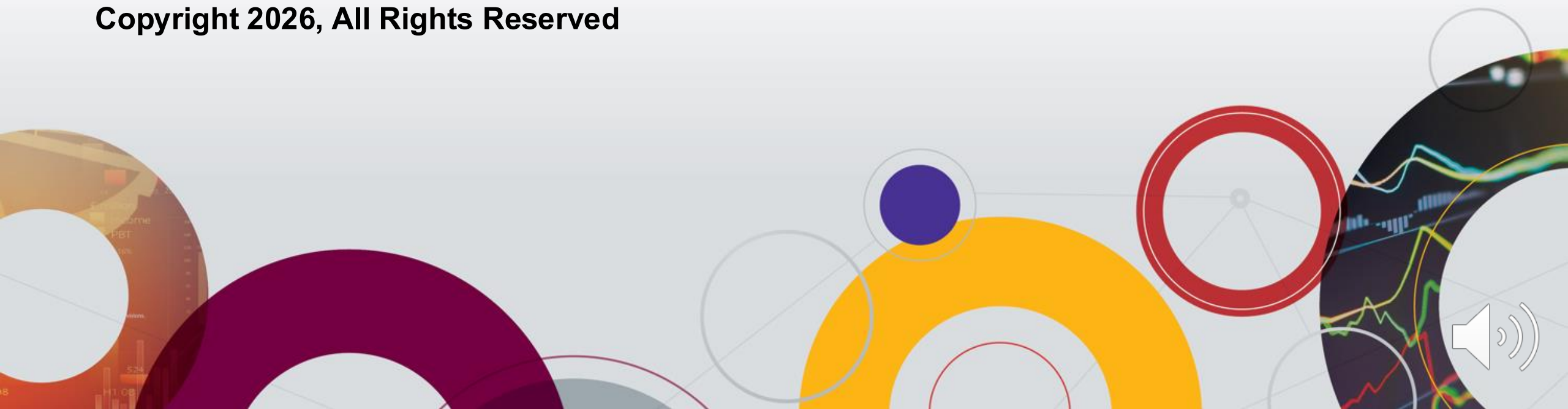


9.6 Omics Data in the EHR and Other Information Systems

10x10 – Introduction to Biomedical Informatics & Artificial Intelligence
William Hersh, MD, FACMI, FAMIA, FIAHSI
Copyright 2026, All Rights Reserved



Omics data in the EHR and other information systems

Genomics-enabled medical records

Annotating the genome

Accessing knowledge-based omics information

Other issues

Toward the genomics-enabled electronic health record (EHR)

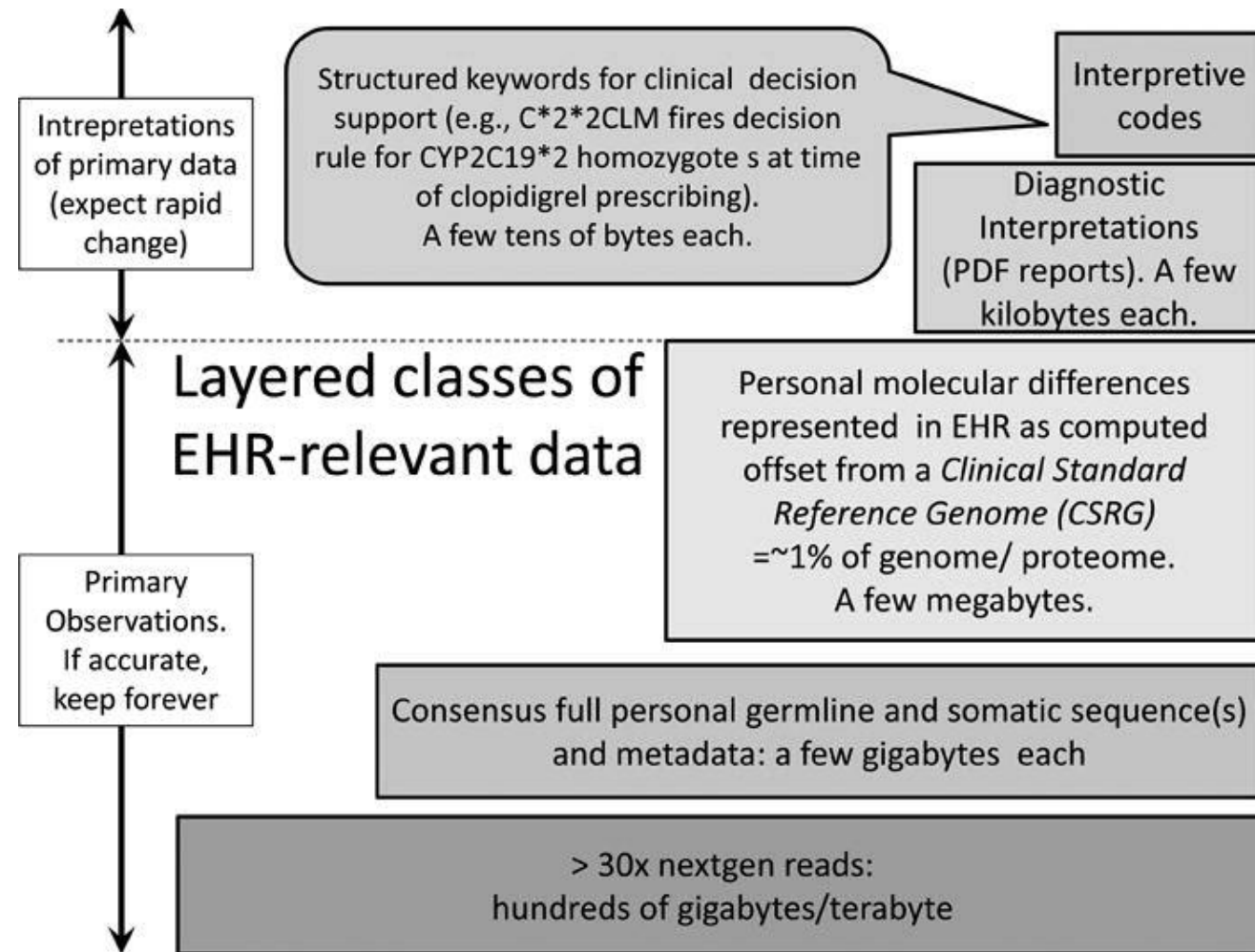
Melding clinical informatics and bioinformatics

