Exploring the Genomic Landscape of Pancreatic Cancer through Mutational Patterns and Survival Outcome Analysis

Tianyi Chu and Xinyue Qie

2024-02-21

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is a formidable challenge in oncology, often characterized by low survival rates, late diagnosis at advanced stages, and very limited treatment options. The majority of patients have unresectable, locally advanced, or metastatic disease at the time of diagnosis. Moreover, traditional treatments such as chemotherapy, surgery, and radiation have not been shown to significantly improve survival. With its aggressive nature and tendency to metastasize early, the five-year survival rate for pancreatic cancer remains dismal at around 10% (Oberstein and Olive 2013; Cancer Genome Atlas Research Network 2017; Sarantis et al. 2020).

The etiology of PDAC is multifactorial, with risk factors including smoking, obesity, chronic pancreatitis, and family history of the disease. However, the exact mechanisms driving pancreatic tumorigenesis are not fully understood. Recent genomic studies have shed light on the molecular landscape of pancreatic cancer, revealing a complex interplay of genetic alterations that drive tumor initiation, progression, and metastasis.

Key mutations in genes such as KRAS, TP53, CDKN2A, SMAD4, and those involved in DNA repair pathways have been identified as common drivers of pancreatic cancer (Saiki et al. 2021). These mutations disrupt crucial cellular processes, including cell cycle regulation, DNA repair mechanisms, and chromatin remodeling, leading to uncontrolled cell proliferation and tumor growth. Targeted therapies directed against these specific mutations, such as inhibitors of the KRAS pathway, hold promise for improving treatment efficacy and patient survival. Therefore, understanding the genetic mutations driving pancreatic cancer is critical for developing personalized treatment approaches and prognostic stratification for improving patient outcomes, and integrating genomic data with clinical parameters can enhance risk assessment and guide therapeutic decisions, ultimately improving patient care and survival rates.

In this study, we aim to explore the genomic landscape of pancreatic cancer, delineate mutational signatures associated with disease progression, and investigate their implications for patient survival. By elucidating molecular mechanisms and identifying prognostic biomarkers, we hope to contribute to enhancing the diagnosis, treatment, and management of pancreatic cancer.

Research Strategy

Significance

In the context of PDAC prognosis, critical mutations in specific genes, notably KRAS, CDKN2A, TP53, and SMAD4, along with the subsequent activation of related signaling pathways, play an essential role in treatment resistance. Among these genes, KRAS stands out for its exceptionally high mutation rate; a cohort study involving whole-exome sequencing of 109 PDAC cases found KRAS mutations in over 90% of the samples. However, it has been established that KRAS mutations alone are insufficient to initiate cancer. The precise mechanisms through which secondary mutations influence PDAC outcomes remain unclear, underscoring the need for further investigation into how these additional genetic alterations impact prognosis.

Past research has demonstrated that mutations in the TP53 gene are present in 50–75% of human pancreatic

ductal adenocarcinomas (PDAC), typically following an initial activating mutation in the KRAS gene (Morton et al. 2010). Efforts to develop treatments have focused on targeting the mutant P53 protein. By examining the Pancreatic Adenocarcinoma TCGA PanCancer dataset, our objective is to identify additional genes that are prone to secondary mutations, similar to TP53. This approach aims to uncover new avenues for therapeutic intervention.

Furthermore, we propose to conduct survival analysis utilizing various models such as the Cox hazard proportional model, random survival forest, and deep learning-based models. This approach aims to illuminate the survival outcomes among distinct patient groups, particularly those with differing gene mutations. Notably, the STG-Cox (Yamada et al. 2020) and LSPIN-Cox (Yang, Lindenbaum, and Kluger 2022) models we intend to employ offer the advantage of feature selection. This capability is crucial as it allows us to distill a concise subset of features from extensive RNA sequencing data, which comprises over 20,000 genes. This targeted selection can offer clinicians valuable insights into the etiology of PDAC. Given that gene information often constitutes weak signals, leveraging these feature selection models on clinical data could further empower clinicians to pinpoint the most relevant features, potentially enhancing the prognosis and diagnosis of PDAC.

Innovation

Prior research has explored the genes frequently mutated in PDAC through both in-vivo and in-vitro studies. In this study, leveraging the TCGA PanCancer dataset, we intend to delineate the genomic landscape of PDAC using a data-driven methodology. Our analysis will particularly concentrate on identifying mutations that occur independently of TP53 mutations, noting that about 30% of secondary mutations remain uncharted in their specific locations.

