

# **CORSO DI METODI MOLECOLARI E BIOINFORMATICA**

**LM Biologia Evoluzionistica, Università di Padova**

**Dr. Silvia Orsi, Dr. Enrico Gaffo,**

**Prof. Stefania Bortoluzzi**

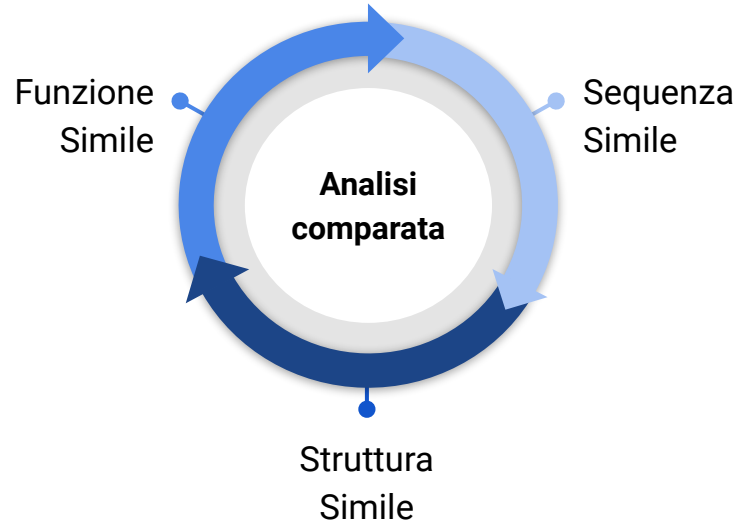
## **Esercitazione 2**

**“Comparazione di biosequenze ed allineamenti multipli”**

*Padova, 02 Novembre 2023*

## Perchè allineare?

- Per fornire una misura di quanto due sequenze nucleotidiche o amminoacidiche abbiano in comune
- Ci permette di fare inferenze biologiche in termini di:
  - relazioni strutturali
  - relazioni funzionali
  - relazioni evolutive



# Allineamento di acidi nucleici: perché?

Ho scoperto un nuovo gene e voglio vedere se:

- il gene è già presente nelle banche dati di acidi nucleici (EMBL/Genbank), anche parzialmente...
- il gene è simile ad altri geni presenti nelle banche dati
- il gene è simile ad una sequenza particolare

# Allineamento di proteine: perché?

- non sono sicuro che la proteina che ho scoperto sia una proteina “nuova”, è già conosciuta?
- la proteina che ho scoperto è una “nuova”, ma assomiglia a qualche proteina nota?
- nella proteina che sto studiando (che non è simile a nessun'altra proteina) esistono porzioni simili a proteine note, o domini funzionali noti anche in altre specie?

# Allineamento di biosequenze: programmi

Pairwise sequence alignment (PSA):

- **BLAST** (NCBI)
- **BLAT** (UCSC)

Multiple sequence alignment (MSA):

- **Muscle**

# BLAST (Basic Local Alignment Search Tool, NCBI)

- Allinea sequenze di DNA, mRNA o proteine
- Permette di comparare sequenze biologiche contro database di sequenze costantemente aggiornati
- Allineamento non preciso come gli algoritmi Needleman-Wunsch e Smith-Waterman ma molto più veloce
- Comprende una serie di programmi che variano in base al tipo di sequenze da comparare (blastn, blastp, tblastn, tblastx).

## Per cosa può essere utile BLAST?

- **Identificare una specie** o una specie **omologa** (per esempio si conosce una sequenza di DNA di una specie sconosciuta)
- Localizzare i **domini funzionali** di una proteina
- Mappare una sequenza di DNA per trovare la **posizione cromosomica**
- **Localizzare geni comuni** tra due specie correlate (può essere usato per “mappare” annotazioni da un organismo all'altro)

## BLAT (Blast-like Alignment Tool, UCSC)

*Algoritmo di allineamento di sequenza a coppie sviluppato da **Jim Kent** presso l'Università della California Santa Cruz (UCSC) all'inizio degli anni 2000 per assistere nell'assemblaggio e nell'annotazione del genoma umano. È stato progettato principalmente per ridurre il tempo necessario per allineare milioni di letture genomiche del topo e tag di sequenza espresse rispetto alla sequenza del genoma umano.*

- Allinea sequenze di DNA, mRNA o proteine con sequenze genomiche
- Algoritmo simile a BLAST
- Cerca similarità tra una sequenza in esame e la sequenza di un genoma
- Efficace per sequenze corte (almeno 40 caratteri e massimo 25.000 nucleotidi o 10.000 amminoacidi)



## Per cosa può essere utile BLAT?

- **Allineamento di sequenze di mRNA** sul genoma per trovare le coordinate del gene che lo codifica
- Mostrare le **sequenze codificanti** di un gene
- Determinare **distribuzione** di esoni ed introni in un gene
- **Analisi tra specie**: allineamento di una proteina o mRNA di una specie su un database di sequenze di un'altra specie
- **Allineare due sequenze proteiche** (meno preciso di BLAST)

## Esercitazione 2: allineamento di sequenze nucleotidiche

1. Trovare **se** la sequenza nucleotidica query “seq1” di *Homo Sapiens* fa parte di un gene conosciuto;
2. Visualizzare le sequenze “seq1” e “seq2” come si allineano sul genoma e qual è il loro ruolo funzionale (codificante/non codificante/regolatoria)

## Esercitazione 2: allineamento di sequenze nucleotidiche

1. Trovare **se** la sequenza nucleotidica query “*seq1*” di *Homo Sapiens* fa parte di un gene conosciuto;
2. Visualizzare le sequenze “*seq1*” e “*seq2*” come si allineano sul genoma e qual è il loro ruolo funzionale (codificante/non codificante/regolatoria)

1

Nucleotide Blast (NCBI)

2

BLAT (UCSC)

3

Protein Blast and Muscle

1 - Scaricare il file da **Guida** in “**Il esercitazione BIOINFORMATICA**”

[http://compgen.bio.unipd.it/~stefania/Didattica/AA2022-2023/MMOL\\_BIOINFO\\_BE/MMOL\\_BIOINFO\\_BE.html](http://compgen.bio.unipd.it/~stefania/Didattica/AA2022-2023/MMOL_BIOINFO_BE/MMOL_BIOINFO_BE.html)

2 - decomprimere il file esercitazione2.zip ed aprire con l'editor di testo il file 3seqs.txt

3 - Scrivere Blast su Google e cliccare su Nucleotide BLAST:

## 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](#)

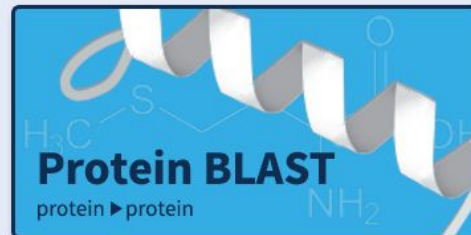
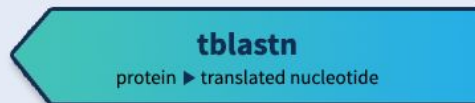
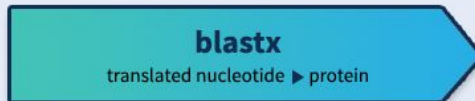
**BLAST now supports the new RefSeq Select databases.**

The RefSeq Select data-set consists of a representative or “Select” transcript for every protein-coding gene.

Tue, 13 Oct 2020 12:00:00 EST

[More BLAST news...](#)

## Web BLAST



## BLAST Genomes

Enter organism common name, scientific name, or tax id

Search

Human

Mouse

Rat

Microbes

2 - Inserire la seq1 come sequenza query e chiamare l'analisi "Seq1":

BLAST<sup>®</sup> » blastn suite [Home](#)

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

**Enter Query Sequence**

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

>Seq1  
actggcgcctaaaaattttgacgtttcctttacaattctcaaaagtctagagccaccotccagggagcaggtagctg  
ctgggctccggggacactttgcgttcgggctgggagcgtgctttccacgacggtgacacgcttccctgattggc  
agccagactgccttccgggtcactgccatggaggagccgcagtcagatcctagcgtcagccccctgagtc

Query subrange [?](#)

From



To

Or, upload file  Nessun file selezionato. [?](#)

Job Title

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

☐ Align two or more sequences [?](#)



- 3 - Controllare che il database selezionato sia “Nucleotide collection (nr/nt)” e che la tipologia sia “Standard databases” (cliccare sul punto di domanda per vedere le caratteristiche del database scelto); scegliere “Homo sapiens” come organismo:

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. more...

Reset page

Bookmark

### Enter Query Sequence

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

Query subrange [?](#)

From

To

Or, upload file  Nessun file selezionato. [?](#)

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

**New** ☐ Experimental databases [Try experimental taxonomic nt databases](#) [Download](#)

For more info see [What are taxonomic nt databases?](#)

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

Homo sapien (taxid:9606) ☐ experimental organism [?](#)

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

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

☐ Sequences from type material

Exclude [Optional](#)

Limit to [Optional](#)

Entrez Query [Optional](#)

[YouTube](#) Create custom database

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

3 - Controllare che il database selezionato sia “Nucleotide collection (nr/nt)” e che la tipologia sia “Standard databases” (clickare sul punto di domanda per vedere le caratteristiche del database scelto); scegliere “Homo sapiens” come organismo:

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. more...

Reset page

Bookmark

### Enter Query Sequence

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

Query subrange [?](#)

From

To

Or, upload file  Nessun file selezionato. [?](#)

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

**New** ☐ Experimental databases [Try experimental taxonomic nt databases](#)

For more info see [What are taxonomic nt databases?](#)

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

Homo sapien (taxid:9606) ☐ exclude organism

**Exclude** [Optional](#)

**Limit to** [Optional](#)

**Entrez Query** [Optional](#)

**Title:** Nucleotide collection (nt)

**Description:** The nucleotide collection consists of GenBank+EMBL+DDBJ+PDB+RefSeq sequences, but excludes EST, STS, GSS, WGS, TSA, patent sequences as well as phase 0, 1, and 2 HTGS sequences and sequences longer than 100Mb. The database is non-redundant. Identical sequences have been merged into one entry, while preserving the accession, GI, title and taxonomy information for each entry.

**Molecule Type:** mixed DNA

**Update date:** 2023/10/30

**Number of sequences:** 99985738



**Megablast è ottimizzato per allineare sequenze quasi uguali**, le cui differenze possano derivare da errori di sequenziamento o polimorfismi.

**discontiguous megablast** viene utilizzato soprattutto per progetti di **genomica comparata**: si confrontano sequenze di specie affini, le cui sequenze non divergono quanto quelle di specie evolutivamente lontane, ma più delle sequenze identiche (*a parte i polimorfismi*) della stessa specie.

