BMC Cancer

RESEARCH ARTICLE   **Open Access**

Integrative genomic expression analysis reveals stable differences between lung cancer and systemic sclerosis

Heng Li[1,2], Liping Ding[1], Xiaoping Hong[1], Yulan Chen[1], Rui Liao[1], Tingting Wang[1,2], Shuhui Meng[1], Zhenyou Jiang[3*] and Dongzhou Liu[1,4*]

Abstract

Background: The incidence and mortality of lung cancer are the highest among all cancers. Patients with systemic sclerosis show a four-fold greater risk of lung cancer than the general population. However, the underlying mechanism remains poorly understood.

Methods: The expression profiles of 355 peripheral blood samples were integratedly analyzed, including 70 cases of lung cancer, 61 cases of systemic sclerosis, and 224 healthy controls. After data normalization and cleaning, differentially expressed genes (DEGs) between disease and control were obtained and deeply analyzed by bioinformatics methods. The gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed online by DAVID and KOBAS. The protein–protein interaction (PPI) networks were constructed from the STRING database.

Results: From a total of 14,191 human genes, 299 and 1644 genes were identified as DEGs in systemic sclerosis and lung cancer, respectively. Among them, 64 DEGs were overlapping, including 36 co-upregulated, 10 co-downregulated, and 18 counter-regulated DEGs. Functional and enrichment analysis showed that the two diseases had common changes in immune-related genes. The expression of innate immune response and response to virus-related genes increased significantly, while the expression of negative regulation of cell cycle-related genes decreased notably. In contrast, the expression of mitophagy regulation, chromatin binding and fatty acid metabolism-related genes showed distinct trends.

Conclusions: Stable differences and similarities between systemic sclerosis and lung cancer were revealed. In peripheral blood, enhanced innate immunity and weakened negative regulation of cell cycle may be the common mechanisms of the two diseases, which may be associated with the high risk of lung cancer in systemic sclerosis patients. On the other hand, the counter-regulated DEGs can be used as novelbiomarkers of pulmonary diseases. In addition, fat metabolism-related DEGs were consideredto be associated with clinical blood lipid data.

Keywords: Lung cancer, Systemic sclerosis, Bioinformatics, Immunity

* Correspondence: tjzhy@jnu.edu.cn; liu_dz2001@sina.com
[3]Department of Microbiology and Immunology, College of Basic Medicine and Public Hygiene, Jinan University, Guangzhou 510632, China
[1]Department of Rheumatology and Immunology, Shenzhen People's Hospital, The Second Clinical Medical College of Jinan University, Shenzhen 518020, China
Full list of author information is available at the end of the article

Li *et al. BMC Cancer*   (2021) 21:259

Page 2 of 16

Background

One in six people dies of cancer worldwide and the cancer burden is increasing each year [1]. In 2018, approximately 9.6 million people worldwide died of cancer, from which lung cancer was the most commonly diagnosed cancer [1, 2]. There is some evidence that patients with autoimmune diseases have an increased risk of cancer due to immune system disorders [3, 4]. Dozen of autoimmune diseases could increase the incidence and mortality of lung cancer [5–9], but the underlying regulations of the associated genes' expression remain unclear. The accumulated gene expression data of related diseases in recent years havebrought opportunities to provide clues of mechanism of expressional regulation.

The immune system plays different roles in combatting different diseases. Autoimmune diseases and cancers have certain correlations and differences. Both lung cancer and systemic sclerosis are associated with pulmonary fibrosis [10, 11], and the early diagnosis is challenging. In addition, pulmonary manifestations of lung cancer and systemic sclerosis are easy to be confused with each other, even with the help of radiological and histopathological evaluation. The major risk factors for lung cancer include external risk factors such as smoking, occupational exposure, air pollution, ionizing radiation and diet, as well as internal risk factors e.g. genetics and lung disease history [12, 13]. Although systemic sclerosis is a rare disease, recent studies have shown that it is linkedto an increased risk of lung cancer [14, 15]. A meta-analysis of 16 studies involving more than 7000 patients revealed that, compared with the general population, scleroderma patients had a significantly higher risk of lung cancer (RR 4.35; 95% CI 2.08, 9.09) [16]. A relationship of immune abnormalities in systemic sclerosis and lung cancer is unclear and needs to be investigated to understand the risk association between these two diseases. Although recent studies have shown certain gene signatures of systemic sclerosis or lung cancer [17–19], the accuracy and stability of the results need to be improved due to their small sample size.

Here, we conducted a comprehensive analysis to clarify the relationship of immune-related genes between systemic sclerosis and lung cancer from an intersection of 14,191 genes from four datasets ($n > 15$). Using data from international sources, we discovered evidence of common gene expressional changes between systemic sclerosis and lung cancer, and identified the corresponding biological processes. Most importantly, we discovered immune genes with opposite expressional trends in systemic sclerosis and lung cancer. Together with rich functional annotation, these findings will help revealing the development mechanisms of the two diseases.

Methods

Data sources and patients

DNA microarray analysis is a reliable technology for analyzing gene expression profiles. The gene expression profiling of lung cancer and systemic sclerosis whole blood samples were used in this study. For systemic sclerosis, which is a rare disease, we used the accumulated data from many years. We considered necessary to have large sample sizes to reduce batch differences (number of patients > 15) and included more healthy controls to provide a reliable reference. The accession and platform information for 61 patients with systemic sclerosis, 70 patients with lung cancer, and 224 normal healthy individuals is presented (Table S1).

Patients with systemic sclerosis from January 1, 2018 to December 1, 2020 in the Department of Rheumatology and Immunology of Shenzhen People's Hospital was diagnosed by two or more experienced doctors according to the 2013 classification criteria for systemic sclerosis. Patients with complications were excluded. The data of healthy controls were collected from the data management platform of Shenzhen People's Hospital and matched with each sample according to age (±1 year).

Integration of microarray data

All packages were executedin R software (version 4.0.1). Data normalization and DEGs identification was performed using limma package (*P* value < 0.05). The RRA package was utilized to analyze common DEGs (*P* value < 0.05). Genes with up- or down-regulated expression in three chips were used in subsequent analysis. The packages are publicly available in the Comprehensive R Network. The 299 and 1644 DEGs were identified in systemic sclerosis and lung cancer, respectively, of which 64 DEGs overlapped, including 36 co-upregulated DEGs, 10 co-downregulated DEGs and 18 DEGs with opposite regulatory trends. Pheatmap package was used to generate heatmaps.

Functional annotation and pathway mapping

DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/) was employed for functional annotation of the DEGs primarily associated with either systemic sclerosis or lung cancer [20]. Pathways were enriched from these DEGs using web-based knowledge databases KEGG, GO and BioCarta [21]. KEGG pathway mapping was utilized to generate toll-like receptors signaling pathways.

Data analysis and statistical analysis

Microsoft Excel 2013 was used to store and organize data. GraphPad Prism 5 was used to generate graphics. The DEGs for each disease were derived by comparison with sex- and age-matched controls (within 10%

difference). DEG was defined as gene with fold change > 1.2 or < 0.8. Mean fold change of important genes was visualized by color intensity ($P < 0.05$), red for up-regulation, and green for down-regulation.

Statistical analysis was conducted using SPSS 17.0 software (SPSS, Chicago, USA). The t-test was applied while comparing groups. The significance level was set at $P < 0.05$.

Results

Improving the robustness of gene expression profiles

The whole process of a comprehensive gene expression analysis of primary systemic sclerosis or primary lung cancer is presentedin Fig. 1. At first, the disease and control data were compared for each group, and then the differences between the two diseases were analyzed. The gene expression profiles of peripheral blood from 61 patients with primary systemic sclerosis, 70 patients with primary lung cancer, and 224 normal healthy individuals were obtained from the Gene Expression Omnibus database (Table S1). The datasets with a number of patients below 15 were excluded to reduce batch differences between different experiments. The collected data were pre-processed to meet analysis requirements. To improve reliability of disease analysis, after standardization of the data (Fig. 2a), a principal component analysis was performed (Fig. 2b). All outliers outside the interface of the two diseases were removed (separated by red dotted lines in Fig. 3, listed in Table S1). The obtained gene expression data enabled us

to systematically identify and mine important disease-related genes.

Genomic expression integration and comparison

Based on the principal component analysis, we examined DEGs identified in four groups screened by the limma package (fold change > 1.20/ fold change < 0.80 and P-value < 0.05) (Fig. 4), and performed a cluster analysis of the top-ranked genes (Fig. S1). The number of upregulated genes was equal to that of downregulated genes in each set of data. To improve the reliability of identification of disease-related genes, the DEGs of each group were analyzed and integrated into co-upregulated or co-downregulated genes according to different disease types (Fig. 5a). Overall, 299 DEGs were screened from the systemic sclerosis data (Group I and Group II), including 228 up-regulated genes and 71 down-regulated genes. Additionally, 1644 DEGs were screened from the lung cancer data, including 991 upregulated genes and 653 downregulated genes (Group III and Group IV). The results showed that DEGs in lung cancer were about five times as those DEGs in systemic sclerosis.

