Bioinformatic evidences and analysis of putative biomarkers in pancreatic ductal adenocarcinoma

Yuan Gu [a,1], Qijin Feng [a,1], Han Liu [b], Qi Zhou [c], Ailing Hu [d], Takuji Yamaguchi [d], Shilin Xia [d,e,*], Hiroyuki Kobayashi [a,**]

[a] Center for Advanced Kampo Medicine and Clinical Research, Graduate School of Medicine, Juntendo University, Tokyo, Japan
[b] Department of Oral Pathology, Dalian Medical University, Dalian, PR China
[c] Institute (College) of Integrative Medicine, Dalian Medical University, Dalian, PR China
[d] Department of Palliative Medicine, Graduate School of Medicine, Juntendo University, Tokyo, Japan
[e] Clinical Laboratory of Integrative Medicine, The First Affiliated Hospital of Dalian Medical University, Dalian, PR China

ARTICLE INFO

ABSTRACT

Background: Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal human cancers. Aberrant expression of genes plays important role in the procession of PDAC. The analysis of gene expression profile will contribute to the research of carcinoma mechanism.

Objective: This present study is focused to investigate the differentially expressed genes (DEGs) from 3 PDAC microarray datasets, which would provide candidate genes for putative biomarkers to understand the mechanism of PDAC and potential targets of treatment.

Method: Based on the overlap genes obtained from 3 GEO datasets, the hub genes were identified using STRING and Cytoscape plugin MCODE. The enrichment and function analysis were applied using DAVID. The protein-protein interaction network was performed using cBioPortal and UCSC Xena. The Oncomine was finally used to determine the candidate gene by analyzing their expression between pancreas sample and PDAC sample.

Results: 25 hub genes were selected from a total of 1006 DEGs from 3 GEO datasets, consisting of 14 upregulated genes and 11 downregulated genes. The overall decline of hub gene expression enriched in G1 phase of cell cycle in other subtypes of pancreatic cancer. Oncomine database was ultimately performed to determine the 8 candidate genes, including CXCL5, CCL20, NMU, F2R, ANXA1, EDNRA, LPAR6, and GNA15.

Conclusions: Conclusively, 8 candidate genes would become the potential PDAC combined biomarkers for diagnosis and therapeutic strategies.

1. Introduction

Pancreatic ductal adenocarcinoma (PDAC) is frequently prescribed as one of the most lethal malignancies, which remains the most difficult to treat carcinoma [1, 2]. It is generally known that aberrant activation and/or inactivation of genes are involved in the carcinogenesis and development of cancer. The previous research on PDAC suggests that abnormal expression of genes play an important role on the occurrence and progression of this neoplasm [3, 4]. Studies over past decades have provided molecular information on PDAC, such as KRAS, PI3K, PTEN, mTOR and signal pathways involved in apoptotic signal, cell cycle regulation signal, cell adhesion pathway and so on [5, 6, 7, 8]. However, it is still hard to interpret the pathological mechanism of PDAC, which leads to absence of effective drug and high medical cost. This paper attempts to provide insights into the exploration of PDAC biomarkers via various databases and bioinformatical approaches.

With rapid development of technology, massive data is generated from high throughout detection, which is widely used to analyze key molecules in various biological progress of neoplasm [9]. Microarray technology is designed to screen the candidate gene, which gradually becomes the foundation of diagnosis and treatment [10, 11]. The information of microarray, especially distinct datasets of carcinoma subtype,

---

is not fully analyzed.

In this paper, three PDAC datasets were firstly reviewed in GEO database. We further sought to obtain the differentially expressed genes (DEGs) using GEO2R, which was a bioinformatical method based on R Language. After we gained 1006 overlap genes among three datasets, the enrichment analysis of biological function and pathway was utilized with Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Subsequently, a holistic approach of protein-protein interaction network was performed with Search Tool for the Retrieval of Interacting Genes (STRING) and Cytoscape. The total of 25 hub genes were finally preferred to analysis from hub genes with cBioPortal and UCSC Xena. To consolidate our findings, the hub gene was input into Oncomine database to determine the most potential gene named candidate gene. In conclusion, the analysis work presented here provides that 8 candidate genes were identified from 25 hub genes, including CXCL5, CCL20, NMU, F2R, ANXA1, EDNRA, LPAR6, and GNA15. Our findings of candidate gene would contribute to the exploration of PDAC biomarker.

