

Answers to assignment 1

Some blast results are shown as well as the sequence alignment with the new sequence added. You may have found different accession numbers for the sequences.

sequence 52A

XM_064892 mRNA
MHDNNLEHLPSQQ

The following seq is from HTGS

>gi|7653447|gb|AC016331.2|AC016331 Homo sapiens clone RP11-30F17, WORKING DRAFT SEQUENCE, 22 unordered pieces
Length = 152195

Score = 103 bits (256), Expect = 2e-21
Identities = 54/55 (98%), Positives = 54/55 (98%)
Frame = +2

Note the Q at the start is an AG intron boundary CAG.

Query: 13 QVLDTGQQLMVPVEVLEVDNKEVLWKFLLSGAMAGAVSRTGTAPLDRKQVYMQV 67
QVLDTGQQLMVPVEVLEVDNKEVLWKFLLSGAMAGAVSRTGTAPLDRKQVYMQV
Sbjct: 127325 QVLDTGQQLMVPVEVLEVDNKEVLWKFLLSGAMAGAVSRTGTAPLDRKQVYMQV 127486

Score = 63.5 bits (153), Expect = 2e-09
Identities = 33/54 (61%), Positives = 40/54 (73%)
Frame = +1

Query: 14 VLDTGQQLMVPVEVLEVDNKEVLWKFLLSGAMAGAVSRTGTAPLDRKQVYMQV 67
VLD GE L VP E + + + WK L++GA+AGAVSRTGTAPLDR KV+MQV
Sbjct: 80917 VLDIGECLTVPDEFKQEKLTGMWWKQLVAGAVAGAVSRTGTAPLDRKQVYMQV 81078

Score = 34.3 bits (77), Expect = 1.3
Identities = 14/14 (100%), Positives = 14/14 (100%)
Frame = +1

Query: 1 MHDNNLEHLPSQQV 14
MHDNNLEHLPSQQV
Sbjct: 126007 MHDNNLEHLPSQQV 126048 note this is an exon 1283bp upstream
The V at the end is the GT boundary. (all valine codons start with GT)

The phase 0 introns are very common and the Q V pair of aa is often seen
At these boundaries. The mRNA sequence supports the intron junction.

52	HUMAN	AC016331.2	GENE1	LHLENVEDVLYFWKHSTVLDIGECLTVPDEFKQEKLTG	0
52A	HUMAN	AC016331.2	gene2	-----MHDNNLEHLPSQQVLDTGQQLMVPVEVLEVDNKEV	0
52B	HUMAN	AL359258.4		LHPATNITEMIHFWKHSTLIDIGEISAIPDEFTEQEKQSG	0

This sequence may be the ortholog of mouse 52C that is also missing the N-terminal

Sequence 52C N-term

This seq from HTGS
>gi|9964809|gb|AC073721.2|AC073721 Mus musculus clone RP23-209N12, WORKING DRAFT SEQUENCE, 14 ordered pieces
Length = 190577

Score = 120 bits (302), Expect = 5e-27
Identities = 71/138 (51%), Positives = 84/138 (60%), Gaps = 29/138 (21%)

Frame = -3
This is seq 54
Query: 8 HLPSQQVLDLTGEQLMVPVEVLEVDNKEVLW-KFLLSGAMAGAVSRTGTAPLDRKVMQV 66
H P QVLD GE L VP E + + +W K L++GA+AGAVSRTGTAPLDR KV+MQV
Sbjct: 69135 HSPCSQVLDIGECLTVPDEFSQEEKLTGMWWKQLVAGAVAGAVSRTGTAPLDRKVMQV 68956
Query: 67 -----YSSKTNFTNLLSGLRTLQEGGVRSLWRGNGI 98
++SK+N N+L GLR ++QEGGV SLWRGNGI
Sbjct: 68955 RGPEGTHPII*AQT*EGTTKPPSLPSQVHASKSNRLNILGGLRNMIQEGGVLSLWRGNGI 68776
Query: 99 NVLKIAPEYAIKFSVCEQ 116
NVLKIAPE AIKF EQ
Sbjct: 68775 NVLKIAPESAIKFMAYEQ 68722

Score = 101 bits (252), Expect(3) = 6e-32
Identities = 49/52 (94%), Positives = 51/52 (97%)
Frame = -1
This is seq 52C
Query: 65 QVYSSKTNFTNLLSGLRTLQEGGVRSLWRGNGINVLKIAPEYAIKFSVCEQ 116
QVYSSK+NF NLLSGLR+LVQEGGVRSLWRGNGINVLKIAPEYAIKFSVCEQ
Sbjct: 53711 QVYSSKSNFRNLLSGLRSLVQEGGVRSLWRGNGINVLKIAPEYAIKFSVCEQ 53556

Score = 97.4 bits (241), Expect = 5e-20
Identities = 48/54 (88%), Positives = 51/54 (93%)
Frame = -3
This is seq 52C
Query: 13 QVLDLTGEQLMVPVEVLEVDNKEVLWKFLLSGAMAGAVSRTGTAPLDRKVMQV 66
QVLDLTGEQLMVPV+VLE +NK LWKFLLSGAMAGAVSRTGTAPLDRA+VYMQV
Sbjct: 55002 QVLDLTGEQLMVPVDVLEENKGTWKFLLSGAMAGAVSRTGTAPLDRARVYMQV 54841

Score = 43.1 bits (100), Expect(3) = 6e-32
Identities = 22/24 (91%), Positives = 22/24 (91%)
Frame = -2
This is seq 52C
Query: 116 QSKNFXXFYGVHSSQLFQERVVAG 139
QSKNF FYGVHSSQLFQERVVAG
Sbjct: 53461 QSKNF--FYGVHSSQLFQERVVAG 53396

Score = 34.3 bits (77), Expect = 0.56
Identities = 17/23 (73%), Positives = 19/23 (81%)
Frame = -3
This is seq 54
Query: 133 QERVVAGSLAVAVSQT LINPMEV 155
QER VAGSLA A +QT+I PMEV
Sbjct: 68481 QERFVAGSLAGATAQTIIYPMEV 68413

Score = 33.5 bits (75), Expect(3) = 6e-32
Identities = 16/16 (100%), Positives = 16/16 (100%)
Frame = -3
This is seq 52C
Query: 139 GSLAVAVSQT LINPME 154
GSLAVAVSQT LINPME
Sbjct: 53397 GSLAVAVSQT LINPME 53350

This is a second blast with the N-terminal of 52B as query expect =100
Score = 28.9 bits (63), Expect = 13
Identities = 9/18 (50%), Positives = 14/18 (77%)
Frame = -1
This is the probable N-terminal exon of 54 since the numbering is close to the next exon

Query: 1 LHPATNITEMIHFVKHST 18
LH N+ ++++FWKHST
Sbjct: 70502 LHSLENVEDVLYFWKHST 70449 first exon 1332bp upstream of exon 2

```
52A HUMAN AC016331.2 gene2 -----MHDNNLEHLPSQQVLDTGELMVPVEVLEVDNKEV 0
52B HUMAN AL359258.4 LHPATNITEMIHFVKHSTLIDIGEISAIPDEFTEQEKQSG 0
52C MOUSE AA492727 -----VLDTGELMVPVDVLEENKGT 0
54 MOUSE W50529 LHSLENVEDVLYFWKHSTVLDIGECLTVPDEFSEQEKLTV 0

52A L-WKFLLSGAMAGAVSRTGTAPLDRKLVMMQVY-----SSKTNF-TNLLGGLQSMVQE----GGFR 0
52B DWWKRLVSAGIASAVARTCTAPLDRKLVMMQVH-----SLKSRK-MRLISGLEQLVKE----GGIF 0
52C L-WKFLLSGAMAGAVSRTGTAPLDRARVYMQVY-----SSKSNF-RNLLSGLRSLVQE----GGVR 0
54 MWWKQLVAGAVAGAVSRTGTAPLDRKLVMMQVH-----ASKSNR-LNILGGLRNMIQE----GGVL 0
```

cannot find first exon of 52C by Blast of two alternative exons from 52A or 52B even at expect 1000.

Bos Taurus (cow) EST AW655355 probable ortholog of mouse 54
AEKILHSMRDGTMIDWQEWDRDHFLLHSLENVEDVLYFWKHSTVLDIGECLTVPDEFSEQEKLTGMWWKQLVAGAVAG
GLCPEQAQPLWTASRSSCRSTPLRPTG*TSWGASGA*SKRGACTPCGVATGLMCSRHLHSRQSSSWPMSRSSGSPSGN
RHCRCRSALWLP

52C Cterm

hybrid with 52A

KTRLTLRFTGQYKGLLDCARQILERDGTALYRGYLPNMLGIIPYACTDLAVYELL
KNFWLKSGRDMGDPGLVSLSSVTLSTTCGQMASYPLTLVTRMQAQTVEGSNPTMR
GVLQRILAQQGWGLYRGMPTLLKVLPAAGGISYVVYEAMKKTGLGItwk

KTRLTLRFTGQYKGLLDCARQILERDGTALYRGYLPNMLGIIPYACTDLAVYELL
QCLWQKLGDRMxxxxxxx
VSLSSVTLSTTCGQMASYPLTLVTRTXMQAQTVEGSNPTMQGVFKRILSQQGWPGLYRGM
TPTLLKVLPAAGGISYLVYEAMKKTGLGVQVLSR*

```
52B KTRLAIGKT-----GEYS-GIIDCGKLLKQEG-VRSFFKGYTPNLLGIVPYAGIDLAVYEILKNFWL 0
52C KTRLTLRFT-----GQYK-GLLDCARQILERDG-TRALYRGYLPNMLGIIPYACTDLAVYELLQCLWQ 0
54 KTRLTLRRT-----GQYK-GLLDCANGLLEREG-PRAFYRGY-PNVLGIIPYAGIDLAVYETLKNRWL 0

52B -----ENYAGNSVNPIMILVGCSTLSNTCGQLASFSVNLIRTRMQASAPVEKKGK-----TTSMI 0
52C -----KLGRDMXXXXXXVSLSSVTLSTTCGQMASYPLTLVTRTXMQAQTVEGS-----NPTMQ 0
54 -----QQYSHESANPGILVLLACGTISSTCGQIASYPLALVTRTRMQAASIEGG-----PQVSMV 0

52B QLIQEITYTEKGLGFYR--GFTSNIKVLPAVGVCVAYEKVKPLFGLTWK*----- 0
52C QGVFKRILSQQGWPGLY--RGMTPTLLKVLPAAGGISYLVYEAMKKTGLGVQVLSR*----- 0
54 GLLRHILSQEGVWGLYR--GIAPNFMKVIPAVSISYVVYENMKQALGVTSGEGLG*----- 0
```

