



Cartography of Pathway Signal Perturbations Identifies Distinct Molecular Pathomechanisms in Malignant and Chronic Lung Diseases

Arsen Arakelyan^{1,2*}, Lilit Nersisyan^{1,2}, Martin Petrek³, Henry Löffler-Wirth⁴ and Hans Binder⁴

¹ Group of Bioinformatics, Institute of Molecular Biology, National Academy of Sciences, Yerevan, Armenia, ² College of Science and Engineering, American University of Armenia, Yerevan, Armenia, ³ Laboratory of Immunogenomics, Department of Pathological Physiology, Faculty of Medicine and Dentistry, Institute of Molecular and Translational Medicine, Palacky University Olomouc, Olomouc, Czech Republic, ⁴ Interdisciplinary Centre for Bioinformatics, University of Leipzig, Leipzig, Germany

OPEN ACCESS

Edited by:

Nicola Mulder,
University of Cape Town, South Africa

Reviewed by:

Gabriele Bucci,
San Raffaele Scientific Institute, Italy
Jie Guo,
Chinese Academy of Sciences, China

*Correspondence:

Arsen Arakelyan
aarakelyan@sci.am

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Genetics

Received: 11 February 2016

Accepted: 20 April 2016

Published: 06 May 2016

Citation:

Arakelyan A, Nersisyan L, Petrek M, Löffler-Wirth H and Binder H (2016) Cartography of Pathway Signal Perturbations Identifies Distinct Molecular Pathomechanisms in Malignant and Chronic Lung Diseases. *Front. Genet.* 7:79. doi: 10.3389/fgene.2016.00079

Lung diseases are described by a wide variety of developmental mechanisms and clinical manifestations. Accurate classification and diagnosis of lung diseases are the bases for development of effective treatments. While extensive studies are conducted toward characterization of various lung diseases at molecular level, no systematic approach has been developed so far. Here we have applied a methodology for pathway-centered mining of high throughput gene expression data to describe a wide range of lung diseases in the light of shared and specific pathway activity profiles. We have applied an algorithm combining a Pathway Signal Flow (PSF) algorithm for estimation of pathway activity deregulation states in lung diseases and malignancies, and a Self Organizing Maps algorithm for classification and clustering of the pathway activity profiles. The analysis results allowed clearly distinguish between cancer and non-cancer lung diseases. Lung cancers were characterized by pathways implicated in cell proliferation, metabolism, while non-malignant lung diseases were characterized by deregulations in pathways involved in immune/inflammatory response and fibrotic tissue remodeling. In contrast to lung malignancies, chronic lung diseases had relatively heterogeneous pathway deregulation profiles. We identified three groups of interstitial lung diseases and showed that the development of characteristic pathological processes, such as fibrosis, can be initiated by deregulations in different signaling pathways. In conclusion, this paper describes the pathobiology of lung diseases from systems viewpoint using pathway centered high-dimensional data mining approach. Our results contribute largely to current understanding of pathological events in lung cancers and non-malignant lung diseases. Moreover, this paper provides new insight into molecular mechanisms of a number of interstitial lung diseases that have been studied to a lesser extent.

Keywords: high-throughput gene expression, biological pathways, pathway signal flow, self-organizing maps, chronic lung diseases, lung cancers

INTRODUCTION

High-throughput gene expression profiling has found wide applications in many areas of lung pathology research, diagnostics, and treatment (Campbell et al., 2011). It has been used for development of biomarker panels allowing for accurate discrimination between diseases, such as chronic obstructive pulmonary disease (COPD), idiopathic pulmonary fibrosis (IPF), and lung cancers (Selman et al., 2006; Wang et al., 2008). The study of gene expression signatures largely contributed to better understanding of molecular pathology of lung diseases (Cancer and Atlas, 2012; Thakur et al., 2014), and to identification of new disease subclasses/entities (Bhattacharjee et al., 2001; West et al., 2012; Li et al., 2014). It also provided new approaches to diagnostics (Buettner et al., 2013; DePianto et al., 2015), and helped to suggest novel therapeutic compounds (Campbell et al., 2012; Gerber et al., 2015). However, most of these studies have been performed in a gene centered fashion, where biological function mining was limited to identification of differentially expressed genes and to enrichment analysis of the obtained gene lists in Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) or any other functional annotation database (Arakelyan et al., 2013). Such simple functional annotations of differentially expressed genes might however have overlooked important information, such as interactions between the components in a biological system and their functional consequences. In this regard, explicit assessment of activity regulation in biological pathways by combining expression data of genes with knowledge about the interactions between their products would represent a more straightforward approach. Biological pathways are sequences of physical molecular interactions that guide information propagations (also called signal flow) leading to regulatory consequences for cell function. Pathways are often branched and have more than one target processes, meaning that activation of different sets of genes within the same pathway may lead to different outcomes. Moreover, genes can be involved in more than one pathway, because different pathways may share common branches, sources, and sinks (Daigle et al., 2010). It has been recognized that pathogenesis of diseases usually involve perturbations occurring at pathways level (Logan and Nusse, 2004; Courtois and Gilmore, 2006). Thus, the explicit analysis of pathway activity deregulation is expected to lead to better understanding of molecular pathomechanisms of diseases.

Pathologic characteristics of lung diseases are very complex due to the involvement of environmental and genetic interactions (Pouladi et al., 2015) and do not always reflect the underlying molecular mechanisms. Dysfunction of a single gene may contribute to multiple lung diseases leading to development of different phenotypes or, vice versa, similar disease phenotypes can be caused by dysfunctions of different genes and (Lewis et al., 2008; Pennings et al., 2008). For example, the WNT

signaling pathway has been shown to be oppositely involved in the pathobiology of COPD and IPF. Despite this difference, both diseases end up with alveolar senescence and lung “premature aging” (Chilosi et al., 2012). Contrarily, typical interstitial lung diseases, such as sarcoidosis and IPF, largely share similarly deregulated genes, especially in extracellular signaling pathways, meanwhile demonstrating significant differences in downstream signaling pathways (Leng et al., 2013).

In this study we have applied a pathway-centered method for mining high-throughput gene expression data aimed at extracting knowledge about pathobiology of a wide spectrum of malignant and chronic lung diseases by providing an extended systems view on pathway deregulation states. We, used a previously developed and intensively benchmarked data mining approach for gene expression analysis using self-organizing maps machine learning (Wirth et al., 2011) in combination with the pathway signal flow (PSF) algorithm, which maps gene expressions to pathway topologies to assess the strength of the signals that propagate along the pathways (Arakelyan et al., 2013; Binder et al., 2014; Nersisyan et al., 2015). It allowed for combining lung diseases into groups of common pathway perturbations and provided new insights on shared and specific pathomechanisms of lung diseases.

MATERIALS AND METHODS

Data and Sources

Six datasets on selected lung diseases extracted from the Gene Expression Omnibus (GEO) public repository were used (Edgar et al., 2002; Barrett et al., 2011). The datasets were downloaded from GEO in the form of Series Matrix Files, which contain metadata about samples, array calibration and normalization methods and actual normalized data, stored in the form of gene expression matrices, where probe IDs are rows and samples are columns (see Supplementary Material Data Sheet 1, Table S1). From these datasets only samples obtained from untreated patients were used in downstream analyses. In total, 948 diseased and normal lung samples constituting 21 disease groups and one healthy lung tissue were included in the analyses (Table 1). The analysis workflow is presented in Figure 1. It was performed using the R-program package oposSOM (Löffler-Wirth et al., 2015) and a series of stand-alone R-scripts (Supplementary Material Data Sheet 2).

Data Preprocessing

For each dataset, microarray probe IDs were converted into Entrez Gene IDs. Microarray probes that did not match any known Entrez ID were discarded. Multiple probe sets for the same gene were averaged. Next, gene expression values in each dataset were inspected for data transformation type (log₂, log₁₀ transformation) using a modified version of the “GEO2R log autocheck” script (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) and were brought to a linear scale. Then we have converted all values to fold change (FC) respective to the mean expression value of the controls (healthy samples) included in the same dataset. These steps cast the data in all datasets to a common scale and type.

Abbreviations: ANOVA, analysis of variance; FC, Fold Change; FDR, False Discovery Rate; GEO, Gene Expression Omnibus; GO, Gene Ontology; ICA, Independent Component Analysis; ILD, Interstitial Lung Disease; KEGG, Kyoto Encyclopedia of Genes and Genomes; PSF, Pathway Signal Flow; SOM, Self-Organizing Maps.

TABLE 1 | Lung diseases and sample sizes used in this study.

