

# 生物信息学 引物设计

张高川

遗传学与生物信息学系 基础医学与生物科学学院

苏州大学医学部

Email: zhanggaochuan@suda.edu.cn

WX: zhanggaochuan770609

Tel: 18962111592 QQ: 2257916241



























All Databases V

Search

National Center for Biotechnology Information

#### NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

#### Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | NCBI News

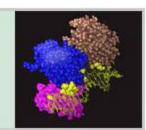
#### **Get Started**

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How Tos: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases

#### 3D Structures

Explore three-dimensional structures of proteins, DNA, and RNA molecules. Examine sequence-structure relationships, active sites, molecular interactions, biological activities of bound chemicals, and associated biosystems.

8



#### Popular Resources

PubMed

Bookshelf

PubMed Central PubMed Health

BLAST

Nucleotide

Genome

SNP Gene

Protein

PubChem

#### NCBI Announcements

NCBI webinar A Submitter's Guide to GenBank, Part 2 on January 7th Dec 31, 2014

On January 7th, NCBI will present the continuation of the December 17th

GenBank release 205.0 is now available via FTP

Dec 16, 2014

Release 205.0 (12/12/2014) has







nucleotide blast

Search a nucleotide database using a nucleotide query

Algorithms: blastn, megablast, discontiguous megablast

Search protein database using a protein query

Algorithms: blastp, psi-blast, phi-blast, delta-blast

blastx

Search protein database using a translated nucleotide query

tblastn

Search translated nucleotide database using a protein query

Search translated nucleotide database using a translated nucleotide query

#### Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with <u>Primer-BLAST</u>
- Cluster multiple sequences together with their database neighbors using MOLE-BLAST
- Find conserved domains in your sequence (cds)
- □ Find sequences with similar conserved domain architecture (cdart)
- Search sequences that have gene expression profiles (GEO)
- Search <u>immunoglobulins and T cell receptor sequences</u> (IgBLAST)
- Screen sequence for vector contamination (vecscreen)
- Align two (or more) sequences using BLAST (bl2seq)
- Search <u>protein</u> or <u>nucleotide</u> targets in PubChem BioAssay
- Search <u>SRA by experiment</u>
- Constraint Based Protein Multiple Alignment Tool
- Needleman-Wunsch Global Sequence Alignment Tool
- Search <u>RefSegGene</u>
- Search trace archives

#### Tip of the Day

### How to save custom search pages.

So you have made a few BLAST searches and after adjusting the database, organism limits and maybe a few Algorithm Parameters you arrive at what you think is a good search strategy.

