UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM GASTROENTEROLOGIA E HEPATOLOGIA

SCREENING DE GENES DIFERENCIALMENTE EXPRESSOS E POTENCIALMENTE PROGNÓSTICOS NO CÂNCER DE PÂNCREAS

Aluna de mestrado: **Mariana dos Santos Lobo**

Orientadora: Dra. Patricia Luciana da Costa Lopez

Co-orientador: Prof. Dr. Eduardo Cremonese Filippi-Chiela

PORTO ALEGRE
2019

MARIANA DOS SANTOS LOBO

SCREENING DE GENES DIFERENCIALMENTE EXPRESSOS E POTENCIALMENTE PROGNÓSTICOS NO CÂNCER DE PÂNCREAS

Dissertação de Mestrado apresentado como requisito parcial para obtenção do título de Mestre em Gastroenterologia e Hepatologia, pelo Programa de Pós-Graduação em Gastroenterologia e Hepatologia da Universidade Federal Do Rio Grande Do Sul - UFRGS.

Orientadora: Dra. Patricia Luciana da Costa Lopez

Co-orientador: Prof. Dr. Eduardo Cremonese Filippi-Chiela

Porto Alegre

2019

CIP - Catalogação na Publicação

Santos-Lobo, Mariana
SCREENING DE GENES DIFERENCIALMENTE EXPRESSOS E POTENCIALMENTE PROGNÓSTICOS NO CÂNCER DE PÂNCREAS / Mariana Santos-Lobo. -- 2019.
76 f.
Orientadora: Patricia Luciana da Costa Lopez.

Coorientadora: Eduardo Cremonese Filippi-Chiela.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Medicina, Programa de Pós-Graduação em Ciências em Gastroenterologia e Hepatologia, Porto Alegre, BR-RS, 2019.

1. Cancer de Pâncreas. 2. Bioinformática. 3. Genética. I. da Costa Lopez, Patricia Luciana, orient. II. Cremonese Filippi-Chiela, Eduardo, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

AGRADECIMENTOS

Agradeço aos meus pais, Ivana Maia dos Santos e Jester Daniel Moraes, por me educarem com tanto amor e dedicação. Por todo carinho, paciência e suporte, por acreditarem nos meus sonhos e me incentivarem a realizá-los.

Aos meus avós, Jussara Silva Maia, Ivan Dias dos Santos e Janete Reis, por me ensinarem a enfrentar obstáculos e por me inspirarem, todos os dias, a ser uma pessoa melhor.

Ao meu namorado, Carlos Henrique Lauermann, por tudo o que é pra mim, e por agregar tanto à minha vida.

Aos amigos e à família, pelo carinho, suporte e estímulo necessários para chegar até aqui.

Ao Núcleo de Bioinformática do CPE-HCPA, pelo tempo e pelo conhecimento compartilhados.

Aos colaboradores do LaBCM, alunos de iniciação científica, colegas de mestrado e professores, pelos ensinamentos e pela parceria neste período em que estivemos juntos.

“A mind that is stretched by a new experience can never go back to its old dimensions.” (Oliver Wendell Holmes).

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

AHCY	Adenosylhomocysteinase
DEA	*Differential Expression Analysis*
DEG	*Differentially Expressed Genes*
FAM45A	Family With Sequence Similarity 45 Member A
GDC	*Genomic Data Commons Data Portal*
LYPLA1	Lysophospholipase I
MAST4	Microtubule-Associated Serine/Threonine Kinase 4
mOS	Média de Sobrevida Global
NT	*Normal Adjacent Tissue*
PanIN	Neoplasias Intraepiteliais Pancreáticas
SPRY3	Sprouty RTK Signaling Antagonist 3
TC2N	Tandem C2 Domains, Nuclear
TCGA	*The Cancer Genome Atlas*
TGM5	Transglutaminase 5
THPA	*The Human Protein Atlas*
TP	*Primary Tumor Tissue*
TRIM67	Tripartite Motif-Containing Protein 67
USP19	Ubiquitin carboxyl-terminal hydrolase 19

SUMÁRIO

APRESENTAÇÃO DA DISSERTAÇÃO

A presente dissertação está estruturada conforme segue. O **resumo** apresenta brevemente a temática deste trabalho. A **introdução** está dividida em duas seções principais: (1) Câncer de Pâncreas, com dados epidemiológicos, terapêuticos e marcadores moleculares; e (2) TCGA e Ferramentas de Bioinformática, onde são expostas as plataformas e *softwares* utilizados neste trabalho. As figuras da introdução estão ordenadas em numerais romanos. Após a introdução, estão descritas **justificativa**, **questão de pesquisa** e **hipóteses**, seguidas de **objetivo geral** e **objetivos específicos**. O artigo original (a ser submetido ao periódico "*Pancreas"*) contém **abstract**, **introduction**, **materials and methods**, **results**, **discussion** e **conclusions**. As figuras do artigo estão ordenadas em números ordinais. Por fim, a partir dos resultados e conclusões, são descritas as **perspectivas**.

RESUMO

O câncer de pâncreas apresenta alta mortalidade, devido a diagnóstico tardio e comportamento biológico agressivo. A taxa de sobrevida é muito baixa e as abordagens terapêuticas ainda não se apresentam efetivas à maioria dos pacientes. Os genes com mutações mais frequentes em tumores pancreáticos são o oncogene *KRAS* e os supressores tumorais *TP53* e *CDK2N*. Estas vias controlam proliferação e morte celular, além do ciclo celular, migração e metabolismo. A utilização de ferramentas de bioinformática como análise integradora pode fornecer informações clinicamente relevantes relacionadas à biologia tumoral e desenvolvimento de potenciais marcadores de diagnóstico, prognóstico e resposta à terapia, com impacto biológico robusto e custo reduzido. O objetivo do presente trabalho foi caracterizar a expressão diferencial da coorte de neoplasias pancreáticas do TCGA com relação ao tecido pancreático normal e buscar potenciais marcadores de prognóstico tumoral. Foram encontrados 235 genes diferencialmente expressos em amostras tumorais em comparação com o tecido normal (NT *versus* TP). Destes, 28 genes apresentaram níveis de expressão pelo menos 3 vezes diferente do tecido normal (3 *downregulated* e 25 *upregulated*); destes, 9 se apresentaram como fatores prognóstico. Níveis elevados de *MAST4, SPRY3, USP19* e *TRIM67* foram associados a melhor prognóstico, enquanto o oposto foi observado para *AHCY, FAM45A, LYPLA1, TC2N* e *TGM5*. Revisando a literatura para o papel destes genes no câncer, observamos uma escassez de referências. Assim, utilizamos as plataformas KEGG (anotação funcional), STRING (redes de interação de proteínas) e THPA (distribuição subcelular) para inferir possíveis funções biológicas dos mesmos. Através desta abordagem identificamos 3 moduladores da função da Ras (*TRIM67, SPRY3* e *LYPLA1*) e 3 moduladores de *TP53* (*AHCY, FAM45A* e *TC2N*). Estes genes também modulam mecanismos de comunicação celular, como a produção de exossomos, migração e metabolismo celular. Considerando o *status* mutacional de *KRAS*, encontramos 952 genes diferencialmente expressos entre amostras tumorais $KRAS^{wt}$ versus $KRAS^{mut}$. Considerando os 9 genes diferencialmente expressos e fatores prognósticos da análise NT *versus* TP, encontramos 6 genes expressos diferencialmente também considerando o status de *KRAS*. Concluindo, nós descrevemos 9 potenciais marcadores prognósticos para o câncer de pâncreas, os quais modulam mecanismos celulares centrais na biologia das células tumorais.

1 INTRODUÇÃO

1.1 Câncer de Pâncreas

1.1.1 O Pâncreas e Epidemiologia dos Cânceres Pancreáticos

O pâncreas é uma glândula mista anexo ao sistema digestório, contendo um componente endócrino e um componente exócrino. O componente endócrino (2% do órgão) corresponde às ilhotas de Langerhans, responsáveis pelo metabolismo da glicose e compostas principalmente por células alfa, beta e delta. As células alfa produzem o hormônio glucagon, enquanto as células beta são produtoras de insulina. O componente exócrino, por outro lado, é constituído por ácinos serosos e ductos excretores que conduzem a secreção pancreática para o duodeno. Esta secreção é composta por cerca de 20 enzimas digestivas, principalmente proteases (como a quimiotripsina e carboxilpeptidase), lipases, amilases, fosfolipases, colesterol esterase, entre outras (KIERSZENBAUM; TRES, 2012).

O pâncreas é composto por 4 porções anatômicas: (1) cabeça, (2) colo, (3) corpo, (4) cauda (**Figura I**). Entre 60 e 70% dos cânceres de pâncreas se localizam na cabeça, e 20 a 25% no corpo. A íntima associação do pâncreas com múltiplos vasos sanguíneos, com drenagem abdominal difusa e extensiva para linfonodos, favorece a disseminação metastática das células tumorais, ao mesmo tempo que dificulta a ressecção cirúrgica de grande parte dos tumores (KIERSZENBAUM; TRES, 2012).

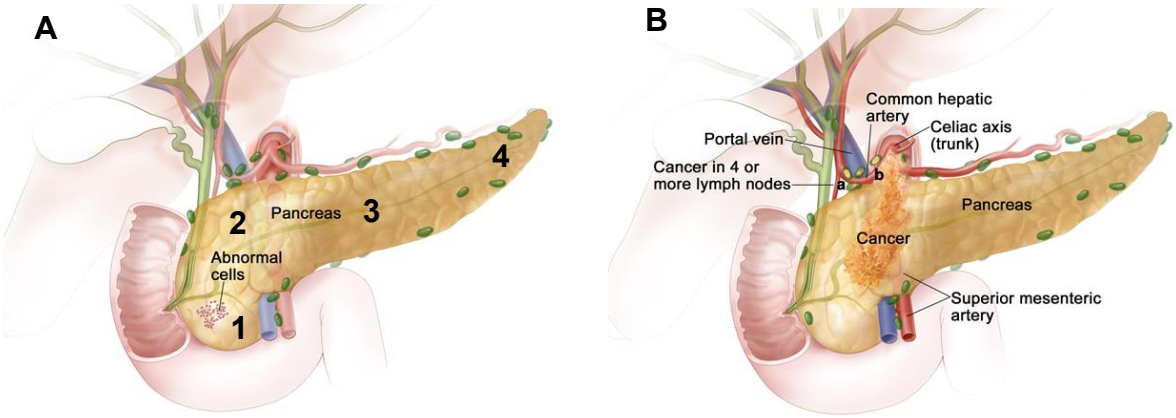


Figura I – Estrutura anatômica do pâncreas e associação com o desenvolvimento tumoral. (A) Regiões anatômicas do pâncreas - (1) cabeça, (2) colo, (3) corpo, (4) cauda. É mostrado na imagem o início de um processo carcinogênico na cabeça do órgão. (B) Representação de um tumor pancreático invasivo (estágio III) (Adaptado de WINSLOW, 2018).

Mais de 95% dos tumores pancreáticos tem origem exócrina, os quais são muito mais agressivos e letais do que tumores de origem endócrina. Entre os tumores exócrinos, o adenocarcinoma ductal pancreático (ADP ou, do inglês, PDAC) é o mais comum e o mais grave, responsável por cerca de 90% casos diagnosticados no Brasil. Entre os outros tipos de tumores exócrinos, destacam-se os tumores mucinosos, enquanto menos de 5% dos cânceres de pâncreas são tumores endócrinos (HAJATDOOST et al., 2018; HRUBAN; HORNSBY, 2018; INSTITUTO NACIONAL DE CÂNCER - INCA, 2018).

O câncer de pâncreas é um dos cânceres mais fatais da atualidade. Em todo o mundo, o câncer de pâncreas corresponde a 3,2% das neoplasias malignas diagnosticadas, com aproximadamente 55.000 novos casos anuais nos EUA. Já considerando a mortalidade, em 2017 foram aproximadamente 44.000 óbitos causados por este tipo tumoral, correspondendo a 7,2% de todas as mortes por câncer. Estes dados o colocam, nos Estados Unidos, apenas como o 11° tipo tumoral em incidência, porém como a 4ª causa de mortes por câncer. No Brasil, em 2015 foram registrados 9.464 novos casos de câncer de pâncreas, e cerca de 4% de mortes por câncer são causadas por esta neoplasia (INSTITUTO NACIONAL DE CÂNCER - INCA, 2018).

Entre os fatores de risco associados ao câncer de pâncreas estão o tabagismo, a pancreatite crônica, a diabetes e a idade avançada, além de síndromes genéticas, como síndrome de Peutz-Jeghers, pancreatite e câncer de pâncreas hereditários, câncer de mama e ovário hereditários. 70% dos casos são diagnosticados em pacientes com 70 anos ou mais. A taxa de sobrevida em 5 anos depende, principalmente, da característica do tumor no momento do diagnóstico, no que diz respeito à invasividade do mesmo. Para tumores localizados esta taxa é de 31%. Por outro lado, para tumores em que há o espalhamento regional (ex.: para linfonodos regionais) ou para tumores metastáticos, estas taxas caem para 11% e 2,7%, respectivamente. O que torna esta estatística ainda mais crítica é o fato de 52% dos cânceres de pâncreas serem detectados já com a presença de metástases, e outros 29% quando já há espalhamento para linfonodos regionais (NATIONAL CANCER INSTITUTE, 2019a).

A abordagem curativa para os pacientes com câncer de pâncreas é a ressecção cirúrgica, e apenas 20% dos pacientes são elegíveis para esta terapia, com tumores em estágio inicial (HAJATDOOST et al., 2018; RYAN; HONG;

BARDEESY, 2014). Na seção a seguir serão descritas algumas características dos quimioterápicos utilizados no câncer de pâncreas.

1.1.2 Terapia do Câncer de Pâncreas

Além da ressecção cirúrgica, existem os procedimentos de radioterapia e quimioterapia que, associados ou não, podem ser utilizados para redução do tamanho do tumor (INSTITUTO NACIONAL DE CÂNCER - INCA, 2018). O análogo de nucleosídeo gemcitabina é a quimioterapia de primeira linha para o câncer de pâncreas metastático (BURRIS et al., 1997), porém o regime de quatro medicamentos, irinotecano, oxaliplatina e fluorouracil (5-FU) modulado por leucovorina (FOLFIRINOX) tem melhorado o prognóstico dos pacientes (CONROY et al., 2011; SHARDA, 2018), com 48,8% de sobrevida em 1 ano evidenciado em ensaio clínico (THOTA; PAUFF; BERLIN, 2014). Outro regime que combina gemcitabina e paclitaxel ligado a nanopartículas de albumina (nab-paclitaxel) elevou a sobrevida de pacientes com câncer de pâncreas metastático, apresentando aumento de 22 pra 35% na sobrevida em 1 ano (CONROY et al., 2011; THOTA; PAUFF; BERLIN, 2014). A quimioterapia adjuvante com 5-flurouracil/leucovorina ou gemcitabina após ressecção cirúrgica reduz a mortalidade em pacientes com câncer de pâncreas (LIAO et al., 2013).

Dentre os estudos relacionados à primeira linha de tratamento, estão os protocolos adjuvantes, neoadjuvantes e não-adjuvantes. Estudos que investigaram quimioterapia adjuvante atribuem a maior média de sobrevida de pacientes com câncer de pâncreas ao tratamento com gemcitabina e tegafur/gimeracil/oteracil (S-1) após a ressecção cirúrgica, com sobrevida de 46,5 meses no estudo com melhor resultado (GOJI et al., 2015; IMAOKA et al., 2016; SHIMODA et al., 2015; UESAKA et al., 2016). O ensaio clínico conduzido por Sherman e colaboradores testou a terapia neoadjuvante de gemcitabina, docetaxel e capecitabine antes da ressecção cirúrgica do câncer de pâncreas, o qual apresentava localmente avançado e irressecável. A média de sobrevida dos pacientes deste estudo foi de 32,5 meses, sendo uma marca promissora, levando em consideração a complexidade do quadro clínico (SHERMAN et al., 2015). O estudo de Hobday e colaboradores apresentou a maior sobrevida de pacientes com câncer de pâncreas no tratamento não-adjuvante combinado de temsirolimus e bevacizumab, com média de 34 meses (HOBDAY et

al., 2015). Os estudos que investigaram a segunda linha de tratamento para o câncer de pâncreas reportaram maior média de sobrevida no protocolos combinados de gemcitabina e cisplatina (35,5 meses) (POSTLEWAIT et al., 2016), e também temsirolimus e bevacizumab (34,0 meses) (HOBDAY et al., 2015). Ainda assim, dada a baixa taxa de sobrevida de pacientes acometidos pelo câncer de pâncreas, mais estudos são necessários para melhor elucidação da temática, focando no diagnóstico precoce, genética e epigenética do desenvolvimento tumoral pancreático (HAJATDOOST et al., 2018).

Embora as células tumorais pancreáticas sejam mais suscetíveis à gemcitabina em comparação com outros agentes anticancerígenos, a maioria dos pacientes desenvolve resistência dentro de semanas do início do tratamento, levando a uma baixa sobrevida. Esta resistência das células tumorais para as drogas anticâncer, muitas vezes ocorre devido a alterações genéticas ou epigenéticas nas células cancerígenas (AMRUTKAR; GLADHAUG, 2017). A **Figura II** apresenta os resultados compilados de ensaios clínicos para terapias adjuvantes, neoadjuvantes e não adjuvantes em pacientes com câncer de pâncreas, exibindo sobrevida global (OS) e Progressão Livre de Doença (PFS) dos pacientes para cada protocolo de terapia.

