Abstract

Aim: This study aims to identify the prognostic and diagnostic significance of protein kinase-coding genes in pancreatic ductal adenocarcinoma (PDAC), products of which constitute one of the main classes of drug targets in cancer treatment.

Material and Methods: Whole-genome gene expression data from seven PDAC cohorts (GSE62452, GSE15471, GSE62165, GSE18670, GSE19280, GSE41368, GSE71989) were included in the integrative transcriptomic analysis (n tumor=252, ncontrol=131). The differentially expressed genes in PDAC compared to controls were identified using random-effects model and were further validated in TCGA (The Cancer Genome Atlas) combined GTEx (Genotype-Tissue Expression) cohort (n tumor=179, ncontrol=171). The prognostic significance of the identified genes was then evaluated by integrating survival and transcriptome data of over 530 (n=530-1302) patients using OSpaad.

Results: The integrative transcriptomic analysis revealed a total of seven down-regulated and 33 up-regulated protein kinase-coding genes in PDAC (adjusted p-value≤0.05, -2≤z-value≤2). The validation analysis using TCGA combined GTEx data confirmed 80% (n=32) of the identified differentially expressed genes in PDAC (p<0.01, and fold change≥2). Amongst, the elevated mRNA expressions of 9 genes (PTK2, TAOK1, CSNK1A1, EIF2AK2, WNK1, CDK12, CDK6, GSK3B, and MAP4K4) were found to be significantly correlated with worse overall survival of patients with PDAC (Logrank p≤0.05, HR>1). Overexpression of SYK and PRKACB were correlated with better overall survival (Logrank p<0.05, HR<1).

Discussion: The results of this study suggest that mRNA expression of the identified eleven protein kinase-coding genes can be used as both prognostic and diagnostic biomarkers for further clinical validation.

Keywords
Pancreatic neoplasms; Protein kinases, Transcriptome
Protein kinase-coding genes in PDAC

Introduction
Protein kinases are proteins that belong to the transferases class of enzymes (EC:2). The human genome has been shown to encode for 518 protein kinases [1]. These enzymes regulate the biological activity of various proteins generally in response to an intracellular or external stimulus by transferring a gamma phosphate group of ATP to serine, threonine, tyrosine, arginine or histidine residues of proteins [2]. Protein kinases are classified by the amino acids they phosphorylate. The two main classes of protein kinases are tyrosine kinases (EC:2.7.10) and serine/threonine kinases (EC:2.7.11). Protein phosphorylation is one of the most frequent and important post-translational modifications and mechanisms of regulation of crucial cellular processes such as proliferation, signal transduction, apoptosis, among others [3-5]. It has been shown that protein kinases are frequently altered, especially in human malignancies [6]. These alterations include genetic aberrations including deregulated expression, amplification, chromosomal translocation, mutation, and epigenetic modifications [6]. Therefore, due to these alterations' high oncogenic potential, protein kinases represent one of the most targeted groups of enzymes in the treatment of diverse types of cancer.

Pancreatic cancer is one of the deadliest malignancies and the fourth common cause of death from cancer [7]. The overall 5-year survival rate for patients with pancreatic cancer remains less than 8%, and a one-year survival is around 18% when all stages are combined [7,8]. Pancreatic ductal adenocarcinoma (PDAC) is the most common form of pancreatic cancer constituting more than 90% of cases with pancreatic cancer [9]. Delayed diagnosis due to the absence of screening methods, and frequent recurrence owing to highly metastatic and chemoresistant nature of pancreatic cancer are the major effectors of its dismal prognosis. Thus, the identification of diagnostic or prognostic biomarkers and targets for effective therapies is an urgent need in the treatment of the disease.

Using an integrative transcriptomic analysis approach, this study aims to identify differentially expressed protein kinase-coding genes in pancreatic ductal adenocarcinoma and their potential to be diagnostic and prognostic biomarkers for PDAC.

