GenBank クイックスタート

GenBankはNLM/NCBIにて維持管理されている核酸配列データベースです。また、GenBankはEMBL, DDBJと三極間で連携しながら国際核酸配列データベースを共同で構築しています。これら三機関はデータを日々交換し続けており、 その規模は160000種にも及ぶ生物種の塩基配列から成り立つまでになっています。

このGenBankクイックスタートでは、

- 1. GenBankの歴史と発展
- 2. タイプの異なるエントリの種類とアクセッション番号との関連性
- 3. GenBankの典型的なエントリのフォーマットとそのアノテーション
- 4. GenBankへの配列登録ツール: BankItおよびSequinの使い方
- 5. GenBankにおいてどのように配列登録が行われるか
- 6. GenBank ftpサイトの概要
- 7. 効率的にGenBankからデータを取得するためのTips

などを紹介していきます。

課題1

■ 課題1-1

Entrez-Nucleotideを用いて"pannexin"という単語を含むエントリを検索してみてください。

何エントリがヒットしましたか?その中でCoreNucleotide、ESTに含まれるエントリはそれぞれいくつですか? 検索結果を1ページ内に全件表示させてください。

■ 課題1-2

CoreNucleotideデータにヒットしたエントリの中から GenBankに登録されたgenomic DNA由来のエントリだけに絞 り込んでください。

GenBankには冗長なエントリが含まれていますか?

pannexinに関して複数のエクソンが含まれるエントリと個々のエクソンが登録されているエントリとを見つけてください。

■ 課題1-3_

絞り込まれた検索結果の中から、アクセッション番号AY048509のエントリを例として以下の質問に答えてみてくだ さい。

いつが最後の更新日ですか?

このエントリは登録後、更新されていますか?

このエントリにはコード領域が完全に含まれていますか?

5'UTRはこのエントリ内のどの部分に相当しますか?

このエントリに含まれるコード領域を翻訳したアミノ酸配列のアクセッション番号は何番ですか?

■ 課題1-4_

Entrezの検索結果に戻ってください。ソースデータベースをRefSeqに変更し、molecular typeをmRNA、さらに生物 種をヒトに限ってください。

いくつのエントリが表示されていますか?これら全部のエントリをFASTA形式で保存してください。

■ 課題1-5_

アクセッション番号NM_015368の更新履歴を見てください。最新の更新はいつですか?登録後何回更新されていま

すか?更新日はいつですか?

課題2

今度は、FOXP2遺伝子を使って上述した手順を繰り返して下さい。

■ 課題2-1_

Entrez-Nucleotideを用いて"FOXP2"という単語を含むエントリを検索してみてください。

何エントリがヒットしましたか?その中でCoreNucleotide、ESTに含まれるエントリはそれぞれいくつですか?

検索結果を1ページ内に全件表示させてください。

■ 課題2-2_

CoreNucleotideデータにヒットしたエントリの中から GenBankに登録されたgenomic DNA由来のエントリだけ に絞り込んでください。

GenBankには冗長なエントリが含まれていますか?

FOXP2に関して複数のエクソンが含まれるエントリと個々のエクソンが登録されているエントリとを見つけてください。

■ 課題2-3_

絞り込まれた検索結果の中から、アクセッション番号AF515032のエントリを例として以下の質問に答えてみ てください。

いつが最後の更新日ですか?

このエントリは登録後、更新されていますか?

このエントリにはコード領域が完全に含まれていますか?

5'UTRはこのエントリ内のどの部分に相当しますか?

このエントリに含まれるコード領域を翻訳したアミノ酸配列のアクセッション番号は何番ですか?

■ 課題2-4

Entrezの検索結果に戻ってください。ソースデータベースをRefSeqに変更し、molecular typeをmRNA、さらに生物種をヒトに限ってください。

いくつのエントリが表示されていますか?これら全部のエントリをFASTA形式で保存してください。

■ 課題2-5_

アクセッション番号NM_148900の更新履歴を見てください。最新の更新はいつですか?登録後何回更新されていますか?更新日はいつですか?

解答・解説

■ 解答1-1_

<u>Entrez-Nucleotideを用いた検索</u>

NCBIトップページを開きましょう。

| S NCPI | National C | enter for l | Biotechnolo | gy Inform | ation | |
|---|---|--|--|--|-----------------------------------|--|
| | National Library o | f Medicine | Nat | ional Institutes | of Health | |
| PubMed All Dat | abases BLAST | OMIM | Books | TaxBrowser | Structure | |
| Search All Database | s 💌 for | | Go | | | |
| SITE MAP | What does NC | BI do? | | Но | t Spots | |
| Alphabetical List Resource Guide | Established in 19 | 88 as a natio | onal resource | for Asser | nbly Archive | |
| About NCBI An introduction to NCBI | molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software | | | S Cluste ortholog | Clusters of orthologous groups | |
| GenBank Sequence submission support and software | tools for analyzing disseminates bior the better understa processes affectin disease. <u>More ab</u> | or Coffee Genes & NCBI Ha | e Break, & Disease, andbook | | | |
| Literature databases PubMed, OMIM, | dbGaP: NCBI's Association Da | s Genome N atabase | Wide | Entrez | z Home | |
| Books, and PubMed Central Molecular databases | NCBI's <u>dbGaP</u> (da Phenotype) provid Association (GW/ elucidate the link b | atabase of Gi les data from A) studies, wi petween geni sers bave as | enotype and Genome Wid hich are helpin es and disease | Entrez Gene Gomnibus Gomnibus | z Tools expression s (GEO) | |
| Sequences, structures, and taxonomy | information about measured and pre between subjects | the phenotype- computed a phenotypes | ic variables associations and genotype: | su PHuma resource s. Influer | n genome es nza Virus | |
| Genomic biology | Concerns also de al | | | Resourc | e | |

画面上部の検索窓を使ってNucleotideデータベースから"pannexin"という単語を含むエントリを検索しま す。画面上部の検索サービスはEntrezと呼ばれ、文献から塩基配列、SNPs、ゲノム、立体構造など様々なデ ータベースへの検索を提供しています。その中のNucleotideデータベースはGenBank, RefSeq, PDBなど様々 なソースから塩基配列を集めたものです。