The acquisition of common DEGs in all patients laid a solid foundation for our analysis of the differences and links between the two diseases. Unlike the above analysis revealing commonalities of all patients with the same disease, not only the same regulatory trend but also the distinction between systemic sclerosis and lung cancer attracted our great interest. We investigated 299 DEGs
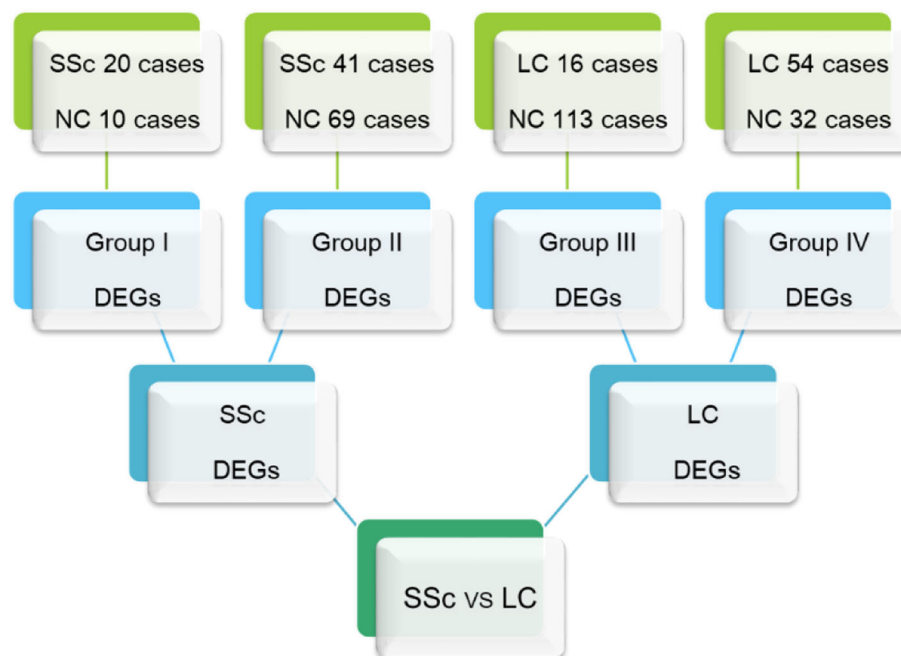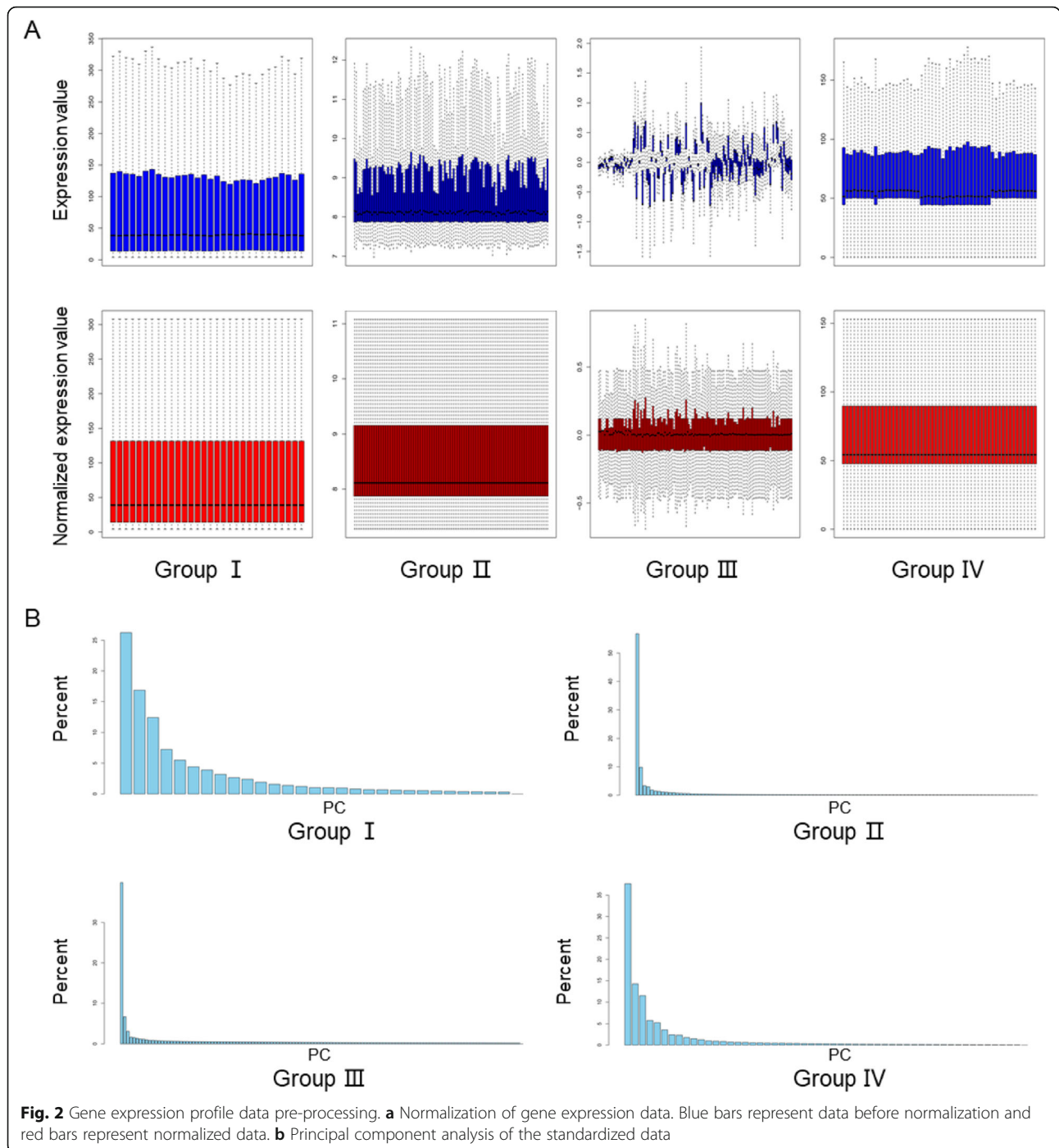

Fig. 1 A scheme of systemic sclerosis (Group I and Group II) and lung cancer (Group III and Group IV) gene expression analysis. SSc: systemic sclerosis, LC: lung cancer, NC: normal control, DEG: differentially expressed gene

Fig. 2 Gene expression profile data pre-processing. **a** Normalization of gene expression data. Blue bars represent data before normalization and red bars represent normalized data. **b** Principal component analysis of the standardized data
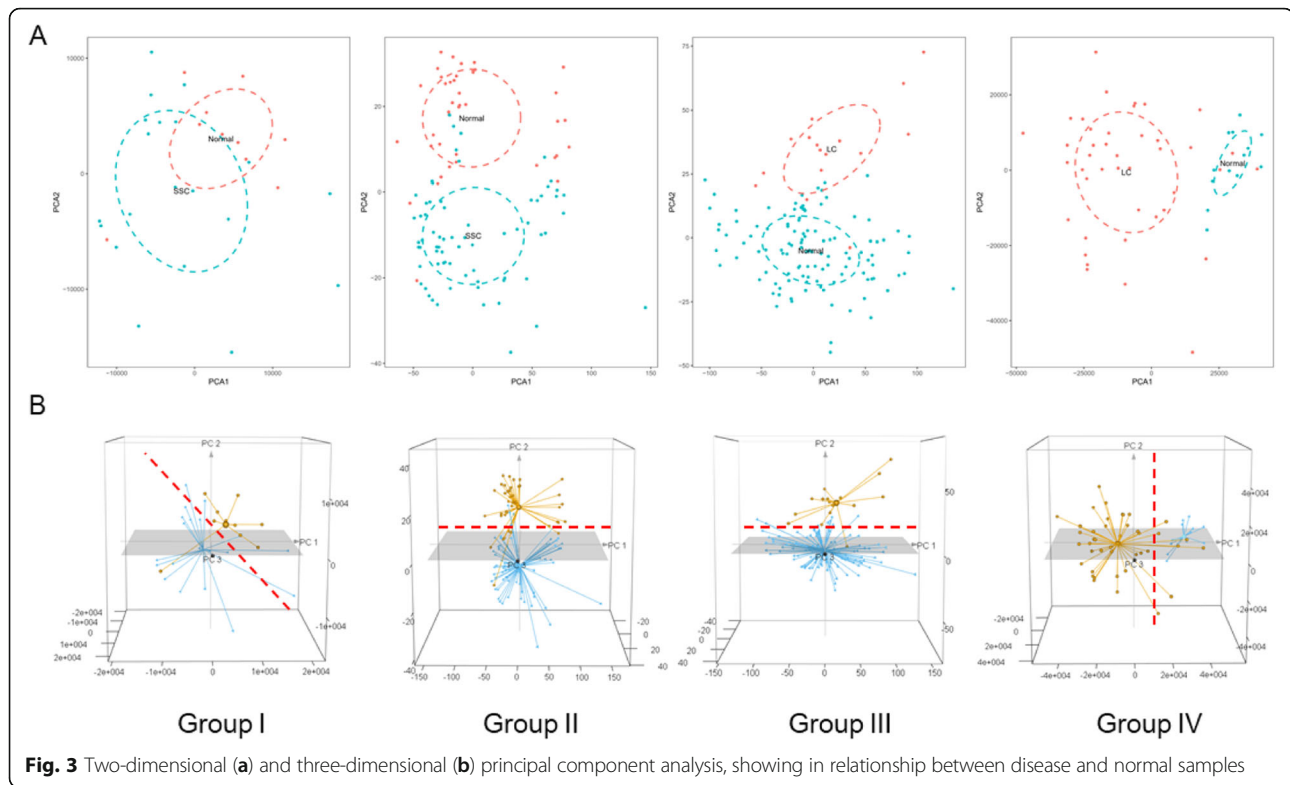
in systemic sclerosis and 1644 DEGs in lung cancer. Finally found 64 overlapping DEGs, which accounted for 22.07 and 4.02% of the total DEGs in systemic sclerosis and lung cancer. It indicated that there is a small proportion of overlapping DEGs between the two diseases. Among the overlapping DEGs, 36 co-upregulated DEGs, 10 co-downregulated DEGs, and 18 DEGs with opposite regulatory trends were identified, and details of the changes were displayed (Fig. 5b and 6). Co-upregulated

DEGs accounted for the largest proportion, while a few genes were co-downregulated. Most of the DEGs with opposite regulatory trends were upregulated in systemic sclerosis but downregulated in lung cancer.

