Digonto Chatterjee

Final Project

Jack Hester

CEPC0904 8/11/2021

# Racial Disparities in Survival Outcomes of Pancreatic Cancer Patients: A Quantitative Genomic Analysis

**Abstract**

**Purpose**: Pancreatic cancer remains a major health concern; currently it is the second leading cause of cancer deaths in the United States. Genetic research on pancreatic cancer has generally only focused on familial pancreatic cancer, which is only 10% of all pancreatic cancer patients. This study focuses on finding genes that impact the survival of pancreatic cancer patients based on race, with the goal of finding genes that are race specific and can be used as biomarkers and potential targets to develop personalized treatment options. **Methods:** I performed a quantitative genomic analysis of pancreatic cancer patients stratified by race categories. **Results**: White, African Americans and Asian have distinct copy number alterations, more specifically amplifications in specific genes, with GATA6, RECQL4, and MIB1 being only altered in White patients, PKD1L1, GARS1, and NEUROD6 only altered in Black & African American patients, and the entire 4p16.3 cytoband of genes (39 in total), only altered in Asian patients. These unique genes were also shown to have an impact on survival. **Conclusion**: Specific race based diagnostic/prognostic biomarkers and targeted therapeutic options can be developed that will not only be personalized for pancreatic cancer patients depending on their race, it might also address some of the health disparities that exist in current treatment modalities.

# Introduction

# Pancreatic cancer is among the most deadly forms of cancer. It is the seventh most common cancer, yet it is the second leading cause of cancer deaths in the United States [1]. It is estimated that in 2021, 48,220 patients will die from pancreatic cancer. Risk factors for pancreatic cancer include smoking, diabetes, obesity, chronic pancreatitis, and family history [2]. Over 80% of the patients present with metastatic disease. Despite advances in chemotherapy, the average survival remains less than 5 years even after surgery [3].

Genetic testing has been primarily focused on familial pancreatic cancer, which only accounts for ten (10%) percent of all pancreatic cancers. A study on non familial pancreatic cancer patients revealed six genes namely CDKN2A; TP53; MLH1; BRCA2; ATM and BRCA1 with significant associations between pancreatic cancer and mutations in these genes [4]. A previous study has shown that patient survival was associated with mutations in KRAS, CDKN2A, SMAD4, and TP53 [5] .

Similar to other common malignancies, pancreatic cancer is associated with disparities by socioeconomic status (SES), ethnic minority status, and insurance [6] [7]. In contrast to other types of cancer (breast, colon) where screening can detect early-stage disease, no screening modality exists for pancreatic cancer. Thus, disparities in outcomes for pancreatic cancer do not result from lack of screening [8]. There are currently limited data on the genetic susceptibility of pancreatic cancer survival based on race. The association of driver gene alterations based on racial category and their association with patient outcomes has not been clearly established. Therefore, I performed a quantitative genomic analysis to find genes associated with patient survival for the pancreatic cancer patients based on their racial categories. The goal of this study is to determine whether White, African American/Black and Asian patients have different gene alterations and whether or not they can be diagnosed and managed based on their specific signature profile. Identification, prevention and management of factors based on race may help find effective strategies for clinical management of pancreatic cancer.

# Methods

* 1. *Clinical Data from cBioportal*

I utilized cBioPortal to access and analyze the public database on cancer tissues generated by the TCGA project (<https://www.cancer.gov/tcga>). The cBioPortal allows users to question datasets across data types including genes and clinical samples, providing an opportunity to investigate a number of different biologically and/or clinically relevant hypotheses. For this study, I selected all datasets for pancreatic cancer. The data was then stratified based on race information for the patients. Three (3) racial categories, 1) White 2) African American and Black, and 3) Asian were chosen. Sample sets were customized based on these three (3) racial categories and virtual datasets were created. Datasets that did not provide race information were not included in our analysis.

# Data Analysis

I analyzed the copy number alteration data (CNAs) in each of our three (3) customized datasets. Within those separate virtual studies, the genes that were amplified in the most patients in terms of CNA (copy number alterations), were recorded and compared among the different virtual datasets.

# Survival Curves

Genes that were altered differently in the racial groups were selected to analyze the genes’ impact on patient survival. Both overall survival (OS) and disease free survival (DFS) curves were computed. This was done by utilizing the inbuilt statistical analysis tools available within the cBioPortal platform. The data was tested for significance using a log-rank test to compare the survival distributions between 2 samples. This was done automatically using cBioPortal.

