CHAPTER 4
IDENTIFICATION OF POTENTIAL ILDS BIOMARKERS
AND THERAPEUTIC TARGETS THROUGH
COMPREHENSIVE ANALYSIS OF NON-CODING RNAs

Summary
ILDs encompass nearly ~200 chronic lung disorders with impenetrable pathological mechanisms. The potential role of deregulation in many molecular mechanisms related to immunity and defense response has been studied in the progression of ILDs. Non-coding RNAs especially microRNA (miRNAs) and long-noncoding RNAs (lncRNAs) are reported for their significance in different diseases pathogenesis including ILDs. Furthermore, these miRNAs and lncRNAs cross-regulate each other towards ceRNAs (competing endogenous RNAs) activity”. These factors can be potentially used for disease diagnostics and observation. Efficient analysis or examination of these regulatory molecules could facilitate promising molecular biomarkers for ILDs. In this study, the roles of significant ncRNAs, differential expression and potential regulatory role in different biological functions and pathways have been studied. Outcomes of this study are validated using reported literature and different customary web resources. This research work revealed the noteworthy involvement of ncRNAs interaction with their target in ILDs pathogenesis. These interactions could lead to novel directions for ILDs management through less-invasive procedures.

4.1 Background
ILDs are a group of numerous chronic respiratory disorders with intricated underlying mechanism of pathogenesis [24]. A multidisciplinary discussion (MDD) among the “pulmonologist, radiologist, and pathologist” is required for the efficient diagnosis of ILDs [24]. Primary ILDs diagnostics approaches rely on the clinical symptoms, histopathological tests results and radiological patterns [103]. Availability of alternative testing such as surgical lung biopsies or bronchoalveolar lavage is difficult. Thus, HRCTs and CXRs are usually considered as a key diagnostic feature for ILDs diagnosis. However, The specificity of CXR and HRCT is low as ILDs can encompass many nonspecific and mimicking patterns associated with other lung diseases as well [103]. Additionally, many thoracic societies recommend low dose CT-protocols to avoid radiation hazards. In the past few decades, use of
less-invasive techniques in ILDs diagnostics and other disease has been popular. These techniques are proven to avoid the side effects of current diagnostic processes such as radiations, other co-morbidities and biopsies [150]. Past researches emphasized the vital role of genetic factors in ILDs etiology and prognosis [30]. Studies have shown the positive impact of molecular biomarkers for ILDs diagnosis, prognosis and treatment response to improve current diagnostic approaches [59].

Many noncoding-RNAs (ncRNAs) were reported for their regulatory roles in pathways and target genes associated with ILDs pathogenesis and diagnosis [32, 60-63]. The ncRNAs can be grouped into many subclasses according to their transcript lengths such as microRNAs (miRNAs), long non-coding RNAs (lncRNAs), siRNAs etc. [151]. Generally, miRNAs are made up of length of 20 to 22 nucleotides, and lncRNAs are made up of approximately 200 nucleotides. These ncRNAs also cross-regulated and the phenomenon is known as ceRNA activity. Many ongoing studies evidently supported the involvement of ncRNAs in the dysregulation of the immune system and defense response. These factors are reported to have potential intervention in many disease progression and pathogenesis including ILDs [61, 64-67].

Role of the differentially regulated miRNAs was already reported in signaling pathways and several biological processes and disease pathogenesis [46]. These miRNAs were already identified to increase lung fibroblast susceptibility and severity [47]. Additionally, few miRNAs are recognized as promising less-invasive therapeutic/biomarkers targets for ILDs (e.g. IPF) and cancers [10, 69]. The lncRNAs are also important for alteration of regular biological processes, genes dysregulation and disease progression to different stages [152, 153]. Role of lncRNAs in adaptive and innate immunity was already reported [61]. Integrated analysis of lncRNAs and disease association were found to promote lncRNAs-based management for many respiratory diseases [72, 73]. Additionally, an association of the ceRNA activities, their post-transcriptional regulation with cancers and fibrotic lungs progression were also implicated [32, 70, 71].

