

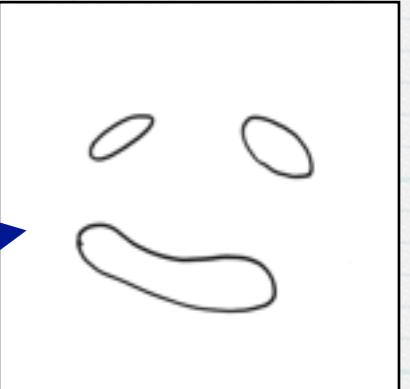


統合データベース講習会
AJACS伊予

塩基配列解析のための データベース・ウェブツール

ライフサイエンス統合データベースセンター (DBCLS)
内藤 雄樹

自己紹介

- 内藤 雄樹 (ないとう ゆうき)
Twitter @meso_cacase → 
- ライフサイエンス統合データベースセンター (DBCLS) 特任助教
- 過去に RNAi メカニズム等の研究
siRNA設計サイト: siDirect 公開

活性が高く、オフターゲット
効果の少ないsiRNAを設計

siRNA設計 : siDirect

檢索

siDirect version 2.0 highly effective, target specific siRNA online design site. [Help](#)

Enter an accession number and retrieve sequence:

[retrieve sequence](#)

or Paste in a nucleotide sequence:

```
>sample sequence
ggctgccaag aacctgcagg aggcagaaga atggtacaaa tccaaatggc ctgacccttc
tgaggctgcc aaccggaca atgacgcctt gcgcggca aaggcaggat ccactgagta
ccggagacag gtgcagtccc tcacctgtga agtggatgcc cttaaaggaa ccaatgatgc
cctggAACgc cagatgcgtg aaatggaa gaacttgcc gttgaagctg ctaacttacca
agacactatt ggccgcctgc aggatgagat tcagaatatg aaggagggaa tggctcgta
ccttcgtgaa taccaagacc tgctcaatgt taagatggcc cttgacattt agattggcac
ctacaggaag ctgcttggaaag gcgaggagag caggattct ctgccttcc caaaactttc
ctccctgaac ctgagggaaa ctaatcttggaa ttcaactccct ctgggttata cccactcaa
aaggacactt ctgattaaga cggttggaaac tagatgttgc caggttatca acgaaacttc
tcagcatcac gatgaccttgc aataaaaatt gcacacactc agtgcagcaa tatattacca
```

[design siRNA](#)

Options: [click here](#)

Effective siRNA candidates

target position	target sequence 21nt target + 2nt overhang	RNA oligo sequences 21nt guide (5'→3') 21nt passenger (5'→3')	functional siRNA selection: Ui-Tei	seed-duplex stability (Tm);		specificity check: minimum number of mismatches against any off-targets;	
				guide	passenger	guide	passenger
24-46	cagaagaatggtacaaatccaag	UGGAUUUGUACCAUUCUUCUG GAAGAAUGGUACAAUCCAAG	U	20.1 °C	12.0 °C	2 [detail]	2 [detail]
159-181	cccttaaaggAACCAatgagtcc	ACUCAUUGGUUCCUUUAAGGG CUUAAAGGAACCAAUGAGUCC	U	18.1 °C	11.0 °C	2 [detail]	2 [detail]
261-283	aggatgagattcagaatatgaag	UCAUAUUCUGAAUCUCAUCU GAUGAGAUUCAGAAUAUGAAG	U	8.7 °C	20.4 °C	2 [detail]	3 [detail]
461-483	ctggttatacccactaaaaag	UUUUGAGUGGGUAUCAACAG GGUUGAUACCCACUCAAAAAG	U	19.2 °C	16.1 °C	3 [detail]	2 [detail]
462-484	tgggtgatacccactaaaaagg	UUUUGAGUGGGUAUCAACCA GUUGAUACCCACUCAAAAAGG	U	12.2 °C	16.1 °C	3 [detail]	2 [detail]
491-513	ctgattaagacggttgaardtag	AGUUUCAACCGUCUUAUCAG GAUUAAGACGGUUGAAACUAG	U	14.9 °C	6.9 °C	3 [detail]	3 [detail]
519-541	gacaggttatcaacgaaacttct	AAGUUUCGUUGAUACCUGUC CAGGUUAUCAACGAAACUUCU	U	19.7 °C	18.5 °C	3 [detail]	3 [detail]
521-543	caggttatcaacgaaacttctca	AGAAGUUUCGUUGAUACCUG GGUUAUCAACGAAACUUCUCA	U	13.3 °C	16.1 °C	2 [detail]	3 [detail]

Graphical view of effective siRNA candidates



遺伝子に関する データベースの基礎

- 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

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 shows the NCBI homepage with a search bar. A dropdown menu is open, listing various databases: All Databases, PubMed, Protein, Nucleotide, GSS, EST, Structure, Genome, BioProject (Genome Project), BioSample, BioSystems, Books, CancerChromosomes, Conserved Domains, dbGaP, dbVar, Epigenomics, Gene, and GENSAT. The 'Nucleotide' option is highlighted with a blue selection bar and a cursor arrow pointing to it.

NCBI Resources ▾ How To ▾

NCBI National Center for Biotechnology Information

NCBI Home

Site Map (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Search ✓ All Databases
PubMed
Protein
Nucleotide
GSS
EST
Structure
Genome
BioProject (Genome Project)
BioSample
BioSystems
Books
CancerChromosomes
Conserved Domains
dbGaP
dbVar
Epigenomics
Gene
GENSAT

Search Clear

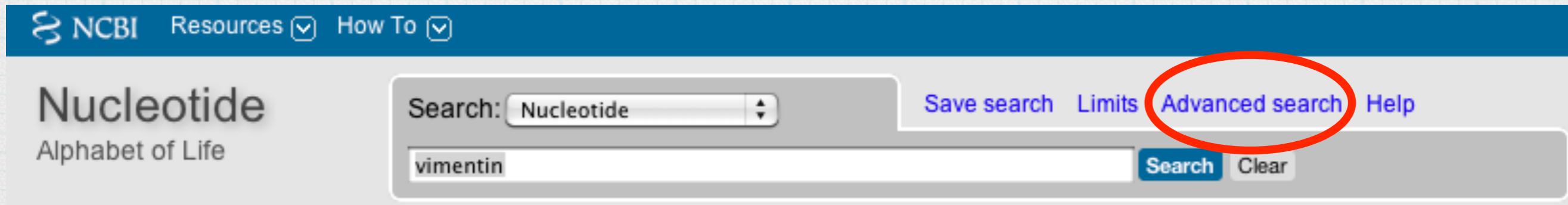
Technology Information advances science and biomedical and genomic information.

