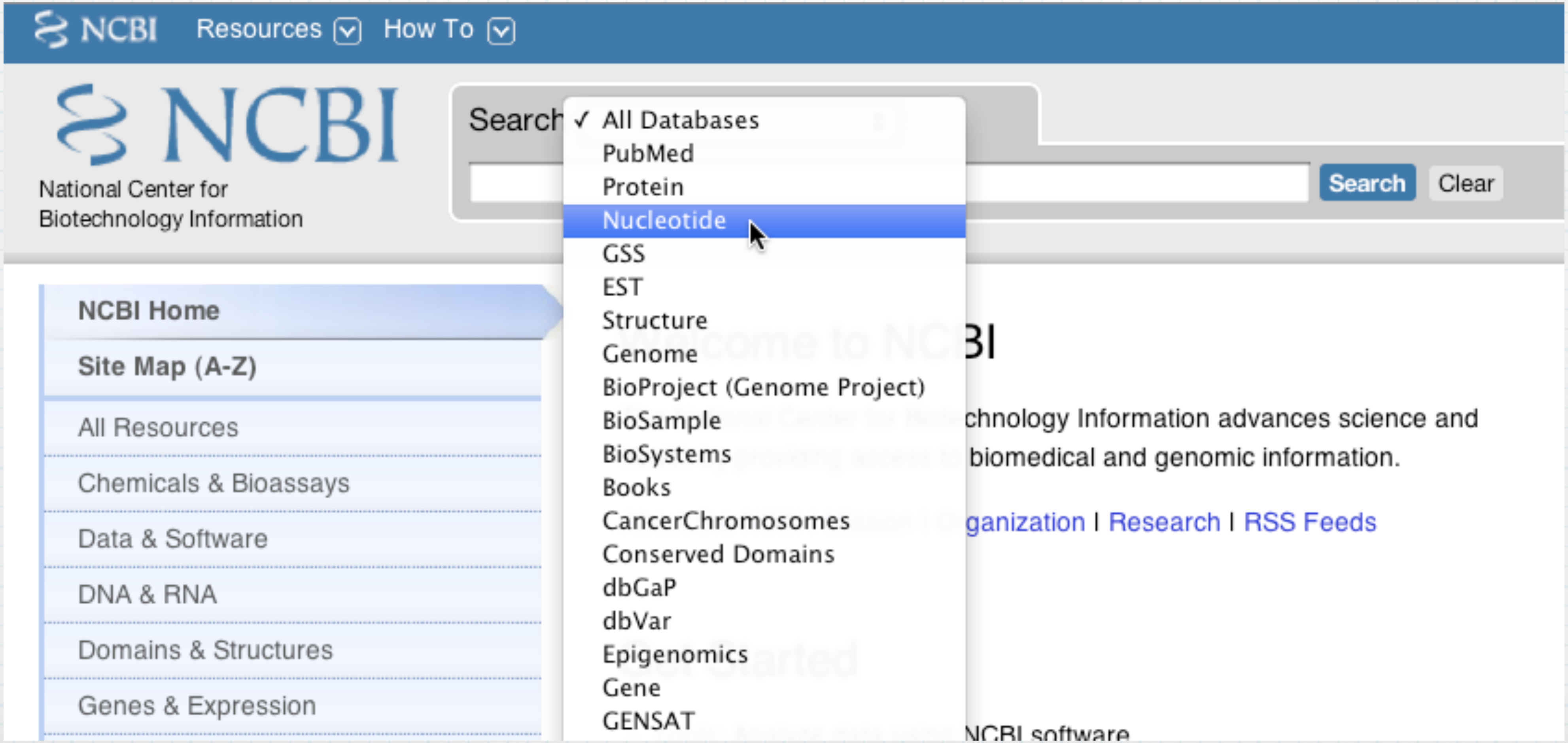
遺伝子のデータベース
基礎の基礎

- ■ 遺伝子名などキーワードで探す
- ■ 遺伝子のさまざまなIDとは？
- ■ 塩基配列から遺伝子を探す
- ■ 統合遺伝子検索GGRNAの紹介

遺伝子をさがす　基礎

- NCBI Entrez　http://www.ncbi.nlm.nih.gov/

（または NCBI でググる）



絞り込み

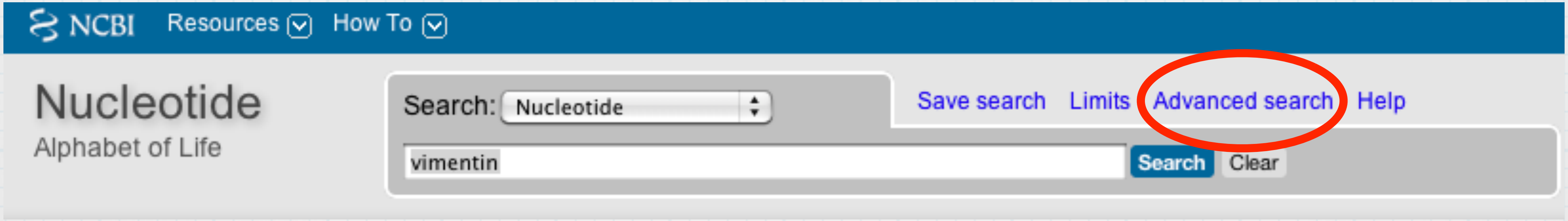
- 検索窓にキーワードを追加していく

  … AND “Homo sapiens”[Organism]

  … AND Vimentin[Gene Name]

  … AND patent[Title]

- または、Advanced searchに行く



Search: Nucleotide

vimentin

遺伝子の ID とは？

- Accession Number
- RefSeq ID
- Gene ID
- Symbol (遺伝子名)

Accession Number

