

 Tweet OK

統合データベース  
講習会 @ 遺伝研

# 塩基配列データベースの検索と CRISPRターゲット配列設計

ライフサイエンス統合データベースセンター (DBCCLS)

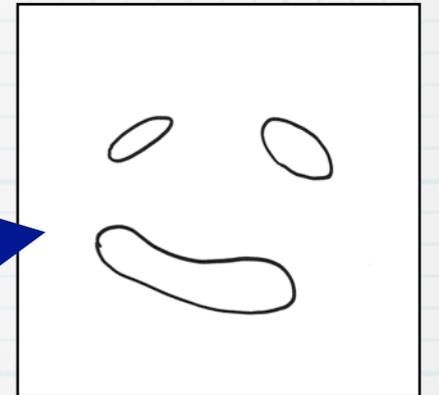
内藤雄樹

# 自己紹介

■ 内藤 雄樹 (ないとう ゆうき)



@meso\_cacase



■ ライフサイエンス統合データベース  
センター (DBCCLS) 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

検索

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 tccaagtttg ctgacctctc
tgaggctgcc aaccggaaca atgacgccct gcgccaggca aagcaggagt cactgagta
ccggagacag gtgcagtccc tcacctgtga agtggatgcc cttaaaggaa ccaatgagtc
cctggaacgc cagatgcgtg aatggaaga gaactttgcc gttgaagctg ctaactacca
agacactatt ggccgctgc aggatgagat tcagaatag aaggaggaaa tggctcgta
ccttcgtgaa taccaagacc tgctcaatgt taagatggcc ctgacattg agattgccac
ctacaggaag ctgctggaag gcgaggagag caggatttct ctgcctctc caaactttc
ctccctgaac ctgagggaaa ctaatctgga ttactccct ctggttgata cccactcaa
aaggacactt ctgattaaga cggttgaaac tagagatgga caggttatca acgaaactc
tcagcatcac gatgaccttg 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: U <sub>i</sub> -Tei	seed-duplex stability (T <sub>m</sub> );		specificity check: minimum number of mismatches against any off-targets;	
				guide	passenger	guide	passenger
24-46	cagaagaatggtacaaatccaag	UGGAUUUGUACCAUUCUUCUG GAAGAAUGGUACAAAUCCAAG	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	UCAUAUUCUGAAUCUCAUCCU GAUGAGAUUCAGAAUAUGAAG	U	8.7 °C	20.4 °C	2 [detail]	3 [detail]
461-483	ctggttgatacccactcaaaaag	UUUUGAGUGGGUAUCAACCAG GGUUGAUACCCACUCAAAAAG	U	19.2 °C	16.1 °C	3 [detail]	2 [detail]
462-484	tggttgatacccactcaaaaagg	UUUUUGAGUGGGUAUCAACCA GUUGAUACCCACUCAAAAAGG	U	12.2 °C	16.1 °C	3 [detail]	2 [detail]
491-513	ctgattaagacggttgaaactag	AGUUUCAACCGUCUUAUCAG GAUUAAGACGGUUGAAACUAG	U	14.9 °C	6.9 °C	3 [detail]	3 [detail]
519-541	gacaggttatcaacgaaacttct	AAGUUUCGUUGAUAAACCUGUC CAGGUUAUCAACGAAACUUCU	U	19.7 °C	18.5 °C	3 [detail]	3 [detail]
521-543	caggttatcaacgaaacttctca	AGAAGUUUCGUUGAUAAACCUG GGUUAUCAACGAAACUUCUCA	U	13.3 °C	16.1 °C	2 [detail]	3 [detail]

## Graphical view of effective siRNA candidates



## 本日のテーマ#1

# 塩基配列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のように探せば...



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"の日本的意識。

## [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; 掲示板. 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、とか間抜けた嘆息を発するのが

## [異性に「月が綺麗ですね」ってメールしてみるwwwwwwwwww:ふえ一速](#)

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

2013/10/13 - 異性に「月が綺麗ですね」ってメールしてみるwwwwwwwwww. 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](#) ▾

# 統合遺伝子検索 <sup>ぐぐる</sup> <sup>な</sup> GGRNA

検索

統合遺伝子検索

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

**GGRNA** ver.2

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プIDから塩基配列を検索
- 「[aa:KDEL](#)」 ..... アミノ酸配列を検索
- 「[caagaagagattg](#)」 ..... 塩基配列を検索
- 「[comp:caagaagagattg](#)」 ..... 相補鎖を検索
- 「[iub:aggtcannrtgacct](#)」 ..... 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](#)