**2.4 Protein Drug Analysis**

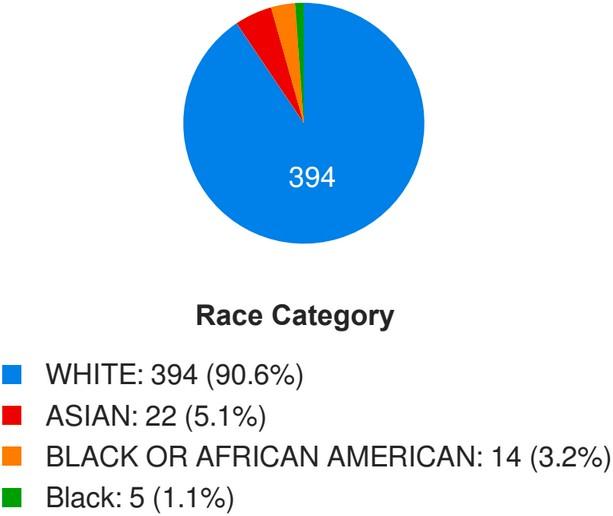
I utilized MD Anderson Cell Lines Project (MCLP) and genecards.org to find protein drug interactions using the genes found in our study as drug targets.

# Results

*Pancreatic Cancer Race Categories:*

I analyzed more than 1200 patient data spanning ten (10) different pancreatic cancer clinical studies. The disease types were 1) Acinar cell carcinoma of the pancreas (JHU, J Pathol 2014), 2) Cystic tumor of the pancreas (Johns Hopkins PNAS 2011), 3) Five (5) Pancreatic Adenocarcinoma studies (ICGC, Nature 2012; QCMG, Nature 2016; TCGA, Firehose Legacy; TCGA PanCancer Atlas; and Pancreatic Cancer UTSW, Nature Communication, 2015), and 4) three (3) Pancreatic Neuroendocrine tumor studies (Shanghai, Nat. Commun. 2013; Johns Hopkins University, Science 2011; Pancreatic Neuroendocrine Tumors (Multi-Institute, Nature 2017). The data was stratified according to three race categories self-identified by patients; 1) White, 2) African American or Black and 3) Asian. There were only 2 samples that identified as Hispanic, therefore the hispanic category was not included in our study. A total of 436 patients samples were customized based on three (3) race categories. The samples were from three (3) Pancreatic Adenocarcinoma studies. Figure 1 shows the customized datasets that were created based on the race categories. For our analysis Black or African American and Black were taken together. The number of samples for White patients was 394, 22 for Asian, and 19 for Black & African American.

**Figure 1**: Customized Dataset created from Pancreatic Cancer Studies in cBioportal



# Copy Number Alterations based on Race Categories

I initially began the analysis by looking for copy number alterations (CNA) at the genomic DNA level for pancreatic cancer samples. Copy number alterations (CNAs) refer to changes in the number of copies of a genetic region caused by deletion or duplication events in the genome. Recent technological advances have enabled the identification of CNAs across the entire genome, associating cancer with the alteration rates of various genes in cancer [9].

I utilized cBioPortal, which contains comprehensive genomic and transcriptomic data from various cancer studies [10, 11]. Focusing on the customized pancreatic cancer datasets created based on race categories as mentioned in the above section I determined the genes with the highest frequency of copy number alterations (CNA) in each race category. I found a few unique genes specific to each race category that had copy number amplifications as shown in Table 1. The altered genes' locations are in different cytobands for each race category. As previously reported, CDKN2A was deleted in 40% of patients in every race category [5], confirming our data analysis approach. Most genes that had deep deletions in copy number were common among the different race categories, therefore were not included in further investigation and analysis.

# Table 1: Genes With Copy Number Alterations Unique to Each Race Category

| **Race** | **Genes with Copy Number Alterations (AMP)** | **Cytoband** | **Alteration Frequency (%)** |
| --- | --- | --- | --- |
| White | GATA6, RECQL4, MIB1 | 18q11.2 (GATA5, MIB1)  8q24.3 (RECQL4) | 14.8 (GATA6) 14.0(RECQL4) 12.8%(MIB1) |
| Black | PKDL1, GARS1, NEUROD3 | 7p12.3(PKDL1)  7p14.3(GARS1, NEUROD3) | 28.6 |
| Asian | FGFR3, ABCA11P, UVSSA(etc.) | 4p16.3 | 27.3 |

**Association of CNA Amplification with Survival Outcome**

I evaluated the association of the gene alterations unique to each category and most frequently amplified in each race category with survival outcome. The most frequently altered genes in each category were found to have a direct association with patient survival. Kaplan- Meyer curves for overall survival (OS) and disease free survival (DFS) were generated using the cbioportal platform for each gene in the specific race categories. P-values were generated by cBioPortal using the logrank test for survival between 2 samples. This was compared to a significance level of 0.05. The survival was calculated in median months, and the 95% confidence interval was given as well.

