Identification of biomarkers associated with clinical severity of chronic obstructive pulmonary disease

Jie Zhang[1][#], Changli Zhu[1][#], Hong Gao[1], Xun Liang[2], Xiaoqian Fan[3], Yulong Zheng[1], Song Chen[4]*, Yufeng Wan[1]*

[1]Department of Respiratory Diseases, the Affiliated Huai'an Hospital of Xuzhou Medical University, Huai'an 223002, China.
[2]College of nursing and midwifery, Jiangsu College of Nursing, Huaian 223300, China.
[3]Department of Emergency Medicine, Suqian First Hospital, Suqian 223899, China。
[4]Institute of Medicinal Biotechnology, Jiangsu College of Nursing, Huaian 223300, China.

#These authors contributed equally to this work.

*Corresponding authors:
Yufeng Wan
No.60, Huaihai South Road, Huai'an, Jiangsu, 223002, China
Email address: ggwanyufeng@163.com
Song Chen
No.2, Huanghe West Road, Huai'an, Jiangsu, 223300, China
Email address: biocs@163.com

Abstract
We sought to identify the biomarkers related to the clinical severity of stage I to stage IV chronic obstructive pulmonary disease (COPD). Gene expression profiles from the blood samples of COPD patients at each of the four stages were acquired from the Gene Expression Omnibus Database (GEO, accession number: GSE54837). Genes showing expression changes among the different stages were sorted by soft clustering. We performed functional enrichment, protein-protein interaction (PPI), and miRNA regulatory network analyses for the differentially expressed genes. The biomarkers associated with the clinical classification of COPD were selected from logistic regression models and the relationships between TLR2 and inflammatory factors were verified in clinical blood samples by qPCR and ELISA. Gene clusters demonstrating continuously rising or falling changes in expression (clusters 1, 2, and 7 and clusters 5, 6, and 8, respectively) from stage I to IV were defined as upregulated and downregulated genes, respectively, and further analyzed. The upregulated genes were enriched in functions associated with defense, inflammatory, or immune responses. The downregulated genes were associated with lymphocyte activation and cell activation. TLR2, HMOX1, and CD79A were hub proteins in the integrated network of PPI and miRNA regulatory networks. TLR2 and CD79A were significantly correlated with clinical classifications. TLR2 was closely associated with inflammatory responses during COPD progression. Functions associated with inflammatory and immune responses as well as lymphocyte activation may play important roles in the progression of COPD from stage I to IV. *TLR2* and *CD79A* may serve as potential biomarkers for the clinical severity of COPD. *TLR2* and *CD79A* may also serve as independent biomarkers in the clinical classification in COPD. TLR2

42 may play an important role in the inflammatory responses of COPD.
43
44

Introduction

46

47 Chronic obstructive pulmonary disease (COPD) is a heterogeneous condition characterized by the
48 progressive, irreversible limitation of a patient's airflow (Agusti 2014; Vestbo et al. 2013). It is a
49 leading cause of chronic morbidity and it has been predicted to be the third leading cause of
50 mortality worldwide by 2030 (Celli et al. 2015; Rodriguez-Roisin et al. 2017). COPD is a
51 progressive disease that typically worsens over time (Rabe et al. 2013). However, the molecular
52 mechanisms underlying its progression are poorly understood (Singh et al. 2018). It has been
53 reported that an early diagnosis and individualized treatments can reduce mortality and the
54 socioeconomic burden caused by the disease. Damage to the airways and lungs may be delayed
55 but not effectively reversed (PAUWELS et al. 2004).
56 Spirometry is used to diagnose the severity of COPD by determining the degree of airflow
57 limitation. Spirometry is based on the forced expiratory volume in one second (FEV1) expressed
58 as a percentage of the predicted normal value for a person's gender, age, weight, and height (Vestbo
59 et al. 2013). COPD is divided into four stages based on FEV1 according to the GOLD guidelines
60 (Vestbo et al. 2013), and COPD exacerbations are classified as episodes of worsening symptoms
61 from stage I to IV (Wedzicha & Seemungal 2007). The treatment recommendations for European
62 and American practitioners are based on FEV1 (Qaseem et al. 2011) but FEV1 does not directly
63 reflect the systemic manifestation of symptoms in COPD patients (Alotaibi & Ansari 2016).
64 Several studies have demonstrated that there is a genetic component to COPD susceptibility (Hersh
65 et al. 2011; McCLOSKEY et al. 2001). Thus, the identification of these genetic markers may help
66 to develop diagnostic and therapeutic targets for the treatment of COPD. Previous genome-wide
67 association studies have identified several genetic loci related to COPD susceptibility (Cho et al.
68 2011; Cho et al. 2014). A recent study investigated the gene expression profiles among patients
69 with frequent COPD exacerbations and identified three genes (*ARHGEF10*, *LAF4*, and *B3GNT*)
70 as predictors of exacerbations (Singh et al. 2014). However, that study did not investigate whether
71 those genes could be used as biomarkers for the clinical classification of COPD.
72 We downloaded the forementioned gene expression profile (GSE54837) (Singh et al. 2014) and
73 investigated the genes showing changes in expression among different stages of COPD. We
74 identified key genes using bioinformatics analyses and then further selected the biomarkers
75 associated with the clinical classification of COPD using logistical regression models.
76

Materials and methods

78

79 *Microarray dataset*

80

81 We downloaded the gene expression profiles with accession number GSE54837 (Singh et al. 2014)
82 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54837) from the Gene Expression
83 Omnibus (GEO) database. These profiles included different clinical stage data, and were obtained
84 based on the Affymetrix Human Genome U133 Plus 2.0 Array. We obtained 226 blood samples
85 from 90 stage I patients, 58 stage II patients, 55 stage III patients, and 13 stage IV patients. The
86 age and sex of each patient is shown in Supplementary Table 1.
87

88

Data preprocessing

90

The original CEL data were preprocessed (background adjustment, quantile normalization, final summarization, and probe ID to gene symbol) using the RMA method (Irizarry et al. 2003) in the Affy package (Gautier et al. 2004). The expression values of the probes were averaged as the final gene expression value when the different probes corresponded to the same gene symbol. After the expression matrix of all genes in each sample was obtained, the gene expression values of each sample group with different clinical grades were averaged according to the clinical classification and the final gene expression differential matrix was obtained.