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 **Megablast (Optimize for highly similar sequences)**

☐ Show results in a new window

La **differenza** tra megablast e blastn è nella scelta della **"word size"**, ovvero la **lunghezza minima della stringa di residui contigui considerata ai fini della valutazione di identità**, che in Megablast è ottimale con valori maggiori o uguali a 16 (multipli di 4). Ciò rende più selettiva l'accettazione di identità e fino a 10 volte più veloce **Megablast che, quindi, si presta meglio all'uso con sequenze molto lunghe ed estremamente simili;**

4 - Ottimizzare per megablast, lasciare i parametri di default dell'algoritmo ed avviare la ricerca:

**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 **Megablast (Optimize for highly similar sequences)**

☐ Show results in a new window



**+ Algorithm parameters**



BLAST result:

ritorna alla pagina della ricerca per poterla rifare cambiando parametri

salva i parametri utilizzati per un'eventuale ricerca futura

mostra una tabella con i parametri utilizzati per la ricerca

BLAST<sup>®</sup> » blastn suite » results for RID-TZSWRRGG013

[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](#)

**i** Your search is limited to records that include: Homo sapiens (taxid:9606)

Job Title	Seq1
RID	<a href="#">TZSWRRGG013</a> Search expires on 11-03 17:46 pm <a href="#">Download All</a> ▼
Program	BLASTN <a href="#">?</a> <a href="#">Citation</a> ▼
Database	nt <a href="#">See details</a> ▼
Query ID	lcl Query_32607
Description	Seq1
Molecule type	dna
Query Length	1115
Other reports	<a href="#">Distance tree of results</a> <a href="#">MSA viewer</a> <a href="#">?</a>

informazioni sulla ricerca

### 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](#)

filtri

scarica le hits selezionate  
in vari formati

permette di scegliere che  
colonne visualizzare

permette di scegliere  
quante righe visualizzare

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download Manage columns Show 100

☒ select all 100 sequences selected

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

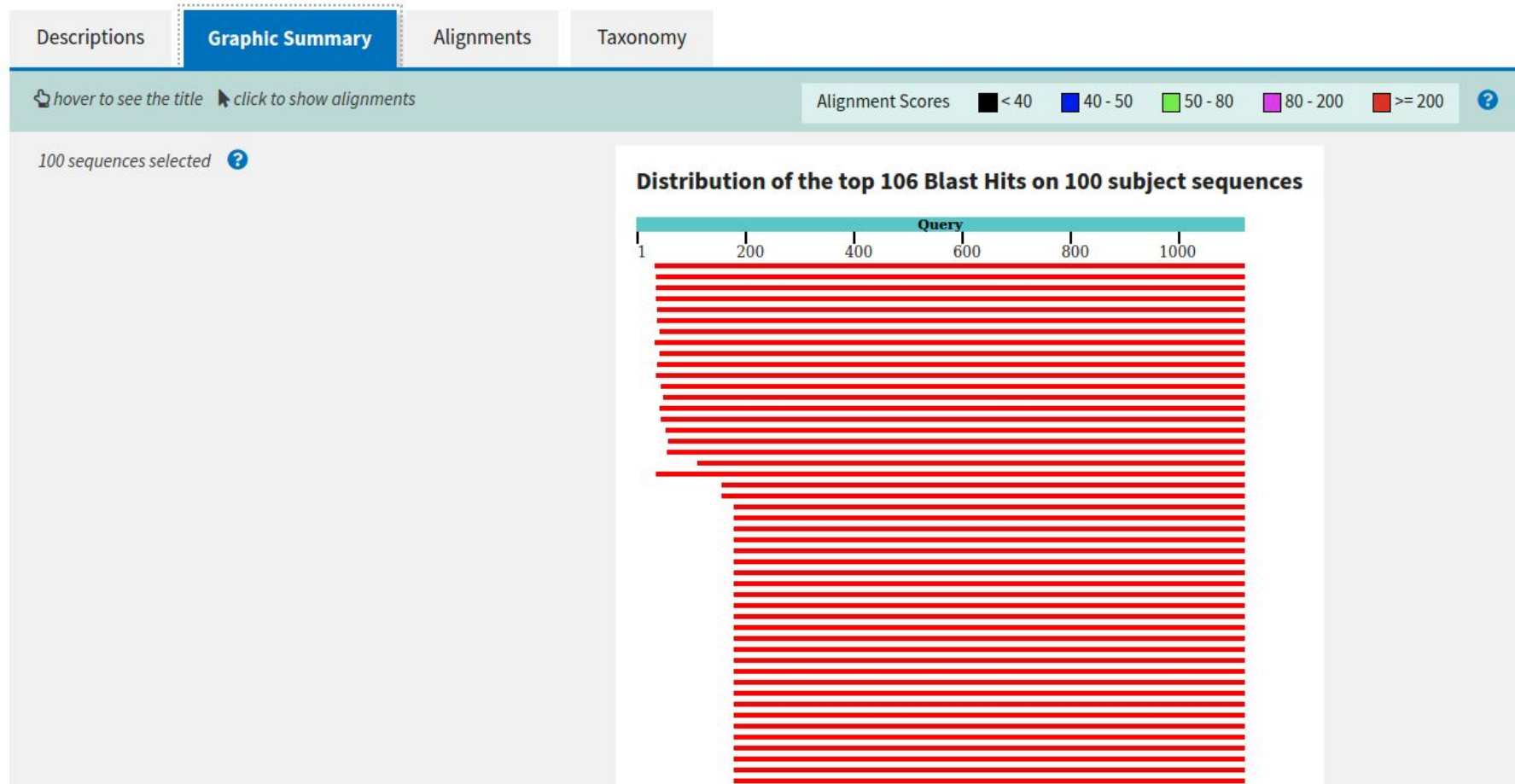
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)</a>	1858	1858	96%	0.0	97.70%	<a href="#">HG975427.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA</a>	1855	1855	96%	0.0	97.69%	<a href="#">NM_000546.6</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens tumor protein p53 (TP53), transcript variant 3, mRNA</a>	1855	1855	96%	0.0	97.69%	<a href="#">NM_001276696.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens tumor protein p53 (TP53), transcript variant 4, mRNA</a>	1855	1855	96%	0.0	97.69%	<a href="#">NM_001276695.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens mRNA for Cellular tumor antigen p53 variant, clone: FCC127A11</a>	1844	1844	96%	0.0	97.59%	<a href="#">AK225838.1</a>
<input checked="" type="checkbox"/>	<a href="#">Human p53 cellular tumor antigen mRNA, complete cds</a>	1844	1844	96%	0.0	97.59%	<a href="#">K03199.1</a>

**Max score:** il più alto score calcolato dell'allineamento, contando premi e penalità.

**Total score:** la somma dei punteggi dell'allineamento di tutti i segmenti della stessa sequenza.

**E-Value (Expected value):** il numero di allineamenti attesi dal caso con lo score calcolato. Per allineamenti significativi l'E-value deve essere molto vicino allo 0.

## 5 - Cliccare su “Graphic Summary” per vedere la qualità degli allineamenti:



Qual è l'allineamento migliore?

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Manage columns

Show100

☒

select all

100 sequences selected

GenBank

Graphics

Distance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)</a>	1858	1858	96%	0.0	97.70%	<a href="#">HG975427.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA</a>	1855	1855	96%	0.0	97.69%	<a href="#">NM_000546.6</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens tumor protein p53 (TP53), transcript variant 3, mRNA</a>	1855	1855	96%	0.0	97.69%	<a href="#">NM_001276696.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens tumor protein p53 (TP53), transcript variant 4, mRNA</a>	1855	1855	96%	0.0	97.69%	<a href="#">NM_001276695.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens mRNA for Cellular tumor antigen p53 variant, clone: FCC127A11</a>	1844	1844	96%	0.0	97.59%	<a href="#">AK225838.1</a>
<input checked="" type="checkbox"/>	<a href="#">Human p53 cellular tumor antigen mRNA, complete cds</a>	1844	1844	96%	0.0	97.59%	<a href="#">K03199.1</a>

Qual è l'allineamento migliore?

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Manage columns

Show100

?

☒ select all

100 sequences selected

GenBank

Graphics

Distance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)</a>	1858	1858	96%	0.0	97.70%	<a href="#">HG975427.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA</a>	1855	1855	96%	0.0	97.69%	<a href="#">NM_000546.6</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens tumor protein p53 (TP53), transcript variant 3, mRNA</a>	1855	1855	96%	0.0	97.69%	<a href="#">NM_001276696.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens tumor protein p53 (TP53), transcript variant 4, mRNA</a>	1855	1855	96%	0.0	97.69%	<a href="#">NM_001276695.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens mRNA for Cellular tumor antigen p53 variant, clone: FCC127A11</a>	1844	1844	96%	0.0	97.59%	<a href="#">AK225838.1</a>
<input checked="" type="checkbox"/>	<a href="#">Human p53 cellular tumor antigen mRNA, complete cds</a>	1844	1844	96%	0.0	97.59%	<a href="#">K03199.1</a>



6 - Cliccare sul primo link di “Description” per vedere l’allineamento della sequenza query con questa sequenza target:

Descriptions Graphic Summary **Alignments** Taxonomy

Alignment view Pairwise ☒ CDS feature ? [Restore defaults](#)

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

**TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)**

Sequence ID: [HG975427.1](#) Length: 2586 Number of Matches: 1

Range 1: 54 to 1135 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1858 bits(1006)	0.0	1061/1086(98%)	9/1086(0%)	Plus/Plus

Query	35	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACT	94
Sbjct	54	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACT	113
Query	95	TTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCTGGATTGGCAGC	154
Sbjct	114	TTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCTGGATTGGCAGC	173
Query	155	CAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGAGTCAGATCCTAGCGTCGAGCCC	214
Sbjct	174	CAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGAGTCAGATCCTAGCGTCGAGCCC	233
Query	215	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCTGAAAACAACGTTCTG	274
Sbjct	234	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCTGAAAACAACGTTCTG	293
Query	275	TCCCCCTTGCCGTCCCAAGCAATGGATGATTGATGCTGTCCCCGGACGATATTGAACAA	334
Sbjct	294	TCCCCCTTGCCGTCCCAAGCAATGGATGATTGATGCTGTCCCCGGACGATATTGAACAA	353
Query	335	TGGTTCACTGAAGACCCAGGTCAGATGAAGCTCCCAGAATGCCAGAGGCTGCTCCCCC	394
Sbjct	354	TGGTTCACTGAAGACCCAGGTCAGATGAAGCTCCCAGAATGCCAGAGGCTGCTCCCCC	413
Query	395	GTGGCCCTGCACACGACGCTCCTACACGGCGGGCCCTGCACAGCCCCCTCCTGGCCC	454
Sbjct	414	GTGGCCCTGCACACGACGCTCCTACACGGCGGGCCCTGCACAGCCCCCTCCTGGCCC	473

modalità di visualizzazione dell'allineamento

informazioni sulla sequenza target

informazioni su allineamento



## 7 - Cliccare su “GenBank” per vedere le informazioni di GenBank relative a questo gene:

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

[Next](#) [Previous](#) [Descriptions](#)

### TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)

Sequence ID: [HG975427.1](#) Length: 2586 Number of Matches: 1

Range 1: 54 to 1135 [GenBank](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score		Expect	Identities	Gaps	Strand
1858 bits(1006)		0.0	1061/1086(98%)	9/1086(0%)	Plus/Plus
Query	35	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACT	94		
Sbjct	54	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACT	113		
Query	95	TTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGC	154		
Sbjct	114	TTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGC	173		
Query	155	CAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGAGTCAGATCCTAGCGTCGAGCCC	214		
Sbjct	174	CAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGAGTCAGATCCTAGCGTCGAGCCC	233		
Query	215	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAAACTACTTCTGAAAACAACGTTCTG	274		
Sbjct	234	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAAACTACTTCTGAAAACAACGTTCTG	293		
Query	275	TCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAA	334		
Sbjct	294	TCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAA	353		

### Related Information

[Gene](#) - associated gene details  
[PubChem BioAssay](#) - bioactivity screening

## 7 - Cliccare su “GenBank” per vedere le informazioni di GenBank relative a questo gene:

### TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)

GenBank: HG975427.1

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

[Go to:](#) ☐

LOCUS HG975427 2586 bp RNA linear PRI 03-MAR-2015  
DEFINITION TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene).  
ACCESSION HG975427  
VERSION HG975427.1  
DBLINK BioProject: [PRJEB6238](#)  
KEYWORDS Third Party Data; TPA; RNACentral; TPA:specialist\_db.  
SOURCE Homo sapiens (human)  
ORGANISM [Homo sapiens](#)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;  
Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Amaral,P.P., Clark,M.B., Gascoigne,D.K., Dinger,M.E. and  
Mattick,J.S.  
TITLE lncRNadb: a reference database for long noncoding RNAs  
JOURNAL Nucleic Acids Res 39 (DATABASE ISSUE), D146-D151 (2011)  
PUBMED [21112873](#)  
REFERENCE 2 (bases 1 to 2586)  
CONSTRM lncRNadb and RNACentral  
TITLE Direct Submission  
JOURNAL Submitted (30-APR-2014)  
COMMENT lncRNadb; 163; p53 mRNA.

