The Investigation of the Correlation between Urine Biomarkers and Pancreatic Ductal Adenocarcinoma

Jiahui Shen[a] [iD]

Institute of Problem Solving, Beijing Normal University Hong Kong Baptist University United International College, Zhuhai, Guangdong, China

Abstract: Pancreatic ductal adenocarcinoma (PDAC)'s low survival rate has long been a world unsolved problem. Many past studies in recent decades proved the possibility of detecting such disease in its early stages, using a new screening panel with several urinary biomarkers. However, limited studies truly focused on the statistical correlation between the fluctuation of urinary biomarker concentrations and PDAC diagnosis status. Our study sought to demonstrate a possible correlation between biomarker concentration values in urine samples and confirmed cases of PDAC that could be used for the early diagnosis of PDAC patients. Based on the correlation of the different biomarker measurements with our investigation, we obtained data from Kaggle originally from an open access paper. We estimated odds ratios (ORs) and 95% CIs in a multinomial logistic regression model. From the analysis of p-value, LYVE1, REG1B, and TFF1 are all possible biomarkers to indicate a patient's PDAC status. Multinomial logistic regression was made to show the correlation between selected biomarkers and diagnosis. Our study suggested that a possible real correlation exists between urinary biomarkers' concentration and PDAC diagnosis status. Our model could be used to detect patients in their early disease stages to some degree.

1 INTRODUCTION

Pancreatic ductal adenocarcinoma, also known as PDAC, is the most common malignant tumor of the pancreas. It arises from cells in the ducts or ducts of the pancreas, hence its name. PDAC has a low survival rate of about 9% at 5 years and is one of the deadliest cancers in the world. Initially, the tumor may not show any signs or symptoms. However, over time, it may cause abdominal pain, nausea, and vomiting, and lead to weight loss and, in most cases, complications that eventually lead to a person's death. In terms of today's medical research developments, complete surgical removal of the tumor is the only chance to cure PDAC. If the disease is detected early, the 5-year survival rate can be increased to 70% when the tumor is still small and resectable. However, because pancreatic ductal adenocarcinoma is difficult to detect at an early stage, many patients are already at an advanced stage of cancer when diagnosed and the disease is already difficult to cure. Therefore, finding a detection method for early PDAC is an important clinical need, which may greatly improve the survival chances of patients.

Since the first risk prediction model for coronary heart disease was introduced in 1976, prediction models for various diseases, including cancer, have in the intervening decades, several tests for Pancreatic ductal adenocarcinoma have emerged, and the methods can be broadly classified into two categories. The first type is based on the use of imaging, where patients can be distinguished from pancreatic ductal adenocarcinoma by the radiomics score (rad-score) using multidetector computed tomography (MDCT), which distinguishes focal-type autoimmune pancreatitis (fAIP) from pancreatic ductal adenocarcinoma (MDCT). adenocarcinoma (PDAC) (Li et al. 2021). The second category is the use of various biomarkers to discriminate PDAC from benign pancreatic disease and healthy individuals. the source of most biomarkers is blood,

---

[a] [iD] https://orcid.org/0000-0003-0924-1786

for example, CA 19-9 and CEA in serum can be used as detection markers (Poruk et al. 2013). Mayerle, J. & Kalthoff, H. et al. generated metabolomic profiles of plasma and serum samples by gas chromatography-mass spectrometry and liquid chromatography-tandem mass spectrometry identification of 477 metabolites and selected nine plasma metabolites, all of which could be identified to distinguish pancreatic cancer from chronic pancreatitis. Serum mRNA for LGALS9 (Galectin-9) was detected to be highly expressed in PDAC patients compared to the normal pancreas and could be used as a new assay biomarker in addition to prognostic for stage IV patients (Seifert et al. 2020). With medical developments in the assay new blood testing techniques with nanoparticles are added, which are expected to distinguish PDAC patients from healthy individuals (Caputo & Caracciolo 2020). For the detection of the resectable phase of PDAC, the exlr signal of PDAC can be detected by analyzing plasma extracellular vesicle long RNA, reported (Yu et al. 2020). In addition to biomarkers in blood, there are volatile organic compounds (VOCs) in alveolar air (Princivalle et al., 2018) and MicroRNA (MiR) in pancreatic fluid (Nakamura et al. 2019) both with good sensitivity and specificity for pancreatic tumors. One more source of biomarkers in urine, and the urine biomarker set, including LYVE1, REG1A, and TFF1, has been shown to be effective in the early detection of PDAC in studies as early as 2015 (Radon et al., 2015). In a recent study, Sahni, S. and Pandya, A. R. et al. used a non-targeted urine metabolic panel to identify novel metabolite biomarker profiles for PDAC diagnosis, and six metabolites were screened and showed very high potential in the detection and diagnosis of PDAC in both early (stages I and II) and late (stages III and IV) patients.

