

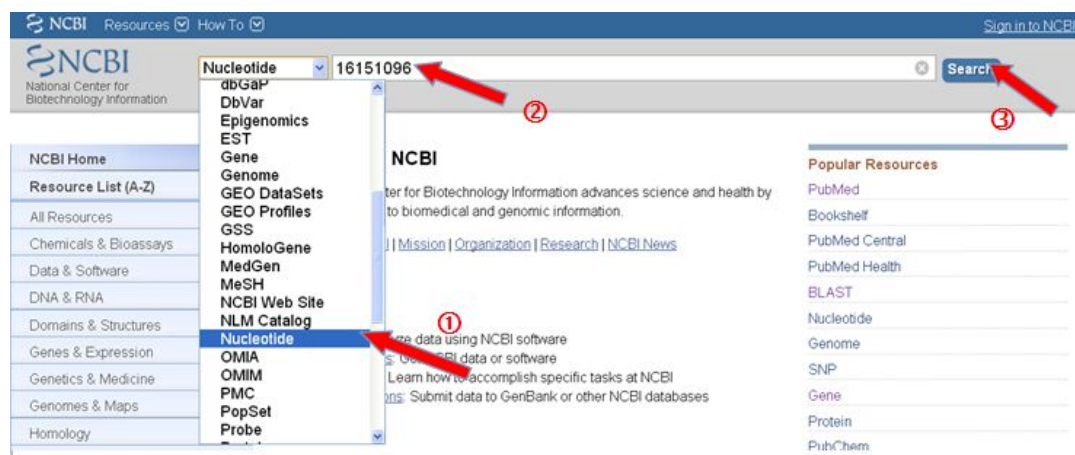
## 如何获得基因序列? ——在 NCBI 中查找目的基因实例

基因, 即具有遗传效应的 DNA 片段, 是控制生物性状的基本遗传单位。基因有两个特点, 一是能忠实地复制自己, 以保持生物的基本特征; 二是基因能够“突变”, 突变绝大多数会导致疾病, 因此研究某种疾病的病理要从基因开始, 研究基因当然就要知道该基因序列了, 如何查找基因就是关键问题了, 这里, 我们将举例详细讲解一下如何查找基因。

### 1. 首先可以根据文献获得目的基因序列

通过阅读文献, 找到你感兴趣的基因, 根据文中提到的该基因在 NCBI 中的 ID 号, 直接打开 <http://www.ncbi.nlm.nih.gov>, 在 All Databases 后的下拉框中选择 Nucleotide, 把基因 ID 号输入 Search 前面的文本框中, 点“Search”, 就可以找到该基因了。

举例说明, 例如: 在 2003 年 JBC 的文章 (Conditional Knock-out of Integrin-linked Kinase Demonstrates an Essential Role in Protein Kinase B/Akt Activation) 中出现了“calreticulin (GenBank accession number gi 16151096)”, 那么把“16151096”输入 Search 前面的文本框中, 点“Search”, 就可以找到该基因了 (当然包括基因序列等相关信息), 见下图。



检索结果界面如下图, 可以看到 GenBank 号为 AY047586 的 CALR 基因的相关信息了。



NCBI Resources How To Sign in to NCBI

Nucleotide **Nucleotide** Search Limits Advanced Help

Display Settings:  GenBank Send:

### Homo sapiens calreticulin (CALR) mRNA, complete cds

GenBank: AY047586.1 

[FASTA](#) [Graphics](#)

Go to:

LOCUS AY047586 1402 bp mRNA linear PRI 15-OCT-2001  
 DEFINITION Homo sapiens calreticulin (CALR) mRNA, complete cds.  
 ACCESSION AY047586  
 VERSION AY047586.1 GI:16151096  
 KEYWORDS .  
 SOURCE Homo sapiens (human)  
 ORGANISM [Homo sapiens](#)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Change region shown

Customize view

Analyze this sequence

- Run BLAST
- Pick Primers
- Highlight Sequence Features
- Find in this Sequence

Articles about the CALR gene

里面有很多基因的信息，再往下是基因的的核酸序列（ORIGIN 之后）。

post-translational cleavage"