Organization | Research | RSS Feeds

NCBI software

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



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

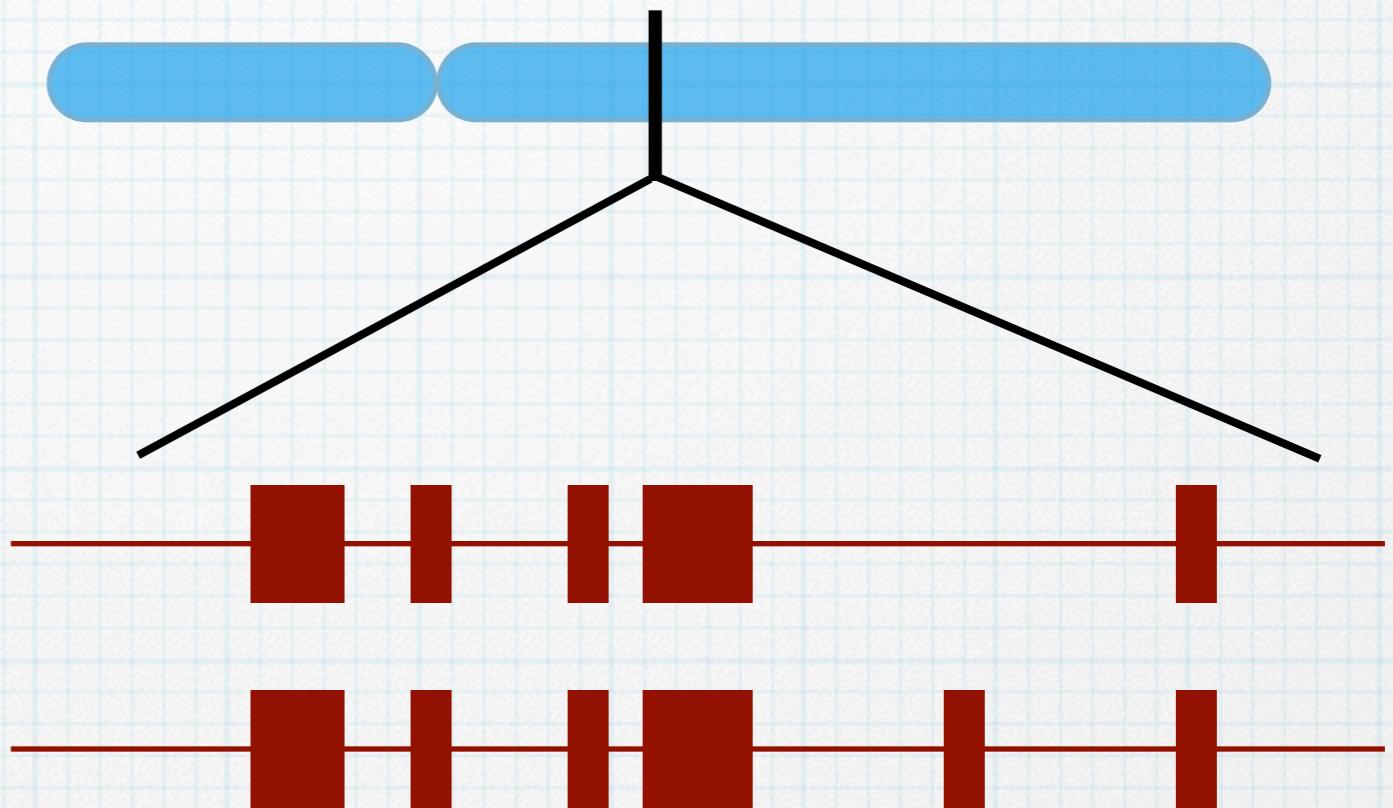
- ・遺伝子ごとに付与される遺伝子名と番号

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)

NCBI BLAST/blastn suite

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

[Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#)

[Clear](#)

Query subrange [?](#)

tgaatgaagacgatcgactcaaattcacagctccacaggatggaaattttttaacaaagctcgacaattcgga

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 (in 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: Assembly: Query type: Sort output: Output type:

tgaatgaagacgatcgactcaaattcacagctccacaggatggaattttttaacaaagctcgacaattcgga

Paste in a query sequence to find its location in 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 submission is 50,000 bases or 25,000 letters.

遺伝子データベースの検索

- 検索ワードが多様（遺伝子名, ID, 遺伝子機能, タンパクのドメイン名, 疾患, 塩基配列, アミノ酸配列,）
 - 入口が異なる。知らないと探せない

- BLAST 等による塩基配列の検索
 - 検索が遅い

Googleのように探せれば...



月が綺麗ですね 検索オプション
 ウェブ全体から検索 日本語のページを検索

約 1,380,000 件 (0.11 秒)

他のキーワード: [夏目漱石](#) [月が綺麗ですね](#) [月が綺麗ですね出典](#)

[月が綺麗ですね \(つきがきれいですね\)とは【ピクシブ百科事典】](#)

[dic.pixiv.net](#) › ... › クリエイター, 作家, 小説家, 文豪, 夏目漱石 ▾

月が綺麗ですねがイラスト付きでわかる！ "I love you" の日本の意訳。

[月が綺麗ですねとは \(ツキガキレイデスネとは\) \[単語記事\] - ニコニコ...](#)

[dic.nicovideo.jp/a/月が綺麗ですね](#) ▾

... 商品; 関連項目; 揭示板. 月が綺麗ですねとは、I love youの慣用表現である。 ... てい
ればマシだが「月が綺麗ですね」は富嶽百景の「みなさん、けふは富士がよく見えます
ね」と一緒だなみんないまさらのごとくやあ、とか、まあ、とか間抜けた嘆声を発する
のが

[異性に「月が綺麗ですね」ってメールしてみろwww.ふえー速](#)

[fesoku.net/archives/6862734.html](#) ▾

2013/10/13 - 異性に「月が綺麗ですね」ってメールしてみろwww. October ...
あなたと見る月は、いつも綺麗です。 35: 以下、 ... 優しい弟だね. 58: 以下、名無しに
かわりましてVIP がお釣りします 2013/10/12 22:36:25 ID:r1jaBvXOi. 妹に送ったら

[【文系集合】「月が綺麗ですね」に対して一番お洒落な返しをした...](#)

[blog.livedoor.jp/schoolunderground/archives/33205964.html](#) ▾

ぐぐるな 統合遺伝子検索GGRNA

統合遺伝子検索
GGRNA ver.2

[Help](#) | [Advanced search](#) | [English](#) | [旧バージョン](#)



検索

検索

Zoo (All organisms in RefSeq)

遺伝子をGoogleのように検索できるサイトです。 [NCBI RefSeq](#) の transcript を全文検索します。

検索例：

- ・ 「[homeobox](#)」 「[claudin](#)」 フリーワード検索
- ・ 「"RNA interference"」 ダブルクオートで囲ってフレーズ検索
- ・ 「[Argonaute](#) "PAZ domain"」 Argonaute かつ "PAZ domain" のAND検索
- ・ 「[NM_001518](#)」 「[10579](#)」 RefSeq IDやGene IDなど各種IDから検索
- ・ 「[symbol:VIM](#)」 遺伝子名(symbolまたはsynonym)から検索
- ・ 「[ref:Naito](#)」 文献情報のなかからフリーワード検索
- ・ 「[1552311_a_at](#)」 マイクロアレイのプローブIDから塩基配列を検索
- ・ 「[aa:KDEL](#)」 アミノ酸配列を検索
- ・ 「[caagaqaqagattq](#)」 塩基配列を検索
- ・ 「[comp:caaaqaqagattq](#)」 相補鎖を検索
- ・ 「[iub:aggcannrtgacct](#)」 N, R, Y 等のあいまいな塩基を含む塩基配列を検索
- ・ [詳細な使い方](#)

新着情報：

- ・ 2013-11-17 データベースをRefSeq rel. 62 (Nov, 2013)に更新。
- ・ 2013-07-24 ソースを公開 - [GitHub](#)
- ・ 2013-07-08 GGRNA ver.2公開。全生物種のRefSeqを検索できます。
- ・ 2012-05-29 [下記論文の日本語による解説を「DBCLSからの成果発信」に掲載。](#)
- ・ 2012-05-29 [GGRNAの論文がNucleic Acids Researchに掲載されました。](#)
- ・ [過去の新着情報](#)

遺伝子名で検索

統合遺伝子検索

GGRNA ver.2

[Help](#) | [Advanced search](#) | [English](#) | [旧バージョン](#)



検索

Zoo (All organisms in RefSeq)

遺伝子をGoogleのように検索できるサイトです。 [NCBI RefSeq](#) の transcript を全文検索します。

検索例：

- ・ 「[homeobox](#)」 「[claudin](#)」 フリーワード検索
- ・ 「"RNA interference"」 ダブルクオートで囲ってフレーズ検索
- ・ 「[Argonaute "PAZ domain"](#)」 Argonaute かつ "PAZ domain" のAND検索
- ・ 「[NM_001518](#)」 「[10579](#)」 RefSeq IDやGene IDなど各種IDから検索
- ・ 「[symbol:VIM](#)」 遺伝子名(symbolまたはsynonym)から検索
- ・ 「[ref:Naito](#)」 文献情報のなかからフリーワード検索
- ・ 「[1552311_a_at](#)」 マイクロアレイのプローブIDから塩基配列を検索
- ・ 「[aa:KDEL](#)」 アミノ酸配列を検索
- ・ 「[caagaqaqagattq](#)」 塩基配列を検索
- ・ 「[comp:caaaqaqagattq](#)」 相補鎖を検索
- ・ 「[iub:aggcannrtqacct](#)」 N, R, Y 等のあいまいな塩基を含む塩基配列を検索
- ・ [詳細な使い方](#)