“The reported outcomes give clear insight on the comprehensive identification of miRNAs, lncRNAs, pathways and ceRNAs that affect molecular pathways, which can be characteristic factors in the etiology of ILDs [32]. The immune mediation and defense response have shown crucial role in ILDs pathogenesis and progression in the previous objective of this research work. Immunological processes were found to have associations with ILDs candidate genes, associated pathways and regulatory networks [65, 126]. The role of ncRNAs
in immunity was also established through previous objective and other supported literature [126]. Thus, the role of biomarkers related to autoimmunity has been well identified for ILDs pathogenesis and diagnosis [61]. Hence, miRNAs, lncRNAs, and ceRNAs are anticipated to be involved in ILDs progression. Nevertheless, very limited miRNAs are in clinical practice as biomarkers or used as therapeutic targets for ILDs subtypes [47].” [154]

“Therefore, functional characterization and verification of ncRNAs and their vital role as ILDs genetic biomarkers and potential therapeutic targets can together formulate significant progress in ILDs management [61, 151]. Considering these factors, this chapter emphasize on comprehensive study of ncRNAs and their associations with ILDs. Enrichment analysis and functional annotation of these ncRNAs, pathways and target genes are also performed. Integrated expression analysis of miRNA is also performed to identify the crucial role of miRNAs in different ILDs. The outcome of this objective would assist in the identification of promising biomarkers and therapeutic targets for ILDs.” [154]

4.2 Materials and Method

The stepwise systematic analysis was performed to identify potential miRNA and lncRNAs as therapeutic targets/biomarkers (Figure 4.1). Subsequent sections provide a better idea about the process, data collection, and their analysis.

4.2.1 Identification of ncRNAs and their association with ILDs

The miRNAs and their most recent information are collected from miRBase [155]. The lncRNAs are collected from lncRNA databases (such as IRNdb [67] and lncRNAdb [156]), for “Homo sapiens”. Identification of ILDs association with the ncRNAs is performed with the help of ILDgenDB [126].

4.2.2 Association analysis of ncRNAs and ILDs-pathways

4.2.2.1 Analysis of miRNAs associations with GO and pathways

Pathways datasets are downloaded from WikiPathways, KEGG, and Reactome. A consistence annotation is produced after removal of redundancy. The miRWalk2.0 comprehensive atlas and literature data are used to perform miRNAs association with pathways [157]. Top 10 ILDs associated pathways and their interactions with miRNAs are studied and identified (FIG 2). Analysis of biological processes, molecular mechanism, and cellular component is done using Gene ontology (GO) enrichment analysis of these miRNAs (Figure 4.1). A cut-off P-value of <0.05 is used to obtain significant enrichment results.
4.2.2 Differential expression analysis of miRNAs

To perform miRNAs expression analysis, Gene expression omnibus (GEO) datasets are used to extract the expression profiles of miRNAs.

“Advance query options were used with keywords like *miRNAs, interstitial lung diseases, idiopathic pulmonary fibrosis*, etc. After checking the availability and type of data available, finally, a total of six datasets (GSE13316, GSE8555, GSE75647, GSE21394, GSE81293, and
GSE27430) are selected based on study design. The GEO2R is used to compare groups of samples (such as control v. disease) for differential expression. A combination ranking approach (proposed by McCarthy et al. [158]) by combining log fold change (LogFC) value (>1.5 for up-regulation and < -1.5 for down-regulation) and P-Value (<0.05) is applied to determine the biologically significant differentially regulated miRNAs.” [154]

4.2.2.3 miRNAs, lncRNAs and pathways association analysis

To analyze the impact of ncRNAs crosstalk, the starBase v2.0 is used [159]. The ILDs association with pathways and lncRNAs are identified using ILDgenDB [126], IRNdb [67] and lncRNAdb [156]. The significantly associated lncRNAs are considered (Figure 4.1) for further analysis.

4.2.2.4 Validation

All the significant associations such as miRNAs/lncRNAs-pathways and miRNAs-lncRNAs are validated with the help of Pubmed. This literature-based approach has established confidence by validating the interactions in ILDs pathogenesis. A systematic literature survey is done for the evaluation of ncRNAs–target gene interaction responsible for ILDs. Advanced search option of PubMed is used to retrieve the potential associations of ILDs:

"microRNAs"[MeSH Terms] OR "miRNAs"[All Fields]) AND "interstitial lung disease "[All Fields]

Only the most significant terms are presented here. Several other query terms are also used such as “pathway name”, “gene name”, “ILD name”, etc. The final conclusion is established after careful screening of the literature. Top-ranked associations have produced few potential ncRNAs that could be used as ILDs specific biomarkers. Few of these biomarkers were already experimentally verified in the literature.

