

---

# Master Biomedizin 2018

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

## 1

- a. Get the fasta sequence of the human (*Homo sapiens*) protein P53 from UniProt (<https://www.uniprot.org/>). Which one of all the isoforms should you download? **P04637**
- 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? **P02340**
- 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? **393 aa**  
And how many nucleotides? **1179 nt**  
Is it a perfect alignment? **No**  
Which is the genomic locus of the target? **Chr17 7669612-7676594**
- d. Visualize and navigate through the P53 genome region, and answer:  
Which genes are around? **ATP1B2 and WRAP53**  
How many exons does it have? **9 exons**
- e. BLAT the mouse P53 against the human genome “hg38”. What do you observe? **The result is worse (85%)**

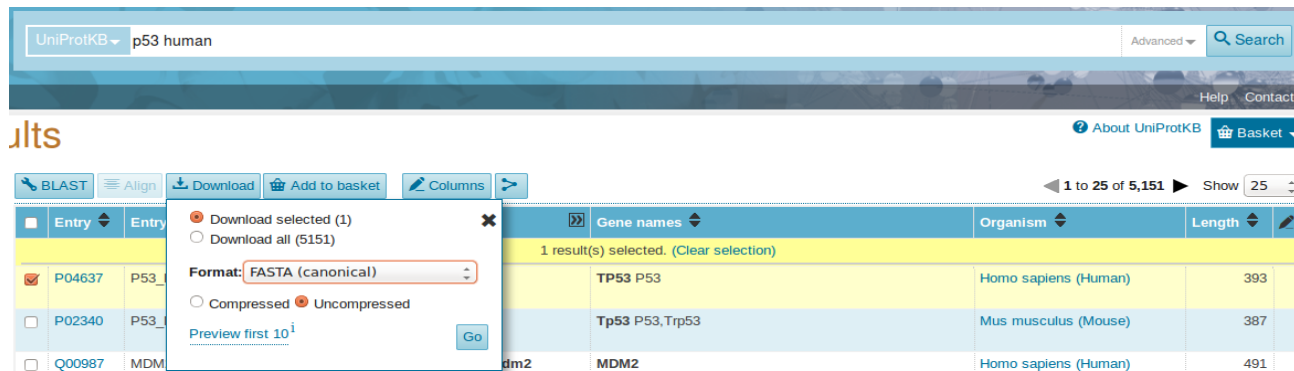


Human  
(*Homo sapiens*)



Mouse  
(*Mus musculus*)

1



UniProtKB p53 human

Advanced Search

Help Contact

About UniProtKB Basket

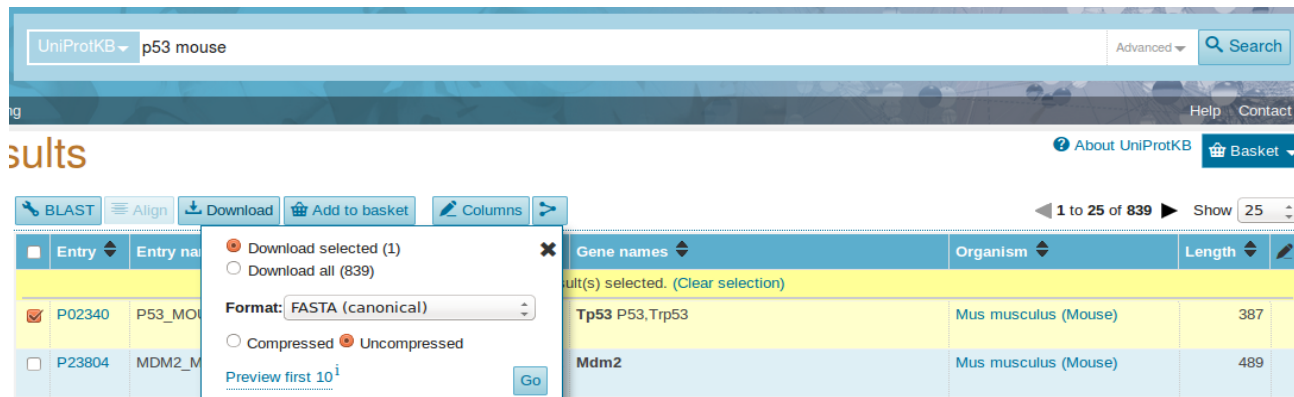
BLAST Align Download Add to basket Columns

