Master Biomedizin 2017

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

Pablo Mier
munoz@uni-mainz.de

1) UCSC & UniProt

Pablo Mier
munoz@uni-mainz.de

*Images from: Wikipedia

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*)

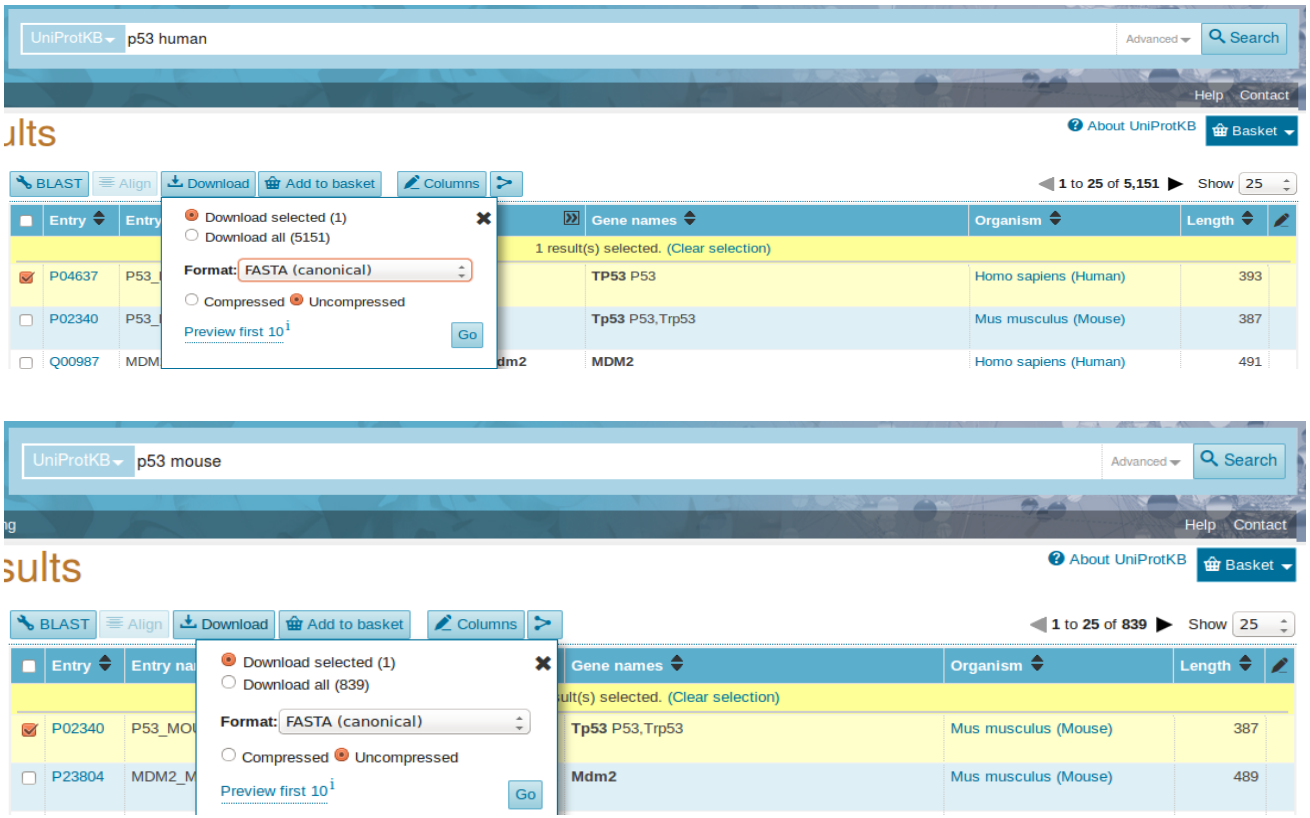
Mouse
(*Mus musculus*)

Pablo Mier
munoz@uni-mainz.de

Genomics

*Images from: UniProt

1



a. P04637 (P53_HUMAN). The canonical.

b. P02340 (P53_MOUSE).