Specialist DB : lncRNadb (The long non-coding RNA database) URL  
: <http://www.lncrnadb.org/>

PRIMARY	TPA_SPAN	PRIMARY_IDENTIFIER	PRIMARY_SPAN	COMP
	1-77	DA932053.1	1-77	
	78-411	BC003596.1	1-334	
	412-1379	X02469.1	350-1317	
	1380-2567	AK223026.1	478-1665	
	2568-2586	DQ186650.1	2257-2275	

8 - cliccare su “Gene” per vedere le informazioni di NCBI Gene relative a questo gene:

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

[Next](#) [Previous](#) [Descriptions](#)

### TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)

Sequence ID: [HG975427.1](#) Length: **2586** Number of Matches: **1**

Range 1: 54 to 1135 [GenBank](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score		Expect	Identities	Gaps	Strand	
1858 bits(1006)		0.0	1061/1086(98%)	9/1086(0%)	Plus/Plus	
Query	35	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACT				94
Sbjct	54	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACT				113
Query	95	TTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGC				154
Sbjct	114	TTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGC				173
Query	155	CAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCC				214
Sbjct	174	CAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCC				233
Query	215	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCTGAAAACAACGTTCTG				274
Sbjct	234	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCTGAAAACAACGTTCTG				293
Query	275	TCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAA				334
Sbjct	294	TCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAA				353

### Related Information

[Gene](#) - associated gene details  
[PubChem BioAssay](#) - bioactivity screening

## *Esercizi di riepilogo*

- Ricavare queste informazioni: in che tessuto è espresso maggiormente TP53? In che pathways è coinvolta?
- Tornare indietro e selezionare le sequenze con almeno il 97% d'identità e il 90% di coverage, e con E-value = 0 (*filtri*). Quante sono?

## Esercitazione 2: allineamento di sequenze nucleotidiche

1. Trovare **se** la sequenza nucleotidica query “*seq1*” di *Homo Sapiens* fa parte di un gene conosciuto;
2. Visualizzare le sequenze “*seq1*” e “*seq2*” come si allineano sul genoma e qual è il loro ruolo funzionale (codificante/non codificante/regolatoria)

1

Nucleotide Blast (NCBI)

2

BLAT (UCSC)

3

Protein Blast and Muscle



Collegarsi a a UCSC Genome Browser, BLAT ed inserire la Seq1 nello spazio apposito. Controllare che l'Assembly del genoma sia quello di Feb. 2009 e che la specie sia uomo. Avviare la ricerca:

Human BLAT Search

### BLAT Search Genome

Genome: ☐ Search all  
Human

Assembly: Feb. 2009 (GRCh37/hg19)

Query type: BLAT's guess

Sort output: query,score

Output type: hyperlink

>Seq1  
actggcgctaaaagttttgagcttttcctttacaattctcaaaagtctagagccaccggtccagggagcaggtagctgctgggctccggggacactttgcttcgggctgggagcgtgctttccacgacggtgacacgcttcctg  
gattggcagccagactgccttcgggtcactgcatggaggagcgcagtcagatcctagctgcagagccctctgagtcaggaaacattttcagacctatggaaactacttctgaaaacaacgttctgtcccttggcgtc  
ccaagcaatggatgatttgatgctgtccccggacgatattgaacaatggttctactgaagaccaggtccagatgaagctcccagaatgccagaggctgctcccccggtggccctgacacgagcagctcctacaccggcgcc  
tgaccagccccctcctggccccctgtcatcttctgtccctccagaaaaactaccaggggacctacggtttccgctggtggttcttgcattctgggacagccaagtctgtgacttgacgctactccctgcctcaacaagt  
gtttgccaactggccaagacctgccccgtgacgtgtgggttgattccacacccccgccccgtttacgcgcatggccatctacaagcagtcacagcacatgacggagggttgtagggcgctgccccaccatgagcgtgct  
cagatagcgtggtgtgccccctcctcagcatcttatccgagtggaaaccttgatttgcgtgtggagtatttggatgacagaaaacattttcgacatagtgtggtggtgccccctttaccatgctgaggttggtctgactgtac  
caccatccactcattaactacatgtgtaacagttcctgcatggcggtcatgaaccggaggccatcctcaccatcatcacactggaagactccagtggttaactactgtggagcgaacagctttgaggtgctgtgttgcctgt  
cctgggagagaccggcgccacagaggaagagaatctccgaagaaaggggagcctcaccacgagctgccccagggagcactaagcgagcactgccaacaacaccag

☐ All Results (no minimum matches)

Submit I'm feeling lucky Clear



## Human (hg19) BLAT Results

### BLAT Search Results

Go back to [chrX:15,578,261-15,621,068](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
<a href="#">browser</a> <a href="#">details</a>	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
<a href="#">browser</a> <a href="#">details</a>	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
<a href="#">browser</a> <a href="#">details</a>	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
<a href="#">browser</a> <a href="#">details</a>	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
<a href="#">browser</a> <a href="#">details</a>	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
<a href="#">browser</a> <a href="#">details</a>	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
<a href="#">browser</a> <a href="#">details</a>	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22



# What's the best alignment?

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Projects](#)[Help](#)[About Us](#)

## Human (hg19) BLAT Results

### BLAT Search Results

Go back to [chrX:15,578,261-15,621,068](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
<a href="#">browser</a> <a href="#">details</a>	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
<a href="#">browser</a> <a href="#">details</a>	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
<a href="#">browser</a> <a href="#">details</a>	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
<a href="#">browser</a> <a href="#">details</a>	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
<a href="#">browser</a> <a href="#">details</a>	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
<a href="#">browser</a> <a href="#">details</a>	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
<a href="#">browser</a> <a href="#">details</a>	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22

# What's the best alignment?

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Projects](#)[Help](#)[About Us](#)

## Human (hg19) BLAT Results

### BLAT Search Results

Go back to [chrX:15,578,261-15,621,068](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
<a href="#">browser details</a>	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
<a href="#">browser details</a>	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
<a href="#">browser details</a>	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
<a href="#">browser details</a>	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
<a href="#">browser details</a>	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
<a href="#">browser details</a>	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
<a href="#">browser details</a>	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22

### 3 - Cliccare su “details” per vedere l’allineamento:

[Home](#) [Genomes](#) [Genome Browser](#) [Tools](#) [Mirrors](#) [Downloads](#) [My Data](#) [Projects](#) [Help](#) [About Us](#)

## Human (hg19) BLAT Results

### BLAT Search Results

Go back to [chrX:15,578,261-15,621,068](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
<a href="#">browser details</a>	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
<a href="#">browser details</a>	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
<a href="#">browser details</a>	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
<a href="#">browser details</a>	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
<a href="#">browser details</a>	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
<a href="#">browser details</a>	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
<a href="#">browser details</a>	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22

## 4 - Cliccare su “together” nel menù a sinistra:

### Side by Side Alignment

```
0000001 actggcgctaaaagttttgagc 0000022
<<<<<<< ||||| <<<<<<<
7590832 actggcgctaaaagttttgagc 7590811

0000035 ttctcaaaagtctagagccaccgtccagggagcaggtagctgctgggctc 0000084
<<<<<<< ||||| <<<<<<<
7590810 ttctcaaaagtctagagccaccgtccagggagcaggtagctgctgggctc 7590761

0000085 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 0000134
<<<<<<< ||||| <<<<<<<
7590760 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 7590711

0000135 cgcttccctggattgg 0000150
<<<<<<< ||||| <<<<<<<
7590710 cgcttccctggattgg 7590695

0000151 cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc 0000200
<<<<<<< ||||| <<<<<<<
7579940 cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc 7579891

0000201 ctagcgtcgagccccctctgagtcaggaacattttcagacctatggaaa 0000250
<<<<<<< ||||| <<<<<<<
7579890 ctagcgtcgagccccctctgagtcaggaacattttcagacctatggaaa 7579841

0000251 ct 0000252
<<<<<<< || <<<<<<<
7579840 ct 7579839
```



## Side by Side Alignment

```
0000001 actggcgctaaaagttttgagc 0000022
<<<<<<< ||||| <<<<<<<
7590832 actggcgctaaaagttttgagc 7590811
```

---

```
0000035 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 0000084
<<<<<<< ||||| <<<<<<<
7590810 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 7590761
```

```
0000085 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 0000134
<<<<<<< ||||| <<<<<<<
7590760 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 7590711
```

```
0000135 cgcttccctggattgg 0000150
<<<<<<< ||||| <<<<<<<
7590710 cgcttccctggattgg 7590695
```

---

```
0000151 cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc 0000200
<<<<<<< ||||| <<<<<<<
7579940 cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc 7579891
```

```
0000201 ctagcgtcgagccccctctgagtcaggaaacattttcagacctatggaaa 0000250
<<<<<<< ||||| <<<<<<<
7579890 ctagcgtcgagccccctctgagtcaggaaacattttcagacctatggaaa 7579841
```

```
0000251 ct 0000252
<<<<<<< || <<<<<<<
7579840 ct 7579839
```

```
0000001 actggcgctaaaagttttgagc 0000022
<<<<<< |||||<
7590832 actggcgctaaaagttttgagc 7590811
```

```
0000035 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 0000084  
<<<<<< ||| | | | | | | | | | | | | | | | | | | | | | <<<<<<  
7590810 ttctcaaaagtctagaqccaccgtccaggaqgcaggtagctgctgggctc 7590761
```

```
0000085 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 0000134
<<<<<< ||||| ||||| ||||| ||||| ||||| ||||| ||||| <<<<<<
7590760 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 7590711
```

```
0000135 cgcttcctggattgg 0000150
<<<<<<< |||||  <<<<<<<
7590710 cgcttcctggattgg 7590695
```

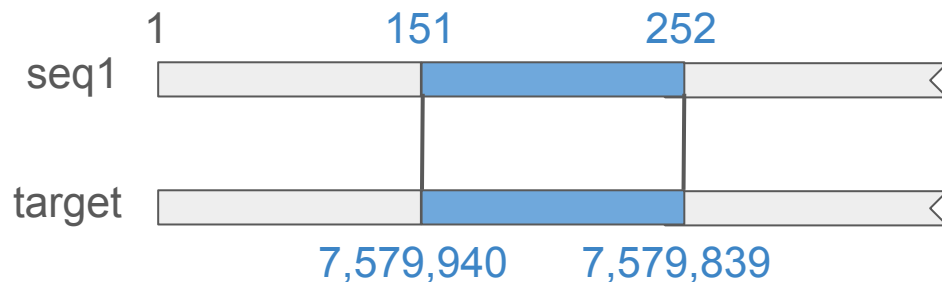
```
0000151   cagccagactgccttcgggtcactgccatggaggagccgcagtcaaatc   0000200  

<<<<<< ||||| |  

7579940   cagccagactgccttcgggtcactgccatggaggagccgcagtcaaatc   7579981
```

```
0000201 ctagcgtcgagccccctctgagtcatgaataattttcacacctatggaaa 0000250  
<<<<<< ||||| |  
7579890 ctagcgtcgagccccctctgagtcatgaataattttcacacctatqaaaa 7579841
```

```
0000251 ct 0000252
<<<<<< || <<<<<<
7579840 ct 7579839
```



```
0000001 actggcgctaaaagttttgagc 0000022
<<<<<< ||||| ||||| ||||| ||||| <<<<<<
7590832 actggcgctaaaagttttgagc 7590811
```

```
0000085 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 0000134
<<<<<< ||||| | | | | | | | | | | | | | | | | | | | | <<<<<<
7590760 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 7590711
```

```
0000135 cgcttcctggattgg 0000150
<<<<<< |||||  <<<<<<
7590710 cgcttcctggattgg 7590695
```

```
0000151   cagccagactgccttcgggtcactgccatggaggagccgcagtcaagtc   0000200  

<<<<<< ||||| |  

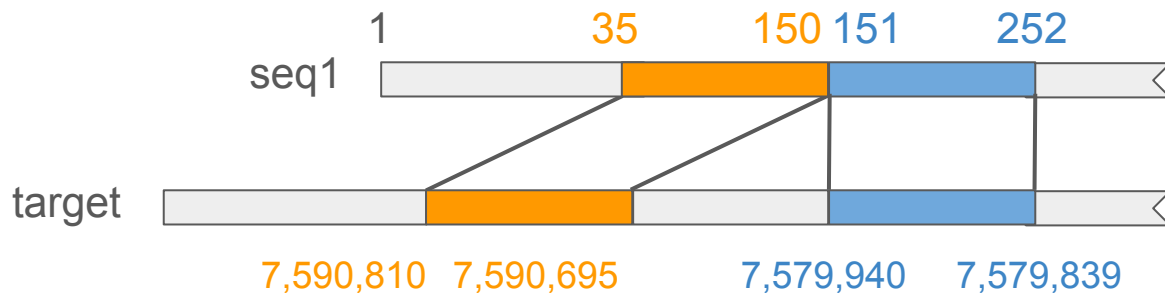
7579940   cagccagactgccttcgggtcactgccatggaggagccgcagtcaagtc   7579891
```

```
0000201 ctagcgtcgagccccctctgagtcaggaacattttcagacctatggaaa 0000250  

<<<<<< ||| | | | | | | | | | | | | | | | | | | | | <<<<<<  

7579890 ctagcgtcgagccccctctgagtcaggaacattttcagacctatqgaaa 7579841
```

```
0000251 ct 0000252
<<<<<< || <<<<<<
7579840 ct 7579839
```



