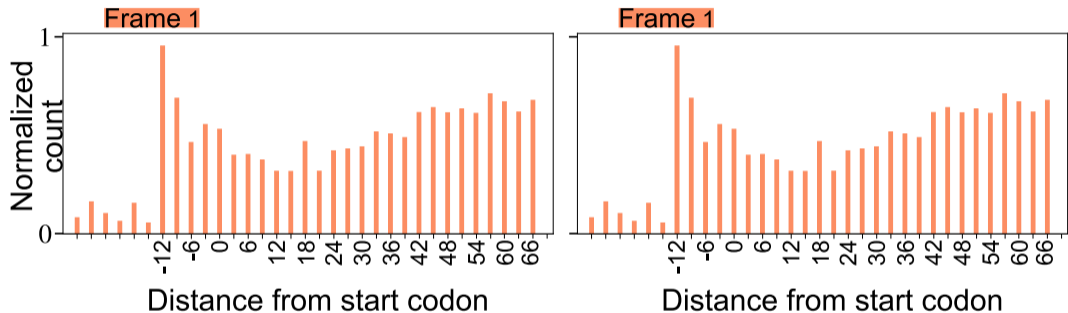


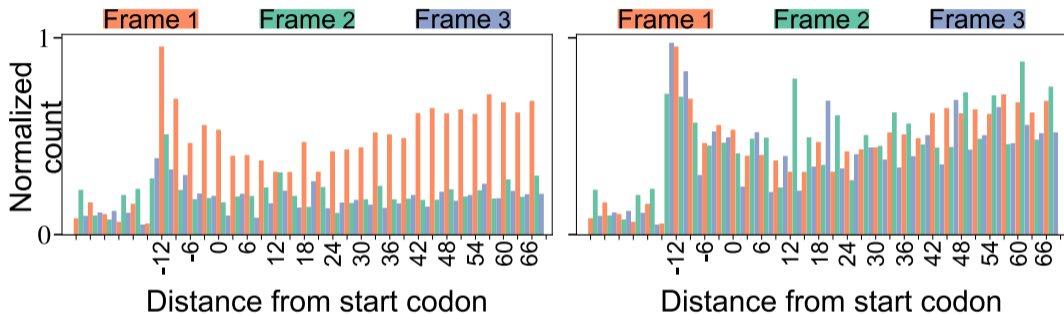
ribotracer: Identifying short and long ORFs under active translation

Saket Choudhary

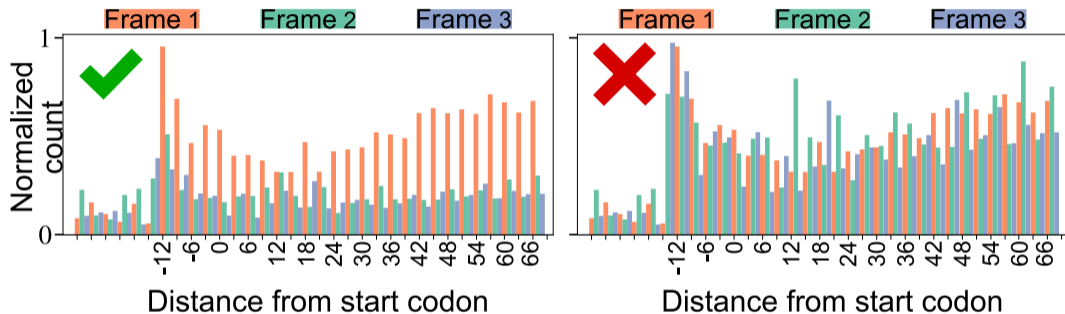
Actively translating fragments show 3-nt periodicity



Actively translating fragments show 3-nt periodicity



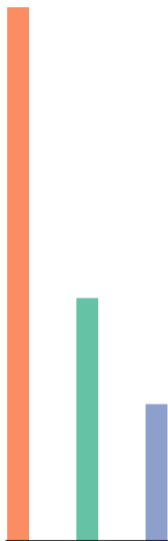
Actively translating fragments show 3-nt periodicity



Problem definition: Differentiate an actively translating profile (left) from non-active (right) in Ribo-seq data

Ribotricker simplifies detecting active ORFs

- Assesses the periodicity of RPF profile by projecting the 3D read count vector of each codon to a 2D unit vector
- Uses qualitative information of “high-low-low” pattern



Ribotracer simplifies detecting active ORFs

- Assesses the periodicity of RPF profile by projecting the 3D read count vector of each codon to a 2D unit vector
- Uses qualitative information of “high-low-low” pattern



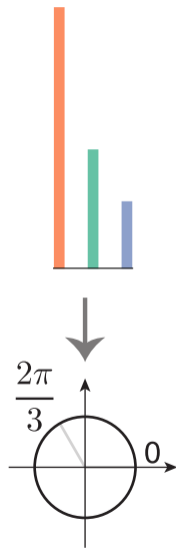
Ribotracer simplifies detecting active ORFs

- Assesses the periodicity of RPF profile by projecting the 3D read count vector of each codon to a 2D unit vector
- Uses qualitative information of “high-low-low” pattern



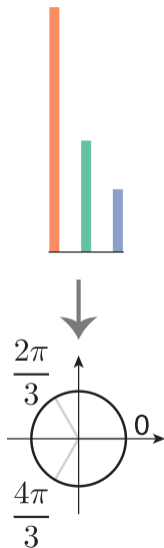
Ribotricker simplifies detecting active ORFs

- Assesses the periodicity of RPF profile by projecting the 3D read count vector of each codon to a 2D unit vector
- Uses qualitative information of “high-low-low” pattern



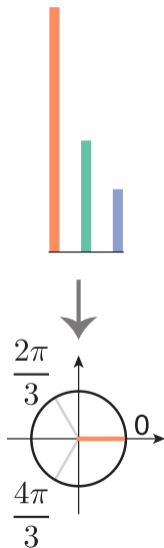
Ribotricker simplifies detecting active ORFs

- Assesses the periodicity of RPF profile by projecting the 3D read count vector of each codon to a 2D unit vector
- Uses qualitative information of “high-low-low” pattern



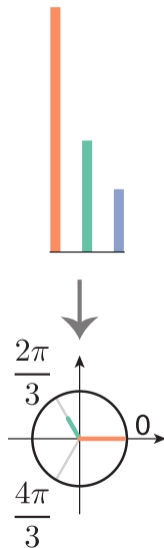
Ribotracer simplifies detecting active ORFs

- Assesses the periodicity of RPF profile by projecting the 3D read count vector of each codon to a 2D unit vector
- Uses qualitative information of “high-low-low” pattern



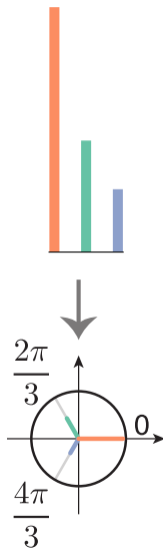
Ribotricker simplifies detecting active ORFs

- Assesses the periodicity of RPF profile by projecting the 3D read count vector of each codon to a 2D unit vector
- Uses qualitative information of “high-low-low” pattern



