

# SCIENTIFIC REPORTS



OPEN

## Comparative Network Analysis of Patients with Non-Small Cell Lung Cancer and Smokers for Representing Potential Therapeutic Targets

Mehrdad Pazhouhandeh<sup>1</sup>, Fatemeh Samiee<sup>2</sup>, Tahereh Boniadi<sup>2</sup>, Abbas Fadaei Khedmat<sup>3</sup>, Ensieh Vahedi<sup>4</sup>, Mahsa Mirdamadi<sup>5</sup>, Naseh Sigari<sup>6</sup>, Seyed Davar Siadat<sup>1,7</sup>, Farzam Vaziri<sup>1,7</sup>, Abolfazl Fateh<sup>1,7</sup>, Faezeh Ajourloo<sup>8</sup>, Elham Tafsiri<sup>9</sup>, Mostafa Ghanei<sup>4</sup>, Fereidoun Mahboudi<sup>10</sup> & Fatemeh Rahimi Jamnani<sup>1,7</sup>

Cigarette smoking is the leading cause of lung cancer worldwide. In this study, we evaluated the serum autoantibody (AAb) repertoires of non-small cell lung cancer (NSCLC) patients and smokers (SM), leading to the identification of overactivated pathways and hubs involved in the pathogenesis of NSCLC. Surface- and solution-phase biopanning were performed on immunoglobulin G purified from the sera of NSCLC and SM groups. In total, 20 NSCLC- and 12 SM-specific peptides were detected, which were used to generate NSCLC and SM protein datasets. NSCLC- and SM-related proteins were visualized using STRING and Gephi, and their modules were analyzed using Enrichr. By integrating the overrepresented pathways such as pathways in cancer, epithelial growth factor receptor, c-Met, interleukin-4 (IL-4) and IL-6 signaling pathways, along with a set of proteins (e.g. phospholipase D (PLD), IL-4 receptor, IL-17 receptor, laminins, collagens, and mucins) into the PLD pathway and inflammatory cytokines network as the most critical events in both groups, two super networks were made to elucidate new aspects of NSCLC pathogenesis and to determine the influence of cigarette smoking on tumour formation. Taken together, assessment of the AAb repertoires using a systems biology approach can delineate the hidden events involved in various disorders.

Lung cancer is recognized as a fatal malignancy, accounting for more than a quarter of all cancer deaths in the United States and 1.6 million deaths every year worldwide<sup>1,2</sup>. Approximately 85% of all new cases of lung cancer are classified as non-small cell lung cancer (NSCLC), including adenocarcinoma, squamous cell carcinoma and large-cell carcinoma<sup>2</sup>. Despite an overall increase in the rate of survival for most cancers, 5-year survival for lung cancer has not improved significantly<sup>1</sup>. On the other hand, smoking is a primary risk factor for lung cancer, and nearly 14–25% of lifetime smokers (SM) are expected to develop lung cancer<sup>2</sup>. Accordingly, the carcinogenic effects of tobacco smoke through deregulation of various cellular pathways and DNA adduct formation have been extensively studied in the past decades<sup>3</sup>.

<sup>1</sup>Human Antibody Lab, Innovation Center, Pasteur Institute of Iran, Tehran, Iran. <sup>2</sup>Department of Microbial Biotechnology, Islamic Azad University, Pharmaceutical Sciences Branch, Tehran, Iran. <sup>3</sup>Department of Pulmonology, Shahid Beheshti University of Medical Sciences, Tehran, Iran. <sup>4</sup>Chemical Injuries Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran. <sup>5</sup>Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran. <sup>6</sup>Internal Medicine Department, Medical Faculty, Kurdistan University of Medical Sciences, Sanandaj, Iran. <sup>7</sup>Microbiology Research Center, Department of Mycobacteriology and Pulmonary Research Pasteur Institute of Iran, Tehran, Iran. <sup>8</sup>Department of Biology, Faculty of Science, Islamic Azad University, East Tehran Branch, Tehran, Iran. <sup>9</sup>Molecular Medicine Department, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran. <sup>10</sup>Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran. Correspondence and requests for materials should be addressed to M.G. (email: [m.ghanei@bmsu.ac.ir](mailto:m.ghanei@bmsu.ac.ir)) or F.R. (email: [Rahimi@pasteur.ac.ir](mailto:Rahimi@pasteur.ac.ir))

The low success rate of lung cancer treatment originates from delayed diagnosis, which reduces the chance of total tumour excision via surgery and acquired resistance to available medication regimens<sup>4,5</sup>. Imaging methods are usually applied in the diagnosis of lung cancer. However, the high costs of these methods, radiation exposure, and false positive or negative results, restrict their applications as screening tests<sup>2,4</sup>. Moreover, resistance to conventional chemotherapeutics, such as platinum agents, antitubulars, topoisomerase II inhibitors, and antimetabolites, is quite common among NSCLC patients who eventually develop resistance to chemotherapeutics<sup>5</sup>.

Autoantibody (AAb) production, as a reaction of the immune system to tumour antigens, is an inevitable phenomenon in cancers<sup>6</sup>. Since AAb production occurs during the early stages of disease, it may serve as a functional tool for NSCLC diagnosis<sup>4,7</sup>. Therefore, detection of AAbs in NSCLC patients has been the subject of a few investigations to determine the AAb repertoires and identify AAbs against particular proteins<sup>4,7-14</sup>. Nevertheless, for comparison of the AAb repertoires and identification of common autoantigens, limited research has been performed on the sera of NSCLC patients and SM.

In contrast to studies assessing lung cancer via bioinformatics tools<sup>15-19</sup>, a few studies have focused on smoking with a systems biology approach; for instance, expression-based network analysis and assessment of protein-protein interactions have revealed changes in human blood<sup>20</sup>, human buccal mucosa<sup>21</sup>, and mouse lung samples<sup>22</sup> due to cigarette exposure. For exhibiting the direct correlation between smoking and lung cancer, Ying Liu *et al.* analysed three protein expression datasets of lung adenocarcinoma and divided them into two subcategories of SM and non-SM. The results revealed Ras signalling pathway (SP) and proteoglycans in cancer as the key pathways in lung cancer development in SM<sup>23</sup>. Furthermore, Wang *et al.* constructed a graphical model to assess gross network alterations during transition from normal to cancer tissue in SM. The findings demonstrated a gain of epithelial growth factor receptor (EGFR) and platelet-derived growth factor receptor A (PDGFRA) modules in addition to a loss of interleukin-6 (IL-6) module in lung cancer<sup>24</sup>.

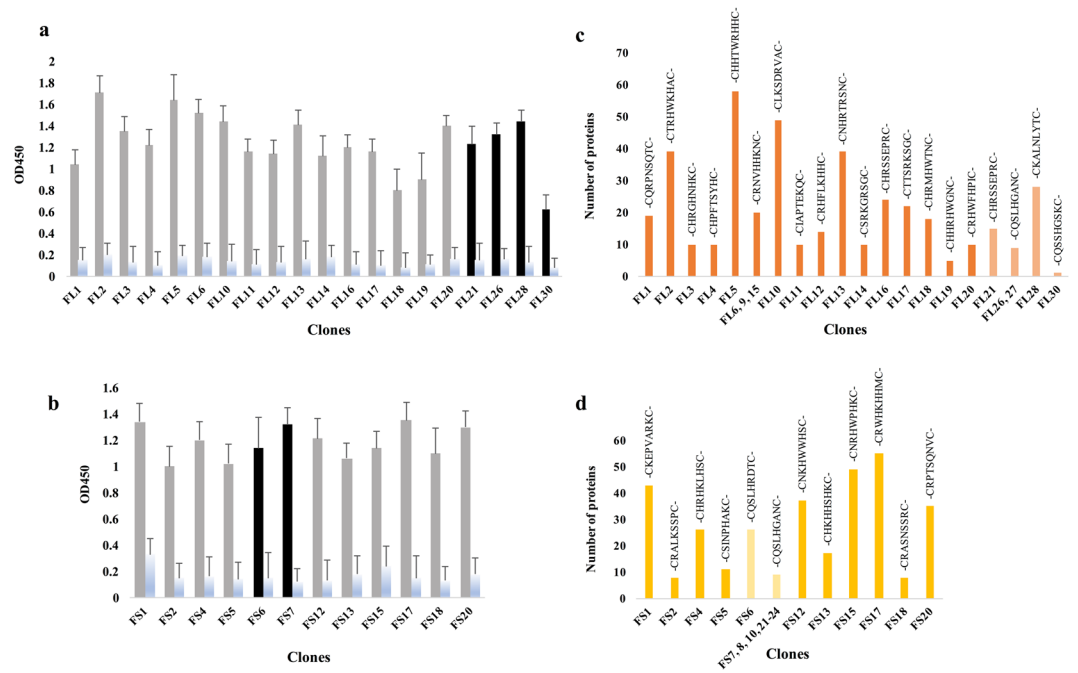
In the present study, we carried out biopanning with a phage-displayed random peptide library on immunoglobulin G (IgG) purified from the sera of NSCLC patients and SM to detect peptides based on their AAbs. NSCLC- and SM-specific proteins, predicted from the detected peptides, were investigated with systems biology tools to demonstrate their involvement as hubs in significant pathways driving lung cancer pathogenesis. Furthermore, they were used to construct two super networks for clarifying multidimensional contribution of smoking to NSCLC.

## Results

**Peptide library enrichment led to the generation of NSCLC and SM protein datasets.** To assess the binding of NSCLC and SM IgG to phage pools after three consecutive rounds of surface- (SU) and solution-phase (SL) biopanning, all the input and output phages were screened via polyclonal phage enzyme-linked immunosorbent assay (ELISA). The highest signals were observed in the third round of SU and SL biopanning (data not shown). Based on monoclonal phage ELISA on random clones obtained from the third round of biopanning, 30 NSCLC (SU, FL<sub>U</sub>1-20; SL, FL<sub>L</sub>21-30) and 15 SM (SU, FS<sub>U</sub>1-5 and FS<sub>U</sub>11-20; SL, FS<sub>L</sub>6-10 and FS<sub>L</sub>21-25) clones were selected, which showed the highest intensities in comparison to the controls (Fig. 1a and b). Among the designated phage clones, three FL<sub>L</sub> and five FS<sub>U</sub> clones failed to produce meaningful sequences. In addition, two NSCLC-specific peptides (FL<sub>U</sub>7 and FL<sub>U</sub>8) containing sequences with three consecutive histidine residues, and one SM-specific peptide (FS<sub>L</sub>25) due to its similarity to another peptide sequence in this group, were omitted. Notably, among NSCLC-specific clones, FL<sub>U</sub>6, FL<sub>U</sub>9, and FL<sub>U</sub>15 (-CRNVVHHKNC-), FL<sub>L</sub>22 and FL<sub>L</sub>29 (-CTRHWPVHC-), in addition to FL<sub>L</sub>26 and FL<sub>L</sub>27 (-CQSLHGANC-), showed similar sequences (Fig. 1c). Additionally, clones FS<sub>L</sub>7, FS<sub>L</sub>8, FS<sub>L</sub>10, and FS<sub>L</sub>21-24 showed an identical amino acid composition (-CQSLHGANC-) to clones FL<sub>L</sub>26 and FL<sub>L</sub>27 (Fig. 1d).

Evaluation of the selected peptides in the MIMOdb demonstrated that 21 NSCLC- and 13 SM-specific peptides, were true target binders. Blasting peptides FL<sub>U</sub>22, FL<sub>U</sub>29, and FS<sub>L</sub>9 (score  $\geq 18.5$ ) followed by an assessment in the UniProt database and review of the literature, did not present major proteins; accordingly, these peptides were deleted from the final peptide lists. However, blasting the remaining peptide sequences led to the prediction of 531 NSCLC and 346 SM proteins with 50 common proteins (Supplementary dataset). Figure 1c and d show the sequence and quantity of proteins predicted from each peptide. Furthermore, the list of proteins found in more than one clone is presented in Table S1. Peptides FL<sub>U</sub>19, FL<sub>L</sub>26, FL<sub>L</sub>27, and FL<sub>L</sub>30 accounted for the lowest number of NSCLC-related proteins (<10), while peptides FL<sub>U</sub>5 and FL<sub>U</sub>10 represented the highest number of NSCLC-related proteins (>40). Moreover, peptides FS<sub>U</sub>1, FS<sub>U</sub>15 and FS<sub>U</sub>17 were individually related to more than 40 proteins whereas a group of SM-specific peptides (FS<sub>U</sub>2 and FS<sub>U</sub>18, and FS<sub>L</sub>7, FS<sub>L</sub>8, FS<sub>L</sub>10, and FS<sub>L</sub>21-24) accounted for less than 10 proteins in the SM protein dataset. Markedly, retinoic acid-induced protein 1 (RAI1), dermatan-sulfate epimerase (DSE), and protein kinase C delta (PRKCD) were proteins predicted from more than one peptide in both NSCLC and SM groups.