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



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

検索

Homo sapiens (human)

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
- 「[caagaagagattg](#)」 ..... 塩基配列を検索
- 「[comp:caagaagagattg](#)」 ..... 相補鎖を検索
- 「[iub:aggtcannrtgacct](#)」 ..... 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



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

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)

agaaccatgaggctcatcctgctgtgggtttgattgctaccactcttgcaattgctcctgtccgcttgacagggagaaggtgtccgctgaagccccaggatgaaaaacaagcagacatcataaaggacttgccaaaac  
caatgagcttgacttctggtatccaggtgccaccaccacgtagctgctaataatgatggtggatttccgagtagtgagaaggaatccaagccatccagctctgccttgatcaaaataaaatgcactatgaaatcttgattca  
tgatcta**caagaagagattg**agaaacagtttgatgtaaagaagatatcccaggcaggcacagctacgaaaatacaataattgggaaaagatttggttgactgaaaagatgatggataagatcctgaaatggtc  
tctcgattaaaattggatctactgttgaagataatccactatatgttctgaagattgggaaaagaatgaaagaagaaaggctattttacggattgtggcattcacgcacgagaatgggtctccccagcattctgccagtgg  
ttgtctatcaggcaacccaaaacttatggg...

position 341

Synonym: MC-CPA

[NM\\_001870.2](#) - Homo sapiens (human) - [NCBI](#) - [UCSC](#) - [RefEx\(expression\)](#)

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

atcctgctggccgagctcgagcagctcaaggccaaggcaagtcgctgctggggacctctacgaggaggagatgctgggagctgctgcccaggtggaccagctaaccaacgacaaagccccgctcgaggtggagcgcg  
acaacctggccgaggacatcatgctcctccgggagaaattgcaggaggagatgcttcagagagaggaagccgaaaacacctgcaatcttcagacaggatggtgacaatgctgctctggcacgtcttgacctgaacgca  
aagtggaatctttg**caagaagagattg**ccttttgaagaaactccacgaagaggaaatccaggagctgcaggctcagattcaggaacagcatgtccaaatcgatgtggatgtttccaagcctgacctcacggctgcctgc  
gtgacgtacgtcagcaatatgaaagtgtggctgccaagaacctgcaggaggcagaagaatggtacaaatccaagtttgctgacctctgaggctgccaaccggaacaatgacgacctgcccaggcaagcaggagctcc  
actgagtaccggagacaggtgcagtcctcacctgtgaa...

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



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

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

tatgggattccacagacaatgagtatcaatggcacaactcattcttgaagccagttcaagaagagactgagtcacgaatgctctaaatgtcacttcacctcatgt

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 //

ctatggttccttagactgcactttgagaactattgacgtacagggaaaaatgctcttgggtgcccaagaagagattccgaagatagaagaggaggcatcaacaccgtgaacatcatggctgagttatgttgagagggata agcctaaagtaaaaaccttgatttgctctgtccttgaa