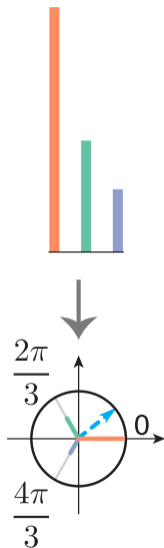
Ribotricker simplifies detecting active ORFs

- Assesses the periodicity of RPF profile by projecting the 3D read count vector of each codon to a 2D unit vector
- Uses qualitative information of “high-low-low” pattern

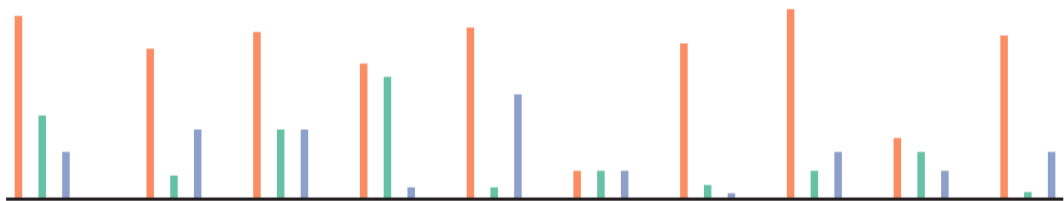


Ribotricker simplifies detecting active ORFs

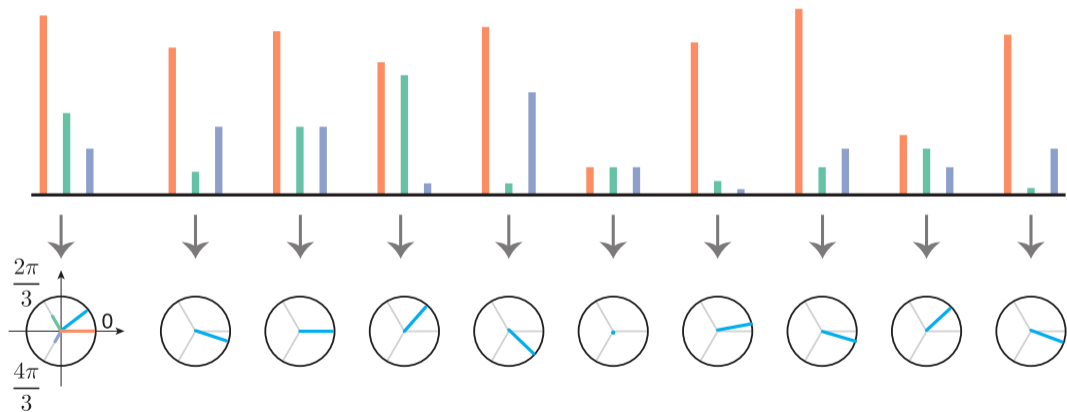
- Assesses the periodicity of RPF profile by projecting the 3D read count vector of each codon to a 2D unit vector
- Uses qualitative information of “high-low-low” pattern



