



Databases

Yazdan Asgari

2019

Database Structure

*Primary, secondary keys, relational
databases*

Spreadsheet

- Simple form of a database
- Columns (fields) contain specific info
 - Name, address, date, number
- Rows (records) represent individual entries
 - Individual person, specific DNA sequence
- Each record has unique identifier
 - Primary key unambiguously identifies record
 - Duplicates of primary key not allowed

Election data/Primary Key

<u>cand</u>	<u>cand/year</u>	Candidate	Year	Party	EV	Pop Vote
wjcl	1992dl	William J. Clinton	1992	Democrat	370	44,908,233
ghwb1	1992rl	George H. W. Bush	1992	Republican	168	39,102,282
hrpl	1992il	H. Ross Perot	1996	Indep.	0	19,741,048
wjcl	1996dl	William J. Clinton	1996	Democrat	379	47,401,185
rdl	1996rl	Robert Dole	1996	Republican	159	39,197,469
hrpl	1996rf1	H. Ross Perot	1996	Reform	0	8,085,294
gwbl	2000rl	George W. Bush	2000	Republican	271	50,456,169
agl	2000dl	Al Gore	2000	Democrat	266	50,996,116
rn1	2000gl	Ralph Nader	2000	Green	0	2,695,696
gwbl	2004rl	George W. Bush	2004	Republican	286	62,040,606
jfk2	2004dl	John F. Kerry	2004	Democrat	252	59,028,109
rn1	2004il	Ralph Nader	2004	Independent	0	411,304
bhol	2008dl	Barack Obama	2008	Democrat	365	69,456,897
jsm1	2008rl	John McCain	2008	Republican	173	59,934,814
bhol	2012dl	Barack Obama	2012	Democrat	332	65,917,119
wmr1	2012rl	Willard Mitt Romney	2012	Republican	206	60,932,078
gjl	2012il	Gary Johnson	2012	Libertarian	0	1,275,886
js1	2012gl	Jill Stein	2012	Green	0	469,015

- What is the primary key in this database?
- How does cand/year differ from cand in uniqueness?



Find Movies, TV shows, Celebrities and more...

All

Movies, TV
& Showtimes

Celebs, Events
& Photos

News &
Community

Watchlist

FULL CAST AND CREW

TRIVIA

USER REVIEWS

IMDbPro

MORE

SHARE

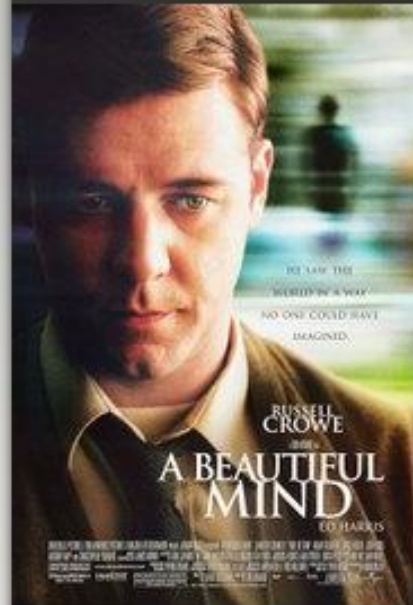


A Beautiful Mind (2001)

★ 8.2
683,141

★ Rate
This

PG-13 | 2h 15min | Biography, Drama | 4 January 2002 (USA)



After [John Nash](#), a brilliant but asocial mathematician, accepts secret work in cryptography, his life takes a turn for the nightmarish.

Director: [Ron Howard](#)

Writers: [Akiva Goldsman](#), [Sylvia Nasar](#) (book)

Stars: [Russell Crowe](#), [Ed Harris](#), [Jennifer Connelly](#) |

[See full cast & crew »](#)

72

Metascore
From [metacritic.com](#)

Reviews
1,200 user | 204 critic



Popularity
582 (▲19)

Oversimplified IMDB

- Primary Key = Movie ID

Movie Title	Year	Movie ID
American Gangster	2007	tt0765429
3:10 to Yuma	2007	tt0381849
A Good Year	2006	tt0401445
Cinderella Man	2005	tt0352248
Master and Commander	2003	tt0311113
A Beautiful Mind	2001	tt0268978
Gladiator	2000	tt0172495
Gladiator	1992	tt0104346

Connected Databases

Movie Title	Year	Movie ID
A Beautiful Mind	2001	tt0268978

Movie ID	Title	Year	Actor	Actor ID
tt0268978	A Beautiful Mind	2001	Russell Crowe	nm0000128
tt0268978	A Beautiful Mind	2001	Ed Harris	nm0000438
tt0268978	A Beautiful Mind	2001	Jennifer Connelly	nm0000124
tt0268978	A Beautiful Mind	2001	Adam Goldberg	nm0004965
tt0268978	A Beautiful Mind	2001	Paul Bettany	nm0079273
tt0268978	A Beautiful Mind	2001	Christopher Plummer	nm0001626

Actor ID	Actor Name	Date of Birth	Country of Birth	Movie Title	Year	Movie ID
nm0000128	Russell Crowe	April 7, 1964	New Zealand	The Nice Guys	2016	tt3799694
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Fathers & Daughters	2015	tt2582502
nm0000128	Russell Crowe	April 7, 1964	New Zealand	The Water Diviner	2014	tt3007512
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Noah	2014	tt1959490
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Winter's Tale	2014	tt1837709
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Man of Steel	2013	tt0770828
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Broken City	2013	tt1235522
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Les Misérables	2012	tt1707386
nm0000128	Russell Crowe	April 7, 1964	New Zealand	The Man with the Iron Fists	2012	tt1258972
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Republic of Doyle	2012	tt1297754
nm0000128	Russell Crowe	April 7, 1964	New Zealand	The Next Three Days	2010	tt1458175
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Robin Hood	2010	tt0955308
nm0000128	Russell Crowe	April 7, 1964	New Zealand	State of Play	2009	tt0473705
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Tenderness	2009	tt0494864
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Body of Lies	2008	tt0758774
nm0000128	Russell Crowe	April 7, 1964	New Zealand	American Gangster	2007	tt0765429
nm0000128	Russell Crowe	April 7, 1964	New Zealand	3:10 to Yuma	2007	tt0381849
nm0000128	Russell Crowe	April 7, 1964	New Zealand	A Good Year	2006	tt0401445
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Cinderella Man	2005	tt0352248
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Master and Commander	2003	tt0311113
nm0000128	Russell Crowe	April 7, 1964	New Zealand	A Beautiful Mind	2001	tt0268978
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Gladiator	2000	tt0172495

Connected Databases

Movie Title	Year	Movie ID
A Beautiful Mind	2001	tt0268978

Primary Key

Secondary Key

Movie ID	Title	Year	Actor	Actor ID
tt0268978	A Beautiful Mind	2001	Russell Crowe	nm0000128
tt0268978	A Beautiful Mind	2001	Ed Harris	nm0000438
tt0268978	A Beautiful Mind	2001	Jennifer Connelly	nm0000124
tt0268978	A Beautiful Mind	2001	Adam Goldberg	nm0004965
tt0268978	A Beautiful Mind	2001	Paul Bettany	nm0079273
tt0268978	A Beautiful Mind	2001	Christopher Plummer	nm0001626

Actor ID	Actor Name	Date of Birth	Country of Birth	Movie Title	Year	Movie ID
nm0000128	Russell Crowe	April 7, 1964	New Zealand	The Nice Guys	2016	tt3799694
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Fathers & Daughters	2015	tt2582502
nm0000128	Russell Crowe	April 7, 1964	New Zealand	The Water Diviner	2014	tt3007512
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Noah	2014	tt1959490
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Winter's Tale	2014	tt1837709
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Man of Steel	2013	tt0770828
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Broken City	2013	tt1235522
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Les Misérables	2012	tt1707386
nm0000128	Russell Crowe	April 7, 1964	New Zealand	The Man with the Iron Fists	2012	tt1258972
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Republic of Doyle	2012	tt1297754
nm0000128	Russell Crowe	April 7, 1964	New Zealand	The Next Three Days	2010	tt1458175
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Robin Hood	2010	tt0955308
nm0000128	Russell Crowe	April 7, 1964	New Zealand	State of Play	2009	tt0473705
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Tenderness	2009	tt0494864
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Body of Lies	2008	tt0758774
nm0000128	Russell Crowe	April 7, 1964	New Zealand	American Gangster	2007	tt0765429
nm0000128	Russell Crowe	April 7, 1964	New Zealand	3:10 to Yuma	2007	tt0381849
nm0000128	Russell Crowe	April 7, 1964	New Zealand	A Good Year	2006	tt0401445
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Cinderella Man	2005	tt0352248
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Master and Commander	2003	tt0311113
nm0000128	Russell Crowe	April 7, 1964	New Zealand	A Beautiful Mind	2001	tt0268978
nm0000128	Russell Crowe	April 7, 1964	New Zealand	Gladiator	2000	tt0172495

Relational Database

- Avoid too many fields
- Databases link to other database
 - Each database called 'relation'
- Secondary key in one database can serve as primary key in another database
- Internet: large relational database
 - Made up of webpages (records)
 - URLs represent primary keys in smaller databases
 - other websites