Gene expression overlap shared by both systemic sclerosis and lung cancer

Considering the above fact that co-up-regulated and co-down-regulated genes were dominant (72%) in the above

Li *et al. BMC Cancer*   (2021) 21:259

Page 5 of 16

Fig. 3 Two-dimensional (**a**) and three-dimensional (**b**) principal component analysis, showing in relationship between disease and normal samples
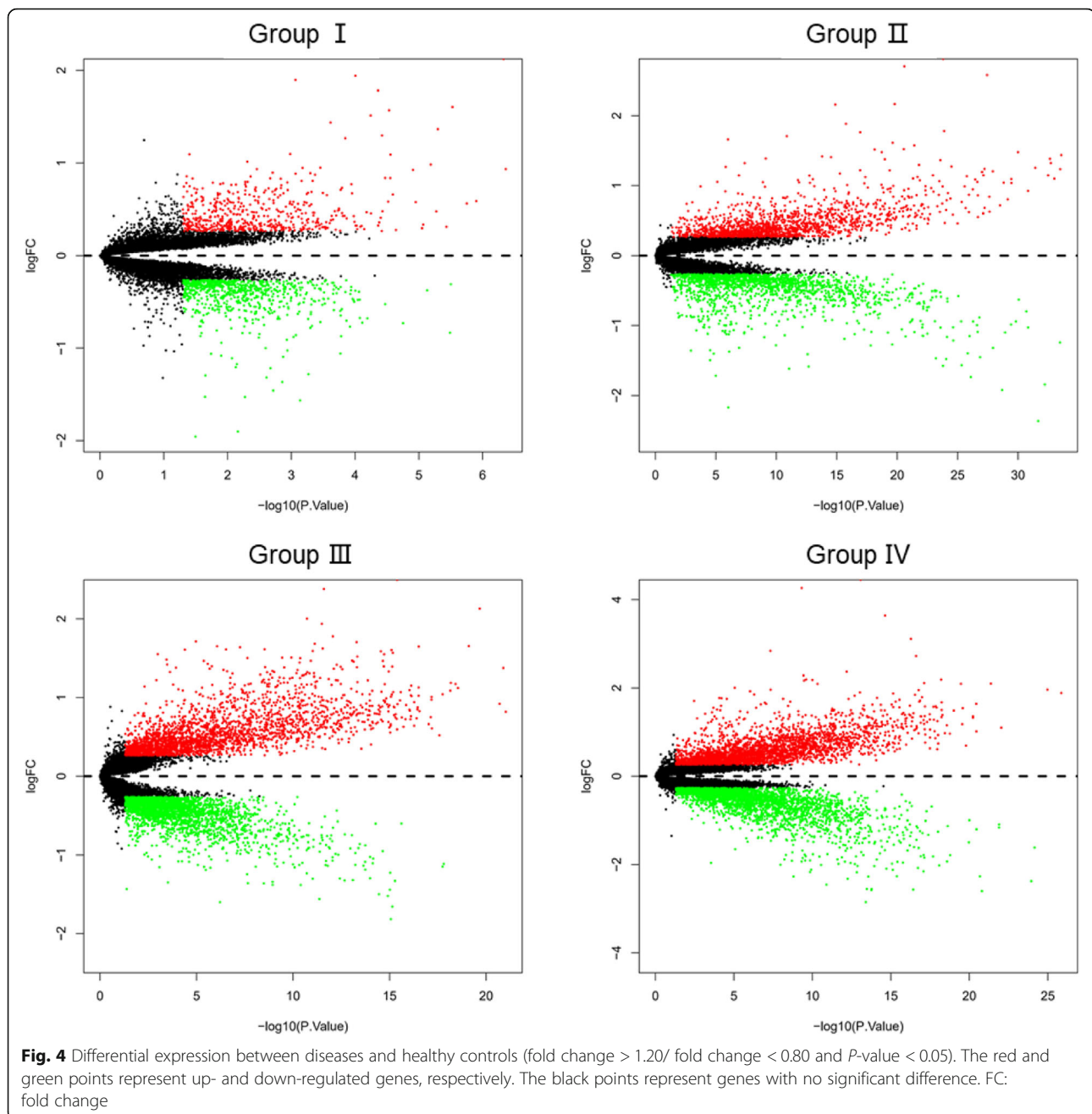
comparative analysis, we subsequently focused on the gene expression overlap between systemic sclerosis and lung cancer. First, we performed a functional enrichment analysis to explore the same intrinsic mechanisms of both diseases. Through the analysis of 36 co-upregulated DEGs distribution, these DEGs were found to come from different pathways (Fig. 7). Most of the DEGs (24/ 36) were related to regulation of innate immune response (GO: 0045088). Except for CISH, the top 12 DEGs were all associated with innate immune response showing its critical role in fighting both diseases. Secondly, these DEGs are mainly mediated by a series of cytokine signaling pathway (GO: 0019221), including regulation of cytokine secretion (GO: 0050707) and cytokine secretion involved in immune response (GO: 0002374). Other GO terms include cellular response to interferon-gamma, response to virus, myeloid leukocyte activation, regulation of B cell proliferation, response to fatty acid, cytokine-cytokine receptor interaction, and regulation of peptidyl-tyrosine phosphorylation. These terms were characterized by antiviral, immunomodulatory, and anti-tumor properties.

Further enrichment analysis showed that the co-upregulated DEGs are mainly enriched in cytokine-mediated signaling pathway, regulation of innate immune response, cellular response to interferon-gamma and response to virus (Fig. 8a). Among them, innate immunity is located in the core position, which connects

regulation of B cell proliferation, myeloid leukocyte activation and cellular response to interferon-gamma (Fig. 8b). These results indicate that cytokine-mediated signaling pathway and regulation of innate immune response were the key biological processes in both systemic sclerosis and lung cancer compared with healthy controls. The co-upregulated DEGs were involved in a series of cytokine-mediated signaling pathways, affecting various receptors on the cell surface. Through KEGG pathway analysis, we found that TLR2, TLR4, TLR6, TLR8, IFNAR1, IFNGR2 and IRF7 in toll-like receptors signaling pathway were significantly upregulated in both diseases, thereby affecting the innate immunity of cells (Fig. 8c).

On the other hand, GO enrichment analysis of 10 co-down-regulated DEGs yielded three terms: signaling by interleukins, response to growth factor and lymphocyte activation, each term contained three DEGs (Fig. 9a), the all three being significant (Fig. 9b). The down-regulation of signaling by interleukins indicates that the communication between lymphocytes may be weakened, which may influencethe function of lymphocytes. Despite lower significance compared with signaling by interleukins, reduced growth factor and lymphocyte activation also weakened the function of immune system. The down-regulation of these terms may affect the maturation, proliferation and activation of immune cells.

Fig. 4 Differential expression between diseases and healthy controls (fold change > 1.20/ fold change < 0.80 and *P*-value < 0.05). The red and green points represent up- and down-regulated genes, respectively. The black points represent genes with no significant difference. FC: fold change
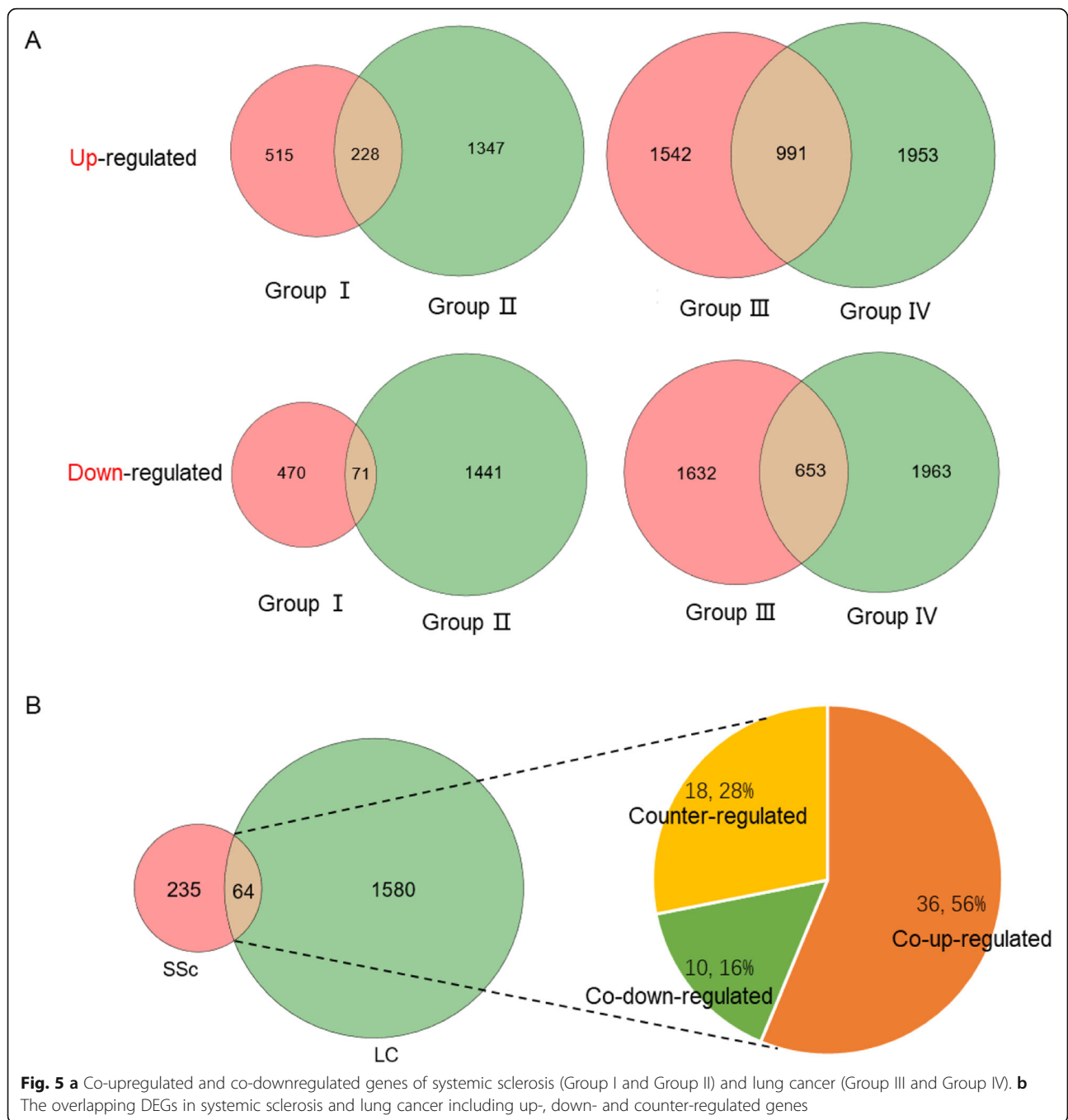
Potential differential pathways between systemic sclerosis and lung cancer

To identify the core gene expression differences between systemic sclerosis and lung cancer, we compared the most important DEGs with the opposite trends. The expression difference of DEGs in the two diseases was visualized by the total lengths of blue and green columns (Fig. 10). Most genes were up-regulated in systemic sclerosis and down-regulated in lung cancer, except for CTSK and MYLIP. The top DEGs with the greatest differences were GIMAP7, CX3CR1, MYLIP and ASF1A.

Among them, GIMAP7 was the most significantly down-regulated in lung cancer, CX3CR1 was the most upregulated and MYLIP was the most downregulated in systemic sclerosis.

Therefore, we performed an enrichment analysis to further identify the potential mechanisms of these associations. Using databases including KEGG, GO, and Bio-Carta, six terms were enriched. The signaling pathways of DEGs were primarily enriched in regulation of mitophagy ($P = 0.027$), followed by chromatin binding, positive regulation of protein targeting to the

Li *et al. BMC Cancer*   (2021) 21:259

Page 7 of 16

Fig. 5 a Co-upregulated and co-downregulated genes of systemic sclerosis (Group I and Group II) and lung cancer (Group III and Group IV). **b** The overlapping DEGs in systemic sclerosis and lung cancer including up-, down- and counter-regulated genes

mitochondrion, phosphoprotein and fatty acid metabolism (Fig. 11a and b). The DEG list was imported into STRING database (http://string-db.org) to explore their interrelationships and calculate the characteristics of the network (Fig. 11c), with a total of 18 DEGs. Except that CTSK is isolated, a complex network was constructed, in which MYLIP and CDC23 had the highest confidence.