2. Materials and methods

2.1. The analysis of microarry data

Three gene expression datasets (GSE101448 [12], GSE91035, GSE71989) were analyzed and downloaded from GEO (http://www.ncbi.nlm.nih.gov/geo) [13], which is public functional genomics data base of high throughout gene expression pool, especially the data of chips and microarrays. The GSE101448 dataset contained 43 tissue including 24 PDAC samples and 19 normal samples. The GSE91035 dataset contained 48 tissue including 25 PDAC samples and 23 normal samples. The GSE71989 dataset contained 21 tissue including 13 PDAC samples and 8 normal samples.

2.2. Identification of DEGs

The DEGs was analyzed by GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r), which is the web tool of GEO platform to compare datasets so that researchers can identify the DEGs under dataset. In the option of GEO2R, Benjamin & Hochberg (False discovery rate) was applied for adjustment to the P-values. The log transformation was applied with Auto-detect. The Category of Platform annotation to display on results was "NCBI generated" (not Submitter supplied). The fold change value (logFC) and adjusted P values (adj. P) were performed to discover the statistically significant genes. The absolute value of fold change$\geq$1.5 and

adj. P value $\leq$ 0.01 were considered statistically significant. Point without gene symbols was removed. The Venn diagram was used to generate overlap results of different gene lists as a graphical output.

2.3. KEGG and GO enrichment analysis of DEGs

The GO and KEGG information of DEGs were achieved from the online cancer biological data repository DAVID (https://david.ncifcrf.gov/home.jsp, version 6.8) [14]. KEGG is a database resource for understanding high-level functions and utilities of the biological system [15]. GO is a major bioinformatics web tool for annotating genes and analyzing biological process from the molecular to the organism level network construction and module analysis [16]. The biological analyses of DEGs was applied using DAVID. We chose TCGA Pancreatic Cancer (PAAD) including 196 samples. The Oncomine databsets used in this study were included Badea Pancreas, Logsdon Pancreas, Ishikawa Pancreas, Iacobuzio-Donahue Pancreas, Grutzmann Pancreas, and Buchholz Pancreas.

2.4. STRING analysis of network

The STRING (http://string-db.org) (version 11.0) is an online database to analyze the functional network between two or multiple proteins, which provides insights into protein-protein interaction and predicts the mechanisms of development of PDAC in present study [17]. The score = 0.900 was considered statistically significant. Protein-protein interaction of DEGs was also analyzed using Cytoscape software (version 3.7.1). Cytoscape is an open source bioinformatics platform which was constructed to apply for the visualization of protein-protein interaction network [18]. The analysis of this network was applied using Cytoscape with Molecular Complex Detection (MCODE) (version 1.5.1). MCODE is a Cytoscape App and this version 1.5.1 was designed by Bader Lab in University of Toronto. This Cytoscape plugin was used to find clusters in a protein-protein interaction network [19]. The configuration of MCODE was as follows: Degree Cutoff = 2, Node score Cutoff = 0.2, K-Core = 2, Max. Depth = 100.

2.5. Hub gene analysis

The network and co-expression of hub genes were performed using cBioPortal (http://www.cbioportal.org) [20]. The cBioPortal provides visualization and analysis of cancer genomics data. Cluster analysis of hub genes was applied on the web of UCSC Xena (https://xenabrowser.net) [21]. Analysis type: cancer vs normal analysis. Cancer type:
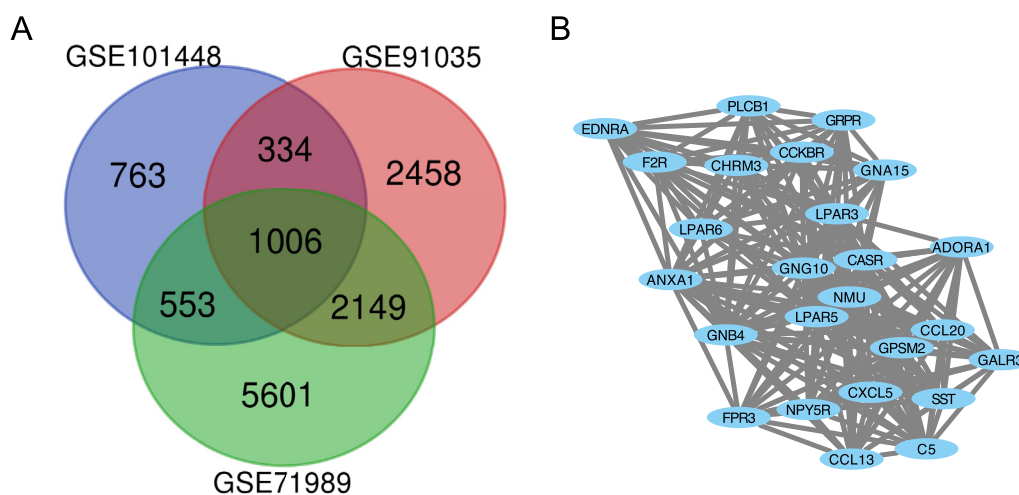