## Side by Side Alignment

```
0000001 actggcgctaaaagttttgagc 0000022
<----- |||-----<-----
7590832 actggcgctaaaagttttgagc 7590811
```

```
0000035 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 0000084
<----- |||-----<-----
7590810 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 7590761
```

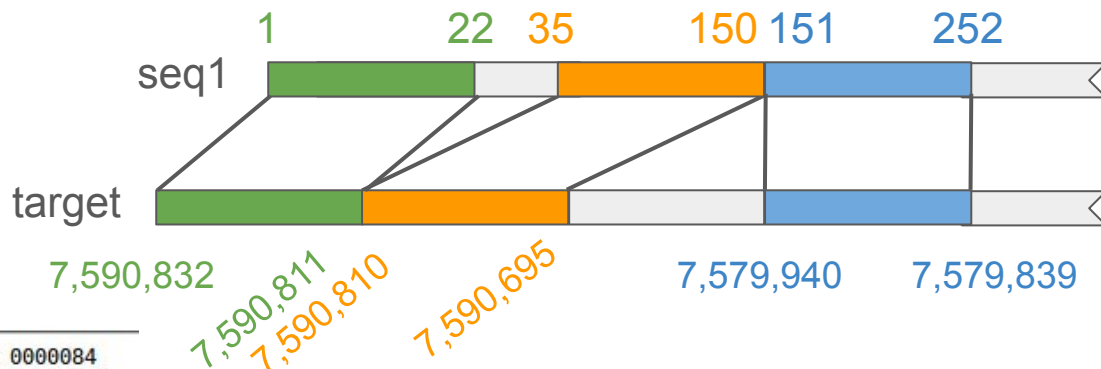
```
0000085 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 0000134
<----- |||-----<-----
7590760 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 7590711
```

```
0000135 cgcttccctggattgg 0000150
<----- |||-----<-----
7590710 cgcttccctggattgg 7590695
```

```
0000151 cagccagactgccttcgggtcactgcatggaggagccgcagtcagatc 0000200
<----- |||-----<-----
7579940 cagccagactgccttcgggtcactgcatggaggagccgcagtcagatc 7579891
```

```
0000201 ctagcgtcgagccccctctgagtcaggaacattttcagacctatggaaa 0000250
<----- |||-----<-----
7579890 ctagcgtcgagccccctctgagtcaggaacattttcagacctatggaaa 7579841
```

```
0000251 ct 0000252
<----- || <-----
7579840 ct 7579839
```

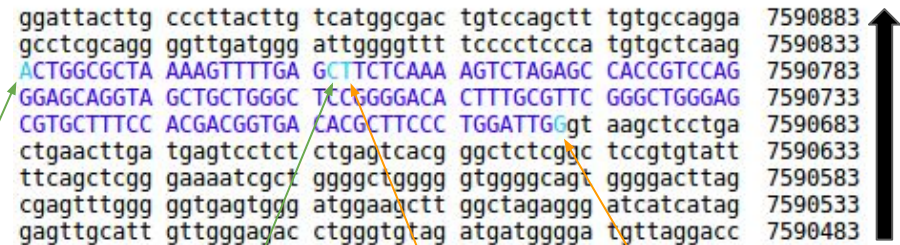




5 - Cliccare su “Seq1” nel menù a sinistra:



**Genomic chr17 (reverse strand):**



7,590,832

7,590,811

7,590,810

7,590,695

## cDNA Seq1

ACTGGCGCTA AAAGTTTTGA Gctttccttt acaaTTCTCA AAAGTCTAGA 50  
GCCACCGTCC AGGGAGCAGG TAGCTGCTGG GCTCCGGGGA CACTTTGCGT 100  
TCGGGCTGGG AGCGTGCTTT CCACGACGGT GACACGCTTC CCTGGATTGG 150  
CAGCCAGACT GCCTTCCGGG TCACTGCCAT GGAGGAGCCG CAGTCAGATC 200  
CTAGCGTCGA GCCCCCTCTG AGTCAGGAAA CATTTTCAGA CCTATGGAAA 250  
CTACTTCCTG AAAACAACGT TCTGTCCCC TTGCCGTCCC AAGCAATGGA 300

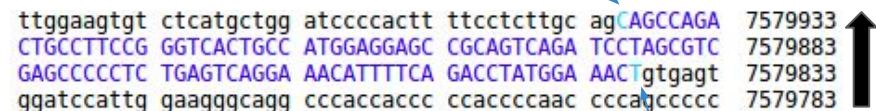


151

252

7,579,940

ttggaagtgt ctcattgctgg atccccactt ttcctcttgc agCAGCCAGA 7579933  
CTGCCTTCCG GGTCACTGCC ATGGAGGAGC CGCAGTCAGA TCCTAGCGTC 7579883  
GAGCCCCCTC TGAGTCAGGA AACATTTTCA GACCTATGGA AACTgtgagt 7579833  
ggatccattg gaagggcagg cccaccaccc ccaccccaac cccagcccc 7579783



7,579,839

6 - Tornare indietro e cliccare su “browser” per vedere la sequenza nel Genome Browser:

[Home](#) [Genomes](#) [Genome Browser](#) [Tools](#) [Mirrors](#) [Downloads](#) [My Data](#) [Projects](#) [Help](#) [About Us](#)

Human (hg19) BLAT Results

## BLAT Search Results

Go back to [chrX:15,578,261-15,621,068](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
<a href="#">browser</a> <a href="#">details</a>	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
<a href="#">browser</a> <a href="#">details</a>	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
<a href="#">browser</a> <a href="#">details</a>	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
<a href="#">browser</a> <a href="#">details</a>	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
<a href="#">browser</a> <a href="#">details</a>	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
<a href="#">browser</a> <a href="#">details</a>	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
<a href="#">browser</a> <a href="#">details</a>	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22

# UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

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

chr17:7,576,908-7,590,832 13,925 bp. enter position, gene symbol, HGVS or search terms

go

chr17 (p13.1) 13.3 13.2 13.1 17p12 17p11.2 17q11.2 17q12 21.31 17q22 24.2q24.3q25.1 17q25.3

Scale 5 kb hg19  
chr17: 7,576,000|7,579,000|7,580,000|7,581,000|7,582,000|7,583,000|7,584,000|7,585,000|7,586,000|7,587,000|7,588,000|7,589,000|7,590,000|  
Seq1 blat on Seq1

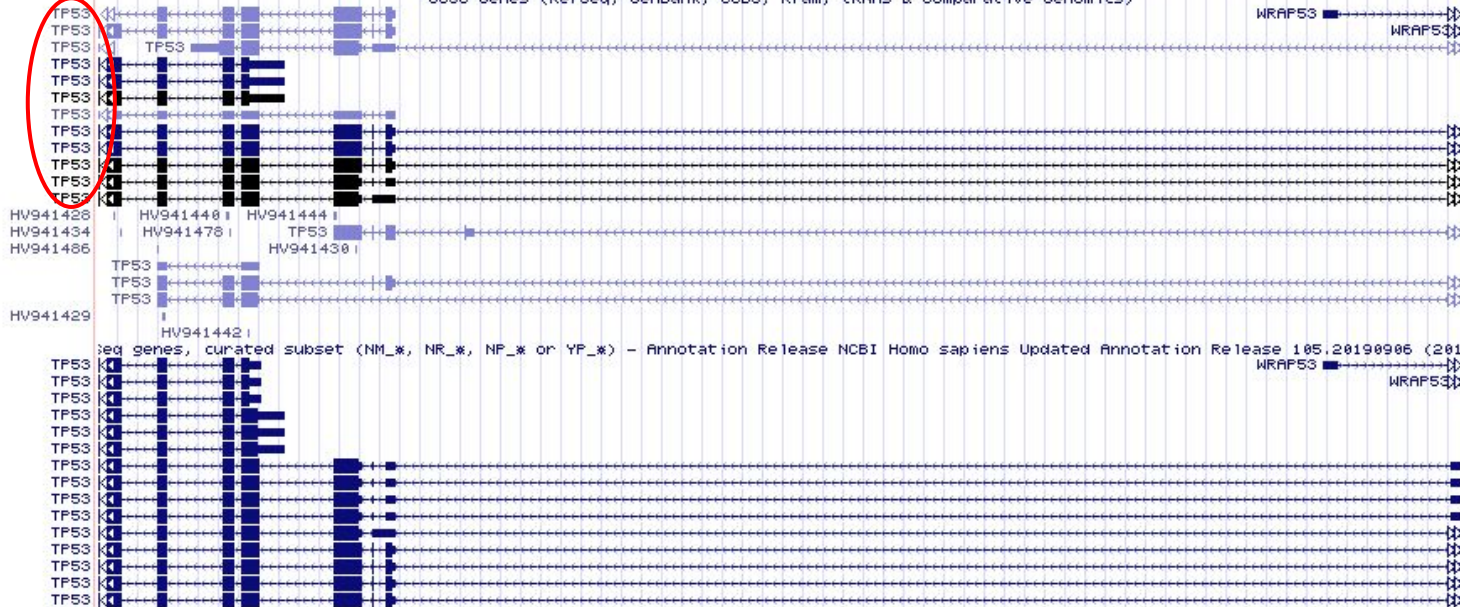
Alt Haplotypes

Seq1

Reference Assembly Fix Patch Sequence Alignments  
Reference Assembly Alternate Haplotype Sequence Alignments