新着情報：

- ・ 2013-11-17 データベースをRefSeq rel. 62 (Nov, 2013)に更新。
- ・ 2013-07-24 ソースを公開 - [GitHub](#)
- ・ 2013-07-08 GGRNA ver.2公開。全生物種のRefSeqを検索できます。
- ・ 2012-05-29 [下記論文の日本語による解説を「DBCLSからの成果発信」に掲載。](#)
- ・ 2012-05-29 [GGRNAの論文がNucleic Acids Researchに掲載されました。](#)
- ・ [過去の新着情報](#)

RefSeqの転写産物

- ・ mRNA ([NM_](#), [XM_](#))
- ・ ncRNA ([NR_](#), [XR_](#))

遺伝子名で検索

統合遺伝子検索

GGRNA

ver.2



[Help](#) | [Advanced search](#) | [English](#) | [日本語](#)

nanog

検索

Homo sapiens (human)

2013-12-06 00:52:08, GGRNA : RefSeq release 62 (Nov, 2013)

Summary:

- [nanog \(41\)](#)
- [INTERSECTION \(41\)](#)

Results:

トップ50件を表示。検索語に色がつきます。重なると色が濃く表示されます。

[Homo sapiens Nanog homeobox \(NANOG\), mRNA.](#) (2098 bp)

tryptase(+) / **Nanog**(+) cells in infantile haemangioma is novel. REFERENCE 2 (bases 1 to 2098) AUTHORS Chu,W.K., Dai,P.M., Li,H.L., Pao,C.C. and Chen,J.K. TITLE **Nanog** expression is negatively regulated by protein kinase C activities in human cancer cell lines JOURNAL Carcinogenesis 34 (7), 1497-1509 (2013) PUBMED 23536578 REMARK GeneRIF: Knockdown of protein kinase C alpha and delta enhances **nanog** expression in cancer cell lines. REFERENCE 3 (bases 1 to 2098) AUTHORS Sun,C., Sun,L., Jiang,K., Gao,D.M., Kang,X.N., Wang,C., Zhang,S., Huang,S., Qin,X., Li,Y. and Liu,Y.K. TITLE **NANOG** promotes liver...

[NM_024865.2 - Homo sapiens \(human\) - NCBI - UCSC - RefEx\(expression\)](#)

[Homo sapiens NANOG neighbor homeobox \(NANOGNB\), mRNA.](#) (907 bp)

LOCUS NM_001145465 907 bp mRNA linear PRI 22-SEP-2013 DEFINITION Homo sapiens **NANOG** neighbor homeobox (**NANOGNB**), mRNA. ACCESSION NM_001145465 XR_040558 XR_040559 XR_040560 VERSION NM_001145465.1 GI:224451052 KEYWORDS RefSeq. SOURCE Homo sapiens (human) ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo. REFERENCE COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final NCBI review. The reference sequence was derived from AY151139.1. On or before Mar... NM_001145465.1 - Homo sapiens (human) - NCBI - UCSC - RefEx(expression)

[PREDICTED: Homo sapiens Nanog homeobox \(NANOG\), transcript variant X1, mRNA.](#) (1023 bp)

LOCUS XM_005253484 1023 bp mRNA linear PRI 13-AUG-2013 DEFINITION PREDICTED: Homo sapiens **Nanog** homeobox (**NANOG**), transcript variant X1, mRNA. ACCESSION XM_005253484 VERSION XM_005253484.1 GI:530399393 DBLINK BioProject: PRJNA168 KEYWORDS RefSeq.

Accession番号で検索

統合遺伝子検索
GGRNA ver.2

TOE PICTURE GALLERY Help | Advanced search | English 日バージョン

遺伝子をGoogleのように検索できるサイトです。 [NCBI RefSeq](#) の transcript を全文検索します。

検索例：

- ・ 「[homeobox](#)」 「[claudin](#)」 フリーワード検索
- ・ 「"[RNA interference](#)"」 ダブルクオートで囲ってフレーズ検索
- ・ 「[Argonaute "PAZ domain"](#)」 Argonaute かつ "PAZ domain" のAND検索
- ・ 「[NM_001518](#)」 「[10579](#)」 RefSeq IDやGene IDなど各種IDから検索
- ・ 「[symbol:VIM](#)」 遺伝子名(symbolまたはsynonym)から検索
- ・ 「[ref:Naito](#)」 文献情報のなかからフリーワード検索
- ・ 「[1552311_a_at](#)」 マイクロアレイのプローブIDから塩基配列を検索
- ・ 「[aa:KDEL](#)」 アミノ酸配列を検索
- ・ 「[caagaqaqagattq](#)」 塩基配列を検索
- ・ 「[comp:caagaqaqagattq](#)」 相補鎖を検索
- ・ 「[iub:aqgtcannrtqacct](#)」 N, R, Y 等のあいまいな塩基を含む塩基配列を検索
- ・ [詳細な使い方](#)

新着情報：

- ・ 2013-11-17 データベースをRefSeq rel. 62 (Nov, 2013)に更新。
- ・ 2013-07-24 ソースを公開 - [GitHub](#)
- ・ 2013-07-08 GGRNA ver.2公開。全生物種のRefSeqを検索できます。
- ・ 2012-05-29 [下記論文の日本語による解説を「DBCLSからの成果発信」に掲載。](#)
- ・ 2012-05-29 [GGRNAの論文がNucleic Acids Researchに掲載されました。](#)
- ・ [過去の新着情報](#)

塩基配列で検索

統合遺伝子検索

GGRNA

ver.2



IMAGE
GALLERY

[Help](#) | [Advanced search](#) | [English](#) | [日本語](#)

NM_003380

検索

Homo sapiens (human)

▼

2013-12-06 02:15:41, GGRNA : RefSeq release 62 (Nov, 2013)

Summary:

- [refid:NM_003380 \(1\)](#)
- [INTERSECTION \(1\)](#)

Results:

トップ50件を表示。検索語に色がつきます。重なると色が濃く表示されます。

[Homo sapiens vimentin \(VIM\), mRNA.](#) (2151 bp)

LOCUS **NM_003380** 2151 bp mRNA linear PRI 02-NOV-2013 DEFINITION Homo sapiens vimentin (VIM), mRNA. ACCESSION **NM_003380** VERSION

NM_003380.3 GI:240849334 KEYWORDS RefSeq. SOURCE Homo sapiens (human) ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata;

Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo. REFERENCE

COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The reference sequence was derived from BQ050765.1 and BC000163.2.

This sequence is a reference standard in the RefSeqGene project. On Jun...

Synonym: CTRCT30

NM_003380.3 - Homo sapiens (human) - [NCBI](#) - [UCSC](#) - [RefEx\(expression\)](#)

Data Export:

下記より最大10000件まで検索結果を取得できます。

- タブ区切りテキスト → [表示](#) | [ダウンロード](#)
エクセル等の表計算ソフトに直接コピペできます。
- JSON形式 → [リンク](#) | [ダウンロード](#)

塩基配列で検索

統合遺伝子検索

GGRNA
ver.2



TOE PICTURE
GALLERY

[Help](#) | [Advanced search](#) | [English](#) | [日本語](#)

caagaagagattg

検索

Homo sapiens (human)

2013-12-06 02:16:05, GGRNA : RefSeq release 62 (Nov, 2013)

Summary:

- [seq:caagaagagattg \(19\)](#)
- [INTERSECTION \(19\)](#)

Results:

トップ50件を表示。検索語に色がつきます。重なると色が濃く表示されます。

[Homo sapiens carboxypeptidase A3 \(mast cell\) \(CPA3\), mRNA. \(1699 bp\)](#)

agaaccatgaggctatcctgcctgtgggtttgattgctaccacttgcattgtctccgtttgacagggagaagggtttccgcgtgaagccccaggataaaaacaaggcataaaggacttggccaaac
caatgagcttgcattctgttatccagggtccaccacgttagtcataatatgtatggattccgagtttagtgagaaggaatcccaagccatccagtctgccttgatcaaataaaatgcactatgattca
tgtatcta**caagaagagattg**agaaacagttgatgttaagaagatatcccaggcaggcacagctacgcaaaatacaataattggaaaagattgtggcttgactgaaaagatgatggataagtatcctgaaatggc
tctcgattaaattggatctactgttgaagataatccactatgttgcattttttacggattgtggatttcacgcacgagaatgggtctcccgacgttgcattctgcaatgg...
position 341

