Genome-Wide Study of miRNA Regulated Gene Expressions Networks in Association with Emphysematous Lung Destruction

Chong Ma[1], Yen-Yi Ho[2], Stephanie Christenson[3] and Richard Nho[4]

[1]PhD Candidate, University of South Carolina
[2]Assistant Professor, University of South Carolina
[3]Assistant Professor, University of California at San Francisco
[4]Assistant Professor, University of Minnesota

Joint Statistical Meetings, August 2, 2017

COPD

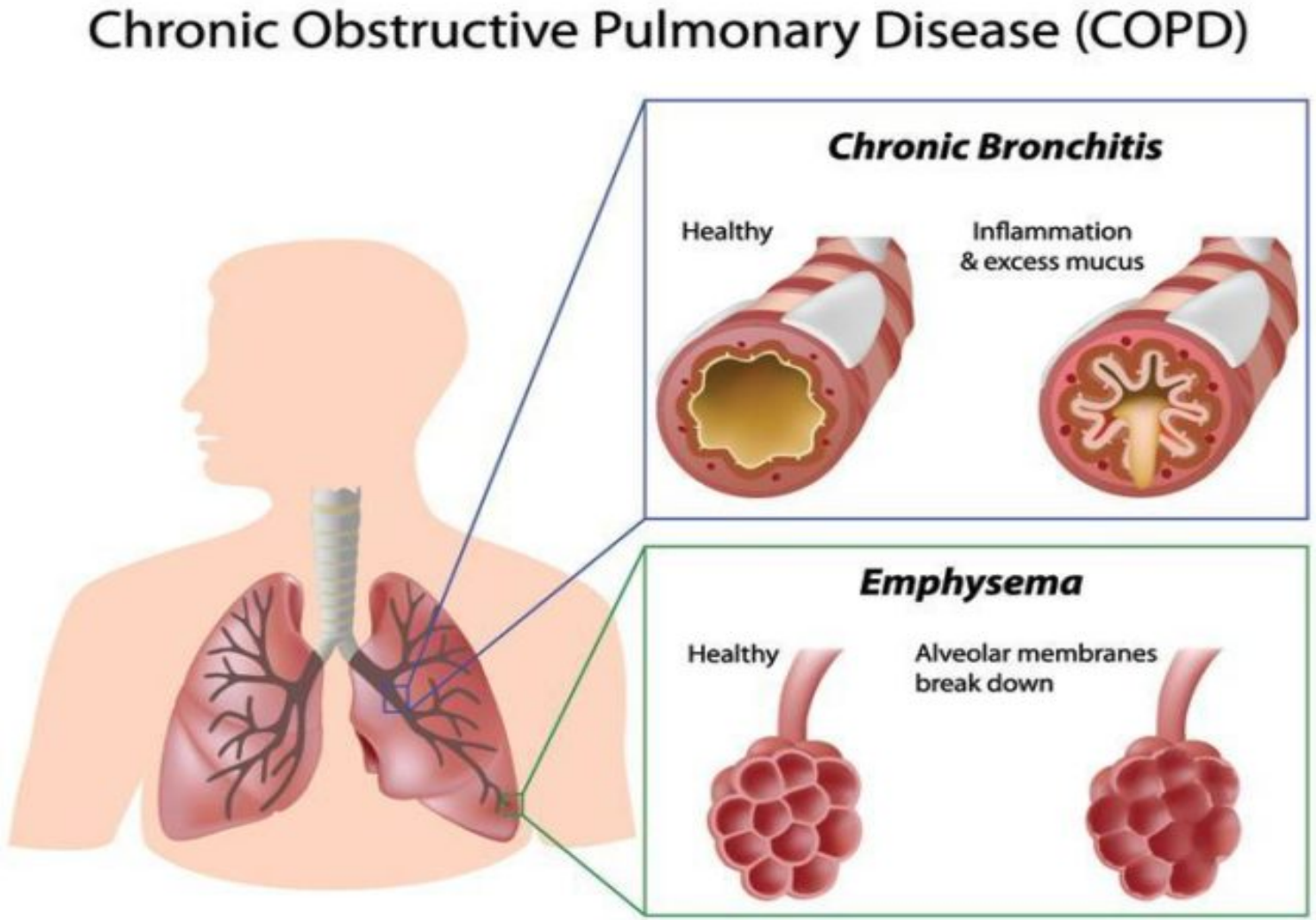


Figure 1: COPD is a progressive fatal lung disease with the potential for major complications and is often eventually fatal.

