Bio-Medical Materials and Engineering 26 (2015) S1903–S1915
DOI 10.3233/BME-151493
IOS Press


# Integration of multi-microarray datasets to identify chronic obstructive pulmonary disease-related miRNAs[1]

Lin Hua[a,b,*], Weiying Zheng [a,b], Hong Xia[a,b], Ping Zhou[a,b] and Li An[c]
[a]*School of Biomedical Engineering, Capital Medical University, Beijing, 100069, China*
[b]*Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Capital Medical University, Beijing, 100069, China*
[c]*Beijing Institute of Respiratory Medicine, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, 100028, China*

**Abstract.** Currently, the mechanisms underlying chronic obstructive pulmonary disease (COPD) remain unclear. As potential biomarkers, microRNAs (miRNAs), which modulate the levels of specific genes and proteins, are important for enhancing our understanding of the mechanisms behind COPD. Although there have been a number of miRNA expression profiling analyses strategies used to document miRNA expression changes during physiological and pathological processes or used to identify differentially expressed miRNAs in disease or control samples, the study results have been inconsistently replicated using different datasets. For this reason, many findings cannot be well synthesized and interpreted. To address this issue, we used a multiple co-inertia analysis (MCIA) method to extract potential COPD-related miRNAs using three COPD microarray datasets. The results showed that miR-223, miR-132, and miR-199a-5p are obviously associated with COPD, and these results are consistent with the highly significant differentially-expressed miRNAs that were observed across three microarray datasets. Moreover, when miR-223, miR-132, and miR-199a-5p are taken as predictors to classify the samples of three datasets, the pooled sensitivity and specificity is 0.96 and 0.75, respectively, thereby suggesting that these three miRNAs can effectively distinguish COPD patients and controls.

Keywords: Chronic obstructive pulmonary disease, multiple co-inertia, miRNAs, differentially expressed

## 1. Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a highly prevalent condition that is currently a major cause of death in developed countries. Furthermore, the disease also imposes a substantial economic burden on healthcare systems as patients may require long hospital stays and/or costly therapeutic interventions for acute exacerbations [1]. The development of COPD is often associated with tobacco smoke, and smoking tobacco continues to be a main risk factor; however, most heavy

[1] Lin Hua, Weiying Zheng and Hong Xia equally contributed to this work.

[*] Address for correspondence: Lin Hua, School of Biomedical Engineering, Capital Medical University, No. 10 Xitoutiao, Youanmen, Fengtai District, Beijing, 100069, China. Tel. +8610 83911567; Fax: 86-10-83911552; E-mail: hualin7750@139.com.

smokers do not develop clinically significant COPD [2]. Therefore, there may be additional factors that determine COPD susceptibility, and further studies are needed to in order to elucidate the mechanisms underlying COPD.

As important biomarkers, microRNAs (miRNAs) that modulate the levels of specific genes and proteins are important in enhancing our understanding of the mechanisms of complex diseases, such as COPD. In the last several years, a number of miRNA profiling strategies have been used to document miRNA expression changes during physiological and pathological processes or have been used to identify differentially expressed miRNAs in disease or control samples [3]. Given the role of miRNAs in lung inflammation and the lung's response to inhaled toxins, miRNAs are likely to be involved in the pathogenesis and progression of COPD. At present, seventy miRNAs have been found to be differentially expressed in lung whole tissue between smoking patients affected by COPD and smokers not affected by COPD [4]. However, many of the study results have only been inconsistently replicated using different miRNA expression datasets [5]. It is for this reason, that many of the findings cannot be well synthesized and interpreted.

Currently, a combination of high-throughput analytical technologies in the fields of transcriptomics, proteomics, and metabolomics has led to a comprehensive understanding of disease occurrence and development. In fact, multiple omics data analysis has become a highly important technology that assists in decreasing false positives. Given multiple transcriptomics data from different microarray studies, the aim of multiple omics data analysis is to discover which platform is the most informative with the highest quality data and to identify robust biomarkers across datasets [6, 7]. Here, we used a recently developed data analysis method-multiple co-inertia analysis (MCIA) [8], which was previously demonstrated to be a powerful method for identifying COPD-related microRNAs from three high dimensional COPD-related microarray datasets. MCIA can simultaneously project several datasets into the same dimensional space and can extract the most variant from each dataset using a covariance optimization criterion. Specially, MCIA can extract important features even when feature annotations are not present across all datasets [8].

Based on our analyzed results, miR-223, miR-132, and miR-199a-5p showed obvious association with COPD. To further approve these identified biomarkers, we extracted differentially expressed miRNAs and then ranked them according to significance for three different datasets. The results showed that miR-223, miR-132, and miR-199a-5p are all highly differentially expressed microRNAs, and each ranked highly in the lists across each of the three datasets. Moreover, when miR-223, miR-132, and miR-199a-5p are taken as predictors to classify the samples of three datasets, the pooled sensitivity and specificity reaches 0.96 and 0.75, respectively, thereby suggesting that these three miRNAs can effectively distinguish COPD patients and healthy controls.

