Significant pathway and biomarker identification of pancreatic cancer associated lung cancer


Abstract

Pancreatic Cancer (PC) is the seventh leading cause of cancer-related death worldwide and one of the deadliest malignant neoplasms. It is difficult to diagnose at an early stage. Lung Cancer (LC) is the second most common cancer and a malignant tumor which spreads into nearby tissues through metastasis. One of the major common risk factors between PC and LC is smoking. Since both PC and LC are malignant they may be associated with one another. The main purpose of this research was to find genetic linkage and association between PC and LC through a bioinformatics approach. We analyzed both gene expression datasets and compared the significant genetic markers to find shared genes between the two cancers. We identified 1686 common significant genes between PC and LC, which is 33.2% of both gene datasets. The enrichment analysis disclosed the biological roles of the selected common genes and revealed the pathways that were mostly associated with the common biomarkers. Using the common genes of both cancers we constructed a PPI network and identified 10 significant hub genes. This drug compound could be useful for treatment of the patient who suffers simultaneously from both cancers. Thus this study identified genetic linkage and elemental functional association between PC and LC, and may help in the development of future medications for patients with both PC and LC.

1. Introduction

Pancreatic Cancer (PC) is the seventh leading cause of cancer-related deaths worldwide and the third most common cancer in the USA according to GLOBOCAN 2018 estimation [1]. The pancreas is a pear-shaped gland that is located in the abdomen between the stomach and the spine [2]. When the healthy pancreatic cells don’t work effectively and grow out of control, then it becomes a malignant cancerous tumor, which means it can develop and spread to other internal organs of the body. Through a process called metastasis, a tumor cell in the pancreas can alter the efficacy of the pancreas, expand into adjacent veins and lymphatic organs, and quickly spread to other body parts [2].

PC remains one of the most deadly malignant carcinomas, with 458,918 confirmed cases and 432,242 fatalities (4.5% of all cancer related incidents) globally in 2018, and it is estimated that there will be 355,317 reported cases by 2040 [1]. PC is slightly more common in men than women: it caused 24,640 male and 22,410 female deaths in 2019 (American Cancer Society) [3]. It is estimated that PC will be diagnosed in 57,600 people in the USA, and more than 47,050 individuals will die from the disease in 2020 [4]. Of all major cancers, PC has the highest mortality rate. The one-year survival rate is 24% and the 5-year relative survival rate is 9% for all combined stages together [1]. The 5-year survival rate is only 37% for the small percentage of people who are diagnosed with other local diseases. For the majority of patients who are...
diagnosed at a distant stage of cancer, the 5-year survival rate is 3% [5]. Pancreatic ductal adenocarcinoma is the most common (80%-85%) type of PC compared to the pancreatic neuroendocrine tumor which is less than 5% [6,7]. The cause of PC is complicated and multifactorial, but smoking tobacco, family history, obesity, and type 2 diabetes are prevalent in this case [8]. There is some evidence that consuming processed meat, red meat, alcoholic drinks, foods that contain saturated fatty acids [9] and drinks containing fructose might increase the risk of PC [10]. Symptoms for PC may include unexplained weight loss, discomfort in the abdomen, back pain, jaundice, light-colored stools, and fatigue [11].

Lung Cancer (LC) is the second most frequent and the leading cause of cancer related death for both men and women [12]. It is estimated that 135,720 deaths (72,500 men and 63,220 women) will occur in 2020 from this disease [4]. Lung Cancer (LC) is a malignant tumor also known as lung carcinoma that can grow in tissues of the lung. Through a process called metastasis, it can spread into nearby tissues, organs, or other body parts beyond the lung [13]. We have two sponge-like organs called lungs in our chest. The right lung has three sections which are called lobes and the left one has two lobes. The right lung is bigger than the left lung because the heart takes up more room on the left side of the body [14].

In 2018, there were an estimated 2.1 million cases of LC and 1.8 million deaths globally [15]. According to the American Cancer Society, 72,500 males and 63,220 females died from LC in 2019. It is estimated that LC will be diagnosed by 228,820 adults (116,300 men and 112,520 women) in the US in 2020 [4]. The 5-year survival rate tells us about the percentage of people who live at least 5 years after cancer has been detected. The percentage of people means how many people out of 100 [12]. The 5-year survival rate is 19% for individuals who have all types of LC [12]. The 5-year survival rate for the localized stage is 17% [16]. The 5-year survival rate for the regional stage is 22% [16]. For NSCLC the 5-year survival rate is 24%. The most common type of LC is NSCLC (84% of all LC diagnoses), compared to less common small cell lung cancer (SCLC) which is 6% of all diagnosed cases [12]. The cause of LC is mainly smoking tobacco. Smoking is the cause of death for about 80% of patients of LC and many others are caused by secondhand smoke exposure [17]. Smokers who are exposed to radon and asbestos are at an even higher risk [17]. Other factors like secondhand smoke, air pollution, exposure to asbestos, diesel exhaust, and genetics probably play a role as well for non-smokers [18-20]. Some common symptoms for LC are persistent coughing, chest pain, throwing up blood, unexplained weight loss, hoarseness, fatigue, breath shortages, and repeated respiratory infections [21].

