Find a Gene Project

BGGN 213

Kalodiah Toma

PID: A07606689

Email: [k1toma@ucsd.edu](mailto: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**](https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=10090&lvl=3&lin=f&keep=1&srchmode=1&unlock). See below for alignment details.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

[**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)  MAADSDDGVPPVPTPSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYPFPEMSTGRMGSKRGSLATHI

Species: Mus musculus

[Eukaryota](https://en.wikipedia.org/wiki/Eukaryote); [Animalia](https://en.wikipedia.org/wiki/Animal); [Chordata](https://en.wikipedia.org/wiki/Chordate); [Mammalia](https://en.wikipedia.org/wiki/Mammal); [Rodentia](https://en.wikipedia.org/wiki/Rodent); [Muridae](https://en.wikipedia.org/wiki/Muridae); [Mus](https://en.wikipedia.org/wiki/Mus_(genus)); [Mus](https://en.wikipedia.org/wiki/Mus_(subgenus)); 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:

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

[**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]

MAADSDDGAVSAPAASDGGVSKSTTSGEELVVQVPVVDVQSNNFKEMWPSLLLAIKTANFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSILSLGLACFKRQPDKGEHSYLAQVFNLTLLCMEEYVIEPKSVQFLIQHGFNFNQQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPESLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLTLEFCNYPSSMRDHIDYRCCLPPATHRPHPTSICDNFSAYGWCPLGPQCPQSHDIDLIIDTDEAAAEDKRRRRRRREKRKRALLNLPGTQTSGEAKDGPPKKQVCGDSIKPEETEQEVAADETRNLPHSKQGNKNDLEMGIKAARPEIADRATSEVPGSQASPNPVPGDGLHRAGFDAFMTGYVMAYVEVSQGPQPCSSGPWLPECHNKVYLSGKAVPLTVAKSQFSRSSKAHNQKMKLTWGSS

>Mus musculus  TOE1 like protein (sequence taken from BLAST result)  MAADSDDGVPPVPTPSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYPFPEMSTGRMGSKRGSLATHI

>Rattus norvegicus XP\_038965533.1 target of EGR1 protein 1 isoform X4 [Rattus norvegicus]

MAADSDDGVPSVPTPSDGGVNRTTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGDSSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKWPMAGAL

>Cricetulus griseus XP\_027258385.1 target of EGR1 protein 1 [Cricetulus griseus]

MAADSDDGVPSVPTTSDGGVNKNTKSAEEFVIRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPETLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKQRAAGTPHLALEFCSYPSSMRGHIDYRCCMSPVTYRRSHTTGICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALLSQPETQTFEEAEDGPPTKQVCEDSLKTEEMEQRVTEGETRDELGSKQAHKSGLEMEHKATSSETVDVATTELPVSQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGPQLCSSRPWLPECHNKVYLSGKTVPLTVAKSQFSHSSKAHNQKMKLAWGSS

>Chionomys nivalis XP\_057640769.1 target of EGR1 protein 1 [Chionomys nivalis]

MAADSDDGVPSVPATSDGDVNKNTKSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKRQATGSPHLALEFCSYPSSMRGHIDYRCCMSPVTSRRSHTAGICTKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDRRKRALLSQPGTGTFEEAEDGPPTKQVCEDSLKTETEQKVTEGESRDQAEDGLPTKQVCEGSLKTEIEQKVTEGESRDQLGSKQCHNSDLEVEHKATSSEIADVAASELPASQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLRHGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAWGSS

>Phodopus roborovskii XP\_051044502.1 target of EGR1 protein 1 [Phodopus roborovskii]

MAADSDDGVPSVPTTSDGGVNKNTKSAEEFVIRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKHQADKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKRRAAGTPHLVLEFCSYPSSMRGHIDYRCCLSPVTYRRSHTTSICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRRRALLSQLGTQNFEEADDGPPTKQVCEDNLKTQELEQRVTERETRDELDSKQGHKSDLEMEHKATGSETADVAISELPVSQASPNPMPGDGLHRAGFDAFMTGYVMAYVGLSQGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSHSSKAHNQKMKLAWSSS

>Peromyscus californicus insignis XP\_052572255.1 target of EGR1 protein 1 [Peromyscus californicus insignis]

MAADSVDGVPSVPTTSDGGVNKNTKSADEFIVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTYRRFHTSGICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALLSQPGAQTFEEAEDGPPTKQVCEDSLKTEEIEQKVTEGETRDQLGSQQGHKSGLAVERKATSSETAEVATSELPVSQANPNPGPGDGLHRAGFDAFMTGYVMAYVGLSQGPQLCSSGPWLPECHNKVYLSGKTVPLTVARSQFSRSSKAHNQKMKLAWGSS