Summary

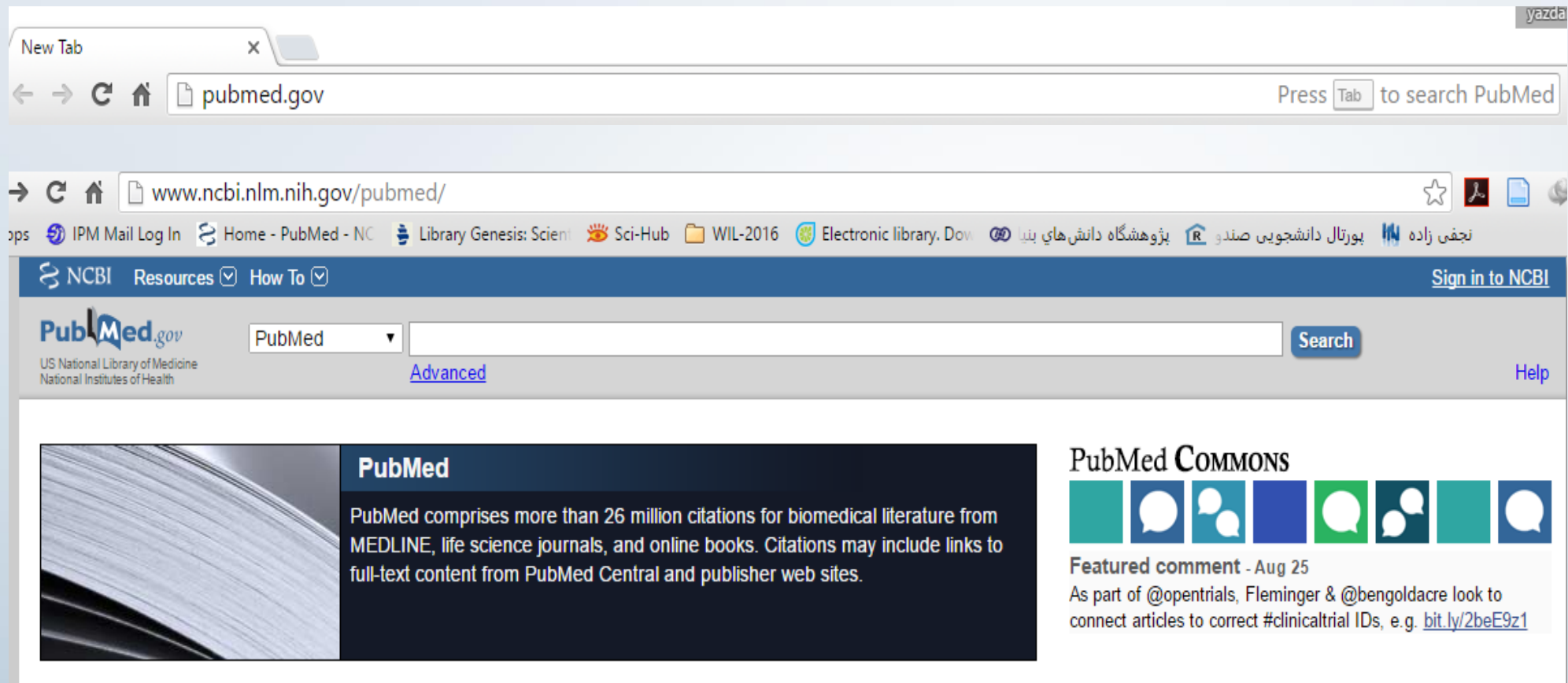
- Recognize primary keys
 - Most unique identifier for a record
 - Usually series of numbers/letters
- Relational databases link related databases
 - IMDB: movie DB and actor DB
 - NCBI: nucleotide DB and protein DB



PubMed

NCBI

- National Center for Biotechnology Education
 - <http://www.ncbi.nlm.nih.gov>
- Large government relational database at National Institutes of Health
- Includes scientific journals, DNA sequences, protein sequences, protein structures



Searching PubMed:

- Kauffman
- Kauffman S
- Kauffman SA

Searching PubMed:

- Kauffman
- Kauffman [author]
- Kauffman [au]
- kauffman AND origin [title]



MeSH

CONTROLLED VOCABULARY

NATIONAL LIBRARY OF MEDICINE

ALLOWS FOR CONSISTENT, COMPLETE SEARCHES

TUTORIAL

[HTTP://WWW.NLM.NIH.GOV/bsd/viewlet/
MESH/SEARCHING/MESH1.HTML](http://www.nlm.nih.gov/bsd/viewlet/mesh/searching/mesh1.html)

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

MeSH MeSH ▾ [Search](#)

[Limits](#) [Advanced](#) [Help](#)

Full ▾ [Send to: ▾](#)

RNA, Small Interfering

Small double-stranded, non-protein coding RNAs (21-31 nucleotides) involved in GENE SILENCING functions, especially RNA INTERFERENCE (RNAi). Endogenously, siRNAs are generated from dsRNAs (RNA, DOUBLE-STRANDED) by the same ribonuclease, Dicer, that generates miRNAs (MICRORNAS). The perfect match of the siRNAs' antisense strand to their target RNAs mediates RNAi by siRNA-guided RNA cleavage. siRNAs fall into different classes including trans-acting siRNA (tasiRNA), repeat-associated RNA (rasiRNA), small-scan RNA (scnRNA), and Piwi protein-interacting RNA (piRNA) and have different specific gene silencing functions.

Year introduced: 2003

[PubMed search builder options](#)
[Subheadings:](#)

[PubMed Search Builder](#)

"RNA, Small Interfering"[Mesh]

[Add to search builder](#) [AND ▾](#)

[Search PubMed](#)

[YouTube](#) [Tutorial](#)

Searching PubMed:

- RNAi
- RNAi AND mello [au]
- "RNA, Small Interfering"[Mesh] mello [au]

NCBI Resources How To Sign in to NCBI

PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

PubMed

[Help](#)

[Show additional filters](#)

Article types
Review
More

Text availability
Abstract available
Free full text available
Full text available

Publication dates
5 years
10 years
Custom range

Species
Humans
Other Animals

[Clear all](#)

[Show additional filters](#)

Display Settings: ☒ Summary, 20 per page, Sorted by Recently Added

Send to: ☐ [Manage Filters](#)

Results: 1 to 20 of 26

1. [The C. elegans CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression.](#)
Seth M, Shirayama M, Gu W, Ishidate T, Conte D Jr, Mello CC.
Dev Cell. 2013 Dec 23;27(6):656-63.
PMID: 24360702 [PubMed - indexed for MEDLINE]
[Related citations](#)

2. [Methylated Alanine-Rich protein Kinase C Substrate \(MARCKS\) expression modulates the metastatic phenotype in human and murine colon carcinoma in vitro and in vivo.](#)
Rombouts K, Carloni V, Mello T, Omenetti S, Galastri S, Madiai S, Gali A, Pinzani M.
Cancer Lett. 2013 Jun 10;333(2):244-52. doi: 10.1016/j.canlet.2013.01.040. Epub 2013 Jan 29.
PMID: 23378641 [PubMed - indexed for MEDLINE]
[Related citations](#)

3. [CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. elegans piRNA precursors.](#)
Gu W, Lee HC, Chaves D, Youngman EM, Pazour GJ, Conte D Jr, Mello CC.
Cell. 2012 Dec 21;151(7):1489-500. doi: 10.1016/j.cell.2012.11.023.

16 free full-text articles in PubMed Central
CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. [Cell. 2012]
Protection from feed-forward amplification in an amplified RNAi mechanism. [Cell. 2012]
PCA3 noncoding RNA is involved in the control of prostate-cancer cell survival. [BMC Cancer. 2012]
[See all \(16\)...](#)

Got a paper in PubMed?

*** Be careful when using filters:**
they stay on for future searches!

[Search details](#)

NCBI Resources How To Sign in to NCBI

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed kinetic* [title] AND modelling [ti] Search

Create RSS Create alert Advanced Help

Article types
Clinical Trial
Review
Customize ...

Text availability
Abstract
Free full text
Full text

PubMed Commons
Reader comments
Trending articles

Publication dates
5 years
10 years

Format: Summary Sort by: Most Recent Send to Filters: Manage Filters

Search results
Items: 1 to 20 of 404

<< First < Prev Page 1 of 21 Next > Last >>

1. [Some aspects on kinetic modeling of evacuation dynamics: Comment on "Human behaviours in evacuation crowd dynamics: From modelling to "big data" toward crisis management" by Nicola Bellomo et al.](#)
Calvo J, Nieto J.
Phys Life Rev. 2016 Aug 24. pii: S1571-0645(16)30081-1. doi: 10.1016/j.plrev.2016.08.008. [Epub ahead of print] No abstract available.
PMID: 27567501
[Similar articles](#)

Find related data
Database: Select
Find items

Search details
(kinetic[title] OR kinetica[title] OR kinetical[title] OR kinetically[title] OR kineticdb[title] OR kinetics[title] OR kineticist's[title] OR

Searching PubMed:

- kinetic* [title] AND modelling [ti]

DATE FIELDS [EDAT] & [PDAT]

YEAR (YYYY) -EG 2014, 1998

MONTH (YYYY/MM) -EG 2014/03

DATE (YYYY/MM/DD) -EG 2014/03/17

RANGE OF DATES

1998/02:2013/09/27

TRANSLATES TO

FEB. 1, 1998 THROUGH SEP. 27, 2013

ALTERNATE QUERY

"LAST 19 DAYS"[EDAT]

Searching PubMed:

- "last 10 days" [edat] AND nature [journal]
- "last 10 days" [edat] AND nature [journal] AND DNA
- "last 10 days" [edat] AND science [journal] AND DNA

PubMed, MEDLINE, and PMC

PubMed citations come from

- 1) MEDLINE indexed journals,
- 2) journals/manuscripts deposited in PMC,
- 3) NCBI Bookshelf.

