

Using BLAST to Compare DNA & Protein Sequences:

Cancer Biology & *BRCA1* Genetic Testing

Understanding *BRCA1* & Cancer

NWABR.ORG
Northwest Association for Biomedical Research

BRCA1 and Breast Cancer



Glossary

DNA
RNA
Genes
Chromosome
Tumor
suppressor
Mutation

Welcome to the BRCA1 gene and protein tutorial.

Click on the button "The function of BRCA1" to begin.
You can also access definitions in the glossary.

- ☐ The function of BRCA1
- ☐ Mutations in BRCA1 Can Cause Cancer

https://nwabr.org/sites/default/files/BRCA1_animation.swf

How Information from DNA is Acquired and Used for Genetic Testing

Patient Sample: Blood or Saliva



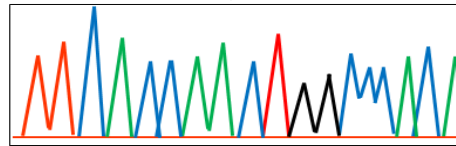
Genetic Counselors work with patients to help them decide whether to have a genetic test, and help them understand the results of the test.

Extract DNA from Cells



Lab Technicians work with patient samples in the lab, purifying and sequencing the DNA.

Sequence DNA



...TTCACCAACAGGCCACACA...

Computational Biologists create computer programs to help biologists analyze data.

Compare Patient DNA Sequence to Reference Sequence

Reference	...TTCACCAAC	ATG	CCCACA...
	F	T	N M P T
Patient	...TTCACCAAC	AGG	CCCACA...
	F	T	N R P T

Biomedical Researchers perform experiments with patient samples to find different variations of genes that might cause disease.

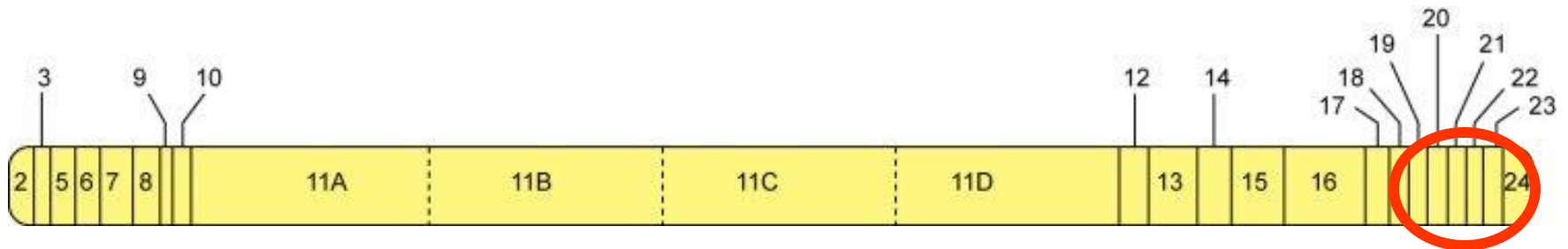
Search Database to Determine whether Patient Mutation is Associated with Disease



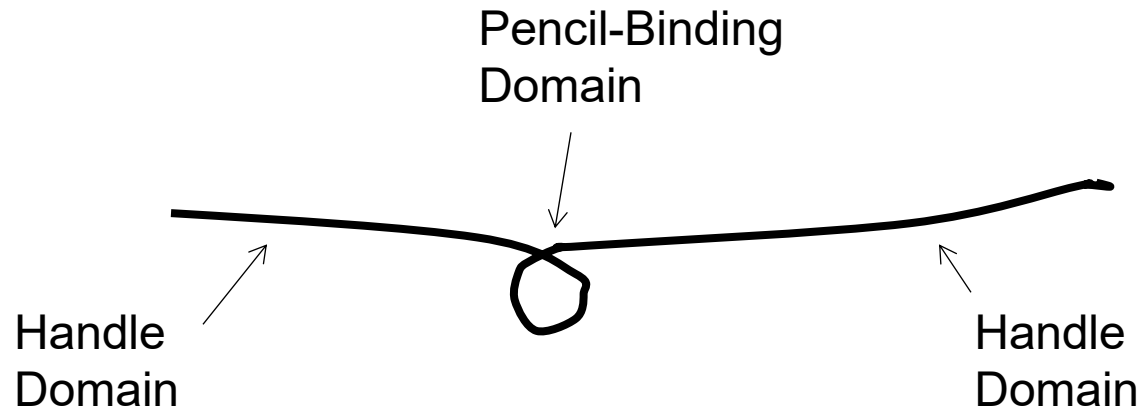
Bioengineers help create tools and treatments for patients with various diseases. *Medical Doctors* and *Veterinarians* use the knowledge gained from genetic testing to care for their patients.

