543-170837543		TCTG	frameshift	HIGH	NPM1	4869	Transcript	NM_0013	protein_coding
13998;COSM1319222;COSM158604;COSM255167;COSM4170215:170837543-170837543		TCTG	frameshift	HIGH	NPM1	4869	Transcript	NM_0025	protein_coding
13998;COSM1319222;COSM158604;COSM255167;COSM4170215:170837543-170837543		TCTG	frameshift	HIGH	NPM1	4869	Transcript	NM_1991	protein_coding

# Types of mutations overview

- **Missense/ con sentido/con cambio de sentido:**  
→ aminoacid change

Missense mutation



- **Nonsense/sin sentido:** premature stop codon

No mutation

Nonsense

DNA level	TTC
mRNA level	AAG
protein level	Lys

ATC
UAG
STOP

- **Frameshift/ cambio en el marco de lectura:** if bases are inserted or removed this will cause a reading frame shift, eventually causing a premature stop codon.

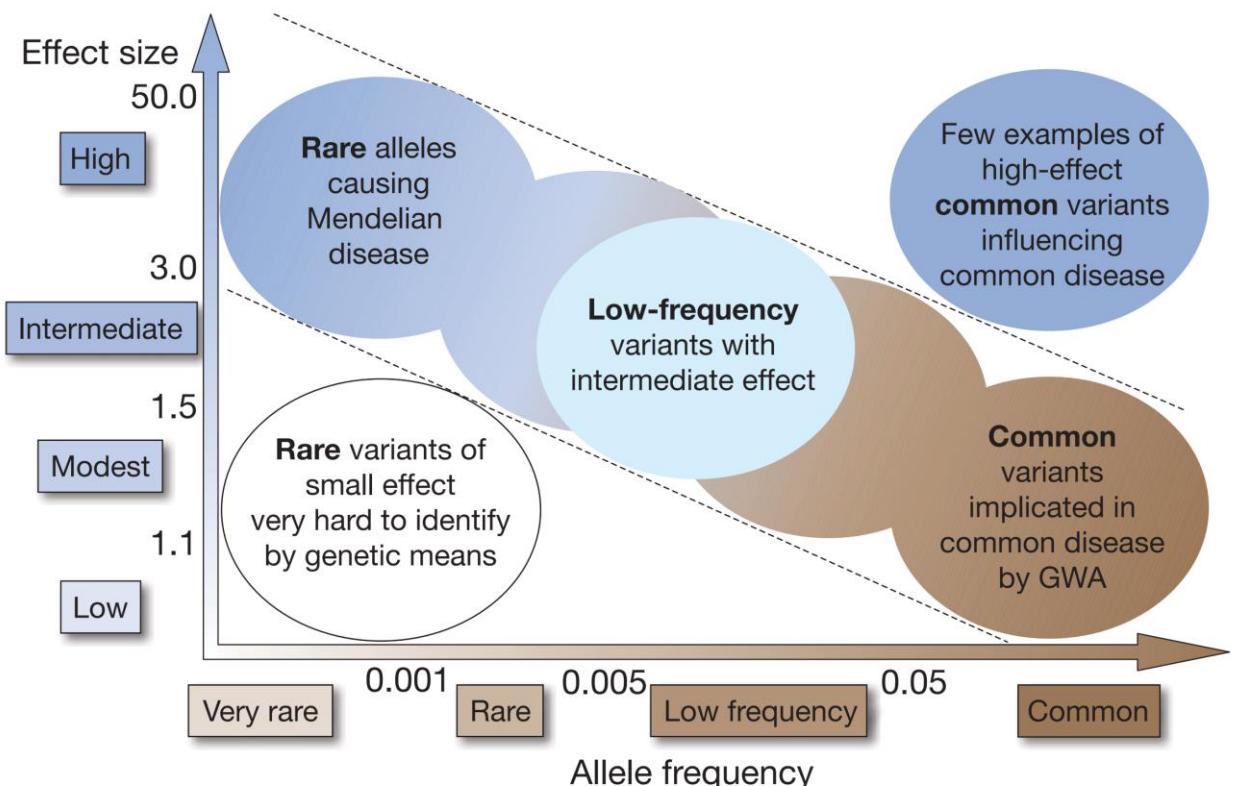
-What happens if the insertion/deletion is a multiple of 3?

- **Silent/ Synonymous/Sinónima/Silenciosa:** there is a change in the DNA but the resulting codon is for the same aminoacid as the original.

No mutation      Silent

DNA level	TTC	TTT
mRNA level	AAG	AAA
protein level	Lys	Lys

# Allelic frequencies



**Figure 1 | Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio).** Most emphasis and interest lies in identifying associations with characteristics shown within diagonal dotted lines. Adapted from ref. 42.

Vol 461|8 October 2009|doi:10.1038/nature08494

The aim of this tutorial is to become familiar with the free computer tools available to obtain information about genes mutated in cancer.

You'll learn to:

- 1-Search for information about a gene and related pathology.
- 2- Find information about a gene and the protein it encodes.
- 3-Determine if a gene variant found is a polymorphism or if it is a mutation, and if it has already been reported.
- 4- Use bioinformatics tools that predict the consequence of a new variant.
- 5- See if there is any specific treatment for tumors that have that variant.

# 1-Search for information about a gene and related pathology.

A-PubMed

B- OMIM

For didactic purposes, suppose that you have started working on lung cancer, you have performed a biopsy of a non-small cell lung tumor (NSCLC) from a patient, finding the following result:

NM\_005228.5(EGFR):c.2369C>T

NP\_005219.2:p.Thr790Met

That is in chromosome 7 , position 55181378 , of GRCh38 Genome Assembly



Always BEWARE to know the Genome Assembly you are using because the positions are NOT the same

## A- PubMed <https://pubmed.ncbi.nlm.nih.gov/>

National Library of Medicine  
National Center for Biotechnology Information

Log in

PubMed.gov

lung cancer egfr mutation

Advanced

PubMed® comprises more than 34 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full text content from PubMed Central and publisher web sites.

