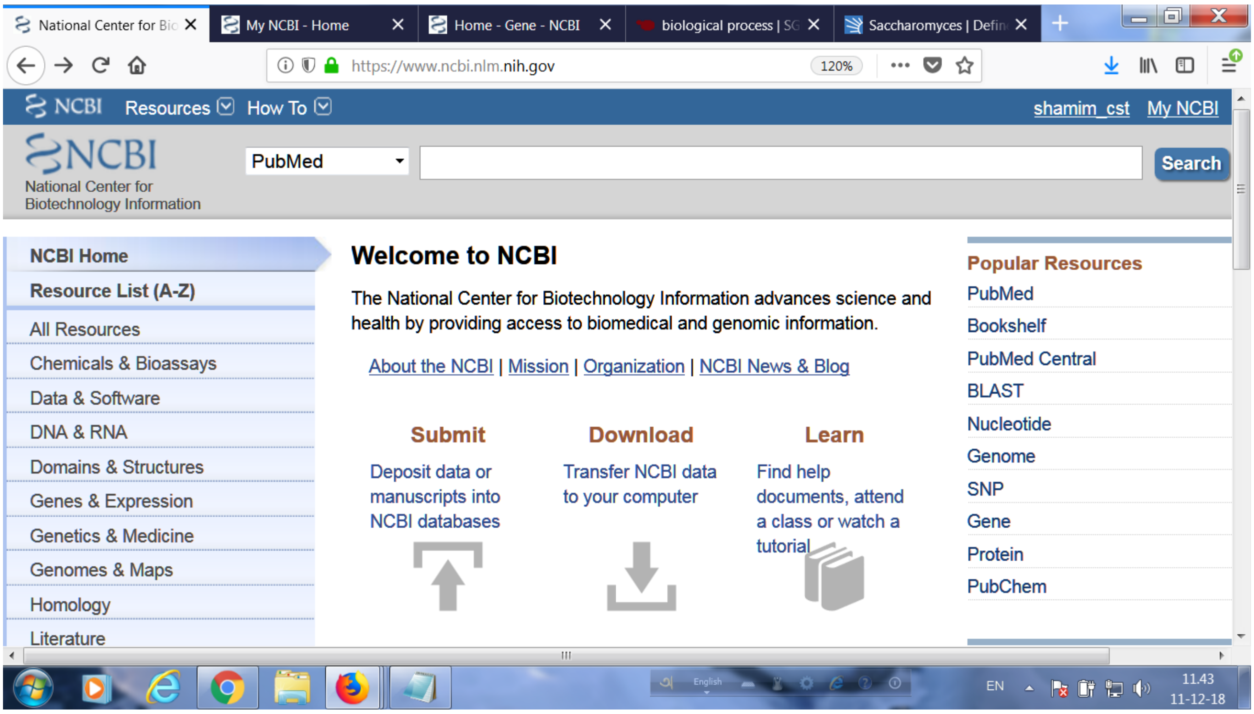
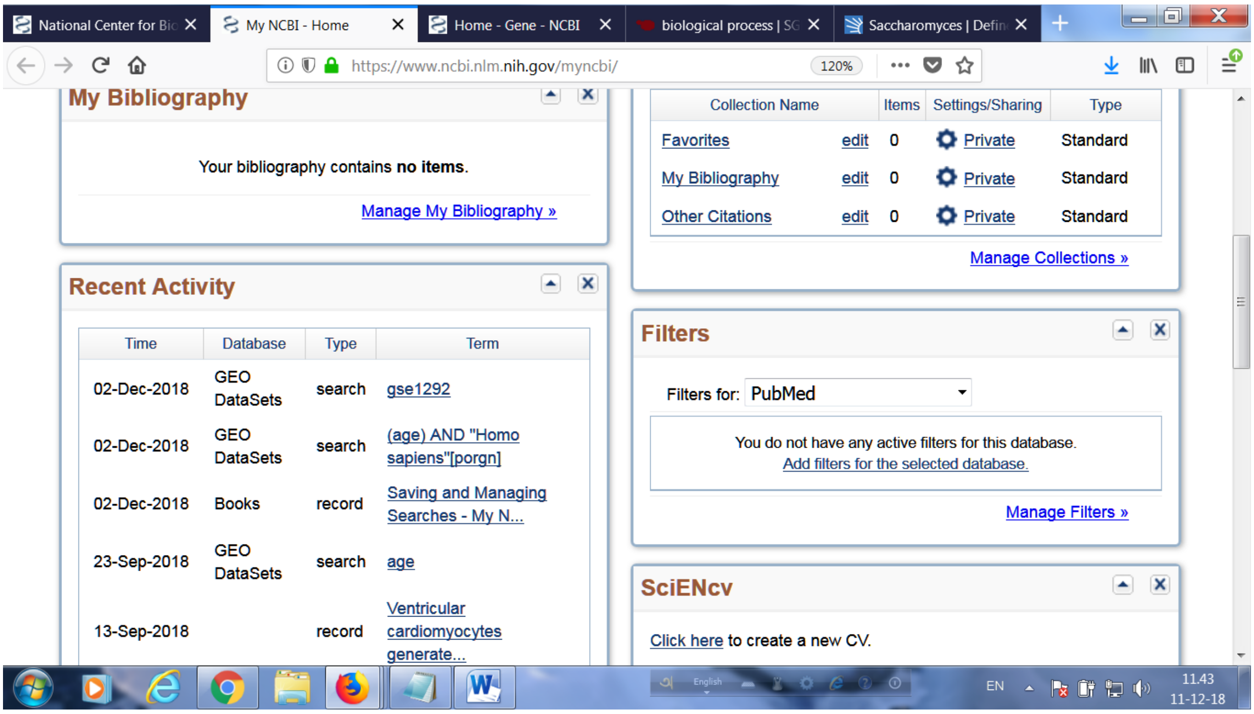


Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit	Download	Learn
Deposit data or manuscripts into NCBI databases	Transfer NCBI data to your computer	Find help documents, attend a class or watch a tutorial



My Bibliography

Your bibliography contains **no items**.

Manage My Bibliography »

Collection Name		Items	Settings/Sharing	Type
Favorites	edit	0	Private	Standard
My Bibliography	edit	0	Private	Standard
Other Citations	edit	0	Private	Standard

Manage Collections »

Recent Activity

Time	Database	Type	Term
02-Dec-2018	GEO DataSets	search	gse1292
02-Dec-2018	GEO DataSets	search	(age) AND "Homo sapiens"[porgn]
02-Dec-2018	Books	record	Saving and Managing Searches - My N...
23-Sep-2018	GEO DataSets	search	age
13-Sep-2018		record	Ventricular cardiomyocytes generate...

Filters

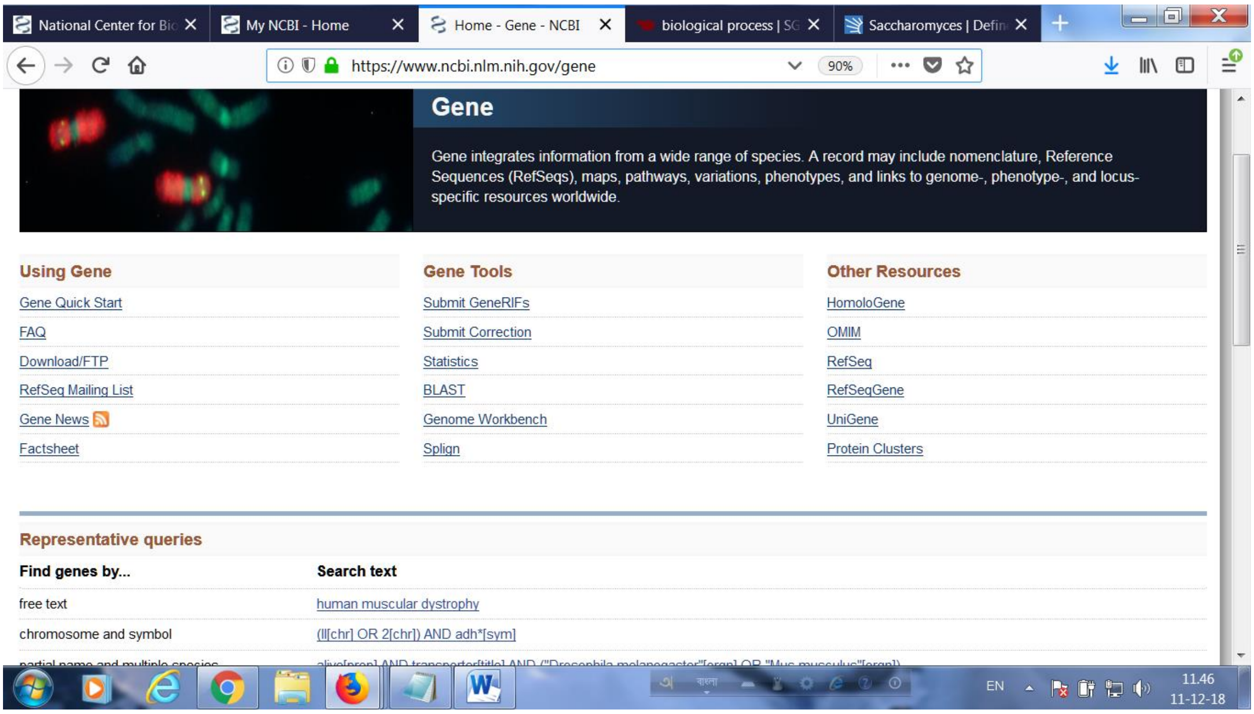
Filters for: PubMed

You do not have any active filters for this database.
Add filters for the selected database.

Manage Filters »

SciENcv

Click here to create a new CV.