The *BRCA1* Gene

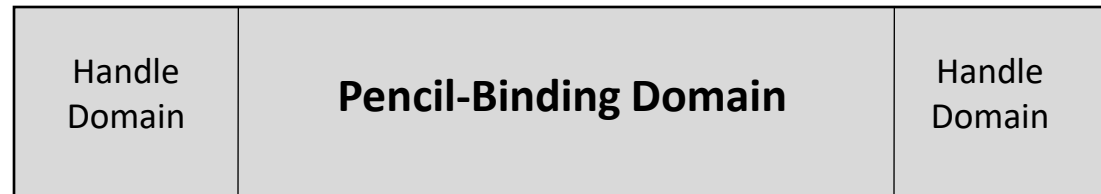
ATGGATTTATCTGCTCTTCGCGTTGAAGAAGTACAAAATGTCATTAATGCTATGCAGAAAATCTTAGAGTGTCCCATCTGTCTGGAGTTGATCAAGGAACCTGTCTCCACAAAGTGTGACCACAT ATTTTGCAAATTTTGCATGCTGAACTTCTCAACCA
GAAGAAAGGGCCTTCACAGTGTCTTTATGTAAGAAATGATATAACCAAAAGGAGCCTACAGAAAGTACGAGATTTAGTCAACTTGTGTAAGAGCTATTAAGAAATCATTGTGCTTTTCAGCTTGACACAGGTTTGAGGATGCAAAACAGCTATAATTTTG
CAAAAAGGAAAAATAACTCTCCTGAACATCTAAAAGATGAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCCAAAAGACTTCTACAGAGTGAACCCGAAAAATCCTTCTCTGCAGGAA ACCAGTCTCAGTGTCCAACCTCTCAACCTTGAAGT
GTGAGAACTCTGAGGCAAAAGCGGATACAACTCTAAAAGAGCTGTCTGTCAATTGAATTGGGATCTGATTTCTCTGAAGATACCGTTAATAAGGCACTTATTGCAGTGTGGGAGATCAAGA ATTTGTTACAAATCACCCCTCAAGGAACCGGATGA
AATCAGTTTGGATCTGCAAAAAAGGCTGCTTGTGAATTTTCTGAGACGGATGTAAACAATACTGAACATCATCAACCCAGTAATAATGATTGTAACACCCTGAGAAGCGTGCAGCTGAGAGGC ATCCAGAAAAGTATCAGGGTAGTTCTGTTTCAACT
TGCATGTGGAGCCATGTGGCACAATACTCATGCCAGCTCATTACAGCATGAGAACAGCAGTTTATTACTCACTAAAGACAGAATGAATGTAGAAAAAGGCTGAATTTCTGTAATAAAAGCAACAG CTTGGCTTAGCAAGAGGCCAACATAACAGATGGGCT
GGAAAGTAAGGAAACATGTAATGATAGGCGGACTCCAGCACAGAAAAAAGGTAGATCTGAATGCTGATCCCTGTGTGAGAGAAAAAGAAATGAATAAGCAGAAACTGCCATGCTCAGAGAATCC TAGAGATACTGAAGATGTTCTTGGATAACACTAAA
TAGCAGCATTCAGAAAGTTAATGAGTGGTTTTCCAGAAGTGATGAAGTGTAGGTTCTGATGACTCACATGATGGGAGTCTGAATCAAAATGCCAAAGTAGTGTATGATTGGACGTTCTAAATG AGGTAGATGAATATTCTGGTTCTCAGAGAAAAATAG
ACTTACTGGCCAGTGTCTCATGAGGCTTAAATATGTAAGGTGAAGAGTTCACCTCCAATCAGTAGAGAGTAATATTGAAGACAAAAATTTGGGAAAAACCTATCGGAAGAAGGCCAAGCCTC CCCAATCTAAGCCATGTAAGTAAAAATCTAATTATA
GGAGCATTTGTTACTGAGCCACAGATAATACAAGAGCGTCCCTCAACAATAAATAAAGCGTAAAAGGAGACCTACATCAGGCTTCTATCCTGAGGATTTTATCAAGAAAGCAGATTGTCAGT TCAAAAGACTCCTGAAATGATAAATCAGGGAACATA
CCAAACGGAGCAGAAATGGTCAAGTATGTAATATTACTAATAGTGTGATGAGAAATAAAACAAAGGTGATTCTATTCAAGATGAGAAAAATCCTAACCCAATAGAAATCACTCGAAAAAGAAATCTG CTTTCAAAACGAAAGCTGAACCTATAAGCAGCAGTA
TAAGCAATATGGAATCGAATTAATATCCACAATTCAAAAGCACCTAAAAGAAATAGGCTGAGGAGGAAGTCTTCTACAGGCATATTCTATGCGCTTGAACATAGTAGT CAGTAGAAAACTAAGCCACCTAATGTACTGAATTGCAAAATGATAGTTGT
TCTAGCAGTGAAGAGATAAAGAAAAAAGTACAACCAATGCCAGTCAGGCACAGCAGAAACCTACAACATCATGGAAGGTAAAGAACCTGCAACTGGAGCCAAGAAGAGTAACAAGCCAAATGA ACAGACAAGTAAAAGACATGACAGCGTACTTTCCC
AGAGCTGAAGTTAACAAATGCACCTGGTTCTTTTACTAAGTGTTCAAATACCACTGAACCTTAAAGAAATTTGTCAATCCTAGCCTTCCAAGAGAGAAAAAGAGAGAACTAGAAACAGT TAAAG TGTCTAATAATGCTGAAGACCCCAAGATCTCATGT
TAAGTGGAGAAAGGTTTTGCAAACTGAAAGATCTGTAGAGAGTAGCAGTATTTTCTTGGTACCTGGTACTGATTATGGCACTCAGGAAAGTATCTCGTTACTGGAAGTTAGCACTCTAGGGAAG GCAAAACAGAACCAATAAATGTGTGAGTCACTGT
GCAGCATTTGAAAGACCCCAAGGAGTAAATCATGGTTGTTCCAAAGATATAGAAATGACACAGAAGGCTTTAAAGTATCCATTGGGACATGAAGTTAACCACAGTCGGGAAACAGCATAGAAAT GGAAGAAAGTGAACCTGTATGCTCAGTATTTGCAGAA
TACATTCAAGGTTTTCAAGAGCCGAGTCAATTTGCTCCGTTTTCAATCCAGGAAATGCAGAAAGAGGAATGTGCAACATTTCTGCCCACCTCTGGGTCTTAAAGAAACAAAGTCCAAAGTCACTT TGAATGTGCAACAAAGGAGAAAACTCAGGAAAGA