Your Sequence from Blat Search

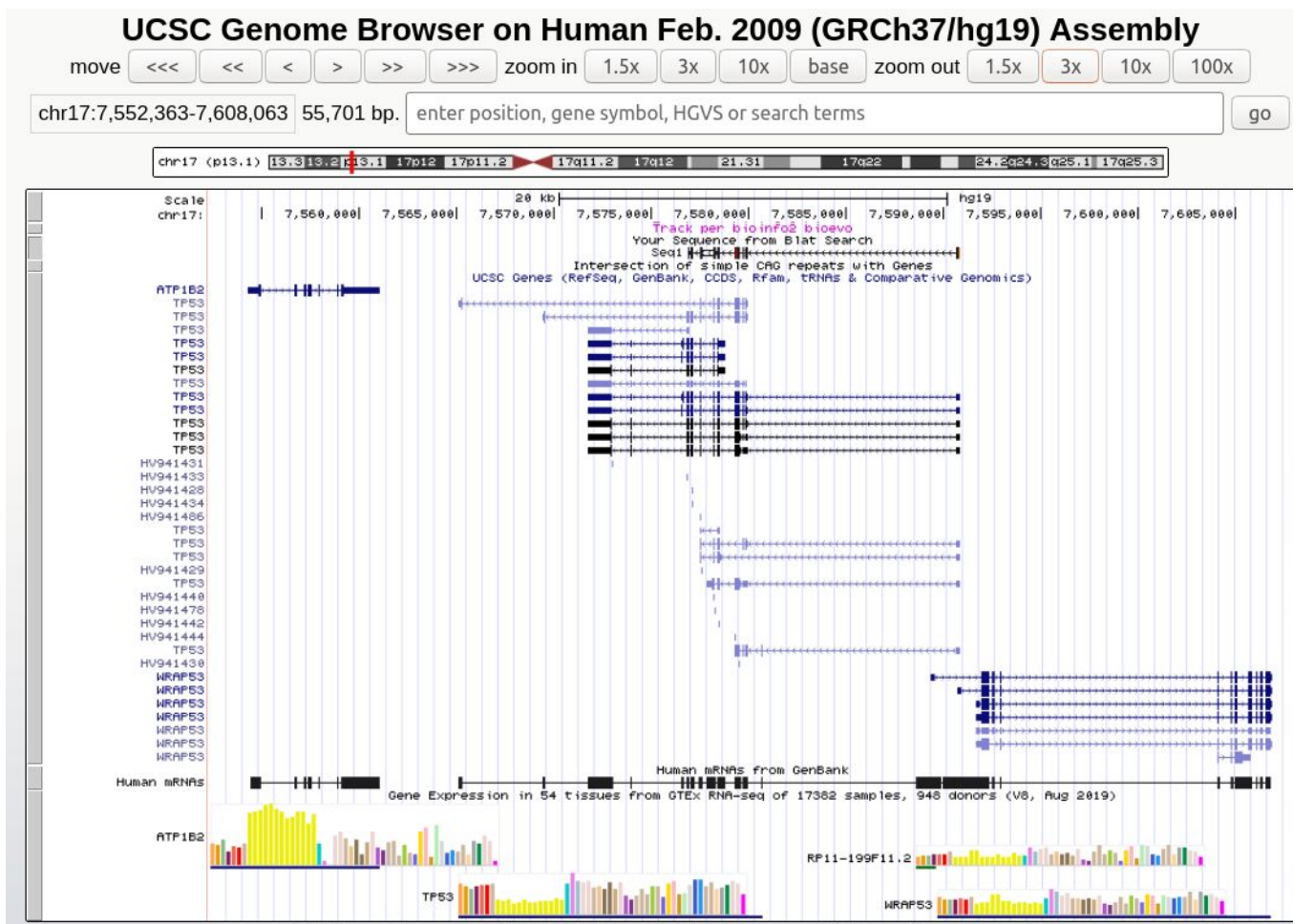
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)



Publications: Sequences in Scientific Articles



7 - Zoomare 2 volte al 3x per vedere a che porzione di gene corrisponde la sequenza query:



8 - Tornare indietro ed inserire la seconda sequenza nel form:

## Human BLAT Search

## BLAT Search Genome

Genome: ☐ Search all

Assembly:

Query type:

Sort output:

Output type:

Human

Feb. 2009 (GRCh37/hg19) ▾

BLAT's guess

query,score

hyperlink

```
>Seq1
```

actggcgcctaaaagttttgagcttttctttacaattttctaaaagcttagagccaccgtccagggnagcaggtagctgctgggctccggggacactttgctgttgggctggggagctgtctttcacgacggtagacacgttc  
cctggattggcagccagagctgctcttccgggtcactgcatggaggagccgagtcagattcttagctctgagcccccctctgagtcaggaaacacttttcagaccattggaaactacttctctgaaacaacactctctgtctccc  
ttggcctgcctaaagcaatggatgattttgatgctgtccctcgagcgaatttgaacaaatggctcactcagaaagccagggtccagatgaagctcccgagatggcagaggtgctctccccctggggccctctgacagctctctac  
atggggcgccctctcagcagagctctctggccctctgactctcttctctctctccttcagaaaacactaccaggggagcctacaggtttccgtctgggctcttctgattctgggagaccgaagctctgtgacttgcagctactccc  
ctgcccctcaacaagatgttttgcgaactggccaagacctgcccctgtgcagctgtgggttgattccacacccccggccggctttacgcgccattggccatctacaagcagtcacagccatatgacggaggttgtgagggcgtg  
ccccaccattgaagctctctcagatagcgatgtctgtggccctctcagactcttatccgagttggaacctatgttgcctctggagattttggaatgacagaaacacttttgcacattagtgtggtgatgtccctttttacca  
cttgagaggttggctctgactgtgtaccaccactcactcttaactacatctgtgaacagttctctgcatggcgccggtatgaaccggggggccatctcaccatcactctctggaagactcagttggtatctactctgggacggaa  
cagcttttgagatctgattttgtctctgtctgggagagagccggcgacagagaagaagaatctcccaagaaagagggagacctaccacagactctccccaggagagcactaagccagcactctcccaacacaccag

```
>Seq2
```

[illegible]☐ All Results (no minimum matches)

Submit

I'm feeling lucky

Clear

Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.

**File Upload:** Rather than pasting a sequence, you can choose to upload a text file containing the sequence.

Upload sequence:

Sfoggia...

Nessun file selezionato.

submit file



## Human (hg19) BLAT Results

## BLAT Search Results

Go back to [chr17:7576908-7590832](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
<a href="#">browser details</a>	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
<a href="#">browser details</a>	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
<a href="#">browser details</a>	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
<a href="#">browser details</a>	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
<a href="#">browser details</a>	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
<a href="#">browser details</a>	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
<a href="#">browser details</a>	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22
<a href="#">browser details</a>	Seq2	910	1	996	996	97.1%	chr17	-	7571736	7572719	984
<a href="#">browser details</a>	Seq2	171	551	856	996	85.0%	chr5	+	49950222	49950507	286
<a href="#">browser details</a>	Seq2	167	547	854	996	85.6%	chr10	-	27408468	27408791	324
<a href="#">browser details</a>	Seq2	158	566	844	996	86.0%	chr22	+	28176923	28177196	274
<a href="#">browser details</a>	Seq2	156	552	834	996	83.6%	chr2	+	37248781	37249047	267
<a href="#">browser details</a>	Seq2	155	553	865	996	85.0%	chr6	-	38263541	38263850	310
<a href="#">browser details</a>	Seq2	149	547	865	996	86.4%	chr10	-	104340658	104365268	24611
<a href="#">browser details</a>	Seq2	148	550	856	996	83.6%	chr15	-	43833643	43833952	310
<a href="#">browser details</a>	Seq2	148	561	865	996	82.6%	chr2	+	26597435	26597736	302
<a href="#">browser details</a>	Seq2	147	550	856	996	86.3%	chr4	+	101191092	101191390	299
<a href="#">browser details</a>	Seq2	146	565	857	996	82.5%	chr9	+	80078922	80079207	286
<a href="#">browser details</a>	Seq2	144	548	851	996	89.2%	chr3	-	73757750	73758057	308
<a href="#">browser details</a>	Seq2	143	552	855	996	85.3%	chr6	+	155350920	155351215	296
<a href="#">browser details</a>	Seq2	143	550	844	996	84.8%	chr15	+	75466123	75466409	287
<a href="#">browser details</a>	Seq2	141	551	855	996	85.8%	chr3	-	167045434	167045723	290

seq1

seq2





## Human (hg19) BLAT Results

## BLAT Search Results

Go back to [chr17:7576908-7590832](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
<a href="#">browser details</a>	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
<a href="#">browser details</a>	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
<a href="#">browser details</a>	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
<a href="#">browser details</a>	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
<a href="#">browser details</a>	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
<a href="#">browser details</a>	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
<a href="#">browser details</a>	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22
<a href="#">browser details</a>	Seq2	910	1	996	996	97.1%	chr17	-	7571736	7572719	984
<a href="#">browser details</a>	Seq2	171	551	856	996	85.0%	chr5	+	49950222	49950507	286
<a href="#">browser details</a>	Seq2	167	547	854	996	85.6%	chr10	-	27408468	27408791	324
<a href="#">browser details</a>	Seq2	158	566	844	996	86.0%	chr22	+	28176923	28177196	274
<a href="#">browser details</a>	Seq2	156	552	834	996	83.6%	chr2	-	37248781	37249047	267
<a href="#">browser details</a>	Seq2	155	553	865	996	85.0%	chr6	-	38263541	38263850	310
<a href="#">browser details</a>	Seq2	149	547	865	996	86.4%	chr10	-	104340658	104365268	24611
<a href="#">browser details</a>	Seq2	148	550	856	996	83.6%	chr15	-	43833643	43833952	310
<a href="#">browser details</a>	Seq2	148	561	865	996	82.6%	chr2	+	26597435	26597736	302
<a href="#">browser details</a>	Seq2	147	550	856	996	86.3%	chr4	+	101191092	101191390	299
<a href="#">browser details</a>	Seq2	146	565	857	996	82.5%	chr9	+	80078922	80079207	286
<a href="#">browser details</a>	Seq2	144	548	851	996	89.2%	chr3	-	73757750	73758057	308
<a href="#">browser details</a>	Seq2	143	552	855	996	85.3%	chr6	+	155350920	155351215	296
<a href="#">browser details</a>	Seq2	143	550	844	996	84.8%	chr15	+	75466123	75466409	287
<a href="#">browser details</a>	Seq2	141	551	855	996	85.8%	chr3	-	167045434	167045723	290