MYLIP is involved in low density lipoprotein receptor (LDLR) degradation, also considering that the co-upregulated DEGs ACADM, NOC3L and AASDH are involved in fatty acid metabolism. This implies that fatty acid metabolism may be associated with the disease process. We retrospectively analyzed lipid data from 32 patients with systemic sclerosis, healthy controls were individually matched according to sex and age. The concentrations of triglycerides, total cholesterol, high density lipoprotein cholesterol (HDL-C), and low density lipoprotein cholesterol (LDL-C) are shown, and overall levels in healthy individuals are within normal reference ranges (Fig. 12).
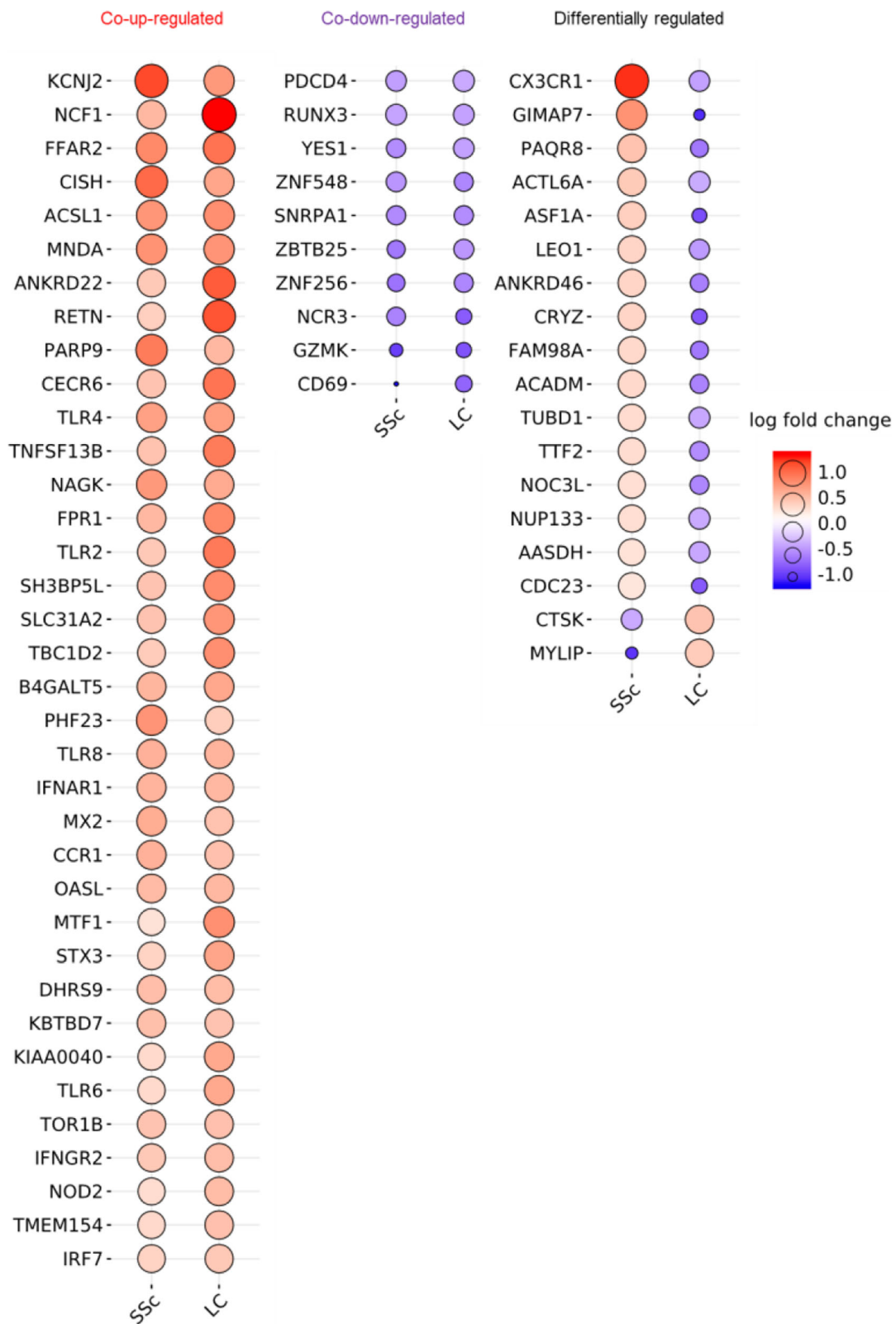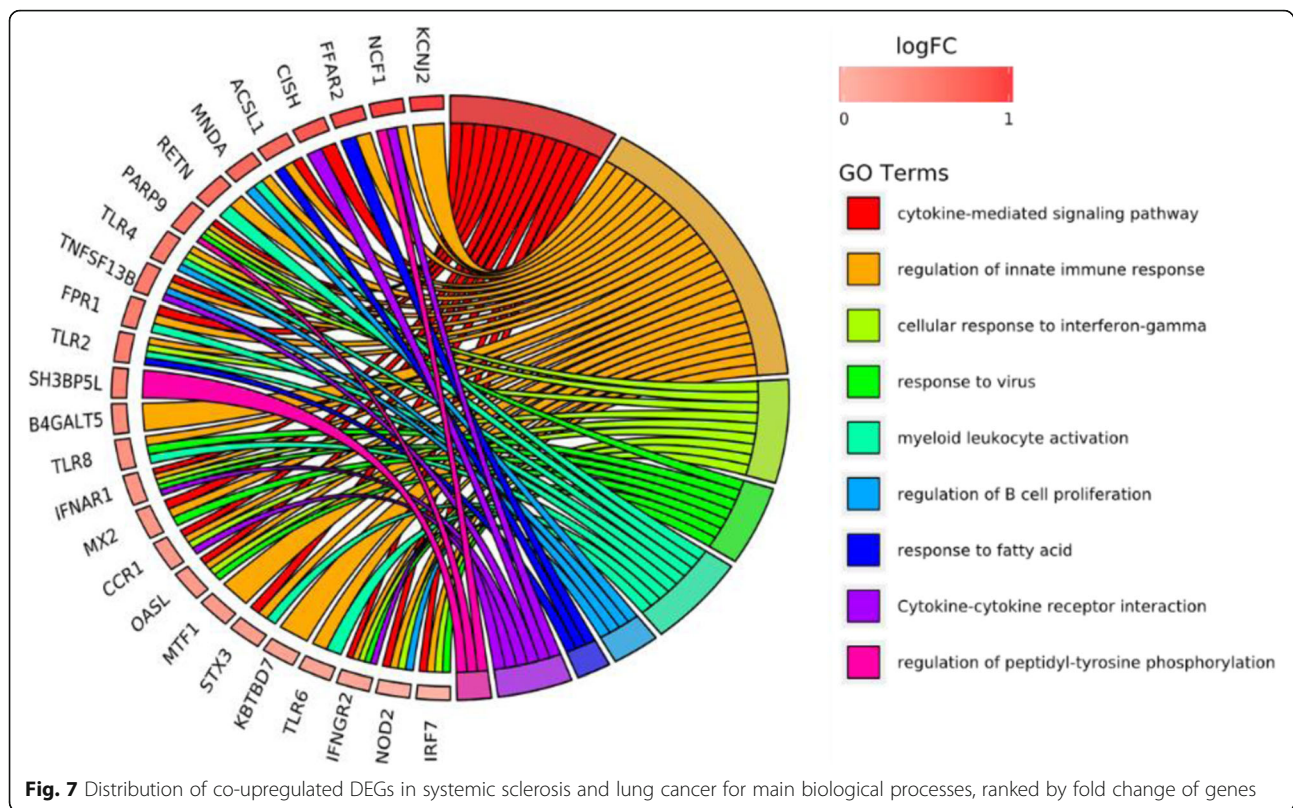
Li et al. BMC Cancer (2021) 21:259

Page 8 of 16

Fig. 6 Overview of DEGs present in both systemic sclerosis and lung cancer. Red and large circles indicate up-regulation, purple and small circles indicate down-regulation.

Li et al. BMC Cancer (2021) 21:259

Page 9 of 16

Fig. 7 Distribution of co-upregulated DEGs in systemic sclerosis and lung cancer for main biological processes, ranked by fold change of genes
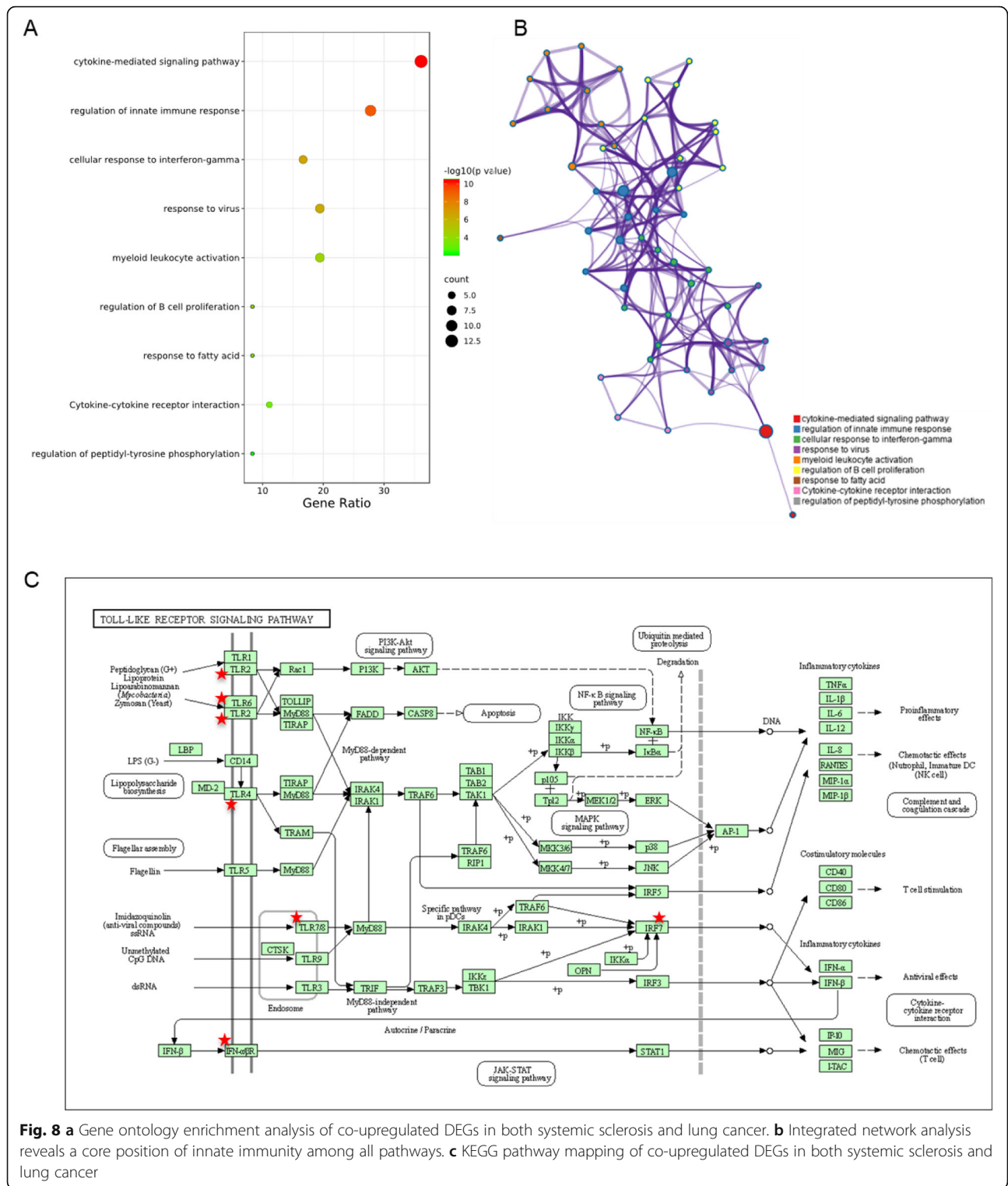
Discussion

