2021年12月16日（木）
AJACSオンライン９ 「疾患に関する多型データを解析する」

# バリアントの機能を推定する

浜松医科大学医化学講座
才津　浩智

# 自己紹介

1998年　　　　　九州大学　産婦人科入局

2001-2006年　京都大学　医学研究科（塩田浩平教授）-- 発生生物学

2006-2015年　横浜市立大学 遺伝学教室（松本直通教授）

2016年-現在　　浜松医科大学 医化学講座

　日本遺伝子診療学会　ジェネティックエキスパート認定制度委員会委員長

**浜松医科大学**



### 第１７回臨床遺伝情報検索講習会（報告）

日　時：　2020年9月16日（水）　10:00 ～ 2020年10月15日（木）23:59
開催形式：　e-ラーニングシステムを利用したウェブ配信

プログラム：

- 次世代シークエンスの基礎知識とメンデル遺伝病におけるバリアントの解釈について
  才津浩智（浜松医科大学医化学講座）

- 臨床遺伝情報検索のweb サイト
  中山智祥（日本大学医学部病態病理学系臨床検査医学分野）

- 単一遺伝子疾患の情報検索（基礎）
  足立香織（鳥取大学研究推進機構 研究基盤センター）

### 第１８回臨床遺伝情報検索講習会（報告）

日　時：　2021年3月8日（月）　0:00 ～ 2021年4月11日（日）23:59
開催形式：　e-ラーニングシステムを利用したウェブ配信

プログラム：

- UCSC Genome Browser を用いたヒトゲノム構成の閲覧
  佐藤謙一（国際医療福祉大学　福岡保健医療学部）

- がんゲノム医療における基礎知識
  雨宮健司（山梨県立中央病院　ゲノム解析センター）

- がん遺伝子パネル検査に関する情報検索と検査結果解釈の実技演習
  柿島裕樹（国立がん研究センター中央病院　臨床検査科）

### 第１９回臨床遺伝情報検索講習会（報告）

日　時：　2021年7月12日（月）　0:00 ～ 2021年8月11日（水）23:59
開催形式：　e-ラーニングシステムを利用したウェブ配信

プログラム：

- 臨床遺伝情報検索のための統合TVの活用法
  小野浩雅（ライフサイエンス総合データベースセンター）

- ACMG-AMPガイドラインによるバリアントの評価
  才津浩智（浜松医科大学　医化学講座）

# Contents

## 1. バリアントの種類

（シノニマス、ミスセンス、ナンセンス、インフレーム/フレームシフト、スプライスサイト、イントロンバリアント）

## 2. バリアントの機能推定（*in silico*解析）の役割

## 3*. in silico*解析プログラムの紹介

SIFT(ミスセンスバリアント)

PolyPhen-2 (ミスセンスバリアント)

M-CAP (ミスセンスバリアント)

PROVEAN (ミスセンスバリアント/インフレーム)

MutationTaster2021 (ミスセンスバリアント/インフレーム/フレームシフト)

CADD (ミスセンスバリアント/インフレーム/フレームシフト/イントロン)

SpliceAI（スプライシング予測）

ESEfinder（スプライシング予測）

# バリアントの大まかな種類

## 一塩基置換(SNV) / 挿入欠失

| 1 bp | 1〜1000 bp |
|---|---|
| **一塩基置換(SNV)** | **挿入欠失** |

一塩基置換(SNV)
```
5'-CTTTGTGCC-3'
3'-GAAACACGG-5'
```
↓ G>A
```
5'-CTTTATGCC-3'
3'-GAAATACGG-5'
```
↓
アミノ酸の置換
早期終止コドンの出現
スプライシング異常

挿入欠失
```
5'-CTTTGTGCC-3'
3'-GAAACACGG-5'
```
↓ Gdel
```
5'-CTTTTGCC-3'
3'-GAAAACGG-5'
```
↓
フレームシフトによる早期終止コドンの出現

## ～数100 Mb

**染色体異数性**　**染色体構造異常**

**21**
21トリソミー

Robertson（型）転座

トンプソン&トンプソン遺伝医学　2009

## 200 bp ～ 30 kb

**反復配列の伸長**

正常　$(CAG)^{12-35}$
異常　$(CAG)^{36-121}$

リピート配列の伸長

## 1 kb～ 数10 Mb

**コピー数変化**
**(CNV, copy number variant）**

遺伝子A　遺伝子B

遺伝子Aの欠失

遺伝子B

↓
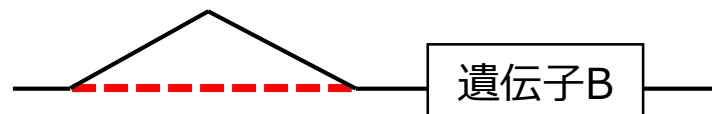
通常2コピーある（常染色体は2本あります）遺伝子が1コピーになって発現量が変わる

**シークエンスバリアント**
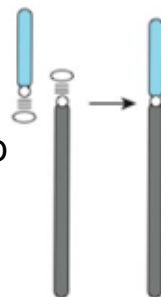
# シークエンスバリアントの分類

## シノニマス(synonymous)バリアント

```
              Gly   Gly   Thr
参照配列  5'-GGT  GGA  ACT -3'
バリアント  5'-GTT  GGG  CTA -3'
                    Gly
```

## ミスセンス(nonsynonymous)バリアント

```
              Gly   Gly   Thr
参照配列  5'-GGT  GGA  ACT -3'
バリアント  5'-GTT  AGA  CTA -3'
                   Arg
```

## ナンセンスバリアント

```
              Gly   Gly   Thr
参照配列  5'-GGT  GGA  ACT -3'
バリアント  5'-GTT  TGA  CTA -3'
                   Stop
```

## フレームシフト/インフレームバリアント(欠失・挿入)

```
              Thr   Ile   Lys   Ile   Gly   Leu   Thr
参照配列  5'-ACC  ATA  AAA  ATT  GGT  TTG  ACT -3'
バリアント1  5'-ACC  ATA  AAA  TTG  GTT  TGA  CT -3'
                               Leu   Val   Stop
```

Leuを含めて3つ目のコドンでストップ

```
              Thr   Ile   Lys   Ile   Gly   Leu   Thr
参照配列  5'-ACC  ATA  AAA  ATT  GGT  TTG  ACT -3'
バリアント2  5'-ACC  ATA  ATT  GGT  TTG  ACT   -3'
              Thr   Ile   Ile   Gly   Leu   Thr
```

1アミノ酸（Lys）の欠失

```
非翻訳領域      1                77   +2              -8           78
（UTR）

----ATG------ACGgt------cggccgcaagAAC---
    開始コドン                                          タンパク質
                         G              A              コーディング領域
                    c.77+2T>G      c.78-8G>A
                 スプライスサイトバリアント  イントロンバリアント
```
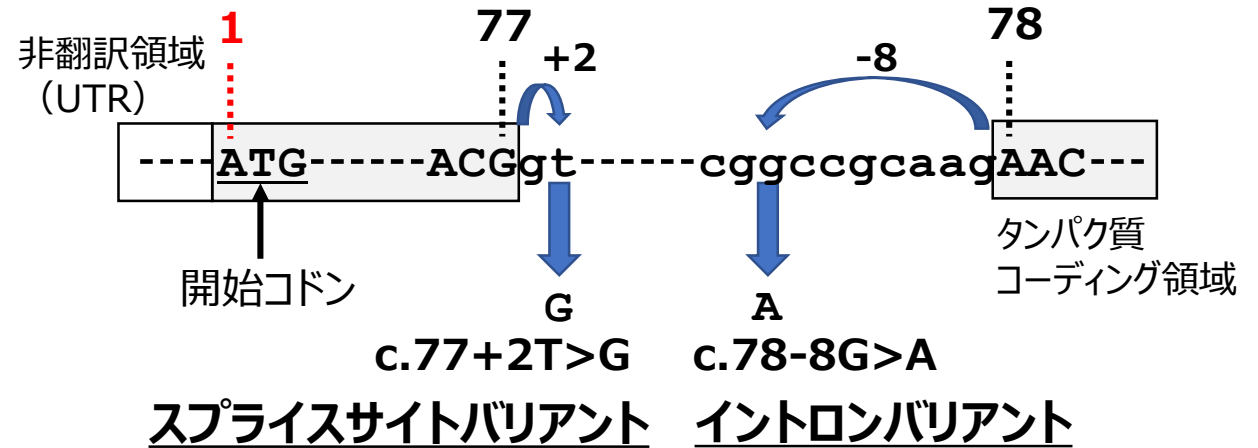
タンパク質の機能に与える影響

**ナンセンス、フレームシフト、スプライスサイト >> ミスセンス、インフレーム >> シノニマス、イントロン**

# バリアントの機能推定（*in silico*解析）の役割

**1．遺伝子診断で既知原因遺伝子に同定されたバリアントの病原性の評価**

**2．網羅的遺伝子解析のバリアントのフィルタリングの一つの指標として用いる**

＜メンデル遺伝病＞

米国臨床遺伝・ゲノム学会(ACMG)と分子病理学協会(AMP)が定めたシークエンス解析で同定された生殖細胞系列バリアント(遺伝する) に対する評価基準
（ACMGガイドライン 2015)

バリアントを5つに分類

(1) pathogenic

(2) likely pathogenic

(3) uncertain significance

(4) likely benign

(5) Benign

エビデンス

| | |
|---|---|
| **ヌルバリアント**<br>（ナンセンス、フレームシフト、スプライス部位） | Very strong |
| ***De novo*であること（要、親子鑑定）**<br>**病的意義が確立されたアミノ酸置換と同一**<br>**バリアントの機能解析で機能低下が証明** | Strong |
| **バリアントのアレル頻度（稀である）**<br>**同じアミノ酸の違うミスセンス変化** | Moderate |
| **表現型との分離**<br>***in silico*の病原性予測** | Supportive |

**病原性予測が高いから病的という使い方よりは、予測が低いので病的でない可能性が高いという使い方**

# バリアントの機能推定（*in silico*解析）の役割

**イントロンバリアントに関しては、スプライス異常を示唆する所見として、**
**<span style="color:red">RNA解析をするかの決定に重要</span>**

*GNAO1* c.724-8G>A

↓

SpliceAIでスプライス異常を予測

↓

末梢血単核球からmRNA抽出



400bp

300bp

MT
WT

Pt　control　control　NC(RT-)
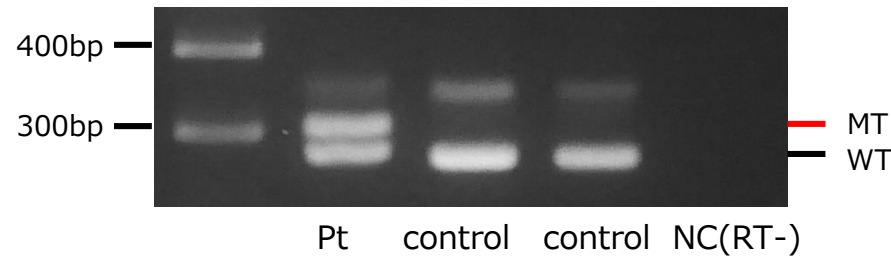
# *in silico*解析プログラムの紹介

**nonsynonymous SNV**
**スプライスサイト SNV**

**dbNSFP**
http://database.liulab.science/dbNSFP

SIFT(ミスセンスバリアント)

PolyPhen-2 (ミスセンスバリアント)

M-CAP (ミスセンスバリアント)

PROVEAN (ミスセンスバリアント/インフレーム)

MutationTaster (ミスセンスバリアント/インフレーム/フレームシフト)

CADD (ミスセンスバリアント/インフレーム/フレームシフト/**イントロン**)

SpliceAI Lookup（スプライシング予測）

ESEfinder（スプライシング予測)

*BRCA2* **NM_000059.3:c.53G>A, p.(Arg18His)**
**NC_000013.11:g.32316513G>A (GRCh38)**
**NC_000013.10:g.32890650G>A (GRCh37)**

# SIFT （https://sift.bii.a-star.edu.sg/）



## Sorting Intolerant From Tolerant

Code　Publications　History　Contact us

This website is maintained by Pauline Ng, creator of SIFT. Pauline developed and maintained the SIFT website at Fred Hutchinson Cancer Researcher Center from 2001-2008, then the J. Craig Venter Institute from 2008-2010, and Genome Institute of Singapore (2010-2017). Pauline is currently an indepndent consultant.

SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids. SIFT can be applied to naturally occurring **nonsynonymous polymorphisms and laboratory-induced missense mutations.**

進化的な配列保存性とアミノ酸の物理的特性

**UPDATE on 6 Dec 2019**

- Change SIFT4G Annotator to browse for genome locally instead of dynamically querying SIFT website due to change in A-STAR security web policy
- Github code for building SIFT 4G databases changed to allow more per missive gff files (start and stop codons no longer required)

### Genome Tools

**SNV / SNP prediction**

SIFT For Genomes Predictions for human build 37, 38, and ≥ 200 genomes

SIFT For Genomes (Online submission) (**Beta**) Predictions for some model organisms (e.g. human, mouse, worm, yeast).

**VCF**

SIFT nonsynonymous single nucleotide variants (genome-scale) (human build 37)

dbSNP rsIDs (SIFT4G predictions)

**INDEL Prediction**
1. Restrict indels to coding
2. Classify coding indels (Insertion/Deletions). Human build 37 and 38)

### Single Protein Tools

SIFT Sequence

SIFT Related Sequence

SIFT Aligned Sequences

If using SIFT 4G, please cite: SIFT missense predictions for genomes. *Nat Protocols* 2016; 11:1-9(Link)

Otherwise, please cite: SIFT web server: predicting effects of amino acid substitutions on proteins.*Nucleic Acids Research*, 2012 Jul; 40

Website is personally maintained by Pauline Ng. Server support by Bioinformatics Institute in Singapore.

# SIFT nonsynonymous single nucleotide variantsの使い方

*BRCA2* NM_000059.3:c.53G>A, p.(Arg18His)
**NC_000013.10:g.32890650G>A** (GRCh37)

## Score < 0.05でintorelant

SIFT nonsynonymous single nucleotide variants (genome-scale)

This page provides SIFT predictions for a list of **chromosome positions and alleles**. Click here to get more information

**User Input**

Homo sapiens (hg19) NCBI 37

Chromosome Coordinates
**Paste in comma separated list of chromosome coordinates, orientation (1,-1) and alleles** see [sample format]

13,32890650,1,G/A

13,32890650,1,G/A

-or-

Upload file containing chromosome coordinates and nucleotide substitutions (size limit: 1M)
ファイルを選択 選択されていません

**Email:**
Please enter your email address if you want your results via email.

**Output Options**

### SIFT version 5.1.1, SIFT database release February 27, 2012

| Coordinates | Codons | Ensembl Transcript ID | RefSeq Transcript ID | Known Transcript ID | CCDS Transcript ID | Ensembl Protein ID |
|---|---|---|---|---|---|---|
| 13,32890650,1,G/A | CGC-CaC | ENST00000380152,ENST00000544455 | NM_000059 | uc001uub.1 | CCDS9344.1 | ENSP00000369497,ENSP00000439902 |

| RefSeq Protein ID | Known Protein ID | Substitution | Region | dbSNP ID | SNP Type | Prediction | Score | Median Info | # Seqs at position |
|---|---|---|---|---|---|---|---|---|---|
| NP_000050 | NP_000050 | R18H | EXON CDS | rs80358762:A | Nonsynonymous | TOLERATED | 0.23 | 2.91 | 18 |

000 Genome Allele Frequencies
☐ All available populations
☐ European population
☐ East Asian population
☐ African population
☐ South Asian population
☐ American population

Select All   UnSelect All

送信   リセット

# SIFT sequenceの使い方

## SIFT Sequence

**SIFT Sequence provides SIFT predictions for a given protein FASTA sequence.** This will take 10-15 min because we must search your protein sequence against a database to pick the related sequences. You can also submit your protein sequence and related sequences or aligned sequences if you already have them.

*Results are deleted after 24 hours, so please save them!*
[Preventing connection failures]

### Protein sequence

Name of file containing protein query sequence (fasta format).
ファイルを選択 選択されていません

-or-

Paste in your protein query sequence (Upload example) (fasta format).

BRCA2のFASTA配列が必要

Enter the substitutions of interest [format]:

R18H

-or-
Upload a file containing substitutions of interest [format]:
ファイルを選択 選択されていません

**BRCA2** NM_000059.3:c.53G>A, p.(Arg18His)
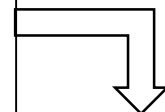
### Parameters

Select database to search UniProt-SwissProt 2010_09
Warnings for you predictions? Change this option

Median conservation of sequences 3.00
*(between 0 and 4.32, we recommend between 2.75 and 3.25)*

Remove sequences more than 90 percent identical to query
*This cleans out any polymorphic alleles (sequences in the database that are nearly identical to your protein but already containing the substitutions of interest). The contaminating sequences may cause the substitution to be predicted as tolerated.*

Enter your email address if you want the results through email : (Optional)

Submit リセット

# UniProt (https://www.uniprot.org/)

# UniProtKB - P51587 (BRCA2_HUMAN)

## Display

**Help video**

| BLAST | Align | Format | Add to basket | History |

Community curation (1)   Add a publication

- Entry
- Publications
- Feature viewer
- Feature table

**None**

- ✓ Function
- ✓ Names & Taxonomy
- ✓ Subcellular location
- ✓ Pathology & Biotech
- ✓ PTM / Processing
- ✓ Expression
- ✓ Interaction
- ✓ Structure
- ✓ Family & Domains
- ✓ Sequences (1+)
- ✓ Similar proteins
- ✓ Cross-references
- ✓ Entry information
- ✓ Miscellaneous

▲ Top

| Protein | **Breast cancer type 2 susceptibility protein** |
| Gene | **BRCA2** |
| Organism | *Homo sapiens* (Human) |
| Status | Reviewed - Annotation score: ●●●●● - Experimental evidence at protein level |

BLAST   Align   Format   Add to basket   History

**View format** ✖

- Text
- **FASTA (canonical)**
- XML
- RDF/XML
- GFF

| Protein | **Breast c...** ...ptibility protein |
| Gene | **BRCA2** |
| Organism | *Homo sap...* |
| Status | Review... ●●●●● - Experiment... |

## Function

Involved in double-strand break repair and/or homologous recombination. Binds RAD51 and potentiates recombinational DNA repair by promoting assembly of RAD51 onto single-stranded DNA (ssDNA). Acts by targeting RAD51 to ssDNA over double-stranded DNA, enabling RAD51 to displace replication protein-A (RPA) from ssDNA and stabilizing RAD51-ssDNA filaments by blocking ATP hydrolysis. Part of a PALB2-scaffolded HR complex containing RAD51C and which is thought to play a role in DNA repair by HR. May participate in S phase checkpoint activation. Binds selectively to ssDNA, and to ssDNA in tailed duplexes and replication fork structures. May play a role in the extension step after strand invasion at replication-dependent DNA double-strand breaks; together with PALB2 is involved in both POLH localization at collapsed replication forks and DNA polymerization activity. In concert with NPM1, regulates centrosome duplication. Interacts with the TREX-2 complex (transcription and export complex 2) subunits PCID2 and SEM1, and is required to prevent R-loop-associated DNA damage and thus transcription-associated genomic instability. Silencing of BRCA2 promotes R-loop accumulation at actively transcribed genes in replicating and non-replicating cells, suggesting that BRCA2 mediates the control of R-loop associated genomic instability, independently of its known role in homologous recombination (PubMed:24896180).

🏷 11 Publications ▾

### GO - Molecular function

- gamma-tubulin binding   🏷 Source: UniProtKB ▾
- H3 histone acetyltransferase activity   🏷 Source: UniProtKB ▾
- H4 histone acetyltransferase activity   🏷 Source: UniProtKB ▾
- identical protein binding   🏷 Source: IntAct ▾
- protease binding   🏷 Source: UniProtKB ▾
- protein C-terminus binding   🏷 Source: MGI ▾
- single-stranded DNA binding   🏷 Source: UniProtKB ▾

Complete GO annotation on QuickGO ...