COPD

- COPD is the third leading cause of death in U.S, but the precise mechanism of the development of COPD has not fully understood yet.
- Recent studies show that several microRNAs are deregulated in COPD patients and microRNA signature becomes altered based on the severity of COPD.
- microRNA is associated with various human diseases such as cardiovascular disease, idiopathic pulmonary fibrosis (IPF) and Alzheimer Disease.

Research Goal:

- whether the progression of severe emphysema from no emphysema alters microRNA signature.
- the relationship of altered microRNAs and their target mRNA changes in emphysema severity within an individual lung.

COPD Data Preprocessing

- six subjects has undergone lung transplantation for sever COPD and two subjects are donors without COPD.
- Each subject has up to 12 slices of lung from apex to bottom.
- For each slice of lung, the emphysema severity, 397 microRNA expression profiles, and 22011 mRNA expression profiles are measured.
- The measurements for some slices are not available.
- The data is available through GEO using GSE27597 and GSE49881.

COPD Data

Response: $y =$ emphysema severity$(\log(\mathrm{Lm}))$

Demographic: $x_1 =$ COPD, $x_2 =$ packyears, $x_3 =$ age
$x_4 =$ sex, $x_5 =$ slices

miRNA: $\mathrm{miRNA}_j, j = 1, 2, \ldots, 397$

mRNA: $\mathrm{mRNA}_k, k = 1, 2, \ldots, 22011$

Where

packyears = The number of packs of cigarettes consumed each year.

slice = A multi-level factor variable ranging from 2 to 13, refering to the position where it locates in lung from apex to base.