>gi|4404988|gb|AI507137.1|AI507137 vi78c06.x1 Stratagene mouse testis (#937308)
Mus musculus cDNA
clone IMAGE:918346 3' similar to TR:O18757 O18757
PEROXISOMAL CA-DEPENDENT SOLUTE CARRIER. ;.
Length = 492

Score = 160 bits (405), Expect = 3e-39
Identities = 79/87 (90%), Positives = 84/87 (95%)
Frame = -1

Query: 74 VSLSSVTLSTTCGQMASYPLTLVTRTRMQAQTVEGSNPTMRGVLQRILAQQGWGLYRGM 133
VSLSSVTLSTTCGQMASYPLTLVTRTRMQAQTVEGSNPTM+GV +RIL+QQGW GLYRGM
Sbjct: 492 VSLSSVTLSTTCGQMASYPLTLVTRTXMQAQTVEGSNPTMQGVFKRILSQQGWPGLYRGM 313

Query: 134 TPTLLKVLPAAGGISYVVYEAMKKTGLGI 160
TPTLLKVLPAAGGISY+VYEAMKKTGLG+

Sbjct: 312 TPTLLKVLPAAGGISYLVYEAMKKTGLV 232

Seq 56 Cterm part

>gi|11656379|gb|BF582661.1|BF582661 602094056F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4208509 5'. Length = 930

Query: 1 DNFADKSVNPGVMVLLGCGALSSTCGQLASYPLALVRTRMQAQAMLEGSPQLNMVGLF-R 59 DNFADKSVNPGVMVLL CGALSSTCGQLASYPLALVRTRMQAQA +EG+PQ F Sbjct: 384 DNFADKSVNPGVMVLLSCGALSSTCGQLASYPLALVRTRMQAQATVEGAPQTEHGWPFS 563

Query: 60 RIISKEGIPGLYRGITPNFMKVLPAVGISYV-VYENMKQTLG 100 + + + L R PN KVLP G V ++MKQ G Sbjct: 564 NRLQRRSVRTLQRH-HPNS*KVLPGRGHQLCGVMKHKQNP 686

There appears to be several frameshifts in this seq

Query: 51 QLNMVGLFRRIISKEGIPGLYRGITPNFMKVLPAVGISYVVYENM-KQTLG 100 +L+MVGLF+RI+SKEG+ GLYRGIT K PAVGISYVV N+ +TLG Sbjct: 535 RLSMVGLFQRIVSKEGVSGLYRGITQIHEKCSPAVGISYVV**NI*SR 687

Probable sequence after correcting for frameshifts

DNFAKDSVNPGVMVLLSCGALSSTCGQLASYPLALVRTRMQAQATVEGAPQ LSMVGLFQRIVSKEGVSGLYRGITPNXXXKVLPAVGISYVVXXXMKQTLG

>gi|11652199|gb|BF578487.1|BF578487 602092909F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4207010 5'. Length = 911

Score = 100 bits (250), Expect = 2e-21 Identities = 48/54 (88%), Positives = 52/54 (95%) Frame = +3

Query: 51 QLNMVGLFRRIISKEGIPGLYRGITPNFMKVLPAVGISYVVYENMKQTLGVTQK 104 QL+MVGLF+RI+SKEG+ GLYRGITPNFMKVLPAVGISYVVYENMKQTLGV QK Sbjct: 3 QLSMVGLFQRIVSKEGVSGLYRGITPNFMKVLPAVGISYVVYENMKQTLGVAQK 164

Probable sequence from both ESTs

DNFAKDSVNPGVMVLLSCGALSSTCGQLASYPLALVRTRMQAQATVEGAPQLSMV GLFQRIVSKEGVSGLYRGITPNFMKVLPAVGISYVVYENMKQTLGVAQK*

Accession numbers BF578487 BF582661

54 -----QQYSHESANPGILVLLACGXISSTCGQIASYPLALVRTRMQAQASIEGG-----PQVSMV 0
54A -----DNFAKDSVNPGVLLGCGALSSTCGQLASYPLALVRTRMQAQAMLEGA-----PQLNMV 0
54B -----DKDGYNHPLTLLAAGAIAGVPAASLVTPADAIAKTRLQVVARSQGT-----TYTG 0
55 -----DNFAKDSVNPGVMVLLGCGALSSTCGQLASYPLALVRTRMQAQAMLEGS-----PQLNMV 0
56 -----DNFAKDSVNPGVMVLLSCGALSSTCGQLASYPLALVRTRMQAQATVEGA-----PQLSMV 0
54 GLLRHILSQEGLVWGLYR--GIAPNFMKVIPAVSISYVVYENMKQALGVTSGEGLG*----- 0
54A GLFRRIISKEGLPGLYR--GITPNFMKVLPAVGISYVVYENMKQTLGVTQK*----- 0
54B DATKKIMAEEGPRAFVK--GTAARVFRSSPQFGVTLVTYELLQRLFYVDFGGTQPKGSEAHKITTPLEQA 0
55 GLFRRIISKEGIPGLYR--GITPNFMKVLPAVGISYVVYENMKQTLGVTQK*----- 0
56 GLFQRIVSKEGVSGLYR--GITPNFMKVLPAVGISYVVYENMKQTLGVAQK*----- 0

56 N-term

Note: the previous N-term was wrong it belongs to seq 51.

>gi|14618687|gb|BI158686.1|BI158686 602921763F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5062095 5'. Length = 776

Score = 183 bits (464), Expect(2) = 4e-65
Identities = 89/93 (95%), Positives = 91/93 (97%)
Frame = +1

Query: 1 FNPVTDIEEIIIRFWKHSTGIDIGDSLTIPEFTEDEKKSQWWRQLLAGGIAGAVSRTST 60
FNPVTDIEEIIIRFWKHSTGIDIGDSLTIPEFTEDEKKSQWWRQLLAGG+AGAVSRTST
Sbjct: 85 FNPVTDIEEIIIRFWKHSTGIDIGDSLTIPEFTEDEKKSQWWRQLLAGGVAVSRTST 264

Query: 61 APLDRLKIMMQVHGSKSDKMNIFFGGFRQMVKEG 93
APLDRLK+MMQVHGSKS MNIFGGFRQMVKEG
Sbjct: 265 APLDRLKVMQVHGSKS--MNIFGGFRQMVKEG 357

Score = 83.6 bits (205), Expect(2) = 4e-65
Identities = 38/41 (92%), Positives = 40/41 (96%)
Frame = +2

Query: 91 KEGGIRSLWRGNGTNVIKIAPETAVKFWAYEQYKLLTEEG 131
++ GIRSLWRGNGTNVIKIAPETAVKFWAYEQYKLLTEEG
Sbjct: 350 RKGIRSLWRGNGTNVIKIAPETAVKFWAYEQYKLLTEEG 472

Accession number BI158686

55	HUMAN	AF123303	FNPVTDIEEIIIRFWKHSTGIDIGDSLTIPEFTEDEKKS	0
56	MOUSE	AA024208	FNPVTDIEEIIIRFWKHSTGIDIGDSLTIPEFTEDEKKS	0
56A	MOUSE	AA212724	LHLENVEDVLYFWKHSTVLDIGECLTVPDEFSQEKLTV	0
55	QWWRQLLAGGIAGAVSRTSTAPLDRLKIMMQVH	-----GSKSDKMNIFFGGFRQMVKE	----GGIR	0
56	QWWRQLLAGGVAVSRTSTAPLDRLKVMQVH	-----GSKS--MNIFGGFRQMVKE	----GGIR	0
56A	MWWKQLVAGAVAGAVSRTGTAPLDRLKVFQVH	-----ASKSNR-LNILGGLRNMIQE	----GGV-	0

seq 54 and 56A shown for comparison

54	MOUSE	W50529	LHLENVEDVLYFWKHSTVLDIGECLTVPDEFSQEKLTV	0
56A	MOUSE	AA212724	LHLENVEDVLYFWKHSTVLDIGECLTVPDEFSQEKLTV	0
54	MWWKQLVAGAVAGAVSRTGTAPLDRL	-----ASKSNR-LNILGGLRNMIQE	----GGVL	0
56A	MWWKQLVAGAVAGAVSRTGTAPLDRL	KVFQVH-----ASKSNR-LNILGGLRNMIQE	----GGVL	0

sequence 56A Note 56A and 54 seem to be the same gene.