ORIGIN

```

1 ctgccggagg gtggttttaa agggcccgcg cgttgccgcc cccctggccc gccatgctgc
61 tatccgtgac gctgctgctc ggccctcctcg gcttgccgct cgcocagcct gccgtctact
121 tcaaggagca gtttctggac ggagacgggt ggacttcccg ctggatcgaa tccaaacaca
181 agtcagattt tggcaaatc gttctcagtt cggcaagtt ctacggtgac gaggagaag
241 ataaagttt gcagacaagc caggatgcac gcttttatgc totgtggccc agtttcagac
301 ctttcagcaa caaaggccag acgctggggt tgcagttcac ggtgaacat gacgagaaca
361 togactgtgg gggcggctat gtaagctgt ttctaatag ttggaccag acagacatgc
421 acggagactc agaatacaac atcatgtttg gtcccagat ctgtggccct ggcaaccaaga
481 agttcatgt catcttcaac tacaaggcca agaactgct gatcaacaag gacatccgtt
541 gcaagatga tgattttaca caoctgtaca cactgattgt gcggccagac aacacctatg
601 aggtgaagat tgacaacagc caggtaggt cggctcctt ggaagacgat tgggacttcc
661 tgcccaccaa gaagataaag gatcctgat ctccaacc ggaagactgg gatgagcgg
721 ccaagatga tgatccaca gactccaagc ctgaggactg ggacaagccc gacatatacc
781 ctgacctga tgctaagaag cccgaggact gggatgaaga gatggacgga gattggaac
841 cccagtgat tcagaacct gactacaagg gtgagtgga gccccggcag atcgacaacc
901 cagattacaa gggcacttgg atccaccag aaattgaca ccccgatgat tctcccgatc
961 ccagtatcta tgccatgat aactttggcg tctgggccc ggaaccttgg caggtcaagt
1021 ctggaccatc ctttgacaac ttctcatca ccaacgatga gccatacgtc gaggagtttg
1081 gcaacgagac gtggggcgta acaaaggcag cagagaaca aatgaaggac aaacaggacg
1141 agagacagag gcttaaggag gaggagaag acaagaaac caaagaggag gaggaggcag
1201 agacacagca gcatatagc gacaaagc agatagaca gcatagagca gacagagagc

```

Homologs of the CALR gene

The CALR gene is conserved in chimpanzee, Rhesus monkey, dog, cow, mouse, rat, chicken, zebrafish, fruit fly, mosquito, and C. elegans.

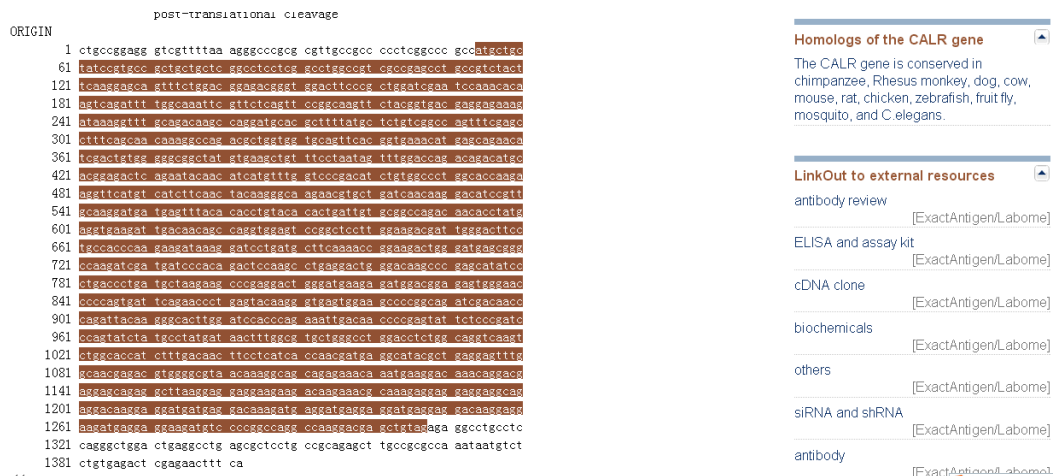
LinkOut to external resources

- antibody review [\[ExactAntigen/Labome\]](#)
- ELISA and assay kit [\[ExactAntigen/Labome\]](#)
- cDNA clone [\[ExactAntigen/Labome\]](#)
- biochemicals [\[ExactAntigen/Labome\]](#)
- others [\[ExactAntigen/Labome\]](#)