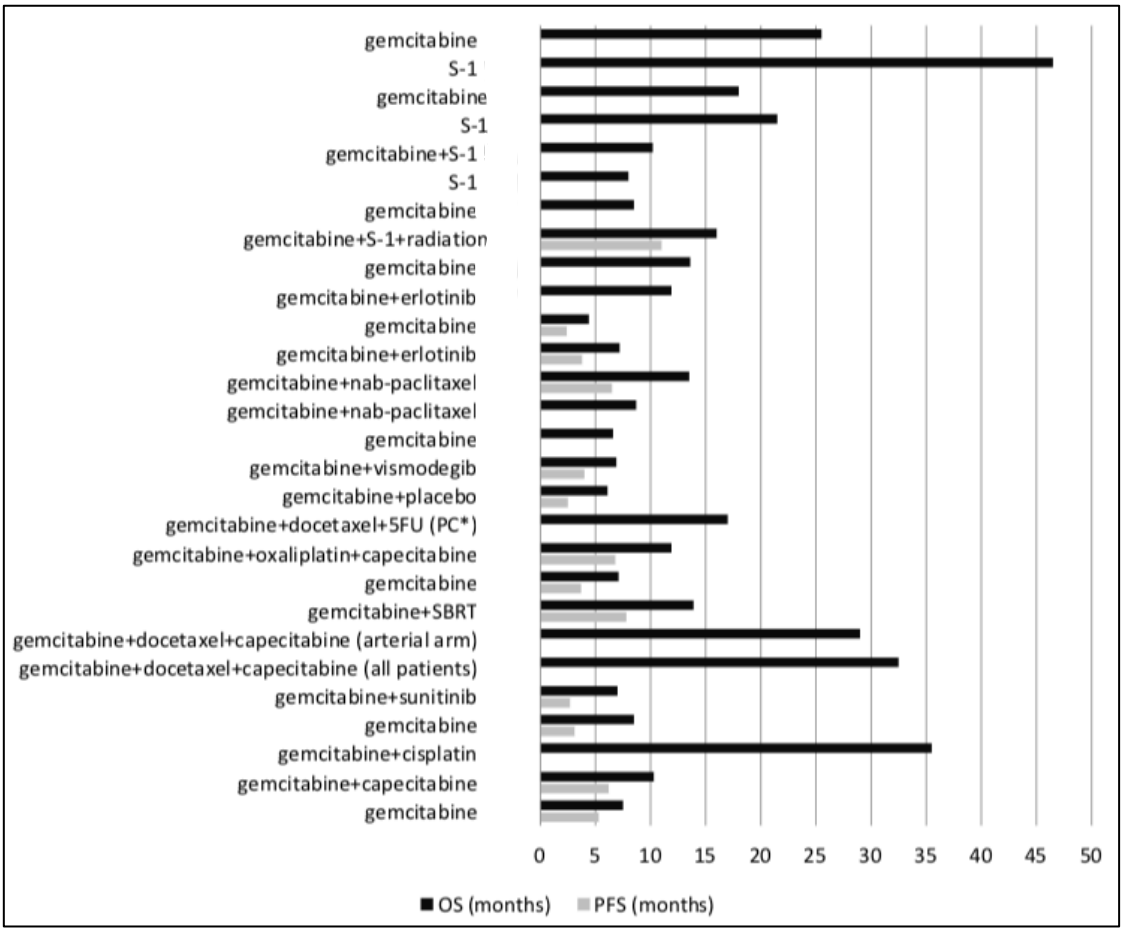


Figura II – Alternativas terapêuticas no câncer de pâncreas. Progressão Livre de Doença (*Progression-Free Disease*, PFS) e Sobrevida Global (*Overall Survival*, OS) para diferentes terapias, em diferentes estudos (Adaptado de HAJATDOOST et al., 2018).

1.1.3 A Biologia Molecular do Câncer de Pâncreas

1.1.3.1 As Vias de Sinalização Alteradas no Câncer de Pâncreas

O câncer de pâncreas se apresenta como uma doença multifatorial com aberrações genéticas e epigenéticas. Mutações em oncogenes e genes supressores tumorais desencadeiam lesões que precedem o câncer de pâncreas, chamadas neoplasias intraepiteliais pancreáticas ou PanIN (ALEMAR; GREGÓRIO; ASHTON-PROLLA, 2015; HRUBAN; WILENTZ; KERN, 2000; IGUCHI et al., 2016; SCHNEIDER et al., 2005). A **Figura III** apresenta uma visão geral da carcinogênese molecular do adenocarcinoma ductal pancreático, o tipo mais frequente e agressivo de câncer de pâncreas.

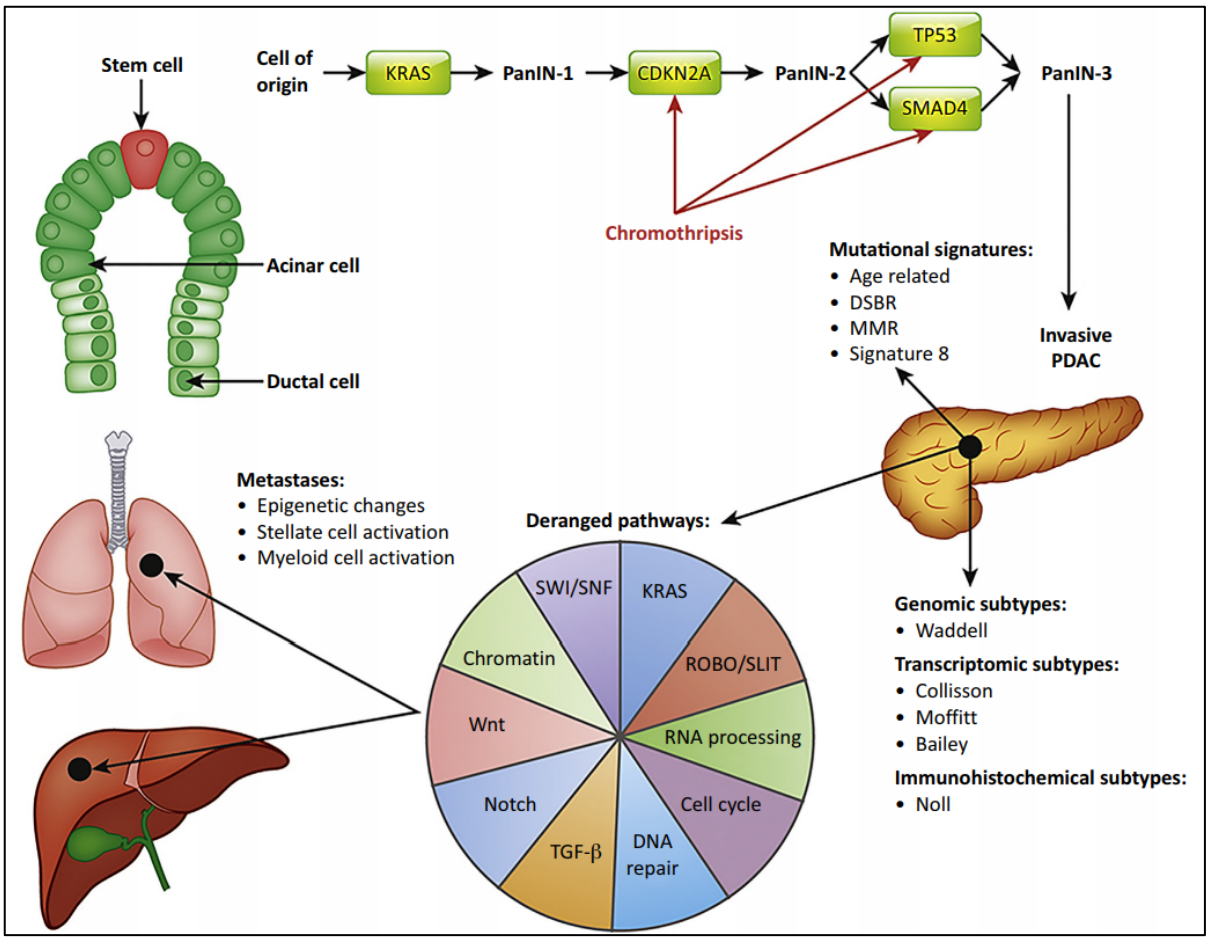


Figura III – Progressão molecular do Adenocarcinoma Ductal de Pâncreas. É mostrada a progressão molecular a partir da célula de origem através de lesões precursores (PanINs) para o câncer ductal invasivo. Na região central são mostradas as principais vias desreguladas ao longo deste processo. À esquerda, abaixo, são mostrados os sítios preferenciais de metástase do adenocarcinoma ductal pancreático (Adaptado de OLDFIELD; CONNOR; GALLINGER, 2017).

A progressão de lesões pré-malignas histologicamente distintas ao carcinoma invasivo em câncer de pâncreas é acompanhada por sucessivas mutações genéticas (KLEEFF et al., 2016; OLDFIELD; CONNOR; GALLINGER, 2017). A grande parte dos pacientes com câncer de pâncreas tem pelo menos uma das quatro mutações comumente encontradas nesse câncer, como o oncogene *KRAS (75,3% mutados)*, os supressores tumorais *TP53 (60,4%), SMAD4 (21,4%) CDKN2A (18,1%)* (GARCEA et al., 2005; OLDFIELD; CONNOR; GALLINGER, 2017; RAPHAEL et al., 2017). Mutações em BRCA2 também apresentam risco aumentado de desenvolver câncer de pâncreas (LE et al., 2016). Essas alterações, juntamente com os diversos processos celulares envolvidos na malignidade e resistência tumoral, são importantes para compreensão do comportamento deste tumor. A **Figura IV** apresenta os genes mais mutados em neoplasias pancreáticas, segundo análise do TCGA (NATIONAL CANCER INSTITUTE, 2019b).

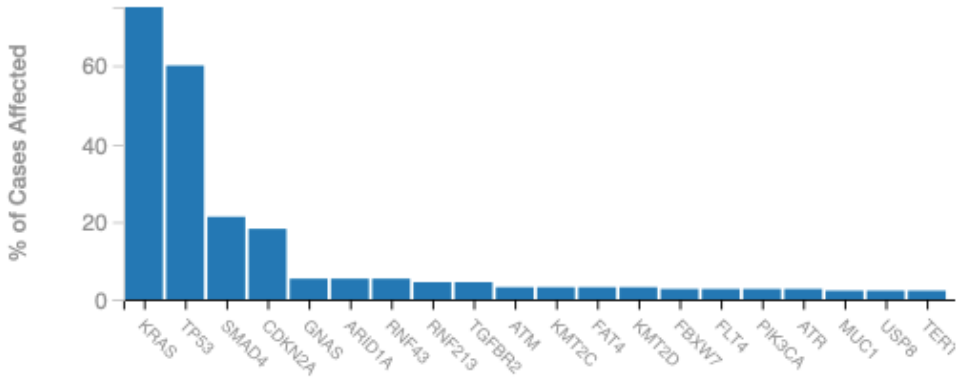


Figura IV - Frequência de mutações gênicas no Adenocarcinoma Ductal Pancreático. (Adaptado de NATIONAL CANCER INSTITUTE, 2019b).

KRAS

O gene *KRAS* participa da sinalização celular como uma chave *on/off* na proliferação celular. Ele pode ser ativado por uma série de sinalizações resultantes de canais iônicos, integrinas e receptores de tirosina cinase, e é importante na sinalização de fatores de transcrição e sobrevivência da célula. Quando livre de mutação, este gene atua automaticamente, ligando e desligando a proliferação celular de acordo com a necessidade. Quando mutado, o desligamento automático não ocorre, estimulando a proliferação celular descontroladamente. Além disso, a mutação em *KRAS* faz com que a proliferação celular passe a ocorrer independente de fatores de transcrição. A **Figura V.a** ilustra as vias que atuam na sinalização da transcrição de genes. A via RAS-RAF-MEK-ERK, na qual *KRAS* atua, participa no crescimento e diferenciação celular. Os fatores de transcrição ativados por *KRAS* também participam de processos como angiogênese, metástase, sobrevivência e atividade anti-apoptótica (RAPHAEL et al., 2017; SIMANSHU; NISSLEY; MCCORMICK, 2017).

TP53

O gene *TP53* é responsável pela codificação da p53, proteína que se localiza no núcleo celular, se liga diretamente ao DNA. O p53 contém domínios que respondem ao *stress* celular e regulam a expressão gênica (**Figura V.b**), induzindo parada no ciclo celular, apoptose, senescência, reparo do DNA ou mudanças no

metabolismo celular. Quando o DNA é danificado, a p53 sinaliza o reparo do mesmo, ou a parada da divisão celular e a morte por apoptose, dependendo da extensão do dano. Impedindo a divisão das células com DNA mutado ou danificado, a p53 previne a proliferação descontrolada e o desenvolvimento tumoral. Mutações no *TP53* estão frequentemente associadas à carcinogênese humana (MARCEL et al., 2010; OLIVIER; HOLLSTEIN; HAINAUT, 2010; YIN et al., 2002).

SMAD4

O gene *SMAD4* codifica proteínas da família Smad (**Figura V.c**), que se localizam no núcleo, ligadas à uma sequência palindrômica específica do DNA, são fosforiladas e ativadas por receptores serina-treonina cinase, regulando a transcrição de genes alvo. O *SMAD4* é um mediador central da sinalização de TGF-β, e é frequentemente inativado, mutado ou deletado em neoplasias, prejudicando a resposta ao dano de DNA e o mecanismo de reparo do mesmo, além de favorecer a instabilidade genômica. A via de sinalização de TGF-β é importante em uma série de processos biológicos, como diferenciação, crescimento e migração celular e apoptose, além do desenvolvimento e progressão tumoral.

Sabe-se que o SMAD4 pode, também, promover a progressão tumoral iniciada por outros genes, como a ativação de *KRAS* em câncer de pâncreas e a inativação de *APC* em câncer colorretal (MIYAKI; KUROKI, 2003; ZHAO; MISHRA; DENG, 2018). A proteína SMAD4 interage com outras SMADs (SMAD1/5 ou 2/3) a partir da ativação de receptores de TGF-β, BMPs, entre outros. Os complexos formados translocam para o núcleo, onde atuam na transcrição de genes envolvidos com migração e proliferação celular, transição epitelio-mesênquima, entre outros. Além de atuarem diretamente na transcrição gênica, também modulam a transcrição por outros fatores, como p53, impedindo a degradação e aumentando a atividade (CHAU et al., 2012).

CDKN2A

O gene *CDKN2A* é localizado no núcleo e atua na formação de diversas proteínas, sendo $p16^{INK4a}$ e $p14^{ARF}$ as mais conhecidas. Ambas funcionam como supressores tumorais, indutores de senescência, reguladores de ciclo celular; tanto

estimulando a progressão do ciclo ($p16^{INK4a}$) quanto estabilizando p53 ($p14^{ARF}$). O silenciamento de *CDKN2A* é associado a alterações na regulação do ciclo celular, justamente por interferir na produção de $p16^{INK4a}$ e $p14^{ARF}$, as quais regulam cinases dependentes de ciclina (CDK) (**Figura V.d**). A perda da função de *CDKN2A* é, desta forma, presente no processo de carcinogênese (OLDFIELD; CONNOR; GALLINGER, 2017; ZHAO et al., 2016).

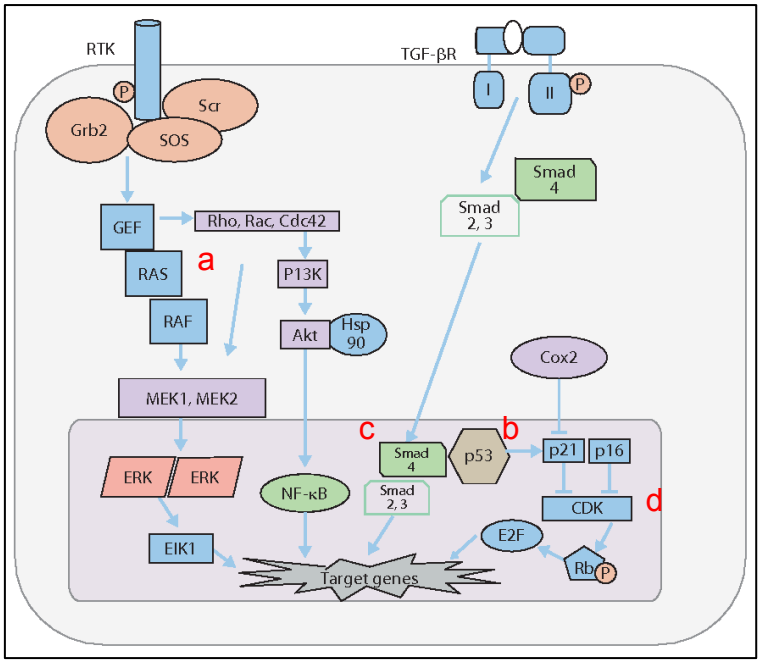


Figura V - Vias de sinalização de genes alvo para crescimento e proliferação celular. (a) via de atuação de *KRAS*; (b) via de atuação de *TP53*; (c) via de atuação de *SMAD4;* via de atuação de *CDKN2A* (Adaptado de MATTHAIOS et al., 2011).

1.1.3.2 Marcadores Prognóstico

O que é, função, importância

Pouco se sabe sobre marcadores prognósticos efetivos em câncer de pâncreas, e embora apresentem limitações, os marcadores séricos CA19-9 e CEA continuam sendo os marcadores prognósticos clínicos da doença (LE et al., 2016). Estudos associam modificações em genes componentes das vias de sinalização de Hedgehog, Notch, Wnt, TGF-β, fatores de crescimento, como o EGF e o respectivo receptor, EGFR, com o desenvolvimento de câncer de pâncreas (ADAMSKA; DOMENICHINI; FALASCA, 2017), e indicam possíveis biomarcadores desta patologia, como MIC-1, MMP7, MUC4, FSCN1 e MUC5AC (HARSHA et al., 2009; LE et al., 2016). Estudos indicam, também, que mutações ou hipermetilação de *TP16* estejam associadas à menor sobrevida de pacientes com câncer de pâncreas.

Há a necessidade de mais estudos que gerem dados de marcadores prognósticos para o câncer de pâncreas. As ferramentas de bioinformática são promissoras alternativas que utilizam como base amostras de pacientes para análise (GERDES et al., 2002; LE et al., 2016; OHTSUBO et al., 2003).