Material and Methods
Selection of Gene Expression Omnibus Datasets and Assessment of Data Quality
Gene Expression Omnibus (GEO) was searched for datasets including mRNA expression data from human-derived pancreatic ductal adenocarcinoma tissues and healthy or adjacent-to-tumor pancreatic tissues. Eligible datasets were subjected to data quality assessment using ExAtlas software [10]. As indicated in the manual of ExAtlas, the correlation of expression of housekeeping genes in the range from 0.5 to 0.95 and the level of standard deviation from the global mean for each set of genes grouped by the average expression is less than 0.3 were considered good quality. Samples of low quality were not included in the analysis.

The Integrative Transcriptomic Analysis
The integrative gene expression analysis was performed by combining gene expression datasets from different studies and platforms using ImaGeo Software [11]. Effect size was selected as a meta-analysis metric and a random effects model was used for effect size estimation. Samples with more than 1% of missing values were ignored during the analysis. Z-values were used to measure differential expression between PDAC and control tissues. The adjusted p-value less than 0.05 and z-value greater than 2 or less than -2 were considered significant.

Determination of the differentially expressed protein kinase-coding genes
The enzyme classes to which the identified differentially expressed genes are assigned were searched manually in KEGG (Kyoto Encyclopedia of Genes and Genomes). The differentially expressed genes encoding protein kinases belong to ‘Transferring phosphorus-containing groups’ (kinases, EC:2.7.) were selected for further analyses.

External Validation of the Identified Differentially Expressed Genes in PDAC
The external validation of the identified differential expression of protein kinase-coding genes in PDAC was generated using GEPIA Database by comparing transcriptomic data from the TCGA PAAD (n=179), and the TCGA normal and GTEx data (n=171) [12]. The method for differential analysis was one-way analysis of variance (ANOVA), using disease state (Tumor or Normal) as a variable for calculating differential expression. P<0.01 and fold changes≥2 was accepted as a statistically significant difference in gene expression.

Prognostic Significance of the identified DEGs in PDAC
The possible association between the mRNA expression level of the identified differentially expressed protein kinase-coding genes in PDAC and overall survival was evaluated using combined long-term follow-up and transcriptomic data of over 530 (n=530-1302) pancreatic carcinoma patients from seven patient cohorts (TCGA PAAD, ICGC_Array, GSE28735, ICGC_Seq, GSE62452, GSE71729, and EMTAB6134). OSpaad was used to generate Kaplan-Meier survival curves, calculate the hazard ratio (HR) with 95% confidence intervals, and log-rank p-value [13]. The ‘upper 50%’ option was selected as the cutoff to split the patient cohort. Log-rank p-value <0.05 was considered statistically significant.

Results
GEO Datasets included in the integrative transcriptomic analysis
A total of ten GEO gene expression microarray datasets including transcriptomic data from 271 PDAC tissues and 150 healthy and adjacent-to-tumor non-tumor pancreatic tissue samples were found to be eligible for the analysis. Three data sets including data from 17 PDAC tissues and 18 control tissues did not pass the quality control and therefore were excluded from the analysis. Three samples from the remaining datasets were not included in the analysis due to poor quality. Consequently, transcriptomic data of 252 PDAC tissues and 131 control samples from seven GEO datasets were included in the study. Healthy pancreatic tissues and adjacent non-tumor pancreatic tissues were used as a control group. The GEO datasets included in the integrative transcriptomic analysis are shown in Table 1.
Protein kinase-coding genes in PDAC

Differentially Expressed Genes in PDAC Compared to Controls
A total of 278 genes were found to be down-regulated and 959 genes were found to be up-regulated in PDAC compared to control tissues (adjusted p-values<0.05, -2≤z-values≤2). The heatmap of the identified top 100 differentially expressed genes in PDAC is shown in Figure 1.

