

# ヒト選択的スプライシングデータベース H-DBASの活用

産業技術総合研究所バイオメディシナル情報研究センター  
武田淳一

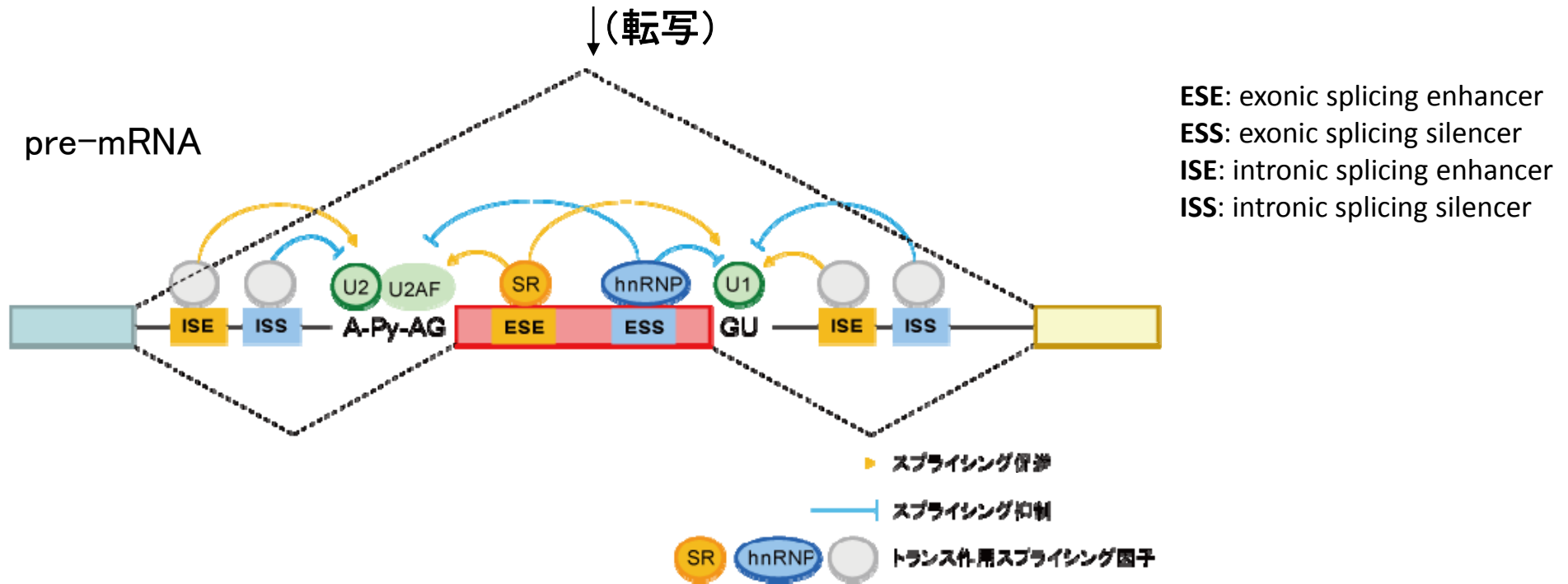
# 目次

- 選択的スプライシングについて
- H-DBASについて
- H-DBASの使い方

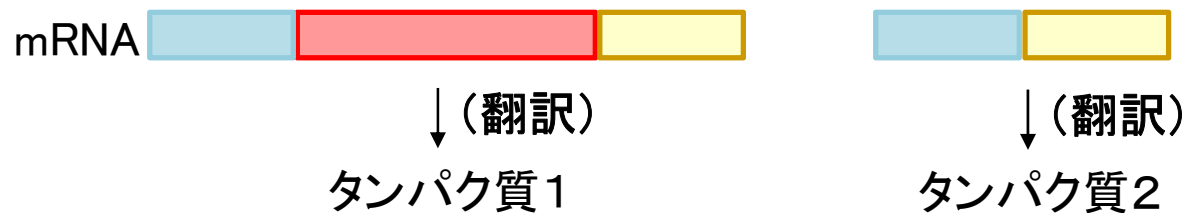
# 目次

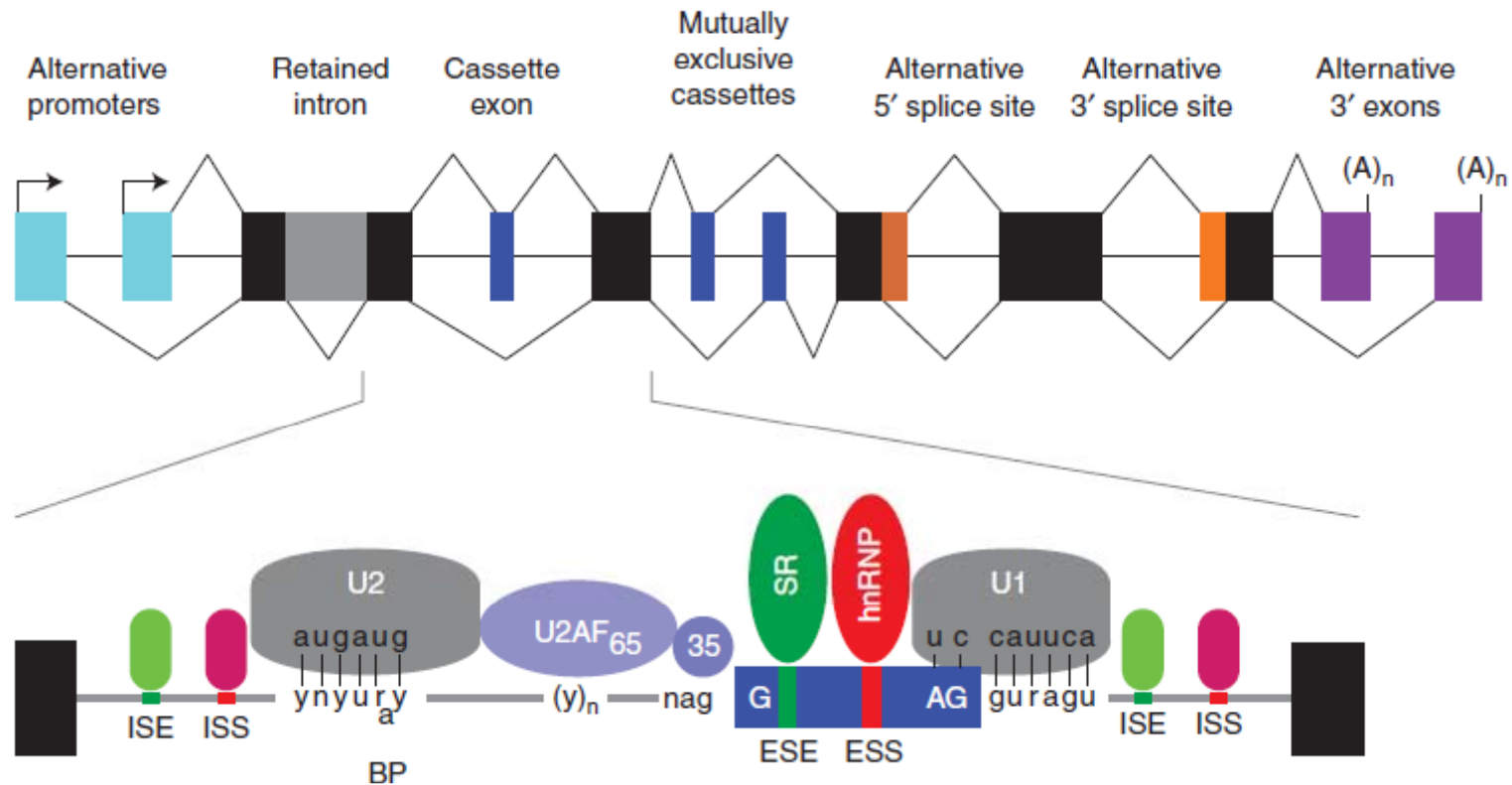
- 選択的スプライシングについて
- H-DBASについて
- H-DBASの使い方

# 選択的スプライシング (Alternative Splicing; AS)とは



↙ ↘ (選択的スプライシング)





**Figure 1.** RNA sequences and factors involved in the regulation of alternative splicing. The upper diagram shows various patterns of alternative splicing using a 12-exon (black and colored boxes) gene that contains alternative promoters (arrows adjacent to turquoise boxes), constitutive exons (black), a retained intron (gray), a cassette and mutual exclusion exons (blue), alternative 5' and 3' splice sites (orange), and alternative 3' exons with polyadenylation sites ( $A_n$ ). The lower diagram highlights the conserved RNA sequence elements at the branch point (BP, U2 binding site) and the 5' (U1 binding site) and 3' (U2AF heterodimer binding site) splice sites of an alternatively spliced cassette exon (blue box). Also shown are the less conserved exonic and intronic splicing enhancer (ESE, ISE) and silencer (ESS, ISS) elements that are recognized by multiple factors including the SR and hnRNP proteins.

