



STNNext®

# GENESEQ ファイル 配列検索

JAICI

# Contents

概要 .....	1
回答表示形式 .....	4
GENESEQ ファイルの配列検索 .....	5
ホモロジー検索 .....	6
参考:コンパクトに表示したい場合はカスタム表示形式を使う .....	13
参考:必要なフィールドを抽出したい場合は ANALYZE (または SELECT) コマンドを使う.....	13
参考:WPI ファイルにクロスオーバーしたい場合は OS フィールドを使う.....	13
完全配列検索・部分配列検索.....	14
塩基コード・アミノ酸コード.....	17
その他の主な検索フィールド.....	19

## 概要

- GENESEQ ファイルは Clarivate Analytics が作成する特許の配列データベースである。

(2022 年 2 月)

製作者	Clarivate Analytics
収録源	WPI ファイルのベーシック特許 (ベーシック特許とは、一つの特許ファミリーのうち Clarivate Analytics が最初に入手した特許)
特長	<ul style="list-style-type: none"><li>・ 特許中の核酸、タンパク質の配列情報を収録する配列専用のファイル</li><li>・ GENESEQ 独自の抄録を作成しているため、配列に関する説明情報が豊富。</li><li>・ 2 種類のホモロジー検索 (BLAST と GETSIM) がある。</li></ul>
レコード単位	配列単位 (同一特許から複数のレコードが作成される場合がある。)
収録件数	核酸 : 40,041,000 件 タンパク質 : 17,337,000 件 合計 : 57,378,000 件
更新頻度	毎週
タイムラグ	約 2 ヶ月以内

- 収録期間・収録基準

	核酸	タンパク質
収録期間	1981 年以降	
収録基準	<ul style="list-style-type: none"><li>・ 塩基数が 10 以上の配列 (プローブ、プライマーに関しては制限なし)</li></ul>	<ul style="list-style-type: none"><li>・ アミノ酸数が 4 以上の配列</li></ul>
	<ul style="list-style-type: none"><li>・ 特許由来の配列<ul style="list-style-type: none"><li>- 1981 年～1994 年 : クレーム中の配列 (クレーム中に記載されている配列は、新規でなくても収録)</li><li>- 1995 年以降 : 実施例、発明の詳細な説明などクレーム外からも収録</li></ul></li></ul>	

- 登録ルール

- ・ 変異した配列やキメラ配列は、新規でなくとも収録される。
- ・ 配列リスト中の配列はすべて収録される。  
(類似性を比較する目的で記載されている配列は収録されない)
- ・ 短い配列・フラグメントを含む長い配列が収録されているが、下記の場合は、例外として、そのフラグメントを別のレコードとして収録することがある。
  - フラグメントに SEQ ID 番号が付与されている場合。
  - 配列のフラグメントがクレームされているが、対応する全長配列がクレームされていない場合。
- ・ 同一の配列でも別の特許に記載されている場合は、別レコードになる
- ・ 「過去の特許から引用している」などの記載がある場合は収録されない。

■ 核酸のレコード例 (ALL 表示形式)

レコード番号 AN BEE04366 GENESEQ ED 20211030 UP 20211030  
 入力日 DED 20170907 Full-text  
 WPI の標題 TI Assessing the risk of developing pneumonia comprises determining the presence of PNEUMONIA Risk Allele(s) in a test sample.

発明者 IN Blein S; Mira J; Pachot A  
 特許出願人 PA BIOMERIEUX SA (INMR)  
 使用言語 LA English  
 資料種類 DT Patent  
 特許情報 PI WO 2017121769 A1 20170720 ● ベーシック特許の特許番号のみ収録されている  
 特許情報のタイプ PIT WOA1 INTERNATIONAL APPLICATION PUBL H REPORT  
 出願情報 AI WO 2017-EP50505 20170111  
 優先権出願情報 PRAI EP 2016-305022 20160112  
 ファイルセグメント FS NUCLEIC; NS  
 クロスリファレンス CR <https://www.ncbi.nlm.nih.gov/snp/?term=rs3004169>  
 WPI のレコード番号 OS 2017-49641D [52]  
 分子タイプ\* MTY DNA

配列情報の記載位置 PSL Claim 16; SEQ ID NO 92; 50pp  
 配列の説明 DESC Human derived gene fragment, SEQ ID 92.  
 キーワード KW SNP; antimicrobial-gen.; ds; pneumonia; prophylactic to disease; pseudomonas aeruginosa infection; respiratory-gen.; single nucleotide polymorphism; snp detection; therapeutic.

生物名 ORGN Homo sapiens.  
 配列の独自の抄録 AB The present invention relates to an in vitro method for assessing the risk of developing pneumonia in a subject. The method involves: determining the presence of pneumonia risk alleles in a sample from the subject, where the presence of the pneumonia risk alleles indicates whether the subject is at increased 配列独自の抄録 risk of developing pneumonia. The risk allele is single nucleotide polymorphism (SNP). pneumonia is Pseudomonas aeruginosa pneumonia. The invention further relates to a kit for performing the method defined above. The invention is also useful for treating or preventing pneumonia in the subject. The present sequence represents a human SNP containing DNA, useful in the method for assessing the risk of developing pneumonia in the subject.

配列番号 SEQN 92  
 配列長 SQL 51  
 配列キー SEQK ecd370d7ede623f9578ad0d246a092f37573054ba8eb152a232d40b1383a86c4  
 配列 SEQ  
 1 ctgaggatct gactgccagt caacartggg ttcaactcctg tgaggaagac  
 51 c

核酸の種類 NA

Code	Count	Percent
A	12	23.5
C	13	25.5
G	14	27.5
T	11	21.6
U	0	0.0
Other	1	2.0

特徴表 FEATURE TABLE: ● 配列の特徴で限定 (/FEAT)

Key	Location	Qualifier
variation	26	*tag= a
		standard_name  "Single nucleotide polymorphism"

\* 分子タイプ (/MTY)  
 DNA, PEPTIDE 以外にも CDNA, RNA, MRNA, PROTEIN などの種類がある。  
 分子タイプの種類は オンライン中に EXPAND で確認できる。

書誌情報 (BIB)

抄録 (ABS)

配列情報 (SQIDE)

■ タンパク質のレコード例 (ALL 表示形式)

AN BED56939 GENESEQ ED 20211030 UP 20211030  
 DED 20170907 Full-text  
 TI Composition used to prevent or treat e.g. tumor, comprises a synthetic peptide comprising an amino acid sequence i.e. homologous to specified number of consecutive amino acids of a portion of cytoplasmic domain of alpha-v beta integrin.  
 IN Byzova T; Mahabaleshwar G H; Feng W  
 PA CLEVELAND CLINIC FOUND (CLCL)  
 LA English  
 DT Patent  
 PI US 20170198023 A1 20170713  
 PIT USA1 FIRST PUBLISHED PATENT APPLICATION [FROM 2001 ONWARDS]  
 AI US 2017-456986 20170313  
 PRAI US 2011-61570433 20111214 (61)  
 US 2012-715132 20121214  
 US 2015-937289 20151110  
 FS PROTEIN: PS  
 OS 2017-478454 [48]  
 MTY peptide  
 PSL Disclosure; SEQ ID NO 22; 57ppDESC Human VEGFR2 peptide, SEQ ID 22.  
 KW VEGFR2 protein; age related macular degeneration; angiogenesis disorder; antiarteriosclerotic; antiarthritic; antiinflammatory; antilipemic;

書誌情報 (BIB)

ORGN Homo sapiens.  
 AB The present invention relates to a novel pharmaceutical composition, useful for preventing or treating tumor, angiogenesis, arthritis, rheumatoid arthritis, psoriasis, atherosclerosis, diabetic retinopathy, age-related macular degeneration, ~~Arterio disease~~ vascular restenosis, meningioma, hemangioma and neovasc. **配列独自の抄録** the pharmaceutical composition comprises a synthetic peptide which is homologous to specified number of consecutive amino acids of a portion of cytoplasmic domain of alpha-v beta integrin that includes a tyrosine residue, where the amino acid of the peptide includes an amino acid that structurally mimics a phosphorylated tyrosine residue that substitutes a corresponding