Lung cancer has the highest morbidity and mortality among all cancers [1, 12], and its early diagnosis has significant challenges. Pulmonary symptoms of various diseases interfere with the diagnosis of lung cancer, especially some autoimmune diseases that can increase the incidence of lung cancer. Blood-based biomarkers would be very useful for early diagnosis of lung cancer. Disease-related gene expression characteristics in peripheral blood mononuclear cells (PBMC) have been described in several types of cancer. However, RNA-stabilized whole blood technology could be more applicable and powerful in the clinic. We conducted a comprehensive study of immune-related genes in systemic sclerosis and lung cancer based on gene expression profiling in whole blood. Our results show that systemic sclerosis and lung cancer share a common trend of change in innate immunity and cell cycle regulation, while in some pathways there are opposite characteristics. These results improve our understanding of the different strategies of the immune system to combat lung cancer and systemic sclerosis, and prompt further development of biomarkers based on gene expression in peripheral blood for early detection of lung diseases.
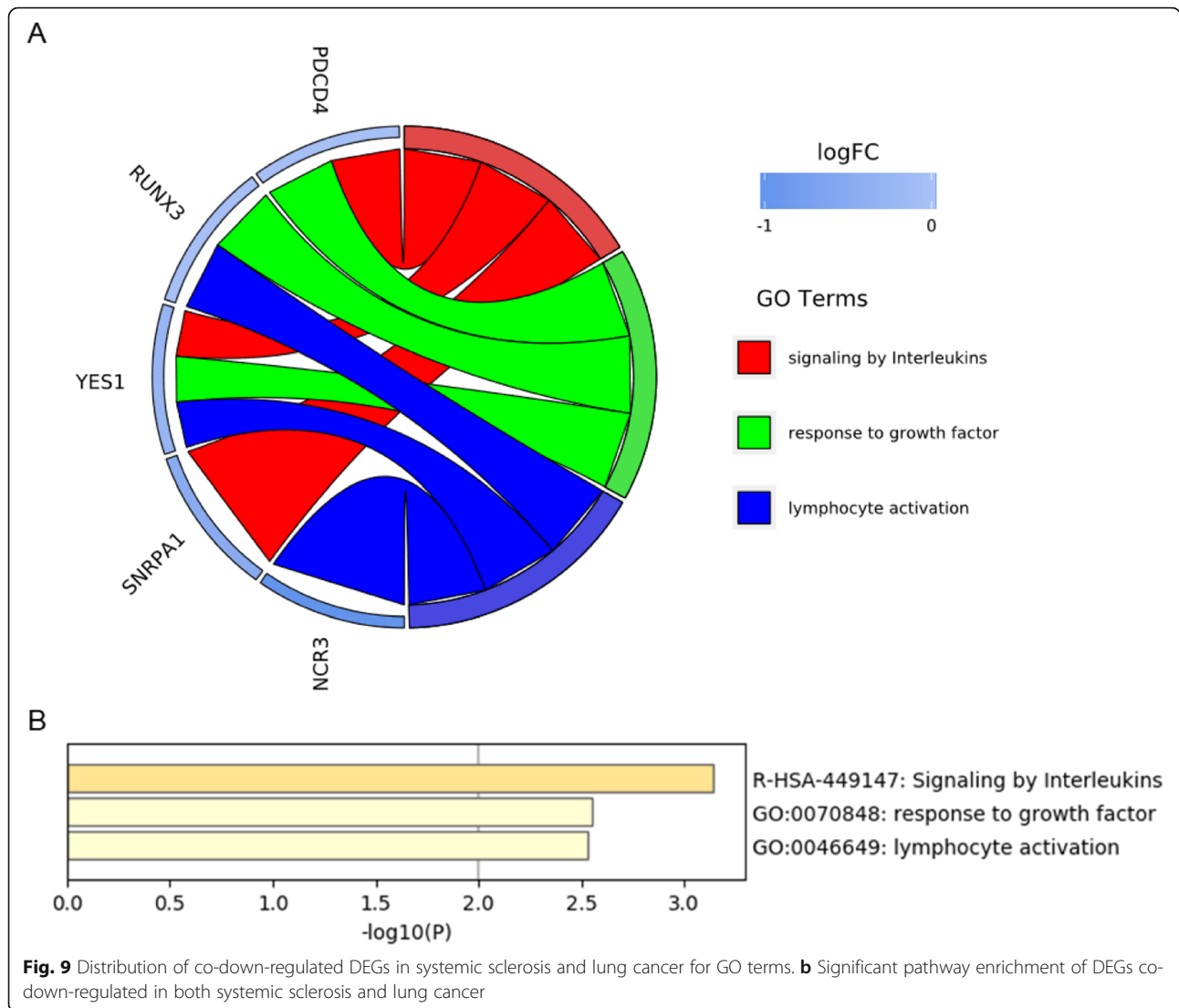
Pulmonary manifestations are known to occur in all autoimmune diseases that have standardized incidence ratios greater than 2.0, indicating that autoimmune

process can lead to lung cancer susceptibility [8]. Systemic sclerosis is one of the connective tissue diseases with the highest mortality rate [11, 22], and it has recently been reported that patients with systemic sclerosis have an increased risk of lung cancer [23, 24], but the intrinsic gene connection is unclear. Our results reveal that genes related to innate immune response and response to virus were co-upregulated significantly in patients with systemic sclerosis and lung cancer. And the expression of genes related to negative regulation of cell cycle and transcription factor activity decreased considerably. In contrast, the expression of genes related to mitophagy regulation and chromatin binding showed a clear opposite trend. Although the fold changes of DEGs just ranged from 0.46 to 2.66, these changes were evident and stable in all patients in this study, providing some reliable support for the intrinsic association between systemic sclerosis and lung cancer. It has been previously reported that certain genes may be critical in the pathogenesis of systemic sclerosis (chemokine CCL2 and CXCL4 [25]) and lung cancer (including EGFR and HER2 [26, 27]), but our results are more convincing in terms of reliability and universality owing to the large sample size. It has also been hypothesized that certain treatments for systemic sclerosis (e.g., immunosuppressants) may lead to a decline in immunity, which in turn increases the cancer risk [28]. However, this cannot

Fig. 8 **a** Gene ontology enrichment analysis of co-upregulated DEGs in both systemic sclerosis and lung cancer. **b** Integrated network analysis reveals a core position of innate immunity among all pathways. **c** KEGG pathway mapping of co-upregulated DEGs in both systemic sclerosis and lung cancer

explain the higher degree of pulmonary fibrosis in lung cancer patients. In summary, this evidence provides some support for the potential immune system association between systemic sclerosis and lung cancer.

In a recent study of the demographic and clinicopathological characteristics of lung cancer in patients with systemic sclerosis, 1.4% patients developed systemic sclerosis before the onset of lung cancer (~ 13 years

Li *et al. BMC Cancer*   (2021) 21:259

Page 11 of 16

Fig. 9 Distribution of co-down-regulated DEGs in systemic sclerosis and lung cancer for GO terms. **b** Significant pathway enrichment of DEGs co-down-regulated in both systemic sclerosis and lung cancer

averagely) [29]. The distribution of histological types of cancer is similar to that of the general population, most of which are adenocarcinomas, followed by squamous and small cell carcinomas. This is consistent with the reported histological pattern of systemic sclerosis-associated lung cancer [30]. Notably, none of these patients ever smoked, suggesting that systemic sclerosis is an independent risk factor for lung cancer. The study confirmed a lack of common cancer driver for gene mutations in these patients, further supporting the possible role of chronic autoimmune inflammation in carcinogenesis of systemic sclerosis-related lung cancer. However, the study did not reveal changes in disease-related genes.

Although innate immunity is enhanced in both diseases, its effect on the diseases is complicated. Excessive innate immunity may negatively affect systemic sclerosis, while it may contribute to the elimination of cancer cells. On the other hand, among the 10 significantly co-downregulated genes, PDCD4 and YES1 are related to apoptosis [31, 32], while NCR3 and CD69 are related to cell killing [33, 34]. It can be speculated that the loss of relevant roles is one of the common pathogenic causes. Both PDCD4 and RUNX3 are tumor suppressor genes [35–37], and their co-downregulation may be one of the causes of lung cancer.

