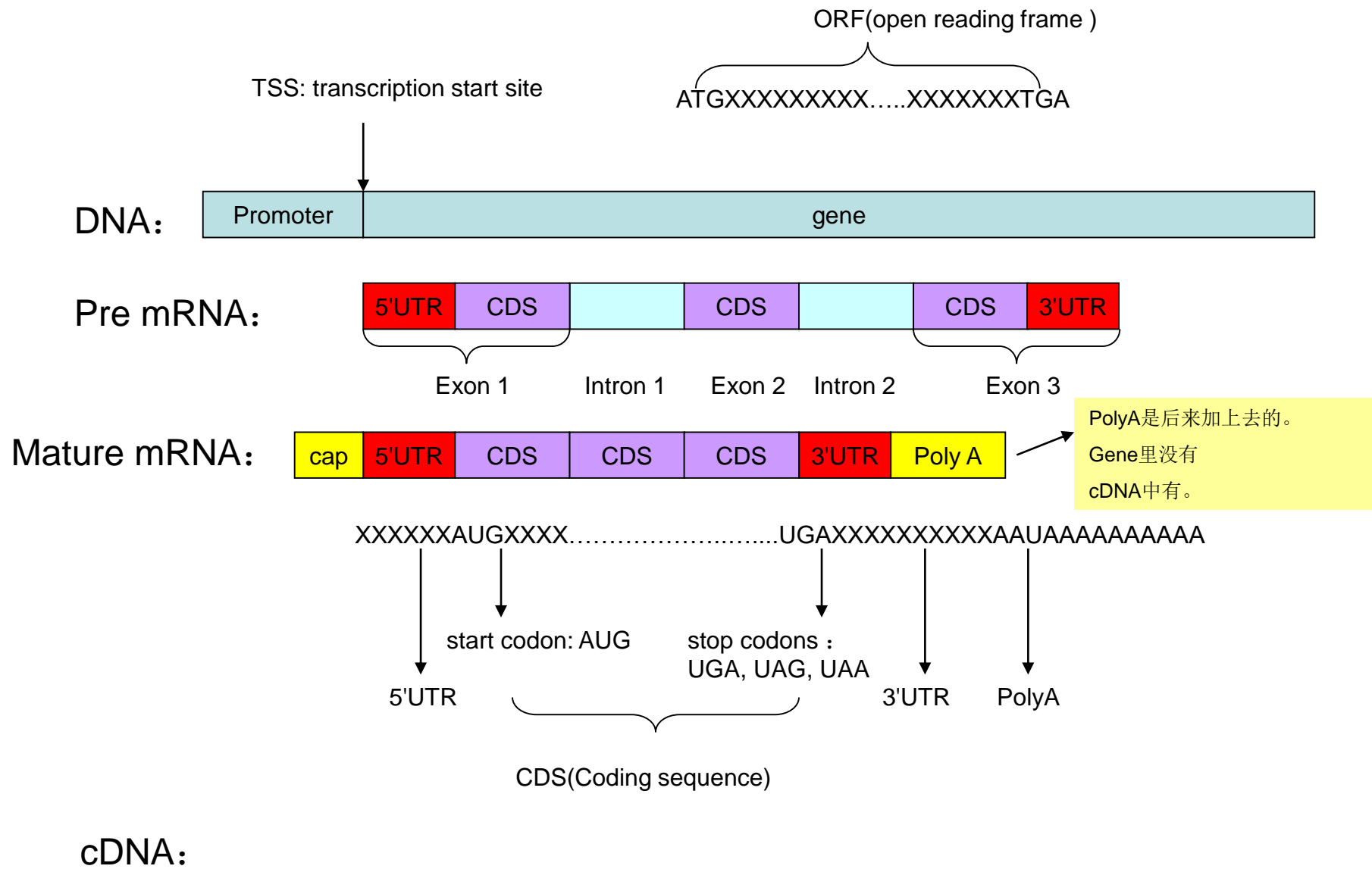


彻底搞清楚promoter, exon, intron, and UTR

- 启动子: RNA聚合酶特异性识别和结合的DNA序列。
- promoter自然不属于intron和Exon的任何一个, 属于noncoding sequence。
- noncoding RNA是现在研究的热点之一。我们常见的MiRNA, SiRNA, antisense RNA tech, 这些都是属于ncRNA的范围。只要你在进一步问下: 这些RNA是哪里来的? 你就知道部分答案, 跟那些看似跟编码蛋白没有关系的DNA序列有关系。这部分DNA有个统称就junk DNA, 垃圾DNA或者冗余DNA, 他们编码的RNA就属于 ncRNA. RNAi就是迄今最经典的ncRNA功能典范。
- An exon is a sequence of DNA that is expressed (transcribed) into RNA and then often, but with many noteworthy exceptions[1], translated into protein. Adjacent exons may be separated by an intron, which is later removed from the RNA transcript via the splicing mechanism. (From Wikipedia)
- UTR (Untranslated Regions)即非翻译区, 是信使RNA (mRNA) 分子两端的非编码片段。
- 5'-UTR从mRNA起点的甲基化鸟嘌呤核苷酸帽延伸至AUG起始密码子, 3'-UTR从编码区末端的终止密码子延伸至多聚A尾巴 (Poly-A) 的末端。

转录 (Transcription) 是遗传信息从DNA到 RNA的转移。即以双链DNA中的一条链为模板, 以ATP、CTP、GTP和UTP 4种核苷三磷酸为原料, 在RNA聚合酶催化下合成RNA的过程。



Example1

IL17A interleukin 17A[Homo sapiens]