this is the mouse ortholog of seq 52. They are almost 100% identical

accession numbers BI220653 BI145091

52	MWWKQLVAGAVAGAVSRTGTAPLDRLKVFQVH	-----ASKTNR-LNILGGLRSMVLE	----GGIR	0
52A	L-WKFLLSGAMAGAVSRTGTAPLDRAKVMQVY	-----SSKTNF-TNLLGGLQSMVQE	----GGFR	0
52B	DWVKRLVSAIASAVARTCTAPLDRLKVMQVH	-----SLKSRK-MRLISGLEQLVKE	----GGIF	0
52C	-----	-----LSGLRTLTVQE	----GGVR	0
54	-----	-----E	----GGVL	0
54A	QWWRQLLAGGIAGAVSRTSTAPLDRLKVMQVH	-----GSKS--MNIFGGFRQMIKE	----GGVR	0
54B	ESSYRFTLGSFAGAVAPTVPIDLVKTRMQNQ	----RAGSYIGEVAY-RNSWDCFKKVVRH	----EGFM	0
55	QWWRQLLAGGIAGAVSRTSTAPLDRLKIMMQVH	-----GSKSDKMNIFFGGFRQMVKE	----GGIR	0
56	-----	-----ADGKG	----RGIR	0
56A	MWWKQLVAGAVAGAVSRTGTAPLDRL	KVFQVH-----ASKSNR-LNILGGLRNMIQE	----GGVL	0
52	SLWRGNGINVLKIAPESAIKFMAYEQIKRAILGQOET	-----LHVQERFVAGSLAGATAQTIIYPMEVL	0	0
52A	SLWRGNGINVLKIAPEYAIKFSVFQCKNYFCGIQGS	-----PPFERLLAGSLAVAISQTLINPMEVL	0	0
52B	SLW*GNGVNLKIAPEATALKVGAYEQYKLLSFDGVH	-----LGILERFIFGSLAGVTAQTCIYPMEVL	0	0
52C	SLWRGNGINVLKIAPEYAIKFSVCEQSKNF	--FYGVH---SSQLFQERVVAGSLAVAVSQTIIYPMEVL	0	0
54	SLWRGNGINVLKIAPEYAIKFMAYEQIKRAIRGQQ	-----ETLHVQERFVAGSLAGATAQTIIYPMEVL	0	0
54A	SLWRGNGTNVIKIAPETAVKFWVYEQYKLLTEEG	-----QKIGTFERFISGSMAGATAQTFIYPMEVM	0	0
54B	GLYRGLLPQLMGVAPEKAIKLTVDLVRDKLTDKK	-----GNIPTWAEVLGGCAGASQVVFVNPLEIV	0	0
55	SLWRGNGTNVIKIAPETAVKFWAYEQYKLLTEEG	-----QKIGTFERFISGSMAGATAQTFIYPMEVM	0	0

```

56 SLWRGNGTNIKIAPETAVKFWAYEQYKLLTEEG-----QKLGTFERFISGSMAGATAQTFIYPMEVL 0
56A SLWRGNGINVLKIAPESAIKFMAYEQIKRAIRGQQ-----ETLHVQERFVAGSLAGATAQTI IYPMEVL 0

52 KTRLTLRRT-----GQYK-GLLDCARRILEREG-PRAFYRGYLPNVLGIIPYAGIDLAVYETLKNWWL 0
52A KTRLTLRRT-----GQYK-GLLDCARQILQREG-TRALYRGYLPNMLGIIPYACTDLAVYEMLQCFWV 0
52B KTRLAIGKT-----GEYS-GIIDCGKLLKQEG-VRSFFKGYTPNLLGIVPYAGIDLAVYEILKNFWL 0
52C KTRLTLRFT-----GQYK-GLLDCARQILERDG-TRALYRGYLPNMLGIIPYACTDLAVYELL----- 0
54 KTRLTLRRT-----GQYK-GLLDCANGLLEREG-PRAFYRGY-PNVLGIIPYAGIDLAVYETLKNRWL 0
54A KTRLAVGKT-----GQYS-GIYDCAKILKYEG-FGAFYKGYVPNLLGIIPYAGIDLAVYELLKSHWL 0
54B KIRLQVAGE-----IAS-GSKIRAWSVRELG-LFGLYKGARACLLRDVPFSAIYFPTYAHTKAMMA 0
55 KTRLAVGKT-----GQYS-GIYDCAKILKHEG-LGAFYKGYVPNLLGIIPYAGIDLAVYELLKSYWL 0
56 KTRLAVAKT-----GQYS-GIYGCAKILKHEG-FGAFYKGYIPNLLGIIPYAGIDLAVYELLKSYWL 0
56A KTRLTLRRT-----GQYK-GLLDCAKRILEREG-PRAFYRGYLPNVLGIIPYAGIDLAVYETLKNRWL 0

52 -----QQYSHDSADPGILVLLACGTISSTCGQIASYPLALVRTRMQAQASIEGGP-----QLSML 0
52A -----KSGRDMGDP SGLVSLSSVTLSTTCGMASYPLTLVRTRMQAQDTVEGS-----NPTMR 0
52B -----ENYAGNSVNP GIMILVGCSTLSNTCGQLASFSVNLIRTRM QASAPVEK GK-----TTSMI 0
52C ----- 0
54 -----QQYSHESANPGILVLLACGXISSTCGQIASYPLALVRTRMQAQASIEGG-----PQVSMV 0
54A -----DNFAKDSVNP GVLVLLGCGALSSTCGQLASYPLALVRTRMQAQAMLEGA-----PQLNMV 0
54B -----DKDGYNHPLTLLAAGAIAGVPAASLVTPADA I KTRLQVVARSGQT-----TYTGVW 0
55 -----DNFAKDSVNP GVMVLLGCGALSSTCGQLASYPLALVRTRMQAQAMLEGS-----PQLNMV 0
56 -----DNFAKDSVNP GVMVLLSCGALSSTC----- 0
56A -----QQYSHESGNPGILVLLACGTISSTCGQIASNPLALVRTxxxAQASIEGGP-----QVSMG 0

52 GLLRHILSQEGMRGLYR--GIAPNFMKVIPAVSISYVYENMKQALGVTSR*----- 0
52A GVLQRILAQQGWLGLYR--GMTPTLLKVLPAAGGISYVVEAMKKT LGI *----- 0
52B QLIQEIYTKEGKLG FYR--GFTSNI IKVLPVAVGVGC VAYEKVKPLFGLTWK*----- 0
52C ----- 0
54 GLLRHILSQEGVWGLYR--GIAPNFMKVIPAVSISYVYENMKQALGVTSGEGLG*----- 0
54A GLFRRIISKEGLPGLYR--GITPNFMKVLPAVGISYVYENMKQTLGVTQK*----- 0
54B DATKKIMAEEGPRAFVK--GTAARVFRSSPQFGVTLVTYELLQRLFYVDFGGTQPKGSEAHKITTPLEQA 0
55 GLFRRIISKEGIPGLYR--GITPNFMKVLPAVGISYVYENMKQTLGVTQK*----- 0
56 -----*----- 0
56A -----*----- 0

```

>gi|14674097|gb|BI220653.1|BI220653 602938749F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5101741

5'.
Length = 817

Score = 79.3 bits (194), Expect = 1e-14
Identities = 37/43 (86%), Positives = 42/43 (97%)
Frame = +2

Query: 13 GAVSRTGTAPLDRLRVFMQVHASKTNRLNILGGLRSMVLEGGI 55
GAVSRTGTAPLDRL+VFMQVHASK+NRLNILGGLR+M+ EGG+
Sbjct: 2 GAVSRTGTAPLDRLKVFMQVHASKSNRLNILGGLRNMIQEGGV 130

GAVSRTGTAPLDRLKVFMQVHASKSNRLNILGGLRNMIQEGGVLSLWRGNGINVLKIAPESAIKFMAYEQIKRAIRGQQ
ETLHVQERFVAGSLAGATAQTI IYPMEVLKTRLTLRRTGQYKGLLDCAKRILEREGPRAFYRGYLPNVLGIIPYAGIDL
AVYETLKNRWLQQYSHESGNPGILVLLGLWHHLQHLWPCQ*PSGNWSVPRMQGPSLH*GLGPQVSMGGVLLRQHSWSQ
EAWNGGPLTRGMGPKIYLKRHSSR*SIFQRWATEN

WRCVTDHRSSSGPTQGIHAGPRLKVQPA*HSRGPEKHDSRRGCLVPLAGQRHQRAQDSPRVCHQIHGL*TDQAGHPGAT
GDTARSGALRGWFPGRGHSSNHHIPHGKTSADSTQNWVQVQAPGLCKADLRT*RAPRLLPWLPA*CAGHHPLCRNRP
SCLRDPESLASAVQPRIR*PRHSRAPRLVAPSPAPVARLPTLWQLVVRTQNARPKPPLRVGVPAGLHGWPSTTFVLPV
GGVEWEGASYPGHGPQNLEASFQPLKHLPTVGYGK

ALCHGQAQLLWTDSTRYSCRSTPQSPGTGLTF*GA*ET*FKKGVSCPSGGATASCTCSR*PPSLPNSWLMNRSSGPGSGNR
RHCTFRSASWLVWPWPGPQLKPSYTPWRY*RLG*LYAELASTRGSWTVQSGS*NVKGPAPSTVATCLMCWASSPMQEST*
LSTRP*RIAGFSSTATNPVTQAFSCS*ACGTISSTCGQIASNPLATGPYPECKAQASIEGWARRSPWVGSYFDNIPGPR
RRGMGGLLPGAWAPKFT*SVI PAVEASSNGGLRK

>gi|14605092|gb|BI145091.1|BI145091 602909168F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5050408 5'.