Differentially Expressed Protein kinase-coding Genes in PDAC
A total of forty protein kinase-coding genes were found to be significantly altered in PDAC. The most significantly altered protein kinase-coding genes in PDAC were found to belong to the protein-serine / threonine kinases (EC:2.7.11., n=35). Five genes (TK2, JAK1, JAK3, PTK2, and SYK) coding protein-tyrosine kinases (EC:2.7.10) were found to be up-regulated in PDAC.

External Validation of the Differential Gene Expression in TCGA-PAAD Data
The differential expression of the identified protein kinase-coding genes (n=40) was further validated in TCGA-PAAD combined GTEx data to increase reliability.

Table 1. Datasets used in the meta-analysis. The table shows the GEO identifier of each dataset and the number of samples assigned to the control group and the PDAC group included in the meta-analysis. ‘Excluded samples’ shows the number of samples that were not included in the meta-analysis due to poor quality.

| GEO Dataset       | Controls (n) | PDAC (n) | Excluded Samples (n) | PMID               |
|-------------------|-------------|----------|----------------------|--------------------|
| GSE62452          | 60          | 69       | 1 control            | PMID: 27197190     |
| GSE15471          | 35          | 36       | 1 control            | PMID: 19260470     |
| GSE62165          | 13          | 117      | 1 case               | PMID: 27520560     |
| GSE18670          | 6           | 6        | 0                    | PMID: 23157946     |
| GSE19280          | 3           | 4        | 0                    | PMID: 23007696     |
| GSE41368          | 6           | 6        | 0                    | PMID: 24120476     |
| GSE71989          | 8           | 14       | 0                    | PMID: 27363020     |

Table 2. The identified differentially expressed protein kinase-coding genes in PDAC. Only genes whose differential expression in PDAC were validated in TCGA-PAAD cohort are listed (ps≤0.01 and fold changes≥2). EC number indicates Enzyme Commission Number. P-val and FDR_p-val are P-values and adjusted P-values, respectively. Z-value is a measure of the differential expression. A positive z-value means that the gene is overexpressed, and a negative z-value means underexpression.