基因的翻译区（CDS）点击 CDS 即可得到。



下图标示的褐色区域序列即为基因的编码区序列。



这里需要指出一下，在显示基因的页面右下侧有一个 LinkOut to external resource，里面是与该基因相关的链接，对于该基因的相关研究是很有用的。



181 agtcagattt tggcaaatc gttctcagtt ccggcaagtt ctacggtagc gaggagaag  
 241 ataaaggttt gcagacaagc caggatgcac gcttttatgc tctgtggcc agtttcgagc  
 301 ctttcagcaa caaaggccag acgctgggtg tgcagttcac ggtgaacat gacgagaaca  
 361 tcgactgtgg gggcggtat gtagagctgt ttctaatag ttggaccag acagacatgc  
 421 acggagactc agaatacaac atcatgtttg gtcccacat ctgtggccct ggcacaaga  
 481 aggttcattg catcttcaac tacaaggcca agaacgtgct gatcaacaag gacatcgttt  
 541 gcaaggatga tgagtttaca cacctgtaca cactgattgt gggccagac aacacctatg  
 601 aggtgaagat tgacaacagc caggtggagt ccgctcctt ggaagacat tgggacttc  
 661 tgccaccocaa gaagataaag gatcctgatg cttcaaaacc ggaagactgg gatgagcggg  
 721 caagatgca tgatccocaa gactccaagc ctgaggactg ggaagaccoc gacatatacc  
 781 ctgacctcga tgctaagaag ccgagagact gggatgaaga gatggacgga gactgggaac  
 841 ccccagtagt tcaagaacct gactacaagg gtgagtgaa gccccggcag atcgacaacc  
 901 cagattacaa gggcacttgg atccaccag aaattgacaa ccccgagtat tctccgagc  
 961 ccagtatcta tgccctatgat aactttggcg tgctgggctt ggacctctgg caggtcaagt  
 1021 ctggcaccat ctttgacaac ttocctatca ccaacgatga ggcatacgt gaggatgttg  
 1081 gcaacgagac gtggggccta acaaaaggcag cagagaacaa aatgaaggac aacaggacg  
 1141 aggacagag gcttaaggag gaggagaag acaagaacg caaagaggag gaggaggcag  
 1201 aggaacaaga gtagtagag gacaagatg aggatgagga gtagtagag gacaaggagg  
 1261 aagatgagga ggaagatgtc cccggccagg ccaagacga gctgtagaga ggcctgcctc  
 1321 caggctgga ctgaggcctg agcgtcctg ccgacagact tgccgcgca aataatgct  
 1381 ctgtgagact cgagaacttt ca

mouse, rat, chicken, zebrafish, fruit fly, mosquito, and C.elegans.

LinkOut to external resources

- antibody review [ExactAntigen/Labome]
- ELISA and assay kit [ExactAntigen/Labome]
- cDNA clone [ExactAntigen/Labome]
- biochemicals [ExactAntigen/Labome]
- others [ExactAntigen/Labome]
- siRNA and shRNA [ExactAntigen/Labome]
- antibody [ExactAntigen/Labome]
- protein and peptide [ExactAntigen/Labome]

如果你只想获得序列（例如去设计 PCR 引物的时候），那就可以选择 FASTA，这样就得到了 FASTA 格式的序列文件，没有其他数字和格式的干扰。



NCBI Resources How To Sign into NCBI

Nucleotide Nucleotide Search Limits Advanced Help

Display Settings: GenBank Send: Change region shown

**Homo sapiens calreticulin (CALR) mRNA, complete cds**

GenBank: AY047586.1

[FASTA](#) [Graphics](#)

Go to:

LOCUS AY047586 1402 bp mRNA linear PRI 15-OCT-2001

DEFINITION Homo sapiens calreticulin (CALR) mRNA, complete cds.