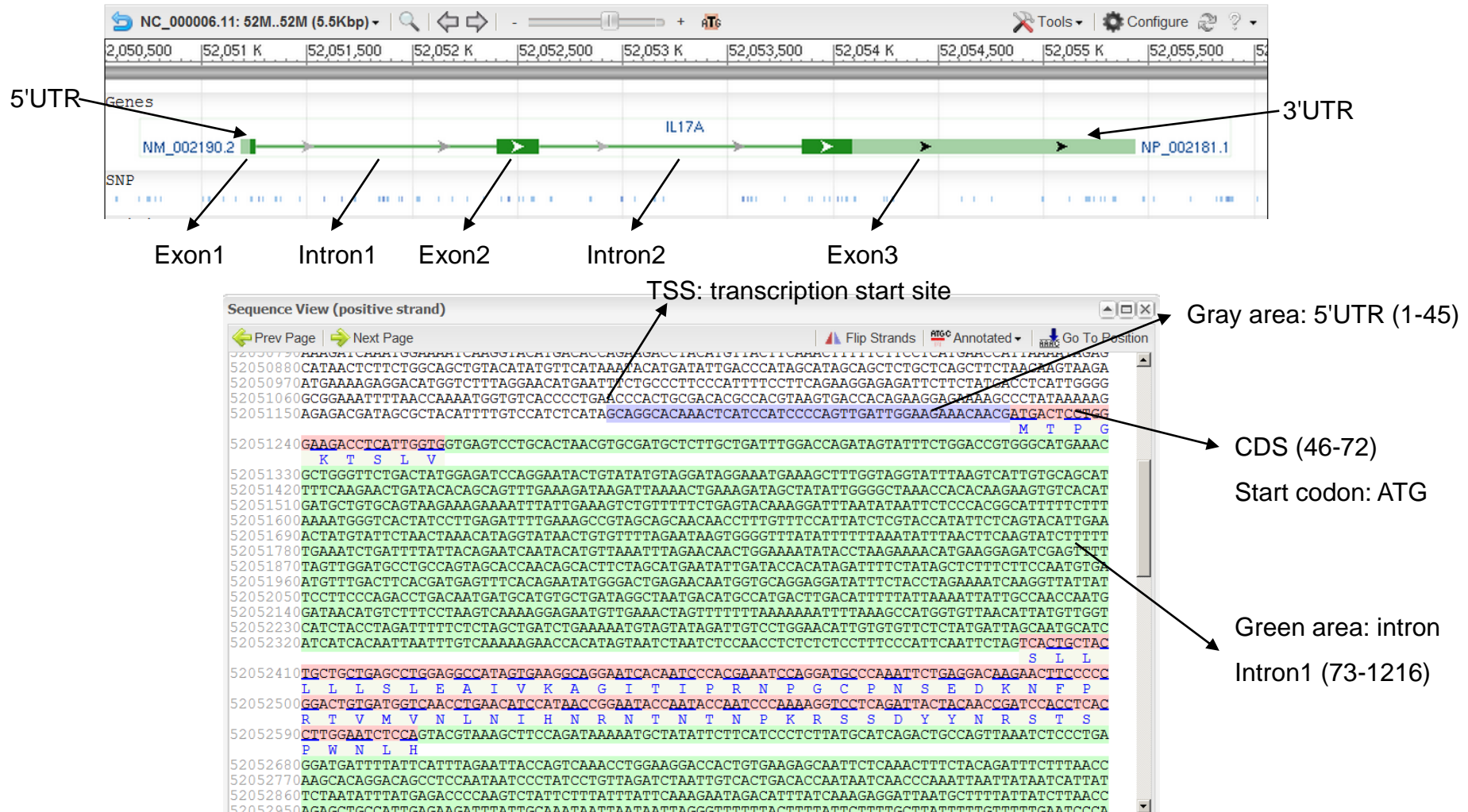
Location : 6p12

Sequence : Chromosome: 6; NC_000006.11 (52051185..52055436)

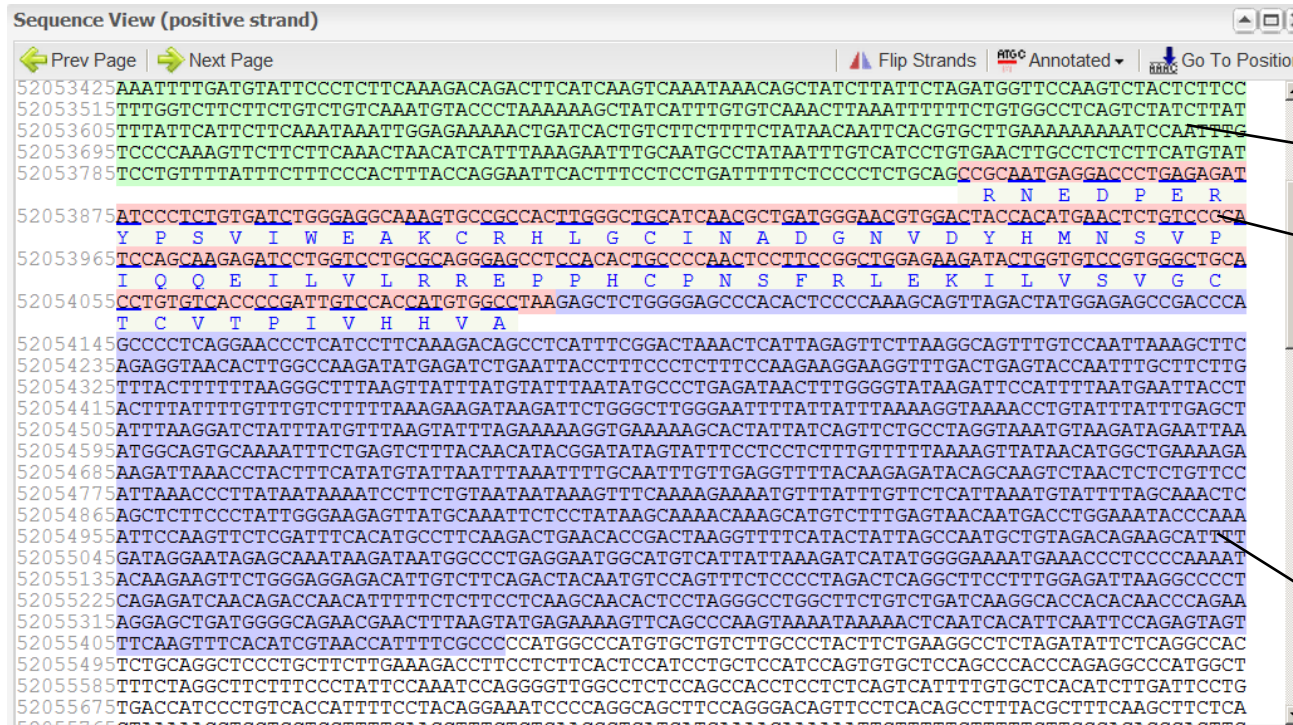
gene 1..4252

mRNA join(1..72,1217..1419,2669..4252)

CDS join(46..72,1217..1419,2669..2906)



CDS join(46..72,1217..1419,2669..2906)



Gray area: 3'UTR (2907-4252)

Example2:

Cx3cr1 chemokine (C-X3-C) receptor 1 [*Mus musculus*]

