

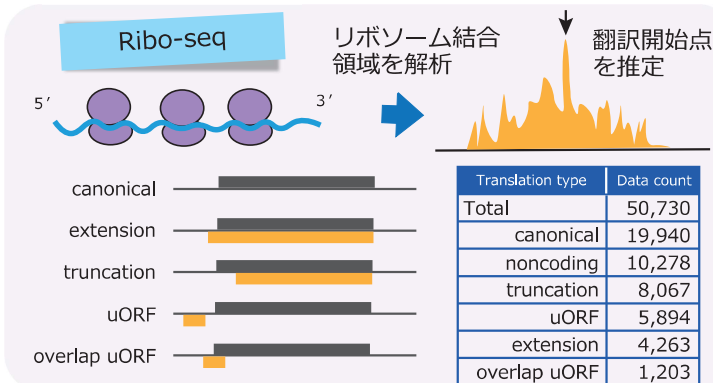
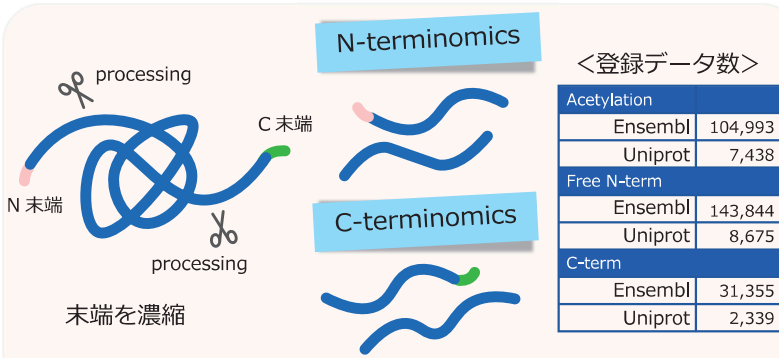
○ 松本雅記¹、高見知代¹、幡野敦¹、奥田悠世²、小林大樹¹、高橋悠志¹、吉沢明康¹、守屋勇樹³、田畑剛⁴、岩崎未央⁴、杉山直幸⁵、田中聡⁶、荒木令江⁷、五斗進³、河野信²、奥田修二郎¹、小寺義男²、石濱泰⁴

1) 新潟大学、2) 北里大学、3) DBCLS、4) 京都大学、5) 国立循環器病研究センター、6) Trans-IT、7) 熊本大学

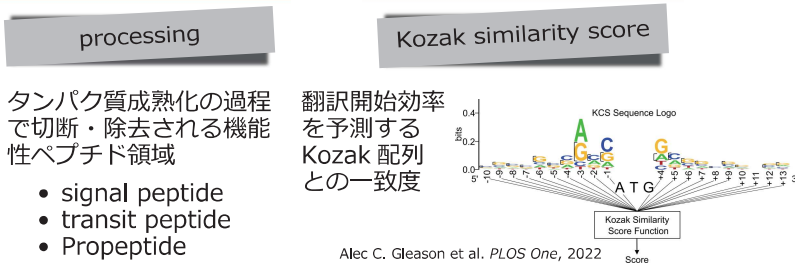
本研究の目的

タンパク質の N 末端・C 末端は機能制御に重要だが、非 AUG 開始、uORF、翻訳後プロセッシング由来の末端多様性は既存 DB で十分に扱われていない。本研究では、末端特化型データベース「EdgeProt」を開発し、プロテオーム、トランスクリプトーム、Ribo-seq 等の情報を統合した。これにより非典型翻訳産物の同定、成熟 N 末端予測、新規 ORF の機能解析を支援し、末端多様性研究の基盤を提供する。

