Find a Gene Project
BGGN 213
Kalodiah Toma
PID: A07606689
Email: k1toma@ucsd.edu

Questions:

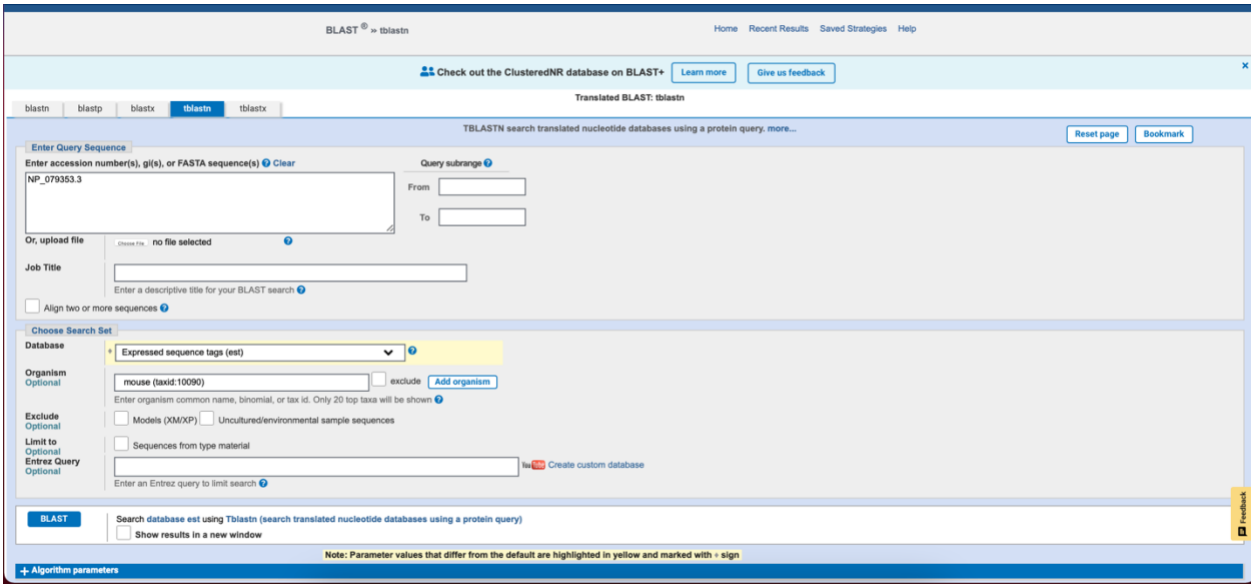
[Q1] Tell me the name of a protein you are interested in. Include the species and the accession number. This can be a human protein or a protein from any other species as long as it's function is known.

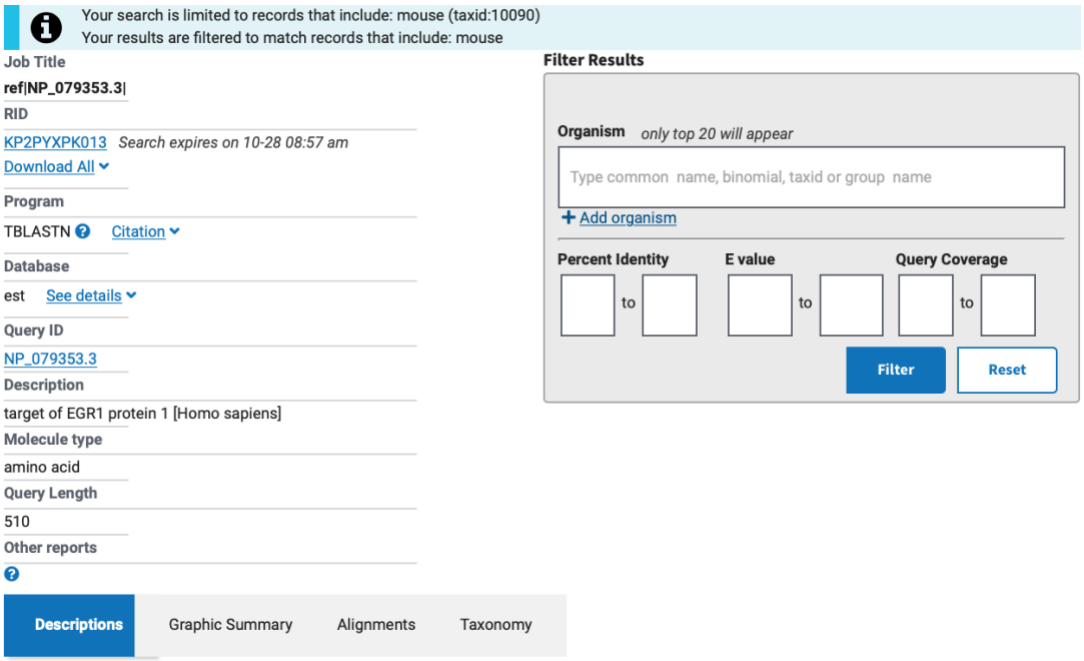
If you do not have a favorite protein, select human RBP4 or KIF11. Do not use beta globin as this is in the worked example report that I provide you with online.

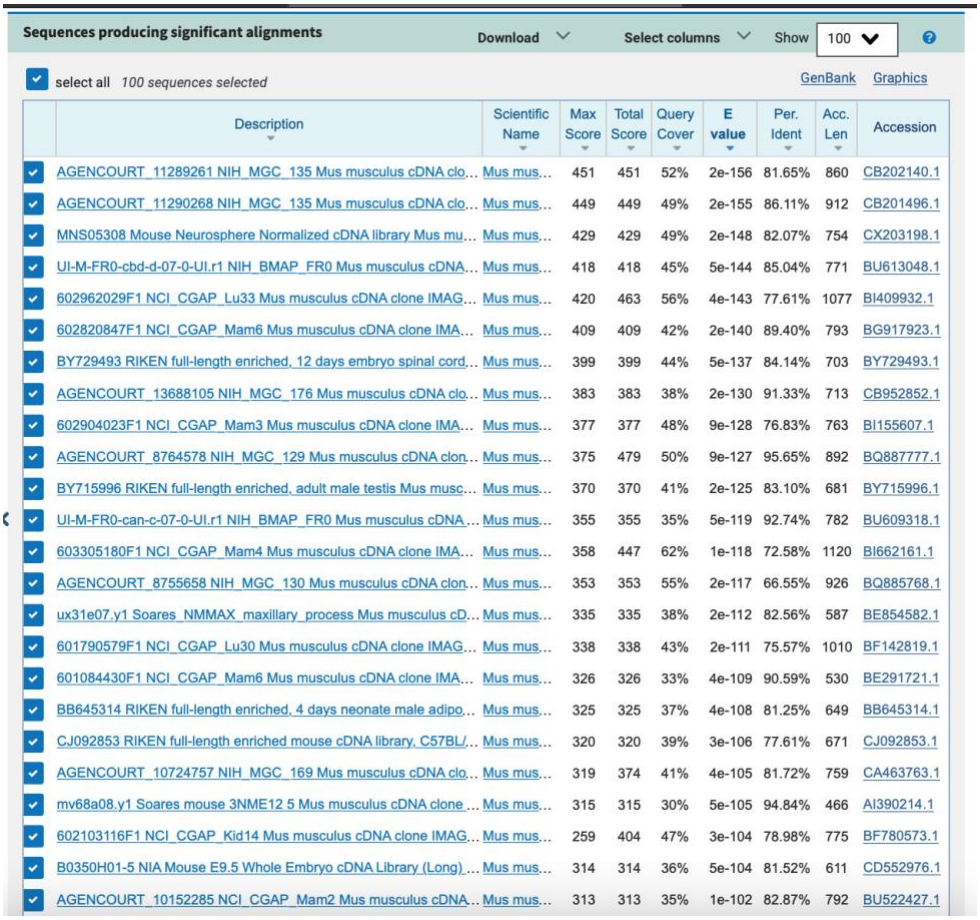
```
Name: TOE1 target of EGR1
Accession: NP_079353.3 (NM_025077.4)
Species: Homo Sapiens
```

[Q2] Perform a BLAST search against a DNA database, such as a database consisting of genomic DNA or ESTs. The BLAST server can be at NCBI or elsewhere. Include details of the BLAST method used, database searched and any limits applied (e.g. Organism).

```
Method: TBLASTN (2.7.1) search against mouse
Database: ESTs Expressed Sequence Tags (est)
Organism: Mouse (Taxid: 10090)
Chosen match: Accession CB202140.1, a 860 base pair clone from Mus musculus.
See below for alignment details.
```



