Expressional correlations between SERPINA6 and pancreatic ductal adenocarcinoma-linked genes

Maxwell Selver¹, Lisa Tibbetts²
¹ Tenafly High School, Tenafly, New Jersey
² Georgetown University, Washington, D.C.

SUMMARY
Pancreatic ductal adenocarcinoma (PDAC) is the most common form of pancreatic cancer, with early diagnosis and treatment challenges. When any of the genes KRAS, SMAD4, TP53, and BRCA2 are heavily mutated, they correlate with PDAC progression. Cellular stress, partly regulated by the gene SERPINA6, also correlates with PDAC progression. When SERPINA6 is highly expressed, corticosteroid-binding globulin inhibits the effect of the stress hormone cortisol. We hypothesized that the expression of SERPINA6 would inversely correlate with the expression of PDAC-linked genes. Healthy pancreatic expression control data was sourced from GTEx, while mutated PDAC experimental expression data was sourced from TCGA. Eight scatterplots with $p$, $R$, and $R^2$ values were produced via Correlation Analysis on gene profiling database GEPIA2. A lack of experimental statistical significance of the KRAS and TP53 scatterplots indicate the genes do not correlate with SERPINA6. SMAD4 scatterplots demonstrated statistical significance on both ends with a direct trend, indicating a potential correlation with SERPINA6. While the control scatterplot of BRCA2 exhibited statistical insignificance, the experimental scatterplot exhibited significance in addition to a hypothesized inverse trend. This may indicate a tumor-specific correlation between BRCA2 and SERPINA6. Further exploring the viability of this potential expressional correlation may build on lacking early diagnosis techniques and further supplement the scientific community’s interest in the mind-body stress connection.

INTRODUCTION
Pancreatic ductal adenocarcinoma, also known as PDAC, is a devastating type of cancer with a history of late prognosis and low survival rates, with an average five-year survival rate of 9% (1). As of 2019, PDAC was the fourth deadliest cancer in the United States, resulting in nearly 40,000 American deaths (2). Presently, PDAC cases are expected to double over the next ten years (3). Currently, even with modern technology, it is challenging to study pancreatic cancer in patients as nearly 85% of those diagnosed have tumors that have already metastasized, resulting in a 4% clinical trial enrollment rate (3). Hence, discovering convenient and concrete ways of detecting and treating PDAC has become a significant area of scientific interest (2, 4). Past studies have illustrated how the increase of biomarker proteins like carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA 19-9) in the blood of PDAC patients indicates disease progression, but these are difficult to trace in early stages of the disease (4, 5). However, recent studies have shown that expression of the mutated genes KRAS, SMAD4, TP53, and BRCA2 may correlate with a higher risk and progression of PDAC (6).

KRAS, located on chromosome 12, is often mutated early in pancreatic carcinogenesis (6). When guanosine triphosphate (GTP) activates KRAS, the produced KRAS protein triggers intracellular flagging pathways and transcription factors that can incite uncontrolled cellular proliferation (7). Recent investigations of ablated KRAS have illustrated a significant decrease in tumor size in mouse models (6). Nearly 90% of patients with pancreatic cancer have mutations in KRAS (8).

SMAD4 is a tumor suppressor located on chromosome 18 that is inactive in more than half of pancreatic cancer cases (2). SMAD4 facilitates signals from growth factor-β ligands. These ligands phosphorylate SMAD proteins that are critical for regulating cell development (9). Lower levels of SMAD4 expression often correlate with tumor prognosis in PDAC (6).

TP53 is a tumor suppressor on chromosome 17 that encodes for protein p53, which incites apoptosis and slows tumor growth (6). TP53 is mutated in nearly 70% of pancreatic cancers and is most often incited by external cellular stress (6). Low TP53 expression is often associated with a worse PDAC prognosis in patients (10).