98

Gene clustering analysis

100

The differential matrix of samples was subjected to noise-robust soft clustering using the fuzzy c-means algorithm in the Mfuzz package (Futschik & Carlisle 2005; Kumar & Futschik 2007). The Fuzzy C-Means (Futschik & Carlisle 2005) clustering method was adopted for clustering analysis on the genes of samples using varying times and changes in expression levels. Multiple clustering results were obtained, and the results were divided according to the trends into two parts: rising and falling. The parameters used were minSTD = 0.1, acore = 0.5, and cOptimal = 10.

107

Functional enrichment analyses

109

The gene sets with significant rising or falling trends were examined using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang et al. 2008). Enrichment results were identified in biological processes (BP), cellular components (CC), molecular functions (MF), and pathways.

115

Protein-Protein Interaction (PPI) analysis and miRNA-target gene regulatory relation analysis

117

PPI pairs were predicted based on the genes with rising and falling trends using the Search Tool for the Retrieval of Interacting Genes (STRING) (Mering et al. 2003). The required confidence of (combined score) > 0.4 was set as the threshold value for PPI relation prediction.
The regulatory miRNAs of genes with PPI relations were predicted using Webgestalt (Zhang et al. 2005) based on the hypergeometric test with a threshold of enriched gene count ≥ 3. The p value was adjusted using the Benjamini-Hochberg (BH) method (Benjamini & Hochberg 1995), and the adjusted p value < 0.05 was set as the threshold.

125

Integration of PPI and miRNA-target gene regulatory relations

127

The PPI pairs and miRNA-target gene regulatory relation pairs were integrated to construct a network using Cytoscape (Shannon et al. 2003). The degree of connectivity in the network was analyzed and the hub protein (the node with a high degree of connectivity) was selected.

131

Influencing factors of the clinical grade of COPD

133

134 We performed logistical regression analysis after obtaining the hub proteins in the network based
135 on the age ($\leq 60 = 0$, $> 60 = 1$), gender (female = 0, male = 1), and the gene corresponding to the
136 hub protein as potential factors influencing the clinical grade to further determine whether the
137 expression of the genes could be used as biomarkers of clinical grade. Univariate logistic
138 regression analysis was conducted to preliminarily screen the factors with p value < 0.05.
139 Multivariate logistic regression analysis was then conducted to further select the factors with p
140 value < 0.05.

141

142 ***Clinical data collections and detections***

143

144 Investigators obtained written informed consent before enrolling participants in the clinical trial.
145 We selected patients who underwent pulmonary function testing at the respiratory function
146 laboratory of the Affiliated Huai'an Hospital of Xuzhou Medical University between January 1,
147 2019 and May 31, 2019. Patients with complications including asthma, pulmonary interstitial
148 fibrosis, bronchial dilatation, other pulmonary diseases, or malignant tumors were excluded.

149 COPD was staged according to the GOLD guidelines: GOLD I, FEV1$\geq$80% predicted; GOLD

150 II, FEV1<80% and $\geq$50% predicted; GOLD III, FEV1<50% and $\geq$30% predicted and GOLD

151 IV, FEV1 <30% predicted. Patients selected for the study were divided into the following four
152 groups: no related diseases non-smoker, no related diseases smoker, COPD stage I-II, and COPD
153 III-IV. Peripheral venous blood samples were collected from all subjects and placed in
154 anticoagulation centrifuge tubes. Mononuclear cells were separated by density gradient
155 centrifugation. CD14$^+$ immunomagnetic beads (Miltenyi, Germany) were sorted by the
156 automatic magnetic bead sorter. TLR2 gene expression in the monocytes was detected using
157 real-time quantitative PCR according to the instructions of the Real-time PCR kit (Takara,
158 Tokyo, Japan). The amplification specificity was confirmed by melting curves and fluorescence
159 was determined at 60°C. PCR conditions were as follows: predegeneration at 95°C for 30 s,
160 followed by 40 cycles for 5 s at 95°C and 31 s at 60°C. qRT-PCR reactions were performed in a
161 total volume of 20 μL. The primer sequences for TLR2 were as follows: F: 5'-
162 TCGGAGTTCTCCCAG TTCTCT-3', R: 5'-TCCAGTGCTTCAACCCACAA -3'. The internal
163 reference was β-actin(F: 5'-CCTGGCACCCAGCACAAT-3', R:5'-
164 GGGCCGGACTCGTCATAC3'. Additional blood samples were taken and centrifuged at 3,000
165 r/min for 10 min, after which the serum was separated and stored in a refrigerator at −80 °C for
166 testing. All experiments were performed in triplicate. Relative quantitative expression levels

167 were calculated using the $2^{-\Delta\Delta Ct}$ method. The levels of IL-6, IL-8, TNF-α, and IFN-γ in the serum

168 samples were detected and compared by an ELISA Kit (Sangon Biotech, Shanghai, China). All
169 experiments were approved by the Ethics Committee of the Affiliated Huai'an Hospital of
170 Xuzhou Medical University. The clinical characteristics of all the subjects are shown in
171 Supplementary Table 2.

172

173 ***Statistical analysis***

174

175 We used SPSS 19.0 and Graphpad Prism 7.0 for statistical analysis. The expression levels of TLR2

176 and inflammatory factors were analyzed using a one-way ANOVA, followed by the least

177 significant difference multiple comparison post-hoc test, when appropriate. We analyzed the

178 correlation between TLR2 expression levels and inflammatory factor levels in all detected samples

179 using Pearson Correlation Analysis. All statistical analyses were conducted with a significance

180 level of $\alpha = 0.05$ ($P < 0.05$).

181

182 ## Results

183

184 ***Gene clustering analysis***

185

186 Opposing trends were observed among clusters 1, 2, and 7 and clusters 5, 6, and 8 as the clinical

187 stage progressed from stage I to stage IV (Figure 1). Clusters 5, 6, and 8 showed significantly

188 increased expression, while clusters 1, 2, and 7 displayed significantly decreased expressions from

189 stage I to IV. The changes in expression in clusters 3, 4, 9, and 10 were disordered. We selected

190 clusters 1, 2, 5, 6, 7, and 8 for further analysis. A total of 78, 42, and 77 genes were included in

191 clusters 5, 6, and 8, respectively, and were defined as upregulated genes. 59, 44, and 56 genes were

192 included in clusters 1, 2, and 7, respectively, and were defined as downregulated genes.

