

# HW1\_forcycleserver

September 24, 2021

## 1 Goal

Retrieve the sequence of all proteins encoded in the human genome.

```
[1]: import numpy as np
import pandas as pd
import csv
```

```
[2]: import Bio
from Bio import SeqIO
```

```
[3]: from Bio.Seq import Seq
```

This is the human genome annotation <https://genome.ucsc.edu/cgi-bin/hgTables>

```
[4]: #HGA = pd.read_csv('human_genome_annotation.txt', sep='\t')
HGA = pd.read_csv('human_genome_annotation.txt', sep='\t')
```

```
[5]: HGA = HGA.head(20)
```

```
[6]: HGA = HGA.drop(['#bin', 'txStart', 'txEnd', 'exonCount', 'score',
↳ 'cdsStartStat', 'cdsEndStat'], axis = 1)
```

```
[7]: alist = []
blist = []
elist = []

for idx, row in HGA.iterrows():

    a = row.exonStarts
    a = a.split(',')
    a = a[:len(a)-1]
    a = list(map(int,a))

    b = row.exonEnds
    b = b.split(',')
    b = b[:len(b)-1]
    b = list(map(int,b))
```

```

    e = row.exonFrames
    e = e.split(',')
    e = e[:len(e)-1]
    e = list(map(int,e))

    alist.append(a)
    blist.append(b)
    elist.append(e)

df = pd.DataFrame({'starts':alist, 'ends':blist, 'frames':elist})

```

```
[8]: HGA = HGA.drop(['exonStarts', 'exonEnds', 'exonFrames'], axis = 1)
```

```
[9]: merged = pd.merge(left=df, left_index=True, right=HGA, right_index=True,
    ↪how='inner')
```

```
[10]: dnaData = pd.read_csv('dnaData.csv')
```

```
[11]: final_df = merged.merge(dnaData, left_on='chrom', right_on='Name')
```

```
[12]: final_df = final_df.dropna()
```

```
[13]: final_df = final_df.apply(lambda x: x.explode() if x.name in ['starts',
    ↪'ends','frames'] else x)
```

```
[14]: final_df = final_df[final_df.frames != -1]
```

```
[15]: final_df = final_df.rename({"name": "name_to_merge"}, axis=1)
```

```
[16]: alist = []
nlist = []
basesData = pd.DataFrame()

for idx,row in final_df.iterrows():
    st = row.starts
    en = row.ends
    cdStart = row.cdsStart
    cdEnd = row.cdsEnd
    nlist.append(row.name_to_merge)

    if (st < cdStart):
        st = cdStart
    if(en > cdEnd):
        en = cdEnd

```

```

actual = row.Bases[st:en]
alist.append(actual)

basesData['Name'] = nlist
basesData['Bases_2'] = alist

```

```
[17]: basesData = basesData.groupby(['Name'])['Bases_2'].apply(lambda x: ''.join(x)).
      ↪reset_index()
```

```
[18]: final_df_2 = merged.merge(basesData, left_on='name', right_on='Name')
```

```
[19]: final_df_2 = final_df_2.dropna()
```

```
[20]: final_df_2 = final_df_2.drop(['starts', 'ends', 'frames', 'cdsStart', '
      ↪cdsEnd'], axis = 1)
```

```
[21]: final_df_2
```

```
[21]:
```

	name	chrom	strand	name2	Name \
0	NM_001276352.2	chr1	-	C1orf141	NM_001276352.2
1	NM_001276351.2	chr1	-	C1orf141	NM_001276351.2
2	XM_011541469.1	chr1	-	C1orf141	XM_011541469.1
3	XM_011541467.1	chr1	-	C1orf141	XM_011541467.1
4	XM_017001276.1	chr1	-	C1orf141	XM_017001276.1
5	XM_011541465.2	chr1	-	C1orf141	XM_011541465.2
6	XM_011541466.2	chr1	-	C1orf141	XM_011541466.2
7	XM_017001277.1	chr1	-	C1orf141	XM_017001277.1
8	XM_011541473.2	chr1	-	C1orf141	XM_011541473.2
9	XM_011541472.1	chr1	-	C1orf141	XM_011541472.1
10	NM_000299.3	chr1	+	PKP1	NM_000299.3
11	NM_001005337.3	chr1	+	PKP1	NM_001005337.3
12	NM_001042682.2	chr1	-	RERE	NM_001042682.2
13	NM_012102.4	chr1	-	RERE	NM_012102.4
14	NM_001042681.2	chr1	-	RERE	NM_001042681.2
15	NM_001281956.1	chr1	-	CSMD2	NM_001281956.1
16	NM_052896.4	chr1	-	CSMD2	NM_052896.4
17	XM_024452878.1	chr1	-	CSMD2	XM_024452878.1
18	XM_017000185.1	chr1	-	CSMD2	XM_017000185.1

```

Bases_2
0  TTATGGGATTTGTGTCCTTTTGTTCCTTTATTTCTATCACCTTCT...
1  TTATGAGGCATTTAAAATTTTCATTTGATAAGTTTAACAAATTATCC...
2  TTATGAGGCATTTAAAATTTTCATTTGATAAGTTTAACAAATTATCC...
3  TTATGAGGCATTTAAAATTTTCATTTGATAAGTTTAACAAATTATCC...
4  TTATGAGGCATTTAAAATTTTCATTTGATAAGTTTAACAAATTATCC...
5  TTATGAGGCATTTAAAATTTTCATTTGATAAGTTTAACAAATTATCC...

```

```

6  TTATGAGGCATTTAAAATTTTCATTTGATAAGTTTAACAAATTATCC...
7  TCATGGAAAATTATGGGATTGTGTCTTTTGTTCCTTTTATTCTATCACCTTCT...
8  TTATGGGATTTGTGTCTTTTGTTCCTTTTATTCTATCACCTTCT...
9  CTAAGAGACCAACAGAATCCAATGGTTTCCTAGGTTTTTTTTTCAAT...
10 ATGAACCACTCGCCGCTCAAGACCGCCTTGGCGTACGAATGCTTCC...
11 ATGAACCACTCGCCGCTCAAGACCGCCTTGGCGTACGAATGCTTCC...
12 TTATAACTGCTTGTGCACCTTCTTTCTTCAGTCGACTGTAATAATCT...
13 TTATAACTGCTTGTGCACCTTCTTTCTTCAGTCGACTGTAATAATCT...
14 TTATAACTGCTTGTGCACCTTCTTTCTTCAGTCGACTGTAATAATCT...
15 CTATACTGCTGTGCACACTGTGCTGACTGTGAACTCCGCCTCGCTG...
16 CTATACTGCTGTGCACACTGTGCTGACTGTGAACTCCGCCTCGCTG...
17 CTATACTGCTGTGCACACTGTGCTGACTGTGAACTCCGCCTCGCTG...
18 CTATACTGCTGTGCACACTGTGCTGACTGTGAACTCCGCCTCGCTG...