Both MEDLINE and other PubMed citations may have links to full-text articles or manuscripts in PMC, NCBI Bookshelf, and publishers' Web sites.

If you limit your PubMed search to MeSH controlled vocabulary or the MEDLINE subset, you will see only MEDLINE citations in your results.



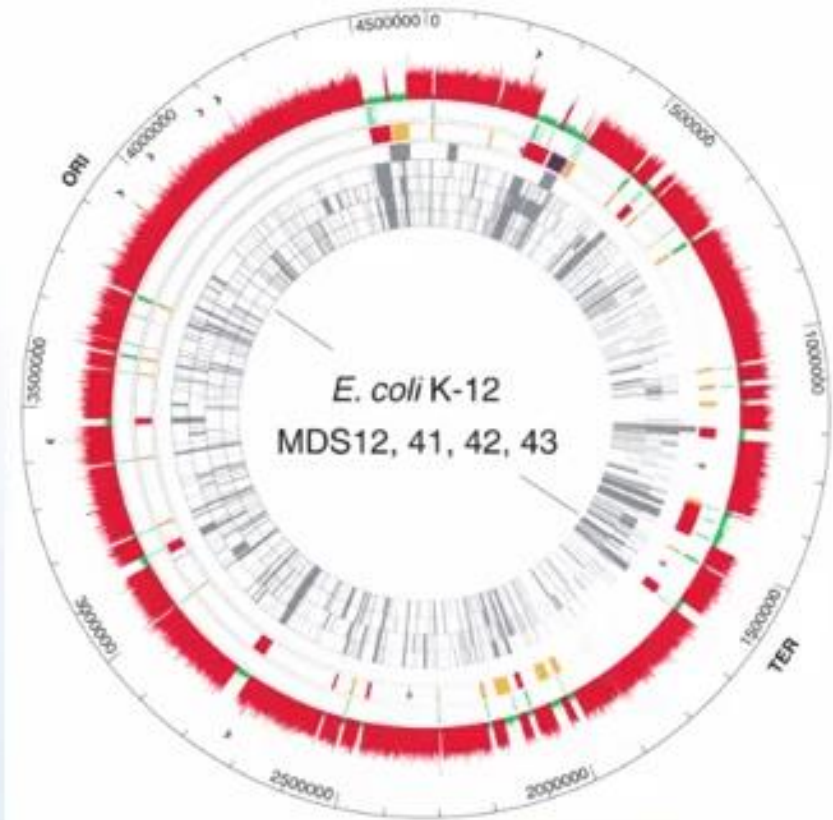
Genome

Typical Prokaryotic Genome

- ✓ Prokaryotes are microscopic organisms
- ✓ They have a circular genome
- ✓ Its length is a few million bp (0.6 – 10 Mb)
- ✓ Prokaryotes have about 1 gene per Kb
- ✓ 70 % of their genome is coding for proteins
- ✓ Their genes do not overlap

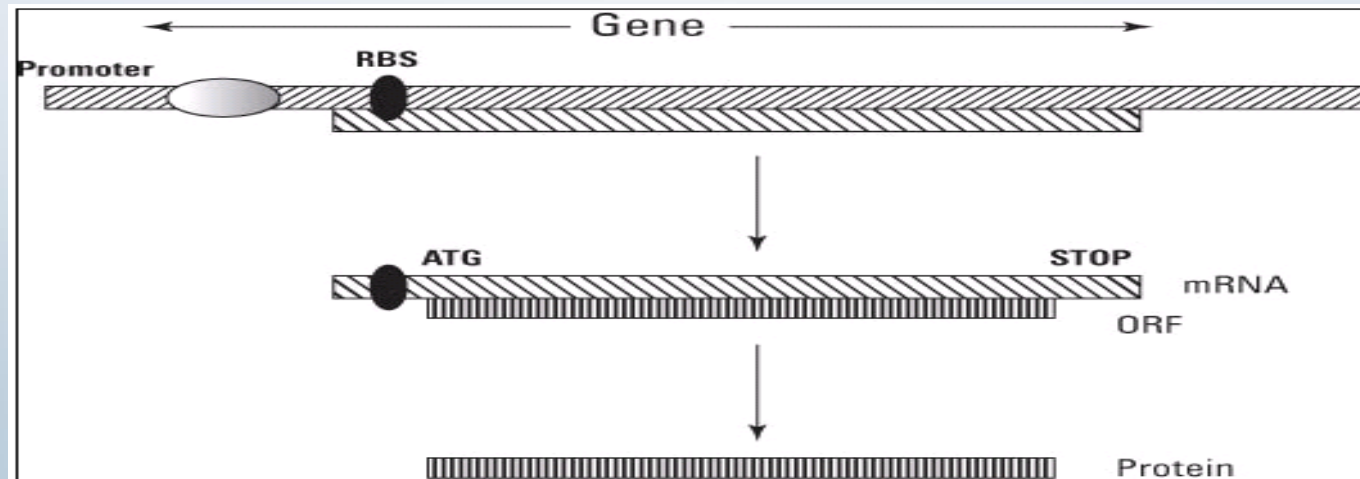
E. coli Genome

- Note position 0 at top of circle
 - Base pairs numbered in clockwise fashion
 - Position 0 represents origin of replication
- + strand genes run clockwise
- - (complementary) strand genes run counterclockwise



Typical Prokaryotic Protein-Coding Gene

- The gene has an uninterrupted sequence
- Prokaryotic mRNA contains
 - The Ribosome Binding Site (RBS)
 - The Open Reading Frame (ORF) **in one piece**
 - In operons, the RNA can contain several ORFs

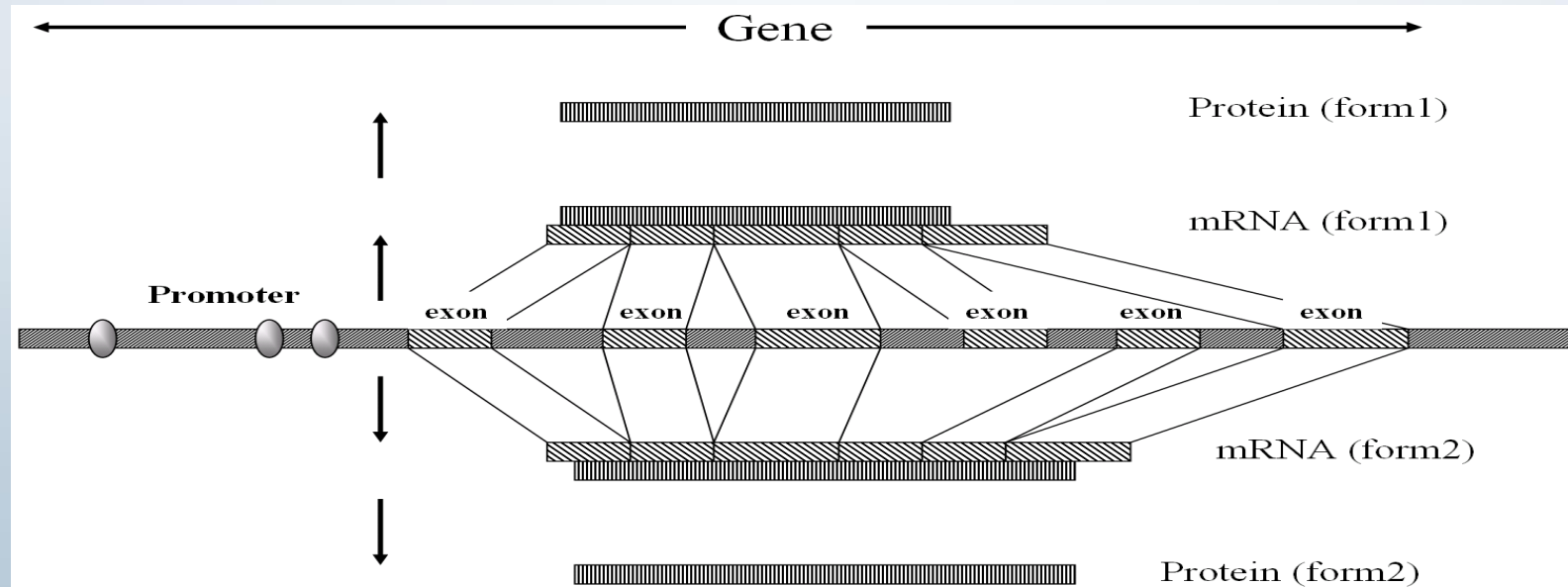


Typical Eukaryotic Genome

- Eukaryotes can be small (yeast) or big (whales)
- Genomes are made of linear pieces of DNA called chromosomes
- One chromosome: 10 to 700 Mb
- The Human Genome
 - Contains 22+1 chromosomes
 - Is 3 Gb long
- One gene every 100 Kb (human)
- 5 % of the genome is coding for proteins