N	Class	Lung disease and control group	Abbreviation	Number of samples
1	Healthy control	Healthy Control	HC	170
2	Chronic lung diseases	Tuberculosis	TB	5
3		Acute interstitial pneumonia	AIP	1
4		Cryptogenic organizing pneumonia	COP	3
5		Chronic obstructive pulmonary disease	COPD	220
6		Desquamative interstitial pneumonia	DIP	4
7		Fibrosis unknown	FU	10
8		Hypersensitive pneumonitis	HP	30
9		Interstitial lung disease other	ILD_OTHER	9
10		Interstitial lung disease unknown	ILD_UNK	14
11		Nonspecific interstitial pneumonia	NSIP	14
12		Respiratory bronchiolitis-interstitial lung disease	RB-ILD	12
13		Pulmonary sarcoidosis	SARC	6
14		Idiopathic pulmonary fibrosis	UIP_IPF	157
15		Lung cancers	Lung adenocarcinoma	ADC
16	Lung cancer basaloid		BAS	39
17	Lung cancer carcinoid		CARCI	24
18	Large cell carcinoma		LCC	3
19	Lung cancer large cell neuroendocrine		LCNE	56
20	Lung cancer other		LCO	4
21	Small cell lung carcinoma		SCC	21
22	Squamous cell carcinoma		SQC	61

Pathway Signal Flow Calculation

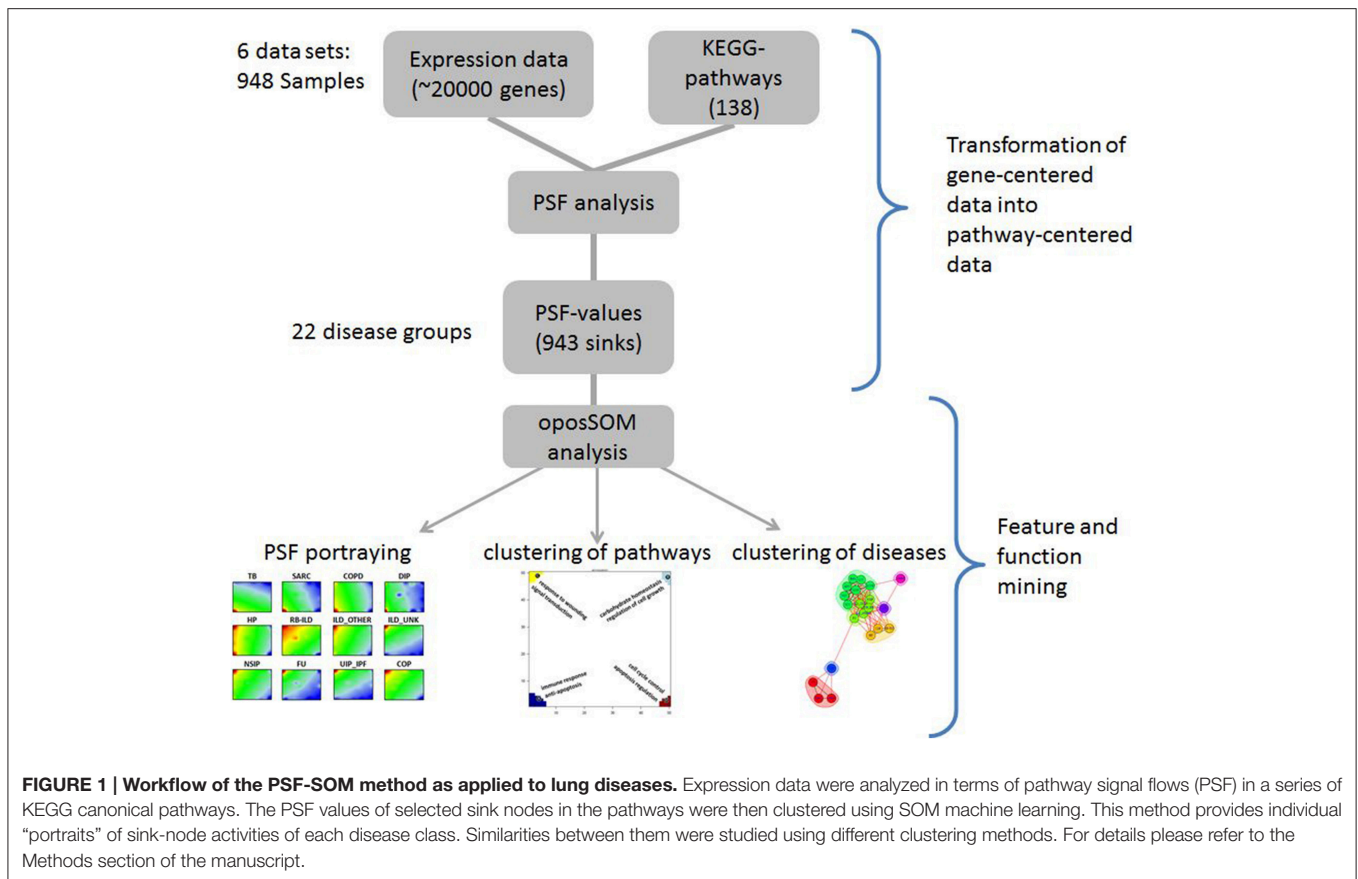
Assessment of pathway activity deregulation was performed using the PSF algorithm described in detail elsewhere (Arakelyan et al., 2013; Binder et al., 2014; Nersisyan et al., 2015). Briefly, this algorithm computes the strength of the signal propagated from the pathway input to the output through interactions of pathway component genes, based on their fold change expression values. Human reference pathway maps were retrieved from the KEGG Pathway database (Kanehisa, 2002). From 291 KEGG Pathways, 138 KEGG metabolic, signaling and organismal pathway maps were selected by exclusion of disease and drug response specific pathways, as well as pathways lacking annotated interactions (for example ko03010, Ribosome). FC values for each gene were mapped to corresponding pathway nodes, and were averaged if a node contained more than one gene. After this step an input signal of unity was applied to the pathway source nodes. Then PSF values were calculated at the output nodes. PSF algorithm is calibrated in a way that gene expression of $FC = 1$ at all nodes (normal gene expression) produces $PSF = 1$ values. Values of

PSF less than unity refer to pathway de-activation, while $PSF > 1$ indicates pathway activation.

Pathways are usually highly branched and can have multiple inputs and outputs, called sources and sinks, respectively. The sinks can be associated with defined biological function and they usually weakly depend on the activity of other sinks in the same pathway. For example, the B cell receptor (BCR) signaling pathway gets input signals from ligands and provides output signals via the sinks *Rac*, which is associated with *Regulation of actin cytoskeleton pathway*, *NFAT*, *API*, and *NFkB*, which lead to expression of genes involved in immune response, and *GSN3B*, which is a protein phosphorylating enzyme (Figure 2A). Different disease-specific expression values of genes involved in this pathway may lead to different PSF values at the sink nodes and thus, also to different activities of the associated molecular processes. For example, the *Rac*-sink of the B cell receptor pathway ends up with low activities in lung cancers (red bars Figure 2B) and high activities in chronic lung diseases (green); whereas the *NFkB*-sink shows nearly an anticorrelated activity profile compared to that of *Rac* (Figure 2B). In total, for each sample included in the analyses, we calculated PSF values for 943 sinks in 138 pathways (on average, seven sinks per pathway).

Pathway Signal Flow—Self Organizing Maps Analysis (PSF-SOM)

The self-organizing maps (SOM) method was developed in the early 1980'ties by T. Kohonen (Kohonen, 1982). First applications to microarray gene expression data were published in 1999 (Tamayo et al., 1999; Törönen et al., 1999), to cluster gene expression profiles into a predefined small number of groups of similarly expressed genes. A sample-centered clustering approach was realized shortly after providing a visual identity of the expression landscapes of each sample (Golub et al., 1999; Covell et al., 2003; Eichler et al., 2003). Several methodical improvements of the SOM-technique were developed enabling more flexible learning and mapping in different applications (see Binder and Wirth, 2014 and references cited therein). We pursued a special implementation of SOM that we called SOM portraying (Wirth et al., 2011; Hopp et al., 2013). It visualizes the “landscapes” of large scale -omics data in a comprehensive and intuitive fashion. This method combines both the sample- and gene-centered perspectives to decode molecular patterns within a two-dimensional image (Huang et al., 2005). It emphasizes easy sample-to-sample comparison and the identification of important characteristics by direct visual inspection. We amended this method by a sophisticated analysis workflow for downstream analyses including differential feature selection, diversity analysis, function mining, and class discovery in high-dimensional molecular-biological data (Wirth et al., 2011; Hopp et al., 2013). In a series of case studies we applied this method to disentangle cancer subtypes (Hopp et al., 2013), for combined mRNA and miRNA expression analysis (Wirth et al., 2014), and to analyze proteome (Wirth et al., 2012b; Binder et al., 2014), genomic (Binder and Wirth, 2014), DNA-methylation (Hopp et al., 2015), and epigenome (Steiner et al., 2012) landscapes in the context of cancer, healthy