プルダウンメニューからNucleotideを選択し、検索窓にpannexinと打ち込んで検索を実行します。

| PubM | ed All Dat | abases | BLAST | OMIM | Books | TaxBrowser | Structure |
|-------------------------------|----------------------------|------------|--|--|--|----------------------------|-----------------------|
| Search | All Database | s 🗸 | for pannexi | n | | Go | |
| SITE M | All Database NCBI Web S | s ^ ite | loes NCB | l do? | | Ho | ot Spots |
| Alphabeti Resource | PubMed Protein | | ed in 1988 | as a natio | onal resourc | e for 🕨 Asser | mbly Archive |
| About N An introdu NCBI | Nucleotide EST GSS | - | biology in abases, c ional biolo | formation, onducts re gy, develo | NCBI creat search in ps software | eS ► Cluste ortholog | ers of Jous groups |
| GenBar | Genome Books | | analyzing g ates biome | enome da edical info | ta, and rmation - all | for Coffee | e Break, |

検索結果の上部から、201エントリにヒットしたことがわかります。(赤四角)

| Limi | s Preview/Index History Clipboard | Details | |
|------------|--|--|--|
| Histor | y has expired. | | |
| Found | 201 nucleotide sequences. Nucleotide [175 | 5] EST [26] | |
| Display | Summary 💌 | Show 20 💌 Sort by 💌 Ser | nd to 💌 |
| All: | 175 Bacteria: 0 RefSeq: 81 mRNA: 98 | * | |
| Item | s 1 - 20 of 175 | Page 1 of 9 Next | ▼ Top Organisms [] |
| This | search in Gene shows 45 results including | | Homo sapiens (35) |
| Ba | (Pattus poperaisus): Papeavin 1 | | Mus musculus (27) Pathis populations (10) |
| <u>Fa</u> | Adve museulus): persevis 1 | | Danio rerio (18) |
| | IXI (Mos moscolos), pannexin 1 | | Pan troglodytes (10) |
| PA | (Homo sapiens), pannexin i | | All other taxa (66) |
| | NM 012762 Reports | Links | more |
| | Rattus norvegicus caspase 1 (Casp1), mRN | NA | Recent Activity |
| | g 31542340 ref NM_012762.2 [31542340] | | т |
| 2: | NM 172454 Reports | Links | 19 |
| | Mus musculus pannexin 3 (Parx3), mRNA gi 86262154 ref NM_1724542 [86262154] | | Q pannexin (175) |
| 3: | NM 019482 Reports | Order cDNA clone, Links | |
| | Mus musculus pannexin 1 (Panx1), mRNA gi 86262133 ref NM_019482.2 [86262133] | | |
| 4 : | NM 199397 Reports | Links | |
| | Rattus norvegicus Pannexin 1 (Panx1), mR gi 40786482 ref NM_199397.1 [40786482] | NA. | |
| 5 : | NM 015368 Reports | Order cDNA clone, Links | |
| | Homo sapiens pannexin 1 (PANX1), mRNA gi[39995063]ref[NM_015368.3][39995063] | | |
| 6: | AB463158 Reports | Links | |
| | Synthetic construct DNA, clone: pF1KB87 gene for pannexin 2, without stop codon, in g 208965313 dbj AB463158.1 [208965313] | 16, Homo sapiens PANX2 n Flexi system | |

また、右横に書かれている情報から201エントリの内訳は、CoreNucleotideに175エントリ、ESTに26エントリ であることが分かります。

デフォルトの検索結果画面ではCoreNucleotideにヒットした175エントリが20件ずつのページに分かれて表示 されています。

検索結果の表示件数を変える

今は、175件のヒットのうち先頭の20件が表示されています。この表示件数を変えて、全件が表示されるよう にしてみましょう。そのためには、画面上部のshowと書かれた右横のプルダウンメニューから 1ページあた りの表示件数を選択します。

| Found 201 nucleotide sequences. Nucleotide [17 | 5] ES | т [<u>26</u> |] | | | | |
|--|-------|-------------------|------|------|------|--------|-------------------------------------|
| Display Summary | Show | 20 | V So | tby | 💌 S | end to | * |
| All: 175 Bacteria: 0 RefSeq: 81 mRNA: 98 | * | 5 10 | | | | | |
| Items 1 - 20 of 175 | Pag | 20 50 | | of 9 | Next | | Top C |
| This search in Gene shows <u>45 results, including:</u> Panx1 (Rattus norvegicus): Pannexin 1 | | 100 200 500 | ĥŝ | | | | Homo sa Mus mus Rattus no |
| Panx1 (Mus musculus): pannexin 1 PANX1 (Homo sapiens): pannexin 1 | | | | | | | Danio rer Pan trogl All other |
| | | | | | | M | ore |

全部で175件なので、200を選択します。

| Display | Summary | 🖌 Show | 200 💌 | Sort by | Send to |
|-------------|--|------------------|-------|------------|----------|
| All: 1 | 75 Bacteria: 0 RefSeq: 81 mRI | VA: 98 🛠 | | | |
| Item | s 1 - 175 of 175 | | | One | page. |
| | | | | | |
| This | search in Gene shows <u>45 results,</u> incl | uding: | | | |
| Par | x1 (Rattus norvegicus): Pannexin 1 | | | | |
| Par | x1 (Mus musculus): pannexin 1 | | | | |
| PAN | IX1 (Homo sapiens): pannexin 1 | | | | |
| □1. | NB 4 010760 Reports | | | | P |
| U 1. | Rattus popularious caspase 1 (Casp) |) mRNA | | | LINKS |
| | gi[31542340]ref[NM_012762.2][31542 | 340] | | | |
| 2: | NM 172454 Reports | | | | Links |
| | Mus musculus pannexin 3 (Panx3), n gi 86262154 ref NM_172454.2 [86262 | 18NA 154] | | | |
| 3: | NM 019482 Reports | | Order | cDNA clone | e, Links |
| | Mus musculus pannexin 1 (Panx1), n gi 86262133 ref NM_019482.2 [86262 | 18NA 133] | | | |
| 4: | NM 199397 Reports | | | | Links |
| | Rattus norvegicus Pannexin 1 (Pan» gi 40786482 ref NM_199397.1 [40786 | 1), mRNA 482] | | | |
| 5 : | NM 015368 Reports | | Order | cDNA clone | e, Links |
| | Homo sapiens pannexin 1 (PANX1), gi 39995063 ref NM_015368.3 [39995 | mRNA 063] | | | |
| | AD4004E0 Devede | | | | |