1.2 TCGA e Ferramentas de Bioinformática

O *The Cancer Genome Atlas* (TCGA) é uma colaboração entre o Instituto Nacional do Câncer dos Estados Unidos (NIH) e o Instituto Nacional de Pesquisa do Genoma Humano dos Estados Unidos. Esta colaboração abrange uma coleção de fenótipos clínicos e moleculares de mais de 11 mil pacientes com tumores, e 33 tipos tumorais diferentes. O TCGA torna público detalhadas características e alterações genômicas e epigenômicas associadas aos tipos tumorais desta coleção, possibilitando o estudo de câncer em larga escala (COLAPRICO et al., 2016; LEE et al., 2015; THE CANCER GENOME ATLAS - TCGA, 2018).

A partir dos dados do TCGA, pode-se testar hipóteses *in silico*, através de ferramentas de bioinformática. O *software* R é uma plataforma que une a linguagem de programação, o desenvolvimento de cálculos estatísticos, análise de dados e geração de gráficos para visualização dos resultados. Esta plataforma conta com bibliotecas, *plug-ins* e pacotes que auxiliam nas análises e na produção de resultados mais específicos para cada área de estudo (RSTUDIO TEAM, 2016).

O TCGAbiolinks é um pacote do *Bioconductor* que facilita a recuperação e processamento de dados do *Genomic Data Commons Data Portal* (GDC) no R. O GDC é outra plataforma do NIH, desenvolvida para pesquisa e download de dados de câncer. Além disso, o TCGAbiolinks fornece métodos de análise, como expressão diferencial de genes, e métodos para visualização de resultados, como gráficos de sobrevida e *volcano plot*, tornando a sua utilização conveniente para as análises *in silico* (GROSSMAN et al., 2016; LAWRENCE et al., 2013).

O TCGA Browser é uma ferramenta que avalia a sobrevida de pacientes de câncer, relacionando a doença a um gene. Pode-se definir o percentil de pacientes que será definido como baixa e alta expressão do gene de interesse, e assim se estabelece o Kaplan-Meier que representa a curva de sobrevida dos pacientes. O navegador também inclui dados como idade, sexo, etnia, estadiamento tumoral e dados clínicos (CHENG; DUMMER; LEVESQUE, 2015).

O *The Human Protein Atlas* é uma plataforma que tem como objetivo mapear proteínas humanas em células, tecidos e órgãos, utilizando uma série de ferramentas experimentais e *in silico*. Através desse navegador, pode-se identificar a distribuição de proteínas nos tecidos e órgãos humanos, além detectar a proteína dentro da célula, em suas diferentes organelas. É possível, também, acessar o impacto da variação nos níveis das proteínas na sobrevida de pacientes com câncer, contribuindo para o entendimento dos componentes que envolvem a carcinogênese (UHLEN et al., 2015).

STRING é uma base de dados que apresenta interações proteína-proteína, sendo elas conhecidas ou previstas. São cerca de 9.643.763 proteínas detectadas através de dados experimentais e análise textual automatizada. As interações incluem associações diretas ou funcionais, baseadas em previsão computacional, co-expressão e estudos com outros organismos, ou incorporadas de outras bases de dados. É uma ferramenta relevante para identificar proteínas que interagem em redes de sinalização distintas no mesmo organismo (SNEL, 2000).

A bioinformática se consolidou como aliada na obtenção de informações sobre a biologia de tumores, por integrar dados genômicos de câncer em escalas moleculares. Mesmo que muitos tumores apresentem eventos genômicos semelhantes e recorrentes, a relação com o fenótipo nem sempre é bem compreendida. Logo, a bioinformática é importante para compreensão da genética e epigenética de tumores, e na elucidação de novas vias biológicas e marcadores diagnósticos para câncer (COLAPRICO et al., 2016; SILVA et al., 2016; THE CANCER GENOME ATLAS - TCGA, 2018).

2 JUSTIFICATIVA

Muito se tem estudado sobre câncer de pâncreas, pois a detecção da doença é complexa, a mortalidade dos pacientes é alta e os tratamentos disponíveis ainda não apresentam resultados de sobrevida satisfatórios. Os mecanismos gênicos e celulares que implicam na formação do câncer de pâncreas, o desenvolvimento de metástases, a resistência à terapia e outros fatores ainda requerem esclarecimentos. É relevante o estudo da expressão diferencial dos genes presentes nos tumores de pâncreas em comparação com tecido pancreático saudável, pois pode permitir o desenvolvimento de novas estratégias diagnósticas, alvos terapêuticos e possíveis biomarcadores.

3 QUESTÃO DE PESQUISA

Quais genes expressos diferencialmente em tumores pancreáticos apresentam potencial como fatores prognósticos nesta neoplasia?

4 HIPÓTESE

Existem genes expressos diferencialmente em tumores pancreáticos que apresentam potencial como fatores prognósticos nesta neoplasia.

5 OBJETIVOS

5.1 Objetivo Geral

Identificar potenciais marcadores de prognóstico tumoral em neoplasias pancreáticas.

5.2 Objetivos Específicos

a) Determinar os genes diferencialmente expressos entre tecido pancreático normal e tumores pancreáticos.
b) Avaliar a influência dos níveis de expressão dos genes diferencialmente expressos na sobrevida dos pacientes.
c) Investigar o papel biológico dos genes que apresentaram fator prognóstico.
d) Avaliar a relação dos genes diferencialmente expressos entre tecido pancreático normal e tumoral com relação ao status de *KRAS*.

REFERÊNCIAS

ADAMSKA, A.; DOMENICHINI, A.; FALASCA, M. Pancreatic Ductal Adenocarcinoma: Current and Evolving Therapies. **International Journal of Molecular Sciences**, v. 18, n. 7, p. 1338, 22 jun. 2017.

ALEMAR, B.; GREGÓRIO, C.; ASHTON-PROLLA, P. miRNAs as Diagnostic and Prognostic Biomarkers in Pancreatic Ductal Adenocarcinoma and Its Precursor Lesions: A Review. **Biomarker Insights**, v. 10, p. BMI.S27679, 7 jan. 2015.

AMRUTKAR, M.; GLADHAUG, I. Pancreatic Cancer Chemoresistance to Gemcitabine. **Cancers**, v. 9, n. 12, p. 157, 16 nov. 2017.

BURRIS, H. A. et al. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. **Journal of clinical oncology : official journal of the American Society of Clinical Oncology**, v. 15, n. 6, p. 2403–13, jun. 1997.

CHAU, J. F. L. et al. A crucial role for bone morphogenetic protein-Smad1 signalling in the DNA damage response. **Nature Communications**, v. 3, n. 1, p. 836, 15 jan. 2012.

CHENG, P. F.; DUMMER, R.; LEVESQUE, M. P. Data mining The Cancer Genome Atlas in the era of precision cancer medicine. **Swiss Medical Weekly**, v. 145, n. September, p. 1–5, 2015.

COLAPRICO, A. et al. TCGAbiolinks : an R/Bioconductor package for integrative analysis of TCGA data. **Nucleic Acids Research**, v. 44, n. 8, p. e71–e71, 5 maio 2016.

CONROY, T. et al. FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. **New England Journal of Medicine**, v. 364, n. 19, p. 1817–1825, 12 maio 2011.

GARCEA, G. et al. Molecular prognostic markers in pancreatic cancer: A systematic review. **European Journal of Cancer**, v. 41, n. 15, p. 2213–2236, out. 2005.

GERDES, B. et al. p16 INK4a is a Prognostic Marker in Resected Ductal Pancreatic Cancer. **Annals of Surgery**, v. 235, n. 1, p. 51–59, 2002.

GOJI, T. et al. A phase I/II study of fixed-dose-rate gemcitabine and S-1 with concurrent radiotherapy for locally advanced pancreatic cancer. **Cancer Chemotherapy and Pharmacology**, v. 76, n. 3, p. 615–620, 29 set. 2015.

GROSSMAN, R. L. et al. Toward a Shared Vision for Cancer Genomic Data. **New England Journal of Medicine**, v. 375, n. 12, p. 1109–1112, 22 set. 2016.

HAJATDOOST, L. et al. Chemotherapy in Pancreatic Cancer: A Systematic Review. **Medicina**, v. 54, n. 3, p. 48, 11 jul. 2018.

HARSHA, H. C. et al. A Compendium of Potential Biomarkers of Pancreatic Cancer. **PLoS Medicine**, v. 6, n. 4, p. e1000046, 7 abr. 2009.

HOBDAY, T. J. et al. Multicenter Phase II Trial of Temsirolimus and Bevacizumab in Pancreatic Neuroendocrine Tumors. **Journal of Clinical Oncology**, v. 33, n. 14, p. 1551–1556, 10 maio 2015.

HRUBAN, R. H.; WILENTZ, R. E.; KERN, S. E. Genetic Progression in the Pancreatic Ducts. **The American Journal of Pathology**, v. 156, n. 6, p. 1821–1825, jun. 2000.

HRUBAN, R.; HORNSBY, E. **The Sol Goldman Pancreatic Cancer Research Center**. Disponível em: <http://pathology.jhu.edu/pc/BasicTypes2.php?area=ba>. Acesso em: 10 nov. 2018.

IGUCHI, E. et al. Pancreatic Cancer, A Mis-interpreter of the Epigenetic Language. **The Yale journal of biology and medicine**, v. 89, n. 4, p. 575–590, dez. 2016.

IMAOKA, H. et al. Clinical outcome of elderly patients with unresectable pancreatic cancer treated with gemcitabine plus S-1, S-1 alone, or gemcitabine alone: Subgroup analysis of a randomised phase III trial, GEST study. **European Journal of Cancer**, v. 54, p. 96–103, fev. 2016.

INSTITUTO NACIONAL DE CÂNCER - INCA. **Câncer de Pâncreas**. Disponível em: <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pancreas>. Acesso em: 12 set. 2017.

KIERSZENBAUM, A. L.; TRES, L. L. **Histologia e Biologia Celular - Uma Introdução À Patologia**. 3. ed. [s.l.] Elsevier, 2012.

KLEEFF, J. et al. Pancreatic cancer. **Nature reviews. Disease primers**, v. 2, p. 16022, 21 abr. 2016.

LAWRENCE, M. et al. Software for Computing and Annotating Genomic Ranges. **PLoS Computational Biology**, v. 9, n. 8, p. e1003118, 8 ago. 2013.

LE, N. et al. Prognostic and predictive markers in pancreatic adenocarcinoma. **Digestive and Liver Disease**, v. 48, n. 3, p. 223–230, mar. 2016.

LEE, H. et al. The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical–genomic driver associations. **Genome Medicine**, v. 7, n. 1, p. 112, 27 dez. 2015.

LIAO, W.-C. et al. Adjuvant treatments for resected pancreatic adenocarcinoma: a systematic review and network meta-analysis. **The Lancet Oncology**, v. 14, n. 11, p. 1095–1103, out. 2013.

MARCEL, V. et al. Δ160p53 is a novel N-terminal p53 isoform encoded by Δ133p53 transcript. **FEBS Letters**, v. 584, n. 21, p. 4463–4468, 5 nov. 2010.

MATTHAIOS, D. et al. Molecular Pathogenesis of Pancreatic Cancer and Clinical Perspectives. **Oncology**, v. 81, n. 3–4, p. 259–272, 2011.

MIYAKI, M.; KUROKI, T. Role of Smad4 (DPC4) inactivation in human cancer. **Biochemical and Biophysical Research Communications**, v. 306, n. 4, p. 799–804, jul. 2003.

NATIONAL CANCER INSTITUTE. **Pancreatic Cancer - Health Professional Version**. Disponível em: <https://www.cancer.gov/types/pancreatic/hp>.

NATIONAL CANCER INSTITUTE. **Genomic Data Commons Data Portal**. Disponível em: <https://portal.gdc.cancer.gov>.

OHTSUBO, K. et al. Abnormalities of tumor suppressor gene p16 in pancreatic carcinoma: immunohistochemical and genetic findings compared with clinicopathological parameters. **Journal of Gastroenterology**, v. 38, n. 7, p. 663–671, 1 jul. 2003.

OLDFIELD, L. E.; CONNOR, A. A.; GALLINGER, S. Molecular Events in the Natural History of Pancreatic Cancer. **Trends in Cancer**, v. 3, n. 5, p. 336–346, maio 2017.

OLIVIER, M.; HOLLSTEIN, M.; HAINAUT, P. *TP53* Mutations in Human Cancers: Origins, Consequences, and Clinical Use. **Cold Spring Harbor Perspectives in Biology**, v. 2, n. 1, p. a001008–a001008, 1 jan. 2010.

POSTLEWAIT, L. M. et al. Combination gemcitabine/cisplatin therapy and ERCC1 expression for resected pancreatic adenocarcinoma: Results of a Phase II prospective trial. **Journal of Surgical Oncology**, v. 114, n. 3, p. 336–341, set. 2016.

RAPHAEL, B. J. et al. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. **Cancer Cell**, v. 32, n. 2, p. 185- 203.e13, ago. 2017.

RSTUDIO TEAM. **RStudio: Integrated Development Environment for R** Boston, MA. RStudio, Inc., , 2016. Disponível em: <http://www.rstudio.com/>

RYAN, D. P.; HONG, T. S.; BARDEESY, N. Pancreatic Adenocarcinoma. **New England Journal of Medicine**, v. 371, n. 11, p. 1039–1049, 11 set. 2014.

SCHNEIDER, G. et al. Pancreatic Cancer: Basic and Clinical Aspects. **Gastroenterology**, v. 128, n. 6, p. 1606–1625, maio 2005.

SHARDA, M. **Treatment Changes for Pancreatic Cancer and Colorectal Cancer**. Disponível em: <https://www.cancer.net/blog/2018-06/asco-annual-meeting-2018-treatment-changes-pancreatic-cancer-and-colorectal-cancer>.

SHERMAN, W. H. et al. Neoadjuvant gemcitabine, docetaxel, and capecitabine followed by gemcitabine and capecitabine/radiation therapy and surgery in locally advanced, unresectable pancreatic adenocarcinoma. **Cancer**, v. 121, n. 5, p. 673–680, 1 mar. 2015.

SHIMODA, M. et al. Randomized clinical trial of adjuvant chemotherapy with S-1 versus gemcitabine after pancreatic cancer resection. **British Journal of Surgery**, v. 102, n. 7, p. 746–754, jun. 2015.

SILVA, T. C. et al. TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. **F1000Research**, v. 5, n. 0, p. 1542, 28 dez. 2016.

SIMANSHU, D. K.; NISSLEY, D. V.; MCCORMICK, F. RAS Proteins and Their Regulators in Human Disease. **Cell**, v. 170, n. 1, p. 17–33, jun. 2017.

SNEL, B. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. **Nucleic Acids Research**, v. 28, n. 18, p. 3442–3444, 15 set. 2000.

THE CANCER GENOME ATLAS - TCGA. **The Cancer Genome Atlas**. Disponível em: <https://cancergenome.nih.gov>. Acesso em: 23 mar. 2018.

THOTA, R.; PAUFF, J. M.; BERLIN, J. D. Treatment of metastatic pancreatic adenocarcinoma: a review. **Oncology (Williston Park, N.Y.)**, v. 28, n. 1, p. 70–4, jan. 2014.

UESAKA, K. et al. Adjuvant chemotherapy of S-1 versus gemcitabine for resected pancreatic cancer: a phase 3, open-label, randomised, non-inferiority trial (JASPAC 01). **The Lancet**, v. 388, n. 10041, p. 248–257, jul. 2016.

UHLEN, M. et al. Tissue-based map of the human proteome. **Science**, v. 347, n. 6220, p. 1260419–1260419, 23 jan. 2015.

WINSLOW, T. **Pancreatic Cancer Staging**. Disponível em: <https://visualsonline.cancer.gov/details.cfm?imageid=12034>.

YIN, Y. et al. p53 stability and activity is regulated by Mdm2-mediated induction of alternative p53 translation products. **Nature Cell Biology**, v. 4, n. 6, p. 462–467, 27 jun. 2002.

ZHAO, M.; MISHRA, L.; DENG, C.-X. The role of TGF-β/SMAD4 signaling in cancer. **International Journal of Biological Sciences**, v. 14, n. 2, p. 111–123, 2018.

ZHAO, R. et al. Implications of Genetic and Epigenetic Alterations of CDKN2A (p16 INK4a) in Cancer. **EBioMedicine**, v. 8, n. 127, p. 30–39, jun. 2016.

6 ARTIGO ORIGINAL

A ser submetido ao periódico 'Pancreas'.

New Potential Prognostic Factors in Pancreatic Cancer

Mariana dos Santos Lobo[1]; Ana Carolina Mello[2]; Tiago Falcon[2]; Eduardo Cremonese Filippi-Chiela[1,2,3*].

1 – Graduate Program Science in Gastroenterology and Hepathology, Faculty of Medicine, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brasil;

2 – Clinicas Hospital of Porto Alegre, Rio Grande do Sul, Brasil;

3 – Morphological Sciences Department, ICBS, UFRGS.

* Corresponding author: eduardochiela@gmail.br

Hospital de Clínicas de Porto Alegre

Rua Ramiro Barcelos, 2500

Zip code: 90035-003 Porto Alegre, RS, Brazil; Phone: +55 51 3308 4420

Abstract

Pancreatic cancer is one of the most lethal solid malignancies with increasing incidence. The poor prognosis is due to late detection, aggressive nature of the tumor, and resistance to chemotherapy and radiotherapy. Characterization of the recurrent genetic alterations in pancreatic cancer has yielded important insights into the biology of this disease, and also supports new early detection strategies and therapy development.

Objectives: characterize the differential expression of the cohort of pancreatic neoplasms of the TCGA in relation to normal pancreatic tissue in the search for potential markers of tumor prognosis.

Methods: Data was analyzed using the R software and the TCGAbiolinks package. The PAAD harmonized total RNA data was obtained through RNA-seq using Illumina HiSeq platform.