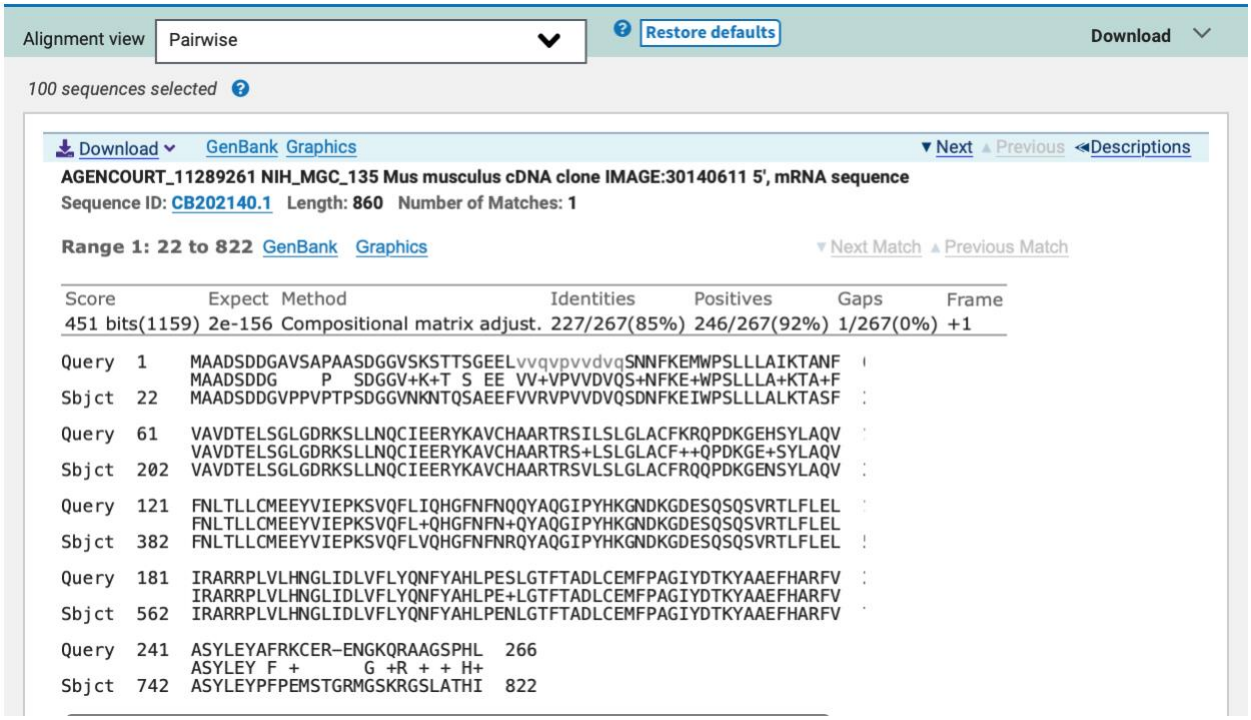




Sequences producing significant alignments — Download — Select columns — Show 100

select all *100 sequences selected* — GenBank — Graphics

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
AGENCOURT_11289261 NIH_MGC_135 Mus musculus cDNA clo...	Mus mus...	451	451	52%	2e-156	81.65%	860	CB202140.1
AGENCOURT_11290268 NIH_MGC_135 Mus musculus cDNA clo...	Mus mus...	449	449	49%	2e-155	86.11%	912	CB201496.1
MNS05308 Mouse Neurosphere Normalized cDNA library Mus mu...	Mus mus...	429	429	49%	2e-148	82.07%	754	CX203198.1
UI-M-FR0-cbd-d-07-0-UI.r1 NIH_BMAP_FR0 Mus musculus cDNA...	Mus mus...	418	418	45%	5e-144	85.04%	771	BU613048.1
602962029F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAG...	Mus mus...	420	463	56%	4e-143	77.61%	1077	BI409932.1
602820847F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMA...	Mus mus...	409	409	42%	2e-140	89.40%	793	BG917923.1
BY729493 RIKEN full-length enriched, 12 days embryo spinal cord...	Mus mus...	399	399	44%	5e-137	84.14%	703	BY729493.1
AGENCOURT_13688105 NIH_MGC_176 Mus musculus cDNA clo...	Mus mus...	383	383	38%	2e-130	91.33%	713	CB952852.1
602904023F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMA...	Mus mus...	377	377	48%	9e-128	76.83%	763	BI155607.1
AGENCOURT_8764578 NIH_MGC_129 Mus musculus cDNA clon...	Mus mus...	375	479	50%	9e-127	95.65%	892	BQ887777.1
BY715996 RIKEN full-length enriched, adult male testis Mus musc...	Mus mus...	370	370	41%	2e-125	83.10%	681	BY715996.1
UI-M-FR0-can-c-07-0-UI.r1 NIH_BMAP_FR0 Mus musculus cDNA ...	Mus mus...	355	355	35%	5e-119	92.74%	782	BU609318.1
603305180F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMA...	Mus mus...	358	447	62%	1e-118	72.58%	1120	BI662161.1
AGENCOURT_8755658 NIH_MGC_130 Mus musculus cDNA clon...	Mus mus...	353	353	55%	2e-117	66.55%	926	BQ885768.1
ux31e07.y1 Soares_NMMAX_maxillary_process Mus musculus cD...	Mus mus...	335	335	38%	2e-112	82.56%	587	BE854582.1
601790579F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAG...	Mus mus...	338	338	43%	2e-111	75.57%	1010	BF142819.1
601084430F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMA...	Mus mus...	326	326	33%	4e-109	90.59%	530	BE291721.1
BB645314 RIKEN full-length enriched, 4 days neonate male adipo...	Mus mus...	325	325	37%	4e-108	81.25%	649	BB645314.1
CJ092853 RIKEN full-length enriched mouse cDNA library, C57BL/...	Mus mus...	320	320	39%	3e-106	77.61%	671	CJ092853.1
AGENCOURT_10724757 NIH_MGC_169 Mus musculus cDNA clo...	Mus mus...	319	374	41%	4e-105	81.72%	759	CA463763.1
mv68a08.y1 Soares mouse 3NME12 5 Mus musculus cDNA clone ...	Mus mus...	315	315	30%	5e-105	94.84%	466	AI390214.1
602103116F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAG...	Mus mus...	259	404	47%	3e-104	78.98%	775	BF780573.1
B0350H01-5 NIA Mouse E9.5 Whole Embryo cDNA Library (Long) ...	Mus mus...	314	314	36%	5e-104	81.52%	611	CD552976.1
AGENCOURT_10152285 NCI_CGAP_Mam2 Mus musculus cDNA...	Mus mus...	313	313	35%	1e-102	82.87%	792	BU522427.1



[Q3] Gather information about this “novel” **protein**. At a minimum, show me the protein sequence of the “novel” protein as displayed in your BLAST results from [Q2] as FASTA format (you can copy and paste the aligned sequence subject lines from your BLAST

result page if necessary) or translate your novel DNA sequence using a tool called EMBOSS Transeq at the EBI. Don't forget to translate all six reading frames; the ORF (open reading frame) is likely to be the longest sequence without a stop codon. It may not start with a methionine if you don't have the complete coding region. Make sure the sequence you provide includes a header/subject line and is in traditional FASTA format.

