

UCSC genome browser

UCSC Genome Browserの豊富なツールを理解する

理化学研究所 生命医科学研究センター
生命医科学大容量データ技術研究チーム
森岡勝樹

本レクチャーの内容

- UCSC Genome Browserの豊富なツールを理解する
 - パワーポイントでツールたちを紹介
- BLAT、In-Silico PCRを使って配列を検索する/Gene Sorterを使って発現データを解析する
 - 資料に沿って実際に使ってみる
- Track Hubを使ってオリジナルデータと公共データを比較した図を作成する
 - 資料に沿って実際に使ってみる

いまさらPCRprimer設計？！

BOCCHI THE FROG PRINCESS ... tg
@

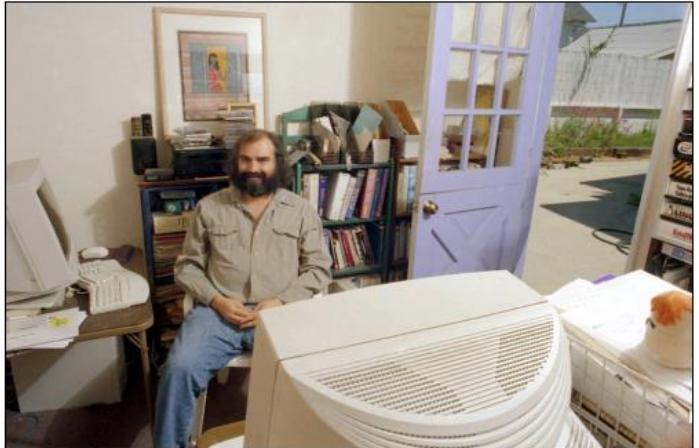
[ゆるぼ] いまだに手作業でPCRプライマーを設計

しない (受託会社の設計プログラム)	3.7%
する (フリーサイト)	33.3%
する (フリーハンド)	22.2%
かえるしとデートしたい	40.7%

27票・残り1日

午後6:38・2023年1月18日・341件の表示

UCSC genome browserの歴史



Jim Kent in his garage sitting next to the computer where he wrote the 10,000 lines of computer code to assemble the first draft assembly of the human genome.

<https://genome.ucsc.edu/goldenPath/history.html>

NEWYORK TIMES

<https://www.nytimes.com/2001/02/13/science/reading-the-book-of-life-grad-student-becomes-gene-effort-s-unlikely-hero.html>

2000年7月7日、世界で初めてヒトのゲノム配列を公開したゲノムブラウザー

ヒトゲノム配列の解読のタイムラインが面白い

- 国際ヒトゲノムプロジェクト（IHGP）の競合であつたセレラ社がそれまでIHGPが計画していたスピードの約半分でヒトゲノム解読を始める。
- 6/22 Kentが、100台のコンピュータを使って、40万コンティグから、1st アッセンブリの作成を終了。
- 6/25の夜 Celera社、1st アッセンブリ完了
- 6/26 ホワイトハウスで、IHGPヘッドのコリンズ、セレラ社代表のベンダーの両者がヒトゲノム配列のソフトアッセンブリ完了を発表

UCSC genome browserのミラーサイト

- 本家: <https://genome.ucsc.edu>
- Europe: <https://genome-euro.ucsc.edu>
- Asia: <https://genome-asia.ucsc.edu>

2016年6月17日
理化学研究所

ゲノムブラウザ・データベースの公式ミラーを開設

理化学研究所(理研)情報基盤センターは6月16日、カリフォルニア大学サンタクルス校(UCSC)と共同し、広く用いられているUCSCゲノムブラウザ・データベースの日本ミラー(英語)を開設しました。欧州に次ぐ二番目の公式ミラーになります。

UCSCゲノムブラウザ・データベースにはおよそ100種もの生物に関するリファレンスゲノム配列が登録されており、[1000人ゲノムプロジェクト\(英語\)](#)、[ENCODE \(Encyclopedia of DNA Elements\) プロジェクト\(英語\)](#)、[国際がんゲノムコンソーシアム\(CGC\)\(英語\)](#)、[国際ヒトエピゲノムコンソーシアム\(IHEC\)\(英語\)](#)、さらに理研が主催する国際FANTOM5プロジェクト*を含むさまざまなデータを自在に閲覧することができる世界で最大規模のゲノム情報データベースです。

ライフサイエンス分野の研究者のみならず、ゲノム情報を用いた医療やサービスを受ける私たちにとっても、近年ますます盛んに蓄積され公開されているゲノミクスデータ(ゲノムDNA塩基配列やその変動、個々のDNA塩基配列の機能的情報)を読み解くことは大変重要です。日本を含むアジア地域からの高速なアクセスが可能になることで、ゲノミクス研究の更なる展開が期待できます。

The screenshot shows the UCSC Genome Browser interface for the Human genome. The main window displays a genomic track for chromosome 9, specifically the region chr9:133,252,000-133,280,861. The track includes various tracks for GENCODE, RefSeq, and OMIM Allelic Variant SNPs. A sidebar on the right lists "Third Party Mirrors" including "US Server", "European Server", and "Japan Server" (which is checked). The URL in the browser's address bar is genome-asia.ucsc.edu/cgi-bin/hgTracks?db=hg.

*国際FANTOM5プロジェクト



注: 本日は、弊所のアジアミラーをご利用ください。

UCSC genome browserの現在

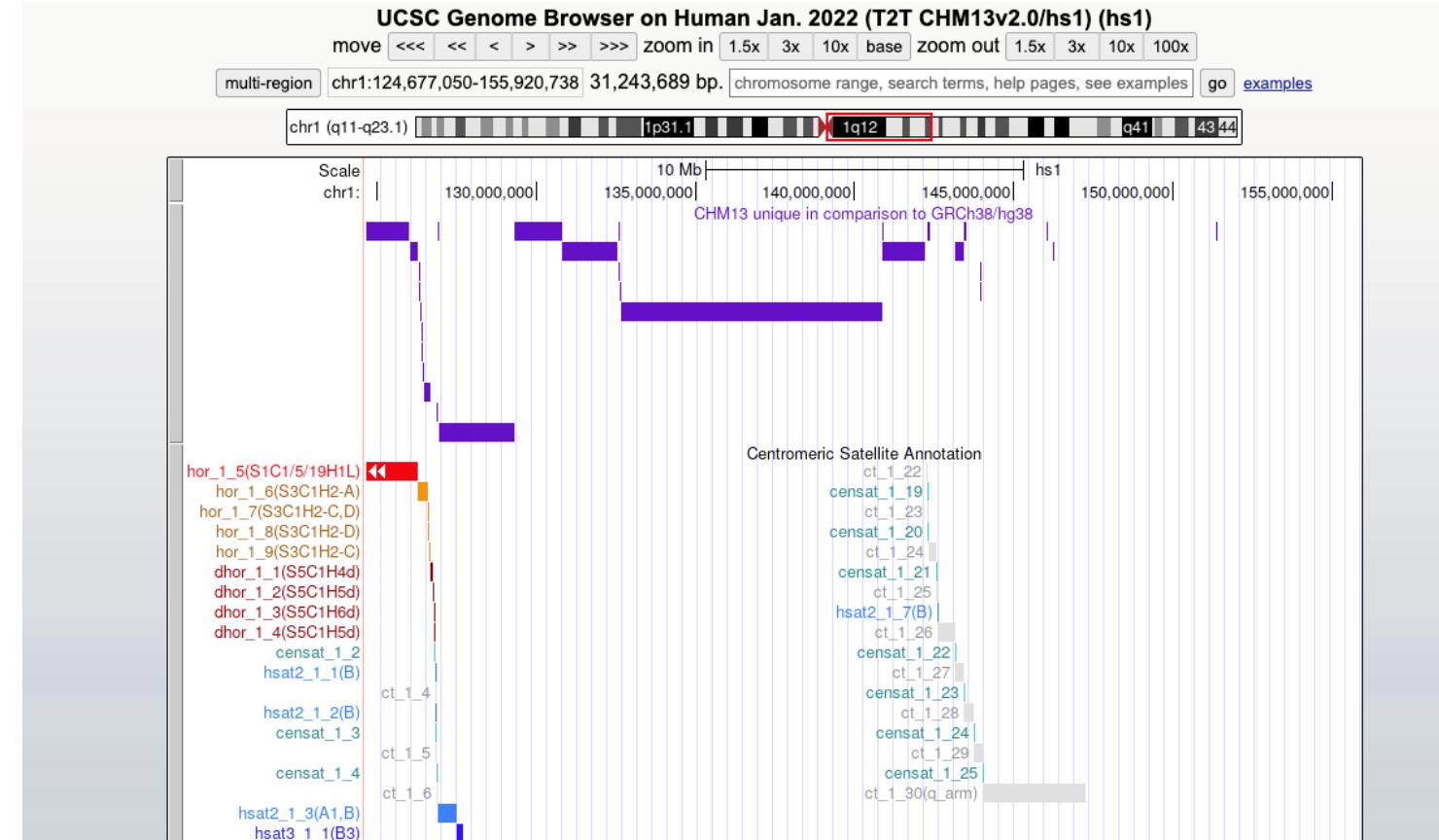
UCSC genome browser

- ヒトをはじめとする100種類以上のゲノム地図を見ることができる。
- 遺伝子位置や塩基配列に紐づいた注釈（アノテーション）情報の検索、表示、抽出ができる。
- ゲノム配列を扱うための各種の細かいスクリプトや、バイオインフォマティクスツールを利用したり、ダウンロードできる。

アップデートが頻繁 (33 updates in 2022!)

- 2022.12.22 470種類の哺乳類の比較ゲノムによる相同性領域のトラック ([470-way Multiz Alignment & Conservation](#)) の追加。
- 2022.12.13 GENCODEの新しいアノテーション (human V42 and mouse VM31) を追加
- 2022.10.24 HGNC and PanelAppの2つのトラックが追加
- 2022.10.19 hs1とmpxvRiversのアッセンブリを追加
- And more!

Human genome hs1



眺めてみたい人もいるかもしれないで、後ほどご案内

最新updateを一つ紹介: 検索機能が強化

TP53というキーワードで検索

Search across the Genome Browser site

Search Results on hg38 (Human Dec. 2013 (GRCh38/hg38))

Search: Human Dec. 2013 (GRCh38/hg38) for TP53

Use the tree to hide/show results from only these categories. Hover your mouse over each category for an explanation:

- GENCODE V41 (351 results)
- hg38 Track Data (672 results)
 - Visible Tracks (59 results)
 - RefSeq Curated (59 results)
 - Currently Hidden Tracks (613 results)
 - Genes and Gene Predictions (575 results)
 - RetroGenes V9 (11 results)
 - Other RefSeq (68 results)
 - MANE v1.0 (3 results)
 - HGNC (13 results)
 - NCBI RefSeq (216 results)
 - All GENCODE (27 results)
 - Phenotype and Literature (1 results)
 - GeneReviews (1 results)
 - mRNA and EST (24 results)
 - Human mRNAs (24 results)
 - Variation (13 results)
 - gnomAD Variants (12 results)
 - hg38 Track Labels/Descriptions (1 results)
 - Public Hubs (2 results)
 - Help Pages (1 results)

Gencode Genes:

 - TP53 (ENST00000269305.9) - chr17:7668421-7687490 - TF
 - TP53TG1 (ENST00000661943.3) - chr7:87322943-8734555
 - TP53TG1 AB007455 ENST00000661943.1
 - TRIAP1 (ENST00000546954.2) - chr12:120443964-1204463
 - apoptosis 1 TRIAP1 mRNA from RefSeq
 - TIGAR (ENST00000179259.6) - chr12:4321213-4360028
 - phosphatase TIGAR mRNA from RefSeq NM_020375
 - TP53TG3F (ENST00000617705.1) - chr16:33459650-33462
 - F TP53TG3F transcript variant 2
 - TP53TG5 (ENST00000372726.5) - chr20:45372557-453783
 - from RefSeq NM_014477 C20orf10 ENST00000372726.1
 - ENSG00000275869 (ENST00000621028.1) - chr16_KI270728v1_random:933862-936467
 - ENSG00000275869 ENST00000621028.1 Homo sapiens TP53-target gene 3 protein LOC102723713 mRNA from RefSeq NM_001368231 NM_001368231
 - ENSG00000274175 (ENST00000616638.1) - chr16_KI270728v1_random:17234-19833
 - ENSG00000274175 ENST00000616638.1 Homo sapiens TP53-target gene 3 protein LOC102723655 mRNA from RefSeq NM_001368237 NM_001368237
 - RPRM (ENST00000325926.4) - chr2:153477338-153478762
 - RPRM ENST00000325926.4 Homo sapiens reproto TP53 dependent G2 arrest mediator homolog RPRM mRNA from RefSeq
 - RRM2B (ENST00000251810.8) - chr8:102204502-102238961
 - RRM2B ENST00000251810.8 Homo sapiens ribonucleotide reductase regulatory TP53 inducible subunit M2B RRM2B transcript variant 1 mRNA from

Show 341 more matches for Gencode Genes

NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*):

 - NM_001126116.2 - chr17:7668421-7675244
 - NM_001276698.3 - chr17:7668421-7675244
 - NM_001126117.2 - chr17:7668421-7675244
 - NM_001276699.3 - chr17:7668421-7675244
 - NM_001126115.2 - chr17:7668421-7675244
 - NM_001276697.3 - chr17:7668421-7675244
 - NM_001126118.2 - chr17:7668421-7687490
 - NM_001126114.3 - chr17:7668421-7687490
 - NM_001276696.3 - chr17:7668421-7687490
 - NM_001126113.3 - chr17:7668421-7687490

Show 75 more matches for NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*)

NCBI RefSeq genes, predicted subset (XM_* or XR_*):

指定したアセンブリについて、トラックデータ、トラック名、トラックの説明、パブリックハブのトラック名、パブリックハブの説明、ヘルプなど横断的に単語とのマッチングを見つけることができるようになった。

各ページ、データに幾つヒットしたのかを階層的に表示

UCSC genome browserの良し悪し

(主観に基づく)

良いところ

- ・ アップデートが頻繁で豊富な情報！
- ・ 豊富でシンプルなツールとその解説！
- ・ ゲノム地図上でデータ横断的に遺伝子情報を調べ、比較することができる。
- ・ カスタマイズすることで、論文に利用できるような図も作ることができる。

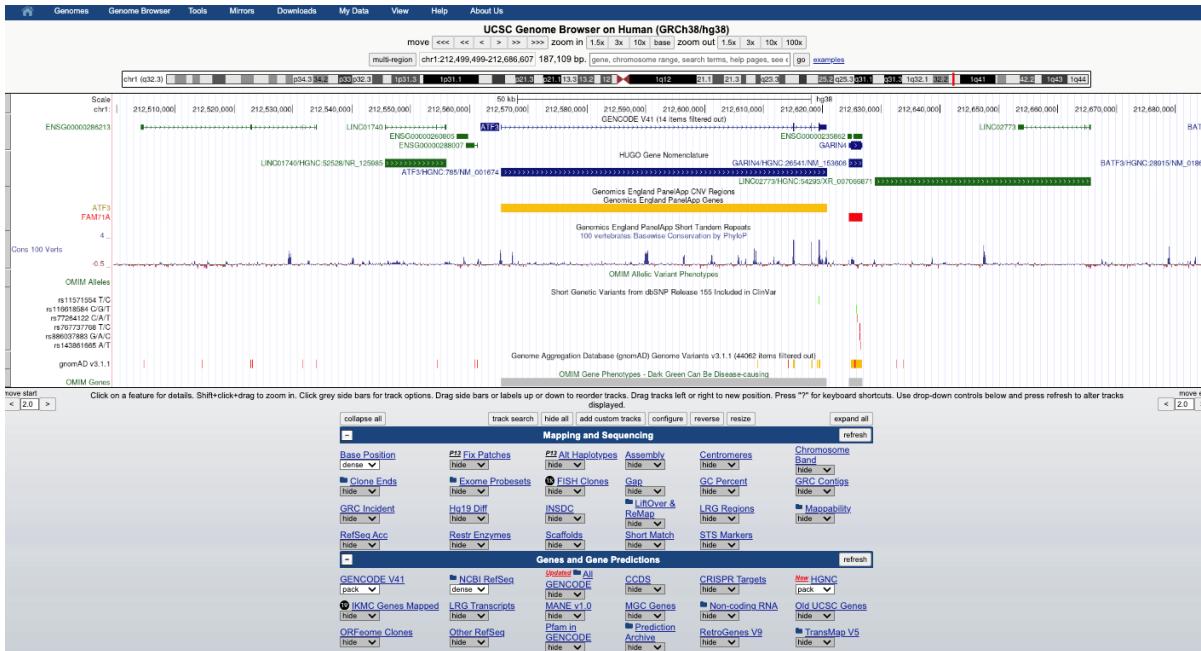
悪いところ

- ・ 見た目がtoo busy!!
- ・ インタラクティブに地図をスクロールできるものの、ラグがある。
- ・ オリジナルデータをインポートするには、ファイルを小さいサイズにする必要がある。

結論

慣れが必要

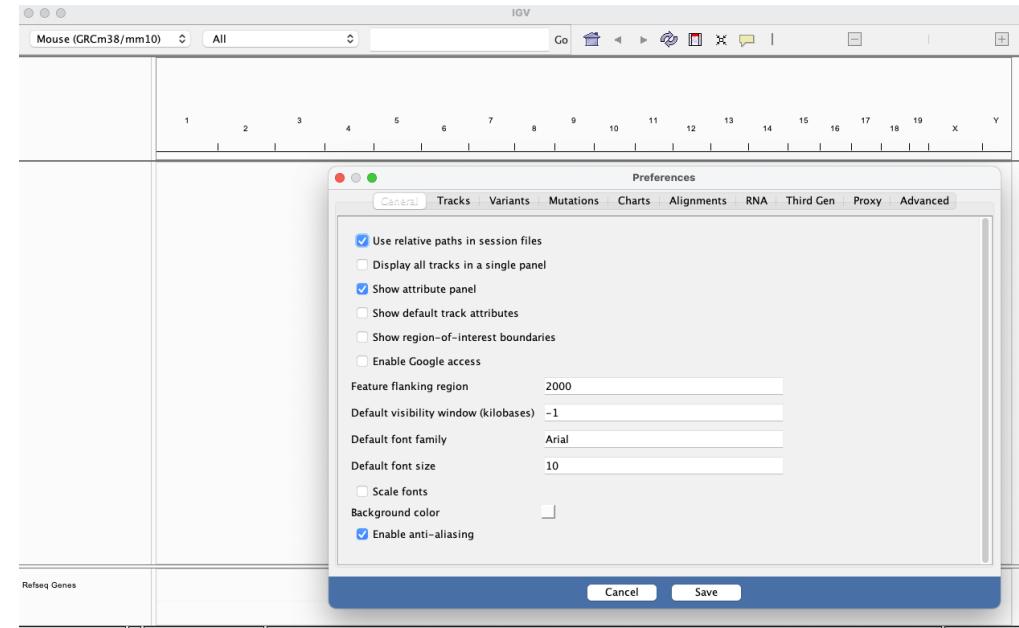
UCSC GB vs. IGV



- 左右スクロールにラグがある
- オリジナルデータをweb上にuploadする必要がある
- 直感的ではない（かも）
- 慣れるとカスタマイズできる
- 事前にたくさんのデータが用意されている

公共データとの比較図作成

取り扱えるファイル形式はほぼ同じ（IGVのtdfファイルは別）。状況による使い分け。



- ファイルサイズが大きくなると重たくなる
- ファイル数が多いと折り返す
- 比較用のデータをほぼ全て用意する必要がある
- 直感的、一方で実現したい可視化に手が届かないこともある
- Uploadの必要がない

個人的にはオリジナルファイルのsanity checkに利用価値大

UCSC genome browserのファイル形式

General formats

- Axt format
- BAM format
- BED format
- **BED detail format**
- bedGraph format
- **barChart and bigBarChart format**
- bigBed format
- **bigGenePred table format**
- **bigPsl table format**
- **bigMaf table format**
- **bigChain table format**
- **bigNarrowPeak table format**
- **bigLolly table format**
- **bigWig format**
- **Chain format**

座標のフォーマット

座標にスコアを載せたフォーマットで、ブラウザ上でグラフ表示をするための形式

変換

座標にスコアを載せた
フォーマット、バイナリー

- CRAM format
- GenePred table format
- **GFF format**
- **GTF format**
- HAL format
- Hic format
- Interact and bigInteract format
- Longrange longTabix format
- MAF format
- Microarray format
- Net format
- Personal Genome SNP format
- **PSL format**
- **VCF format**
- **WIG format**

遺伝子構造などのアノテーションを記載する形式。GFFの拡張形式としてGTFがある。

BLATの結果の一行出力形式

バリエント情報を記載する一般的な形式

ENCODE-specific formats

- ENCODE broadPeak format
- ENCODE gappedPeak format
- ENCODE narrowPeak format
- ENCODE pairedTagAlign format
- ENCODE peptideMapping format
- ENCODE RNA elements format
- ENCODE tagAlign format

BLATを行うときにデータベースとなる配列情報（合計4Gbまで）をコンパクトにして検索可能にしたフォーマット

Download-only formats

- **.2bit format**
- **.fasta format**
- .fastQ format
- .nib format

今日デモを行わないツールについての簡単な紹介

UCSC genome browserのツール群

Main tools

- [Genome Browser](#) • ゲノム地図の可視化ツール. さまざまな情報を地図上で閲覧、抽出ができるツール
- [BLAT](#) • 塩基配列を問い合わせ（クエリ）文字列として与え、ゲノムに対して相同性検索を行う
- [In-Silico PCR](#) • PCRプライマーペアをクエリとして、その配列で囲まれた塩基配列をfastaフォーマットで出力するツール
- [Table Browser](#) • データベース上にあるゲノムデータをはじめとするさまざまなアノテーション情報にアクセスして、必要な情報を部分抽出することが可能なツール
- [LiftOver](#) • ゲノム座標のバージョン間での対応関係に従って、与えられた配列を目的のゲノムバージョンへ変換することができる。
- [Gene Sorter](#) • ターゲットの遺伝子とその他の遺伝子間の発現相関を調べ、さまざまな情報で抽出するツール
- [Genome Graphs](#) • G-band上に、グラフを表示できるツール
- [Data Integrator](#) • クエリとして与えた座標、遺伝子名などの用語に対して、データベース内のアノテーション情報を連結したテーブルを作成するツール

UCSC genome browserのツール群

- UnShER
 - SARS-CoV-2のツールで、塩基配列を入れるとSAR-CoV-2の系統樹に配置してくれる。
- VisiGene
 - マウスとカエルのin situ hybridizationによる発現組織の可視化画像データベース
- Gene Interactions
 - PINやPubMed共起に基づく遺伝子間の関係をグラフで描くツール
- Variant Annotation Integrator
 - バリアントによるタンパク質の同義・非同義置換を予測するさまざまなツールの予測情報をまとめたデータベースを用いて、バリアントのアノテーションを行うツール
- DNA Duster and Protein Duster
 - DNAやタンパク質の配列の文字列中にあるファイル形式に合っていない文字を取り除くツール
- Phylogenetic Tree PNG Maker
 - 与えられた条件に従って系統樹を作成し、PNGファイルを作ってくれるツール
- Executable and Source Code Downloads
 - UCSC Genome Browserを開発する過程で利用されているスクリプト群、バイナリーファイルも用意されている
- その他
 - 非公式ツール群、RやPythonにUCSCのファイルを導入する時のパッケージや、BEDtoolsなどのUCSCのファイルを扱うツールなど

LiftOver

Lift Genome Annotations

This tool converts genome coordinates and annotation files between assemblies. The input data can be entered into the text box or uploaded as a file. For files over 500Mb, use the command-line tool described in our [LiftOver documentation](#). If a pair of assemblies cannot be selected from the pull-down menus, a sequential lift may still be possible (e.g., mm9 to mm10 to mm39). If your desired conversion is still not available, please [contact us](#).

Original Genome:
Human

Original Assembly:
Dec. 2013 (GRCh38/hg38)

New Genome:
Human

New Assembly:
Jan. 2022 (T2T-CHM13v2.0/GCA_009914755.4)

Minimum ratio of bases that must remap: 0.1

BED 4 to BED 6 Options

Allow multiple output regions:



0
0

Minimum hit size in query:

Minimum chain size in target:

BED 12 Options

Min ratio of alignment blocks or exons that must map: 1

If thickStart/thickEnd is not mapped, use the closest mapped base:

Paste in data below, one position per line. You can use the [BED format](#) (e.g. "chr4 100000 100001", 0-based) or the format of the position box ("chr4:100,001-100,001", 1-based). See the [documentation](#). We do not recommend liftOver for SNPs that have rsIDs. See our [FAQ](#) for more information.



Submit
Clear

Or upload data from a file ([BED](#) or chrN:start-end in plain text format):

ファイルを選択 選択されていません Submit File

Command Line Tool

To lift genome annotations locally on Linux systems, download the [LiftOver](#) executable and the appropriate [chain file](#). Run `liftOver` with no arguments to see the usage message. See the [LiftOver documentation](#).

Genome Graphs

Human Genome Graphs

clade: Mammal genome: Human assembly: Dec. 2013 (GRCh38/hg38) 2種類に色分け可能

graph: -- nothing -- in blue , -- nothing -- in red
upload import configure correlate significance threshold: 3.5 browse regions sort genes Gene sorterによるその領域の遺伝子抽出
オリジナルデータのファイルの入力

No graph data is available for this assembly. Upload your own data or import from a table or custom track.

Using Genome Graphs

- G-band上に、グラフを表示できるツール
- Gene Sorterと連携、SNP to gene抽出などに使えるか？

Input format

```
chr2 100100000 2.3
chr2 100100500 4.5
chr2 100101000 1.2
```

rs1007298	+1
rs1007863	+1
rs10154509	+1
rs10154678	+1
rs10154785	+1
rs1018448	+1
rs10212022	+1
rs1022478	+1
rs1042311	+1
rs1042435	+1

Data Integrator

Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us

Data Integrator Undo Redo

Select Genome Assembly and Region

group genome assembly
Mammal Human Dec. 2013 (GRCh38/hg38)

region to annotate
position or search term chr1:212565334-212620777 ← 検索座標を指定（遺伝子名OK）

Configure Data Sources

please add at least one data source

Add Data Source

track group track
Custom Tracks test 1 (ct_test1_8193) View table schema Add ← トラックグループとその中にあるトラックを指定
"Add"が表示されるかぎり追加可能

get more data:
track hubs custom tracks

Output Options

At least one data source must be selected

Select Genome Assembly and Region

group genome assembly
Mammal Human Feb. 2009 (GRCh37/hg19)

region to annotate
position or search term chr3:12393001-12475855

Configure Data Sources

[] GTEx Gene V8 View table schema
[] adipose tissue - omental fat pad signal View table schema
[] Caltech RNA-seq - GM78 1x75D - 1 View table schema
[] Caltech RNA-seq - GM78 1x75D - 2 View table schema
[] Caltech RNA-seq - H1ES 1x75D - 1 View table schema

Add Data Source

track group track view subtrack
Expression Expression View table schema Add

get more data:
track hubs custom tracks

出力結果

追加、連結したデータたち

```
# hgIntegrator: database=hg19 region=chr3:12393001-12475855 Tue Jan 17 21:47:19 2023
#gtexGeneV8.chrom gtxGeneV8.chromStart gtxGeneV8.chromEnd gtxGeneV8.name gtxGeneV8.score gtxGeneV8.strand gtxGeneV8.genelid gtxGeneV8.geneType gtxGeneV8.expCount gtxGeneV8.expScores hub_2441861_f354a5d08a4396adc2cce0a04fb70841.chrom
hub_2441861_f354a5d08a4396adc2cce0a04fb70841.chromStart hub_2441861_f354a5d08a4396adc2cce0a04fb70841.valueAverage wgEncodeCaltechRnaSeqGm12878R1x75dTh1014UMinusRawRep1V4.chrom
wgEncodeCaltechRnaSeqGm12878R1x75dTh1014UMinusRawRep1V4.chromStart wgEncodeCaltechRnaSeqGm12878R1x75dTh1014UMinusRawRep1V4.chromEnd wgEncodeCaltechRnaSeqGm12878R1x75dTh1014UMinusRawRep1V4.valueAverage wgEncodeCaltechRnaSeqGm12878R1x75dTh1014UMinusRawRep2V4.chrom
wgEncodeCaltechRnaSeqGm12878R1x75dTh1014UMinusRawRep2V4.chromStart wgEncodeCaltechRnaSeqGm12878R1x75dTh1014UMinusRawRep2V4.chromEnd wgEncodeCaltechRnaSeqGm12878R1x75dTh1014UMinusRawRep2V4.valueAverage wgEncodeCaltechRnaSeqGm12878R1x75dTh1014UMinusRawRep1V4.chrom
wgEncodeCaltechRnaSeqH1hescR1x75dTh1014UMinusRawRep1V4.chromStart wgEncodeCaltechRnaSeqH1hescR1x75dTh1014UMinusRawRep1V4.chromEnd wgEncodeCaltechRnaSeqH1hescR1x75dTh1014UMinusRawRep1V4.valueAverage
chr3 12328866 12475855 PPARG 441+ ENSG00000132170.19 protein_coding 54
111.1,100.8,2.817,4.412,11.615,4.447,15.31,1.268,1.835,0.962,2.774,2.425,1.834,2.175,1.344,1.058,1.546,0.712,1.473,0.827,71.26,0.31,20.425,14.11,8.176,5.916,21.575,3.985,5.033,3.615,7.896,5.217,4.716,7.852,8.148,3.55,19.455,4.966,2.097,13.79,17.47,0.893,0.456,5.365,4.717,6.024,9.239,11.03,10.46,
7.023,18.51,10.755,9.384,1.777, chr3 12393000 12475855 0.044918 chr3 12428944 12429019 -0.031200
```

- バイオインフォマティクス解析用データ連結プログラム
- UCSC GBの中にあるトラックであれば次々に追加できる
- 検索は、ゲノム座標であるところがユニーク

UShER

SARS-CoV-2系統樹に、入力した配列のデータを配置する高速プログラム

UShER: Ultrafast Sample placement on Existing tRee

Place your sequences in a global phylogenetic tree

Choose your pathogen: **SARS-CoV-2**

Select your FASTA, VCF or list of sequence names/IDs: **ファイルを選択** 選択されていません

or paste in sequence names/IDs:

Phylogenetic tree version:

6,653,976 genomes from GenBank, COG-UK and CNCB (2023-01-17); sarscov2phylo 13-11-20 tree with newer sequences added by UShER

Number of samples per subtree showing sample placement: 50

Upload **Upload Example File** **More example files**

More information

Upload your SARS-CoV-2 sequence (FASTA or VCF file) to find the most similar complete, high-coverage samples from **GISAID** or from public sequence databases (INSDC: GenBank/ENA/DDBJ accessed using **NCBI Virus**, **COG-UK** and the **China National Center for Bioinformation**), and your sequence placement in the phylogenetic tree generated by the **sarscov2phylo** pipeline. Placement is performed by **Ultrafast Sample placement on Existing tRee (UShER)** (Turakhia et al.). UShER also generates local subtrees to show samples in the context of the most closely related sequences. The subtrees can be visualized as Genome Browser custom tracks and/or using **Nextstrain's** interactive display which supports **drag-and-drop** of local metadata that remains on your computer.

GISAID data displayed in the Genome Browser are subject to GISAID's **Terms and Conditions**. SARS-CoV-2 genome sequences and metadata are available for download from **GISAID EpiCoV™**.

COVID-19 Pandemic Resources at UCSC

UShER: Ultrafast Sample placement on Existing tRee

[View in Genome Browser](#) | [View downloaded global tree in Nextstrain](#) | [View subtree 1 in Nextstrain](#) | [View subtree 2 in Nextstrain](#)

If you have metadata you wish to display, click a View subtree in Nextstrain button, and then you can drag on a CDS file to add it to the tree view.

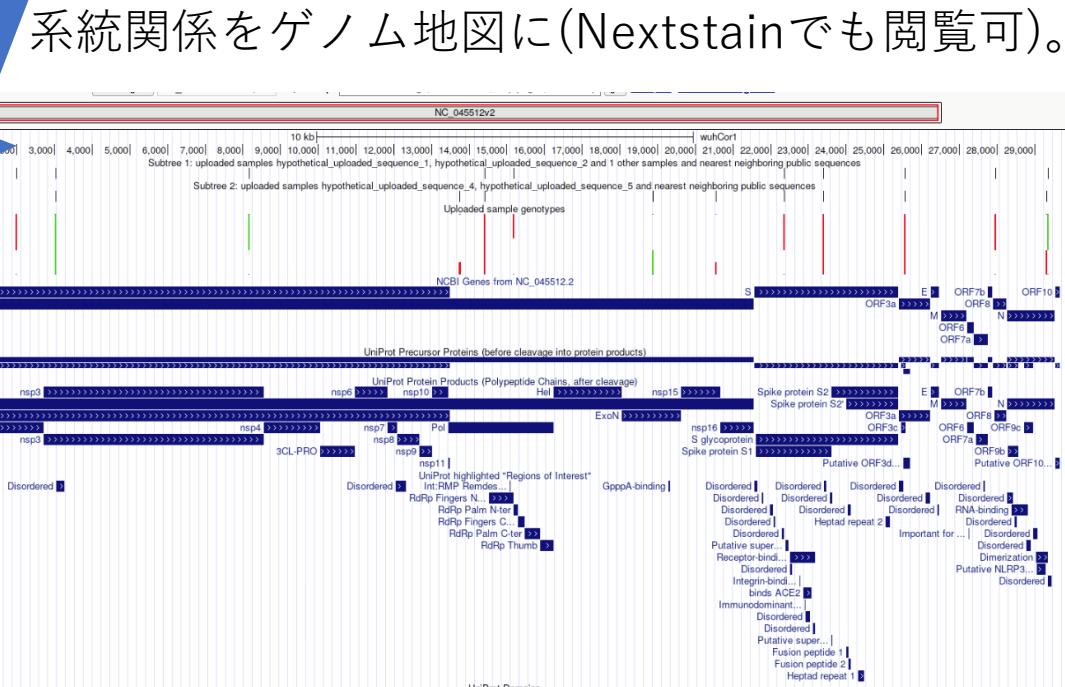
Note: The Nextstrain subtree views, and Download files below, are temporary files and will expire within two days. Please download the Nextstrain subtree JSON files if you will want to view them again in the future. The JSON files can be drag-dropped onto <https://usepileup.us/>.

Downloads: [Global phylogenetic tree with your sequences](#) | [TSV summary of sequences and placements](#) | [TSV summary of Spike mutations](#) | [ZIP of subtree JSON and Newick files](#)

Fasta Sequence	Size (kb)	Nts (n)	#Mined (t)	Bases aligned (t)	Inserted bases (t)	Deleted bases (t)	#BMs used for placement (t)	#Masked SNVs (t)	Nearest clade (t)	Pango lineage (t)	Neighboring sample in tree (t)	Lineage of neighbor (t)	#Imputed values for mixed bases (t)	#Maximally parsimonious placements (t)	Parsimony score (%)	Subtree number (t)
hypothetical_uploaded_sequence_1	29803	0	0	29803 (t)	0	0	11 (t)	0	20C	B.1.424	USA/MN-GDX-1722/2020 USA/MN-90506.1/2020-03-16	B.1.424	0	2	3	1 (view in Nextstrain)
hypothetical_uploaded_sequence_2	29803	0	0	29803 (t)	0	0	12 (t)	0	20C	B.1.424	USA/MN-GDX-1722/2020 USA/MN-90506.1/2020-03-16	B.1.424	0	1	1	1 (view in Nextstrain)
hypothetical_uploaded_sequence_3	29803	0	0	29803 (t)	0	0	12 (t)	0	20C	B.1.424	USA/MN-GDX-1722/2020 USA/MN-90506.1/2020-03-16	B.1.424	0	1	0	1 (view in Nextstrain)
hypothetical_uploaded_sequence_4	29803	0	0	29803 (t)	0	0	7(t)	0	20A	B.1	USA/MM-024/2020/2020 MW/90506.1/2020-12-22	B.1	0	5	1	2 (view in Nextstrain)
hypothetical_uploaded_sequence_5	29803	0	0	29803 (t)	0	0	9(t)	0	20A	B.1	USA/MM-024/2020/2020 MW/90506.1/2020-12-22	B.1	0	1	2	2 (view in Nextstrain)

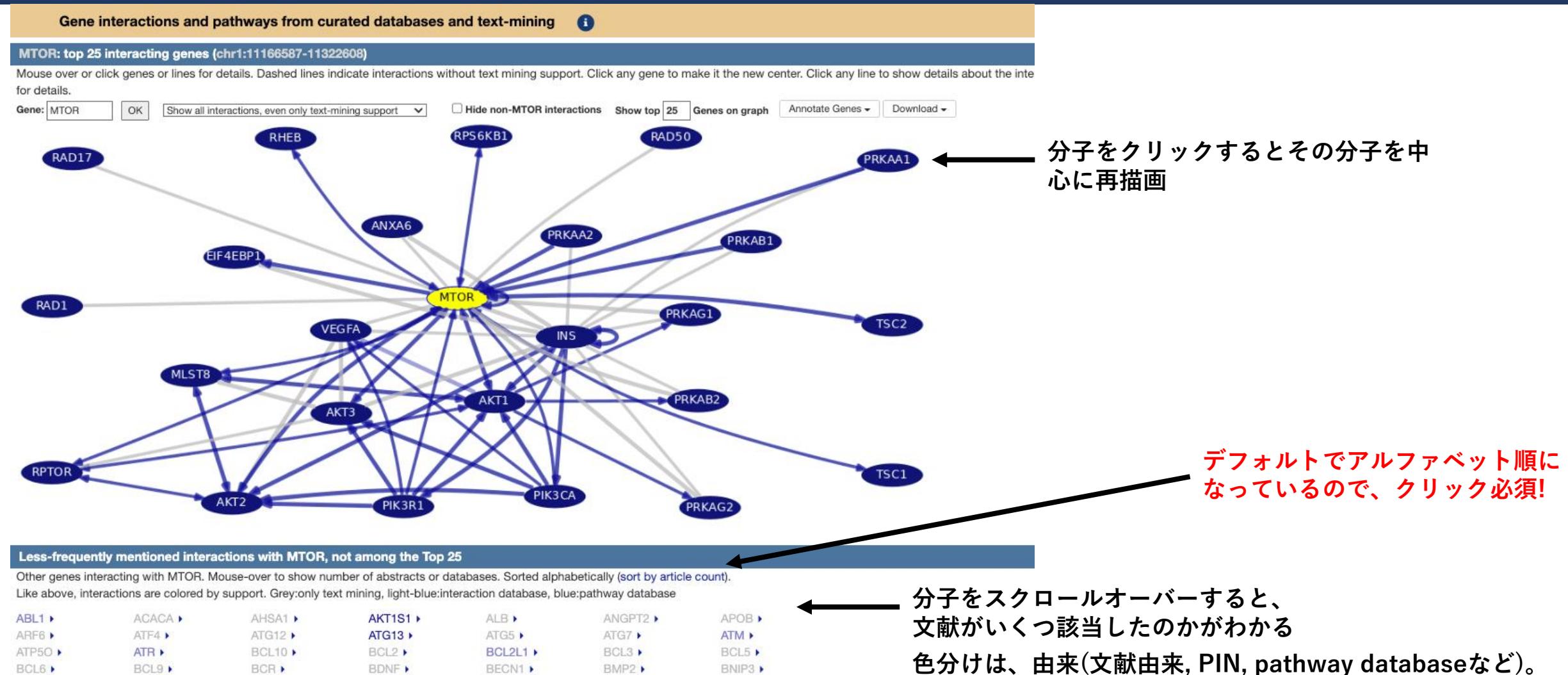
Subtree 1: 3 related samples
[View subtree 1 in Nextstrain](#)
 ↓ hypothetical_uploaded_sequence_2 USA/MN-GDX-1722/2020/MW/90506.1/2020-03-16 B.1.424
 ↓ hypothetical_uploaded_sequence_3 USA/MN-GDX-1722/2020/MW/90506.1/2020-03-16 B.1.424
 ↓ hypothetical_uploaded_sequence_1 USA/MN-GDX-1722/2020/MW/90506.1/2020-03-16 B.1.424
 Differences from the reference genome (NC_045512.2): C241T, C109F, G109A, C303T, T8167A, C1449T, Q2254A, A2340S, G2565T, C2796T
 Mutations along the path from the root of the phylogenetic tree to hypothetical_uploaded_sequence_1:
 C1449T > C241T > C303T > A2340S > G2565T > C2796T > G109A > T8167A, G2234A, C2993T
 Nearest neighboring public sequence already in phylogenetic tree: USA/MN-GDX-1722/2020/MW/90506.1/2020-03-16, lineage B.1.424
 2 nearest samples:
 • hypothetical_uploaded_sequence_2
 • hypothetical_uploaded_sequence_3
 Differences from the reference genome (NC_045512.2): C241T, C109F, G109A, C303T, T8167A, C1449T, Q2254A, A2340S, G2565T, C2796T, C2993T

結果の画面



Gene interactions

文献ベース（アブストラクト由来）を中心とする
分子共起ネットワーク



VisiGene Image Browser

マウス胎児のIn Situ Hybridization Image

VisiGene Image Browser

VisiGene is a virtual microscope for viewing *in situ* images. These images show where a gene is used in an organism, sometimes down to cellular resolution. With VisiGene users can retrieve images that meet specific search crit

ATF3

Good search terms include gene symbols, authors, years, body parts, organisms, GenBank and UniProt accessions, Known Gene descriptive terms, [Theiler stages for mice](#), and [Nieuwkoop/Faber stages for frogs](#). The wildcard character (*) must match.

Sample queries

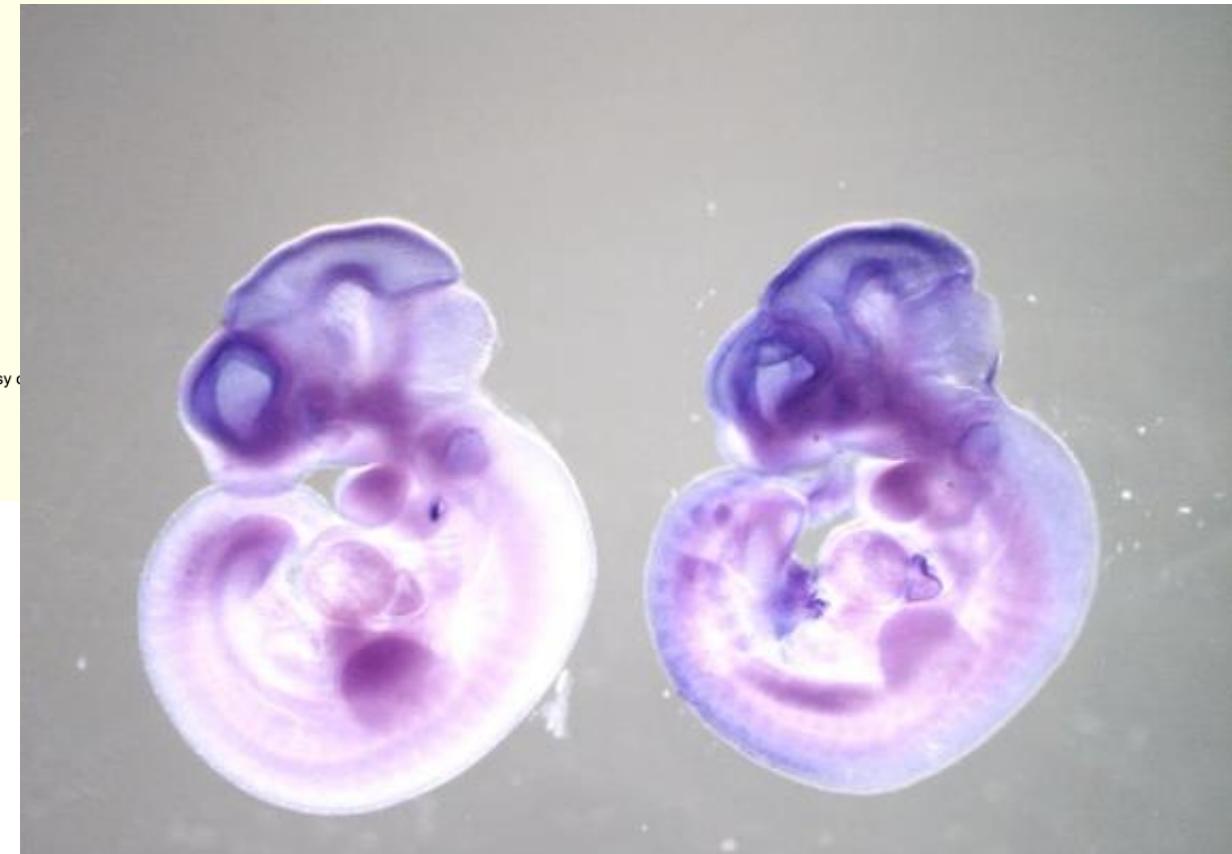
Request:	VisiGene Response:
nkx2-2	Displays images associated with the gene nkx2-2
hoxa*	Displays images of all genes in the Hox-A cluster (Note: * works only at the end of the word)
NM_007492	Displays images associated with accession NM_007492
theiler 22	Displays all images that show Theiler stage 22
vgPrb_16	Displays images associated with VisiGene probe ID 16
allen institute	Displays all images from the Allen Brain Atlas
mouse	Displays all mouse images
xenopus	Displays all images associated with frogs of genus Xenopus
mouse midbrain	Displays mouse images that show expression in the midbrain
smith jc 1994	Displays images contributed by scientist J.C. Smith in 1994

Images Available

The following image collections are currently available for browsing:

- High-quality high-resolution images of eight-week-old male mouse sagittal brain slices with reverse-complemented mRNA hybridization probes from the [Allen Brain Atlas](#), courtesy of the Allen Institute for Brain Science
- Mouse *in situ* images from the [Jackson Lab Gene Expression Database](#) (GXD) at MGI
- Transcription factors in mouse embryos from the Mahoney Center for Neuro-Oncology
- Mouse head and brain *in situ* images from the [Gene Expression Nervous System Atlas](#) (GENSAT) database
- Xenopus laevis *in situ* images from the [National Institute for Basic Biology](#) (NIBB) XDB project

- 転写因子で検索
- ワイルドカードで、ファミリーごと検索できる
- 結果に元となっているメタデータがしっかり記載
- 画像もハイクオリティでダウンロードが可能



source: Mahoney Lab source: MGI Reference: Mouse Brain Organization Revealed Through Direct Genome-Scale TF Expression Analysis.
Year: 2004 Contributors: Gray P.A.,Fu H.,Luo P.,Zhao Q.,Yu J.,Ferrari A.,Tenzen T.,Yuk D.I.,Tsung E.F.,Cai Z.,Alberta J.A.,Cheng L.P.,Liu Y.,Stenman J.M.,Valerius M.T.,Billings N.,Kim H.A.,Greenberg M.E.,McMahon A.P.,Rowitch D.H.,Stiles C.D.,Ma Q.
Gene: [Atf3](#) Probe: [RNA from primers](#) GenBank: [NM_007492](#)
Organism: Mus musculus Sex: n/a Strain: C57BL Genotype: wild type
Stage: 10.5 day old embryo (Theiler 17) Body Part: whole
Expression: central nervous system(0.17) Section Type: whole mount
Acknowledgements: Thanks to Paul Gray for transferring the images.
Acknowledgements: Thanks to the Gene Expression Database group at Mouse Genome Informatics (MGI) for collecting, annotating and sharing this image. The MGI images were last updated in VisiGene on March 28, 2006.
Additional and more up to date annotations and images may be available directly at [MGI](#).

Variant Annotation Integrator

バリアントにアノテーションを加える

The screenshot shows the Variant Annotation Integrator interface. At the top, there's a navigation bar with links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Projects, Help, and About Us. Below the navigation bar, the main content area is titled "Variant Annotation Integrator".

Select Genome Assembly and Region

clade: Mammal; genome: Human; assembly: Dec. 2013 (GRCh38/hg38)

region to annotate: position or search term: chr2:25,160,915-25,168,903

Select Variants

Your session doesn't have any custom tracks or hub tracks in pgSnp or VCF format. [add pgSnp or VCF custom track](#) [add track hub](#)

variants: Artificial Example Variants ▾
maximum number of variants to be processed: 10,000 ▾

[manage custom tracks](#) [track hubs](#) To reset all user cart settings (including custom tracks), [click here](#).

Select Genes

The gene predictions selected here will be used to determine the effect of each variant on genes, for example intronic, missense, splice site, intergenic etc.
NCBI RefSeq HGMD subset: transcripts with clinical variants in HGMD ▾

Select Regulatory Annotations

The annotations in this section provide predicted regulatory regions based on various experimental data. When a variant overlaps an annotation selected here, the consequence term [regulatory_region_variant](#) will be assigned. Follow the links to description pages that explain how each dataset was constructed. Some datasets cover a significant portion of the genome and it may be desirable to filter these annotations by cell type and/or score in order to avoid an overabundance of hits.

DNase I Hypersensitivity Peak Clusters from ENCODE (95 cell types) filter items

Select More Annotations (optional)

Database of Non-synonymous Functional Predictions (dbNSFP)

dbNSFP ([Liu et al. 2015](#)) release 3.1a provides pre-computed scores and predictions of functional significance from a variety of tools. Every possible coding change to transcripts in Gencode release 22 (Ensembl 79, Mar. 2015) gene predictions has been evaluated. Note: This may not encompass all transcripts in your selected gene set.

Variant Effect Scoring Tool (VEST) (scores [0-1] predict confidence that a change is deleterious)
 SIFT (D = damaging, T = tolerated)
 PolyPhen-2 with HumDiv training set (D = probably damaging, P = possibly damaging, B = benign)
 PolyPhen-2 with HumVar training set (D = probably damaging, P = possibly damaging, B = benign)
 MutationTaster (A = disease causing automatic, D = disease causing, N = polymorphism, P = polymorphism automatic)
 MutationAssessor (high or medium: predicted functional; low or neutral: predicted non-functional)
 Likelihood ratio test (LRT) (D = deleterious, N = Neutral, U = unknown)
 InterPro protein domains
 GERP++ Rejected Substitutions (RS)
 GERP++ Neutral Rate (NR)

VCFファイルを入力とする

dbSNPから情報を検索

Variant登録にはほぼ必須となっている機能予測プログラム群の結果をバリアントの情報として付加できる。

DNA Duster and Protein Duster

配列のコピペをしたときに混じっているゴミをとってくれる

DNA Duster

by Jim Kent

uff from a sequence. It outputs in a variety of formats including as translated protein.

```
GGGTG CCCGTC GAGNTC CCCCC
```

Submit

Output format:

spaces line breaks: show numbers case

strand

translate start at AUG lower case is intron

DNA Dusterの結果

```
GGGTGCCCGTCGAGNTCCCCCC
```

- よくある5塩基フォーマットの空白を除いてくれる
- Nは除かなかった

Executable and Source Code Downloads

Jim KentのUtility tools

Name	Last modified	Size	Description
Parent Directory	-		
external.i386/	2017-05-25 15:04	-	
external.x86_64/	2021-11-01 14:04	-	
linux.x86_64.v369/	2019-02-01 18:05	-	
linux.x86_64.v385/	2021-08-04 10:50	-	
linux.x86_64/	2023-01-17 16:04	-	
macOSX.arm64/	2022-12-21 13:12	-	
macOSX.x86_64/	2023-01-17 21:59	-	
testFiles/	2022-11-14 16:52	-	
userApps.archive/	2023-01-17 16:41	-	
userApps.src.tgz	2023-01-17 16:40	23M	

#####
This file is from:

http://hgdownload.soe.ucsc.edu/admin/exe/macOSX.x86_64/README.txt

This directory contains applications for stand-alone use,
built on a Mac OSX 10.15.7 intel machine. (Catalina)
Darwin Kernel Version 19.6.0, gcc version:
Apple clang version 12.0.0 (clang-1200.0.32.29)
Target: x86_64-apple-darwin19.6.0
Thread model: posix

kent source tree v442 January 2023.

For help on the bigBed and bigWig applications see:
<http://genome.ucsc.edu/goldenPath/help/bigBed.html>
<http://genome.ucsc.edu/goldenPath/help/bigWig.html>

View the file 'FOOTER.txt' to see the usage statement for
each of the applications.

The shared libraries used by these binaries are: (from: otool -L <binary>)

/usr/lib/libc++.1.dylib (compatibility version 1.0.0, current version 902.1.0)
/usr/lib/libSystem.B.dylib (compatibility version 1.0.0, current version 1281.100.1)
/usr/lib/libz.1.dylib (compatibility version 1.0.0, current version 1.2.11)

神

- UCSC GBを作成・利用する上で、必要なスクリプトが維持されている
- 各種バイナリファイルが用意されていて、コンパイルの必要がない
- 高速で、バグが少ない

参考となるウェブサイト

本家のYouTubeチャンネル

The screenshot shows the UCSC Genome Browser YouTube channel page. At the top, there's a video player showing a genomic track for H3K27Ac marks. Below the video, there are several video thumbnails. One thumbnail is titled "Introduction to the UCSC SARS-CoV-2 Genome Browser" and another shows a DNA sequence with protein substitutions. The channel stats at the bottom show 2830 subscribers and 431 video views. The channel name is "UCSC Genome Browser" and the handle is "@ucscgenomebrowser". A Twitter link (@GenomeBrowser) and a bio describing the browser as a public, open-source web-based graphical viewer for genome sequences and annotations are also present.

TogoTV (<https://togotv.dbcls.jp/>)

The screenshot shows the search results for "UCSC" on TogoTV. The results are organized into three categories: 動画マニュアル (37), 講演 (7), and ハンズオン講習 (8). The first result is a video titled "FANTOM5における UCSC Genome Browser Track Hubの利用とその展開" by 川路英哉, posted on 2022-09-03. The second result is "UCSC Genome Browserを使って ClinVarに登録されている SNVとCNVを検索する" posted on 2022-03-23. Other results include videos on gnomAD SVs v2.1, UCSC Table Browser, Data Integrator, and UCSC Xena.