## 2. Materials and methods

### *2.1. Data source*

In the present study, we used three COPD miRNA expression profiling data sets (GSE34466 [9], GSE24709 [10], and GSE38974 [11]) to implement our analysis. The GSE34466 dataset includes 12 healthy controls and 35 asthma patients. The GSE24709 dataset includes 19 normal controls and 24 COPD samples. Finally, the GSE38974 dataset includes 17 normal samples and 42 COPD samples. The number of miRNAs for GSE34466, GSE24709, and GSE38974 are 862, 863, and 860, respectively. The miRNA microarray of GSE34466 was performed using Agilent custom UCSF

miRNA multispecies 8 × 15K Ink-jet arrays, and the data were normalized using a robust multichip averaging method. For GSE24709, it was performed using febit Homo sapiens miRBase 13.0, and the data were normalized using the quantile method. For GSE38974, it was performed using miRCURY LNA microRNA Array, and the data were normalized using the quantile method and transformed into $\log_2$ scale. In addition, the probe with the highest average value was retained when multiple probes mapped to the same miRNA. In order to ensure that the order of the samples in all three datasets is the same, which is the basic condition for performing MCIA, we performed a match analysis for the three datasets, and the final three datasets each includes 8 normal samples and 19 COPD samples.

### *2.2. Simple introduction of MCIA*

Prior to other general meta-analyses of microarrays, MCIA projects the original data onto a lower dimensional space, which can then maximize the covariance of each dataset with respect to the reference data structure. The two-step process is described simply in the following [8]:

Step 1. Centered data:

Given a matrix $M=[m_{ij}]$, a new matrix $X=[x_{ij}]$ can be derived by defining as follows:

$$x_{ij} = \frac{p_{ij}}{r_i} - c_j \tag{1}$$

Where $p_{ij} = \frac{m_{ij}}{m_{++}}$, (2)

$$r_i = \frac{m_{i+}}{m_{++}}, \tag{3}$$

$$c_j = \frac{m_{+j}}{m_{++}}. \tag{4}$$

where $m_{i+}$ and $m_{+j}$ indicate the sums of the row and the column of M, respectively. $m_{++}$ is the grand total.

Step 2. The generalization of the co-inertia analysis (CIA):

Given a set of statistical triplets ($X_k$, $Q_k$, and D) where $Q_k$ indicates the hyperspace of features metrics, and D is an identity matrix. The sum of the squared covariance between the scores of each table with synthetic axes $v$ is maximized [8]:

$$f(\mu_1, \mu_2, ..., \mu_K, v) = \sum_{k=1}^{K} w_k \operatorname{cov}^2(X_k Q_k \mu_k, v) \tag{5}$$

where $\operatorname{cov}^2(X_k Q_k \mu_k, v)$ indicates the square of covariance, and $\omega_k$ indicates the weight of each table. The $v$ indicates the synthetic center or reference data structure, and $\mu_k$ are auxiliary axes. $X$ and $Q$ can be concatenated as:

$$X = [w_1^{\frac{1}{2}} X_1 | ... | w_K^{\frac{1}{2}} X_K], \quad Q = [Q_1 | ... | Q_K] \tag{6}$$

The first order solutions of $\mu_i^1 (i = 1,2,...,K)$ and $v^1$ can be obtained by the first principal component of the eigen-system as:

$$wXQX^T Dv = \lambda v, \quad \mu_k^1 = \frac{X_K^T Dv^1}{\left\| X_K^T Dv^1 \right\|_{Q_k}} \text{ (k=1, 2, …, K)} \tag{7}$$

With the aid of the calculation of the first order solution, when the order axes of previous sets are remained orthogonal, the subsequent solutions can be obtained with the following residual matrices:

$$v^{jT} Dv^s = 0, \ \mu_k^{jT} Q_k \mu_k^s = 0 (1 \le j < s) \tag{8}$$

Then, the second order solution is as follows:

$$X_1(order2) = X_1 - X_1 P_k^1 \tag{9}$$

$P_k^1$ is a projection matrix and

$$P_k^1 = \mu_k^1 (\mu_k^1 Q_k \mu_k^{1T})^{-1} \mu_k^1 Q_k \tag{10}$$

The two steps above are repeated until the desired number of principal components is generated. In the current MCIA analysis, miRNAs or samples sharing similar trends will be closely projected. We used the omicade4 [8] package of R software (http://www.r-project.org) to implement this analysis [12, 13].

### *2.3. Validation using differentially expressed miRNAs analysis*

In order to validate the identified miRNAs that were obtained by MCIA, we employed the most basic microarray analysis approach, SAM (significance analysis of microarrays) [14], as a filter to extract statistically significant differential expressions of miRNAs that distinguish the normal controls and the COPD patients for the three respective datasets. The repeated permutations of the data are used to determine if the expression of any particular miRNA is significantly related to the COPD phenotype. Furthermore, $P<0.05$ and false discovery rates (FDR) $<0.1$ were used as the filter criterion to select differentially expressed miRNAs. For each generated list of differentially expressed miRNAs, the miRNAs were ranked according to their p-values, with the top-ranked miRNA having the smallest p-value and thus being the most significantly differentially expressed miRNA.

## 3. Results and discussion

### *3.1. Hierarchical clustering for visualizing sample similarities*

Before performing the MCIA, we used hierarchical clustering [15] to gain a general understanding of the similarities between samples. The clustering result is shown in Figure 1A. From Figure 1A, it is apparent that GSE24709 and GSE38974 have a good clustering effect, in which the same phenotype samples were clustered. However, the clustering effect of GSE34466 is obviously lower than that of

both GSE24709 and GSE38974. For example, in the GSE34466 clustering result, it can be seen that some of the healthy controls and COPD samples were clustered into the same group. For validation, the sDmatCluster method [16] was also used to cluster samples for three COPD-related microRNA expression profiling datasets. sDmatCluster partitions the trained map into clusters using a region-growing algorithm to ensure that each cluster is continuous. We then used the supraHex package [16] of R software (http://www.r-project.org) to graph the clustering results (see Figure 1B), and similar clustering effects were found. From Figure 1B, it is clear that the GSE24709 dataset displays the best clustering effect (the samples were clustered into 4 groups), followed by the GSE38974 dataset (the samples were clustered into 5 groups), and finally, the clustering effect of the GSE34466 dataset is the worst (the samples were clustered into 8 groups). In the following study, we will compare these clustering results with MCIA analysis.

