San Jose State University
SJSU ScholarWorks

Master's Projects | Master's Theses and Graduate Research

Fall 12-19-2020

The Use of Evidential Reasoning Model with Biomarkers in Pancreatic Cancer Prediction

Qianhui Fan
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_projects

 Part of the Artificial Intelligence and Robotics Commons, and the Other Computer Sciences Commons

Recommended Citation

Fan, Qianhui, "The Use of Evidential Reasoning Model with Biomarkers in Pancreatic Cancer Prediction" (2020). *Master's Projects*. 966.
DOI: https://doi.org/10.31979/etd.cdzu-9qy7
https://scholarworks.sjsu.edu/etd_projects/966

This Master's Project is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Projects by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

DocuSign Envelope ID: 55FE336B-F760-4058-91D0-6C691D181C1D

Qianhui Fan

has passed the defense for the project

The Use of Evidential Reasoning Model with Biomarkers in Pancreatic Cancer Prediction

DocuSigned by:
Leonard Wesley
D9F85983BBCE4B1...

12/11/2020

Advisor's Signature — Date

DocuSigned by:
Yulia Newton
65A3150EE8C7460...

12/12/2020

Committee Member's Signature — Date

DocuSigned by:
William Andreopoulos
3CFA6E56A8E74FF...

12/12/2020

Committee Member's Signature — Date

NOTE: The advisor should send the final report to the graduate coordinator so that the student can be cleared for graduation

The Use of Evidential Reasoning Model with Biomarkers in Pancreatic Cancer Prediction

CS298 Graduation Research

Presented to

Department of Computer Science

San José State University

In Partial Fulfillment

Of the Requirements for the CS298

By

Qianhui Fan

November, 2020

ABSTRACT

In this project, an evidential reasoning model is built to amalgamate factors that could be used in early detection of pancreatic cancer. Our machine learning model outputs a probability of a given patient having prostate cancer based on various input variables. These variables include health history factors, such as smoking and medical history, technical artifacts, such as biopsy sequencing technology, and genomic biomarkers such as mutational, transcriptional and methylomic profiles, cfDNA, and copy number variation. The dataset used in this project is a part of The Cancer Genome Atlas (TCGA) project and was collected from the National Cancer Institute (NIH) Genomic Data Commons (GDC). The model is tested by varying input propositions and probability mass functions of input frames to create different combinations of input factors. Baseline prediction results in (0.084, 0.19) of not having pancreatic cancer. Prediction results were compared to the baseline prediction and a set of positive control expectations. For example, medium to high smoking history, medium to high drinking history with some cancer history will increase the posterior belief of a patient having pancreatic cancer to (0.091, 0.208). Presence of prognostic biomarkers will also increase the support for having pancreatic cancer, having medium impact DNA methylation and medium impact mRNA expression can lead the belief of having pancreatic cancer increase to (0.167, 0.273).

Keywords – pancreatic cancer, evidential reasoning, biomarkers

TABLE OF CONTENTS

LIST OF TABLES

LIST OF Figures

I. Introduction

Pancreatic cancer is the result of uncontrolled division of malignant cells in the pancreas. With its late and few symptoms, pancreatic cancer is ranked as the fourth leading cause of cancer-related mortality worldwide [1]. The death toll recorded in 2015 from pancreatic cancer alone, among all the different types of cancer, was 411,600 deaths globally [2]. In the United States, it is the third-most-common cause of death [3]. It is more prevalent within the developed countries, accounting for about 70% of new cases recorded in 2012 [4].

It is quite a rare occurrence for patients below the ages of 40 to be diagnosed with this disease condition, while more than half of patients diagnosed with pancreatic adenocarcinoma are over 70 [5]. Risk factors associated with it are some rare genetic predisposition conditions, tobacco smoking, obesity and diabetes.

Although expressed in various types, pancreatic cancer can be categorized into two groups, the exocrine and endocrine groups.

- **Exocrine group**: This group accounts for the vast majority of pancreatic cancer cases and occurs within the part of the pancreas responsible for digestive enzyme production, called the exocrine component. And among all types within this group, pancreatic ductal adenocarcinoma (PDAC) constitute over 90%, making it by far the most common type of pancreatic cancer cases [1]. This is quite the case despite the fact that the tissue it arises from constitutes only 10% of the pancreas cell volume, because it is just the duct within the pancreas [6]. The next most common type, representing about 5% of the exocrine group of cancers is called the acinar cell carcinoma[7]. Accounting for 1% of pancreatic

cancer cases is another type known as Cystadenocarcinomas [7]. Compared to other types of exocrine cancers, it has a better prognosis.

- **Endocrine group**: As for the second group, endocrine group of cancers, they account for the small minority types of cancer and are also called pancreatic neuroendocrine tumors (PanNet) [8]. These diverse groups of sometimes benign tumors arise from the body's neuroendocrine cells, which serves to integrate the endocrine and nervous systems. There are two types of endocrine group of cancers, the functioning and nonfunctioning types. The difference between both is the amount of hormones they secret. Functioning types, secret hormones in large quantities that often results in serious symptoms, favoring early detection. As for the second type, nonfunctioning PanNet, because they do not secrete hormones in sufficient quantities, it doesn't result in overt clinical symptoms and therefore are only diagnosed after it has spread to other body organs [8].

The prognosis for pancreatic adenocarcinoma is very poor, as 25% survive one year after diagnosis and 5% live for five years; if detected early, it increases to 20% [4][9]. As for neuroendocrine types of cancer, they've got a better survival rate, of which after five years of diagnosis, 65% are living depending on the type of tumor [4]. Aside from the fact that the symptoms of PDAC do not usually appear at an early stage, they are not individually distinctive. Symptoms vary according to the cancer's location in the pancreas. While tumors in the body and tail of the pancreas express painful symptoms, those at the head typically cause dark urine, jaundice, loss of appetite and so on.

To confirm diagnosis and its resectability, medical imaging techniques such as endoscopic ultrasound and computed tomography (CT scan) are employed. Abdominal ultrasound tends to miss small tumors but is effective in identifying cancers that have spread to the liver [10]. Pancreatic cancer is tackled using radiotherapy, chemotherapy, palliative care and undergoing surgery.

II. Background

Pancreatic cancer has such an alarming death rate because about 80-85% of patients are diagnosed when the disease has gotten to its late stage, and often already spread to other organs [11]. At such a stage, curative surgical resection is not possible. However, it takes a period of one or two decades between chronic pancreatitis diagnosis and an overt tumor [12]. Capitalizing on this long latency period will really curb the mortality rate when diagnosed early and treatment is commenced.

A. Detection with Biomarker

Compared to other tumors, such as those for lungs, breast, cervix and colon, the screening program for PDAC remains a challenge. Barrier hindering its progress includes the specificity of the test. This has led to lots of false cases because it requires a high performing screening test that's having a very high sensitivity. As a result of this, there has been an ongoing intensive effort aimed at discovering pancreatic cancer-associated biomarkers meant to assist in early detection, diagnosis and predicting response to treatment. These efforts have focused on serum biomarkers.

A biomarker is any substance, molecule that is measured in the body and that can be used to predict the incidence or outcome of a disease. The absence of reliable biomarkers to capture

the early development of this disease, oftentimes, has resulted in patients being diagnosed when it's at an advanced or critical stage. So far, the US Food and Drug Administration (FDA) has approved only one biomarker associated with pancreatic cancer treatment, serum protein-carbohydrate antigen 19-9 (CA19-9 or sialylated Lewis antigen).

Unfortunately, CA19-9 has not really been effective for early detection due to its low specificity and sensitivity. With a sensitivity of 79-81% and a specificity of 82-90%, it has a poor predictive value in asymptomatic patients [13]. Not only that other types of cancer can lead to elevated CA19-9 levels but about 10% of the Caucasian population lacks CA19-9 on their red blood cells [14]. Therefore, up to date, there's no reliable biomarker approved for early diagnosis of pancreatic cancer in clinical settings.