| Enzyme Sub-Subclass | EC Number | Gene Symbol | FDR_p-val | P-val | Z-value | Gene_Name |
|----------------------|-----------|-------------|-----------|-------|---------|-----------|
| EC2.7.10. Protein-Tyrosine Kinases | EC2.7.10.1 | TK2         | 0.0044    | 0.0015 | 3.2     | thymidine kinase 2, mitochondrial |
|                      | EC2.7.10.2 | JAK1        | 0.021     | 0.0096 | 2.6     | Janus kinase 1 |
|                      | EC2.7.10.2 | JAK3        | 0.027     | 0.013  | 2.5     | Janus kinase 3 |
|                      | EC2.7.10.2 | PTK2        | 0.00087   | 0.0002 | 3.7     | protein tyrosine kinase 2 |
|                      | EC2.7.10.2 | SYK         | 0.0015    | 0.00039| 3.5     | spleen associated tyrosine kinase |
|                      | EC2.7.10.1 | TAO1K1      | 0.0004    | 0.000073| 4       | TAO kinase 1 |
|                      | EC2.7.11.1 | MKNK1       | 3.20E-06  | 2.00E-07| -5.2    | MAP kinase interacting serine/threonine kinase 1 |
|                      | EC2.7.11.1 | BMP2K       | 0.0045    | 0.0015 | 3.2     | BMP2 inducible kinase |
|                      | EC2.7.11.1 | CSNK1A1     | 5.70E-07  | 2.60E-08| 5.6     | casein kinase 1 alpha 1 |
|                      | EC2.7.11.1 | EIF2AK2     | 0.00064   | 0.00014 | 3.8     | eukaryotic translation initiation factor 2 alpha kinase 2 |
|                      | EC2.7.11.1 | MAP4K4      | 0.00011   | 9.40E-07| 4.9     | mitogen-activated protein kinase kinase 4 |
|                      | EC2.7.11.1 | MLK1        | 0.00049   | 5.50E-06| 4.5     | mixed lineage kinase domain like pseudokinase |
|                      | EC2.7.11.1 | MKK1        | 7.60E-06  | 5.80E-07| 5       | p21 (RAC1) activated kinase 1 |
|                      | EC2.7.11.1 | RIPK3       | 0.0019    | 0.00054 | 3.5     | receptor interacting serine/threonine kinase 3 |
|                      | EC2.7.11.1 | RPS6KA3     | 0.027     | 0.013  | 2.5     | ribosomal protein S6 kinase A3 |
|                      | EC2.7.11.1 | SIK2        | 0.00064   | 0.00014 | 3.8     | SIK protein kinase 2 |
|                      | EC2.7.11.1 | STK10       | 0.00036   | 0.000066| 4       | serine/threonine kinase 10 |
|                      | EC2.7.11.1 | STK11B     | 0.00064   | 0.00013 | 3.8     | serine/threonine kinase 17b |
|                      | EC2.7.11.1 | STK4        | 0.0001    | 0.000014| 4.3     | serine/threonine kinase 4 |
|                      | EC2.7.11.1 | UHK1        | 2.30E-06  | 1.40E-07| 5.3     | U2AF homology motif kinase 1 |
|                      | EC2.7.11.1 | WNK1        | 2.20E-07  | 8.50E-09| 5.8     | WNK lysine deficient protein kinase 1 |
|                      | EC2.7.11.1 | AKT3        | 0.0027    | 0.00082 | 3.3     | AKT serine/threonine kinase 3 |
|                      | EC2.7.11.1 | LATS2       | 0.00075   | 0.00017 | 3.8     | large tumor suppressor kinase 2 |
|                      | EC2.7.11.1 | PRKACB      | 0.00069   | 0.00015 | 3.8     | protein kinase cAMP-activated catalytic subunit beta |
|                      | EC2.7.11.1 | PRKD3       | 0.029     | 0.014  | 2.5     | protein kinase D3 |
|                      | EC2.7.11.1 | MYLK        | 0.00029   | 0.000049| 4.1     | myosin light chain kinase |
|                      | EC2.7.11.2 | CDK12       | 0.00087   | 0.0002  | 3.7     | cyclin dependent kinase 12 |
|                      | EC2.7.11.2 | CDK19       | 0.0013    | 0.00032 | 3.6     | cyclin dependent kinase 19 |
|                      | EC2.7.11.2 | CDK6        | 0.0011    | 0.00028 | 3.6     | cyclin dependent kinase 6 |
|                      | EC2.7.11.2 | MAP3K2      | 0.00082   | 0.00019 | 3.7     | mitogen-activated protein kinase kinase 2 |
|                      | EC2.7.11.2 | MAP3K8      | 0.0041    | 0.0013  | 3.2     | mitogen-activated protein kinase kinase 8 |
|                      | EC2.7.11.2 | GSK3B       | 0.037     | 0.019  | 2.3     | glycogen synthase kinase 3 beta |

605 | Annals of Clinical and Analytical Medicine
The analysis resulted in the validation of the differential expression of 32 (80%) protein kinase-coding genes identified in PDAC (p≤0.01). The identified and externally validated differentially expressed protein kinase-coding genes in PDAC are shown in Table 2.

Prognostic Value of the Significantly altered Protein kinase-coding Genes in PDAC

The prognostic significance of the identified and validated a total of 32 differentially expressed kinase-coding genes in PDAC was evaluated using integrated long-term follow-up and transcriptome data from over 530 (n=530-1302) patients with PDAC.