Results: 235 differentially expressed genes were found in tumor samples compared to normal tissue. 28 genes had expression levels at least 3-fold different from normal tissue; 9 presented as prognostic factors. Elevated levels of *MAST4, SPRY3, USP19* and *TRIM67* were associated with better prognosis, whereas the opposite was observed for *AHCY, FAM45A, LYPLA1, TC2N* and *TGM5*.

Conclusions: Through TCGA databank we described 9 potential prognostic factors in PAAD, and also explore the biological role and its relevance in these tumors' biology.

Introduction

Pancreatic cancer has one of the biggest mortality rates and aggressive features among all cancers. Its detection is difficult due to complex biological behavior. The survival rate is significantly low, and the therapeutic strategies still need improvements, because the patient's prognosis is still poor. Studies show that *KRAS*, *TP53*, *SMAD4* and *CDKN2A* are the main genes involved pancreatic carcinogenesis, and its pathways' components may be potential targets for future treatments[1–3]. Despite the improvement in the knowledge of basic cancer biology that has led to increments in patient's survival, the prognosis of patients diagnosed with pancreatic cancer has not advanced for decades. Thus, searching for new early diagnostic and prognostic markers or therapeutic targets can contribute to the clinical management of these cancers.

Bioinformatics integrates the omic data regarding tumor development. Through bioinformatics data it is possible to determine differential expression profiles among tumor and normal tissue, seek for novel biomarkers for diagnosis and prognosis, develop therapeutic strategies and understand the gene expression role on survival rates, producing substantial biological impact at low costs, considering that most of the data is available through free access in public database. TCGA is a collaboration between institutions from United States and Canada that covers a collection of clinical and molecular phenotypes of more than 11,000 tumor patients, and 33 different tumor types. The TCGA discloses detailed genomic and epigenomic characteristics and alterations associated with this collection's tumor types, making possible the study of cancer in large scale. The TCGA's cohort of pancreatic samples, normal and tumor, is called PAAD[4–7].

The aim of this study was to identify novel prognostic markers for pancreatic neoplasias. We also evaluated the differential expression genes considering *KRAS* mutation in PAAD. From more than 250 differentially expressed genes in tumor samples in comparison to normal tissue, 9 of them showed a significant prognostic factor. None of these genes was previously associated with PAAD. These genes are related to the modulation of the main pathways altered in PAAD (i.e. *KRAS*, *TP53* and *SMAD*), thus controlling cell growth, death, migration and metabolism.

Materials and Methods

All data used in this study is available at TCGA via GDC Data Portal website[8] and was analyzed using the R software (version 1.1.383)[9]. The TCGAbiolinks package (available at Bioconductor, version release 3.9)[6] was used to download and preprocess the samples. The differential expression analysis (DEA) and volcano plots were also performed using TCGAbiolinks. At first, a DEA was performed among normal tissue samples (NT) and primary solid tumor (TP). Subsequently, the DEA was performed among tumor samples that presented mutation in *KRAS* oncogene and tumor samples without the *KRAS* mutation.

Download and Preprocessing

The PAAD harmonized total RNA data (hg38), downloaded from TCGA, was obtained through RNA-seq using Illumina HiSeq platform. First, a query was set using the function *GDCquery* with the following options *project="TCGA-PAAD", data.category="Transcriptome Profiling", data.type="Gene" Expression Quantification", workflow.type="HTSeq-counts", legacy=FALSE".*

The samples download was performed using the function *GDCdownload* with the previously set query as the main variable, and the option *method="api"*. Next, the samples were prepared in order to perform the analyses, using the function *GDCprepare* with the previously set query. Then, the data was normalized using the function *TCGAanalyze_Normalization*. It was first performed by GC content with the options *geneInfo="geneInfoHT"* and *method="gcContent",* and then by gene length changing the method option to *"geneLenght"*. After this step, a quantile filter was applied using the function *TCGAanalyze_Filtering* with the option *method="quantile"* and the *qnt.cut=0.25*. A total of 181 samples were downloaded (177 TP and 4 NT), including 29,959 genes.

NT *versus* TP Differential Expression

The DEA was performed with the function *TCGAanalyze_DEA* and the options *fdr.cut=0.01*, *logFC.cut=3* and *method="glmLRT"*. All samples downloaded were analyzed. A logFC cut of |±3| was considered to define the differentially expressed genes. A volcano plot was performed using the *gplot* package, also considering a logFC cut of |±3|.

KRAS^{wt} *versus KRAS*^{mut} Differential Expression

The DEA was performed with the function *TCGAanalyze_DEA* and the options *fdr.cut=0.01*, *logFC.cut=3* and *method="glmLRT"*. All samples downloaded were analyzed. A logFC cut of |±3| was considered to define the differentially expressed genes. A volcano plot was performed using the *gplot* package, also considering a logFC cut of |±3|.

The Human Protein Atlas (THPA)

The Human Protein Atlas website (version 18.1)[10] was used to filter the genes obtained after the DEA described in the previous section. The website determines the best separation cutoff for the survival curve and establishes a p-value for the genes. It was considered a p-value ≤0.05 to include the genes on this project.

TCGA Browser

To identify the survival curve of PAAD patients, the TCGA Browser platform was used, selecting "Pancreatic Adenocarcinoma" under the tab "cancer", selecting the gene of interest on the tab "RNAseq explorer", and scrolling through the percentile tab to define the best option. The percentiles selected were 10%, 25% and 50%.

String

In order to assess the protein network around the remaining proteins of NT *versus* TP DEA, String website[11] was used. We performed the analysis using data from experimental evidence only, a minimum required interaction score of 0.7 (high confidence) and the "Homo-sapiens" organism[11,12].

Venny

The Venny website[13] was used to access the genes concomitants in both NT *versus* TP and *KRAS*^{wt} *versus KRAS*^{mut} analysis.

RESULTS

Pipeline

Our strategy of action is shown in **Figure 1**. Briefly, TCGA data were obtained and DEA was performed to compare NT and TP. In addition to this, a DEA was also performed to compare TP samples with normal *versus* mutated *KRAS*. A Venn diagram was build to compare the results obtained from both DEA.

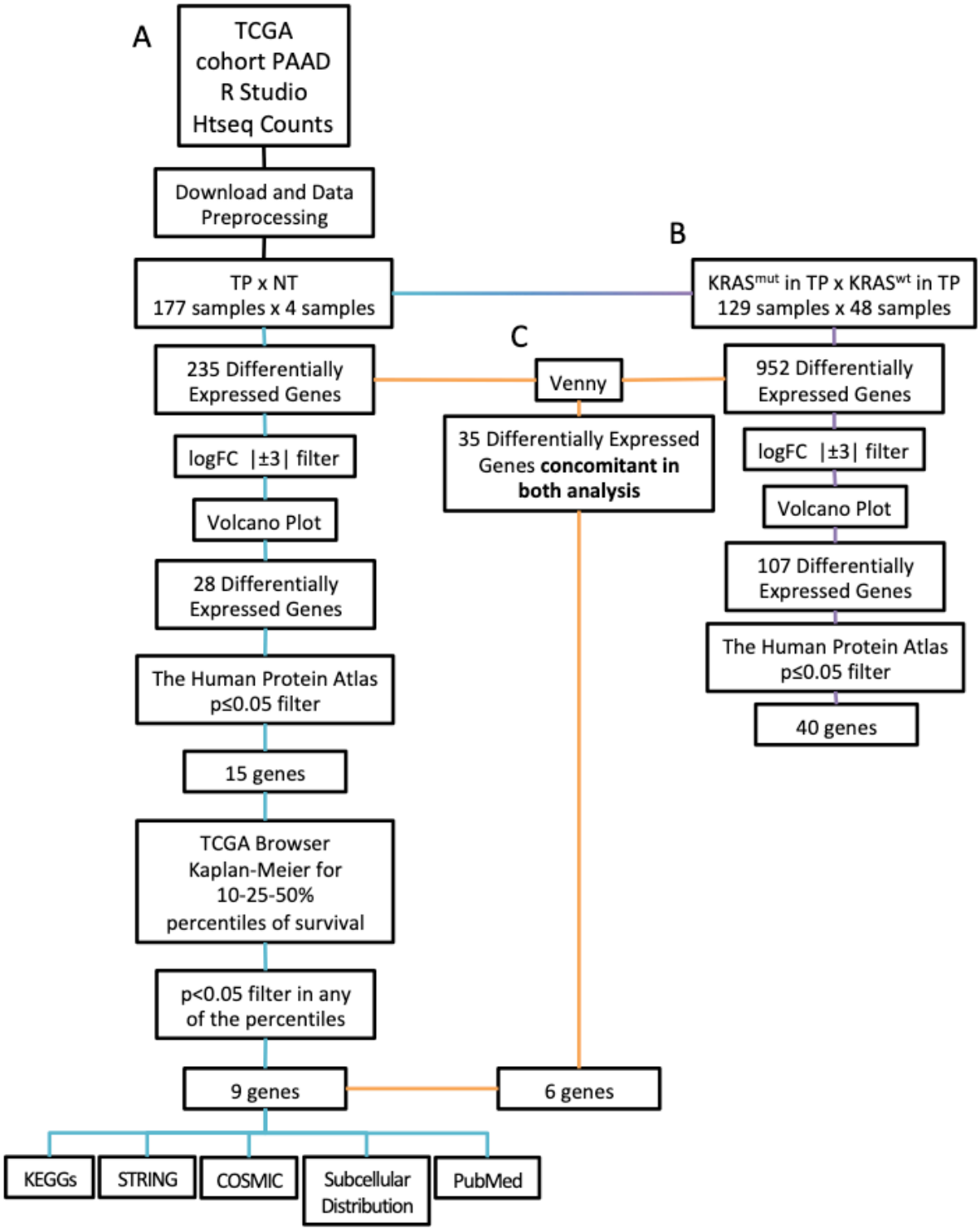


Figure 1 – Pipeline of the study. This workflow presents the steps of this project's analysis. The lines in blue refer to the NT *versus* TP data (A). The lines in purple refer to the *KRAS*^{wt} *versus KRAS*^{mut} data (B). The lines in orange refer to the data concomitant in both DEA (C).

NT *versus* TP DEA

From the total of 29.959 genes, 28 were differentially expressed (hereafter called DEGs) considering a logFC cut of |±3|. The volcano plot (**Figure 2**) shows the distribution of the differentially expressed genes. In green are the 25 increased DEGs (logFC ≥3), and in red are the 3 decreased DEGs (logFC ≤-3). The list of genes, along with the respective logFC, gene code and p-value are available on **Supplementary Table 1**.

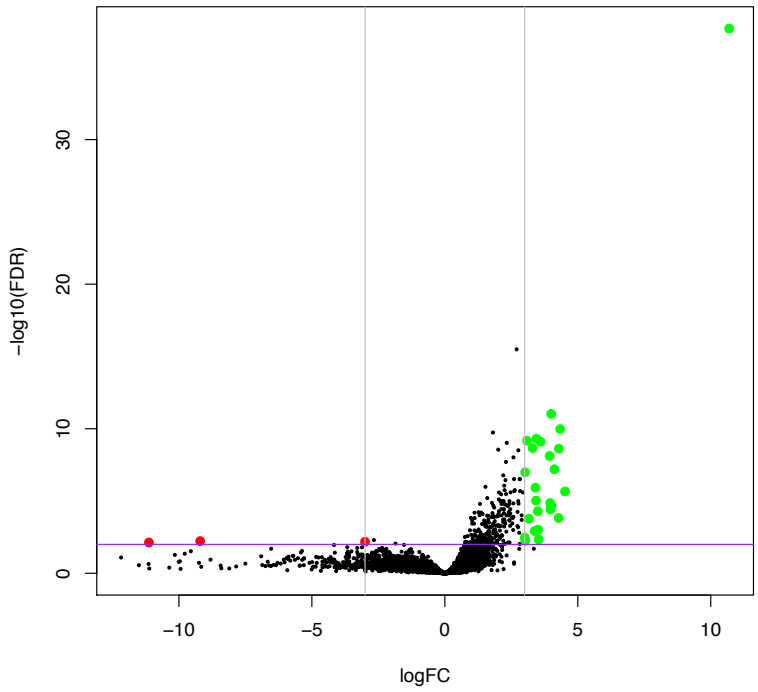


Figure 2 – NT *versus* TP Differential Expression Analysis. This volcano plot shows the DEG distribution. In red are the decreased DEGs. In green are the increased DEGs. The line in purple refers to the adjusted p-value of 0,01. The lines in grey refer to the logFC cut of |±3|.

The Prognostic Role Of Differentially Expressed Genes

We firstly assessed the influence of DEGs in the survival of PAAD patients using the THPA website. THPA calculates Kaplan-Meier survival plots considering the separation of patients expressing high or low levels of a given gene based on two

conditions: the median or the ‘best separation’ cutoff, i.e. the separation that shows the highest statistical significance between groups.

The genes remaining from the logFC |±3| cut (28 DEGs for NT *versus* TP) were tested for their influence in the survival of PAAD using the best separation cutoff in THPA. To the TP x NT analysis, expression levels of 15 from 28 DEGs influenced the survival of PAAD (**Table 1 - "THPA p-value" column**). After filtering DEGs using the THPA, we next assessed the influence of expression levels of DEGs from NT *versus* TP in the survival of PAAD patients using the TCGA Browser (**Table 1**). This browser allows the separation of patients (in low expression level and high expression level) based on specific percentiles. Then, the survival curve of PAAD patients was assessed for each remaining gene from the previous steps using the percentiles 10% (**Supplementary Figure 1**), 25% (upper and lower quartiles, **Figure 3**) and 50% (Median, **Supplementary Figure 2**). Through this analysis, we found that 9 from 15 DEGs influenced the survival of PAAD in at least one percentile. From these 9 genes, 4 showed a statistical difference to all percentiles (TRIM67, TC2N, AHCY, MAST4).

Table 1 – Influence of DEG's expression levels in the survival of patients according to the THPA or TCGABrowser.

		Percentil 10%			Percentil 25%			Percentil 50%		
		mOS			mOS			mOS		
	THPA p-value	High	Low	p-value	High	Low	p-value	High	Low	p-value
TRIM67	4.90E-10	5.57	0.78	*0.000*	5.57	1.04	*0.000*	2.5	1.33	*0.001*
AHCY	0.042	1.21	3.63	*0.028*	1.74	4.67	*0.000*	2.13	3.12	*0.013*
MAST4	0.05	2.01	5.91	*0.006*	2.47	5.54	*0.008*	2.4	3.2	*0.02*
TC2N	0.000	1.42	NA	0.056	1.62	3.65	*0.016*	1.49	2.02	*0.01*
LYPLA1	0.000	1.42	2.9	0.12	1.08	2.02	*0.003*	1.49	1.82	*0.035*
USP19	0.023	NA	1.36	*0.002*	5.97	1.64	*0.02*	1.81	1.62	0.36
TGM5	0.022	1.25	5.57	*0.007*	1.36	2.5	0.064	1.79	1.65	0.98
FAM45A	0.045	3.19	2.06	0.24	3.19	2.34	0.23	3.24	2.34	*0.04*
SPRY3	0.012	1.42	1.81	0.45	1.92	1.36	*0.053*	1.74	1.65	0.16
RILPL1	0.023	NA	NA	0.92	2.02	1.42	0.14	1.68	1.63	0.23
TIAF1	0.006	1.79	1.62	0.31	1.79	1.63	0.14	1.87	1.63	0.078
PXN	0.027	0.8	2.02	0.13	1.3	1.79	0.12	1.62	1.79	0.19
MTMR11	0.029	1.46	NA	0.4	1.49	1.79	0.13	1.49	1.74	0.42
DEAF1	0.003	NA	1.52	0.12	1.79	1.42	0.13	1.9	1.62	0.06
SLCO2A1	0.02	3.63	2.84	0.61	2.57	2.1	0.83	2.9	2.4	0.85

Red, blue, purple and gray lines: DEGs showing prognostic value to three, two, one or no percentiles tested, respectively.

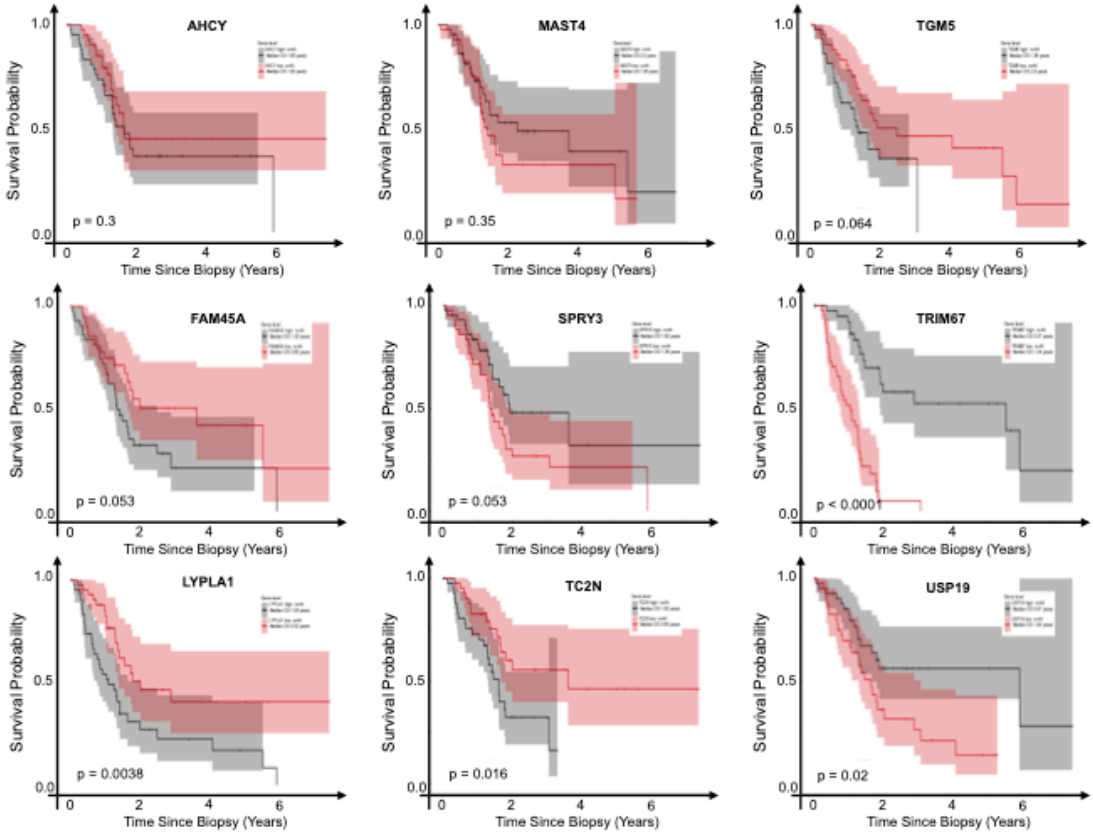


Figure 3 – Kaplan-Meier survival analysis to the 9DEG^{NTxTP}. These graphs show the survival probability of PAAD patients (n=88, 44 as high expression group and 44 as low expression group) based on the 9DEGs selected from NT *versus* TP DEA. The red lines present the low level genes and the black lines present the high gene levels. The each analysis p-value is presented on the inferior left corner of the respective gene graph.

