



Identification and Analysis of microRNA-Disease Associations with Kernelized Bayesian Matrix Factorization

Ahmet Toprak^{1*}, Esmâ Eryılmaz Doğan²

^{1*} Selçuk Üniversitesi, Elektrik ve Enerji Bölümü, Konya, Türkiye, (ORCID: 0000-0003-3337-4917), ahmettoprak42@gmail.com

² Selçuk Üniversitesi, Teknoloji Fakültesi, Biyomedikal Mühendisliği, Konya, Türkiye (ORCID: 0000-0001-6809-7513), eeryilmaz@selcuk.edu.tr

(1st International Conference on Applied Engineering and Natural Sciences ICAENS 2021, November 1-3, 2021)

(DOI: 10.31590/ejosat.980257)

ATIF/REFERENCE: Toprak, A. & Eryılmaz Doğan, E. (2021). Identification and Analysis of microRNA-Disease Associations with Kernelized Bayesian Matrix Factorization. *European Journal of Science and Technology*, (28), 40-45.

Abstract

MicroRNA (miRNA) molecules, which are effective on the initiation and progression of many different diseases, are a type of non-coding RNA with a length of about 22 nucleotides. Scientists have reported the importance of miRNAs in the prevention, diagnosis, and treatment of complex human diseases. Therefore, in the last decade, researchers have been working hard to find potential miRNA-disease associations. Many computational techniques have been developed because of the experimental techniques are time-consuming and expensive used to find new relationships between miRNAs and diseases. In this study, we suggested Kernelized Bayesian matrix factorization (KBMF) technique to predict new miRNA-disease relationships. We applied 5-fold cross validation technique and obtained an average value AUC of 0.9450. Also, we applied case studies based on breast, lung, and colon neoplasms to prove the performance of KBMF technique. The results showed that KBMF can be used as a reliable computational model to reveal possible miRNA-disease relationships.

Keywords: miRNA, disease, miRNA-Disease Association, similarity measure.

Kernelized Bayesian Matris Faktörizasyonu ile mikroRNA-Hastalık İlişkilerinin Tanımlanması ve Analizi

Öz

Birçok farklı hastalığın başlamasında ve ilerlemesinde etkili olan mikroRNA (miRNA) molekülleri yaklaşık 22 nükleotid uzunluğunda kodla yapmayan bir RNA türüdür. Bilim insanları karmaşık insan hastalıklarının önlenmesi, teşhisi ve tedavisinde miRNA'ların önemini açıklamıştır. Bu nedenle son yıllarda araştırmacılar potansiyel miRNA-hastalık ilişkilerini bulmak için çok çalışmaktalar. miRNA'lar ve hastalıklar arasında yeni ilişkiler bulmak için kullanılan deneysel tekniklerin zaman alıcı ve pahalı olması nedeniyle birçok hesaplama tekniği geliştirilmiştir. Bu çalışmada yeni miRNA-hastalık ilişkilerini tahmin etmek için Kernelized Bayesian matrix factorization (KBMF) tekniğini önerdik. 5-katlı çapraz doğrulama tekniği uyguladık ve 0.9450 ortalama AUC değeri elde ettik. Ayrıca KBMF tekniğinin performansını kanıtlamak için meme, akciğer ve kolon neoplazmalarına dayalı vaka çalışmaları uyguladık. Sonuçlar KBMF'nin olası miRNA-hastalık ilişkilerini ortaya çıkarmak için güvenilir bir hesaplama modeli olarak kullanılabileceğini gösterdi.

Anahtar Kelimeler: miRNA, hastalık, miRNA-hastalık ilişkisi, benzerlik ölçümü.

* Sorumlu Yazar: ahmettoprak42@gmail.com

1. Introduction

MicroRNAs are a type of non-coding RNAs with a length of about 22 nucleotides. Several research studies have concluded that miRNAs have important functions in various basic biological processes such as cell development, signal transduction, proliferation, apoptosis, differentiation, viral infection, metabolism, and aging (Bartel, 2009; Chen, Zhou, & Zhao, 2018; Lan et al., 2018; Tang, Zhou, Zheng, Zhang, & Sha, 2019). With the development of molecular biology and biotechnology, researchers have revealed that miRNAs have an important links with many diseases (X. Chen et al., 2016; Kim, 2015). For example, miRNA de-regulation triggers the development of various cancers such as breast, skin, lung, colon, prostate etc. Furthermore, down-regulations of miRNA-143 and miRNA-145 have been observed in colorectal tumors and especially breast cancer (Espinosa & Slack, 2006).

