Human Absorbable MicroRNA Prediction and Dietary MicroRNA Databases (DMD)

Jiang Shu, Kevin Chiang, Dongyu Zhao, Juan Cui

Dept. of Computer Science and Engineering, University of Nebraska-Lincoln, NE

Abstract

MicroRNAs, a class of short non-coding RNAs, are able to regulate more than half of human genes and affect many fundamental biological processes. It has been long considered synthesized endogenously until very recent discoveries showing that human can absorb exogenous microRNAs from dietary sources. This finding has raised a challenge scientific question: which exogenous microRNAs can be integrated into human circulation and possibly exert functions in human? Here we present an integrated study that includes three components: 1) a well-designed manifold ranking model for identifying human absorbable exogenous microRNAs; 2) the Dietary MicroRNA Databases that archives the published and novel microRNAs discovered in dietary resources; 3) an efficient computational model to predict the cooperative regulatory modules among miRNAs.

Specially, we have comparatively analyzed a total of 34,612 microRNAs from 194 species in terms of 1,102 sequence and structural features of miRNAs. 8 groups of discriminative features have been extracted to identify the human absorbable exogenous microRNAs based on a manifold ranking model. Moreover, we have also launched the first repository, the Dietary MicroRNA Databases, for archiving and analyzing 5,217 microRNAs from 15 food species. Furthermore, by utilizing bi-clustering algorithm, we predicted 173 microRNA regulatory modules that multiple microRNAs co-regulate same set of target genes.

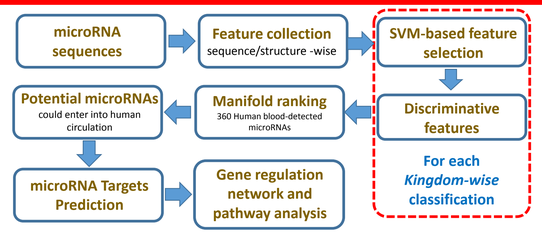
Datasets

The data for this study was collected from the following resources:

- 63,033 miRNA sequences from m*iRBase*, categorized into 5 kingdoms
- 360 human blood miRNAs from literatures
- 5,217 dietary miRNAs from *Dietary micorRNA databases* (DMD)
- 18,514 high-confidence miRNA-mRNA interactions from CLASH dataset

	Animalia	*Plantae*	*Fungi*	*Viruses*	*Protista*	*Total*
Mature microRNAs	*26,705*	*7,645*	*84*	*152*	*26*	***34,612***
Precursor miRNAs	*21,257*	*6,990*	*53*	*91*	*30*	***28,421***
Species	*111*	*71*	*5*	*5*	*2*	***194***

Human Absorbable miRNAs Prediction



- **Kingdom-wise Classification:** *Support vector machine (SVM)*
- **Feature Selection:** *SVM-based feature elimination strategy*
- **Prediction:** *Manifold Ranking*

Discriminative features for Manifold ranking:

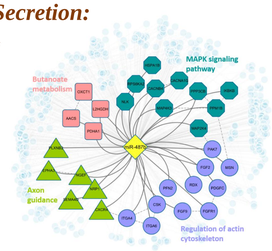
8 groups of discriminative features were selected to characterize human circulating miRNAs

Feature groups	*#*	*Feature list*
Frequency in seed region	28	AG, AGGU, C, CAGC, CAUC, CC, CCA, CCAG, CCAU, CCCA, CUUC, GA, GAG, GAGG, GCA, GCAG, GGU, GGUA, GU, GUA, GUAG, UA, UAG, UCC, etc.
Frequency in mature miRNA	63	ACG, ACGG, AG, AGC, AGCU, C, CAGU, CAUA, CC, CCG, CCGA, CG, CGA, CGAC, CGG, CGGA, GCAC, GCUC, GGG, GGUA, GGUU, GU, GUA, GUAG, GUU, etc.
Frequency in precursor sequence	80	ACCC, ACG, ACGA, ACGG, C, CACG, CAG, CAGG, CAGU, CC, CCA, CCG, CCGA, GCUC, GCUG, GGCC, GGCG, GGU, GGUA, GGUU, GU, GUA, GUAG, GUUG, etc.
3 nucleotides in stem loop structure	16	A(((, A((., A(.(, A.((, A.(., A..., C(((, C(.(, C.((, C..., G(((, G((., G(.(, G(.., G..., U(((
Structure indicators	14	MFE, NMFE, EFE, NEFE, freqMFEStructures, MFEI1, MFEI3, MFEI4, etc.
Stems/Pairs	12	%pairAU, %pairGC, %pairGU, max_stem_length, %G+C_stem, pairs, stems, etc
Percentage of nucleotides	4	%A+U_P, %A+U_m, %G+C content_P, %G+C content_m
Length/Palindromes	4	Length_m, length_P, palindromes_P, palindromes_seed