Synonym: MC-CPA

NM_001870.2 - Homo sapiens (human) - [NCBI](#) - [UCSC](#) - [RefEx\(expression\)](#)

[Homo sapiens vimentin \(VIM\), mRNA. \(2151 bp\)](#)

atcctgctggccgagctcgagctcaaggccaaggcaagtgcgcctggggaccttacgaggaggatgcgggagctgcgcggcagggtggaccagctaaccacgacaaagccgcgtcgagggtggagcg
acaacccgtggcggaggacatcatgcgcctccggagaaaattgcaggaggagatgctcagagagaggaagccgaaaacaccctgcaatcttcagacaggatgttgcataatgcgtctggcacgtctggaccttgc
aagtggaatcttg**caagaagagattg**ccttttgaagaaactccacgaagagaaaatccaggagctgcaggctcagattcaggaacagcatgtccaaatcgatgtggatgtttcaagcctcacggctgccctgc
gtgacgtacgtcagcaatatgaaagtgtggctgccaagaacctgcaggaggcagaagaatggtacaaatccaagttgtgacctctgtgaggctgccaaccggaaacaatgacgcctgcgcccaggcaagcaggatcc
actggataccggagacaggtgcagtccctcacctgtgaa...
position 1095

Synonym: CTRCT30

NM_003380.3 - Homo sapiens (human) - [NCBI](#) - [UCSC](#) - [RefEx\(expression\)](#)

塩基配列で検索

統合遺伝子検索

GGRNA

ver.2



TOE PICTURE
GALLERY

[Help](#) | [Advanced search](#) | [English](#) | [日本語](#)

caagaagaga

検索

Homo sapiens (human)

▼

2013-12-06 02:16:23, GGRNA : RefSeq release 62 (Nov, 2013)

Summary:

- [seq:caagaagaga \(1297\)](#)
- [INTERSECTION \(1297\)](#)

Results:

トップ50件を表示。検索語に色がつきます。重なると色が濃く表示されます。

[Homo sapiens microRNA 5197 \(MIR5197\), microRNA. \(112 bp\)](#)

AUTHORS Griffiths-Jones,S., Grocock,R.J., van Dongen,S., Bateman,A. and Enright,A.J. TITLE miRBase: microRNA sequences, targets and gene nomenclature JOURNAL Nucleic Acids Res. 34 (DATABASE ISSUE), D140-D144 (2006) PUBMED 16381832

tatggattccacagacaatgaggatcaatggcacaaactcatttgaagccagtt**caagaagaga**actgagtcatcaatgtcacctcatgt

position 63

[NR_049829.1 - Homo sapiens \(human\) - NCBI - UCSC - RefEx\(expression\)](#)

[PREDICTED: Homo sapiens uncharacterized LOC101927342 \(LOC101927342\), transcript variant X2, misc_RNA. \(180 bp\)](#)

analysis using gene prediction method: Gnomon. Supporting evidence includes similarity to: 1 EST, and 100% coverage by RNAseq alignments, including 1 sample with support for all introns" /db_xref="GeneID:101927342" ORIGIN //

ctatggttccttagactgcactttgagaactattgacgtacaggaaaaatgcttttgtgc**caagaagaga**ttccgaagatagaagaggaggcatcaacaccgtgaacatcatggctgagtttttgagaggata
agcctaaagtaaaaacccctggatttgcctctgtccttcaa

position 63

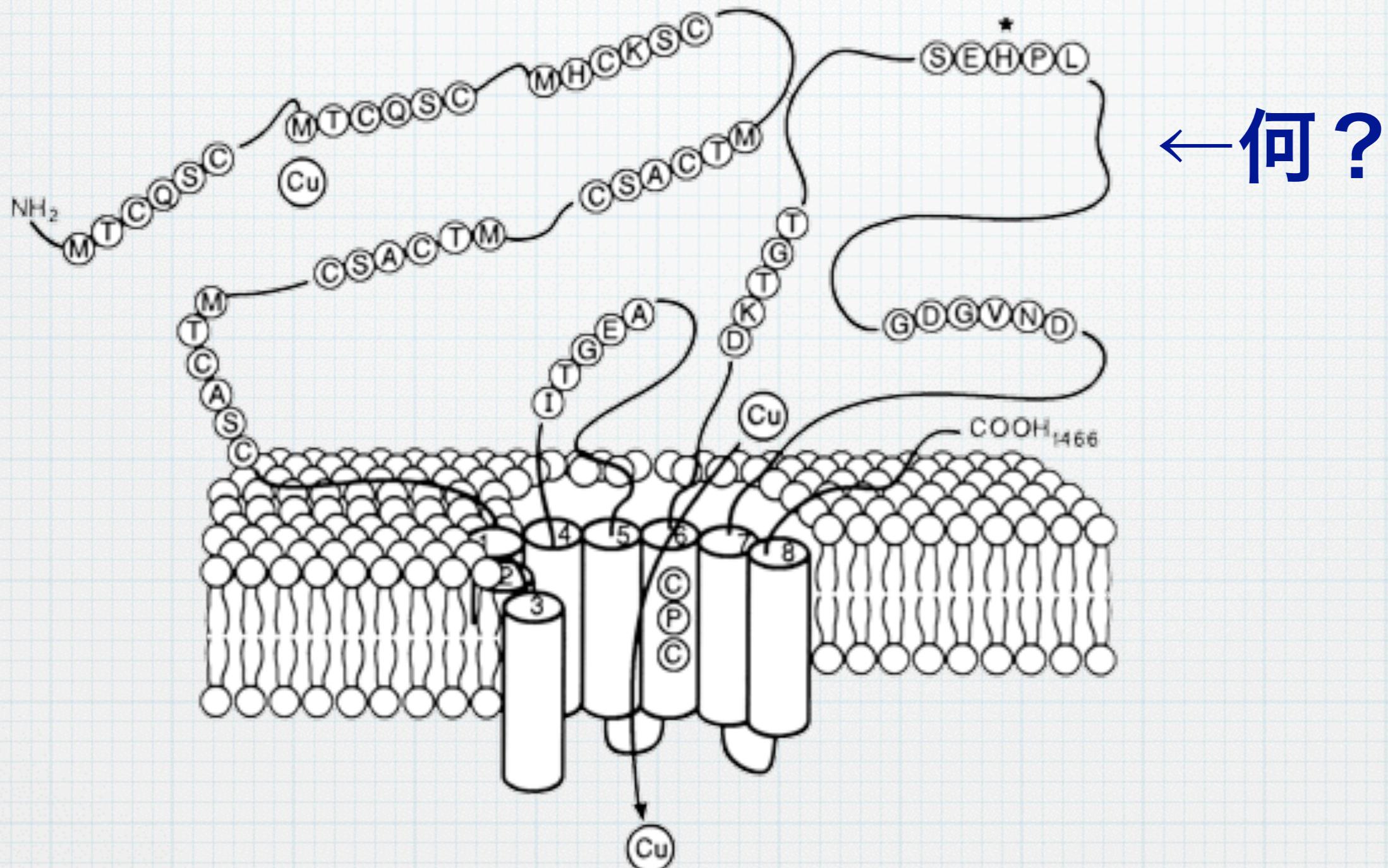
[XR_250764.1 - Homo sapiens \(human\) - NCBI - UCSC - RefEx\(expression\)](#)

[PREDICTED: Homo sapiens uncharacterized LOC101927342 \(LOC101927342\), transcript variant X1, misc_RNA. \(265 bp\)](#)

method: Gnomon. Supporting evidence includes similarity to: 2 ESTs, and 66% coverage by RNAseq alignments" /db_xref="GeneID:101927342" ORIGIN //

gatctccatgacctctatggttccttagactgcactttgagaactattgacgtacaggaaaaatgcttttgtgc**caagaagaga**ttccgaagatagaagaggaggcatcaacaccgtgaacatcatggctgagtttttgagaggata
gatctccatgacctctatggttccttagactgcactttgagaactattgacgtacaggaaaaatgcttttgtgc**caagaagaga**ttccgaagatagaagaggaggcatcaacaccgtgaacatcatggctgagtttttgagaggata