# Figure 2: Kaplan-Meyer Curves For Disease Free Survival In Each Race Category

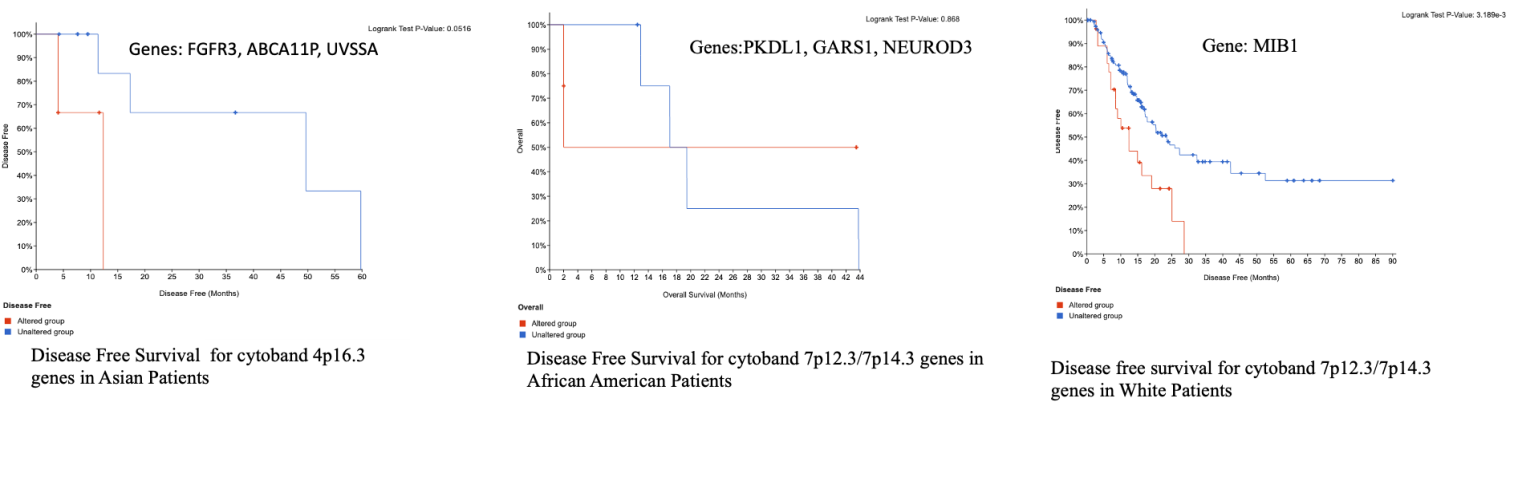


Figure 2 shows representative Kaplan-Meier curves for disease free survival (DFS) outcomes associated with alterations of genes in patients from different race categories. The p-values are 0.0516 for Asian patients, 0.868 for African American & Black patients, and 0.003189 for White patients.

# Table 2: Survival Outcome Associated With Altered Genes In Each Race Category

| **Race Category** | **Gene Altered** | **Overall Survival (OS) Unaltered** | **Overall Survival (OS)**  **Altered** | **Disease Free Survival (DFS)- Unaltered** | **Disease Free Survival (DFS)-Altered** |
| --- | --- | --- | --- | --- | --- |
| White | GATA6 | 20.17 | 20.35 | 20.4 | 12.43 |
| RECQL4 | 20.35 | 15.11 | 20.37 | 9.57 |
| MIB1 | 20.19 | 20.34 | 23.52 | 12.42 |
| African American & Black | PKD1L1, GARS1, NEUROD6 | 17.03 | 2.01 | NA | NA |
| Asian | FGFR3, ABCA11P, UVSSA,  (etc.) | 66.89 | 11.61 | 12.32 | 49.68 |

My results showed that there are unique genes that are associated with each race category which have an effect on the survival outcomes of patients. For patients that had expression of these genes the median months of survival was less compared to those who did not have amplified expression.

*White Patients:* For the genes amplified in the White race category, the effect on OS was minimal while the DFS was significantly worse than in patients where there was no amplification. The genes amplified in the White race category were from the cytoband 18p11.2 (GATA6, MIB1) and 8q24.3 (RECQL4). It is notable that the genes that were from the same cytoband showed nearly identical data.