Elevated mRNA expressions of 11 genes were found to be significantly associated with the overall survival rate of patients with PDAC (Logrank p≤0.05). Amongst, two genes PTK2 (protein tyrosine kinase 2) and SYK (spleen associated tyrosine kinase) belong to protein-tyrosine kinases sub-subclass. Others were MAP4K4 (mitogen-activated protein kinase kinase kinase kinase 4), TAOK1 (TAO kinase 1), CSNK1A1 (casein kinase 1 alpha 1), EIF2AK2 (eukaryotic translation initiation factor 2 alpha kinase 2), WNK1 (WNK lysine deficient protein kinase 1), PRKACB (protein kinase cAMP-activated catalytic subunit beta), GSK3B (glycogen synthase kinase 3 beta), CDK12 (cyclin- dependent kinase 12), and CDK6 (cyclin- dependent kinase 6) belong to the protein serine/threonine kinases sub-subclass of transferases. Overexpression of nine genes (PTK2, TAOK1, CSNK1A1, EIF2AK2, WNK1, CDK12, CDK6, GSK3B, and MAP4K4) was found to be significantly correlated with worse overall survival of patients with PDAC (Logrank p≤0.05, HR>1). Elevated gene expression of SYK and PRKACB were correlated with better overall survival (Logrank p≤0.05, HR<1). The Kaplan-Meier survival curves for the identified prognostic genes are shown in Figure 2A-K.

Discussion

In this study, the differentially expressed protein kinase-coding genes in pancreatic ductal adenocarcinoma were determined using an integrative transcriptome meta-analysis approach. The identified genes were then externally validated in TCGA PAAD combined GTEx data involving data from PDAC samples and healthy pancreatic tissues to increase the reliability of the findings. Non-validated genes were excluded from further analysis. Eighty percent (n=32) of the identified differentially expressed genes were validated in TCGA combined GTex cohort, suggesting that mRNA expression of these genes may have the potential to be diagnostic biomarkers for PDAC. Since protein-kinases are preferred targets for cancer therapies, these dysregulated genes also merit further study as potentially promising candidates for the development of more effective treatment strategies for PDAC.

The identified protein kinase-coding genes were further evaluated in terms of their prognostic significance in PDAC. Among 32 differentially expressed protein kinase-coding genes, over-expression of nine genes (PTK2, TAOK1, CSNK1A1, EIF2AK2, WNK1, CDK12, CDK6, GSK3B, and MAP4K4) were found to be significantly associated with worse overall survival. Furthermore, overexpression of SYK and PRKACB correlated with better overall survival in the combined seven patient cohorts in OSpaad. Among the identified prognostic genes, a significant relationship between overexpression of eight genes (PTK2, TAOK1, CSNK1A1, EIF2AK2, WNK1, CDK12, CDK6, and PRKACB) and overall survival rate of patients with PDAC has not yet been reported. The knowledge about the functional significance of these identified dysregulated genes in the pathogenesis of PDAC is limited and deserves further investigation.

In this study, MAP4K4 was the gene most significantly associated with the overall survival of patients with PDAC. MAP4K4, belongs to the mammalian STE20/MAP4K family, is involved in important cellular processes such as migration and proliferation [14,15]. Elevated protein expressions of MAP4K4 have been associated with worse prognosis in pancreatic ductal adenocarcinoma before [16]. However, the prognostic significance of MAP4K4 has been evaluated in only stage II PDAC [16]. In this study, overexpression of MAP4K4 was found to be significantly associated with a decreased overall survival rate of patients with PDAC. Therefore, the results of the presented study indicate that MAP4K4 is overexpressed in PDAC compared to control tissues, and elevated mRNA expression of MAP4K4 may have the potential to be a prognostic biomarker for PDAC, inspiring further clinical investigation.

Another gene, which was found to be overexpressed and
Protein kinase-coding genes in PDAC