193

194 ***Functional enrichment analyses***

195

196 The upregulated genes were significantly related to GO functions including defense, inflammatory,

197 wounding, and immune responses, the positive regulation of biosynthetic processes, system lupus

198 erythematosus pathways, the toll-like receptor signaling pathway, and cytokine-cytokine receptor

199 interactions (Table 1). The downregulated genes were involved in GO functions including

200 lymphocyte, cell, and leukocyte activation, the external side of the plasma membrane, type I

201 diabetes mellitus pathway signaling, Jak-STAT signaling, and T cell receptor signaling (Table 2).

202

203 ***PPI and miRNA-target gene regulatory relations prediction***

204

205 100 pairs were predicted from upregulated genes and 40 PPI pairs were predicted from

206 downregulated genes using STRING. 64 and 92 miRNA-target gene regulatory relation pairs were

207 predicted from upregulated and downregulated genes, respectively (Table 3).

208

209 ***Network integration analysis***

210

211 Two integrated networks were constructed based on the obtained PPI and miRNA-target gene

212 regulatory relation pairs (Figure 2). A constructed network of upregulated genes displayed the

213 following five nodes with degrees ≥ 6: toll like receptor 2 (TLR2), matrix metallopeptidase 9

214 (MMP9), heme oxygenase 1 (HMOX1), and C-C motif chemokine receptor 1 (CCR1) (Figure 2A).

215 In the network of downregulated genes, the CD79a molecule (CD79A) and phospholipase C

216 gamma 1 (PLCG1) had degrees ≥ 6 and were considered to be hub nodes (Figure 2B).

217

Influencing factors of the clinical grade of COPD

219

220 Univariate logistic regression analysis results revealed that the factors of age, *TLR2*, *MMP9*,
221 *CCR1*, *CD79A*, and *PLCG1* were significantly associated with clinical classification (Table 4).
222 These factors were further analyzed by multivariate logistic regression analysis and *TLR2* was
223 found to be significantly positively correlated and *CD79A* was significantly negatively correlated
224 with clinical classification, suggesting that they could be used as independent biomarkers of
225 clinical classification (Table 5).

226

The relationships between TLR2 and inflammatory factors in COPD progression

228

229 TLR2 can trigger signal transduction and lead to the release of inflammatory mediators. COPD is
230 defined as a chronic inflammatory lung disease. We chose IL-6, IL-8, TNF-α, and IFN-γ to assess
231 the severity of COPD as they are known inflammatory factors in COPD progression. We found
232 that the levels of IL-6, IL-8, TNF-α, and IFN-γ in the smoker and COPD III-IV groups were
233 significant higher than those in the non-smoker and COPD I-II groups, and TLR2 in mononuclear
234 cells of the peripheral blood was significantly correlated with the clinical classification of the
235 patients (Figure 3A-E). We also found that TLR2 and the expressions of inflammatory markers
236 were more significant in the GOLDIII/VI group than those in the healthy smoker group. We used
237 the ROC curves to identify the sensitivity and specificity of the expression levels. The results
238 showed that the five markers, IL-6, IL-8, TNF-α, IFN-γ and TLR2, have diagnostic value for
239 COPD. The areas under the curve (AUC) were 0.7782, 0.8857, 0.9318, 0.9276 and 0.7647,
240 respectively (P<0.0001, figures 3F-3J). The correlations between the TLR2 expression and
241 inflammatory factor levels was analyzed by Pearson correlation analysis. The expression levels of
242 TLR2 were correlated with the serum levels of IL-6 (standardized regression coefficient = 0.5594,
243 P < 0.0001, Figure 3K), IL-8 (standardized regression coefficient = 0.6498, P < 0.0001, Figure
244 3L), TNF-α (standardized regression coefficient = 0.5126, P < 0.0001, Figure 3M), and IFN-γ
245 (standardized regression coefficient = 0.4973, P < 0.0001, Figure 3N).

246

Discussion

248 COPD is a growing global health problem. As the population ages, life expectancy among the
249 elderly is expected to rise in relation to the number of smokers and the prevalence of COPD.
250 Researchers are working to identify and diagnose COPD using easily available biomarkers to
251 facilitate the early detection of fluid build-up in COPD and improve diagnostic accuracy. Previous
252 studies focused on finding new biomarkers in the blood for the early diagnosis of COPD and many
253 of these potential diagnostic markers need to be verified in additional studies.
254 We studied gene clusters with continuously rising (clusters 5, 6, and 8) or falling (clusters 1, 2,
255 and 7) trends in expression changes from stage I to stage IV. The upregulated genes, such as *TLR2*,
256 and *HMOX1*, were significantly enriched in functions associated with defense, inflammatory, or
257 immune responses, while the downregulated genes, such as *CD79A*, were associated with the
258 activation of lymphocytes, cells, and leukocytes. TLR2, HMOX1, and CD79A were hub proteins