Various types of databases such as miRBase (Kozomara & Griffiths-Jones, 2013), HMDD (Y. Li et al., 2014), miR2Disease (Q. Jiang et al., 2009), miRGen (Alexiou et al., 2009), deepBase (J.-H. Yang, Shao, Zhou, Chen, & Qu, 2009), and dbDEMC (Z. Yang et al., 2010) have been developed to store miRNA related data. These databases include human miRNA-disease relationships and provide differentially expressed miRNAs for cancers.

Computational techniques are used to estimate potential miRNA-disease relationships, as determination of miRNAs-disease relationships by biological experimental techniques is time consuming and very expensive (Mugunga, Ju, Liu, & Huang, 2017). In recent years, many new computational techniques such as WBSMDA (X. Chen et al., 2016), RKNMMDA (Chen, Wu, & Yan, 2017), BNPMMDA (X. Chen, D. Xie, et al., 2018), EGBMMDA (Chen, Huang, Xie, & Zhao, 2018), LRSSLMDA (Chen & Huang, 2017), NDAMDA (Chen, Wang, & Huang, 2018), MCMMDA (J.-Q. Li, Rong, Chen, Yan, & You, 2017), NSEMMDA (C. C. Wang, Chen, Yin, & Qu, 2019), HDMP (Xuan et al., 2013), MaxFlow (Yu, Chen, & Lu, 2017), and SACMDA (Shao, Liu, & Yan, 2018) have been developed by scientists to predict possible miRNA-disease relationships. Understanding the complex disease mechanism at the molecular level of miRNA may be possible by predicting new miRNA-disease relationships. Thus, the diagnosis, treatment, prognosis, and prevention of diseases can be provided with its contribution to the field of personalized medicine (Chen, 2015; Chen, Huang, Wang, You, & Chan, 2016).

In this study, first of all we calculated the functional similarity (*FS*) for each miRNA and the semantic similarity (*SS*) for each disease. Secondly, we calculated the Gaussian Interaction Profile (GIP) kernel similarities for both miRNA and disease. Thirdly, we integrated the miRNA GIP kernel similarity with *FS* and the disease GIP kernel similarity with *SS*. Lastly, we estimated the potential relationships between miRNAs and diseases by analyzing these data with the Kernelized Bayesian Matrix Factorization (KBMF) method. We evaluated the success of our model with the most widely used 5-fold cross-validation technique and several case studies.

2. Material and Method

We obtained 495 miRNAs and 383 diseases data set from the HMDD v2.0. This data set includes 5430 miRNA-disease

relationships that were experimentally verified (Y. Li et al., 2014). The relationships between miRNAs and diseases are represented by constructing an adjacency matrix ($A_{495 \times 383}$) matrix.

2.1. miRNA Functional Similarity (*FS*) and Disease Semantic Similarity (*SS*)

The functional similarity scores calculated using the method proposed by Wang et al. (D. Wang, Wang, Lu, Song, & Cui, 2010) were obtained from the web address <http://www.cuilab.cn/files/images/cuilab/misim.zip>.

To calculate disease semantic similarity, we first downloaded the Medical Subject Headings (MeSH) definitions from the web page of National Library of Medicine (<http://www.nlm.nih.gov>). The connections between various diseases can be explained by using Directed Acyclic Graph (DAG). In DAG_A , when calculating the semantic value of disease A , the contribution of other diseases can be expressed as given below:

$$\begin{cases} D_A(A) = 1 \\ D_A(t) = \max\{\Delta * D_A(t') | t' \in \text{children of } t\} \quad \text{if } t \neq A \end{cases} \quad (1)$$

where Δ represents the contribution coefficient between disease t with its child disease t' . Moreover, the semantic value (*DV*) of disease A can be described with Eq. 2.

$$DV(A) = \sum_{t \in T_A} D_A(t) \quad (2)$$

The *SS* between disease A and disease B can be computed as shown below:

$$SS(A, B) = \frac{\sum_{t \in T_A \cap T_B} (D_A(t) + D_B(t))}{DV(A) + DV(B)} \quad (3)$$

where all ancestor of disease A and disease B , including disease A and disease B themselves, are represented by T_A and T_B . $DA(t)$ and $DB(t)$ represent the semantic value of disease t associated with disease A and disease B , respectively. Semantic similarity of each disease is calculated according to the Eq. 3 (D. Wang et al., 2010).

