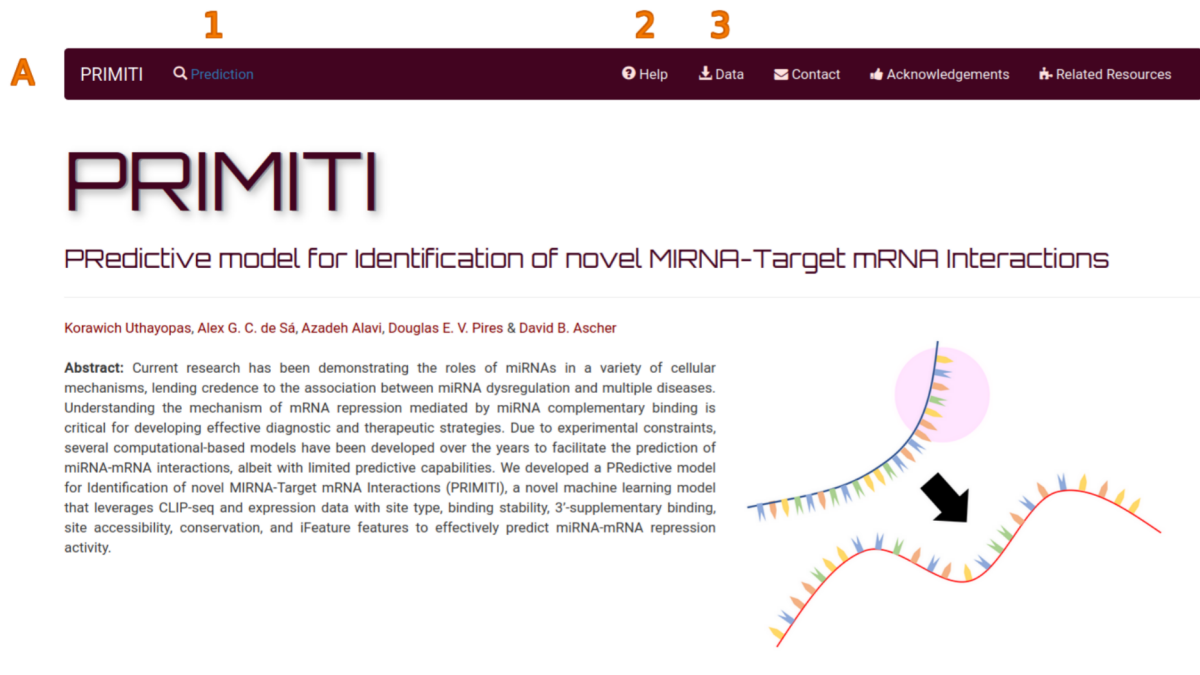


User Manual of PRIMITI

About PRIMITI



1 **2** **3**

A PRIMITI Prediction Help Data Contact Acknowledgements Related Resources

PRIMITI

PRedictive model for Identification of novel MIRNA-Target mRNA Interactions

Korawich Uthayopas, Alex G. C. de Sá, Azadeh Alavi, Douglas E. V. Pires & David B. Ascher

Abstract: Current research has been demonstrating the roles of miRNAs in a variety of cellular mechanisms, lending credence to the association between miRNA dysregulation and multiple diseases. Understanding the mechanism of mRNA repression mediated by miRNA complementary binding is critical for developing effective diagnostic and therapeutic strategies. Due to experimental constraints, several computational-based models have been developed over the years to facilitate the prediction of miRNA-mRNA interactions, albeit with limited predictive capabilities. We developed a PRedictive model for Identification of novel MIRNA-Target mRNA Interactions (PRIMITI), a novel machine learning model that leverages CLIP-seq and expression data with site type, binding stability, 3'-supplementary binding, site accessibility, conservation, and iFeature features to effectively predict miRNA-mRNA repression activity.

PRedictive model for Identification of novel MIRNA-Target mRNA Interaction (PRIMITI) is a machine learning model that utilizes diverse CLIP-seq and expression data to predict novel miRNA-target mRNA interactions. The main focus of PRIMITI model is a miRNA-induced post-transcriptional repression caused by canonical binding between miRNAs and target sites in 3'-UTR of protein-coding transcripts. That is because the canonical binding in 3'-UTR is considered as a functional site type that is more likely to lead to transcriptional repression.