>Mesocricetus auratus XP\_005072112.3 target of EGR1 protein 1 [Mesocricetus auratus]

MAADSDDGVPSVPTTSDGGVNKNTKSAEKFVIRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENCYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPETLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKQRAAGTPHLALEFCSYPSSMRSHIDYRCCISPVTYRRSHTTGICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALLSQPGTQTFEEAEDGPPTKQVCEDNLKTEEIEQRVTEGEIRDELGSKQAHKSGSEMEHKATSSETVDVATSELPVSQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSHSSKAHNQKMKLAWGSS

>Myodes glareolus XP\_048286556.1 target of EGR1 protein 1 [Myodes glareolus]

MAADSDDGVPSVPATSDGDVNKTTKSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKRQAAGSPHLALEFCSYPSSMRGHIDYRCCMSPITSRRSHTTGICTKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDRRKRALLNQPGTETFEEAEDGPPTKQVCEDSLKTEIEQKVTEGETRDQLGSNQDHKSDLEVEHKATSSEIADVAASELPVSQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLRHGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAWGSS

>Mastomys coucha XP\_031234176.1 target of EGR1 protein 1 isoform X1 [Mastomys coucha]

MAADSADGMPSVPSSSDGGVNKNTQSAKEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPEKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFEARFVASYLEYAFRKCERENGRQQAVGSPHLALEFCSYPSSMRGHIDYRCCMSPVTCRRSYTTGICDKFSAYGWCPLGPECPQSHDIDLIIDTDEAVAEDKRRRRWQRRKDKRKRALQSQPGTQTLEEAEGGPPTKQVCEDSLKAEKMEQKVAEGDAGDQLGSRQGHTGSLEIAHRRTSAETADVAPSELPVSQASTNPLPGDGLHRAGFDAFMTGYVMAYVGLSQGLQLCSSEPWLPKCHNKVYLSGKTVPLTVAKSQFSHPSKAHKQKMKLAWGSS

>Microtus ochrogaster XP\_026633303.1 target of EGR1 protein 1 isoform X1 [Microtus ochrogaster]

MAADSDDGVPSVSATSDGDVNKNTKPAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKRQAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTSRRSHTTGICTRFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDRRKRALLSQPGTETFEEAEDGPPIKQVCEDSLKTEIEQKVTEGESRDQLGSKQDQNSDLEVEHKATSSEIADVAASELPVSQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLRHGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAWGSS

>Acomys russatus XP\_051027553.1 target of EGR1 protein 1 [Acomys russatus]

MAADSGDGVPLVPKTSDGGVNKSTSAEEFVARVPVVDVQTDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKQRAAGGPHLALEFCSYPSSMRGHIDYRCCMSPVTDRRSYATGICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRKRRRKDKRRRALLRQPGTQTFEEGEDGPPTKQVREDSLNTENTEQKVAEGETSDQLGSKQGHKGGLEMKHEATGSETADVATSELQVNQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGLQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGNS

>Psammomys obesus XP\_055474418.1 target of EGR1 protein 1 [Psammomys obesus]

MAADSDDGGLSVPPTSDGVVNKNTKSEEEFVIRVPVVDVQTDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTYHRSHTGGICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRRRALLSQPGMQTFEEAEEGPPTKQVCEDSLKTENTEQKVAEGETRDQPGSKQGHKGGLEMEHEAVSSEIADVATSELPVNQTSPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGTQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Neotoma lepida OBS83234.1 hypothetical protein A6R68\_22771 [Neotoma lepida]

MAADSDDGVPSVPTTSDDGVNKNTKSADEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQNQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFFCSYPSSMRDHIDYRCCMSPVTYRRSHTTGICNKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRRRALLSQPETQTFEEAEDGPPTKQVCEDSLKTEEIEQKVTEGETRDQLGSQQGHKSSLEIEYKATSSKIADVATSELPVSQASPNPMPGDGLHRAGFDAFMTGYVMAYVGLSQGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAWGSS

>Mus caroli XP\_021015031.1 target of EGR1 protein 1 [Mus caroli]

MAADSDDGVPPVPTSSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKNLLNQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRMLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYASEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCTSPGTCRRSRPTGICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRSLQSQPGTQALAEAEDGPPTKQVCEDSLKTEKMEQKVAEGEAGDQPGSREGHTSSLEMAHRRTSAETADVATSELLVNQASTNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGLQLCSSEPWLPECHNKVYLSGKTVPLTVTKSQFSRPSKAHNQKMKLAWGSS