Another vital role biomarkers play is that they can be used to monitor treatment efficacy and any resurgence of resected tumors. In fact, this is the function CA 19-9 currently serves, as they're used to provide valuable information regarding the patient's response to pancreatic cancer treatment. After the patient has been diagnosed of having pancreatic cancer, the healthcare team takes a baseline (initial) measurement of CA 19-9. As the patient commences treatment, the CA 19-9 levels are continuously measured till completion of treatment. The changes recorded in these levels enable the doctors to ascertain if the tumor is growing, diminishing or staying the same size.

Another biomarker that’s often under investigation is known as KRAS mutations, which occur in Kirsten Rat Sarcoma Viral Oncogene Homolog (KRAS) gene. Although these occur frequently in pancreatic cancer, its diagnostic accuracy is not sufficient enough for clinical

utilization. This is the result of non-specificity of KRAS mutations as they are observed in many tumor types. From the extensive studies carried out, low levels of cell-free circulating DNA in serum were observed, thereby limiting the use of non-invasive assays for clinical diagnostics [15]. Other biomarkers under investigation include various microRNAs, Macrophage inhibitory cytokine 1 (MIC1), Peptidylglycine Alpha-Amidating Monooxygenase (PAM 4), Glypican (GPCX), Osteopontin (SPP1), DNA methylation and RNA transcriptional profiles, copy number variation information, and signaling pathway-level aberrations.

B. Sequencing

Advances in DNA and RNA sequencing technologies provided scientists and clinicians with ability to understand the molecular differences between normal and malignant tissues. Sequencing is the method of determining the sequence arrangement of DNA or RNA nucleotides. Different sequencing strategies have assisted with revealing new transformations tumor cells.

Next-generation sequencing technology (NGS) has brought many important discoveries to cancer research and treatment. Finding the tumor driver genes and related signal transduction pathways has become a new strategy for current clinical tumor assessment and treatment. For example, study [16] includes a total of 22 patients with pancreatic cancer and 42 genes and 61 loci were detected by NGS to have DNA mutations. The mutation rate of TP53, KRAS, CDKN2A, SMAD4 genes is significantly higher than other genes. Researchers in [17] extract DNA from plasma samples and use NGS analysis methods to detect allelic mutation frequencies. The three most common KRAS mutations in blood samples of pancreatic cancer were screened by droplet digital PCR (dPCR). Through multivariate analysis, it is revealed that ctDNA in the

blood is a prognostic biomarker for patients with pancreatic cancer and is related to the development of the disease and the degree of tumor differentiation. As the tumor progresses, it sheds some of its DNA into the bloodstream. Blood-based assays can be utilized to detect DNA that came from malignant cells in pancreas.

Advances in next-generation sequencing technologies also promote the development of whole-genome-sequencing (WGS) and whole-exome-sequencing (WES). While WES is capable of detecting mutations in the tumor exome, WGS can be used to detect all somatic mutations in tumor samples.

C. Machine Learning

Massive cancer data has been generated as recent advances have taken place in the field of medicine. However, predicting an outcome correctly still remains challenging and fascinating for doctors and biomedical researchers. This has made Machine Learning (ML) methods an informative tool for medical sciences. ML techniques can model trends and associations from diverse data sets while accurately predicting diagnostic, therapeutic, and prognostic outcomes for each specific cancer type.

Prognosis and prediction of diseases with various techniques and feature selection algorithms have widely applied in the last two decades [18]. Most of these initiatives use ML approaches to model cancer development and detect useful factors subsequently used in ML classifications. These methods use clinical, histological, and genomic data to model tumorigenesis. For almost three decades, leading ML techniques like Artificial Neural Networks (ANNs) and Decision Trees (DT) were used in cancer detection. Based on the investigation done by [19], more than 7,510 papers have been published to date on the topic of ML and cancer.

Most publications use ML algorithms and incorporate heterogeneous tumor detection data and cancer prediction/forecast data. In the past decade, a trend has developed towards cancer detection and prediction, and other controlled teaching techniques have been noted. All these classification algorithms have been used across cancer types.

D. Research on evidential reasoning

When multimillion-dollar decisions are about to be made, hinging such decisions on results obtained from traditional decision-analytic and probabilistic approaches is quite risky. This is because, in practice, the true probability distribution of all factors might never be known even if the requirements and assumptions the calculus requires are known. For better decision making, the evidential-based approach was developed, with the works of Arthur Dempster and Glenn Shafer serving as its basis [20]. The term "evidence" denotes that data are best treated such that they either tend to support or refute to varying degrees probabilistic arguments of different alternatives. Unlike traditional probabilistic techniques that require point estimations, this technique makes use of interval estimations. A great advantage is that the prior data required is more intuitive and easy to obtain.

This project is an extension of former projects completed by Chandratre, and Sharghi [21][22]. The genomic dataset used in their studies are taken from the National Cancer Institute (NIH) Genomic Data Commons (GDC). Not only the outcome of the machine learning model with Support Vector Classifier (SVC) is used, their evidential reasoning (ER) model also takes other factors that may affect the predictions. For example, NGS technology used and sequencing reads could affect the outcome of machine learning prediction. The list of all inputs of their ER model is: NGS technology, sequence read, smoking history, drinking history, family medical

history, patient medical history, biopsy site cell result, and amount of genetic material. Experiments done by adjusting the proposition and mass of input frames is consistent with intuitive judgement of the prediction [21][22].

Chandratre built on Sharghi's project by creating an improved version of machine learning classifier by taking all projects in the NIH GDC portal into account. This includes 22,872 genes and 3,142,246 total mutations related to these genes [22]. The improved model also considered the lethality of mutations, the impact of each mutation on the mutated protein can be classified as VEP, SIFT, and PolyPhen in NIH GDC portal. Sharghi's model reported a high SVC prediction accuracy of ~91%, but only used gene-mutation combinations occurred in 185 cases of the TCGA-PAAD dataset. Result of the improved model shows a low prediction accuracy ~85% with small standard deviation by considering VEP impact [22].

E. Remaining technical gaps

Current diagnostic approaches of detecting pancreatic cancer neglect to analyze the pancreatic disease in its early phase, bringing about a reduction in this current condition's survival rate. In addition, in the case of a fatal disease like pancreatic cancer, it has been proven unreliable and imprecise to count on the limited resources available for analysis, including the Artificial Intelligence (AI)-assisted genomic data analysis and the few biomarkers known to be causative.

The reasoning for this is that several variables contribute to the cell mutation in pancreas. These reasons might be personal habits (dietary habits, use of cigarettes, liquor intake, and so forth) or even inherited (hereditary conditions caused by certain germline genetic mutations). All these factors along with the available analytical data can be ambiguous, inconsistent, uncertain,

or even deficient. In correlation, if data regarding age and race of higher-risk people were to be considered, maybe only a certain group of the larger population will be ultimately diagnosed [21].

To maximize the chance of correctly prognosis of pancreatic cancer, all different factors that could contribute to cancer development should be taken into consideration. Evidential reasoning model and the Belief Function (BF) thus become a crucial part to formulate rational and sound conclusions in relation to the probability of developing pancreatic cancer. This is accomplished more straightforwardly and less confining than standard analytical and probabilistic strategies [20].

III. Approach

With a considerable development of research suggesting the presence of change of certain genetic material may lead to malignant tumors in their pancreas. This project tries to bring the gap in pancreatic cancer research to make predictions based not only on environmental factors, and factors that can affect machine learning outcome, but also try to incorporate biomarkers in addition to the work done by Chandratre and Sharghi.

The biomarkers used in this project are namely: novel/rare mutation, pathway mutation, DNA methylation, cfDNA/cfDNA methylation, mRNA expression, exosome-miRNA, circulating tumor cell (CTC), and copy number variation (CNV). (Check Appendix C to J for more information regarding each biomarker).

Each biomarker is used as an input frame in the project, frames are mutually exclusive random variables. The collection of frames is called a galley. This project contains a galley file

that records information for each frame: frame name, frame type, propositions, input node, output node, and compatrel relationship between frames. Proposition is a value of the frame, the set of propositions in a frame is mutually exclusive and exhaustive in the BF world. Frame type indicates if the propositions are discrete or continuous. For example, the outcome of a machine learning prediction can be PC or NOT_PC, they are discrete propositions of ML_PREDICTION frame. For each frame, a probability value representing a degree of belief denoted as "mass" is assigned to the subset of propositions of the frame. The baseline proposition and mass can be found in Appendix A Table I. For example, the baseline mass 0.5 is assigned to LOW_IMPACT proposition of the DNA methylation frame, it represents there's 0.5 confidence from the evidence that supports DNA methylation frame has low impact. The remaining mass of 0.5 means DNA methylation could be either LOW_IMPACT, MEDIUM_IMPACT, or HIGH_IMPACT. In another case, cfDNA/cfDNA methylation frame has a vacuous mass of 1.0 assigned to the proposition set (LOW_IMPACT, MEDIUM_IMPACT, HIGH_IMPACT), which means that mass 1.0 is assigned to a disjunction of of the set of all propositions in the frame.