Among the counter-regulated DEGs, ACTL6A and CTSK belonged to mitophagy regulation, ASF1A, NOC3L and ACTL6A belonged to chromatin binding. Although ACTL6A is involved in both pathways, in fact it is a component of chromatin remodeling complexes and is indirectly involved in transcriptional activation or repression [38], the significance remains to be determined. CTSK displays potent endoprotease activity against fibrinogen [39], plays an important role in extracellular matrix degradation, its counter-regulation in

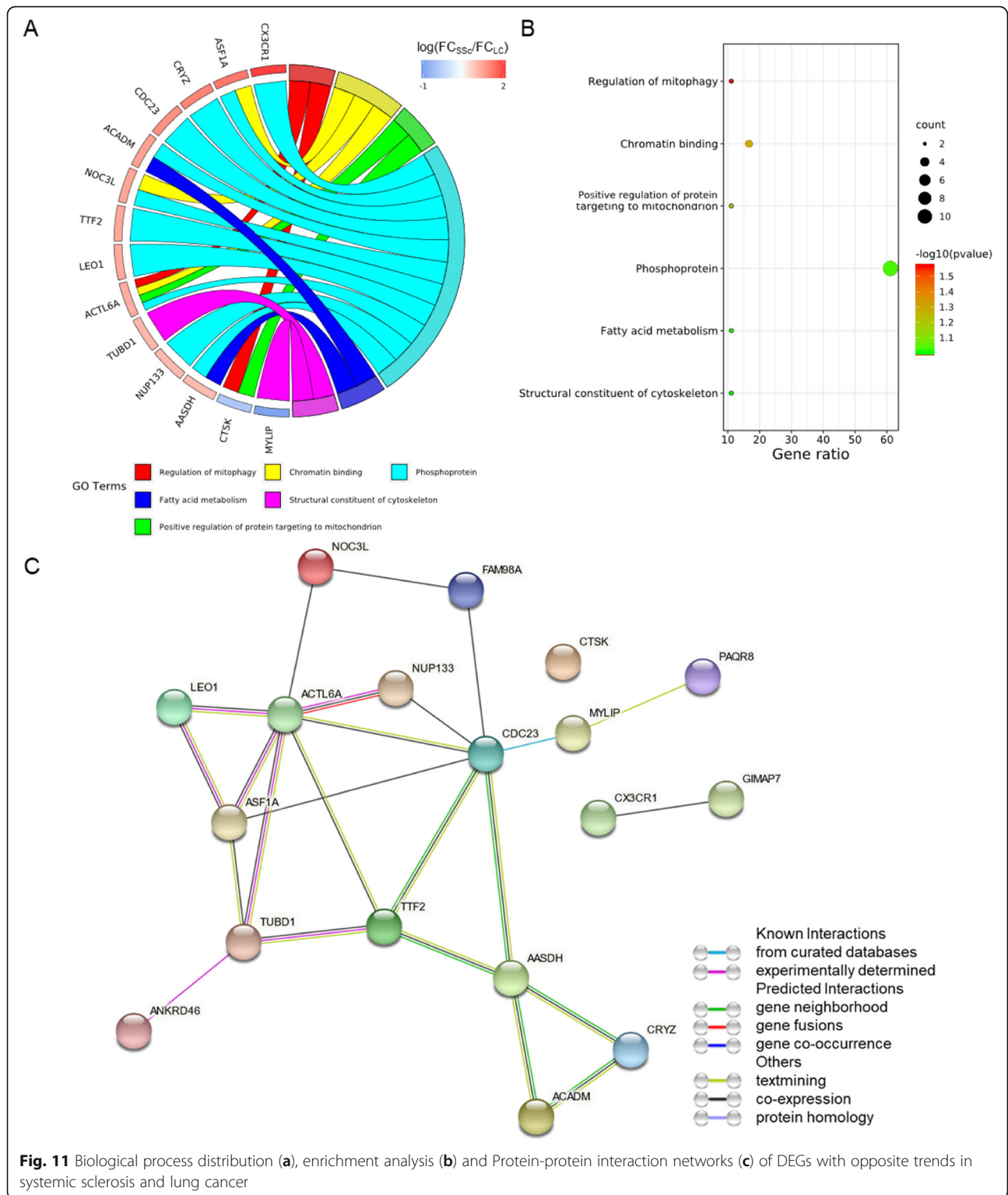**Fig. 10** KEGG analysis revealed associations between genes or pathways
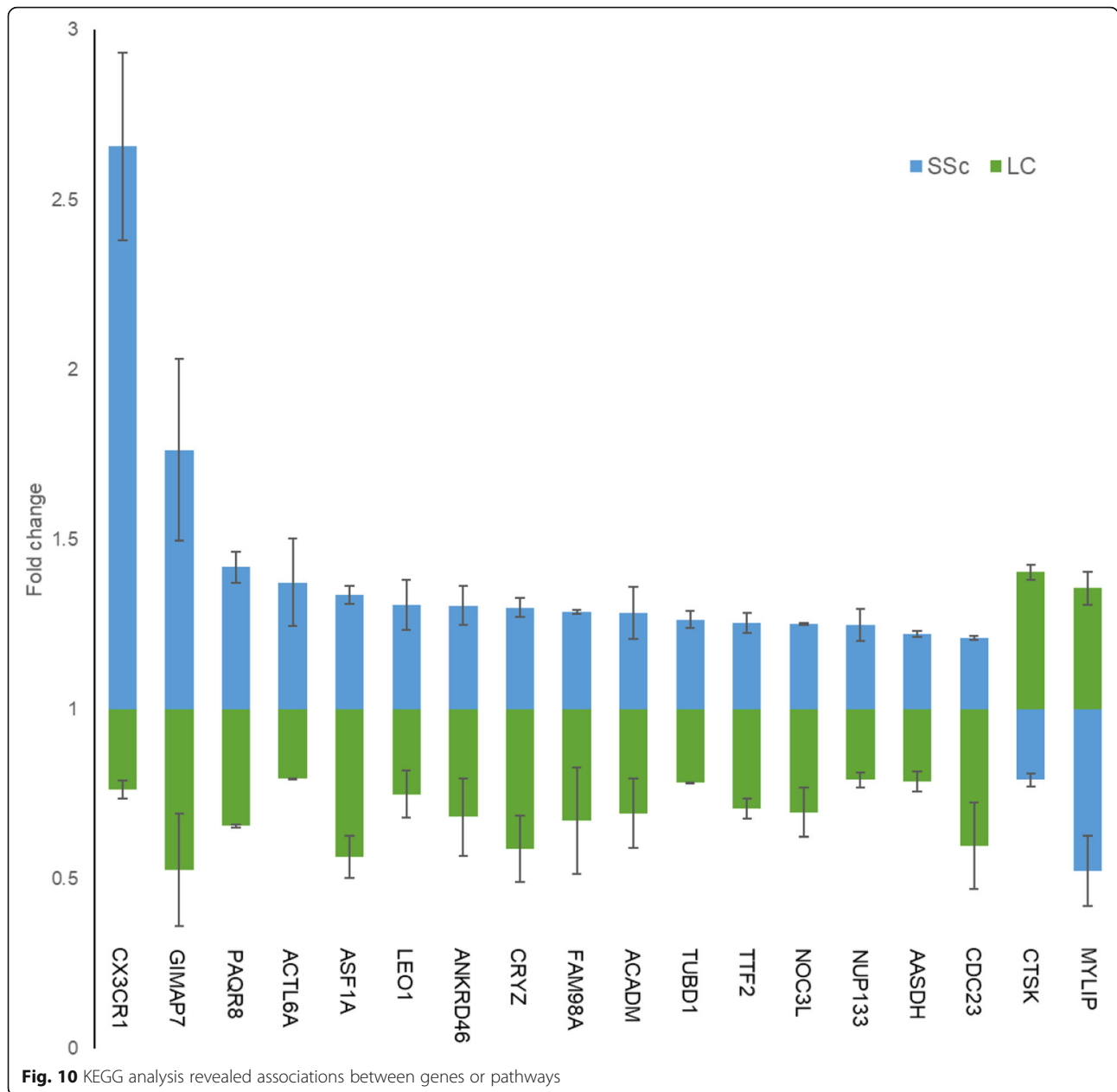
lung cancer and systemic sclerosis may lead to different effects on the extracellular matrix. NOC3L is a homolog of nucleolar complex protein 3, may be required for adipogenesis. ASF1A functions as a histone chaperone in nucleosome assembly and disassembly [40]. As chromatin binding proteins, NOC3L and ASF1A may be indirectly related to diseases.

GIMAP (GTPase of the immunity-associated protein) gene family encodes unique GTPases, most of which functions are unknown. Abnormal expression of GIMAP family members in cancer tissues has also been reported in non-small cell lung cancer, and qPCR analysis showed downregulation of GIMAP6 and GIMAP8 in lung cancer
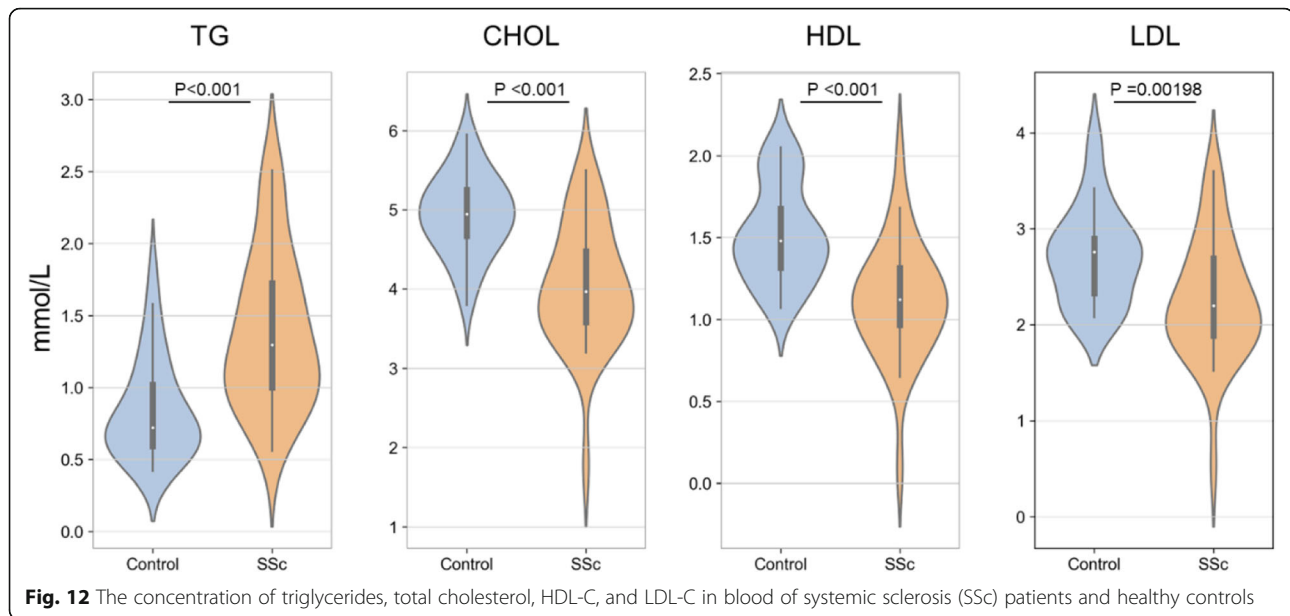
tissues [41]. Another report showed that downregulation of GIMAP7 at both protein and mRNA levels observed in serum and tissue samples of oral cancer patients may imply that GIMAP7 has an anticancer effect [42]. Downregulation of GIMAP gene may regulate immune cell viability or development affecting cancer progression [43], however, high expression of GIMAP7 in systemic sclerosis has not been reported, which needs further exploration.