**Protein complexes in the NSCLC and SM groups.** Gene set enrichment by Consensus PathDB (CPDB) revealed multiple physically associated protein complexes, with at least two members in the NSCLC or SM protein dataset. Proline-rich tyrosine kinase 2 (PYK2)/Src homology region 2-containing protein tyrosine phosphatase 2 (SHP2) complex (two out of two components; 2/2), ADP-ribosylation factor 6 (ARF6)/Guanosine triphosphate (GTP)/Nucleoside diphosphate kinase A (NME1)/T-lymphoma invasion and metastasis-inducing protein 1 (Tiam1) complex (2/3), proto-oncogene tyrosine-protein kinase Src (SRC)-PRKCD-CUB domain-containing protein 1 (CDCP1) complex (2/3), and Janus kinase 1 (JAK1): Interleukin-13 receptor subunit alpha-1 (IL-13RA-1): Interleukin-4 receptor alpha (IL-4RA): Tyrosine kinase 2 (TYK2) complex (2/4) were among the most significant complexes in the NSCLC group (*p*-value < 0.05). APC-stimulated guanine nucleotide exchange factor 2 (ASEF2)/Adenomatous polyposis coli (APC) complex (2/2), Early elongation complex with hyperphosphorylated RNA polymerase II C-terminal domain (Pol II CTD) (4/4), Positive transcription elongation factor



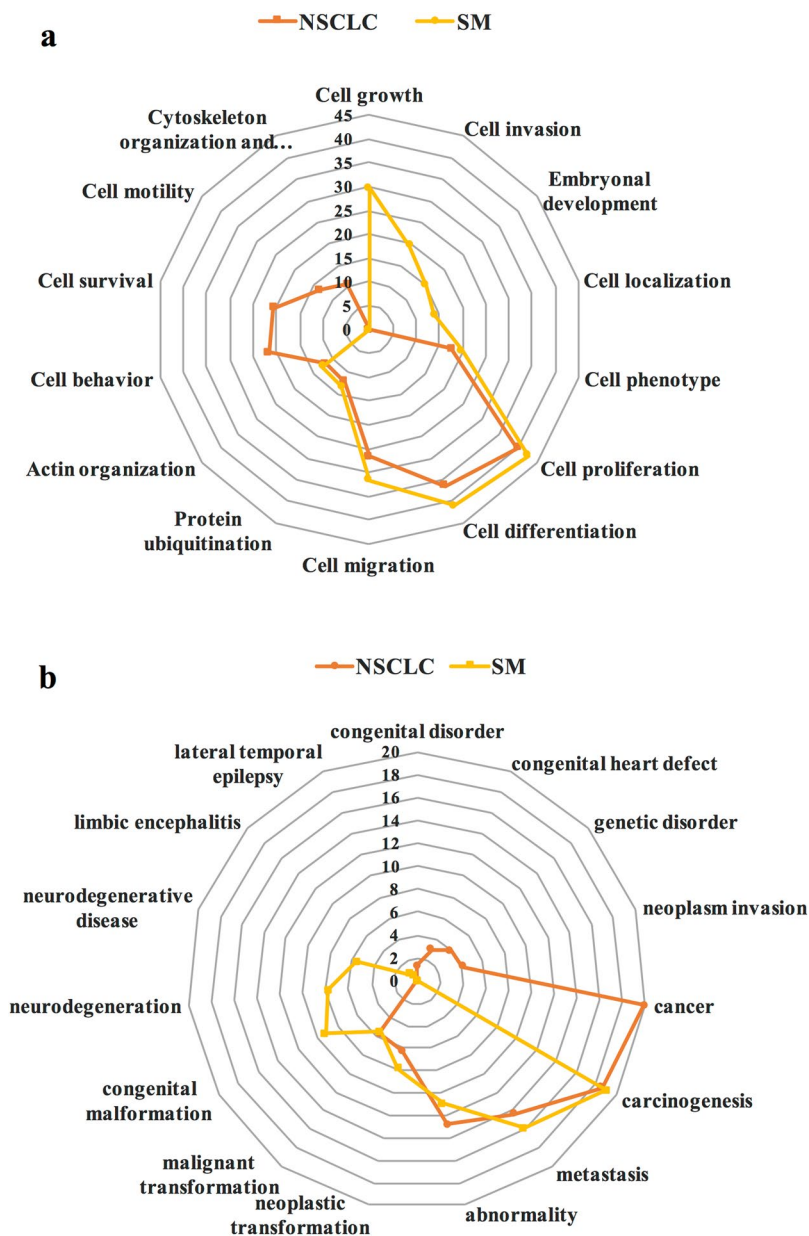
**Figure 1.** Monoclonal phage ELISA and number of proteins predicted from each selected peptide. (a) Randomly selected NSCLC and (b) SM clones were screened by monoclonal phage ELISA; intensities are reported as mean  $\pm$  SD. Columns 1–10 (grey) and 11–20 (black) belong to clones from SU and SL biopanning with the highest intensities, respectively, and the blue columns indicate the controls (bovine serum albumin (BSA)). Peptides identified with the purified IgG of (c) NSCLC patients (orange) and (d) SM (yellow) are plotted against the number of human proteins which matched to each sequence after blasting with a cut-off of 18.5 and manual deletion of unrelated entities. Mild orange (c) and (d) mild yellow indicate the peptides obtained from SL biopanning.

(P-TEFb) complex (2/3), and Tankyrin 1-tankyrin 2-telomeric repeat binding factor (TRF1) complex (2/3) were the most significant complexes in the SM group ( $p$ -value  $< 0.05$ ). Histone deacetylase 2 (HDAC2) and lysine-specific demethylase 1 (LSD1) complexes were common in the datasets. Table S2 shows the overlapped components and  $p$ -value of each complex.

**Primary classification of all proteins via Pathway Studio®.** Network analysis of NSCLC and SM protein datasets using Pathway Studio® presented a similar pattern in some cell processes, including proliferation (40%,  $p$ -value:  $1.7E-13$ ; and 42%,  $p$ -value:  $7.1E-15$ , respectively), differentiation (36%,  $p$ -value:  $3.7E-23$ ; and 41%,  $p$ -value:  $2.9E-26$ , respectively), and migration (27%,  $p$ -value:  $5.0E-19$ ; and 32%,  $p$ -value:  $2.9E-23$ , respectively) as the top three processes associated with the highest proteins count (Fig. 2a). Although cell growth seemed to be an important phenomenon in the SM group, lack of cell survival arm was witnessed.

Diseases analysis demonstrated cancer (20%,  $p$ -value:  $3.1E-6$ ), carcinogenesis (19%,  $p$ -value:  $2.1E-19$ ), metastasis (14%,  $p$ -value:  $3.2E-10$ ), abnormality (13%,  $p$ -value:  $5.6E-10$ ), neoplastic transformation (6%,  $p$ -value:  $6.8E-7$ ), malignant transformation (6%,  $p$ -value:  $4.3E-7$ ), and neoplasm invasion (4%,  $p$ -value:  $4.1E-5$ ), as the most important diseases linked to the NSCLC group (Fig. 2b). Moreover, carcinogenesis ( $p$ -value:  $1.9E-14$ ) was a major event, involving 19% of SM proteins. Other events in this group included metastasis (16%,  $p$ -value:  $7.3E-10$ ), abnormality (11%,  $p$ -value:  $7.0E-5$ ), congenital malformation (9%,  $p$ -value:  $1.8E-8$ ), neoplastic transformation (8%,  $p$ -value:  $5.0E-8$ ), neurodegeneration (8%,  $p$ -value:  $1.5E-5$ ), malignant transformation (5%,  $p$ -value:  $5.9E-5$ ), and neurodegenerative disease (5%,  $p$ -value:  $6.3E-5$ ). Although carcinogenesis, metastasis, and abnormality owned similar shares in both groups, cancer state and neurodegenerative disease were major disorders which were specifically associated with the NSCLC and SM groups, respectively. The percentage indicates the ratio of involved proteins to the total number of proteins in each dataset.

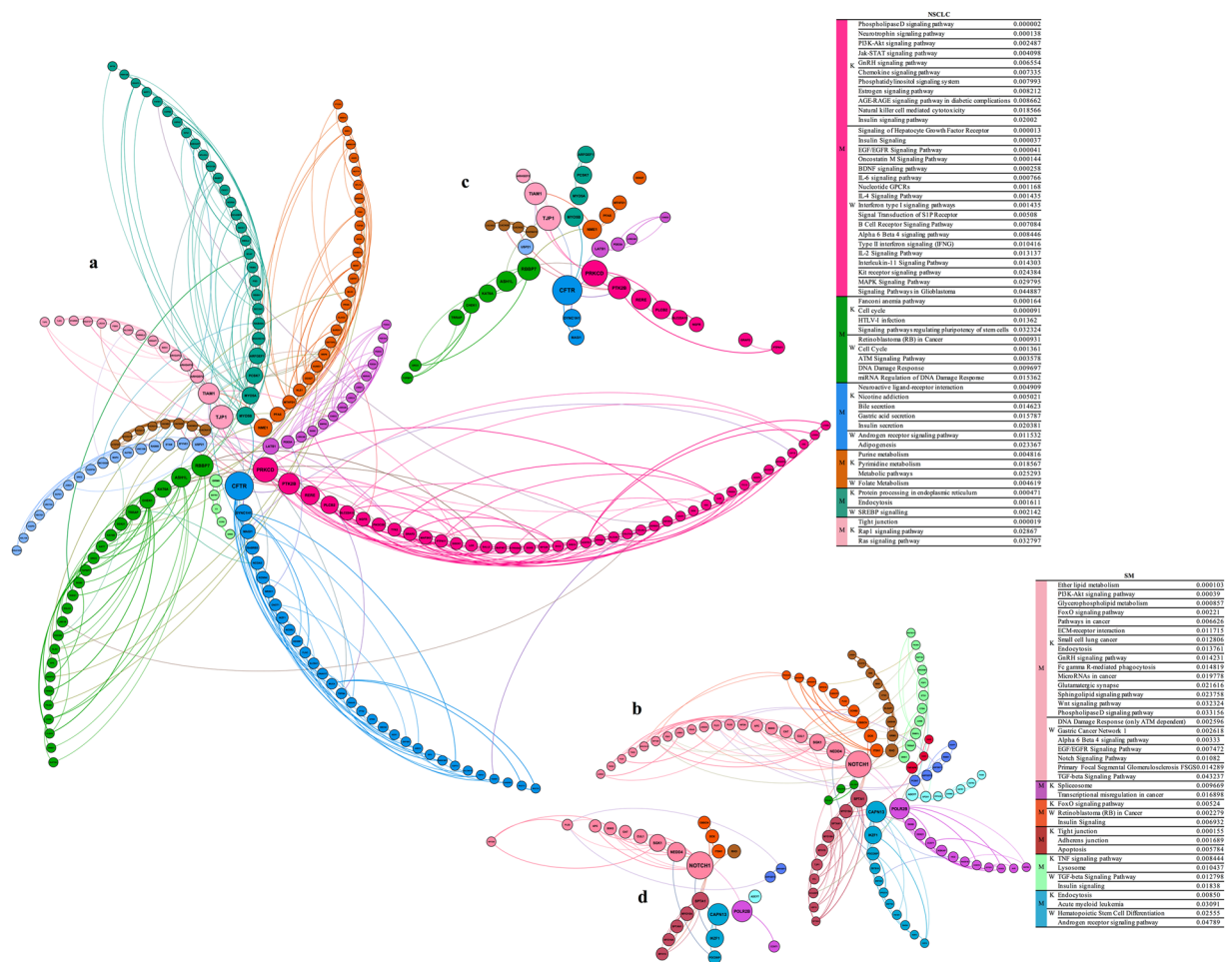
**Gene Ontology (GO) and pathway enrichment analysis of NSCLC and SM protein datasets.** The NSCLC network constructed via Gephi contained 316 nodes and 513 edges, while the SM network consisted of 179 nodes and 255 edges (Fig. 3a). GO enrichment of the top six NSCLC- and SM-related modules by Metascape exhibited terms such as transmembrane receptor protein tyrosine kinase SP (GO:0007169), phospholipid metabolic process (GO:0006644), intrinsic apoptotic SP in response to oxidative stress (GO:0008631), collagen trimer (GO:0005581), DNA metabolic process (GO:0006259), signal transduction by p53 class mediator (GO:0072331), chromatin remodelling (GO:0006338), integrin-mediated SP (GO:0007229), cellular response to organic cyclic compound (GO:0071407), O-glycan processing (GO:0016266), regulation of GTPase activity (GO:0043087), phospholipid binding (GO:0005543), and apical junction complex (GO:0043296) in the NSCLC



**Figure 2.** Top NSCLC- and SM-related cell processes and diseases based on Pathway Studio®. **(a)** The most significant cell processes. **(b)** The most significant diseases. Significant cell processes and diseases associated with the NSCLC (orange) and SM (yellow) groups were selected based on the network analysis function in Pathway Studio®. They are reported as the percentage of the number of proteins involved in the cell process or disease, to the total number of proteins in the related dataset (NSCLC or SM;  $100 \times \frac{\text{involved proteins}}{\text{total protein in dataset}}$ ).

group, as well as regulation of cell migration (GO:0030334), inositol lipid-mediated signalling (GO:0048017), regulation of canonical Wnt SP (GO:0060828), DNA repair (GO:0006281), mitotic cell cycle process (GO:1903047), regulation of small GTPase-mediated signal transduction (GO:0051056), kinase activity (GO:0016301), calmodulin binding (GO:0005516), and toll-like receptor SP (GO:0002224) in the SM group. Furthermore, there were some terms related to cardiovascular diseases and diabetes which are highly prevalent disorders among Iranian population (as also confirmed among the participants of this study)<sup>25,26</sup>.