Although there are many studies on biomarker, the current researches show different degrees of understanding of the role of each biomarker in pancreatic cancer prediction. A discount value is assigned to each frame to decrease the impact of conviction for the predisposition of this frame to pancreatic cancer. For example, CNV is assigned a higher discount rate, because research papers like [23] mention that the correlation between CNV and pancreatic cancer need to be elucidated by further investigation. Compared to NGS technology which provides concrete data, smoking history or drinking history has a higher discount rate, because they could be inaccurate and subjective. The value of discount rate is selected based on

subjective estimation. When a discount rate of 0.7 is applied, only 70% of the mass will be used in the calculation.

The evidential model sets up the conviction of predicting pancreatic malignancy relying upon the combination of propositions and their masses as inputs. Frames are combined based on Dempster's rule of combination. Some frames map directly to calculating the chance of having pancreatic cancer, while a disjunction of related frames can also form intermediate nodes. These intermediate nodes aggregate the evidence provided by the input nodes. For example, an intermediate node is created using BIOPSY_SITE_CELL_RESULT and AMOUNT_GEN_MATERIAL as direct input nodes, since the different amount of material taken at a biopsy site may affect sequencing and ML outcome. There's also an intermediate node created between SMOKING_HISTORY, DRINKING_HISTORY, FAMILY_MED_HISTORY, and PATIENT_MED_HISTORY, because as [24] has shown, there is a high co-occurrence between smoking and drinking behaviors, and smoking and drinking habits will tend to result in medical conditions. Figure 1 and Figure 2 shows all the input frames and intermediate nodes for this project.

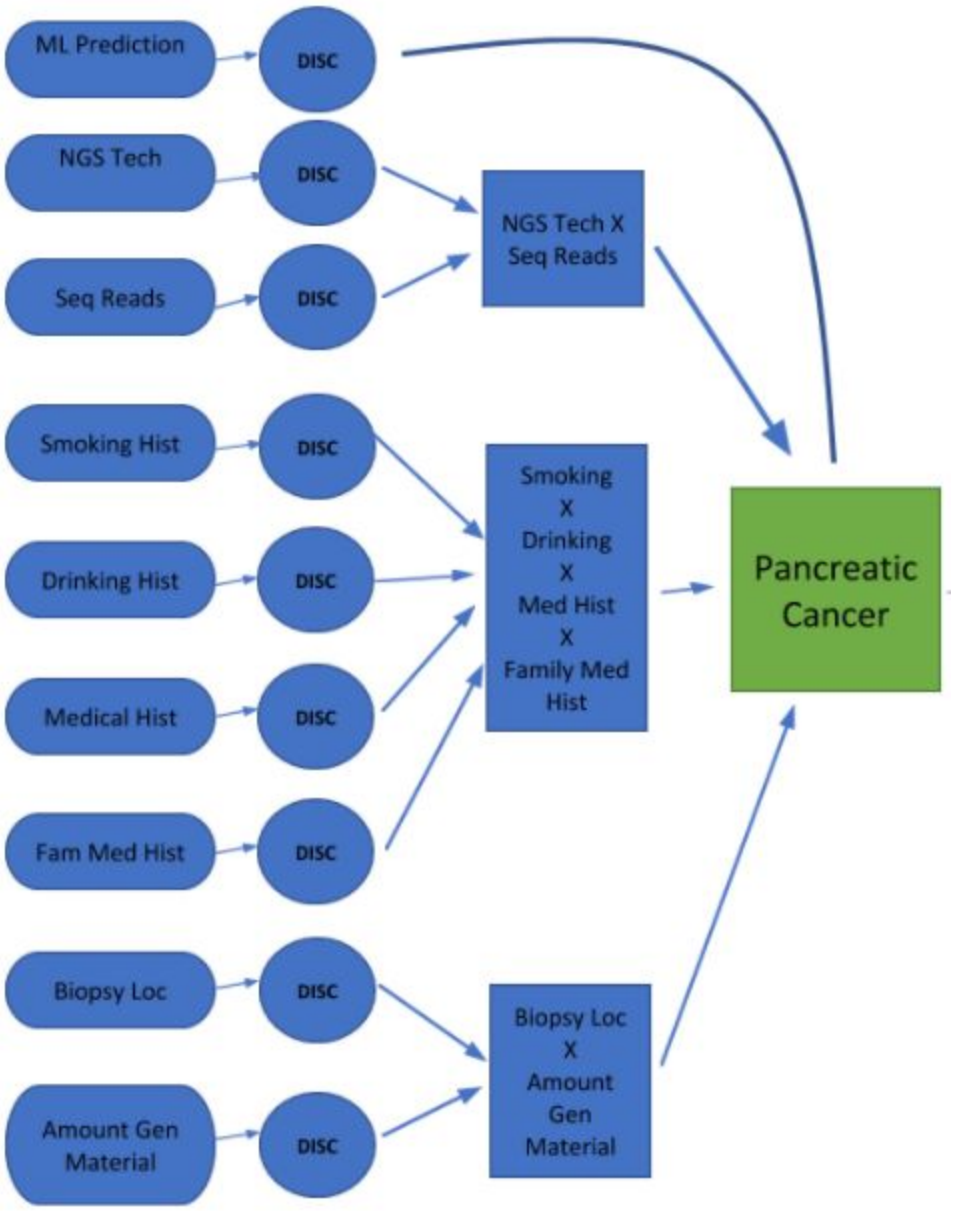

Figure 1. Evidential Reasoning Model (left)