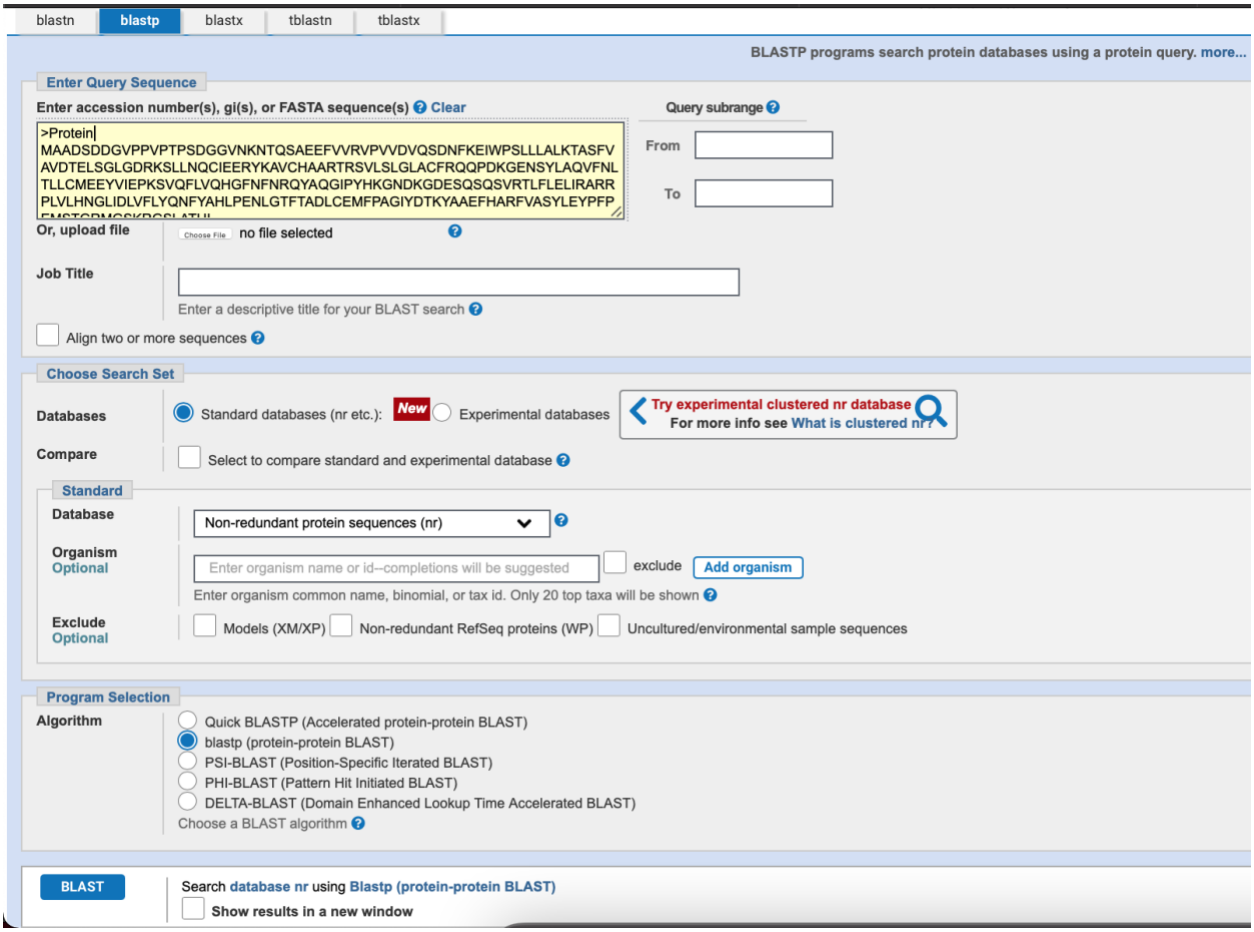
Chosen sequence:
>A. TOE1 like protein (sequence taken from BLAST result)
MAADSDDGVPPVPTPSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYPFPEMSTGRMGSKRGSLATHI
Species: Mus musculus
Eukaryota; Animalia; Chordata; Mammalia; Rodentia; Muridae; Mus; Mus;
M. musculus.

[Q4] Prove that this gene, and its corresponding protein, are novel. For the purposes of this project, "novel" is defined as follows. Take the protein sequence (your answer to [Q3]), and use it as a query in a blastp search of the nr database at NCBI.

- If there is a match with 100% amino acid identity to a protein in the database, from the same species, then your protein is NOT novel (even if the match is to a protein with a name such as "unknown"). Someone has already found and annotated this sequence, and assigned it an accession number.
- If the top match reported has less than 100% identity, then it is likely that your protein is novel, and you have succeeded.
- If there is a match with 100% identity, but to a different species than the one you started with, then you have likely succeeded in finding a novel gene.
- If there are no database matches to the original query from [Q1], this indicates that you have partially succeeded: yes, you may have found a new gene, but no, it is not actually homologous to the original query. You should probably start over.

A BLASTP search against NR database yielded a top hit result is to a protein from Rattus norvegicus (Norway rat).

See additional screen shots below for top hits and selected alignment details:



Details: The top result is to a protein from Rattus norvegicus (Norway rat), see second screen shot below for alignment details:

Sequences producing significant alignments Download Select columns Show 100

select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
target of EGR1 protein 1 isoform X4 [Rattus norvegicus]	Rattus no...	504	504	92%	1e-179	97.18%	257	XP_038965533.1
target of EGR1 protein 1 [Cricetulus griseus]	Cricetulu...	503	503	100%	2e-175	91.01%	511	XP_027258385.1
target of EGR1 protein 1 [Chionomys nivalis]	Chionom...	502	502	92%	2e-174	96.77%	541	XP_057640769.1
target of EGR1 protein 1 [Phodopus roborovskii]	Phodopu...	501	501	100%	3e-174	90.64%	511	XP_051044502.1
target of EGR1 protein 1 [Peromyscus californicus insignis]	Peromys...	501	501	100%	3e-174	90.64%	511	XP_052572255.1
target of EGR1 protein 1 [Mesocricetus auratus]	Mesocric...	500	500	100%	4e-174	90.26%	511	XP_005072112.3
target of EGR1 protein 1 [Peromyscus maniculatus bairdii]	Peromys...	500	500	100%	4e-174	90.26%	511	XP_006986621.1
target of EGR1 protein 1 [Cricetulus griseus]	Cricetulu...	500	500	100%	7e-174	90.26%	511	XP_003497425.1
target of EGR1 protein 1 [Myodes glareolus]	Myodes ...	499	499	92%	2e-173	96.37%	510	XP_048286556.1
target of EGR1 protein 1 [Peromyscus eremicus]	Peromys...	499	499	100%	2e-173	90.26%	511	XP_059110794.1
target of EGR1 protein 1 isoform X1 [Mastomys coucha]	Mastomy...	498	498	92%	7e-173	95.97%	513	XP_031234176.1
target of EGR1 protein 1 isoform X1 [Microtus ochrogaster]	Microtus ...	496	496	92%	2e-172	95.97%	510	XP_026633303.1
target of EGR1 protein 1 isoform X3 [Microtus ochrogaster]	Microtus ...	496	496	92%	4e-172	95.97%	541	XP_005370077.1
target of EGR1, member 1 (nuclear), isoform CRA_c [Mus mu...	Mus mus...	488	488	100%	6e-172	92.51%	329	EDL30586.1
Target of EGR1, member 1 (nuclear) [Mus musculus]	Mus mus...	494	494	100%	2e-171	93.26%	511	AAH23109.1
target of EGR1, member 1 (nuclear) (predicted), isoform CRA...	Rattus no...	485	485	100%	3e-171	91.39%	296	EDL90253.1
target of EGR1 protein 1 [Acomys russatus]	Acomys r...	493	493	100%	3e-171	90.64%	510	XP_051027553.1
target of EGR1, member 1 (nuclear) (predicted), isoform CRA...	Rattus no...	484	484	100%	5e-171	91.39%	293	EDL90254.1
target of EGR1 protein 1 isoform 2 [Mus musculus]	Mus mus...	489	489	100%	5e-171	92.51%	410	NP_001406115.1
unnamed protein product [Mus musculus]	Mus mus...	484	484	100%	1e-170	92.13%	329	BAB30482.1
target of EGR1 protein 1 [Psammomys obesus]	Psammo...	491	491	100%	2e-170	89.51%	511	XP_055474418.1
target of EGR1, member 1 (nuclear) (predicted), isoform CRA...	Rattus no...	484	484	100%	4e-170	91.39%	359	EDL90251.1
hypothetical protein A6R68_22771 [Neotoma lepida]	Neotoma...	489	489	91%	6e-170	95.88%	481	OBS83234.1

Download GenPept Graphics ▼Next ▲Previous ◀Descriptions

target of EGR1 protein 1 isoform X4 [Rattus norvegicus]

Sequence ID: XP_038965533.1 Length: 257 Number of Matches: 1

Range 1: 1 to 248 GenPept Graphics ▼Next Match ▲Previous Match

Score	Expect	Method	Identities	Positives	Gaps
504 bits(1299)	1e-179	Compositional matrix adjust.	241/248(97%)	245/248(98%)	0/248(0%)

```
Query  1    MAADSDDGVPPVPTPSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
            MAADSDDGVP VPTPSDGGVN+ TQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
Sbjct  1    MAADSDDGVPSVPTPSDGGVNRTTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF

Query  61   VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQV
            VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACF+QQPDKG++SYLAQV
Sbjct  61   VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGDSSYLAQV

Query  121  FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
            FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Sbjct  121  FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL

Query  181  IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
            IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Sbjct  181  IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV

Query  241  ASYLEYPF  248
            ASYLEY F
Sbjct  241  ASYLEYAF  248
```

Related Information

Gene - associated gene details

Genome Data Viewer - aligned genomic context

[**Q5**] Generate a multiple sequence alignment with your novel protein, your original query protein, and a group of other members of this family from different species. A typical number of proteins to use in a multiple sequence alignment for this assignment purpose is a minimum of 5 and a maximum of 20 - although the exact number is up to you. Include the multiple sequence alignment in your report. Use Courier font with a size appropriate to fit page width.