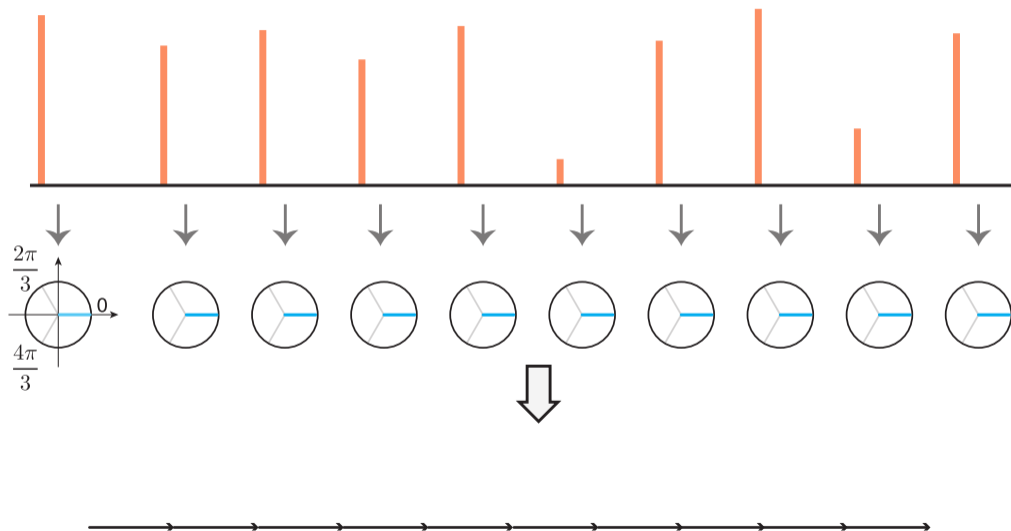
Ribotricker's phase score calculation



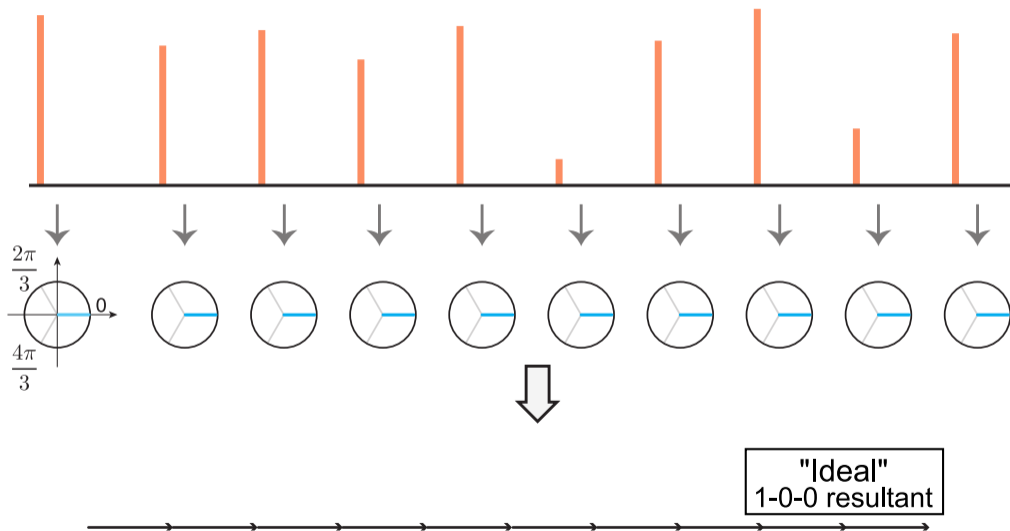
Ribotricer's phase score calculation



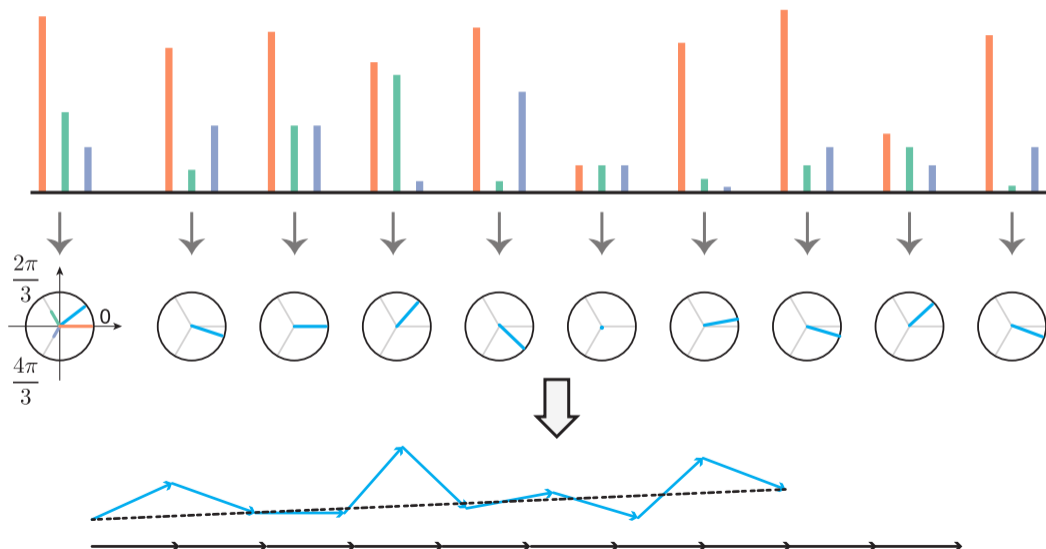
Ribotricer's phase score calculation



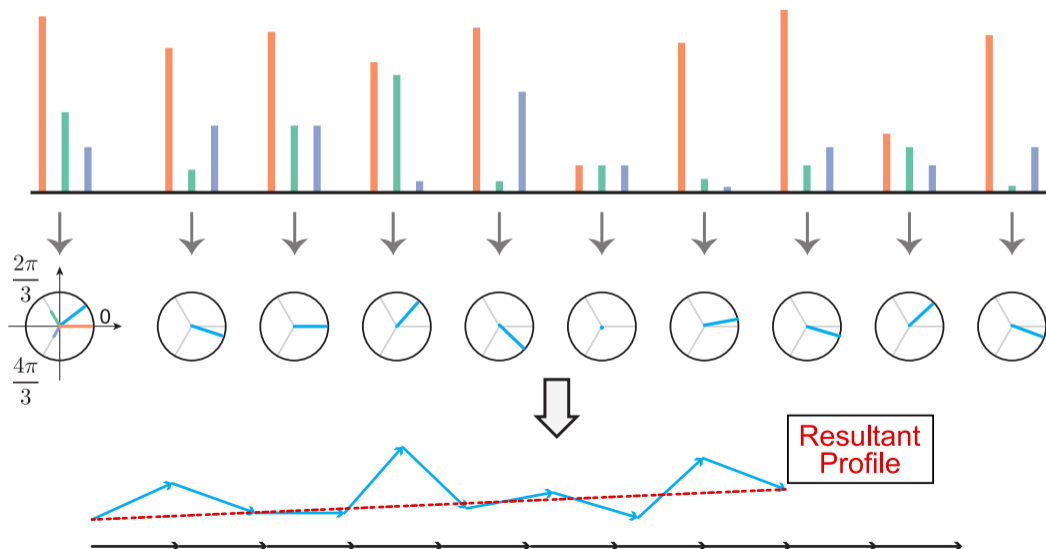
Ribotricer's phase score calculation



Ribotricer's phase score calculation



Ribotricker's phase score calculation



Ribotricer's phase score calculation

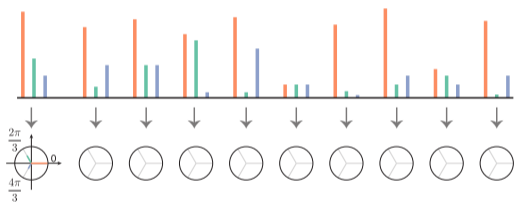
$$\mathbf{x}_i = \left(x_{i1} \quad x_{i2} \quad x_{i3} \right)^T$$



Ribotricer's phase score calculation

$$\mathbf{x}_i = (x_{i1} \quad x_{i2} \quad x_{i3})^T$$

$$\mathbf{W} = \begin{pmatrix} 1 & 0 \\ \cos(-2\pi/3) & \sin(-2\pi/3) \\ \cos(-4\pi/3) & \sin(-4\pi/3) \end{pmatrix}$$

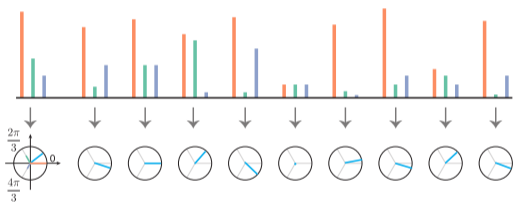


Ribotricer's phase score calculation

$$\mathbf{x}_i = \begin{pmatrix} x_{i1} & x_{i2} & x_{i3} \end{pmatrix}^T$$

$$\mathbf{W} = \begin{pmatrix} 1 & 0 \\ \cos(-2\pi/3) & \sin(-2\pi/3) \\ \cos(-4\pi/3) & \sin(-4\pi/3) \end{pmatrix}$$

$$\phi_i = \frac{\mathbf{W}^T \mathbf{x}_i}{\|\mathbf{W}^T \mathbf{x}_i\|}$$



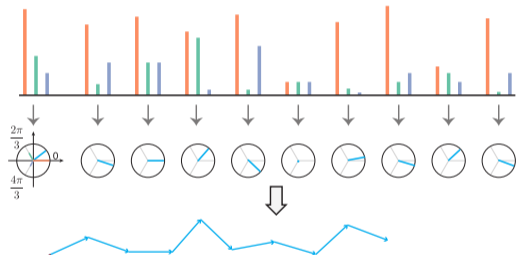
Ribotricer's phase score calculation

$$\mathbf{x}_i = (x_{i1} \quad x_{i2} \quad x_{i3})^T$$

$$\mathbf{W} = \begin{pmatrix} 1 & 0 \\ \cos(-2\pi/3) & \sin(-2\pi/3) \\ \cos(-4\pi/3) & \sin(-4\pi/3) \end{pmatrix}$$

$$\phi_i = \frac{\mathbf{W}^T \mathbf{x}_i}{\|\mathbf{W}^T \mathbf{x}_i\|}$$

$$\bar{\phi} = \frac{1}{N} \sum_{i=1}^N \phi_i$$



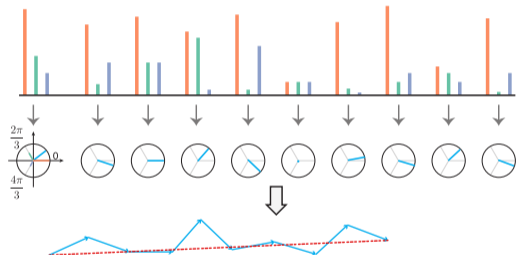
Ribotricer's phase score calculation

$$\mathbf{x}_i = (x_{i1} \quad x_{i2} \quad x_{i3})^T$$

$$\mathbf{W} = \begin{pmatrix} 1 & 0 \\ \cos(-2\pi/3) & \sin(-2\pi/3) \\ \cos(-4\pi/3) & \sin(-4\pi/3) \end{pmatrix}$$