```

```

[22]: f=open('foo.fa', "w")

for idx,row in final_df_2.iterrows():
    strand = row.strand
    bases = row.Bases_2
    bases = Seq(bases)
    name1 = row.Name
    name2 = row.name2

    if (strand == '-'):
        bases = bases.reverse_complement()

    my_seq = bases.translate(to_stop=True)
    fasta_format_string = ">%s:%s" % (name1, name2)
    print(fasta_format_string, file=f)
    print(fasta_format_string)

    fasta_format_string = "%s" % my_seq
    print(fasta_format_string, file=f)
    print(fasta_format_string)

f.close()

```

```

>NM_001276352.2:C1orf141
MAEKILEKLDVLDKQAEIILARRTKINRLQSEGRKTTMAIPLTFDFQLEFEEALATSASKAISKIKEDKSCSITKSKMHV
SFKCEPEPRKSNFEKSNLRPFFIQTNVKNKESESTEPVEEHLKRSRISRPYLYLKDTTEMENAGPLNVLYSQHRQACRRSL
GSTDFSPMFNIQSNNAHKKEKSTLFTAQIEKKPRKPLDSVGLLEGDRNKRKRKTQIP
>NM_001276351.2:C1orf141
MAEKILEKLDVLDKQAEIILARRTKINRLQSEGRKTTMAIPLTFDFQLEFEEALATSASKAISKIKEDKSCSITKSKMHV
SFKCEPEPRKSNFEKSNLRPFFIQTNVKNKESESTAQIEKKPRKPLDSVGLLEGDRNKRKKSPQMNDFNKENKSVRNYQ
LSKYRSVRKKSLLPLCFEDELKNPHAKIVNVSPKTVTSHMEQKDTNP IIFHDTEYVRMLLLTKNRFSSHPLENENIYPH
KRTNFILERNCEILKSIIGNQSISLFPKQKTMPTVQRKDIQIPMSFKAGHTTVDDKLKKKTNKQTLNRSWNTLYNFSQN
FSSLTKQFVGYLDAVIHEMSAQTGKFERMFSAGKPTSIPTSSALPVKCYSKPFKYIYELNNVTPLDNLLNLSNEILNAS

```

>XM\_011541469.1:C1orf141

MENAGPLNVLYSQHRQACRRSLGSTDFSPMFNIQSNNAHKKKEDSTLFTAQIEKKPRKPLDSVGLLEGDRNKRKKSPQMND  
FNIKENKSVRNYQLSKYRSVRKKSLLPLCFEDELKNPHAKIVNVSPKTVTSHMEQKDTNPIIFHDTEYVRMLLLTKNRF  
SSHPLENENIYPHKRTNFILERNCEILKSIIGNQSIISLFPKQKTMTPTVQRKDIQIPMSFKAGHTTVDDKLKKTKNQTL  
NRSWNTLYNFSQNFSSLTQFVGYLDKAVIHEMSAQTGKFERMFSAGKPTSIPTSSALPVKCYSKPFKYIYELNNVTPLD  
NLLNLSNEILNAS

>XM\_011541467.1:C1orf141

MAEKILEKLDVLDKQAEIILARRTKINRLQSEGRKTTMAIPLTFDFQLEFEEALATSASKAISKIKEDKSCSITKSKMHV  
SFKCEPEPRKSNFEKSNLRPFFIQTNVKNKESESTEPVEEHLKRSIRPYLYLKDTTEHRQACRRSLGSTDFSPMFNIQ  
NAHKKKEDSTLFTAQIEKKPRKPLDSVGLLEGDRNKRKKSPQMNDFNIKENKSVRNYQLSKYRSVRKKSLLPLCFEDEL  
KNPHAKIVNVSPKTVTSHMEQKDTNPIIFHDTEYVRMLLLTKNRFSSHPLENENIYPHKRTNFILERNCEILKSIIGNQ  
SIISLFPKQKTMTPTVQRKDIQIPMSFKAGHTTVDDKLKKTKNQTLNRSWNTLYNFSQNFSSLTQFVGYLDKAVIHEMSA  
QTGKFERMFSAGKPTSIPTSSALPVKCYSKPFKYIYELNNVTPLDNLLNLSNEILNAS

>XM\_017001276.1:C1orf141

MAEKILEKLDVLDKQAEIILARRTKINRLQSEGRKTTMAIPLTFDFQLEFEEALATSASKAISKIKEDKSCSITKSKMHV  
SFKCEPEPRKSNFEKSNLRPFFIQTNVKNKESESTEPVEEHLKRSIRPYLYLKDTTEMENAGPLNVLYSQHRQACRRSL  
GSTDFSPMFNIQSNNAHKKKEDSTLFTAQIEKKPRKPLDSVGLLEGDRNKRKKSPQMNDFNIKENKSVRNYQLSKYRSVRK  
KSLLPLCFEDELKNPHAKIVNVSPKTVTSHMEQKDTNPIIFHDTEYVRMLLLTKNRFSSHPLENENIYPHKRTNFILER  
NCEILKSIIGNQSIISLFPKQKTMTPTVQRKDIQIPMSFKAGHTTVDDKLKKTKNQTLNRSWNTLYNFSQNFSSLTQFV  
GYLDKAVIHEMSAQTGKFERMFSAGKPTSIPTSSALPVKCYSKPFKYIYELNNVTPLDNLLNLSNEILNAS

>XM\_011541465.2:C1orf141

MAEKILEKLDVLDKQAEIILARRTKINRLQSEGRKTTMAIPLTFDFQLEFEEALATSASKAISKIKEDKSCSITKSKMHV  
SFKCEPEPRKSNFEKSNLRPFFIQTNVKNKESESTEPVEEHLKRSIRPYLYLKDTTEMENAGPLNVLYSQHRQACRRSL  
GSTDFSPMFNIQSNNAHKKKEDSTLFTAQIEKKPRKPLDSVGLLEGDRNKRKKSPQMNDFNIKENKSVRNYQLSKYRSVRK  
KSLLPLCFEDELKNPHAKIVNVSPKTVTSHMEQKDTNPIIFHDTEYVRMLLLTKNRFSSHPLENENIYPHKRTNFILER  
NCEILKSIIGNQSIISLFPKQKTMTPTVQRKDIQIPMSFKAGHTTVDDKLKKTKNQTLNRSWNTLYNFSQNFSSLTQFV  
GYLDKAVIHEMSAQTGKFERMFSAGKPTSIPTSSALPVKCYSKPFKYIYELNNVTPLDNLLNLSNEILNAS

>XM\_011541466.2:C1orf141

MAEKILEKLDVLDKQAEIILARRTKINRLQSEGRKTTMAIPLTFDFQLEFEEALATSASKAISKIKEDKSCSITKSKMHV  
SFKCEPEPRKSNFEKSNLRPFFIQTNVKNKESESTEPVEEHLKRSIRPYLYLKDTTEMENAGPLNVLYSQHRQACRRSL  
GSTDFSPMFNIQSNNAHKKKEDSTLFTAQIEKKPRKPLDSVGLLEGDRNKRKKSPQMNDFNIKENKSVRNYQLSKYRSVRK  
KSLLPLCFEDELKNPHAKIVNVSPKTVTSHMEQKDTNPIIFHDTEYVRMLLLTKNRFSSHPLENENIYPHKRTNFILER  
NCEILKSIIGNQSIISLFPKQKTMTPTVQRKDIQIPMSFKAGHTTVDDKLKKTKNQTLNRSWNTLYNFSQNFSSLTQFV  
GYLDKAVIHEMSAQTGKFERMFSAGKPTSIPTSSALPVKCYSKPFKYIYELNNVTPLDNLLNLSNEILNAS

>XM\_017001277.1:C1orf141

MAEKILEKLDVLDKQAEIILARRTKINRLQSEGRKTTMAIPLTFDFQLEFEEALATSASKAISKIKEDKSCSITKSKMHV  
SFKCEPEPRKSNFEKSNLRPFFIQTNVKNKESESTEPVEEHLKRSIRPYLYLKDTTEMENAGPLNVLYSQHRQACRRSL  
GSTDFSPMFNIQSNNAHKKKEDSTLFTAQIEKKPRKPLDSVGLLEGDRNKRKKSPQMNDFNIKENKSVRNYQLRTKGHKSH  
NFP

>XM\_011541473.2:C1orf141

MAEKILEKLDVLDKQAEIILARRTKINRLQSEGRKTTMAIPLTFDFQLEFEEALATSASKAISKIKEDKSCSITKSKMHV  
SFKCEPEPRKSNFEKSNLRPFFIQTNVKNKESESTAQIEKKPRKPLDSVGLLEGDRNKRKNRTQIP

>XM\_011541472.1:C1orf141

MAEKILEKLDVLDKQAEIILARRTKINRLQSEGRKTTMAIPLTFDFQLEFEEALATSASKAISKIKEDKSCSITKSKMHV  
SFKCEPEPRKSNFEKSNLRPFFIQTNVKNKESESTEPVEEHLKRSIRPYLYLKDTTELKLKKNLGNHWILLVS