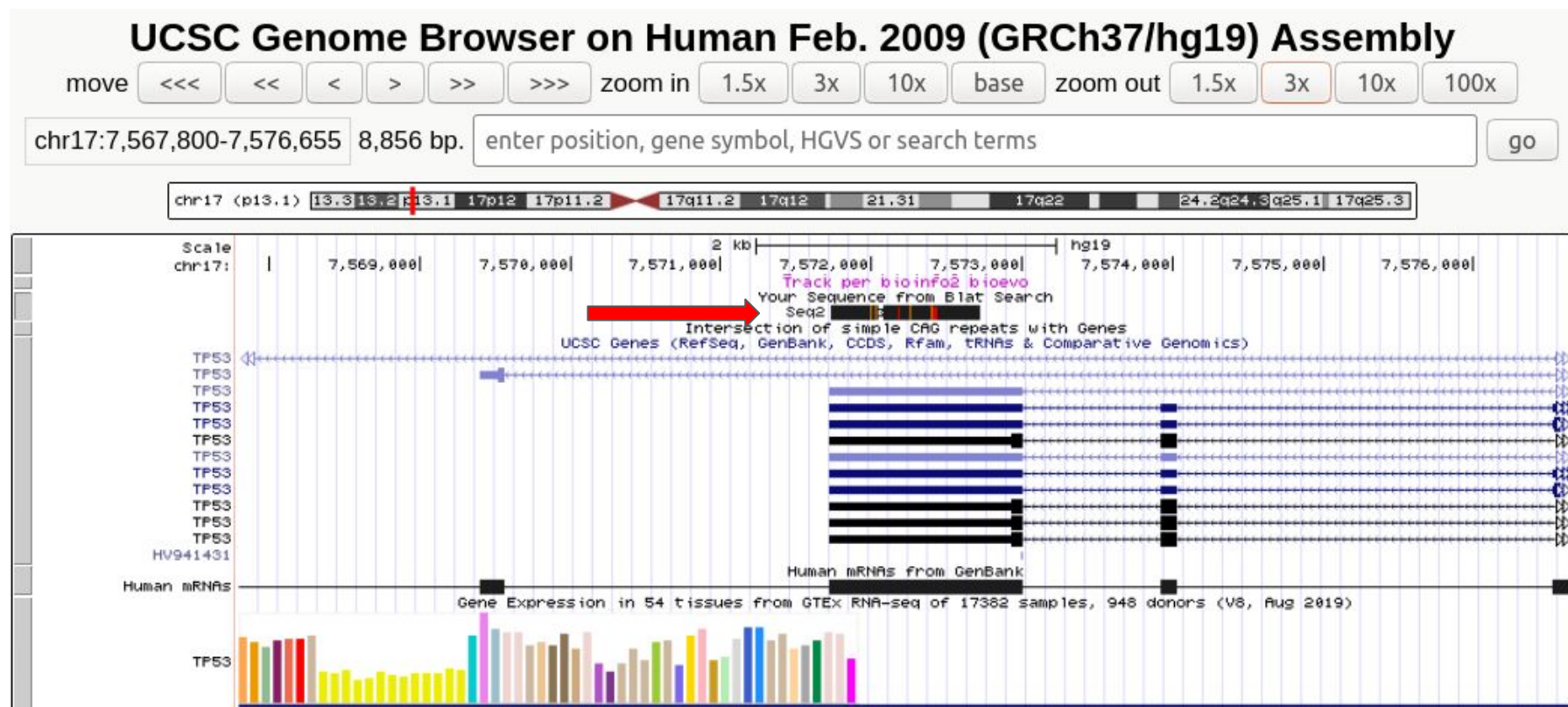
seq1

seq2

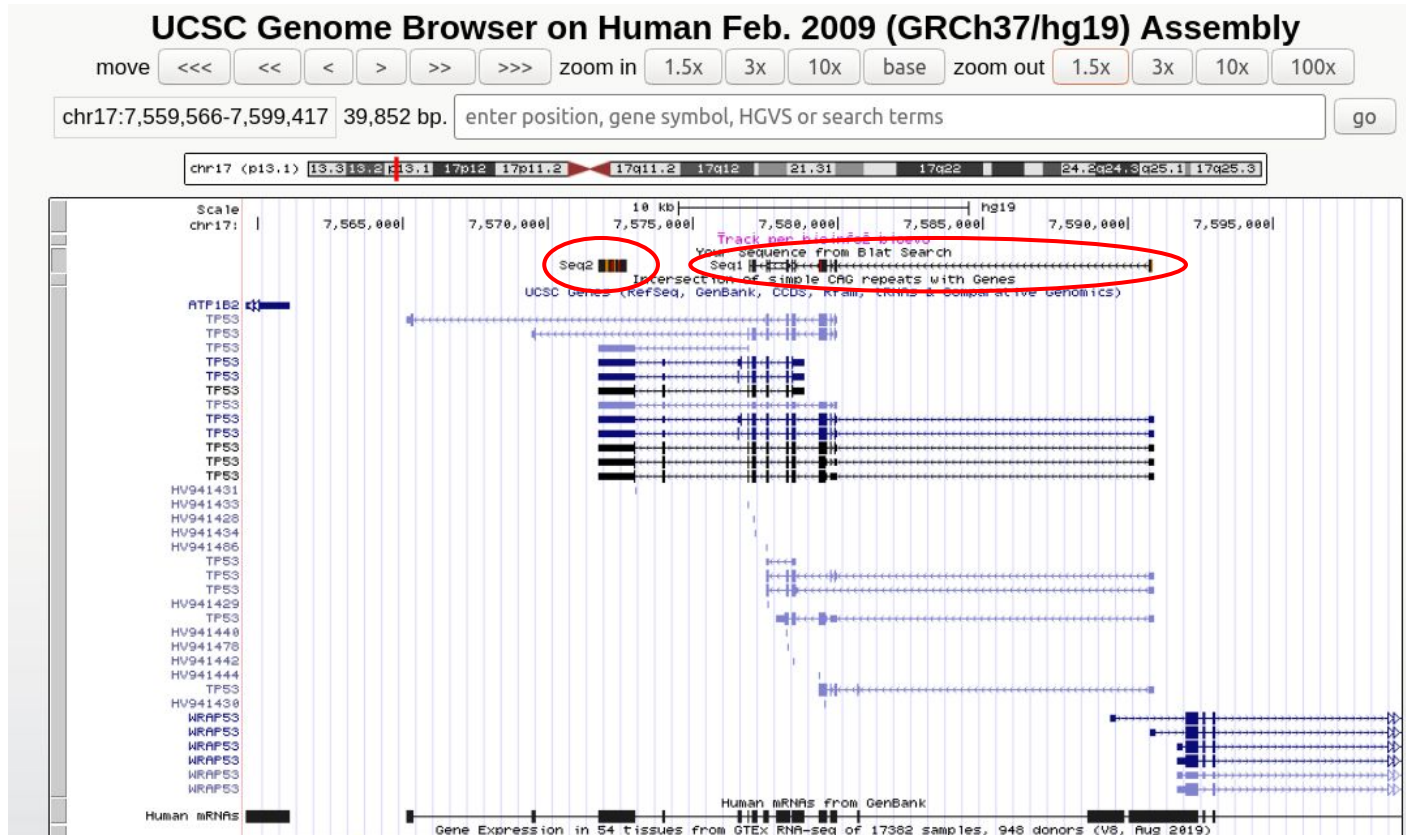




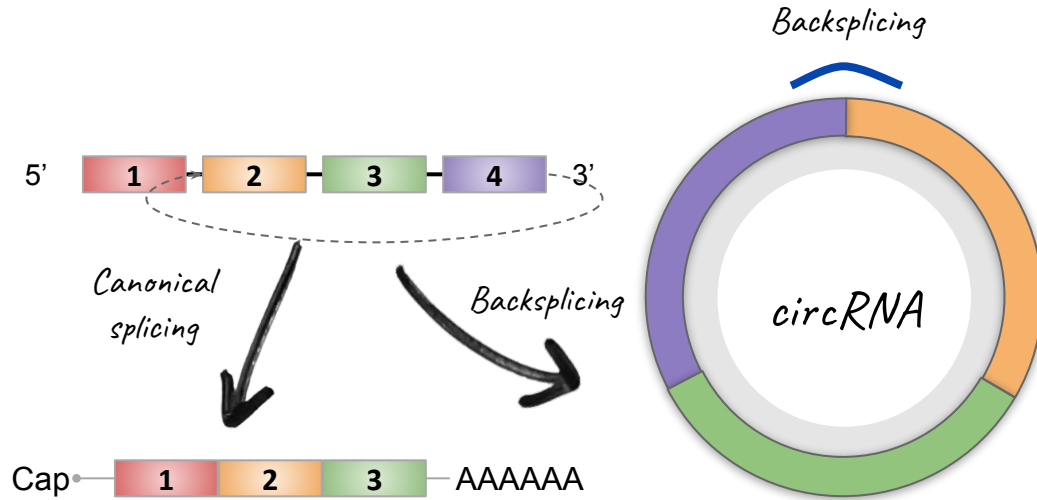
9 - Zoomare 2 volte al 3x per vedere a che porzione di gene corrisponde la sequenza query:



10 - Spostare un po' il browser verso sinistra e zoomare prima al 3x e poi al 1.5x per vedere come si posizionano le due sequenze sul gene TP53:

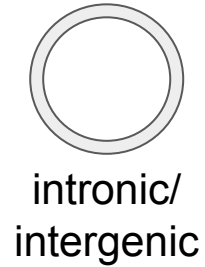
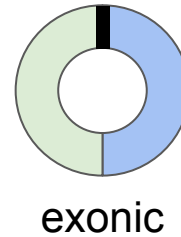
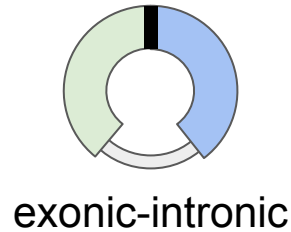


# circular RNA



- Covalently closed RNA molecules
- Ubiquitary
- CircRNA structure evolutionarily conserved
- Expression varied with tissue and developmental stage
- More stable than linear RNAs
- Detected in plasma and saliva
- Conferred to extracellular bodies e.g. exosomes

- Dove mappano sul genoma umano?
- Derivano da regioni introniche/esoniche/intergeniche?



# *Esercizio di riepilogo*

## STEP 1

Per ogni sequenza  
nucleotidica della  
regione del  
backsplicing

- **circ1** e **circ2** nella cartella  
esercitazione2 con sequenze  
nucleotidiche ottenute dal  
genoma di riferimento **hg38**

## STEP 1

Per ogni sequenza  
nucleotidica della  
regione del  
backsplicing

## STEP 2

Ottenere  
l'allineamento della  
sequenza in **BLAT**

## STEP 1

Per ogni sequenza  
nucleotidica della  
regione del  
backsplicing

## STEP 2

Ottenere  
l'allineamento della  
sequenza in **BLAT**

## STEP 3

Analizzare risultato  
nel browser per  
caratterizzare i  
circRNA



## Esercitazione 2: allineamento multiplo di sequenze proteiche

1. Data una sequenza proteica query di *Homo sapiens* identificare a quale proteina appartiene ed evidenziare se appartiene ad una famiglia di proteine
2. Allineare la proteina con alcuni altri membri della famiglia appartenenti anche ad altre specie (*Tursiops truncatus*, *Rattus norvegicus*, *Pan troglodytes*) presenti nel database Reference proteins (RefSeq protein)
3. Ottenere ed analizzare il cladogramma e l'albero filogenetico risultante dall'allineamento multiplo

## Esercitazione 2: allineamento multiplo di sequenze proteiche

1. Data una sequenza proteica query di *Homo sapiens* identificare a quale proteina appartiene ed evidenziare fa parte di una famiglia di proteine
2. Allineare la proteina con alcuni altri membri della famiglia appartenenti anche ad altre specie (*Tursiops truncatus*, *Rattus norvegicus*, *Pan troglodytes*) presenti nel database Reference proteins (RefSeq protein)
3. Ottenere ed analizzare il cladogramma e l'albero filogenetico risultante dall'allineamento multiplo

1

Nucleotide Blast (NCBI)

2

BLAT (UCSC)

3

Protein Blast and Muscle

## 1 - Entrare in Protein BLAST:

### 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](#)

NEW

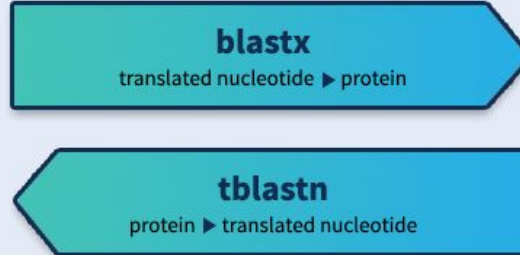
**BLAST now supports the new RefSeq Select databases.**

The RefSeq Select data-set consists of a representative or “Select” transcript for every protein-coding gene.

Tue, 13 Oct 2020 12:00:00 EST

[More BLAST news...](#)

### Web BLAST



### BLAST Genomes

Enter organism common name, scientific name, or tax id

Search

Human

Mouse

Rat

Microbes

2 - Incollare la sequenza 3 in formato FASTA, scegliere il database UniProtKB/Swiss-Prot ed avviare la ricerca con i parametri di default:

Standard Protein BLAST

blastn blastp **blastx** tblastn tblastx

BLASTP programs search protein databases using a protein query sequence.

### Enter Query Sequence

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

From

To

**>seqP**  
VDRPGRIPITSSRCNTIPDDNOVVSLSPGSRVYVLRPVCFEKGNTYVRLPYITSSSDVESPYT  
LIDSLVLMYPYCKSLDFTYGGSGGVYTNASWETEDRYRCLENSRSVYKTRMTDVCNIIETISALLHQ  
TGLACECDPQGLSSVCDPNGGQCCRPNVVGRICNRICAPGTEGEGSGCKPCECHLG

Or, upload file  Nessun file selezionato.