>Mus pahari XP\_021055456.1 target of EGR1 protein 1 [Mus pahari]

MAADSDDGVPPVATSSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFRQQPDKGENSYLAQVFNLTLLCVEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPGTCRRSHTTGICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALQSQPGTQTLAEAEDGPPTKQVCEDSLKTEKMEQKVAEGEAGDQPGSRQDHTGSLEMEHRRTRAETAEVATSELLVSQARTDPVPGDGLHRAGFDAFMTGYVMAYVGLSQGLQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Rattus rattus XP\_032755624.1 target of EGR1 protein 1 [Rattus rattus]

MAADSDDGVPSVPTPSDGGVNRTTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGDSSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKQRAAGSPHVALEFCSYPSSMRGHIDYRCCMSPVSCRRSHTTGICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALQSQPGTQNLEEAEDGPPTKQVCEDSLKTEKIEQKVAEGDQLGSTQGHKDSLEMACKRTADVPTSELLVNQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLSKGLQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Apodemus sylvaticus XP\_052033054.1 target of EGR1 protein 1 [Apodemus sylvaticus]

MAADSDDGVPSVPTSSDGGVNKNTQTAEEFVVRVPVVDVQNDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLAQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSNMRGHIDYRCCMPPVTCRPPHTTGICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALQSQLGTQTLEEAEDGPPTKQVCEDNVKTEKVEQKVAEGEAGDELGSRQGHTGSPEMAHRTSADTADVATSELPVNQANANPVPGDGLHRAGFDAFMTGYVMAYVGLSQGLQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Meriones unguiculatus XP\_021489081.1 target of EGR1 protein 1 [Meriones unguiculatus]

MAADSVDGGLSVPPTSDGVVNKNTKSEEEFVIRVPVVDVQTDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLVQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTYHRSHTSGICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRRRALLSQPGMQTFEEAEEGPPTKQVCEDSLKTENPEQKVAEGETRDQVGSKQGHEGGLEMEHEAPSSEIADVATSELPVNQASPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGTQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Arvicanthis niloticus XP\_034357978.1 target of EGR1 protein 1 [Arvicanthis niloticus]

MAADSDDGVPSVPTSSDGGVNKNTQSAEEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPDKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDQGDESQNQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYVFRKCERENGKQRAAGSPHLALEFCSYPSSMRGHIDYRCCMSPVTCRRSHTTGICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALQSQPQTQTLEEAEDGPPTKQVCEDSLQTEKIEQIMAEGEAKDQLGSKQGHTGSLVMAHKRTSSETADMATSDLLVNQGSTNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGLQLCSSEPWLPECHNKVYLSGKTVPLTVAKSQFSRPSKAHNQKMKLAWGSS

>Onychomys torridus XP\_036033233.1 target of EGR1 protein 1 [Onychomys torridus]

MAADSDDGVPSVPTTSDGGVNKNTKSADEFVVRVPVVDVQSDNFKEIWPSLLLALKTASFVAVDTELSGLGDRKSLLNQCIEERYKAVCHAARTRSVLSLGLACFKQQPAKGENSYLTQVFNLTLLCMEEYVIEPKSVQFLVQHGFNFNRQYAQGIPYHKGNDKGDESQSQSVRTLFLELIRARRPLVLHNGLIDLVFLYQNFYAHLPENLGTFTADLCEMFPAGIYDTKYAAEFHARFVASYLEYAFRKCERENGKQRAAGSPHLTLEFCSYPSSMGGHIDYRCCMSPVTHRRSHTSSICDKFSAYGWCPLGPQCPQSHDIDLIIDTDEAVAEDKRRRRRRKDKRKRALLSQPGAQTFEEAEDGPPTKQICEDSLKTEETEQKVTEGETKDQLGSQQDHKSGLAIRRKATSSETADVATSELPVSQANPNPVPGDGLHRAGFDAFMTGYVMAYVGLSQGPQLCSSGPWLPECHNKVYLSGKTVPLTVAKSQFSRSSKAHNQKMKLAWGSS

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.

A chart of a tree

Description automatically generated with medium confidence

[**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.

A red and orange squares with white text

Description automatically generated

[**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 proteingiven the high sequence similarity (>76%).

A close-up of a protein

Description automatically generated

[**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. A screenshot of a computer

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a report card

Description automatically generated

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.