上図のように全件が1ページに収まって表示できているのが確認できます。

■ 解答1-2____

検索結果を絞り込む

続いて、ヒットしたエントリからgenomic DNA由来のエントリに絞り込んでみましょう。

検索結果の絞込みには画面上部の"Limits"と書かれたタブから行うのが一番簡単になります。クリックしてみましょう。

| Limits | Preview/Index History Clipboard Details | | | | | |
|--|---|--------|--|--|--|--|
| Found 2 | 01 nucleotide sequences. Nucleotide [<u>175</u>] EST [<u>26</u>] | | | | | |
| Display | Summary Show 20 💌 Sort by | Send t | | | | |
| All: 1 | 75 Bacteria: 0 RefSeq: 81 mRNA: 98 😿 | | | | | |
| Items | 1 - 20 of 175 Page 1 of 9 | Next | | | | |
| This search in Gene shows <u>45 results,</u> including: <u>Panx1</u> (Rattus norvegicus): Pannexin 1 <u>Panx1</u> (Mus musculus): pannexin 1 | | | | | | |
| PAN | X1 (Homo sapiens): pannexin 1 | | | | | |
| [] 1:] | <u>NM 012762</u> Reports Rattus norvegicus caspase 1 (Casp1), mRNA gi 31542340 ref NM_012762.2 [31542340] | Links | | | | |
| 2: | <u>NM 172454</u> Reports Mus musculus pannexin 3 (Panx3), mRNA gil86262154]refINM 1724542[[86262154] | Links | | | | |

すると下図のように様々な条件をプルダウンメニューで選んで選択し、その条件で絞り込んでから先ほどと 同じキーワード(例ですとpannexin)で検索が可能です。

| or pannexin | Go Clear |
|---|--|
| Limits Preview/Index History Clipb | oard Details |
| Use All Fields pull-down menu to spe If search fields tags are used enclose More help on using limits is available | ecify a field. e in square brackets, e.g., rubella [ti]. <u>here</u> . |
| mited to: | |
| Fields | |
| All Fields | |
| Exclude: | |
| STSs working draft TPA | patents |
| and the second se | |
| Molecule: | Gene Location: |
| Any | Gene Location: |
| Any Segmented Sequences: | Gene Location: Any |
| Any Segmented Sequences: Any | Gene Location: Any Only from: Any |
| Molecule: Any Segmented Sequences: Any Published in the last: | Gene Location: Any Only from: Any |
| Molecule: Any Segmented Sequences: Any Published in the last: Any Date | Gene Location: Any Only from: Any |
| Molecule: Any Segmented Sequences: Any Published in the last: Any Date Modified in the last: | Gene Location: Any Only from: Any |

まず、genomic DNA由来に絞り込むには、 Moleculeと書かれたプルダウンメニューから"genomic DNA/RNA"を、

| -Molecule: | Gene Location: |
|------------------------|----------------|
| Any 🔽 | Any |
| Any as: | Only from: |
| mRNA SS. | Anu V |
| rRNA | Any |
| cRNA | |
| Any Date | |
| -Modified in the last: | |
| | |

GenBankに登録されたエントリに絞り込むには、Only fromと書かれたプルダウンメニューから"GenBank"を選択し、

| v | Any 💌 |
|----------|------------|
| | Any RefSer |
| ~ | GenBank |
| | |
| ~ | PDB |

再度上の検索窓に"pannexin"というキーワードが入っていることを確認して検索を実行してください。

| ✓ for pannexin キーワードを確認 Limits Preview/Index History Clipboard D | Go Clear 検索実行 |
|--|---------------------------------------|
| Limits: Genomic DNA/RNA, GenBank | |
| Use All Fields pull-down menu to specify a fie If search fields tags are used enclose in squar More help on using limits is available <u>here</u>. | ld. re brackets, e.g., rubella [ti |
| Limited to: Fields All Fields | |
| Exclude: STSs working draft TPA patents | |
| Molecule: Genomic DNA/RNA マ Segmented Sequences: | Gene Location: Any |
| Any | GenBank 💌 |

"Limits"タブにチェックが入って、Limits:Genomic DNA/RNA, GenBankと書かれていることから絞込み検索が 行われたことを確認してください。(下図の赤く囲んだ領域)

| Searc | h Nucleotide 🛛 🔽 for pannexin | |
|----------|--|-------|
| I | imits Preview/Index History Clipboard Details | |
| Limits | s: Genomic DNA/RNA, GenBank | |
| Found | 15 hucleotide sequences. Nucleotide [<u>15</u>] | |
| Display | / Summary 🛛 💙 Show 20 💌 Sort by | 🖌 Se |
| All: | 15 Bacteria: 0 RefSeq: 0 mRNA: 0 🛠 | |
| Item | ns 1 - 15 of 15 One | page. |
| This | search in Gene shows 45 results, including: | |
| 11115 | nut (Detter servering), Deservin 1 | |
| Pa | nx1 (Rattus norvegicus): Pannexin 1 | |
| Pa | nx1 (Mus musculus): pannexin 1 | |
| PA | <u>NX1</u> (Homo sapiens): pannexin 1 | |
| □1: | CH473993 Reports | Links |
| | Rattus norvegicus 181000120144526 genomic scaffold, whole geno | me |
| | shotgun sequence | |
| | gi/4325685[gb]CH473993.2[L74325685] | |
| | CH474096 Reports | Links |
| | shotgun sequence | me |
| | gi[74325592]gb[CH474096.2][74325592] | |
| 3: | CH466550 Reports | Links |
| | Mus musculus 232000009743879 genomic scaffold, whole genome | |
| | shotgun sequence gi 74140747 gb CH466550.2 [74140747] | |

