

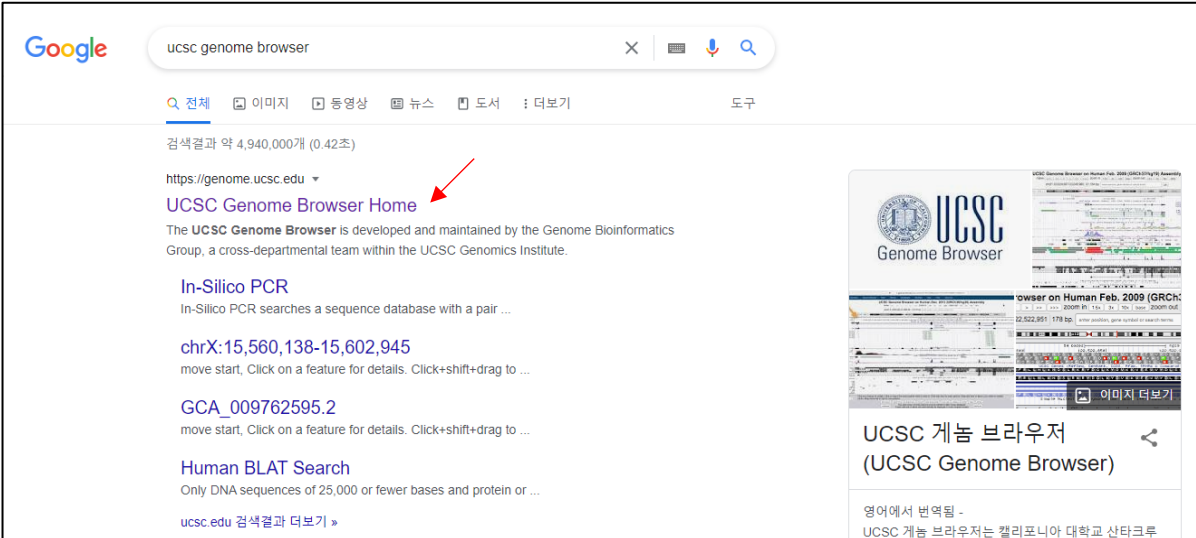
Molecular Biology Laboratory

Bioinformatics and Genomics Lab.

Week1. DNA / RNA Sequence Alignment (BLAST) & Genome Browser

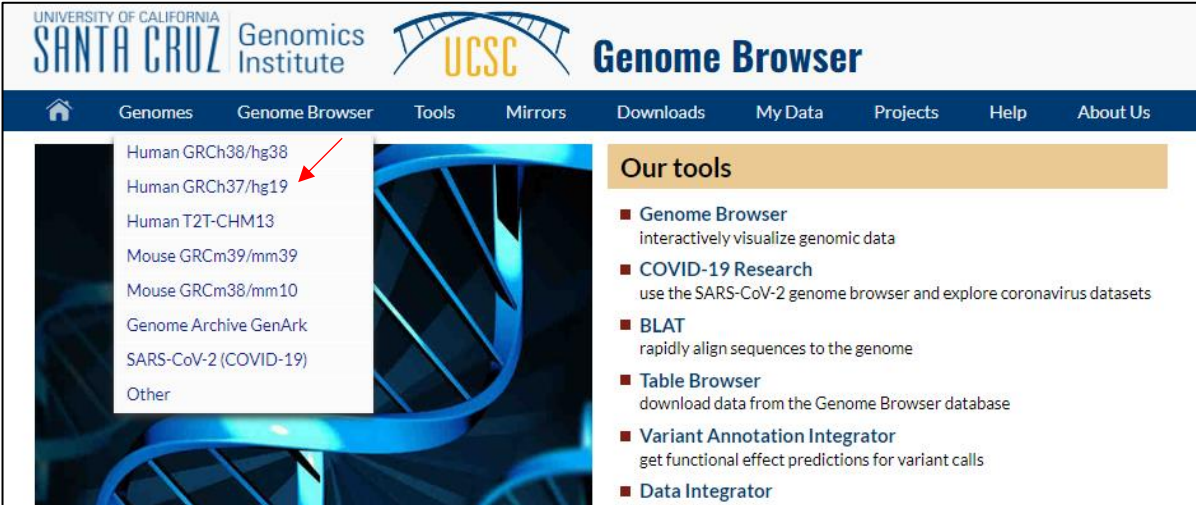
1. Practice how to use UCSC Genome Browser

- Access to the UCSC Genome Browser and find human insulin gene
 - Search "UCSC Genome Browser" in google and access to genome browser.



The screenshot shows a Google search results page for the query "ucsc genome browser". The search bar at the top contains the text "ucsc genome browser". Below the search bar, there are navigation options: "전체" (All), "이미지" (Images), "동영상" (Videos), "뉴스" (News), "도서" (Books), and "더보기" (More). The search results show approximately 4,940,000 results. The top result is "https://genome.ucsc.edu" with the title "UCSC Genome Browser Home". A red arrow points to this link. Below the main result, there are several related links: "In-Silico PCR", "chrX:15,560,138-15,602,945", "GCA_009762595.2", and "Human BLAT Search". On the right side of the search results, there is a preview of the UCSC Genome Browser interface, showing a genomic track and a search bar. Below the preview, there is a section titled "UCSC 게놈 브라우저 (UCSC Genome Browser)" with a share icon and a note that the page is in English.