DEGs Protein Networks and KEGG Functional Annotation

To shed some light on the biological function of the 9DEGs in cancer, we: (1) assessed the KEGG functional annotation; (2) look for molecular connections using the STRING browser; (3) checked for the subcellular distribution; (4) reviewed the literature.

We firstly explore the KEGG Functional Annotation to assess the biological functions of the proteins encoded by the $9DEG^{NTxTP}$. As shown in **Table 2**, three proteins are related to the production of exosomes, membrane trafficking or exocytosis (AHCY, TC2N and TRIM67), two are related to specific aspects of cell metabolism (LYPLA1 and AHCY), two others are related to the ubiquitination process (USP19 and TRIM67) and two are related to basic cell signaling (SPRY3 and TC2N).

To improve the knowledge about the biological role of these proteins in cancer we search the literature (**Table 2**). We firstly search for the biological role of each DEG-encoded protein in general pathophysiology (not only in cancer). To this, we search for papers including in the title or the abstract the DEG symbols according to the *Hugo Gene Nomenclature* or the full name of each DEG. Then, we filtered our search only for cancer-related articles, using the gene symbol AND 'cancer' or the full name of the gene AND 'cancer'. Except to AHCY, we found an average of 1 or 2 studies to each DEG, none in PAAD. These manuscripts are also limited concerning the mechanisms underlying the role of DEGs in the models studied. Additional integration of data from literature about the role of $9DEG^{NTxTP}$ in cancer is presented in the discussion section.

Table 2 – Biological functions of 9DEG^{NTxTP} inferred from KEGG functional annotation and literature review.

Gene	KEGG Functional Annotation	#Refs/ #Refs in Cancer*	Cancer Types*	Role in Cancer*	Ref*
AHCY	**a.** Exosomal proteins of fluids (saliva and urine) **b.** Protein phosphatase-1 interacting proteins **c.** Cysteine and methionine metabolism	16/10	Hypopharyngeal, ovarian, breast, liver, colorectal and colon cancer; neuroblastoma	p53 negative regulation; cellular methylation, cell cycle arrest, DNA damage repair, anti-proliferative effect; target of MYC (biomarker of MYC amplification)	14–19
FAM45A	Not annotated	2/1	Osteosarcoma	p53 negative regulation	20
LYPLA1	**a.** Glycerophospholipid metabolism **b.** Choline metabolism in cancer	11/2	Breast and endometrial cancer	Phospholipid metabolism; Ras positive regulation	21,22
MAST4	**a.** Serine/threonine kinases	1/1	Breast cancer	Breast cancer diagnostic biomarker in urine	23
SPRY3	**a.** Signaling proteins (kinase)	7/4	Acute Myeloid Leukemia; Acute Lymphoblastic Leukemia; Hepatocellular carcinoma (HCC)	Ras negative regulation, Resistance to Quizartinib; not altered in HCC in relation to normal tissue	24–27
TC2N	**a.** Exosomal proteins	5/2	Lung cancer; breast cancer	Oncogene in lung cancer (p53 suppressor); tumor suppressor in breast cancer (PI3K-AKT suppressor)	28
TGM5	**a.** Protein-glutamine ɣ-glutamyltransferase	3/2	Lung cancer; Head and neck	No significant association between TGM5 polymorphism and lung cancer	29
TRIM67	**a.** Ca^{2+}-dependent exocytosis **b.** Ubiquitin system **c.** Membrane trafficking	3/1	Pulmonary adenocarcinoma	Ras negative regulation	30–32
USP19	**a.** Ubiquitin-specific proteases (UBPs) **b.** Cysteine Peptidases **c.** Ubiquitinyl hydrolase 1	11/2	Ewing Sarcoma, prostate cancer	Regulates oncoprotein EWS-FLI1, anti-proliferative effect, DNA damage repair regulation	33–36

** Data from PUBMED search, in 06/01/2019*

The third approach to better understand the role of DEGs in cancer was to search for molecular connectors through building a network of protein interactions to these genes using the STRING platform. The **Figure 4** shows the protein network of each 9DEG-encoded protein. Four proteins (SPRY3, MAST4, TC2N and TGM5) no network was build, suggesting that no direct interactor is known to these proteins.

USP19 is part of a network related to E3 ubiquitin-protein ligase through direct interacting with SIAH1 and SIAH2, E3 ligase involved in ubiquitination and proteasome-mediated degradation of specific proteins[33-36].

TRIM67 interacts with DCC (Deleted in Colorectal Carcinoma). DCC is a transmembrane protein that also interacts, through the cytosolic tail, with Src and FAK proteins to control cell adhesion. It also acts as a tumor suppressor, being frequently mutated or downregulated in colorectal cancer and esophageal carcinoma[30-32].

AHCY interacts with XIAP, an inhibitor of apoptosis that suppress the activation of various caspases. AHCY also interacts with Heat Shock Protein B1 (HSPB1), which act as a molecular chaperone that plays a role in stress resistance and actin remodeling. Finally, AHCY also interacts with metabolism-related proteins, ALDH4A1 and LDHB. ALDHA1 participates on the prolin metabolism, while LDHB in involved in the interconversion of pyruvate and lactate in a post-glycolysis process[14-19].

Finally, LYPLA1 interacts with CSPO, a cysteine proteinase and a member of the papain superfamily involved in protein degradation and turnover. LYPLA1 also interacts with CDA, an enzyme that catalyzes the irreversible hydrolytic deamination of cytidine and deoxycytidine to uridine and deoxyuridine, respectively, being responsible for maintaining the cellular pyrimidine pool. Another interactor of LYPLA1 is OXCT1, a mitochondrial matrix enzyme central to the extrahepatic ketone body

catabolism by catalyzing the reversible transfer of coenzyme A from succinyl-CoA to acetoacetate. LYPLA1 also interacts with TCC38, whose function is not known[21,22].

FAM45 is connected to COMM proteins, which are proteins involved in neural development during embryogenesis. It controls key aspects such as axon growth, intracellular trafficking in neurons, cell migration, among others[20].

Importantly, we also evaluated to possible interactions among the 9DEGs but no interaction was found (data not shown).

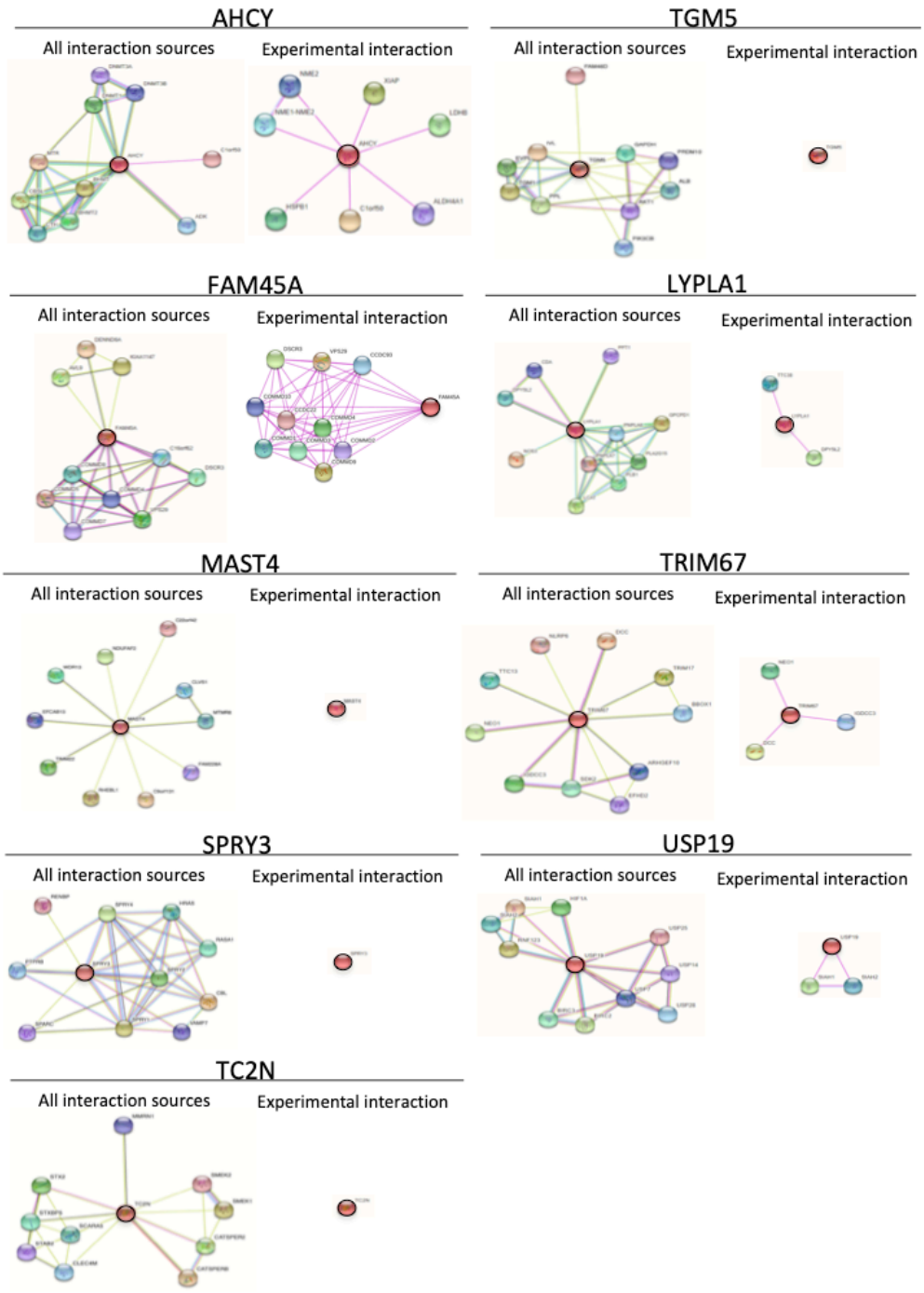


Figure 4 – DEGs encoded protein networks. These STRING networks present the connections between each one of the 9DEGs. Lines in pink refer to experimentally determined interactions; lines in yellow refer to textmining interactions; lines in green refer to gene neighborhood interactions; lines in light blue refer to curated database interactions; lines in dark blue refer to gene co-occurrence interactions.

Subcellular Distribution

Finally, we assessed the subcellular distribution of DEG-coding proteins through assessing the THPA subcellular distribution tool. We found data for 6 from 9

proteins in the THPA, as shown in **Supplementary Figure 3** and summarized in **Table 3**. From these, 3 proteins are mainly cytosolic (AHCY, FAM45A and SPRY3), but 2 also showed nuclear dots (AHCY and FAM45A). FAM45 shows a clear cytosolic punctiform pattern. On the other hand, MAST4, TC2N and LYPLA1 are located mainly in the nucleus; LYPLA1 show a diffuse patter of nuclear distribution, while MAST4 and TC2N showed a punctiform pattern. MAST4 also appeared in cytosolic dots. We did not find the subcellular distribution of USP19, TRIM67 and TGM5 in the THPA, so that we search for this in the literature. We found that TRIM67 and USP19 appeared diffusely in the cytosol of cerebellar neurons and Hela cells. Finally, the subcellular distribution of TGM5 is not known.

Table 3 – Subcellular distribution of DEGs.

Protein	Cell type	Nucleus (pattern)	Cytosol
AHCY	CaCo2 (colorectal cancer); HepG2 (hepatocellular carcinoma)	Low (foci)	High (dots)
FAM45A	A-431 (skin epidermoid carcinoma); U-2OS (Osteosarcoma); U251 (glioma)	Low (foci)	High (dots)
SPRY3	A-431 (skin epidermoid carcinoma); U-2OS (Osteosarcoma); U251 (glioma)	Absent	High (diffuse)
MAST4	A-431 (skin epidermoid carcinoma); U-2OS (Osteosarcome); U251 (glioma)	High (diffuse + dots)	Medium (dots)
TC2N	A-431 (skin epidermoid carcinoma); U251 (glioma)	High (diffuse + dots)	Low (dots)
LYPLA1	A-431 (skin epidermoid carcinoma); SH-SY5Y (neuroblastoma)	High (diffuse)	High (diffuse)
USP19	Hela	Absent	High (diffuse)
TRIM67	Neurons	Absent	High (diffuse + dots)
TGM5	-	Not available	Not available

* Data interpretation based on immunocytochemistry images from the THPA (for AHCY, FAM45A, SPRY3, MAST4, TC2N and LYPLA1) or from literature[31,37].

KRAS^{wt} *versus KRAS*^{mut} Differential Expression Analysis

We next tested for the differential expression of 9DEG^{NTxTP} in tumor samples considering the mutation status of *KRAS*. Initially, we check for the differential expression between *KRAS* wild type (*KRAS*^{wt}) and *KRAS* mutated (*KRAS*^{mut}) tumor samples. From the total of 29,959 genes, expression levels of 917 genes significantly differed between the conditions.

Considering the 917 DEGs in the *KRAS*^{wt} *versus KRAS*^{mut} analysis, 107 genes showed a difference of at least three times in the LogFC (cut of |±3|) (**Figure 5 -** in green are the 4 upregulated genes [LogFC ≥3], and in red are the 103 downregulated genes [LogFC ≤-3]). From these genes, 103 were downregulated in *KRAS*^{mut} in comparison to *KRAS*^{wt} (**Supplementary Table 2**).

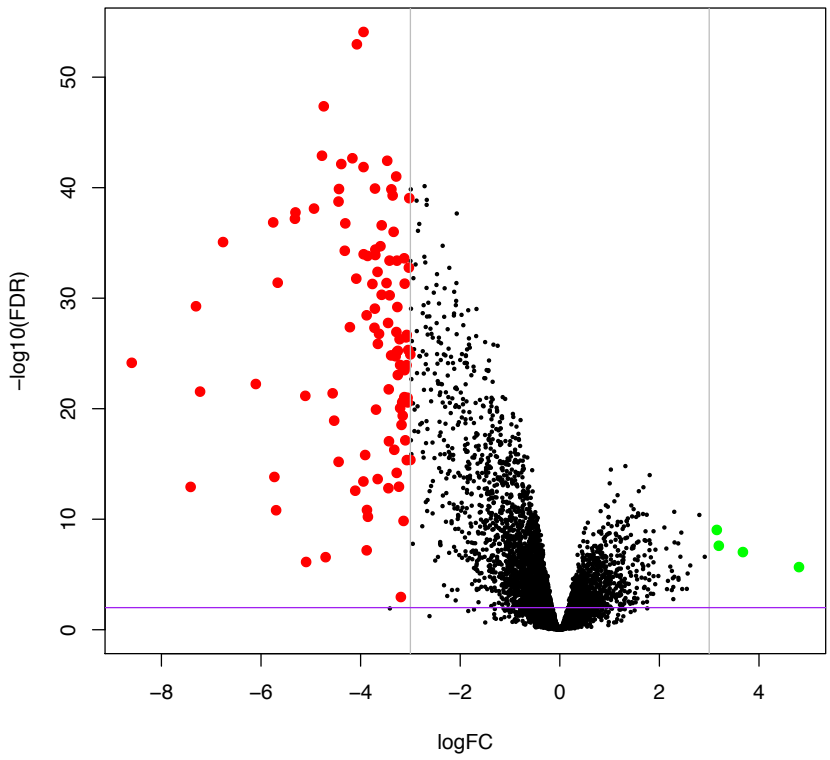


Figure 5 – *KRAS*^{wt} *versus KRAS*^{mut} Differential Expression Analysis. This volcano plot shows the DEG distribution. In red are the decreased DEGs. In green are the increased DEGs. The line in purple refers to the adjusted p-value of 0,01. The lines in grey refer to the logFC cut of |±3|.

Comparing the list of DEGs from NT *versus* TP and *KRAS*^{wt} *versus KRAS*^{mut} (in general, not only considering the cut of |±3|) we found 35 DEGs in common, which represents 3% (**Figure 6; Supplementary Table 3)**.

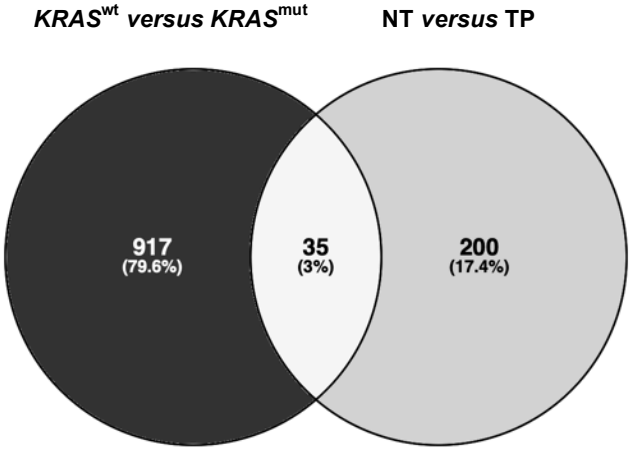


Figure 6 – Venn Diagram to differential expression analysis. This Venn Diagram shows the DEGs specific *KRAS*^{wt} *versus KRAS*^{mut} (black) and NT *versus* TP DEAs (gray). In the white area is shown the number of differentially expressed genes concomitant in both analyses. The percentage of genes inside each condition represent the amount related to the global gene number.