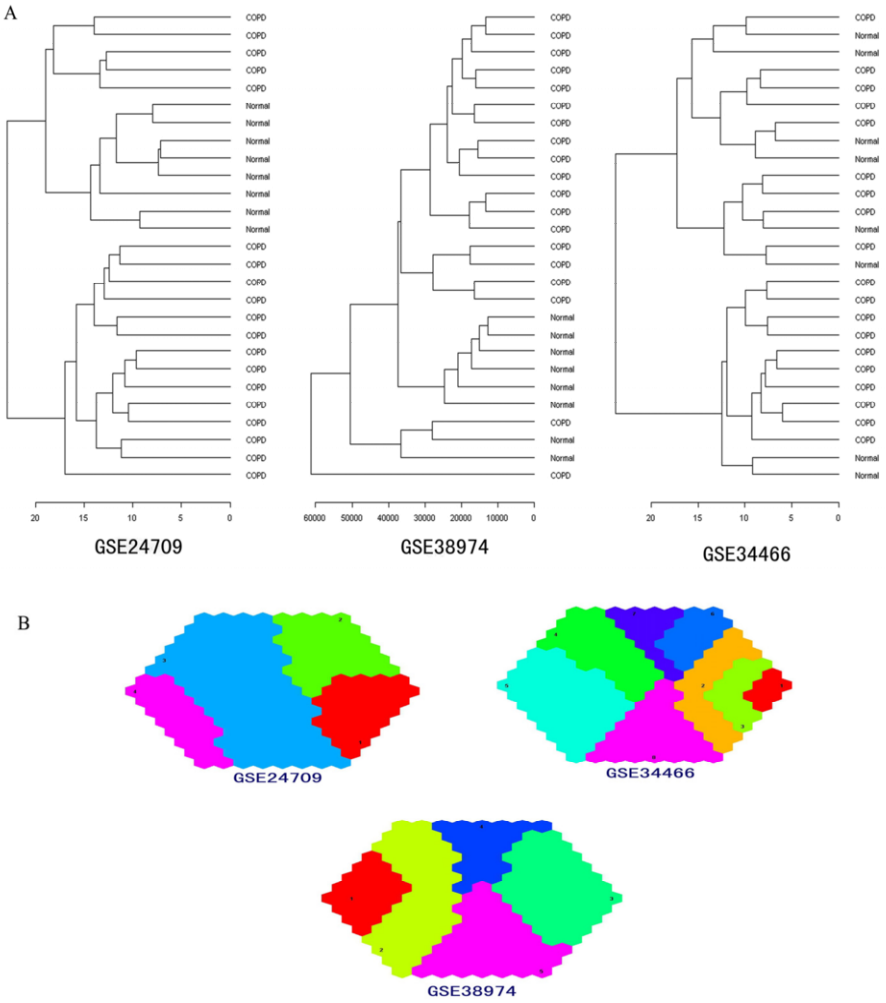


Fig. 1. (A) The hierarchical clustering for three miRNA expression profiling datasets: GSE24709, GSE38974, and GSE34466, respectively. (B) The clustering effects of GSE24709, GSE34466, and GSE38974 produced via the sDmatCluster method. Each cluster is filled with the same continuous color. The different colors indicate the different clusters. The cluster index is marked in the seed node.

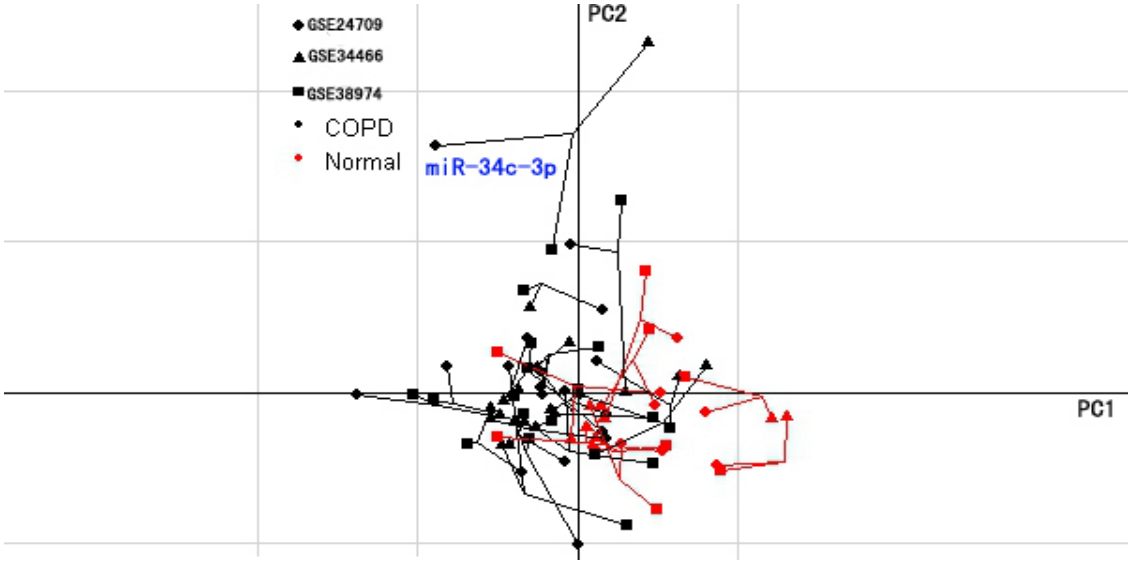


Fig. 2. The projection of samples onto the first two principal components (PC1 and PC2) of MCIA. Different shapes represent the respective microarray datasets and are connected by lines. The length of the line is proportional to the divergence between the data from the same sample. Black and red colors represent COPD and Normal samples, respectively.