- Hover the mouse pointer to "Genomes" and select genome version. We will use "Human GRCh38/hg38" in our practical exercise.



The screenshot shows the UCSC Genome Browser website. The header includes the logos for the University of California Santa Cruz Genomics Institute and UCSC, along with the text "Genome Browser". Below the header is a navigation menu with the following items: "Genomes", "Genome Browser", "Tools", "Mirrors", "Downloads", "My Data", "Projects", "Help", and "About Us". The "Genomes" menu is open, showing a list of genome versions: "Human GRCh38/hg38", "Human GRCh37/hg19", "Human T2T-CHM13", "Mouse GRCm39/mm39", "Mouse GRCm38/mm10", "Genome Archive GenArk", "SARS-CoV-2 (COVID-19)", and "Other". A red arrow points to "Human GRCh38/hg38". To the right of the menu is a section titled "Our tools" with a list of tools: "Genome Browser" (interactively visualize genomic data), "COVID-19 Research" (use the SARS-CoV-2 genome browser and explore coronavirus datasets), "BLAT" (rapidly align sequences to the genome), "Table Browser" (download data from the Genome Browser database), "Variant Annotation Integrator" (get functional effect predictions for variant calls), and "Data Integrator".

- Type "INS" search window and select "Homo sapiens insulin".

UCSC Genome Browser on Human (GRCh38/hg38)

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

multi-region chr11:10,472,002-10,486,431, 14,430 bp. go examples

INS (Homo sapiens insulin (INS), transcript variant 2, mRNA. (from RefSeq NM_001185097))

INS-IGF2 (Homo sapiens INS-IGF2 readthrough (INS-IGF2), transcript variant 2, mRNA. (from RefSeq NM_001042376))

INSC (Homo sapiens INSC spindle orientation adaptor protein (INSC), transcript variant 2, mRNA. (from RefSeq NM_001042536))

INSIG1 (Homo sapiens insulin induced gene 1 (INSIG1), transcript variant 1, mRNA. (from RefSeq NM_005542))

INSIG1-DT (INSIG1-DT (from geneSymbol))

INSIG2 (Homo sapiens insulin induced gene 2 (INSIG2), transcript variant 1, mRNA. (from RefSeq NM_016133))

INSL3 (Homo sapiens insulin like 3 (INSL3), transcript variant 2, mRNA. (from RefSeq NM_005543))

INSL4 (Homo sapiens insulin like 4 (INSL4), mRNA. (from RefSeq NM_002195))

INSL5 (Homo sapiens insulin like 5 (INSL5), mRNA. (from RefSeq NM_005478))

INSL6 (Homo sapiens insulin like 6 (INSL6), mRNA. (from RefSeq NM_007179))

INSM1 (Homo sapiens INSM transcriptional repressor 1 (INSM1), mRNA. (from RefSeq NM_002196))

INSM2 (Homo sapiens INSM transcriptional repressor 2 (INSM2), mRNA. (from RefSeq NM_032594))

INSR (Homo sapiens insulin receptor (INSR), transcript variant 2, mRNA. (from RefSeq NM_001079817))

INSRR (Homo sapiens insulin receptor related receptor (INSRR), mRNA. (from RefSeq NM_014215))

INSYN1 (Homo sapiens inhibitory synaptic factor 1 (INSYN1), transcript variant 5, mRNA. (from RefSeq NM_001384353))

INSYN1-AS1 (Homo sapiens INSYN1 antisense RNA 1 (INSYN1-AS1), transcript variant 2, long non-coding RNA. (from RefSeq NR_120353))

INSYN2A (Homo sapiens inhibitory synaptic factor 2A (INSYN2A), mRNA. (from RefSeq NM_001039762))

INSYN2B (Homo sapiens inhibitory synaptic factor family member 2B (INSYN2B), transcript variant 2, mRNA. (from RefSeq NM_001346304))

Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 2019)

- Compare Isoforms of the insulin gene and look through other information.

UCSC Genome Browser on Human (GRCh38/hg38)

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

multi-region chr11:2,159,418-2,161,582, 2,165 bp. go examples

Reference Assembly Fix Patch Sequence Alignments

Reference Assembly Alternate Haplotype Sequence Alignments

GENCODE V41

INS

INS

INS

INS

INS

OMIM Allelic Variant Phenotypes

Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 2019)

- Make custom gene annotation and visualize

- Click the marked insulin isoform, "ENST00000397262.5", and look through the information of the isoform.

Human Gene INS (ENST00000397262.5) from GENCODE V41

Description: Homo sapiens insulin (INS), transcript variant 4, mRNA. (from RefSeq NM_001291897)

RefSeq Summary (NM_001185098): This gene encodes insulin, a peptide hormone that plays a vital role in the regulation of carbohydrate and lipid metabolism.