Recently, DA Barth, C Brenner et al. (2020) [22] aimed to assess and externally validate the effectiveness of the advanced lung cancer inflammation index (ALI) as a prognostic biomarker in a cohort of PC patients for cancer specific survival in their study. In their single centered cohort study, they analyzed 429 PC patients to find the association between ALI and PC using Spearman’s rank correlation analysis, ROC-curve analysis, Kaplan-Meier estimator, Hazard ratios, Uni- and multivariate Cox Proportional models on clinicopathological parameters of PC patients. The result showed that although ALI was found a reliable and significant predictor in LC (NSCLC and SCLC) and other cancers such as colorectal and esophageal but ALI was not found to be a prognosis biomarker for PC determination. Meanwhile, Kurahara H, Maemura K et al. (2020) [23] said that the lung is one of the major metastatic sites of PC. They aimed to evaluate the functionality and prognosis of PC patients with lung metastasis and also wanted to identify clinicopathological features of lung metastasis. In the study, they enrolled 183 PC patients who had undergone surgery and clinically examined them including FDG-PET, CT, and MRI. In histological examination, all resected lung metastasis specimen were examined and analyzed with primary PC specimen. For statistical analysis, they performed Fisher’s exact test, Mann-Whitney U test, Kaplan-Meier method for survival curve and analyzed it with the log-rank test. As a result, they found that patients with single lung metastasis showed a suitable prognosis that underwent surgical resection, and single lung metastasis was associated significantly with lower FDG-PET SUV_{max} of primary PC tumor. Also, Morimoto D., Yamada S., et al. (2020) [24] reported that lung metastasis works as an initial recurrence pattern in PC patients after curative resection of PC usually occurs in the patient’s body. They assessed their study with clinicological factors and survival durations for patients with lung metastasis and compared them with other recurrence patients. In all these studies they didn’t show the genetic association between PC and LC. Their methods were clinical examination, statistical and cohort study based.

In this analysis, we showed the genetic linkage and connections between PC and LC by identifying common associated genes between these two selected cancer diseases. Using these common genes we have performed GO (Gene Ontology) term analysis, pathway enrichment analysis to identify enriched pathways by the identified common genes. Network analysis, hub gene identification, cluster identification, and drug target analysis were also performed for further analysis. Thus this study aimed to find genetic linkage and association between PC and LC through a bioinformatics approach.

2. Methodology

Gene collection, common gene identification, enrichment analysis, PPI analysis, drug target identification and analysis were performed to complete this study. All phases of the proposed methodologies are illustrated below and a short visualization of this research work is exhibited in Fig. 1.

2.1. Gene collection

DisGeNET is a consistent tool that helps in easy analysis and interpretation of human gene-disease networks [25]. From DisGeNET database we have collected genes for PC and LC which is human associated. After that we filtered the common genes from both selected diseases using an online tool called ‘Venny’ for further analysis [26].

2.2. Enrichment analysis

Enrichment analysis is an important set of computational methods which is conducted to know the biological roles of the selected genes like Gene Ontology (GO) annotation and KEGG pathway analysis [27,28]. With the common genes identified from both diseases, we have collected the associated data of GO terms: biological process (BP), molecular function (MF), cellular component (CC) from the GO database, and

![Fig. 1. Flowchart of the proposed methodology.](image-url)
human associated pathway dataset was collected from KEGG 2016 database using an online web based enrichment analysis tool called ‘Enrichr’ [29]. The cut-off criterion for all enrichment analysis on GO terms and KEGG pathways was considered with a p-value < 0.05. To perform enrichment analysis on gene sets is one of the main uses of Gene Ontology. GO terms are divided into three categories: MF, CC, and BP. MF (Molecular Function) represents activities that occur at the molecular level by gene products. CC (Cellular Component) represents the location that relates to cellular structures where functions are performed by gene products. BP (Biological Process) represents the broader operations that are conducted by various molecular activities [30]. KEGG is a database collection that deals with genomes, diseases, drugs, biological pathways, and systemic functional information [31].