In contrast to previously published articles on the detection of Pancreatic ductal adenocarcinoma, this study changes the detection of PDAC from traditional imaging methods (e.g., CT) or blood markers to urine biomarkers. While imaging methods are expensive and require training of dedicated personnel for testing, while many biomarkers in the blood (e.g., CA19-9 serum test) can be used to diagnose PDAC, urine instead of blood allows for completely non-invasive sampling, high volume collection, and easily repeatable measurements, with a smaller dynamic range and less complex proteome than blood. In addition, continuous ultrafiltration of blood is expected to result in the accumulation of at least some biomarkers in the urine, leading to higher concentrations. Therefore, sensitivity, specificity, positive and negative predictive values are superior to conventional methods. Second, previous studies have used urine biomarkers to detect PDAC, and they used REG1A in urine as one of the biomarkers. expression of REG1A increased with the progression of PanINs to cancer, but REG1B was highly expressed in the earliest PanINs, showing a better difference. Despite their similar performance, this study confirmed that REG1B was superior to REG1a in comparisons between control samples and stage I-IIA PDAC samples. therefore, all subsequent experiments in this study used REG1B as a component of the biomarker set.

In this paper, A secondary data analysis is conducted to study the statistical correlation between four urine bio-markers and Pancreatic ductal adenocarcinoma, and to predict whether this patient has pancreatic problems and determine whether he has pancreatic cancer at an early stage. First, the type of variable was determined after obtaining the data and corresponding to obtaining the p-value. Continuous variables were assessed using ANOVA, while categorical variables were tested with the χ^2-test. Then a polynomial logistic regression model was built to come and correlations were assessed by the above.

2 DATA SOURCE

The data was obtained from Kaggle. It is a community that allows users to find and publish datasets, explore and build models in a web-based data science environment. It is also a community where one can collaborate with other data scientists and machine learning engineers as well as participate in competitions to solve data science challenges. The data was uploaded to Kaggle by John Davis and the data was initially derived from an open-access paper by Silvana Debernardi and a colleague in PLoS medicine published on December 10, 2020.We selected this secondary data based on the relevance of the different biomarker measurements to our study and the number of reported NAs. These clinical specimens come from multiple centers, such as Barts Pancreas Tissue Bank（BPTB), University College London (UCL), University of Liverpool (LIV), Spanish National Cancer Research Center (ESP), the University of Cambridge Hospital, and the University of Belgrade. All samples were collected prior to surgery or chemotherapy treatment and were potentially age and sex matched. The data including a total of 590 biomarker panels tested on urine samples, 332 of which were collected in 2013 by Vanessa W and colleagues to study the association

between pancreatic ductal adenocarcinoma and urinary metabolic features, and the latter 258 samples were collected by Debernardi and colleagues collected additional samples at the time of the study. The majority of samples were from BPTB and LIV with 409 and 132 samples, respectively, with the remaining 8% of samples originating from other centers. The first category was 183 control samples, with no known pancreatic disease or malignancy confirmed. The second category was benign disease samples, which included 119 cases of chronic pancreatitis, 54 cases of gallbladder disease, 20 cystic lesions of the pancreas, and 15 cases with abdominal pain and gastrointestinal symptoms, for a total of 208 samples. The remaining 199 were all PDAC patients. The male to female ratio remained essentially 1:1 at 291 and 299 respectively, and the mean age of the sample is 59.1 years old.