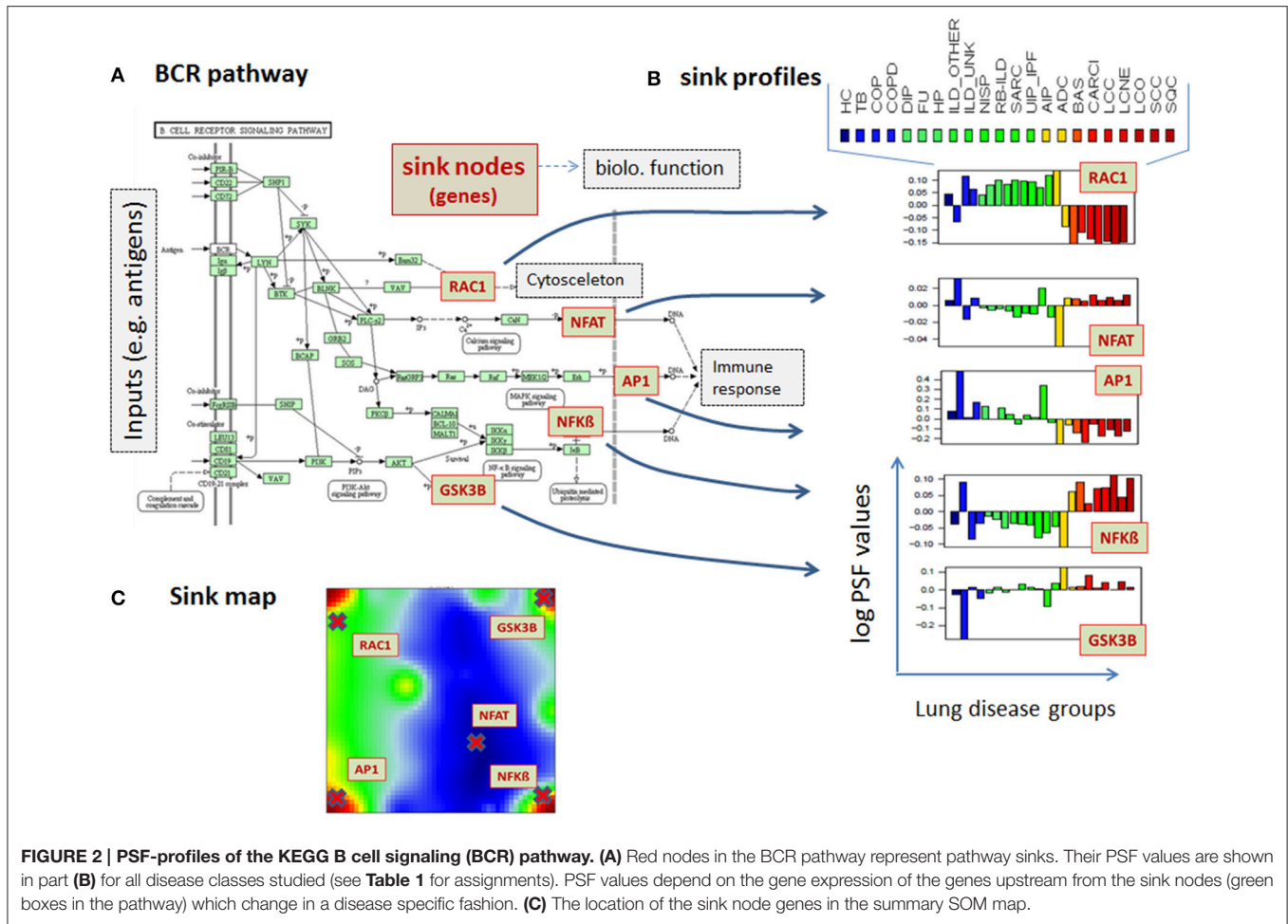


populations, microorganisms, tissue, and cell experiments. Strengths of SOM-portraying applications to molecular omics data are dimensionality-reduction and intuitive visualization of high-dimensional data, “personalized” views and detection of outliers and methodical biases in the data and their classification owing to re-weighting in SOM-space (Binder et al., 2015). Disadvantages often ascribed to SOM machine learning such as high computational efforts, loss of resolution, and the problematic choice and adjustment of parameters defining SOM space (e.g., size, topology, boundary, and initiating conditions) and the non-linear data scaling are of minor importance in our applications to molecular *-omics* data. In oposSOM package typical SOM runs require from minutes to, at maximum a few hours of computer time. SOM-space was optimized in extensive pre-studies (see Binder and Wirth, 2014 and references cited therein), performed in a similar and thus comparable way in a large number of applications. Here we combined the SOM portraying method with PSF analysis to achieve a pathway-centered view on pathomechanisms of lung diseases. PSF analysis generated profiles of PSF values for all 943 sinks of all the pathways across all the samples studied. The PSF values of each sink were centralized with respect to the sink-wise global mean over all samples to focus rather on changes of PSF-values than on absolute PSF levels.

We used SOM machine learning implemented in “som” R package as a core for oposSOM package (Löffler-Wirth et al., 2015) to disentangle the multivariate structure inherent in the

PSF-data (Wirth et al., 2011). The SOM algorithm arranged PSF profiles onto two-dimensional 50×50 grids (maps), where similar profiles are combined into micro-clusters called meta-PSFs, by analogy with meta-genes in the original application of the algorithm to gene expression data. The similarity of PSF-profiles and SOM node weights was calculated using the Euclidean distance formula. The SOM learning rate was set to 0.02, the constant in the inverse learning rate function was set to 0.01 and the radius of the neighborhood was 3. SOM was trained using default parameters implemented in “som” package which is comprised of 12 training epochs for a complete dataset and required about 9 h on a personal computer (core-i5, 8 Gb RAM). This SOM configuration enables the robust identification of spot modules inherent in the data (Wirth et al., 2011).

PSF landscape of each disease group is described by the meta-PSF expression values (“personal” SOM portrait). They are arranged according to the underlying SOM grid and visualized by an appropriate color gradient. The color patterns emerge as smooth textures representing the fingerprint of pathway activity perturbation for each disease. The individual portraits are mutually comparable. The SOM algorithm arranges similar meta-PSF profiles in neighboring tiles of the map whereas more different ones are located more distantly. Meta-PSFs located in the same spot-like region of the map are concertedly deregulated across the diseases. Overexpression (underexpression) spots in each “personal SOM” portrait were determined as clusters of meta-PSFs, which have values above (below) the predefined



threshold (90 percent of meta-PSF extremes), or by applying K-means or correlation clustering to the meta-PSF profiles as described in detail elsewhere (Wirth et al., 2011). All spots detected in the individual portraits are transferred into one master map to visualize the global spot patterns.

Significance analysis for differential PSF-values use a shrinkage *t*-test and false discovery rate (FDR) estimation for multiple test correction as implemented in oposSOM and described in detail elsewhere (Wirth et al., 2011, 2012a). This analysis provided lists of pathway sinks with co-regulated activities in the lung diseases and healthy controls studied.

A separate part of analyses focused on assessment of pathway activity changes in lung diseases compared to healthy lungs. For this, the difference of mean-centered PSF values between diseases and healthy controls was computed ($PSF_{disease} - PSF_{healthy lung}$), representing log fold difference of pathway deregulation in diseased over healthy state. A pathway was considered as significantly deregulated, if its *p* value was less than 0.05 and FDR *q* value was less than 0.2.

Clustering of Diseases Based on the Pathway Activation States

Similarity analysis of lung diseases was performed based on their SOM portraits using so-called second level SOM, independent

component analysis (ICA), and hierarchical clustering methods implemented in oposSOM (Wirth et al., 2011). In addition we applied graph community search with random walktrap algorithm implemented in igraph R package using Fruchterman-Reingold layout (Csardi and Nepusz, 2006).

RESULTS

Pathway Activity Portraits of Lung Diseases

PSF profiles were evaluated for 138 KEGG metabolic, signaling, and organismal pathways in 948 diseased and normal lung samples constituting 21 disease groups and one healthy lung tissue (Table 1). Each of the samples was characterized by its state of pathway (de-)regulation, which is defined as a set of PSF values at all 943 pathway sinks. In order to evaluate similarities and differences between the states of pathway deregulation of diseased and healthy lungs, we applied machine learning using self-organizing maps (SOM). This method transforms the multidimensional PSF data into a series of two-dimensional images, called “portraits,” which visualize the activities of the sink nodes in each disease studied (Figure 3). These PSF-activity portraits show blue and red spot like regions corresponding to down- and up-regulated sink nodes, respectively. The complete

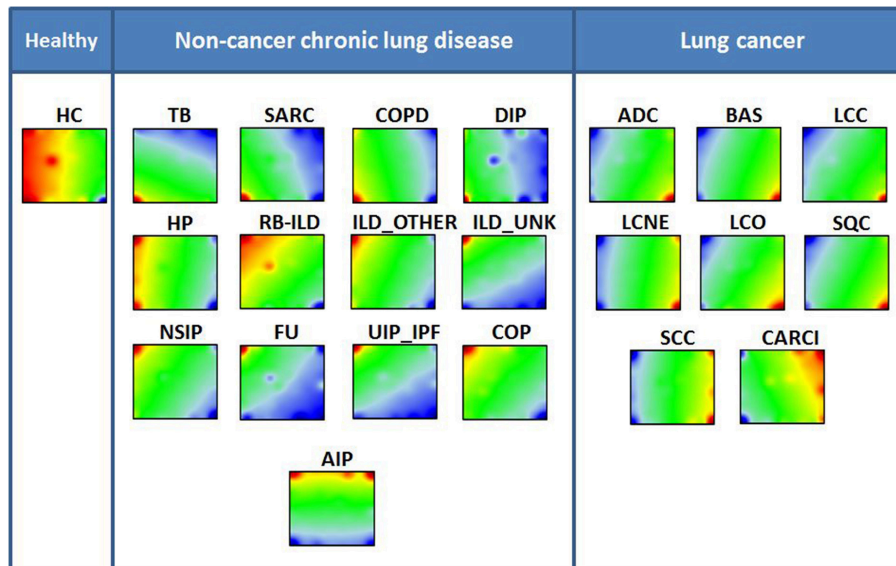


FIGURE 3 | Disease specific SOM portraits. Each lung disease and healthy control is characterized by its PSF portrait visualizing the pathway activities. The diseases are sorted into three groups. The color key for the summary SOM portrait is blue-yellow-red, where blue colors represent lowest values, green-yellow transition includes intermediate values and the red colors represent highest values. Note the marked difference between groups, especially between cancer and chronic lung diseases.