- Clinical care
 - Early vision (Ashley, 2010)
 - Now use of genomic screening with EHR data (Williams, 2018; Buchanan, 2020), polygenic risk scores (Lambert, 2021), etc.
- Research
 - Extracting the phenotype (Hripcask, 2018; Kashyap, 2020)
 - Associating with actionable therapies (Joyner, 2019)

Issues

- Data to include/exclude
- Implementing genomic clinical decision support
- Leveraging for research

Classes of EHR-relevant genomics data (Masys, 2012)



Desiderata for the integration of genomic data into EHRs (Masys, 2012)

Maintain separation of primary molecular observations from the clinical interpretations of those data

Support lossless data compression from primary molecular observations to clinically manageable subsets

Maintain linkage of molecular observations to the laboratory methods used to generate them

Support compact representation of clinically actionable subsets for optimal performance

Simultaneously support human-viewable formats and machine-readable formats in order to facilitate implementation of decision support rules

Anticipate fundamental changes in the understanding of human molecular variation

Support both individual clinical care and discovery science

Association for Molecular Pathology statement on challenges and opportunities for EHR to support genomic medicine (Carter, 2022)

Challenges	Opportunities
EHRs are not yet ready to send accurate, coded, and appropriately granular clinical history, signs, symptoms, family history, and other broad sets of data elements to laboratories without generating significant burden on providers.	Develop stakeholder consensus-derived and standardized methods to apply accurate concept codes to the clinical notes, signs, symptoms, and family histories as well as specimen locations, such that necessary information could be sent by the provider using an automated or semi-automated import into the electronic order for genomic testing.
EHRs lack sufficient information about genetic test orders and generally do not have discrete variant result data to facilitate appropriate test ordering and utilization.	Establish a consensus standard for minimum discrete data required to define a genomic variant in an EHR.*
Absence of standardized genomic variant data structures from the LIS to the EHR.	Establish a consensus standard for minimum discrete data needed to define a genetic/genomic test order.
Interoperability standards for genomics are currently limited in several ways, including the use of syntax with limited hierarchy (HL7 version 2.x) and inadequate coding systems for genomic orders and results.	Before mandating changes to interoperability requirements, governments and regulators should carefully review the cost and burden to laboratories as well as the safety of existing coding standards that are currently inadequate for genomic data. Such work should occur after establishment of a consensus standard for minimum discrete data required to define a genomic variant.
Genomic reports between laboratories are variable in structure and content.	Develop recommendations for content and structure of genomic reports that support providing the established consensus standard for minimum discrete data required to define a genomic variant in an EHR.*
Implementation of molecular pathology and genomics professional societies multi-organizational consensus standards and guidelines for variant nomenclature, hierarchical result structure, and genomic report formats has been modest.	Encourage laboratories and EHRs to support and implement molecular pathology and genomics professional societies multi-organizational consensus standards and guidelines for report content and structure to standardize structure and content between laboratories.*
There are no standards for how to display aggregated variant data over time and between sample sources, tests, and laboratories that keep clinical context and associated interpretation intact.	Professional organizations should consider further developing consensus report structures that use sound principles of human factors engineering and usability.
Consensus guidelines on best practice for requesting and providing reclassification of variants is not currently available.	Standards for aggregation of genomic data over time and between sample sources, tests, and laboratories should be developed that keep clinical context and associated interpretation intact.
Genomic reports are difficult for medical personnel and patients to understand, and in the United States, patients have the right to immediately access their genomic test reports on request.	Establish consensus guidelines on best practice for requesting and providing reclassification of variants.*
Absence of national and international standards for CDS rules.	Professional organizations should consider further developing consensus report structures that use sound principles of human factors engineering and usability that enable understanding by most patients.
Currently, textual data from genomic reports are exported and released to authorized parties, and analysis of such data is limited by its highly variable format and lack of structure.	Establish evidence-based international recommendations for CDS in genomic test orders and in drug orders impacted by genomic results for safety and consistency in practice with a prerequisite minimum discrete genomic variant data consensus standard established.
Current technology lacks sufficient functionality to ensure that release of genomic data to appropriate health care organizations, research studies, or clinical repositories requires and receives informed consent from the patient or legal guardian, and/or institutional board review.	Establish and integrate appropriate, safe, and functional international standards for interoperability and data retrieval.
Current coding and interoperability standards are not adequate for genomic data.	Develop technology and cybersecurity functions in EHRs that ensure that genomic data are released as authorized after informed consent from the patient along with tools to educate patients about informed consent.
	Develop safe, complete, and accurate standards for coding and interoperability of genomic data, per the future established minimum discrete variant data standard.

Implementing clinical decision support (CDS)

Implementations of genomic CDS integrated into EHR

- CYP2C19 variants in response to clopidogrel (Pulley, 2012)
- 48 alerts from 11 variants detected by exome sequencing (Nishimura, 2015)
- Screening of newborns for 156 early-onset genetic conditions with established interventions found 3.7% actionable (Ziegler, 2025)

Tools still challenging for use (Sebastian, 2021)

Need to account for changing knowledge of variants in disease, e.g., from ClinVar (Yaury, 2022)

Can obtain high adherence to alerts with education of clinicians and demonstration of value (Nguyen, 2022)

CDS in genomics and precision medicine

Must build on and integrate standards

- HL7 Infobutton – adding clinical genomics resources, e.g., ClinGen (Heale, 2016)
- SMART on FHIR
 - Initial application to genomics (Alterovitz, 2015)
 - SMART Precision Cancer Medicine app provides information tailored to patients' genomics findings (Warner, 2016)
 - ONC Sync for Genes – leveraging FHIR Resources (Garcia, 2020) and APIs (Garcia, 2022)
 - Continued adherence to all standards essential (Versel, 2022)
 - FHIR-based Genomic Reporting Implementation Guide (Stellmach, 2022)
 - FHIR Genomics Operations (Dolin, 2023)

ACMG guidelines for genetic data in EHR (Grebe, 2020)

Genetically-guided precision medicine CDS tools (Johnson, 2024)

GMXIII Short Term Goals

Systematic Review Findings

GMXIII Long Term Goals

Developing new and existing tools and workflows to address systemic bias

Revising the Technical Desiderata

Developing user-friendly clinician and patient centered genomic based tools and workflows

