Further investigation of the gene which causes muscle and connective tissue to turn to bone

Alexia McGregor, s1789372 November 2017

Introduction

Orthologue: *"Orthologs are genes in different species that evolved from a common ancestral gene by speciation. Normally, orthologs retain the same function in the course of evolution. Identification of orthologs is critical for reliable prediction of gene function in newly sequenced genomes." [1]*

Homologue: *"A gene related to a second gene by descent from a common ancestral DNA sequence. The term, homolog, may apply to the relationship between genes separated by the event of speciation (see ortholog) or to the relationship betwen genes separated by the event of genetic duplication (see paralog)." [1]*

Question 1

Using a BLAST query to find potential orthologues for the FOP gene

To investigate whether there were any orthologues present in other species, I took the following steps:

- I navigated to the BLAST webpage [10] and selected nucleotide BLAST search, navigated to the blastp tab, entered the protein accession number (NP_001096.1 [5]) into the query sequence, chose the reference proteins database, selected to show a maximum of 50 aligned sequences and changed the max e threshold to 10^{-10} [2] from the algorithm parameters section, and hit the search button.
- The search returned names of genes to me with an associated maximum score, total score, query cover, e-value and percentage identity .
- If the BLAST search returned genes with e-values smaller than 10^{-10} [2], then I concluded that the two were potential orthologues. All of the returned genes from my BLAST search had an e-value of 0 (at least according to BLAST - in reality, they were probably extremely small values which were close to 0), so I concluded that as far as this search was concerned, they could all be considered potential orthologues. The returned BLAST search containing all of these genes is listed in the appendix.

Using an NCBI Homologene query to find potential homologues for the FOP gene

To investigate whether there were any homologues present, I took the following steps:

- Searched for ACVR1 on the Homologene database on the NCBI website [5]. Under the 'Genes' section, there was list of genes which have been selected as probable homologues by Homologene.
- From that page, I clicked on pairwise alignment scores to take a look at the protein identity percentages and investigate whether they matched the percentages from my BLAST search.

These were the results I got from Homologene: [3]

Pairwise Alignment Scores

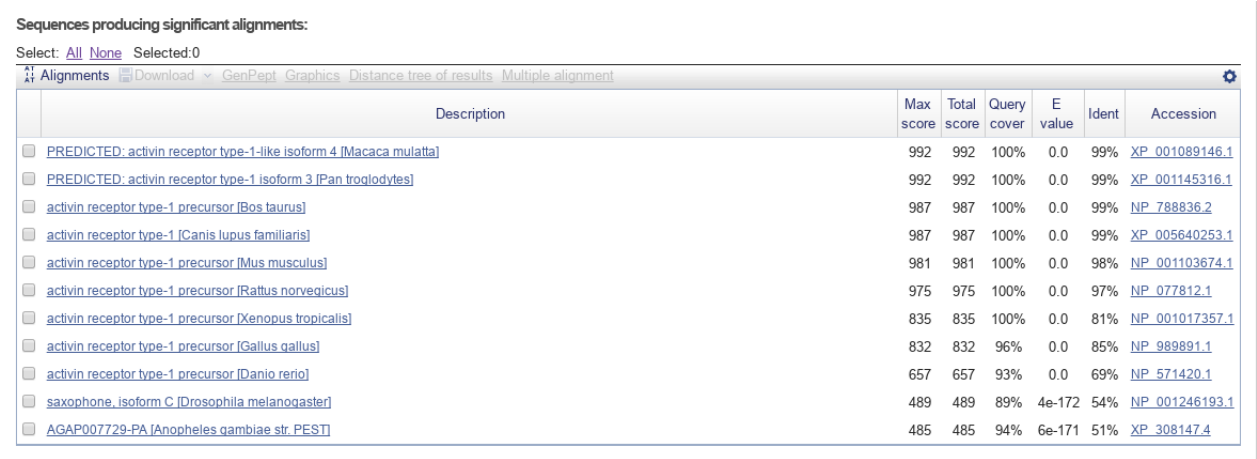
Gene		Identity (%)	
Species	Symbol	Protein	DNA
H.sapiens	**ACVR1**		
vs. P.troglodytes	ACVR1	99.8	99.6
vs. M.mulatta	ACVR1	99.8	98.7
vs. C.lupus	ACVR1	99.2	94.0
vs. B.taurus	ACVR1	99.2	93.1
vs. M.musculus	Acvr1	98.4	90.6
vs. R.norvegicus	Acvr1	97.2	89.5
vs. G.gallus	ACVR1	85.6	77.5
vs. X.tropicalis	acvr1	81.9	73.9
vs. D.rerio	acvr1l	70.3	65.6
vs. D.melanogaster	sax	57.1	57.1
vs. A.gambiae	AgaP_AGAP007729	54.5	56.0