Location : 9 F4; 9

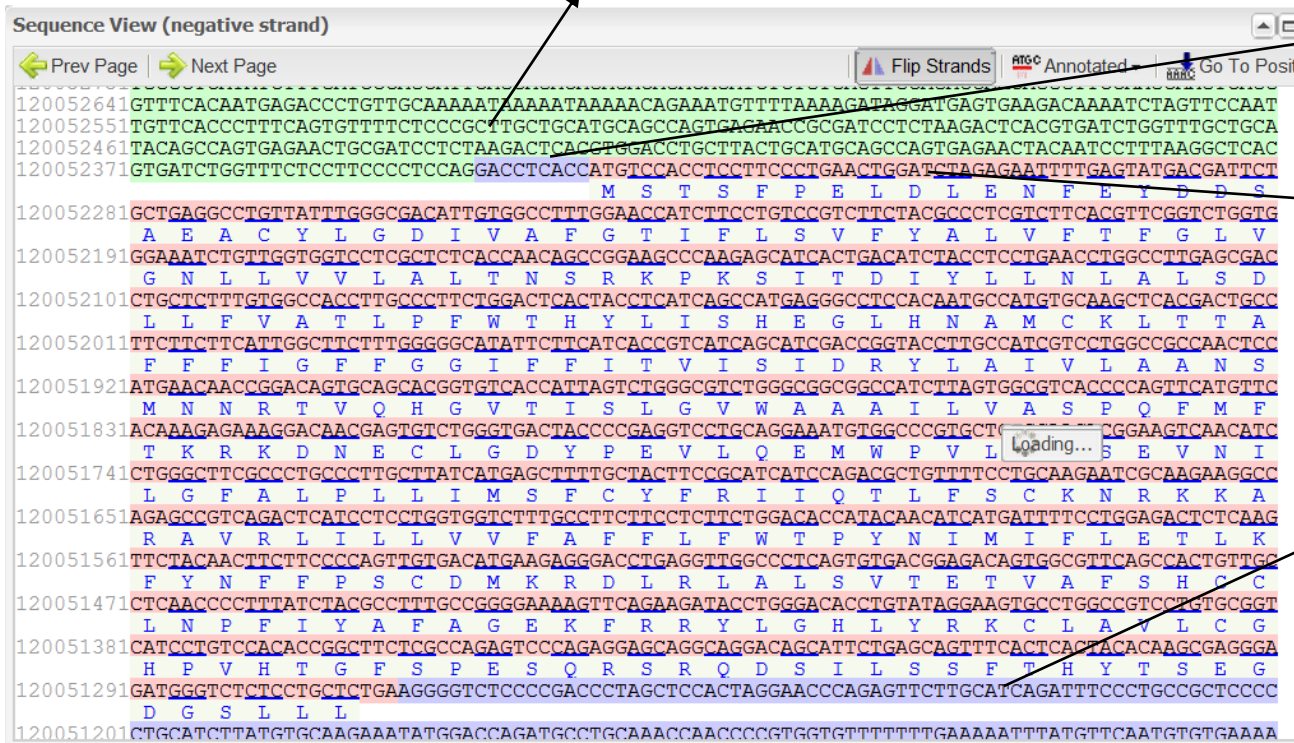
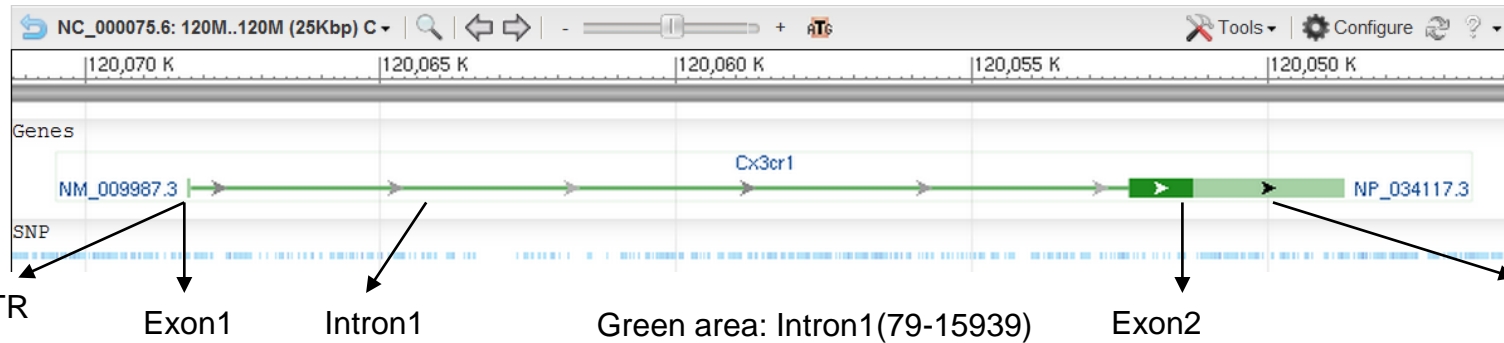
Sequence : Chromosome: 9; NC_000075.6 (120048693..120068282, complement)

gene 1..19590

mRNA join(1..78,15940..19590)

CDS 15949..17013

互补链上



Gray area: 5'UTR
(15940-15948) in exon2

CDS (15949-17013)

Start codon: ATG

End codon: TGA

Gray area: 3'UTR
(17014-19590) in exon2

常见问题：

问：Promoter在DNA序列中是算内含子还是外显子？

答：都不是。属于noncoding sequence。

问：UTR在DNA序列中是算内含子还是外显子？

答：外显子。

问：起始密码子ATG在gene的哪个位置？

答：外显子内，CDS以ATG开始。

问：每个基因之间都有一定的间隔序列，这些基因间的间隔序列应该不属于内含子吧？

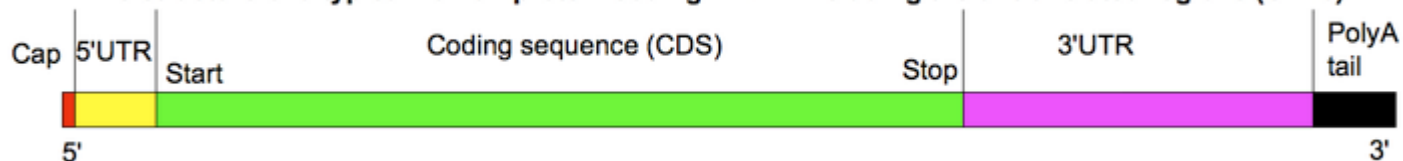
答：是的。属于noncoding sequence。

问：如何寻找启动子区域和预测转录因子结合位点？

答：见我的帖子<http://www.dxy.cn/bbs/topic/23466429>



The structure of a typical human protein coding mRNA including the untranslated regions (UTRs)



目的：

- 寻找promoter区域
- 预测Transcription factor binding site

举例：

预测人基因ANKH上游2000bp启动子区域中NF-kB的结合位点

寻找promoter区域

1. 用NCBI: <http://www.ncbi.nlm.nih.gov/>
2. 用UCSC: <http://www.genome.ucsc.edu/>
3. 用Ensembl: <http://www.ensembl.org/index.html>
4. 用公司信息（只包含公司拥有promoter clones的信息）：
<http://www.genecopoeia.com/>

寻找promoter区域

- NCBI <http://www.ncbi.nlm.nih.gov/pubmed/>
- 选择Gene, 输入ankh, 点击search
- 选择第一项, 人类Homo sapiens的ANKH
- Chromosome 5 location 14704909-14871887, complement(反义链)即-14871887 到 -14704909为基因范围

此例中选取-14873887 到-14871887 约2000bp核苷酸序列作为启动子区域

ANKH ankylosis, progressive homolog (mouse) [*Homo sapiens*]