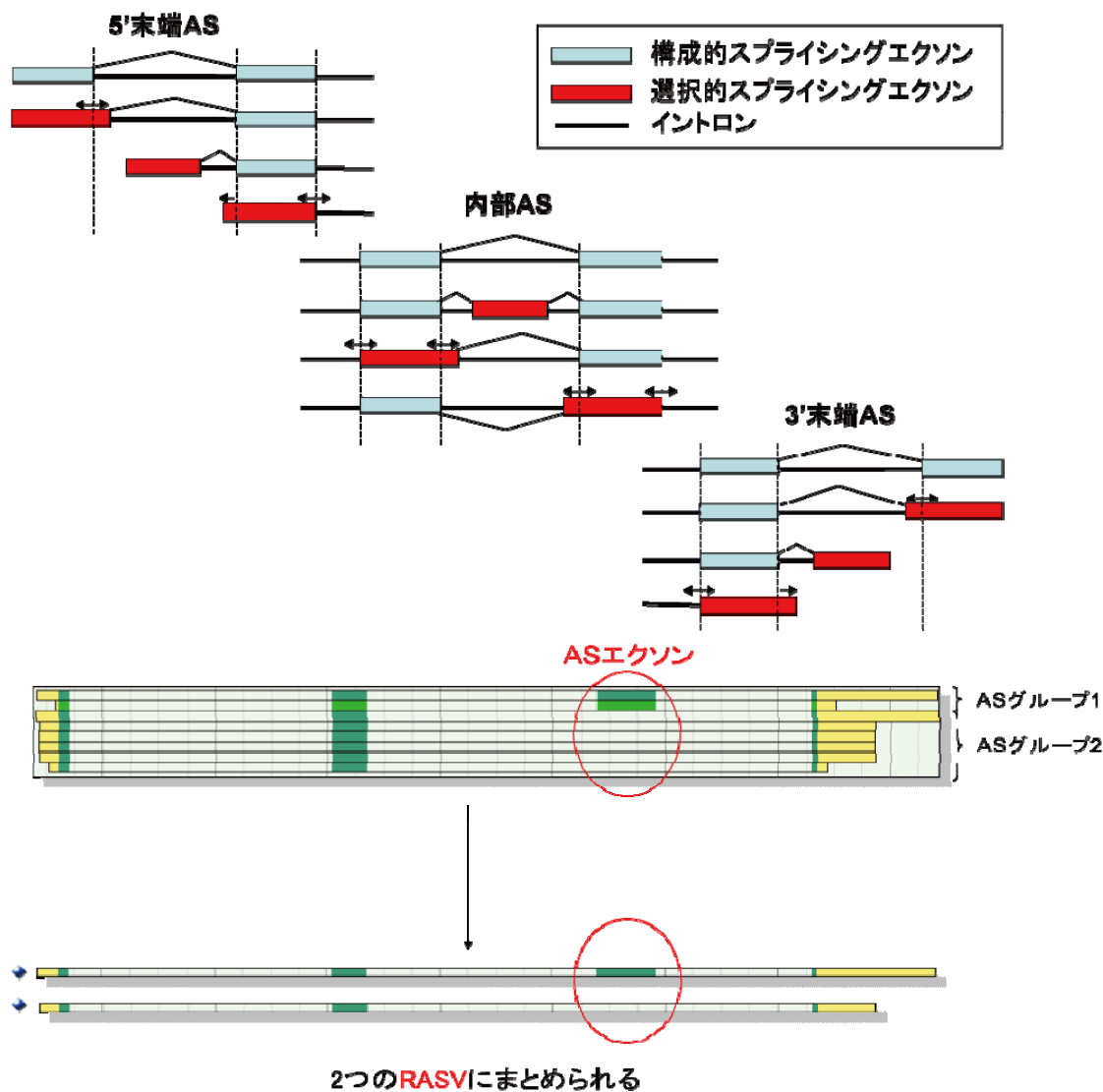
# 目次

- 選択的スプライシングについて
- H-DBASについて
- H-DBASの使い方

# H-DBASの特徴

- 6つの哺乳類モデル生物種(ヒト・マウス・ラット・チンパンジー・アカゲザル・イヌ)の転写物から代表ASバリエント(Representative AS variant, **RASV**)を同定して公開。ヒトとマウスについては、完全長cDNAのみのデータセットもある
- ヒトと他生物種のRASVについて、ゲノムを介した比較によって種間保存の同ースプライシングバリエント(equally-spliced variant, **ESV**)を同定して公開。**比較ゲノム解析**