Expectedly, pathways enrichment analysis based on the KEGG and WikiPathways databases in Enrichr presented a group of important pathways in the top six modules of NSCLC, including phospholipase D (PLD) SP, phosphoinositide 3-kinase-Akt (PI3K-Akt) SP, JAK-signal transducer and activator of transcription (JAK-STAT) SP, gonadotropin-releasing hormone (GnRH) SP, chemokine SP, natural killer cell (NKC) mediated cytotoxicity, pathways in cancer, signalling of hepatocyte growth factor receptor (HGFR), EGF/EGFR SP, IL-4 SP, IL-6 SP, type II interferon signalling (IFN $\gamma$ ), mitogen-activated protein kinase (MAPK) SP, signalling pathways regulating pluripotency of stem cells, ataxia telangiectasia mutated (ATM) SP, Fanconi anaemia pathway, nicotine addiction, gamma-aminobutyric acid (GABAergic) synapse, and Ras SP (Fig. 3a). Similarly, critical pathways were found



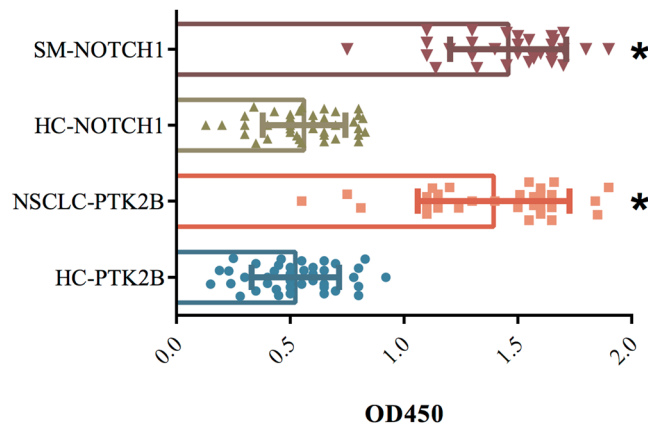
**Figure 3.** Network visualization, modularity and pathway analysis of NSCLC and SM. The node size demonstrates the betweenness centrality and the edge thickness shows the combined score of STRING. Pathway enrichment analysis of the top six modules (M) of (a) NSCLC and (b) SM was performed using Enrichr (based on the KEGG and WikiPathways databases). (c) NSCLC- and (d) SM-specific hubs are also displayed, respectively.

in the SM group, such as PI3K-Akt SP, pathways in cancer, extracellular matrix (ECM)-receptor interaction, Fc gamma R-mediated phagocytosis, microRNAs in cancer, glutamatergic synapse, Wnt SP, PLD SP, DNA damage response (DDR) only ATM dependent, alpha 6 Beta 4 SP, EGF/EGFR SP, neurogenic locus notch homolog protein (NOTCH) SP, transforming growth factor beta (TGF- $\beta$ ) SP, transcriptional misregulation in cancer, Forkhead box O (FoxO) SP, tight junction, and tumour necrosis factor (TNF) SP (Fig. 3b). Table S3 summarizes the information on some significant pathways identified via data mining.

**CFTR, PRKCD, PTK2B, CHEK1, and NOTCH1 as functional hubs.** According to features such as betweenness centrality and degree implemented in Gephi, several NSCLC- and SM-specific hubs were identified (Table S4a). Proteins such as cystic fibrosis transmembrane conductance regulator (CFTR), PRKCD, tight junction protein ZO-1 (TJP1), PTK2B, TIAM1, nucleoside diphosphate kinase A (NME1), phospholipase C beta 2 (PLCB2), brefeldin A-inhibited guanine nucleotide-exchange protein 1 (ARFGEF1), and checkpoint kinase 1 (CHEK1) were NSCLC-specific hubs. Furthermore, NOTCH1, calpain-13 (CAPN13), DNA-directed RNA polymerase II subunit RPB2 (POLR2B), E3 ubiquitin-protein ligase NEDD4 (NEDD4), serine/threonine-protein kinase Sgk1 (SGK1), serine/threonine-protein kinase Sgk3 (SGK3), APC, PLD2, ARFGEF1, and regulatory-associated protein of mTOR (RPTOR) were highly ranked proteins in the SM group.

Assessment of NSCLC-specific hubs in the WebGestalt database demonstrated disorders such as lung neoplasms ( $p$ -value: 0.014), epithelial cancers ( $p$ -value: 0.014), lung diseases ( $p$ -value: 0.015), and carcinoma ( $p$ -value: 0.025), as well as several malignancies and disorders related to cancer. Evaluation of SM-specific hubs revealed confirmative results such as precancerous conditions ( $p$ -value: 0.003), substance-related disorders ( $p$ -value: 0.008), chromosome aberrations ( $p$ -value: 0.016), and stress ( $p$ -value: 0.021), in addition to adenocarcinoma and neoplasms.

Potential drugs, which could modulate NSCLC hubs, were determined by searching DrugBank and Reaxys<sup>®</sup> (Table S4b). Among NSCLC-specific hubs, a group of drugs was found to target PRKCD, PTK2B, and CHEK1, as indicated in several types of malignancies. In this regard, drugs such as bosutinib, dasatinib, axatinib, and



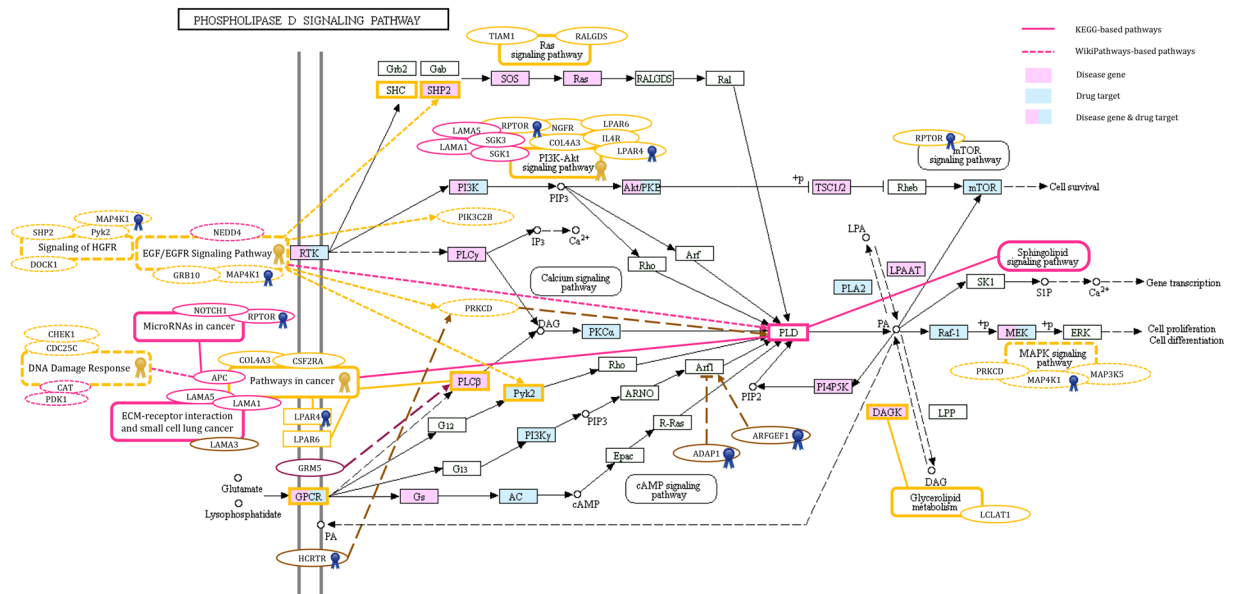
**Figure 4.** Binding assessment of the selected hubs to the NSCLC and SM sera. The results of PTK2B and NOTCH1 binding to the sera of 30 NSCLC patients (peach), 30 SM (red), and 40 age-matched healthy subjects (HC, as the controls) are presented (\* $p$ -value < 0.01).

lapatinib have been developed to concurrently inhibit these three kinases. Astonishingly, PTK2B inhibitors, crizotinib and mereletinib, are prescribed for patients with metastatic NSCLC showing specific molecular characteristics<sup>27,28</sup>.

**Ability of the selected hubs for discrimination of NSCLC, SM, and control sera.** Based on features such as betweenness centrality and degree, PTK2B and NOTCH1 were selected from the NSCLC and SM groups, respectively. The data exhibited specific binding of two hubs to the corresponding NSCLC and SM sera ( $p$ -value < 0.01). In contrast, they showed poor binding to the sera of age-matched healthy subjects (Fig. 4).