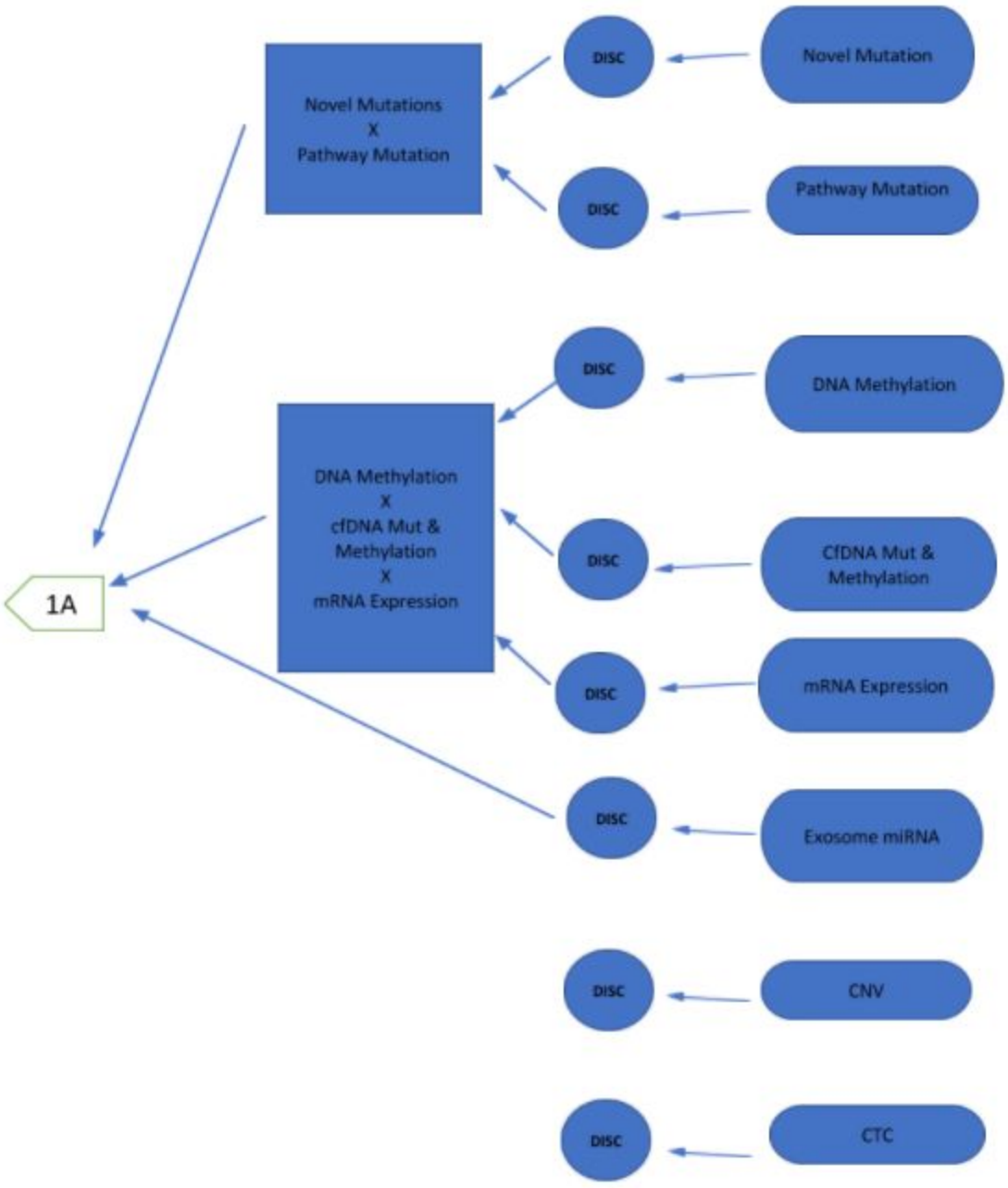

Figure 2. Evidential Reasoning Model (right)

IV. EXPERIMENTS

Our evidential reasoning model is tested by adjusting the proposition and mass of the input frames to see if the prediction is expected. The ER prediction using all frames with baseline inputs mass, proposition, and discount rate gives the following output:

Belief Of Having Pancreatic Cancer Lies Between: (0.084, 0.19) (0)|**--------|(1)
Belief Of Not Having Pancreatic Cancer Lies Between: (0.809, 0.915) (0)|--------**|(1)

The following experiments aim to demonstrate how adjusting input factors affect the final prediction. A updating program is used to update the mass and propositions for input frames, information about how the program is used can be found in Appendix K.

A. ER Experiment 1

Since there's an intermediate node created for BIOPSY_SITE_CELL_RESULT and AMOUNT_GEN_MATERIAL. This experiment aims to demonstrate how change in biopsy sites, and the amount of genetic material taken from the biopsy site affects the result of the prediction. Adjust the two frames and keep the values of all other input parameters the same as the baseline values. The prediction results are shown in the following table:

Biopsy site		Amount of genetic material		Belief Of Having Pancreatic Cancer Lies Between:	Belief Of Not Having Pancreatic Cancer Lies Between:
Proposition	Mass	Proposition	Mass		
NEAR_PAN_IRREG	0.7	SMALL	0.7	(0.09, 0.205) (0)|***-------|(1)	(0.794, 0.909) (0)|-------***|(1)
NEAR_PAN_IRREG	0.3	MEDIUM	0.3	(0.116, 0.226) (0)|-**-------|(1)	(0.773, 0.883) (0)|-------**-|(1)
NEAR_PAN_IRREG	0.7	MEDIUM	0.7	(0.29, 0.381) (0)|--**------|(1)	(0.619, 0.709) (0)|------**--|(1)

NEAR_PAN_IRREG	0.7	HIGH	0.7	(0.291, 0.381) (0)|--**------|(1)	(0.619, 0.708) (0)|------**--|(1)
NOT_NEAR_PAN_IRREG	0.7	SMALL	0.7	(0.09, 0.202) (0)|***-------|(1)	(0.797, 0.909) (0)|-------***|(1)
NOT_NEAR_PAN_IRREG	0.7	LARGE	0.7	(0.089, 0.202) (0)|***-------|(1)	(0.797, 0.91) (0)|-------***|(1)
NEAR_PAN_REG	0.7	SMALL	0.7	(0.039, 0.09) (0)|*---------|(1)	(0.909, 0.96) (0)|---------*|(1)
NEAR_PAN_REG	0.7	LARGE	0.7	(0.09, 0.204) (0)|***-------|(1)	(0.795, 0.909) (0)|-------***|(1)
NOT_NEAR_PAN_REG	0.7	SMALL	0.7	(0.016, 0.04) (0)|*---------|(1)	(0.96, 0.983) (0)|---------*|(1)
NOT_NEAR_PAN_REG	0.7	LARGE	0.7	(0.09, 0.202) (0)|***-------|(1)	(0.797, 0.909) (0)|-------***|(1)

Table 1. Prediction result of various input parameters for biopsy site and amount of material

From the table above we can see that when the biopsy site is NEAR_PAN_IRREG, NEAR_PAN_REG, or NOT_NEAR_PAN_REG, change the amount of genetic material from SMALL to MEDIUM or HIGH, will lead to an increase in the belief of having pancreatic cancer. When the biopsy site is NOT_NEAR_PAN_IRREG with mass 0.7, change the amount of material from SMALL to LARGE, the prediction doesn't change. When the biopsy site is NEAR_PAN_IRREG with MEDIUM amount of genetic material, increasing the mass from 0.3 to 0.7 also leads to an increase in the belief of having pancreatic cancer.

B. ER Experiment 2

An intermediate node is created between SMOKING_HISTORY, DRINKING_HISTORY, FAMILY_MED_HISTORY, and PATIENT_MED_HISTORY. This

experiment tests how varying mass distribution and proposition affects the result of prediction. Adjust the four frames and keep the values of all other input parameters the same as the baseline values. The prediction results are shown in the following table:

Smoking history		Drinking history		Family med. history		Patient med. history	
Proposition	Mass	Proposition	Mass	Proposition	Mass	Proposition	Mass
LOW	0.3	LOW	0.3	NO_CANCER	0.3	NO_CANCER	0.3
Belief Of Having Pancreatic Cancer Lies Between: (0.092, 0.207)(0)|***-------|(1) Belief Of Not Having Pancreatic Cancer Lies Between: (0.792, 0.907)(0)|-------***|(1)							
LOW	0.7	LOW	0.7	NO_CANCER	0.7	NO_CANCER	0.7
Belief Of Having Pancreatic Cancer Lies Between: (0.085, 0.191)(0)|**--------|(1) Belief Of Not Having Pancreatic Cancer Lies Between: (0.808, 0.914)(0)|--------**|(1)							
LOW	0.7	LOW	0.7	SOME_CANCER	0.7	SOME_CANCER	0.7
Belief Of Having Pancreatic Cancer Lies Between: (0.092, 0.207)(0)|***-------|(1) Belief Of Not Having Pancreatic Cancer Lies Between: (0.792, 0.907)(0)|-------***|(1)							
MEDIUM	0.3	MEDIUM	0.3	SOME_CANCER	0.3	SOME_CANCER	0.3
Belief Of Having Pancreatic Cancer Lies Between: (0.091, 0.208)(0)|***-------|(1) Belief Of Not Having Pancreatic Cancer Lies Between: (0.791, 0.908)(0)|-------***|(1)							
MEDIUM	0.7	MEDIUM	0.7	SOME_CANCER	0.7	SOME_CANCER	0.77
Belief Of Having Pancreatic Cancer Lies Between: (0.092, 0.208)(0)|***-------|(1) Belief Of Not Having Pancreatic Cancer Lies Between: (0.791, 0.907)(0)|-------***|(1)							
HIGH	0.3	HIGH	0.3	CANCER	0.3	CANCER	0.3
Belief Of Having Pancreatic Cancer Lies Between: (0.092, 0.207)(0)|***-------|(1) Belief Of Not Having Pancreatic Cancer Lies Between: (0.792, 0.907)(0)|-------***|(1)							
HIGH	0.7	HIGH	0.7	CANCER	0.7	CANCER	0.7

Belief Of Having Pancreatic Cancer Lies Between: (0.109, 0.223)(0)|-**-------|(1) Belief Of Not Having Pancreatic Cancer Lies Between: (0.776, 0.89)(0)|-------**-|(1)

Table 2. Prediction result of various input parameters for smoking history, drinking history, family medical history, and patient medical history

From the above table, we can see that when propositions are LOW or NO_CANCER, increasing the mass of the four propositions from 0.3 to 0.7, will lead to an increase in the belief of not having pancreatic cancer. When the propositions are of high correlation will cause pancreatic cancer, increasing the mass from 0.3 to 0.7 increases the prediction of probability of having pancreatic cancer. When the propositions are of MEDIUM or SOME_CANCER, change in mass doesn't lead to a change in the prediction. In addition, we can observe that when keeping mass the same for all four propositions, changing propositions from MEDIUM to HIGH lead to an increase in the chance of having pancreatic cancer. The results match with expectation.