3 RESEARCH VARIABLES

The pancreas plays a vital role in exocrine function, helping digest protein, cholesterol, and fat. Pancreatic cancer can severely impair the normal function of the pancreas. Therefore, we selected four biomarkers from urine biomarkers that are closely related to pancreas and used their values as independent variables in this study. The following are the main characteristics of the four urine biomarkers and they are all continuous variables. Variable，creatinine, as a urinary biomarker of kidney function, is a protein often used as an indicator of kidney function. In patients with PDAC, decreased protein digestion may lead to increased urinary creatinine production. Variable, lymphatic vessel endothelial hyaluronan receptor 1 (LYVE1) is a protein that may play a role in tumor metastasis, growing tumors may produce large amounts of YVLE1 for cell metastasis. Variable, REG1B stand for regenerating family member 1 beta is a Protein Coding gene. It may be associated with pancreas regeneration; damaged pancreas tissue may release large amounts of REG1B during regeneration. TFF1 is trefoil factor 1, serves as a variable that may be associated with regeneration and repair of the urinary tract. Increasing TFF1 in the gastrointestinal mucosa can help repair the damaged digestive tract. In addition, age and gender were also independent variables in this study. Age was a continuous variable and gender was used as a categorical variable, with M for male and F for female.

The dependent variable in the study is diagnosis, which is a categorical variable. There are three diagnostic classifications, 1 represents control (no pancreatic disease), 2 stands for benign hepatobiliary disease (119 of which are chronic pancreatitis), and 3 for pancreatic ductal adenocarcinoma, i.e., pancreatic cancer. Through the correlation between these four urine biomarkers and the patient's diagnosis and data analysis, the early diagnosis of PDAC will be more accurate.

Table 1: Summary of Data Collected, including the type of independent variable, the number of independent variables in the dependent variable category, the mean, standard deviation and median, and the relative p-value of each independent variables.

	Type	no pancreatic disease (N=183)	benign hepatobiliary disease (N=208)	Pancreatic ductal adenocarcinoma (N=199)	P-value
Age (years) Mean (SD) Median [Min, Max]	Continuous	56.3 (12.2) 57.0 [26.0, 89.0]	54.7 (13.3) 54.0 [26.0, 82.0]	66.2 (10.5) 67.0 [29.0, 88.0]	<0.001
Gender Female Male	Categorical	115 (62.8%) 68 (37.2%)	101 (48.6%) 107 (51.4%)	83 (41.7%) 116 (58.3%)	<0.001
Creatinine Mean (SD) Median [Min, Max]	Continuous	0.798 (0.559) 0.713 [0.0679, 3.45]	0.848 (0.616) 0.746 [0.0566, 3.34]	0.916 (0.724) 0.724 [0.0792, 4.12]	0.189
Lymphatic vessel endothelial hyaluronan receptor 1 (LYVE1) Mean (SD) Median [Min, Max]	Continuous	1.21 (1.92) 0.146 [0.000129, 8.32]	2.08 (2.37) 1.21 [0.000226, 11.0]	5.79 (3.78) 5.62 [0.00127, 23.9]	<0.001

320

Regenerating family member 1 beta (REG1B) Mean (SD) Median [Min, Max]	Continuous	41.3 (61.9) 17.6 [0.00110, 544]	64.2 (116) 20.2 [0.00280, 864]	226 (277) 23 [1.65, 1400]	<0.001
Trefoil Factor 1 (TFF1) Mean (SD) Median [Min, Max]	Continuous	169 (278) 59.8 [0.00529, 1880]	448 (646) 210 [0.0132, 4460]	1150 (1430) 723 [0.0212, 13300]	<0.001

4 STATISTICAL ANALYSIS

We performed a preliminary processing of the data. As shown in Table 1, continuous variables are expressed as mean (SD), while categorical variables are reported as absolute numbers and percentages. Firstly, the correlation between the number of pancreatic disease-healthy, benign hepatobiliary disease-noncancerous and PDAC—pancreatic cancer) is evaluated by ANOVA test while categorical variables like age and sex are done by χ^2-test. The level of significance, α was set at 0.05.

Next, we produced density plots of biomarkers and then we performed logarithmic transformation on the four urine biomarkers. At last, we established a multinomial logistic regression model to predict outcomes as we have more than 2 categorical outcomes (healthy, benign, PDAC) that cannot be put into meaningful orders.