Gene ID: 56172, updated on 10-Jul-2012

Summary	
Official Symbol	ANKH provided by HGNC
Official Full Name	ankylosis, progressive homolog (mouse) provided by HGNC
Primary source	HGNC:15492
Locus tag	UNQ241/PRO274
See related	Ensembl:ENSG00000154122 ; HPRD:05509 ; MIM:605145 ; Vega:OTTHUMG00000090539
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	ANK; CMDJ; HANK; MANK; CCAL2; CPPDD
Summary	This gene encodes a multipass transmembrane protein that is expressed in joints and other tissues and controls pyrophosphate levels in cultured cells. Progressive ankylosis-mediated control of pyrophosphate levels has been suggested as a possible mechanism regulating tissue calcification and susceptibility to arthritis in higher animals. Mutations in this gene have been associated with autosomal dominant craniometaphyseal dysplasia. [provided by RefSeq, Jul 2008]

Genomic context

Location: 5p15.1

Sequence: Chromosome: 5; NC_000005.9 (14704909..14871887, complement)

See ANKH in [Epigenomics](#), [MapView](#)

寻找promoter区域

- 点击Graphics----Tools-----Sequece Text View

ANKH gene在反义链上，所以用负数表示

输入序列可以查询染色体位置

图形显示

FASTA格式显示的核苷酸序列

可以查询具体核苷酸序列

Genomic re...s, transcripts, and products

Genomic Sequence NC_000005 chromosome 5 reference chr37.p5 Primary Assembly

Go to [reference sequence details](#)

Go to nucleotide [Graphics](#) [FASTA](#) [GenBank](#)

NC_000005.9 -15M...-15M (217Kb-) Find on Sequence:

Genes

NM_054027.4 UQCRBP3

MIR4637 NR_039780.1

RBBP4P1

ANKH

LOC100128508 XR_132508.1

Genes - tRNA

SNP

Clinical Channel

Cited Variants

Association Results

Tools

Go To

Search

Flip Strands

Markers

Set Origin

Sequence Text View

Add new Panel

Add new Panel on Range

BLAST and Primer Search

Download

Printer-Friendly Page

寻找promoter区域

- 点击Go To Position, 输入-14873887, 点击Prev Page找到具体位置
- 复制白底黑色区域即为promoter区域。



白底黑字
为启动子
区域

紫底黑字
为基因区
域

粉底黑字为编码区，ATG为
 启示密码子

寻找promoter区域

- 在前两张幻灯片中选择FASTA
- 在右边Change region shown输入14871887到14873887
- Display options选择Show reverse complement
- 可以直接得到FASTA格式的promoter核苷酸序列（似乎有一个bp的差距，可以输入14871887到14873886）

Display Settings: FASTA

Send:

Showing 2.00kb region from base 14871887 to 14873886.

Homo sapiens chromosome 5, GRCh37.p5 Primary Assembly

NCBI Reference Sequence: NC_000005.9

[GenBank](#) [Graphics](#)

```
>gi|224589817:c14873886-14871887 Homo sapiens chromosome 5, GRCh37.p5 Primary Assembly
TTTGTTTTAGGAAAAAAAAAACCCTCTTTCCAGAGAAATCCITTTAAATAACAAGCTCCCTTTTGTATAT
TTAAGTTAGATTCAAATATATTAACACACACCTTTTCCAGGAACCTTGTCTCTGAAGACAGTCCGTGCCACAT
GAACCTCCTGTTTGTCTGGGAAATCTTTTAAATACAGTATTGCTATTTGACACATTAAACAATAAATACT
TACCAAGTTTAATTTTAATCATAGGGAATGTAATTTGAGGTTACATCAGAGCTACAAGGTAAATTAGCTT
CCTATTTTTCAAAGCCTGCTCAAATTTGTTGGAATTTCAAAAAAGTATTTTAAAAAGGATGCAAAATATG
GAAATATTTTATTTATTTATTTATTTTATTTTGTAGAGGGAGTCTCACTCTGTCAACCCAGGCTGGAGTGCCA
TGGGTGACGCTGCAACCTCCGCTCTTCCGGGTTCAAACCATTTCTCTGCTTAGCCTCTTGAGTAGCTGGG
GCTACCGGGGCCACCGCTCACGCCGGCCTATGGAACATTTAACACTTTCAAAGTAGTTGGGAAATAC
CTGAACCCAGGATATTGAGGTCTCAAGCTCACTATTTACAAAAGCTTTACAAAGCTAAATTGTAGTTAA
TCGGTTCAITCTAGGTACAAGATAGTTTCTTACTTTTAAATAAATAAATTACTTTTACTTCGTTATTTTATT
AAATGCAGAGATGTGTGTAGAGTTGGTATAAAATAGTAAGACCAGATAGGCTACATACTATGCCAGGC
CCACTACAAAATGAAAATGCAGAGACCTTGTTCAAATTCAGAGTGCACAGCAGCGCGTTAGTCCTTT
TAAGTGGGGCCCTCTGTGACTTCACAGATCGCACGCCGTGAAGCCAGCCCTGCCAGAGTCCCTTTTC
ATTCACTGGGTCTCTCCCTCGCGTCACAGTCTGGGGTGGCGGGATGCTGGCGTTGCTGTCACTGTC
CGAGCTTTAGGTCAAACTCTCAGCACCGGACTAGTTACCTGAAATCACTGGGGCCCGGTATTACAAA
CAGAACACACTTTTCTCTTCTTCCCTTTTATTTAAAAAGAGATGATAGGCTTTACGAGTTATTTTAAATA
GAAATTTTACAAATGAAAAGAAATTCGGGCGTACTGTTTGCAAACTGCAGCAGACAGGCTGAGCGCCCC
AGAAAGCGCGCGGCCAGCCTGGGAGCCTGGCGGGACCACTCCCTGCTGCCCTGGCGCATGTCCTCCAG
GGACACGGGCCACCGTTGCCCTGGAACCTGGGGGACAGGCTCCCTCCTGGGCGACCTTTCCACGTTT
```

Change region shown

- ☐ Whole sequence
☒ Selected region

from: 14871887 to: 14873886

Update View

Customize view

Display options

- ☒ Show reverse complement

Update View

Analyze this sequence

[Run BLAST](#)

[Pick Primers](#)

Related information

[Assembly](#)

[BioAssay by RNA target](#)

寻找promoter区域

- <http://www.genome.ucsc.edu/> 选择genomes
- 在clade选择Mammal, genome选择Human, assmeby选择最新的数据库, gene中输入ANKH
- 点击Tables
- 在track中选择RefSeq Genes, 在output format中选择sequence
- 点击get output。 选择genomic。

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequences. For more information, see this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation. To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and [Annotation Downloads](#) page.

clade: genome: assembly:

group: track:

table:

region: ☐ genome ☐ ENCODE Pilot regions ☒ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: Send output to ☐ [Galaxy](#) ☐ [GREAT](#)

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

寻找promoter区域

- 选择Promoter/Upstream by 2000 bases
- Exons in upper case, everything else in lower case外显子大写，其他小写

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

RefSeq Genes Genomic Sequence

Sequence Retrieval Region Options:

☒ Promoter/Upstream by bases

☒ 5' UTR Exons

☒ CDS Exons

☒ 3' UTR Exons

☒ Introns

☐ Downstream by bases

☒ One FASTA record per gene.