BRCA2, often associated with pancreatic and breast cancers, is a tumor suppressor located on chromosome 13. When mutated, BRCA2 alters cell proliferation and gene-transcriptional regulation (13). Nearly 10% of PDAC patients have some form of a BRCA mutation, illustrating its importance in genetic analysis for PDAC and breast cancer risk profiles (1). Assessments are particularly beneficial for those with Ashkenazi Jewish ancestry or predetermined BRCA mutations (1).

In recent years, a near consensus has been reached amongst scientists and oncologists that various forms of external cellular stress may contribute to the progression of many disorders and diseases, including various cancers (12, 13). One study found that in mice, chronic exposure to glucocorticoids, which are stress-induced steroid hormones that influence bodily systems, alter gene expression but do not change DNA sequences (14). Another recent finding indicates that beta-blockers, which inhibit stress hormones, may contribute to better pancreatic cancer prognosis (15, 16). Stress-driven responses from extracellular scenarios have been the subject of investigation for many years, and increasing evidence for their role in the progression of physiological disorders like cancer is leading to a broader
acceptance of notions such as the mind-body connection. Newfound interest in these responses by the scientific community has started to lay the path for a new field of study that extends beyond holistic notions (17, 13).

One of the most examined genes that correlates with stress is SERPNA6, that encodes for the protein corticosteroid-binding globulin (CBG) which is produced in the liver. CBG binds to the stress hormone cortisol, which plays an instrumental role in regulating blood sugar, inflammation, and metabolism (18). When CBG is bound to cortisol, typically when it is in the process of traveling through the bloodstream, the hormone’s function is suppressed (18). Although cortisol is a necessary component for the regular function of bodily processes, one study suggests that particularly elevated levels of cortisol, especially for long periods in adults, can trigger mental health issues like anxiety and depression, and potentially lead to other physiological ramifications, such as cancer (12, 13). When SERPNA6 is highly expressed, activated cortisol levels decrease, and overall stress levels along with it (18).

We established this study to determine if increased stress levels increase the gene expression of PDAC-linked genes. Both chronic cellular stress, represented by low expression of the gene SERPNA6, and mutations in the genes KRAS, TP53, SMAD4, and BRCA2 have been proven to correlate with PDAC risk and progression. Hence, to potentially identify a correlation between these linkages, we hypothesized that the expression of SERPNA6 would inversely contribute to a higher expression of genes linked to PDAC risk and progression, such as KRAS, TP53, SMAD4, and BRCA2.

RESULTS

We utilized the free online database Gene Expression Profiling Interactive Analysis 2, or GEPIA2 (19). GEPIA2 is a reliable resource for gene expression data from both the Cancer Genome Atlas (TCGA) program and the Genotype-Tissue Expression (GTEx) projects. Differential genes, expression DIY, survival analysis, isoform details, correlation analysis, similar genes detection, and dimensionality reduction are some of the features that are provided to help analysts investigate biological mechanisms of various cancer genes and subtypes. GEPIA2 uses a customized Python package to aid in accessible command-line analyses and visualizations. Specifically, the Correlation Analysis tool was utilized to compare each PDAC-linked gene expression profile datapoint to each SERPNA6 expression profile datapoint. Although we did not draw regression lines, we analyzed proper p, R, and R² values. We considered statistical significance and plot linearity when comparing each expressional correlation.

The data in the experimental scatterplot comparing KRAS and SERPNA6 is less linear than the control. For the control (Figure 1A), a p-value of 0.00025 renders the set statistically significant (p-values of < 0.05 entail significance in biological studies), meaning the genes interact with each other and is not by chance. The R-value of 0.28 indicates that there is a weak positive relationship present between KRAS and SERPNA6. For the experimental scatterplot (Figure 1B), the p-value of 0.19 renders the set statistically insignificant, meaning there is no expressional correlation between KRAS and SERPNA6. The R-value of -0.098 indicates that despite its negative correlation value, there is almost no visible trend in the set. We found 159 points for the control scatterplot and 177 points for the experimental scatterplot.

Figure 1: KRAS vs. SERPNA6 expression in control healthy pancreatic cells (top) and experimental PDAC tumor cells (bottom). Data is represented by transcripts per million (TPM), a unit of expressional measure. A log2 transformation was applied to the data to condense graphical visualizations. (A) 159 points are exhibited. (B) 177 points are exhibited. P-values, R-values, and R² values are listed at the top of the figure.