The important role of chemokine receptors in disease pathogenesis has caused much attention. In peripheral blood, chemokine receptor CX3CR1 is present on the cell membrane of monocytes and leukocytes [44, 45], it is not

Fig. 11 Biological process distribution (**a**), enrichment analysis (**b**) and Protein-protein interaction networks (**c**) of DEGs with opposite trends in systemic sclerosis and lung cancer

only involved in cell chemotaxis, but is also related to cell adhesion [46]. The arrival of inflammatory cells from the peripheral circulation to the site of inflammation is a dynamic, multistep process in which CX3CR1 plays a crucial role [47]. An increased level of CX3CR1 in systemic sclerosis may help to recruit immune cells to the site of inflammation, giving CX3CR1 an important role in the pathogenesis and development of inflammatory response in autoimmune diseases. Animal experiments and preclinical studies with blocking CX3CR1 signaling showed good

Fig. 12 The concentration of triglycerides, total cholesterol, HDL-C, and LDL-C in blood of systemic sclerosis (SSc) patients and healthy controls

anti-inflammatory effects. However, a decreased level of CX3CR1 in lung cancer may be related to immune weakness and attenuate cell adhesion, which may be associated with cell metastasis. The co-expression characteristics of CX3CR1 and GIMAP7 may play an overlapping role in the occurrence and development of both diseases. Anti-silencing function protein 1 homolog A (ASF1A) is an abundant histone remodeling chaperone in meiosis phase 2. Recent studies uncover Asf1a as a tumor-intrinsic suppressor of immune checkpoint blockade through suppression of GM-CSF expression [48]. Functional analysis of another study showed that the interaction between ASF1A and E2 ubiquitin-binding enzymes was associated with tumorigenesis [49].

Myosin regulatory light chain interacting protein (MYLIP, also named IDOL) is an important ubiquitin E3 ligase, which mediates low density lipoprotein receptor (LDLR) degradation through ubiquitination reaction and affect blood lipid levels [50]. Previous reports indicated the LDLR-related protein (LRP1B) was discovered as a putative tumor suppressor and frequently inactivated in lung cancer cells [51, 52]. It may be related to the upregulation of protein MYLIP which mediates LDLR degradation. In this study, dyslipidemia was found to be fairly common in systemic sclerosis patients, with an average increase in triglycerides of 64.4% ($P = 0.012$) and a predominant decrease (17.0–26.7%) in total cholesterol, HDL-C and LDL-C, which may be associated with changes of lipoprotein receptor. A recent study in Cell Reports revealed that short chain fatty acids in the gut and blood can regulate macrophages to suppress bacterial infection by activating free fatty acid receptor 2 (FFAR2) [53]. Therefore, the relationship between fat metabolism and immune regulation requires more investigation. In addition, the increased level of chain fatty acid CoA ligase 1 (ACSL1) may contribute to abnormal lipid level.

Conclusion

Our study detected the association and differences in immune-related genes between systemic sclerosis and lung cancer using multiple cohorts based on different populations, which will help us understand the relationship of autoimmune diseases and cancers, as well as the mechanisms by which the immune system responds to different diseases. The DEG overlap between systemic sclerosis and lung cancer may partly explain the clinical lung phenotypic association, and the counter-regulated DEGs may provide potential molecular diagnostic markers and biological clues for the two diseases. However, further studies are needed to reveal how the identified gene signatures relate to the pathogenesis of both diseases.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-021-07959-6.

Additional file 1: Table S1. Data of 61 patients with systemic sclerosis, 70 patients with lung cancer and 224 normal healthy individuals. Underlined accessions are removed by principal component analysis. **Fig. S1.** Hierarchical clustering heatmap of top-ranked DEGs screened from systemic sclerosis (Group I and Group II) and lung cancer (Group III and Group IV).

Availability of data and materials
All data of this study are included in this published article.

Declarations
The authors report no conflicts of interest in this work.

Ethics approval and consent to participate
The study was conducted in accordance with the Helsinki Declaration and approved by the Medical Ethics Committee of Shenzhen People's Hospital. Patient personal information is not involved in this paper and only overall results are reported. All patients signed a written informed consent before collecting clinical data.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interest.

Author details
[1]Department of Rheumatology and Immunology, Shenzhen People's Hospital, The Second Clinical Medical College of Jinan University, Shenzhen 518020, China. [2]Integrated Chinese and Western Medicine Postdoctoral Research Station, Jinan University, Guangzhou 510632, China. [3]Department of Microbiology and Immunology, College of Basic Medicine and Public Hygiene, Jinan University, Guangzhou 510632, China. [4]The First Affiliated Hospital (Shenzhen People's Hospital) Southern University of Science and Technology, Shenzhen 518055, China.

References
1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ, Wu YL, Paz-Ares LJL. Lung cancer: current therapies and new targeted treatments. 2017; 389(10066):299–311.
3. Yves A, Jeremy K, Anne-Marie B, Valérie J, Nicolas C, Baptiste H, Yoland S, Aude R, Laurent G, Lucile M High risk of cancer in autoimmune necrotizing myopathies: usefulness of myositis specific antibody. Brain. 2016;139(Pt 8): 2131–5.
4. Haanen J, Ernstoff MS, Wang Y, Menzies AM, Puzanov I, Grivas P, Larkin J, Peters S, Thompson JA, Obeid M. Autoimmune diseases and immune-checkpoint inhibitors for cancer therapy: review of the literature and personalized risk-based preventionstrategy. Ann Oncol. 2020;31(6):724–44.
5. Edson-Heredia, Zhu, Lefevre, Wang, Barrett, Bushe CJ, Cox JJ. DERMATOL M-CJJEA: Prevalence and incidence rates of cardiovascular, autoimmune, and other diseases in patients with psoriatic or psoriatic arthritis: a retrospective study using clinical practice research Datalink; 2015.
6. Lorenzo R, Argibay A, Sousa A, Val N, Baroja A, Maure B, Freire M, Rivera A. AB0501Sjgren Syndrome, Cancer Incidence and Mortality in Vigo Area. JAotRD. 2016;75(Suppl 2):1076.1074–7.
7. Telomeres Mendelian Randomization Collaboration, Haycock PC, Burgess S, Nounu A, Zheng J, Okoli GN, Bowden J, Wade KH, Timpson NJ, Evans DM, Willeit P, Aviv A, Gaunt TR, Hemani G, Mangino M, Ellis HP, Kurian KM, Pooley KA, Eeles RA, Lee JE, Fang S, Chen WV, Law MH, Bowdler LM, Iles MM, Yang Q, Worrall BB, Markus HS, Hung RJ, Amos CI, Spurdle AB, Thompson DJ, O'Mara TA, Wolpin B, Amundadottir L, Stolzenberg-Solomon R, Trichopoulou A, Onland-Moret NC, Lund E, Duell EJ, Canzian F, Severi G, Overvad K, Gunter MJ, Tumino R, Svenson U, van Rij A, Baas AF, Bown MJ, Samani NJ, van t'Hof FNG, Tromp G, Jones GT, Kuivaniemi H, Elmore JR, Johansson M, Mckay J, Scelo G, Carreras-Torres R, Gaborieau V, Brennan P, Bracci PM, Neale RE, Olson SH, Gallinger S, Li D, Petersen GM, Risch HA, Klein AP, Han J, Abnet CC, Freedman ND, Taylor PR, Maris JM, Aben KK, Kiemeney LA, Vermeulen SH, Wiencke JK, Walsh KM, Wrensch M, Rice T, Turnbull C, Litchfield K, Paternoster L, Standl M, Abecasis GR, SanGiovanni JP, Li Y, Mijatovic V, Sapkota Y, Low SK, Zondervan KT, Montgomery GW, Nyholt DR, van Heel DA, Hunt K, Arking DE, Ashar FN, Sotoodehnia N, Woo D, Rosand J, Comeau ME, Brown WM, Silverman EK, Hokanson JE, Cho MH, Hui J, Ferreira MA, Thompson PJ, Morrison AC, Felix JF, Smith NL, Christiano AM, Petukhova L, Betz RC, Fan X, Zhang X, Zhu C, Langefeld CD, Thompson SD, Wang F, Lin X, Schwartz DA, Fingerlin T, Rotter JI, Cotch MF, Jensen RA, Munz M, Dommisch H, Schaefer AS, Han F, Ollila HM, Hillary RP, Albagha O, Ralston SH, Zeng C, Zheng W, Shu XO, Reis A, Uebe S, Hüffmeier U, Kawamura Y, Otowa T, Sasaki T, Hibberd ML, Davila S, Xie G, Siminovitch K, Bei JX, Zeng YX, Försti A, Chen B, Landi S, Franke A, Fischer A, Ellinghaus D, Flores C, Noth I, Ma SF, Foo JN, Liu J, Kim JW, Cox DG, Delattre O, Mirabeau O, Skibola CF, Tang CS, Garcia-Barcelo M, Chang KP, Su WH, Chang YS, Martin NG, Gordon S, Wade TD, Lee C, Kubo M, Cha PC, Nakamura Y, Levy D, Kimura M, Hwang SJ, Hunt S, Spector T, Soranzo N, Manichaikul AW, Barr RG, Kahali B, Speliotes E, Yerges-Armstrong LM, Cheng CY, Jonas JB, Wong TY, Fogh I, Lin K, Powell JF, Rice K, Relton CL, Martin RM, Davey Smith G. Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. JAMA Oncol. 2017; 3(5):636–51.
8. Hemminki K, Liu X, Ji J, Sundquist J, Sundquist K. Effect of autoimmune diseases on risk and survival in histology-specific lung cancer. Eur Respir J. 2012;40(6):1489–95.
9. Onishi A, Sugiyama D, Kumagai S, Morinobu A. Cancer incidence in systemic sclerosis: meta-analysis of population-based cohort studies. Arthritis Rheum. 2013;65(7):1913–21.
10. Tzouvelekis A, Gomatou G, Bouros E, Trigidou R, Tzilas V, Bouros DJC. Common Pathogenic Mechanisms Between Idiopathic Pulmonary Fibrosis and Lung Cancer. Chest. 2019;156(2):383–91.
11. Denton CP, Khanna D. Systemic sclerosis. Lancet (London, England). 2017; 390(10103):1685–99.
12. Islami F, Goding Sauer A, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ, McCullough ML, Patel AV, Ma J, Soerjomataram I, Flanders WD, Brawley OW, Gapstur SM, Jemal A. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. CA Cancer J Clin. 2018;68(1):31–54.
13. Villalobos P, Wistuba II. Lung Cancer biomarkers. Hematol Oncol Clin North Am. 2017;31(1):13–29.
14. Okafor BC. Incidence of Menière's disease. J Laryngol Otol. 1984;98(8):775–9.
15. Maria ATJ, Partouche L, Goulabchand R, Rivière S, Rozier P, Bourgier C, Le Quellec A, Morel J, Noël D, Guilpain P. Intriguing Relationships Between Cancer and Systemic Sclerosis: Role of the Immune System and Other Contributors. Front Immunol. 2019;9:3112.
16. Bonifazi M, Tramacere I, Pomponio G, Gabrielli B, Avvedimento EV, La Vecchia C, Negri E, Gabrielli A. Systemic sclerosis (scleroderma) and cancer risk: systematic review and meta-analysis of observational studies. Rheumatology (Oxford, England). 2013;52(1):143–54.
17. Showalter K, Spiera R, Magro C, Agius P, Martyanov V, Franks JM, Sharma R, Geiger H, Wood TA, Zhang Y, et al. Machine learning integration of scleroderma histology and gene expression identifies fibroblast polarisation as a hallmark of clinical severity and improvement. Ann Rheum Dis. 2021; 80(2):228–37.
18. Taroni JN, Greene CS, Martyanov V, Wood TA, Christmann RB, Farber HW, Lafyatis RA, Denton CP, Hinchcliff ME, Pioli PA. A novel multi-network approach reveals tissue-specific cellular modulators of fibrosis in systemic sclerosis. Genome Med. 2017;9(1):27.
19. Dashtban M, Balafar M, Suravajhala P. Gene selection for tumor classification using a novel bio-inspired multi-objective approach. Genomics. 2018;110(1):10–7.
20. Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4(1):44–57.
21. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1): D545–d551.