position 63

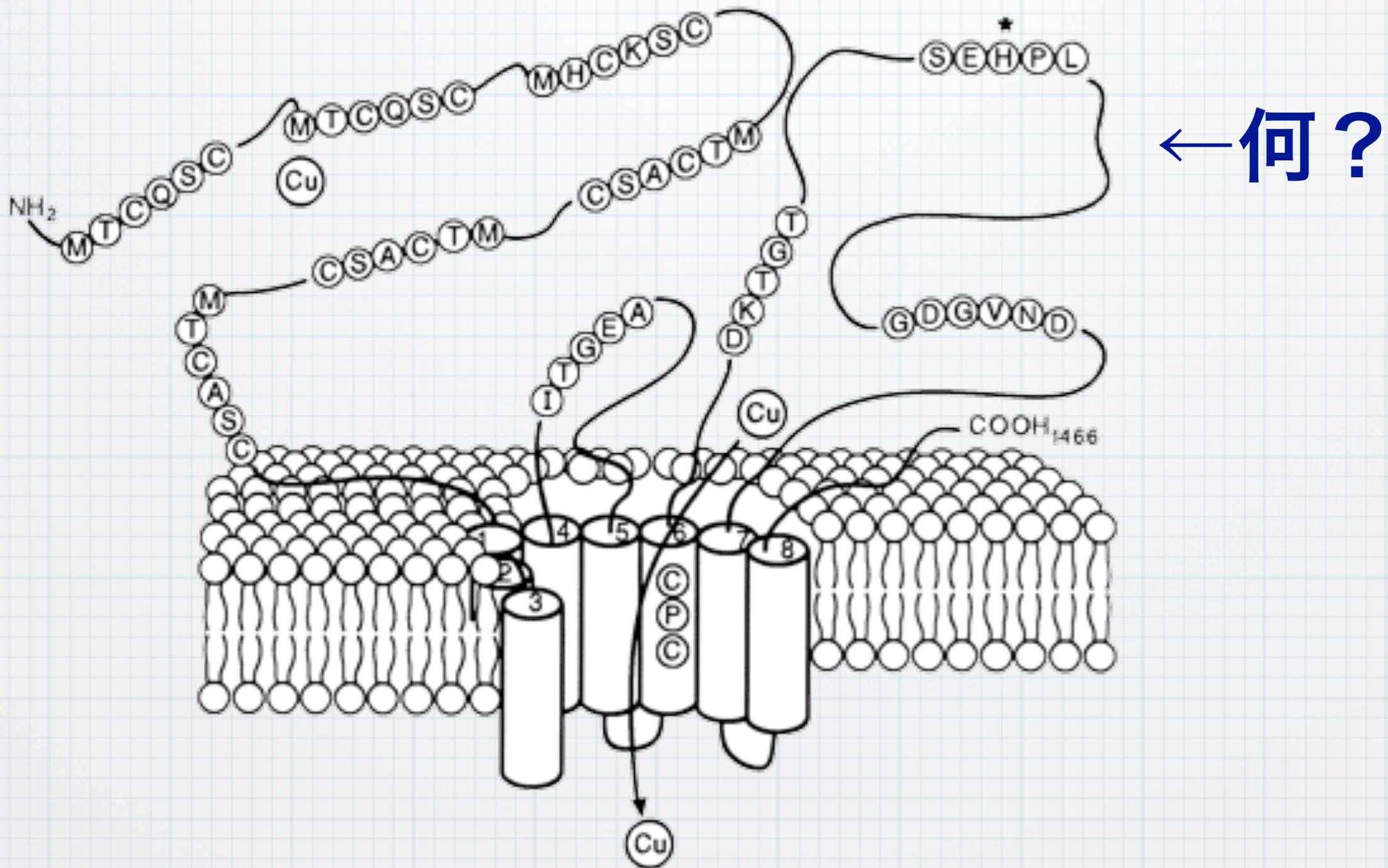
[XR\\_250764.1 - Homo sapiens \(human\) - NCBI - UCSC - RefEx\(expression\)](#)

[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 //

gatctccatgacctctatggttccttagactgcactttgagaactattgacgtacagggaaaaatgctcttgggtgtccagttaagcacagaataaaaatctggactgtgtcagtttgtaaagatcttatgtctacttggtcatcag