C. ER Experiment 3

Since there's an intermediate node created for NOVEL_MUTATION and PATHWAY_MUTATION. This experiment aims to demonstrate how novel mutation, pathway mutation affects the result of prediction. Adjust the two frames and keep the values of all other input parameters the same as the baseline values. The prediction results are shown in the following table:

Novel mutation		Pathway mutation		Belief Of Having Pancreatic Cancer Lies Between:	Belief Of Not Having Pancreatic Cancer Lies Between:
Proposition	Mass	Proposition	Mass		
NOT_NOVEL	0.7	LOW_ IMPACT	0.7	(0.09, 0.203) (0)|***-------|(1)	(0.796, 0.909) (0)|-------***|(1)

NOVEL	0.7	LOW_ IMPACT	0.7	(0.089, 0.201) (0)|***-------|(1)	(0.798, 0.91) (0)|-------***|(1)
NOT_NOVEL	0.3	MEDIUM_ IMPACT	0.3	(0.09, 0.204) (0)|***-------|(1)	(0.795, 0.909) (0)|-------***|(1)
NOT_NOVEL	0.7	MEDIUM_ IMPACT	0.7	(0.09, 0.202) (0)|***-------|(1)	(0.797, 0.909) (0)|-------***|(1)
NOVEL	0.7	MEDIUM_ IMPACT	0.7	(0.09, 0.203) (0)|***-------|(1)	(0.796, 0.909) (0)|-------***|(1)
NOT_NOVEL	0.7	HIGH_ IMPACT	0.7	(0.089, 0.205) (0)|***-------|(1)	(0.794, 0.91) (0)|-------***|(1)
NOVEL	0.7	HIGH_ IMPACT	0.7	(0.09, 0.202) (0)|***-------|(1)	(0.797, 0.909) (0)|-------***|(1)

Table 3. Prediction result of various input parameters for novel mutation and pathway mutation

Compared to the baseline prediction when propositions are NOT_NOVEL and LOW_IMPACT with both masses being 0.5, increasing mass to 0,7 caused a slight increase in the belief of not having pancreatic cancer. But in general not much variation is observed across all predictions, this is because there's not much variation in the compatrel relationships of the intermediate frame BIO_LOC_X_AMT. MEDIUM_IMPACT or HIGH_IMPACT pathway mutation could both lead to pancreatic cancer. Mutation being NOVEL or NOT_NOVEL could both lead to pancreatic cancer.

D. ER Experiment 4

Since there's an intermediate node created for DNA_METHYLATION, cfDNA_METHYLATION and mRNA_EXPRESSION. The proposition of cfDNA methylation is always (LOW_IMPACT, MEDIUM_IMPACT, HIGH_IMPACT) is always 1.0 (Check Appendix F for more details). This experiment aims to demonstrate how changes in DNA

methylation, mRNA expression affects the result of prediction. Adjust the three frames and keep the values of all other input parameters the same as the baseline values. The prediction results are shown in the following table:

DNA methylation		mRNA expression		Belief Of Having Pancreatic Cancer Lies Between:	Belief Of Not Having Pancreatic Cancer Lies Between:
Proposition	Mass	Proposition	Mass		
LOW_ IMPACT	0.3	LOW_ IMPACT	0.3	(0.09, 0.204) (0)|***-------|(1)	(0.795, 0.909) (0)|-------***|(1)
LOW_ IMPACT	0.7	LOW_ IMPACT	0.7	(0.09, 0.203) (0)|***-------|(1)	(0.796, 0.909) (0)|-------***|(1)
MEDIUM_ IMPACT	0.3	MEDIUM_ IMPACT	0.3	(0.101, 0.213) (0)|-**-------|(1)	(0.786, 0.898) (0)|-------**-|(1)
MEDIUM_ IMPACT	0.7	MEDIUM_ IMPACT	0.7	(0.167, 0.273) (0)|-**-------|(1)	(0.727, 0.832) (0)|-------**-|(1)
HIGH_ IMPACT	0.3	HIGH_ IMPACT	0.3	(0.101, 0.213) (0)|-**-------|(1)	(0.786, 0.898) (0)|-------**-|(1)
HIGH_ IMPACT	0.7	HIGH_ IMPACT	0.7	(0.166, 0.273) (0)|-**-------|(1)	(0.727, 0.833) (0)|-------**-|(1)

Table 4. Prediction result of various input parameters for DNA methylation, cfDNA methylation, and RNA expression.

When the propositions are of LOW_IMPACT, an increase in the mass from 0.3 to 0.7 doesn't lead to much change in the belief of not having pancreatic cancer. When the proposition is of MEDIUM_IMPACT or HIGH_IMPACT, an increase in the mass from 0.3 to 0.7 will cause an increase in the belief of having pancreatic cancer. The belief of having pancreatic cancer is higher when the proposition is HIGH_IMPACT or MEDIUM_IMPACT, compared to when the

proposition is LOW_IMPACT. The prediction doesn't change much when changes made from MEDIUM_IMPACT to HIGH_IMPACT.

E. ER Experiment 5

Circulating Tumor Cells (CTC) is a new biomarker frame added to the ER model with no intermediate nodes between other frames. This experiment demonstrates how adjusting this frame affects the final result when keeping other values the same. Set CTC to the following parameters and run the program:

CTC		Belief Of Having Pancreatic Cancer Lies Between:	Belief Of Not Having Pancreatic Cancer Lies Between:
Proposition	Mass		
LOW	0.3	(0.203, 0.349) (0)|--**------|(1)	(0.65, 0.796) (0)|------**--|(1)
LOW	0.7	(0.114, 0.198) (0)|-*--------|(1)	(0.801, 0.886) (0)|--------*-|(1)
MEDIUM	0.3	(0.363, 0.522) (0)|---***----|(1)	(0.477, 0.637) (0)|----***---|(1)
MEDIUM	0.7	(0.578, 0.684) (0)|-----**---|(1)	(0.315, 0.422) (0)|---**-----|(1)
HIGH	0.3	(0.361, 0.52) (0)|---***----|(1)	(0.479, 0.639) (0)|----***---|(1)
HIGH	0.7	(0.577, 0.684) (0)|-----**---|(1)	(0.315, 0.423) (0)|---**-----|(1)

Table 5. Prediction result of various input parameters for CTC.

When the proposition is of LOW, an increase in the mass from 0.3 to 0.7 will cause an increase in the belief of not having pancreatic cancer. When the proposition is of MEDIUM or HIGH, an increase in the mass from 0.3 to 0.7 will cause an increase in the belief of having

pancreatic cancer. The belief of having pancreatic cancer is higher when the proposition is HIGH or MEDIUM, compared to when the proposition is LOW. The prediction doesn't change much when changes made from MEDIUM to HIGH.

F. ER Experiment 6

CNV is a new biomarker frame added to the ER model with no intermediate nodes between other frames. This experiment demonstrates how adjusting this frame affects the final result when keeping other values the same. Set CNV to the following parameters and run the program:

CNV		Belief Of Having Pancreatic Cancer Lies Between:	Belief Of Not Having Pancreatic Cancer Lies Between:
Proposition	Mass		
LOW_IMPACT	0.3	(0.643, 0.72) (0)|------**--|(1)	(0.28, 0.356) (0)|--**------|(1)
LOW_IMPACT	0.7	(0.578, 0.649) (0)|-----**---|(1)	(0.351, 0.421) (0)|---**-----|(1)
MEDIUM_IMPACT	0.3	(0.679, 0.76) (0)|------**--|(1)	(0.24, 0.32) (0)|--**------|(1)
MEDIUM_IMPACT	0.7	(0.678, 0.761) (0)|------**--|(1)	(0.239, 0.321) (0)|--**------|(1)
HIGH_IMPACT	0.3	(0.716, 0.787) (0)|-------*--|(1)	(0.212, 0.284) (0)|--*-------|(1)
HIGH_IMPACT	0.7	(0.771, 0.829) (0)|-------**-|(1)	(0.17, 0.228) (0)|-**-------|(1)

Table 6. Prediction result of various input parameters for CNV.

