Anexo PEC 2 - Genómica Computacional

Alejandro Keymer

10/12/2019

Ejercicio 1. Estrategias de alineamiento [50%]

Secuencias CDS de la Proteína de ratón y humano del gen `TMEM106B`

```
>mouse
MGKSLSHLPLHSNKEDGYDGVTSTDNMRNGLVSSEVHNEDGRNGDVSQFP
YVEFTGRDSVTCPTCQGTGRIPRGQENQLVALIPYSDQRLRPRRTKLYVM
ASVFVCLLLSGLAVFFLFPRSIEVKYIGVKSAYVSYDAEKRTIYLNITNT
LNITNNNYYSVEVENITAQVQFSKTVIGKARLNNITNIGPLDMKQIDYTV
PTVIAEEMSYMYDFCTLLSIKVHNIVLMMQVTVTTAYFGHSEQISQERYQ
YVDCGRNTTYQLAQSEYLNVLQPQQ
```

```
>human
MGKSLSHLPLHSSKEDAYDGVTSENMRNGLVNSEVHNEDGRNGDVSQFPY
VEFTGRDSVTCPTCQGTGRIPRGQENQLVALIPYSDQRLRPRRTKLYVMA
SVFVCLLLSGLAVFFLFPRSIDVKYIGVKSAYVSYDVQKRTIYLNITNTL
NITNNNYYSVEVENITAQVQFSKTVIGKARLNNITIIGPLDMKQIDYTVP
TVIAEEMSYMYDFCTLISIKVHNIVLMMQVTVTTTYFGHSEQISQERYQY
VDCGRNTTYQLGQSEYLNVLQPQQ
```

Secuencias CDS de la Nucleótidos de ratón y humano del gen `TMEM106B`

```
>human
ATGGGAAAGTCTCTTTCTCATTTGCCTTTGCATTCAAGCAAAGAAGATGC
TTATGATGGAGTCACATCTGAAAACATGAGGAATGGACTGGTTAATAGTG
AAGTCCATAATGAAGATGGAAGAAATGGAGATGTCTCTCAGTTTCCATAT
GTGGAATTTACAGGAAGAGATAGTGTCACCTGCCCTACTTGTCAGGGAAC
AGGAAGAATTCCTAGGGGGCAAGAAAACCAACTGGTGGCATTGATTCCAT
ATAGTGATCAGAGATTAAGGCCAAGAAGAACAAAGCTGTATGTGATGGCT
TCTGTGTTTGTCTGTCTACTCCTTTCTGGATTGGCTGTGTTTTTCCTTTT
CCCTCGCTCTATCGACGTGAAATACATTGGTGTAAAATCAGCCTATGTCA
GTTATGATGTTCAGAAGCGTACAATTTATTTAAATATCACAAACACACTA
AATATAACAAACAATAACTATTACTCTGTCGAAGTTGAAAACATCACTGC
CCAAGTTCAATTTTCAAAAACAGTTATTGGAAAGGCACGCTTAAACAACA
TAACCATTATTGGTCCACTTGATATGAAACAAATTGATTACACAGTACCT
ACCGTTATAGCAGAGGAAATGAGTTATATGTATGATTTCTGTACTCTGAT
ATCCATCAAAGTGCATAACATAGTACTCATGATGCAAGTTACTGTGACAA
CAACATACTTTGGCCACTCTGAACAGATATCCCAGGAGAGGTATCAGTAT
GTCGACTGTGGAAGAAACACAACTTATCAGTTGGGGCAGTCTGAATATTT
AAATGTACTTCAGCCACAACAGTAA
```

```
>mouse
ATGGGAAAGTCTCTTTCTCACTTACCTTTGCATTCAAATAAAGAAGATGG
CTATGATGGCGTTACATCGACAGACAATATGAGAAATGGATTGGTTAGCA
GTGAAGTGCACAACGAAGACGGAAGAAATGGAGATGTCTCTCAGTTCCCA
TATGTGGAATTTACTGGAAGAGATAGTGTCACTTGTCCCACTTGCCAAGG
AACAGGAAGAATTCCTAGGGGACAAGAAAACCAACTGGTGGCATTGATTC
```

```
CATATAGTGATCAGCGGTTACGGCCAAGAAGAACAAAGCTGTATGTGATG
GCGTCTGTGTTTGTCTGCCTGCTCCTGTCTGGATTGGCTGTGTTTTTTCT
TTTCCCTCGATCTATTGAGGTGAAGTACATTGGAGTAAAATCAGCCTATG
TCAGCTACGACGCTGAAAAGCGAACCATATATTTAAATATCACGAACACA
CTAAATATAACAAATAATAACTATTATTCTGTTGAAGTTGAAAACATCAC
TGCTCAAGTCCAGTTTTCAAAAACCGTGATTGGAAAGGCTCGTTTAAACA
ACATAACTAACATTGGCCCACTTGATATGAAGCAGATTGATTATACGGTA
CCCACAGTTATTGCAGAGGAAATGAGTTACATGTATGATTTCTGTACACT
GCTCTCCATCAAAGTGCACAACATAGTACTCATGATGCAAGTTACTGTAA
CAACAGCATACTTTGGACACTCTGAGCAGATATCTCAGGAAAGGTACCAG
TATGTCGACTGTGGAAGGAACACGACTTACCAGTTGGCCCAGTCTGAGTA
TCTAAATGTCCTTCAGCCACAACAATAA
```

3-5

GENEID

```
## date Sun Dec 15 17:16:17 2019
## source-version: geneid v 1.2 -- geneid@imim.es
# Sequence human - Length = 37571 bps
# Optimal Gene Structure. 2 genes. Score = 31.87
# Gene 1 (Forward). 11 exons. 622 aa. Score = 31.58
    First     157     286   9.81   + 0 1    8.07    2.83   20.67   0.00   AA    1: 44 human_1
  Internal  10376   10458   1.45   + 2 0    5.58    2.65    3.77   0.00   AA   44: 71 human_1
  Internal  12800   12857   0.89   + 0 1    3.87    3.09    5.54   0.00   AA   72: 91 human_1
  Internal  15504   15655  -0.00   + 2 0    0.91    4.65    5.41   0.00   AA   91:141 human_1
  Internal  16764   16828   1.03   + 0 2    4.32    1.67    6.10   0.00   AA  142:163 human_1
  Internal  17225   17406   5.73   + 1 1    3.69    3.72   15.71   0.00   AA  163:224 human_1
  Internal  23771   23865  -1.35   + 2 0   -0.44    3.68    4.25   0.00   AA  224:255 human_1
  Internal  25045   25142   2.96   + 0 2    3.54    0.05   14.52   0.00   AA  256:288 human_1
  Internal  26262   26281   2.17   + 1 1    6.90    4.53    0.77   0.00   AA  288:295 human_1
  Internal  27296   27427   2.70   + 2 1    0.45    5.26   10.70   0.00   AA  295:339 human_1
  Terminal  28008   28858   6.20   + 2 0    4.56    0.00   21.14   0.00   AA  339:622 human_1
```

```
>human_1|geneid_v1.2_predicted_protein_1|622_AA
MAPAMQPAEIQFAQRLASSEKGIRDRAVKKLRQYISVKTQRETGGFSQEELLKIWKGLFY
CMWVQDEPLLQEELANTIAQLVHAVNNSAAQACVWFFSRIKVFLDVLMKEVLCPESQSPN
GVRFHFIDIYLDELSKVGGKELLADQNLKFIDPFCKIAAKTKDHTLVQTIARGVFEAIVD
QSPFVPEETMEEQKTKVGDGDLSAEEIPENEVSLRRAVSKKKTALGKNHSRKDGLSDERG
RDDCGTFEDTGPLLQFDYKAVADRLLEMTSRKNTPHFNRKRLSKLIKKFQDLSEGSSISQ
LSFAEDISADEDDQILSQGKHKKKGNKLLEKTNLEKEKGSRVFCVEEEDSESSLQKRRRK
KKKKHHLQPENPGPGGAAPSLEQNRGREPEASGLKALKARVAEPGAEATSSTGEESGSEH
PPAVPMHNKRKRPRKKSPRAHREMLESAVLPPEDMSQSGPSGSHPQGPRGSPTGGAQLLK
RKRKLGVVPVNGSGLSTPAWPPLQQEGPPTGPAEGANSHTTLPQRRRLQKKKAGPGSLEL
CGLPSQKTASLKKRKKMRVMSNLVEHNGVLESEAGQPQALVRWEHPQASSPQRHSLASMG
LHCLLRGRVGAGGQASGLSSS*
```

```
# Gene 2 (Forward). 4 exons. 141 aa. Score = 0.28
    First    30518   30529  -2.92   + 0 0    1.42    1.23    1.21   0.00   AA    1:  4 human_2
  Internal  30780   30932   0.68   + 0 0    2.81    3.31    5.01   0.00   AA    5: 55 human_2
  Internal  31931   31994   2.93   + 0 1    4.88    4.80    5.31   0.00   AA   56: 77 human_2
  Terminal  33682   33875  -0.40   + 2 0   -0.70    0.00   12.53   0.00   AA   77:141 human_2
```

```
>human_2|geneid_v1.2_predicted_protein_2|141_AA
MKIKGSSGTCSSLKKQKLRAESDFVKFDTPFLPKPLFFRRAKSSTATHPPGPAVQLNKTP
```