Considering the 28 DEG that differ at least 3 times from NT *versus* TP (see Table S1), 17 genes were also differentially expressed in *KRAS*^{wt} *versus KRAS*^{mut} analysis (data not shown). Note, however, that none of these 17 genes differed more than 3 times between the groups. Actually, from the 9DEG^{NTxTP} whose levels influenced in the survival of PAAD patients (see Table 1), 6 showed significant differential expression between *KRAS*^{wt} *versus KRAS*^{mut} (TGM5, TC2N, MAST4, USP19, LYPLA1 and SPRY3), while the levels of the other 3 DEGs^{NTxTP} (AHCY, FAM45A and TRIM67) did not differ between *KRAS*^{wt} *versus KRAS*^{mut} patients.

Mutations in DEGs

Finally, we tested for genetic mutations in 9DEG^{NTxTP} through assessing the Cosmic database. As shown in **Table 4**, we found a low frequency of mutations in

these genes, from 0 (to FAM45A and LYPLA1) to 1.22 percent (MAST4). In MAST4, which was the most mutated DEG, 4 mutations were detected in two samples each one. Interestingly, according to the UNIPROT database, all these mutations occur in the kinase domain of the protein (**Supplementary Table 4**).

Altogether, with data from mutations analysis, we can infer that DEGs are differentially expressed in tumor cells in comparison to normal cells, but that these differences are not due to genetic alterations.

Table 4 – Genetic mutations of 9DEG^{NTxTP}

Gene	Samples Tested	Mutated Samples	% Mutation
AHCY	1820	3	0.16
FAM45A	1820	0	0
LYPLA1	1820	0	0
MAST4	1798	22	1.22
SPRY3	1798	0	0
TC2N	1820	1	0.05
TGM5	1820	2	0.11
TRIM67	1820	3	0.16
USP19	1823	3	0.16

Genetic mutations of 9DEG^{NTxTP} from Cosmic database, considering the samples tested and samples mutated.

Discussion

Here, we search for new potential prognostic factors for PAAD. We found 235 DEGs in tumor samples in comparison to normal tissue. From these, 28 showed expression levels that differ at least 3 times from normal tissue, being 25 upregulated and 3 downregulated. From these 28 genes, 9 influenced the patient's survival, all of them being upregulated. We also assessed the differential expression of these genes considering the mutational status of *KRAS*. We found 952 DEGs between *KRAS*^{wt} and *KRAS*^{mut} samples. Thirty-five genes were also present in the list of DEG^{NTxTP}. From the 952 DEGs, 107 showed expression levels that differ at least 3 times between *KRAS*^{wt} and *KRAS*^{mut} samples, and 40 of them were prognostic factors

according to the THPA. Seventeen DEGs appeared as prognostic factors in both NT *versus* TP and *KRAS*^{wt} *versus KRAS*^{mut} analysis. Complementarily, 23 DEGs were prognostic factors only for *KRAS*^{wt} *versus KRAS*^{mut} and 11 were exclusive for NT *versus* TP analysis, independent of *KRAS* mutation.

All the 9DEG^{NTxTP} were upregulated in tumor tissue in comparison to normal pancreas. From these genes, high expression of MAST4, SPRY3, USP19 and TRIM67 were associated with favorable prognosis, while high levels of the 5 others (AHCY, FAM45A, LYPLA1, TC2N and TGM5) were associated with poorer prognosis. The biological functions of 9DEG^{NTxTP} are poorly or nothing described in cancer. To shed some light on the biological function of these genes in cancer, we: (1) look for molecular connections using the STRING browser; (2) assessed the KEGG functional annotation; (3) checked for the subcellular distribution; (4) reviewed the literature. Through this we found that proteins encoded by the 9DEG^{NTxTP} modulate key proteins related to the carcinogenic process of PAAD (**Figure 7**).

The most important oncogenic driver pathway in PAAD is *KRAS*. Among the 9DEG^{NTxTP}, **TRIM67** has been described as a negative regulator of Ras in non-small cell lung cancer. In this cancer, TRIM67 is downregulated, which may contribute to the overactivation of Ras and cancer aggressiveness[30]. Corroborating the tumor suppressor activity of TRIM67, *Do et al.* found autoantibodies targeting TRIM67 at high concentration in the serum and the cerebrospinal fluid of patients with lung adenocarcinoma[32]. Here, we found that high levels of TRIM67 were associated with favorable prognosis in PAAD, which could be related to the suppression of Ras pathway. In agreement to this, the ectopic expression of TRIM67 increased the differentiation and suppressed the growth of neuroblastoma cell lines through the negative regulation of Ras[31].

Sprouty proteins are encoded by SPRY genes, and negatively regulates Ras pathway leading to cell growth inhibition[38]. Here, we found an increase of SPRY3 levels in PAAD in comparison to normal tissue. In leukemia, loss of function of **SPRY3** mediates the resistance of cancer cells to quizartinib, an inhibitor of FLT3 tyrosine kinase receptor that leads to the inhibition of the Ras-Raf-Mek pathway. As a consequence of SPRY3 loss, Ras pathway may be upregulated, thus causing the resistance[24]. As observed for TRIM67 in lung cancer, SPRY2 regulates oncogenic *KRAS* during lung tumorigenesis in a genetic mouse model harboring a germline oncogenic *KRAS*^{G12D} mutation. Similar to our data, authors found an upregulation of SPRY2 during the carcinogenic process and hypothesized that the tumor suppressive activity of this gene may be triggered to limit tumor number and overall tumor burden[39]. Indeed, here we found that high levels of SPRY3 were associated with favorable prognosis, which could also be related to the inhibition of Ras signaling.

The third 9DEG^{NTxTP} involved in the control of Ras pathway is **LYPLA1,** which encodes the Lysophospholipase A1. This enzyme plays a critical role in both depalmitoylating and lysophospholipase activity. Cycles of depalmitoylation and repalmitoylation control the steady-state localization and function of several proteins, including Ras. A small molecule inhibitor of lysophospholipase A1 directly affects the pro-tumor role of Ras protein, suggesting that LYPLA1 favors the tumorigenesis and tumor aggressiveness. In agreement to this, the suppression of LYPLA1 through iRNA in lung cancer cells reduced cell proliferation and invasion[40]. Our data corroborate this hypothesis, with higher levels of LYPLA1 associated to poorer the prognosis. Furthermore, the potential of lipid metabolism in the early detection of PAAD was recently shown by Fahrmann et al. Authors found that

lysophosphatidilcoline, which is degraded by Lysophpospholipase A1, is reduced in early stages of pancreatic cancer in comparison to normal tissue, being a potential marker of early carcinogenesis. This could be explained by the upregulation of LYPLA1, as found here and raised by authors[41]. LYPLA1 expression is also increased in endometrial[22] and breast cancer cells[21]. Elevated levels of metabolites derived from Lysophospholipase A1 activity are strongly correlated with the malignancy[21], in agreement with our data.

The second most mutated gene in PAAD is *TP53*. Three from the $9DEG^{NTxTP}$ we found are modulators of *TP53* (AHCY, FAM45A and TC2N). High levels of these genes were associated with poor prognosis, which could be related to the inhibition of *TP53*. The protein encoded by these three genes showed nuclear foci at different extent (TC2N > AHCY > FAM45A). Adenosylhomocysteinase (**AHCY**) is a conserved and ubiquitously expressed enzyme that catalyzes the hydrolysis of *S*-adenosyl-L-homocysteine into adenosine and homocysteine. Here, we found an upregulation of AHCY in PAAD, and the higher the levels, the poorer the survival rate. AHCY gene has been described as upregulated also in hypopharyngeal, ovarian, colon and colorectal cancers[15,16,18]. In these tumors, the reduction of AHCY activity causes positive modulation of p53 activity, with activation of the DNA damage response, leading to cell cycle arrest, decreased proliferation rate and DNA damage[14,17,42]. Thus, it is plausible to infer that high levels of AHCY, as observed here, may lead to increased cell proliferation and suppression of cell cycle arrest, favoring tumor growth.

Another DEG^{NTxTP} that modulate the p53 signaling is **TC2N**, which is present almost exclusively in the nucleus both in a diffuse pattern of distribution and as foci. Recently, TC2N was described as a new oncogene that accelerates lung cancer

progression and is associated with poor prognosis. The mechanism underlying this effect involves the suppression of p53 pathway, thus leading to increased cell proliferation and resistance to apoptosis[28]. Here, we also found increased levels of TC2N in PAAD samples in comparison to normal pancreas, and the higher the levels of TC2N expression the poorer the prognosis of patients. Furthermore, in lung cancer the downregulation of TC2N with subsequent activation of p53 pathway is involved in the apoptosis triggered by doxorubicin, suggesting that high levels of TC2N may also be related to resistance to therapy. Finally, the mRNA or the protein encoded by TC2N is present in extracellular vesicles (mainly exosomes) secreted by mammary gland cells, colorectal cancer cells, urothelial cells and mesenchymal stem cells[43]. In milk, these exosomes are able to modulate the immune system[44,45], while in colorectal cancer these microvesicles trigger the proliferation of endothelial cells[46]. In PAAD, exosomes play a role in the carcinogenesis, metastasis, resistance to therapy, modulation of immune system, among other functions[47,48]. One should explore whether TC2N is involved in these processes since the detection of circulating tumor markers, including extracellular vesicles, has been proposed as alternatives to early detection through liquid biopsy.

The third DEG^{NTxTP} that modulates the p53 pathway is **FAM45A**. In osteosarcoma, authors found a fusion transcript of the first exon of *TP53* to the fifth exon of FAM45A, leading to the inactivation of *TP53*[20]. Thus, the upregulation of FAM45A could increase the availability of its mRNA, thus facilitating the defective splicing of *TP53*. Here we found that high levels of FAM45A were associated with a poorer prognosis, which could be related to a suppression of p53 signaling. Recently, Mello et al described a new pathway involved in the tumor suppressive role of p53 in PAAD, involving the negative regulation of Yap oncoprotein by the p53 target Ptpn14.

Yap is a transcription factor which stimulates cell proliferation, invasion, stemness and migration. Among the molecular targets of Yap is FAM45A[49]. Indeed, here we found that high levels of FAM45A (which suggest low levels of *TP53* and, as a consequence, high levels of Yap activation) are associated with poor prognosis.

Another DEG found here that modulates cell division, DNA damage response, cell death and proliferation is **USP19**. The protein encoded by this gene play a critical role in the DNA damage response (DDR) and the maintenance of genome stability. After irradiation, USP19 translocates to the nucleus and modulates the activity of HDAC1/2, which is crucial to the DDR. USP19 is commonly lower or deleted in several cancer types such as kidney renal clear cell carcinoma, stomach adenocarcinoma, cervical squamous cell carcinoma, esophageal carcinoma and brain lower grade glioma[36]. It is positively correlated with the tumor suppressor gene BAP1 in invasive breast carcinoma, uveal melanoma, and colon adenocarcinoma[35]. If, in one hand, USP19 acts as an anti-tumor gene that contributes to the maintenance of genome integrity, in cancer cells this gene plays an opposite role. USP19 is positively involved in cell proliferation and resistance to apoptosis, through mechanisms involving the degradation of inhibitor of cyclin-CDK complexes, like p27 (which controls the G1/S progression)[34], the suppression of oncoproteins, like EWS-FL1 in Ewing Sarcoma[33] and the stabilization of inhibitors of apoptosis[50]. USP19 also controls autophagy through the deubiquitination of beclin-1, thus avoiding its degradation and increasing basal autophagic flux (which may be cytoprotective in PAAD), and negatively regulates type I IFN signaling pathway[51]. Here, we found that high expression of USP19 was associated with favorable prognosis. Thus, at least in PAAD, the anti-tumoral role of USP19 may be dominant in relation to the pro-tumoral role of this protein.

Increased levels of the protein encoded by **MAST4** have been proposed as a diagnostic marker in breast cancer[23]. Here, we also found higher levels of MAST4 expression in PAAD in comparison to normal tissue. Thus, it is plausible to assume that increasing levels of this protein may not be exclusive to a specific tumor type. High levels of MAST4 were associated with a favorable prognosis in PAAD, while its upregulation was associated with pre-invasive breast cancer[23]. MAST4 is a member of the microtubule-associated kinase family of proteins[52] ubiquitously expressed in human tissues including pancreas, in which has low levels (data from THPA). This protein is present in the cytosol as bright dots and in the nucleus in a diffuse pattern and also as dots. MAST4 interacts with SMAD1[53], which is involved in the progression of pancreatic cancer through promoting the downregulation of E-cadherin and the upregulation of matrix metalloproteinase-2, leading to increased invasiveness and cancer progression[54]. In this context, since high levels of MAST4 were associated with a favorable prognosis, it is plausible to assume that MAST4 may be a negative regulator of SMAD1 signaling.

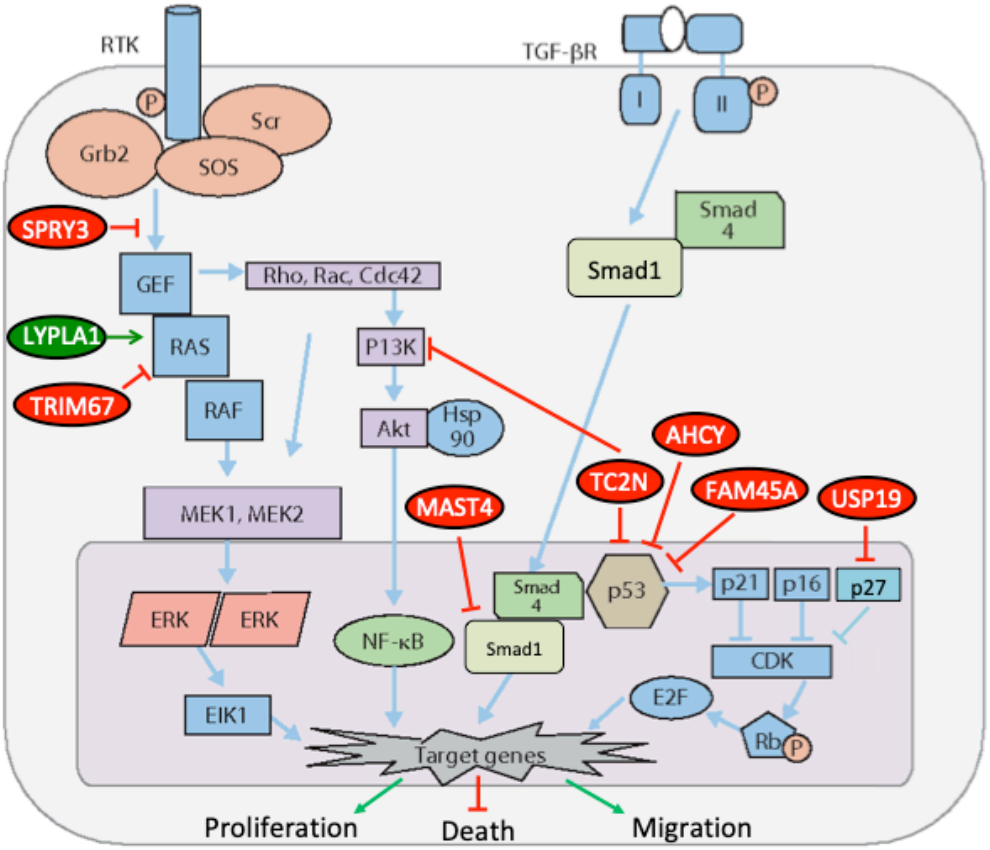


Figure 7 - Cell signalling pathway for proliferation, death and cell migration. The proteins of this study's interest are indicated in red when related to inhibition, and in green when related to activation. Adapted from MATTHAIOS et al., 2011[55].

Conclusions

The genomic *era* provides a large amount of data relating to the molecular biology of cancer with clinical outcome. Here, we explore the TCGA databank and describe 9 new genes showing potential as prognostic factors in PAAD, and also explore the biological role and its relevance in the biology of these tumors. We shed some light in a dark area of clinical cancer, as observed by the scarcity of literature about prognostic markers in PAAD. Predicting the expected outcome of patients diagnosed with this cancer type is a critical step to define the treatment.

Acknowledgments

We thank to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and CNPq for the grant. Funding agencies had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

Authors declare no conflict of interest.

Disclaimer

We are aware that our study has limitations, such as the small number of normal tissue (NT) to compare to tumor tissue (TP), which restrains the comparison analysis of NT *versus* TP pancreatic samples.

REFERÊNCIAS

1. Kunovsky L, Tesarikova P, Kala Z, et al. The Use of Biomarkers in Early Diagnostics of Pancreatic Cancer. *Can J Gastroenterol Hepatol*. 2018;2018:1-10. doi:10.1155/2018/5389820
2. Shen Q, Yu M, Jia J-K, Li W-X, Tian Y-W, Xue H-Z. Possible Molecular Markers for the Diagnosis of Pancreatic Ductal Adenocarcinoma. *Med Sci Monit*. 2018;24:2368-2376. doi:10.12659/msm.906313
3. Raphael BJ, Hruban RH, Aguirre AJ, et al. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell*. 2017;32(2):185-203.e13. doi:10.1016/j.ccell.2017.07.007
4. American Cancer Society. Cancer Facts and Figures 2017. American Cancer Society. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2017.html. Published 2017. Accessed November 10, 2018.
5. Bye A, Wesseltoft-Rao N, Iversen PO, et al. Alterations in inflammatory biomarkers and energy intake in cancer cachexia: a prospective study in patients with inoperable pancreatic cancer. *Med Oncol*. 2016;33(6):54. doi:10.1007/s12032-016-0768-2
6. Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks : an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016;44(8):e71-e71. doi:10.1093/nar/gkv1507
7. The Cancer Genome Atlas - TCGA. The Cancer Genome Atlas. TCGA. https://cancergenome.nih.gov. Published 2018. Accessed March 23, 2018.
8. National Cancer Institute. Genomic Data Commons Data Portal. Harmonized Cancer Datasets. https://portal.gdc.cancer.gov. Published 2019.