607

associated with poor overall survival was GSK3B. GSK3B plays important role in the regulation of cell cycle, transcription, proliferation, differentiation, and apoptosis. Overexpression and activity of GSK3B have been reported in pancreatic cancer cells compared to non-neoplastic cells [17]. Moreover, it has been shown that inhibition of GSK3B significantly reduced proliferation and survival of cancer cells, sensitized them to gemcitabine and ionizing radiation, and attenuated their migration and invasion in vitro [17,18]. Furthermore, GSK3B gene expression has been shown to be increased by Ras, which is frequently (>90%) mutated in pancreatic cancer, through Raf/MEK/ERK signaling [19]. GSK3B promotes constitutive NF-κB signaling, which is an important pathway in cancer cell survival, growth and responses to chemotherapeutic agents [20]. In accordance, it has been reported that inhibition of GSK3B arrests pancreatic tumor growth in vivo and decreases NF-kappaB-mediated pancreatic cancer cell survival and proliferation in established tumor xenografts [21]. However, the higher protein level of GSK3B has been shown to be associated with a better survival rate of patients with pancreatic cancer (n=163) [22]. In this study, GSK3B was found to be up-regulated in PDAC compared to controls including adjacent to tumor non-cancerous tissues and normal pancreatic tissues (n=252 tumor vs. n=131 controls), and the identified high expression of GSK3B in PDAC was validated in an external cohort including 179 pancreatic cancer patients and healthy controls (n=171). In addition, the correlation between the identified high mRNA expression of GSK3B and overall survival of patients was evaluated in a cohort including a total of 1301 patients with PDAC. It was found that high tumoral mRNA expression of GSK3B is significantly associated with lower overall survival rate of patients with PDAC. Overall, these indicate that the mRNA level of GSK3B has the potential to be both diagnostic and prognostic biomarker for PDAC. However, the correlation between mRNA and protein levels of GSK3B as well as the potential difference between their prognostic values in PDAC should be investigated in further studies.

Furthermore, the results of the presented study were in

Figure 2. The identified differentially expressed protein kinase-coding genes whose mRNA levels significantly associated with overall survival of patients with PDAC. Kaplan Meier survival plots were created based on the TCGA PAAD dataset and ranked in ascending order by p-value separately for genes associated with good prognosis and poor prognosis. A-I shows Kaplan Meier survival curves for genes whose high expression is associated with worse overall survival. J and K show Kaplan Meier survival curves for the identified genes whose mRNA expression correlated with better overall survival. P<0.05 was accepted statistically significant.
Protein kinase-coding genes in PDAC

according with the results of previous bioinformatics study reporting that overexpression of SYK in pancreatic cancer patients is correlated with better overall survival, which highlights the prognostic value of SYK for PDAC [23].

Conclusion

Overall, the results of this study revealed protein kinase-coding genes whose altered mRNA expression levels can serve as both diagnostic and prognostic biomarkers for PDAC. Additional studies are necessary to validate the suggested diagnostic and prognostic significance of differentially expressed genes in pancreatic ductal adenocarcinoma.

Scientific Responsibility Statement

The authors declare that they are responsible for the article’s scientific content including study design, data collection, analysis and interpretation, writing, some of the main line, or all of the preparation and scientific review of the contents and approval of the final version of the article.

Animal and human rights statement

All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. No animal or human studies were carried out by the authors for this article.

Funding: None

Conflict of interest

None of the authors received any type of financial support that could be considered potential conflict of interest regarding the manuscript or its submission.

References

1. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science (New York, NY). 2002;298(5600):1912-34.
2. Bhullar KS, Lagarón NO, McGowan EM, Parmar I, Jha A, Hubbard BP, et al. Kinase-targeted cancer therapies: progress, challenges and future directions. Mol Cancer. 2018;17(1):48.
3. Schenk PW, Smaar-Jagalska BE. Signal perception and transduction: the role of protein kinases. Biochim Biophys Acta. 1999;1449(1):1-24.
4. Levin ME, Watters D, Song Q. Role of protein kinase activity in apoptosis. Expert Rev. 1996;52(10-11):979-94.
5. Benoni A, Agnoletto C, De Marchi E, Marchi S, Paterngrafi S, Bonora M, et al. Protein Kinases and Phosphatases in the Control of Cell Fate. Enzyme Research. 2020;2020:11(6):603-608.

How to cite this article:

Sevcan Atay. Protein kinase-coding genes as novel diagnostic and prognostic biomarkers for pancreatic ductal adenocarcinoma. Ann Clin Anal Med 2020;11(6):603-608