259   in the integrated network. Logistic regression analysis showed that *TLR2* was positively correlated
260   and *CD79A* was negatively correlated with clinical classification.
261   Our results showed that the upregulated genes, such as *TLR2*, and *HMOX1*, were significantly
262   associated with the inflammatory and immune responses. Exposure to inhaled pollutants can result
263   in chronic airway inflammation in COPD by activating structural and inflammatory cells within
264   the lungs (Rovina et al. 2013). COPD exacerbation is defined as an acute worsening of the
265   symptoms that are implicated in increased systemic and airway inflammation (Wedzicha &
266   Seemungal 2007). The inflammatory and immune responses are known to play a critical role in
267   COPD. Dave Singh, et al. (2014b) found that systemic immune function is associated with the
268   exacerbation of COPD. Our study further suggested the important role of inflammatory and
269   immune responses in the progression of COPD from stage I to stage IV. IL-6, IL-8, TNF-α and
270   IFN-γ are important inflammatory mediators during the progression of COPD and IL-6 participates
271   in the inflammatory response. It can further aggravate the oxidative stress and inflammatory
272   responses by promoting the proliferation and maturation of T lymphocytes and B lymphocytes.
273   IL-8 is an inflammatory mediator that promotes the activation of neutrophils and the release of
274   various enzymes that damage the bronchi. Excessive levels of IL-8 can aggravate the inflammatory
275   response and destroy the lung tissue (Farahi et al. 2017; Nakamoto et al. 2019). By acting on
276   neutrophils, TNF-α enhances the expression of leukocyte adhesion molecules and activates other
277   factors involved in the inflammatory response. Higher levels of TNF-α and IFN-γ in COPD also
278   indicate increased inflammation (Xu et al. 2019; Zhang et al. 2016). Our findings about these
279   inflammatory factors in COPD were consistent with previous studies.
280   TLR2 was among the top two hub proteins in the network. TLR2 is a member of the toll-like
281   receptor family, which plays a fundamental role in the activation of innate immunity (Da Silva et
282   al. 2008). Haw et al. (2018) found that TLR2 mRNA expression was increased in the epithelium
283   and parenchyma of the mouse airway when chronically exposed to cigarette smoke. These results
284   were replicated in human COPD patients. Our results showed that increased *TLR2* was an
285   independent clinical classification biomarker in COPD. The results of some studies do not support
286   this research. A previous study reported that the alveolar macrophages in smokers and COPD
287   patients presented an equally decreased surface expression of TLR2 compared to non-smokers
288   (Droemann et al. 2005). The role of TLR2 in the pathogenesis of COPD is also controversial in
289   previous literature (Freeman et al. 2013; Haw et al. 2018; Simpson et al. 2013; von Scheele et al.
290   2011). These conflicting results may be due to differences in experimental methods (e.g. peripheral
291   blood monocytes vs. macrophages) or cohorts of patients with varying medical backgrounds.
292   We found that the downregulated genes were involved in functions associated with lymphocyte
293   activation, which may suggest that lymphocyte activation plays a role in COPD development.
294   COPD has been thought to possess an autoimmune component (Agusti et al. 2003; Lee et al. 2007),
295   but the antigenic stimulation responsible for lymphocyte activation is unclear. Previous studies
296   demonstrated that the numbers of B lymphocytes and CD8+ T cells are increased in the airways
297   of COPD patients compared with healthy controls (Gosman et al. 2006; Saetta et al. 1999). The
298   number of lymphocytes has been shown to increase with the severity of the disease (Hogg et al.
299   2004). We found that the hub protein, CD79A, was enriched with lymphocyte activation. CD79A
300   has been demonstrated to play diverse roles in the development and function of B lymphocytes
301   (Pelanda et al. 2002) and may be downregulated in severe COPD when compared with controls
302   (Cockayne et al. 2011). Our results showed that CD79A was negatively correlated with the clinical
303   classification of COPD. Thus, lymphocyte activation and *CD79A* may be associated with the
304   clinical severity of COPD.

305 Our study is limited because we analyzed the differentially expressed genes in COPD patients
306 using a single set of GEO data. The sample capacity was small and more GEO data should be used
307 in further studies. Secondly, by extracting the total RNA content in blood monocytes, we found
308 that the Ct value was relatively large, indicating that the expression of TLR2 in blood was lower,
309 and the micro expression levels may have caused experimental errors. We performed each test at
310 least three times to improve the repeatability of the results. Finally, we only studied the expression
311 of TLR2 in the clinical population but its specific mechanism and inconsistency in previous studies
312 was well explained and requires additional experimentation. Our results show that the
313 inflammatory and immune responses, as well as lymphocyte activation, may play important roles
314 in COPD development from stage I to stage IV. *TLR2*, and *CD79A* may serve as potential
315 biomarkers in the exacerbation of COPD, and *TLR2* and *CD79A* may also serve as independent
316 biomarkers for the clinical classification of COPD.

317

Acknowledgements

319 We wish to thank Mr. Lizhang Xun for assisting with the collection of clinical blood samples and
320 Prof. Song Chen for their assistance with the experiments.

321

322

References.

324 Agusti A. 2014. The path to personalised medicine in COPD. *Thorax* 69:857-864.

325 Agusti A, MacNee W, Donaldson K, and Cosio M. 2003. Hypothesis: does COPD have an autoimmune component? :
326   BMJ Publishing Group Ltd.

327 Alotaibi N, and Ansari K. 2016. The Body Mass Index, Airflow Obstruction, Dyspnea and Exercise Capacity (BODE)
328   Index in Chronic Obstructive Pulmonary Disease for Saudi Population. *J Pulm Respir Med* 6:2.

329 Benjamini Y, and Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to
330   Multiple Testing. *Journal of the Royal Statistical Society* 57:289-300.

331 Celli BR, Decramer M, Wedzicha JA, Wilson KC, Agustí A, Criner GJ, Macnee W, Make BJ, Rennard SI, and Stockley
332   RA. 2015. An Official American Thoracic Society/european Respiratory Society Statement: Research
333   questions in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 191:e4-e27.

334 Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, Himes BE, Sylvia JS, Klanderman BJ, and Ziniti JP.
335   2011. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13.
336   *Human molecular genetics* 21:947-957.

337 Cho MH, McDonald M-LN, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, DeMeo DL, Sylvia JS, Ziniti J, and Laird NM.
338   2014. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-
339   analysis. *The lancet Respiratory medicine* 2:214-225.

340 Cockayne DA, Cheng DT, Waschki B, Hilton H, Kourteva G, Sridhar S, Ravindran P, Bitter H, Visvanathan S, and Mueller
341   K-C. 2011. Whole Blood Gene Expression And Serum Biomarker Profiling In A COPD Cohort Identifies
342   Markers Related To Neutrophilic Inflammation, Wound Healing And Altered B Cell Signaling. *C22*
343   *PATHOBIOLOGY OF COPD: LESSONS FROM INFLAMMATORY MECHANISMS AND GENOMIC STUDIES*: Am

344   Thoracic Soc, A4091-A4091.

345 Da Silva CA, Hartl D, Liu W, Lee CG, and Elias JA. 2008. TLR-2 and IL-17A in chitin-induced macrophage activation and
346   acute inflammation. *The Journal of Immunology* 181:4279-4286.

347 Droemann D, Goldmann T, Tiedje T, Zabel P, Dalhoff K, and Schaaf B. 2005. Toll-like receptor 2 expression is
348   decreased on alveolar macrophages in cigarette smokers and COPD patients. *Respiratory research* 6:68.