Identifying semantic and syntactic gaps related to the representation of genomic information in existing clinical data standards and models

Developing & implementing genomics-based computational tools that automatically extract clinical genomic evidence from variant interpretation to reduce manual curation

Generating evidence from research focused on end-user prioritized use cases to standardize annotation and interpretation for clinical care

1. User-Centered Design strategies are underutilized in gCDS system design

2. Genetic data are difficult to work with and need standardization for storage and data transfer

3. Few CDS systems are utilizing standards-based approaches for storing and transmitting genetic information

4. Implementation frameworks are underutilized in guiding gCDS systems and aiding the findings, including measuring underserved populations

5. Clinicians lack understanding or clinical relevance of genetic information

6. Multi-disciplinary teams enhance the development of gCDS systems

7. Evidence-based and personalized recommendations are effective for facilitating a genetic diagnosis

Developing efficient, portable, and user-friendly storage of genomic and genomics-based information for clinical use

Studying genomics-based CDS tools and resources that ensure equitable implementations

Developing and assessing the value proposition for genomics-based clinical informatics systems

Developing and implementing genomics-based interfaces between laboratory information systems and the electronic health record

Identifying and addressing barriers related to the establishment and implementation of genomic exchange system

Annotating function of the genome – Gene Ontology (GO; Pitarch, 2025; GO Consortium, 2026)

Gene function is classified by

- Molecular function – what gene product does, e.g., adenylate cyclase enzyme
- Biological process – biological objective of gene, e.g., pyrimidine metabolism
- Cellular component – where function and process take place, e.g., rough endoplasmic reticulum

Each assigned code is associated with evidence code that describes “level” (or quality) of evidence, e.g.,

- Traceable author statement (TAS)
- Inferred from direct assay (IDA)

Similar to other names, gene names can be problematic

Problems include (Chen, 2005)

- Synonym/polysem problem
- Names may be case- and/or style-sensitive
- Gene and corresponding protein may have same name

Even worse, applications such as auto-formatting in Microsoft Excel may change gene names (Ziemann, 2016)

- e.g., gene APR-1 → date 1-Apr
- Problem has persisted despite knowledge about it (Abeysooriya, 2021)
- Microsoft finally added option to disable (Davis, 2023)

Human gene names standardized by [Human Genome Organization \(HUGO\) Nomenclature Committee](#) (Seal, 2026)

Bioinformatics information resources

Much activity led by [National Center for Biotechnology Information](#) (NCBI)

- Overview of resources (Sayers, 2026)

Overviews and updates of major resources provided annually in open-access journal [Nucleic Acids Research](#)

- Database issue covers available databases (Rigden, 2026)
- Web server issue covers accessible resources

Some other major omics resources

- [UCSC Browser](#) (Casper, 2026)
- [Ensembl](#) (Yates, 2026) from EMBL-EBI (Thakur, 2026)

Bioinformatics information resources (cont.)

Bioinformatics community has long-standing commitment to data sharing and linkage

- Many journals require deposit of data into public databases as precondition for publication of papers
- Clinical informatics does not have similar commitment

Can navigate across variety of information resources

- From data about genes (Gene), variations (dbVar), patients (dbGap), and sequences (GenBank) to roles in diseases (OMIM)

Will look at gene EGFR, associated response to treatment in lung cancer

- You can try this at home!

9.6 - Hersh - AMIA

Search

Substances

Deposited substance and chemical information

Search for EGFR

**National Library of Medicine**
National Center for Biotechnology Information

Log in

Search NCBI

egfr

Search

Results found in 28 databases

GENE

Was this helpful?  

EGFR – epidermal growth factor receptor

Homo sapiens (human)

Also known as: ERBB, ERBB1, ERBP, HER1, NISBD2, NNCIS, PIG61, mENA

Gene ID: 1956

[RefSeq products](#) [Orthologs](#) [Genome Data Viewer](#)

New - [Visualize gene across multiple species](#)

RefSeq Sequences +

Literature

Bookshelf	5,960
MeSH	26
NLM Catalog	111
PubMed	138,237
PubMed Central	354,079

Genes

Gene	11,537
GEO DataSets	63,867
GEO Profiles	326,114

Proteins

Conserved Domains	126
Identical Protein Groups	862
Protein	27,920
Protein Family Models	52
Structure	555

Search for EGFR (cont.)

Genomes	Clinical	PubChem
Assembly / Genome NCBI Datasets	ClinicalTrials.gov 0	BioAssays 25,541
BioCollections 0	ClinVar 3,999	Compounds 308
BioProject 1,422	dbGaP 32	Pathways 80
BioSample 5,422	dbSNP 85,415	Substances 1,363
Nucleotide 47,797	dbVar 1,185	
SRA 18,850	GTR 232	
Taxonomy 0	MedGen 341	
	OMIM 256	

NCBI Datasets All assembled genome data is now available through NCBI Datasets

FOLLOW NCBI



Connect with NLM


National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894

Web Policies
FOIA
HHS Vulnerability
Disclosure

Help
Accessibility
Careers

NLM | NIH | HHS | USA.gov

EGFR gene summary

 **National Library of Medicine**
National Center for Biotechnology Information

Log in

Gene [Advanced](#) [Help](#)

Full Report ▾ Send to: ▾ Hide sidebar >>

EGFR epidermal growth factor receptor [*Homo sapiens* (human)] Download Datasets

Gene ID: 1956, updated on 7-May-2024

Summary

Official Symbol EGFR provided by HGNC

Official Full Name epidermal growth factor receptor provided by HGNC

Primary source [HGNC:HGNC:3236](#)

See related [Ensembl:ENSG00000146648](#) [MIM:131550](#); [AllianceGenome:HGNC:3236](#)

Gene type protein coding

RefSeq status REVIEWED

Organism [Homo sapiens](#)

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

Also known as ERBB; ERBP; HER1; mENA; ERBB1; PIG61; NISBD2

Summary The protein encoded by this gene is a transmembrane glycoprotein that is a member of the protein kinase superfamily. This protein is a receptor for members of the epidermal growth factor family. EGFR is a cell surface protein that binds to epidermal growth factor, thus inducing receptor dimerization and tyrosine autophosphorylation leading to cell proliferation. Mutations in this gene are associated with lung cancer. EGFR is a component of the cytokine storm which contributes to a severe form of Coronavirus Disease 2019 (COVID-19) resulting from infection with severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). [provided by RefSeq, Jul 2020]

Annotation information Note: This gene has been reviewed for its involvement in coronavirus biology, and is involved in cytokine storm inflammatory response.