Li *et al. BMC Cancer*   (2021) 21:259

Page 16 of 16

22. Murdaca G, Noberasco G, Battaglini A, Vassallo C, Giusti F, Greco M, Schiavi C, Sticchi L, Icardi G, Orsi A. Systemic Sclerosis and Vaccinations: A Register-Based Cohort Study about Seasonal Influenza and *Streptococcus pneumoniae* Vaccination Rate and Uptake from Liguria Regional Center, Northwest Italy. Vaccines (Basel). 2020;8(2):204.

23. David B-B, de Tena JG, Castillo A G-d. Reviews AJA: Novel risk factors related to cancer in scleroderma; 2017.

24. Szekanecz É, Szamosi S, Horváth Á, Németh Á, Juhász B, Szántó J, Szücs G, Szekanecz Z. Malignancies associated with systemic sclerosis. Autoimmun Rev. 2012;11(12):852–5.

25. Matsushita T, Takehara K. An update on biomarker discovery and use in systemic sclerosis. Expert Rev Mol Diagn. 2017;17(9):823–33.

26. Sharma SV, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer. 2007;7(3):169–81.

27. Bu S, Wang R, Pan Y, Yu S, Shen X, Li Y, Sun Y, Chen H. Clinicopathologic characteristics of patients with HER2 insertions in non-small cell lung Cancer. Ann Surg Oncol. 2017;24(1):291–7.

28. Bhat M, Mara K, Dierkhising R, Watt KDS. Immunosuppression, Race, and Donor-Related Risk Factors Affect De novo Cancer Incidence Across Solid Organ Transplant Recipients. Mayo Clin Proc. 2018;93:1236.

29. Chen M, Liu X, Xu Y, Zhou Q, Wang M. Clinicopathological characteristics of lung cancer in patients with systemic sclerosis. Clin Respir J. 2020;14:1131.

30. Saijo A, Hanibuchi M, Goto H, Toyoda Y, Tezuka T, Nishioka Y. An analysis of the clinical features of lung cancer in patients with connective tissue diseases. Respir Investig. 2017;55(2):153–60.

31. Li Z, Zhou Y, Zhang L, Jia K, Wang S, Wang M, Li N, Yu Y, Cao X, Hou JJO: microRNA-199a-3p inhibits hepatic apoptosis and hepatocarcinogenesis by targeting PDCD4. Oncogenesis. 2020;9(10):95.

32. Li L, He L, Zhao JL, Xiao J, Liu M, Li X, Tang H. MiR-17-5p Up-Regulates YES1 to Modulate the Cell Cycle Progression and Apoptosis in Ovarian Cancer Cell Lines. J Cell Biochem. 2015;116(6):1050–9.

33. Warren HS, Jones AL, Freeman C, Bettadapura J, Parish CR. Evidence That the Cellular Ligand for the Human NK Cell Activation Receptor NKp30 Is Not a Heparan Sulfate Glycosaminoglycan. J Immunol. 2005;175(1):207.

34. Vujanovic L, Stahl EC, Pardee AD, Geller DA, Butterfield LH. Tumor-derived α-fetoprotein directly drives human natural killer cell activation and subsequent cell death. Cancer Immunol Res. 2017;5(6):493.

35. Ouyang M, Li Y, Ye S, Ma J, Lu L, Lv W, Chang G, Li X, Li Q, Wang S, et al. MicroRNA profiling implies new markers of chemoresistance of triple-negative breast cancer. PLoS One. 2014;9(5):e96228.

36. Tsai HP, Huang SF, Li CF, Chien HT, Chen SC. Differential microRNA expression in breast cancer with different onset age. PLoS One. 2018;13(1):e0191195.

37. Sheedy FJ, Palsson-McDermott E, Hennessy EJ, Martin C, O'Leary JJ, Ruan Q, Johnson DS, Chen Y, O'Neill LA. Negative regulation of TLR4 via targeting of the proinflammatory tumor suppressor PDCD4 by the microRNA miR-21. Nat Immunol. 2010;11(2):141–7.

38. Nishimoto N, Watanabe M, Watanabe S, Sugimoto N, Fujita M. Heterocomplex formation by Arp4 and β-actin is involved in the integrity of the Brg1 chromatin remodeling complex. J Cell Sci. 2012;125(Pt 16):3870–82.

39. Shepard A, Clark AF, Jacobson N. Use of cathepsin k inhibitors for the treatment of glaucoma; 2006.

40. Alsford S, Horn D. Cell-cycle-regulated control of VSG expression site silencing by histones and histone chaperones ASF1A and CAF-1b in Trypanosoma brucei. Nucleic Acids Res. 2012;40(20):10150–60.

41. Shiao YM, Chang YH, Liu YM, Li JC, Su JS, Liu KJ, Liu YF, Lin MW, Tsai SF. Dysregulation of GIMAP genes in non-small cell lung cancer. Lung Cancer. 2008;62(3):287–94.

42. Usman M, Ilyas A, Hashim Z, Zarina S. Identification of GIMAP7 and Rabl3 as putative biomarkers for Oral squamous cell carcinoma through comparative proteomic approach. Pathol Oncol Res. 2020;26(3):1817–22.

43. Saunders A, Webb LMC, Janas ML, Hutchings A, Butcher GW. Putative GTPase GIMAP1 is critical for the development of mature B and T lymphocytes. Blood. 2010;115(16):3249–57.

44. Foussat A, Coulomb-L'Hermine A, Gosling J, Krzysiek R, Emilie D. Fractalkine receptor expression by T lymphocyte subpopulations and in vivo production of fractalkine in human. Eur J Immunol. 2015;30(1):87–97.

45. Birgitte J, Vennervald DW. Diseases DJCOil: morbidity in schistosomiasis: an update; 2004.

46. Pearce EJ, Macdonald AS. The immunobiology of schistosomiasis. Nat Rev Immunol 2:499–511. Nat Rev Immunol. 2002;2(7):499–511.

47. Nishimura M, Kuboi Y, Muramoto K, Kawano T, Imai T. Chemokines as Novel Therapeutic Targets for Inflammatory Bowel Disease. Ann N Y Acad Sci. 2010;1173:350–6.

48. Li F, Huang Q, Luster TA, Hu H, Wong KK. In Vivo Epigenetic CRISPR Screen Identifies Asf1a as an Immunotherapeutic Target in Kras-Mutant Lung Adenocarcinoma. Cancer Discov. 2019;10(2):CD-19-0780.

49. Wang C, Chang JF, Yan H, Wang DL, Sun FL. A conserved RAD6-MDM2 ubiquitin ligase machinery targets histone chaperone ASF1A in tumorigenesis. Oncotarget. 2015;6(30):29599–613.

50. Adi D, Lu XY, Fu ZY, Wei J, Luo J. IDOL G51S Variant Is Associated With High Blood Cholesterol and Increases Low-Density Lipoprotein Receptor Degradation. Arterioscler Thromb Vasc Biol. 2019;39(12):2468–79.

51. Li Y, Knisely JM, Lu W, Mccormick LM, Wang J, Henkin J, Schwartz AL, Bu G. Low density lipoprotein (LDL) receptor-related protein 1B impairs urokinase receptor regeneration on the cell surface and inhibits cell migration. J Biol Chem. 2002;277(44):42366–71.

52. Beer AG, Zenzmaier C, Schreinlechner M, Haas J, Dietrich MF, Herz J, Marschang P. Expression of a recombinant full-length LRP1B receptor in human non-small cell lung cancer cells confirms the postulated growth-suppressing function of this large LDL receptor family member. Oncotarget. 2016;7(42):68721–33.

53. Sencio V, Barthelemy A, Tavares LP, Machado MG, Soulard D, Cuinat C, Queiroz-Junior CM, Noordine ML, Salomé-Desnoulez S, Deryuter L. Gut Dysbiosis during Influenza Contributes to Pulmonary Pneumococcal Superinfection through Altered Short-Chain Fatty Acid Production. Cell Rep. 2020;30(9):2934–2947.e2936.

Publisher's Note