$$\phi_i = \frac{\mathbf{W}^T \mathbf{x}_i}{\|\mathbf{W}^T \mathbf{x}_i\|}$$

$$\bar{\phi} = \frac{1}{N} \sum_{i=1}^N \phi_i$$





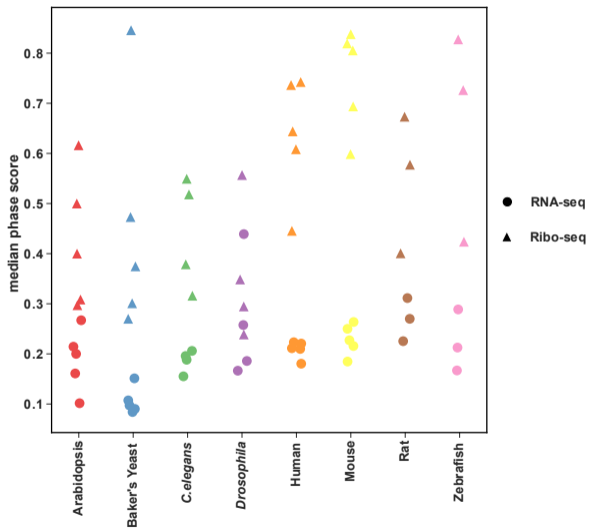
- Focus on expressed Coding Domain Sequences (CDS)
- Ribo-seq profiles from CDS exons: True positives
- RNA-seq profiles from CDS exons: True negatives



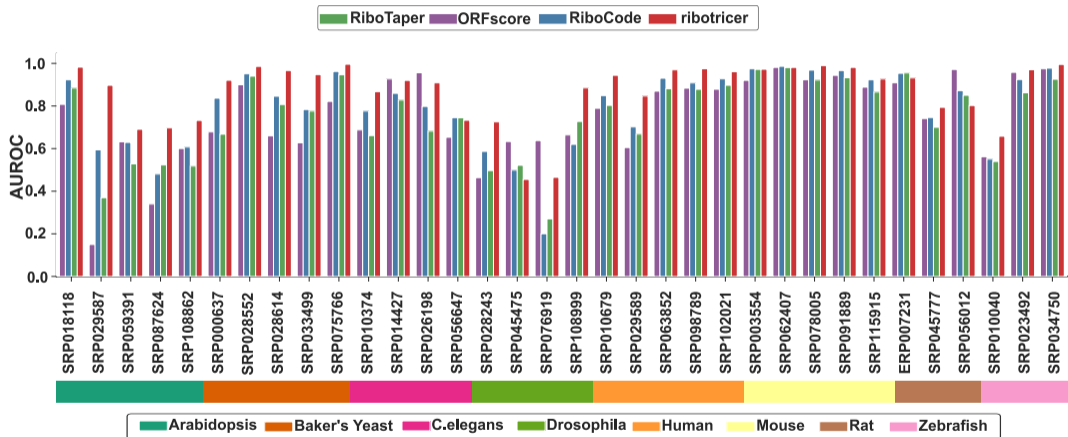
- Focus on expressed Coding Domain Sequences (CDS)
- Ribo-seq profiles from CDS exons: True positives
- RNA-seq profiles from CDS exons: True negatives

Goal: Detect actively translating ORFs with minimal dependence on the ORF length.

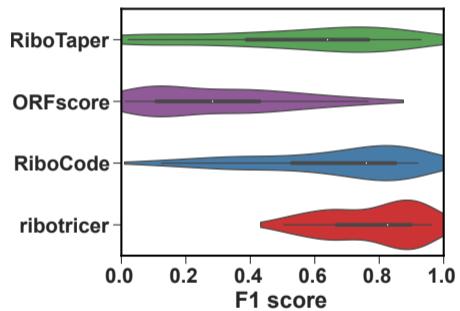
Ribo-seq data is heterogeneous across species



Ribotricker achieves highest AUROC across datasets



Ribotricer achieves highest F1 score across datasets

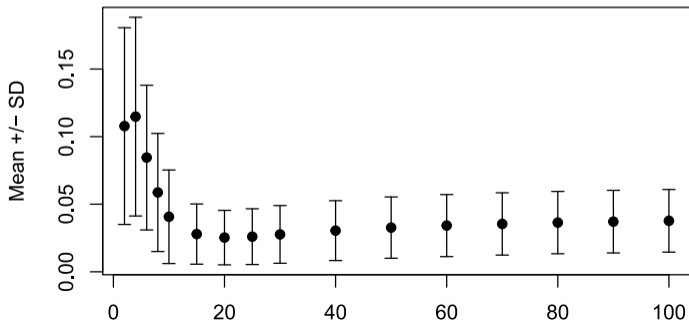


Ribotracer can detect ORFs as short as 20 codons

- Calculate the deviation of phase score between a full length exon and its shorter variant created by subsampling codons

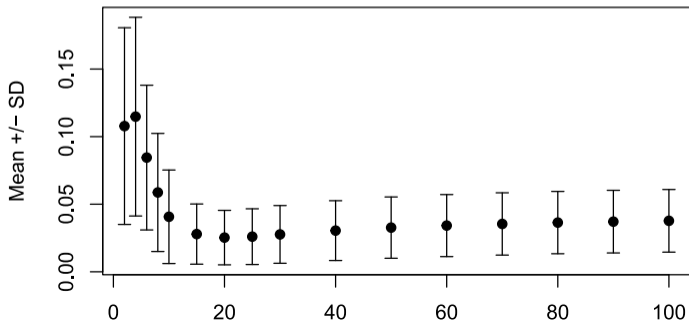
Ribotricker can detect ORFs as short as 20 codons

- Calculate the deviation of phase score between a full length exon and its shorter variant created by subsampling codons



Ribotricker can detect ORFs as short as 20 codons

- Calculate the deviation of phase score between a full length exon and its shorter variant created by subsampling codons
- Ribotricker can accurately detect ORFs as short as 20 codons



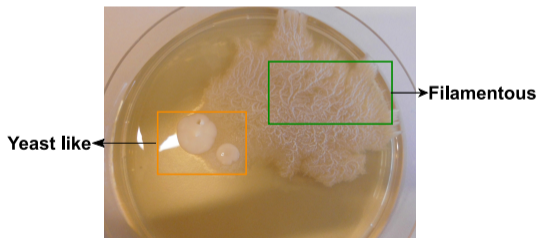
Identifying actively translating regions using ribotracer

Yeast-filament transition in *C. albicans*

- *C. albicans*: a fungal pathogen that inhabits the mucosal surfaces of most healthy individuals