ATGAGTCTAATATCAAGCCTGTACAGACAGTTAATATCACTGCAGGCTTTCTGTGGTTGGTCAGAAAGATAAGCCAGTTGATAATGCCAAATGTAGTATCAAAAGGAGGCTCTAGGTTTTGTCTA TCATCTCAGTTCAGAGGCAACGAAACTGGACTCATT
ACTCCAAATAAACATGGACTTTTACAAAACCCATATCGTATACCACCACTTTTTCCCATCAAGTCATTTGTTAAACTAAATGTAAGAAAAATCTGCTAGAGGAAAACTTTGAGGAACATCAAT GTCACCTGAAAGAGAAATGGGAAATGAGAACATTC
AAGTACAGTGAGCACAATTAGCGTAATAACATTAGAGAAAAATGTTTTAAAGAGCCAGCTCAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGGCTCCAGTATTAATGAATAG GTTCCAGTGTAGAAAACATTCAAGCAGAACTAGSTA
GAAACAGAGGGGCAAAATGGAATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGCTTATAAACAAAGTCTTCCCTGGAAGTAAATGTAAGCATCCTGAAATAAAAAAGCAAGAATATGAAGAA TAGTTTCAGATGTTAAATACAGATTTCTCTCATAT
CTGATTTAGATAAATTAGAACAGCCTATGGGAAGTAGTCATGCATCTCAGGTTTGTCTGAGACACCTGATGACCTGTTAGATGATGGTGAATAAAGGAAGATACTAGTTTTGCTGAAAAATGA CATTAAGGAAAGTTCTGCTGTTTTAGCAAAAGCGT
CCAGAAAGGAGAGCTTAGCAGGAGTCTAGCCCTTTACCCATACACATTTGGCTCAGGTTTACCAGAGAGGGGCCAAGAAATAGAGTCTCAGAAAGAACTTATCTAGTGAGGATGAAGAGC TTCCCTGCTTCCAACACTTGTATTTTGGTAAAGTAA
ACAATATACCTTCTAGCTTACTAGGCATAGCACCGTTGTACCGAGGTGTCTGTCTAAGAACACAGAGGAGAAATTTATTAATGAAAGATGCTTAAATGACTGCAGTAACCCAGGTAATATTG CAAAGGAGCATCTCAGGAACATCACTGTGAGGAA
ACAAAATGTTCTGCTAGCTTGTCTTCTCAGTGCAGTGAATTTGGAAGACTTGACTGCAATACAAACACCCAGGATCCTTTCTTGATTGGTTCTTCCAACAAATGAGGCATCAGTCTGAAG CCAGGAGTGGTCTGAGTGACAAGGAATTTGGTTTC
AGATGATGAAGAAAGAGGAACGGGCTTGGAAGAAAAATCAAGAAAGAGCAAGCATGGATTCAAACCTTAGGTGAAGCAGCATCTGGGTGTGAGAGTGAACCAAGCGTCTCTGAAGACTGCTCAG GGCTATCTCTCAGAGTGACATTTTAACCACTCAGC
AGAGGATACCATGCAACATAAAGCTGATAAAGCTCCAGCAGGAAATGGCTGAACAGAGCTGTGTTAGAACAGCATGGGAGCCAGCCTTCTAACAGCTACCTTCCATCATAAAGTGACTCTTCTGCCCTTGAGGACCTGCGAAATCCAGAACAGCACA
TCAGAAAAAGCAGTATTAACCTCACAGAAAAAGTAGTGAATACCTATAAGCCAGAAATCCAGAAAGGCTTTCTGCTGACAAGTTTGAGGTGTCTGCAGATAGTTCTACCAGTAAAAATAAAGAAC AGGAGTGGAAAGGTATCCCTTCTAAATGCCCATC
ATTAGATGATAGGTGATGATGCACAGTGTCTCTGGGAGTCTTCAGAAATAGAACTACCCATCTCAAGCAGGAGCTGATTAAGCTTCTGAGGAGCAACAGCTGGAAGAGTCTGGGCCAC ACGATTTGACGGAACATCTTACTTGCCAAAGGCAAG
ATCTAGAGGGAACCCCTTACCTGCAATTTGCAATGACCTTCTCTGATGACCTGATCTGTTCTGGAAGACAGAGCCAGCAGTCACTGTTGTCGCAACATACCATCTTCAACCTCTGCAATTAAGCTTCTCCCAATTGAAAGTTGCAGAACTC
GCCAGAGCTGCACTGCTGCTCATACTACTGATGACTGCTGGGTATAATGGAAGAAAGTGTGAGCAGGGGAGGCAAGCAATTTAGTACAGCTTCAACAGAAAGGGTCAACAAAGAAATGTCCAT GGTGGTGTCTGGCCTGACCCAGAAATTTATGCT
GTTGTACAAGTTTGCCAGAAAAACACCACATCACTTTAACTAATCTAATTACTGAAGAGCACTACTCATGTTGTTATGAAAAACAGATGCTGAGTTTGTGTGTGAACGGCACTGAAATATTTCTAG GAATTGCGGGAGGAAAAATGGGTAGTTAGCTATTTC
GGGTGACCCAGCTATTATTAAGAAAGAAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGGAGATGTGGTCAATGGAAGAAACCAAGGTCCAAAGCAGCAGAGAAAGATCCAGGACAGAAAG ATCTTCAGGGGGCTAGAAATCTGTTGCTATGGGCC
TTCAACCAATGCCACAGATCACTGGAATGGATGGTACAGCTGTGTGTGCTCTGTGGTGAAGGAGCTTTCATCATCCCTGGCAGAGGTGTCACCCCAATGTGGTTGTGTCAGCCAGA TGCCCTGGACAGAGGCAATGGCTTCCATGCAATTTGG
GCAGATGCTGAGGCACCTGTGGTGACCCGAGAGTGGGTGTTGGACAGTGTAGCACTCTACCAGTGCCAGGAGCTGGACACCTACCTGATACCCAGATCCCCACAGCCACTACTGA