Length = 741

Score = 101 bits (251), Expect(2) = 4e-40
Identities = 51/55 (92%), Positives = 53/55 (95%)
Frame = +1

Query: 50 PQVSMGGLLRHILSQEGMRGLYRGIAPNFMKVIPAVSISYVVYENMKQALGVTSR 104
P+VSM GLLRHILSQEG+ GLYRGIAPNFMKVIPAVSISYVVYENMKQALGVTSR
Sbjct: 289 PKVSMVGLLRHILSQEGVWGLYRGIAPNFMKVIPAVSISYVVYENMKQALGVTSR 453

Score = 82.8 bits (203), Expect(2) = 4e-40
Identities = 42/49 (85%), Positives = 43/49 (87%)
Frame = +3

Query: 1 QQYSHESGNPILVLLACGTISSTCGQIASNPLALVRTRMQAQASIEGG 49
QQYSHES NPGILVLL CGTISSTCGQIAS PLALVRTRMQAQ + GG
Sbjct: 165 QQYSHESANPILVLLGCGTISSTCGQIASYPLALVRTRMQAQ-GLHGG 308

Sequence 78B

>gi|14572586|emb|AL157871.5|CNS01RG9 Human chromosome 14 DNA sequence BAC R-638I2 of library RPCI-11 from

chromosome 14 of Homo sapiens (Human), complete sequence
Length = 180688

Query: 8 IGGVCGVAVGYPLDTVKVR 26
+ GVCVAVGYPLDTVKVR
Sbjct: 133753 LAGVCGVAVGYPLDTVKVR 133809

Query: 23 VKVRIQTEPKYTGIIWHCVRDITYHRERVWGLYKGLG 57
++VRIQTEPKYTGIIWHCVRDITYHRERV GL G G
Sbjct: 134123 LQVRIQTEPKYTGIIWHCVRDITYHRERV-GLGPGAG 134224

Query: 50 WGLYKGLGSPLMGLTFINALVFGVQGNTLRALGHDSPLNQFLAGAAPSLLQVFLTSPTEV 109
W +GLG+ +T N + G Q + PL PSLQVFLTSPTEV
Sbjct: 136566 WASLQGLGAPPEVTVSNMRLVGGVQAVGRATGG*PL-----LPSLLQVFLTSPTEV
136724

Query: 110 AKVRLQTQ-----MCPVPPACPEPKYRGLHCLATVAREEGLCGL 149
AKVRLQTQ MCPVPPACPEPKYRGLHCLATVAREEGLCGL
Sbjct: 136725 AKVRLQTQTQAQKQRRLSASGPLAVPPMCPVPPACPEPKYRGLHCLATVAREEGLCGL
136904

Query: 150 YKGSSALVLRDGHFSFATYFLSYAVLCEWLS 179
YKGSSALVLRDGHFSFATYFLSYAVLCEWLS
Sbjct: 136905 YKGSSALVLRDGHFSFATYFLSYAVLCEWLS 136994

FKSIIKQESVL
GLYKGLGSPLMGLTFINALVFGVQGNTLRALGHDSPLNQFLAGAAAGAIQCVICC

HLALRPGYVSPRARRSGARGWVGRQGGDAGPASAAQQVMPGSPGLSLSLTFHMGIVLPIISWVPLRSGARTSGS*RGLRR
RFPPPAEGLEVAFRSLKVGGGTHALHFCLLQARHPLKAVLGGVFRV*LRDLSGAVLGDILLRGLGAWCRPGASL*KQQW
QNMLLLVHCQARSGLRL*SQHFGRPKQADHLWSGVQDKPCQHGETPSLLKIQKLAGGGGRHL*SQWLRRLRQENRNLN
GCGGCSELRSCHCTPAWAKKSSSSSSWAGLGLVWLFRSPLVTSILQQRKLRP*GRE*PSRGPTAECGAGLSVGSGLG
WLLLLAGP*PDLDPVC*VWGFYRGLSLPVCVSLVSSVSFGTYRHCLAHCRLRYGNPDAKPTKADITLSGCASGLVRV
SRGSQGGEGPRETAGSERLSILAGGETEA**GQTEAMHGVRGCGRRTRIPRRLGLEPLSSTQMRGSLSCPARPPQEG
GCDQRDSPCHGQKVT*KPPWGAE*AGPGHSLQARRCLNTVGHCMWAFLLHPQLCMEHLRCPALREPLLPQGRQTFAL
GDSTEPAGAAQQLKAWPRGPDHQAQLSALTLMMS*ATLGKQRLLLESLHL*EA**EFSEIMRVKPPAHCIVRHE*VP
TKWQPVCLVIFFFFEETESCSVTQASV*WHHLRSLQPLRPRFKQFSCLSLLSSWDYRHAPPSPANFCIFSRDGVSPILA

RLVSN*SRDPPALTSQSAGITGVSRCAWPWYFIADMLLNTFNLGHGNYLTRPRRSLAGWPGRDVAAEEAWALALGG
RSWLWAPWPPSPGPACRVSGPNLLRSP*AT*GWWEARSRLWAEPQVADPCFLLCSRCS*RRPLRWPKSAC

IWHCVRDYTHRERVGLGPGAG*GDREGLGLPPLPSR*CQAPRWVSALSPTWE*FCQSHGFL*GLEPGHLAARGASEG
GSHLQQRV*R*RSGL*KSEEGPMHCISVSSRPGTP*RQCWVESLGSSETCPEQYLATFSYVDLGLGAGQGLVCESSSG
RTCCY*FIVRPGAVAYACNPSTLGGRSRQITCGQEFKTSLANMVKPRLY*KYKN*LEVAVGTCNRSGSGG*GRRIA*TW
DAEVAVS*DRATALQPGQKKVHRHQAGLWAWSSGSSGPPL*PPSYSRGN*GPEAVNDLPEVRQRSAEQG*VWAPSG
GSSWRGSLTWILSARCGASTGASRCPCARCPWYLPCLLAPTATAWRTSAGSGTATLTPSPRPTSRSRDAPPASSA*
VGAARVGAQERLRGLRGYPSWQVQGLRLDEGRHWRPCMESGAVGEEELASRLGWAWNPAAHR*GGRCCPAQHGPLRKG
AVTKGTLPVTVRRSPRSHHGELNRQAPGTHGSRHGDA*TQSAIAGCGHFSILNCAWSTCGARLPSGSPCSRGRDRLSHL
VTARSQLEQPSS*RPQPGVQTRPSSQL*PS*AEQPWASKGDSWFCLSICKRHENLVKS*G*SPQHTYVESDMNEY
QSGSQYV*FFFFFFLRQSLARLPRPVFSGTISAHYNYLVPSSNSPASAS*VAGTTGTRHQARLIFVFLVEMGFHPYWP
GWSQTPDLVIRLP*PPKVLGLQA*AAAPGHGIIILLICY*IRS*TWVMEIILHVPGEAWQAGLGEMWLPKHKHGLWHWVG
GAGFGPRGPPALGQPPAGSRGPTS*GHREQHEAGRPGPGCGQSHRWLTPASFSAAGVDPVAH*GGQSPL

SGTASGIRITESA*VWGQGLGRETGRGCWACLRCRCPAGDARLPAGSQLSHLPHGNSSANLMGSSEVWSQDIWQLEGPQKE
VPTSSRGRSGSVQVFESRRDPCTAFLSPPGQAPLEGSVGSWL*GLAQRVRSSTWRHSPTWTWGLVQARG*FVKAABA
EHAVTSSLSGQERWLTVPVIALWEAEAGRSPVVRSSRQALPTW*NPVSTKNTKISWRWWQAPVIAVAQEAEGESLEPG
MRRLQ*VKIVPLHSSLGKKKFKIVIKQLGWVGPGLALQVPPCDLHLTAETEALRP*MTFPRSDSGVRSRAECGLPRV
APPPGALA*PGSCLLVGLLPGPLAARVHGVPGIFRVFVHLLPPLPGAHLPAVRQP*RQAHQGRHHALGMRLRPRPRE
*GQPGWGRPKRDCGV*EVIHPGRWGN*GLMRADTGGHAWSQLWEKNHPA*AGPGTPQHTDEGVAAVLPSTAPSGRG
L*PKGLSLSRSEGHLEATMGS*IGRPRALMAPGTEPEHSRPLLDVGISSSTVHGAPAVPACPGALAPAGETDFRTW
*QH GASWSSPAAEGLAQGSRPPGALSSDPHELNLGQARETLGASVSPSVRGMRI**NHEGKAPSTHMWSPT*MSTH
KVAASMSDDFFFF*DRVLLGYPGQCLVAPSPLTSTSTSQVQAILLPQPE*LGLQARATKPG*FLYF**RWGFTHIGQ
AGLKLIS*SACLDLPCWDYRREPLRLAMVLYC*YVTKYVLEPGSWKLSYTSQEKPGRLAWERCGRGSMGFIGWA
ELALGPVAPQPWASLQGLGAQPPEVTVSNMRLVGGQVQAVGRATGG*PLLPSLLQVFLTSPTEVAKVRL

78A GLYKGLGSPLMGLTFINALVFGVQGNLRLALGHDS-----LNQFLAGAAAGAIQCVCIPMELA 0
78B GFYRGLSLPVCTVSLVSSVSFGTYRHCLAHCRLRYG-NPDAKPTKADITLSCASGLVRVFLTSPTEVA 0
78C GLYRGISSPMGGIGLVNAIVFGVYGNVQRLSNPNL-----TSHFFAGSIAGVAQGFVCAPMELA 0

Sequence 90

AK019150 BC013845 BC016932 AW825118 note this is a 100% match to human

89 HUMAN NM_003705 LADIERIAPLAEGALPYNLAELQRQ-QSPGLGRPIWLQIA 0
90 MOUSE W77458 LADIERIAPLAEGALPYNLAELQRQ-QSPGLGRPIWLQIA 0
90A DROSO AC020102 -----MAGQQHDISHAK 0
90B HUMAN AL079303 AL079304-----MSAKPEVSLVR 0
91 HUMAN AK000766.1 LADIERIAPLEEGTLPFNLAEAQRQKASGDSARPVLLQVA 0
92 MOUSE W76821 LADIERIAPLEEGMLPFNLAEARQKASGDAARPFLLQLA 0

Sequence 94B extension

>gi|12845460|dbj|AK010193.1|AK010193 Mus musculus adult male tongue cDNA, RIKEN
full-length enriched
library, clone:2310076D23, full insert sequence
Length = 2531

Score = 206 bits (523), Expect = 1e-53
Identities = 110/135 (81%), Positives = 113/135 (83%), Gaps = 20/135 (14%)
Frame = +3

Query: 1 KIQLQDAGRIAAQRKILA-----PRSTATQLTRDLLRSRGIAGLY 40
KIQLQDAGRIAAQRKILA PR TATQLTRDLLR+ GIAGLY
Sbjct: 588 KIQLQDAGRIAAQRKILAAQAQLSAQGAQPSVEAPAPPRPTATQLTRDLLRNHGIAGLY 767

Query: 41 KGLGATLLRDVPFVSFVYFPLFANLNQLGRPASEEKSPFYVSFLAGCVAGSAAAVAVNPCD 100
KGLGATLLRDVPFVS+VYFPLFANLNQLGRP+SEEKSPFYVSFLAGCVAGSAAAVAVNPCD
Sbjct: 768 KGLGATLLRDVPFVSIVYFPLFANLNQLGRPSSEEKSPFYVSFLAGCVAGSAAAVAVNPCD 947