Yeast-filament transition in *C. albicans*

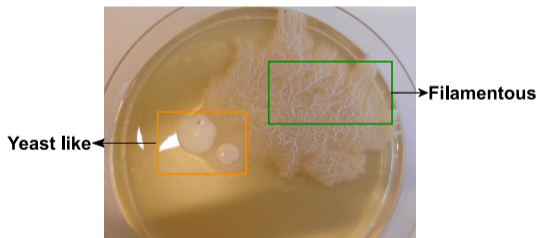
- *C. albicans*: a fungal pathogen that inhabits the mucosal surfaces of most healthy individuals
- Reversible morphological transition from single budding yeasts to continuously branching filaments increases virulence



By Garnhami - Own work, CC BY-SA 4.0

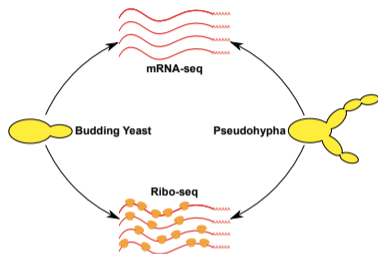
Yeast-filament transition in *C. albicans*

- *C. albicans*: a fungal pathogen that inhabits the mucosal surfaces of most healthy individuals
- Reversible morphological transition from single budding yeasts to continuously branching filaments increases virulence
- Translational landscape changes unknown: Avenues for improving drug efficacy



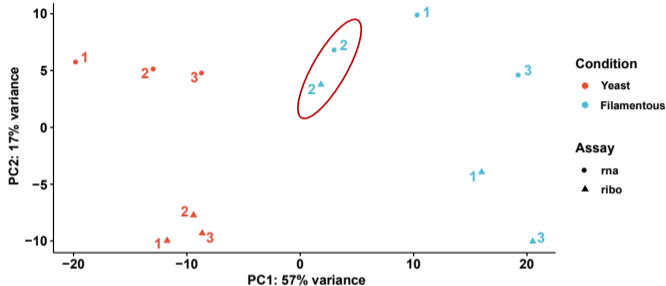
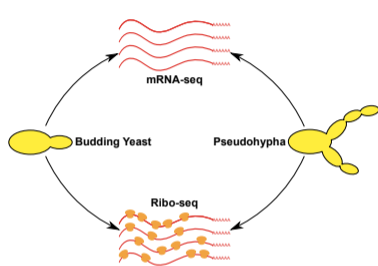
By Garnhami - Own work, CC BY-SA 4.0

Ribotricker enables recovery of ORFs with low signal to noise ratio



- High rRNA contamination leads to a very few shallow sequenced samples

Ribotricker enables recovery of ORFs with low signal to noise ratio

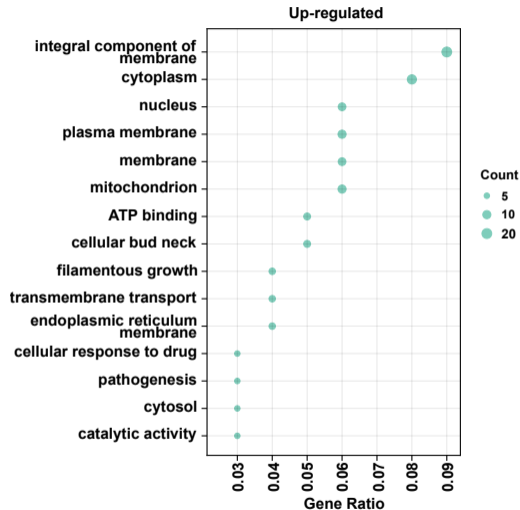


- High rRNA contamination leads to a very few shallow sequenced samples
- Most ORF detection methods fail at detecting any translating regions because of low signal to noise ratio

Ribotricker identifies genes playing differential role at translational level

Ribotricker detects hundreds of genes playing a differential role at the translational level in yeast to filamentous transition.

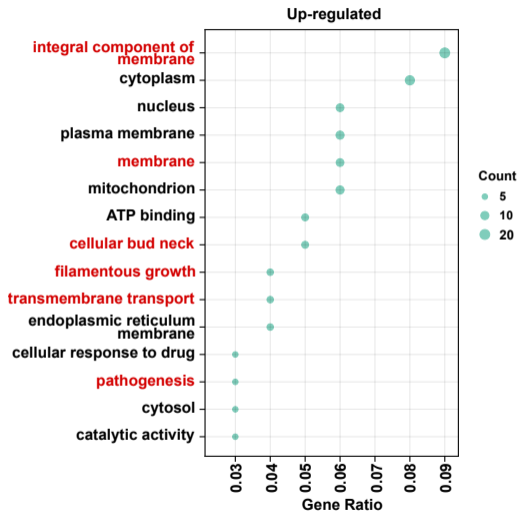
Gene	Function	FC
HMS1	TF required for morphogenesis	3.6
ERK1	Kinase required for yeast-hyphal	3.5
	switch	
PTC8	Required for hyphal growth	3.1
RAX2	Involved in establishment of bud sites and hyphal growth	2.4



Ribotricker identifies genes playing differential role at translational level

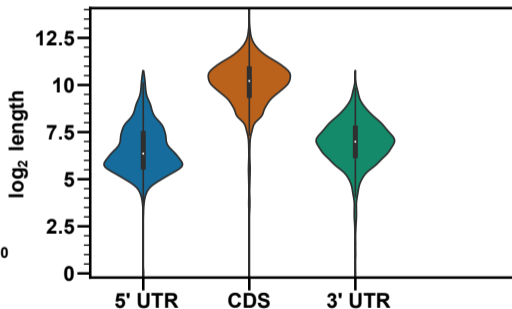
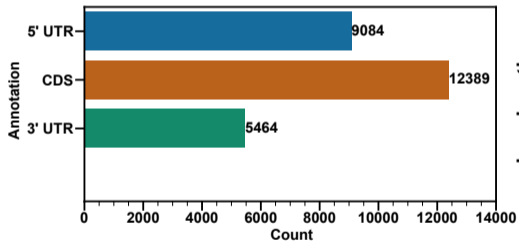
Ribotricker detects hundreds of genes playing a differential role at the translational level in yeast to filamentous transition.

Gene	Function	FC
HMS1	TF required for morphogenesis	3.6
ERK1	Kinase required for yeast-hyphal switch	3.5
PTC8	Required for hyphal growth	3.1
RAX2	Involved in establishment of bud sites and hyphal growth	2.4



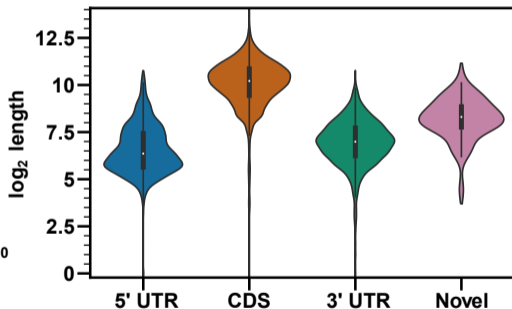
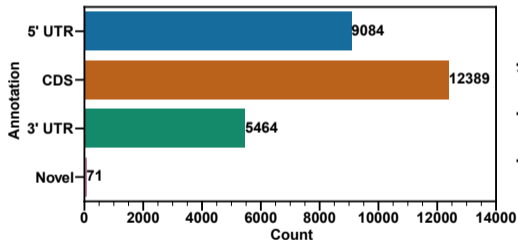
Ribotricker enables discovery of novel ORFs