Expression Broad expression in placenta (RPKM 36.6), skin (RPKM 15.6) and 22 other tissues [See more](#)

Orthologs [mouse](#) [all](#)

NEW Try the new [Gene table](#)
Try the new [Transcript table](#)

Genomic context

Location: 7p11.2

Table of contents

[Summary](#)

[Genomic context](#)

[Genomic regions, transcripts, and products](#)

[Bibliography](#)

[Phenotypes](#)

[Variation](#)

[HIV-1 interactions](#)

[Pathways from PubChem](#)

[Interactions](#)

[General gene information](#)
[Markers, Homology, Gene Ontology](#)

[General protein information](#)

[NCBI Reference Sequences \(RefSeq\)](#)

[Related sequences](#)

[Additional links](#)

Genome Browsers

[Variation Viewer \(GRCh37.p13\)](#)

[Variation Viewer \(GRCh38\)](#)

[1000 Genomes Browser \(GRCh37.p13\)](#)

[Ensembl](#)

EGFR genomic context

**National Library of Medicine**
National Center for Biotechnology Information

Log in

Gene

Advanced Help

Full Report Send to:

EGFR epidermal growth factor receptor [*Homo sapiens* (human)]

Gene ID: 1956, updated on 7-May-2024

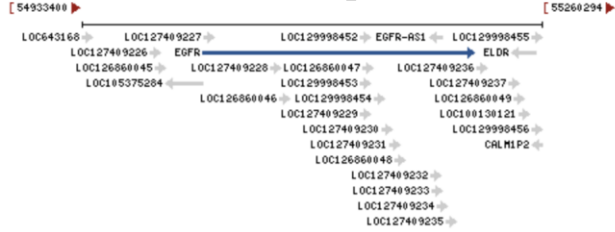
Summary

Genomic context

Genomic regions, transcripts, and products

Location: 7p11.2
Exon count: 32

Annotation release	Status	Assembly	Chr	Location
RS_2023_10	current	GRCh38.p14 (GCF_000001405.40)	7	NC_000007.14 (55019017..55211628)
RS_2023_10	current	T2T-CHM13v2.0 (GCF_009914755.1)	7	NC_060931.1 (55178937..55372056)
105.20220307	previous assembly	GRCh37.p13 (GCF_000001405.25)	7	NC_000007.13 (55086710..55279321)

Chromosome 7 - NC_000007.14


Genomic regions, transcripts, and products

Genome Browsers

Variation Viewer (GRCh37.p13)
Variation Viewer (GRCh38)
1000 Genomes Browser (GRCh37.p13)
Ensembl

EGFR regions, transcripts, and products

NIH National Library of Medicine
National Center for Biotechnology Information

Gene [Advanced](#) [Help](#)

Full Report

EGFR epidermal growth factor receptor [*Homo sapiens* (human)]

Gene ID: 1956, updated on 7-May-2024

☒ Summary ☐ Genomic context ☐ Genomic regions, transcripts, and products

Go to [reference sequence details](#)

Genomic Sequence:

Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)

NC_000007.14 Find:

Genes, MANE Project (release v1.3)

NCBI RefSeq Annotation GCF_000001405.40-RS_2023_10

LOC105375284
XR_007060328.1

NP_005228.5
NP_001346899.2
NP_001346941.2
NP_001346898.2
NP_001346897.2
NP_201284.2
NP_201282.2
NP_201283.2

NP_005219.2
NP_001333828.1
NP_001333870.1
NP_001333827.1
NP_001333826.1
NP_958441.1
NP_958439.1
NP_958440.1
XP_047275909.1
XP_047275908.1
NP_001333829.1

LOC1001301

EGFR-AS1
NR_047551.1

Table of contents

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- Phenotypes
- Variation
- HIV-1 interactions
- Pathways from PubChem
- Interactions
- General gene information
 - Markers, Homology, Gene Ontology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences
- Additional links

Genome Browsers

- Variation Viewer (GRCh37.p13)
- Variation Viewer (GRCh38)
- 1000 Genomes Browser (GRCh37.p13)
- Ensembl

Nucleotide sequence for EGFR

**National Library of Medicine**
National Center for Biotechnology Information

Log in

Nucleotide Nucleotide Search

Advanced Help

FASTA

Send to:

Change region shown

☐ Whole sequence
☒ Selected region
from: 55019017 to: 55211628 Update View

Customize view

Analyze this sequence

[Run BLAST](#)

[Pick Primers](#)

Related information

[Assembly](#)

[BioProject](#)

[Protein](#)

[PubMed](#)

[Taxonomy](#)

[Components \(Core\)](#)

[Full text in PMC](#)

[Gene](#)