抄録 (ABS)

SEQN 22  
 SQL 20  
 SEQK 02abf2a96ac3cfd3f954c3cc4b6c2b795f04eab12f81d613c57c1ed1a55a9801

配列情報 (SQIDE)

SEQ 1 lgaspepgvk ideefcrrlk

アミノ酸の種類

Code	Count	SQ3IDE 表示形式の場合 : アミノ酸が 3 文字コードで表示される SEQ3
A	1	1 Leu-Gly-Ala-Ser-Pro-Glu-Pro-Gly-Val-Lys-
R	2	11 Ile-Asp-Glu-Glu-Phe-Cys-Arg-Arg-Leu-Lys
N	0	0.0
D	1	5.0
B	0	0.0
C	1	5.0
Q	0	0.0
E	3	15.0
Z	0	0.0
G	2	10.0
H	0	0.0
:		
V	1	5.0
Others	0	0.0

FEATURE TABLE:

Key	Location	Qualifier
Modified-site 6	note	"Gamma-carboxyglutamic acid"

## 回答表示形式

### ■ 表示形式 (    は利用頻度の高い表示形式)

- ・ TRIAL 表示形式では、配列長を確認することは出来るが、配列は表示されない。
- ・ ALIGN 表示形式は、完全配列・部分配列検索ではヒットした配列を表示し、ホモロジー検索ではアライメントを表示する。
- ・ 書誌情報、抄録、配列情報を確認するには ALL 表示形式を利用する。

表示形式		内容
一般	BIB IBIB* <sup>1</sup>	書誌情報 (AN, ED, UP, DED, DUPD, TI, PA, PACO, LA, DT, PI, PIT, AI, PRAI, FS, CR, OS, MTY, PSL, DESC)(デフォルト)
	SCAN	ED, UP, DED, DUPD, TI, MTY, DESC * 回答番号なしのランダム表示
	TRIAL	レコード番号 (AN), 標題 (TI), 分子タイプ (MTY), 配列の説明 (DESC), キーワード (KW), 配列長 (SQL)
	ABS	レコード番号 (AN), 抄録 (AB)
	ALL IALL* <sup>1</sup>	書誌情報, 抄録, 配列情報およびキーワード (KW)
	PSL	特許中の配列情報の記載位置
	APPS	出願情報 (AI), RLI
配列関連	SQIDE	レコード番号 (AN), 分子タイプ (MTY) と配列情報 (ORGN, SEQC, SEQN, SQL, SEQK, SEQ (1 文字コード), AA または NA, FEAT)
	SQ3IDE	レコード番号 (AN), 分子タイプ (MTY) と配列情報 (ORGN, SEQC, SEQN, SQL, SEQK, SEQ (3 文字コード), AA または NA, FEAT)
	ALIGN* <sup>2</sup>	(完全配列, 部分配列検索結果の場合) 配列とヒット情報 (ホモロジー検索結果の場合) ヒットした配列と質問配列のアライメント情報 (一致 / 不一致 / ギャップ)
	ALIGNG* <sup>2</sup>	アライメントを画像データで表示 (表示内容は ALIGN と同じ)
	FASTA	FASTA フォーマット
	SCORE	相同性スコア
ほか	HIT	ヒットタームを含むフィールド
	KWIC* <sup>3</sup>	ヒットタームの前後 20 語
	OCC	ヒットしたフィールド名とヒットターム数

\*1 完全なフィールド名が表示されるインデント型表示形式

\*2 オンラインでのみ有効。PRINT コマンドで指定しても入手できない。

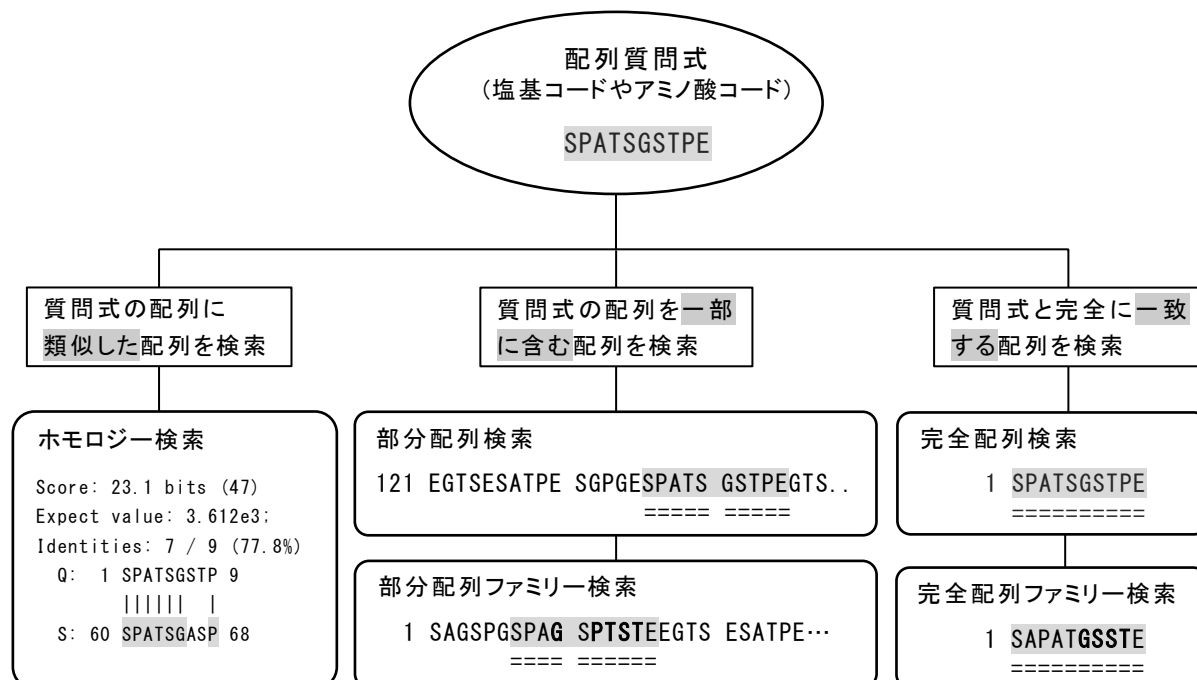
\*3 配列情報を表示する場合、KWIC 表示形式は HIT 表示形式と同じ結果になる。

### ■ 基本的な回答表示の方法

- ・ => D TRIAL ALIGN で回答の適合性を確認したのち、完全・部分配列検索では => D ALL, ホモロジー検索では => D ALL ALIGN で書誌情報, 抄録, 配列情報を表示する。

## GENESEQ ファイルの配列検索

- GENESEQ ファイルでは、2 種類のホモロジー検索 (BLAST, GETSIM) と完全配列検索, 部分配列検索を行うことができる。



### ■ ホモロジー検索プログラムの違い

プログラム	BLAST	GETSIM
概要	Basic Local Alignment Search Tool 最もよく利用されている配列検索プログラム。他のプログラムに比べて高速処理できる。ギャップをあまり考慮しないため、検出感度や選択性が低いと考えられがちだが、実際には他と比べてそれほど遜色はない。	FASTA 系列のプログラム。 データベース中のすべての配列との間で忠実にアラインメントを行ってホモロジースコアを算定する。BLAST ホモロジー検索で回答が得られない場合でも、GETSIM ホモロジー検索で回答が得られることがある。
処理速度	速い	遅い*
類似性の高い配列	○	○
類似性の低い配列	△	○
比較方法	短い部分配列を比較 ギャップはあまり考慮されない	配列全体を比較 ギャップも考慮される
感受性	デフォルトの設定では低い	高い