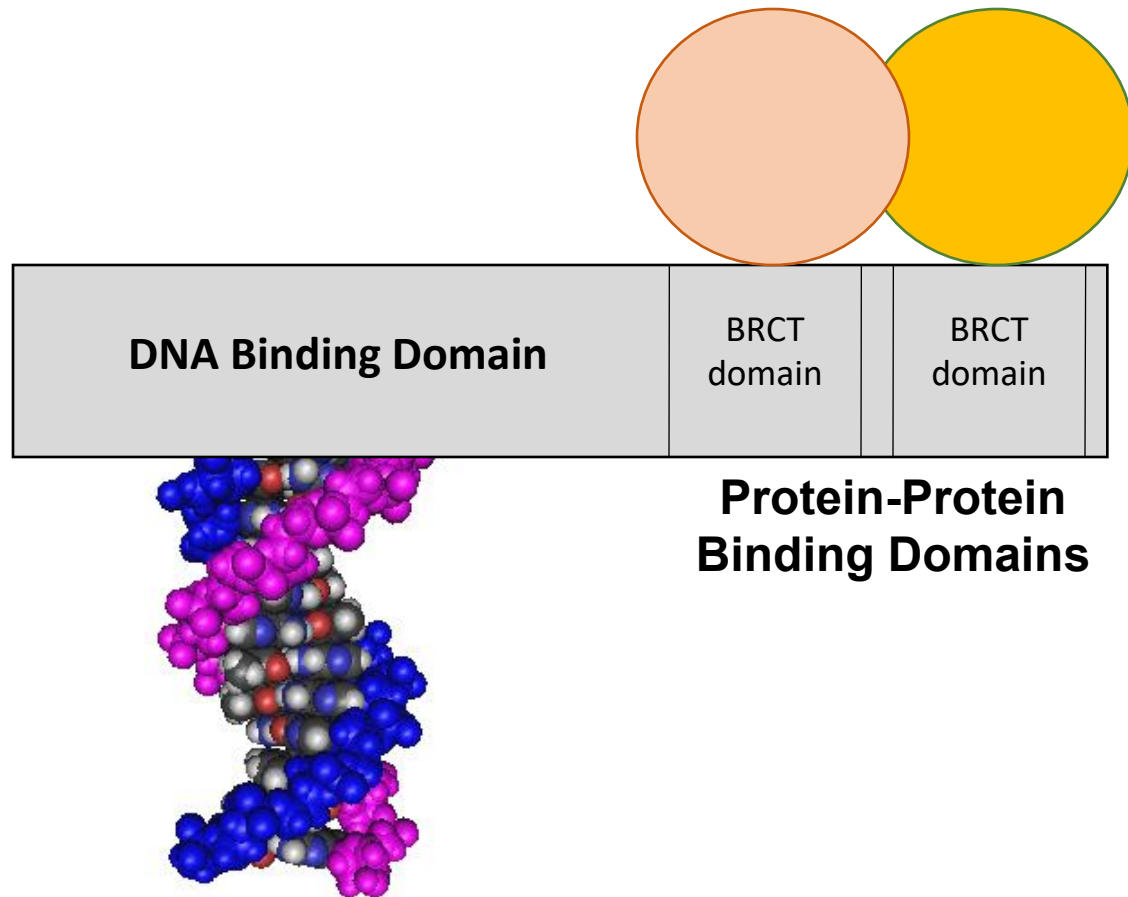
Pencil Transferase Domains

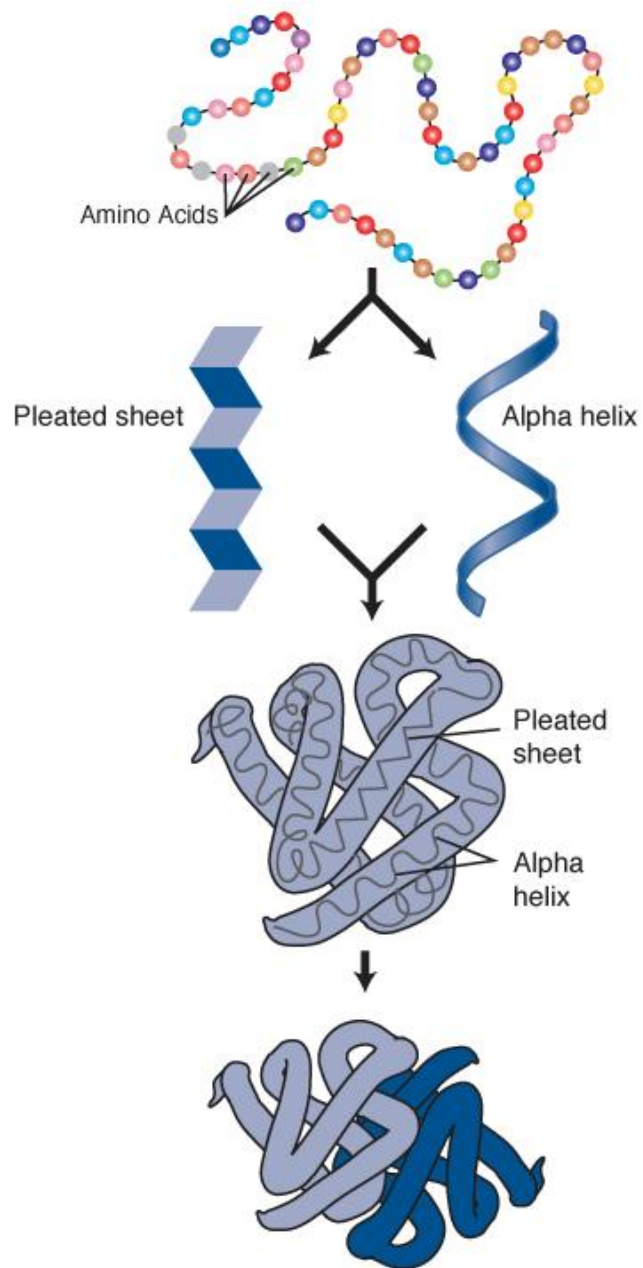


*Linear
Structure:*



BRCA1 Protein Domains





Levels of protein organization

Primary protein structure

is sequence of a chain of amino acids

Secondary protein structure

occurs when the sequence of amino acids are linked by hydrogen bonds


Tertiary protein structure

occurs when certain attractions are present between alpha helices and pleated sheets.

Quaternary protein structure

is a protein consisting of more than one amino acid chain.

Basic Local Alignment Search Tool

 U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

BLAST®

Home Recent Results


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](#)

NEWS

QuickBLASTP
Try **QuickBLASTP** for a fast protein search of nr.
Tue, 23 May 2017 13:00:00 EST
[More BLAST news...](#)

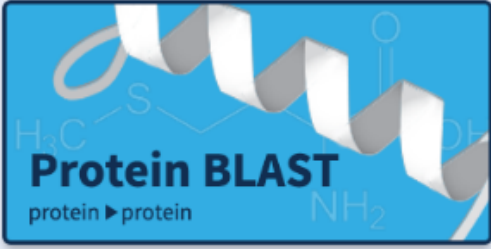
Web BLAST



Nucleotide BLAST
nucleotide ► nucleotide

blastx
translated nucleotide ► protein

tblastn
protein ► translated nucleotide



Protein BLAST
protein ► protein

BLAST Genomes

Human Mouse Rat Microbes

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

A Tool for Comparing Primary Sequence Information

Compare Two or More
Sequences to One Another

Compare a Sequence “Query”
to an NCBI Database

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [Graphics](#)

☐ [Father BRCA1 DNA Sequence](#)

[Alignments](#)

		Description
Query 1	GTGTACAAGTTTGCAGAAAAACACCATCACTTTAACTAATCTAATTACTGAAGAGACT	60
49893 1	60
Query 61	ACTCATGTTGTTATGAAAACAGATGCTGAGTTTGTGTGTGAACGGACACTGAAATATTTT	120
49893 61	120
Query 121	CTAGGAATTCGGGAGGAAAATGGGTAGTTAGCTATTTCTGGGTGACCCAGTCTATTAAA	180
49893 121	180
Query 181	GAAAGAAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGGAGATGTGTCATGGAAGA	240
49893 181	240
Query 241	AACCACCAAGGTCCTCAAGCGAGCAAGAGAATCCCAGGACAGAAAGATCTTCAGGGGGCTA	300
49893 241	300
Query 301	GAAATCTGTTGCTATGCGCCCTTCACCAACATGCCACAGATCAACTGGAATGGATGGTA	360
49893 301	360
Query 361	CAGCTGTGTGGTGCTTCTGTGGTGAAGGAGCTTTCATCAITCACCTTGGCAGAGTGTG	420
49893 361	420
Query 421	CACCCAATTGTGTTGTGTCAGCCAGATGCCTGACAGAGGACAATGGCTTCCATGCAATT	480
49893 421	480
Query 481	GGGCAGATGTGTGAGGCACCTGTGTTGACCCGAGAGTGGGTGTTGGACAGTGTAGCACTC	540
49893 481	540
Query 541	TACCAAGTCCAGGAGCTGGACACCTACCTGATACCCAGATCCCCACAGCCACTACTGA	600
49893 541	600

BLAST Basic Local Alignment Search Tool

Query ID: [id11559](#) Database Name: [nr](#)

Description: [None](#) Description: [All GenBank+EMBL+DDBJ+PDB sequences \(but not EST, STS, GSS, environmental samples or phage C, I or L HTGS sequences\)](#)

Molecule type: [nucleic acid](#) Program: [BLASTN 2.2.3+](#) [Blast](#)

Query Length: [1542](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Sequence logos of results](#)

[Graphic Summary](#)

Dist. of 100 Blast Hits on the Query Sequence (X)

Mouse over to see the details, click to show alignments

Color key for alignment scores

Query 0 300 600 900 1200 1500

Descriptions

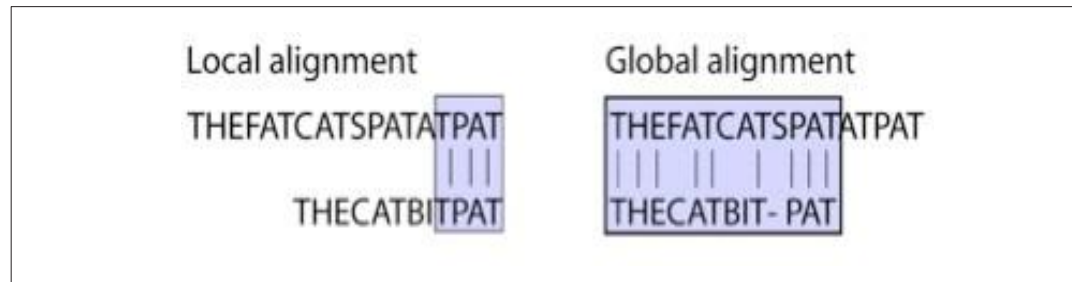
Legend for links to other resources: [UniProt](#) [GenBank](#) [EMBL](#) [DDBJ](#) [PDB](#) [RefSeq](#) [Ensembl](#) [NCBI](#)