- GenBank/EMBL/DDBJ 国際塩基配列データベースに登録された塩基配列のID
- A12345 や AB123456 の形をしている
- A12345.1 のようにバージョンを表示。UTRが延長されたりエラーが修正されてA12345.2 のようにアップデートされる
- GenBankのAccessionと呼ばれることも...

×

RefSeq ID

- 三大データバンクの配列を元にtranscriptごとに1個登録 → RefSeq データベース（遺伝子の百科事典のようなもの）
- 選択的スプライシングで生じるvariantには別々のIDが付与されている
- NM_012345.6 の形式をしている。広義には（実用上は）Accession番号の一種

Symbol, Gene ID

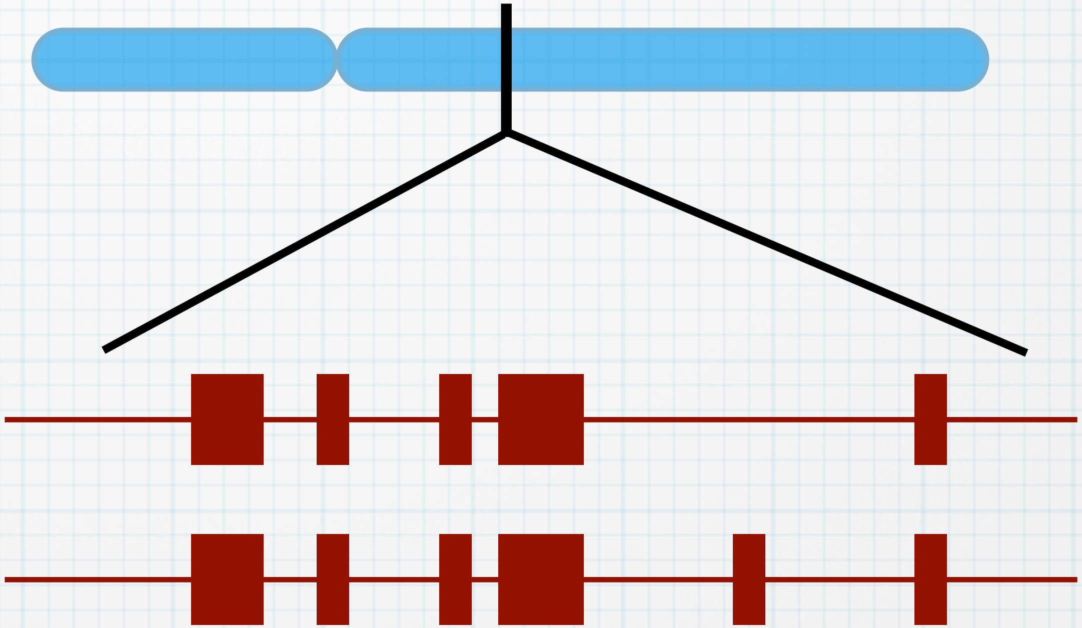
- 遺伝子ごとに付与される遺伝子名と番号

慣用名	Symbol	Gene ID
ヒトcadherin	CDH1	999
マウスcadherin	Cdh1	12550
ラットcadherin	Cdh1	83502