The protein percentage identities for these are the same as in my BLAST search, except I notice that BLAST rounds to the nearest integer and HomoloGene rounds to one decimal place.

Comparing results from both queries

When I checked whether the genes from the search in HomoloGene were also in my original BLAST search, I found that the top few were, and then from G.gallus onwards, they were not. I speculated that this was simply because I only returned 100 possible orthologues from the BLAST search, and had I not set this constraint, I would have found all the genes in my BLAST search.

To test this theory out, I repeated the protein BLAST search as before, except this time I clicked on 'align more than two sequences', and I put the accession number of the proteins of each of the genes which were returned to my on HomoloGene. The results of this BLAST search were as follows:



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: activin receptor type-1-like isoform 4 [Macaca mulatta]	992	992	100%	0.0	99%	XP_001089146.1
PREDICTED: activin receptor type-1 isoform 3 [Pan troglodytes]	992	992	100%	0.0	99%	XP_001145316.1
activin receptor type-1 precursor [Bos taurus]	987	987	100%	0.0	99%	NP_788836.2
activin receptor type-1 [Canis lupus familiaris]	987	987	100%	0.0	99%	XP_005640253.1
activin receptor type-1 precursor [Mus musculus]	981	981	100%	0.0	98%	NP_001103674.1
activin receptor type-1 precursor [Rattus norvegicus]	975	975	100%	0.0	97%	NP_077812.1
activin receptor type-1 precursor [Xenopus tropicalis]	835	835	100%	0.0	81%	NP_001017357.1
activin receptor type-1 precursor [Gallus gallus]	832	832	96%	0.0	85%	NP_989891.1
activin receptor type-1 precursor [Danio rerio]	657	657	93%	0.0	69%	NP_571420.1
saxophone, isoform C [Drosophila melanogaster]	489	489	89%	4e-172	54%	NP_001246193.1
AGAP007729-PA [Anopheles gambiae str. PEST]	485	485	94%	6e-171	51%	XP_308147.4

The e-values for all of these were very small (less than the threshhold of 10^{-10}), and so I conclude that these are all potential orthologues.

Are there any true orthologues?

From my searches and just looking at e-values, I did not discard ant potential orthologues and I came to the conlcusion that all the 100 genes returned to my in the BLAST search and all the results in the HomoloGene search were potential orthologues.