2.3. PPI network construction and module analysis

PPI (Protein-Protein Interaction) is referred to as the graph of nodes connected with edges where nodes and edges represent proteins and their interaction with each other respectively [32]. From the selected diseases (PC and LC) a PPI network of common genes has been developed using online bioinformatics web based application STRING (http://string-db.org/) with a confidence score 0.40 which is the minimum required interaction score [33]. After constructing the PPI network using STRING, we analyzed the network with the help of a Cytoscape software tool called Network Analyzer. Cytoscape is an open source platform that helps in visualization, analysis, and biological network modeling [34]. We identified the top 10 hub genes based on higher connectivity and through 11 topological analysis methods using the cytoHubba plugin tool of Cytoscape [35]. The nodes which are highly connected with other nodes are called hub genes and they are mostly responsible for the disease to occur. Then to identify the complex part of the network, top modules/clusters of the PPI network were identified using MCODE (Molecular Complex Detection) method with basic cut-off criterion degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and maximum depth = 100 [36]. Then some significant module networks were illustrated by STRING where the minimum required interaction score was 0.4 which is a default in settings.

2.4. Identification of drug targets and analysis

We collected significant drug targets for the top 10 hub genes from DSigDB (Drug Signatures Database) using a web based tool called ‘Enrichr’ with hub genes as input genes. DSigDB is an online resource that relates drugs, small molecules, and target genes for GSEA (gene set enrichment analysis) [37].

3. Result

We collected significant drug targets for the top 10 hub genes from DSigDB (Drug Signatures Database) using a web based tool called ‘Enrichr’ with hub genes as input genes. DSigDB is an online resource that relates drugs, small molecules, and target genes for GSEA (gene set enrichment analysis) [37].

3.1. Gene collection

In this research analysis, all responsible genes for PC and LC were collected from the open source database DisGeNET. We collected a total of 2689 genes for PC and 4081 genes for LC. Then to identify the common genes between PC and LC, we used an online tool Venny to compare and identify the common genes. As a result we found 1686 common genes between LC and PC (Fig. 2).

3.2. Enrichment analysis

Enrichment analysis is covered by GO term analysis and KEGG pathway analysis. Gene Ontology (GO) enrichment analysis revealed that the common gene set of PC and LC associate with the regulation of the apoptotic process, regulation of cell proliferation, negative regulation of the apoptotic process, and positive regulation of intracellular signal transduction for BP, focal adhesion, chromatin, nuclear chromosome part and secretory granule lumen for CC, protein kinase activity, protein kinase binding, transcription regulatory region DNA binding, and protein tyrosine kinase activity for MF significantly (Table 1, Fig. 3). The KEGG pathway analysis revealed that the common genes of PC and LC associates with significant pathways involving MicroRNAs in cancer, Pathways in cancer, and Proteoglycans in cancer and Pancreatic cancer (Table 2, Fig. 3).

Table 1

<table>
<thead>
<tr>
<th>Category</th>
<th>GO ID</th>
<th>GO Term</th>
<th>Overlap</th>
<th>P- value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological Process</td>
<td>GO:0042981</td>
<td>regulation of apoptotic process</td>
<td>276</td>
<td>3.13E-99</td>
</tr>
<tr>
<td></td>
<td>GO:0042127</td>
<td>regulation of cell proliferation</td>
<td>246</td>
<td>5.13E-86</td>
</tr>
<tr>
<td></td>
<td>GO:0043066</td>
<td>negative regulation of apoptotic process</td>
<td>195</td>
<td>3.11E-84</td>
</tr>
<tr>
<td></td>
<td>GO:1902533</td>
<td>positive regulation of intracellular signal transduction</td>
<td>185</td>
<td>2.22E-76</td>
</tr>
<tr>
<td>Molecular Function</td>
<td>GO:0043069</td>
<td>negative regulation of programmed cell death protein kinase activity</td>
<td>163</td>
<td>1.01E-69</td>
</tr>
<tr>
<td></td>
<td>GO:0004672</td>
<td>protein kinase activity</td>
<td>149</td>
<td>2.14E-43</td>
</tr>
<tr>
<td></td>
<td>GO:0019901</td>
<td>protein kinase binding</td>
<td>144</td>
<td>4.93E-42</td>
</tr>
<tr>
<td></td>
<td>GO:0042412</td>
<td>transcription regulatory region DNA binding</td>
<td>116</td>
<td>6.64E-37</td>
</tr>
<tr>
<td></td>
<td>GO:0004713</td>
<td>protein tyrosine kinase activity</td>
<td>61</td>
<td>1.15E-27</td>
</tr>
<tr>
<td></td>
<td>GO:0003677</td>
<td>DNA binding</td>
<td>174</td>
<td>1.37E-26</td>
</tr>
<tr>
<td>Cellular Component</td>
<td>GO:0005925</td>
<td>focal adhesion</td>
<td>99</td>
<td>2.13E-27</td>
</tr>
<tr>
<td></td>
<td>GO:0000785</td>
<td>chromatin</td>
<td>77</td>
<td>1.16E-19</td>
</tr>
<tr>
<td></td>
<td>GO:0044454</td>
<td>nuclear chromosome part</td>
<td>90</td>
<td>7.39E-19</td>
</tr>
<tr>
<td></td>
<td>GO:0034774</td>
<td>secretory granule lumen</td>
<td>78</td>
<td>2.34E-18</td>
</tr>
<tr>
<td></td>
<td>GO:0000790</td>
<td>nuclear chromatin</td>
<td>65</td>
<td>1.64E-16</td>
</tr>
</tbody>
</table>
3.3. PPI network construction and module analysis