アミノ酸配列で検索



Schaefer et al. (1999) IV. Wilson's disease and Menkes disease.
Am. J. Physiol. Gastrointest. Liver Physiol. 276, G311-G314

PCRのプライマー

ヒトのある遺伝子に対して
RT-PCRをかけようとしたら
なぜかバンドが2本・・・

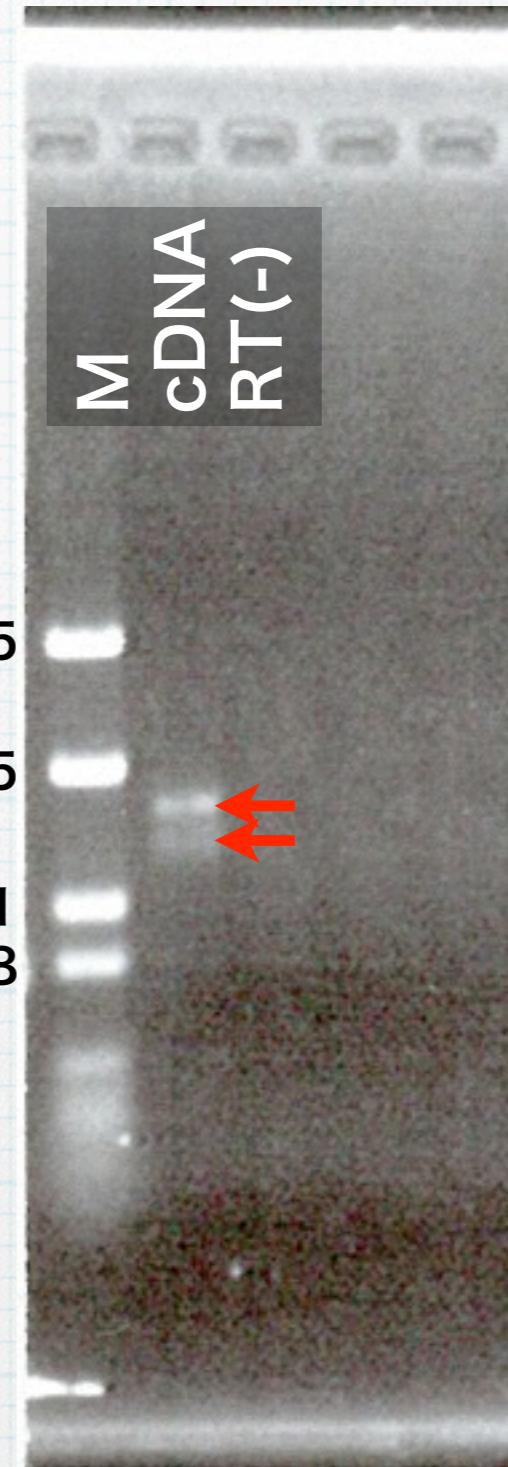
primer-F:

agctcattactttatcagtgcA

primer-R:

tgacgtattcaactcttctggtt

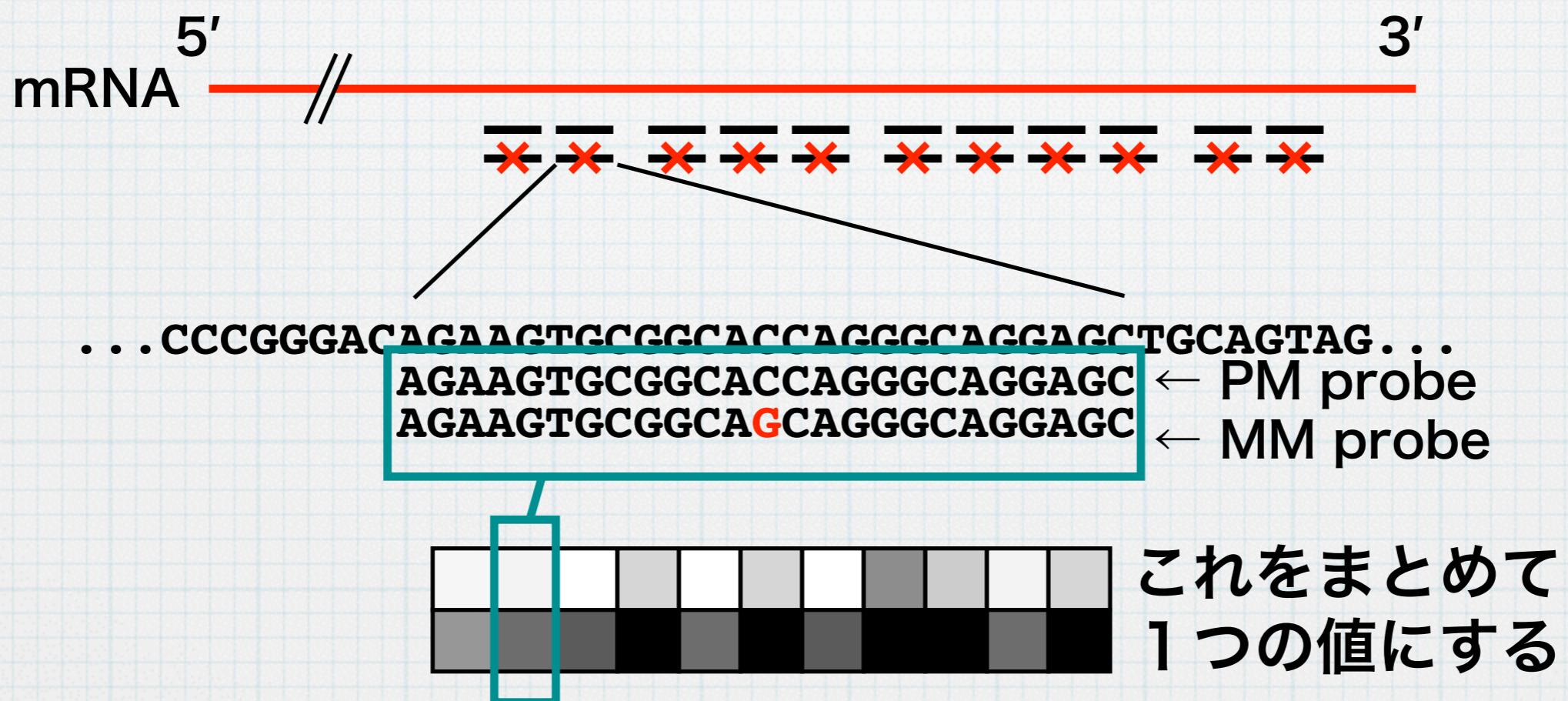
増幅遺伝子は何か？
予想されるバンドのサイズは？



※架空のデータです

マイクロアレイのプローブID

Affymetrix社GeneChipマイクロアレイ
1遺伝子につき25塩基 × 11箇所のプローブ
→ 「プローブセット」 例) 1552311_a_at



マイクロアレイのプローブID

統合遺伝子検索
GGRNA ver.2

TOME PICTURE GALLERY Help | Advanced search | English | 日バージョン

1552311_a_at Homo sapiens (human) ▾

2013-12-05 21:34:57, GGRNA : RefSeq release 62 (Nov, 2013)

Summary:

- [seq:TCTCCACAAACGTTTTAAAATGTG \(2\)](#)
- [seq:GGGACACGGCAGTAAGCACAAGAAA \(2\)](#)
- [seq:GGCAGGAGCTGCAGTAGCTACCCCTC \(2\)](#)
- [seq:GCATGGATGGGACAGTCTGGGCCA \(2\)](#)
- [seq:ATGTGCCGGTGTACTGGTGCACAC \(2\)](#)
- [seq:AGGTCACCCCATCTCTAGGCGGCAC \(2\)](#)
- [seq:AGATCACTCCCAGATCACCAAGGTCA \(2\)](#)
- [seq:AGAAGTGGGCACCAGGGCAGGAGC \(2\)](#)
- [seq:ACGGCAGTAAGCACAAGAAAGATT \(2\)](#)
- [seq:AATGTCACCGCACACCAGGCAGTGG \(2\)](#)
- [seq:AAAATGTGCCGGTGTACTGGTGCA \(2\)](#)
- **INTERSECTION (2)**

Results:

トップ50件を表示。検索語に色がつきます。重なると色が濃く表示されます。

[Homo sapiens retina and anterior neural fold homeobox 2 \(RAX2\), mRNA. \(2190 bp\)](#)

cagccgcacggccaaagagactctgagcaaggaggctgtcgccccggcagaaccggctggactggcaagcacggctggccagtgcagcaggagggggccctgaggcatggatggacagtctggccacgc
ccacctccgggacagaagtgcggcaccaggcaggagctgcagtagctaccctcccgctccagcctggctccccagatcactcccagatcaccaggtaaccatctctaggcggcacccacaca
ccagtcctgtggccaacgccccccatcaccaatgtcaccgcacaccaggcagtgggacacggcagtaagcacaagaagattttttttaaagctaaaccaggccaggtgcggctcatgcctgta
atcccagtgtttggaggctgaggtggaggattgtgagaccagcctgggtgacacagcaagacccatctccacaaacgttttaaatgtgccgggtgtactggtgacacactgtcatccagtacccaag
aagctgaggcaagaggatcacttgagccagaaggctcgaggctgcaggagctgtgatcacactg...

position 1592 1634 1650 1698 1717 1783 1807 1812 1955 1972 1975

Synonym: ARMD6; CORD11; QRX; RAXL1

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

統合遺伝子検索ぐぐるな GGRNA

- 遺伝子をGoogleのように検索
- 検索対象はRefSeqの転写産物
(mRNA, ncRNA)
- さまざまなキーワードに対応
- 塩基配列、アミノ酸配列も素早く検索
(とくに短い配列の検索が得意)

Naito & Bono, GGRNA: an ultrafast, transcript-oriented search engine for genes and transcripts.
Nucleic Acids Res. **40**, W592-6 (2012)

高速配列検索GGGenome

超絶高速ゲノム配列検索
GGGenome

[Help](#) | [English](#)

検索



検索

Human genome, GRCh37/hg19 (Feb, 2009)

許容するミスマッチ/ギャップの数 : (検索する塩基配列の長さの20%まで)

検索例 :

- [[TTCATTGACAACATT](#)] 塩基配列を検索
- [詳細な使い方](#)

検索結果へのリンク :

- [http://GGGenome.dbcls.jp/db/k/sequence\[.format\]\[.download\]](http://GGGenome.dbcls.jp/db/k/sequence[.format][.download])
 - db → hg19, mm10, rn5, galGal4, xenTro3, danRer7, ci2, dm3, ce10, TAIR10, rice, bmor1, refseq, ddbj。省略時は hg19
 - k → 許容するミスマッチ/ギャップの数。あまり大きいとしほうする。省略時は 0
 - sequence → 塩基配列。大文字・小文字は区別しない
 - format → html, txt, json。省略時は html
 - download → URLの最後に付加すると検索結果をファイルとしてダウンロードできる
- 例1 : <http://GGGenome.dbcls.jp/TTCATTGACAACATT>
 - ヒトゲノム hg19 (省略可) で
 - ミスマッチ/ギャップを許容せず (省略可)
 - [TTCATTGACAACATT](#) を検索し
 - html 形式 (省略可) で結果を返す
- 例2 : <http://GGGenome.dbcls.jp/mm10/2/TTCATTGACAACATTGCGT.txt>
 - マウスゲノム mm10 で
 - 2 ミスマッチ/ギャップまで許容して
 - [TTCATTGACAACATTGCGT](#) を検索し
 - txt 形式 (タブ区切りテキスト) で結果を返す

新着情報 :

- 2013-10-02 TogoGenome from RefSeq 60 (Jul, 2013) に更新。
- 2013-09-20 RefSeq rel. 61 (Sep, 2013) に更新。

高速配列検索GGGenome

げ げ げ

超絶高速ゲノム配列検索

[Help](#) | [English](#)

GGGenome

TTCACTGACAACATTGAGTA

検索

Human genome, GRCh37/hg19 (Feb, 2009)

許容するミスマッチ/ギャップの数 : (検索する塩基配列の長さの20%まで)

2013-12-06 06:20:53, GGGenome : Human genome, GRCh37/hg19 (Feb, 2009)

Summary:

- [TTCACTGACAACATTGAGTA \(36\)](#)
- [TACTCAATGTTGTCAGTGAA \(33\)](#)

Results:

+鎖および-鎖それぞれ50件まで表示。検索語に色がつきます（ミスマッチ・挿入欠失）。

[chr1:77930953-77930971](#)

▼77930953

AGCAGTACTCAAGATCACCTTTTAAAGTTCACTG-CAACATT~~CAGTA~~GATGAAAAATGTGTAGCTGTAATTCAATTGA

[chr1:106947513-106947532](#)

▼106947513

TAGAAAACATACCAACATAAAGAACATCTTTCACTG~~CACAA~~-ATTGAGTATAAAATTAAACCCGTTGAAAGGGTCATTA

[chr2:30898313-30898332](#)

▼30898313

GCCCCCAGCCACTTATATGTATTTTAAATTTCACTGACAACATTGAGTAAGAAAGATAATTTTTTTTGAGACAG

[chr2:141517099-141517116](#)

▼141517099

TTTATTTCTTAACACAGCCTTGATAAAATTCACTGAC-ACATTGAGT-TAGCCATAGTGTGAATGGGAGACTGCCA

[chr2:163535086-163535103](#)

▼163535086

TGAGAAAGAGGACACCCTAAGAATTATGGCTTCACTGACAACATT-AG-A~~GA~~AACATATCAAACGAATATTCTGG

[chr2:186291305-186291324](#)

▼186291305

ATTATGAACAGAAAAAGCAAAGGTTACTTTCACTGAG~~CA~~ACATTGAG-ATAGAGTTCTTGACACTTAATTGTGTA

GGRNA REST API

http://GGRNA.dbcls.jp/spe/query+string[.format][.download]

GGGenome REST API

http://GGGenome.dbcls.jp/db/k/sequence[.format][.download]

- 検索結果とURLとが1対1対応
- テキスト形式やJSON形式でも出力

GGGenome REST API



超絶高速ゲノム配列検索

[Help](#) | [English](#)

GGGenome

TTCACTGACAACATTGAGTA

検索

Human genome, GRCh37/hg19 (Feb, 2009)

▼

許容するミスマッチ/ギャップの数 : (検索する塩基配列の長さの20%まで)

2013-12-06 07:36:41, GGGenome : Human genome, GRCh37/hg19 (Feb, 2009)

Summary:

- [TTCACTGACAACATTGAGTA \(36\)](#)
- [TACTCAATGTTGTCAGTGAA \(33\)](#)

Results:

+鎖および-鎖それぞれ50件まで表示。検索語に色がつきます（ミスマッチ・挿入欠失）。

[chr1:77930953-77930971](#)

▼77930953

AGCAGTACTCAAGATCACCTTTTTAAAG **TTCACTG**-CAACATT**CAGTA**GATGAAAAATGTGTAGCTGTAATTCAATTGA

[chr1:106947513-106947532](#)

▼106947513

TAGAAAACATAACCACATAAAGAACATCT**TTCACTG****CACAA**-ATTGAGTATAAAATTAAACCGTTGAAAGGGTCATTA

[chr2:30898313-30898332](#)

▼30898313

GCCCCCAGCCACTTATATGTATTTTAAAT**TTCACTGACAACATTGAGTA**GAAAAGATAATTTTTTTTGAGACAG

[chr2:141517099-141517116](#)

▼141517099

TTTATTTCTTAACACAGCCTTGATAAA**TTCACTGAC**-ACATTGAGT-TAGCCATAGTGTGTAATGGGAGACTGCCA

[chr2:163535086-163535103](#)