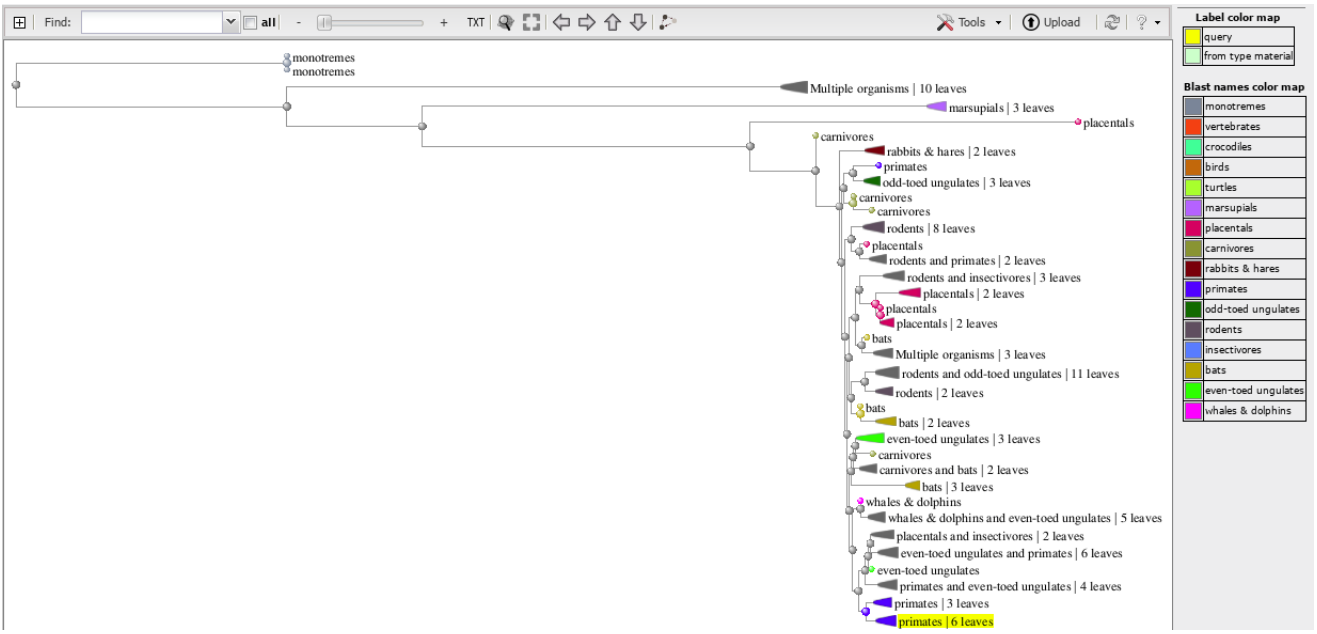
When I search for the ACVR1 gene on the NCBI site [5], and scroll down to 'General Gene Information', I see that there are 210 orthologues in other species from the 'Annotation Pipeline'. These will be homologues from speciation, and most likely the genes carry out the same functions.

However, it's important to note that we cannot definitively say that two genes are orthologues unless we can present the common ancestral gene that they both evolved from, as was said in the lectures .

Question 2

Creating a BLAST tree

To create a BLAST tree, I clicked 'distance of results' at the top of the page displaying the results of a similar BLAST search from question 1 using [10](except I used 100 orthologues not 50 in order to get more of a feel of the tree). Note that this is not the entire tree, as I restricted the number of orthologues to 100 in my BLAST search to make it manageable. The tree is below:



This tree shows that there are closely related genes in primates, or at least their alignments are most similar to the ACVR1 gene in humans. The tree shows only vertebrates, or animals with exoskeletons, which is what is expected since the ACVR1 gene's "helps to control the growth and development of the bones and muscles" [8].

Are there orthologues in the house mouse, fruit fly or yeast?