- タンパク機能(タンパク機能モチーフ・遺伝子オントロジー・細胞内局在化シグナル・膜タンパクドメイン)に影響を与えるRASVの観察が可能(ヒトのみ)。**タンパク機能アノテーション解析**
- 細胞画分(細胞質・核・ポリソーム)ごとに検出されたRNA-Seqタグを用い、核由来特異的なASジャンクションを持つバリエントを公開(ヒトのみ)。**RNA-Seq解析**

# ASの判定と、代表ASバリエント(RASV)の同定



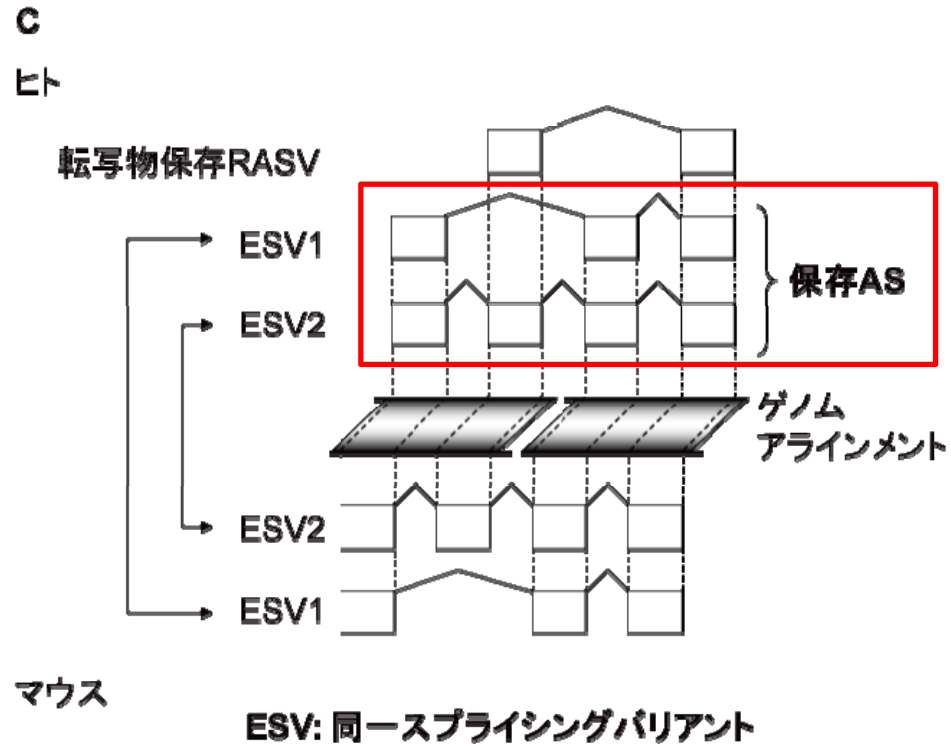
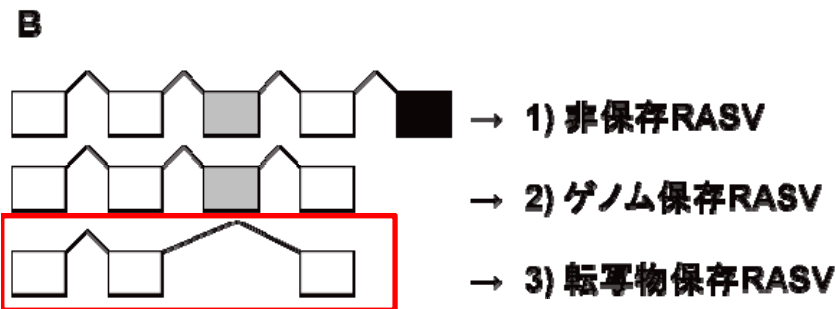
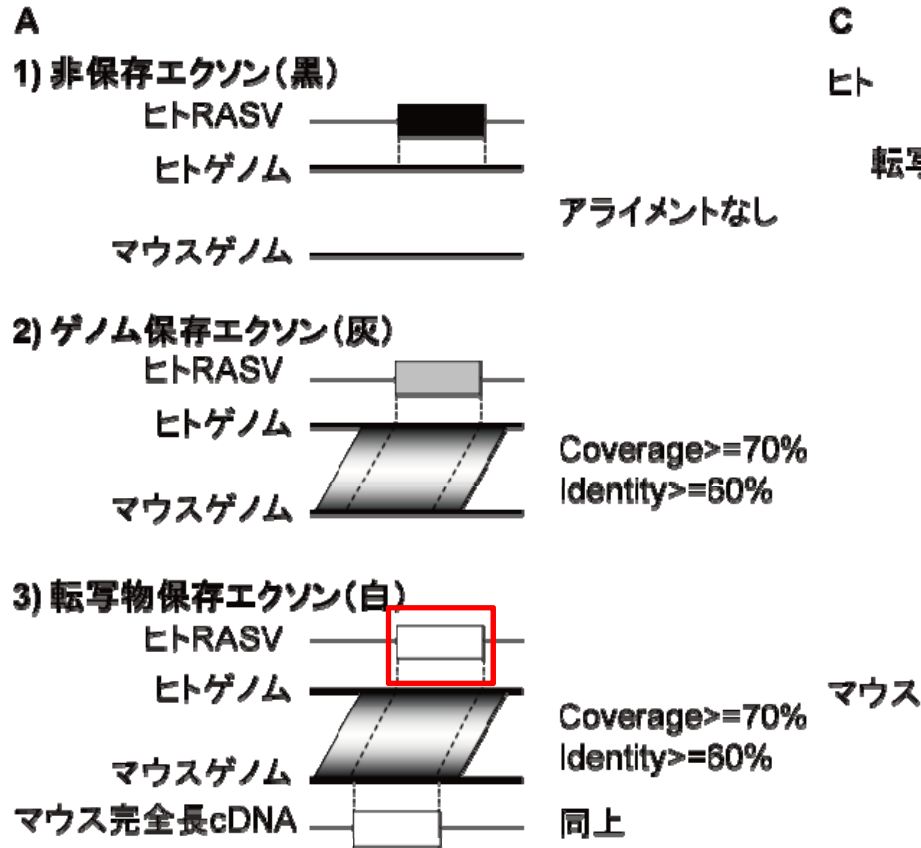
1. 転写物のエクソン構造を総当りで比較し、ASかどうかを判定