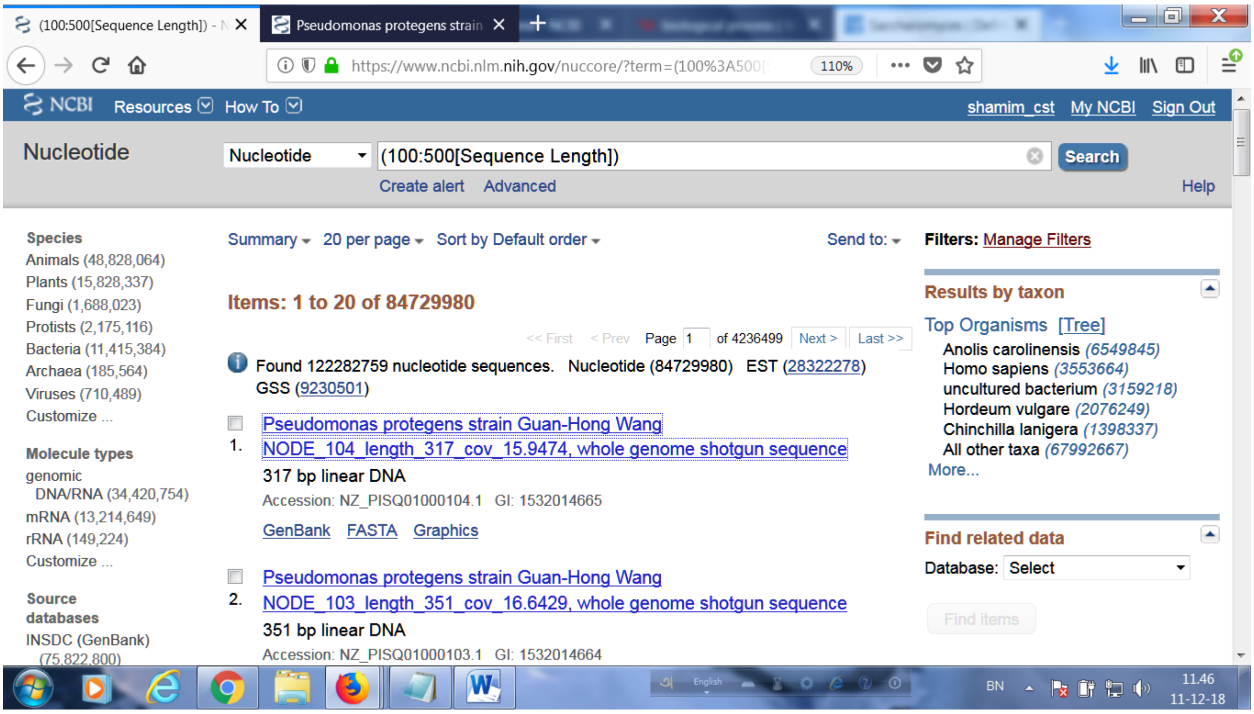
Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.

Using Gene	Gene Tools	Other Resources
Gene Quick Start	Submit GeneRIFs	HomoloGene
FAQ	Submit Correction	OMIM
Download/FTP	Statistics	RefSeq
RefSeq Mailing List	BLAST	RefSeqGene
Gene News	Genome Workbench	UniGene
Factsheet	Splign	Protein Clusters

Representative queries

Find genes by...	Search text
free text	human muscular dystrophy
chromosome and symbol	(II[chr] OR 2[chr]) AND adh*[sym]



Nucleotide

Nucleotide (100:500[Sequence Length])

Species
Animals (48,828,064)
Plants (15,828,337)
Fungi (1,688,023)
Protists (2,175,116)
Bacteria (11,415,384)
Archaea (185,564)
Viruses (710,489)
Customize ...

Molecule types
genomic
DNA/RNA (34,420,754)
mRNA (13,214,649)
rRNA (149,224)
Customize ...

Source databases
INSDC (GenBank)
(75,822,800)

Items: 1 to 20 of 84729980

Page 1 of 4236499

Found 122282759 nucleotide sequences. Nucleotide (84729980) EST (28322278)
GSS (9230501)

☐ 1. Pseudomonas protegens strain Guan-Hong Wang NODE_104_length_317_cov_15.9474, whole genome shotgun sequence
317 bp linear DNA
Accession: NZ_PISQ01000104.1 GI: 1532014665
GenBank FASTA Graphics

☐ 2. Pseudomonas protegens strain Guan-Hong Wang NODE_103_length_351_cov_16.6429, whole genome shotgun sequence
351 bp linear DNA
Accession: NZ_PISQ01000103.1 GI: 1532014664

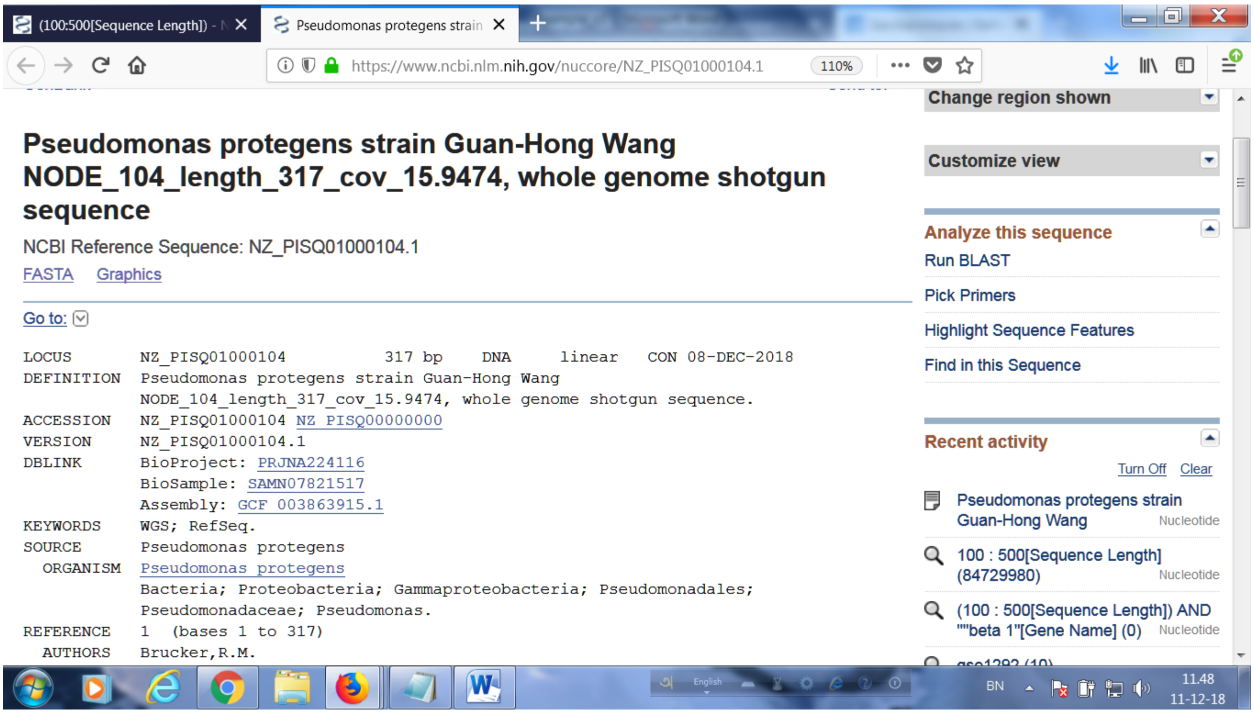
Filters: Manage Filters

Results by taxon

Top Organisms [Tree]
Anolis carolinensis *(6549845)*
Homo sapiens *(3553664)*
uncultured bacterium *(3159218)*
Hordeum vulgare *(2076249)*
Chinchilla lanigera *(1398337)*
All other taxa *(67992667)*
More...

Find related data

Database: Select



Pseudomonas protegens strain Guan-Hong Wang NODE_104_length_317_cov_15.9474, whole genome shotgun sequence

NCBI Reference Sequence: NZ_PISQ01000104.1

FASTA Graphics

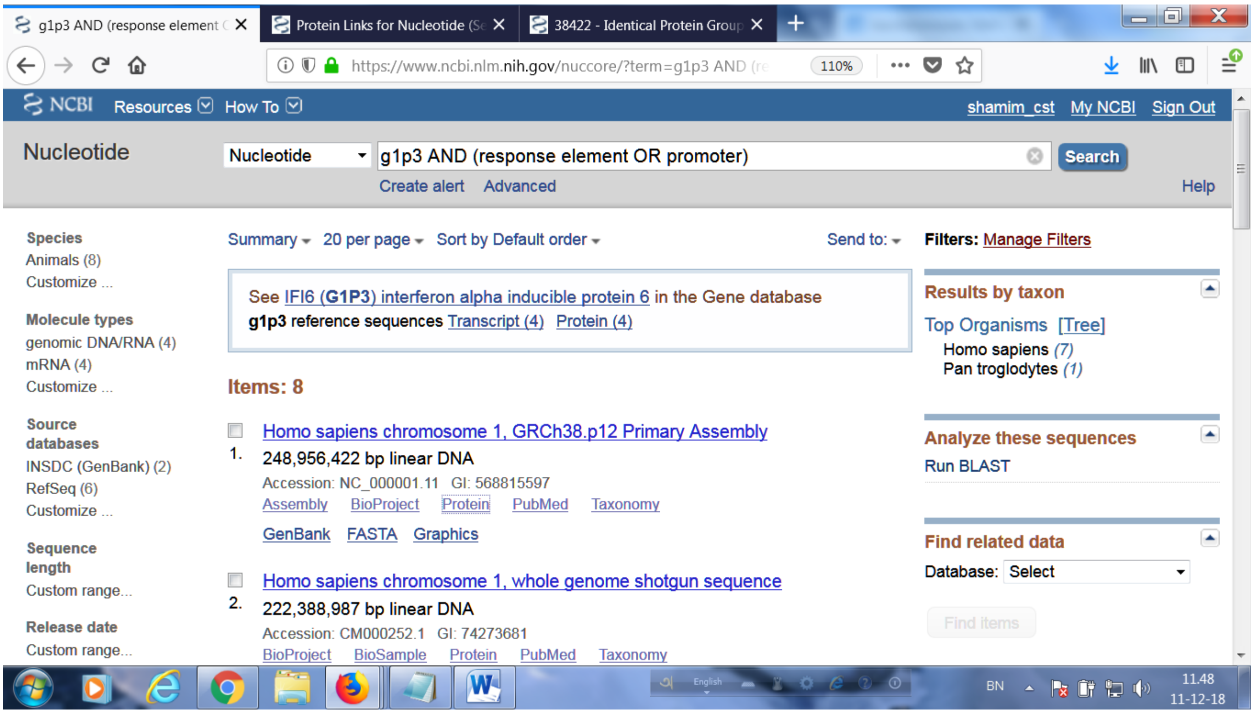
Go to:

```
LOCUS       NZ_PISQ01000104          317 bp    DNA     linear   CON 08-DEC-2018
DEFINITION  Pseudomonas protegens strain Guan-Hong Wang
            NODE_104_length_317_cov_15.9474, whole genome shotgun sequence.
ACCESSION   NZ_PISQ01000104 NZ_PISQ00000000
VERSION     NZ_PISQ01000104.1
DBLINK      BioProject: PRJNA224116
            BioSample: SAMN07821517
            Assembly: GCF_003863915.1
KEYWORDS    WGS; RefSeq.
SOURCE      Pseudomonas protegens
  ORGANISM  Pseudomonas protegens
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Brucker,R.M.
```

Pseudomonas protegens strain Guan-Hong Wang — Nucleotide

100 : 500[Sequence Length] (84729980) — Nucleotide

(100 : 500[Sequence Length]) AND ""beta 1"[Gene Name] (0) — Nucleotide



Nucleotide | g1p3 AND (response element OR promoter)

Species
Animals (8)
Customize ...