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

☐ Align two or more sequences

### Choose Search Set

Database **UniProtKB/Swiss-Prot(swissprot)** [+](#)

Organism [Optional](#)  ☐ exclude [+](#)

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

Exclude [Optional](#) ☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

### Program Selection

Algorithm ☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

[Choose a BLAST algorithm](#)

**BLAST** Search database swissprot using Blastp (protein-protein BLAST)

☐ Show results in a new window

[+ Algorithm parameters](#) **Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign**

### 3 - Analizzare il risultato della ricerca di similarità, per capire a quale proteina umana corrisponde il frammento di sequenza in analisi (vedere “Descriptions”, “Alignments”, “GenPept” e “Gene”):

<div> <div>Descriptions</div> <div>Graphic Summary</div> <div>Alignments</div> <div>Taxonomy</div> </div>									
Sequences producing significant alignments						Download ▾	Manage columns ▾	Show	100 ▾ ?
<input checked="" type="checkbox"/> select all 72 sequences selected						<a href="#">GenPept</a>	<a href="#">Graphics</a>	<a href="#">Distance tree of results</a>	<a href="#">Multiple alignment</a>
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession		
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Laminin subunit beta-1; AltName: Full=Laminin B1 chain; AltName: Full=Laminin-1 subunit beta; AltName: Full=Laminin-10 subunit</a>	410	635	100%	4e-132	100.00%	<a href="#">P07942.2</a>		
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Laminin subunit beta-1; AltName: Full=Laminin B1 chain; AltName: Full=Laminin-1 subunit beta; AltName: Full=Laminin-10 subunit</a>	393	615	100%	3e-126	94.42%	<a href="#">P02469.3</a>		
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Laminin subunit beta-2; AltName: Full=Laminin-11 subunit beta; AltName: Full=Laminin-14 subunit beta; AltName: Full=Laminin-15</a>	176	361	100%	9e-50	45.64%	<a href="#">Q61292.2</a>		
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Laminin subunit beta-2; AltName: Full=Laminin chain B3; AltName: Full=Laminin-11 subunit beta; AltName: Full=Laminin-14 subunit</a>	175	362	100%	1e-49	45.69%	<a href="#">P15800.1</a>		
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Laminin subunit beta-2; AltName: Full=Laminin B1s chain; AltName: Full=Laminin-11 subunit beta; AltName: Full=Laminin-14 subunit</a>	164	354	100%	9e-46	44.28%	<a href="#">P55268.2</a>		
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Laminin subunit beta-4; Flags: Precursor [Danio rerio]</a>	124	410	94%	1e-31	37.43%	<a href="#">Q8JHV6.1</a>		
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Laminin subunit beta-1; AltName: Full=Laminin B1 chain; Flags: Precursor [Drosophila melanogaster]</a>	120	347	98%	1e-30	37.50%	<a href="#">P11046.4</a>		
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Laminin subunit beta-4; AltName: Full=Laminin beta-1-related protein; Flags: Precursor [Homo sapiens]</a>	102	341	88%	4e-24	36.57%	<a href="#">A4D0S4.1</a>		



### 3 - Analizzare il risultato della ricerca di similarità, per capire a quale proteina umana corrisponde il frammento di sequenza in analisi (vedere “Descriptions”, “Alignments”, “GenPept” e “Gene”):

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download ▾

Manage columns ▾

Show

100 ▾

?

**LAMB1** laminin subunit beta 1 [*Homo sapiens* (human)]

Gene ID: 3912, updated on 20-Sep-2020

Summary

⌵ ?

**Official Symbol**

LAMB1 provided by [HGNC](#)

**Official Full Name**

laminin subunit beta 1 provided by [HGNC](#)

**Primary source**

[HGNC:HGNC:6486](#)

**See related**

[Ensembl:ENSG00000091136](#) [MIM:150240](#)

**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**

CLM; LIS5

**Summary**

Laminins, a family of extracellular matrix glycoproteins, are the major noncollagenous constituent of basement membranes. They have been implicated in a wide variety of biological processes including cell adhesion, differentiation, migration, signaling, neurite outgrowth and metastasis. Laminins are composed of 3 non identical chains: laminin alpha, beta and gamma (formerly A, B1, and B2, respectively) and they form a cruciform structure consisting of 3 short arms, each formed by a different chain, and a long arm composed of all 3 chains. Each laminin chain is a multidomain protein encoded by a distinct gene. Several isoforms of each chain have been described. Different alpha, beta and gamma chain isomers combine to give rise to different heterotrimeric laminin isoforms which are designated by Arabic numerals in the order of their discovery, i.e. alpha1beta1gamma1 heterotrimer is laminin 1. The biological functions of the different chains and trimer molecules are largely unknown, but some of the chains have been shown to differ with respect to their tissue distribution, presumably reflecting diverse functions in vivo. This gene encodes the beta chain isoform laminin, beta 1. The beta 1 chain has 7 structurally distinct domains which it shares with other beta chain isomers. The C-terminal helical region containing domains I and II are separated by domain alpha, domains III and V contain several EGF-like repeats, and domains IV and VI have a globular conformation. Laminin, beta 1 is expressed in most tissues that produce basement membranes, and is one of the 3 chains constituting laminin 1, the first laminin isolated from Engelbreth-Holm-Swarm (EHS) tumor. A sequence in the beta 1 chain that is involved in cell attachment, chemotaxis, and binding to the laminin receptor was identified and shown to have the capacity to inhibit metastasis. [provided by RefSeq, Aug 2011]

**Expression**

Broad expression in placenta (RPKM 75.7), fat (RPKM 65.2) and 23 other tissues [See more](#)

**Orthologs**

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

## Esercitazione 2: allineamento multiplo di sequenze proteiche

1. Data una sequenza proteica query di *Homo sapiens* identificare a quale proteina appartiene ed evidenziare fa parte di una famiglia di proteine
2. Allineare la proteina con alcuni altri membri della famiglia di altre specie di mammiferi (*Tursiops truncatus*, *Rattus norvegicus*, *Pan troglodytes*) presenti nel database RefSeq protein
3. Ottenere ed analizzare il cladogramma e l'albero filogenetico risultante dall'allineamento multiplo



- 1 - Recuperare la sequenza di riferimento (Accession number, FASTA), salvarla in un file di testo e cambiare l'header in modo da renderlo più leggibile, es. ">LAMB1\_HUMAN\_precursor":

```
*Documento senza titolo 1
>LAMB1_HUMAN_precursor
MGLLQLLAFSFLALCRARVRAQEPEFSYGCAEGSCYPATGDLIGRAQKLSVTSTCGLHKPEPYCIVSHL
QEDKKCFICNSQDPYHETLNPDSHLIENVVTTFAPNRLKIWWQSENGVENVTIQLDLEAEFHFTHLIMTF
KTRPAAMLIERSDFGKTWGVYRYFAYDCEASFPGISTGPMKKVDDIICDSRYSDIEPSTEGEVIFRAL
DPAFKIEDPYPSPRIQNLLKITNLRIKFVKLHLLGDNLLDSRMEIREKYYYAVYDMVVRGNCFCYGHASEC
APVDGFGNEEVEGMVHGCMCRHNTKGLNCEL CMDFYHDL PWRPAEGRNSNACKKCNENEHSISCHFDMAV
YLATGNVSGGVCDQCQNTMGRNCEQCKPFYQHPERDIRDPNFCERCTCDPAGSQNEGICDSYTDFTSG
LIAGQCRCKLNVEGEHCDVCKEGFYDLSSDPFGCKSCACNPLGTIPGGNPCDSETHCYCKRLVTGQHC
DQCLPEHWGLSNDLDGCRPCDCDLGGALNNSCAESGQCSCRPHMIGRQCNEVEPGYYFATLDHYLYEAE
EANLPGVSIIVERQYIQDRIPSWTGAGFVRVPEGAYLEFFIDNIPYSMEYDILIRYEPQLPDHWEKAVIT
VQRPGRIPITSSRCGNTIPDDDNQVVSLSPGSRVVLPRPVCFEKGTNYTVRLELPQYTSDDSDVESPYTL
IDSLVLMFYCKSLDIFTVGGSGDGVVTSAWETFRQYRCLENSRSVVKTPMTDVCNIIIFSISALLHQTG
LACECDPQGSLSVCDPNGGQCQCRPNVVGRTCNRCAPGTGFGPSGCKPCECHLQGSVNAFCNPVTGQC
HCFQGVYARQCDRCLPGHWGFPSCQPCQCNHADDQDVTGECLNCQDYTMGHNCERCLAGYYGDP IIGS
GDHCRPCPCPDGPDGRQFARSCYQDPVTLQLACVCDPGYIGSRCDDCASGYFGNPSEVGGSCQPCQCHN
NIDTTDPEACDKETGRCLKLYHTEGEHCQFCRFGYYGDALQQDCRKCVCNYLGTVQEHNGSDCQCDKA
TGQCLCLPNVIGQNCDRCAPNTWQLASGTGDCPCNCAAHSGPSCNEFTGQCQCMFGFGGRTCTSECQEL
FWGDPDVECRACDCDPRGIETPQCDQSTGQCVCVEGVEGPRCDKCTRGYSGVFPDCTPCHQCFALWDVII
AELTNRTHRFLKAKALKISGVIGPYRETVDSEVERKVSEIKDILAQSPAAEPLKNIGNLFEAEKLIKDV
TEMAQVEVKLSDTTSQSNSTAKELDSLQTEAESLDNTVKELAEQLEFIKNSDIRGALDSITKYFQMSLE
AEERVNASTTEPNSTVEQSALMRDRVEDVMERESQFKEKQEEQARLLDELAGKLQSLDLASAAEMTCGT
PPGASCSETECGGPNCRTDEGERKCGGPGCGGLVTVAHNAWQKAMDLDQDVL SALAEVEQLSKMVSEAKL
RADEAKQSAEDILLKTNATKEKMDKSNEELRNLIKQIRNFLTQDSADLDSIEAVANEVLKMEMPSTPQQL
QNLTEDIRERVESLSQVEVILQHSAAIARAEMLL EEAKRASKSATDVKVTADMVKEALEEAEKAQVAAE
KAIKQADEDIQGTQNLTSIESETAASEETLFNASQRISELERNVEELKRKAAQNSGEAEYIEKVYTVK
QSAEDVKKTLDGELDEKYKKVENLIAKKTEESADARRKAEMLQNEAKTL LAQANSKLQLLKD LERKYEDN
QRYLEDKAQELARLEGEVRSLLKDISQKVAVYSTCL
```

2 - Ritornare a Protein BLAST, usare come query la sequenza appena recuperata limitando la ricerca a “Reference proteins (refseq protein)” di proteine nei gruppi tassonomici *Homo Sapiens*, *Tursiops truncatus*, *Rattus norvegicus* e *Pan troglodytes* ed avviare la ricerca:

Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#)

### Enter Query Sequence

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

Query subrange [?](#)

From

To

MGLLQLLAFLALCRARVRAQEPEFSYGCAEGSCYPATGDLIGRAQKLSV  
TSTCGLHKPEPYCIVSHLQEDKKCFICNSQDPYHETLNPDSHLIENVVTFAP  
NRLKIWWQSENGVENVTIQDLEAEHFHFLIMTFKTRPAAMLIERSSDFCK

Or, upload file  Nessun file selezionato [?](#)

Job Title

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

☐ Align two or more sequences [?](#)

### Choose Search Set

Database  [?](#)

Organism Optional

<input type="text" value="Homo sapiens (taxid:9606)"/>	<input type="checkbox"/> exclude	<a href="#">Add organism</a>
<input type="text" value="Tursiops truncatus (taxid:9739)"/>	<input type="checkbox"/> exclude	
<input type="text" value="Rattus norvegicus (taxid:10116)"/>	<input type="checkbox"/> exclude	
<input type="text" value="Pan troglodytes (taxid:9598)"/>	<input type="checkbox"/> exclude	

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

Exclude Optional

☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

### Program Selection

Algorithm

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

QuickBLASTP is an accelerated version of BLASTP that is very fast and works best if the target percent identity is 50% or more.

BlastP simply compares a protein query to a protein database.

PSI-BLAST allows the user to build a PSSM (position-specific scoring matrix) using the results of the first BlastP run.

PHI-BLAST performs the search but limits alignments to those that match a pattern in the query.

DELTA-BLAST constructs a PSSM using the results of a Conserved Domain Database search and searches a sequence database.

3 - Selezionare le sequenze di *Homo sapiens*, *Pan troglodytes*, *Tursiops truncatus*, *Rattus norvegicus* della famiglia delle laminine **beta 1 e 3** (se ci sono più isoforme, scegliere “precursor” per *Hs* e “X1” per *Rn*)

4- Recuperarle in formato **FASTA** (“**Download**”) e copiarle in un file di testo.