2

```
SSSKKVTFGLNRNMTAEFKKTDKSILVSPTGPSRVAFDPEQKPLHGVLKTPTSSPASSPL
VAKKPLTTTPRRRPRAMDFF*
```

GENESCAN

View gene model output: PS | PDF

GENSCAN 1.0 Date run: 15-Dec-119   Time: 10:56:32

Sequence /tmp/12_15_19-10:56:32.fasta : 37576 bp : 47.68% C+G : Isochore 2 (43 - 51 C+G%)

Parameter matrix: HumanIso.smat

Predicted genes/exons:

Gn.Ex	Type	S	.Begin	...End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P....	Tscr..
1.01	Init	+	162	291	130	2	1	107	80	324	0.752	33.81
1.02	Intr	+	10381	10463	83	2	2	94	92	26	0.829	2.96
1.03	Intr	+	12805	12862	58	0	1	97	99	62	0.963	6.66
1.04	Intr	+	14367	14452	86	1	2	47	95	49	0.678	1.04
1.05	Intr	+	15133	15194	62	0	2	53	86	51	0.694	-1.07
1.06	Intr	+	15531	15660	130	0	1	27	99	108	0.642	6.50
1.07	Intr	+	16769	16833	65	1	2	78	83	73	0.995	3.32
1.08	Intr	+	17230	17411	182	1	2	77	91	192	0.962	17.91
1.09	Intr	+	23776	23870	95	2	2	37	94	55	0.688	0.68
1.10	Intr	+	25050	25147	98	2	2	64	26	129	0.640	3.11
1.11	Intr	+	26267	26286	20	2	2	91	100	-1	0.600	-2.35
1.12	Intr	+	27301	27432	132	2	0	41	121	120	0.872	11.22
1.13	Intr	+	27668	27856	189	0	0	51	67	92	0.625	2.96

```
1.14 Intr +  28013  28737  725  0  2   85   95   470 0.762  38.55

1.15 Intr +  30241  30385  145  0  1   71   48    71 0.368   1.26

1.16 Intr +  30594  30676   83  1  2   30   51    91 0.478  -1.04

1.17 Intr +  30785  30937  153  1  0  100  101   109 0.999  13.67

1.18 Intr +  31936  31999   64  0  1  114  131    52 0.996  10.39

1.19 Term +  33687  33880  194  1  2   52   55   187 0.999   9.38

1.20 PlyA +  35505  35510    6                              -0.45
```

Suboptimal exons with probability > 1.000

```
Exnum Type S .Begin ...End .Len Fr Ph B/Ac Do/T CodRg P.... Tscr..

----- ---- - ------ ------ ---- -- -- ---- ---- ----- ----- ------
```

NO EXONS FOUND AT GIVEN PROBABILITY CUTOFF

Predicted peptide sequence(s):

```
>/tmp/12_15_19-10:56:32.fasta|GENSCAN_predicted_peptide_1|897_aa
MAPAMQPAEIQFAQRLASSEKGIRDRAVKKLRQYISVKTQRETGGFSQEELLKIWKGLFY
CMWVQDEPLLQEELANTIAQLVHAVNNSAAQHLFIQTFWQTMNREWKGIDRLRLDKYYML
IRLVLRQSFEVLKRNGWEESRIKVFLDVLMKEVLCPESQSPNGVRFHFIDIYLDELSKVG
GKELLADQNLKFIDPFCKIAAKTKDHTLVQTIARGVFEAIVDQSPFVPEETMEEQKTKVG
DGDLSAEEIPENEVSLRRAVSKKKTALGKNHSRKDGLSDERGRDDCGTFEDTGPLLQFDY
KAVADRLLEMTSRKNTPHFNRKRLSKLIKKFQDLSEGSSISQLSFAEDISADEDDQILSQ
GKHKKKGNKLLEKTNLEKEKGKQELQGALGGGCLMTTRDLWFLPLSPKISGNGTISVPYV
FINGQKEGFQSQLGMEEVGPDDKGSRVFCVEEEDSESSLQKRRRKKKKKHHLQPENPGPG
GAAPSLEQNRGREPEASGLKALKARVAEPGAEATSSTGEESGSEHPPAVPMHNKRKRPRK
KSPRAHREMLESAVLPPEDMSQSGPSGSHPQGPRGSPTGGAQLLKRKRKLGVVPVNGSGL
STPAWPPLQQEGPPTGPAEGANSHTTLPQRRRLQKKKAGPGSLELCGLPSQKTASLKKRK
```

4

```
KMRVMSNLVEHNGVLESEAGQPQALAAHLNLPEPPVCRQRHWAAHTSESQVRDPVSLWVA
VSCCTRNECPGPASVVLCVKPELCRMEGLSASAVRKTAGRRGSSGTCSSLKKQKLRAESD
FVKFDTPFLPKPLFFRRAKSSTATHPPGPAVQLNKTPSSSKKVTFGLNRNMTAEFKKTDK
SILVSPTGPSRVAFDPEQKPLHGVLKTPTSSPASSPLVAKKPLTTTPRRRPRAMDFF
```