1 to 25 of 5,151 Show 25

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

Entry	Gene names	Organism	Length
TP53 P53	Homo sapiens (Human)	393	
Tp53 P53, Trp53	Mus musculus (Mouse)	387	
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 10  
 Go

Entry	Gene names	Organism	Length
Tp53 P53, Trp53	Mus musculus (Mouse)	387	
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  
 SVTCTYSPALNKMFCQLAKTCVQLWVDSTPPGTRVRAMAIYKQSQHMTVEVRRCPHHE  
 RCDSDGLAPQHLIRVEGNLRVEYLDNRNTRFHSVVPYEPPEVGSQDCTTIHYNMYCNS  
 SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRRRTEENLRKKGEPPHHEP  
 PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG  
 GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
<a href="#">browser</a> <a href="#">details</a>	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")

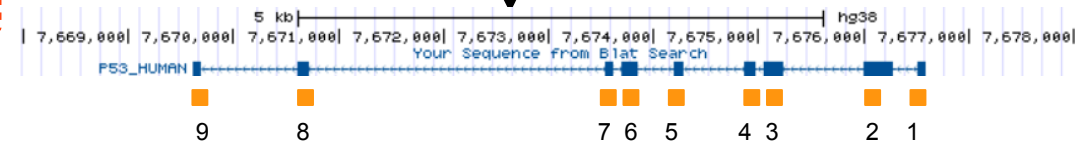
chr17  
7669612-7676594

### P53\_HUMAN

```

MEEPQSDPSV EPPLSQETFS DLWKLlpenn vLSPLSQAM DDLMLSPDDI EQWFTEDPGP 60
DEAPRMPEAA PPVAPAPAAP TPAAPAPAPS WPLSSSVPSQ KTYQGSYGFR LGFLHSGTAK 120
SVTCTYSPAL NKMFCQLAKT CPVQLWVDST PPPGTRVRAM AIYKQSQHMT EVVRRCPHHE 180
RCDSDGLAPL PQHLIRVEGN LRVEYLDNRN TFRHSVVPVY EPPEVGSQDCT TIHYNMYCNS 240
SCMGGMNRRP ILTIITLED5 SGNLLGRNSF EVRVCACPGR DRRTEENLR KKGEPHHELP 300
PGSTKRALPN NTSSSPQPKK KPLDGEYFTL QIRGRERFEM FRELNEALEL KDAQAGKEPG 360
GSRAHSSHLK SKKGQSTSRH KKLMEFKTEG 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  
 SGAPAAQDPVTETPGVAPAPATPWPLSSFVPSQKTYQGNYGFLGFLGTAQSVMTCTY  
 SPPLNKLFCQLAKTCVQLWVSATPPAGSRVRAMAIYKKSQHMTEVVRRCPPHHERCSDGD  
 GLAPPQHLIRVEGNLYPEYLEDROTFRHSVVPYEPPEAGSEYTTIHYKYMNSSCMGGM  
 NRRPILTIITLEDSSGNLLGRDSFEVRVCACPGRRRTEENFRKKEVLCPPELPPGSAKR  
 ALPTCTASAPPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAHATEESGDSRAHS  
 SYLTKTKGQSTSRHKKTMVKVGPDS

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
<a href="#">browser</a> <a href="#">details</a>	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 VPSQKTYQGN YGFhLGFLqS GTAKSVmCTY 120
SPpLNKLFCQ LAKTCVQLW VsaTPPaGsR VRAMAIYKks QHMTVEVRRRC PHHERCSDgD 180
GLAPPQHLIR VEGNlypEYL eDRqTFRHSV VVPYEPPEaG SeyTTIHykY MCNSSCMGGM 240
NRRPILTIIT LEDSSGNLLG RdSFEVRVCAC PGRRDRTEE ENFRKKEvlc pELPPGSAkR 300
alptctsasp 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. **One protein (SwissProt): Q9BZZ5**
- b. And how many UniProt entries? **At least four UniProt entries**

2

The screenshot shows the UniProtKB search results for the query "name:'apoptosis inhibitor 5' organism:human". The results table lists five entries:

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

On the left, the 'Filter by' section shows 1 Reviewed (Swiss-Prot) and 4 Unreviewed (TrEMBL) entries. The 'Popular organisms' section lists Human (5). The 'Search terms' section shows the filter "human" as: organism.

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