Gencode Transcript: ENST00000397262.5

Gencode Gene: ENSG00000254647.7

Transcript (Including UTRs)

Position: hg38 chr11:2,159,779-2,161,204 **Size:** 1,426 **Total Exon Count:** 2 **Strand:** -

Coding Region

Position: hg38 chr11:2,159,852-2,160,971 **Size:** 1,120 **Coding Exon Count:** 2

Page Index	Sequence and Links	UniProtKB Comments	MalaCards	CTD	RNA-Seq Expression
Microarray Expression	RNA Structure	Protein Structure	Other Species	GO Annotations	mRNA Descriptions
Pathways	Other Names	GeneReviews	Methods		

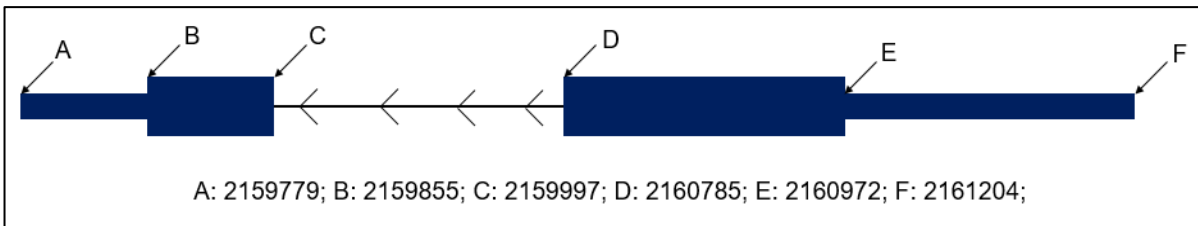
Data last updated at UCSC: 2022-05-14 09:57:26

- Open notepad and save it as "Custom_INS_Annotation.gtf".
- Fill the blank with reference to the below information.

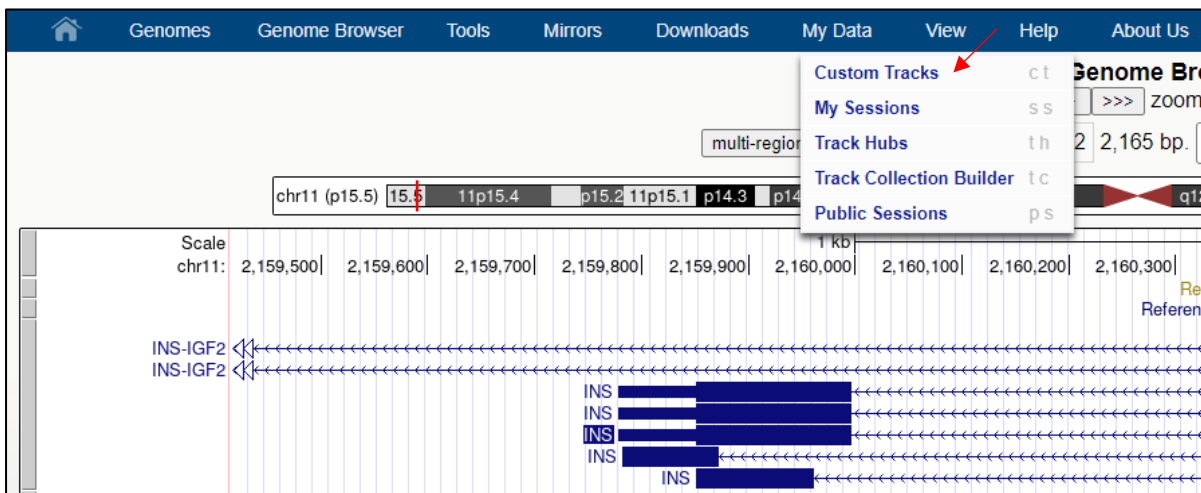
```

Week1_Custom_INS_Annotation_Answer.gtf - Windows 메모장
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
chr11 TEST transcript [ ] [ ] . [ ] [ ] . gene_id "INS"; transcript_id "INS";
chr11 TEST exon [ ] [ ] . [ ] [ ] . gene_id "INS"; transcript_id "INS";
chr11 TEST UTR [ ] [ ] . [ ] [ ] . gene_id "INS"; transcript_id "INS";
chr11 TEST CDS [ ] [ ] . [ ] [ ] . gene_id "INS"; transcript_id "INS";
chr11 TEST exon [ ] [ ] . [ ] [ ] . gene_id "INS"; transcript_id "INS";
chr11 TEST CDS [ ] [ ] . [ ] [ ] . gene_id "INS"; transcript_id "INS";
chr11 TEST UTR [ ] [ ] . [ ] [ ] . gene_id "INS"; transcript_id "INS";

```



- Cautions
 - ✓ Each row is arranged in ascending order based on the start position.
 - ✓ There is a difference of 1 between the end position of UTR and the end position of CDS.
 - ✓ Each column is separated by a tab, but gene id and transcript id are written in one column and they are separated by space.
- For uploading custom an annotation file to the genome browser, hover the mouse on "My Data" and click "Custom Tracks".