FGNESH

```
FGENESH 2.6 Prediction of potential genes in Homo_sapiens genomic DNA
 Time    :   Sun Dec 15 10:58:15 2019
 Seq name: human
 Length of sequence: 37571
 Number of predicted genes 1: in +chain 1, in -chain 0.
 Number of predicted exons 16: in +chain 16, in -chain 0.
 Positions of predicted genes and exons: Variant   1 from   1, Score:115.993872
   G Str    Feature    Start        End     Score        ORF           Len

   1 +    1 CDSf      157 -       286    28.30      157 -       285    129
   1 +    2 CDSi    10376 -     10458     7.43    10378 -     10458     81
   1 +    3 CDSi    12800 -     12857     6.33    12800 -     12856     57
   1 +    4 CDSi    14362 -     14447     3.34    14364 -     14447     84
   1 +    5 CDSi    15128 -     15189     2.40    15128 -     15187     60
   1 +    6 CDSi    15526 -     15655     6.39    15527 -     15655    129
   1 +    7 CDSi    16764 -     16828     6.62    16764 -     16826     63
   1 +    8 CDSi    17225 -     17406    12.24    17226 -     17405    180
   1 +    9 CDSi    23771 -     23865     2.10    23773 -     23865     93
   1 +   10 CDSi    25045 -     25142     0.60    25045 -     25140     96
   1 +   11 CDSi    26262 -     26281    -1.21    26263 -     26280     18
   1 +   12 CDSi    27296 -     27427     8.29    27298 -     27426    129
   1 +   13 CDSi    28008 -     28732    33.29    28010 -     28732    723
   1 +   14 CDSi    30780 -     30932     9.89    30780 -     30932    153
   1 +   15 CDSi    31931 -     31994    13.04    31931 -     31993     63
   1 +   16 CDSl    33682 -     33875     1.90    33684 -     33875    192
   1 +      PolA    33923                -4.47
```

```
Predicted protein(s):
>FGENESH:[mRNA]   1  16 exon (s)    157  -  33875  2277 bp, chain +
ATGGCCCCCGCCATGCAGCCGGCCGAGATCCAATTTGCCCAGCGGCTGGCGTCCAGCGAG
AAGGGCATCCGGGACCGAGCGGTGAAGAAGCTGCGCCAGTACATCAGCGTGAAGACGCAG
AGGGAGACAGGAGGTTTCAGTCAGGAAGAACTTCTGAAAATCTGGAAGGGGCTCTTCTAC
TGCATGTGGGTGCAGGATGAACCCCTTCTACAGGAAGAGCTCGCCAACACCATTGCACAG
CTAGTCCATGCTGTTAACAACTCAGCGGCTCAACACCTGTTCATTCAGACCTTTTGGCAA
ACCATGAATCGAGAATGGAAAGGAATAGACAGGCTACGCCTGGACAAATACTATATGCTG
ATTCGTCTGGTCCTGAGGCAGTCCTTTGAAGTCTTGAAGCGAAATGGCTGGGAAGAAAGC
CGAATCAAGGTTTTCTTGGATGTCCTGATGAAGGAGGTCCTGTGTCCTGAGAGTCAGTCT
CCTAATGGAGTGAGATTCCACTTCATTGATATTTACCTGGATGAACTCTCCAAAGTCGGG
GGGAAGGAGCTTTTAGCAGATCAGAATCTCAAGTTTATCGATCCATTCTGCAAAATTGCT
GCGAAGACGAAGGACCACACCCTGGTACAGACCATAGCTCGGGGTGTCTTCGAAGCTATC
GTAGATCAGTCTCCTTTTGTGCCTGAAGAGACGATGGAGGAACAGAAGACAAAAGTGGGT
GATGGTGACCTCTCTGCTGAGGAGATACCTGAAAATGAGGTATCCTTGAGAAGAGCTGTC
AGTAAAAAGAAGACAGCACTGGGCAAAAACCATTCCAGAAAAGATGGACTCAGTGATGAA
AGAGGAAGAGATGACTGTGGAACCTTTGAGGACACAGGGCCCCTTCTCCAGTTTGACTAT
AAGGCTGTTGCTGATCGACTCCTGGAAATGACCAGCAGGAAGAACACGCCCCACTTCAAC
AGGAAGCGCCTCTCCAAACTCATCAAGAAATTCCAAGACCTTTCTGAAGGAAGCAGTATA
TCTCAACTCAGTTTTGCGGAGGACATTTCTGCTGATGAAGATGACCAAATCCTCAGTCAA
GGAAAGCATAAGAAGAAAGGAAATAAACTTTTAGAGAAAACTAACTTGGAAAAGGAGAAA
```

```
GGAAGCAGAGTCTTTTGTGTAGAGGAAGAGGACAGTGAAAGCAGTCTTCAAAAGAGAAGA
AGGAAGAAGAAGAAGAAGCACCACCTGCAGCCTGAAAATCCAGGCCCAGGGGGTGCAGCC
CCATCCCTGGAACAGAACCGGGGCAGGGAGCCCGAGGCCTCTGGGCTGAAAGCCCTGAAG
GCACGTGTGGCCGAGCCAGGTGCAGAGGCCACGTCCAGCACTGGGGAGGAGAGTGGCTCC
GAGCATCCTCCAGCCGTCCCCATGCACAATAAAAGGAAACGGCCACGGAAGAAGAGCCCG
AGGGCCCACAGGGAAATGTTGGAATCAGCAGTGTTGCCCCCAGAGGACATGTCTCAGAGT
GGCCCGAGTGGCAGTCATCCTCAGGGACCTAGAGGGTCCCCGACAGGTGGAGCCCAACTC
CTAAAAAGGAAGCGGAAACTTGGAGTTGTGCCCGTCAATGGCAGTGGCCTGTCCACGCCG
GCCTGGCCTCCATTGCAGCAGGAAGGCCCTCCCACAGGCCCCGCAGAGGGGGCGAACAGC
CACACCACGCTGCCCCAGCGCAGGAGGCTGCAGAAAAAGAAGGCAGGGCCCGGCAGCCTG
GAGCTCTGTGGCCTGCCCAGCCAGAAAACAGCAAGTTTGAAAAAGAGGAAGAAAATGAGA
GTGATGTCAAACTTGGTGGAGCACAACGGGGTGCTGGAGTCCGAAGCTGGGCAACCCCAG
GCTCTGGGAAGCAGTGGGACTTGCAGTTCCCTGAAGAAGCAGAAGCTGAGGGCAGAGAGC
GACTTTGTGAAGTTTGACACCCCCTTCTTACCAAAGCCCCTGTTCTTCAGAAGAGCCAAG
AGCAGCACTGCCACCCACCCTCCAGGCCCTGCCGTCCAGCTAAACAAGACACCATCCAGC
TCCAAGAAAGTCACCTTTGGGCTGAACAGAAACATGACTGCCGAATTCAAGAAGACAGAC
AAGAGTATCTTGGTCAGTCCCACGGGCCCTTCTCGAGTGGCCTTCGACCCTGAACAGAAG
CCCCTCCACGGGGTGCTGAAGACCCCCACCAGCTCACCTGCCAGCTCACCCCTGGTGGCC
AAGAAGCCCCTGACCACCACACCAAGGAGAAGGCCCAGGGCTATGGATTTCTTCTGA
```

>FGENESH: 1 16 exon (s)   157 - 33875   758 aa, chain +
MAPAMQPAEIQFAQRLASSEKGIRDRAVKKLRQYISVKTQRETGGFSQEELLKIWKGLFY
CMWVQDEPLLQEELANTIAQLVHAVNNSAAQHLFIQTFWQTMNREWKGIDRLRLDKYYML
IRLVLRQSFEVLKRNGWEESRIKVFLDVLMKEVLCPESQSPNGVRFHFIDIYLDELSKVG
GKELLADQNLKFIDPFCKIAAKTKDHTLVQTIARGVFEAIVDQSPFVPEETMEEQKTKVG
DGDLSAEEIPENEVSLRRAVSKKKTALGKNHSRKDGLSDERGRDDCGTFEDTGPLLQFDY
KAVADRLLEMTSRKNTPHFNRKRLSKLIKKFQDLSEGSSISQLSFAEDISADEDDQILSQ
GKHKKKGNKLLEKTNLEKEKGSRVFCVEEEDSESSLQKRRRKKKKHHLQPENPGPGGAA
PSLEQNRGREPEASGLKALKARVAEPGAEATSSTGEESGSEHPPAVPMHNKRKRPRKKSP
RAHREMLESAVLPPEDMSQSGPSGSHPQGPRGSPTGGAQLLKRKRKLGVVPVNGSGLSTP
AWPPLQQEGPPTGPAEGANSHTTLPQRRRLQKKKAGPGSLELCGLPSQKTASLKKRKKMR
VMSNLVEHNGVLESEAGQPQALGSSGTCSSLKKQKLRAESDFVKFDTPFLPKPLFFRRAK
SSTATHPPGPAVQLNKTPSSSKKVTFGLNRNMTAEFKKTDKSILVSPTGPSRVAFDPEQK
PLHGVLKTPTSSPASSPLVAKKPLTTTPRRRPRAMDFF

7.

Resultado de alineamiento con tblastn

```
RID: Z9YBNTWX114
Job Title:Protein Sequence
Program: TBLASTN
Query: None ID: lcl|Query_24331(amino acid) Length: 274
Subject:genomicC ID: lcl|Query_24333(dna) Length: 21894

Sequences producing significant alignments:
                                                   Max    Total Query  E   Per.
Description                                        Score  Score cover Value Ident  Acces
genomicC                                           114    537   93%   1e-31 75.68  Query_

Alignments:
>genomicC
Sequence ID: Query_24333 Length: 21894
Range 1: 8096 to 8317
```

```
Score:114 bits(285), Expect:1e-31,
Method:Compositional matrix adjust.,
Identities:63/74(85%), Positives:67/74(90%), Gaps:1/74(1%)


Query  1      MGKslshlplhssKEDAYDGVT-SENMRNGLVNSEVHNEDGRNGDVSQFPYVEFTGRDSV  59
              MGKSLSHLP+H+ KED YDG T S+NMRNGLV+SE H EDGR GDVSQFPYVEFTGRDSV
Sbjct  8096   MGKSLSHLPIHTCKEDGYDGGTVSDNMRNGLVHSESHGEDGRCGDVSQFPYVEFTGRDSV  8275


Query  60     TCPTCQGTGRIPRG  73
              TCPTCQGTGRIPRG
Sbjct  8276   TCPTCQGTGRIPRG  8317




Range 2: 16460 to 16606

Score:86.7 bits(213), Expect:3e-22,
Method:Compositional matrix adjust.,
Identities:40/49(82%), Positives:46/49(93%), Gaps:0/49(0%)


Query  226    LMMQVTVTTTYFGHSEQISQERYQYVDCGRNTTYQLGQSEYLNVLQPQQ  274
              + ++VTVTT+YFGHSEQIS+E+YQYVDCG NTTYQLGQSEYLNVLQP Q
Sbjct  16460  VFLRVTVTTSYFGHSEQISREKYQYVDCGGNTTYQLGQSEYLNVLQPPQ  16606




Range 3: 13800 to 13919

Score:71.2 bits(173), Expect:7e-19,
Method:Compositional matrix adjust.,
Identities:35/40(88%), Positives:37/40(92%), Gaps:0/40(0%)


Query  159    SVEVENITAQVQFSKTVIGKARLNNITIIGPLDMKQIDYT  198
              SVEV NITAQVQFSKTVIGKARLNNIT IGPLDMKQ++ T
Sbjct  13800  SVEVANITAQVQFSKTVIGKARLNNITNIGPLDMKQVNRT  13919




Range 4: 14014 to 14061

Score:29.6 bits(65), Expect:7e-19,
Method:Compositional matrix adjust.,
Identities:13/16(81%), Positives:14/16(87%), Gaps:0/16(0%)


Query  195    IDYTVPTVIAEEMSYM  210
              IDY VPTVI +EMSYM
Sbjct  14014  IDYMVPTVIQDEMSYM  14061
```