Molecule types
genomic DNA/RNA (4)
mRNA (4)
Customize ...

Source databases
INSDC (GenBank) (2)
RefSeq (6)
Customize ...

Sequence length
Custom range...

Release date
Custom range...

See IFI6 (**G1P3**) interferon alpha inducible protein 6 in the Gene database
g1p3 reference sequences Transcript (4) Protein (4)

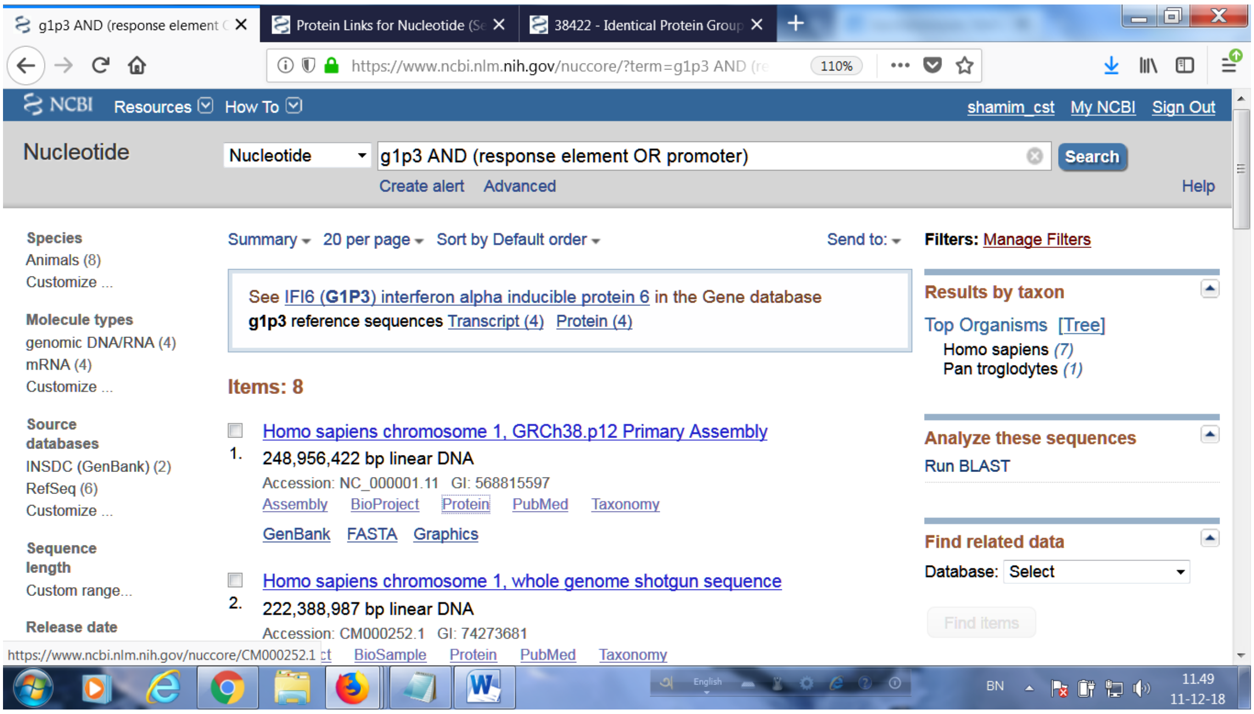
Items: 8

☐ 1. Homo sapiens chromosome 1, GRCh38.p12 Primary Assembly
248,956,422 bp linear DNA
Accession: NC_000001.11 GI: 568815597
Assembly BioProject Protein PubMed Taxonomy
GenBank FASTA Graphics

☐ 2. Homo sapiens chromosome 1, whole genome shotgun sequence
222,388,987 bp linear DNA
Accession: CM000252.1 GI: 74273681
BioProject BioSample Protein PubMed Taxonomy

Results by taxon
Top Organisms [Tree]
Homo sapiens *(7)*
Pan troglodytes *(1)*

Database: Select



Nucleotide: g1p3 AND (response element OR promoter)

Species
Animals (8)
Customize ...

Molecule types
genomic DNA/RNA (4)
mRNA (4)
Customize ...

Source databases
INSDC (GenBank) (2)
RefSeq (6)
Customize ...

Sequence length
Custom range...

Release date

Summary 20 per page Sort by Default order Send to:

See IFI6 (**G1P3**) interferon alpha inducible protein 6 in the Gene database
g1p3 reference sequences Transcript (4) Protein (4)

Items: 8

☐ 1. Homo sapiens chromosome 1, GRCh38.p12 Primary Assembly
248,956,422 bp linear DNA
Accession: NC_000001.11 GI: 568815597
Assembly BioProject Protein PubMed Taxonomy
GenBank FASTA Graphics

☐ 2. Homo sapiens chromosome 1, whole genome shotgun sequence
222,388,987 bp linear DNA
Accession: CM000252.1 GI: 74273681
BioSample Protein PubMed Taxonomy

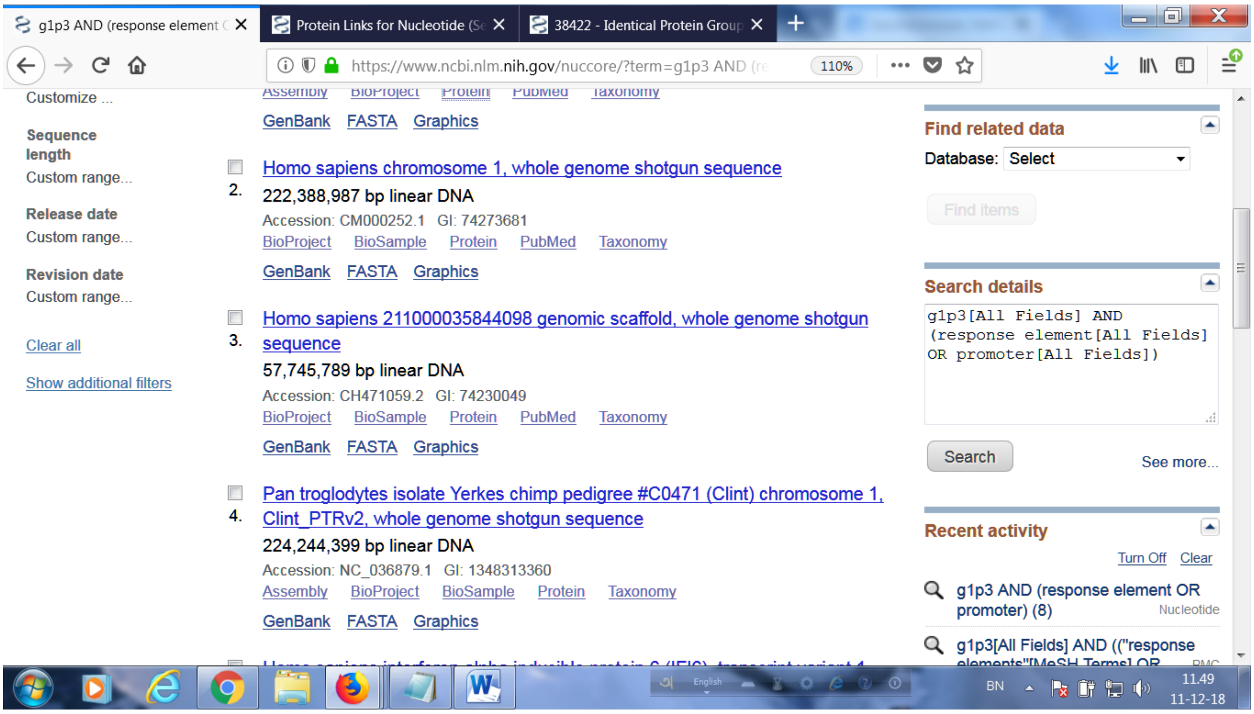
https://www.ncbi.nlm.nih.gov/nuccore/CM000252.1

Filters: Manage Filters

Results by taxon
Top Organisms [Tree]
Homo sapiens (7)
Pan troglodytes (1)

Analyze these sequences
Run BLAST

Find related data
Database: Select
Find items



Customize ...

Sequence length
Custom range...

Release date
Custom range...

Revision date
Custom range...