Accession	Description	Hit #	Score	Query Coverage	E-value	Max. Match	Uniq.
G00001	Gorilla gorilla mitochondrial DNA, complete genome	236	2762	100%	0.0	100%	
G00001	Gorilla gorilla mitochondrial DNA, complete genome	232	2772	100%	0.0	99%	
G00001	Gorilla gorilla mitochondrial genome	212	2715	100%	0.0	99%	
G00001	Pan troglodytes chromosome 7 clone G001-275010, complete sequence	2167	2167	99%	0.0	91%	
G00001	Pan troglodytes BAC clone G001-250214 from chromosome 7, complete sequence	2128	2127	99%	0.0	91%	
G00001	Pan paniscus mitochondrial DNA, complete sequence	2128	2129	99%	0.0	90%	
G00001	Pan troglodytes valisus isolate 102 H0412 mitochondrial, complete genome	2145	2144	99%	0.0	90%	
G00001	Pan troglodytes troglodytes isolate 102 H0412 mitochondrial, complete genome	2111	2111	99%	0.0	90%	

Terminology

- **Query Sequence:**
 - Same root as “question”
 - Sometimes called a “reference sequence”
 - The sequence to which other sequences are compared
 - Independent or control variable.
- **Subject Sequence:**
 - The sequence being compared
 - Dependent or experimental variable
- **BLAST Scores**
 - **Max Score, Total Score**
 - **Query Coverage**
 - **Percent Identity**

The “Local” in Local Search Tool



- Sequences are broken into “words”
- Defaults vary by program (11-28 bases or amino acids)
- BLAST looks for word matches, then tries to extend those words
- Proteins:
 - BLAST Scoring Table contains the probability of each amino acid being replaced with another amino acid (ex: ELVIS vs. ELVES)
- Nucleotides:
 - Plus points where Nucleotides match, minus points where they do not match [default is adjustable]

How to BLAST

Comparing Two or More Sequences to One Another

This screenshot shows the 'Align Sequences Nucleotide BLAST' interface. It features two main input sections: 'Enter Query Sequence' and 'Enter Subject Sequence'. Each section has a text area for 'Enter accession number(s), gi(s), or FASTA sequence(s)', a 'Clear' button, and a 'Query subrange' or 'Subject subrange' section with 'From' and 'To' input fields. Below these are options to 'Or, upload file' with a 'Browse...' button and a 'Job Title' field. A checkbox labeled 'Align two or more sequences' is checked. At the bottom, there is a 'Program Selection' section with radio buttons for 'Optimize for' (Highly similar sequences (megablast), More dissimilar sequences (discontiguous megablast), Somewhat similar sequences (blastn)) and a 'BLAST' button. A link for 'Algorithm parameters' is at the bottom left.

Compare a Sequence “Query” to an NCBI Database

This screenshot shows the 'Basic Local Alignment Search Tool' interface. It includes a navigation bar with 'Home', 'Recent Results', 'Saved Strategies', and 'Help'. The main section is titled 'NCBI BLAST/ blastn suite'. It has tabs for 'blastn', 'blastp', 'blastx', 'tblastn', and 'tblastx'. The 'Enter Query Sequence' section includes a text area for 'Enter accession number, gi, or FASTA sequence', a 'Clear' button, and a 'Query subrange' section with 'From' and 'To' input fields. Below this are options to 'Or, upload file' with a 'Browse...' button and a 'Job Title' field. A checkbox labeled 'Align two or more sequences' is unchecked. The 'Choose Search Set' section includes a 'Database' dropdown menu (currently showing 'Nucleotide collection (nr/nt)'), an 'Organism' dropdown menu (currently showing 'Human genomic plus transcript'), and an 'Exclude' checkbox. The 'Program Selection' section includes a 'Program Select' dropdown menu (currently showing 'Nucleotide collection (nr/nt)') and an 'Optimize for' section with radio buttons for 'Highly similar sequences (megablast)', 'More dissimilar sequences (discontiguous megablast)', and 'Somewhat similar sequences (blastn)'. A 'BLAST' button is at the bottom.

Available Genomes & Databases

BLAST Assembled Genomes

Find Genomic BLAST pages:

Enter organism name or id—completions will be suggested

GO

☐ [Human](#)
☐ [Mouse](#)
☐ [Rat](#)
☐ [Cow](#)
☐ [Pig](#)
☐ [Dog](#)

☐ [Rabbit](#)
☐ [Chimp](#)
☐ [Guinea pig](#)
☐ [Fruit fly](#)
☐ [Honey bee](#)
☐ [Chicken](#)

☐ [Zebrafish](#)
☐ [Clawed frog](#)
☐ [Arabidopsis](#)
☐ [Rice](#)
☐ [Yeast](#)
☐ [Microbes](#)

The screenshot shows the NCBI BLAST web interface. At the top, there's a navigation bar with 'Home', 'Recent Results', 'Saved Strategies', and 'Help'. Below this, the 'NCBI/BLAST/blastn suite' is selected. The 'Standard Nucleotide BLAST' dropdown menu is open, displaying a list of databases. The 'Nucleotide collection (nr/nt)' is highlighted in blue. Other databases listed include 'Genomic plus Transcript', 'Human genomic plus transcript (Human G+T)', 'Mouse genomic plus transcript (Mouse G+T)', 'Reference RNA sequences (refseq_ma)', 'Reference genomic sequences (refseq_genomic)', 'RefSeq Representative genomes (refseq_representative_genomes)', 'NCBI Genomes (chromosome)', 'Expressed sequence tags (est)', 'Genomic survey sequences (gss)', 'High throughput genomic sequences (HTGS)', 'Patent sequences(pat)', 'Protein Data Bank (pdb)', 'Human ALU repeat elements (alu_repeats)', 'Sequence tagged sites (dbsts)', 'Whole-genome shotgun contigs (wgs)', 'Transcriptome Shotgun Assembly (TSA)', '16S ribosomal RNA sequences (Bacteria and Archaea)', and 'Sequence Read Archive (SRA)'. The 'Nucleotide collection (nr/nt)' is also shown at the bottom of the dropdown menu.