### GO - Biological process

- brain development   🏷 Source: Ensembl
- cell aging   🏷 Source: Ensembl
- centrosome duplication   🏷 Source: UniProtKB
- DNA damage response, signal transduction by p53 class mediator resulting in transcription of p21 class mediator   🏷 Source: Ensembl
- double-strand break repair   🏷 Source: UniProtKB ▾
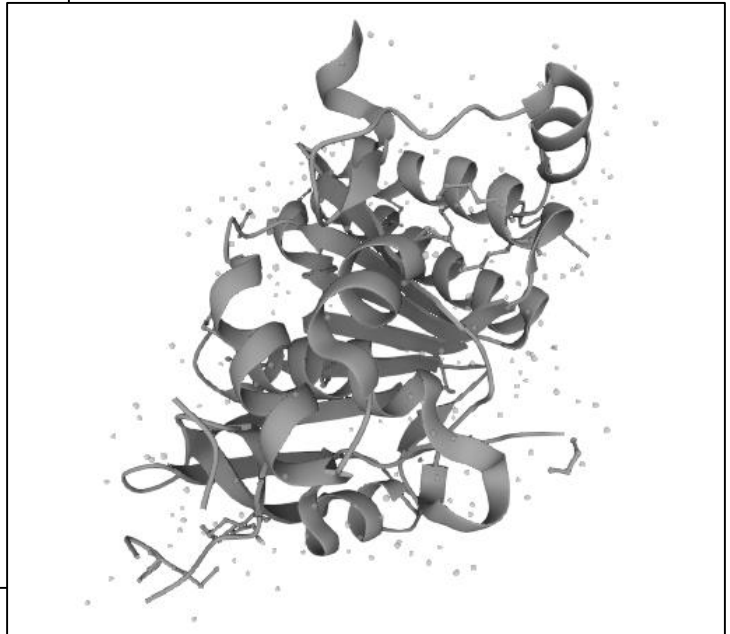
```
>sp|P51587|BRCA2_HUMAN Breast cancer type 2 susceptibility protein OS=Homo sapiens OX=9606 GN=BRCA2 PE=1 SV=3
MPIGSKERPTFFEIFKTRQNKADLGPISLNWFEELSSEAPPYNSEPAEESEHKNNNYEPN
LFKTPQRKPSYNQLASTPIIFKEQGLTLPLYQSPVKELDKFKLDLGRNVPNSRHKSLRTV
KTKMDQADDVSCPLLNSCLSESPVVLQCTHVTPQRDKSVVCGSLFHTPKFVKGRQTPKHI
SESLGAEVDPDMSWSSSLATPPTLSSTVLIVRNEEASETVFPHDTTANVKSYFSNHDESL
KKNDRFIASVTDSENTNQREAASHGFKGTSGNSFKVNSCKDHIGKSMPNVLEDEVYETVV
DTSEEDSFSLCFSKCRTKNLQKVRTSKTRKKIFHEANADECEKSKNQVKEKYSFVSEVEP
NDTDPLDSNVANQKPFESGSDKISKEVVPSLACEWSQLTLSGLNGAQMEKIPLLHISSCD
QNISEKDLLDTENKRKKDFLTSENSLPRISSLPKSEKPLNEETVVNKRDEEQHLESHTDC
ILAVKQAISGTSPVASSFQGIKKSIFRIRESPKETFNASFSGHMTDPNFKKETEASESGL
EIHTVCSQKEDSLCPNLIDNGSWPATTTQNSVALKNAGLISTLKKKTNKFIYAIHDETSY
KGKKIPKDQKSELINCSAQFEANAFEAPLTFANADSGLLHSSVKRSCSQNDSEEPTLSLT
SSFGTILRKCSRNETCSNNTVISQDLDYKEAKCNKEKLQLFITPEADSLSCLQEGQCEND
PKSKKVSDIKEEVLAAACHPVQHSKVEYSDTDFQSQKSLLYDHENASTLITPTSKDVLS
NLVMISRGKESYKMSDKLKGNNYESDVELTKNIPMEKNQDVCALNENYKNVELLPPEKYM
RVASPSRKVQFNQNTNLRVIQKNQEETTSISKITVNPDSEELFSDNENNFVFQVANERNN
LALGNTKELHETDLTCVNEPIFKNSTMVLYGDTGDKQATQVSIKKDLVYFQVLAEENKNSVK
QHIKMTLGQDLKSDISLNIDKIPEKNNDYMNKWAGLLGPISNHSFGGSFRTASNKEIKLS
EHNIKKSKMFFKDIEEQYPTSLACVEIVNTLALDNQKKLSKPQSINTVSAHLQSSVVVSD
CKNSHITPQMLFSKQDFNSNHNLTPSQKAEITELSTILEESGSQFEFTQFRKPSYILQKS
TFEVPENQMTILKTTSEECRDADLHVIMNAPSIGQVDSSKQFEGTVEIKRKFAGLLKNDC
NKSASGYLTDENEVGFRGFYSAHGTKLNVSTEALQKAVKLFSDIENISEETSAEVHPISL
SSSKCHDSVVSMFKIENHNDKTVSEKNNKCQLILQNNIEMTTGTFVEEITENYKRNTENE
DNKYTAASRNSHNLEFDGSDSSKNDTVCIHKDETDLLFTDQHNICLKLSGQFMKEGNTQI
KEDLSDLTFLEVAKAQEACHGNTSNKEQLTATKTEQNIKDFETSDTFFQTASGKNISVAK
ESFNKIVNFFDQKPEELHNFSLNSELHSDIRKNKMDILSYEETDIVKHKILKESVPVGTG
NQLVTFQGQPERDEKIKEPTLLGFHTASGKKVKIAKESLDKVKNLFDEKEQGTSEITSFS
HQWAKTLKYREACKDLELACETIEITAAPKCKEMQNSLNNDKNLVSIETVVYPPKLLSDNL
CRQTENLKTSKSIFLKVKVHENVEKETAKSPATCYTNQSPYSVIENSALAFYTSCSRKTS
VSQTSLLEAKKWLREGIFDGQPERINTADYVGNYLYENNSNSTIAENDKNHLSEKQDTYL
SNSSMSNSYSYHSDEVYNDSGYLSKNKLDSGIEPVLKNVEDQKNTSFSKVISNVKDANAY
PQTVNEDICVEELVTSSSPCKNKNAAIKLSISNSNNFEVGPPAFRIASGKIVCVSHETIK
KVVKDIFTDSFSKVIKENNENKSKICQTKIMAGCYEALDDSEDILHNSLDNDECSTHSHKV
FADIQSEEILQHNQNMSGLEKVSKISPCDVSLETSDICKCSIGKLHKSVSSANTCGIFST
ASGKSVQVSDASLQNARQVFSEIEDSTKQVFSKVLFKSNEHSDQLTREENTAIRTPEHLI
SQKGFSYNVVNSSAFSGFSTASGKQVSILESSLHKVKGVLEEFDLIRTEHSLHYSPTSRQ
NVSKILPRVDKRNPEHCVNSEMEKTCSKEFKLSNNLNVEGGSSENNHSIKVSPYLSQFQQ
DKQQLVLGTKVSLVENIHVLGKEQASPKNVKMEIGKTETFSDVPVKTNIEVCSTYSKDSE
NYFETEAVEIAKAFMEDDELTDSKLPSHATHSLFTCPENEEMVLSNSRIGKRRGEPLILV
GEPSIKRNLLNEFDRIIENQEKSLKASKSTPDGTIKDRRLFMHHVSLEPITCVPFRTTKE
RQEIQNPNFTAPGQEFLSKSHLYEHLTLEKSSSNLAVSGHPFYQVSATRNEKMRHLITTG
RPTKVFVPPFKTKSHFHRVEQCVRNINLEENRQKQNIDGHGSDDSKNKINDNEIHQFKN
NSNQAAAVTFTKCEEEPLDLITSLQNARDIQDMRIKKKQRQRVFPQPGSLYLAKTSTLPR
ISLKAAVGGQVPSACSHKQLYTYGVSKHCIKINSKNAESFQFHTEDYFGKESLWTGKGIQ
LADGGWLIPSNDGKAGKEEFYRALCDTPGVDPKLISRIWVYNHYRWIIWKLAAMECAFPK
EFANRCLSPERVLLQLKYRYDTEIDRSRRSAIKKIMERDDTAAKTLVLCVSDIISLSANI
SETSSNKTSSADTQKVAIIELTDGWYAVKAQLDPPLLAVLKNGRLTVGQKIILHGAELVG
SPDACTPLEAPESLMLKISANSTRPARWYTKLGFFPDPRPFPLPLSSLFSDGGNVGCVDV
IIQRAYPIQWMEKTSSGLYIFRNEREEEKEAAKYVEAQQKRLEALFTKIQEEFEEHEENT
TKPYLPSRALTRQQVRALQDGAELYEAVKNAADPAYLEGYFSEEQLRALNNHRQMLNDKK
QAQIQLEIRKAMESAEQKEQGLSRDVTTVWKLRIVSYSKKEKDSVILSIWRPSSDLYSLL
TEGKRYRIYHLATSKSKSKSERANIQLAATKKTQYVQQLPVSDEILFQIYQPREPLHFSKF
LDPDFQPSCSEVDLIGFVVSVVKKTGLAPFVYLSDECYNLLAIKFWIDLNEDIIKPHMLI
AASNLQWRPESKSGLLTLFAGDFSVFSASPKEGHFQETFNKMKNTVENIDILCNEAENKL
MHIHANDPKWSTPTKDCTSGPYTAQIIPGTGNKLMSSPNCEIYYQSPLSLCMAKRKSV
STPVSAQMTKSKSCKGEKEIDDQKNCKKRRALDFLSRLPLPPPVSPICTFVSPAAQKAFQP
PRSCGTKYETPIKKKELNSPQMTPFKKFNEISLLESNSIADEEALINTQALLSGSTGEK
QFISVSESTRTAPTSSEDYLRLKRRCTTSLIKEQESSQASTEECEKNKQDTITTKKYI
```

# SIFT Results (Protein Sequences)

Your job id is abab675a9d and is currently running.

If your browser times out before results are shown, please go to https://sift.bii.a-star.edu.sg//sift-bin/format.pl?abab675a9d_sequences in **20 minutes**. Opening it up before 20 minutes will throw an error (in which case, just refresh).

Problems? Contact us with your job id.
Your results will be available here for the next 24 hours.
**4** sequences were selected to be closely related to your query sequence.
PSIBLAST alignment of submitted sequences
Alignment in FASTA format (for modification)
The alignment taken from PSIBLAST is returned in msf format.
**Note: X**es are placeholders at the beginning and end of sequences. While **-** means a gap in the alignment an X means a lack of information such as a partial alignment or incomplete sequence and do not contribute to the prediction.
*Please check the sequences that have been chosen. If the sequences are too diverged from your query or the alignment is questionable, we suggest you modify the fasta-formatted file above and* _resubmit_.

SIFT amino acid predictions for:
Positions 1 to 100
Positions 101 to 200
Positions 201 to 300
Positions 301 to 400
Positions 401 to 500
Positions 501 to 600
Positions 601 to 700
Positions 701 to 800
Positions 801 to 900
Positions 901 to 1000
Positions 1001 to 1100
Positions 1101 to 1200
Positions 1201 to 1300
Positions 1301 to 1400
Positions 1401 to 1500
Positions 1501 to 1600
Positions 1601 to 1700
Positions 1701 to 1800
Positions 1801 to 1900
Positions 1901 to 2000
Positions 2001 to 2100
Positions 2101 to 2200
Positions 2201 to 2300
Positions 2301 to 2400
Positions 2401 to 2500
Positions 2501 to 2600
Positions 2601 to 2700
Positions 2701 to 2800
Positions 2801 to 2900
Positions 2901 to 3000
Positions 3001 to 3100
Positions 3101 to 3200
Positions 3201 to 3300
Positions 3301 to 3400
Positions 3401 to 3500

Scaled Probabilities for Entire Protein
***May take some time to load!!*** *Please be patient if you do not see the table immediately.*
Amino acids with probabilities < .05 are predicted to be deleterious.

Predictions of substitutions entered

*** The following sequences have been removed because they were found to be over 90% identical with your protein query: P51587,.