2. 同じエクソン構造のASバリエントをグループ化し、その中から代表ASバリエント(RASV)を選択

同定されたRASVのうち、タンパク機能アノテーション解析に用いられるのは、CDS(開始コドンで始まり、終止コドンで終わる80aa以上のORF)を含むもの



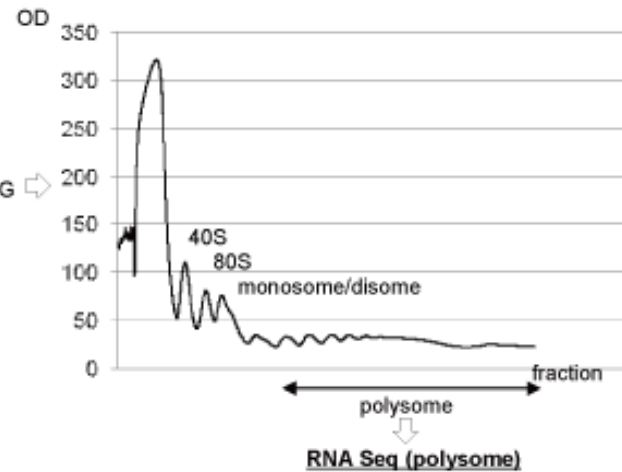
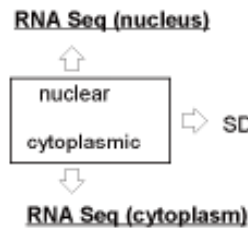
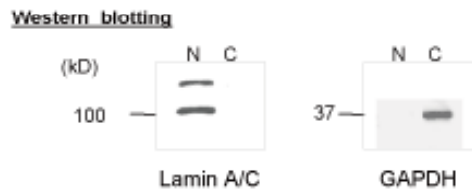
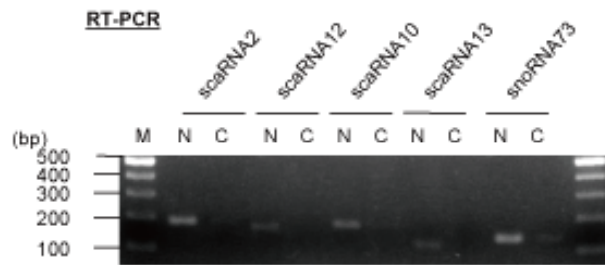
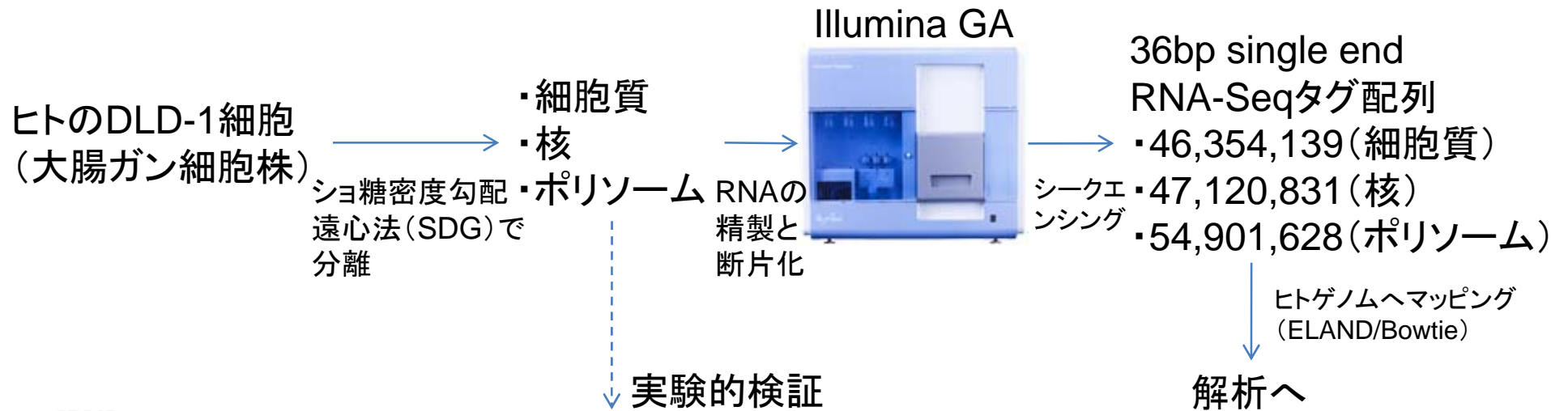
# 同ースプライシングバリエント(ESV)の同定