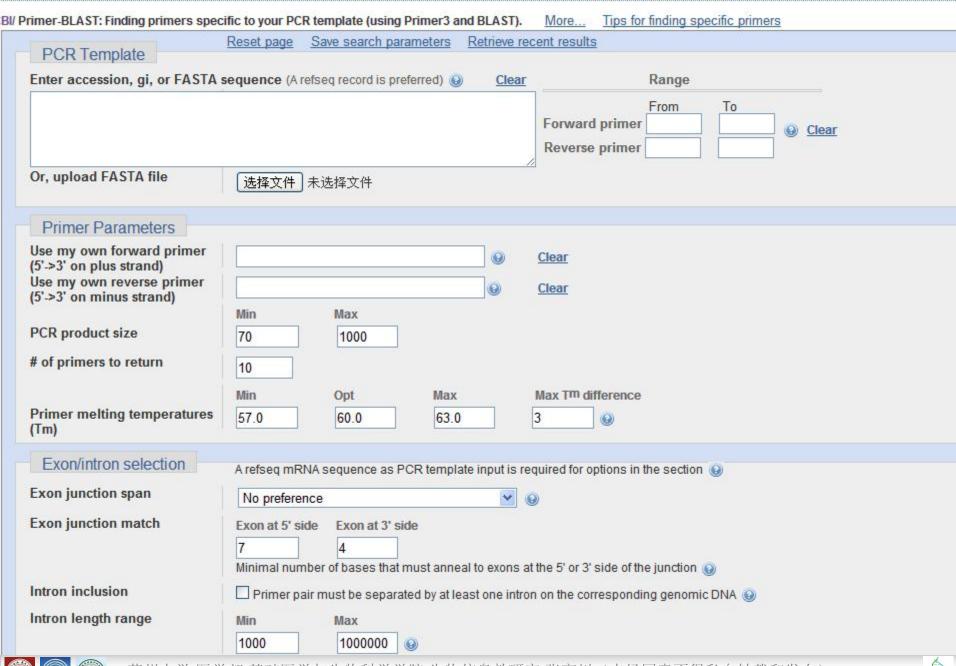








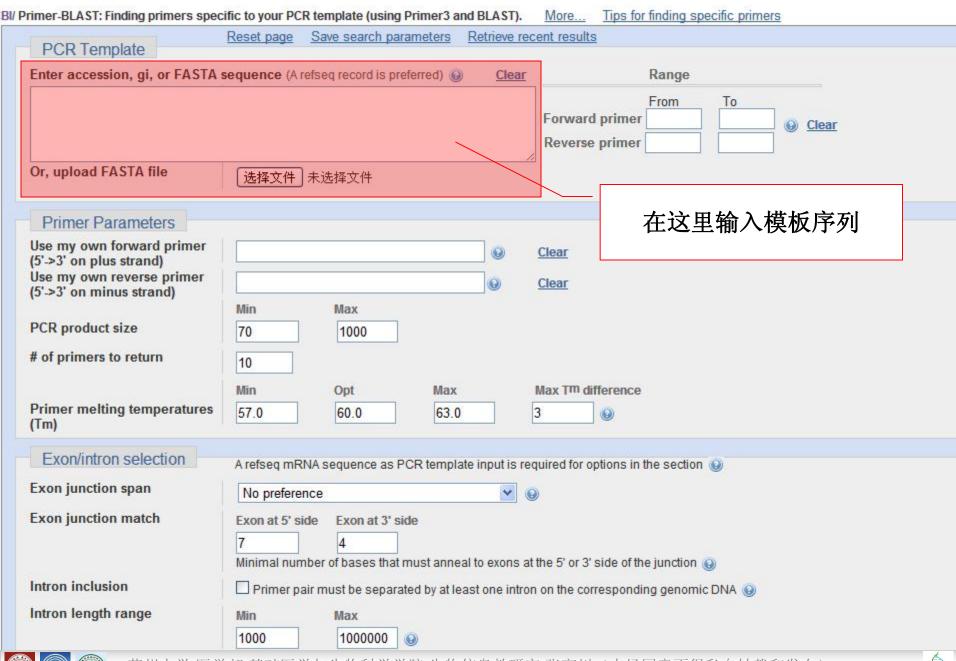








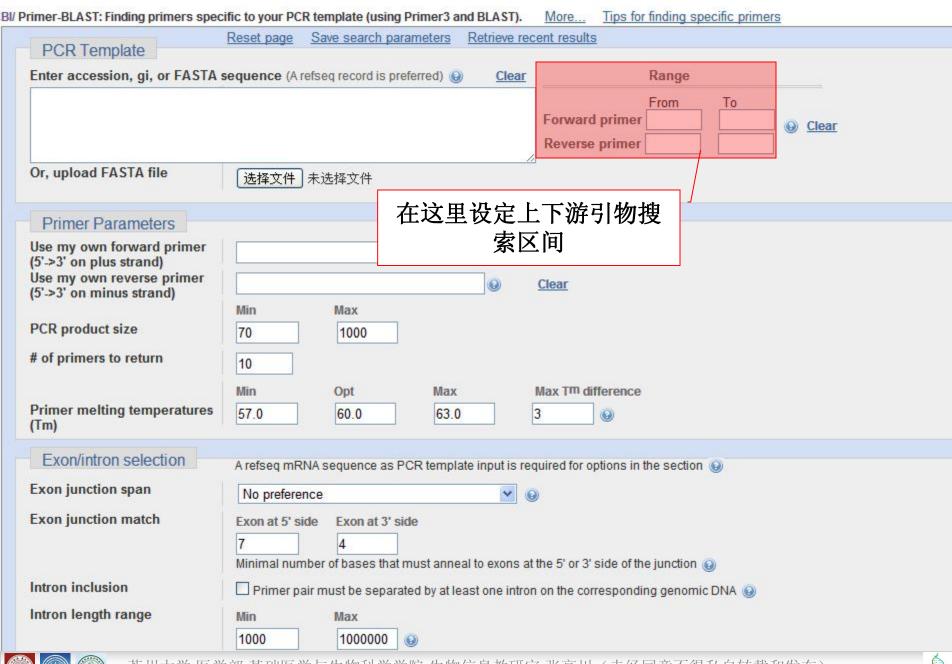








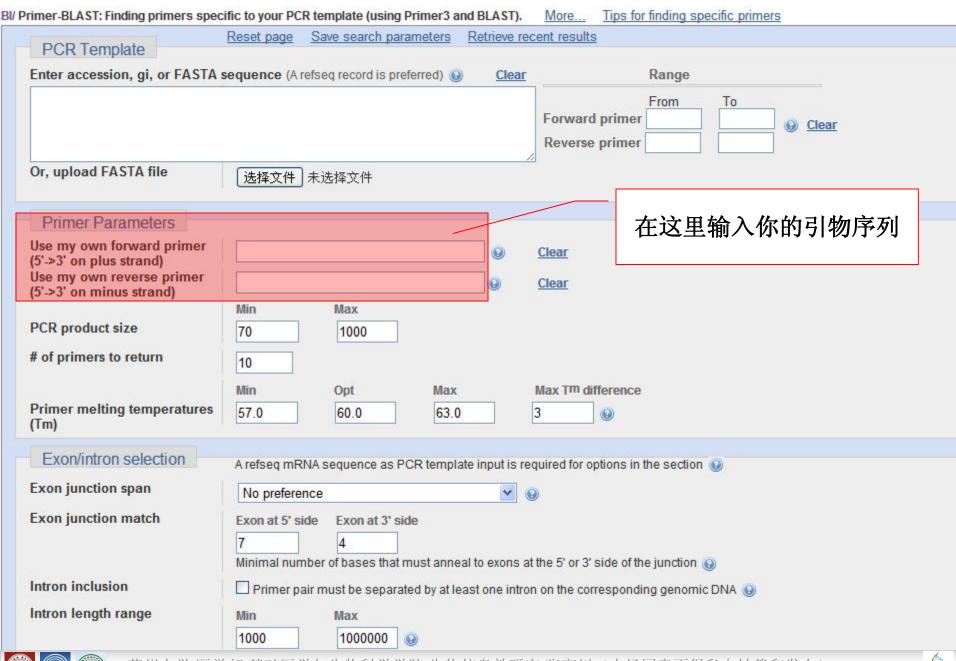








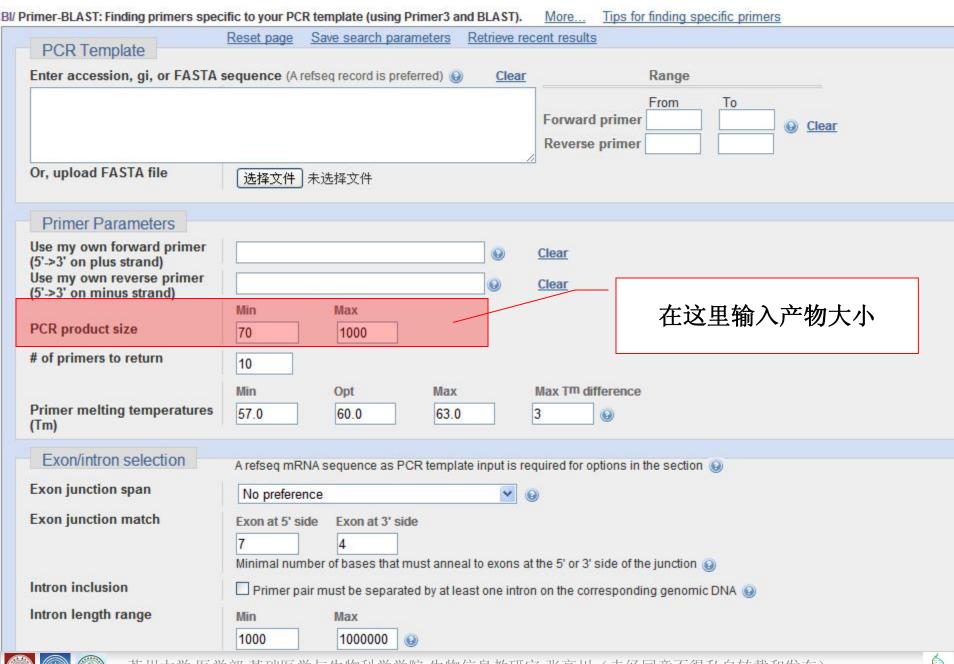






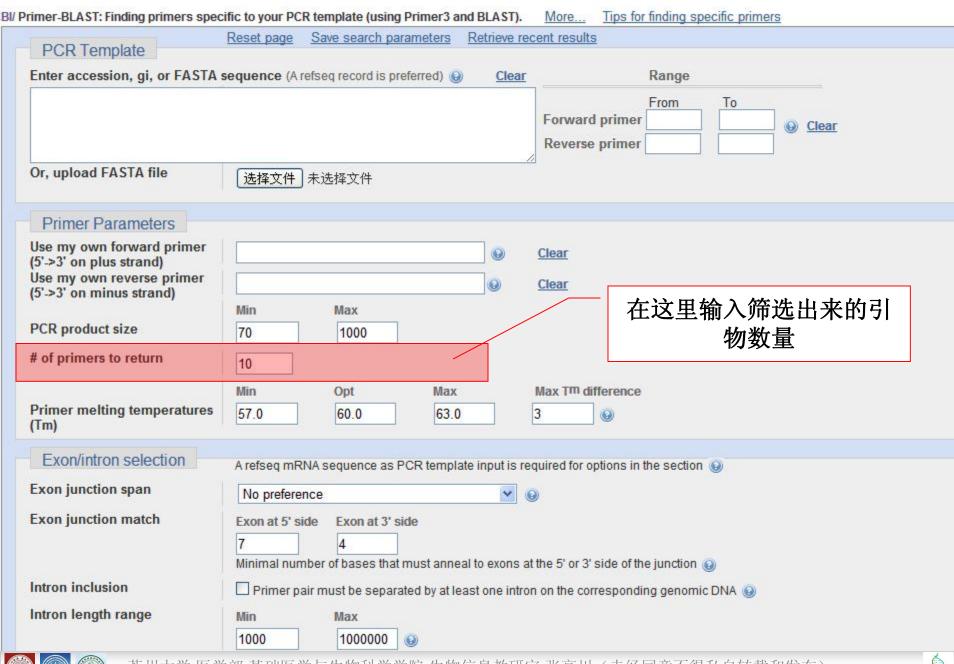