Fig. 1. Venn diagram and the most significant module of DEGs. (A) DEGs were selected with the absolute value of fold change$\geq$1.5 and adj. P value $\leq$ 0.01 among the mRNA expression profiling datasets GSE101448, GSE91035 and GSE71989. The 3 datasets showed an overlap of 1006 genes. (B) The most significant module was obtained from Cytoscape plugin MCODE.

Pancreatic ductal adenocarcinoma. The candidate genes were selected from hub genes using Oncomine (http://www.oncomine.com), [22], which is an online database with performance of unified searches. The lowest P value and highest fold change of dataset were the limitations for analyzing the differential analysis between normal and carcinoma data.

2.6. Statistics analysis

GraghPad Prism 8.0 was used for statistics analysis. The Mann-Whitney test was assessed for expression of candidate gene in PDAC compared to that in Pancreas.

3. Results

3.1. Identification of DEGs in HCC

The results of DEGs were downloaded and standarded from GEO

dataset with adj.P value and logFC. The total of 3788 in GSE101448, 14188 in GSE91035, 28990 in GSE71989 were identified. The Venn Diagram was used to display the 1006 overlap genes among the three GSE datasets between normal sample and PDAC sample (Fig. 1 A).

3.2. GO and KEGG enrichment analyses of DEGs

After identifying the overlap genes, the functional annotation of DAVID was performed to analysis GO and KEGG from DEGs with uploading of overlap gene symbol. The biological processes (BP), molecular function (MF) and cell component (CC) of DEGs were illustrated as GO analysis. The change of BP was mainly enriched in cell adhesion, extracellular matrix organization, and signal transduction (Supplementary Table 1). The change of CC was apparently enriched in extracellular space, extracellular exosome, and extracellular matrix (Supplementary Table 2). The MF was significantly in protein binding, receptor binding, and calcium ion binding (Supplementary Table 3). The change of KEGG
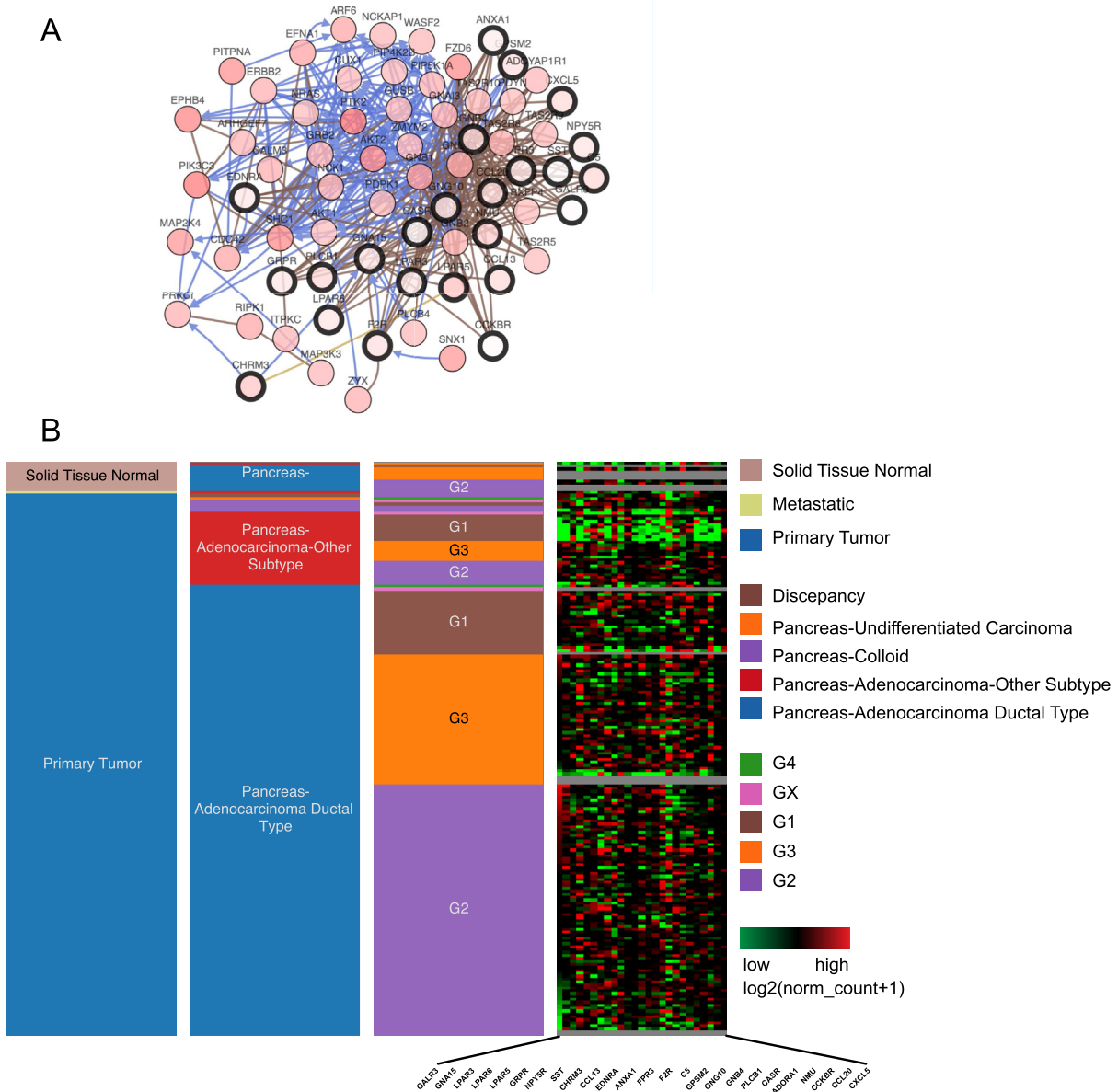