Clear all

Show additional filters

Assembly BioProject Protein PubMed Taxonomy
GenBank FASTA Graphics

☐ 2. Homo sapiens chromosome 1, whole genome shotgun sequence
222,388,987 bp linear DNA
Accession: CM000252.1 GI: 74273681
BioProject BioSample Protein PubMed Taxonomy
GenBank FASTA Graphics

☐ 3. Homo sapiens 211000035844098 genomic scaffold, whole genome shotgun sequence
57,745,789 bp linear DNA
Accession: CH471059.2 GI: 74230049
BioProject BioSample Protein PubMed Taxonomy
GenBank FASTA Graphics

☐ 4. Pan troglodytes isolate Yerkes chimp pedigree #C0471 (Clint) chromosome 1, Clint_PTRv2, whole genome shotgun sequence
224,244,399 bp linear DNA
Accession: NC_036879.1 GI: 1348313360
Assembly BioProject BioSample Protein Taxonomy
GenBank FASTA Graphics

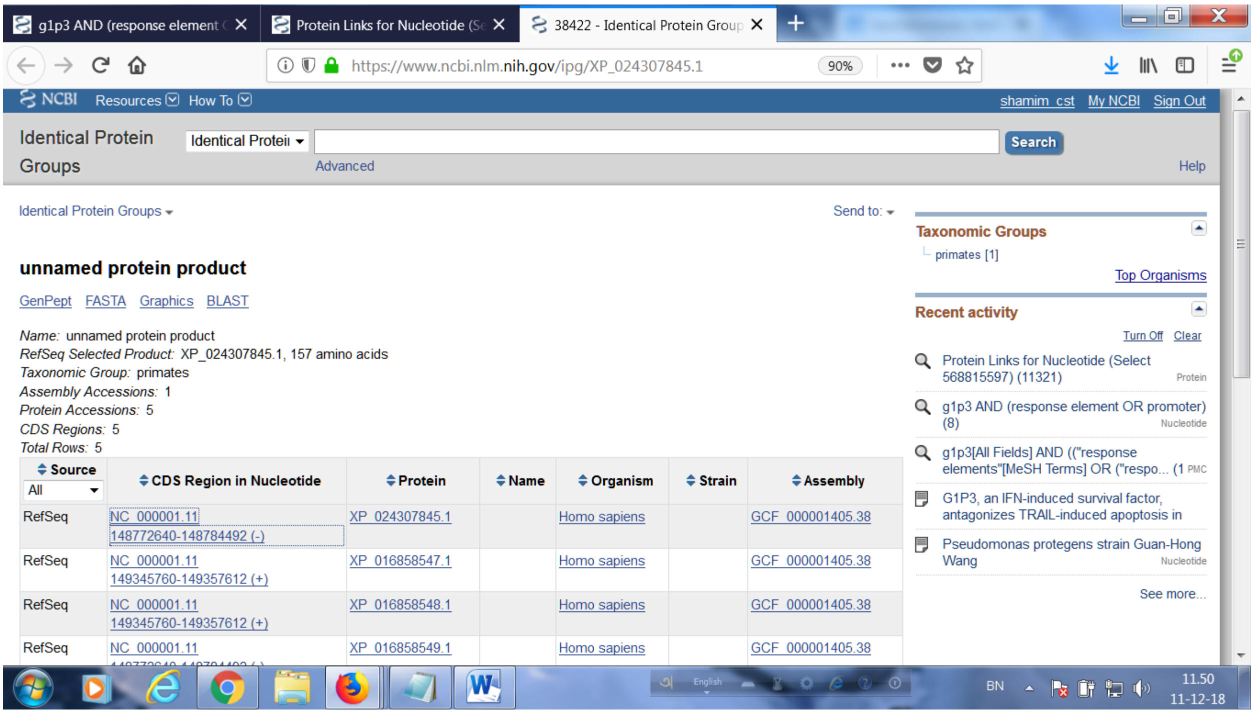
Find related data
Database: Select
Find items

Search details

```
g1p3[All Fields] AND
(response element[All Fields]
OR promoter[All Fields])
```

Search See more...

Recent activity
Turn Off Clear
g1p3 AND (response element OR promoter) (8) Nucleotide
g1p3[All Fields] AND (("response elements"[MeSH Terms] OR



unnamed protein product

GenPept FASTA Graphics BLAST

Name: unnamed protein product
RefSeq Selected Product: XP_024307845.1, 157 amino acids
Taxonomic Group: primates
Assembly Accessions: 1
Protein Accessions: 5
CDS Regions: 5
Total Rows: 5

Source All	CDS Region in Nucleotide	Protein	Name	Organism	Strain	Assembly
RefSeq	NC_000001.11 148772640-148784492 (-)	XP_024307845.1		Homo sapiens		GCF_000001405.38
RefSeq	NC_000001.11 149345760-149357612 (+)	XP_016858547.1		Homo sapiens		GCF_000001405.38
RefSeq	NC_000001.11 149345760-149357612 (+)	XP_016858548.1		Homo sapiens		GCF_000001405.38
RefSeq	NC_000001.11	XP_016858549.1		Homo sapiens		GCF_000001405.38

Taxonomic Groups

primates [1]

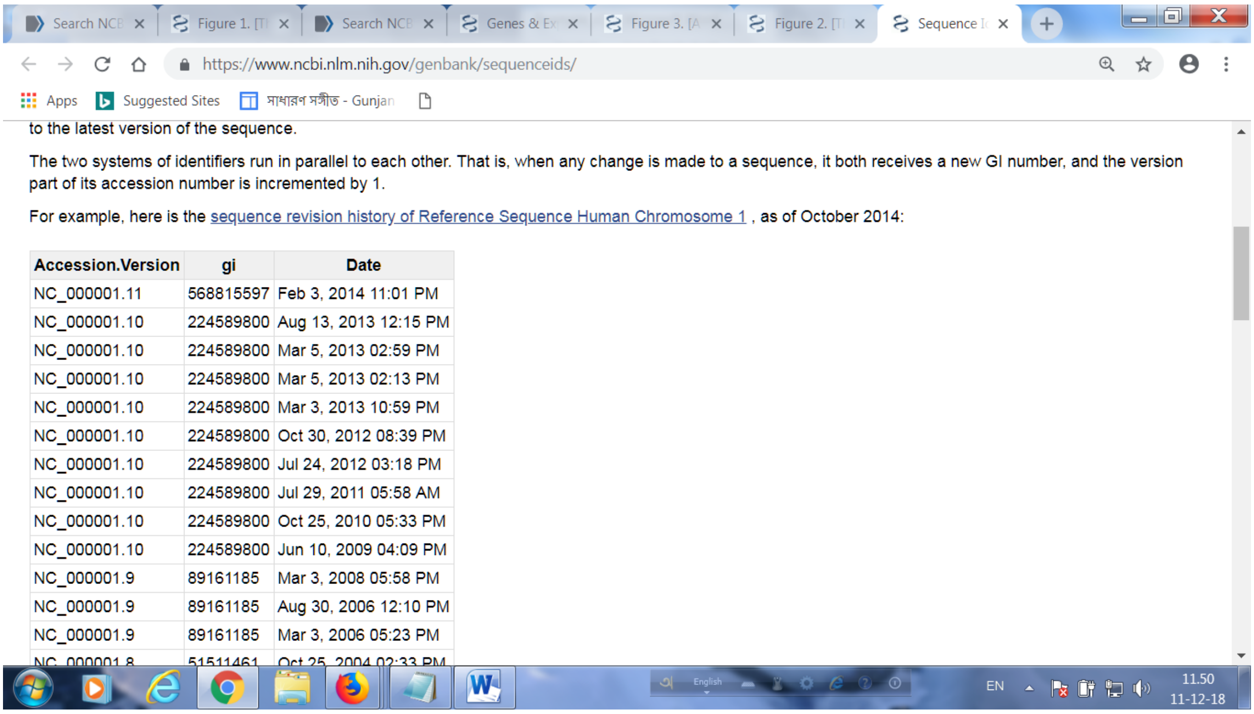
Top Organisms

Recent activity

Turn Off Clear

- Protein Links for Nucleotide (Select 568815597) (11321) Protein
- g1p3 AND (response element OR promoter) (8) Nucleotide
- g1p3[All Fields] AND (("response elements"[MeSH Terms] OR ("respo... (1 PMC
- G1P3, an IFN-induced survival factor, antagonizes TRAIL-induced apoptosis in
- Pseudomonas protegens strain Guan-Hong Wang Nucleotide

See more...

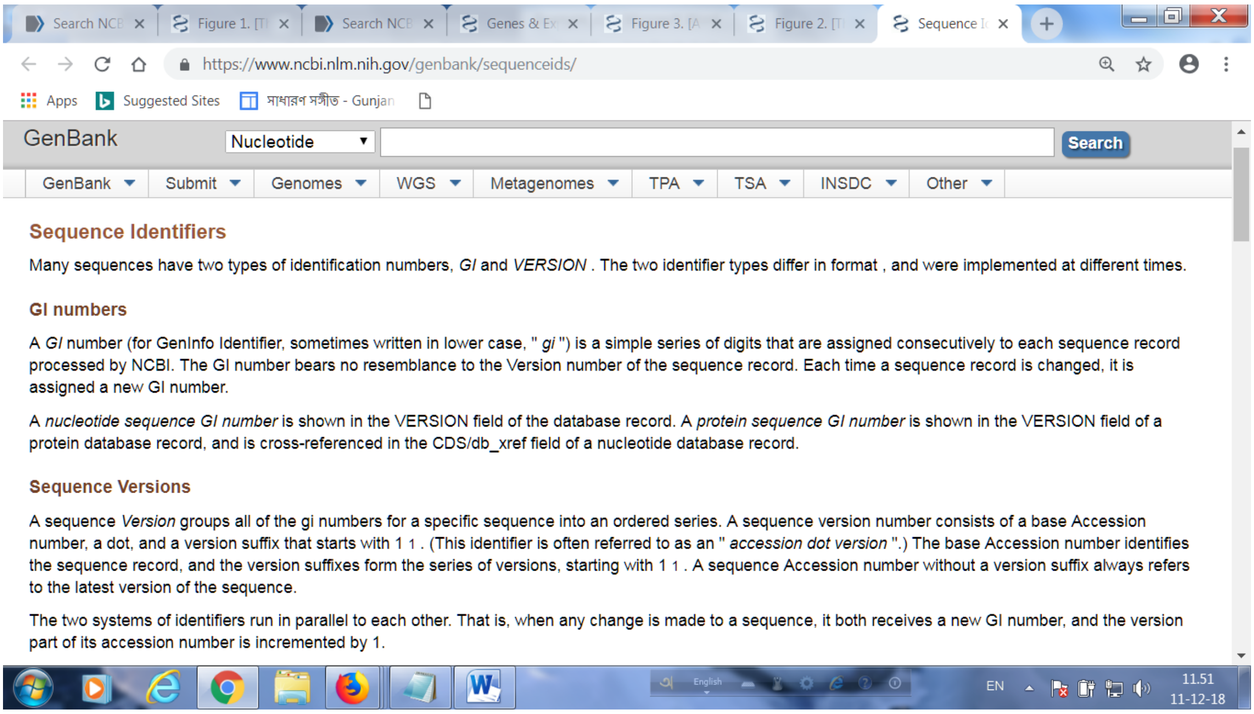


to the latest version of the sequence.