Nonlinearity between age, packyears and lLm

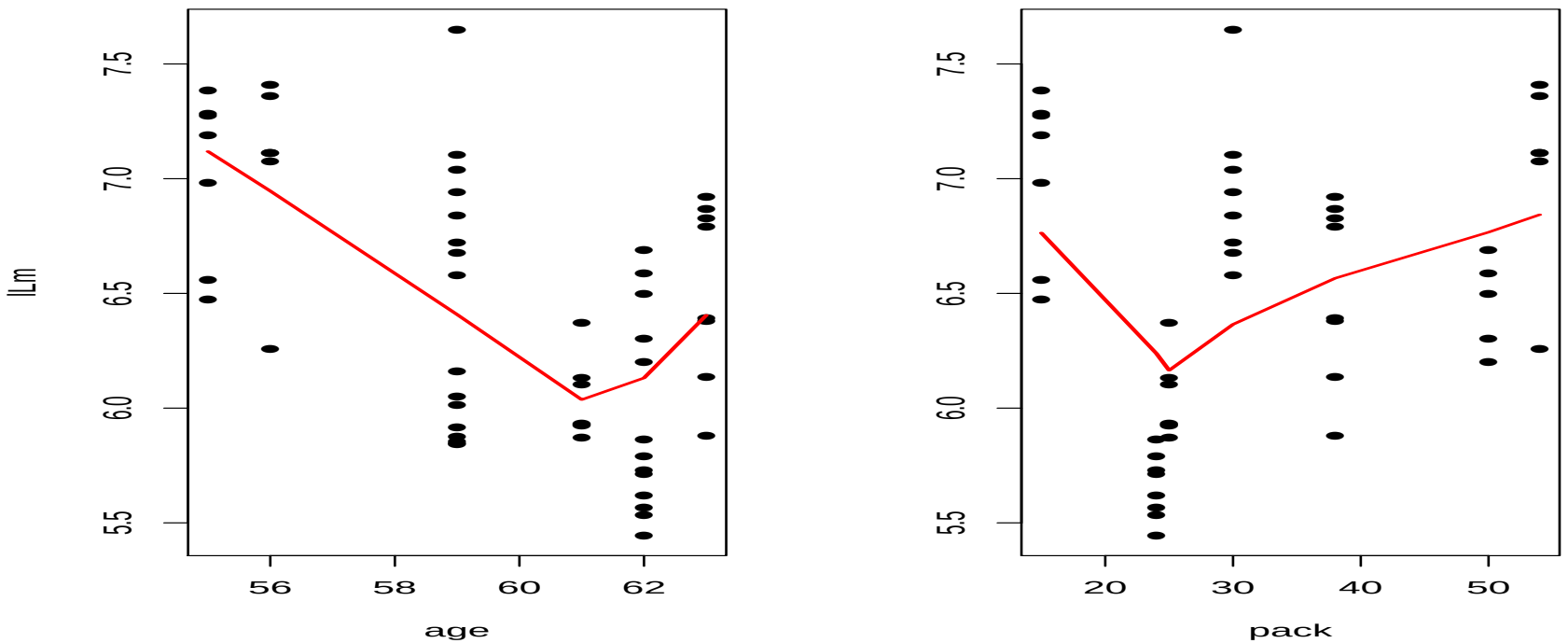


Figure 2: lowess fitting for $\log$(Lm) with packyears and age respectively

Random Intercepts of log(Lm)