2.2. Gaussian Interaction Profile (GIP) Kernel Similarity for miRNAs and Diseases

Assuming that similar diseases tend to be associated with miRNAs with similar functions, we have created miRNA GIP kernel similarity and disease GIP kernel similarity (Lan et al., 2018; van Laarhoven, Nabuurs, & Marchiori, 2011). The GIP kernel similarity value (*GM*) between miRNA $m(i)$ and miRNA $m(j)$ can be computed with Eq. 4.

$$GM(m_i, m_j) = \exp\left(-\gamma_m \|IP(m_i) - IP(m_j)\|^2\right) \quad (4)$$

Similarly, the GIP kernel similarity value (*GD*) between disease $d(i)$ and disease $d(j)$ can be computed with Eq. 5.

$$GD(d_i, d_j) = \exp\left(-\gamma_d \|IP(d_i) - IP(d_j)\|^2\right) \quad (5)$$

Here, γ_m and γ_d parameters control kernel bandwidth and can be obtained from the Eq. 6 and Eq. 7.

$$\gamma_m = \frac{\delta_m}{\frac{1}{n_m} \sum_{i=1}^{n_m} \|IP(m_i)\|^2} \quad (6)$$

$$\gamma_d = \frac{\delta_d}{\frac{1}{n_d} \sum_{i=1}^{n_d} \|IP(d_i)\|^2} \quad (7)$$

The parameters δ_m and δ_d and represent the new bandwidth parameters and were adjusted to 1 according to (van Laarhoven et

al., 2011) for simplicity. In addition, all miRNA numbers and all disease numbers were indicated by n_m and n_d , respectively.

2.3. Integrated Similarity for miRNAs and Diseases

We have integrated the functional similarity and the GIP kernel similarity of miRNAs using Eq. 8.

$$SM = \beta xGM + (1 - \beta)xFS \quad (8)$$

Similarly, we have combined the semantic similarity and the GIP kernel similarity of diseases with Eq. 9.

$$SD = \beta xGD + (1 - \beta)xSS \quad (9)$$

where, β was assumed to be 0.5.

2.4. Kernelized Bayesian Matrix Factorization (KBMF)

The KBMF described in detail in the study reported by Gönen et al. (Gönen, Khan, & Kaski, 2013) is an effective way to obtain a bipartite graph by multiple data source integration. We assume that miRNAs and diseases come from two domains: $X = \{m_1, m_2, \dots, m_{N_x}\}$ and $Z = \{d_1, d_2, \dots, d_{N_z}\}$, respectively. In order to calculate potential interaction between miRNAs and diseases, we have multiple kernel matrix, namely $\{K_{X,m}: X \times X \rightarrow \mathbb{R}\}_{m=1}^{P_x}$ and $\{K_{Z,n}: Z \times Z \rightarrow \mathbb{R}\}_{n=1}^{P_z}$ by calculating miRNA similarity domain and disease similarity domain, respectively. The number of miRNA kernel matrix and the number of disease kernel matrix have given with P_x and P_z . Adjacency matrix $A \in \{0,1\}^{N_x \times N_z}$ shows known miRNAs-diseases interactions (Ammad-Ud-Din et al., 2014; Gönen et al., 2013).

$$A_j^i = \begin{cases} +1 & \text{if } x_i \text{ and } z_j \text{ are interacting} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

3. Results

3.1. 5-fold Cross Validation (CV) Technique

We tested the predictive ability of the KBMF method we used in this study with a 5-fold cross validation technique. In this validation technique, we divided all known miRNA-disease relationships into five subgroups. For testing of the model four of the 5 subgroups were used as training data and one as test data. Additionally, we have computed false positive rate (FPR) and true positive rate (TPR) and have plotted the receiver operator characteristics (ROC) curve according to the results, then have computed the area under the ROC curve (AUC) for performance evaluations.

