

TSMDA: Target and symptom-based computational model for miRNA-disease association prediction

Korawich Uthayopas^{1,2,3}, Alex G.C. de Sá^{1,2,3,4}, Azadeh Alavi^{1,2,3}, Douglas E.V. Pires^{1,2,3,5*}, David B. Ascher^{1,2,3,4,6*}

¹Structural Biology and Bioinformatics, Department of Biochemistry, University of Melbourne, Parkville 3052, Victoria, Australia

²Systems and Computational Biology, Bio21 Institute, University of Melbourne, Parkville 3052, Victoria, Australia

³Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute, Melbourne 3004, Victoria, Australia

⁴Baker Department of Cardiometabolic Health, Melbourne Medical School, University of Melbourne, Parkville 3010, Victoria, Australia

⁵School of Computing and Information Systems, University of Melbourne, Parkville 3052, Victoria, Australia

⁶Department of Biochemistry, University of Cambridge, 80 Tennis Ct Rd, Cambridge CB2 1GA

TSMDA Help Page

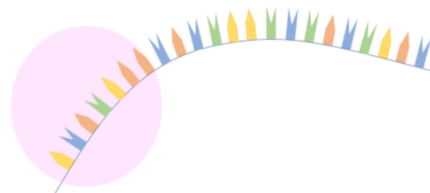
About TSMDA



TsmDA target and symptom-based computational model for miRNA-disease association prediction

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

Abstract: The emergence of high-throughput sequencing techniques has revealed a primary role of miRNAs in a wide range of diseases, including cancers and neurodegenerative disorders. Understanding novel relationships between miRNAs and diseases can potentially unveil complex pathogenesis mechanisms, leading to effective diagnosis and treatment. The investigation of novel miRNA-disease associations, however, is currently costly and time consuming. Over the years, several computational models have been proposed to prioritize potential miRNA-disease associations, however, with limited usability or predictive capability. In order to fill this gap, we introduce TSMDA, a novel machine learning method that leverages target and symptom information and negative sample selection to predict miRNA-disease association. TSMDA significantly outperforms similar methods, achieving an Area Under the ROC curve (AUC) of 0.989 and 0.982 under 5-fold cross-validation and blind test, respectively. We also demonstrate the capability of the method to uncover potential miRNA-disease associations in breast, prostate, and lung cancers, as case studies. We believe TSMDA will be an invaluable tool for the community to explore and prioritise potentially new miRNA-disease association for further experimental characterization.



TSMDA is a machine learning-based model that leverages target and symptom information and two robust negative sample selection approaches to accurately predict potential miRNA-disease associations. This model was created based on available known associations in HMDD v2.0.

(A) represents the main page of TSMDA:

- Users are directed to the submission page when clicking on **“Prediction”** at the top menu (1).
- Data used to train and validate the model can be found on the **“Data”** tab (2).
- In case you experience any trouble using TSMDA or if you have any suggestions or comments, please do not hesitate to contact us either via [Biosig’s webpage](#) (Subject: TSMDA) by clicking on the **“Contact”** page (3) or by [email](#).
 - If you are contacting us regarding a job submission, please include its details such as input information and the job identifier.

Submission Page

B

TSM DA ⚡ Prediction

Help Data Contact Acknowledgements Related Resources

TSM DA target and symptom-based computational model for miRNA-disease association prediction

Step 1: Please provide a set of miRNAs

miRNAs file **1A**

Choose File No file chosen

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

OR

miRNA string **1B**

hsa-miR-548bc

Step 2: Please provide a set of diseases (by using MESH IDs)

MESH ID file **2A**

Choose File No file chosen

Files are expected to have a header "mesh_id" identifying the MESH ID column [MESH File Example].

OR

MESH ID string **2B**

D001943

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

Obs: The number of miRNAs times the number of diseases must not exceed 100.

E-mail address (for sending a notice with the result link): **3**

thanks_for_using_tsm da@unimelb.edu.au

Predict miRNA-disease associations **4**

(B) depicts the submission page.

- Firstly, users need to provide a set of miRNAs as a file (**1A**) **OR** as a single string (**1B**). Only miRNA IDs in [miRBase](#) are acceptable.
- Similarly, users must submit a set of diseases using the **Medical Subject Heading (MeSH)** format. This set can be defined using a file (**2A**) **OR** a single string (**2B**). The MeSH IDs are available at [MeSH Browser \(nih.gov\)](#).
- Users can fill the **email addresses (3)** for sending a notice with the result link when TSM DA prediction is done.
- By clicking on **"Predict miRNA-disease associations" (4)**, the provided list of miRNA and disease will be submitted to TSM DA webserver for prediction.

Waiting Page


C

TSMDA ⚡ Prediction

[Help](#) [Data](#) [Contact](#) [Acknowledgements](#) [Related Resources](#)

TSM_{DA}

 is processing your submission...



We are processing your submission.

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.)

(C) represents the waiting step:

- In this step, the disease-association query is being processed by TSM_{DA}. The maximum waiting time is shown for each submission. Users can bookmark this page and come back later to check the results. If a valid email address was filled in on the submission page, a notice with the result link will be received through it.