349 Farahi N, Paige E, Balla J, Prudence E, Ferreira RC, Southwood M, Appleby SL, Bakke P, Gulsvik A, Litonjua AA, Sparrow
350   D, Silverman EK, Cho MH, Danesh J, Paul DS, Freitag DF, and Chilvers ER. 2017. Neutrophil-mediated IL-6
351   receptor trans-signaling and the risk of chronic obstructive pulmonary disease and asthma. *Hum Mol Genet*
352   26:1584-1596. 10.1093/hmg/ddx053

353 Freeman CM, Martinez FJ, Han MK, Washko GR, Jr., McCubbrey AL, Chensue SW, Arenberg DA, Meldrum CA,
354   McCloskey L, and Curtis JL. 2013. Lung CD8+ T cells in COPD have increased expression of bacterial TLRs.
355   *Respir Res* 14:13. 10.1186/1465-9921-14-13

356 Futschik ME, and Carlisle B. 2005. Noise-robust soft clustering of gene expression time-course data. *Journal of*
357   *bioinformatics and computational biology* 3:965-988.

358 Gautier L, Cope L, Bolstad BM, and Irizarry RA. 2004. affy—analysis of Affymetrix GeneChip data at the probe level.
359   *Bioinformatics* 20:307-315. 10.1093/bioinformatics/btg405

360 Gosman MM, Willemse BW, Jansen DF, Lapperre TS, van Schadewijk A, Hiemstra PS, Postma DS, Timens W, and
361   Kerstjens H. 2006. Increased number of B-cells in bronchial biopsies in COPD. *European Respiratory Journal*
362   27:60-64.

363 Haw TJ, Starkey MR, Pavlidis S, Fricker M, Arthurs AL, Nair PM, Liu G, Hanish I, Kim RY, Foster PS, Horvat JC, Adcock
364   IM, and Hansbro PM. 2018. Toll-like receptor 2 and 4 have opposing roles in the pathogenesis of cigarette
365   smoke-induced chronic obstructive pulmonary disease. *Am J Physiol Lung Cell Mol Physiol* 314:L298-l317.
366   10.1152/ajplung.00154.2017

367 Hersh CP, Hokanson JE, Lynch DA, Washko GR, Make BJ, Crapo JD, Silverman EK, and Investigators C. 2011. Family
368   history is a risk factor for COPD. *Chest* 140:343-350.

369 Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, Cherniack RM, Rogers RM, Sciurba FC, and Coxson HO.
370   2004. The nature of small-airway obstruction in chronic obstructive pulmonary disease. *New England*
371   *Journal of Medicine* 350:2645-2653.

372 Huang DW, Sherman BT, and Lempicki RA. 2008. Systematic and integrative analysis of large gene lists using DAVID
373   bioinformatics resources. *Nat Protocols* 4:44-57.
374   http://www.nature.com/nprot/journal/v4/n1/suppinfo/nprot.2008.211_S1.html

375 Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, and Speed TP. 2003. Exploration,
376   normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249-
377   264. 10.1093/biostatistics/4.2.249

378 4/2/249 [pii]

379 Kumar L, and Futschik ME. 2007. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation*
380   2:5.

381 Lee S-H, Goswami S, Grudo A, Song L-z, Bandi V, Goodnight-White S, Green L, Hacken-Bitar J, Huh J, and Bakaeen F.
382   2007. Antielastin autoimmunity in tobacco smoking–induced emphysema. *Nature medicine* 13:567-569.

383 McCLOSKEY SC, Patel BD, Hinchliffe SJ, Reid ED, Wareham NJ, and Lomas DA. 2001. Siblings of patients with severe
384   chronic obstructive pulmonary disease have a significant risk of airflow obstruction. *American journal of*

385   *respiratory and critical care medicine* 164:1419-1424.

386   Mering Cv, Huynen M, Jaeggi D, Schmidt S, Bork P, and Snel B. 2003. STRING: a database of predicted functional
387   associations between proteins. *Nucleic acids research* 31:258-261.

388   Nakamoto K, Watanabe M, Sada M, Inui T, Nakamura M, Honda K, Wada H, Ishii H, and Takizawa H. 2019.
389   Pseudomonas aeruginosa-derived flagellin stimulates IL-6 and IL-8 production in human bronchial epithelial
390   cells: A potential mechanism for progression and exacerbation of COPD. *Exp Lung Res* 45:255-266.
391   10.1080/01902148.2019.1665147

392   PAUWELS RA, BUIST AS, CALVERLEY PMA, JENKINS CR, and HURD SS. 2004. Global Strategy for the Diagnosis,
393   Management, and Prevention of Chronic Obstructive Pulmonary Disease: GOLD Executive Summary
394   Updated 2003. *Copd-journal of Chronic Obstructive Pulmonary Disease* 1:105.

395   Pelanda R, Braun U, Hobeika E, Nussenzweig MC, and Reth M. 2002. B cell progenitors are arrested in maturation
396   but have intact VDJ recombination in the absence of Ig-α and Ig-β. *The Journal of Immunology* 169:865-872.

397   Qaseem A, Wilt TJ, Weinberger SE, Hanania NA, Criner G, van der Molen T, Marciniuk DD, Denberg T, Schünemann
398   H, Wedzicha W, MacDonald R, Shekelle P. 2011. Diagnosis and management of stable chronic obstructive
399   pulmonary disease: A clinical practice guideline update from the american college of physicians, american
400   college of chest physicians, american thoracic society, and european respiratory society. *Annals of internal*
401   *medicine* 155:179-191. 10.7326/0003-4819-155-3-201108020-00008

402   Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguezroisin R, and Weel CV.
403   2013. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary
404   Disease. *American Journal of Respiratory & Critical Care Medicine* 176:532.

405   Rodriguez-Roisin R, Rabe KF, Vestbo J, Vogelmeier C, and Agustí A. 2017. Global Initiative for Chronic Obstructive
406   Lung Disease (GOLD) 20th anniversary: a brief history of time. Eur Respiratory Soc.

407   Rovina N, Koutsoukou A, and Koulouris NG. 2013. Inflammation and immune response in COPD: where do we stand?
408   *Mediators of inflammation* 2013.

409   Saetta M, Baraldo S, Corbino L, Turato G, Braccioni F, Rea F, CAVALLESCO G, Tropeano G, Mapp CE, and Maestrelli
410   P. 1999. CD8+ ve cells in the lungs of smokers with chronic obstructive pulmonary disease. *American Journal*
411   *of Respiratory and Critical Care Medicine* 160:711-717.