As a result, we plotted the ROC curve shown in Fig. 1, and computed AUC value of 0.9450 for 5-fold cross-validation. We compared the results with the other eight techniques of WBSMDA (X. Chen et al., 2016), MCMDA (J.-Q. Li et al., 2017), RKNNMDA (Chen et al., 2017), NSEMDA (C. C. Wang et al., 2019), BNPMDA (X. Chen, D. Xie, et al., 2018), EGBMMDA (X. Chen, L. Huang, et al., 2018), LRSSLMDA (Chen & Huang, 2017), and NDAMDA (X. Chen, L. Y. Wang, et al., 2018) to prove the performance of KBMF method we used. The other eight methods, using 5-fold cross-validation technique, had AUC values of 0.8185, 0.8767, 0.6723, 0.8878, 0.8980, 0.9048, 0.9181, and 0.8935, respectively. Figure 2 shows the comparative AUC

values. Here, it is clearly seen that KBMF method gives a better result than the other eight compared methods.

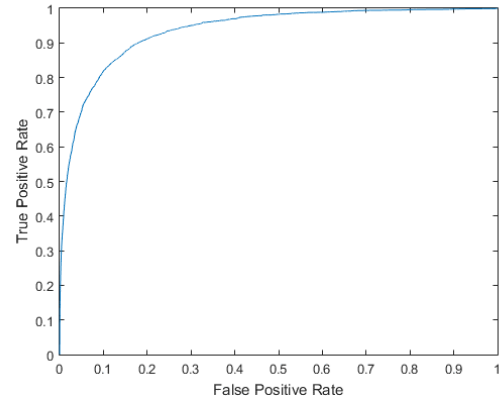


Figure 1. AUC value of KBMF

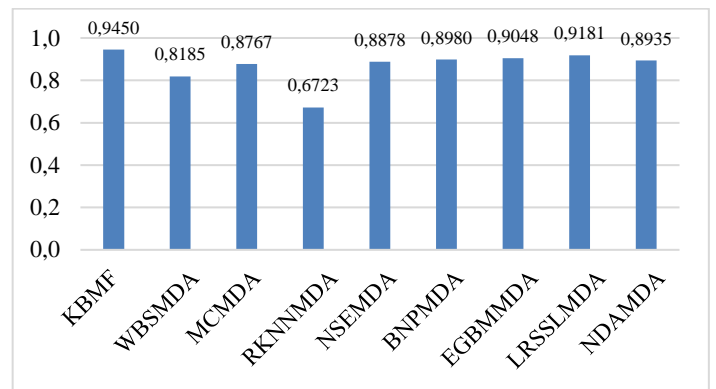


Figure 2. Comparison of the AUC values of KBMF with the other eight methods

3.2. Case Studies

In order to predictive accuracy demonstration of the KBMF method, three case studies have been conducted based on breast, lung, and colon neoplasms from databases of miR2Disease, dbDEMOC, and HMDD v2.0. In our model, we used 5430 known miRNA-disease relationships from HMDD v2.0 as a training set. In this training set, we made the known relationships for each disease to zero. Based on the results obtained, all miRNAs for three diseases were ranked according to their scores. Then, the first 20 miRNAs predicted for each disease were confirmed from three different databases mentioned above.

Breast neoplasms, which cause many deaths each year, are the most common type of female cancer among female cancers and comprise approximately 22% of female cancers (Al-Hajj, Wicha, Benito-Hernandez, Morrison, & Clarke, 2003; Qinghua Jiang et al., 2010). The list of the top 20 candidate miRNAs predicted by our method for breast neoplasm is shown in Table 1. We have confirmed all 20 miRNAs are 100% associated with breast cancer from the mentioned databases.

Lung neoplasms are one of the main factors in cancer-related deaths in the world. We have listed the top 20 candidate miRNAs for lung neoplasm using the same method (shown in Table 2). Examining the table, it can be seen that 19 out of 20 candidate miRNAs were confirmed to be associated with lung cancer. In other words, we can say that 95% of candidate miRNAs are associated with lung cancer.

Table 1. Prediction results of the top 20 predicted miRNAs related with breast neoplasm

Disease	miRNA	Database
Breast Neoplasms	hsa-let-7a	dbDEMC,miR2Disease,HMDD
	hsa-let-7b	dbDEMC,HMDD
	hsa-let-7c	dbDEMC,HMDD
	hsa-let-7d	dbDEMC,miR2Disease,HMDD
	hsa-mir-200c	dbDEMC,miR2Disease,HMDD
	hsa-let-7e	dbDEMC,HMDD
	hsa-let-7f	dbDEMC,miR2Disease,HMDD
	hsa-let-7g	dbDEMC,HMDD
	hsa-mir-200b	dbDEMC,miR2Disease,HMDD
	hsa-let-7i	dbDEMC,miR2Disease,HMDD
	hsa-mir-200a	miR2Disease,HMDD
	hsa-mir-126	miR2Disease,HMDD
	hsa-mir-145	miR2Disease,HMDD
	hsa-mir-34c	dbDEMC,HMDD
	hsa-mir-1	dbDEMC
	hsa-mir-141	dbDEMC,miR2Disease,HMDD
	hsa-mir-218	HMDD
	hsa-mir-100	dbDEMC,HMDD
	hsa-mir-34b	HMDD
	hsa-mir-221	miR2Disease,HMDD