Manifold Ranking Prediction on the Likelihood of Secretion:

- 345 dietary microRNAs have been predicted as highly transportable candidates
- 117 of them have identical human homologs
- 9 cow-milk miRNAs have been validated in human plasma using microRNA-sequencing analysis
  - bta-miR-487b, miR-181b, and miR-421.

```
bta-miR-487b      1  aaucguacagggucauccacuu  22
                     ||||||||||||||||||||||
hsa-miR-487b-3p   1  aaucguacagggucauccacuu  22
```

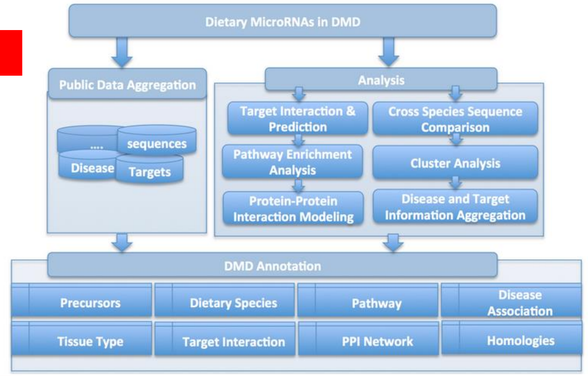


Dietary MicroRNA Databases (DMD)

The DMD can be accessed at http://sbbi.unl.edu/dmd/.

- 5,217 miRNAs from 15 dietary species
- Multiple resources of experimental validated targets and predicted targets
- 1,203 KEGG pathways are included and 87,694 distinct gene-pathway relationships
- 21,478 disease entries, of which some are associated with dietary microRNAs.

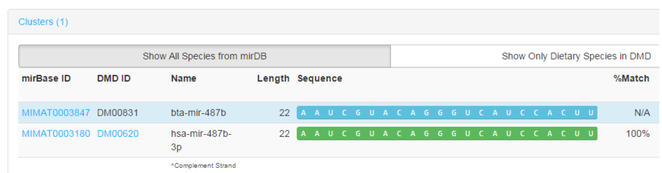
The following workflow diagram illustrates the process of data aggregation and analysis using cross-species sequence comparison and pathway enrichment analysis to produce the structured data that is presentable to the users of the database.

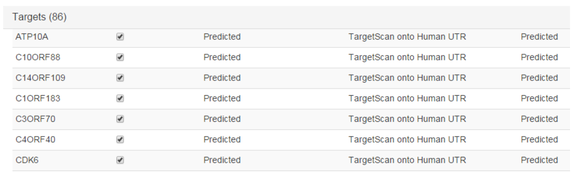


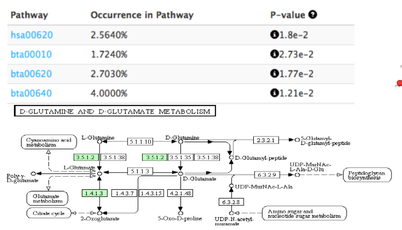
	Species (15)	*Mature (5,217)*	*Precursor (5,865)*
Plant	Apple	203	202
	Banana	360	180
	Corn	309	166
	Grape	180	157
	Orange	61	57
	Rice	634	526
	Soybean	620	554
	Tomato	96	68
	Wheat	111	108
Animal	Breast milk	434	402
	Cow's milk	243	245
	Cow's fat	205	229
	Atlantic Salmon	498	371
	Chicken	994	740
	Pig	411	382

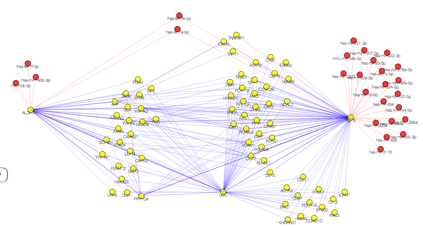
Analytic Tools:

- Cross species sequences comparison
- Targets identification
- Pathway enrichment analysis based on miRNA targets
- PPI networks
- Disease association





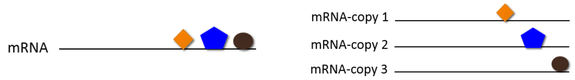




miRNA Regulatory Modules Identification

CLASH data:

- 18,514 high-confidence miRNA-mRNA interactions with the validated binding site



Analysis:

- Calculate the abundances of reported miRNAs and mRNAs
- Assess the statistical confidence of each miRNA-mRNA interaction to identify the functional/meaningful interactions from the random cases

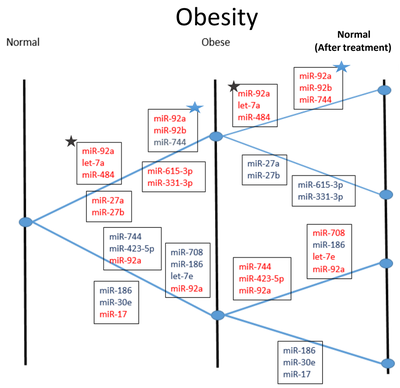
$$\binom{C}{C_{ij}}\left(\frac{m_i}{M}\frac{g_j}{G}\right)^{C_{ij}}\left(1-\frac{m_i}{M}\frac{g_j}{G}\right)^{C-C_{ij}}$$

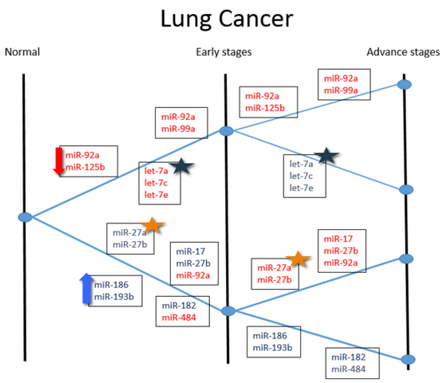
- Use bi-clustering to predict the miRNA regulatory modules on all high-confidence interactions

Results:

- By utilizing bi-clustering, we identified 173 miRNA regulatory modules:
  - 2~5 miRNAs (average: 3.0)
  - 5~26 mRNAs (average: 7.5)
- The modules cover 31 miRNAs and 378 mRNAs
  - the modules potentially functional related:
    - eg: miR-18a and miR-92a regulate genes that involved in Insulin pathway and immunity pathways
    - eg: miR-15a and miR-92a repress CCND2 that involved in Jak-STAT signaling pathway, which is associated with leptin signaling.

	ID	miR-18a	miR-92a	miR-15a
Primary immunodeficiency	ADA	1	1	0
	ATP1A2	1	1	0
	CDC25A	1	1	0
	CSDE1	1	1	0
	DDX50	1	1	0
	DIAPH2	1	1	0
	EEF1A2	1	1	0
	EEF2K	1	1	0
	ETNK1	1	1	0
	HERC2	1	1	0
Insulin Signaling Pathway	IRS2	1	1	0
	KIAA0141	1	1	0
	MSN	1	1	0
Insulin Signaling Pathway	PDK2	1	1	0
	POP1	1	1	0
	RPL7L1	1	1	0
	TNPO2	1	1	0
	TWF1	1	1	0
	ZC3H3	1	1	0
	ZNF100	1	1	0
	ZNF704	1	1	0
	BUB1	0	1	1
Jak-STAT signaling pathway	CCND2	0	1	1
	FASTK	0	1	1
	KIF1B	0	1	1
	NACA	0	1	1
	TRIM33	0	1	1





References & Acknowledgement

Baier, S. R., Nguyen, C., Xie, F., Wood, J. R., & Zempleni, J. (2014). MicroRNAs are absorbed in biologically meaningful amounts from nutritionally relevant doses of cow milk and affect gene expression in peripheral blood mononuclear cells, HEK-293 kidney cell cultures, and mouse livers. The Journal of nutrition, 144(10), 1495-1500.

Chiang K, Shu J, Zempleni J, Cui J (2015) Dietary MicroRNA Database (DMD): An Archive Database and Analytic Tool for Food-Borne microRNAs. PLoS ONE 10(6): e0128089. doi:10.1371/journal.pone.0128089

Helwak, A., Kudla, G., Dudnakova, T., & Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding.*Cell*, *153*(3), 654-665.

Shu, J.,Chiang, K., Zempleni, J. & Cui, J. Computational characterization of exogenous microRNAs transfer into human circulation.(Under review)

* This study receives funding from Nebraska Center for the Prevention of Obesity Diseases through Dietary Molecules (NPOD) through a grant from NIH's Center of Biomedical Research Excellence (NIH COBRE 1P20GM104320).