Range 5: 12835 to 12972

Score:75.1 bits(183), Expect:3e-18,
Method:Compositional matrix adjust.,
Identities:36/46(78%), Positives:41/46(89%), Gaps:0/46(0%)

Query 95   KLYVMASVFVCLLLSGLAVFFLFPRSIDVKYIGVKSAYVSYDVQKR 140
              KLYV ASV VCLLLSGLAVFFLFPRS+DV+YIGVKS YV+Y+ +R
Sbjct 12835 KLYVTASVIVCLLLSGLAVFFLFPRSVDVEYIGVKSVYVNYEQSRR 12972

Range 6: 11376 to 11441

Score:48.5 bits(114), Expect:1e-09,
Method:Compositional matrix adjust.,
Identities:22/22(100%), Positives:22/22(100%), Gaps:0/22(0%)

Query 73   GQENQLVALIPYSDQRLRPRRT 94
              GQENQLVALIPYSDQRLRPRRT
Sbjct 11376 GQENQLVALIPYSDQRLRPRRT 11441

Range 7: 15653 to 15718

Score:39.3 bits(90), Expect:1e-06,
Method:Compositional matrix adjust.,
Identities:17/22(77%), Positives:19/22(86%), Gaps:0/22(0%)

Query 208   SYMYDFCTLISIKVHNIVLMMQ 229
              S+ DFCTL SIKVHNIV+MMQ
Sbjct 15653 SFFSDFCTLASIKVHNIVVMMQ 15718

Range 8: 11710 to 11805

Score:18.9 bits(37), Expect:4.4,
Method:Compositional matrix adjust.,
Identities:8/32(25%), Positives:16/32(50%), Gaps:0/32(0%)

Query 107   LLSGLAVFFLFPRSIDVKYIGVKSAYVSYDVQ 138
              L+ L +   RSI++   + S ++ D+Q
Sbjct 11805 LI*TLIIKMHLLRSINILQRAIHSCIINIDIQ 11710

Range 9: 5896 to 5970

```
Score:18.1 bits(35), Expect:6.1,
Method:Compositional matrix adjust.,
Identities:7/25(28%), Positives:11/25(44%), Gaps:0/25(0%)


Query   185    TIIGPLDMKQIDYTVPTVIAEEMSY   209
               T  GP  +  +  T+P V +    Y
Sbjct   5970   TPPGPFPLHSLPATLPQVCSIAWGY   5896




Range 10: 792 to 857


Score:17.7 bits(34), Expect:8.5,
Method:Compositional matrix adjust.,
Identities:7/22(32%), Positives:11/22(50%), Gaps:0/22(0%)


Query   247    RYQYVDCGRNTTYQLGQSEYLN   268
               R+Q   CG + T + G  +  N
Sbjct   857    RFQVEFCGSDMTSESGSCKSWN   792




Range 11: 12045 to 12113


Score:17.7 bits(34), Expect:9.5,
Method:Compositional matrix adjust.,
Identities:9/23(39%), Positives:13/23(56%), Gaps:0/23(0%)


Query   93     RTKLYVMASVFVCLLLSGLAVFF   115
               R +L +    +VC LL  + VFF
Sbjct   12113  RFRLLLHQIYYVCYLLCEVTVFF   12045
```

9.

Promotores (5000 bases) del gen TMEM106B segun especies

>human
```
ttttttatttgctgtttccctaggtttagtgttctttaaatttgttttt
tgttttgttttgttttaggattcactgaatttcttgaatatatggattt
acatattcagtcagttttggagatttctaagttttatcttttcaatatt
tcttctatcccctttctgctccttggattccaagtacatatgtgctaaat
gatttgatattatcctaaagatatcagatgctcccttccatttgtccttg
attttgctctttatattcagtttatataatttctactgaccgtctcaagt
ttacaggttttttttccctctgttatgtccaatttggtgataatgctaaaa
agtgaacttttacttctaatatttatttctaatatttcaacttgctttt
tataaatagttcccatctttttgctaaaattccctgtctgttcatgcatg
ttttccattttttatacagatgctttaacatatctatcaaagtgatttta
aagtctctgccagataataccaaatctgggccatttctgtattttcccct
attaatcattttctctcttgaccatggctcacattttcttgctccttcat
gtctcataatctttattatatgctagatattgtataaaagaaaacagtaa
ataacatgtacctcaaaaaacatcttgccattttttctatcaagattatt
gtttggaaggctgaatcaaatctttcaacagtgagtcaaaaacccacatt
ctgggttttgttgaagttttgattcagtttatttactattgctttcaaat
```