Multinomial Logistic regression analysis was used to estimate the odds ratio (OR) and 95% confidence interval (95% CIs) are calculated. A 2-sided P-value less than 0.05 was considered significant. Data management and statistical analyses were performed using R, version 4.1.1.
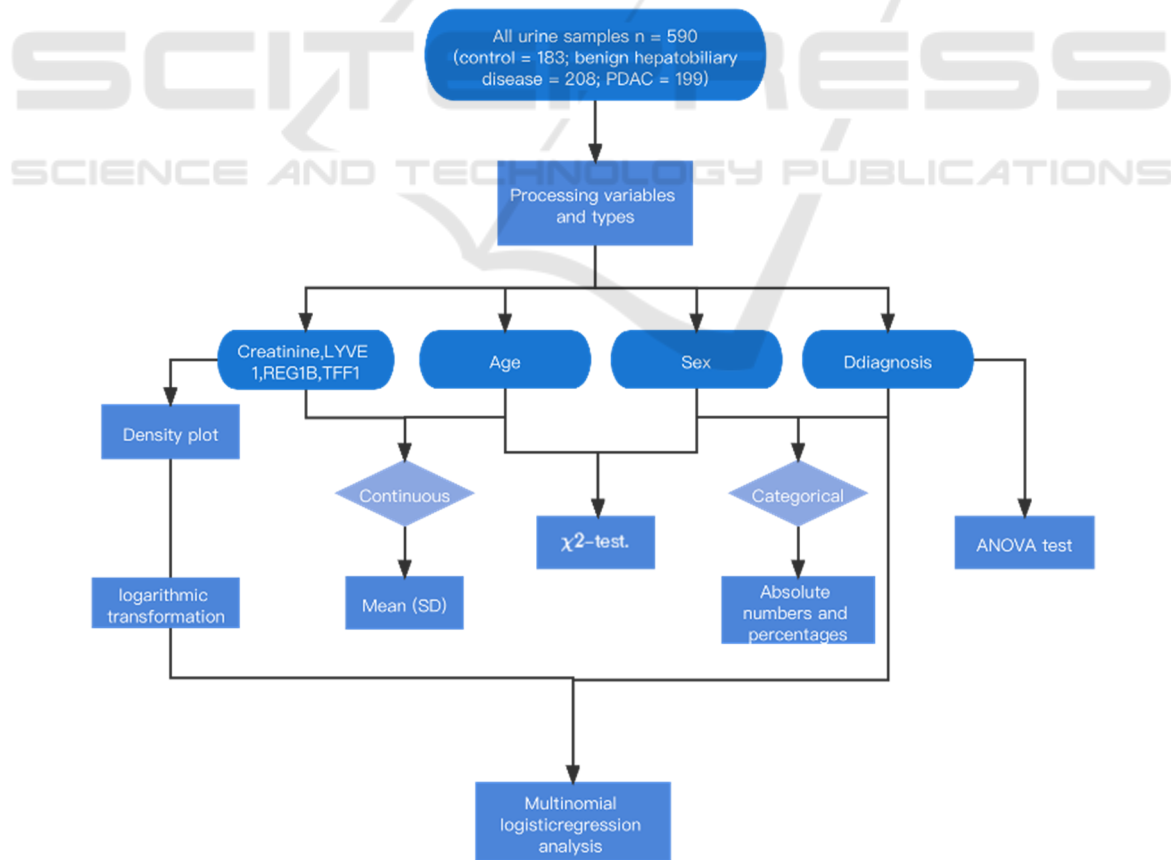

Figure 1: The process of data analysis.

5 RESULTS

According to table-1, P-value for creatinine is 0.189, which is higher than the level of significance. This indicates that there is no significant statistical correlation between levels of creatinine and diagnosis (cancerous, non-cancerous condition, healthy). Therefore, we will remove creatinine in further studies and continue to investigate the correlation between the remaining 5 independent variables and diagnosis. The p-value for age, sex, LYVE1, REG1B, and TFF1 all have p-value way smaller than the alpha level, indicating there is a correlation between these biomarkers and confounders and diagnosis.

LYVE1's mean and median are slightly different while the other 2 biomarkers having a relatively large mean median difference and SD, indicating that LYVE1 is less likely to be affected by individual differences. The large difference will be discussed further in the discussion session.