Side-note: Indicate your sequence in the alignment by choosing an appropriate name for each sequence in the input unaligned sequence file (i.e. edit the sequence file so that the species, or short common, names (rather than accession numbers) display in the output alignment and in the subsequent answers below). The goal in this step is to create an interesting an alignment for building a phylogenetic tree that illustrates species divergence.

```
>Homo sapiens NP_079353.3 target of EGR1 protein 1 [Homo sapiens]
MAADSDDGAVSAPAASDGGVSKSTTSGEELVVQVPVVDVQSNNFKEMWPSLLLAIKTANFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSILSLGLACFKRQPDKGEHSYLAQVFNLTLLCMEEYVIEPKSVQFLIQHGFNFNQQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPESLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLTLEFCNYPSSMRDHIDYRCCLPPATHRPHPTSICDNFSAYGW
CPLGPQCPQSHDIDLIIDTDEAAAEDKRRRRRRREKRKRALLNLPGTQTSGEAKDGPPKKQVCGDSIKPEETEQEVA
ADETRNLPHSKQGNKNDLEMGIKAARPEIADRATSEVPGSQASPNPVPGDGLHRAGFDAFMTGYVMAYVEVSQGPQP
CSSGPWLPECHNKVYLSGKAVPLTVAKSQFSRSSKAHNQKMKLTWGSS

>Mus musculus  TOE1 like protein (sequence taken from BLAST result)
MAADSDDGVPPVPTPSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYPFPEMSTGRMGSKRGSLATHI

>Rattus norvegicus XP_038965533.1 target of EGR1 protein 1 isoform X4 [Rattus
norvegicus]
MAADSDDGVPSVPTPSDGGVNRTTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGDSSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKWPMAGAL

>Cricetulus griseus XP_027258385.1 target of EGR1 protein 1 [Cricetulus
griseus]
MAADSDDGVPSVPTTSDGGVNKNTKSAEEFVIRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPETLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKQRAAGTPHLALEFCSYPSSMRGHIDYRCCMSPVTYRRSHTTGICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALLSQPETQTFEEAEDGPPTKQVCEDSLKTEEMEQRV
TEGETRDELGSKQAHKSGLEMEHKATSSETVDVATTELPVSQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGPQ
LCSSRPWLPECHNKVYLSGKTVPLTVAKSQFSHSSKAHNQKMKLAWGSS

>Chionomys nivalis XP_057640769.1 target of EGR1 protein 1 [Chionomys
nivalis]
MAADSDDGVPSVPATSDGDVNKNTKSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
```

AAEFHARFVASYLEYAFRKCERENGKRQATGSPHLALEFCSYPSSMRGHIDYRCCMSPVTSRRSHTAGICTKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDRRKRALLSQPGTGTFEEAEDGPPTKQVCEDSLKTETEQKVT
EGESRDQAEDGLPTKQVCEGSLKTEIEQKVTEGESRDQLGSKQCHNSDLEVEHKATSSEIADVAASELPASQASPNP
VPGDGLHRAGFDAFMTGYVMAYVGLRHGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAWG
SS

>Phodopus roborovskii XP_051044502.1 target of EGR1 protein 1 [Phodopus roborovskii]
MAADSDDGVPSVPTTSDGGVNKNTKSAEEFVIRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKHQADKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKRRAAGTPHLVLEFCSYPSSMRGHIDYRCCLSPVTYRRSHTTSICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRRRALLSQLGTQNFEEADDGPPTKQVCEDNLKTQELEQRV
TERETRDELDSKQGHKSDLEMEHKATGSETADVAISELPVSQASPNPMPGDGLHRAGFDAFMTGYVMAYVGLSQGPQ
LCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSHSSKAHNQKMKLAWSSS

>Peromyscus californicus insignis XP_052572255.1 target of EGR1 protein 1 [Peromyscus californicus insignis]
MAADSVDGVPSVPTTSDGGVNKNTKSADEFIVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTYRRFHTSGICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALLSQPGAQTFEEAEDGPPTKQVCEDSLKTEEIEQKV
TEGETRDQLGSQQGHKSGLAVERKATSSETAEVATSELPVSQANPNPGPGDGLHRAGFDAFMTGYVMAYVGLSQGPQ
LCSSGPWLPECHNKVYLSGKTVPLTVARSQFSRSSKAHNQKMKLAWGSS

>Mesocricetus auratus XP_005072112.3 target of EGR1 protein 1 [Mesocricetus auratus]
MAADSDDGVPSVPTTSDGGVNKNTKSAEKFVIRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENCYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPETLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKQRAAGTPHLALEFCSYPSSMRSHIDYRCCISPVTYRRSHTTGICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALLSQPGTQTFEEAEDGPPTKQVCEDNLKTEEIEQRV
TEGEIRDELGSKQAHKSGSEMEHKATSSETVDVATSELPVSQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGPQ
LCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSHSSKAHNQKMKLAWGSS

>Myodes glareolus XP_048286556.1 target of EGR1 protein 1 [Myodes glareolus]
MAADSDDGVPSVPATSDGDVNKTTKSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKRQAAGSPHLALEFCSYPSSMRGHIDYRCCMSPITSRRSHTTGICTKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDRRKRALLNQPGTETFEEAEDGPPTKQVCEDSLKTEIEQKVT
EGETRDQLGSNQDHKSDLEVEHKATSSEIADVAASELPVSQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLRHGPQL
CSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAWGSS

>Mastomys coucha XP_031234176.1 target of EGR1 protein 1 isoform X1 [Mastomys coucha]
MAADSADGMPSVPSSSDGGVNKNTQSAKEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPEKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFEARFVASYLEYAFRKCERENGRQQAVGSPHLALEFCSYPSSMRGHIDYRCCMSPVTCRRSYTTGICDKFSAYG
WCPLGPECPQSHDIDLIIDTDEAVAEDKRRRRWQRRKDKRKRALQSQPGTQTLEEAEGGPPTKQVCEDSLKAEKMEQ
KVAEGDAGDQLGSRQGHTGSLEIAHRRTSAETADVAPSELPVSQASTNPLPGDGLHRAGFDAFMTGYVMAYVGLSQG
LQLCSSEPWLPKCHNKVYLSGKTVPLTVAKSQFSHPSKAHKQKMKLAWGSS

>Microtus ochrogaster XP_026633303.1 target of EGR1 protein 1 isoform X1 [Microtus ochrogaster]

MAADSDDGVPSVSATSDGDVNKNTKPAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKRQAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTSRRSHTTGICTRFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDRRKRALLSQPGTETFEEAEDGPPIKQVCEDSLKTEIEQKVT
EGESRDQLGSKQDQNSDLEVEHKATSSEIADVAASELPVSQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLRHGPQL
CSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAWGSS

>Acomys russatus XP_051027553.1 target of EGR1 protein 1 [Acomys russatus]
MAADSGDGVPLVPKTSDGGVNKSTSAEEFVARVPVVDVQTDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLN
QCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQG
IPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYA
AEFHARFVASYLEYAFRKCERENGKQRAAGGPHLALEFCSYPSSMRGHIDYRCCMSPVTDRRSYATGICDKFSAYGW
CPLGPQCPQSHDIDLIIDTDEAVAEDKRRKRRRKDKRRRALLRQPGTQTFEEGEDGPPTKQVREDSLNTENTEQKVA
EGETSDQLGSKQGHKGGLEMKHEATGSETADVATSELQVNQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGLQL
CSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGNS

>Psammomys obesus XP_055474418.1 target of EGR1 protein 1 [Psammomys obesus]
MAADSDDGGLSVPPTSDGVVNKNTKSEEEFVIRVPVVDVQTDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTYHRSHTGGICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRRRALLSQPGMQTFEEAEEGPPTKQVCEDSLKTENTEQKV
AEGETRDQPGSKQGHKGGLEMEHEAVSSEIADVATSELPVNQTSPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGTQ
LCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Neotoma lepida OBS83234.1 hypothetical protein A6R68_22771 [Neotoma lepida]
MAADSDDGVPSVPTTSDDGVNKNTKSADEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQNQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFFCSYPSSMRDHIDYRCCMSPVTYRRSHTTGICNKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRR
RRRRKDKRRRALLSQPETQTFEEAEDGPPTKQVCEDSLKTEEIEQKVTEGETRDQLGSQQGHKSSLEIEYKATSSKI
ADVATSELPVSQASPNPMPGDGLHRAGFDAFMTGYVMAYVGLSQGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQ
FSRSSKAHNQKMKLAWGSS

>Mus caroli XP_021015031.1 target of EGR1 protein 1 [Mus caroli]
MAADSDDGVPPVPTSSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKNLL
NQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRMLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
ASEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCTSPGTCRRSRPTGICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRSLQSQPGTQALAEAEDGPPTKQVCEDSLKTEKMEQKV
AEGEAGDQPGSREGHTSSLEMAHRRTSAETADVATSELLVNQASTNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGLQ
LCSSEPWLPECHNKVYLSGKTVPLTVTKSQFSRPSKAHNQKMKLAWGSS

>Mus pahari XP_021055456.1 target of EGR1 protein 1 [Mus pahari]
MAADSDDGVPPVATSSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQVFNLTLLCVEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPGTCRRSHTTGICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALQSQPGTQTLAEAEDGPPTKQVCEDSLKTEKMEQKV
AEGEAGDQPGSRQDHTGSLEMEHRRTRAETAEVATSELLVSQARTDPVPGDGLHRAGFDAFMTGYVMAYVGLSQGLQ
LCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Rattus rattus XP_032755624.1 target of EGR1 protein 1 [Rattus rattus]
MAADSDDGVPSVPTPSDGGVNRTTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGDSSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY

AAEFHARFVASYLEYAFRKCERENGKQRAAGSPHVALEFCSYPSSMRGHIDYRCCMSPVSCRRSHTTGICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALQSQPGTQNLEEAEDGPPTKQVCEDSLKTEKIEQKV
AEGDQLGSTQGHKDSLEMACKRTADVPTSELLVNQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLSKGLQLCSSEPW
LPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Apodemus sylvaticus XP_052033054.1 target of EGR1 protein 1 [Apodemus sylvaticus]
MAADSDDGVPSVPTSSDGGVNKNTQTAEEFVVRVPVVDVQNDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSNMRGHIDYRCCMPPVTCRPPHTTGICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALQSQLGTQTLEEAEDGPPTKQVCEDNVKTEKVEQKV
AEGEAGDELGSRQGHTGSPEMAHRTSADTADVATSELPVNQANANPVPGDGLHRAGFDAFMTGYVMAYVGLSQGLQL
CSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Meriones unguiculatus XP_021489081.1 target of EGR1 protein 1 [Meriones unguiculatus]
MAADSVDGGLSVPPTSDGVVNKNTKSEEEFVIRVPVVDVQTDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLVQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTYHRSHTSGICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRRRALLSQPGMQTFEEAEEGPPTKQVCEDSLKTENPEQKV
AEGETRDQVGSKQGHEGGLEMEHEAPSSEIADVATSELPVNQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGTQ
LCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Arvicanthis niloticus XP_034357978.1 target of EGR1 protein 1 [Arvicanthis niloticus]
MAADSDDGVPSVPTSSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDQGDESQNQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYVFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTCRRSHTTGICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALQSQPQTQTLEEAEDGPPTKQVCEDSLQTEKIEQIM
AEGEAKDQLGSKQGHTGSLVMAHKRTSSETADMATSDLLVNQGSTNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGLQ
LCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Onychomys torridus XP_036033233.1 target of EGR1 protein 1 [Onychomys torridus]
MAADSDDGVPSVPTTSDGGVNKNTKSADEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLL
NQCIEERYKAVCHAARTRSVLSLGLACFKQQPAKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQ
GIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKY
AAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLTLEFCSYPSSMGGHIDYRCCMSPVTHRRSHTSSICDKFSAYG
WCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALLSQPGAQTFEEAEDGPPTKQICEDSLKTEETEQKV
TEGETKDQLGSQQDHKSGLAIRRKATSSETADVATSELPVSQANPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGPQ
LCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAWGSS

Alignment:

```
Homo-sapiens                        MAADSDDGAVSAPAASDGGVSKSTTSGEELVVQVPVVDVQSNNFKEMWPSLLLAIKTANF
Psammomys-obesus                    MAADSDDGGLSVPPTSDGVVNKNTKSEEEFVIRVPVVDVQTDNFKEIWPSLLLALKTASF
Meriones-unguiculatus               MAADSVDGGLSVPPTSDGVVNKNTKSEEEFVIRVPVVDVQTDNFKEIWPSLLLALKTASF
Acomys-russatus                     MAADSGDGVPLVPKTSDGGVNKST-SAEEFVARVPVVDVQTDNFKEIWPSLLLALKTASF
Mastomys-coucha                     MAADSADGMPSVPSSSDGGVNKNTQSAKEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
Phodopus-roborovskii                MAADSDDGVPSVPTTSDGGVNKNTKSAEEFVIRVPVVDVQSDNFKEIWPSLLLALKTASF
Cricetulus-griseus                  MAADSDDGVPSVPTTSDGGVNKNTKSAEEFVIRVPVVDVQSDNFKEIWPSLLLALKTASF
Mesocricetus-auratus                MAADSDDGVPSVPTTSDGGVNKNTKSAEKFVIRVPVVDVQSDNFKEIWPSLLLALKTASF
Neotoma-lepida                      MAADSDDGVPSVPTTSDDGVNKNTKSADEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
Onychomys-torridus                  MAADSDDGVPSVPTTSDGGVNKNTKSADEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
Peromyscus-californicus-insignis    MAADSVDGVPSVPTTSDGGVNKNTKSADEFIVRVPVVDVQSDNFKEIWPSLLLALKTASF
Mus-caroli                          MAADSDDGVPPVPTSSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
Mus-pahari                          MAADSDDGVPPVATSSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
```

```
Arvicanthis-niloticus                MAADSDDGVPSVPTSSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
Myodes-glareolus                     MAADSDDGVPSVPATSDGDVNKTTKSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
Chionomys-nivalis                    MAADSDDGVPSVPATSDGDVNKNTKSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
Microtus-ochrogaster                 MAADSDDGVPSVSATSDGDVNKNTKPAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
Apodemus-sylvaticus                  MAADSDDGVPSVPTSSDGGVNKNTQTAEEFVVRVPVVDVQNDNFKEIWPSLLLALKTASF
Mus-musculus                         MAADSDDGVPPVPTPSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
Rattus-norvegicus                    MAADSDDGVPSVPTPSDGGVNRTTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
Rattus-rattus                        MAADSDDGVPSVPTPSDGGVNRTTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASF
                                     ***** **   .. .**. *...* . .::: .*******.:****:*******:***.*

Homo-sapiens                         VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSILSLGLACFKRQPDKGEHSYLAQV
Psammomys-obesus                     VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLAQV
Meriones-unguiculatus                VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLVQV
Acomys-russatus                      VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQV
Mastomys-coucha                      VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPEKGENSYLAQV
Phodopus-roborovskii                 VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKHQADKGENSYLTQV
Cricetulus-griseus                   VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQV
Mesocricetus-auratus                 VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENCYLTQV
Neotoma-lepida                       VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQV
Onychomys-torridus                   VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPAKGENSYLTQV
Peromyscus-californicus-insignis     VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQV
Mus-caroli                           VAVDTELSGLGDRKNLLNQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQV
Mus-pahari                           VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQV
Arvicanthis-niloticus                VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQV
Myodes-glareolus                     VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQV
Chionomys-nivalis                    VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQV
Microtus-ochrogaster                 VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQV
Apodemus-sylvaticus                  VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLAQV
Mus-musculus                         VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQV
Rattus-norvegicus                    VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGDSSYLAQV
Rattus-rattus                        VAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGDSSYLAQV
                                     **************.*********************:********..*. **: .**.**

Homo-sapiens                         FNLTLLCMEEYVIEPKSVQFLIQHGFNFNQQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Psammomys-obesus                     FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Meriones-unguiculatus                FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Acomys-russatus                      FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Mastomys-coucha                      FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Phodopus-roborovskii                 FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Cricetulus-griseus                   FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Mesocricetus-auratus                 FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Neotoma-lepida                       FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQNQSVRTLFLEL
Onychomys-torridus                   FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Peromyscus-californicus-insignis     FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Mus-caroli                           FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRMLFLEL
Mus-pahari                           FNLTLLCVEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Arvicanthis-niloticus                FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDQGDESQNQSVRTLFLEL
Myodes-glareolus                     FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Chionomys-nivalis                    FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Microtus-ochrogaster                 FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Apodemus-sylvaticus                  FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Mus-musculus                         FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Rattus-norvegicus                    FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
Rattus-rattus                        FNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLEL
                                     *******:*************.*******.*************:*****.**** *****

Homo-sapiens                         IRARRPLVLHNGLIDLVFLYQNFYAHLPESLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Psammomys-obesus                     IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Meriones-unguiculatus                IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Acomys-russatus                      IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Mastomys-coucha                      IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFEARFV
Phodopus-roborovskii                 IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Cricetulus-griseus                   IRARRPLVLHNGLIDLVFLYQNFYAHLPETLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Mesocricetus-auratus                 IRARRPLVLHNGLIDLVFLYQNFYAHLPETLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Neotoma-lepida                       IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARF-
Onychomys-torridus                   IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Peromyscus-californicus-insignis     IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Mus-caroli                           IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYASEFHARFV
Mus-pahari                           IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Arvicanthis-niloticus                IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Myodes-glareolus                     IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Chionomys-nivalis                    IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Microtus-ochrogaster                 IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Apodemus-sylvaticus                  IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Mus-musculus                         IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
Rattus-norvegicus                    IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
```

```
Rattus-rattus                       IRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFV
                                    ******************************.**********************:** ***

Homo-sapiens                        ASYLEYAFRKCERENGKQRAAGSPHLTLEFCNYPSSMRDHIDYRCCLPPAT-HRPHPTSI
Psammomys-obesus                    ASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTYHRSHTGGI
Meriones-unguiculatus               ASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTYHRSHTSGI
Acomys-russatus                     ASYLEYAFRKCERENGKQRAAGGPHLALEFCSYPSSMRGHIDYRCCMSPVTDRRSYATGI
Mastomys-coucha                     ASYLEYAFRKCERENGRQQAVGSPHLALEFCSYPSSMRGHIDYRCCMSPVTCRRSYTTGI
Phodopus-roborovskii                ASYLEYAFRKCERENGKRRAAGTPHLVLEFCSYPSSMRGHIDYRCCLSPVTYRRSHTTSI
Cricetulus-griseus                  ASYLEYAFRKCERENGKQRAAGTPHLALEFCSYPSSMRGHIDYRCCMSPVTYRRSHTTGI
Mesocricetus-auratus                ASYLEYAFRKCERENGKQRAAGTPHLALEFCSYPSSMRSHIDYRCCISPVTYRRSHTTGI
Neotoma-lepida                      ----------------------------FCSYPSSMRDHIDYRCCMSPVTYRRSHTTGI
Onychomys-torridus                  ASYLEYAFRKCERENGKQRAAGSPHLTLEFCSYPSSMGGHIDYRCCMSPVTHRRSHTSSI
Peromyscus-californicus-insignis    ASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTYRRFHTSGI
Mus-caroli                          ASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCTSPGTCRRSRPTGI
Mus-pahari                          ASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPGTCRRSHTTGI
Arvicanthis-niloticus               ASYLEYVFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTCRRSHTTGI
Myodes-glareolus                    ASYLEYAFRKCERENGKRQAAGSPHLALEFCSYPSSMRGHIDYRCCMSPITSRRSHTTGI
Chionomys-nivalis                   ASYLEYAFRKCERENGKRQATGSPHLALEFCSYPSSMRGHIDYRCCMSPVTSRRSHTAGI
Microtus-ochrogaster                ASYLEYAFRKCERENGKRQAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTSRRSHTTGI
Apodemus-sylvaticus                 ASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSNMRGHIDYRCCMPPVTCRPPHTTGI
Mus-musculus                        ASYLEYPFPEM-------------------------------------------------
Rattus-norvegicus                   ASYLEYAFRK--------------------------------------------------
Rattus-rattus                       ASYLEYAFRKCERENGKQRAAGSPHVALEFCSYPSSMRGHIDYRCCMSPVSCRRSHTTGI

Homo-sapiens                        CDNFSAYGWCPLGPQCPQSHDIDLIIDTDEAAAEDKRRRR--RRREKRKRALLNLPGTQT
Psammomys-obesus                    CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRRRALLSQPGMQT
Meriones-unguiculatus               CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRRRALLSQPGMQT
Acomys-russatus                     CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRKR--RRKDKRRRALLRQPGTQT
Mastomys-coucha                     CDKFSAYGWCPLGPECPQSHDIDLIIDTDEAVAEDKRRRRWQRRKDKRKRALQSQPGTQT
Phodopus-roborovskii                CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRRRALLSQLGTQN
Cricetulus-griseus                  CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRKRALLSQPETQT
Mesocricetus-auratus                CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRKRALLSQPGTQT
Neotoma-lepida                      CNKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRRRALLSQPETQT
Onychomys-torridus                  CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRKRALLSQPGAQT
Peromyscus-californicus-insignis    CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRKRALLSQPGAQT
Mus-caroli                          CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRKRSLQSQPGTQA
Mus-pahari                          CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRKRALQSQPGTQT
Arvicanthis-niloticus               CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRKRALQSQPQTQT
Myodes-glareolus                    CTKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDRRKRALLNQPGTET
Chionomys-nivalis                   CTKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDRRKRALLSQPGTGT
Microtus-ochrogaster                CTRFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDRRKRALLSQPGTET
Apodemus-sylvaticus                 CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRKRALQSQLGTQT
Mus-musculus                        ------------------------------------------------------------
Rattus-norvegicus                   ------------------------------------------------------------
Rattus-rattus                       CDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRR--RRKDKRKRALQSQPGTQN

Homo-sapiens                        SGEAKDGPPKKQVCGDSIKPE-----------------------------ETEQEVAAD
Psammomys-obesus                    FEEAEEGPPTKQVCEDSLKTE-----------------------------NTEQKVAEG
Meriones-unguiculatus               FEEAEEGPPTKQVCEDSLKTE-----------------------------NPEQKVAEG
Acomys-russatus                     FEEGEDGPPTKQVREDSLNTE-----------------------------NTEQKVAEG
Mastomys-coucha                     LEEAEGGPPTKQVCEDSLKAE-----------------------------KMEQKVAEG
Phodopus-roborovskii                FEEADDGPPTKQVCEDNLKTQ-----------------------------ELEQRVTER
Cricetulus-griseus                  FEEAEDGPPTKQVCEDSLKTE-----------------------------EMEQRVTEG
Mesocricetus-auratus                FEEAEDGPPTKQVCEDNLKTE-----------------------------EIEQRVTEG
Neotoma-lepida                      FEEAEDGPPTKQVCEDSLKTE-----------------------------EIEQKVTEG
Onychomys-torridus                  FEEAEDGPPTKQICEDSLKTE-----------------------------ETEQKVTEG
Peromyscus-californicus-insignis    FEEAEDGPPTKQVCEDSLKTE-----------------------------EIEQKVTEG
Mus-caroli                          LAEAEDGPPTKQVCEDSLKTE-----------------------------KMEQKVAEG
Mus-pahari                          LAEAEDGPPTKQVCEDSLKTE-----------------------------KMEQKVAEG
Arvicanthis-niloticus               LEEAEDGPPTKQVCEDSLQTE-----------------------------KIEQIMAEG
Myodes-glareolus                    FEEAEDGPPTKQVCEDSLKTE-----------------------------IEQKVTEG
Chionomys-nivalis                   FEEAEDGPPTKQVCEDSLKTETEQKVTEGESRDQAEDGLPTKQVCEGSLKTEIEQKVTEG
Microtus-ochrogaster                FEEAEDGPPIKQVCEDSLKTE-----------------------------IEQKVTEG
Apodemus-sylvaticus                 LEEAEDGPPTKQVCEDNVKTE-----------------------------KVEQKVAEG
Mus-musculus                        ------------------------------------------------------------
Rattus-norvegicus                   ------------------------------------------------------------
Rattus-rattus                       LEEAEDGPPTKQVCEDSLKTE-----------------------------KIEQKVAEG

Homo-sapiens                        ETRNLPHSKQGNKNDLEMGIKAARPEIADRATSEVPGSQASPNPVPGDGLHRAGFDAFMT
Psammomys-obesus                    ETRDQPGSKQGHKGGLEMEHEAVSSEIADVATSELPVNQTSPNPVPGDGLHRAGFDAFMT
Meriones-unguiculatus               ETRDQVGSKQGHEGGLEMEHEAPSSEIADVATSELPVNQASPNPVPGDGLHRAGFDAFMT
Acomys-russatus                     ETSDQLGSKQGHKGGLEMKHEATGSETADVATSELQVNQASPNPVPGDGLHRAGFDAFMT
```