ACCESSION AY047586

VERSION AY047586.1 GI:16151096

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1402)

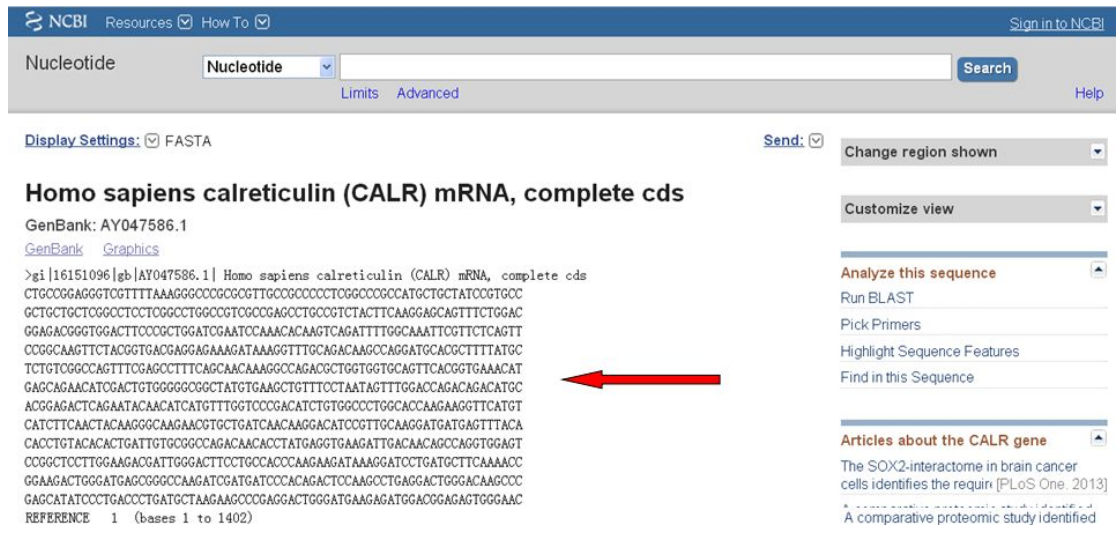
Analyze this sequence

- Run BLAST
- Pick Primers
- Highlight Sequence Features
- Find in this Sequence

Articles about the CALR gene

- The SOX2-interactome in brain cancer cells identifies the requir [PLoS One. 2013]
- A comparative proteomic study identified

这就是 FASTA 格式的序列：



NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Limits Advanced Help

Display Settings: FASTA Send: Change region shown

**Homo sapiens calreticulin (CALR) mRNA, complete cds**

GenBank: AY047586.1  
[GenBank](#) [Graphics](#)