>NM\_000299.3:PKP1

MNHSPLKTALAYECFQDQDNSTLALPSDQKMKGTGSGRQVRVQEVMVTVRQKSKSSQSSTLSHSNRGSMYDGLADNYNY  
GTTSSSSYYSKFQAGNGSWGYPYINGTLKREPDRRFFSSYSQMENWSRHYPRGSCNTTGAGSDICFMQKIKASRSEPDLY

CDPRGTLRKGTGSGKGQKTTQNRYSFYSTCSGQKAIKKCPVRPPSCASKQDPVYIPPI SCNKDLSFGHSRASSKICSEDI  
ECSGLTIPKAVQYLSSQDEKYQAIGAYYIQTCTCFQDESAKQVYQLGGICKLVDLLRSPNQNVQAAAGALRNLVFRSTT  
NKLETRRQNGIREAVSLLRRTGNAEIQKQLTGLLWNLSSTDELKEELIADALPVLADRVIIPFSGWCDGNSNMSREVVD  
EVFFNATGCLRKRLGMRLLALVPQRATSSRVNLSSADAGRQTMNRNYSGLIDSLMAYVQNCVAASRCDDKSVENCMCVLH  
NLSYRLDAEVPTRYRQLEYNARNAYTEKSSTGCFSNKSDDMMNNNYDCPLPEEETNPKGSGWLYHSDAIRTYLNLMGKSK  
KDATEACAGALQNLTASKGLMSSGMSQLIGLKEKGLPQIARLLQSGNSDVVRSGASLLSNMSRHPLLRVMGNQVFPEV  
TRLLTSHTGNTSNESEDILSSACYTVRNLMASQPQLAKQYFSSSMLNNIINLCRSSASPKAAEAARLLLSDMWSSKELQGV  
LRQQGFDRNMLGTLAGANSLRNFTSRF

>NM\_001005337.3:PKP1

MNHSPLKTALAYECFQDQDNSTLALPSDQKMKGTSGRQVRVQEQVMMTVKRQKSKSSQSSTLSHSNRGSMYDGLADNYNY  
GTTSRSSYYSKFQAGNGSWGYPYINGTLKREPDRRFFSSYSQMENWSRHYPRGSCNTTGAGSDICFMQKIKASRSEPDLY  
CDPRGTLRKGTGSGKGQKTTQNRYSFYSTCSGQKAIKKCPVRPPSCASKQDPVYIPPI SCNKDLSFGHSRASSKICSEDI  
ECSGLTIPKAVQYLSSQDEKYQAIGAYYIQTCTCFQDESAKQVYQLGGICKLVDLLRSPNQNVQAAAGALRNLVFRSTT  
NKLETRRQNGIREAVSLLRRTGNAEIQKQLTGLLWNLSSTDELKEELIADALPVLADRVIIPFSGWCDGNSNMSREVVD  
EVFFNATGCLRNLSADAGRQTMNRNYSGLIDSLMAYVQNCVAASRCDDKSVENCMCVLHNLNLSYRLDAEVPTRYRQLEYNA  
RNAYTEKSSTGCFSNKSDDMMNNNYDCPLPEEETNPKGSGWLYHSDAIRTYLNLMGKSKKDATEACAGALQNLTASKGL  
MSSGMSQLIGLKEKGLPQIARLLQSGNSDVVRSGASLLSNMSRHPLLRVMGNQVFPEVTRLLTSHTGNTSNESEDILSSA  
CYTVRNLMASQPQLAKQYFSSSMLNNIINLCRSSASPKAAEAARLLLSDMWSSKELQGVLRQQGFDRNMLGTLAGANSLR  
NFTSRF

>NM\_001042682.2:RERE

MFKPVKEEDDGLSGKHSMTRRSRGSMSTLRSRKKQPASPDGRTSPINEDIRSSGRNSPSAASTSSNDSKAETVKKSAK  
KVKEEASSPLKSNKRQREKVASDTEEADRTSSKTKTQEISRPNPSEGESESSDSRSVNDEGSSDPKDIDQDNRSTSPS  
IPSPQDNESDSSAQQQMLQAQPPALQAPTGTVPAPSSAPPGTPLPTPGPTPSATAVPPQGSPTASQAPNQPPAPTAP  
VPHTHIQAPALHPQRPPSPHPPHPSHPPLQPLTGSAQPSAPSHAQPPPLHGQGPFGPHSLQAGPLLQHPGPPQPFGL  
PPQASQGAQPLGTSPAAAYPHTSLQLPASQSALQSQQPPREQPLPPAPLAMPHIKPPPTTPIPQLPAPQAHKHPHLSGP  
SPFSMNANLPPPPALKPLSSLSTHPPSAHPPPLQLMPQSQPLPSSPAQPPGLTQSQNLPPPPASHPPTGLHQVAPQPPF  
AQHPFVPGGPPPIPTPTCPSTSTPPAGPGTSAQPPCSGAAASGGSIAGGSSCPLPTVQIKEEALDDAEEPESPPPPRSP  
SPEPTVVDTPSHASQSARFYKHLDRGYNSCARTDLYFMPLAGSKLAKKREEAIEKAKREAEQKAREEREREKEKERER  
EREREREAEAAKASSSAHEGRLSDPQLSGPGHMRPSFEPPPTTIAAVPPYIGPDTPALRTLSEYARPHVMSPTNRNHPF  
YMPLNPTDPLLAYHMPGLYNVDPTIRERELREREIREREIRERELRERMKPGFEVKPELDPLHPAANPMEHFARHSALT  
IPTAGPHPFASFHPGLNPLERERLALAGPQLRPEMSYPDLAAERIIHAERMASLTSDPLARLQMFNVTPHHHQHSHIHS  
HLHLHQQDPLHQGSAGPVHPLVDPLTAGPHLARFPYPPGTLPNLLGQPPHEHEMLRHPVFGTPYPRDLPGAIPPPMSAA  
HQLQAMHAQSAELQRLAMEQWLHGHPMHGGHLPSQEDYYSRLKKEGDKQL

>NM\_012102.4:RERE

MTADKDKDKKEKDRDRDREREKRDKARESENSRPRRSCLEGGAKNYAESDHSEDEDNDNNSATAEESTKKNKKKPP  
KKKSRYERTDTGEITSYITEDDVYRPGDCVYIESRRPNTPYFICSIQDFKL VHNSQACCRSPTPALCDPPACSLPVASQ  
PPQHLSEAGRGPVGSKRDLHLMNVKWWYRQSEVPDSVYQHLVQDRHNENDSGRELVITDPVIKNRELFI SDYVDTYHAAA  
LRGKCNISHFSDIFAAREFKARVDSFFYILGYNPETRRLNSTQGEIRVGPSHQAKLPDLQFPSPDGDVTVTQHEELVWMP  
GVNDCDLLMYLRAARSMAAFAGMCDGGSTEDGCVAASRDDTTLNALNTLHESGYDAGKALQRLVKKPVKLEKCWTEDE  
VKRFVKGLRQYGKNFFRIRKELLPNKETGELITFYYYWKKTPEAASSRAHRRHRRQAVFRRIKTRTASTPVNTPSRPPSS  
EFLDLSSASEDDFSEDSEQLKGYACRHCFTTTSKDWHHGGRENILLCTDCRIHFKKY GELPPIEKPVDPPPFMFKPVK  
EEDDGLSGKHSMTRRSRGSMSTLRSRKKQPASPDGRTSPINEDIRSSGRNSPSAASTSSNDSKAETVKKSAKKVKEEA  
SSPLKSNKRQREKVASDTEEADRTSSKTKTQEISRPNPSEGESESSDSRSVNDEGSSDPKDIDQDNRSTSPSIPSPQD  
NESDSDSSAQQQMLQAQPPALQAPTGTVPAPSSAPPGTPLPTPGPTPSATAVPPQGSPTASQAPNQPPAPTAPVPHTHI  
QQAPALHPQRPPSPHPPHPSHPPLQPLTGSAQPSAPSHAQPPPLHGQGPFGPHSLQAGPLLQHPGPPQPFGLPPQASQ  
GQAPLGTSPAAAYPHTSLQLPASQSALQSQQPPREQPLPPAPLAMPHIKPPPTTPIPQLPAPQAHKHPHLSGPSFSMN  
ANLPPPPALKPLSSLSTHPPSAHPPPLQLMPQSQPLPSSPAQPPGLTQSQNLPPPPASHPPTGLHQVAPQPPFAQHPFV  
PGGPPPIPTPTCPSTSTPPAGPGTSAQPPCSGAAASGGSIAGGSSCPLPTVQIKEEALDDAEEPESPPPPRSPSPEPTV