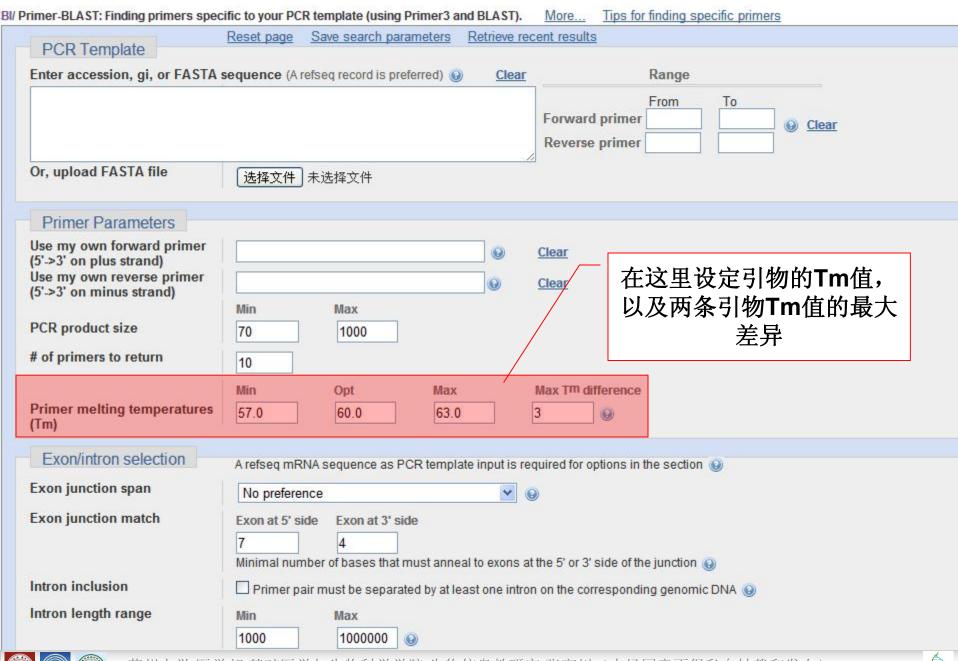








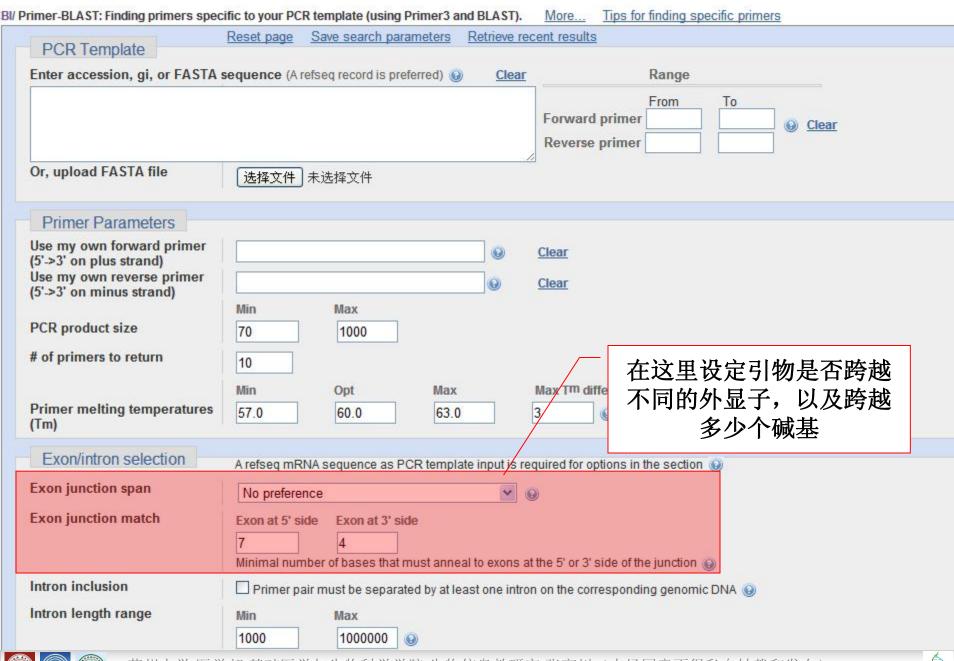








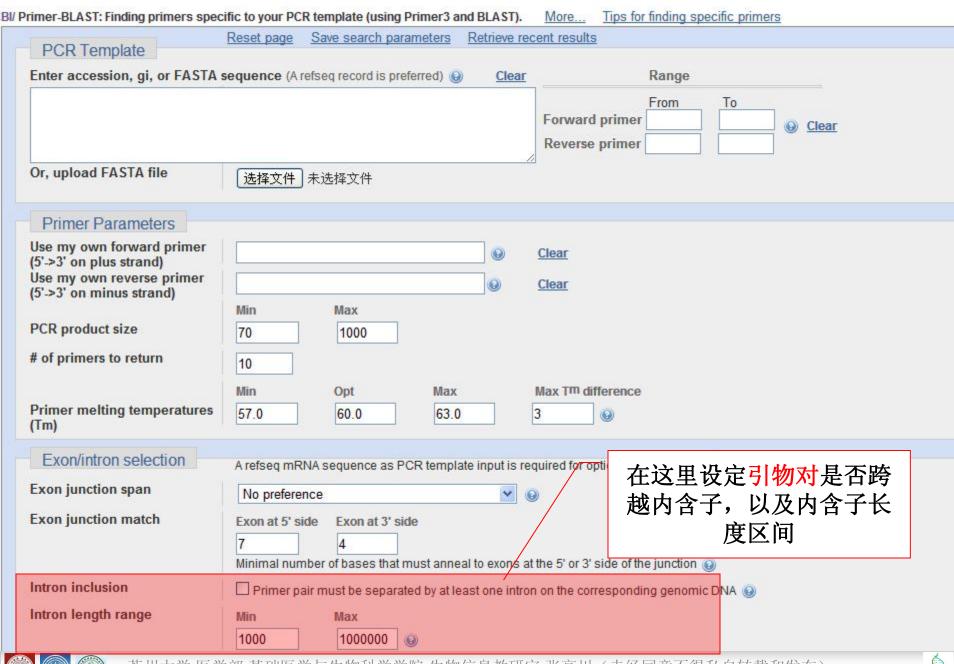








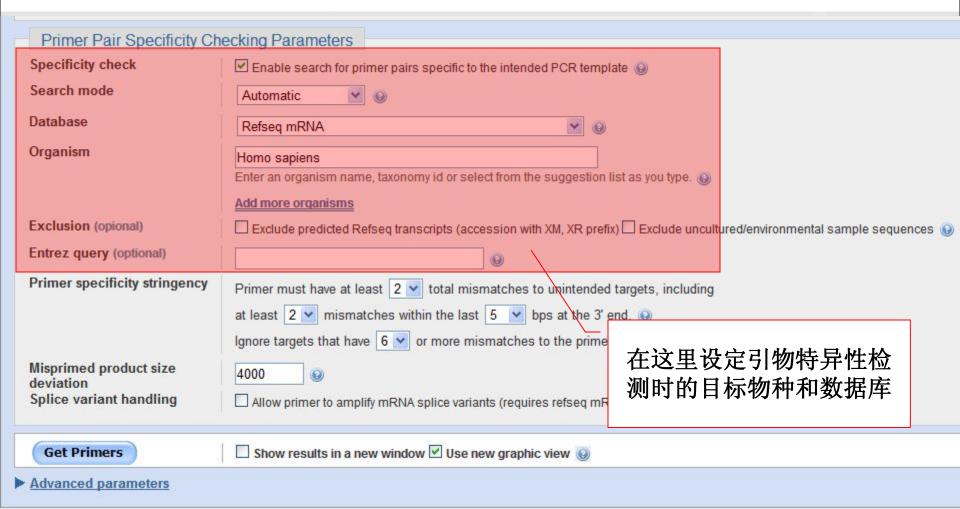












Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.