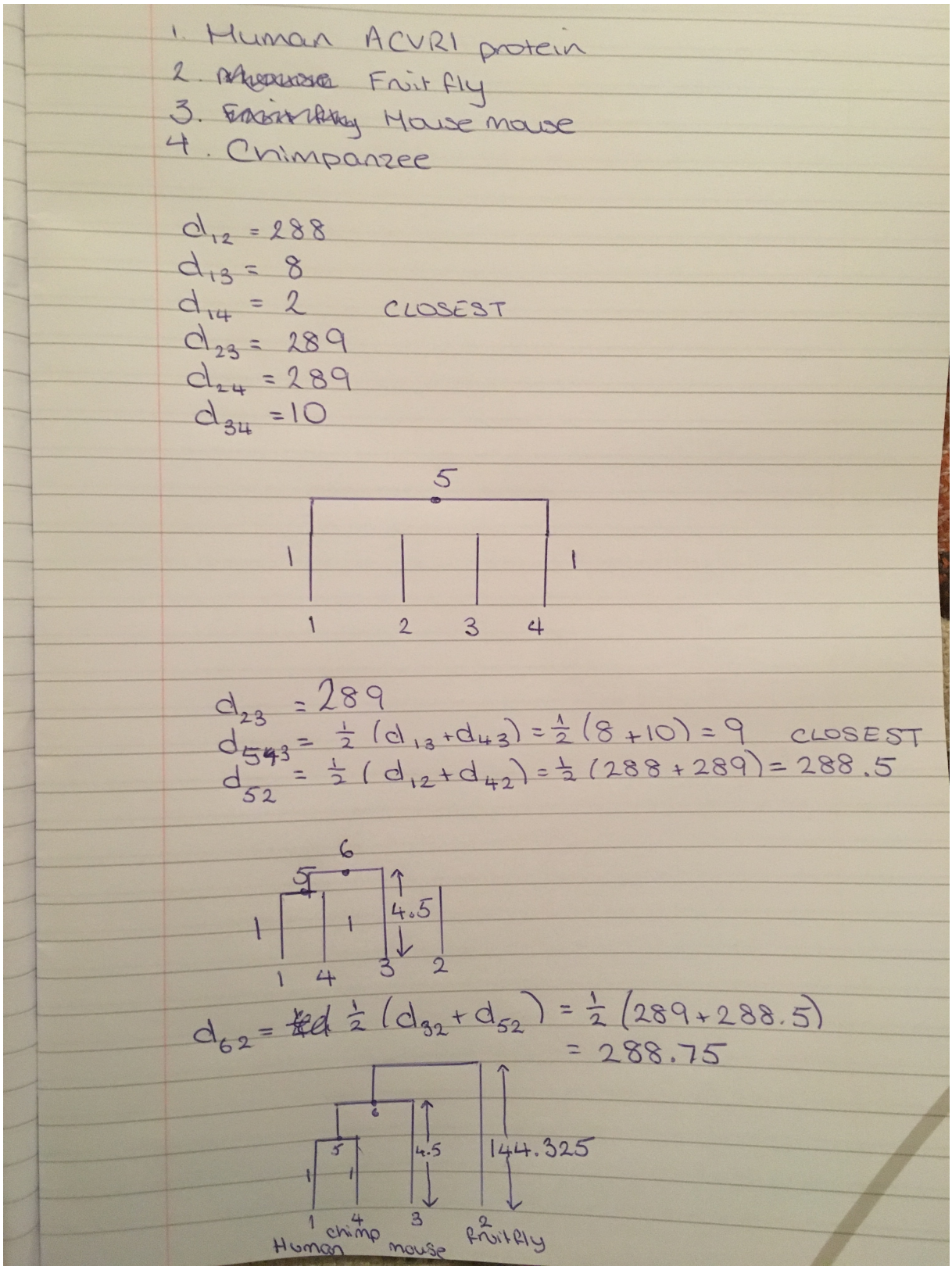
From the above investigation, I concluded that there were orthologues in the house mouse and fruit fly. When I repeated the BLAST protein search and included yeast, the yeast gene was not returned to me as having a significant alignment, which means it did not have an e value lower than 10^{-10}. Therefore I conclude there is no orthologue in yeast. This would make sense since the ACVR1 gene "helps to control the growth and development of the bones and muscles, including the gradual replacement of cartilage by bone (ossification)." [8], and yeast has no bones.

Is it feasible to study the gene in one of these orthologues?

It is feasible. "Activin receptor type I is found in many tissues of the body including skeletal muscle and cartilage" [8], and since a mouse has these things, it is a good candidate for modelling the disease, as was done in [9].

Creating a rooted phylogenetic tree using UPGMA algorithm

I constructed this tree using the protein sequences of the genes and the algorithms shown in class. I obtained the protein sequences from clicking on the protein accesion numbers beginning with NP from each of the genes' NCBI page [5] and displaying them as fasta files. To get the initial differences between the sequences, I copied and pasted the sequences in turn into [**?**]. Then I constructed the tree by hand. My working and final tree are below.