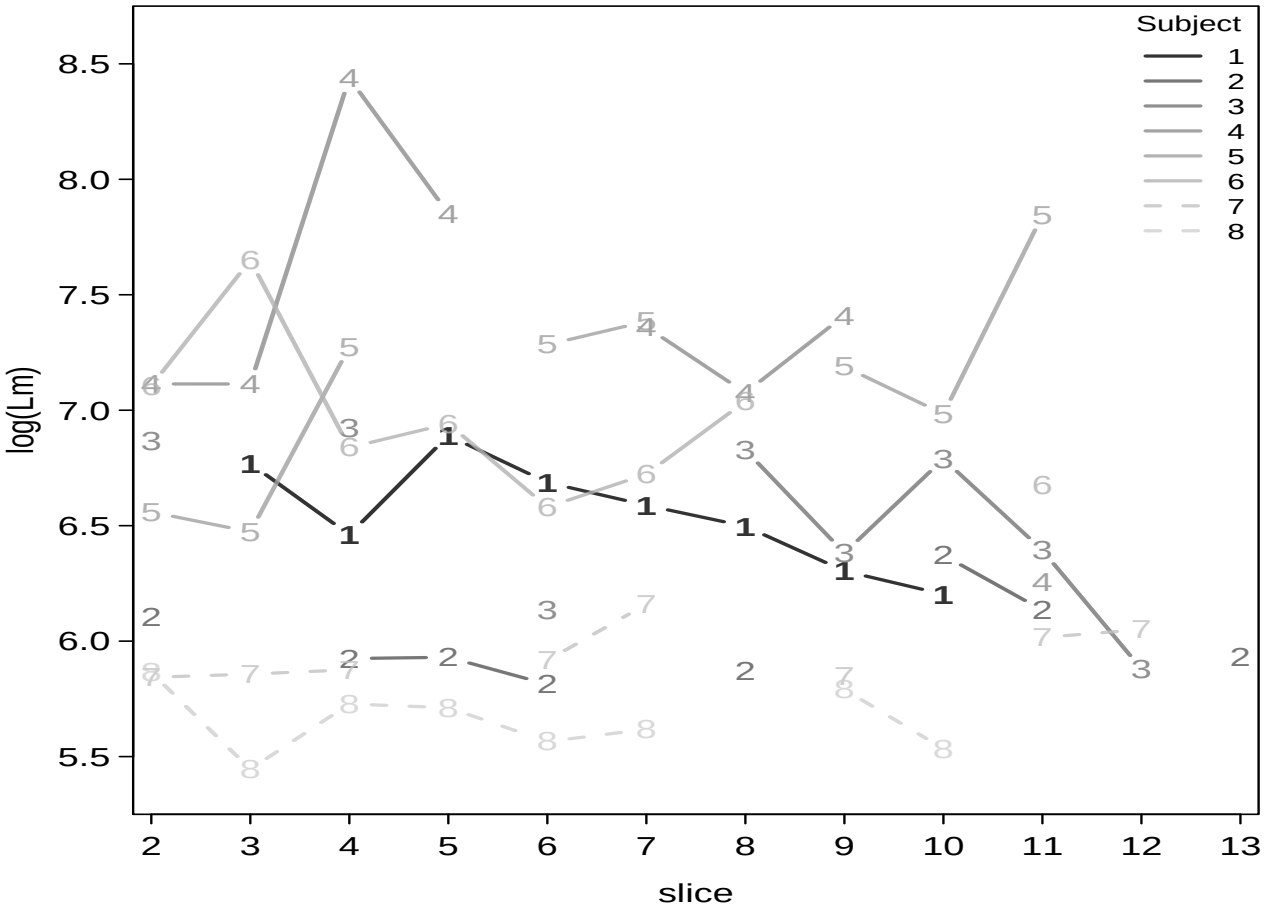


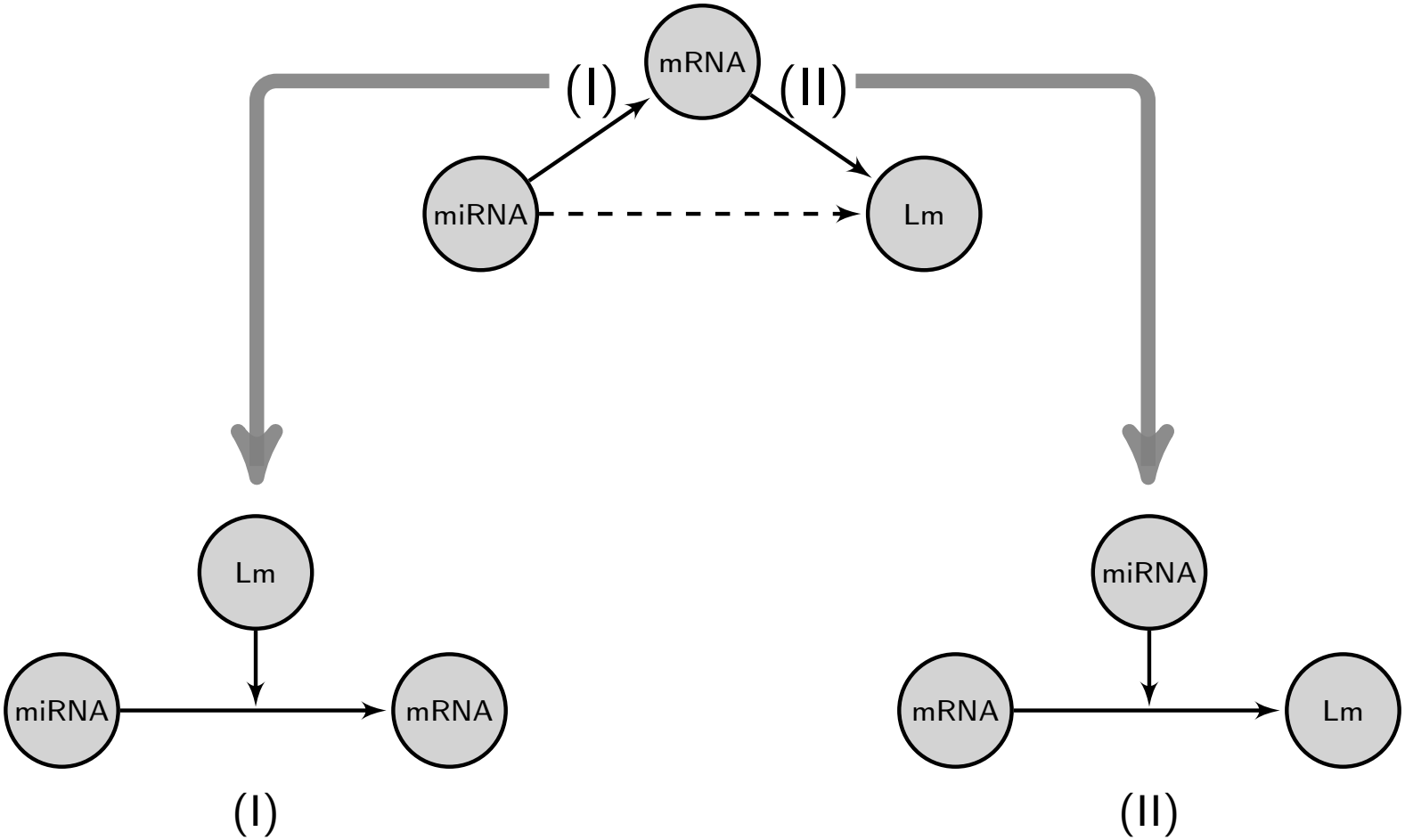
Figure 3: Spaghetti plot for $\log$(Lm) with slice for each subject. Slice 2 is the apex in lung and slice 13 is the bottom in lung. Subject 1 to 6 represent patients with COPD and subject 7 and 8 represent donors without COPD.