atgcttgagttcttggcactggcaaaatctcaagattaatgtgcacctac
tagctcattccaccaaactattcaatactttcttagctcaaacacctagc
ttccaatatctccacaggttgttctgtttactctccagtcttgccggtca
tttctgtgcctcaggagaatcctgggcagcctgccaccctgcctccagcc
ttcacagagctgctgcagtacactggttaaaggcctagaatctgcaaaga
aatttcttcagcattcatgccctaccccatgcctttaacatatgacattt
gagcacttaatgaagacctttaggagagaattggagagtgagtgcacact
cgttctatcactggaacgacatgagattgtaatctatcatcctggtccac
acacaggggtgtcattcggcaggtctcccttcggccctgtttatggtgca
ttcctccttctcccactgttttgtaggtgattaaagcaactctaattcta
ttctctcttaggaagctctccttatcttttggatttaagtgtacttaaat
ttgcgtcctcatttctttgataagttttttaaaaatgtttaaaaaaaaaa
aaaccttctctagcttgtgtttgttgtgaagaaacagtcttttataacat
aaatcctaattggactctgtgttttccttattttatattcggaacacctc
ctcaaggatttctggacccaattctcctcttttcagcagcagtgatgatc
acttagcaaaagagtgaaaataccctcttttccccccctaaccagttaac
tttactcaccgtcatcacttatctttgtgtagtaaaagttgaggacagaa
ttcactcttccaaatttaatacagatattcatgccacagtggttaatagc
tgtagctttgcagtcaaaatatccgagttctagttgctagctgtgtggtc
atgtgcaagtcataatacctaggatcctctatttccactgtttgaaataa
gatatgggataataacagaacctaatccatagcatttctataaaaattag
atgatataaactactttacacaatgcctggcacaacaaattcattcattg
gccaagtgttattgaactcctaaaattattcttactagaaattatgatta
tgacaaaaaccctacaaaagatgacttgaatatttgggaaagaaagtttt
aacacatgggaaagattagctccatgcaggtgtgaagcaagatgaatgag
acagaggaaatatagaagatcctgcctaagaaggagtacagagacataac
caagagtttcagataggtaagttatttgcagaaattaataaagacctctg
ctttaaaataattggaatgacaaagatgtgtcaatagagtcaggagttt
actgcctgtgaatgtaagtcaggtggcacagcgggaaatggtagttttgc
taatggtagttttgctatgtatctggggcactatctctgaaaatcagcca
gacaaacttcaatcttttcaatccctggtttccaatgtaacctatccagg
tttctaacataatctgaccagcatatcagtggaaggggggtgagctgagag
agacttatatattaacatttgacaaattatatgtttcaaaataattatgt
tccttttctgcttataattttttaacagctaatttctagaaaagtaattttt
tgcaaaattttttagttatggctgactgctattaggatttgggtgatttac
cattaaaagagataaaagagaattgaaatgagagataaggctgttttcaa
gatttagtacaacctagagaaatcccttagcttcagcatcatactaaata
taaccagaaatcattaagaaagaatgttcttctattcaaacttatcccag
actgttaaaattactggcctttatatctgatagcacctagaccactactg
aactcggaactcaagaaatcactgttaatcaaataattggaagctttttca
gaaaattttcctttcaaaatgtgcttcaatgaaaaataaggcataggaac
aaatgtaccaaacattcaatatctgtctttaaagtacttccttgttatat
ggatgataaatttgtaatcataatatctaatatttactgattgctctgca
ggtatcagctcccatccataaccctttgagacagacattattactatctt
cattttgcagataataaaaattaggtataaagaaggctaagtattttgcc
cactcttacagagattttgtatatctctgtcaggctgggctcattttcag
gccatatggctcccaaggctgtgttcccaaccctgctactgaactgcttc
ttacagcaacattctcccaggaagacttattcctatgtgaattccaggaa
tttggataaggaactgtgctatatggccagcacgtttttaagttaaatgat
gaggttgagaaaatactaaatactcccttaagaattaaagaaaaaagtaa
cttttttcacttaagaataatatatattacattatagaaccatttgcctat
ttaacaataataaagaaaaaacatccatgctctcaaactacaggcaactg
tttatattttgtatatttctttccatttcaaggcagatatgtaatatatt
tatatctatatgttatctatatgtgaatatagtttacatcctattacaat

```
acctcatatcctatactttgcatcctgattttaattttatatcactccgc
gtataacagccttataagatggtttttactggctttataaaaactcaata
gatggctattccataattttaaaaactcagtaattccaaggggaacatca
tatctccacaggtggactgtcaataattttgagaaaaattttcccccatg
attctgatttgttttaaccaacctattaattttgaaatttcaaacagtgt
taatggaagacaccctgacccttccctttggaaaatatgggctaaataaa
agcggttgaaaaataatgattttatcaatgtcccttgttggacattcagt
ttatttaaatggatacacctcatacccaaatctttttagagttctcatta
ggtttaggatacttctgaaagggtggaattgctggaccaaacgatgtgaa
cattttaatagtcttgatacatactgactgtctgccagaaaacttaggag
ttttaaatctccgacaacattgtatgagaaagctcattatctcacatatg
gatctacagtaatcctcttaaattgtatatcgttctgacttgtatttcct
acttttacatagcttcatttcccaattcctgttagaattccctgattcta
acaggatcacactgatcctgttagaagagagaaatatatgactctcactt
ccttgctcctgtgactgtaagaatgcctaccattcattaaacgttaaact
ctccgcaggaatgagactagagtgcccttaatattattaccggcatctca
ttgatccctcaaaacgactccttgtcgttgttcttatcttctaagtgaaa
tgacctagacgttggggggagggcagccccaacccacagcgaagagcttaa
gtactgggatccgagacgggatttcacacacctgtaactagcaccagagt
taaggggtggggaggggcagcgtgggcaaagcgaaacgaaaccgagcaaa
aacaaaaactacggccttgccgcttgctgtctccttctagttggccagtt
tctaaattaaaatagaggtagctttgctgctgtcagcggcgtgtgtacgg
accagtttattccattcctaaggaccagactgtcagcgacaactcgcatt
gcgtgagcgaggaggctgctcagctgaaagctgaggcagatggtgcacat
ttataccctttgcaaaaggaaatggtctgatatgggaaataggaggagcg
actgttcctaggtaattcgtggctacagtggacactttgaagtctggggg
acgcgaaggtaatgaaacctgaagagcgcacaatctcacacgatcgacgc
ctctgggctcaaacaggaggaagaaggacgcatgcgtcacgcgcacgcca
ggggtgtccctccgggagcgctctgcgcaggcgcggacgcaggttacagc
agcgcttggcctctgctgatgccgtcgttatcctacccctcccccgtccc


>mouse
tgtgtgtatgtgtttgtatgtgtgtatgtgtgtgtatgtgtgtgtgtatg
tgtgtgtatgtgtgtatgtgtgtgtgtgtatgtgtgtttatgtgtgtgtg
tgtgtatgtgtgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgttttct
ttataaagctcccctgagaaatgtttgttgtcttagttcagatatttctg
agctgtacttctgatctaatcagtattagggtgtgcatacaataaaccat
ccctgtctgactgagatcggtgttcattcgtatggtttgtagtgctactc
tcagccctcaacaatatattttaattctttatttggacaaaatatctgct
attccaagtttaaaaaaattattcatgccacagtggtgaataacatgtgc
ttttctgatttctagttatacaatttgatagctgtgtgttcatatacaag
ttacataacctagaatcttctatcttctttttttcttctgatgtgtgcat
gtttataagtatgcagggatgtgtgtatggaaatttgggacaatacaggg
tttctttttcaaacattctctaccatagctatagtaagataggatctctc
actgagtctggacctcaccaatttgccagtgagcaccactgatccttgca
tctctgttttccagagctgagatcacaaataccaacctctaggcttaaat
atttaatgtgtgttctggtgaattgaactagtttcttcatacttatggaa
caaacactttacagactgagtgatctctcccgtctttagtttctctgttt
aaaataaaattcagataataatagaatgtaatccaaagacttcccataga
aattaaaaagagatgaactgaaatacttcatatactgcatgacacaacaa
actcattcatttattaaacatgacctattttgtaacattttttcactaac
ttattatgacaaagactctgtaaaagtttaaaatcaataaagttagttat
tggtttaaacatcttaatacatacagagaaatttgcctctctaggtttga
```

```
agcaatgaagatgtaagtcaggatatacagaagattcttatgctaaaaat
ggtttcttagggatgtgaaccaaaggctcaaagggatgagatgtttgcag
aatgtaacaaaggatattgaagagaggaaacacaaaggtgtagtcacaga
accaagaacttgctatcagtgaatgtagggcaaatgaccacagggaataa
tacttttgctatgtatctgggaaactatcacaggaattcagccagaaaca
ctccagttctttaacttctaattttctattatagaaggttgagtaaaata
agccacaccaaaaagacaaaaagaaaattctttatatacacagagtataa
gtttgtgagtgcatctataaaaataattaaactcagtttgctaatgttac
ccagccagcttgaaacttaaaagtgcaggctacgacagacctgcatacta
acaataaaccatatacttcagaacgattgagagagtaaattatgaattat
ttgtctgtttataagtttagcaacaaatttctaaggaagtgacatttga
aatttggctgccagctgttaatatgtgggtgagtctcattcaaggaaatt
atgagaaacatggaaggaacatgcatgttttgaagatttaacctagcata
gaagaataccttatttcaacacccacgaaaaagaggtttaatcagatcac
aacttctgttcagttacctcatactatttaaaacattgtttctaatatct
aatagcatgtagactgacatggatctcaaaatttcaaagaaaaagttgtt
gacattttccaaaattttcttttaaaatattctttaatgattaaaaaaa
gaagaattggagaaagtctgtctagcattaaagatctatctttaaaatac
tttactgtttgtggataaaaactttgtcacaataatatctaattgttgtg
ttttataaaaatataaagagaaggaatatgtttggtggaggtaaggtagg
tgtgggggaaggaggtgcctctgcagggccatgctgaggtattgcttccc
ctggagggaccagccatgcaatgtattcagggcatggggaggggagttga
gacagaaaaaggtaggggaagagagagagaaagagagagagagagagaga
gagagagagagagagagagagagagagagagagagagagagaacaagagt
agaggctgtgaccacttgcagagagaggggaaggggaatgcggagcctgt
gtaaaggccaatgggcagagctagagcaagaaggcaagagagagagagag
aggaggggacaagcagccccttttatagagtcaggcacacctggcaacag
gtaactgcggggcagagcctgacaaaatgccaacactaacagctgctata
cttacaaccttttaaaacaagctccaacagggccataccttctaatcgtg
ccactccttgggcagagcatatacacaccatcacattctgcttcctggcc
cccataggcttgtccaaacatatgaatctatgggggccatacctaaacat
agcataataaaaccatcacagtgtatgagcatgtaactgaggcactatag
aaaaataatatagaaagacaggtgtaattatggaggtgggggaaaagacc
cttactctggaggctagttggagtctgtcttcagttgagagttaaggatt
ccagcataattaacacttactgtgatgttgcagacatagaaacaatactc
attttaaaactgtaaaaataacaacaaaaacaaaacacccaatagaaaga
gtgcataagacatcatgtatgttctcagtccttgatattttaagaagctt
tcacaaacgtgtcttttttttttttttttttttttttttttttttttggt
aaacttttatcttaactttctgcacccctaggtcgctttacattgtgagt
tgtaatcctattcaagactccacaggagctctaatcccatttacgttcct
taatgaacctttaaagattactgaaaacccaccccaaaatgtatgtgtgg
ctactataaagacaatgaagtgtgattctctcttgaggagcccacagcca
tgttctcagatgagatttatacattctctgagtccctttcatttctattt
aagtctatgcttaaactctatgtttaaatgccttttaatacacatcaagt
gggcttcatattggatcttattgaagggaaggaaggctgttctcacagca
agttttagggcaactagtgtgcaacccaaggaaccacatctcaaaagcaa
tcattttatcaagtgcactgctggtgtgacttaatcaggtatttagtttta
tgtgtatttagtttccttgcagtggtttgaataggaacagcctctggacc
ctgagtgtgaaggattggcccaaagggagtagcactattagggagggtgt
ggcttaaatggaggaggtgctgactagggagagggccttggggtctaagg
agctaattcacttccctgctgtctctgatgtcgaactctaagctcctcca
gtatcatgtctgcctggttgctgccatgtttcctgccaggagcataacgg
gctaaacctctgaaactgtaagccaaccccaaataaatgtattcttttata
agagttgccatggtctacttcgaagctattgttaactctattgaaatcct
```