412   Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T. 2003. Cytoscape:
413   A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*
414   13:2498-2504. 10.1101/gr.1239303

415   Simpson JL, McDonald VM, Baines KJ, Oreo KM, Wang F, Hansbro PM, and Gibson PG. 2013. Influence of age, past
416   smoking, and disease severity on TLR2, neutrophilic inflammation, and MMP-9 levels in COPD. *Mediators*
417   *Inflamm* 2013:462934. 10.1155/2013/462934

418   Singh D, Fox SM, Tal-Singer R, Bates S, Riley JH, and Celli B. 2014. Altered gene expression in blood and sputum in
419   COPD frequent exacerbators in the ECLIPSE cohort. *PloS one* 9:e107381.

420   Singh S, Verma S, Kumar S, Ahmad M, Nischal A, Singh S, and Dixit R. 2018. Correlation of severity of chronic
421   obstructive pulmonary disease with potential biomarkers. *Immunology letters* 196:1-10.

422   Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, and Nishimura
423   M. 2013. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary
424   disease: GOLD executive summary. *American journal of respiratory and critical care medicine* 187:347-365.

425   von Scheele I, Larsson K, Dahlén B, Billing B, Skedinger M, Lantz AS, and Palmberg L. 2011. Toll-like receptor

426   expression in smokers with and without COPD. *Respir Med* 105:1222-1230. 10.1016/j.rmed.2011.02.012

427 Wedzicha JA, and Seemungal TA. 2007. COPD exacerbations: defining their cause and prevention. *Lancet* 370:786-

428   796.

429 Xu W, Li R, and Sun Y. 2019. Increased IFN-γ-producing Th17/Th1 cells and their association with lung function and

430   current smoking status in patients with chronic obstructive pulmonary disease. *BMC Pulm Med* 19:137.

431   10.1186/s12890-019-0899-2

432 Zhang B, Kirov S, and Snoddy J. 2005. WebGestalt: an integrated system for exploring gene sets in various biological

433   contexts. *Nucleic acids research* 33:W741-W748.

434 Zhang L, Gu H, Gu Y, and Zeng X. 2016. Association between TNF-α -308 G/A polymorphism and COPD susceptibility:

435   a meta-analysis update. *Int J Chron Obstruct Pulmon Dis* 11:1367-1379. 10.2147/copd.S105394

436

Figure 1

Results of soft clustering for genes.

Changes in colors (red-blue-green) indicate the coincidence degree of a gene change with the central variation of the cluster. Red indicates a high degree and green indicates a low degree of coincidence.Cluster1-10 (A-J).

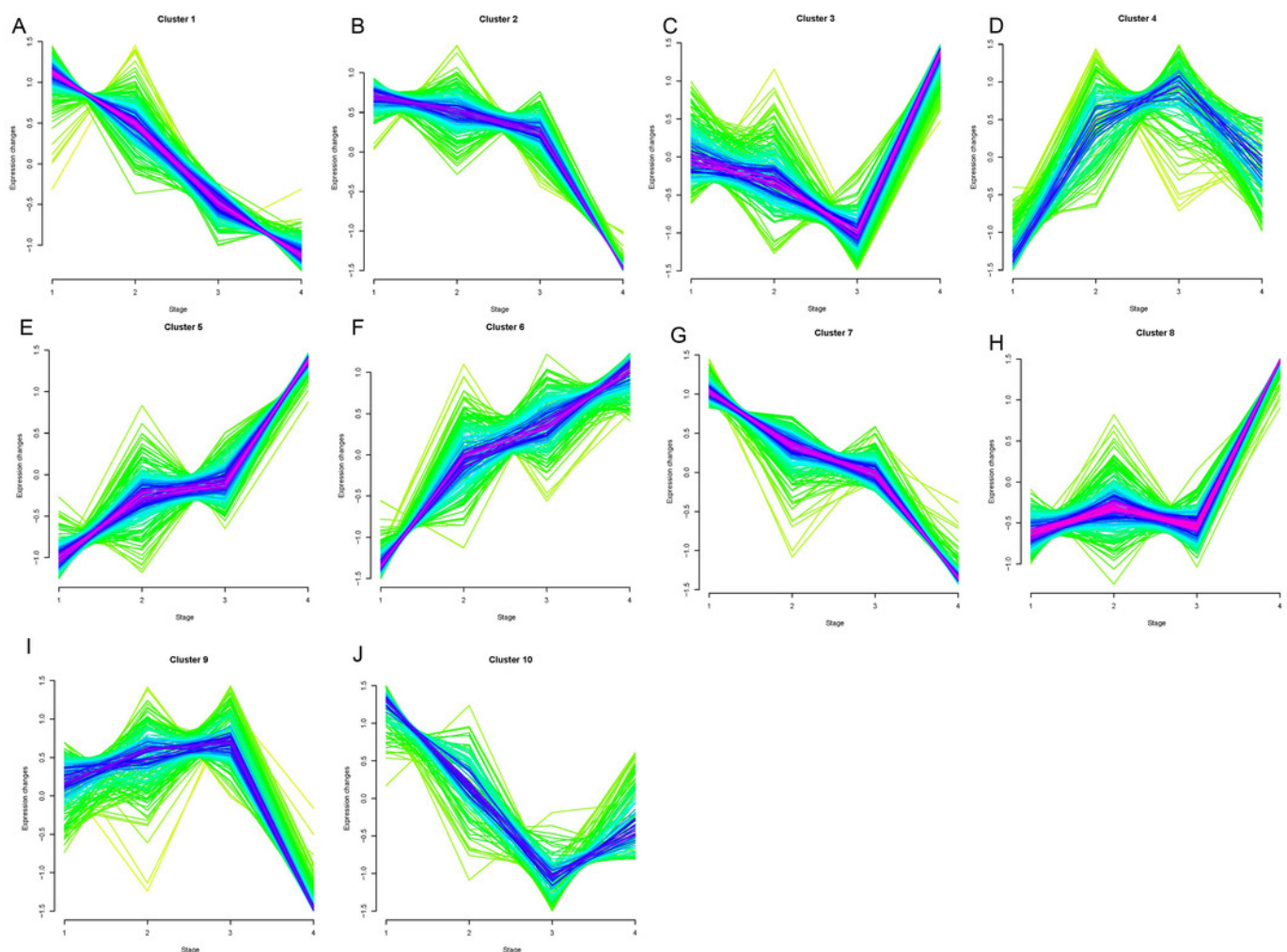



Figure 2

The integrated network of protein-protein interactions and miRNA-target gene regulatory relations

The integrated network of protein-protein interactions and miRNA-target gene regulatory relations for upregulated genes (A) and downregulated genes (B). Red circles indicate upregulated genes, green circles indicate downregulated genes, and blue diamonds indicate miRNA.