A: エクソン単位で判定  
B: Aを元にバリエント単位でも判定  
C: Bを元に種間保存ASを同定

□ ABCの各カテゴリーで最も高い保存度

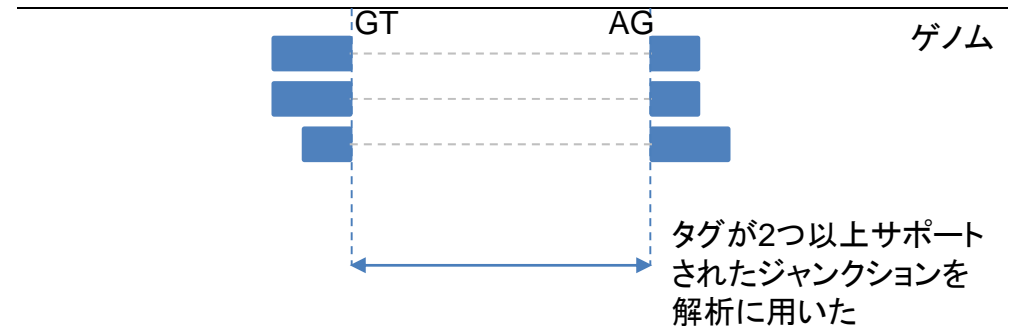
# 細胞画分ごとのRNA-Seqタグ生成



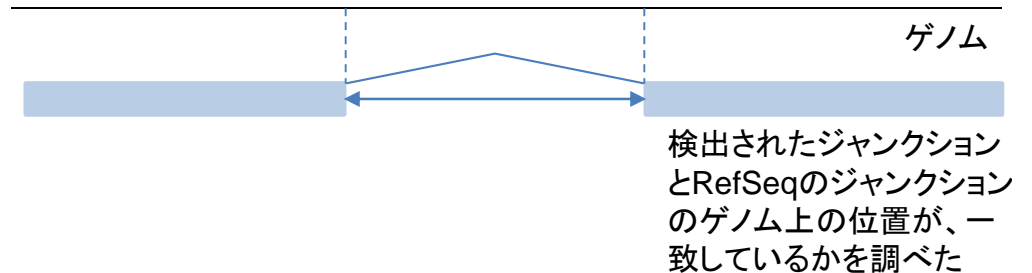
sno/scaRNAのRT-PCRとラミンA/Cのウェスタンブロッティングにより核(N)であること、GAPDHのウェスタンブロッティングにより細胞質(C)であること、SDG後の吸光度(OD)からポリソームであることを確認

# 翻訳しない(核のみに存在する) ASバリエーションの同定

①RNA-Seqタグ配列をゲノムにマッピングしてジャンクションを検出(TopHatを使用)



②既存の遺伝子としてRefSeq転写物を用いそのジャンクションとゲノム上の位置を比較



①のRNA-Seqタグ配列が核のみで検出、かつ②のジャンクションがASジャンクションの場合

翻訳しないと考えられるASバリエーション(翻訳は核外のポリソーム画分で行われるため)

# 目次

- 選択的スプライシングについて
- H-DBASについて
- H-DBASの使い方

# ASビューワの利用

**H-DBAS** Human-transcriptome DataBase for Alternative Splicing

Human (Full-length cDNA) [ヘルプ] [English]

Go Clear

トップ ASの機構 データと解析方法 統計 ダウンロード 用語集 リンク

## H-DBAS - ヒトの選択的スプライシングデータベース

### お知らせ

- 2010/09/09: H-DBASバージョン6を公開しました

クリックで直接ASローカス構造図へ飛びます

### 検索と解析ページ

- 右上の簡易検索例(クリックで直接飛びます): [HIX0004914](#), PI3K, PI3-kinase p85 subunit alpha
- 検索ページ: [詳細検索](#), [BLAST検索](#)
- 解析ページ: [比較ゲノム解析](#), [RNA-Seq解析](#)

### H-DBASについて

- H-DBASは、[H-InvDB](#)の配列データに基づく選択的スプライシング (alternative splicing, AS) のデータベースです。H-DBASの特徴は以下の通りです
  - ヒト・マウス・ラット・チンパンジー・アカゲザル・イヌの6生物種から成る8つのデータセットから、[代表ASバリエント \(Representative AS variant, RASV\)](#)を同定して公開しています。データセットの内容と生物種との対応は以下の通りです