Substitution at pos 18 from R to H is predicted to AFFECT PROTEIN FUNCTION with a score of 0.00.
    Median sequence conservation: 3.63
    Sequences represented at this position:4
WARNING!!   This substitution may have been predicted to affect function just because the sequences used were not diverse enough.  **There is LOW CONFIDENCE in this prediction.**

ゲノムでやった結果とアミノ酸配列からやったデータで矛盾がある

15

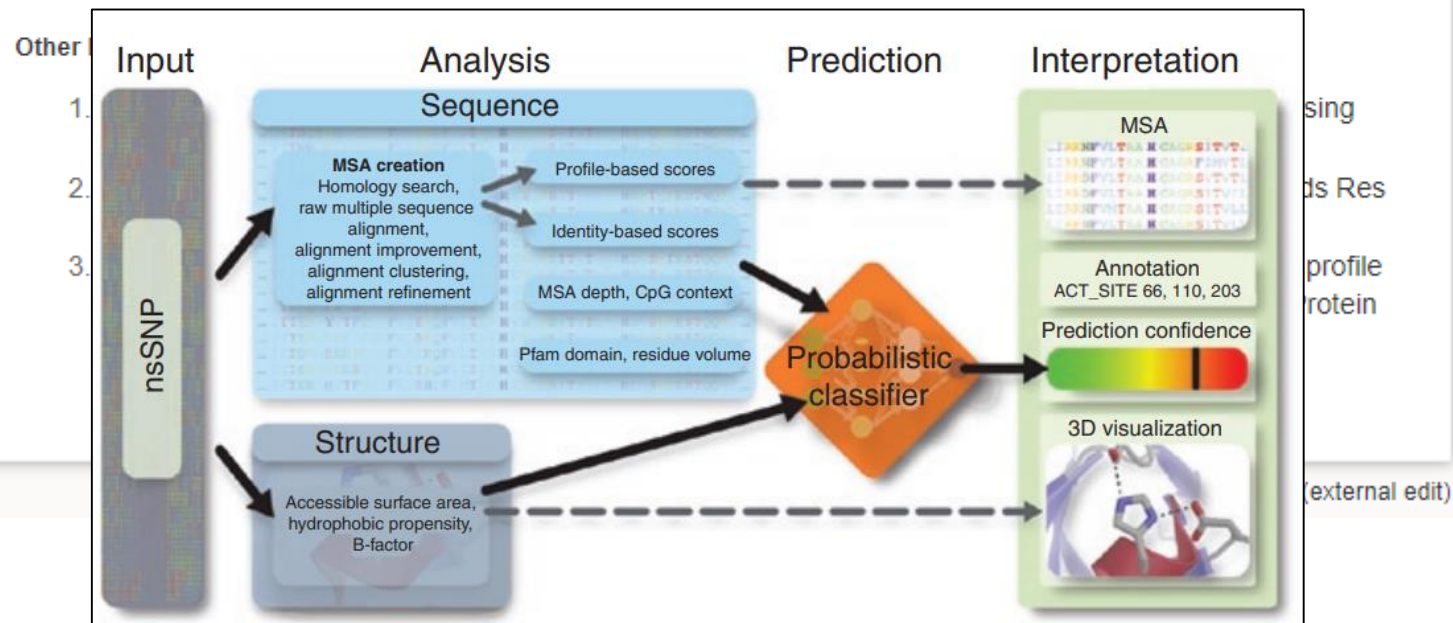# PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/)

## About

**PolyPhen-2** is a new development of the **PolyPhen** tool for annotating coding nonsynonymous SNPs. Some of the highlights of the new version are:

- High quality multiple sequence alignment pipeline
- Probabilistic classifier based on machine-learning method
- Optimized for high-throughput analysis of the next-generation sequencing data

マルチプルアライメントによる進化的保存性とタンパク質構造を加味した機械学習によるアノテーション

- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. Nat Methods 7(4):248-249 (2010). ⊕ PubMed 🔲 PDF 🔲 Supplemental Information

# PolyPhen-2の使い方

*BRCA2* NM_000059.3:c.53G>A, p.(**Arg18His**)



ミスセンス(nonsynonymous)バリアントのみ

**HumDiv- and HumVar-trained PolyPhen-2モデルを選ぶ必要ある**

**HumDiv** - **3,155 damaging alleles in Mendelian diseases, 6,321 differences between human protein and mammalian homologs(non-damaging)**

- **Genome-wide association studies**

**HumVar** – **13,032 human disease-causing mutations, 8,946 human nonsynonymous SNV without annotated involvement of disease(non-damaging)**

– **Diagnostic of Mendelian diseases**

# マルチプルアライメントの図と（利用可能な場合は）3次元構造の図が見られる



STXBP1のC180Y

# Local serverでscoreをつける

**1. 右のsource codeをダウンロードしてインストール（ハードル高い）**

**2.** ミスセンスバリアントだけなので、
**Annovar** (https://annovar.openbioinformatics.org/en/latest/)
**SnpEff** (http://pcingola.github.io/SnpEff/)
で**dbNSFP**を利用する

## Downloads

Downloadable material.

## Software

### Licensing

The software provided herein is free for academic instruction and research use only. Commercial licenses are available to legal entities, including companies and organizations (both for-profit and non-profit), requiring the software for general commercial use. To obtain a commercial license please, contact us via ✉ e-mail.

### Disclaimer

This software is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. The code should not be modified and/or redistributed without the permission of the authors.

| Item | File(s) |
|---|---|
| PolyPhen-2 standalone source code, see INSTALL file included for installation instructions | polyphen-2.2.3r407.tar.gz |

## Databases

| Item | File(s) |
|---|---|
| PolyPhen-2 bundled databases (6.5G) | polyphen-2.2.3-databases-2021_05.tar.bz2 |
| Precomputed MLC protein alignments (2.4G) | polyphen-2.2.2-alignments-mlc-2011_12.tar.bz2 |
| Precomputed MultiZ genomic alignments (1.8G) | polyphen-2.2.2-alignments-multiz-2009_10.tar.bz2 |
| PDB/DSSP structural databases snapshot (38G) | polyphen-2.2.3-databases-pdb-dssp-2021_05.tar.bz2 |
| MD5 checksums for all databases tarballs | md5sum.txt |

## Datasets

| Item | File(s) |
|---|---|
| HumDiv & HumVar training sets | README |
| Whole human exome sequence space annotations | README |

## Mendelian Clinically Applicable Pathogenicity (M-CAP) Score

M-CAP is the first pathogenicity classifier for rare missense variants in the human genome that is tuned to the high sensitivity required in the clinic (see Table). By combining previous pathogenicity scores (including SIFT, Polyphen-2 and CADD) with novel features and a powerful model, we attain the best classifier at all thresholds, reducing a typical **進化的保存性** exome/genome rare (<1%) missense variant (VUS) list from 300 to 120, while never mistaking 95% of known pathogenic variants as benign. Further details can be found here.

3/25/2019: M-CAP score file with an interpretable sensitivity score can be found here! Sensitivity score ≤ 0.95 is possibly pathogenic.

| Method | Authors' Recommended Pathogenicity threshold | Misclassified known pathogenic variants |
|---|---|---|
| SIFT | < 0.05 | 38% |
| Polyphen-2 | > 0.8 | 31% |
| CADD | > 20 | 26% |
| MetaLR | > 0.5 | 27% |
| M-CAP | > 0.025 | 5% |

**機械学習**

### Score a variant

Enter the GRCh37/hg19 coordinate for a missense variant to retrieve its M-CAP score.

| GRCh37/hg19 | e.g. 16:76,486,622 | Go! | Demo! |

### How to cite

Jagadeesh, K., Wenger, A., Berger, M., Guturu, H., Stenson, P., Cooper, D., Bernstein, J., and Bejerano, G. (2016). M-CAP eliminates a majority of variants with uncertain significance in clinical exomes at high sensitivity. Nature Genetics, 2016. 48 (12) 1581 DOI: 10.1038/ng.3703

**稀 (≤ 1% allele frequency)**
**ミスセンス(nonsynonymous)バリアントのみ**

### Download M-CAP Scores

M-CAP only scores rare missense variants: hg19, ENSEMBL 75 missense, ExAC v0.3 where no super population has minor allele frequency above 1%. If a missense variant has no M-CAP score, the M-CAP prediction should be assumed to be likely benign.

- M-CAP v1.0 scores 10/24/2016 please do not use
- M-CAP v1.3 raw scores 10/30/2018

**スコアをダウンロード可能**

- M-CAP v1.4 raw and normalized scores 3/25/2019 (v1.3 raw score of 0.025 converted to sensitivity score 0.95, to facilitate interpretation)

# M-CAPの使い方

## WEBで一個ずつ

**BRCA2 NM_000059.3:c.53G>A, p.(Arg18His)**
**NC_000013.10:g.32890650G>A (GRCh37)**



## Local serverで

1.  **Annovar** (https://annovar.openbioinformatics.org/en/latest/)
    **SnpEff** (http://pcingola.github.io/SnpEff/)
    で**dbNSFP**を利用

2.  **mcap_v1_4.txt**を使って、**Annovar**
    のgeneric databaseとしてアノテーション

# PROVEAN
## (http://provean.jcvi.org/index.php)

アミノ酸置換（ミスセンスバリアント）とインフレームの挿入や欠失も判断可能

バリアント配列とタンパク質配列ホモログ間の類似性

アミノ酸配列からでも、ゲノム配列からでもどちらでも解析可能

SIFTの開発者が2008-2010年にジョン・クレイグ・ヴェンター研究所に在籍していたためか、SIFTのスコアも出力される

# PROVEAN Protein Batchの使い方

## PROVEAN HUMAN PROTEIN BATCH

This tool provides PROVEAN prediction for all human proteins and variants. It also shows SIFT predictions when precomputed scores are available.

- Input: A list of **human protein variants**. See example.
- Output: PROVEAN scores and predictions along with available SIFT predictions. See example.

### PROVEAN Tools
- PROVEAN Protein
- PROVEAN Protein Batch
  - Human
  - Mouse
- PROVEAN Genome Variants
  - Human
  - Mouse
→ About
→ FAQ
→ News
→ Download
→ Help
→ Contact Us
→ Related Links

## Enter a list of human protein variants

Paste in your protein variants: [format]

```
P51587 18 R H
P51587 18 R .
```

Example (upload example)

```
ENSP00000224605 63 A S
ENSP00000224605 55 D G
ENSP00000443112 651 T .
ENSP00000359240 59 Q QA
NP_000483.3 508 F .
P13569 508 F .
```

**BRCA2** NM_000059.3:c.53G>A, p.(Arg18His)
**UniProt P51587**

Each varaint is represented in comma-separated (or space-separated) values as the following:
**<protein ID>,<position>,<reference amino acids>,<variant amino acids>,<comment(optional)>**

- **Protein ID:**
  Ensembl Protein ID, NCBI RefSeq ID, or UniProt Accession ID (for human)
  Ensembl Protein ID (for mouse)
- **Position:**
  Reference position, with the 1st amino acid having position 1.
- **Reference amino acids:**
  One or more amino acids in the query protein sequence. The value in the "Position" field refers to the position of the first amino acid in this field. At least one amino acid is required.
- **Variant amino acids:**
  Amino acids for variant protein sequence. The amino acids in the "Reference amino acid" field are replaced by the amino acids in this field. A deletion can be described by using a period ("."), which means no amino acids (empty).

Or upload a file containing variants (1MB limit):
ファイルを選択 選択されていません

## Email (optional)

If provided, results will also be sent via email.

送信 リセット

25

# PROVEAN Protein Batchの結果

→ PROVEAN Tools
   PROVEAN Protein
   PROVEAN Protein Batch
     Human
     Mouse
   PROVEAN Genome Variants
     Human
     Mouse
→ About
→ FAQ
→ News
→ Download
→ Help
→ Contact Us
→ Related Links

## Results

**Total number of input variants: 2**

| View result table | Download |

- Job ID: 2423309096385730
- Submitted at 0:28:39 EST, Thursday, Dec 2, 2021
- Started at 00:28:41 EST, Thursday, Dec 2, 2021
- Finished at 00:28:42 EST, Thursday, Dec 2, 2021

* The results are kept for 48 hours.

## Cutoff=-2.5 - deleterious or neutral

| VARIATION | | PROTEIN SEQUENCE CHANGE | | | |
|---|---|---|---|---|---|
| ROW_NO. | INPUT | PROTEIN_ID | POSITION | RESIDUE_REF | RESIDUE_ALT |
| 1 | P51587 18 R H | P51587 | 18 | R | H |
| 2 | P51587 18 R . | P51587 | 18 | R | . |

| PROVEAN PREDICTION | | | | SIFT PREDICTION | | | |
|---|---|---|---|---|---|---|---|
| SCORE | PREDICTION (cutoff=-2.5) | #SEQ | #CLUSTER | SCORE | PREDICTION (cutoff=0.05) | MEDIAN_INFO | #SEQ |
| -1.17 | Neutral | 147 | 30 | 0.310 | Tolerated | 2.97 | 42 |
| -3.63 | Deleterious | 147 | 30 | NA | NA | NA | NA |

**SIFT nonsynonymous single nucleotide variants 0.23**

PROVEAN Genome Variantsの使い方と結果

**BRCA2** NM_000059.3:c.53G>A, p.(Arg18His)
NC_000013.10:g.32890650G>A (GRCh37)

SIFT scoreはツールで異なるので注意が必要

27

**1. 右のsource codeをダウンロードしてインストール（ハードル高い）**

**2.** ミセンスバリアントだけを考えるなら
**Annovar** (https://annovar.openbioinformatics.org/en/latest/)
**SnpEff** (http://pcingola.github.io/SnpEff/)
で**dbNSFP**を利用する

DOWNLOADS

PROVEAN software

- PROVEAN source code
- README for installation and usage instructions
- LICENSE

PROVEAN precomputed scores

- PROVEAN scores (v1.1) on all possible single AA substitutions and deletions in human proteins from Ensembl 66 (SIFT scores are also available for single AA substitutions)
- PROVEAN scores (v1.1) on *Human dbSNP 137* coding variants
- PROVEAN scores (v1.0) on *UniProt human polymorphisms and disease mutations* dataset (Release: 2011_09, 20821 disease variants and 36825 polymorphisms)
- PROVEAN scores (v1.0) on indel dataset collected from UniProt/Swiss-Prot database (Release: 2011_09)
- PROVEAN scores (v1.0) on indel dataset collected from 1000 Genomes Project (Release: August 2010)

Prerequisite software and database

- NCBI BLAST (download)
- CD-HIT (download, we recommend not using v4.6 or v4.6.1 (known issue))
- NCBI nr protein database (download latest release or download August 2011 release used in our publication)

# MutationTaster
## (https://www.genecascade.org/MutationTaster2021/#transcript)

進化的保存性

CliVar, HGMD(Human Gene Mutation Database)で病的と判断されるバリアントを病的バリアントして、ホモ接合性バリアントを持つ人がいるバリアントを良性バリアントとして学習

スプライシング予測はMaxEntScan(2004)を使用

VCFファイルを取り扱える

*W446–W451 Nucleic Acids Research, 2021, Vol. 49, Web Server issue*

# Specific transcriptでの解析

*BRCA2* NM_000059.3:c.53G>A, p.(Arg18His)
NC_000013.10:g.32890650G>A (GRCh37)



古い方のMutationTaster(https://www.mutationtaster.org/)



**サンガーシークエンスのデータがあるときに、バリアントの記述が正しいかの確認にも使える**



32

# CADD
**https://cadd.bihealth.org/**

**SNVだけでなくinsertion/deletionもスコアをつけることが可能**

**シュミレーションで生じるバリアントと自然選択を生き延びてきた実際に認められるバリアントを対比することで、複数のアノテーションを統合**

**コーディング領域だけでなく、ノンコーディング領域もスコアを付けられる**

**CADD v1.6ではDeep learningによるスプライスへの影響予測が改善**
MMSplice (Cheng et al. Genome Biology, 2019)
SpliceAI (Jaganathan et al. Cell, 2019)

---

CADD - Combined Annotation Dependent Depletion

News  Score  SNV  Downloads  About  (Alt. site: 🇺🇸)

## What is Combined Annotation Dependent Depletion (CADD)?

CADD is a tool for scoring the deleteriousness of single nucleotide variants as well as insertion/deletions variants in the human genome.

While many variant annotation and scoring tools are around, most annotations tend to exploit a single information type (e.g. conservation) and/or are restricted in scope (e.g. to missense changes). Thus, a broadly applicable metric that objectively weights and integrates diverse information is needed. Combined Annotation Dependent Depletion (CADD) is a framework that integrates multiple annotations into one metric by contrasting variants that survived natural selection with simulated mutations.

C-scores strongly correlate with allelic diversity, pathogenicity of both coding and non-coding variants, and experimentally measured regulatory effects, and also highly rank causal variants within individual genome sequences. Finally, C-scores of complex trait-associated variants from genome-wide association studies (GWAS) are significantly higher than matched controls and correlate with study sample size, likely reflecting the increased accuracy of larger GWAS.

CADD can quantitatively prioritize functional, deleterious, and disease causal variants across a wide range of functional categories, effect sizes and genetic architectures and can be used prioritize causal variation in both research and clinical settings.

In addition to this website, CADD has been described in three publications. The most recent manuscript describes CADD-Splice (CADD v1.6), the latest extension of CADD to improve its predictions of splicing effects:

Rentzsch P, Schubach M, Shendure J, Kircher M.
*CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores.*
Genome Med. 2021 Feb 22. doi: 10.1186/s13073-021-00835-9.
PubMed PMID: 33618777.

Our second manuscript describes the updates between the initial publication and CADD v1.4, introduces CADD for GRCh38 and explains how we envision the use of CADD. It was published by Nucleic Acids Research in 2018:

Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M.
*CADD: predicting the deleteriousness of variants throughout the human genome.*
Nucleic Acids Res. 2018 Oct 29. doi: 10.1093/nar/gky1016.
PubMed PMID: 30371827.

The original manuscript describing the method was published by Nature Genetics in 2014:

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J.
*A general framework for estimating the relative pathogenicity of human genetic variants.*
Nat Genet. 2014 Feb 2. doi: 10.1038/ng.2892.
PubMed PMID: 24487276.

33

まとめて解析

GRCh37/38
バージョンを指定可能

You successfully uploaded variants and the scoring of your variants finished.

You can download your output file here. Make sure that your browser does not alter the file extension (.tsv.gz) during download; otherwise your operating system will not be able to automatically pick the right programs for opening the output.

Legal notice / Impressum

| #CHROM | POS | ID | REF | ALT | | RawScore | PHRED | |
|---|---|---|---|---|---|---|---|---|
| 7 | 5999131 | rs376258383 | C | T | | 2.32179 | 21.7 | PMS2:NM_001322008:exon4:c.364G>A:p.G122S |
| 9 | 95479063 | . | G | C | ミスセンス (nonsynonymous) | 1.743024 | 17.32 | PTCH1:NM_000264:exon8:c.1152C>G:p.H384Q |
| 9 | 95508222 | . | C | A | | 2.1063 | 20.2 | PTCH1:NM_000264:exon1:c.140G>T:p.R47L |
| 13 | 32316513 | rs80358762 | G | A | | 3.466603 | 24.6 | BRCA2:NM_000059:exon2:c.53G>A:p.R18H |
| 19 | 2111827 | rs747848580 | T | C | | 2.529657 | 22.5 | AP3D1:NM_003938:exon23:c.2603A>G:p.K868R |
| 20 | 63419667 | rs201750561 | C | A | | 2.757074 | 22.9 | KCNQ2:NM_172107:exon12:c.1253G>T:p.G418V |
| 14 | 87939915 | rs138577661 | A | G | | 4.045992 | 27.3 | GALC:NM_001201401:exon15:c.1832T>C:p.L611S |
| 17 | 45983222 | rs151115928 | C | T | | 0.938738 | 10.86 | MAPT:NM_001377265:exon5:c.643C>T:p.P215S |
| 16 | 1474841 | . | GCAGCCCCAGGCC | G | コーディングのインフレーム欠失 | 1.892505 | 18.43 | CLCN7:NM_001114331:exon1:c.122_133del:p.G41_A44del |
| 16 | 72788110 | rs727502780 | TGCTGCTGCTGCTGCTGTAGTTGCC | T | | 2.086341 | 19.91 | ZFHX3:NM_001164766:exon9:c.7400_7423del:p.R2467_Q2474del |
| 12 | 31084974 | rs778497594 | GGAA | G | | 1.9976 | 19.24 | DDX11:NM_001257144:exon5:c.487_489del:p.E166del |
| 14 | 58364422 | . | AAAC | A | | 0.494642 | 6.448 | ARID4A:NM_002892:exon20:c.2334_2336del:p.T779del |
| 9 | 131228152 | . | T | C | | 1.975673 | 19.07 | NM_005085:exon33:c.5903-8T>C |
| 11 | 65660221 | rs746773448 | C | T | | -0.216828 | 0.47 | NM_001243985:exon5:c.336-6G>A |
| 5 | 6623292 | rs368018490 | GA | G | | 0.147205 | 2.547 | NM_001193455:exon4:c.361-8T>- |
| 19 | 13919211 | rs371243328 | T | C | イントロン | -0.730796 | 0.037 | NM_017721:exon11:c.1222+9T>C |
| 18 | 70060049 | rs141673274 | T | C | | 2.649794 | 22.7 | NM_001318520:exon35:c.2012-7A>G |
| 1 | 11802878 | rs1413355 | T | C | | 2.089058 | 19.93 | NM_005957:exon2:c.236+3A>G |
| 18 | 58735336 | . | T | C | | 0.969444 | 11.24 | NM_006785:exon13:c.1603+7T>C |
| 5 | 147828020 | rs148954387 | A | G | | 4.06441 | 27.5 | NM_003122:exon4:c.194+2T>C |
| 2 | 201187781 | rs768931544 | G | A | | -0.456302 | 0.147 | |
| 19 | 15238842 | . | G | A | | -0.276793 | 0.349 | |
| 17 | 58358303 | . | A | G | Synonymous | -0.36671 | 0.226 | |
| 3 | 50350599 | . | C | G | | 1.282571 | 14.26 | |
| 19 | 13208933 | . | C | T | | 0.641051 | 7.906 | |
| 20 | 8788472 | rs577076166 | C | A | | 0.205429 | 3.186 | |

# CADDスコアの評価

| RawScore | PHRED |
|---:|---:|
| 2.32179 | 21.7 |
| 1.743024 | 17.32 |
| 2.1063 | 20.2 |
| 3.466603 | 24.6 |
| 2.529657 | 22.5 |
| 2.757074 | 22.9 |
| 4.045992 | 27.3 |
| 0.938738 | 10.86 |
| 1.892505 | 18.43 |
| 2.086341 | 19.91 |
| 1.9976 | 19.24 |
| 0.494642 | 6.448 |
| 1.975673 | 19.07 |
| -0.216828 | 0.47 |
| 0.147205 | 2.547 |
| -0.730796 | 0.037 |
| 2.649794 | 22.7 |
| 2.089058 | 19.93 |
| 0.969444 | 11.24 |
| 4.06441 | 27.5 |
| -0.456302 | 0.147 |
| -0.276793 | 0.349 |
| -0.36671 | 0.226 |
| 1.282571 | 14.26 |
| 0.641051 | 7.906 |
| 0.205429 | 3.186 |

Raw scoreが高いほど、実際の集団中には見られず、シュミレーションで認められるバリアントである
⇒ "deleterious effects"を持つ可能性が大

総バリアントをランキング化し、スケール化されたCADDスコアが"PHRED"

| Phred score | 塩基の 正確性 | | CADD PHREDの場合 |
|---|---|---|---|
| 40 | 99.99 % | | Top 0.01% -- CADD-40 |
| 30 | 99.9 % | | Top 0.1% -- CADD-30 |
| 20 | 99 % | | Top 1% -- CADD-20 |
| 10 | 90 % | | Top 10% -- CADD-10 |

**Cut off値はカットオフを10から20の間のどこかに置くことを推奨**
15 (v1.0でのスプライスサイトバリアントとミスセンスバリアントの中央値）

# SNVは個別入力も可能

**BRCA2** NM_000059.3:c.53G>A, p.(Arg18His)
NC_000013.11:g.32316513G>A (GRCh38)
NC_000013.10:g.32890650G>A (GRCh37)

## CADD - Combined Annotation Dependent Depletion

News　Score　SNV　Downloads　About　(Alt. site: 🇺🇸)

Single
Range

### Single nucleotide variant (SNV) lookup

This form allows you to quickly access the score (and annotation) of a single nucleotide variant (SNV) or all scores at a specific genomic position. If you are investigating multiple or even ranges of CADD SNV scores, please have a look at our Multi-SNV scoring form. Please note that copying and downloading of the annotation tables is not working in Internet Explorer.

CADD scores are freely available for all non-commercial applications. If you are planning on using them in a commercial application, please contact us.

| Chromosome: | 22 | Position: | 43451447 |
| | T | Alt (optional): | A |
| GRCh37-v1.6 ▼ | | INCLUDE ANNOTATIONS | TRANSPOSE TABLE |

### Single nucleotide variant (SNV) lookup

This form allows you to quickly access the score (and annotation) of a single nucleotide variant (SNV) or all scores at a specific genomic position. If you are investigating multiple or even ranges of CADD SNV scores, please have a look at our Multi-SNV scoring form. Please note that copying and downloading of the annotation tables is not working in Internet Explorer.

CADD scores are freely available for all non-commercial applications. If you are planning on using them in a commercial application, please contact us.

| Chromosome: | 13 | Position: | 32316513 |
| Ref (optional): | T | Alt (optional): | A |
| CADD model: | GRCh38-v1.6 ▼ | INCLUDE ANNOTATIONS | TRANSPOSE TABLE |

**LOOKUP VARIANT(S)**

**The (maximum) CADD score at this position is 24.8. (Request: Chromosome 13, Position 32316513, CADD GRCh38-v1.6)**

Details:

COPY TABLE    SAVE TABLE

| Chrom | Pos | Ref | Alt | RawScore | PHRED |
|-------|-----|-----|-----|----------|-------|
| 13 | 32316513 | G | A | 3.466603 | 24.6 |
| 13 | 32316513 | G | C | 3.532795 | 24.8 |
| 13 | 32316513 | G | T | 2.795310 | 23.0 |

...cess the score (and annotation) of multiple single nucleotide variants (SNV). If you are investigating longer ranges of ...interested in our UCSC Genome Browser Tracks for hg19/GRCh37 and hg38/GRCh38. Please note that copying and ...bles is not working in Internet Explorer.

...for all non-commercial applications. If you are planning on using them in a commercial application, please contact us.

| 22 | | |
| 43451447 | End: | 43451450 |
| GRCh37-v1.6 ▼ | INCLUDE ANNOTATIONS | TRANSPOSE TABLE |

**LOOKUP RANGE**

# Local serverで自分のVCFファイルにアノテーションすることも可能

## Developmental release: v1.6 [release notes]
### Offline scoring scripts for v1.6

conda環境とsnakemakeが必要

| Scripts and README for offline installation | file (360K) |
|---|---|

## Genome build GRCh38 / hg38

| Description | Link (Size) | Tabix Index (Size) |
|---|---|---|
| All possible SNVs of GRCh38/hg38 | US \| DE (81G) | US \| DE (2.6M) |
| All possible SNVs of GRCh38/hg38 incl. all annotations | US \| DE (313G) | US \| DE (2.7M) |
| All gnomad release 3.0 InDels to initiate a local setup | US \| DE (1.1G) | US \| DE (1.8M) |
| All gnomad release 3.0 InDels incl. all annotations to initiate a local setup | US \| DE (7.2G) | US \| DE (2.5M) |
| All gnomad release 3.0 SNVs | US \| DE (5.9G) | US \| DE (2.4M) |
| MD5Sums | US \| DE | |
| All GRCh38 annotations required for offline installation | US \| DE (206G) | |

## Genome build GRCh37 / hg19

# SpliceAI Lookup （https://spliceailookup.broadinstitute.org/）

Enter a variant below to see its SpliceAI scores. This can be any SNV or simple InDel within an exon or intron defined by GENCODE v38.

**Translational Genomics Group**

This web interface and the underlying SpliceAI server are being developed by the TGG at the Broad Institute. The SpliceAI server optionally uses pre-computed scores provided by Illumina for SNVs and small InDels, but resorts to running the SpliceAI model directly for scores not available in the pre-computed files - such as larger InDels or non-default "Max distance" values.

NOTE: The API currently only supports interactive use. For automated batch processing, please use the source code to set up your own local instance of the API server or just run the underlying SpliceAI tool.

For more information on the SpliceAI model, see Jaganathan et al. Cell 2019 and github.com/Illumina/SpliceAI, as well as in this recorded talk by Kishore Jaganathan.

Examples (on hg38):

    chr8-140300616-T-G

    chr8  140300616  T  G

    NM_000552.4(VWF):c.3797C>A (p.Pro1266Gln)

more examples...

| 8:140300616 T>G | 'chr' prefix is optional |
| 1:930130:C:G | position with overlapping genes |
| 1-1042601-A-AGAGAG | insertion of GAGAG |
| 1-1042466-GGGC-G | deletion of GGC |
| NM_000249.4(MLH1):c.116G>A | another HGVS exapmle |

Enter a variant...

*What's New*

To post issues or feature requests, please use SpliceAI-lookup/issues

**June 2, 2021**
- show RefSeq ids for the subset of Ensembl ENST ids (~30%) that have one or more matching RefSeq NM ids according to Ensembl. See issue #8 for details.
- update to Gencode v38 (was previously on v37)

**April 11, 2021**
- show gray background for non-coding transcripts
- switch to showing all non-coding transcripts. (Previously, transcripts with Gencode biotypes like *lncRNA*, *processed_pseudogene*, *processed_transcript*, *retained_intron*, and *nonsense_mediated_decay* were filtered out if they overlapped *protein_coding* transcripts). See issue #6 for details.
- update to Gencode v37 (was previously on v36)

**Related web tools:**
liftover - simple hg19 <=> hg38 conversion for variants, positions, intervals
CMA search - search OMIM by genomic region, gene name, phenotype, etc.

Genome version:  ◯ hg19  ● hg38

Submit

Score type:  ◯ masked  ● raw  ← **正常のスプライシングが強くなり、新たなスプライシングが弱くなるような予測をマスクするか**

Max distance:  50  ← **バリアントによって影響を受ける（消失あるいは出現）スプライス部位をどの範囲で調べるか**
**Max 10,000**

Use Illumina's pre-computed scores:  ◯ yes  ● no  ?
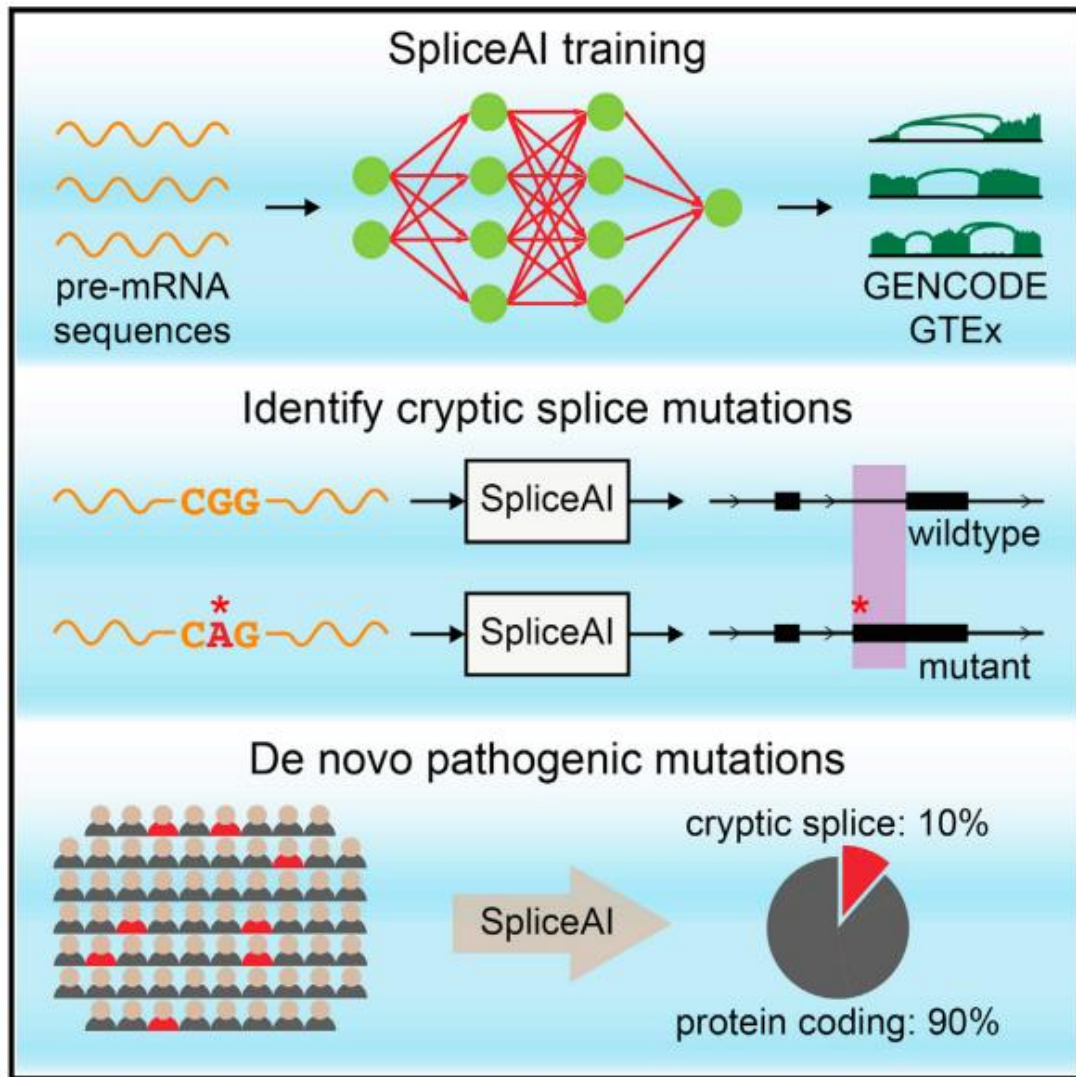
**イルミナ社が作ったプログラム**

**Deep learning (Deep neural network)で、pre-mRNA配列からスプライシングを予測**

**GENCODE-annotated pre-mRNA配列を使ってトレーニング**

**予測された潜在的(cryptic)なスプライス部位を使う転写産物の75%を、RNA-seqで確認**

**潜在的なスプライス部位を使うような異常な転写は、神経発達障害の~10%を占めると考えられる**

Jaganathan et al, Cell. 2019 24;176(3):535-548.e24

**GNAO1** NM_020988.3:c.724-8G>A
NC_000016.10:g.56351376G>A(GRCh38)
NC_000016.9:g.56385288G>A(GRCh37)

----ACGgt--tccttcctgc**g**gccgcagAAC---

-8    5789

**A**gCCGCAGAAC

16:56351376G>A

**Genome version:** ○ hg19 ● hg38

**Score type:** ● masked ○ raw ⑦

**Max distance:** 50 ⑦

**Use Illumina's pre-computed scores:** ○ yes ● no ⑦

Submit

| variant | gene | Δ type | Δ score ⑦ | pre-mRNA position ⑦ |
|---|---|---|---|---|
| 16-56351376-G-A | GNAO1 (ENSG00000087258.16 / ENST00000262493.12 / NM_020988.3) biotype: protein coding | Acceptor Loss | 0.92 | 8 bp |
| | | Donor Loss | 0.00 | |
| UCSC, gnomAD | OMIM, GTEx, gnomAD, ClinGen, Ensembl, Decipher, GeneCards | Acceptor Gain | 1.00 | 2 bp |
| | | Donor Gain | 0.00 | |

| GAA | GAC | GAA | ACC | ACG | CCG | CAG | AAC | CGC | ATG | CAC | GAG | TCT | CTC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Glu | Asp | Glu | Thr | Thr | Pro | Gln | Asn | Arg | Met | His | Glu | Ser | Leu |
| 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 |

G AAG AC G AAA CCA CG C CG CAG AA C C GC ATG CAC GA G TCT CT C

exon6 ←←6bp→→ exon7

6bp insertion

バリアント±50-bpでAcceptorあるいはDonorが消失・出現するかを計算し、最もスコアが高い位置を表示。

正の値が3′側、負の値が5′側の位置を示す。

**Δ scoreの
カットオフ値**
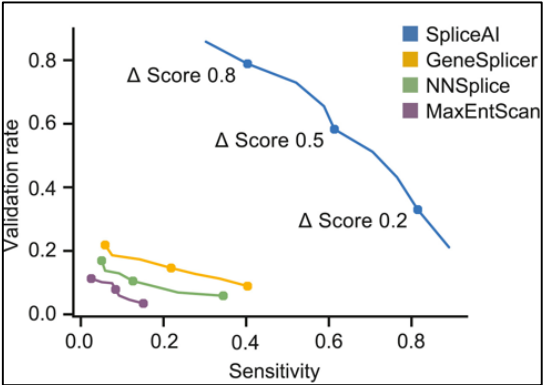
0.2 (感度、再現率が高い)
0.5 (推奨)
0.8 (適合度、精度が高い)

Jaganathan et al, Cell. 2019 24;176(3):535-548.e24

**FBN1** NM_000138.4:c.5789-15G>A
NC_000015.10:g.48445519C>T(GRCh38)
NC_000015.9:g.48737716C>T(GRCh37)

-15　　　5789
--TAGgt--aggattgttg**g**ggattttctgcagATG--
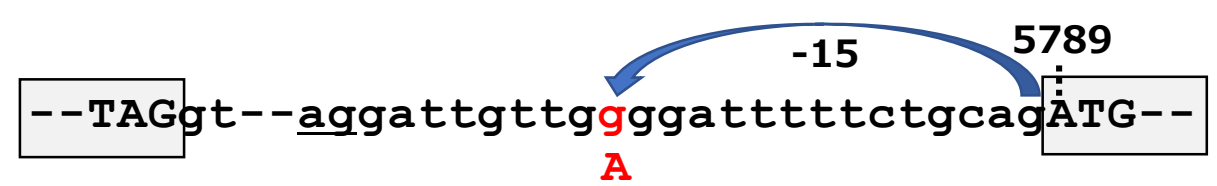
**イントロン保持**

15:48445519C>T

Genome version: ○ hg19 ● hg38    **Submit**

Score type: ● masked ○ raw ⑦

Max distance: 50  ⑦

Use Illumina's pre-computed scores: ○ yes ● no ⑦

FBN1はアンチセンス鎖に位置しているので、ゲノムで上流15-bpが通常のアクセプター部位

| variant | gene | Δ type | Δ score ⑦ | pre-mRNA position ⑦ |
|---|---|---|---|---|
| 15-48445519-C-T | FBN1 (ENSG00000166147.15 / ENST00000316623.10 / NM_000138.5) | Acceptor Loss | 0.24 | -15 bp |
| UCSC, gnomAD | biotype: protein coding canonical transcript | Donor Loss | 0.00 | |
| | OMIM, GTEx, gnomAD, ClinGen, Ensembl, Decipher, GeneCards | Acceptor Gain | 0.08 | 8 bp |
| | | Donor Gain | 0.00 | |

Max distance: 1000

**Max distanceは大きい方が良さそう**

| variant | gene | Δ type | Δ score ⑦ | pre-mRNA position ⑦ |
|---|---|---|---|---|
| 15-48445519-C-T | FBN1 (ENSG00000166147.15 / ENST00000316623.10 / NM_000138.5) | Acceptor Loss | 0.24 | -15 bp |
| UCSC, gnomAD | biotype: protein coding canonical transcript | Donor Loss | 0.00 | |
| | OMIM, GTEx, gnomAD, ClinGen, Ensembl, Decipher, GeneCards | Acceptor Gain | 0.62 | 56 bp |
| | | Donor Gain | 0.00 | |

43

POLR3A NM_007055.3:c.645+312C>T
NC_000010.11:g.78024237G>A(GRCh38)
NC_000010.10:g.79783995G>A(GRCh37)

Wild-type | 5 | gt ·········ttttttttgagatggaagaca ······gggattacaggcatgtgcca ···········ag | 6

Mutant | 5 | gt ·········ttttttttgagatggaagaca ······gggattacaggtatgtgcca ···········ag | 6

129bp insertion

Acceptor gain site          Donor gain site

10:78024237G>A

Genome version:  ○ hg19  ● hg38          Submit

Score type:  ● masked  ○ raw  ⑦

Max distance:  1000  ⑦

Use Illumina's pre-computed scores:  ○ yes  ● no  ⑦

| variant | gene | Δ type | Δ score ⑦ | pre-mRNA position ⑦ |
|---|---|---|---|---|
| 10-78024237-G-A | POLR3A (ENSG00000148606.13 / ENST00000372371.8 / NM_007055.4) | Acceptor Loss | 0.00 | |
| UCSC, gnomAD | biotype: protein coding canonical transcript | Donor Loss | 0.00 | |
| | OMIM, GTEx, gnomAD, ClinGen, Ensembl, Decipher, GeneCards | Acceptor Gain | 0.55 | 130 bp |
| | | Donor Gain | 0.77 | 2 bp |

44

# Local serverで自分のVCFファイルにアノテーションすることも可能

## SpliceAI: A deep learning-based tool to identify splice variants

`release v1.3.1` `license GPLv3` `downloads 115k`

This package annotates genetic variants with their predicted effect on splicing, as described in Jaganathan et al, Cell 2019 in press. The annotations for all possible substitutions, 1 base insertions, and 1-4 base deletions within genes are available here for download. These annotations are free for academic and not-for-profit use; other use requires a commercial license from Illumina, Inc.

### License

SpliceAI source code is provided under the GPLv3 license. SpliceAI includes several third party packages provided under other open source licenses, please see NOTICE for additional details. The trained models used by SpliceAI (located in this package at spliceai/models) are provided under the CC BY NC 4.0 license for academic and non-commercial use; other use requires a commercial license from Illumina, Inc.

### Installation

The simplest way to install SpliceAI is through pip or conda:

```
pip install spliceai
# or
conda install -c bioconda spliceai
```

Alternately, SpliceAI can be installed from the github repository:

```
git clone https://github.com/Illumina/SpliceAI.git
cd SpliceAI
python setup.py install
```

SpliceAI requires `tensorflow>=1.2.0`, which is best installed separately via pip or conda (see the TensorFlow website for other installation options):

```
pip install tensorflow
# or
conda install tensorflow
```

- ✓ **Python環境が必要**
- ✓ **pipやcondaで楽にインストール可能**
- ✓ **機械学習用のtensorflowを使って動く**
- ✓ **インプットのVCFをアウトプットのVCFを指定するだけで動く（シンプル）**
- ✓ **結果はINFO fieldに記載される**
- ✓ **MaxDistance(-D)を増やすと時間はかかる**

45

# ESEfinder
(http://krainer01.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home)

**Exonic splicing enhancer (ESE)を見つけるプログラムで、ESEsに結合するSR proteinを見つけるのが主な目的**

**SRProteinとSpliceSiteの2つのマトリックスがある**

**スプライス部位の予測もやってくれる**

**<5000-bpで配列を張り付ける必要がある**

last update: 02/27/2007

...AG ESE Finder GU...

release 3.0

Search | background | matrices & thresholds | input & output | caveats

Go to ESEfinder2.0

**Matrix infomration**

Please choose a matrix library: [SpliceSites ▾]

Choose the matrix and the threshold to be used:

| Matrices (select one or more) | Threshold |
|---|---|
| ☑ 5SS_U2_human (5' splice sites (donor) of human (U2 type)) | 6.67 |
| ☑ 3SS_U2_human (3's splice sites (acceptor) of human (U2 type)) | 6.632 |
| ☑ 5SS_U2_mouse (5' splice sites (donor) of mouse (U2 type)) | 6.679 |
| ☑ 3SS_U2_mouse (3's splice sites (acceptor) of mouse (U2 type)) | 6.724 |
| ☑ BranchSite (Mammalian branch site (U2 type)) | 0 |
| | Reset thresholds |

**Sequence infomration**

Enter here your input data in FASTA or MULTI-FASTA format (**<5000nt, accept both 'T' and 'U' as being equivalent**)
(please read important information about search format description)

**Output information**

Override the thresholds above and use one of the following two options:
☐ Report only the best hit in each sequence (instead of hits above the thresholds)
☐ Report all scores in each sequence (instead of hits above the thresholds)

(**Note**: if you check both, the program will report the best hit in each sequence.)

☐ Output as a plain text file

☐ please send the results to the following e-mail address: [_____]

[Send] [Restore defaults]

alternatively, upload a text file: [ファイルを選択] 選択されていません

# UCSC genome browser(https://genome.ucsc.edu/)でバリアント前後の配列を取得

*GNAO1* NM_020988.3:c.724-8G>A
NC_000016.10:g.56351376G>A(GRCh38)
NC_000016.9:g.56385288G>A(GRCh37)



当該遺伝子がアンチセンス鎖にある場合は
チェックを入れた方が分かりやすい

`--ACGgt--tcccttcctgcggccgcagAAC---`

```
>hg38_dna range=chr16:56350376-56352376 5'pad=1000 3'pad=1000 strand=+ repeatMasking=lower
TTCCTCCCACTCCCTGTGCAGAAATGGATTCGTTCACAGAGAAAGATGCC
CTCTGTGCTGCTGAGAGACAGACTCTGCCCAAGTGAATGAGCTCAGCGCC
CCAAAGCAGAGTTTTTGCCCAGGGCAAGAGGCTGCCCCCGTTCGGCATCT
GTGGGGCGCCCATCACGTTCATCCATTTGCTAAACAGGCTCCTGTGTGCA
GGGCAGCCCTCCACATCTGCCTTGCTTAGAACATCATGCAGTCTGTCAGATG
CAGCTGAGTAGGACCCAGCCAGAGAGGTGTGGCCAGTCCCCCAGGTGCT
CCTGGAGGTGGCCTGCCAGGACCTCTCCCCTGAGTGGTGGCACCCAGG
ATGCTGGCAGGACCTGGCATGCTTCCCTGAGACCATAGGAAAGGGCCCG
GAGTCTCTCACCACTCCCGAGAACATCCCCCTACCCAGGATGGGCCCAG
GAGGAGAGGTCTAGGCAGGAAGGGCCAAAGCCCAGCTCTCCGGCTCCGGT
CCAGACTCCCCAACGCCcacacacacaggcacacacacaggcacgcaca
cacactcaagcacatggatgcacacaggcatgcacacacaggtgcacaca
cataggcacacatgcacacgaacacacacaggcacacataacaggtgcat
gcacacacacacaggcacatgcacacagccacacacacGTGTGCTGGGAG
ACCCTCACATAGCTGGGAACTCTGCAGTGGGGAGAGCCCGGGTCACCCTG
TCCTCAGCCTCCCCAGAGAGCCATGGCTGGTGTCCAGTCCAGTCTCTGCC
TCTGGCCCCGACATGACGGGTTCTCCGGAAGCAGGGAATGAGATGTGCTG
TGAGGTGTAACTGACTCAGGTTCCATCTGGCAGCCTCTCGGAGGAGCTGC
CGAGTAGCCCAGTCCCTCTCTGTCAAGCCTAATTCtctccttctctttcc
ctgtctctgtgtctccctcccgctgtctgtcctctctcCTCCCTTCCTGC
GGCCGCGAACCGCATGCACGAGTCTCTCATGCTCTTCGACTCCATCTGT
AACAACAAGTTCTTCATCGATACCTCCATCATTCTCTTCCTCAACAAGAA
AGATCTCTTTGGCGAGAAGATCAAGAAGTCACCTTTGACCATCTGCTTTC
CTGAATACACAGGTGGGTGCCAGGCAGTCCTGTGCAGGGGGGAAGCCTGCC
CCTGACAGGGACCCATGGCTCAGAGAACAGGCGTCGGCCAGCACTGCT
GGGGCTAGCTGGGAGGACCCATTGCAGGAGACTGGTGGGGCGCAAA
AGAAAAACAGTGCTGTGCCACCAGGTTTGGGCTTCCCAGGACACAGCCCT
CAGCGTGGTGGAAATGGCCCCTCCTAAGATATATGTGTTAGGACCAAGTG
ACTCAGGAGTGGTGGGGAGCAGGGGCAGGACAGGATCACAGGCTTCCCC
CTTCCTCCTGGTCACCCTCTAGAAGAGGTCTCAGGACAGCCTCCTGTAAT
CTCAGCAGTGGCCTCCCCTCACCCCTGCCCAGAGCCCCACAGCACCTTTG
CATGAAACGAAAAGTGGACACGTGCCTGTGAACAGGGCCTGGACTGCC
TCTTGGCCCACCTGCTTTCCCTGCTCCTGCTTTGGTGCCCGGAACATG
CTTCATCTCTGTGCAGCGGCCCAGACAGACTCCACCCATCCCACATC
GTGCCTGCCCTCCCTCCCAACCCCAGCCCCATTGCGCCTCCTCTCTCCC
CTTCTTCCTTCGTCCCTGGTCTTTGCCCCAAAATCCAACCACTCCAGCTAT
CTGCAAGCCTCTGTATCACTGCCTCCCTTTCTTCTGCCTTCTCTGGGGC
TTTCTTCTTGCTCTCAGGGGAATGTGTCACAGAACCTTCCCCCCCACCCGG
CTCAATGGCCTCTTTCCAAATGGATGGATGAGGGATGGTGGCTTCCGTG
GAATGGGCCAACTCTGGGCTACTGCTAATGCCTAAAAGCAGCACTGATGG
G
```

**Matrix infomration**

Please choose a matrix library: [SpliceSites ▾]

Choose the matrix and the threshold to be used:

バリアントによって予測スコアが低下すること
を確認するために閾値は0とする

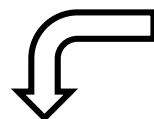| Matrices (select one or more) | Threshold |
|---|---|
| ☑ **5SS_U2_human** (5' splice sites (donor) of human (U2 type)) | 0 |
| ☑ **3SS_U2_human** (3's splice sites (acceptor) of human (U2 type)) | 0 |
| ☑ **5SS_U2_mouse** (5' splice sites (donor) of mouse (U2 type)) | 0 |
| ☑ **3SS_U2_mouse** (3's splice sites (acceptor) of mouse (U2 type)) | 0 |
| ☑ **BranchSite** (Mammalian branch site (U2 type)) | 0 |
| | Reset thresholds |

**Sequence infomration**

Enter here your input data in FASTA or MULTI-FASTA format (<5000nt, accept both 'T' and 'U' as being
(please read important information about search format description)

```
>WT
--gtctctgtgtctccctcccgctgtctgtcctctctcCTCCCTTCCTGC
GGCCGCAGAACCGCATGCACGAGTCTCTCATGCTCTTCGACTCCATCTGT
AACAACAAGTTCTTCATCGATACCTCCATCATTCTCTTCCTCAACAA----
>Variant
--gtctctgtgtctccctcccgctgtctgtcctctctcCTCCCTTCCTGC
aGCCGCAGAACCGCATGCACGAGTCTCTCATGCTCTTCGACTCCATCTGT
AACAACAAGTTCTTCATCGATACCTCCATCATTCTCTTCCTCAACAA---
```

```
                         -8   5789
--ACGgt--tcccttcctgcggccgcagAAC--

AgCCGCAGAAC
```

| 5SS_U2_human threshold: 0 | | | 3SS_U2_human threshold: 0 | | | |
|---|---|---|---|---|---|---|
| Position*/Site/Score | | | Position*/Site/Score | | | |
| 1565 (-437) | GTGGACACGTGGCCTGTGAACAGGGCCTGG | 3.10990 | 985 (-1017) | tctcCTCCCTTCCTGCGGCCGCAGAACCGC | 2.42960 | WT |
| 1622 (-380) | CTGCTCCCTGCTTTGGTGCCCGGAACATGC | 0.56590 | 994 (-1008) | TTCCTGCGGCCGCAGAACCGCATGCACGAG | 8.33480 | |
| 1686 (-316) | CACCCATCCCACATCGTGCCTGCCCTCCCT | 1.37990 | 1084 (-918) | CTCTTCCTCAACAAGAAAGATCTCTTTGGC | 3.30600 | |
| 1565 (-437) | GTGGACACGTGGCCTGTGAACAGGGCCTGG | 3.10990 | 985 (-1017) | tctcCTCCCTTCCTGCaGCCGCAGAACCGC | 2.37420 | Variant |
| 1622 (-380) | CTGCTCCCTGCTTTGGTGCCCGGAACATGC | 0.56590 | 988 (-1014) | cCTCCCTTCCTGCaGCCGCAGAACCGCATG | 12.37800 | |
| 1686 (-316) | CACCCATCCCACATCGTGCCTGCCCTCCCT | 1.37990 | 994 (-1008) | TTCCTGCaGCCGCAGAACCGCATGCACGAG | 8.04330 | |
| 1754 (-248) | CTTCCTTCGTCCCTGGTCTTTGCCCCAAAT | 0.27270 | 1084 (-918) | CTCTTCCTCAACAAGAAAGATCTCTTTGGC | 3.30600 | |

**994が正常のスプライス部位のスコア 8.33480⇒8.04330 微減**
**988に12.37800と高いスコアで新しいアクセプター部位が出現**
**⇒ 新しいアクセプター部位の出現を高い自信をもって予測できている**

**FBN1** NM_000138.4:c.5789-15G>A
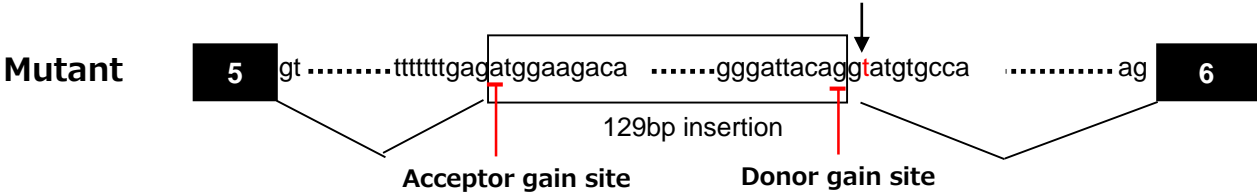NC_000015.10:g.48445519C>T(GRCh38)
NC_000015.9:g.48737716C>T(GRCh37)

-15    5789

`--TAGgt--aggattgttggggattttttctgcagATG--`

A

イントロン保持

| 3SS_U2_human threshold: 0 | | |
|---|---|---|
| Position*/Site/Score | | |
| 1001 (-1001) | GGGATTTTTCTGCAGATGTTGATGAATGTG | 7.20780 |

**WT**

1001が正常のスプライス部位のスコア　7.20780 ⇒ 6.86040 微減
新たなスプライス部位の出現なし

| 1001 (-1001) | aGGGATTTTTCTGCAGATGTTGATGAATGTG | 6.86040 |
|---|---|---|

**Variant**

---

**POLR3A** NM_007055.3:c.645+312C>T
NC_000010.11:g.78024237G>A(GRCh38)
NC_000010.10:g.79783995G>A(GRCh37)

Mutant

| 5 | gt ········ tttttttttgagatggaagaca ········ gggattacaggtatgtgcca ········ ag | 6 |

129bp insertion

Acceptor gain site        Donor gain site

| 5SS_U2_human threshold: 0 | | | 3SS_U2_human threshold: 0 | | |
|---|---|---|---|---|---|
| Position*/Site/Score | | | Position*/Site/Score | | |
| 985 (-1017) | tagctgggattacaggcatgtgccaccatg | 3.94440 | 856 (-1146) | tttttttttttttgagatggaagacagagtc | 7.06220 |

**WT**

| 985 (-1017) | tagctgggattacaggTatgtgccaccatg | 10.88050 | 856 (-1146) | tttttttttttttgagatggaagacagagtc | 7.06220 |

**Variant**

985が新たに出現するドナー部位　3.94440 ⇒ 10.88050 劇的にスコアが増加
856が新たに使われるアクセプター部位 7.06220で変化ないが、スコアはもともと高く潜在的なアクセプター部位

# Take-home message

- ✓ 様々な*in silico*解析プログラムが開発されており、バリアントの解釈や絞り込みに利用可能

- ✓ 機械学習や**Deep learning**といった人工知能を利用するプログラムが最近では主流

- ✓ ただ、*in silico*解析はあくまで**supportive**なエビデンスであることを認識する必要がある（ミスセンスバリアントの解釈の**strong**エビデンスは機能解析で、ハードルが高い）

- ✓ スプライス異常は、逆転写**PCR**で証明可能であり、機能解析を実行するハードルは低い。スプライス異常を起こしうるバリアントをピックアップするためには、*in silico*解析が重要