- Ribotricker can be used to discover novel ORFs that are unannotated



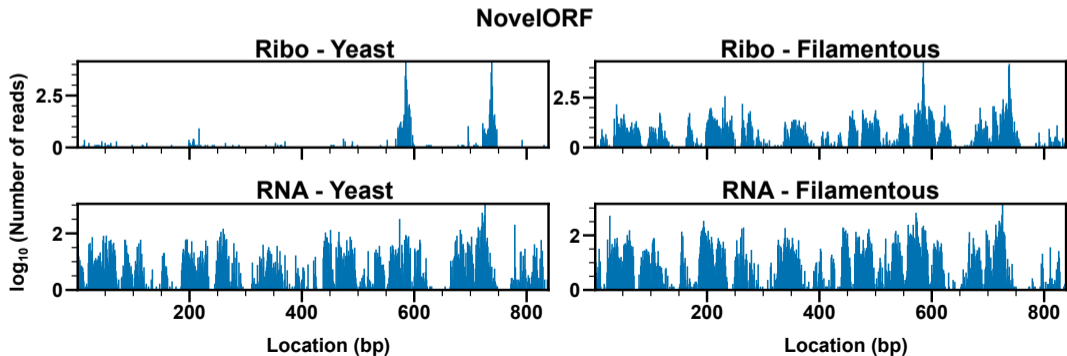
Ribotricker enables discovery of novel ORFs

- Ribotricker can be used to discover novel ORFs that are unannotated
- Ribotricker discovered 71 novel ORFs that are currently missing from *C. albicans*' annotation

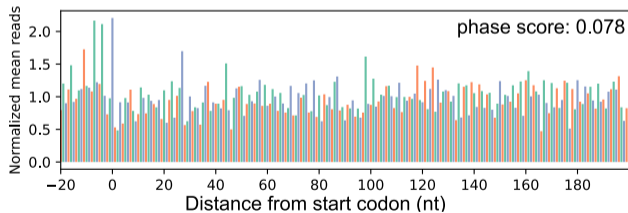
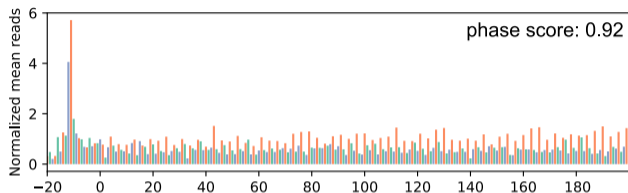


Ribotricer enables discovery of novel ORFs

- Ribotricer can be used to discover novel ORFs that are unannotated
- Ribotricer discovered 71 novel ORFs that are currently missing from *C. albicans*' annotation

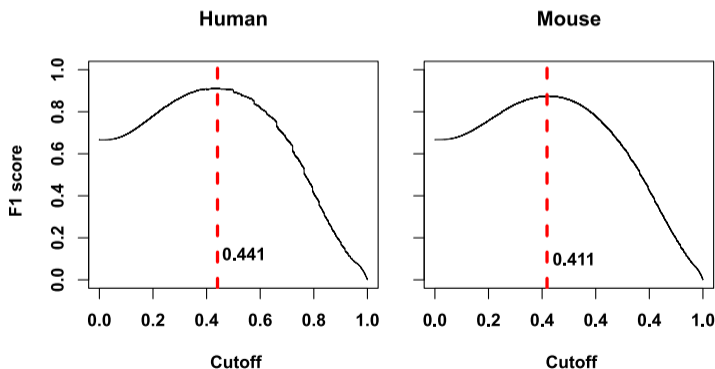


Ribotricker's phase score captures periodicity in Ribo-seq profile

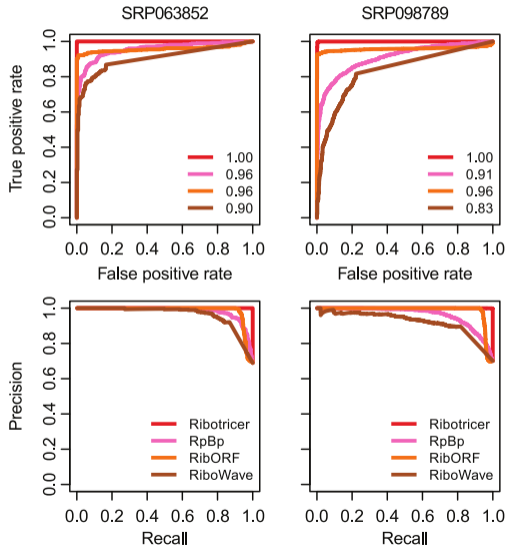


Ribotricker cutoff learned by maximizing the F1 score

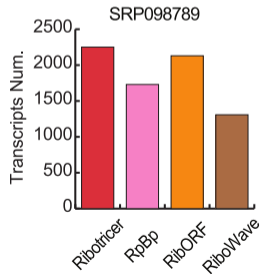
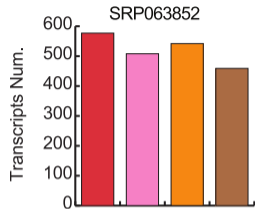
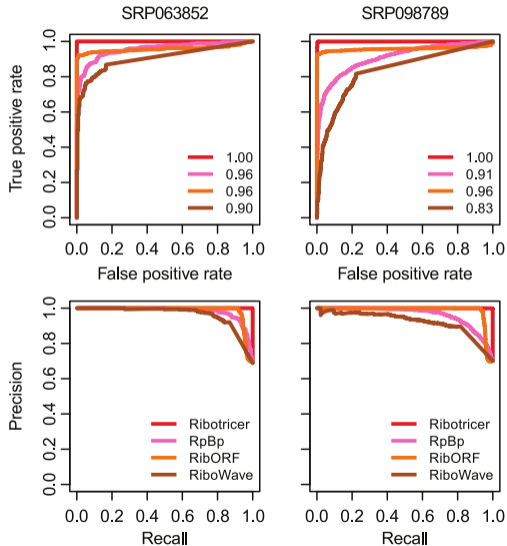
- F1 score = harmonic mean of precision and recall
- Using public datasets in human and mouse, we choose the phase-score cutoff as the score resulting in maximum F1 score



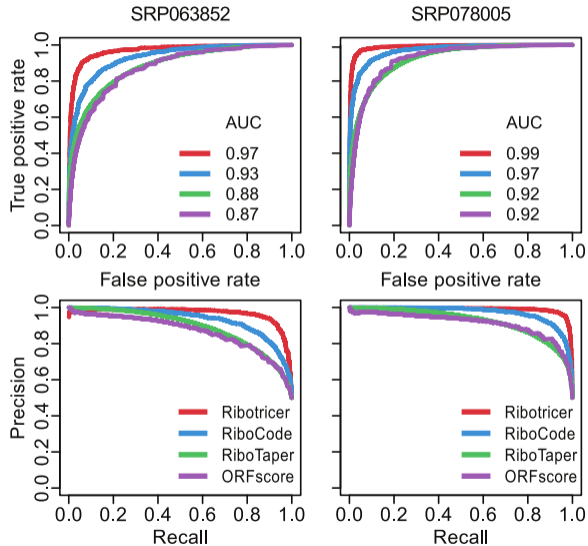
Ribotricer outperforms other methods at the isoform level