# ASローカス構造図

メイン画面の  
表示形式操作

タンパク機能アノテーション解析



Human (Full-length cDNA) [ヘルプ] [English]  
Go Clear

トップ ASの機構 データと解析方法 統計 ダウンロード 用語集 リンク

## ASローカス構造図

ダウンロード ワイド表示

HIX0004914: Phosphatidylinositol 3-kinase regulatory subunit alpha (PI3-kinase p85 subunit alpha) (PtdIns-3-kinase p85-alpha) (PI3K).  
染色体: 5 位置: 67511604..67597648 (86.0Kbp) スtrand: + RASVの数: 6

**アノテーション コントローラー**  
PUSH  
タンパク機能アノテーションなどを  
ユーザーが任意に表示可能

ORF: ■ 完全長 ■ 不完全 ■ NMD ■ UTR  
スプライスサイト: ■ GT-AG ■ GC-AG ■ AT-AC ■ その他 ■ NAGNAG  
AS: □ ASエクソン ▲ ASジャンクション 多型: ■ SNP

種間比較  
■ エクソンの保存度 (エクソン下部に表示)  
■ 転写物保存 ■ ゲノム保存  
■ 非保存

スケール:  
● エクソン中心  
○ 全体  
x1 x3 x10 塩基 アミノ酸

## Human RASV

RASV: (6)  Other ASV: (0)  RefSeq: (3)  Ensembl: (4)



比較ゲノム解析

\*デフォルトは構成的イントロンを小さくして表示  
(スケールのエクソン中心)しています

メイン画面

# アノテーションコントローラーと種間比較を使用

The screenshot displays the 'Annotation Controller' (アノテーションコントローラー) interface. The main window title is 'アノテーションコントローラー'. The left sidebar shows a tree view with 'タンパク機能アノテーション' selected, and a sub-item 'タンパク機能に影響を与えるAS' circled in red. The main content area shows a list of AS (Alternative Splicing) events for 'Human RASV' on chromosome 5 (chr5:67511604-67595055). The list includes entries like 'HIT000002596 (AK000121)', 'HIT000019640 (AK094785)', 'HIT000041152 (BC030815) ESV 1', 'HIT000046218 (AK126345)', 'HIT000335096 (BC094795) ESV 2', and 'HIT000489543 (AK294919)'. Below this, a genomic track shows the alignment of these AS events across the human and mouse genomes. The mouse genome section shows 'chr13:102538168-102454037 (-)' with entries 'AK171366', 'BC026146 ESV 2', and 'BC051106 ESV 1'. A red arrow points from the 'タンパク機能に影響を与えるAS' header to the 'ESV クラスター: 2 (保存ASがあります)' label. Another red arrow points from the 'タンパク機能アノテーション' header to the 'タンパク機能アノテーション解析の結果' text. A third red arrow points from the 'タンパク機能アノテーション' header to the 'タンパク機能アノテーション解析の結果' text. The interface also includes a search bar, a 'ダウンロード' button, and a 'ワイド表示' button. The bottom of the interface shows a legend for 'タンパク機能アノテーション' and 'タンパク機能アノテーション解析の結果'.

タンパク機能アノテーション解析の結果

比較ゲノム解析の結果

タンパク機能アノテーション解析の結果

# 検索の利用

\*簡易検索は常に右上に表示されます

The screenshot shows the top of the H-DBAS website. On the left is the logo 'H-DBAS Human-transcriptome DataBase for Alternative Splicing'. On the right is a search bar with a dropdown menu set to 'Human (Full-length cDNA)', a search input field, and 'Go' and 'Clear' buttons. Below the search bar is a navigation menu with links: 'トップ', 'ASの機構', 'データと解析方法', '統計', 'ダウンロード', '用語集', and 'リンク'.

## H-DBAS - ヒトの選択的スプライシングデータベース

### お知らせ

- 2010/09/09: H-DBASバージョン6を公開しました

クリックでそれぞれの検索ページへ  
飛びます (BLAST検索は別ウィンドウ)

### 検索と解析ページ

- 右上の簡易検索例(クリックで直接飛びます): [HIX0004914](#), PI3K, PI3-kinase p85 subunit alpha
- 検索ページ: [詳細検索](#) [BLAST検索](#)
- 解析ページ: [比較ゲノム解析](#), [RNA-Seq解析](#)

### H-DBASについて

- H-DBASは、[H-InvDB](#)の配列データに基づく選択的スプライシング (alternative splicing, AS) のデータベースです。H-DBASの特徴は以下の通りです
  1. ヒト・マウス・ラット・チンパンジー・アカゲザル・イヌの6生物種から成る8つのデータセットから、[代表AS/バリエーション \(Representative AS variant, RASV\)](#)を同定して公開しています。データセットの内容と生物種との対応は以下の通りです



# 検索ページ

## 詳細検索: 計22の検索項目から絞り込み

### 詳細検索

#### 基本情報

- データセット: Human (Full-length cDNA)
- ローカス転写物情報: Keyword

#### ゲノム構造

- 染色体番号: All
- ストランド: + - 両方
- ゲノム位置: from to bp
- スプライスサイト配列: All
- レトロトランスポゾン: LTR LINE SINE SINE (Aluのみ)

#### ASゲノム構造

- RASVの数: from to
- ASロケーション: 5末端 内部 3末端
- ASパターン:
  - カセット型エクソン (複数カセット型を含む)
  - 選択的3'スプライス
  - 選択的5'スプライス
  - 相互排他的エクソン
  - 選択的保持イントロン
  - 選択的第一エクソン 5'-end
  - 選択的最終エクソン 3'-end

#### タンパク機能アノテーション

- ORF長: from to aa
- タンパク機能モチーフ: InterPro ID
- 遺伝子オントロジー (GO): GO ID
- 細胞内局在化シグナル:
  - WoLF PSORT All
  - TargetP All
- 膜タンパクドメイン:  TMHMM  SOSUI
- 異常なORF:  NMD (PTC)