## ホモロジー検索

### ■ ホモロジー検索の入力方法

=> RUN BLAST コード/検索フィールド パラメータ

=> RUN GETSIM コード/検索フィールド パラメータ

- ・ コードは 5 つ以上を入力する（ギャップ記号, 特殊記号, アミノ酸の 3 文字コードは利用できない）。
- ・ 配列質問式の長さの制限値は検索タイプによって異なる。=> ? QLIMIT で確認する。
- ・ 回答数の上限のデフォルトは 15,000 件。パラメータで最大 10 万件に変更できる。

### ■ ホモロジー検索の検索フィールドと入力例

- ・ BLAST の場合は下表の全検索フィールドを使用可能。GETSIM の場合は, /SQN, /TSQN, /SQP の 3 タイプのみ使用できる。

例 1: フィルタリングを行わず, 入力したコードの鎖のみを検索する (相補鎖を含まない)。

=> RUN BLAST GCTCCCAGAATGCCAGAGGCTGC/SQN -F F -S SIN

例 2: フィルタリングを行わず, 回答上限を 10 万件に変更して検索する。

=> RUN BLAST L1/SQN -F F -MAXSEQ 100000

例 3: 短い配列質問式 (30 配列未満) で検索するためパラメータを変更して検索する。

=> RUN BLAST L1/SQN -F F -E 1000 -W 7

=> RUN BLAST L1/SQP -F F -M PAM30 -E 20000 -W 2

検索フィールド	内容	質問式	回答
/SQN	塩基配列の質問式に類似した塩基配列を検索	塩基配列	塩基配列
/TSQN	データベース中の塩基配列をアミノ酸に翻訳した配列の中からアミノ酸配列の質問式に類似した配列を検索	アミノ酸配列	塩基配列
/SQP	アミノ酸配列の質問式に類似したアミノ酸配列を検索	アミノ酸配列	アミノ酸配列
/SQM	非常に類似した (種内などの) 配列用に最適化された BLASTn (megaBLAST)	塩基配列	塩基配列
/SQDM	一部の塩基を無視し (多少のミスマッチを許容し), より離れた (種間などの) 配列を検索するために最適化された BLASTn (discontiguous megaBLAST)	塩基配列	塩基配列
/TSQP	塩基配列の質問式をアミノ酸配列に翻訳してこれに類似したアミノ酸配列を検索 (BLASTx)	塩基配列	アミノ酸配列
/TSQNX	塩基配列の質問式をアミノ酸配列に翻訳してこれに類似したアミノ酸配列に翻訳された塩基配列を検索 (tBLASTx)	塩基配列	塩基配列



■ ホモロジー検索のパラメータ

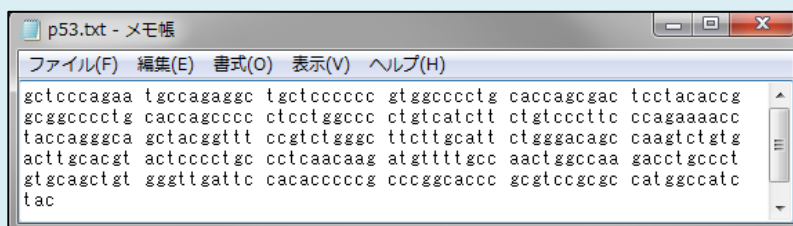
- ・ デフォルトでは [フィルタリングを行う][回答上限 15,000 件] などデフォルトのパラメータに設定されている。変更する場合は、RUN コマンドと同じ行に入力する。

パラメータ	入力形式	指定できる数値・項目
回答数の上限	-maxseq 数値	<数値> ・ 101 (指定できる最小値) ・ 15000 (デフォルト) ・ 100000 (指定できる最大値)
フィルタ (Filter)	-f 項目	配列データのフィルタリングの指定 <項目> ・ T: フィルタリングを行う (デフォルト) ・ F: フィルタリングを行わない
期待値 (Expectation Value)	-e 数値	データベース中の配列に対してマッチする際の統計的有意性の閾値のこと。0 より大きい数値を入力する。期待値が 0 に近づくほど類似性が高くなり回答件数は少なくなる。デフォルトは 10 短い配列質問式で検索する場合は、期待値を大きくする。
Word Size	-w 数値	<数値> ・ 塩基酸配列の質問式使用時は 7-23 (デフォルトは 11) ・ アミノ酸配列の質問式使用時は 2-3 (デフォルトは 3)
核酸検索時の相補鎖検索の有無 (Strand)	-s 項目	<項目> ・ SIN : 入力した配列コードを検索する。 ・ COM : 入力した配列コードの相補鎖を検索する。 ・ BOTH : 入力した配列コードと相補鎖を検索する (デフォルト)。
	-s 数値	<数値>    ・ 1 : single                    ・ 2 : complementary ・ 3 : both (デフォルト)
置換行列 (Matrix)	-m 項目	スコア値を計算する置換行列の指定 (アミノ酸配列質問式の使用時) <BLAST> ・ BLOSUM45            ・ BLOSUM50            ・ BLOSUM62 (デフォルト) ・ BLOSUM80            ・ BLOSUM90 ・ PAM30                ・ PAM70                ・ PAM250 <GETSIM> ・ BL50 (デフォルト) ・ BL62                    ・ BL80 ・ MD10                ・ MD20                ・ MD40 ・ OPT5                ・ P120 ・ P250                ・ VT160
Gap Penalty	-g 数値	・ BLAST のデフォルトは、核酸 5, タンパク質 11 ・ GETSIM のデフォルトは、核酸 12, タンパク質 12
Gap Extension	-x 数値	・ BLAST のデフォルトは核酸 2, タンパク質 1 ・ GETSIM のデフォルトは核酸 4, タンパク質 2
Penalty for Nucleotide mismatch	-q 数値	・ BLAST のデフォルトは -3 ・ GETSIM のデフォルトは -2
Reward for nucleotide match	-r 数値	・ BLAST のデフォルトは 1 ・ GETSIM のデフォルトは 3