Results Page

D

TSMDA ⚡ Prediction

Help Data Contact Acknowledgements Related Resources

TSMDA predictions for the miRNA-disease associations

1

Show 10 entries

Search:

miRNA ID	MeSH ID	Disease	Associated? ✓ ✗	Association Confidence	Evidence: MNDR i	Evidence: dbDEMC i
hsa-miR-17-5p	D013274	Stomach Neoplasms	Yes	98.64	[20234369 '21415212' '21703006]	Not found
hsa-miR-16-5p	D013274	Stomach Neoplasms	Yes	98.64	[18449891 '21081469' '21415212]	Not found
hsa-miR-20a-5p	D013274	Stomach Neoplasms	Yes	98.53	Not found	Not found
hsa-miR-155-5p	D013274	Stomach Neoplasms	Yes	98.46	[21415212 '22426647' '24222951]	Not found
hsa-miR-21-5p	D013274	Stomach Neoplasms	Yes	98.46	[18794849 '21081469' '21415212]	[25167801]
hsa-miR-146a-5p	D013274	Stomach Neoplasms	Yes	98.38	[21347720 '21632853' '22020746]	Not found
hsa-miR-125b-5p	D013274	Stomach Neoplasms	Yes	97.92	[21703006 '28672982]	Not found
hsa-miR-16-5p	D008113	Liver Neoplasms	Yes	97.53	Not found	Not found
hsa-miR-17-5p	D008113	Liver Neoplasms	Yes	97.53	[21861697 '25706130]	Not found
hsa-miR-20a-5p	D008113	Liver Neoplasms	Yes	97.34	Not found	Not found

Showing 1 to 10 of 30 entries

Previous 1 2 3 Next

2 3 4

Run another prediction Download results Download error logs

(D) presents the result page for TSMDA.

- The prediction results are shown as a table (1). The information included miRNA ID, MeSH ID, Disease, Predicted Association, Association Confidence, and Evidence in MNDR and dbDEMC databases in terms of PubMed IDs (PMIDs).
- PMIDs are provided only if miRNA-disease pairs have been confirmed to be associated in MNDR and/or dbDEMC databases.
- Users have the options to run another prediction (2), download the result in a comma-separated values (CSV) format (3), and/or download the error logs (4). Frequently errors are found in the format of miRNAs and diseases (i.e., in the respective miRNA and/or MeSH IDs).

Supplementary Information

Medical Subject Headings (MeSH) ID Retrieval Tutorial

E

1

pancreatic cancer

FullWord ▾

2

Exact Match All Fragments Any Fragment

Sort by: Relevance ▾

Results per Page: 20 ▾

☒ All Terms

☐ Main Heading (Descriptor) Terms

☐ Qualifier Terms

☐ Supplementary Concept Record Terms

☐ MeSH Unique ID

☐ Search in all Supplementary Concept Record Fields

☐ Heading Mapped To

☐ Indexing Information

☐ Pharmacological Action

☐ Search Related Registry and CAS Registry/EC Number/UNII Code/NCBI Taxonomy ID Number (RN)

☐ Related Registry Search

☐ CAS Registry/EC Number/UNII Code/NCBI Taxonomy ID Number (RN)

☐ Search in all Free Text Fields

☐ Annotation

☐ ScopeNote

☐ SCR Note

6 results in 0.563 seconds

1 pages

3

Pancreatic Neoplasms **Descriptor**

Cancer of Pancreas

Neoplasms, Pancreatic

Pancreas Cancer

Pancreatic Cancer

pancreatic cancer-associated antigen MUSE11 **Supplementary Concept Record**

(E) illustrates the “Search” page in [MeSH Browser](#):

- Users need to fill the box with the name of the disease of interest (1). Next, by clicking on either **Exact Match**, **All Fragments**, OR **Any Fragment Boxes** (2), you will define the type of the search. As a result, the list of terms related to the query will be shown (3). By clicking on the selected term, users are redirected to the detailed data of that particular disease term.

F

Pancreatic Neoplasms MeSH Descriptor Data 2021

Details	Qualifiers	MeSH Tree Structures	Concepts
MeSH Heading	Pancreatic Neoplasms		
Tree Number(s)	C04.588.274.761 C04.588.322.475 C06.301.761 C06.689.667 C19.344.421		
1	Unique ID	D010190	
RDF Unique Identifier	http://id.nlm.nih.gov/mesh/D010190		
Annotation	coord IM with histol type of neopl (IM); available are ALPHA-CELL TUMOR see GLUCAGONOMA; BETA-CELL TUMOR see INSULINOMA; and pancreatic delta-cell tumor see SOMATOSTATINOMA		
Scope Note	Tumors or cancer of the PANCREAS. Depending on the types of ISLET CELLS present in the tumors, various hormones can be secreted: GLUCAGON from PANCREATIC ALPHA CELLS; INSULIN from PANCREATIC BETA CELLS; and SOMATOSTATIN from the SOMATOSTATIN-SECRETING CELLS. Most are malignant except the insulin-producing tumors (INSULINOMA).		
Entry Version	PANCREATIC NEOPL		
Entry Term(s)	Cancer of Pancreas Cancer of the Pancreas Neoplasms, Pancreatic Pancreas Cancer Pancreas Neoplasms Pancreatic Cancer		
NLM Classification #	WI 810		
Date Established	1966/01/01		
Date of Entry	1999/01/01		
Revision Date	2012/07/03		

(F) shows the “Details” for the searched disease term on MeSH Browser:

- In this page, the detailed information of the specified disease term is provided. The respective MeSH ID of that particular disease can be found at (1).