Query: 101 VVKTRLQSLERGVNE 115
VVKTRLQSLERGVNE

Sbjct: 948 VVKTRLQSLERGVNE 992

MADKQI
SLPAKLINGGIAGLIGVTCVFPIDLAKTRLQNGQRMYSMSDCLIKTIRSEGYF
GMYRGAAVNLTIVTPEKAIKLAANDFFRHQLSKDQKLTLPKEMLAGCGAGTCQVIVTTPMEML
KIQLQDAGRIAAQRKILAAQAQLSAQGAQPSVEAPAP
PRPTATQLTRDLLRNHGIAGLYKGLGATLLRDVFPFSIVYFPLFANLNQLG
RPSSEKSPFYVSFLAGCVAGSAAAVAVNPCDVVKTRLQSLERGVNEDTYSGFLDCARKIW
RHEGPSAFLKGAYCRALVIAPLFGIAQVVYFLGIAESLLGLLQEPQA

94 HUMAN AK023106.1 -----MADKQI 6
94B MOUSE -----MADKQI 0
94A DROSO AC018265 gene 2 -----MLEQVEQKNQEQQKPKQKF 0
94 SLPAKLINGGIAGLIGVTCVFPIDLAKTRLQNGQRMYSMSDCLIKTVRS---EGYF 63
94B SLPAKLINGGIAGLIGVTCVFPIDLAKTRLQNGQRMYSMSDCLIKTIRS---EGYF 0
94A NVFPKIINGGVAGIIGVACVYPLDMVKTRLQNGQRMYSMSDCLIKTVRS---EGYF 0
94 GMYRGAAVNLTIVTPEKAIKLAANDFFRHQLSKDQKLTLPKEMLAGCGAGTCQVIVTTPMEML 112
94B GMYRGAAVNLTIVTPEKAIKLAANDFFRHQLSKDQKLTLPKEMLAGCGAGTCQVIVTTPMEML 0
94A GMYRGSAVNIVLITPEKAIKLTANDFFRYHLASDDG-----VIPLSRATLAGGLAGLFQIVVTPMELL 0

There is an insert of 20 aa in the middle of this section (see gap)

94 KIQLQDAGRIAAQRKILAAQAQLSAQGAQPSVEAPAPPRPTATQLTRDLLRSRGIAGLYKGLGATLLRDVFPFSIVYFPLFANLNQLG-0
94B KIQLQDAGRIAAQRKILAAQAQLSAQGAQPSVEAPAPPRPTATQLTRDLLRNHGIAGLYKGLGATLLRDVFPFSIVYFPLFANLNQLG-0
94A KIQLQDAGRVDRAAGREVKTITALGLTKTLRERGIAGLYKGLGATLLRDVFPFSIVYFPLMAWINDQG-0
94 -----RPSSEKSPFYVSFLAGCVAGSAAAVAVNPCDVVKTRLQSLERGVNE---DTYSGIL 0
94B -----RPSSEKSPFYVSFLAGCVAGSAAAVAVNPCDVVKTRLQSLERGVNE---DTYSGFL 0
94A -----PRKSDGSGEAVFYWSLIAGLLSGMTSAFVMVTPFDVVKTRLQADGEKK-----FKGIM 0

112 N-term

>gi|15398277|dbj|BB624380.1|BB624380 BB624380 RIKEN full-length enriched, 15 days embryo male testis Mus
musculus cDNA clone 8030476B22 5'.
Length = 647

Score = 251 bits (640), Expect = 6e-67
Identities = 128/142 (90%), Positives = 128/142 (90%), Gaps = 11/142 (7%)
Frame = +1

Query: 1 MSQRDTLVHLFAGGCGGTGAILTCPLEVVKTRLQSSSVTLYISEVQLNT----- 50
MSQRDTLVHLFAGGCGGTGAILTCPLEVVKTRLQSSSVTLYISEVQLNT
Sbjct: 85 MSQRDTLVHLFAGGCGGTGAILTCPLEVVKTRLQSSSVTLYISEVQLNTMAGASVNRVV 264
Query: 51 -PGPLHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAYSNCKEKLNDFDPDSTQV 109
PGPLHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAYSNCKEKLNDFDPDSTQV
Sbjct: 265 SPGPLHCLKVILEKEGPRSLFRGLGPNLVGVAPSRAIYFAAYSNCKEKLNDFDPDSTQV 444
Query: 110 HMISAAMAGFTAITATNPIWLI 131
HM SAAMAGFTAITATNPIWLI
Sbjct: 445 HMASAAMAGFTAITATNPIWLI 510

ACCESSION NUMBER BB624380

111 HUMAN NM_018155.1 -----MSQR 0
112 MOUSE W53812 -----MSQR 0
111 DTLVHLFAGGCGGTGAILTCPLEVVKTRLQSS--SVTLYISEVQLNT-PGPLHCLKVILEK---EGPR 0
112 DTLVHLFAGGCGGTGAILTCPLEVVKTRLQSS--SVTLYISEVQLNT-PGPLHCLKVILEK---EGPR 0

111 SLFRGLGPNLVGVAPSRAIYFAAYSNCCKEKLNDVFD-----PDSTQVHMISAAMAGFTAITATNPIWLI 14
112 SLFRGLGPNLVGVAPSRAIYFAAYSNCCKELNGVFD-----PDSTQVHMISAAMAGFTAITATNPIWLI 0

seq 112 C-term

111 LEYKTASTMENDEESVKEASDFVGMMLAAATSKTCATTIAYPHEVVRTRLREEGTKYRSFF----- 133
112 LECKTASMMETDEESVKEASDFVRMMLAAATSKTCATTIAYPHEVVRTRLREEGTKYRSFF----- 0
112ALEQRNQRHT-----DTKGSRDFLEFMMAGAVSKTIASCIPYPHEVARTRLREEGNKYNSFW----- 0

111 QTLSELLVQEEGYGSLYR--GLTTHLVRQIPNTAIMMATYELVVYLLNG*----- 0
112 QTLSLIVQEEGYGSLYR--GLTTHLVRQIPNTAIMMATYELVVYLLNG*----- 0
112AQLHTVWKEEGRAGLYR--GLATQLVRQIPNTAIMMATYEAVVYVLTTRRFNNKSNEFYDF*----- 0

>gi|16464438|dbj|BB626464.1|BB626464 BB626464 RIKEN full-length enriched, adult male diencephalon Mus musculus cDNA clone 9330189G22 5'.
Length = 647

Score = 149 bits (375), Expect = 7e-36
Identities = 74/91 (81%), Positives = 79/91 (86%)
Frame = +2

QTLSLIVQEEGYGSL 62
QTLSLIVQEEGYGSL
QTLSLIVQEEGYGSL 463

Query: 63 YRGLTTHLVRQIPNTAIMMATYELVVYLLNG 93
YRGLTTHLVRQIPNTAIMMATYELVVYLLNG
Sbjct: 464 YRGLTTHLVRQIPNTAIMMATYELVVYLLNG 556

Seq 114

most like human 113 hybrid shown below for searching
Some corrections were made to the C-terminal of 114

SLFRGLGPNLVGVAPSRAVYFACYSKAKEQFNGIFVPNSNIVHIFSAGSAAFITNSLMNPIWMV
KTRMQLEQKVRGSKQMNTLQCARYVYQTEAFVGFYRGLTASYAGISETIICFAIYESLKKYL
KEAPLASSANGTEKNSTSFGLMAAAALSKGCAPAIAYPHEVIRTRLR QEGSKYRSFVQVVR

>gi|7046137|gb|AW476031.1|AW476031 uq83e04.x1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:2937726 3'
similar to WP:T09F3.2 CE02342 CARRIER PROTEIN C2 ;.
Length = 673

Score = 231 bits (589), Expect = 2e-60
Identities = 116/126 (92%), Positives = 119/126 (94%)
Frame = +2

Query: 63 MVKTRMQLEQKVRGSKQMNTLQCARYVYQTEAFVGFYRGLTASYAGISETIICFAIYESL 122
MVKTRMQLEQKVRGSKQMNTLQCARYVYQTE GFYRGLTASYAGISETIICFAIYESL
Sbjct: 47 MVKTRMQLEQKVRGSKQMNTLQCARYVYQTEGIRGFYRGLTASYAGISETIICFAIYESL 226

Query: 123 KKYLKEAPLASSANGTEKNSTSFGLMAAAALSKGCAPAIAYPHEVIRTRLRQEGSKYRS 182
KKYLKEAPLASSANGTEKNSTSFGLMAAAALSKGCA IAYPHEVIRTRLR+EG+KY+S
Sbjct: 227 KKYLKEAPLASSANGTEKNSTSFGLMAAAALSKGCASCIAYPHEVIRTRLREEGTKYKS 406

Query: 183 FVQVVR 188
FVQ R
Sbjct: 407 FVQTAR 424

NSDAWGMVKTRMQLEQKVRGSKQMNTLQCARYVYQTEGIRGFYRGLTASYAGISETIICFAIYESLKKYLKEAPLASSA
NGTEKNSTSFGLMAAAALSKGCASCIAYPHEVIRTRLREEGTKYKSFVQTAR
LVFREEGYLAFYRGLCAQLIRQIPNTAIVLSTYELIVYLLEDRTQ*

Accession numbers for ESTs BE690351 AW475180
Accession numbers for genomic DNA AL607078 AL626808 chr 4 same gene
There are some consistent disagreements
between sequences so there are probably two different genes about 89%
identical.