Fig. 2. Interaction network and biological process analysis of 25 hub genes. (A) Hub genes and their co-expression genes were analyzed using cBioPortal. Nodes with bold black outline represent hub genes. Nodes with thin black outline represent the co-expression genes. (B) Hierarchical clustering of 25 hub genes was performed using UCSC Xena. In the first column graph on the right, upregulation of genes is marked in red; downregulation of genes is marked in green. In the second column graph on the right, there are three phases of cycles, including G1 (brown), G2 (purple), and G3 (yellow).

was mainly enriched in pathways in pancreatic secretion, protein digestion and absorption, and focal adhesion (Supplementary Table 4).

3.3. Protein-protein interaction analysis

The total of 1006 DEGs were selected from the GEO database. After STRING analysis, DEGs were analyzed and visualized with the Cytoscape software. There were several PPI prediction clusters. Finally, we selected the first cluster of 25 genes from the list of 26 PPI clusters using Cytoscape plugin MCODE (Fig. 1 B). There are 14 upregulated genes and 11 downregulated genes. The names, abbreviations, regulation direction and functions of these 25 hub genes were shown in Table 1, with 14 upregulated genes including GNA15, LPAR5, CCL13, CXCL5, CCL20, GPSM2, NMU, FPR3, GNG10, EDNRA, F2R, ANXA1, GNB4 and LPAR6, and 11 downregulated genes including NPY5R, GALR3, LPAR3, PLCB1, ADORA1, CASR, C5, CHRM3, SST, GRPR and CCKBR.

3.4. Hub gene analysis

The network and co-expression of 25 hub genes was revealed and visualized using cBioPortal platform (Fig. 2 A). The cluster analysis was applied for differentiating the genes among subtypes and grades (Fig. 2 B). These results, as shown in Fig. 2B, indicate that the expression of hub genes in PDAC seemed higher than that in other subtypes of pancreatic cancer. The most striking observation was that low expression section mainly enriched in G1 phase of the Pancreas Adenocarcinoma Other Subtype, not G2 or G3 (Fig. 2 B). Subsequently, it was considered with low P value and high fold change to select 8 candidate genes with at least 3 PDAC datasets (Fig. 3). These candidate genes included CXCL5, CCL20, NMU, F2R, ANXA1, EDNRA, LPAR6, and GNA15. The 8 candidate genes were selected and analyzed the differential expression of pancreas compared with that of PDAC (Fig. 4). The most fold change of each candidate gene was as follows: CXCL5: P = 3.37E-13, fold change = 13.978. CCL20: P = 0.147, fold change = 1.670. NMU: P = 9.95E-4, fold change = 6.394. F2R: P = 7.08E-13, fold change = 5.095. ANXA1: P = 3.33E-10, fold change = 4.489. EDNRA: P = 4.87E-16, fold change = 3.711. LPAR6: P = 1.64E-9, fold change = 3.059. GNA15: P = 3.09E-9, fold change = 2.212.