- Upload the custom annotation file and click "Submit".

Add Custom Tracks

clade genome assembly

Display your own data as custom annotation tracks in the browser. Data must be formatted in [bigBed](#), [GTF](#), [hic](#), [interact](#), [MAF](#), [narrowPeak](#), [Personal Genome SNP](#), [PSL](#), or [WIG](#) formats.

- You can paste just the URL to the file, without a "track" line, for bigBed, bigWig, bigGenePred, BAM
- To configure the display, set [track](#) and [browser](#) line attributes as described in the [User's Guide](#). Examples are [here](#). If you do not have web-accessible data storage available, please see the [Hosting](#)

Please note a much more efficient way to load data is to use [Track Hubs](#), which are loaded from the [I](#)

Paste URLs or data: Or upload: Week1_Cust...on_Answer.gtf

- After uploading, click "go" to visualize the custom annotation track.

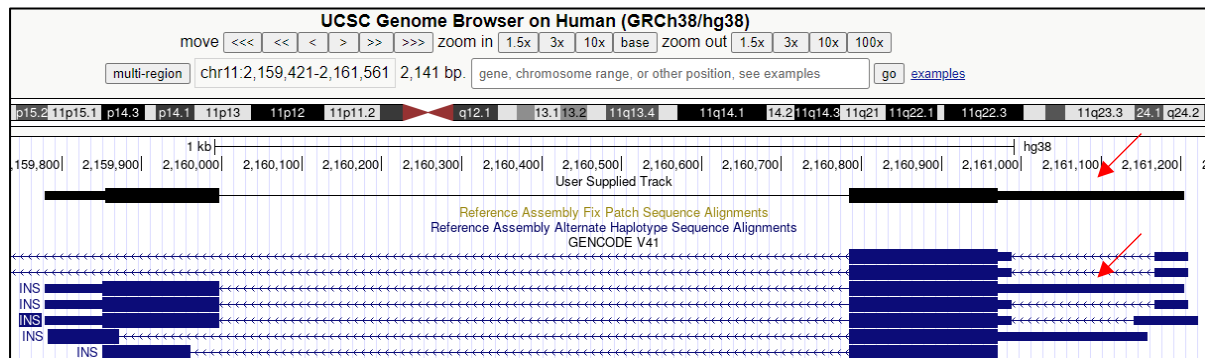
Manage Custom Tracks

genome: Human assembly: Dec. 2013 (GRCh38/hg38) [hg38]

Name	Description	Type	Doc	Items	Pos	delete
User Track	User Supplied Track	gtf		1	chr11:	<input type="checkbox"/>

view in

- Compare the custom annotation track and marked Insulin isoform.



2. Practice how to get sequence data and use BLASTN

- Find GFP gene sequence and align the sequence

➤ Search "NCBI" in google and access NCBI.

Google search results for 'ncbi'. The top result is the National Center for Biotechnology Information website. A red arrow points to the URL <https://www.ncbi.nlm.nih.gov>.

Search results for 'ncbi':

- 검색결과 약 608,000,000개 (0.52초)
- <https://www.ncbi.nlm.nih.gov>
- National Center for Biotechnology Information**
- Welcome to NCBI. The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.
- BLAST**
BLAST finds regions of similarity between biological sequences.
- PubMed**
PubMed® comprises more than 34 million citations for biomedical ...
- Nucleotide**
The Nucleotide database is a collection of sequences from ...
- Gene**
The .gov means it's official. Federal government websites often end ...

Right sidebar information:

미국 국립생물공학정보센터
회사

미국 국립생물공학정보센터는 미국 보건성 산하 국립 의학도서관의 운영 분야 중 하나이다. NCBI는 1988년 미국 메릴랜드주에 설립되었다. 위키백과

창립자: 클로드 페퍼
창립: 1988년 11월 4일
본사: 미국 메릴랜드 베서스다
모회사: 미국 국립 의학 도서관

➤ Search "GFP Aequorea victoria" and click "Nucleotide".

Search NCBI: GFP Aequorea victoria [Search]

Results found in 13 databases

Literature	Genes	Proteins
Bookshelf: 11	Gene: 0	Conserved Domains: 0
MeSH: 0	GEO DataSets: 56	Identical Protein Groups: 6
NLM Catalog: 2	GEO Profiles: 0	Protein: 715
PubMed: 439	HomoloGene: 0	Protein Family Models: 0
PubMed Central: 2,358	PopSet: 2	Structure: 307
Genomes	Clinical	PubChem
Assembly: 0	ClinicalTrials.gov: 0	BioAssays: 0
BioCollections: 0	ClinVar: 0	Compounds: 0
BioProject: 0	dbGaP: 0	Pathways: 0
BioSample: 8	dbSNP: 0	Substances: 1
Genome: 0	dbVar: 0	
Nucleotide: 146	GTR: 0	
SRA: 8	MedGen: 0	
Taxonomy: 0	OMIM: 0	

A red arrow points to the 'Nucleotide' database in the Genomes section, which shows 146 results.