結果が15件に絞り込まれたことがわかります。

| 7 : | AF398508 Reports | Links |
|--------------|--|----------|
| | Homo sapiens pannexin 1 (PANX1) gene, exons 3, 4, 5, and | complete |
| | cds | |
| | gi 15193202 gb AF398508.1 AF398506S3[15193202] | |
| 8: | AF398507 Reports | Links |
| | Homo sapiens pannexin 1 (PANX1) gene, exon 2 gi 15193201 gb AF398507.1 AF398506S2[15193201] | |
| 9: | AF398506 Reports | Links |
| | Homo sapiens pannexin 1 (PANX1) gene, exon 1 gi 15193200 gb AF398506.1 AF398506S1[15193200] | |
| 1 0: | AH010983 Reports | Links |
| | Homo sapiens pannexin 1 (PANX1) gene, complete cds gi 15193199 gb AH010983.1 SEG_AF398506S[15193199] | |
| 11 : | AY048509 Reports | Links |
| | Homo sapiens pannexin 1 (PANX1) gene, partial cds gi 15808666 gb AY048509.1 [15808666] | |
| 1 2: | AF283663 Reports | Links |
| | Homo sapiens pannexin 1 (PANX1) gene, exons, partial cds gi 14794510 gb AF283663.1 AF283661S3[14794510] | |
| 1 3: | AF283662 Reports | Links |
| | Homo sapiens pannexin 1 (PANX1) gene, exon gi 14794509 gb AF283662.1 AF283661S2[14794509] | |
| 1 4: | AF283661 Reports | Links |
| | Homo sapiens pannexin 1 (PANX1) gene, exon gi 14794508 gb AF283661.1 AF283661S1[14794508] | |
| 1 15: | AH010945 Reports | Links |
| | Homo sapiens pannexin 1 (PANX1) gene, partial cds gi 14794507 gb AH010945.1 SEG_AF283661S[14794507] | |

結果を下のほうにスクロールしてみていくと、PANX1遺伝子のエクソンが複数登録されていることがわかりま す。このようにGenBankには冗長度が含まれているため、ユーザ側でどのエントリを用いるかをある程度判断 することも必要です。例えば、赤く囲んだアクセッション番号AF398508のエントリには、 PANX1遺伝子のエ クソン3,4,5番目が完全に含まれたゲノム断片が登録されていることがわかりますし、 AF398507にはエクソ ン2番目のみを含むゲノム断片が登録されていることがわかります。

■ 解答1-3____

<u>エントリの更新状況を知る</u>

では、検索結果からアクセッション番号AY048509のエントリをクリックして詳細を表示させてみてください。

| 1: <u>AY04</u> | 18509. Reports Homo sapiens pann[gi:15808666] |
|---|---|
| Features | Sequence |
| LOCUS DEFINITION ACCESSION VERSION KEYWORDS | AYO48509 1430 bp DNA linear PRI 29-SEP-2001 Homo sapiens pannexin 1 (PANX1) gene, partial cds. AYO48509 AYO48509.1 GI:15808666 |
| SOURCE ORGANISM | Homo sapiens (human) <u>Homo sapiens</u> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo. |
| REFERENCE AUTHORS | 1 (bases 1 to 1430) Baranova,A., Ivanov,D., Skoblov,M., Pestova,A., Kelmanson,I., Shagin,D., Usman,N., Lukyanov,S. and Panchin,Y. Mammalian nannevin family boxelogous to invertebrate gan-junction |
| | Unpublished 2 (bases 1 to 1430) |
| JOURNAL | Ivanov,D. and baranova,A. Direct Submission Submitted (27-JUL-2001) Genome Analysis Laboratory, Vavilov Institute of General Genetics, Gubkina 3, Moscow 119991 |
| FEATURES source | Location/Gualifiers 11430 /organism="Homo sapiens" /mol_type="genomic DNA" |
| | /db_xref="taxon: <u>9606</u> " /chromosome="11" /map="between D11S917 and D11S4176" /clone="LANL LA1139-020" |
| gene | 606>1430 //mmo="PANY1" |
| <u>mRNA</u> | 606>1170 /gene="PANX1" /product="pannexin 1" |
| <u>exon</u> | 6061170 /gene="PANX1" /number=1 |
| <u>CDS</u> | 990>1170 /gene="PANX1" /note="putative gap junction protein; contains four nutative transmembrane domaine: similar to innexin san |
| | Junction gene family /codon_start=1 /product="pannexin 1" |
| | /protein_id= <u>AALU66044.1</u> /db_xref="GI:15808667" /translation="MAIAQLATEYVFSDFLLKEPTEPKFKGLRLELAVDKMVTCIAVG LPLLISLAFAQEISI" |
| ORIGIN 1 a | cctgcagge ecceptetet etgetggete eeggaacaag getetgattg ggatggeaga |
| 61 g 121 t 181 c 241 c 301 c | gaaagagaa acgggaccag cagogotact caggootoga aactocacac toactacogt toogogoca coototoacg oggagotoot ggaagggtoo agogotoot otgoaaggoa ogtoggtot acacogotog attggoottt tatggoactg actgttottt gaagottttg agotootto taaccogatt toacogtgtt ggaaatgttg ggggtggggo glaaaggggg calcoogto cacagaaaga gtogottgoa gaaatgattt taagactooo gegtencage |

このように各エントリは、はじめの部分でエントリに関する情報が記述された後、最後に配列が書かれてい ます。

最終の更新日は、エントリの1行目LOCUSと書かれた行の一番右側に表示されています。

| LOCUS | AY048509 | 1430 bp | DNA | linear | PRI | 29-SEP-2001 |
|------------|-------------------------|---------|------------|----------|-----|-------------|
| DEFINITION | Homo sapiens pannexin l | TPANX11 | gene, part | ial cds. | | |

2001年9月21日に最終更新されたことがわかります。

次に、登録後このエントリが更新されたかは、VERSIONと書かれた箇所を見ることでわかります。

GenBankでは、登録された時にアクセッション番号が発行されますが(例ではAY048509)、その際に VERSION 番号というものも発行されます。VERSIONは登録時が1で更新されるたびに2,3と数字が一つずつ増えていきま す。この数字はアクセッション番号の後ろに.1,.2というように付け加えられています。このエントリの VERSIONは AY048509.1ですので、登録時から更新されていないことがわかります。