4. Discussions

Pancreas carcinoma is one of the most malignant tumors. PDAC, the

pancreatic adenocarcinoma ductal subtype, carries the five-year survival rate of 8%, [23], which regards PDAC as one of the most leading cause of cancer-related deaths worldwide. Prior studies have noted the importance of major mutated genes in the PDAC research [24]. Abnormal expression level of genes is being investigated, thus the underlying mechanism of pancreas carcinoma, especially PDAC, will be elaborated. However, the molecular mechanisms of PDAC remain poorly understood. This study set out with the aim of determining hub genes of PDAC and analyzing the protein-protein interaction network based on the enrichment analysis. Then, the candidate genes selected from hub genes were finally identified with Oncomine analysis. This study would be conducive to molecular research of PDAC and pharmaceutical advantage in clinical treatment of PDAC.

The current study found the DEGs between normal samples and PDAC samples from 3 microarray datasets. The total of DEGs obtained from overlap genes is 1006 genes among these datasets. During the initial stage of this research, we actually found 5 microarray datasets in GEO including GSE101448, GSE91035, GSE71989, GSE101462 and GSE77858. There was no gene of GSE101462 and GSE77858 datasets with adjusted P-values of 0.01 (data not shown). The adjusted P-values combined with fold change is important for screening DEGs. Therefore, the DEGs were finally screened from these 3 datasets. According to the absolute value of fold change≥1.5 and adj. P value ≤ 0.01, the analysis of 3 GES datasets provided 1006 genes. This enormous overlap data could be attributed to consistency of distinct datasets, with the reliability of these 3-research works. After MCODE analysis of Cytoscape software, we found a total of 25 hub genes among overlap data. There are 14 upregulated genes and 11 downregulated genes. These hub genes with their interaction network would be potential combined biomarkers with relevant treatment of PDAC.

We applied for the UCSC Xena to show the map of 25 hub genes expression under 3 conditions. There was no different between normal tissue and pancreas primary tumor. However, it was found that there was an overall decline in the expression of hub genes in other subtype of pancreas adenocarcinoma than that in PDAC. With the further analysis of tumor grade, it was apparent that the hub gene expression had a process of decrease in low grade of pancreas adenocarcinoma other subtype not grades of PDAC. It was suggested that the hub genes play distinct roles among different pancreas adenocarcinoma subtypes. Therefore, there should be further objective research of PDAC not pan-pancreas carcinoma.

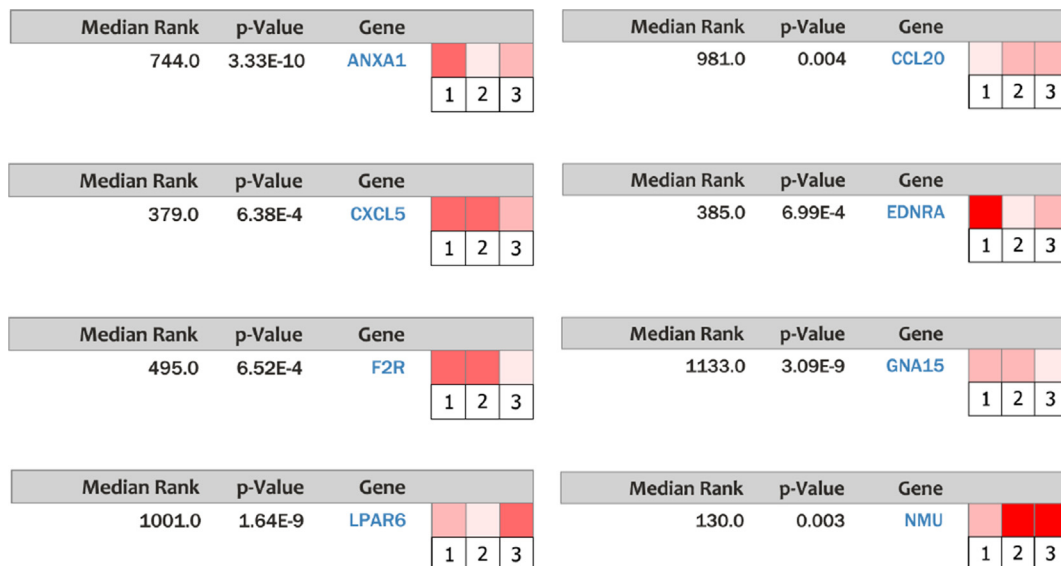To validate the role of hub gene in PDAC, the 25 hub genes were