### *3.2. Data exploration with Multiple Co-inertia Analysis (MCIA)*

At this point, MCIA was applied as an exploratory analysis of three COPD-related miRNA expression profiling datasets. The projection of samples onto the first two principal components (PCs) of MCIA can be seen in Figure 2.

In Figure 2, the different shapes represent samples in different platforms (microarray datasets). Samples that are attributed to the same class are linked by lines, and a shorter line is indicative of a better correlation of samples between different platforms. In other words, a relatively high correlation between microarray datasets is depicted by short lines. From Figure 2, it is apparent that miRNAs originating from the COPD sample were projected close to each other and converged into one group. Similarly, miRNAs originating from the normal healthy control were projected close to each other and converged into the other group. In particular, the miRNAs with the strongest association to a class are those projected in the same direction and have the highest weights (a greater distance from the origin) in the MCIA plots. A large variation was observed between some special miRNAs with other miRNAs in the COPD sample. For example, miR-34c-3p is located far from the other miRNAs, suggesting that miR-34c-3p is obviously differently expressed from the other miRNAs. Moreover, some evidence has confirmed that differences in miR-34c family expression are associated with emphysema severity in COPD patients. Recently, a COPD study that focused on the screening of serum miRNAs found that miR-34c-5p is significantly dys-regulated in COPD patients as compared to the healthy controls [17]. In addition, miR-34c modulates the expression of its putative target gene, SERPINE1, *in vitro* in respiratory cell lines and *ex vivo* in emphysematous lung tissue [18].

To explore each microarray dataset's contribution to the total variance, MCIA pseudo-eigenvalues were extracted to examine to what extent each microarray dataset deviates or agrees with what the majority of the datasets support. Figure 3A shows the pseudo-eigenvalues associated with the first two principal components of each dataset. This graph indicates how much the variance of an eigenvalue is contributed by each dataset. It was observed that GSE34466 is highly weighted on the second axis

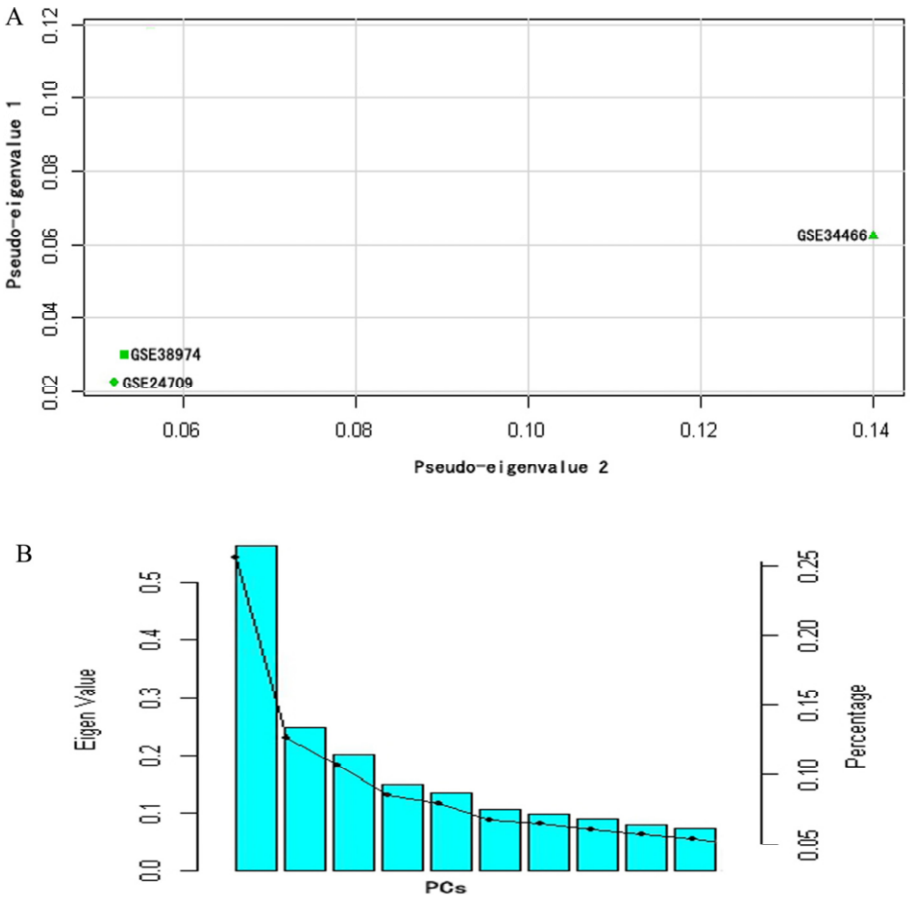


Fig. 3. (A) The concordance analysis between three COPD-related miRNA expression datasets using pseudo-eigenvalue space. The pseudo-eigenvalue space represents the overall co-structure between the datasets and displays which dataset contributes more to the total variance. (B) Pseudo-eigenvalues of three COPD-related miRNA expression datasets. Each pseudo-eigenvalue is associated with a principal component (PC) indicating the variance explained by each PC. The blue bars show the pseudo-eigenvalues (left axis), and the black line corresponds to the percentage of variance of each PC (right axis), calculated as the eigenvalue divided by the sum of all eigenvalues.