ゲノムエントリからエクソン情報を知る

次に、このエントリに書かれたエクソン(遺伝子)情報を見ていきましょう。エントリのどの領域にエクソンが含まれるかなどのアノテーション情報はFEATURESに書かれています。

| | Institute of General Genetics, Gubkina 3, Moscow (1999) |
|----------|--|
| FEATURES | Location/Qualifiers |
| source | 11430 |
| | /organism="Homo sapiens" |
| | /mol type="genomic DNA" |
| | /db xref="taxon:9606" |
| | /chromosome="11" |
| | /map="between D11S917 and D11S4176" |
| | /clone="LANL_LA1139-020" |
| gene | 606>1430 |
| | /gene="PANX1" |
| mRNA | 606>1170 |
| | /gene="PANX1" |
| | /product="panneyin 1" |
| evon | 6061170 |
| 0.001 | /vone="PANX1" |
| | /number=1 |
| 200 | 990 \1170 |
| 000 | /wano="DAMY1" |
| | /actor mutative gap junction protein: contains four |
| | note putative gap junction protein, contains rou |
| | junctive transmentrane domarns, shirtar to threath gap |
| | Junction gene raining |
| | /coun_start-1 |
| | /product = pannexin i |
| | Alberta Contraction and a contraction of the second s |
| | |
| | TERMINIA CONTRACTOR |
| 001010 | LPLLLIOLAF AGE IOI |

まず、簡単に情報の読み方を説明します。各情報はフィールド名、その位置、説明という基本構成をしており、 "/"で始まる行はその上の行の続きであることを示しています。

位置は[start]..[end]という形式が基本形でエントリの何bp(start)~何bp(end)にこのフィールドがアノテートされていることを示します。派生形として、[start]..>[end]や[start]<..[end]といった形式があり、少なくともこのフィールドが [start]から[end]までは続いているがさらに下流や上流に続いていることを示しています。

説明は基本的に[a]=[b]という形式をとっており、[a]という特徴は[b]です。と読みます。

実際に見てみましょう。

| gene | 7cTone= LANL LATT39-020 606>1430 | |
|------|-------------------------------------|--|
| | /gene="PANX1" | |

geneフィールドから、このエントリの606bp~1430bpにPANX1という遺伝子が含まれていることがわかりま す。また、"...>1430"と書かれているのでこの遺伝子はさらに下流まで続いており、一部分がエントリに含ま れていることがわかります。同時にstart側には<がついていないので、このエントリの606bpが遺伝子の5'端 ということもわかります。

| <u>exon</u> | 6061170 /gene="PANX1" | |
|-------------|--------------------------|--|
| | /number=1 | |

exonフィールドから、このエントリの606bp~1170bpに1番目のエクソンが含まれていることがわかります。

| CDS | 990>1170 |
|-----|--|
| | /gene="PANX1" |
| | /note="putative gap junction protein; contains four |
| | putative transmembrane domains; similar to innexin gap |
| | junction gene family" |
| | /codon start=1 |
| | /product="pannexin 1" |
| | /protein_id="AAL06604.1" |
| | /db_xref="GI:15808667" |
| | /translation="MAIAQLATEYVFSDFLLKEPTEPKFKGLRLELAVDKMVTCIAVG |
| | LPLLLISLAFAQEISI" |

CDSフィールドからは、990bp~1170bpにCDSの頭の部分が含まれており、さらに下流に続いていることがわかります。exonフィールドの情報と組み合わせて考えると、ファーストエクソンは606bp~1170bpであり990bpからがスタートコドンで翻訳領域となり、606bp~989bpは5'UTRになることがわかります。また、翻訳されたアミノ酸配列は AAL06604として登録されていることも分かります。

■ 解答1-4_

検索結果を絞り込む(その2)

再度検索結果を絞り込んで見ましょう。pannexinというキーワードで検索し、Limitsタブをクリックして絞 込みのページに移動します。

| All Fields 🛛 💌 | |
|--------------------------------------|----------------|
| Exclude: STSs Sworking draft DTPA | patents |
| Volecule: | Gene Location: |
| Any 💌 | Any |
| Segmented Sequences: | Only from: |
| Any 💌 | Any 💌 |
| Published in the last: | |
| Any Date 💌 | |
| Modified in the last: | |
| Any Date | |

まず、Only fromよりソースをRefSeqに設定し、

| nces: | -Only from: Any |
|-------|--------------------------|
| st: | Any RefSeq GenBank |
| t: | DDBJ PDB |

MoleculeよりmRNAに設定します。

| Any 🗸 | | Any 🗸 |
|--|---|----------------------|
| Any Genomic DNA/RNA es: mRNA rRNA cRNA | | Only from: RefSeq |
| Any Date | ~ | |

生物種をヒトに限る方法ですが、上部の検索窓に"pannexin"の後ろに" AND human[organism]"と入力して AND検索により実現します。

| * | for | pannexin AND human[organism] | Go | Clear |
|---|-----|------------------------------|----|-------|
| | | | | |

すると下図のように3件のヒットが確認できます。

| | 3 nucleotide sequences. Nucleotide [3] | |
|----------|---|----------------------------|
| Display | / Summary 💌 | Show 20 💌 Sort by 💌 Send t |
| All: | 3 Bacteria: 0 RefSeq: 3 mRNA: 3 🛠 | |
| Iten | ns 1 - 3 of 3 | One page. |
| This | search in Gene shows 8 results including: | |
| PA | NX1 (Homo saniens): pannexin 1 | |
| PA | NX2 (Homo sapiens): pannexin 1 | |
| PA | NX3 (Homo sapiens): pannexin 3 | |
| 1: | NM 015368 Reports | Order cDNA clone, Links |
| | Homo sapiens pannexin 1 (PANX1), mRNA | Α |
| | g 39995063 ref NM_015368.3 [39995063] | |
| | NM 052959 Reports | Links |
| 2: | Homo sapiens pannexin 3 (PANX3), mRNA | д. |
| 2: | gi 39995066 ref NM_052959.2 [39995066] | |
| 2: 3: | gi[39995066]ref[NM_052959.2][39995066] NM 052839 Reports | Links |