- Symbolは慣用名と一致しないこともあり（ヒトp53 → TP53）種でダブる可能性も
- Gene ID は生物種と遺伝子を特定できる

それぞれの関係

ヒト Chr22 (q11)



RefSeq ID:

NM_001190326

NM_022720

transcriptごと

（塩基配列ごと）

Symbol: DGCR8

Gene ID: 54487

遺伝子（locus）ごと

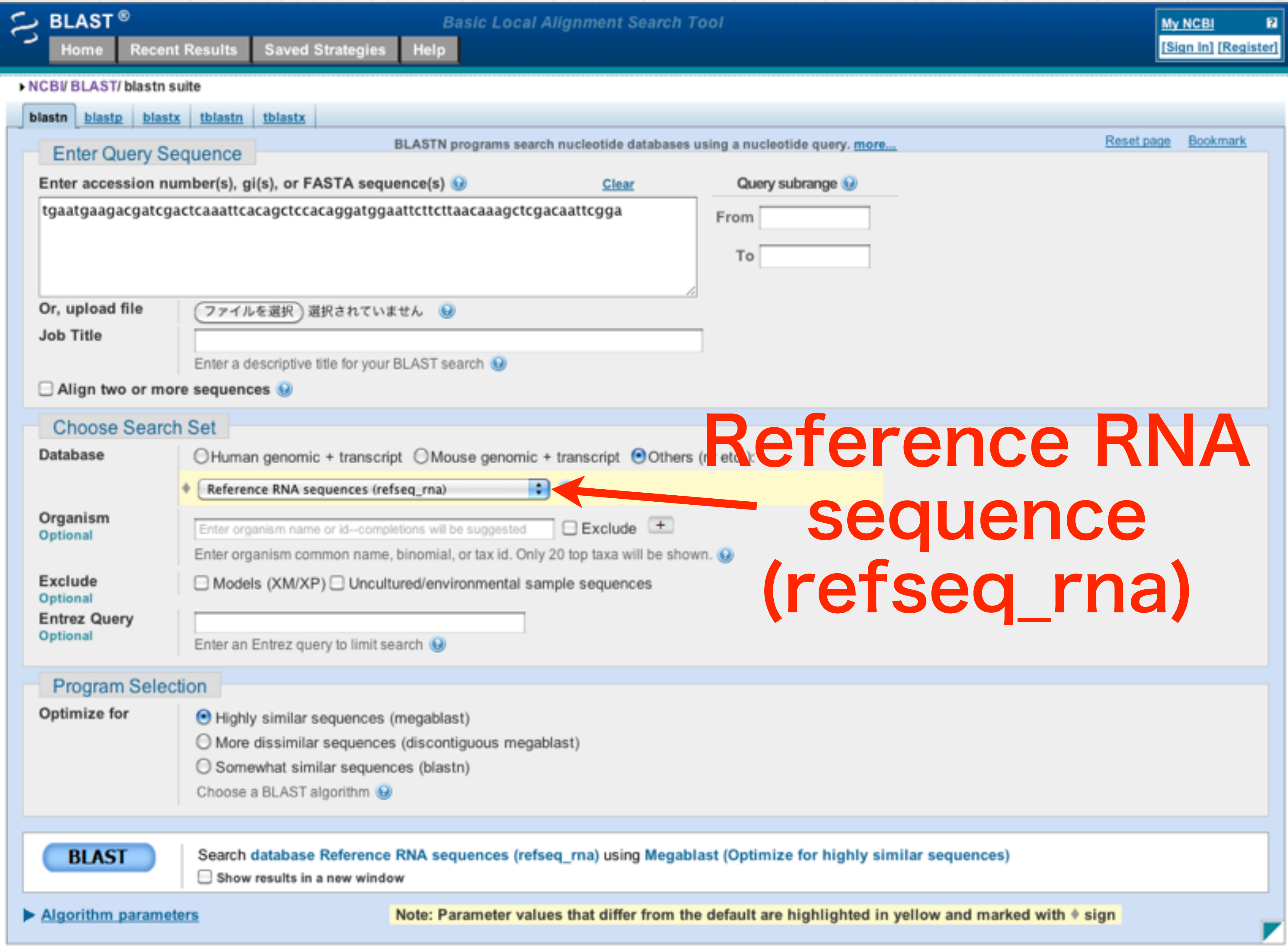
配列から遺伝子をさがす

- NCBI BLAST

  http://www.ncbi.nlm.nih.gov/BLAST/
  （または BLAST でググる）

- UCSC BLAT

  http://genome.ucsc.edu/ → BLATへ
  （または BLAT でググる）



Enter Query Sequence

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

```
tgaatgaagacgatcgactcaaattcacagctccacaggatggaattcttcttaacaaagctcgacaattcgga
```

Query subrange

From

To

Or, upload file （ファイルを選択） 選択されていません

Job Title

Enter a descriptive title for your BLAST search