Disease	miRNA	Database
Colon Neoplasms	hsa-mir-155	dbDEMC,miR2Disease
	hsa-mir-17	HMDD
	hsa-mir-20a	dbDEMC,miR2Disease
	hsa-let-7a	dbDEMC,miR2Disease
	hsa-let-7b	dbDEMC,miR2Disease
	hsa-let-7d	dbDEMC
	hsa-mir-18a	dbDEMC,miR2Disease
	hsa-let-7c	dbDEMC
	hsa-let-7e	dbDEMC
	hsa-mir-19b	dbDEMC,miR2Disease
	hsa-let-7f	miR2Disease
	hsa-mir-92a	dbDEMC
	hsa-let-7i	<i>unconfirmed</i>
	hsa-mir-19a	dbDEMC,miR2Disease
	hsa-let-7g	miR2Disease
	hsa-mir-21	dbDEMC,miR2Disease
	hsa-mir-146a	<i>unconfirmed</i>
	hsa-mir-221	dbDEMC,miR2Disease
	hsa-mir-15a	<i>unconfirmed</i>
	hsa-mir-125b	dbDEMC

Table 2. Prediction results of the top 20 predicted miRNAs related with lung neoplasm

Disease	miRNA	Database
Lung Neoplasms	hsa-mir-200a	dbDEMC,miR2Disease,HMDD
	hsa-mir-200b	dbDEMC,miR2Disease,HMDD
	hsa-mir-200c	dbDEMC,miR2Disease,HMDD
	hsa-mir-141	dbDEMC,miR2Disease
	hsa-mir-429	dbDEMC,miR2Disease
	hsa-mir-199a	dbDEMC,miR2Disease,HMDD
	hsa-mir-34b	dbDEMC,HMDD
	hsa-mir-183	dbDEMC,miR2Disease
	hsa-mir-375	HMDD
	hsa-mir-1	dbDEMC,miR2Disease,HMDD
	hsa-mir-21	miR2Disease,HMDD
	hsa-mir-100	HMDD
	hsa-mir-182	dbDEMC,miR2Disease,HMDD
	hsa-mir-196a	HMDD
	hsa-mir-133a	HMDD
	hsa-mir-126	miR2Disease,HMDD
	hsa-mir-31	miR2Disease,HMDD
	hsa-mir-34c	dbDEMC,HMDD
	hsa-mir-194	<i>unconfirmed</i>
	hsa-mir-143	miR2Disease,HMDD

Colon neoplasm, one of the greatest threats to human life, is the most common type of cancer in the world. About half of colon cancer patients die within 5 years after diagnosis (Drusco et al., 2014; Ogata-Kawata et al., 2014; Phipps et al., 2013; Torre et al., 2015). In this research, the first 20 miRNAs estimated by KBMF method are listed in Table 3. It can see that 17 of the 20 predicted colon cancer associations have been confirmed from the databases mentioned above. We can say that the predicted success rate for colon cancer is 85%.

Table 3. Prediction results of the top 20 predicted miRNAs related with colon neoplasm

4. Conclusions and Recommendations

Related research studies have shown that miRNAs play a major role in various biological processes (Bartel, 2009; Xing Chen et al., 2018; Lan et al., 2018; Tang et al., 2019). Therefore, it is important to determine miRNA-disease relationships by computational techniques before applying costly and time-consuming experimental techniques. In this study, we used the Kernelized Bayesian matrix factorization method to predict possible relationships between miRNAs and diseases. In our previous study (Toprak & Eryilmaz, 2020), high quality results were obtained by using similar approaches with a different method. We evaluated the predictive performance of our model with 5-fold cross validation technique and several case studies. The calculated AUC value of the 5-fold cross validation technique is 0.9450. Also, we conducted three case studies like breast, lung, and colon neoplasms to further verify the performance of KBMF and validated the results with dbDEMC, miR2Disease and HMDD databases. The results indicates that the KBMF method can be used as an efficient method to predict possible relationships between miRNAs and diseases.