The two systems of identifiers run in parallel to each other. That is, when any change is made to a sequence, it both receives a new GI number, and the version part of its accession number is incremented by 1.

For example, here is the sequence revision history of Reference Sequence Human Chromosome 1 , as of October 2014:

Accession.Version	gi	Date
NC_000001.11	568815597	Feb 3, 2014 11:01 PM
NC_000001.10	224589800	Aug 13, 2013 12:15 PM
NC_000001.10	224589800	Mar 5, 2013 02:59 PM
NC_000001.10	224589800	Mar 5, 2013 02:13 PM
NC_000001.10	224589800	Mar 3, 2013 10:59 PM
NC_000001.10	224589800	Oct 30, 2012 08:39 PM
NC_000001.10	224589800	Jul 24, 2012 03:18 PM
NC_000001.10	224589800	Jul 29, 2011 05:58 AM
NC_000001.10	224589800	Oct 25, 2010 05:33 PM
NC_000001.10	224589800	Jun 10, 2009 04:09 PM
NC_000001.9	89161185	Mar 3, 2008 05:58 PM
NC_000001.9	89161185	Aug 30, 2006 12:10 PM
NC_000001.9	89161185	Mar 3, 2006 05:23 PM
NC_000001.8	51511461	Oct 25, 2004 02:33 PM



Sequence Identifiers

Many sequences have two types of identification numbers, *GI* and *VERSION* . The two identifier types differ in format , and were implemented at different times.

GI numbers

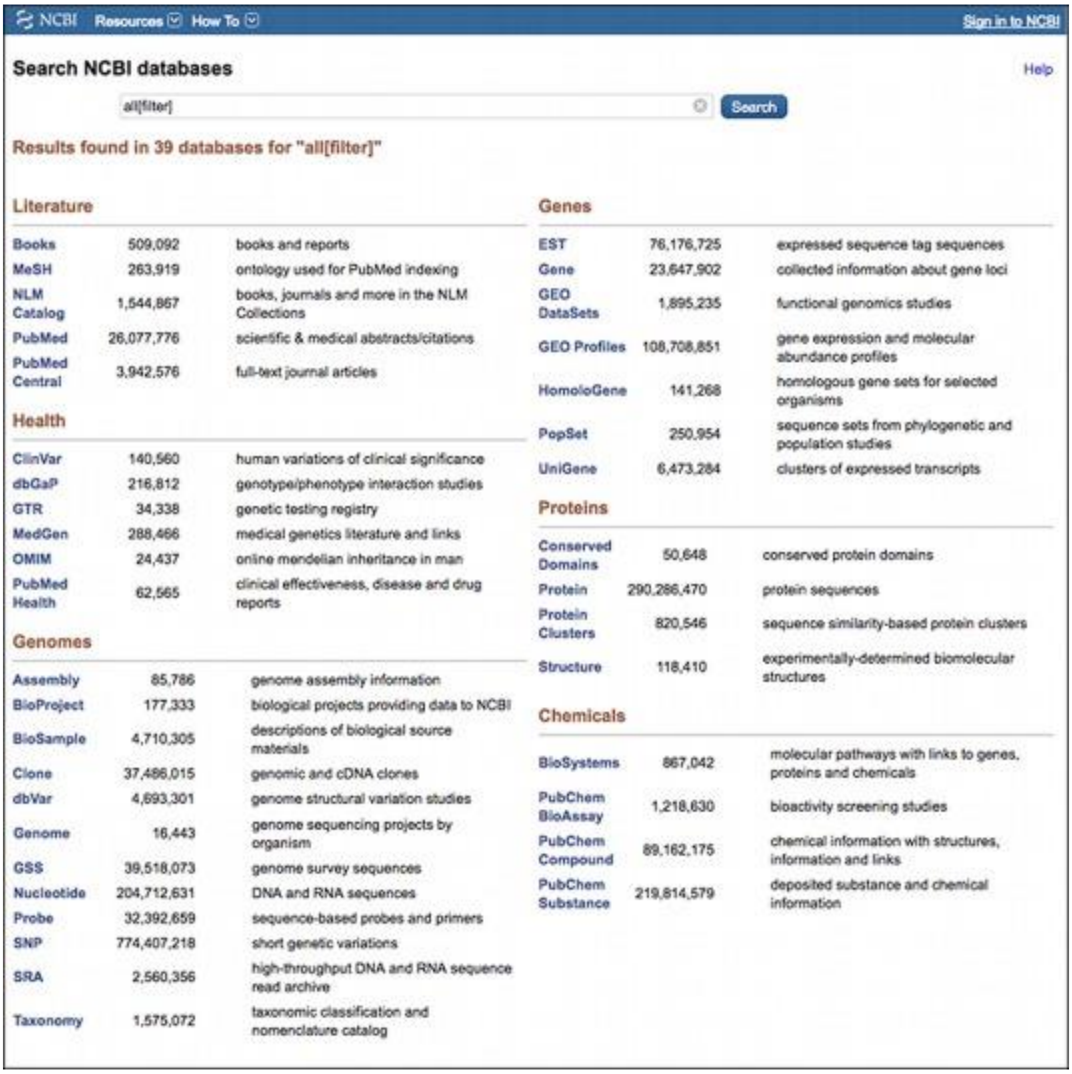
A *GI* number (for GenInfo Identifier, sometimes written in lower case, " *gi* ") is a simple series of digits that are assigned consecutively to each sequence record processed by NCBI. The GI number bears no resemblance to the Version number of the sequence record. Each time a sequence record is changed, it is assigned a new GI number.

A *nucleotide sequence GI number* is shown in the VERSION field of the database record. A *protein sequence GI number* is shown in the VERSION field of a protein database record, and is cross-referenced in the CDS/db_xref field of a nucleotide database record.

Sequence Versions

A sequence *Version* groups all of the gi numbers for a specific sequence into an ordered series. A sequence version number consists of a base Accession number, a dot, and a version suffix that starts with 1 1 . (This identifier is often referred to as an " *accession dot version* ".) The base Accession number identifies the sequence record, and the version suffixes form the series of versions, starting with 1 1 . A sequence Accession number without a version suffix always refers to the latest version of the sequence.

The two systems of identifiers run in parallel to each other. That is, when any change is made to a sequence, it both receives a new GI number, and the version part of its accession number is incremented by 1.



Search NCBI databases

all[filter]

Results found in 39 databases for "all[filter]"

Literature

Books	509,092	books and reports
MeSH	263,919	ontology used for PubMed indexing
NLM Catalog	1,544,867	books, journals and more in the NLM Collections
PubMed	26,077,776	scientific & medical abstracts/citations
PubMed Central	3,942,576	full-text journal articles

Health

ClinVar	140,560	human variations of clinical significance
dbGaP	216,812	genotype/phenotype interaction studies
GTR	34,338	genetic testing registry
MedGen	288,466	medical genetics literature and links
OMIM	24,437	online mendelian inheritance in man
PubMed Health	62,565	clinical effectiveness, disease and drug reports

Genomes

Assembly	85,786	genome assembly information
BioProject	177,333	biological projects providing data to NCBI
BioSample	4,710,305	descriptions of biological source materials
Clone	37,486,015	genomic and cDNA clones
dbVar	4,693,301	genome structural variation studies
Genome	16,443	genome sequencing projects by organism
GSS	39,518,073	genome survey sequences
Nucleotide	204,712,631	DNA and RNA sequences
Probe	32,392,659	sequence-based probes and primers
SNP	774,407,218	short genetic variations
SRA	2,560,356	high-throughput DNA and RNA sequence read archive
Taxonomy	1,575,072	taxonomic classification and nomenclature catalog

Genes

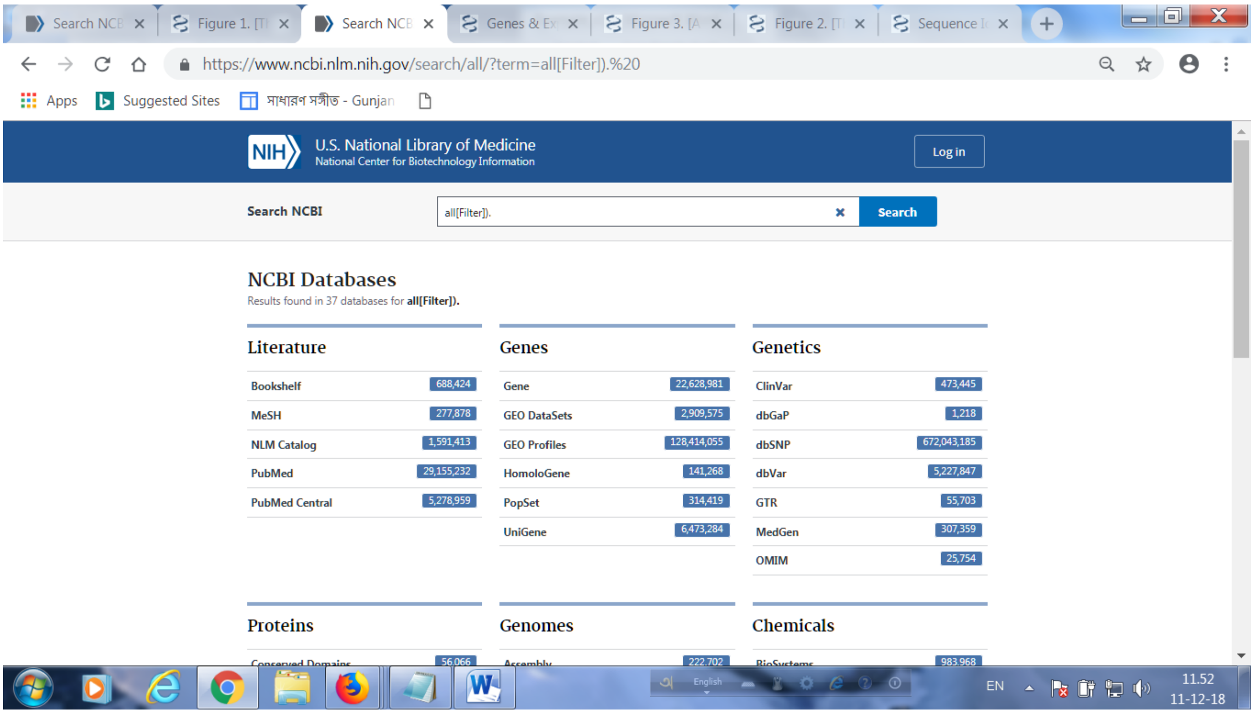
EST	76,176,725	expressed sequence tag sequences
Gene	23,647,902	collected information about gene loci
GEO DataSets	1,895,235	functional genomics studies
GEO Profiles	108,708,851	gene expression and molecular abundance profiles
HomoloGene	141,268	homologous gene sets for selected organisms
PopSet	250,954	sequence sets from phylogenetic and population studies
UniGene	6,473,284	clusters of expressed transcripts