PRIMITI utilizes 22 distinct features to describe physical miRNA-target site binding/duplex pairing and 4 features to describe miRNA-target mRNA repression.

Marked sections in the above toolbox, include:

(A) depicts the main page of PRIMITI. Users are directed to the job submission page by clicking on **Prediction (1)** page at the top menu. **Help page (2)** also available to instruct the users how to use PRIMITI. **Data page (3)** provides the user with all data used in PRIMITI training, (cross-validation) and evaluation.

Job Submission Page

B PRIMITI Q Prediction Help Data Contact Acknowledgements Related Resources

PRIMITI

PRedictive model for Identification of novel MIRNA-Target mRNA Interactions

Step 1: Please provide a set of miRNAs

1 miRNAs file **OR** **2** miRNA string (miRBase format)

Files are expected to have a header "miRNA" identifying the miRNAs column [miRNA File Example].

Step 2: Please provide a set of transcripts (ENST IDs)

3 Transcript ID file **OR** **4** transcript string (ENST IDs)

Files are expected to have a header "transcript_id" identifying the transcript ID column [Transcript File Example].

Step 3: Fill your email address (optional) AND/OR press the button below for predicting the interactions

Obs: The number of miRNAs times the number of targets must not exceed 30,000.

5 E-mail address (for sending a notice with the result link):

In the **Job submission page (B)**, users need to provide identifiers for **miRNAs** and candidate **target messenger RNAs (mRNAs)**. Users may provide the information as a **file (1/3)** or as a **string (2/4)**. The example files for both miRNA and protein-coding transcript (linked to mRNA) can be downloaded from the buttons.

For transcripts, **Ensembl transcript (ENST) form for inputs** are highly preferred, as our model makes predictions using the information of the transcript. Nevertheless, our web server still supports several gene ID formats, and users can instead provide gene identifiers in place of transcripts. All the protein-coding transcripts associated with a given gene will be examined. For instance, If ENSG00000141510 (TP53) is provided, all the protein-coding transcripts for TP53 will be analysed in the PRIMITI model.

The list of supported input format for transcript/gene are as following:

- 1) **Ensembl Transcript** : ENST00000340080
- 2) **Ensembl Gene** : ENSG00000141510
- 3) **NCBI RefSeq Transcript** : NM_001003806, NP_001263195
- 4) **NCBI Gene name** : SIDT2, TP53
- 5) **NCBI Gene ID** : NCBIGeneID:7157, NCBIGeneID:1956
- 6) **MIM Gene ID** : MIM:165280, MIM:613919
- 7) **HGNC Gene ID** : HGNC:11547, HGNC:46137

Please note that the analysis **request** that **contains a gene identifier will take longer to process when compared to those with only containing transcripts**. If users are interested in the analysis of both gene and transcript, we recommend submitting the jobs individually.

An **email address** can be also provided by the user (in an optional form) **(5)**, so that the user can be notified with an access link when the analysis is finished.

Notably, our web server only accepts **the number of candidate interactions** (number of miRNAs x number of transcripts) up to **30,000**. If the user's data is larger than the limitation, please consider submitting them separately. In a case that gene identifiers are provided, the number of candidate interactions will be multiplied by 3. (For example, 1 miRNA and 1 gene = 3 candidates interactions)

Waiting Page

C

PRIMITI Prediction Help Data Contact Acknowledgements Related Resources

PRIMITI is processing your submission...

We are processing your submission.

Sorry! We are taking longer than expected to process your inputs. We added more 20 min. to the total processing time of your queries.

You can bookmark (★) this page and come back later to check the results and/or wait for an email, if you set it on the previous page.

This page will automatically refresh after 10 seconds and your results will be displayed as soon as they are ready.

Max. Waiting Time (17 min.)

After submitting the analysis, users will be redirected to a **waiting page (C)**. If the email address is filled in **(B)**, users can receive the link to the result page in the email.