VDTPSHASQSARFYKHLDRGYNSCARTDLYFMPLAGSKLAKKREEAIEKAKREAEQKAREEREREKEKEKERERERERER  
EAERAAKASSSAHEGRLSDPQLSGPGHMRPSFEPPPTTIAAVPPYIGPDTPALRTLSEYARPHVMSPTNRNHPFYMPLNP  
TDPLLAYHMPGLYNVDPTIRERELREREIREREIRERELRERMKPGFEVKPPELDPLHPAANPMEHFARHSALTIPPTAG  
PHPFASFHPGLNPLERERLALAGPQLRPEMSYPDLAAERIHAERMASLTSDPLARLQMFNVTPHHHQSHIHSHLHLHQ  
QDPLHQGSAGPVHPLVDPLTAGPHLARFPYPPGTLPNPLLQGPPHEHEMLRHPVFGTPYPRDLPGAIPPPMSAAHQLQAM  
HAQSAELQRLAMEQQWLHGHPMHGGHLPSQEDYY SRLKKEGDKQL

>NM\_001042681.2:RERE

MTADKDKDKDEKDRDRDREREKRDKARESENSRPRRSCTLEGGAKNYAESDSEDEDNDNNSATAEESTKKNKKKPP  
KKKSRYERTDTGEITSYITEDDVVYRPGDCVYIESRRPNTPYFICSIQDFKLHNSQACCRSPTPALCDPPACSLPVASQ  
PPQHLSEAGRGPVGSKRDLMLNVKWWYRQSEVPDSVYQHLVQDRHNENDSGRELVITDPVIKNRELFISDYVDTYHAAA  
LRGKCNISHFSDFIAAREFKARVDSFFYILGYNPETRRLNSTQGEIRVGPSHQAKLPDLQFPSPDGDVTVTQHEELVWMP  
GVNDCDLLMYLRAARSMAAFAGMCDGGSTEDGVAASRDDTTLNALNTLHESGYDAGKALQRLVKKPVPKLIEKCWTEDE  
VKRFVKGLRQYGNFFRIRKELLPNKETGELITFYYYWKKTPEAASSRAHRRHRRQAVFRRIKTRTASTPVNTPSRPPSS  
EFLDLSSASEDDFSEDESEQLKGYACRHCFTTTSKDWHHGGRENILLCTDCRIHFKKYGELEPPIEKVPDPPPFMFKPVK  
EEDDGLSGKHSMTRRSRGSMSTLRSRKKQPASPDGRTPINEDIRSSGRNSPSAASTSSNDSKAETVKKSAKKVKEEA  
SSPLKSNKRQREKVASDTEEADRTSSKKTQTEISRPNSPSEGESESSDSRSVNDEGSSDPKDIDQDNSTSPSIPSPQD  
NESDSDSSAQQMLQAQPPALQAPTGVTPAPSSAPPGTPLPTPGTPSATAVPPQGSPTASQAPNQPAQTAPVPHTHI  
QQAPALHPQRPPSPHPPHPSHPPLQPLTGSAGQPSAPSHAQPPLHGQGGPHSLQAGPLLQHPGPPQPFGLPPQASQ  
GQAPLGTSPAAAYPHTSLQLPASQSALQSQPPREQPLPPAPLAMPHIKPPPTTIPQLPAPQAHKHPHLSGSPSFSMN  
ANLPPPPALKPLSSLSTHPPSAHPPPLQLMPQSQPLPSSPAQPPGLTQSQNLPPPASHPPTGLHQVAPQPPFAQHFPV  
PGGPPITPPTCPSTSTPPAGPGTSAQPPCSGAAASGGSIAGSSCPLPTVQIKEEALDDAEEPESPPPPRSPSPSEPTV  
VDTPSHASQSARFYKHLDRGYNSCARTDLYFMPLAGSKLAKKREEAIEKAKREAEQKAREEREREKEKEKERERERERER  
EAERAAKASSSAHEGRLSDPQLSGPGHMRPSFEPPPTTIAAVPPYIGPDTPALRTLSEYARPHVMSPTNRNHPFYMPLNP  
TDPLLAYHMPGLYNVDPTIRERELREREIREREIRERELRERMKPGFEVKPPELDPLHPAANPMEHFARHSALTIPPTAG  
PHPFASFHPGLNPLERERLALAGPQLRPEMSYPDLAAERIHAERMASLTSDPLARLQMFNVTPHHHQSHIHSHLHLHQ  
QDPLHQGSAGPVHPLVDPLTAGPHLARFPYPPGTLPNPLLQGPPHEHEMLRHPVFGTPYPRDLPGAIPPPMSAAHQLQAM  
HAQSAELQRLAMEQQWLHGHPMHGGHLPSQEDYY SRLKKEGDKQL

>NM\_001281956.1:CSMD2

MPSRSGRELGRCGCPAGRARGETGISALVPGAGSRWGRPPPTPPPLLLLLGCGLLSVSAAAGQNCTFQLHGPNGTVESP  
GFPYGYPNYANCTWTITAEQHRIQLVFQSFAL EEDFDVLSVFDGPPQENLRTRLTGFLPATIVSAATTLRLISDY  
AVSAQGFIATYEVLPSTCGNPGRLPNGIQGGSTFNLGDKVRYSCNLGFFLEGHAVLTCHAGSENSATWDFPLPSCRADD  
ACGGTLRGQSGIISSPHFPSEYHNNADCTWTILAE LGDTIALVFIDFQLEDGYDFLEVGTGEGSSLWFTGASLPAPVISS  
KNWRLRLHFTSDGNHRQRGFS AQYQVKKQIELKSRGVKLMPSKDNSQKTSVLTQVGVSQGHNMCPDPGIPERGKRLGSDFR  
LGSSVQFTCNEGYDLQGSKRITCMKVSDMFAAWS DHRPVCRARMCD AHLRGPSGIITSPNFP IQYDNNAHC VWIITALNP  
SKVIKLA FEEDFLERGYDTLTVGDGGQDGDQTVLYILTGTSVPDLIVSTNHQM WLLFQTDGSGSSLGFKASYEEIEQGS  
CGDPGIPAYGRREGSRFHGDTLKFECQPA FELVGQKAITCQKNNQWSAKKPGCVFSCFFNFTSPSGVVLSPNYPEDYGN  
HLHC VWLILARPESRIHLAFNDIDVEPQDFDLVIKDGATAEAPVLGTFSGNQLPSSITSSGHVARLEFQTDHSTGKRGFN  
ITFTTFRHNECPDPGPVNGKRF GDSLQLGSSISFLCDEGFLGTQGSETITCVLKEGSVWNSAVLRCEAPCGGHLTSPS  
GTILSPGWPGFYKDALS CAWVIEAQPGYPIKITFDRFKTEVNYDTLEVRDGR TYSAPLIGVYHGTQVPQFLISTS NYLYL  
LFSTDKSHSDIGFQLRYETITLQSDHCLDPGIPVNGQRHGNDFYVGALVTFSCDSGYTLSDGELECEPNFQWSRALPSC  
EALCGGFIQSSGTILSPGFDPDFYPNNLNCTWIIETSHGKG VFFTFHTFHLES GHDYLLITENG SFTQPLRQLTGSRLPA  
PISAGLYGNFTAQVRFISDFSMSYEGFNITFSEYDLEPCEEVPAYSIRKGLQFGVGDTLTFSCFPGRLEG TARITCL  
GRRRLWSSPLPRCVAECGNSVTGTGTLLSPNFPVNYNNNHECIYSIQTPGKGIQLKARAFELSEG DVLKVYDGNNS  
ARLLGVF SHSEMMGVTLNSTSSSLWLD FITDAENTSKGFELHFSSFELIKCEDPGTPKFGYKVHDEGHFAGSSVSFSCDP  
GYSLRGSEELLCLSGERRTWDRPLPTCVAECGGTVRGEVSGQVLSPGYPAPYEHNLCIWTIEAEAGCTIGLHFLVFDTE  
EVHDLVLRWDGPVESGVLLKELSGPALPKDLHSTFNSVVLQFSTDFFTSKQGFAIQFSVSTATSCNDPGIPQNGSRSGDS  
WEAGDSTVFQCDPGYALQGS AEISCVKIENRFFWQPSPTCIA PCGGDLTGPSGVILSPNYPEPYPPGKECDWKVTVSPD  
YVIALVFNIFNLEPGYDFLHIYDGRDSLSP LIGSFYGSQLPGRIESSNSLFLAFRSDASVS NAGFVIDYTENPRESCFD