This tree mimics the one seen in question 2 (although a lot more compact since there's only 4 total orthologues tested), in the sense that here, the chimpanzee is closest to the human gene, and in question 2, primates were closest. Next closest in this example was the house mouse, and in question 2, rodents were further away than primates to the human gene. The furthest away here was the fruit fly, which wasn't even present in the tree in question 2 as I only selected the top 100 orthologues, and the fruit fly was not present in that query.

References

[1] Chris Lewis at the U of S. (n.d.). Retrieved November 18, 2017, from http://homepage.usask.ca/ ctl271/857/def_homolog.shtml

[2] Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. Current protocols in bioinformatics, 3-1.

[3] HomoloGene - NCBI. (n.d.). Retrieved November 18, 2017, from https://www.ncbi.nlm.nih.gov/homologene/?term=ACVR1

[4] Common chimpanzee. (2017, November 17). Retrieved November 18, 2017, from https://en.wikipedia.org/wiki/Common_chimpanzee

[5] ACVR1 activin A receptor type 1 [Homo sapiens (human)] - Gene - NCBI. (n.d.). Retrieved October 24, 2017, from https://www.ncbi.nlm.nih.gov/gene/90

[6] House mouse. (2017, October 25). Retrieved October 26, 2017, from https://en.wikipedia.org/wiki/House_mouse

[7] Drosophila melanogaster. (2017, October 22). Retrieved October 26, 2017, from https://en.wikipedia.org/wiki/Drosophila_melanogaster

[8] ACVR1 gene - Genetics Home Reference. (n.d.). Retrieved November 22, 2017, from https://ghr.nlm.nih.gov/gene/ACVR1

[9] Chakkalakal, S. A., Zhang, D., Culbert, A. L., Convente, M. R., Caron, R. J., Wright, A. C., ... & Shore, E. M. (2012). An Acvr1 R206H knock-in mouse has fibrodysplasia ossificans progressiva. Journal of bone and mineral research, 27(8), 1746-1756.

[10] BLAST: Basic Local Alignment Search Tool. (n.d.). Retrieved November 22, 2017, from https://blast.ncbi.nlm.nih.gov/Blast.cgi

[11] Member, T. P. (n.d.). Levenshtein Distance. Retrieved November 22, 2017, from https://planetcalc.com/1721/