Result Page – Main

D PRIMITI Prediction

Help Data Contact Acknowledgements Related Resources

PRIMITI Prediction result

Show 10 entries

miRNA 1	Provided ID 2	Transcript ENST 3	Gene ENSG 4	Interaction 5	Interaction Confidence 6	No. of target sites 7	More details 8
hsa-let-7d-5p	ENSG00000141510	ENST00000619186	ENSG00000141510	Yes	0.942	1	View Details
hsa-let-7d-5p	ENSG00000141510	ENST00000619485	ENSG00000141510	Yes	0.942	1	View Details
hsa-let-7d-5p	ENSG00000141510	ENST00000620739	ENSG00000141510	Yes	0.942	1	View Details
hsa-let-7d-5p	ENSG00000141510	ENST00000622645	ENSG00000141510	Yes	0.942	2	View Details
hsa-let-7a-5p	MIM:613919	ENST00000287152	ENSG00000164627	Yes	0.94	1	View Details
hsa-let-7a-5p	ENST00000340080	ENST00000340080	ENSG00000164654	Yes	0.939	1	View Details
hsa-let-7d-5p	MIM:613919	ENST00000287152	ENSG00000164627	Yes	0.935	1	View Details
hsa-let-7a-5p	NCBIGeneID:1956	ENST00000455089	ENSG00000146648	No	0.417	1	View Details
hsa-let-7d-5p	NCBIGeneID:1956	ENST00000455089	ENSG00000146648	No	0.309	1	View Details
hsa-let-7d-5p	ENST00000340080	ENST00000340080	ENSG00000164654	No	0.075	1	View Details

Showing 31 to 40 of 42 entries

Previous 1 2 3 4 5 Next

9 [Run another prediction](#)

10 [Download results - target mRNA](#)

11 [Download results - target site](#)

12 [Download error logs](#)

The **Result Page (D)** provides the information of predicted miRNA-target mRNA. Pairs of miRNA and transcript (representing mRNAs) that contain at least one canonical binding site are shown in the table of this page. The identifiers for **miRNAs (1)** and **input transcripts/genes (2)** are provided along with **Ensembl Transcript (ENST) (3)** and **Ensembl Gene (ENSG) (4)**.

If a miRNA-transcript pair is predicted to exhibit a post-transcriptional repression, the **Interaction (5)** will be shown as "Yes", otherwise "No". The confidence score for each pair is provided in the column **Interaction Confidence (6)**. To allow a high-confidence prediction, a cut-off of 0.9 was set for predicting the repression. The numbers of all candidates canonical target sites are shown in **No. of target sites (7)**. If users are interested in more detailed information regarding the physical binding in each target site, please click on the button in

The next table provides the following information for each target site **(4)**: **Binding status (5)**, **Binding Confidence score (6)**, **Position in terms of 3'-UTR (7)** or **cDNA (8)**, and **the type of canonical binding (6-mer, 7-mer-A1, 7-mer-m8, and 8-mer) (9)**.

Predicted miRNA-target site duplex structure (10) is shown as a picture. The upper sequence is miRNA sequence (5' to 3') and the bottom is mRNA sequence (3' to 5'). Across the sequence, the predictions of the base pairwise interactions are shown. One line represents one base pairwise. Additionally, as the presence of adenine (A) nucleobase in the first position of mRNA (3') strongly correlates with a significantly higher miRNA-mRNA duplex stability, Adenine nucleotide will be depicted in brown if it is present at the first position.

Error Page



The screenshot shows the PRIMITI web application interface. At the top, there is a dark blue navigation bar with the PRIMITI logo on the left and links for Help, Data, Contact, Acknowledgements, and Related Resources on the right. Below the navigation bar, the PRIMITI logo is displayed in a large, stylized font. Underneath the logo, the full name of the application is written: "PRedictive model for Identification of novel MIRNA-Target mRNA Interactions". A red "Error" heading is followed by a message: "You have not provided any miRNA (string) or a list of miRNAs (file), neither any single transcript (in string) or a list of transcript (ID file)! Please, provide them first." At the bottom of the error message, there is a button labeled "← Back to submission".