Typical Eukaryotic Protein-Coding Gene

- The coding sequences are made of coding **exons** separated by **introns**
- Introns are spliced out and exons glued together to make the ORF
- One gene can code for several alternative proteins: **alternative splicing**



Prokaryotes vs. Eukaryotes

- Prokaryotes

- Genome=one large circular chromosome + a few small circular chromosomes (plasmids)
- 0.5 to 8 Mb / chromosome
- Genes in one piece
- 70% of the genome is coding
- 1 gene / Kb

- Eukaryotes

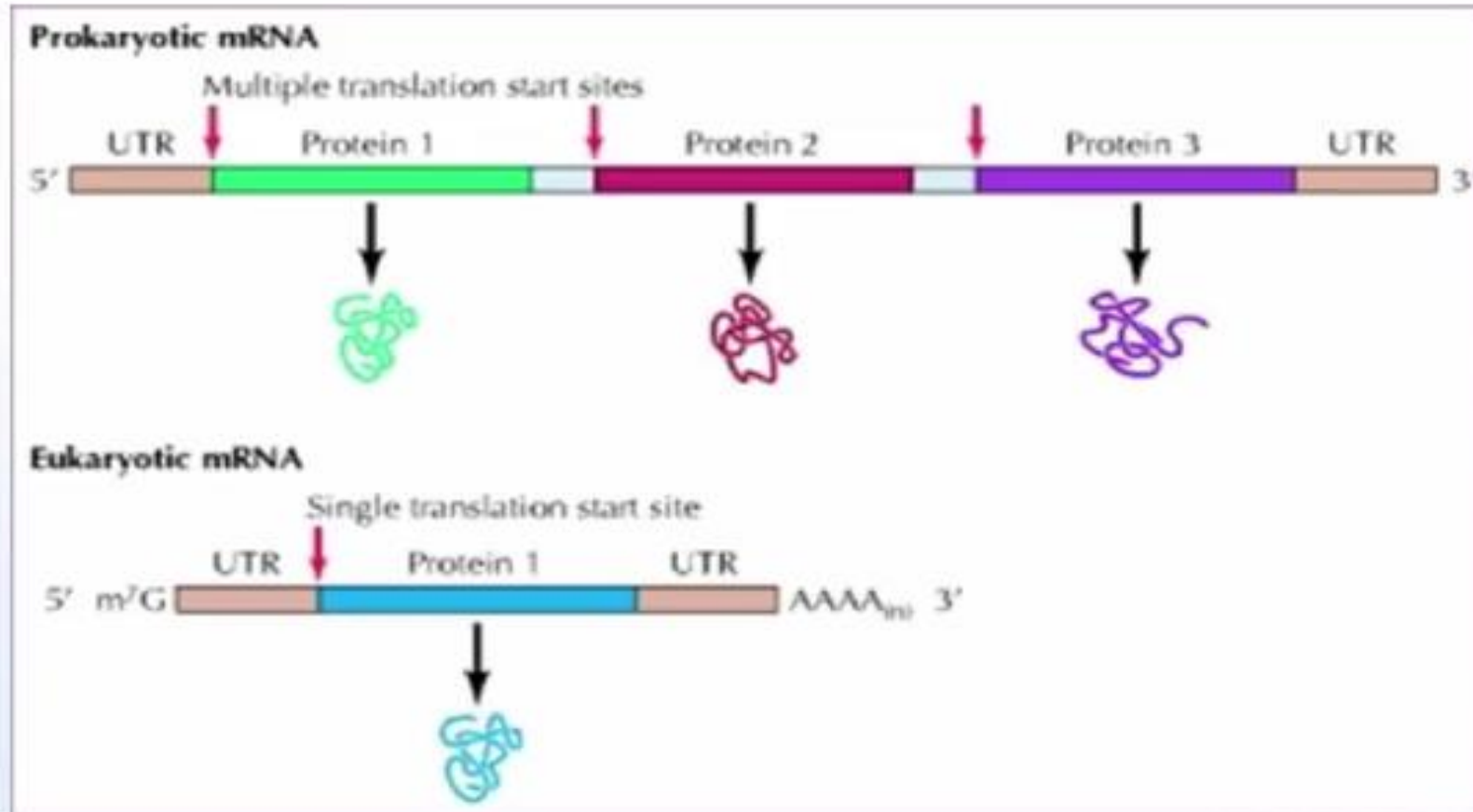
- Genome= many large linear chromosomes
- 10 to 700 Mb / chromosome
- Genes split
- 5% of the genome is coding
- 1 gene/ 100 Kb (Human)

Some Definitions

Polycistronic mRNA

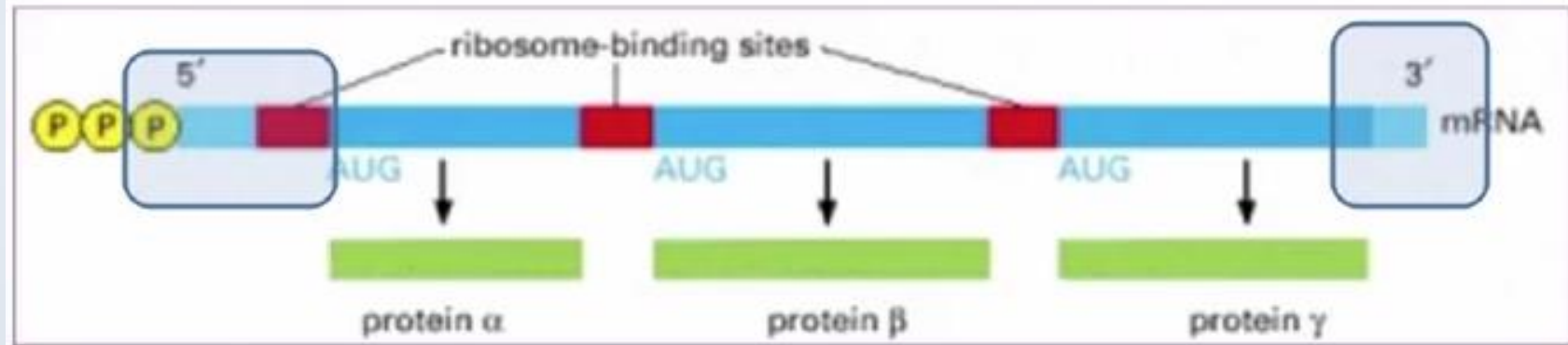
- Eukaryotic transcripts usually have only one coding region
 - One mRNA makes one protein
 - Actually many molecules of same single protein
- Prokaryotic transcripts often have more than one coding region
 - One mRNA make multiple proteins
 - Actually many molecules of each protein

Comparison



Some Definitions

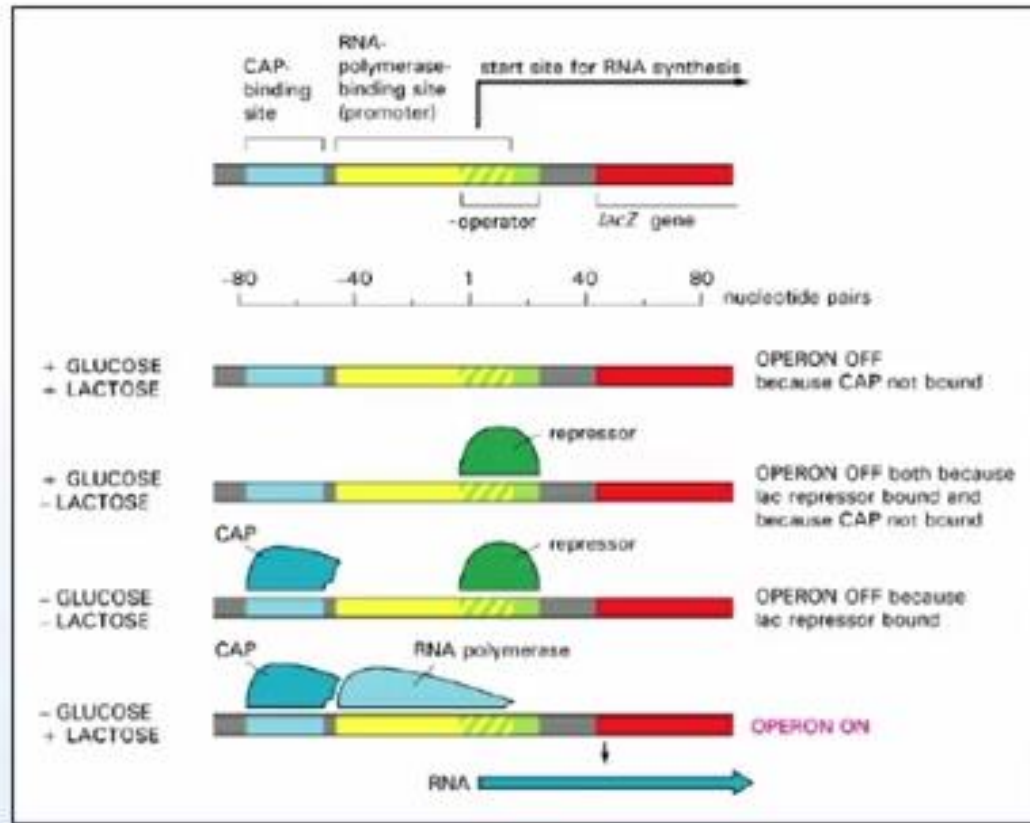
UTRs



- Untranslated region before first start codon
– 5' UTR
- Another after last stop codon
– 3' UTR