#### The PLD SP and inflammatory cytokines network as overriding events in the NSCLC and SM groups.

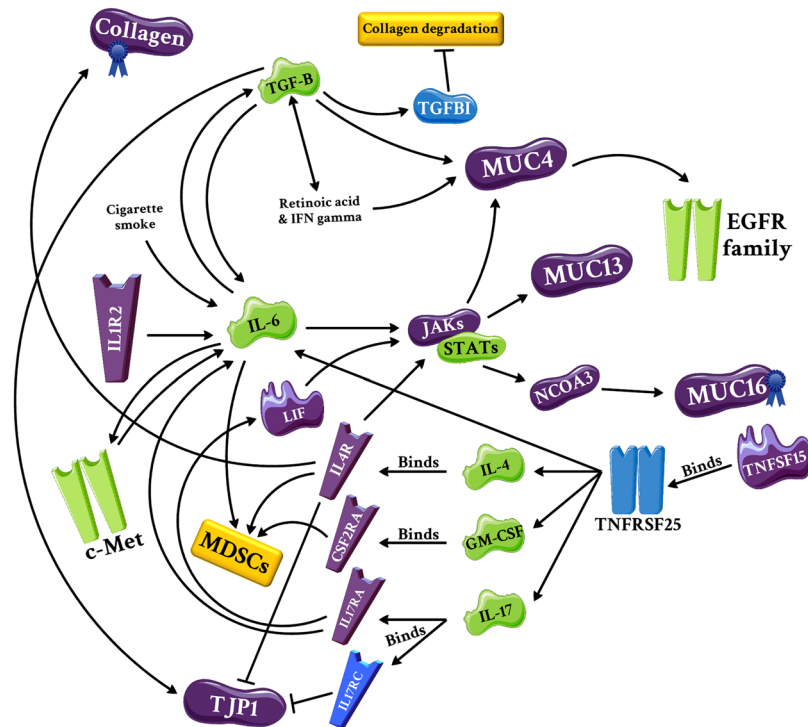
The PLD SP was undoubtedly the most significant NSCLC pathway. Remarkably, PLD was found among SM-specific proteins. Furthermore, phospholipid metabolic process and phospholipid binding were over-represented GO terms identified through NSCLC-related module analysis. According to the literature review and evaluation of pathways involved in tumour formation and development, the PLD SP took centre stage in the pathogenesis of NSCLC in this study. Curation of the PLD SP derived from the KEGG database with the information attained through data-mining and Pathway Studio<sup>®</sup> led to the formation of a super network constructed with several NSCLC (e.g. PRKCD, PTK2B, PLCB2, and ARFGF1) and SM (e.g. PLD2, ARFGF1, and RPTOR) hubs. Among the mentioned hubs, RPTOR, a SM-specific hub, was also found in the NSCLC protein dataset. As exhibited in Fig. 5, a group of NSCLC proteins such as SHC-transforming protein 4 (SHC4), SHP2, mitogen-activated protein kinase kinase kinase 1 (MAP4K1), phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta (PIK3C2B), diacylglycerol kinase theta (DGKQ), laminin subunit alpha-3 (LAMA3), arf-GAP with dual PH domain-containing protein 1 (ADAP1), ral guanine nucleotide dissociation stimulator (RALGDS), and G-protein-coupled receptors (GPCRs, including lysophosphatidic acid receptor 4 (LPAR4), lysophosphatidic acid receptor 6 (LPAR6), and orexin receptor type 1 (HCRTR1)) linked to the PLD SP, were observed in this network. Among the mentioned proteins, LPAR4, HCRTR1, and MAP4K1 were present in the SM protein dataset, as well. In addition, there were several key proteins such as NOTCH1, APC, pyruvate dehydrogenase (acetyl-transferring) kinase isozyme 1, mitochondrial (PDK1), laminin subunit alpha-1 (LAMA1), laminin subunit alpha-5 (LAMA5), and metabotropic glutamate receptor 5 (GRM5) which were only found in the SM protein dataset. Based on the KEGG and WikiPathways databases, significant SPs such as EGF/EGFR SP, HGFR SP, MAPK SP, PI3K-Akt SP, Ras SP, glycerophospholipid metabolism, glutamatergic synapse, ECM-receptor interaction, small cell lung cancer, microRNAs in cancer, DDR, and pathways in cancer identified via analysis of NSCLC and SM modules-related pathways, were also integrated in this super network. The members of this super network were linked to each other, resulting in a cascade of events. Binding of lysophosphatidic acid (LPA) to its receptors, LPAR4 and LPAR6, activates GPCR-related pathways which stimulate PLD<sup>29,30</sup>. Two other GPCRs, HCRTR1 activated by neuropeptide orexin and GRM5 activated by glutamate (released due to the indirect action of HIFs<sup>31</sup> like HIF3A observed in the SM group), promote various pathways leading to PLD activation<sup>32,33</sup>. Another remarkable receptor in this network, EGFR, stimulates PLD via multiple routes containing proteins, some of which were found in this study, as well (Fig. 5)<sup>30</sup>. Among proteins involved in pathways leading to PLD activation, ARF1 acts as a key protein activated and inactivated by ARFGF1 and ADAP1, respectively<sup>34</sup>. Several pathways such as glycerophospholipid metabolism, sphingolipid, mechanistic target of rapamycin (mTOR), and MAPK SPs, are induced through phosphatidic acid (PA) produced by the enzymatic activity of PLD<sup>30,35,36</sup>. Components of ECM (LAMA1, 3, and 5) induce the synthesis of PA by activation of PLD which stimulates matrix metalloproteinases (MMPs) and enhances the invasive capacity of tumor cells<sup>37</sup>. Among pathways connected to the PLD SP, pathways in cancer was one of the critical events observed in both NSCLC and SM groups and was linked to other significant pathways such as microRNAs in cancer, DDR, and small cell lung cancer. A set of proteins, including LPAR4, LPAR6, collagen alpha-3(IV) chain-like (COL4A3), PLCB2, and granulocyte-macrophage colony-stimulating factor receptor subunit alpha (CSF2RA) from the NSCLC protein



**Figure 5.** The super network constructed based on the PLD SP. The network was established based on the PLD SP in the KEGG database<sup>30</sup> and was expanded by the addition of some pathways and neighbours via WikiPathways and data mining in the literature. As the identified pathways are shown in this figure, the FcεRI SP has been deleted from the PLD SP. LPAR4, LPAR6, GRM5, HCRTR1, and EGFR are the known receptors in this network. Different genes involved in the PLD SP and other related pathways are presented as coloured boxes. Yellow and pink colours indicate NSCLC- and SM-specific proteins in significant pathways found in this study, respectively. Brown and plum colours correspond to proteins from the NSCLC and SM protein datasets, respectively. Additionally, blue and yellow ribbons show common proteins and pathways between the two groups, respectively.

dataset as well as APC, LAMA1, LAMA5, and PLD1 from the SM protein dataset, were involved in pathways in cancer. All NSCLC- and SM-specific proteins playing roles in the above-mentioned SPs are presented in Fig. 5.

The second super network was constructed to describe how cytokines alter the tumour microenvironment (e.g. ECM proteins) and how tumour recruits them to fabricate the immune system (Fig. 6). This network was originated from multiple significant pathways such as JAK-STAT SP, chemokine SP, NKC-mediated cytotoxicity, signalling of HGFR (c-Met), EGF/EGFR SP, IL-4 SP, IL-6 SP, and type II interferon signalling obtained from the NSCLC module analysis and ECM-receptor interaction, Fc gamma R-mediated phagocytosis, TGF- $\beta$  SP, tight junction, and TNF SP obtained from the SM module analysis. It consisted of various proteins, including TJP1 (NSCLC hub), leukemia inhibitory factor (LIF), IL-4R, IL-1 receptor type 2 (IL-1R2), IL-17 receptor A (IL-17RA), CSF2RA, TNF ligand superfamily member 15 (TNFSF15), TYK2, PIK3C2B, nuclear receptor coactivator 3 (NCOA3), mucin-4 (MUC4), MUC13, MUC16, COL4A3, collagen alpha-1(XXVI) chain (COL26A1), and collagen alpha-1(XXVII) chain (COL27A1) which were present in the NSCLC protein dataset. SM-specific proteins such as TNF receptor superfamily member 25 (TNFRSF25), IL-17 receptor C (IL-17RC), TGF-beta-induced protein ig-h3 (TGFB1), MUC16, collagen alpha-1(III) chain (COL3A1), collagen alpha-1(XVI) chain (COL16A1), and COL27A1, were also integrated into the network. A remarkable relationship was found among the predicted proteins in this network. One of the most important pathways in the NSCLC group was IL-6 SP. IL-6 production is induced by IL-1R2<sup>38</sup>, IL-17<sup>39,40</sup> and cigarette smoking<sup>41</sup>. There is a two-way interaction between IL-6 and HGF, resulting in increased invasiveness of a lung cancer cell line<sup>42</sup> and an autocrine loop between IL-6 and TGF- $\beta$ 1 which its dysregulation causes an array of inflammatory disorders<sup>43</sup>. As shown in Fig. 6, IL-6 can augment the expression of MUC4 and MUC13 through JAK/STAT SP<sup>44,45</sup>. Production of LIF, an IL-6 family cytokine, by synovial fibroblasts is also induced by IL-17<sup>39</sup>. Another interleukin, IL-4, can increase the transcription of MUC4 (via JAK/STAT) and MUC16 (through JAK/STAT and NCOA3)<sup>44,46,47</sup>. Furthermore, IL-4 triggers M2 macrophages to produce much more collagen a3 (VI)<sup>48</sup>. Expression of TJP1 is also decreased by IL-4, and its distribution at the interface of adjacent cells is disrupted by IL-17 receptor C (IL-17RC)<sup>49,50</sup>. The major cytokine in this network, TGF- $\beta$ , contributes to the modification of tumour microenvironment via MUC4 upregulation in an interdependent manner with retinoic acid and interferon type II (IFN $\gamma$ )<sup>51</sup>. TGF- $\beta$  increases TJP1 expression<sup>52</sup> and inhibits collagens degradation through activation of TGFB1<sup>53</sup>. Locally-secreted cytokines not only modulate the structure of tumour microenvironment, but also disrupt the protective function of the immune system. TGF- $\beta$  signalling plays critical roles in the inhibition of tumour-specific CD8 T cell- and NK cell-mediated cytotoxicity through regulatory T cells<sup>54,55</sup>. In addition, IL-4, IL-6, and granulocyte-macrophage colony-stimulating factor (GM-CSF) increase the population of myeloid-derived suppressor cells (MDSCs) which interrupt the identification of tumour antigens by T cells through induction of nitric oxide synthase and consequent nitration of T cell receptor<sup>56</sup>. The final link is related to the TNF family which seems to be involved in tumour suppression<sup>57</sup>.



**Figure 6.** The super network constructed based on the relationship between inflammatory cytokines and tumour microenvironment. Similar to the first super network, the second network was created by proteins, particularly inflammatory cytokines, involved in significant NSCLC and SM pathways. IL-4, IL-6, IL-17, GM-CSF, and TGF- $\beta$  are the most important cytokines in this network. Purple, blue, and green colours indicate presence of protein in the NSCLC protein dataset, presence of protein in the SM dataset, and unavailability of proteins in the datasets, respectively. Blue ribbons also show the presence of protein in the SM protein dataset.

However, the binding of TNFSF15 to its receptor, TNFRSF25, amplifies the expression of IL-4, IL-6, GM-CSF, and IL-17<sup>58,59</sup>, which may explain the overactivity of these tumour-promoting cytokines in our network.

## Discussion

A combination of events such as cigarette smoking, nicotine addiction, delayed diagnosis, and resistance development during chemotherapy, has led to major concerns about lung cancer. Therefore, it can be useful to compare the AAb repertoires of NSCLC patients with SM to discover proteins involved in lung cancer pathogenesis<sup>60,61</sup>. According to the literature, AAbs exist in the plasma of patients several months to years before the emergence of any clinical signs, reflecting the functional immunoreactivity against tumour antigens<sup>62</sup>. AAbs against tumour-specific antigens generated due to antigen mutation, overexpression, or altered structure, have been recruited to discover cancer biomarkers<sup>63</sup>. Aside from the important role of AAbs in the discovery of tumour antigens for the development of T cell-based vaccines, they can act as therapeutic antibodies by either blocking tumour antigens or recruiting the immune system<sup>63</sup>. Various studies have assessed cancer-specific antibodies in different types of malignancies<sup>63</sup>. Similarly, several studies have examined NSCLC patients and have identified a panel of AAbs binding to rho-associated protein kinase 1, protein kinase C beta, complement factor H, c-myc, cyclin A, B1 and D1, cyclin-dependent kinase 2, MUC1, Dickkopf-related protein 1, p53, NY-ESO-1, and annexins (I and II)<sup>8-14</sup>. Furthermore, Yao, *et al.* introduced four combined AAb biomarkers identifying nucleolar and coiled-body phosphoprotein 1, metastasis-associated lung adenocarcinoma transcription 1, hyaluronan mediated motility receptor, and spermine oxidase for early detection of NSCLC<sup>7</sup>.

In the present study, classification of the predicted proteins into significant cellular processes revealed some overlapped areas as well as dissimilarities in the cellular behaviour of NSCLC and SM. Network enrichment analysis of NSCLC- and SM-related modules resulted in the identification of remarkable pathways which play key roles in cancer initiation, progression, or suppression. NSCLC-specific pathways included MAPK<sup>64,65</sup>, PI3K-Akt<sup>66</sup>, PLD<sup>36</sup>, Ras<sup>67</sup>, EGF/EGFR<sup>68-71</sup>, c-Met<sup>72-74</sup>, LPARs<sup>75</sup>, IL-4<sup>76</sup> and IL-6<sup>77-79</sup> SPs which play significant roles in tumorigenesis and cancer development. Remarkably, we found tumour suppressor pathways such as Fanconi anaemia pathway<sup>80</sup> and ATM SP<sup>81</sup> in the NSCLC group, which might be activated as feedbacks to the overactivated pathways in cancer cells.

Considerably, analysis of SM-related modules revealed multiple tumour-promoting pathways such as PI3K-Akt, PLD, and EGF/EGFR SPs which were commonly found in the NSCLC group. Conversely, activation of tumour suppressor pathways, including DDR and Foxo<sup>82</sup> SPs as well as negative regulation of Wnt SP<sup>83</sup> in the SM group, show activation of cellular defensive mechanisms against the carcinogenic compounds of cigarette smoke. Moreover, we found TGF- $\beta$  and TNF SPs in this group, known as “double-edged swords”, which can promote



or suppress cancers<sup>84,85</sup>. As the results demonstrate, carcinogenic events are strongly significant in SM and may accelerate the emergence of disorders such as cancer.