National Library of Medicine  
National Center for Biotechnology Information

Log in

PubMed.gov

lung cancer egfr mutation

Advanced Create alert Create RSS User Guide

Save Email Send to Sorted by: Best match Display options

MY NCBI FILTERS  15,140 results Page 1 of 1,514 < >

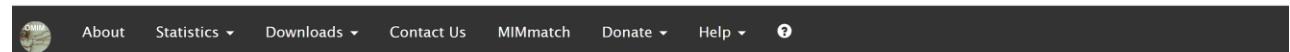
RESULTS BY YEAR

1985 2022

Share  **EGFR mutations and lung cancer.**  
da Cunha Santos G, Shepherd FA, Tsao MS.  
Annu Rev Pathol. 2011;6:49-69. doi: 10.1146/annurev-pathol-011110-130206.  
PMID: 20887192 Review.  
Given that more than 60% of non-small cell **lung** carcinomas (NSCLCs) express **EGFR**, **EGFR** has become an important therapeutic target for the treatment of these tumors. ...We review the role of **EGFR mutations** in the diagnosis and management of NSCLC ...

# B-OMIM Online Mendelian Inheritance in Man

<http://www.omim.org>



## OMIM®

### An Online Catalog of Human Genes and Genetic Disorders

Updated October 7, 2022

Search OMIM for clinical features, phenotypes, genes, and more... 

Advanced Search : OMIM, Clinical Synopses, Gene Map

Need help? : Example Searches, OMIM Search Help,  OMIM Video Tutorials

Mirror site : <https://mirror.omim.org>



Search OMIM...



Options 

Display:  Highlights

\*131550  
Table of Contents

Title

Gene-Phenotype Relationships

Text

Description

Cloning and Expression

Gene Structure

Mapping

Biochemical Features

Gene Function

Cytogenetics

Molecular Genetics

Animal Model

Allelic Variants

Table View

\* 131550

### EPIDERMAL GROWTH FACTOR RECEPTOR; EGFR

Alternative titles: symbols

V-ERB-B AVIAN ERYTHROBLASTIC LEUKEMIA VIRAL ONCOGENE HOMOLOG ONCOGENE ERBB  
ERBB1  
HER1  
SPECIES ANTIGEN 7; SA7

Other entities represented in this entry:

EGFR/SEPT14 FUSION GENE, INCLUDED

HGNC Approved Gene Symbol: EGFR

#### External Links

▶ Genome

▶ DNA

▶ Protein

▶ Gene Info

▶ Clinical Resources

#### ▼ Variation

ClinVar

gnomAD

GWAS Catalog

GWAS Central

HGMD

NHLBI EVS

PharmgKB

#### ▶ Animal Models

## Gene-Phenotype Relationships

Location	Phenotype	<a href="#">View Clinical Synopses</a>	Phenotype MIM number	Inheritance	Phenotype mapping key
7p11.2	?Inflammatory skin and bowel disease, neonatal, 2 		616069	AR	3
	Adenocarcinoma of lung, response to tyrosine kinase inhibitor in		211980	AD, SMu	3
	Non-small cell lung cancer, response to tyrosine kinase inhibitor in		211980	AD, SMu	3
	{Non-small cell lung cancer, susceptibility to}		211980	AD, SMu	3

PheneGene Graphics 



2- Find information about a gene and the protein it encodes.

- C) Genecards
- D) Protein Data Bank

If I want to know information about a particular gene, I can use several free tools available online:

## C- Genecards

<http://www.genecards.org>

The screenshot shows the GeneCards homepage. At the top, there's a navigation bar with links like GeneCards Suite, GeneCards, GeneCaRNA, MalaCards, PathCards, VarElect, GeneAnalytics, GeneALaCart, and GenesLikeMe. Below the navigation is a banner for "GeneCards THE HUMAN GENE DATABASE". The main content area features a search bar with "Keywords" and "Search Term" fields, and a "GO" button. A large orange DNA helix icon is prominently displayed. To the left, under "Explore a Gene", there's a search input field containing "EGFR" which is highlighted with a red box. Below it is a link "Jump to section for this gene:". To the right, there are sections for "NGS Analysis" (with a VarElect link) and "Affiliated Databases" (listing MalaCards, PathCards, GeneALaCart, and GeneCaRNA).

This site provides a lot of information from different sites, in it we can know information about the position of the gene, the length of the protein it encodes, what plasmids are available to express the protein, etc.

Using the information available on this site, download the protein sequence encoded by the EGFR gene in FASTA format, since we will need it later.

# D- Protein Data Bank

<https://www.rcsb.org/>

The screenshot shows the RCSB PDB homepage. On the left, there's a sidebar with links for Welcome, Deposit, Search, Visualize, and Analyze. The main content area features a search bar with "EGFR" entered. Below the search bar, a list of search results includes "in Additional Structure Keywords" and "in Structure Title". To the right, there's a large molecular model of EGFR with a blue domain and orange ligand.

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB Contact us

**PDB** PROTEIN DATA BANK 196,108 Structures from the PDB 1,000,361 Computed Structure Models (CSM)

**PDB-101** **wwPDB** **EMDataResource** **Nucleic Acid Database** **wwPDB Foundation**

**NEW! Computed Structure Models (CSM)**

RCSB Protein Data Bank (RCSB PDB) enables breakthrough science and education by providing access and tools for visualization, and analysis of:

- Experimentally-determined 3D structures from the PDB archive
- Computed Structure Models (CSM) from Alpha ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.

MAX2

If we enter one of the crystallographic structures we can see:

This screenshot shows the detailed view for structure 5XDK. It includes a 3D ribbon model of the protein, experimental data snapshot, validation metrics, and a 3D report link.

Biological Assembly 1 5XDK

Crystal structure of EGFR 696-1022 T790M in complex with CO-1686

PDB DOI: 10.2210/pdb5XDK/pdb

Classification: TRANSFERASE  
Organism(s): Homo sapiens  
Expression System: Spodoptera frugiperda  
Mutation(s): Yes

Deposited: 2017-03-28 Released: 2017-12-13  
Deposition Author(s): Yan, X.E., Yun, C.H.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION  
Resolution: 2.35 Å  
R-Value Free: 0.248  
R-Value Work: 0.217  
R-Value Observed: 0.219

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree	18	0.248
Clashscore	0	18
Ramachandran outliers	14.9%	0
Sidechain outliers	13.0%	0
RSRZ outliers	0	13.0%

3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

Display Files Download Files

3D Report Full Report

3-Determine if a gene variant found is a polymorphism or if it is a mutation, and if it has already been reported.