Some Definitions

Lac operon



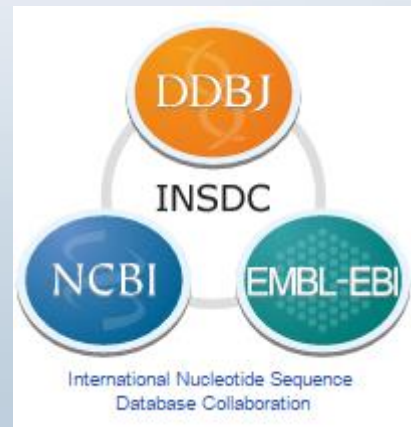
- Glucose preferred energy source
- Lactose used in absence of glucose
- +1 = transcription start site
 - -40 = 40 bases “upstream”
 - +80 = 80 bases “downstream”



DNA Databases

Most Important DNA Databases

- GenBank (NCBI): <http://www.ncbi.nlm.nih.gov/genbank/>
- EMBL-EBI (Ensembl): <http://www.ebi.ac.uk/services>
- DNA Data Bank of Japan (DDBJ): <http://www.ddbj.nig.ac.jp/>





International Nucleotide Sequence Database Collaboration

- The International Nucleotide Sequence Database Collaboration (INSDC) is a long-standing foundational initiative that operates between [DDBJ](#), [EMBL-EBI](#) and [NCBI](#). INSDC covers the spectrum of data raw reads, though alignments and assemblies to functional annotation, enriched with contextual information relating to samples and experimental configurations.

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



Looking for DNA Sequences

GenBank (NCBI)

GenBank

- Housed by the National Center for Biotechnologies (NCBI)
- GenBank is the memory of biological science
- Contains **EVERY** DNA sequence ever published
- GenBank is the original information source for most biological databases
- GenBank is more complicated to use than gene-centric databases



NCBI Nucleotide Search

Example: *BRCA1*

Species

Animals (29,334)
Plants (4,912)
Fungi (9,124)
Protists (2,929)
Bacteria (140,758)
Archaea (310)
Viruses (158)
Customize ...

Molecule types

genomic DNA/RNA (168,542)
mRNA (19,637)
Customize ...

Source databases

INSDC (GenBank) (88,845)
RefSeq (99,930)
Customize ...

Genetic compartments

Mitochondrion (1)
Plasmid (195)

Sequence length

Custom range...

Release date

Custom range...

Revision date

Custom range...

Summary ▾ 20 per page ▾ Sort by Default order ▾

Send to: ▾

See [BRCA1 BRCA1. DNA repair associated](#) in the Gene database
brca1 reference sequences [Genomic \(1\)](#) [Transcript \(6\)](#) [Protein \(5\)](#)

Items: 1 to 20 of 188854

<< First < Prev Page 1 of 9443 Next > Last >>

i Found 189427 nucleotide sequences. Nucleotide (188854) EST ([381](#)) GSS ([192](#))

☐ [Pan troglodytes BRCA1 \(BRCA1\) gene, complete cds](#)

1. 80,966 bp linear DNA
Accession: AY365046.1 GI: 37953274
[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Macropus rufus BRCA1 \(BRCA1\) gene, partial cds](#)

2. 698 bp linear DNA
Accession: AY211956.1 GI: 28199994
[GenBank](#) [FASTA](#) [Graphics](#)

Custom range

7224 to 7225

Apply

Clear

4. 750 bp linear DNA
Accession: AY211954.1 GI: 28199990
[GenBank](#) [FASTA](#) [Graphics](#)

[Resources](#)
[How To](#)

Nucleotide

NCBI is phasing out sequence GI numbers in September 2016. Please use accession.version! [Read more...](#)

GenBank
Send:

Homo sapiens BRCA1, DNA repair associated (BRCA1), transcript variant 1, mRNA

NCBI Reference Sequence: NM_007294.3

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

Go to: ☐

LOCUS	NM_007294	7224 bp	mRNA	linear	PRI 03-SEP-2016
DEFINITION	Homo sapiens BRCA1, DNA repair associated (BRCA1), transcript variant 1, mRNA.				
ACCESSION	NM_007294				
VERSION	NM_007294.3 GI:237757283				
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 7224)				
AUTHORS	Orban TI and Olah E.				
TITLE	Emerging roles of BRCA1 alternative splicing				
JOURNAL	MP, Mol. Pathol. 56 (4), 191-197 (2003)				
PUBMED	12890739				
REMARK	Review article				

exon

PPP1R53; PSCP; RNF53"
/note="upstream in-frame stop codon"
214..312

CDS

/gene="BRCA1"
/gene_synonym="BRCAI; BRCC1; BROVCA1; FANCS; IRIS; PNCA4;
PPP1R53; PSCP; RNF53"
/inference="alignment:Splign:1.39.8"
233..5824
/gene="BRCA1"
/gene_synonym="BRCAI; BRCC1; BROVCA1; FANCS; IRIS; PNCA4;
PPP1R53; PSCP; RNF53"
/note="isoform 1 is encoded by transcript variant 1;
BRCA1/BRCA2-containing complex, subunit 1; RING finger
protein 53; breast cancer type 1 susceptibility protein;
breast and ovarian cancer susceptibility protein 1;
protein phosphatase 1, regulatory subunit 53; Fanconi
anemia, complementation group 5; early onset breast cancer
1; breast cancer 1, early onset"
/codon_start=1
/product="breast cancer type 1 susceptibility protein
isoform 1"
/protein_id="[NP_009225.1](#)"
/db_xref="GI:6552299"
/db_xref="CCDS:[CCDS11453.1](#)"
/db_xref="GeneID:672"
/db_xref="HGNC:[HGNC:1100](#)"
/db_xref="HPRD:[00218](#)"
/db_xref="MIM:[113705](#)"
/translation="MDLSALRVEEVONVINAMOKILECPICLELIKEPVSTKCDHIFC

Assembling CDSs from a GenBank Entry

- The **gene**, **mRNA**, and **CDS** sections tell you which segments of which entry must be joined to reconstruct the gene, the mRNA, or the CDS

<u>gene</u>	<pre>order(AF018429.1:<1..1735,1..1177,AF018431.1:1..45, AF018432.1:658..732,AF018432.1:884..954, AF018432.1:1391..>1447) /gene="DUT"</pre>
<u>mRNA</u>	<pre>join(AF018429.1:<282..561,AF018429.1:1034..1172, 560..651,AF018431.1:1..45,AF018432.1:658..732, AF018432.1:884..954,AF018432.1:1391..>1447) /gene="DUT" /product="dUTPase" /note="alternatively spliced; encodes mitochondrial form of the protein"</pre>
<u>CDS</u>	<pre>join(AF018429.1:282..561,AF018429.1:1034..1172,560..651, AF018431.1:1..45,AF018432.1:658..732,AF018432.1:884..954, AF018432.1:1391..1447)</pre>

polyA site

ORIGIN

PPP1R53; PSCP; RNF53"

7207

/gene="BRCA1"