☐ One FASTA record per region (exon, intron, etc.) with extra bases upstream (5') and extra

☐ Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they

Sequence Formatting Options:

☒ Exons in upper case, everything else in lower case.

☐ CDS in upper case, UTR in lower case.

☐ All upper case.

☐ All lower case.

☐ Mask repeats: ☒ to lower case ☐ to N

寻找promoter区域

- 选择Promoter/Upstream by 2000 bases
- Exons in upper case, everything else in lower case外显子大写，其他小写

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

RefSeq Genes Genomic Sequence

Sequence Retrieval Region Options:

☒ Promoter/Upstream by bases

☒ 5' UTR Exons

☒ CDS Exons

☒ 3' UTR Exons

☒ Introns

☐ Downstream by bases

☒ One FASTA record per gene.

☐ One FASTA record per region (exon, intron, etc.) with extra bases upstream (5') and extra

☐ Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they

Sequence Formatting Options:

☒ Exons in upper case, everything else in lower case.

☐ CDS in upper case, UTR in lower case.

☐ All upper case.

☐ All lower case.

☐ Mask repeats: ☒ to lower case ☐ to N

寻找promoter区域

```
>hg19_refGene_NM_054027 range=chr5:14704909-14873887 5'pad=0 3'pad=0 strand=- repeatMasking=none
```

```
atttgttttaggaaaaaaaaaacctctttcccagagaattccttttaaata  
acaagctccctttttgtatatttaagtttagattcaaatatattaaacac  
acctttcaggaaacttgctcctgaagacagtcctgtccacatgaacctcct  
gttttctgctgggaaatcttttaaatacagtattgctatttgacacatttaa  
caataaatacttaccaagttttaattttaatcataggggaatgtaatttgag  
gttacatcagagctacaaaggttaatttagcttcctatttttcaaagcctgc  
tcaaatgtgtggaatttcaaaaaagtattttaaaaggatgcaaaatat  
ggaaatattttattttattttatttttttggagaggagctcactcct  
gtcaccacaggtggagtgccatgggtgacgctgcaacctccgtcttcggg  
gttcaaaccattctcctgccttagcctcttgagtagctggggctacccggc  
gcccacgctcacgccccggcctatggaaacatttaacactttcaaagtag  
ttgggaaatacctgaaccaggatattgaggtctcaagctcactatttaac  
aaaagctttacaaagctaaattgtgagtttaacggttcattctaggtaca  
agatagtttcttacttttaataataaattactttacttcgttatatttat  
taaatgcagagatgtgtgtgagagttggtataaaatagtaagaccagat  
aggctacatactatgccaggccactacaaaatgaaaatgcagagaccct  
tgttcaaattcaagagtgcaacagcagcgcttagtccttttaagtgcgg  
gccctctgtgacttcacagatcgcaagccgctgaagccagccctgccag  
agtcctctttcattcactgggtcctcctcctcgcgtcacagtcctgggg  
tgccgggatgctggcggtgctgtcacactgccagcttttaggtcaaactc  
tcagcacccgcgactagttaccctgaaatcactggggcccggtattacaaa  
acagaacacactttcctctttcttccctttattttaaagagatgatagg  
ctttacgagttatttttaatagaaattttttacaatgaaaagaattcggg  
cgtactgtttgcaaaactgcagcagacaggtctgagcgccccagaaggcgg  
ccggccccagcctgggagcctgcccgggaccactccctgctgccctggcgc  
atgtcctccagggaacagggccaccgttgccctggaaaccctgggggac  
aggctccctcctgggagcctttccacgtttctttcgcccccccgaggc  
ggcacctccatacctccggggccgctagtccccgaagcagggccccagcc  
ccggcgcaagctgcctcctccgcagtccccgcccgcgcagcccgccca  
ccgcccggagcggcccttcccagccccgcgccccgctccccagccctcct  
cgcgctccgagcccacgtcgcttgccccggctgccaagtcttccagacg  
gccggggccctcccgccacagcgccgggaacccccgcccccgccccaggg  
acgagggtacagggccccggccggcgagccagggcaacccccgggtctcca  
ggcgggccacccgctcaccccccacccggcgtccccaaagccccagaccg  
tccccgcgcgcctcagcgccgcgcgcccccgggccggccgaggtccgc  
ggcgagtccecgccagcgccccggaggcgccagccccacggccccagcgt  
cgcgagcgcccccccgcgcccgccgaagcgacggcgacaggaaag  
gaggcccgggcgcgccccggccggccccctccccagcccgccccggggc  
cgctggcggtgtccgcgcctccgcgcgccccctgatttcctccggggcg  
cggcgggcgggcgggcgccctcgctgcgcgcggttattttttccggggcccc  
AGAGCGCTTATAATGGAGCCGCTGTGAGCAGAACCTTCTGCCGCCCGCGC  
CGCCGCCCGCGCTCCCTCCTCTTTTTTTTCCCGGCAGATCTTTGTTGTGTG  
GGAGGGCAGCAGGGATGGACTTGAGCTTGCAGATCCCCTGCTAGAGCAGC  
CGCGCTCGGAGAAGGCGCCGCGAGCCGAGGAGAGCGCCGCGCCGCGCG  
CCCGAGGCCCCCGCGCCCGCGCGCCTCTGTGCGCCCGCGCCCCGCTCGCCC  
CGTCCGCCCGTCCGCCCTCGCCTCCCGCAGAGTCCCCTCGCGGCAGCAG  
ATGTGTGTGGGGTCAGCCACGCGGGGACTATGGTGAATTCCTGGCGC  
TCACGCACTACTGGCCCTGATCCGGTCTTGGTGCCCTGGGCATCACC  
AACATAGCCATCGACTTCGGGGAGCAGtaagccccggccccggccccca  
-----
```


小写字母为promoter区域

大写字母为基因区域，与NCBI结果相同

ATG为CDS区起始密码子

寻找promoter区域

- <http://www.ensembl.org/index.html> 选择human 输入 ankh
- 选择Gene, 点击 GeneID [ENSG00000154122](#)
- 点击左边的Export data

 BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (GRCh37) ▾ Location: 5:14,704,910-14,871,887 Gene: ANKH

Gene-based displays

- Gene summary
- Splice variants (8)
- Supporting evidence
- Sequence
- External references
- Regulation
- Comparative Genomics
- Phenotype
- Genetic Variation
 - Variation Table
 - Variation Image
 - Structural Variation
- External Data
 - Personal annotation
- ID History
 - Gene history

Configure this page

Manage your data

Export data

Bookmark this page

Gene: ANKH ENSG00000154122

Description ankylosis, progressive homolog (mouse) [Source:HGNC Symbol;Acc:15492]

Location [Chromosome 5: 14,704,910-14,871,887](#) reverse strand.