Furthermore, we plan to conduct exploratory analyses on samples with the identified secondary mutations. Through survival analysis, we aim to assess the impact of these mutations on patient survival, comparing the outcomes of patients with these genetic alterations against those without, and against patients with TP53 mutations. Our findings will elucidate the influence of gene mutations on survival prospects, highlighting how secondary mutations might differentiate survival outcomes in PDAC patients.

Concurrently, we will examine the nature of mutations within these frequently mutated genes. By integrating mutation data from the TCGA dataset with mRNA sequencing data, we will analyze patient samples based on gene expression levels of these mutated genes. This will allow us to classify patients with secondary mutations and examine how different mutations within the same gene affect patient survival. This comprehensive analysis will provide insights into whether specific types of mutations in the same genes influence the survival rates of PDAC patients differently.

Research Plan

In our project, we will utilize the Pancreatic Adenocarcinoma dataset from The Cancer Genome Atlas (TCGA) PanCancer Atlas project. This dataset provides comprehensive genomic and clinical information on pancreatic adenocarcinoma, a particularly aggressive and challenging cancer type. The TCGA PanCancer data offers a wealth of molecular profiling data, including somatic mutations, gene expression patterns, DNA copy number alterations, and clinical outcomes, collected from a large cohort of pancreatic cancer patients.

Specifically, we will focus on the mutation dataset, the mRNA sequencing z-score dataset, the survival status of patients, and the clinical record of the patients which contains demographic information. Our variables of interest will be time-to-event (death), mutation locations, gene expression and the demographics of patients.

By leveraging this rich dataset, we aim to gain deeper insights into the molecular mechanisms driving pancreatic adenocarcinoma pathogenesis, identify key genetic drivers associated with disease progression and treatment response, and explore potential biomarkers for prognostication and personalized therapy. In particular, we aim to harness this information to analyze mutations and survival outcomes for patients using their clinical features and bulk RNA sequencing data.

Specific Aims

Hypothesis

We would like to divide the PDAC patient data into 4 groups based on their mutation data: (1) patients with no mutations in KRAS, (2) patients with only first-hit mutations in KRAS, (3) patients with both mutations in KRAS and TP53, and (4) patients with mutations in KRAS but not in TP53.

We believe that PDAC patients with mutations in KRAS but not TP53 (Group 4 patients) are likely to have mutations in other certain genes, which we denote as frequent genes for secondary mutations. We hypothesize that patients having mutations in these frequent genes have less optimal survival when compared to patients in group 1 and 2. We also hypothesize that patients in group 4 have different survival when compared to those in group 3.

Specifically, our hypothesis posits that the presence of specific mutations in key genes is correlated with lower survival outcomes in pancreatic cancer patients. We expect that individuals harboring mutations in these genes will exhibit reduced overall survival rates compared to those without such mutations.

Rationale

We intend to perform exploratory analyses on PDAC patient samples and identify frequent locations for secondary mutations other than TP53. By employing survival analysis, our goal is to evaluate how these mutations affect patient survival. We will compare the survival rates of patients harboring these genetic alterations with those who do not possess them, as well as with patients who have TP53 mutations. Our research will shed light on the role of gene mutations in influencing survival rates, emphasizing the distinct impact that secondary mutations have on the survival outcomes of PDAC patients.

Our rationale for this hypothesis is grounded in extensive literature documenting the critical roles of KRAS, CDKN2A, TP53, and SMAD4 in pancreatic tumorigenesis and disease progression. These genes are frequently mutated in pancreatic cancer and are known to exert significant influence over crucial cellular processes such as cell cycle regulation, DNA repair mechanisms, and signaling pathways involved in cancer development and metastasis.

Studies have demonstrated that mutations in KRAS, for instance, are among the earliest and most common genetic alterations in pancreatic cancer, contributing to tumor initiation and progression. Similarly, alterations in CDKN2A, TP53, and SMAD4 have been implicated in the development of aggressive tumor phenotypes, treatment resistance, and metastatic dissemination.

Given the pivotal roles of these genes in pancreatic cancer pathogenesis, it is plausible to hypothesize that the presence of mutations in KRAS, CDKN2A, TP53, and SMAD4 may confer a survival disadvantage to patients by promoting tumor aggressiveness, treatment resistance, and disease recurrence.