4.3 Results and Discussion

Non-coding RNAs were found to have an important role in disease mechanism and pathogenesis, including ILDs [69, 159-161]. However, the majority of the studies were conducted mainly for cancers, and only a small number of ncRNAs-based clinical trials are carried out for ILDs and other disease diagnosis and prognosis [61]. This association analysis of ncRNAs along with a role in mRNAs regulation can result in less-invasive ncRNAs based diagnostic biomarker for pulmonary diseases like ILDs [10, 59, 151, 162]. To develop improved indulgent on these regulatory elements associations and their role in ILDs,
integrated analyses of ncRNAs, genes and pathways are performed and provided in a comprehensive manner (Figure 4.1).

In-house PERL scripting is used for identifying the ncRNAs and ceRNAs associations with ILDs specific genes and pathways. Many relevant bioinformatics tools, web resources are used for the enrichment analysis and literature from Pubmed are used for the results validation.

“In total, 2701, 669 and 4568 associations are identified in miRNA-gene-pathways, IncRNAs-pathways and miRNA-IncRNA (ceRNAs), respectively. The miRNAs interacting with many genes can alter their expressions and are believed to have a potential role as biomarkers and therapeutic targets in different studies [153, 163]. Potential diagnostics/therapeutic targets were proposed based on their implication in ncRNAs-pathways associations, IncRNAs-miRNAs associations, and dysregulated miRNAs (Figure 4.1).” [154]

4.3.1 Association analysis among ncRNAs, ILD genes, and pathways

4.3.1.1 miRNAs and pathways associations

Role of miRNAs in the alteration of several regulatory pathways and genes expression in ILDs cases are already studied [32]. These associations are identified by validating the curated miRNAs, genes and pathways associated with ILDs (Figure 4.2). In total, 127 ILD-specific genes reported in ILDgenDB are found to be associated with 228 different pathways. These pathways were retrieved from “KEGG”, “Reactome” and “WikiPathways” knowledgebase [126].

“Total of 10 different ILDs pathways and associated ncRNAs are identified and demonstrated to prove their significant role in ILDs pathogenesis (Table 4.2 and Figure 4.2). Outcomes of this study indicated that majority of the miRNAs are involved in inflammatory pathways, which suggest the potential role of inflammation in disease pathogenesis. Immunological processes are also identified for their potential role in ILDs pathogenesis. ncRNAs, ceRNAs and associated pathways are provided in Figure 4.2. Pathways such as TGF-β Signaling, Chemokine signaling, Cytokine-cytokine receptor interaction are identified as highly associated pathways to ncRNAs. Many pathways such as JAK/STAT, MAPK and TGF-β signaling pathways were already studied for their role in fibroblast activation which leads to ILDs [67, 162]. These results substantiate the impending role of miRNAs in immunological processes. The miRNAs like miR-335-5p, miR-1, etc. have shown the
significant association to ten different ILDs pathways. These miRNAs can be explored for their significant roles in future studies.” [154]

![Figure 4.2](image)

**Figure 4.2** “Number of non-coding RNAs associated with pathways predicted to be involved in ILDs pathogenesis” [154]

### 4.3.1.2 The miRNAs’ differential expressions

Role of differently expressed miRNAs in ILD pathogenesis is already studied. Rapid and precise measurability of miRNAs along with their highly stable nature makes them a promising molecular biomarker. miRNAs are proved to have significant roles not only for diagnosis but also in different phases of disease mechanism [48, 164]. To further explore the potential role of differentially expressed miRNAs in ILDs, seven different gene expression omnibus (GEO) datasets are analyzed (Table 4.1). The miRNAs which passed through the combined ranking were then analyzed for their potential roles in ILDs pathways (Section 4.3.1.1).