When the proposition is of LOW_IMPACT, an increase in the mass from 0.3 to 0.7 will cause an increase in the belief of not having pancreatic cancer. When the proposition is of MEDIUM_IMPACT, the prediction remains the same as the mass changes. When the proposition is of HIGH_IMPACT, an increase in the mass from 0.3 to 0.7 will cause an increase in the belief of having pancreatic cancer. The belief of having pancreatic cancer increases when the impact of proposition increases when keeping the mass the same.

V. Conclusion and Analysis

Overall, the outcome of ER prediction matches with expectation, that increase in the mass of a proposition, doesn't change the direction of prediction, but will increase the belief in the prediction. Sometimes changes in the proposition do not cause change in the prediction. This is because the different combination of the compateral relationship could lead to the same prediction in the galley definition. The hyperparameters can be fine tuned by assigning a frame continuous proposition type and may create more variation in the compateral relationships. It is observed that variation in frames that do not connect to intermediate nodes create a greater impact in the magnitude in the prediction, compared to those connected to intermediate frames. This may be caused by modulation in the intermediate frames which reduce the impact of changes.

VI. Future Work

Some other prognostic factors that lead to pancreatic cancer could also be added to the current model. Studies of proteomics, metabolomics have revealed certain proteins and miRNA can be potential biomarkers for detecting pancreatic cancer. Other factors including dietary habits, allergies, and some ecological factors may also lead to a higher chance of getting pancreatic cancer, which hasn't been considered in our project. Secondly, the genomic data of this project is gathered from The Cancer Genome Atlas Program (TCGA), which only has 185 cases in the pancreatic adenocarcinoma project. This creates an imbalance in the dataset, as pancreatic cancer cases are only a small portion in the whole dataset [22]. Therefore, more data for pancreatic cancer data is desired. Due to COVID-19 social distancing restrictions, our project is not able to sequence real pancreatic cancer tissue from the California State University (CSU) East Bay lab. In the future, the credibility of this ER model could be further validated if genomic data from real tumor samples could be used. Moreover, since the discount value decreases the credibility of each input frame in prediction, more research needs to be done to justify if the discount values are appropriate. In addition, there are so many different combinations of propositions and masses of different frames. From the experiments we can observe that some frames have greater impacts on the magnitude of prediction than others. More parameter tuning needs to be done to find the desired mass for each frame, or even for each proposition. More in depth research may help to find the data to back up the mass value assigned to each frame.The next phase of this project should also collect genomic data for testing each biomarker to verify if the prediction is correct.

APPENDIX A

Frames	Assigned Proposition	Support
ML_PREDICTION	NOT_PC	0.5
NGS_TECH	ionTorrent	0.5
SEQ_READ	LOW_GC_x_LOW_HMR	0.5
SMOKING_HISTORY	LOW	0.5
DRINKING_HISTORY	LOW	0.5
FAMILY_MED_HISTORY	NO_CANCER	0.5
PATIENT_MED_HISTORY	NO_CANCER	0.5
BIOPSY_SITE_CELL_RESULT	NOT_NEAR_PAN_IRREG	0.5
AMOUNT_GEN_MATERIAL	SMALL	0.5
NOVEL_MUTATION	NOT_NOVEL	0.5
PATHWAY_MUTATION	LOW_IMPACT	0.5
DNA_METHYLATION	LOW_IMPACT	0.5
cfDNA_METHYLATION	(LOW_IMPACT, MEDIUM_IMPACT, HIGH_IMPACT)	1.0
mRNA_EXPRESSION	LOW_IMPACT	0.5
EXOSOME_MiRNA	LOW_IMPACT	0.5
CNV	LOW_IMPACT	0.5
CTC	LOW	0.5

Table 7. Baseline Propositions and Corresponding Support

APPENDIX B

Frames	Discount Rate
ML_PREDICTION	0.1
NGS_TECH	0.2
SEQ_READ	0.1
NGS_X_SEQ_READ	0.1
SMOKING_HISTORY	0.3
DRINKING_HISTORY	0.3
FAMILY_MED_HISTORY	0.2
PATIENT_MED_HISTORY	0.1
BIOPSY_SITE_CELL_RESULT	0.2
AMOUNT_GEN_MATERIAL	0.1
NOVEL_MUTATION	0.3
PATHWAY_MUTATION	0.1
DNA_METHYLATION	0.2
cfDNA_METHYLATION	0.1
mRNA_EXPRESSION	0.2
EXOSOME_MiRNA	0.1
CNV	0.5
CTC	0.1

Table 8. Discount Rates and Corresponding Frames

APPENDIX C

Novel Mutation

There are more and more discoveries of novel mutations that would increase the chance of developing pancreatic cancer. [25] describes one of the largest whole-genome association studies, where more than 11.3 million variants were analyzed in more than 21,536 persons. Among these genetic donors, 9,040 are pancreatic cancer patients and 12,946 are healthy individuals of European descent. It involved researchers from the National Cancer Institute, Johns Hopkins Kimmel Cancer Center and other collaborators from over 80 other worldwide institutions.

From the study, five novel genetic changes were identified to have linkage to increasing pancreatic cancer risk. Commenting on this study, Alison Klein, a professor of oncology, pathology and epidemiology, highlighted that a holistic consideration of all five variants is vital in understanding pancreatic cancer development. On an individual basis, though these variants can result in modest changes, they aren't sufficient as indicators of pancreatic cancer detection.

These variants are located on the human chromosomes 1, 7, 8, 17. One of the variants was found in a protein-coding gene, NOC2L, which binds directly to gene p53, which drives pancreatic cancer development. This then binds with another variant, a tumor gene, p63, that's associated with increased pancreatic cancer risk.

The third variant was identified in the HNF4G and HNF1B genes, they are growth factors that regulate cell growth. The next variant was found near the GRP gene, which is associated with the regulation of gastrointestinal hormones. And the final variant was found on the TNS3

gene, which possibly aids metastasis. According to the study leader, Alison Klein, PH.D., M.H.S., changes of these five identified regions increases the risk of getting pancreatic cancer but yet, there's still a whole lot more to learn about hereditary factors associated with high-risk patients.

Appendix D

Pathway mutation

Pancreatic cancer genomes are characterized by several genetic mutations which aid the development and progression of neoplastic lesions. Driver mutations such as K-Ras, initiate the disease development while passenger mutations like phosphoinositide-3-kinase (PI3K), CDKN2A and ERBB2, amplify its progression.

PDAC are genetically heterogeneous tumors of which K-Ras are the first detected major driver mutations during progression in over 90% of patients, constituting the most frequently mutated oncoprotein [26]. Mutated K-Ras activates several downstream effector-signalling pathways like PI3K, which are linked with migration, proliferation and metastasis. In PDAC patients, the most common K-Ras mutation points are G12D and G12V.

As for PI3K signalling pathway, studies have indicated that they inhibit cellular apoptosis and stimulate the proliferation of cancerous cells. An estimate of about 60% of PDAC patients have a deregulated PI3K/Akt signalling pathway [27]. There are three classes of PI3Ks, grouped as Class I, Class II and Class III. And studies have indicated that the Class I PI3K are more responsible for the proliferation of pancreatic cancers. It modulates downstream signalling cascades in response to stimuli from several growth factor receptors on the cancerous cell surface [27].

Also, serine-threonine kinase (Akt) functions as regulators of diverse cellular processes that are very vital for cell growth. Development of a screening methodology that's highly sensitive in analyzing the high incidences of mutations in the PI3K signalling pathway will aid

the early detection of pancreatic cancers. Also, asides from early detection, understanding thoroughly these signalling pathways can aid to improve the therapeutic options available.