BLAST Results Homepage

blast.ncbi.nlm.nih.gov/Blast.cgi

An official website of the United States government [Here's how you know](#)

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST® » blastn suite-2sequences » results for RID-8G1THAX3114

Home Recent Results Saved Strategies Help

[← Edit Search](#) [Save Search](#) [Search Summary](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Job Title BRCA1_Reference_DNA_Sequence

Program Blast 2 sequences [Citation](#)

Query ID lc|Query_19059 (dna)

Query Descr BRCA1_Reference_DNA_Sequence

Query Length 600

Subject Descr Deborah_BRCA1_DNA_Sequence

Subject Length 600

Other reports [MSA viewer](#)

Filter Results

Percent Identity to

E value to

Query Coverage to

[Filter](#) [Reset](#)

Descriptions [Graphic Summary](#) [Alignments](#) [Dot Plot](#)

Sequences producing significant alignments [Download](#) [Select columns](#) [Show](#) 100 [?](#)

☒ select all 1 sequences selected [Graphics](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
-------------	-----------------	-----------	-------------	-------------	---------	------------	----------	-----------

Feedback

BLAST Results: Descriptions & Scores

Subject Descr [See details ▼](#)

Subject 2807

Length

Descriptions

Graphic Summary

Alignments

Sequences producing significant alignments

Download ▼

Select columns ▼

Show 100 ▼



☒ select all 6 sequences selected

[Graphics](#)

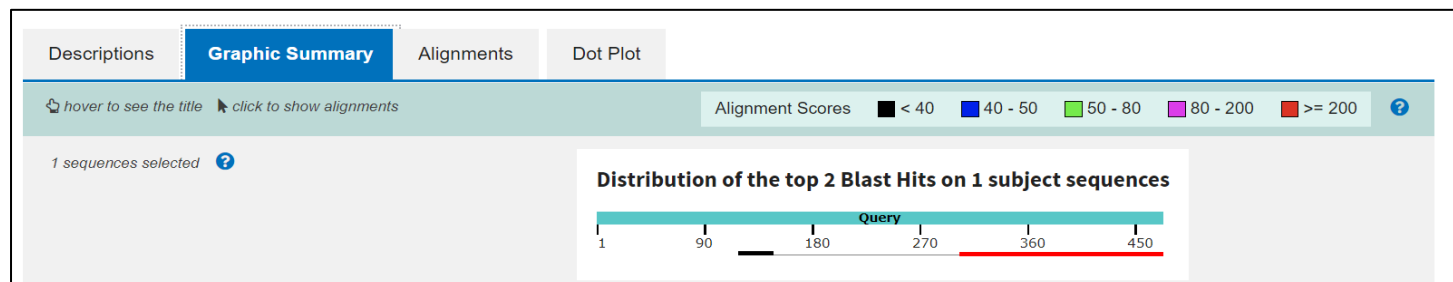
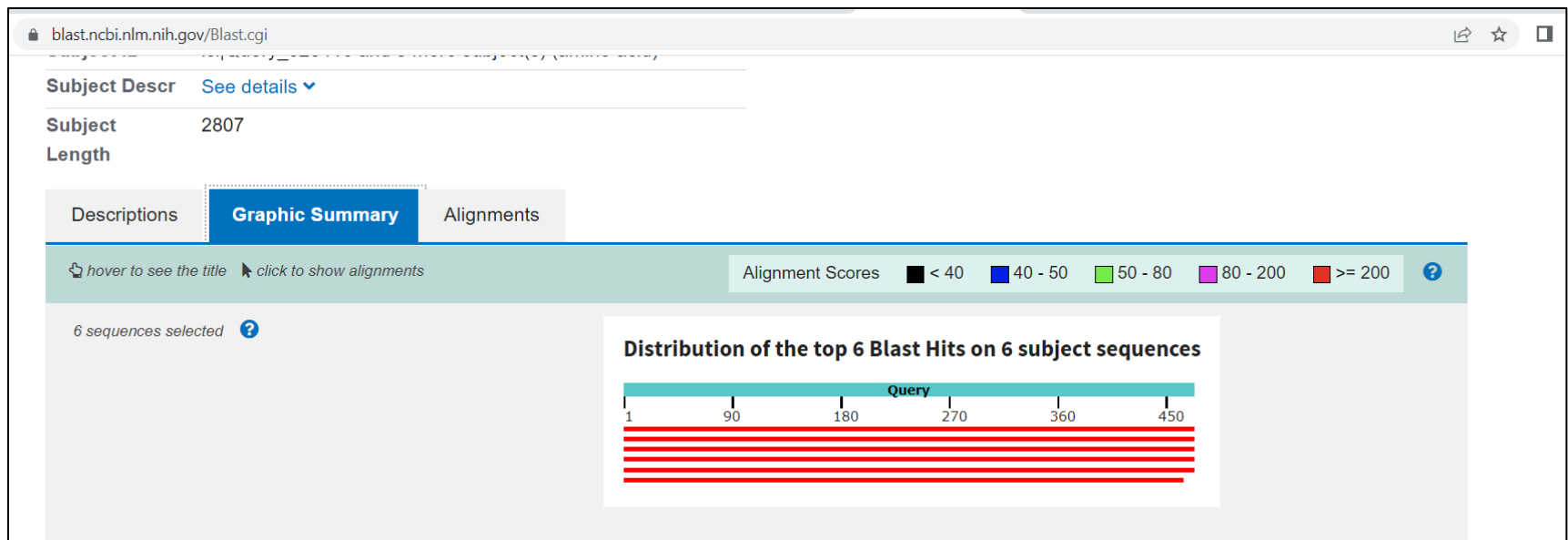
[Distance tree of results](#)

[Multiple alignment](#)

[MSA Viewer](#)

	Description ▼	Scientific Name ▼	Max Score ▼	Total Score ▼	Query Cover ▼	E value ▼	Per. Ident ▼	Acc. Len ▼	Accession
<input checked="" type="checkbox"/>	Patient A17		956	956	100%	0.0	99.79%	469	Query_325421
<input checked="" type="checkbox"/>	Patient A13		956	956	100%	0.0	99.79%	469	Query_325420
<input checked="" type="checkbox"/>	Patient A12		956	956	100%	0.0	99.79%	469	Query_325419
<input checked="" type="checkbox"/>	Patient A11		956	956	100%	0.0	99.79%	469	Query_325418
<input checked="" type="checkbox"/>	A/California/07/2009(H1N1)-2011-2016 Vaccine		941	941	100%	0.0	97.01%	469	Query_325416
<input checked="" type="checkbox"/>	A/Brisbane/59/2007(H1N1)-2008-2010 Vaccine		773	773	98%	0.0	80.74%	462	Query_325417

BLAST Results: Graphic Summary – Multiple Sequence Alignment of Sample Sequences



BLAST Results: Graphic Summary – Identifying a Sequence in the NCBI Database

blast.ncbi.nlm.nih.gov/Blast.cgi

An official website of the United States government [Here's how you know](#)

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST® » blastp suite » results for RID-6AZ66FHP013

Home Recent Results Saved Strategies Help

[Edit Search](#) Save Search Search Summary ▾

How to read this report? BLAST Help Videos Back to Traditional Results Page

Job Title A/Michigan/45/2015 (H1N1)-2017-Present

RID 6AZ66FHP013 Search expires on 04-26 00:37 am [Download All](#) ▾

Program BLASTP [Citation](#) ▾

Database nr [See details](#) ▾

Query ID IclQuery_87755

Description A/Michigan/45/2015 (H1N1)-2017-Present Vaccine

Molecule type amino acid

Query Length 469

Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) ?