検索結果をFASTAファイルで保存する

続いて絞り込まれた3件のエントリをFASTA形式で保存してみましょう。画面上部Display横のプルダウンメニ ューから"FASTA"を選択し、



multi FASTA形式で表示させた後、

| Display | FASTA | 💙 Show | v 20 | 🖌 Send to | ▼ |
|--|---|---|--|--|---|
| | | | | | Item 1 - 3 of 3 |
| | | | | | |
| □1: <u>N</u> | I <u>M 015368</u> . R | eports Ho | mo sa | piens pann | [gi:39995063] |
| >gi 3999 GGGCGGCG GCGGGAGA AGGAAGCO CAGCGGACO AAGCGAAA CCGGACCI CTCGGATI | 15063 ref NM GCGGAGGGGCAGO GGCGCGGATCCC GGCGCGCGGG GTCGCTGGGGAGCO GCCGCGCCCCCCCGAGCO ICCTGGTCGAGCO ITCTTGCTGAAGO | 015368.3 GCCAGAGG GAGTGCCGCCC TCCCGCCC TGAGGCAC GCCGGTGA GCCCGCGCGC GAGCCCACG | Homo GAAGCG GCGCGG CGCCCC CGAGAC CTGGGT CGCAGC GAGCCC | Sapiens pa CTTTGTTCCGC CCCGGGGACTT GCCCCGCCGGC ACAAAGGCAGG GAAGGCGCCGC CATGGCCATCG AAGTTCAAGGG | nnexin 1 (PANX1), mRNA GCGTGGTTCCCGCGCCTGGGGGTGC GCACGGCGTGCGCGGGGTGCAACCGC CGCGATGCCGCAGCGACGCGAGAGCC CGGGATGCGGGAGCAGGCAAAGGGA GCACCTTTCCCCGACGCCGGCTGTAC CTCAACTGGCCACCGAGTACGTGTT GCTGCGACTGGAGCTGGCTGTGGAC |
| AAGATGGI CGATIGGI ATATIGCI CATAAGTI TCGCAGCI TGCAATTA ACCGAGAA ACTIGAAQ CATIATAQ | ICACGTGCATTGC IACACAGATAAGC IGGGCGGCTGTTC ITTTCCCCTACAT IGCTCCTCATATT AAGGCTGCAAAGA ACTTAGGGCAAAA BACAAAGAAAAAT SACAAAGAAAAAA | CGGTGGGGGC CTGTTTCTC CAGCAGAAG CCCTGCTGC TGCTCAGA CTGCTCAGA CTGCGCGCT CTCTAATAA | FGCCCC FCCAAG AACTCA FCTTTG CTTGAA GACCTT AGGTAT FTTAAT | TGCTGCTCATC TCTTCTCCCT CTGCAGAGGGA COGATCCTCCTG GATCATCATGG GACATGAGGA CTGAAAGCCAC CATCAAGTACA | TCGCTGGCCTTCGCGCAGGAGATCT GGCGTCAGGCTGCCTTTGTGGATTC GTCTGGAAACCTCCCACTGTGGCTG TACCTGCCCCCGCTGTTCTGGCGTT AAGAACTTGACAAAGTTTACAACCG TGGAGCCTGCTCAGTTCCAGGTGTT TTCAAGTACCCAATTGTGGAGCAGT TTAGCTGCCGCCTGCTGACACTCAT GLCCTCACTCTCAGACACTUGG |
| TGCAGCA1 CCGTGGGG CTACACGG GATGTTC1 | ICAAATCAGGGAT CATCTTCCAGTTC CTGTTTGTTCCAT IGCATTTCAAATC | CCTGAGAA CTCAGTGT TCCGACAG TGAAGGGT | ACGACA CATTAA AAGACA ACAACG | GCACCGTGCCC CCTTGTGGTTT GATGTTCTCAA | GATCAGTTTCAGTGCAAACTCATTG ATGTCCTGCTGGCTCCCGTGGTTGT AGTGTACGAAATCCTCCCCACTTTT TACAATCTCTTCTGGAGGGAAAATA |
| CCCAATGO GCAGAGAGA AAGCAAAA GAGCTGTA TGGTTTGO TCAACAAA GCCAGTAA | AGI CHANGI CA IAU DTACTOCTGACAA IGAGAGAGGAGCA IAATGGAGAGAAGAA IAATCTGTGACTI CAATAAATGGTTC AAGGTTATGGAAG IACTATGTGAAGI IACACCCTCCCAA | CAAGIGICI ACCTTGGC GGGGAACC GAATGCCCG CTGCGACA CTGGGGGAGA CTTGGTGGA GAATGGTTT CCTCATCA GCAATACCT | TAAGGT ATGATC AGACGG ACAGAG IGGGAT GATACT ATGAAC ATGAAC ITCIGT | ACT GGAGAATA CAGAGCTCCAA CAGAGCTCCAA CACTTCTGGATT TTAATTTGGCT GAGCATGTCTT CCCATAGGA AAGCAGAAAGA AACCAGTGGTGA | TTAAGAGGGTGGTGGTCAGGGGATCGA TGTTGATGGCAAAACTCCCCATGTCT GGTATGAACATAGACAGTGAAACTA CTTCTTGCTGATGATTTTTTCCTT AAAGCACCCCTGTTGGTTTCACAGC ATTGAGTCCCTAATGGAAATGGTGA AGCACCTGAGAGATAGTAAACTGCA CAAGAACAATTAGTCAAGAGCAGTA GACAAGACCCAGAGCTACTGGAAAA |
| CAAGCACI | ITTGGAAGATTT(IGCCCTTAAAGA(| TTTTGTTT | TCATGG | AGCATTCACTG | GTCAGGGTATAATTTAACGTGAGTT GCTAATAATCCATAGGTCGACCTAT |

Showの二つとなりのプルダウンメニューから"File"を選択し、適切な名前をつけることで検索結果をFASTA形式で保存することができます。

| Display | FASTA | ~ | Show | 20 | ~ | Send to | * | |
|--------------|--------------------|---------|------|-------|-----|-------------------|-----|-------------|
| | | | | | | Send to Text | | Item 1 - 3 |
| □1: <u>N</u> | <u>VM 015368</u> . | Reports | Hom | io sa | pie | File Clipboard |] | 9995063] |
| >gi 399 | 95063 ref | NM_0153 | 68.3 | Home | | apiens pa | nne | xin 1 (PANX |

