
Master Biomedizin 2017

- 1) UCSC & UniProt
- 2) Homology
- 3) MSA
- 4) Phylogeny

1) UCSC & UniProt

1

- a. Get the fasta sequence of the human (*Homo sapiens*) protein P53 from UniProt (<http://www.uniprot.org/>). Which one of all the isoforms should you download?
- b. Find the P53 protein from mouse (*Mus musculus*). As you see, there is more than one entry for mouse. Which UniProt entry should you select?
- c. BLAT the human P53 using “hg38” as database (in UCSC, <http://genome.ucsc.edu/cgi-bin/hgBlat>), and answer:
 - How many amino acids has the query sequence?
 - And how many nucleotides?
 - Is it a perfect alignment?
 - Which is the genomic locus of the target?
- d. Visualize and navigate through the P53 genome region, and answer:
 - Which genes are around?
 - How many exons does it have?
- e. BLAT the mouse P53 against the human genome “hg38”. What do you observe?



Human
(*Homo sapiens*)



Mouse
(*Mus musculus*)

1

UniProtKB p53 human

Advanced Search

Help Contact

About UniProtKB Basket

BLAST Align Download Add to basket Columns

1 to 25 of 5,151 Show 25

Download selected (1)
☐ Download all (5151)
 Format: FASTA (canonical)
☐ Compressed ☒ Uncompressed
 Preview first 10ⁱ Go

Entry	Gene names	Organism	Length
P04637	TP53 P53	Homo sapiens (Human)	393
P02340	Tp53 P53, Trp53	Mus musculus (Mouse)	387
Q00987	MDM2	Homo sapiens (Human)	491

a. P04637 (P53_HUMAN). The canonical.

UniProtKB p53 mouse

Advanced Search

Help Contact

About UniProtKB Basket

BLAST Align Download Add to basket Columns

1 to 25 of 839 Show 25

Download selected (1)
☐ Download all (839)
 Format: FASTA (canonical)
☐ Compressed ☒ Uncompressed
 Preview first 10ⁱ Go

Entry	Gene names	Organism	Length
P02340	Tp53 P53, Trp53	Mus musculus (Mouse)	387
P23804	Mdm2	Mus musculus (Mouse)	489

b. P02340 (P53_MOUSE).

1

BLAT Search Genome

Genome: Assembly: Query type: Sort output: Output type:

>sp|P04637|P53_HUMAN Cellular tumor antigen p53 OS=Homo sapiens GN=TP53 PE=1 SV=4
 MEEPQSDPSVEPPLSQETFSDLWKLLENVLSFLPSQAMDDLMLSPDDIEQWFTEDPGP
 DEAPRMPEAAPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRGLHSGTAK
 SVTCTYSPALNKMFCQAKTQVQLWVDSSTPPGTRVRAMAIYKQSQHMTVEVRRCPHHE
 RCSDSDGLAPQHLIRVEGNLRVEYLDNRNTRFHSVVPYEPPEVGSQDCTTIHYNMNCNS
 SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRRRTEENLRKKGEPPHHELP
 PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
 GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	P53_HUMAN	1149	1	393	393	100.0%	17	+-	7669612	7676594	6983

c. 393 amino acids
393*3 = 1179 nucleotides

Not a perfect alignment
("lpennvl")

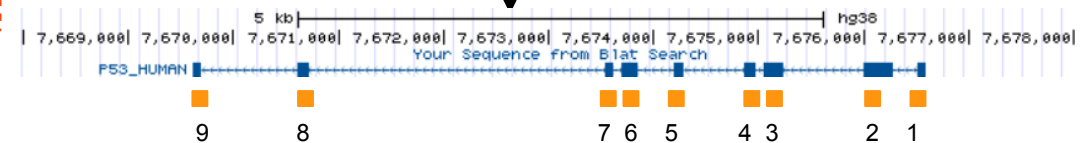
chr17
7669612-7676594

P53_HUMAN

```

MEEPQSDPSV EPPLSQETFS DLWKLlpenn vLSPLSQAM DDLMLSPDDI EQWFTEDPGP 60
DEAPRMPEAA PPVAPAPAAP TPAAPAPAPS WPLSSSVPSQ KTYQGSYGFR LGFLHSGTAK 120
SVTCTYSPAL NKMFCQAKT CPVQLWVDST PPPGTRVRAM AIYKQSQHMT EVVRRCPHHE 180
RCSDSDGLAP QHLIRVEGN LRVEYLDNRN TFRHSVVPVY EPPEVGSQDCT TIHYNMNCNS 240
SCMGGMNRRP ILTIITLED5 SGNLLGRNSF EVRVCACPGR DRRTEENLR KKGEPHHELP 300
PGSTKRALPN NTSSSPQPKK KPLDGEYFTL QIRGRERFEM FRELNEALEL KDAQAGKEPG 360
GSRAHSSHLK SKKGQSTSRH KKLMLFKTEG DSD
  
```

d. ATP1B2 and WRAP53. 9 exons (9 blocks).