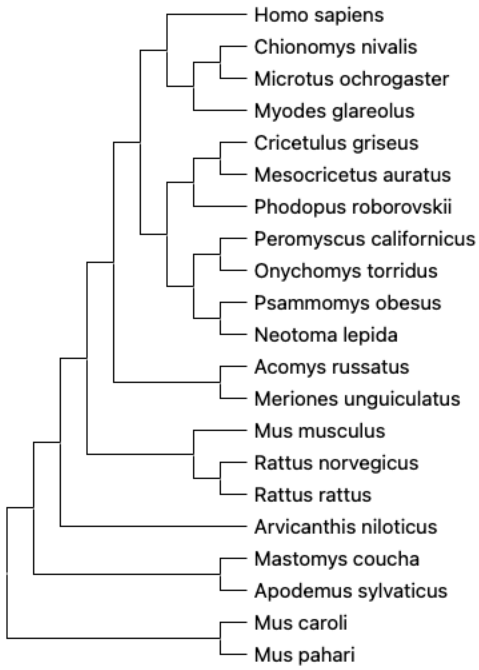
```
Mastomys-coucha                  DAGDQLGSRQGHTGSLEIAHRRTSAETADVAPSELPVSQASTNPLPGDGLHRAGFDAFMT
Phodopus-roborovskii             ETRDELDSKQGHKSDLEMEHKATGSETADVAISELPVSQASPNPMPGDGLHRAGFDAFMT
Cricetulus-griseus               ETRDELGSKQAHKSGLEMEHKATSSETVDVATTELPVSQASPNPVPGDGLHRAGFDAFMT
Mesocricetus-auratus             EIRDELGSKQAHKSGSEMEHKATSSETVDVATSELPVSQASPNPVPGDGLHRAGFDAFMT
Neotoma-lepida                   ETRDQLGSQQGHKSSLEIEYKATSSKIADVATSELPVSQASPNPMPGDGLHRAGFDAFMT
Onychomys-torridus               ETKDQLGSQQDHKSGLAIRRKATSSETADVATSELPVSQANPNPVPGDGLHRAGFDAFMT
Peromyscus-californicus-insignis ETRDQLGSQQGHKSGLAVERKATSSETAEVATSELPVSQANPNPGPGDGLHRAGFDAFMT
Mus-caroli                       EAGDQPGSREGHTSSLEMAHRRTSAETADVATSELLVNQASTNPVPGDGLHRAGFDAFMT
Mus-pahari                       EAGDQPGSRQDHTGSLEMEHRRTRAETAEVATSELLVSQARTDPVPGDGLHRAGFDAFMT
Arvicanthis-niloticus            EAKDQLGSKQGHTGSLVMAHKRTSSETADMATSDLLVNQGSTNPVPGDGLHRAGFDAFMT
Myodes-glareolus                 ETRDQLGSNQDHKSDLEVEHKATSSEIADVAASELPVSQASPNPVPGDGLHRAGFDAFMT
Chionomys-nivalis                ESRDQLGSKQCHNSDLEVEHKATSSEIADVAASELPASQASPNPVPGDGLHRAGFDAFMT
Microtus-ochrogaster             ESRDQLGSKQDQNSDLEVEHKATSSEIADVAASELPVSQASPNPVPGDGLHRAGFDAFMT
Apodemus-sylvaticus              EAGDELGSRQGHTGSPEMAHR-TSADTADVATSELPVNQANANPVPGDGLHRAGFDAFMT
Mus-musculus                     -STGRMGSKRG----------------------------------------------------
Rattus-norvegicus                ----------------------------------------------------------------
Rattus-rattus                    ---DQLGSTQGHKDSLEMACKRT----ADVPTSELLVNQASPNPVPGDGLHRAGFDAFMT


Homo-sapiens                     GYVMAYVEVSQGPQPCSSGPWLPECHNKVYLSGKAVPLTVAKSQFSRSSKAHNQKMKLTW
Psammomys-obesus                 GYVMAYVGLSQGTQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAW
Meriones-unguiculatus            GYVMAYVGLSQGTQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAW
Acomys-russatus                  GYVMAYVGLSQGLQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAW
Mastomys-coucha                  GYVMAYVGLSQGLQLCSSEPWLPKCHNKVYLSGKTVPLTVAKSQFSHPSKAHKQKMKLAW
Phodopus-roborovskii             GYVMAYVGLSQGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSHSSKAHNQKMKLAW
Cricetulus-griseus               GYVMAYVGLSQGPQLCSSRPWLPECHNKVYLSGKTVPLTVAKSQFSHSSKAHNQKMKLAW
Mesocricetus-auratus             GYVMAYVGLSQGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSHSSKAHNQKMKLAW
Neotoma-lepida                   GYVMAYVGLSQGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAW
Onychomys-torridus               GYVMAYVGLSQGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAW
Peromyscus-californicus-insignis GYVMAYVGLSQGPQLCSSGPWLPECHNKVYLSGKTVPLTVARSQFSRSSKAHNQKMKLAW
Mus-caroli                       GYVMAYVGLSQGLQLCSSEPWLPECHNKVYLSGKTVPLTVTKSQFSRPSKAHNQKMKLAW
Mus-pahari                       GYVMAYVGLSQGLQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAW
Arvicanthis-niloticus            GYVMAYVGLSQGLQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAW
Myodes-glareolus                 GYVMAYVGLRHGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAW
Chionomys-nivalis                GYVMAYVGLRHGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAW
Microtus-ochrogaster             GYVMAYVGLRHGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAW
Apodemus-sylvaticus              GYVMAYVGLSQGLQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAW
Mus-musculus                     ------------------------------------SLATHI------------------
Rattus-norvegicus                --------------------W---------------PMA--------------------
Rattus-rattus                    GYVMAYVGLSKGLQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAW
                                                                        .::

Homo-sapiens                     GSS
Psammomys-obesus                 GSS
Meriones-unguiculatus            GSS
Acomys-russatus                  GNS
Mastomys-coucha                  GSS
Phodopus-roborovskii             SSS
Cricetulus-griseus               GSS
Mesocricetus-auratus             GSS
Neotoma-lepida                   GSS
Onychomys-torridus               GSS
Peromyscus-californicus-insignis GSS
Mus-caroli                       GSS
Mus-pahari                       GSS
Arvicanthis-niloticus            GSS
Myodes-glareolus                 GSS
Chionomys-nivalis                GSS
Microtus-ochrogaster             GSS
Apodemus-sylvaticus              GSS
Mus-musculus                     ---
Rattus-norvegicus                GAL
Rattus-rattus                    GSS
```

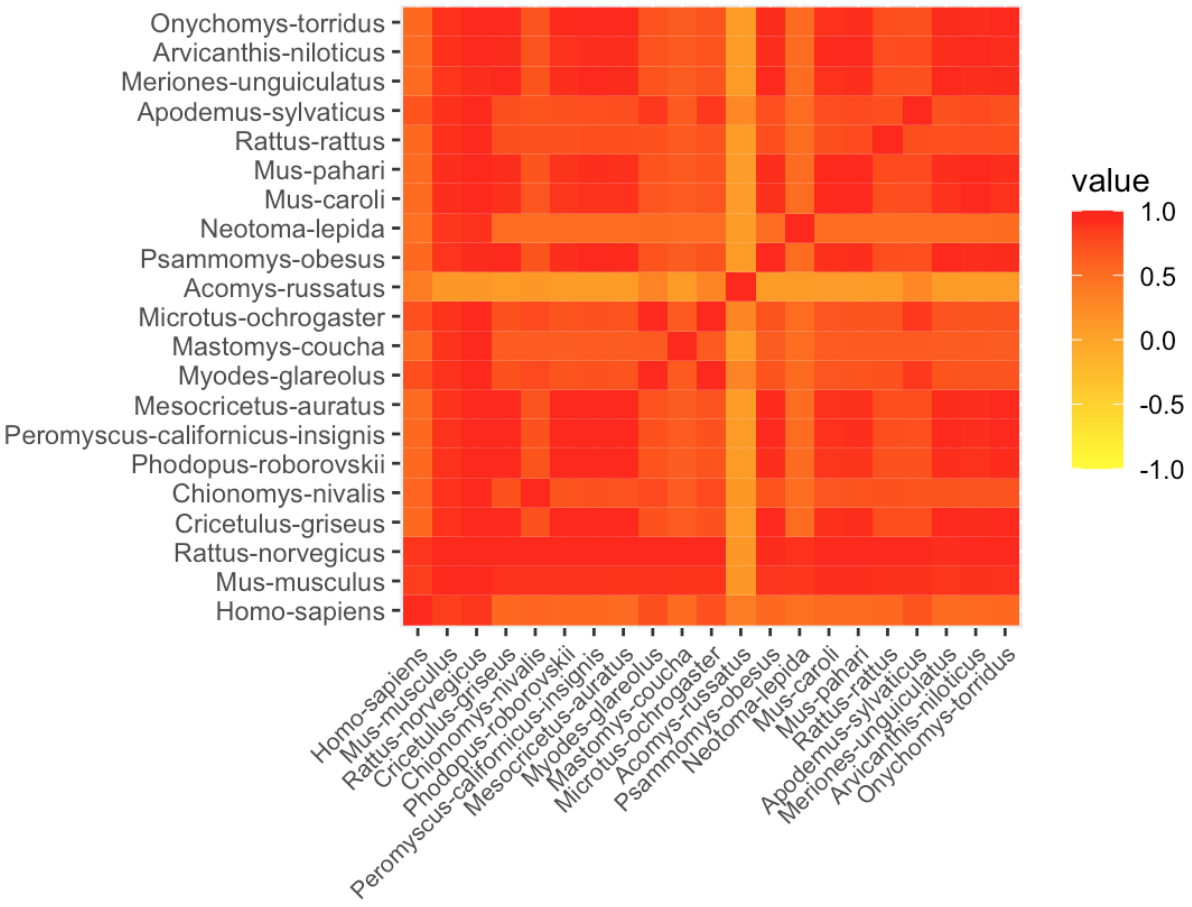
[Q6] Create a phylogenetic tree, using either a parsimony or distance-based approach. Bootstrapping and tree rooting are optional. Use “simple phylogeny” online from the EBI or any respected phylogeny program (such as MEGA, PAUP, or Phylip). Paste an image of your Cladogram or tree output in your report.

Tree created through MEGA.



[**Q7**] Generate a sequence identity based **heatmap** of your aligned sequences using R.

If necessary convert your sequence alignment to the ubiquitous FASTA format (Seaview can read in clustal format and “Save as” FASTA format for example). Read this FASTA format alignment into R with the help of functions in the **Bio3D package**. Calculate a sequence identity matrix (again using a function within the Bio3D package). Then generate a heatmap plot and add to your report. Do make sure your labels are visible and not cut at the figure margins.