\* パラメータの詳細はオンラインヘルプ (=> [HELP 1OPTIONS](#) と => [HELP 2OPTIONS](#)) で確認できる。

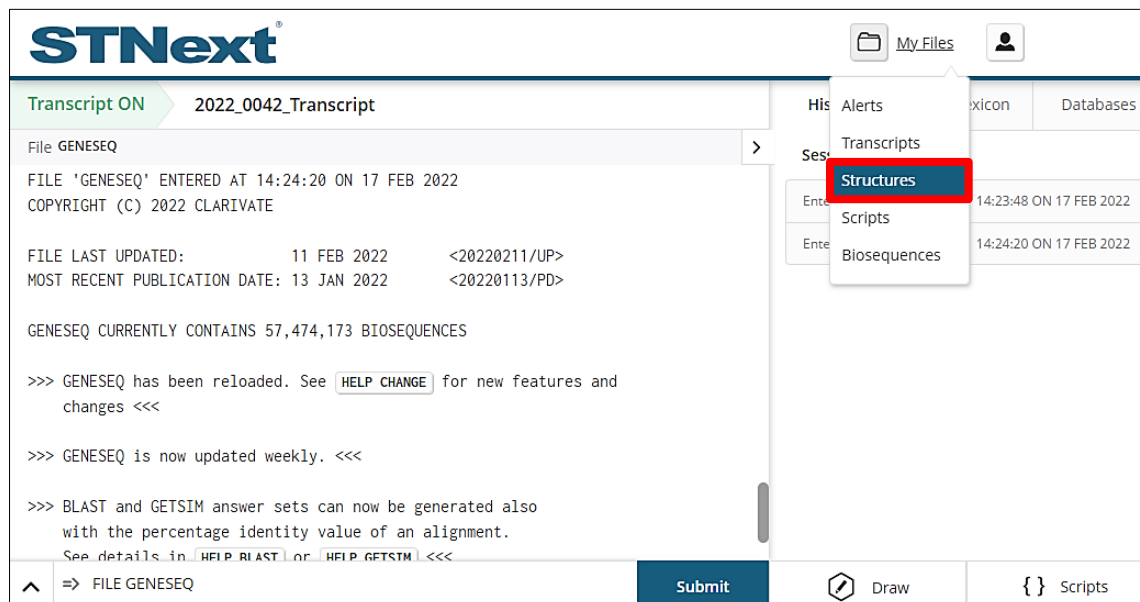
■ 検索例: 癌抑制遺伝子 p53 に類似する核酸の検索

事前準備: 配列質問式のテキストファイル (.txt) を作成する。

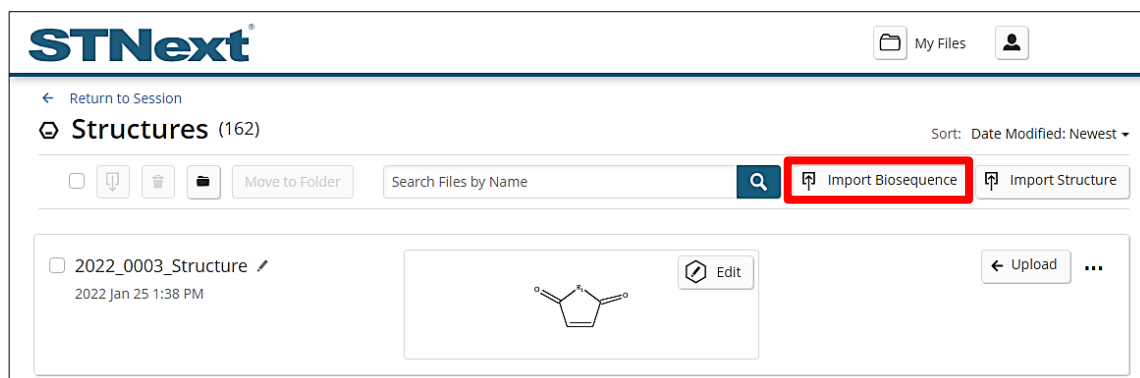


ログイン～配列質問式のアップロード

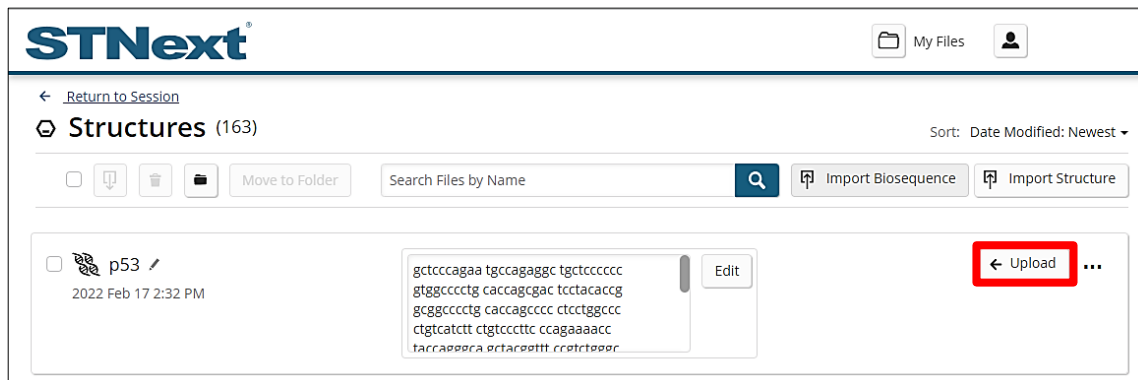
- GENESEQ ファイルに入り, 画面右上の My Files のプルダウンメニューから Structures を選択する。



- 「Import Biosequence」をクリックし, 事前に作成した配列質問式のテキストファイルを呼び出して, OK をクリックする。



- ・ 配列質問式がインポートされたら, Upload をクリックする.



- ・ アップロードした配列質問式の L 番号が表示される. この L 番号を用いて配列検索を行う.



## 配列検索の実行

=> `RUN BLAST L1/SQN -F F` ← フィルタをはずして BLAST ホモロジー検索を実行する

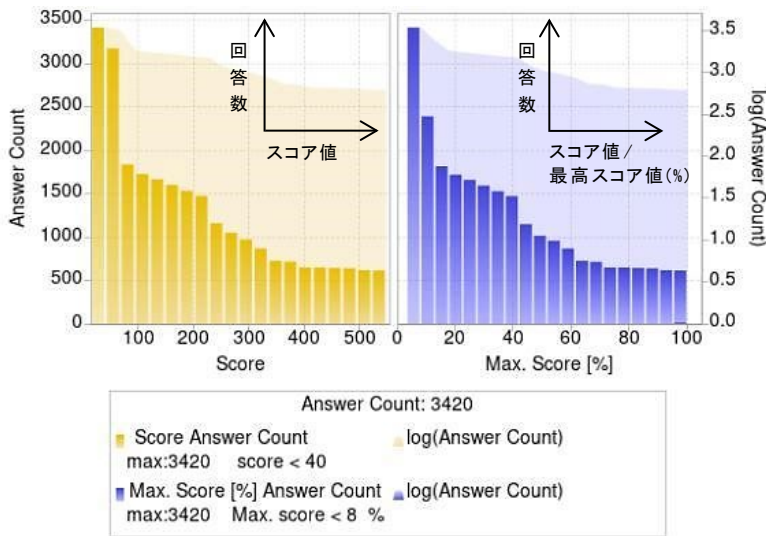
Algorithm: BLAST - BLASTN. Version: 2.12.0+

The BLAST software is used herein with permission of the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM). See also: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.

Database: GENESEQ  
 Posted Date: Feb 17, 2022 02:00 AM CET  
 Number of Hits to DB: 3420  
 Expectation Value: 10.0  
 Number of Sequences: 40117120  
 Length of query: 303  
 Length of database: 22553464291  
 Search space: 6833699680173  
 Effective search space: 5829239206577  
 Lambda: 0.625  
 Kappa: 0.41  
 Entropy: 0.78

Highest possible score value: 547.7  
 Best answer score value: 547.7 ← 最も類似度が高い回答のスコア値

3420 ANSWERS FOUND BELOW EXPECTATION VALUE OF: 10.0  
GENESEQ



Query time: 681  
 ENTER EITHER "ALL" TO KEEP ALL ANSWERS  
 OR ENTER THE MINIMUM SCORE VALUE YOU WISH TO KEEP  
 OR ENTER THE MINIMUM PERCENT OF SCORE FOLLOWED BY "% SCORE"  
 OR ENTER THE MINIMUM PERCENT OF IDENTITY FOLLOWED BY "% IDENT"  
 OR COMBINE MINIMUM PERCENT OF SCORE AND IDENTITY AS "X% SCORE Y% IDENT"  
 OR ENTER "END". "END" MUST BE ENTERED TO COMPLETE THE RUN COMMAND.  
 ENTER (ALL) OR ?:ALL

回答集合のオプションを指定する (ALL: 全件)

```
L2 RUN STATEMENT CREATED
L2 3420 GCTCCCAGAATGCCA
      TCCTACACGGGCGGC
      CTGTCCCTTCCCAGA
      TTCTTGCAATTCTGGG
      CCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGT
      GGGTTGATTCCACACCCCGCCGGCAGCCGGCTCCGGCCATGGCCATC
      TAC/SQN -F F
```

- 回答全件 (ALL と入力)
- スコア値の最低値 (数字を入力)
- スコア値パーセント (スコア値/最高スコア値) の最低値
- 同一性パーセント (一致コード数/Alignment コード数) の最低値
- スコア値パーセントの最低値と同一性パーセントの最低値