Proteins

Conserved Domains	50,648	conserved protein domains
Protein	290,286,470	protein sequences
Protein Clusters	820,546	sequence similarity-based protein clusters
Structure	118,410	experimentally-determined biomolecular structures

Chemicals

BioSystems	867,042	molecular pathways with links to genes, proteins and chemicals
PubChem BioAssay	1,218,630	bioactivity screening studies
PubChem Compound	89,162,175	chemical information with structures, information and links
PubChem Substance	219,814,579	deposited substance and chemical information



NCBI Databases

Results found in 37 databases for **all[Filter]).**

Literature

Bookshelf	688,424
MeSH	277,878
NLM Catalog	1,591,413
PubMed	29,155,232
PubMed Central	5,278,959

Genes

Gene	22,628,981
GEO DataSets	2,909,575
GEO Profiles	128,414,055
HomoloGene	141,268
PopSet	314,419
UniGene	6,473,284

Genetics

ClinVar	473,445
dbGaP	1,218
dbSNP	672,043,185
dbVar	5,227,847
GTR	55,703
MedGen	307,359
OMIM	25,754

Proteins

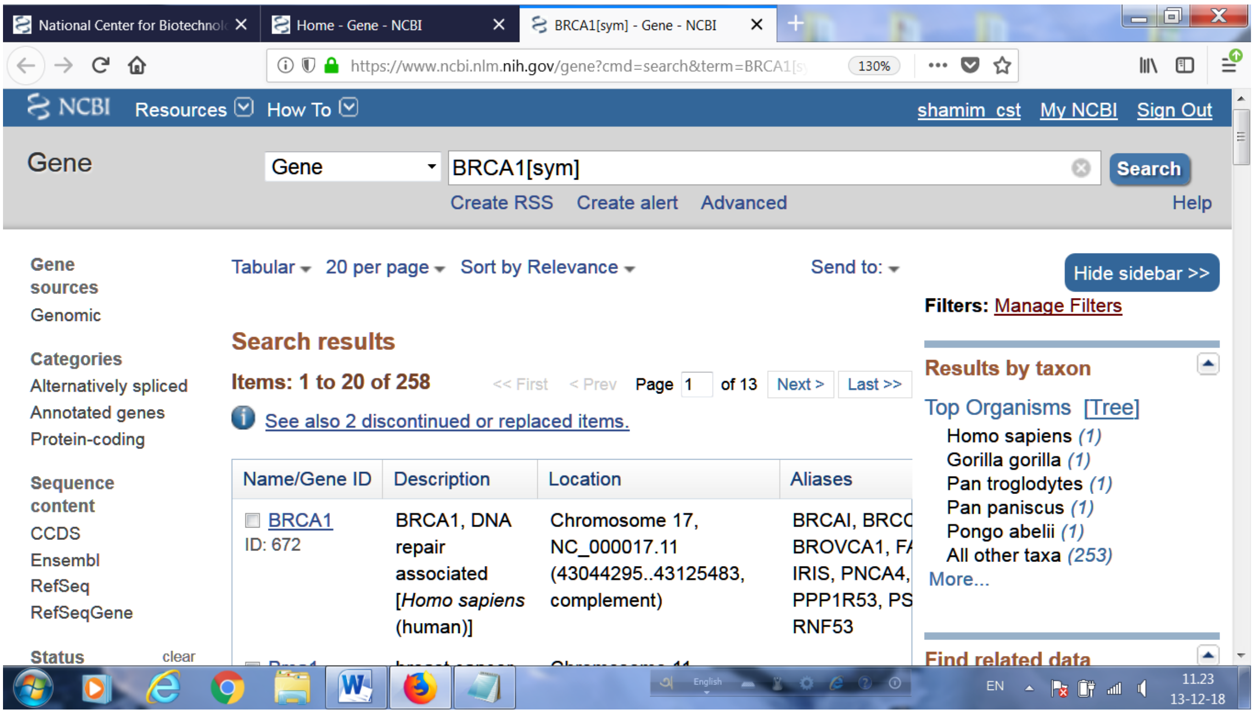
Conserved Domains	56,066

Genomes

Assembly	222,702

Chemicals

BioSystems	983,968



Gene

Gene ▾ BRCA1[sym]

Gene sources
Genomic

Categories
Alternatively spliced
Annotated genes
Protein-coding

Sequence content
CCDS
Ensembl
RefSeq
RefSeqGene

Status clear

Search results

Items: 1 to 20 of 258

<< First < Prev Page 1 of 13 Next > Last >>

See also 2 discontinued or replaced items.

Name/Gene ID	Description	Location	Aliases
☐ BRCA1 ID: 672	BRCA1, DNA repair associated [*Homo sapiens* (human)]	Chromosome 17, NC_000017.11 (43044295..43125483, complement)	BRCAI, BRCC BROVCA1, F IRIS, PNCA4, PPP1R53, PS RNF53

Filters: Manage Filters

Results by taxon

Top Organisms [Tree]

- Homo sapiens *(1)*
- Gorilla gorilla *(1)*
- Pan troglodytes *(1)*
- Pan paniscus *(1)*
- Pongo abelii *(1)*
- All other taxa *(253)*

More...

Find related data

https://www.ncbi.nlm.nih.gov/gene

Gene

Gene: (BRCA1[sym]) AND "Homo sapiens"[porgn:__txid9606]

Showing Current items.

BRCA1 BRCA1, DNA repair associated [*Homo sapiens* (human)]

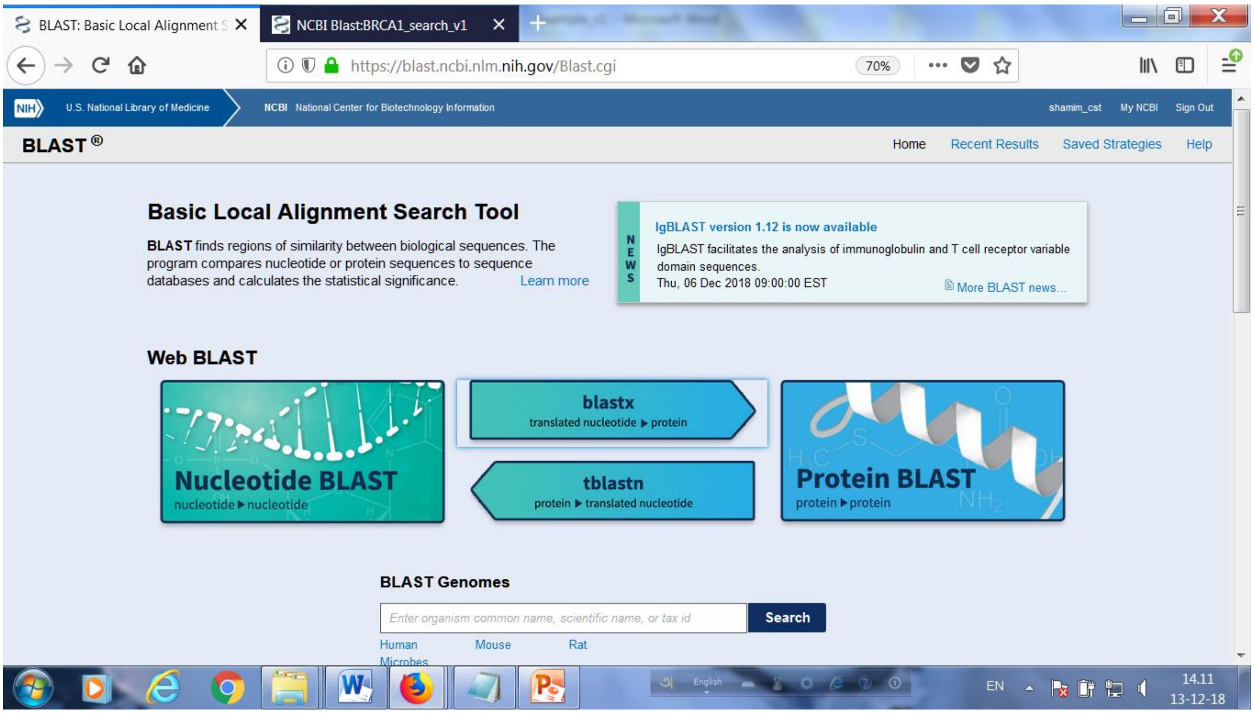
Gene ID: 672, updated on 9-Dec-2018

Summary

Official Symbol	BRCA1 provided by HGNC
Official Full Name	BRCA1, DNA repair associated provided by HGNC
Primary source	HGNC:HGNC:1100
See related	Ensembl:ENSG00000012048 MIM:113705
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	IRIS; PSCP; BRCAI; BRCC1; FANCS; PNCA4; RNF53; BROVCA1; PPP1R53

Table of contents

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Expression
- Bibliography
- Phenotypes
- Variation
- HIV-1 interactions
- Pathways from BioSystems
- Interactions
- General gene information
  - Markers, Related pseudogene(s), Potential



Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. Learn more

IgBLAST version 1.12 is now available

IgBLAST facilitates the analysis of immunoglobulin and T cell receptor variable domain sequences.