9. RStudio Team. RStudio: Integrated Development Environment for R. 2016. http://www.rstudio.com/.

10. Uhlen M, Fagerberg L, Hallstrom BM, et al. Tissue-based map of the human proteome. *Science (80-)*. 2015;347(6220):1260419-1260419. doi:10.1126/science.1260419

11. Snel B. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res*. 2000;28(18):3442-3444. doi:10.1093/nar/28.18.3442

12. Brohée S, van Helden J. Evaluation of clustering algorithms for protein-protein interaction networks. *BMC Bioinformatics*. 2006;7(1):488. doi:10.1186/1471-2105-7-488

13. Oliveros JC. Venny: An interactive tool for comparing lists with Venn's diagrams. BioinfoGP Service Centro Nacional de Biotecnología. http://bioinfogp.cnb.csic.es/tools/venny/index.html. Published 2015.

14. Belužić L, Grbeša I, Belužić R, et al. Knock-down of AHCY and depletion of adenosine induces DNA damage and cell cycle arrest. *Sci Rep*. 2018;8(1):14012. doi:10.1038/s41598-018-32356-8

15. Fan J, Yan D, Teng M, et al. Digital Transcript Profile Analysis with aRNA-LongSAGE Validates FERMT1 As a Potential Novel Prognostic Marker for Colon Cancer. *Clin Cancer Res*. 2011;17(9):2908-2918. doi:10.1158/1078-0432.CCR-10-2552

16. Kim H-J, Kang HJ, Lee H, et al. Identification of S100A8 and S100A9 as Serological Markers for Colorectal Cancer. *J Proteome Res*. 2009;8(3):1368-1379. doi:10.1021/pr8007573

17. Park SJ, Kong HK, Kim YS, Lee YS, Park JH. Inhibition of S-

adenosylhomocysteine hydrolase decreases cell mobility and cell proliferation through cell cycle arrest. *Am J Cancer Res*. 2015;5(7):2127-2138. http://www.ncbi.nlm.nih.gov/pubmed/26328244.

18. Wu P, Tang Y, Fang X, et al. Metformin Suppresses Hypopharyngeal Cancer Growth by Epigenetically Silencing Long Non-coding RNA SNHG7 in FaDu Cells. *Front Pharmacol*. 2019;10(February):1-11. doi:10.3389/fphar.2019.00143
19. Zhong X, Liu Y, Liu H, Zhang Y, Wang L, Zhang H. Identification of Potential Prognostic Genes for Neuroblastoma. *Front Genet*. 2018;9(November):1-10. doi:10.3389/fgene.2018.00589
20. Barøy T, Chilamakuri CSR, Lorenz S, et al. Genome Analysis of Osteosarcoma Progression Samples Identifies FGFR1 Overexpression as a Potential Treatment Target and CHM as a Candidate Tumor Suppressor Gene. Heymann D, ed. *PLoS One*. 2016;11(9):e0163859. doi:10.1371/journal.pone.0163859
21. Glunde K, Jie C, Bhujwalla ZM. Molecular Causes of the Aberrant Choline Phospholipid Metabolism in Breast Cancer. *Cancer Res*. 2004;64(12):4270-4276. doi:10.1158/0008-5472.CAN-03-3829
22. Trousil S, Lee P, Pinato DJ, et al. Alterations of Choline Phospholipid Metabolism in Endometrial Cancer Are Caused by Choline Kinase Alpha Overexpression and a Hyperactivated Deacylation Pathway. *Cancer Res*. 2014;74(23):6867-6877. doi:10.1158/0008-5472.CAN-13-2409
23. Beretov J, Wasinger VC, Millar EKA, Schwartz P, Graham PH, Li Y. Proteomic Analysis of Urine to Identify Breast Cancer Biomarker Candidates Using a Label-Free LC-MS/MS Approach. Aboussekhra A, ed. *PLoS One*. 2015;10(11):e0141876. doi:10.1371/journal.pone.0141876
24. Hou P, Wu C, Wang Y, et al. A Genome-Wide CRISPR Screen Identifies Genes

Critical for Resistance to FLT3 Inhibitor AC220. *Cancer Res*. 2017;77(16):4402-4413. doi:10.1158/0008-5472.CAN-16-1627

25. Atak ZK, Keersmaecker K De, Gianfelici V, et al. High Accuracy Mutation Detection in Leukemia on a Selected Panel of Cancer Genes. Hoheisel JD, ed. *PLoS One*. 2012;7(6):e38463. doi:10.1371/journal.pone.0038463
26. Ning Z, McLellan AS, Ball M, et al. Regulation of SPRY3 by X chromosome and PAR2-linked promoters in an autism susceptibility region. *Hum Mol Genet*. 2015;24(18):5126-5141. doi:10.1093/hmg/ddv231
27. Sirivatanauksorn Y, Sirivatanauksorn V, Srisawat C, Khongmanee A, Tongkham C. Differential expression of sprouty genes in hepatocellular carcinoma. *J Surg Oncol*. 2012;105(3):273-276. doi:10.1002/jso.22095
28. Hao X, Han F, Zhang N, et al. TC2N, a novel oncogene, accelerates tumor progression by suppressing p53 signaling pathway in lung cancer. *Cell Death Differ*. 2018;26(7):1235-1250. doi:10.1038/s41418-018-0202-8
29. Zhang MMY, Zhang A, Shi MMH, Kong MMX. Association between TGM5, PPAP2B and PSMA4 polymorphisms and NSCLC in never-smoking Chinese population. *J Cancer Res Ther*. 2013;9(4):660. doi:10.4103/0973-1482.126473
30. Zhan W, Han T, Zhang C, et al. TRIM59 Promotes the Proliferation and Migration of Non-Small Cell Lung Cancer Cells by Upregulating Cell Cycle Related Proteins. Ahmad A, ed. *PLoS One*. 2015;10(11):e0142596. doi:10.1371/journal.pone.0142596
31. Yaguchi H, Okumura F, Takahashi H, et al. TRIM67 Protein Negatively Regulates Ras Activity through Degradation of 80K-H and Induces Neuritogenesis. *J Biol Chem*. 2012;287(15):12050-12059. doi:10.1074/jbc.M111.307678

32. Do LD, Gupton SL, Tanji K, et al. TRIM9 and TRIM67 Are New Targets in Paraneoplastic Cerebellar Degeneration. *The Cerebellum*. 2019;18(2):245-254. doi:10.1007/s12311-018-0987-5

33. Gierisch ME, Pedot G, Walser F, et al. USP19 deubiquitinates EWS-FLI1 to regulate Ewing sarcoma growth. *Sci Rep*. 2019;9(1):951. doi:10.1038/s41598-018-37264-5

34. Lu Y, Adegoke OAJ, Nepveu A, et al. USP19 Deubiquitinating Enzyme Supports Cell Proliferation by Stabilizing KPC1, a Ubiquitin Ligase for p27Kip1. *Mol Cell Biol*. 2009;29(2):547-558. doi:10.1128/MCB.00329-08

35. Shahriyari L, Abdel-Rahman M, Cebulla C. BAP1 expression is prognostic in breast and uveal melanoma but not colon cancer and is highly positively correlated with RBM15B and USP19. Ahmad A, ed. *PLoS One*. 2019;14(2):e0211507. doi:10.1371/journal.pone.0211507

36. Wu M, Tu H, Chang Y, et al. USP19 deubiquitinates HDAC1/2 to regulate DNA damage repair and control chromosomal stability. *Oncotarget*. 2017;8(2):2197-2208. doi:10.18632/oncotarget.11116

37. Lee J-G, Kim W, Gygi S, Ye Y. Characterization of the Deubiquitinating Activity of USP19 and Its Role in Endoplasmic Reticulum-associated Degradation. *J Biol Chem*. 2014;289(6):3510-3517. doi:10.1074/jbc.M113.538934

38. Gross I, Bassit B, Benezra M, Licht JD. Mammalian Sprouty Proteins Inhibit Cell Growth and Differentiation by Preventing Ras Activation. *J Biol Chem*. 2001;276(49):46460-46468. doi:10.1074/jbc.M108234200

39. Shaw AT, Meissner A, Dowdle JA, et al. Sprouty-2 regulates oncogenic K-ras in lung development and tumorigenesis. *Genes Dev*. 2007;21(6):694-707. doi:10.1101/gad.1526207

40. Mohammed A, Zhang C, Zhang S, et al. Inhibition of cell proliferation and migration in non-small cell lung cancer cells through the suppression of LYPLA1. *Oncol Rep*. November 2018:973-980. doi:10.3892/or.2018.6857

41. Fahrmann JF, Bantis LE, Capello M, et al. A Plasma-Derived Protein-Metabolite Multiplexed Panel for Early-Stage Pancreatic Cancer. *JNCI J Natl Cancer Inst*. 2019;111(4):372-379. doi:10.1093/jnci/djy126

42. Leal JF, Ferrer I, Blanco-Aparicio C, et al. S-adenosylhomocysteine hydrolase downregulation contributes to tumorigenesis. *Carcinogenesis*. 2008;29(11):2089-2095. doi:10.1093/carcin/bgn198

43. Kalra H, Simpson RJ, Ji H, et al. Vesiclepedia: A Compendium for Extracellular Vesicles with Continuous Community Annotation. *PLoS Biol*. 2012;10(12):e1001450. doi:10.1371/journal.pbio.1001450

44. van Herwijnen MJC, Zonneveld MI, Goerdayal S, et al. Comprehensive Proteomic Analysis of Human Milk-derived Extracellular Vesicles Unveils a Novel Functional Proteome Distinct from Other Milk Components. *Mol Cell Proteomics*. 2016;15(11):3412-3423. doi:10.1074/mcp.M116.060426

45. Admyre C, Johansson SM, Qazi KR, et al. Exosomes with Immune Modulatory Features Are Present in Human Breast Milk. *J Immunol*. 2014;179(3):1969-1978. doi:10.4049/jimmunol.179.3.1969

46. Hong B, Cho J-H, Kim H, et al. Colorectal cancer cell-derived microvesicles are enriched in cell cycle-related mRNAs that promote proliferation of endothelial cells. *BMC Genomics*. 2009;10(1):556. doi:10.1186/1471-2164-10-556

47. Yan Y, Fu G, Ming L. Role of exosomes in pancreatic cancer (Review). *Oncol Lett*. 2018;15(5):7479-7488. doi:10.3892/ol.2018.8348

48. Armstrong EA, Beal EW, Chakedis J, et al. Exosomes in Pancreatic Cancer:

from Early Detection to Treatment. *J Gastrointest Surg*. 2018;22(4):737-750. doi:10.1007/s11605-018-3693-1

49. Mello SS, Valente LJ, Raj N, et al. A p53 Super-tumor Suppressor Reveals a Tumor Suppressive p53-Ptpn14-Yap Axis in Pancreatic Cancer. *Cancer Cell*. 2017;32(4):460-473.e6. doi:10.1016/j.ccell.2017.09.007

50. Mei Y, Hahn AA, Hu S, Yang X. The USP19 Deubiquitinase Regulates the Stability of c-IAP1 and c-IAP2. *J Biol Chem*. 2011;286(41):35380-35387. doi:10.1074/jbc.M111.282020

51. Jin S, Tian S, Chen Y, et al. USP19 modulates autophagy and antiviral immune responses by deubiquitinating Beclin-1. *EMBO J*. 2016;35(8):866-880. doi:10.15252/embj.201593596

52. Sun L, Gu S, Li X, et al. Identification of a novel human MAST4 gene, a new member of human microtubule associated serine/threonine kinase family. *Mol Biol*. 2006;40(5):724-731. doi:10.1134/S0026893306050062

53. Colland F, Jacq X, Trouplin V, et al. Functional Proteomics Mapping of a Human Signaling Pathway. *Genome Res*. 2004;14(7):1324-1332. doi:10.1101/gr.2334104

54. Gordon KJ, Kirkbride KC, How T, Blobe GC. Bone morphogenetic proteins induce pancreatic cancer cell invasiveness through a Smad1-dependent mechanism that involves matrix metalloproteinase-2. *Carcinogenesis*. 2009;30(2):238-248. doi:10.1093/carcin/bgn274

55. Matthaios D, Zarogoulidis P, Balgouranidou I, Chatzaki E, Kakolyris S. Molecular Pathogenesis of Pancreatic Cancer and Clinical Perspectives. *Oncology*. 2011;81(3-4):259-272. doi:10.1159/000334449

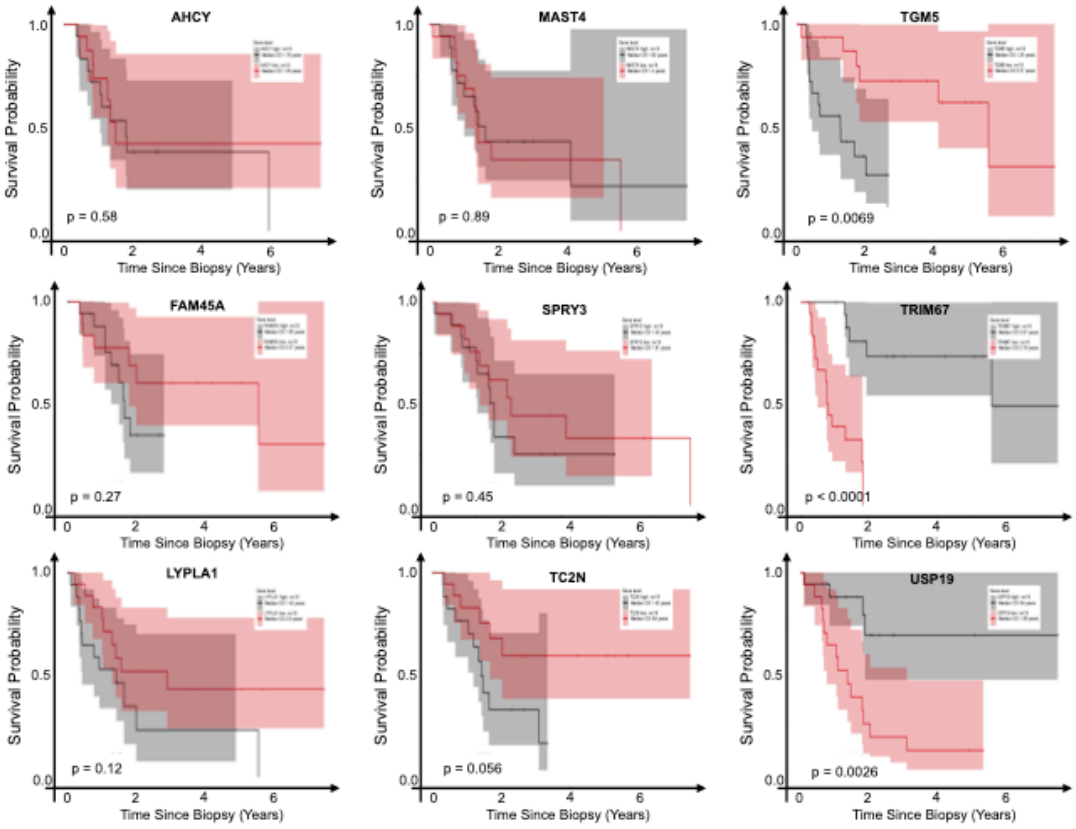
SUPPLEMENTARY MATERIAL

Supplementary Table 1 - List of the 28 DEGs in NT *versus* TP

Gene Name	Gene Code	logFC	p-value
ACSM1	ENSG00000166743	-11.1	0.00018819
MRAS	ENSG00000158186	-9.2	0.000141288
SLCO2A1	ENSG00000174640	-3.0	0.000160998
SNORD115-28	ENSG00000200801	3.0	6.90E-05
HLA-DOB	ENSG00000241106	3.0	2.13E-10
FAM45A	ENSG00000119979	3.0	0.000128812
PXDC1	ENSG00000168994	3.1	5.46E-13
MTMR11	ENSG00000014914	3.2	1.81E-06
AHCY	ENSG00000101444	3.3	2.50E-12
TC2N	ENSG00000165929	3.4	1.91E-05
TGM5	ENSG00000104055	3.4	3.56E-09
ZNF350	ENSG00000256683	3.4	4.15E-08
NDST3	ENSG00000164100	3.4	3.42E-13
PXN	ENSG00000089159	3.5	3.66E-07
RILPL1	ENSG00000188026	3.5	1.46E-05
TRIM67	ENSG00000119283	3.5	0.000100699
PGM5P2	ENSG00000277778	3.6	7.23E-13
KIAA0556	ENSG00000047578	3.9	1.23E-11
TIAF1	ENSG00000221995	4.0	7.43E-08
ANKRD34C-AS1	ENSG00000259234	4.0	2.79E-07
IL17RA	ENSG00000177663	4.0	3.23E-15
MIR3937	ENSG00000263730	4.0	1.32E-07
LYPLA1	ENSG00000120992	4.1	1.24E-10
DEAF1	ENSG00000177030	4.3	1.55E-06
USP19	ENSG00000172046	4.3	3.03E-12
HESX1	ENSG00000163666	4.3	4.86E-14
MAST4	ENSG00000069020	4.5	7.80E-09
SPRY3	ENSG00000168939	10.7	2.39E-42

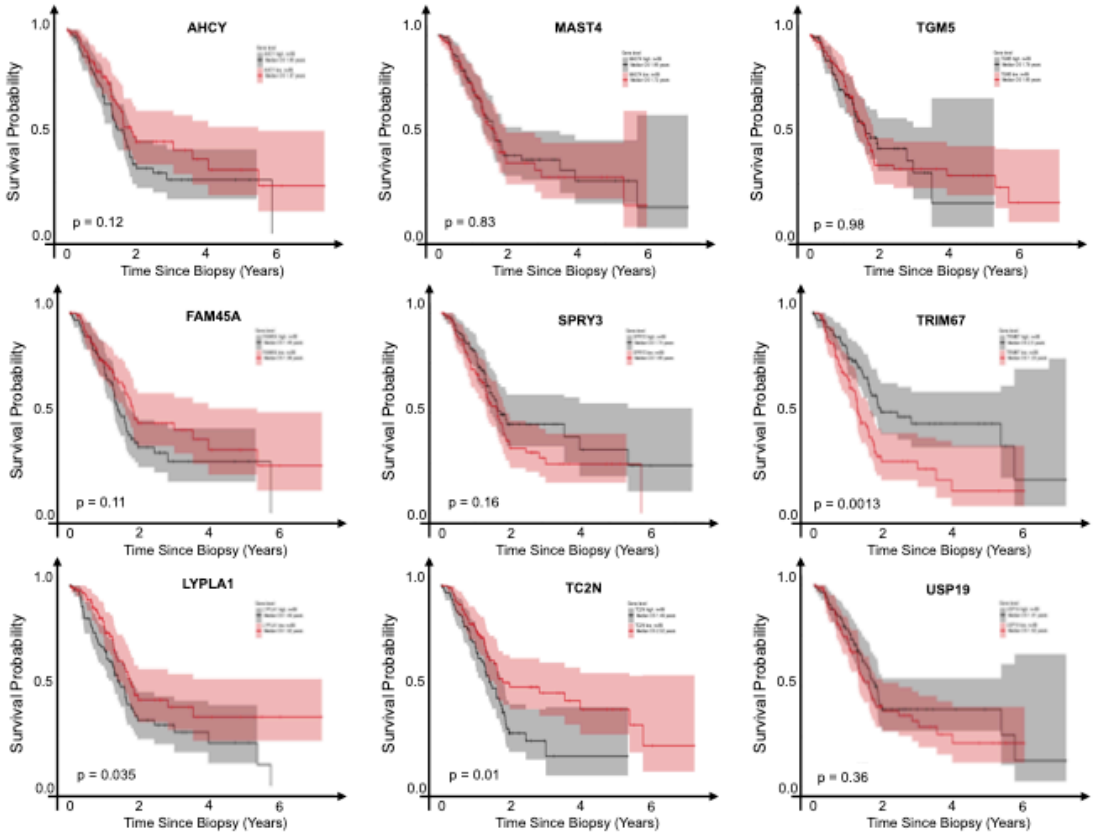
List of the 28 DEGs in NT *versus* TP. In red are the decreased DEGs and in green are the increased DEGs, disposed in ascending order of logFC.