Among NSCLC-specific hubs, there are four proteins which can be considered as the key players involved in lung cancer development. The first hub is CFTR which its dysregulation has been reported in a group of cancers such as breast, gastric, prostate, lung, and colorectal cancers (CRC)<sup>86</sup>. Moreover, a significant correlation has been reported between CFTR downregulation and lung cancer progression and metastasis<sup>86,87</sup>. In contrast to studies reporting CFTR suppression in epithelial cells due to exposure to cigarette smoke, Li, *et al.* showed CFTR upregulation through nicotine-derived nitrosamine ketone (NNK), as one of the most toxic tobacco-specific carcinogens in cigarette smoke<sup>88</sup>. CFTR as an ABC transporter promotes protection against tobacco-specific carcinogens in cigarette smoke<sup>88</sup>. On the other hand, AAbs can lead to the development of lung cancer by blocking CFTR.

The second hub, PTK2B, is a part of protein complexes found in this study. This hub plays major roles in cancer cells due to its connections to critical pathways such as PLD, EGFR, and c-Met SPs<sup>30</sup>. It has been reported that PTK2B is upregulated in NSCLC. In fact, its higher expression and greater activity lead to the progression of NSCLC through modulating extracellular signal-regulated kinase 1/2 (ERK1/2)<sup>89</sup>. Furthermore, wound-induced activation of PTK2B stimulates EGFR activation in the epithelial cells of lung lesions<sup>90</sup>. PTK2B was also detected as a sunitinib-specific target in NSCLC patients with kirsten rat sarcoma viral oncogene (KRAS) mutations<sup>91</sup>. Likewise, prescription of crizotinib and merlotinib in NSCLC patients underscores the importance of PTK2B in the pathogenesis of lung cancer. KRAS-activated PRKCD promotes growth, invasion, and migration of cancer cells, through MAPK signalling in NSCLC cell lines (and mouse model), whereas PRKCD activation in cells without active KRAS leads to tumour suppression<sup>92</sup>. In this regard, the PRKCD inhibitor, rottlerin, can effectively potentiate chemotherapy-induced apoptosis, particularly if it is used concomitantly with trastuzumab<sup>93</sup>. Liu, *et al.* found that CHEK1 overexpression is associated with poor overall survival in NSCLC patients, and inhibition of CHEK1 by miR-195 suppresses tumour cell growth, migration, and invasion<sup>94</sup>. In this regard, Xiaojie, *et al.* revealed that miR-195 synergizes with microtubule-targeting agents (eribulin and paclitaxel) via CHEK1 regulation, leading to the inhibition of NSCLC cells growth<sup>95</sup>. In agreement with other studies, we confirmed the involvement of CFTR, PTK2B, PRKCD, and CHEK1 in NSCLC pathogenesis.

In terms of major cancer-related proteins among SM-specific hubs, NOTCH1, ARFGEF1, PLD2, and RPTOR were strongly aligned with hallmarks of NSCLC. NOTCH1 contributes to cell proliferation, invasion, and chemoresistance. Its involvement in tumor progression and prognosis of NSCLC has been demonstrated in different studies<sup>96,97</sup>. Mutations and upregulation of *Notch1* have been reported in 10% and 30% of NSCLC patients, respectively<sup>97,98</sup>. Konishi, *et al.* showed that suppression of NOTCH signaling inhibits the growth of NSCLC<sup>99</sup>. Furthermore, NOTCH signaling can crosstalk with other SPs such as Wnt and TGF- $\beta$  driving cancer development<sup>98</sup>. Therefore, targeting NOTCH signaling by therapeutic agents like antibodies which can inhibit delta-like protein 4 (Dll4) binding to NOTCH1, may be advantageous for NSCLC patients<sup>96,98</sup>. In this regard, a humanized anti-Dll4 antibody, demcizumab, is currently under evaluation in three Phase 1b studies on pancreatic cancer, CRC, and NSCLC<sup>98</sup>.

As illustrated in the results section, three interconnected proteins, ARFGEF1, PLD2, and RPTOR, were successively activated, stimulating the cancer driver mTOR SP. Intriguingly, mutations in ARFGEF1, an activator of ARF1, have been reported in lung cancer<sup>100</sup>. Among these three proteins, PLDs have been garnering attention as potential contributors to cellular signalling of tumours<sup>101</sup>. Surprisingly, PLD represented an important intersection for other pathways in this study, and the first super network was constructed based on this enzyme. The existence of PLD isozymes (PLD1 and PLD2) in the SM protein dataset indicates the role of these junction boxes in the initiation of tumorigenesis. PLD is activated through a variety of signals such as neurotransmitters, hormones, and growth factors, transducing signals into pivotal cellular events, including proliferation, secretion, respiratory burst, and actin cytoskeletal reorganization<sup>102</sup>. As mentioned in the results section, PLD network involves proteins such as LPAR4, LPAR6, HCRTR1, and GRM5, as well as pathways like EGF/EGFR SP, which their roles have been proven in initiation and progression of several cancers. Phosphatidic acid, the product of PLD enzymatic activity, can be also converted to LPA and form a positive loop of LPA-LPARs<sup>30</sup>. LPA is involved in various cancer-related processes such as proliferation, growth, and survival<sup>75</sup>. Even though LPAR4 and LPAR6 are newly-discovered members of LPAR family, their contributions to hepatocellular carcinoma<sup>103</sup> and thyroid cancer<sup>104</sup> have been demonstrated, respectively. Although HCRTR1 stimulation with high doses of orexin causes apoptosis in colon and prostate cancers<sup>105,106</sup>, elevated serum orexin level has not been detected in cancer patients<sup>107</sup>. It seems that the normal serum level of orexin is not sufficient to induce apoptosis in cancer cells. Physiologically activated HCRTR1 may activate some pathways other than apoptosis, such as the PLD SP which contributes to tumour progression<sup>33,105</sup>. Therefore, pharmacologically activated HCRTR1 may be promising in cancer prevention or treatment.

Hypoxia, as a common feature of tumour microenvironment, increases the transcription of glutamate receptors, promotes the activation of GRM1, 3, 4, and 5, and supports proliferation and survival of several types of tumours through MAPK and PI3K/Akt SPs<sup>31,107</sup>. Noteworthy, GRM4 and 8 have been reported as susceptibility genes in NSCLC<sup>31</sup>. One of the critical pathways in both NSCLC and SM groups is the EGFR SP. More than half of lung tumours express receptors from EGFR family and one-fifth of lung tumours carry EGFR mutations, causing proliferation, migration, and metastasis<sup>68,71</sup>. Despite the preliminary response of all mutated lung tumours to tyrosine kinase inhibitors, they essentially acquire resistance<sup>68</sup>. Activation of mTOR, a master regulator of cell growth and proliferation, is one of the most important consequences of PLD activation. Accordingly, combination therapy of mTOR antagonists with standard regimens for lung cancer has enhanced treatment efficacy in early-stage trials<sup>108</sup>. Similarly, the MAPK SP, triggered by both PA and PLD, is associated with P53 degradation and MMPs induction, leading to cell survival and motility, respectively<sup>36</sup>. Overall, the PLD super network clarifies why treatment fails in this type of lung cancer and presents new insights into NSCLC chemotherapy. Moreover, this network underscores the pivotal function of PLDs in the occurrence of lung cancer in smokers.

Another super network constructed in this study is a complicated linkage between inflammatory cytokines and tumour microenvironment. Inflammation is a mixed blessing in cancer and the T helper 2 response deteriorates the state of cancer through suppression of cellular immunity<sup>56</sup>. Incidentally, the significance of IL-4 and IL-6 SPs in the present study may suggest inappropriate activation of this unfavourable arm of immunity. IL-4 SP promotes metastasis by changing the microenvironment of epithelial tumours<sup>76</sup>. It has been reported that serum IL-6 levels in NSCLC patients (especially metastatic cases) are higher than healthy subjects<sup>77,78</sup>, and blockade of IL-6/STAT3 signalling results in the inhibition of lung tumour growth<sup>109</sup>. Besides, STAT3 is persistently activated in GPCR C5a-knockout (KO) mice, a spontaneous lung cancer developing model, and the use of an antibody against LIF leads to STAT3 inactivation in this model<sup>110</sup>. The IL-6 and c-Met loop causes cell invasion through MMP activation in A549 cells, as well<sup>42,72</sup>. Proteins such as IL-6 and c-Met, in cooperation with IL-17 and TGF- $\beta$ , play various roles in the regulation of proteins such as mucins, collagens, and TJP1. It has been demonstrated that mucins such as MUC4 and MUC13 are associated with cancer development; the circulating level of MUC16 is even used to monitor patients with ovarian cancer<sup>111–113</sup>. Regarding collagens, their increased and decreased expressions are correlated with malignancy<sup>114</sup>; accordingly, collagen XXIII has been suggested as a biomarker for the diagnosis of NSCLC<sup>115</sup>. Another study showed that COL4A3 cleavage by MMP-9 produces tumstatin which is a circulating anti-angiogenic factor<sup>116</sup>. Hence, AAb against COL4A3 may worsen the state of cancer by preventing the generation of tumstatin. Additionally, COL27A1 is aberrantly expressed in idiopathic pulmonary fibrosis<sup>117</sup>, a condition associated with lung cancer<sup>118</sup>. Ni and colleagues found that the expression of TJP1 was decreased in NSCLC tissue in comparison to adjacent non-tumour tissue and its overexpression was correlated with better prognosis in NSCLC patients<sup>119</sup>. Considerably, increased MDSCs population as well as dysfunction of NKCs and CD8<sup>+</sup> T cells highlight an inefficient immune system in NSCLC<sup>120–122</sup>. Notably, our network exhibits how IL-4, IL-6, GM-CSF, and TGF- $\beta$  predispose these phenomena.

In conclusion, given the rapid emergence of lung cancer in smoker populations across the world, new aspects of NSCLC pathogenesis and effects of cigarette smoke should be addressed. Tailoring the AAb profiling with a systems biology approach provides an excellent image of the network of AAbs targeting proteins which may be key drivers of cancer-interacting SPs such as PLD, PI3K-Akt, MAPK, EGF/EGFR, c-Met, Ras, DDR, IL-4, IL-6, and TGF- $\beta$  SPs in the NSCLC and SM groups. Based on the PLD SP and inflammatory cytokines network, two super networks were constructed with NSCLC- and SM-specific proteins such as PLD1/2, PRKCD, PTK2B, LPARs, RPTOR, NOTCH1, TJP1, laminins, mucins, and collagens as well as pathways found in this study. Focusing on these super networks will help to design novel therapeutic agents targeting proteins involved in the pathogenesis of lung cancer.

## Methods

**Sample collection.** For obtaining a repertoire of NSCLC AAbs, 67 blood samples of untreated newly-diagnosed adenocarcinoma and squamous cell carcinoma patients (stages I, Ib, II, and IIIa) were collected from different centres, including Baqiyatallah and Masih Daneshvari hospitals<sup>123</sup>, Tehran, Iran, as well as Kurdistan Hospital in Sanandaj, Iran. The patients aged 22–77 years (mean, 50.7 years) were all confirmed for NSCLC via histopathological tests and were classified into different stages based on the tumour/node/metastasis scale (TNM) staging system. Furthermore, 57 subjects (age range, 20–72 years; mean, 40.0 years) were allocated to the SM group according to the complete blood cell count, blood level of C-reactive protein ( $\leq 6$  mg/dL), erythrocyte sedimentation rate ( $\leq 32$  mm/h), rheumatoid factor (negative), and chest X-ray imaging to evaluate their overall health. All the subjects in the SM group were current smokers with different smoking patterns, ranging from a few cigarettes per week to heavy smoking.

A control group of 92 healthy non-SM volunteers (range, 11–80 years; mean, 38.9 years) were also analysed, based on the same criteria considered for the subjects in the SM group and were enrolled in the study. All healthy participants avoided anti-inflammatory drugs for three days prior to blood collection. The demographic information was collected from the subjects, and they were interviewed for potential risk factors for cancers and autoimmune diseases, including personal and family medical history. This study was done in accordance with the Helsinki Declaration<sup>124,125</sup> and was approved by the Ethics Committee of Baqiyatallah University of Medical Sciences, the Ethics Committee of Masih Daneshvari Hospital<sup>123</sup>, and the Ethics Committee of Kurdistan University of Medical Sciences. All participants provided written informed consent before enrollment. The characteristics of all three groups are presented in Table S5.

**IgG purification from the sera of healthy control (HC), NSCLC patients, and SM.** The sera of each group (HC, NSCLC, and SM) were mixed separately and IgG purification was performed using Melon<sup>TM</sup> Gel IgG Purification Kit (Pierce, Rockford, IL) according to the manufacturer's instructions. The purification accuracy was confirmed by reducing SDS-PAGE analysis.