Fig. 3. Analysis of PDAC vs. normal tissue of 8 candidate genes from 3 datasets in Ocomine. Heat maps of 8 candidate genes expression in PDAC samples vs. normal samples.
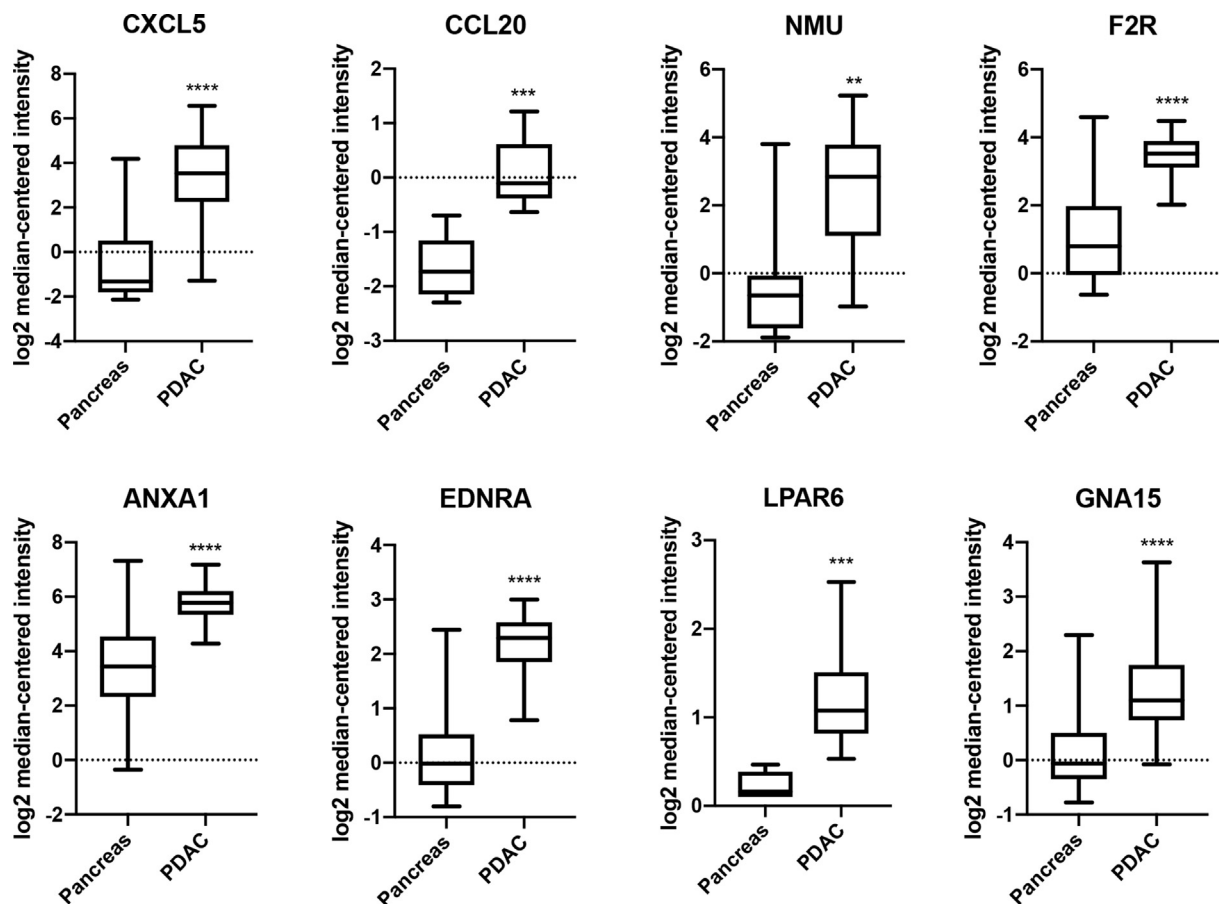
Fig. 4. Expression analysis of PDAC vs. normal tissue of 8 candidate genes in the most significant dataset. mRNA expression of candidate genes in PDAC is compared to that in pancreas. $**P < 0.01$, $***P < 0.001$, and $****P < 0.0001$.