Appendix

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession
activin receptor type-1 precursor [Homo sapiens]	1063	1063	100%	0.0	100%	NP_001096.1
PREDICTED: activin receptor type-1 [Rhinopithecus roxellana]	1062	1062	100%	0.0	99%	XP_010351652.1
PREDICTED: activin receptor type-1 [Saimiri boliviensis boliviensis]	1062	1062	100%	0.0	99%	XP_003922005.1
activin receptor type-1 precursor [Macaca mulatta]	1061	1061	100%	0.0	99%	NP_001247690.1
activin receptor type-1 isoform X2 [Sus scrofa]	1061	1061	100%	0.0	99%	XP_001927833.3
PREDICTED: activin receptor type-1 [Vicugna pacos]	1061	1061	100%	0.0	99%	XP_006196248.1
PREDICTED: activin receptor type-1 [Pan troglodytes]	1061	1061	100%	0.0	99%	XP_009441795.1
PREDICTED: activin receptor type-1 [Camelus dromedarius]	1060	1060	100%	0.0	99%	XP_010983904.1
PREDICTED: activin receptor type-1 isoform X1 [Miniopterus natalensis]	1060	1060	100%	0.0	99%	XP_016058436.1
PREDICTED: activin receptor type-1 [Condylura cristata]	1060	1060	100%	0.0	99%	XP_004674692.1
activin receptor type-1 [Delphinapterus leucas]	1059	1059	100%	0.0	99%	XP_022425478.1
PREDICTED: activin receptor type-1 [Lipotes vexillifer]	1059	1059	100%	0.0	99%	XP_007446510.1
PREDICTED: activin receptor type-1 [Balaenoptera acutorostrata scammoni]	1059	1059	100%	0.0	99%	XP_007183175.1
PREDICTED: activin receptor type-1 [Tupaia chinensis]	1059	1059	100%	0.0	99%	XP_006166312.1
PREDICTED: activin receptor type-1 [Camelus ferus]	1059	1059	100%	0.0	99%	XP_006176838.1
PREDICTED: activin receptor type-1 [Ovis aries]	1059	1059	100%	0.0	99%	XP_004004730.1
PREDICTED: activin receptor type-1 [Nannospalax galili]	1059	1059	100%	0.0	99%	XP_008839969.1
PREDICTED: activin receptor type-1 [Odobenus rosmarus divergens]	1059	1059	100%	0.0	99%	XP_004394876.1
PREDICTED: activin receptor type-1 [Panthera pardus]	1058	1058	100%	0.0	99%	XP_019312432.1
PREDICTED: activin receptor type-1 isoform X2 [Equus caballus]	1058	1058	100%	0.0	99%	XP_005601558.1
PREDICTED: activin receptor type-1 [Trichechus manatus latirostris]	1058	1058	100%	0.0	99%	XP_004375461.1
activin receptor type-1 [Odocoileus virginianus texanus]	1058	1058	100%	0.0	99%	XP_020756072.1
PREDICTED: activin receptor type-1 [Leptonychotes weddellii]	1058	1058	100%	0.0	99%	XP_006734908.1
PREDICTED: activin receptor type-1 [Orcinus orca]	1058	1058	100%	0.0	99%	XP_004276837.1
activin receptor type-1 isoform X1 [Carlito syrichta]	1057	1057	100%	0.0	99%	XP_008054349.1
PREDICTED: activin receptor type-1 isoform X1 [Myotis lucifugus]	1057	1057	100%	0.0	99%	XP_006091305.1
PREDICTED: activin receptor type-1 isoform X1 [Callithrix jacchus]	1057	1057	100%	0.0	99%	XP_002749466.1
activin receptor type-1 precursor [Bos taurus]	1057	1057	100%	0.0	99%	NP_788836.2
activin receptor type-1 [Canis lupus familiaris]	1057	1057	100%	0.0	99%	XP_005640253.1
PREDICTED: activin receptor type-1 isoform X1 [Loxodonta africana]	1056	1056	100%	0.0	99%	XP_010584602.1
activin receptor type-1 [Ictidomys tridecemlineatus]	1056	1056	100%	0.0	99%	XP_005315836.1