#### ASタンパク機能アノテーション

- RASVのアミノ酸長の差: from to aa
- ASリジョン:  5'UTR  CDS  3'UTR
- タンパク機能に影響を与えるAS:
  - タンパク機能モチーフ  GO  細胞内局在化シグナル  膜タンパクドメイン
- 複雑なASパターン:  ブリッジ  ネスト  二重CDS

#### AS比較ゲノム解析

- 比較生物種: Mouse (Full-length cDNA)
- RASVの保存度:
  - 非保存  ゲノム保存  転写物保存
  - 同スプライシング/リリアント (ESV)  保存AS

どの組み合わせも可能。検索条件はANDのみ

Search Reset

## BLAST検索: ヒトRASVに対する配列相同性

[\[English\]](#)

プログラムと検索するデータベースを選んで下さい:

- プログラム: blastn
- データベース:
  - 塩基配列データベース
    - ヒト(完全長cDNA)
    - ヒト(mRNA)
  - アミノ酸配列データベース
    - ヒト(完全長cDNA)
    - ヒト(mRNA)

FASTA形式での配列の入力

ディスクからの読みこみ

参照...

入力配列のクリア 検索実行

デフォルトでは **フィルター** がONになっており、相同性を判断するのにあまり意味がない配列は無効化されます。

- フィルター**  低複雑配列のマスキング  "lookup table" 作成時のみ

- 期待値: 10
- アミノ酸置換テーブル: BLOSUM62
- ギャップ無しの計算

- コドン表: Standard (1)
- フレームシフトペナルティ: No OOF

その他のオプション:

- アラインメント結果のグラフィカル表示
- アラインメント表示方法: Pairwise
- 簡易結果の表示件数: 100
- アラインメントの表示件数: 50
- 色彩のスキーマ: No color schema

入力配列のクリア 検索実行

# RNA-Seq解析の利用



H-DBAS Human-transcriptome DataBase for Alternative Splicing

Human (Full-length cDNA) [ヘルプ] [English]

Go Clear

トップ ASの機構 データと解析方法 統計 ダウンロード 用語集 リンク

## H-DBAS - ヒトの選択的スプライシングデータベース

### お知らせ

- 2010/09/09: H-DBASバージョン6を公開しました

### 検索と解析ページ

- 右上の簡易検索例(クリックで直接飛びます): [HIX0004914](#), [PI3K](#), [PI3-kinase p85 subunit alpha](#)
- 検索ページ: [詳細検索](#), [BLAST検索](#)
- 解析ページ: [比較ゲノム解析](#), [RNA-Seq解析](#)

クリックでRNA-Seq解析ページへ飛びます(別ウィンドウ)

### H-DBASについて

- H-DBASは、[H-InvDB](#)の配列データに基づく選択的スプライシング (alternative splicing, AS) のデータベースです。H-DBASの特徴は以下の通りです
  1. ヒト・マウス・ラット・チンパンジー・アカゲザル・イヌの6生物種から成る8つのデータセットから、[代表ASバリエント \(Representative AS variant, RASV\)](#)を同定して公開しています。データセットの内容と生物種との対応は以下の通りです

# RNA-Seq解析ページ

Search item1: ID or name

Search item2: AS pattern of RefSeq junction

Search item3: Expression of RefSeq junction  Cytoplasm  Nuclear  Polysome

## RNA-Seq analysis

Single-end 36bp tags (RNA-Seq tags) extracted from cytoplasm, nuclear and polysome in human DLD-1 cell were used. By detecting junctions from them by [TopHat](#), we identified translationable and non-translationable AS junctions in RefSeq transcripts (release23) mapped on human genome (hg19).

Expression confirmation

[Separation of sub-cellular fractions](#)

Download

- RNA-Seq tag sequence data: [Cytoplasm](#), [Nuclear](#), [Polysome](#)
- RefSeq RASV sequence data: [RASV human refseq.fa.gz](#)
- RefSeq RASV position data: [refFlat RASV human refseq.txt.gz](#)

RNA-Seq statistics

	RNA-Seq tags	Mapped tags on genome*	Detected junctions**
Cytoplasm	46354137	28634895	548974
Nuclear	47120831	28808545	582576
Polysome	54901628	29485229	647181

# 核のみに存在するASジャンクションの例

Search item1: ID or name Protein name

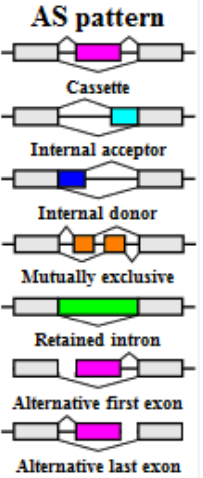
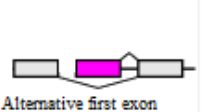
Search item2: AS pattern of RefSeq junction All