- E) Human Gene Mutation Database
- F) gnomAD
- G) db SNP
- H) Variation Viewer
- I) ClinVar
- J) COSMIC
- K) CBio Portal

# E-The Human Gene Mutation Database (HGMD)

<http://www.hgmd.cf.ac.uk/ac/index.php>

The screenshot shows the HGMD website homepage. At the top, there's a navigation bar with links for Home, Search, help, Statistics, New genes, What is new, Background, Publications, Contact, Register, Login, LSDBs, and Other links. The main content area features a table titled "Mutation types" with columns for "Table:", "Description:", "Public entries:", and "Total entries:". The table lists various mutation types with their counts. A note at the bottom left mentions the database is maintained by Cardiff University. A note at the bottom right says "HGMD Professional 2016.2".

Table:	Description:	Public entries: This site, Academic/non-profit users only	Total entries: HGMD Professional 2016.2
	<b>Mutation totals (as of 2016-09-10)</b>	<b>134732</b>	<b>187995</b>
Gene symbol	The gene description, gene symbol (as recommended by the HUGO Nomenclature Committee) and chromosomal location is recorded for each gene. In cases where a gene symbol has not yet been made official, a provisional symbol has been adopted which is denoted by lower-case letters.	5158	7473
cDNA sequence	cDNA reference sequences are provided, numbered by codon.	5076	7709
Genomic coordinates	Genomic (chromosomal) coordinates have been calculated for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	165408
HGVS nomenclature	Standard HGVS nomenclature has been obtained for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	166488
Missense/nonsense	Single base-pair substitutions in coding regions are presented in terms of a triplet change with an additional flanking base included if the mutated base lies in either the first or third position in the triplet.	74707	105236
Splicing	Mutations with consequences for mRNA splicing are presented in brief with information specifying the relative position of the lesion with respect to a numbered intron donor or acceptor splice site. Positions given as positive integers refer to a 3' (downstream) location, negative integers refer to a 5' (upstream) location.	12498	17120
Regulatory	Substitutions causing regulatory abnormalities are logged in with thirty nucleotides flanking the site of the mutation on both sides. The location of the mutation relative to the transcriptional initiation site, initiation codon, polyadenylation site or termination codon is given.	2644	3522
Small deletions	Micro-deletions (20 bp or less) are presented in terms of the deleted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^).	20803	27975
Small insertions	Micro-insertions (20 bp or less) are presented in terms of the inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^).	8565	11668
	Indels (20 bp or less) are presented in terms of the deleted/inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion.	2001	2662

This database has two versions, a Public and a Professional, the first being free and the second paid.

In order to access this database it is necessary to register with your data, request an institutional email.

We will use

User name: cmatho@fcien.edu.uy

Password: bgc2022

Once registered we can search for the gene of interest and it will show us the number of mutations reported according to the classification: missense, nonsense, etc.:

The Human Gene Mutation Database  
at the Institute of Medical Genetics in Cardiff

Home Search help Statistics New genes What is new Background Publications Contact Register Login LSDBs Other links Edit details Logout

Gene symbol: EGFR Go!

Symbol:  Missense/nonsense Go!

Gene Symbol	Chromosomal location	Gene name	cDNA sequence	Extended cDNA	Mutation viewer
EGFR (Aliases: available to <a href="#">subscribers</a> )	7p12	Epidermal growth factor receptor (Aliases: available to <a href="#">subscribers</a> )	NM_005228.5	Not available	Available to <a href="#">subscribers</a> 
Mutation type		Number of mutations	Mutation data by type ( <a href="#">register</a> or <a href="#">log in</a> )		
Missense/nonsense		18	<a href="#">Get mutations</a>		
Splicing		1	<a href="#">Get mutations</a>		
Regulatory		2	<a href="#">Get mutations</a>		
Small deletions		1	<a href="#">Get mutations</a>		
Small insertions		0	No mutations		
Small indels		0	No mutations		
Gross deletions		0	No mutations		
Gross insertions/duplications		0	No mutations		
Complex rearrangements		0	No mutations		
Repeat variations		2	<a href="#">Get mutations</a>		
Get all mutations by type			Available to <a href="#">subscribers</a> 		
<b>Public total</b> (HGMD Professional 2021.4 total)		<b>24 (60)</b>			

If we enter, for example, the MISSENSE /NONSENSE category we will see the list of mutations and the article(s) where it was reported.

NM\_005228.5 Gene symbol: [EGFR](#) Extended cDNA not available

**Database: Missense/nonsense** - Single base-pair substitutions in coding regions are presented in terms of a triplet change with an additional flanking base included if the mutated base lies in either the first or third position in the triplet. There are currently 18 mutations available in this category.

Missense/nonsense	Splicing	Regulatory	Small deletions	Small insertions	Small indels	Gross deletions	Gross insertions	Complex	Repeats
50 mutations in HGMD <a href="#">professional</a> 2021.4	1 mutation in HGMD <a href="#">professional</a> 2021.4	2 mutations in HGMD <a href="#">professional</a> 2021.4	4 mutations in HGMD <a href="#">professional</a> 2021.4	No mutations	No mutations	No mutations	No mutations	3 mutations in HGMD <a href="#">professional</a> 2021.4	

Further options available in [HGMD professional 2021.4](#)

Accession Number	Codon change	Amino acid change	Codon number	Genomic coordinates & HGVS nomenclature	Phenotype	Reference	Comments
CM187886	<a href="#">CGA-TGA</a>	<a href="#">Arg-TerN</a>	98	Available to <a href="#">subscribers</a> 	Ectodermal dysplasia with severe skin defects and gastrointestinal dysfunction	<a href="#">Hayashi (2018) Hum Genome Var 5, 11</a>	
CM1615752	<a href="#">TGT-TTT</a>	<a href="#">Cys-Phe</a>	326	Available to <a href="#">subscribers</a> 	Lhermitte-Duclos disease	<a href="#">Colby (2016) Cold Spring Harb Mol Case Stud 2, 001230</a>	Functional s...
CM187888	<a href="#">ATC-ARC</a>	<a href="#">Ile-Asn</a>	365	Available to <a href="#">subscribers</a> 	Ectodermal dysplasia with severe skin defects and gastrointestinal dysfunction	<a href="#">Hayashi (2018) Hum Genome Var 5, 11</a>	
CM149155	<a href="#">GGC-GAC</a>	<a href="#">Gly-Rsp</a>	428	Available to <a href="#">subscribers</a> 	Epithelial inflammation	<a href="#">Campbell (2014) J Invest Dermatol 134, 2570</a> Additional report available to <a href="#">subscribers</a> Additional phenotype report available to <a href="#">subscribers</a> Functional characterisation report available to <a href="#">subscribers</a> Additional report available to <a href="#">subscribers</a>	Functional s...
						<a href="#">Morai (1994) Proc Natl Acad Sci U S A 91, 10217</a>	