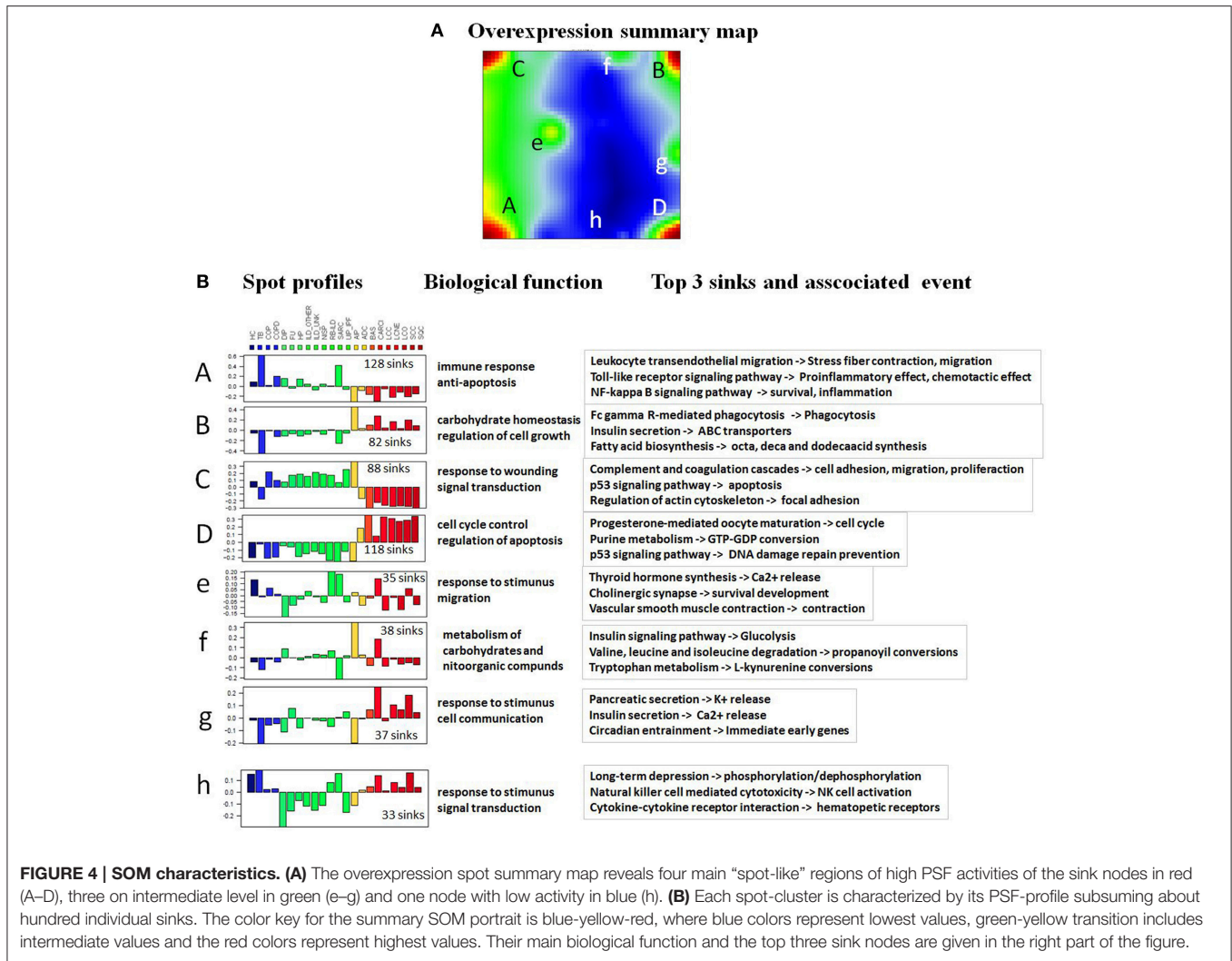
list of pathway sinks associated with each spot in each disease groups in presented in Supplementary Material Data Sheet 3.

It is apparent that healthy lung and non-cancer diseases have clearly distinct pathway deregulation patterns compared to lung cancers (Figure 3). Most cancer diseases are characterized by an upregulated red spot located in the right lower corner, in combination with a down regulated blue spot in the left upper corner of the portrait. Non-malignant lung diseases, in turn, show a distribution of up- and down-regulated spots that virtually mirror that of cancer diseases. Further, this group of diseases showed higher variability of their spot distributions compared to cancers. Here, generally, two different patterns in portraits can be distinguished. One pattern can be specified by the presence of an upregulated spot in the bottom left corner (TB, SARC, COPD, DIP), while the second pattern is defined by an upregulated spot in the upper left corner (NSIP, FU, UIP/IPF, COP, ILD_OTHER, ILD_UNK, RB_ILD). Meanwhile, the spot distribution in HP portrait shows transition between the first and the second chronic lung disease patterns, while the spot distribution on the AIP portrait can be considered as a transition between non-cancer and cancer disease patterns. Finally, it is worthy to note that FU, UIP/IPF, and ILD_UNK share a down-regulated spot with lung malignancies at the bottom left corner of the portraits.

To have a detailed look on pathways involved in formation of these disease-specific portraits, we referred to the overexpression summary map in Figure 4A. It integrates spots from all individual lung disease portraits and thus it provides an overview of all relevant regions becoming activated in the data set. The spots represent clusters of correlated and thus concertedly deregulated pathway sink profiles in one or several diseases. The overexpression spots (A, B, C, and D) on the corners of

the summary map represent the main features allowing for distinguishing between non-cancer and malignant lung diseases. The rest of the spots provide finer structure of pathway activation states in different diseases. Since the amplitude of meta-PSF values of these intermediate spots was much lower than that of the major overexpression spots, we decided to concentrate on the latter's in subsequent analyses. Overall, 51 pathways characterized with at least one deregulated sink have been found to be associated with these main overexpression spots (Supplementary Material Data Sheet 1, Table S2). Next, we proceeded to more thorough functional annotation of all pathway sinks located in each overexpression spot of the summary map. Because pathway sinks are gene products directly associated with some functional event or molecular process, we have performed GO term enrichment analysis of sink genes in spots using WebGestalt (Zhang et al., 2005; Figure 4B). The results showed that pathway sinks associated with spots A and C (upregulated in non-cancer diseases) are mainly enriched with GO terms related to immune/inflammatory response, proliferation, and anti-apoptosis, while spots B and D (upregulated in cancer diseases) are linked to cell cycle regulation, apoptosis, and carbohydrate metabolism. The complete list of GO terms associated with pathway sinks and the enrichment significance can be found in the Supplementary Material Data Sheets 4–11.

The pathways “Cytokine-cytokine receptor interaction,” “MAPK signaling pathway,” “Cell adhesion molecules (CAMs),” “FoxO signaling pathway” and “Dopaminergic synapse” were found to be significantly associated with all the overexpression spots and thus all the investigated lung conditions; however, with disease-specific molecular processes associated with each sink in these pathways (for details on these pathways and their specific sinks, associated with each spot and their significance



values, see to Supplementary Material Data Sheet 1, Table S3). While pathway sinks associated with immune response were located in spots A and C and activated mainly in non-malignant diseases, the cancer-associated spots B and D contained pathway sinks related to signaling, proliferation, and cancer related cell communication. These results show that different sink nodes of a pathway can be located in the same spot or in different ones reflecting thus correlated, anti-correlated, or non-correlated profiles. The examples in **Figure 5** illustrate two different situations. Difference of sink profiles in the same pathway can be caused by multiple inputs, multibranching, existence of inhibitory interactions, and/or deregulated expression of intermediate nodes (**Figure 5A**). On the other hand, unbranched pathway topologies tend to show more concerted profiles of their sink nodes, which, in consequence, accumulate in the same region of the map (**Figure 5B**).

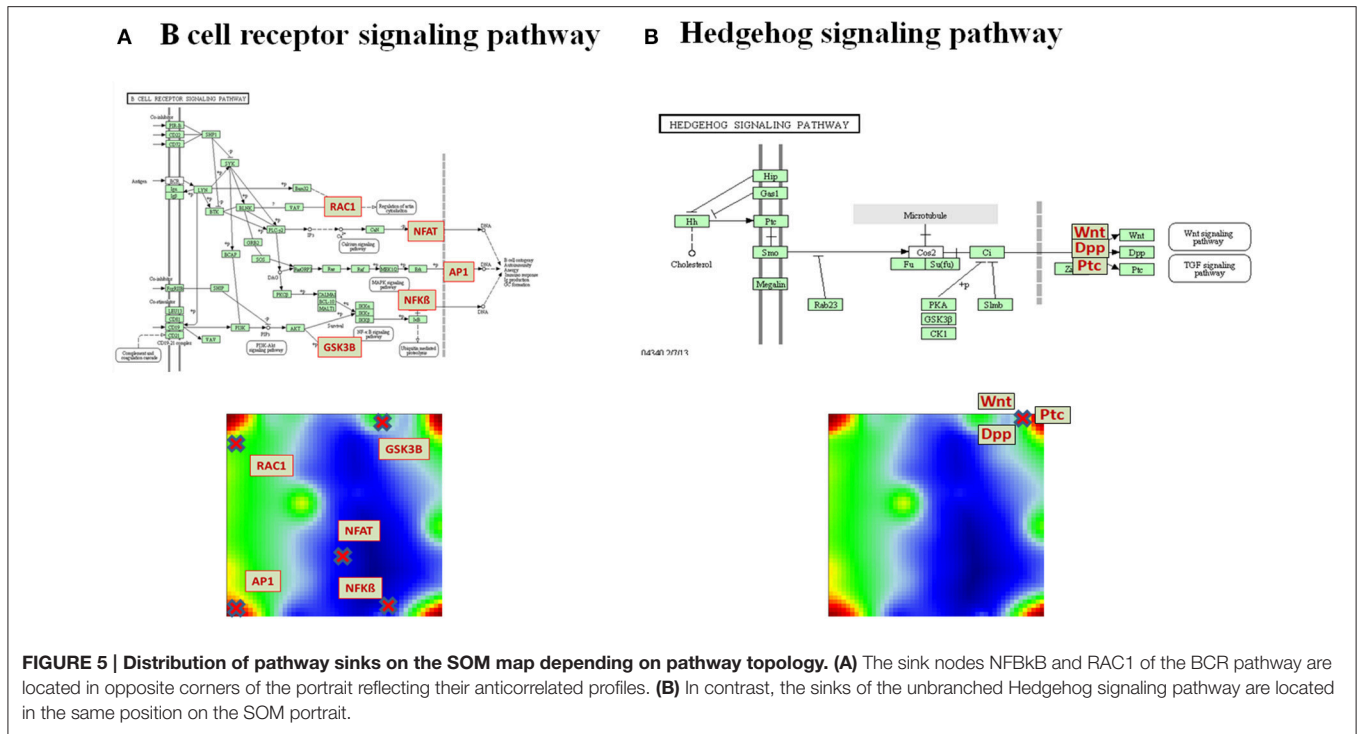
Clustering of Disease Groups Based on Pathway Deregulation Activities

Visual inspection of the SOM portraits revealed considerable similarities among non-cancer, as well as cancer diseases.