Genomics

*Images from: UCSC

1

BLAT Search Genome

Genome: Human | Assembly: Dec. 2013 (GRCh38/hg38) | Query type: BLAT's guess | Sort output: query,score | Output type: hyperlink

```
>sp|P04637|P53_HUMAN Cellular tumor antigen p53 OS=Homo sapiens GN=TP53 PE=1 SV=4
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP
PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD
```

submit | I'm feeling lucky | clear

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

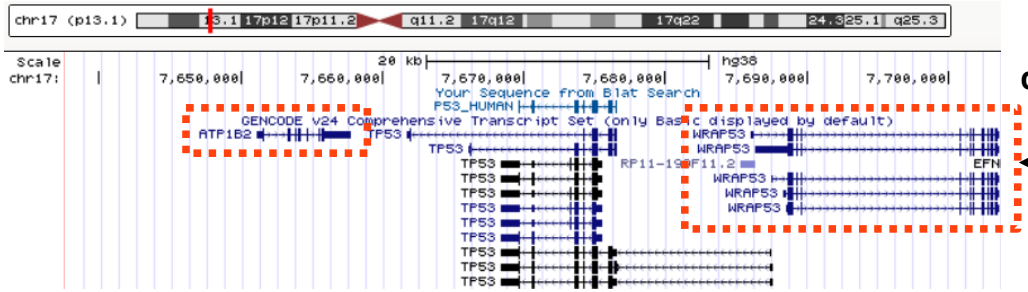
c. 393 amino acids
393*3 = 1179 nucleotides

Not a perfect alignment
("*lpennvl*")

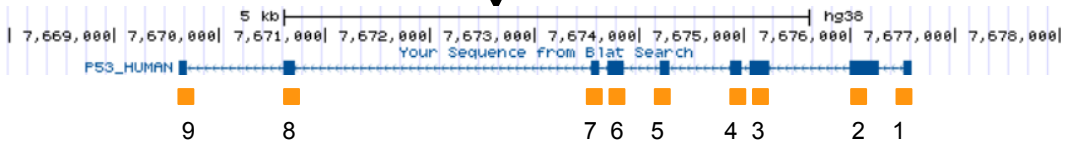
chr17
7669612-7676594

P53_HUMAN

```
MEEPQSDPSV EPPLSQETFS DLWKLlpenn vlSPLPSQAM DDLMLSPDDI EQWFTEDPGP  60
DEAPRMPEAA PPVAPAPAAP TPAAPAPAPS WPLSSSVPSQ KTYQGSYGFR LGFLHSGTAK  120
SVTCTYSPAL NKMFCQLAKT CPVQLWVDST PPPGTRVRAM AIYKQSQHMT EVVRRCPHHE  180
RCSDSDGLAP PQHLIRVEGN LRVEYLDDRN TFRHSVVVPY EPPEVGSDCT TIHYNYMCNS  240
SCMGGMNRRP ILTIITLEDS SGNLLGRNSF EVRVCACPGR DRRTEEENLR KKGEPHHELP  300
PGSTKRALPN NTSSSPQPKK KPLDGEYFTL QIRGRERFEM FRELNEALEL KDAQAGKEPG  360
GSRAHSSHLK SKKGQSTSRH KKLMFKTEGP DSD
```



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



BLAT Search Genome

Genome: Human | Assembly: Dec. 2013 (GRCh38/hg38) | Query type: BLAT's guess | Sort output: query,score | Output type: hyperlink

```
>sp|P02340|P53_MOUSE Cellular tumor antigen p53 OS=Mus musculus GN=Tp53 PE=1 SV=3
MEESQSDISLELPLSQETFSGLWKLLPPEDILPSPHCMDDLLLPQDVEEFFEGPSEALRV
SGAPAAQDPVTETPGPVAPAPATPWPLSSFVPSQKTYQGNYGFHLGFLQSGTAKSVMCTY
SPPLNKLFCQLAKTCPVQLWVSATPPAGSRVRAMAIYKKSQHMTEVVRRCPHHERCSDGD
GLAPPQHLIRVEGNLYPEYLEDRQTFRHSVVVPYEPPEAGSEYTTIHYKYMCNSSCMGGM
NRRPILTIITLEDSSGNLLGRDSFEVRVCACPGRDRRTEEENFRKKEVLCPELPPGSAKR
ALPTCTSASPPQKKKPLDGEYFTLKIRGRKRFEMFRELNEALELKDAHATEESGDSRAHS
SYLKTKKGQSTSRHKKTMVKKVGPDSD
```