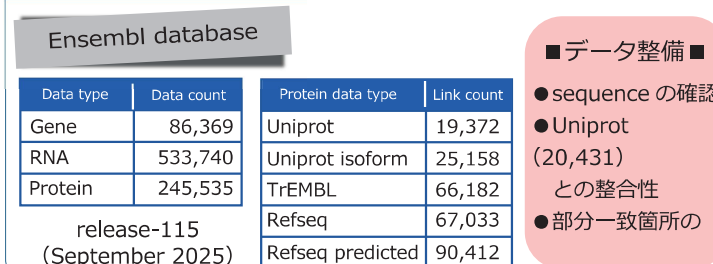
実測データ



付加情報



Public database (Human)



EdgeProt : <https://edgeprot.jpostdb.org/>

<Transcript list view>



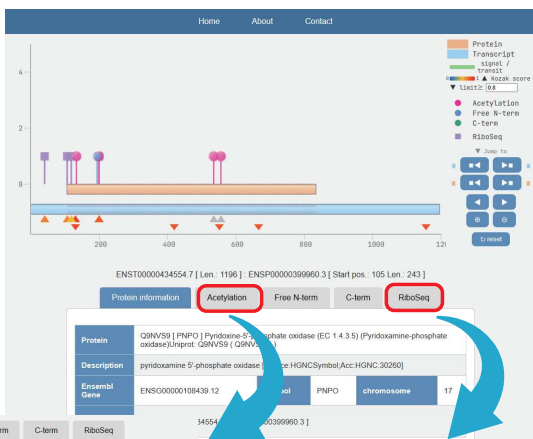
Ensembl Gene の Transcript Id を一覧表示。Isoform 間のデータ分布が参照可能。TranscriptId をクリックすると Protein view を開く。

<Acetylation list>

開始コドン前後 10 塩基を記載。翻訳開始点が ATG もしくは near-cognate start の場合のみ Kozak similarity score を計算。

RNA pos.	Protein pos.	Modified sequence	Codon sequence	Kozak score
133 - 201	10 - 32	[Acetyl]ATFGRAEPYGLYLHC[Carbamidomethyl]GRSAAM	GCAGGCGGCTC ACG [GGG]ACGTTCCG	0.89*
199 - 219	32 - 38	[Acetyl]MDLGPMPR	CAGTGTGGCC [ATG] GACCTGGGAC	0.79*
199 - 243	32 - 46	[Acetyl]MDLGPMPK[Dimethyl]SYRGDRE	CAGTGTGGCC [ATG] GACCTGGGAC	0.79*
199 - 255	32 - 50	[Acetyl]MDLGPMPK[Dimethyl]SYRGDREAFEE	CAGTGTGGCC [ATG] GACCTGGGAC	0.79*
535 - 584	144 - 163	[Acetyl]PK[Dimethyl]SSQIGAVSHGSSVPPDR	CCAGTCGGCC [GGC] AAGAGAGCC	

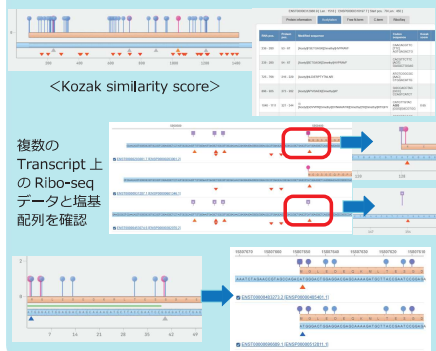
<Protein view>



<Ribo-seq list>

RNA pos.	Data type	Codon sequence	Kozak score
40	extension	ATT	0.69
106	canonical	ATG	0.65
118	truncation	CTG	0.50
199	truncation	ATG	0.79

複数のデータを重ねることでデータの信頼性を確認



terminocys で同定されていない Ribo-seq データ

Ribo-seq データから fasta を作成することで同定される可能性がある

今後の展望

jPOST 再解析プロテオームデータやパートナー研究室との連携により、品質を担保したマルチオミクスデータを継続的に蓄積するとともに、マウスをはじめとする他生物種へのデータ拡張を進める。さらに、翻訳開始点および Processing 部位について、プロテオームデータや Ribo-seq などの複数のエビデンスに基づくスコアリング機能を実装し、データベースの信頼性・網羅性・有用性を一層向上させることを目指す。