/gene_synonym="BRCAI; BRCC1; BROVCA1; FANCS; IRIS; PNCA4; PPP1R53; PSCP; RNF53"

```

1 gtacctgat ttcgtattc gagaggctg tgcttagcgg tagcccttg gtttccgtg
61 caacggaaaa gcgcgggaat tacagataa ttaaaactgc gactgcgcgg cgtgagctcg
121 ctgagacttc ctggacgggg gacaggctgt ggggtttctc agataactgg gcccttcgac
181 tcaggaggcc ttcacctctc gctctgggta aagttcattg gaacagaaag aaatggattt
241 atctgctctt cgcgttgagg aagtacaaaa tgcattaat gctatgcaga aaatcttaga
301 gtgtcccatc tgtctggagt tgatcaagga acctgtctcc acaaagtgtg accacatatt
361 ttgcaaat ttgcatgctga aacttctcaa ccagaagaaa gggccttcac agtgcctttt
421 atgtaagaat gatataacca aaaggagcct acaagaaagt acgagattta gtcaacttgt
481 tgaagagcta ttgaaatca tttgtgcttt tcagcttgac acagggttgg agtatgcaa
541 cagctataat ttgcaaaaa aggaataaa ctctcctgaa catctaaaag atgaagtctt
601 tatcatccaa agtatgggct acagaaccg tgccaaaaga ctctacaga gtgaaccgga
661 aaatccttcc ttgaggaaa ccagctcag tgcctcaact tctaacttg gaactgtgag
721 aactctgagg acaaagcagc ggatacaacc tcaaaagacg tctgtctaca ttgaattggg
781 atctgattct tctgaagata ccgttaataa ggcaacttat tgcaagtgtg gagatcaaga
841 attgttaca atcacccctc aaggaaccag ggatgaaatc agtttggatt ctgcaaaaaa
901 ggctgcttgt gaattttctg agacggatgt aacaaatct gaacatcatc aaccagtaa
961 taatgattt aacaccactg agaagcgtgc agctgagagg catccagaaa agtatcaggg
1021 tagttctgtt tcaaaactgc atgtggagcc atgtggcaca aatactcatg ccagctcatt
1081 acagcatgag aacagcagtt tattactcac taaagacaga atgaatgtag aaaaggctga
1141 attctgtaat aaaagcaac agcctggctt agcaaggagc caacataaca gatgggctgg
1201 aagtaaggaa acatgtaatg ataggcggac tccagcaca gaaaaaaagg tagatctgaa
1261 tgctgatccc ctgtgtgaga gaaaagaatg gaataagcag aaactgccat gctcagagaa
1321 tcctagagat actgaagatg ttccttgat aactactaat agcagcattc agaaagttaa
1381 tgagtgggtt tccagaagtg atgaactgtt aggttctgat gactcacatg atggggagtc
1441 tgaatcaaa gccaaagtag ctgatgtatt ggacgttcta aatgaggtag atgaatattc
1501 tggttcttca gaaaaatag acttactggc cagtgtacct catgaggcct taatatgtaa
1561 aagtgaaga gttcaactca aatcagtaga gagtaattt gaagacaaaa tatttgggaa
1621 aacctatcg aagaaggcaa gcctcccaa cttaagcat gtaactgaaa atctaattat
1681 aggagcattt gttactgagc cacagataat acaagagcgt cccctcaca ataaattaaa
1741 gcgtaaaagg agacctacat caggccttca tctgaggat tttatcaaga aagcagattt
1801 ggcattcaa agactctg aatgataa tcaaggact aacaaaggc agcagatg

```

CDS

Feature

1 of 1

NM_007294 : 1 segment

Details

Display: FASTA GenBank Help

/db_xref="HPRD: 00218 "

/db_xref="MIM: 113705 "

/translation="MDLSALRVEEVQNVINAMQKILECPICLLEIKPEVSTKCDHIFC
KFCMLKLLNQKGPSQCPLKNDITKRSLQESTRFSQLVEELLKICAFQLDTGLEYA
NSYNFAKKENNSPEHLKDEVSIQSMGYRNRKRLQSEPNPSLQETSLSVQLSNLG
TVRTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATYCSVGDQELLQITPQGTREISL
DSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPEKYQGSVSNLHVEPCGT
NTHASSLQHENSLLLT KDRMNVKAEFCNKSKQGLARSQHNRWAGSKETCNDRRTP
STEKKVDLNADPLCERKEWIKQLPCSENPRDTEVPWITLNSISQKVNWFNRSDLE
LGSDSDSHGGESENAKVADVLVDLNEVDEYSGSSEKIDLLASDPHEALICKSERVHSK
SVESNIEDKIFGKTYRKKASLPNLSHVTENLIIGAFVTEPQIIQERPLTNLKRKRPP
TSGLHPEDFIKKADLAVQKTPMINQGTNQTEQNGQVMNITNSGHENKTGDSIQNEK
NPNPIESLEKESAFKTKAEPISSSISNMELELNHNSKAPKNNLRKRKSSRTHIALE
LVVSRNLSPPNCTELQIDSCSSSEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKKSN
KPNQTSKRHSDSTFPELKL TNAPGSFTKCSNTSELKEFVNPSLPREEKEKLETVKV
SNNAEDPKDMLSGERVLTQTERSVESSSISLVPGTGYGTQESISLLEVSTLGAKTEP
NKCVSQCAAFENPKGLIHGSKDNDRDTEGFKYPLGHEVNHRSRETSIEMEESLDAQY
LQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKKQSPKVTFECEQKEENQGKNE
SNIKPVQTVNITAGFPVVGQDKPVDNAKCSIKGGSRFCLSSQFRGNETGLITPNKHG
LLQNYPYRIPPLPIKSFVKTKCKNLLLEENFEHSMSPEREMGNENIPSTVSTISRNN
IRENVFKEASSNINEVGSSSTNEVGSSSINEIGSSDENIQAE LGRNRGPKLNAHLRLGV
LQPEVYKQSLPGSNCKHPKIKKQYEEVVQTVNTDFSPYLISDNLEQPMGSSHASQVC
SETPDLLDDGEIKEDTSFAENDIKESSAVFSKSVQKGLSRSPSPFTHTLAQGYRR
GAKKLESSEENLSEDEELPCFQHL LFGKVMNIPSQSTRHSTVATECLSKNTEENLLS
LKNSLNDCSNQVILAKASQEHHLSEETKCSALFSSQCELEDTANTNTQDPFLIGS
SKQMRHQSESQGVGLSDKELVSDDEERGTLGLENQEEQSMDSNLGEAASGCESETSV
SEDCSGLSSQSDILTTQQRDTMQHNLIKLQQEMAELEAVLEQHGSQPSNSYPSIISDS
SALEDLRNPEQSTSEKAVLTSQKSSEYPSIQNPEGLSADKFEVSADSSSKNKEPGVE
RSSPSKCPSLDDRWMHSCSGSLQNRNYPQEEELIKVVDVEEQLEESGPHDLTETSY
LPRQDLEGTPLYESGISLFSDDPEDSPSEDRAPE SARVGNIPSSSALKVPQLKVAES
AQSPAAAHTDTAGYNAMESVSREKPELTASTERNVNRHSMVVSGLTPEEFMLVVKF
ARKHHITLNLITEETTHVMKTD AEFVCERTLKYFLGIAGGKMWVSFWNTQSIKER
KMLNEHDFEVRGDVVNGRHHQGPKRARESDRKIFRGL EICCYGPFNTMPDQLEHMMV
QLCGASVVKELSSFTLGTGVHPIVVVQPDWATEDNGFHAIGQNCAPVVTREWLDSV
ALYQCQELDTYLIPQIPHSY"

FASTA ▾

Format

- ☐ Summary
- ☐ GenBank
- ☐ GenBank (full)
- ☒ FASTA
- ☐ FASTA (text)
- ☐ Graphics
- ☐ ASN.1
- ☐ Revision History
- ☐ Accession List
- ☐ GI List

BRCA1, DNA repair associated (BRCA1), transcript variant 1,

Accession: NM_007294.3

NM_007294.3 | Homo sapiens BRCA1, DNA repair associated (BRCA1),
mRNA

```
AGAGGCTGCTGCTTAGCGGTAGCCCCTTGGTTTCCGTGGCAACGGAAAA
TAAACTGCGACTGCGCGGCGTGAGCTCGCTGAGACTTCCTGGACGGGG
GATAACTGGGCCCTGCGCTCAGGAGGCCTTACCCTCTGCTCTGGGTA
AATGGATTTATCTGCTCTTCGCGTTGAAGAAGTACAAAATGTCATTAAT
GCTATGCAGAAAATCTTAGAGTGTCCCATCTGTCTGGAGTTGATCAAGGAACCTGTCTCCACAAAGTGTG
ACCACATATTTTGCAAATTTGCATGCTGAAACTTCTCAACCAGAAGAAAGGGCCTTCACAGTGTCTTT
ATGTAAGAATGATATAACCAAAAGGAGCCTACAAGAAAGTACGAGATTTAGTCAACTTGTTGAAGAGCTA
TTGAAAATCATTTGTGCTTTTCAGCTTGACACAGGTTTGGAGTATGCAAACAGCTATAATTTTGCAAAAA
AGGAAAATAACTCTCCTGAACATCTAAAAGATGAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCG
TGCCAAAAGACTTCTACAGAGTGAACCCGAAAATCCTTCCTTGCAAGAAACCAGTCTCAGTGTCCAACCTC
TCTAACCTTGGAAGTGTGAGAACTCTGAGGACAAAGCAGCGGATACAACCTCAAAAGACGTCTGTCTACA
TTGAATTGGGATCTGATTCTTCTGAAGATACCGTTAATAAGGCAACTTATTGCAGTGTGGGAGATCAAGA
ATTGTTACAAATCACCCCTCAAGGAACCAGGGATGAAATCAGTTTGGATTCTGCAAAAAAGGCTGCTTGT
GAATTTTCTGAGACGGATGTAACAAATACTGAACATCATCAACCCAGTAATAATGATTTGAACACCACTG
AGAAGCGTGCAGCTGAGAGGCATCCAGAAAAGTATCAGGGTAGTTCTGTTTCAAACCTGTCATGTGGAGCC
ATGTGGCAGAAATACTCATGCCAGCTCATTACAGCATGAGAACAGCAGTTTATTACTCACTAAAGACAGA
ATGAAATGTAGAAAAGGGCTGAATTCTGTAAATAAAAGCAAAACAGCCTGGCTTAGCAAGGAGGCCAATATAACA
```