Filter Results

Organism only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[Add organism](#)

Percent Identity to **E value** to **Query Coverage** to

[Filter](#) [Reset](#)

Compare these results against the new Clustered nr database ? [BLAST](#)

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download ▾ Select columns ▾ Show 100 ▾ ?

☒ select all 100 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

	Description ▾	Scientific Name ▾	Max Score ▾	Total Score ▾	Query Cover ▾	E value ▾	Per. Ident ▾	Acc. Len ▾	Accession ▾
<input checked="" type="checkbox"/>	neuraminidase [Influenza A virus (A/India/Pun151245/2015(H1N1))]	Influenza A virus (A/India/Pun151245/2015(H1N1))	957	957	100%	0.0	100.00%	469	AKE37534.1
<input checked="" type="checkbox"/>	neuraminidase [Influenza A virus (A/North Dakota/01/2016(H1N1))]	Influenza A virus (A/North Dakota/01/2016(H1N1))	957	957	100%	0.0	99.79%	469	ANE27703.1
<input checked="" type="checkbox"/>	neuraminidase [Influenza A virus (A/Sistan/57/2015(H1N1))]	Influenza A virus (A/Sistan/57/2015(H1N1))	957	957	100%	0.0	99.79%	469	AMA65394.1
<input checked="" type="checkbox"/>	neuraminidase [Influenza A virus (A/Montana/03/2016(H1N1))]	Influenza A virus (A/Montana/03/2016(H1N1))	957	957	100%	0.0	99.79%	469	ANE27769.1
<input checked="" type="checkbox"/>	neuraminidase [Influenza A virus (A/Viroinia/15/2016(H1N1))]	Influenza A virus (A/Viroinia/15/2016(H1N1))	957	957	100%	0.0	99.79%	469	ANM90834.1
<input checked="" type="checkbox"/>	neuraminidase [Influenza A virus (A/India/DRDE_GWL989/2015(H1N1))]	Influenza A virus (A/India/DRDE_GWL989/2015(H1N1))	957	957	100%	0.0	99.79%	469	ANC28534.1
<input checked="" type="checkbox"/>	neuraminidase [Influenza A virus (A/Indiana/01/2016(H1N1))]	Influenza A virus (A/Indiana/01/2016(H1N1))	957	957	100%	0.0	99.79%	469	ANE28407.1
<input checked="" type="checkbox"/>	neuraminidase [Influenza A virus (A/California/123/2016(H1N1))]	Influenza A virus (A/California/123/2016(H1N1))	957	957	100%	0.0	99.79%	469	AOZ86020.1
<input checked="" type="checkbox"/>	neuraminidase [Influenza A virus (A/Idaho/43/2015(H1N1))]	Influenza A virus (A/Idaho/43/2015(H1N1))	957	957	100%	0.0	99.79%	469	ANE27318.1
<input checked="" type="checkbox"/>	neuraminidase [Influenza A virus (A/Rhode Island/03/2016(H1N1))]	Influenza A virus (A/Rhode Island/03/2016(H1N1))	957	957	100%	0.0	99.79%	469	ANE29500.1
<input checked="" type="checkbox"/>	neuraminidase [Influenza A virus (A/Czech Republic/7/2016(H1N1))]	Influenza A virus (A/Czech Republic/7/2016(H1N1))	957	957	100%	0.0	99.79%	469	AMP44463.1

Pairwise Comparisons & Default Alignment Format

Descriptions

Graphic Summary

Alignments

Alignment view

Pairwise

Restore defaults

6 sequences selected

Download

Graphics

Patient A17
Sequence ID: **Query_325421** Length: **469** Number of Matches: **1**

Range 1: 1 to 469

Next Match

Previous Match

Score	Expect	Method	Identities	Positives	Gaps
956 bits(2470)	0.0	Compositional matrix adjust.	468/469(99%)	469/469(100%)	0/469(0%)
Query 1	MNPNQKIITIGSICHTIGMANLILQIGNIISIWVSHSIQIGNQSQIETCNQSVITYENNT	60			
Sbjct 1	MNPNQKIITIGSICHTIGMANLILQIGNIISIWVSHSIQIGNQSQIETCNQSVITYENNT	60			
Query 61	WVNQTYVNISNTNFAAGQSVSVKLAGNSSLCPVSGWAIYSKDNSVRIGSKGDVFIREF	120			
Sbjct 61	WVNQTYVNISNTNFAAGQSVSVKLAGNSSLCPVSGWAIYSKDNSVRIGSKGDVFIREF	120			
Query 121	FISCSPLCERTFFLTQGALLNDKHSNGTIKDRSPYRTLMSCPIGEVPSPYNSRFESVAWS	180			
Sbjct 121	FISCSPLCERTFFLTQGALLNDKHSNGTIKDRSPYRTLMSCPIGEVPSPYNSRFESVAWS	180			
Query 181	ASACHDGINLWTIGISGPDGAVAVLYNGIITDTIKSWRNILRTQSEACVNGSCFT	240			
Sbjct 181	ASACHDGINLWTIGISGPDGAVAVLYNGIITDTIKSWRNILRTQSEACVNGSCFT	240			
Query 241	IMTDGPGDQASYKIFRIEKGKIIKSVEHKAPNYHYEECSYPSSEITCVCRDNIHGSN	300			
Sbjct 241	IMTDGPGDQASYKIFRIEKGKIIKSVEHKAPNYHYEECSYPSSEITCVCRDNIHGSN	300			
Query 301	RPWVSFNQLEYQMGYICSGVFGDNPRPNDKTGSCGPVSNANGVKGFsfkyGNGVNIIG	360			
Sbjct 301	RPWVSFNQLEYQMGYICSGVFGDNPRPNDKTGSCGPVSNANGVKGFsfkyGNGVNIIG	360			
Query 361	RTKSISRKGFEMIWDPNNGWTGTDNKFSIKQDIVGINEWSGYSGSFVQHPELTGLDCIRP	420			
Sbjct 361	RTKSISRKGFEMIWDPNNGWTGTDNKFSIKQDIVGINEWSGYSGSFVQHPELTGLDCIRP	420			
Query 421	CFWVELIRGRPEENTIWTSGSSISFCGVNSDVTGWSWPDGAELPFTIDK	469			
Sbjct 421	CFWVELIRGRPEENTIWTSGSSISFCGVNSDVTGWSWPDGAELPFTIDK	469			

Download

Graphics

Patient A13
Sequence ID: **Query_325420** Length: **469** Number of Matches: **1**