“Several miRNAs (44) that are identified as significant in this study were also reported in previous studies for their promising roles in lung disease. For example, hsa-miR-1821 and has-miR-199a are studied for their potential role in ILDs and IPF, respectively [163][165]. Similarly, hsa-miR-363 and miR-588 were studied for their role in lung cancer [166, 167].
all the seven datasets of this study, mir-575, mir-4417, mir-665, etc. have shown the significant up-regulation in diseased condition. This altered expression of miRNAs makes them promising candidates for ILD-specific biomarkers. Similar studies could be performed on other ILDs to identify subtype-specific biomarkers. Therefore, differential expression analysis of miRNAs might be a great tool for ILDs diagnostics.” [154]

4.3.1.3 lncRNA and ILD pathways association

Reported case studies have shown the negative impact of lncRNAs in ample of dysregulated pathways. These results suggested the use of lncRNAs as a novel tool for diagnostics, prognostics, and treatment of different cancers [69]. Identification of regulatory lncRNAs and their pathway association analysis provide links to promising lncRNAs that are involved in various ILDs pathways (Figure 4.2.).

“Total of seventy-three lncRNAs is found to interact with 91 ILDs genes and 196 pathways, which leads to 669 interactions among lncRNAs, pathways, and genes. Some potential lncRNAs such as XIST (5), MALAT1 (4), CTA-204B4.6 (5), etc. are mapped with many immunological pathways and ILDs genes. The detailed summary of lncRNAs and miRNAs association is given in Table 4.2. The genes targeted by these lncRNAs were found to be associated with ILDs pathogenesis (Figure 4.3), and they have exhibited the lncRNA-mediated differential expression in host defense and inflammation [161, 168]. Promising roles of these lncRNAs in a number of pathophysiological processes is already suggested in ongoing studies [61]. However, the involvement of lncRNAs in the pathogenesis of ILDs is not well known. Identified lncRNAs mediated regulation of genes and pathways related to ILDs could provide a greater insight into the disease pathogenesis; therefore, these associations could be a good therapeutic or diagnostic target for ILDs.” [154]
4.3.1.4 Interaction between miRNA- lncRNA (ceRNA)

“The crosstalk between miRNA and lncRNA knew as “ceRNAs” interrupt the miRNAs bindings to the target [161]. Regulatory role of this ceRNAs activity was already studied in ILD like IPF [32, 168]. Total of 108 miRNAs associated to ILDs and 909 lncRNAs have produced the 4568 ceRNAs interactions in this study. Targeted ILDs genes used in this study are literature curated and have a potential role in disease pathogenesis and progression (Table 4.1). Table 4.1 also provides a number of lncRNAs important for ceRNAs interactions, and the number of literature verified pathways involved in ILDs. Identified top ten miRNAs associated pathways related to ILDs are namely, Asthma, Cytokine-cytokine receptor interaction, Chemokine signaling pathway, Inflammatory response pathway, Cytokines and inflammatory response, MAPK signaling pathway, JAK/STAT signaling pathway, Toll-like receptor signaling, TGF-β signaling pathway and Wnt signaling pathway. The potentially significant lncRNAs with the higher number of miRNAs interactions (ceRNAs interaction) are identified (Table 4.2). These top-ranked lncRNAs identified as biomarkers are associated with five different pathways involved in ILDs. These five pathways are namely, MAPK signaling, Cytokine-cytokine receptor, TGF-β signaling, Toll-like receptor signaling, and Chemokine signaling. The significant involvement of ceRNAs activities in cellular immune responses, reduction in anti-thrombogenic agents, inflammatory/pro-inflammatory pathways, growth factors, etc. is already reported [63, 160].” [154]
### Table 4.1 The miRNAs predicted as potential biomarkers or therapeutic targets