*African American and Black Patients:* For genes amplified in the African American & Black race category, the effect on overall survival for patients that had amplified genes in the 7p cytoband had significantly poor outcomes (2 months) with alterations in the genes PKD1L1, GARS1, NEUROD6 compared to patients with no amplification (17 months). There was no data for disease free survival due to the limited number of samples. However, it is notable to mention that data for the African American and Black and Asian race categories were limited (African American/Black, n=19 and Asian n=22), and were not always statistically significant.

*Asian Patients:* For genes amplified in the Asian race category, the survival outcomes for patients with amplified genes was significantly worse for both overall survival and disease free survival compared to patients who did not have amplified genes. The genes in the Asian race category were all from the same cytoband (4p16.3), and while only 3 were listed in Table 2, there were 39 genes from the cytoband which were all amplified and showed the same survival outcomes.

**Protein Expression and Protein Drug Interaction**

Using the Protein Atlas, Protein-Drug Interaction MCLP dataset and genecards.org data, I analyzed the role of the genes in cancers that are the focus of our study and determined drug candidates that can target the proteins encoded by the genes. I propose potential drug candidates that can be further tested in clinical trials to provide gene and race specific therapeutic options for pancreatic cancer patients. Table 3 shows the role of the genes in cancers, their role as prognostic markers and drugs that can be used to target these genes. The top 2 drugs targeting each gene were included.

**Table 3: Role of the Genes in Cancers and Potential Drug Candidates**

| **Genes/Race Categories** | **Role in Pancreatic Cancer** | **Role in Other Cancers (Prognostic Value)** | **Drug Candidates** |
| --- | --- | --- | --- |
| GATA6 (White) | No | Renal (unfavorable) (detected in many) | Spautin.1; parbendazole |
| RECQL4 (White) | No | Liver (unfavorable) (detected in all) | Rec 15/2615 dihydrochloride |
| MIB1 (White) | No | None (detected in all) | SIB 1893 |
| PKD1L1 (Black) | No | None (detected in many) | JKC 363; HS 014 |
| GARS1 (Black) | NA | NA | MRS 1220; RS 100329 hydrochloride |
| NEUROD6 (Black) | No | None (not detected) | SB 224289 hydrochloride; Pancuronium dibromide |
| FGFR3 (Asian) | No | Endometrial (unfavorable) (detected in many) | PD 173074 |
| ABCA11P (Asian) | NA | NA | Bobcat339, TBCA |
| UVSSA (Asian) | No | Renal (unfavorable) urothelial (favorable) detected in all | Iressa, ASB 14780 |

# Conclusion

* 1. *Overall Findings*

This study demonstrates that there are unique sets of genes commonly associated with self-identified race. Alterations in these genes are associated with patient outcomes in patients with pancreatic cancer. These genes serve as potential biomarkers for specific race groups to predict patient outcomes. Additionally, several potential drug candidates proposed in our study targeting the specific proteins encoded by the genes specific to the individual race categories can be further studied in clinical settings to develop more personalized race based therapeutic options. Additionally, understanding the molecular events and mechanisms that determine patient outcomes has the potential to develop new and improved treatment approaches for patients with pancreatic cancer.

# *4. 2 Results in Context*

My results are correlated by other studies, as GATA6 is now a known oncogene [12], meaning it has the potential to cause cancer. Similarly, studies have shown copy number alterations in the cytoband 8q24.3 having a role in cancer [13]. The role of genes and their respective cytobands from the Black & African American as well as Asian race category was more novel. Of the genes found to be amplified in these 2 race categories, only FGFR3 is known to be an oncogene [14]. It is notable that the cytoband 4p16.3 which was amplified frequently in the Asian race category, showing 39 genes all with frequent alteration, causes Wolf-Hirschhorn syndrome when deleted [15], however there has been little investigation on its amplification and no study on its unique amplification in the Asian race category or pancreatic cancer.

# *Shortcomings*

The biggest shortcoming was that the sample size was low, especially for the Black and African American and Asian race categories. Despite this, the results can be used as a good starting point for further, more robust investigation.

# F*uture goals*

Future goals include developing new and improved treatment for patients with pancreatic cancer, potentially by targeting the genes that were noted in this study. Another potential goal would be to do statistical modeling and determine if these genes can be used for diagnosis of

pancreatic cancer, which would be especially noteworthy as pancreatic cancer is known to be hard to diagnose early. Furthermore, the specific role of each cytoband that was largely amplified and its association with pancreatic cancer can be investigated.