Note 114a is more like the human sequence from one EST AW476031.
114 is from genomic sequence, supported by two independent clones.

```
113 SLFRGLGPNLVGVAPSRAVYFACYSKAKEQFNGIFV-----PNSNIVHIFSAGSAAAFITNSLMNPIWMV      0
114 SLFRGLGPNLVGVAPSRAVYFACYSKAKEQFNGIFV-----PNSNTVHILSAGSAAAFVTNTLMNPIWMV      0
114a                                     MV

113 KTRMQLE-----QKVRGSKQMNTLQCARYVYQTEA-FVGFYRGLTASYAGISE-TIICFAIYESLKKYL-      23
114 KTRMQLE-----RKVRGCKQMNTLQCARRVYQTEG-VRGFYRGLTASYAGISE-TIICFAIYESLKKCL-      0
114aKTRMQLE-----QKVRGSKQMNTLQCARYVYQTEG-IRGFYRGLTASYAGISE-TIICFAIYESLKKYL-

113 KEAPLASSA--NGTEKNSTSFGLMAAAALSKGCAPAIAYPHEVIRTRLREEGTKYKSFV-----      81
114 KDAPIVSST---DGAEKSSSGFFGLMAAAAVSKGCASCIAYPHEVIRTRLREEGTKYKSFV-----      0
114aKEAPLASSA--NGTEKNSTSFGLMAAAALSKGCASCIAYPHEVIRTRLREEGTKYKSFV-----

113 QTARLVFREEGYLAFYR--GLFAQLIRQIPNTAIVLSTYELIVYLLEDRTQ*-----      130
114 QTARLVFREEGYLAFYR--GLFAQLIRQIPNTAIVLSTYELIVYLLGERA*-----      0
114aQTARLVFREEGYLAFYR--GLCAQLIRQIPNTAIVLSTYELIVYLLEDRTQ*-----
```

>gi|18072546|emb|AL607078.11|AL607078 Mus musculus chromosome 4 clone RP23-
422H14, *** SEQUENCING IN PROGRESS
***, in unordered pieces
Length = 235987

```
Query: 1      SLFRGLGPNLVGVAPSRAVYFACYSKAKE 29
             SLFRGLGPNLVGVAPSR   AC  +A E
Sbjct: 34009 SLFRGLGPNLVGVAPSR---*ACNCRASE 34086

Query: 15     PSRAVYFACYSKAKEQFNGIFVPNSNTVHILSAGSA 50
             P+RAVYFACYSKAKEQFNGIFVPNSNTVHILSAGSA
Sbjct: 36226 PTRAVYFACYSKAKEQFNGIFVPNSNTVHILSAGSA 36333

Query: 51     AFVTNTLMNPIWMVKTRMQLERK 73
             AFVTNTLMNPIWMVKTRMQLERK
Sbjct: 37662 AFVTNTLMNPIWMVKTRMQLERK 37730

Query: 73     KVRGCKQMNTLQCARRVYQTEGVRGFYRGLTASYAGISETIICFAIYESLKKYLKEAPLA 132
             +VRGCKQMNTLQCARRVYQTEGVRGFYRGLTASYAGISETIICFAIYESLKK LK+AP+
Sbjct: 40802 RVRGCKQMNTLQCARRVYQTEGVRGFYRGLTASYAGISETIICFAIYESLKKCLKDAPIV 40981

Query: 133    SSANGTEKNSTSFGLMAAAALSKGCASCIAYPH 166
             SS +G EK+S+ FFGLMAAAA+SKGCASCIAYPH
Sbjct: 40982 SSTDGAEKSSSGFFGLMAAAAVSKGCASCIAYPH 41083

Query: 165    PHEVIRTRLREEGTKYKSFVQTARLVFREEGYLAFYRGLFAQLIRQIPNTAIVLSTYELI 224
             P EVIRTRLREEG+KY+SFVQTARLVFREEGYLAFYRGLFAQLIRQIPNTAIVLSTYE I
Sbjct: 45189 PAEVIRTRLREEGSKYRSFVQTARLVFREEGYLAFYRGLFAQLIRQIPNTAIVLSTYEFI 45368

Query: 225    VYLLEDR 231
             VYLL +R
Sbjct: 45369 VYLLGER 45389
```

sequence 124

hybrid with 123

MTGQGHASAGSSAWSTVFRH
VRYENLVAGVSGGVLNLAHPLDLVKIRFAVSDGLELRPKYNGILHCLTTIWKLDGLR
GLYQGVTPNIWGAGLSWGLYFFFYNAIKSYKTEGRAEYLEATEYLVSAAEAGAMTLCITNPLWVT
KTRLMLQYDAVVNSPHRQY 164 KGMFDALVKIYKYEGRGLYKGFVPGFLFGTSHGALQFMAYELLKLY
NKHINRLPEAQLSTAHEYISVAALSKIFAVAATYPYQVV 248
RARLQDQHMFYSGVI DVITKTWRKEGVGGFYKGIAPNLRVTPACCITFVVYENVSHFLDLREKRK

>gi|16459502|dbj|BB620475.1|BB620475 BB620475 RIKEN full-length enriched, 13
days embryo forelimb Mus
musculus cDNA clone 5930424M21 5'.
Length = 637

Score = 381 bits (978), Expect = e-105
Identities = 183/205 (89%), Positives = 190/205 (92%)
Frame = +3

Query: 1 MTGQGHASAGSSAWSTVFRHVRYENLVAGVSGGVLNLAHPLDLVKIRFAVSDGLELRP 60
MTGQG SA+GS+AWS VFRHVRYENLVAGVSGGVLNLAHPLDLVKIRFAVSDGLE+RP
Sbjct: 21 MTGQQQSAAGSAAWSAVFRHVRYENLVAGVSGGVLNLAHPLDLVKIRFAVSDGLEVRP 200

Query: 61 KYNGILHCLTTIWKLDGLRGLYQGVTPNIWGAGLSWGLYFFFYNAIKSYKTEGRAEYLEA 120
KY GILHCL TIWK+DGLRGLYQGVTPN+WGAGLSWGLYFFFYNAIKSYKTEGRAE LE
Sbjct: 201 KYKGIHCLATIWKVDGLRGLYQGVTPNVWGAGLSWGLYFFFYNAIKSYKTEGRAEQLEP 380

Query: 121 TEYLVSAAEAGAMTLCITNPLWVTKTRLMLQYDAVVNSPHRQYKGMFDALVKIYKYEGRV 180
EYLVSAAEAGAMTLCITNPLWVTKTRLMLQY V + RQYKGMFDALVKIYKYEGRV
Sbjct: 381 LEYLVSAAEAGAMTLCITNPLWVTKTRLMLQYGGVASPSQRQYKGMFDALVKIYKYEGRV 560

Query: 181 GLYKGFVPGFLFGTSHGALQFMAYEL 205
GLYKGF PGLFGTSHGALQFM +EL
Sbjct: 561 GLYKGFAPGLFGTSHGALQFMGFEL 635

>gi|9979752|gb|BE653839.1|BE653839 UI-M-AN1-afa-a-05-0-UI.r1 NIH_BMAP_MBG_N Mus
musculus cDNA clone
UI-M-AN1-afa-a-05-0-UI 5'.
Length = 526

Score = 320 bits (819), Expect = 7e-87
Identities = 157/171 (91%), Positives = 161/171 (93%)
Frame = +1

Query: 144 TKTRLMLQYDAVVNSPHRQYKGMFDALVKIYKYEGRVGLYKGFVPGFLFGTSHGALQFMAY 203
TKTRLMLQY V + RQYKGMFDALVKIYKYEGRVGLYKGFVPGFLFGTSHGALQFMAY
Sbjct: 1 TKTRLMLQYGGVASPSQRQYKGMFDALVKIYKYEGRVGLYKGFVPGFLFGTSHGALQFMAY 180

Query: 204 ELLKLYNKHINRLPEAQLSTAHEYISVAALSKIFAVAATYPYQVVRARLQDQHMFYSGVI 263
ELLKLYNKHINRLPEAQLSTAHEYISVAALSKIFAVAATYPYQVVRARLQDQH+ Y GV
Sbjct: 181 ELLKLYNKHINRLPEAQLSTAHEYISVAALSKIFAVAATYPYQVVRARLQDQHVSYGGVT 360

Query: 264 DVITKTWRKEGVGGFYKGIAPNLRVTPACCITFVVYENVSHFLDLREKR 314
DVITKTWRKEG+GGFYKGIAPNLRVTPACCITFVVYENVSHFL DLREK+
Sbjct: 361 DVITKTWRKEGIGGFYKGIAPNLRVTPACCITFVVYENVSHFLYDLREKK 513

MTGQQQSAAGSAAWSAVFRH
VRYENLVAGVSGGVLNLAHPLDLVKIRFAVSDGLEVRPKYKGIHCLATIWKVDGLR
GLYQGVTPNVWGAGLSWGLYFFFYNAIKSYKTEGRAEQLEPLEYLVSAAEAGAMTLCITNPLWVT
KTRLMLQYGGVASPSQRQY
164 KGMFDALVKIYKYEGRVGLYKGFVPGFLFGTSHGALQFMAYELLKLY
NKHINRLPEAQLSTAHEYISVAALSKIFAVAATYPYQVV 248
RARLQDQHVSYGGVTDVITKTWRKEGIGGFYKGIAPNLRVTPACCITFVVYENVSHFLYDLREKKVS*