(Pseudo-eigenvalue 2). Therefore, from among the three datasets, this dataset contributes the most variance on this axis. It was also observed that in the plot GSE38974 and GSE24709 are located close to each other (Figure 3A), which means the data structures of GSE38974 and GSE24709 are highly similar. However, GSE38974 and GSE24709 are both located far from GSE34466 in the plot (Figure 3A), suggesting that the data structures of these two datasets are very different from that of GSE34466.

Ten eigenvectors were kept, and the top two axes have a relatively large eigenvalue. The first two PCs of the MCIA accounted for 25.3% and 13.9% of the variance, respectively. In other words, this observation that the first two PCs capture nearly 40% of the structure in the datasets reflects the complexity that is inherent to different samples. Generally, the similar trend of principal components reflects the similar variant sources of biological information. In the MCIA plot (Figure 3B), the first two principal components show similar trends, which indicates that the most variant sources of biological information were similar.

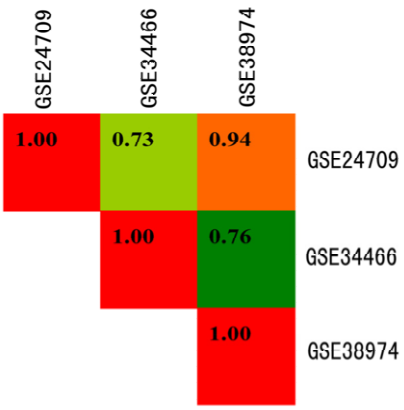


Fig. 4. The heatmap of the RV coefficients between each pair of miRNA expression datasets. A higher RV coefficient represents a higher similarity degree in the structure among different miRNA expression datasets. The deeper color indicates the higher RV coefficient.

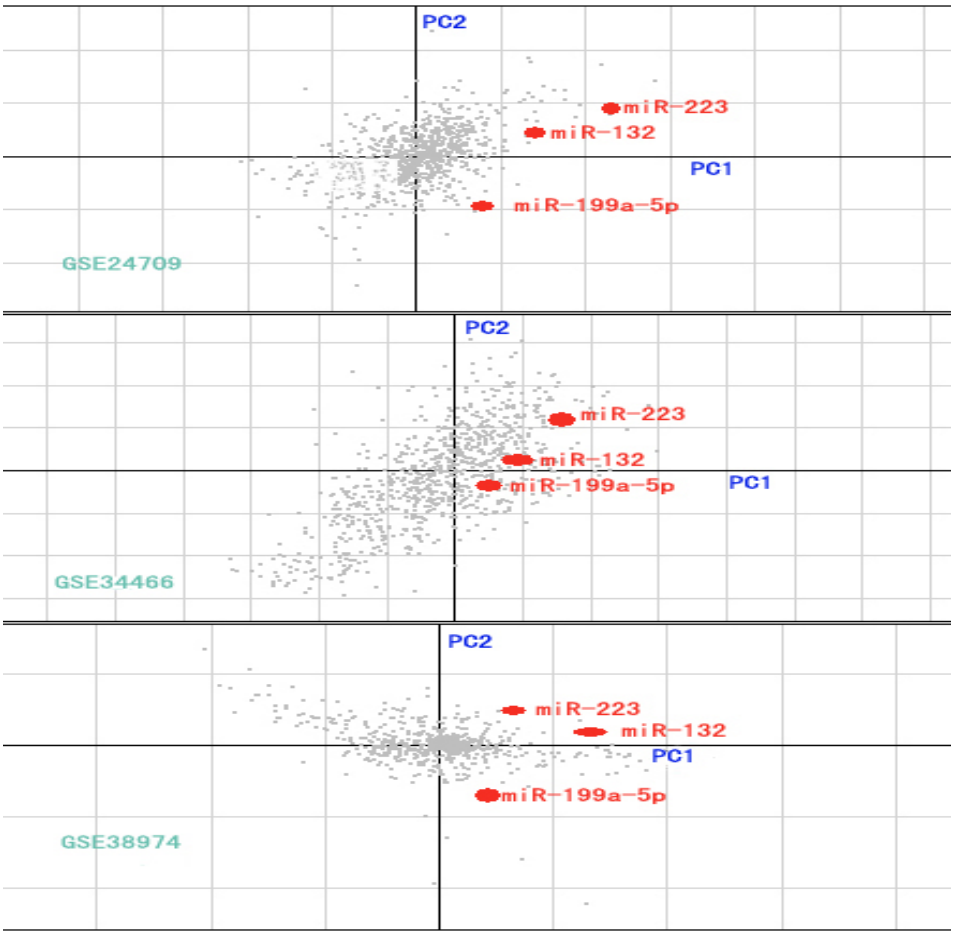


Fig. 5. Visualization of miRNAs of interest. miR-223 and miR-132 had high weights on the positive side of PC1, whereas miR-199a-5p had higher weights on the negative side of PC2.

To estimate the overall correlation between each pair of high dimensional miRNA expression profiling data, we calculated the multivariate generalization of the squared Pearson correlation coefficient, otherwise known as the pair-wise RV coefficient [8]. For each pair of miRNA expression datasets, and from the individual analysis, the RV-coefficient is calculated as the sum of the eigenvalues of the product of two cross-product matrices divided by the square root of the product of the sum of the eigenvalues. The results showed that the overall similarity in structure between GSE24709 and GSE38974 (RV=0.94) was higher than the similarity between either GSE24709 and GSE34466 (RV=0.73) or GSE38974 and GSE34466 (RV=0.76). The heatmap of the RV coefficients between each pair of miRNA expression datasets is shown in Figure 4.