BLAT Search Genome

Genome: Assembly: Query type: Sort output: Output type:

>sp|P02340|P53_MOUSE Cellular tumor antigen p53 OS=Mus musculus GN=TP53 PE=1 SV=3
 MEESQSDISLEPLSQETFSGLWKLPPEDILPSPHCMDDLLLPQDVEEFFEGPSEALRV
 SGAPAAQDPVTETPGVAPAPATPWPLSSFVPSQKTYQGSYGFRGLHSGTAKSVMTCTY
 SPPLNKLFCQAKTQVQLWVSATPPAGSRVRAMAIYKKSQHMTEVVRRCPPHHERCSDGD
 GLAPPQHLIRVEGNLYPEYLEDROTFRHSVVPYEPPEAGSEYTTIHYKYMNCSSCMGM
 NRRPILTIITLEDSSGNLLGRDSFEVRVCACPGRRRTEENFRKKEVLCPCLPPGSAKR
 ALPTCTASAPPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAHATEESGDSRAHS
 SYLTKTKGQSTSRHKKTMVKVGPDS

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	P53_MOUSE	548	74	360	387	85.0%	17	+-	7670611	7676131	5521

e. The result is worse (85%).

P53_MOUSE

```

meesqsdisl elplsgetfs glwklpped ilpsphcmdd lllpqdveef fegpsealrv 60
sgapaaqdpv tetPpVAPA PATpWPLSf VPSQKTYQGS YGfHGLfLqS GTAKSVmCTY 120
SPpLNKLFCQ LAKTQVQLW VsaTPPaGSR VRAMAIYKKS QHMTVEVRRRC PHHERCSDgD 180
GLAPPQHLIR VEGNlypEYL eDRqTFRHSV VVPYEPPEaG SeyTTIHyKY MCNSSCMGM 240
NRRPILTIIT LEDSSGNLLG RdSFEVRVCAC PGRRRTEEE ENfRKKEvlc pELPPGSAkR 300
alptctasap pqkkkpldge yftLkIRGrK RFEMFRELNE ALELKDAHAt eESgdSRAHS 360
syltktkgqs tsrhkktmvk kvgpdsd
  
```

2

- a. How many “Apoptosis inhibitor 5” (api5) proteins are there in human (*Homo sapiens*)? Use UniProt.
- b. And how many UniProt entries?

2

UniProtKB name:"apoptosis inhibitor 5" organism:human

UniProtKB results

Filter by ⁱ

- Reviewed (1) Swiss-Prot
- Unreviewed (4) TrEMBL

Popular organisms

Human (5)

Search terms

Filter "human" as: organism

Entry	Entry name	Protein names	Gene names	Organism	Length
<input type="checkbox"/> Q9BZZ5	API5_HUMAN	Apoptosis inhibitor 5	API5 MIG8	Homo sapiens (Human)	524
<input type="checkbox"/> G3V1C3	G3V1C3_HUMAN	Apoptosis inhibitor 5	API5	Homo sapiens (Human)	510
<input type="checkbox"/> E9PQK6	E9PQK6_HUMAN	Apoptosis inhibitor 5	API5	Homo sapiens (Human)	123
<input type="checkbox"/> H0YER7	H0YER7_HUMAN	Apoptosis inhibitor 5	API5	Homo sapiens (Human)	294
<input type="checkbox"/> B4DDR3	B4DDR3_HUMAN	cDNA FLJ52148, highly similar to Ap...		Homo sapiens (Human)	331

a. One protein (SwissProt): Q9BZZ5.

b. At least four UniProt entries.

Q9BZZ5	API5_HUMAN	1		524
G3V1C3	G3V1C3_HUMAN	1		510
H0YER7	H0YER7_HUMAN	1		294
E9PQK6	E9PQK6_HUMAN	1		123