



Computational Characterization of Exogenous MicroRNAs Transfer into Human Circulation

Jiang Shu¹, Kevin Chiang¹, Janos Zemleni², Juan Cui¹

¹Dept. of Computer Science and Engineering; ²Dept. of Nutrition and Health Sciences; University of Nebraska-Lincoln, NE

Background

MicroRNAs have been long considered synthesized endogenously until very recent discoveries that human can absorb dietary microRNAs from foods of animal sources. Such compelling evidences, incorporated with controversial discoveries on plant-borne microRNAs, has created high interests in systematic research to study microRNA cross-species transfer and explore which and how exogenous microRNAs can be integrated into human circulation and possibly exert functions in humans. Here we present an integrated study where genomics analysis and computational modeling were applied to discover molecular features associated with the underlying mechanism of cross-species transportation.

Specially, we have comparatively analyzed a total of 34,612 microRNAs from 194 species in terms of sequence and structural features of these microRNA. 8 groups of discriminative features have been extracted to characterize human circulating microRNAs and therefore to base our computational inference of the likelihood of exogenous microRNAs getting integrated into human circulation. For example, 345 dietary microRNAs have been predicted as highly transportable candidates where 73 of them are exosome associated, e.g. bta-miR-487b, ssc-miR-29a, gga-miR-29a-5p, bta-miR-421 and bta-miR-29b which has been experimentally validated, and showed implications for health-related processes based on target and pathway analysis.

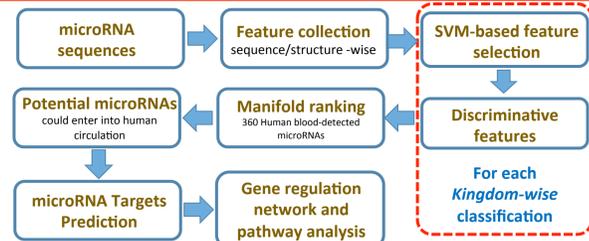
Datasets

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

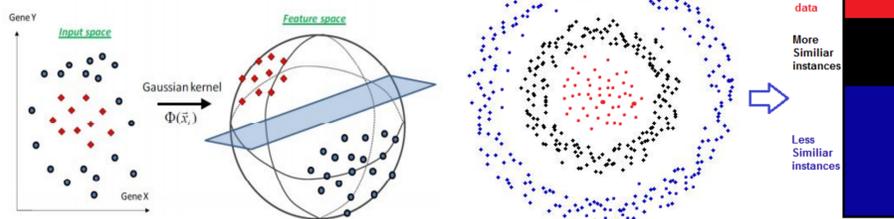
- 63,033 miRNA sequences from miRBase, categorized into 5 kingdoms
- 360 human blood miRNAs from literatures
- 370 exosomal miRNAs from ExoCarta exosome molecule database
- 5,217 dietary miRNAs from Dietary microRNA databases (DMD)

	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

Analytics workflow and methods



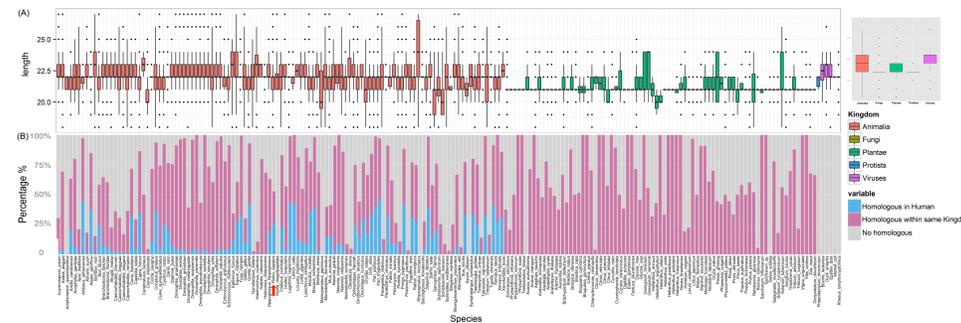
- Kingdom-wise Classification: Support vector machine (SVM)**
 - builds a model based on the labeled training instances
 - assigns new examples into one category or another
- Feature Selection: SVM-based feature elimination strategy**
 - recursive feature elimination removes most irrelevant features
 - obtain a minimal subset of features that produces a decent accuracy
- Prediction: Manifold Ranking**
 - the computational prediction model to detect the highly transportable miRNAs among all miRNAs.



