#Background

* Pancreatic cancer is an extremely deadly type of cancer. Once diagnosed, the five-year survival rate is less than 10%. However, if pancreatic cancer is caught early, the odds of surviving are much better. Unfortunately, many cases of pancreatic cancer show no symptoms until the cancer has spread throughout the body. A diagnostic test to identify people with pancreatic cancer could be enormously helpful.

#The data

A multi-national team of researchers sought to develop an accurate diagnostic test for the most common type of pancreatic cancer, called pancreatic ductal adenocarcinoma or PDAC. They gathered a series of biomarkers from the urine of three groups of patients:

* Healthy controls
* Patients with non-cancerous pancreatic conditions, like chronic pancreatitis
* Patients with pancreatic ductal adenocarcinoma

The key features are four urinary biomarkers: creatinine, LYVE1, REG1B, and TFF1.

* Creatinine is a protein that is often used as an indicator of kidney function.

**Creatinine** is measured in milligrams per deciliter (mg/dL). Here are the **normal values** by age: 0.9 to 1.3 mg/dL for adult males. 0.6 to 1.1 mg/dL for adult females.

* YVLE1 is lymphatic vessel endothelial hyaluronan receptor 1, a protein that may play a role in tumor metastasis
* REG1B is a protein that may be associated with pancreas regeneration
* TFF1 is trefoil factor 1, which may be related to regeneration and repair of the urinary tract .
* Age and sex, both included in the dataset, may also play a role in who gets pancreatic cancer. The dataset includes a few other biomarkers as well, but these were not measured in all patients (they were collected partly to measure how various blood biomarkers compared to urine biomarkers).

#Features

* Sample ID: Unique string identifying each subject
* Patient's Cohort:
  + Cohort 1, previously used samples;
  + Cohort 2, newly added samples
* Sample Origin:
  + BPTB: Barts Pancreas Tissue Bank, London, UK; :(1)
  + ESP: Spanish National Cancer Research Centre, Madrid, Spain; (2)
  + LIV: Liverpool University, UK; (3)
  + UCL: University College London, UK (4)
* Age: Age in years
* Sex:M = male, F = female
* Diagnosis (1=Control, 2=Benign, 3=PDAC):

1 = control (no pancreatic disease), 2 = benign hepatobiliary disease (119 of which are chronic pancreatitis), 3= pancreatic ductal adenocarcinoma

* Stage:For those with pancratic cancer, what stage was it? One of I, IA, IB, II,IIA, IIB, III, IV
* Benign Samples Diagnosis: For those with a benign, non-cancerous diagnosis, what was the diagnosis?
* Plasma CA19-9 U/ml:Blood plasma levels of CA 19–9 monoclonal antibody that is often elevated in patients with pancreatic cancer. Only assessed in 350 patients (one goal of the study was to compare ).The upper limit of the normal reference **value** for **CA19**-**9** is 37 U/mL.
* Creatinine mg/ml:Urinary biomarker of kidney function
* LYVE1 ng/ml: Urinary levels of Lymphatic vessel endothelial hyaluronan receptor 1, a protein that may play a role in tumor metastasis
* REG1B ng/ml:Urinary levels of a protein that may be associated with pancreas regeneration.
* TFF1 ng/ml: Urinary levels of Trefoil Factor 1, which may be related to regeneration and repair of the urinary tract
* REG1A ng/ml: Urinary levels of a protein that may be associated with pancreas regeneration. Only assessed in 306 patients (one goal of the study was to assess REG1B vs REG1A)
* benign\_sample\_diagnosis:
  + Abdominal Pain :1
  + Biliary Stricture (Secondary to Stent) :2
  + Cholecystitis:3
  + Cholecystitis (Chronic) :4
  + Cholecystitis (Chronic) Cholelithiasis:5
  + Cholecystitis (Chronic) Cholesterolsis:6
  + Choledochal Cyst :7
  + Choledocholiathiasis:8
  + Cholelithiasis with adenomyomatous hyperplasia :9
  + Duodenal Stricture:10
  + Duodenitis:11
  + Gallbladder polyps :12
  + Gallbladder Porcelain:13
  + Gallstones:14
  + Gallstones - Incidental :15
  + Gastritis :16
  + Gastritis and Reflux :17
  + Ill defined lesion in uncinate process:18
  + Ischaemic Common Bile Duct Stricture:19
  + Pancreatitis (Pseudocyst):20
  + Pancreatitis:21
  + Pancreatitis (Abscess):22
  + Pancreatitis (Acute):23
  + Pancreatitis (Alcohol-Chronic-Pseuodcyst):24
  + Pancreatitis (Alcohol-Chronic):25
  + Pancreatitis (Alcohol):26
  + Pancreatitis (Autoimmune) :27
  + Pancreatitis (Chronic-Pseudocyst) :28
  + Pancreatitis (Chronic) :29
  + Pancreatitis (Chronic) (Later became PDAC):30
  + Pancreatitis (Chronic) Choledocholithiasis:31
  + Pancreatitis (Gallstone-Alcohol-Pseudocyst):32
  + Pancreatitis (Gallstone-Pseudocyst):33
  + Pancreatitis (Gallstone) :34
  + Pancreatitis (Hereditary-Chronic) :35
  + Pancreatitis (Hypertriglyceridemia):36
  + Pancreatitis (Idiopathic):37
  + Pancreatitis (Pseudocyst):38
  + Pancreato-jejunostomy Anastomoses Stricture:39
  + Premalignant lesions-Adenoma-NOS:40
  + Premalignant lesions-Mucinous cystadenocarcinoma-noninvasive:41
  + Premalignant lesions-Mucinous cystadenoma-NOS:42
  + Premalignant lesions-Tubular adenoma-NOS:43
  + Premalignant lesions-Tubulovillous adenoma-NOS:44
  + Premalignant lesions-Villous adenoma-NOS:45
  + Serous cystadenoma - NOS:46
  + Serous microcystic adenoma:47
  + Simple benign liver cyst:48