In a MCIA analysis, a miRNA that is highly expressed will be displayed a greater distance from the origin, thus suggesting that it is more strongly associated with the disease phenotype. Therefore, to identify those miRNAs that are highly associated with the COPD phenotype, we examined the feature space of miRNAs that were projected in the same direction and space. It was found that miR-223 and miR-132 had high weights on the positive side of PC1. For the negative side of PC2, miR-199a-5p had a higher weight than any of the other miRNAs (See Figure 5). Moreover, previous evidence has confirmed that these identified miRNAs are associated with a COPD-related phenotype or traits. For example, from among the seventy miRNAs that were found to be differentially expressed between the lung tissue from subjects with COPD and from smokers without COPD, Ezzie *et al.* found that miR-223 was the most affected miRNA in subjects with COPD as compared with smokers without airway obstruction [19]. Moreover, by analyzing miRNA clusters, a recent study reported that the miR-132-212 cluster was up-regulated in patients with COPD, thereby suggesting that miRNA-controlled networks involved in the pathogenesis of adenocarcinoma and COPD share common risk factors (i.e. tobacco smoking) [20]. In addition, Chatila *et al.* confirmed the function of miR-199a-5p in an *in-vitro* loss-of-function cell model by running TaqMan® arrays of the human TGF-β pathway; their work suggested that the abnormal repression of miR-199a-5p in patients with COPD compared to unaffected smokers may be involved in modulating the adaptive immune balance to be in favor of a Th1 and Th17 response [21]. Hassan *et al.* also found that miR-199a-5p is a key regulator of the unfolded protein response in AAT-deficient monocytes, and epigenetic silencing of its expression regulates this process in COPD [22]. These evidences lend further support to the idea that miRNAs identified by MCIA analysis might be potential therapeutic targets for COPD.

### *3.3. Validation using differentially expressed miRNAs analysis*

To validate the identified miRNAs obtained by the MCIA analysis, we used a SAM filter to extract differentially expressed miRNAs for the three respective miRNA expression datasets. According to the filter criterion, 163, 339, and 53 differentially expressed miRNAs were extracted for GSE24709, GSE38974, and GSE34466, respectively (See Supplementary file 1). Each list of differentially expressed miRNAs was then ranked according to their p-values. The highly ranked miRNAs across the three miRNA expression datasets are consistent with those miRNAs that were identified by MCIA analysis. The results showed that miR-223, miR-132, and miR-199a-5p are highly significant differentially expressed miRNAs that all display at the top of the ranked lists. The ranks and the p-values of the three miRNAs (miR-223, miR-132, and miR-199a-5p) in the different datasets can be viewed in Table 1 and Supplementary file 2. From Table 1, it can be seen that miR-223 ranked in the top ten in both GSE24709 and GSE38974, whereas miR-132 ranked in the top ten in both GSE24709 and GSE34466. Finally, miR-199a-5p ranked first and seventh in GSE34466 and GSE38974, respectively.

Table 1

The rank and the p-value of three miRNAs in differentially expressed miRNAs lists

| miRNAs | GSE24709 | | GSE38974 | | GSE34466 | |
|---|---|---|---|---|---|---|
| | Rank | p-value | Rank | p-value | Rank | p-value |
| miR-223 | 5 | 2.21E-06 | 10 | 1.10E-07 | 27 | 3.03E-02 |
| miR-132 | 8 | 6.36E-06 | 43 | 2.77E-05 | 9 | 7.58E-03 |
| miR-199a-5p | 16 | 4.32E-05 | 7 | 8.45E-08 | 1 | 5.55E-04 |

### *3.4. Validation of classification accuracy rate*

Next, the three identified miRNAs (miR-223, miR-132, and miR-199a-5p), were used as predictors to classify samples. We implemented this process for the three datasets, respectively. In order to achieve the best performance, the Support Vector Machine (SVM) [23] was selected for classifying samples. In the SVM program, Radial Basis Function (RBF) kernel was selected. Then, MetaDiSc software [24] was used to evaluate the synthetic classification performance of the three miRNAs as MetaDiSc can summarize data from each individual study and compute the pooled sensitivity and specificity. The computed results indicated that the pooled sensitivity is at 0.96 (95%CI: 0.88-1.00) (Figure 6A), and the pooled specificity is 0.75 (95%CI: 0.53-0.90) (Figure 6B). The classification accuracy rates are 81.5%, 96.3%, and 92.6% for GSE34466, GSE24709, and GSE38974, respectively (Figure 6C). These results suggest that miR-223, miR-132, and miR-199a-5p are three important biomarkers that can classify samples well when they are taken as prediction variables.

Currently, significant developments have been made towards understanding the contribution of miRNAs, which play important roles in the inflammation of lung diseases, to COPD. Importantly, there is evidence that supports inflammation as an important determinant feature in COPD [25]. In our analysis, the three identified miRNAs were all approved as associated with inflammation. For example, low levels of miR-132 have been found in rheumatoid arthritis (RA) patients, thus suggesting that miR-132 helps to regulate inflammation in healthy joints [26]. Moreover, Dalal *et al.* found that the expression levels of miR-199a-5p were elevated in the blood of ulcerative colitis (UC) patients as compared to controls, which indicates that miR-199a-5p may be involved in the regulation of a generalized inflammatory state [27]. Therefore, this study may be extended to other studies involved with inflammation-related complex diseases.