ENTER EITHER "ALL" TO KEEP ALL ANSWERS  
 OR ENTER THE MINIMUM SCORE VALUE YOU WISH TO KEEP  
 OR ENTER THE MINIMUM PERCENT OF SCORE FOLLOWED BY "% SCORE"  
 OR ENTER THE MINIMUM PERCENT OF IDENTITY FOLLOWED BY "% IDENT"  
 OR COMBINE MINIMUM PERCENT OF SCORE AND IDENTITY AS "X% SCORE Y% IDENT"  
 OR ENTER "END". "END" MUST BE ENTERED TO COMPLETE THE RUN COMMAND.  
 ENTER (ALL) OR ?:END ← 回答集合のオプションは複数回指定できる  
 終了する場合は END を入力する

### (任意) 回答集合の保存

=> SAVE L2 P53/A TEMP ← 配列検索結果を保存する場合は回答を並べ替える前の L 番号を指定する  
 ANSWER SET L2 HAS BEEN SAVED AS 'P53/A'

### (任意) 回答の並べ替え

=> SORT L2 1- SCORE D IDENT D ← スコア値 (SCORE) の降順 (D) かつ同一性 (IDENT) の降順 (D)  
 で回答を並べ替える

```
PROCESSING COMPLETED FOR L2
L3 3420 SORT L2 1- SCORE D IDENT D
```

回答表示

=> D L3 1 3000 3085 BIB ALIGN ← 書誌情報とアライメントを表示する

L3 ANSWER 1 OF 3420 GENESEQ COPYRIGHT 2022 CLARIVATE on STN.
AN AYM36275 GENESEQ ED 20211030 UP 20211030
DED 20110120 Full-text
TI Evaluating a patient with acute lymphoblastic leukemia (ALL) that is characterized by the presence of Philadelphia chromosome comprises generating an expression profile of ALL biomarkers from a test biological sample.
IN Zuo Z; Luthra R
PA UNIV TEXAS SYSTEM (TEXA)
LA English
DT Patent
PI WO 2010138843 A2 20101202
PIT WOA2 INTERNATIONAL APPLICATION PUBLISHED WITHOUT INTERNATIONAL SEARCH REPORT or INTERNATIONAL APPLICATION PUBLISHED WITH DECLARATION UNDER ARTICLE 17 (2) (A) [FROM 20090101 ONWARDS]
AI WO 2010-US36623 20100528
PRAI US 2009-182228P 20090529
FS NUCLEIC; NS
CR http://www.ncbi.nlm.nih.gov/protein/339815; http://www.ncbi.nlm.nih.gov/protein/339816
OS 2010-P75161 [82]
MTY cDNA
PSL Disclosure; SEQ ID NO 101; 43pp
DESC Acute lymphoblastic leukemia prognosis determining DNA marker, SEQ 101.
ALIGN

スコア値が高い(類似性が高い)回答

Query Length: 303; Sequence Length: 1303; ← 配列質問式の長さ; 回答配列の配列長
Score: 547.7 bits (606), 100.0% of highest possible score 547.7; ← スコア値
Expect value: 7.751e-153; ← 期待値
Identities: 303 / 303 (100.0%); ← 同一性パーセント

Strand: Plus / Plus; Alignment Length: 303;
Q: 1 GCTCCCAGAATGCCAGAGGCTGCTCCCCCGTGGCCCTGCACCAGCGACTCCTACACCG 60
S: 308 GCTCCCAGAATGCCAGAGGCTGCTCCCCCGTGGCCCTGCACCAGCGACTCCTACACCG 367
Q: 61 GCGGCCCTGCACCAGCCCTCCTGGCCCTGTCATCTTCTGTCCCTCCAGAAAACC 120
S: 368 GCGGCCCTGCACCAGCCCTCCTGGCCCTGTCATCTTCTGTCCCTCCAGAAAACC 427
Q: 121 TACCAGGGCAGCTACGGTTTCCGTCTGGGCTTCTTGCACTTCTGGGACAGCCAAGTCTGTG 180
S: 428 TACCAGGGCAGCTACGGTTTCCGTCTGGGCTTCTTGCACTTCTGGGACAGCCAAGTCTGTG 487
Q: 181 ACTTGCACGTA CTGCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCT 240
S: 488 ACTTGCACGTA CTCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCT 547
Q: 241 GTGCAGCTGTGGGTTGATTCCACACCCCGCCGGCACC CGGTCCGGCCATGGCCATC 300
S: 548 GTGCAGCTGTGGGTTGATTCCACACCCCGCCGGCACC CGGTCCGGCCATGGCCATC 607
Q: 301 TAC 303
S: 608 TAC 610

全配列中の位置

アライメント表示
Q (配列質問式: Query)
S (回答の配列: Subject)
| は一致しているコード

L3 ANSWER 3000 OF 3420 GENESEQ COPYRIGHT 2022 CLARI  
AN ACA61112 GENESEQ ED 20211030 UP 20211030  
TI New nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide, useful as a component of an adenovector or DNA plasmid vaccine for preventing or treating hepatitis C virus.  
IN Emini EA; Kaslow DC; Bett AJ; Shiver JW; Nicosia A; Lahm A; Luzzago A; Cortese R; Colloca S  
PA MERCK & CO INC (MERI)  
IST RICERCHE BIOL MOLECOLARE ANGELETTI (RICE-N)  
LA English  
DT Patent  
PI WO 2003031588 A2 20030417  
PIT WOA2 INTERNATIONAL APPLICATION PUBLISHED WITHOUT INTERNATIONAL SEARCH REPORT or INTERNATIONAL APPLICATION PUBLISHED WITH DECLARATION UNDER ARTICLE 17 (2) (A) [FROM 20090101 ONWARDS]  
AI WO 2002-US32512 20021010  
PRAI US 2001-328655P 20011011  
US 2002-363774P 20020313  
FS NUCLEIC; NS  
CR ABU09574  
OS 2003-381708 [36]  
MTY cDNA  
PSL Claim 12; Fig 3; 231pp  
DESC HCV cDNA encoding Met-NS3-NS4A-NS4B-NS5A-NS5B (inactivated), humanised.

スコア値が低い (類似性が低い) 回答

ALIGN

Query Length: 303; Sequence Length: 5965;  
Score: 41.9 bits (45) , 7.6% of highest possible score 547.7;  
Expect value: 1.458;  
Identities: 40 / 51 (78.4%);  
Strand: Plus / Plus; Alignment Length: 51;  
Q: 33 GGCCCTGCACCGACTCCTACACCGGCGCCCTGCA---CCAGCCCC 80  
||| ||| | ||| | || ||||| ||||| |||||  
S: 3103 GGCACCTTCCCATCAACGCCTACACCGGCCCCTGCACCCCAGCCCC 3153

L3 ANSWER 3085 OF 3420 GENESEQ COPYRIGHT 2022 CLARI  
AN AAV21525 GENESEQ ED 20211030 UP 20211030  
TI Human breast cancer related genes - used for diagnosis and therapy of breast and ovarian cancers.  
IN Bergmann JE; Preddie ER  
PA BERGMANN J E (BERG-I)  
PREDDIE E R (PRED-I)  
LA English  
DT Patent  
PI WO 9807851 A2 19980226  
PIT WOA2 INTERNATIONAL APPLICATION PUBLISHED WITHOUT INTERNATIONAL SEARCH REPORT or INTERNATIONAL APPLICATION PUBLISHED WITH DECLARATION UNDER ARTICLE 17 (2) (A) [FROM 20090101 ONWARDS]  
AI WO 1997-EP4600 19970822  
PRAI CA 1996-2183900 19960822  
FS NUCLEIC; NS  
OS 1998-169156 [15]  
MTY cDNA  
PSL Disclosure; Fig 10; 96pp  
DESC L-oncogene primer BC531L (reverse).