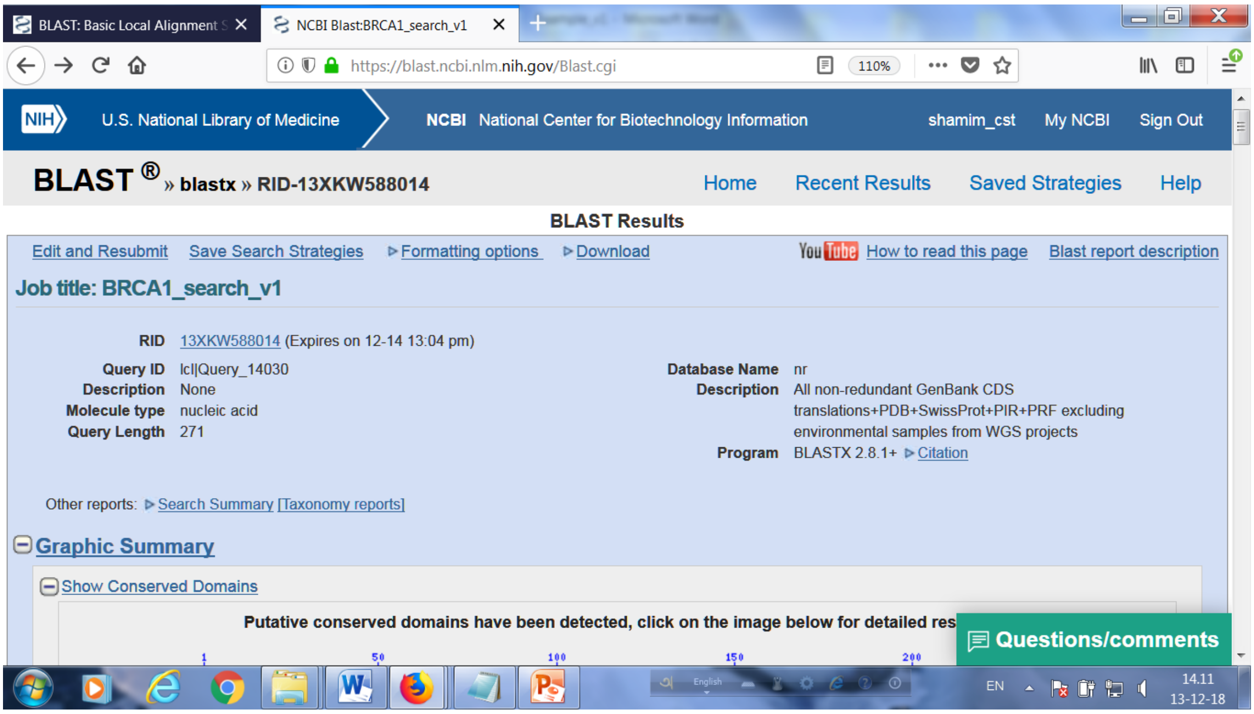
Thu, 06 Dec 2018 09:00:00 EST

More BLAST news...

Web BLAST

BLAST Genomes

Enter organism common name, scientific name, or tax id



BLAST® » blastx » RID-13XKW588014

BLAST Results

Job title: BRCA1_search_v1

RID 13XKW588014 (Expires on 12-14 13:04 pm)

Query ID lcl|Query_14030
Description None
Molecule type nucleic acid
Query Length 271

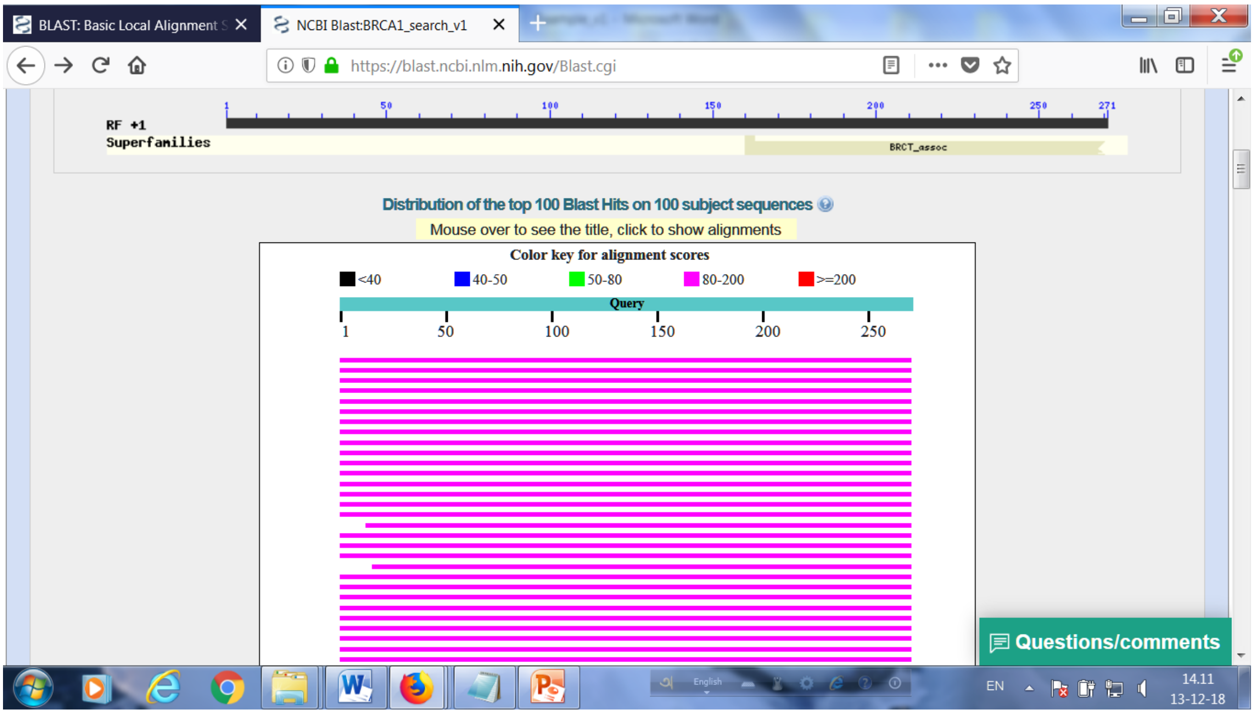
Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTX 2.8.1+ ▷ Citation

Other reports: ▷ Search Summary [Taxonomy reports]