The data in the experimental scatter plot comparing SMAD4 and SERPNA6 is less linear than that of the control. The p-value of the control is extremely significant with a value of 9.4E-12, meaning the genes expressionaly correlate (Figure 2A). The control also exhibits an R-value of 0.5, indicating there is a moderate positive trend. The experimental p-value is also statistically significant with a value of 0.037, indicating there is an expressional correlation present (Figure 2B). The experimental R-value of 0.16 indicates a slight positive correlation between the expression of the two genes. We found 161 points for the control scatterplot and 176 points for the experimental scatterplot.

Akin to the correlation between SMAD4 and SERPNA6, the data in the experimental scatterplot comparing TP53 and SERPNA6 is less linear than the control dataset. The p-value of the control (Figure 3A), 3.1E-9, is extremely significant, entailing an expressional correlation. The control exhibits a moderate positive trend with an R-value of 0.44. For the experimental scatterplot (Figure 3B), the p-value of 0.16 renders the set statistically insignificant, indicating no correlation in PDAC gene expression. Still, the R-value of -0.11 indicates that there is still a weak negative trend between the two variables. We found 162 points for the control scatterplot.
and 182 points for the experimental scatterplot.

The BRCA2 vs. SERPINA6 correlation varied most out of the performed correlation analyses, with the largest difference in \( p \)-values and less significance in the control than the experimental data. The \( p \)-value of 0.63 for the control renders the set statistically insignificant, meaning no expressional correlation is present (Figure 4A). The R-value of 0.037 illustrates that there is little to no trend in the control scatterplot, albeit positive. However, the \( p \)-value of the experimental set (Figure 4B), 0.036, is statistically significant, illustrating an expressional correlation. This is accompanied by the lowest R-value exhibited in the study, -0.16. Although the R-value does not represent an extremely strong trend, a negative correlation is exhibited. This may suggest a PDAC tumor-specific expressional correlation. We found 162 points for the control scatterplot and 178 points for the experimental scatterplot.

**DISCUSSION**

Regarding the KRAS vs. SERPINA6 correlation analyses (Figures 1A and 1B), due to a lack of statistical significance on the experimental end, we can conclude that there is no expressional correlation between the two genes. For the SMAD4 vs. SERPINA6 correlations (Figures 2A and 2B), both experimental and control models exhibited statistical significance. The experimental graph’s R-value is positive, indicating a direct relationship instead of a hypothesized inverse one. For the TP53 vs. SERPINA6 correlation analyses (Figures 3A and 3B), an expressional correlation is not present due to a lack of statistical significance on the experimental end. For the BRCA2 vs. SERPINA6 correlation analyses (Figures 4A and 4B), an expressional correlation is not present on the control end as there is a lack of statistical significance. The experimental end of this analysis, however, exhibits significance and a negative trend, entailing an expressional correlation. This is notable as this correlation is exhibited in PDAC tumors and not in healthy pancreatic cells, meaning it may be tumor-specific.

In terms of overall results for the correlation analyses performed, \( R^2 \) values for all correlations were indicative of a lack of linearity in each plot. Reductions in R-value across all correlations may prompt future reexaminations that identify other, possibly non-linear trends throughout the data. Signs of clumping throughout the data could also be examined in further investigations.

The data presented cannot definitively support an expressional correlation between SERPINA6 and each PDAC-linked gene studied. Still, it is within the realm of possibility that the expression levels of mutated KRAS, SMAD4, TP53, and BRCA2 may affect SERPINA6 expression. Future

---

**Figure 2: SMAD4 vs. SERPINA6 expression in control healthy pancreatic cells (top) and experimental PDAC tumor cells (bottom).** Data is represented by transcripts per million (TPM), a unit of expressional measure. A log2 transformation was applied to the data to condense graphical visualizations. (A) 161 points are exhibited. (B) 176 points are exhibited. \( p \)-values, \( R \)-values, and \( R^2 \) values are listed at the top of the figure.