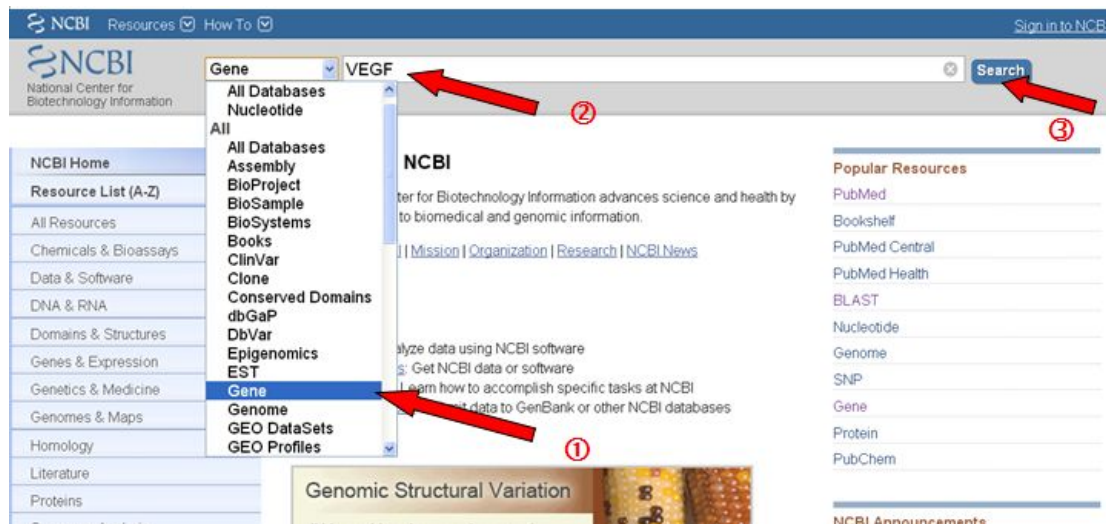
>gi|116151096|gb|AY047586.1| Homo sapiens calreticulin (CALR) mRNA, complete cds  
 CTGCCGAGGGTGGTTTTAAAGGGCCCGCGGTTGCCGCCCTGGGCCGCATGCTGCTATCCGGTCC  
 GCTGCTGCTGGGCTCTGGGCTGGCGCTGGCCGAGCTGCGCTCACTTCAAGGAGCAGTTTCTGGAC  
 GGAGACGGGTGGACTTCCCGCTGGATCGAATCCAAACACAAGTCAGATTTTGGCAATTGCTTCTCAGTT  
 CCGCAAGTTCTACGGTGACGAGGAAAGATAAAGGTTTGCAGACAAGCCAGGATGCACGCTTTATGC  
 TCTGTGGCCAGTTTCAGCTTTTCAGCAACAAGGCCAGACGCTGGTGGCAGTTACGGTGAACAT  
 GAGCAGAATCGACTGTGGGGGGGCTATGTGAAGCTGTTTCTAATAGTTTGGACACAGACATCC  
 ACGGAGACTCAGAAATACAACATCATGTTTGTGGCCGACATCTGTGGCCCTGGCACCAGAAAGTTTCATGT  
 CACTTCAACTACAAGGGCAAGAACGTGCTGATCAACAAGGACATCCGTTGCAAGGATGATGAGTTTACA  
 CACTGTACACACTGATTGTGGCCGACAGACAACCTATGAGCTGAAGATTGACAACAGCCAGTGGAGT  
 CCGGCTCTTGAAGACGATTGGGACTTCTGCCACCAAGAAGATAAAGGATCTGATGCTTCAAACC  
 GGAAGACTGGGATGAGCGGGCCAGATGATGATCCACAGACTCCAAAGCTGAGGACTGGGACAAGCC  
 GAGCATATCCCTGACCTGATGCTAAGAAGCCGAGGACTGGGATGAAGAGATGGACGGAGCTGGGAAC  
 REFERENCE 1 (bases 1 to 1402)

Analyze this sequence  
 Run BLAST  
 Pick Primers  
 Highlight Sequence Features  
 Find in this Sequence

Articles about the CALR gene  
 The SOX2-interactome in brain cancer cells identifies the require [PLoS One. 2013]  
 A comparative proteomic study identified

## 2. 根据已经获得的基因的相关信息查找

如果只是知道基因的名字，怎么查序列呢？还是举例说明，比如研究的基因名称是人的 VEGF 基因，那么怎么在 NCBI 中找到它呢？首先打开 <http://www.ncbi.nlm.nih.gov/> 在 All Databases 的下拉框中选择 Gene，然后在中间的文本框中输入基因名称“VEGF”，点击 Search...



NCBI Resources How To Sign in to NCBI

Gene VEGF Search

All Databases Nucleotide

All Databases  
 Assembly  
 BioProject  
 BioSample  
 BioSystems  
 Books  
 ClinVar  
 Clone  
 Conserved Domains  
 dbGaP  
 DbVar  
 Epigenomics  
 EST  
 Gene  
 Genome  
 GEO DataSets  
 GEO Profiles

NCBI  
 Center for Biotechnology Information advances science and health by  
 to biomedical and genomic information.  
 | Mission | Organization | Research | NCBI News

Popular Resources  
 PubMed  
 Bookshelf  
 PubMed Central  
 PubMed Health  
 BLAST  
 Nucleotide  
 Genome  
 SNP  
 Gene  
 Protein  
 PubChem

NCBI Announcements

搜索结果如下：



结果有很多条，哪一条是我想要的基因呢？这时候要根据自己研究的基因所属物种来选择，如研究的是人属（*Homo sapiens*）的，则点击第四条，出现如下界面。



里面是这个基因的详细信息，需要指出的是，在 NCBI 中，基因有很多别名（Aliases），你得到的基因名和 NCBI 中记录的名称有可能不一致。比如在这里，VEGFA 是 NCBI 中记录的基因名称，而它还有很多别名，比如 VPF，VEGF（这就是我们要找的基因名称），MVCD1。

