**A novel computational method for the identification of potential miRNA-disease association based on symmetric non-negative matrix factorization and Kronecker regularized least square**

**Yan Zhao1, Xing Chen1,\*, Jun Yin1**

1School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, 221116, China

**\* Correspondence:**Xing Chen

xingchen@amss.ac.cn

**Keywords: microRNA; disease; association prediction; matrix factorization; Kronecker regularized least square**

# Supplementary Table

**Supplementary Table 1.** We applied SNMFMDA to prioritize all the candidate miRNA-disease pairs based on all the known miRNA-disease associations in HMDD database as training samples. This prediction result is released for further experimental validation and research.

# The calculation of miRNA functional similarity

In order to calculate the functional similarity between the two miRNAs, it is necessary to consider the diseases associated with these two miRNAs. We need to define the semantic similarity between disease and disease group at first. Here, we use $d$ to indicate one disease and $D=\{d(1),d(2),…d(t)\}$ to represent one disease group containing $t$ diseases. We define the semantic value between the disease $d$ and disease groups $D$ as the maximum value of the semantic similarity between the disease $d$ and all diseases in the disease group $D$. Therefore, the semantic similarity $S(d,D)$ between disease $d$ and disease groups $D$ is calculated as follows:

$$S\left(d,D\right)=\max\_{1\leq i\leq m}S(d,d(i))$$

For miRNA $m(1)$, we define the disease group $D1=\{d\_{1}(1),d\_{1}(2),…d\_{1}(m)\}$ whose element $d\_{1}(i)$ was associated with miRNA $m(1)$. Similarly, we define another disease group $D2=\{d\_{2}(1),d\_{2}(2),…d\_{2}(n)\}$ for miRNA $m(2)$, and there is known association between disease $d\_{2}(j)$ and miRNA $m(2)$. The functional similarity between $m(1)$ and $m(2)$ could be calculated as follow:

$$FS\left(m\left(1\right),m\left(2\right)\right)=\frac{\sum\_{1\leq i\leq m}^{}S(d\_{1}\left(i\right),D2)+\sum\_{1\leq j\leq n}^{}S(D1,d\_{2}(j))}{m+n}$$