- Select one from the list. We will use 4th one, "Aequorea victoria green-fluorescent protein (GFP) mRNA, complete cds" with 922 bp linear mRNA in a practical exercise.

[<< First](#)
[< Prev](#)
 Page of 8
 [Next >](#)
[Last >>](#)

- [Synthetic construct for Aequorea victoria partial gfp gene for GFP](#)
 1. 735 bp linear other-genetic
 Accession: LN515608.1 GI: 690969141
[PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- [Aequorea victoria clone EM1 GFP gene, partial cds](#)
 2. 135 bp linear DNA
 Accession: JN800726.1 GI: 353442080
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- [Aequorea victoria isolate sGFP-206 green fluorescent protein \(gfp\) gene, partial cds](#)
 3. 753 bp linear DNA
 Accession: JX472995.1 GI: 506954254
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)
- [Aequorea victoria green-fluorescent protein \(GFP\) mRNA, complete cds](#) 
 4. 922 bp linear mRNA
 Accession: L29345.1 GI: 606383
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

- Click "FASTA" to get the gene sequence of GFP.

Aequorea victoria green-fluorescent protein (GFP) mRNA, complete cds

GenBank: L29345.1
[FASTA](#) [Graphics](#)

Go to:

LOCUS AEVGFP 922 bp mRNA linear INV 30-DEC-1994
 DEFINITION Aequorea victoria green-fluorescent protein (GFP) mRNA, complete cds.
 ACCESSION L29345
 VERSION L29345.1
 KEYWORDS aequorin; bioluminescence; energy acceptor protein; green-fluorescent protein.
 SOURCE Aequorea victoria
 ORGANISM [Aequorea victoria](#)
 Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidolina; Leptothecata; Aequoreidae; Aequorea.
 REFERENCE 1 (bases 1 to 922)
 AUTHORS Inouye,S. and Tsuji,F.I.
 TITLE Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein
 JOURNAL FEBS Lett. 341 (2-3), 277-280 (1994)
 PUBMED [8137953](#)
 REFERENCE 2 (bases 1 to 922)
 AUTHORS Inouye,S. and Tsuji,F.I.
 TITLE Evidence for redox forms of the Aequorea green fluorescent protein
 JOURNAL FEBS Lett. 351 (2), 211-214 (1994)
 PUBMED [8082767](#)

Aequorea victoria green-fluorescent protein (GFP) mRNA, complete cds

GenBank: L29345.1

[GenBank](#) [Graphics](#)

>L29345.1 Aequorea victoria green-fluorescent protein (GFP) mRNA, complete cds

```
TACACACGAATAAAAAGATAACAAAAGATGAGTAAAGGAGAAGAAGCTTTTCACTGGAGTTGTCCCAATTCCTT
GTTGAATTAGATGGCGATGTTAATGGGCAAAAATTTCTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACAT
ACGAAAAAATTACCCTAAAATTTATTTGCACTACTGGGAAGCTACCTGTTCCATGGCCAACTTGTCCAC
TACTTTCTCTTATGGTGTTCATGCTTTTCAAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAG
AGTGCCATGCCCGAAGGTTATGTACAGGAAAGAAGCTATATTTTACAAAAGATGACGGGAAGCTACAAGACAC
GTGCTGAAGTCAAGTTTGAAGGTGATACCCCTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGA
AGATGGAACATTCTTGGACACAAAATGGAATACAACATAAATCACAATATGTATACATCATGGCAGAC
AAACCAAAGAATGGAATCAAAGTTAACTTCAAATAGACACAACATTAAAGATGGAAGCGTTCAATTAG
CAGACCATTAACAACAAAATCTCCAATTGGCGATGGCCCTGTCCTTTTACCAGCAACCATACCTGTG
CACACAATCTGCCCTTTCCAAAGATCCCAACGAAAAGAGAGATCAGATGATCCTTCTTGAGTTTGTACA
GCTGCTGGGATTACACATGGCATGGATGAACTATACAAAATAAATGTCCAGACTTCCAATTGACACTAAAG
TGTCCGAACAATTACTAAATCTCAGGGTCTCTGGTAAATTCAGGCTGAGACTTTATTTATATTTAT
AGATTCATTAATAATTTATGAATAATTTATGATGTTAATAAGGGCTATTTTCTATTAAATAGGCT
ACTGGAGTGTAT
```

➤ Search "BLAST" in google and access BLAST.

The screenshot shows a Google search for "blast". The search bar contains "blast" and the search button is visible. Below the search bar, there are navigation options like "전체", "이미지", "뉴스", "동영상", "도서", and "더보기". The search results show approximately 728,000,000 results (0.44초). The top result is "BLAST: Basic Local Alignment Search Tool" with a red arrow pointing to the title. Below the title, there is a description: "The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to ...". There are several sub-links: "BLAST", "Nucleotide BLAST", "Standard Nucleotide BLAST", and "Align two or more sequences". On the right side, there is a preview of the BLAST website interface.

➤ Click "Nucleotide BLAST" (BLASTN).

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

BLAST+ 2.13.0 is here!

Starting with this release, we are including the blastn_vdb and tblastn_vdb executables in the BLAST+ distribution.