PGSIKNGTRVGSDDLKLGSSVTYYCHGGYEVEGTSTLSCILGPDGKPVWNNPRPVCTAPCGGQYVVGSDGVVLSNPYPQNYT  
SGQICLYFVTVPKDYVVFVGQFAFFHTALNDVVEVHDGHSQHSRLSSLSGSHGESLPLATSNQVLKFSKGLAPARGF  
HFVYQAVPRTSATQCSSVPEPRYGRKLGSDFSVGAIVRFECNSGYALQGSPEIECLPVPGALAQWNVSAPTCVVPCGGNL  
TERRGTILSPGFPEPYLNSLNCVWKIVVEGAGIQVVSFVTEQNWDSLEVFDGADNTVTMLGSFSGTTVPALLNSTSN  
QLYLHFYSDISVSAAGFHLEYKTVGLSSCEPAVPSNGVKTGERYLNDVVSFQCEPGYALQGHAIHISCMPGTVRRWNYP  
PPLCIAQCGGTVEEMEGVILSPGFPGNYPNSMDCSWKIALPVGFGAHIQFLNFSTEPNHDIYIEIRNGPYETSRMMGRFSG  
SELPSSLLSTSHETTVYFHSQNRPGFKLEYQAYELQECPDPEPFANGIVRGAGYNVQSVTFECLPGYQLTGHPVLT  
CQHGTNRNWDHPLPKCEVPCGGNITSSNGTVYSPGFSPYSSSQDCVWLITVPIGHGVRLNLSLLQTEPSGDFITIWGDP  
QQTAPRLGVFTRSMAKKTVQSSSNQVLLKFHRDAATGGIFAIAFSAYPLTKCPPPTILPNAEVVTENEENIGDIVRYRC  
LPGFTLVGNEILTCKLGTYLQFEGPPPICEVHCPTNELLTDSTGVILSQSYPGSYPPQFQTCSWLVRVEPDYNISLTVYEF  
LSEKQYDEFEIFDGPSPGQSPLLKALSGNYSAPLIVTSSNSVYLRWSSDHAYNRKGFKIRYSAPYCSLPRAPLHGFILGQ  
TSTQPGGSIHFGCNAGYRLVGHSMAICTRHPQGYHLWSEAIPLCQALSCGLPEAPKNGMVFGKEYTVGKAMYSCEGYH  
LQAGAEATAECLDTGLWSNRNVPPQCVPTCPDVSSISVEHGRWRLIFETQYQFQAQLMLICDPGYYYTGQRVIRCQANG  
KWSLGDSTPTCRIISCGELPIPPNGHRIGTLSVYGATAIFSCNSGYTLVGSRVRECMANGLWSGSEVRCLATQTKLHSIF  
YKLLFDVLSSPSLTKAGHCGTPEPIVNGHINGENYSYRGSVVYQCNAFRLIGMSVRICQDDHHWSGKTPFCVPITCGHP  
GNPVNGLTQGNQFNLNDVVKFVCNPGYMAEGAARSQCLASGQWSDMLPTCRIINCTDPGHQENSVRQVHASGPHRFSFGT  
TVSYRCNHGFYLLGTPVLSCQGDGTWDRPRPQCLLVSCGHPGSPHQSMSGDSYTVGAVVRYSCIGKRTLGVNSTRMCGL  
DGHWTGSLPHCSGTSVGVCGDPGIPAHGIRLGDSTDPGTVMRFSCEAGHVLRGSSERTCQANGSWSGSQPECGVISCGNP  
GTPSNARVVFSDGLVFSSSIYVECREGYYATGLLSRHCSVNGTWTGSDPECLVINCDDPGIPANGLRLGNDFRYNKTVTY  
QCVPGYMMESHRSVLSCTKDRTWNGTKPVCKALMCKPPPLIPNGKVVGSDFMWGSSVTYACLEGYQLSLPAVFTCEGNG  
SWTGELPQCFPVFCGDPGVPSRGRREDRGFSYRSSVSFSCHPPLVLVGSPRRFCQSDGTWSGTQPSCIDPTLTTCADPGV  
PQFGIQNNSQGYQVGSTVLFRCQKGYLLQGSTRITCLPNLTWSGTPPDCVPHHCRQPETPTANVGALDLPMSGYTLIYS  
CQEGFSLKGGSEHRTCKADGSWTGKPPICLEVRPSGRPINTAREPPLTQALIPGDVFAKNSLWKAYEYQKKQPAMLRV  
TGFQVANSKVNATMIDHSGVELHLAGTYKKEDFHLLQVYQITGPVEIFMNKFKDDHWALDGHVSSSESSGATFIYQGSVK  
GQGFQGFQRLDLRLLES DPESIGRHFASNSSSVAAAAILVPFIALIIAGFVLYLYKHRRRPKVPFNGYAGHENTNVRAT  
FENPMYDRNIQPTDIMASEAEFTVSTVCTAV