[Q8] Using R/Bio3D (or an online blast server if you prefer), search the main protein structure database for the most similar atomic resolution structures to your aligned sequences.

List the top 3 *unique* hits (i.e. not hits representing different chains from the same structure) along with their Evalue and sequence identity to your query. Please also add annotation details of these structures. For example include the annotation terms PDB identifier (structured), Method used to solve the structure (experimental Technique), resolution (resolution), and source organism (source).

```
Similar atomic resolution structures:

I used two methods to validate this date, PDB website and the NCBI Blast
website. Both methods only provided these three structures.
```

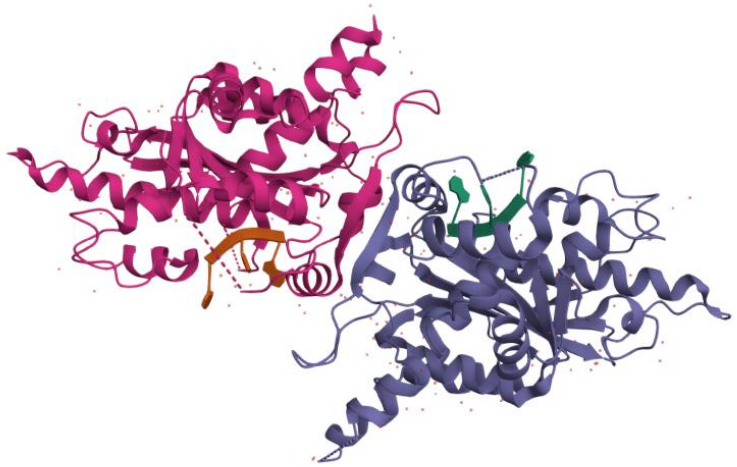
ID	Technique	Resolution	Source	Evalue	Identity
2FC6	SOLUTION NMR		Homo sapiens	1.99e-12	76

3D45	X-RAY DIFFRACTION	3 Å	Mus musculus	7.51e-7	26
2A1S	X-RAY DIFFRACTION	2.6 Å	Homo sapiens	1.068e-7	29

[Q9] Generate a molecular figure of one of your identified PDB structures using **VMD**. You can optionally highlight conserved residues that are likely to be functional. Please use a white or transparent background for your figure (i.e. not the default black).

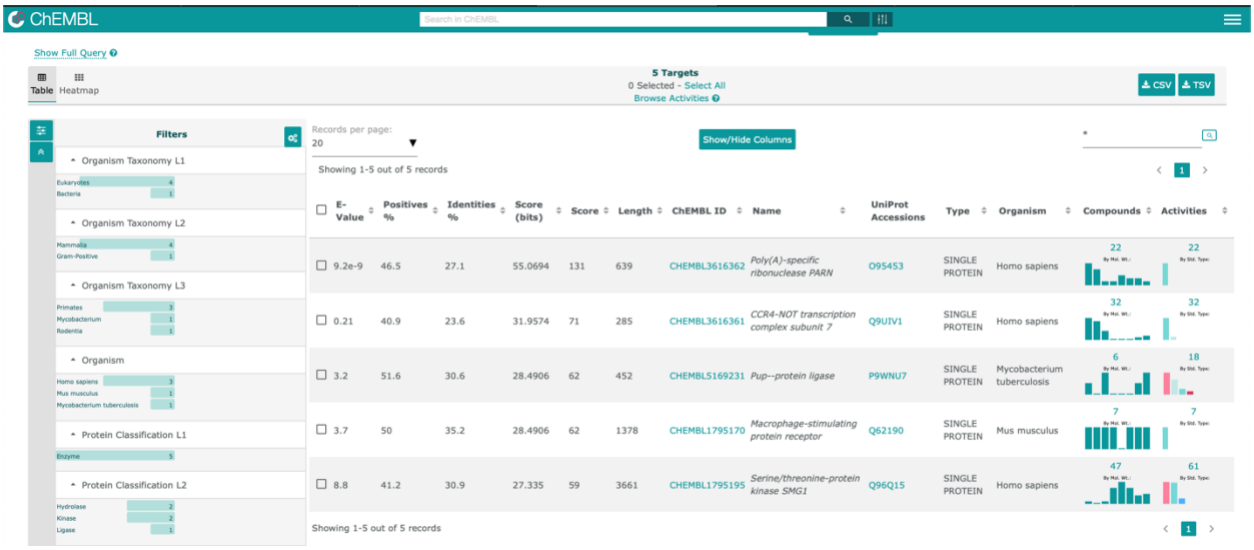
Based on sequence similarity. How likely is this structure to be similar to your "novel" protein?

2FC6 is very likely to be similar in structure to Mus musculus TOE1 like protein given the high sequence similarity (>76%).



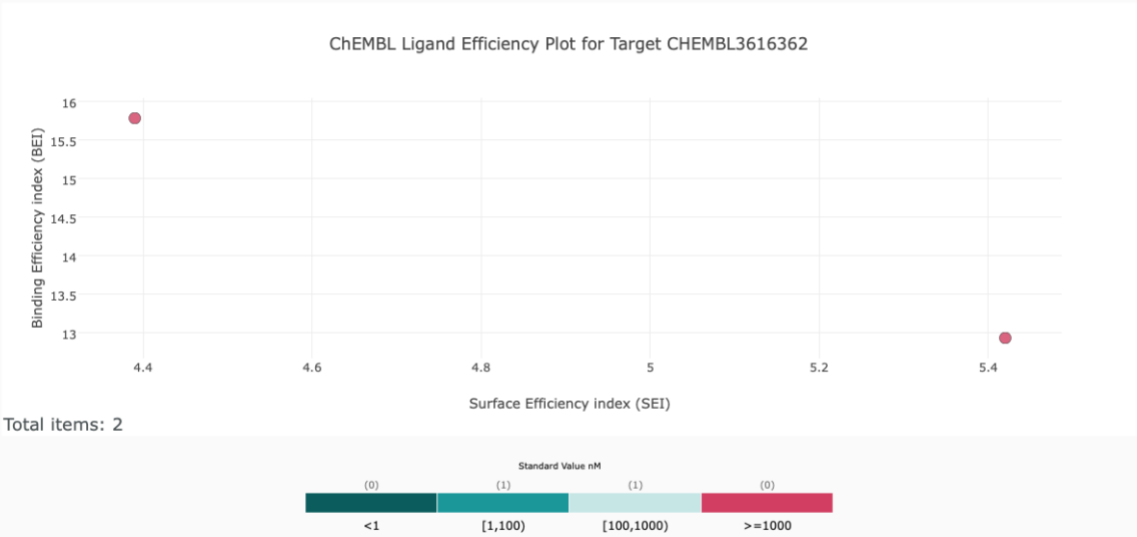
[Q10] Perform a "Target" search of ChEMBEL (https://www.ebi.ac.uk/chembl/) with your novel sequence. Are there any **Target Associated Assays** and **ligand efficiency data** reported that may be useful starting points for exploring potential inhibition of your novel protein?

Yes, there are five different targets. Within those targets, there are several assays and ligand efficiency data that could be useful starting points for exploring potential inhibition of the novel protein. One strong contender is purine-2,6-dione which inhibited PARN. PARN has a similar structure to the novel protein.



Ligand Efficiencies

See all bioactivities for target CHEMBL3616362 used in this visualisation



Total items: 2

The Ligand Efficiency chart plots Binding Efficiency Index (BEI) against Surface Efficiency Index (SEI), where:

SEI = (-log10(Standard Value*10^-9))*100/PSA
BEI = (-log10(Standard Value*10^-9))*1000/MWT

Target Report Card

Name And Classification

ID:	CHEMBL3616362
Type:	SINGLE PROTEIN
Preferred Name:	Poly(A)-specific ribonuclease PARN
Synonyms:	DAN, Deadenylating nuclease, Deadenylation nuclease, PARN, Polyadenylate-specific ribonuclease, Poly(A)-specific ribonuclease PARN
Organism:	Homo sapiens
Species Group:	No
Protein Target Classification:	- Enzyme > Hydrolase

Assay Report Card

Basic Information

Assay ID:	CHEMBL3619398
Type:	Binding
Description:	Inhibition of PARN (unknown origin)
Format:	BAO_0000357
Journal:	Bioorg Med Chem Lett (2015) 25:4219-4224
Organism:	Homo sapiens
Strain:	---
Tissue:	---
Cell Type:	---
Subcellular Fraction:	---
Target:	CHEMBL3616362
Document:	CHEMBL3616430
Cell:	
Tissue:	

Discovery, synthesis and biochemical profiling of purine-2,6-dione derivatives as inhibitors of the human poly(A)-selective ribonuclease Caf1

https://doi.org/10.1016/j.bmcl.2015.07.095.