Thu, 17 Mar 2022 12:00:00 EST

[More BLAST news...](#)

Web BLAST

Nucleotide BLAST
nucleotide ▶ nucleotide

blastx
translated nucleotide ▶ protein

tblastn
protein ▶ translated nucleotide

Protein BLAST
protein ▶ protein

- Copy and paste the GFP sequence to the query sequence box, and set the job title. When pasting the sequence, we must paste the header part too. There are many options, but we will use the default setting in the practical experiment. After then, select the program "BLASTN" and click "BLAST" for alignment.

BLAST® >> blastn suite

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Database Standard databases (nr etc.): rRNA/ITS databases Genomic + transcript databases Betacoronavirus

Nucleotide collection (nr/nt) [?](#)

Organism exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude Models (XM/XP) Uncultured/environmental sample sequences

Limit to Sequences from type material

Entrez Query [YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search [?](#)

Program Selection

Optimize for Highly similar sequences (megablast) More dissimilar sequences (discontiguous megablast) Somewhat similar sequences (blastn) [?](#)

Choose a BLAST algorithm [?](#)

BLAST Search database Nucleotide collection (nr/nt) using Blastn (Optimize for somewhat similar sequences)

Show results in a new window

- We can see the alignment result. There are many similar sequences in the database, but the first one is the best-matched sequence. Select it and see "Graphic Summary" and "Alignments".

Job Title **GFP BLAST**

RID [HD72483P013](#) Search expires on 09-07 09:03 am [Download All](#) [v](#)

Program **BLASTN** [Citation](#) [v](#)

Database **nt** [See details](#) [v](#)

Query ID **lcl|Query_20431**

Description **L29345.1 Aequorea victoria green-fluorescent protein (GF...**

Molecule type **dna**

Query Length **922**

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

Filter Results

Organism *only top 20 will appear* exclude

[+ Add organism](#)

Percent Identity to E value to Query Coverage to

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

Descriptions Graphic Summary Alignments Taxonomy

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

select all *1 sequences selected*

[GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Aequorea victoria green-fluorescent protein (GFP) mRNA, complete cds	Aequorea victoria	1663	1663	100%	0.0	100.00%	922	L29345.1
<input type="checkbox"/> Aequorea victoria green-fluorescent protein mRNA, complete cds	Aequorea victoria	1558	1558	100%	0.0	97.51%	966	M62653.1
<input type="checkbox"/> Synthetic construct from Aequorea victoria GFPS (gfps) mRNA, complete cds	synthetic construct	1498	1498	96%	0.0	97.21%	998	AF183395.1
<input type="checkbox"/> Cloning vector p5xHis-GFP, complete sequence	unidentified cloni...	1494	1494	96%	0.0	97.09%	5271	U89936.1

Job Title GFP BLAST
RID HD72483P013 Search expires on 09-07 09:03 am [Download All](#) ▼
Program BLASTN [Citation](#) ▼
Database nt [See details](#) ▼
Query ID lc|Query_20431
Description L29345.1 Aequorea victoria green-fluorescent protein (GF| ...
Molecule type dna
Query Length 922
Other reports [Distance tree of results](#) [MSA viewer](#) ?

Filter Results
Organism only top 20 will appear exclude

[+ Add organism](#)
Percent Identity to
E value to
Query Coverage to

Descriptions **Graphic Summary** Alignments Taxonomy

Alignment Scores < 40 40 - 50 50 - 80 80 - 200 >= 200 ?

1 sequences selected ?

Distribution of the top 1 Blast Hits on 1 subject sequences

Descriptions Graphic Summary **Alignments** Taxonomy

Alignment view CDS feature ? [Restore defaults](#) [Download](#) ▼

1 sequences selected ?

[Download](#) ▼ [GenBank](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

Aequorea victoria green-fluorescent protein (GFP) mRNA, complete cds
Sequence ID: [L29345.1](#) Length: 922 Number of Matches: 1

Range 1: 1 to 922 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1663 bits(1844)	0.0	922/922(100%)	0/922(0%)	Plus/Plus

```

Query 1  TACACACGAATAAAAAGATAACAAGATGAGTAAAAGAGAAAGACTTTTCACTGGAGTTGT 60
          |||
Sbjct 1  TACACACGAATAAAAAGATAACAAGATGAGTAAAAGAGAAAGACTTTTCACTGGAGTTGT 60

Query 61  CCCAATTCTTGTGAATTAGATGGCGATGTTAATGGGCAAAAATTCCTGTCAGTGGAGA 120
          |||
Sbjct 61  CCCAATTCTTGTGAATTAGATGGCGATGTTAATGGGCAAAAATTCCTGTCAGTGGAGA 120

Query 121  GGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGGAA 180
          |||
Sbjct 121  GGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGGAA 180

Query 181  GCTACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTCAATGCTTTTC 240
          |||
Sbjct 181  GCTACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTCAATGCTTTTC 240
  
```

- Find insulin gene sequences of human and pig, and compare the two sequences
 - Search "human insulin" in NCBI.

Search NCBI

Results found in 29 databases

GENE Was this helpful?