# F-gnomAD: genome aggregation database

<https://gnomad.broadinstitute.org/>

v2 data set (GRCh37/hg19) → 125,748 exomes and 15,708 genomes

v3 data set (GRCh38) → 76,156 genomes

v4 data set (GRCh38) → 730,947 exomes and 76,215 genomes

The screenshot shows the gnomAD browser homepage. At the top, there's a navigation bar with links for About, Team, Stats, Policies, Publications, Blog, Changelog, Downloads, Forum, Contact, Help/FAQ, and a search bar. Below the navigation bar, a blue header bar says "gnomAD v4 is here! Read our [blog post](#) for more details". The main content area features the "gnomAD" logo with a blue mountain graphic and the text "Genome Aggregation Database". Below the logo is another search bar labeled "Search by gene, region, or variant". An orange arrow points from the text "Here we write the name of the gene or position in the genome and we can also select the version we want to use" to this search bar.

Here we write the name of the gene or position in the genome and we can also select the version we want to use

The screenshot shows the gnomAD browser gene page for EGFR. The top navigation bar is identical to the homepage. The main content area shows the gene name "EGFR epidermal growth factor receptor" and a "Dataset" dropdown set to "gnomAD v4.0.0" and "gnomAD SVs v4.0". Below this, a blue header bar says "gnomAD v4 is here! Read our [blog post](#) for more details".

## EGFR epidermal growth factor receptor

Dataset [gnomAD v4.0.0](#) [gnomAD SVs v4.0](#) [?](#)

**Genome build** GRCh38 / hg38  
**Ensembl gene ID** ENSG00000146648.20  
**MANE Select transcript** [?](#) ENST00000275493.7 / NM\_005228.5  
**Ensembl canonical transcript** [?](#) ENST00000275493.7  
**Other transcripts**  
ENST00000342916.7, ENST00000344576.7, and 7 more  
**Region** [7:55019017-55211628](#)  
**External resources** Ensembl, UCSC Browser, and more

Category	Expected SNVs	Observed SNVs	Constraint metrics
Synonymous	616.2	613	Z = 0.07 o/e = 0.99 (0.93 - 1.06) 0
Missense	1645.9	1307	Z = 3.05 o/e = 0.79 (0.76 - 0.83) 0
pLoF	151.9	58	pLI = 0.98 o/e = 0.38 (0.31 - 0.47) 0

Constraint metrics based on MANE Select transcript (ENST00000275493.7).

Viewing full gene. [Zoom in](#)

exome genome Metric: Over 20 [Save plot](#)

			intron		
7-55221630-GTGC-G	G	c.748-70_748-68delTGC	● intron		1
7-55221790-T-C	E	p.Asp278Asp	● synonymous	Likely benign	2
7-55221829-C-T	E	p.Cys291Cys	● synonymous	Likely benign	3
7-55223534-G-T	E	p.Val301Leu	● missense		1
7-55223675-C-T	E	c.1006+36C>T	● intron		31
7-55224270-A-G	E	p.Ile351Val	● missense	Uncertain significance	5
7-55227903-G-A	E	p.Ser457Asn	● missense		1
7-55227904-T-C	E	p.Ser457Ser	● synonymous	Likely benign	13
7-55227907-T-C	E	p.Asp458Asp	● synonymous	Likely benign	2
7-55229347-C-T	E	c.1631+23C>T	● intron		2
7-55233090-G-A	E	p.Gly614Ser	● missense	Conflicting interpret...	5
7-55233090-G-C	E	p.Gly614Arg	● missense	Uncertain significance	1
7-55240688-G-A	E	p.Pro644Pro	● synonymous	Likely benign	18
7-55240688-G-T	E	p.Pro644Pro	● synonymous		1
7-55240790-G-T	E	p.Thr678Thr	● synonymous	Likely benign	5
7-55240790-G-A	E	p.Thr678Thr	● synonymous	Likely benign	7
7-55249071-C-T	E	p.Th <b>790</b> Met	● missense	drug response	10
7-55249072-G-A	E	p.Th <b>790</b> Thr	● synonymous	Likely benign	7

## Single nucleotide variant: 7-55249071-C-T(GRCh37)

[Copy variant ID](#)

Dataset [gnomAD v2.1.1](#) ▾

Filters	Exomes	Genomes	Total	<a href="#">External Resources</a>
	Pass	Pass		
<a href="#">Allele Count</a>	7	3	10	
<a href="#">Allele Number</a>	251464	31398	282862	
<a href="#">Allele Frequency</a>	0.00002784	0.00009555	0.00003535	
<a href="#">Popmax Filtering At (95% confidence)</a>	0.00002132	0.00004078		
<a href="#">Number of homozygotes</a>	0	0	0	
<a href="#">Mean depth of coverage</a>	93.0	32.4		

### Feedback

[Report an issue with this variant](#)

## Population Frequencies ⓘ

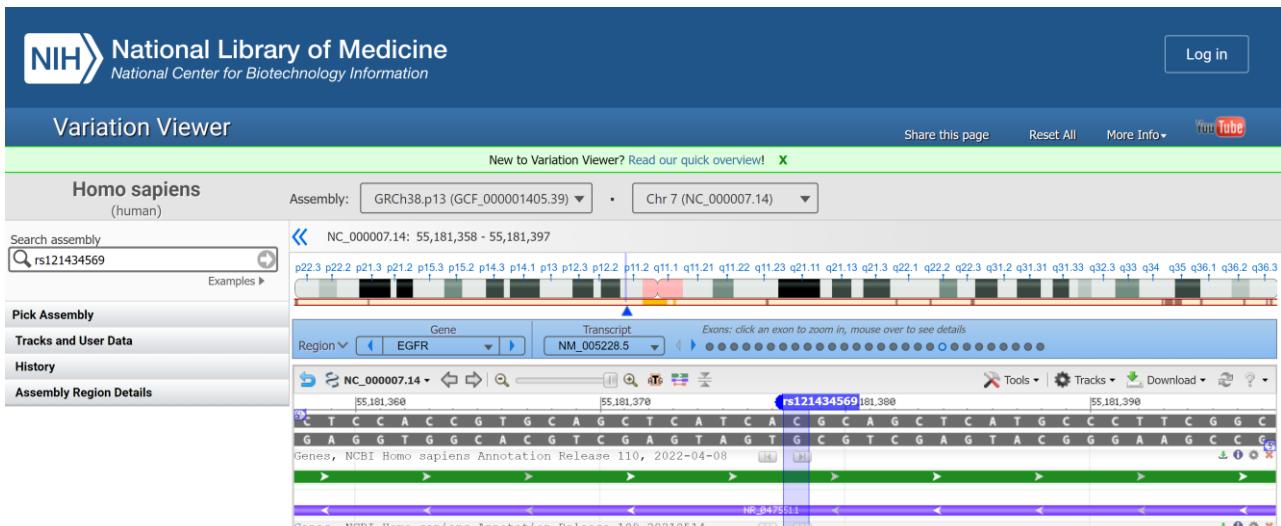
Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
African/African American	4	24970	0	0.0001602
European (non-Finnish)	5	129164	0	0.00003871
Latino/Admixed American	1	35438	0	0.00002822
Ashkenazi Jewish	0	10370	0	0.000
East Asian	0	19952	0	0.000
European (Finnish)	0	25124	0	0.000
Other	0	7228	0	0.000
South Asian	0	30616	0	0.000
XX	5	129476	0	0.00003862
XY	5	153386	0	0.00003260
<b>Total</b>	<b>10</b>	<b>282862</b>	<b>0</b>	<b>0.00003535</b>