>NM\_052896.4: CSMD2

MRLISIAQGPWRQVQSHHPRSAGQNCTFQLHGPNGTVESPGFPYGYPNYANCTWTITAEQHRIQLVFQSFAL EEDFDVL  
SVFDGPPQPENLRTRLTGFLPATIVSAATTLRLISDYAVSAQGFHATYEVLPSTHCGNPGRLPNGIQQGSTFNLGDK  
VRYSCNLGFFLEGHAVLTCHAGSENSATWDFPLPSCRADDACGGTLRGQSGIISPHFPSEYHNNADCTWTILAE LGDTI  
ALVFIDFQLEDGYDFLEVTGTGESSLWFTGASLPAPVISSKNWRLRLHFTSDGNHRQRGFSAYQVKKQIELKSRGVKLM  
SKDNSQKTSVLTQVGVSQGHNMCPDPGIPERGKRLGSDFRLGSSVQFTCNEGYDLQGSKRITCMKVSDMFAAWS DHRPVC  
RARMCDALHRGPSGIITSPNFPIQYDNNAHCVIITALNPSKVIKLAEEFDLERYDTLTVGDGGQDGDQKT VLYILTG  
TSVPDLIVSTNHQMWLLFQTDGSGSSLGFKASYEEIEQGSCGDPGIPAYGRREGSRFHHGDTLKFECQPAFELVGQKAIT  
CQKNNQWSAKKPGCVFSCFFNFTSPSGVVLSPNYPEDYGNHLHCVWLILARPESRIHLAFNDIDVEPQDFLVIKDGATA  
EAPVLGTFSGNQLPSSITSSGHVARLEFQTDHSTGKRGFNITFTTFRHNECPDPGVPVNGKRFGDSLQLGSSISFLCDEG  
FLGTQGSETITCVLKEGSVVWNSAVLRCEAPCGHLTSPSGTILSPGWPGFYKDALS CAWVIEAQPGYPIKITFDRFKTE  
VNYDTLEVRDGRYTSAPLIGVYHGTQVPQFLISTSNYLYLLFSTDKSHSDIGFQLRYETITLQSDHCLDPGIPVNGQRHG  
NDFYVGALVTFSCDSGYTLSDGEPLCEPNFQWSRALPSCEALCGGFIQSSGTILSPGFDPDFYPNNLNCTWIIETSHGK  
GVFFTFTHTFHLES GHDYLLITENGSTQPLRQLTGSRLPAPISAGLYGNFTAQVRFISDFSMSYEGFNITFSEYDLEPCE  
EPEVPAYSIRKGLQFGVGDTLTFSCFPGYRLEGTARITCLGRRRLWSSPLPRCVAECGNSVTGTQGTLLSPNFPVNYNN  
NHECIYSIQTPGKGIQLKARAFELSEGDLVKVYDGNNSARLLGVFSSHSEMMGVTLNSTSSSLWLD FITDAENTSKGFE  
LHFSSFELIKCEDPGTPKFGYKVHDEGHFAGSSVSFSCDPGYSLRGSEELLCLSGERRTWDRPLPTCVAECGGTVRGEVS  
GQVLSPGYPAPYEHNLNCIWTIEAEAGCTIGLHFLVFDTEEVHDVLRWDGPVESGVLLKELSGPALPKDLHSTFNSVVL  
QFSTDFFTSKQGFQAIQFSVSTATSCNDPGIPQNGSRSGDSWEAGDSTVFQCDPGYALQGS AEISCVKIENRFFWQPSPT  
CIAPCGGDLTGPSGVILSPNYPEPYPGKECDWKVTVSPDYVIALVFNIFNLEPGYDFLHIYDGRDLSPLIGSFYGSQ  
PGRIESSNSLFLAFRSDASVSNAGFVIDYTENPRESCFDPGSIKNGTRVGSDDLKLGSSVTYYCHGGYEVEGTSTLSCIL  
GPDGKPVWNNPRPVCTAPCGGQYVVGSDGVVLSNPYPQNYTSGQICLYFVTVPKDYVVFVGQFAFFHTALNDVVEVHDGHSQ



HSRLSSSLSGSHTGESLPLATSNQVLKFSAGKLAPARGFHFVYQAVPRTSATQCSSVPEPRYGKRLGSDFSVGAIVRFE  
CNSGYALQGSPEIECLPVPGALAQWNVSAPTQVPCGGNLTERRGITILSPGFPEPYLNSLNCVWKIVVEGAGIQIVVSV  
FVTEQNWDSLEVFDGADNTVTMLGSFSGTTVPALLNSTSNQLYLFYSDISVSAAGFHLEYKTVGLSSCPEPAVPSNGVK  
TGERYLVDVVSFQCEPGYALQGHAIHSCMPGTVRWNYPPLCIAQCGGTVEEMEGVILSPGFPGNYPNSMDCSWKIAL  
PVGFGAHIQFLNFSTEPNHDYIEIRNGPYETSRMMGRFSGSELPSLLSTSHETTVYFHSQNRPGFKLEYQDLTYSH  
QISSFLRGFDLSELERTNSTPPVAASYVWDLDPGCEAYELQECDPPEPFANGIVRGAGYNVQSVTFECLPGYQLTGHPV  
LTCQHGTNRNWDHPLPKCEVPCGGNITSSNGTVYSPGFSPYSSSQDCVWLITVPIGHGVRLNLSLLQTEPSGDFITIW  
GPQQTAPRLGVFTRSMAKKTVQSSSNQVLLKFHRDAATGGIFAIAFSAYPLTKCPPPTILPNAEVTENEFEINIGDIVRY  
RCLPGFTLVGNEILTCKLGTYLQFEGPPPICEVHCTNELLTDSTGVILSQSYPGSYPQFQTCWLVRVEPDYNISLTV  
YFLSEKQYDEFEIFDGPSPQSPLLKALSGNYSAPLIVTSSNSVYLRWSSDHAYNRKGFKIRYSAPYCSLPRAPLHGFIL  
GQTSTQPGGSIHFGCAGYRLVGHSMACITRHPQGYHLWSEAIPLCQALSCGLPEAPKNGMVFGKEYTVGKAMYSCEG  
YHLQAGAEATAECLDTGLWSNRNVPPQCVPTCPDVSSISVEHGRWRLIFETQYQFQAQLMLICDPGYYYTGQRVIRCQA  
NGKWSLGDSTPTCRIISCGELPIPPNGHRIGTILSVYGATAIFSCNSGYTLVGSRVRECMANGLWSGSEVRCLAGHCGTPE  
PIVNGHINGENYSYRGSVVYQCNAGFRLIGMSVRICQDDHHWSGKTPFCVLVSCGHPGSPPHSQMSGDSYTVGAVVRYSC  
IGKRTLGVNSTRMCGLDGHWGSLPHCSGTVGCGDGPAPAHGIRLGDSFDPGTVMRFSCEAGHVLRGSSERTCQANGS  
WSGSQPEGCVISCGNPGTPSNARVVFSDGLVFSSIVYECREGYATGLLSRHCSVNGTWTGSDPECLVINCGDPGIPAN  
GLRLGNDFRYNKTVTYQCVPGYMMESHRSVLSCTKDRTWNGTKPVCKALMCKPPPLIPNGKVVGSDFMWGSSVTYACLE  
GYQLSLPAVFTCEGNGSWTGELPQCFPVFCGDPGVPSRGRREDRGFSYRSSVSFSCHPPLVLVGSPPRRFCQSDGTWSGTQ  
PSCIDPTLTTCADPGVPQFGIQNNSQGYQVGSTVLFRCQKGYLLQGSTTRTCLPNLTWSGTPPDCVPHHCRQPETPTAN  
VGALDPLSMGYTLIYSCQEGFSLKGGSEHRTCKADGWSGKPPICLEVPSGRPIINTAREPPLTQALIPGDVFAKNSLWK  
GAYEYQKKQPAMLRVTGFQVANSKVNATMIDHSGVELHLAGTYKKEDFHLLQVYQITGPVEIFMKNFKDDHWALDGHV  
SSESSGATFIYQGSVKGGQFGQFGRQLDLRLLES DPESIGRHFASSSSVAAA ILVPIALIIAGFVLYLYKHRRRPKV  
PFNGYAGHENTNVRATFENPMYDRNIQPTDIMASEAEFTVSTVCTAV