# References

1. Siegel, Rebecca L., et al. “Cancer Statistics, 2019.” CA: A Cancer Journal for Clinicians, vol. 69, no. 1, Jan. 2019, pp. 7–34. PubMed, <https://doi.org/10.3322/caac.21551>.
2. Ryan, David P., et al. “Pancreatic Adenocarcinoma.” *The New England Journal of Medicine*, vol. 371, no. 11, Sept. 2014, pp. 1039–49. *PubMed*, <https://doi.org/10.1056/NEJMra1404198>.
3. Noel, Marcus, and Kevin Fiscella. “Disparities in Pancreatic Cancer Treatment and Outcomes.” *Health Equity*, vol. 3, no. 1, 2019, pp. 532–40. *PubMed*, <https://doi.org/10.1089/heq.2019.0057>.
4. Hu, Chunling, et al. “Association Between Inherited Germline Mutations in Cancer Predisposition Genes and Risk of Pancreatic Cancer.” *JAMA*, vol. 319, no. 23, June 2018, pp. 2401–09. *Silverchair*, <https://doi.org/10.1001/jama.2018.6228>.
5. Qian, Zhi Rong, et al. “Association of Alterations in Main Driver Genes With Outcomes of Patients With Resected Pancreatic Ductal Adenocarcinoma.” *JAMA Oncology*, vol. 4, no. 3, Mar. 2018, pp. e173420–e173420. *Silverchair*, <https://doi.org/10.1001/jamaoncol.2017.3420>.
6. Shapiro, Mia, et al. “Associations of Socioeconomic Variables With Resection, Stage, and Survival in Patients With Early-Stage Pancreatic Cancer.” *JAMA Surgery*, vol. 151, no. 4, Apr. 2016, pp. 338–45. *PubMed*, <https://doi.org/10.1001/jamasurg.2015.4239>.
7. Khawja, Shumaila N., et al. “Pancreatic Cancer Disparities in African Americans.” *Pancreas*, vol. 44, no. 4, May 2015, pp. 522–27. *PubMed*, <https://doi.org/10.1097/MPA.0000000000000323>.
8. Zhang, Qiubo, et al. “Pancreatic Cancer Epidemiology, Detection, and Management.” *Gastroenterology Research and Practice*, vol. 2016, 2016, p. 8962321. *PubMed*, <https://doi.org/10.1155/2016/8962321>.
9. Shlien, Adam, and David Malkin. “Copy Number Variations and Cancer.” *Genome Medicine*, vol. 1, no. 6, June 2009, p. 62. *PubMed*,<https://doi.org/10.1186/gm62>.
10. Cerami, Ethan, et al. “The CBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data.” *Cancer Discovery*, vol. 2, no. 5, May 2012, pp. 401–04. *PubMed*,<https://doi.org/10.1158/2159-8290.CD-12-0095>.
11. Gao, Jianjiong, et al. “Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the CBioPortal.” *Science Signaling*, vol. 6, no. 269, Apr. 2013, p. pl1. *PubMed*,<https://doi.org/10.1126/scisignal.2004088>.
12. Lin, Lin, et al. “Activation of GATA Binding Protein 6 (GATA6) Sustains Oncogenic Lineage-Survival in Esophageal Adenocarcinoma.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 11, Mar. 2012, pp. 4251–56. *PubMed Central*,<https://doi.org/10.1073/pnas.1011989109>.
13. Brusselaers, Nele, et al. “Copy Number of 8q24.3 Drives HSF1 Expression and Patient Outcome in Cancer: An Individual Patient Data Meta-Analysis.” *Human Genomics*, vol. 13, no. 1, Nov. 2019, p. 54. *BioMed Central*,<https://doi.org/10.1186/s40246-019-0241-3>.
14. Lafitte, Marie, et al. “FGFR3 Has Tumor Suppressor Properties in Cells with Epithelial Phenotype.” *Molecular Cancer*, vol. 12, no. 1, July 2013, p. 83. *BioMed Central*,<https://doi.org/10.1186/1476-4598-12-83>.
15. Bernardini, Laura, et al. “Small 4p16.3 Deletions: Three Additional Patients and Review of the Literature.” *American Journal of Medical Genetics. Part A*, vol. 176, no. 11, Nov. 2018, pp. 2501–08. *PubMed*,<https://doi.org/10.1002/ajmg.a.40512>.