# 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:

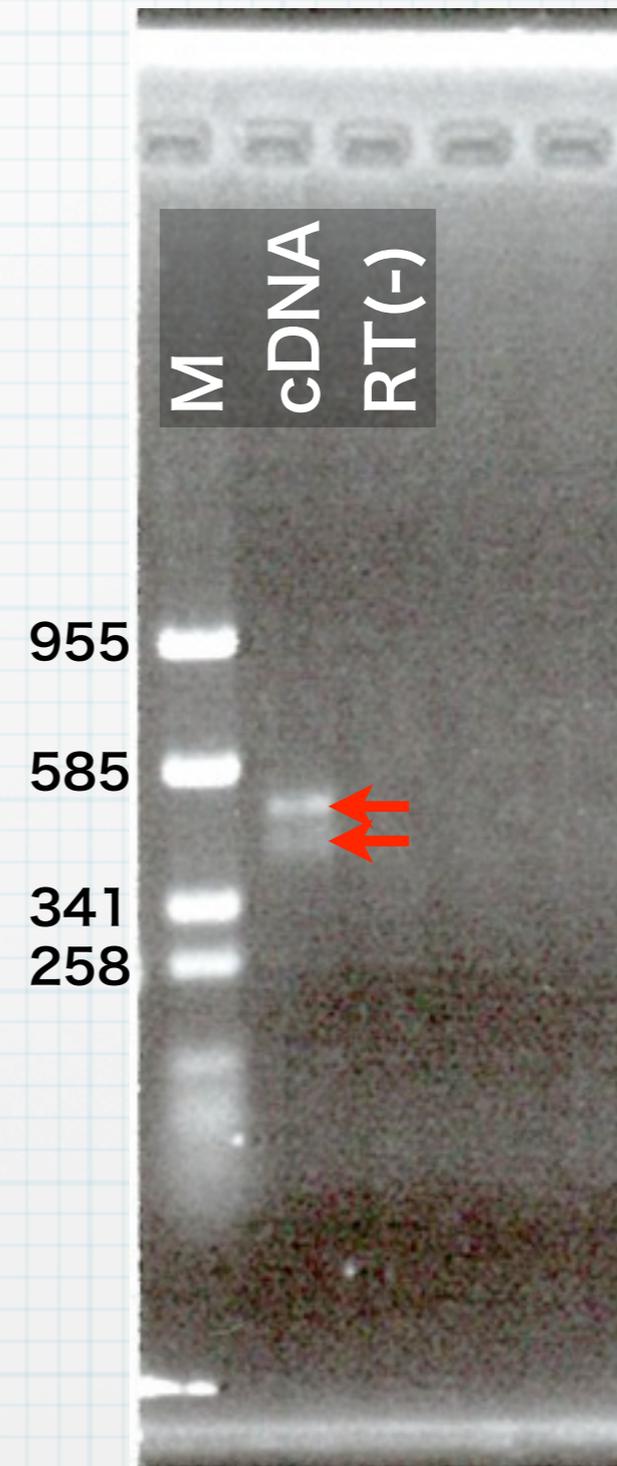
agctcattactttatcagtgca

primer-R:

tgacgtattcactcttctggtt

増幅遺伝子は何か？

予想されるバンドのサイズは？

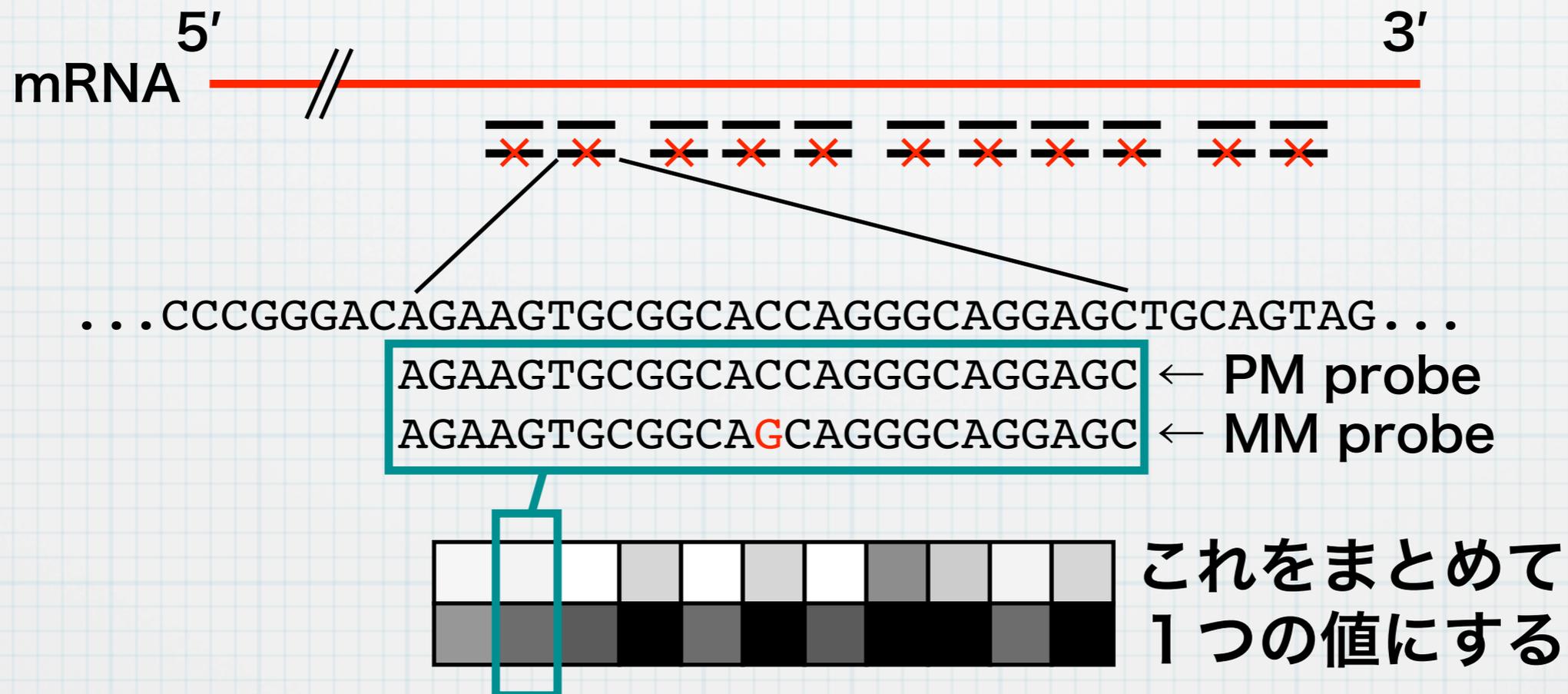


※架空のデータです

# 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



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

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

## Summary:

- [seq:TCTCCACAAACGTTTTTAAAATGTG \(2\)](#)
- [seq:GGGACACGGCAGTAAGCACAAGAAA \(2\)](#)
- [seq:GGCAGGAGCTGCAGTAGCTACCCTC \(2\)](#)
- [seq:GCATGGGATGGGACAGTCTGGGCCA \(2\)](#)
- [seq:ATGTGCCGGGTGTACTGGTGCACAC \(2\)](#)
- [seq:AGGTCACCCCATCTCTAGGCGGCAC \(2\)](#)
- [seq:AGATCACTCCCAGATCACCAGGTCA \(2\)](#)
- [seq:AGAAGTGCGGCACCAGGGCAGGAGC \(2\)](#)
- [seq:ACGGCAGTAAGCACAAGAAAGATTT \(2\)](#)
- [seq:AATGTCACCGCACACCAGGCAGTGG \(2\)](#)
- [seq:AAAATGTGCCGGGTGTACTGGTGCA \(2\)](#)
- **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)

cagccgcatcgcccaagagctctgagcaaggaggctgtcgcggggcccagaaaccgctgggactggcaagcacggctggcccagtgagcaggagggggccctgaggcatgggatgggacagtctgggccagcg  
ccacctcccgggacagaagtgcggcaccagggcaggagctgcagtagctaccctcccgtctccagcctgggctccccagatcactcccagatcaccaggtcaccatctctaggcggcaccctcacaca  
ccagtctgtggtccaacgccccgccatcaccatgatgacaccagggcagtggggacacggcagtaagcacaagaagatttttttttaagctaaaccaggccaggtgcggtggctcatgctgta  
atcccagtgcttgggaggctgaggtgggaggattgcttgagaccagcctgggtgacacagcaagacccca**ctccacaaacgttttaaatgtgccgggtgtactgggtgcacac**ctgtcatcccagctaccaag  
aagctgaggcaagaggatcacttgagcccagaaggtcgaggctgcagggagctgtgatcacactg...

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 <sup>ぐぐる</sup> <sup>な</sup> 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 げげげ GGGenome

検索

超絶高速ゲノム配列検索

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

## GGGenome

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。あまり大きいとしぼうする。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 げげげ GGGenome

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

AGCAGTACTCAAGATCACCTTTTTTTAAAGTTCACTG-CAACATT**C**AGTAGATGAAAAATGTGTAGCTGTAATTCATTGA

[chr1:106947513-106947532](#)

▼ 106947513

TAGAAAACATACCACATAAAGAAGCAATCTTTCACTG**C**ACAA-ATTGAGTATAAAATTTAACCCGTTTGAAAGGGTCATTA

[chr2:30898313-30898332](#)

▼ 30898313

GCCCCAGCCACTTATATGTATTTTTTAAATTTCACTGACAACATTGAGTAGAAAAGATAATTTTTTTTTTTTGGAGACAG

[chr2:141517099-141517116](#)

▼ 141517099

TTATTTTCTTAAACACAGCCTTTGATAAAATTTCACTGAC-ACATTGAGT-TAGCCATAGTGTGTAATGGGAGACTGCCCA

[chr2:163535086-163535103](#)

▼ 163535086

TGAGAAAGAGGACACCCTAAGAATTATGGCTTTCACTGACAACATT-AG-A GAACATATCAAACGAATATTCTGG

[chr2:186291305-186291324](#)