>XM\_024452878.1:CSMD2

MPRSRGRELGRCGCPAGRARGETGISALVPGAGSRWGRPPPTPPPLLLLLGCGLLSVSAAAGQNCTFQLHGPNGTVESP  
GFPYGYPNYANCTWTITAEQHRIQLVFQSFAL EEDFDVLSVFDGPPQPENLRTRLTGFLPATIVSAATLSLRLISDY  
AVSAQGFHATYEVLPSTCGNPGRLPNGIQGGSTFNLGDKVRYSCNLGFFLEGHAVLTCHAGSENSATWDFPLPSCRADD  
ACGGTLRGQSGIISSPHFPSEYHNNADCTWTILAE LGDTIALVFIDFQLEDGYDFLEVGTGEGSSLWFTGASLPAPVISS  
KNWRLHLFTSDGNHRQRGFSAQYQVKKQIELKSRGVKLMPSKDNSQKTSVLTQVGVVSQGHNMCPDPGIPER GKRLGSDFR  
LGSSVQFTCNEGYDLQGSKRITCMKVSDMFAAWS DHRPVCARMCD AHLRGPSGIITSPNFIQYDNNAHCWII TALNP  
SKVIKLA FEEDLERGYDTLTVGDGGQDGDQKTVLYILGTSTVPDLIVSTNHQM WLLFQTDGSGSSLGFKASYEEIEQGS  
CGDPGIPAYGRREGSRFHGDTLKFECQPAFELVGQAITCQKNNQWSAKKPGCVFSCFFNFTSPSGVVLSPNYPEDYGN  
HLHCVWLILARPESRIHLAFNDIDVEPQDFDLVIKD GATAEAPVLGTFSGNQLPSSITSSGHVARLEFQTDHSTGKRGFN  
ITFTTFRHNECPDPGVVNGKRFGDSLQLGSSISFLCDEGFLGTQGSETITCVLKEGSSVVWNSAVLRCEAPCGH LTPS  
GTILSPGWPGFYKDALS CAWVIEAQPGYPIKITDFRKFTEVNYDTLEVRDGRITYSAPLIGVYHGTQVPQFLISTSNYLYL  
LFSTDKSHSDIGFQLRYETITLQSDHCLDPGIPVNGQRHGNDFYVGALVTFSCDSGYTLSDGEPLCEPNFQWSRALPSC  
EALCGGFIQSSGTTILSPGFPDFYPPNNLNCTWIIETSHGKG VFFTFHTFHLES GHDYLLITENGSTFTQLRQLTGSRLPA  
PISAGLYGNFTAQVRFISDFSMSYEGFNITFSEYDLEPCEEPEVPAYSIRKGLQFGVGDTLTFSCFPGRLEG TARITCL  
GRRRLWSSPLPRCVAECGNSVTGTQGTLLSPNFPVNYNNNHECIYSIQTPGKGIQLKARAFELSEGDVLKVYDGNNS  
ARLLGVFSSHSEMMGVTLNSTSSSLWLD FITDAENTSKGFELHFSSFELIKCEDPGTPKFGYKVHDEGHFAGSSVSFSCDP  
GYSLRGSEELLCLSGERRTWRPLPTCVAECGGTVRGEVSGQVLSPGYPAPYEHNLCIWTIEAEAGCTIGLHFLVFDTE  
EVHDLRIWDGPVESGVLLKELSGPALPKDLHSTFNSVVLQFSTDFFTSKQGF AIQFSVSTATSCNDPGIPQNGSRSGDS  
WEAGDSTVFQCDPGYALQGS AEISCVKIENRFFWQPSPTCIA PCGGDLTGPSGVILSPNYPEPYPPGKECDWKVTVSPD  
YVIALVFNIFNLEPGYDFLHIYDGRDSLSP LIGSFYGSQLPGRIESSNSLFLAFRSDASVSNAGFVIDY TENPRESCFD  
PGSIKNGTRVGSDDLKLGSSVTYYCHGGYEVEGTSTLSCILGPDGKPVWNNPRPVCTAPCGGQYVGS DGVVLSPNYPQNYT  
SGQICLYFVTVPKDYVVFQGFAFFHTALNDVVEVHDGHSQHSRLSSSLSGSHTGESLPLATSNQVLKFSAGKLAPARGF  
HFVYQAVPRTSATQCSSVPEPRYGKRLGSDFSVGAIVRFE CNSGYALQGSPEIECLPVPGALAQWNVSAPTQVPCGGNL  
TERRGITILSPGFPEPYLNSLNCVWKIVVEGAGIQIVVSVFVTEQNWDSLEVFDGADNTVTMLGSFSGTTVPALLNSTSN

QLYLHFYSDISVSAAGFHLEYKTVGLSSCPEPAVPSNGVKTGERYL VNDVVSFQCEPGYALQGHAIHISCMPGTVRRWNYP  
PPLCIAQCQGGTVEEMEGVILSPGFPGNYPNSMDCSWKIALPVGFGAHIQFLNFSTEPNHDIYIEIRNGPYETSRMMGRFSG  
SELPSSLLSTSHETTVYFHSQNRPGFKLEYQAYELQECDPPEPFANGIVRGAGYNVGGSVTFECLPGYQLTGHPVLT  
CQHGTNRNWDHPLPKCEVPCGGNITSSNGTVYSPGFSPYSSSQDCVWLITVPIGHGVRLNLSLLQTEPSGDFITIWGDP  
QQTAPRLGVFTRSMAKKTVQSSSNQVLLKFHRDAATGGIFAIASAYPLTKCPPPTILPNAEVVTENEENIGDIVRYRC  
LPGFTLVGNEILTCKLGTYLQFEGPPPICEVHCPTNELLTDSTGVILSQSYPGSYPQFQTCSWLVRVEPDYNISLTV EYF  
LSEKQYDEFEIFDGPSPQSPLLKALSGNYSAPLIVTSSNSVYLWSSDHAYNRKGFKIRYSAPYCSLPRAPLHGFILGQ  
TSTQPGGSIHFGCNAGYRLVGHSMIAICTRHPQGYHLWSEAIPLCQALSCGLPEAPKNGMVFGKEYTVGTKAMYSCSEGYH  
LQAGAEATAECLDTGLWSNRNVPPQCVPTCPDVSSISVEHGRWRLIFETQYQFQAQLMLICDPGYYYTGQRVIRCQANG  
KWSLGDSTPTCRIISCGELPIPPNGHRIGTSLVYGATAIFSCNSGYTLVGSRVRECMANGLWSGSEVRCLATQTKLHSIF  
YKLLFDVLLSSPLTKAGHCGTPEPIVNGHINGENYSYRGSVVYQCNAGFRLIGMSVRICQDDHHWSGKTPFCVPITCGHP  
GNPVNGLTQGNQFNLNDVVKFVCNPGYMAEGAARSQCLASGQWSDMLPTCRIINCTDPGHQENSVRQVHASGPHRFSFGT  
TVSYRCNHGFYLLGTPVLSCQGDGTWDRPRPQCLLVSCGHPGSPHPSQMSGDSYTVGAVVRYSCIGKRTL VGNSTRMCGL  
DGHWTGSLPHCSGTSVGVCGDPGIPAHGIRLGD SFDPGTVMRFSCEAGHVLRGSSERTCQANGSWSGSQPECGVISCGNP  
GTPSNARVVFSDGLVFSSSI VYECREGYATGLLSRHCSVNGTWTGSDPECLVINCGDPGIPANGLRLGNDFRYNKTVTY  
QCVPGYMMESHRSVLSCTKDRTWNGTKPVCKALMCKPPPLIPNGKVVGSDFMWGSSVTYACLEGYQLSLPAVFTCEGNG  
SWTGELPQCFPVFCGDPGVPSRGRREDRGFSYRSSVSFSCHPPLVLVGSPRRFCQSDGTWSGTQPSCIDPTLTTCADPGV  
PQFGIQNNSQGYQVGSTVLFRCQKGYLLQGSTTRTCLPNLTWSGTPDCVPHHCRQPETPTHANVGALDLP SMGYTLIYS  
CQEGFSLKGGSEHRTCKADGSWTGKPPICLEVRPSGRPINTAREPPLTQALIPGDVFAKNSLWK GAYEYQKKQPA MLRV  
TGFQVANSKVNATMIDHSGVELHLAGTYKKEDFHLLQVYQITGPVEIFMKNFKDDHWALDGHVSSSESSGATFIYQGSVK  
GQFGQGFGFQRLDLRLLES DPESIGRHFASNSSSVAAILVPFIALIIAGFVLYLYKHRRRPKVPFNGYAGHENTNVRAT  
FENPMYDRNIQPTDIMASEAEFTVSTVCTAV