**Biopanning.** Using Ph.D.<sup>TM</sup>-C7C Phage Display Peptide Library Kit (New England Biolabs, Beverly, MA), SU and SL biopanning were carried out for three successive cycles on the purified IgG of NSCLC and SM groups. For SU biopanning, 96-well Nunc MaxiSorp flat-bottom plates were coated with IgG of HC, NSCLC and SM groups. In addition, for SL biopanning, Protein G Mag Sepharose Xtra (GE healthcare, US) was employed to expose IgG, according to the manufacturer's instructions. In both procedures, the phages of three inputs were subtracted on HC IgG before being transferred to the wells coated with IgG (NSCLC or SM IgG) or microtubes containing IgG (NSCLC or SM IgG) according to the instructions of Ph.D.<sup>TM</sup>-C7C Phage Display Peptide Library Kit and Protein G Mag Sepharose Xtra).

**Polyclonal and monoclonal phage ELISA.** Polyclonal phage ELISA was conducted on phages obtained from three rounds of SU and SL biopanning on NSCLC and SM IgG. The specificity of peptides displayed on 60 selected phages (30 clones from SU biopanning and 30 clones from SL biopanning) to NSCLC and SM IgG was individually assayed by monoclonal phage ELISA according to the manufacturer's instructions (Ph.D.<sup>TM</sup>-C7C kit) (see Supplementary information).

**DNA sequencing.** To extract the single-stranded phage DNAs, 20 and 10 clones obtained from SU and SL biopanning on NSCLC IgG, along with 15 and 10 clones obtained from SU and SL biopanning on SM IgG, were selected, respectively, according to the Ph.D.<sup>TM</sup>-C7C kit instructions. After sequencing and inference of amino acid sequences by Gene Runner version 5.0, the peptides were analysed in the Biopanning Data Bank (MimoDB) (<http://immunet.cn/bdb/>) to discard target-unrelated peptides<sup>126</sup>. Based on default parameters for the Blastp (score  $\geq 18.5$ ), the selected peptides were blasted against the Refseq\_protein database for the *Homo sapiens*<sup>127</sup>.

**Determination of protein complexes in the NSCLC and SM protein datasets.** NSCLC- and SM-specific proteins obtained through blasting (score  $\geq 18.5$ ) and preliminary analysis of the literature and UniProt database, were assessed in CPDB (<http://cpdb.molgen.mpg.de>) to find significant protein complexes with at least two members in the NSCLC or SM protein dataset ( $p$ -value  $< 0.05$ ).

**Primary evaluation of NSCLC and SM protein datasets in Pathway Studio®.** Pathway Studio® 11.2.5.9 (<https://mammalcedfx.pathwaystudio.com>) was used to exhibit the top cellular processes and diseases in two datasets. For this purpose, the NSCLC and SM protein datasets were separately introduced to Pathway Studio® and the Summary Analysis function was used to present a quick analytical summary of the datasets.

**GO and pathway-enrichment analysis of predicted proteins.** For visualizing molecular interaction networks, the NSCLC and SM protein datasets were introduced to STRING 10.0 (<http://string-db.org>)<sup>128</sup> and the nodes (proteins) and edges (protein-protein interactions) were extracted and transferred to Gephi 0.9.1<sup>129</sup>. The combined scores (from STRING) were defined as the edge weights. Modularity function implemented in Gephi was used for clustering, and the nodes were coloured differentially based on their modules. To identify GO terms and pathways related to the top six NSCLC and SM modules, Metascape (<http://metascape.org>)<sup>130</sup> and Enrichr web tools (<http://amp.pharm.mssm.edu/Enrichr/>)<sup>131</sup> were recruited. Disease Association Analysis in the WebGestalt database (<http://webgestalt.org>)<sup>132</sup> was performed for 25 NSCLC- and 26 SM-specific nodes (hubs) which showed the highest scores of betweenness centrality and degree. Target-based search were also implemented with DrugBank 5.0 (<https://drugbank.ca>)<sup>133</sup> and Reaxys® (<https://reaxys.com>) to identify marketed drugs affecting NSCLC-specific hubs.

To present a final image of NSCLC pathogenesis in this study, two super networks were constructed by combining pathways obtained from the KEGG PATHWAY and WikiPathways databases. Considering the PLD pathway as a central event in the NSCLC group, a super network was built on the PLD SP taken from the KEGG database<sup>30</sup>. This super network was developed by trimming the unrelated branches and adding connections based on the literature review.

Another super network was constructed, considering the effects of inflammatory cytokines on tumour microenvironment. For this purpose, proteins involved in GO terms and pathways related to inflammation or ECM, were introduced to Pathway Studio® for generating a primary network. Using “the properties function” defined in Pathway Studio®, each network component was assessed, leading to the generation of a secondary network expanded with NSCLC key proteins. Data mining was also carried out to verify this super network with the deliberated connections.

**Assessment of the selected hubs by ELISA.** PTK2B (ORIGENE) and NOTCH1 (R&D systems), as the representatives of NSCLC and SM groups, were selected for ELISA, respectively (see Supplementary information).

**Statistical analysis.** The statistical analyses were carried out using GraphPad prism®. The data were stated as the mean  $\pm$  SD and the statistical significance ( $p$ -value  $< 0.05$ ) was determined by two-tailed Student's  $t$ -test.

## References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2015. *CA Cancer J Clin* **65**, 5–29, <https://doi.org/10.3322/caac.21254> (2015).
2. Gridelli, C. *et al.* Non-small-cell lung cancer. *Nat Rev Dis Primers* 15009, <https://doi.org/10.1038/nrdp.2015.9> (2015).
3. Hecht, S. S. & Szabo, E. Fifty years of tobacco carcinogenesis research: from mechanisms to early detection and prevention of lung cancer. *Cancer Prev Res (Phila)* **7**, 1–8, <https://doi.org/10.1158/1940-6207.CAPR-13-0371> (2014).
4. Wu, L. *et al.* Development of autoantibody signatures as novel diagnostic biomarkers of non-small cell lung cancer. *Clin Cancer Res* **16**, 3760–3768, <https://doi.org/10.1158/1078-0432.ccr-10-0193> (2010).
5. Chang, A. Chemotherapy, chemoresistance and the changing treatment landscape for NSCLC. *Lung Cancer* **71**, 3–10, <https://doi.org/10.1016/j.lungcan.2010.08.022> (2011).
6. Zaenker, P., Gray, E. S. & Ziman, M. R. Autoantibody Production in Cancer—The Humoral Immune Response toward Autologous Antigens in Cancer Patients. *Autoimmun Rev* **15**, 477–483, <https://doi.org/10.1016/j.autrev.2016.01.017> (2016).
7. Yao, Y. *et al.* Potential application of non-small cell lung cancer-associated autoantibodies to early cancer diagnosis. *Biochem Biophys Res Commun* **423**, 613–619, <https://doi.org/10.1016/j.bbrc.2012.06.050> (2012).
8. Leidinger, P. *et al.* Toward an early diagnosis of lung cancer: An autoantibody signature for squamous cell lung carcinoma. *Int J Cancer* **123**, 1631–1636, <https://doi.org/10.1002/ijc.23680> (2008).
9. Amornsiripanitch, N. *et al.* Complement Factor H Autoantibodies Are Associated with Early Stage NSCLC. *Clin Cancer Res* **16**, 3226–3231, <https://doi.org/10.1158/1078-0432.ccr-10-0321> (2010).
10. Rom, W. N. *et al.* Identification of an autoantibody panel to separate lung cancer from smokers and nonsmokers. *BMC cancer* **10**, 1 (2010).
11. Hirasawa, Y. *et al.* Natural Autoantibody to MUC1 Is a Prognostic Indicator for Non-Small Cell Lung Cancer. *Am J Respir Crit Care Med* **161**, 589–594, <https://doi.org/10.1164/ajrccm.161.2.9905028> (2000).