□ Align two or more sequences

Choose Search Set

Database ○ Human genomic + transcript ○ Mouse genomic + transcript ● Others (r etc.):

♦ Reference RNA sequences (refseq_rna)

Organism
Optional

Enter organism name or id--completions will be suggested □ Exclude +

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude
Optional

□ Models (XM/XP) □ Uncultured/environmental sample sequences

Entrez Query
Optional

Enter an Entrez query to limit search

Program Selection

Optimize for

● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST

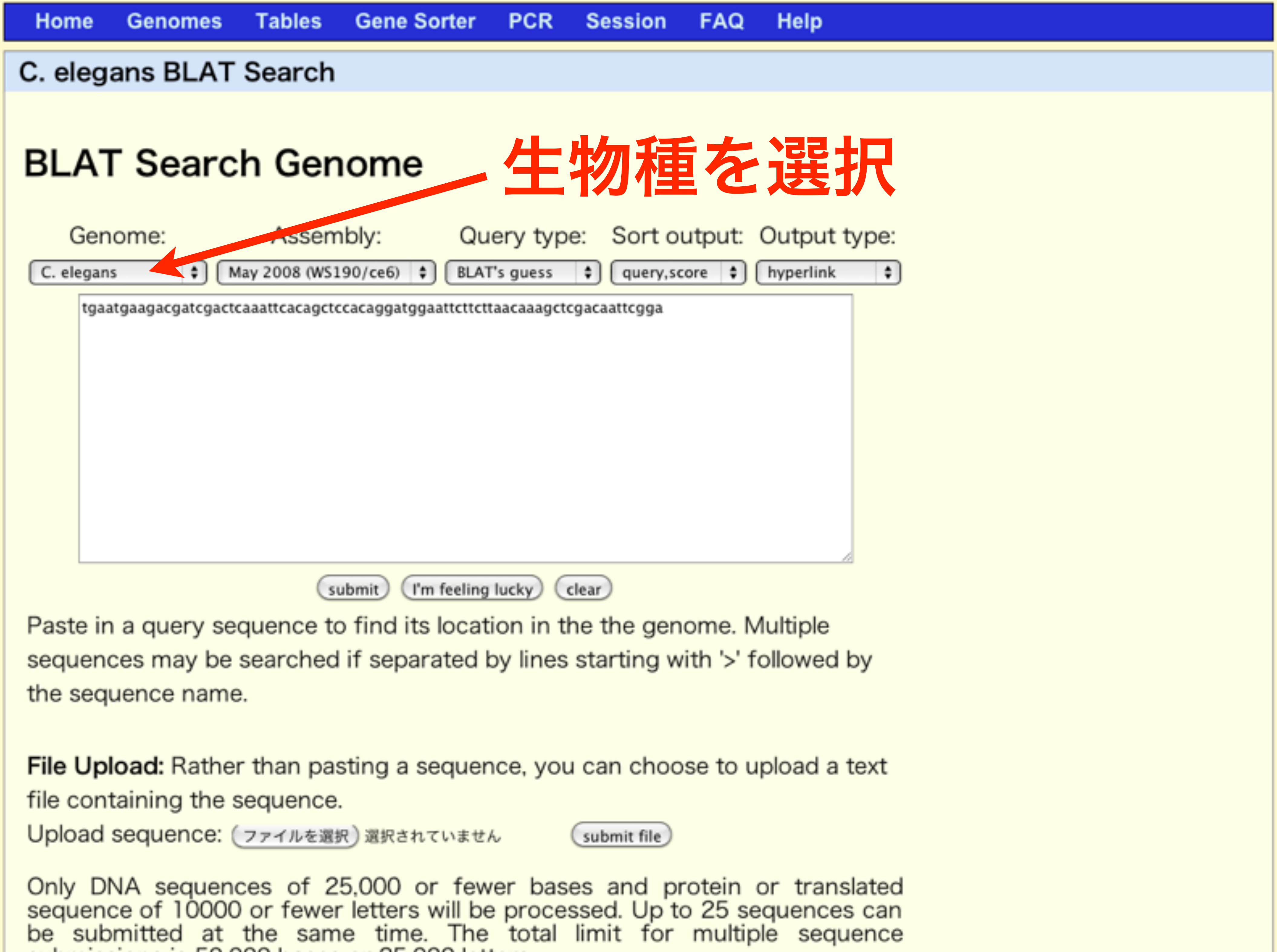
Search database Reference RNA sequences (refseq_rna) using Megablast (Optimize for highly similar sequences)