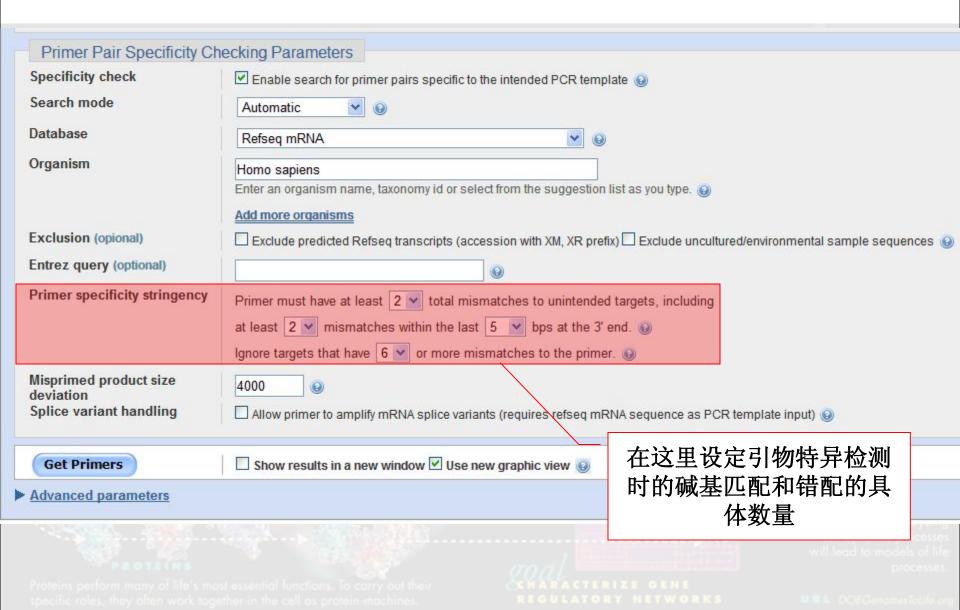
these dynamic processe will lead to models of lif processes

URL DOEGenomesToLife.org





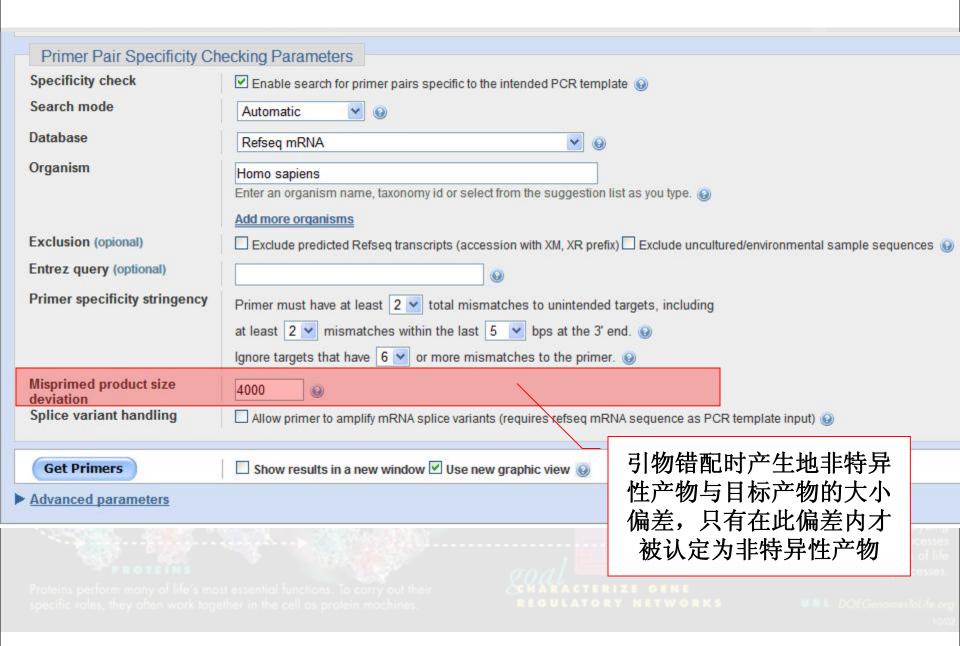








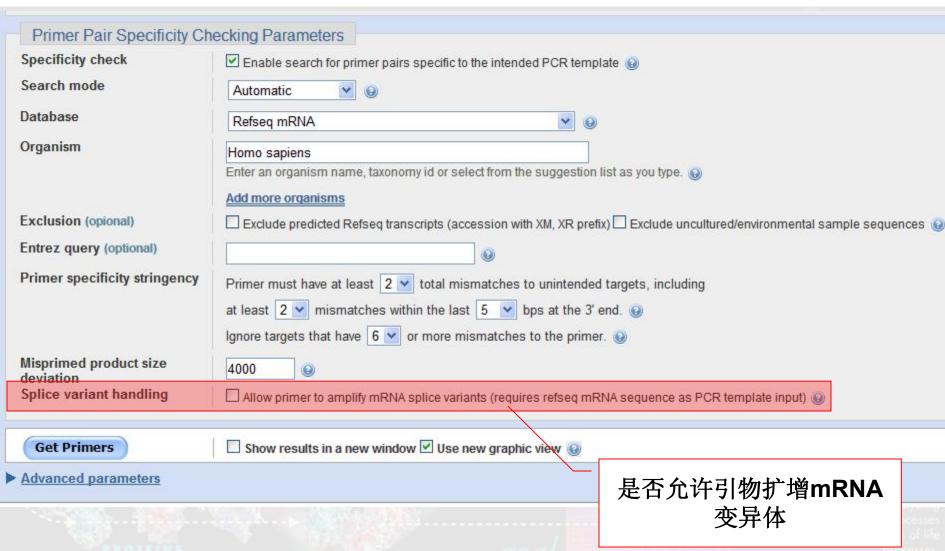




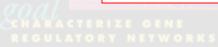








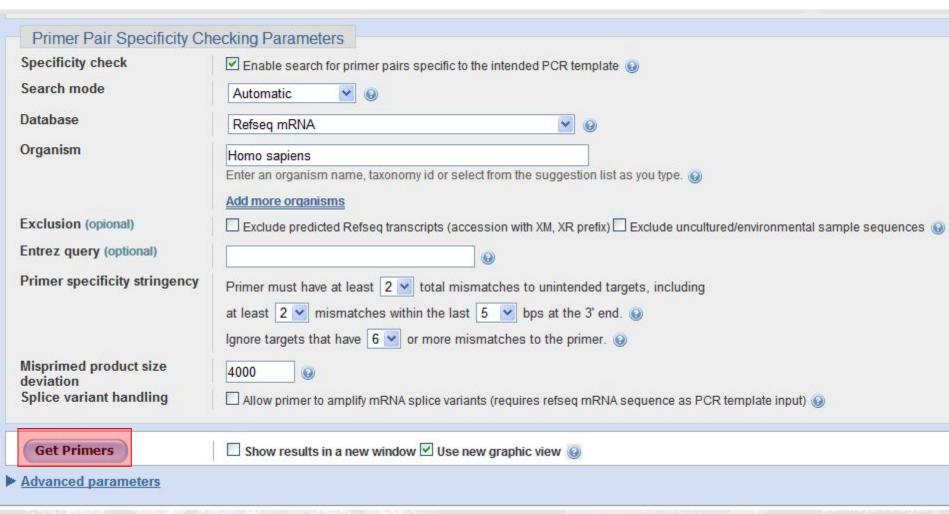
Proteins perform many of life's most essential functions. To carry out the specific roles, they often work together in the cell as protein machines.