**VEGFA vascular endothelial growth factor A [ *Homo sapiens* (human) ]**

Gene ID: 7422, updated on 21-Aug-2013

**Summary**

<b>Official Symbol</b>	VEGFA provided by HGNC
<b>Official Full Name</b>	vascular endothelial growth factor A provided by HGNC
<b>Primary source</b>	HGNC:12889
<b>Locus tag</b>	RP1-261G23.1
<b>See related</b>	Ensembl:ENSG00000112715, HPRD:01889, MIM:192240, Vega:OTTHUMG00000014745
<b>Gene type</b>	protein coding
<b>RefSeq status</b>	REVIEWED
<b>Organism</b>	<i>Homo sapiens</i>
<b>Lineage</b>	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominoidea; Homo
<b>Also known as</b>	VPF; VEGF; MVCD1
<b>summary</b>	This gene is a member of the PDGF/VEGF growth factor family and encodes a protein that is often found as a disulfide linked homodimer. This protein is a glycosylated mitogen that specifically acts on endothelial cells and has various effects, including mediating increased vascular permeability, inducing angiogenesis, vasculogenesis and endothelial cell growth, promoting cell migration, and inhibiting apoptosis. Elevated levels of this protein is linked to POEMS syndrome, also known as Crow-Fukase syndrome. Mutations in this gene have been associated with proliferative and nonproliferative diabetic retinopathy. Alternatively spliced transcript variants, encoding either freely

Summary
Genomic context
Genomic regions, transcripts, and products
Bibliography
Phenotypes
HIV-1 protein interactions
Interactions
Pathways
General gene information
Gene Ontology
General protein information
Reference sequences
Related sequences
Additional links
Locus-specific Databases

<b>Related information</b>
Order cDNA clone
3D structures
BioAssay

再往下看，可以看到里面可以看到该基因再染色体上的位置，以及基因在转录时有几个剪切体等信息。

这个基因有很多转录本（isoform a 到 isoform r），可以看到其的 mRNA 的链接（如 [NM\\_001025366.2](#)）和蛋白质的链接（如 [NP\\_001020537.2](#)）

**mRNA and Protein(s)**

[NM\\_001025366.2](#) → [NP\\_001020537.2](#) **vascular endothelial growth factor A isoform a**

Status: REVIEWED

<b>Description</b>	Transcript Variant: This variant (1) represents the <b>longest</b> transcript. This variant can initiate translation from four non-AUG (CUG) sites, and also from a downstream, in-frame AUG. The isoform (a) represented in this RefSeq is derived from the 5'-most CUG start codon. This is the <b>longest</b> isoform.
<b>Source sequence(s)</b>	<a href="#">AF024710</a> , <a href="#">BF700556</a> , <a href="#">BM661679</a> , <a href="#">BU153227</a> , <a href="#">CN256173</a> , <a href="#">DA237741</a> , <a href="#">S85192</a>
<b>Consensus CDS</b>	<a href="#">CCDS34457.1</a>
<b>UniProtKB/Swiss-Prot</b>	<a href="#">P15692</a>

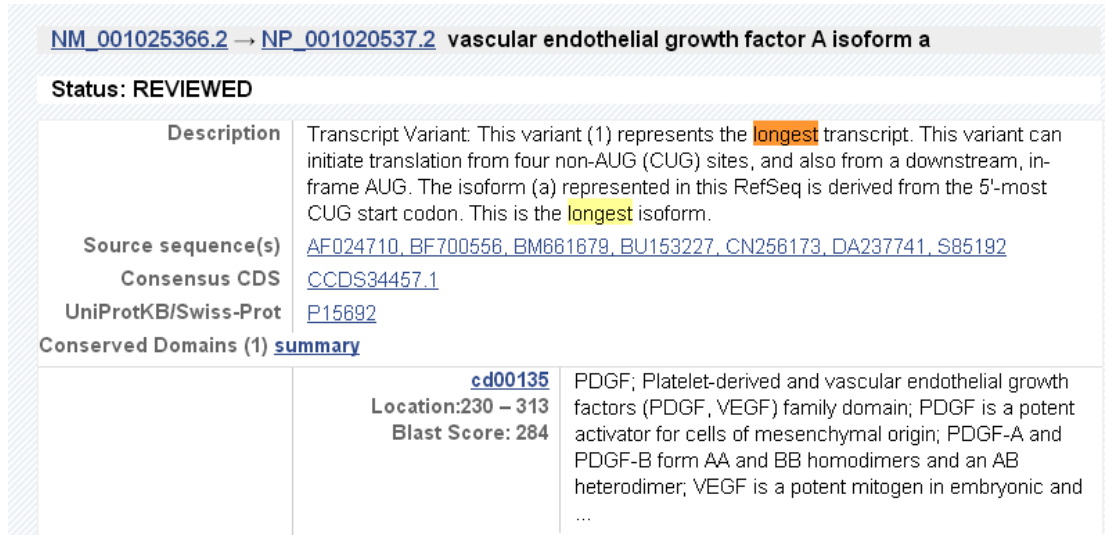
Conserved Domains (1) [summary](#)