**Figure 3: TP53 vs. SERPINA6 expression in control healthy pancreatic cells (top) and experimental PDAC tumor cells (bottom).** Data is represented by transcripts per million (TPM), a unit of expressional measure. A log2 transformation was applied to the data to condense graphical visualizations. (A) 162 points are exhibited. (B) 182 points are exhibited. \( R \)-values, and \( R^2 \) values are listed at the top of the figure.
investigations may entail comparing SERPINA6 expression between patients with wildtype and mutant versions of PDAC-linked genes. Furthermore, searching for genes that co-express with SERPINA6 holds a point of interest in potentially identifying other expressional correlations, and identifying rates of survival based on patients’ genetic profiles may help illuminate disparities in expression compared to stress levels (as represented by SERPINA6).

In terms of limitations, statistical significance posed a large issue to drawing viable conclusions. Variations among n-values (number of points) when comparing expressional correlations may have limited comparisons between healthy pancreatic patients and PDAC patients. Ideally, the control and experimental n-values would be identical. Regression lines were also not able to be incorporated into the graph-generating pipeline, which may have been useful when comparing linearity with $R^2$ values.

In the future, these findings may point to looking at the progression of SERPINA6 progression throughout various stages of PDAC. Additionally, analyzing the correlation between other previously identified measures of stress in the form of hormones called catecholamines and other PDAC-linked genes, like PALB2 and ATM, holds promise considering the stress basis of cancer progression (15). If a stronger correlation is found between SERPINA6, cortisol levels, catecholamines, or other stress-related genes or chemicals and PDAC-linked genes, developing treatments that specifically target sources of chronic stress in cancer patients to improve their prognosis could be further investigated in clinical trials. These new treatments could improve chances of survival and provide insight on how effective a more stress-free patient may affect the prognosis of a particularly aggressive disease.

MATERIALS AND METHODS

To determine the presence of a potential expressional correlation between SERPINA6 and each PDAC-linked gene, GEPIA2’s Correlation Analysis tool plugin was used to compare expressional profiles of PDAC patients pertaining to each gene (19). Each Correlation Analysis (SERPINA6 v. each PDAC-linked gene) produced one control plot and one experimental (tumor) plot. Control scatterplots consisted of non-malignant pancreatic gene expression data from GTEx (The Genotype-Tissue Expression Project), and each experimental scatterplot consisted of mutated PDAC (called PAAD on GEPIA2’s interface) gene expression data from TCGA (The Cancer Genome Atlas Project). For proper analysis, a Pearson coefficient, explicitly used to measure the degree of linearly related variables, was applied to all scatterplots. Additionally, a log2 transformation was applied to the scatterplots to produce more condensed visualizations. The Correlation Analysis pipeline directly produced eight total expression scatterplots with respective $p$, $R$, and $R^2$ values. $P$-values were analyzed for statistical significance to determine whether each analysis exhibited an expressional correlation. If a model was deemed statistically significant ($p < 0.05$), $R$-values were analyzed to measure the degree of direct or inverse trends for each expressional correlation (on a scale of -1 to 1). $R^2$ values were also analyzed to determine the degree of linearity for each expressional correlation to see how closely the model adhered to a general linear trend.

Received: March 10, 2021  
Accepted: August 14, 2021  
Published: October 06, 2021

REFERENCES


Figure 4: BRCA2 vs. SERPINA6 expression in control healthy pancreatic cells (top) and experimental PDAC tumor cells (bottom). Data is represented by transcripts per million (TPM), a unit of expresional measure. A log2 transformation was applied to the data to condense graphical visualizations. (A) 162 points are exhibited. (B) 178 points are exhibited. $R$-values, and $R^2$ values are listed at the top of the figure.


Copyright: © 2021 Selver and Tibbetts. All JEI articles are distributed under the attribution non-commercial, no derivative license (http://creativecommons.org/licenses/by-nc-nd/3.0/). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.