Proteins perform many of life's most essential functions. To carry out thei specific roles, they often work together in the cell as protein machines.



these dynamic processes will lead to models of life processes

URL DOEGenomesToLife.org







Hide sidebar >>

Help

Save search Advanced

Location

Chromosome 17,

Chromosome 10.

Chromosome 5.

Chromosome 6,

NC 000006.12

NC 007116.6

(56187013..56198449)

(23582427...23594007)

(34236800..34246231)

(7668402...7687550,

NC 000017.11

complement)

NC 005109.4

▼ Top Organisms [Tree]

Homo sapiens (976) Mus musculus (36)

Rattus norvegicus (35)

Pan troglodytes (10)

All other taxa (996)

Bos taurus (11)

Find related data

Search

Display Settings: Tabular, 20 per page, Sorted by Relevance Send to: 

✓

Clear all

Gene

Show additional filters

Gene sources Genomic

Categories Alternatively spliced Annotated genes

Non-coding Protein-coding Pseudogene Sequence content

CCDS Ensembl RefSeq

Status ✓ Current only Chromosome locations

clear

Clear all

Select ...

Show additional filters

Results: 1 to 20 of 2064 Filters activated: Current only. Clear all to show 2067 items.

\_\_\_\_\_Tp53

ID: 24842

Name/Gene ID Description TP53 tumor protein ID: 7157

p53 [Homo sapiens (human)]

tumor protein p53 [Rattus norvegicus (Norway rat)]

tp53 tumor protein p53 [Danio rerio ID: 30590

(zebrafish)]

HMGA1 high mobility ID: 3159 group AT-hook 1 [Homo sapiens (human)]

TP53 tumor protein Chromosome 5, NC 006587.3 ID: 403869 p53 [Canis lupus

Filters: Manage Filters << First < Prev Page 1 of 104 Next > Last >>

brp53, drp53,

etID22686.5.

fb40d06, p53,

wu:fb40d06.

zgc:111919

RP11-

513115.2.

HMG-RA.

HMGIY,

HMGA1

P53

MIM Aliases 191170 BCC7, LFS1,

P53, TRP53 Trp53, p53

600701

Database: Select

More...

Search details

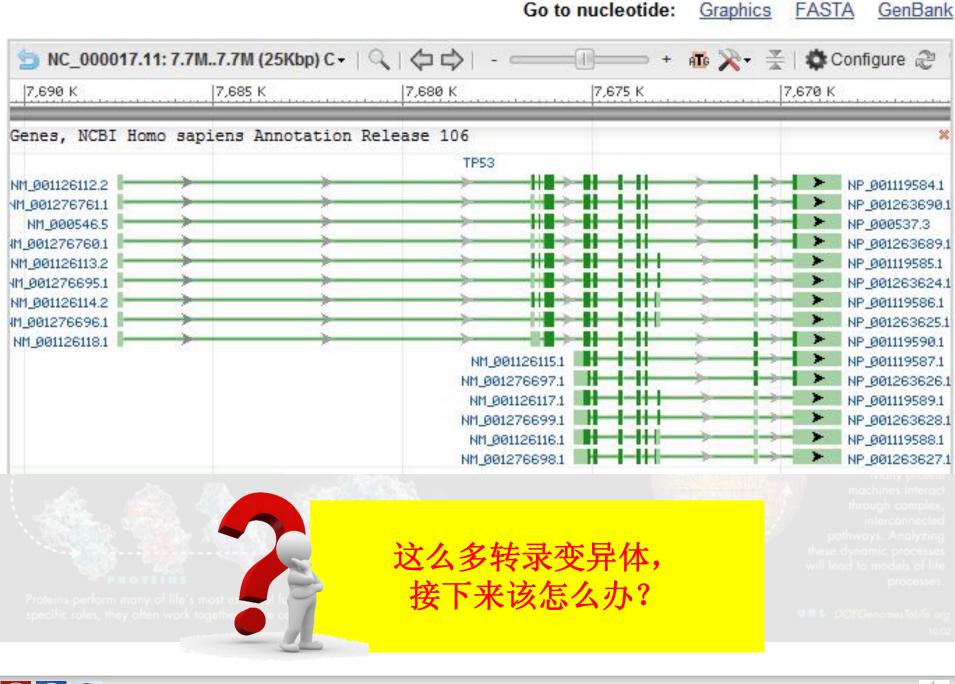
(TP53[All Fields] AND ("Homo sapiens" [Organism] OR human [All Fields])) AND alive[property]

Search

See more...













```
/organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /chromosome="17"
                                                                                                    Related information
                /map="17p13.1"
                                                                                                    Related Sequences
                1..2591
gene
                                                                                                    Annotated Genomic
                /gene="TP53"
                /gene_synonym="BCC7; LFS1; P53; TRP53"
                                                                                                    BioSystems
                /note="tumor protein p53"
                                                                                                    CCDS
                /db xref="GeneID:7157"
                /db xref="HGNC:HGNC:11998"
                                                                                                    Components (Core)
                /db xref="MIM:191170"
                                                                                                    Components (EST)
                1..174
exon
                /gene="TP53"
                                                                                                    Full text in PMC
                /gene_synonym="BCC7; LFS1; P53; TRP53"
                                                                                                    Gene
                /inference="alignment:Splign:1.39.8"
misc_feature
                95..97
                                                                                                    GeneView in dbSNP
                /gene="TP53"
                                                                                                    HomoloGene
                /gene_synonym="BCC7; LFS1; P53; TRP53"
                /note="upstream in-frame stop codon"
                                                                                                    Map Viewer
                175..276
exon
                                                                                                    MIMO
                /gene="TP53"
                /gene_synonym="BCC7; LFS1; P53; TRP53"
                                                                                                    Probe
                /inference="alignment:Splign:1.39.8"
                                                                                                    Protein
CDS
                203..1384
                /gene="TP53"
                                                                                                    PubMed
                /gene_synonym="BCC7; LFS1; P53; TRP53"
                                                                                                    PubMed (RefSeq)
                /note="isoform a is encoded by transcript variant 1; tumor
                protein 53; mutant tumor protein 53; cellular tumor
                                                                                                    PubMed (Weighted)
                antigen p53; phosphoprotein p53; transformation-related
                                                                                                    SNP
                protein 53; p53 tumor suppressor; antigen NY-CO-13"
                /codon_start=1
                                                                                                    Taxonomy
                /product="cellular tumor antigen p53 isoform a"
                                                                                                    UniGene
                /protein_id="NP_000537.3"
                /dh xref="GT:120407068"
                                                                                                    0000 0
```