References

Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences*, 100(7), 3983-3988. doi:10.1073/pnas.0530291100

Alexiou, P., Vergoulis, T., Gleditzsch, M., Prekas, G., Dalamagas, T., Megraw, M., . . . Hatzigeorgiou, A. G. (2009). miRGen 2.0: a database of microRNA genomic information and regulation. *Nucleic Acids Research*, 38(suppl_1), D137-D141. doi:10.1093/nar/gkp888

Ammad-Ud-Din, M., Georgii, E., Gonen, M., Laitinen, T., Kallioniemi, O., Wennerberg, K., . . . Kaski, S. (2014). Integrative and personalized QSAR analysis in cancer by kernelized Bayesian matrix factorization. *Journal of chemical*

- information and modeling, 54(8), 2347-2359. doi:10.1021/ci500152b
- Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *cell*, 136(2), 215-233. doi:10.1016/j.cell.2009.01.002
- Chen, X. (2015). KATZLDA: KATZ measure for the lncRNA-disease association prediction. *Scientific reports*, 5, 16840. doi:10.1038/srep16840
- Chen, X., & Huang, L. (2017). LRSSLMDA: Laplacian Regularized Sparse Subspace Learning for MiRNA-Disease Association prediction. *PLoS Computational Biology*, 13(12), e1005912. doi:10.1371/journal.pcbi.1005912
- Chen, X., Huang, L., Xie, D., & Zhao, Q. (2018). EGBMMDA: Extreme Gradient Boosting Machine for MiRNA-Disease Association prediction. *Cell Death & Disease*, 9(1), 3. doi:10.1038/s41419-017-0003-x
- Chen, X., Huang, Y.-A., Wang, X.-S., You, Z.-H., & Chan, K. C. (2016). FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model. *Oncotarget*, 7(29), 45948. doi:10.18632/oncotarget.10008
- Chen, X., Wang, L. Y., & Huang, L. (2018). NDAMDA: Network distance analysis for MiRNA-disease association prediction. *Journal of Cellular and Molecular Medicine*, 22(5), 2884-2895. doi:10.1111/jcmm.13583
- Chen, X., Wu, Q. F., & Yan, G. Y. (2017). RKNMMDA: Ranking-based KNN for MiRNA-Disease Association prediction. *RNA biology*, 14(7), 952-962. doi:10.1080/15476286.2017.1312226
- Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z. H., & Liu, H. (2018). BNPMDA: Bipartite Network Projection for MiRNA-Disease Association prediction. *Bioinformatics*, 34(18), 3178-3186. doi:10.1093/bioinformatics/bty333
- Chen, X., Yan, C. C., Zhang, X., You, Z. H., Deng, L., Liu, Y., . . . Dai, Q. (2016). WBSMDA: Within and Between Score for MiRNA-Disease Association prediction. *Scientific reports*, 6, 21106. doi:10.1038/srep21106
- Chen, X., Zhou, Z., & Zhao, Y. (2018). ELLPMDA: ensemble learning and link prediction for miRNA-disease association prediction. *RNA biology*, 15(6), 807-818. doi:10.1080/15476286.2018.1460016
- Drusco, A., Nuovo, G. J., Zanesi, N., Di Leva, G., Pichiorri, F., Volinia, S., . . . Bottoni, A. (2014). MicroRNA profiles discriminate among colon cancer metastasis. *PLoS One*, 9(6), e96670. doi:10.1371/journal.pone.0096670
- Espinosa, C. E. S., & Slack, F. J. (2006). Cancer issue: the role of microRNAs in cancer. *The Yale journal of biology and medicine*, 79(3-4), 131-140.
- Gönen, M., Khan, S., & Kaski, S. (2013). Kernelized Bayesian matrix factorization. Paper presented at the International Conference on Machine Learning.
- Jiang, Q., Hao, Y., Wang, G., Juan, L., Zhang, T., Teng, M., . . . Wang, Y. (2010). Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Systems Biology*, 4(1), S2. doi:10.1186/1752-0509-4-s1-s2
- Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., . . . Liu, Y. (2009). miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*, 37(Database issue), D98-104. doi:10.1093/nar/gkn714
- Kim, Y.-K. (2015). Extracellular microRNAs as biomarkers in human disease. *Chonnam medical journal*, 51(2), 51-57. doi:10.4068/cmj.2015.51.2.51