Supplementary Figure 1 - Survival curve of PAAD patients: percentile 10%



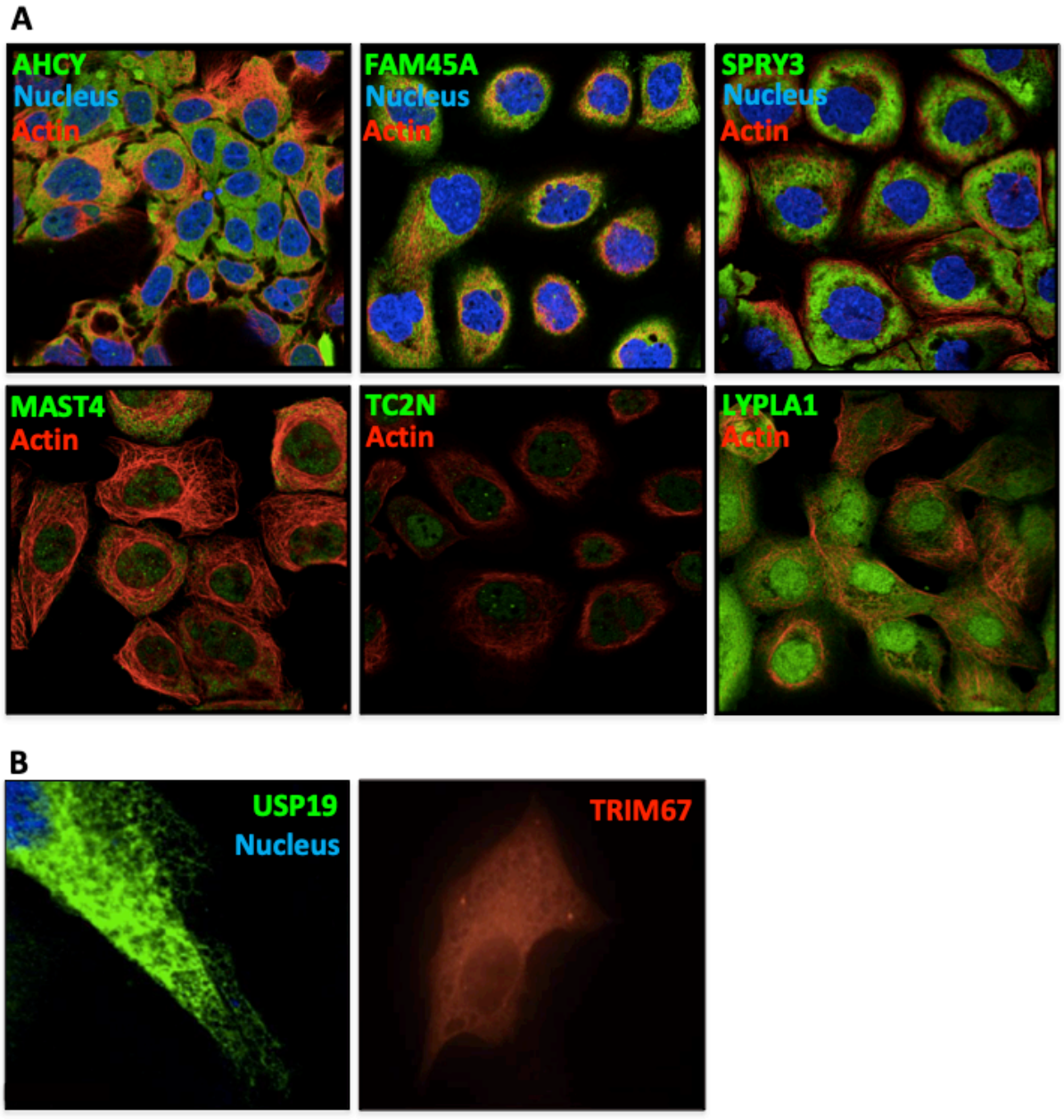
These graphs show the survival probability of PAAD patients (n=36, 18 as high expression group and 18 as low expression group) based on the 9 DEGs selected from NT *versus* TP DEA. The red lines present the low level genes and the black lines present the high gene levels. Each p-value is presented on the inferior left corner of the respective gene graph.

Supplementary Figure 2 - Survival curve of PAAD patients: percentile 50%



These graphs show the survival probability of PAAD patients (n=178, 89 as high expression group and 89 as low expression group) based on the 9 DEGs selected from NT *versus* TP DEA. The red lines present the low level genes and the black lines present the high gene levels. Each p-value is presented on the inferior left corner of the respective gene graph.

Supplementary Figure 3 - Subcellular distribution of DEG-coding proteins



Immunocytochemistry images from (A) THPA (for AHCY (CaCo2), FAM45A (A431), SPRY3 (A431), MAST4 (A431), TC2N (A431) and LYPLA1 (A431)) and (B) literature (USP19 (HeLa) – Lee et al, 2014; and TRIM67 (HeLa) – Yaguchi et al, 2012). (A) In green are the protein of interest, in blue are the nucleus (first row only) and in red are the actin filaments. (B) In green is the marking of protein USP19 and in blue is the nucleus marking. In red is the marking of TRIM67.

Supplementary Table 2 - List of the 103 downregulated in *KRAS*^{wt} *versus KRAS*^{mut}

Gene	logFC	PValue
ZNF286B	-8.6	1.10E-26
CHCHD7	-7.4	6.34E-15
CD207	-7.3	4.25E-32
EIF4G2	-7.2	5.76E-24
PCDHB9	-6.8	2.97E-38
STRC	-6.1	1.12E-24
SNORD115-17	-5.8	3.97E-40
RELT	-5.7	7.17E-16
MBL1P	-5.7	1.01E-12
SLC6A1	-5.7	2.37E-34
ACAT2	-5.3	1.78E-40
LINC00634	-5.3	4.39E-41
ACY1	-5.1	1.44E-23
ANKLE2	-5.1	1.05E-07
MXI1	-4.9	1.88E-41
PITPNM3	-4.8	5.90E-47
LRP5	-4.7	1.50E-51
SHANK2-AS1	-4.7	3.53E-08
BRICD5	-4.6	8.42E-24
TATDN3	-4.5	3.29E-21
RAET1G	-4.4	3.94E-42
CLEC4M	-4.4	2.62E-17
INS	-4.4	1.82E-43
RAD18	-4.4	5.83E-46
GLB1	-4.3	2.05E-37
FOXF1	-4.3	5.06E-40
ZCCHC24	-4.2	4.19E-30
HSDL1	-4.2	1.25E-46
PAPOLG	-4.1	1.42E-14
ZNF778	-4.1	9.99E-35
MIR1224	-4.1	2.44E-57
CHFR	-3.9	1.93E-15
CDC23	-3.9	1.26E-45
NPVF	-3.9	9.30E-59
ARNT	-3.9	4.37E-37
SAT2	-3.9	5.79E-18
MIR7-3HG	-3.9	3.17E-31
KLHL22	-3.9	7.56E-09
FOXD3	-3.9	9.51E-13
EFHD1	-3.9	6.55E-37

Gene	logFC	PValue
RALGAPA1	-3.9	4.24E-12
TGS1	-3.8	3.32E-34
ZNF77	-3.7	4.82E-30
CX3CL1	-3.7	7.19E-32
MIR4635	-3.7	1.51E-43
IFNL2	-3.7	5.25E-37
MIR3065	-3.7	1.56E-37
METRNL	-3.7	3.01E-22
AIRE	-3.7	2.30E-35
SEMA6A	-3.7	1.15E-15
GCSAM	-3.7	1.69E-28
MIR4529	-3.6	1.78E-29
PSORS1C1	-3.6	7.53E-38
SLCO1B7	-3.6	3.52E-33
MIR3621	-3.6	8.22E-40
TMEM234	-3.5	2.59E-34
PFDN2	-3.5	2.52E-46
ZC3H12B	-3.4	1.66E-30
MIGA1	-3.4	8.38E-15
SNORD121A	-3.4	3.54E-24
THUMPD3	-3.4	2.90E-19
ASIC1	-3.4	1.87E-36
SMAD1-AS2	-3.4	3.98E-33
ADAMTS6	-3.4	2.19E-27
MIR4461	-3.4	2.27E-43
PPP3R2	-3.4	8.74E-43
USP9X	-3.3	3.40E-39
ARHGAP26-AS1	-3.3	1.85E-18
TRIP6	-3.3	2.70E-27
ABCC5	-3.3	1.01E-44
PTPRZ1	-3.3	1.22E-29
MAD2L2	-3.3	2.94E-16
PIGQ	-3.3	1.02E-27
NPIPA8	-3.3	1.93E-36
SYCE2	-3.3	5.05E-32
DANCR	-3.3	1.63E-25
FAM234A	-3.3	7.82E-28
CCDC177	-3.2	6.08E-15
CCDC9B	-3.2	6.09E-29
POLR1E	-3.2	1.84E-26
OR2AK2	-3.2	2.17E-22
MIR659	-3.2	0.000333
ERF	-3.2	8.39E-21

Gene	logFC	PValue
MIR941-4	-3.2	6.28E-23
RDH16	-3.2	1.12E-21
MLXIPL	-3.2	4.85E-23
RNF222	-3.1	1.06E-11
PRAMEF26	-3.1	1.13E-36
USP18	-3.1	1.90E-23
FRG2B	-3.1	5.62E-26
E2F4	-3.1	3.05E-34
MIR6775	-3.1	2.35E-19
NAPEPLD	-3.1	3.99E-29
COL27A1	-3.1	1.91E-26
OR5H6	-3.1	2.35E-29
SRBD1	-3.1	1.79E-17
MEPCE	-3.1	2.20E-23
SLC5A2	-3.1	6.16E-23
POMGNT2	-3.0	6.55E-28
PRIMPOL	-3.0	9.03E-36
UBE4B	-3.0	1.63E-42
RORA-AS1	-3.0	1.73E-27
C1QB	-3.0	1.70E-17

List of the 103 DEGs in *KRAS*^{wt} *versus KRAS*^{mut}, disposed in ascending order of logFC.

Supplementary Table 3 - Genes in common from Venn's Diagram

R Studio		THPA
GENE	**logFC**	**p-value**
BRSK1	-1,149783237	0,0001
LYPLA1	-1,962756643	0,00023
FAM129C	-1,131241095	0,00095
TC2N	-1,458669972	0,00099
RTN4RL1	-1,089425289	0,0026
PACSIN3	-1,464019145	0,0039
TIAF1	-1,622374095	0,0061
TRIB1	-1,106239115	0,0068
MRPS26	-1,007781313	0,0096
SPRY3	-2,952146652	0,012
SLC39A13	-1,377069517	0,014
B3GNT7	-1,375560244	0,022
TGM5	-1,12894153	0,022
USP19	-1,846290576	0,023
PXN	-1,772530224	0,027
MTMR11	-1,346346297	0,029
MAST4	-1,666552653	0,05
SLC30A10	-1,169556782	0,051
ASB6	-1,023838897	0,077
HESX1	-1,314625997	0,095
ABHD14A-ACY1	-1,395424012	0,11
TIPIN	-1,13968379	0,13
ZNF350	-1,297115299	0,18
IL17RA	-1,3618477	0,19
ACSM1	1,751039872	0,21
TMEM30A	-1,081198548	0,21
LDLRAD4	-1,008632881	0,23
KIAA0556	-1,317492932	0,29
KCNJ9	-1,00183791	0,34
BDH2	-1,051526256	0,44
ANKRD34C-AS1	-1,700592188	N/A
DEFB127	-1,286382046	N/A
MIR3937	-1,57556068	N/A
MIR758	-1,037388592	N/A
SNORD115-28	-1,512964967	N/A

List of the 35 genes in common from Venn's Diagram (NT *versus* TP and *KRAS*^{wt} *versus KRAS*^{mut}), disposed in ascending order of THPA's p-value. N/A: not applicable.

Supplementary Table 4 - Mutations according to UNIPROT database

Gene	AA Mutation	CDS Mutation	Histology Subtype
AHCY	p.V68I	c.202G>A	Acinar Carcinoma
	p.E59D	c.177G>T	Ductal Carcinoma
	p.N27N	c.81C>T	Ductal Carcinoma
FAM45A	-	-	-
LYPLA1	-	-	-
MAST4	p.S213N	c.638G>A	Ductal Carcinoma
	p.K312T	c.935A>C	NS
	p.L559V	c.1675C>G	Ductal Carcinoma
	p.L559V	c.1675C>G	Ductal Carcinoma
	p.T657I	c.1970C>T	Acinar Carcinoma
	p.D815Y	c.2443G>T	Ductal Carcinoma
	p.L846S	c.2537T>C	Ductal Carcinoma
	p.L846S	c.2537T>C	Ductal Carcinoma
	p.R874W	c.2620C>T	Ductal Carcinoma
	p.R874W	c.2620C>T	Ductal Carcinoma
	p.E952V	c.2855A>T	PanIN
	p.E952V	c.2855A>T	Ductal Carcinoma
	p.H998fs*45	c.2988delC	Acinar Carcinoma
	p.R1246Q	c.3737G>A	Carcinoid-endocrine tumor
	p.R1347W	c.4039C>T	Ductal Carcinoma
	p.A1410T	c.4228G>A	Ampulla of Vater
	p.K1579T	c.4736A>C	Carcinoid-endocrine tumor
	p.S1924N	c.5771G>A	Acinar Carcinoma
	p.E2123Q	c.6367G>C	Ductal Carcinoma
	p.P2167S	c.6499C>T	Ductal Carcinoma
	p.S2256F	c.6767C>T	Ductal Carcinoma
	p.A2382T	c.7144G>A	Ductal Carcinoma
SPRY3	-	-	-
TC2N	p.E98K	c.292 G>A	Acinar Carcinoma
TGM5	p.R595H	c.1784 G>A	Ductal Carcinoma
	p.K465R	c.1394 A>G	Ductal Carcinoma
TRIM67	p.T531T	c.1593 G>A	Ductal Carcinoma
	p.D552D	c.1656 C>T	Ductal Carcinoma
	p.A687V	c.2060C>T	Ampulla of Vater
USP19	p.D1011N	c.3031 G>A	Ampulla of Vater
	p.T983A	c.2947 A>G	Ductal Carcinoma
	p.L764F	c.2292 G>C	Ductal Carcinoma

Genetic mutations in 9DEG^{N1xIP} assessed on Cosmic database. Mutational analyses correspond to different histological subtypes of pancreatic neoplastic conditions. AA Mutation: change in the peptide sequence. CDS Mutation: change in the nucleotide sequence. NS: not specified.

7 PERSPECTIVAS

- Avaliar os níveis de expressão de *MAST4, SPRY3, USP19, TRIM67, AHCY, FAM45A, LYPLA1, TC2N* e *TGM5* em cultura celular comercial (linhagens) e culturas primárias de adenocarcinoma pâncreas e associar os níveis de expressão destes genes com a ativação das vias que estes genes potencialmente modulam.
- Silenciar, através de ferramenta lentiviral, os genes que apresentaram o maior fator prognóstico (*TRIM67, AHCY, MAST4* e *TC2N*) e avaliar a consequência deste silenciamento na proliferação, migração e resposta à gemcitabina, 5-FU e oxaliplatina.
- Avaliar o efeito do silenciamento na produção de exossomas pelas células tumorais.
- Avaliar a presença destes marcadores em exossomas isolados a partir de amostras de biópsias líquidas de pacientes com adenocarcinoma de pâncreas.
- Utilizar outras bases de dados para comparar NT *versus* TP, como GEO.
- Avaliar o fator prognóstico associado à mutações em *TP53* e *KRAS*