Homo sapiens chromosome 7, GRCh38.p14 Primary Assembly
NCBI Reference Sequence: NC_000007.14
[GenBank](#) [Graphics](#)
>NC_000007.14:55019017-55211628 Homo sapiens chromosome 7, GRCh38.p14 Primary Assembly
AGACGTCCGGGCAGCCCCGGCGCAGCGCGGCCGACAGCCTCCGCCCCCGACGGTGTGAGCGCCCC
ACGCGGCCGAGCGCGCCGAGTCCCGAGCTAGCCCCGGCGGCCGCCGCCAGACCGGACGACAGGCC
ACCTCGTGGCGTCCGCCCAGTCCCGCCTCGCCGCCAACGCCACAACCACGCGCAGGCCCCCTGAC
TCCGTCCAGTATTGATCGGGAGAGCCGGAGCGAGCTCTTCGGGGAGCAGCGATGCGACCCCTCCGGGACGG
CCGGGGCAGCGCTCCTGGCGCTGCTGGCTGCGCTCTGCCCGCGAGTCCGGCTCTGGAGGAAAAGAAAGG
TAAGGGCGTGTCTCGCCGGCTCCCGCGCCGCCCGGATCGCGCCCGGACCCGCGAGCCGCCCAACCG
CGCACCGGCGCAGCGCTCGGCGCCCGGCCCGCCGCTCTTCTGTTCTTCTTGAGATCAGTGC
CGCCGACCGGACCGCGGAGGAACGGGACGTTTCGTTCTTCGGCCGGGAGAGTCTGGGCGGGCGGAGG
AGGAGACGCGTGGGACACCGGGTGCAGGCCAGGCGGGGAACGGCGCGGGACCTCCGGCGCCCCGAAC
CGCTCCCACTTTCTTCTCACTTTCCCGCCAGCTGCGCAGGATCGGCGTCAGTGGGCGAAAGCCGG
GTGCTGGTGGGCGCTGGGCGCGGGTCCCGCAGTGCGCCCGCGCTGTCTCCAGGGCGCGACGGGG
TCCTGGCGCGCACCCGAGGGGGCGGGCGCTGCCACCCGCCGAGACTGCACTGTTTAGGGAAGCTGAGGAA
GGAACCAAAAATACAGCTCCCCCTCGGACCCCGCGGGACAGGCGGCTTTCTGAGAGGACCTCCCCGCT
CCGCCCTCCGCGCAGGTCTCAAACTGAAGCCGGCGCCCGCAGCTGGCCCGGCCCTCTCCAGGTCC
CGCGATCCTCGTTCCTCAGTGTGAGTGCAGCCTCGACCTGGGAGCTGGGAGAACTCGTCTACCACAC
CTGCGGCTCCCGGGGAGGGGTGGTGTGGCGCGGTTAGTTTCTCGTTGGCAAAAGGAGGTGGGGTCC
GACCCGCCCTTGGGCGCAGACCCCGGCGCTCGCTCGCCCGGTGCGCCCTCGTCTTGCTATCCAAGA
GTGCCCCCACCTCCCGGGGACCCAGCTCCCTCCTGGGCGCCCGCGCGAAAGCCAGGCTCTCCTTC
GATGGCCGCTCGCGGAGACGTCCGGGTCTGCTCCACCTGCAGCCCTTCGGTTCGCGCTGGGCTTCGCG
TGGAGCGGACCGGCTGTCCGGCACTGCAGGGGGGGATCGCGGACTTTGAGCGGAAGCCCGGAAAG
CAGAGCTCATCTGGCCAACACCATGGTGTTCAAATGGGGCTCACAGCAAACTTCTCTCAAAACCCG
GAGACTTTCTTTCTGGATGTCTTTTTGCTGTTGAAGAAATTTGAGCCAACCAAAATATTAACCTGT
CTTACACACACACACACACACACACACACACACCGGATTGCTGTCCCTGGTTCAAGTGTGCCAAGT
GTGCAGACAGAATGAGCGAGTCTGGCTTCGTGACTACCGACATAAACCCACTTGACAGGGGAAACAT

Genbank record for EGFR

NCBI [Resources](#) [How To](#) [Sign in to NCBI](#)

Nucleotide

Nucleotide

Search

Advanced

Help

The Nucleotide database will include EST and GSS sequences in early 2019. [Read more.](#)

GenBank

Homo sapiens chromosome 7, GRCh38.p12 Primary Assembly

NCBI Reference Sequence: NC_000007.14

[FASTA](#) [Graphics](#)

LOCUS NC_000007 189060 bp DNA linear CON 26-MAR-2018

DEFINITION Homo sapiens chromosome 7, GRCh38.p12 Primary Assembly.

ACCESSION [NC_000007](#) REGION: 55019021..55208080

VERSION NC_000007.14

DBLINK BioProject: [PRJNA168](#)

Assembly: [GCF_000001405.38](#)

KEYWORDS RefSeq.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 189060)

CONSTRM International Human Genome Sequencing Consortium

TITLE Finishing the euchromatic sequence of the human genome

JOURNAL Nature 431 (7011), 931-945 (2004)

PUBMED [15496913](#)

REFERENCE 2 (bases 1 to 189060)

AUTHORS Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H., Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R., Wyllie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E., Fewell,G.A., Delehaunty,K.D., Miner,T.L., Nash,W.E., Cordes,M., Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S., Isak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamar,B., Courtney,L., Kalicki,J., Ozersky,P., Bielicki,L., Scott,K., Holmes,A.,

Send to:

Change region shown

☐ Whole sequence (abbreviated view)

☒ Selected region

from: 55019021 to: 55208080

Update View

Customize view

☐ Abbreviated view

☒ Customize

Basic Features

☐ All features

☒ Gene, RNA, and CDS features only

Display options

☒ Show sequence

☐ Show reverse complement

Update View

Analyze this sequence

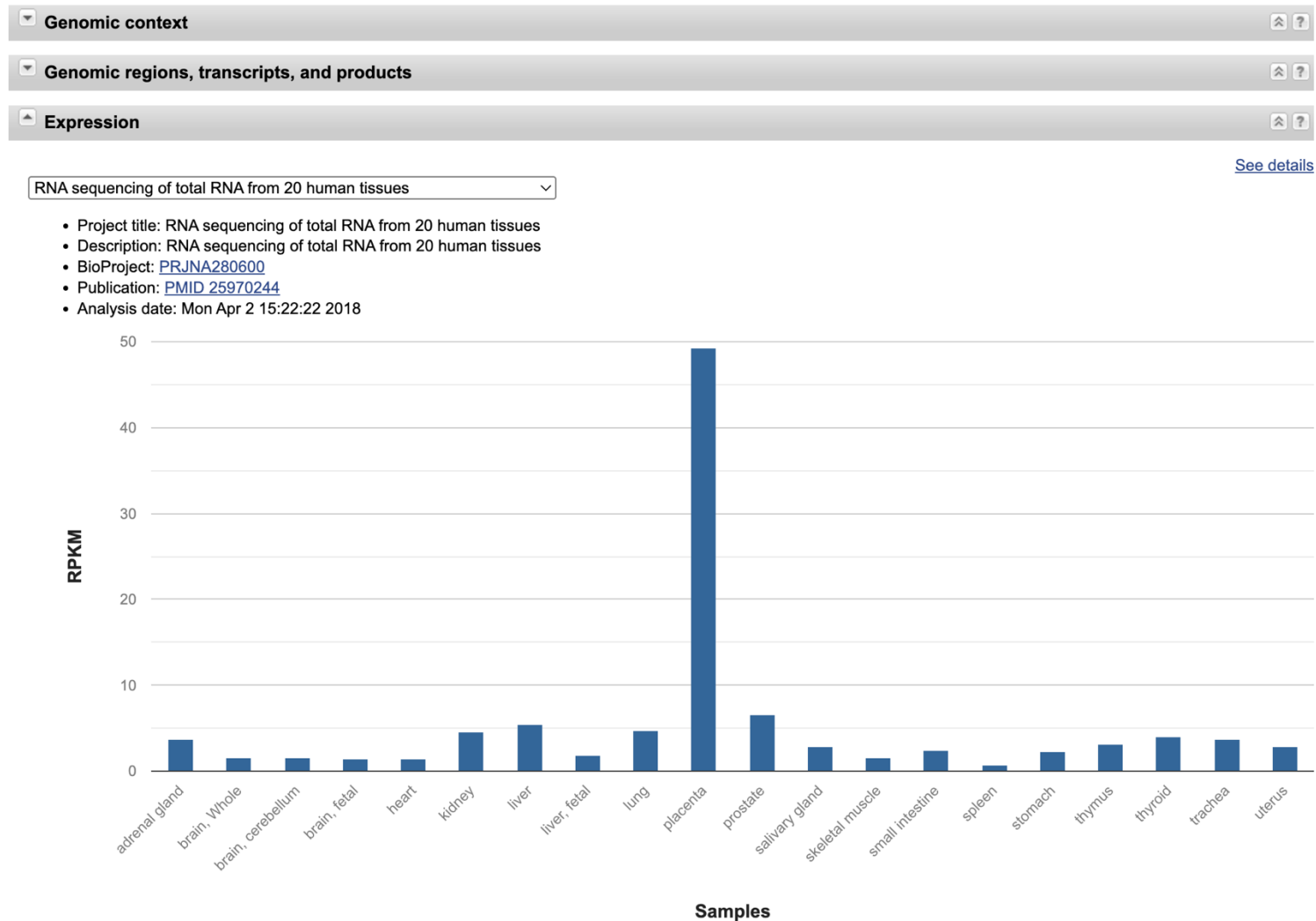
[Run BLAST](#)