We conclude that LYVE1 may be impacted more by PDAC than the other 2, but since all 3 biomarkers show statistical correlation with the diagnosis, we should use 3 together when making predictions on a patient's health conditions. Through the density plot of 3 urine biomarkers, it was found that the distributions of the four urine biomarkers are all lognormal distributions.
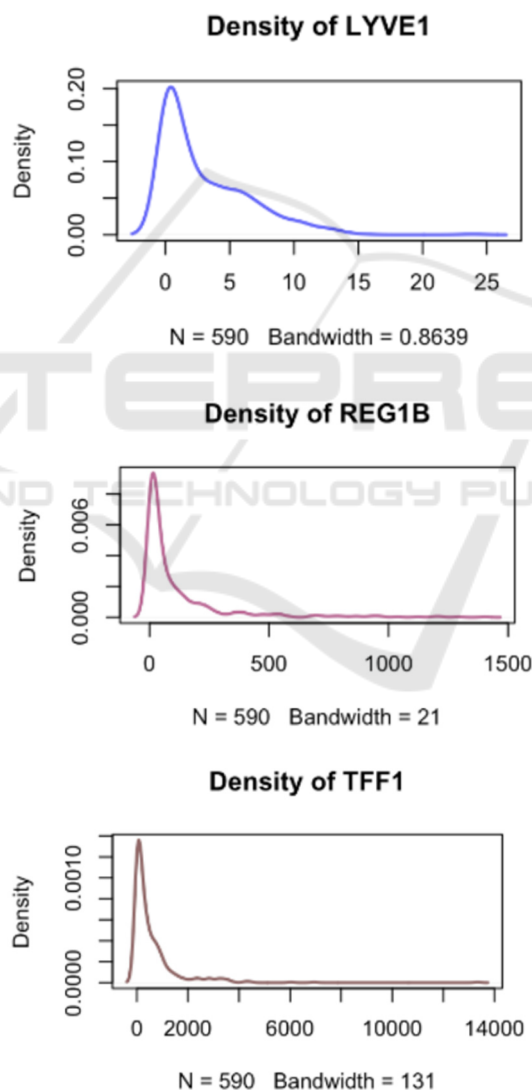

Figure 2: Diagrams of Density of 3 Urine Biomarkers.

There are also a noticeable number of outliers for samples under all 3 diagnoses, as can be seen among 3 box plots. Because we're investigating the possible correlation between biomarkers and PDAC thus no patients' samples, regardless of their underlying health conditions, should be ignored.
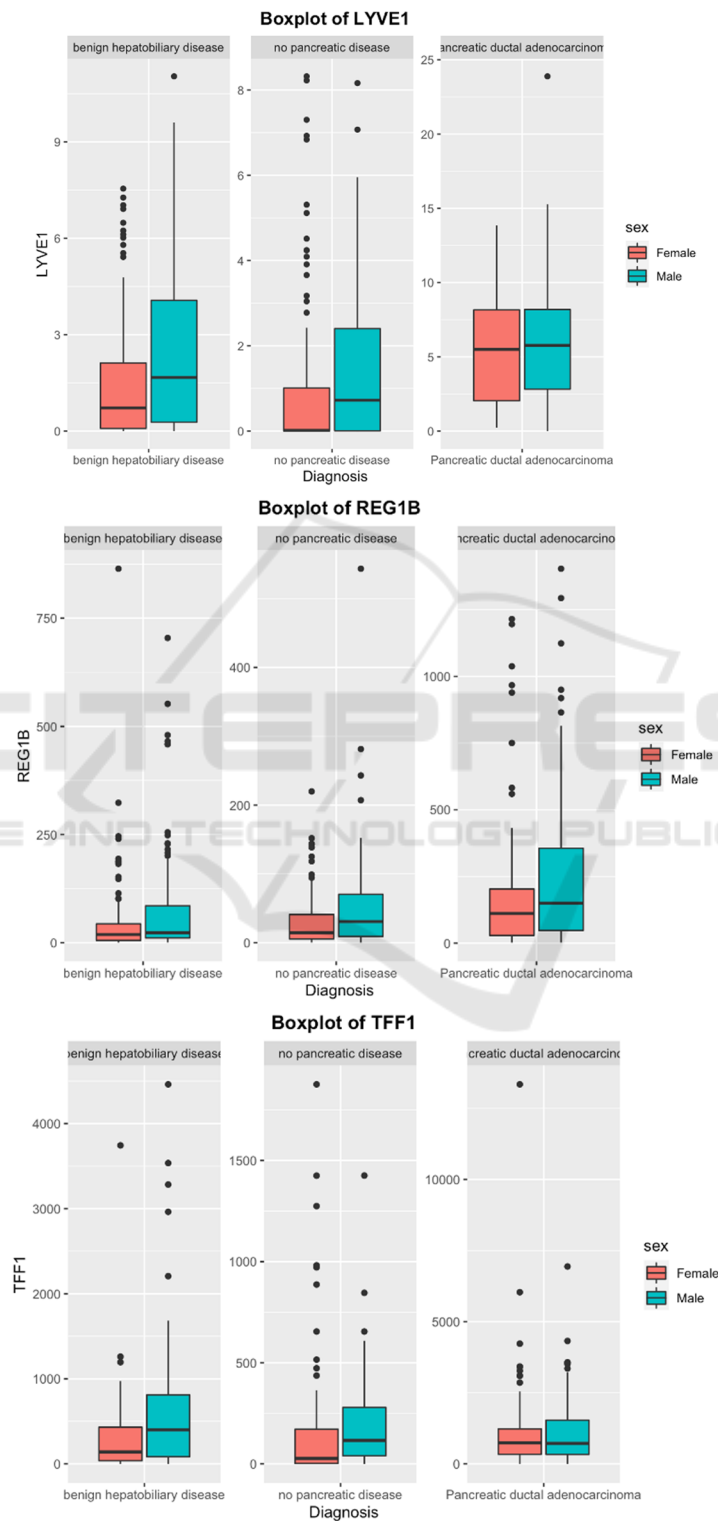