Results

Sequence conservation analysis:

- 16,458 (~48%) microRNAs have their homologs in other species
- 903 (~35%) human microRNAs are identical with 5,662 microRNAs
 - From 96 *Animalia* species: cow (255), pig (205), chicken (103), etc.
 - none in *other kingdoms*



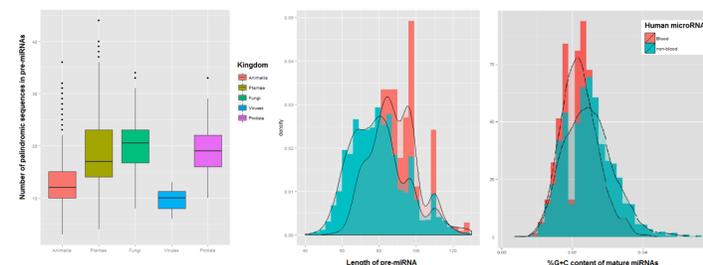
Features collection:

A total of 1,102 features were collected from two domains:

- Sequence features (1,031): mainly from statistical calculation on sequences
 - palindromes, %G+C content, %G+C content in seed region, lengths, di/triplet/four-nucleotide composition, etc.
- Structural features (71): based on structure prediction of pre-miRNAs
 - minimum free energy(MFE), Shannon Entropy (Q), Frobenius Norm (F), %pair, A(((, %GC-pair, etc.

Features assessment:

- Features show differential pattern among different kingdoms
 - number of palindromic sequence of pre-miRNA
- Features that may possibly recognize the blood-circulating miRNAs in Human
 - Length of pre-miRNA and %G+C content of pre-miRNA



Features selection:

	Selected Features	ACC	SEN	SPE	MCC
Animalia	166	93.29%	96.46%	90.11%	86.75%
Plantae	147	93.28%	89.71%	96.86%	86.79%
FPV	126	89.79%	87.39%	92.19%	79.68%

Discriminative features for Manifold ranking:

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

Feature groups	#	Feature list
Frequency in seed region	28	AG, AGGU, C, CAGC, CAUC, CC, CCA, CCG, 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, GGC, 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, GCUC, 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... 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:

- predicts the likelihood of an exogenous miRNAs getting integrated into human circulation.
- 221 overlapped discriminative features
- 360 known human blood miRNAs as true positive samples

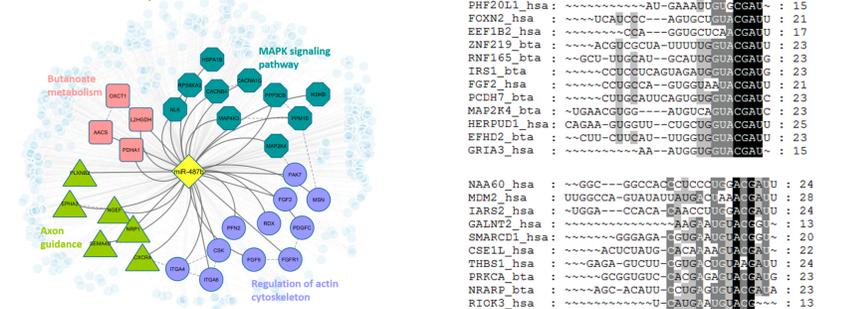
Results:

- ranking result shows that the highly transportable miRNAs are from *Animalia*, *Plantae* and *Viruses*
- from statistical perspective, *Animalia* and *Viruses* miRNAs have significantly higher rates of occurrence among the top predicted candidates when compared to the original dataset

	Animalia	Plantae	Viruses	Fungi	Protista
Original	26,705 (77.16%)	7,645 (22.09%)	152 (0.44%)	84 (0.24)	26 (0.08%)
Top-1K	962 (96.2%)	30 (3%)	8 (0.8%)	0	0
Top-3K	2,812 (93.7%)	163 (5.43%)	25 (0.87%)	0	0
Top-5K	4,678 (93.56%)	295 (5.9%)	27 (0.54%)	0	0
Top-10K	9,269 (92.69%)	670 (6.7%)	55 (0.55%)	4	2

Moreover, according to the result, we found 345 dietary miRNAs are likely to transfer into human blood, e.g. bta-miR-487b, ssc-miR-29a, gga-miR-29a-5p, bta-miR-421, etc.

- 254 of them are from 3 three *Animalia* species
 - cow(80), chicken(123), pig(91)
- 73 of 345 are exosome associated
 - e.g. bta-miR-487b, bta-miR-182, ssc-miR-125a, gg-miR-17
- bta-miR-487b as an example:
 - diet-born (cow's milk)
 - identical with human miRNA *has-miR-487*
 - could regulate human genes (464 targets)
 - could get involved in human pathways, e.g. MAPK Signal Pathway, Butanoate Metabolism, etc.



References & Acknowledgement

Shu, J., Chiang, K., Zemleni, J. & Cui, J. Computational characterization of exogenous microRNAs transfer into human circulation. (Submitted to NAR)

Baier, S. R., Nguyen, C., Xie, F., Wood, J. R., & Zemleni, 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.

Chen, X., Gao, C., Li, H., Huang, L., Sun, Q., Dong, Y., ... & Zhang, C. Y. (2010). Identification and characterization of microRNAs in raw milk during different periods of lactation, commercial fluid, and powdered milk products. *Cell research*, 20(10), 1128-1137.

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).