スコア値は低い (類似性が低い) が同一性が高い回答

ALIGN

Query Length: 303; Sequence Length: 22;  
Score: 41.0 bits (44) , 7.5% of highest possible score 547.7;  
Expect value: 2.725;  
Identities: 22 / 22 (100.0%);  
Strand: Plus / Plus; Alignment Length: 22;  
Q: 72 ACCAGCCCCTCCTGGCCCTG 93  
|||||||  
S: 1 ACCAGCCCCTCCTGGCCCTG 22

参考:コンパクトに表示したい場合はカスタム表示形式を使う

=> D L3 1-10 PI PA PSL ← 特許番号, 出願人, 配列の記載位置のみ表示

L3 ANSWER 1 OF 3420 GENESEQ COPYRIGHT 2022 CLARIVATE on STN.  
PI WO 2010138843 A2 20101202  
PA UNIV TEXAS SYSTEM (TEXA)  
PSL Disclosure; SEQ ID NO 101; 43pp

L3 ANSWER 2 OF 3420 GENESEQ COPYRIGHT 2022 CLARIVATE on STN.  
PI WO 2008091911 A2 20080731  
PA UNIV CALIFORNIA (REGC)  
DEPT VETERANS AFFAIRS (VETE-N)  
PSL Disclosure; SEQ ID NO 16; 37pp

L3 ANSWER 3 OF 3420 GENESEQ COPYRIGHT 2022 CLARIVATE on STN.  
PI WO 2006124700 A2 20061123  
PA INTROGEN THERAPEUTICS INC (INTR-N)  
PSL Disclosure; SEQ ID NO 1; 173pp

:

参考:必要なフィールドを抽出したい場合は ANALYZE (または SELECT) コマンドを使う

=> ANA L3 PN ← 特許番号, 出願人, 配列の記載位置のみ表示

L4 ANALYZE L3 1- PN : 861 TERMS

=> D 1-

L4 ANALYZE L3 1- PN : 861 TERMS

TERM #	# OCC	# DOC	% DOC	PN
1	506	506	14.80	JP2003265187
2	192	192	5.61	W02013151666
3	136	136	3.98	CN111321149
4	67	67	1.96	US20030175862
5	54	54	1.58	W02009001359
6	39	39	1.14	W02010064838
7	38	38	1.11	W02017062726
8	35	35	1.02	W02021015986
9	31	31	0.91	W02017191274
10	30	30	0.88	W02018102584

:

参考:WPI ファイルにクロスオーバーしたい場合は OS フィールドを使う

=> FILE WPINDEX

=> TRA L3 OS /AN ← GENESEQ ファイルの OS フィールドを抽出して, WPINDEX ファイルの /AN (レコード番号) フィールドで検索

L5 TRANSFER L3 1- OS : 862 TERMS

L6 862 L4/AN

ALL TERMS IN L4/AN RETRIEVED.

=> D L6 1 TRI

L6 ANSWER 1 OF 862 WPINDEX COPYRIGHT 2022 CLARIVATE ANALYTICS on STN  
AN **2021-F0903C** [2022008] WPINDEX [Full-text](#)  
TT TT: NEW POLYNUCLEOTIDE SEQUENCE ENCODE CIRCULAR RNA MOLECULAR COMPRISE PROTEIN CODE  
NUCLEIC ACID INTERNAL RIBOSOME ENTER SITE REGION LINK PRODUCE CELL

:

\* WPI ファイルへクロスオーバーする際に, 特許番号 (PN) を用いると, GENESEQ ファイルでヒットした配列情報を含む特許に加え, 関連レコード (分割出願, 継続出願など) がヒットする可能性がある.

## 完全配列検索・部分配列検索

### ■ 完全配列検索・部分配列検索の入力方法

=> RUN GETSEQ コード/検索フィールド

- ・ 回答数の上限は 250,000 件.
- ・ コード以外にギャップ記号や特殊記号を利用することもできる. 使用できる記号の一覧は => ? SQQ で確認する.
- ・ 配列質問式の長さの制限値は検索タイプによって異なる. => ? QLIMIT で確認する.

### ■ 完全配列検索・部分配列検索の検索フィールドと入力例

	検索フィールド	検索タイプ	検索フィールド	検索タイプ
核酸	/SQEN	完全配列	/SQSN	部分配列
タンパク質	/SQEP	完全配列	/SQSP	部分配列
	/SQEFP	完全配列ファミリー	/SQSFP	部分配列ファミリー

- ・ 完全配列検索 : 配列質問式に完全に一致した配列がヒットする.
- ・ 部分配列検索 : 配列質問式を一部に含む配列がヒットする.
- ・ ファミリー検索 : 等価なアミノ酸も含めた配列がヒットする.

=> RUN GETSEQ AATCCGTCAG/SQEN ← 核酸の完全配列検索  
=> RUN GETSEQ MALWMRLLPL/SQEP ← タンパク質の完全配列検索  
=> RUN GETSEQ GFF.[2,10]LSS/SQSP ← ギャップ記号を使った検索  
=> RUN GETSEQ (TGACGGAT|ATCCGTCA)/SQSN ← 特殊記号を使った検索

- ・ アミノ酸の 3 文字コードを用いて検索する場合は, ハイフンで繋ぎ, 全体を ' で囲む.

=> RUN GETSEQ 'ASP-HIS-ALA-ILE-HIS'/SQEP

### ■ 相補鎖の検索

- ・ 核酸の完全配列 (/SQEN) および部分配列検索 (/SQSN) では, 相補鎖を含めて検索する /しないを指定できる.

=> RUN GETSEQ AATCCGTCAG/SQSN -S BOTH ← 相補鎖を含めて検索  
=> RUN GETSEQ AATCCGTCAG/SQSN -S SIN ← 部分配列検索 (一本鎖のみ)  
=> RUN GETSEQ AATCCGTCAG/SQSN -S COM ← 部分配列検索 (相補鎖のみ)  
=> RUN GETSEQ AATCCGTCAG/SQSN ← -S BOTH と同じ