Figure 3: Mean level of different urine biomarkers displayed by different sexes under 3 diagnoses.

As represented by TFF1 and REG1B in the box plot (Figure 3), gender differences were shown under the same diagnosis: male patients showed higher levels of the 2 biomarkers than females, while both sexes showed an increasing trend in the levels of the 2 biomarkers at the diagnosis of PDAC.

We also observed an increasing trend for LYVE1 when patients were diagnosed with PDAC. However, unlike the other 2 biomarkers, the number of extremes for this marker decreased when diagnosed with PDAC.

We also noticed that for REG1B, dispersion of data increased for both females and males when diagnosed with PDAC while the dispersion for creatinine remain similarly across diagnosis and gender. Therefore, we assume that these data may not follow a normal distribution and need to use a different approach.

LYVE1 is also a significant indicator for health (OR: 1.202, 95% CI: 1.099-1.315) and PDAC (OR: 2.418, 95% CI: 1.874-3.118). REG1B was identified as a significant risk factor for health (OR: 0.834, 95% CI: 0.716–0.973) and PDAC (OR: 1.247, 95% CI: 1.004–1.550). TFF1 was identified as a significant risk factor for health (OR: 1.187, 95% CI: 1.080–1.305) and PDAC (OR: 1.166, 95% CI: 1.013–1.343).

$$\log\left(\frac{\hat{\pi}_{\text{benign hepatobiliary disease}}}{\hat{\pi}_{\text{no pancreatic disease}}}\right) = \beta_{g0} + \beta_{g1}x_{age} + \beta_{g2}x_{female} + \beta_{g3}x_{LYVE1} + \beta_{g4}x_{REG1B} + \beta_{g5}x_{TFF1}$$
$$= 0.966 - 0.012x_{age} - 0.509x_{female} + 0.184x_{LYVE1} - 0.181x_{REG1B} + 0.172x_{TFF1}$$

$$\log\left(\frac{\hat{\pi}_{\text{Pancreatic ductal adenocarcinoma}}}{\hat{\pi}_{\text{no pancreatic disease}}}\right) = \beta_{g0} + \beta_{g1}x_{age} + \beta_{g2}x_{female} + \beta_{g3}x_{LYVE1} + \beta_{g4}x_{REG1B} + \beta_{g5}x_{TFF1}$$
$$= -4.726 + 0.045x_{age} - 0.438x_{female} + 0.883x_{LYVE1} + 0.154x_{REG1B} + 0.003x_{TFF1}$$

Formula 1: Summary of Multinomial Regression Model Used.

6 DISCUSSION

Even though we concluded from our analysis on data that LYVE 1 impacted most by PDAC, we should still consider using all 3 biomarkers when analyzing urine samples collected from clinics. The differences of mean and standard deviation for REG1B and TFF1 among different diagnoses, according to several past studies that also used urine biomarkers, were a likely outcome of patients' other health conditions. Also, as we discovered that the data obtained followed a log-normal distribution, thus the extreme values we thought to be outliers are normal under such contribution.