## G) db SNP

<http://www.ncbi.nlm.nih.gov/SNP/rs121434569>

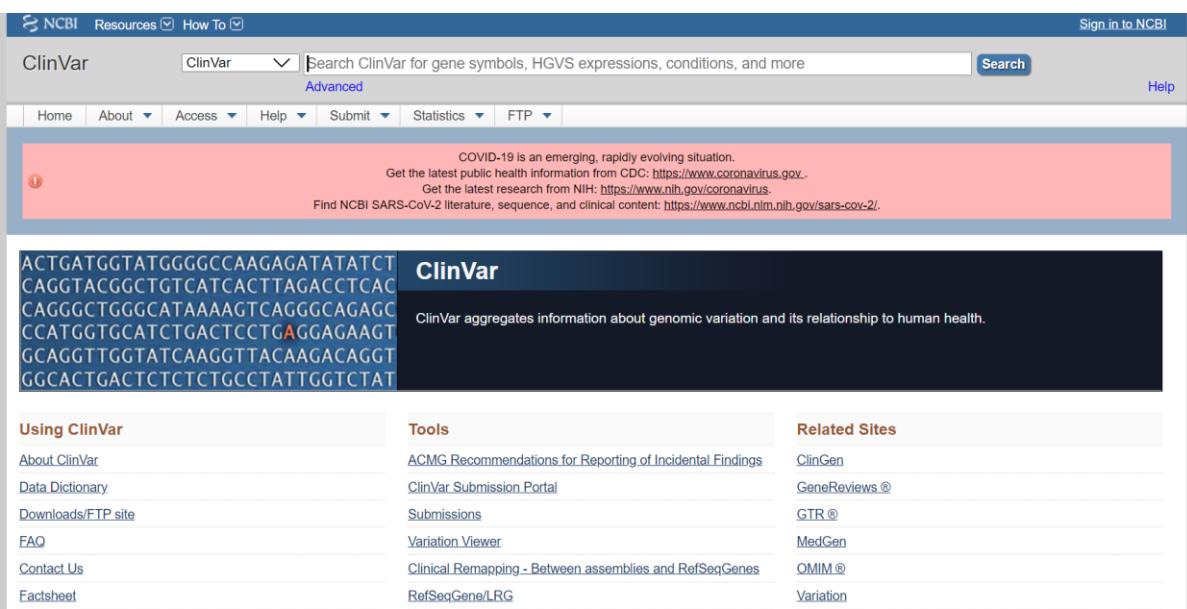
## H) Variation Viewer

<http://www.ncbi.nlm.nih.gov/variation/view/>



## I) Clin Var

<http://www.ncbi.nlm.nih.gov/clinvar/>



## J) COSMIC

### Catalog of Somatic Mutations in Cancer

<http://cancer.sanger.ac.uk/cosmic>

The screenshot shows the COSMIC homepage. At the top, there's a banner announcing the launch of Version v99. Below it, a section for 'COSMIC v99, released 28-NOV-23' provides an overview of the resource. A search bar at the bottom left allows users to search for specific genes or mutations. To the right, there's a 'COSMIC News' section featuring a blog post by Dr. Burcak Otu about topographical features in the genome. Another section below it is dedicated to Sir Professor Mike Stratton's retirement.

**COSMIC v99, released 28-NOV-23**

COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

Start using COSMIC by searching for a gene, cancer type, mutation, etc. below.

eg Braf, COLO-829, Carcinoma, V600E, BRCA-UK, Campbell

**COSMIC v99: The release stats**

Category	Value	Description
New coding mutations	208,168	New genomic variants
New non-coding mutations	274,853	New samples
New whole genomes	1,303	New samples

**COSMIC**  
Catalogue Of Somatic Mutations In Cancer

**COSMIC News**

**Topographical features in the genome and how they're helping us understand tumorigenesis**

In a special guest-edition to our blog, Dr Burcak Otu from the Cancer Grand Challenges Mutographs team talks us through the new topographical analysis data being added to COSMIC following the recent publication of the team's latest paper! [More...](#)

**Sir Professor Mike Stratton**

As he prepared to step down as director of The Sanger Institute we

## K) CBio Portal

### <http://www.cbioportal.org/>

The screenshot shows the cBioPortal homepage. It features a search bar at the top and a main area for selecting studies for visualization and analysis. A sidebar on the right displays a news feed from the cBioPortal Twitter account, mentioning a job opening at Dana-Farber and encouraging sign-up for email alerts.

**Select Studies for Visualization & Analysis:** 0 studies selected (0 samples)

Study Type	Count	Studies
PanCancer Studies	10	TCGA PanCancer Atlas Studies
Pediatric Cancer Studies	13	Curated set of non-redundant studies
Immunogenomic Studies	8	MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017)
Cell lines	3	Metastatic Solid Cancers (UMich, Nature 2017)
Adrenal Gland	3	MSS Mixed Solid Tumors (Broad/Dana-Farber, Nat Genet 2018)
Ampulla of Vater	1	SUMMIT - Neratinib Basket Study (Multi-Institute, Nature 2018)
Biliary Tract	15	TMB and Immunotherapy (MSKCC, Nat Genet 2019)
Bladder/Urinary Tract	18	Tumors with TRK fusions (MSK, Clin Cancer Res 2020)

Please cite: Cerami et al., 2012 & Gao et al., 2013

**What's New**

cBioPortal Retweeted  
Dana-Farber Dat... @dfcidat... Jun 13

Awesome chance to join our @cbioportal team! Apply: social.icims.com/viewjob/pt1654...