## 4. Conclusion

Currently, although there are a certain number of miRNAs that have been found to be differentially expressed between COPD samples and normal lung tissues, and there are still some miRNAs that have been inconsistently replicated using different miRNA expression datasets. The existing meta-microarray analysis methods were recently developed, and the power of these methods is not stable and strongly depends upon the complexity of the data structure and the analysis type. We specially used the MAMA package of R software (http://www.r-project.org) to compute the similarity score of the differentially expressed miRNAs obtained from the three datasets. This similarity score is based on the number of overlapping miRNAs in the top ranks calculated by comparing the ranked miRNAs lists and computing the size of overlap for each rank. The computed similarity score was only 72.34 (p=0.92), suggesting a lower consistency of the differentially expressed miRNAs obtained from three datasets. Therefore, in the present study, we used the MCIA method, which is prior to other meta-

microarray analysis, to identify COPD-related microRNAs from three high dimensional COPD miRNA expression profiling datasets. The results showed that miR-223, miR-132, and miR-199a-5p are obviously associated with COPD phenotype, and these results are consistent with the observed highly significant differentially expressed miRNAs across three microarray datasets. For further validation, it was concluded that miR-223, miR-132, and miR-199a-5p can well distinguish COPD patients and healthy controls when they were taken as predictors to classify the samples of the three datasets. By not requiring mapping or pre-filtering the features to all datasets, MCIA increases the coverage and allows the integration of multiple microarray datasets so as to identify disease relevant biomarkers. However, there are some limitations inherent to this study. Firstly, careful statistical analysis is very important for such integration since each of the datasets as well as the specific data pre-process methodologies may harbor different levels of noise and variations. Secondly, some miRNAs, which display high weights on the positive side or negative side of PC1 and PC2, are not approved to be associated with COPD-related phenotype or traits. Therefore, in future studies, many data sources and extensive biological network context should be integrated to MCIA, which will aide in improving the accuracy of identifying COPD-related miRNAs. Finally, the identification of potential COPD-related miRNAs that can be confirmed by multiple studies is important for future research and for the advancement of COPD-related research.

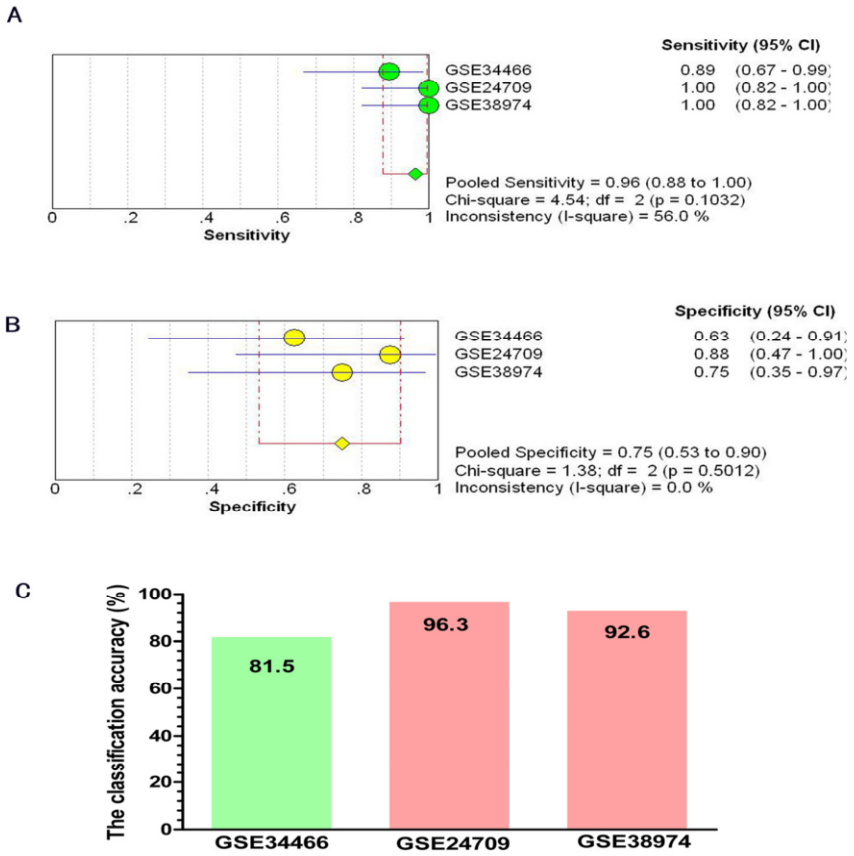


Fig. 6. (A) The pooled sensitivity of three datasets using miR-223, miR-132, and miR-199a-5p as predictors to classify samples. (B) The pooled specificity of three datasets using miR-223, miR-132, and miR-199a-5p as predictors to classify samples. (C) The classification accuracy rate of three datasets when miR-223, miR-132, and miR-199a-5p were taken as predictors to classify samples.

## Acknowledgment

This work is supported by the Beijing Natural Science Foundation (Grant No. 7142015), National Natural Science Foundation of China (Grant Nos. 31100905), and the Science Technology Development Project of Beijing Municipal Commission of Education (SQKM201210025008). This study is also funded by the excellent talent cultivation project of Beijing (2012D005018000002) and the young backbone teacher's cultivation project of the Beijing Municipal Commission of Education, and it is supported by the foundation-clinical cooperation project of Capital Medical University (14JL43).