Hierarchical cluster analysis of the SOM portraits using either Pearson correlation or Euclidian distances of meta-PSF values as similarity metrics revealed two main clusters of diseases. In both cases, the first cluster grouped lung cancer diseases, while the second cluster contained non-cancer diseases and healthy controls (**Figures 6A,B**). For a higher resolution of the diversity analysis we applied SOM mapping of the samples (so-called 2nd level SOM) and ICA as described in Wirth et al. (2012a). The SOM mapping reveals that especially the cluster of non-malignant lung diseases further disentangles into different subclusters labeled 1–3 (**Figures 7A,B**). Similarly, the 2nd level SOM, ICA, which assumes non-Gaussian distribution of distances between samples, clearly separates cancer from non-cancer lung diseases. TB, SARC, and AIP appear as outliers that are separated from the other diseases.

Pathway Deregulation Profiles in Lung Diseases Compared to Healthy Controls

In order to assess pathway activity deregulations in lung diseases compared to healthy state, we calculated differential PSF values respective to healthy controls. Most of the obtained differential



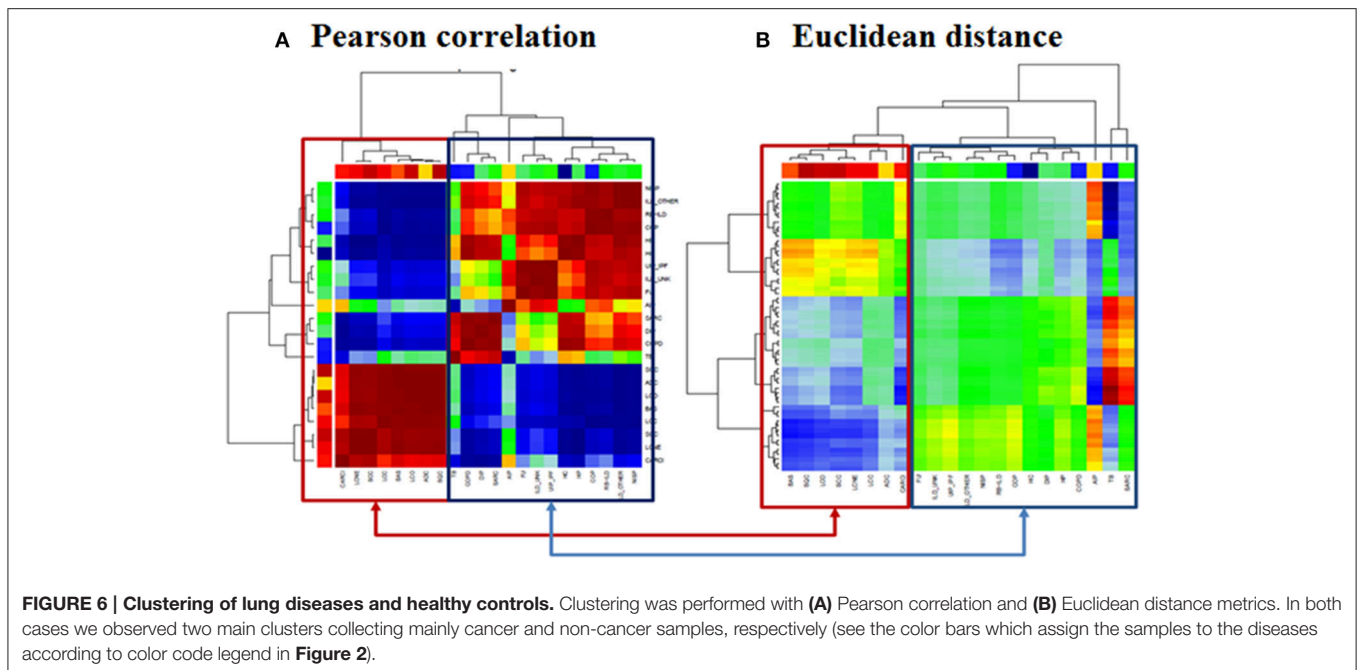
SOM portraits showed similar spot distributions compared to the original ones (see Supplementary Material Data Sheet 12, Supplementary Figure 1). Next, we created a graph object with nodes representing diseases and edges associated with the number of co-regulated pathways between each pair of disease nodes. Finally, we performed a graph community search with random walktrap algorithm in order to identify diseases sharing maximal similarities in pathway deregulation states. This algorithm analyzes the graph by random walks to identify densely connected subgraphs, called communities. In total, we identified 4 communities (or clusters) containing three or more diseases (Figure 8 and Table 2, clusters 1–4). Each multi-disease community is specifically described by shared pathway sinks, similarly deregulated in all members of a given community (Table 2). Particularly, we observed that diseases in the cancer cluster contained 14 shared pathway deregulation states, while chronic lung disease clusters showed less number of shared states (between three and five).

The first community (Cluster 1—red) of non-cancer lung diseases contained COPD, sarcoidosis (SARC), and tuberculosis (TB). Numerous experimental studies, including our own, implicated immune/inflammatory pathways such as Toll-like receptor signaling, phagocytosis, and chemokine signaling in sarcoidosis, tuberculosis, and COPD (Arakelyan et al., 2009; Haspel and Choi, 2011; Kriegova et al., 2011; An et al., 2012; Maertzdorf et al., 2012; Pabst et al., 2013; Pugazhendhi et al., 2013). Moreover, differential gene expression as well as gene polymorphisms associated with these diseases demonstrated considerable mutual overlap (Arakelyan et al., 2009; Haspel and Choi, 2011). Furthermore, the pathophysiology of TB, SARC, and COPD is very similar, often making them hard to distinguish

(Maertzdorf et al., 2012). Thus, our findings are in line with previous findings confirming the presence of shared deregulation of immune system pathways in these diseases.

The second community (Cluster 2—orange) joined respiratory bronchiolitis interstitial lung disease (RB-ILD) (Leslie, 2009; Meyer et al., 2012), hypersensitivity pneumonia (HP), and cryptogenic organizing pneumonia (COP). The molecular pathways involved in the development of these diseases are poorly understood and there is not much data on gene expression changes in these disorders. Nevertheless, the clustering of RB-ILD, HP, and COP together based on global gene expression has also been described previously (Cho et al., 2011; Lee and Yang, 2013). Our results showed that, these diseases were characterized by shared downregulation of anti-apoptotic and proinflammatory branches of Hippo signaling pathway. This pathway is an important regulator of tissue growth. It has been studied largely in tumor development, while results on non-cancer lung diseases are mostly limited to animal models (Gomez et al., 2014). It has been shown that the activation of Hippo signaling pathway in non-cancer diseases is beneficial because it promotes tissue regeneration through proliferation and inhibition of apoptosis (Halder and Johnson, 2011). By contrast, inhibition of Hippo signaling can promote substitution of healthy tissues with fibrotic tissue during cardiac remodeling (Xin et al., 2013). Thus, we can speculate that this pathway may be implicated in progression of fibrosis in these diseases.

The third and the biggest community of interstitial lung diseases (Cluster 3—light green) included interstitial pneumonias, such as UIP/IPF, NSIP, and unclassified interstitial lung diseases (ILDs). This cluster was characterized by activation of various cytokine-cytokine receptor interactions,



and upregulation of WNT signaling pathway. Moreover, it has been noticed that this cluster serves as a link between cancer and other ILD diseases (Archontogeorgis et al., 2012; Vancheri, 2013). The observed association between ILD, and especially UIP/IPF with cancer, has been reported previously (reviewed in Archontogeorgis et al., 2012; Vancheri, 2013). It has been noted that IPF and lung malignancies share molecular pathomechanisms, including epigenetic changes, delayed apoptosis and changes in cell-cell interactions (Vancheri, 2013). Our results demonstrated that the link can be realized through common upregulation in cell cycle related pathways (Hippo, p53), signal transduction (cytokine, thyroid hormone), and metabolism of nucleic acids (purines and pyrimidines).

Finally, cancer diseases (Cluster 4—green) were clustered into one highly homogeneous community characterized by downregulation of immune system related pathways and upregulation of cell cycle, proliferation, and metabolism pathways, as expected (Han et al., 2014; Domagala-Kulawik, 2015). From all lung cancer types included in our dataset, the lung carcinoid tumor (CARCI) demonstrated a different pathway deregulation profile (Anbazhagan et al., 1999), which may indicate its neuroendocrine origin (Rekhtman, 2010). Indeed, our data shows that it shares considerable amount of similarities with large cell neuroendocrine cancer (Figure 8).

Besides the multi-disease communities, we observed the presence of “linker” disease nodes (DIP, CARCI, and AIP) in the similarity graph. Though they are not included in any cluster, they provide links between different clusters.

The community detection analysis confirmed the diverse nature of cancer and non-malignant lung diseases, but it also provided additional information that wasn't obvious using other similarity analysis methods, such as clustering, second level SOM or ICA. The main advantage here was that the graph object

allowed for direct assessment of similarly deregulated pathways between all pairs of studied diseases, with reference to the healthy state. We have noticed that pathway deregulation states in fibrotic diseases (cluster 3) serve as a central hub that shares considerable similarities with inflammatory (cluster 1, 2) and cancer diseases (cluster 4). Moreover, our results suggest that interstitial lung diseases seem to constitute a more heterogeneous group of diseases in terms of molecular mechanisms underlying their pathology. Even if it is well known that inflammation and fibrosis are the main characteristics of interstitial lung diseases (Bourke, 2006), our data suggested that the drivers of these processes are different and are associated with perturbations in different pathways.