<input type="radio"/>	<input type="radio"/>	1	6552298	Aug 23, 2004 11:53 AM
<input type="radio"/>	<input type="radio"/>	1	6552298	Dec 20, 2003 07:49 AM
<input type="radio"/>	<input type="radio"/>	1	6552298	Oct 4, 2003 03:37 AM
<input type="radio"/>	<input type="radio"/>	1	6552298	Sep 6, 2003 06:48 PM
<input type="radio"/>	<input type="radio"/>	1	6552298	Apr 6, 2003 09:44 AM
<input type="radio"/>	<input type="radio"/>	1	6552298	Jan 15, 2003 04:15 PM
<input type="radio"/>	<input type="radio"/>	1	6552298	Nov 5, 2002 01:21 PM
<input type="radio"/>	<input type="radio"/>	1	6552298	Aug 27, 2002 05:01 PM
<input type="radio"/>	<input type="radio"/>	1	6552298	Feb 6, 2001 12:05 PM
<input type="radio"/>	<input type="radio"/>	1	6552298	Nov 2, 2000 08:54 AM
<input type="radio"/>	<input type="radio"/>	1	6552298	Jan 6, 2000 12:20 PM
<input type="radio"/>	<input type="radio"/>	1	6552298	Dec 9, 1999 05:10 PM

Accession [NM_007294](#) was first seen at NCBI on Dec 9, 1999 05:10 PM

Revision History ▾

Send: ▾

Show difference between **I** and **II** as

[Close Comparison](#)

Gi	Version	Update Date
237757283	3	May 19, 2009 12:30 AM
63252871	2	May 17, 2009 11:04 AM

LOCUS NM_007294 7224 bp mRNA linear PRI 19-MAY-2009

LOCUS NM_007294 7191 bp mRNA linear PRI 17-MAY-2009

DEFINITION Homo sapiens breast cancer 1, early onset (BRCA1), transcript variant 1, mRNA.

ACCESSION NM_007294

VERSION NM_007294.3 GI:237757283

VERSION NM_007294.2 GI:63252871

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 7224)

[Open Comparison in separate window](#)

[Homo sapiens BRCA1, DNA repair associated \(BRCA1\), transcript variant 1, mRNA](#)

7,224 bp linear mRNA

Related information



[Annotated Genomic](#)

[BioSystems](#)

[CCDS](#)

[Components \(Core\)](#)

[Components \(EST\)](#)

[Full text in PMC](#)

[Gene](#)

[Map Viewer](#)

[OMIM](#)

[Protein](#)

[PubMed](#)

[PubMed \(RefSeq\)](#)

[PubMed \(Weighted\)](#)

[SNP](#)

[Taxonomy](#)

[UniGene](#)

[GEO Profiles](#)

GenPept ▾

Send to: ▾

Change region shown ▾

Customize view ▾

breast cancer type 1 susceptibility protein isoform 1 [Homo sapiens]

NCBI Reference Sequence: NP_009225.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to: ☐

LOCUS NP_009225 1863 aa linear PRI 03-SEP-2016
DEFINITION breast cancer type 1 susceptibility protein isoform 1 [Homo sapiens].
ACCESSION NP_009225
VERSION NP_009225.1 GI:6552299
DBSOURCE REFSEQ: accession [NM_007294.3](#)
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 (residues 1 to 1863)
AUTHORS Orban TI and Olah E.
TITLE Emerging roles of BRCA1 alternative splicing
JOURNAL MP, Mol. Pathol. 56 (4), 191-197 (2003)
PUBMED [12890739](#)
REMARK Review article
REFERENCE 2 (residues 1 to 1863)
AUTHORS Orban TI and Olah E.
TITLE Expression profiles of BRCA1 splice variants in asynchronous and in G1/S synchronized tumor cell lines

Analyze this sequence ▴

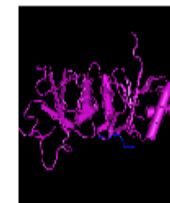
Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Protein 3D Structure ▴



Structure Of Brca1 Brct
Domains In Complex With
Abraxas Single
PDB: 4Y2G
Source: Homo sapiens
Method: X-Ray Diffraction
Resolution: 2.5 Å

[See all 27 structures...](#)

Articles about the BRCA1 gene ▴

Incidence of BRCA1 somatic mutations and

Region

23..68

/region_name="RING"

/note="RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc; defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-

H-X(2-3)-(N/C/H)-X2-C-X(4-48)C-X2-C; probably involved in...; cd00162"

ORIGIN

```
1 mdlsalrvee vqnvinamqk ilecpiclel ikepystkcd hifckfcmk llnqkkgpsq
61 cplcknditk rslqestrfs qlveellkii cafqldtgle yansynfakk ennspehlkd
121 evsiiiqsmgy rnrakrllqs epenpslqet slsvqlsnlg tvrtlrkqr iqpkqtsvyi
181 elgsdssedt vnkatycsvg dqellqitpq gtrdeislss akkaacefse tdtvntehhq
241 psnndlnnte kraaerhpek yqgssvsnlh vepcgtnttha sslqhenssl lltkdrmnve
301 kaefcnkskq pglarsqhnw wagsketend rrtptstekkv dlnadplcer kewnkqklpc
361 senprdtedv pwitlnssiq kvnewfsrsd ellgsddshd gesesnakva dvldvlnev
421 eysgssekid llsdpheal ickservhsk svesniedki fgktyrkas lplnshvten
481 liigafvtep qiiqerpltn klkrkrpts glhpedefikk adlavqktpe minqgtntqe
541 qngqvmnitn sghenktkgd siqneknpnp ieslekesaf ktkaeπισs ismmelelni
601 hnskapkknr lrrksstrhi halelvvsn lsspncitelq idscsseei kkkkynqmpv
661 rhrnlqlme gkepatgakk snkpneqtsk rhdsdtfpel kltnapgsft kcsntselke
721 fvnpslpree keekletvkv snnaedpkl mlsgerlvt ersvessis lvpqtdygtq
781 esisllevst lgkaktepnk cvsqcaafen pkglihgcsk dnrndtegf yplghevnh
841 retsiemees eldaqylqnt fkvskrqsf pfsnpgnaee ecatsahsg slkkqspkvt
901 feceqkeenq gknesnikpv qtvnitagfp vvgqkdkpvd nakcsikggs rfclssqfrg
961 netglitpnk hglqnpyri pplfpiksfv ktkckknle enfeehsmep eremngenip
1021 stvstisrnn irenvfkeas ssninevgss tnevgssine igssdeniqa elgrnrgpkl
1081 namrlrlglvq pevykqslpg snckhpeikk qeyeevvtv ntdfspyli dnleqpmgss
1141 hasqvcsetp ddllldgeik edtsfaendi kessavfsks vqkgelsrsp spfththlaq
1201 gyrrgakkle sseenlssed eelpcfqhl1 fgkvnnpisq strhstvat clsknteenl
1261 lslknslncl snqvilakas qehhlseetk csaslfssqc seleditant ntqdpfligs
1321 skqmrhqsas qvgvlsdkel vsddeertg leennqeeqs mdsnlgeaas gcesetsvse
1381 dcsqslssqd ilttqqrtdm qhnliklqqe maeleavleq hgsqpsnsyp siisdssale
1441 dlrnpeqsts ekavitsqks seypisqnpe glsadmfevs adsstsknke pgversspsk
1501 cpslldrwm hscsgslqnr nypsqeelik vvdveeqle esgphdltet sylprqdeg
1561 tpylesgisf fsddpesdps edrapesarv gnipsstsal kvpqkvaes aqspaaahtt
1621 dtagunamae svceaknalt asterunkem smvscaltse afmlyukfse kbhitltoli
```

23..68

/region_name="RING"

/note="RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc; defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48)C-X2-C; probably involved in...; cd00162"

/db_xref="CDD: [238093](#) "

Feature 1 of 5 NP_009225 : 1 segment

Details

Display: [FASTA](#) [GenBank](#) [Help](#)

GenPept

Format

- Summary
- GenPept**
- GenPept (full)
- FASTA
- FASTA (text)
- Graphics
- ASN.1
- Revision History
- Accession List
- GI List
- Identical Protein Report

breast cancer type 1 susceptibility protein isoform 1 [Homo sapiens]

Accession: NP_009225.1

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

1863 aa linear PRI 03-SEP-2016

breast cancer type 1 susceptibility protein isoform 1 [Homo sapiens]

GI: 6552299

Accession [NM_007294.3](#)

KEYWORDS

RefSeq.