PREDICTED: activin receptor type-1 [Hipposideros armiger]	1055	1055	100%	0.0	99%	XP_019492408.1
activin receptor type-1 isoform X1 [Castor canadensis]	1055	1055	100%	0.0	99%	XP_020023116.1
PREDICTED: activin receptor type-1 [Bos mutus]	1055	1055	100%	0.0	99%	XP_005906436.1
activin receptor type-1 [Meriones unguiculatus]	1054	1054	100%	0.0	99%	XP_021494426.1
PREDICTED: activin receptor type-1 isoform X1 [Cebus capucinus imitator]	1054	1054	100%	0.0	99%	XP_017402735.1
PREDICTED: activin receptor type-1 [Otolemur garnettii]	1054	1054	100%	0.0	99%	XP_003796009.1
PREDICTED: activin receptor type-1 [Rousettus aegyptiacus]	1053	1053	100%	0.0	99%	XP_016005598.1
PREDICTED: activin receptor type-1 [Sorex araneus]	1053	1053	100%	0.0	99%	XP_004608169.1
PREDICTED: activin receptor type-1 [Ceratotherium simum simum]	1053	1053	100%	0.0	99%	XP_004428350.1
PREDICTED: activin receptor type-1 isoform X2 [Pteropus alecto]	1053	1053	100%	0.0	99%	XP_006921158.1
activin receptor type-1 [Heterocephalus glaber]	1053	1053	100%	0.0	99%	XP_021121722.1
PREDICTED: activin receptor type-1 [Fukomys damarensis]	1052	1052	100%	0.0	99%	XP_010605080.1
PREDICTED: activin receptor type-1 [Oryctolagus cuniculus]	1052	1052	100%	0.0	99%	XP_002712242.1
PREDICTED: activin receptor type-1 [Elephantulus edwardii]	1052	1052	100%	0.0	99%	XP_006880855.1
PREDICTED: activin receptor type-1 [Propithecus coquereli]	1052	1052	100%	0.0	99%	XP_012496237.1
activin receptor type-1 isoform X2 [Microcebus murinus]	1051	1051	100%	0.0	99%	XP_012600715.1
PREDICTED: activin receptor type-1 [Octodon degus]	1051	1051	100%	0.0	99%	XP_004629962.1
PREDICTED: activin receptor type-1 isoform X1 [Pteropus alecto]	1050	1050	100%	0.0	99%	XP_015452268.1
PREDICTED: activin receptor type-1 [Jaculus jaculus]	1050	1050	100%	0.0	99%	XP_004651917.1
activin receptor type-1 precursor [Mus musculus]	1050	1050	100%	0.0	98%	NP_031420.2
PREDICTED: activin receptor type-1 isoform X1 [Equus przewalskii]	1050	1050	100%	0.0	98%	XP_008523662.1
activin receptor type-1 [Mus caroli]	1049	1049	100%	0.0	98%	XP_021005492.1
PREDICTED: activin receptor type-1 [Erinaceus europaeus]	1049	1049	100%	0.0	98%	XP_007524357.1
PREDICTED: activin receptor type-1 [Pan paniscus]	1049	1049	100%	0.0	99%	XP_008956376.1
PREDICTED: activin receptor type-1 isoform X1 [Dipodomys ordii]	1048	1048	100%	0.0	98%	XP_012885892.1
PREDICTED: activin receptor type-1 [Eptesicus fuscus]	1048	1048	100%	0.0	99%	XP_008136890.1
PREDICTED: activin receptor type-1 [Peromyscus maniculatus bairdii]	1048	1048	100%	0.0	98%	XP_006984711.1
PREDICTED: activin receptor type-1 isoform X3 [Chinchilla lanigera]	1048	1048	100%	0.0	98%	XP_005393460.1
PREDICTED: activin receptor type-1 [Cavia porcellus]	1048	1048	100%	0.0	99%	XP_003478621.2
PREDICTED: activin receptor type-1 [Manis javanica]	1048	1048	100%	0.0	98%	XP_017508310.1
PREDICTED: activin receptor type-1 [Ochotona princeps]	1047	1047	100%	0.0	98%	XP_004577203.1