FEATURES

source





UCSC Genome Browser

**IUCSC Genome Browse** 

Location/Qualifiers

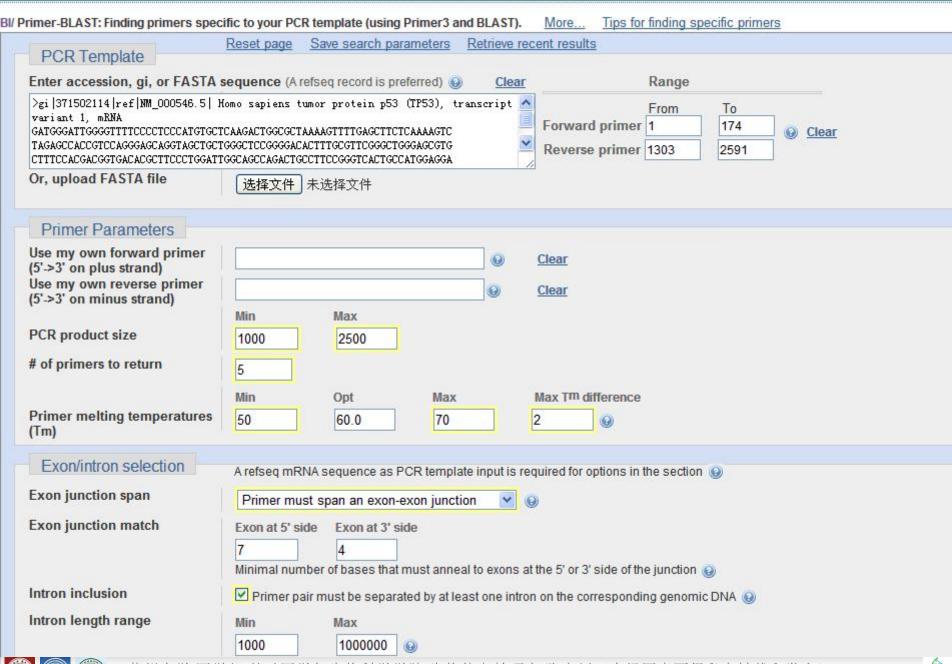
1..2591

>gi|371502114|ref|NM\_000546.5| Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA GATGGGATTGGGGTTTTCCCCTCCCATGTGCTCAAGACTGGCGCTAAAAGTTTTGAGCTTCTCAAAAGTC TAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGTTCGGGCTGGGAGCGTG GCCGCAGTCAGATCCTAGCGTCGAGCCCCCTCTGAGTCAGGAAACATTTTCAGACCTATGGAAACTACTT  $\tt CCTGAAAACAACGTTCTGTCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATA$ TTGAACAATGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATGCCAGAGGCTGCTCCCCCGT GGCCCCTGCACCAGCAGCTCCTACACCGGCGGCCCCTGCACCAGCCCCCTCCTGGCCCCTGTCATCTTCT AGTCTGTGACTTGCACGTACTCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGT GCAGCTGTGGGTTGATTCCACACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTACAAGCAGTCA  ${\tt CAGCACATGACGGAGGTTGTGAGGCGCTGCCCCACCATGAGCGCTGCTCAGATAGCGATGGTCTGGCCC}$  $\tt CTCCTCAGCATCTTATCCGAGTGGAAGGAAATTTGCGTGTGGAGTATTTGGATGACAGAAACACTTTTCG$ ACATAGTGTGGTGCCCTATGAGCCGCCTGAGGTTGGCTCTGACTGTACCACCATCCACTACAACTAC ATGTGTAACAGTTCCTGCATGGGCGGCATGAACCGGAGGCCCATCCTCACCATCATCACACTGGAAGACT CACAGAGGAAGAATCTCCGCAAGAAAGGGGAGCCTCACCACGAGCTGCCCCCAGGGAGCACTAAGCGA TTCAGATCCGTGGGCGTGAGCGCTTCGAGATGTTCCGAGAGCTGAATGAGGCCTTGGAACTCAAGGATGC ACCTCCGCCATAAAAAACTCATGTTCAAGACAGAAGGGCCTGACTCAGACTGACATTCTCCACTTCTTG AAGTTGGCCTGCACTGGTGTTTTGTTGTGGGGAGGAGGATGGGGAGTAGGACATACCAGCTTAGATTTTA AGGTTTTTACTGTGAGGGATGTTTGGGAGATGTAAGAAATGTTCTTGCAGTTAAGGGTTAGTTTACAATC AGCCACATTCTAGGTAGGGGCCCACTTCACCGTACTAACCAGGGAAGCTGTCCCTCACTGTTGAATTTTC TAATGAAATAATGTACATCTGGCCTTGAAACCACCTTTTATTACATGGGGTCTAGAACTTGACCCCCTTG AGGGTGCTTGTTCCCTCTCCCTGTTGGTCGGTGGGTTGGTAGTTTCTACAGTTTGGCCAGCTGGTTAGGTA GAGGGAGTTGTCAAGTCTCTGCTGGCCCAGCCAAACCCTGTCTGACAACCTCTTGGTGAACCTTAGTACC TAAAAGGAAATCTCACCCCATCCCACACCCTGGAGGATTTCATCTCTTGTATATGATGATCTGGATCCAC  $\tt CTGGGTCTCGCTTTGTTGCCCAGGCTGGAGTGGAGTGGCGTGATCTTGGCTTACTGCAGCCTTTGCCTCC$  $\tt CCGGCTCGAGCAGTCCTGCCTCAGCCTCCGGAGTAGCTGGGACCACAGGTTCATGCCACCATGGCCAGCC$ AACTTTTGCATGTTTTGTAGAGATGGGGTCTCACAGTGTTGCCCAGGCTGGTCTCAAACTCCTGGGCTCA GGCGATCCACCTGTCTCAGCCTCCCAGAGTGCTGGGATTACAATTGTGAGCCACCACGTCCAGCTGGAAG GGTCAACATCTTTTACATTCTGCAAGCACATCTGCATTTTCACCCCACCCTTCCCCTCCTTCTCCCTTTT TATATCCCATTTTTATATCGATCTCTTATTTTACAATAAAACTTTGCTGCCACCTGTGTGTCTGAGGGGGT G