Accessions BB620475 BE653839

123 HUMAN AC012213.3 -----MTGQGHASAGSSAWSTVFRH 0

```

123A DROSO AC017981 -----MNPIKAQSTGSPKFFNVFAH 0
123B DROSO AC019525 -----MVAPSKLSQVLSY 0
124 MOUSE FLX1 FORM2 -----MTGQQQSAAGSAAWSAVFRH 0

123 VRYENLVAGVSGGVLNLAHPLDLVKIRFAVS-----DG-LELRPKY-NGILHCLTTIWKL----DGLR 0
123AVKYEHLVAGVSGGVVSTLILHPLDLKIRFAVN-----DGRTATVPQY-RGLSSAFTTIFRQ----EGFR 0
123BQNFVHAVSAGAGGCIAMSTFYPLDVTVSRLQL-----EEAGDV-RSTRQVIKEIVLG----EGFQ 0
124 VRYENLVAGVSGGVLNLAHPLDLVKIRFAVS-----DG-LEVRPKY-KGILHCLATIWKV----DGLR 0

123 GLYQGVTPNIWGAGLSWGLYFFFYNAIKSYKTEGRA-----EYLEATEYLVSAAEAGAMTLCITNPLWVT 0
123AGLYKGVTPNVWGSWGLYFMYNTIKTFIQGGNT-----TMPLGPTMNMLAAAESGILTLLTNPIWVV 0
123BSLYRGLGPVLQSLCISNFVYFYTFHALKAVASGGSP-----SQHSALKDLLLLGSIAGIINVLTTTPFWVV 0
124 GLYQGVTPNVWAGLSWGLYFFFYNAIKSYKTEGRA-----EQLEPLEYLVSAAEAGAMTLCITNPLWVT 0

123 KTRLMLQYDA-VVNSPHRQYKGMFDTLVKIYKYEGR-VRGLYKGFVPGFLFGTSH-GALQFMAYELLKLY- 0
123AKTRLCLQCA----ASSAEYRGMihalGQIYKEEG-IRGLYRGFVPGMLGVSH-GAIQFMTYEELKNAY- 0
123BNTRLRMRNVAGTSDEVNKHYNLLEGLKYVAEKEG-IAGLWSGTIPSLMLVSN-PALQFMMEYMLKRNI- 0
124 KTRLMLQYGG-VASPSQRQYKGMFDALVKIYKYEGR-VRGLYKGFVPGFLFGTSH-GALQFMAYELLKLY- 0

123 NQHINRL-----PEAQLSTVEYISVAALSKIFAATAATYPYQVVRARLQDQHMF-----YSGVI 0
123ANEYRKLK-----IDTKLATTEYLFAAAVSKLIAAAATYPYQVVRARLQDHHHR-----YNGTW 0
123BMRFTGGE-----MGSLSFFFIGAIAKAFATVLTYPQLVQTKQRHRSKESDSAGSTPRTESTL 0
124 NKHINRL-----PEAQLSTAIEYISVAALSKIFAATAATYPYQVVRARLQDQHVS-----YGGVT 0

123 DVITKTWRKEGVGGFYK--GIAPNLIRVTPACCITFVYENVSHFLLDLREKVK*----- 0
123ADCIKQTRWYERMRGFYK--GLKASLTRVVPACMVTFVYENVSHFLLARRKRIETKEDASDV*----- 0
123BELMISILQHQIRGLFR--GLEAKILQTVLTAALMFAYEKIAGTVGMLLKRN*----- 0
124 DVITKTWRKEGIGGFYK--GIAPNLIRVTPACCITFVYENVSHFLYDLREKVK*----- 0

```

Seq 133

ACCESSION NUMBERS BB654057 BE649081 AF361699 NM_026331 and NM_016612

```

132 HUMAN AF223466.1 MELRSGSVGSQAVARRMDGDSRDGG-GGKDATGSEYENLPTSA 0
133 MOUSE AA014938 MELRRGGVGNQAAGRMDGDCRDGGCGSKDA-GSEYENLPTSA 0

132 SVSTHMTAGAMAGILEHSMYPVDSVKTRMQSL-----SPDPKAQY-TSIYGALKKIMRT----EASG 63
133 SVSTHMTAGXXAGILEHSMYPVDSVKTRMQSL-----NPDPKARY-TSIYGALKRIMHT----EGFW 0

132 RPLRGNVNMIMGAGPAHAMYFACYENMKRTLNDVFH----HQGNSHLANGIAGSMATLLHDAVMNPAEVV 0
133 RPLRGLNVMMGAGPAHAMYFACYENMKRTLNDVFS----HQGNSHLANGVAGSMATLLHDAVMNPAEVV 0

132 KQRLQMY-----NSQHRS AISCI RTVWRTEG-LGAFYRSYTTQLTMNIPFQSIHFITYEFLQEQVN 0
133 KQRLQMY-----NSQHRS AISCI RTVWRTEG-LGAFYRSYTTQLTMNIPFQSIHFITYEFLQEQVN 0

132 -----PHRTYNPQSHIISGGLAGALAAAATPLDVCKTLLNTQENVALSANISGRLSGMA 0
133 -----PRRDYNPQSHIISGGLAGALAAAATPLDVCKTLLNTQENMALSANVSGRLSGMA 0

132 NAFRTVYQLNGLAAYFK--GIQARVIYQMPSTAISWSVYEFFKYFLTKRQWENRAPY*----- 0
133 NAFRTVYQLNGLAGYFK--GIQARVIYQMPSTAISWSVYEFFKYILTKRQLENRTLY*----- 0

```

>gi|16755527|gb|AF361699.1|AF361699 Mus musculus mitochondrial carrier-like protein mRNA, complete cds;
nuclear gene for mitochondrial product
Length = 5869

Score = 330 bits (845), Expect = 2e-90
Identities = 164/171 (95%), Positives = 164/171 (95%)
Frame = +1

```

Query: 1 KQRLQMYNSQHSAFSCIRTVWRTEGLGAFYRSYTTQLTMNIPFQSIHFITYEFLQEQVN 60
KQRLQMYNSQHSAFSCIRTVWRTEGLGAFYRSYTTQLTMNIPFQSIHFITYEFLQEQVN
Sbjct: 2374 KQRLQMYNSQHSAFSCIRTVWRTEGLGAFYRSYTTQLTMNIPFQSIHFITYEFLQEQVN 2553

Query: 61 PHRTYNPQSHIISGGLAGALAAAATPLDVCKTLLNTQENMALSANVSGRLSGMANAFR 120

```

P R YNPQSHIISGGLAGALAAAATPLDVCKTLLNTQENMALSLANVSGRLSGMANAFR
Sbjct: 2554 PRRDYNPQSHIISGGLAGALAAAATPLDVCKTLLNTQENMALSLANVSGRLSGMANAFR 2733

Query: 121 TVYQLNGLAAYFKGIQARVIYQMPSTAISWSVYEFFKYFLTKRQWENRAPY 171
TVYQLNGLA YFKGIQARVIYQMPSTAISWSVYEFFKY LTKRQ ENR Y
Sbjct: 2734 TVYQLNGLAGYFKGIQARVIYQMPSTAISWSVYEFFKYILTKRQLENRTLY 2886

>gi|9974905|gb|BE649081.1|BE649081 UI-M-AP1-agf-a-12-0-UI.r1 NIH_BMAP_MST_N Mus
musculus cDNA clone
UI-M-AP1-agf-a-12-0-UI 5'.
Length = 127

Score = 76.3 bits (186), Expect = 6e-14
Identities = 39/42 (92%), Positives = 41/42 (96%), Gaps = 1/42 (2%)
Frame = +2

Query: 75 AAATPLDVCKTLLNTQENVALS-ANISGRLSGMANAFRTVY 115
AAATPLDVCKTLLNTQEN+ALS AN+SGRLSGMANAFRTVY
Sbjct: 2 AAATPLDVCKTLLNTQENMALSLANVSGRLSGMANAFRTVY 127

>gi|16487885|dbj|BB654057.1|BB654057 BB654057 RIKEN full-length enriched, adult
male liver tumor Mus
musculus cDNA clone C730046J21 5'.
Length = 678

Score = 270 bits (690), Expect(2) = 1e-76
Identities = 134/151 (88%), Positives = 139/151 (91%)
Frame = +2

Query: 1 MDGDCRDGVTSSKDAGSEYENLPTSASVSTHMTAGAMAGILEHSIMYPVDSVKTRMQSL 60
MDGDCRDG SKDAGSEYENLPTSASVSTHMTAGAMAGILEHSIMYPVDSVKTRMQSL
Sbjct: 170 MDGDCRDGGCGSKDAGSEYENLPTSASVSTHMTAGAMAGILEHSIMYPVDSVKTRMQSL 349

Query: 61 SPDPKAQYTSIYGALKKIMRTEASGRPLRGVNVMMMGAGLAHAMYFAFYENMKRTLNDVF 120
+PDPKA+YTSIYGALK+IM TE RPLRG+NVMMMGAG AHAMYFA YENMKRT NDV
Sbjct: 350 NPDPKARYTSIYGALKRIMHTEGFWRPLRGLNVMMMGAGPAHAMYFACYENMKRTXNDVX 529

Query: 121 HHQGNSHLANGIAGSMATLLHDAVMNPAEVV 151
HHQGNSHLANG+AGSMATLLHDAVMNPAEVV
Sbjct: 530 SHQGNSHLANGVAGSMATLLHDAVMNPAEVV 622

Score = 37.7 bits (86), Expect(2) = 1e-76
Identities = 16/19 (84%), Positives = 18/19 (94%)
Frame = +1

Query: 152 KQRLQMYNSQHRS AIS CIR 170
+QRLQMYNSQH+SA SCIR
Sbjct: 622 EQRLQMYNSQHQS AFS CIR 678

>gi|16479103|dbj|BB644566.1|BB644566 BB644566 RIKEN full-length enriched, adult
male corpora
quadrigemina Mus musculus cDNA clone B230333D09 5'.
Length = 666

Score = 239 bits (610), Expect = 1e-62
Identities = 118/132 (89%), Positives = 122/132 (92%)
Frame = +1

Query: 1 MDGDCRDGVTSSKDAGSEYENLPTSASVSTHMTAGAMAGILEHSIMYPVDSVKTRMQSL 60
MDGDCRDG SKDAGSEYENLPTSASVSTHMTAGAMAGILEHSIMYPVDSVKTRMQSL
Sbjct: 64 MDGDCRDGGCGSKDAGSEYENLPTSASVSTHMTAGAMAGILEHSIMYPVDSVKTRMQSL 243

Query: 61 SPDPKAQYTSIYGALKKIMRTEASGRPLRGVNVMMMGAGLAHAMYFAFYENMKRTLNDVF 120

+PDPKA+YTSIYGALK+IM TE RPLRG+NVMMMGAG AHAMYFA YENMKRTLNDVF
Sbjct: 244 NPDPKARYTSIYGALKRIMHTEGFWRPLRGLNVMMMGAGPAHAMYFACYENMKRTLNDVF 423

Query: 121 HHQGNSHLANGI 132
HHQGNSHLANGI
Sbjct: 424 SHQGNSHLANGI 459

>gi|6516145|gb|AW210205.1|AW210205 ul52b03.y1 Rashbass mouse MOC 11 5 optic cup
Mus musculus cDNA
clone IMAGE:2101901 5' similar to WP:W02B12.9 CE03769
MITOCHONDRIAL RNA SPLICING MSR4 LIKE PROTEIN ;.
Length = 512