aagaaactggttacacactgcacacaactttactttagaattctcattat
aatttaggatgaatctcacagagtgaaattgccggacctaaggaggtgaa
tgttttaaatagtcttaaaacatcctgcttatgtagatggaaaatacaca
gctctcacctcaaaaggacaaagagtttattctaaagccaatatgaggaa
ccatgacctgaaaacaaagacttgggctattccaaataccaatgtcccaa
tgtgataacaatttcatgaaacttttatatacatagaagatggtcataaa
tcaaaatactttccaaatacatacaaggcagtttagaagagcgggacaaa
attgggagctcctcgcccagtgcagttgctttttccagtagagaaggaa
aaaaaaaaaaaaaaaaagaaaaagaaaaaaaaaaaaaaacacaattaag
agatgcttacttcccagttataactctgttaattttctaggaatttcacc
agaaaccccatttatagggaagctattaaactcctttagcaaccgcatct
gttcttgctttcaatttttatcttattataatcccatccaggaatcttatg
agccacttaatacatcttgattatcaggaagacagaaaagaggtcataac
aacctttgactataacactggcgagccactaaaatgagccactgtagcct
caactcactgtaaattatgactggacaatgaggtgcatggattgggacac
tcctttaacaactcacttggtcatgagctttggggtacaactcttattgc
atgcactgaataatgtgaatttaagtacaaaatcaaaacaagattaaacc
aaggcggggggcggggggtgcgtcttcccggtgacattaatttctaaatgaa
aactcaataacaaaattctggcctgcaactctgccagaaatgtgcaccaa
ccggctcattctaaccccaggaaggaagttttgtgtctgcaaatctagac
ttgagaacacgatcgctcaaaattgcgacggtgcacattcctccccttta
caaaaggaatcagcctggtggaggggaaaagggaaaaccgattattcgta
agccaggcgcaattagtactattttgaaactttccagtctagagaagaga
caggtgacaaaaccagaagctttcaccctggcaacttcccttcaaaccgg
tgacagtgcgcatgcgtagtgtccattccagggcctagcccctccttcct

>rat
gccaggaaacaccagaaggcagagttgagcaggcaggggcagagagctga
ggcaaaagacaccagaatgaaagcaagaaaggaactgggtgaaagtatgt
ctgcaccagtcccaggcagggcagggccactggttttgttcccttccctc
ccctgcctgctgccctccctgtcgcccagatggccagatacagcagtggt
gacagcagcagcagcagaagttaaaattattaactagaaaatggcagcca
ggccagatcaccaggtggtcacatggtgaacttggggagtgagctctggg
ccgcgactgtaggtaaacagttcacctatgcctaggatttggcaactaca
agaaggaactacctcagaggagagaggaaaatctcgcccaagagaaaatg
tcctagaatgagggtcagtgaaaagaccaagtcataaagaccgatatctc
atcctcaagggcagcacaggggttgctctgacaagtaggaggagggggagg
ttctctgaagtttgggctaaagagcaaaggcctctctatgtggtggtggg
gggtggggggaggtgggatgtgaacctgaggagtttctaatccaagtat
cggctccaaatgtgtttttataaaaatatagagagaaggaatatgtttgct
ggaagttaaggtaaggtgtggaggaatggggtgcctttgcaggcccattg
ctggcgcatcgctttcccacctgaggaccaacaccacaggacagttgcta
tggttcctttagtagctcgatatatcaggcatgagcgacgtgggttggtg
ggaggtgaggtgtggagtgttgagggtgggttcgaggggggtattttagtg
ctaggttttggtgtcagggtttaggactatttctgaggagagagtagggc
gagtgtgtgtgatctctgtgagagagatgaggtagcttcttctttcttgtagg
tgaaggagtgggtggtgactgtttaccttgtggggtcatcagtttgtga
catggtgttgctggctgatttttctgtttaattttaactatctgtgctgct
gtctgtgacctgctgtatctgaggaagtggcatggaggggggagagggagg
ggcgaggagggtgggagggacatggaagatccggagccatagatgataga
tgggtggtgacacagatgaactatgaggggggtacaaggattgatattact
gatggtttacatttggtgtttttgtagtttgccgagagagggaaaggggt
aagaggacagcgctagagtaagaaagcaagaggtgggtggggcaagcagc

```
ctggtttatagtgagtcaggcacacctgactgttgccaggtaactgggggg
gtggagcttagacaaaatgccagcactaataatgctagacctacgaccct
gtaaaacaagcattattgctatctcaattctaaaagcaatagacatatgt
ataaagaatgaggttttctttattttccaggctgtttacctactttgtag
cttagttcatcctcaggcaagcagccttctaaggcaataccactaaattt
acaggacttaccttctagagattcattattgaattctagtcacgtggata
aagactgccaaaatgacaatattgatagagacaattctgtgaaaatacac
ccttagatattcctttaaggaaggacttattttgcattcaagaataatac
aggtcagcctatagaaataggcacaacttcaacattgtatacaataaaat
atttattttaaagtctctcaaaaggcaactggtggggatttctgtccct
tgtccaaaaaaaaaaaaaaaaaatagtctaattaagagattggaaattgt
taccctgaccccattccagcttcctcttagctcggggtcattataacctt
catgaaaacagtttagatgacaccgagggcaaagtttattgctgaagaat
gtgaggcattccacctactgagcatggacaaaggtcaagagaccactgca
gaggaaagctggctaagggattagggtgatgatgtccagagaccttgtgc
taactgacctacatgttcaggaaatctttgtctgaaacaacttcttgaaa
tggtatttccctatccaggtgtgcttcaggtccattttgagtaactctta
aggtatgtgtgtgcacgtgggtgtcttaggggtttactgctgtgaacaga
catcatgaccaggacaactcttacaaggacaacgtttaattgggactggc
ttacaggttcagaggttcagtccattatcacccaggtagaacatggcagc
ctcaaggcaggcctggcacaggaggaactgagagccttacatcttcatct
tcaggctgctagcagaatactggcttcagctaggatgagggtcttatagc
cacagtgacacacctactccaacagggccacactttctaatcatgccatt
ccctgggcaaagcatatacaaaccattaatcctacttcctggtccccata
ggcttgtccaaacatgtgagtaaacacagcataataaagccatcacagtg
ggtgagggtgttattgaggcactgtacaaaaataatacagaaaggcaggt
gaagttatggggtcgggggtggggtgctccttattctggaatctagctga
aatctttcaccagcttagagttaagggttccagcacaattaatattcact
gtgatgttgcaaatagaaacattatccatttaaaagctgtaaaaataaca
acaaaaacaaaatactcaaaagaaggagtgcataagacttcatgcatgtt
ctcagtccttgatattttacaaagttcttaagaacaatacttttctttta
ggtaaagatatcttaatcttctgcacccctaggttgctttatattgtcag
ttgtaatcctactcaagactccacaggagctacccgccagccaatcccat
ttaaattccttaatgaggctttaaagactactgaaaacccaccccaaaag
gtatgtgtcgctacagtaaagacaatgactatttcctcagaagtccacag
ccatgttctcagatgagatttatacattctttgagcccctttccttttca
taaaagtctgtgcttgaaatatatgtttaaatgccctttaatacacttca
agtaggcttcatattggatcttattgaagggcaggaagactgttctcgca
gcaagttttagggcagttagtgcaacacaaaggaccacatctcaaaagct
attaattcacccagtgcactactggtgtgactaatcaggtatttagtttta
tgtgactggtcacagtggtttgaataggaatagcctcctcaacctgtttg
aaggtttggctcatagggagtggcacagttaggaggcgtggctttagaat
gggtgtggctttgttgggggaggtgctgactaggaagtaagctttgaggt
gtaaggagccaattcactttctgttgtatgtgaactctcaagcagaacca
gcaccatctgcctgcaggaagataatggactaaacctttgaaactgtaag
ccaggcccaaagagttgccgtggtcgtggtccacttcgcagcaatagaaa
tcctaagatattggtcacccactgcacgcaactttactttggaattctca
ttatcattcaggatgaattctacagtggacctgaagagatgaatgtttta
gtcttgaaacttccttctgtagatggaaaatacacagctctcaccttaaa
aggtctaagagtttattttggagccaaatgtgagggaccatgacctgaaa
acaaaaactcgggctattctgaataccatgtcccaacgtgtaacaatttc
atgaaacttatacacataaaagaagatggtcataaattaagatactttcc
aaataaaaacaaggcagtttagaagagcggggcaaaattggtcgctcctc
gcccagtccggttgcttttttgccagcagagaagaaagaacaacacaatca
```

```
atggatgcttaacccttttcccacttagttatagtccattaattttctagg
aatttcaccagtaaccccatttataggaaattattaaactttagcagct
gcatcagttcttggtttcaatttatcgagattataattccatccagtcct
cttttgagccacttaatatatcttgattatcaggaagacaggaaaaaaaa
aagtaaccttttaagataacactcgagattcactaaaatgagccacagta
gcctcaactcacagtgagttatgactggacaatgaggtgtgtgggtggag
aggtcctgaggatactcctcaagaactcacttggtcatgagtattgggg
tacaactcttattgcacacacagaataatgtgaacttaagtacaaaatca
aaacaagattaaaccaatgcggggggtgggggtgcgttttcccagtgacag
aaatttctaaatgaaaacaataacaagatgtgtgtaactctgtcagagat
gtgcaccaactggttcatccttaccccagggaggaagtcttgcggctgca
aatgcagcattgaaaacaaggccgctcaaaactgcgacagtacacatttc
tccccttcacaaaagaaatcagccttatggaagggaaaaggaaaaccgat
ttttcgtaaggcagacgcaataagtactatttctaaacttgccagtctag
agaacagacaggaggcgaaaccaggagccttcccccagcaacttcccttc
agggtgccaaagctgggaaagggcgcatgcgtactacctattccaggctc
ccagctcctccttccatgggcggactacgcaggcgcacgcgcaggctccc
ggggtacccggctcagctgatgcagttgttgtcctaccctcccccgccg
ctctcccggcctcactcgggccggcccttagccccgccctgggaccggga
ctccgtgaccccgtcctgcggctcgccagcgcaggagcatcccgggaccg
```