Transcripts ☐ This gene has 8 transcripts

Show/hide columns

Filter

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
ANKH-001	ENST00000284268	8203	ENSP00000284268	492	Protein coding	CCDS3885
ANKH-201	ENST00000535119	1722	ENSP00000442524	294	Protein coding	-
ANKH-002	ENST00000505140	1866	ENSP00000426332	126	Nonsense mediated decay	-
ANKH-004	ENST00000503389	287	No protein product	-	Processed transcript	-
ANKH-007	ENST00000503939	610	No protein product	-	Processed transcript	-
ANKH-003	ENST00000513115	697	No protein product	-	Retained intron	-
ANKH-005	ENST00000502585	830	No protein product	-	Retained intron	-
ANKH-006	ENST00000515517	615	No protein product	-	Retained intron	-

Transcript and Gene level displays

In Ensembl we provide displays at two levels:

寻找promoter区域

- 5 Flanking sequence 输入2000
- Options for FASTA sequence中Genomic选5 Flanking sequence,
- deselect all
- 点击Next

The screenshot shows the Ensembl 'Export data' interface. The 'Export Configuration - Feature List' section is active, displaying the 'Gene to export' as ENSG00000154122 (ANKH). The 'Output' is set to 'FASTA sequence' and the 'Strand' is 'Feature strand'. The '5 Flanking sequence (upstream)' field is set to 2000, which is circled in red. The '3 Flanking sequence (downstream)' field is set to 0. A 'Next >' button is visible. Below this, the 'Options for FASTA sequence' section shows a dropdown menu for 'Genomic' set to '5 Flanking sequence'. Under 'Select/deselect all:', there are checkboxes for 'cDNA:', 'Coding sequence:', 'Peptide sequence:', '5' UTR:', '3' UTR:', 'Exons:', and 'Introns:', all of which are currently unchecked.

Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (G) Export data

Gene-based

Export Configuration - Feature List

Gene to export: ENSG00000154122 (ANKH)

Output: FASTA sequence *

Strand: Feature strand *

5' Flanking sequence (upstream): 2000 * (Maximum of 1)

3' Flanking sequence (downstream): 0 * (Maximum of 1)

Next >

Fields marked * are required

Options for FASTA sequence

Genomic: 5' Flanking sequence *

Select/deselect all:

cDNA: ☐

Coding sequence: ☐

Peptide sequence: ☐

5' UTR: ☐

3' UTR: ☐

Exons: ☐

Introns: ☐

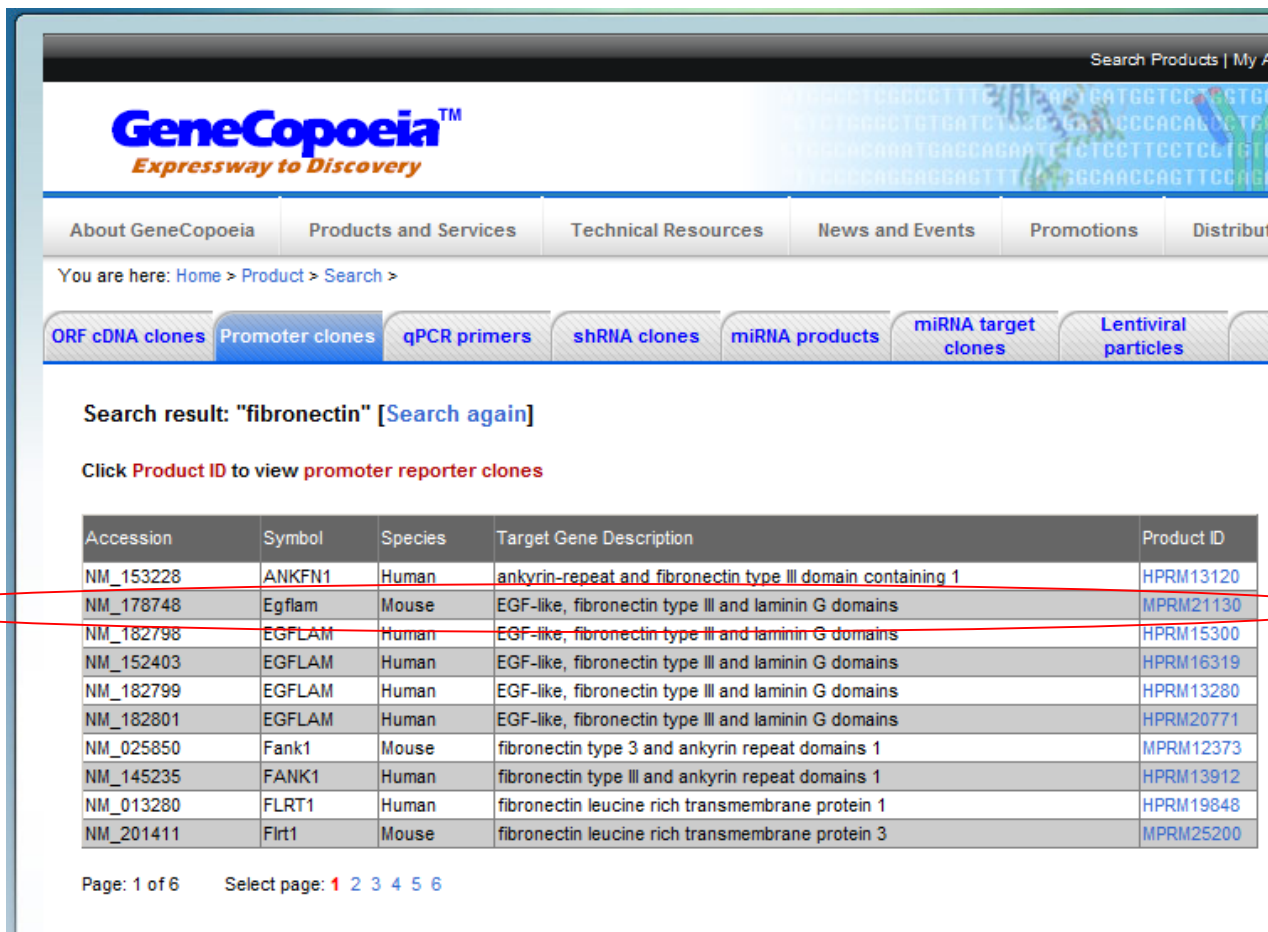
寻找promoter区域

- 得到2000 bases 的核苷酸序列

```
>5' Flanking sequence chromosome:GRCh37:5:14871888:14873887:-1
ATTTGTTTATAGGAAAAAAAAAACCTCTTTCCAGAGAATTCCTTTAAATAACAAGCTCCC
TTTTTGATATTTAAGTTAGATTCAAATATATTTAAACACACCTTTCAGGAAGTTGCTCC
TGAAGACAGTCCTGTCCACATGAACCTCCTGTTTTGCTGGGAAATCTTTAAATACAGTA
TTGCTATTTGACACATTTAAACAATAAATACTTACCAAGTTTAATTTAATCATAGGGAAT
GTAATTTGAGGTTACATCAGAGCTACAAGGTAAATTAGCTTCCTATTTTTCAAAGCCTGC
TCAAATTGTGTGAATTTCAAAAAAGTATTTTAAAAAGGATGCAAATATGGAAATATTT
ATTTATTTATTTATTTATTTTTGAGAGGGAGTCTCACTCTGTCACCCAGGCTGGAGTGCC
ATGGGTGACGCTGCAACCTCCGTCTTCCGGGTTCAAACCATTCCTGCCTTAGCCTCTT
GAGTAGCTGGGGCTACCGGCGCCACCGCTCACGCCCGGCCTATGGAAACATTTAACACT
TTCAAAGTAGTTGGGAAATACCTGAACCCAGGATATTGAGGTCTCAAGCTCACTATTTAC
AAAAGCTTTACAAAGCTAAATTGTGAGTTAATCGGTTCACTTAGGTACAAGATAGTTTC
TTACTTTTAATAATAAATTACTTTACTTCGTTATATTTATTAATGCAGAGATGTGTGTG
AGAGTTGGTATAAAATAGTAAGACCACGATAGGCTACATACTATGCCAGGCCCACTACAA
AATGAAAATGCAGAGACCCTTGTTCAAATTCAAGAGTGCAACAGCAGCGCGTTAGTCCTT
TTAAGTGCGGGCCCTCTGTGACTTCACAGATCGCACGCCCGTGAAGCCAGCCCTGCCCAG
AGTCCCTTTTCATTCACTGGGTCTCTCCCTCGCGTCACAGTCCTGGGGTGCGGGGATG
CTGGCGTTGCTGTCACTGCCGAGCTTTAGGTCAAACCTCTCAGCACCGCGACTAGTTAC
CCTGAAATCACTGGGGCCCGGTATTACAAAACAGAACACACTTTCTCTTTCTTCCCCTT
TATTTAAAAGAGATGATAGGCTTTACGAGTTATTTTAAATAGAAATTTTTACAATGAAA
AGAATTCGGGCGTACTGTTTGCAAACCTGCAGCAGACAGGTCTGAGCGCCCCAGAAAGCGG
CCGGCCCCAGCCTGGGAGCCTGCGGGGACCACTCCCTGCTGCCCTGGCGCATGTCTCCA
GGGACACGGGCCACCGTTGCCCTGGAAACCTGGGGGACAGGCTCCCTCCTGGGCGACC
TTTCCACGTTTCTTTCGCCCCCCCCGGAGCGGCACCTCCATACCTCCCGGGCCGCTAGTC
CCCCGAAGCAGGCCCCAGCCCCGGCGCAAGCTGCCTCCTCCGAGTCCCCGCCCCCGGCC
AGCCCGGCCACGCCGACGGCGCCCTTCCAGCCCCGCGCCCCGCTCCCCAGCCCTCCT
CGCGCGTCCGAGCCACGTCGCTTGCCCCGGCTGCCAAGTCTTCCAGACGGCCGGGCCCT
CCGGGCACAGCGCGGAACCCCGGCCCGGCCCGGCCAGGGACGAGGGCTACAGGGCCCGG
CCGGCGAGCCAGGGCACCCCGGGGTCTCAGGCGGCCACCGCCCTACCCCCACCGGC
GTCCCCAAGCCCCAGACCGTCCCCGCGCGCCCTCAGCGCGCGCCCCGGGGCGGGC
GAGGCTCCGCGCGAGTCCCGCCAGCGCCCCGAGGCGCCAGCCCCACGGCCCGAGCGT
GCGCAGCGCCCCCGGGCGCGCCAAGCGCAGGCGACGGCACAGGAAAGGAGGCGCGGG
CGCGCCCGGCCCGGCCCTCCCGAGCCCCCGGGGCGGCTGGCGGTGTCCGCGCCT
CCGCGCCCGCCCTGATTTCTCGGGCGGGCGGGCGGGCGGGCGCGCTCGTGCAGCG
GTTATTTATTTCCGGGCCCC
```

寻找promoter区域

- <http://www.genecopoeia.com/>
- 点击search product, 选择promoter clones, 因为没有ANKH的信息, 此处输入FIBRONECTIN
- 选择目的基因



Search Products | My Account

GeneCopoeia™
Expressway to Discovery

About GeneCopoeia | Products and Services | Technical Resources | News and Events | Promotions | Distributors

You are here: [Home](#) > [Product](#) > [Search](#) >

[ORF cDNA clones](#) | [Promoter clones](#) | [qPCR primers](#) | [shRNA clones](#) | [miRNA products](#) | [miRNA target clones](#) | [Lentiviral particles](#)

Search result: "fibronectin" [[Search again](#)]

Click **Product ID** to view promoter reporter clones

Accession	Symbol	Species	Target Gene Description	Product ID
NM_153228	ANKFN1	Human	ankyrin-repeat and fibronectin type III domain containing 1	HPRM13120
NM_178748	Egflam	Mouse	EGF-like, fibronectin type III and laminin G domains	MPRM21130
NM_182798	EGFLAM	Human	EGF-like, fibronectin type III and laminin G domains	HPRM15300
NM_152403	EGFLAM	Human	EGF-like, fibronectin type III and laminin G domains	HPRM16319
NM_182799	EGFLAM	Human	EGF-like, fibronectin type III and laminin G domains	HPRM13280
NM_182801	EGFLAM	Human	EGF-like, fibronectin type III and laminin G domains	HPRM20771
NM_025850	Fank1	Mouse	fibronectin type 3 and ankyrin repeat domains 1	MPRM12373
NM_145235	FANK1	Human	fibronectin type III and ankyrin repeat domains 1	HPRM13912
NM_013280	FLRT1	Human	fibronectin leucine rich transmembrane protein 1	HPRM19848
NM_201411	Flrt1	Mouse	fibronectin leucine rich transmembrane protein 3	MPRM25200

Page: 1 of 6 Select page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

寻找promoter区域

- 点击click here to view the promoter sequence
- 得到promoter信息

ORF cDNA clones Promoter clones qPCR primers shRNA clones miRNA products miRNA target clones

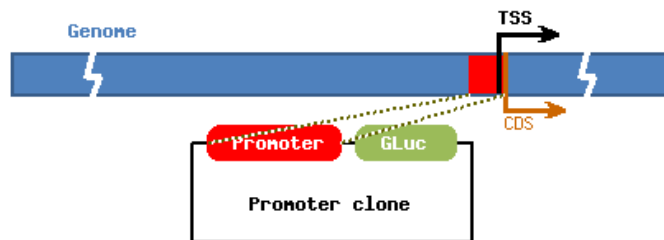
Select promoter reporter clones

Product ID: MPRM21130 (click here to view the promoter sequence)