filtered in different datasets in Oncomine. Among these hub genes, 8 candidate genes were determined with significant over-expression in PDAC in different datasets of Oncomine database. The Oncomine result consolidated the administration of hub gene screen. Meanwhile, this analysis was available for exploration of more concrete targets. Combined with Oncomine analysis, it would provide more credible and meaningful data for the novel biomarker. It was somewhat surprising that all candidate genes were upregulated in the PDAC, which directly prompted the development of carcinoma.

Among the candidate genes, there are 5 genes associates with the function of guanine-nucleotide-binding proteins (G-protein), as follows: GNA15, CXCL5, EDNRA, F2R and LPAR6. This result is in accord with recent studies indicating that G-protein coupled receptor family involved in growth and metastasis of some types of tumors [25, 26]. CXCL5, YAP/TEAD-regulated genes, has been reported to be associated with unfavorable survival of PDAC patients [27]. F2R was also revealed as hub gene in a microarray analysis of insulinoma [28]. In a previous research, EDNRA was found to upregulate in pancreatic neuroendocrine tumors (PanNET) [29]. GNA15 was not expressed in normal neuroendocrine cells but was overexpressed in gastroenteropancreatic neuroendocrine neoplasia (GEP-NEN) cell lines [30]. There are still limit findings of these candidate genes in PDAC research. The findings reported here still need to be investigated with in vivo and in vitro experiments to verify the functional role of candidate genes.

5. Conclusions

The evidence from this study suggested that the candidate gene might be exploited as a biomarker for diagnosis and therapeutic selection. We found that 5 of 8 candidate genes were related with G-protein signal,

which would contribute to research and development of new drug, such as plozalizumab. This study was one of the first attempts to enable bioinformatic data to prospective drug research for malignant pancreatic cancer.

In conclusion, the aim of the present research was to identify the DEGs which were potentially involved in the progression of PDAC. A total of 1006 DEGs and 25 hub genes were selected and analyzed with bioinformatic tools. The 8 candidate genes were finally determined with datasets from Oncomine, which would be regarded as combined biomarkers for PDAC. A further study could assess the biological function and interaction network of these candidate genes in order to elucidate the systemic mechanism of PDAC.

Declarations

Author contribution statement

Shilin Xia: Conceived and designed the experiments; Wrote the paper.
Hiroyuki Kobayashi: Conceived and designed the experiment.
Yuan Gu: Performed the experiments; Wrote the paper.
Han Liu: Performed the experiments; Contributed reagents, materials, analysis tools or data.
Qijin Feng: Analyzed and interpreted the data; Wrote the paper.
Qi Zhou, Ailing Hu: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.
Takuji Yamaguchi: Analyzed and interpreted the data.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Data associated with this study has been deposited at https://www.ncbi.nlm.nih.gov/geo/ under the accession numbers GSE101448, GSE91035, GSE71989.

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2019.e02378.

References

[1] S.R. Chandana, H.M. Babiker, D. Mahadevan, Therapeutic trends in pancreatic ductal adenocarcinoma (PDAC), Expert Opin. Investig. Drugs 28 (2019) 161–177.

[2] D.S. Longnecker, Pancreatic Ductal Adenocarcinoma: recent updates, Am. J. Pathol. 189 (2019) 6–8.

[3] A.A. Tesfaye, A.S. Azmi, P.A. Philip, miRNA and gene expression in Pancreatic Ductal Adenocarcinoma, Am. J. Pathol. 189 (2019) 58–70.

[4] L. Zhao, H. Zhao, H. Yan, Gene expression profiling of 1200 pancreatic ductal adenocarcinoma reveals novel subtypes, BMC Canc. 18 (2018) 603.

[5] D. Hashimoto, K. Arima, N. Yokoyama, et al., Heterogeneity of KRAS mutations in Pancreatic Ductal Adenocarcinoma, Pancreas 45 (2016) 1111–1114.

[6] A.L. Windon, A. Loaiza-Bonilla, C.E. Jensen, et al., A KRAS wild type mutational status confers a survival advantage in pancreatic ductal adenocarcinoma, J. Gastrointest. Oncol. 9 (2018) 1–10.