■ 検索例:GCCCAAGCTGGCATCCGTCA を含む配列長 2,000 以上でクレームされている核酸を調べる.

```

=> FILE GENESEQ                ← GENESEQ ファイルに入る

=> RUN GETSEQ (GCCCAAGCTGGCATCCGTCA)/SQSN ← RUN GETSEQ /SQSN で部分配列検索を行う
L1  RUN STATEMENT CREATED
L1  492 (GCCCAAGCTGGCATCCGTCA)/SQSN

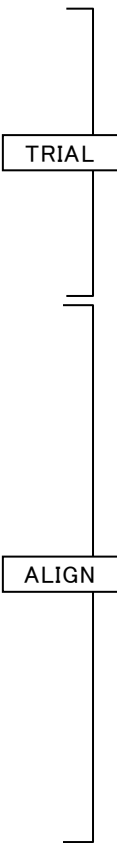
=> S L1 AND SQL=>2000          ← 配列長が 2000 以上のレコードに限定する
L2  201 L1 AND SQL=>2000

=> S L2 AND CLAIM?/PSL        ← 配列がクレームされているレコードに限定する
L3  105 L2 AND CLAIM?/PSL

=> D L3 1-105 TRIAL ALIGN      ← TRIAL ALIGN 表示形式を用いて、標題、配列長、ヒットした
                               コード部分を確認する
L3  ANSWER 1 OF 105 GENESEQ COPYRIGHT 2022 CLARIVATE on STN.
AN  AJQ34458 GENESEQ
TI  New array comprising a transcriptome from a diseased tissue, useful for diagnosing
    a pathological condition or for determining responsiveness of a patient afflicted
    with a pathological condition to a therapeutic agent.
MTY  cDNA
DESC  Human diseased colorectal tissue cDNA transcript SEQ ID NO:26105.
KW  Diagnosis; prognosis; screening; neoplasm; inflammation; degeneration; cancer;
    arthritis; asthma; neurodegenerative disease; cardiovascular disease;
    hypertension; psychiatric disorder; infectious disease; metabolic disorder; immune
    disorder; ss
SQL  3024
    :

L3  ANSWER 17 OF 105 GENESEQ COPYRIGHT 2018 CLARIVATE ANALYTICS on STN
AN  BES08390 GENESEQ
TI  New isolated protein useful for modulating cell proliferation or
    apoptosis, diagnosing or determining risk of colon cancer, and treating
    colon cancer, comprises specific amino acid sequence.
MTY  cDNA
DESC  Human colon cancer associated target (CCAT) cDNA, SEQ ID 1741.
KW  apoptosis modulation; cell proliferation; colon tumor; cytostatic;
    diagnostic test; drug screening; gastrointestinal-gen.; prognosis;
    protein detection; ss; therapeutic; tumor marker.
SQL  2984 ●————— 配列長
ALIGN
    Sequence Length: 2984;
    Strand: Plus / Plus;
    Hits at: 2101-2120 ●—————
        1 GCGCCCCAGT CGAGCGTGAG CTCCTCTGCT ACTCAGAGTT GCAACCTCAG CCTCGCTATG
        61 GCTCCAGCA GCCCCCGCC CGCGCTGCCG GCACCTCTGG TCCTGCTCGG GGCCTGTTC
        121 CCAGGACCTG GCAATGCCCA GAGA ヒット位置
        181 GGCTCCGTGC TGGTGACATG CAGG
        241 ACCCGTTGC CTA AAAAGGA GTTG
    :
    1801 TACAACAGCA TTTGGGGCCA TGGTACCTGC ACACCTAAAA CACTAGGCCA CGCATCTGAT
    1861 CTGTAGTCAC ATGACTAAGC CAAGAGGAAG GAGCAAGACT CAAGACATGA TTGATGGATG
    1921 TTAAAGTCTA GCCTGATGAG AGGGGAAGTG GTGGGGGAGA CATAGCCCA CCATGAGGAC
    1981 ATACAACTGG GAA 相補鎖でヒット CTATTGGGTA TGCTGAGGCC CCACAGACTT
    2041 ACAGAAGAAG TGG TAGCATCAA ACACAAAGGC CCACACTTCC
    2101 TGACGGATGC CAGCTTGGGC ACTGCTGTCT ACTGACCCCA ACCCTTGATG ATATGTATT
    =====
    2161 ATTCATTGT TATTTACCA GCTA ヒットしたコードに二重下線 (=) が付く ACAT
    :

```



=> D 1 ALL ← 書誌情報, 抄録, 配列情報を確認したい場合は, ALL 表示形式を利用する

L3 ANSWER 1 OF 105 GENESEQ COPYRIGHT 2022 CLARIVATE on STN.  
 AN AJQ34458 GENESEQ ED 20211123 UP 20211123  
 DED 20080207 Full-text  
 TI New array comprising a transcriptome from a diseased tissue, useful for diagnosing a pathological condition or for determining responsiveness of a patient afflicted with a pathological condition to a therapeutic agent.  
 IN Harkin P; Johnston P; Mulligan P  
 PA ARRADX LTD (ARRA-N)  
 LA English  
 DT Patent  
 PI WO 2006048291 A2 20060511

ベーシック特許の特許情報のみ表示される  
 ・ 対応特許情報を表示したい場合は, WPI ファイルへ  
 クロスオーバー検索して確認する

MTY cDNA  
 PSL **Claim 4**; SEQ ID NO 26105; 64pp  
 DESC Human diseased colorectal  
 KW Diagnosis; prognosis; screening; arthritis; asthma; neurodegenerative disease; cardiovascular disease; hypertension; psychiatric disorder; infectious disease; metabolic disorder; immune disorder; ss  
 ORGN Homo sapiens  
 AB The invention relates to a new array comprising a diseased tissue. The diseased tissue comprises a tissue affected by a disease, an inflammatory disease or a degenerative disease, including cancer, arthritis, asthma, neurodegenerative diseases, cardiovascular diseases, hypertension, psychiatric disorders, infectious diseases, metabolic disorders and  
 SEQN 26105  
 SQL **3024**  
 SEQK 949513f305c043acba540628f3f7924f03cfe5b57542e511e8cb3d13a5458f24

特許中の配列情報の記載位置  
 ・ クレーム 4 に記載されており, 配列番号は 26105 である

WPI ファイルの抄録と異なる配列  
 独自の抄録を読むことができる

SEQ  
 1 ctataaagga tcacgcgccc cagtcgacgc tgagctcctc tgctactcag  
 51 agttgcaacc tcagcctcgc tttgacgg atgccagctt gggcactgct gtctactgac  
 2101 aggccacac ttccatgacgg atgccagctt gggcactgct gtctactgac  
 2151 cccaaccctt gatgatatgt atttattcat ttgtattttt accagctatt  
 2201 tattgagtgt cttttatgta ggctaaatga acataggtct ctggcctcac  
 2251 ggagctccca gtcctaatac cattcaaggt caccaggtag agttgtacag  
 2301 gttgtacact goaggagagt gcctggcaaa aagatcaaat ggggctggga  
 2351 cttctcattg gccaacctgc cttccccag aaggagtgat ttttctatcg  
 2401 gcacaaaagc actatatgga ctggtaatgg ttacaggttc agagattacc  
 2451 cagtgaggcc ttattcctcc cttccccca aaactgacac ctttgttagc  
 2501 cacctcccca ccacataca tttctgccag tgttcacaat gacactcagc  
 2551 ggtcatgtct ggacatgagt gccagggaa tatgccaag ctatgccttg  
 2601 tcctctgtgc ctgcttgcac ttcactggga gcttgacta tgcagctcca  
 2651 gtttcctgca gtgatcaggg tcctgcaagc agtggggaag ggggccaagg  
 2701 tattggagga ctccctcca gctttggaag cctcatccgc gtgtgtgtgt  
 2751 gtgtgtatgt gtagacaagc tctogctctg tcaccaggc tggagtgcag  
 2801 tgggtcaatc atggttcaat gcagcttga ccttttgagc tcaagtgatc  
 2851 ctcccacctc agcctcctga gtagctggga ccataggctc acaacaccac  
 2901 acctggcaaa tttgattttt ttttttttc cagagacggg gtcttgcaac  
 2951 attgcccaga cttcctttgt gttagttaat aaagctttct caactgocaa  
 3001 aaaaaaaaaa aaaaaaaaaa aaaa

ALL 表示形式ではヒットしたコードがハイライトされない

NA

Code	Count	Percent
A	719	23.8
C	878	29.0
G	801	26.5
T	626	20.7
U	0	0.0
Other	0	0.0

## 塩基コード・アミノ酸コード

### ■ 核酸の塩基コード

- ・ 塩基コード (Symbol) と曖昧コード (Ambiguity Codes) がある. 利用できるコードは => ? NUC で確認できる.
- ・ 配列のコーディング (コードの並び順) は, 5' 末端から 3' 末端の順.

塩基コード	塩基名	完全配列でヒットするコード	部分配列でヒットするコード
A	adenine	A	A
C	cytosine	C	C
G	guanine	G	G
T	thymine (in DNA)	T	T
U	uracil (in RNA)	U	U
曖昧コード	定義	完全配列でヒットするコード	部分配列でヒットするコード
Y	pyrimidine (C or T/U)	Y	Y
R	purine (A or G)	R	R
M	amino (A or C)	M	M
K	keto (G or T/U)	K	K
S	strong interaction (3 H bonds) (C or G)	S	S
W	weak interaction (2 H bonds) (A or T/U)	W	W
B	not-A (C, G or T/U)	B	B
V	not-T/U (A, C or G)	V	V
D	not-C (A,G or T/U)	D	D
H	not-G (A,C or T/U)	H	H
N	unknown or any nucleotide	N	N

### ■ 参考: イノシンの検索

- ・ GENESEQ ファイルではイノシンの塩基コードは使用されていない.
- ・ レコード内ではイノシンに対して, 曖昧コード N が利用されており, 特徴表に N がイノシンであることが示されている. そのため, 検索には下記の式が有効である.