Graphic Summary

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed res

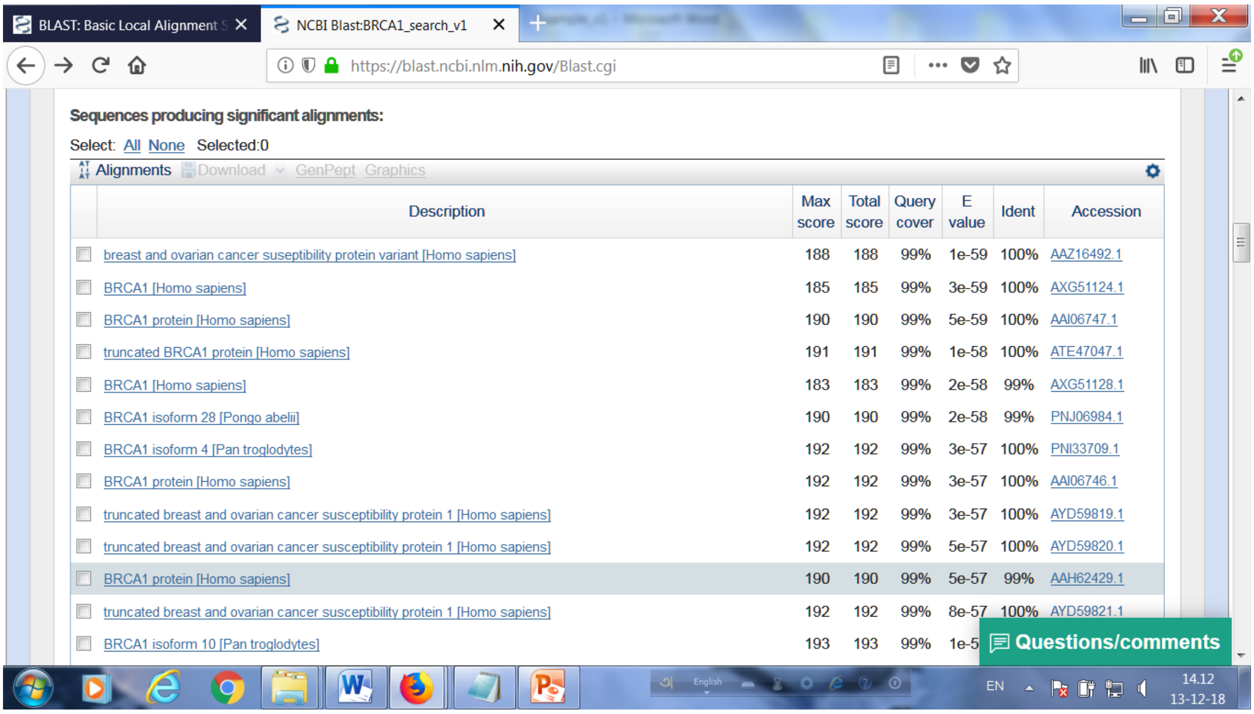


Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

■ <40 🟦 40-50 🟩 50-80 ■ 80-200 🟥 >=200

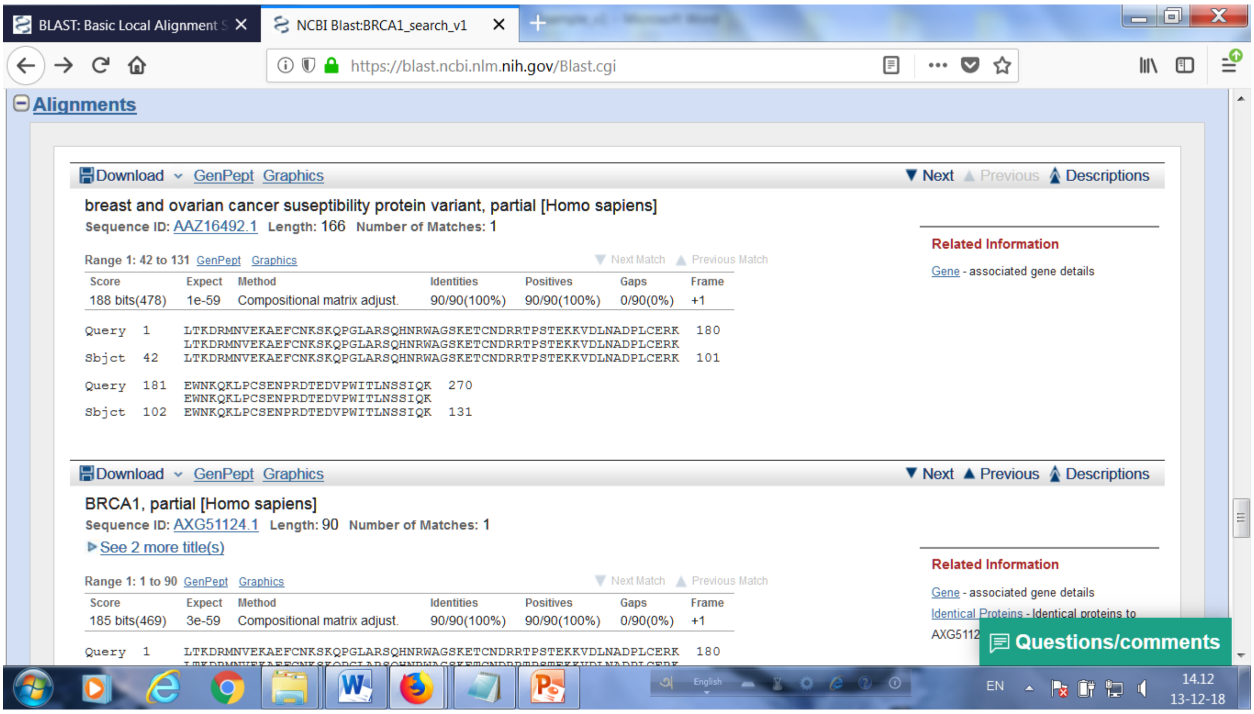


Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	breast and ovarian cancer suseptibility protein variant [Homo sapiens]	188	188	99%	1e-59	100%	AAZ16492.1
☐	BRCA1 [Homo sapiens]	185	185	99%	3e-59	100%	AXG51124.1
☐	BRCA1 protein [Homo sapiens]	190	190	99%	5e-59	100%	AAI06747.1
☐	truncated BRCA1 protein [Homo sapiens]	191	191	99%	1e-58	100%	ATE47047.1
☐	BRCA1 [Homo sapiens]	183	183	99%	2e-58	99%	AXG51128.1
☐	BRCA1 isoform 28 [Pongo abelii]	190	190	99%	2e-58	99%	PNJ06984.1
☐	BRCA1 isoform 4 [Pan troglodytes]	192	192	99%	3e-57	100%	PNI33709.1
☐	BRCA1 protein [Homo sapiens]	192	192	99%	3e-57	100%	AAI06746.1
☐	truncated breast and ovarian cancer susceptibility protein 1 [Homo sapiens]	192	192	99%	3e-57	100%	AYD59819.1
☐	truncated breast and ovarian cancer susceptibility protein 1 [Homo sapiens]	192	192	99%	5e-57	100%	AYD59820.1
☐	BRCA1 protein [Homo sapiens]	190	190	99%	5e-57	99%	AAH62429.1
☐	truncated breast and ovarian cancer susceptibility protein 1 [Homo sapiens]	192	192	99%	8e-57	100%	AYD59821.1
☐	BRCA1 isoform 10 [Pan troglodytes]	193	193	99%	1e-5		



breast and ovarian cancer suseptibility protein variant, partial [Homo sapiens]

Sequence ID: AAZ16492.1 **Length:** 166 **Number of Matches:** 1

Range 1: 42 to 131 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps	Frame
188 bits(478)	1e-59	Compositional matrix adjust.	90/90(100%)	90/90(100%)	0/90(0%)	+1

```
Query  1    LTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTEKKVDLNADPLCERK  180
            LTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTEKKVDLNADPLCERK
Sbjct  42   LTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTEKKVDLNADPLCERK  101

Query  181  EWNKQKLPCSENPRDTEDVPWITLNSSIQK  270
            EWNKQKLPCSENPRDTEDVPWITLNSSIQK
Sbjct  102  EWNKQKLPCSENPRDTEDVPWITLNSSIQK  131
```

Related Information

Gene - associated gene details

BRCA1, partial [Homo sapiens]

Sequence ID: AXG51124.1 **Length:** 90 **Number of Matches:** 1

See 2 more title(s)

Range 1: 1 to 90 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps	Frame
185 bits(469)	3e-59	Compositional matrix adjust.	90/90(100%)	90/90(100%)	0/90(0%)	+1

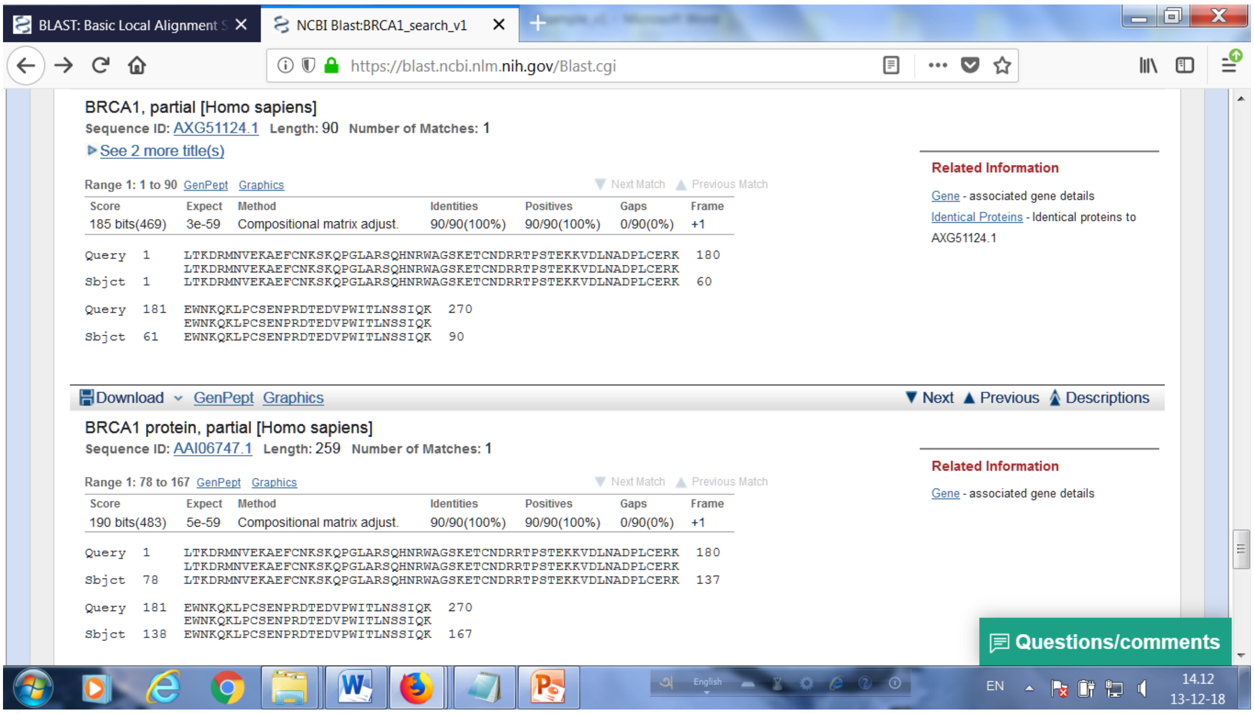
```
Query  1    LTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTEKKVDLNADPLCERK  180
```

Related Information

Gene - associated gene details

Identical Proteins - Identical proteins to

AXG5112



BRCA1, partial [Homo sapiens]

Sequence ID: AXG51124.1 **Length:** 90 **Number of Matches:** 1

See 2 more title(s)

Range 1: 1 to 90 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps	Frame
185 bits(469)	3e-59	Compositional matrix adjust.	90/90(100%)	90/90(100%)	0/90(0%)	+1

```
Query  1    LTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTEKKVDLNADPLCERK  180
            LTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTEKKVDLNADPLCERK
Sbjct  1    LTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTEKKVDLNADPLCERK  60

Query  181  EWNKQKLPCSENPRDTEDVPWITLNSSIQK  270
            EWNKQKLPCSENPRDTEDVPWITLNSSIQK
Sbjct  61   EWNKQKLPCSENPRDTEDVPWITLNSSIQK  90
```

Related Information

Gene - associated gene details

Identical Proteins - Identical proteins to AXG51124.1

BRCA1 protein, partial [Homo sapiens]

Sequence ID: AAI06747.1 **Length:** 259 **Number of Matches:** 1

Range 1: 78 to 167 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps	Frame
190 bits(483)	5e-59	Compositional matrix adjust.	90/90(100%)	90/90(100%)	0/90(0%)	+1

```
Query  1    LTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTEKKVDLNADPLCERK  180
            LTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTEKKVDLNADPLCERK
Sbjct  78   LTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTEKKVDLNADPLCERK  137

Query  181  EWNKQKLPCSENPRDTEDVPWITLNSSIQK  270
            EWNKQKLPCSENPRDTEDVPWITLNSSIQK
Sbjct  138  EWNKQKLPCSENPRDTEDVPWITLNSSIQK  167
```

Related Information

Gene - associated gene details