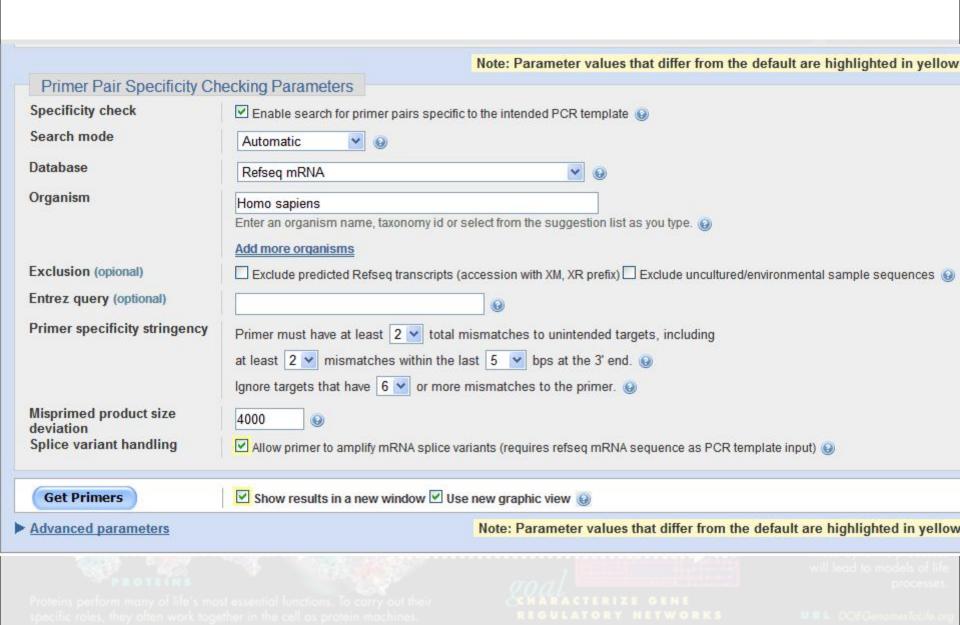














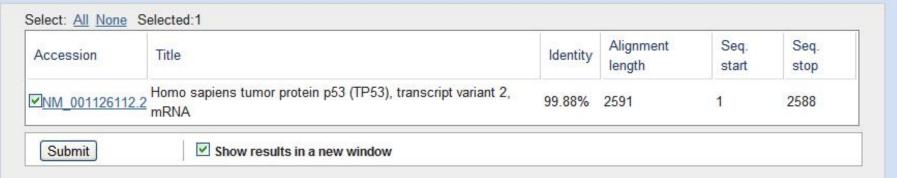


NCBI/ Primer-BLAST: Finding primers specific to your PCR template (using Primer3 and BLAST).

Input PCR template Range NM 000546.5 Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA 1 - 2591

For better specificity checking, we have substituted the PCR template with the GenBank refseq record NM\_000546.5 which is identical to your input template

Your PCR template is highly similar to the following sequence(s) from the search database. To increase the chance of finding specific primers, please review the list below and select all sequences (within the given sequence ranges) that are intended or allowed targets.



PROTEINS

Proteins perform many of life's most essential functions. To carry out the specific roles, they often work together in the cell as protein machines.

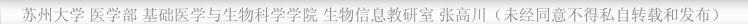
CHARACTERIZE GENE REGULATORY NETWORKS through complex interact through complex interconnected pathways. Analyzing these dynamic processes will lead to models of life processes

URL DOEGenomesTolife.org











NCBI/ Primer-BLAST: results: Job id=BR7uwCnKgO6\_3LvS2vyJrMHRu73Szqa4

A No primers satisfying exon/intron selection parameters were found. Try to loosen the selection criteria and/or increase value for "Max primer

Warnings:

**Input PCR template** NM 000546.5 Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA

1 - 2591Range

▶ Search Summary Other reports

For better specificity checking, we have substituted the PCR template with the GenBank refseq record NM 000546.5 which is identical to your input template

No primers were found...see explanation below: Primer3 info:

Left primer: considered 1705, GC content failed 5, low tm 127, high tm 147, ok 1426.

Right primer: considered 10428, too many Ns 343 (This could be due to low complexity and/or repeat filtering. Try search with filtering off), \$60 Primer pairs: considered 10943124, no overlap of required point 10943124, ok 0

Some or all specific primers may have been excluded due to one or more of following restrictions: user specified primer ranges. Try search w

No primers were found...see explanation below: Primer3 info:

Left primer: considered 1705, GC content failed 5, low tm 127, high tm 147, ok 1426.

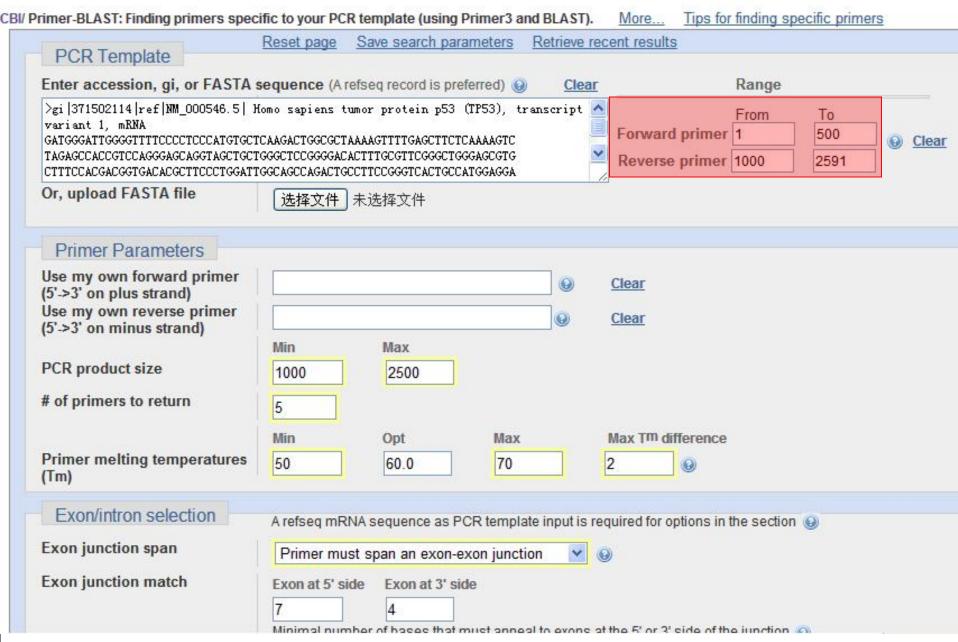
Right primer: considered 10428, too many Ns 343 (This could be due to low complexity and/or repeat filtering. Try search with filtering off), GC content failed 157, low tm 2181, high tm 53, long poly-x seq 20, ok 7674.

Primer pairs: considered 10943124, no overlap of required point 10943124, ok 0. Some or all specific primers may have been excluded due to one or more of following restrictions: user specified primer ranges. Try search with less restrictions.















► NCBI/ Primer-BLAST: results: Job id=nld3DfMHWiNhm2WVBLtX6x-WZfoMiXj more.

Input PCR template NM 000546.5 Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA

Range 1 - 2591

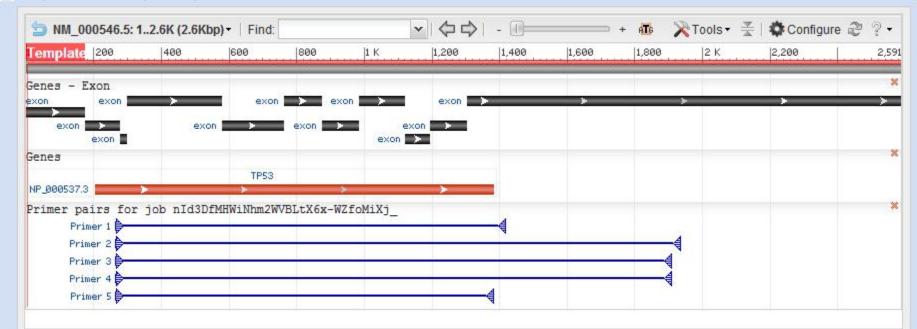
Specificity of primers Primer pairs are specific to input template as no other targets were found in selected database: Refseq mRNA

(Organism limited to Homo sapiens)

Other reports > Search Summary

① For better specificity checking, we have substituted the PCR template with the GenBank refseq record NM\_000546.5 which is identical to your input template