A PPI network was constructed with the common 1686 genes from PC and LC using STRING. For further analysis Cytoscape software tools like Network Analyzer, cytoHubba, and MCODE method were used on the PPI network. This PPI network has 1451 nodes and 51915 edges which indicates the connection between PC and LC (Fig. 4 A). The top 10 hub genes TP53, AKT1, EGFR, TNF, IL6, ALB, EGF, MYC, VEGFA, and GAPDH have been identified using the degree/connectivity method with the cytoHubba plugin tool (Fig. 4 B, Table 3). Then we performed cluster analysis using the MCODE method on the PPI network and found some significant module networks within the PPI network (Fig. 5).

3.4. Identification of drug targets and analysis

In this part of the study, we have identified drug targets for the top 10 hub genes from the DSigDB database. For the significant hub genes, we took 10 drug signatures where the p-value < 0.05 and overlap >5. From the significant 10 drug signatures ‘Capsaicin CTD 00005570’ was found to be associated with all the significant hub genes (TP53, AKT1, EGFR, TNF, IL6, ALB, EGF, MYC, VEGFA, and GAPDH). The drug signatures for significant hub genes were listed in Table 4.
4. Discussion

PC is often called ‘Silent Disease’ because it does not show symptoms that are noticeable before an early stage [38]. The symptoms it shows are almost similar to the symptoms of other medical problems like ulcer, gastric, and pancreatitis. PC is mostly diagnosed at an advanced stage of cancer spreading because PC is clinically silent at an early stage due to non-specific symptoms [38]. PC is one of the deadliest malignant neoplasms that spread to other organs like the lungs, liver, abdomen, bone, and even brain [39]. One common and main cause of both LC and PC is tobacco smoking [7,17]. PC can be categorized into four stages, stage1 (not spread, resectable or removable) when cancer is limited within the pancreas; stage2 (local spread, marginal removable) when cancer spread through the pancreas and adjacent lymph nodes; stage3 (spread widely and irremovable) when the cancer spreads to the pancreas completely and expanded to nearby blood vessels; stage4 (metastasized) when the cancer spread to the nearby and distant organ or body parts [40].