Symbol: [Egflam](#)
Species: Mouse
Target Gene Accession: [NM_178748](#) (click here to view gene annotation page)
Gene Description: EGF-like, fibronectin type III and laminin G domains
Delivery format: 10 µg purified plasmid
Estimated Delivery: Please email sales@geneCopoeia.com or call 301-762-0888

Custom promoter

Want to customize the own vector?
[Request a quote here.](#)



GeneCopoeia™
Expressway to Discovery

About GeneCopoeia Products and Services Technical Resources News and Events

You are here: [Home](#) > [Product](#) > [Search](#) >

ORF cDNA clones Promoter clones qPCR primers shRNA clones miRNA products

Sequence [Back]

>MPRM21130 promoter

```
gggtaggggaaagatccagggtagaaatcacaggacaacactcagctctcctaataatggaag
attaacaaggggacaggaagctgctctgccaggtggggcggtggcgatcacccctcagacgt
ttgccagtgaaccacagtttctatctgcggggcacttcctagggcctttgatctttccag
gattctggcgagggtcgttttcttttagattattattgtatctgttttcacgaaggc
ttggggtcagaaagccagtttgctcgggagctcacacccttggggcagtactgtaaact
ccagccagctgcacagatctcttggaaggagggggtcttctcctcagctaggagctaa
atgctggctcccgcatgcccctactcccagtcocggcctggccgcaagcgctcccttt
atagaggaccttctcagcccaagaagtttagactttgccacaataacctggactttttt
ctcttttctgggggaggggggtgtctaaagtggcaccaccaccagaaaggggacctgg
ccccaggcaccagtccttaattctcatttggcctttgctctgagtgccctagcctcag
ctctccatctgtgoccttaaatccatttatctcttagcaccgaagacttgccagcg
ccgcctcctcctcctgcttgaaaattacagccctcctctgctcttctccaagagtgc
ctctacagagtagagaaacctttgaggtattaggaagtggtttaaactgggagcaagc
cacctagaggggtatgggggaagaaaggtccagccagcgaaaccgaagcaccaaacaaa
agttaaggggttaggatccagccgagggacctgggcccagctctgagtcagggaacacgc
ggcgagtgtcgaagtcggtcccaccccgcccccaatctaggtcaccgcggggaagggc
tgagctgggggctgtcggcgccggaggtggcgcgagtggaattgattcagaccctgcc
ctcgttcttccactagattgtaattccatctcctcggggtcgagccgcctcctccggc
ggccagtgcgctccctccctcggctctcagcgtccactccccagccccgaaagccttgc
cgctgcgagcccgctccgctcaagcacgccagcttcccactgtgccttgccctcgg
actcccgagtggtcctgggtcagctcctgggtctcccacttctcctcctcggctcac
```

预测Transcription factor binding site

1. 用Jaspar <http://jaspar.genereg.net/>
2. 用PROMO http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3
3. 用TFSEARCH（据说用的是TRANSFAC很旧的数据库）
<http://www.cbrc.jp/research/db/TFSEARCH.html>
4. 用商业数据库TRANSFAC（要付费）
<http://www.gene-regulation.com/pub/databases.html/>

预测Transcription factor binding site

<http://jaspar.genereg.net/> 点击JASPAR CORE vertebrata

左边转录因子选择MA0061.1 NF-kappaB，右边输入ANKH启动子区域，点击
SCAN

结果得到5个 Transcription factor binding site，其中Strand -1没有特殊意义，另外三个GGGAAATACC得分最高

5 putative sites were predicted with these settings (80%) in sequence named gil224589817:c14873887-14871887							
Model ID	Model name	Score	Relative score	Start	End	Strand	predicted site sequence
MA0061.1	NF-kappaB	6.814	0.800574249111552	34	43	1	GAGAATTCCT
MA0061.1	NF-kappaB	8.606	0.843971565747803	553	562	-1	GGTATTTCCT
MA0061.1	NF-kappaB	10.772	0.896426134851665	553	562	1	GGGAAATACC
MA0061.1	NF-kappaB	6.807	0.800404728343442	1554	1563	1	GGGCCCTCCC
MA0061.1	NF-kappaB	7.528	0.817865367458808	1751	1760	-1	GGGACTCGCC

预测Transcription factor binding site

http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3

Step1 selectspecies选择human

Step1 SelectFactors选择NF-kappaB [T00590]

Step2 SearchSites输入ANKH的promoter区域

结果中有一个位点TGGGAAATACCT，与JASPAR结果中得分最高的相同

0 NF-kappaB [T00590] was predicted in:

gi|224589817:c14873887-14871887 Homo sapiens chromosome 5, GRCh37.p5 Primary Assembly TGGGAAATACCT

552 563

Dissimilarity

9.69%

RE equally

0.01777

RE query

0.02072

Consensus sequence and matrix:

A	1	2	1	10	5	7	0	0	0	0	0	3
C	2	0	0	0	1	1	0	0	13	13	13	6
G	8	11	12	3	4	1	0	0	0	0	0	3

预测Transcription factor binding site

<http://www.cbrc.jp/research/db/TFSEARCH.html>

Enter a label for the sequence: 输入基因名字ANKH

Enter your DNA sequence 输入ANKH的promoter区域

点击submit

结果中有2个NF- κ p位点，其中正义链GGGAAATACC，与JASPAR结果中得分最高的相同

```

551 CGCTCAGGCC CGGCCTATGG AACATTAA CACTTTCAAA GTAGTTGGGA entry score
----- M00087 Ik-2 88.6
----- M00120 dl 88.2
----- M00054 NF-kap 88.2
<-- M00053 c-Rel 87.6
-----> M00029 HSF 86.9
----- M00029 HSF 86.9
----- M00109 C/EBPb 86.2
<-- M00052 NF-kap 86.1
-----> M00028 HSF 85.9
----- M00028 HSF 85.9
----- M00141 Lyf-1 85.7
<----- M00033 p300 85.1

601 AATACCTGAA CCCAGGATAT TGAGGTCICA AGCTCACTAT TTACAAAAGC entry score
-----> M00253 cap 92.0
-----> M00076 GATA-2 90.1
-----> M00271 AML-1a 88.7
----> M00087 Ik-2 88.6
-----> M00120 dl 88.2
-----> M00054 NF-kap 88.2
-----> M00101 CdxA 87.9
-----> M00053 c-Rel 87.6
-----> M00100 CdxA 87.2
-> M00029 HSF 86.9
<----- M00253 cap 86.5
-----> M00211 Poly A 86.4
-----> M00029 HSF 86.3
-----> M00109 C/EBPb 86.2
-----> M00052 NF-kap 86.1
-> M00028 HSF 85.9
-----> M00074 c-Ets- 85.8
-> M00141 Lyf-1 85.7
-----> M00075 GATA-1 85.7

```