Search item3: Expression of RefSeq junction  Cytoplasm  Nuclear  Polysome

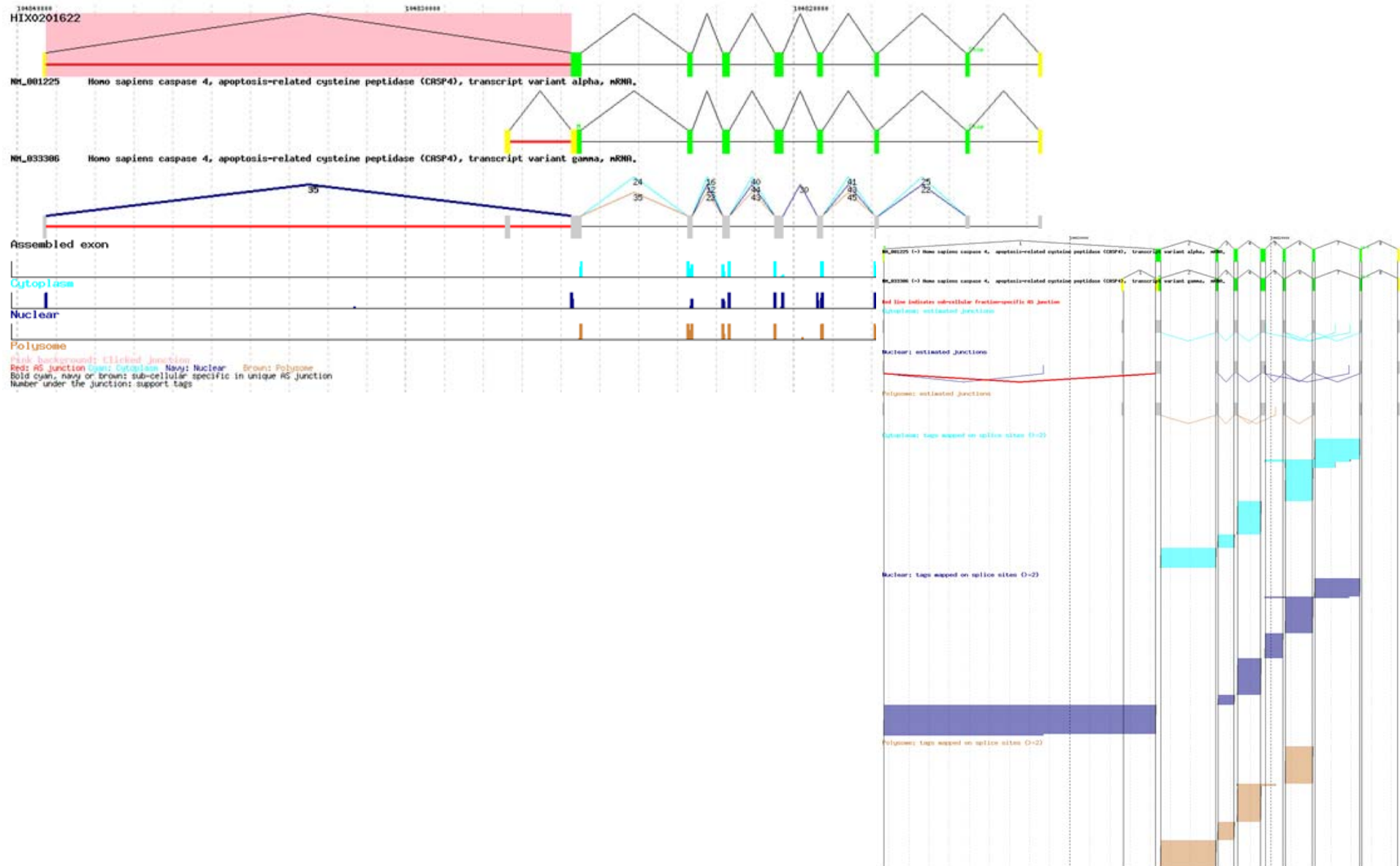
## Search result

protein = CASP4; as pattern = all; subcell fraction = cyt nuc pol. [return to the default page](#)

hit no: 1

Chr./Band	Locus (Strand)	RefSeq	HUGO	Protein	AS junction NO/ genome start/ genome end	AS pattern 	Expression in sub-cellular fractions: Cytoplasm Nuclear Polysome (Support tags [ $\geq 2$ ]) * indicates specific
chr11 -	<a href="#">HIX0201622 (-)</a> <a href="#">RNA Seq viewer</a> <a href="#">RNA Seq overview</a> <a href="#">GBrowse2</a>	<a href="#">NM_001225</a>	<a href="#">CASP4</a>	Homo sapiens caspase 4, apoptosis-related cysteine peptidase (CASP4), transcript variant alpha, mRNA.	1 104825728 104839246	 Alternative first exon	Cytoplasm <b>Nuclear* (35)</b> Polysome

# RNA-Seq viewer (左) と overview (下)



# GBrowse2 (RNA-Seqタグの配列も表示)

**領域**

← 104800k 104900k

HIX (H-Inv loci)

HIX0035901 HIX0010072 HIX0201622 HIX0201719 HIX0035831 HIX0035832

**詳細ビュー**

104840k 104830k 104820k

HIX (H-Inv loci)

HIX0201622

**RefSeq transcripts (Release 23)**

NM\_001225  
Caspase-4 precursor (EC 3.4.22.57) (CASP-4) (ICH-2 protease) (TX protease) (ICE(rel)-II) [Contains: Caspase-4 subunit 1; Ca

NM\_033306  
Caspase 4 isoform gamma precursor.

**RefSeq AS junctions**

**Junctions in Cytoplasm (support tags >=2)**

**Junctions in Nucleoplasm (support tags >=2)**

**Junctions in Polysome (support tags >=2)**

**chr11**

**893057-2 Details**

Name: 893057-2  
 Class: match  
 Type: match  
 Description: sam/bam  
 Source: sam/bam  
 Position: chr11:104839245..104839276 (- strand)  
 Length: 32  
 Query: 893057-2  
 Matches: 893057-2:1..32 (-)

```

chr11 104839276 CCTGCCATAGGGAACAGCCTCTGTCCCTTTTTT 104839245
|-----|
893057-2 1 TCTGCCATAGGGAACAGCCTCTGTCCCTTTTTT 32
  
```

Score: 0  
 MD: 32  
 NM: 0  
 X0: 2  
 primary\_id: 893057-2;0;104839245;104839276;-1  
 gbrowse\_dbid: maseq\_nuc\_chr11:database

>893057-2 class=match position=chr11:104839245..104839276 (- strand)  
 AAAAAAGGAC AGAGGCTGTT CCCTATGGCA GG