#opensource #genomics

Sign up for low-volume email news alerts

**Example Queries**

- Primary vs. metastatic prostate cancer
- RAS/RAF alterations in colorectal cancer
- BRCA1 and BRCA2 mutations in ovarian cancer
- POLE hotspot mutations in endometrial cancer
- TP53 and MDM2/4 alterations in GBM
- PTEN mutations in GBM in text format
- Patient view of an endometrial cancer case

Non-Small Cell Lung Cancer (MSK, Cancer Cell 2018) 

Whole-exome sequencing of 75 tumor/normal NSCLC pairs treated with PD-1 plus CTLA-4 blockade.. PubMed

Click gene symbols below or enter here

Query

Summary

Clinical Data

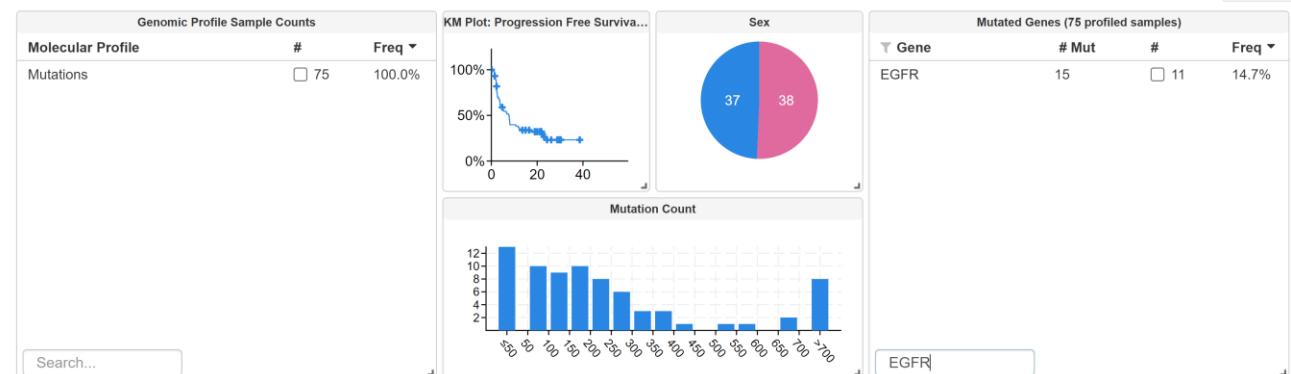
Selected: 75 patients | 75 samples



Custom Selection

Charts

Groups

Study Page Help 

## 4- Use bioinformatics tools that predict the consequence of a new variant.

- L) Mutation Taster
- M) Polyphen
- N) SNP's and Go
- O) Human Splicing Finder

For this section we must be clear about what reference sequence we are talking about, know transcript numbers, etc.

Therefore, it is always advisable to have all the information beforehand.

## L) Mutation Taster

<http://www.mutationtaster.org/>



## mutation t@sting

- [documentation](#) | [FAQs](#)
- [single query](#)
- [query chromosomal positions](#)
- [QueryEngine](#)
- [MutationDistiller \(stable\)](#)
- [RegulationSpotter \(stable\)](#)
- [other applications](#) | [team](#)
- [slides ESHG2017 Copenhagen](#)

Gene  HGNC gene symbol, NCBI Gene ID, Ensembl gene ID [show available transcripts](#)

Transcript  Ensembl transcript ID

Position / snippet refers to  coding sequence (ORF)  transcript (cDNA sequence)  gene (genomic sequence)

Alteration [all types by sequence](#)

enter a few bases around your alteration

Format:  
ACTGTC[A/*T*] GTGTF      A substituted by *T*  
ACTGTC[AG/*T*] GTGTF      AG substituted by *T*  
ACTGTC[ACGT/-]*T* GTGTF      ACGT deleted  
ACTGTC[-AA] GTGTF      AA inserted

show nucleotide alignment

single base exchange by position

enter position  
and new base

insertion or deletion by position

enter positions of

## Results can be:

*disease causing* - i.e. probably deleterious

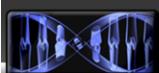
*disease causing automatic* - i.e. known to be deleterious

*polymorphism* - i.e. probably harmless

*polymorphism automatic* - i.e. known to be harmless

## M) Polyphen

<http://genetics.bwh.harvard.edu/pph2/>

 PolyPhen-2 prediction of functional effects of human nsSNPs

Home About Help Downloads Batch query WHESS.db

PolyPhen-2 (Polymorphism Phenotyping v2) is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations. Please, use the form below to submit your query.

21-Jun-2021: Server has been migrated to new hardware. Note, all queries were terminated and user sessions data discarded in the process, hence you will need to resubmit your query if affected. We apologize for the inconvenience caused.

**Query Data**

Protein or SNP identifier

Protein sequence in FASTA format

Position

Substitution   
AA<sub>1</sub> A R N D C E Q G H I L K M F P S T W Y V  
  
AA<sub>2</sub> A R N D C E Q G H I L K M F P S T W Y V

Query description

Guardado en Este PC     
 Display advanced query options

3 possible scores: **benign**, **possibly damaging**, or **probably damaging**

<http://genetics.bwh.harvard.edu/pph2/dokuwiki/overview>

## Prediction

PolyPhen-2 predicts the functional significance of an allele replacement from its individual features by Naïve Bayes classifier trained using supervised machine-learning.

Two pairs of datasets were used to train and test PolyPhen-2 prediction models. The first pair, **HumDiv**, was compiled from all damaging alleles with known effects on the molecular function causing human Mendelian diseases, present in the UniProtKB database, together with differences between human proteins and their closely related mammalian homologs, assumed to be non-damaging. The second pair, **HumVar**, consisted of all human disease-causing mutations from UniProtKB, together with common human nsSNPs (MAF>1%) without annotated involvement in disease, which were treated as non-damaging.

The user can choose between HumDiv- and HumVar-trained PolyPhen-2 models. Diagnostics of Mendelian diseases requires distinguishing mutations with drastic effects from all the remaining human variation, including abundant mildly deleterious alleles. Thus, HumVar-trained model should be used for this task. In contrast, HumDiv-trained model should be used for evaluating rare alleles at loci potentially involved in complex phenotypes, dense mapping of regions identified by genome-wide association studies, and analysis of natural selection from sequence data, where even mildly deleterious alleles must be treated as damaging.