[Pick Primers](#)

[Highlight Sequence Features](#)

[Find in this Sequence](#)

Gene expression for EGFR



Bibliography for EGFR

 **National Library of Medicine**
National Center for Biotechnology Information

Log in

Gene

Advanced Help

Full Report ▾ Send to: ▾ Hide sidebar >>

EGFR epidermal growth factor receptor [*Homo sapiens* (human)] Download Datasets

Gene ID: 1956, updated on 7-May-2024

Summary

Genomic context

Genomic regions, transcripts, and products

Bibliography

Related articles in PubMed

- [Differential clinicopathological features, treatments and outcomes in patients with Exon 19 deletion and Exon 21 L858R EGFR mutation-positive adenocarcinoma non-small-cell lung cancer.](#)
Batra U, *et al.* BMJ Open Respir Res, 2023 Jun. PMID 37321664, [Free PMC Article](#)
- [Does EGFR Signaling Mediate Orexin System Activity in Sleep Initiation?](#)
Kniazkina M, *et al.* Int J Mol Sci, 2023 May 30. PMID 37298454, [Free PMC Article](#)
- [The Role of EGFR Amplification in Deep Venous Thrombosis Occurrence in IDH Wild-Type Glioblastoma.](#)
Kaye B, *et al.* Curr Oncol, 2023 May 12. PMID 37232831, [Free PMC Article](#)
- [Advanced Non-small Cell Lung Cancer: EGFR Mutation Analysis Using Pyrosequencing and the Fully Automated qPCR-Based Idylla\(TM\) System.](#)
Boustany Y, *et al.* Cancer Control, 2023 Jan-Dec. PMID 37196225, [Free PMC Article](#)
- [An association of epidermal growth factor receptor mutation subtypes with prognostic prediction and site-specific recurrence in advanced stage lung cancer patients.](#)
Joshi J, *et al.* Mol Biol Rep, 2023 Jun. PMID 37099232

[See all \(6512\) citations in PubMed](#)

Table of contents

Summary

Genomic context

Genomic regions, transcripts, and products

Bibliography

Phenotypes

Variation

HIV-1 interactions

Pathways from PubChem

Interactions

General gene information

Markers, Homology, Gene Ontology

General protein information

NCBI Reference Sequences (RefSeq)

Related sequences

Additional links

Genome Browsers

Variation Viewer (GRCh37.p13)

Variation Viewer (GRCh38)

1000 Genomes Browser (GRCh37.p13)

Ensembl

9.6 - Hersh - AMIA

23

Phenotypes for EGFR

 **National Library of Medicine**
National Center for Biotechnology Information

Log in

Gene

Advanced Help

Full Report ▾ Send to: ▾ Hide sidebar >>

EGFR epidermal growth factor receptor [*Homo sapiens* (human)] Download Datasets

Gene ID: 1956, updated on 7-May-2024

Summary

Genomic context

Genomic regions, transcripts, and products

Bibliography

Phenotypes

[BioGRID CRISPR Screen Phenotypes \(161 hits/1334 screens\)](#)

[Find tests for this gene in the NIH Genetic Testing Registry \(GTR\)](#)

[Review eQTL and phenotype association data in this region using PheGenI](#)

Associated conditions

Description	Tests
Cetuximab response MedGen: C0077967 , GeneReviews: Not available	Compare labs
Inflammatory skin and bowel disease, neonatal, 2 MedGen: C4015130 , OMIM: 616069 , GeneReviews: Not available	Compare labs
Lung cancer MedGen: C0242379 , OMIM: 211980 , GeneReviews: Not available	Compare labs

Table of contents

Summary

Genomic context

Genomic regions, transcripts, and products

Bibliography

Phenotypes

Variation

HIV-1 interactions

Pathways from PubChem

Interactions

General gene information

Markers, Homology, Gene Ontology

General protein information

NCBI Reference Sequences (RefSeq)

Related sequences

Additional links

Genome Browsers

Variation Viewer (GRCh37.p13)

Variation Viewer (GRCh38)

1000 Genomes Browser (GRCh37.p13)

Ensembl

EGFR variants

**National Library of Medicine**
National Center for Biotechnology Information

Log in

Gene

Advanced Help

Full Report ▾ Send to: ▾ Hide sidebar >>

EGFR epidermal growth factor receptor [*Homo sapiens* (human)]
Gene ID: 1956, updated on 7-May-2024

Summary

Genomic context

Genomic regions, transcripts, and products

Bibliography

Phenotypes

Variation

[See variants in ClinVar](#)

[See studies and variants in dbVar](#)