## References

[1] E.P. Vilanova, P. Ausin, J.M. Llorens, J. Gea and E. Barreiro, Do epigenetic events take place in the vastus lateralis of patients with mild chronic obstructive pulmonary disease? PLoS One **9** (2014), e102296.
[2] J.M. Sethi and C.L. Rochester, Smoking and chronic obstructive pulmonary disease, Clinics in Chest Medicine **21** (2000), 67-86.
[3] K. William, Z.J. Jian, H.E. Lili and Q.C. Jin, Strategies for profiling MicroRNA expression, Journal of Cellular Physiology **218** (2009), 22–25.
[4] K. Chen and N. Rajewsky, The evolution of gene regulation by transcription factors and microRNAs, Nature Reviews Genetics **8** (2007), 93–103.
[5] G.R.V. Pottelberge, P. Mestdagh, K.R. Bracke, O. Thas, Y.M.V. Durme, J. Vandesompele and G.G. Brusselle, MicroRNA expression in induced sputum of smokers and patients with chronic obstructive pulmonary disease, American Journal of Respiratory and Critical Care Medicine **183** (2011), 898–906.
[6] Z. Wang, F.A. Lucas, Q. Peng and L. Yin, Improving the sensitivity of sample clustering by leveraging gene co-expression networks in variable selection, BMC Bioinformatics **15** (2014), 153.
[7] Z. Wang, Q. Peng and L. Yin, Spectral feature selection and its application in high dimensional gene expression studies, Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, 2014, pp. 314–320.
[8] M. Chen, K. Bernhard, C.C. Aedín and M.G. Amin, A multivariate approach to the integration of multi-omics datasets, BMC Bioinformatics **15** (2014), 162.
[9] O.D. Solberg, E.J. Ostrin, M.I. Love, J.C. Peng et al., Airway epithelial miRNA expression is altered in asthma, American Journal of Respiratory and Critical Care Medicine **186** (2012), 965–974.
[10] http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24709, last access: Mar. 6th, 2015.
[11] M.E. Ezzie, M. Crawford, J.H. Cho, R. Orellana et al., Gene expression networks in COPD: MicroRNA and mRNA regulation, Thorax **67** (2012), 122–131.
[12] S. Dray and A.B. Dufour, The ade4 package: Implementing the duality diagram for ecologists, Journal of Statistical Software **22** (2007), 1–20.
[13] S. Dray, A.B. Dufour and D. Chessel, The ade4 package-ii: Two-table and k-table methods, R News **7** (2007), 47–52.
[14] V.G. Tusher, R. Tibshirani and G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, Proceedings of the National Academy of Sciences of the United States of America **98** (2001), 5116–5121.
[15] G.J. Székely and M.L. Rizzo, Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method, Journal of Classification **22** (2005), 151–183.
[16] F. Hai and G. Julian, SupraHex: An R/Bioconductor package for tabular omics data analysis using a supra-hexagonal map, Biochemical and Biophysical Research Communications **443** (2014), 285–289.
[17] A. Fahri, C. Ender, A. Engin, M.O. Yasemin and Y. Pinar, Analysis of serum micro-RNAs as potential biomarker in chronic obstructive pulmonary disease, Experimental Lung Research **38** (2012), 286–294.
[18] S.M.F. Savarimuthu, M.R. Davidson, M.E. Tan, C.M. Wright, B.E. Clarke, E.E. Duhig et al., MicroRNA-34c is associated with emphysema severity and modulates SERPINE1 expression, BMC Genomics **15** (2014), 88.
[19] M.E. Ezzie, M. Crawford, J.H. Cho, R. Orellana, S. Zhang, R. Gelinas et al., Gene expression networks in COPD: MicroRNA and mRNA regulation, Thorax **67** (2012), 122–131.
[20] S.P. Molina, M.D. Pastor, R. Suarez, B.R. Romero, M.P.L.D. González, A. Salinas et al., MicroRNA clusters: Dysregulation in lung adenocarcinoma and COPD, European Respiratory Journal **43** (2014), 1740–1749.

[21] W.M. Chatila, G.J. Criner, W.W. Hancock, T. Akimova, B. Moldover, J.K. Chang et al., Blunted expression of miR-199a-5p in regulatory T cells of patients with chronic obstructive pulmonary disease compared to unaffected smokers, Clinical & Experimental Immunology **177** (2014), 341–352.
[22] T. Hassan, T.P. Carroll, P.G. Buckley, R. Cummins, S.J. O'Neill, N.G. McElvaney and C.M. Greene, MiR-199a-5p silencing regulates the unfolded protein response in chronic obstructive pulmonary disease and α1-antitrypsin deficiency, American Journal of Respiratory and Critical Care Medicine **189** (2014), 263–273.
[23] I. Guyon, J. Weston and S. Barnhill, Gene selection for cancer classification using support vector machines, Machine Learning **46** (2002), 389–422.
[24] Z. Javier, A. Victor, M. Alfonso, K. Khalid and C. Arri, Meta-DiSc: A software for meta-analysis of test accuracy data, BMC Medical Research Methodology **6** (2006), 31.
[25] T. Hassan, P.J. McKiernan, N.G. McElvaney, S.A. Cryan and C.M. Greene, Therapeutic modulation of miRNA for the treatment of proinflammatory lung diseases, Expert Review of Anti-infective Therapy **10** (2012), 359–368.
[26] K. Murata, H. Yoshitomi, S. Tanida, M. Ishikawa, K. Nishitani, H. Ito and T. Nakamura, Plasma and synovial fluid microRNAs as potential biomarkers of rheumatoid arthritis and osteoarthritis, Arthritis Research & Therapy **12** (2010), R86.
[27] S.R. Dalas and J.H. Kwon, The role of microRNA in inflammatory bowel disease, Gastroenterology & Hepatology (NY) **6** (2010), 714–722.