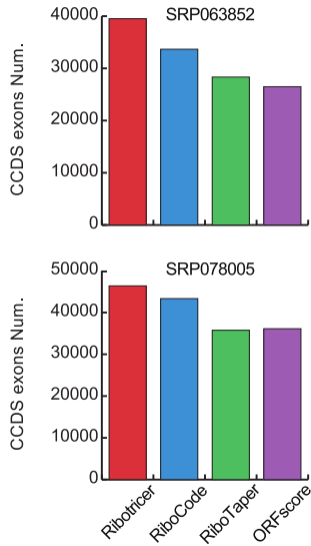
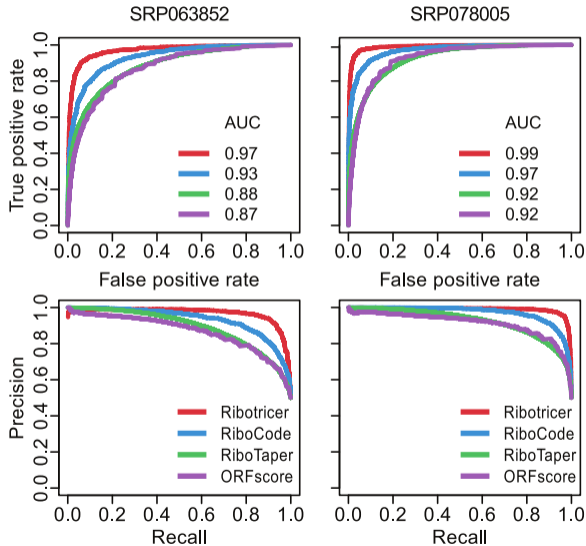
Ribotricker outperforms other methods at the isoform level



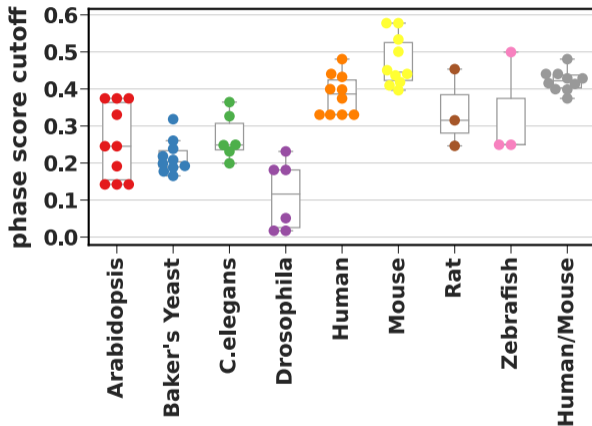
Ribotricker outperforms other methods at the exon level



Ribotricker outperforms other methods at the exon level

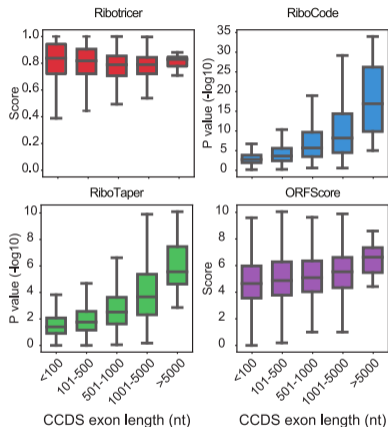


Species-specific cutoffs can vary with datasets

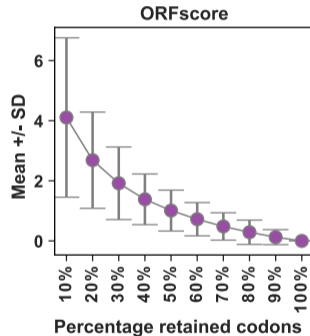
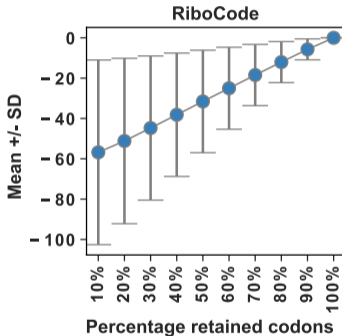
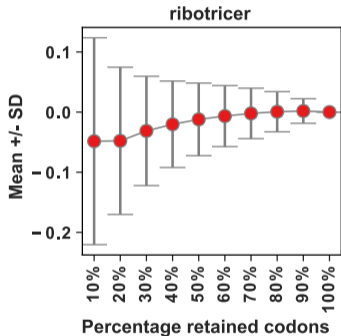


Ribotricker's performance is consistent across short and long ORFs

- Ribotricker's performance unaffected by signal strength
- Particularly advantageous in regions of low signal to noise ratio