Appendix E

DNA methylation

DNA methylation is a biological process that occurs when methyl group bonds with the carbon 5 of cytosines from a DNA molecule, to form 5-methylcytosine (5mC), due to a DNA methyltransferases mediated covalent addition [28].

DNA methylation is vital in the progression of cancers. When found at the regulatory regions of a gene, it results in the transcription of such a gene being suppressed. It alters the chromatin structure and silences the tumor suppressor gene or activates oncogenes.

The use of DNA methylation analysis promises to be effective because cfDNA are more informative, sensitive and carries methylation markers that makes it easier to identify tissue-specific cell death. It is interesting to note that DNA methylation profiles appear very similar when originating from the same tissue. So scientists are able to decipher the heterogeneous signals emanating from a cfDNA pool to locate the originating tissues [29].

Some notable methods used in the assessment of DNA methylation include Methylation-specific PCR (MSP), Quantitative Methylation Specific PCR (QMSP), Targeted Amplicon Sequencing, and a whole lot of others [30]. A large number of these methods are dependent on the relatively fast deamination of unmethylated cytosines into uracils [28]. A recent study [30] of the methylation profile of genes ADAMTS1 and BNC1, found the promoter methylation of the two genes could be vital biomarkers for the early detection of pancreatic cancer.

Appendix F

Cell-free DNA/cfDNA methylation

In carrying out screening for pancreatic cancers, the analysis of pancreatic juice can provide more information when compared to other forms of analysis. But it is a more cumbersome and invasive means of obtaining specimen samples. This has led to the exploration of other sensitive and non-invasive methods for early detection.

Right now one worthy approach involves the study of circulating cell-free DNA (cfDNA), which are carriers of specific markers that help to identify certain cell death. cfDNA refers to degraded DNA fragments that are released into the blood plasma. It is made up of short double-stranded segments of nucleic acids. Analysis of cfDNA presents a minimally non-invasive approach as they are present as polynucleotide chains of 0.1-20 kilobase-base pairs in the plasma and serum [29].

Increased levels of cfDNA in the plasma can be as a result of stroke, trauma or even strenuous exercises, but its concentration can't be compared to those of a cancer patient. While a healthy individual's cfDNA concentration can range from 0 to 100 ng/ml of blood, a cancerous patient presents an average of 4 to 40 times increased levels [29]. So invariably, the higher the plasmatic cfDNA levels observed, the higher the cancer's cellular turnover. cfDNA are carriers of markers for KRAS and DNA methylation signatures. Once cfDNA is sequenced, the genetic information of a tumor becomes available since all tumors exhibit genetic alterations.

A major hurdle encountered with the circulating cfDNA is that certain factors limit the information provided by the genetic sequence. Together, factors such as heterogeneous genetic

background, inter-individual variability of plasmatic levels and having diverse origin from both healthy and neoplastic cells, makes the interpretation of its sequence data cumbersome. Nevertheless, analyzing the sequence data obtained from the circulating cfDNA is a method that holds a lot of promise in the prior diagnosis of pancreatic cancer.

Appendix G

mRNA-expression

In addition to DNAs, protein-coding mRNAs from the tumor tissues are released into the blood, and can reflect changes in tumor specific gene expression. Combined with advances in molecular diagnostics, systematic profiling of cell free m-RNA can improve our understanding of cancer pathology and identify novel biomarkers for early detection, without the need for invasive biopsy.

Certain cell-free m-RNA species are in complexed forms that protect them from degradation by RNases [31]. This ensures their stability in the circulation, in contrast to complex-free RNA, which is rapidly degraded. Therefore, key challenges in the cell free m-RNA testing include its extremely low abundance, susceptible to degradation, relatively unstable and poor extraction efficiency. Circulating cell free m-RNA carries information from human tissues; the pattern of cell free m-RNA expression reflects cancer cell growth and reproduction, dysfunction of cancer immunity, which makes cell free m-RNA expression signature a promising biomarker for early diagnostic, prognostic and therapeutic purposes [32].

While several research reports have shown impressive promises of circulating cell free m-RNA, there's still a lot more need to be learned in this field. The advantage of using circulating cell free m-RNA is that, compared to protein biomarkers, PCR can be used in detecting circulating cell free m-RNA at a single molecule level [33].

Appendix H

Exosomes-miRNA

Exosomes are membrane-bound extracellular vesicles that are generated by the endosomal compartment, functioning as prominent mediators of intercellular communication [34]. So, through the transfer of biological materials, they instruct, regulate and re-educate their microenvironment and target organs. Some of the components of exosomes include lipids, proteins and nucleic acids such as DNA, microRNA (miRNA) and so on. Analysis of exosomes offers a minimally invasive or non-invasive method because they are easily isolated and identified in body fluids.

For early detection, exosome-microRNA holds a lot of potential as an effective biomarker. miRNA is a small non-coding RNA molecule of about 19-25 nucleotides, invariably regulating gene expression in almost all cellular processes, including carcinogenesis [34]. Their dysfunction often results in the initiation, growth and spread of cancer.

The signature profile of a pancreatic cancer's miRNA is very different when compared with that of a normal pancreatic cell. And they tend to be upregulated in patients diagnosed of having pancreatic cancers than those with benign pancreatic disorders. So when achieved, an exosomal miRNA-based detection can aid in early detection of dysfunctional miRNA at its localized stage, using a minimally invasive or non-invasive method.

Appendix I

CTC

Circulating tumor cells(CTC) refers to detached cells emanating from a primary tumor that's in circulation in the bloodstream. CTCs are capable of metastasizing in distant organs as offshoots of their primary tumor. Invariably, this means that CTCs tend to be upregulated as the disease progresses, serving as an indicator for cancer progression.

Although, due to metastatic inefficiency, only an estimated 0.01% of CTCs progresses to form metastases, accounting for its rarity in per millimetre of blood [35]. Still yet, notwithstanding the fact that it can't exclusively indicate clinically significant macro-metastases, it certainly indicates the presence of malignant tumors which can aid in the early prognosis of pancreatic cancer.

As a result of their low frequency in per milliliter of blood, detection and isolation are quite difficult. Thus, it employs a two-step procedure for isolation:

Step1. CTC Enrichment: CTC enrichment technique isolates CTCs based on either through their surface antibody or their electric charge or cell size. CellSearch is an example of a surface antibody-based technique. Referred to as the "gold standard" by authors, it is the only method that's approved by the FDA for diagnosis of breast, colorectal and prostate cancers [36].

Step 2. CTC Detection: The detection of enriched CTCs is carried out either through morphological examination, mRNA analysis or mutational analysis of the DNA.

Some technologies developed for CTC isolation include ScreenCell, ISET, ApoStream, ClearCell FX System and so on.

Appendix J

CNV

Copy number variation (CNV) refers to an occurrence where sections of a genome are repeated, and between individuals, there's a variation of the number of repeats in the genome [23]. They are vital in displaying variations within a population, modulating gene expression and even disease phenotype. In pancreatic cancer, several genes have been differentially expressed due to copy number alterations. Familial pancreatic cancers are as a result of heritable alterations in at least a rare major gene.

In the structural architecture of copy number variations, researchers have identified hotspot regions where copy number variations are more enriched in a genome. In these regions, they have an increased rate of chromosomal rearrangements that's also responsible for genetic diseases. These CNV hotspots are consistent in many populations regardless of your ethnographic origin [23].

For example, analyzing the CNV of mutations in the PRSS1 gene can assist in identifying high-risk patients regardless of the ethnographic origin [37]. Developing a screening methodology that'll identify the CNV hotspot regions for pancreatic cancers will greatly aid the early detection of pancreatic cancers.

