

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

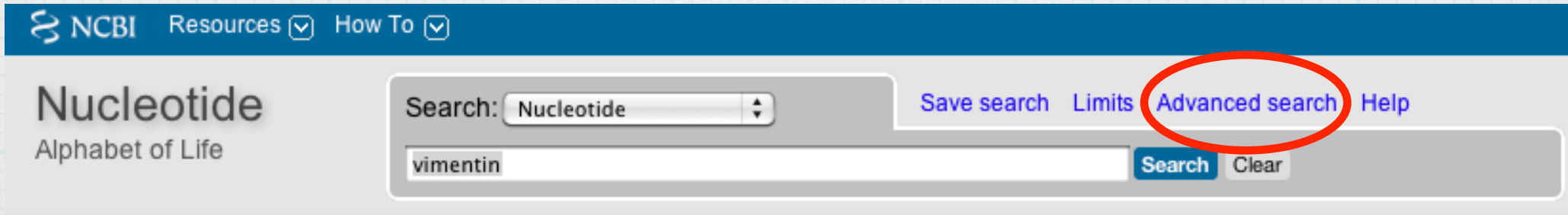
遺伝子をさがす 基礎

- NCBI Entrez <http://www.ncbi.nlm.nih.gov/>
(または NCBI でググる)

The screenshot displays the NCBI Entrez homepage. At the top, there is a navigation bar with the NCBI logo, "Resources" (with a dropdown arrow), and "How To" (with a dropdown arrow). Below this is the main header area featuring the NCBI logo and the text "National Center for Biotechnology Information". A search bar is present with a "Search" button and a "Clear" button. A dropdown menu is open over the search bar, listing various database options: "All Databases" (checked), "PubMed", "Protein", "Nucleotide" (highlighted with a mouse cursor), "GSS", "EST", "Structure", "Genome", "BioProject (Genome Project)", "BioSample", "BioSystems", "Books", "CancerChromosomes", "Conserved Domains", "dbGaP", "dbVar", "Epigenomics", "Gene", and "GENSAT". On the left side, there is a vertical navigation menu with links to "NCBI Home", "Site Map (A-Z)", "All Resources", "Chemicals & Bioassays", "Data & Software", "DNA & RNA", "Domains & Structures", and "Genes & Expression". The main content area on the right is partially visible, showing the text "Welcome to NCBI" and "National Center for Biotechnology Information advances science and biomedical and genomic information." along with links for "Organization", "Research", and "RSS Feeds".

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



NCBI Resources ▾ How To ▾

Nucleotide
Alphabet of Life

Search: Nucleotide

Save search Limits **Advanced search** Help

vimentin Search Clear

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

- 三大データベースの配列を元に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

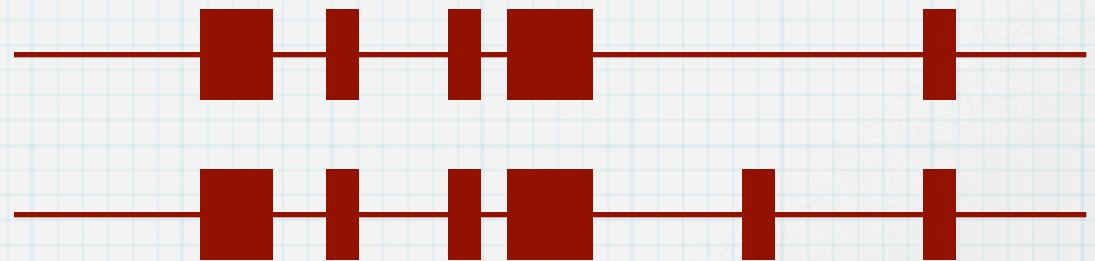
ヒト Chr22 (q11)



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)

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

Enter Query Sequence

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

Query subrange
From
To

Or, upload file
Job Title
Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database: Reference RNA sequences (refseq_rna)
Organism:
Exclude:
Entrez Query:

Reference RNA sequence (refseq_rna)

Program Selection

Optimize for: Highly similar sequences (megablast)

BLAST

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

C. elegans BLAT Search

BLAT Search Genome

生物種を選択

Genome: Assembly: Query type: Sort output: Output type:

C. elegans May 2008 (WS190/ce6) BLAT's guess query,score hyperlink

```
tgaatgaagacgatcgactcaaattcacagctccacaggatggaattcttctaacaagctcgacaattcgga
```

submit I'm feeling lucky clear

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: 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 submissions is 50,000 bases or 25,000 letters.

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
3ミスマッチを許容
- 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

統合遺伝子検索
GGRNA by @meso_cacase

1552311_a_at [再検索] Homo sapiens (human)

Summary:

search term:	hits:	results:
seq:GCATGGGATGGACAGTCTGGGCCA	1	NM_032753.3
seq:AGAAATGCGGCACACAGGGCAGGAC		RefSeqGene
seq:GGCAGGAGCTGCAGTAGCTACCCCTC		
seq:AGATCACTCCAGATCACCAGGTCA		
seq:AGGTCACCCCTCTCTAGGCGGCAC		
seq:AATGTCACCGCACACAGGCGAGTGG		
seq:GGGACACGGCAGTAAGACAAGAAA		
seq:ACGGCAGTAAGACAAGAAAGATT		
seq:TCTCCACAACGTTTTTAAATGTG		
seq:AAAATGCGCGGTGACTGGTGCA		
seq:ATGTGCGGGGTGACTGGTGACAC		

[AND]

Results:

[Homo sapiens retina and anterior neural fold homeobox 2 \(RAX2\), mRNA](#)

LOCUS NM_032753.3 - Homo sapiens (human) - NCBI - UCSC - RefSeq検索

DEFINITION: **Homo sapiens retina and anterior neural fold homeobox 2 (RAX2), mRNA**; linear; PRI 12-FEB-2011

ACCESSION: NM_032753.3

VERSION: 3

KEYWORDS: RefSeqGene project.

OR chromosomes to daughter cells. The d in mitosis and in response to activation t, and disappears when cells transition to ith a mitotic regulator, and its expression ly activate the spindle checkpoint. It is 2 kinase activity during spindle checkpoint tions that fuse this gene and the adjacent T cell leukemias, and are thought to arise D)-J recombination events. Multiple ding different isoforms have been found for RefSeq).

ent (4789)
ZIL"
onym="DKFZp686009161; MCPH7; SIL"
"tt"
dbSNP:34771841"
IL"
onym="DKFZp686009161; MCPH7; SIL"
IL"
onym="DKFZp686009161; MCPH7; SIL"
tccagagggcggggcgggagtgggcggctcttgaggc
gcaacagactcctcagctccgggaagtgggactgagatt
tccacagacatcaggacctatatactcttggcggccaga
ttcaagaagagtggtaccttccacttctccatacaaatgac
ccaactggagatctcactcaactcaactcaactcaactca
tgagaagccatccgacttgcttatgctcatgtaagcagaataaaaaa
ctttacttggtctctgacagcagcagaagtgaagaaggtgaactga
gctttgatcctggcggagagtagctgaactgaactgaactgaactgaactg
ggggacttttgatccactcaactcaactcaactcaactcaactgaaga
gttccactgtgactcaactcaactcaactcaactcaactcaactcaactga
aaagatccttgactggtaagctgcttccctcaagactcaactcaactcaactga
gttgacagctgggaattgactgcatgggcagcagtagctcaactcaactcaact

用データ:

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**
- 塩基配列から検索。 **seq1:** や **seq2:** を
付けると1, 2ミスマッチを含む配列もヒット
→ 完全一致と比べてどのくらい増える？
- さきほどの配列をさがしてみる