[7] H. Ying, K.G. Elpek, A. Vinjamoori, et al., PTEN is a major tumor suppressor in pancreatic ductal adenocarcinoma and regulates an NF-kappaB-cytokine network, Cancer Discov. 1 (2011) 158–169.

[8] S. Iriana, S. Ahmed, J. Gong, et al., Targeting mTOR in pancreatic ductal adenocarcinoma, Front. Oncol. 6 (2016) 99.

[9] R. Alanni, J. Hou, H. Azzawi, et al., A novel gene selection algorithm for cancer classification using microarray datasets, BMC Med. Genom. 12 (2019) 10.

[10] M. Daoud, M. Mayo, A survey of neural network-based cancer prediction models from microarray data, Artif. Intell. Med. (2019).

[11] N.C.M. Visser, A.A.M. van der Wurff, J.M.A. Pijnenborg, et al., Tissue microarray is suitable for scientific biomarkers studies in endometrial cancer, Virchows Arch. 472 (2018) 407–413.

[12] H. Klett, H. Fuellgraf, E. Levit-Zerdoun, et al., Identification and validation of a diagnostic and prognostic multi-gene biomarker panel for pancreatic ductal adenocarcinoma, Front. Genet. 9 (2018) 108.

[13] R. Edgar, M. Domrachev, A.E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, Nucleic Acids Res. 30 (2002) 207–210.

[14] D.W. Huang, B.T. Sherman, Q. Tan, et al., The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists, Genome Biol. 8 (2007) R183.

[15] M. Tanabe, M. Kanehisa, Using the KEGG database resource, Curr. Protoc. Bioinform. (2012). Chapter 1: Unit1 12.

[16] H. Al-Mubaid, Gene multifunctionality scoring using gene ontology, J. Bioinform. Comput. Biol. 16 (2018) 1840018.

[17] D. Szklarczyk, A.L. Gable, D. Lyon, et al., STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res. 47 (2019) D607–D613.

[18] N.T. Doncheva, J.H. Morris, J. Gorodkin, et al., Cytoscape StringApp: network analysis and visualization of proteomics data, J. Proteome Res. 18 (2019) 623–632.

[19] W.P. Bandettini, P. Kellman, C. Mancini, et al., MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study, J. Cardiovasc. Magn. Reson. 14 (2012) 83.

[20] J. Gao, B.A. Aksoy, U. Dogrusoz, et al., Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, Sci. Signal. 6 (2013) pl1.

[21] M. Haeussler, A.S. Zweig, C. Tyner, et al., The UCSC Genome Browser database: 2019 update, Nucleic Acids Res. 47 (2019) D853–D858.

[22] L. Badea, V. Herlea, S.O. Dima, et al., Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia, Hepato-Gastroenterology 55 (2008) 2016–2027.

[23] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2017, CA A Cancer J. Clin. 67 (2017) 7–30.

[24] B. Zhou, A. Irwanto, Y.M. Guo, et al., Exome sequencing and digital PCR analyses reveal novel mutated genes related to the metastasis of pancreatic ductal adenocarcinoma, Cancer Biol. Ther. 13 (2012) 871–879.

[25] D. Wang, L. Hu, G. Zhang, et al., G protein-coupled receptor 30 in tumor development, Endocrine 38 (2010) 29–37.

[26] Q. Liu, Z. Chen, G. Jiang, et al., Epigenetic down regulation of G protein-coupled estrogen receptor (GPER) functions as a tumor suppressor in colorectal cancer, Mol. Cancer 16 (2017) 87.

[27] E. Rozengurt, J. Sinnett-Smith, G. Eibl, Yes-associated protein (YAP) in pancreatic cancer: at the epicenter of a targetable signaling network associated with patient survival, Signal Transduct. Target Ther. 3 (2018) 11.

[28] W. Zhou, L. Gong, X. Li, et al., Screening key candidate genes and pathways involved in insulinoma by microarray analysis, Medicine (Baltim.) 97 (2018), e10826.

[29] D. Speisky, A. Duces, I. Bieche, et al., Molecular profiling of pancreatic neuroendocrine tumors in sporadic and Von Hippel-Lindau patients, Clin. Cancer Res. 18 (2012) 2838–2849.

[30] S. Zanini, F. Giovinazzo, D. Alaimo, et al., GNA15 expression in small intestinal neuroendocrine neoplasia: functional and signalling pathway analyses, Cell. Signal. 27 (2015) 899–907.