▼163535086

TGAGAAAGAGGACACCCCTAAGAATTATGGC**TTCACTGACAACATT**-AG-AGAACATATCAAACGAATATTCTGG

表計算ソフトでGGGenome

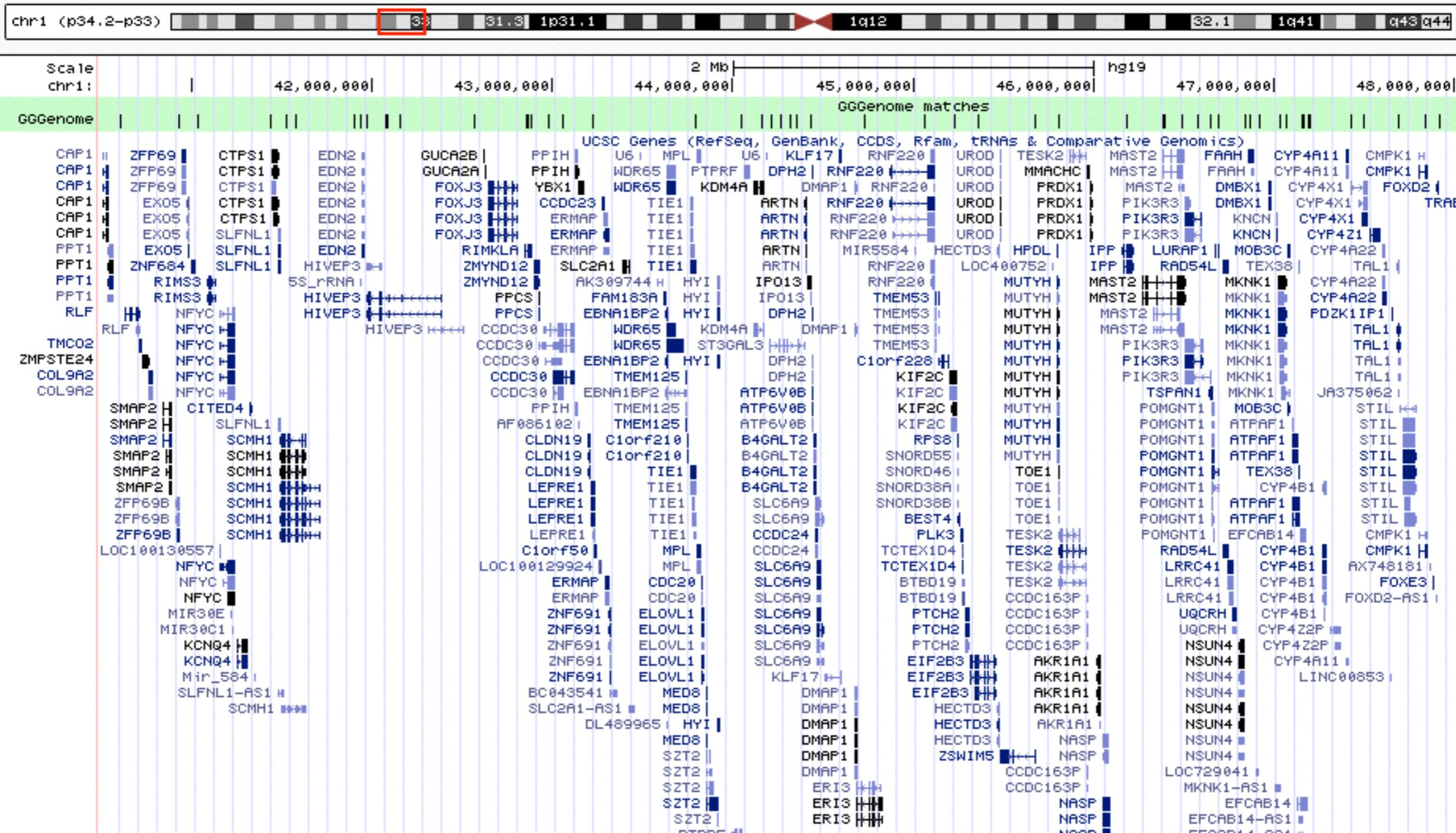
ヒットをゲノムブラウザ上に表示

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr1:40,500,001-49,500,000 9,000,000 bp. enter position, gene symbol or search terms

go



高速配列検索 GGGenome

- 塩基配列の検索に特化
(1kb 以下、短い配列が得意)
- 約200種のゲノムおよび転写産物
- ミスマッチ・ギャップに対応
- REST API が便利
R, Ruby, Galaxy でゲノム検索

CRISPRターゲット配列設計

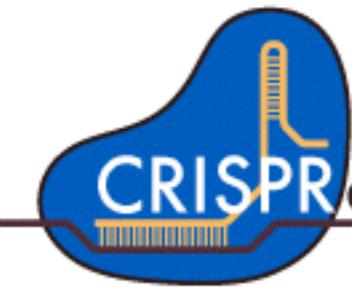
- NGGに隣接する20塩基を選択
- オフターゲットの少ない配列を選択

→ CRISPRdirect を公開



Naito, Hino, Bono & Ui-Tei,
CRISPRdirect: software for designing CRISPR/
Cas guide RNA with reduced off-target sites.
Bioinformatics, 31, 1120-1123 (2015)

CRISPRdirect



— Rational design of CRISPR/Cas target. [Help](#)

Enter an accession number (e.g. NM_006299) or genome location (e.g.

[retrieve sequence](#)

or Paste a nucleotide sequence: [?](#)

```
>sample sequence  
atgcccgcgtcgccgaccagagaaggcaagttcgagaacgaggagtttttaggaag  
ctgagccgcgtgtgagattaagtacacgggcttcagggaccggccccacgaggaacgc  
caggcacgcttccagaacgcctgccgcacggccgctcgaaatcgctttgtggccaca  
ggaaccatatctgtctccagttttccggccagctggcagggagaacagcgacaaaaca  
cctagccgagagtatgtcgacttagaaagagaaggcaggcaaggatatttgaaggctccc  
atgattctgaatggagtctgtgttatctggaaaggctggattgtctccaaagactggat  
ggtagatggctgtggagttgtgaggagcggccaggcaggaggatgcattgcacaa  
caggcccttgaagaggctcgagaaaggacacgcgaatttgaagatagagacaggctcat  
cgggaggaaatggaggcaagaagacaacaagaccctagtccctggccaatttaggtgt  
ggtagatgaccaaacttcgttaa
```

or upload sequence file: [?](#) [ファイルを選択](#) 選択されていません

PAM sequence requirement: (e.g. NGG, NRG) [?](#)

Specificity check: [?](#)

[design](#)

<u>ヒト</u> (hs) TOPO PICTURE GALLERY	<u>マウス</u> (mm) TOPO PICTURE GALLERY	<u>ラット</u> (rn) TOPO PICTURE GALLERY	<u>ニワトリ</u> (gg) TOPO PICTURE GALLERY
<u>ツメガエル</u> (xt) TOPO PICTURE GALLERY	<u>ゼブラフィッシュ</u> (dr) TOPO PICTURE GALLERY	<u>ホヤ</u> (ci) TOPO PICTURE GALLERY	<u>ショウジョウバエ</u> (dm) TOPO PICTURE GALLERY
<u>線虫</u> (ce) TOPO PICTURE GALLERY	<u>シロイヌナズナ</u> (at) TOPO PICTURE GALLERY	<u>イネ</u> (os) TOPO PICTURE GALLERY	
<u>出芽酵母</u> (sc) TOPO PICTURE GALLERY	<u>分裂酵母</u> (sp) TOPO PICTURE GALLERY		

+ マーモセット、
ブタ、カイコ、
ソルガム

CRISPRdirect

show **highly specific** target only

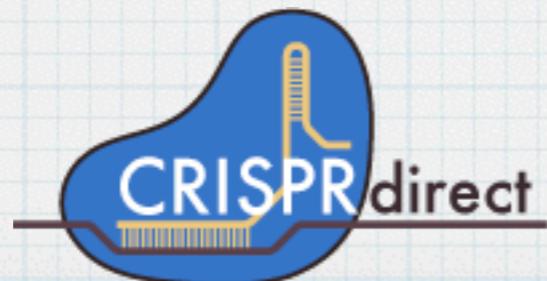
Show 50 entries

Search:

position		target sequence	sequence information			number of target sites <small>?</small>		
start	- end	20mer+PAM (total 23mer)	GC% of 20mer	Tm of 20mer	TTTT in 20mer	20mer +PAM	12mer +PAM	8mer +PAM
4 - 26	-	ccg cgcgtcggtccccgaccagag [gRNA]	75.00 %	82.34 °C	-	1 [detail]	3 [detail]	103 [detail]
16 - 38	-	ccc gaccagagaaggcaagttcga [gRNA]	50.00 %	71.17 °C	-	1 [detail]	26 [detail]	4354 [detail]
17 - 39	-	ccg accagagaaggcaagttcgag [gRNA]	50.00 %	71.53 °C	-	1 [detail]	32 [detail]	5910 [detail]
21 - 43	-	cca gagaaggcaagttcgagaacg [gRNA]	50.00 %	69.25 °C	-	1 [detail]	24 [detail]	8744 [detail]
24 - 46	+	gagaaggcaagttcgagaacg agg [gRNA]	50.00 %	69.25 °C	-	1 [detail]	2 [detail]	306 [detail]
35 - 57	+	tcgagaacgaggagttttt agg [gRNA]	40.00 %	66.49 °C	+	1 [detail]	25 [detail]	9901 [detail]
66 - 88	-	ccg cgagtgtgagattaagtaca [gRNA]	40.00 %	65.70 °C	-	0 [detail]	3 [detail]	1097 [detail]
69 - 91	+	cgagtgtgagattaagtaca cgg [gRNA]	40.00 %	65.70 °C	-	0 [detail]	26 [detail]	3790 [detail]
70 - 92	+	gagtgtgagattaagtacac ggg [gRNA]	40.00 %	65.27 °C	-	0 [detail]	3 [detail]	3604 [detail]
77 - 99	+	agattaagtacacgggcttc agg [gRNA]	45.00 %	70.40 °C	-	1 [detail]	6 [detail]	1306 [detail]
78 - 100	+	gattaagtacacgggcttca ggg [gRNA]	45.00 %	70.01 °C	-	1 [detail]	9 [detail]	6616 [detail]
83 - 105	+	agtacacgggcttcagggac cgg [gRNA]	60.00 %	78.32 °C	-	1 [detail]	63 [detail]	6397 [detail]
93 - 115	+	cttcagggaccggccccacg agg [gRNA]	75.00 %	84.21 °C	-	1 [detail]	7 [detail]	2568 [detail]
102 - 124	+	ccggccccacgaggaacgcc agg [gRNA]	80.00 %	86.67 °C	-	1 [detail]	1 [detail]	1244 [detail]
102 - 124	-	ccg gccccacgaggaacgccagg [gRNA]	75.00 %	84.31 °C	-	1 [detail]	19 [detail]	3777 [detail]
106 - 128	-	ccc cacgaggaacgccaggcacg [gRNA]	70.00 %	80.55 °C	-	1 [detail]	1 [detail]	1098 [detail]
107 - 129	-	ccc acgaggaacgccaggcacgc [gRNA]	70.00 %	82.51 °C	-	1 [detail]	2 [detail]	1116 [detail]
108 - 130	-	cca cgaggaacgccaggcacgc [gRNA]	70.00 %	82.52 °C	-	1 [detail]	2 [detail]	1011 [detail]
120 - 142	-	cca ggcacgcctccagaacgcct [gRNA]	65.00 %	80.61 °C	-	1 [detail]	6 [detail]	1510 [detail]
130 - 152	+	tccagaacgcctgccgcga cgg [gRNA]	65.00 %	80.58 °C	-	1 [detail]	4 [detail]	289 [detail]
132 - 154	-	cca gaacgcctgccgcacggcc [gRNA]	80.00 %	86.25 °C	-	1 [detail]	2 [detail]	937 [detail]
138 - 160	+	cgcctgccgcacggccgc tgg [gRNA]	85.00 %	90.21 °C	-	1 [detail]	5 [detail]	815 [detail]
140 - 162	-	cct gccgcgcacggccgc tggaa [gRNA]	80.00 %	86.71 °C	-	1 [detail]	1 [detail]	1044 [detail]
144 - 166	-	ccg cgacggccgc tggaaatcg [gRNA]	70.00 %	79.76 °C	-	1 [detail]	16 [detail]	628 [detail]
153 - 175	+	ccgctcgaaatcgcttttg tgg [gRNA]	55.00 %	72.84 °C	+	0 [detail]	7 [detail]	801 [detail]
153 - 175	-	ccg ctcggaaatcgctttgtgg [gRNA]	50.00 %	70.06 °C	-	0 [detail]	1 [detail]	805 [detail]
160 - 182	+	aaaaatcaactttataaccac agg [gRNA]	50.00 %	71.37 °C	+	0 [detail]	37 [detail]	7499 [detail]

CRISPRdirect

- NGGに隣接する20塩基を選択
- ゲノムのほかの領域とは塩基配列が一致せず、標的とする部位とだけ完全一致する配列を選択
- とくにPAM近傍の12merまたは8merの特異性を重視



Naito, Hino, Bono & Ui-Tei,
CRISPRdirect: software for designing CRISPR/
Cas guide RNA with reduced off-target sites.
Bioinformatics, 31, 1120-1123 (2015)

One more thing ...

テキスト比較ツール diff《デュフ》ver.6

“下の枠に比較したい文章を入れてください。差分 (diff) を表示します”

■ diff《デュフ》とは? → DEMO

■ 開発の経緯

- ・論文執筆（推敲）を効率化したい
- ・とくに共著者の修正をすばやく確認したい
- ・（はじめての）ウェブツールを書いてみたい

■ diff《デュフ》のしくみ

- ・UNIXのdiffコマンドを利用

テキスト比較ツール diff《デュフ》ver.6

“下の枠に比較したい文章を入れてください。差分 (diff) を表示します”

■ diff《デュフ》のしくみ

- UNIXのdiffコマンド(行ごとに比較)を利用

I|s||t|h|i{s}||a||p|e{n}|?|!|?|?|!

- 改行を挿入 → diffコマンド
- ただし一時ファイルは作成しない
- クエリのログも取らない

テキスト比較ツール diff《デュフ》ver.6

“下の枠に比較したい文章を入れてください。差分(diff)を表示します”

■ なんで《デュフ》と読むの？

- diff file → diff(ディフ・エフ) → デュフW

Hiromasa ONO
@h_ono

しかし、めそさんのデュフは役に立つなあ

返信 リツイート済み ★ お気に入り ... その他

4 リツイート 9 お気に入り

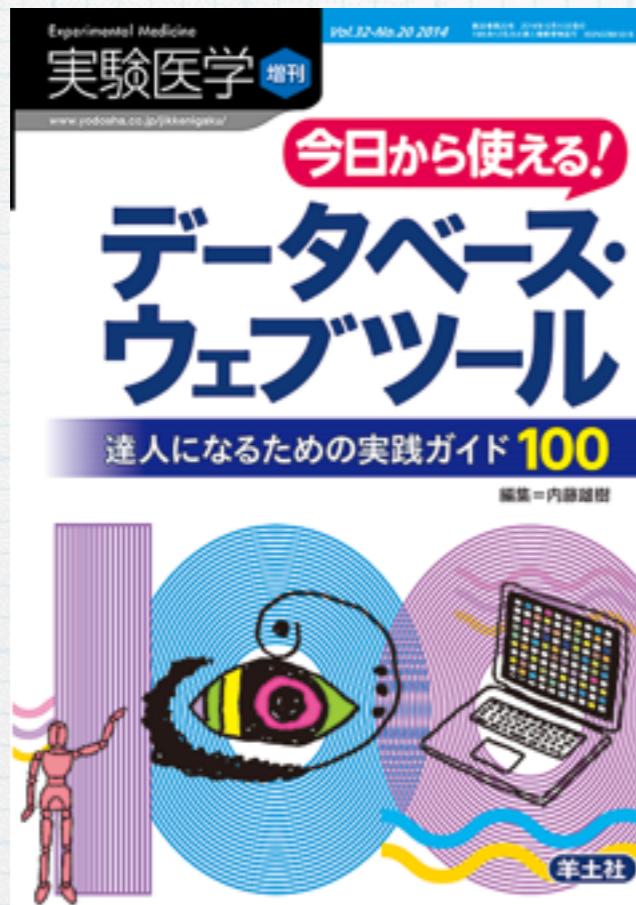
17:39 - 2012年4月10日

テキスト比較ツール diff《デュフ》ver.6

“下の枠に比較したい文章を入れてください。差分 (diff) を表示します”

■ diff《デュフ》の便利機能

- ・文字数・単語数のカウント
- ・モノクロ印刷モード
- ・ひとつのHTMLに保存。再開も可能
- ・結果を公開（3日間限定）



← 昨年12月発行
生命科学研究を支える
便利なデータベースやツール
約100個を簡潔に紹介

ありがとうございました