DISCUSSION

Using our novel PSF-SOM method we have evaluated similarities of pathway deregulation profiles in a large spectrum of lung pathologies by mining high-dimensional gene expression data and topologies of signaling and metabolic pathways. We pursued a systems biology view to identify most prominent properties of groups of lung diseases in terms of pathway deregulation patterns. Our results revealed considerable differences in pathway deregulations implicated in cancer and non-cancer lung diseases. While lung cancers were characterized by pathways implicated in cell proliferation, metabolism, non-malignant lung diseases were characterized by deregulations in pathways involved in immune/inflammatory response and fibrotic tissue remodeling. Moreover, we observed considerable heterogeneity in terms of pathway deregulation in interstitial lung diseases (ILD). We were able to identify three groups of ILDs and showed that the development of characteristic pathological processes

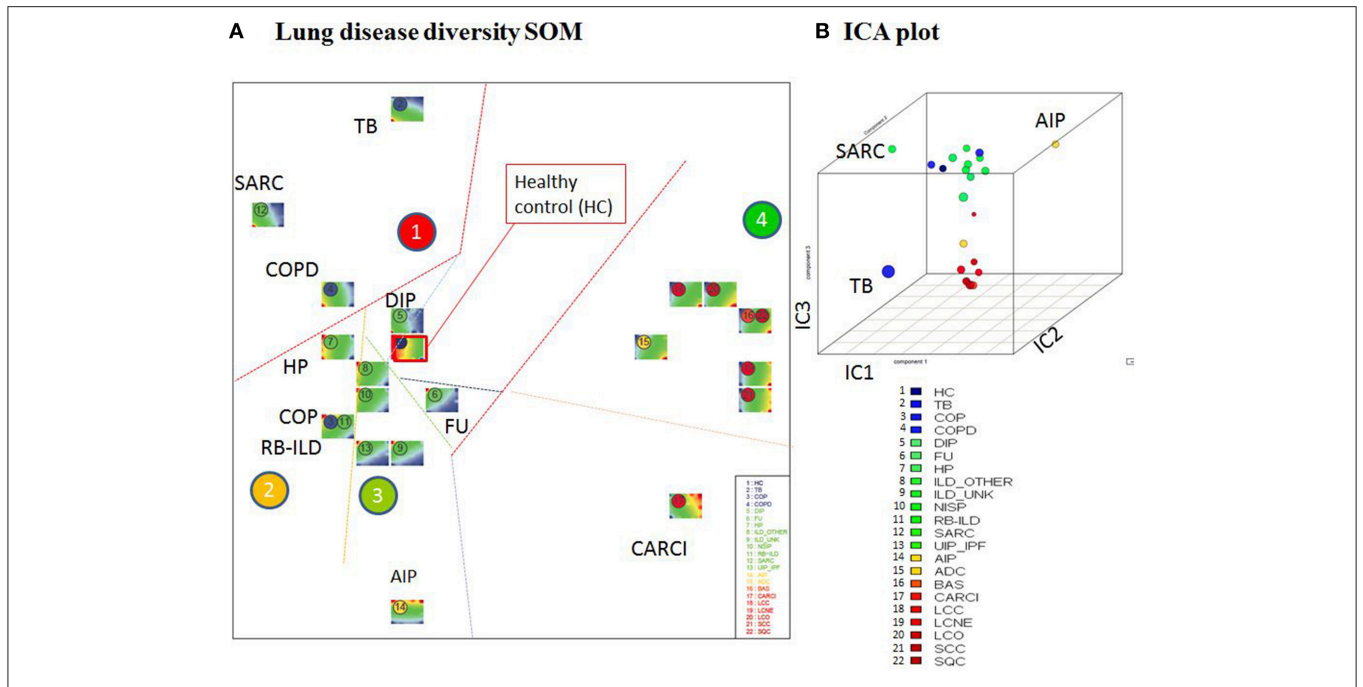


FIGURE 7 | Diversity analysis of lung diseases. (A) The sample-diversity SOM projects the multidimensional diversity landscape into a two-dimensional plot. Each disease class is visualized by its sample portraits. Borders between the clusters 1–4 serve as guides for the visual inspection. (B) The three dimensional ICA plot shows that independent component 3 (IC3) mainly separates lung cancers from chronic lung diseases.

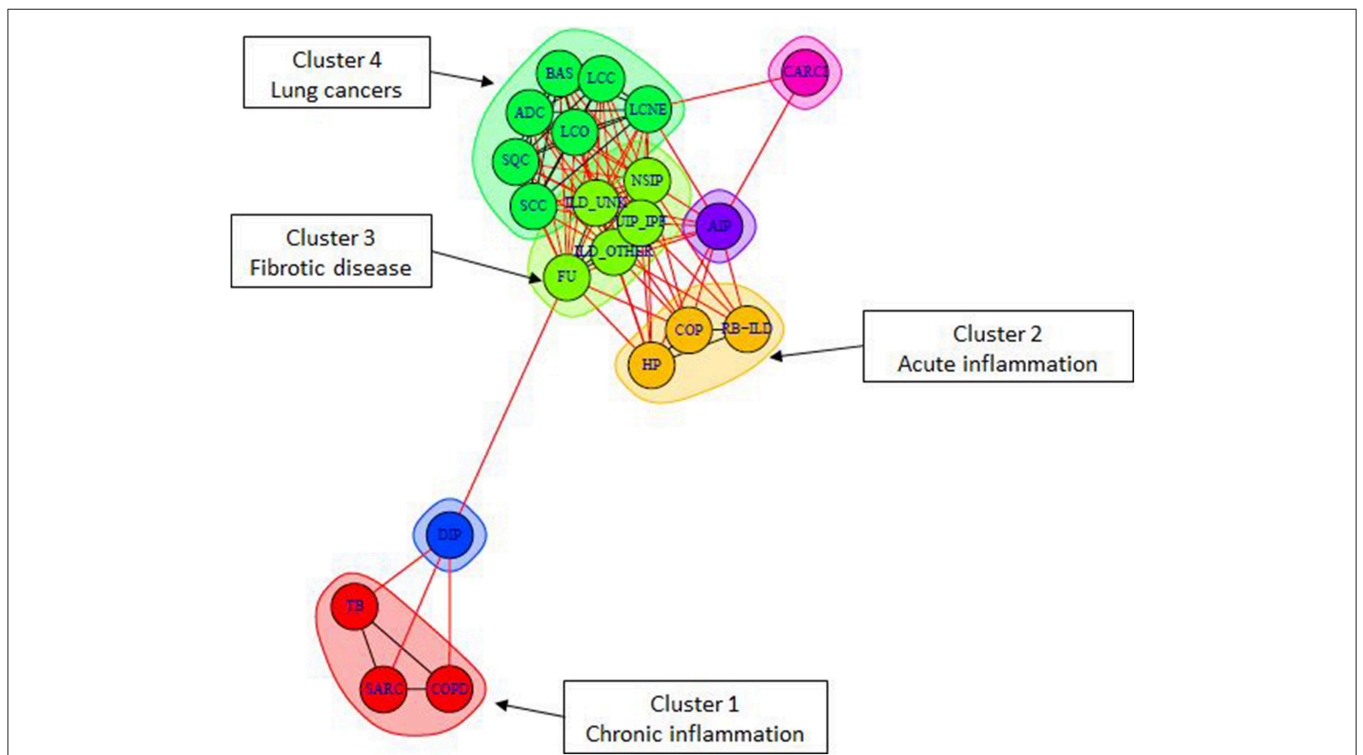


FIGURE 8 | Community detection in the lung disease graph. Each color represents one community, which gathers diseases with the highest pairwise similarity in pathway deregulation states (see also Table 2). The graph layout is generated with Fruchterman–Reingold algorithm where edge length is inversely correlated with node connectivity.

TABLE 2 | Assessment of pathway common deregulation profiles in lung disease communities.

Community (color*)	Diseases	Deregulated pathway	Sink (GeneID/Symbol)	Associated molecular process
Cluster 1 (Red)	SARC TB COPD	B cell receptor signaling pathway	FOS/2353	Immune response (UP)
		Chemokine signaling pathway	WAS/7454	Regulation of cytoskeleton (UP)
		NF-kappa B signaling pathway	CCL19/6363	Lymphoid-tissue homing (UP)
		Toll-like receptor signaling pathway	CCL3L3/414062	Chemotactic effect (UP)
Cluster 2 (Orange)	RB-ILD HP COP	Hippo signaling pathway	SERPINE1/5054	Antiapoptosis (DN)
Cluster 3 (Light green)	UIP_IPF ILD_OTHER ILD_UNK NSIP FU	Cytokine-cytokine receptor interaction	TNFRSF17/608	TNF Family (UP)
		Cytokine-cytokine receptor interaction	CXCR5/643	CXC Family (UP)
		Wnt signaling pathway	MMP7/4316	Cell cycle (UP)
Cluster 4 (Green)	ADC SQC LCC LCO BAS LCNE SCC	Alanine, aspartate, and glutamate metabolism	GFPT1/2673	Amino sugar metabolism (UP)
		Alanine, aspartate, and glutamate metabolism	ASPA/443	L-aspartate (UP)
		Glycine, serine, and threonine metabolism	AOC2/314	Methylglyoxal (DN)
		Glycolysis/Gluconeogenesis	PGK1/5230	Glycerate1,3P (UP)
		Purine metabolism	PGM1/5236	Ribose-5P (UP)
		Purine metabolism	GDA/9615	Guanine (UP)
		Pyrimidine metabolism	TK1/7083	dUMP(UP)
		Adrenergic signaling in cardiomyocytes	AKT3/10000	Apoptosis (DN)
		p53 signaling pathway	IGF1/3479	Apoptosis (DN)
		p53 signaling pathway	GTSE1/51512	DNA damage prevention & repair (UP)
		Progesterone-mediated oocyte maturation	ANAPC10/10393	Metaphase arrest (DN)
Progesterone-mediated oocyte maturation	CDK1/983	Cell cycle (UP)		

*DIP, CARCI, and AIP formed separate single member communities; UP, up-regulation; DN, downregulation.

such as fibrosis can be initiated by deregulations in different signaling pathways. These results raise a question whether it is appropriate to combine ILDs into single group of diseases. Finally, we also detected substantial similarities between cancers and a subset of interstitial lung diseases characterized by fibrosis development, which suggests the presence of shared molecular mechanisms involved in their pathogenesis.