5- Modificare gli headers delle sequenze in modo da renderli brevi, informativi ed univoci (ad es. “>Hs\_lamb1\_pre”, “>Pt\_lamb1”)

Sequences producing significant alignments									
Download		New	Select columns	Show	100				
<input type="checkbox"/> select all 8 sequences selected		GenPept	Graphics	Distance tree of results	Multiple alignment	New	MSA Viewer		
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">laminin subunit beta-1 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	3705	3705	100%	0.0	100.00%	1786	<a href="#">NP_002282.2</a>
<input checked="" type="checkbox"/>	<a href="#">laminin subunit beta-1 [Pan troglodytes]</a>	<a href="#">Pan troglodytes</a>	3691	3691	100%	0.0	99.61%	1786	<a href="#">XP_001165667.4</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-1 isoform X1 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	3678	3678	99%	0.0	99.72%	1810	<a href="#">XP_016867690.1</a>
<input checked="" type="checkbox"/>	<a href="#">laminin subunit beta-1 [Tursiops truncatus]</a>	<a href="#">Tursiops truncatus</a>	3492	3492	100%	0.0	93.17%	1786	<a href="#">XP_033718774.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-1 isoform X2 [Rattus norvegicus]</a>	<a href="#">Rattus norvegicus</a>	3438	3438	100%	0.0	92.61%	1786	<a href="#">XP_006240058.1</a>
<input checked="" type="checkbox"/>	<a href="#">laminin subunit beta-1 isoform X1 [Rattus norvegicus]</a>	<a href="#">Rattus norvegicus</a>	3437	3437	100%	0.0	92.61%	1834	<a href="#">XP_003750185.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-1 isoform X2 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	2368	2688	65%	0.0	99.56%	1212	<a href="#">XP_016867691.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-2 isoform X2 [Tursiops truncatus]</a>	<a href="#">Tursiops truncatus</a>	1907	1907	98%	0.0	51.28%	1803	<a href="#">XP_033720801.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-2 isoform X1 [Rattus norvegicus]</a>	<a href="#">Rattus norvegicus</a>	1905	1905	99%	0.0	51.18%	1801	<a href="#">XP_006243771.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-2 precursor [Rattus norvegicus]</a>	<a href="#">Rattus norvegicus</a>	1905	1905	99%	0.0	51.18%	1801	<a href="#">NP_037106.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-2 isoform X2 [Rattus norvegicus]</a>	<a href="#">Rattus norvegicus</a>	1901	1901	99%	0.0	51.18%	1800	<a href="#">XP_038936809.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-2 isoform X1 [Tursiops truncatus]</a>	<a href="#">Tursiops truncatus</a>	1899	1899	98%	0.0	51.10%	1809	<a href="#">XP_033720799.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-2 [Pan troglodytes]</a>	<a href="#">Pan troglodytes</a>	1882	1882	98%	0.0	50.65%	1798	<a href="#">XP_016796574.2</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-2 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	1878	1878	98%	0.0	50.65%	1798	<a href="#">NP_002283.3</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-2 isoform X3 [Rattus norvegicus]</a>	<a href="#">Rattus norvegicus</a>	1763	1763	93%	0.0	50.51%	1667	<a href="#">XP_038936810.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-4 isoform 1 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	1358	1358	98%	0.0	41.31%	1761	<a href="#">NP_001304975.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-4 isoform X1 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	1336	1336	95%	0.0	41.32%	1772	<a href="#">XP_011514277.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-4 isoform X2 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	1336	1336	95%	0.0	41.32%	1753	<a href="#">XP_016867368.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-4 isoform X3 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	1334	1334	95%	0.0	41.31%	1723	<a href="#">XP_011514280.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-4 isoform X4 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	1334	1334	96%	0.0	41.22%	1703	<a href="#">XP_011514281.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-4 isoform X6 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	1312	1312	86%	0.0	43.34%	1564	<a href="#">XP_011514282.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-4 isoform X5 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	1311	1311	86%	0.0	43.34%	1567	<a href="#">XP_016867369.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-4 isoform 2 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	1081	1375	64%	0.0	49.77%	1101	<a href="#">NP_001304976.1</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-4 [Pan troglodytes]</a>	<a href="#">Pan troglodytes</a>	739	1274	62%	0.0	49.93%	772	<a href="#">XP_016813513.2</a>
<input type="checkbox"/>	<a href="#">laminin subunit beta-4 isoform 3 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	736	1271	62%	0.0	49.79%	772	<a href="#">NP_001304977.1</a>
<input checked="" type="checkbox"/>	<a href="#">laminin subunit beta-3 precursor [Rattus norvegicus]</a>	<a href="#">Rattus norvegicus</a>	416	1133	87%	3e-121	43.54%	1172	<a href="#">NP_001094311.1</a>
<input checked="" type="checkbox"/>	<a href="#">laminin subunit beta-3 [Pan troglodytes]</a>	<a href="#">Pan troglodytes</a>	410	1046	87%	2e-119	41.56%	1171	<a href="#">XP_003949737.2</a>
<input checked="" type="checkbox"/>	<a href="#">laminin subunit beta-3 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	408	1091	87%	1e-118	41.19%	1172	<a href="#">NP_000219.2</a>
<input checked="" type="checkbox"/>	<a href="#">laminin subunit beta-3 [Tursiops truncatus]</a>	<a href="#">Tursiops truncatus</a>	400	927	85%	5e-116	42.25%	1172	<a href="#">XP_033708792.1</a>

## Esercitazione 2: allineamento multiplo di sequenze proteiche

1. Data una sequenza proteica query di *Homo sapiens* identificare a quale proteina appartiene ed evidenziare la sua eventuale appartenenza ad una famiglia di proteine
2. Allineare la proteina con alcuni altri membri della famiglia (proteine appartenenti anche ad altri organismi)
3. Ottenere ed analizzare il cladogramma e l'albero filogenetico risultante dall'allineamento multiplo

# Muscle

- Fa allineamenti multipli di sequenze biologiche
- Altamente ottimizzato sia in termini di tempo che di precisione dei risultati (può essere usato per allineare fino ad oltre 1000 sequenze)
- Basato su allineamento progressivo seguito da raffinamenti iterativi

1 - Collegarsi all'home page dell'EBI dedicata agli allineamenti multipli (<http://www.ebi.ac.uk/Tools/msa/>) e scegliere **Muscle**.

2 - Scegliere il file fasta con le 8 sequenze ed avviare l'analisi con i parametri di default:

[Input form](#) [Web services](#) [Help & Documentation](#) [Bioinformatics Tools FAQ](#) [Feedback](#)

Tools > Multiple Sequence Alignment > MUSCLE

## Multiple Sequence Alignment

MUSCLE stands for **M**ultiple **S**equences **C**omparison by **L**og-**E**xpectation. MUSCLE is claimed to achieve both better average accuracy and better speed than [ClustalW2](#) or T-Coffee, depending on the chosen options.

**Important note:** This tool can align up to 500 sequences or a maximum file size of 1 MB.

STEP 1 - Enter your input sequences

Enter or paste a set of sequences in any supported format:

Or upload a file:  sequence (4).fasta [Use a example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

STEP 2 - Set your Parameters

OUTPUT FORMAT:

ClustalW

The default settings will fulfill the needs of most users.

(Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

☐ Be notified by email (Tick this box if you want to be notified by email when the results are available)



3 - Cliccare su “Show colors” per visualizzare un colore per ciascun gruppo di amminoacidi (polari, non polari, acidi):

Rn\_b1X1  
Tt\_b1  
Hs\_b1p  
Pt\_b1  
Rn\_b3p  
Tt\_b3  
Pt\_b3  
Hs\_b3p

```

MERPLSSLPTFAFSPLYLSRKEDTKKRAACLPRPPSRAASPRLQGLDMGVLQVFAFGVL
-----MGRLQMFAFSFL
-----MGLLQLLAFSFL
-----MWLLQLLAFSFL
-----MMTALFLLWLALP
-----MRPLLLLYFVL
-----MRPFLLCFALP
-----MRPFLLCFALP
*  :  :  :  :
  
```

“\*” = stesso amminoacido  
per tutte le sequenze

Rn\_b1X1  
Tt\_b1  
Hs\_b1p  
Pt\_b1  
Rn\_b3p  
Tt\_b3  
Pt\_b3  
Hs\_b3p

```

ALWGTRVCAQEPEFSYGCAEGSCYPATGDLIGRAQKLSVTSTCGLHKPEPYCIVSHLQE
ALCGAQVSAQEPEFSYGCAEGSCYPATGDLIGRAQKLSVTSTCGLYKPEPYCIVSHLQE
ALCRARVRAQEPEFSYGCAEGSCYPATGDLIGRAQKLSVTSTCGLHKPEPYCIVSHLQE
ALCRARVRAQEPEFSYGCAEGSCYPATGDLIGRAQKLSVTSTCGLHKPEPYCIVSHLQE
GL----LCAQQP----CSRGACYPVGDLLIGRTQLLRASSTCGLTKPETYC-TQYGQW
SV----LCAQQA----CSRGACYPVGDLLIGRTFLRASSTCGLAKPETYC-TQYGEW
GL----LHAQQA----CSRGACYPVGDLLVGRTRFLRASSTCGLTKPETYC-TQYGEW
GL----LHAQQA----CSRGACYPVGDLLVGRTRFLRASSTCGLTKPETYC-TQYGEW
.:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
  
```

“:” = sostituzione  
conservativa

Rn\_b1X1  
Tt\_b1  
Hs\_b1p  
Pt\_b1  
Rn\_b3p  
Tt\_b3  
Pt\_b3  
Hs\_b3p

```

DKKCFICDSRDPYHETLNPDSHLIENVVTFAPNRLKIWWQSENGVENVTIQLDLAEAFH
DKKCFICNSQDPYHETLNPDSHLIENVVTFAPNRLKIWWQSENGVENVTIQLDLAEAFH
DKKCFICNSQDPYHETLNPDSHLIENVVTFAPNRLKIWWQSENGVENVTIQLDLAEAFH
DKKCFICNSQDPYHETLNPDSHLIENVVTFAPNRLKIWWQSENGVENVTIQLDLAEAFH
QMKCKCKDSRLPH---NYNSHRVENVSSSGPMR---WWQSQNDVSPVSLQLDLDRKFQ
QMKCKCKDSRLPH---NYNSHRVENVSSSGPMR---WWQSQNDVSPVSLQLDLDRKFQ
QMKCKCKDSRQPH---NYNSHRVENVSSSGPMR---WWQSQNDVNPVSLQLDLDRRFQ
QMKCKCKDSRQPH---NYNSHRVENVSSSGPMR---WWQSQNDVNPVSLQLDLDRRFQ
.:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
  
```

“.” = sostituzione  
semi-conservativa

“ ” = sostituzione non  
conservativa



4 - Cliccare su “Phylogenetic Tree” per visualizzare l’albero filogenetico:

Alignments Result Summary **Phylogenetic Tree** Results Viewers Submission Details

Download Phylogenetic Tree Data

## Phylogenetic Tree

*This is a Neighbour-joining tree without distance corrections.*

Branch length: ☒ Cladogram ☐ Real



4 - Cliccare su “Phylogenetic Tree” per visualizzare l’albero filogenetico:

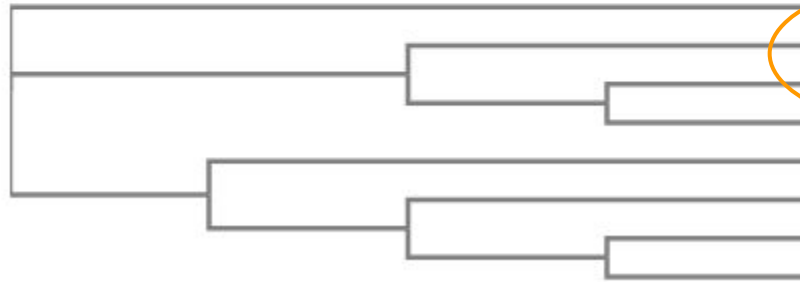
Alignments Result Summary **Phylogenetic Tree** Results Viewers Submission Details

Download Phylogenetic Tree Data

## Phylogenetic Tree

*This is a Neighbour-joining tree without distance corrections.*

Branch length: ☒ Cladogram ☐ Real



Rn\_b1X1 0.03705  
Tt\_b1 0.03651  
Hs\_b1p 0.0014  
Pt\_b1 0.00252

*laminin subunit beta-1*

Rn\_b3p 0.08657  
Tt\_b3 0.07491  
Pt\_b3 0.00732  
Hs\_b3p 0.00805

*laminin subunit beta-3*

4 - Cliccare su “Phylogenetic Tree” per visualizzare l’albero filogenetico:

Alignments Result Summary **Phylogenetic Tree** Results Viewers Submission Details

Download Phylogenetic Tree Data

## Phylogenetic Tree

*This is a Neighbour-joining tree without distance corrections.*

Branch length: ☐ Cladogram ☒ Real