Revision History

Show difference between **I** and **II** as

GenBank/GenPept

Compare

[breast cancer type 1 susceptibility protein isoform 1 \[Homo sapiens\]](#)

1863 aa protein

Accession: NP_009225.1 GI: 6552299

Current status: live

I	II	Version	Gi	Update Date	Action
<input checked="" type="radio"/>	<input type="radio"/>	1	6552299	Sep 3, 2016 02:17 AM	
<input type="radio"/>	<input checked="" type="radio"/>	1	6552299	May 25, 2016 12:31 AM	
<input type="radio"/>	<input type="radio"/>	1	6552299	Jan 26, 2016 12:41 AM	
<input type="radio"/>	<input type="radio"/>	1	6552299	Nov 18, 2015 12:38 AM	
<input type="radio"/>	<input type="radio"/>	1	6552299	Nov 14, 2015 12:31 AM	

Looking for DNA Sequences

ENTREZ Gene

Using Gene-centric Databases: Entrez Gene

- Entrez Gene can be accessed from the NCBI
- In GenBank, each entry is one **sequence from one publication**
- In Entrez Gene, each entry is one **gene**
- Entrez Gene is built with GenBank data

Comparison (Nucleotide vs Gene Search)

NCBI Resources How To

Nucleotide Nucleotide HIV1
Create alert Advanced

NCBI is phasing out sequence GI numbers in September 2016. Please use accession.version! [Read more...](#)

Species Summary 20 per page Sort by Default order Send to:

- Animals (1,517)
- Plants (1)
- Fungi (3)
- Protists (6)
- Viruses (654,581)
- Customize ...

Items: 1 to 20 of 655338

<< First < Prev Page 1 of 32767 Next > Last >>

Found 655348 nucleotide sequences. Nucleotide (655338) EST (10)

NCBI Resources How To

Gene Gene HIV1
Create RSS Create alert Advanced

Gene sources Tabular 20 per page Sort by Relevance Send to:

Genomic

Categories

- Alternatively spliced
- Annotated genes
- Protein-coding

Sequence content

- CCDS
- Ensembl
- RefSeq
- RefSeqGene

Search results

Items: 1 to 20 of 32

Showing Current items.

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> env ID: 155971	gp160; envelope glycoprotein [<i>Human immunodeficiency virus 1</i>]	NC_001802.1 (5771..8341)	HIV1gp8	

Looking for DNA Sequences

EMBL-EBI (ENSEMBL)



Services

[Overview](#) | [A to Z](#) | [Data submission](#) | [Support](#)

Databases and tools

EMBL-EBI offers a comprehensive range of **freely available**, up-to-date [molecular databases](#). You can work locally by downloading our data and software, or use our [web services](#) to access our resources programmatically. You can read more about our services in the journal [Nucleic Acids Research](#).

Our services are made possible by scientists all over the world sharing their data. This increases the impact of individual research projects, and allows science to progress more rapidly. To submit your data to the public archives, visit our [Data Submission Wizard](#).

DNA & RNA

genes, genomes & variation

Gene expression

RNA, protein & metabolite expression

Proteins

sequences, families & motifs

Structures

Molecular & cellular structures


Systems

reactions, interactions & pathways


Chemical biology


chemogenomics & metabolomics

Popular

 Ensembl

 UniProt


 PDB

 ArrayExpress

 ChEMBL

 BLAST

 Europe PMC

 Reactome

 Train online

 Support

Service news



Search: for

Go

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

Browse a Genome

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Popular genomes



Human

GRCh38.p7



Human

GRCh37



Mouse

GRCm38.p4



Zebrafish

GRCz10

★ [Log in to customize this list](#)

All genomes

-- Select a species --

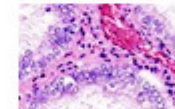
Still using Human GRCh37?



Variant Effect Predictor



Gene expression in different tissues



Find SNPs and other variants for my gene

GTGTATACATT
CTTAAAGTCTT
CTTCTAAATTCT
GTAACATTTTCC

Retrieve gene sequence

GCCTGACTTCGGGTGG
GGGCTTGTGGCGGAGC
GCGCCTCTGCTGCGCT
AGGGACAGATTTGTGA
CACCTCTGGAGCGGTT
CCCACTCCAGCGTGGCG

Compare genes across species



Use my own data in Ensembl

ENCODE data in Ensembl

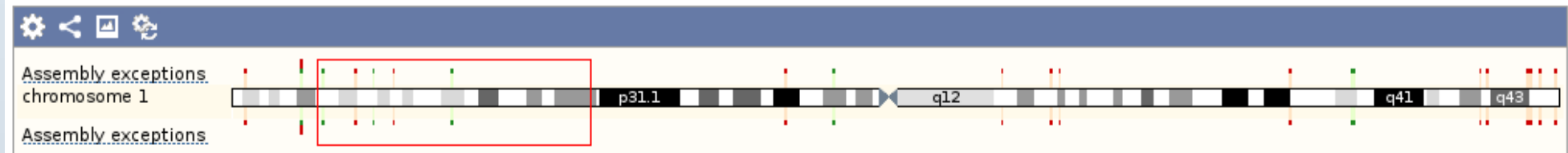
ENSEMBL

- ENSEMBL is a database of eukaryotic genomes
 - Annotated entries
 - Wide range of examples: human, mouse, dog, and so on
- ENSEMBL annotation is mostly automated
- ENSEMBL contains tools to
 - Browse the complete genome
 - Search the complete genome with BLAST
 - Visualize the position of a gene
 - Visualize all experimental information on this gene (transcripts)

Visualizing Human Chromosomes

- By pointing on a chromosome region you can zoom inside the chromosome
- All genes are cross-indexed with databases so you can find all related experimental information

Chromosome 1: 15,996,032-67,183,334

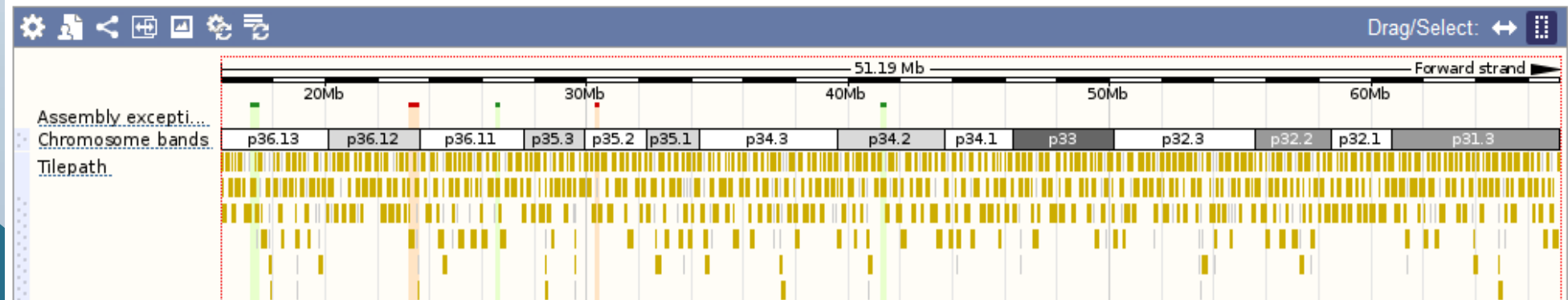


Region overview ?

Location:

Gene:

Navigation controls:



Looking for DNA Sequences

DDBJ

About DDBJ

How to Use

Report/Statistics

FAQ

Contact Us

DDBJ Data Analysis Challenge

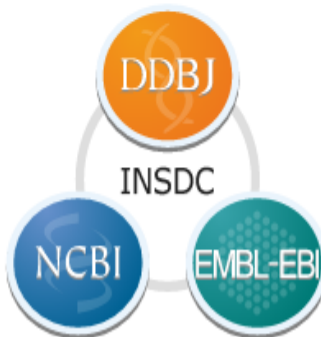
July 6 Kickoff
August 31 Submission deadline



RSS



DDBJ Twitter



International Nucleotide Sequence
Database Collaboration

DDBJ Service



Data Submission



Search / Analysis



Super Computer



ftp.ddbj.nig.ac.jp

Hot Topics

News Archive

News

Release

PR

Maintenance

Operation

All

2016.09.09

PDB 2016-08-25 released

Search and Analysis

Database Search

[getentry](#)

Data retrieval by accession numbers, etc.

[ARSA](#)

All-round Retrieval of Sequence and Annotation

[TXSearch](#)

Retrieval of unified taxonomy database

[BLAST](#)

Homology Search

[VecScreen](#)

Vector Search to screen contamination in nucleic acid sequences.

[DRA Search](#)

Search metadata by keywords and retrieve data

Phylogenetics

[ClustalW](#)

Multiple alignment and Tree-making

WABI (Web API for Biology)

Next Generation Sequence Analysis

[DDBJ Read Annotation Pipeline](#)

High-throughput data analysis of next generation sequence data (Login ID is required)

Genome Analyses

[MiGAP](#)

Mechanical annotation tool for microbial genomes(Login ID is required.)

[MiGAP-OLD \(Result before Feb. 2012. Read-only\)](#)

To login, DBCLS OpenID is required.

Links for DBCLS* Search Tools

[AOE](#)

Statistics and trends of gene expression data

[CRISPRdirect](#)

Designing CRISPR/Cas9 guide RNA with reduced off-target sites

[DBCLS SRA](#)

Statistics and trends of SRA data

[Gendoo](#)

Functional profiling of gene and disease features for omics analysis

Going Further

- University of California at Santa Cruz: <http://genome.ucsc.edu/>
- The DoE Joint Genome Institute: <http://jgi.doe.gov/>
- ...