submit | I'm feeling lucky | clear

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 elplsqetfs glwkllpped ilpsphcmdd lllpqdveef fegpsealrv  60
sgapaaqdpv tetPgPvAPA PAtpWPLSSf VPSQKTYQGn YGFhLGFLqS GTAKSVmCTY  120
SPpLNKlFCQ LAKTCPVQLW VsaTPPaGsR VRAMAIYKkS QHMTEVVRRC PHHERCSDgD  180
GLAPPQHLIR VEGNLypEYL eDRqTFRHSV VVPYEPPEaG SeyTTIHYkY MCNSSCMGGM  240
NRRPILTIIT LEDSSGNLLG RdSFEVRVCA CPGRDRRTEE ENfRKKevlc pELPPGSaKR  300
alptctsasp pqkkkpldge yftLkIRGRk RFEMFRELNE ALELKDAhAt eEsGdSRAHS  360
sylktkkgqs tsrhkktmvk kvgpdsd
```

UniProt database

2

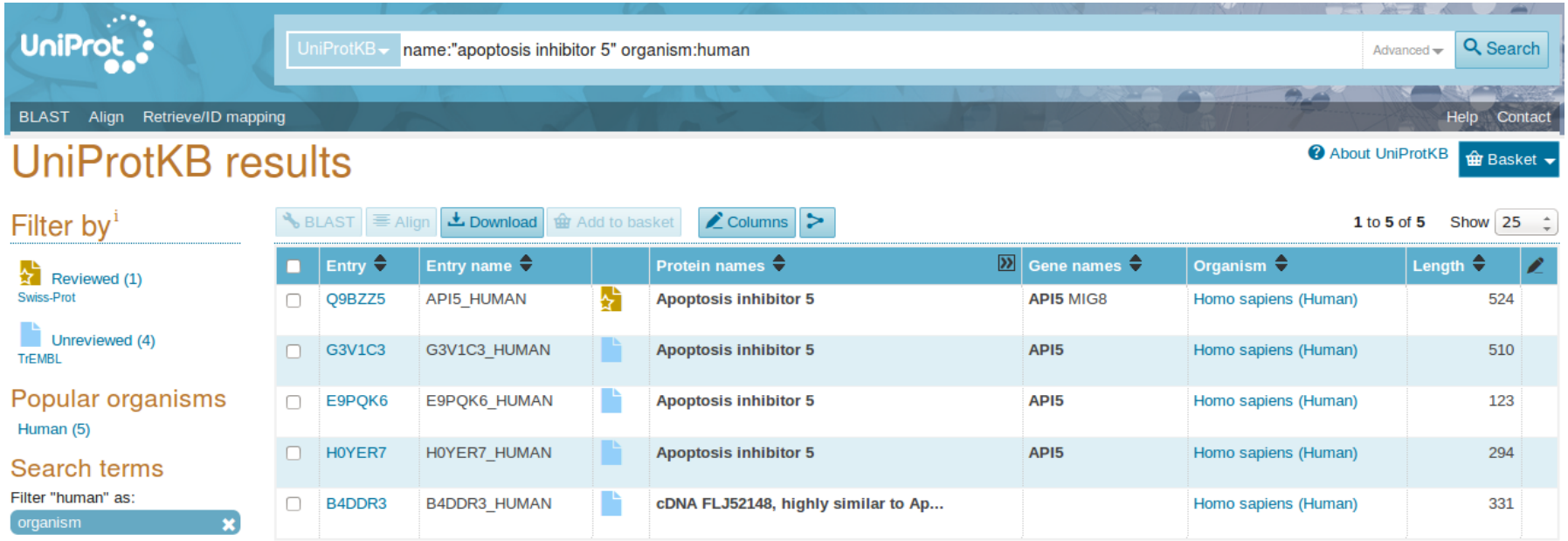
a. How many “Apoptosis inhibitor 5” (api5) proteins are there in human (*Homo sapiens*)? Use UniProt.

b. And how many UniProt entries?

UniProt database

*Images from: UniProt

2



a. One protein (SwissProt): Q9BZZ5.

b. At least four UniProt entries.