>chicken
```
cagctgcaattgctgctatgtttcttaggtaaagacaagtggtttggaat
gtgatttgtaataaccgtgttctgaaagaagctattcagaagcattaaac
tataatcactgggtcattcacccaactgatcagaagatttattattcttt
gtctagatctgtgggttgttggcattttgtatctccccatcaatatcca
gtttcttgaatttccagtttgttttttttttcctcctccagaagaaaggtg
agaatatacaaacagaattaagatgggctattttttcttcaactcctggga
gcacacaatgcagcaagcaagccatggataatacatagcatcagaactgt
ggaggactgagcacaatttctaattcattttcatctatctacagaacat
aatcatctcattcgtcaactaatgcaaaatttaaggtttgcacacgtaat
gaacaaggagaaactataaagtatacactgcagaagaccatagagtgttg
caagaatgacagttgtcggtagctgtgcaatatatcagaaatacatcagt
tataaaaggctgtttatgtttttattcattcaaaataactcagagcacagg
ccgtatcatcagaaaatacatagcaatgaaagtacctgtgacacaccatg
cttaagcattaaatgcatctaaaaatgtttaccacagaatatcattaaaa
tcagaaattctgaaatacagagaattcttcacataggtggaaaaggatgt
ttcaaaaacagaactcttaaaaatccgaacagaccaaataaagacagtgg
aattttttgatattcatattttttaattcttgtaattctaatttaattagta
caggcacagctcaaaggaacactggggtgcagtaacaggtatgttgctga
tcagataacaggcagggtcatgagctatgttgtcctgtgcctggaataga
tttgcctcagattacatacacctacatgagcgtcgcaactattgtataga
ttggtttgatgtgttctgggattaggatgatttctgaataaaataagact
gttgcccttgtaatagatcaagggcaaaatcaataaacaattatagttgt
tcttaattttttatcaggatagcttatacctctcttcctgagtttctcaa
aaagaaaatcagagtaaagggaatcagcattttcacttggttgtcccagt
gagcaggtgatagaagtctaaaaaaattgttcctgtgactgatggagatg
aaggaagaaagtagtccacataggcagctgtaacagatgtgcaaagccta
ctgtgcccttatataaaccagtggtaataaccagtctaaagcagttacaag
ctttcttaaaggccttctcccctatcagcaactggtcagtgcccacattc
tctcacattgctgttttatattctagattagatagaaatcattgctttta
tgttgctttatttatttattttttgtggggtttggtagtttaggttcttttc
tgatttgtttgatttttttgctgaggagaagcttggatatatgttagagat
```

aaaaactgagctaaaatgctgtatctacctcagcattcagatagcatgcc
acccattttcaggctagcagactgtatgaaaacaagactcagtaactttg
aaggtgttagccagtttcagctgagtcacagagataagaaagtttataca
aatttcataaagatagataaccaggcagaagagagacaagggattgaatg
tttggaagccattgtgtaataactttctatgcagtaagcacatttctaac
aaattctatttgtgaacatggcagcctgcatgttactgtcagaatactct
gcaggaaaagacatgagttgtatgcagttctttttgttacaatttgggtaa
acagtcatctgagaggtagatctataggaaatgaggctgcattagtttca
gtgctcaccaaaccagtagctttatttctgaagggtgaaataaatgcag
tcagtttgaagtttcttcggataagcaatgcaatgaggttgaagaagaaa
taaaattaattagtctctcaagcatctgaggtcatgcagatatgcactgg
gagtctgcacattggtgcagttgattgctcttgagattgctaattccagc
tcttgcaagaacctgactcacttgtcatatccgagccacaaaactcaact
tgaaacctgattcatcacttggctttctttactgatatactactggaagt
ataaaatgcacctggttactgatctatgagaactaagagcacttgcagct
gtcccataactcagttcaaaggtgatgatctcaaattccctgttcctatc
atttgattgtttccatttcatacccctcacagtgccatttgtgctgtaca
tcacttgcctatcatgcatttttattattttttttgtatctgtgtttaagt
atgcaatcccagtccagtggtggttagcacagtccaatcagtttcaggaa
ctggcaatattggatgctgtgcagtgtaaaatgtgtaaaatactccttct
ataaccaaaggtattttgtttcaatcagtccaagctcagtggcccacttc
agtgacgctttcaggcatctcatcccatatttaagcacttctaatgtact
tgtaatgtgtacatcttgttatcggtaataccacatatttctcatgtttt
tcatgcttgcatagtgagcgtaagactttacaggaccttcatattattgc
ttattctcaattcttatcacataactttagcctataacaaataaatgata
tttaagacaactattaaggagcttaaggcacataattggtgagcaagatt
tttctagtgtttcaacttaaattttctttaacatgcactttctaaatgaa
aaacaaagcataaaacaaagaggaggaggggggaaaggggaggtgggggggg
gagagtaggagagagaggagagagggaggggaaggaaggaggaaaaaggcaaat
gaagggaaaaagctttcctaatgtatcaaatctacagaaggaaaggctat
ggcagtttggaaatgtaattaatacaaaacagtttattgtttaaagtaac
agactcaaaacctcatgcagatttcaactgttgggatctatcatggaaag
agaaacaatgtctcaggcaactatgtaacatcctgatcggcatttggcct
tctagataagtttcaatgttggatgaaaactgctgcaatcaaaggtgtaa
aactgtaatttgacccaagactgcactgtatttcaaacttgctaggagaa
taaatttatattgagatactctgtcagagtaaataactctgagtagccag
aattcagtgtgaattcccaaagccattccaaaaacctttctgctagactt
actcctcacgtctcctgtagtataatgattcagatcaatctgaatgactg
acattattatctgatactattcaaaaggcggagactgtgttctcattttc
tgtttggatagctggaaactctgatactgtttgatgtggatcaatataaa
cttgagactgaaccaaaagtcaatagccagtgaagccagtagccaggctt
accaaattcacctcatctgacacttgttgctgaacatgtcatgttcatac
aggttccaaacaagccagggttctgttcaaaggttttttatgtataatttt
gtatgtaggaacccacagagtaccatggtcatccatgctgcactggcagc
tttgatggggacctcaggaaacttgaagacctgcaaaagtgttggctgag
gagggacaaagaagaccatgacatggcagctttgccactactttcagaca
cttgaccatgcctgactgcagaggtggcatcaccagcaagcagcacacag
gcctctcctttacatgaactggctaaccaggagtgaggatgattgctaaa
caagagcttttaaactgccagaaaaaaaaaaaaaaaaaaaaaaaaagaaaga
aaaaaaacccaggaagttccaaggataaagaaaataaatgggggtgacat
tggctcaggtcatgatttgtgcagaccattttttttctgagtgcgtatttt
tgcaggtgagttattttctttgcctattctttcctggacacagagcccaa
catgagcaatactgccagcagggctcgcacctgcagcagctctggactgg
tgttagaggtgtcgtgcctccagagcagcctggagccatgccctgtcctc

```
atgctcaccccatgctgtcataatgcctatttccttatcctctcccttga
agtctgcccttatgtttgcccttggaattccgtacatggcctccttcagg
ccctccctcccacagccccaggtctgatccatttctgctgttggcctgag
gctcccaagctgggctgagttcctccctagccctccaaactccttcatcc
ccagtcgtgccccactgcatggcctcatgtcccttggctgccccacactc
ctcgagctctctccctgggcagccatttcacagaccagctgcaggcacac
agggagagcaacaacctcaagttcattccattccagtgtgctttattttg
ttccggtaaaatgcctggaaatcatctcactgctcatgccctgagtacac
gtgggttgtgcatgatccttgtaaaacgcgatgtgagttacagaaccaaa
tggccactcccaacagaggcactgtgcctgtttcctactgtctttatgcc
ggtctgtcacagctctccttgccctcagttctcacaggagcgcagtgcgg
taatttacctccactatcaccaatcaaacagagaagagaaatcatgaca
gtgcctactatgacaacaaacaagcccacctggccttccctgtcctgcac
agcgcagggcaccccgctcctgccgtgcaacagccaccgccgaagcgctc
ccccccccgcagcggcaccgccccccgcgcctgcgcacggcccggatct
```

9.