INS - insulin

Homo sapiens (human)
Also known as: IDDM, IDDM1, IDDM2, ILPR, IRDN, MODY10, PNDM4
Gene ID: 3630
[RefSeq transcripts \(4\)](#) [RefSeq proteins \(4\)](#) [RefSeqGene \(2\)](#) [PubMed \(970\)](#)

RefSeq Sequences

RefSeq transcripts
INS - 3 of 4 transcripts

Transcript	Isoform	Len (nt)
NM_000207.3		465
NM_001185097.2		491
NM_001291897.2		525

[View full table](#) NCBI Datasets

- Click "INS" of pig and get gene sequence in the same way.

Species	Gene	Architecture	aa
<input type="checkbox"/> <i>Homo sapiens</i> human	INS insulin		110
<input type="checkbox"/> <i>Mus musculus</i> house mouse	Ins2 insulin II		110
<input type="checkbox"/> <i>Rattus norvegicus</i> Norway rat	Ins2 insulin 2		119
<input type="checkbox"/> <i>Bos taurus</i> cattle	INS insulin		221
<input type="checkbox"/> <i>Sus scrofa</i> pig	INS insulin		175
<input type="checkbox"/> <i>Gallus gallus</i> chicken	INS insulin		107

Sus scrofa isolate TJ Tabasco breed Duroc chromosome 2, Sscrofa11.1, whole genome shotgun sequence

NCBI Reference Sequence: NC_010444.4

[GenBank](#) [Graphics](#)

>NC_010444.4:c1498052-149842 Sus scrofa isolate TJ Tabasco breed Duroc chromosome 2,

Sscrofa11.1, whole genome shotgun sequence

```
TGGCCCTCAGGCCGCTCGTTAAGACTCTAATGACCTCAAGGCCCCAGAGGGCGCTGATGACCCACGGAGA
TGATCCCGCAGGCCCTGGCAGCAGGAAATGATCCAGAAAGTGCCACCTCAGCCCCAGCCATCTGCCACC
CACCTGGAGGCCCTCAGGGCCCGGGCCCGGGGGGCAAGGCGCTATAAAGCCGGCCGGGCCAGCCGCCCC
CAGCCCTCTGGGACCAAGCTGCGTCCCAGGCCCGGGCAAGCAGGTCTGTCCCCTGGGCTCCCGTCAGC
TGGGTCTGGGCTGTCTGCTGGGGCCAGGGCATCTCGGCAAGGAGCAGTGGGCTCCTCTCTGGAGCCCT
TGGGGGTGAGGCTGGTGGGGCTGCAGGTGCCCTGGCTGGCCTCAAAGCCCGCCCGTCCCCAGGTCCT
CACCCCCGCCATGGCCCTGTGGACGGCCTCCTGCCCTGCTGGCCCTGCTGGCCCTCTGGGCGCCCGC
CCCGGCCAGGCCCTTCGTGAACCAAGCACCCTGTGCGGCTCCACCTGGTGGAGGCGCTGTACCTGGTGTGC
GGGGAGCGCGGCTCTTCTCTACAGCCCAAGGCCCGTGGGAGGGCGGAGAACCTCAGGCTGAGCCGAGGG
GGCGTCCCGGAGCGGTGGGGGAGTTTTTAAAGGAGAAATGGTAAAAGTGACCAACTCCCTGGGAGCT
GAGCCAGAGACACCCCTCCACGCCCGGTCCTCGCTCGAGAAAGCCCCCTTCCCTCCCTCTCCCGAG
GCGGCTCAGGGAGGAAATCTACGGAGTCAAGGCCCGGGTCCCGCTGGTCTCCGAGTACATGGCCGTGG
TGTCCCGTCTGCCGGCCACATGCCCGTGGGAGATGCCCATCCCGCTGGGAGGGGGCCCGTGGCGGGC
AGGACGCGGAGGCCAGGACCGGGGCTGCTGCGGCTTCCACTCCAGGGTGGGCGGGTGGGGGTGGC
TGTCTCTGTGTGACCGGCTCTCCCGCAGCAGGTGCCGTGGAAGTGGGCGGAGGCTGGGCGGCTGCAAG
GCCCTGGCGCTGAGGGGCCCGCAGAGCGTGGCATCGTGGAGCAGTGTGACCCAGCATCTGTCCC
TCTACCACTGGAGACTAGCAACTAGGCCCGCCCTGAGGGCGCTGCTGCTCCCGCACCCCAAAAC
CCAATAAAGTCTGAATGAGC
```

- Go to BLASTN and check the "Align two or more sequences". Paste the human sequence in the query sequence and paste the pig's sequence in the subject sequence. After then, select "blastn" for the program, and click "BLAST".