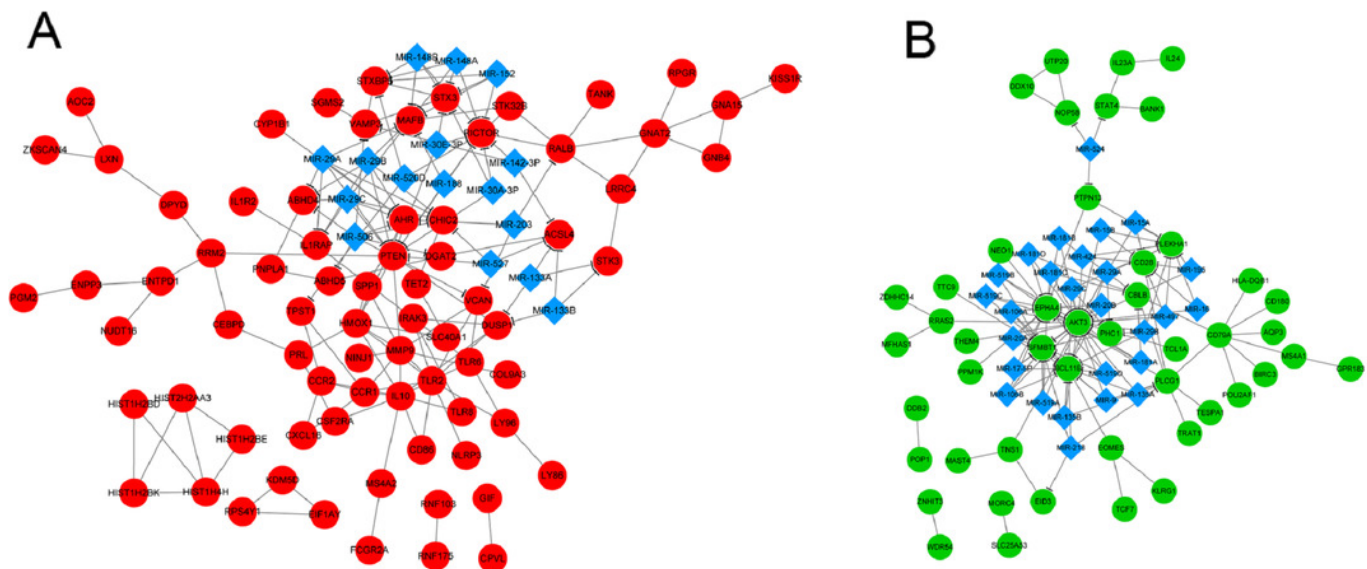
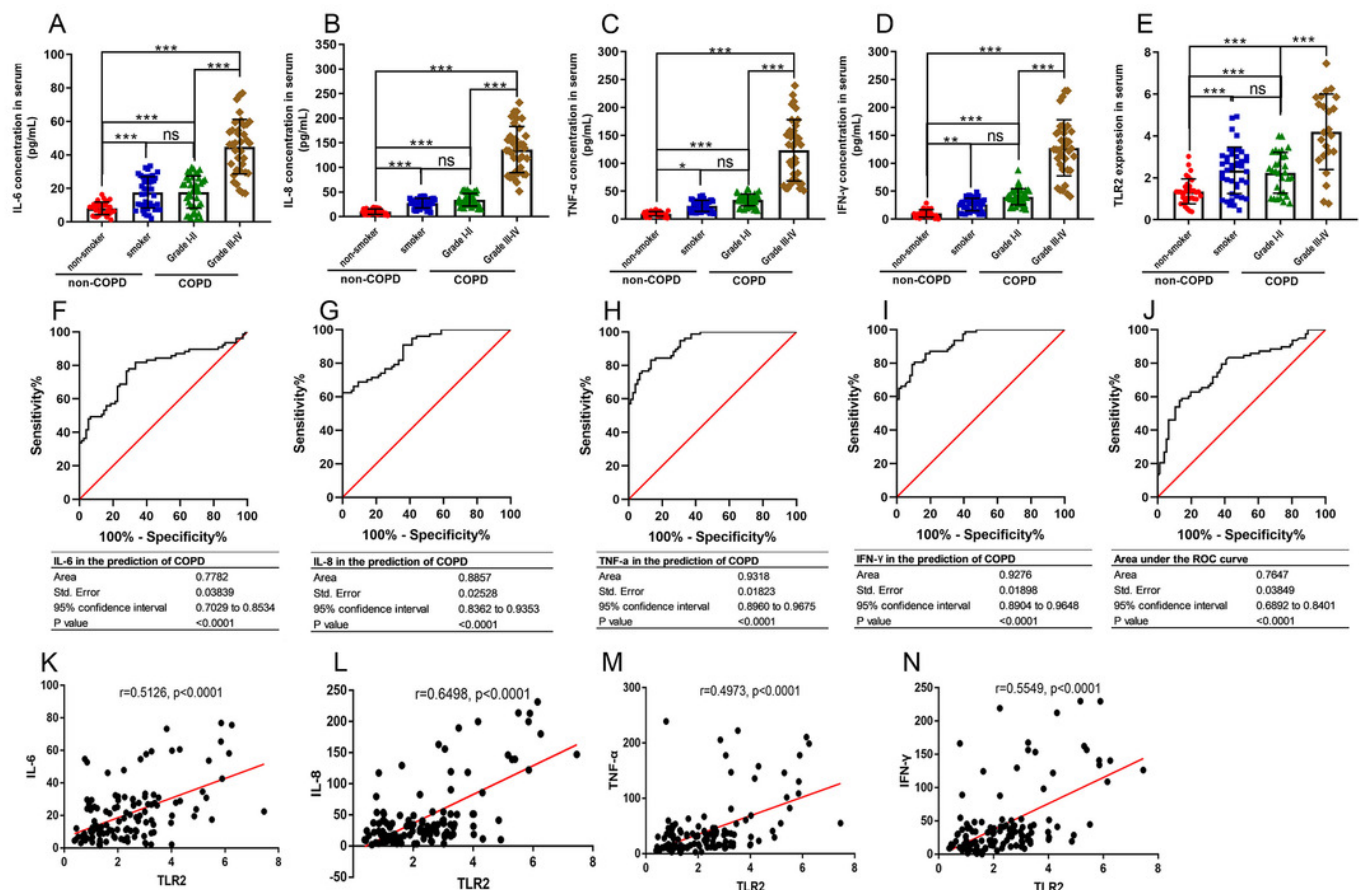



Figure 3

The relationships between TLR2 and inflammatory factors in COPD progression.