Proteinas según los diferentes alineamientos

>GENEID
```
MAPAMQPAEIQFAQRLASSEKGIRDRAVKKLRQYISVKTQRETGGFSQEELLKIWKGLFY
CMWVQDEPLLQEELANTIAQLVHAVNNSAAQACVWFFSRIKVFLDVLMKEVLCPESQSPN
GVRFHFIDIYLDELSKVGGKELLADQNLKFIDPFCKIAAKTKDHTLVQTIARGVFEAIVD
QSPFVPEETMEEQKTKVGDGDLSAEEIPENEVSLRRAVSKKKTALGKNHSRKDGLSDERG
RDDCGTFEDTGPLLQFDYKAVADRLLEMTSRKNTPHFNRKRLSKLIKKFQDLSEGSSISQ
LSFAEDISADEDDQILSQGKHKKKGNKLLEKTNLEKEKGSRVFCVEEEDSESSLQKRRRK
KKKKHHLQPENPGPGGAAPSLEQNRGREPEASGLKALKARVAEPGAEATSSTGEESGSEH
PPAVPMHNKRKRPRKKSPRAHREMLESAVLPPEDMSQSGPSGSHPQGPRGSPTGGAQLLK
RKRKLGVVPVNGSGLSTPAWPPLQQEGPPTGPAEGANSHTTLPQRRRLQKKKAGPGSLEL
CGLPSQKTASLKKRKKMRVMSNLVEHNGVLESEAGQPQALVRWEHPQASSPQRHSLASMG
LHCLLRGRVGAGGQASGLSSS*
MKIKGSSGTCSSLKKQKLRAESDFVKFDTPFLPKPLFFRRAKSSTATHPPGPAVQLNKTP
SSSKKVTFGLNRNMTAEFKKTDKSILVSPTGPSRVAFDPEQKPLHGVLKTPTSSPASSPL
VAKKPLTTTPRRRPRAMDFF
```

>GENSCAN
```
MAPAMQPAEIQFAQRLASSEKGIRDRAVKKLRQYISVKTQRETGGFSQEELLKIWKGLFY
CMWVQDEPLLQEELANTIAQLVHAVNNSAAQHLFIQTFWQTMNREWKGIDRLRLDKYYML
IRLVLRQSFEVLKRNGWEESRIKVFLDVLMKEVLCPESQSPNGVRFHFIDIYLDELSKVG
GKELLADQNLKFIDPFCKIAAKTKDHTLVQTIARGVFEAIVDQSPFVPEETMEEQKTKVG
DGDLSAEEIPENEVSLRRAVSKKKTALGKNHSRKDGLSDERGRDDCGTFEDTGPLLQFDY
KAVADRLLEMTSRKNTPHFNRKRLSKLIKKFQDLSEGSSISQLSFAEDISADEDDQILSQ
GKHKKKGNKLLEKTNLEKEKGKQELQGALGGGCLMTTRDLWFLPLSPKISGNGTISVPYV
FINGQKEGFQSQLGMEEVGPDDKGSRVFCVEEEDSESSLQKRRRKKKKKHHLQPENPGPG
GAAPSLEQNRGREPEASGLKALKARVAEPGAEATSSTGEESGSEHPPAVPMHNKRKRPRK
KSPRAHREMLESAVLPPEDMSQSGPSGSHPQGPRGSPTGGAQLLKRKRKLGVVPVNGSGL
STPAWPPLQQEGPPTGPAEGANSHTTLPQRRRLQKKKAGPGSLELCGLPSQKTASLKKRK
KMRVMSNLVEHNGVLESEAGQPQALAAHLNLPEPPVCRQRHWAAHTSESQVRDPVSLWVA
VSCCTRNECPGPASVVLCVKPELCRMEGLSASAVRKTAGRRGSSGTCSSLKKQKLRAESD
FVKFDTPFLPKPLFFRRAKSSTATHPPGPAVQLNKTPSSSKKVTFGLNRNMTAEFKKTDK
SILVSPTGPSRVAFDPEQKPLHGVLKTPTSSPASSPLVAKKPLTTTPRRRPRAMDFF
```

>FGENESH
```
MAPAMQPAEIQFAQRLASSEKGIRDRAVKKLRQYISVKTQRETGGFSQEELLKIWKGLFY
```

```
CMWVQDEPLLQEELANTIAQLVHAVNNSAAQHLFIQTFWQTMNREWKGIDRLRLDKYYML
IRLVLRQSFEVLKRNGWEESRIKVFLDVLMKEVLCPESQSPNGVRFHFIDIYLDELSKVG
GKELLADQNLKFIDPFCKIAAKTKDHTLVQTIARGVFEAIVDQSPFVPEETMEEQKTKVG
DGDLSAEEIPENEVSLRRAVSKKKTALGKNHSRKDGLSDERGRDDCGTFEDTGPLLQFDY
KAVADRLLEMTSRKNTPHFNRKRLSKLIKKFQDLSEGSSISQLSFAEDISADEDDQILSQ
GKHKKKGNKLLEKTNLEKEKGSRVFCVEEEDSESSLQKRRRKKKKKHHLQPENPGPGGAA
PSLEQNRGREPEASGLKALKARVAEPGAEATSSTGEESGSEHPPAVPMHNKRKRPRKKSP
RAHREMLESAVLPPEDMSQSGPSGSHPQGPRGSPTGGAQLLKRKRKLGVVPVNGSGLSTP
AWPPLQQEGPPTGPAEGANSHTTLPQRRRLQKKKAGPGSLELCGLPSQKTASLKKRKKMR
VMSNLVEHNGVLESEAGQPQALGSSGTCSSLKKQKLRAESDFVKFDTPFLPKPLFFRRAK
SSTATHPPGPAVQLNKTPSSSKKVTFGLNRNMTAEFKKTDKSILVSPTGPSRVAFDPEQK
PLHGVLKTPTSSPASSPLVAKKPLTTTPRRRPRAMDFF

>RRP1B
MAPAMQPAEIQFAQRLASSEKGIRDRAVKKLRQYISVKTQRETGGFSQEE
LLKIWKGLFYCMWVQDEPLLQEELANTIAQLVHAVNNSAAQHLFIQTFWQ
TMNREWKGIDRLRLDKYYMLIRLVLRQSFEVLKRNGWEESRIKVFLDVLM
KEVLCPESQSPNGVRFHFIDIYLDELSKVGGKELLADQNLKFIDPFCKIA
AKTKDHTLVQTIARGVFEAIVDQSPFVPEETMEEQKTKVGDGDLSAEEIP
ENEVSLRRAVSKKKTALGKNHSRKDGLSDERGRDDCGTFEDTGPLLQFDY
KAVADRLLEMTSRKNTPHFNRKRLSKLIKKFQDLSEGSSISQLSFAEDIS
ADEDDQILSQGKHKKKGNKLLEKTNLEKEKGSRVFCVEEEDSESSLQKRR
RKKKKKHHLQPENPGPGGAAPSLEQNRGREPEASGLKALKARVAEPGAEA
TSSTGEESGSEHPPAVPMHNKRKRPRKKSPRAHREMLESAVLPPEDMSQS
GPSGSHPQGPRGSPTGGAQLLKRKRKLGVVPVNGSGLSTPAWPPLQQEGP
PTGPAEGANSHTTLPQRRRLQKKKAGPGSLELCGLPSQKTASLKKRKKMR
VMSNLVEHNGVLESEAGQPQALGSSGTCSSLKKQKLRAESDFVKFDTPFL
PKPLFFRRAKSSTATHPPGPAVQLNKTPSSSKKVTFGLNRNMTAEFKKTD
KSILVSPTGPSRVAFDPEQKPLHGVLKTPTSSPASSPLVAKKPLTTTPRR
RPRAMDFF
```