Align Sequences Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide subjects using a nucleotide query.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

>NC_000011.10:c2161209-2159779 Homo sapiens chromosome 11,
GRCh38 p14 Primary Assembly
AGCCCTCCAGGACAGGCTGCATCAGAAGAGGCCATCAAGCAGTCTGTT
CCAAGGCCCTTTCGTCAGGT

Query subrange [?](#)

From

To

Or, upload file [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

>NC_010444.4:c1498052-1496842 Sus scrofa isolate TJ Tabasco breed
Duroc chromosome 2, Sscrofa11.1, whole genome shotgun sequence
TGGCCCTCAGGCCGCTCGTTAAGACTCTAATGACCTCAAGGCCCCAGAG
GGCGCTGATGACCCACGGAGA

Subject subrange [?](#)

From

To

Or, upload file [?](#)

Program Selection

Optimize for

Highly similar sequences (megablast)

More dissimilar sequences (discontiguous megablast)

Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

BLAST Search nucleotide sequence using Blastn (Optimize for somewhat similar sequences)

Show results in a new window

- We can see how similar human's and pig's insulin gene sequence from the result.

Job Title **Human INS vs. Pig INS**

RID [HD8JZJPW114](#) Search expires on 09-07 09:29 am [Download All](#) [?](#)

Program Blast 2 sequences [Citation](#) [?](#)

Query ID lcl|Query_43979 (dna)

Query Descr NC_000011.10:c2161209-2159779 Homo sapiens chrom ...

Query Length 1431

Subject ID lcl|Query_43981 (dna)

Subject Descr NC_010444.4:c1498052-1496842 Sus scrofa isolate TJ Te ...

Subject Length 1211

Other reports [MSA viewer](#) [?](#)

Filter Results

Percent Identity to

E value to

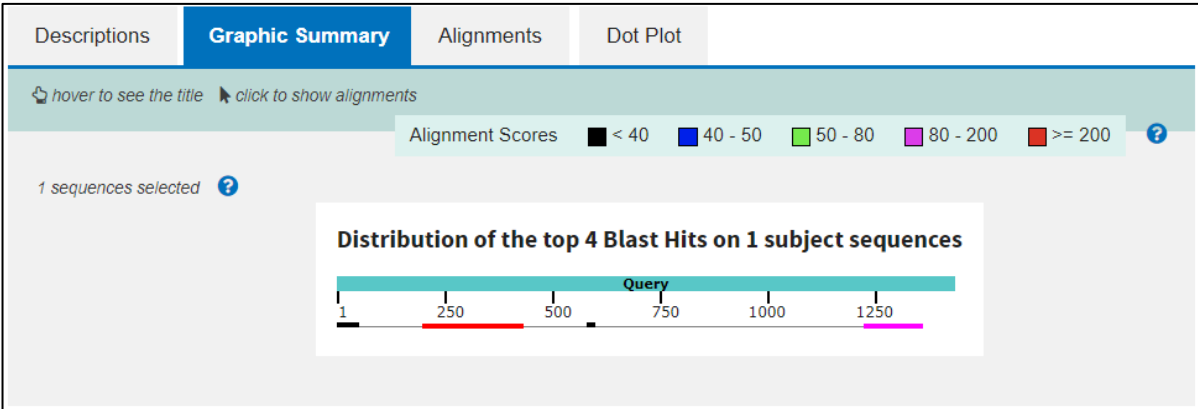
Query Coverage to

Descriptions Graphic Summary Alignments Dot Plot

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

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

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> NC_010444.4:c1498052-1496842 Sus scrofa isolate TJ Tabasco breed Duroc chromosome 2, Sscrofa11.1, whole genom ...		250	476	31%	8e-70	83.83%	1211	Query_43981



Download Graphics Sort by: E value Next Previous Descriptions

NC_010444.4:c1498052-1496842 Sus scrofa isolate TJ Tabasco breed Duroc chromosome 2, Sscrofa11.1, whole genome shotgun sequence

Sequence ID: Query_43981 Length: 1211 Number of Matches: 4

Range 1: 392 to 625 Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
250 bits(276)	8e-70	197/235(84%)	1/235(0%)	Plus/Plus
Query 198	GCCTCAGCCCTGCCTGTCTCCAGATCACTGTCCTTCTGCCATGGCCCTGTGGATGCGCC	25		
Sbjct 392	GCCTCAACGCCGCCCGTCCCCAAGGTC-CTACCCCCCGCCATGGCCCTGTGGACGCGCC	45		
Query 258	TCCTGCCCTGCTGGCGCTGCTGGCCCTCTGGGGACCTGACCCAGCCGACGCTTTGTGA	31		
Sbjct 451	TCCTGCCCTGCTGGCCCTGCTGGCCCTCTGGGCGCCCGCCCGGCCAAGGCTTCGTGA	51		
Query 318	ACCAACACCTGTGCGGCTCACACCTGGTGAAGCTCTCTACCTAGTGTGCGGGGAACGAG	37		
Sbjct 511	ACCAGCACCTGTGCGGCTCCCACCTGGTGAAGGCGCTGTACCTGGTGTGCGGGGAGCGCG	57		
Query 378	GCTTCTTACACACCCAAGACCCGCGGGAGGCAGAGGACCTGCAAGGTTAGCC	432		
Sbjct 571	GCTTCTTACACGCCCAAGGCCCGTGGGAGGCGGAGAACCTCAGGTTAGCC	625		