Experimental Approach

We plan to compare the effect of different mutations on survival outcomes by benchmarking various computational survival methodologies and models. This includes, but is not limited to, the Cox proportional hazard model (CPH) (Katzman et al. 2018), the random survival forest model (RSF) (Ishwaran et al. 2008), and deep learning-based Cox regression models (Katzman et al. 2018). Additionally, we will explore models incorporating different regularization methods to potentially explore dimensionality reduction and feature selection techniques. These methods aim to identify features that are related to or contribute significantly to survival outcomes. By employing a diverse range of computational approaches, we seek to comprehensively analyze the impact of mutations on patient survival and uncover novel insights into the prognostic factors associated with pancreatic adenocarcinoma.

Interpretation of Results

We anticipate that our analysis will reveal a significant association between mutations in KRAS, CDKN2A, TP53, and SMAD4 and reduced survival outcomes in pancreatic cancer patients. The identification of specific

genetic markers associated with poor prognosis could inform personalized treatment strategies and facilitate risk stratification for improved patient management.

After identifying the frequent locations for secondary mutations, we will examine and compare the survival status of patients with or without mutations in these genes. We anticipate that two patient groups will have different survival time and that the patients without these secondary mutations tend to have longer time-to-event. This result will provide insights into clinical research and prognosis because clinicians can look for mutations in these genes as a first step in diagnosis, and treatment specifically targeting these genes may be more efficient.

Potential Problems and Alternative Approaches

Challenges may arise in the identification and validation of mutations, sample size limitations, and confounding factors affecting survival outcomes. Alternative approaches may involve integrating multi-omics data, employing machine learning algorithms for predictive modeling, and validating findings in independent cohorts to enhance the robustness and generalizability of results.

References

Cancer Genome Atlas Research Network. 2017. "Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma." *Cancer Cell* 32 (2): 185–203.e13. https://doi.org/10.1016/j.ccell.2017.07.007.

Ishwaran, Hemant, Udaya B. Kogalur, Eugene H. Blackstone, and Michael S. Lauer. 2008. "Random Survival Forests." *The Annals of Applied Statistics* 2 (3). https://doi.org/10.1214/08-AOAS169.

Katzman, Jared L., Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. 2018. "DeepSurv: Personalized Treatment Recommender System Using a Cox Proportional Hazards Deep Neural Network." *BMC Medical Research Methodology* 18 (1): 24. https://doi.org/10.1186/s12874-018-0482-1.

Morton, Jennifer P., Paul Timpson, Saadia A. Karim, Rachel A. Ridgway, Dimitris Athineos, Brendan Doyle, Nigel B. Jamieson, et al. 2010. "Mutant P53 Drives Metastasis and Overcomes Growth Arrest/Senescence in Pancreatic Cancer." *Proceedings of the National Academy of Sciences* 107 (1): 246–51. https://doi.org/10.1073/pnas.0908428107.

Oberstein, Paul E., and Kenneth P. Olive. 2013. "Pancreatic Cancer: Why Is It so Hard to Treat?" *Therapeutic Advances in Gastroenterology* 6 (4): 321–37. https://doi.org/10.1177/1756283X13478680.

Saiki, Yuriko, Can Jiang, Masaki Ohmuraya, and Toru Furukawa. 2021. "Genetic Mutations of Pancreatic Cancer and Genetically Engineered Mouse Models." *Cancers* 14 (1): 71. https://doi.org/10.3390/cancers14010071.

Sarantis, Panagiotis, Evangelos Koustas, Adriana Papadimitropoulou, Athanasios G Papavassiliou, and Michalis V Karamouzis. 2020. "Pancreatic Ductal Adenocarcinoma: Treatment Hurdles, Tumor Microenvironment and Immunotherapy." *World Journal of Gastrointestinal Oncology* 12 (2): 173–81. https://doi.org/10.4251/wjgo.v12.i2.173.

Yamada, Yutaro, Ofir Lindenbaum, Sahand Negahban, and Yuval Kluger. 2020. "Feature Selection Using Stochastic Gates." In *International Conference on Machine Learning*, 10648–59. PMLR. https://proceedings.mlr.press/v119/yamada20a.html.

Yang, Junchen, Ofir Lindenbaum, and Yuval Kluger. 2022. "Locally Sparse Neural Networks for Tabular Biomedical Data." In *International Conference on Machine Learning*, 25123–53. PMLR. https://proceedings.mlr.press/v162/yang22i.html.