<b>cd00135</b>	PDGF; Platelet-derived and vascular endothelial growth factors (PDGF, VEGF) family domain; PDGF is a potent activator for cells of mesenchymal origin; PDGF-A and PDGF-B form AA and BB homodimers and an AB heterodimer; VEGF is a potent mitogen in embryonic and ...
<b>Location:230 – 313</b>	
<b>Blast Score: 284</b>	

[NM\\_001025367.2](#) → [NP\\_001020538.2](#) **vascular endothelial growth factor A isoform c**

Status: REVIEWED

isoform a 到 isoform r 哪个是自己想找的基因呢？这就需要根据自己查阅的文献以及在这些基因序列后面的解释来确定了。客户提供基因的 ID 号更好，如果不提供，那么我们一般选择众多 mRNA 转录本中最长的转录本（longest isoform），即下图中所标示的 isoform a。



**NM\_001025366.2** → **NP\_001020537.2** vascular endothelial growth factor A isoform a

Status: REVIEWED

<b>Description</b>	Transcript Variant: This variant (1) represents the <b>longest</b> transcript. This variant can initiate translation from four non-AUG (CUG) sites, and also from a downstream, in-frame AUG. The isoform (a) represented in this RefSeq is derived from the 5'-most CUG start codon. This is the <b>longest</b> isoform.
<b>Source sequence(s)</b>	<a href="#">AF024710</a> , <a href="#">BF700556</a> , <a href="#">BM661679</a> , <a href="#">BU153227</a> , <a href="#">CN256173</a> , <a href="#">DA237741</a> , <a href="#">S85192</a>
<b>Consensus CDS</b>	<a href="#">CCDS34457.1</a>
<b>UniProtKB/Swiss-Prot</b>	<a href="#">P15692</a>
<b>Conserved Domains (1) summary</b>	
	<a href="#">cd00135</a> <b>Location:230 – 313</b> <b>Blast Score: 284</b>
	PDGF; Platelet-derived and vascular endothelial growth factors (PDGF, VEGF) family domain; PDGF is a potent activator for cells of mesenchymal origin; PDGF-A and PDGF-B form AA and BB homodimers and an AB heterodimer; VEGF is a potent mitogen in embryonic and ...

如果要找的基因是第一个序列即 isoform a，就可以点击 [NM\\_001025366.1](#)，得到如下基因的信息界面：



NCBI Resources How To Sign in to NCBI

Nucleotide  Search Limits Advanced Help

Display Settings:  GenBank Send:

**Homo sapiens vascular endothelial growth factor A (VEGFA), transcript variant 1, mRNA**

NCBI Reference Sequence: [NM\\_001025366.2](#)

[FASTA](#) [Graphics](#)

Go to:

LOCUS [NM\\_001025366](#) 3677 bp mRNA linear PRI 12-AUG-2013

DEFINITION Homo sapiens vascular endothelial growth factor A (VEGFA), transcript variant 1, mRNA.

ACCESSION [NM\\_001025366](#)

VERSION [NM\\_001025366.2](#) GI:284172447

KEYWORDS RefSeq.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the VEGFA gene

The role of VEGF and other parameters in tracking the clinical course [J BUON. 2013]

A systematic review of vascular endothelial growth factor expression [Tumour Biol. 2013]

点击 [NP\\_001020537.2](#) 就可以获得该转录本基因翻译的蛋白信息。