LYVE1 itself was discovered to be a protein that played a role in the autocrine regulation of cell growth and tumor metastasis; in the meantime, the other 2 biomarkers are more associated with other organs and tissues. These urine biomarkers are not unique to PDAC, thus could be affected by cancer or disease. For instance, TFF1 is an indicator of urinary canal's self-repair, but also present in normal breast tissues; thus, situations like a male patient with prostate carcinoma, a prostate cancer, near the male urinary canal, may have a significantly higher level of TFF1 than other patients. In the meantime, breast cancer in women can also significantly increase the TFF1 levels shown by a 2017 Japanese study. Serum TFF1 and TFF3 but not TFF2 is higher in women with breast cancer than in women without breast cancer. Thus, the great differences in levels of TFF1 among individuals under the same pancreatic diagnosis may not be due to PDAC we invested in but other cancers not indicated in the study when samples were collected. Therefore, the 3 biomarkers (L, R, and T) should be analyzed together when testing and predicting the diagnosis of PDAC for susceptible patients in the clinic. Future studies could focus on distinguishing between biomarkers or find biomarkers that are uniquely correlated to PDAC when making predictions in the clinics.

7 CONCLUSIONS

In summary, we performed secondary data analysis with data obtained on Kaggle, including categorical variables, chi-square test, and estimated dominance ratio (OR) using multinomial logistic regression analysis and calculated 95% confidence intervals (95% CI). Ultimately, a correlation between the number of biomarkers in urine (continuous variable) and different diagnoses (no pancreatic disease health, benign hepatobiliary disease non-cancer, and PDAC pancreatic cancer) was successfully demonstrated and urine biomarkers (LYVE1, REG1B, and TFF1) could be used to screen for no pancreatic disease, benign The urine biomarkers (LYVE1, REG1B, and TFF1) can be used to screen for no pancreatic disease, benign hepatobiliary disease, and pancreatic ductal adenocarcinoma, cancer, providing a completely non-invasive and convenient method for detecting PDAC.

REFERENCES

Abdou, A. G., Aiad, H. A., & Sultan, S. M. (2008). pS2 (TFF1) expression in prostate carcinoma: correlation with steroid receptor status. APMIS: acta pathologica, microbiologica, et immunologica Scandinavica,

116(11), 961–971. https://doi.org/10.1111/j.1600-0463.2008.01009.x

Caputo, D., & Caracciolo, G. (2020). Nanoparticle-enabled blood tests for early detection of pancreatic ductal adenocarcinoma. Cancer letters, 470, 191–196. https://doi.org/10.1016/j.canlet.2019.11.030

Davis, V. W., Schiller, D. E., Eurich, D., Bathe, O. F., & Sawyer, M. B. (2013). Pancreatic ductal adenocarcinoma is associated with a distinct urinary metabolomic signature. Annals of surgical oncology, 20(3), S415–S423. https://doi.org/10.1245/s10434-012-2686-7

Debernardi, S., O'Brien, H., Algahmdi, A. S., Malats, N., Stewart, G. D., Plješa-Ercegovac, M., Costello, E., Greenhalf, W., Saad, A., Roberts, R., Ney, A., Pereira, S. P., Kocher, H. M., Duffy, S., Blyuss, O., & Crnogorac-Jurcevic, T. (2020). A combination of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer: A case-control study. PLoS medicine, 17(12), e1003489. https://doi.org/10.1371/journal.pmed.1003489

Ishibashi, Y., Ohtsu, H., Ikemura, M., Kikuchi, Y., Niwa, T., Nishioka, K., Uchida, Y., Miura, H., Aikou, S., Gunji, T., Matsuhashi, N., Ohmoto, Y., Sasaki, T., Seto, Y., Ogawa, T., Tada, K., & Nomura, S. (2017). Serum TFF1 and TFF3 but not TFF2 are higher in women with breast cancer than in women without breast cancer. Scientific reports, 7(1), 4846. https://doi.org/10.1038/s41598-017-05129-y