■ 解答1-5___

エントリの更新履歴を見る

検索結果に戻ってください。

| All: 3 Bacteria: 0 RefSeq: 3 mRNA: 3 Items 1 - 3 of 3 Items 1 - 3 of 3 One page. This search in Gene shows <u>6 results</u> , including: PANX1 (Homo sapiens): pannexin 1 PANX2 (Homo sapiens): pannexin 2 PANX3 (Homo sapiens): pannexin 3 I: NM 015368 Reports Order cDNA clone, Links Homo sapiens pannexin 1 (PANX1), mRNA gi[39995063] Links Pano sapiens pannexin 3 (PANX3), mRNA gi[39995066] ref[NM_052959.2][39995066] Links Borno sapiens pannexin 3 (PANX3), mRNA gi[39995066] ref[NM_052959.2][39995066] Links | Display | Summary | | * | Show | 20 | Sort by | Send to |
|--|-------------------|--|---|--|------|----|---------------|-----------------------------|
| Items 1 - 3 of 3 One page. This search in Gene shows <u>6 results</u> , including: PANX1 (Homo sapiens): pannexin 1 PANX2 (Homo sapiens): pannexin 2 PANX3 (Homo sapiens): pannexin 3 I: NM 015368 Reports Order cDNA clone, Links Homo sapiens pannexin 1 (PANX1), mRNA gi[39995063] Links Panx 3 (Homo sapiens pannexin 3 (PANX3), mRNA gi[39995066] ref[NM_015368.3][39995066] Links Homo sapiens pannexin 3 (PANX3), mRNA gi[39995066] ref[NM_052959.2][39995066] Links | All: 3 | Bacteria: 0 | RefSeq: 3 | mRNA: 3 🛠 | | | | |
| This search in Gene shows 6 results, including: PANX1 (Homo sapiens): pannexin 1 PANX2 (Homo sapiens): pannexin 2 PANX3 (Homo sapiens): pannexin 3 1: NM 015368 Reports Homo sapiens pannexin 1 (PANX1), mRNA gi[39995063]ref[NM_015368.3][39995063] 2: NM 052959 Reports Homo sapiens pannexin 3 (PANX3), mRNA gi[39995066]ref[NM_052959.2][39995066] 3: NM 052839 Reports Links | Item | s 1 - 3 of 3 | | | | | On | e page. |
| PANX1 (Homo sapiens): pannexin 1 PANX2 (Homo sapiens): pannexin 2 PANX3 (Homo sapiens): pannexin 3 I1: NM 015368 Reports Homo sapiens pannexin 1 (PANX1), mRNA gi[39995063]ref[NM_015368.3][39995063] I2: NM 052959 Reports Homo sapiens pannexin 3 (PANX3), mRNA gi[39995066]ref[NM_052959.2][39995066] I3: NM 052839 Reports Links | This | search in Gene | shows <u>6 res</u> | ults, including: | | | | |
| PANX2 (Homo sapiens): pannexin 2 PANX3 (Homo sapiens): pannexin 3 1: NM 015368 Reports Homo sapiens pannexin 1 (PANX1), mRNA gi[39995063]ref[NM_015368.3][39995063] 2: NM 052959 Reports Homo sapiens pannexin 3 (PANX3), mRNA gi[39995066]ref[NM_052959.2][39995066] 3: NM 052839 Reports Homo sapiens pannexin 2 (PANX2), mRNA | PA | NX1 (Homo sapier | ns): pannexin 1 | | | | | |
| PANX3 (Homo sapiens): pannexin 3 1: NM 015368 Reports Order cDNA clone, Links Homo sapiens pannexin 1 (PANX1), mRNA gi[39995063]ref[NM_015368.3][39995063] Links 2: NM 052959 Reports Links Homo sapiens pannexin 3 (PANX3), mRNA gi[39995066]ref[NM_052959.2][39995066] Links 3: NM 052839 Reports Links Homo sapiens pannexin 2 (PANY2) mRNA Links | PA | NX2 (Homo sapier | ns): pannexin 2 | | | | | |
| 1: NM 015368 Reports Order cDNA clone, Links Homo sapiens pannexin 1 (PANX1), mRNA gi[39995063]ref[NM_015368.3][39995063] 2: NM 052959 Reports Links Homo sapiens pannexin 3 (PANX3), mRNA gi[39995066]ref[NM_052959.2][39995066] Links 3: NM 052839 Reports Links Homo sapiens pannexin 2 (PANX2), mRNA Links | PA | NX3 (Homo sapier | ns): pannexin 3 | | | | | |
| gi[163659919]ref NM_052839.3][163659919] | □1: □2: □3: | NM 015368 Rep Homo sapiens gi 39995063 re NM 052959 Rep Homo sapiens gi 39995066 re NM 052839 Rep Homo sapiens gi 163659919 re | orts pannexin 1 (I f NM_015368 orts pannexin 3 (I f NM_052959 ports pannexin 2 (I ef NM_05283 | PANX1), mRNA 3 [39995063] PANX3), mRNA 2 [39995066] PANX2), mRNA 9.3 [163659919 | | On | der cDNA clor | ie, Links Links Links |
| | | | | | | | | |

エントリNM_015368をクリックして詳細を見てみましょう。



LOCUS行の情報から最新の更新日は2008年10月12日であることがわかります。

続いて、更新履歴を見てみましょう。更新履歴は、エントリ番号の右にある"Reports"をクリックして表示されるメニューから"Revision History"を選ぶことで表示されます。

| □1: <u>NM(</u> | 015368. Repor | s Homo sapiens Reports | pann[gi:39995 | i063] | | Oriclo |
|----------------|---------------------------|---------------------------|-----------------|------------|-----------------|--------|
| Comment | Features Se | ▶ ASN.1 | | | | |
| | | ▶ XML | | | | |
| LOCUS | NM_015368 | Summary | 2 bp mRNA | linear | PRI 12-0CT-2008 | |
| DEFINITION | Homo sapien | ▶ Brief | NXI), mRNA. | | | |
| VERSION | NM 015368.3 | ▶ FASTA | | | | |
| KEYWORDS | | ▶ TinySeq XML | | | | |
| SOURCE | Homo sapien | ▶ GenBank | | | | |
| URGANISM | Homo sapien Eukaryota; | ▶ INSDSeq XML | a; Craniata; Ve | ertebrata | ; Euteleostomi; | |
| | Mammalia; E | GenBank(Full) | ntoglires; Prim | mates; Hap | olorrhini; | |
| DECEDENCE | Catarrhini; | ▶ GI List | • | | | |
| AUTHORS | Iglesias,R. | ▶ Graphic | que,A., Alberto | o,A.P., Da | ahl,G., | |
| | Spray,D.C. | Revision History | | | | |
| TITLE | PZX/ recept | or-Pannexinl cor | lex: pharmacol | logy and s | signaling | |

クリックしてみましょう。

| Revision history for <u>NM_015368</u> | | | | | | | | |
|---------------------------------------|-----------|-----------------------------|--------------------|---------|----------------------|--|--|--|
| GI | Version | Update Date | Status | Ι | II | | | |
| 39995063 | 3 | Oct 12 2008 10:43 AM | Live | \odot | 0 | | | |
| 39995063 | 3 | May 11 2008 10:51 AM | Dead | 0 | $\overline{\bullet}$ | | | |
| 39995063 | 3 | Mar 27 2008 12:39 AM | Dead | 0 | 0 | | | |
| 39995063 | 3 | Feb 11 2008 12:49 AM | Dead | 0 | 0 | | | |
| 39995063 | 3 | Dec 9 2007 11:09 AM | Dead | 0 | 0 | | | |
| 39995063 | 3 | Sep 3 2007 3:02 AM | Dead | 0 | 0 | | | |
| 39995063 | 3 | <u>Jun 26 2007 10:53 PM</u> | Dead | 0 | 0 | | | |
| 39995063 | 3 | <u>Jun 22 2007 11:28 PM</u> | Dead | 0 | \bigcirc | | | |
| 39995063 | 3 | <u>Mar 25 2007 11:40 AM</u> | Dead | 0 | \bigcirc | | | |
| 39995063 | 3 | <u>Jan 14 2007 12:11 PM</u> | Dead | 0 | \bigcirc | | | |
| 39995063 | 3 | Dec 3 2006 10:40 AM | Dead | 0 | \bigcirc | | | |
| 39995063 | 3 | Nov 17 2006 11:49 PM | Dead | 0 | \bigcirc | | | |
| 39995063 | 3 | <u>Nov 3 2006 1:07 PM</u> | Dead | 0 | \bigcirc | | | |
| 39995063 | 3 | Oct 29 2006 10:26 AM | Dead | 0 | \bigcirc | | | |
| 39995063 | 3 | Oct 15 2006 10:59 AM | Dead | 0 | \bigcirc | | | |
| 39995063 | 3 | Aug 6 2006 2:43 PM | Dead | \circ | \bigcirc | | | |
| 39995063 | 3 | <u>Jan 29 2006 11:11 AM</u> | Dead | 0 | \bigcirc | | | |
| 39995063 | 3 | Sep 24 2005 6:44 AM | Dead | 0 | \circ | | | |
| 39995063 | 3 | <u>Jul 8 2005 1:43 PM</u> | Dead | \circ | \bigcirc | | | |
| 39995063 | 3 | Apr 23 2005 9:43 AM | Dead | 0 | \circ | | | |
| 39995063 | 3 | <u>Mar 2 2005 1:02 PM</u> | Dead | 0 | \circ | | | |
| 39995063 | 3 | Oct 27 2004 9:40 PM | Dead | 0 | \bigcirc | | | |
| 39995063 | 3 | Aug 23 2004 3:34 PM | Dead | \circ | \bigcirc | | | |
| 39995063 | 3 | <u>Jan 27 2004 9:22 AM</u> | Dead | \circ | \bigcirc | | | |
| 39995063 | 3 | Dec 23 2003 12:52 PM | Dead | 0 | \circ | | | |
| 39995063 | 3 | Dec 17 2003 5:18 PM | Dead | 0 | \bigcirc | | | |
| 29837657 | 2 | Oct 5 2003 3:40 PM | Dead | \circ | \bigcirc | | | |
| 29837657 | 2 | <u>Sep 7 2003 4:39 PM</u> | Dead | 0 | \bigcirc | | | |
| 29837657 | 2 | <u>May 7 2003 4:21 AM</u> | Dead | 0 | \bigcirc | | | |
| 29837657 | 2 | <u>Apr 15 2003 3:31 AM</u> | Dead | 0 | \bigcirc | | | |
| 7662507 | 1 | <u>Apr 7 2003 11:18 AM</u> | Dead | 0 | \circ | | | |
| 7662507 | 1 | Dec 23 2002 12:19 PM | Dead | 0 | \bigcirc | | | |
| 7662507 | 1 | <u>Nov 5 2002 10:47 AM</u> | Dead | 0 | \bigcirc | | | |
| 7662507 | 1 | Aug 27 2002 3:38 PM | Dead | 0 | 0 | | | |
| 7662507 | 1 | <u>May 14 2002 2:03 PM</u> | Dead | 0 | \bigcirc | | | |
| 7662507 | 1 | Feb 11 2002 3:04 AM | Dead | 0 | \bigcirc | | | |
| 7662507 | 1 | May 24 2001 4:16 AM | Dead | 0 | \bigcirc | | | |
| 7662507 | 1 | Nov 2 2000 6:44 AM | Dead | 0 | 0 | | | |
| 7662507 | 1 | Aug 18 2000 10:14 AM | Dead | 0 | \bigcirc | | | |
| 7662507 | 1 | Apr 28 2000 9:30 AM | Dead | 0 | 0 | | | |
| Accession N | VI 015368 | was first seen at NCBI on A | Apr 28 2000 9:30 / | AM | | | | |

極めて多くの更新がなされていることがわかります。配列が変わることによるVersionは3回しか更新されていませんが、FEATURESなどに変更があったため非常に多くの更新があったことがわかります。その日付も合わせてわかります。

解答・解説2

課題1の手順でpannexinのところをFOXP2に変えて同じ事を実行してみてください。

原文更新日: 2005年2月15日 日本語版更新日: 2008年10月31日

All Rights Reserved, Copyright(C) 1997-2008 Japan Science and Technology Agency(JST)