Our results replicate existing knowledge on pathological processes in lungs. The inflammation and fibrosis related pathways identified to be deregulated in non-malignant chronic lung diseases considerably overlapped with the ones reported by Cho et al. (2011) using gene-centered analysis approaches (see Additional File 1 of Cho et al., 2011). Moreover, deregulations of metabolic activities and cell cycle control have also been reported previously in lung cancers (Cantor and Sabatini, 2012).

Furthermore, our findings are in line with the results of recent publication by Kim et al. (Cantor and Sabatini, 2012) on interstitial disease subgrouping based in molecular phenotypes. However, we believe that our study has a number of important differences compared to the previously reported results. In their paper, authors used ANOVA significance value cut-offs for identification of differentially expressed genes and have applied topology free enrichment analysis for functional assessment of biological processes implicated in lung diseases. Consequently, all the criticism we have mentioned in the Introduction, also applies here. In contrast, our results directly extend current understanding of molecular events associated with various aspects of lung pathology. We did not limit our study with obstructive or interstitial lung diseases only, but aggregated lung cancers, chronic non-cancer lung diseases and, especially, a wide

spectrum of interstitial lung diseases. Moreover, we have directly evaluated pathway activity deregulation states for each ILD type separately, rather than obtaining results about interstitial diseases in general. This is a special feature of this work and our data may provide additional insight into pathobiology of these frequently neglected diseases, which are often difficult to diagnose and treat.

From methodological point of view this paper demonstrates the power of our pathway centered analysis. While usual gene-centered studies provide lists of genes and associated functional categories, our approach more closely refers to systems level definitions and provides a rich outlook on actual molecular events associated with studied conditions. We have combined PSF and SOM algorithms that previously have been extensively benchmarked (Wirth et al., 2011, 2012a; Binder et al., 2014; Löffler-Wirth et al., 2015; Nersisyan et al., 2015). The PSF algorithm integrates pathway knowledge and topology information with gene expression data to identify deregulated branches of pathways. On the other hand, SOM analysis allows for mining for pathways with similar deregulation patterns across datasets.

This study has a number of limitations that is worth to discuss here. First, the data analyzed in this study refer to endpoint pathologies, and not to initial steps of disease development. While the latter is important for assessing the causes of a disease, the former describes the state at which most therapeutic interventions are targeted. Thus, assessment of disease similarities based on end-point phenotypes is not suited to study initiation and early genesis of lung diseases. Second, our study was limited to KEGG pathways that contain previously curated information (Ogata et al., 1999). This approach does not allow for identification of new gene interactions or pathways implicated in lung diseases. On the other hand, it provides

the advantage of using “gold standard” pathway topologies, which are enriched with information about known functional outcomes.

In conclusion, this paper describes the pathobiology of lung diseases from systems viewpoint using a pathway centered high-dimensional data mining approach. Our results largely contribute to current understanding of pathological events in lung cancers and non-malignant lung diseases. Moreover, our results provide new insight into molecular mechanisms of a number of interstitial lung diseases that have been studied to a lesser extent compared to interstitial pulmonary fibrosis, COPD, and sarcoidosis.

AUTHOR CONTRIBUTIONS

AA, LN, and HB initiated the study. AA and LN performed calculations with contribution from HB and HL. AA, LN, HB, HL, and MP contributed to results interpretation and manuscript writing. All authors read edited and approved the final manuscript.

ACKNOWLEDGMENTS

Grant in acknowledgment: IGA_PU LF 2015_020 (MP), LO1304, SCS AM 13YR-1F0022 (AA), CapSys BMBF (FKZ: 01ZX1304A) (HB).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fgene.2016.00079>

REFERENCES

- An, C. H., Wang, X. M., Lam, H. C., Ifedigbo, E., Washko, G. R., Ryter, S. W., et al. (2012). TLR4 deficiency promotes autophagy during cigarette smoke-induced pulmonary emphysema. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 303, L748–L757. doi: 10.1152/ajplung.00102.2012
- Anbazhagan, R., Tihan, T., Bornman, D. M., Johnston, J. C., Saltz, J. H., Weigering, A., et al. (1999). Classification of small cell lung cancer and pulmonary carcinoid by gene expression profiles. *Cancer Res.* 59, 5119–5122.
- Arakelyan, A., Aslanyan, L., and Boyajyan, A. (2013). “High-throughput gene expression analysis concepts and applications,” in *Genomics II - Bacteria, Viruses and Metabolic Pathways* (iConcept Press Ltd.), 71–95. Available online at: <https://www.iconceptpress.com/books/genomics-ii--bacteria-viruses-and-metabolic-pathways/14410061/>
- Arakelyan, A., Kriegova, E., Kubistova, Z., Mrazek, F., Kverka, M., du Bois, R. M., et al. (2009). Protein levels of CC chemokine ligand (CCL)15, CCL16 and macrophage stimulating protein in patients with sarcoidosis. *Clin. Exp. Immunol.* 155, 457–465. doi: 10.1111/j.1365-2249.2008.03832.x
- Archontogeorgis, K., Steiropoulos, P., Tzouvelekis, A., Nena, E., and Bouros, D. (2012). Lung cancer and interstitial lung diseases: a systematic review. *Pulm. Med.* 2012:315918. doi: 10.1155/2012/315918
- Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., et al. (2011). NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res.* 39, D1005–D1010. doi: 10.1093/nar/gkq1184
- Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., et al. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13790–13795. doi: 10.1073/pnas.191502998
- Binder, H., Hopp, L., Lembcke, K., and Wirth, H. (2015). “Personalized disease phenotypes from massive OMICs data” (IGI Global). Available online at: <http://www.igi-global.com/chapter/personalized-disease-phenotypes-from-massive-omics-data/121465/> (Accessed April 6, 2016).
- Binder, H., and Wirth, H. (2014). “Analysis of large-scale OMIC data using self organizing maps,” in *Encyclopedia of Information Science and Technology, 3rd Edn.*, ed M. Khosrow-Pour (Hershey, PA: IGI Global), 1642–1654.
- Binder, H., Wirth, H., Arakelyan, A., Lembcke, K., Tiys, E., Ivanisenko, V., et al. (2014). Time-course human urine proteomics in space-flight simulation experiments. *BMC Genomics* 15:S2. doi: 10.1186/1471-2164-15-S12-S2
- Bourke, S. J. (2006). Interstitial lung disease: progress and problems. *Postgrad. Med. J.* 82, 494–499. doi: 10.1136/pgmj.2006.046417
- Buettner, R., Wolf, J., and Thomas, R. K. (2013). Lessons learned from lung cancer genomics: the emerging concept of individualized diagnostics and treatment. *J. Clin. Oncol.* 31, 1858–1865. doi: 10.1200/JCO.2012.45.9867
- Campbell, J. D., McDonough, J. E., Zeskind, J. E., Hackett, T. L., Pechkovsky, D. V., Brandsma, C.-A., et al. (2012). A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. *Genome Med.* 4:67. doi: 10.1186/gm367
- Campbell, J. D., Spira, A., and Lenburg, M. E. (2011). Applying gene expression microarrays to pulmonary disease. *Respirology* 16, 407–418. doi: 10.1111/j.1440-1843.2011.01942.x
- Cancer, T., and Atlas, G. (2012). Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489, 519–525. doi: 10.1038/nature11404