Model Selection

Parsimonious Model for $\log$(Lm)	logLik	AIC
COPD + packyears* + age* + sex + slice	-0.70	41.39
COPD + packyears* + age* + sex	-8.14	34.28
COPD + packyears* + age*	-13.86	43.72
COPD + packyears*	-17.71	47.41
COPD	-18.92	45.84

Table 1: Fits of parsimonious longitudinal models for $\log$(Lm) with various demographic variable including packyears(number of packs of cigarettes consumed each year), age, sex and slice. Note age* = age + (age − 61)$_+$ and packyears* = packyears + (packyears − 25)$_+$ are used in model fittings, motivated by the exploratory study in figure 2.

Weight-adjusted p-value method



Weight-adjusted p-value method

Step 1. Obtain p-values for 397 miRNAs, denoted by $p_j, j = 1, 2, \ldots, 397$, using likelihood ratio test based on the linear mixed effect models (1), (2).

$$\mathrm{lLm} = u + \mathbf{x}'\boldsymbol{\beta} + \varepsilon, \tag{1}$$

$$\mathrm{lLm} = u + \mathbf{x}'\boldsymbol{\beta} + \beta_{\mathrm{miR}_j,\mathrm{Lm}}\mathrm{miRNA}_j + \varepsilon, \tag{2}$$

Weight-adjusted p-value method

Step 2. Calculate the weight matrix **W** in which each element w_{jk} represents to the extent that the jth miRNA is associated with the emphysema severity lLm by regulating the kth mRNA.

$$w_{jk} = \underbrace{\left(\frac{\hat{\beta}_{\mathrm{miR}_j,\mathrm{mR}_k}}{SE(\hat{\beta}_{\mathrm{miR}_j,\mathrm{mR}_k})}\right)^2}_{w_{\mathrm{miR}_j\mathrm{mR}_k}} \times \underbrace{\left(\frac{\hat{\beta}_{\mathrm{mR}_k,\mathrm{lLm}}}{SE(\hat{\beta}_{\mathrm{mR}_k,\mathrm{lLm}})}\right)^2}_{w_{\mathrm{mR}_k\mathrm{lLm}}}$$

Where $\hat{\beta}_{\mathrm{miR}_j,\mathrm{mR}_k}$ and $\hat{\beta}_{\mathrm{mR}_k,\mathrm{lLm}}$ are obtained from the linear mixed effect model (3) and (4), accordingly.

$$\mathrm{mRNA}_k = u + \mathbf{x}'\boldsymbol{\beta} + \beta_{\mathrm{miR}_j,mR_k}\mathrm{miRNA}_j + \beta_{\mathrm{lLm}}\mathrm{lLm} + \varepsilon \quad (3)$$

$$\mathrm{lLm} = u + \mathbf{x}'\boldsymbol{\beta} + \beta_{\mathrm{mR}_k,\mathrm{lLm}}\mathrm{mRNA}_k + \beta_{\mathrm{miR}_j}\mathrm{miRNA}_k + \varepsilon \quad (4)$$