=>S L# AND (MODIFIED(S)BASE(L)I)/FEAT

■ アミノ酸コード（一般アミノ酸コード）

- ・ 1 文字コードと 3 文字コードが利用できる。利用できるコードは => ? AAC で確認できる。
- ・ 配列のコーディング（コードの並び順）は、N 末端（NH<sub>2</sub>）から C 末端（COOH）の順。

1 文字コード	3 文字コード	アミノ酸名	完全配列でヒットするコード	部分配列でヒットするコード
A	ALA	Alanine	A	A
B*	ASX	Aspartic acid or Asparagine	B	B
C	CYS	Cysteine	C	C
D	ASP	Aspartic acid	D	D
E	GLU	Glutamic acid	E	E
F	PHE	Phenylalanine	F	F
G	GLY	Glycine	G	G
H	HIS	Histidine	H	H
I	ILE	Isoleucine	I	I
J	XLE	Leucine or Isoleucine		
K	LYS	Lysine	K	K
L	LEU	Leucine	L	L
M	MET	Methionine	M	M
N	ASN	Asparagine	N	N
P	PRO	Proline	P	P
Q	GLN	Glutamine	Q	Q
R	ARG	Arginine	R	R
S	SER	Serine	S	S
T	THR	Threonine	T	T
U	SEC	Selenocysteine		
V	VAL	Valine	V	V
W	TRP	Tryptophan	W	W
X	Xaa	Unspecified (A to Y) or Unknown or Other	X	X
Y	TYR	Tyrosine	Y	Y
Z*	GLX	Glutamic acid or Glutamine	Z	Z

\* B, J, Z は検索タイプによりヒットするコードが異なる。

	ヒットするコード	
	/SQSP	/SQEFP, /SQSFP
B	B	B, D, N, E, Q, Z
Z	Z	B, D, N, E, Q, Z

■ 特殊・未定義のアミノ酸コード

- ・ 特殊・未定義のアミノ酸コードは X (XXX) で検索する。
  - 特徴表（FEAT フィールド）に X に対応する特殊・未定義アミノ酸情報が収録されている。

■ ファミリー検索時に適用されるアミノ酸の等価表

等価クラス	アミノ酸コード
親水性, 酸・アミド	D, E, N, Q, B, Z
親水性, 塩基	H, K, R
弱疎水性, 中性	A, G, P, S, T
疎水性	I, L, M, V
疎水性, 芳香族	F, Y, W
橋かけ形状	C

## その他の主な検索フィールド

フィールド	検索項目	句/単語/ 数値	前方 一致	後方・中間 一致 <sup>*1</sup>	入力例
なし または /BI	基本索引	単語	○	×	S INSULIN (L) PREP? S JUNIPERUS VIRGINIANA S EPIDERM? (L) GROWTH?
/AA	アミノ酸の種類	句	×	×	S S/AA AND T/AA
/AA.CNT <sup>*3</sup>	アミノ酸の数	数値	×	×	S S/AA (S) AA.CNT>9
/AA.PER <sup>*3</sup>	アミノ酸の比率	数値	×	×	S T/AA (S) 25-30/AA.PER
/AC	出願国	句	○	×	S L1 AND AT/AC S L1 AND AUSTRIA/AC
/AD <sup>*3</sup>	出願日	数値	×	×	S 7 JUN 1999/AD
/AP <sup>*4</sup>	出願番号	句	○	×	S AU94-64520/AP S 94AU-0064520/AP
/AY <sup>*3</sup>	出願年	数値	×	×	S L1 AND 2000/AY
/CR	関連する配列レコードの クロスリファレンス	句	○	×	S ABU09574/CR
/DESC	配列の説明	句/単語	○	×	S DYNORPHIN/DESC
/DED <sup>*3</sup>	データの入力日	数値	×	×	S DED>20140103
/DUPD <sup>*3</sup>	データの更新日	数値	×	×	S DUPD>20000101
/ED	入力日	数値	×	×	S 20211030/ED
/FEAT	特徴表	単語	○	○	S L1 AND CDS/FEAT S L2 AND ?CYTE?/FEAT
/FS	ファイルセグメント	句	○	×	S L10 AND NS/FS S L10 AND PROTEIN/FS
/IN	発明者	句	○	×	S GOSSEN M/IN
/MTY	分子タイプ	句	○	×	S L1 AND CDNA/MTY S L7 AND RNA/MTY
/NA	塩基の種類	句	×	×	S C/NA
/NA.CNT <sup>*3</sup>	塩基の数	数値	×	×	S A/NA (S) NA.CNT>10
/NA.PER <sup>*3</sup>	塩基の比率	数値	×	×	S G/NA (S) 60-70/NA.PER
/ORGN	生物名	句/単語	○	×	S ADENOVIRUS/ORGN S (SIMIAN AND ADENOVIRUS)/ORGN
/OS	WPI のレコード番号	句	○	×	S 1994-151326/OS

フィールド <sup>②</sup>	検索項目	句/単語/ 数値	前方 一致	後方・中間 一致 <sup>*1</sup>	入力例
/PA <sup>*2</sup>	特許出願人	句/単語	○	×	S CHIRON SPA/PA
/PACO	特許出願人コード <sup>③</sup>	句	○	×	S BADI/PACO
/PC	特許発行国	句	○	×	S L1 AND US/PC
/PK	特許種別コード <sup>③</sup>	句	○	×	S L1 AND JPA/PK
/PN <sup>*4</sup>	特許番号	句	○	×	S DE4244565/PN S EP-348819/PN
/PSL	特許中の配列情報の 記載位置	単語	○	×	S DISCLOSURE/PSL S CLAIM?/PSL S (CLAIM?(S)1)/PSL
/PD <sup>*3</sup>	特許発行日	数値	×	×	S L2 AND 1 NOV 2001/PD
/PY <sup>*3</sup>	特許発行年	数値	×	×	S L3 AND 1997-1999/PY
/PRC	優先権主張国	句	○	×	S L1 AND JP/PRC
/PRD <sup>*3</sup>	優先権主張日	数値	×	×	S 19930907/PRD
/PRDF <sup>*3</sup>	最先の優先権主張日	数値	×	×	S 17 JUNE 1993/PRDF
/PRN <sup>*4</sup>	優先権出願番号	句	○	×	S US93-78471/PRN S 93US-0078471/PRN
/PRY <sup>*3</sup>	優先権主張年	数値	×	×	S L1 AND 1995-2000/PRY
/PRYF <sup>*3</sup>	最先の優先権主張年	数値	×	×	S L1 AND 1995-2000/PRYF
/SEQK	配列キー	句	○	×	S 0702CCB35274BA330A6D0F5D C8D75884CEDE8D95A06A051381 70003CE5EC921E/SEQK
/SEQN	配列番号	数値	×	×	S 337/SEQN
/SQL <sup>*3</sup>	配列長	数値	×	×	S 70-90/SQL
/TI	WPI の標題	単語	○	×	S HUMAN INSULIN/TI
/UP	更新日	数値	×	×	S 20220204/UP

\*1 中間一致・後方一致検索を利用する場合は、入力語は最低 4 文字は必要

\*2 (S) 演算子はスペースで代用できる。

\*3 数値演算子または範囲指定による検索が可能な数値検索フィールド

\*4 STN 形式またはダウエント形式のどちらでも利用できる。



# JAICI

化学情報協会

情報事業部

〒113-0021 東京都文京区本駒込6-25-4 中居ビル

TEL: 0120-003-462 FAX: 03-5978-4090

URL: [www.jaici.or.jp](http://www.jaici.or.jp)

E-mail: [support@jaici.or.jp](mailto:support@jaici.or.jp)