▼ 186291305

ATTATGAACAGAAAAAAGCAAAGGTTACTTTCACTGAG**G**CAACATTGAG-A TAGAGTTTCTTTGGACTTAATTTTGTA

# GGRNA REST API

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

# GGGenome REST API

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



# 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

chr1 (p34.2-p33) 31.3 1p31.1 1q12 32.1 1q41 q43 q44

Scale chr1: 42,000,000 43,000,000 44,000,000 45,000,000 46,000,000 hg19 47,000,000 48,000,000

GGGenome matches

UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)

The image shows a detailed view of the UCSC Genome Browser interface for a specific region of chromosome 1 (hg19 assembly). The track displays a dense collection of gene annotations, including CAP1, ZFP69, CTPS1, EDN2, GUCA2B, PPIH, U6, MPL, U6, KLF17, RNF220, UROD, TESK2, MAST2, FAAH, CYP4A11, CMPK1, SKIN, and many others. The genes are represented by vertical bars with arrows indicating their orientation. The track is color-coded by gene type and includes a scale bar at the top. The interface also shows navigation controls for moving and zooming through the genome browser.

# 高速配列検索 <sup>げげげ</sup>GGGenome

- 塩基配列の検索に特化  
(1kb 以下、短い配列が得意)
- モデル生物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 でゲゲゲノム検索

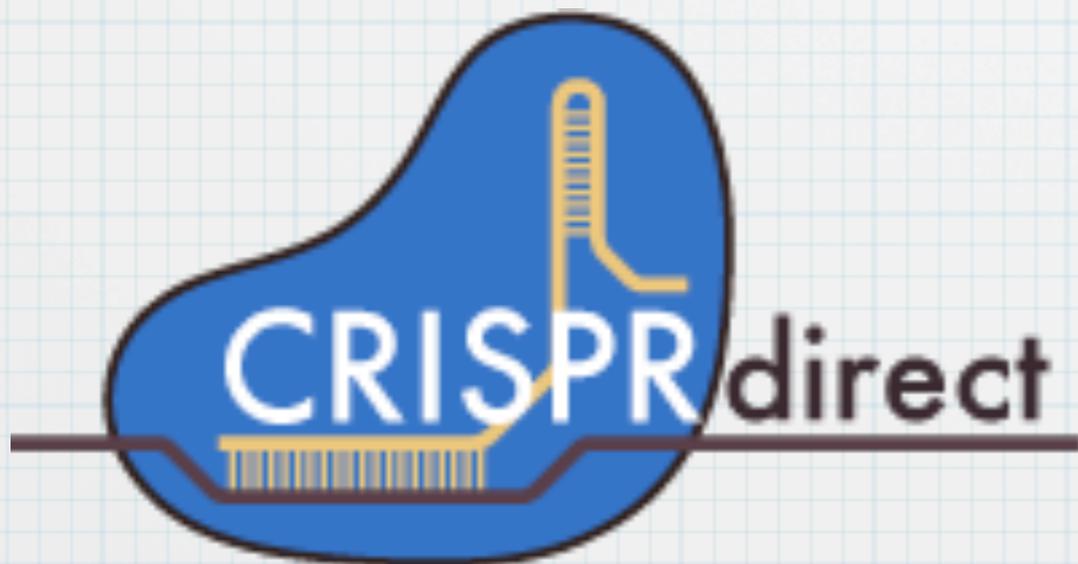
# GGRNA/GGGGenome 拡張計画

- GenBank/EMBL/DDDBJ の  
国際塩基配列データベース (INSDC)  
全体をすばやく検索したい
- 非モデル生物ゲノムもどんどん追加
- データベース更新についていけるか？
- ランキングの工夫