PREDICTED: activin receptor type-1 isoform X2 [Chinchilla lanigera]	1047	1047	100%	0.0	98%	XP_005393456.1
activin receptor type-1 isoform X1 [Mus pahari]	1046	1046	100%	0.0	98%	XP_021048150.1
PREDICTED: activin receptor type-1 [Microtus ochrogaster]	1045	1045	100%	0.0	98%	XP_005346487.1
PREDICTED: activin receptor type-1 isoform X1 [Chinchilla lanigera]	1045	1045	100%	0.0	98%	XP_005393455.1
PREDICTED: activin receptor type-1 [Cricetulus griseus]	1044	1044	100%	0.0	98%	XP_003503456.1
PREDICTED: activin receptor type-1 [Myotis davidii]	1043	1043	100%	0.0	98%	XP_006767448.1
activin receptor type-1 precursor [Rattus norvegicus]	1042	1042	100%	0.0	97%	NP_077812.1
PREDICTED: activin receptor type-1 [Chrysochloris asiatica]	1041	1041	100%	0.0	98%	XP_006872097.1
PREDICTED: activin receptor type-1 isoform X1 [Rattus norvegicus]	1031	1031	100%	0.0	94%	XP_008760086.1
PREDICTED: activin receptor type-1 [Dasypus novemcinctus]	1023	1023	100%	0.0	95%	XP_004449191.1
PREDICTED: activin receptor type-1 isoform X1 [Bos taurus]	1018	1018	96%	0.0	99%	XP_010800264.1
PREDICTED: activin receptor type-1 isoform X2 [Loxodonta africana]	1018	1018	95%	0.0	99%	XP_010584605.1
PREDICTED: activin receptor type-1 isoform X2 [Cebus capucinus imitator]	1012	1012	95%	0.0	99%	XP_017402760.1
PREDICTED: activin receptor type-1 isoform X2 [Dipodomys ordii]	1009	1009	95%	0.0	99%	XP_012885919.1
PREDICTED: activin receptor type-1 isoform X2 [Rattus norvegicus]	1008	1008	96%	0.0	98%	XP_008760089.1
PREDICTED: activin receptor type-1 isoform X3 [Equus przewalskii]	995	995	93%	0.0	99%	XP_008523665.1
PREDICTED: activin receptor type-1 isoform X3 [Loxodonta africana]	995	995	93%	0.0	99%	XP_010584606.1
PREDICTED: activin receptor type-1 isoform X2 [Callithrix jacchus]	994	994	93%	0.0	99%	XP_017828958.1
PREDICTED: activin receptor type-1 isoform X2 [Tursiops truncatus]	994	994	93%	0.0	99%	XP_019778665.1
PREDICTED: activin receptor type-1 isoform X2 [Myotis lucifugus]	994	994	93%	0.0	99%	XP_014313473.1
activin receptor type-1 isoform X3 [Sus scrofa]	973	973	96%	0.0	95%	XP_020932110.1
activin receptor type-1 isoform X1 [Sus scrofa]	972	972	96%	0.0	95%	XP_020932107.1
activin receptor type-1 [Phascolarctos cinereus]	961	961	100%	0.0	90%	XP_020860555.1
PREDICTED: activin receptor type-1 [Sarcophilus harrisii]	958	958	100%	0.0	89%	XP_003763925.1
PREDICTED: activin receptor type-1 [Monodelphis domestica]	957	957	100%	0.0	89%	XP_007494258.1
activin receptor type-1 isoform X2 [Enhydra lutris kenyoni]	955	955	100%	0.0	92%	XP_022373891.1
PREDICTED: activin receptor type-1 isoform X2 [Ornithorhynchus anatinus]	948	948	100%	0.0	88%	XP_001509625.1
PREDICTED: activin receptor type-1 isoform X1 [Ornithorhynchus anatinus]	940	940	100%	0.0	86%	XP_007669326.1
PREDICTED: activin receptor type-1 isoform X2 [Gavialis gangeticus]	914	914	100%	0.0	86%	XP_019377919.1
PREDICTED: activin receptor type-1 isoform X3 [Gavialis gangeticus]	914	914	100%	0.0	86%	XP_019377920.1
PREDICTED: activin receptor type-1 isoform X1 [Gavialis gangeticus]	913	913	100%	0.0	86%	XP_019377918.1
PREDICTED: activin receptor type-1 [Tauraco erythrolophus]	913	913	100%	0.0	86%	XP_009977967.1
PREDICTED: activin receptor type-1 isoform X1 [Crocodylus porosus]	912	912	100%	0.0	85%	XP_019409683.1
PREDICTED: activin receptor type-1 isoform X3 [Crocodylus porosus]	912	912	100%	0.0	85%	XP_019409685.1
PREDICTED: activin receptor type-1 isoform X2 [Crocodylus porosus]	912	912	100%	0.0	85%	XP_019409684.1
PREDICTED: activin receptor type-1 isoform X1 [Chrysemys picta bellii]	912	912	100%	0.0	86%	XP_005296984.1
PREDICTED: activin receptor type-1 [Chelonia mydas]	911	911	100%	0.0	86%	XP_007056728.1
PREDICTED: activin receptor type-1 isoform X2 [Chrysemys picta bellii]	911	911	100%	0.0	86%	XP_005296985.1