The relationships between TLR2 and inflammatory factors in COPD progression. The levels of IL-6 (A), IL-8 (B), TNF-α (C), and IFN-γ (D) in the serum and TLR2 in the peripheral blood mononuclear cells of the patients (E). The sensitivity and specificity of IL-6 (F), IL-8 (G), TNF-α (H), IFN-γ (I) and TLR2 (J) in the prediction of COPD . The TLR2 expression levels were correlated with the serum levels of IL-6 (K), IL-8 (L), TNF-α (M), and IFN-γ (N).

Upregulated genes with enriched functions of Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG)

Category	Term	Description	Count	P value
BP	GO:0006952	defense response	22	3.84E-07
BP	GO:0006954	inflammatory response	16	5.18E-07
BP	GO:0009611	response to wounding	20	7.08E-07
BP	GO:0006955	immune response	22	2.45E-06
CC	GO:0009891	positive regulation of biosynthetic process	35	0.002989929
CC	GO:0051173	positive regulation of nitrogen compound metabolic process	22	0.004929606
CC	GO:0031328	positive regulation of cellular biosynthetic process	22	0.006361208
CC	GO:0045935	positive regulation of nucleobase, nucleoside, nucleotide, and nucleic acid metabolic process	51	0.006484061
KEGG	hsa05322	Systemic lupus erythematosus	8	4.11E-04
KEGG	hsa04620	Toll-like receptor signaling pathway	6	0.012939
KEGG	hsa04060	Cytokine-cytokine receptor interaction	9	0.028464

2 BP: biological process; CC: cellular component

Downregulated genes with enriched functions of Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG)

Category	Term	Description	Count	P value
BP	GO:0046649	lymphocyte activation	11	1.89E-06
BP	GO:0001775	cell activation	12	7.79E-06
BP	GO:0045321	leukocyte activation	11	1.08E-05
BP	GO:0048534	lymphoid organ development	10	1.19E-04
BP	GO:0002520	immune system development	10	1.87E-04
BP	GO:0030097	hemopoiesis	9	3.33E-04
BP	GO:0030098	lymphocyte differentiation	6	9.42E-04
BP	GO:0042110	T cell activation	6	0.002303352
BP	GO:0002521	leukocyte differentiation	6	0.0027276
BP	GO:0042113	B cell activation	4	0.018182557
BP	GO:0030217	T cell differentiation	3	0.081968615
CC	GO:0009897	external side of plasma membrane	4	0.120963049
CC	GO:0009986	cell surface	5	0.235650749
CC	GO:0044459	plasma membrane part	24	0.033231996
CC	GO:0031226	intrinsic to plasma membrane	14	0.088759042
CC	GO:0005886	plasma membrane	34	0.099440061
CC	GO:0005887	integral to plasma membrane	13	0.136364905
KEGG	hsa04940	Type I diabetes mellitus	4	0.001194
KEGG	hsa04630	Jak-STAT signaling pathway	5	0.007356
KEGG	hsa05330	Allograft rejection	3	0.014236
KEGG	hsa05332	Graft-versus-host disease	3	0.016588
KEGG	hsa04660	T cell receptor signaling pathway	4	0.016985

2   BP: biological process; CC: cellular component

Regulatory microRNAs of gene sets

sequence	microRNA	count
Hsa-TGGTGCT	miR-29a, miR-29b, miR-29c	6
Hsa-AAGGGAT	miR-188	3
Hsa-GTGCCTT	miR-506	7
Hsa-ACACTAC	miR-142-3p	3
Hsa-TGCACTG	miR-148a, miR-152, miR-148b	4
Hsa-ACTGAAA	miR-30a-3p, miR-30e-3p	3
Hsa-GGGACCA	miR-133a, miR-133b	3
Hsa-CTTTGCA	miR-527	3
Hsa-CATTTCA	miR-203	3
Hsa-TTTGTAG	miR-520d	3
Hsa-TAGCTTT	miR-9	4
Hsa-AAGCACA	miR-218	4
Hsa-TGGTGCT	miR-29a, miR-29b, miR-29c	4
Hsa-GCACTTT	miR-17-5p, miR-20a, miR-106a, miR-106b, miR-20b, miR-519d	4
Hsa-AAGCCAT	miR-135a, miR-135b	3
Hsa-TGCACTT	miR-519c, miR-519b, miR-519a	3
Hsa-CTTTGTA,	miR-524	3
Hsa-TGAATGT	miR-181a, miR-181b, miR-181c, miR-181d	3
Hsa-TGCTGCT	miR-15a, miR-16, miR-15b, miR-195, miR-424, miR-497	3

Results of univariate logistic regression analysis

Factor	Coefficient	P value
Age	0.2796	0.0389 *
Gender	0.1545	0.2383
TLR2	0.9265	2.1068E-07 ***
IL-10	0.3250	0.1524
MMP9	0.2702	0.0059 **
HMOX1	0.3163	0.0713
CCR1	0.3503	0.0298 *
CD79A	-0.3443	0.0002 ***
PLCG1	-0.7284	0.0037 **

2 Factor: factors that may affect clinical grading; coefficient: regression coefficient (> 0, positive
3 correlation; < 0, negative correlation).
4
5

Results of multivariate logistic regression analysis

Factor	Coefficients	P value
Age	0.1069	0.4220
TLR2	0.6894	0.0027 **
MMP9	0.0374	0.7310
CCR1	0.0282	0.8695
CD79A	-0.1930	0.0450 *
PLCG1	-0.1267	0.6488

2   Factor: factors that may affect clinical grading; coefficient: regression coefficient (> 0,
3 positive correlation; < 0, negative correlation).
4