<table>
<thead>
<tr>
<th>S. No.</th>
<th>miRNA</th>
<th>Targeting ILD genes</th>
<th>No. of associated lncRNAs</th>
<th>Number of associated Pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hsa-miR-1</td>
<td>AP3B1, BDNF, CCL2, EDN1, F2, HMOX1, IL6, IL8, ITGA3, TLR4</td>
<td>33</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>hsa-miR-124-3p</td>
<td>BMP6, CAMP, CAV1, CCL2, DNAH5, EDN1, HMOX1, IL6, IL8, ITGA3, NME4, PGM1, PIK3C2A, PLA2G7, SERPINE1, SERPINH1</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>hsa-miR-125b-5p</td>
<td>BMP1B, IL1RN, ADM, BMP1B, SLC9A3R2, STAT3, VDAC1, VDR</td>
<td>26</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>hsa-miR-155-5p</td>
<td>CAT, CCL2, EDN1, F5, FGF2, IFNGR1, IL6, IL8, NEU1, SLC9A3R2, SMAD3, STAT3, TTF1</td>
<td>46</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>hsa-miR-21-5p</td>
<td>BMP2, DERL1, FAS, FASLG, PIK3C2A, PLAT, SOD3, STAT3, TGF-β1, TLR4</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>hsa-miR-26b-5p</td>
<td>ADM, BMP2, BMP2R, C3, CAV1, CCL2, CCL7, CXCL9, CXCR1, DERL1, FASLG, FGF23, GUSB, HMOX1, IFNG, IFNGR2, INHA, ITGA3, MPP8, PDGF, SERPINE1, STX1A, TLR1</td>
<td>48</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>hsa-miR-335-5p</td>
<td>ABCA3, ACE, ATF6, BMP2, CCR4, CCR7, CD14, CD27, CFTR, CLCA4, CXCL9, CXCR1, CXCR2, CYP2E1, FGF23, GC, GUC2A2B, HMOX1, HSPG2, IL17A, IL1A, IL4, IL5, IL6, IL8, LTB1P2, P2RY2, PDE5A, PIK3C2A, PLA2G7, PLAT, SERPINE1, SHH, SLC6A4, SMAD3, STX1A, TERT, THBD, TLR1, TLR2, TLR4, TNC, TREAT, TSC1, VIP</td>
<td>32</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>hsa-miR-93-5p</td>
<td>BMP2, DCTN4, STAT3, TOLLIP</td>
<td>74</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>hsa-miR-19a</td>
<td>BMP2, TLR2, BMP2R, DERL1, TLR2</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>hsa-miR-15a-5p</td>
<td>IFNG, PLA2G2D</td>
<td>145</td>
<td>7</td>
</tr>
</tbody>
</table>

### Table 4.2 Top ranked lncRNAs predicted as potential biomarkers or therapeutic target

<table>
<thead>
<tr>
<th>S. No.</th>
<th>lncRNAs</th>
<th>Association with No. of miRNAs</th>
<th>Association with No. of Pathways</th>
<th>Association with No. of ILD-Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>XIST</td>
<td>87</td>
<td>5</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>CTA-204B4.6</td>
<td>58</td>
<td>5</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>MALAT1</td>
<td>42</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>KCNO10T1</td>
<td>41</td>
<td>3</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>ZNF518A</td>
<td>40</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>NEAT1</td>
<td>34</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>7</td>
<td>SNHG16</td>
<td>33</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>8</td>
<td>HCG18</td>
<td>32</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>OIP5-AS1</td>
<td>29</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>10</td>
<td>RP11-361F15.2</td>
<td>29</td>
<td>5</td>
<td>21</td>
</tr>
</tbody>
</table>

Imperative role of these processes in ILDs was also established [126]. Further analyses of the proposed ceRNAs interactions may assist in functional similarity networks and ILDs-lncRNAs associations, which may assist to comprehend the mechanism of ILDs [69]. Top 10
ceRNAs interactions obtained from 10 different disease candidate genes and pathways are anticipated as potential biomarkers (Table 4.1, Table 4.2).

4.3.2 Association of ILDs genes and ncRNAs involved in immune processes

Association analysis of miRNAs, GO and ILDs genes are performed to identify the miRNAs’ regulatory roles in various biological processes. Analysis of miRNAs with GO terms yielded a total of 2364 interactions. The miRNAs such as miR-335-5p and miR-26b-5p are found to be associated with the highest number of GO terms, which are 20 and 9, respectively. Hypoxia and cytokine activities related GO terms are predominant in this study, where former is the characteristic symptom of lung disorder like ILDs [169]. Inflammation, wounding and defense response is the predominant biological processes identified through miRNAs-GO term analysis. These outcomes have suggested a lucid indication that ncRNAs are associated with immunological processes in ILDs pathogenesis.