For a mutation, PolyPhen-2 calculates Naïve Bayes posterior probability that this mutation is damaging and reports estimates of false positive rate (FPR, the chance that the mutation is classified as damaging when it is in fact non-damaging) and true positive rate (TPR, the chance that the mutation is classified as damaging when it is indeed damaging). A mutation is also appraised qualitatively, as **benign**, **possibly damaging**, or **probably damaging** based on pairs of false positive rate (FPR) thresholds, optimized separately for each model (e.g., HumDiv and HumVar).

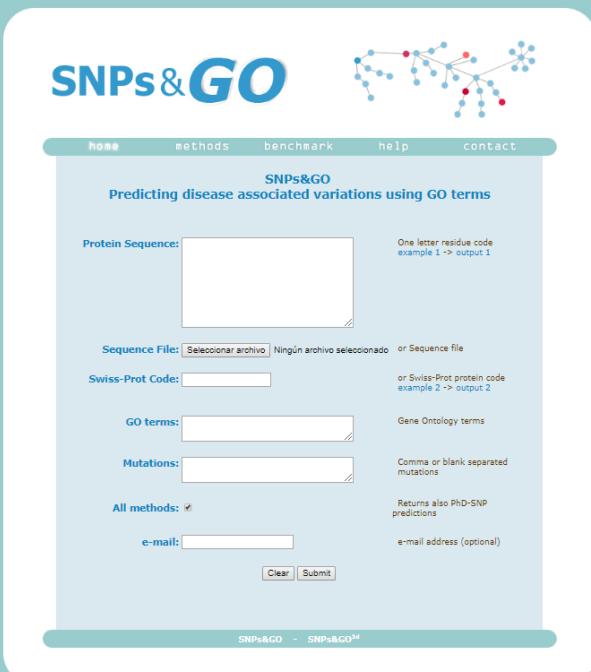
Current version 2.2 of the PolyPhen-2 uses 5% / 10% FPR for **HumDiv** model and 10% / 20% FPR for **HumVar** model as the thresholds for this ternary classification. Mutations with their posterior probability scores associated with estimated false positive rates at or below the first (lower) FPR value are predicted to be **probably damaging** (more confident prediction). Mutations with the posterior probabilities associated with false positive rates at or below the second (higher) FPR value are predicted to be **possibly damaging** (less confident prediction). Mutations with estimated false positive rates above the second (higher) FPR value are classified as **benign**.

If the lack of data does not allow to make a prediction then the outcome is reported as **unknown**.

## N) SNP's and Go

### Predicting disease associated variations using GO terms

<http://snps.biogrid.org/snps-and-go/snps-and-go.html>



The screenshot shows the SNPs&GO web interface. At the top, there is a navigation bar with links for 'home', 'methods', 'benchmark', 'help', and 'contact'. Below the navigation bar is a header with the text 'SNPs&GO' and 'Predicting disease associated variations using GO terms'. The main form area contains several input fields:

- Protein Sequence:** A text input field with placeholder text 'One letter residue code example 1-> output 1'.
- Sequence File:** A file input field with placeholder text 'Seleccionar archivo Ningún archivo seleccionado or Sequence file'.
- Swiss-Prot Code:** A text input field with placeholder text 'or Swiss-Prot protein code example 2-> output 2'.
- GO terms:** A text input field with placeholder text 'Gene Ontology terms'.
- Mutations:** A text input field with placeholder text 'Comma or blank separated mutations'.
- All methods:** A checkbox labeled 'All methods'.
- e-mail:** An input field for an e-mail address.

At the bottom of the form are 'Clear' and 'Submit' buttons, and a footer with the text 'SNPs&GO - SNPs&GO<sup>3d</sup>'.

We need the Swiss prot code or the sequence

The result looks like this:

Mutation	Prediction	RI	Probability	Method
R500G	Neutral	5	0.271	PhD-SNP: F[R]=16% F[G]=3% Nali=90
	Neutral	6	0.208	PANTHER: F[R]=19% F[G]=4%
	Disease	7	0.826	SNPs&GO

**Mutation:** WT+POS+NEW  
WT: Residue in wild-type protein  
POS: Residue position  
NEW: New residue after mutation

**Prediction:**  
**Neutral:** Neutral variation  
**Disease:** Disease associated variation

**RI:** Reliability Index

**Probability:** Disease probability (if >0.5 mutation is predicted Disease)

**Method:** SVM type and data  
PANTHER: Output of the PANTHER algorithm  
PhD-SNP: SVM input is the sequence and profile at the mutated position  
SNPs&GO: SVM input is all the input in PhD-SNP, PANTHER and GO terms features

**F[X]:** Frequency of residue X in the sequence profile  
**Nali:** Number of aligned sequences in the mutated site

SNPs&GO - SNPs&GO<sup>3d</sup>

It also has the option of sending us the result via email

# O) Human Splicing Finder

<http://www.umd.be/HSF3/>

The screenshot shows the homepage of the Human Splicing Finder (HSF) website. At the top, there is a blue navigation bar with links for Home, Analyse Now!, What's New?, Help & Tutorials, Credits & Publications, Our Other Tools, and Contact Us. Below the navigation bar, there are two main sections: 'Description' on the left and 'Get Started' on the right. The 'Description' section contains text about the completion of the Human Genome Project and the identification of mutations in diagnostic and research laboratories. It also mentions the creation of the HSF website to help study pre-mRNA splicing. The 'Get Started' section features a large button with the text 'Start an Analysis with HSF 3.1' and a sunburst graphic.

**Description**

With the completion of the Human Genome Project our vision of human genetic diseases has changed. Thousands of mutations are identified in diagnostic and research laboratories yearly. The knowledge of these mutations associated with clinical and biological data is essential for clinicians, geneticists and researchers.

In order to better understand intronic and exonic mutations leading to splicing defects, we decided to create the **Human Splicing Finder** website. This tool is aimed to help studying the pre-mRNA splicing [more about splicing background].

**Get Started**

Start an Analysis with  
**HSF 3.1**

The Human Splicing Finder system is licensed to the **GENOMNIS SAS** company, which developed the **HSF Pro system**.

The Genomnis SAS company, in order to support research worldwide, has decided to provide free access to its products to academic researchers by giving them free tickets. Concomitantly, other users have the choice between immediate access by obtaining tickets or standard access through specific annual licenses whose prices are decreasing according to the annual activity.