Weight-adjusted p-value method

$$\mathbf{W} = \overbrace{\begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} & \cdots & w_{1,22011} \\ w_{2,1} & w_{2,2} & w_{2,3} & \cdots & w_{2,22011} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{396,1} & w_{396,2} & w_{396,3} & \cdots & w_{396,22011} \end{bmatrix}}^{\text{mRNA}} \Bigg\} \text{microRNA}$$

step 3. Calculate $w_{\mathrm{miR}_j} = \max_k w_{jk}$ and assign the average scaled weight $w^*_{\mathrm{miR}_j} = \frac{w_{\mathrm{miR}_j}}{\overline{w_{\mathrm{miR}}}}$ to the jth miRNA, such that $w^*_{\mathrm{miR}} > 0$ and the average of all weights $\overline{w}^*$ is 1 and $\overline{w_{\mathrm{miR}}}$ is the average of all w_{miR_j}. Then, the weight-adjusted p-value for the jth miRNA is

$$\text{Padj} = \frac{p_j}{w^*_{\mathrm{miR}_j}}$$

miRNA-mRNA heatmap

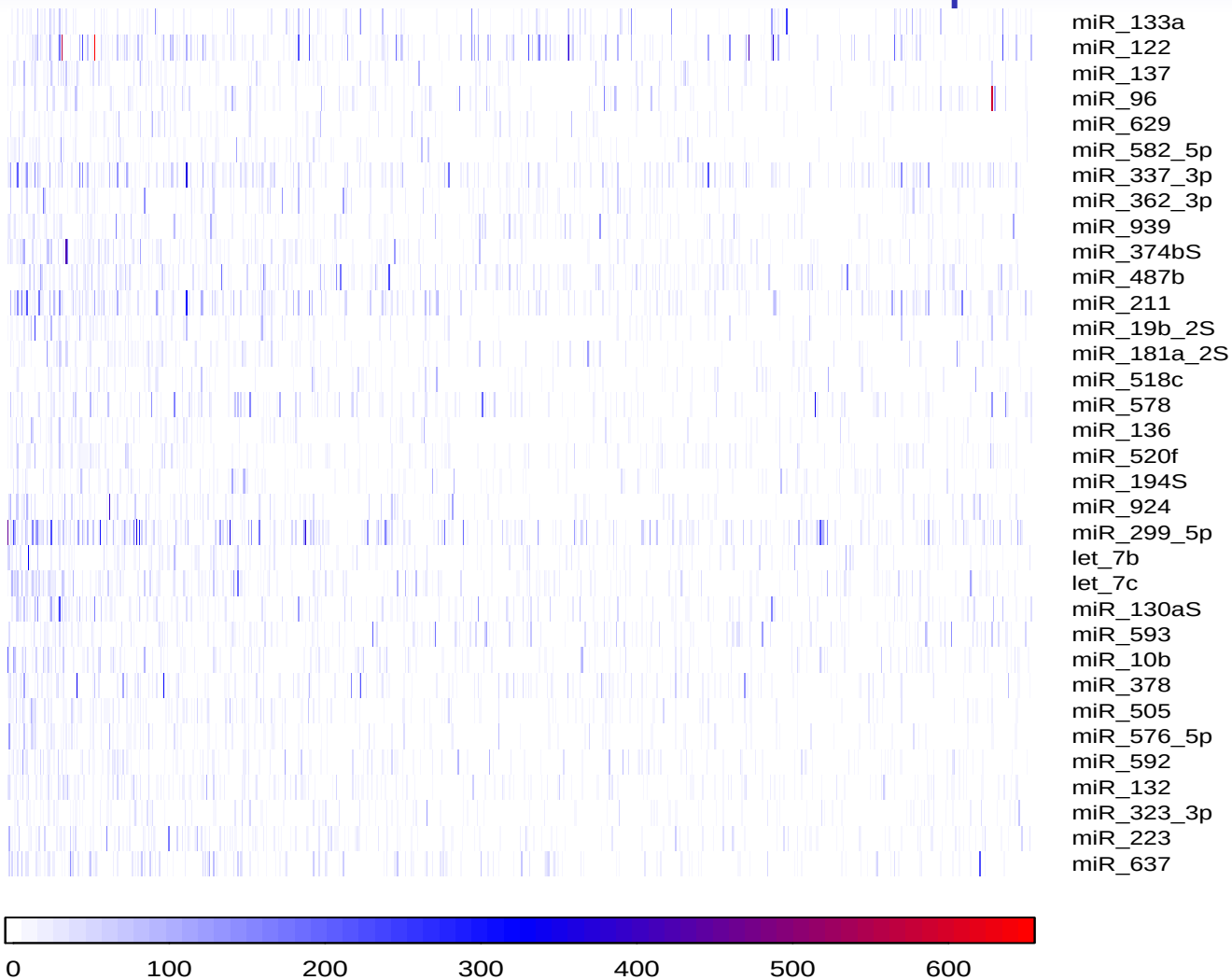
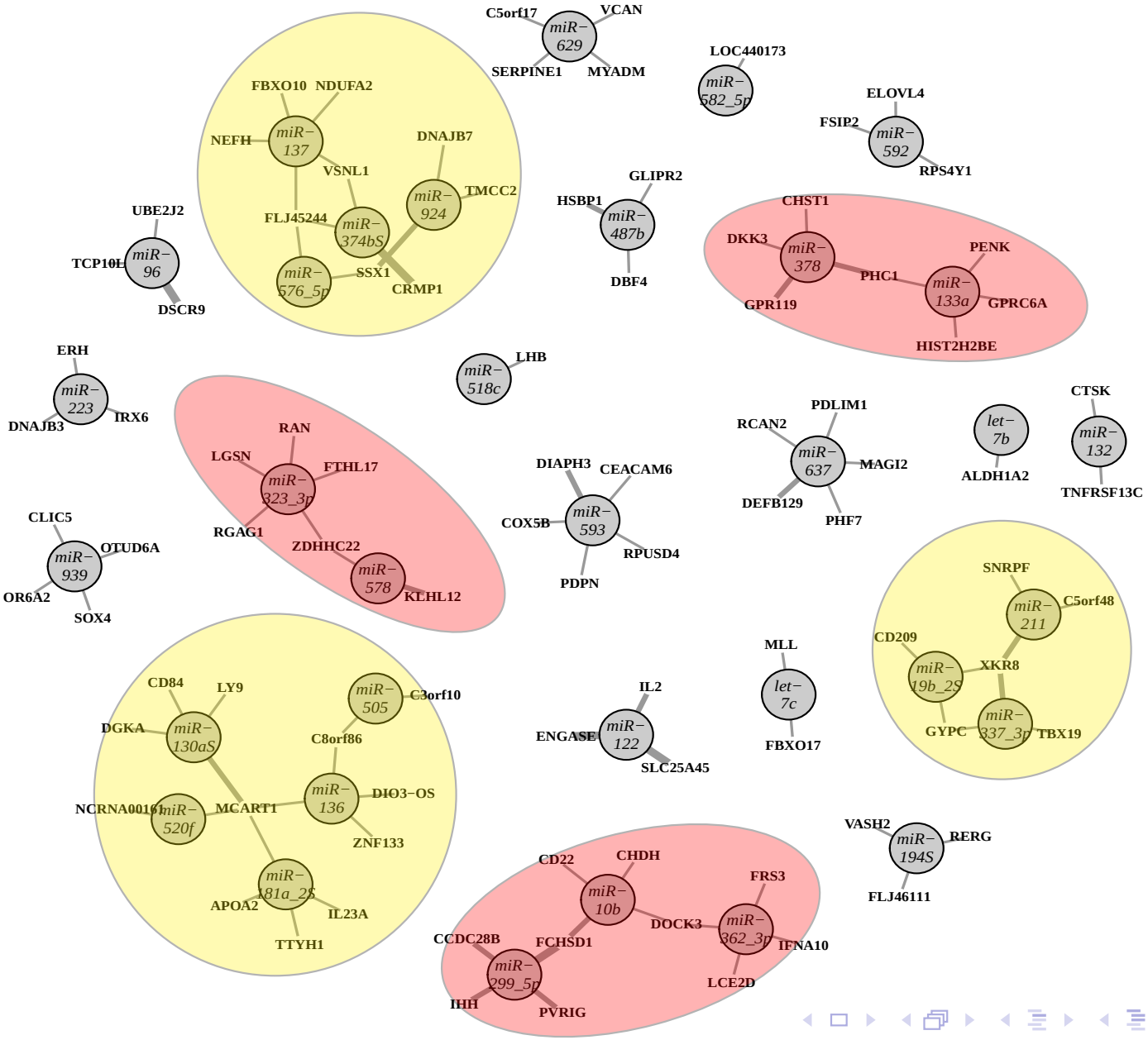