12. Yao, X. *et al.* Dickkopf-1 autoantibody is a novel serological biomarker for non-small cell lung cancer. *Biomarkers: biochemical indicators of exposure, response, and susceptibility to chemicals* **15**, 128–134, <https://doi.org/10.3109/13547500903325662> (2010).
13. Chapman, C. J. *et al.* Autoantibodies in lung cancer: possibilities for early detection and subsequent cure. *Thorax* **63**, 228–233, <https://doi.org/10.1136/thx.2007.083592> (2008).
14. Brichory, F. M. *et al.* An immune response manifested by the common occurrence of annexins I and II autoantibodies and high circulating levels of IL-6 in lung cancer. *Proceedings of the National Academy of Sciences* **98**, 9824–9829, <https://doi.org/10.1073/pnas.171320598> (2001).
15. Wang, Y. C. & Chen, B. S. A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. *BMC Med Genomics* **4**, 2, <https://doi.org/10.1186/1755-8794-4-2> (2011).
16. Li, X. J. *et al.* A blood-based proteomic classifier for the molecular characterization of pulmonary nodules. *Sci Transl Med* **5**, 207ra142, <https://doi.org/10.1126/scitranslmed.3007013> (2013).
17. Liu, B. *et al.* Screening of diagnostic biomarkers for lung cancer by bioinformatics analysis. *International Journal Of Clinical And Experimental Medicine* **10**, 2737–2745 (2017).
18. Wang, L. Q., Zhao, L. H. & Qiao, Y. Z. Identification of potential therapeutic targets for lung cancer by bioinformatics analysis. *Mol Med Rep* **13**, 1975–1982, <https://doi.org/10.3892/mmr.2015.4752> (2016).
19. Young, J. H. *et al.* Computational discovery of pathway-level genetic vulnerabilities in non-small-cell lung cancer. *Bioinformatics* **32**, 1373–1379, <https://doi.org/10.1093/bioinformatics/btw010> (2016).
20. Huan, T. *et al.* A whole-blood transcriptome meta-analysis identifies gene expression signatures of cigarette smoking. *Hum Mol Genet* **25**, 4611–4623, <https://doi.org/10.1093/hmg/ddw288> (2016).
21. Cao, C. *et al.* Bioinformatics Analysis of the Effects of Tobacco Smoke on Gene Expression. *PLoS One* **10**, e0143377, <https://doi.org/10.1371/journal.pone.0143377> (2015).
22. Elamin, A. *et al.* Quantitative proteomics analysis using 2D-PAGE to investigate the effects of cigarette smoke and aerosol of a prototypic modified risk tobacco product on the lung proteome in C57BL/6 mice. *J Proteomics* **145**, 237–245, <https://doi.org/10.1016/j.jpro.2016.05.037> (2016).
23. Liu, Y. *et al.* Identification of feature genes for smoking-related lung adenocarcinoma based on gene expression profile data. *Oncotargets Ther* **9**, 7397–7407, <https://doi.org/10.2147/OTT.S114230> (2016).
24. Wang, Y., Qian, W. & Yuan, B. A graphical model of smoking-induced global instability in lung cancer. *IEEE/ACM Trans Comput Biol Bioinform*, <https://doi.org/10.1109/TCBB.2016.2599867> (2016).
25. Finegold, J. A., Asaria, P. & Francis, D. P. Mortality from ischaemic heart disease by country, region, and age: statistics from World Health Organisation and United Nations. *Int J Cardiol* **168**, 934–945, <https://doi.org/10.1016/j.ijcard.2012.10.046> (2013).
26. Majeed, A. *et al.* Diabetes in the Middle-East and North Africa: an update. *Diabetes Res Clin Pract* **103**, 218–222, <https://doi.org/10.1016/j.diabres.2013.11.008> (2014).
27. Poon, C. C. & Kelly, J. J. Development of crizotinib, a rationally designed tyrosine kinase inhibitor for non-small cell lung cancer. *Int J Cancer* (2016).
28. Wang, S., Cang, S. & Liu, D. Third-generation inhibitors targeting EGFR T790M mutation in advanced non-small cell lung cancer. *J Hem Onc* **9**, 1 (2016).
29. Shi, M., Zheng, Y., Garcia, A., Xu, L. & Foster, D. A. Phospholipase D provides a survival signal in human cancer cells with activated H-Ras or K-Ras. *Cancer Lett* **258**, 268–275, <https://doi.org/10.1016/j.canlet.2007.09.003> (2007).
30. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353–D361, <https://doi.org/10.1093/nar/gkw1092> (2017).
31. Hu, H. *et al.* Hypoxia-inducible factors enhance glutamate signaling in cancer cells. *Oncotarget* **5**, 8853–8868 (2014).
32. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*, gkv1070 (2015).
33. Jantti, M. H., Putula, J., Somerharju, P., Frohman, M. A. & Kukkonen, J. P. OX1 orexin/hypocretin receptor activation of phospholipase D. *Br J Pharmacol* **165**, 1109–1123, <https://doi.org/10.1111/j.1476-5381.2011.01565.x> (2012).
34. Cherfils, J. & Zeghouf, M. Regulation of Small GTPases by GEFs, GAPs, and GDIs. *Physiol Rev* **93**, 269–309, <https://doi.org/10.1152/physrev.00003.2012> (2013).
35. Sun, Y. & Chen, J. mTOR signaling: PLD takes center stage. *Cell Cycle* **7**, 3118–3123, <https://doi.org/10.4161/cc.7.20.6881> (2008).
36. Bruntz, R. C., Lindsley, C. W. & Brown, H. A. Phospholipase D signaling pathways and phosphatidic acid as therapeutic targets in cancer. *Pharmacol Rev* **66**, 1033–1079, <https://doi.org/10.1124/pr.114.009217> (2014).
37. Reich, R., Blumenthal, M. & Liscovitch, M. Role of phospholipase D in laminin-induced production of gelatinase A (MMP-2) in metastatic cells. *Clin Exp Metastasis* **13**, 134–140, <https://doi.org/10.1007/BF00133618> (1995).
38. Mar, A.-C. Abstract 1361: The decoy receptor interleukin-1 receptor type 2 acts as an angiogenic factor in human colorectal cancer. *Cancer Res* **75**, 1361–1361, <https://doi.org/10.1158/1538-7445.am2015-1361> (2015).
39. Chabaud, M., Fossiez, F., Taupin, J.-L. & Miossec, P. Enhancing Effect of IL-17 on IL-1-Induced IL-6 and Leukemia Inhibitory Factor Production by Rheumatoid Arthritis Synoviocytes and Its Regulation by Th2 Cytokines. *J Immunol* **161**, 409–414 (1998).
40. Wang, L. *et al.* IL-17 can promote tumor growth through an IL-6–Stat3 signaling pathway. *J Exp Med* **206**, 1457–1464, <https://doi.org/10.1084/jem.20090207> (2009).
41. Lee, H. *et al.* Cigarette smoke-mediated oxidative stress induces apoptosis via the MAPKs/STAT1 pathway in mouse lung fibroblasts. *Toxicol Lett* **240**, 140–148, <https://doi.org/10.1016/j.toxlet.2015.10.030> (2016).
42. To, Y. *et al.* A two-way interaction between hepatocyte growth factor and interleukin-6 in tissue invasion of lung cancer cell line. *Am J Respir Cell Mol Biol* **27**, 220–226 (2002).
43. Liton, P. B., Li, G., Luna, C., Gonzalez, P. & Epstein, D. L. Cross-talk between TGF- $\beta$ 1 and IL-6 in human trabecular meshwork cells. *Mol Vis* **15**, 326–334 (2009).
44. Carraway, K. L. & Carothers Carraway, C. A. MUC4 and membrane receptors in cancer. *Mucins and Cancer* 68–81, <https://doi.org/10.2217/fmeb2013.13.69> (2013).
45. Gupta, B. K. *et al.* Functions and regulation of MUC13 mucin in colon cancer cells. *J Gastroenterol* **49**, 1378–1391, <https://doi.org/10.1007/s00535-013-0885-z> (2014).
46. Arimura, A., van Peer, M., Schröder, A. J. & Rothman, P. B. The transcriptional co-activator p/CIP (NCoA-3) is up-regulated by STAT6 and serves as a positive regulator of transcriptional activation by STAT6. *Journal of Biological Chemistry* **279**, 31105–31112 (2004).
47. Kumar, S. *et al.* NCOA3-mediated upregulation of mucin expression via transcriptional and post-translational changes during the development of pancreatic cancer. *Oncogene* (2014).
48. Chen, P., Cescon, M. & Bonaldo, P. Collagen VI in cancer and its biological mechanisms. *Trends Mol Med* **19**, 410–417, <https://doi.org/10.1016/j.molmed.2013.04.001> (2013).
49. Chen, Y., Yang, P., Li, F. & Kijlstra, A. The Effects of Th17 Cytokines on the Inflammatory Mediator Production and Barrier Function of ARPE-19 Cells. *PLoS ONE* **6**, e18139, <https://doi.org/10.1371/journal.pone.0018139> (2011).
50. Ahdieh, M., Vandenbos, T. & Youakim, A. Lung epithelial barrier function and wound healing are decreased by IL-4 and IL-13 and enhanced by IFN- $\gamma$ . *Am J Physiol Cell Physiol* **281**, C2029–C2038 (2001).
51. Chaturvedi, P., Singh, A. P. & Batra, S. K. Structure, evolution, and biology of the MUC4 mucin. *FASEB J* **22**, 966–981 (2008).

52. Lee, S. H. *et al.* Tight junction protein 1 is regulated by transforming growth factor- $\beta$  and contributes to cell motility in NSCLC cells. *BMB Rep* **48**, 115–120, <https://doi.org/10.5483/BMBRep.2015.48.2.035> (2015).
53. Wen, G. *et al.* TGFBI expression reduces *in vitro* and *in vivo* metastatic potential of lung and breast tumor cells. *Cancer Lett* **308**, 23–32, <https://doi.org/10.1016/j.canlet.2011.04.010> (2011).
54. Chen, M.-L. *et al.* Regulatory T cells suppress tumor-specific CD8 T cell cytotoxicity through TGF- $\beta$  signals *in vivo*. *Proc Natl Acad Sci USA* **102**, 419–424, <https://doi.org/10.1073/pnas.0408197102> (2005).
55. Wahl, S. M., Wen, J. & Moutsopoulos, N. M. The kiss of death: interrupted by NK-cell close encounters of another kind. *Trends Immunol* **27**, 161–164, <https://doi.org/10.1016/j.it.2006.02.002> (2006).
56. O'Callaghan, D. S., O'Donnell, D., O'Connell, F. & O'Byrne, K. J. The Role of Inflammation in the Pathogenesis of Non-small Cell Lung Cancer. *J Thorac Oncol* **5**, 2024–2036, <https://doi.org/10.1097/JTO.0b013e3181f387e4> (2010).
57. Wang, X. & Lin, Y. Tumor necrosis factor and cancer, buddies or foes? *Acta Pharmacol Sin* **29**, 1275–1288, <https://doi.org/10.1111/j.1745-7254.2008.00889.x> (2008).
58. Ślebioda, T. J. & Kmiec, Z. Tumour necrosis factor superfamily members in the pathogenesis of inflammatory bowel disease. *Mediators Inflamm* **2014** (2014).
59. Jin, S. *et al.* TL1A/TNFSF15 directly induces proinflammatory cytokines, including TNF $\alpha$ , from CD3 $^{+}$ ; CD161 $^{+}$ ; T cells to exacerbate gut inflammation. *Mucosal Immunol* **6**, 886–899 (2013).
60. Kim, E. S. In *Lung Cancer and Personalized Medicine: Current Knowledge and Therapies* (eds Aamir Ahmad & Shirish Gadgil) 189–209 (Springer International Publishing, 2016).
61. Alberg, A. J., Brock, M. V., Ford, J. G., Samet, J. M. & Spivack, S. D. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: american college of chest physicians evidence-based clinical practice guidelines. *Chest* **143**, e1S–e29S, <https://doi.org/10.1378/chest.12-2345> (2013).
62. Zaenker, P. & Ziman, M. R. Serologic autoantibodies as diagnostic cancer biomarkers—a review. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **22**, 2161–2181, <https://doi.org/10.1158/1055-9965.EPI-13-0621> (2013).
63. Järås, K. & Anderson, K. Autoantibodies in cancer: prognostic biomarkers and immune activation. *Expert Rev Proteomics* **8**, 577–589 (2011).
64. Vicent, S. *et al.* ERK1/2 is activated in non-small-cell lung cancer and associated with advanced tumours. *Br J Cancer* **90**, 1047–1052, <https://doi.org/10.1038/sj.bjc.6601644> (2004).
65. Roberts, P. J. & Der, C. J. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* **26**, 3291–3310, <https://doi.org/10.1038/sj.onc.1210422> (2007).
66. Yip, P. Y. Phosphatidylinositol 3-kinase-AKT-mammalian target of rapamycin (PI3K-Akt-mTOR) signaling pathway in non-small cell lung cancer. *Transl Lung Cancer Res* **4**, 165–176, <https://doi.org/10.3978/j.issn.2218-6751.2015.01.04> (2015).
67. Mebratu, Y. & Tesfaigzi, Y. How ERK1/2 Activation Controls Cell Proliferation and Cell Death Is Subcellular Localization the Answer? *Cell cycle (Georgetown, Tex.)* **8**, 1168–1175 (2009).
68. Siegelin, M. D. & Borczuk, A. C. Epidermal growth factor receptor mutations in lung adenocarcinoma. *Lab Invest* **94**, 129–137, <https://doi.org/10.1038/labinvest.2013.147> (2014).
69. Cortas, T. *et al.* Activation state EGFR and STAT-3 as prognostic markers in resected non-small cell lung cancer. *Lung cancer* **55**, 349–355 (2007).
70. Mitsudomi, T. & Yatabe, Y. Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. *The FEBS journal* **277**, 301–308, <https://doi.org/10.1111/j.1742-4658.2009.07448.x> (2010).
71. Normanno, N. *et al.* Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene* **366**, 2–16, <https://doi.org/10.1016/j.gene.2005.10.018> (2006).
72. Landi, L., Minuti, G., D'Incecco, A. & Cappuzzo, F. Targeting c-MET in the battle against advanced nonsmall-cell lung cancer. *Curr Opin Oncol* **25**, 130–136, <https://doi.org/10.1097/CCO.0b013e32835daf37> (2013).
73. Liu, X., Yao, W., Newton, R. C. & Scherle, P. A. Targeting the c-MET signaling pathway for cancer therapy. *Expert Opin Investig Drugs* **17**, 997–1011, <https://doi.org/10.1517/13543784.17.7.997> (2008).
74. Gherardi, E., Birchmeier, W., Birchmeier, C. & Vande Woude, G. Targeting MET in cancer: rationale and progress. *Nat Rev Cancer* **12**, 89–103, <https://doi.org/10.1038/nrc3205> (2012).
75. Leblanc, R. & Peyruchaud, O. New insights into the autotaxin/LPA axis in cancer development and metastasis. *Exp Cell Res* **333**, 183–189, <https://doi.org/10.1016/j.yexcr.2014.11.010> (2015).
76. Bankaitis, K. V. & Fingleton, B. Targeting IL4/IL4R for the treatment of epithelial cancer metastasis. *Clin Exp Metastasis* **32**, 847–856, <https://doi.org/10.1007/s10585-015-9747-9> (2015).
77. Kayacan, O. *et al.* Impact of TNF-alpha and IL-6 levels on development of cachexia in newly diagnosed NSCLC patients. *Am J Clin Oncol* **29**, 328–335, <https://doi.org/10.1097/01.coc.0000221300.72657.e0> (2006).
78. Yanagawa, H. *et al.* Serum levels of interleukin 6 in patients with lung cancer. *Br J Cancer* **71**, 1095–1098 (1995).
79. Guo, Y., Xu, F., Lu, T., Duan, Z. & Zhang, Z. Interleukin-6 signaling pathway in targeted therapy for cancer. *Cancer Treat Rev* **38**, 904–910, <https://doi.org/10.1016/j.ctrv.2012.04.007> (2012).
80. Moldovan, G.-L. & D'Andrea, A. D. How the Fanconi Anemia pathway guards the genome. *Annu Rev Genet* **43**, 223–249, <https://doi.org/10.1146/annurev-genet-102108-134222> (2009).
81. Cremona, C. A. & Behrens, A. ATM signalling and cancer. *Oncogene* **33**, 3351–3360, <https://doi.org/10.1038/onc.2013.275> (2014).
82. Coomans de Brachène, A. & Demoulin, J.-B. FOXO transcription factors in cancer development and therapy. *Cell Mol Life Sci* **73**, 1159–1172, <https://doi.org/10.1007/s00018-015-2112-y> (2016).
83. Stamos, J. L. & Weis, W. I. The beta-catenin destruction complex. *Cold Spring Harb Perspect Biol* **5**, a007898, <https://doi.org/10.1101/cshperspect.a007898> (2013).
84. Mocellin, S. & Nitti, D. TNF and cancer: the two sides of the coin. *Front Biosci* **13**, 2774–2783 (2007).
85. Neuzillet, C. *et al.* Targeting the TGFbeta pathway for cancer therapy. *Pharmacol Ther* **147**, 22–31, <https://doi.org/10.1016/j.pharmthera.2014.11.001> (2015).
86. Tu, Z. *et al.* CFTR is a potential marker for nasopharyngeal carcinoma prognosis and metastasis. *Oncotarget* **7**, 76955–76965, <https://doi.org/10.18632/oncotarget.12762> (2016).
87. Li, J. *et al.* The cystic fibrosis transmembrane conductance regulator as a biomarker in non-small cell lung cancer. *Int J Oncol* **46**, 2107–2115, <https://doi.org/10.3892/ijo.2015.2921> (2015).
88. Li, C., Schuetz, J. D. & Naren, A. P. Tobacco carcinogen NNK transporter MRP2 regulates CFTR function in lung epithelia: implications for lung cancer. *Cancer Lett* **292**, 246–253, <https://doi.org/10.1016/j.canlet.2009.12.009> (2010).
89. Zhang, S., Qiu, X., Gu, Y. & Wang, E. Up-regulation of proline-rich tyrosine kinase 2 in non-small cell lung cancer. *Lung Cancer* **62**, 295–301, <https://doi.org/10.1016/j.lungcan.2008.05.008> (2008).
90. Block, E. R., Tolino, M. A. & Klarlund, J. K. Pyk2 Activation Triggers Epidermal Growth Factor Receptor Signaling and Cell Motility after Wounding Sheets of Epithelial Cells. *J Biol Chem* **285**, 13372–13379, <https://doi.org/10.1074/jbc.M109.083089> (2010).
91. Gridling, M. *et al.* Identification of kinase inhibitor targets in the lung cancer microenvironment by chemical and phosphoproteomics. *Mol Cancer Ther* **13**, 2751–2762, <https://doi.org/10.1158/1535-7163.MCT-14-0152> (2014).
92. Symonds, J. M. *et al.* Protein kinase C delta is a downstream effector of oncogenic K-ras in lung tumors. *Cancer Res* **71**, 2087–2097, <https://doi.org/10.1158/0008-5472.CAN-10-1511> (2011).