In this study, we have analyzed the dataset of PC and LC to investigate genetic links between these diseases and we found 1686 common genes which were 33.2% of both gene datasets of PC and LC which indicated strong interaction between PC and LC. We have performed GO term analysis and KEGG pathway analysis with the help of Enrichr using the common genes. Enrichment analysis revealed that the common gene set significantly associates with the regulation of the apoptotic process, regulation of cell proliferation, negative regulation of the apoptotic process, and positive regulation of intracellular signal transduction for BP, focal adhesion, chromatin, nuclear chromosome part and secretory granule lumen for CC, protein kinase activity, protein kinase binding, transcription regulatory region DNA binding, and protein tyrosine kinase activity for MF. The KEGG pathway analysis revealed genes associated with a significant pathway involving MicroRNAs in cancer, Pathways in cancer, Proteoglycans in cancer, PI3K-Akt signaling pathway, Prostate cancer, AGE-RAGE signaling pathway in diabetic complications, Pancreatic cancer, etc. MicroRNAs are short RNA molecules that target several cellular regulators and inhibit their expression [41]. MicroRNA profiling has been found in PC patient’s blood and tumor cells. MicroRNA plays an important role in the initiation, development, progression, and metastasis of PC [42]. Pathways in cancer involve Wnt/Notch signaling, TGF-b signaling, regulation of invasion, KRAS signaling, etc. was found in the pancreas cancer regulatory process or pathway [43]. Proteoglycans perform multiple functions in cancer which are key molecular effectors of the cell surface. Proteoglycans such as glypicans and syndecans can directly impact the development of cancer by regulating key signaling pathways [44]. Heparan sulfate proteoglycans (HSPGs) glypican-1 regulates the action of growth factors in pancreatic cancer cells and tumors [45]. PI3K-Akt signaling pathway is atypically activated in several cancers. Receptor tyrosine kinases and somatic mutations in particular signaling pathway elements trigger two of the most widely discovered PI3K-Akt activation mechanisms in human cancers [46]. The inhibition of the PI3K-Akt signaling pathway is involved with PC tumor suppression [47]. The AGE-RAGE signaling influences apoptosis and autophagy in numerous cancer cells. The AGE-RAGE system plays a vital role in the development of various types of cancers.
of cancers like PC, LC, and other pathological diseases including diabetes and cardiovascular diseases [48].

After employing PPI network of the common genes, we have identified significant hub genes (TP53, AKT1, EGFR, TNF, IL6, ALB, MYC, VEGFA, and GAPDH) based on their connectivity/degree value. TP53 gives instructions for making a tumor protein called p53. PDAC (Pancreatic ductal adenocarcinoma) is a cancer type where mutant TP53 affects disease progression. Mutant TP53 promotes invasion and metastasis and it is mutated in 75% of cases in PDAC [49]. AKT is a serine protein kinase that plays a role in the majority of human cancers by inhibiting apoptosis and mediating cell proliferation. Isoforms of AKT are frequently activated in PC. The hyper-activation of AKT1 has been involved in the formation of PC that’s why it is suggested that through inhibiting AKT1 there may be a therapeutic possibility to cure pancreatic cancer [50]. EGF (Epidermal growth factor) is a multifunctional cytokine that plays significant roles in different cellular events like cell survival, proliferation, differentiation, and dissolution [52]. TNF-α is a crucial member of the TNF-α superfamily that plays roles in immunity, cellular remodeling, cell survival, and apoptosis [53]. TNF-α plays an important role in the inflammatory etiology of PC. It was found that TNF-α supports PC cell growth through EGFR and TGF-α (Transforming growth factor) expression [54]. TNF-α (308 A) allele promotes LC development and progression [55]. IL6 (Interleukin-6) is a pro-inflammatory cytokine that is elevated in particular with PC. IL6 partially promotes metastatic invasion by activating pro-invasive GTPase signaling, priming tumor cells for metastatic invasion [56]. IL6 Trans signaling activates Stat3 in the pancreas that promotes PDAC development and PanIN progression [57].

After drug target analysis we have found one drug signature ‘Capsaicin CTD 00005570’ for the common genes of PC and LC that are associated and connected with all the hub genes. Capsaicin is the main pungent component of red-chili pepper which induces apoptosis in pancreatic cancer cells. Capsaicin is an effective inhibitor of pancreatic cancer cells in vitro and in vivo growth [58]. Capsaicin displays powerful antineoplastic activity in a large variety of human cancer cells and it induces apoptosis in human SCLC through the TRPV6 receptor and the Calpain pathway [59].

5. Conclusion

In this study, we have analyzed the respective gene dataset of PC and LC to identify common genes between these two cancer diseases. We integrated these common genes for GO terms analysis, pathway enrichment analysis and found the molecular functions, cellular components, biological processes, and significant pathways of the common genes. We constructed a PPI network with the common genes and identified the top 10 significant hub genes for both PC and LC including TP53, AKT1, GAPDH, EGFR, MYC, VEGFA, IL6, ALB, TNF, and EGF.
which play a significant role in these two cancer disease progressions. Then we identified significant modules from the PPI network. Furthermore, we have identified candidate drug compounds for the significant pathways, and drug signatures associated with both PC and LC, which can be helpful in the treatment of patients suffering from both PC and LC cancers.

Funding
Conflict of interest: None declared.

Declaration of competing interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement
Our deepest gratitude to the editor and anonymous reviewers whose thoughtful suggestions helped us to improve this paper.

References


[26] GSK 938-94.

[27] T. Khan et al. 