>XM\_017000185.1:CSMD2

MPRSRGRELGRCGCPAGRARGETGISALVPGAGSRWGRPPPTPPPLLLLLGCGLLSVSAAAGQNCTFQLHGPNGTVESP  
GFPYGYPNYANCTWTITAEQHRIQLVFSFALEEDFDVLSVFDGPPQPENLRRLTGFLPATIVSAATLSLRLISDY  
AVSAQGFHATYEVLP SHTCGNPGRLPNGIQGGSTFNLGDKVRYSCNLGFFLEGHAVLTCHAGSENSATWDFPLPSCRADD  
ACGGTLRGQSGIISPHFPSEYHNNADCTWTILAE LGDTIALVFIDFQLEDGYDFLEV TGTEGSSLWFTGASLPAPVISS  
KNWLRHLFTSDGNHRQRGFSAYQVKKQIELKSRGVKLMP SKDNSQKTSVLTQVGVSQGHNMCPDPGIPER GKRLGSDFR  
LGSSVQFTCNEGYDLQGSKRITCMKVSDMFAAWS DHRPVCRRMCD AHLRGPSGIITSPNFIQYDNNAHCWII TALNP  
SKVIKLA FEEFDLERYD TLTVGDGGQDGDQKT VLYILTGT SVPDLIVSTNHQM WLLFQTDGSGSSLGFKASYEEIEQGS  
CGDPGIPAYGRREGSRFHGDTLKFECQPA FELVGKAITCQKNNQWSAKKPGCVFSCFFNFTSPSGVVLSPNY PEDYGN  
HLHCVWLILARPESRIHLAFNDIDVEPQDFLVIKDGATAEAPVLGTFSGNQLPSSITSSGHVARLEFQTDHSTGKRGFN  
ITFTTFRHNECPDPGPVPVNGKRF GDSLQLGSSISFLCDEGFLGTQGSETITCVLKEGSSVVWNSAVLRCEAPCGH LTSPS  
GTILSPGWPGFYKDALS CAWVIEAQPGYPIKITFDRFKTEVNYDTLEVRDGR TYSAPLIGVYHGTQVPQFLISTSNYLYL  
LFSTDKSHSDIGFQLRYETITLQSDHCLDPGIPVNGQRHGNDFYVGALVTFSCDSGYTLSDGEPLECEPNFQWSRALPSC  
EALCGGFIQSSSGTILSPGFPDFYPNNL NCTWIIETSHGKG VFFTFTHTFHLES GHDYLLITENG SFTQPLRQLTG SRLPA  
PISAGLYGNFTAQVRFISDFSMSYEGFNITFSEYDLEPCEEPEVPAYSIRKGLQFGVGDTLTFSCFPGYRLEGTARITCL  
GRRRLWSSPLPRCVAECGNSVTGTGTLLSPNFPVNYNNNHECIYSIQTPGKGIQLKARAFELSEG DVLKVYDGNNS  
ARLLGVF SHSEMMGVTLNSTSSSLWLD FITDAENTSKGFELHFSSFELIKCEDPGTPKFYKVHDEGHFAGSSVSFSCDP  
GYSLRGSEELLCLSGERRTWDRPLPTCVAECGGTVRGEVSGQVLSPGYPAPYEHNLCIWTIEAEAGCTIGLHFLVFDTE  
EVHDVLR IWDGPVESGVLLKELSGPALPKDLHSTFNSVVLQFST DFFTSKQGF AIQFSVSTATSCNDPGIPQNGSRSGDS  
WEAGDSTVFQCDPGYALQGS AEISCVKIENRFFWQPSPTCIAPCGDLTGPSGVILSPNYPEPYPPGKECDWKVTVSPD  
YVIALVFNIFNLEPGYDFLHIYDGRDSLPLIGSFYGSQLP GRIESSNSLFLAFRSDASVSNAGFVIDY TENPRESCFD  
PGSIKNGTRVGS DLKLGSSVTYYCHGGYEVEGTSTLSCILGPDGKP VNNPRPVCTAPCGGQYVGS DGVVLSPNYPQNYT  
SGQICLYFVTVPKDYVVFQGFAFFHTALNDVVEVHDGHSQHSRLSSLSGSH TGESLPLATSNQVL IKFSAKGLAPARGF  
HFVYQAVPRTSATQCSSVPEPRYGRKLGSDFSVGAIVRFECNSGYALQGSPEIECLPVPGALAQWNVSAPTCVVPCG GNL  
TERRGTILSPGFPEPYLNSLNCVWKIVPEGAGIQIVVSFVTEQNWDSLEVFDGADNTVTMLGSFSGTTVPALLNSTSN  
QLYLHFYSDISVSAAGFHLEYKTVGLSSCPEPAVPSNGVKTGERYL VNDVVSFQCEPGYALQGHAIHISCMPGTVRRWNYP

PPLCIAQCGGTVEEMEGVILSPGFPGNYPNSMDCSWKIALPVGFGAHIQFLNFSTEPNHDYIEIRNGPYETSRMMGRFSG  
SELPSSLLSTSHETTVYFHSQNRPGFKLEYQAYELQECPDPEPFANGIVRGAGYNVGQSVTFECLPGYQLTGHPVLT  
CQHGTNRNWDHPLPKCEVPCGGNITSSNGTVYSPGFPSYSSSQDCVWLITVPIGHGVRLNLSLLQTEPSGDFITIWGDP  
QQTAPRLGVFTRSMAKKTVQSSSNQVLLKFHRDAATGGIFAIAFSAYPLTKCPPPTILPNAEVVTENEENIGDIVRYRC  
LPGFTLVGNEILTCKLGTYLQFEGPPPICEVHCPTNELLTDSTGVILSQSYPGSYPQFQTCSWLVRVEPDYNISLTVEYF  
LSEKQYDEFEIFDGPSPGQSPLLKALSGNYSAPLIVTSSSNSVYLRWSSDHAYNRKGFKIRYSAPYCSLPRAPLHGFI LGQ  
TSTQPGGSIHFGCNAGYRLVGHSMAICTRHPQGYHLWSEAIPLCQALSCGLPEAPKNGMVFGKEYTVGTKAMYSCSEGYH  
LQAGAEATAECLDTGLWSNRNVPPQCVPTCPDVSSISVEHGRWRLIFETQYQFQAQLMLICDPGYYYTGQRVIRCQANG  
KWSLGDSTPTCRIISCGELPIPPNGHRIGTLSVYGATAIFSCNSGYTLVGSRVRECMANGLWSGSEVRCLAGHCGTPEPI  
VNGHINGENYSYRGSVVYQCNAGFRLIGMSVRICQQDHHWSGKTPFCVPITCGHPGNPVNGLTQGNQFNLNDVVKFVCNP  
GYMAEGAARSQCLASGQWSDMLPTCRIINCTDPGHQENSVRQVHASGPHRFSFGTTVSYRCNHGFYLLGTPVLSCQGDGT  
WDRPRPQCLLVSCGHPGSPPHSQMSGDSYTVGAVVRYSCIGKRTL VG NSTRMCGLDGHW TGSLPHCSGTSVGVC G DPGIP  
AHGIRLGDSFDPGTVMRFSCEAGHVLRGSSERTCQANGSWSGSQPECGVISCGNPGTPSNARVVFSDGLVFSSSI VYECR  
EGYYATGLLSRHCSVNGTWTGSDPECLVINCGDPGIPANGLRLGNDFRYNKT VTYQCVPGYMMESH RVSVLSCTKDRTWN  
GTKPVCKALMCKPPPLIPNGKVVGSDFMWGSSVTYACLEGYQLSLPAVFTCEGNGSWTGELPQCFPVFCGDPGVPSRGR  
EDRGFSYRSSVSFSCHPPLVLVGSPRRFCQSDGTWSGTQPSCIDPTLTTCADPGVPQFGIQNNSQGYQVGSTVLFRCQKG  
YLLQGSTTRTCLPNLTWSGTPDCVPHHCRQPETPTHANVGALDLP SMGYTLIYSCQEGFSLKGGSEHRTCKADGSWTGK  
PPICLAEVRPSGRPINTAREPPLTQALIPGDVFAKNSLWKGAYEYQGGKQPAMLRVTGFQVANSKVNATMIDHSGVELHL  
AGTYKKEDFHLLQVYQITGPVEIFMNKFKDDHWALDGHVSSSESSGATFIYQGSVKGGFGGFGFQRLDLRLLESDPESI  
GRHFASNSSSVAAAAILVPFIALIIAGFVLYLYKHRRRPKV PFNGYAGHENTNVRATFENPMYDRNIQPTDIMASEAEFTV  
STVCTAV

[ ]: