

はじめに

この資料はRefseqや「NM_」といった遺伝子解析において必要とされる情報やリファレンス配列の変換によるバリエーション表記の確認などについて説明しています。

検査依頼時や結果の解釈でお困りの際は下記をご参照ください。

- [Refseq・codingDNAについて・・・2～5ページ](#)

Refseqの説明、Accession番号、codingDNAの番号、NMの重要性について案内しています。

- [バリエーションの位置検索・確認方法・・・6～13ページ](#)

TransVarを用いたバリエーション位置の検索、Accession番号の確認、バリエーション情報のリファレンス配列変換とMutalyzerでの確認方法についてをご案内します。

RefSeqとは

- Refseq (Reference Sequence) とは、NCBIが提供しているゲノムやタンパク質、転写産物などをコードしている基準となる塩基配列やアミノ酸配列などの参照配列が集約されたデータベースです。 <https://www.ncbi.nlm.nih.gov/refseq/>
- Refseqに登録された配列にはAccession番号といわれるIDが割り振られます。Accession番号には「NM_」や「XM_」などの種類があり、下記のように配列の種類によって接頭語が異なります。 <https://www.ncbi.nlm.nih.gov/books/NBK21091/>

| 接頭語 | 対象分子、種類 |
|-----|---------------------|
| NC_ | 完全なゲノム領域、染色体 |
| NG_ | 不完全なゲノム領域 |
| NM_ | タンパクをコードする転写産物、mRNA |
| NR_ | タンパクをコードしない転写物、RNA |
| NP_ | NM_に関連付けられたタンパク |

Accession番号「NM_」について

- 1つの遺伝子につき、Accession番号が1つとは限らず、下記の*ATP7B*のように基準となるcDNAの参照配列が複数登録されています。(Mutation @ a glanceより引用)
- 弊所では複数のAccession IDのうち、[HGMD\(The Human Gene Mutation Database\)](#)で採用されたものを使用して解析をしております。

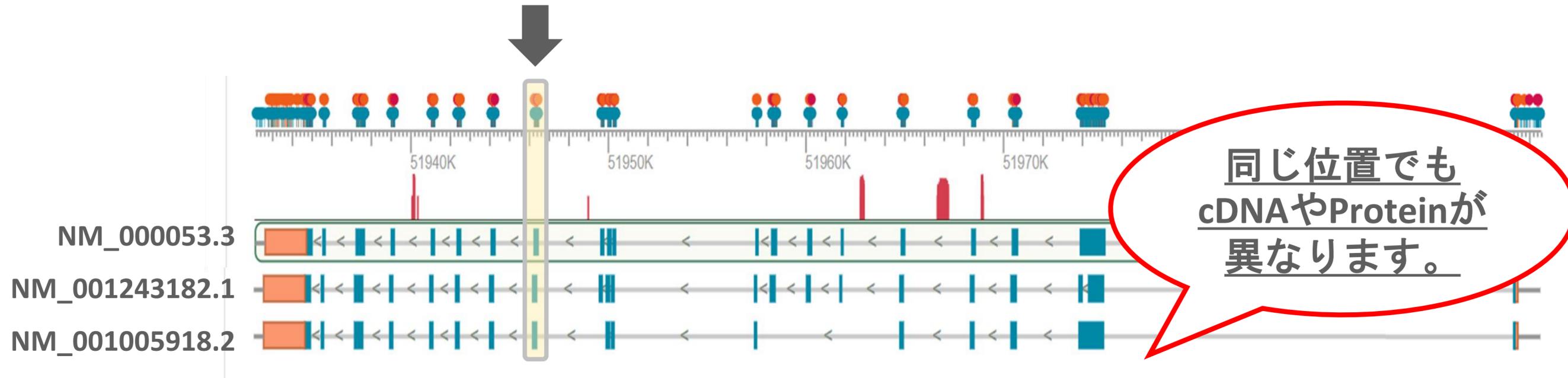
ATP7B ATPase copper transporting beta



矢印で示すような青四角がコーディングシーケンス (CDS) を表しています。

「NM_」の情報がなぜ必要なのか

- 例として、*ATP7B* のchr13:g.51946441A>G のバリエントを調べてみます。



| gDNA (hg38) | Accession ID | cDNA | Protein | Exon位置 |
|---------------|----------------|-----------|---------|---------|
| g.51946441A>G | NM_000053.3 | c.2903T>C | p.I968T | Exon 13 |
| | NM_001243182.1 | c.2570T>C | p.I857T | Exon 14 |
| | NM_001005918.2 | c.2282T>C | p.I761T | Exon 9 |

Accession IDが異なれば同じゲノムポジションでもcDNAやタンパクが異なり、「NM_」の情報がないと正しいバリエントの位置や情報を認識することができません。

リファレンス (hg19、hg38) とは

- ゲノム解析の際に解析データと比較するための基準となる塩基配列のことです。2009年にリリースされた古いものをhg19、2013年の新しいものをhg38といい、配列情報がhg19とhg38間で異なります。
- 前の例として、NM_000053.3 : c.2903T>Cをhg19とhg38で比較すると、リファレンスが異なるとゲノム位置も異なります。

同じバリエーションでも
ゲノムポジションが
異なります。

| リファレンス | Accession ID | gDNA | cDNA | Protein | Exon位置 |
|--------|--------------|---------------|-----------|---------|---------|
| hg38 | NM_000053.3 | g.51946441T>C | c.2903T>C | p.I968T | Exon 13 |
| hg19 | NM_000053.3 | g.52520577A>G | c.2903T>C | p.I968T | Exon 13 |

バリアントの位置やリファレンスの変換方法

- RefseqのAccession IDや「hg38」などのリファレンス（参照配列）が異なることで正しくバリアントを認識できないため、それらを調べる方法をご紹介します。

1. TransVar

- ・ バリアント情報からAccession ID、ゲノムポジションの確認
- ・ リファレンス配列「hg38」および「hg19」の変換

2. Mutalyzer

- ・ TransVarで得られたゲノムポジションの確認

バリアントの位置を検索

- ゲノム位置検索の機能の一例としてMD Anderson Cancer Centerが提供するTransVarを使って説明致します。

<https://bioinformatics.mdanderson.org/transvar/>

THE UNIVERSITY OF TEXAS
MDAnderson
Cancer Center

FACULTY RESEARCH PUBLIC SOFTWARE PUBLIC DATASETS PROGRAM SUPPORT

Department of Bioinformatics and Computational Biology

Home > Public Software > TransVar

TransVar

TransVar is a multi-way annotator for genetic elements and genetic variations.

It operates on genomic coordinates (e.g., chr3:g.178936091G>A) and transcript-dependent cDNA as well as protein coordinates (e.g., PIK3CA:p.E545K or PIK3CA:c.1633G>A, or NM_006218.2:p.E545K, or NP_006266.2:p.G240Afs*50), and was designed to resolve ambiguous mutation annotations arising from differential transcript usage.

TransVar supports

- HGVS nomenclature
- both left-alignment and right-alignment convention in reporting indels.
- annotation of a region based on a transcript dependent characterization
- single nucleotide variation (SNV), insertions and deletions (indels) and block substitutions
- mutations at both coding region and intronic/UTR regions
- transcript annotation from commonly-used databases such as Ensembl, NCBI RefSeq and GENCODE etc
- UniProt protein id as transcript id
- GRCh36, 37, 38
- forward annotation.

TransVar

Overview

Description TransVar is a multi-way annotator for genetic elements and genetic variations.

Development Information

GitHub [zwdzwd/transvar](#)

URL <https://bioinformatics.mdanderson.org/transvar/>

Language Python

Current version 2.4.0

License [The MIT License](#)

Status Active

Last updated September 7, 2018

References

- *ATP7B*, NM_000053.3, c.2903T>C, p.I968Tのバリアントの情報を例としてTransVarでの検索例をご紹介します。

バリアントの位置を検索

- *ATP7B*、NM_000053.3、c.2903T>C、p.I968Tのバリアントを例に Accession IDの確認の手順をご紹介します。

TransVar Annotate About

Select a task:

Reverse Annotation: Protein Reverse Annotation: cDNA Forward Annotation: gDNA Codon Search: Protein

Select a reference genome:

GRCh38/hg38 GRCh37/hg19 GRCh36/hg18

Select one or more annotation databases:

Ensembl CCDS RefSeq GENCODE UCSC AceView

Upload a **plain text** file of identifiers, one identifier per line, case sensitive

Submit

お持ちのバリアント情報を選択します。
HGVS.cのバリアントから検索します。

Upload a **plain text** file of identifiers, one identifier per line, case sensitive

検索及び変換したい参照配列を選択します。

ATP7B:c.2903T>C

検索したいバリアントの情報を入力します。
例・・・遺伝子名：c.100A>T

バリアントの情報をどのデータベースの結果として
得るかを選択します。
Accession IDを入手するため、Refseqを選択します。

バリアントの位置を検索

- 「hg38」をリファレンスとしたATP7Bのc.2903T>CのAccession番号やゲノムポジションなどのバリアント情報が表示されます。

c.2903T>Cのバリアント情報にヒットしたAccession IDが表示されます。

coordinatesにgDNA、cDNA、proteinが表記されます。お持ちの情報と比較・ご確認ください。

Search:

| input | transcript | gene | strand | coordinates(gDNA/cDNA/protein) | region | in |
|-----------------|-------------------------------|-------|--------|---------------------------------------|-------------------------|----|
| ATP7B:c.2903T>C | XM_005266430 (protein_coding) | ATP7B | - | chr13:g.51946441A>G/c.2903T>C/p.I968T | inside_[cds_in_exon_14] | CS |
| ATP7B:c.2903T>C | XM_011535120 (protein_coding) | ATP7B | - | chr13:g.51942481A>G/c.2903T>C/p.V968A | inside_[cds_in_exon_12] | CS |
| ATP7B:c.2903T>C | XM_011535119 (protein_coding) | ATP7B | - | chr13:g.51946441A>G/c.2903T>C/p.I968T | inside_[cds_in_exon_13] | CS |
| ATP7B:c.2903T>C | NM_000053 (protein_coding) | ATP7B | - | chr13:g.51946441A>G/c.2903T>C/p.I968T | inside_[cds_in_exon_13] | CS |

Showing 1 to 4 of 4 entries

New Search

Accession ID : NM_000053の情報を用いて、hg19での位置に変換してみます。
リファレンス配列のみ「hg19」に変換して同様に検索してください。

バリアントの位置を検索

- 「hg38」でのATP7Bのc.2903T>Cのバリアントに対して「hg19」へ変換するとゲノムポジションが表のようになります。

TransVar Results

Search:

| input | transcript | gene | strand | coordinates(gDNA/cDNA/protein) | region | int |
|-----------------|-------------------------------|-------|--------|---------------------------------------|-------------------------|-----|
| ATP7B:c.2903T>C | XM_005266430 (protein_coding) | ATP7B | - | chr13:g.52520577A>G/c.2903T>C/p.I968T | inside_[cds_in_exon_14] | CS |
| ATP7B:c.2903T>C | NM_000053 (protein_coding) | ATP7B | - | chr13:g.52520577A>G/c.2903T>C/p.I968T | inside_[cds_in_exon_13] | CS |

Showing 1 to 2 of 2 entries

| Accession ID | gDNA (hg38) | gDNA (hg19) | Exon位置 | Protein |
|--------------|-------------------------|-------------------------|---------|---------|
| NM_000053.3 | g.51946441A>G | g.52520577A>G | Exon 13 | p.I968T |

表のように「hg38」と「hg19」などの参照配列が異なると、ゲノムポジションは異なり、正しいバリアントの位置を認識できません。

Mutalyzerでの確認

- Mutalyzerを用いて、TransVarで得られた結果の確認方法をご紹介します。

<https://mutalyzer.nl/>

LUMC Mutalyzer DNA tools ▾ Batch Jobs ▾ Web Services External links ▾ Help ▾ About

Welcome to the Mutalyzer website

The aim of this program suite is to support checks of [Sequence Variant Nomenclature](#) according to the guidelines of the [Human Genome Variation Society](#).

Name Checker

The Name Checker takes the complete sequence variant description as input and checks whether it is correct.

Examples: `AB026906.1:c.40_42del` , `NG_012337.1(SDHD_v001):c.274G>T` , `LRG_24t1:c.159dup`

Variant description using HGVS format

Syntax Checker

Takes the complete sequence variant description as input and checks whether the syntax is correct.

Position Converter

Converts chromosomal positions to transcript orientated positions and vice versa.

SNP Converter

Allows you to convert a dbSNP rsId to HGVS notation.

Name Generator

A user friendly interface that helps to make a valid HGVS variant description.

Mutalyzerでの確認

- Build（緑枠）の中から、解析したいリファレンス配列を選択します。
Variant description（赤枠）に、Accession IDとバリアントの情報を入力します。
* NM_00001.○ の○部分の数字がVersionを表しており、稀に対応していないことがありますので、その場合はページ下部に「We found these versions:NM_~」と表示されますので、versionを変えてお試しください。

Position Converter

Please supply the genome assembly which you want to use to convert your position.

Note: The Position Converter does NOT check the description or normalize it to HGVS. Use the [Name Checker](#) for this.

Build

Homo sapiens – GRCh38 (hg38)

Canis lupus familiaris – Broad CanFam3.1 (canFam3)

Homo sapiens – NCBI36 (hg18)

Homo sapiens – GRCh37 (hg19)

Homo sapiens – GRCh38 (hg38)

Mus musculus – GRCm38 (mm10)

Saccharomyces cerevisiae – SacCer_Apr2011 (sacCer3)

Variant description

Variant description using HGVS format

Examples: [NM_003002.3:c.274G>T](#), [LRG_11t1:c.274G>T](#), [chr11:g.111959693G>T](#), [NC_000011.9:g.111959693G>T](#)

Convert variant description

Help

Mutalyzerでの確認

- リファレンス配列にそれぞれ「hg38」と「hg19」を選択し、AccessionIDはNM_000053.3を選択し、c.2903T>Cについて検索した結果、TransVarで得られた結果と同じであることを確認できます。

Build

Homo sapiens – GRCh38 (hg38)

Variant description

NM_000053.3:c.2903T>C

Examples: NM_003002.3:c.274G>T, LRG_11t1:c.274G>T, chr11:

Convert variant description

Chromosomal variant

NC_000013.11:g.51946441A>G

Build

Homo sapiens – GRCh37 (hg19)

Variant description

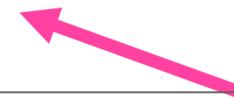
NM_000053.3:c.2903T>C

Examples: NM_003002.3:c.274G>T, LRG_11t1:c.274G>T, chr1

Convert variant description

Chromosomal variant

NC_000013.10:g.52520577A>G



| Accession ID | gDNA (hg38) | gDNA (hg19) | Exon位置 | Protein |
|--------------|---------------|---------------|---------|---------|
| NM_000053.3 | g.51946441A>G | g.52520577A>G | Exon 13 | p.I968T |