## 本日のテーマ#2

# 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*, in press (2014)

# CRISPRdirect



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

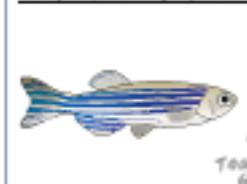
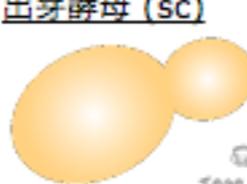
```
atgccgcgctcgtgcccagaccagagaagcaagttcgagaacgaggagtttttaggaag
ctgagccgcgagtgtagattaagtacacgggcttcagggaccggccccacgaggaacgc
caggcacgcttcagaacgcctgcccgcgacggccgctcggaaatcgcttttgtggccaca
ggaaccaatctgtctctccagtttttccggccagctggcagggagaacagcgacaaca
cctagccgagagtatgtcgacttagaaagagaagcaggcaaggtatattgaaggctccc
atgattctgaatggagctctgtgtatctggaaggctggattgatctccaaagactggat
ggtatgggctgtctggagtttgatgaggagcgagccagcaggaggatgcattagcaca
caggcctttgaagaggctcggagaaggacacgcgaatttgaagatagagacaggtctcat
cgggaggaatggaggcaagaagacaacaagaccctagtcctggtccaatttaggtggt
ggtgatgacctcaacttcgtaa
```

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 (hs)</u>  TOOO PICTURES GALLERY	<u>マウス (mm)</u>  TOOO PICTURES GALLERY	<u>ラット (rn)</u>  TOOO PICTURES GALLERY	<u>ニワトリ (gg)</u>  TOOO PICTURES GALLERY
<u>ツメガエル (xt)</u>  TOOO PICTURES GALLERY	<u>ゼブラフィッシュ (dr)</u>  TOOO PICTURES GALLERY	<u>ホヤ (ci)</u>  TOOO PICTURES GALLERY	<u>ショウジョウバエ (dm)</u>  TOOO PICTURES GALLERY
<u>線虫 (ce)</u>  TOOO PICTURES GALLERY	<u>シロイヌナズナ (at)</u>  TOOO PICTURES GALLERY	<u>イネ (os)</u>  TOOO PICTURES GALLERY	
<u>出芽酵母 (sc)</u>  TOOO PICTURES GALLERY	<u>分裂酵母 (sp)</u>  TOOO PICTURES 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 ?		
start - end	+ -	20mer+PAM (total 23mer)	GC% of 20mer	Tm of 20mer	TTTT in 20mer	20mer +PAM	12mer +PAM	8mer +PAM
4 - 26	-	ccg cgcgtcgtgcccgaccagag [gRNA]	75.00 %	82.34 °C	-	1 [detail]	3 [detail]	103 [detail]
16 - 38	-	ccc gaccagagaagcaagttcga [gRNA]	50.00 %	71.17 °C	-	1 [detail]	26 [detail]	4354 [detail]
17 - 39	-	ccg accagagaagcaagttcgag [gRNA]	50.00 %	71.53 °C	-	1 [detail]	32 [detail]	5910 [detail]
21 - 43	-	cca gagaagcaagttcgagaacg [gRNA]	50.00 %	69.25 °C	-	1 [detail]	24 [detail]	8744 [detail]
24 - 46	+	gagaagcaagttcgagaacg 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 cgaggaacgccaggcacgct [gRNA]	70.00 %	82.52 °C	-	1 [detail]	2 [detail]	1011 [detail]
120 - 142	-	cca ggcacgcttcagaacgcct [gRNA]	65.00 %	80.61 °C	-	1 [detail]	6 [detail]	1510 [detail]
130 - 152	+	ttccagaacgcctgccgcga cgg [gRNA]	65.00 %	80.58 °C	-	1 [detail]	4 [detail]	289 [detail]
132 - 154	-	cca gaacgcctgccgcgacggcc [gRNA]	80.00 %	86.25 °C	-	1 [detail]	2 [detail]	937 [detail]
138 - 160	+	cgctgccgcgacggccgct cgg [gRNA]	85.00 %	90.21 °C	-	1 [detail]	5 [detail]	815 [detail]
140 - 162	-	cct gccgcgacggccgctcggaa [gRNA]	80.00 %	86.71 °C	-	1 [detail]	1 [detail]	1044 [detail]
144 - 166	-	ccg cgacggccgctcggaaatcg [gRNA]	70.00 %	79.76 °C	-	1 [detail]	16 [detail]	628 [detail]
153 - 175	+	ccgctcggaaatcgcttttg tgg [gRNA]	55.00 %	72.84 °C	+	0 [detail]	7 [detail]	801 [detail]
153 - 175	-	ccg ctcggaaatcgcttttg tgg [gRNA]	50.00 %	70.06 °C	-	0 [detail]	1 [detail]	805 [detail]
160 - 182	+	aaatcactttttataaccac aga [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*, in press (2014)

***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

Is | this | a | pen? | ! | ? | ? | !

- 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



# 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



← 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