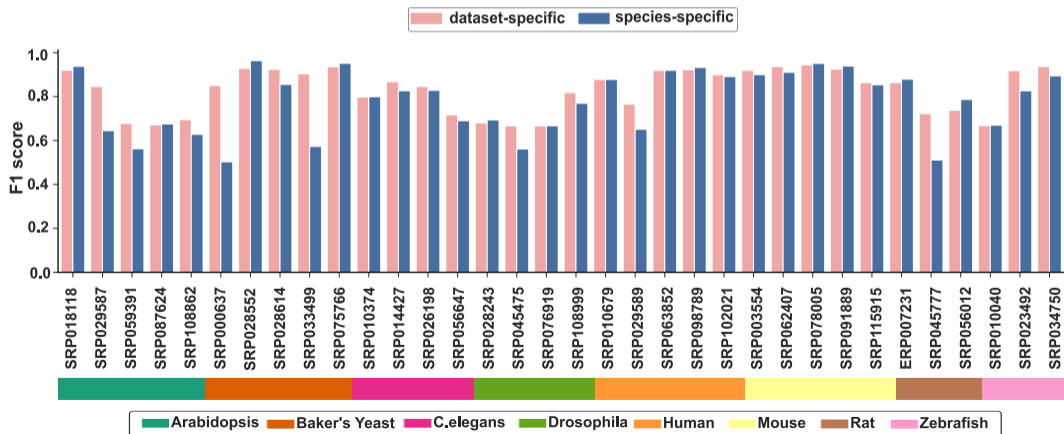


Ribotricker's performance is minimally dependent on ORF length

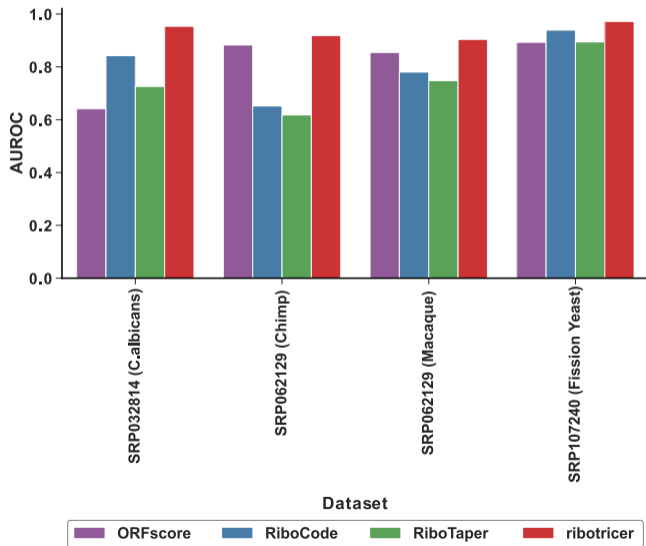


Dataset specific cutoffs can give marginal improvement

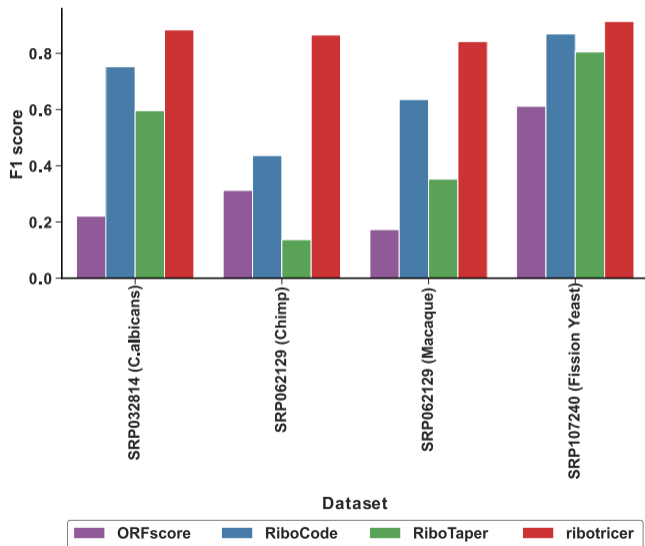
Ribotricker can also learn dataset-specific cutoffs based on availability of RNA- and Ribo-seq dataset



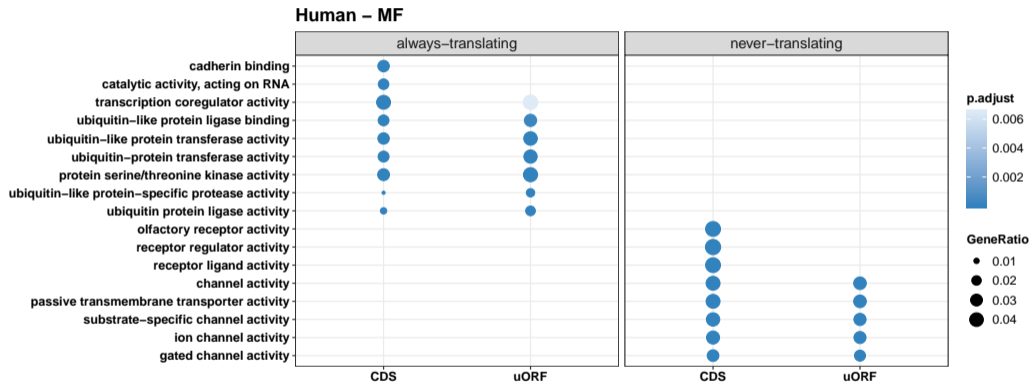
Ribotricker achieves highest AUROC in independent species



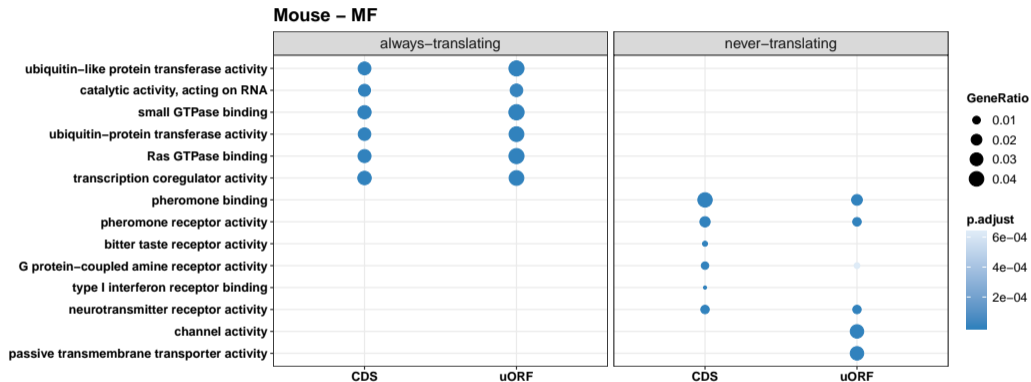
Ribotricker achieves highest F1 score in independent species



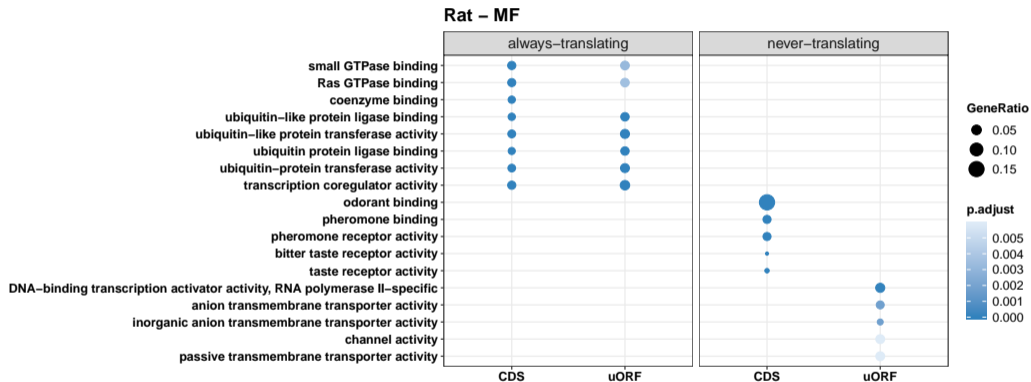
Membrane proteins seem to be not-profiled by Ribo-seq



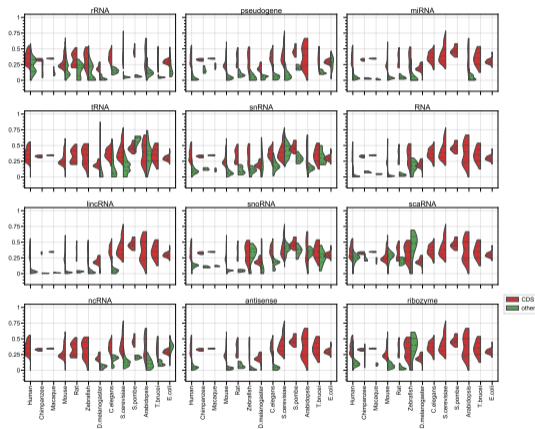
Membrane proteins seem to be not-profiled by Ribo-seq



Membrane proteins seem to be not-profiled by Ribo-seq



Phase scores with transcript category



Phase scores correlation conservation