- Cantor, J. R., and Sabatini, D. M. (2012). Cancer cell metabolism: one hallmark, many faces. *Cancer Discov.* 2, 881–898. doi: 10.1158/2159-8290.CD-12-0345
- Chilosi, M., Poletti, V., and Rossi, A. (2012). The pathogenesis of COPD and IPF: distinct horns of the same devil? *Respir. Res.* 13:3. doi: 10.1186/1465-9921-13-3
- Cho, J.-H., Gelinias, R., Wang, K., Etheridge, A., Piper, M. G., Batte, K., et al. (2011). Systems biology of interstitial lung diseases: integration of mRNA and microRNA expression changes. *BMC Med. Genomics* 4:8. doi: 10.1186/1755-8794-4-8
- Courtois, G., and Gilmore, T. D. (2006). Mutations in the NF-kappaB signaling pathway: implications for human disease. *Oncogene* 25, 6831–6843. doi: 10.1038/sj.onc.1209939
- Covell, D. G., Wallqvist, A., Rabow, A. A., and Thanki, N. (2003). Molecular classification of cancer: unsupervised self-organizing map analysis of gene expression microarray data. *Mol. Cancer Ther.* 2, 317–332.
- Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJ. Complex Syst.* 1695. Available online at: http://www.interjournal.org/manuscript_abstract.php?361100992
- Daigle, B. J., Srinivasan, B. S., and Flannick, J. A. (2010). “Current progress in static and dynamic modeling of disease networks,” in *Signalling Networks*. Available at: http://link.springer.com/chapter/10.1007%2F978-1-4419-5797-9_2
- DePianto, D., Chandriani, S., Abbas, A., Jia, G., N’Diaye, E., Caplazi, P., et al. (2015). Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis. *Thorax* 70, 48–56. doi: 10.1136/thoraxjnl-2013-204596
- Domagala-Kulawik, J. (2015). The role of the immune system in non-small cell lung carcinoma and potential for therapeutic intervention. *Transl. lung cancer Res.* 4, 177–190. doi: 10.3978/j.issn.2218-6751.2015.01.11
- Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210. doi: 10.1093/nar/30.1.207
- Eichler, G. S., Huang, S., and Ingber, D. E. (2003). Gene Expression Dynamics Inspector (GEDI): for integrative analysis of expression profiles. *Bioinformatics* 19, 2321–2322. doi: 10.1093/bioinformatics/btg307
- Gerber, D., Oxnard, G., and Govindan, R. (2015). ALCHEMIST: bringing genomic discovery and targeted therapies to early-stage lung cancer. *Clin. Pharmacol. Ther.* 97, 447–450. doi: 10.1002/cpt.91
- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537. doi: 10.1126/science.286.5439.531
- Gomez, M., Gomez, V., and Hergovich, A. (2014). The Hippo pathway in disease and therapy: cancer and beyond. *Clin. Transl. Med.* 3:22. doi: 10.1186/2001-1326-3-22
- Halder, G., and Johnson, R. L. (2011). Hippo signaling: growth control and beyond. *Development* 138, 9–22. doi: 10.1242/dev.045500
- Han, J., Tang, F., Pu, D., Xu, D., Wang, T., and Li, W. (2014). Mechanisms underlying regulation of cell cycle and apoptosis by hnRNP B1 in human lung adenocarcinoma A549 cells. *Tumori* 100, 102–111. doi: 10.1700/1430.15824
- Haspel, J. A., and Choi, A. M. K. (2011). Autophagy: a core cellular process with emerging links to pulmonary disease. *Am. J. Respir. Crit. Care Med.* 184, 1237–1246. doi: 10.1164/rccm.201106-0966CI
- Hopp, L., Willscher, E., Löffler-Wirth, H., and Binder, H. (2015). Function shapes content: DNA-methylation marker genes and their impact for molecular mechanisms of glioma. *J. Can. Res. Updates* 4, 127–148. doi: 10.6000/1929-2279.2015.04.04.1
- Hopp, L., Wirth, H., Fasold, M., and Binder, H. (2013). Portraying the expression landscapes of cancer subtypes: a glioblastoma multiforme and prostate cancer case study. *Syst. Biomed.* 1, 99–121. doi: 10.4161/sysb.25897
- Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D. E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* 94:128701. doi: 10.1103/physrevlett.94.128701
- Kanehisa, M. (2002). The KEGG database. *Novartis Found. Symp.* 247, 91–101. discussion 101–103, 119–128, 244–252. doi: 10.1002/0470857897.ch8
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 43, 59–69. doi: 10.1007/BF00337288
- Kriegova, E., Fillerova, R., Tomankova, T., Hutyrova, B., Mrazek, F., Tichy, T., et al. (2011). T-helper cell type-1 transcription factor T-bet is upregulated in pulmonary sarcoidosis. *Eur. Respir. J. Off. J. Eur. Soc. Clin. Respir. Physiol.* 38, 1136–1144. doi: 10.1183/09031936.00089910
- Lee, I. T., and Yang, C. M. (2013). Inflammatory signalings involved in airway and pulmonary diseases. *Mediators Inflamm.* 2013:791231. doi: 10.1155/2013/791231
- Leng, D., Huan, C., Xie, T., Liang, J., Wang, J., Dai, H., et al. (2013). Meta-analysis of genetic programs between idiopathic pulmonary fibrosis and sarcoidosis. *PLoS ONE* 8:e71059. doi: 10.1371/journal.pone.0071059
- Leslie, K. O. (2009). My approach to interstitial lung disease using clinical, radiological and histopathological patterns. *J. Clin. Pathol.* 62, 387–401. doi: 10.1136/jcp.2008.059782
- Lewis, C. C., Yang, J. Y., Huang, X., Banerjee, S. K., Blackburn, M. R., Baluk, P., et al. (2008). Disease-specific gene expression profiling in multiple models of lung disease. *Am. J. Respir. Crit. Care Med.* 177, 376–387. doi: 10.1164/rccm.200702-333OC
- Li, J., Li, D., and Wei, X., S. Y. (2014). *In silico* comparative genomic analysis of two non-small cell lung cancer subtypes and their potentials for cancer classification. *Cancer Genomics Proteomics* 11, 303–310.
- Löffler-Wirth, H., Kalcher, M., and Binder, H. (2015). oposSOM: R-package for high-dimensional portraying of genome-wide expression landscapes on bioconductor. *Bioinformatics* 31, 3225–3227. doi: 10.1093/bioinformatics/btv342
- Logan, C. Y., and Nusse, R. (2004). The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* 20, 781–810. doi: 10.1146/annurev.cellbio.20.010403.113126
- Maertzdorf, J., Weiner, J., Mollenkopf, H.-J., Network, T., Bauer, T., Prasse, A., et al. (2012). Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7853–7858. doi: 10.1073/pnas.1121072109
- Meyer, K. C., Raghu, G., Baughman, R. P., Brown, K. K., Costabel, U., Du Bois, R. M., et al. (2012). An official American Thoracic Society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. *Am. J. Respir. Crit. Care Med.* 185, 1004–1014. doi: 10.1164/rccm.201202-0320ST
- Nersisyan, L., Johnson, G., Riel-Mehan, M., Pico, A., and Arakelyan, A. (2015). PSFC: a pathway signal flow calculator app for cytoscape [v1; ref status: approved 1]. *F1000Research* 4:480. doi: 10.12688/f1000research.6706.1
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34. doi: 10.1093/nar/27.1.29
- Pabst, S., Bradler, O., Gillissen, A., Nickenig, G., Skowasch, D., and Grohe, C. (2013). Toll-like receptor-9 polymorphisms in sarcoidosis and chronic obstructive pulmonary disease. *Adv. Exp. Med. Biol.* 756, 239–245. doi: 10.1007/978-94-007-4549-0_30
- Pennings, J. L. A., Kimman, T. G., and Janssen, R. (2008). Identification of a common gene expression response in different lung inflammatory diseases in rodents and macaques. *PLoS ONE* 3:e2596. doi: 10.1371/journal.pone.0002596
- Pouladi, N., Bime, C., Garcia, J. G. N., and Lussier, Y. A. (2015). Complex genetics of pulmonary diseases: lessons from genome-wide association studies and next-generation sequencing. *Transl. Res.* 168, 22–39. doi: 10.1016/j.trsl.2015.04.016
- Pugazhendhi, S., Jayakanthan, K., Pulimood, A. B., and Ramakrishna, B. S. (2013). Cytokine gene expression in intestinal tuberculosis and Crohn’s disease. *Int. J. Tuberc. Lung Dis.* 17, 662–668. doi: 10.5588/ijtld.12.0600
- Rekhtman, N. (2010). Neuroendocrine tumors of the lung: an update. *Arch. Pathol. Lab. Med.* 134, 1628–1638. doi: 10.1043/2009-0583-RAR.1
- Selman, M., Pardo, A., Barrera, L., Estrada, A., Watson, S. R., Wilson, K., et al. (2006). Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity pneumonitis. *Am. J. Respir. Crit. Care Med.* 173, 188–198. doi: 10.1164/rccm.200504-644OC
- Steiner, L., Hopp, L., Wirth, H., Galle, J., Binder, H., Prohaska, S. J., et al. (2012). A global genome segmentation method for exploration of epigenetic patterns. *PLoS ONE* 7:e46811. doi: 10.1371/journal.pone.0046811
- Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., et al. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 96, 2907–2912. doi: 10.1073/pnas.96.6.2907
- Thakur, R., Yadav, V. K., Kumar, A., Basundra, R., Kar, A., Halder, R., et al. (2014). Functional genomics of lung cancer progression reveals mechanism of

- metastasis suppressor function. *Mol. Cytogenet.* 7:19. doi: 10.1186/1755-8166-7-S1-19
- Törönen, P., Kolehmainen, M., Wong, G., and Castrén, E. (1999). Analysis of gene expression data using self-organizing maps. *FEBS Lett.* 451, 142–146. doi: 10.1016/S0014-5793(99)00524-4
- Vancheri, C. (2013). Common pathways in idiopathic pulmonary fibrosis and cancer. *Eur. Respir. Rev.* 22, 265–272. doi: 10.1183/09059180.00003613
- Wang, I.-M., Stepaniants, S., Boie, Y., Mortimer, J. R., Kennedy, B., Elliott, M., et al. (2008). Gene expression profiling in patients with chronic obstructive pulmonary disease and lung cancer. *Am. J. Respir. Crit. Care Med.* 177, 402–411. doi: 10.1164/rccm.200703-390OC
- West, L., Vidwans, S. J., Campbell, N. P., Shrager, J., Simon, G. R., Bueno, R., et al. (2012). A novel classification of lung cancer into molecular subtypes. *PLoS ONE* 7:e31906. doi: 10.1371/journal.pone.0031906
- Wirth, H., Cakir, M. V., Hopp, L., and Binder, H. (2014). Analysis of microRNA expression using machine learning. *Methods Mol. Biol.* 1107, 257–278. doi: 10.1007/978-1-62703-748-8_16
- Wirth, H., Löffler, M., von Bergen, M., and Binder, H. (2011). Expression cartography of human tissues using self organizing maps. *BMC Bioinformatics* 12:306. doi: 10.1186/1471-2105-12-306
- Wirth, H., von Bergen, M., and Binder, H. (2012a). Mining SOM expression portraits: feature selection and integrating concepts of molecular function. *BioData Min.* 5:18. doi: 10.1186/1756-0381-5-18
- Wirth, H., von Bergen, M., Murugaiyan, J., Rösler, U., Stokowy, T., and Binder, H. (2012b). MALDI-typing of infectious algae of the genus *Prototheca* using SOM portraits. *J. Microbiol. Methods* 88, 83–97. doi: 10.1016/j.mimet.2011.10.013
- Xin, M., Kim, Y., Sutherland, L. B., Murakami, M., Qi, X., McAnally, J., et al. (2013). Hippo pathway effector Yap promotes cardiac regeneration. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13839–13844. doi: 10.1073/pnas.1313192110
- Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* 33, W741–W748. doi: 10.1093/nar/gki475

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Arakelyan, Nersisyan, Petrek, Löffler-Wirth and Binder. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.