Range 1: 1 to 469

Next Match

Previous Match

Score	Expect	Method	Identities	Positives	Gaps
956 bits(2470)	0.0	Compositional matrix adjust.	468/469(99%)	469/469(100%)	0/469(0%)
Query 1	MNPNQKIITIGSICHTIGMANLILQIGNIISIWVSHSIQIGNQSQIETCNQSVITYENNT	60			
Sbjct 1	MNPNQKIITIGSICHTIGMANLILQIGNIISIWVSHSIQIGNQSQIETCNQSVITYENNT	60			
Query 61	WVNQTYVNISNTNFAAGQSVSVKLAGNSSLCPVSGWAIYSKDNSVRIGSKGDVFIREF	120			

Reformatting Results Creates Custom Views: Query-Anchored with Dots for Identities

The screenshot displays a web-based alignment viewer interface. At the top, there are three tabs: 'Descriptions', 'Graphic Summary', and 'Alignments', with 'Alignments' being the active tab. Below the tabs, there's a header bar with 'Alignment view' on the left, a 'Line length' dropdown set to '60', a 'Restore defaults' button, and a 'Download' button. A dropdown menu is open under 'Alignment view', showing several options: 'Query-anchored with dots for identities' (selected), 'Pairwise', 'Pairwise with dots for identities', 'Query-anchored with letters for identities', 'Flat query-anchored with dots for identities', and 'Flat query-anchored with letters for identities'. The main content area shows three query ranges. Each range has a 'Download' button and navigation links ('Next', 'Previous', 'First Range').

Query range 1:

Query	1GNQSQIETCNQSVITYENNT	60
Query_325421	1	60
Query_325420	1	60
Query_325419	1	60
Query_325418	1	60
Query_325416	1V.....I.....L.....N.....	60
Query_325417	1SIA..IIS.M.....A.....T.S.NNTGI...RI.....S.	60

Query range 2: 61 to 120

Query	61	WVNQTYVNISNTNFAAGQSVVSVKLAGNSSLCVPVSGWAIYSKDNSVRIGSKGDVFIREF	120
Query_325421	61	120
Query_325420	61	120
Query_325419	61	120
Query_325418	61	120
Query_325416	61	120
Query_325417	61	...H.....N...VV...EDKT...T.....SI.....T...I.....	120

Query range 3: 121 to 180

Query	121	FISCSPLECRTFFLTQGALLNDKHSNGTIKDRSPYRTLMSCPIGEVPSPYNSRFESVAWS	180
Query_325421	121	180
Query_325420	121	180
Query_325419	121	180
Query_325418	121	180
Query_325416	121	180
Query_325417	121H.....V.....A.....L..A.....K.....	180

Examples of Query Sequences

- **Genetic Testing:** Comparing a patient's *BRCA1* gene sequence to a sequence known *not* to contain cancer-causing mutations
- **DNA Barcoding:** Comparing your DNA sequence to all of the DNA sequences in the NCBI database to identify your unknown sample
- **Evolution:** Compare the human sickle cell allele [hemoglobin genotype AS] to the hemoglobin gene from other species to predict if other species get sickle cell disease

BLAST Scores: Query Coverage

Query Coverage: Where does the match start and where does it end?

Which alignment has the highest query coverage?

1. ATGGATACGT
TGAGATGATC

2. ATGCCGACAG
AGGGCAACAG

3. ATGGATAAGT
TGGGATGATC

4. ATGCCGATTC
AGGGCAACAG

BLAST Scores: Query Coverage

Query Coverage: Where does the match start and where does it end?

Which alignment has the highest query coverage?

1. ATGGATACGT

TGAGATGATC

Query Coverage = $3/10 = 30\%$

2. ATGCCGACAG

AGGGCAACAG

Query Coverage = $10/10 = 100\%$

3. ATGGATAAGT

TGGGATGATC

Query Coverage = $6/10 = 60\%$

4. ATGCCGATTC

AGGGCAACAG

Query Coverage = $7/10 = 70\%$

BLAST Scores: Percent Identity

Percent Identity: Where the query and subject sequences align, how much of the alignment is identical?

Which alignment has the highest percent identity?

(= Number of matches / Length of Query Coverage)

1. ATGGATACGT

TGAGATGATC

Query Coverage = 3/10 = 30%

2. ATGCGACAG

AGGCAACAG

Query Coverage = 10/10 = 100%

3. ATGGATAAGT

TGGGATGATC

Query Coverage = 6/10 = 60%

4. ATGCGATTC

AGGCAACAG

Query Coverage = 7/10 = 70%

BLAST Scores: Percent Identity

Percent Identity: Where the query and subject sequences align, how much of the alignment is identical?

Which alignment has the highest percent identity?

(= Number of matches / Length of Query Coverage)

1. ATGGATACGT

TGAGATGATC

Query Coverage = 3/10 = 30%

Percent Identity = 3/3 = 100%

2. ATGCCGACAG

AGGCAACAG

Query Coverage = 10/10 = 100%

Percent Identity = 7/10 = 70%

3. ATGGATAAGT

TGGGATGATC

Query Coverage = 6/10 = 60%

Percent Identity = 5/6 = 83%

4. ATGCCGATTC

AGGCAACAG

Query Coverage = 7/10 = 70%

Percent Identity = 4/7 = 57%

BLAST Scores: Use Together

Look at query coverage and percent identity together

Which alignment shows the best match?

1. ATGGATACGT

TGAGATGATC

Query Coverage = $3/10 = 30\%$

Percent Identity = $3/3 = 100\%$

3. ATGGATAGT

TGGGATATC

Query Coverage = $6/10 = 60\%$

Percent Identity = $5/6 = 83\%$

2. ATGCGACAG

AGGCAACAG

Query Coverage = $10/10 = 100\%$

Percent Identity = $7/10 = 70\%$

4. ATGCGATTC

AGGCAACAG

Query Coverage = $7/10 = 70\%$

Percent Identity = $4/7 = 57\%$

BLAST Scores: E-Values

E-Value: How likely is our result? Could it be due to chance?

An E-value is essentially the same as a p-value.

$p = 0.05$ means that 5% of the time (1 in 20) you will see this result even if it's not "true" (it's due to random chance)

Descriptions

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	pLemon-YFP What is the likelihood that this result is due to chance?	486	486	100%	4e-180	98%	47194
<input type="checkbox"/>	mTomato-RFP	70.9	70.9	100%	2e-19	28%	47195
<input type="checkbox"/>	mGrape1-CFP	62.8	62.8	95%	1e-16	25%	47196
<input type="checkbox"/>	pLime-GFP	487	487	99%	1e-180	99%	47197
<input type="checkbox"/>	pBlueberry-BFP	489	489	100%	0.0	98%	47198
<input type="checkbox"/>	mTangerine1.5	70.1	70.1	100%	3e-19	29%	47199
<input type="checkbox"/>	mCherry-RFP	71.6	71.6	100%	8e-20	29%	47200
<input type="checkbox"/>	mOrange-OFP	74.7	74.7	100%	8e-21	29%	47201