The Error Page (F) is shown when your input is incorrect. Please check the example files for an input format. Common problems are 1) Wrong format in input files 2) Swap miRNA and transcript input files

If you have any inquiry, please do not hesitate to contact us via our **Group website** (<http://biosig.unimelb.edu.au/biosig/contact>).

Downloadable results

G

index	miRNA	Provided ID	Transcript ENST	Gene ENSG	Interaction	Interaction Prob. Score [0-1]	No. of target sites
0	hsa-let-7a-5p	ENSG00000141510	ENST00000269305	ENSG00000141510	Yes	0.942	1
1	hsa-let-7a-5p	MIM:613919	ENST00000287152	ENSG00000164627	Yes	0.94	1
2	hsa-let-7a-5p	NCBIGeneID:3479	ENST00000337514	ENSG0000017427	Yes	0.942	1
3	hsa-let-7a-5p	ENST00000340080	ENST00000340080	ENSG00000164654	Yes	0.939	1
4	hsa-let-7a-5p	ENSG00000141510	ENST00000420246	ENSG00000141510	Yes	0.942	2
6	hsa-let-7a-5p	ENSG00000141510	ENST00000445888	ENSG00000141510	Yes	0.942	1
7	hsa-let-7a-5p	NCBIGeneID:1956	ENST00000455089	ENSG00000146648	No	0.417	1
8	hsa-let-7a-5p	ENSG00000141510	ENST00000455263	ENSG00000141510	Yes	0.942	2
10	hsa-let-7a-5p	ENSG00000141510	ENST00000504290	ENSG00000141510	Yes	0.942	2
12	hsa-let-7a-5p	ENSG00000141510	ENST00000504937	ENSG00000141510	Yes	0.942	1
13	hsa-let-7a-5p	ENSG00000141510	ENST00000510385	ENSG00000141510	Yes	0.942	2
15	hsa-let-7a-5p	MIM:601945	ENST00000538548	ENSG00000061936	Yes	0.942	1
16	hsa-let-7a-5p	NP_001263195	ENST00000560572	ENSG00000140254	No	0.058	1

The **predicted miRNA-target mRNA repressions (G)** can be retrieved from a **button in Result Page - Main (D-10)** as a csv file. The provided information is the same as in the result table in **(D)**. It contains identifiers for miRNA, Provided ID, Transcript ENST, Gene ENSG, Interaction status (Yes: post-transcriptional repression, No: no repression), Confidence score, and Number of target sites.

H

miRNA	Provided ID	Transcript ENST	Gene ENSG	Interaction	Interaction Prob. Score [0-1]	No. of target sites
hsa-let-7a-5p	ENSG00000141510	ENST00000269305	ENSG00000141510	Yes	0.942	1
hsa-let-7a-5p	MIM:613919	ENST00000287152	ENSG00000164627	Yes	0.94	1
hsa-let-7a-5p	NCBIGeneID:3479	ENST00000337514	ENSG00000017427	Yes	0.942	1
hsa-let-7a-5p	ENST00000340080	ENST00000340080	ENSG00000164654	Yes	0.939	1
hsa-let-7a-5p	ENSG00000141510	ENST00000420246	ENSG00000141510	Yes	0.942	2
hsa-let-7a-5p	ENSG00000141510	ENST00000420246	ENSG00000141510	Yes	0.942	2

I

No. of target sites	Target site no.	Position in 3'-UTR	Position in cDNA	Binding	Binding Prob. Score [0-1]	Binding type	miRNA sequence	Bound transcript sequence
1	1	396-417	1720-1741	Yes	0.939	8-mer	UGAGGUAGUAGGUU GUAUAGUU	GCUGGCAUUUGC ACCUACCUCA
1	1	1955-1976	4498-4519	Yes	0.893	7-mer-A1	UGAGGUAGUAGGUU GUAUAGUU	GCAGUGGUGCGA CCAUACCUCA
1	1	1297-1318	1941-1962	Yes	0.939	8-mer	UGAGGUAGUAGGUU GUAUAGUU	AAU AUGUCAUCU ACCUACCUCA
1	1	57-78	3094-3115	Yes	0.815	7-mer-A1	UGAGGUAGUAGGUU GUAUAGUU	CAUAGCUCAGAA ACAUACCUCA
2	1	220-241	1379-1400	Yes	0.929	7-mer-m8	UGAGGUAGUAGGUU GUAUAGUU	CAAAAAGGGUCA GUCUACCUCC
2	2	685-706	1844-1865	Yes	0.939	8-mer	UGAGGUAGUAGGUU GUAUAGUU	GCUGGCAUUUGC ACCUACCUCA