□ Show results in a new window

Algorithm parameters

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

Reference RNA sequence (refseq_rna)



C. elegans BLAT Search

BLAT Search Genome

生物種を選択

Genome: C. elegans

Assembly: May 2008 (WS190/ce6)

Query type: BLAT's guess

Sort output: query,score

Output type: hyperlink

```
tgaatgaagacgatcgactcaaattcacagctccacaggatggaattcttcttaacaaagctcgacaattcgga
```

Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.

File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.

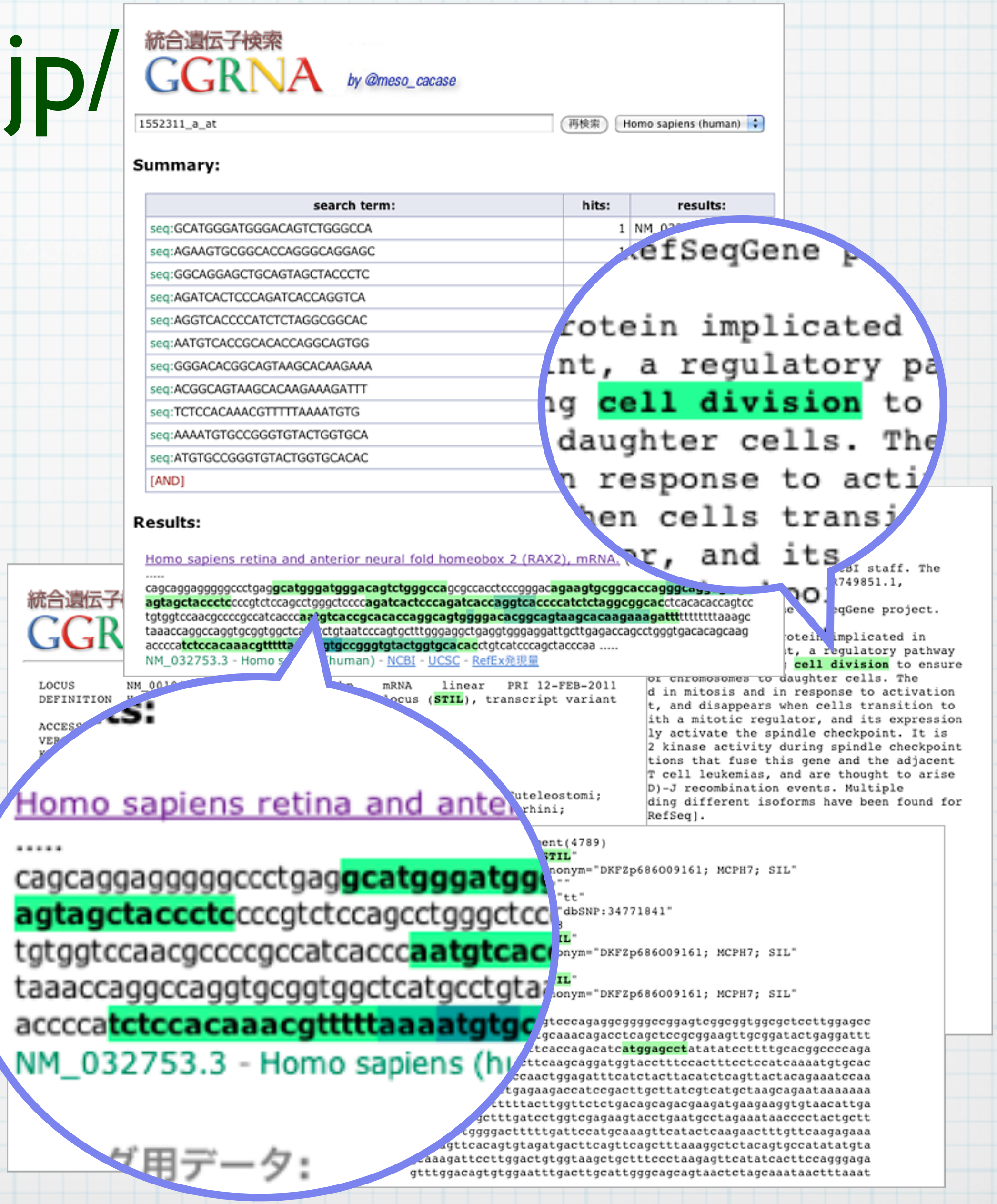
Upload sequence: ファイルを選択 選択されていません

Only DNA sequences of 25,000 or fewer bases and protein or translated sequence of 10000 or fewer letters will be processed. Up to 25 sequences can be submitted at the same time. The total limit for multiple sequence

統合遺伝子検索GGRNA

http://GGRNA.dbcls.jp/

- RefSeqを全文検索
- 塩基配列も簡単検索
- ヒト、マウス、ラット、ニワトリ、ツメガエル、ゼブラ、ホヤ、ハエ、線虫、シロイヌナズナ、イネ、出芽酵母、分裂酵母



Summary:

search term:	hits:	results:
seq:GCATGGGATGGGACAGTCTGGGCCA	1	NM_03
seq:AGAAGTGCGGCACCAGGGCAGGAGC		