Li, J., Liu, F., Fang, X., Cao, K., Meng, Y., Zhang, H., Yu, J., Feng, X., Li, Q., Liu, Y., Wang, L., Jiang, H., Shao, C., Lu, J., & Bian, Y. (2021). CT Radiomics Features in Differentiation of Focal-Type Autoimmune Pancreatitis from Pancreatic Ductal Adenocarcinoma: A Propensity Score Analysis. Academic radiology, S1076-6332(21)00209-9. Advance online publication. https://doi.org/10.1016/j.acra.2021.04.014

Mayerle, J., Kalthoff, H., Reszka, R., Kamlage, B., Peter, E., Schniewind, B., González Maldonado, S., Pilarsky, C., Heidecke, C. D., Schatz, P., Distler, M., Scheiber, J. A., Mahajan, U. M., Weiss, F. U., Grützmann, R., & Lerch, M. M. (2018). Metabolic biomarker signature to differentiate pancreatic ductal adenocarcinoma from chronic pancreatitis. Gut, 67(1), 128–137. https://doi.org/10.1136/gutjnl-2016-312432

Nakamura, S., Sadakari, Y., Ohtsuka, T., Okayama, T., Nakashima, Y., Gotoh, Y., Saeki, K., Mori, Y., Nakata, K., Miyasaka, Y., Onishi, H., Oda, Y., Goggins, M., & Nakamura, M. (2019). Pancreatic Juice Exosomal MicroRNAs as Biomarkers for Detection of Pancreatic Ductal Adenocarcinoma. Annals of surgical oncology, 26(7), 2104–2111. https://doi.org/10.1245/s10434-019-07269-z

Poruk, K. E., Gay, D. Z., Brown, K., Mulvihill, J. D., Boucher, K. M., Scaife, C. L., Firpo, M. A., & Mulvihill, S. J. (2013). The clinical utility of CA 19-9 in pancreatic adenocarcinoma: diagnostic and prognostic updates. Current molecular medicine, 13(3), 340–351. https://doi.org/10.2174/1566524011313030003

Princivalle, A., Monasta, L., Butturini, G., Bassi, C., & Perbellini, L. (2018). Pancreatic ductal adenocarcinoma can be detected by analysis of volatile organic compounds (VOCs) in alveolar air. BMC cancer, 18(1), 529. https://doi.org/10.1186/s12885-018-4452-0

Radon, T. P., Massat, N. J., Jones, R., Alrawashdeh, W., Dumartin, L., Ennis, D., Duffy, S. W., Kocher, H. M., Pereira, S. P., Guarner posthumous, L., Murta-Nascimento, C., Real, F. X., Malats, N., Neoptolemos, J., Costello, E., Greenhalf, W., Lemoine, N. R., & Crnogorac-Jurcevic, T. (2015). Identification of a Three-Biomarker Panel in Urine for Early Detection of Pancreatic Adenocarcinoma. Clinical cancer research: an official journal of the American Association for Cancer Research, 21(15), 3512–3521. https://doi.org/10.1158/1078-0432.CCR-14-2467

Sahni, S., Pandya, A. R., Hadden, W. J., Nahm, C. B., Maloney, S., Cook, V., Toft, J. A., Wilkinson-White, L., Gill, A. J., Samra, J. S., Dona, A., & Mittal, A. (2021). A unique urinary metabolomic signature for the detection of pancreatic ductal adenocarcinoma. International Journal of Cancer, 148(6), 1508. https://doi.org/10.1002/ijc.33368

Seifert, A. M., Reiche, C., Heiduk, M., Tannert, A., Meinecke, A. C., Baier, S., von Renesse, J., Kahlert, C., Distler, M., Welsch, T., Reissfelder, C., Aust, D. E., Miller, G., Weitz, J., & Seifert, L. (2020). Detection of pancreatic ductal adenocarcinoma with galectin-9 serum levels. Oncogene, 39(15), 3102–3113. https://doi.org/10.1038/s41388-020-1186-7

Yu, S., Li, Y., Liao, Z., Wang, Z., Wang, Z., Li, Y., Qian, L., Zhao, J., Zong, H., Kang, B., Zou, W. B., Chen, K., He, X., Meng, Z., Chen, Z., Huang, S., & Wang, P. (2020). Plasma extracellular vesicle long RNA profiling identifies a diagnostic signature for the detection of pancreatic ductal adenocarcinoma. Gut, 69(3), 540–550. https://doi.org/10.1136/gutjnl-2019-318860