Score = 184 bits (466), Expect = 6e-46
Identities = 91/114 (79%), Positives = 96/114 (83%)
Frame = +1

Query: 64 PKAQYTSIYGALKKIMRTEASGRPLRGVNVMMMGAGLAHAMYFAFYENMKRTLNDVFFHQ 123
PK + + IM TE RPLRG+NVMMMGAG AHAMYFA YENMKRTLNDVF HQ
Sbjct: 169 PKPGIQASMAPSRGIMHTEGFWRPLRGLNVMMMGAGPAHAMYFACYENMKRTLNDVFSHQ 348

Query: 124 GNSHLANGIAGSMATLLHDAVMNPAEVVKQRLQMYNSQHRSAISCIRTVWRTEG 177
GNSHLANG+AGSMATLLHDAVMNPAEVVKQRLQMYNSQH+SA SCIRTVWRTEG
Sbjct: 349 GNSHLANGVAGSMATLLHDAVMNPAEVVKQRLQMYNSQHSAFSCIRTVWRTEG 510

Score = 129 bits (324), Expect = 2e-29
Identities = 63/66 (95%), Positives = 66/66 (99%)
Frame = +3

Query: 12 SKDAGSEYENLPTSASVSTHMTAGAMAGILEHSIMYPVDSVKTRMQSLSPDPKAQYTSI 71
SKDAGSEYENLPTSASVSTHMTAGAMAGILEHSIMYPVDSVKTRMQSL+PDPKA+YTSI
Sbjct: 12 SKDAGSEYENLPTSASVSTHMTAGAMAGILEHSIMYPVDSVKTRMQSLNPDPKARYTSI 191

Query: 72 YGALKK 77
YGALK+
Sbjct: 192YGALKR 209

>gi|9817629|gb|BE573907.1|BE573907 601331185F1 NCI_CGAP_Mam6 Mus musculus cDNA
clone IMAGE:3708765 5'.
Length = 617

Score = 112 bits (281), Expect = 2e-24
Identities = 60/84 (71%), Positives = 65/84 (76%), Gaps = 2/84 (2%)
Frame = +2

Query: 131 GIAGSMATLLHDAVMNPAEVVKQRLQMYNSQHRSAISCIRTVWRTEGLGAFYRSYT-TQL 189
G+AGSMATLLHDAVMNPAEVVKQRLQMYNSQH+SA SCIRTVWRTEGLGAFYRSY T
Sbjct: 227 GVAGSMATLLHDAVMNPAEVVKQRLQMYNSQHSAFSCIRTVWRTEGLGAFYRSYNHTS* 406

Query: 190 TMNIPFQSIHFIT-YEFLQEQVNP 212
P IHF + F + +V P
Sbjct: 407 P*ISPSSQIHFPSPMSFCRSKVEP 478

Score = 48.1 bits (113), Expect = 5e-05
Identities = 34/63 (53%), Positives = 37/63 (57%), Gaps = 3/63 (4%)
Frame = +1

Query: 188 QLTMNIPFQSIHF-ITYEFLQEQVNP-HRTYNPQSHIISG-GLAGALAAAATTPLDVCKT 244
+LTMNIPFQS ITYEFLQEQ T N HIIS G A TTP++V KT
Sbjct: 400 ELTMNIPFQSNLSITYEFLQEQRRLAGTNNHSLHIISARPWPGTWPAGCTTPVEVWKT 579

Query: 245 LLN 247
L N

Sbjct: 580 LSN 588

Sequence 135

Accession numbers BG964218 BI103329 BI854643

MELEGRSAGGVAGGPAAGPGRSPGESALLDGLWLLAVAWAWGAGGGEA probable N-term

Note the N-term of 134 human was probably incorrect due to a frameshift
It now matches the mouse

134 HUMAN AL353719.5	LQRGVGRGAGGGEAGACRPPVRQDPDS--GPDYEALPAGA	0
135 MOUSE HOMOLOG W46092	LAVAWAWGAGGGEAGAYQPPVRLDPES--GPEYEALPAGA	0
135A DROSO AC010580	-----MAAELGLESAAGSVAIKMQEPVNKL	0

3 frames translation

EAPGSLGRA*GGRNGAGRRQWWRQRRRVWAPRLAEVDGVVTGL*VQACCVLGVMTLLGPGGAPGSRRRCWTGGCSGAWA
GGPAAGRRGPIISPLYGWIRSPARNTKRCRLEPLSPRTWWRAPWQGSWSIA*CTRSTASRPGCRAYSILTQPPAIGTCWRL
SGES*GQRACGGPCGG*TSQQQARGLPTPSILPATKS*KRH*VT*STQGAIAILPMVAAGCVATLLS*CRHESGSRQA
EDADVQLAVPPDRDLCSGSVGMQAGALLTAATR

GGPGLSGPRLRRTQRGRGAAAVVAAAAARVGAAAG*SGWSCDGAVSAGVLCAGSDHAAGPGRSPGESALLDGLWLRQGVG
RGAGGGEAGAYQPPVRLDPESGPEYEALPAGATVTHMVAVAGVAGILEHCVMYPIDCVKTRMQSLQDPDPAARYRNVLEA
LWRIMRTEGLWRPMRGLNVTATGAGPAHALYFACYEKLKKTLSVDVIHPGGNSHIANGCSRMCGDITFMMQA*IQRKSSS
RGCRCCTTRRTTA*QTVFGQCWQNAAGGAFNRSYTN

RRPRALWAAPEADATGPGGGGSGGSGGACGRRGWLKWMELEGRSAGGVAGGPAAGPGRSPGESALLDGLWLLAVAWAWGAG
GGRRRRGGGGLSAPCTAGSGVRPGIRSAAGWSHCHHAHGGRRGRDPGALRDVDPDLRQDPDAEPTA*PSRPLSERVGG
SLENHEDRGPVEAHAGAERHSNRGACPRPLFCLLRKVKKIDIE*RNPPRGQ*PYCQWLQPDVWRHYFHDAGMNPAEVVK
QRMQMYNSPYHRVTDCVRAVLAKCRGRGF*PQLHE

3 frames translation

PGLSGPRLRRTQRGRGAAAVVAAACGACGRRGWPKWMELEGRSAGGVAGGPAAGPGRSPGESALLDGLWLLAVAWAWGAG
GGEAGAYQPPVLAGSGVRPGIRSAAGWSHCHHAHAGVAVAGVAGILEHCVMYPIDCVKTRMQSLQDPDPAARYRNVLEALWRI
MRTEGLWRPMRGLNVTATGAGPAHALYFACYEKLKKTLSVDVIHPGGK*PYCQWY*AFLCWSPLPNSLALLSVLPGSHG
LRAASFAGGKPGLYGGGAPPKTCTVASRDSGACFNRYLPMGRASRKVPSVAPPLHPGWDFRSISLSIGSCGSVFFVE
PYDPA

PRALWAAPEADATGPGGGGSGGGSVRRVWAPRLAEVDGVGGAECRRRGGRTSCWARAEPRGVGAAGRVAASGGVGLGGR
RRGGGGLSAPCTGWIRSPARNTKRCRLEPLSPRTCGRRGRDPGALRDVDPDLRQDPDAEPTA*PSRPLSERVGGSL
HEDRGPVEAHAGAERHSNRGACPRPLFCLLRKVKKIDIE*RNPPRGQIAILPMVLSLPVLVPTSKLSSFAAVSASRLAW
FARCQLWRWG*ARIIWGGGSPQNLYSVSRQGVF*PQISPYGARLTQGSQCCSPLTPRLGFPLDIVDWFMWLFCIFR
TL*SR*

APGSLGRA*GGRNGAGRRQWWRQRAARVGAAGRSGWSWRGGVQAAWREDQLLGGGAPGSRRRCWTGGC*RWRGPGGP
AAGRRGPIISPLYWLDPESEGPEYEALPAGATVTHMWRAPWQGSWSIA*CTRSTASRPGCRAYSILTQPPAIGTCWRLS
S*GQRACGGPCGG*TSQQQARGLPTPSILPATKS*KRH*VT*STQGANSHIANGIEPSCAGPHFQTL*LCCCQCFQARM
VCALPALALGVSPDYMGGLPKPVQLRLATAGRVLPDISLWLGAPHARFPVLLPPYTQVGISARYRCRLVHVALYFL*
NPMIPL

sequence 138

accession BC019156

137 HUMAN PET8 HOMOL	-----MDRP	4
138 MOUSE R74620 PET8 HOMOL	-----MDAP	0
139 PET8 ChrXIV U02536	-----MN	2

137 GFVALLVAGGVAGVSDVLDLIFPLDTIKTRLQSP	-----QGFSKA	---GGFH	47
138 GFTASLVAGGVAGVSDVLDLIFPLDTIKTRLQSP	-----QGFNKA	---GGFR	40
139 TFFLSLLSGAAAGTSTDLVFFPIDTIKTRLQAK	-----GGFFAN	---GGYK	45

>gi|17512386|gb|BC019156.1|BC019156 Mus musculus, Similar to RIKEN cDNA
4930433D19 gene, clone
IMAGE:5066084, mRNA
Length = 1101

Score = 66.2 bits (160), Expect = 2e-11
Identities = 33/37 (89%), Positives = 33/37 (89%)
Frame = +1

Query: 1 MDRPGLVALLVAGGVAGVSVDLILFPLDTIKTRLQSP 37
MD PG A LVAGGVAGVSVDLILFPLDTIKTRLQSP
Sbjct: 82 MDAPGFTASLVAGGVAGVSVDLILFPLDTIKTRLQSP 192

sequence 166

accession BC018161

165	HUMAN	PHOSPHATE CARRIER	DGLGDLRSSSPGPTGQPRRPRNLAAA	VEE-YSCEFGSAK	13
166	Phosphate carrier mouse	DGLSGPRS----	PPAPRRSRHLAAA	VEE-YSCEFGSMK	0
167	HUMAN P04 CARRIER 2	DGLGDLRSSSPGPTGQPRRPRNLAAA	VEEQYSCDYGSGR		0