#### Graphical view of primer pairs



#### Detailed primer reports









|                           | Sequenc   | e (5'->3' | )                          | Template<br>strand | Length     | Start S   | top Tm    | GC%     | Self<br>complementarity | Self 3'<br>complementarity |
|---------------------------|---|-----------|----------------------------|--------------------|------------|-----------|-----------|---------|-------------------------|----------------------------|
| Forward primer            | ACCTAT  | GGAAAC    | TACTTCCTGAAAA              | Plus               | 25         | 264 2     | 88 57.9   | 1 36.00 | 4.00                    | 1.00                       |
| Reverse<br>primer         | GCTGTC  | AGTGG     | GGAACAAGA                  | Minus              | 20         | 14151     | 396 59.89 | 955.00  | 5.00                    | 0.00                       |
| Product<br>length         | 1152  |           |                            |                    |            |           |           |         |                         |                            |
| Exon<br>junction<br>Total | 276/277   | (forward  | primer) on template        | NM_00054           | 6.5        |           |           |         |                         |                            |
| intron<br>size            | 5687 (be  | tween po  | s. 7187395 and 71          | 80703 on <u>N</u>  | T_01071    | 8.17)     |           |         |                         |                            |
|                           | on intend   |           | et<br>iens tumor protein i | n53 (TP53)         | transcri   | nt varia  | nt 2 mRN  | IA      |                         |                            |
|                           |   | 1000      |                            | ,,,                | transcrip  | or varia  |           |         |                         |                            |
| product                   | 125 - 1 |           |                            |                    |            |           |           |         |                         |                            |
| Forward<br>Template       | -   |           | ACCTATGGAAACT.             |                    |            | 25<br>285 |           |         |                         |                            |
| Reverse                   | primer  | 1         | GCTGTCAGTGGG               | GAACAAGA           | 20         |           |           |         |                         |                            |
| Template                  | -   |           |                            |                    |            |           |           |         |                         |                            |
| >NM_0005                  | 46.5 Homo   | sapiens   | tumor protein p53          | (TP53), tran       | nscript va | ariant 1  | , mRNA    |         |                         |                            |
| product                   | length  | = 1152    |                            |                    |            |           |           |         |                         |                            |
| Forward                   | primer  | 1         | ACCTATGGAAACT.             | ACTTCCTG           | AAAA:      | 25        |           |         |                         |                            |
| Template                  | ∋   | 264       |                            |                    |            | 288       |           |         |                         |                            |
| Reverse                   | primer  | 1         | GCTGTCAGTGGG               | GAACAAGA           | 20         |           |           |         |                         |                            |
| Template                  |   |           |                            |                    |            |           |           |         |                         |                            |
| Products                  | on allow  | ed trans  | cript variants             |                    |            |           |           |         |                         |                            |
| >NM_0011                  | 26114.2 H   | omo sap   | iens tumor protein         | p53 (TP53),        | transcrip  | ot varia  | nt 3, mRN | IA      |                         |                            |
| product                   |   |           |                            |                    |            |           |           |         |                         |                            |
| Forward                   |   |           | ACCTATGGAAACT.             |                    |            |           |           |         |                         |                            |
| Template                  | 3   | 264       |                            |                    |            | 288       |           |         |                         |                            |
| Reverse<br>Template       |   |           | GCTGTCAGTGGG               |                    |            |           |           |         |                         |                            |
| remprace                  | =   | 1346      |                            |                    | 1323       |           |           |         |                         |                            |
| > <u>NM_0011</u>          | 26113.2 H   | omo sap   | iens tumor protein į       | p53 (TP53),        | transcrip  | ot varia  | nt 4, mRN | IA      |                         |                            |
| product                   |   |           |                            |                    |            |           |           |         |                         |                            |
| Forward                   |   |           | ACCTATGGAAACT.             |                    |            |           |           |         |                         |                            |
| Template                  | 3   | 264       |                            |                    |            | 288       |           |         |                         |                            |
|                           |   |           | GCTGTCAGTGGG               |                    |            |           |           |         |                         |                            |
| Template                  | <b>=</b>  | 1475      |                            |                    | 1456       |           |           |         |                         |                            |



Primer pair 2











### Gen Script Real-time PCR (TaqMan) Primer Design https://www.genscript.com/ssl-bin/app/primer

\*\* This online tool helps you to design primers and probes for your Real-time PCR (TaqMan) experiments. You can customize the potential PCR amplicon's size range, Tm (melting temperature) for the primers and probes, as well as the organism. You can also decide how many Primer/Probe sets you want the tool to return to you. We recommend you using the GenBank Accession to input your target sequence. However, you can choose to input the sequences manually in raw format.

Number of Primer/Probe Sets to Return (Limit 20 Sets) 3

PCR Amplicon Size Range 50-150

Primer Tm Minimum 58 Optimum 59 Maximum 60 °C

Probe Tm Minimum 68 Optimum 69 Maximum 70 °C

Organism: Human Pick Primer/Probe Crossing Exon Junction

When Pick Primer/Probe Crossing Exon Junction is selected, the exon regions must be defined. If only raw sequence is provided, the sequence will be mapped on the genome (human, mouse or rat at present) sequences to locate the exon boundaries. Besides, there are other two ways to specify exon regions:

- The best way is to fill in GenBank Accession with the accession number (e.g. NM\_145027, please note the accession number is
  case-sensitive and should not contain version suffix like '.1' or '.2'). The sequence will be fetched from NCBI data center, and the exon
  regions are already defined in sequence annotations.
- . Manually specify the exon junctions with "in the input sequence (e.g. 'ACGCGCG:CGTACG')

#### Target Nucleotide Sequences:

- GenBank Accession:
- · or Paste in the DNA Sequence in raw format:







## 拓展阅读和实践练习

查阅引物设计方面的文献资料

尝试练习各种引物设计软件的使用

on how and when

PROTEIN MACHINES COMPLEX BIOLOGICAL

roteins perform many of life's most essential functions. To carry out the pecific roles, they often work together in the cell as protein machines.

OAL HARACTERIZE GENE

URL DOEGenomes Tolife.c

♥ R L DOEGenomesTolife.or







### 第一章 绪论

生物信息学 基本概念

生命科学研究的三重层次 【个体-细胞-分析分子】

人类基因组计划

生物信息学的研究内容和方向

生物信息学在生命科学研究中的应用 【复杂疾病的早期诊断、药物研发、遗传育种、进化分析】









## 第二章 常用数据库介绍

NCBI-Gebank 【综合】 Entrez 【PubMed,Gene,OMIM】

EMBL-EBI【综合】

### DDBJ 【综合】

on how and when to build proteins

PROTEIN MACHINES ExPASy 【蛋白质】

BIOLOGICAL

COAL CHARACTERIZE GENE

machines interactions interactions interconnected pathways. Analyzing these dynamic processe will lead to models of life processes

URL DOEGenomesTolife.org

Proteins perform many of life's most essential functions. To ca







## 第四章 转录组学与调控网络分析

转录组(transcriptome)和转录组学(transcriptomics) 基本概念

传统生化与分子生物学方法

高通量技术和方法

DNA microarray (DNA chip or gene chip)

RNA-Seq

Serial analysis of gene expression (SAGE)

**EST/cDNA** library







# 第四章 转录组学与调控网络分析

Genbank之GEO数据库及其在线分析工具

Genbank之EST和UniGene数据库, 以及在线DDD分析工具

基因功能注释【Gene Ontology, KEGG=>DAVID】

蛋白相互作用【STRING】

引物设计【Primer-Blast】









# Thanks for your attention!