seq:GGCAGGAGCTGCAGTAGCTACCCTC		
seq:AGATCACTCCCAGATCACCAGGTCA		
seq:AGGTCACCCCATCTCTAGGCGGCAC		
seq:AATGTCACCGCACACCAGGCAGTGG		
seq:GGGACACGGCAGTAAGCACAAGAAA		
seq:ACGGCAGTAAGCACAAGAAAGATTT		
seq:TCTCCACAAACGTTTTTAAAATGTG		
seq:AAAATGTGCCGGGTGTACTGGTGCA		
seq:ATGTGCCGGGTGTACTGGTGCACAC		
[AND]		

Results:

Homo sapiens retina and anterior neural fold homeobox 2 (RAX2), mRNA.

```
.....
cagcaggaggggccctgaggcatgggatgggacagtctgggccagcgccacctcccgggacagaagtgcggcaccagggcagg
agtagctaccctccccgtctccagcctgggctccccagatcactcccagatcaccaggtcaccccatctctaggcggcacctcacacaccagtcc
tgtggtccaacgccccgccatcacccaatgtcaccgcacaccaggcagtggggacacggcagtaagcacaagaaagatttttttttttaaagc
taaaccaggccaggtgcggtggctcatgcctgtaatcccagtgctttgggaggctgaggtgggaggattgcttgagaccagcctgggtgacacagcaag
accccatctccacaaacgtttttaaaatgtgccgggtgtactggtgcacacctgtcatcccagctacccaa .....
```

NM_032753.3 - Homo sapiens (human) - NCBI - UCSC - RefEx発現量

実習：簡単な検索例

- 遺伝子名、フレーズ、各種IDで検索

  例）claudin, “RNA interference”,
  NM_001518, 10579,
  VIM(ヒット多し)→ symbol:VIM

- プローブのIDでさがしてみる

  例）A_23_P101434