[See Variation Viewer \(GRCh37.p13\)](#)

[See Variation Viewer \(GRCh38\)](#)

Genotypes

[See SNP Geneview Report](#)

[See 1000 Genomes Browser \(GRCh37.p13\)](#)

Table of contents

Summary

Genomic context

Genomic regions, transcripts, and products

Bibliography

Phenotypes

Variation

HIV-1 interactions

Pathways from PubChem

Interactions

General gene information

Markers, Homology, Gene Ontology

General protein information

NCBI Reference Sequences (RefSeq)

Related sequences

Additional links

Genome Browsers

ClinVar for EGFR



National Library of Medicine
National Center for Biotechnology Information

ClinVar

ClinVar

Search ClinVar for gene symbols, HGVS expressions, conditions, and more

Search

Advanced

Help

Home

About

Access

Help

Submit

Statistics

FTP

Classification type

☐ Germline (3,449)

☐ Somatic (32)

Germline classification

☐ Conflicting classifications (154)

☐ Benign (227)

☐ Likely benign (1,429)

☐ Uncertain significance (1,571)

☐ Likely pathogenic (51)

☐ Pathogenic (81)

Types of conflicts

☐ P/LP vs LB/B (2)

☐ P/LP vs VUS (19)

☐ VUS vs LB/B (137)

Molecular consequence

☐ Frameshift (41)

☐ Missense (1,576)

☐ Nonsense (42)

☐ Splice site (41)

☐ ncRNA (221)

☐ Near gene (0)

☐ UTR (71)

Links from Gene

Display options

Sort by Relevance

Download

Items: 1 to 100 of 3468

<< First

< Prev

Page 1 of 35

Next >

Last >>

Variation	Gene (Protein Change)	Type (Consequence)	Condition	Classification, Review status
☐ NM_005228.5(EGFR):c.3350C>A (p.Pro1117His)	EGFR (P1072H +3 more)	Single nucleotide variant (missense variant)	Hereditary cancer-predisposing syndrome	G Uncertain significance ★
☐ NM_005228.5(EGFR):c.3518A>T (p.Gln1173Leu)	EGFR (Q906L +3 more)	Single nucleotide variant (missense variant)	Hereditary cancer-predisposing syndrome	G Uncertain significance ★
☐ NM_005228.5(EGFR):c.3383C>G (p.Pro1128Arg)	EGFR (P1083R +3 more)	Single nucleotide variant (missense variant)	Hereditary cancer-predisposing syndrome	G Uncertain significance ★
☐ NM_005228.5(EGFR):c.2643G>T (p.Met881Ile)	EGFR (M614I +3 more)	Single nucleotide variant (missense variant)	Hereditary cancer-predisposing syndrome	G Uncertain significance ★
☐ NM_005228.5(EGFR):c.2328C>G (p.Arg776=)	EGFR, EGFR-AS1	Single nucleotide variant (non-coding transcript variant +1 more)	Hereditary cancer-predisposing syndrome	G Likely benign ★
☐ NM_005228.5(EGFR):c.3321T>C (p.Asn1107=)	EGFR	Single nucleotide variant (synonymous variant)	Hereditary cancer-predisposing syndrome	G Likely benign ★

EGFR general gene information and Gene Ontology terms

General gene information

Markers

Homology

Gene Ontology

[NCBI Orthologs](#)

[Orthologs from OrthoDB](#)

[Provided by GOA](#)

Items 1 - 25 of 30 < Prev Page 1 of 2 Next >		
Function	Evidence Code	Pubs
enables ATP binding	IEA	
enables ATPase binding	ISS	
enables MAP kinase kinase kinase activity	NAS	PubMed
enables actin filament binding	IDA	PubMed
enables cadherin binding	HDA	PubMed
enables calmodulin binding	IEA	

General protein information

Preferred Names

Names

epidermal growth factor receptor

EGFR vIII
avian erythroblastic leukemia viral (v-erb-b) oncogene homolog
cell growth inhibiting protein 40
cell proliferation-inducing protein 61
epidermal growth factor receptor tyrosine kinase domain
erb-b2 receptor tyrosine kinase 1
proto-oncogene c-ErbB-1
receptor tyrosine-protein kinase erbB-1

NP_001333826.1

EC [2.7.10.1](#)

NP_001333827.1

EC [2.7.10.1](#)

Variation viewer for EGFR

National Library of Medicine
National Center for Biotechnology Information

[Log in](#)

Variation Viewer

[Share this page](#)
[Reset All](#)
[More Info](#)

New to Variation Viewer? [Read our quick overview!](#)

Homo sapiens
(human)

Assembly: GRCh38.p14 (GCF_000001405.40) • Chr 7 (NC_000007.14)

Search assembly

Examples ▶

▶ Pick Assembly
▶ Tracks and User Data
▶ History
▶ Assembly Region Details

NC_000007.14: 54,999,755 - 55,230,888

Region ▼ Gene EGFR Transcript NM_005228.5 Exons: click an exon to zoom in, mouse over to see details

NC_000007.14 ◀ ▶ 🔍 ⏪ ⏩ 🗨️ 📄 🧬 🛠️ Tools ▾ ⚙️ Tracks 📁 Download 📄 ?

55 M 55,020 K 55,040 K 55,060 K 55,080 K 55,100 K 55,120 K 55,140 K 55,160 K 55,180 K 55,200 K 55,220 K

NCBI RefSeq Annotation GCF_000001405.40-RS_2023_10
EGFR (+22) LOC105375284 (+31) EGFR-AS1 NR_047551.1

Clinical, dbSNP b156 v2
Live RefSNPs, dbSNP b156 v2
dbVar Clinical Structural Variants (nstd102)
dbVar Non-Pathogenic Clinical Structural Variants (subset of nstd102)
dbVar Pathogenic Clinical Structural Variants (subset of nstd102)

nsv3909087 (+1)
nsv3919826 (+1)
nsv3888815 (+1)
nsv6314254 (+2)
nsv3917263 (+1)
nsv3924854 (+1)
nsv3828531 (+1)
nsv5381722 (+1) nsv5381722 (+1)
nsv7897379 (+1) nsv6312545 (+1)
nsv3908592 (+2)
nsv6315223 (+1)
nsv4352523 (+1)