J

Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12	Col13	Col14	Col15
U	G	A	G	G	U	A	G	U	A	G	G	U	U	G
A	C	U	C	C	A	U	C	C	A	C	G	U	U	U
U	G	A	G	G	U	A	G	U	A	G	G	U	U	G
													:	
A	C	U	C	C	A	U	A	C	A	C	C	A	G	C
U	G	A	G	G	U	A	G		U	A	G	G	U	U
												:		
A	C	U	C	C	A	U	C	C	A	U	C	U		A
U	G	A	G	G	U	A		G	U					A
A	C	U	C	C	A	U	A	C	A	A	A	G	A	C
U	G	A	G	G	U	A	G	U	A	G	G	U	U	G
C	C	U	C	C	A	U	C	U	G	A	C	U	G	G
U	G	A	G	G	U	A	G	U	A	G	G	U	U	G
A	C	U	C	C	A	U	C	C	A	C	G	U	U	U

The predicted miRNA-target site interactions (H/I/J) can be retrieved using the dedicated download button in **Result Page - Main (D/11)** as a csv file. The provided information contains predicted miRNA-target mRNA repression (same as in **(D)**), and data in each target site, including the Predicted binding status (Yes: miRNA can bind with target site, No: they are not binding), Binding confidence (range from No (0) to Yes (1)), Binding type (6-mer, 7-mer-A1, 7-mer-m8, and 8-mer), mRNA sequence, Bound transcript sequence, and Predicted duplex structure calculated by IntaRNA.

K # No miRNA errors were found.
The following input_1(s) was (were) not supported --- The types of inputs are not recognized. Please check if the inputs are in correct forms:
MIM:Trash, HGNC:MIM_Trash, HGNC:Trash, Another_type_of_Trash_ID_3, ENSG0000Trash, HGNC:MIM:NCBIGeneID_Trash, Trash_ID_1, Trash_ID_2, NCBIGeneIDTrash, NM_Trash, TP53Trash, MIM:NM_Trash

The following input_2(s) was (were) not supported --- The types of inputs are recognized; however, no matched supported ENSTs are found. Please check if the inputs specify protein-coding transcripts:
MIR3165, NM_001290307, HGNC:46137, ALOX12B, RPL17P8

The following input_3(s) was (were) not supported --- The provided ENST are not supported. Please check if the inputs are in correct forms and specify protein-coding transcripts:
ENST0000Trash

An error log (K) can be downloaded using the dedicated **button in Result Page - Main (D-12)**. The error entries and related causes will be reported in this file. Please note that the analysis submission will still be done even though some error entries are found. That is, the error entries will be ignored and the analysis will be accomplished with the remaining correct entries. To illustrate, if users provide 100 transcripts, in which 5 of them are not correct, the analysis will be performed with 95 transcripts.

4 Type of errors will be reported:

- **1) Errors in miRNAs:** most likely occur from incorrect miRNA format.
- **2) Errors in transcript/gene inputs:** The type of the inputs is not recognizable. It is more likely that the inputs are in incorrect forms. For example, Trash_ID_1 and Trash_ID_2
- **3) Errors in transcript/gene inputs:** The type of inputs is recognizable, but it does not specify any protein-coding transcripts. For example, RPL17PB specifies a pseudogene, which can be found in the database, but not supported in the model.
- **4) Errors in ENST inputs:** The provided ENST are not supported. It can be caused by 1) incorrect forms of ENST (For example, ENST0000Trash) 2) Not a protein-coding transcript.