Figure 4: Weight matrix. Rows are 34 top significant miRNAs in terms of p-values. Columns are 758 mRNAs which are the combined top 5 mRNAs that are most associated with the 397 miRNAs.

miRNA-mRNA regulation network



Gene Enriched Analysis

	ID	PathName	Pvalue	Odds	Expected
1	05320	Autoimmune thyroid disease	0.023	9.346	0.238
2	04672	Intestinal immune network for IgA production	0.024	9.117	0.243
3	05012	Parkinson's disease	0.025	5.265	0.633
4	04060	Cytokine-cytokine receptor interaction	0.041	3.374	1.324
5	04630	Jak-STAT signaling pathway	0.043	4.198	0.786

Table 2: Enriched KEGG pathway using 145 top genes regulated by the 34 microRNAs that are most associated with ILm.

Conclusion

1. Proposed a new linear mixed model to analyze COPD data by integrating the genotype and phenotype information.
2. Used the weight-adjusted p-value method to detect the significant microRNAs that are most associated with COPD while controlling the nominal Family-Wise Error Rate.
3. Discovered a new sparse regulation network between microRNA and mRNA that might be beneficial to unveil the pathogenesis of COPD.
4. Provided a new way to discover the regulation network between microRNA and mRNA for cancer disease.

Reference I

Ho,Y.Y., Baechler, E.C., Ortmann, W., et al (2014) Using Gene Expression to Improve the Power of Genome-Wide Association Analysis, *Human Heredity*, **78(2)**, 94-103.

Roeder K., Bacanu SA, Wasserman, L., Devlin B. (2006) Using linkage genome scans to improve power of association in genome scans, *Am J Hum Genet*, **78**, 243-252.

Roeder K., Wasserman L. (2009) Genome-wide significance levels and weighted hypothesis testing, *Stat Sci*, **24**, 398-413.

Christenson, SA, Brandsma, CA., Campbell DJ., et al (2013) miR-638 regulates gene expression networks associated with emphysematous lung destruction, *Genome Medicine*, **5(12)**, 114.

Home S. (1979) A simple sequentially rejective multiple test procedure, *Scand J Stat*, **6**, 65-70.

Reference II

Benjaminin Y., Hochberg Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J R Statis Soc*, **57**, 289-300.

Bhattacharya S, Srisuma S, Demeo DL, et al (2009) Molecular biomarker for quantitative and discrete COPD phenotypes, *Am J Respir Cell Mol Biol*, **40**, 359-367.

Ezzie ME, Crawford M, Cho JH, et al (2012) Gene expression networks in COPD: MicroRNA and mRNA regulation, *Thorax*, **67**, 122-131.

Pottelberge GR., Mestdagh P., Bracke KR, et al (2011) MicroRNA expression in induced sputum of smokers and patients with chronic obstructive pulmonary disease, *Am J Respir Crit Care Med*, **183**, 898-906.

The End

Thank you!