Appendix K

The program for this project is written in Python. Values for mass, discount and propositions can be defined and modified for a particular frame to test how the changes affect the final prediction. Values for propositions and masses for different frames are defined in md_input.txt. Discount values for various frames are defined in the pc_discount_rates.py. Code in update_md_input.py serves to automate the process of adjusting the mass and proposition values for different frames. The program reads definition from md_input_uddate.txt to update the values in md_input.txt. In md_input_uddate.txt, an initial mass value, mass increment rule, and a list of propositions can be provided for different frames. The automate updating program reads the list of propositions for each frame, and updates the mass for each proposition based on mass increment rule starting from the initial mass. For example, if for the frame DRINKING_HISTORY, the list of propositions given are {(LOW), (MEDIUM), (HIGH)} with initial mass 0.2, and mass updating rule is +0.2, then the program will automatically update the values for the frame DRINKING_HISTORY in md_input.txt, starting from proposition LOW with mass 0.2, then increase mass to 0.4, and keeping increment mass as long as its <= 1.0. After that, the program will use the next proposition MEDIUM in the list with the initial mass 0.2, and repeat the process to update mass, and this process will be repeated for proposition HIGH as well. After each update, the program will also run the evidential reasoning program predict_pac_new.py, which reads input parameters from md_input.txt to make predictions and record the input mass, propositions for the updated frame along with the outcomes in result.txt.

If multiple frames are defined in md_input_uddate.txt, the updating program will update each frame just like the above example in md_input.txt simultaneously, and record outcomes in

result.txt. Mass updating rule can be set as incrementing using “+” sign or decrementing with “-” sign in front of the numeric value, if the initial mass is 0.7, and mass update rules is -0.2, then for each update, the program will decreatement mass for each proposition till >= 0.0. When defining propositions to be updated for each frame, each proposition should be surrounded by a pair of and separated by a space.

REFERENCES

[1] M. Orth et al., "Pancreatic ductal adenocarcinoma: biological hallmarks, current status, and future perspectives of combined modality treatment approaches", Radiation Oncology, vol. 14, no. 1, 2019. Available: 10.1186/s13014-019-1345-6.

[2] H. Wang et al., "Global, regional and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015", *Lancet*, Oct. 2016.

[3] "Cancer of the Pancreas - Cancer Stat Facts", SEER, 2020. [Online]. Available: https://seer.cancer.gov/statfacts/html/pancreas.html. [Accessed: 29- Nov- 2020].

[4] "WHO | World Cancer Report 2014", Who.int, 2020. [Online]. Available: https://www.who.int/cancer/publications/WRC_2014/en/. [Accessed: 29- Nov- 2020].

[5] D. P. Ryan, T. S. Hong, N. Bardeesy, "Pancreatic adenocarcinoma", *The New England Journal of Medicine*, Sep. 2014.

[6] T. Vincent et al., "DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology", Nov. 2018.

[7] J. S. Tobias, D. Hochhauser, "Cancer and its Management (7th ed.)", p. 297, 2014.

[8] W. R. Burns, B. H. Edil, "Neuroendocrine pancreatic tumors: guidelines for management and update", *Current Treatment Options in Oncology*, Mar. 2012.

[9] "Pancreatic Cancer Treatment (Adult) (PDQ®)–Health Professional Version", National Cancer Institute, 2020. [Online]. Available:https://www.cancer.gov/types/pancreatic/hp/pancreatic-treatment-pdq#section/a ll. [Accessed: 29- Nov- 2020].

[10] G. Bond-Smith et al., "Pancreatic adenocarcinoma", May 2012.

[11] L. Kunovsky et al., "The Use of Biomarkers in Early Diagnostics of Pancreatic Cancer", *Canadian Journal of Gastroenterology and Hepatology*, vol. 2018, pp. 1-10, 2018.

[12] S. Raimondi et al., "Pancreatic Cancer In Chronic Pancreatitis; Aetiology, Incidence, And Early Detection", *Best Practice. Red. Clin. Gastroenterol*, 2010.

[13] U. K. Ballehaninna, “The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence-based appraisal”, *J Gastrointest Oncol*, 2012.

[14] K. S. Goonetilleke, “Systematic review of carbohydrate antigen (ca 19-9) as a biochemical marker in the diagnosis of pancreatic cancer”, *European Journal of Surgical Oncology (EJSO)*, 2007.

[15] A. Nygaard, "Levels of Cell-free DNA And Plasma KRAS During Treatment of Advanced NSCLC", *Oncology Reports*, vol. 31, no. 2, pp. 969-974, 2013.

[16] D. He, “Analytical Application of Next - Generation Sequencing Technology For Detection of Gene Mutations In Pancreatic Cancer”, 2019.

[17] D. Pietrasz et al., "Plasma Circulating Tumor DNA in Pancreatic Cancer Patients Is a Prognostic Marker", *Clinical Cancer Research*, vol. 23, no. 1, pp. 116-123, 2016.

[18] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis", *Cancer Informat*, p. 59, 2006.

[19] K. K et al., “Machine Learning Applications In Cancer Prognosis And Prediction”, *Comput Struct Biotechnol J.*, Nov. 2014.

[20] L. Wesley, K. Graham, “Evidence-Based Decision Support For The Biopharmaceutical Industry”, 2006.

[21] O. Sharghi, “Towards Early Detection of Pancreatic Cancer”, *San Jose State University*, 2019.

[22] R. D. Chandratre, “Evidence-Based Detection of Pancreatic Cancer”, *San Jose State University*, 2020.

[23] S. A. McCarroll and D. M. Altshuler DM. "Copy-number Variation And Association Studies of Human Disease", *Nat. Genetics*, Jun. 2007.

[24] D. Foster, M. Zvolensky, L. Garey, J. Ditre and N. Schmidt, "Relationships Between Drinking Motives and Smoking Expectancies Among Daily Smokers Who Are Also Problem Drinkers", *Journal of Dual Diagnosis*, vol. 10, no. 3, pp. 118-129, 2014.

[25] A. P. Kelvin et al., “Genome-wide Meta-analysis Identifies Five New Susceptibility Loci for Pancreatic Cancer”, *Nat. Commun.*, 2018.

[26] J. Lennerz and A. Stenzinger, “Allelic Ratio of KRAS Mutations in Pancreatic Cancer”, *The Oncologist*, 2014.

[27] V. M. Bondar et al., “Inhibition of The Phosphatidylinositol 3'-kinase-AKT Pathway Induces Apoptosis in Pancreatic Carcinoma Cells In Vitro and In Vivo”, *Mol. Cancer Ther.*, p. 989–997, 2002.

[28] L. Moore, T. Le and G. Fan, "DNA Methylation and Its Basic Function", Neuropsychopharmacology, vol. 38, no. 1, pp. 23-38, 2012. Available: 10.1038/npp.2012.112.

[29] M. Brancaccio, F. Natale, G. Falco and T. Angrisano,. “Cell-Free DNA Methylation: The New Frontiers of Pancreatic Cancer Biomarkers’ Discovery”. *Genes*, 11(1), p.14., 2019.

[30] M. Eissa et al., "Promoter methylation of ADAMTS1 and BNC1 as potential biomarkers for early detection of pancreatic cancer in blood", *Clinical Epigenetics*, vol. 11, no. 1, 2019.

[31] Chen. H. Yeh, “Enabling Circulating Cell-free mRNA Profiling to Empower Cancer Early Detection”, *Genet Med*, Vol. 14:S3, 2020.

[32] M. F. Souza et al., "Circulating mRNAs and miRNAs as Candidate Markers for The Diagnosis and Prognosis of Prostate cancer", *Plos One*, 2017.

[33] A. Turchinovich, A. Baranova, O. Drapkina and A. Tonevitsky. "Cell-Free Circulating Nucleic Acids as Early Biomarkers for NAFLD and NAFLD-Associated Disorders", *Frontiers in Physiology*, 2018.

[34] K. Ranganathan and V. Sivasankar, “MicroRNAs - Biology and Clinical Applications”, *Journal of Oral and Maxillofacial Pathology*, 18(2), p.229, 2014.

[35] A. Kowalik, M. Kowalewska, and S. Góźdź, “Current Approaches for Avoiding The Limitations of Circulating Tumor Cells Detection Methods—Implications for Diagnosis And Treatment of Patients With Solid Tumors”, *Trans. Res.*, 185, pp.58-84, e15, Jul. 2017

[36] D. K.Yadav et al., “Liquid Biopsy in Pancreatic Cancer: The Beginning of A New Era”. *Oncotarget*. 2018.

[37] M. Emmanuelle M. et al., “PRSS1 Copy Number Variants And Promoter Polymorphisms In Pancreatitis: Common Pathogenetic Mechanism, Different Genetic Effects”, *Gut*., Mar. 2018.