- Kozomara, A., & Griffiths-Jones, S. (2013). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, 42(D1), D68-D73. doi:10.1093/nar/gkt1181
- Lan, W., Wang, J., Li, M., Liu, J., Wu, F. X., & Pan, Y. (2018). Predicting MicroRNA-Disease Associations Based on Improved MicroRNA and Disease Similarities. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(6), 1774-1782. doi:10.1109/TCBB.2016.2586190
- Li, J.-Q., Rong, Z.-H., Chen, X., Yan, G.-Y., & You, Z.-H. (2017). MCMDA: Matrix completion for MiRNA-disease association prediction. *Oncotarget*, 8(13), 21187. doi:10.18632/oncotarget.15061
- Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T., & Cui, Q. (2014). HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Research*, 42(Database issue), D1070-D1074. doi:10.1093/nar/gkt1023
- Mugunga, I., Ju, Y., Liu, X., & Huang, X. (2017). Computational prediction of human disease-related microRNAs by path-based random walk. *Oncotarget*, 8(35), 58526. doi:10.18632/oncotarget.17226
- Ogata-Kawata, H., Izumiya, M., Kurioka, D., Honma, Y., Yamada, Y., Furuta, K., . . . Sonoda, H. (2014). Circulating exosomal microRNAs as biomarkers of colon cancer. *PLoS One*, 9(4), e92921. doi:10.1371/journal.pone.0092921
- Phipps, A. I., Lindor, N. M., Jenkins, M. A., Baron, J. A., Win, A. K., Gallinger, S., . . . Newcomb, P. A. (2013). Colon and rectal cancer survival by tumor location and microsatellite instability: the Colon Cancer Family Registry. *Diseases of the colon and rectum*, 56(8), 937. doi:10.1097/DCR.0b013e31828f9a57
- Shao, B., Liu, B., & Yan, C. (2018). SACMDA: MiRNA-Disease Association Prediction with Short Acyclic Connections in Heterogeneous Graph. *Neuroinformatics*, 16(3-4), 373-382. doi:10.1007/s12021-018-9373-1
- Tang, C., Zhou, H., Zheng, X., Zhang, Y., & Sha, X. (2019). Dual Laplacian regularized matrix completion for microRNA-disease associations prediction. *RNA biology*, 16(5), 601-611. doi:10.1080/15476286.2019.1570811
- Toprak, A., & Eryilmaz, E. (2020). Prediction of miRNA-disease associations based on Weighted K-Nearest known neighbors and network consistency projection. *Journal of Bioinformatics and Computational Biology*, 18(6), 2050041. doi:10.1142/s0219720020500419
- Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. *CA: a cancer journal for clinicians*, 65(2), 87-108. doi:10.3322/caac.21262
- van Laarhoven, T., Nabuurs, S. B., & Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics*, 27(21), 3036-3043. doi:10.1093/bioinformatics/btr500
- Wang, C. C., Chen, X., Yin, J., & Qu, J. (2019). An integrated framework for the identification of potential miRNA-disease association based on novel negative samples extraction strategy. *RNA biology*, 16(3), 257-269. doi:10.1080/15476286.2019.1568820
- Wang, D., Wang, J., Lu, M., Song, F., & Cui, Q. (2010). Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics*, 26(13), 1644-1650. doi:10.1093/bioinformatics/btq241
- Xuan, P., Han, K., Guo, M., Guo, Y., Li, J., Ding, J., . . . Huang, Y. (2013). Prediction of microRNAs associated with human

- diseases based on weighted k most similar neighbors. *PLoS One*, 8(8), e70204. doi:10.1371/journal.pone.0070204
- Yang, J.-H., Shao, P., Zhou, H., Chen, Y.-Q., & Qu, L.-H. (2009). deepBase: a database for deeply annotating and mining deep sequencing data. *Nucleic Acids Research*, 38(suppl_1), D123-D130. doi:10.1093/nar/gkp943
- Yang, Z., Ren, F., Liu, C., He, S., Sun, G., Gao, Q., . . . Zhao, H. (2010). dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genomics*, 11 Suppl 4, S5. doi:10.1186/1471-2164-11-S4-S5
- Yu, H., Chen, X., & Lu, L. (2017). Large-scale prediction of microRNA-disease associations by combinatorial prioritization algorithm. *Scientific reports*, 7, 43792. doi:10.1038/srep43792