93. Clark, A. S., West, K. A., Blumberg, P. M. & Dennis, P. A. Altered Protein Kinase C (PKC) Isoforms in Non-Small Cell Lung Cancer Cells PKC $\delta$  Promotes Cellular Survival and Chemotherapeutic Resistance. *Cancer Res* **63**, 780–786 (2003).
94. Liu, B. *et al.* MiR-195 suppresses non-small cell lung cancer by targeting CHEK1. *Oncotarget* **6**, 9445–9456 (2015).
95. Yu, X., Zhao, Z., Ma, X., Du, L. & Pertsemliadis, A. miR-195 represses the tumorigenesis of non-small cell lung cancer and synergizes with microtubule targeting agents. *Cancer Res* **76**, 1063–1063 (2016).
96. Yuan, X. *et al.* Notch signaling: an emerging therapeutic target for cancer treatment. *Cancer Lett* **369**, 20–27, <https://doi.org/10.1016/j.canlet.2015.07.048> (2015).
97. Kikuchi, H. *et al.* Expression of Notch1 and Numb in small cell lung cancer. *Oncotarget* **8**, 10348–10358, <https://doi.org/10.18632/oncotarget.14411> (2017).
98. Yang, Y. L., Jablons, D. & You, L. An alternative way to initiate Notch1 signaling in non-small cell lung cancer. *Transl Lung Cancer Res* **3**, 238–241, <https://doi.org/10.3978/j.issn.2218-6751.2013.12.06> (2014).
99. Konishi, J. *et al.* Gamma-secretase inhibitor prevents Notch3 activation and reduces proliferation in human lung cancers. *Cancer Res* **67**, 8051–8057, <https://doi.org/10.1158/0008-5472.CAN-07-1022> (2007).
100. Lanzetti, L. & Di Fiore, P. P. Endocytosis and cancer: an 'insider' network with dangerous liaisons. *Traffic* **9**, 2011–2021, <https://doi.org/10.1111/j.1600-0854.2008.00816.x> (2008).
101. Gomez-Cambornero, J. & Carman, G. M. Thematic minireview series on phospholipase D and cancer. *J Biol Chem* **289**, 22554–22556, <https://doi.org/10.1074/jbc.R114.593137> (2014).
102. Ha, S. H. *et al.* PLD2 forms a functional complex with mTOR/raptor to transduce mitogenic signals. *Cell Signal* **18**, 2283–2291, <https://doi.org/10.1016/j.cellsig.2006.05.021> (2006).
103. Mazzocca, A. *et al.* Lysophosphatidic acid receptor LPAR6 supports the tumorigenicity of hepatocellular carcinoma. *Cancer Res* **75**, 532–543, <https://doi.org/10.1158/0008-5472.CAN-14-1607> (2015).
104. Pan, W. *et al.* Whole exome sequencing identifies lncRNA GAS8-AS1 and LPAR4 as novel papillary thyroid carcinoma driver alternations. *Hum Mol Genet*, doi:<https://doi.org/10.1093/hmg/ddw056> (2016).
105. Alexandre, D. *et al.* The orexin type 1 receptor is overexpressed in advanced prostate cancer with a neuroendocrine differentiation, and mediates apoptosis. *Eur J Cancer* **50**, 2126–2133, <https://doi.org/10.1016/j.ejca.2014.05.008> (2014).
106. Laburthe, M. & Voisin, T. The orexin receptor OX(1)R in colon cancer: a promising therapeutic target and a new paradigm in G protein-coupled receptor signalling through ITIMs. *Br J Pharmacol* **165**, 1678–1687, <https://doi.org/10.1111/j.1476-5381.2011.01510.x> (2012).
107. Prickett, T. D. & Samuels, Y. Molecular pathways: dysregulated glutamatergic signaling pathways in cancer. *Clin Cancer Res* **18**, 4240–4246, <https://doi.org/10.1158/1078-0432.CCR-11-1217> (2012).
108. Pal, S. K., Figlin, R. A. & Reckamp, K. L. The role of targeting mammalian target of rapamycin in lung cancer. *Clin Lung Cancer* **9**, 340–345, <https://doi.org/10.3816/CLC.2008.n.049> (2008).
109. Lee, J. J. *et al.* A high-affinity protein binder that blocks the IL-6/STAT3 signaling pathway effectively suppresses non-small cell lung cancer. *Mol Ther* **22**, 1254–1265, <https://doi.org/10.1038/mt.2014.59> (2014).
110. Chen, Y. *et al.* Gprc5a Deletion Enhances the Transformed Phenotype in Normal and Malignant Lung Epithelial Cells by Eliciting Persistent Stat3 Signaling Induced by Autocrine Leukemia Inhibitory Factor. *Cancer Res* **70**, 8917–8926, <https://doi.org/10.1158/0008-5472.can-10-0518> (2010).
111. Kufe, D. W. Mucins in cancer: function, prognosis and therapy. *Nat Rev Cancer* **9**, 874–885, <https://doi.org/10.1038/nrc2761> (2009).
112. Chauhan, S. C. *et al.* Expression and Functions of Transmembrane Mucin MUC13 in Ovarian Cancer. *Cancer Res* **69**, 765–774, <https://doi.org/10.1158/0008-5472.can-08-0587> (2009).
113. Lakshmanan, I. *et al.* Mucins in lung cancer: diagnostic, prognostic, and therapeutic implications. *J Thorac Oncol* **10**, 19–27, <https://doi.org/10.1097/JTO.0000000000000404> (2015).
114. Egeblad, M., Rasch, M. G. & Weaver, V. M. Dynamic interplay between the collagen scaffold and tumor evolution. *Curr Opin Cell Biol* **22**, 697–706, <https://doi.org/10.1016/j.ceb.2010.08.015> (2010).
115. Spivey, K. A. *et al.* Collagen XXIII: a potential biomarker for the detection of primary and recurrent non-small cell lung cancer. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **19**, 1362–1372, <https://doi.org/10.1158/1055-9965.EPI-09-1095> (2010).
116. Hamano, Y. *et al.* Physiological levels of tumstatin, a fragment of collagen IV  $\alpha 3$  chain, are generated by MMP-9 proteolysis and suppress angiogenesis via  $\alpha V\beta 3$  integrin. *Cancer Cell* **3**, 589–601, [https://doi.org/10.1016/s1535-6108\(03\)00133-8](https://doi.org/10.1016/s1535-6108(03)00133-8) (2003).
117. Selman, M. *et al.* Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity pneumonitis. *Am J Respir Crit Care Med* **173**, 188–198, <https://doi.org/10.1164/rccm.200504-644OC> (2006).
118. Samet, J. M. Does idiopathic pulmonary fibrosis increase lung cancer risk? *Am J Respir Crit Care Med* **161**, 1 (2000).
119. Ni, S. *et al.* Increased ZO-1 expression predicts valuable prognosis in non-small cell lung cancer. *Int J Clin Exp Pathol* **6**, 2887–2895 (2013).
120. Domagala-Kulawik, J. The role of the immune system in non-small cell lung carcinoma and potential for therapeutic intervention. *Transl Lung Cancer Res* **4**, 177–190, <https://doi.org/10.3978/j.issn.2218-6751.2015.01.11> (2015).
121. Carrega, P. *et al.* Natural killer cells infiltrating human nonsmall-cell lung cancer are enriched in CD56 bright CD16(–) cells and display an impaired capability to kill tumor cells. *Cancer* **112**, 863–875, <https://doi.org/10.1002/cncr.23239> (2008).
122. Prado-Garcia, H., Romero-Garcia, S., Aguilar-Cazares, D., Meneses-Flores, M. & Lopez-Gonzalez, J. S. Tumor-Induced CD8+ T-Cell Dysfunction in Lung Cancer Patients. *Clin Dev Immunol* **2012**, 11, <https://doi.org/10.1155/2012/741741> (2012).
123. Tafsiiri, E., Darbouy, M., Shadmehr, M. B., Cho, W. C. & Karimipoor, M. Abberent expression of oncogenic and tumor-suppressive microRNAs and their target genes in human adenocarcinoma alveolar basal epithelial cells. *J Cancer Res Ther* **12**, 395–400, <https://doi.org/10.4103/0973-1482.148673> (2016).
124. Morton, L. M. *et al.* Etiologic heterogeneity among non-Hodgkin lymphoma subtypes. *Blood* **112**, 5150–5160 (2008).
125. Zwick, C. *et al.* Autoantigenic targets of B-cell receptors derived from chronic lymphocytic leukemias bind to and induce proliferation of leukemic cells. *Blood* **121**, 4708–4717 (2013).
126. Vodnik, M., Zager, U., Strukelj, B. & Lunder, M. Phage display: selecting straws instead of a needle from a haystack. *Molecules* **16**, 790–817, <https://doi.org/10.3390/molecules16010790> (2011).
127. Liu, X. *et al.* Serum Antibody Repertoire Profiling Using In Silico Antigen Screen. *PLoS One* **8**, e67181 (2013).
128. Szklarczyk, D. *et al.* STRINGv10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447–452, <https://doi.org/10.1093/nar/gku1003> (2015).
129. Bastian, M., Heymann, S. & Jacomy, M. Gephi: an open source software for exploring and manipulating networks. *ICWSM* **8**, 361–362 (2009).
130. Tripathi, S. *et al.* Meta- and Orthogonal Integration of Influenza “OMICs” Data Defines a Role for UBR4 in Virus Budding. *Cell host & microbe* **18**, 723–735, <https://doi.org/10.1016/j.chom.2015.11.002> (2015).
131. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, gkw377 (2016).
132. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based gene set analysis toolkit (WebGestalt): update 2013. *Nucleic Acids Res* **41**, W77–W83 (2013).
133. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* **34**, D668–D672 (2006).

## Acknowledgements

The authors would like to thank the personnel of Baqiyatallah, Masih Daneshvari, and Kurdistan hospital for their assistance with sample collections. This work was supported by grants from the Iran national science foundation (grant number 92002095).

## Author Contributions

F.R.J. and M.G. conceived and designed the experiments. F.M. helped to design the study. F.R.J. and M.P. performed the data analysis and interpretation. F.R.J., F.S., T.B., and F.A. performed the experiments. A.F.K., E.V., M.M., N.S., and E.T. contributed to the acquisition and interpretation of data. F.R.J. and M.P. wrote the paper. S.D.S., F.V., and A.F. were involved in the manuscript preparation. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-14195-1>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017