OMIM for EGFR in lung cancer

#211980
Table of Contents

- Title
- Phenotype–Gene Relationships
- Clinical Synopsis
- Text
 - Description
 - Clinical Features
 - Inheritance
 - Population Genetics
 - Pathogenesis
 - Clinical Management
 - Mapping
 - Molecular Genetics
 - Cytogenetics

- See Also
- References
- Contributors
- Creation Date
- Edit History

211980

LUNG CANCER

Other entities represented in this entry:

ALVEOLAR CELL CARCINOMA, INCLUDED
ADENOCARCINOMA OF LUNG, INCLUDED
NONSMALL CELL LUNG CANCER, INCLUDED
LUNG CANCER, PROTECTION AGAINST, INCLUDED

Phenotype–Gene Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
1q24.3	{Lung cancer, susceptibility to}	211980	AD, SMu	3	FASLG	134638
2q33.1	{Lung cancer, protection against}	211980	AD, SMu	3	CASP8	601763
3q26.32	Nonsmall cell lung cancer, somatic	211980		3	PIK3CA	171834
5q31.1	Nonsmall cell lung cancer, somatic	211980		3	IRF1	147575
6q26	Adenocarcinoma of lung, somatic	211980		3	PRKN	602544
7p11.2	Nonsmall cell lung cancer, response to tyrosine kinase inhibitor in	211980	AD, SMu	3	EGFR	131550
7p11.2	{Nonsmall cell lung cancer, susceptibility to}	211980	AD, SMu	3	EGFR	131550
7p11.2	Adenocarcinoma of lung, response to tyrosine kinase inhibitor in	211980	AD, SMu	3	EGFR	131550
7q34	Adenocarcinoma of lung, somatic	211980		3	BRAF	164757
7q34	Nonsmall cell lung cancer, somatic	211980		3	BRAF	164757
10p11.23	Lung cancer, somatic	211980		3	MAP3K8	191195
10q11.23	{Lung cancer, susceptibility to}	211980	AD, SMu	3	ERCC6	609413
11p15.4	Lung cancer, somatic	211980		3	SLC22A1L	602631
11q23.1	Lung cancer, somatic	211980		3	PPP2R1B	603113
12p12.1	Lung cancer, somatic	211980		3	KRAS	190070
17q12	Adenocarcinoma of lung, somatic	211980		3	ERBB2	164870
19q13.2	{Lung cancer, resistance to}	211980	AD, SMu	3	CYP2A6	122720

Clinical Synopsis

PheneGene Graphics

?

▼ TEXT

A number sign (#) is used with this entry because mutations in several genes are associated with lung cancer. Both germline and somatic mutations have been identified in the EGFR (131550) and p53 (TP53; 191170) genes, and somatic mutations have been identified in the KRAS (190070), BRAF (164757), ERBB2 (164870), MET (164860), STK11 (602216), PIK3CA (171834), and PARK2 (602544) genes. Amplification of several genes, including EGFR, ERBB2, MET, PIK3CA, and NKX2-1 (600635),

ICD+

▼ External Links

► Protein

▼ Clinical Resources

Clinical Trials

Genetic Alliance

► MedlinePlus Genetics

GTR

► Animal Models



ClinGen (Rehm, 2015; Dalton, 2023)



Get StartedAbout UsCuration ActivitiesWorking GroupsExpert PanelsDocuments & AnnouncementsTools

Gene Symbol

Enter a gene symbol or HGNC ID (Examples: ADNP, HGNC:15766)

Search

Curated Genes

Gene-Disease Validity

Dosage Sensitivity

Clinical Actionability

Curated Variants

Statistics

Downloads

More

?

EGFR

1

Gene-Disease Validity Classifications

2

Dosage Sensitivity Classifications

0

Clinical Actionability Assertions

0

Variant Pathogenicity Assertions

0 / 2

CPIC / PharmGKB High Level Records

Follow Gene

View Gene Facts

Curation Summaries

Status and Future Work0

GenomeConnect3

External Genomic Resources

ClinVar Variants

Gene-Disease Validity

Group By ActivityGroup By Gene-Disease Pair

Dosage Sensitivity

Gene	Disease	Working Group	HI Score & TS Score	Report & Date
EGFR		Dosage Sensitivity WG	0 (No Evidence for Haploinsufficiency)	05/11/2022
			0 (No Evidence for Triplosensitivity)	05/11/2022

Pharmacogenomics - PharmGKB

Gene	Drug	Highest Level of Evidence	Last Curated	Information
EGFR	gefitinib	Level 1A	05/28/2024	View
	erlotinib	Level 2B	03/24/2021	View

Other clinical resources of note

[Clinical Interpretation of Variants in Cancer \(CIViC\)](#)

[Catalogue Of Somatic Mutations In Cancer \(COSMIC\)](#)

[Precision Medicine Knowledgebase \(PMKB\)](#)

[OncoKB Precision Oncology Knowledge Base](#)

[Monarch Initiative](#) – connecting phenotypes and genotypes across species

[Biomedical Data Translator](#) – translate among various data resources and types (Fecho, 2022)

Other integrated multi-omics resources

[MetaGraph](#) – indexing “all life’s” sequences (Karasikov, 2024) and searching them efficiently (Karasikov, 2025)

- “Google for DNA” (Offord, 2024)

Spatial multi-omics integrates with other cellular elements and their spatial locations (Kiessling, 2024)

Allowing development of [Human Cell Atlas](#) (Rood, 2025) requiring many computational methods (Dance, 2024)

- “Google Maps” for the human body or “Wikipedia for cells” (Nature, 2024)

Variety of other clinical issues

We are entering era of “multi-omics” and growing use clinically (Misra, 2018; Subramanian, 2020)

How do we educate clinicians (Rao, 2020; Hull 2020)?

How do we serve the diversity of the world?

- Health and Human Heredity in Africa (H3Africa) creating more representative tools (e.g., chips) (Banks, 2020) and training scientists and informaticians (LeMieux, 2021)

Must also be cognizant of self-reported race and other groupings applied to people based on existing genetic preconceptions (NAM, 2023)

Potential for Gattaca (1997) still exists (Ogbunugafor, 2022; Greenbaum, 2022)

Future directions for research informatics

Quantity of data will continue to grow

Informatics tools will both help manage data as well as generate new knowledge

- Researchers will likely be unable to justify experiments without delving into databases first

As in all areas of computing, user-centered and standards-based approaches are essential while maintaining privacy, ethics, and equity

AI likely to play increasing role