GENOMNIS      HSF Pro system      Sales Policy

• RegRNA: A Regulatory RNA motifs and Elements Finder  
• EBI Splice Signal Analysis  
• GeneSplicer

Director: Christophe BERROUD

Need to register individually, limited analyses for the free version



## Access Human Splicing Finder

The GENOMNIS company has developed a **professional version** of the HSF system, which is available for all users. If you are using HSF for **academic research only**, you are eligible for a **free access** (limited to a standard annual use of the system), otherwise for **public or private diagnostic** and **private research** activities, you need to subscribe a **license**.

The licensing model is based on the **Software as a Service** (SaaS) also known as on-demand software, hosted software or web-based software. The pricing is based on a pay per use model and is based on a per patient cost related to the type of analysis (small gene panels, medium gene panels and large gene panels and exomes). To get pricing details, please contact us [here](#).

For users willing to get predictions for a handful of mutations, you can use our e-commerce options to buy credits for few analysis. This option is also helpful to evaluate the new system.

The **professional version** of the HSF system gives you access to a new world:



You need to login to buy tickets



Improved Predictions Accuracy



New ESE/ESS Predictions



Ability to analyse large VCF files

# 5- See if there is any specific treatment for tumors that have that variant

<https://www.pharmgkb.org/>

New to the site? [Take a short tour](#) of the chemical overview page. X

 PHARMGKB erlotinib X Add a term to make a combination... Search icon ☰ Menu Help

## erlotinib

OVERVIEW PRESCRIBING INFO DRUG LABEL ANNOTATIONS CLINICAL ANNOTATIONS PATHWAYS

Prescribing Info 2 5 5 2

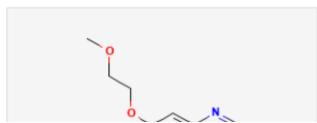
Drug Label Annotations 5

Clinical Annotations 5

Variant Annotations 2

Structure Email icon

[large version](#)  
[3D version](#)  
source: PubChem



- Clinical trials: <https://clinicaltrials.gov/>

We're building a better [ClinicalTrials.gov](#). Check it out and tell us what you think!

 U.S. National Library of Medicine [ClinicalTrials.gov](#) Find Studies About Studies Submit Studies Resources About Site PRS Login

Home > Search Results

Modify Search Start Over + icon

196 Studies found for: **egfr t790m**  
Also searched for **Epidermal Growth Factor Receptor, T790m, and EGFR Gene.** [See Search Details](#)

List By Topic On Map Search Details

Hide Filters Download Subscribe to RSS Show/Hide Columns

Showing: 1-10 of 196 studies 10 studies per page

Row	Saved	Status	Study Title	Conditions	Interventions	Locations
1	<input type="checkbox"/>	Completed	ARTEMIS DIANE T790M (An Amino Acid Substitution at Position	• Locally Advanced or	• Procedure:	• Research Site

# Clinical Interpretation of Variants in Cancer (CIViC) knowledgebase

<https://civicdb.org/welcome>

The screenshot shows the CIViC knowledgebase homepage. On the left is a dark sidebar with the CIViC logo and links to various databases: Assertions, Evidence, Genes, Variants, Variant Groups, Clinical Trials, Diseases, Drugs, Phenotypes, Sources, Variant Types, and Activity. The main content area has a purple header with the text "Participate with colleagues to add variants and support for cancer-related mutations.". Below this is a "Knowledgebase Statistics" section with a table of counts for assertions, evidence, genes, variants, diseases, drugs, phenotypes, sources, and variant types. There are also sections for "News & Events" and "Live Curation Activity". A top navigation bar includes a search bar, "Home", "About CIViC", "Help", and "Sign In / Sign Up".

Total Assertions	Total Evidence	Total Genes	Total Variants	Total Contributors
53	9,302	479	3,337	331
Total Diseases	Total Drugs	Total Sources	Total Revisions	Total Comments
341	494	3,239	34,140	61,845

The screenshot shows a detailed view of a variant in the CIViC knowledgebase. The variant is EGFR T790M, with the ID RS121434569. The top navigation bar includes a search bar, "Home", "About CIViC", "Help", and "Sign In / Sign Up". The sidebar on the left is identical to the homepage sidebar. The main content area shows the variant summary, including its description, gene information (EGFR), representative variant coordinates (GRCH37), and various scores and IDs from ClinVar and Ensembl. It also shows the variant's sources (PubMed articles), aliases (THR790MET, RS121434569), and a "MyVariant.info" section with links to ClinVar, gnomAD, EXAC, CADD, EGL, and other resources. At the bottom, it provides the MyVariant.info ID and ClinVar ID.

Variant: EGFR T790M, RS121434569

Description: EGFR T790M was one of the very first mutations recognized to confer resistance to targeted therapies in non-small cell lung cancer. While successful in amplified EGFR, the efficacy of the first and second generation TKI's (erlotinib, gefitinib, neratinib) in treating patients harboring this mutation before treatment is notably lower. This lack of efficacy can likely be to blame for the poorer prognosis for patients with this mutation as compared to patients with wildtype EGFR or other types of EGFR mutations. Approximately half of EGFR mutant tumors with acquired resistance to TKI inhibition have been shown to harbor this mutation, implicating it as a mechanism of acquired therapy resistance. A third generation TKI (osimertinib) has been approved for the treatment of EGFR T790M mutant NSCLC. Patients positive for T790M in a plasma-based test have similar outcomes like those with tumor biopsy testing.

Gene: EGFR

Representative Variant Coordinates:

Ref. Build	GRCH37	Ensembl Version	75
Coordinates			
Chr.	7	Start	55249071
Stop	55249071	Ref. Bases	C
Var. Bases	T	Transcript	ENST00000275493.2

MyVariant.info

Overview ClinVar gnomAD (2.1.1) EXAC (0.3.1) CADD EGL ...

MyVariant.info ID: chr7:g.55249071C>T ClinVar ID: 16613

**We saw some tools, there are many more places to obtain information from, some of them:**

**-Gene information**

Ensembl <http://www.ensembl.org/index.html>

UCSC Genome Browser: <https://genome.ucsc.edu/>

-Allelic frequencies

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

**-Protein Data**

UniProt: <http://www.uniprot.org>

Human Genome Variation Society, nomenclature information at  
<https://varnomen.hgvs.org/>

Plasmids to buy:<http://www.addgene.org>